From Clocks to Coordinates: Controlled Human Perturbations Reveal Hidden Directions in ECG Biological Age

Jaron Mohammed

July 13, 2026

Abstract

Neural "ECG age" and organ-age models return a single scalar, collapsing biophysically distinct processes into one number. This work shows that a phase-resolved ECG-age representation contains mechanism-aligned physiological *directions* hidden by its scalar output. Four phase-of-heartbeat age clocks (P, AV, QRS, ST-T) compress into a *shared coordinate* A and a contrast-space *disagreement radius* D, with the geometry frozen on CODE-15 *before any outcome was read*. The radius D was null on its single prespecified confirmatory mortality test (hazard ratio 1.01 per SD, 95% CI 0.98 to 1.04, $p = 0.48$; $n = 74\,715$, 2583 deaths), the honest negative that motivates a signed analysis. A randomized I_{Kr} blocker (dofetilide) displaced the ST-T repolarization clock (+4.2 SD·h, FDR-adjusted $q = 0.002$), defining a covariance-scaled signed *direction* $\mathbf{w}_{I_{\mathrm{Kr}}}$, the coordinate the unsigned radius discards. A second I_{Kr} blocker (quinidine, +0.83 on the frozen axis) gave within-cohort held-mechanism support, while the active multi-channel comparators ranolazine and verapamil projected more weakly with confidence intervals spanning zero; the direction-derivation is exploratory (Holm-adjusted permutation $p = 0.069$ across the two screened perturbations). Frozen and carried into an independent external cohort (Chapman–Shaoxing/Ningbo, $n = 44\,550$, 386 QT-extension cases), the direction added *conditional* information about physician-assigned QT-interval extension beyond conventional intervals, A and D (fully adjusted odds ratio 1.23 per SD, 95% CI 1.10 to 1.36, $p = 1.9 \times 10^{-4}$; marginally null, odds ratio 1.00, an adjustment-dependent, potentially model-dependent association consistent with statistical suppression). The shared coordinate A showed a development-split mortality association (hazard ratio 1.22 per SD). At another scale, an analogous organ-system shared coordinate was mortality-associated in NHANES (+1.21 per SD); the underlying whole-ECG clock additionally reconstructed on a geographically distinct cohort ($r = 0.646$ with chronological age), whereas A itself, a residual over age-adapted gaps, correlates with age only weakly there ($r = 0.26$). At brain scale, four modality-specific AABC clocks recovered chronological age on held-out data ($r = 0.50$ to 0.59), but the prespecified brain-disagreement score showed no detectable association with subsequent four-metre walk-performance change ($\beta_D = -0.021$ per SD, 95% CI $[-0.051, +0.009]$, $p = 0.17$; $n = 295$). In a distinct open-data T1 analysis, older LEMON participants showed greater cross-view structural-aging dispersion than younger participants (Cohen's $d = 0.84$); the contrast remained positive in all 5000 balanced-group resamples (median $d = 0.60$). At anatomical scale, a volume-only multi-region CT model ($n = 1227$) achieved a cross-validated MAE of 8.56 years, with cardiovascular volumes providing the strongest system-level age estimate and exploratory structure-level residuals producing an anatomically patterned hypothesis map. These cross-scale analyses separate constructability, localization and outcome utility rather than independently confirming the electrical perturbation result. A separate exploratory evolutionary-genomics screen scored 1955 substitutions from the 312-element Zoonomia HAR catalogue and found no matched enrichment near skeletal GWAS loci ($p = 0.24$), yielding hypotheses rather than validated regulatory findings. Phase-resolved

biological-age representations can therefore be interrogated as coordinates in a state space rather than compressed to one scalar.

1 Introduction

Biological age is usually reported as one number. The central result of this study is that it is better read as a *direction* in a state space: a controlled human perturbation of a single ion channel displaces the heartbeat's electrical age along a specific, reproducible axis that a scalar age, and even an unsigned measure of subsystem disagreement, discards. That perturbation discovery, developed in full for the electrical heartbeat, is the paper's contribution; the whole-body and deep-time analyses that follow are a supporting multiscale atlas, posing related coordinate-based questions at other scales to show the stance is not an idiosyncrasy of the electrocardiogram, not independent confirmation of the electrical result. Within a single lifetime, physiological subsystems age at different rates, and the *pattern* of that disagreement, not only its average, may carry information; the electrical analysis makes this concrete and testable, and the structural (multi-region CT) and genomic (human-accelerated DNA) scales, with the skeleton as a literal bridge organ between the within-lifetime and deep-time views, use scale-appropriate shared-axis/disagreement representations under one methodological stance: rigorous, outcome-blind, open-data analysis, with no causal chain forced between scales. The electrical analysis occupies the remainder of this introduction; the atlas scales follow in Parts II and III.

Machine-learning models that predict chronological age from physiological measurements such as the 12-lead electrocardiogram, blood proteomics and brain MRI have become a standard instrument of aging research [1–4]. The gap between predicted and true age is treated as a biomarker: a positive gap means the system "looks older" than it is, and larger gaps predict disease and death. Almost universally, however, these models return a single scalar. A whole-ECG age clock compresses atrial, conduction, depolarization and repolarization physiology, processes with distinct cellular substrates and distinct disease associations, into one number, and a multi-organ age panel is typically read one organ at a time or averaged into a composite [3, 4]. That organ systems and brain-imaging views age heterogeneously, and that such heterogeneity tracks mortality and disease, is by now well established across organ-specific plasma clocks [3, 4], Mahalanobis physiological dysregulation [5], multimodal brain-age gaps [6], and composite multi-system phenotypes that already span brain, body and ECG [7]; ECG-waveform localization of aging effects has likewise been reported [8]. A scalar cannot express a hypothesis that is natural once several clocks exist: that the *pattern of disagreement* among subsystems, not only their common drift, might carry information.

That intuition is formalized geometrically. Given k subsystem age gaps measured on the same individual, standardized against a healthy reference, two coordinates are defined. The *shared axis* A is the projection onto the direction common to all subsystems, the coordinated "how old does the whole system look" signal. The *disagreement radius* D is the covariance-normalized Mahalanobis length of the residual after the shared axis is removed, an unsigned measure of *how much the subsystems disagree*, calibrated so that its scale reflects the natural covariation among clocks in health. D is a score, not a direction: it discards the sign of the disagreement and reports only its magnitude relative to the healthy covariance ellipsoid. The shared direction and contrast subspace are orthogonal by construction; the scalar summaries A and D are not themselves two orthogonal axes. Whether those summaries answer different questions is the empirical question of the paper.

The idea of separating a shared component from an orthogonal residual is not itself new: singular-value decompositions of mortality surfaces, residual scores in multi-biomarker aging panels, and orthogonalized brain-age models all exploit a version of it [5, 9]. Neither is ECG age, nor

multi-organ age, nor multimodal brain-age, new [1–4, 6, 7, 10], and the broad claim that organs and imaging views age heterogeneously is by now crowded. What has not, to the author's knowledge, been reported is (i) to train *independent* age clocks for the individual electrical phases of a single heartbeat, (ii) to compress them into a shared axis and a contrast-space disagreement radius under a geometry frozen *before any outcome is read*, (iii) to ask, across multiple cohorts and physiological perturbations, whether the shared axis and the disagreement radius carry *dissociable* information, and (iv) to recover the *signed* coordinate the unsigned radius discards as a covariance-scaled direction learned outcome-blind from a controlled ion-channel perturbation, gated for stability and frozen for a prespecified external test. That recovered perturbation *direction*, the coordinate the unsigned radius discards, frozen and shown to transport to an external clinical phenotype, not any single clock, atlas or score, is the contribution.

The dissociation is reported here under a strict outcome-blind protocol. The four phase clocks, the shared axis A, the disagreement radius D and its healthy-reference covariance were all fixed on the CODE-15 cohort without reading any mortality outcome; a four-volume physical firewall and a split-before-partition patient hash enforced the separation [11]. The frozen coordinates were then evaluated on: a locked CODE-15 development partition (mortality); the EchoNext cohort (echocardiographically confirmed structural disease) [12, 13]; a QT-prolonging drug challenge (ECGRDVQ) [14, 15] and controlled coronary occlusion (STAFF III) [16, 17] as within-person physiological perturbations; the Chapman–Ningbo cohort (geographic transfer) [18]; and, at a different biological scale entirely, NHANES organ-system aging linked to mortality [19]. The single prespecified confirmatory mortality test for the disagreement radius was executed once on the locked CODE-15 final-test partition. The shared coordinate showed consistent associations across development and secondary analyses; the electrical disagreement radius showed no detectable association with structure or acute stress across three prespecified tests *and* was null on the confirmatory mortality test; and that a mixed model on the locked split assigns larger coefficients to specific phase gaps (AV conduction and ST-T repolarization) than to their disagreement, a hypothesis an external identifiable test did not support. Biological age is not one clock but a direction in a state space.

2 Part I: The electrical witness, a shared coordinate and a contrast-space disagreement radius

Part I develops the study's methodological core in full: separately trained age clocks for the four electrical phases of the heartbeat, summarized by one shared coordinate and one radius in its orthogonal contrast subspace under a frozen, outcome-blind geometry. The supporting scales then pose analogous shared-axis/disagreement questions with scale-appropriate constructions, the brain-imaging arm reusing the same covariance-normalized Mahalanobis radius, the others adapting it, rather than instantiating one identical geometry throughout.

2.1 Two coordinates from four phase clocks

Each median heartbeat was partitioned into four temporally disjoint windows corresponding to atrial depolarization (P), atrioventricular conduction (AV, measured on the $\mathrm{P_{off}} \to \mathrm{R_{on}}$ window), ventricular depolarization (QRS) and ventricular repolarization (ST-T), and one convolutional age regressor was trained per window (Fig. 6A,B; Methods). Each clock produces a per-subsystem age gap; standardizing the four gaps against a healthy reference yields a vector $\mathbf{z} = (z_P, z_{AV}, z_{QRS}, z_{ST\text{-}T})$. A unit-norm shared-axis direction $u = \frac{1}{2}(1,1,1,1)^\top$ was fixed, defining the shared axis as its projection $A = u^\top \mathbf{z} = 2\bar{z}$ (twice the mean gap), and the disagreement radius $D = \sqrt{\mathbf{q}^\top \Sigma_q^{-1} \mathbf{q}}$ as the

covariance-normalized length of the contrast vector $\mathbf{q} = C\mathbf{z}$ that lives in the subspace orthogonal to u, where Σ_q is the contrast covariance estimated on the healthy subset used to fit the covariance ($n = 9338$; Ledoit–Wolf shrinkage; distinct from the larger $n = 9887$ healthy calibration cohort on which the standardized scores are later centred and scaled; Methods). By construction $Cu = 0$ and $CC^\top = I_3$, so D is a radius *within the contrast subspace orthogonal to the shared direction*; it carries no shared-aging component, and D^2 follows a χ^2_3 ideal under a standard-normal reference. (As a non-negative scalar length, D is not itself "orthogonal" to the scalar A; the orthogonality is between the shared direction and the contrast subspace D summarizes.) All of these quantities, including the clocks, u_0, C, Σ_q and the standardization constants, were frozen before any outcome was read (Fig. 6C). A "concordant" heart whose four gaps agree sits near the centre of the healthy ellipsoid (D small) regardless of how old it looks; a "discordant" heart in which one phase contrast dominates has a large D (Fig. 6C).

2.2 The shared axis A is consistently associated with aging across cohorts and scales

The shared axis was associated with aging and its outcomes in every setting examined (Fig. 7); these associations were consistent across development, secondary and transfer analyses, but A was not itself the single prespecified confirmatory endpoint (D was), so its incremental contribution was not adjudicated on a locked split and the models below do not isolate A's independent effect. On the locked CODE-15 development partition ($n = 61\,905$, 2292 deaths, median follow-up 3.6 years), A carried mortality risk with a hazard ratio of 1.22 per SD (95% CI 1.17 to 1.27, $p = 1.4 \times 10^{-20}$) adjusting for age and sex. In the EchoNext cohort A was associated with echocardiographically confirmed structural heart disease (odds ratio 1.38 per SD in the frozen development model). Under controlled balloon coronary occlusion in STAFF III, A rose sharply during ischemia (+0.64 SD, 95% CI 0.33 to 0.96, $p < 0.001$; whole-ECG apparent age +3.8 years, 95% CI 1.67 to 6.00, $p = 0.0008$). The frozen geometry also reconstructed without any refitting on the geographically distinct Chapman–Ningbo cohort: the whole-ECG age clock that anchors the construction transferred at correlation $r = 0.646$ with chronological age ($n = 44\,595$ with usable age; predictive $R^2 = 0.377$, computed as $1 - \mathrm{SSE/SST}$ on the transferred age scale, lower than r^2 because the untrained transfer carries a calibration slope and offset). The shared axis A is by construction a residual over age-adapted phase gaps rather than an age predictor, so its own correlation with chronological age in this cohort is low ($r = 0.26$); the transfer evidence here is therefore that the frozen clocks and geometry reconstruct out of distribution, not that A predicts age. Finally, at a completely different biological scale, an analogous shared axis built from six organ-system age gaps in NHANES carried mortality risk (hazard ratio 1.21 per SD, 95% CI 1.13 to 1.29, $p = 2.0 \times 10^{-8}$, survey-weighted; §2.8). Across mortality, structure, acute stress and organ scale, the coordinate common to all subsystems was consistently associated with aging and its outcomes, and the frozen construction additionally reconstructed on a geographically distinct cohort.

2.3 The disagreement radius D shows no detectable association with structure or response to acute stress

The hypothesis had been prespecified that the disagreement radius would respond to physiological stress and to structural disease, that stress or remodeling would drive the phase clocks apart. In three independent prespecified tests the radius showed no detectable association with either, and the nulls are reported in full (Fig. 8).

First, in EchoNext, the frozen four-phase disagreement score D carried no association with

structural heart disease: adjusted odds ratio 1.00 per SD (95% CI 0.97 to 1.04), likelihood-ratio $p = 0.89$, change in area under the ROC curve -0.0001 ($n = 9687$ final split, 3794 structural-disease events). Second, in the ECGRDVQ crossover study, a supratherapeutic dose of the QT-prolonging drug dofetilide did not move the composite radius D relative to placebo (area under the effect curve -1.88, 95% CI $[-4.56, 0.75]$, $p = 0.16$). The signed phase decomposition behind D, however, showed a phase-selective pattern consistent with the drug's known target: dofetilide is a selective blocker of the rapid delayed-rectifier potassium current I_{Kr}, which prolongs repolarization, and it displaced the ST-T repolarization clock specifically (standardized AUEC +4.16, 95% CI 2.43 to 5.86, false-discovery-rate $q = 0.002$; $n = 22$ paired crossovers) while the prespecified ventricular-depolarization (QRS) negative control did not move (+0.28, $q = 0.84$). The disagreement radius is null because it is the covariance-normalized *magnitude* of the phase-gap contrast vector: the unsigned radius discards the *direction* of a coherent one-phase displacement and retains only its magnitude, so a physiologically-directed shift of a single phase clock need not register as increased disagreement. Third, in STAFF III, acute ischemia produced no detectable change in D (inflation minus baseline +0.14 SD, 95% CI $[-0.30, 0.57]$, $p = 0.54$; recovery within the prespecified ± 1 SD equivalence margin by a two one-sided test). The disagreement radius was also reproducible under repeated measurement on placebo within the short ECGRDVQ crossover window (intraclass correlation, absolute agreement, 0.83; 95% CI 0.78 to 0.87). This argues against complete short-window instability but does not exclude measurement error as a contributor to the null results. Across repeat CODE-15 examinations, D had the lowest observed within-patient reproducibility (ICC = 0.18; §2.7). Acquisition intervals were unavailable, so this value quantifies repeat-exam reproducibility rather than stability or biological drift over a known interval.

Together, these prespecified tests bound the interpretation of D: the covariance-normalized disagreement score showed no detectable association with structural disease and no detectable change under two acute stressors. These are null tests, not demonstrated equivalence: the STAFF interval (-0.30 to +0.57 SD) still permits a moderate ischemia effect, and the drug-challenge and imaging arms had no prespecified equivalence margin, so the claim is the absence of a detectable effect, not proof of exact invariance. As §2.6 shows, acute ischemia is in fact a coherent shift *along* the shared axis, accompanied by no detectable change in the contrast-space disagreement radius. The single prespecified confirmatory mortality test, reported in §2.4, was null in the same way: on the locked CODE-15 final split the electrical disagreement radius carried no mortality information beyond age, sex, A and a whole-ECG age clock (HR = 1.01 per SD, $p = 0.48$). The electrical disagreement composite therefore showed no detectable association across the prespecified structure, acute-stress and mortality tests. This consistency motivates separate evaluation of the shared coordinate and signed phase-direction hypotheses without treating a null radius as proof that those alternatives carry the signal. A prespecified secondary sensitivity score that excluded the atrial (P) phase showed a small association with structural disease (odds ratio 1.05 per SD, 95% CI 1.02 to 1.09, $p = 0.0021$, Holm-adjusted ≈ 0.006) but no improvement in discrimination (change in AUC +0.0004, confidence interval including zero); per the frozen protocol it was not promoted to a confirmatory score, and is reported as an exploratory, phase-selective hypothesis for external testing.

2.4 The confirmatory mortality test is null

On the locked CODE-15 development partition, the disagreement radius had added a statistically significant but quantitatively modest amount of mortality information beyond age, sex, the shared axis A and a previously published whole-ECG age clock [1]: adding D improved fit by a likelihood-ratio $\chi^2 = 9.77$ ($p = 0.0018$), hazard ratio 1.049 per SD (95% CI 1.021 to 1.079). This

development association was used only to freeze the model and was never confirmatory. It already attenuated when the four phase gaps were modeled individually ($p = 0.26$), hinting that the signal was concentrated in specific phases rather than in "disagreement" as such.

The single prespecified confirmatory test resolved that question. Executed once, on the untouched final-test partition, behind the outcome firewall (Methods), the primary M2-vs-M1 Cox likelihood-ratio test for D was **null**: hazard ratio 1.010 per SD (95% CI 0.983 to 1.039), likelihood-ratio $\chi^2 = 0.50$, $p = 0.48$ ($n = 74\,715$ complete cases, 2583 deaths, median follow-up 3.4 years). The frozen success criterion ($p < 0.05$ and HR > 1) was not met. This is an informative rather than an inconclusive null: the confirmatory 95% confidence interval (ceiling 1.039 per SD) lies *below* the development point estimate (1.049), so the final split excludes a disagreement–mortality effect as large as the one development had suggested. The prespecified sensitivity scores were null in the same way (P-wave-excluded three-phase score $p = 0.20$; AV-excluded $p = 0.97$), the diagnosis-adjusted model was null ($p = 0.54$), proportional hazards held (Schoenfeld $p = 0.29$), and adding D did not improve discrimination (development-frozen concordance $0.8071 \rightarrow 0.8072$, $\Delta C = +0.0001$).

A prespecified phase-resolved decomposition (part of the frozen reveal, reported as a secondary rather than as the single confirmatory primary) generates a hypothesis about which phase gaps carry the association. In a secondary joint phase-gap model adjusted for age, sex, D and the four individual phase age gaps, but *not* the separately trained whole-ECG clock, the atrioventricular-conduction and ST-T-repolarization coefficients were positive and large (AV hazard ratio 1.24 per SD, 95% CI 1.19 to 1.29, $p = 3.1 \times 10^{-26}$; ST-T hazard ratio 1.13 per SD, 95% CI 1.08 to 1.17, $p = 1.8 \times 10^{-9}$), whereas the atrial (P) and QRS coefficients were null (both $p > 0.5$); both large coefficients recur on an internal robustness partition ($p = 3.7 \times 10^{-21}$ and $p = 2.2 \times 10^{-7}$). This parameterization must be read with care. Because the four gaps span the shared axis $A = \frac{1}{2}(z_P + z_{AV} + z_{QRS} + z_{STT})$, the model contains the shared direction, but an individual phase coefficient is not an effect at fixed A: raising one gap while holding the others constant also changes A. The positive AV and ST-T coefficients therefore *suggest* that directional, phase-specific information carries risk, but this mixed shared-plus-contrast parameterization does not establish phase-specific effects at fixed A or beyond a whole-ECG clock. The rigorous identifiable test, $M_{\text{base}} = \text{age} + \text{sex} + \text{whole-ECG age} + A$ versus $M_{\text{direction}} = M_{\text{base}} + q_1 + q_2 + q_3$ with a three-degree-of-freedom joint likelihood-ratio test, requires re-opening the CODE-15 mortality labels, which the single-use outcome firewall has now spent (Methods); it is prespecified future work. An external, explicitly post-reveal version of that test was run in the independent SaMi-Trop cohort (next paragraph), and was null. The complementary geometric reading is that D, an unsigned Mahalanobis magnitude over the contrast subspace, collapses the signed phase contrasts into a single non-negative distance and so discards exactly the directional information the phase-gap model highlights; a radial disagreement summary can compress away directional signal, which is consistent with the confirmatory D null.

To probe the phase-direction hypothesis without re-opening the spent CODE-15 firewall, the frozen geometry was carried unchanged to an independent external cohort and tested once as a declared post-reveal, exploratory analysis (protocol hash-locked before outcomes were read; Methods). In SaMi-Trop [20, 21], a Chagas-disease cohort from Brazil ($n = 1439$ after a prespecified plausibility filter, 85 deaths, median follow-up 2.1 years), the identifiable directional model was fit: $M_{\text{base}} = \text{age} + \text{sex} + \text{whole-ECG gap} + A$ versus $M_{\text{direction}} = M_{\text{base}} + q_1 + q_2 + q_3$. The three-degree-of-freedom joint likelihood-ratio test was **null** ($\chi^2 = 2.75$, $p = 0.43$); no signed contrast reached significance, and data-motivated AV-versus-others and ST-T-versus-others contrasts were null after Holm correction (both $p > 0.5$). The unsigned radius D was likewise null in this cohort (hazard ratio 1.03 per SD, $p = 0.36$). The development-grade AV/ST-T signal therefore did not replicate in an external cohort under a frozen, identifiable model. With only 85 deaths, this test is under-

powered. The confidence intervals on the signed contrasts are wide and do not exclude a moderate effect, so the analysis provides no positive external support for a transportable phase-direction effect rather than firm evidence against one; a better-powered external replication is required.

2.5 An orientation the radius discards defines a transportable perturbation direction

That the radius D showed no detectable association across the prespecified structure, acute-stress and mortality tests (§2.3, §2.4), including the single confirmatory mortality test, while the signed decomposition recovers the drug's mechanism motivates a sharper construction: if the sign carries the information the magnitude throws away, that sign should define a fixed, transportable axis rather than a per-analysis finding. In whitened coordinates $\hat{\mathbf{u}} = \Sigma_q^{-1/2}\mathbf{q}$ the radius is the length $D_{\mathrm{raw}} = \|\hat{\mathbf{u}}\|$; a signed perturbation score is the projection of $\hat{\mathbf{u}}$ onto a fixed unit axis, so $|S_{\mathrm{raw}}| \le D_{\mathrm{raw}}$ for every exam by construction (Fig. 9c). This bound is a geometric fact about the *raw* whitened coordinates S_{raw} and D_{raw}; the reported standardized scores $S_{I_{\mathrm{Kr}}}$ and D are each separately z-scored on the frozen calibration cohort, so the inequality need not hold between the standardized values. The question is whether a perturbation with known biology yields a *stable enough* axis to freeze and carry into an untouched cohort.

A covariance-scaled signed direction was learned outcome-blind from the I_{Kr} blockade contrast (dofetilide minus placebo, ECGRDVQ, $n = 22$ participants): $\mathbf{w}_{I_{\mathrm{Kr}}} = \Sigma_q^{-1}\mathbf{m}/\sqrt{\mathbf{m}^\top\Sigma_q^{-1}\mathbf{m}}$, scaled so that $\mathbf{w}^\top\Sigma_q\mathbf{w} = 1$. Its mean displacement was dominated by the ST-T-versus-QRS contrast coordinate, consistent with selective repolarization prolongation. Two candidate perturbations were screened for a stable, freezable direction, the I_{Kr} blockade contrast and, below, acute ischemia, so this direction-derivation step is treated as **exploratory** and its statistics are multiplicity-corrected across the two candidates; the confirmatory weight rests on the independent external phenotype test that follows, not on the derivation. Stability was read out on four descriptive criteria: pipeline yield, the participant-bootstrap covariance-aware cosine $\cos_\Sigma$ with the full-sample direction and its lower bound, sign stability, and a treatment-label sign-flip permutation. The I_{Kr} direction was highly stable in direction: the bootstrap median $\cos_\Sigma$ with the full-sample axis was 0.94 (95% CI 0.63 to 1.00), and 9999 of 10 000 participant bootstraps reproduced its sign without any alignment (99.99 % sign-stable; bootstrap directions were aligned only for the visualization in Fig. 9a, never for this count). Its magnitude was also non-random, though only nominally so once multiplicity is accounted for: the sign-flip permutation gave $p = 0.034$, which is $p = 0.069$ after Holm correction for the two screened directions, suggestive rather than confirmatory, and the reason the derivation is labelled exploratory. It was confirmed within the same crossover cohort by a second, held-mechanism I_{Kr} blocker, quinidine: the mean quinidine projection on the frozen direction was +0.84 (95% CI 0.52 to 1.20, bootstrap $p < 10^{-4}$), the two blockers were aligned in the whitened metric ($\cos_\Sigma = 0.94$), and a leave-one-participant-out reconstruction was highly stable (median $\cos_\Sigma = 0.997$, minimum 0.95). Two non-I_{Kr} active comparators projected weakly and with intervals including zero (ranolazine +0.21; verapamil +0.12), consistent with an I_{Kr}-dominant direction rather than a response to drug exposure in general (Fig. 9b).

The same gate was applied to acute ischemia and it *failed* (bootstrap median $\cos_\Sigma = 0.63$ with an interval spanning zero; sign stable in only 79 % of bootstraps; permutation $p = 0.84$): ischemia displaces the phase-age vector coherently along the shared axis A (§2.6), leaving no stable contrast-subspace direction to freeze. Per the frozen protocol no substitute direction was constructed, so the transportable "atlas" contains a single retrospectively derived, exploratory perturbation direction ($k = 1$), frozen before the external transport test but not itself a prespecified hypothesis. The

signed score was then standardized (mean μ, SD σ) on the frozen healthy calibration cohort, the larger $n = 9887$ cohort on which D is also standardized, not the smaller $n = 9338$ subset used to estimate Σ_q ($\mu = 0.008$, $\sigma = 1.01$, $n = 9887$).

The frozen axis was then carried, unchanged, into an independent external cohort with a clinical phenotype: the combined Chapman–Shaoxing and Ningbo 12-lead databases ($n = 44\,550$ with usable records; 386 carrying the physician-assigned QT-interval-extension label, SNOMED 111975006). The prespecified test was a one-degree-of-freedom likelihood-ratio comparison for the standardized I_{Kr} score in a logistic model of QT extension adjusting for age, sex, heart rate, QTc, QRS, whole-ECG age, A and D. The result is a *conditional* (suppression) effect, and the adjustment ladder is the point: marginally the frozen score is null (odds ratio 1.00 per SD, $p = 0.996$), it strengthens monotonically as conventional context is added (age/sex OR 1.09, $p = 0.09$; +intervals OR 1.15, $p = 0.007$), and in the fully adjusted model it is clearly non-null (OR 1.23 per SD, 95% CI 1.10 to 1.36, likelihood-ratio $p = 1.9 \times 10^{-4}$; identical under a Firth penalized fit; Fig. 10). In other words the frozen I_{Kr} direction adds information about physician-assigned QT extension *beyond* conventional intervals, the shared displacement A and the radial disagreement D; it is a shape coordinate unmasked by adjustment, not a marginally separable one. The pooled association was driven primarily by the larger Ningbo site (329 cases, OR 1.20, $p = 0.002$), which is part of the combined primary cohort and so is a site-stratified estimate, not an independent replication; Chapman–Shaoxing pointed the same way but is underpowered (57 cases, OR 1.44, 95% CI 1.06 to 1.97, reported without a pass/fail verdict). It was not, however, independent of rhythm context: it held when atrial flutter/fibrillation were excluded ($p = 0.014$) but attenuated to null when both flutter and sinus bradycardia were removed ($p = 0.48$, 86 cases), a limitation reported in full. A reliable-QT subgroup filter (QTc 350 ms to 600 ms, heart rate 40 bpm to 120 bpm) is a post-hoc robustness check, labelled as such and not prespecified. The frozen score is not claimed to be *stronger* than QTc, A or D: its incremental information is real (bootstrap incremental likelihood-ratio CI excludes zero) but is not statistically distinguishable from those scalars. The defensible claim is narrow and external: a direction frozen from a randomized drug perturbation carries conditional information about an independent clinical QT phenotype, beyond conventional measurements.

2.6 Acute ischemia is a coherent shift of the shared axis, not a loss of coordination

STAFF III records the 12-lead ECG before, during and after elective balloon coronary occlusion, providing a within-person model of acute, reversible ischemia (91 occlusions in 70 patients). This experiment resolves *why* the disagreement radius shows no detectable change under stress (Fig. 11). During occlusion all four phase clocks moved in the same direction by similar amounts: the AV, QRS and ST-T gaps rose significantly (AV +0.52 SD, false-discovery-rate $q = 0.009$; QRS +0.49, $q = 0.004$; ST-T +0.47, $q = 0.037$) and the atrial gap rose non-significantly (+0.40, $q = 0.20$). A coordinated positive displacement of all four clocks is, by construction, a shift *along* the shared axis A (which duly rose +0.64 SD) and *not* a change in their disagreement, accompanied by no detectable change in the contrast-space radius D. The apparent phase-clock displacement persisted into the 30 to 40-second reperfusion window (A recovery minus baseline +0.59 SD, $p < 0.001$), while the recovery-versus-baseline change in D met the prespecified equivalence criterion within the ± 1.0 SD margin (0.00 SD, two one-sided tests, 90 % CI -0.32 to 0.31); this does not establish literal equality to zero. This is, to be clear, an *apparent* displacement of an age-trained readout under acute ischemia, not literal biological aging over minutes. The observed pattern is predominantly shared-axis displacement accompanied by no detectable change in the disagreement radius; it does not establish a mechanism.

2.7 The electrical construct is internally valid, externally transferable and differs in reproducibility

The analyses in this subsection were computed *after* the single prespecified confirmatory test and are reported as post-hoc; each is either a re-presentation of the already-locked mortality numbers or a no-outcome computation on frozen scores, so none re-opens the confirmatory endpoint. All ran through the frozen scoring pipeline with no refitting (geometry hash `5a3d8800...ee8e3e0`; Methods).

First, the phase clocks encode more than the classical interval durations they might be suspected of merely re-reading. After adjusting for age and sex, each predicted phase age was only weakly correlated with its matched standard interval (partial Spearman ρ, CODE-15 calibration set, $n = 74\,715$): P-clock versus P-duration $\rho = -0.11$, AV-clock versus PR-interval $\rho = -0.13$, ST-T-clock versus QTc $\rho = +0.08$, and only the QRS-clock read its interval appreciably ($\rho = +0.20$). The AV clock, which carried the largest development-split mortality coefficient, is, if anything, *weakly negatively* related to the PR interval, so what it contributes is not a re-badging of classical PR prolongation (Supplementary Fig. S8b).

Second, the frozen geometry reconstructs on an untrained external cohort. Applying the fixed standardization, contrast basis and disagreement metric to Chapman–Shaoxing/Ningbo (a 12-lead cohort used to train no clock and fit no calibration constant), the whole-ECG age clock transferred at $r = 0.64$ and all four phase clocks transferred positively ($n = 44\,595$ with usable age; QRS $r = 0.57$, P $r = 0.52$, ST-T $r = 0.46$, AV $r = 0.34$). On the subset entering the disagreement-geometry analysis ($n = 44\,550$ after D-score quality control), the shared axis and disagreement radius were empirically nearly uncorrelated ($r = -0.17$). This is descriptive; mathematical orthogonality applies to the shared direction and contrast subspace, not to the scalar summaries themselves. External disagreement was lowest in atrial-fibrillation/flutter records, consistent with delineation-driven behavior rather than a coordination signal. These cohorts carry no mortality follow-up comparable to CODE-15, so this establishes generality of the age-structure geometry only, not an external mortality replication (Supplementary Fig. S8c).

Third, the equal-weight shared axis is close to the direction the data would have chosen. The prespecified axis uses fixed weights $u = \frac{1}{2}(1, 1, 1, 1)$; a data-driven first principal component of the four calibration z-scores (development split, $n = 61\,905$; 51.5% of variance) has weights $(0.45, 0.51, 0.51, 0.53)$, cosine similarity 0.998 with the fixed axis, and per-exam scores correlating at $r = 0.9995$. The choice made before any outcome was read is essentially the empirical principal direction, so the axis is both interpretable and data-consistent.

Finally, the coordinates differ sharply in within-patient reproducibility. Across 66 474 CODE-15 patients with ≥ 2 QC-passing ECGs (177 336 exams), the composite disagreement radius D was the *least* reproducible coordinate the geometry defines (test–retest ICC$(1, 1) = 0.18$), well below the shared axis A (0.41) and every single-phase gap (QRS 0.47, ST-T 0.38, P 0.32, AV 0.31). A Mahalanobis radius over a mixture of signal and noise channels compounds measurement error, so this is a mechanistic complement to the null: D is not only uninformative for mortality, it is the noisiest coordinate defined. The AV gap carried a large development-split coefficient despite only moderate reproducibility (ICC = 0.31), which is consistent with a reproducible within-patient measurement rather than a reliability artifact, though the external identifiable test did not support it. CODE-15 lacks per-exam acquisition dates, so this quantifies within-patient measurement stability, not calendar drift of the coordinate; longitudinal trajectory modeling of A requires dated repeat exams and is left to future work (Supplementary Fig. S8d).

2.8 The same multiscale question extends across the body

If the shared-axis/disagreement-radius decomposition captures something general about biological aging rather than an idiosyncrasy of the electrocardiogram, the same construction should behave similarly at a completely different scale. An analogous BodyVector was built from six organ-system age gaps in NHANES 2005–2010 linked to 2019 mortality (14 362 participants, 1739 deaths, 160 493 person-years, median follow-up 11.3 years), using survey-weighted Cox regression with cross-fitted clocks (Fig. 12; Methods). This analysis used the same geometric construction but produced a different cross-scale pattern. The shared body axis A_{body} carried mortality (hazard ratio 1.21 per SD, 95% CI 1.13 to 1.29, $p = 2.0 \times 10^{-8}$), and the body disagreement radius D_{body} added mortality information beyond it (hazard ratio 1.21 per SD, 95% CI 1.108 to 1.328, $p = 3.1 \times 10^{-5}$; cross-validated concordance improvement +0.0041, 95% CI 0.0029 to 0.0054).[1] The concordance ladder rose from 0.8147 (age and sex) to 0.8288 (adding A_{body}) to 0.8329 (adding D_{body}). The disagreement signal was robust to leaving out any single organ system (all six design-based $p < 0.05$). This NHANES analysis is reported as a secondary, cross-scale arm of the A/D construction: its mortality outcome was inspected in a prior analysis, so cross-fitting makes the clock construction leakage-free but does not make the arm outcome-naive, and it is not a substitute for, nor a rescue of, the confirmatory CODE-15 test, which for the electrical disagreement radius was null (§2.4). What it does show is that an organ-system disagreement radius behaves differently from the electrical one, a cross-scale contrast revisited in the Discussion. The broader observation is that the same two-coordinate construction can be tested at multiple biological resolutions without requiring the coordinates to behave identically.

2.9 Brain imaging separates constructability from prospective utility

The cross-scale analysis asked two distinct questions: whether a multimodal brain-aging decomposition can be constructed, and whether its disagreement coordinate forecasts subsequent functional change. The Aging Brain Cohort (AABC / HCP-Aging Release 2) contained 1396 adults and 2878 study visits; 2469 visits had complete predictions from all four imaging channels and entered the frozen geometry. Separately trained clocks represented cortical structure, T1w/T2w myelin, arterial-spin-labeling perfusion and resting-state functional amplitude. Each recovered chronological age on held-out data (Pearson $r = 0.59, 0.51, 0.51,$ and 0.50), clearing the prespecified $r \geq 0.20$ gate. Across complete visits, the scalar shared axis and disagreement radius were empirically nearly uncorrelated ($r = -0.03$; Fig. 1). This empirical correlation is descriptive: mathematical orthogonality applies to the shared direction and contrast subspace, not necessarily to the two scalar summaries.

The prespecified primary test found no detectable association between baseline brain-channel disagreement and subsequent annualized change in four-metre walk performance after adjustment for the shared axis, baseline performance, age, sex, education, site, height and BMI: $\beta_D = -0.021$ per SD (95% CI $[-0.051, +0.009]$, $p = 0.17$; $n = 295$ longitudinal holdout participants). The cross-sectional secondary test was also null ($p = 0.78$), as were the secondary motor and cognitive outcomes after false-discovery correction and the mandatory alternative constructions. A single planted-effect fixture and a single null fixture served only as code sanity checks, not as power

[1] The D_{body} interval 1.108 to 1.328 is read directly from the fitted design-based survey Cox model (`svycoxph`, combined-cycle MEC weights, robust standard errors, design degrees of freedom 47; $n = 14\,362$, 1739 deaths, 93 clusters), which is the appropriate model for the NHANES Linked Mortality File. Hazard ratio, robust standard error, interval and p are taken verbatim from the released canonical model table (`results/act2_nhanes/nhanes_canonical_cox_table.csv`), not reconstructed in prose; a naïve non-survey Cox model yields a narrower nominal interval that is not used here.

or type-I-error calibration. Thus the four-channel representation was technically viable, but its disagreement radius was not established as a prospective marker of gait decline in this cohort.

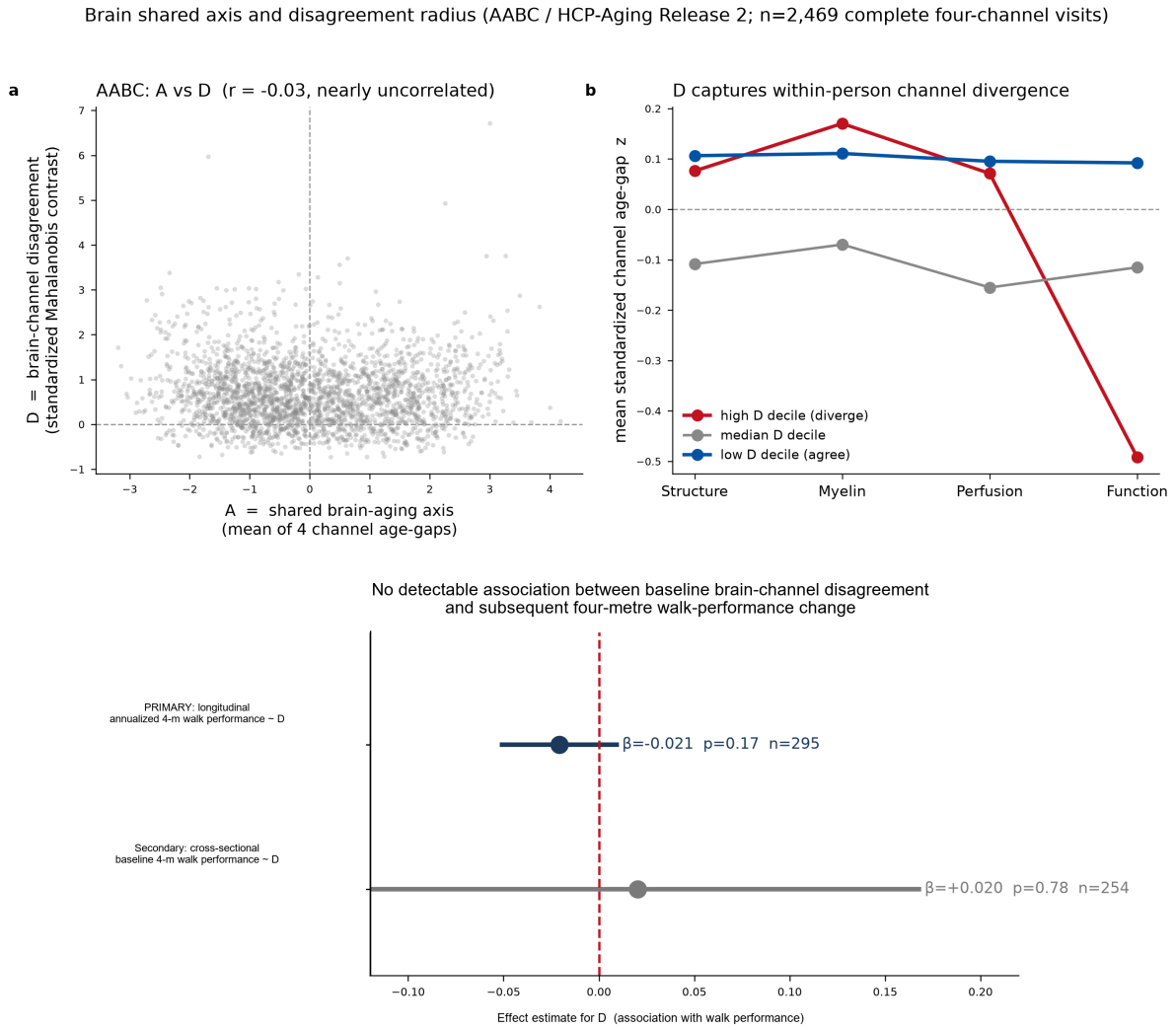

Figure 1: A multimodal brain-aging geometry was constructible, but its prespecified functional test was null (AABC Release 2). Top: Across 2469 complete four-channel visits, the shared-axis and disagreement summaries were empirically nearly uncorrelated ($r = -0.03$). This is descriptive: mathematical orthogonality applies to the shared direction and contrast subspace, not necessarily to scalar A and scalar D. **Bottom:** Baseline brain-channel disagreement showed no detectable association with subsequent four-metre walk-performance change in the longitudinal holdout ($n = 295$; $\beta_D = -0.021$ per SD, 95% CI $[-0.051, +0.009]$, $p = 0.17$). The cross-sectional secondary analysis was also null. These estimates establish neither equivalence nor absence of every clinically relevant effect.

A separate open-data analysis addressed a different construct question using raw anatomy. All 220 archived LEMON MP2RAGE scans were processed into four complementary T1-derived age-sensitive views: deep image-derived brain age, gray-matter fraction, ventricle–brain ratio and brain-tissue fraction. Because these views come from one acquisition and are correlated, their within-person standard deviation is denoted S_{T1} rather than D; it is a cross-view dispersion score, not the covariance-normalized AABC disagreement radius. Under pooled-cohort standardization, mean S_{T1} was 0.517 in older participants and 0.340 in younger participants (difference +0.177,

bootstrap 95% CI $[0.112, 0.242]$; one-sided Mann–Whitney $p = 4.3 \times 10^{-8}$; Cohen's $d = 0.84$). The contrast survived label permutation, removal of any one view and adjustment for absolute overall old-appearance.

Because the groups were unequal in size, an additional imbalance diagnostic repeatedly sampled 69 younger participants to match the 69 older participants and recomputed every view's standardization. The older–younger contrast remained positive in all 5000 resamples, with median Cohen's $d = 0.60$ (95% resampling interval 0.44 to 0.78). Equal-group scaling gave $d = 0.62$ ($p = 5.5 \times 10^{-5}$). By contrast, standardizing each age group separately removed and modestly reversed the difference. The appropriate interpretation is therefore that the four T1-derived views shift by different amounts between younger and older groups; this analysis does not establish excess age-residualized disorganization within older individuals or prospective functional utility. The anatomical result complements, but does not overturn, the prospective AABC functional null (Figs. 2 and 3).

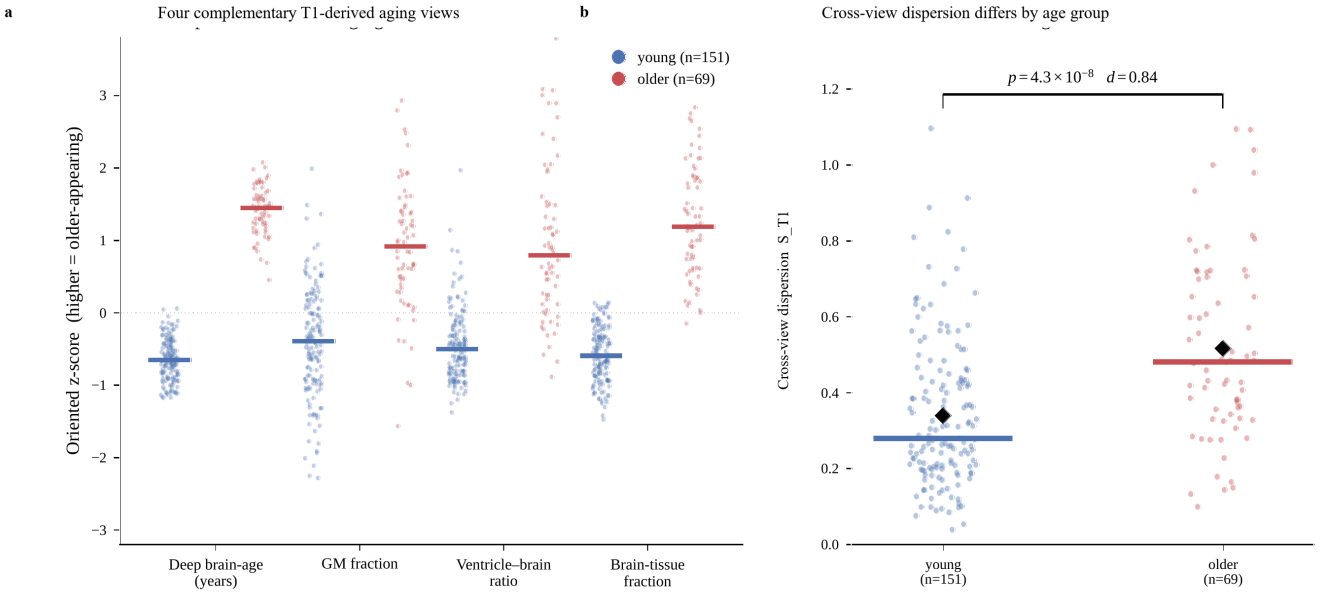

Figure 2: Cross-view T1-derived structural-aging dispersion differs between age groups (LEMON). All 220 open-access MP2RAGE scans were processed into four complementary, correlated image-derived views. **(a)** Each view was oriented so that higher values represent older appearance and standardized across the pooled cohort. **(b)** Cross-view dispersion S_{T1} was greater in older than younger participants (0.517 versus 0.340; difference +0.177, bootstrap 95% CI $[0.112, 0.242]$; Cohen's $d = 0.84$). The contrast survived permutation and leave-one-view-out analyses and remained positive under balanced-group resampling. Because within-group standardization removed the contrast, the result is interpreted as differential age-group shifts across T1-derived views, not as validated age-residualized dysregulation or prospective functional prediction.

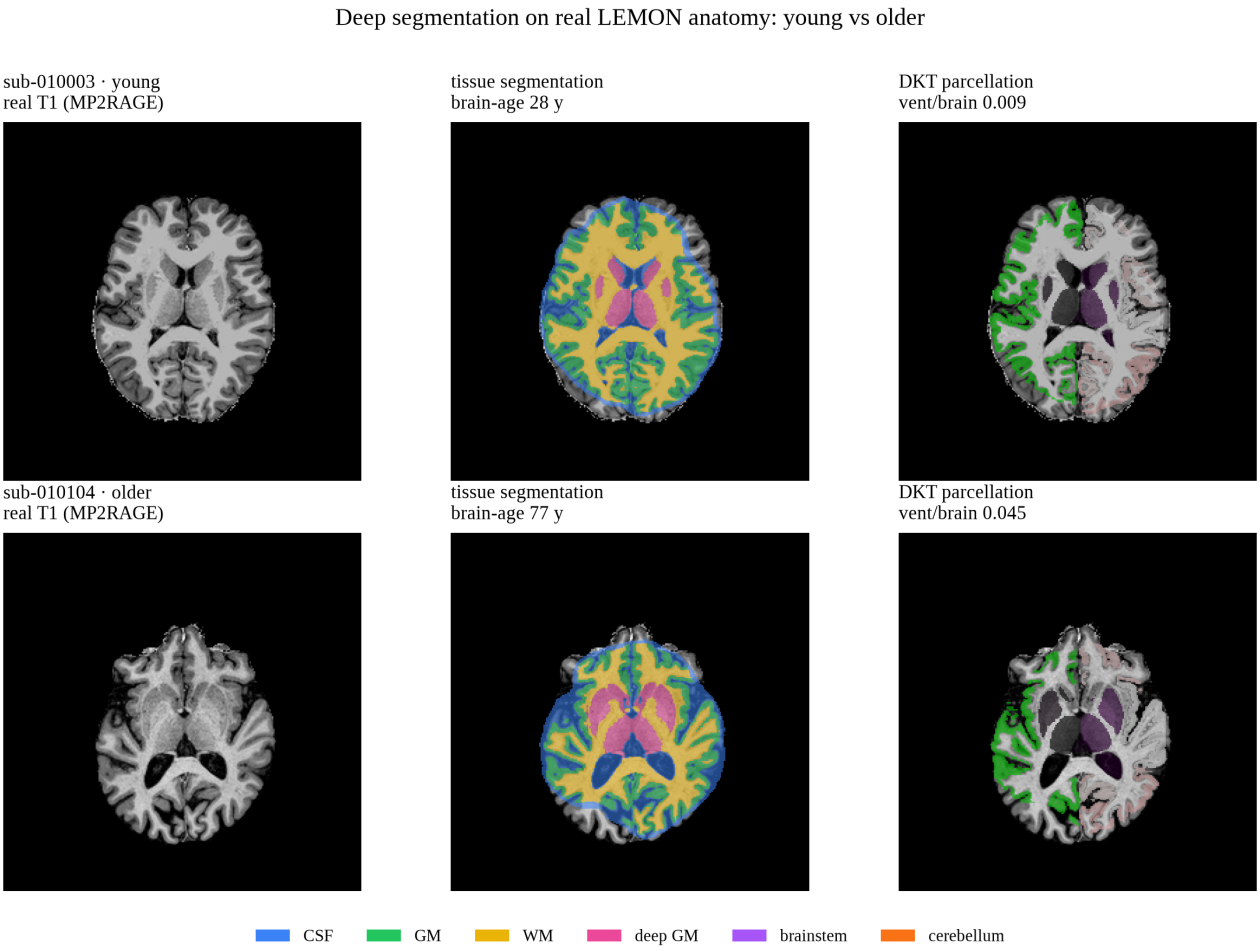

Figure 3: Illustrative anatomy underlying the T1-derived views. Open-access MP2RAGE examples from one younger and one older LEMON participant are shown as raw T1, six-tissue segmentation and DKT parcellation. The older example has a larger ventricle–brain ratio (0.045 versus 0.009) and more prominent sulci. These examples illustrate image features used to derive the quantitative views; they were not used to select, validate or causally explain the cohort-level dispersion result.

3 Part II: Anatomical localization in a heterogeneous multi-region CT atlas

Part II asks a different question from the electrical analysis: where does a volume-based age estimator obtain information across anatomy? It does not apply the shared-axis/disagreement or signed-direction geometry of Part I and is not an independent confirmation of the electrical result. Instead, it provides a secondary anatomical-localization analysis across a whole-body label set measured in heterogeneous routine-care CT examinations.

3.1 A volume-only multi-region CT age model

Across 1227 TotalSegmentator v2 examinations with recorded chronological age [22], acquired on 20 scanner models from three manufacturers at ten contributing institutions, a volume-only model using approximately 115 segmented structures attained a five-fold cross-validated mean absolute

error of 8.56 years (95 % CI 8.14 to 8.97) and an R^2 of 0.432 (CI 0.381 to 0.479). The examinations span thoracic, abdominal, pelvic, neck, angiographic and polytrauma protocols rather than a uniform whole-body acquisition. Accordingly, this estimate describes performance in a heterogeneous multi-region collection and may contain information from both anatomy and acquisition coverage. The model is volume-only: it includes no intensity-derived bone-density, muscle-attenuation or fat-attenuation feature.

When the same modeling procedure was restricted to prespecified anatomical systems, the cardiovascular subset provided the strongest age estimate on volume alone (MAE 9.71 years, $R^2 = 0.288$; 18 structures). The skeletal and digestive subsets each achieved an MAE of 11.20 years ($R^2 = 0.111$ and 0.107, respectively). These comparisons rank predictive performance under unequal feature counts and scan coverage; they do not estimate or compare biological rates of aging.

3.2 The fitted model relies strongly on aortic volume

Permutation analysis identified aortic volume as the feature with the largest effect on the fitted model's error: shuffling it increased MAE by 2.87 years, compared with 0.73 years for the right subclavian artery, 0.63 years for the left iliac artery and 0.49 years for the brachiocephalic trunk (Fig. 4A). Cardiovascular structures were prominent throughout the ranking. This model-reliance ranking differed from the univariate volume–age ranking: aortic volume had a modest Spearman correlation with age ($\rho = 0.30$), whereas the brachiocephalic trunk, right subclavian artery and left iliac artery had correlations of 0.43, 0.40 and 0.34, respectively. Thus multivariable model reliance and univariate age association capture different properties. Because structure volumes and field-of-view indicators are correlated, permutation importance cannot establish an independent biological contribution or show that a structure ages fastest.

3.3 Exploratory structure-gap differences across pathology labels

In the exploratory structure-level analysis, 12 structure–label comparisons met Benjamini–Hochberg FDR < 0.05 (Fig. 4B). Several of the largest sample effects involved vascular labels and large arteries: right iliac artery, Cohen's $d = 0.66$; left iliac artery, $d = 0.596$; brachiocephalic trunk, $d = 0.47$; and aorta, $d = 0.39$, $p_{\mathrm{FDR}} = 0.0016$. Right-humerus gap also differed by trauma label ($d = 0.60$). The retained set included both anatomically plausible and less-specific pairings. These comparisons were not externally validated or matched for sex, body size, anatomical coverage, site or clinical indication; they therefore generate localization hypotheses rather than establish pathology specificity or organ-specific clinical meaning.

3.4 A descriptive spatial map of bone-volume associations

At group level, positive cross-sectional volume–age associations were largest for the femur ($\rho = 0.261$), pelvis ($\rho = 0.215$), lumbar spine ($\rho = 0.175$), sacrum ($\rho = 0.172$) and thoracic spine ($\rho = 0.145$), whereas the humerus ($\rho = 0.024$), cervical spine ($\rho = 0.017$) and shoulder girdle ($\rho = 0.017$) were near zero (Fig. 5). This apparent lower-body and axial pattern is shown as a descriptive anatomical hypothesis map. It is based on segmented volume, not bone density or mineral content, and is unadjusted for sex, body size, anatomical coverage, site and clinical indication. It therefore neither estimates longitudinal bone loss nor demonstrates that particular bones age faster.

The group-level cohort statistics were painted onto the segmented anatomy of reference examination s1045 for visualization. The colors do not represent disease, aging or a trajectory in that

individual. The skeleton provides a visual bridge to Part III, not evidence of a causal connection between the CT and evolutionary-genomics analyses.

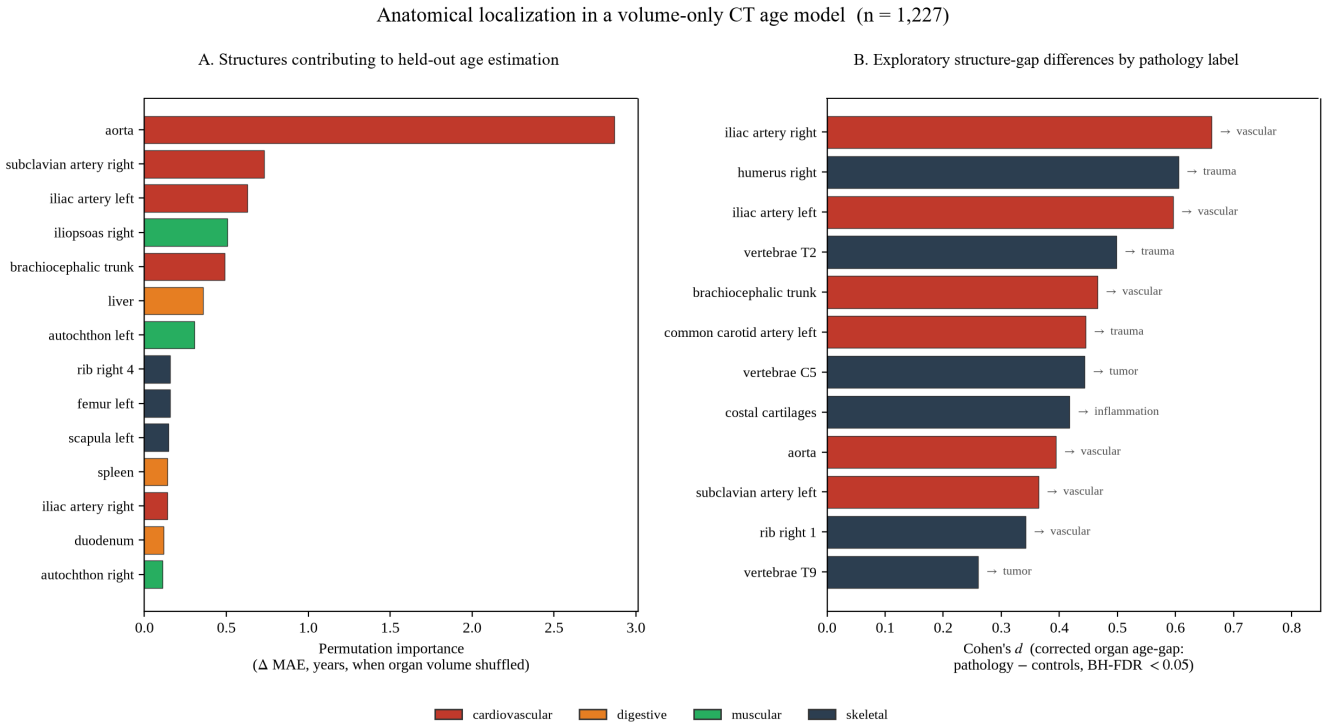

Figure 4: Anatomical localization in a volume-only CT age model (n = 1227). **(A)** Structures ranked by fitted-model reliance, measured as the increase in cross-validated MAE after permuting one volume feature. Aortic volume produced the largest single-feature effect (2.87 years). Permutation effects describe this model's reliance under correlated anatomy and heterogeneous scan coverage; they do not identify the fastest-aging structure or an independent biological contribution. **(B)** Twelve exploratory structure–pathology-label comparisons retained after Benjamini–Hochberg correction. Bars show unadjusted sample Cohen's d. The comparisons were not externally validated or matched for anatomical coverage, site or clinical indication and are presented as localization hypotheses rather than pathology-specific biomarkers.

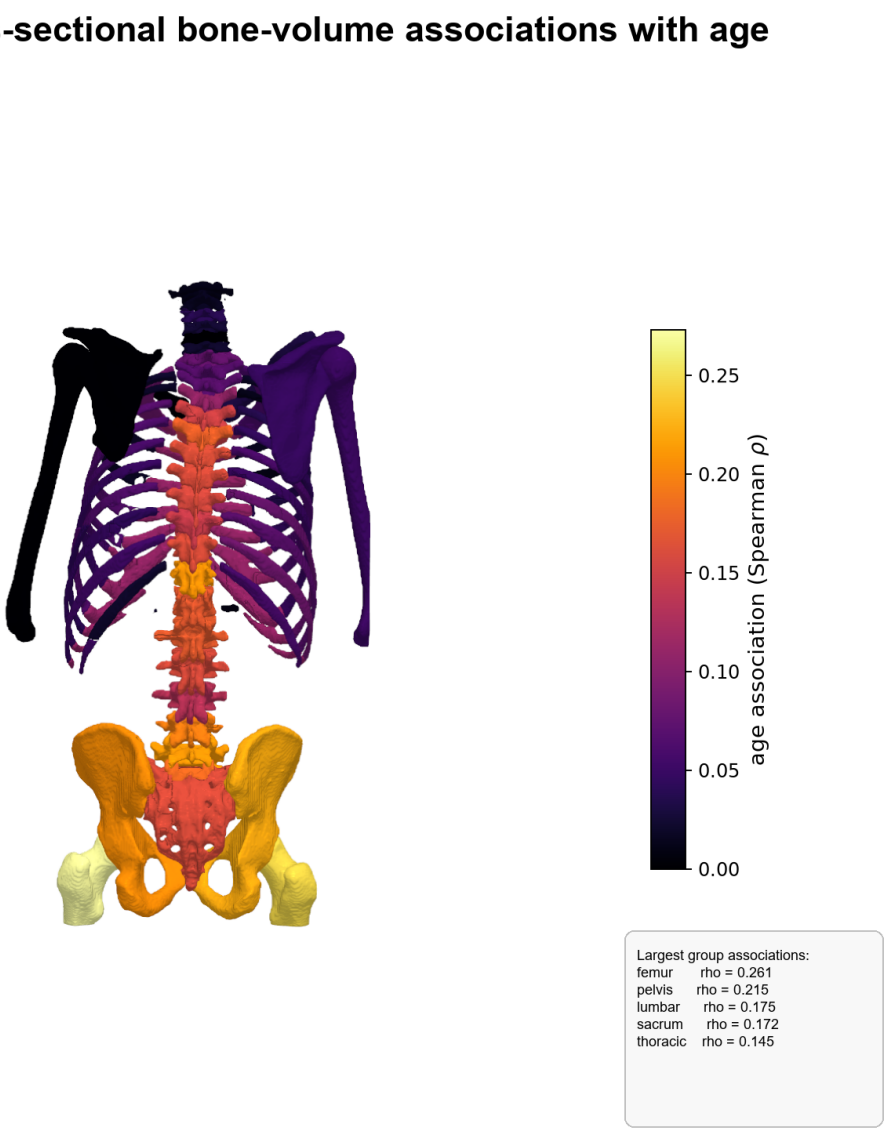

Figure 5: Group-level cross-sectional bone-volume associations with chronological age. Spearman correlations between age and segmented volume for prespecified skeletal groups are painted onto the real TotalSegmentator v2 anatomy of reference examination s1045. The largest positive group-level correlations were femur ($\rho = 0.261$), pelvis ($\rho = 0.215$), lumbar spine ($\rho = 0.175$), sacrum ($\rho = 0.172$) and thoracic spine ($\rho = 0.145$); shoulder-girdle and cervical-spine correlations were approximately 0.017. The colors are cohort-level statistics, not measurements of aging or disease in s1045. These volume associations are cross-sectional and unadjusted for sex, body size, anatomical coverage, site and clinical indication; they are not bone-density estimates or longitudinal aging rates.

4 Part III: Supporting evolutionary-genomics screen of predicted skeletal-lineage effects in the Zoonomia HAR catalogue

This genomic scale of the supporting atlas reads the body across the deepest timescale in this study, where the human genome changed fastest, rather than which subsystems age fastest. It is evolutionary genomics, not an aging clock, and shares with Parts I and II only a theme and a method, never a fitted causal path, and is not independent confirmation of the electrical result (see Synthesis).

Human Accelerated Regions (HARs) are conserved sequences that acquired an excess of substitutions on the human lineage after divergence from chimpanzee, and they are strongly enriched for gene-regulatory function. Functional dissection of HARs has to date concentrated on neural contexts; the skeleton, the literal bridge organ of this manuscript, has not been screened at the level of individual HAR substitutions. As a supporting evolutionary-genomics analysis, all 1955 human-specific substitutions represented in the 312-element Zoonomia HAR catalogue [23] (mean 6.27 per HAR, maximum 38) were scored using a sequence-to-coverage model. The analysis asks whether model-predicted regulatory effects can prioritise skeletal-lineage hypotheses. It is not an aging clock, an experimental regulatory assay, or independent confirmation of the ECG result; the candidates below are prioritised hypotheses awaiting experimental validation.

A descriptive skeletal-minus-neural coverage contrast. For each substitution, Borzoi-predicted [24] reference-to-alternate coverage changes were aggregated across a prespecified set of 16 skeletal-lineage accessibility tracks, predicted chromatin accessibility in osteoblast, mesenchymal-stem-cell, embryonic arm- and leg-bone, limb and skeletal-muscle contexts, and 81 neural accessibility tracks. Their difference, termed the skeletal-minus-neural predicted-coverage contrast, is a *descriptive model output*. It does not establish tissue specificity, enhancer activity, or causal regulatory function, and is not a measured accessibility effect, a demonstrated skeletal-regulatory effect, a validated functional variant, or a biological discovery. Across the substitution catalogue the contrast spans -2.45 to 4.69 (predicted skeletal-loss to skeletal-gain), a heavy-tailed distribution centred near zero (Fig. 13A).

Evolutionary constraint and a gBGC-susceptibility flag. Two filters separate candidate regulatory substitutions from the confounded majority. First, per-base constraint from the Zoonomia 241-way mammalian alignment [25]: 1208 substitutions (61.8 %) fall at positions with phyloP $>$ 2.27, and 1463 (74.8 %) at phyloP > 1.6. Second, a GC-biased gene conversion (gBGC) susceptibility flag [26]. Because gBGC over-transmits strong (G/C) over weak (A/T) alleles in recombination hotspots, weak-to-strong substitutions can mimic the substitution excess that defines a HAR without any positive selection. Because allelic direction alone cannot establish that gBGC occurred, weak-to-strong substitutions were treated as gBGC-susceptible rather than classified as artifacts; recombination context and direct evolutionary evidence would be needed to attribute any individual substitution to gBGC. Classifying every substitution by allelic direction gives weak-to-strong 936 (48 %), strong-to-weak 688 (35 %) and neutral 331 (17 %) (Fig. 13B). Requiring both constraint and a substitution not flagged by the allelic-direction rule leaves 722 candidate skeletal-regulatory substitutions.

The central genomic result: no enrichment near skeletal GWAS loci. The 722 candidates were intersected with GWAS credible-set variants for two skeletal phenotypes, osteoarthritis (OA; 936 variants) and bone mineral density (BMD; 903 variants). The strongest inferential result of this

arm is not the candidate list; it is a null. Ten HARs were located within 25 kb of an OA or BMD credible-set variant, but none overlapped at the exact base (0 exact-base overlaps). Their predicted skeletal-lineage contrasts were not enriched relative to matched HARs: a matched permutation null (50 000 draws, per-HAR skeletal-specificity taken as the maximum absolute skeletal-minus-neural score over each region's candidate substitutions, matched on candidate-substitution count) placed the ten proximal HARs' mean score (0.220) only marginally above the matched-null expectation (0.179), with a permutation $p = 0.24$ (and $p = 0.26$–0.28 for region-length-matched and unmatched nulls). The screen therefore generated ranked hypotheses without demonstrating disease-locus enrichment.

A blind GDF5 filter-logic control. Before any score was computed, three substitutions at the *GDF5* growth-and-arthritis locus were frozen as a blind control with pre-specified pass/reject calls on the constraint-and-gBGC filters. The GDF5 substitutions were prespecified as filter-logic controls: the two constrained, strong-to-weak variants (GROW1 enhancer rs4911178, chr20, phyloP 7.04; *GDF5* 5′UTR rs143384, phyloP 4.96) passed, while the unconstrained, gBGC-favoured promoter variant rs6060369 (phyloP −0.07, weak-to-strong) failed, exactly as prespecified (Fig. 13A). The two surviving variants are strong-to-weak and highly constrained, the signature expected for the derived alleles shown to be under ancient selection at this locus [27]. No signed Borzoi effect was prespecified, and these controls do not validate the magnitude or direction of the predicted accessibility contrast, indeed one retained control has a near-zero skeletal-minus-neural score. The frozen quantity is the constraint-and-gBGC filter outcome only.

Future experimental prioritisation. Separately, and several inferential steps removed from the HAR screen itself, a genetic-score × tractability composite over skeletal genes is provided in the supplement as a forward-looking prioritisation aid; it is not part of the HAR result and is not claimed to rediscover drug targets. It makes no claim that HAR substitutions cause any age-related skeletal phenotype.

To the author's knowledge, this analysis applies one prespecified scoring workflow to all human-specific substitutions represented in the selected 312-element Zoonomia HAR catalogue; the concentration of prior HAR functional work on neural contexts, leaving the skeletal question comparatively unexamined, is empirically the case.

5 Cross-scale synthesis

One electrical result and a supporting multiscale atlas motivate a common question: when does a structured age representation contain information that a single scalar obscures? The electrical analysis is the load-bearing finding. It resolves phase-age output into a shared coordinate and a radius in its contrast subspace; the radius showed no detectable association across the prespecified tests, including the single confirmatory mortality test. A separate exploratory projection, anchored to a controlled perturbation, retains one orientation within that contrast space and carries conditional information in an external phenotype analysis. The supporting arms do not instantiate one identical geometry or establish a causal chain. NHANES tests blood-biomarker system coordination; AABC separates constructability from prospective outcome utility; LEMON tests a related raw-image dispersion construct; CT localizes the anatomical information used by a volume-age model; movement probes a technically limited structure–function boundary; and the genomic screen reports a null enrichment while prioritizing experimental hypotheses.

The unification is deliberately *thematic and methodological, not causal*. The skeleton is a visual bridge *organ*, Part II maps unadjusted cross-sectional bone-volume associations within a lifetime, whereas Part III probes how skeletal regulatory DNA evolved across species, but no claim is made, and none is supported, that the Human Accelerated Region substitutions of Part III cause the bone-aging pattern of Part II. These analyses use different cohorts, different species-scales and different readouts; the value is in the shared analytic stance applied honestly across all three, not in a path fitted between them. A forward-looking, non-causal target-prioritisation aid connecting the structural and genomic arms is provided in the supplement; it is hypothesis-generating only and infers no mechanism.

Finally, the boundary between what was reachable and what was not is itself a result. Three modalities of open data, including hundreds of thousands of ECGs, more than a thousand multi-region CTs, and the complete Human Accelerated Region catalogue with mammalian-alignment constraint, were analyzed end to end, and the frozen perturbation direction was carried into a fourth open cohort (Chapman–Shaoxing/Ningbo) for an external phenotype test, whereas gated resources (biobank-scale imaging, linked-mortality intensive-care records) remained out of reach within the build window. That open-and-instant versus gated-and-slow divide bounds what augmented, multi-scale analysis can currently reach.

6 Discussion

This study tested whether a phase-resolved ECG-age representation can be interrogated as coordinates rather than compressed to one scalar, then posed analogous questions at organ-system, imaging and anatomical scales. Across development and secondary analyses, the electrical shared coordinate A showed associations with mortality, echocardiographically confirmed structural disease and acute ischemia. The frozen clocks and geometry also reconstructed across geography, although A itself correlated only weakly with chronological age in that cohort. The electrical disagreement radius D behaved differently: it showed no detectable association with structural disease, no detectable change under two acute physiological stressors and, on the single prespecified confirmatory test, no incremental mortality information beyond the shared coordinate and a whole-ECG age clock. These results support neither a universal positive disagreement biomarker nor proof of literal invariance.

When the single disagreement scalar is replaced with the four individual phase age gaps, the residual mortality signal appears concentrated: the atrioventricular-conduction and ST-T-repolarization coefficients are positive and large, while the P and QRS coefficients are null. This is a prespecified secondary decomposition, not the confirmatory primary, and it must be read cautiously, the model conditions on the shared axis only implicitly (the four gaps span it) but not on the separately trained whole-ECG clock, and because raising one gap at fixed others also moves A, its coefficients do not isolate phase-specific effects at fixed A. The rigorous identifiable test needs the spent CODE-15 firewall re-opened; run instead as a declared post-reveal analysis in the external SaMi-Trop cohort, it was null, so the phase-direction hypothesis is not yet externally supported. The complementary geometric reading is that D is an unsigned Mahalanobis radius, and taking that non-negative magnitude over a subspace whose signed directions are mixed discards the directional information the phase-gap model highlights. A radial disagreement summary may therefore compress away directional signal, a more precise reading than treating D as a simple average of its channels. The signed perturbation analysis makes this concrete and turns it into a usable instrument: a controlled I_{Kr} blockade defines a specific, covariance-scaled direction in the same contrast subspace. The direction passed its outcome-blind stability gate, received within-cohort

held-mechanism support from a second I_{Kr} blocker and, carried unchanged into an independent external cohort, adds conditional information about a clinical QT-extension phenotype beyond conventional intervals. Acute ischemia, by contrast, produced no stable contrast direction and a phase-clock pattern predominantly aligned with the shared coordinate, accompanied by no detectable change in the disagreement radius. The evidence therefore motivates further study of the shared coordinate and specific signed phase directions, while the tested data do not support the disagreement composite as an independent predictor of the studied outcomes.

Against that backdrop the organ-system result is a genuine cross-scale contrast rather than a confirmation. In NHANES an analogous organ-system disagreement radius did carry a secondary, non-confirmatory mortality association beyond its shared axis ($p = 3.1 \times 10^{-5}$), with a small cross-validated concordance gain (+0.0041). This should not be over-read: it is a secondary arm on a cohort whose outcome had been inspected in earlier work, and it is not a rescue of the electrical confirmatory null. But it does suggest that "disagreement" is not one thing, a covariance-normalized residual among slow organ-system clocks may behave differently from one among fast electrical-phase clocks, and that where a disagreement coordinate is informative is itself an empirical question, to be settled by prespecified tests rather than assumed. A prospective brain-imaging arm makes the same point in the opposite direction: applying the same mathematical construction, independently calibrated in the brain cohort, to four separately trained, channel-specific clocks across complementary imaging modalities of brain aging (structure, myelin, perfusion, function) and a longitudinal gait outcome, under a hash-locked outcome-blind protocol, found no detectable association between brain-channel disagreement and prospective walk-performance change. In LEMON, a different and simpler quantity constructed directly from raw T1 images, namely cross-view structural-aging dispersion S_{T1}, was greater in older than younger participants ($d = 0.84$), and the contrast remained positive under balanced resampling. Within-group standardization removed that difference. The defensible interpretation is therefore nonuniform age-group shifts across complementary T1-derived views, not validation of the AABC radius, age-residualized tissue disorganization or functional prognosis.

The CT atlas contributes anatomical localization rather than another test of the shared/disagreement geometry. A volume-only model recovered chronological-age information from 1227 heterogeneous routine-care examinations; the cardiovascular subset supplied the strongest system-level estimate and aortic volume produced the largest single-feature permutation effect. Exploratory gap comparisons also placed several of the larger vascular-label differences in major arteries. These observations identify where this fitted model relied on anatomy and where follow-up hypotheses arose; they do not show that the aorta or any bone ages fastest. Forty percent of structure-by-scan cells represented anatomy outside the acquired field of view, and the current analysis lacked a coverage-only comparator. The skeletal rendering is consequently an unadjusted group-level map of cross-sectional volume associations, not bone-density loss or longitudinal skeletal aging.

Across scales, then, the analyses give different answers to related questions: a disagreement coordinate carried a secondary mortality association among organ systems, was null for mortality among electrical phases, and was null for motor decline among brain-imaging views. That heterogeneity is the finding, not a coordinate that always predicts, but a decomposition whose two halves must each be tested, and whose disagreement half is informative only in some scale–outcome pairings. The ceiling is stated plainly throughout: even where a disagreement radius is associated with mortality, its incremental discrimination is small, it is not a stand-alone screening test, and almost all discrimination comes from the shared axis and chronological age. The contribution of this work is conceptual and methodological, a decomposition of subsystem age into interpretable coordinates, frozen before outcomes were read, and a demonstration that those coordinates can behave differently across prespecified tests, including a confirmatory null that motivated signed

phase-direction hypotheses, rather than a claim to a better risk model.

Several boundaries deserve emphasis. First, the electrical disagreement radius D showed no detectable association in the prespecified structure, acute-stress and mortality tests, including the single confirmatory mortality test; the development mortality association ($p = 0.0018$) was hypothesis-generating, used only to freeze the model, and did not survive on the locked final split. In the non-identifiable joint phase-gap parameterization, the association appeared concentrated in the AV and ST-T coefficients. The external identifiable SaMi-Trop test did not support that pattern, so it remains a phase-selective hypothesis rather than a validated predictor. A symmetric caveat applies to the shared axis: A is the headline coordinate, but it was not itself the single prespecified confirmatory endpoint, the locked reveal was designed to test D. The evidence that A shows consistent associations is therefore assembled from development-split, secondary and external analyses (CODE-15 development, EchoNext, secondary STAFF, previously outcome-exposed NHANES, and the Chapman age-transfer), not from a dedicated locked A-test. The one locked final-split quantity that bears on A is that the reference model containing age, sex, A and the whole-ECG clock reaches a Harrell concordance of 0.807 on the untouched CODE-15 final partition, against which adding D produced no improvement; a dedicated confirmatory test of A's incremental value is left to future prespecified work. Second, the CODE-15 development and final-test partitions are non-overlapping splits of one cohort, so the confirmatory test is an internal one; the NHANES organ-system arm is an independent analog at a different scale, but its mortality outcome had been inspected in earlier work and its (secondary, non-confirmatory) disagreement–mortality association is reported as such, not as a replication of an electrical finding that itself did not confirm. Third, the atrial (P) clock separates structural disease least of the four phases (Cohen's $d = 0.17$ versus 0.47 for AV and ST-T) and dilutes the four-phase disagreement score; a prespecified P-excluded sensitivity score showed a small structural association that was not promoted. Fourth, the "aging" displacement seen under acute ischemia is an apparent shift of an age-trained readout over minutes, not literal biological aging, and no mechanistic claim is made beyond the geometry. Finally, the shared-axis/orthogonal-residual architecture is itself established in other domains [5, 9]; the novelty here is its application to independently trained per-phase clocks under an outcome-blind freeze and the demonstration of dissociation, not the decomposition idea in the abstract.

Whether a related coordinate-based description extends to slower, non-electrical scales is an open question only begun to be probed here. In a limited pilot, an analogous structure–function construction was applied in NHANES 2005–2006 (whole-body DXA versus seven-day accelerometry, $n = 2891$; Supplementary S4), asking whether a "locomotor reserve" contrast predicted lower-body physical limitation. The prespecified test did not reach significance (odds ratio 1.10 per contrast unit, 95% CI 0.99 to 1.23, $p = 0.070$), but the pilot has technical limitations, accelerometry processed without a validated wear-time algorithm and DXA imputations averaged rather than combined by Rubin's rules, so no substantive conclusion is drawn from it and it is treated only as a motivating observation for a properly rebuilt version. It is at most suggestive that not every modality is best summarized as "how old does it look," and that a mature account of biological state may need age-derived coordinates alongside direct measures of function; establishing that will require a methodologically sound structure–function analysis left to future work.

The two-coordinate view suggests concrete next steps. Because the confirmatory test sent the residual electrical mortality signal to specific phases, the sharpest follow-up is a phase-resolved one: the AV-conduction and ST-T-repolarization age gaps show large coefficients that recur on an internal robustness partition, but because that parameterization does not isolate phase-specific effects, and the external SaMi-Trop test of the identifiable model was null, they remain specific, externally testable hypotheses rather than established effects, as is the phase-selective structural

signal (concentrated in the same AV and ST-T phases). The cross-scale contrast, namely an electrical disagreement radius that is null for mortality while an organ-system one is not, motivates testing the decomposition in other multi-clock settings (proteomic, imaging, or multi-organ panels) with prespecified endpoints, to map *where* a disagreement coordinate is informative rather than assuming it always is. Biological age, on this evidence, is not one clock; it is at least two coordinates, and they mean different things.

7 Methods

7.1 Phase clocks

The 12-lead ECG was delineated on lead II with a discrete-wavelet delineator to locate P, QRS and T boundaries, and each median beat was partitioned into four temporally disjoint windows: atrial depolarization (P), atrioventricular conduction (AV, the $\mathrm{P_{off}} \to \mathrm{R_{on}}$ segment), ventricular depolarization (QRS) and ventricular repolarization (ST-T). One convolutional age regressor was trained per window on the PTB-XL corpus against chronological age, together with a whole-ECG clock and shuffled-label negative controls. The clock historically labeled "PR" is trained on the AV-conduction window and is referred to throughout as the AV clock to reflect what it measures. Delineation accuracy was validated externally against cardiologist annotations in LUDB without any retuning (Supplementary Fig. S1): conditional boundary errors were 4–8 ms for the P wave, 36 ms for QRS onset (a systematic early placement), 10 ms for QRS offset and 12 ms for the T offset, with a 1.7% missed-beat rate and a 1.1% in-span false-detection rate over 200 records and 1800 matched beats.

7.2 From predicted phase ages to standardized residuals

The raw output of phase clock k was first mapped to the chronological-age scale by a frozen quadratic calibration $a_{ik} = c_{k2}\,\hat{y}_{ik}^2 + c_{k1}\,\hat{y}_{ik} + c_{k0}$ (the coefficients c_k were fit once on a held-out age-adaptation partition and frozen). Each calibrated phase age a_{ik} was then converted to a conditional standardized residual against a healthy reference, so that z captures how far a subsystem's apparent age departs from the age- and sex-specific healthy expectation, not a bare predicted-minus-chronological gap. With the design row $\mathbf{x}_i = \big[1,\ \frac{\mathrm{age}_i-50}{10},\ (\frac{\mathrm{age}_i-50}{10})^2,\ s_i,\ s_i\frac{\mathrm{age}_i-50}{10}\big]$ ($s_i = 1$ female, 0 male), the conditional mean and standard deviation are $m_{ik} = \mathbf{x}_i^\top \boldsymbol{\beta}_k^{\mathrm{mean}}$ and $\sigma_{ik} = \max\big(\exp(\frac{1}{2}\mathbf{x}_i^\top \boldsymbol{\beta}_k^{\mathrm{logvar}}),\ 0.25\big)$, and the standardized residual is

$$z_{ik} = \frac{a_{ik} - m_{ik}}{\sigma_{ik}}. \tag{1}$$

The calibration coefficients $\boldsymbol{\beta}_k^{\mathrm{mean}}$, $\boldsymbol{\beta}_k^{\mathrm{logvar}}$ and the variance floor were fit on the healthy reference and frozen; all constants are provided in the frozen definitions file (Code availability).

7.3 The shared axis and disagreement radius

With the four residuals stacked as $\mathbf{z}_i \in \mathbb{R}^4$ (phase order P, AV, QRS, ST-T), the shared-axis direction was fixed as $u = \frac{1}{2}(1,1,1,1)^\top$ (unit norm) and the shared axis defined as its projection

$$A_{\mathrm{raw},i} = u^\top \mathbf{z}_i = \tfrac{1}{2}\sum_{k=1}^{4} z_{ik} = 2\,\bar{z}_i, \tag{2}$$

i.e. twice the mean residual. A fixed orthonormal contrast basis $C \in \mathbb{R}^{3\times 4}$ spanning the complement of u gives the contrast vector $\mathbf{q}_i = C\mathbf{z}_i$ with $Cu = 0$ and $CC^\top = I_3$, so $\mathbf{q}$ lives in the three-dimensional subspace orthogonal to the shared direction. The disagreement score is the covariance-normalized radius *within* that subspace,

$$D_{\mathrm{raw},i} = \sqrt{\mathbf{q}_i^\top \, \Sigma_q^{-1} \, \mathbf{q}_i}, \tag{3}$$

with $\mu_q = 0$ (uncentered) and Σ_q the contrast covariance estimated on the healthy subset used for covariance estimation ($n = 9338$; Ledoit–Wolf shrinkage [28] 0.0177, covariance condition number 1.497). This covariance-estimation subset is distinct from, and smaller than, the $n = 9887$ frozen healthy calibration cohort on which A, D and the signed score are standardized below. D is a non-negative scalar length, not a signed axis; it is therefore not itself "orthogonal" to A. The precise statement is that D is a radius in the contrast subspace orthogonal to the shared direction, so it discards the shared-aging component by construction and also discards the direction of disagreement (only its normalized magnitude is retained). Under a standard-normal reference $D^2 \sim \chi^2_3$ as a construction ideal (the empirical tails are mildly over-dispersed, so it is described as an ideal, not a fitted distribution). A and D were standardized on the calibration partition to the calibration-reference SD,

$$D_{\mathrm{std}} = \frac{D_{\mathrm{raw}} - 1.568794}{0.757840}, \tag{4}$$

(A SD 1.4654); "one SD" throughout denotes one calibration-reference SD, not the empirical SD of any confirmation cohort. Every element of this geometry, including the clocks, the calibration functions, u, C, Σ_q and all standardization constants, was frozen before any mortality outcome was read.

7.4 Signed perturbation directions and the transport gate

The radius D retains the magnitude $\|\hat{\mathbf{u}}\|$ of the whitened contrast vector $\hat{\mathbf{u}} = \Sigma_q^{-1/2}\mathbf{q}$ and discards its direction. A *signed* perturbation score restores that discarded direction as the projection onto a fixed unit axis, so $|S_{\mathrm{raw}}| \le D_{\mathrm{raw}}$ for every exam in the *raw* whitened geometry. This bound relates the raw coordinates S_{raw} and D_{raw} only; the standardized scores $S_{I_{\mathrm{Kr}}}$ and D reported below are each separately z-scored on the frozen calibration cohort, so it is a geometric property of the raw coordinates, not of the standardized scores. For a controlled perturbation k with mean contrast-subspace displacement $\mathbf{m}_k = \mathbb{E}[C(\mathbf{z}_{\mathrm{post}} - \mathbf{z}_{\mathrm{baseline}})]$, the covariance-scaled signed direction and its whitened representation are

$$\mathbf{w}_k = \frac{\Sigma_q^{-1}\mathbf{m}_k}{\sqrt{\mathbf{m}_k^\top \Sigma_q^{-1}\mathbf{m}_k}}, \qquad \hat{\mathbf{v}}_k = \frac{\Sigma_q^{-1/2}\mathbf{m}_k}{\|\Sigma_q^{-1/2}\mathbf{m}_k\|}, \tag{5}$$

with the scaling chosen so that $\mathbf{w}_k^\top \Sigma_q \mathbf{w}_k = 1$ (verified to $<10^{-4}$); the patient score is $S_k = \mathbf{w}_k^\top \mathbf{q} = \hat{\mathbf{v}}_k \cdot \hat{\mathbf{u}}$. For I_{Kr} blockade the per-participant response was the triplicate-averaged, baseline-corrected (pre-dose -0.5 h), 0.5 h to 8 h time-averaged (trapezoidal integral divided by 7.5 h) paired dofetilide-minus-placebo contrast, and $\mathbf{m}_{I_{\mathrm{Kr}}}$ was the mean over the $n = 22$ participants; for acute ischemia it was the inflation-minus-baseline occlusion contrast averaged within patient before pooling (91 occlusions, 70 patients).

Two candidate perturbations were screened for a freezable direction (I_{Kr} blockade and acute ischemia), so direction derivation is an **exploratory** step and its permutation p-values are Holm-corrected across the two candidates; confirmatory weight is carried by the independent external phenotype test, not by this screen. Stability was read out on four descriptive criteria: pipeline yield $\ge$

0.70; participant/patient-bootstrap median covariance-aware cosine $\cos_\Sigma(\mathbf{a},\mathbf{b}) = \mathbf{a}^\top\Sigma_q^{-1}\mathbf{b}/\sqrt{(\mathbf{a}^\top\Sigma_q^{-1}\mathbf{a})(\mathbf{b}^\top\Sigma_q^{-1}\mathbf{b})} \geq$ 0.80 with lower 95% bound > 0; sign stability $\geq 90\,\%$; and a treatment-label sign-flip permutation test (statistic $\|\mathbf{m}\|_\Sigma$). Sign stability was computed as the fraction of bootstrap directions whose *signed* covariance-aware cosine with the full-sample direction was positive, with **no alignment applied**, so the count is not tautological; bootstrap directions were sign-aligned only for the compass visualization (Fig. 9a), never for the stability statistic. The resampling unit was the participant or patient, carrying all periods, timepoints and occlusions; all resampling used 10 000 replicates with `numpy` default-RNG seed 20 260 712. Criteria were applied and read out with no verdict predeclared; for I_{Kr} the sign-flip permutation gave $p = 0.034$ ($p = 0.069$ Holm-adjusted for the two screened directions), so the derivation is reported as suggestive/exploratory rather than confirmatory. Confirmation of the I_{Kr} direction used quinidine, a second I_{Kr} blocker in the same crossover cohort, hence a within-cohort held-mechanism confirmation rather than independent external validation, requiring a positive mean quinidine projection with a bootstrap CI excluding zero, positive $\cos_\Sigma(\mathbf{m}_{\mathrm{dof}}, \mathbf{m}_{\mathrm{quin}})$, and a leave-one-participant-out reconstruction. Ranolazine and verapamil were projected as secondary active-drug specificity controls, neither required null nor used to define the direction. The signed score was standardized on the frozen healthy calibration cohort (the same larger $n = 9887$ population on which D is standardized, not the $n = 9338$ subset used to estimate Σ_q). Sign-flipped, frozen sign-flip and 200 random Σ_q-normalized directions were retained as diagnostic empirical nulls only; they cannot be searched for a better direction, redefine $\mathbf{w}$, change the endpoint or gate the conclusion. A scorer, its fixture and an independent verifier (asserting $Cu = 0$, $CC^\top = I$, Σ_q symmetric positive-definite, $\mathbf{w}^\top\Sigma_q\mathbf{w} = 1$, exact fixture reproduction and that no outcome column is ever loaded) were frozen alongside a protocol lock carrying a canonical content hash and SHA-256 digests of C, Σ_q, the input feature tables and the scorer.

The external phenotype transport test applied the frozen I_{Kr} score, unchanged, to the combined Chapman–Shaoxing and Ningbo 12-lead databases (*ecg-arrhythmia* 1.0.0), an independent Chinese arrhythmia cohort whose records carry age, sex and SNOMED-CT diagnostic headers only, no mortality and no imaging. All 45 152 records were scored with the same STRIP pipeline that produced the calibration Σ_q (extraction QC pass rate 99.4 %; $n = 44\,550$ with usable records), so the score is in-metric; scorer faithfulness against the frozen definition was verified exactly (z-difference 0). The endpoint was the physician-assigned QT-interval-extension label (SNOMED 111975006; 386 cases), fixed before scoring. The prespecified inference is a one-degree-of-freedom likelihood-ratio test for the standardized I_{Kr} score in a logistic model of QT extension adjusting for age, sex, heart rate, QTc, QRS, whole-ECG age, A and D, with a marginal-to-fully-adjusted ladder reported in full because the effect is conditional (suppression). Secondary, post-hoc robustness analyses comprised a Firth penalized fit, atrial-flutter/fibrillation and sinus-bradycardia exclusion sensitivities, per-site (Ningbo, Chapman–Shaoxing) estimates, and a bootstrap comparison of incremental information against QTc, A and D; these were not part of a timestamped pre-result lock. A reliable-QT subgroup (QTc 350 ms to 600 ms, heart rate 40 bpm to 120 bpm) is a *post-hoc* robustness check and labelled as such. The label is imperfect, 42 % of raw cases have QTc below 350 ms in rhythms (flutter, bradycardia) where automated QT delineation is unreliable, so the QT-label analysis is reported as an external association test, not a gold-standard validation.

7.5 Outcome-blind protocol and firewall

Throughout, "prespecified" denotes an analysis whose hypothesis, model and success criterion were fixed in a protocol file, committed with a content hash, and frozen before any label or outcome data were accessed; the scoring pipeline was likewise hash-locked before the reveal. This is a content-hash-locked, outcome-blind commitment enforced by the firewall below, *not* a public, timestamped

registration in an external registry, and it is described as such rather than as formal preregistration. The CODE-15 cohort (233 770 patients, 345 779 exams; 12-lead, 400 Hz) was split by a salted patient hash *before* partitioning, yielding disjoint age-adaptation, calibration, survival-development and final-test partitions; patients with legacy exposure were forced into development and never into final-test. Mortality outcomes were held in physically separate storage volumes from the score-construction data, and the final-test outcome was read exactly once, at the single authorized confirmatory reveal (below). The development survival analysis set comprised 61 905 patients and 2292 deaths (median follow-up 3.6 years). The single prespecified confirmatory test, a primary M2-vs-M1 Cox likelihood-ratio test for D on the locked final-test partition, was then executed once on that partition ($n = 74\,715$ complete cases, 2583 deaths).

7.6 Cohorts and analyses

CODE-15 [11] provided the primary mortality analysis. The confirmatory model is a Cox proportional-hazards model

$$h_i(t) = h_0(t) \exp\{\beta_a a_i + \beta_s s_i + \beta_A A_{\mathrm{std},i} + \beta_G G_i + \beta_D D_{\mathrm{std},i}\}, \tag{6}$$

where a_i is chronological age, s_i sex, G_i the published whole-ECG age comparator [1], and A_{std}, D_{std} the frozen shared axis and disagreement radius. The prespecified test isolates the incremental value of D by comparing the nested models

$$M_1 : \beta_D = 0, \qquad M_2 = M_1 + \beta_D D_{\mathrm{std}}, \tag{7}$$

via the likelihood-ratio statistic $\Lambda = 2\{\ell(M_2) - \ell(M_1)\}$, $p = P(\chi^2_1 \geq \Lambda)$, with $\mathrm{HR}_D = e^{\beta_D}$. The frozen success rule is $p < 0.05$ *and* $\mathrm{HR}_D > 1$. Age and sex alone (M0) are reported as a reference. This single M2-vs-M1 test on the locked final-test partition is the sole confirmatory estimand; it was executed once, on an immutable, network-isolated compute image with the mortality vault mounted read-only, after a fresh-context internal attestation of the frozen analysis script and a create-once outcome-vault build performed by a custodian that emitted no outcome values. The score-to-outcome join was one-to-one and complete ($n = 74\,715$); sensitivities (P-wave-excluded and AV-excluded three-phase scores, a diagnosis-adjusted model, a non-linearity test) and an out-of-sample predictor-robustness partition were reported but never allowed to gate the verdict. **SaMi-Trop** [20, 21] provided an external, explicitly post-reveal test of the phase-direction hypothesis, run only after the CODE-15 mortality vault was spent and reported as exploratory (it does not alter the single confirmatory result). The frozen clocks, contrast basis, covariance and standardization constants were applied without any refitting through the same scoring library validated bit-exact against 44 832 Chapman rows (frozen-definition hash `0f6ef5d1...c6331 70`; protocol hash `1bbbb246...f960e5`). One index ECG per participant was scored; a prespecified, outcome-blind plausibility filter required all four predicted phase ages to fall in $[0, 120]$ years and complete model covariates with follow-up time >0, retaining 1439 of 1609 records (170 excluded, the AV clock failing to transfer to Chagas ECGs with predictions spanning -171 to $+518$ years). The mortality time origin was the index ECG with administrative censoring as released; 85 deaths accrued over a median 2.1 years. The prespecified primary was a three-degree-of-freedom partial likelihood-ratio test of the signed contrasts q_1, q_2, q_3 added to a base model age + sex + whole-ECG gap + A (Cox proportional hazards, Efron ties); D entered only a separate sensitivity model, and two data-motivated single-contrast tests (AV-versus-others, ST-T-versus-others) were Holm-corrected and labeled post-hoc. Per-contrast hazard ratios were q_1 1.05 (95% CI 0.93 to 1.18), q_2 0.91 (0.79 to 1.06), q_3 0.98 (0.86 to 1.12); D sensitivity 1.03 (0.97 to 1.09); proportional hazards were checked by

scaled Schoenfeld residuals. **EchoNext** (Elias & Finer; data staged from the study team's private repository, not fetched from a public server) [12, 13] provided the structural-disease test: a logistic model for echocardiographically confirmed structural heart disease adjusted for age, sex, heart rate and standard intervals, with the frozen D as the prespecified term (final split $n = 9687$, 3794 events). Dev-to-final agreement within this single cohort is described as internal reproduction, not replication. **ECGRDVQ** [14, 15] provided a pharmacological specificity control. For participant i on arm d, the baseline-corrected effect curve is $\Delta D_{id}(t) = D_{id}(t) - D_{id}(t_0)$, and the estimand is the trapezoidal area under it, $\mathrm{AUEC}_{id} = \int_{0.5}^{8} \Delta D_{id}(t)\, dt$, evaluated over 0.5 to 8 hours; the primary contrast is the paired within-participant difference $\mathrm{AUEC}_{i,\mathrm{dofetilide}} - \mathrm{AUEC}_{i,\mathrm{placebo}}$, tested by paired bootstrap. Placebo test–retest reliability of D was high (intraclass correlation, absolute agreement, 0.83). **STAFF III** [16, 17] provided within-person acute ischemia (91 occlusions, 70 patients; 1 kHz; a disclosed, hashed 9-to-12-lead adapter reconstructed the augmented leads via the Goldberger relations). For occlusion j in patient i the within-occlusion contrast is $\Delta D_{ij} = D_{ij,\mathrm{inflation}} - D_{ij,\mathrm{baseline}}$ (and analogously for recovery minus baseline); because a patient may contribute multiple occlusions, inference used a patient-clustered bootstrap that resamples patients (carrying all their occlusions), and per-phase displacements were false-discovery-rate adjusted. **Chapman–Ningbo** [18] provided geographic transfer: the frozen clocks were applied without refitting and age correlation was measured (outcome-blind; no mortality claim). **NHANES** 2005–2010 linked to 2019 mortality [19] provided the cross-scale analog with an analogous whole-body BodyVector; its construction, survey handling and concordance estimation are given in full in the next subsection.

7.7 Whole-body BodyVector (NHANES)

The cross-scale analog mirrors the phase-clock geometry with organ-system age gaps. **Cohort and systems.** NHANES cycles 2005–2006, 2007–2008 and 2009–2010 (ages 20–79, MEC-examined, mortality-eligible) were pooled, giving $n = 14\,362$ with 1739 deaths over 160 493 person-years (median follow-up 11.3 years). Six organ systems were defined *a priori* from standard clinical assays: *Cardiovascular* (systolic and diastolic blood pressure, pulse pressure, pulse, total cholesterol); *Metabolic* (HbA1c, BMI, waist circumference, serum glucose); *Renal* (creatinine, urea nitrogen, albumin, urate); *Hepatic* (ALT, AST, GGT, alkaline phosphatase, total bilirubin, total protein, globulin); *Immune* (CRP, white-cell count, neutrophils, lymphocytes, monocyte %, neutrophil-to-lymphocyte ratio); and *Hematologic* (red-cell count, hemoglobin, hematocrit, MCV, red-cell distribution width, platelets, mean platelet volume). Right-skewed markers (CRP, alkaline phosphatase, AST, ALT, glucose, GGT, bilirubin, urate, neutrophil-to-lymphocyte ratio) were log-transformed. Estimated GFR was excluded *a priori* both as a feature and as a healthy-gate criterion because CKD-EPI embeds chronological age; the renal panel retains only directly measured markers. **Per-system clocks.** For each system a separate age regressor was trained on its panel by elastic-net regression against chronological age (`ElasticNetCV`, ℓ_1-ratio and penalty chosen by internal cross-validation minimizing mean squared error). Median imputation, standardization, elastic-net fitting and a linear age-bias correction were all fit *in-fold* on the training-fold healthy subset only and applied out-of-fold, in a 5-fold cross-fit, so every participant's six age gaps are out-of-sample. **Healthy reference.** The outcome-blind healthy calibration set ($n = 8691$) was defined by disease- and medication-free status across twelve standard flags (no reported cardiovascular disease, diabetes, cancer, or relevant medications; sex-specific in-range creatinine replacing any age-based renal gate). Each system gap was standardized to the mean and SD of this set. **Geometry.** The six standardized gaps $\mathbf{z}^{\mathrm{body}} \in \mathbb{R}^6$ define the shared axis $A_{\mathrm{body}} = \mathbf{z}^{\mathrm{body}\top} u_0$ with $u_0 = \mathbf{1}/\sqrt{6}$ (equal-weight mean acceleration), and, via a fixed 6×5 orthonormal contrast basis C_6 with $C_6 u_0 = 0$, the disagreement radius $D_{\mathrm{body}} = \sqrt{\mathbf{q}^\top \Sigma_{\mathrm{body}}^{-1} \mathbf{q}}$, $\mathbf{q} = C_6 \mathbf{z}^{\mathrm{body}}$, with Σ_{body} the 5×5 contrast covariance on the healthy

set (mean $D^2 = 5.00$, matching the χ^2_5 construction ideal). A_{body} and D_{body} are standardized per SD for hazard-ratio reporting. **Survival model and survey design.** Survey-weighted Cox models (`svycoxph`) used the combined three-cycle MEC weight (`WTMEC2YR`/3), primary sampling units (`SDMVPSU`) nested in strata (`SDMVSTRA`) and lonely-PSU adjustment; the design supported 47 degrees of freedom (primary sampling units minus strata), which govern the survey-weighted variance estimates. The ladder adds A_{body} then D_{body} to an age-and-sex baseline. Reported coefficient p-values are the design-based Wald tests returned by `svycoxph` ($z = \hat{\beta}/\text{se}$ referred to the standard normal), so the reported A_{body} ($p = 2.0 \times 10^{-8}$) and D_{body} ($p = 3.1 \times 10^{-5}$) values are normal-approximation Wald p-values, not t_{47} values; a finite-sample t_{47} reference would give more conservative but still significant values ($\approx 1 \times 10^{-6}$ and $\approx 1.3 \times 10^{-4}$ respectively). The metabolic panel uses serum glucose from the standard biochemistry profile (`LBXSGL`), a measurement taken on all mobile-examination-center participants rather than the morning fasting subsample; the mobile-center weight (`WTMEC2YR`, combined across three cycles as `WTMEC2YR`/3) is therefore the correct design weight for the full analytic sample and no fasting-subsample weighting is required. Incremental concordance (ΔC) was estimated by survey-weighted 5-fold cross-validation with PSU-level folds and by bootstrap, and a leave-one-system-out analysis dropped each system in turn. **Firewall.** No mortality label entered clock training, standardization, geometry or calibration; mortality (`MORTSTAT`/`PERMTH`) was read only at the final Cox stage. This NHANES arm had been outcome-inspected in a prior unweighted analysis and is therefore reported as a secondary conceptual replication, never as a primary or a rescue. All systems, panels, constants, the contrast basis C_6 and Σ_{body} are provided in the frozen BodyVector definition file (Code availability).

7.8 Structure–function pilot (MotionVector)

As an exploratory probe of the decomposition's boundary, a similar construction was applied to a structure–function pair in NHANES 2005–2006, the only cycle with both whole-body dual-energy X-ray absorptiometry (DXA) and minute-level accelerometry in the same participants. Two outcome-blind cross-fitted ElasticNet age clocks were trained, one on regional DXA bone-mineral, bone-content and lean-mass measures ("skeletal"), one on seven-day accelerometry summaries (steps, activity-intensity minutes, sedentary fraction and day-to-day variability; "locomotor"), on a healthy calibration subset. A shared structure–function axis was fixed as the first principal axis of the two standardized age gaps on the calibration set, and a "locomotor reserve" was defined as the calibration-orthogonal contrast of the locomotor gap (a contrast, not a globally-orthogonal residual: it correlates $r = -0.15$ with the shared axis on the full sample). The clocks, axis and reserve were frozen (protocol SHA-256 hash recorded) before any outcome was read. The prespecified primary test was whether the reserve predicted a lower-body physical-limitation score beyond age, sex, BMI and the shared axis (logistic likelihood-ratio test; success defined as $p < 0.05$ in the protective direction). The analysis set was $n = 2891$ (ages 20–69, 378 with limitation). This arm is reported as a *technically limited pilot* rather than as evidence for or against the thesis, for several reasons that a rebuilt version (V2) would need to fix. First, the accelerometry summaries were computed from the raw minute-level intensity counts without a validated non-wear/wear-time algorithm or valid-day filtering, so the activity features are not comparable to the accepted NHANES accelerometry processing pipeline. Second, the five multiply-imputed DXA replicates were *averaged* into a single record rather than analyzed separately and combined by Rubin's rules, which understates measurement variance contrary to NHANES guidance. Third, the outcome is a self-reported functional limitation (soft) in a young DXA-eligible cohort (20–69, few deaths, so the prespecified mortality test was not run), and the implemented locomotor feature set was only a subset of the locked list (cadence-distribution, walking-bout, fragmentation and cosinor-amplitude

features were not computed and, the outcome now being unblinded, were deliberately not added post hoc). Given these limitations the pilot's null is not interpreted as a validated boundary result; it is retained only as a motivating observation and does not support any claim in the main text.

7.9 Brain-imaging decomposition (NeuroMotionVector)

To test the decomposition prospectively at a third scale, the identical shared-axis/disagreement construction was applied to brain imaging in the Aging Brain Cohort (AABC / HCP-Aging Release 2; 1396 adults, 2878 visits), under a controlled-access data-use agreement. Governance was strict: participant-level imaging data were analyzed only inside a user-approved, access-controlled environment, no participant scan was ever rendered, and only code, schema, checksums, group-level maps and aggregate results left that environment; the anatomical figures use the open-access LEMON dataset [29], not AABC scans. **Channels.** Four imaging-derived phenotype sets, each yielding one age clock: *structure* (HCP-MMP cortical thickness and volume plus FreeSurfer subcortical volumes), *myelin* (T1w/T2w areal myelin, globally covariate-corrected), *perfusion* (arterial spin-labeling cerebral blood flow and arterial transit time) and *function* (resting-state fMRI amplitudes aggregated to networks). No scan images were used, only imaging-derived phenotypes. **Splits and clocks.** Participants were hash-split 60/20/20 (development/calibration/final-test) with all visits of a person kept in one partition; one ElasticNet age clock per channel was trained on development only, with age-bias correction and standardization fit on calibration. All four clocks reached held-out $r \geq 0.49$ (structure 0.587, myelin 0.505, perfusion 0.514, function 0.497), clearing the pre-specified $r \geq 0.20$, $\geq 3/4$ gate. **Geometry.** The four standardized channel gaps $\mathbf{z}^{\text{brain}} \in \mathbb{R}^4$ define $A_{\text{brain}} = \frac{1}{4}\mathbf{1}^\top \mathbf{z}^{\text{brain}}$ and, via a fixed 3×4 Helmert contrast C with $C\mathbf{1} = 0$ and $CC^\top = I_3$, the disagreement radius $D_{\text{brain}} = \sqrt{\mathbf{q}^\top \Sigma_q^{-1} \mathbf{q}}$, $\mathbf{q} = C\mathbf{z}^{\text{brain}}$, with Σ_q a Ledoit–Wolf covariance on the calibration contrasts. The geometry (clocks, bias correction, covariance, standardization) was frozen and hashed on 2469 complete four-channel visits before any outcome was opened. **Outcome-blind firewall.** The analysis code refuses to run the outcome stage unless the frozen-geometry SHA-256 verifies and $\geq 3/4$ clocks pass; this was confirmed to block tampered, unfrozen and under-powered configurations, an auditable rather than asserted blinding. **Availability gate and primary test.** Because a 20% final-test partition (≤ 279) cannot reach the original locomotor-power threshold, an outcome-blind rule decided from availability counts alone before any outcome was read, triggered the longitudinal primary on the 40% non-training holdout (clocks still frozen on the untouched development set), requiring ≥ 150 complete longitudinal participants and otherwise falling back to cross-sectional. The primary model regressed annualized change in four-metre walk performance on baseline D_{brain}, adjusting for A_{brain}, baseline gait, a smooth function of age, sex, education, site, height and BMI, with a one-degree-of-freedom nested test on the disagreement coefficient (two-sided $\alpha = 0.05$). Secondary motor and cognitive outcomes (two-minute walk, grip, MoCA) used Benjamini–Hochberg control, and a favorable secondary was pre-specified never to replace a null primary. The analysis code was sanity-checked on two synthetic fixtures with known ground truth (planted effect recovered at $p = 0.045$; null data returned at $p = 0.23$); these single fixtures check code correctness and do not calibrate statistical power or type-I error. A 7T MR-spectroscopy molecular layer was specified as a further secondary but was not run in this analysis. This arm is an independent brain-imaging extension and is explicitly not a substitute for the CODE-15 confirmatory test.

Cross-view dispersion from raw voxels (LEMON)

As an independent construct check built from images rather than precomputed clinical tables, all 220 available skull-stripped MP2RAGE T1 volumes of the open-access LEMON cohort (PDDL license; GWDG mirror; 151 younger, 20–35 y and 69 older, 55–80 y by the released 5-year age bins) were processed end-to-end. Raw volumes were streamed one at a time through the sandbox to durable storage and deleted locally, keeping a single-file on-disk footprint; five subjects absent at the ses-01 mirror path (HTTP 404) were excluded, and all 220 archived subjects completed the pipeline. Four *complementary* T1-derived estimates of structural brain age were computed per subject with the ANTsX ecosystem [30]: a deep brain-age regression [31]; gray-matter fraction from a six-tissue `deep_atropos` segmentation; the ventricle–brain ratio from a DKT parcellation; and the brain-tissue fraction (segmented tissue over total intracranial volume). Each view was oriented so that higher values represented older appearance and standardized across the pooled cohort. Cross-view structural-aging dispersion, S_{T1}, was defined as the within-person population standard deviation across the four oriented scores. This is deliberately simpler than the covariance-normalized Mahalanobis radius D_{brain} used in AABC. The four views are complementary rather than independent because all derive from the same T1-weighted acquisition. Because the deep brain-age clock was trained on conventional T1w and LEMON is MP2RAGE, absolute brain-age carries a domain-shift bias (rank Pearson $r = 0.95$ against binned age but a compressed slope near 0.79 with a positive offset); the primary test was therefore pre-specified on the bias-invariant younger-versus-older *difference* in dispersion, which is insensitive to any additive or multiplicative per-cohort bias. Significance used a one-sided Mann–Whitney test with a bootstrap confidence interval and Cohen's d; robustness comprised a 5000-fold label-permutation null, a leave-one-view-out re-computation, an alternative range-based dispersion form, an extremity-adjusted ordinary-least-squares model ($S_{\mathrm{T1}} \sim |\text{overall level}| + \text{older}$) that removes uniform old-appearance before testing group membership, and a within-older age-gradient test. Because pooled standardization with unequal group sizes can amplify group contrasts, a post-hoc imbalance diagnostic repeated the full standardization in 5000 balanced 69-versus-69 samples and under equal-group weighting. A separate within-group-standardization diagnostic tested whether the result represented excess within-age-group dispersion rather than differential shifts between groups. Age bin midpoints were used as an age proxy for the continuous checks only; the primary contrast is binary and unaffected by binning. This analysis is a cross-sectional construct-validity demonstration on open data and is deliberately distinct from, and not a substitute for, the prespecified prospective AABC gait test above.

7.10 CT structural age clock

Cohort and features. The cohort is the publicly released TotalSegmentator v2 image collection [22]: clinical CT examinations acquired in routine care across ten contributing institutions on 20 scanner models (Siemens, Philips and GE), with heterogeneous anatomical coverage (thorax, abdomen, pelvis, neck, CT-angiography and polytrauma protocols). Of the 1228 released examinations, the $n = 1227$ with a recorded chronological age were used. Each examination is a distinct subject identifier in the released metadata; the collection is treated as one scan per subject, but the release carries no explicit patient key, so repeat-subject leakage cannot be positively excluded (see cross-validation note below). Each examination was segmented into approximately 115 anatomic structures (organs, bones, muscles and vessels) using TotalSegmentator, and each structure contributed a single feature, its segmented volume; no Hounsfield-unit or density-derived feature was used. Structures outside a scan's field of view were encoded as zero volume (40 % of structure-by-

scan entries), so partial-body coverage enters the model as explicit absence rather than imputation. Because zero coding makes anatomical coverage observable to the estimator, the reported performance may combine anatomy, protocol and clinical indication. No coverage-only comparator or protocol-stratified validation was available; this CT arm therefore does not isolate tissue-intrinsic aging. This volume-only design is a deliberate lower bound: intensity features such as vertebral trabecular attenuation (bone-mineral density) and muscle and fat attenuation are known to be age-informative and are left as a documented, un-executed upgrade.

Clock and evaluation. A histogram-based gradient-boosted regression tree ensemble (scikit-learn `HistGradientBoostingRegressor`, library-default hyperparameters, fixed random seed; native handling of the zero-coded out-of-field-of-view structures) mapped the vector of structure volumes to chronological age. It was evaluated by five-fold cross-validation with shuffled exam-level folds (`KFold`, shuffle, fixed seed), reporting mean absolute error and R^2 with bootstrap 95 % confidence intervals (2000 resamples of the out-of-fold residuals). Folds were assigned at the examination level; because the released metadata carries no patient identifier, a patient-grouped split could not be constructed, and the reported errors should be read as exam-level cross-validation under the assumption, not positively verified, of one scan per subject. Per-system clocks were fit identically on the structures of each organ system. Predicted age is subject to the standard regression-to-the-mean bias (over-prediction in the young, under-prediction in the old); a linear age-bias model (slope -0.561, intercept 35.12) was fit on 404 disease-free control subjects and used to residualize predicted age before any age-gap was computed, following established age-prediction bias-correction practice [32, 33].

Structure importance. Model reliance was summarized by the increase in cross-validated MAE after permuting one structure-volume feature while leaving the fitted model and all other features unchanged. Correlated anatomy and correlated field-of-view patterns can redistribute or inflate permutation effects; the measure is therefore not interpreted as an independent biological contribution or an anatomical aging rate.

Exploratory structure–pathology comparisons. For each structure, the bias-corrected organ-age gap was compared between subjects carrying a given pathology label (vascular, trauma, tumor or inflammation) and disease-free controls, with effect size expressed as Cohen's d. Significance was assessed across all structure×pathology combinations under Benjamini–Hochberg false-discovery-rate control at 0.05. The comparisons were not adjusted for sex, body size, site, protocol, anatomical coverage or clinical indication and were not externally validated.

Skeletal association map. The frozen atlas table supplied one cross-sectional volume–age Spearman ρ for each prespecified skeletal group. The same group value was assigned to every mesh belonging to that group and painted onto the real segmented anatomy of reference examination s1045. The visualization is group-level rather than a set of independent left/right or vertebra-specific estimates. No adjustment was made for sex, body size, site, protocol, anatomical coverage or clinical indication.

7.11 SKELETOME: HAR skeletal variant-effect screen

HAR set and human-specific substitutions. The screen used the 312-region Zoonomia HAR call set (UCSC `hars312`; 23, doi:10.1126/science.abm1696). Within each HAR, positions carrying a

human-specific derived allele relative to the inferred mammalian ancestral state were enumerated, giving 1955 human-specific substitutions (mean 6.27 per HAR, maximum 38); three additional *GDF5*-locus substitutions were carried as blind controls with pre-specified pass/reject calls on the constraint-and-gBGC filter (not a frozen predicted effect direction; 1958 scored in total).

Skeletal-minus-neural coverage scoring (exact configuration). Each substitution was scored with Borzoi [24], an open (MIT-licensed) sequence-to-coverage model, using the public PyTorch port `borzoi_pytorch` and the pretrained weights `johahi/borzoi-replicate-0` loaded via `Borzoi.from_pretrained`, with no model retraining or fine-tuning. A *single* replicate was used; predictions were *not* ensembled across Borzoi replicates and were *not* averaged over the forward and reverse-complement strands. Reference sequence was hg38 (UCSC `hg38.fa`, faidx-indexed); the exact reference file and its checksum are recorded in the released run manifest, and the substitution manifest (`manifest.tsv`) fixes every `har_id`, `chrom`, `pos_hg38`, `ref` and `alt`. Each variant was centred in a 524 288-bp window; the derived and ancestral alleles were passed through the model, and for each track the predicted-coverage difference (derived minus ancestral) was summed over the 32 output bins spanning ±512 bp around the variant (central bin OUT_BINS/2 = 3072, flank ±16 bins of 32 bp). Per-track summed deltas were then averaged *across* the tracks within each group (skeletal, neural). The 16 skeletal-lineage tracks, ENCODE DNase-seq experiments, given with their Borzoi output-head indices in `results/skeletome_borzoi_track_indices.csv`, comprise osteoblast (index 1319, ENCFF300IKA), bone-marrow stromal cell (1334, ENCFF747REW), skeletal-muscle myoblast, myotube and skeletal-muscle cell (1392/1393/1423), villous-mesenchyme fibroblast (1395), embryonic fore-/hind-limb muscle (1523/1634), embryonic arm- and leg-bone (1578/1654/1667/1940), embryonic limb (1709/1852) and mesenchymal stem cell (1809/1854). The 81 neural comparator tracks are the model's DNase/ATAC output heads whose annotations match neural, neuron, brain, cortex, astrocyte, forebrain or hindbrain (immune-lineage heads such as CD34, pDC, GMP and megakaryocyte excluded); the full index list is released alongside the score tables. Track labels denote the literal Borzoi readout; no cell type is inferred beyond the track's own annotation. The reported skeletal-minus-neural contrast is the mean skeletal-track delta minus the mean neural-track delta, computed on the model's native predicted-coverage scale with no per-track normalisation (no standardisation, quantile or log transform). Because the skeletal and neural sets differ in track count and baseline predicted signal, this is a difference of raw group-mean coverage deltas, not a variance-standardised effect size, and is treated throughout as a descriptive model output rather than a measured or validated effect.

Constraint and gBGC-susceptibility filters. Per-base evolutionary constraint was taken from the Zoonomia 241-way Cactus mammalian alignment phyloP track [25], with a substitution deemed constrained at phyloP > 2.27 (a second, permissive threshold of 1.6 is also reported). Each substitution was assigned a GC-biased gene conversion class [26] from its allelic direction, weak-to-strong (A/T $\rightarrow$ G/C), strong-to-weak (G/C $\rightarrow$ A/T) or neutral. Because allelic direction alone cannot establish that gBGC occurred at a given site, weak-to-strong substitutions were flagged as *gBGC-susceptible* rather than classified as artifacts; attributing any individual substitution to gBGC would require recombination-rate context and direct evolutionary evidence not used here. Candidate skeletal-regulatory substitutions were those passing both the constraint threshold and not carrying the gBGC-susceptible flag (722 of 1955).

GWAS crossing, the enrichment null, and ranking. Candidate HARs were intersected with osteoarthritis and bone-mineral-density credible-set variants (936 OA and 903 BMD credible-set

positions) using a 25 kb proximity window; exact-base coincidence was recorded separately (0 exact overlaps). The central inferential test was whether the ten HARs proximal to an OA/BMD locus are more skeletal-specific than matched non-proximal HARs. Per-HAR skeletal specificity was the maximum absolute skeletal-minus-neural contrast over that HAR's candidate substitutions; a matched permutation null drew, for 50 000 iterations (fixed seed `SKELETOME_NPERM` run recorded in the release), sets of HARs matched on candidate-substitution count and compared the proximal-set mean (0.220) with the null mean (0.179), giving $p = 0.24$ (region-length-matched and unmatched nulls give $p = 0.26$–0.28). The permutation script, seed and matching variables are released with the score tables. Candidate rankings are *descriptive*: no per-variant p-value or false-discovery control is computed, and the ranked list is a preliminary screen of prioritised hypotheses, not a set of statistically significant findings, a limitation stated as such. A supplementary genetic-score $\times$ tractability composite over nearby skeletal genes is provided only as a forward-looking prioritisation aid. A blinded audit manifest for the *GDF5*/GROW1 filter-logic control contained variant identities and pass/reject calls, but no predicted effect direction; it was frozen before scoring and is reported without post-hoc adjustment. The HAR call set, phyloP track, alignment build and all thresholds are recorded with the released summary tables.

7.12 Post-hoc construct-validity, transfer and reproducibility analyses

The following analyses were run after the confirmatory reveal and are post-hoc. For each, the four phase-age predictions were recomputed through the frozen pipeline without refitting, quadratic domain adaptation, heteroscedastic age/sex calibration (sex coded female = 1, age centered in decades at 50 y), shared axis $A = u^\top \mathbf{z}/2$ with $u = (1, 1, 1, 1)$, contrast $\mathbf{q} = C\mathbf{z}$ and $D = \sqrt{\mathbf{q}^\top \Sigma_q^{-1} \mathbf{q}}$; reproduction of the frozen calibration z-scores was verified to $< 1 \times 10^{-6}$ (A) and $< 2 \times 10^{-4}$ (D). *Interval anchoring:* partial Spearman correlation between each predicted phase age and its matched interval (P-duration, PR interval, QRS duration, QTc), controlling for age and sex by rank-residualization, with Fisher-z confidence intervals. *External transfer:* frozen inference on Chapman–Shaoxing/Ningbo (Zenodo, open), restricted to successfully delineated records with known age and sex; age transfer reported as Pearson r of predicted versus chronological age. *Fixed versus data-driven axis:* PC1 of the four development-split z-scores (mean-centered SVD), compared by cosine similarity of loadings and correlation of resulting per-exam scores. *Test–retest reproducibility:* one-way random-effects ICC$(1, 1)$ across patients with ≥ 2 QC-passing exams; because CODE-15 lacks per-exam dates this measures within-patient stability, not calendar drift.

7.13 Software and reproducibility

All clocks, geometry constants and analysis scripts were frozen as versioned artifacts with content hashes; the confirmatory reveal script is hash-locked, and its prespecified inputs and success criterion were checked by a fresh-context reviewer with no access to the outcome labels. Figures were rendered from the frozen result tables. Anatomical renderings in Figs. 6 and 11 use CT-segmented cardiac chamber meshes and are illustrative of anatomy, not of any individual participant's scan.

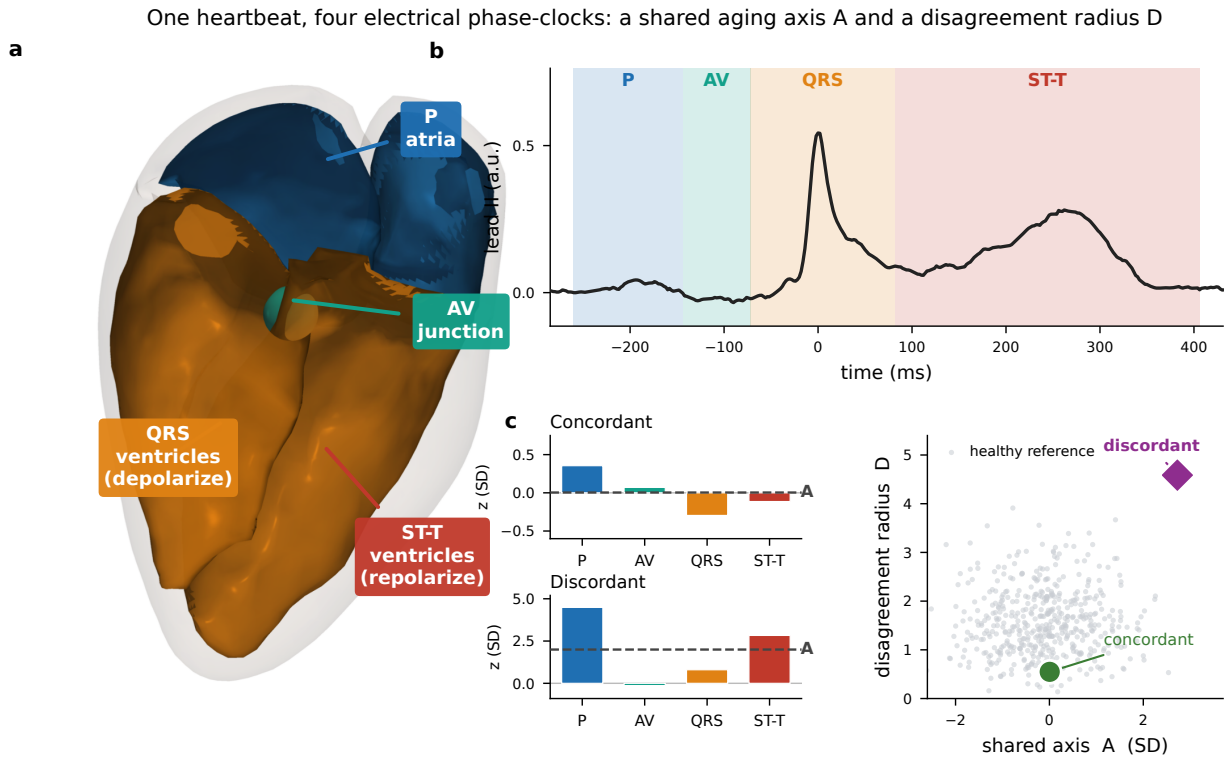

Figure 6: Biological age as a vector: two coordinates from four phase clocks. **(A)** The four electrical phases of the heartbeat mapped onto their anatomical territories on a CT-segmented four-chamber heart (translucent epicardium with the two atrial and two ventricular cavities): P (atria), AV conduction (atrioventricular junction), QRS (ventricular depolarization) and ST-T (ventricular repolarization). **(B)** A median beat partitioned into the four temporally disjoint windows, color-matched to (A); one independently trained age clock reads each window. **(C)** The four standardized phase age gaps define two coordinates: a shared axis A (their coordinated component, the projection onto the shared direction) and a disagreement radius D (the covariance-normalized Mahalanobis length of the residual contrast in the subspace orthogonal to that direction). A concordant heart sits near the centre of the healthy ellipsoid (D small) whatever its A; a discordant heart, in which one phase contrast dominates, has a large D. Both coordinates were frozen outcome-blind. Anatomy in (A) is illustrative, not a study participant's scan.

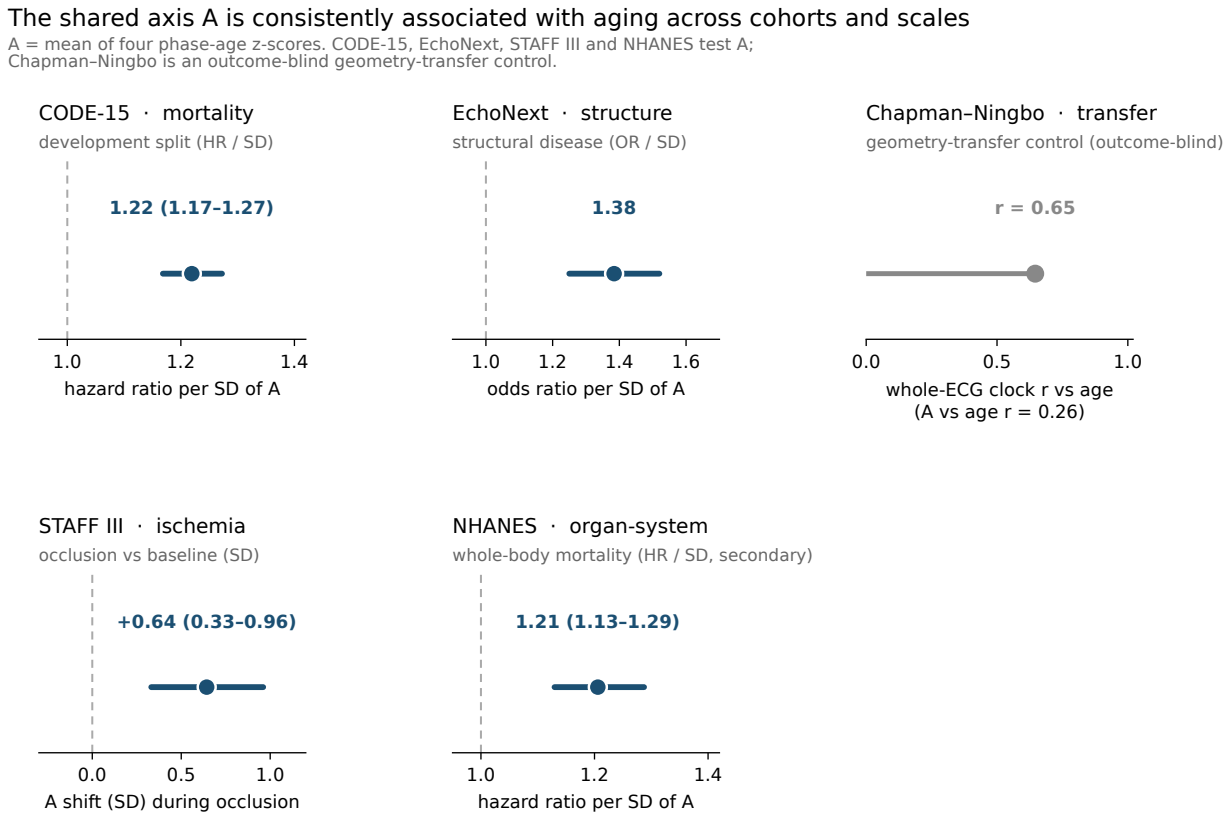

Figure 7: The shared axis A is consistently associated with aging across cohorts and biological scales. Effect of A (per SD unless noted) in four settings that directly test A: mortality (CODE-15 development, hazard ratio 1.22 per SD), echocardiographically confirmed structural disease (EchoNext, odds ratio 1.38 per SD), acute ischemia (STAFF III, +0.64 SD) and whole-body organ-system mortality (NHANES, hazard ratio 1.21 per SD). The fifth bar is a geometry-transfer control, not a test of A: on the untrained Chapman–Ningbo cohort the whole-ECG clock that anchors the construction reconstructs at age $r = 0.646$ (A itself, a residual over age-adapted gaps, correlates with age at only $r = 0.26$ here). Because the effect measures differ in kind (hazard ratio, odds ratio, SD shift, correlation), they are shown on separate, labeled scales rather than pooled. Chapman–Ningbo is outcome-blind; NHANES is a secondary cross-scale analysis. Bars/points show 95% confidence intervals where applicable.

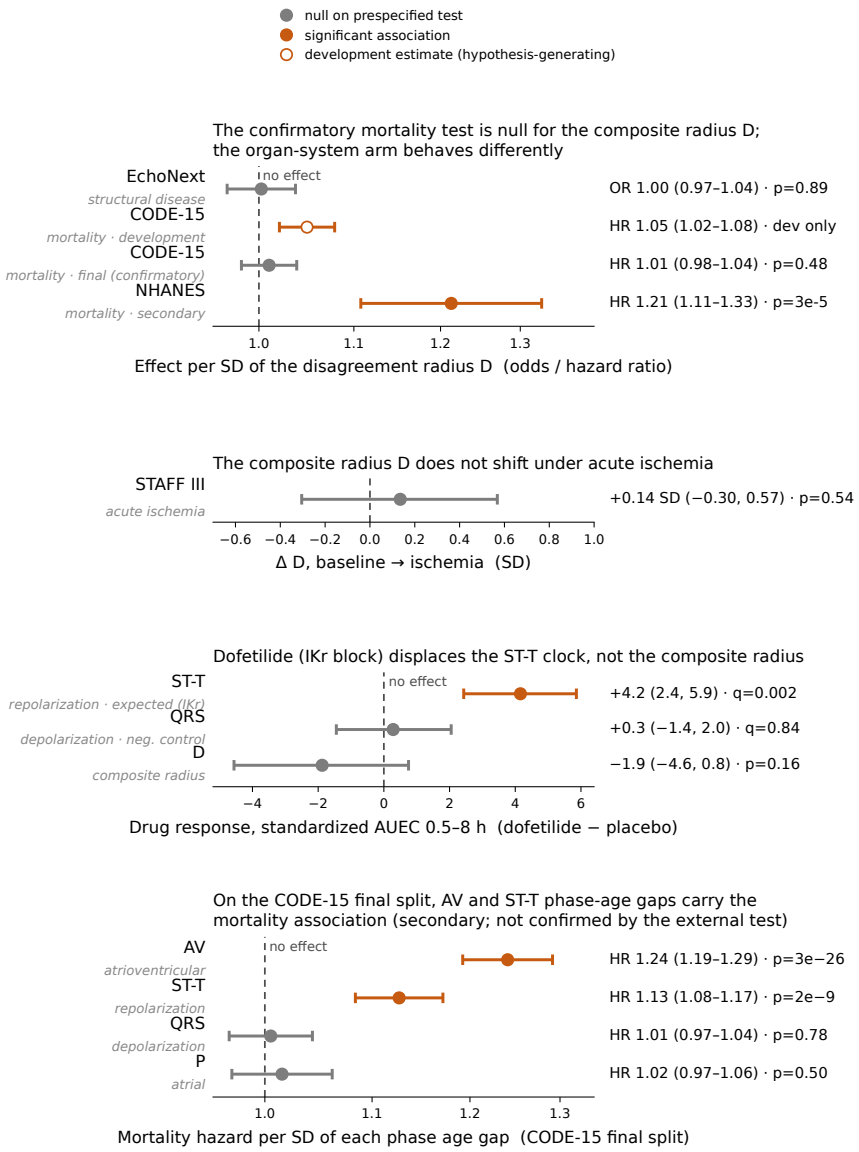

Figure 8: The electrical disagreement radius D is null on every prespecified confirmatory-grade test, yet the signed phase decomposition behind it recovers expected pharmacology and generates phase-gap hypotheses that an external test did not support. (Row 1) Effect of the composite radius D per SD across endpoints: structural disease is null (EchoNext, odds ratio 1.00 per SD, $p = 0.89$); the development mortality association (hollow, hazard ratio 1.05 per SD) is hypothesis-generating and was used only to freeze the model; and the single prespecified confirmatory test on the locked CODE-15 final split is null (hazard ratio 1.01 per SD, 95% CI 0.98 to 1.04, $p = 0.48$), its interval excluding the development estimate. A whole-body organ-system disagreement radius does carry a secondary, non-confirmatory association (NHANES, hazard ratio 1.21 per SD, $p = 3.1 \times 10^{-5}$), so "disagreement" behaves differently across scales. **(Row 2)** The composite radius does not shift under acute ischemia (STAFF III, +0.14 SD, $p = 0.54$). **(Row 3)** Under the I_{Kr} blocker dofetilide the composite radius does not move (−1.9 standardized AUEC, $p = 0.16$), but the signed decomposition does: the drug displaces the ST-T repolarization clock specifically (+4.2, false-discovery-rate $q = 0.002$) while the prespecified QRS depolarization negative control does not ($q = 0.84$), matching the drug's mechanism. **(Row 4)** With the four individual phase age gaps on the CODE-15 final split, the atrioventricular (hazard ratio 1.24 per SD, $p = 3 \times 10^{-26}$) and ST-T (1.13 per SD, $p = 2 \times 10^{-9}$) gaps show large coefficients while the P and QRS gaps are null; because this mixed model does not isolate phase-specific effects at fixed A and an external identifiable test (SaMi-Trop) was null, these are hypotheses rather than established effects.

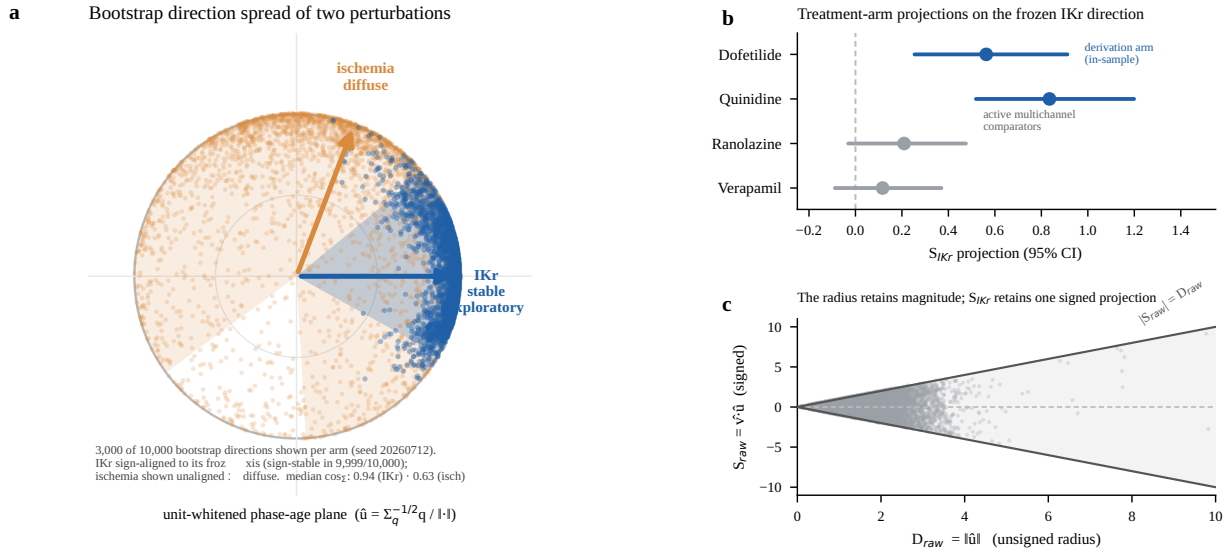

Figure 9: An outcome-blind, signed perturbation direction recovers a mechanism that the unsigned radius discards, and is frozen for external transport. A controlled human ion-channel perturbation defines a fixed axis in the contrast subspace: the covariance-scaled signed direction $\mathbf{w}_{I_{\mathrm{Kr}}} = \Sigma_q^{-1}\mathbf{m}/\sqrt{\mathbf{m}^\top\Sigma_q^{-1}\mathbf{m}}$, where $\mathbf{m}$ is the mean dofetilide-minus-placebo phase-contrast displacement (ECGRDVQ, $n = 22$ participants). The external patient score is the projection $S_{I_{\mathrm{Kr}}} = \mathbf{w}_{I_{\mathrm{Kr}}}^\top\mathbf{q}$. **(a)** Bootstrap direction spread, shown in a two-dimensional slice of the three-dimensional whitened contrast space ($\hat{\mathbf{u}} = \Sigma_q^{-1/2}\mathbf{q}$): the plane is spanned by the frozen I_{Kr} axis (horizontal) and the orthogonal component of the ischemia axis (vertical), and each bootstrap direction is drawn as its unit vector projected onto that plane. The I_{Kr}-blockade direction (blue) is tightly concentrated, bootstrap median $\cos_\Sigma = 0.94$ with the full-sample axis, and 9999 of 10 000 bootstraps reproduce its sign with no alignment (99.99 %; bootstrap directions were sign-aligned for this panel only, not for the stability count). Its magnitude permutation test is only nominal once multiplicity is accounted for ($p = 0.034$; $p = 0.069$ Holm-adjusted for the two screened directions), so the derivation is exploratory. The acute-ischemia direction (orange) is diffuse ($\cos_\Sigma = 0.63$, 79 % sign-stable, $p = 0.84$), because ischemia moves the vector along the shared axis A rather than within the contrast subspace. Only the I_{Kr} direction was frozen; the atlas contains one transportable direction. **(b)** Projections of four active drugs on the frozen direction: the two I_{Kr} blockers separate from zero (dofetilide, used to derive the direction, +0.56, 95% CI 0.25 to 0.91; quinidine, a within-cohort held-mechanism confirmation, +0.83, 95% CI 0.52 to 1.20), while the non-I_{Kr} comparators do not (ranolazine +0.21, verapamil +0.12; intervals include zero). **(c)** Geometry: in the raw whitened coordinates, for every exam the signed score is bounded by the unsigned radius, $|S_{\mathrm{raw}}| \le D_{\mathrm{raw}}$ (the standardized $S_{I_{\mathrm{Kr}}}$ and D are separately z-scored, so the bound is a fact about the raw geometry, not the standardized scores). The radius retains contrast magnitude while discarding all angular orientation in the whitened contrast space; $S_{I_{\mathrm{Kr}}}$ retains one exploratory projection onto the perturbation-defined direction, one of the orientations the radius collapses. The frozen axis was carried into an independent external cohort (Chapman–Shaoxing/Ningbo), where it added conditional information about physician-assigned QT-interval extension beyond conventional intervals, A and D (fully adjusted odds ratio 1.23 per SD, $p = 1.9 \times 10^{-4}$; §2.5).

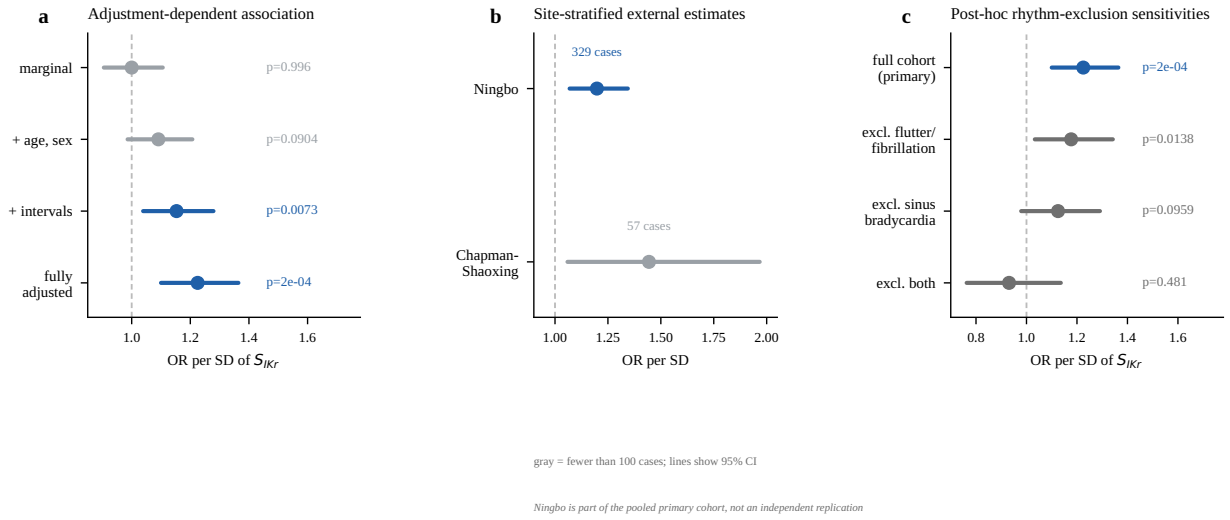

Figure 10: External-cohort transport of the frozen I_{Kr} direction (Chapman–Shaoxing/Ningbo, $n = 44\,550$, 386 QT-extension cases). (a) Adjustment-dependent association: the standardized I_{Kr} score is marginally null (OR 1.00, $p = 0.996$) and the odds ratio grows only as age, sex, conventional intervals, A and D are added, reaching a fully adjusted OR 1.23 per SD ($p = 1.9 \times 10^{-4}$), the model-dependent pattern of a suppressed coordinate. **(b)** Site-stratified external estimates; gray marks the site with fewer than 100 cases and lines show 95 % CIs. The pooled association is driven primarily by the larger Ningbo site, which is part of the combined primary cohort rather than an independent replication. **(c)** Post-hoc rhythm-exclusion sensitivities: the estimate holds when atrial flutter/fibrillation are excluded but attenuates to null when both flutter and sinus bradycardia are removed (86 cases), a limitation reported in full. Blue marks the frozen primary estimate; post-hoc exclusions are shown in a neutral color.

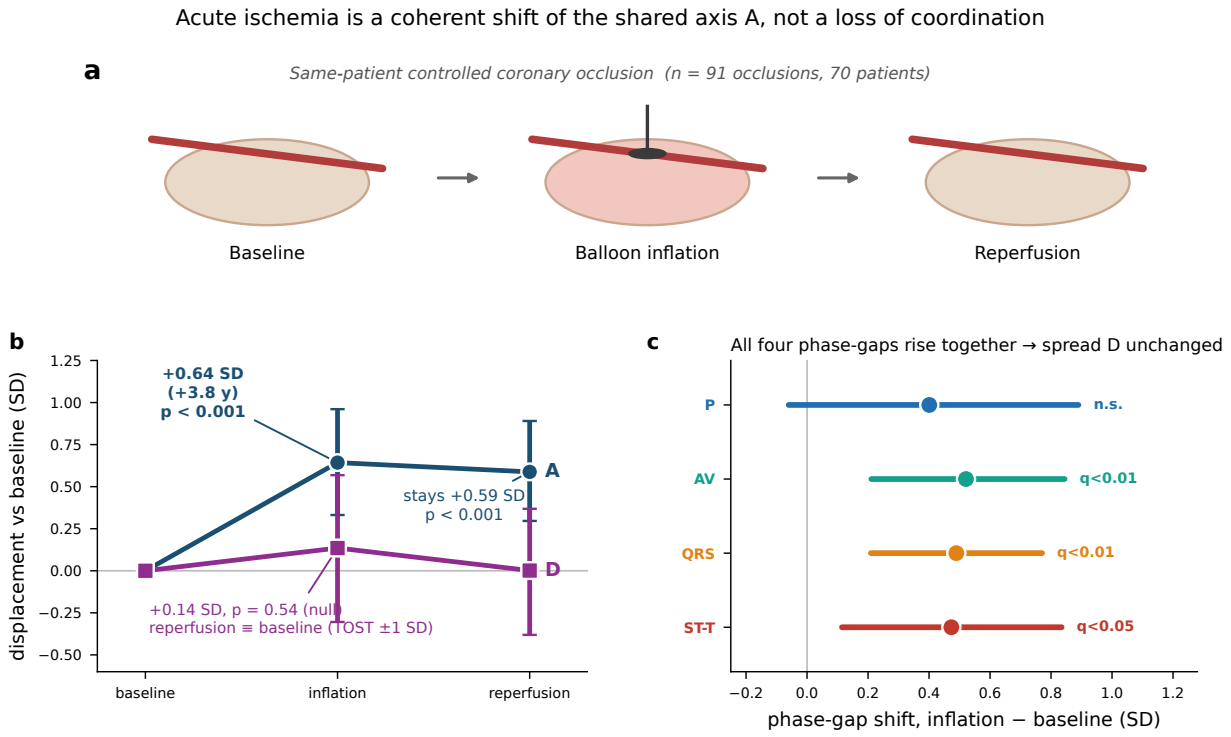

Figure 11: Acute ischemia is a coherent shift of the shared axis, not a loss of coordination. During controlled balloon coronary occlusion (STAFF III; 91 occlusions, 70 patients), all four phase clocks move in the same direction by similar amounts (AV +0.52 SD, QRS +0.49, ST-T +0.47, all false-discovery-rate significant; P +0.40, not significant). This coordinated displacement is a shift along the shared axis A (+0.64 SD, $p < 0.001$; whole-ECG apparent age +3.8 years) and was accompanied by no detectable change in the contrast-space disagreement radius D (+0.14 SD, $p = 0.54$). The shared-axis elevation persists into early reperfusion (+0.59 SD, $p < 0.001$); the corresponding D estimate remains imprecise rather than proving unchanged coordination. The displacement is an apparent shift of an age-trained readout under ischemia, not literal aging over minutes.

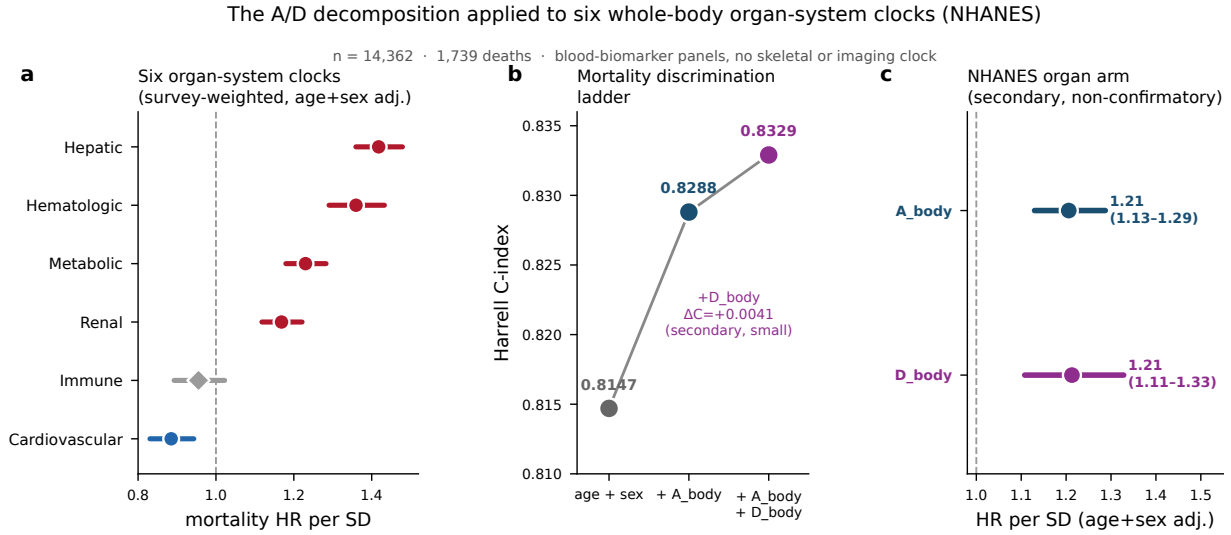

Figure 12: A scale-appropriate shared-axis/disagreement construction applied to six whole-body organ-system clocks (NHANES). All six systems are defined from standard blood-chemistry and clinical-assay panels; there is no skeletal or imaging clock in this mortality model. **(a)** Per-system mortality hazard ratios (per SD, survey-weighted, age- and sex-adjusted): the hepatic and hematologic clocks show the strongest positive mortality associations (hazard ratio 1.42 and 1.36 per SD), the cardiovascular panel runs protective after adjustment (0.89), and the immune panel is null (grey diamond). **(b)** The survey-weighted concordance ladder rises 0.8147→0.8288→0.8329 on adding the shared body axis A_{body} then the body disagreement radius D_{body}; the D_{body} increment is small and secondary. **(c)** In this cohort (14 362 participants, 1739 deaths) both A_{body} (hazard ratio 1.21 per SD, 95% CI 1.13 to 1.29) and D_{body} (1.21 per SD, 95% CI 1.11 to 1.33, $p = 3.1 \times 10^{-5}$, adjusted for age, sex and A_{body}) are associated with mortality. This is a *secondary, non-confirmatory* cross-scale arm: the NHANES mortality outcome was inspected in prior work, so this is not a substitute for, nor a rescue of, the electrical confirmatory CODE-15 test, which was null (§2.4). Its interest is the cross-scale contrast: an organ-system disagreement radius behaves differently from the electrical one.

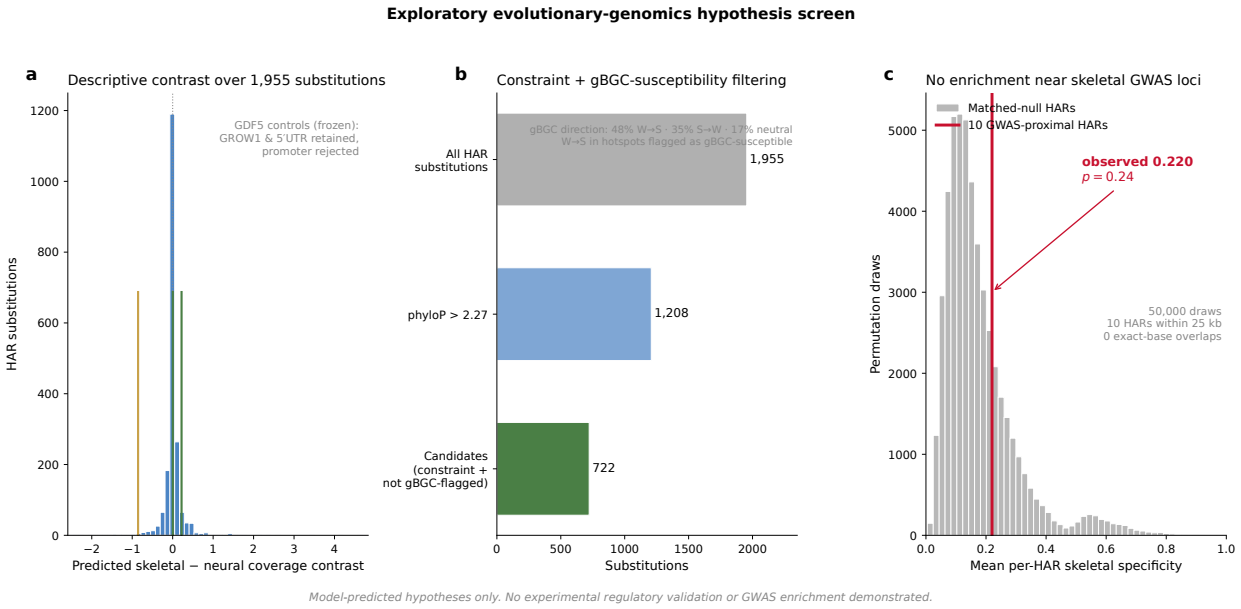

Figure 13: Exploratory evolutionary-genomics hypothesis screen of Human Accelerated Regions. A supporting analysis (evidence level 4); model-predicted hypotheses only, with no experimental regulatory validation and no demonstrated GWAS enrichment. **(a)** Distribution of the descriptive skeletal-minus-neural predicted-coverage contrast (derived versus ancestral allele) across the 1955 human-specific HAR substitutions. Positive values denote a predicted skeletal-lineage coverage gain, negative a predicted loss. The three frozen *GDF5*-locus filter-logic controls are marked: the promoter control (rs6060369) rejected as unconstrained and gBGC-favoured, and the GROW1 enhancer (rs4911178) and *GDF5* 5′UTR (rs143384) controls retained; the retained controls validate the pass/reject filter logic only, not the magnitude or sign of the predicted contrast. **(b)** Filtering flow: the 1955 substitutions are reduced to 722 candidates by requiring both per-base constraint (phyloP > 2.27; 1208 pass) and a substitution not flagged as gBGC-susceptible by the allelic-direction rule (gBGC directions: 48 % weak-to-strong, 35 % strong-to-weak, 17 % neutral). **(c)** The central inferential result was null. The ten HARs within 25 kb of an osteoarthritis or bone-mineral-density credible-set variant (0 exact-base overlaps) are not more skeletal-specific than size- and substitution-matched HARs: observed mean 0.220 versus a matched permutation null (50 000 draws), $p = 0.24$. Ranked top candidates are descriptive hypotheses and are shown in Supplementary Fig. S9.

Data availability

CODE-15 is publicly available (Zenodo record 4916206; CC BY 4.0). SaMi-Trop 12-lead ECG traces with age and mortality annotations are openly available (Zenodo record 4905618). PTB-XL, LUDB, ECGRDVQ and STAFF III are available from PhysioNet under their respective licenses. The Chapman–Shaoxing and Ningbo databases used for the external phenotype transport test are openly available (PhysioNet *ecg-arrhythmia* 1.0.0); the frozen I_{Kr} scorer, protocol lock and the QT-extension result table are provided as versioned artifacts. EchoNext (Elias & Finer) is available under its data-use agreement; the bytes analyzed here were staged from the study team's access-controlled private repository and were not redistributed. NHANES and the NCHS Linked Mortality File are US public-domain data. The Aging Brain Cohort (AABC / HCP-Aging Release 2) is available under registered academic access from BALSA [34, 35]; participant-level AABC data were analyzed only within an approved, access-controlled environment and were not redistributed. Public artifacts are limited to code, frozen-geometry hashes, cohort-level summaries and disclosure-checked aggregate tables. The LEMON dataset used for the anatomical figures and for the image-derived disagreement analysis (§2.9) is openly available [29]. Raw T1 volumes are not redistributed; the per-subject four-view feature table, aggregate sensitivity analyses and statistical result are provided as versioned artifacts. Frozen score definitions, calibration constants and result tables are provided as versioned artifacts. A claim-to-artifact ledger maps every headline quantity in this manuscript to the specific versioned result file it was read from. Per-visit AABC standardized channel scores remain inside the approved controlled-access environment.

The TotalSegmentator CT dataset v2.0.1, comprising 1228 examinations and segmentations of 117 anatomical structures, is openly available under CC BY 4.0 from Zenodo (record 10047292; doi:10.5281/zenodo.10047292). Source images and masks were not redistributed; reference examination s1045 supplies visualization geometry only. The current release contains disclosure-safe aggregate CT performance and ranking tables, methods, provenance receipts and static figures, but not the scan-level feature matrix, cross-validation folds, source images or masks. The CT atlas remains exploratory because it is cross-sectional and volume-only, not because its aggregate results are unavailable.

Code availability

The frozen phase clocks, the BodyVector geometry (shared axis and disagreement radius), the outcome-blind firewall, and all analysis and figure code are released under the MIT license at `https://github.com/jim4226/ClaudeScienceHackathon_2026` (subdirectory `electrical-body-clock/`), as content-hashed, versioned artifacts; the confirmatory reveal script is hash-locked, and its pre-specified inputs and success criterion were verified by a fresh-context reviewer with no access to the outcome labels. The frozen signed-perturbation direction is released as a protocol lock carrying a canonical content hash and SHA-256 digests of the contrast basis, the contrast covariance, the input feature tables and the scorer, together with the scorer module, its fixture and an independent verifier; the executed external transport test (Chapman–Shaoxing/Ningbo QT-extension phenotype) is included in full. An earlier mortality-transport protocol (MIMIC) was specified but abandoned before execution and is not part of the reported results.

Acknowledgments

The investigators and participants of the CODE-15, SaMi-Trop, PTB-XL, LUDB, ECGRDVQ, STAFF III, Chapman–Ningbo, EchoNext, NHANES, AABC / HCP-Aging and LEMON studies are gratefully acknowledged.

Author contributions

Conceptualization, methodology and study design; software and frozen-pipeline engineering; formal analysis and statistical modeling; data curation and access governance; visualization; writing: original draft; and writing: review and editing. The author had full access to all data in the study and takes responsibility for its integrity and the accuracy of the analysis.

Ethics and consent

This study is a secondary analysis of previously collected, de-identified data. Each source cohort was approved by its originating institutional review board with participant informed consent obtained by the primary investigators: CODE-15, SaMi-Trop, Chapman–Ningbo, LUDB, PTB-XL, ECGRDVQ and STAFF III are released through Zenodo or PhysioNet under their respective data-use agreements; NHANES and the NCHS Linked Mortality File were approved by the NCHS Research Ethics Review Board; EchoNext was analyzed under its data-use agreement; and the Aging Brain Cohort (AABC / HCP-Aging Release 2) was analyzed under an approved controlled-access data-use agreement within a secure environment. No new human-subjects data were collected for this work; no individual-level or identifiable data are reported.

Competing interests

The author declares no competing interests.

References

[1] Emilly M. Lima, Antônio H. Ribeiro, Gabriela M. M. Paixão, Manoel Horta Ribeiro, Marcelo M. Pinto-Filho, Paulo R. Gomes, Derick M. Oliveira, Ester C. Sabino, Bruce B. Duncan, Luana Giatti, Sandhi M. Barreto, Wagner Meira Jr, Thomas B. Schön, and Antonio Luiz P. Ribeiro. Deep neural network-estimated electrocardiographic age as a mortality predictor. *Nature Communications*, 12(1):5117, 2021. doi: 10.1038/s41467-021-25351-7.

[2] Zachi I. Attia, Paul A. Friedman, Peter A. Noseworthy, Francisco Lopez-Jimenez, Dorothy J. Ladewig, Gaurav Satam, Patricia A. Pellikka, Thomas M. Munger, Samuel J. Asirvatham, Christopher G. Scott, Rickey E. Carter, and Suraj Kapa. Age and sex estimation using artificial intelligence from standard 12-lead ECGs. *Circulation: Arrhythmia and Electrophysiology*, 12 (9):e007284, 2019. doi: 10.1161/CIRCEP.119.007284.

[3] Ye Ella Tian, Vanessa Cropley, Andrea B. Maier, Nicola T. Lautenschlager, Michael Breakspear, and Andrew Zalesky. Heterogeneous aging across multiple organ systems and prediction of chronic disease and mortality. *Nature Medicine*, 29(5):1221–1231, 2023. doi: 10.1038/s41591-023-02296-6.

[4] Hamilton Se-Hwee Oh, Jarod Rutledge, Daniel Nachun, Róbert Pálovics, Olamide Abiose, Patricia Moran-Losada, Divya Channappa, Deniz Yagmur Urey, Kate Kim, Yun Ju Sung, Lihua Wang, Jigyasha Timsina, Dan Western, Menghan Liu, Pat Kohlfeld, John Budde, Edward N. Wilson, Yann Guen, Taylor M. Maurer, Michael Haney, Andrew C. Yang, Zihuai He, Michael D. Greicius, Katrin I. Andreasson, Sanish Sathyan, Erica F. Weiss, Sofiya Milman, Nir Barzilai, Carlos Cruchaga, Anthony D. Wagner, Elizabeth Mormino, Benoit Lehallier, Victor W. Henderson, Frank M. Longo, Stephen B. Montgomery, and Tony Wyss-Coray. Organ aging signatures in the plasma proteome track health and disease. *Nature*, 624(7990):164–172, 2023. doi: 10.1038/s41586-023-06802-1.

[5] Alan A. Cohen, Emmanuel Milot, Jian Yong, Christopher L. Seplaki, Tamàs Fülöp, Karen Bandeen-Roche, and Linda P. Fried. A novel statistical approach shows evidence for multisystem physiological dysregulation during aging. *Mechanisms of Ageing and Development*, 134 (3-4):110–117, 2013. doi: 10.1016/j.mad.2013.01.004.

[6] Junhao Wen et al. The genetic architecture of multimodal human brain age. *Nature Communications*, 15:2604, 2024. doi: 10.1038/s41467-024-46796-6.

[7] Ye Ella Li et al. Synergistic and heterogeneous aging using composite phenotypes and multiorgan clocks. *Communications Medicine*, 5, 2025. doi: 10.1038/s43856-025-01222-w.

[8] Philip Hempel et al. Explainable AI associates ECG aging effects with increased cardiovascular risk in a longitudinal population study. *npj Digital Medicine*, 8:25, 2025. doi: 10.1038/s41746-024-01428-7.

[9] Meng Hao et al. Human aging reflects increases in entropy across organ networks. *medRxiv*, 2025. doi: 10.1101/2025.05.06.25327108. Preprint.

[10] Adetola O. Ladejobi, Jose R. Medina-Inojosa, Michal Shelly Cohen, Zachi I. Attia, Christopher G. Scott, Nathan K. LeBrasseur, Bernard J. Gersh, Peter A. Noseworthy, Paul A. Friedman, Suraj Kapa, and Francisco Lopez-Jimenez. The 12-lead electrocardiogram as a biomarker of biological age. *European Heart Journal – Digital Health*, 2(3):379–389, 2021. doi: 10.1093/ehjdh/ztab043.

[11] Antônio H. Ribeiro, Gabriela M. M. Paixão, Emilly M. Lima, Manoel Horta Ribeiro, Marcelo M. Pinto Filho, Paulo R. Gomes, Derick M. Oliveira, Wagner Meira Jr, Thomas B. Schön, and Antonio Luiz P. Ribeiro. CODE-15%: a large scale annotated dataset of 12-lead ECGs. Zenodo, 2021. CC BY 4.0.

[12] Pierre Elias and Joshua Finer. EchoNext: A dataset for detecting echocardiogram-confirmed structural heart disease from ECGs. PhysioNet, 2026. Version 1.1.1.

[13] Timothy J. Poterucha, Linyuan Jing, Rachel P. Ricart, Medina Adjei-Mosi, Joshua Finer, Dustin Hartzel, Christopher Kelsey, Aaron Long, Daniel Rocha, Jeffrey A. Ruhl, David van Maanen, Marc A. Probst, Brock Daniels, Sneha D. Joshi, Olivier Tastet, Daniel Corbin, Robert Avram, John P. Barrios, Geoffrey H. Tison, I-Min Chiu, David Ouyang, Alexander Volodarskiy, Christopher M. Haggerty, and Pierre Elias. Detecting structural heart disease from electrocardiograms using AI. *Nature*, 644(8075):221–230, 2025. doi: 10.1038/s41586-025-09227-0.

[14] Jose Vicente, Lars Johannesen, and David G. Strauss. ECG effects of ranolazine, dofetilide, verapamil, and quinidine in healthy subjects. PhysioNet, 2016.

[15] Lars Johannesen, Jose Vicente, Jay W. Mason, Carlos Sanabria, Kristin Waite-Labott, Mira Hong, Ping Guo, Jay Lin, Jens Stampe Sørensen, Loriano Galeotti, Jeffry Florian, Martin Ugander, Norman Stockbridge, and David G. Strauss. Differentiating drug-induced multichannel block on the electrocardiogram: Randomized study of dofetilide, quinidine, ranolazine, and verapamil. *Clinical Pharmacology & Therapeutics*, 96(5):549–558, 2014. doi: 10.1038/clpt.2014.155.

[16] Juan Pablo Martínez, Olle Pahlm, Michael Ringborn, Stafford Warren, Pablo Laguna, and Leif Sörnmo. STAFF III database. PhysioNet, 2017.

[17] Juan Pablo Martínez, Olle Pahlm, Michael Ringborn, Stafford Warren, Pablo Laguna, and Leif Sörnmo. The STAFF III database: ECGs recorded during acutely induced myocardial ischemia. *Computing in Cardiology*, 44:1–4, 2017. doi: 10.22489/CinC.2017.266-133.

[18] Jianwei Zheng, Hangyuan Guo, and Huimin Chu. A large scale 12-lead electrocardiogram database for arrhythmia study (version 1.0.0). PhysioNet, 2022. CC BY 4.0.

[19] National Center for Health Statistics (NCHS). National health and nutrition examination survey (nhanes) 2005–2010, linked to the NCHS 2019 public-use linked mortality file. Centers for Disease Control and Prevention, 2019. `https://www.cdc.gov/nchs/nhanes/`; `https://www.cdc.gov/nchs/data-linkage/mortality.htm`.

[20] Clareci Silva Cardoso, Ester Cerdeira Sabino, Claudia Di Lorenzo Oliveira, Lea Campos de Oliveira, Ariela Mota Ferreira, Edecio Cunha-Neto, Ana Luiza Bierrenbach, Antônio Luiz Pinho Ribeiro, et al. Longitudinal study of patients with chronic chagas cardiomyopathy in brazil (sami-trop project): a cohort profile. *BMJ Open*, 6(5):e011181, 2016. doi: 10.1136/bmjopen-2016-011181.

[21] Antônio H. Ribeiro, Gabriela M. M. Paixão, Emilly M. Lima, Manoel Horta Ribeiro, Marcelo M. Pinto Filho, Paulo R. Gomes, Derick M. Oliveira, Ester C. Sabino, Bruce B. Duncan, Luana Giatti, Sandhi M. Barreto, Wagner Meira Jr, Thomas B. Schön, and Antonio Luiz P. Ribeiro. Sami-trop: 12-lead ecg traces with age and mortality annotations, 2021.

[22] Jakob Wasserthal, Hanns-Christian Breit, Manfred T. Meyer, Maurice Pradella, Daniel Hinck, Alexander W. Sauter, Tobias Heye, Daniel T. Boll, Joshy Cyriac, Shan Yang, Michael Bach, and Martin Segeroth. TotalSegmentator: Robust segmentation of 104 anatomic structures in CT images. *Radiology: Artificial Intelligence*, 5(5):e230024, 2023. doi: 10.1148/ryai.230024.

[23] Kathleen C. Keough, Sean Whalen, Fumitaka Inoue, Pawel F. Przytycki, Tyler Fair, Chengyu Deng, Marilyn Steyert, Hane Ryu, Kerstin Lindblad-Toh, Elinor Karlsson, Tomasz Nowakowski, Nadav Ahituv, and Katherine S. Pollard. Three-dimensional genome rewiring in loci with human accelerated regions. *Science*, 380(6643):eabm1696, 2023. doi: 10.1126/science.abm1696.

[24] Johannes Linder, Divyanshi Srivastava, Han Yuan, Vikram Agarwal, and David R. Kelley. Predicting RNA-seq coverage from DNA sequence as a unifying model of gene regulation. *Nature Genetics*, 57:949–961, 2025. doi: 10.1038/s41588-024-02053-6.

[25] Matthew J. Christmas, Irene M. Kaplow, Diane P. Genereux, Michael X. Dong, Graham M. Hughes, Xue Li, Patrick F. Sullivan, Allyson G. Hindle, Gregory Andrews, Joel C. Armstrong, et al. Evolutionary constraint and innovation across hundreds of placental mammals. *Science*, 380(6643):eabn3943, 2023. doi: 10.1126/science.abn3943.

[26] Laurent Duret and Nicolas Galtier. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annual Review of Genomics and Human Genetics*, 10:285–311, 2009. doi: 10.1146/annurev-genom-082908-150001.

[27] Terence D. Capellini, Hao Chen, Jiaxue Cao, Andrew C. Doxey, Ata M. Kiapour, Michael Schoor, and David M. Kingsley. Ancient selection for derived alleles at a GDF5 enhancer influencing human growth and osteoarthritis risk. *Nature Genetics*, 49(8):1202–1210, 2017. doi: 10.1038/ng.3911.

[28] Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411, 2004. doi: 10.1016/S0047-259X(03)00096-4.

[29] Anahit Babayan, Miray Erbey, Deniz Kumral, et al. A mind-brain-body dataset of mri, eeg, cognition, emotion, and peripheral physiology in young and old adults. *Scientific Data*, 6:180308, 2019. doi: 10.1038/sdata.2018.308.

[30] Nicholas J Tustison, Philip A Cook, Andrew J Holbrook, Hans J Johnson, John Muschelli, Gabriel A Devenyi, Jeffrey T Duda, Sandhitsu R Das, Nicholas C Cullen, Daniel L Gillen, et al. The ANTsX ecosystem for quantitative biological and medical imaging. *Scientific Reports*, 11(1):9068, 2021. doi: 10.1038/s41598-021-87564-6.

[31] Vishnu M Bashyam, Guray Erus, Jimit Doshi, Mohamad Habes, Ilya M Nasrallah, Monica Truelove-Hill, Dhivya Srinivasan, Liz Mamourian, Raymond Pomponio, Yong Fan, et al. MRI signatures of brain age and disease over the lifespan based on a deep brain network and 14,468 individuals worldwide. *Brain*, 143(7):2312–2324, 2020. doi: 10.1093/brain/awaa160.

[32] Ann-Marie G. de Lange and James H. Cole. Commentary: Correction procedures in brain-age prediction. *NeuroImage: Clinical*, 26:102229, 2020. doi: 10.1016/j.nicl.2020.102229.

[33] Alan Le Goallec, Samuel Diai, Sasha Collin, Jean-Baptiste Prost, Théo Vincent, and Chirag J. Patel. Using deep learning to predict abdominal age from liver and pancreas magnetic resonance images. *Nature Communications*, 13:1979, 2022. doi: 10.1038/s41467-022-29525-9.

[34] Susan Y. Bookheimer, David H. Salat, Melissa Terpstra, et al. The lifespan human connectome project in aging: An overview. *NeuroImage*, 185:335–348, 2019. doi: 10.1016/j.neuroimage.2018.10.009.

[35] Michael P. Harms, Leah H. Somerville, Beau M. Ances, et al. Extending the human connectome project across ages: Imaging protocols for the lifespan development and aging projects. *NeuroImage*, 183:972–984, 2018. doi: 10.1016/j.neuroimage.2018.09.060.

Supplementary Information

Supplementary figures S1–S3 document the electrical measurement pipeline (delineation accuracy, cross-cohort transfer and exploratory per-phase effect sizes); S4–S8 cover the cross-scale pilots; S9 collects the post-hoc electrical construct-validity panels.

S1 Delineation accuracy (LUDB)

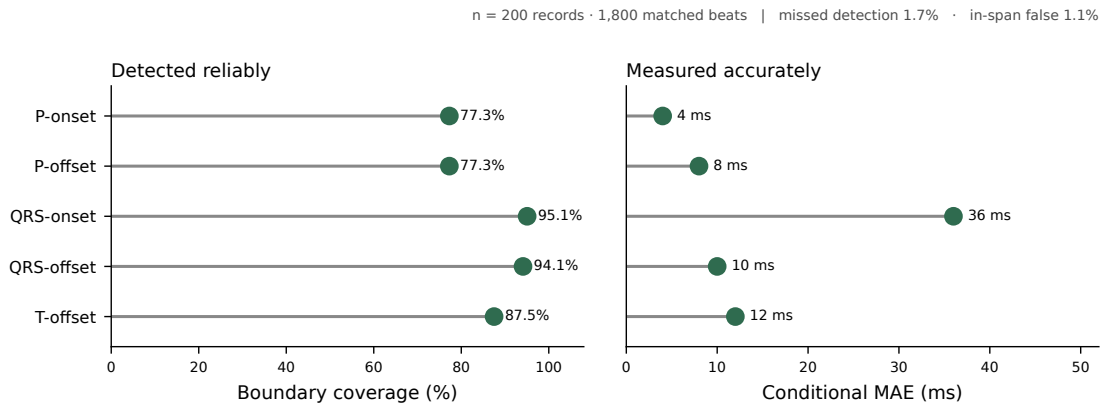

Figure S1: Delineation accuracy sets the measurement floor (LUDB). The wavelet delineator was validated against cardiologist annotations in the Lobachevsky University database (200 records, 1800 matched beats) with no retuning. Left, per-boundary detection coverage; right, conditional mean absolute error against the reference annotation. QRS onset is placed systematically early (36 ms); all other boundaries are within 12 ms. Missed-beat rate 1.7%, in-span false-detection rate 1.1%.

S2 Cross-cohort transfer of the phase clocks (EchoNext)

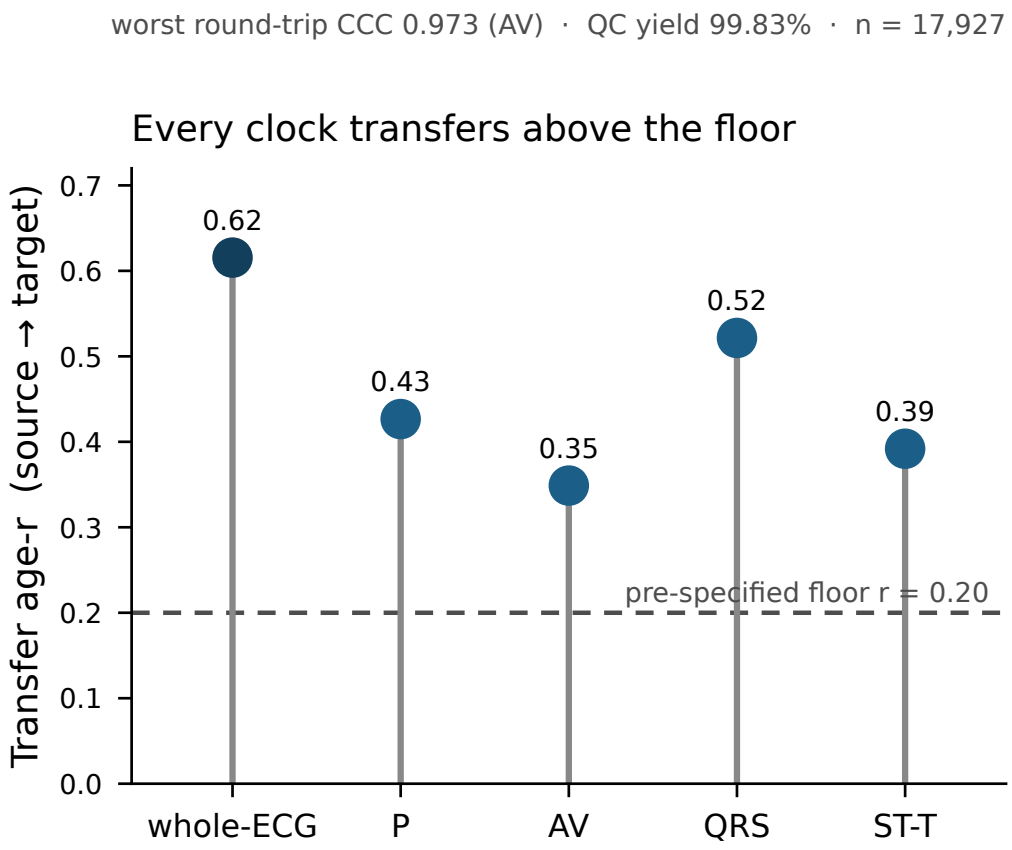

Figure S2: Every phase clock transfers across cohorts (EchoNext). Correlation between source-trained and target-cohort predicted age for the whole-ECG clock and each of the four phase clocks (n = 17 927). All five exceed the pre-specified transfer floor of $r = 0.20$ (whole-ECG 0.62, QRS 0.52, P 0.43, ST-T 0.39, AV 0.35). Worst round-trip concordance correlation 0.973 (AV); QC yield 99.83%.

S3 Exploratory per-phase effect sizes in structural disease (EchoNext)

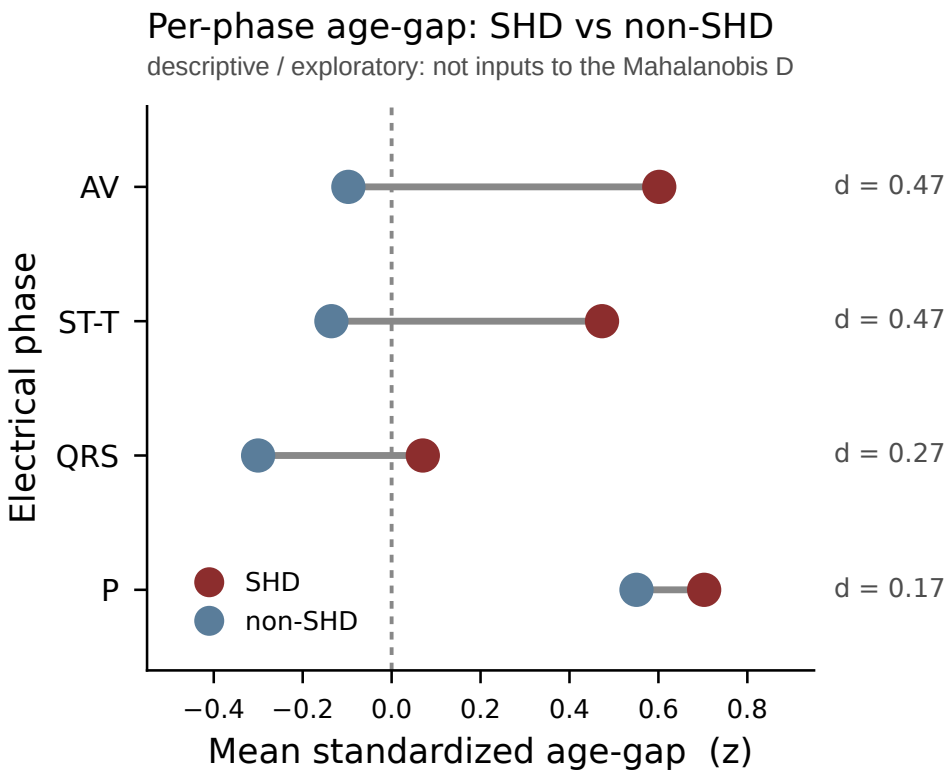

Figure S3: Per-phase age gaps in structural disease (EchoNext, exploratory). Mean standardized age gap for each phase in participants with versus without echocardiographically confirmed structural heart disease, ordered by effect size. The AV and ST-T phases separate the groups most (Cohen's $d = 0.47$ each), the atrial phase least ($d = 0.17$). This descriptive per-phase view is exploratory and is not an input to the Mahalanobis disagreement radius D; it motivates the phase-selective, externally testable hypothesis discussed in the main text.

S4 Structure–function pilot (MotionVector): a technically limited probe

As an exploratory probe of whether the shared-axis/orthogonal-residual construction might extend to a structure–function pair, skeletal aging (whole-body DXA) versus seven-day physical activity (accelerometry), an analogous decomposition was built in the NHANES 2005–2006 participants with both modalities ($n = 2891$; Methods). The prespecified primary test, whether a calibration-orthogonal "locomotor reserve" predicts lower-body physical limitation beyond age, sex, BMI and the shared axis, did not reach significance: odds ratio 1.10 per contrast unit (95% CI 0.99 to 1.23), likelihood-ratio $p = 0.070$, in the non-protective direction.

This result is not interpreted as a validated boundary finding. The pilot has two processing limitations that must be fixed before any conclusion is drawn (a rebuilt V2 is future work): (i) the accelerometry features were computed from raw minute-level intensity counts without a validated non-wear/wear-time algorithm or valid-day filtering, so they are not comparable to accepted NHANES accelerometry processing; and (ii) the five multiply-imputed DXA replicates were averaged into one record rather than analyzed separately and combined by Rubin's rules, understating variance contrary to NHANES guidance. The implemented locomotor feature set was also only a subset of the locked list (cadence-distribution, walking-bout, fragmentation and cosinor-amplitude features were not computed and, the outcome now being unblinded, were deliberately not added post hoc), and the outcome is self-reported in a young cohort (ages 20–69).

With those caveats, the descriptive pattern (Fig. S4) is the following. The two age gaps were nearly uncorrelated ($r = 0.07$; Fig. S4a). The locomotor age clock was a weak age predictor (out-of-fold $R^2 = 0.17$) whose gap correlated *positively* with daily steps ($r = +0.30$; Fig. S4b), more activity mapping to an older apparent locomotor age in this age range, opposite to the protective association of activity with function. Post-hoc, the raw ingredients were associated with limitation in the expected directions (skeletal age gap odds ratio 1.14 per SD, $p = 0.019$; raw daily step count odds ratio 0.81 per 1000 steps, $p < 1 \times 10^{-32}$; Fig. S4c). These exploratory associations were computed after the primary outcome was unblinded and are labeled as such. The pilot does not alter the main-text A-versus-D findings and is independent of the CODE-15 confirmatory test.

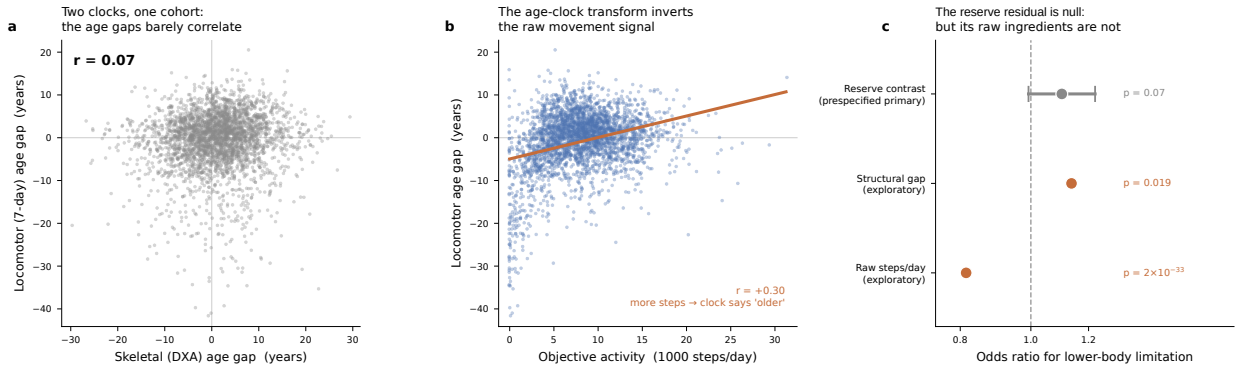

Figure S4: Structure–function pilot (NHANES 2005–2006, $n = 2891$): a technically limited probe, not a validated result. Descriptive patterns only; see caveats in §7.13. **(a)** The skeletal (DXA) and locomotor (seven-day accelerometry) age gaps are nearly uncorrelated ($r = 0.07$). **(b)** The locomotor age-gap correlates positively with daily steps ($r = +0.30$), more activity mapping to an older apparent locomotor age, opposite to the protective role of activity. **(c)** The prespecified "locomotor reserve" contrast does not reach significance for lower-body limitation (grey; odds ratio 1.10 per unit, $p = 0.070$, confidence interval crossing 1); the post-hoc raw-ingredient associations (orange; skeletal gap $p = 0.019$, raw steps $p < 1 \times 10^{-32}$) were computed after unblinding and are exploratory. Accelerometry was processed without a validated wear-time algorithm and the DXA imputations were averaged, so these patterns are provisional.

S5 Brain-imaging decomposition (NeuroMotionVector)

As a prospective test of the decomposition at a third biological scale, the same shared-direction/contrast-space construction was applied to four brain-imaging channels in the Aging Brain Cohort (AABC / HCP-Aging Release 2) with a longitudinal gait outcome and a hash-locked, outcome-blind protocol (§7.9; §2.9). The prespecified test found no detectable association between baseline brain-channel disagreement and subsequent four-metre walk-performance change (Fig. 1). The four brain-age clocks nonetheless recovered age on held-out data (r 0.50 to 0.59), and the scalar shared-axis and disagreement summaries were empirically nearly uncorrelated. This scalar correlation is descriptive; the construction makes the shared direction orthogonal to the contrast subspace, not necessarily scalar A orthogonal to scalar D. The null concerns the tested disagreement–function link specifically, not technical clock constructability.

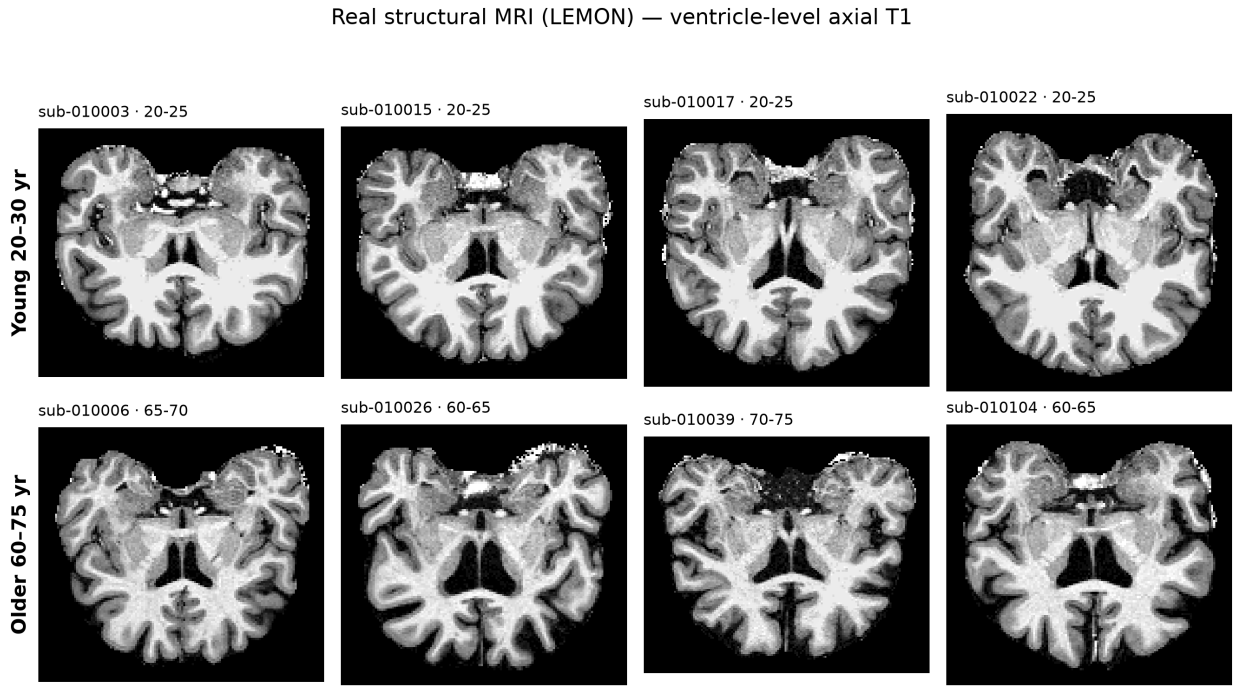

Figure S5: The visible substrate of brain aging (real structural MRI, LEMON). Ventricle-level axial T1 slices for younger (20–30 y) versus older (60–75 y) adults from the open-access LEMON dataset (skull-stripped MP2RAGE, 256^3 at 1 mm). Older brains show wider sulci and larger ventricles. These open-data images illustrate anatomy; no controlled-access AABC scan is shown anywhere in this work.

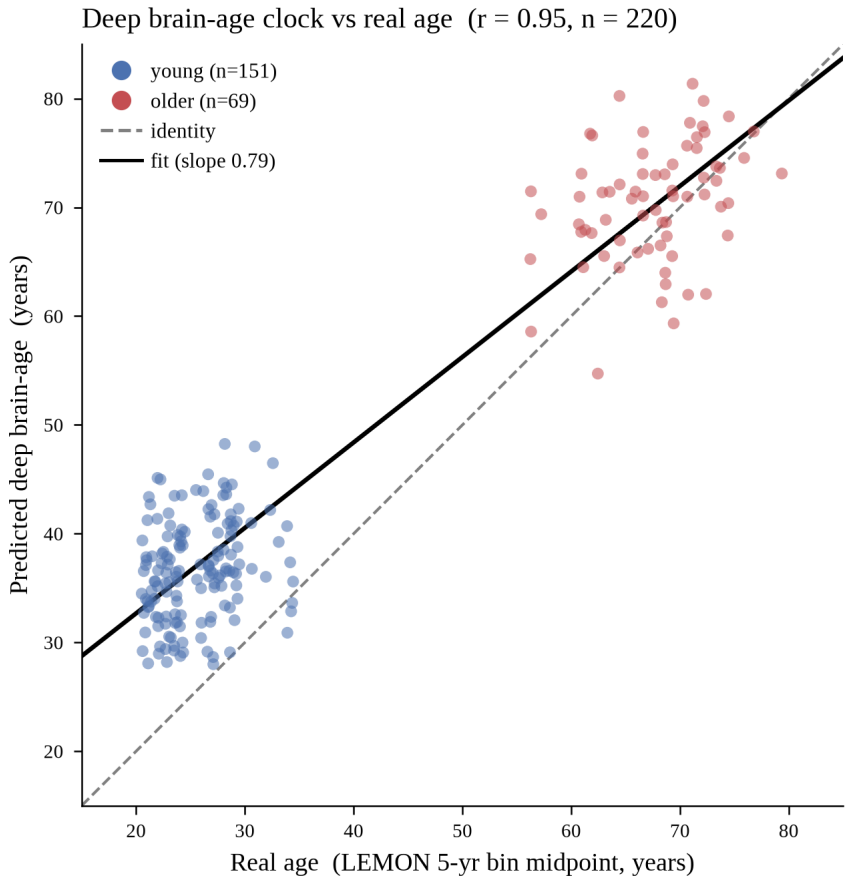

Figure S6: Deep brain-age clock validation and its domain-shift offset (LEMON). Predicted deep brain-age versus real age (LEMON 5-year bin midpoints) for all 220 scans. The clock tracks age closely (Pearson $r = 0.95$) but with a compressed regression slope (0.79) and a positive offset relative to the identity line, a textbook domain-shift bias from applying a conventional-T1w-trained model to MP2RAGE data. Because absolute brain-age is therefore not comparable across the acquisition shift, the image-derived dispersion analysis (§2.9, Fig. 2) is specified on the younger-versus-older *difference* in cross-view dispersion rather than on absolute values.

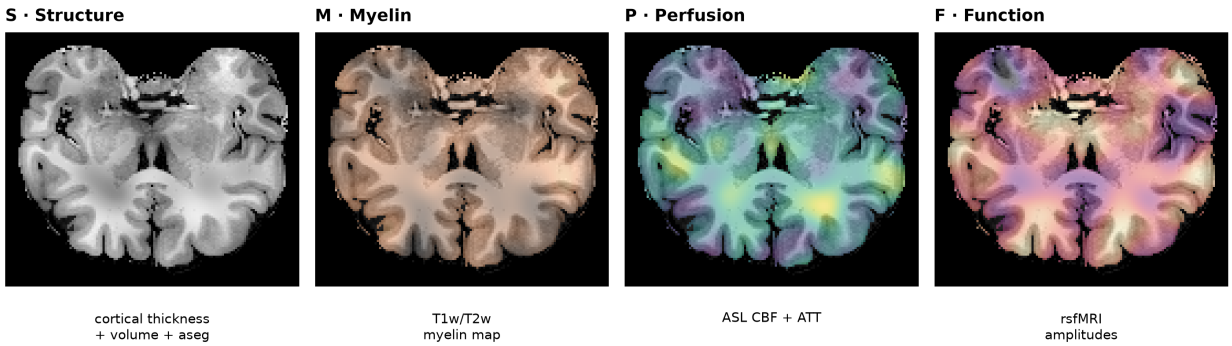

Figure S7: Four imaging channels, one brain. The four brain-aging views used by NeuroMotionVector, cortical structure, T1w/T2w myelin, arterial spin-labeling perfusion, and resting-state functional amplitude, shown as schematic overlays on a real LEMON T1 slice. The overlays are schematic, not measurements from the displayed participant and not controlled-access AABC scans. Each AABC channel is read by a separately trained age clock; their standardized age gaps form the shared axis and disagreement radius.

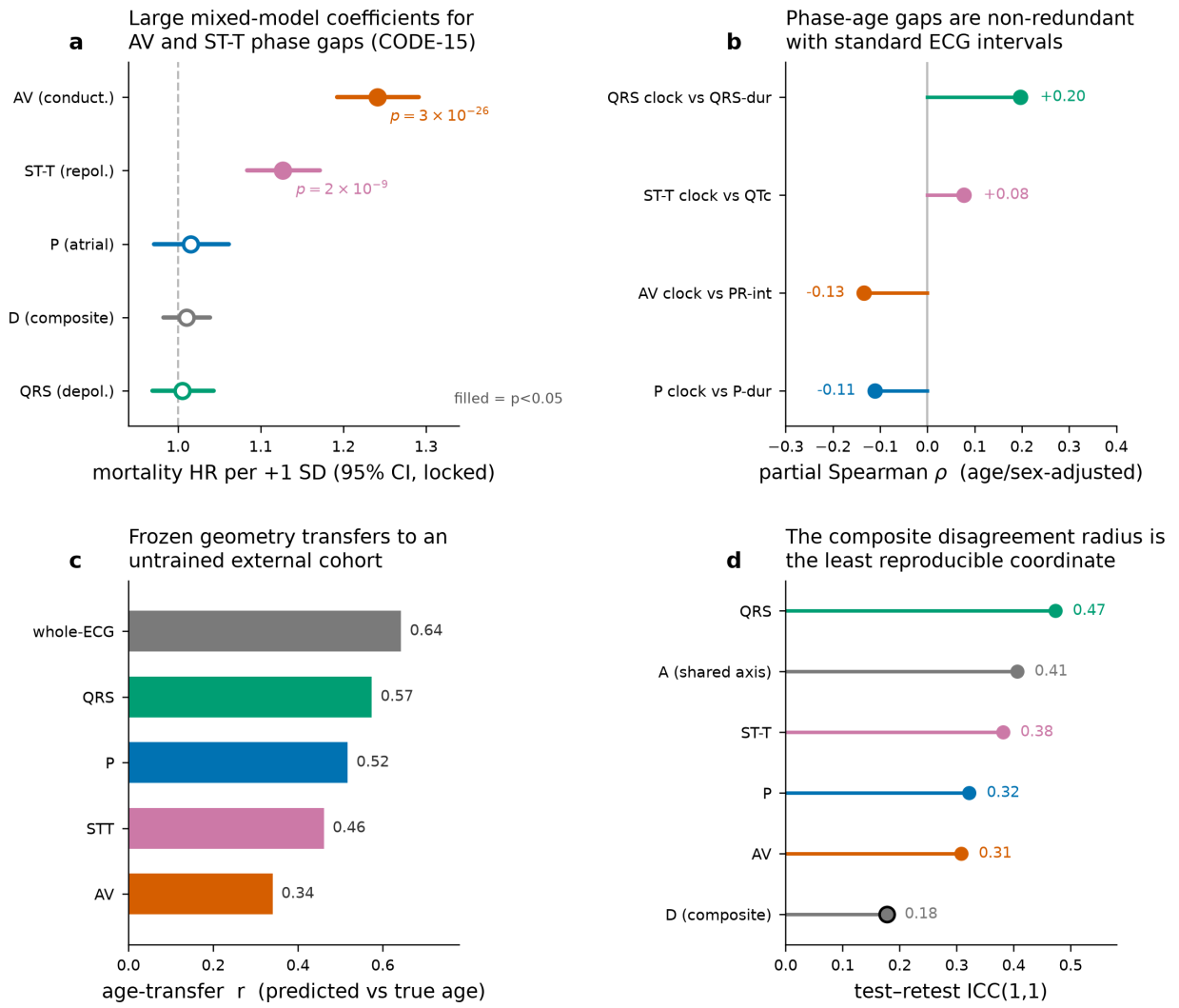

Figure S8: Post-hoc construct-validity, transfer and reproducibility of the electrical decomposition. All panels use the frozen scoring pipeline with no refitting and were computed after the confirmatory reveal. **(a)** In a mixed shared-plus-contrast Cox model on the locked CODE-15 final split, the atrioventricular (AV, HR 1.24, $p = 3 \times 10^{-26}$) and ST-T repolarization (HR 1.13, $p = 2 \times 10^{-9}$) phase-age gaps carry large coefficients, while the atrial (P), ventricular-depolarization (QRS) and composite disagreement radius (D) coordinates are null (filled markers $p < 0.05$; numbers re-presented from the locked reveal, not recomputed). Because the four gaps jointly span the shared axis A, this model does not isolate phase-specific effects at fixed A, and the external identifiable test (SaMi-Trop) was null; these are therefore hypotheses rather than established effects. **(b)** After adjusting for age and sex, each predicted phase age is only weakly related to its matched standard interval (partial Spearman ρ; CODE-15 calibration, $n = 74\,715$), i.e. the phase clocks are non-redundant with standard intervals rather than carrying incremental outcome information; the AV clock is weakly *negative* against PR interval, so its signal is not classical PR prolongation. **(c)** The frozen geometry transfers to the untrained external Chapman–Shaoxing/Ningbo cohort ($n = 44\,595$ with usable age): predicted-vs-chronological age r for the whole-ECG clock and each phase clock (no mortality follow-up available, so this shows age-structure generality only). **(d)** Within-patient test–retest reproducibility (ICC$(1,1)$; 66 474 patients, 177 336 exams): the composite disagreement radius D is the least reproducible coordinate (0.18), below the shared axis A (0.41) and every single-phase gap, a mechanistic complement to the confirmatory null.

S10 Ranked skeletal-lineage HAR candidates (descriptive)

The main-text genomic figure (Fig. 13) reports the screen's distribution, filtering flow and the central GWAS-proximity null. The ranked candidate list below is provided for completeness as a set of model-predicted hypotheses for experimental follow-up; the ranking is descriptive and no statistical significance is assigned to any individual candidate.

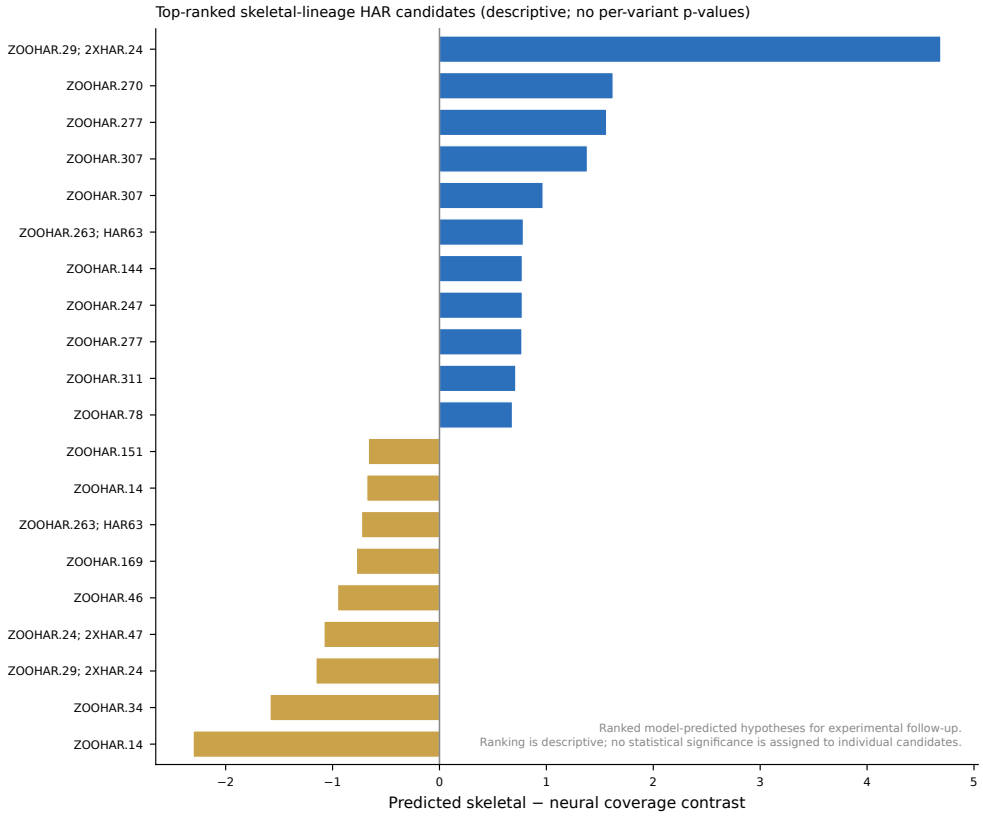

Figure S9: Top-ranked skeletal-lineage HAR candidates from the exploratory genomics screen. The twenty substitutions with the largest predicted skeletal-minus-neural coverage contrast, labelled by ZOOHAR identifier; bar direction gives the sign of the predicted change. These are descriptive model-predicted hypotheses with no per-variant p-values and no experimental validation; they are not claimed as regulatory findings.