Supplementary Appendix

Decoding biological age from face photographs using deep learning

Osbert Zalay*, Dennis Bontempi*, Danielle S Bitterman, Nicolai Birkbak, Derek Shyr, Fridolin Haugg, Jack M Qian, Hannah Roberts, Subha Perni, Vasco Prudente, Suraj Pai, Andre Dekker, Benjamin Haibe-Kains, Christian Guthier, Tracy Balboni, Laura Warren, Monica Krishan, Benjamin H Kann, Charles Swanton, Dirk De Ruysscher, Raymond H Mak^, Hugo JWL Aerts^+

Corresponding author: haerts@bwh.harvard.edu

Supplementary Information 1 | Datasets

Discovery Datasets | For training we used the IMDb-Wiki database (*42*), a publicly available age-labeled online database of 523,051 face images altogether. An independently curated database of age-labeled face images (that also contained gender and ethnicity labels), the UTKFace (UTK) database (*18*), was used to evaluate the technical performance of the model. The UTK database contains 24,109 images in total, subdivided into three datasets. Together, these training and testing databases contain photos of known individuals (in particular politicians, actors, professional athletes and other well-known people) in addition to photos of other people in the public domain whose birth dates can be verified. All photographs are labeled with the photo date and birthdate of the individual so that the age at the moment of the photograph can be determined. For the training dataset, 56,304 images from the reference IMDB-Wiki database were selected after applying exclusion criteria, using randomization and augmentation with rebalancing, and performing manual quality assurance on images with age labels of 60 years or older. No clinical patient datasets were used in model training. We randomly rebalanced the training dataset with augmentation using coordinate deformation, horizontal flips, and up to 20 degrees rotation either way, to create a uniformly distributed training set over the age range of 18 to 105, targeting a per-age-year sample size of between 600-700 images. As the dataset was too large to perform manual quality assurance on all images, manual curation and image quality assurance was performed on the training images with age labels of 60 or older because that age group is the most relevant clinically with respect to the oncology datasets we tested (comprising of ~15,000 of the training images), and to ensure the model would perform at its best over in this age range. In terms of criteria for manual quality assurance, we removed images that were of poor resolution, had artifacts or distortions, or in which the face was covered either completely or partially, or in which there was no face present. For technical validation we assessed the performance of FaceAge across genders and ethnicities in the presumed healthy individuals included in the UTK dataset. After manual quality assurance and curation, data of 2,547 individuals were included in subsequent analyses. Only age-labeled photos of real people were used in model training and validation; the website https://thispersondoesnotexist.com served to generate example face photos for illustrative purposes and figure creation, so as to not publish face photos of real people, but was not used in any technical capacity.

Clinical Datasets | Three large retrospective oncology datasets from separate institutions were used for testing of the FaceAge algorithm totaling 6,196 cancer patients in the final analysis. Two smaller datasets of non-cancerous patients totaling 535 patients were used as a control for validation purposes. Cancer patients were allowed to have had multiple courses of radiotherapy, as well as surgery and/or systemic therapy, although curatively treated patients had only a single course of radiation treatment. All face photographs used for the analyses of patients treated curatively were acquired prior to the patient's first treatment. Patients were excluded if no treatment registration photographs were available or of poor-quality, or if their registration date, treatment date, or photo date did not correspond within three months of each other.

MAASTRO Cohort: The first clinical dataset consists of 6,835 patients with a cancer diagnosis of which data was prospectively collected and included in the MAASTRO Biobank (Maastricht, The Netherlands). These patients were treated with both curative and palliative intent between 2006 and 2019. The predominant primary malignancies amongst these patients were breast,

colorectal, prostate, lung and head and neck cancer. After eliminating records of patients with missing face images, duplicate records, records of patients without follow-up information, and manual image quality assessment, a total of 5,498 entries remained. After removing records of metastatic and/or patients treated with palliative intent or for ductal carcinoma in-situ of the breast (DCIS), the final cohort contained data of 4,906 patients.

Harvard Thoracic Cohort: The second clinical dataset consists of 2035 records of thoracic cancer patients who had their most recent treatment with radiotherapy at Dana Farber - Brigham and Women's Cancer Center between 2008 and 2018. The predominant histology was adenocarcinoma (a form of non-small cell lung cancer) and most patients had Stage III cancer (based on AJCC 7th edition). After eliminating duplicate records and applying exclusion criteria, 802 records remained. Manual image quality assurance and curation reduced the number to a final analysis cohort of 573 patients.

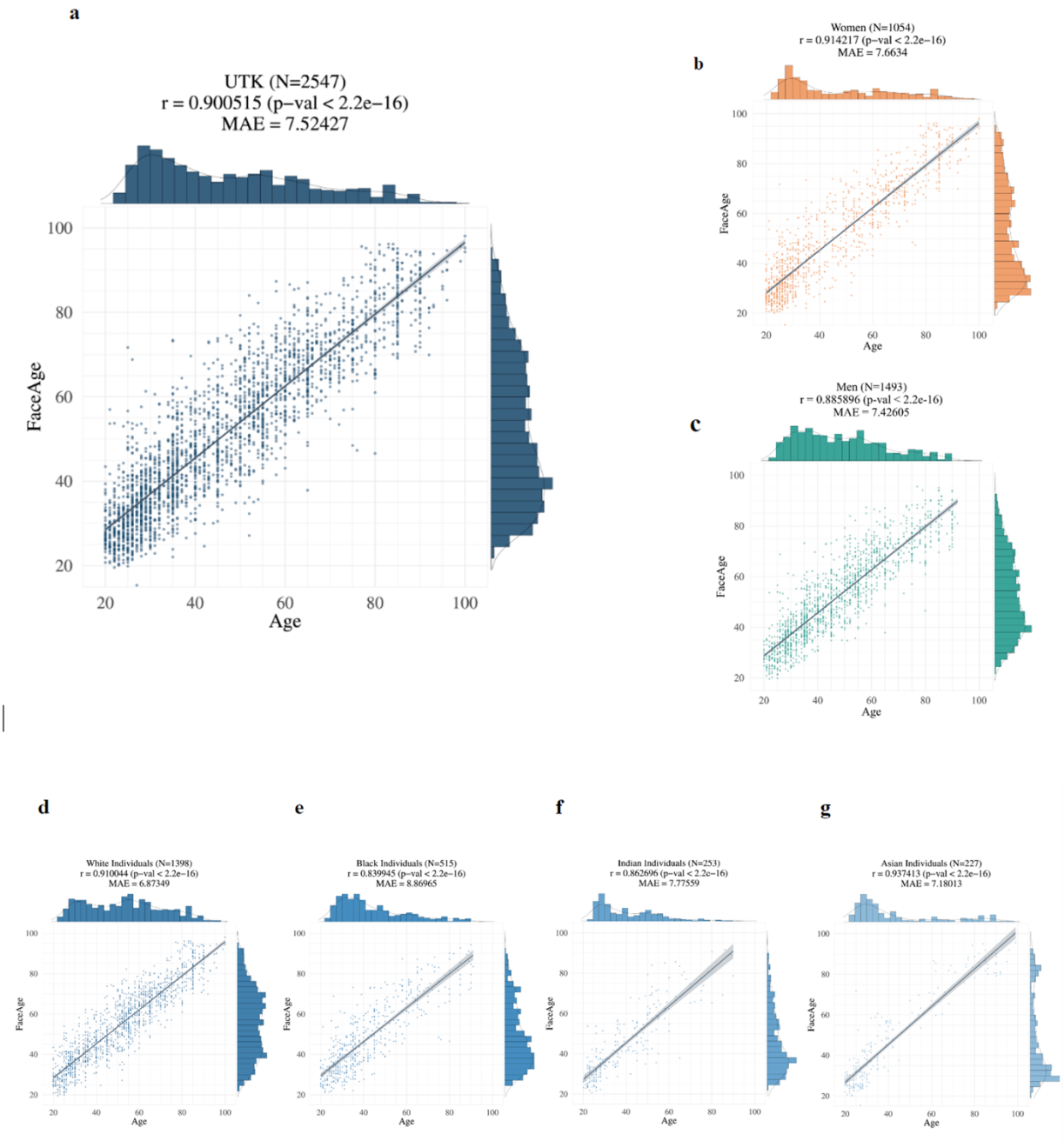
Harvard Palliative Cohort: The third clinical dataset consists of 1775 records of palliative patients with metastatic disease seen for consideration of palliative-intent treatment at Dana Farber - Brigham and Women's Cancer Center between 2008 and 2020. The predominant primary malignancies amongst these palliative patients were lung, breast, prostate and colorectal cancer. After removing duplicate records, records of patients who ended up not receiving treatment, and records with inconsistent dates and/or missing or poor-quality face images, 717 patients remained for subsequent analyses.

Harvard Non-cancerous Cohorts: Two smaller cohorts of patients who had their face photographs taken in a clinical setting as part of routine workflow were used as a non-cancerous control to evaluate FaceAge model age predictions, which could then be compared with the predictions from the oncology cohorts. The first cohort consisted of patients treated with benign conditions including, keloids, heterotopic ossification, benign intracranial tumors such as meningiomas and vestibular schwannomas, and cardiovascular conditions, and the second cohort consisted of patients with ductal carcinoma *in situ* of the breast, a precancerous condition that if left untreated leads to development of invasive breast cancer in approximately 30% of patients. The datasets were generated using queries of the electronic medical record systems of Dana Farber - Brigham and Women's Cancer Center based on clinical indications for radiation therapy, and face photographs were collected between 2009-2023 as part of routine clinical care. The same quality assurance procedure was applied to such datasets before processing (e.g., removal of images with face partially covered with a face mask, error during face extraction phase (MTCNN), etc.), leading to exclusion of 62 and 46 patients for the benign and DCIS cohorts, respectively, resulting in the final inclusion of 112 patients in the benign cohort and 423 patients in the DCIS cohort.

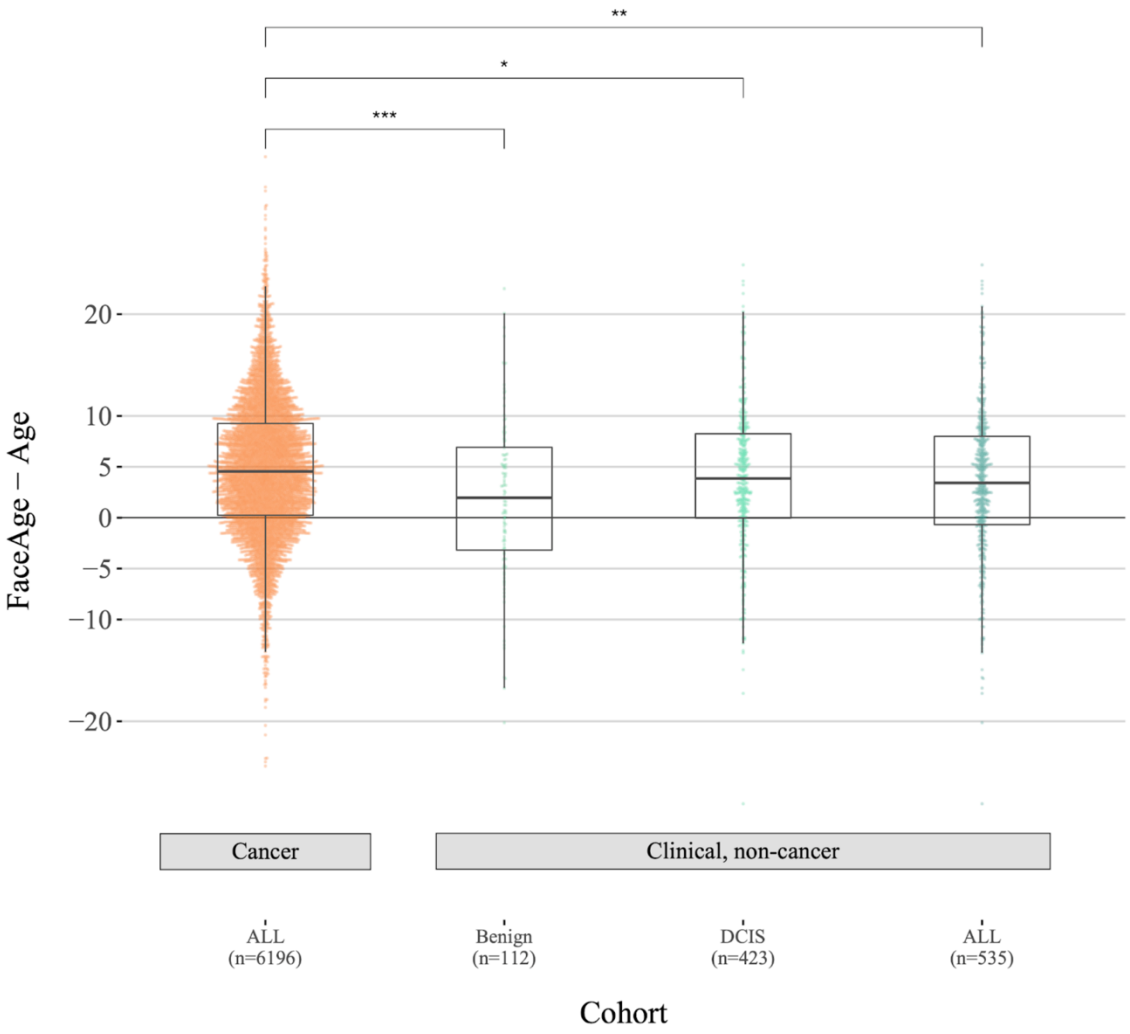
Supplementary Information 2 | Physician Survey

Physician Survey and Comparison of Human to Machine Performance | A survey was conducted to assess the performance of oncologists and palliative care physicians in estimating the apparent age and 6-month survival of $n = 100$ randomly-selected palliative cancer patients from the Harvard Palliative database, and to compare their performance against FaceAge directly, and to a Cox proportional hazards survival model based on FaceAge. The survey was sent to attending physicians, residents and lay researchers at Harvard-affiliated hospitals. A total of 10 survey participants were enlisted: 5 attending staff, 3 residents and 2 lay researchers. The survey consisted of two parts, administered two weeks apart to reduce memory bias. The first part of the survey presented survey takers with the face photograph of each of the 100 patients, and no accompanying chart information, and the survey-taker then asked to estimate the age of the patient (by decade) and whether the patient would be alive in 6 months' time (*53–55*). The second part of the survey presented survey-takers with the face photograph accompanied by chart information (without identifiers) that contained the same clinical information available to a Cox PH survival risk model incorporating FaceAge. This risk model was used in the survey to compute a predicted probability of death with respect to time, incorporating the clinical covariates of the TEACHH database, using FaceAge in place of chronologic age, with the same covariates made available to clinicians for survival prediction. The FaceAge risk model was fitted to the remainder of the Harvard Palliative cohort excluding the 100 randomly-selected survey cases, using forward and backward selection of covariates with p-value cutoff of 0.2 (see appendix p. 16-17). During the second part of the survey, survey-takers were asked to estimate the probability (in increments of 10%) that the given patient would be alive in 6 months, with all chart information provided. Once their response was given, the FaceAge risk model individualized survival probability curve was then presented to them, and the survey-taker asked to give their estimate of survival probability again, with the survey-taker having the choice of ignoring the new information provided by the FaceAge risk model or modifying their answer accordingly. The area under the receiver operating characteristic curve (AUC) and concordance index (C-index) were used to evaluate and compare the estimates of survey-takers and the FaceAge risk model against ground truth, and groups were compared using the non-parametric two-sided paired Wilcoxon signed rank test. A mock survey case with parts 1 and 2 is presented in the appendix for reference (p. 26). A post-survey questionnaire gathering demographics about the survey-taker, including whether they were an attending or resident, and years of experience, was also included.

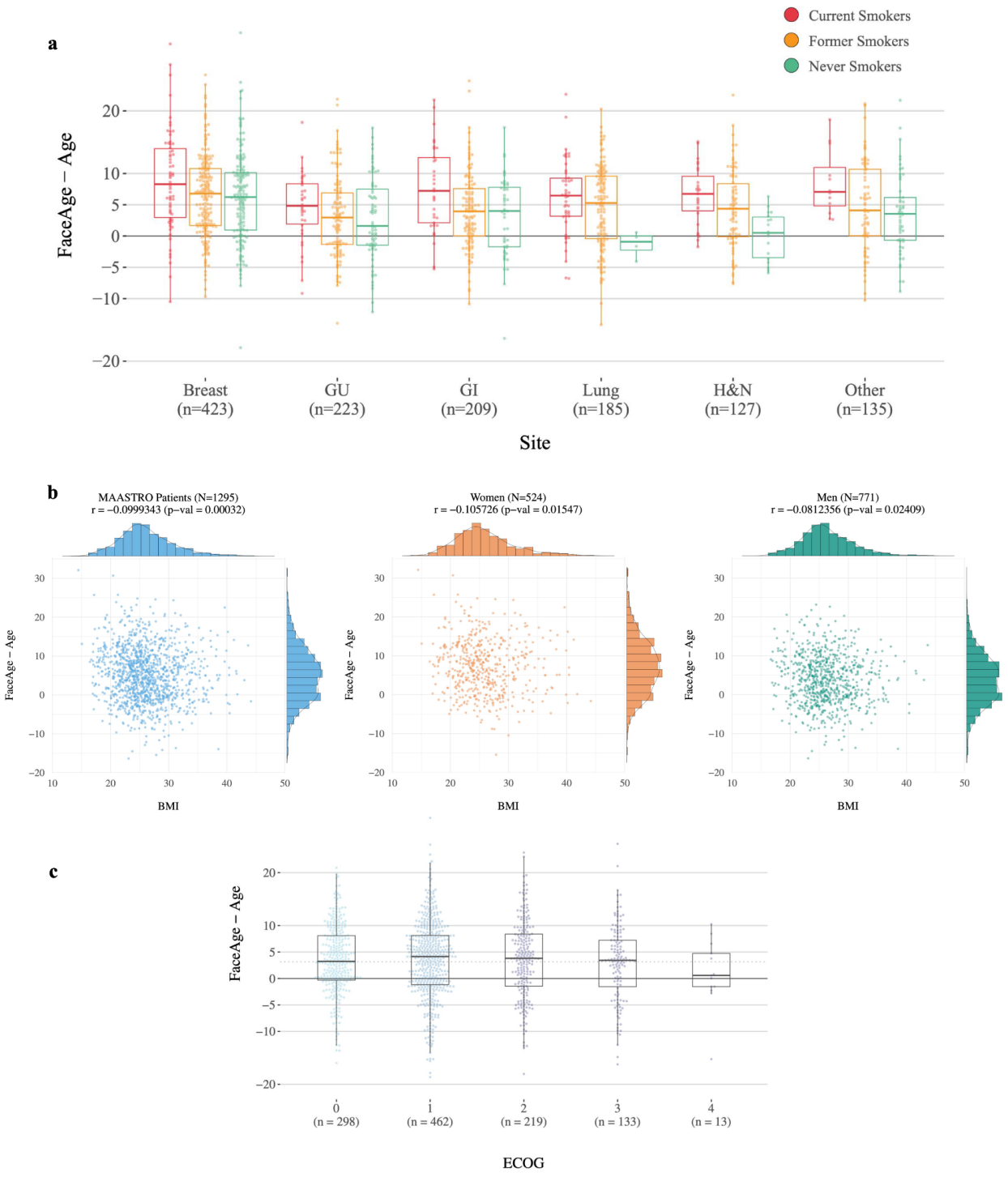
Supplementary Information 3 | Results



Supplementary Figure 1.
Performance of the FaceAge algorithm in the independent UTK test dataset, for all (**a**), women (**b**), men (**c**), as well as for white (**d**), black (**e**), indian (**f**), and asian (**g**) ethinic subgroups. The model performance is similar and significant across all the groups. *MAE*: Mean Absolute Error, *r*: Pearson R.



Supplementary Figure 2.
Difference between FaceAge and age in healthy and clinical non-cancer cohorts, to investigate if individuals look older or younger compared to their age. Analyzing all cancer patients included in our analysis, we found that, on average, cancer patients look older than their age (mean 4.79 years, $P < 0.001$). This larger FaceAge-to-chronologic age gap was significant when comparing cancer patients to the reference UTK dataset of healthy individuals 60 years and older ($P < 0.001$) and to the two clinical non-cancer datasets (benign patients: $P < 0.0001$ and DCIS patients: $P < 0.019$) acquired in the same clinical settings and with the same equipment as that of cancer patients, demonstrating that cancer patients look older than those who do not have cancer.



Supplementary Figure 3.

Association of the FaceAge algorithm with lifestyle factors. **a)** Difference between FaceAge and age with the smoking history for different types of cancer patients in the MAASTRO cohort. **b)** Association of the difference between FaceAge and age with body mass index (BMI), for all, women, and men in the MAASTRO cohort. **c)** Association of the difference between FaceAge and age with performance status (Eastern Cooperative Oncology Group (ECOG)) in the HARVARD cohort. We found no significant differences between the groups (unpaired two-sided t-test, $P > 0.165$). (*r*: pearson R.)

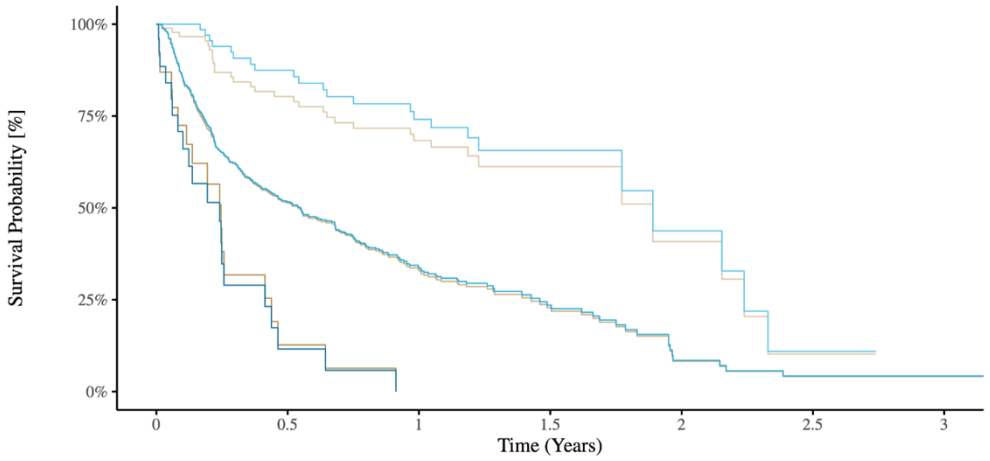
a

TEACHH Model	Median Survival (years)			Concordance	Log-likelihood Ratio (2df)
	Low Risk	Medium Risk	High Risk		
Age	1.89	0.556	0.241	0.587 (SE: 0.012)	63.3 ($p < 0.001$)
FaceAge	2.153	0.578	0.208	0.595 (SE: 0.011)	75.1 ($p < 0.001$)

b

TEACHH Model	Low Risk			Medium Risk			High Risk		
	HR	95% CI	P value	HR	95% CI	P value	HR	95% CI	P value
Age	0.2751	0.1797–0.4434	< 0.001	n/a	Reference	Reference	2.4286	1.6667–3.5389	< 0.001
FaceAge	0.2157	0.1209–0.3846	< 0.001	n/a	Reference	Reference	2.7449	1.9346–3.8944	< 0.001

c



Number at risk

Age L	90	58	40	10	4	1	0
Age M	449	173	83	25	6	3	3
Age H	23	2	0	0	0	0	0
FaceAge L	70	50	34	10	4	1	0
FaceAge M	466	181	89	25	6	3	3
FaceAge H	26	2	0	0	0	0	0

Supplementary Figure 4.

Cox regression analysis of age and FaceAge TEACHH models in the HARVARD palliative cohort. Using FaceAge instead of chronologic age as a covariate in the TEACHH model increases its discriminatory power, as quantified by decreased median survival and increased hazard ratio (HR) of the highest of the risk group, and increased median survival and decreased HR of the lowest risk group. c) TEACHH model Kaplan-Meier survival curves (all-cause mortality) obtained by using chronologic age (solid line) and FaceAge (dashed line) as covariates with 60 years threshold for both. Substituting age with FaceAge, significantly increased the discriminatory power of the model, increasing hazard ratio (HR) of the highest of the risk groups and decreasing the HR of the lowest risk group.

Supplementary Figure 5.
Association of FaceAge and Chronological Age with senescence genes. GeneMania network of senescence genes. Red edges indicate physical interaction, green genetic interactions, and blue pathway. The size of the node represents the score assigned by label propagation algorithms reflecting the strength of association between the node and the input list of genes, i.e., TERT, ATM, CDKN1A, CDKN2B, TP53, IGFBP7, and MAPK10.

a

Characteristics		Number	Percent
Sex	Female	2310	47.1
Age at Treatment	Median	66.0	
	Range	22.0 – 94.0	
Main Tumor Group	Breast cancer	1337	27.3
	Gastrointestinal cancer	1003	20.4
	Genitourinary cancer	843	17.2
	Lung cancer	737	15.0
	Head and neck cancer	456	9.3
	Other types of cancer	530	10.8
Smoking History	Yes (current/former/never)	1302	26.5
BMI	Yes	1297	26.4
ECOG	Yes	1170	23.8

b

Characteristics		Number	Percent
Sex	Female	303	52.8
Age at Treatment	Median	69.0	
	Range	33.3 – 93.2	
Median Overall Survival (months)	Crude	14.4	
	Actuarial	16.9	
Diagnosis	Non-small cell lung cancer	450	78.5
	Small cell lung cancer	49	10.9
	Other types of cancer	74	12.9
Treatment Intent	Curative non-SBRT	433	75.6
	Curative SBRT	106	18.5
	Palliative	34	5.9
Clinical Stage	I	145	25.3
	II	70	12.2
	III	279	48.7
	IV	70	12.2
	Not specified	9	1.6
Tumor Grade	≥ 2	322	56.2
Race	Caucasian	493	86.0
Smoking History	Yes (current/former/never)	509	88.8
BMI	Yes	106	18.5
ECOG	≥ 2	109	19.0

Supplementary Table 1.
Clinical characteristics of the a) MAASTRO Cohort (n=4,906), and b) HARVARD thoracic cohort (N=573). *SBRT*: Stereotactic Body Radioherapy.

Characteristics		Number	Percent
Sex	Female	384	53.6
Age at Treatment	Median	65.2	
	Range	19.6 – 97.6	
Median Overall Survival (months)	Crude	4.5	
	Actuarial	8.2	
Time from Diagnosis to Metastasis	Median	1.0 months	
	Range	0 – 26.8 years	
Time from Metastasis to RT Consult	Median	4.5 months	
	Range	0 – 33.0 years	
Primary Diagnosis	Lung	201	28.0
	Breast	118	16.4
	Prostate	58	8.1
	Colorectal	44	6.1
	Gynecologic	43	6.0
	Melanoma	38	5.3
	Esophagus	24	3.3
	Renal	24	3.3
	Sarcoma	24	3.3
	Unknown primary	20	2.8
	Bladder	19	2.6
	Head and neck	10	2.6
	Pancreas	18	2.5
	Neuroendocrine	12	1.4
	Cholangiocarcinoma	9	1.3
	Hepatocellular	8	1.1
	Non-melanoma skin	8	1.1
	Stomach	8	1.1
	GU (non-bladder/testicular)	6	0.8
	Mesothelioma	5	0.7
	Testicular	4	0.6
	Small bowel	3	0.4
	Other	6	0.8
Metastases	Bone	344	48.0
	Brain	265	37.0
	Spine	256	35.7
	Lung	223	31.1
	Liver	222	31.0
	Lymph	219	30.5
	Adrenal	65	9.1
	Other	153	21.3
ECOG	0 – 1	309	43.1
	2	135	18.8
	3	110	15.3
	4	11	1.5
Prior Palliative RT (courses)	0	627	87.5
	1	72	10.0
	≥2	14	2.0
Prior Palliative Chemo (courses)	0	280	39.1
	1	197	27.1
	≥2	237	33.1
Hospital Admissions	0	370	51.6
	1	291	40.6
	≥2	55	7.7
ER Visits	0	421	58.7
	1	233	32.5
	≥2	62	8.7

Supplementary Table 2.
Clinical characteristics of the HARVARD palliative cohort (N=717).

Variable	Univariate			Multivariate		
	HR	95% CI	P value	HR	95% CI	P value
FaceAge (decade)	1.428	1.351 – 1.510	< 0.001	1.151	1.057 – 1.254	0.00129
Chronologic Age (decade)	1.427	1.365 – 1.501	< 0.001	1.244	1.150 – 1.346	< 0.001
Sex						
Female	1 (ref.)	–	–	1 (ref.)	–	–
Male	1.539	1.390 – 1.705	< 0.001	1.359	1.204 – 1.533	< 0.001
Site						
Breast	1 (ref.)	–	–	1 (ref.)	–	–
Gastrointestinal	3.162	2.578 – 3.879	< 0.001	2.297	1.839 – 1.869	< 0.001
Genitourinary	1.671	1.331 – 2.096	< 0.001	1.019	0.788 – 1.318	0.885
Lung	9.156	7.526 – 11.140	< 0.001	6.468	5.228 – 8.002	< 0.001
Head and neck	3.278	2.600 – 4.133	< 0.001	2.484	1.929 – 3.198	< 0.001
Other	3.770	3.017 – 4.709	< 0.001	3.394	2.703 – 4.261	< 0.001

Supplementary Table 3.
Univariate and multivariate survival analyses of FaceAge on the MAASTRO dataset. FaceAge univariate and multivariate Cox regression analysis for the MAASTRO cohort (N=4,906). Since FaceAge, age, sex, and cancer site have a p-value of less than 0.001 in the univariate analysis, we include all the available covariates in the multivariate model. FaceAge remains statistically significant after adjusting for all the aforementioned covariates. CI: Confidence Interval; *HR*: Hazard Ratio.

Variable	Univariate			Multivariate (FaceAge)			Multivariate (Age)		
	HR	95% CI	P value	HR	95% CI	P value	HR	95% CI	P value
FaceAge (decade)	1.086	0.99 – 1.19	0.084	1.148	1.03 – 1.28	0.011	–	–	–
Age (decade)	1.044	0.95 – 1.15	0.37	–	–	–	1.081	0.97 – 1.21	0.16
BMI (unit)	1.00083	0.96 – 1.04	0.97	–	–	–	–	–	–
Sex									
Female	1 (ref.)	–	–	1 (ref.)	–	–	1 (ref.)	–	–
Male	1.273	1.03 – 1.57	0.024	1.315	1.06 – 1.63	0.012	1.284	1.04 – 1.59	0.021
Smoking									
Never	1 (ref.)	–	–	1 (ref.)	–	–	1 (ref.)	–	–
Current/former	1.261	0.84 – 1.87	0.26	1.014	0.67 – 1.54	0.95	1.064	0.70 – 1.62	0.77
Ethnicity									
Caucasian	1 (ref.)	–	–	1 (ref.)	–	–	1 (ref.)	–	–
Non-Caucasian	1.137	0.82 – 1.56	0.43	1.024	0.74 – 1.43	0.89	1.041	0.75 – 1.45	0.81
Treatment Intent									
Palliative	1 (ref.)	–	–	1 (ref.)	–	–	1 (ref.)	–	–
Curative	0.458	0.3 – 0.69	< 0.001	0.538	0.34 – 0.86	0.0096	0.554	0.35 – 0.89	0.014
Tumor Grade									
1	1 (ref.)	–	–	1 (ref.)	–	–	1 (ref.)	–	–
> 1	1.157	0.94 – 1.43	0.18	1.086	0.86 – 1.37	0.49	1.069	0.85 – 1.35	0.58
ECOG									
≤ 1	1 (ref.)	–	–	1 (ref.)	–	–	1 (ref.)	–	–
> 1	1.295	1.01 – 1.66	0.039	1.340	1.02 – 1.76	0.034	1.375	1.05 – 1.80	0.021
Overall Stage									
I	1 (ref.)	–	–	1 (ref.)	–	–	1 (ref.)	–	–
II	0.822	0.59 – 1.14	0.24	1.179	0.79 – 1.75	0.42	1.129	0.76 – 1.68	0.55
III	1.293	1.05 – 1.59	0.016	1.667	1.25 – 2.23	< 0.001	1.605	1.20 – 2.15	0.0015
IV	1.542	1.12 – 2.12	0.0074	2.029	1.33 – 3.09	0.0010	2.000	1.30 – 3.05	0.0016
Histology									
Other	1 (ref.)	–	–	1 (ref.)	–	–	1 (ref.)	–	–
NSCLC	1.053	0.84 – 1.32	0.66	1.221	0.92 – 1.62	0.16	1.222	0.92 – 1.62	0.16
SCLC	1.577	1.12 – 2.21	0.009	1.785	1.16 – 2.74	0.0083	1.781	1.16 – 2.74	0.0085

Supplementary Table 4.

Univariate and multivariate survival analyses of FaceAge on the HARVARD Thoracic dataset. FaceAge univariate and multivariate Cox regression analysis for the HARVARD Thoracic cohort (N=573). In the multivariate model, FaceAge remains statistically significant after adjusting for sex, smoking history, ethnicity, treatment intent, tumour grade, ECOG, overall stage, and histology. The same covariates were used to fit a model with age, where age did not have a significant effect. Note, that the multivariate HR for FaceAge in the HARVARD Thoracic dataset is close to the one computed in the multivariate analysis on the MAASTRO cohort (Extended Data Figure 3). BMI was excluded from multivariate analysis as this information was available for only 18.5% of patients. *BMI*: Body Mass Index; *ECOG*: Eastern Cooperative Oncology Group scale; *CI*: Confidence Interval; *HR*: Hazard Ratio.

Variable	Univariate			Multivariate (FaceAge)			Multivariate (Age)		
	HR	95% CI	P value	HR	95% CI	P value	HR	95% CI	P value
				The multivariate analysis includes all the covariates for which p < 0.2 in the univariate analysis					
FaceAge (decade)	1.102	1.01 – 1.21	0.035	1.117	1.02 – 1.23	0.021	–	–	–
Age (decade)	1.065	0.98 – 1.16	0.16	–	–	–	1.076	0.98 – 1.18	0.11
Number of ER Admits	2.323	1.86 – 2.90	< 0.001	1.519	1.14 – 2.02	0.0039	1.493	1.12 – 1.99	0.0058
Number of Hospital Admits	1.678	0.35 – 2.09	< 0.001	1.020	0.77 – 1.34	0.89	1.014	0.77 – 1.34	0.92
Number of Prior RT	1.443	1.04 – 2.00	0.027	1.425	1.00 – 2.04	0.053	1.445	1.01 – 2.07	0.045
Number of Prior Chemo	1.244	0.95 – 1.63	0.12	1.854	1.32 – 2.60	< 0.001	1.910	1.36 – 2.68	< 0.001
Time to Consult (years)	0.919	0.86 – 0.98	0.011	0.906	0.83 – 0.99	0.024	0.900	0.83 – 0.98	0.015
Time to 1st Metastasis (years)	0.92	0.92 – 0.95	< 0.001	0.944	0.91 – 0.98	0.0026	0.948	0.91 – 0.98	0.0043
Sex									
Female	1 (ref.)	–	–	1 (ref.)	–	–	1 (ref.)	–	–
Male	1.021	0.82 – 1.27	0.85	–	–	–	–	–	–
Ethnicity									
Caucasian	1 (ref.)	–	–	1 (ref.)	–	–	1 (ref.)	–	–
Non-Caucasian	1.053	0.79 – 1.41	0.73	–	–	–	–	–	–
ECOG									
≤ 1	1 (ref.)	–	–	1 (ref.)	–	–	1 (ref.)	–	–
2	1.852	1.45 – 2.37	< 0.001	2.282	1.73 – 3.00	< 0.001	2.282	1.73 – 3.01	< 0.001
3-4	2.547	1.97 – 3.29	< 0.001	2.762	2.03 – 3.75	< 0.001	2.825	2.08 – 3.83	< 0.001
Tumor Site									
Breast	1 (ref.)	–	–	1 (ref.)	–	–	1 (ref.)	–	–
Lung	1.463	1.16 – 1.85	0.0016	1.690	1.98 – 2.65	0.022	1.783	1.14 – 2.28	0.011
Prostate	0.446	0.28 – 0.72	< 0.001	1.150	0.63 – 2.09	0.65	1.189	0.65 – 2.16	0.65
Colorectal	0.674	0.43 – 1.05	0.081	1.097	1.62 – 1.96	0.75	1.150	0.65 – 2.05	0.57
Gynecologic	0.727	0.44 – 1.20	0.21	–	–	–	–	–	–
Other	1.574	1.15 – 1.78	< 0.001	2.128	1.43 – 3.16	< 0.001	2.176	1.47 – 3.23	< 0.001
Bone Metastases									
No	1 (ref.)	–	–	1 (ref.)	–	–	1 (ref.)	–	–
Yes	1.103	0.89 – 1.37	0.38	–	–	–	–	–	–
Lung Metastases									
No	1 (ref.)	–	–	1 (ref.)	–	–	1 (ref.)	–	–
Yes	1.202	0.96 – 1.51	0.11	1.364	1.06 – 1.75	0.015	1.360	1.06 – 1.75	0.016
Liver Metastases									
No	1 (ref.)	–	–	1 (ref.)	–	–	1 (ref.)	–	–
Yes	1.321	1.06 – 1.65	0.015	1.458	1.15 – 1.86	0.0022	1.445	1.13 – 1.84	0.0029
Brain Metastases									
No	1 (ref.)	–	–	1 (ref.)	–	–	1 (ref.)	–	–
Yes	1.734	1.37 – 2.20	< 0.001	1.505	1.13 – 2.01	0.0054	1.508	1.13 – 2.01	0.0052
Spine Metastases									
No	1 (ref.)	–	–	1 (ref.)	–	–	1 (ref.)	–	–
Yes	1.277	1.02 – 1.59	0.03	1.189	0.93 – 1.52	0.17	1.190	0.93 – 1.52	0.17
Lymph Nodes Metastases									
No	1 (ref.)	–	–	1 (ref.)	–	–	1 (ref.)	–	–
Yes	0.947	0.75 – 1.20	0.65	–	–	–	–	–	–
Adrenal Metastases									
No	1 (ref.)	–	–	1 (ref.)	–	–	1 (ref.)	–	–
Yes	2.217	1.57 – 3.13	< 0.001	1.495	1.02 – 2.18	0.037	1.466	1.01 – 2.14	0.047
Other Metastases									
No	1 (ref.)	–	–	1 (ref.)	–	–	1 (ref.)	–	–
Yes	1.324	1.03 – 1.71	0.031	1.088	0.83 – 1.43	0.55	1.112	0.85 – 1.46	0.45

Supplementary Table 5.

Univariate and multivariate survival analyses for FaceAge and Age on the HARVARD Palliative dataset. FaceAge univariate and multivariate Cox regression analysis for the HARVARD Palliative cohort (N=717). The covariates for the multivariate model are selected using a forward selection procedure with a p-value cutoff of $P < 0.2$. In the final multivariate model, FaceAge remains statistically significant after adjusting for sex, smoking history, ethnicity, treatment intent, tumour grade, ECOG, overall stage, and histology, while age was not. Furthermore, the multivariate HR for FaceAge in the HARVARD Palliative dataset is similar to the multivariate analysis on both the MAASTRO and HARVARD Thoracic cohorts (Extended Data Figures 5 and 6). BMI: Body Mass Index; ECOG: Eastern Cooperative Oncology Group scale; CI: Confidence Interval; HR: Hazard Ratio.