Systems Age: A single blood methylation test to quantify aging heterogeneity across 11 physiological systems

Raghav Sehgal[1], Margarita Meer[2], Aladdin H. Shadyab[3], Ramon Casanova[4], JoAnn E. Manson[5], Parveen Bhatti[6,7], Eileen M. Crimmins[8], Themistocles L. Assimes[9], Eric A. Whitsel[10], Albert T. Higgins-Chen[11,12], Morgan Levine[2,12*]

[*] Corresponding author
mlevine@altoslabs.com

[1] Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA
[2] Altos Labs, San Diego Institute of Sciences, San Diego, CA, USA
[3] Herbert Wertheim School of Public Health and Human Longevity Science, University of California, San Diego, La Jolla, CA, USA
[4] Department of Biostatistics and Data Science, School of Medicine, Wake Forest University, Winston-Salem, NC 27157, USA
[5] Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[6] Cancer Control Research, BC Cancer, Vancouver, BC, Canada.
[7] School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada
[8] Leonard Davis School of Gerontology, University of Southern California, Los Angeles, CA, USA
[9] Department of Medicine (Division of Cardiovascular Medicine) and Epidemiology and Population Health, Stanford University School of Medicine, Stanford, CA, USA
[10] Department of Epidemiology, Gillings School of Public Health and Department of Medicine, School of Medicine, University of North Carolina, Chapel Hill, NC, USA
[11] Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA
[12] Department of Pathology, Yale University School of Medicine, New Haven, CT, USA

Keywords: aging; biomarker; epigenetic clock; organ aging; aging subtypes

Abstract

Individuals, organs, tissues, and cells age in diverse ways throughout the lifespan. Epigenetic clocks attempt to quantify differential aging between individuals, but they typically summarize aging as a single measure, ignoring within-person heterogeneity. Our aim was to develop novel systems-based methylation clocks that, when assessed in blood, capture aging in distinct physiological systems. We combined supervised and unsupervised machine learning methods to link DNA methylation, system-specific clinical chemistry and functional measures, and mortality risk. This yielded a panel of 11 system-specific scores– Heart, Lung, Kidney, Liver, Brain, Immune, Inflammatory, Blood, Musculoskeletal, Hormone, and Metabolic. Each system score predicted a wide variety of outcomes, aging phenotypes, and conditions specific to the respective system, and often did so more strongly than existing epigenetic clocks that report single global measures. We also combined the system scores into a composite Systems Age clock that is predictive of aging across physiological systems in an unbiased manner. Finally, we showed that the system scores clustered individuals into unique aging subtypes that had different patterns of age-related disease and decline. Overall, our biological systems based epigenetic framework captures aging in multiple physiological systems using a single blood draw and assay and may inform the development of more personalized clinical approaches for improving age-related quality of life.

Introduction

The geroscience hypothesis states that directly targeting the biology of aging can improve human health and delay the onset of multiple chronic diseases. Yet, to truly test this hypothesis, reliable biomarkers must be developed that reflect valid age-related changes and responses to interventions.[1] “Epigenetic clocks”, based on DNA methylation (DNAm), are among the most studied aging biomarkers to date. Multiple versions have been developed, some to predict different age-related outcomes.[2–4] Most utilize information on a few hundred CpGs to report a single biological age value for each individual that is meant to reflect how the individual’s degree of biological aging compares to a reference population. For a number of epigenetic clocks, discordance between predicted and observed age has been shown to be biologically meaningful, as it is predictive of age-related morbidity and mortality, and correlates with other age-related phenotypes.[5–7] More recently, DNAm markers have been built to predict longitudinal changes in clinical indicators of aging, referred to as pace of aging, and these too show associations with age-related health outcomes.[8,9]

Existing DNAm clocks report an individual’s overall degree or pace of aging as a single value, capturing the heterogeneity between individuals. However, there is also heterogeneity in the aging process within individuals, at various levels of biological organization. For instance, there is variation in the rate of aging between organ systems, organs, tissues, and even cells.[10–13] Existing blood-based DNAm clocks do not address system-specific differences in aging. Though blood methylation is utilized in numerous aging studies due to ease of access, it remains unclear how much

information about various other organ systems can be gleaned from blood methylation alone. Attempts have been made to build DNAm biomarkers that capture some cardiac or metabolic disease risk, though they tend to only be targeted to a single system.[14–16] Another important caveat is that physiological systems can function or decline independently of each other, as well as in concert through their interactions. This gives rise to a third type of heterogeneity in aging - heterogeneity in aging that manifests in the overall pattern of decline across physiological systems. This is important for geroscience applications and for the prevention of multimorbidity via targeting biological aging, as diseases are often caused by a combination of malfunctioning in specific biological systems. For example, arthritis involves musculoskeletal deterioration and inflammation; while stroke may be caused by a combination of cardiovascular, metabolic, inflammatory, and neurological factors. These patterns across physiological systems may give rise to aging subtypes that predispose an individual to a subset of specific aging conditions.

Failing to capture the different levels of heterogeneity in the aging process has practical consequences for the assessment of aging patients and populations. Two individuals can have different DNAm profiles that produce the exact same epigenetic age as calculated by blood-based epigenetic clocks, yet they may be physiologically deteriorating in entirely different systems. Additionally, two individuals may have the exact same age of a system as calculated by blood-based methylation clocks yet they may be predisposed to different diseases depending on the co-occurrence of aging across other systems in their body. Not only will this refinement of aging indicators help

predict disease-specific differential risks across individuals, but our understanding of heterogeneity in aging is also a key step in the eventual development of targeted interventions based on personalized aging characteristics.

The overarching aim of this study was to construct systems-specific aging scores from DNA methylation data derived from whole blood. While clinical biomarkers and functional measures themselves provide direct assessment of specific physiological systems, we reasoned systems-specific methylation predictors would have two major advantages: 1) it is a single standardized assay that is comparable between studies, whereas the set of clinical biomarkers and functional measures can markedly differ between aging studies, and 2) DNA methylation is closer to root causes of aging[17]. Our study demonstrates it is possible to capture heterogeneity across many physiological systems using a single blood DNA methylation test, in turn predicting decline and disease specific to each system, as well as clustering individuals into diverse yet distinct epigenetic aging subtypes.

Results

Systems Age Pipeline for modeling systems-specific aging

Systems Age was constructed in a five step process (Figure 2, Methods). Briefly, we first mapped clinical chemistry and hematology biomarkers available in the Health and Retirement study (HRS) to specific biological systems (Supplementary Table 1). In addition to blood-based measures, we incorporated relevant functional assessments and disease status. Second, we performed principal component analysis (PCA) on measures within each system to identify latent signals captured by system-specific

principal components (PCs). Third, we predicted these system PCs using methylation PCs selected via elastic net regression within HRS. We utilized methylation PCs for predicting these measures given that use of PCs over individual CpGs increases test-retest reliability without sacrificing validity[18] (Supplementary Figure 3). Fourth, we calculated the predicted DNAm system PCs (using the models trained in HRS) in the Framingham heart Study (FHS) and trained a mortality prediction model for each system via elastic net Cox penalized regression. We referred to the resulting scores as 'system scores', meant to estimate aging in a particular system. Finally, we trained an elastic net Cox model by incorporating PCs from all systems into a unified whole body score called 'Systems Age'. Both individual system scores and overall System Age were scaled to the expected age range for interpretability. Given the complexity of the training, we wanted to determine which clinical chemistry and functional biomarkers contributed the most to each of the system scores to aid in interpretation. For this purpose we associated each system score with its corresponding biomarkers in HRS to highlight the biomarkers which had the greatest impact on the specific scores (Supplementary Figure 1).

System scores capture meaningful and specific aging signals

Specificity of system scores was assessed in an independent sample from 3 cohorts of the Women's Health Initiative (WHI BAA23, AS311 and EMPC, total N = ~5,600). Details about each cohort are available in the Methods section. Cohorts were stratified by race (except WHI AS311, which included few black and hispanic participants), for a total of 7 groups. Results from multivariate analyses adjusting for chronological age

were meta-analyzed via a fixed effects model with inverse variance weights (Figure 3 as well as Supplementary Table 2 and 3). We tested associations with disease incidence (using cox proportional hazard models), disease prevalence (using logistic regression models), and functional parameters of aging (using Ordinary Least Squares regression models).

Our results suggested high levels of specificity of system scores to the expected organ system (Figure 4 as well as Supplementary Table 3 and 4). This was true across baseline and future conditions as well as functional phenotypes and diseases. For functional phenotypes at baseline, the Brain score showed the strongest association with cognitive function (meta z-score = 3.51), and the Musculoskeletal score had the strongest association with physical function (meta z-score = 8.53). The Heart score was most strongly associated with time-to-CHD events (meta z-score = 8.29) as well as with time to myocardial infarction (meta z-score = 6.30). The Lung score was most strongly associated with time to lung cancer (meta z-score = 9.69) with the Heart score coming a close second (meta z-score = 9.49), which shares risk factors such as smoking with cardiovascular disease. For time to stroke, Heart score (meta z-score = 3.40) was the most strongly associated, with the Metabolic score coming a close second (meta z-score = 3.35). The blood score was most strongly associated with time to Leukemia (meta z-score = 4.88).

In almost all conditions and disease outcomes, we observed the expected directionality–with increased age indicators associated with increased risk. The one

exception to this was reproductive organ cancers, in which we observed negative associations with all of the systems. Hormone (meta z-value = -3.04) and Blood (meta z-value = -2.22) were the most negatively associated with breast and endometrial cancer, respectively.

For diseases at baseline, Musculoskeletal was most strongly associated with diabetes (meta z-value = 10.90) and arthritis (meta z-value = 5.15). Hormone had the strongest association with thyroid disease (meta z-value = 2.65) and for cataract, both Heart (meta z-value = 3.16) and Liver (meta z-value = 3.12) had the strongest associations. Finally, the total comorbidities variable at baseline was most strongly associated with Inflammation (meta z-value = 7.65).

Compared to existing clocks, system scores better capture the multifactorial nature of aging

We compared system scores and Systems Age to previously trained epigenetic clocks using meta z-scores (Figure 4, Supplementary Figure 2, Supplementary Table 2, 3, 4, 5, 6 and 7), estimated in the same manner as was done for system scores in the previous section (Figure 3). We focused on three prominent epigenetic clocks that have previously been demonstrated to be strongly associated with aging outcomes: PCPhenoAge, DNAmGrimAge and DunedinPACE.

The most relevant system scores had a higher meta Z-score based on effect sizes to existing epigenetic clocks for 10 of the 14 diseases. For example, system scores had

stronger associations with diabetes at baseline (Musculoskeletal meta z-score 10.90; DunedinPACE 8.96), time to leukemia (Blood meta z-score 4.88; PCPhenoage 2.74), cognitive function (Brain meta z-score 3.51; PCPhenoAge 2.95), cataract at baseline (Liver meta z-score 3.12; PCPhenoAge 2.59), time to breast cancer (Hormone absolute meta z-score 3.01;PCPhenoage 1.32), time to endometrial cancer (Blood absolute meta z-score 2.22; PCPhenoage 1.19), arthritis at baseline (Musculoskeletal meta z-score 5.15; DunedinPACE 4.96), time to myocardial infarction (Heart meta z-score 6.30; DNAmGrimAge 6.09), time to CHD (Heart meta z-score 8.29; DNAmGrimAge 8.11), and time to colorectal cancer (Blood meta z-score 2.08; DNAmGrimAge 1.88). For the remaining 4 diseases and conditions, the clocks were a close second to existing clocks–physical function (musculoskeletal meta z-score 9.46; DunedinPACE 9.47), time to stroke (Heart meta z-score 3.40; DunedinPACE 3.45), thyroid disease at baseline (Hormone meta z-score 2.65; DNAmGrimAge 2.92), and time to lung cancer (Lung meta z-score 9.69; DNAmGrimAge 12.11). For the three whole body metrics, system specific scores had second highest Z-scores including for time to death or Mortality (Heart meta z-score 15.17; DNAmGrimAge 16.81), total comorbidities at baseline (Inflammation meta z-score 7.65; DunedinPACE 9.08), and disease free at baseline (Heart meta z-score 3.98; DunedinPACE 4.28).

To determine whether system scores were significantly better at outcome prediction than existing epigenetic clocks, we performed receiver operating characteristic curve (ROC) analysis (Figure 5, Supplementary Tables 13 and 14) on each cohort comparing the best system score to the different clocks. The system scores had a significantly

better area under the curve compared to some clocks when it came to prediction of /associations with Mortality (3.55 and 3.84 Meta z-score for Heart compared to DunedinPACE and PCPhenoAge ROC, respectively), total comorbidities at baseline (2.39 and 2.17 Meta z-score for Inflammation compared to DNAmGrimAge and PCPhenoAge respectively), CHD (2.15 Meta z-score for Heart compared to PCPhenoAge), MI (2.5 Meta z-score for Heart compared to PCPhenoAge), leukemia (5.63 and 6.13 Meta z-score for Immune compared to DunedinPACE and DNAmGrimAge respectively), lung cancer (3.13 and 3.92 Meta z-score for Lung compared to DunedinPACE and PCPhenoAge respectively), cognitive function (2.99, 3.87, and 3.23 Meta z-score for Brain compared to DunedinPACE, DNAmGrimAge, and PCPhenoAge respectively), physical function (3.53 and 2.50 Meta z-score for MusculoSkeletal compared to DNAmGrimAge and PCPhenoAge respectively), arthritis (3.62 and 3.43 Meta z-score for MusculoSkeletal compared to DNAmGrimAge and PCPhenoAge respectively), and diabetes (2.66, 5.17, and 5.66 Meta z-score compared to DunedinPACE, DNAmGrimAge, and PCPhenoAge respectively). We did not find statistically significant improvements for cataracts or stroke.

Overall, these results suggest that having different scores for each system may more precisely capture disease relevant risk and facilitate personalized interventions compared to a single global metric.

Systems Age: the golden mean

When training clocks, each may be biased towards predicting specific aspects of aging based on the combination of variables and datasets used for training. Since each

systems score showed superior or equivalent associations with specific diseases and aging phenotypes, we hypothesized that combining them into a single Systems Age score would lead to a more uniform prediction across all diseases and aging phenotypes. Indeed, we found that Systems Age was not biased to a specific dimension of aging and performed well across a variety of diseases and conditions. Of the 14 different conditions we tested for which every clock showed significant associations (Figure 3, Supplementary figure 2, Supplementary Table 2, Supplementary Table 3 and Supplementary File 1), Systems Age had the strongest associations of all clocks for four conditions, including cataract (Systems Age 3.06; PCPhenoAge 2.59), CHD (Systems Age 8.43; DNAmGrimAge 8.11), myocardial infarction (Systems Age 6.36; DNAmGrimAge 6.09), and leukemia (Systems Age 2.95; PCPhenoAge 2.74). For 8 of the conditions, Systems Age was second best, as in the case for time to stroke (Systems Age 3.33; DunedinPACE 3.45), disease free at baseline (Systems Age 3.87; DunedinPACE 4.28), physical function (Systems Age 8.91; DunedinPACE 9.47), cognitive function (Systems Age 2.68; PCPhenoAge 2.95), time to death (Systems Age 15.35; DNAmGrimAge 16.81), thyroid disease at baseline (Systems Age 2.56; DNAmGrimAge 2.92), arthritis at baseline (Systems Age 3.23; DNAmGrimAge 4.96), and time to lung cancer (Systems Age 9.18; DNAmGrimAge 12.19). DunedinPACE and DNAmGrimAge did perform best with a few of the conditions (3 and 6 resp.) but was ranked last, second to last, or was not significant in nearly all others (11 and 8 resp.), indicating they were biased towards certain dimensions of aging .

To determine whether Systems Age was significantly better at outcome prediction than existing epigenetic clocks, we performed ROC analysis (Figure 5, Supplementary Tables 11 and 12) on each cohort comparing Systems Age to the different clocks (similar to above for system specific scores). ROC analysis Z-scores followed similar trends to the association analysis, but in some cases the improvement was not significantly higher. Nevertheless, Systems Age did outperform other clocks significantly in many outcomes including mortality (3.89 and 4.32 Meta z-score compared to DunedinPACE and PCPhenoAge respectively), comorbidities at baseline (2.05 and 1.97 Meta z-score compared to DNAmGrimAge and PCPhenoAge respectively), CHD (2.49 Meta z-score compared to PCPhenoAge), MI (2.78 Meta z-score compared to PCPhenoAge), lung cancer (2.94 and 4.09 Meta z-score compared to DunedinPACE and PCPhenoAge respectively), cognitive function (2.43 Meta z-score compared to DNAmGrimAge), physical function (3.10 Meta z-score compared to DNAmGrimAge) and arthritis (2.36 Meta z-score compared to DNAmGrimAge).

Overall, this suggests the Systems Age training paradigm enables more uniform and unbiased prediction across many dimensions of aging compared to existing state-of-the-art clocks.

Capturing aging signal beyond smoking

Smoking is well known to affect DNA methylation and epigenetic clocks, [21] as well as disease incidence (especially cardiopulmonary diseases and cancer), aging phenotypes, and mortality. To determine the degree, smoking impacts the associations with clocks, we calculated meta z-scores for system score and existing clocks while

adjusting for smoking status (Supplementary Figure 2, Supplementary Table 4 and Supplementary Table 5). For example, the Metabolic system score was strongly associated with stroke, and this association changed minimally when adjusting for smoking status (meta z-score 3.35 as compared to meta z-score 3.20 when adjusted for smoking status). On the other hand, GrimAge's association with time to stroke decreased by more than 50% and was no longer significant when adjusting for smoking (meta z-score 2.75 as compared to meta z-score 1.23 when adjusted for smoking status). Reduced associations with CHD and time to myocardial infarction were also observed across all clocks after adjusting for smoking (as expected given that smoking is a major risk factor for heart disease), but Systems Age and Heart score retain much of their association after smoking status adjustment (meta Z-score in CHD before and after smoking status adjusted for Heart 8.29 vs 6.57, Systems Age 8.43 vs 6.77, DNAmGrimAge 8.11 vs 5.80; meta Z-score in myocardial infarction before and after smoking status adjusted for Heart 6.30 vs 5.04, Systems Age 6.36 vs 5.18, DNAmGrimAge 6.09 vs 4.34). Similar impacts of smoking were seen for time to lung cancer and time to death. For other diseases and aging phenotypes, Systems Age and system scores retained their prediction after adjusting for smoking, indicating they captured epigenetic signals beyond just cigarette exposure. We also performed an analysis among never smokers (Supplementary Figure 2 as well as Supplementary Table 6 and 7) finding that system scores performed the best across all the conditions in which at least one score was significant.

System scores capture distinct dimensions of aging

Systems do not work independently of each other. Some systems are closely related, sharing many associations with a given disease, condition, or aging phenotype. Indeed, looking at correlations between different age-adjusted system scores across the WHI cohorts (Figure 6a and Supplementary table 8), we found some systems to be highly correlated with each other, such as Heart and Lung ($r = 0.759$), or Inflammation and Musculoskeletal ($r = 0.716$). Hierarchical clustering revealed that Heart and Lung formed a cluster, Liver, Brain and Blood formed a second cluster, and Metabolic, Inflammation, and Kidney formed a third cluster. The latter two clusters formed a super-cluster that also included Heart and Musculoskeletal. Given these patterns, we hypothesized that there is predictive value in examining not just heterogeneity within a system or across systems, but also to test whether individuals can be grouped based on their system scores to generate aging subtypes with distinct predisposition to aging related diseases and conditions.

To test this, we first examined whether there existed individuals with the same chronological age and overall Systems Age yet different age-accelerated system scores. One example of this was observed for three individuals (Figure 6b), with the same chronological age, Systems Age, gender, and race. However, they had entirely different patterns of system scores.

To test whether these patterns of system scores constituted biologically relevant aging subtypes, we clustered individuals using adaptive hierarchical clustering in the WHI

EMPC cohort. This analysis identified 9 unique groups or clusters, with each cluster showing distinct patterns across the different system aging scores, as well as different associations with the prevalence or future occurrence of certain diseases (Figure 6c and 5d, Supplementary table 9 and 10). For example, Cluster 8 and 9 were found to have a high mean Lung aging score (Lung age accelerations 1.06 and 0.71) and a higher prevalence of future lung cancer events (pval: 0.01, 0.04). These groups also had an overrepresentation of smokers (pval: 0.001, 0.009), indicating they captured smoking pathophysiology. As expected, their distinct clustering stemmed from other pathophysiology: Group 8 had fewer obese individuals (p= 0.01) while Group 9 was enriched for future CHD events (p = 0.02). All of this indicated that the 2 groups were capturing distinct aging subtypes within smokers. At the same time, Cluster 3 demonstrated an increased prevalence of MI (p= 0.02) yet it showed low Lung aging and decreased prevalence of smokers (p= 0.03). It also demonstrated increased Metabolic (average age acceleration 0.55) and Inflammation (average age acceleration 0.34) aging. Thus, while Cluster 3 and 9 are both at risk for cardiovascular diseases, the risk stems from different sources (smoking vs. metabolic and inflammatory aging). Overall, the system's scores were capturing relevant aging subtypes that had distinct behavioral and genetic patterns predisposing individuals to certain types of aging phenotypes and diseases.

Discussion

In the past decade, various epigenetic clocks have been developed to predict chronological age, composite measures of biological age, single-cell epigenetic age,

and aging in various tissues.[22–24] Yet, what remained missing was a method to capture aging in different biological systems, independently and interactively, using a single blood test. While it was unclear to what extent signals in other organ systems could be captured in blood DNAm, our results suggest that this is possible. Systems Age is the first measure to capture heterogeneity in aging across different biological systems using a single assay performed on samples from a single blood draw.

We showed that Systems Age scores were not only predictive of a wide variety of aging conditions and phenotypes at both baseline and follow-up, but were also specific to the pathophysiology of their intended system. For example, the Heart score was most associated with heart disorders CHD and MI,[25] as well as overall mortality reflecting that cardiovascular disease is the leading cause of mortality worldwide.[26] Heart was also strongly associated with thyroid disease, lung cancer, stroke, cataracts, reduced physical function, and total comorbidities, reflecting disease and treatment complications, shared risk factors, and shared pathophysiology.[27–32] However, Heart demonstrated specificity in that it was only very weakly associated with diseases such as baseline arthritis or time-to-leukemia (instead these were most strongly associated with Musculoskeletal and Blood respectively).[33] Likewise, the Inflammation score was strongly associated with time-to-CHD[34], baseline arthritis[35] and baseline physical and cognitive functioning [36,37], which are expected based on known pathophysiology. Inflammation was the most strongly associated system with total number of comorbidities at baseline, consistent with inflammation driving many diseases of aging.[38] The Brain score was associated with baseline cognitive functioning and

time-to-stroke, but much less with most other phenotypes. The Musculoskeletal score was strongly associated with physical function and baseline arthritis as expected, as well as total comorbidities and baseline diabetes which can worsen musculoskeletal function,[39,40] but was far less predictive of other phenotypes than other system scores. Thus, blood DNA methylation data alone can be used to derive many different specific aging scores for various physiological systems, rather than just a single blood-specific or whole-body aging process.

We also observed unexpected results, such as negative associations of epigenetic aging with breast cancer risk in women, with the strongest negative association being Hormone. However, prior literature has shown that epigenetic clock acceleration is associated with a younger age at menopause.[41] Simultaneously, later menopause has been linked to higher risk of reproductive organ cancers.[20,42] Thus, women who epigenetically age faster may in fact have lower chances of getting reproductive organ cancer. In moving forward, there may be a further need to develop reproductive system specific scores for different sexes as well as expand the Hormone score.

Given the interconnectedness of aging, it was unclear prior to our study whether a given aging phenotype would be better predicted by epigenetic clocks that detected more global aging signals, or clocks trained on a limited set of clinical biomarkers specifically related to that phenotype. Our results demonstrated the advantages of the latter approach. Other recently developed clocks that strongly associated with specific aging conditions and phenotypes included PCPhenoAge (trained on a composite biological

age measure that involves multiple systems), DNAmGrimAge (trained using smoking and proteins each involved in multiple systems), and DunedinPACE (trained on longitudinal changes across many systems).[6–8] Thus, these clocks are intended to capture global aging signals that are not limited to any particular system. For 10 of the 14 system specific phenotypes and diseases we tested, the most relevant system score had higher meta-Z scores than all three of these clocks. ROC analysis demonstrated that many of these improvements were statistically significant. While the relevant system score did not surpass the predictive ability of all existing clocks for other outcomes, the system scores were nearly as predictive, while being interpretable and granular - they reveal which systems are related to which phenotypes. Another advantage of system scores became apparent when looking across many phenotypes simultaneously. For example, DNAmGrimAge predicted mortality, cardiovascular outcomes, lung cancer, and thyroid dysfunction particularly well, but was less predictive of cognitive function, comorbidities, arthritis, diabetes or leukemia risk. Interestingly, PCPhenoAge showed nearly the opposite pattern as DNAmGrimAge. DunedinPACE's pattern of association resembled a mixture between PCPhenoAge and DNAmGrimAge, but was still not predictive of some phenotypes like thyroid dysfunction and leukemia risk. This suggested that epigenetic clocks that are directly trained to predict global proxies of aging can introduce biases in which aging phenotypes they are related to. In contrast, Systems Age showed more uniform prediction across all phenotypes, showing either the strongest or second-strongest associations with 12 of the 14 conditions, compared to the other three clocks. Thus, Systems Age appeared to not be strongly biased by a particular dimension of aging, which is likely the result of first training

predictors of mortality in each physiological system independently before combining them. In further support of this idea, system scores and Systems Age remain highly associated with all these phenotypes after correction for smoking.

In addition to heterogeneity at the systems level, we examined the heterogeneity that arose due to the interaction of different systems. Interestingly, some system scores were more correlated with each other than others. Heart and Lung were highly related, consistent with their common vulnerability to smoking, shared pulmonary circulation and combined function in oxygenating the body. The Liver and Blood scores showed strong correlations, potentially reflecting the liver's blood filtration function and production of blood products. Both Liver and Blood were highly correlated with Brain, potentially reflecting known contributions of anemia and altered levels of liver products to brain aging[46–49]. Metabolic, Inflammation, and Kidney were highly correlated, reflecting numerous links between metabolism and inflammation[50] as well as the inclusion of IL-6 and CRP in both systems, and the contributions of both to kidney aging[51,52]. The super-cluster involving Heart, Musculoskeletal, Liver, Blood, Brain, Metabolic, Inflammation, and Kidney can be similarly ascribed to numerous known interactions between systems as well as shared risk factors. Of note, while the correlations between systems do likely reflect true physiological interactions, it is also possible that some of the correlation structure can be attributed to similar mechanisms by which they impact the blood methylome and vice versa.

We defined subtypes as groups of individuals with similar systems specific aging scores that have an over or under representation of specific diseases and conditions. We showed the existence of 9 such distinct clusters or aging subtypes that had predisposition to very distinct diseases and conditions. Demonstration of the existence of aging subtypes has already been documented in the literature.[10] However, we provided evidence that these distinct groups can be observed even when using a single assay–in our case, DNA methylation assessed in blood. In the future, information on longitudinal changes will be critical for further defining age subtypes and their relevant disease risks. This could facilitate sub-classification of aging conditions and eventually inform targeted therapies based on aging subtypes.

Previous attempts have been made at capturing system-specific aging using metabolomic[53], proteomic , clinical biomarker[54] , and other multi-Omic [55] data; however, the translatability of these models to clinic and wider application is a challenge given the limitations with usage of these data types in clinic due to technological limitations[56] or harmonization across 100s of clinical biomarkers. Here we were able to accomplish a similar goal, while only relying on a single assay. Thus, the standardization offered by DNA methylation data generation allows for both potential wide-spread usage and ease of cross-comparison across experiments. Additionally, with recent developments, the decrease in cost of epigenetic data generation further shows its versatility as a superior source of biological information.

Systems Age as a framework for capturing heterogeneity shows a lot of promise, yet limitations remain. Due to lack of clinical measures for certain domains, we were not able to capture certain systems that are known to be affected by aging. Take for example, reproductive aging which has been shown to be a critical dimension of aging but was only captured indirectly in our framework. Clearly, there is a need for reproductive systems specific scores, something that can be developed in later iterations of such frameworks. Additionally, while we were able to approximate metabolic and hepatic aging, there is no score that shines light on aging in other digestive organ systems such as stomach, pancreas, colon and more. All of these organ systems have distinct aging related conditions which are unfortunately not captured in the present scores. Similarly, sensory aging such as those pertaining to vision, hearing, and sensation are not captured in the system scores. All of these are potential future system scores that can be built into the framework. It is also important to note that one could add more clinical biomarkers and phenotypes to the systems, and there are other ways to map biomarkers to systems. Rather, it is unknown how robust these system scores are to changes in the systems biomarkers used for training and thus there needs to be further testing done to shine light on these aspects. Another important future direction for Systems Age could be to detangle genetic predisposition to aging in certain systems as opposed to environmental effects on aging of systems. A very clear example of this is smoking, which leads to accelerated aging in certain systems and predisposition to specific aging related diseases. Conversely, There may also be genetic factors, which predispose certain systems to be more or less vulnerable than others. [57] Another caveat along the same lines is that it is unknown why DNAm in blood

reflects aging in other physiological systems, and what is the molecular relationship between the clinical biomarkers, disease states and blood DNAm. It could reflect shared genetic variation, exposures, age-related patterns between tissues. Alternatively it could involve intercellular signaling influencing DNAm in blood (either directly through epigenetic regulators or via changes in blood cell proportions), or blood DNAm reflecting processes by which immune cells affect aging in those systems. Further analysis needs to be performed to understand these relationships better. Finally, Systems Age uses only clinical data to first generate scores that are then predicted from epigenetic data. Other data types, such as proteomics, metabolomics, or imaging, may be highly informative when it comes to capture more diverse dimensions of aging.

Overall, we highlight the importance of capturing heterogeneity in aging while also building a reusable framework for quantifying multifactorial aging phenotypes. We show that this level of dimensionality can be estimated from a single data source–in this case DNA methylation in blood. The scores built using our approach perform as well, or in many cases better than, presently available epigenetic clocks, while simultaneously providing the potential to identify individuals with distinct aging subtypes for clinical healthcare and drug development purposes.

Acknowledgments

This work was supported by the National Institute on Aging (NIA:1R01AG065403-03, 1R01AG057912-05 to M.E.L. and A.H.C.). It was also supported by the Impetus Grant (R.S.), the Gruber Science Fellowship at Yale University (R.S.), the Thomas P. Detre

Fellowship Award in Translational Neuroscience Research from Yale University (to A.H.C.), and the Medical Informatics Fellowship Program at the West Haven, CT Veterans Healthcare Administration (to A.H.C.). The HRS study was supported by NIA R01AG060110-05, R01AG068937-02 and U01AG009740. Additionally, we would like to recognize the support of the WHI Program Office: (National Heart, Lung, and Blood Institute, Bethesda, Maryland) Jacques Rossouw, Shari Ludlam, Joan McGowan, Leslie Ford, and Nancy Geller, the WHI Clinical Coordinating Center: (Fred Hutchinson Cancer Research Center, Seattle, WA) Garnet Anderson, Ross Prentice, Andrea LaCroix, and Charles Kooperberg, the WHI Investigators and Academic Centers: (Brigham and Women's Hospital, Harvard Medical School, Boston, MA) JoAnn E. Manson; (MedStar Health Research Institute/Howard University, Washington, DC) Barbara V. Howard; (Stanford Prevention Research Center, Stanford, CA) Marcia L. Stefanick; (The Ohio State University, Columbus, OH) Rebecca Jackson; (University of Arizona, Tucson/Phoenix, AZ) Cynthia A. Thomson; (University at Buffalo, Buffalo, NY) Jean Wactawski-Wende; (University of Florida, Gainesville/Jacksonville, FL) Marian Limacher; (University of Iowa, Iowa City/Davenport, IA) Jennifer Robinson; (University of Pittsburgh, Pittsburgh, PA) Lewis Kuller; (Wake Forest University School of Medicine, Winston-Salem, NC) Sally Shumaker; (University of Nevada, Reno, NV) Robert Brunner Women’s Health Initiative Memory Study: (Wake Forest University School of Medicine, Winston-Salem, NC) Mark Espeland. Data from the Epigenetic Mechanisms of PM-Mediated CVD Risk (EMPC) were generated under National Institute of Environmental Health Sciences grant R01-ES020836

Conflicts of Interest Statement

The methodology described in this manuscript is the subject of an invention declaration at Yale and a provisional patent application for which M.E.L, R.S., A.H.C. and M.M. are named as inventors and Yale University is named as owner. In the past, M.E.L. was a Scientific Advisor for Elysium Health from July 2019 to October 2021. M.E.L. also holds licenses for epigenetic clocks that she has developed. A.H.C. has received consulting fees from TruDiagnostic and FOXO Biosciences. R.S. has received consulting fees from the LongevityTech.fund and Cambrian BioPharma unrelated to this publication. M.E.L. is an employee of Altos Labs. All other authors report no biomedical financial interests or potential conflicts of interest.

Author Contributions

M.E.L. conceived the project. R.S. and M.E.L. conceived the study design. R.S. and M.E.L. built the pipeline for training Systems Age. R.S. performed analysis in WHI to validate Systems Age. A.H.C. and R.S. performed reliability analysis. R.S., A.H.C. and M.E.L. interpreted the results. A.H.C. and M.E.L. provided supervision and managed the overall project. Other authors contributed data and analyses related to WHI (A.S., R.C., J.E.M., P.B., T.A., E.A.W.) and HRS (E.M.C.). All authors reviewed and contributed to the manuscript.

Data Availability Statement

HRS data is available at hrsdata.isr.umich.edu. FHS data is available at dbGaP, accession number: phs000724.v7.p11. Access to WHI data is available upon review and subsequent approval of proposals submitted through the study website (https://www.whi.org/propose-a-paper). The reliability dataset is available on GEO, accession ID GSE55763.

Code Availability Statement

Code to Systems Age will be available upon publication.

Licensing System Age

Systems Age or any of its components have not yet been licensed in any form to any third party. However, Systems Age is available to license. For more details, please email: lolahon.kadiri@yale.edu.

Systems Age code was first developed in July 2021. Since the first development, the Systems Age code has been updated with new biomarkers and systems, and has now been validated in an independent dataset (WHI).

References

1. Kritchevsky, S. B. & Justice, J. N. Testing the Geroscience Hypothesis: Early Days. *J. Gerontol. A Biol. Sci. Med. Sci.* **75**, 99–101 (2020).
2. Jylhävä, J., Pedersen, N. L. & Hägg, S. Biological age predictors, EBioMedicine 21

(2017) 29--36.

3. Ferrucci, L., Levine, M. E., Kuo, P.-L. & Simonsick, E. M. Time and the Metrics of Aging. *Circulation Research* vol. 123 740–744 Preprint at https://doi.org/10.1161/circresaha.118.312816 (2018).
4. Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* **19**, 371–384 (2018).
5. Oblak, L., van der Zaag, J., Higgins-Chen, A. T., Levine, M. E. & Boks, M. P. A systematic review of biological, social and environmental factors associated with epigenetic clock acceleration. *Ageing Res. Rev.* **69**, 101348 (2021).
6. Lu, A. T. *et al.* DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging* **11**, 303–327 (2019).
7. Levine, M. E. *et al.* An epigenetic biomarker of aging for lifespan and healthspan. *Aging* **10**, 573–591 (2018).
8. Belsky, D. W. *et al.* DunedinPACE, a DNA methylation biomarker of the pace of aging. *Elife* **11**, (2022).
9. Belsky, D. W. *et al.* Quantification of the pace of biological aging in humans through a blood test, the DunedinPoAm DNA methylation algorithm. *eLife* vol. 9 Preprint at https://doi.org/10.7554/elife.54870 (2020).
10. Ahadi, S. *et al.* Personal aging markers and ageotypes revealed by deep longitudinal profiling. *Nat. Med.* **26**, 83–90 (2020).
11. Liu, Z. *et al.* Underlying features of epigenetic aging clocks in vivo and in vitro. *Aging Cell* **19**, e13229 (2020).
12. Li, Y. *et al.* A programmable fate decision landscape underlies single-cell aging in

yeast. *Science* **369**, 325–329 (2020).

13. Trapp, A., Kerepesi, C. & Gladyshev, V. N. Profiling epigenetic age in single cells. *Nature Aging* **1**, 1189–1201 (2021).
14. Fernández-Sanlés, A. *et al.* DNA methylation biomarkers of myocardial infarction and cardiovascular disease. *Clin. Epigenetics* **13**, 86 (2021).
15. Westerman, K. *et al.* Epigenomic Assessment of Cardiovascular Disease Risk and Interactions With Traditional Risk Metrics. *J. Am. Heart Assoc.* **9**, e015299 (2020).
16. Hidalgo, B. A. *et al.* A 6-CpG validated methylation risk score model for metabolic syndrome: The HyperGEN and GOLDN studies. *PLoS One* **16**, e0259836 (2021).
17. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194–1194 (2013).
18. Higgins-Chen, A. T., Thrush, K. L., Wang, Y. & Minteer, C. J. A computational solution for bolstering reliability of epigenetic clocks: Implications for clinical trials and longitudinal tracking. (2022).
19. Wu, Y., Sun, W., Liu, H. & Zhang, D. Age at Menopause and Risk of Developing Endometrial Cancer: A Meta-Analysis. *Biomed Res. Int.* **2019**, 8584130 (2019).
20. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *Lancet Oncol.* **13**, 1141–1151 (2012).
21. Gao, X., Zhang, Y., Breitling, L. P. & Brenner, H. Relationship of tobacco smoking and smoking-related DNA methylation with epigenetic age acceleration. *Oncotarget* **7**, 46878–46889 (2016).
22. Seale, K., Horvath, S., Teschendorff, A., Eynon, N. & Voisin, S. Making sense of the

ageing methylome. *Nat. Rev. Genet.* (2022) doi:10.1038/s41576-022-00477-6.

23. Yousefi, P. D. *et al.* DNA methylation-based predictors of health: applications and statistical considerations. *Nat. Rev. Genet.* **23**, 369–383 (2022).
24. Thrush, K. L., Higgins-Chen, A. T., Liu, Z. & Levine, M. E. R methylCIPHER: A Methylation Clock Investigational Package for Hypothesis-Driven Evaluation & Research. *bioRxiv* 2022.07.13.499978 (2022) doi:10.1101/2022.07.13.499978.
25. Aronow, W. S. Heart disease and aging. *Med. Clin. North Am.* **90**, 849–862 (2006).
26. Roth, G. A. *et al.* Global Burden of cardiovascular diseases and risk factors, 1990-2019: Update from the GBD 2019 Study. *J. Am. Coll. Cardiol.* **76**, 2982–3021 (2020).
27. Campbell, B. C. V. *et al.* Ischaemic stroke. *Nat Rev Dis Primers* **5**, 70 (2019).
28. Cappola, A. R. *et al.* Thyroid and Cardiovascular Disease Research Agenda for Enhancing Knowledge, Prevention, and Treatment. *Circulation* (2019) doi:10.1161/CIRCULATIONAHA.118.036859.
29. Handy, C. E. *et al.* Synergistic Opportunities in the Interplay Between Cancer Screening and Cardiovascular Disease Risk Assessment: Together We Are Stronger. *Circulation* **138**, 727–734 (2018).
30. Buddeke, J. *et al.* Comorbidity in patients with cardiovascular disease in primary care: a cohort study with routine healthcare data. *Br. J. Gen. Pract.* **69**, e398–e406 (2019).
31. Nemet, A. Y., Vinker, S., Levartovsky, S. & Kaiserman, I. Is cataract associated with cardiovascular morbidity? *Eye* **24**, 1352–1358 (2010).
32. Hu, X. *et al.* Physical Function and Subsequent Risk of Cardiovascular Events in

Older Adults: The Atherosclerosis Risk in Communities Study. *J. Am. Heart Assoc.* **11**, e025780 (2022).

33. Grais, I. M. & Sowers, J. R. Thyroid and the heart. *Am. J. Med.* **127**, 691–698 (2014).
34. Bartekova, M., Radosinska, J., Jelemensky, M. & Dhalla, N. S. Role of cytokines and inflammation in heart function during health and disease. *Heart Fail. Rev.* **23**, 733–758 (2018).
35. Simmonds, R. E. & Foxwell, B. M. Signalling, inflammation and arthritis: NF-κB and its relevance to arthritis and inflammation. *Rheumatology* **47**, 584–590 (2008).
36. Walker, K. A. *et al.* Systemic inflammation during midlife and cognitive change over 20 years: The ARIC Study. *Neurology* **92**, e1256–e1267 (2019).
37. Brinkley, T. E. *et al.* Chronic inflammation is associated with low physical function in older adults across multiple comorbidities. *J. Gerontol. A Biol. Sci. Med. Sci.* **64**, 455–461 (2009).
38. Franceschi, C., Garagnani, P., Parini, P., Giuliani, C. & Santoro, A. Inflammaging: a new immune-metabolic viewpoint for age-related diseases. *Nat. Rev. Endocrinol.* **14**, 576–590 (2018).
39. Perry, B. D. *et al.* Muscle atrophy in patients with Type 2 Diabetes Mellitus: roles of inflammatory pathways, physical activity and exercise. *Exerc. Immunol. Rev.* **22**, 94–109 (2016).
40. Solomon, D. H. *et al.* Factors Associated With 10-Year Declines in Physical Health and Function Among Women During Midlife. *JAMA Netw Open* **5**, e2142773 (2022).

41. Levine, M. E. *et al.* Menopause accelerates biological aging. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 9327–9332 (2016).

42. Trichopoulos, D., MacMahon, B. & Cole, P. Menopause and breast cancer risk. *J. Natl. Cancer Inst.* **48**, 605–613 (1972).

43. Leng, S. X. *et al.* Serum levels of insulin-like growth factor-I (IGF-I) and dehydroepiandrosterone sulfate (DHEA-S), and their relationships with serum interleukin-6, in the geriatric syndrome of frailty. *Aging Clin. Exp. Res.* **16**, 153–157 (2004).

44. Grimberg, A. Mechanisms by which IGF-I may promote cancer. *Cancer Biol. Ther.* **2**, 630–635 (2003).

45. Mahmud, K. Hormones and breast cancer: can we use them in ways that could reduce the risk? *Oncol. Rev.* **2**, 146–153 (2008).

46. Winchester, L. M., Powell, J., Lovestone, S. & Nevado-Holgado, A. J. Red blood cell indices and anaemia as causative factors for cognitive function deficits and for Alzheimer's disease. *Genome Med.* **10**, 51 (2018).

47. Nho, K. *et al.* Association of Altered Liver Enzymes With Alzheimer Disease Diagnosis, Cognition, Neuroimaging Measures, and Cerebrospinal Fluid Biomarkers. *JAMA Netw Open* **2**, e197978 (2019).

48. Min, J.-Y. *et al.* Chronic Status of Serum Albumin and Cognitive Function: A Retrospective Cohort Study. *J. Clin. Med. Res.* **11**, (2022).

49. Horowitz, A. M. *et al.* Blood factors transfer beneficial effects of exercise on neurogenesis and cognition to the aged brain. *Science* **369**, 167–173 (2020).

50. Wu, H. & Ballantyne, C. M. Metabolic Inflammation and Insulin Resistance in

Obesity. *Circ. Res.* **126**, 1549–1564 (2020).

51. Romagnani, P. *et al.* Chronic kidney disease. *Nat Rev Dis Primers* **3**, 17088 (2017).
52. Sepe, V., Libetta, C., Gregorini, M. & Rampino, T. The innate immune system in human kidney inflammaging. *J. Nephrol.* **35**, 381–395 (2022).
53. Buergel, T. *et al.* Metabolomic profiles predict individual multidisease outcomes. *Nat. Med.* (2022) doi:10.1038/s41591-022-01980-3.
54. Tian, Y. E. *et al.* Biological aging of human body and brain systems. *bioRxiv* (2022) doi:10.1101/2022.09.03.22279337.
55. Nie, C. *et al.* Distinct biological ages of organs and systems identified from a multi-omics study. *Cell Rep.* **38**, 110459 (2022).
56. Frantzi, M., Bhat, A. & Latosinska, A. Clinical proteomic biomarkers: relevant issues on study design & technical considerations in biomarker development. *Clin. Transl. Med.* **3**, 7 (2014).
57. Kuo, C.-L., Pilling, L. C., Liu, Z., Atkins, J. L. & Levine, M. E. Genetic associations for two biological age measures point to distinct aging phenotypes. *Aging Cell* **20**, e13376 (2021).

Methods

Datasets used for training Systems Age

Two different longitudinal studies were used for training Systems Age: the Health and Retirement Study (HRS) and Framingham Heart Study (FHS). We previously utilized and described methylation data from these datasets in a separate study (Higgins-Chen et al. 2022). Briefly, as stated on their website, HRS is a nationally representative sample of Americans over age 50 years. HRS had biomarker information available for 9,933 participants of which Infinium Methylation EPIC BeadChip data was available for 4,018 individuals (Crimmins, Thyagarajan, and Levine 2021). Out of the 4018 individuals only 3,593 had clinical data (age range 51–100 years) which were used for training of Systems Age. All participants provided written informed consent. The study was approved by the Institutional Review Board (IRB) at the University of Michigan (HUM00061128).

FHS includes 2,748 FHS Offspring cohort participants attending the eighth exam cycle (2005–2008) and 1,457 Third Generation cohort participants attending the second exam cycle (2005–2008), who consented to provide their DNA for genomic research (Kannel et al. 1979; Splansky et al. 2007). DNA methylation was assayed with the Infinium HumanMethylation450 BeadChip and is available in dbGaP (accession no. phs000724.v7.p11). For the purpose of training Systems Age FHS Offspring data was used but for scaling of system scores and systems age to age range both the Offspring

and Third generation data was used. Deaths of FHS participants occurring before 1 January 2014 were ascertained by routine contact with participants, surveillance at the local hospital, local obituaries and queries to the National Death Index dates. Causes of death were reviewed by an endpoint panel of three investigators. The study protocol was approved by the IRB at Boston University Medical Center. All participants provided written informed consent at the time of each examination visit.

Dataset	**Total number of samples**	**Female**	**Age**	**Deaths and follow-up**
HRS biomarkers	9,933	52%	63.1 ± 14.4	NA
HRS methylation	4018	41%	69.6 ± 9.7	NA
FHS methylation	3935	47%	58.49 ± 13.2	319 with 6.8 ± 2.2 years

Table 1: Datasets used for Systems Age training, including information on total number of samples, female percentage, age distribution, deaths and follow-up years.

Systems Age pipeline

Term	Definition
Methylation PCs	Principal components generated by performing PCA on methylation data with 125K CpGs from HRS
System biomarker PCs	Principal components generated by performing PCA on groups of Biomarkers in HRS
DNAm system PCs	Methylation PC based surrogates of system biomarker PCs
System scores	Cox elastic net regression model scores generated by combining DNAm system PCs for each system
System score PCs	Principal components generated by performing PCA on the 11 system scores plus a chronological age prediction score in FHS
Systems Age	Whole body aging score generated by combining system score PCs using a cox elastic net regression model

Glossary 1: terms used to describe Systems Age and intermediate values derived during Systems Age calculation

Step 1: Grouping biomarkers into systems

We utilized molecular and cellular biomarker data from the Health and Retirement Study (HRS) 2016 Venous Blood Study (VBS), for which a subset also has paired DNA methylation data. We assessed the available biomarkers, and manually annotated them as biomarkers for specific physiological systems, totaling 11 systems. To each system we added functional biomarkers (e.g. grip strength) and system-specific disease and condition history (e.g. history of stroke or chronic lung disease).

Our goal was to develop epigenetic aging clocks that are interpretable in terms of physiological systems for clinical and epidemiological applications. Thus, for manual annotation, we required biomarkers to fulfill at least one of two criteria to be assigned to a system: 1) Is there evidence that the biomarkers predict risk of age-related diseases for that physiological system? 2) Would a clinician utilize the biomarker in assessing the status of that physiological system? Annotations were done by multiple team members supported by literature searches to validate disease prediction and clinical interpretations. The team included a gerontologist (M.E.L.) and a physician-scientist (A.H.C.). Most of the biomarkers were transformed and thresholded such that their distribution is more normal. The biomarker-to-system mapping, dataset-specific variable names, and transformations used can be found in Supplementary Table 1.

There is no gold standard list of biomarkers for each physiological system and there is often not a clear delineation between systems because of their biological integration. We do not claim that these are the only 11 systems or the only correct mapping of the Biomarkers to these 11 systems. The Systems Age pipeline can be easily adapted to other biomarker-to-system mappings. Our work here is intended as a proof-of-concept that omics clocks can capture aging in specific physiological systems, and thus the most important validation of our chosen mappings is the high specificity of the System scores in our WHI validation dataset, rather than the exact list of starting biomarkers.

Step 2: Principal component analysis of system biomarkers and DNA methylation data

We previously found that performing principal component analysis (PCA) and then using principal components (PCs) as input into supervised machine learning models produces more robust and reliable epigenetic clocks (Higgins-Chen et al. 2022) . PCA removes collinearity, reduces dimensionality of the data, and better separates signal from technical noise. Thus, for each system, we performed PCA on the selected system biomarkers. Before performing PCA, the biomarkers were first transformed to have a normal distribution as described in supplementary table 1 as well as scaled before inputting into the prcomp function (stats 4.1.1) in R. In parallel, we performed PCA on DNA methylation data as previously described (Higgins-Chen et al. 2022), utilizing 125,175 CpGs that 1) are in all of our training and validation data and 2) present on commercially available methylation arrays including the Infinium HumanMethylation450 BeadChip and Infinium Methylation EPIC BeadChip. Practically, this was done using the prcomp function in R. This yielded two sets of PCs (see Glossary 1 for terms): 1) system biomarker PCs (the number of PCs per system is equivalent to the number of biomarkers for each system, since number of samples is greater than number of features) and 2) 4,017 DNA methylation PCs (one less than number of samples, since the number of samples is less than number of features). We did not filter out low-variance PCs (for example using scree plots or random matrix theory methods). Low-variance PCs can still capture relevant variation for prediction (Jolliffe 1982; Yan et al. 2020; Aschard et al. 2014; Tarashansky et al. 2019), while those that are irrelevant are removed or minimized at later supervised machine learning steps. Thus, when predicting system biomarker PCs from DNA methylation PCs (Step 3), we can predict

both dominant, shared signals between biomarkers (high-variance PCs) as well as more subtle variations.

Step 3: Building DNAm surrogates of system PCs

We utilize elastic net regression to train a model using methylation PCs to predict each system biomarker PC using the glmnet 4.1-4 package in R. We refer to the resulting models as DNAm system PCs. This was done as described previously (Higgins-Chen et al. 2022). The L1 to L2 regularization ratio was 1 ($\alpha = 0.5$), the λ tuning parameter was selected via tenfold cross-validation, and the final methylation PC was excluded as it is not meaningful in cases where the number of samples is less than the number of features. Not all system PCs are predicted well using methylation PCs. We retained DNAm system PCs with at least 20 DNAm PCs being used at the minimum mean cross-validated error in the model and at least 5 DNAm PCs at the cross-validated error one standard error from the minimum mean cross-validated error in the model. This allows us to take only well predicted DNAm system PCs to the next step.

Step 4: Building system scores

To build system scores we first calculate DNAm system PCs in FHS based on parameters previously trained in HRS (first calculating methylation PCs, then predicting system PCs). Then, for each system separately, we predicted mortality using DNAm system PCs in a Cox elastic net mortality prediction model using the glmnet 4.1-4

package. The L1 to L2 regularization ratio was 1 (α = 0.5), the λ tuning parameter was selected via tenfold cross-validation. This yielded 11 separate mortality prediction models that we term system scores, and can serve as a measure of mortality-related deterioration of each system.

Step 5: Building Systems Age

To build Systems Age, we first perform PCA on the DNAm system scores and age prediction score using the prcomp(stats 4.1.1) function in R, as the system scores and age prediction score are partially correlated with one another.

The age prediction score is built specifically to predict chronological age and was trained in HRS. The DNAm PCs in HRS were first used to predict chronological age. The scores thus generated were then used to predict chronological age again but instead now using a second degree polynomial function fitted to the 5 year interval averages of the predicted chronological age score (previous step) predicting for the 5 year interval averages of chronological age. The score obtained from the second degree polynomial is referred to as age prediction in our model.

Using all system score PCs, we then predict mortality using another Cox elastic net mortality prediction model using glmnet 4.1-4 package in R. The L1 to L2 regularization ratio was 1 (α = 0.5), the λ tuning parameter was selected via tenfold cross-validation, Again, using PCs as input is intended to reduce redundancy, increase reliability, and

allow for more subtle variations in system scores to have an important role in the overall model.

Step 6: Scaling scores to age range

The 11 system scores and Systems Age are first standardized to have mean 0 and standard deviation 1. They are then scaled to match the mean and standard deviation of chronological age for the 3935 samples from FHS Offspring and Gen3 cohorts.

Association meta-analysis in Women's Health Initiative cohorts

The Women's Health Initiative (WHI) is a long-term national health study (The Women's Health Initiative Study Group 1998). WHI is funded by the National Heart, Lung, and Blood Institute, or NHLB and ran from the early 1990s to 2005. Post 2005, there have been Extension Studies, which continue to collect data on health outcomes annually. We used 3 WHI cohorts which had methylation data available. In each WHI cohort (Table 2) we calculated system scores then regressed all epigenetic aging clocks on chronological age using a linear regression model and defined clock age acceleration as the corresponding residual. We then calculated associations between these clock accelerations and different diseases and aging phenotypes in all WHI cohorts. We stratified the cohorts by race (except WHI AS311 where analysis of the Black and Hispanic populations would be underpowered), for a total of 7 groups. Depending on

condition and disease we either built linear regression models (cognitive function, physical function, comorbidities and more), cox prediction models (Lung Cancer, Breast Cancer, Leukemia, CHD, MI and more) or logistic regression models (Thyroid disease, Diabetes, Arthritis, Cataract and more) to look at associations with the age accelerated scores. An example of the formula used is as follows: *Cognitive function ~ AgeAccel + Age*. In certain cases, additional factors such as Education level (for Cognitive function) were also added to the models. Sex was not a covariate as all WHI participants are female. We combined the associations from the different cohorts and racial groups in a fixed effects model meta-analysis with inverse variance weights, obtaining meta-analysis Z-scores for the associations (package metafor 2.4-0, function rma and forest). Forest plots, z-scores, heterogeneity p-values and other meta-analysis results are provided in Supplementary Figure 2 and 4 and Supplementary Tables 2-7.

WHI cohort	Number of samples	White	Black	Hispanic	Current Smokers	Past smokers
BAA23	2107	998	676	433	213	756
AS311	855	760	70	25	73	393
EMPC	2167	1207	606	354	181	840

Table 2: WHI cohorts used for testing with racial distribution, percent current and percent past smokers

We performed additional analyses adjusting for smoking status by adding smoking status (present smoker, ex-smoker or never smoked) into the linear (package stats

4.0.2, function lm), Cox (package survival 3.2-11, function coxph) and logistic models (package stats 4.0.2, function glm(family="binomial")). We also examined non-smokers separately. A list of the variables used from WHI are shown in Table 3. It is important to note that even though multiple disease variables were available we could not test for a majority of the variables because they were underpowered. Rather, for many of the variables we calculated z-scores and showed them in our supplementary data. Variables which had insufficient N have not been listed below. (Table 3)

Condition	Variable used	Available in	Additional notes
Cognitive Function	100 - **F393MSE**	Form 39 - Cognitive Assessment	F393MSE score mapped using the year at which blood was drawn for methylation
Physical function	100 - **PHYSFUN**	Form 38 - Daily Life	PHYSFUN score mapped using the year at which blood was drawn for methylation
Mortality	**DEATHALL, ENDFOLLOWALLDY**	Adjudicated Outcomes as of	Corrected for time to date at which

		Mar 6, 2021	blood for methylation was drawn
CHD	**CHD, CHDDY**	Adjudicated Outcomes as of Mar 6, 2021	Corrected for time to date at which blood for methylation was drawn
MI	**MI, MIDY**	Adjudicated Outcomes as of Mar 6, 2021	Corrected for time to date at which blood for methylation was drawn
Lung Cancer	**LUNG, LUNGDY**	Adjudicated Outcomes as of Mar 6, 2021	Corrected for time to date at which blood for methylation was drawn
Breast Cancer	**BREAST, BREASTDY**	Adjudicated Outcomes as of Mar 6, 2021	Corrected for time to date at which blood for

			methylation was drawn
Endometrial Cancer	**ENDMTRL, ENDMTRLDY**	Adjudicated Outcomes as of Mar 6, 2021	Corrected for time to date at which blood for methylation was drawn
Leukemia	**LEUKEMIA, LEUKEMIADY**	Adjudicated Outcomes as of Mar 6, 2021	Corrected for time to date at which blood for methylation was drawn
Skin Cancer	**MELANOMA, MELANOMADY**	Adjudicated Outcomes as of Mar 6, 2021	Corrected for time to date at which blood for methylation was drawn
Colorectal Cancer	**COLORECTAL, COLORECTALDY**	Adjudicated Outcomes as of Mar 6, 2021	Corrected for time to date at which blood for methylation was

			drawn
Stroke	**STROKE, STROKEDY**	Adjudicated Outcomes as of Mar 6, 2021	Corrected for time to date at which blood for methylation was drawn
Thyroid Disease	**THYROID**	Form 30 Medical History	At baseline
Cataract	**CATARACT**	Form 30 Medical History	At baseline
Arthritis	**ARTHRIT**	Form 30 Medical History	At baseline
Diabetes	**DIAB**	Form 30 Medical History	At baseline
Any condition	**NACOND**	Form 30 Medical History	At baseline
Total Comorbidities	Calculated using all available variables in Form 30 such as **CVDEVER, DIAB, ALS,**	Form 30 Medical History	At baseline

	ARTHRIT, THYROID, CANC_F30 and more		
Smoking status	SMOKING	Form 34 - Personal Habits	
Age	AGE	Demographics and Study Membership	Corrected for time to date at which blood for methylation was drawn

Table 3: WHI variables used for testing associations of scores

ROC meta-analysis in Women's Health Initiative cohorts

For performing ROC meta analysis, we first converted all our variables to binary format. Continuous variables such as cognitive function and physical function were converted to binary variables by marking the lowest quintile as diseased and the rest as healthy. For total comorbidities we marked greater than 2 comorbidities as diseased. For time to event variables such as mortality, CHD, MI and more, we censored at t=6,000 days. The censoring time was determined manually by looking at the longest follow-up times for event-free individuals. Models for ROC analysis were built including the score along with age. Each ROC curve for a clock was either compared with Systems Age or best system score using roc.test() function in the pROC package in R. Meta-analysis was

then performed by analyzing Z-scores comparing each clock to Systems Age or best system score for all cohorts using Stouffer's method.

Calculating different clocks

In addition to Systems Age, we calculated a large number of additional existing clocks for comparison. We used the following packages or sources to do so (Table 4).

Clock	**Package/Source**
PC clocks (Higgins-Chen et al. 2022)	PC-clocks R package - https://github.com/MorganLevineLab/PC-Clocks
Methylation clocks except DunedinPACE and GrimAge (Thrush et al. 2022)	methylCIPHER R package - https://github.com/MorganLevineLab/methylCIPHER
DunedinPACE (Belsky et al. 2022)	DunedinPACE R package - https://github.com/danbelsky/DunedinPACE
GrimAge (Lu et al. 2019)	Pre-calculated values as provided on the WHI server

Table 4: Packages used for calculating epigenetic clocks

Aging Subtypes and overrepresentation of diseases in subtypes

Age-adjusted system scores were used to perform adaptive hierarchical clustering using the Dynamic Tree Cut library (dynamicTreeCut 1.63-1, function cutreeDynamicTree) in R. Parameters used other than default settings included minModuleSize which was set at 100. Based on the most stable node distance, 9 clusters were identified. Average score for each system for each cluster was plotted on polar spider plots. An over representation analysis comparing occurrence of disease in the cluster compared to the whole population was performed using Fisher's exact test. Binary disease status variables were used without transformation, continuous variables such as cognitive function and physical function were converted into binary variables by marking values lesser than 1 standard deviation from mean as disease states. For time-to-event variables, the model was built only for individuals who were alive until the 7 year follow-up or died because of the condition.

Association of Biomarkers with System specific scores in HRS

We wanted to show how the different Biomarkers used in training were associated with the methylation based system scores. For this purpose we built linear regression models associating the system score with each biomarker. The z-score of association was then plotted on the y axis for each score in supplementary figure 1.

Test-retest reliability analysis

Reliability was calculated as described before (Higgins-Chen et al. 2022). Briefly, reliability was calculated in GSE55763 which consisted of 36 whole-blood samples measured in duplicate (age range 37.3 to 74.6). We used the icc function in the irr R package version 0.84.1, using a single-rater, absolute-agreement, two-way random-effects model (Koo and Li 2016).

Belsky, Daniel W., Avshalom Caspi, David L. Corcoran, Karen Sugden, Richie Poulton, Louise Arseneault, Andrea Baccarelli, et al. 2022. "DunedinPACE, a DNA Methylation Biomarker of the Pace of Aging." ***eLife*** **11 (January). https://doi.org/10.7554/eLife.73420.**

Crimmins, E. M., B. Thyagarajan, and M. E. Levine. 2021. "Associations of Age, Sex, Race/ethnicity, and Education with 13 Epigenetic Clocks in a Nationally Representative US Sample: The Health and Retirement Study." ***The Journals of*****. https://academic.oup.com/biomedgerontology/article-abstract/76/6/1117/6102583.**

Higgins-Chen, A. T., K. L. Thrush, Y. Wang, and C. J. Minteer. 2022. "A Computational Solution for Bolstering Reliability of Epigenetic Clocks: Implications for Clinical Trials and Longitudinal Tracking." https://www.nature.com/articles/s43587-022-00248-2.

Kannel, William B., Manning Feinleib, Patricia M. McNamara, Robert J. Garrison, and William P. Castelli. 1979. "An Investigation of Coronary Heart Disease in Families: The Framingham Offspring Study." ***American Journal of Epidemiology*** **110 (3): 281–90.**

Lu, Ake T., Austin Quach, James G. Wilson, Alex P. Reiner, Abraham Aviv, Kenneth Raj, Lifang Hou, et al. 2019. "DNA Methylation GrimAge Strongly Predicts Lifespan and Healthspan." ***Aging*** **11 (2): 303–27.**

Splansky, Greta Lee, Diane Corey, Qiong Yang, Larry D. Atwood, L. Adrienne Cupples, Emelia J. Benjamin, Ralph B. D'Agostino, et al. 2007. "The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: Design, Recruitment, and Initial Examination." ***American Journal of Epidemiology*** **165 (11): 1328–35.**

The Women's Health Initiative Study Group. 1998. "Design of the Women's Health Initiative Clinical Trial and Observational Study." ***Controlled Clinical Trials*** **19 (1): 61–109.**

Thrush, Kyra L., Albert T. Higgins-Chen, Zuyun Liu, and Morgan E. Levine. 2022. "R methylCIPHER: A Methylation Clock Investigational Package for Hypothesis-Driven Evaluation & Research." ***bioRxiv*****. https://doi.org/10.1101/2022.07.13.499978.**

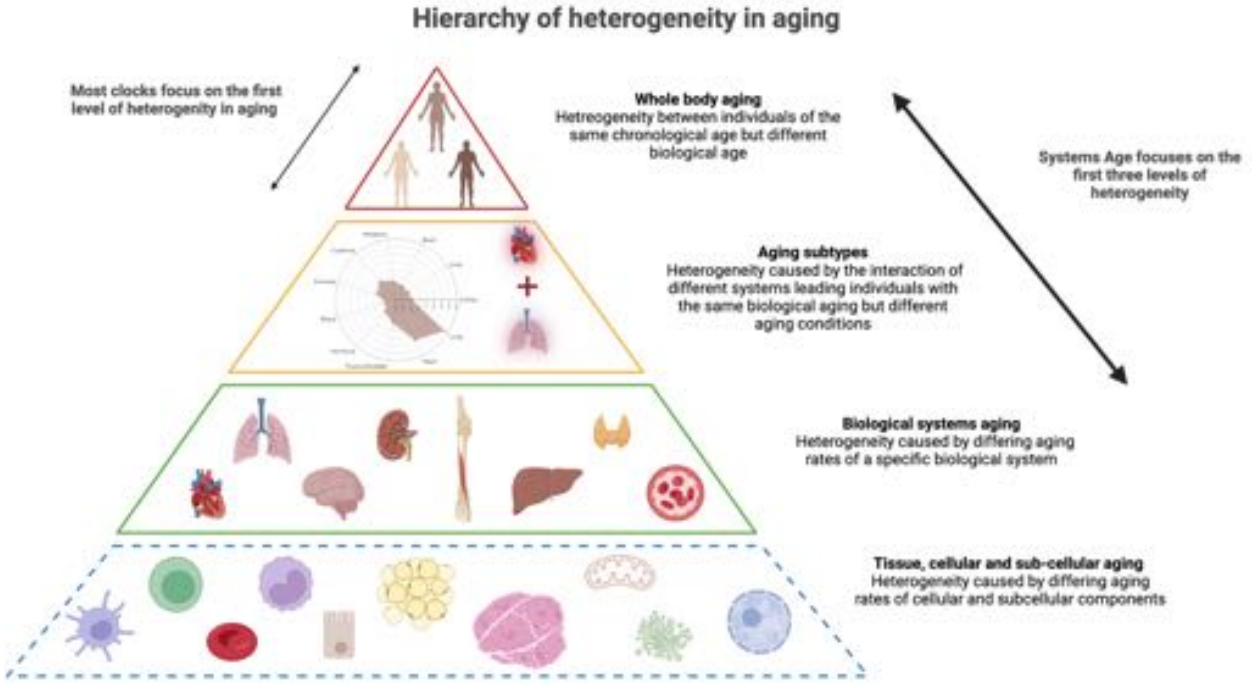

Figure 1 : Hierarchy of heterogeneity in aging. Heterogeneity in aging starts at the very cellular and subcellular levels due to genetic and environmental factors. These variations in aging go on to accumulate at the tissue, organ and the biological system level causing differences in the rates of aging of different systems within an individual. These systems do not behave independently of each other and this leads to certain common patterns of deterioration across systems giving rise to aging subtypes. Eventually, all of these variations accumulate at the whole body level to cause variations in overall aging rates across individuals. Most epigenetic aging clocks typically focus on the whole body aging level of heterogeneity, Systems Age attempts to capture both the systems level heterogeneity and aging subtypes other than the whole body aging itself. Image created using Biorender.com.

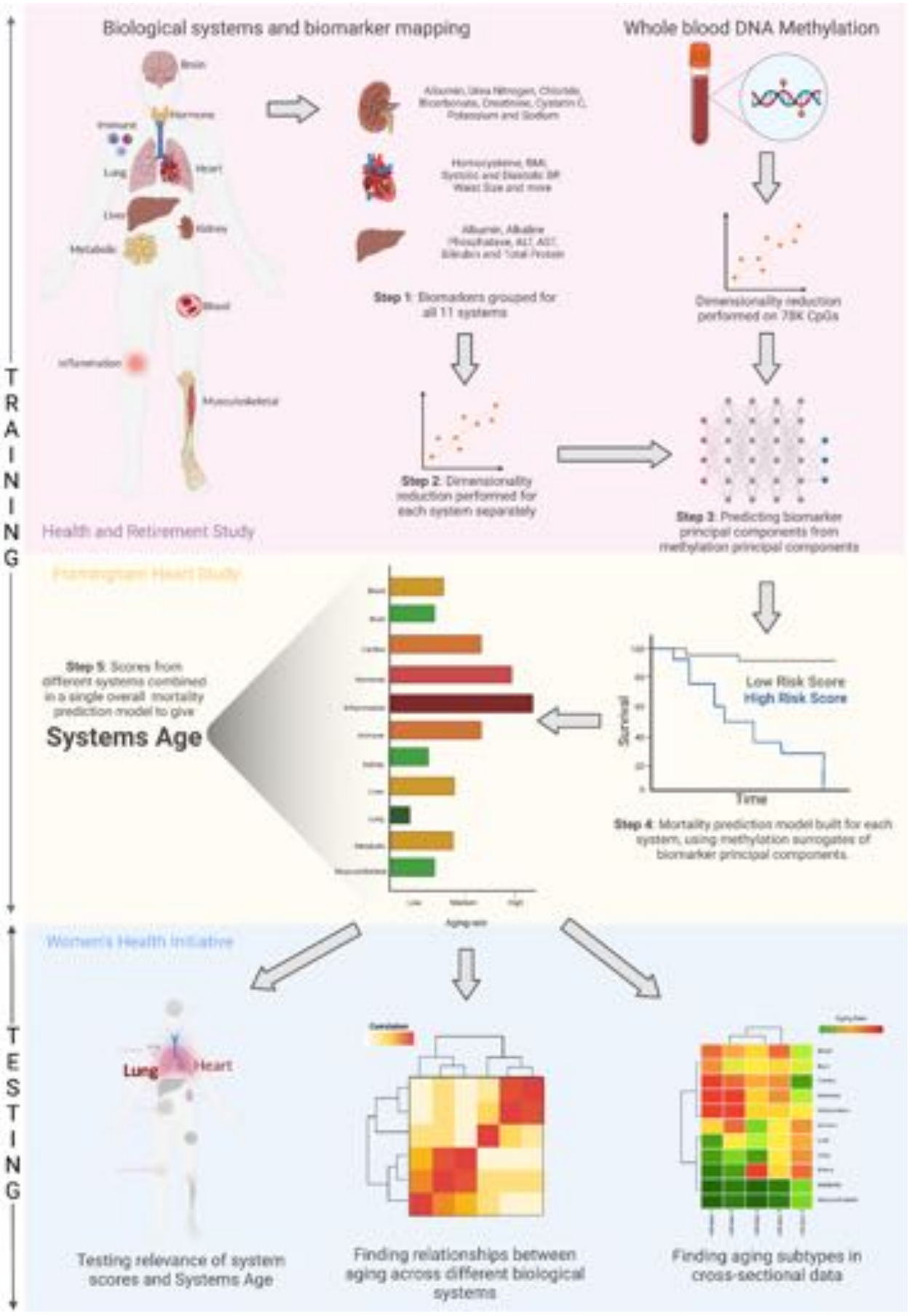

Figure 2: Analysis pipeline: Step 1 - Grouping Biomarkers into systems; Step 2 - Deconvoluting systems into principal components; Step 3 - Building DNAm surrogates of system PCs using ElasticNet regression; Step 4 - Building system scores by combining system PCs using Cox ElasticNet regression; Step 5 - Building Systems Age by combining system scores using Cox ElasticNet regression. Training done in HRS and FHS datasets while testing for specificity and aging subtypes done in WHI. Image created using Biorender.com.

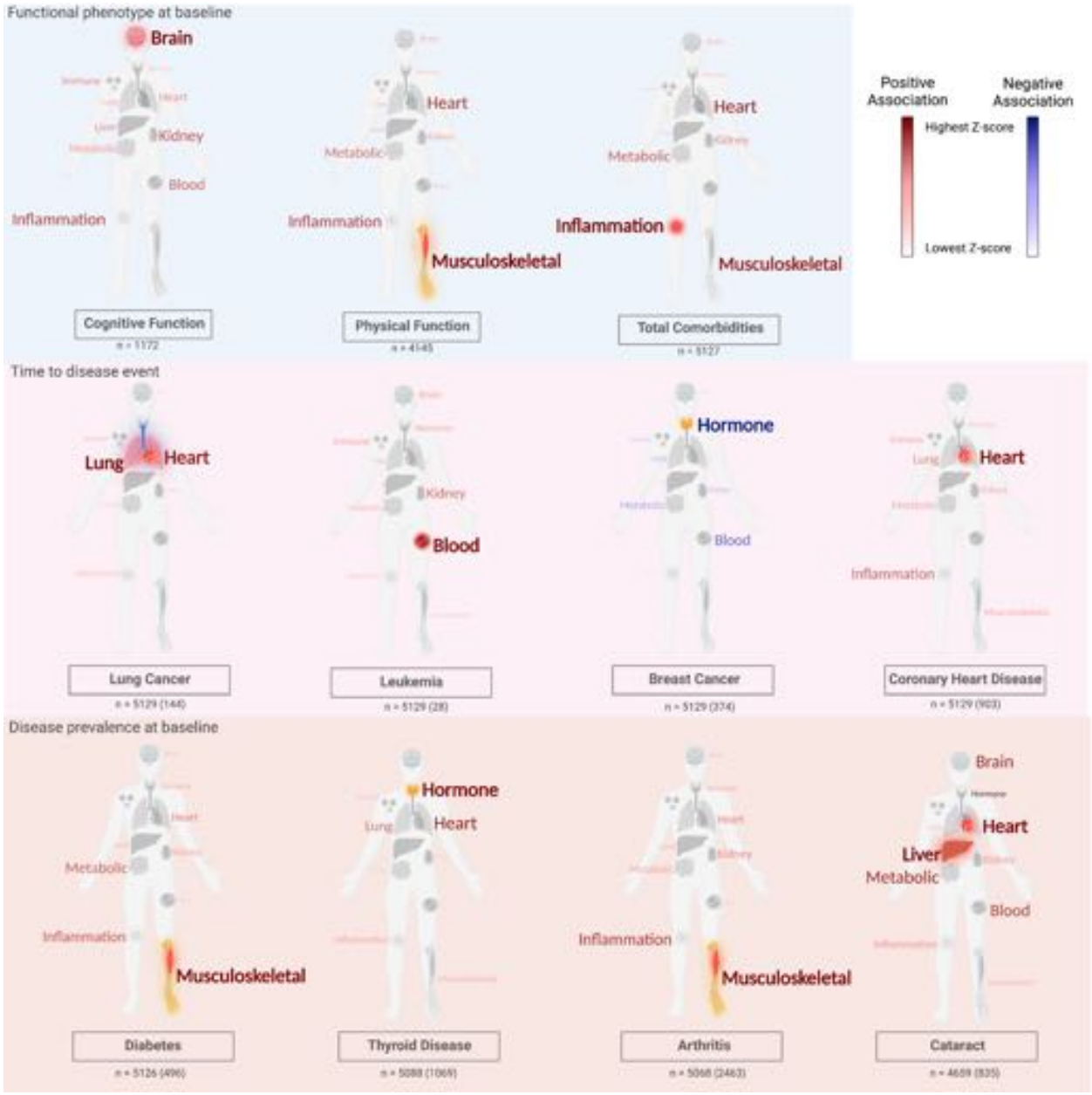

Figure 3 : Meta-analysis associations (z-scores calculated using a race stratified analysis of 3 WHI datasets) for specific diseases and aging phenotypes with system score age accelerations depicted with text size and color. Additionally the system(s) with

the highest positive association (or lowest in case of negative association) is bolded and the organ colored on the human figure. N for functional phenotypes ranges between 1172 and 5127. For time to disease events and disease prevalence at baseline, total N as well as number of events or individuals with diseases has been provided (in brackets). Total N for the time to disease events and disease prevalence at baseline is typically around 5000. For functional phenotypes at baselines Ordinary Least Squares regression model was used, for time to disease events cox proportional hazard models were used and for disease prevalence at baseline logistic regression models were used. Models built for each racial group separately and then meta-analyzed via a fixed effects model with inverse variance weights. *Exact z-scores as well as heterogeneity p-values and other phenotypes are given in supplementary table 2 and 3. Image created using Biorender.com.*

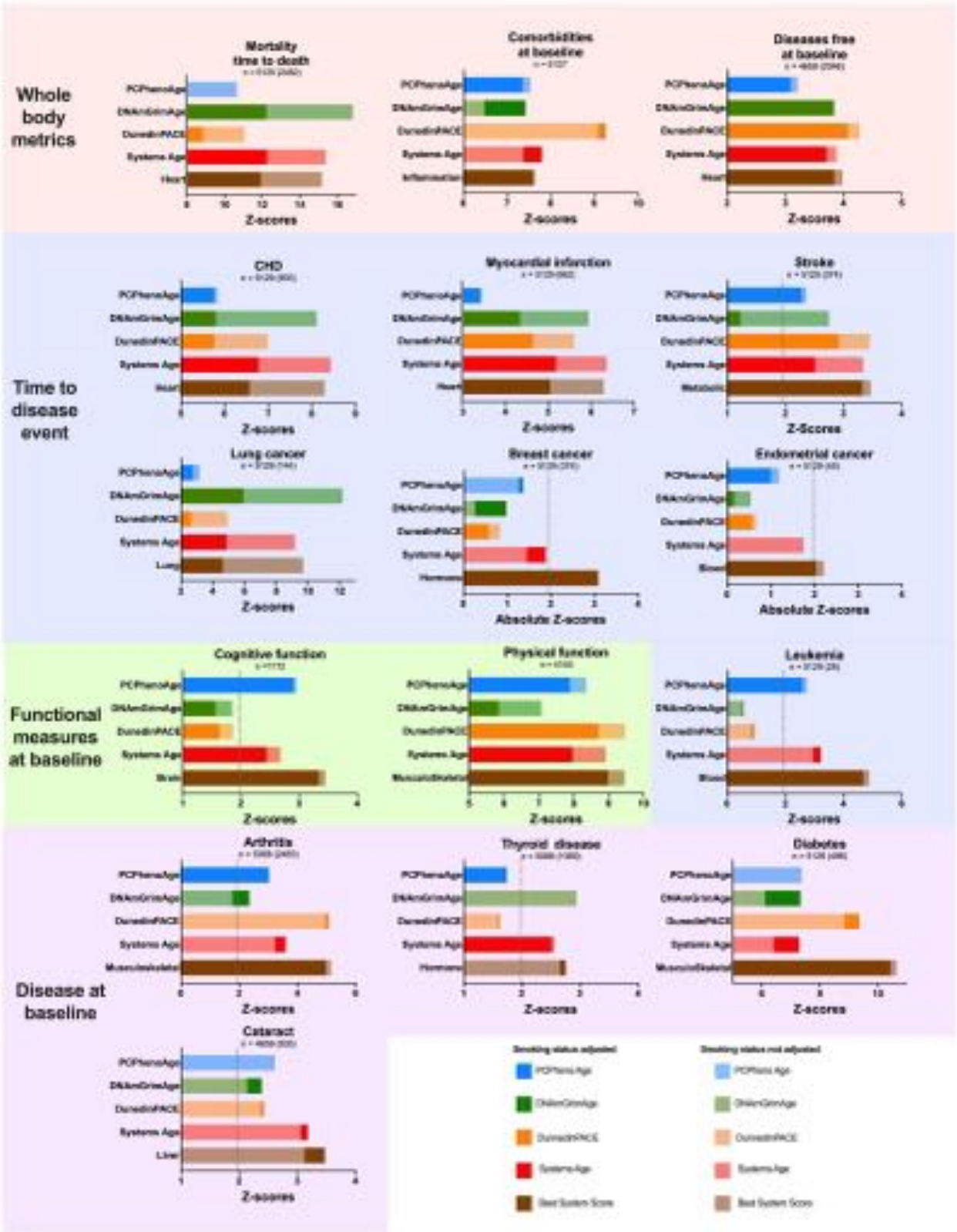

Figure 4: Meta-analysis associations (z-scores calculated using a race stratified analysis of 3 WHI datasets) for specific diseases and aging phenotypes with age accelerations of different clocks, Systems Age and the best system score plotted for smoking status adjusted (in darker shades) and no smoking status adjusted (lighter shades). N for functional measures ranges between 1172 and 4145. For time to disease events and disease prevalence at baseline, total N as well as number of events or individuals with diseases has been provided (in brackets). Total N for the time to disease events and disease prevalence at baseline is typically around 5000. For functional measures at baseline Ordinary Least Squares regression model was used, for time to disease events cox proportional hazard models were used and for disease prevalence at baseline logistic regression models were used. Models built for each racial group separately and then meta-analyzed via a fixed effects model with inverse variance weights. *Exact z-scores as well as heterogeneity p-values are given in supplementary table 2, 3, 4 and 5. Plots generated using Prism 9.*

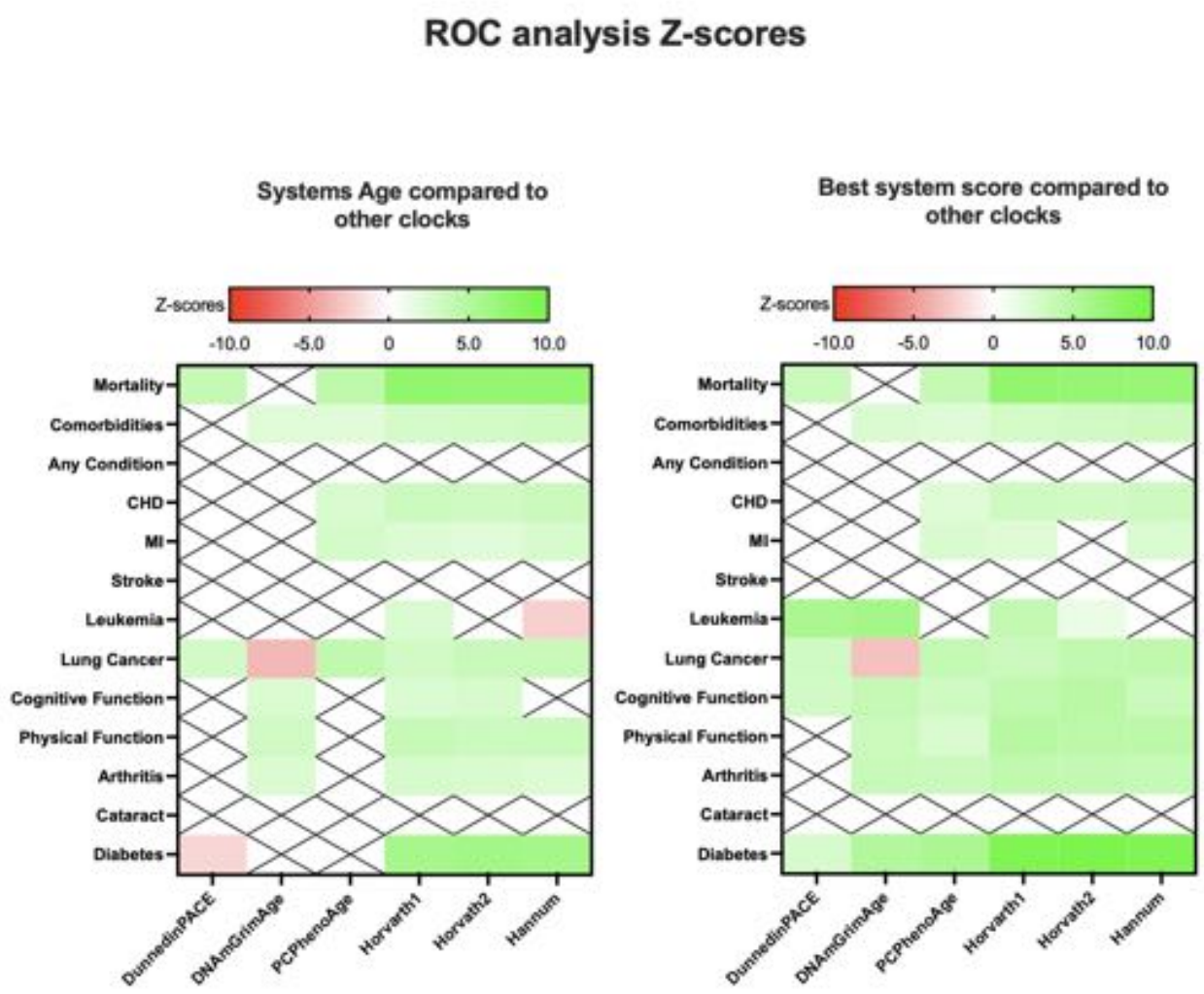

Figure 5: ROC analysis comparing each clock to Systems Age or best system score. Z-scores were calculated using cohort stratified meta analysis using Stouffer's method on 3 WHI cohorts: EMPC, AS311 and BAA23. Scores with no significant difference have been crossed out. Significance is defined as absolute z-scores with values greater than 1.96. Plots generated using Prism 9.

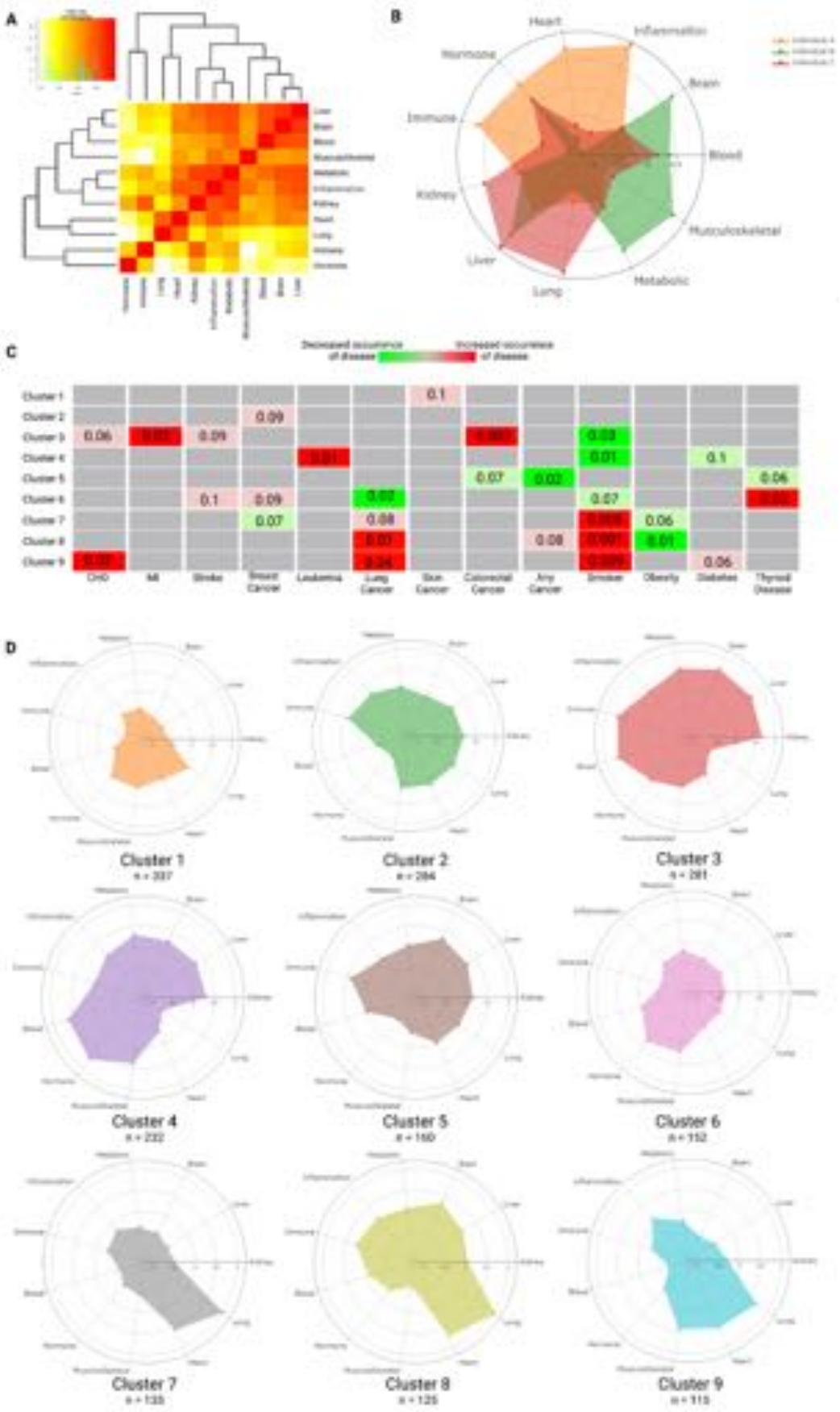

Figure 6: Discovering Aging Subtypes. A) Correlations between system scores across all WHI cohorts corrected for batch effects (N = 5129). Exact correlations provided in supplementary table 8 B) Three chronological age matched individuals with the same race and gender as well as similar age-accelerated Systems Age having very different age-accelerated system scores. C) Overrepresentation analysis of presence or absence of diseases amongst individuals from 9 different clusters. P Values have been calculated using fisher's exact test and are available in supplementary table 9. D) Mean age accelerated score has been depicted for each cluster using a spider plot and is also available in supplementary table 10. Image created using Biorender.com.