

# Thymic health consequences in adults

<https://doi.org/10.1038/s41586-026-10242-y>

Received: 13 January 2025

Accepted: 5 February 2026

Published online: 18 March 2026

Open access

 Check for updates

Simon Bernatz<sup>1,2,3,4,18</sup>, Vasco Prudente<sup>1,2,3,18</sup>, Suraj Pai<sup>1,2,3,18</sup>, Asbjørn K. Attermann<sup>1,5,6,7</sup>, Yumeng Cao<sup>8</sup>, Jiachen Chen<sup>8</sup>, Asya Lyass<sup>9</sup>, Borek Foldyna<sup>1,10</sup>, Leonard Nürnberg<sup>1,2,3</sup>, Keno Bressemer<sup>1,2,11,12</sup>, Christopher Abbosh<sup>11</sup>, Charles Swanton<sup>11,12,13</sup>, Mariam Jamal-Hanjani<sup>11,13,14</sup>, Michael T. Lu<sup>1,10</sup>, Joanne M. Murabito<sup>15,16,17</sup>, Kathryn L. Lunetta<sup>8,15,19</sup>, Nicolai J. Birkbak<sup>5,6,7,19</sup> & Hugo J. W. L. Aerts<sup>1,2,3,10,19</sup>✉

The thymus is essential for establishing T cell diversity early in life, but undergoes profound involution with age and has therefore traditionally been regarded as largely nonfunctional in adults<sup>1,2</sup>. Here we propose that preserving thymic functionality is integral to adult health and longevity. We developed a deep learning framework to quantify thymic health from routine radiographic images and evaluated its association with longevity and risk of major age-associated diseases in two large prospective cohorts of asymptomatic adults: the National Lung Screening Trial ( $n = 25,031$ ) and the Framingham Heart Study ( $n = 2,581$ ). In both cohorts, thymic health varied markedly across the population. In the National Lung Screening Trial, higher thymic health was consistently associated with lower all-cause mortality, reduced lung cancer incidence and lower cardiovascular mortality over 12 years of follow-up after adjustment for age, sex, smoking and comorbidities. In the independent Framingham Heart Study cohort, higher thymic health was significantly associated with reduced cardiovascular mortality, independent of age, sex and smoking. Thymic health was further linked to systemic inflammation and metabolic dysregulation, and associated with modifiable lifestyle factors including smoking, obesity and physical activity. Together, these findings reposition the thymus as a central regulator of immune-mediated ageing and disease susceptibility in adulthood, highlighting its potential as a target for preventive and regenerative strategies to promote healthy ageing and longevity.

The thymus is a specialized immune organ responsible for maturing T cells, thereby producing a diverse T cell repertoire crucial for mounting an adaptive immune response<sup>1,2</sup>. The thymus itself decays with age and eventually transforms entirely into adipose tissue through a process known as thymic involution<sup>3</sup>. While the absence of a functioning thymus in children is associated with profound immunodeficiency<sup>4</sup>, the consequences of thymic decay in adulthood are more subtle<sup>5,6</sup>. Indeed, it was long believed that once the thymus generates a sufficiently diverse T cell repertoire in childhood, the T cell repertoire could be peripherally maintained to support an adaptive immune response against a diverse array of pathogens<sup>2,7</sup>. For this reason, the thymus has long been considered largely nonfunctional in adults.

However, a growing body of evidence challenges this notion<sup>2,6,8–17</sup>. In a recent landmark study by Kooshesh et al.<sup>8</sup> investigating the impact of thymectomy on long-term health, the authors found that adults who had their thymus removed experienced adverse health consequences across multiple diseases and outcomes, where penetrance can

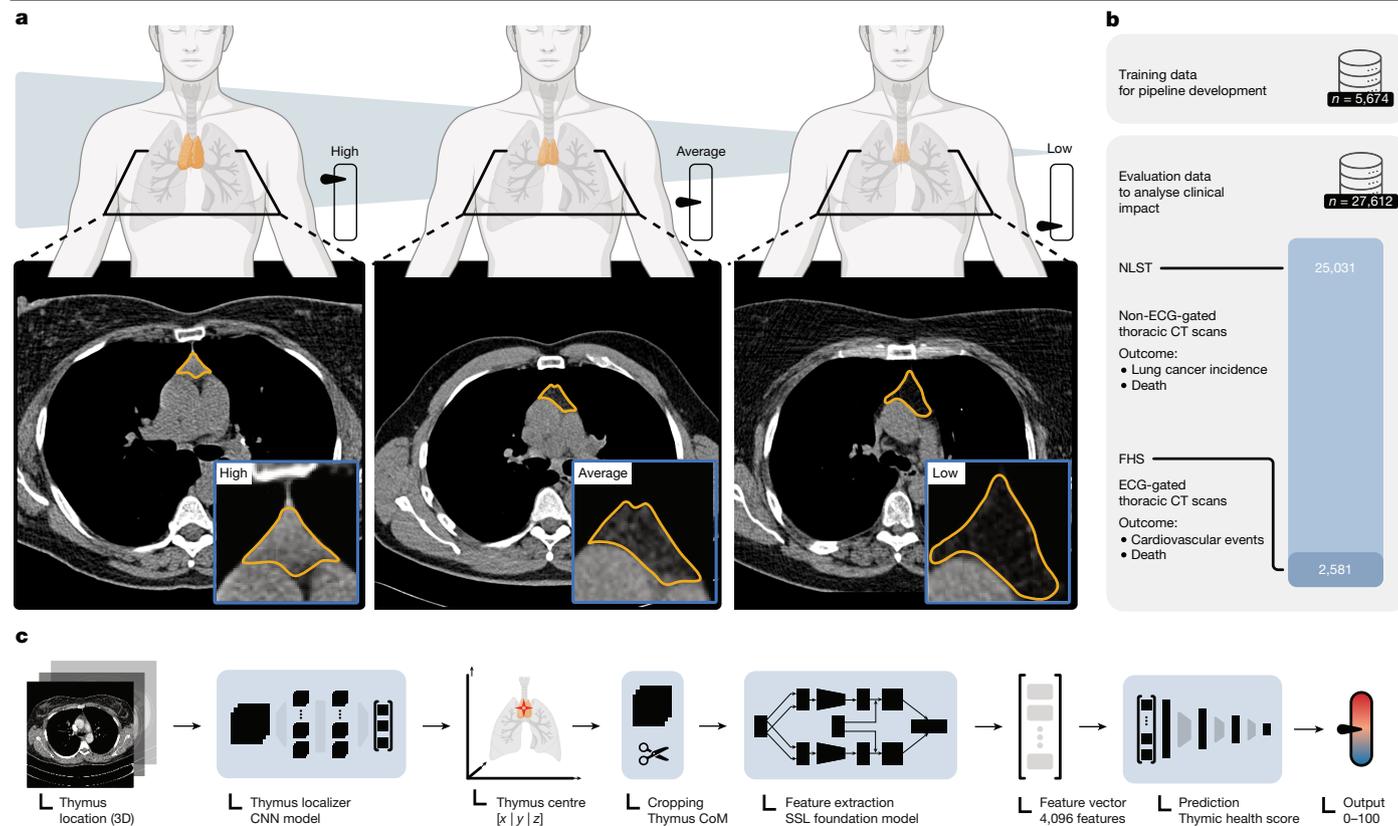
be decades after thymectomy<sup>8</sup>. While the consequences of thymectomy are impactful, only a small fraction of the population is exposed to this procedure, whereas individual and lifestyle-dependent differences in thymic decay affect everyone.

Although the impact of thymic function is increasingly recognized in ageing and across a wide range of clinical settings, thymic decay across the population has been incompletely explored<sup>2,13</sup>. However, there is increasing evidence that the rate of thymic involution varies among individuals<sup>18–20</sup>. Indeed, given the role of the thymus in maintaining an adaptive immune response, the individualized rate of thymic decay may be a major driver of age-associated diseases, such as cardiovascular disease and cancer<sup>13</sup>.

In this study, we investigated the impact of thymic functionality, here called thymic health, in adults. For this purpose, we analysed two prospectively collected cohorts of 27,612 individuals enrolled in the Framingham Heart Study (FHS) and National Lung Screening Trial (NLST). We developed a deep learning system to automatically

<sup>1</sup>Artificial Intelligence in Medicine (AIM) Program, Mass General Brigham & Harvard Medical School, Boston, MA, USA. <sup>2</sup>Department of Radiation Oncology, Brigham and Women's Hospital and Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA. <sup>3</sup>Radiology and Nuclear Medicine, GROW & CARIM, Maastricht University, Maastricht, The Netherlands. <sup>4</sup>Department of Radiology and Nuclear Medicine, Goethe University, Frankfurt am Main, Germany. <sup>5</sup>Department of Molecular Medicine, Aarhus University Hospital, Aarhus, Denmark. <sup>6</sup>Department of Clinical Medicine, Aarhus University, Aarhus, Denmark. <sup>7</sup>Bioinformatics Research Center, Aarhus University, Aarhus, Denmark. <sup>8</sup>Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA. <sup>9</sup>Department of Mathematics and Statistics, Boston University, Boston, MA, USA. <sup>10</sup>Cardiovascular Imaging Research Center, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. <sup>11</sup>Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, London, UK. <sup>12</sup>Cancer Evolution and Genome Instability Laboratory, The Francis Crick Institute, London, UK. <sup>13</sup>Department of Medical Oncology, University College London Hospitals NHS Foundation Trust, London, UK. <sup>14</sup>Cancer Metastasis Laboratory, University College London Cancer Institute, London, UK. <sup>15</sup>Framingham Heart Study, National Heart, Lung and Blood Institute and Boston University Chobanian and Avedisian School of Medicine, Framingham, MA, USA. <sup>16</sup>Department of Medicine, Boston University Chobanian and Avedisian School of Medicine, Boston, MA, USA. <sup>17</sup>Section of General Internal Medicine, Boston Medical Center, Boston, MA, USA. <sup>18</sup>These authors contributed equally: Simon Bernatz, Vasco Prudente, Suraj Pai. <sup>19</sup>These authors jointly supervised this work: Kathryn L. Lunetta, Nicolai J. Birkbak, Hugo J. W. L. Aerts.

✉e-mail: haerts@bwh.harvard.edu



**Fig. 1 | Overview of study design.** **a**, Illustration of thymic health, that is, an imaging-based proxy of thymic functionality and three representative examples of individuals with high, average or low thymic health. The thymus bed is outlined in orange. **b**, The model was developed on 5,674 CT scans and validated on 27,612 independent CT scans from the FHS and NLST. **c**, Illustration of the deep learning pipeline, which takes a CT scan as input and outputs a continuous

quantification of thymic health after automatically localizing and quantifying the thymus on the basis of self-supervised learning (SSL). CNN, convolutional neural network; CoM, centre of mass; ECG, electrocardiogram. Illustrations in **a** and **c** created in BioRender; Birkbak, N. <https://biorender.com/bd3dmmr> (2026).

quantify thymic health on computed tomography (CT) scans. Our results demonstrate that thymic health varies between individuals and is impacted by sex, age and lifestyle habits. Notably, we show that individuals with low thymic health, that is, lost thymic functionality, have a shorter lifespan and an increased risk of cancer and cardiovascular diseases. These findings strongly suggest that thymic health is crucial for long-term health and lifespan.

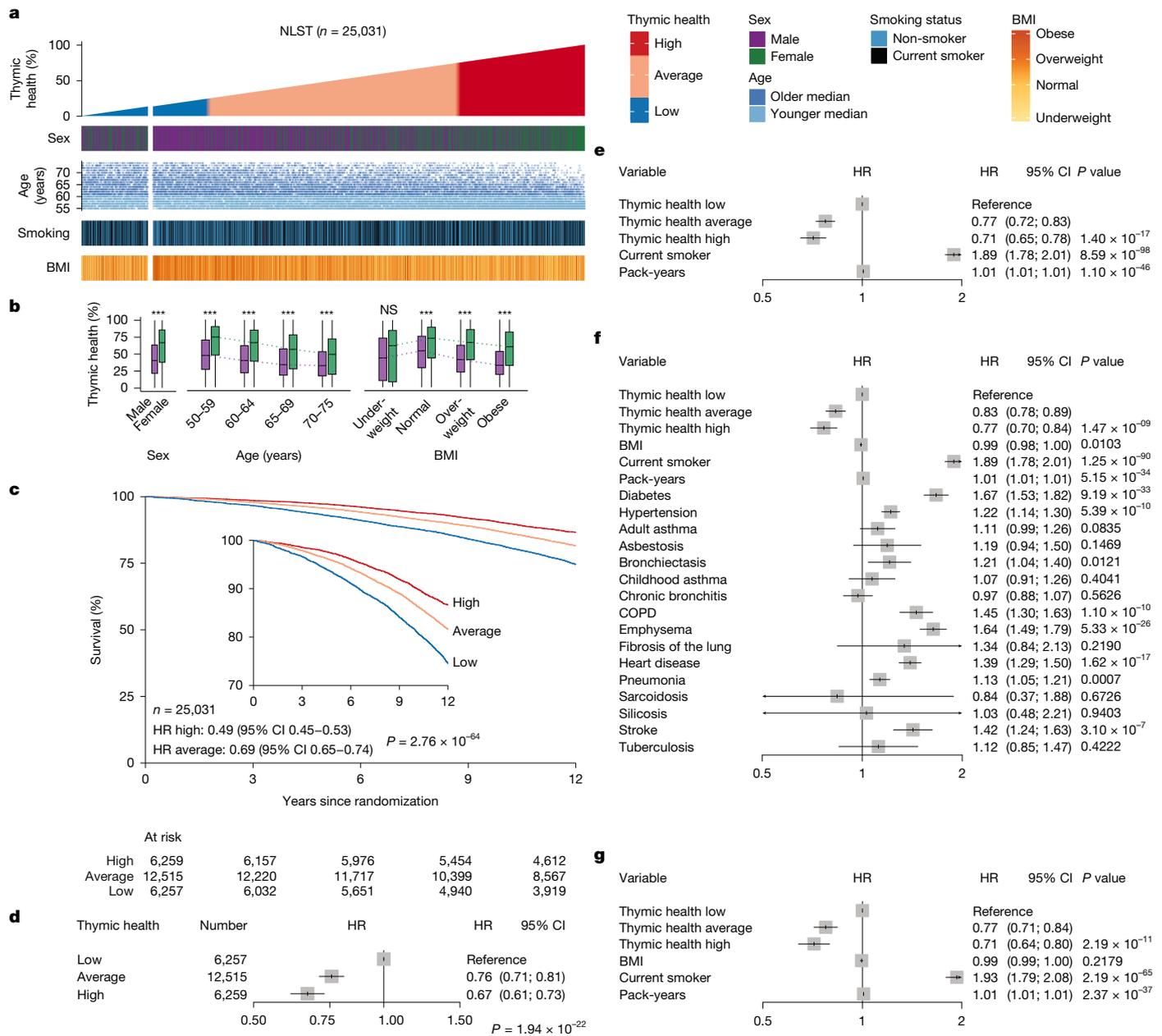
### Quantification of thymic health

For quantification of thymic health, we developed a deep learning system using an independent dataset of 5,674 individuals to determine compositional radiographic characteristics of the thymus as a proxy for its functionality (Fig. 1, Methods and Supplementary Fig. 9). The system takes a CT scan as input and provides the automatic continuous thymic health estimate as output. We applied the system to prospectively collected data from a total of 27,612 individuals from two cohorts, including 2,581 participants in the FHS and 25,031 participants in the NLST (Fig. 1). Clinical characteristics are shown in Extended Data Table 1. For outcome analyses, participants were categorized as low, average or high thymic health based on the bottom 25%, middle 50% and top 25% of the population. These thresholds were supported by cut point iterations (Extended Data Fig. 1). Individuals with low thymic health are assumed to have lost most of their thymic functionality, while individuals with average to high thymic health have preserved their thymic functionality throughout ageing to an increasing degree. As expected<sup>21,22</sup>, thymic health was higher in female than male participants and significantly declined with age (Fig. 2a,b, Extended Data Fig. 2a,b and Supplementary Fig. 2).

Furthermore, thymic health was lower in individuals with higher body mass index (BMI) (Fig. 2b, Extended Data Fig. 2b and Supplementary Fig. 2).

### Thymic health and risk of death

To investigate the associations of thymic health with clinical outcomes, we assessed all-cause mortality throughout a 12-year follow-up period using Kaplan–Meier and Cox proportional hazards analysis. For the NLST, participants with higher thymic health showed lower mortality than those with low thymic health (high versus low thymic health: Kaplan–Meier mortality estimate at 12 years 13.4% versus 25.5%; hazard ratio (HR) 0.49; 95% confidence interval (CI) 0.45–0.53; Fig. 2c). The association was preserved in a Cox analysis adjusted for smoking status and pack-years, and stratified by sex and 5-year age bins (Fig. 2d). To verify that these results were not confounded by age, we repeated the analysis using narrower 3-year age bins (Supplementary Fig. 3a) and, in a separate model, used age as the time scale to implicitly adjust for the effect of ageing (Fig. 2e). To further evaluate the potential impact of comorbidities, we extensively adjusted for clinical variables and known diseases in a Cox model (Fig. 2f), and further repeated the analysis in a healthier sub-cohort where individuals with prior occurrence of cancer or major comorbidities were excluded (Fig. 2g). Throughout all analyses, thymic health remained significantly associated with outcomes after controlling for the potential confounding effects of either age or comorbidities (all type III  $P < 0.001$ ), firmly indicating a prognostic value of thymic health beyond clinical variables. Finally, excluding thymic health from multivariate models did

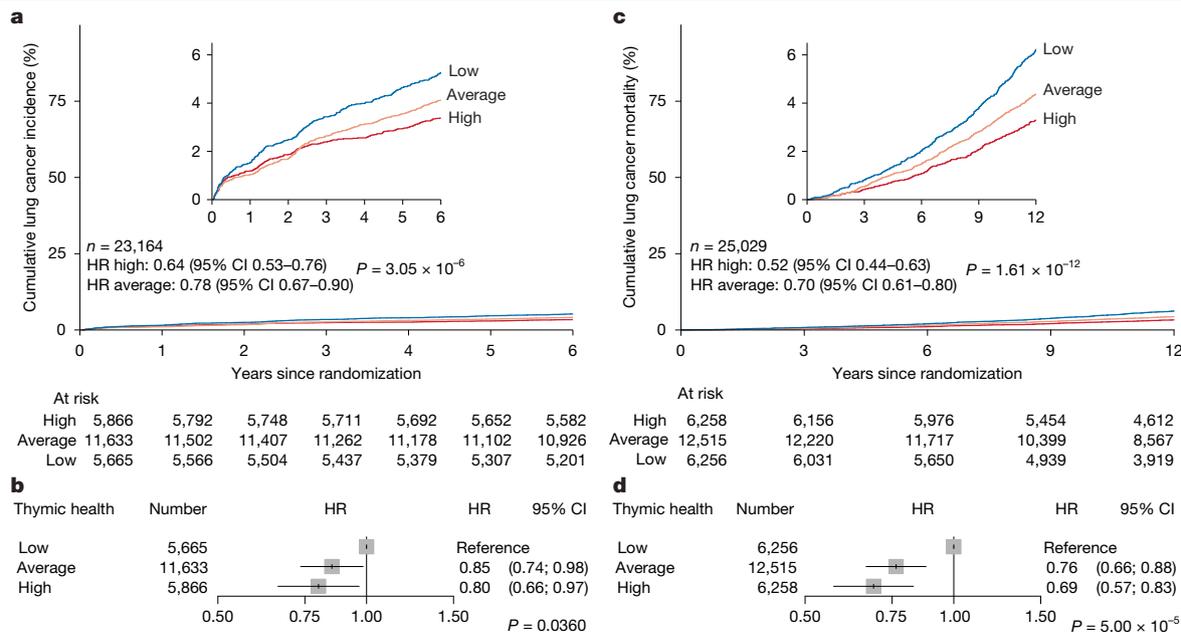


**Fig. 2 | Association of thymic health with long-term mortality.** **a**, Overview of the NLST. Data are sorted by ascending thymic health where each column represents one patient. Thymic health is categorized into low, average and high based on the bottom 25% (blue), middle 50% (orange) and top 25% (red) of the population. The fractional left split of the NLST represents patients who were defined as having low thymic health through automatic quality control. **b**, Associations between thymic health and sex ( $n = 25,031$ ) in the NLST, across age groups in years ( $n = 25,031$ ) and body mass categories ( $n = 24,948$ , missingness  $n = 83$ ).  $***P < 2 \times 10^{-16}$ ; NS, not significant with  $P = 0.2505$ . **c**, Kaplan-Meier plots for overall survival outcomes in the NLST across thymic health categories. The inset in the plot shows the same data on an expanded y-axis. Unadjusted HRs are shown on the bottom left. **d**, HR for thymic health categories adjusted for pack-years and smoking status and stratified by sex and age

binned at 5 years. **e-g**, HRs of all-cause death for participants in the NLST, using continuous age as the time scale to account for potential residual confounding by age ( $n = 25,027$ ; missingness  $n = 4$ ) adjusted for pack-years and smoking status and stratified by sex (**e**); adjusted for the shown clinical and epidemiological covariates and stratified by sex and age binned at 5 years ( $n = 24,597$ ; missingness  $n = 434$ ) (**f**); adjusted for the shown clinical and epidemiological covariates and stratified by sex and age binned at 5 years in the subgroup of the NLST in which participants with a history of cancer ( $n = 575$ ), childhood or adult asthma, diabetes, asbestosis, bronchiectasis, lung fibrosis, sarcoidosis, silicosis or tuberculosis ( $n = 5,500$ ) were excluded from the analysis ( $n = 18,565$ ; missingness  $n = 54$ ) (**g**). COPD, chronic obstructive pulmonary disease.

not markedly alter the HR estimates of the remaining covariates, indicating that thymic health provides independent and potentially complementary prognostic information (Extended Data Fig. 3a,e). These findings were supported in the FHS with similar effect size and direction (high versus low thymic health: 3.9% versus 14.5% Kaplan-Meier mortality estimate at 12 years; HR 0.24; 95% CI 0.16–0.38;

average versus low thymic health: 9.6% versus 14.5%; HR 0.63; 95% CI 0.48–0.83; Extended Data Fig. 2c) (type III  $P < 0.001$ ), although statistical significance was not reached after multivariate modelling in this smaller cohort (Extended Data Fig. 2d) (type III  $P = 0.254$ ). Continuous and detailed cutoff analyses testing a wide range of thresholds confirmed these results, showing gradually improving survival



**Fig. 3 | Association of thymic health with long-term risk of lung cancer and lung cancer-specific mortality. a**, Percentage of individuals who did develop lung cancer. **b**, HR of new lung cancer stratified by sex and age and adjusted for pack-years and smoking status. **c**, Lung cancer-specific mortality. **d**, HR of death from lung cancer stratified by sex and age and adjusted for pack-years and smoking status. The insets in the inverted Kaplan–Meier plots show the same data on an expanded y-axis. **a–d**, Cox proportional hazards regression

was used to estimate HRs. In the forest plots, the centre of each box represents the estimated HR, and the whiskers denote the corresponding 95% CI; shaded box size is for visualization only and does not encode statistical weight. The overall contribution of thymic health to uni- or multivariable models was evaluated using likelihood ratio tests ( $\chi^2$  tests) comparing full models with nested models excluding thymic health (type III test, two-sided) without adjustments for multiple comparisons.

independent of sex and age with increasing thymic health (Extended Data Fig. 1).

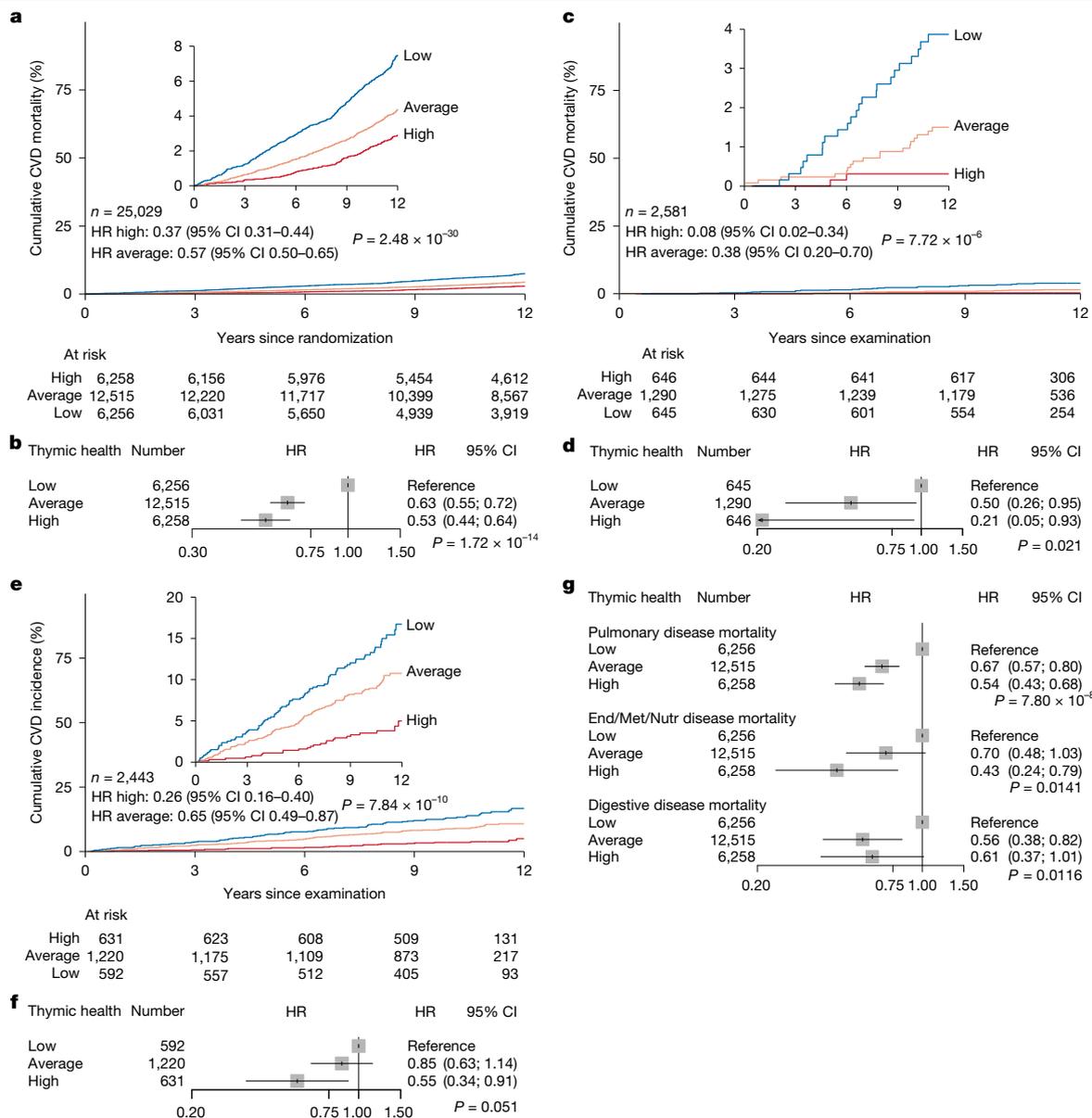
### Lung cancer incidence and mortality

We investigated the associations of thymic health with lung cancer mortality and lung cancer incidence in the NLST, the main endpoints of the trial. Participants with high and average thymic health were less likely to develop lung cancer than participants with low thymic health (high versus low thymic health: 3.4% versus 5.3% Kaplan–Meier incidence estimate at 6 years; HR 0.64; 95% CI 0.53–0.76; average versus low thymic health: 4.1% versus 5.3%; HR 0.78; 95% CI 0.67–0.90; Fig. 3a). These associations were preserved in sex and age-stratified analyses with adjustments for smoking status and pack-years (type III  $P = 0.036$ ) (Fig. 3b). Again, these associations were consistent across further analyses, extensively controlled for the confounding effects of age and comorbidities (Extended Data Figs. 3b,f, 4a,d and 5a and Supplementary Fig. 3b) (all type III  $P < 0.05$ ). Furthermore, participants with higher thymic health had lower lung cancer mortality risks relative to participants with low thymic health (high versus low thymic health: 1.1% versus 2.0% Kaplan–Meier mortality estimate at 6 years; HR 0.52; 95% CI 0.44–0.63; average versus low thymic health: 1.5% versus 2.0%; HR 0.70; 95% CI 0.61–0.80; HR reported through 12 years; Fig. 3c). Again, these associations were preserved across the respective multivariate models (all type III  $P < 0.001$ ) (Fig. 3d, Extended Data Figs. 3c,g, 4b,e and 5b and Supplementary Fig. 3c). We further stratified the cohort into current and former smokers. For lung cancer incidence, the associations with thymic health showed similar trends in both groups but did not reach statistical significance (current smokers type III  $P = 0.051$ ; former smokers type III  $P = 0.25$ ). For lung cancer mortality, the associations were preserved in both current (type III  $P = 0.009$ ) and former smokers (type III  $P = 0.003$ ) (Supplementary Fig. 4). While the NLST was designed to screen for lung cancer, pan-cancer mortality was also recorded (Supplementary Table 1). Again, we observed that participants

with higher thymic health had lower cancer-specific mortality risks relative to participants with low thymic health, which were also preserved in multivariate models (all type III  $P < 0.02$ ) (Extended Data Figs. 3c and 6a,b and Supplementary Fig. 3d).

### Cardiovascular mortality and incidence

We investigated the associations of thymic health with cardiovascular-specific mortality in the NLST, where participants with high and average thymic health had lower cardiovascular-specific mortality risks than participants with low thymic health (high versus low thymic health: 2.9% versus 7.5% cardiovascular-specific mortality estimates at 12 years; HR 0.37; 95% CI 0.31–0.44; average versus low thymic health: 4.4% versus 7.5%; HR 0.57; 95% CI 0.50–0.65) (Fig. 4a and Supplementary Table 2). Again, these associations were preserved in sex and age-stratified analyses with adjustments for smoking status and pack-years (Fig. 4b) and also consistent across further sensitivity analyses, extensively controlled for the confounding effects of age and comorbidities (Extended Data Figs. 3d,h, 4c,f and 5d and Supplementary Fig. 3e) (all type III  $P < 0.001$ ). Similar results were obtained in the independent FHS. Participants with higher thymic health had lower cardiovascular-specific mortality relative to those with low thymic health (high versus low thymic health: 0.3% versus 3.9% cardiovascular-specific mortality estimate at 12 years; HR 0.08; 95% CI 0.02–0.34, average versus low thymic health: 1.5% versus 3.9%; HR 0.38; 95% CI 0.20–0.70; Fig. 4c). These associations also were preserved in multivariate models (type III  $P = 0.021$ ) (Fig. 4d). Further, participants with higher thymic health had lower cumulative incidence of cardiovascular-specific diseases relative to participants with low thymic health (high versus low thymic health: 5% versus 16.7% Kaplan–Meier estimate for cardiovascular disease at 12 years; HR 0.26; 95% CI 0.16–0.40, average versus low thymic health: 10.8% versus 16.7%; HR 0.65; 95% CI 0.49–0.87; Fig. 4e). However, these latter associations were partly attenuated in sex and age-stratified



**Fig. 4 | Association of thymic health with long-term risk of CVD-specific mortality, CVD incidence and disease-specific mortalities. a**, Percentage of participants in the NLST who died from CVD. **b**, HR of death from CVD adjusted for pack-years and smoking status, and stratified by sex and age. **c**, Percentage of participants in the FHS who died from CVD. **d**, HR of death from CVD adjusted for smoking status and stratified by sex and age. **e**, Percentage of participants in the FHS who had a new CVD-specific event, such as myocardial infarction, congestive heart failure or cerebral embolism. **f**, HR of a new CVD-specific event adjusted for smoking status and stratified by sex and age. **g**, HR of death from the specified disease groups among participants in the NLST stratified by sex and age and adjusted for pack-years and smoking status. **a–g**, Follow-up for

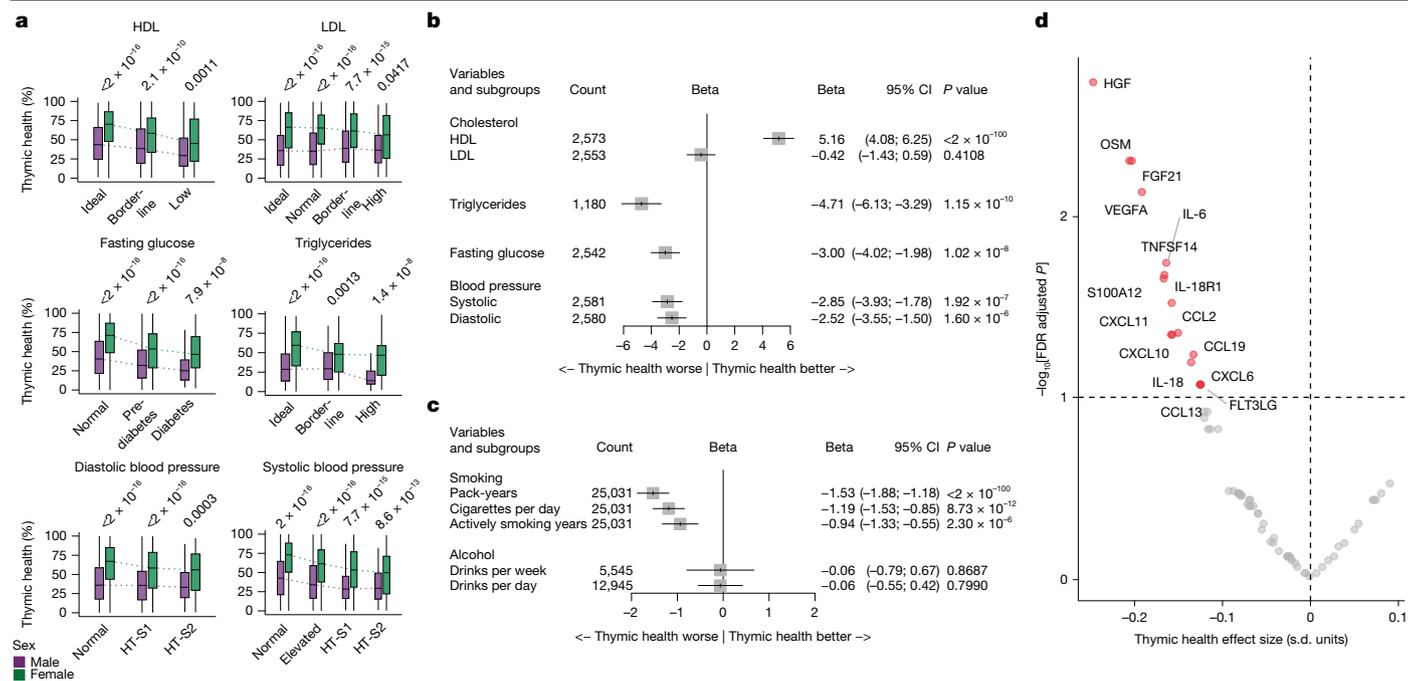
analyses with adjustment for smoking status (type III  $P = 0.051$ ) (Fig. 4f).

### Mortality by disease type

To investigate the relevance of thymic health in different disease types, aside from cancer and cardiovascular disease, we analysed the NLST, where the cause of death was recorded through a 12-year follow-up. Across all investigated disease-specific causes of death, participants with high or average thymic health were less likely to die as compared

all analyses was 12 years. The insets in the inverted Kaplan–Meier plots show the same data on an expanded y axis. Cox proportional hazards regression was used to estimate HRs. In the forest plots, the centre of each box represents the estimated HR, and the whiskers denote the corresponding 95% CI; arrowheads indicate that the 95% CI extends beyond the visualized limits; shaded box size is for visualization only and does not encode statistical weight. The overall contribution of thymic health to uni- or multivariable models was evaluated using likelihood ratio tests ( $\chi^2$  tests) comparing full models with nested models excluding thymic health (type III test, two-sided) without adjustments for multiple comparisons. End, endocrine; Met, metabolic; Nutr, nutritional.

with participants with low thymic health (Extended Data Fig. 6), and all associations remained statistically significant in sex and age-stratified analyses with adjustments for smoking status and pack-years (Fig. 4g and Supplementary Fig. 3f) and in multivariate analyses using continuous age as the time scale to implicitly adjust for the effect of ageing (Extended Data Fig. 5e–g) (type III  $P < 0.05$  for all). Mortality from pulmonary disease for high and average thymic health was 61% and 40% lower, respectively, compared to those with low thymic health (Extended Data Fig. 6c and Supplementary Table 3). Likewise, mortality from endocrine, nutritional and metabolic diseases, including



**Fig. 5 | Associations of thymic health with metabolism, lifestyle and inflammation in the FHS. a**, Associations of metabolically relevant variables with thymic health, stratified by sex with female participants shown in green and male participants in purple. **b**, Respective associations adjusted for sex, age and smoking status. **c**, Associations of smoking-related factors and weekly consumption of alcoholic beverages with thymic health adjusted for sex and age. **d**, Association of Olink-based plasma protein levels with thymic health, adjusted for sex, age and smoking status ( $n = 317$ ). In **a**, box plots show the median (centre line), interquartile range (25th–75th percentiles; box), and whiskers extending to the minimum and maximum values within  $1.5 \times$  the interquartile range. Statistical comparisons between male and female participants were performed using two-sided Wilcoxon rank sum tests. Patient counts shown in **a** correspond to the number (Count) of patients depicted in the respective rows

metabolic disorders such as diabetes mellitus, was 68% and 37% lower for individuals with high and average thymic health, respectively, compared to those with low thymic health (Extended Data Fig. 6d and Supplementary Table 4). Finally, mortality from diseases of the digestive system that included liver, gallbladder or pancreatic diseases was 54% and 47% lower for individuals with high and average thymic health, respectively, compared to those with low thymic health (Extended Data Fig. 6e and Supplementary Table 5). Together, these data provide evidence that thymic health is prognostic across diverse diseases, indicating disease-agnostic relevance for health.

### Impact of metabolic health and lifestyle

Next, we investigated the associations of cholesterol, triglycerides, fasting glucose and blood pressure with thymic health in the FHS (Fig. 5). Across all clinically relevant categories, female participants had higher thymic health as compared with male participants, and thymic health was positively associated with metabolic health (Fig. 5a,b). In sex-, age- and smoking status-adjusted analyses, and across sex strata, we consistently found that high-density lipoprotein (HDL) had a significant positive association with thymic health, while the common variables of metabolic syndrome, including triglyceride levels, fasting glucose and blood pressure, showed negative associations with thymic health (Fig. 5b and Supplementary Fig. 5).

We assessed the association of actionable lifestyle factors, such as smoking and consumption of alcohol, with thymic health. In sex- and age-adjusted analysis, both smoking duration (actively smoking years)

and intensity (packs smoked per day) were negatively associated with thymic health, as also reflected in the strong negative association with the composite measure pack-years (Fig. 5c). The drinking of alcoholic beverages showed no association with thymic health. Furthermore, we used the Fried frailty phenotype<sup>23</sup> to investigate whether individuals with low thymic health are at increased risk for future disabilities and reduced quality of life. We found a significant association between low thymic health and increased frailty scores ( $P < 0.001$ ), independent of sex, age and smoking status (Supplementary Table 6). Exploring the individual components of the Fried frailty index, we found that this signal was particularly driven by slower walking speed ( $P = 0.032$ ), lower physical activity index ( $P = 0.010$ ) and increased exhaustion ( $P = 0.021$ ). Together, these data indicate a strong relationship between metabolic health and lifestyle with thymic health, and that thymic health may directly be associated with an individual's quality of life and long-term risk of disability.

and intensity (packs smoked per day) were negatively associated with thymic health, as also reflected in the strong negative association with the composite measure pack-years (Fig. 5c). The drinking of alcoholic beverages showed no association with thymic health.

Furthermore, we used the Fried frailty phenotype<sup>23</sup> to investigate whether individuals with low thymic health are at increased risk for future disabilities and reduced quality of life. We found a significant association between low thymic health and increased frailty scores ( $P < 0.001$ ), independent of sex, age and smoking status (Supplementary Table 6). Exploring the individual components of the Fried frailty index, we found that this signal was particularly driven by slower walking speed ( $P = 0.032$ ), lower physical activity index ( $P = 0.010$ ) and increased exhaustion ( $P = 0.021$ ). Together, these data indicate a strong relationship between metabolic health and lifestyle with thymic health, and that thymic health may directly be associated with an individual's quality of life and long-term risk of disability.

### Chronic inflammation and thymic health

As immune response and inflammation are closely related, we explored whether low thymic health is associated with dysregulated inflammatory processes. For this, we investigated the association of blood inflammatory proteins with thymic health. First, we analysed proteomic blood plasma data obtained from 317 individuals included in FHS (Supplementary Table 7) before the CT scans (mean time difference, 10.4 years). Plasma samples were analysed using the Olink inflammation panel comprising 92 proteins, with 68 passing quality control.

Of these, 16 out of 68 (24%) proteins were negatively associated with thymic health (Fig. 5d), including major mediators and regulators of inflammation, such as vascular endothelial growth factor A (VEGFA), interleukin-6 (IL-6), IL-18, hepatocyte growth factor (HGF), oncostatin M (OSM) and C-X-C motif chemokine ligand (CXCL) family members 10 and 11. These results show that individuals with higher systemic levels of inflammation had lower thymic health, indicating an interplay between inflammatory processes and thymic health.

To investigate this further, we assessed whether chronic inflammation lasting for a prolonged period of time correlated with thymic health. For this, we collected longitudinal C-reactive protein (CRP) measures over 5 to 10 years for 1,156 participants in the FHS (Supplementary Table 8). We found that participants in the FHS with chronic systemic inflammation, quantified by consistently high CRP levels of greater than or equal to 3 mg l<sup>-1</sup> (140 out of 1,156, 12.1%) over multiple longitudinal blood measurements, had substantially lower thymic health independent of sex, age and smoking status ( $P = 0.0012$ ).

Taken together, these data highlight that inflammation and thymic health are closely related; one potentially driving and potentiating the other, with associated negative health consequences.

### Thymic health model stability

To assess model stability, we performed stability investigations that demonstrated excellent test–retest stability and robustness to inter-reader input variations (Extended Data Fig. 7). Furthermore, to investigate which anatomical regions the model relies on for making its quantifications, we performed activation mapping, which demonstrated specific attention of the model to the thymic bed, while giving minor attention to adjacent structures, indicating knowledge of the anatomical context (Extended Data Fig. 8). Taken together, these results demonstrate that the performance of the thymic health model is robust against input variations, and that it captures contextual information directly from the anatomical region of the thymus. Extensive details regarding model development and testing can be found in Supplementary Methods S1 and S2, Supplementary Figs. 9–17 and Supplementary Tables 9–13.

### Discussion

Our results show that thymic decay in adults is highly individualized and that loss of thymic health increases mortality and disease incidence, including cancer and cardiovascular diseases. These findings were investigated in two independent prospectively collected clinical studies of asymptomatic adults, the NLST and the FHS. Among more than 25,000 participants in the NLST, higher thymic health was consistently associated with a significantly lower risk of mortality by any cause, lung cancer and cardiovascular diseases, independent of sex, age, smoking, prior diseases or cancer history. Similar results were found in the independent FHS, where participants with high thymic health had a significantly reduced risk of death from cardiovascular disease, a primary endpoint of the study, independent of sex, age and smoking. Considering the prevailing perception of the limited role of the thymus in adults, our findings may change our understanding of human health, emphasizing the complex yet critical role of the immune system in long-term well-being and longevity.

We found that thymic health was associated with critical health consequences. Presumed healthy participants in the NLST with high thymic health had an approximately 50% reduction in the risk of death, were 36% less likely to develop lung cancer and nearly 50% less likely to die from lung cancer, as compared to participants with low thymic health. Further, the significant impact of high thymic health on cardiovascular disease (CVD) was consistent across participants in the NLST and FHS, with risk reductions in CVD mortality ranging from 63% to 92%. Given the critical role of the thymus in generating a diverse T cell repertoire<sup>24</sup>,

these results support the broad impact of a sustained adaptive immune system to combat disease and promote longevity.

Our results challenge the established notion that the cessation of thymic output in ageing adults is inconsequential as it is naturally replaced by peripheral expansion of T cells<sup>7</sup>. Rather, our results suggest that loss of thymic tissue in adults may forecast higher risks of disease and death. The main function of the thymus is to generate a diverse T cell repertoire, which provides adaptive immunity throughout life<sup>7</sup>. While the relevance and abundance of the T cell repertoire at a young age are well documented<sup>7</sup>, our results indicate that the thymus retains a continued role in T cell production throughout adulthood and that the pattern of decline of thymic function in adults is associated with poorer health outcomes.

A recent landmark study demonstrated an association between thymectomy and reduced lifespan and increased risk of cancer, among other effects<sup>8</sup>. While thymectomy is rare, we demonstrate that the thymic decay is highly individualized even in presumed healthy adults, indicating that thymic function can also be substantially reduced in individuals who did not have their thymus surgically removed. Our work therefore impacts the wider population and aligns with the well-established age-related decline in immune system function<sup>6,25,26</sup>. This is also consistent with previous work modelling declining T cell output as a major risk factor for age-related increase of cancer development<sup>13</sup>.

Lifestyle and metabolic health measures, such as smoking, physical activity or HDL levels, showed strong associations with thymic health. Likewise, in longitudinal blood-based evaluations, we found individuals with chronic inflammation, a hallmark of immunosenescence and commonly associated with chronic stress, carbohydrate-rich diet and obesity, had lower thymic health.

Among presumed healthy individuals from the FHS, lower thymic health was indeed associated with pro-inflammatory modifications of blood plasma protein levels, consistent with the presence of chronic inflammation. The pro-inflammatory pattern included increased levels of cytokines IL-6, IL-18 and OSM, as well as several CXCL chemokines, all of known relevance in systemic inflammatory diseases such as atherosclerosis, age-associated diseases such as arthritis, and cancer<sup>27</sup>.

Our findings are further supported by studies in which thymic involution was associated with immunosenescence and inflammation, contributing to illnesses such as metabolic or cardiovascular disease<sup>12</sup>. Metabolic syndrome affects more than a third of all adults in the USA, with continuously increasing prevalence<sup>28</sup>. Our results demonstrate significant associations between metabolic and thymic health. These findings are consistent with those of recent studies in which fatty degeneration of the thymus was associated with obesity<sup>18–20</sup> and smoked pack-years<sup>19,20</sup>. Together, these findings suggest a profound impact of actionable lifestyle choices on thymic health and may further clarify why healthy behaviour improves well-being and lifespan.

In addition, the global increase in early-onset cancers might be linked to an accelerated rate of thymic decay, potentially driven by factors such as smoking, minimal physical activity and overall unhealthy lifestyles, leading to more inflammation in the body. Indeed, thymic health could refine disease and cancer screening strategies, especially for high-risk individuals.

Here we examined the impact of thymic health on lung cancer incidence separately in current and former smokers. Associations were primarily driven by current smokers, likely reflecting both higher statistical power due to higher event rates and the greater biological relevance of thymic function under ongoing tobacco exposure and thereby continuously increasing neoantigen load, making thymic health particularly relevant for immune surveillance in this setting<sup>29</sup>.

Previous studies have attempted to quantify thymic characteristics using imaging techniques, using conventional approaches such as estimating the proportion of fatty degeneration or measuring attenuation density<sup>18–20</sup>. However, these studies were limited in scope, focusing

on the exploration of associations between thymic features and basic clinical and epidemiological measures. Previous studies did not find any associations between thymic imaging characteristics and outcomes such as survival or disease incidence. These previous investigations reported presumed residual thymic tissue in only a minor fraction of the examined population ranging from 41%<sup>18</sup> to 26%<sup>19,20</sup>, that is, the studies estimated a fully fatty degenerated thymus in approximately 60–75% of individuals<sup>18</sup>. We find these results contradictory to biological associations of sustained T cell output in visually fully fatty degenerated thymic glands<sup>18</sup>, and we could corroborate that our measure of thymic health had preserved impact on outcomes across individuals who had no appreciable visual thymic tissue according to independent visual scoring, indicating insufficient thymic quantification using a visual scoring system. Our deep learning model found favourable thymic functionality among individuals with average and high thymic health, representing 75% of participants, consistent across both independent study populations and paralleled by substantial health benefits in these participants. Indeed, our findings are supported by Kooshesh et al.<sup>8</sup>, who found substantial negative health consequences after thymectomy in adults, indicating relevant thymic activity throughout life across the population.

Participants included in this study were prospectively enrolled in the FHS and NLST, covering a wide age range for both sexes. They are, however, predominantly white, and further validation in more ethnically diverse populations is warranted. While we observed an association between thymic health and overall survival in both cohorts, this did not remain significant following thorough multivariate adjustment in the smaller FHS. This further indicates that long-term outcomes may be influenced by dynamic lifestyle choices in this population that benefits from regular medical examinations and follow-up, and supports the notion that active engagement in health-promoting behaviours may potentially attenuate the long-term negative impact of unhealthy characteristics associated with low thymic health<sup>30</sup>. Taken together, our results provide evidence of multifaceted associations between thymic health and clinically relevant health consequences across clinical, epidemiological and biological characteristics and strongly suggest the critical relevance of the thymus for adult health.

Before the thymic health deep learning model developed here can be applied in a clinical setting, it is essential to prove generalizability. Development and application of the thymic health model were performed in fully independent datasets, with high robustness as demonstrated by test–retest stability. Given substantial differences in acquisition protocols, scanners and population characteristics between the FHS and the NLST, the thymic health analyses were conducted using population-specific thresholding, and no universal cut-offs can be assumed. Future studies with international and external validation are needed to explicitly address batch and scanner variability and enable direct cross-cohort comparisons and generalizable thresholds.

Our results connect thymic health with longevity and lower disease incidence. Our results suggest that inflammation potentially drives an accelerated rate of thymic decay. This finding may provide new opportunities for preventive strategies that aim to reduce thymic decay and potentially even reverse it. These can include the use of anti-inflammatory and anti-obesity drugs and open new avenues of drug development that aim to improve long-term immune health. Furthermore, even practical approaches, such as lifestyle changes including exercise and sleep, as well as healthy food choices and supplement intake, are likely to notably impact thymic health.

While this study investigated the role of thymic health in healthy individuals, the state of the immune system could also have an important role in individuals with disease. This is particularly relevant for treatments that rely on triggering an immune response, such as immunotherapies used to treat patients with cancer. However, it may also be relevant for other diseases: for example, during the recent COVID-19

pandemic, the response to the virus varied significantly between sex and age groups, with older men particularly affected<sup>31</sup>.

Our analysis reveals previously unrecognized, possible negative consequences of reduced thymic health. While thymic health declines with age, we also find considerable variation in thymic health within age groups, indicating that the rate of decay varies considerably between individuals. When we analysed blood inflammatory proteins, we found an association between increased levels of inflammatory proteins and thymic decay. Although this analysis was limited by a relatively long interval of 10.4 years between the CT scans used for thymic health assessment and blood samples used to assess protein content, these results were supported by an orthogonal assessment of CRP levels, supporting that chronic inflammation may be a likely driver of poor thymic health. It is also likely that a genetic component exists, predisposing certain individuals to increased or decreased rates of thymic decay. A recent large-scale study highlighted substantial individual variability in immune resilience and its link to long-term health outcomes<sup>32</sup>. Gene expression profiles associated with immune competence and low inflammation were linked to longevity, while pro-inflammatory signatures correlated with poorer outcomes. These findings align with our observations that thymic health varies widely across individuals and is negatively associated with chronic inflammation. Awareness of a genetic predisposition may lead to preventive measures and increased surveillance, and this should be elucidated in future studies that match single nucleotide polymorphisms with rates of thymic decay. Other limitations of this study include the older age and heavy smoking status of individuals in the NLST cohort relative to the FHS cohort, both likely to affect thymic health.

An important implication of our study is that the retrospective observational design does not allow conclusions about causality. It is possible that lower thymic health contributes to adverse outcomes by weakening immune resilience, but it is also possible that pathological processes leading to reduced health status, mortality and disease drive thymic decline. Clarifying the direction and nature of these associations will require future prospective and mechanistic studies and will be essential to determine whether thymic health can serve as a target for prevention or intervention strategies. Recent large-scale initiatives, such as the Advanced Research Projects Agency for Health-funded Thymus Rejuvenation programme, underscore growing interest in developing regenerative approaches to restore thymic function, further highlighting the high clinical relevance of this work<sup>33</sup>.

In summary, this study underscores the highly personalized nature of thymic health and emphasizes the previously unrecognized possible critical role of maintaining thymic health to preserve an agile, adaptive immune response that will accommodate long-term well-being and longevity. Today, thymic assessments do not have an established clinical standard, and the thymus is not examined in routine clinical care. The extent to which the adult appearance of the thymus is associated with health and whether actionable lifestyle or risk factors may be harnessed to improve thymic health was unknown. By analysing 27,612 individuals, our results provide evidence that thymic health is directly associated with critical outcomes and diseases and may be directly targetable by various approaches, such as smoking reduction and weight loss in overweight and obese individuals.

Our results underline the relevance of the thymus throughout life. Clinical investigations of preventive or regenerative strategies will be vital to help us understand how to harness the thymic potential to improve population health<sup>34–38</sup>.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions

and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-026-10242-y>.

1. Goldrath, A. W. & Bevan, M. J. Selecting and maintaining a diverse T-cell repertoire. *Nature* **402**, 6–13 (1999).
2. Miller, J. F. A. P. The function of the thymus and its impact on modern medicine. *Science* **369**, eaba2429 (2020).
3. Palmer, D. B. The effect of age on thymic function. *Front. Immunol.* **4**, 316 (2013).
4. Collins, C., Sharpe, E., Silber, A., Kulke, S. & Hsieh, E. W. Y. Congenital athymia: genetic etiologies, clinical manifestations, diagnosis, and treatment. *J. Clin. Immunol.* **41**, 881–895 (2021).
5. Rezzani, R., Nardo, L., Favero, G., Peroni, M. & Rodella, L. F. Thymus and aging: morphological, radiological, and functional overview. *Age* **36**, 313–351 (2014).
6. Lynch, H. E. et al. Thymic involution and immune reconstitution. *Trends Immunol.* **30**, 366–373 (2009).
7. Kumar, B. V., Connors, T. J. & Farber, D. L. Human T cell development, localization, and function throughout life. *Immunity* **48**, 202–213 (2018).
8. Kooshesh, K. A., Foy, B. H., Sykes, D. B., Gustafsson, K. & Scadden, D. T. Health consequences of thymus removal in adults. *N. Engl. J. Med.* **389**, 406–417 (2023).
9. Murray, J. M. et al. Naive T cells are maintained by thymic output in early ages but by proliferation without phenotypic change after age twenty. *Immunol. Cell Biol.* **81**, 487–495 (2003).
10. Williams, K. M., Hakim, F. T. & Gress, R. E. T cell immune reconstitution following lymphodepletion. *Semin. Immunol.* **19**, 318–330 (2007).
11. Weng, N.-P. Aging of the immune system: how much can the adaptive immune system adapt? *Immunity* **24**, 495–499 (2006).
12. Thomas, R., Wang, W. & Su, D.-M. Contributions of age-related thymic involution to immunosenescence and inflammaging. *Immun. Ageing* **17**, 2 (2020).
13. Palmer, S., Albergante, L., Blackburn, C. C. & Newman, T. J. Thymic involution and rising disease incidence with age. *Proc. Natl Acad. Sci. USA* **115**, 1883–1888 (2018).
14. Vadasz, Z., Haj, T., Kessel, A. & Toubi, E. Age-related autoimmunity. *BMC Med.* **11**, 94 (2013).
15. Nunes-Alves, C., Nobrega, C., Behar, S. M. & Correia-Neves, M. Tolerance has its limits: how the thymus copes with infection. *Trends Immunol.* **34**, 502–510 (2013).
16. Douek, D. C. et al. Changes in thymic function with age and during the treatment of HIV infection. *Nature* **396**, 690–695 (1998).
17. Wang, W., Thomas, R., Sizova, O. & Su, D.-M. Thymic function associated with cancer development, relapse, and antitumor immunity—a mini-review. *Front. Immunol.* **11**, 773 (2020).
18. Sandstedt, M. et al. Complete fatty degeneration of thymus associates with male sex, obesity and loss of circulating naive CD8+ T cells in a Swedish middle-aged population. *Immun. Ageing* **20**, 45 (2023).
19. Araki, T. et al. Normal thymus in adults: appearance on CT and associations with age, sex, BMI and smoking. *Eur. Radiol.* **26**, 15–24 (2016).
20. Ackman, J. B. et al. Sex difference in normal thymic appearance in adults 20–30 years of age. *Radiology* **268**, 245–253 (2013).
21. Pido-Lopez, J., Imami, N. & Aspinall, R. Both age and gender affect thymic output: more recent thymic migrants in females than males as they age: age and gender affect thymic output. *Clin. Exp. Immunol.* **125**, 409–413 (2001).
22. Heng, T. S. P. et al. Effects of castration on thymocyte development in two different models of thymic involution. *J. Immunol.* **175**, 2982–2993 (2005).
23. Fried, L. P. et al. Frailty in older adults: evidence for a phenotype. *J. Gerontol. A* **56**, M146–M157 (2001).
24. Klein, L., Kyewski, B., Allen, P. M. & Hogquist, K. A. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nat. Rev. Immunol.* **14**, 377–391 (2014).
25. Britanova, O. V. et al. Age-related decrease in TCR repertoire diversity measured with deep and normalized sequence profiling. *J. Immunol.* **192**, 2689–2698 (2014).
26. Mittelbrunn, M. & Kroemer, G. Hallmarks of T cell aging. *Nat. Immunol.* **22**, 687–698 (2021).
27. West, N. R., Owens, B. M. J. & Hegazy, A. N. The oncostatin M-stromal cell axis in health and disease. *Scand. J. Immunol.* **88**, e12694 (2018).
28. Moore, J. X., Chaudhary, N. & Akinymijiu, T. Metabolic syndrome prevalence by race/ethnicity and sex in the United States, National Health and Nutrition Examination Survey, 1988–2012. *Prev. Chronic Dis.* **14**, E24 (2017).
29. Alexandrov, L. B. et al. Mutational signatures associated with tobacco smoking in human cancer. *Science* **354**, 618–622 (2016).
30. Allaire, S. H. et al. Evidence for decline in disability and improved health among persons aged 55 to 70 years: the Framingham Heart Study. *Am. J. Public Health* **89**, 1678–1683 (1999).
31. Cummings, M. J. et al. Epidemiology, clinical course, and outcomes of critically ill adults with COVID-19 in New York City: a prospective cohort study. *Lancet* **395**, 1763–1770 (2020).
32. Ahuja, S. K. et al. Immune resilience despite inflammatory stress promotes longevity and favorable health outcomes including resistance to infection. *Nat. Commun.* **14**, 3286 (2023).
33. Jenkins, A. ARPA-H funds project to restore immune system function. ARPA-H <https://arpa-h.gov/news-and-events/arpa-h-funds-project-restore-immune-system-function> (28 September 2023).
34. Duggal, N. A., Pollock, R. D., Lazarus, N. R., Harridge, S. & Lord, J. M. Major features of immunosenescence, including reduced thymic output, are ameliorated by high levels of physical activity in adulthood. *Aging Cell* **17**, e12750 (2018).
35. Duah, M. et al. Thymus degeneration and regeneration. *Front. Immunol.* **12**, 706244 (2021).
36. Chaudhry, M. S., Velardi, E., Dudakov, J. A. & van den Brink, M. R. M. Thymus: the next (re) generation. *Immunol. Rev.* **271**, 56 (2016).
37. Fahy, G. M. et al. Reversal of epigenetic aging and immunosenescent trends in humans. *Aging Cell* **18**, e13028 (2019).
38. Buono, M., Thézénas, M.-L., Ceroni, A., Fischer, R. & Nerlov, C. Bi-directional signaling by membrane-bound KitL induces proliferation and coordinates thymic endothelial cell and thymocyte expansion. *Nat. Commun.* **9**, 4685 (2018).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026

## Methods

### Study cohorts

We used two independent prospectively enrolled cohort studies ( $n = 27,612$ ), consisting of the FHS ( $n = 2,581$ ) and the NLST ( $n = 25,031$ ), to conduct our retrospective secondary analysis (Supplementary Fig. 6).

The FHS is a longitudinal community-based prospective cohort study. It started enrolling the original participants in 1948 and has since been followed by consecutive enrolments of the Offspring cohort (children of the original participants and their spouses) in the 1970s and the Third-Generation cohort (children of the Offspring participants) from 2002 to 2005. Eligibility for enrolment in the FHS CT study required an age of 35 years or older for male participants and 40 years or older for female participants. All participants had regular follow-up examination cycles every 4–8 years, and they provided written informed consent for the CT study and each attended examination. For the current study, we identified 2,581 participants from the Offspring and Third-Generation cohorts who had non-contrast-enhanced, non-gated, full-thoracic CT scans covering the entire thymic bed between 2005 and 2011. Participants are under ongoing surveillance for cardiovascular disease endpoints and death. An endpoint review committee of senior investigators using all available data (for example, hospital records) adjudicated the FHS study endpoints using standardized criteria<sup>39</sup>. For subanalyses, Offspring participants who attended Exam 7 (1998–2001), 8 (2005–2008) or 9 (2011–2014) for Olink inflammation proteomics panel, longitudinal CRP and Fried frailty analyses were identified. The measurement of thymic health occurred an average of 10.4 and 3.75 years after Exams 7 and 8, respectively, and 2 years before Exam 9, in our study population.

The NLST is a longitudinal randomized trial of screening for lung cancer with the use of low-dose chest CT as compared with chest radiography. It enrolled participants from 2002 to 2004 with screening examinations from 2002 to 2007. Primary event follow-up was carried out until 31 December 2009, and extended follow-up for overall survival and disease-specific survival was registered until 2015. Eligibility for enrolment required an age between 55 and 74 at randomization and a smoking history of a minimum of 30 pack-years without quitting smoking for more than 15 years before enrolment. Each participant provided written informed consent. For the current study, we identified 25,031 participants from the first low-dose CT screening exam (T0), which was performed soon after the time of randomization in Institutional Review Board-approved centres of the Lung Screening Study or in centres of the American College of Radiology Imaging Network, which were responsible for collecting participant data.

Our retrospective secondary analysis of the FHS and NLST was reviewed and approved by their corresponding review boards. All participants provided written informed consent at study enrolment. Each participant provided written informed consent during each examination attended as part of the FHS, a procedure evaluated and approved by the Institutional Review Board at Boston University Medical Center.

### Artificial intelligence-based thymic health assessment

We developed a deep learning system that automatically extracts a thymic health score ranging from zero (complete thymic decay) to one (high thymic health) from a given CT scan that covers the thoracic region. A detailed description of the model development, including training data and architecture, can be found in Supplementary Methods S1. In short, first, the system identifies the location of the thymic bed to generate a centre-of-mass (CoM). Next, a deep learning model performs a thymic health assessment using a second model that leverages self-supervised learning in which information from unlabelled imaging data was used to quantify CT characteristics of the thymic area. Model development was done in a collection of 5,674 thoracic CT scans (independent of FHS and NLST), covering a wide range of acquisition settings of patients that were imaged for various diseases,

including cancer and infectious diseases. The system was fine-tuned against expert image annotations of thymic involution to obtain an objective measure of remaining thymic soft tissue, hypothesized to reflect thymic functionality, on the basis of an existing protocol<sup>18–20</sup>. FHS and NLST were defined as test sets and remained unseen until the fully independent and externally developed model was locked. A technical evaluation of the performance of the end-to-end deep learning system can be found in Supplementary Methods S2. An evaluation of thymic health against the previously proposed manual protocol of thymus scoring<sup>18–20</sup> is depicted in Supplementary Fig. 1, and an evaluation against thymic bed volume is shown in Supplementary Figs. 7 and 8.

### Assessment of inter-reader stability

To evaluate the stability of the end-to-end system, a sub-sample of 5,081 CT scans was selected, consisting of 2,500 random scans from the NLST and 2,581 random scans from the FHS. The CoM coordinates were perturbed by shifting the CoM input in all three directional axes, adding a random amount of simulated inter-reader noise sampled from a Gaussian distribution with a mean of 0 and a variance of 16 mm. The results were bootstrapped for 50 iterations, and the concordance index was calculated between predicted thymic scores for a pair of trials and averaged over all trials to assess the robustness of the system's performance against input perturbations. Further details for this assessment can be found in Supplementary Methods S2.2.3.

### Analysis of test–retest variability

To assess the test–retest variability of the system, the Reference Image Database to Evaluate Therapy Response dataset was used. This dataset consists of 31 pairs of CT scans acquired approximately 15 min apart from patients with non-small cell lung cancer. The intraclass correlation coefficient and Cohen's Kappa were computed for both the automatic quality control and thymic health model, grouped by predicted scores and categories, respectively. The intraclass correlation coefficient was used to evaluate the consistency of the automatic quality control scores, while Cohen's Kappa was used to measure the agreement between the thymic health categories assigned by the model across the test and retest scans. Furthermore, the  $R^2$  coefficient was calculated to compare the model's predictions across the test and retest scans, attesting to the model's reliability and reproducibility in the context of short-term variability inherent in the dataset.

### Quality control of the artificial intelligence system

To investigate the associations between thymic health scores and the CT-based input, that is, the thymic bed area, we used Shapley value analysis and saliency maps. Shapley value analysis is a game-theoretic approach that assigns importance scores to each input feature, indicating its contribution to the model's output. This method allowed us to identify the most attributable features for a predicted thymic health score, providing insights into the model's decision-making process. After determining the most attributable features using Shapley value analysis, we selected the feature with the largest contribution to the output prediction. Next, we determined regions in the input volume that affected this feature substantially. We used occlusion sensitivity, where sliding windows of 7 mm<sup>3</sup> were occluded across the volume, and the difference in feature values was observed. Intuitively, windows that change the feature value the most are the most salient regions. These visualizations aid in understanding the artificial intelligence system's measures, demonstrating a visual representation of the imaging regions that contributed the most to the model's predictions.

### Statistical analysis

We scaled the raw output of the thymic health classification model, after supervised learning, that is, the scalar probability corresponding to the likelihood of the thymus being not 'fully-fatty degenerated', to match the percentile distribution ranks of the examined study populations

to facilitate thymic health interpretation and potential clinical translation. By doing so, thymic health values range from 0 to 100 and can be easily interpreted, such as a value of 50 ranking a participant exactly in the median of the study population distribution. Further, using the first and third quartiles in each respective study population, thymic health was ranked into low (less than or equal to 25), average (25–75) and high (greater than 75).

Summary statistics used mean, median or range for continuous variables and proportions or percentages for categorical variables. Overall survival was calculated from the date of the CT examination (FHS) or from the date of randomization (NLST) until the date of death. For time-to-event analyses, we used a cutoff at 12 years of follow-up for both FHS and NLST for all survival and incidence analyses except for lung cancer incidence, for which data were only available for a follow-up of 6 years after randomization. If participants were alive or did not have an event after 12 or 6 years of follow-up, respectively, they were censored at that time. Association analyses used Wilcoxon rank sum test or linear regression as appropriate. For linear regression analyses, clinical variables were z-score normalized and treated as predictors for thymic health. The beta coefficients represent the change in thymic health per one s.d. increase in each predictor. Time-to-event distributions were analysed with the Kaplan–Meier estimator and Cox proportional hazards models as appropriate for univariate and multivariate testing. The Schoenfeld residuals were assessed for inspection of the proportional hazards assumption. Initial Cox proportional hazards models used thymic health as predictor, with adjustments for sex and age, which violated the proportional hazards assumptions. We addressed this by implementing sex- and age-stratified analyses in which age was categorized into distinct bins using the age distribution of the studied population (FHS, younger than 45 years, 45–48 years, 48–51 years, 51–54 years, 54–57 years, 57–60 years, 60–63 years, 63–66 years, 66–69 years, 69–72 years, 72–75 years and 75 years or older; NLST, 55–59 years, 60–64 years, 65–69 years and 70–75 years) and this stratification approach was used in all sex- and age-adjusted analysis throughout the article if not described otherwise. Both the Akaike Information Criterion and Bayesian Information Criterion were inspected, and the model fit and parsimony were improved using the implemented stratified models.

Unless explicitly stated otherwise, the following apply: box plots show the median (centre line), interquartile range (25th–75th percentiles; box), and whiskers extending to the minimum and maximum values within 1.5× the interquartile range. Statistical comparisons between groups were performed using two-sided Wilcoxon rank sum tests with no adjustment for multiple comparisons. *P* values are interpreted against a significance level that is defined as less than 0.05. Cox proportional hazards regressions were used to estimate HRs, and the corresponding 95% CI are provided for all statistical values. In the forest plots, the centre of each box represents the estimated HR, and the whiskers denote the corresponding 95% CI; arrowheads indicate that the 95% CI extends beyond the visualized limits; shaded box size is for visualization only and does not encode statistical weight. The overall contribution of thymic health to uni- or multivariable models was evaluated using likelihood ratio tests ( $\chi^2$  tests) comparing full models with nested models excluding thymic health (type III test, two-sided). Statistical significance of individual covariate coefficients was assessed using two-sided Wald z-tests with no adjustments for multiple comparisons.

There were no missing values in all time-to-event analyses if not otherwise described. In the case of missing values in the association analyses, values were assumed to be missing completely at random, and the analyses handled missing data by their row-wise exclusion. No statistical methods were used to predetermine sample size. Statistical analyses were done in R v.4.2.2 (R Project for Statistical Computing).

## Inclusion and ethics statement

This retrospective secondary analysis of NLST data was reviewed and approved by their corresponding review boards and approved under

NLST-367 and NLST-374. FHS analysis was reviewed and approved by the Boston University Medical Center International Review Board, protocol number H-32132.

## Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

NLST data may be requested from the National Cancer Institute (<https://biometry.nci.nih.gov/cdas/nlst/>), imaging data from the Imaging Data Commons portal (<https://portal.imaging.datacommons.cancer.gov/>) and the Thymic Health scores for NLST are available at Zenodo (<https://doi.org/10.5281/zenodo.18306999>)<sup>40</sup>. Data from the FHS are available from the online repositories BioLINCC and dbGap, or from the submission of a research proposal via FHS ResApp (<https://www.framinghamheartstudy.org/fhs-for-researchers/research-application/>). An overview of public imaging data collections used for the development of the deep learning model, which were downloaded from the Imaging Data Commons (<https://portal.imaging.datacommons.cancer.gov/>), can be found in Supplementary Table 9.

## Code availability

The software used in the publication is available on GitHub for academic, non-commercial use in our GitHub ([https://github.com/AIM-Harvard/thymus\\_health\\_deeplearning\\_system.git](https://github.com/AIM-Harvard/thymus_health_deeplearning_system.git)). Additional technical details about both the development and evaluation of our deep learning framework can also be found in the Supplementary Information. The models' weights are subject to intellectual property obligations and cannot be shared publicly, but may be made available through academic collaboration. For more details, please contact the corresponding author.

39. Vasan, R. S., Enserro, D. M., Xanthakis, V., Beiser, A. S. & Seshadri, S. Temporal trends in the remaining lifetime risk of cardiovascular disease among middle-aged adults across 6 decades: the Framingham Study. *Circulation* **145**, 1324–1338 (2022).
40. Prudente, V. C. G. et al. Resources “Thymic health consequences in adults”. *Zenodo* <https://doi.org/10.5281/zenodo.18306999> (2026).

**Acknowledgements** S.B., V.P. and S.P. contributed equally as leading authors to this work. K.L.L., N.J.B. and H.J.W.L.A. acted jointly as a supervisory team. We thank the National Cancer Institute for collecting and making the data from the NLST accessible, and The Cancer Imaging Archive and the Imaging Data Commons for making this and other imaging collections used for developing our deep learning model available on their platforms. H.J.W.L.A. acknowledges financial support from NIH (HA: NIH-USA U24CA194354, NIH-USA U01CA190234, NIH-USA U01CA209414 and NIH-USA R35CA22052; BHK: NIH-USA K08DE030216-01) and the European Union–European Research Council (HA: 866504). K.L.L., J.M.M., Y.C. and J.C. received financial support from NIH-USA R01AG067457). S.B. acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—502050303. K.B. acknowledges funding from Bayern Innovativ, German Federal Ministry of Research, Technology and Space, Max Kade Foundation and Wilhelm-Sander Foundation. N.J.B. acknowledges funding from the Lundbeck Foundation (R272-2017-4040), the Novo Nordisk Foundation (NNF21OC0071483 and NNF23OC0085954) and Savvaerksejer Jeppe Juhl og Hustru Ovita Juhl Research Stipend. M.J.-H. has received funding from CRUK, NIH National Cancer Institute, IASLC International Lung Cancer Foundation, Lung Cancer Research Foundation, Rosetrees Trust, UKI NETs and NIHR. The FHS is funded by a contract from the National, Heart, Lung, and Blood Institute (75N92019D0031, contract number 75N92019D0031).

**Author contributions** Conceptualization: S.B., V.P., S.P., H.J.W.L.A. and N.J.B. Methodology: S.B., V.P., S.P., A.K.A., Y.C., J.C., L.N., M.J.-H., C.A., C.S., K.L.L., J.M.M., H.J.W.L.A. and N.J.B. Software: V.P. and S.P. Validation: S.B., V.P., S.P., Y.C., J.C., K.B., K.L.L., A.L., H.J.W.L.A. and N.J.B. Formal analysis: S.B., V.P., S.P., Y.C., J.C., K.L.L. and A.L. Resources: H.J.W.L.A., K.L.L. and J.M.M. Data curation: S.B., V.P., S.P., Y.C., J.C., K.L.L., J.M.M., B.F., M.T.L. and A.L. Writing—original draft: S.B., V.P., S.P., H.J.W.L.A. and N.J.B. Writing—review and editing: all authors. Visualization: S.B., V.P., S.P., A.K.A., L.N., H.J.W.L.A. and N.J.B. Supervision: H.J.W.L.A., N.J.B., K.L.L. and J.M.M. Project administration: H.J.W.L.A. and N.J.B. Funding acquisition: S.B., H.J.W.L.A., K.L.L. and J.M.M.

**Competing interests** M.J.-H. has consulted for Astex Pharmaceutical and Achilles Therapeutics, is a member of the Achilles Therapeutics Scientific Advisory Board and Steering Committee, and has received speaker honoraria from Pfizer, Astex Pharmaceuticals, Oslo Cancer Cluster, Bristol Myers Squibb and Genentech. M.J.-H. is listed as a co-inventor on a

# Article

European patent application relating to methods to detect lung cancer (PCT/US2017/028013), this patent has been licensed to commercial entities and, under terms of employment, M.J.-H. is due a share of any revenue generated from such license(s), and is also listed as a co-inventor on the GB priority patent application (GB2400424.4) with title *Treatment and Prevention of Lung Cancer*. C.S. reports receiving support from the Francis Crick Institute and the Royal Society; has received grants or contracts from AstraZeneca, Boehringer Ingelheim, Bristol Myers Squibb, Invitae (formerly Archer Dx), Ono Pharmaceuticals, Pfizer and Roche-Ventana; has received consulting fees from Bicycle Therapeutics, Genentech, Medicxi, Metabomed, Novartis, and is a member of the GRAIL SAB, Relay Therapeutics SAB, China Innovation Center of Roche (CiCoR), SAGA Diagnostics SAB and the Sarah Cannon Research Institute; has received honoraria from Amgen, AstraZeneca, Bristol Myers Squibb, Illumina, GlaxoSmithKline, MSD, Roche-Ventana and Pfizer; holds patents, planned, issued, or pending, including PCT/GB2017/053289, PCT/EP2016/059401, PCT/EP2016/071471, PCT/GB2018/052004, PCT/GB2020/050221, PCT/GB2018/051912, PCT/US2017/28013 and PCT/GB2018/051892; has leadership or fiduciary roles with Cancer Research UK and AACR; holds stock or stock options in Apogen Biotech, Epic Biosciences, GRAIL and Achilles Therapeutics; and has other financial or non-financial interests with AstraZeneca and GRAIL Bio UK. M.T.L. reports grants research funding

from the American Heart Association, AstraZeneca, Ionis, Johnson & Johnson Innovation, Kowa Pharmaceuticals America, MedImmune, National Academy of Medicine, the National Heart, Lung, and Blood Institute and the Risk Management Foundation of the Harvard Medical Institutions outside the submitted work. S.B. reports consulting fees from Ambient. H.J.W.L.A. reports consulting fees and/or stock from Onc.AI, Love Health, Sphera, Health-AI, Ambient, and AstraZeneca. N.J.B. reports consulting fees from Ambient, is listed as a co-inventor on a patent application (PCT/GB2020/050221) on methods for cancer prognostication and a patent on methods for predicting anti-cancer response (US14/466.208). The other authors declare no competing interests.

## Additional information

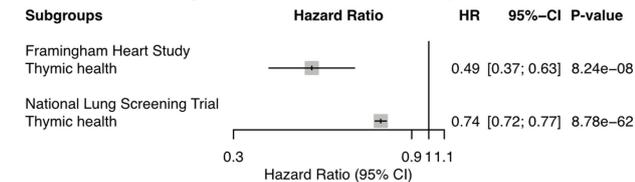
**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41586-026-10242-y>.

**Correspondence and requests for materials** should be addressed to Hugo J. W. L. Aerts.

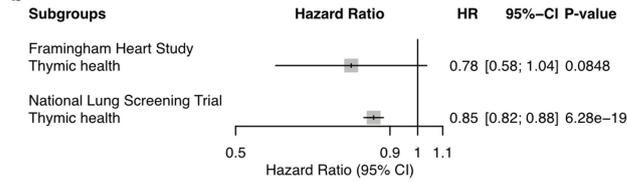
**Peer review information** *Nature* thanks Bharat Thyagarajan and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

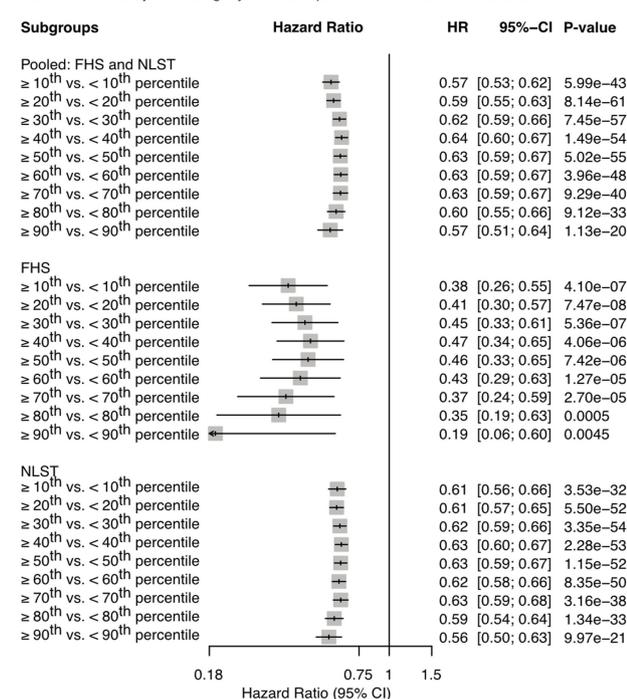
**a** Overall survival according to a one SD increase of continuous thymic health



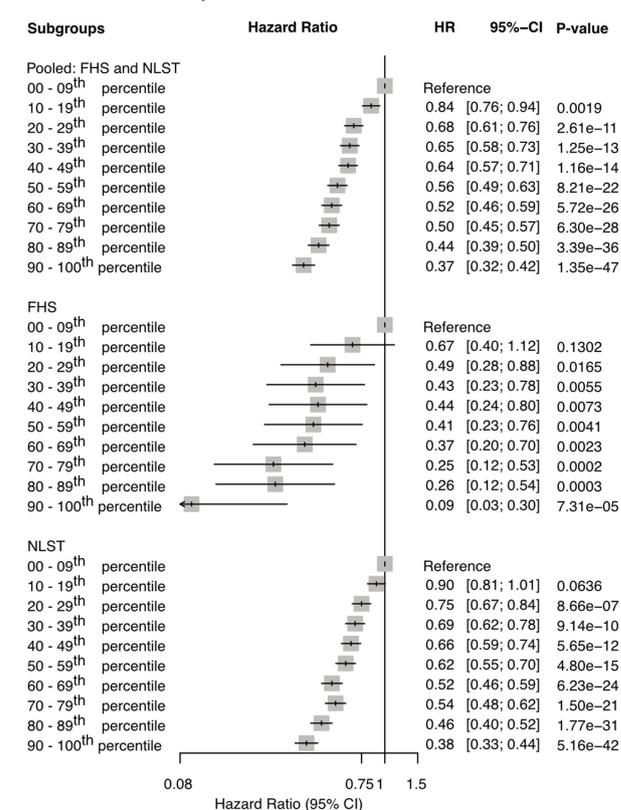
**b** Overall survival according to a one SD increase of continuous thymic health, adjusted



**c** Overall survival by increasing thymic health percentile binarization thresholds

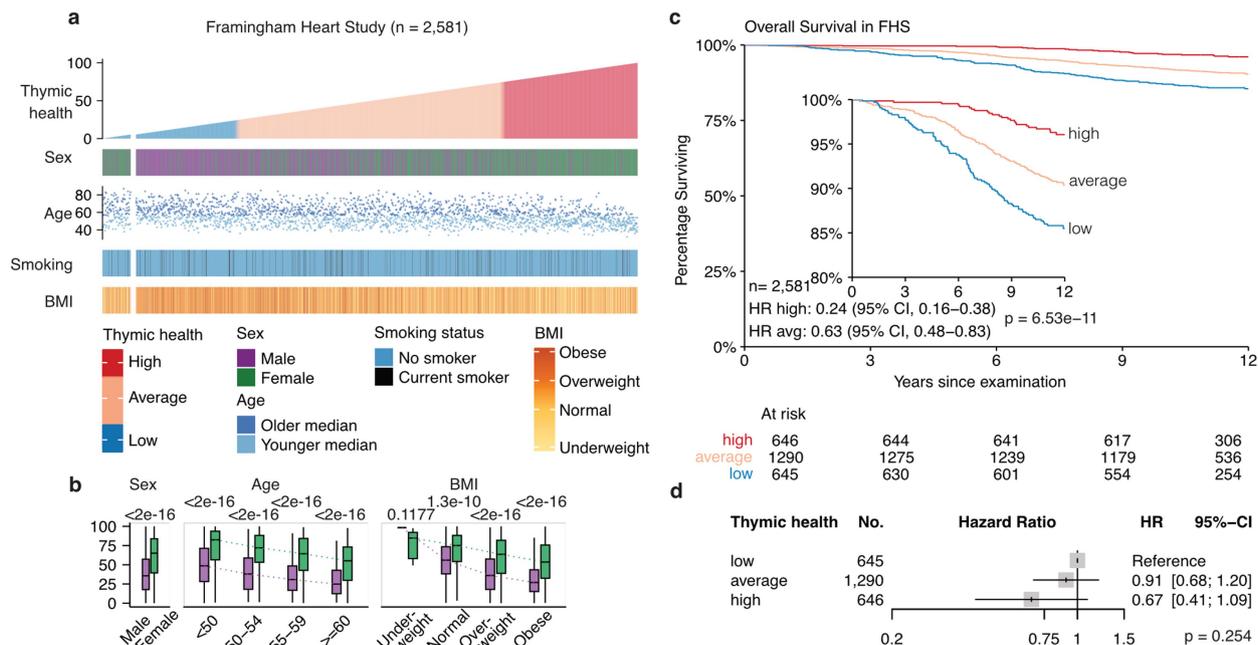


**d** Overall survival in each thymic health decile versus the lowest decile as reference



**Extended Data Fig. 1 | Long-term risk of mortality by increasing continuous thymic health, by increasing thymic health percentiles cut-off thresholds, and in each thymic health decile. a**, Risk of death according to a one standard deviation increase of the continuous thymic health measure, and **b**, with stratification by sex and age. **c**, Risk of death by increasing thymic health percentile thresholds in the pooled (FHS and NLST) (upper panel) and independent FHS (middle panel) and NLST (lower panel) cohorts. **d**, Risk of death in each thymic health decile versus the lowest decile, as reference, in the pooled (upper panel) and independent FHS (middle panel) and NLST (lower panel) cohorts. **a-d**, FHS,  $n = 2,581$ ; NLST,  $n = 25,031$ . Cox proportional hazards

regression was used to estimate HRs. In the forest plots, the center of each box represents the estimated hazard ratio, and the whiskers denote the corresponding 95% CI; arrowheads indicate that the 95% CI extends beyond the visualized limits; shaded box size is for visualization only and does not encode statistical weight. Statistical significance of the continuous or binarized thymic health covariate coefficients at the different cut points was assessed using two-sided Wald z-tests without adjustments for multiple comparisons. CI Confidence Interval, FHS Framingham Heart Study, HR Hazard Ratio, NLST National Lung Screening Trial, SD standard deviation.

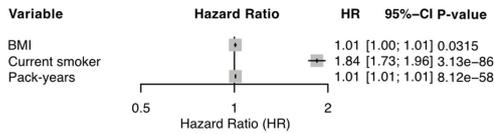


**Extended Data Fig. 2 | Association of thymic health with long-term mortality in the Framingham Heart Study.** **a**, Overview of the Framingham Heart Study (FHS). Each column represents one patient. The data is sorted by ascending thymic health, i.e., an imaging-based proxy for thymic functionality, from left to right. Thymic health is categorized into low, average, and high based on the bottom 25%, middle 50%, and top 25% of the population. The fractional left split of the FHS represents the patients who were defined as having low thymic health through automatic quality control. Thymic health quantification and basic clinical information are shown. **b**, Associations between thymic health and sex in FHS ( $n = 2,581$ ), across age groups in years ( $n = 2,581$ ) and body mass categories ( $n = 2,577$ , missingness  $n = 4$ ) using pairwise Wilcoxon Rank Sum test. Box plots show the median (center line), interquartile range (25th–75th percentiles; box), and whiskers extending to the minimum and maximum values within  $1.5 \times$  the interquartile range. Statistical comparisons between groups were performed using two-sided

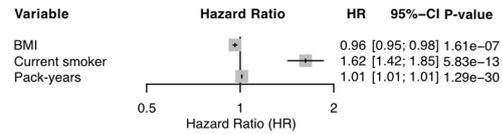
Wilcoxon rank-sum tests without adjustment for multiple comparisons. **c**, Survival of participants from the FHS without adjustments. **d**, Hazard ratio (HR) of death from any cause for FHS participants adjusted for smoking status and stratified by sex and age binned at 3-years. Follow-up for all analyses was 12 years. The inset in the Kaplan-Meier plot shows the same data on an expanded y-axis. **c,d**, Cox proportional hazards regression was used to estimate HRs. In the forest plots, the center of each box represents the estimated hazard ratio, and the whiskers denote the corresponding 95% CI; arrowheads indicate that the 95% CI extends beyond the visualized limits; shaded box size is for visualization only and does not encode statistical weight. The overall contribution of thymic health to uni- or multivariable models was evaluated using likelihood ratio tests ( $\chi^2$  tests) comparing full models with nested models excluding thymic health (type III test, two-sided) with no adjustments for multiple comparisons. BMI Body Mass Index, CI Confidence Interval, FHS Framingham Heart Study, HR Hazard Ratio.

Clinical variables

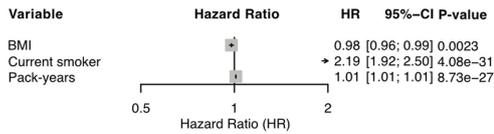
**a** Overall survival



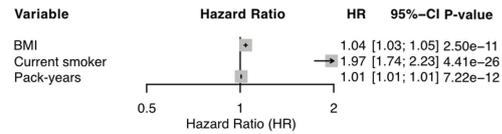
**b** Lung cancer incidence



**c** Lung cancer mortality

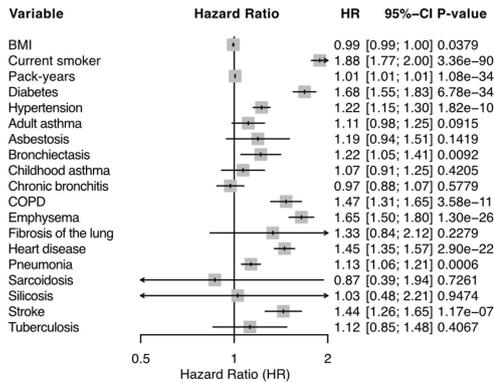


**d** CVD mortality

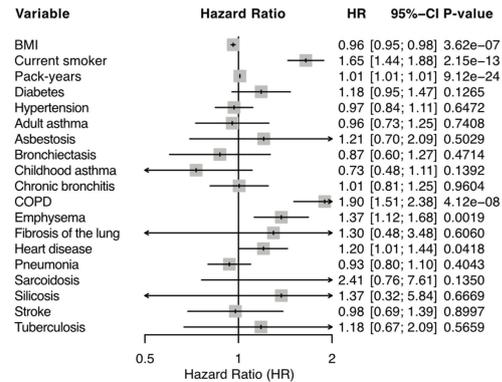


Clinical variables and prior diseases

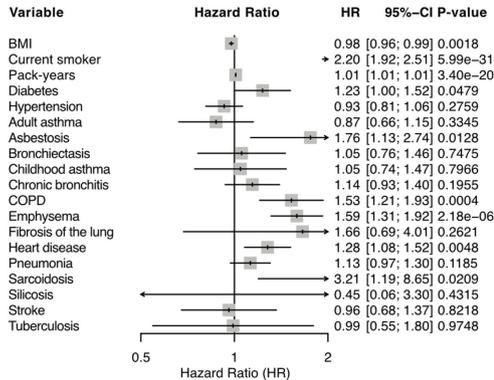
**e** Overall survival



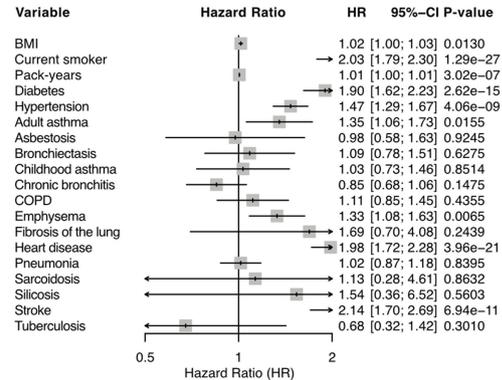
**f** Lung cancer incidence



**g** Lung cancer mortality



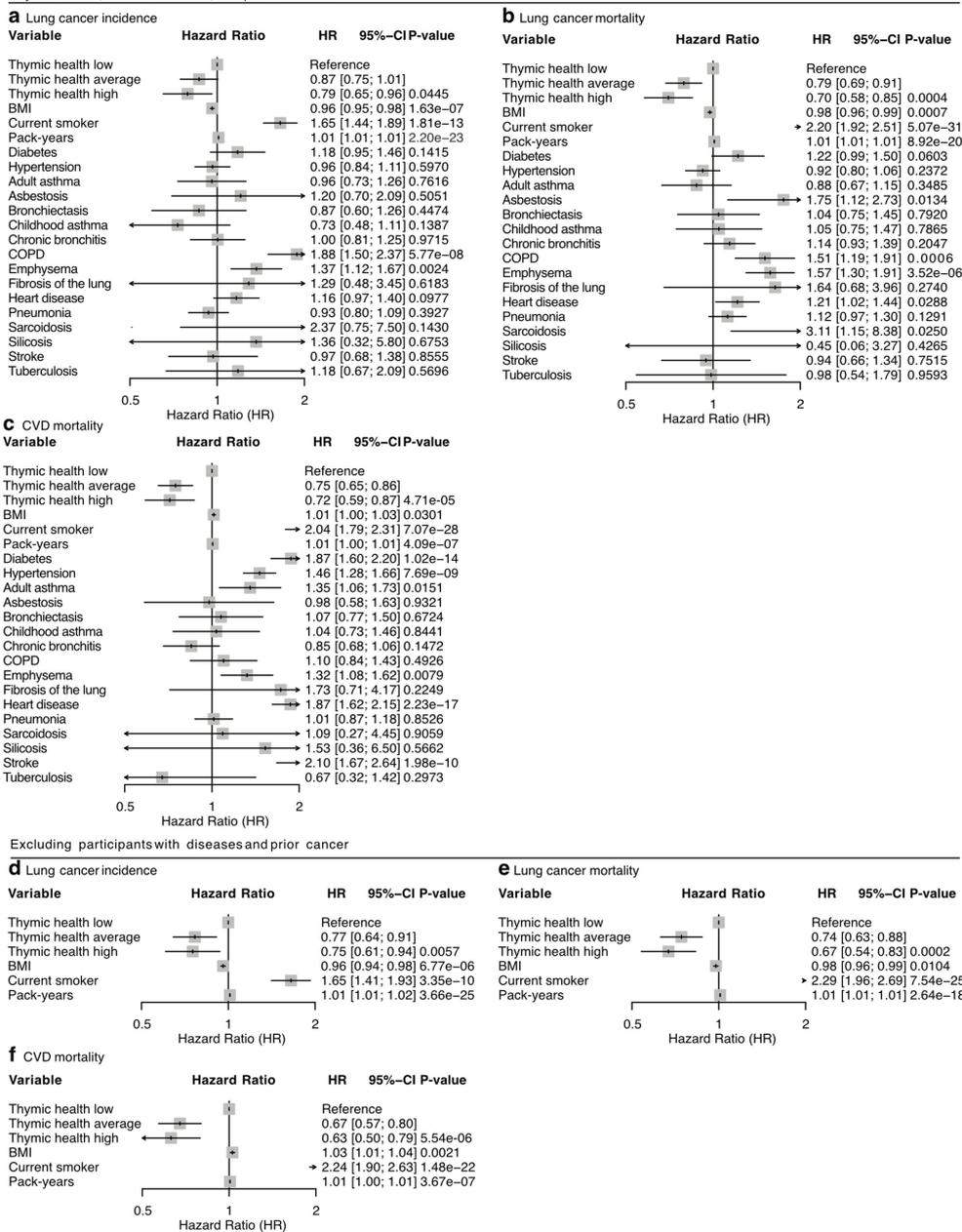
**h** CVD mortality



**Extended Data Fig. 3 | Association of clinical variables and known diseases prior to trial enrollment with health outcomes in NLST. a–d,** Association analyses between the baseline clinical variables body mass index (BMI), smoking status, and pack-years with **a**, overall survival, **b**, lung cancer incidence, **c**, lung cancer-specific mortality, and **d**, cardiovascular disease (CVD)-specific mortality in NLST, stratified by sex and age. 83 participants were excluded due to missing BMI values, resulting in 24,948 individuals. Exact cohort sizes after outcome-specific missingness were: overall survival ( $n = 24,948$ ; 83 missing), lung cancer incidence ( $n = 23,087$ ; 1,944 missing), and lung cancer mortality and CVD mortality (both  $n = 24,946$ ; 85 missing). **e–h,** Association analyses between the baseline clinical variables BMI, smoking status, and pack-years together with diagnosed diseases prior to trial enrollment with **e**, overall survival, **f**, lung cancer incidence, **g**, lung cancer-specific mortality, and **h**, CVD-specific mortality in NLST, stratified by sex and age. 434 participants

were excluded due to missing values in BMI and diagnosed diseases, resulting in 24,597 individuals. Exact cohort sizes after outcome-specific covariate missingness were: overall survival ( $n = 24,597$ ; 434 missing), lung cancer incidence ( $n = 22,776$ ; 2,255 missing), and lung cancer mortality and CVD mortality (both  $n = 24,595$ ; 436 missing). **a–h,** Cox proportional hazards regression was used to estimate HRs. In the forest plots, the center of each box represents the estimated hazard ratio, and the whiskers denote the corresponding 95% CI; arrowheads indicate that the 95% CI extends beyond the visualized limits; shaded box size is for visualization only and does not encode statistical weight. Statistical significance of individual covariate coefficients was assessed using two-sided Wald z-tests with no adjustments for multiple comparisons. CI Confidence Interval, HR Hazard Ratio, NLST National Lung Screening Trial.

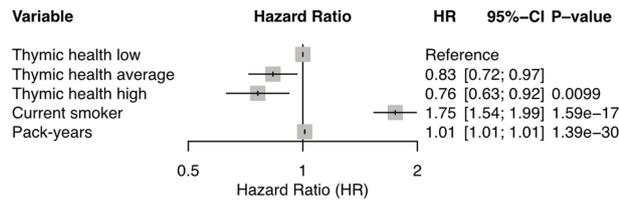
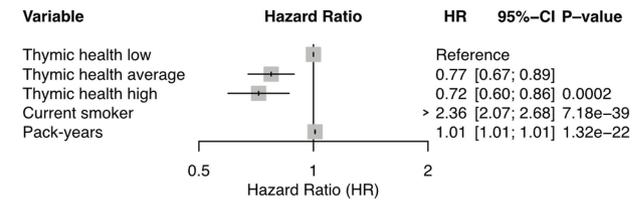
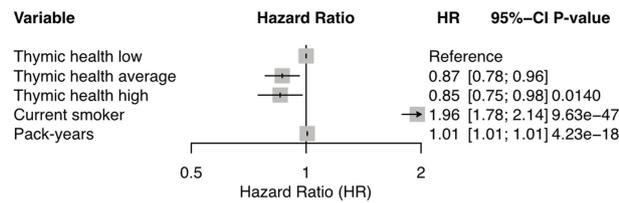
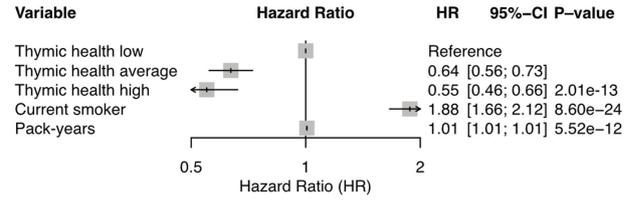
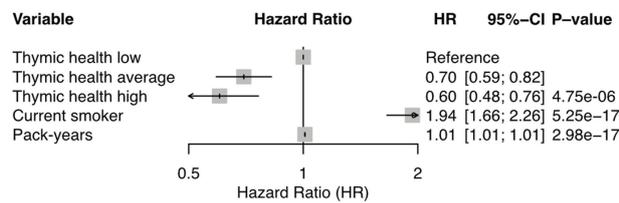
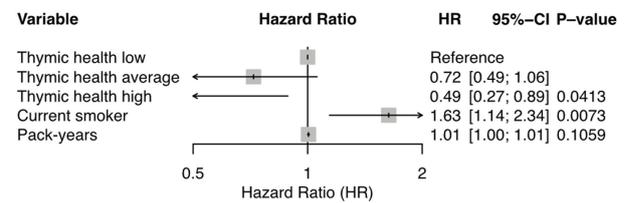
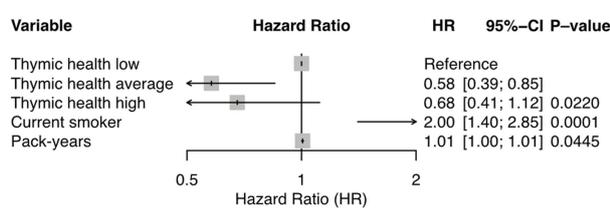
Thymic health, clinical variables, and prior diseases



**Extended Data Fig. 4 | Association of thymic health, clinical variables, and known diseases prior to trial enrollment with health outcomes in NLST.**

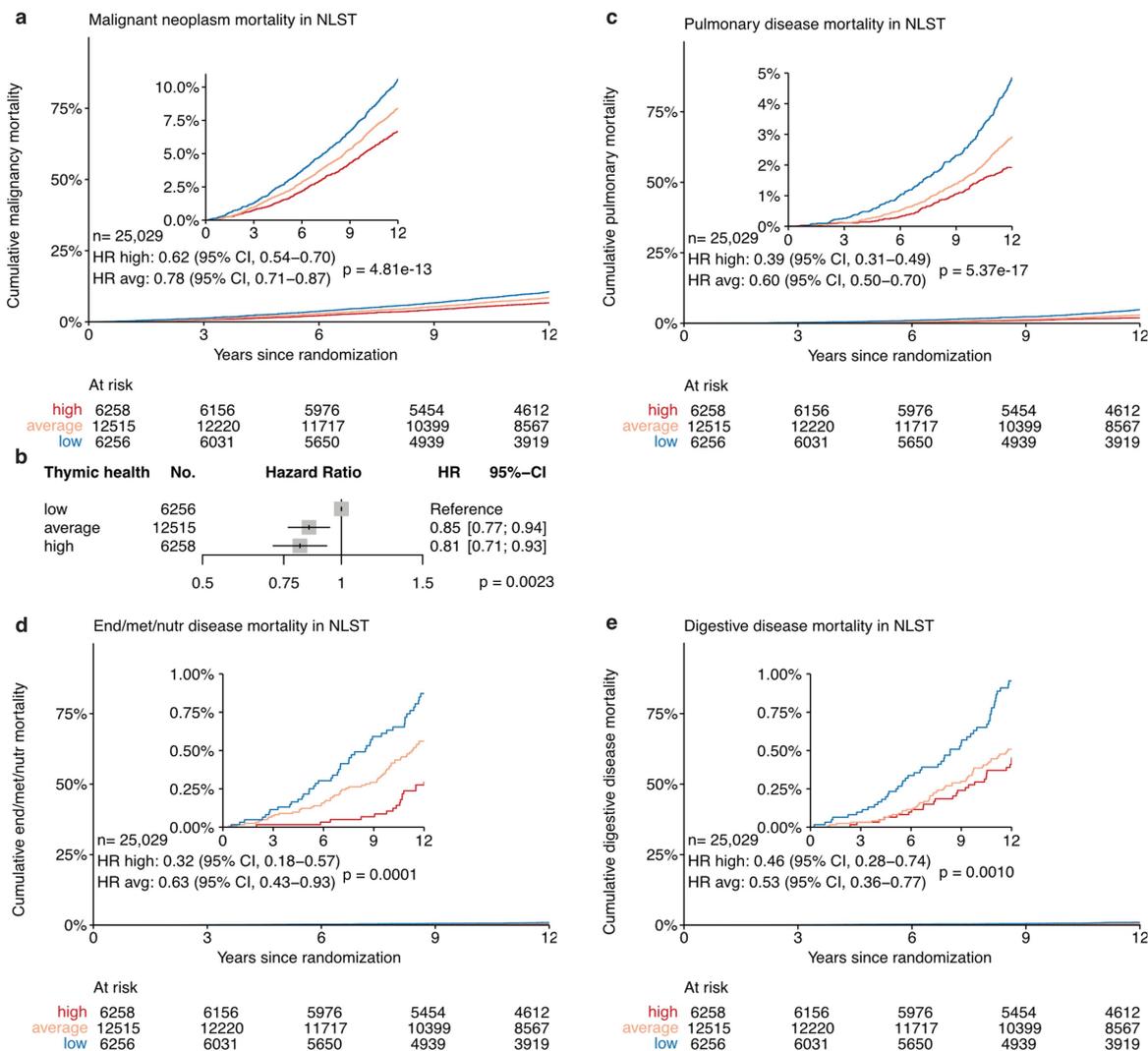
**a-c**, Association analyses between thymic health, the baseline clinical variables body mass index (BMI), smoking status, and pack-years together with diagnosed diseases prior to trial enrollment with **a**, lung cancer incidence, **b**, lung cancer-specific mortality, and **c**, cardiovascular disease (CVD)-specific mortality in NLST, stratified by sex and age. 434 participants were excluded due to missing values in BMI and diagnosed diseases, resulting in 24,597 individuals. Exact cohort sizes after outcome-specific missingness were: overall survival ( $n = 24,597$ ; 434 missing), lung cancer incidence ( $n = 22,776$ ; 2,255 missing), and lung cancer and CVD mortality ( $n = 24,595$ ; 436 missing). **d-f**, Association analyses between thymic health and the baseline clinical variables BMI, smoking status, and pack-years, stratified by sex and age excluding participants with known diseases or cancer prior to trial enrollment with **d**, lung cancer incidence, **e**, lung cancer-specific mortality, and **f**, CVD-specific mortality in the subgroup of the NLST in which participants with a history of cancer

( $n = 575$ ), childhood or adult asthma, diabetes, asbestosis, bronchiectasis, lung fibrosis, sarcoidosis, silicosis, or tuberculosis ( $n = 5,500$ ) were excluded from the analysis ( $n = 18,619$ ). Exact cohort sizes after outcome-specific missingness were: overall survival ( $n = 18,565$ ; 54 missing), lung cancer incidence ( $n = 17,280$ ; 1,339 missing), and lung cancer and CVD mortality ( $n = 18,564$ ; 55 missing). Cox proportional hazards regression was used to estimate HRs. In the forest plots, the center of each box represents the estimated hazard ratio, and the whiskers denote the corresponding 95% CI; arrowheads indicate that the 95% CI extends beyond the visualized limits; shaded box size is for visualization only and does not encode statistical weight. The overall contribution of thymic health to multivariable models was evaluated using likelihood ratio tests ( $\chi^2$  tests) comparing full models with nested models excluding thymic health (type III test, two-sided). Statistical significance of individual covariate coefficients was assessed using two-sided Wald z-tests with no adjustments for multiple comparisons. CI Confidence Interval, HR Hazard Ratio, NLST National Lung Screening Trial.

**a Lung cancer incidence****b Lung cancer mortality****c Malignant neoplasm mortality****d CVD mortality****e Pulmonary disease mortality****f End/Met/Nutr disease mortality****g Digestive disease mortality**

**Extended Data Fig. 5 | Association of thymic health with health outcomes in NLST using continuous age as the time scale.** Analyses were adjusted for smoking status and continuous pack-years and stratified by sex. Modeling age as the underlying time scale accounts for potential residual confounding by age. Hazard ratios of **a**, lung cancer incidence ( $n = 23,163$ ), **b**, lung cancer mortality ( $n = 25,027$ ), **c**, malignant neoplasm mortality ( $n = 25,025$ ), **d**, cardiovascular-disease specific mortality ( $n = 25,025$ ), **e**, pulmonary-disease specific mortality ( $n = 25,025$ ), **f**, endocrine, nutritional, or metabolic-disease specific mortality ( $n = 25,025$ ), and **g**, digestive-disease specific mortality ( $n = 25,025$ ). Follow-up for all analyses was 12 years. Cox proportional hazards regression was used to estimate HRs. In the forest plots, the center of each box represents the

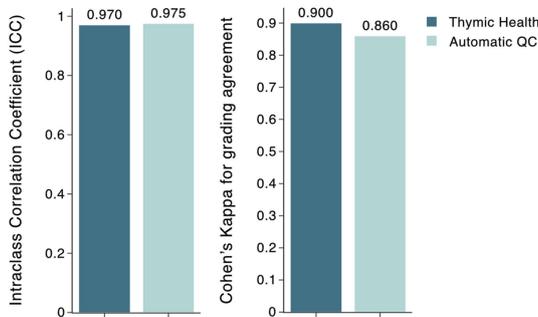
estimated hazard ratio, and the whiskers denote the corresponding 95% CI; arrowheads indicate that the 95% CI extends beyond the visualized limits; shaded box size is for visualization only and does not encode statistical weight. The overall contribution of thymic health to multivariable models was evaluated using likelihood ratio tests ( $\chi^2$  tests) comparing full models with nested models excluding thymic health (type III test, two-sided). Statistical significance of individual covariate coefficients was assessed using two-sided Wald z-tests with no adjustments for multiple comparisons. CI Confidence Interval, End/Met/Nutr endocrine, nutritional, and metabolic diseases, HR Hazard Ratio, NLST National Lung Screening Trial.



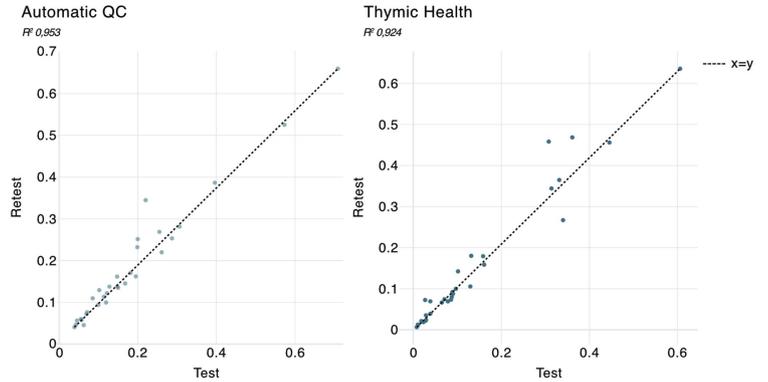
**Extended Data Fig. 6 | Impact of thymic health on long-term risk of disease-specific mortality and clinically relevant cardiovascular disease.** Percentage of individuals who died from **a**, any malignancy, **b**, adjusted for sex, age, and smoking status. Percentage of individuals who died from **c**, pulmonary disease, **d**, endocrine, metabolic, or nutritional disease; or **e**, digestive disease. Adjusted analyses of **c, d, e**, see Fig. 4g. Follow-up for all analyses was 12 years. The insets in the inverted Kaplan-Meier plots show the same data on an expanded y-axis. **a–e**, Cox proportional hazards regression was used to estimate HRs. In the forest plot, the center of each box represents the estimated hazard ratio, and

the whiskers denote the corresponding 95% CI; arrowheads indicate that the 95% CI extends beyond the visualized limits; shaded box size is for visualization only and does not encode statistical weight. The overall contribution of thymic health to uni- or multivariable models was evaluated using likelihood ratio tests ( $\chi^2$  tests) comparing full models with nested models excluding thymic health (type III test, two-sided) with no adjustments for multiple comparisons. CI Confidence Interval, FHS Framingham Heart Study, HR Hazard Ratio, NLST National Lung Screening Trial.

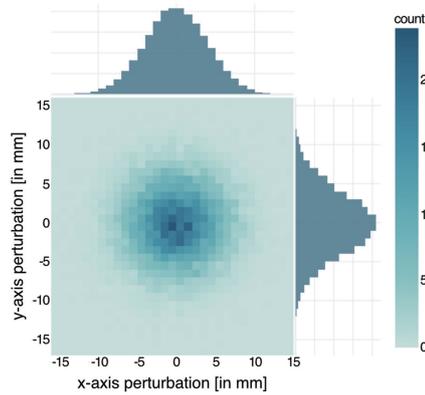
**a** Test-retest stability of continuous model predictions and three predicted thymus grades using ICC and Cohen's Kappa respectively



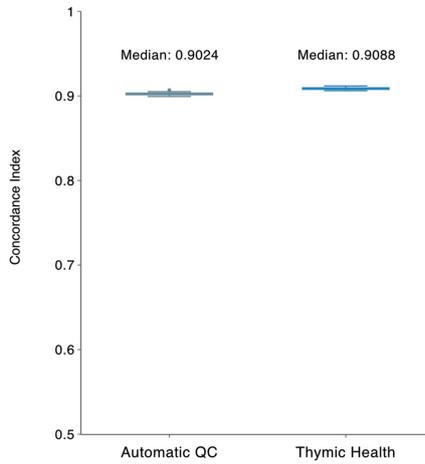
**b** Scatter plot of test-retest model predictions for the 0vs123 and 7vsrest prediction models



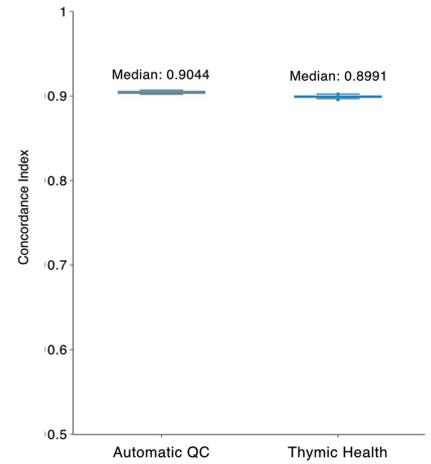
**c** Sampling distribution for testing stability to input perturbation



**d** Stability to input perturbation for automatic QC and thymic health predictors for FHS cohort.

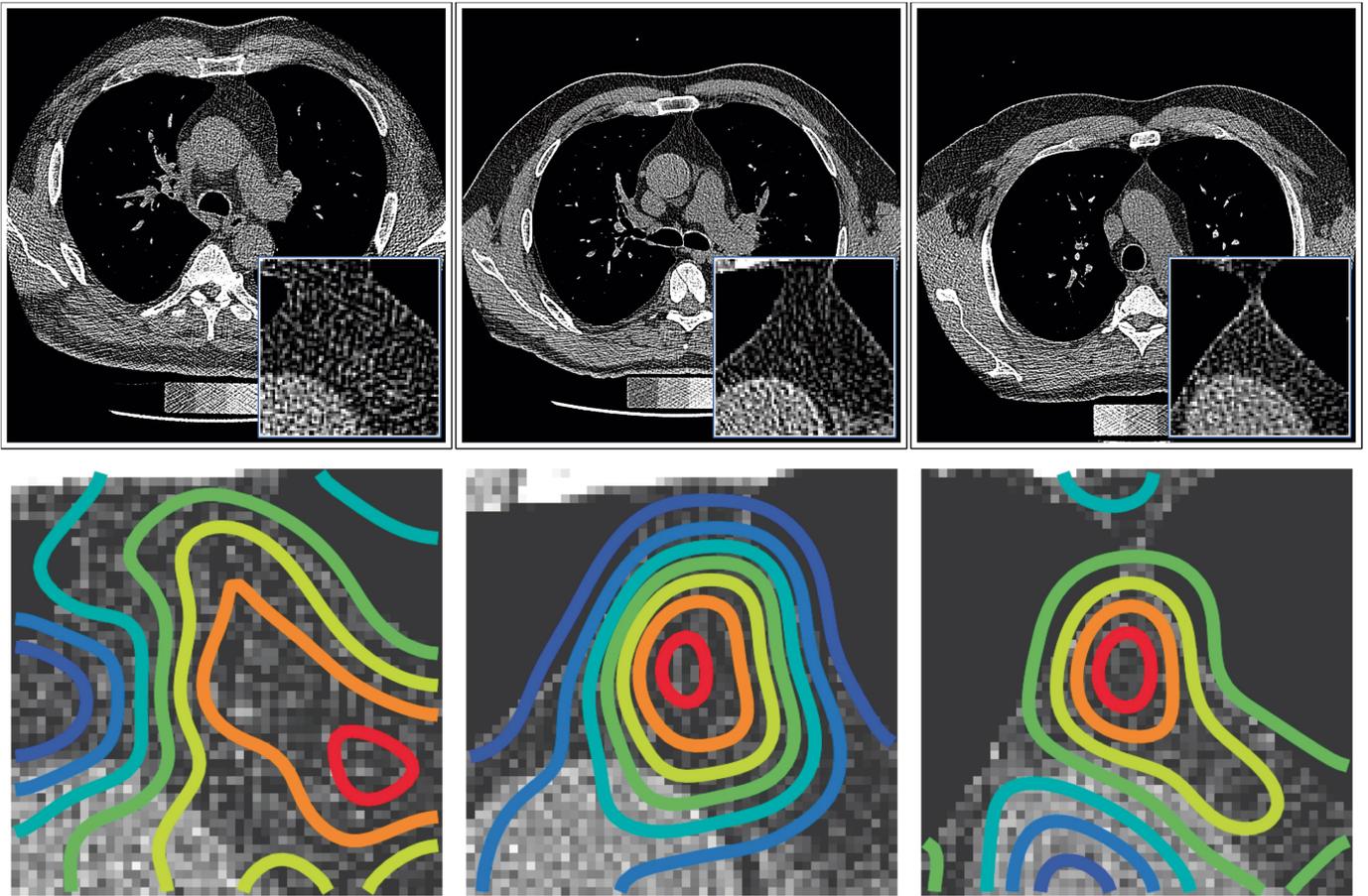


**e** Stability to input perturbation for automatic QC and thymic health predictors for NLST cohort.



**Extended Data Fig. 7 | Stability assessment through test-retest and input variability analysis.** **a**, Test-retest stability of the two prediction models (QC model and 0 vs 1 grading) is analyzed first through intraclass correlation coefficient for continuous measures ( $n = 31$ ), followed by Cohen's kappa for grade stability. **b**, Test-retest predictions ( $n = 31$ ) for both models are displayed as a scatter plot to show their linearity. **c**, Sampling distribution for generating input perturbations for the 50 perturbation trials, where for each scan (patient), the seed point is perturbed with a drawn sample. Although a 2D distribution is

shown, the perturbations are performed in 3 dimensions. Stability to input variation/perturbation is shown through the concordance between predicted scores across different trials for the **d**, FHS ( $n = 2,581$ ), and **e**, NLST ( $n = 2,500$ ) cohorts. The box plots show the median (center line), interquartile range (25th–75th percentiles; box), and whiskers extending to the minimum and maximum values within 1.5 x the interquartile range. ICC Intra-class Correlation Coefficient, FHS Framingham Heart Study, NLST National Lung Screening Trial, QC Quality Control.



**Extended Data Fig. 8 | Explainability: Saliency Maps.** Representative CT scans from the FHS cohort overlaid with the generated occlusion-driven saliency maps using the most important feature, highlighting the regions of the thymus

that the model focuses on when making its predictions. The saliency maps are displayed on the central slice of the thymus, and the jet color scale is used to indicate contribution intensity.

**Extended Data Table 1 | Study population clinical and epidemiological characteristics**

Characteristics	Pooled (% , range)	FHS (% , range)	NLST (% , range)
Number of participants ( <i>No. (%)</i> )	27,612 (100.0%)	2,581 (100.0%)	25,031 (100%)
Sex ( <i>No. (%)</i> )			
<i>Male</i>	16,084 (58.3%)	1,241 (48.1%)	14,843 (59.3%)
<i>Female</i>	11,528 (41.7%)	1,340 (51.9%)	10,188 (40.7%)
Age (median [range])	60 [32, 87]	56 [32, 87]	60 [55, 74]
Race ( <i>No. (%)</i> )		*	
<i>White</i>			22,855 (91.3%)
<i>Not white</i>			2,082 (8.3%)
BMI (median [range])	27.3 [13.3, 65]	27.7 [16.3, 54.7]	27.3 [13.3, 65]
Smoking status ( <i>No. (%)</i> )			
<i>No current smoker</i>	15,431 (55.9%)	2,415 (93.6%)	13,016 (52%)
<i>Current smoker</i>	12,181 (44.1%)	166 (6.4%)	12,015 (48%)

\* Race data is not available for participants of the Framingham Heart Study (FHS). The FHS is a longitudinal community-based prospective cohort study. It started enrolling the original participants in 1948 and has since been followed by consecutive enrollments of the Offspring cohort (children of the original participants and their spouses) in the 1970s and the Third-Generation (Gen 3) cohort (children of the Offspring participants) from 2002-2005. All original participants are of White/European ancestry. Few exceptions might be spouses or adoptees. The degree of missing data for Race and Smoking status is the difference between the summed values and the total number of participants in each cohort. The degree of missingness of BMI was  $n=4$  for FHS and  $n=83$  for NLST. Percentages refer to the data after the exclusion of missing values. BMI Body mass index, FHS Framingham Heart Study, NLST National Lung Screening Trial.

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | n/a                                 | Confirmed  |
|-------------------------------------|--|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of all covariates tested   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated  |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

```
survminer_0.4.9
ggpubr_0.6.0
ggplot2_3.4.4
survival_3.5-7
dplyr_1.1.4

Python 3.8. Libraries:
brotlipy==0.7.0
certifi==2023.5.7
cffi==1.15.0
charset-normalizer==2.0.4
cryptography==39.0.1
fire==0.4.0
idna==3.4
imageio==2.28.1
importlib-resources==5.12.0
lazy_loader==0.2
loguru==0.7.0
mkl-fft==1.3.1
mkl-random==1.2.2
mkl-service==2.4.0
monai==1.1.0
networkx==3.1
nibabel==5.1.0
nptyping==2.5.0
numpy==1.24.3
packaging==23.1
pandas==2.0.1
Pillow==9.4.0
pip==23.0.1
psutil==5.9.5
pycparser==2.21
pynrrd==1.0.0
pyOpenSSL==23.0.0
PySocks==1.7.1
python-dateutil==2.8.2
pytorch-ignite==0.4.12
pytz==2023.3
PyWavelets==1.4.1
PyYAML==6.0
requests==2.29.0
scikit-image==0.20.0
scipy==1.9.1
setuptools==66.0.0
SimpleITK==2.2.1
six==1.16.0
termcolor==2.3.0
tifffile==2023.4.12
torch==1.12.0
torchaudio==0.12.0
torchvision==0.13.0
tqdm==4.65.0
typing_extensions==4.5.0
tzdata==2023.3
urllib3==1.26.15
wheel==0.38.4
zipp==3.15.0
foundation-cancer-image-biomarker==0.0.1a22
imblearn==0.0
ipykernel==6.29.4
optuna-dashboard==0.15.1
pip-chill==1.0.3
wandb==0.17.0
```

The software used in the publication is available on GitHub for academic, non-commercial use in our GitHub ([https://github.com/AIM-Harvard/thymus\\_health\\_deeplearning\\_system.git](https://github.com/AIM-Harvard/thymus_health_deeplearning_system.git)). We provide package management through Python requirements to ensure exact versioning of packages. Additionally, the Supplementary Information was expanded to also include a list of the used packages (section 1.7.).

All statistical code and code to reproduce the figures will be available in the paper's respective Zenodo repository (<https://doi.org/10.5281/zenodo.18306999>).

Code is provided on a public GitHub repository ([https://github.com/AIM-Harvard/thymus\\_health\\_deeplearning\\_system.git](https://github.com/AIM-Harvard/thymus_health_deeplearning_system.git)).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

NLST data, including raw CT images, may be requested from the National Cancer Institute (<https://biometry.nci.nih.gov/cdas/nlst/>). The Framingham Heart Study makes data available through the online repositories BioLINCC and dbGap, or through the submission of a research proposal via FHS ResApp. (<https://www.framinghamheartstudy.org/fhs-for-researchers/research-application/>). Public imaging data collections used for the development of the deep learning model, and overview of these datasets can be found in Supplementary Information Table 9. The data availability statement was updated to reflect these details.

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Sex was used as biological attribute and considered in the study design. Analyses were adjusted or stratified by sex whenever appropriate. We provide sex data of the study population in Extended Data Table 1.

Reporting on race, ethnicity, or other socially relevant groupings

We reported race data of the study population in Extended Data Table 1. We clearly state in our discussion section the following: Participants included in this study were prospectively enrolled in the FHS and NLST, covering a wide age range for both sexes. They are, however, predominantly caucasian, and further validation in more diverse ethnic populations is warranted.

Population characteristics

We describe the covariate-relevant population characteristics of the human research participants in Extended Data Table 1. Here, the most relevant covariates were sex, age, and smoking status.

Recruitment

We performed retrospective secondary analysis of the FHS and NLST.

The FHS is a longitudinal community-based prospective cohort study. It started enrolling the original participants in 1948 and has since been followed by consecutive enrollments of the Offspring cohort (children of the original participants and their spouses) in the 1970s and the Third-Generation (Gen 3) cohort (children of the Offspring participants) from 2002-2005. Eligibility for enrollment in the FHS CT study required an age of  $\geq 35$  years for males and  $\geq 40$  years for female participants. All participants had regular follow-up examination cycles every 4 to 8 years, and they provided written informed consent for the CT study and each attended examination. For the current study, we identified 2,581 participants from the Offspring and Third-Generation cohorts who had non-contrast-enhanced, non-gated, full thoracic CT scans covering the entire thymic bed between 2005 - 2011. Participants are under ongoing surveillance for cardiovascular disease endpoints and death. An endpoint review committee of senior investigators using all available data (e.g. hospital records) adjudicated the FHS study endpoints using standardized criteria. For subanalyses, Offspring participants who attended Exam 7 - 9 for Olink inflammation proteomics panel, longitudinal CRP, and Fried frailty analyses were identified.

The NLST is a longitudinal randomized trial of screening for lung cancer with the use of low-dose chest CT as compared with chest radiography. It enrolled participants from 2002 through 2004 with screening examinations from 2002 through 2007. Primary event follow-up was done through December 31, 2009, and extended follow-up for overall survival and disease-specific survival was registered through 2015. Eligibility for enrollment required an age between 55 and 74 at randomization and a smoking history of a minimum of 30 pack-years without quitting smoking for more than 15 years before enrollment. Each participant provided written informed consent. For the current study, we identified 25,031 participants from the first low-dose CT screening exam (T0), which was performed soon after the time of randomization in Institutional Review Board-approved centers of the Lung Screening Study (LSS) or in centers of the American College of Radiology Imaging Network (ACRIN) which were responsible for collecting participant data.

Ethics oversight

Our retrospective secondary analysis of the FHS and NLST was reviewed and approved by their corresponding review boards. All participants provided written informed consent at study enrollment. Each participant provided written informed consent during each examination attended as part of the Framingham Heart Study (FHS), a procedure evaluated and approved by the Institutional Review Board (IRB) at Boston University Medical Center (BUMC).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was chosen based on all available data from the FHS and NLST. The cohort comprised a total of 27,612 participants (2,581 in FHS; 25,031 in NLST). In our study we performed a secondary analysis of the two observational cohort studies (FHS, NLST). Sample size was not determined by calculation as this was not a prospective study.
Data exclusions	We provide flow charts of study population inclusion in Supplementary Data Figure S5. All exclusion criteria were pre-established before data analysis was started. All patients that were available from FHS and NLST were eligible for study inclusion. As shown in Supplementary Data Figure S5 individuals who failed quality control metrics including corrupted CT scans, slice thickness of 3.2 mm or higher, missing slides, or missing clinical data were excluded as this was established as quality requirements.
Replication	We performed a secondary use analysis of two observational studies. As such, the results shown here are not the result of an experimental setup. We share the code to reproduce our analyses and also demonstrate through stability assessments of the deep learning model using test-retest and input variability analysis that our investigation is repeatable and reproducible.
Randomization	This is a secondary analysis of two observational studies using preexisting data. Randomization was therefore not relevant.
Blinding	Blinding was not relevant as this was a secondary analysis of two observational studies. In NLST patients were initially randomized in a low dose CT arm and in an X-Ray arm. For our study, we could only leverage the CT arm, as we required CT data for the thymic health assessment. Therefore, in our secondary analysis, patients were not allocated to any intervention. No biomarker result was reported back to patients, so the individuals could not change their behavior based on our findings. The deep learning model development and application was performed without knowing the outcome status of the patients, which represents a form of blinding.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration	The original clinical registration numbers are: ClinicalTrials.gov: FHS, NCT00005121; NLST, NCT00047385.
Study protocol	No study protocol was required for our secondary use analysis. This was not a clinical trial but a retrospective study using pre-existing clinical datasets. The study protocols are described in the manuscript and can be found online for the public dataset.
Data collection	The FHS is a longitudinal community-based prospective cohort study. It started enrolling the original participants in 1948 and has since been followed by consecutive enrollments of the Offspring cohort (children of the original participants and their spouses) in the 1970s and the Third-Generation (Gen 3) cohort (children of the Offspring participants) from 2002-2005. Eligibility for enrollment in the FHS CT study required an age of $\geq 35$ years for males and $\geq 40$ years for female participants. All participants had regular follow-up examination cycles every 4 to 8 years, and they provided written informed consent for the CT study and each attended examination. For the current study, we identified 2,581 participants from the Offspring and Third-Generation cohorts who had non-contrast-enhanced, non-gated, full thoracic CT scans covering the entire thymic bed between 2005 - 2011. Participants are under ongoing surveillance for cardiovascular disease endpoints and death. An endpoint review committee of senior investigators using all available data (e.g. hospital records) adjudicated the FHS study endpoints using standardized criteria. For subanalyses, Offspring participants who attended Exam 7 - 9 for Olink inflammation proteomics panel, longitudinal CRP, and Fried frailty analyses were identified.  The NLST is a longitudinal randomized trial of screening for lung cancer with the use of low-dose chest CT as compared with chest radiography. It enrolled participants from 2002 through 2004 with screening examinations from 2002 through 2007. Primary event

follow-up was done through December 31, 2009, and extended follow-up for overall survival and disease-specific survival was registered through 2015. Eligibility for enrollment required an age between 55 and 74 at randomization and a smoking history of a minimum of 30 pack-years without quitting smoking for more than 15 years before enrollment. Each participant provided written informed consent. For the current study, we identified 25,031 participants from the first low-dose CT screening exam (T0), which was performed soon after the time of randomization in Institutional Review Board-approved centers of the Lung Screening Study (LSS) or in centers of the American College of Radiology Imaging Network (ACRIN) which were responsible for collecting participant data.

## Outcomes

The main clinical outcomes are:

Overall survival - measured from the date of the CT examination (FHS) or from the date of randomization (NLST) until the date of death. For time-to-event analyses, we used a cut-off at 12 years of follow-up for both FHS and NLST for all survival and incidence analyses except for lung cancer incidence, for which data was only available for a follow-up of 6 years after randomization. If participants were alive or did not have an event after 12 or 6 years of follow-up, respectively, they were censored at that time.

Disease incidence, including lung cancer incidence and soft CVD incidence - measured from the date of the CT examination (FHS) or from the date of randomization (NLST) until the date the participant was last known to be free of lung cancer or free of CVD.

Disease specific-mortality (NLST) - measured from the date of randomization until the date of death from the prespecified causes based on ICD-10 codes as detailed in Supplementary Data Tables S1 - S5.

## Plants

### Seed stocks

*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*

### Novel plant genotypes

*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*

### Authentication

*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.*