nature aging

Technical Report

https://doi.org/10.1038/s43587-025-00886-2

A transcriptome-based human universal senescence index (hUSI) robustly predicts cellular senescence under various conditions

Received: 2 February 2024

Accepted: 28 April 2025

Published online: 29 May 2025

Check for updates

Jing Wang[1,7], Xiaolan Zhou[1,7], Peng Yu[1,3,7], Jun Yao[1,4], Pengfei Guo[2], Qiushi Xu[1], Yuqi Zhao[2], Guanlong Wang[1], Qianru Li[1], Xiaofeng Zhu[1], Gang Wei[1] ✉, Weixu Wang[1,5,6] ✉ & Ting Ni[1,2] ✉

Despite the manifestation and contribution of cellular senescence to aging and various diseases, accurate identification of heterogeneous senescent cells remains challenging. Current senescence evaluation methods rely mainly on limited markers or homogeneous samples, which might fail to capture universal senescence features, limiting their generalizability. Here we developed the human universal senescence index (hUSI), an accurate and robust senescence evaluation method for diverse cells and samples. Based on features learned from the most comprehensive cellular senescence-associated transcriptome data so far, hUSI demonstrated its convincing connections with senescence phenotypes and superior robustness in predicting senescence state. Using hUSI, we discovered potential senescence regulators and mapped senescent cell accumulation across cell types in coronavirus disease 2019 (COVID-19). The method also facilitates decoding heterogeneous senescence states in melanoma tumors, identifying prognosis-associated signaling pathways. Overall, hUSI demonstrates its utility in characterizing cellular senescence across biological contexts, with broad applications in aging research and clinical practice.

Cellular senescence (CS) characterized by irreversible cell cycle arrest is considered a critical factor for aging and age-related diseases[1]. For instance, by presenting senescence-associated secretory phenotypes (SASPs) including increased secretion of pro-inflammatory proteins and other paracrine factors (such as transforming growth factor β (TGFβ) family ligands, vascular endothelial growth factor (VEGF) and chemokines)[2], senescent cells can stimulate immune response and cell–cell communication, leading to pleiotropic effects in various tissues[3]. Targeted clearance of accumulated senescent cells using senolytic drugs[4] or inducing tumor cells into senescence[5] has demonstrated benefits for disease prognosis and healthy lifespan. Depending on the real situations, CS in diverse cell types can be induced by various intrinsic and extrinsic stressors, such as replicative stress, oncogene activation, chemotherapeutic drugs and ionizing radiation[6]. However, despite the use of several morphological markers (such as flattening and enlarging[7]) and molecular markers (such as p16 (ref. 8), p21 (ref. 9), p15 (ref. 10) and p27 (ref. 10)) to characterize senescent cells, identification of in vivo senescent cells still remains a great challenge due to their heterogeneity[11]. Therefore, there is an urgent need for a universal method that can accurately predict the senescence state in multiple scenarios, especially in the era where single-cell transcriptome technology has been widely applied[12,13].

Developing a senescence scoring method based on transcriptome data is an efficient way to monitor senescence state in disease progression[14]. However, the relative scarcity of reliable senescence markers keeps limiting distinguishing and targeting senescent cells both in vitro

A full list of affiliations appears at the end of the paper. ✉e-mail: gwei@fudan.edu.cn; weixu.wang@helmholtz-munich.de; tingni@fudan.edu.cn

and in vivo[1,11,15,16]. For example, in single-cell studies, the differential expression level of classical senescence markers is often insufficient to identify senescent cells due to the heterogeneity of both cell types and senescence types[17]. Considering that there is no one-size-fits-all marker for CS, several studies evaluated senescence by compiling a series of senescence-associated genes derived from differential analysis or literature studies, giving rise to several gene sets (such as CellAge[18] and SenMayo[19]) and scoring methods (such as *DAS_MSS*[14] and *CS_score*[20]). However, these gene sets and methods were usually developed based on genes associated with certain aspects of senescence, such as activated secretory phenotype or replicative-related changes[21]. Notably, single-sample gene set enrichment analysis (ssGSEA) is widely used as the quantification means for gene-set-based methods, although performing ssGSEA for each cell in large-scale single-cell atlases is extremely time-consuming, due to the substantial number of permutations required to accurately estimate the nominal *P* value[22].

Recently, to evaluate senescence in a more unbiased and convenient manner, machine learning methods were used to learn senescence features and predict senescent state, resulting in a higher accuracy in identifying senescent cells, such as *lassoCS*[23], *SENCAN*[24] and *SenCID*[25]. However, these methods, developed from limited senescence samples or pre-defined cell-type-specific senescence signatures, might not capture universal senescence features across diverse samples. For instance, methods relying solely on tumor samples might introduce bias in evaluating senescence in normal senescence samples or real aging tissues. In addition, despite the increase of senescence transcriptome data, identifying universal senescence features remained a challenge due to the inherent heterogeneity present in both mechanisms and technologies[26]. Most machine-learning-based senescence evaluation tools are derived from binary classification models, which face the challenge of a shortage of high-quality and consistent negative (non-senescent) samples in real-world senescence-related studies. For example, some senescence signals (such as senescence-associated β-galactosidase (SA-β-gal) and $p16^{INK4a}$) also present in non-senescent cells[27], and cells at different conditions (such as untreated, mature, proliferation or quiescence), all have been selected as non-senescent samples. Meanwhile, binary classification models primarily focus on classifying samples as positive or negative, which may bias the learning toward features prevalent in non-senescent samples, impacting the generalization ability of these models[28].

To overcome the limitations mentioned above, we curated a comprehensive senescence transcriptome profile from publicly available RNA sequencing (RNA-seq) data and employed the One-Class Logistic Regression (OCLR) model to learn universal senescence features due to its robustness in negative samples and mixed signals[29]. Ultimately, we successfully developed a reliable data-driven senescence scoring method that we termed the human universal senescence index (hUSI). We conducted benchmarking of hUSI against 31 collected methods in both bulk and single-cell transcriptome samples under various conditions, confirming the outperformance of hUSI. Furthermore, we validated the senescence repressors identified by hUSI and confirmed the usability of hUSI in complex pathological tissues (coronavirus disease 2019 (COVID-19) lung tissue and melanoma tissue). Overall, we demonstrated that hUSI is a robust method for evaluating senescence state, enabling more comprehensive investigations into senescence in various experimental and clinical contexts.

Results

Overview of hUSI

To develop hUSI, we devised a workflow comprising data collection, feature learning, hUSI computation and application (Fig. 1). Initially, we collected CS-related RNA-seq data from 73 studies[24,30–98] to construct, to our knowledge, the most comprehensive human senescence transcriptome dataset so far (Supplementary Table 1). Given the batch effect in experimental methods and sequencing protocols of the collected data, we standardized the pre-processing of all raw data using a uniform pipeline. Rigorous quality control was then implemented to generate a high-quality senescence gene expression profile, which included 770 samples (385 senescent samples and 385 non-senescent samples) across 34 cell types and 13 senescence types from 64 studies (Supplementary Table 2). Through applying the OCLR model to this profile, we obtained a gene weights vector and confirmed its representation for CS features (Supplementary Table 4). In the final stage of our method, we used the Spearman correlation between the OCLR-learned senescence features and gene expression profiles to assess the senescence states of bulk or single-cell samples. The resulting scores were scaled and denoted as hUSI. To evaluate the CS predictive performance of hUSI, we conducted a comprehensive benchmarking against 31 CS scoring methods, including canonical CS-related markers, CS-related gene sets, traditional CS scores and machine-learning-based CS scores. Upon confirming the superior performance of hUSI, we then explored the potential of hUSI for multiple downstream applications. We discerned the accumulation changes of senescent cells between normal and pathological samples and characterized senescent cells and their interactions with neighboring cells, their relationships with clinical traits and potential genes whose perturbation could induce a cell into senescent state.

Development and validation of hUSI

After quality control processing, we also confirmed the significantly elevated expression of senescence markers *CDKN1A* (National Center for Biotechnology Information (NCBI) Gene: 1026, encodes p21) and *CDKN2B* (NCBI Gene: 1030, encodes p15) in senescent samples (Extended Data Fig. 1a,b). Then, to reduce heterogeneity attributable to varying cell types, the top 85 cell-type-specific genes were removed using the adjusted Rand index (ARI) from the senescence gene expression profile (Extended Data Fig. 1c–e). The hUSI development strategy was optimized via 10-fold cross-validation out-of-distribution (OOD) testing[99], and the final OCLR model (L2 regularization = 1, Spearman correlation) achieved superior performance (area under the curve (AUC) > 0.97) (Extended Data Figs. 1f–h and 2a). In addition, hUSI demonstrated robustness against several data quality problems (including batch effects, sparsity and outlier) in simulations using five independent RNA-seq datasets[81,100–102] (AUC > 0.95) (Extended Data Fig. 2b–d and Supplementary Table 3). Furthermore, compared to widely used binary classification models (including random forest (RF), elastic net (EN) and support vector machine (SVM)), OCLR exhibited higher accuracy and stability across cell types in OOD testing (Extended Data Fig. 2e–g). Even binarization via sum-of-square error (SSE), hUSI maintained a low false-positive rate (0.07) and a high AUC (0.90) (Extended Data Fig. 2h), which are crucial for model performance. All these results underscore that hUSI, which is based on comprehensive senescence transcriptome profile, OCLR model[29] and non-parametric rank-based correlation[103], is stable and robust as a CS prediction tool.

Next, we investigated the biological interpretability of hUSI by examining the weight distribution of genes learned by the OCLR model (Supplementary Table 4). We found that classical senescence-associated markers and SASP genes ranked in the top 10% of weights, whereas canonical cell cycle markers were in the lowest 10% (Extended Data Fig. 2i). Through comparing enrichment levels of OCLR-learned features among 268 cell types from the CellMarker database[104] and a widely used senescence-associated gene set (SenUP)[105] by GSEA, we observed a more significant positive enrichment for SenUP compared to cell type marker sets, indicating that the learned features prefer CS signatures to cell type signatures (Extended Data Fig. 2j). Subsequently, we conducted GSEA on OCLR-learned features across multiple databases, including Hallmark[106], Gene Ontology (GO)[107,108], Reactome[109] and Kyoto Encyclopedia of Genes and Genomes (KEGG)[110] (Extended Data Fig. 2k). We found that, taking Hallmark as an example, multiple senescence-associated gene sets were positively enriched, including interferon-gamma response[111], KRAS

https://doi.org/10.1038/s43587-025-00886-2

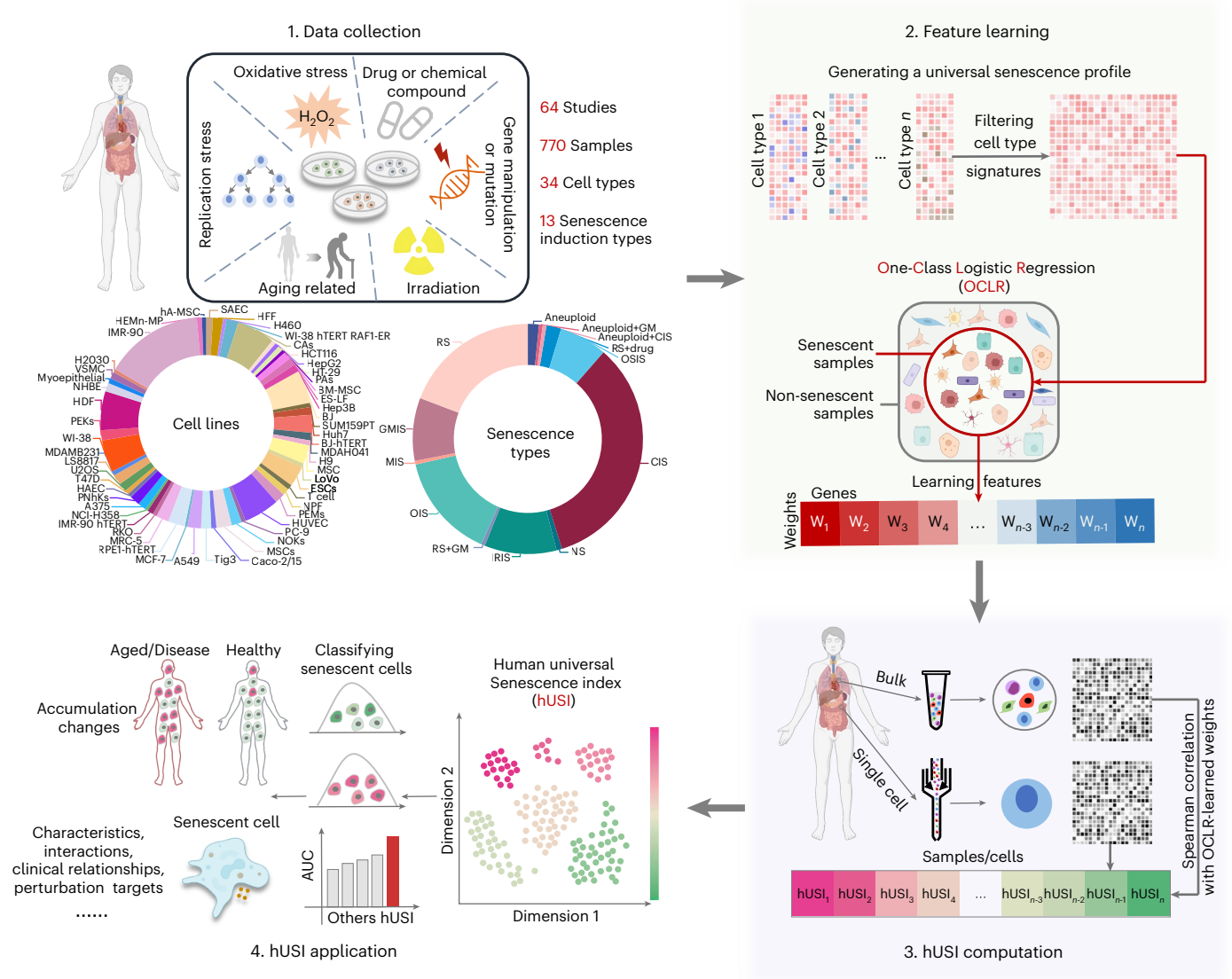


Fig. 1 | Overall workflow of hUSI. The workflow of hUSI encompasses four major steps: data collection, feature learning, hUSI computation and hUSI application. 1. Data collection: We collected bulk RNA-seq data (843 samples) from 73 studies. After quality control, 770 samples (385 senescent and 385 non-senescent) from 64 studies were retained across 34 distinct cell types and 13 senescence types. The distribution of cell lines and senescence types is shown in pie charts (lower left and lower right, respectively). 2. Feature learning: A uniform RNA-seq data processing pipeline was employed to create a standardized gene expression matrix. We selected senescent samples and filtered cell type signatures to generate a universal senescence transcriptome profile. Then, an OCLR model was applied to this profile and ultimately produced a weighted gene vector, referred to as the OCLR-learned senescence features. 3. hUSI computation: hUSI can be calculated for both bulk and single-cell samples by calculating the Spearman correlation coefficient between the OCLR-learned senescence features and gene expression value. 4. hUSI application: The effectiveness of hUSI was validated in distinguishing senescent samples across multiple bulk and single-cell datasets, demonstrating superior performance compared to other methods. Applying hUSI to various datasets can elucidate characteristics, interactions, clinical relationships and regulators of senescent cells as well as their accumulative changes in diseases. Figure created with BioRender.

signaling[112], inflammatory response[113] and p53 pathway[114,115] (Extended Data Fig. 2l). In contrast, proliferation-associated pathways (such as G2M checkpoint[116], E2F targets[117], mitotic spindles[118] and MYC targets[119]) were in the negative enrichment terms (Extended Data Fig. 2m). These results support that the OCLR-learned features are representative in reflecting senescence (Supplementary Table 5).

hUSI robustly predicts senescence across diverse samples

We then conducted a systematic benchmarking of hUSI and published senescence evaluation tools. Multiple bulk and single-cell gene expression datasets derived from in vitro or in vivo sources were used for the comparison (Supplementary Table 3). We included 31 methods categorized into four types: 13 classical gene markers for senescent cells (referred to as SenMarker), eight widely used senescence-associated gene sets (referred to as SenSet)[18,19,21,105,120–122], four senescence scoring methods based on empirical senescence markers and traditional calculation approaches (referred to as TraditionScore)[14,20] and six senescence scoring methods developed through various machine learning models (referred to as MachineScore)[20,23–25,123] (Supplementary Tables 6 and 7).

To evaluate the performance of hUSI in bulk samples, we compared hUSI with other methods across six in vitro bulk microarray datasets[124–129], encompassing six cell types and three senescence types (Supplementary Table 3). The result shows that hUSI is the only method capable of accurately identifying senescent samples across all datasets (AUC = 1.0) (Fig. 2a and Supplementary Table 8). Subsequently, we applied hUSI to in vivo bulk samples from the Genotype-Tissue Expression (GTEx)[130] and The Cancer Genome Atlas (TCGA) databases. Consistent with age-related accumulation of cellular senescence[131,132],

https://doi.org/10.1038/s43587-025-00886-2

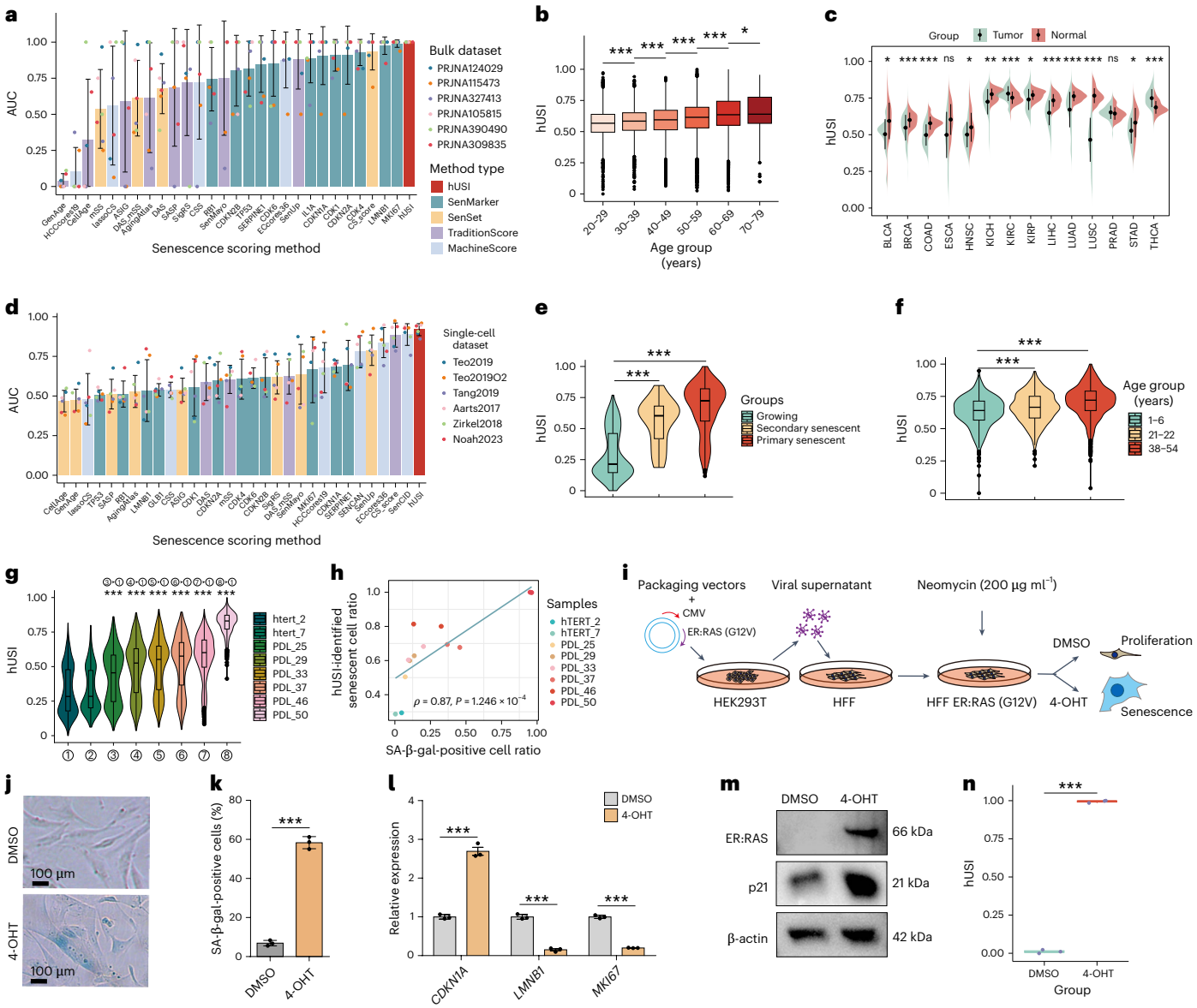


Fig. 2 | hUSI enables robustly evaluating senescence for diverse samples. **a**, Mean AUC values across six microarray datasets, with sample sizes detailed in Supplementary Table 3, presented as mean ± s.d. **b**, hUSI values in GTEx samples grouped by age (n = 17,382 samples; two-tailed Wilcoxon test). **c**, hUSI comparison between tumor and matched normal tissues (n = 1,291 samples; two-tailed t-test). **d**, Similar to **a**, benchmarking results of all methods in six scRNA-seq datasets (the number of cells is provided in Supplementary Table 3). **e**, hUSI levels of IMR90 cells were compared across three groups, including growing (untreated), primary senescent (4-OHT-induced) and secondary senescent (SASP-mediated) (n = 480 cells; two-tailed Wilcoxon test). **f**, hUSI levels of pancreas cells from donors aged 1–6, 21–22 and 38–54 years (n = 2,544 cells; two-tailed Wilcoxon test). **g**, hUSI levels in cells at different immortalization stages (hTERT_2 and hTERT_7) and passage stages (PDL_25 to PDL_50) were compared to the hTERT_2 stage (n = 12,019 cells; two-tailed Wilcoxon test). **h**, Spearman correlation between the proportion of hUSI-identified senescent cells and SA-β-gal-positive cells. **i**, Workflow for generating HFF ER:RAS cells. **j**, SA-β-gal staining in HFF cells after oncogene activation. **k**, The percentage of SA-β-gal-positive cells (n = 3 biological replicates; two-tailed t-test). Data are presented as mean ± s.e.m. **l**, Expression levels of *CDKN1A*, *LMNB1* and *MKI67* detected by qRT–PCR upon oncogene activation (n = 3 biological replicates; two-tailed t-test). Data are presented as mean ± s.e.m. **m**, Western blot analysis of ER:RAS and p21 protein in OIS HFF cells, with β-actin serving as the loading control. **n**, Boxplot showing hUSI levels in proliferation (treated with DMSO) versus senescence (treated with 4-OHT) groups (n = 3 biological replicates; two-tailed t-test.). The box in the boxplot (**b** and **e**–**g**) represents the interquartile range (IQR), with its lower and upper edges indicating the 25th and 75th percentiles, respectively. The medium value (50th percentile) is shown within the box, and the whiskers extend to the minimum and maximum values within 1.5 times the IQR of the quartiles. For all statistical tests: $^{*}P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 1 \times 10^{-3}$. The exact P values are reported in Source Data for Fig. 2.

hUSI exhibited significantly incremental levels with age and a higher correlation with age, suggesting that hUSI might be more predictive to aging process compared to other methods (Fig. 2b and Extended Data Fig. 3a,b). This finding was also supported by a significant positive correlation between hUSI levels and age in a human dermal fibroblast RNA-seq dataset[133] (Extended Data Fig. 3c,d). We next compared hUSI levels of tumor samples with corresponding normal samples from the same tissues. We found that most tumor types exhibited significantly lower hUSI levels, which is consistent with higher proliferation ability of tumor cells[20] (Fig. 2c). These results confirm the utility of hUSI in evaluating senescence state of bulk samples from various sources.

To evaluate the performance of hUSI in single-cell samples, we compared hUSI with existing methods across six in vitro single-cell RNA sequencing (scRNA-seq) datasets[67,78,134–136], comprising four cell

types and six senescence types (Supplementary Table 3). Given that scRNA-seq can capture heterogeneous senescence states, we initially assessed the distribution of cells and observed distinct clusters of senescent and non-senescent cells within each dataset, with all senescent groups displaying significantly higher hUSI levels compared to their non-senescent counterparts (Extended Data Fig. 3e,g–k). Subsequently, we evaluated the accuracy of hUSI and other methods in predicting senescent cells across these datasets, where hUSI still demonstrated superior performance (Fig. 2d and Supplementary Table 8). Notably, hUSI levels varied among senescent cells subjected to different treatments—for instance, cells undergoing primary senescence exhibited higher hUSI levels than those undergoing secondary senescence, highlighting the potential of hUSI to delineate heterogeneous senescence states[134] (Fig. 2e and Extended Data Fig. 3g–k). Furthermore, we applied hUSI to an in vivo scRNA-seq dataset from human pancreas and observed a significant increase in hUSI levels with age[137] (Fig. 2f and Extended Data Fig. 3f), affirming the reliability of hUSI in assessing senescence state at single-cell level.

Validation of hUSI-identified senescent cells

We next validated the connection between hUSI-identified senescent cells and experimental CS phenotypes. We first applied hUSI to an scRNA-seq dataset[137], which encompassed human lung fibroblasts at different proliferation and passage stages, along with corresponding senescence SA-β-gal activity data, a widely used senescence marker[138]. We observed a significant increase of hUSI along the passage numbers, indicating the sensitivity of hUSI in reflecting senescence state of cells under different replicative stress (Fig. 2g and Extended Data Fig. 3l). We also observed a positive Spearman correlation between the ratio of SA-β-gal-positive cells and hUSI-identified senescent cells in this dataset, suggesting the connection between hUSI levels and the SA-β-gal phenotype at the single-cell level (Fig. 2h).

Then, we conducted experiments on human foreskin fibroblasts (HFFs) transduced with an estrogen receptor fused to the H-RAS^{G12V} protein (ER:RAS), whose expression can be induced by 4-hydroxytamoxifen (4-OHT), ultimately establishing a model of oncogene RAS-induced cellular senescence (Fig. 2i). The senescence state of 4-OHT-treated cells was confirmed by classical CS phenotypes, including increased SA-β-gal activity, significant upregulation of the canonical senescence marker both in mRNA (*CDKN1A*) and protein (p21) and significant lower expression of cycling-associated genes (*LMNB1* (NCBI Gene: 4001) and *MKI67* (NCBI Gene: 4288)) compared to proliferative cells (treated with DMSO) (Fig. 2j–m). We next applied hUSI to RNA-seq data derived from cells in proliferation and senescence states. The result showed significantly higher hUSI scores in senescent group (4-OHT) compared to proliferation group (DMSO) (Fig. 2n). These results further validated the connections between hUSI-identified senescent cells and canonical CS phenotypes.

hUSI helps identify potential senescence regulators

We next explored whether hUSI could aid the discovery of potential senescence regulators by using only gene expression profile as the molecular phenotype. To achieve this, we employed hUSI on a Perturb-seq dataset from human retinal pigment epithelial cells (RPE1 cell line)[139]. This dataset was generated by single-cell CRISPR screening technologies, including 228,790 cells with 2,393 perturbations (Fig. 3a). We calculated hUSI score for each cell and compared the difference between perturbed cells and control cells, thereby identifying a list of genes whose perturbation led to significantly higher hUSI levels (Supplementary Table 9). We found that these genes primarily enriched in pathways related to cell cycle and DNA replication, which is consistent with the irreversible cell cycle arrest of senescent cells[11] (Fig. 3b). We subsequently focused on the top 10 genes (*NACA* (NCBI Gene: 4666), *ECT2* (NCBI Gene: 1894), *PWP2* (NCBI Gene: 5822), *PRIM1* (NCBI Gene: 5557), *PSMA7* (NCBI Gene: 5688), *INCENP* (NCBI Gene: 3619), *PSMD13* (NCBI Gene: 5719), *PSMD2* (NCBI Gene: 5708), *DDX49* (NCBI Gene: 54555) and *PDCD11* (NCBI Gene: 22984)) with the highest hUSI difference (ΔhUSI) (Fig. 3c). The inactivation of these genes led to a significant elevation of hUSI scores, suggesting that their downregulation may play a critical role in inducing senescence. Among these genes, *NACA*[140] and *INCENP*[141] have been reported whose knockdown could induce senescence, whereas other genes currently have not been reported to link to CS.

To validate that these genes indeed regulate senescence, we conducted experimental knockdowns of the top 10 genes in ARPE19 cells (a similar cell line to RPE1). Each gene was targeted by two distinct small interfering RNAs (siRNAs) (Supplementary Table 10) and exhibited significant reduction in its mRNA expression after knockdowns (Fig. 3d,h,l and Extended Data Fig. 4), except for *PWP2*, for which it is technically challenging to design specific siRNAs (Extended Data Fig. 4a). Subsequently, we examined the CS phenotypes induced by downregulating these genes. The result showed that eight out of nine candidate genes (*NACA*, *ECT2*, *PSMD13*, *PRIM1*, *PSMA7*, *INCENP*, *PSMD2* and *DDX49*) caused senescence-associated phenotypes after knocking down, including increased SA-β-gal activity (Fig. 3e,i,m and Extended Data Fig. 4), upregulation of senescence marker at both mRNA (*CDKN1A*) (Fig. 3f,j,n and Extended Data Fig. 4) and protein (p21) levels (Fig. 3g,k,o and Extended Data Fig. 4) and downregulation of cell-cycle-related genes (*CDK1* (NCBI Gene: 983), *LMNB1* and *MIK67*) (Fig. 3f,j,n and Extended Data Fig. 4). These results suggest that, with the application of Perturb-seq technology, hUSI enables the identification of potential senescence regulators.

hUSI aids in evaluating senescence burden in COVID-19

It has been reported that increased accumulation of senescent cells, or senescence burden, could elevate the susceptibility of patients with COVID-19 by promoting severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)-mediated hyperinflammation and cytokine storm[142,143]. Targeted elimination of these senescent cells was suggested as a viable therapeutic approach for COVID-19 (refs. 142,143). However, detailed insights into the accumulation and characteristics of senescent cells in various cell types in COVID-19 remain elusive. Therefore, we applied hUSI to a single-nuclei RNA sequencing (snRNA-seq) dataset of COVID-19, which included 116,313 nuclei from both infected and uninfected lungs, encompassing nine major cell types subdivided into 41 fine cell types[144] (Fig. 4a,b).

Our analysis revealed that most epithelial cells, endothelial cells, fibroblasts and myeloid cells from patients with COVID-19 exhibited significantly higher hUSI values compared to those from uninfected donors (Fig. 4c), aligning with the reported increase in CS in response to SARS-CoV-2 infection[145]. Then, we employed SSE[146] to categorize cells based on hUSI within each fine cell type into 'Normal'-like and 'Senescent'-like groups (Fig. 4d and Extended Data Fig. 5a). We observed that the Senescent-like group exhibited elevated expression of senescence-associated genes and reduced expression of proliferation markers across most cell types (Fig. 4e). Then, we performed differential gene expression analysis between these two groups of each cell type and found that genes significantly highly expressed in the Senescent-like group were enriched in pathways related to viral infection, cellular senescence and immune activation across multiple cell types (Fig. 4f). These findings are consistent with the reports that senescent cells may contribute to immune imbalance and mediate both local and systemic damage in SARS-CoV-2-infected tissues, with the potential to lead to severe symptoms[142].

To further elucidate the variations in senescent cell accumulation associated with COVID-19, we quantified the proportional differences in Senescent-like cells between COVID-19 and normal cases as well as between moderate and severe COVID-19 cases. Our analysis revealed that lung tissues from patients with COVID-19 exhibited higher proportion of Senescent-like cells across most cell subtypes (36/41) (Extended

https://doi.org/10.1038/s43587-025-00886-2

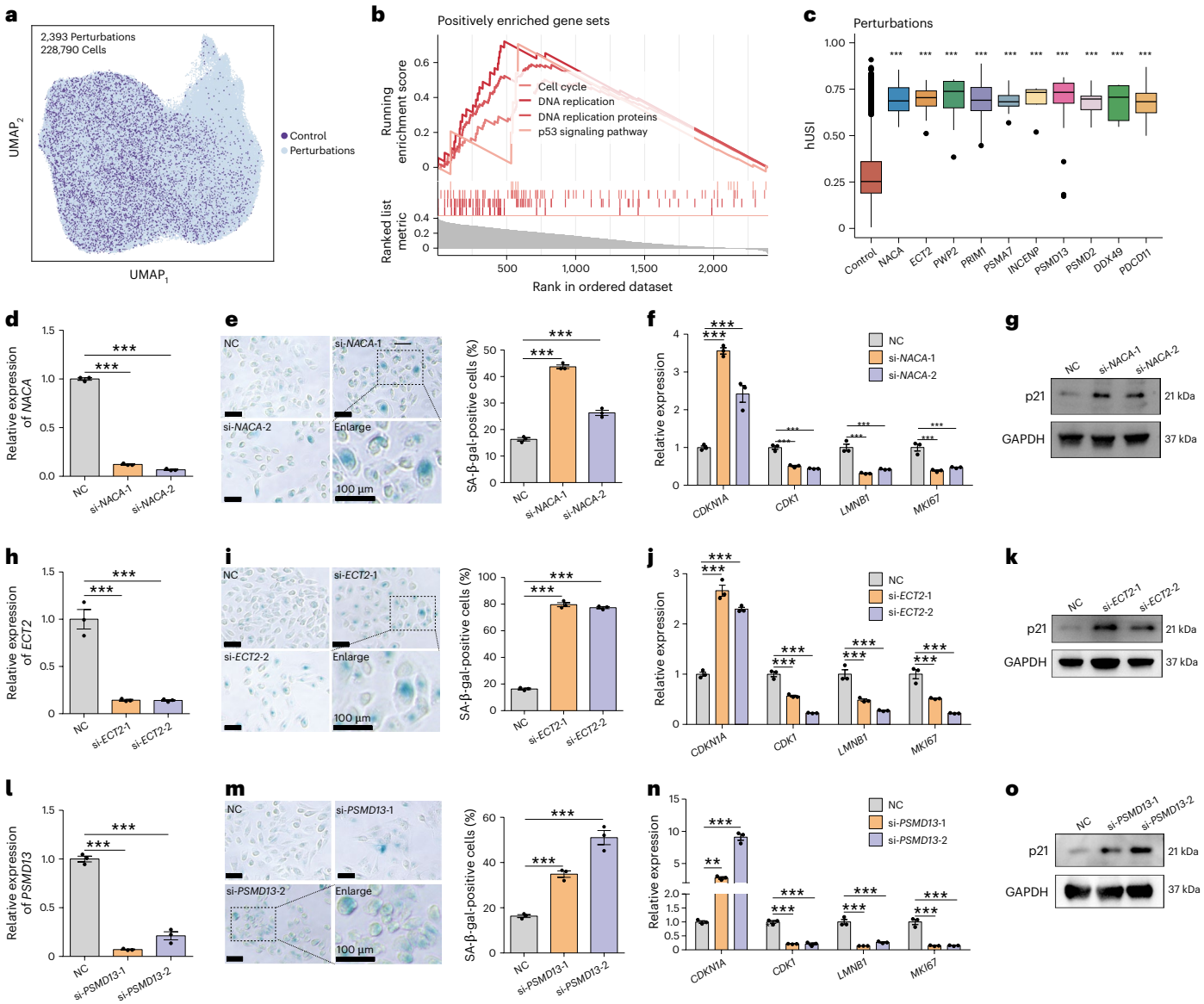


Fig. 3 | hUSI revealed potential senescence regulators in a Perturb-seq dataset. **a**, UMAP plot showing the distribution of cells in the Perturb-seq dataset. **b**, GSEA results for hUSI-identified genes whose perturbations could induce cell into senescence state. Each line corresponds to a different Hallmark gene set, with line color shading indicating NES. **c**, Boxplot showing the hUSI levels of cells targeted by the top 10 hUSI-identified genes compared to that of the control cells ($n = 11{,}705$ cells; $^{***}P < 1 \times 10^{-3}$, two-tailed Wilcoxon test with Benjamini–Hochberg correction). The box represents the IQR, with its lower and upper edges indicating the 25th and 75th percentiles, respectively. The medium value (50th percentile) is shown within the box, and the whiskers extend to the minimum and maximum values within 1.5 times the IQR of the quartiles. **d**–**g**, Senescence validation of *NACA* knockdown in ARPE19 cells ($n = 3$ biological replicates; two-tailed *t*-test). Data are presented as mean ± s.e.m. *NACA* was knocked down by two siRNAs (si-*NACA*-1 and si-*NACA*-2) (**d**). **e**, Left: the senescence state is demonstrated by representative images of SA-β-gal staining. Right: the bar chart shows the percentage of SA-β-gal-positive cells ($n = 3$ biological replicates). Scale bars, 100 μm. The gene expression levels of *CDKN1A*, *CDK1*, *LMNB1* and *MKI67* evaluated by qRT–PCR (**f**) and the western blot of p21 protein levels before and after *NACA* knockdown in ARPE19 cells, with GAPDH serving as the internal control (**g**). **h**–**k** and **l**–**o** present the senescence validation results for *ECT2* and *PSMD13* (as those in **d**–**g**), respectively. Scale bars, 100 μm. For all statistical tests: $^{**}P < 0.01$ and $^{***}P < 1 \times 10^{-3}$. The exact *P* values are reported in Source Data for Fig. 3. NC, negative control.

Data Fig. 5b), and a similar trend was also observed in severe COVID-19 cases (26/41), suggesting that more extensive accumulation of senescent cells might aggravate this disease (Extended Data Fig. 5c). These findings indicate that hUSI can identify senescent cells that accumulate detrimentally in the lungs of patients with COVID-19 across diverse cell types, providing potential targets for therapeutic interventions.

hUSI identifies clinically related senescent cells in melanoma

Given the pivotal role of cellular senescence in tumor progression and its potential to trigger immune responses[147], we further extended the application of hUSI to tumor samples. Recent advances in melanoma immunotherapy, particularly through the use of immune checkpoint inhibitors such as PD-L1 and CTLA-4 antibodies, have achieved durable responses and therapeutic efficacy[148,149]. Despite these successes, the mechanisms driving immunotherapy responses remain only partially elucidated. Previous research established a link between senescent tumor cells and immune recognition[68,150–152], indicating that senescent cells within melanoma tumors can serve as potential immunotherapeutic targets.

By applying hUSI on a melanoma scRNA-seq dataset[153], we evaluated the senescence state of tumor cells (Fig. 5a,b). Then, we observed a pronounced decline in hUSI in the immunotherapy-treated group,

https://doi.org/10.1038/s43587-025-00886-2

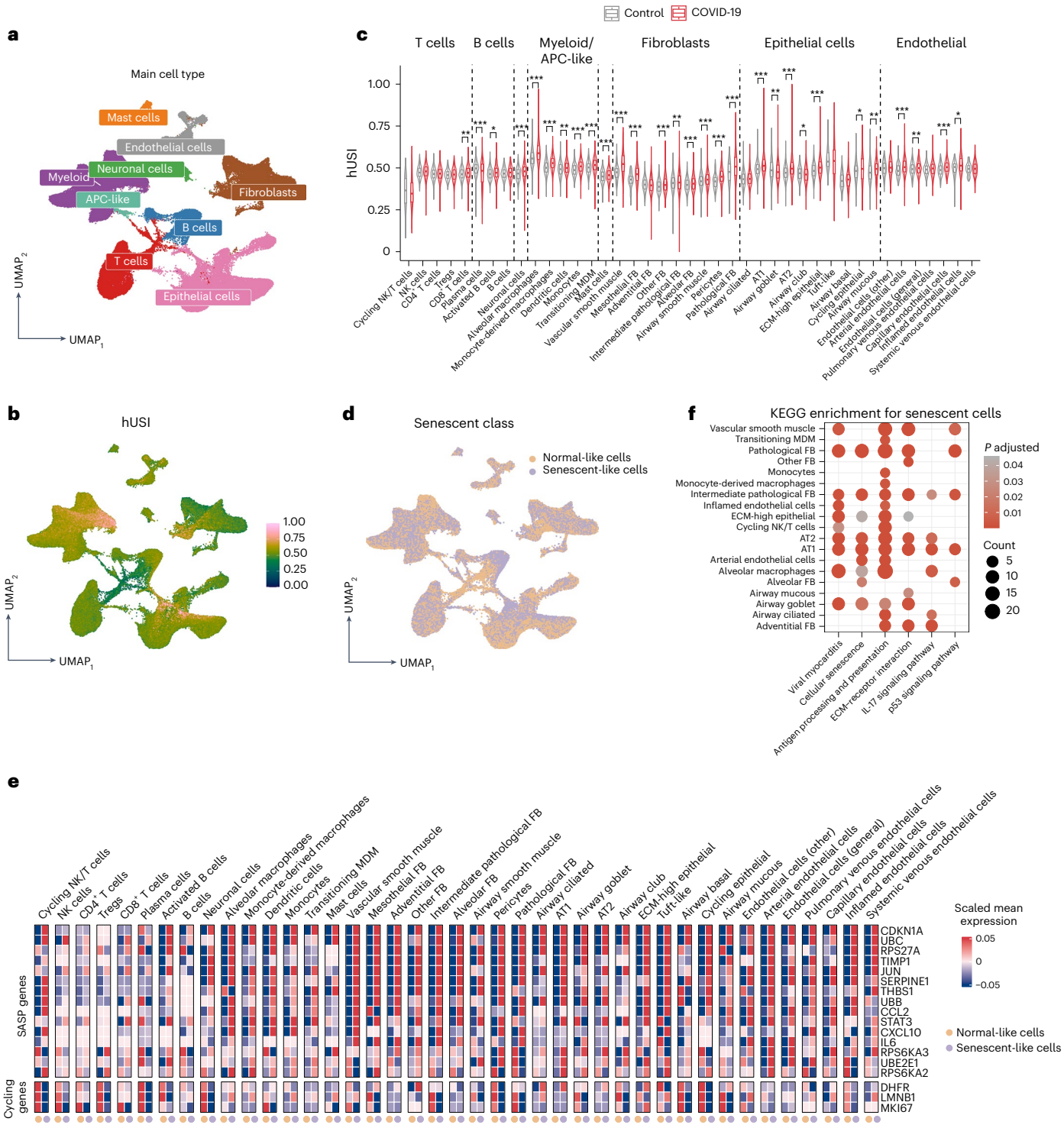


Fig. 4 | hUSI enables the evaluation of senescence burden in COVID-19. **a**, UMAP plot showing the distribution of nine annotated cell types in the snRNA-seq dataset derived from COVID-19 lung tissues (GSE171524). **b**, The distribution of hUSI value for each cell in the COVID-19 dataset projected onto the UMAP. **c**, Violin plot showing the comparison of hUSI levels between control and COVID-19 samples across 41 fine cell types (two-tailed *t*-test; $^{*}P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 1 \times 10^{-3}$; the exact *P* values and cell numbers of each cell type are reported in Source Data for Fig. 4). The box in the boxplot represents the IQR, with its lower and upper edges indicating the 25th and 75th percentiles, respectively. The medium value (50th percentile) is shown within the box, and the whiskers extend to the minimum and maximum values within 1.5 times the IQR of the quartiles. **d**, UMAP plot showing the distribution of hUSI-identified Senescent-like cells and Normal-like cells. **e**, Heatmap showing the scaled mean expression of senescence-related and cycling-related genes in Normal-like and Senescent-like groups across each fine cell type. **f**, KEGG enrichment results of significant (the hypergeometric test with Benjamini–Hochberg correction) highly expressed genes in Senescent-like cells across fine cell types. Only gene sets with the highest representation across cell types are shown. The *x* axis represents gene set names, and the *y* axis represents cell types.

which in line with therapy-resistant malignant cells exhibiting attenuated CS phenotypes[154] (Fig. 5c). To better reveal the heterogeneity of CS in tumor, we next used Gaussian mixture modeling (GMM)[155] to classify tumor cells based on hUSI, identifying three senescence subpopulations ('Cycling', 'Transitional' and 'Senescent') with progressively increased hUSI scores (Extended Data Fig. 6a). These subpopulations mirrored gene signatures from independent senescent melanocyte datasets[126] (Extended Data Fig. 6b). As expected, the treated group displayed a markedly reduced proportion of Senescent tumor cells, supporting the diminished presence of senescent cells in melanoma after immunotherapy (Fig. 5d). Trajectory analysis reconstructed via co-expression modules (ICAnet[156]) and dimension reduction (phateR[157]) revealed a progression from Cycling to Senescent subpopulation (Extended Data Fig. 6c), accompanied by increasing expression levels of two canonical senescence markers, *CDKN1A* and *SERPINE1* (NCBI Gene: 5054) (Fig. 5e,f), suggesting the underlying dynamic CS progression in melanoma.

After performing GSEA of genes highly expressed in each subpopulation, we found increased immune activity in Senescent tumor cells (Fig. 5g). To further explore the relationship between Senescent tumor cells and immune response, we deconvoluted the RNA-seq data from the TCGA-skin cutaneous melanoma (SKCM) cohort using EpiDISH[158] to estimate the proportion of each subpopulation and quantified the abundance of 22 immune cell types using CIBERSORT[159] in individual patients. The result showed a higher positive correlation between the proportion of Senescent cells and several types of immune cells (including M1 macrophages, $CD8^+$ T cells, gamma delta T cells and various activated immune cells (including activated CD4 memory T cells, activated Mast cells and activated natural killer (NK) cells)) compared to the Cycling subpopulation (Fig. 5h). We subsequently assessed the effect of Senescent tumor cells on survival of patients with melanoma. Survival analysis showed that the higher proportion of Senescent cells in a patient, the more favorable for the patient's survival, whereas the Cycling subpopulation is the opposite (Fig. 4i and Extended Data Fig. 6d,e). These results suggest that hUSI enables the distinguishing and characterization of heterogenous senescence states of melanoma tumor cells, providing a potential prognostic biomarker for melanoma therapy.

hUSI reveals senescent-specific signal pathways in melanoma

Because tumor microenvironment (TME) is important for tumor genesis, progress and therapy[160], we further investigated the cell–cell communication characteristics of hUSI-identified tumor senescence subpopulations. Previous studies reported that senescent cells can communicate with neighbor cells and modulate their behavior through SASP. For instance, SASP presented by senescent tumor cells contributes to the communication with immune system by attracting immune cells (such as T cells and NK cells) and then leading to the clearance of tumor cells[68,151,152]. In addition, senescence-associated communication has been speculated to regulate immune surveillance and influence tumorigenesis[161].

We analyzed the cell–cell communication among the three tumor cell subpopulations (Cycling, Transitional and Senescent) and their neighboring cells (including endothelial cells (Endo), B cells (B), T cells (T), NK cells, macrophages (Macro) and cancer-associated fibroblasts (CAFs)) using CellChat[162]. The results showed that the communication strength of hUSI-identified Senescent subpopulation was higher than Cycling and Transitional (Fig. 5j and Extended Data Fig. 6f). Given the SASP of senescent cells, we further analyzed the global output communication patterns of these three subpopulations and uncovered two different signaling patterns, with pattern 1 corresponding to the Senescent tumor subpopulation and pattern 6 corresponding to the Cycling and Transitional subpopulations (Extended Data Fig. 6g). Within pattern 1, we found several SASP-associated genes, including *SPP1* (NCBI Gene: 6696), *AGRN* (NCBI Gene: 375790) and *LIFR* (NCBI Gene: 3977), underscoring their potential roles in mediating cell–cell interactions in senescent melanoma cells. Nevertheless, additional studies are required to clarify the specific functions of given genes within this pattern in senescent melanoma cells.

To analyze which pathways are responsible for potentially regulating senescent tumor cells in the TME, we compared the communication strength of incoming signaling pathways. We identified five key pathways, including TGFβ, chondroitin sulfate proteoglycan 4 (CSPG4), CD6, CD46 and chemokine signaling (CCL), which exhibited significant signal strength in the Senescent subpopulation but were not detected in the Cycling subpopulation (Fig. 5k), suggesting that these signaling pathways are more likely to specifically function in senescent tumor cells. TGFβ is well documented for its role in inducing senescent phenotype of tumor cells, primarily secreted by macrophages originating in tumor stroma[163,164]. Our network analysis also revealed that Senescent tumor cells receive TGFβ signals from macrophages (Fig. 5l), in line with findings in lymphoma[164]. Interestingly, Senescent tumor cells received more TGFβ signals from CAFs, indicating that, as a solid tumor, melanoma differs from lymphoma in microenvironment by the presence of a high number of fibroblasts[165] (Fig. 5l). In addition, we found that the Senescent subpopulation specifically received CCL signals and CD46 signals from macrophages and endothelial cells, respectively, and primarily received CSPG4 and CD6 signals from CAFs and T cells, respectively (Fig. 5k and Extended Data Fig. 6h). Although these four signaling pathways were linked to tumor progression[166–170], their specific roles in the context of tumor cell senescence require further investigation.

We also investigated the impact of these specific signaling pathways on patient survival. Initially, we selected four receptors—*TGFBR2* (NCBI Gene: 7048), *ACVR1* (NCBI Gene: 90), *ALCAM* (NCBI Gene: 214) and *CCR10* (NCBI Gene: 2826)—which are highly specifically expressed in the Senescent subpopulation (Fig. 5m). Then, we conducted survival analysis on the TCGA-SKCM cohort and found that highly expressing these genes benefits patient survival (Extended Data Fig. 6i–l) as well as a significantly higher survival probability in groups with high enrichment scores of these four genes (Fig. 5n). This analysis suggests that upregulating these genes in senescent melanoma cells could enhance their interactions within the TME, potentially inhibiting tumor progression. Overall, these findings underscore the clinical implications of using hUSI to identify specific targets of senescent melanoma tumor cells in the TME.

Discussion

We developed hUSI, a robust and versatile transcriptome-based tool for evaluating cellular senescence across diverse conditions and cell types. By leveraging the most comprehensive senescence-associated RNA-seq dataset so far, we generated a universal senescence gene expression profile and optimized hUSI for generalization, ensuring resilience to batch effects, sparsity and outliers. Comparative analyses confirmed the superior performance of hUSI across cell types, senescence types and sequencing platforms, with a positive correlation to age in in vivo samples further validating its applicability. Using Perturb-seq technology, hUSI enabled high-throughput identification of potential senescence regulators, which were experimentally validated. Beyond this, hUSI revealed distinct roles of senescence in disease: in COVID-19, it highlighted the detrimental impact of senescence-driven inflammation and tissue damage, and, in melanoma, it suggested that senescent tumor cells may enhance immune responses and improve patient survival. These findings demonstrate the potential of hUSI to advance senescence research and guide therapeutic strategies.

We also highlighted that hUSI, as a data-driven senescence scoring method based on the OCLR model and the most comprehensive gene expression datasets on cellular senescence so far, exhibited its unprecedented advantages in learning senescence features and extensibility. Although binary classification-based methods have benefited

https://doi.org/10.1038/s43587-025-00886-2

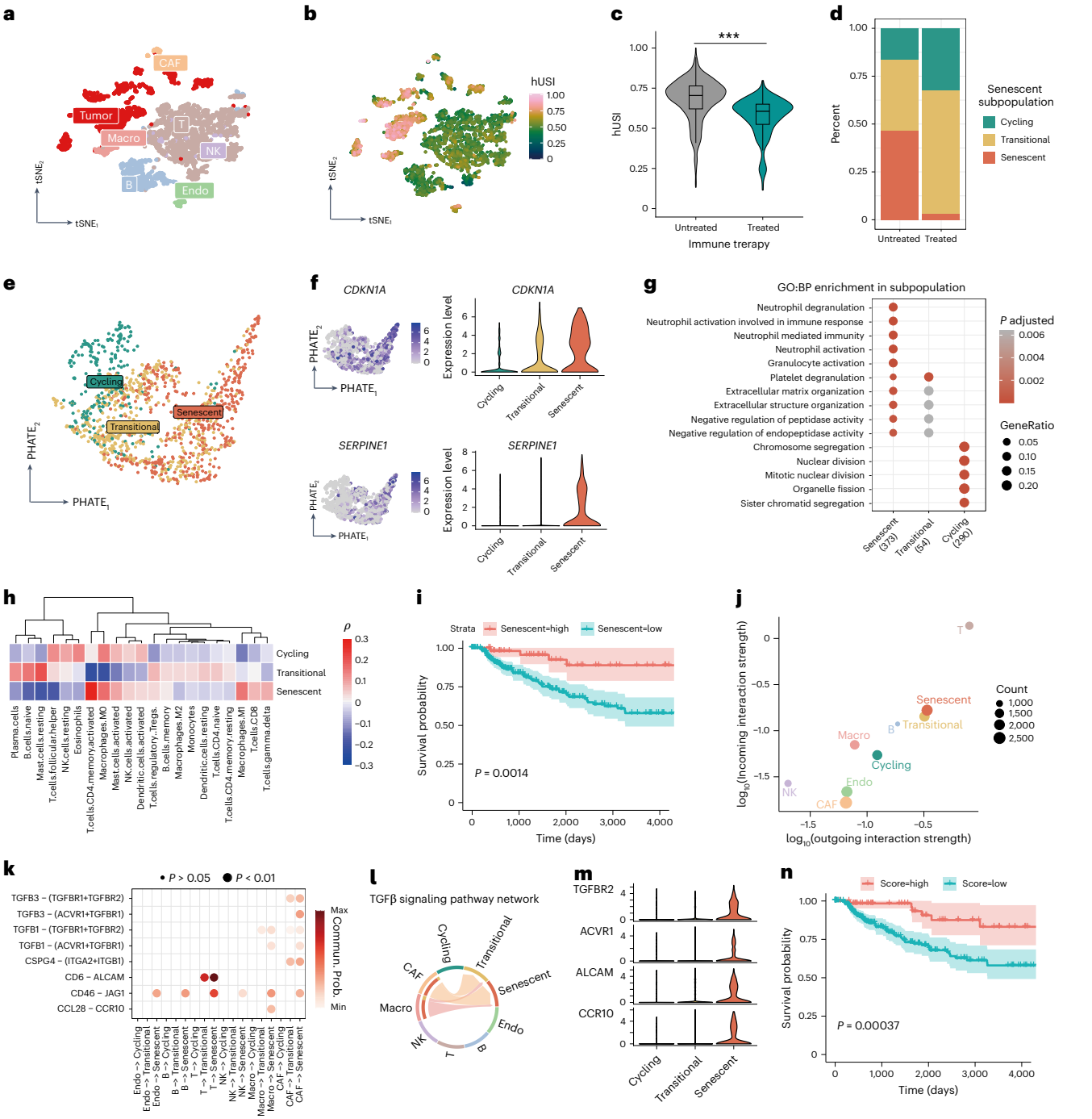


Fig. 5 | hUSI identifies prognosis-related senescent tumor cells in melanoma. **a**, tSNE plot showing the distribution of seven cell types identified within the melanoma dataset (GSE72056). **b**, The UMAP displays the distribution of hUSI value of each cell in the melanoma dataset. **c**, Violin plot compares hUSI levels between untreated and treated groups in immune therapy (two-tailed Wilcoxon test). The box in the boxplot represents the IQR, with its lower and upper edges indicating the 25th and 75th percentiles, respectively. The medium value (50th percentile) is shown within the box, and the whiskers extend to the minimum and maximum values within 1.5 times the IQR of the quartiles. $^{***}P < 1 \times 10^{-3}$. The exact *P* values are reported in Source Data for Fig. 5. **d**, The proportions of three tumor senescence subpopulations of untreated and treated groups in immune therapy. **e**, Senescence trajectory of melanoma tumor cells in PHATE space. **f**, Increased expression levels of aging-related marker genes *CDKN1A* and *SERPINE* along the trajectory in **e** and in different senescence subpopulations. **g**, Different GO terms (biological process (BP)) characterize Cycling, Transitional and Senescent subpopulations, with replication-associated terms enriched by the Cycling subpopulation and immunity activation-related terms enriched by the Senescent subpopulation (the hypergeometric test with Benjamini–Hochberg correction). The exact FDR-corrected *P* values are reported in Source Data for Fig. 5. **h**, Heatmap of Spearman correlation coefficient between the abundance of the three subpopulations and that of 22 immune cell types. **i**, Survival curves of patients with melanoma with low or high proportion of Senescent subpopulation in the TCGA-SKCM cohort (log-rank test). **j**, Dot plot showing the interaction strength of three tumor subpopulations and six cell types in the TME. **k**, Dot plot of communication probabilities for signaling pathways presenting in the Senescent subpopulation while absent in the Cycling subpopulation (permutation test, $n = 100$ iterations). **l**, Chord plots showing that Macro and CAFs interact with Transitional and Senescent subpopulations by the TGFβ signaling pathway. **m**, Violin plot showing the expression levels of four receptors specifically expressed in the Senescent subpopulation. **n**, Survival curves of patients with melanoma with low or high enrichment score of *TGFBR2*, *ACVR1*, *ALCAM* and *CCR10* in the TCGA-SKCM cohort (log-rank test). tSNE, t-distributed stochastic neighbor embedding.

from their superior ability to capture data features[171], improving the identification of senescent cells, classification of samples as negative, that is, non-senescent cells, requires careful evaluation. A suboptimal selection of negative samples can affect the generalizability of the model. To overcome these challenges, hUSI adopted the OCLR machine learning algorithm[29], which is more robust in dealing with the shortage and confounder signals of negative samples by learning the features based merely on positive samples. We further confirmed the biological interpretability of OCLR-learned features through examining the weights of classical senescence-associated genes and preference for known senescence-associated pathways. Notably, due to its straightforward model training and implementation strategy, hUSI offers potential extensibility in evaluating senescence in other types of senescence samples as well as samples from other species by recombining or recollecting training set samples.

Although hUSI has demonstrated superior performance in multiple aspects, limitations still exist in its current status. First, with increased accumulation of senescence-associated transcriptome data, the quality and quantity of training sets still have room for further improvement. In particular, due to the heterogeneity of senescence state across various cell types, the cell-type-specific inclusion criteria for senescent samples in the training set deserve further consideration. Second, the integration strategy for incorporating additional transcriptome data into the training set requires further investigation to optimize or customize hUSI in the future. Third, although hUSI characterized senescent tumor cells and identified key signaling pathways in melanoma, its generalizability across tumor types requires further investigation. In this study, our training set incorporated a part of tumor cell lines to help learn universal senescence features via OCLR, and we acknowledge that there might be higher heterogeneity of tumor cells compared to normal. In the future, more tumor samples could be separately collected to develop a tumor-specific hUSI applicable across all tumor types. Lastly, with the advantages of spatial transcriptome technology in the study of cellular senescence[172], additional efforts should be made to explored the application of hUSI within this context.

To our knowledge, this research introduces the first unified cellular senescence prediction model in human. In the future, leveraging advancements in foundation models, such as SCimilarity[173], which provide harmonized cellular representations, can allow hUSI to perform robust cellular senescence predictions across diverse biological systems. This integration can enhance generalizability, deepen insights into conserved and context-specific senescence mechanisms and pave the way for transformative therapeutic innovations targeting cellular aging and related diseases.

Methods

Data collection and pre-processing

In this study, we collected 843 samples from 73 senescence-related studies[24,30–98] with both senescent and non-senescent samples for model building. The sample size was determined through comprehensive data collection and quality control. We also collected their senescence-associated phenotypes from original publications and divided them into primary evidence (including SA-β-gal activity, cell cycle arrest, increasing level of p16, p21 or core SASP and decreasing level of lamin B1) and secondary evidence (including morphological changes, chromatin reorganization, DNA damage response, increasing level of senescence-associated genes and decreasing level of cycling-associated genes). Then, FASTQ files of all 843 samples (Supplementary Table 1) were downloaded through accession numbers in the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) or the European Molecular Biology Laboratory European Bioinformatics Institute (EMBL-EBI) database (https://www.ebi.ac.uk/ebisearch/about). The R package GEOquery (version 2.58.0) was used to collect metadata of samples through GEO accession numbers if provided.

For RNA-seq data pre-processing, Trim Galore (version 0.6.6) was used to filter out low-quality reads and bases in the 3′ end of FASTQ files, and then the clean data files were consequentially aligned to human GRCh38 reference genome by STAR (version 2.2.1)[174], and only unique mapping reads were included for further analysis. StringTie (version 2.2.1)[175] was used to quantify the gene expression level and normalize by transcripts per million (TPM). Only protein-coding genes (annotated by GENCODE version 31) with TPM > 3 in 99% of samples were included for next analysis. Then, samples with at least two senescence phenotypes (including at least one primary evidence) were selected to build the representative senescence profiles as training set. We also included only cell lines that contain both senescent and non-senescent samples, resulting in 385 senescent samples and 385 non-senescent samples from 54 cell lines (34 cell types). Considering bias introduced by batch effect, we used log-space transformation, scaling and mean centering to further reduce the disturbances for model training.

To eliminate the potential disturbance from cell type preference, we decided to filter cell type signatures of cell types included in our training set. To achieve this, we initially used the raw counts of 385 non-senescent samples to identify cell-type-specific genes by DESeq2 (version 1.42.1)[176], and only genes with adjusted $P < 0.05$ and $\log_2$ fold change > 1 were kept. Subsequently, we generated new log (TPM + 1) profiles by removing varying numbers (1–300) of cell-type-specific genes for each cell type. Using these new profiles, we clustered the cells employing the Louvain algorithm via the FindClusters function of the Seurat package (version 4.4.0)[177], with the resolution parameter set to 0.6. Finally, we calculated the ARI between the clustering results and the corresponding cell type labels. Because reducing cell type signatures could reduce heterogeneity of gene expression profiles, thereby resulting in a lower ARI, we selected the number of cell-type-specific genes to remove based on the elbow point of the ARI curves, which corresponds to 85 genes for each cell type.

Model training and hyperparameter optimization

The processed gene expression matrix of senescent samples was used to learn senescence features by the OCLR algorithm[29] implemented by the 'gelnet' function of the gelnet package (version 1.2.1) in the R (version 4.0.5) platform. The mathematics details are provided in Supplementary Note 1. In this function, each gene was weighted iteratively, and the final gene weight vector represented the maximum likelihood estimation of gene expression distribution in senescent samples. In this study, we tested different L2 penalty values (0.001, 0.01, 0.1, 1, 5 and 10) in the gelnet function (with parameter x = processed gene expression matrix, parameter y = NULL and parameter L1 = 0) to learn senescence features. Considering that different quantification methods can be used to calculate senescence score based on the OCLR-learned senescence features, six common quantification methods (including Spearman correlation, Pearson correlation, logistic function, dot product, ssGESA and mean expression) were also tested. We describe the detailed implementation of different quantification methods in Supplementary Note 2.

Through the 10-fold cross-validation OOD test (Supplementary Note 3), we evaluated the performance of hUSI generated by different combinations of OCLR models (decided by L2 values) and quantification methods, ultimately finding that L2 = 1 and Spearman correlation yielded the most advantageous results. Finally, by applying the OCLR model to the gene expression profiles of all 385 senescent samples, we obtained the OCLR-learned senescence features (the output weighted gene vector). We then used Spearman correlation to score each sample, scaling the resulting coefficient values to derive the hUSI score. We next evaluated the robustness of hUSI in batch effect, data sparsity and outlier (Supplementary Note 4) and compared OCLR models with three binary classification models, including RF, EN and SVM (Supplementary Note 5).

GSEA for OCLR-learned senescence features

To demonstrate that hUSI prefers senescence state rather than cell types, we performed GSEA using the R package fgsea (version 1.16.0)[178] on OCLR-learned features in the CellMarker database (http://xteam.xbio.top/CellMarker/), which includes 268 common human cell type signatures. The normalized enrichment score (NES) was chosen to compare the enrichment degree across different gene sets. Additionally, we compared the enrichment results of all cell types to a senescence-related gene set, SenUp[105]. We also performed GSEA for genes in OCLR-learned features ranked by their weights across four widely used gene set databases: Hallmark gene sets from the Molecular Signatures Database (MSigDB)[106] (https://www.gsea-msigdb.org/gsea/msigdb), GO[107,108] (https://www.geneontology.org/), KEGG[110] (https://www.genome.jp/kegg/) and the Reactome database[109] (https://reactome.org/).

Model benchmark in in vitro datasets

To evaluate the performance of hUSI in classifying senescent samples or cells, we benchmarked hUSI against other 31 senescence scoring methods both in bulk (GSE19864 (ref. 124), GSE16058 (ref. 125), GSE83922 (ref. 126), GSE11954 (ref. 127), GSE100014 (ref. 128) and GSE77239 (ref. 129)) and in single-cell in vitro (GSE115301 (ref. 134), GSE119807 (ref. 135), GSE94980 (ref. 78), GSE102090 (ref. 67) and GSE226225 (ref. 136)) datasets. The sample size was defined by the source publications, and only datasets with at least three replicates per condition were included in the analysis. For the bulk datasets, we downloaded the normalized microarray gene expression profiles using GEO accession numbers. For single-cell datasets, we used the NormalizeData function of Seurat (version 4.2.1)[177] to normalize the raw counts matrix downloaded using GEO accession numbers. We include 31 methods into benchmarking, which are categorized into SenMarker (including *GLB1* (NCBI Gene: 2720), *TP53* (NCBI Gene: 7157), *CDKN1A*, *CDKN2A* (NCBI Gene: 1029), *LMNB1*, *IL1A* (NCBI Gene: 3552), *RB1* (NCBI Gene: 5925), *CDK1*, *CDK4* (NCBI Gene: 1019), *CDK6* (NCBI Gene: 1021), *MKI67*, *CDKN2B* and *SERPINE1*); SenSet (including *SenMayo*[19], *CellAge*[18], *GenAge*[120], *ASIG*[121], *SASP* pathway (downloaded from the MSigDB under accession ID R-HSA-2559582), *AgingAtlas*[122], *SenUp*[105] and *SigRS*[21]); TraditionScore (including *DAS*, *mSS*, *DAS_mSS*[14] and *CS_score*[20]); and MachineScore (including *CSS*[20], *lassoCS*[23], *ECcores36* (ref. 123), *HCCcores19* (ref. 179), *SENCAN*[24] and *SenCID*[25]). Detailed implementation of these methods is provided in Supplementary Note 6. After obtaining senescence scores for samples generated by each method, we calculated the AUC values for each method across all datasets. Additionally, we computed the mean AUC values across all datasets for each method to represent their overall performance.

Validation of hUSI on in vivo datasets

To validate the performance of hUSI on in vivo datasets, we applied hUSI to both bulk and single-cell in vivo datasets. We computed the hUSI for bulk samples from GTEx[130] and TCGA databases using TPM normalized gene expression matrix. The gene expression matrix of TCGA samples was collected from UCSC Xena (http://xena.ucsc.edu/public/). The gene expression matrix of GTEx samples and corresponding metadata were collected from GTEx Portal (version 8) (https://www.gtexportal.org/home/downloads/adult-gtex). We then compared hUSI of GTEx samples sourced from different age groups and hUSI of TCGA tumor samples with corresponding GTEx normal samples from the same tissues across various cancer types. We used a *t*-test to calculate significance as we assumed the distribution of these samples to be normal, but this was not formally tested. We also applied hUSI to an independent human skin bulk dataset obtained from the GEO under accession number GSE113957 (ref. 133) along with the associated age information. Then, we applied hUSI to an scRNA-seq dataset from in vivo human pancreas cells (GSE81547 (ref. 137)). The raw gene expression profiles were obtained through GEO accession number, and the processes of data normalization, cell clustering and visualization were all performed using the Seurat package. The hUSI scores for single cell were calculated based on gene expression matrices normalized by counts per million (CPM).

Classification of senescent cells based on hUSI

We recommend two methods based on different assumptions. (1) If assuming that there are only two cell states, senescence and non-senescence, in the dataset, we use a variance-based method—SSE—for thresholding. SSE, as the well-performed binary clustering method for one-dimensional vector[146], can divide cells into senescence and non-senescence groups by minimizing the variance of hUSI of two groups (implementation details are provided in Supplementary Note 7). (2) If assuming that there are uncertain senescence states in the dataset, we use GMM[155] to estimate the optimal number of senescence states and the probability that each cell belongs to a specific state. Because GMM is frequently used as the optimal thresholding tool in various scenarios[180,181], it is helpful to determine cellular senescence heterogeneity (implementation details are provided in Supplementary Note 7).

SA-β-gal scRNA-seq dataset analysis

Following a similar scRNA-seq data analysis pipeline as mentioned above, we applied hUSI to an scRNA-seq dataset (GSE175533 (ref. 137)) and identified senescent cells using the SSE method in each sample. We then calculated the Spearman correlation coefficient between the proportion of senescent cells and the percentage of cells exhibiting positive SA-β-gal activity.

Oncogene-induced senescent HFF cells RNA-seq analysis

We performed experiments to generate RNA-seq data of oncogene-induced senescence (OIS) and growing HFF cells (Supplementary Notes 8 and 14) and applied hUSI to these samples. The sample size for experiments was determined based on repeatability, with a minimum of three replicates per condition to ensure statistical robustness. Cells from the same passage were randomly allocated to experimental conditions, with each condition biologically replicated at least three times. Representative results shown in the figures were randomly selected from all technical replicates. Experimental statistical data distribution was assumed to be normal as we used a simple randomization strategy, but this was not formally tested. Data collection and analysis were not performed blinded to the conditions of the experiments.

Perturb-seq analysis of RPE1 cells

We obtained the Perturb-seq dataset for RPE1 cells from a previous study[139], comprising a total of 228,790 cells and 2,393 genetic perturbation annotations. The control group consists of cells with non-targeting control single guide RNAs (sgRNAs), and the perturbation group comprises cells with sgRNAs specifically targeting a gene of interest. For the perturbation group, we retained cells exhibiting expression levels of the targeted gene at less than 1/4 of those observed in the control group using Python. hUSI-identified senescence regulators were defined as those where perturbed cells showed a significantly higher hUSI (false discovery rate (FDR)-corrected Wilcoxon $P < 0.05$) and a greater than 1.5-fold increase in the hUSI compared to the control group. Raw unique molecular identifier (UMI) counts were normalized to log (CP10K + 1) using Scanpy (version 1.10.4). We selected the top 2,000 highly variable genes using the 'highly_variable_genes' function of Scanpy. We then constructed a nearest neighbor graph with default parameters and performed uniform manifold approximation and projection (UMAP) embedding using the top 10 principal components (PCs). Cells in the perturbation group were ranked based on the fold change of hUSI for subsequent KEGG pathway analysis. We further conducted experimental validation for top hUSI-identified senescence regulators, with details provided in Supplementary Notes 8 and 13.

Analysis of snRNA-seq data of COVID-19-infected samples

The raw expression matrix was analyzed using Seurat (version 4.2.1) in R (version 4.0.5). Then, we normalized the raw counts using the NormalizeData function and calculated the hUSI score for each cell. hUSI levels were compared between control and COVID-19 groups across cell types, using a *t*-test to calculate significance as we assumed the distribution of cells was normal, but this was not formally tested. Subsequently, cells were categorized using the 'cell_type_fine' label, resulting in 41 cell types, and SSE was employed to classify senescent groups in each cell type. We compiled a list of senescence-associated genes (including *CDKN1A*, *IL6* (NCBI Gene: 3569), *CCL2* (NCBI Gene: 6347), *CXCL10* (NCBI Gene: 3627), *SERPINE1*, *THBS1* (NCBI Gene: 7057), *TIMP1* (NCBI Gene: 7076), *UBC* (NCBI Gene: 7316), *RPS27A* (NCBI Gene: 6233), *JUN* (NCBI Gene: 3725), *UBB* (NCBI Gene: 7314), *STAT3* (NCBI Gene: 6774), *UBE2E1* (NCBI Gene: 7324), *RPS6KA2* (NCBI Gene: 6196) and *RPS6KA3* (NCBI Gene: 6197)) and cycling-associated genes (including *DHFR* (NCBI Gene: 1719), *MKI67* and *LMNB1*) and examined their expression levels in both normal and senescent groups. The FindMarkers function was used to calculate differentially expressed genes (DEGs) in each cell type (senescent group versus normal group), which were subsequently used to perform GSEA in the KEGG database using the 'enricher' function. For dividing the patients with COVID-19 into moderate and severe groups, we used interval days between death to symptom onset ('interval_death_symptoms_onset_days'). If interval days is less than the 0.25 quantile of all patients' interval days, it is classified as severe; if it is greater than the 0.75 quantile, it is classified as moderate.

Analysis of scRNA-seq data of melanoma samples

The processed melanoma single-cell matrices, retaining only defined tumor cells (malignant = 2) and non-tumor cells (malignant = 1) (including T cells, B cells, macrophages, endothelial cells, CAFs and NK cells), were analyzed using the R package Seurat (version 4.2.1). To validate the reliability of inferred tumor subpopulations, we identified specifically highly expressing genes ($\log_2$ fold change > 0.1) of each subpopulation and overlapped them with DEGs derived from melanoma microarray data (RS versus Growing), which were calculated using the linear models 'lm (gene expression-pheno)'. The R function 'phyper()' was used to test the overlapping significance. To observe the positional relationships of the different subpopulations of CS states in the projected space, for each gene, we calculated the Pearson correlation coefficient between hUSIs and gene expression values of all cells and selected the top 1,500 genes ranked by absolute correlation coefficient values as hUSI-related genes. Then, using the tool ICAnet (version 0.0.9)[156], the tumor cells were integrated based on the 89 co-expression modules of the hUSI-related genes above. The phateR package (version 1.0.7)[157] was used to infer the senescence trajectory based on five PCs of the 89 co-expression modules. Highly expressed genes of each tumor subpopulation were enriched using the R package clusterProfiler (version 3.18.1.)[182] on the GO database, using $\log_2$ fold change as the ranking metric. Cell communication analysis was carried out using the R package CellChat (version 1.0.0.)[162], and the communication intensity between tumor subpopulations and different non-tumor cell types within a signaling network was quantified. When filtering pathways specific for Senescent tumor cells, we set three tumor subpopulations as 'target' and T cell, NK cell, macro cell and CAF cell as 'source'. Only signaling pathways with no interaction strength in the Cycling subpopulation were selected.

Data analysis of melanoma patient cohort in TCGA

The normalized level 3 RNA-seq data of a melanoma patient cohort (SKCM), along with associated clinical data, were downloaded from the TCGA database using the R package TCGAbiolinks (version 2.18.0)[183]. To analyze the proportion of cells with different senescence degrees, three tumor subpopulations were used as reference to deconvolute the RNA-seq data of the SKCM patient cohort using the R package EpiDISH (version 2.6.1)[184]. The abundance of the 22 immune components was calculated using CIBERSORT (R script version 1.04.)[159]. Survival analysis was performed using the R packages 'survival' (version 3.4-0) and 'survminer' (version 0.4.9) with 'OS = vital_status' and 'OS.time = days_to_last_followup' as the parameters. The enrichment score of four Senescent subpopulation-specific receptor genes (*TGFBR2*, *ACVR1*, *ALCAM* and *CCR10*) was calculated by the R package GSVA, using the 'gsva' function with 'method = zscore'.

Statistics and reproducibility

No statistical method was used to predetermine sample size, but our sample sizes are similar to those generally employed in the field[25]. For training set, the sample size for the training set was determined through comprehensive data collection (public databases and publications were systematically searched using keywords such as 'senescence', 'RNA-seq' and 'human' to identify relevant datasets) and quality control (only datasets containing both senescent and control samples that met predefined standards were retained for analysis). For publicly available datasets, the sample size was defined by the source publications. Only datasets with at least three replicates per condition were included in the analysis. In this study, the sample size for experiments was determined based on repeatability, with a minimum of three replicates per condition to ensure statistical robustness. In experiments, cells from the same passage were randomly allocated to experimental conditions at the seeding stage following the simple randomization strategy, with each condition biologically replicated at least three times. Representative results shown in the figures were randomly selected from all technical replicates. In data analysis, we uniformly sampled the data when performing the 10-fold cross-validation OOD test. Data exclusion standards for model training are mentioned in the Methods, and no data were excluded from the experiments. For downstream analyses, low-quality cells were excluded from the scRNA-seq and Perturb-seq analyses based on pre-established quality control criteria to ensure the reliability of the data. Although the data distribution was assumed to be normal, this assumption was not formally tested, and most statistical tests employed in the study were non-parametric. Statistical analyses were conducted using appropriate statistical tests. Further details on the statistical methods can be found in the figure legends.

Reporting summary

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

Data availability

All RNA-seq data generated by this study can be downloaded under GEO accession number GSE282274. The published datasets included in this article are recorded in Supplementary Table 1 for hUSI development and in Supplementary Table 3 for hUSI validation. The processed RNA-seq profile used to develop hUSI is deposited in a figshare database (https://figshare.com/s/0335252fca31a5303510). The GRC38 reference genome used to align the raw sequence files was downloaded from the NCBI (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.26/). The gene annotation file was downloaded from GENCODE (https://www.gencodegenes.org/human/release_31.html). The processed melanoma profiles were downloaded from the GEO under accession number GSE72056 (ref. 153). The raw single-nuclei counts matrix of normal and COVID-19 patients' lung tissues was downloaded from the GEO under accession number GSE171524 (ref. 144) along with the corresponding metadata. All source data used to reproduce the results in this article can be found in a GitHub repository (https://github.com/WJPina/HUSI/tree/main/Data). Any other data reported in this paper are available from the lead contact upon reasonable request.

Code availability

hUSI can be implemented in both R and Python. The data and codes used to reproduce our analysis results are provided in a GitHub repository (https://github.com/WJPina/HUSI) along with a detailed usage guideline.

References

1. van Deursen, J. M. The role of senescent cells in ageing. *Nature* **509**, 439–446 (2014).
2. Acosta, J. C. et al. A complex secretory program orchestrated by the inflammasome controls paracrine senescence. *Nat. Cell Biol.* **15**, 978–990 (2013).
3. Birch, J. & Gil, J. Senescence and the SASP: many therapeutic avenues. *Genes Dev.* **34**, 1565–1576 (2020).
4. Chaib, S., Tchkonia, T. & Kirkland, J. L. Cellular senescence and senolytics: the path to the clinic. *Nat. Med.* **28**, 1556–1568 (2022).
5. Wang, L., Lankhorst, L. & Bernards, R. Exploiting senescence for the treatment of cancer. *Nat. Rev. Cancer* **22**, 340–355 (2022).
6. Huang, W., Hickson, L. J., Eirin, A., Kirkland, J. L. & Lerman, L. O. Cellular senescence: the good, the bad and the unknown. *Nat. Rev. Nephrol.* **18**, 611–627 (2022).
7. Cho, K. A. et al. Morphological adjustment of senescent cells by modulating caveolin-1 status. *J. Biol. Chem.* **279**, 42270–42278 (2004).
8. Petrova, N. V., Velichko, A. K., Razin, S. V. & Kantidze, O. L. Small molecule compounds that induce cellular senescence. *Aging Cell* **15**, 999–1017 (2016).
9. Rossiello, F., Herbig, U., Longhese, M. P., Fumagalli, M. & D'Adda, D. F. F. Irreparable telomeric DNA damage and persistent DDR signalling as a shared causative mechanism of cellular senescence and ageing. *Curr. Opin. Genet. Dev.* **26**, 89–95 (2014).
10. Munoz-Espin, D. et al. Programmed cell senescence during mammalian embryonic development. *Cell* **155**, 1104–1118 (2013).
11. Hernandez-Segura, A., Nehme, J. & Demaria, M. Hallmarks of cellular senescence. *Trends Cell Biol.* **28**, 436–453 (2018).
12. Han, X. et al. Construction of a human cell landscape at single-cell level. *Nature* **581**, 303–309 (2020).
13. He, S. et al. Single-cell transcriptome profiling of an adult human cell atlas of 15 major organs. *Genome Biol.* **21**, 294 (2020).
14. Lafferty-Whyte, K. et al. Scoring of senescence signalling in multiple human tumour gene expression datasets, identification of a correlation between senescence score and drug toxicity in the NCI60 panel and a pro-inflammatory signature correlating with survival advantage in peritoneal mesothelioma. *BMC Genom.* **11**, 532 (2010).
15. Calcinotto, A. et al. Cellular senescence: aging, cancer, and injury. *Physiol. Rev.* **99**, 1047–1078 (2019).
16. Gorgoulis, V. et al. Cellular senescence: defining a path forward. *Cell* **179**, 813–827 (2019).
17. Wiley, C. D. et al. Analysis of individual cells identifies cell-to-cell variability following induction of cellular senescence. *Aging Cell* **16**, 1043–1050 (2017).
18. de Magalhaes, J. P., Curado, J. & Church, G. M. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics* **25**, 875–881 (2009).
19. Saul, D. et al. A new gene set identifies senescent cells and predicts senescence-associated pathways across tissues. *Nat. Commun.* **13**, 4827 (2022).
20. Wang, X. et al. Comprehensive assessment of cellular senescence in the tumor microenvironment. *Brief. Bioinform.* **23**, bbac118 (2022).
21. Reyfman, P. A. et al. Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am. J. Resp. Crit. Care* **199**, 1517–1536 (2019).
22. Nighat, N., Zhenqing, Y., Yidong, C., Xiaojing, W. & Siyuan, Z. Signature-scoring methods developed for bulk samples are not adequate for cancer single-cell RNA sequencing data. *eLife* https://doi.org/10.7554/eLife.71994 (2022).
23. Gong, Q., Jiang, Y., Xiong, J., Liu, F. & Guan, J. Integrating scRNA and bulk-RNA sequencing develops a cell senescence signature for analyzing tumor heterogeneity in clear cell renal cell carcinoma. *Front. Immunol.* **14**, 1199002 (2023).
24. Jochems, F. et al. The Cancer SENESCopedia: a delineation of cancer cell senescence. *Cell Rep.* **36**, 109441 (2021).
25. Tao, W., Yu, Z. & Han, J. J. Single-cell senescence identification reveals senescence heterogeneity, trajectory, and modulators. *Cell Metab.* **36**, 1126–1143 (2024).
26. Sun, Y. An updated landscape of cellular senescence heterogeneity: mechanisms, technologies and senotherapies. *Transl. Med. Aging* **7**, 46–51 (2023).
27. Gil, J. The challenge of identifying senescent cells. *Nat. Cell Biol.* **25**, 1554–1556 (2023).
28. Mauceri, S., Sweeney, J., Nicolau, M. & McDermott, J. Feature extraction by grammatical evolution for one-class time series classification. *Genet. Program Evolable Mach.* **22**, 267–295 (2021).
29. Sokolov, A., Paull, E. O. & Stuart, J. M. One-class detection of cell states in tumor subtypes. *Pac. Symp. Biocomput.* **21**, 405–416 (2016).
30. Park, Y. J. et al. Senescent melanocytes driven by glycolytic changes are characterized by melanosome transport dysfunction. *Theranostics* **13**, 3914–3924 (2023).
31. Gallanis, G. T. et al. Stromal senescence following treatment with the CDK4/6 inhibitor palbociclib alters the lung metastatic niche and increases metastasis of drug-resistant mammary cancer cells. *Cancers* **15**, 1908 (2023).
32. Marin, I. et al. Cellular senescence is immunogenic and promotes antitumor immunity. *Cancer Discov.* **13**, 410–431 (2023).
33. Chen, F. et al. Silencing circSERPINE2 restrains mesenchymal stem cell senescence via the YBX3/PCNA/p21 axis. *Cell. Mol. Life Sci.* **80**, 325 (2023).
34. Roh, K. et al. Lysosomal control of senescence and inflammation through cholesterol partitioning. *Nat. Metab.* **5**, 398–413 (2023).
35. Cho, H. et al. Nintedanib induces senolytic effect via STAT3 inhibition. *Cell Death Dis.* **13**, 760 (2022).
36. Da Silveira, W. A., Renaud, L., Hazard, E. S. & Hardiman, G. miRNA and lncRNA expression networks modulate cell cycle and DNA repair inhibition in senescent prostate cells. *Genes (Basel)* **13**, 208 (2022).
37. Sturmlechner, I. et al. Senescent cells limit p53 activity via multiple mechanisms to remain viable. *Nat. Commun.* **13**, 3722 (2022).
38. Urata, R. et al. Senescent endothelial cells are predisposed to SARS-CoV-2 infection and subsequent endothelial dysfunction. *Sci. Rep.* **12**, 11855 (2022).
39. Goncalves, S. et al. COX2 regulates senescence secretome composition and senescence surveillance through PGE_2. *Cell Rep.* **34**, 108860 (2021).
40. St-Jean, S. et al. NCOR1 sustains colorectal cancer cell growth and protects against cellular senescence. *Cancers* **13**, 4414 (2021).
41. Uryga, A. K. et al. Telomere damage promotes vascular smooth muscle cell senescence and immune cell recruitment after vessel injury. *Commun. Biol.* **4**, 611 (2021).
42. Martinez-Zamudio, R. I. et al. Senescence-associated beta-galactosidase reveals the abundance of senescent $CD8^+$ T cells in aging humans. *Aging Cell* **20**, e13344 (2021).
43. DePianto, D. J. et al. Molecular mapping of interstitial lung disease reveals a phenotypically distinct senescent basal epithelial cell population. *JCI Insight* **6**, e143626 (2021).

44. Soshnikova, N. V. et al. PHF10 subunit of PBAF complex mediates transcriptional activation by MYC. *Oncogene* **40**, 6071–6080 (2021).
45. Deryabin, P. et al. 'All-in-one' genetic tool assessing endometrial receptivity for personalized screening of female sex steroid hormones. *Front. Cell Dev. Biol.* **9**, 624053 (2021).
46. Kolesnichenko, M. et al. Transcriptional repression of *NFKBIA* triggers constitutive IKK- and proteasome-independent p65/RelA activation in senescence. *EMBO J.* **40**, e104296 (2021).
47. Schwartz, R. E. et al. Insights into epithelial cell senescence from transcriptome and secretome analysis of human oral keratinocytes. *Aging* **13**, 4747–4777 (2021).
48. Wang, R. W., Vigano, S., Ben-David, U., Amon, A. & Santaguida, S. Aneuploid senescent cells activate NF-κB to promote their immune clearance by NK cells. *EMBO Rep.* **22**, e52032 (2021).
49. Zhang, X. et al. The loss of heterochromatin is associated with multiscale three-dimensional genome reorganization and aberrant transcription during cellular senescence. *Genome Res.* **31**, 1121–1135 (2021).
50. Zhang, B. et al. KDM4 orchestrates epigenomic remodeling of senescent cells and potentiates the senescence-associated secretory phenotype. *Nat. Aging* **1**, 454–472 (2021).
51. Sturmlechner, I. et al. p21 produces a bioactive secretome that places stressed cells under immunosurveillance. *Science* **374**, eabb3420 (2021).
52. Gu, J. et al. IL-6 derived from therapy-induced senescence facilitates the glycolytic phenotype in glioblastoma cells. *Am. J. Cancer Res.* **11**, 458–478 (2021).
53. Sati, S. et al. 4D genome rewiring during oncogene-induced and replicative senescence. *Mol. Cell* **78**, 522–538 (2020).
54. Chan, K. T. et al. A functional genetic screen defines the AKT-induced senescence signaling network. *Cell Death Differ.* **27**, 725–741 (2020).
55. Han, L. et al. Senescent stromal cells promote cancer resistance through SIRT1 loss-potentiated overproduction of small extracellular vesicles. *Cancer Res.* **80**, 3383–3398 (2020).
56. Simmnacher, K. et al. Unique signatures of stress-induced senescent human astrocytes. *Exp. Neurol.* **334**, 113466 (2020).
57. Wang, S. et al. Rescue of premature aging defects in Cockayne syndrome stem cells by CRISPR/Cas9-mediated gene correction. *Protein Cell* **11**, 1–22 (2020).
58. Zhao, Z. et al. Dynamic transcriptome profiling in DNA damage-induced cellular senescence and transient cell-cycle arrest. *Genomics* **112**, 1309–1317 (2020).
59. Casella, G. et al. Transcriptome signature of cellular senescence. *Nucleic Acids Res.* **47**, 7294–7305 (2019).
60. Pantazi, A. et al. Inhibition of the 60S ribosome biogenesis GTPase LSG1 causes endoplasmic reticular disruption and cellular senescence. *Aging Cell* **18**, e12981 (2019).
61. Mongiardi, M. P. et al. Axitinib exposure triggers endothelial cells senescence through ROS accumulation and ATM activation. *Oncogene* **38**, 5413–5424 (2019).
62. Bianco, J. N. et al. Overexpression of Claspin and Timeless protects cancer cells from replication stress in a checkpoint-independent manner. *Nat. Commun.* **10**, 910 (2019).
63. Wang, C. et al. Inducing and exploiting vulnerabilities for the treatment of liver cancer. *Nature* **574**, 268–272 (2019).
64. Fu, L. et al. Up-regulation of FOXD1 by YAP alleviates senescence and osteoarthritis. *PLoS Biol.* **17**, e3000201 (2019).
65. De Cecco, M. et al. L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature* **566**, 73–78 (2019).
66. Lau, L., Porciuncula, A., Yu, A., Iwakura, Y. & David, G. Uncoupling the senescence-associated secretory phenotype from cell cycle exit via interleukin-1 inactivation unveils its protumorigenic role. *Mol. Cell. Biol.* **39**, e00586–18 (2019).
67. Zirkel, A. et al. HMGB2 loss upon senescence entry disrupts genomic organization and induces CTCF clustering across cell types. *Mol. Cell* **70**, 730–744 (2018).
68. Ruscetti, M. et al. NK cell-mediated cytotoxicity contributes to tumor control by a cytostatic drug combination. *Science* **362**, 1416–1422 (2018).
69. Ohashi, M. et al. Loss of MECP2 leads to activation of P53 and neuronal senescence. *Stem Cell Reports* **10**, 1453–1463 (2018).
70. Raman, A. T. et al. Apparent bias toward long gene misregulation in MeCP2 syndromes disappears after controlling for baseline variations. *Nat. Commun.* **9**, 3225 (2018).
71. Wang, S. et al. ATF6 safeguards organelle homeostasis and cellular aging in human mesenchymal stem cells. *Cell Discov.* **4**, 2 (2018).
72. Saunderson, E. A. et al. Hit-and-run epigenetic editing prevents senescence entry in primary breast cells from healthy donors. *Nat. Commun.* **8**, 1450 (2017).
73. Muniz, L. et al. Control of gene expression in senescence through transcriptional read-through of convergent protein-coding genes. *Cell Rep.* **21**, 2433–2446 (2017).
74. Yang, J. et al. Genetic enhancement in cultured human adult stem cells conferred by a single nucleotide recoding. *Cell Res.* **27**, 1178–1181 (2017).
75. Santaguida, S. et al. Chromosome mis-segregation generates cell-cycle-arrested cells with complex karyotypes that are eliminated by the immune system. *Dev. Cell* **41**, 638–651 (2017).
76. Lenain, C. et al. Massive reshaping of genome–nuclear lamina interactions during oncogene-induced senescence. *Genome Res.* **27**, 1634–1644 (2017).
77. Kovatcheva, M. et al. ATRX is a regulator of therapy induced senescence in human cells. *Nat. Commun.* **8**, 386 (2017).
78. Aarts, M. et al. Coupling shRNA screens with single-cell RNA-seq identifies a dual role for mTOR in reprogramming-induced senescence. *Genes Dev.* **31**, 2085–2098 (2017).
79. Wang, L. et al. High-throughput functional genetic and compound screens identify targets for senescence induction in cancer. *Cell Rep.* **21**, 773–783 (2017).
80. Tasdemir, N. et al. BRD4 connects enhancer remodeling to senescence immune surveillance. *Cancer Discov.* **6**, 612–629 (2016).
81. Hoare, M. et al. NOTCH1 mediates a switch between two distinct secretomes during senescence. *Nat. Cell Biol.* **18**, 979–992 (2016).
82. Marthandan, S. et al. Conserved senescence associated genes and pathways in primary human fibroblasts detected by RNA-seq. *PLoS ONE* **11**, e154531 (2016).
83. Crowe, E. P. et al. Changes in the transcriptome of human astrocytes accompanying oxidative stress-induced senescence. *Front. Aging Neurosci.* **8**, 208 (2016).
84. Nelson, D. M. et al. Mapping H4K20me3 onto the chromatin landscape of senescent cells indicates a function in control of cell senescence and tumor suppression through preservation of genetic and epigenetic stability. *Genome Biol.* **17**, 158 (2016).
85. Marthandan, S. et al. Hormetic effect of rotenone in primary human fibroblasts. *Immun. Ageing* **12**, 11 (2015).
86. Marthandan, S. et al. Similarities in gene expression profiles during in vitro aging of primary human embryonic lung and foreskin fibroblasts. *Biomed. Res. Int.* **2015**, 731938 (2015).
87. Herranz, N. et al. mTOR regulates MAPKAPK2 translation to control the senescence-associated secretory phenotype. *Nat. Cell Biol.* **17**, 1205–1217 (2015).
88. Purcell, M., Kruger, A. & Tainsky, M. A. Gene expression profiling of replicative and induced senescence. *Cell Cycle* **13**, 3927–3937 (2014).
89. Alspach, E. et al. p38MAPK plays a crucial role in stromal-mediated tumorigenesis. *Cancer Discov.* **4**, 716–729 (2014).

90. Rai, T. S. et al. HIRA orchestrates a dynamic chromatin landscape in senescence and is required for suppression of neoplasia. *Genes Dev.* **28**, 2712–2725 (2014).
91. Neyret-Kahn, H. et al. Sumoylation at chromatin governs coordinated repression of a transcriptional program essential for cell growth and proliferation. *Genome Res.* **23**, 1563–1579 (2013).
92. Hernandez-Segura, A. et al. Unmasking transcriptional heterogeneity in senescent cells. *Curr. Biol.* **27**, 2652–2660 (2017).
93. Russo, T. et al. The SATB1-MIR22-GBA axis mediates glucocerebroside accumulation inducing a cellular senescence-like phenotype in dopaminergic neurons. *Aging Cell* **23**, e14077 (2024).
94. Lin, C. et al. Arginine hypomethylation-mediated proteasomal degradation of histone H4—an early biomarker of cellular senescence. *Cell Death Differ.* **27**, 2697–2709 (2020).
95. Gao, J. et al. Mutual regulation between GDF11 and TET2 prevents senescence of mesenchymal stem cells. *J. Cell. Physiol.* **238**, 2827–2840 (2023).
96. Zhang, X. et al. Telomere-dependent and telomere-independent roles of RAP1 in regulating human stem cell homeostasis. *Protein Cell* **10**, 649–667 (2019).
97. Balakrishnan, I. et al. Senescence induced by BMI1 inhibition is a therapeutic vulnerability in H3K27M-mutant DIPG. *Cell Rep.* **33**, 108286 (2020).
98. Birger, A. et al. Human iPSC-derived astrocytes from ALS patients with mutated C9ORF72 show increased oxidative stress and neurotoxicity. *EBioMedicine* **50**, 274–289 (2019).
99. Wu, J. et al. Towards open vocabulary learning: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **46**, 5092–5113 (2024).
100. Georgilis, A. et al. PTBP1-mediated alternative splicing regulates the inflammatory secretome and the pro-tumorigenic effects of senescent cells. *Cancer Cell* **34**, 85–102 (2018).
101. Guerrero, A. et al. Cardiac glycosides are broad-spectrum senolytics. *Nat. Metab.* **1**, 1074–1088 (2019).
102. Parry, A. J. et al. NOTCH-mediated non-cell autonomous regulation of chromatin structure during senescence. *Nat. Commun.* **9**, 1840 (2018).
103. Somekh, J., Shen-Orr, S. S. & Kohane, I. S. Batch correction evaluation framework using a priori gene–gene associations: applied to the GTEx dataset. *BMC Bioinform.* **20**, 268 (2019).
104. Zhang, X. et al. CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Res.* **47**, D721–D728 (2019).
105. Chatsirisupachai, K., Palmer, D., Ferreira, S. & de Magalhaes, J. P. A human tissue-specific transcriptomic analysis reveals a complex relationship between aging, cancer, and cellular senescence. *Aging Cell* **18**, e13041 (2019).
106. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
107. Ashburner, M. et al. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
108. Aleksander, S. A. et al. The Gene Ontology knowledgebase in 2023. *Genetics* **224**, iyad031 (2023).
109. Croft, D. et al. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* **39**, D691–D697 (2011).
110. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
111. Kim, K. S., Kang, K. W., Seu, Y. B., Baek, S. H. & Kim, J. R. Interferon-γ induces cellular senescence through p53-dependent DNA damage signaling in human endothelial cells. *Mech. Ageing Dev.* **130**, 179–188 (2009).
112. Cisowski, J., Sayin, V. I., Liu, M., Karlsson, C. & Bergo, M. O. Oncogene-induced senescence underlies the mutual exclusive nature of oncogenic KRAS and BRAF. *Oncogene* **35**, 1328–1333 (2016).
113. Lasry, A. & Ben-Neriah, Y. Senescence-associated inflammatory responses: aging and cancer perspectives. *Trends Immunol.* **36**, 217–228 (2015).
114. Artandi, S. E. & Attardi, L. D. Pathways connecting telomeres and p53 in senescence, apoptosis, and cancer. *Biochem. Biophys. Res. Commun.* **331**, 881–890 (2005).
115. Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic *ras* provokes premature cell senescence associated with accumulation of p53 and $p16^{INK4a}$. *Cell* **88**, 593–602 (1997).
116. Oshi, M. et al. G2M cell cycle pathway score as a prognostic biomarker of metastasis in estrogen receptor (ER)-positive breast cancer. *Int. J. Mol. Sci.* **21**, 2921 (2020).
117. Narita, M. et al. Rb-mediated heterochromatin formation and silencing of E2F target genes during cellular senescence. *Cell* **113**, 703–716 (2003).
118. Dikovskaya, D. et al. Mitotic stress is an integral part of the oncogene-induced senescence program that promotes multinucleation and cell cycle arrest. *Cell Rep.* **12**, 1483–1496 (2015).
119. Wu, C. H. et al. Cellular senescence is an important mechanism of tumor regression upon c-Myc inactivation. *Proc. Natl Acad. Sci. USA* **104**, 13028–13033 (2007).
120. Tacutu, R. et al. Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Res.* **46**, D1083–D1090 (2018).
121. Saul, D. & Kosinsky, R. L. Single-cell transcriptomics reveals the expression of aging- and senescence-associated genes in distinct cancer cell populations. *Cells* **10**, 3126 (2021).
122. Consortium, A. A. Aging Atlas: a multi-omics database for aging biology. *Nucleic Acids Res.* **49**, D825–D830 (2020).
123. Park, H. S. & Kim, S. Y. Endothelial cell senescence: a machine learning-based meta-analysis of transcriptomic studies. *Ageing Res. Rev.* **65**, 101213 (2021).
124. Chicas, A. et al. Dissecting the unique role of the retinoblastoma tumor suppressor during cellular senescence. *Cancer Cell* **17**, 376–387 (2010).
125. Garbe, J. C. et al. Molecular distinctions between stasis and telomere attrition senescence barriers shown by long-term culture of normal human mammary epithelial cells. *Cancer Res.* **69**, 7557–7568 (2009).
126. Orfanidis, K., Waster, P., Lundmark, K., Rosdahl, I. & Ollinger, K. Evaluation of tubulin β-3 as a novel senescence-associated gene in melanocytic malignant transformation. *Pigment Cell Melanoma Res.* **30**, 243–254 (2017).
127. Krizhanovsky, V. et al. Senescence of activated stellate cells limits liver fibrosis. *Cell* **134**, 657–667 (2008).
128. Yuan, L. et al. Switching off IMMP2L signaling drives senescence via simultaneous metabolic alteration and blockage of cell death. *Cell Res.* **28**, 625–643 (2018).
129. Costarelli, L. et al. Different transcriptional profiling between senescent and non-senescent human coronary artery endothelial cells (HCAECs) by Omeprazole and Lansoprazole treatment. *Biogerontology* **18**, 217–236 (2017).
130. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
131. Tabula Muris Consortium. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature* **583**, 590–595 (2020).
132. Karin, O., Agrawal, A., Porat, Z., Krizhanovsky, V. & Alon, U. Senescent cell turnover slows with age providing an explanation for the Gompertz law. *Nat. Commun.* **10**, 5495 (2019).
133. Fleischer, J. G. et al. Predicting age from the transcriptome of human dermal fibroblasts. *Genome Biol.* **19**, 221 (2018).
134. Teo, Y. V. et al. Notch signaling mediates secondary senescence. *Cell Rep.* **27**, 997–1007 (2019).

135. Tang, H. et al. Single senescent cell sequencing reveals heterogeneity in senescent cells induced by telomere erosion. *Protein Cell* **10**, 370–375 (2019).
136. Wechter, N. et al. Single-cell transcriptomic analysis uncovers diverse and dynamic senescent cell populations. *Aging* **15**, 2824–2851 (2023).
137. Enge, M. et al. Single-cell analysis of human pancreas reveals transcriptional signatures of aging and somatic mutation patterns. *Cell* **171**, 321–330 (2017).
138. Ashraf, H. M., Fernandez, B. & Spencer, S. L. The intensities of canonical senescence biomarkers integrate the duration of cell-cycle withdrawal. *Nat. Commun.* **14**, 4527 (2023).
139. Replogle, J. M. et al. Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell* **185**, 2559–2575 (2022).
140. Wu, H. et al. Improving dermal fibroblast-to-epidermis communications and aging wound repair through extracellular vesicle-mediated delivery of *Gstm2* mRNA. *J. Nanobiotechnology* **22**, 307 (2024).
141. Sun, M. et al. Targeting the chromosomal passenger complex subunit INCENP induces polyploidization, apoptosis, and senescence in neuroblastoma. *Cancer Res.* **79**, 4937–4950 (2019).
142. Nehme, J., Borghesan, M., Mackedenski, S., Bird, T. G. & Demaria, M. Cellular senescence as a potential mediator of COVID-19 severity in the elderly. *Aging Cell* **19**, e13237 (2020).
143. Lipskaia, L. et al. Evidence that SARS-CoV-2 induces lung cell senescence: potential impact on COVID-19 lung disease. *Am. J. Resp. Cell Mol.* **66**, 107–111 (2022).
144. Melms, J. C. et al. A molecular single-cell lung atlas of lethal COVID-19. *Nature* **595**, 114–119 (2021).
145. Schmitt, C. A. et al. COVID-19 and cellular senescence. *Nat. Rev. Immunol.* **23**, 251–263 (2023).
146. Jenssen, R. & Eltoft, T. A new information theoretic analysis of sum-of-squared-error kernel clustering. *Neurocomputing* **72**, 23–31 (2008).
147. Burton, D. & Stolzing, A. Cellular senescence: immunosurveillance and future immunotherapy. *Ageing Res. Rev.* **43**, 17–25 (2018).
148. Lo, J. A. & Fisher, D. E. The melanoma revolution: from UV carcinogenesis to a new era in therapeutics. *Science* **346**, 945–949 (2014).
149. Robert, C. et al. Nivolumab in previously untreated melanoma without *BRAF* mutation. *N. Engl. J. Med.* **372**, 320–330 (2015).
150. Hoenicke, L. & Zender, L. Immune surveillance of senescent cells—biological significance in cancer- and non-cancer pathologies. *Carcinogenesis* **33**, 1123–1126 (2012).
151. Kang, T. W. et al. Senescence surveillance of pre-malignant hepatocytes limits liver cancer development. *Nature* **479**, 547–551 (2011).
152. Xue, W. et al. Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas. *Nature* **445**, 656–660 (2007).
153. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
154. Jerby-Arnon, L. et al. A cancer cell program promotes T cell exclusion and resistance to checkpoint blockade. *Cell* **175**, 984–997 (2018).
155. Teschendorff, A. E. & Enver, T. Single-cell entropy for accurate estimation of differentiation potency from a cell's transcriptome. *Nat. Commun.* **8**, 15599 (2017).
156. Wang, W. et al. Independent component analysis based gene co-expression network inference (ICAnet) to decipher functional modules for better single-cell clustering and batch integration. *Nucleic Acids Res.* **49**, e54 (2021).
157. Moon, K. R. et al. Visualizing structure and transitions in high-dimensional biological data. *Nat. Biotechnol.* **37**, 1482–1492 (2019).
158. Teschendorff, A. E., Breeze, C. E., Zheng, S. C. & Beck, S. A comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies. *BMC Bioinform.* **18**, 105 (2017).
159. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
160. de Visser, K. E. & Joyce, J. A. The evolving tumor microenvironment: from cancer initiation to metastatic outgrowth. *Cancer Cell* **41**, 374–403 (2023).
161. Biran, A. et al. Senescent cells communicate via intercellular protein transfer. *Genes Dev.* **29**, 791–802 (2015).
162. Jin, S. et al. Inference and analysis of cell–cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).
163. Senturk, S. et al. Transforming growth factor-beta induces senescence in hepatocellular carcinoma cells and inhibits tumor growth. *Hepatology* **52**, 966–974 (2010).
164. Reimann, M. et al. Tumor stroma-derived TGF-β limits Myc-driven lymphomagenesis via Suv39h1-dependent senescence. *Cancer Cell* **17**, 262–272 (2010).
165. Chandra, J. B., Sarkar, S., Rout, L. & Mandal, M. The transformation of cancer-associated fibroblasts: current perspectives on the role of TGF-β in CAF mediated tumor progression and therapeutic resistance. *Cancer Lett.* **520**, 222–232 (2021).
166. Korbecki, J. et al. CC chemokines in a tumor: a review of pro-cancer and anti-cancer properties of the ligands of receptors CCR1, CCR2, CCR3, and CCR4. *Int. J. Mol. Sci.* **21**, 8412 (2020).
167. Price, M. A. et al. CSPG4, a potential therapeutic target, facilitates malignant progression of melanoma. *Pigment Cell Melanoma Res.* **24**, 1148–1157 (2011).
168. Gurrea-Rubio, M. & Fox, D. A. The dual role of CD6 as a therapeutic target in cancer and autoimmune disease. *Front. Med.* **9**, 1026521 (2022).
169. Zhang, C. et al. STAT3 activation-induced fatty acid oxidation in $CD8^+$ T effector cells is critical for obesity-promoted breast tumor growth. *Cell Metab.* **31**, 148–161 (2020).
170. Donnelly, O. G. et al. Measles virus causes immunogenic cell death in human melanoma. *Gene Ther.* **20**, 7–15 (2013).
171. Hughes, B. K., Wallis, R. & Bishop, C. L. Yearning for machine learning: applications for the classification and characterisation of senescence. *Cell Tissue Res.* **394**, 1–16 (2023).
172. Ma, S. et al. Spatial transcriptomic landscape unveils immunoglobin-associated senescence as a hallmark of aging. *Cell* **187**, 7025–7044 (2024).
173. Heimberg, G. et al. A cell atlas foundation model for scalable search of similar human cells. *Nature* **638**, 1085–1094 (2024).
174. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
175. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
176. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
177. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).
178. Gennady, K., Vladimir, S. & Alexey, S. Fast gene set enrichment analysis. Preprint at *bioRxiv* https://doi.org/10.1101/060012 (2019).
179. Zhang, S. et al. Cellular senescence-related gene signature as a valuable predictor of prognosis in hepatocellular carcinoma. *Aging* **15**, 3064–3093 (2023).
180. Yeung, K. Y., Fraley, C., Murua, A., Raftery, A. E. & Ruzzo, W. L. Model-based clustering and data transformations for gene expression data. *Bioinformatics* **17**, 977–987 (2001).

181. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
182. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
183. Colaprico, A. et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44**, e71 (2016).
184. Zheng, S. C., Breeze, C. E., Beck, S. & Teschendorff, A. E. Identification of differentially methylated cell types in epigenome-wide association studies. *Nat. Methods* **15**, 1059–1066 (2018).

Acknowledgements

The design and launching of this study were financially supported by the National Key Research and Development Program of China (2023YFC3603300, awarded to T.N.). The National Natural Science Foundation of China (92249302, awarded to T.N.; 32370592, awarded to T.N.) financially supported continuous investigation of this study. The preparation of the manuscript and the decision to publish were financially supported by the Shanghai Municipal Science and Technology Major Project (grant no. 2023SHZDZX02, awarded to T.N.).

Author contributions

The paper was written by J.W., W.W. and J.Y. and was revised by G.W., W.W. and T.N. The method was conceived by W.W., J.Y., J.W. and T.N. The algorithm was implemented by J.W. and X.Z. The algorithm evaluations and computational analyses were conducted by J.W., X.Z. and W.W. The experiments were performed and analyzed by P.Y., P.G. and Q.X. The public data used were collected by X.Z., Y.Z., G.W., Q.L. and X.Z. This work was supervised by T.N., W.W. and G.W.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s43587-025-00886-2.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43587-025-00886-2.

Correspondence and requests for materials should be addressed to Gang Wei, Weixu Wang or Ting Ni.

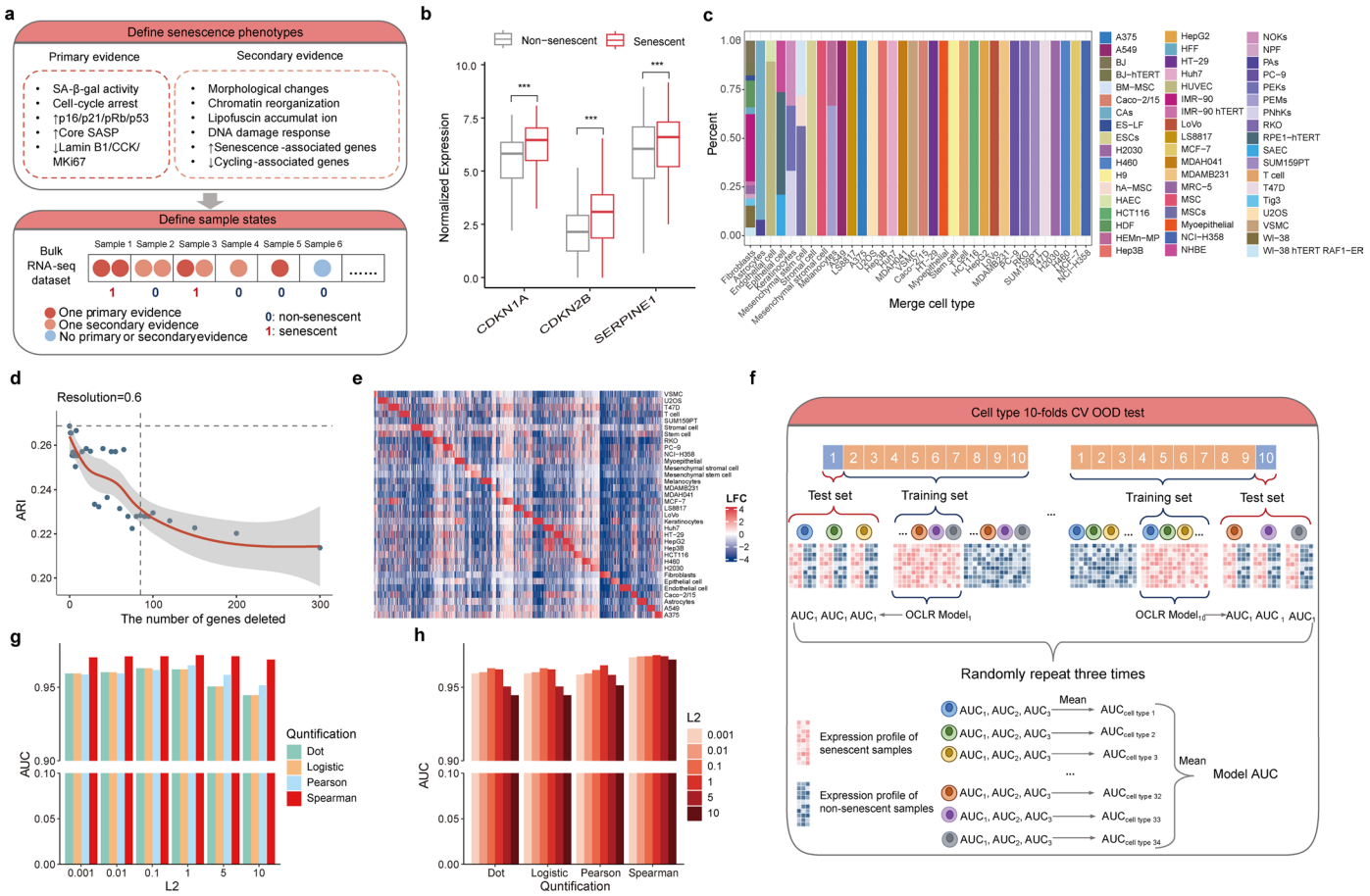
Peer review information *Nature Aging* thanks Jesus Gil and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

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

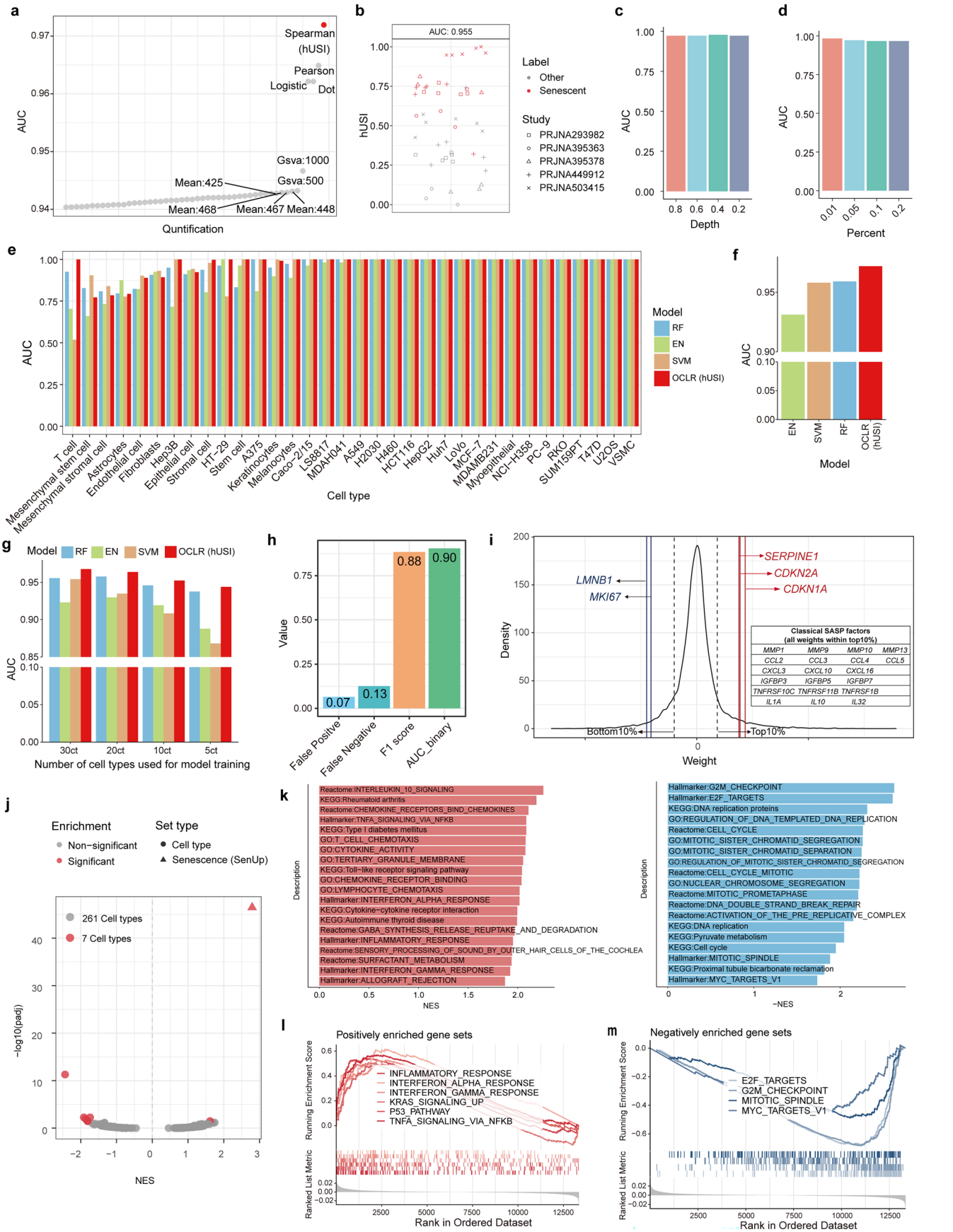
[1]State Key Laboratory of Genetics and Development of Complex Phenotypes, National Clinical Research Center for Aging and Medicine, Huashan Hospital, Collaborative Innovation Center of Genetics and Development, Human Phenome Institute, Center for Evolutionary Biology, Shanghai Engineering Research Center of Industrial Microorganisms, School of Life Sciences, Fudan University, Shanghai, China. [2]State Key Laboratory of Reproductive Regulation and Breeding of Grassland Livestock, Institutes of Biomedical Sciences, School of Life Sciences, Inner Mongolia University, Hohhot, China. [3]Present address: Department of Cardiology, Zhongshan Hospital, Fudan University, Shanghai Institute of Cardiovascular Diseases, Shanghai, China. [4]Present address: Department of Data System, 3D Medicines, Inc., Shanghai, China. [5]Present address: Institute of Computational Biology, Helmholtz Center Munich, Munich, Germany. [6]Present address: TUM School of Life Sciences Weihenstephan, Technical University of Munich, Munich, Germany. [7]These authors contributed equally: Jing Wang, Xiaolan Zhou, Peng Yu. ✉e-mail: gwei@fudan.edu.cn; weixu.wang@helmholtz-munich.de; tingni@fudan.edu.cn

https://doi.org/10.1038/s43587-025-00886-2



Extended Data Fig. 1 | Details of the dataset used for model training and parameter testing. a. Criteria for selecting senescent samples in the training set. **b**. Boxplot showing that senescent samples included in the training set have significantly higher expression of *CDKN1A*, *CDKN2B*, and *SERPINE1* compared to non-senescent samples (n = 770 samples; 385 senescent samples and 385 non-senescent samples; two-tailed Wilcoxon test; *** denotes p < 1e-3; The exact p values are reported in Source Data for Extended Data Fig. 1). **c**. Histogram showing the proportion of cell lines within each cell type, explicitly excluding merging of cancer cell lines due to their high heterogeneity. **d**. ARI values plotted against numbers of removed cell-type-specific genes at a clustering resolution of 0.6, with the elbow point at 85 highlighted by a dashed line. **e**. Log2 fold change of the top 85 cell-type-specific genes across various cell types. **f**. Conceptual diagram illustrating the 10-folds cross-validation OOD (10-folds CV OOD) test methodology (Supplementary Information Note 3). **g**, **h**. AUC values for models using different L2 values and quantification methods in the OOD test.

https://doi.org/10.1038/s43587-025-00886-2

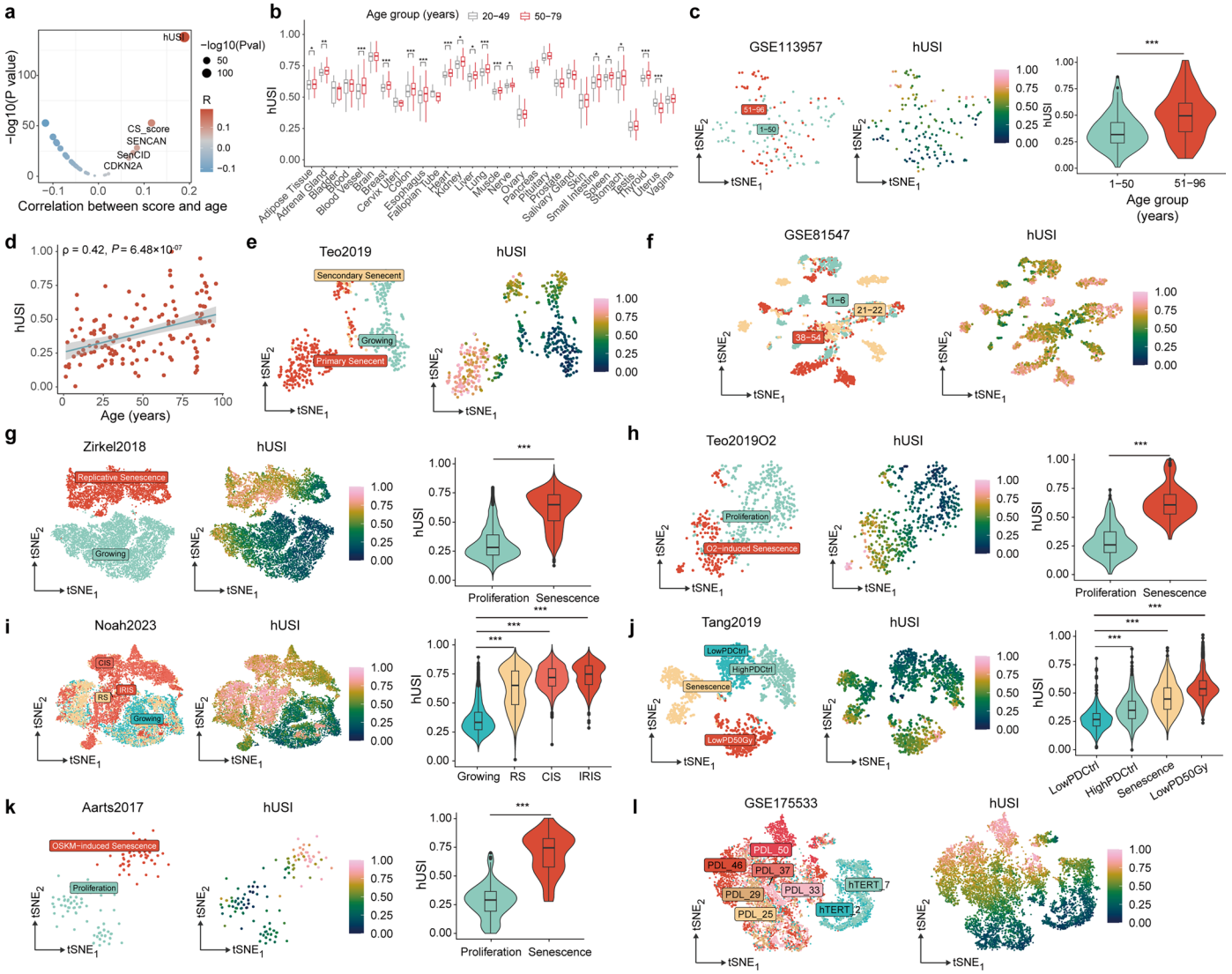


Extended Data Fig. 2 | See next page for caption.

https://doi.org/10.1038/s43587-025-00886-2

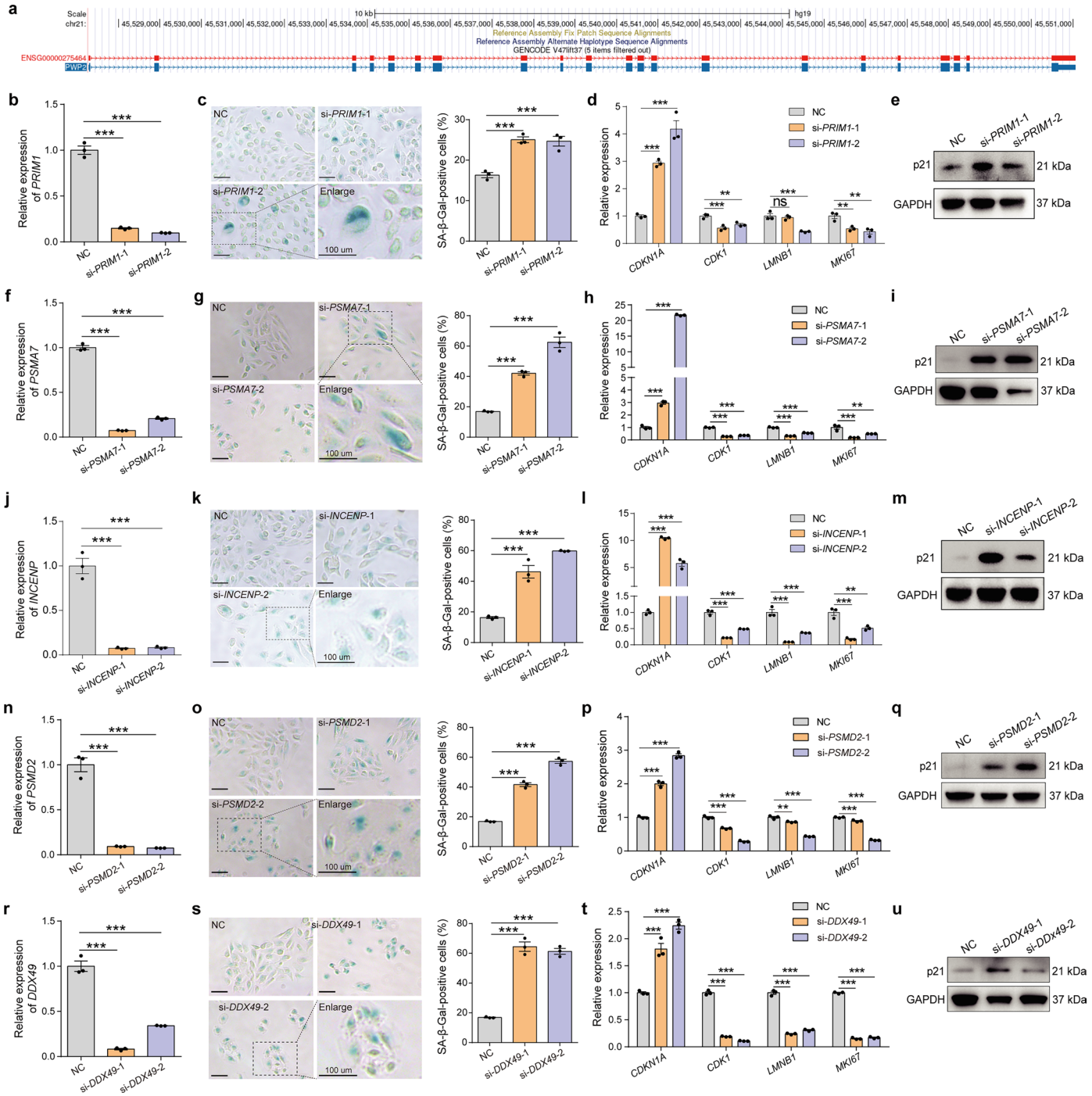
Extended Data Fig. 2 | Performance comparison of models and distribution of the OCLR-learned features. a. AUC values for models with L2 = 1 across various quantification methods (Dot, Logistic, Pearson, Spearman, Gsva, and Mean). Only the top 50 methods were displayed and only the top 10 were labeled. The red points represent the best performed methods—Spearman correlation—which is used for hUSI computation. **b**. AUC values of hUSI applied to integrated bulk gene expression profiles from IMR90 cells across five studies (PRJNA395363, PRJNA395378, PRJNA503415, PRJNA449912, PRJNA293984), distinguishing non-senescent (grey) and OIS (red) samples. **c, d**. Bar plots showing the mean AUC values across five studies (mentioned in (**b**)) with different sparsity levels (**c**) and outlier percentages (**d**) in gene expression profiles. **e**, **f**. The AUC values of OCLR versus RF, EN, and SVM models in OOD test, both overall (**f**) and by cell type (**e**). **g**. The AUC values of OCLR and two-class models (RF, EN, SVM) with different cell type numbers. **h**. The false positive rate, false negative rate, F1 score and AUC of SSE binary hUSI in 10-folds CV OOD test. **i**. Distribution of gene weights in OCLR-learned senescence features, highlighting genes related to senescence (red), cycling (blue), and SASP. **j**. Enrichment analysis of OCLR-learned senescence features on 268 cell type marker sets from the CellMark database and one senescence-associated gene set (SenUP) for comparison. P-values were calculated by permutation test (n = 1,000 iterations) using GSEA function, and adjusted by the Benjamini-Hochberg (BH) method. **k**. Bar plots exhibiting gene sets (top 5 for each database: HALLMARKER, GO, KEGG, and Reactome) positively (left) and negatively (right) enriched by the OCLR-learned senescence features. **l, m**. OCLR-learned senescence features were significantly positively enriched in senescence associated gene sets (**l**) and negatively enriched in proliferation gene sets (**m**).

https://doi.org/10.1038/s43587-025-00886-2



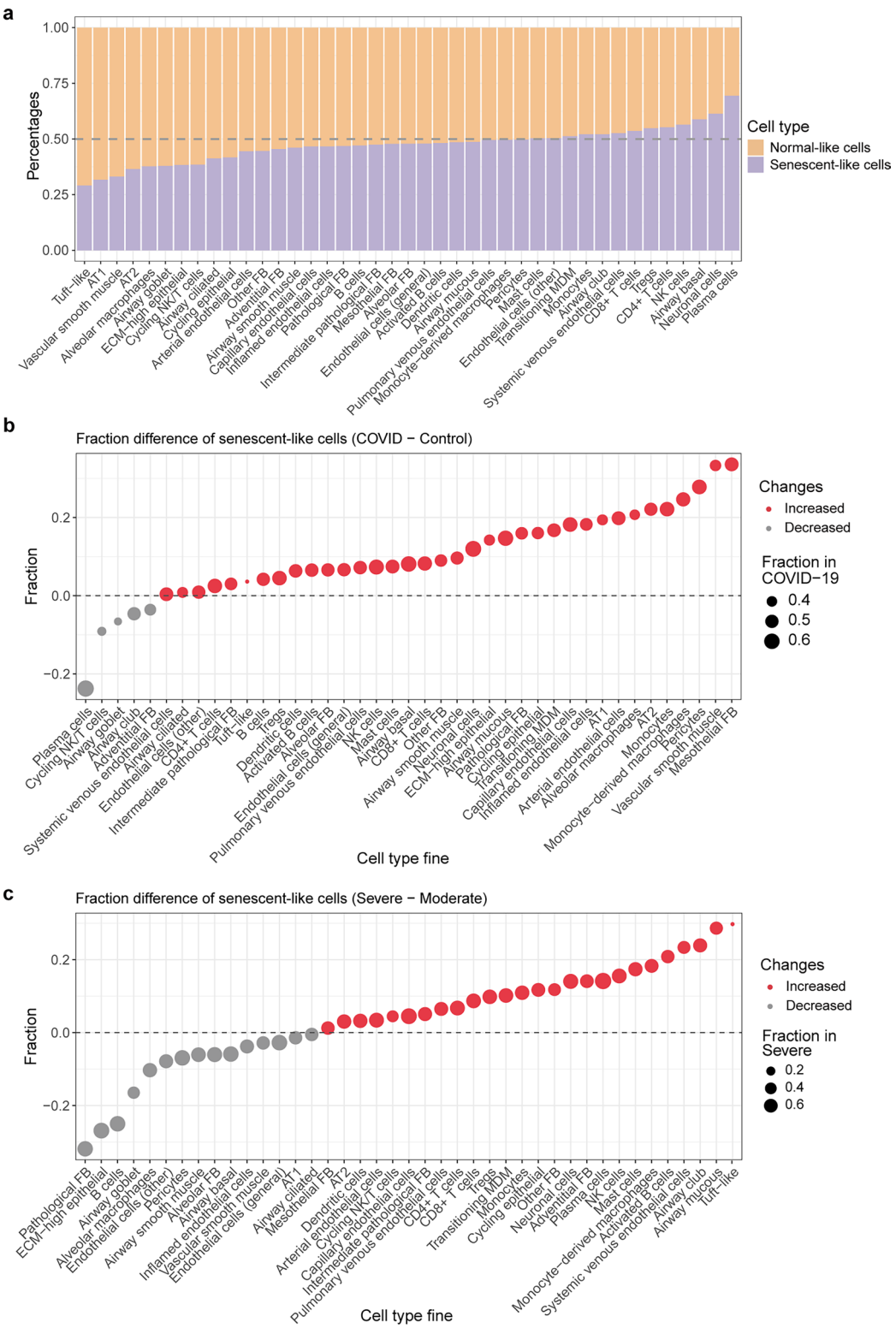
Extended Data Fig. 3 | The distribution of hUSI across diverse datasets. a. Scatter plot showing the significance of Spearman correlation between hUSI scores and the ages of GTEx samples. Each point represents a different method, with the top five correlated methods labeled. P-values were calculated using cor.test function. **b.** Box plot of hUSI levels in GTEx samples stratified across various tissues by age groups (n = 17382 samples; 5345 donors aged 20 to 49 and 12037 donors aged 50 to 70). Asterisks denote significance levels, (two-tailed *t*-test). **c.** t-SNE distribution of bulk samples from donors aged from 1 to 96 years, which were clustered into two age groups (left panel). The middle panel shows hUSI values for each sample in UAMP, while the right panel shows the comparison of hUSI levels between the two age groups (n = 143 samples; 83 donors aged 1 to 50 and 60 donors aged 51 to 96; two-tailed Wilcoxon test). **d.** Scatter plot showing the Spearman correlation between hUSI scores and age in the human skin dataset. **e.** Cell distribution in the Teo2019 dataset. The left panel shows the t-SNE distribution of different cell groups, and the right panel displays hUSI values for each cell. **f.** Cell distribution in the human pancreas dataset. **g-k.** Five single-cell RNA-seq datasets included for benchmarking (the number of cells for each dataset is provided in Supplementary Data 3). The left panel shows the t-SNE distributions of different cell groups, the middle panel displays hUSI values for each cell, and the right panel presents the comparison of hUSI levels across different groups (two-tailed Wilcoxon test). **l.** The t-SNE distribution of cells at different immortalization stages (hTERT_2 and hTERT_7) and passage stages (PDL_25, PDL_29, PDL_33, PDL_37, PDL_46, and PDL_50). The boxplots in the upper panels (**b-c**, **g-k**) represents the interquartile range (IQR), with its lower and upper edges indicating the 25th and 75th percentiles, respectively. The medium value (50th percentile) is shown within the box, and the whiskers extend to the minimum and maximum values within 1.5 times the IQR of the quartiles. For all statistics, * denotes p < 0.05, ** denotes p < 0.01 and *** denotes p < 1e-3; The exact p values are reported in Source Data for Extended Data Fig. 3.

https://doi.org/10.1038/s43587-025-00886-2



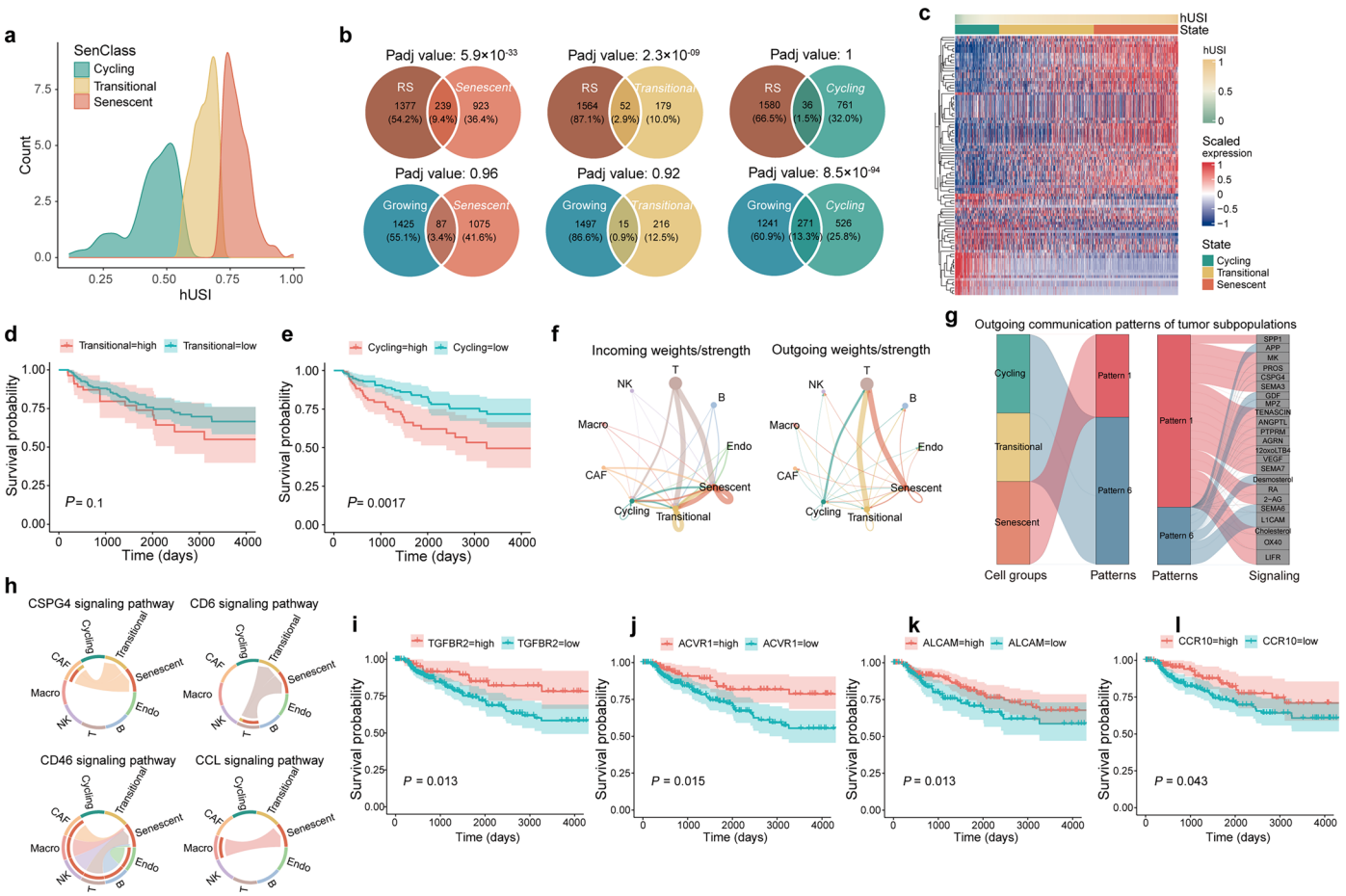
Extended Data Fig. 4 | hUSI revealed potential senescence regulators. a. A high similarity transcript (ENSG00000275464) hinders the design of siRNA for the knockdown of *PWP2*. **b-e**. Senescence validation of *PRIM1* knockdown in ARPE19 cells (n = 3 biological replicates; two-tailed *t*-test, the exact p values were shown in Source Data Extended Data Fig. 4). *PRIM1* was knocked down by two siRNAs (si-*PRIM1*-1 and si-*PRIM1*-2) (**b**). The senescence status was demonstrated by representative images of SA-β-Gal staining (**c**). The gene expression of *CDKN1A*, *CDK1*, *LMNB1* and *MKI67* evaluated by qRT-PCR (n = 3 biological replicates) (**d**) and the Western blot of p21 protein levels (**e**) before and after *PRIM1* knockdown in ARPE19 cells, with GAPDH serving as the internal control. **f-i**, **j-m**, **n-q** and **r-u** present the senescence validation results for *PSMA7*, *INCENP*, *PSMD2*, and *DDX49* (as those in **b-e**), respectively. Data are presented as mean ± SEM. *, **, and *** denote $p < 0.05$, $p < 0.01$ and $p < 1e\text{-}3$, respectively, two-tailed *t*-test. The exact p values of **f-h**, **j-l**, **n-p**, **r-t** were shown in Source Data for Extended Data Fig. 4.

https://doi.org/10.1038/s43587-025-00886-2



Extended Data Fig. 5 | The accumulation of senescent cells in COVID-19. a. Bar plot showing the percentage of senescent subpopulations across various cell types. **b, c**. Dot plots displaying the fractional difference of *Senescent-like* cells (y-axis) between the COVID-19 and Control groups (**b**) and that between Severe and Moderate COVID-19 groups (**c**) across all fine cell types.

https://doi.org/10.1038/s43587-025-00886-2



Extended Data Fig. 6 | hUSI identifies prognosis-related senescent tumor cells in melanoma. a. The distribution of hUSI across different senescent subpopulations. **b**. Venn diagram showing the validation of subpopulation of different senescence degrees (*Cycling*, *Transitional* and *Senescent*) by overlapping their specifically highly expressed genes with genes up-regulated in replicative senescent (RS) or growing samples. P-values were calculated by hypergeometric distribution test. **c**. Heatmap showing the activity values of 89 ICAnet modules across cells, sorted by hUSI value. **d**. Survival curves of melanoma patients with low or high proportion of *Transitional* subpopulation in TCGA-SCKM cohort. **e**. Survival curves of melanoma patients with low or high proportion of *Cycling* subpopulation in TCGA-SCKM cohort. **f**. Incoming (left panel) and outgoing (right panel) network in each tumor subpopulation and other cell types, with point size indicating interaction strength and arrow thickness representing weight size. **g**. River plot showing the outgoing pattern 1 and 6 in three tumor subpopulations (left panel) and the corresponding genes involved in these two patterns were labeled out (right panel). **h**. Chord plots showing the interaction between *Senescent* subpopulation and other cell types in tumor microenvironment by CSPG4, CD6, CD46 and CCL signaling pathways, respectively. **i-l**. Survival curves of melanoma patients with low or high expression of *TGFBR2* (**i**), *ACVR1* (**j**), *ALCAM* (**k**), and *CCR10* (**l**) in TCGA-SCKM cohort.

nature portfolio

Corresponding author(s): Ting Ni, Weixu Wang, Gang Wei

Last updated by author(s): Mar 7, 2025

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
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ 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)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ 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	R (v.4.0.5), R package GEOquery (v2.58.0),R package TCGAbiolinks (v.2.18.0)
Data analysis	R (v.4.0.5), TrimGalore (v.0.6.6), STAR (v.2.2.1), StringTie (v2.2.1),DESeq2 (v1.42.1), Seurat (v.4.4.0), gelnet (v.1.2.1), fgsea (v.1.16.0), Seurat (v.4.2.1), ICAnet (v.0.0.9), phateR package (v.1.0.7), clusterProfiler (v.3.18.1.), CellChat (v.1.0.0.), EpiDISH (v.2.6.1), CIBERSORT (R script v1.04.), survival (v.3.4-0), survminer (v.0.4.9), GSVA (v.1.38.2), pROC (v.1.18.0), randomForest (v.4.7.1.1), glmnet (v.4.1.8), e1071 (1.7.14), mclust (v.5.4.10), python (v3.12.2), custom codes (https://github.com/WJPina/HUSI), OCLR was implemented by gelnet::gelnet()

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

All RNA-seq data generated by this study can be downloaded under GEO accession numbers (GSE282274). The published datasets included in this study were

recorded in the Supplementary Table (Supplementary Table 1 for hUSI development and Supplementary Table 3 for hUSI validation). The processed RNA-seq profile used to develop hUSI are deposited in figshare database (https://figshare.com/s/0335252fca31a5303510). The GRC38 reference genome used to aligned the raw sequence files was download from National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.26/). The gene annotation file was download from GENECODE (https://www.gencodegenes.org/human/release_31.html). The processed melanoma profiles were downloaded from GEO under accession number

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	Not applicable.
Reporting on race, ethnicity, or other socially relevant groupings	Not applicable.
Population characteristics	Not applicable.
Recruitment	Not applicable.
Ethics oversight	Not applicable.

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

Life sciences study design

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

Sample size	For training set,the sample size for the training set was determined through comprehensive data collection and quality control:data collection (public databases and publications were systematically searched using keywords such as "senescence," "RNA-seq," and "human" to identify relevant datasets.);quality control (only datasets containing both senescent and control samples that met predefined standards were retained for analysis.) For publicly available datasets, the sample size was defined by the source publications. Only datasets with at least three replicates per condition were included in the analysis. In this study, the sample size for experiments was determined based on repeatability, with a minimum of three replicates per condition to ensure statistical robustness.
Data exclusions	Data exclusions standards for model training were included in the main text and no data were excluded from the experiments. For downstream analyses and, low-quality cells were excluded from the scRNA-seq and Perturb-seq analyses based on pre-established quality control criteria to ensure the reliability of the data.
Replication	Reported results were consistently replicated across multiple experiments with all replicates generating similar results. Experiments were repeated at least 3 times.
Randomization	Randomization was applied throughout the cell biology experiments. At the seeding stage, cells from the same passage were randomly allocated to experimental conditions, with each condition biologically replicated at least three times. Representative results shown in the figures were randomly selected from all technical replicates.
Blinding	Blinding is not typically used as the all the results are quantitative and do not require subjective judgment or interpretation.

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
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

p21 (#CTS2947, Cell Signaling Technology, Lot:14), RAS (#CST3965, Cell Signaling Technology, lot:4), HRP-conjugated Beta Actin monoclonal antibody (#HRP-60008, Proteintech, Lot:21014779), p21 (#sc-6242, Santa Cruz, Lot:A3124), HRP-conjugated GAPDH monoclonal antibody (#HRP-60004, Proteintech, Lot:21010938).

Validation

p21 (https://www.cellsignal.cn/products/primary-antibodies/p21-waf1-cip1-12d1-rabbit-mab/2947), RAS (https://www.cellsignal.cn/products/primary-antibodies/ras-antibody/3965), HRP-conjugated Beta Actin monoclonal antibody (https://ptglab.com/products/Beta-Actin-Antibody-HRP-60008.htm), p21 (https://www.scbt.com/p/p21-antibody-f-5), HRP-conjugated GAPDH monoclonal antibody (https://www.ptglab.com/products/GAPDH-Antibody-HRP-60004.htm).

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)

The HEK 293T (human renal epithelial cell line, GNHu44) and ARPE-19 (a human retinal pigment epithelial cell line with differentiated properties, SCSP-5277) cells were originally obtained from the Cell Bank of the Chinese Academy of Sciences. Primary HFF (human foreskin fibroblasts, HFF-1) cells were originally obtained from ATCC (SCRC-1041).

Authentication

All cells were authenticated using STR profiling by the supplier, the Cell Bank of the Chinese Academy of Sciences or ATCC.

Mycoplasma contamination

All cells were assayed for mycoplasma contamination by the supplier, the Cell Bank of the Chinese Academy of Sciences or ATCC. Mycoplasma contamination was not detected.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell lines were used in this study.

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, mosiacism, off-target gene editing) were examined.