Deep learning assessment of senescence-associated nuclear morphologies in mammary tissue from healthy female donors to predict future risk of breast cancer: a retrospective cohort study

Indra Heckenbach, Mark Powell, Sophia Fuller, Jill Henry, Sam Rysdyk, Jenny Cui, Amanuel Abraha Teklu, Eric Verdin, Christopher Benz, Morten Scheibye-Knudsen

Summary

Background Cellular senescence has been associated with cancer as either a barrier mechanism restricting autonomous cell proliferation or a tumour-promoting microenvironmental mechanism that secretes proinflammatory paracrine factors. With most work done in non-human models and the heterogeneous nature of senescence, the precise role of senescent cells in the development of cancer in humans is not well understood. Furthermore, more than 1 million non-malignant breast biopsies are taken every year that could be a major resource for risk stratification. We aimed to explore the clinical relevance for breast cancer development of markers of senescence in mammary tissue from healthy female donors.

Methods In this retrospective cohort study, we applied single-cell deep learning senescence predictors, based on nuclear morphology, to histological images of haematoxylin and eosin-stained breast biopsy samples from healthy female donors at the Komen Tissue Bank (KTB) at the Indiana University Simon Cancer Center (Indianapolis, IN, USA). All KTB participants (aged ≥18 years) who underwent core biopsies for research purposes between 2009 and 2019 were eligible for the study. Senescence was predicted in the epithelial (terminal duct lobular units [TDLUs] and non-TDLU epithelium), stromal, and adipose tissue compartments using validated models, previously trained on cells induced to senescence by ionising radiation (IR), replicative exhaustion (or replicative senescence; RS), or antimycin A, atazanavir–ritonavir, and doxorubicin (AAD) exposures. To benchmark our senescence-based cancer prediction results, we generated 5-year Gail scores—the current clinical gold standard for breast cancer risk prediction—for participants aged 35 years and older on the basis of characteristics at the time of tissue donation. The primary outcome was estimated odds of breast cancer via logistic modelling for each tissue compartment based on predicted senescence scores in cases (participants who had been diagnosed with breast cancer as of data cutoff, July 31, 2022) and controls (those who had not been diagnosed with breast cancer).

Findings 4382 female donors (median age at donation 45 years [IQR 34–57]) were eligible for the study. As of data cutoff (median follow-up of 10 years [7–11]), 86 (2·0%) had developed breast cancer a mean of 4·8 years (SD 2·84) after date of donation and 4296 (98·0%) had not received a breast cancer diagnosis. Among the 86 cases, we found significant differences in adipose-specific IR and AAD senescence prediction scores compared with controls. Risk analysis showed that individuals in the upper half (above the median) of scores for the adipose tissue IR model had higher odds of developing breast cancer (odds ratio [OR] 1·71 [95% CI 1·10–2·68]; p=0·019), whereas the adipose AAD model revealed a reduced odds of developing breast cancer (OR 0·57 [0·36–0·88]; p=0·013). For the other tissue compartments and the RS model, no significant associations were found (except for stromal tissue via the IR model, had higher odds of developing breast cancer [OR 1·59, 1·03–2·49]). Individuals with both of the adipose risk factors had an OR of 3·32 (1·68–7·03; p=0·0009). Participants with 5-year Gail scores above the median had an OR for development of cancer of 2·33 (1·46–3·82; p=0·0012) compared with those with scores below the median. When combining Gail scores with our adipose AAD risk model, we found that individuals with both of these predictors had an OR of 4·70 (2·29–10·90; p<0·0001). When combining the Gail score with our adipose IR model, we found that individuals with both predictors had an OR of 3·45 (1·77–7·24; p=0·0002).

Interpretation Assessment of senescence-associated nuclear morphologies with deep learning allows prediction of future cancer risk from normal breast biopsy samples. The combination of multiple models improved prediction of future breast cancer compared with the current clinical benchmark, the Gail model. Our results suggest an important role for microscope image-based deep learning models in predicting future cancer development. Such models could be incorporated into current breast cancer risk assessment and screening protocols.

Funding Novo Nordisk Foundation, Danish Cancer Society, and the US National Institutes of Health.

Lancet Digit Health 2024; 6: e681–90

Center for Healthy Aging, Department of Cellular and Molecular Medicine, University of Copenhagen, Copenhagen, Denmark (I Heckenbach PhD, A A Teklu MSc, M Scheibye-Knudsen MD DMSc); **Buck Institute for Research on Aging, Novato, CA, USA** (I Heckenbach, M Powell MD, Prof E Verdin MD, Prof C Benz MD); **Zero Breast Cancer, San Rafael, CA, USA** (M Powell); **Graduate Group in Biostatistics, School of Public Health, University of California, Berkeley, Berkeley, CA, USA** (S Fuller MA); **Susan G Komen Tissue Bank at the IU Simon Comprehensive Cancer Center, Indiana University School of Medicine, Indianapolis, IN, USA** (J Henry MBA, S Rysdyk JD, J Cui MS); **Department of Biochemistry and Molecular Biology, College of Health Sciences, Mekelle University, Mekelle, Ethiopia** (A A Teklu)

Correspondence to:
Dr Morten Scheibye-Knudsen, Center for Healthy Aging, Department of Cellular and Molecular Medicine, University of Copenhagen, Copenhagen 2200, Denmark
mscheibye@sund.ku.dk

Introduction

Breast cancer is one of the most common forms of cancer worldwide and results in substantial mortality, with 287 850 cases and 43 250 deaths annually in the USA alone.[1] Numerous genetic and environmental factors contribute to the risk of breast cancer, but the primary risk factor is age, with over 80% of cases occurring after age 50 years, such that most countries begin regular mammographic screening in all women by this age.[2] Like many diseases strongly correlated with ageing, breast cancer has been associated with cellular senescence,[3–5] by which aged or damaged cells cease dividing but remain metabolically active.[6] Senescence was first identified as a mechanism to limit proliferation of cells[7] and later shown to operate via multiple pathways that respond to molecular damage to prevent excessive division, purportedly as a tumour suppressor mechanism.[8] Paradoxically, senescent cells can also promote tumour development by emitting a variety of paracrine factors in what is known as the senescence-associated secretory phenotype (SASP), which produces an inflammatory state to signal immune clearance.[9]

Despite its relevance to numerous diseases including cancer, cellular senescence is poorly characterised in human tissues because of the absence of specific and universal biomarkers.[10] Although diverse markers are associated with senescence, they also indicate other biological processes, such as DNA damage, inflammation, and cell cycles, and no known marker is exclusively associated with the senescent state. Best practice approaches to overcome this challenge include staining for multiple markers simultaneously, although little consensus exists regarding the right combination, which would also vary by tissue. Hence, initiatives have been launched, such as SenNet by the US National Institutes of Health (NIH), to map senescence in human tissues using the latest multiomics methods.[11,12]

Given the observation that senescent cells display altered nuclear morphology,[13] we recently showed that senescence can be detected using deep learning on tissue micrographs showing nuclear morphology.[14] In the current study, we applied the nuclear senescence predictor (NUSP) to images of haematoxylin and eosin (H&E)-stained breast biopsy samples from healthy female donors who were followed up for breast cancer, to investigate whether senescence is associated with future breast cancer development. We generated predicted senescence scores for nuclei in four tissue compartments and analysed the spatial distribution of predicted senescent cells to explore how senescence arises in tissue and how it might relate to future breast cancer development.

Methods

Study design and population

In this retrospective cohort study, we used data from participants of the Komen Tissue Bank (KTB) at the Indiana University Simon Cancer Center (Indianapolis, IN, USA), one of the largest biorepositories that collects,

Research in context

Evidence before this study

Cellular senescence has been associated with cancer as both a mechanism to limit proliferation and a factor promoting malignancy through the senescence-associated secretory phenotype. However, whether senescence in tissue is related to individual-level future likelihood of developing cancer or otherwise influences the risk of future cancer is unknown. We searched PubMed on May 22, 2023 , with no date or language restrictions for publications using the terms "cellular senescence" AND "breast cancer" AND "risk" and identified 45 results. We did not identify any studies investigating senescence as a prognostic marker for future breast cancer development. Accordingly, to our knowledge, no study has shown that senescence in healthy breast tissue is a risk factor for developing cancer and no study has been conducted in a large cohort and focused on morphology identified via haematoxylin and eosin (H&E)-stained biopsy images of healthy breast tissue.

Added value of this study

In our retrospective cohort study, we investigated the role of cellular senescence in the development of breast cancer in healthy individuals. We used deep learning methods to predict senescence on the basis of the nuclear morphology in H&E-stained biopsy images. By the use of high-precision senescence scores for more than 32 million individual nuclei, we identified spatial distributions of senescent cells in tissue types within breast biopsies. We found that, with multiple deep learning models, senescence is both positively and negatively associated with cancer development, depending on the type of senescence. By combining two senescence models, we obtained an odds ratio that considerably exceeded that of the Gail score, which is the clinical gold standard for assessing breast cancer risk.

Implications of all the available evidence

Our results suggest that deep learning models can be used to assess senescence-associated nuclear morphology to predict the risk of developing breast cancer using H&E-stained images obtained from negative breast biopsies. These findings could have clinically significant applications for patients who have negative breast biopsies, providing insight into their future cancer risk and informing both patients and clinicians on the need for more aggressive cancer screening and prevention.

stores, and annotates breast tissue, with a focus on female donors who exhibit no signs of breast disease.[15] Tissue cores from the upper outer quadrant of either breast were acquired from consenting donors using 10-gauge needles and immediately processed as snap-frozen tissues or using 10% formalin fixation and paraffin embedding or the PAXgene tissue preservation system before paraffin embedding. Tissues were collected and processed by standardised KTB operating protocols and samples were archived by the KTB.[16,17]

Donors supplied demographic data (including sex and race or ethnicity) and reproductive and medical histories through the completion of questionnaires at the time of donation and annual follow-up questionnaires. All KTB participants who underwent core biopsies for research purposes between 2009 and 2019 (more exact dates were not available) were eligible for this study. Participants were followed up from the time of biopsy up to July 31, 2022. Additional details are provided in the appendix (p 2).

KTB participants and biopsy donors were initially recruited after providing written informed consent under a protocol approved by the Indiana University Institutional Review Board. KTB participants who were eligible for this study had provided consent for biopsy samples to be used in future research and hence specific consent for this study was not needed.

Image segmentation and senescence prediction

H&E-stained whole slide images were provided by the KTB as tiled TIF (SVS) files. To assist in the analysis, these large images were split into tiles of 1024×1024 pixels. Nuclei were identified using a U-Net model trained on ten sample image tiles.[18] Tissue was also segmented by training of two U-Net models to detect epithelial and adipose regions, where each model was trained on 20 annotated tiles that were randomly selected to include multiple individuals and different regions of tissue. Terminal duct lobular units (TDLUs), which are believed to be the primary site for origination of breast malignancies,[19] were segmented using a published model also based on U-Net.[20] Additional details are provided in the appendix (p 2).

Cellular senescence was predicted from nuclear morphology in the epithelial (TDLU and non-TDLU epithelium), stromal, and adipose compartments in images of H&E-stained tissue using NUSP, which is a set of senescence prediction models that were previously trained on human fibroblasts in cell cultures. These fibroblasts were induced to senescence by ionising radiation (IR), replicative exhaustion (replicative senescence or RS), antimycin A, atazanavir–ritonavir, or doxorubicin exposures, as well as all five inducers together (the unified model), as previously described.[14] Due to high correlations, a new model (AAD) was trained using the same images from the drug treatments only (ie, including antimycin A, atazanavir–ritonavir, and doxorubicin). Senescence scores were generated using multiple models (IR, RS, and AAD) for each of the nuclei recognised, and the precise locations of predicted senescent cells in the tissue images were tracked. Spatial analyses were performed by classifying nuclei as most likely senescent (above 90th percentile of scores) and non-senescent (below 10th percentile of scores) and evaluating mean senescence scores of other nuclei of the same tissue type by distance, categorised into 20 buckets of equal width radial segments. Further details are provided in the appendix (pp 2–3).

To study nuclei classification by tissue type, we applied Uniform Manifold Approximation and Projection (UMAP) to project morphological metrics (ie, area, convexity, and aspect), colour intensities, and nuclei density and count, and three predicted senescence scores (IR, RS, and AAD) by tissue and individual. To benchmark our senescence-based cancer prediction results, we generated 5-year Gail scores for all participants aged 35 years and older on the basis of characteristics at the time of tissue donation. The Gail model or Breast Cancer Risk Assessment Tool—the current clinical gold standard for breast cancer risk prediction—uses a woman's medical and reproductive history and the history of breast cancer among her first-degree relatives to estimate the probability (%) of developing invasive breast cancer in the next 5 years.

See Online for appendix

Statistical analysis

The primary outcome was to estimate the odds of breast cancer for each tissue compartment via logistic regression. Donors were categorised as cases or controls on the basis of whether they had been diagnosed with breast cancer by the time of data cutoff July 31, 2022. Demographic and clinical characteristics were summarised for cases and controls using descriptive statistics, with characteristics compared using χ^2 test for categorical variables and Student's t test for continuous variables.

Individuals were compared by outcome (ie, breast cancer diagnosis) using mean senescence scores and 95% CIs, calculated by bootstrapping. Unadjusted p values were calculated using a two-sided Student's t test and the Benjamini–Hochberg method was used to adjust p values for multiple testing.

Odds ratios (ORs) and 95% CIs were calculated using logistic regression models to analyse the association between senescence scores and breast cancer incidence. Breast tissue was segmented into adipose, TDLU, non-TDLU epithelial, and stromal tissue for senescence scoring, and three senescence models (IR, RS, and AAD) were applied to the four tissue types. Results were generated as above or below the median of senescence scores with 95% CIs and the Benjamini-Hochberg method was applied for multiple testing. Examination by quartiles and continuous (linear) association was performed but was limited in statistical strength by the relatively small number of cases.

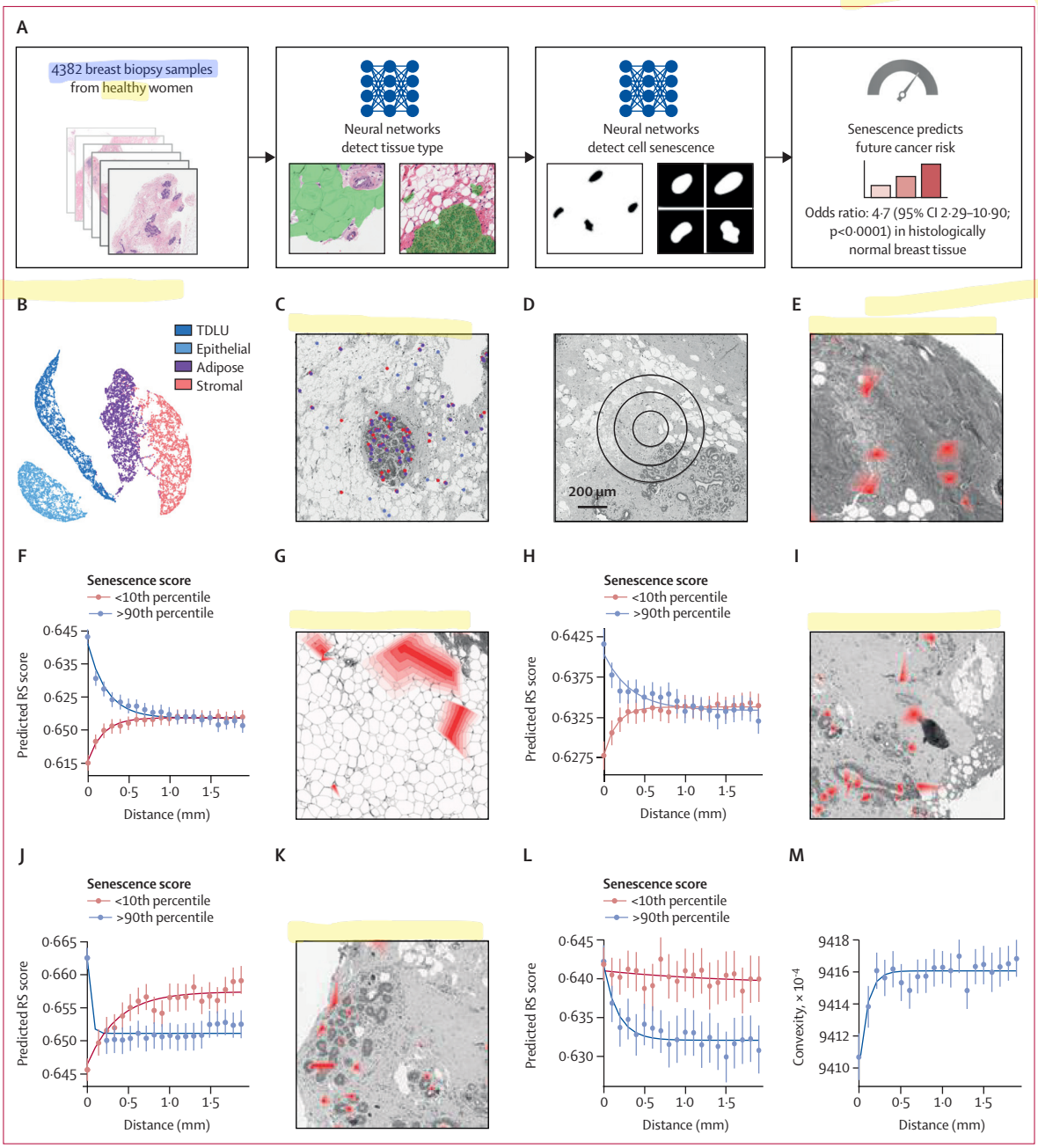

Figure 1: **Spatial and tissue-specific organisation of cellular senescence**

(A) Workflow for predicting senescence in tissue images. (B) Uniform Manifold Approximation and Projection of morphological metrics (nuclei area, nuclei convexity, nuclei aspect, IR score, AAD score, RS score, red intensity, green intensity, blue intensity, density, and count) by individual tissue, coloured to indicate predicted tissue. (C) Representative histological images showing nuclei with high predicted senescence (>90th percentile of predicted senescence scores). Models trained on different senescence inducers: IR in red, RS in purple, and AAD in blue. (D) Representative image illustrating the distance cutoffs where senescence is measured. (E) Representative image of stroma showing clusters (in red) of nuclei that have higher predicted scores for replicative senescence. (F) Mean RS scores of cells in stromal tissue with increasing distance from senescent (score >90th percentile) and non-senescent (score <10th percentile) cells (exponential curve fit: at >90th percentile, R^2=0·95; at <10th percentile, R^2=0·98). (G) Representative image of adipose tissue sample showing regions of high-scoring RS senescent nuclei in red. (H) Mean RS scores of cells in adipose tissue with increasing distance from senescent (score >90th percentile) and non-senescent (score <10th percentile) cells (exponential curve fit: at >90th percentile, R^2=0·87; at <10th percentile, R^2=0·97). (I) Representative image of epithelial tissue sample showing regions of high-scoring RS senescent nuclei in red. (J) Mean RS scores of epithelial cells with increasing distance from senescent (score >90th percentile) and non-senescent (score <10th percentile) cells (exponential curve fit: at >90th percentile, R^2=0·89; at <10th percentile, R^2=0·90). (K) Representative image of TDLU sample showing regions of high-scoring RS senescent nuclei in red. (L) Mean RS scores of cells in TDLUs with increasing distance from senescent (score >90th percentile) and non-senescent (score <10th percentile) cells (exponential curve fit: at >90th percentile, R^2=0·86; at <10th percentile, R^2=0·14). (M) Mean convexity of nuclei in adipose tissue near high (>90th percentile) IR-scoring nuclei (exponential curve fit: R^2=0·83). AAD=antimycin A, atazanavir–ritonavir, and doxorubicin. IR=ionising radiation. RS=replicative senescence. TDLU=terminal duct lobular unit.

For model results by tissue compartment with significant associations with breast cancer, we did a cross-classification analysis looking at the risk for female donors with neither risk factor, with one of the risk factors, and with both risk factors. For this analysis, logistic regression models were used to calculate ORs with 95% CIs, comparing individuals with one or both risk factors to the reference group who had neither. These and all other models were adjusted for the following covariates: age at donation, race and ethnicity, parity, BMI, age at menarche, smoking history, alcohol intake, and family history of breast cancer, all of which have independent associations with breast cancer. Additional details are provided in the appendix (pp 3–4).

All analyses were performed in R (version 4.2.0).

Role of the funding source

The funders had no role in study design, data analysis, data collection, data interpretation, writing of the report, or decision to submit for publication.

Results

We applied the NUSP to H&E-stained healthy breast tissue biopsy images from 4382 KTB donors of various ages (median age 45 years [IQR 34–57]; figure 1A). As of data cutoff (July 31, 2022; median follow-up of 10 years [IQR 7–11]), 86 donors had developed breast cancer (mean time to cancer diagnosis 4·8 years [SD 2·84]). Baseline demographic and clinical characteristics of cases (donors diagnosed with breast cancer) and controls (no breast cancer diagnosis) are shown in the table.

Although the NUSP has been shown to apply to multiple cell types and image preparation methods,[1] it produces scores on different scales per tissue and imaging context. To reduce confounding factors and explore tissue-specific senescence, we segmented breast tissue into three primary tissue types (stromal, adipose, and epithelial, which was subcategorised into TDLU and non-TDLU epithelial). By automatically recognising tissue regions in each image of each biopsy sample, nuclei could be classified by the region where they were found. The NUSP was used to generate senescence scores using multiple models for each of the 32 857 482 nuclei recognised across images from the 4382 biopsy samples. To evaluate potential misclassification of segmented tissue types, we used UMAP to project several metrics, and found separation of clusters by tissue type, indicating that the segmentation method is robust (figure 1B).

Visualising nuclei with high predicted senescence (scores over the 90th percentile) revealed spatial distribution throughout the tissues (figure 1C). We found substantial clustering in epithelial regions and TDLUs, although this might be due to the higher density of nuclei and generally smaller spatial regions in the tissue. Previous work has suggested that senescent cells can induce senescence in neighbouring cells.[21] We evaluated the spatial distribution of senescence by examining senescence scores organised by distance buckets from low or high scoring nuclei (figure 1D). For all three models (IR, RS, and AAD) and the three main tissue types, nuclei were surrounded by other nuclei with similar scores, indicating clustering in which senescent nuclei were found near other high-scoring nuclei. As the distance from senescent nuclei increased, the mean senescence score generally decreased, indicating an increasing rate of non-senescent nuclei (figure 1E–I; appendix p 5). For most tissues and models, the clustering of high-scoring senescent nuclei and low-scoring nuclei suggests that causal factors induce senescence in groups of cells or that senescent cells propagate their state through contact or by SASP inflammatory factors, forming regions of increased senescence. Observing the steep decline in mean senescence with distance, we fit an inverse exponential curve,

	Controls (n=4296)	Cases (n=86)	p value
Age at tissue donation, years	44·6 (14·6)	53·1 (10·5)	<0·0001
Race and ethnicity			
Non-Hispanic White	2948 (68·6%)	65 (75·6%)	0·029
Non-Hispanic Black	767 (17·9%)	19 (22·1%)	..
Hispanic	371 (8·6%)	1 (1·2%)	..
Asian	162 (3·8%)	1 (1·2%)	..
Other or unknown*	48 (1·1%)	0	..
Parity†	1·3 (1·3)	1·6 (1·4)	0·078
Age at first birth, years	25·4 (5·7)	26·1 (5·7)	0·39
BMI, kg/m²	29·7 (7·7)	30·7 (6·8)	0·052
Age at menarche, years	12·5 (1·6)	12·3 (1·4)	0·53
Menopause			
Post-menopause	1630 (37·9%)	53 (61·6%)	<0·0001
Pre-menopause	2546 (59·3%)	26 (30·2%)	..
Uterine ablation	120 (2·8%)	7 (8·1%)	..
Smoking history			
Never smoker	3203 (74·6%)	60 (69·8%)	0·084
Past smoker	832 (19·4%)	24 (27·9%)	..
Current smoker	261 (6·1%)	2 (2·3%)	..
Current drinker			
No	1495 (34·8%)	28 (32·6%)	0·67
Yes	2801 (65·2%)	58 (67·4%)	..
Number of first-degree family members with a history of breast cancer			
0	3388 (78·9%)	62 (72·1%)	0·20
1	838 (19·5%)	22 (25·6%)	..
≥2	70 (1·6%)	2 (2·3%)	..
Number of previous biopsies			
0	3577 (83·3%)	64 (74·4%)	0·049
1	522 (12·1%)	14 (16·3%)	..
≥2	197 (4·6%)	8 (9·3%)	..
Gail 5-year risk score‡	1·47 (1·08)	1·85 (1·20)	0·0024

Data are mean (SD) or n (%). Controls had not been diagnosed with breast cancer and cases had been diagnosed with breast cancer as of data cutoff. *Including 15 unknown race or ethnicity, one Pacific Islander, 22 Native American, and 10 Native Hawaiian. †2715 of 4296 controls and 50 of 86 cases had a livebirth. ‡The Gail model uses key risk factors for breast cancer to estimate the probability of developing invasive breast cancer in the next 5 years.

***Table*: Donor characteristics at the time of tissue donation, by breast cancer diagnosis at data cutoff**

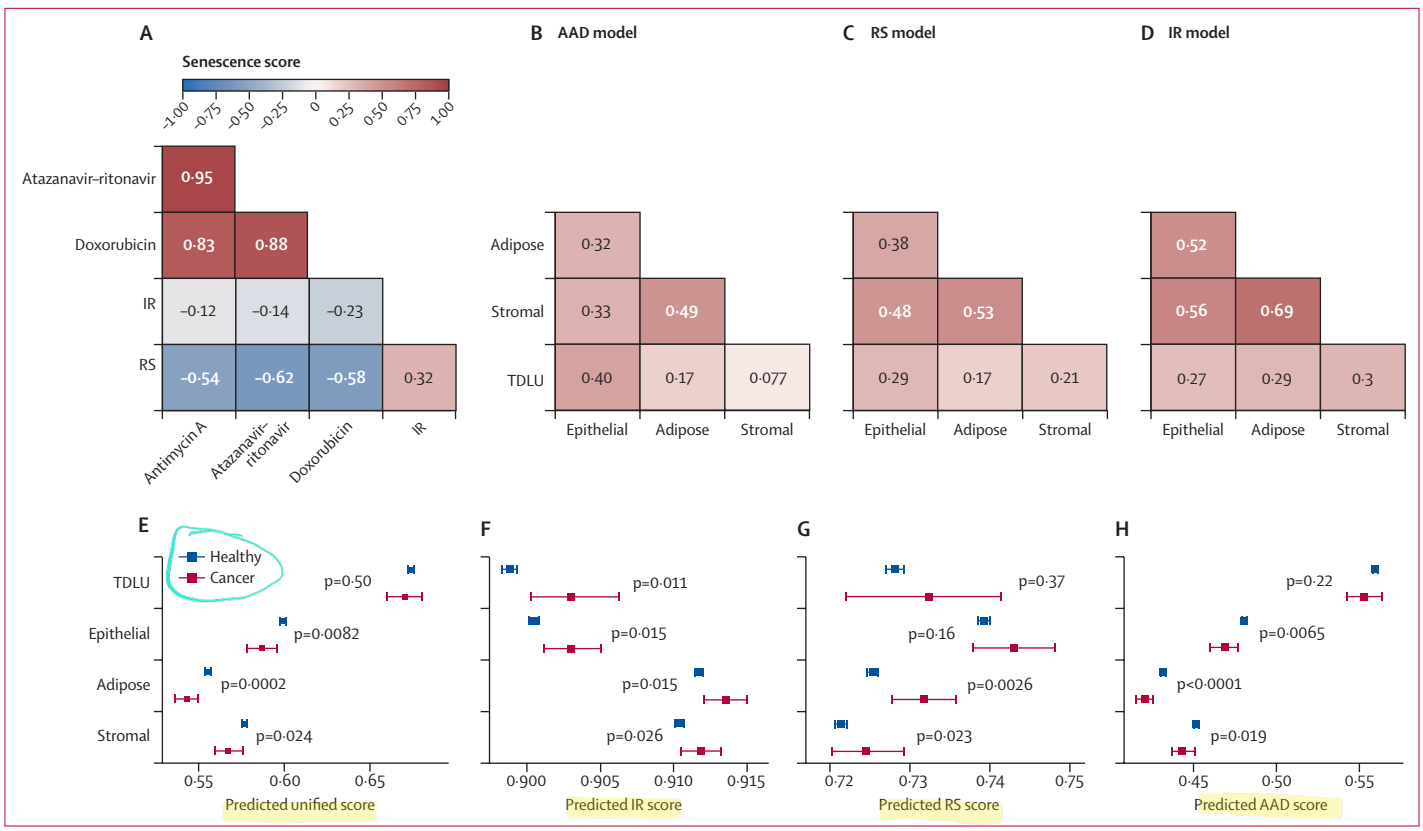

Figure 2: **Tissue classification and future cancer diagnosis**
(A) Correlation between senescence models per nuclei. (B) Correlation of senescence between tissues per individual for AAD model. (C) Correlation of senescence between tissues per individual for RS model. (D) Correlation of senescence between tissues per individual for IR model. Mean senescence scores for individuals per tissue for cases *vs* controls, using the unified (E), IR (F), RS (G), and AAD (H) models, with datapoints showing mean scores with errors bars showing 95% CIs. Multiple testing adjustments are presented in the appendix (p 6). AAD=antimycin A, atazanavir–ritonavir, and doxorubicin. IR=ionising radiation. RS=replicative senescence. TDLU=terminal duct lobular unit.

obtaining good fits with high R^2 values (figure 1F, H, J, L). The spatial distribution of senescence near high-scoring senescent nuclei resembles exponential decay, further supporting the hypothesis that paracrine SASP factors propagate senescence by local diffusion.

Considering that cell density could influence senescence and vice versa, we tested whether the mean count of nuclei differs in the vicinity of high-scoring senescent nuclei and low-scoring nuclei. The AAD model suggested increased density of cells in relatively close proximity to high-scoring nuclei, and the IR model showed higher counts for epithelial but lower counts for adipose cells, while the RS model did not show differences (appendix p 5). Although speculative, these results could indicate that the higher density of senescent nuclei near high senescent scoring nuclei than in other regions is related to an inflammatory response, where immune cells are recruited by factors in the SASP.

To better characterise predicted senescence, we investigated morphological metrics near high-scoring senescent nuclei. We evaluated convexity and aspect ratio, which were shown to indicate senescence (figure 1M; appendix p 8).[14] Indeed, we found substantially higher aspect and reduced convexity, both indicating senescence, in the cells closest to high-scoring senescent cells. These patterns were consistent for all tissue types and both the IR and RS models, indicating clustering of senescent cells.

NUSP provides five individual models trained on different senescence inducers (IR, RS, antimycin A, atazanavir-ritonavir, and doxorubicin). To better understand their predictions, we evaluated correlation between models per nuclei (figure 2A) and found high correlation between antimycin A, atazanavir–ritonavir, and doxorubicin, leading us to create the new combined model AAD. The three resulting models (IR, RS, and AAD) represent diverse inducers of senescence, and we interpret their correlation as common cell states (irreversible growth arrest) with different likelihoods of inducing a senescence associated secretory phenotype (SASP). We also found that IR and RS were both negatively correlated with AAD and only weakly correlated with each other (figure 2A). This finding led us to speculate that the models were capturing three relatively distinct phenotypes of senescence. We also evaluated

correlation between mean senescence scores per tissue type across individuals, and we found moderate correlation between several tissues—most notably, adipose and stroma—for all three models (figure 2B–D). We speculate that this finding might be related to individuals' systemic exposure or their response to stressors.

To investigate whether senescence relates to cancer formation, we looked for differences in the 86 cases who developed cancer compared with the 4296 controls in our cohort. Notably, future cancer cases appeared to have a smaller number of nuclei in all tissue types, with a small but significant difference in adipose, epithelial, and stromal tissue, and a larger difference in TDLU (appendix p 8). These findings could be related to the senescence response, with reduced immune recruitment, a reduced protective effect of senescence, or other subtle changes to tissue.

Because we observed spatial differences in predicted senescence for multiple tissue types, we investigated whether tissue-specific senescence could be a prognostic marker for breast cancer. Using the unified model, we found that the cases had mean predicted senescence scores significantly different from those of controls, with cases having lower mean predicted senescence in three tissue compartments, stroma, adipose, and non-TDLU epithelium (figure 2E). Senescence appears to have a protective effect in adipose, stroma, and non-TDLU epithelium, reducing the risk of developing cancer. Importantly, biopsies were taken a mean of 4·8 years (SD 2·84) before diagnosis, suggesting that the predicted senescence scores might be a crucial indicator of the risk for developing malignancy.

We also applied the senescence prediction models that were trained on several specific types of stress to evaluate whether different types of senescence can influence carcinogenesis. We observed significant differences in mean predicted senescence scores for all cell types with IR, for adipose and stromal tissue with RS, and for epithelial, adipose, and stromal tissue with AAD (figure 2F–H). We also applied adjustments to account for multiple testing with different models for each tissue, finding that all significant results maintained significance (appendix p 6). Although the AAD model shows a protective role for senescence, such that cases tended to have lower scores than controls, the IR and RS models indicated that cases tended to have higher senescence scores than did controls, indicating that different types of senescence might influence the likelihood of carcinogenesis differently.

Evaluating overall senescence per tissue per individual and adjusting for batch effects, we found a significant decrease in the odds of developing breast cancer for individuals with high AAD scores in adipose tissue (scores above the median) compared with individuals with low scores (below the median; OR 0·57 [95% CI 0·36–0·88]; p=0·013), indicating that participants with senescence scores below the median had significantly

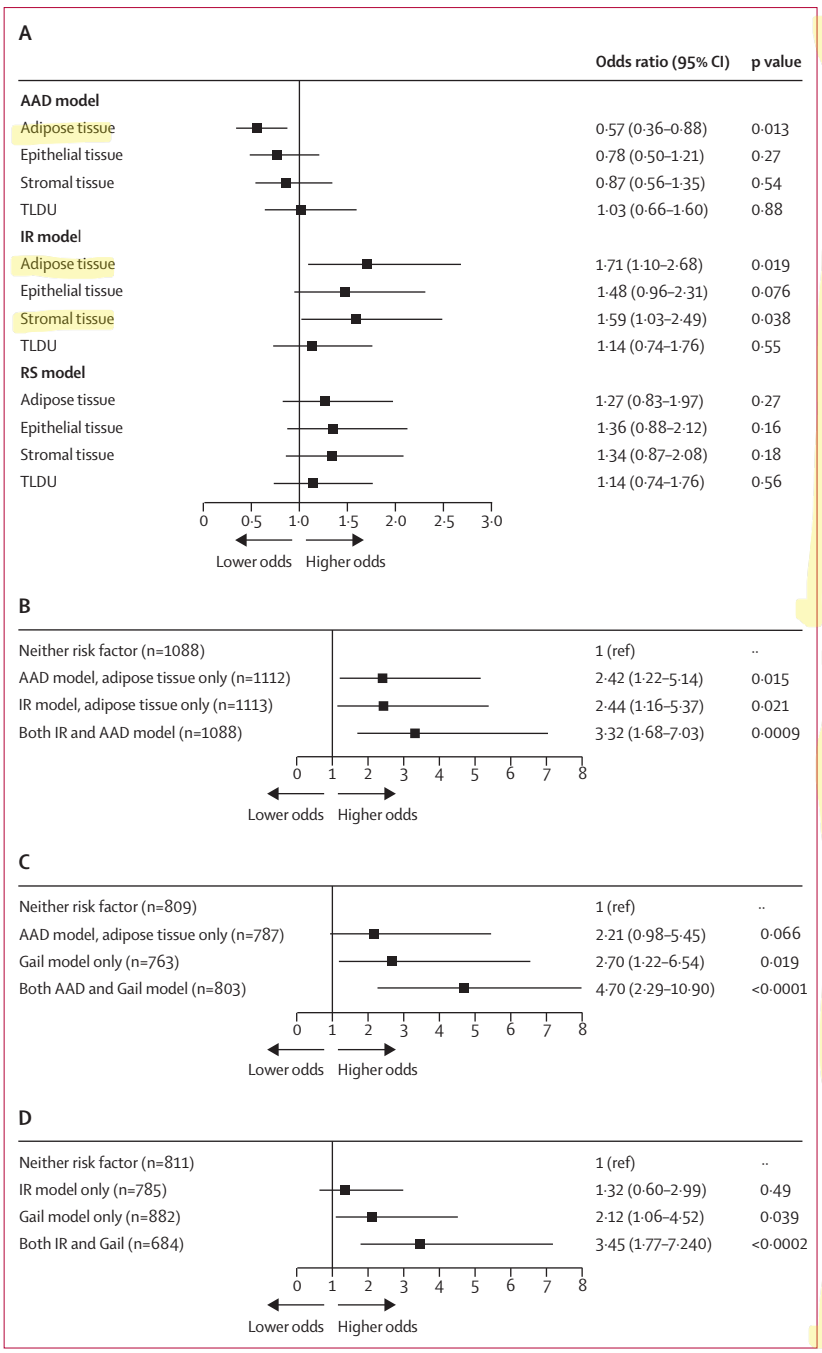

Figure 3: **Assessment of different forms of senescence to predict future odds of breast cancer**
(A) Estimated odds ratios for donors with senescence scores above the median (ie, upper half of scores) compared with those below the median (ie, lower half of scores) for different models and tissue types; donors with median scores were excluded. (B) Cross-classification analysis showing odds ratios for donors with senescence scores above versus below the median when combining the IR and AAD models for adipose tissue; excluding donors with median scores. (C) Cross-classification analysis showing odds ratios for donors with senescence scores above versus below the median when combining AAD for adipose tissue with 5-year Gail scores; excluding donors with median scores, and Gail models were only calculated for donors aged 35 years and older. (D) Cross-classification analysis showing odds ratios for donors with senescence scores above versus below the median when combining IR for adipose tissue with 5-year Gail scores; excluding donors with median scores, and Gail models were only calculated for donors aged 35 years and older. Multiple testing adjustments are presented in the appendix (p 6). AAD=antimycin A, atazanavir–ritonavir, and doxorubicin. IR=ionising radiation. RS=replicative senescence. TDLU=terminal duct lobular unit.

higher odds of developing breast cancer than those above the median (figure 3A). Conversely, with the IR model, we found significantly increased odds of cancer development for individuals with high predicted senescence scores in adipose tissue (OR 1·71 [95% CI 1·10–2·68]; p=0·019) and in stromal tissue (1·59 [1·03–2·49]; p=0·038). All other model and tissue results were not significant (figure 3A). The results were adjusted for multiple testing with different models, and the differences remained significant for the AAD and IR models for adipose tissue (appendix p 6). We also evaluated odds of breast cancer by quartiles, although this analysis was underpowered because of the low sample size (86 cases; appendix p 7).

Cross-classification analyses suggest that the odds of breast cancer indicated by each model were largely independent (figure 3B). Individuals with both low AAD model scores in adipose tissue and high IR model scores in adipose tissue had an OR of 3·32 (95% CI 1·68–7·03; p=0·0009; figure 3B), suggesting that multiple forms of senescence relate to odds of breast cancer differently, with the AAD model capturing a protective effect and the IR model being associated with increased risk. We performed a similar analysis using individuals' 5-year Gail scores, comparing those with scores above the median to those below, which gave an OR of 2·33 (1·46–3·82; p=0·0012). Remarkably, individuals with both high Gail scores and low AAD model scores in adipose tissue had an OR of 4·70 (2·29–10·90; p<0·0001; figure 3C), while individuals with both high Gail scores and high IR model scores in adipose tissue had an OR of 3·45 (1·77–7·24; p=0·0002; figure 3D). Importantly, these results indicate little overlap in the risk derived from the different predicted senescence models and the Gail score.

Discussion

In this study, we applied deep learning models to predict senescence per nucleus in images of H&E-stained biopsy samples of breast tissue from 4382 healthy donors. Compared with individuals who did not develop breast cancer as of data cutoff, the NUSP identified significant differences in the mean senescence score for the 86 individuals in our cohort who were later diagnosed with breast cancer at a mean interval of 4·8 years after their non-malignant breast biopsy. Our risk analysis suggested that individuals with predicted senescence scores above the median (ie, upper half of scores) for adipose tissue using the AAD model had reduced odds of developing cancer (OR 0·57 [95% CI 0·36–0·88]; p=0·013), whereas individuals with senescence scores above the median for the IR model in adipose tissue had increased odds (OR 1·71 [1·10–2·68]; p=0·019). These results suggest that some forms of cellular senescence in specific tissue types can have a protective effect that can substantially reduce cancer risks, while other forms of senescence might promote cancer development. Individuals with both risk factors had an OR for breast cancer of 3·32 ([1·68–7·03]; p=0·0009), which exceeds the OR determined using the clinical gold standard in the field, the Gail model, which had an OR of 2·33 for those with scores above the median versus below the median. Combining a single senescence risk factor (AAD model in adipose tissue) with the Gail model resulted in an OR of 4·70 ([2·29–10·90]; p<0·0001), which was considerably higher than that predicted with the Gail score alone. The ability to predict risk could probably be further improved by combining multiple senescence models with Gail scores to enhance the prediction of breast cancer risk. Our findings suggest dual roles for senescence in maintaining tissue health while also being associated with malignancy, and we introduce new senescence-based models that can be applied directly to H&E-stained images.

The clinical relevance of these findings depends on reliable detection by biopsy. We found evidence that the extent of markers of different forms of senescence relates to an individual's risk of future cancer. Whether the extent of senescence reflects an individual's propensity to undergo senescence or their systemic exposure to senescence-inducing factors is unclear. Our results show that predicted senescence of tissues moderately correlates per individual, suggesting that the scope at least spans multiple tissue types in a biopsy sample. Although individual propensity and the scope of effects in the body are not well understood, the extent of senescence can be related to age[22] and sex,[23] and might be influenced by diverse environmental factors such as smoking,[24] sleep deprivation,[25] and stress.[26] Ultimately, regardless of the cause, our data suggest that differences in predicted senescence, based on nuclear morphology, is a clinically significant risk factor in the development of breast cancer.

The precise role of senescence in cancer is complex and not fully understood; it might have causative effects or simply correlate with underlying processes involved in cancer development, or both.[27,28] NUSP models trained on various senescence inducers indicated that different forms of senescence can promote or protect against cancer. The IR model, trained on senescent cells with severe DNA damage induced by radiation, showed higher senescence in the stromal and adipose tissues of those who later developed breast cancer than in those who did not develop cancer, whereas the AAD model showed a protective effect of senescence in adipose tissue. The positive association between the IR model and breast cancer could be due to the effect of the pro-inflammatory SASP, which has been shown to be particularly prominent in IR-induced senescent cells.[29] Senescence arising from substantial DNA damage and captured by the IR model could also be related to a higher rate of mutations that could contribute to malignant transformation. Conversely, the AAD model was trained on images of cells that had senescence induced by diverse drug treatments, with antimycin A inducing

mitochondrial dysfunction and atazanavir–ritonavir inducing proteotoxic stress. Increased rates of these types of senescence could provide a protective effect by controlling proliferation without mutational load and SASP effects. The dual roles of senescence in promoting and controlling cancer are widely recognised,[30,31] and our study provides evidence for both roles in breast cancer using high-precision machine learning methods to analyse a large cohort. We speculate that the IR model captures a stronger senescence phenotype with a more developed SASP, whereas the AAD model, having been trained on diverse drug treatments, might reflect a weaker or earlier senescence phenotype with a reduced SASP.[32,33] Our observations might indicate that differences in the senescence subtype can affect cancer initiation, possibly due to differences in the SASP (appendix p 5); however, further research, including spatial omics, would be needed to answer these questions in greater detail.

The NUSP models have been shown previously to match diverse senescence markers (p16, p21, p53, β-galactosidase staining, and ethynyl-2□-deoxyuridine incorporation),[14] but they are ultimately based on nuclear morphology. Despite the validation of NUSP, predicted senescence metrics might directly reflect not senescence but other closely related biological processes, such as nuclear changes related to DNA damage, and those underlying processes might be related to cancer development. The low correlation between scores from the IR and RS models per nucleus and negative correlation between those scores and scores from the AAD model supports the hypothesis that these models are detecting different forms of senescence.

The use of biopsy sample images from more than 4000 female donors to the KTB with associated covariates was a strength of the study, because all female donors in the KTB had biopsies at the time of enrolment, and the development of breast cancer occurred after biopsy in all study participants. The single-centre study design, specific race and ethnicity mix, and the inclusion of only female participants who were willing to donate healthy breast tissue are study characteristics that might limit generalisability. Furthermore, although the effect size was large enough to find statistical significance in breast cancer cases as a whole, the low number of cases did prevent subcategorisation by breast cancer type or hormone receptor status. Selection bias was also a possible limitation because not all women are willing to undergo breast biopsy on a voluntary basis, and the race and ethnicity of the KTB participants might limit generalisability to the overall population or to under-represented subgroups. Despite including all available covariates with documented associations with breast cancer (eg, age, ethnicity and race, age at menarche, BMI, family history of breast cancer, history of smoking and alcohol consumption), the risk of uncontrolled confounding still exists that could have biased the results in unknown directions.

Our spatial analysis also uncovered several notable aspects of cellular senescence in breast tissues. The SASP induces changes through paracrine factors and, for all three major tissue types, senescent nuclei were generally found near similar nuclei, and the proportion of senescent nuclei declined with distance, suggesting clusters of senescence. We also found that senescent epithelial (including TDLU) nuclei exist in regions of higher density, which might suggest that higher density induces senescence or that these regions of higher density have increased susceptibility to senescence. Comparing the density between cases and controls revealed a lower density for cases across all three tissue compartments, as well as TDLUs, supporting the hypothesis that, at least locally, senescence can be a barrier to cancer.

In summary, the extent of senescence predicted from nuclear morphology in images of H&E-stained healthy breast tissue is a substantial predictor of breast cancer development. This association was found for individuals overall and per tissue. In our cohort, the combination of multiple models improved risk prediction compared with the current clinical benchmark, the Gail score. Notably, our deep learning senescence models predicted risk of breast cancer orthogonally to the Gail model and when the AAD adipose model was added to the Gail model, participants with both risk factors had an OR of 4·70 versus an OR of 2·33 in those with just the Gail risk factor. High-precision predicted senescence scores provide new insights into breast cancer risk and have the potential to be applied in the clinical assessment and screening of individuals with non-malignant breast biopsies that are not otherwise useful and might be potentially misleading with respect to the assessment of future breast cancer risk. The results from improved risk assessment could lead to more targeted screening, thus reducing unnecessary testing in those found to be at low risk, with reduced costs and anxiety for these individuals. Those found to be at increased risk could have more aggressive testing and follow-up, potentially leading to more cases being diagnosed at an earlier stage, and decreased morbidity and mortality.

Contributors

IH wrote the manuscript, applied the deep learning models, and analysed the data. MP conceived, designed, and executed the study, supervised the analysis, and edited the manuscript. SF analysed the KTB data. JH, SR, and JC provided access to necessary KTB images and covariates. AAT managed clinical samples and edited the manuscript. EV advised the project on senescence. CB contributed to the KTB collaboration, consulted on the study results, and edited the manuscript. MS-K supervised the project and edited the manuscript. MP and MS-K accessed and verified the data used for this study. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of interests

IH, MP, and MS-K have filed a provisional patent related to risk prediction methods. All other authors declare no competing interests.

Data sharing

Histology images and clinical data are available from the KTB with an appropriate data sharing agreement. Source code for the senescence

For the **source code for senescence predictor** see https://github.com/scheibye-knudsen-lab/sen_predictor

For the **source code for additional methods** see https://github.com/scheibye-knudsen-lab/cancer_sen

predictor was made available online under our previous publication[1] and source code for additional methods is available online.

Acknowledgments

MS-K was supported by the Novo Nordisk Foundation Challenge Programme (NNF17OC0027812) and the Danish Cancer Society (grant numbers R368-A21521, R302-A17379_001, and R167-A11015_001). CB was supported by the US National Institutes of Health (grant U54 AG075932). We thank contributors, including those at Indiana University, who collected data used in this study, as well as donors and their families, whose help and participation made this work possible.

References

1 Giaquinto AN, Sung H, Miller KD, et al. Breast cancer statistics, 2022. *CA Cancer J Clin* 2022; **72:** 524–41.
2 Kamińska M, Ciszewski T, Łopacka-Szatan K, Miotła P, Starosławska E. Breast cancer risk factors. *Prz Menopauzalny* 2015; **14:** 196–202.
3 Benz CC. Impact of aging on the biology of breast cancer. *Crit Rev Oncol Hematol* 2008; **66:** 65–74.
4 Pare R, Yang T, Shin J-S, Lee CS. The significance of the senescence pathway in breast cancer progression. *J Clin Pathol* 2013; **66:** 491–95.
5 Shen J, Song R, Fuemmeler BF, McGuire KP, Chow W-H, Zhao H. Biological aging marker $p16^{INK4a}$ in T cells and breast cancer risk. *Cancers (Basel)* 2020; **12:** 3122.
6 González-Gualda E, Baker AG, Fruk L, Muñoz-Espín D. A guide to assessing cellular senescence in vitro and in vivo. *FEBS J* 2021; **288:** 56–80.
7 Hayflick L, Moorhead PS. The serial cultivation of human diploid cell strains. *Exp Cell Res* 1961; **25:** 585–621.
8 Serrano M, Lin AW, McCurrach ME, Beach D, Lowe SW. Oncogenic *ras* provokes premature cell senescence associated with accumulation of p53 and $p16^{INK4a}$. *Cell* 1997; **88:** 593–602.
9 Campisi J, Andersen JK, Kapahi P, Melov S. Cellular senescence: a link between cancer and age-related degenerative disease? *Semin Cancer Biol* 2011; **21:** 354–59.
10 Sharpless NE, Sherr CJ. Forging a signature of in vivo senescence. *Nat Rev Cancer* 2015; **15:** 397–408.
11 SenNet Consotium. NIH SenNet Consortium to map senescent cells throughout the human lifespan to understand physiological health. *Nat Aging* 2022; **2:** 1090–100.
12 Gurkar AU, Gerencser AA, Mora AL, et al. Spatial mapping of cellular senescence: emerging challenges and opportunities. *Nat Aging* 2023; **3:** 776–90.
13 Matias I, Diniz LP, Damico IV, et al. Loss of lamin-B1 and defective nuclear morphology are hallmarks of astrocyte senescence in vitro and in the aging human hippocampus. *Aging Cell* 2022; **21:** e13521.
14 Heckenbach I, Mkrtchyan GV, Ezra MB, et al. Nuclear morphology is a deep learning biomarker of cellular senescence. *Nat Aging* 2022; **2:** 742–55.
15 Sherman ME, Figueroa JD, Henry JE, Clare SE, Rufenbarger C, Storniolo AM. The Susan G. Komen for the Cure Tissue Bank at the IU Simon Cancer Center: a unique resource for defining the "molecular histology" of the breast. *Cancer Prev Res (Phila)* 2012; **5:** 528–35.
16 The Komen Tissue Bank at the IU Simon Cancer Center. Standard operating procedure (SOP) 001V8.0: acquisition of normal breast tissue and blood at a tissue collection event. https://cancer.iu.edu/pdf/ktb/research/sop-001v8.0-acquisition-of-normal-breast-tissue-at-a-tissue-collection-event.pdf (accessed Jan 17, 2023).
17 The Komen Tissue Bank at the IU Simon Cancer Center. Standard operating procedure (SOP) 005V7.0: processing and storage of breast tissue. https://cancer.iu.edu/pdf/ktb/research/sop-005v7.0-processing-and-storage-of-breast-tissue.pdf (accessed Jan 17, 2023).
18 Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, eds. Medical image computing and computer-assisted intervention—MICCAI 2015. Cham: Springer International Publishing, 2015: 234–41.
19 Tabár L, Dean PB, Tucker FL, et al. Breast cancers originating from the terminal ductal lobular units: in situ and invasive acinar adenocarcinoma of the breast, AAB. *Eur J Radiol* 2022; **152:** 110323.
20 Wetstein SC, Onken AM, Luffman C, et al. Deep learning assessment of breast terminal duct lobular unit involution: towards automated prediction of breast cancer risk. *PLoS One* 2020; **15:** e0231653.
21 Nelson G, Wordsworth J, Wang C, et al. A senescent cell bystander effect: senescence-induced senescence. *Aging Cell* 2012; **11:** 345–49.
22 Idda ML, McClusky WG, Lodde V, et al. Survey of senescent cell markers with age in human tissues. *Aging (Albany NY)* 2020; **12:** 4052–66.
23 Ng M, Hazrati L-N. Evidence of sex differences in cellular senescence. *Neurobiol Aging* 2022; **120:** 88–104.
24 Astuti Y, Wardhana A, Watkins J, Wulaningsih W. Cigarette smoking and telomere length: a systematic review of 84 studies and meta-analysis. *Environ Res* 2017; **158:** 480–89.
25 Timonina D, Hormazabal GV, Heckenbach I, et al. Chronically disrupted sleep induces senescence in the visceral adipose tissue of C57BL/6 mice. *bioRxiv* 2023; published online Oct 19. https://doi.org/10.1101/2023.10.17.562803 (preprint).
26 Yegorov YE, Poznyak AV, Nikiforov NG, Sobenin IA, Orekhov AN. The link between chronic stress and accelerated aging. *Biomedicines* 2020; **8:** 198.
27 Wyld L, Bellantuono I, Tchkonia T, et al. Senescence and cancer: a review of clinical implications of senescence and senotherapies. *Cancers (Basel)* 2020; **12:** 2134.
28 Domen A, Deben C, Verswyvel J, et al. Cellular senescence in cancer: clinical detection and prognostic implications. *J Exp Clin Cancer Res* 2022; **41:** 360.
29 Rodier F, Coppé J-P, Patil CK, et al. Persistent DNA damage signalling triggers senescence-associated inflammatory cytokine secretion. *Nat Cell Biol* 2009; **11:** 973–79.
30 Coppé J-P, Desprez P-Y, Krtolica A, Campisi J. The senescence-associated secretory phenotype: the dark side of tumor suppression. *Annu Rev Pathol* 2010; **5:** 99–118.
31 Davalos AR, Coppe J-P, Campisi J, Desprez P-Y. Senescent cells as a source of inflammatory factors for tumor progression. *Cancer Metastasis Rev* 2010; **29:** 273–83.
32 Heckenbach I, Scheibye-Knudsen M. Tracking the dynamics of cellular senescence. *Aging (Albany, NY)* 2023; **15:** 3219–20.
33 Basisty N, Kale A, Jeon OH, et al. A proteomic atlas of senescence-associated secretomes for aging biomarker development. *PLoS Biol* 2020; **18:** e3000599.