



# Inference of age-associated transcription factor regulatory activity changes in single cells

Alok K. Maity<sup>1</sup>, Xue Hu<sup>1</sup>, Tianyu Zhu<sup>1</sup> and Andrew E. Teschendorff<sup>1,2</sup>✉

**Transcription factors (TFs) control cell identity and function. How their activity is altered during healthy aging is critical for an improved understanding of aging and disease risk, yet relatively little is known about such changes at cell-type resolution.** Here we present and validate a TF activity estimation method for single cells from the hematopoietic system that is based on TF regulons, and apply it to a mouse single-cell RNA-sequencing atlas, to infer age-associated differentiation activity changes in the immune cells of different organs. This revealed an age-associated signature of macrophage dedifferentiation, which is shared across tissue types, and aggravated in tumor-associated macrophages. By extending the analysis to all major cell types, we reveal cell-type and tissue-type-independent age-associated alterations to regulatory factors controlling antigen processing, inflammation, collagen processing and circadian rhythm, that are implicated in age-related diseases. Finally, our study highlights the limitations of using TF expression to infer age-associated changes, underscoring the need to use regulatory activity inference methods.

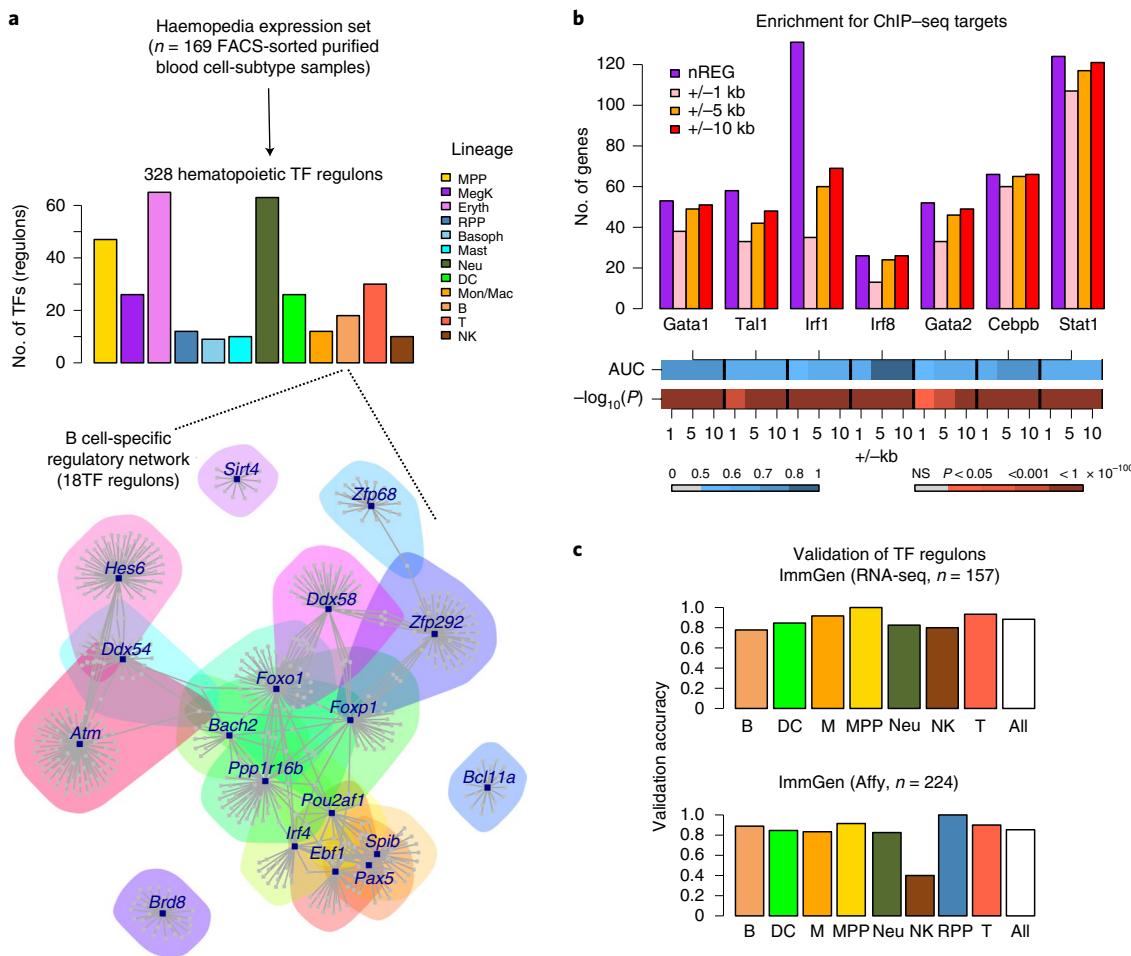
**A**ging phenomena such as immunosenescence, inflammation, fibrosis and stem cell exhaustion are intrinsic processes of normal aging that contribute to disease risk, morbidity and mortality<sup>1–5</sup>. Given that TFs play a key role in determining cell identity and function<sup>6,7</sup>, it is particularly important to elucidate how the regulatory interactions they form are disrupted in aging<sup>8</sup>. Because bulk tissue is composed of many different cell types, efforts to map age-associated molecular changes should ideally be carried out at the highest possible cellular resolution, that is, the single cell. For instance, a recent meta-analysis of transcriptomic aging signatures across bulk tissues revealed little overlap between organs<sup>9</sup>. To address this, the Tabula Muris Senis (TMS) consortium recently generated a single-cell RNA-sequencing (scRNA-seq) atlas comprising 350,000 mouse cells from 23 different organs across 6 age time points<sup>10</sup>, an invaluable resource for investigating the effects of age on regulatory networks and cellular function across different cell types and tissues.

In principle, one way to infer putative changes to regulatory factors is to study the age-associated changes in TF expression. However, this is problematic for two reasons. First, transcriptomic changes associated with healthy aging will be much smaller compared to the differences in expression associated with cell type or disease, making it much harder to identify. Exacerbating this, age-associated changes are characterized by an increased cell-to-cell variability<sup>11</sup> and stochasticity<sup>12</sup>, even within the same cell type. Second, scRNA-seq data are noisy and suffer from many dropouts, which can significantly hamper the inference of small effect sizes. In view of these challenges, we and others have advocated a different strategy that substantially improves the inference of transcription factor activity (TFA) in single cells<sup>13–15</sup>. This strategy relies on the concept of a ‘TF regulon’ defined as a set of carefully selected downstream (that is, direct and indirect) TF targets that provide a faithful measure of upstream regulatory activity<sup>16–18</sup>. Inference of regulatory activity is then derived from the coordinated expression patterns of the TF’s regulon genes. By anchoring the inference on a pool of downstream targets, as opposed to a single gene

(that is, the TF itself), one can overcome the challenge posed by a high dropout rate of scRNA-seq data, leading to more reliable estimates of a TF’s regulatory activity. For instance, we have recently demonstrated how TF regulons can detect inactivation of known tumor suppressor genes (for example, *KLF5* and *ATOH1* in colon cancer), which would not have been detected had we just relied on TF expression levels<sup>13</sup>.

Given the higher sensitivity to detect regulatory activity changes in single cells with TF regulons, we here decided to apply two such regulon-based approaches to the TMS scRNA-seq dataset, with the aim of exploring age-associated regulatory activity changes across different cell types and tissues. To build the TF regulons, we follow two different strategies, depending on the aim and cell types being considered. To study age-associated differentiation defects in immune cells, we adapt our SCIRA algorithm<sup>13</sup> to build regulons for hundreds of hematopoietic TFs, which we accomplish by leveraging high-quality expression datasets that are purified by fluorescence-activated cell sorting (FACS) and profile all major immune cell types. This derivation of TF regulons from FACS-sorted data constitutes a new adaptation of the SCIRA algorithm, as the original implementation is tailored to multi-tissue bulk data. By construction, the SCIRA regulons measure regulatory activity of the TF’s function in differentiation, hence we refer to this as differentiation activity. To study age-associated regulatory activity changes of general TFs in other cell types (for example, endothelial and epithelial cells), we make use of the TF regulons from the DOROTHEA database<sup>19</sup>, a highly curated mouse and human regulon database that has been shown to lead to improved estimates of regulatory activity in scRNA-seq data<sup>14</sup>. The DOROTHEA regulons provide a more general measure of a TF’s regulatory activity, potentially reflecting a TF’s functional pleiotropy, and are not cell- or tissue-type specific. By applying the SCIRA and DOROTHEA regulons to the TMS data, we identify many cell-type and tissue-type-independent age-associated signatures of regulatory activity change, targeting key age-related pathways.

<sup>1</sup>CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China. <sup>2</sup>UCL Cancer Institute, Paul O’Gorman Building, University College London, London, UK. ✉e-mail: andrew@picb.ac.cn

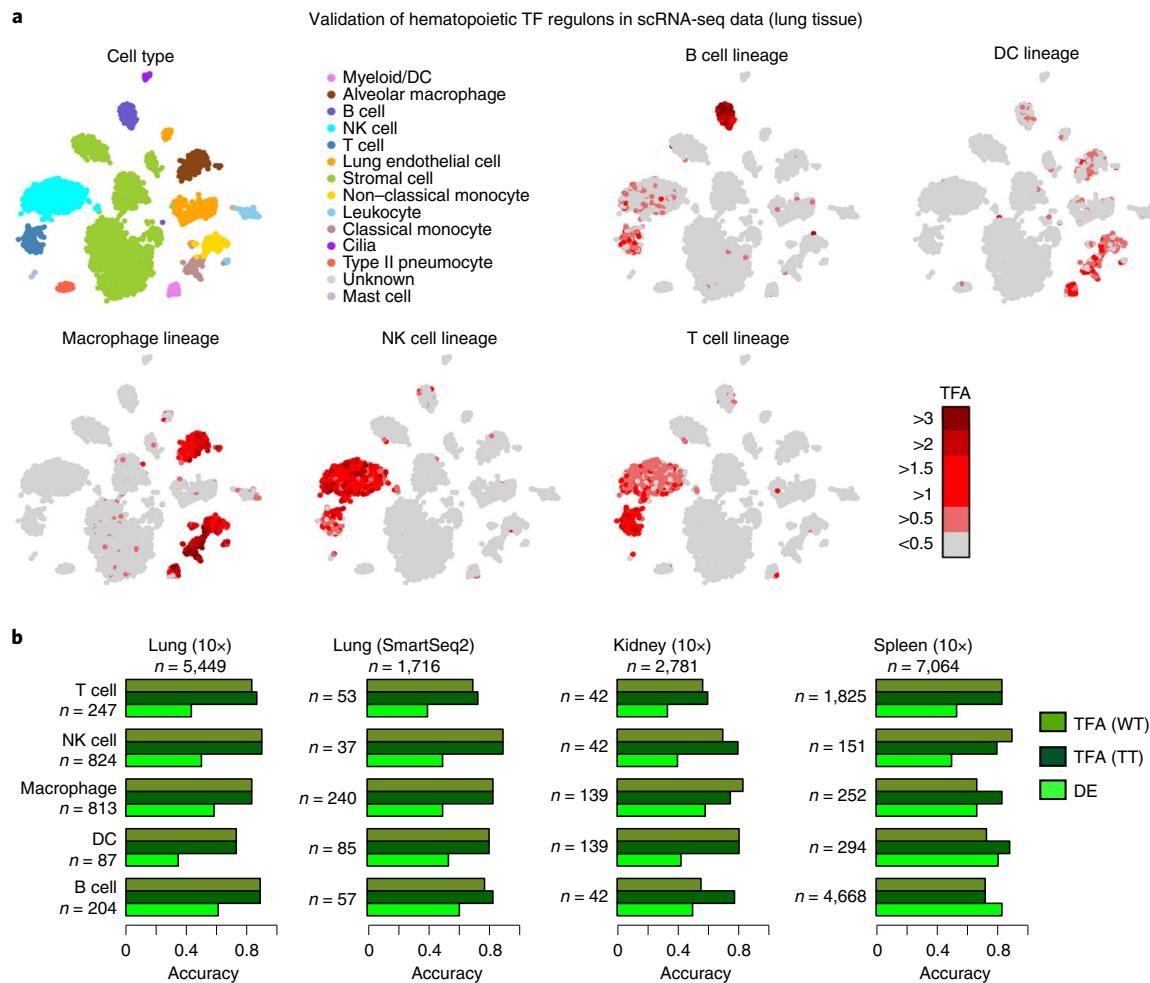


**Fig. 1 | Derivation and validation of hematopoietic transcription factor regulons.** **a**, From the Haemopedia mouse mRNA expression dataset profiling 169 cell-sorted samples, we derived 328 lineage-specific TFs and their associated TF regulons, that is, sets of direct and indirect targets of the TFs that measure upstream regulatory activity in relation to the TF's functions in differentiation. Bar plot displays the number of lineage-specific TFs for 12 different lineages, including multipotent progenitor (MPP), megakaryocyte (MegK), erythrocyte (Eryth), restricted potential progenitor (RPP), basophil (Basoph), mast (Mast), neutrophil (Neu), DCs, monocyte/macrophage (Mon/Mac), B cell, T cell and NK cell. The B cell-specific regulatory network displaying 18 B cell-specific TFs and their regulon genes is displayed. The derivations of this regulatory network and those of the other 11 lineages are described in Methods. As expected, regulons from different TFs may exhibit overlap. **b**, Enrichment of ChIP-seq targets among specific TF regulons for which corresponding ChIP-seq data were available from the ChIP-seq atlas. Bar plot displays the number of genes in the TF regulon (nREG) and the subsets of these that map to binding sites of the TF using  $\pm 1$ -kb, 5-kb and 10-kb windows from the ChIP-seq binding peak to determine overlap with a gene target. Bar plots at the bottom display the area under the curve (AUC) statistic comparing the binding-intensity signals of TF regulon genes versus non-regulon genes, as derived from a one-tailed Wilcoxon rank-sum test, and the corresponding one-tailed  $P$  values. **c**, Validation of the 328 hematopoietic TFs in independent cell-sorted mouse RNA-seq and Affymetrix gene expression datasets from the Immunogenomics project. The number of cell-sorted samples is indicated above the plots. Bar plots display the validation accuracy for each blood lineage, defined as the fraction of blood lineage-specific TFs that display a significantly higher TFA in the corresponding cell-sorted samples of that lineage compared to all other lineages. These validation accuracies are only displayed for those lineages with reasonable numbers of cell-sorted samples, otherwise there is little power to detect differences. The last bar is the average validation accuracy over all assessed lineages. Statistical significance for each individual TF regulon was determined by a one-sided  $t$ -test comparing TFA of the cell-sorted samples against that of all other samples, and using nominal  $P$  values  $< 0.05$  to declare significance. NS, not significant.

## Results

**Construction and validation of hematopoietic transcription factor regulons.** To study age-associated changes affecting the differentiation state of immune cells in mice, we first set out to derive high-quality regulons for blood cell-type-specific TFs, using murine mRNA expression data derived from 169 FACS-sorted purified blood cell types, representing 13 different lineages within the hematopoietic system, as generated by the Haemosphere/Haemopedia (Fig. 1a and Methods)<sup>20</sup>. Starting out from a set of 1,675 TFs in mice (<http://genome.gsc.riken.jp/TFdb/>)<sup>21</sup>, we applied SCIRA to the Haemopedia dataset, to infer blood cell-lineage-specific TFs and their regulons (Methods). This identified 369 TFs specific

to at least one of 13 main blood cell-type lineages and for each of which we could identify at least 10 regulon target genes. To verify the consistency of these regulons, we estimated the TFs differentiation activity in the same Haemopedia dataset, which resulted in 328 (that is, 89%) consistent TF regulons (Fig. 1a, Methods and Supplementary Data 1). Regulon sizes ranged from 10 to 198, with a mean of 54 gene targets, and with a grand total of 17,729 TF-target gene interactions, of which the majority (72%) represented positive regulations. Regulons exhibited strong enrichment for binding targets of the corresponding TFs (Fig. 1b), as assessed using chromatin immunoprecipitation followed by sequencing (ChIP-seq) data from the ChIP-seq Atlas<sup>22</sup> (Methods).

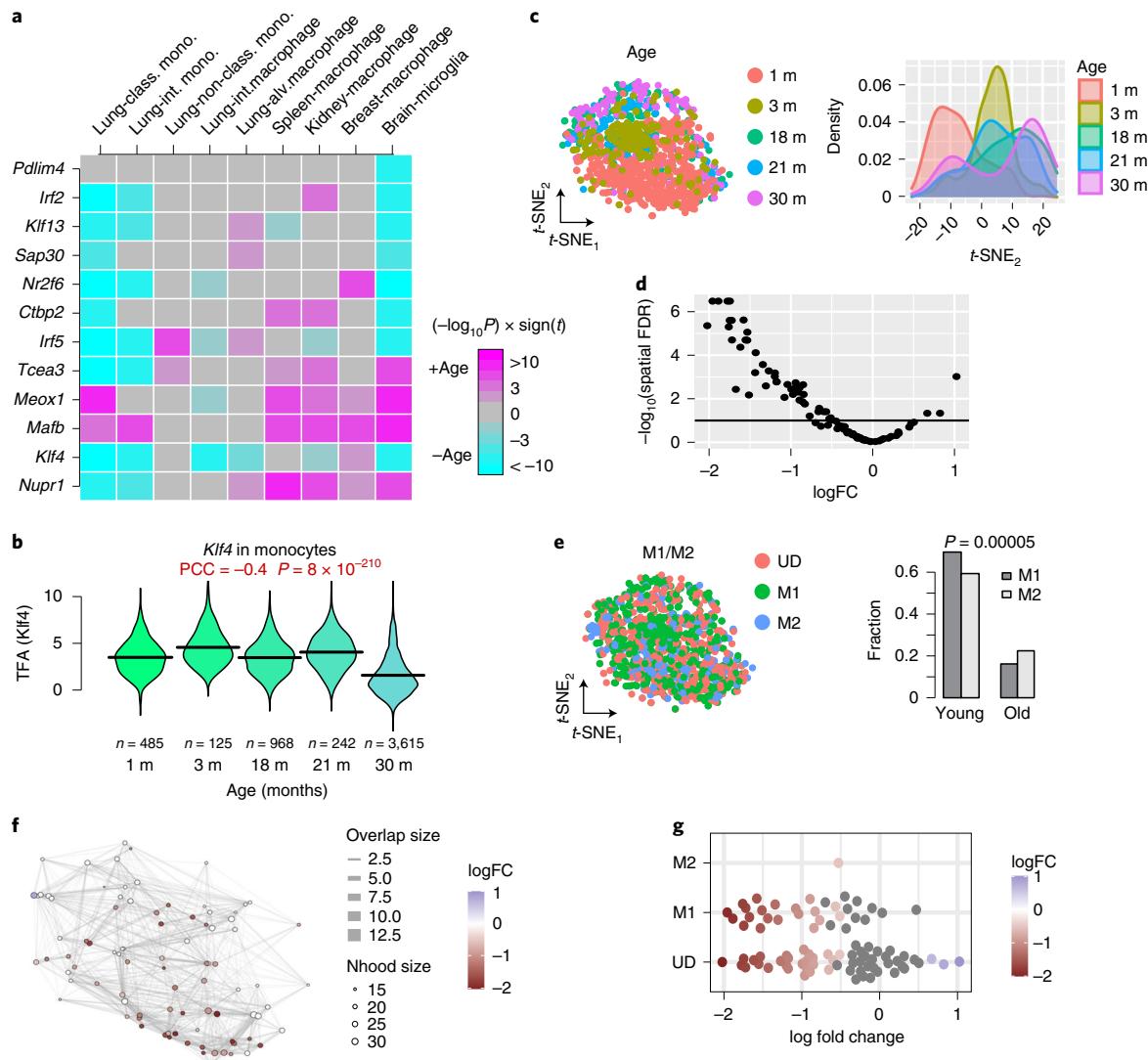


**Fig. 2 | Validation of hematopoietic transcription factor regulons in scRNA-seq data.** **a**, Left t-SNE diagram shows the TMS lung tissue 10x dataset (mice aged 3 months), displaying clusters annotated to various cell types. Other plots show the same t-SNE diagrams but with cells colored by the estimated TFA as derived with hematopoietic TF regulons. The displayed TFA is the mean TFA averaged over the corresponding lineage-specific TFs (B cell, DC, macrophage, T cell and NK cell). **b**, Bar plots display the validation accuracies of our hematopoietic TFs in scRNA-seq datasets from TMS: lung 10x, lung SmartSeq2, kidney 10x and spleen 10x. These tissues were selected because they contained large numbers of cells annotated to corresponding immune cell subsets. Bar plots compare the validation accuracy obtained using estimated TFA values from corresponding lineage-specific TF regulons (TFA) to the accuracy obtained by DE. For TFA, statistical significance was assessed using either a one-tailed t-test (TT) or a one-tailed Wilcoxon test (WT). For DE, we used a one-tailed Wilcoxon rank-sum test. For the bar plots, the total numbers of immune cells are indicated above and the numbers of cells in the immune cell categories evaluated are shown on the left.

To validate the TF regulons, we estimated differentiation activity of these 328 TFs in two independent FACS-sorted bulk expression datasets from the Immunological Genome Project (ImmGen)<sup>23,24</sup>. As each of the 328 TFs is specific to one blood cell type, we assessed if the estimated differentiation activity of the TF is significantly higher in the corresponding blood cell type compared to the other immune cell subtypes. Validation accuracies were 88% and 85%, respectively (Fig. 1c). Of note, validation accuracies of our cell-type-specific TF regulons were substantially higher than those obtained using generic TF regulons from the DOROTHEA database<sup>4,19</sup> (Extended Data Fig. 1a and Methods), thus justifying our approach to derive immune cell-specific TF regulons from FACS-sorted data.

**Validation of hematopoietic transcription factor regulons in single-cell RNA-seq data.** To validate our TF regulons in scRNA-seq data, we estimated differentiation activity of blood cell-type-specific TFs in the TMS scRNA-seq dataset<sup>25</sup>, where mRNA expression was measured in various tissues from 3-month-old mice. At this age, mice are young adults, which is an appropriate time point in which

to assess the validity of our TF regulons, without confounding by early developmental effects or by effects associated with aging. For the validation, we focused on tissue types like lung that contained relatively high numbers of immune cell subtypes, to allow for an objective evaluation (Fig. 2a). Estimating differentiation activity for the hematopoietic TFs in lung tissue cells revealed substantially higher activity in the immune cells in a cell-lineage-specific manner (Fig. 2a). For instance, B cell-specific TFs exhibited much higher activity in B cells, while those specific for dendritic cells (DCs) were highly active in the lung myeloid cells, a predominantly DC population (Fig. 2a). In general, over 80% of the TF regulons exhibited a significantly higher TFA in the single cells of the corresponding immune cell type compared to other cell types, as evaluated using t-tests (Fig. 2b). We observed the same strong validation accuracy in the lung scRNA-seq dataset profiled with a different technology (SmartSeq2), and in 10x scRNA-seq sets from other tissue types with sufficient numbers of immune cells (Fig. 2b). In addition, we validated differentiation activity of our hematopoietic TFs in a 10x scRNA-seq dataset of early hematopoiesis, encompassing 4,142



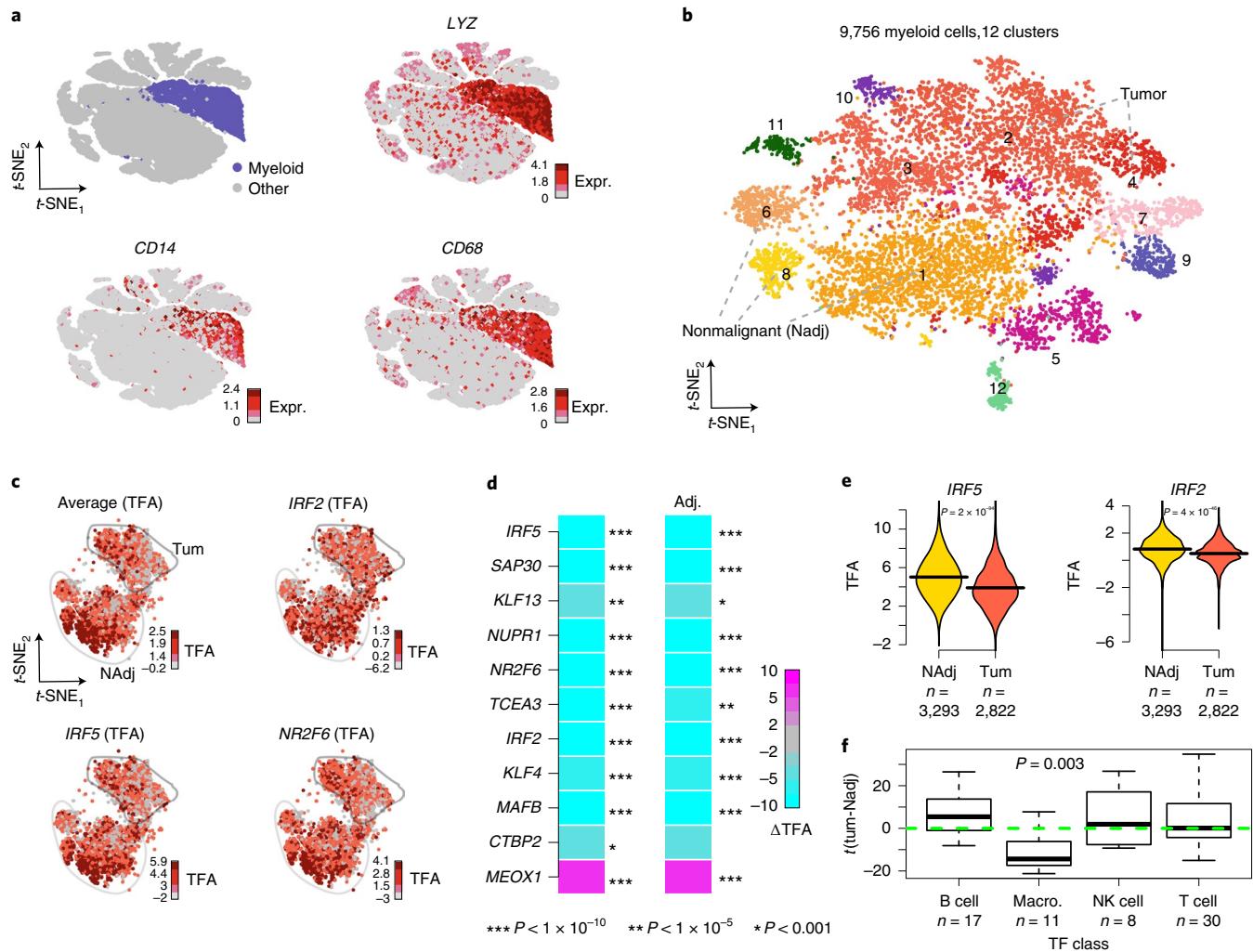
**Fig. 3 | Age-associated differentiation activity changes in myeloid cells.** **a**, Heat map displaying age-associated differentiation activity (TFA) changes for monocyte/macrophage cells from lung, kidney, spleen, breast and brain tissue, as inferred from the TMS dataset. *P* values are two sided. **b**, Violin plots displaying estimated TFA versus age group for *Klf4* in lung monocytes. Pearson correlation coefficient (PCC) and associated two-sided *P* value are shown. **c**, t-SNE map of 1,261 lung alveolar macrophage cells annotated by age (months), with the density plot depicting the age distributions along *t*-SNE<sub>2</sub>. **d**, Volcano plot of the spatial FDRs of cellular neighborhoods, reflecting differential abundance in relation to age. **e**, As in **c**, but with cells annotated by macrophage polarization (M1/M2) or in the case that M1/M2 could not be unambiguously assigned, we annotated the cell as undetermined (UD). The bar plot displays the relative proportions of M1 and M2 cells between young (1 and 3 months) and old (21 and 30 months) mice. *P* value derived from a chi-squared test. **f**, Graph representing the cellular neighborhoods, with edge width reflecting overlap size, node size reflecting neighborhood size and with the nodes colored by the log fold change (FC) of differential abundance in relation to age. **g**, Beeswarm plot of the neighborhoods annotated to their main cell type and their corresponding log fold changes displaying differential abundance (larger negative values indicate neighborhoods of the given cell type where there was a more significant depletion for older cells).

human bone marrow cells<sup>26</sup> (Extended Data Fig. 1b). These data demonstrate that our hematopoietic TF regulons are valid at the single-cell level, and that they can be used to infer differentiation activity of these TFs in immune cells across different organs and species (human and mouse).

Of note, using TF expression levels directly, we observed a significantly lower validation accuracy (Fig. 2b), and therefore a lower level of consistency with purified FACS-sorted bulk data. For instance, comparing differential TFA to differential expression (DE) using Wilcoxon rank-sum tests revealed that, of 20 comparisons, differential TFA displayed higher accuracy than DE in 17 of them, with equal performance in 1 comparison, and only 2 where DE performed better, a highly significant skew (binomial *P*<0.0001). This

indicates that regulons substantially improve the reliability and sensitivity to detect true differentiation activity changes in scRNA-seq data, as compared to TF expression alone (Fig. 2b), in line with previous findings<sup>13</sup>. This justifies the use of SCIRA-derived TF regulons to infer differentiation activity in scRNA-seq data representing immune cells.

**Age-associated changes in lymphoid cells across tissues.** We applied our validated immune cell TF regulons to the TMS data, to infer age-associated changes of TFA within the immune cell compartment of tissues. We selected tissues and cell types that were abundantly profiled across as many of the six age groups (1, 3, 18, 21, 24 and 30 months) as possible. Associations were inferred for



**Fig. 4 | Age-associated macrophage signature in lung tumor macrophages.** **a**, t-SNE-diagram of over 50,000 cells (10x scRNA-seq data from Lambrecht et al.) from the lung tumor microenvironment, with myeloid cells indicated in blue, or with cells labeled by the normalized expression of macrophage markers LYZ, CD14 and CD68. *P* values are two sided. **b**, t-SNE diagram of the 9,756 myeloid cells, displaying 12 clusters, with those clusters strongly enriched for tumor (2 and 4) and nonmalignant normal adjacent (1, 6 and 8) samples highlighted with orange and yellow backgrounds, respectively. **c**, t-SNE diagram as in **b**, but with the macrophages now colored by the estimated TFA levels. Average indicates the average TFA over the 11 macrophage-specific TFs, and diagrams for the TFA of IRF5, IRF2 and NR2F6 are shown. **d**, Color bars displaying the *t*-statistics of differential TFA ( $\Delta$ TFA) between the tumor and normal macrophages for each of the 11 macrophage-specific TFs. The color bar displays the *t*-statistics from a linear regression adjusting for the level of expression of the macrophage markers LYZ, CD68 and CD14 to ensure that results were not driven because of residual differences in macrophage numbers between normal adjacent and tumor samples. **e**, Violin plots displaying the TFA of IRF2 and IRF5 against normal tumor status. Number of normal and tumor macrophages indicated below. Two-sided *P* value shown is the same as that depicted in **d** and was derived from a linear regression *t*-test between TFA and normal cancer status. **f**, Box plots of *t*-statistics of differential TFA between tumor and normal macrophages for TFs that are specific to different immune cell subsets. The number of TFs specific to each immune cell type is indicated below the plots. For the 11 macrophage-specific TFs, the corresponding *P* value was derived from a one-tailed Wilcoxon rank-sum test to determine if the distribution of *t*-statistics was significantly less than 0. Whiskers extend to 1.5 times the interquartile range.

TFs of the five hematopoietic lineages (B cells, T cells, natural killer (NK) cells, monocytes/macrophages and neutrophils) using a linear regression of TFA against chronological age of the cell while adjusting for gender. Among T cells from lung tissue, we observed a particularly strong age-related decrease in differentiation activity for *Lef1* in Cd4<sup>+</sup> T cells ( $P=2\times 10^{-12}$ ; Extended Data Fig. 2a,b) as well as for *Ankrd10* ( $P=1\times 10^{-6}$ ), a target of *Lef1*. This pattern was also observed in the T cells from other tissue types (Extended Data Fig. 2a). We note that *Lef1* is a marker of naïve Cd4<sup>+</sup> T cells, and decreased differentiation activity of *Lef1* may reflect a reduction in naïve Cd4<sup>+</sup> T cells with age, as indicated recently by Elyahu and others<sup>27–32</sup>. Indeed, consistent with Elyahu et al., we also observed

*Lef1* expression to decrease with age in Cd4<sup>+</sup> T cells ( $P=3\times 10^{-9}$ ), yet expression did not have the sensitivity to detect changes for *Ankrd10*. The overall pattern of TFA change in T cells between tissues like lung and spleen was remarkably consistent (Extended Data Fig. 2c). For instance, *Batf* displayed increased differentiation activity in the T cells of both lung and spleen (Extended Data Fig. 2a). Notably, *Batf* is an essential regulator of T cell differentiation whose increased regulatory activity with age has been validated using the assay for transposase-accessible chromatin with sequencing (ATAC-seq) and RNA-seq data in two independent studies<sup>29,30</sup>, in support of an age-related shift of T cells from a naïve undifferentiated state to a more differentiated and exhausted state<sup>33,34</sup>. In B cells,

TFA of *Spib*, encoding an essential regulator for B cell differentiation and a tumor suppressor in pre-B cell acute lymphoblastic leukemia<sup>35</sup>, decreased with age in lung (Extended Data Fig. 2d) and spleen (Extended Data Fig. 2a).

#### Age-associated dedifferentiation of myeloid cells across tissues.

Neutrophil-specific TFs exhibited a clear skew toward inactivation with age with a highly consistent pattern between lung and spleen, the only tissues where neutrophils were abundantly profiled (Extended Data Fig. 3). This is consistent with the well-known decline of neutrophil function and increased disease susceptibility with age<sup>36</sup>. Monocyte- and macrophage-specific TFs also displayed numerous changes (Fig. 3a). For instance, in classical monocytes from lung tissue, we observed a strong decline in differentiation activity for factors such as *Klf4* (Fig. 3b), *Tcea3*, *Irf5*, *Nr2f6* and *Irf2* ( $P < 10^{-50}$ ; Fig. 3a). Of note, *Klf4* is a critical regulator of monocyte differentiation<sup>37,38</sup>, and so the observed decrease in differentiation activity may underlie an age-related dedifferentiation program. To explore potential mechanisms associated with this signature, we analyzed a large genome-wide DNA methylation (DNAm) dataset profiling over 1,000 purified monocyte samples<sup>39</sup>. For all 11 macrophage-specific TFs, we observed highly significant DNAm changes, with nine of these (*IRF5*, *KLF4*, *SAP30*, *NR2F6*, *IRF2*, *KLF13*, *CTBP2*, *MAFB* and *TCEA3*) displaying significant age-associated hypermethylation near the promoter (Supplementary Table 1 and Extended Data Fig. 4), an indication that epigenetic changes could be contributing to dysregulation of these TFs with age.

Besides lung, *Klf4* also displayed reduced differentiation activity in macrophages from kidney and brain (Fig. 3a). Although *Klf4* has been reported to induce both pro-inflammatory and anti-inflammatory macrophage polarization depending on biological context<sup>40–42</sup>, *Klf4*'s age-associated decline in differentiation activity may signal a skew toward an M2 tumor-promoting polarization program<sup>41</sup>. We posited that such an age-related shift from an M1 to an M2 polarization program may be directly detectable from the relative ratio of M1 to M2 cells. Focusing on alveolar macrophages, the only macrophage population in lung tissue with sufficient cell numbers to assess this, we observed that one of the t-stochastic neighborhood embedding (t-SNE) dimensions correlated strongly with age, albeit not strong enough to induce separate clusters (Fig. 3c). We thus applied the miloR algorithm<sup>43</sup> to test if cellular neighborhoods correlated significantly with age, which revealed a relatively strong association (Fig. 3d). We devised a method to annotate alveolar macrophages into three subtypes (M1, M2 and undetermined), revealing a subtle yet significant shift toward a lower M1 (and higher M2) proportion in old mice (chi-squared test,  $P = 0.00005$ ; Fig. 3e and Methods). Using miloR, which takes the variability of different but identically aged mice into account<sup>43</sup>, we observed a considerable number of M1 cellular neighborhoods statistically depleted for cells from old mice (Fig. 3f,g), supporting the view that the relative M2 fraction increases with age.

**Age-associated macrophage polarization signature in lung tumor-associated macrophages.** Given that macrophage polarization has been linked to cancer development and prognosis, we asked

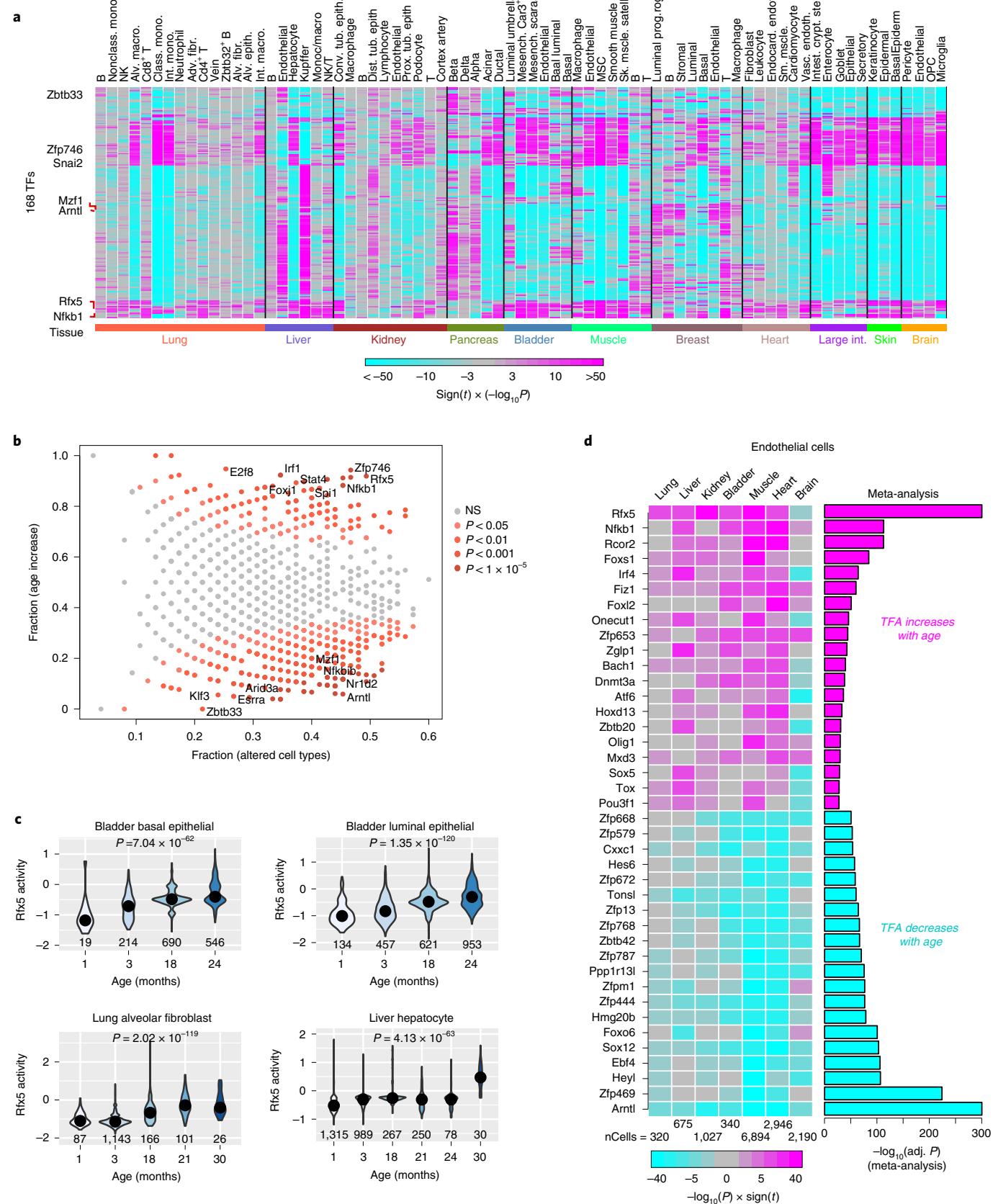
if age-associated TFA changes displayed by macrophage-specific TFs in lung are also seen when comparing lung tumor-associated macrophages (TAMs) to age-matched macrophages from normal adjacent tissue. To this end, we analyzed a 10x scRNA-seq data from Lambrecht et al.<sup>44</sup>, which profiled approximately 10,000 myeloid cells from malignant and normal adjacent tissue from five individuals with lung cancer (Fig. 4a). We verified that the myeloid cell cluster is composed primarily of macrophages (Fig. 4a). Using the annotation from Lambrecht et al., we identified macrophage clusters containing predominantly only TAMs or macrophages from normal adjacent tissue (Fig. 4b). To estimate TFA in this human dataset, we first translated the mouse hematopoietic TFs and regulons into human homologs, subsequently validating them in cell-sorted (FACS) human mRNA expression data from Haemopedia<sup>20</sup>, as well as in a 10x scRNA-seq dataset profiling 68,000 peripheral blood mononuclear cells (PBMCs; Extended Data Fig. 5a, b and Methods). We next asked if macrophage-specific TFs displayed differential TFA between the lung TAMs and those from adjacent normal tissue (Fig. 4c). We observed a strong skew toward lower TFA in TAMs, with 10/11 (that is, 91%) TFs displaying a statistically significant decrease (Fig. 4d,e). Of note, this list included *KLF4*, *IRF5* and *IRF2*. As a negative control, we compared the statistics of differential TFA of macrophage-specific TFs against those of TFs specific to other immune cell types, computing all of these statistics in the same macrophage population. Only for macrophage TFs did we observe a decrease in TFA in the TAMs (Fig. 4f). Thus, these data point toward the age-associated macrophage signature being aggravated in TAMs. To study potential clinical relevance, we evaluated this signature in The Cancer Genome Atlas (TCGA) lung adenocarcinoma (LUAD) RNA-seq dataset<sup>45</sup>. Estimated TFAs for the macrophage-specific TFs revealed patterns that were highly consistent with those observed at single-cell resolution (Extended Data Fig. 5c). For instance, nine TFs exhibited lower activity in LUAD samples compared to their matched normal adjacent tissue. These results were robust when adjusting for macrophage marker expression (*LYZ* and *CD14*; Extended Data Fig. 5d), supporting the view that the decreased TFA in LUAD is not driven by a tumor-associated decrease in macrophage frequency, but by an intrinsic effect within the TAM population, or potentially also within the cancer cells themselves. When correlating the estimated TFA profiles to clinical outcome, we observed a relatively strong association between low TFA of *IRF2* with poor overall survival, even when adjusting for clinical stage and macrophage (*LYZ* and *CD14*) expression (Extended Data Fig. 5e). These observations are noteworthy given that *IRF2* downregulation, which is frequently observed across many cancer types including lung, promotes immune evasion through decreased major histocompatibility complex (MHC) class I antigen presentation and increased programmed death ligand 1 expression<sup>46</sup>.

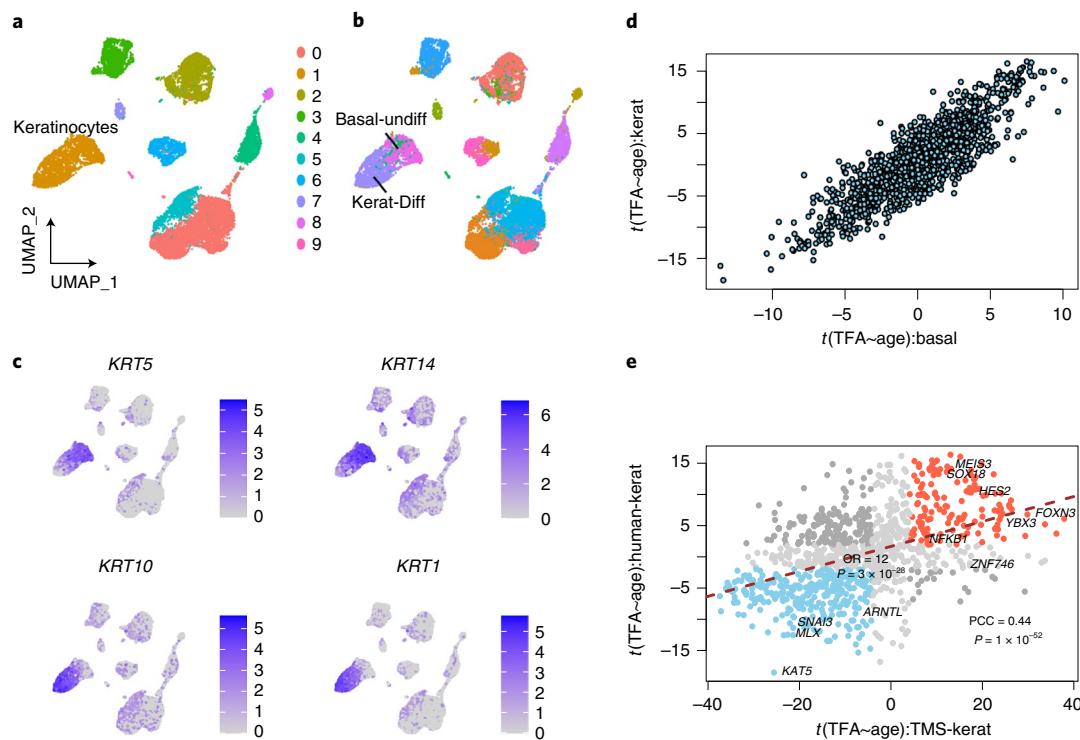
**Cell-type and tissue-type-independent age-associated patterns.** To infer regulatory activity changes for general TFs across a broader spectrum of cell types (for example, fibroblasts and epithelial cells), we used the nonspecific TF regulons from the DOROTHEA database<sup>19</sup>. As shown<sup>14</sup>, these regulons can be used in conjunction with

**Fig. 5 | Consistent age-associated increased or decreased regulatory activity patterns across cell and tissue types.** **a**, Heatmap displays the signed significance levels ( $\text{sign}(t\text{-statistic}) \times (-\log_{10}P \text{ value})$ ) of associations between TFA and age for 168 TFs that display consistent increased or decreased regulatory activity with age across 75 cell types from 11 different tissue types. The TFs have been clustered using hierarchical clustering over the  $t$ -statistics.  $P$  values are two-sided. **b**, Scatterplot of the fraction of cell types in which a TF exhibited a significant association with age (x axis) against the fraction of these associations where the correlation was positive. Each data point represents one or more TFs from a total of 1,152. Data points have been colored according to statistical significance of the skew toward increased or decreased regulatory activity (Binomial test  $P$  value). **c**, Violin plots of *Rfx5* activity against age group for four different cell types.  $P$  value (two-sided) is from a linear regression of TFA against age. **d**, Heatmap of ( $\text{sign}(t\text{-statistic}) \times (-\log_{10}P \text{ value})$ ) values of association between TFA and age for a selected set of TFs that displayed consistent increased or decreased regulatory activity patterns across the endothelial cells from different tissue types, as indicated.  $P$  values are two-sided.

the VIPER algorithm<sup>47</sup> to yield single-cell TFA estimates. We first performed a systematic analysis of the TMS dataset for 1,155 TFs across 75 cell types from 11 tissues (Methods). TF regulons from

DOROTHEA are significantly larger than those from SCIRA (Extended Data Fig. 6a), reflecting the functional pleiotropy of TFs, and in line with this, we observed a weak but significant association





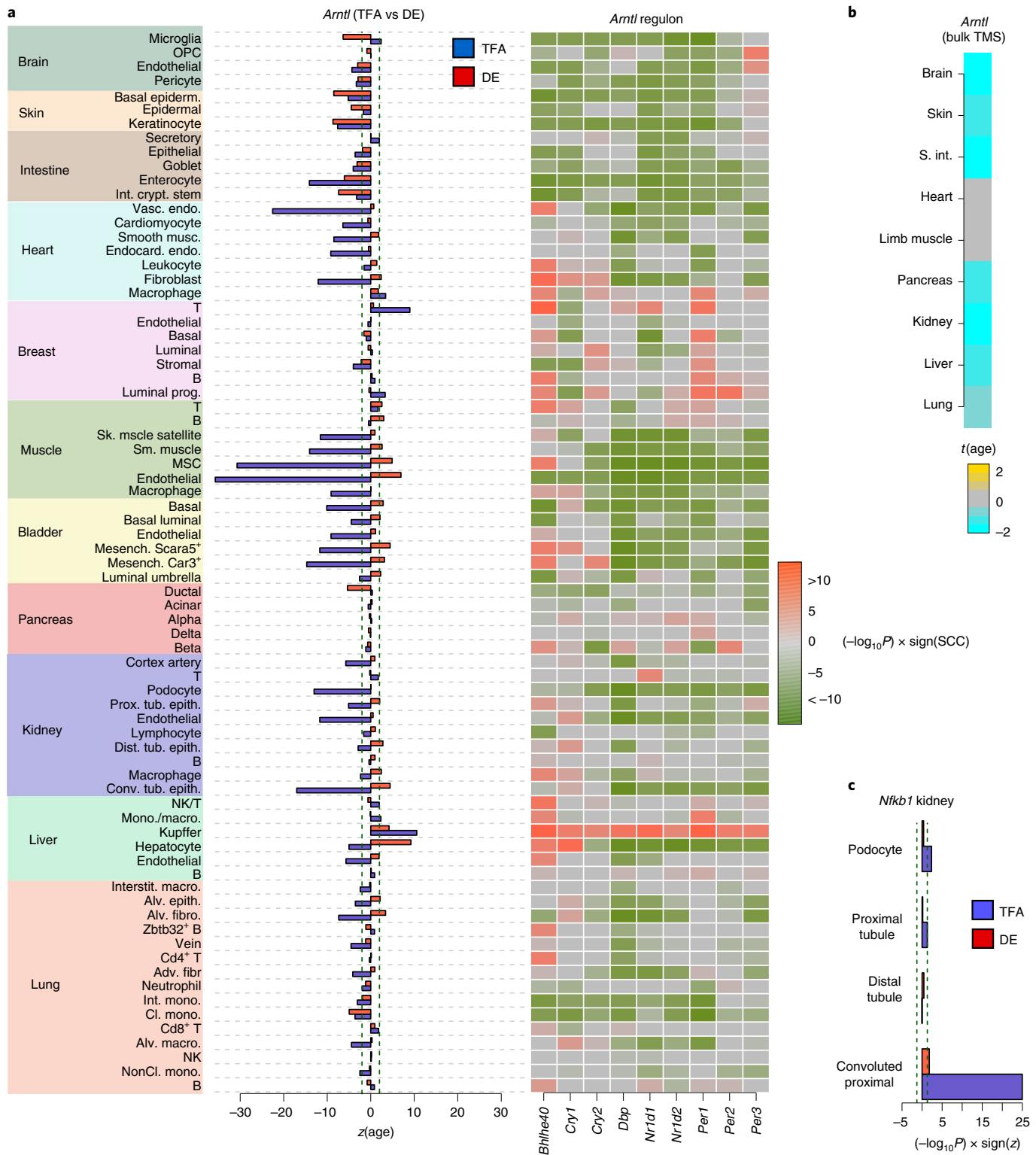
**Fig. 6 | Validation of age-associated DOROTHEA results in an independent skin aging 10x scRNA-seq dataset.** **a**, Uniform manifold approximation and projection (UMAP) plot of a 10x scRNA-seq dataset of skin aging (Solé-Baldo et al.<sup>55</sup>) for a total of over 15,000 cells, with cells annotated by cluster ID. The cluster defining keratinocytes is indicated. **b**, As in **a**, but with cells now annotated by cell type, as provided by Solé-Baldo et al. Color corresponds to cell type, and we mark the undifferentiated and differentiated keratinocyte populations. **c**, As in **a**, but with cells annotated according to the expression levels of two basal (undifferentiated keratinocyte) markers (KRT5 and KRT14), and of two differentiated keratinocyte markers (KRT10 and KRT1). **d**, Scatterplot of *t*-statistics of association between TFA and age for the DOROTHEA TF regulons in the undifferentiated keratinocytes (basal; x axis) versus the corresponding statistics in the differentiated keratinocytes (kerat; y axis). **e**, Validation plot displaying the age-associated *t*-statistics in the differentiated keratinocytes of the TMS data (x axis; TMS-kerat) versus the corresponding *t*-statistics in the differentiated keratinocytes of Solé-Baldo et al. (y axis; human-kerat). Validation strength is indicated in two ways: (i) PCC and associated *P* value, (ii) odds ratio (OR) and associated one-tailed Fisher test *P* value. Each data point represents a DOROTHEA regulon, and the colored ones denote directionally consistent and significant ones in both analyses.

between DOROTHEA regulon size and its chance of being associated with age (Extended Data Fig. 6b), in stark contrast to SCIRA regulons, which did not display such an association (Extended Data Fig. 6c). All 1,155 TFs displayed a significant association with age in at least one of the 75 cell types (Supplementary Tables 2 and 3). We identified a subset of 168 TFs that displayed a significant skew (Binomial  $P < 0.001$ ) for increased or decreased regulatory activity across cell types (Fig. 5a and Supplementary Table 4). For each TF, we plotted the fraction of cell types in which its TFA changes significantly with age against the fraction of these associations where the correlation with age is positive (Fig. 5b). Among TFs displaying frequent increased regulatory activity with age, we observed *Zfp746*, encoding a TF implicated in neurodegeneration<sup>48</sup>, whose activity increased in 33 cell types; *Rfx5*, encoding a ubiquitously expressed TF that activates MHC class II genes, which increased in 34 cell types, and the pro-inflammatory factor encoded by *Nfkbia*, which increased in 30 cell types (Extended Data Fig. 7). Among those displaying reduced activity with age, we observed *Arntl*, encoding a co-activator of the circadian clock; *Mzf1*, encoding a TF implicated in cell-mediated collagen uptake and degradation<sup>49</sup>; and *Nfkbb1*, encoding a repressor of *Nfkbia* (Fig. 5 and Extended Data Fig. 7). Violin plots confirmed the cell-type independence of the age-associated changes, as illustrated for *Rfx5* (Fig. 5c). *Rfx5*'s TFA also increased in fibroblast and smooth muscle cells from heart (Extended Data Fig. 8a). Of note, *Col1a2*, which is involved in collagen production in heart smooth muscle cells and is repressed

by *Rfx5* (ref. <sup>50</sup>), displayed decreased expression with age, while *Rfx5* activity also increased independently of *Col1a2* expression (Extended Data Fig. 8b), thus revealing a link between increased *Rfx5* activity and lower collagen production in aged heart smooth muscle cells, which could underpin the age-related increased risk of rupture of the atherosclerotic plaque<sup>50</sup>.

For endothelial cells, age-associated changes also displayed broad consistency across tissues, although tissue-specific differences were also evident (Fig. 5d). A meta-analysis identified 286 TFs displaying highly significant and consistent trends across the endothelial cells from at least five different tissues (Fig. 5d, Supplementary Table 5 and Methods), which included *Nfkbia*, *Rfx5* and *Arntl*. Focusing on liver endothelial cells, we verified that approximately 46 of the 184 *Nfkbia* regulon genes (that is, 25%) were significantly associated with age, with 40 of these 46 displaying increased expression with age. We observed a strong enrichment for TNF-signaling genes (Supplementary Table 6; false discovery rate (FDR)  $< 1 \times 10^{-8}$ ), consistent with *Nfkbia* increased activity inducing a pro-inflammatory program that may lead to age-related endothelial dysfunction<sup>51</sup>.

**Validation of age-associated regulatory activity patterns.** Several of the above findings have experimental support: for instance, increased age-associated regulatory activity of *Nfkbia* in kidney convoluted tubular and podocyte cells (Fig. 5a and Extended Data Fig. 7) has been experimentally verified in kidney using ChIP-seq data<sup>52,53</sup>. To further validate the patterns for *Nfkbia*, we obtained



**Fig. 7 | Improved sensitivity of transcription factor activity over transcription factor expression to detect age-related variation.** **a**, Bar plot contrasts the age-associated  $z$ -statistics of differential TFA and expression (DE) for *Arntl*, as derived from the scRNA-seq TMS dataset, revealing an overall trend for TF inactivation with age, not seen for DE. Vertical green dashed lines represent the Bonferroni-adjusted significance level. Heat map depicts the age-associated significance statistics for each of the *Arntl* regulon genes (1 of the 10 regulon genes (*Avb*) is not depicted due to non-variable expression in many cell types) in each cell type.  $P$  values are two-sided. **b**, Color bar displaying the  $t$ -statistics of association of *Arntl* expression with age in the bulk RNA-seq TMS dataset, also revealing a trend toward lower activity with age. **c**, Bar plot contrasts the significance levels of the age-associated statistics of differential TFA and expression (DE) for *Nfkbia*, as derived from the scRNA-seq TMS kidney dataset, and for four kidney epithelial cell subtypes. Green vertical dashed lines represent the nominal 0.05 significance level.

*NFKB1* ChIP-seq profiles from four CD4<sup>+</sup> T cell samples from the ChIP-seq atlas<sup>22</sup>, and asked if these targets displayed increased expression in the T cells from tissues where our algorithm had predicted a strong increase in regulatory activity (Extended Data Fig. 7). For all such tissues, *NFKB1* ChIP-seq targets displayed a significant skew toward increased expression with age, even after removing the few targets that overlapped with the *NfkB1* DOROTHEA regulon (Extended Data Fig. 9). Consistent with an age-related decrease of *Mzf1* activity (Extended Data Fig. 7), *Mzf1* has been independently observed to be downregulated with age in fibroblasts from lung and heart<sup>49</sup>, a process thought to underpin age-related fibrosis. Another example is *Rfx5*'s increased regulatory activity in Cd4<sup>+</sup> T cells (Fig. 5a and Extended Data Fig. 7), which has been experimentally validated with ATAC-seq data by demonstrating that *Rfx5*'s binding sites become more open with increased age<sup>30</sup>. Given *Rfx5*'s main role in activating MHC class II genes, we also asked if the age-associated increase in its regulatory activity reflects increased activity of the antigen processing machinery (APM). To test this, we collated a set of 16 APM-related genes from the literature (Methods)<sup>54</sup>, which did not include *Rfx5* or any of *Rfx5*'s regulon genes, and asked if the expression of these genes increased with age in those cell types where we observed increases in *Rfx5*'s TFA. Expression patterns of APM genes were generally also positively correlated with age (Extended Data Fig. 10a), with 17 cell types exhibiting significant consistent association between *Rfx5* activity and APM gene expression, and with only four cell types exhibiting an inconsistent pattern (Spearman correlation coefficient (SCC)=0.36,  $P=0.004$ ; Extended Data Fig. 10b). To further demonstrate that our results generalize to independent datasets, we focused on skin, an organ that has recently been profiled at scRNA-seq level in both young and old humans<sup>55</sup>. Because our TMS skin analysis was restricted to keratinocytes (both basal and differentiated; Extended Data Fig. 7), we calculated the *t*-statistics of association between TFA and age for all DOROTHEA TF regulons in both basal and differentiated keratinocyte populations (Fig. 6a–c), revealing a very strong correlation (Fig. 6d), which allowed us to merge all keratinocytes together. Comparing the *t*-statistics between the skin TMS and the human dataset demonstrated a very strong validation (odds ratio (OR)=12, Fisher test  $P=3 \times 10^{-28}$ ; Fig. 6e). For instance, as predicted in the TMS dataset, *NFKB1* and *ARNTL* displayed increased and decreased regulatory activity with age, respectively (Fig. 6e). In addition, this analysis revealed an age-associated decrease for *Kat5*/*KAT5* in both mouse and human keratinocytes (Fig. 6e). Of note, *KAT5*, also known as *TIP60*, is a major histone acetyl-transferase that has been implicated in aging<sup>56</sup>.

**Regulatory activity displays strong consistency with bulk RNA-seq.** Finally, we observed a number of important examples that further support the view that TF regulons display improved sensitivity over TF expression. For instance, the age-associated decreases in *Arntl* regulatory activity across different cell and tissue types are consistent with the observed underexpression of *Arntl* with age in the TMS bulk RNA-seq data from many different organs (Fig. 7a,b)<sup>57</sup>. In lung, at least seven cell types exhibited a significant decrease in *Arntl* TFA with age, in contrast to *Arntl* expression for which only classical monocytes exhibited a decrease. In liver, the TFA of *Arntl* decreased with age in hepatocytes, the dominant cell population, while *Arntl* expression increased. The latter is inconsistent with hepatocytes making up the bulk of liver tissue and the observed decrease of *Arntl* expression in the bulk RNA-seq TMS liver set (Fig. 7b). Studying the age-associated gene expression patterns of *Arntl*'s regulon genes confirmed coordinated patterns of downregulation in many cell types (Fig. 7a). Another example is *NfkB1* in kidney tissue: using ChIP-seq data it has been shown that *NfkB1* regulatory activity increases with age in kidney epithelial cells, including tubular and podocyte cells<sup>52,53</sup>. Comparing TFA to

TF expression revealed that for two of the four kidney epithelial cell types (convoluted proximal tubules and podocytes), TFA predicted a significant increase with age, while TF expression failed to detect a significant increase in all four cell types (Fig. 7c).

## Discussion

While the TMS study reported age-associated patterns of gene expression across cell types and tissues<sup>10</sup>, here we focused on TF regulatory activity patterns. The data presented supports the view that regulatory activity of TFs in single cells provides highly non-redundant information, displaying substantially improved sensitivity over TF expression to detect true patterns of regulatory activity. This finding is consistent with several previous studies<sup>13,14</sup>, and is particularly pertinent for studying age-associated changes, as these are characterized by relatively small effect sizes, which are more challenging to discern. Indeed, many of the insights gained in this study (for example, the cell-type-independent age-related decrease and increase of *Arntl* and *Rfx5*, respectively) would not have been obtained had we just relied on TF expression levels. This highlights the critical importance of using computational methods based on the framework of TF regulons, where inference of regulatory activity is made based on the coordinated expression pattern of downstream target genes, thus rendering it more robust to the high dropout rate of scRNA-seq data.

Among the biological insights gained, it is worth highlighting the following. First, we have identified several macrophage/monocyte-specific TFs, including notably *Klf4* and *Irf5*, that displayed decreased differentiation activity with age in various solid tissues, including lung, kidney and brain. Although the precise role of these two TFs in macrophage polarization is complex, most of the existing literature points toward *Klf4* promoting an M1 polarization program<sup>40,41,58</sup>. *Klf4*'s decreased differentiation activity could thus signal an age-associated skew toward M2 polarization, a prediction that we were able to confirm in the lung alveolar macrophage population by direct annotation of cells into M1/M2 subtypes. This is relevant for age-related diseases like cancer<sup>59–61</sup>, as shown here in the context of lung cancer where TAMs displayed lower TFA of these TFs compared to age-matched normal lung macrophages. Another notable finding is the observed age-associated decrease of *Irf2* activity in microglia: given that in brain *Irf2* favors an anti-inflammatory M2 polarization program<sup>62</sup> that promotes tissue regeneration and recovery from stroke injury<sup>63,64</sup>, the age-related decreased *Irf2* activity in microglia could enhance a pro-inflammatory M1 program that is associated with a larger infarct and worse stroke outcome. The decreased differentiation activity of *Spib* in B cells is also noteworthy: *Spib* deletion (in conjunction with *Pu.1* deletion) has been shown to impair differentiation and to lead to a 100% incidence rate of pre-B acute lymphocytic leukemia (ALL) in mice younger than 7 months<sup>35</sup>. In the context of pediatric pre-B ALL, characterized by the t(12;21) chromosomal translocation that results in overexpression of the ETV6–RUNX1 fusion gene, *SPIB* is downregulated by direct repression of ETV6–RUNX1 (ref. <sup>65</sup>). Thus, it is plausible that decreased TFA of *Spib*/*SPIB* with age could contribute to ALL risk in the smaller proportion of ALL cases that occur in adults. Our analysis also revealed a largely cell-type- and tissue-type-independent increase of *NfkB1* activity with age. While *NfkB1* has been traditionally viewed as a suppressor of inflammation, aging and cancer<sup>66</sup>, some recent studies indicate a more complex role for *NfkB1* with increased regulatory activity promoting cancer development<sup>67</sup>. One of the TFs exhibiting the most significant and consistent changes in regulatory activity was *Rfx5*. Given the role of *Rfx5* in regulating MHC class II genes, the age-associated increase of *Rfx5* activity across so many cell types and tissues may reflect an age-associated increase in antigen processing, a finding supported by the upregulation of many APM genes. The TF exhibiting the most significant and consistent decrease in regulatory activity with age was *Arntl*

(Bmal1), a co-activator of the circadian clock. This is interesting because alterations to Arntl and the circadian rhythm have not only been linked to aging<sup>68</sup>, but also to infertility<sup>69</sup>, to impaired gluconeogenesis and lipogenesis<sup>70</sup>, and with altered sleep patterns and neurodegenerative diseases<sup>71–73</sup>. Thus, our finding that *Arntl* exhibits decreased regulatory activity across so many different cell and tissue types could represent a common mechanism that unifies these different manifestations of aging<sup>74,75</sup>. To explore this further will require more in-depth scRNA-seq studies profiling a larger number of cells within specific tissues.

Despite the TMS being an excellent resource, it is worth noting some limitations that also affect our study. First, we were limited by the availability of sufficient cell numbers per cell type across age groups, thus preventing us from performing a more systematic cross-tissue analysis. Another limitation is the cell-type annotation and unaccounted cell-type heterogeneity; although we used the cell-type annotations provided by the TMS study, it is conceivable that specific cell clusters were not homogeneous, being composed of further subtypes. Thus, studying age-related patterns within a given cell type could be confounded by age-related variations in cell-subtype proportions. However, such variations are also informative, as the case of *Lef1* in Cd4<sup>+</sup> T cells demonstrates, where the reduction in *Lef1* activity reflects a lower naïve Cd4<sup>+</sup> T cell population with increased age<sup>27,28</sup>. Another inevitable limitation is that cells from each age group are drawn from different mice. While sampling from different mice is critically important to increase power and to capture inter-subject variability in aging, such biological variability may inevitably be confounded with technical batch effects. In this regard, it is worth pointing out that adjusting for mouse ID in these analyses may over-adjust and lead to negative findings<sup>10</sup>. A concrete example is *Lef1* in Cd4<sup>+</sup> T cells, where naïve adjustment for mouse ID in the linear model would lead to a predicted increase in *Lef1* activity, which would be inconsistent with the well-known reduction in naïve Cd4<sup>+</sup> T cells with age<sup>27–32</sup>. Although residual technical batch effects can't be excluded, these are unlikely to represent a major confounder, because we have been able to validate many of our key findings using either independent scRNA-seq datasets or independent ChIP-seq data (for example, NfkB1 in T cells). Moreover, we have focused on reporting age associations that are consistent across different tissue types, which, given the TMS study design, ensures findings are relatively robust to the choice of mice used in each tissue type. A limitation specific to this study is that, like TF expression itself, TFA estimates need to be interpreted with caution. This is because TFs may regulate different subsets of a regulon set depending on which TF function is being exerted. This is particularly pertinent for DOROTHEA-derived regulons, because these are on average much larger and not tissue or function specific. Finally, although factors such as gender have been shown to play an important role in modulating the aging process<sup>76,77</sup>, the study of sex-specific age-associated changes in regulatory activity was beyond the scope of this study.

In summary, by applying state-of-the-art computational methods to infer age-associated regulatory activity changes in the TMS scRNA-seq dataset, we have here performed single-cell resolution analyses into how aging affects regulatory factors. Our analysis has revealed many age-associated TFA patterns that are independent of cell and tissue type, and which shed new insights into diverse age-related phenomena including immune suppression, inflammation, fibrosis and disruption of the circadian rhythm.

## Methods

**FACS-sorted murine bulk expression datasets.** Haemopedia mouse RNA-seq dataset<sup>20</sup>. This mouse expression dataset was used to build the hematopoietic TF regulons. The normalized data were downloaded from <https://www.haemosphere.org/>. Ensembl IDs were converted to gene symbols, resulting in a data matrix defined over 169 purified blood cell types and 18,330 genes. Of the 169 samples, 15 represented cell types from the B cell lineage, 2 from the basophil lineage, 12 from

the DC lineage, 6 from the eosinophil lineage, 17 from the erythrocyte lineage, 11 from the macrophage lineage, 2 from mast cells, 9 from the megakaryocyte lineage, 17 from the MPP lineage, 5 from the neutrophil lineage, 2 from the NK cell lineage, 28 from the RPP lineage and 43 from the T cell lineage.

**Immunological Genome Project datasets<sup>3,24</sup>.** These mouse datasets were used for validation of the hematopoietic TF regulons. One dataset was generated with RNA-seq ( $n=145$ ) and another with Affymetrix ( $n=216$ ) microarrays. In the case of the RNA-seq dataset, the TPM-normalized data were downloaded from the Haemosphere website (<https://www.haemosphere.org/>). Ensembl IDs were transformed to gene symbols. Of the 145 hematopoietic samples, 33 mapped to cell types from the B cell lineage, 8 from the DC lineage, 7 from the macrophage lineage, 1 from the mast cell lineage, 11 from MPPs, 6 from the neutrophil lineage, 20 from the NK cell lineage and 59 from the T cell lineage. In the case of the Affymetrix set, the normalized data were downloaded from the same Haemosphere website. Ensembl IDs were transformed to gene symbols. Of the 216 hematopoietic samples, 26 were from B cells, 36 from the DC lineage, 23 from the macrophage lineage, 7 from the MPPs, 6 from neutrophils, 22 from NK cells, 7 from RPPs and 89 from T cells.

**Construction of hematopoietic transcription factor regulons.** We adapted the machine-learning algorithm implemented in our SCIRA algorithm<sup>13,18</sup> to infer blood cell-lineage-specific TFs and their regulons. While the previous implementation of SCIRA was in the context of a multi-tissue bulk RNA-seq dataset (GTEx), here we adapted it to derive cell-lineage-specific TFs and their regulons from the Haemopedia mouse RNA-seq data representing 169 FACS-sorted purified samples representing 13 different blood cell-type lineages. To infer regulons, we first identified the TFs using as input the TF list from RIKEN (<http://genome.gsc.riken.jp/TFdb/>)<sup>21</sup>, which consists of 1,675 TFs. We note that of the 1,675 TFs, 1,121 passed quality control in the Haemopedia mouse expression dataset. From the Haemopedia dataset, we filtered all genes based on variability, requiring at least a standard deviation larger than 0.25, which resulted in a data matrix of 14,442 genes, including 926 TFs. Next, we computed Pearson correlations between each TF and the 13,516 target genes, transforming these to Fisher z-scores and computing P values by comparison to a normal null distribution of mean 0 and standard deviation equal to  $1/\sqrt{169} - 3$ . Significant marginal associations were declared using a P-value threshold of  $1 \times 10^{-6}$ . The aim of the above marginal correlation analysis is to identify genes that are correlated univariately with TFs, to later run a multivariate (that is, partial correlations) analysis to identify the most likely regulators of each gene. Given that target genes of different regulators can overlap, it is critical to perform the partial correlation analysis to identify the more likely regulatory interactions. At this stage, TFs were also filtered further to only allow those with at least ten significant target genes. Next, for each target gene with at least two regulators from the marginal analysis, we ran a multivariate partial correlation model, and partial correlations with absolute values larger than 0.2 were declared as significant interactions. The partial correlation threshold of  $\pm 0.2$  is sensible, given the number of samples ( $n=169$ ). Finally, only TFs with at least ten significant gene targets from the partial correlation analysis were kept, resulting in a regulon matrix defining over 606 mouse TFs and 8,841 target genes, that is, target genes with at least one TF as a regulator. Partial correlations were computed using the corpcor\_1.6.9 R package (<https://cran.r-project.org/web/packages/corpcor/>). To identify the blood cell-type-specific TFs, we used the limma empirical Bayes framework<sup>78,79</sup> to perform 13 DE analyses, in each comparing one of the 13 cell-lineage samples to all others. We defined cell-lineage-specific markers by requiring that the FDR < 0.05 and that the median expression level in the given lineage was maximal among all lineages. Among the obtained markers, we then selected the TFs present in our regulon matrix constructed above. This resulted in 369 blood cell-type-specific TFs and associated TF regulons.

**Estimation of transcription factor differentiation activity and validation of hematopoietic transcription factor regulons.** We performed validations in the two ImmGen expression datasets described above, as well as in the scRNA-seq lung TMS dataset, which profiles the largest and most extensive set of immune cells. In all cases, a TF regulon was evaluated by first estimating its TFA in each sample. This was done by performing a linear regression between a sample's or cell's expression profile and the binding profile of the TF as specified by its TF regulon, encoding positive targets by +1, non-targets by 0 and negative targets by -1. The TFA was defined as the t-statistic from this regression. Finally, for a TF that is specific to blood cell type 'X', we compared the estimated TFA values for samples/cells belonging to cell type 'X' against all other cell types. Because TFA values are approximately normally distributed and to ensure greater power, we used t-tests to assess statistical significance. Thus, for each blood cell lineage, we could compute an accuracy score, that is, the fraction of cell-lineage-specific TFs that exhibit higher TFA values (one-tailed t-test P value < 0.05) in samples/cells of the corresponding lineage in the independent datasets (ImmGen and lung TMS).

The hematopoietic TF regulons were also validated against corresponding ChIP-seq binding profiles, as obtained from the ChIP-seq atlas (<https://chip-atlas.org/>)<sup>22</sup>. This validation was done by checking that the binding intensity (as derived from the TF ChIP-seq profile) of the predicted TF regulon genes was higher than

for non-regulon genes using a Wilcoxon rank-sum test. We also report the AUC statistic from this test. This analysis was performed for a number of TFs with available ChIP-seq data (Gata1, Tal1, Irf1, Irf8, Gata2, Cebpb and Stat1).

**Single-cell RNA-seq datasets.** We here analyzed the following scRNA-seq datasets:

*Tabula Muris Senis.* This scRNA-seq mouse dataset<sup>10</sup> encompasses many different tissue types with samples collected at six different ages: 1, 3, 18, 21, 24 and 30 months. Data object files were downloaded from figshare at <https://doi.org/10.6084/m9.figshare.8273102.v2>.

We used the normalized data as provided in the h5ad files.

*Bone marrow early hematopoiesis dataset.* This scRNA-seq 10x Chromium data were derived from Setty et al.<sup>26</sup>, profiling approximately 25,000 CD34<sup>+</sup>-enriched bone marrow single cells. We downloaded the high-quality subset, consisting of 4,142 cells and 16,106 genes from the Code Ocean capsule (<https://codeocean.com/capsule/2535144>) and processed the data and annotated cell types following work by Setty et al.<sup>26</sup>. In total, 1,070 cells were annotated as hematopoietic stem cells (HSCs), 971 as general precursors, 1,209 as monocytes, 260 as DCs and 632 as erythroid cells.

*Lung tissue and tumor microenvironment dataset.* This scRNA-seq 10x Chromium dataset was derived from a study by Lambrechts et al.<sup>44</sup>, which profiled malignant and nonmalignant lung samples from five patients. We downloaded all .Rds files available from ArrayExpress (E-MTAB-6149), which included the processed data and t-SNE coordinates, as well as cluster cell-type assignments. After quality control, a total of 52,698 single cells remained, of which 1,709 were annotated as alveolar cells, 5,603 as B cells, 1,592 as endothelial cells, 1,465 as fibroblasts, 9,756 as myeloid cells, 24,911 as T cells and 7,450 as tumor epithelial cells.

*Skin aging dataset.* This scRNA-seq 10x dataset was derived from a study by Solé-Baldo et al.<sup>55</sup>, which profiled over human 15,000 cells from young and old adults. We downloaded the .Rds file from the Gene Expression Omnibus (GEO; [GSE130973](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE130973)) containing the normalized Seurat object. Cell annotation was provided in the .Rds file, and cell types included differentiated and undifferentiated keratinocytes. To verify the cell annotation, we reran the Seurat pipeline (Seurat 4.0.3) with our own parameters on the count matrix, retrieving similar clusters to those in work by Solé-Baldo et al. Using the same markers as in Solé-Baldo et al., we identified clear clusters defining differentiated and undifferentiated keratinocytes. Finally, we declared keratinocytes as differentiated or undifferentiated based on the overlap between our cluster assignment and the annotation provided by Solé-Baldo et al.

**Comparison of differential transcription factor activity to differential expression analysis.** We compared differential TFA (SCIRA) to ordinary DE analysis in the scRNA-seq TMS lung 10x dataset. For DE, we used a Wilcoxon rank-sum test for binary phenotypes, or non-parametric Spearman rank correlations for ordinal phenotypes (for example, multiple time points or stages), as is customary for scRNA-seq data<sup>80</sup>. This is because a non-parametric test, which is free of distributional assumptions, can be applied in the context of high dropout rates<sup>80</sup>. In the case of TFA, differential analysis can be performed using an ordinary t-test. This is because the TFA values, being t-statistics themselves, are not subject to the high sparsity and dropout rate of scRNA-seq data. This can be seen as another advantage of using TFA over TF expression, because estimation of TFA is permissive of more powerful parametric testing. Thus, when comparing differential TFA to DE, we implement what are in effect the best possible methods for calling differential TFA and DE. Finally, for TFs specific to cell type 'X' we compute the fraction of TFs for which the estimated TFA is significantly higher in the single cells of cell type 'X' when compared to all other cell types (using a one-tailed t-test P value and declaring significance at a nominal  $P < 0.05$  level). This fraction is called the validation accuracy score. Analogously, for TF expression and a cell type 'X', we also compute an accuracy score by comparing expression of the TF in the single cells of cell type 'X' to that of all other cells, using a one-tailed Wilcoxon rank-sum test and declaring significance at a  $P < 0.05$  level.

**Estimation of transcription factor activity with VIPER and DOROTHEA.** The previously described SCIRA strategy to estimate TFA only works for TFs that are blood cell-type specific and which are relevant for processes that functionally define these cell types (for example, differentiation). To obtain TFA estimates for all TFs and for more general cell types, we used the DOROTHEA regulon database (for mouse)<sup>19</sup> in conjunction with VIPER<sup>4,47</sup>. Briefly, VIPER was run using the non-likelihood-based DOROTHEA regulons, and with the following options: dnull = FALSE, pleiotropy = FALSE, nes = TRUE, method = none and minsize = 5.

**Age association analysis in Tabula Muris Senis data.** With TFA estimates derived with either SCIRA or VIPER-DOROTHEA, we found correlations with age by performing linear regressions between TFA values and age, while adjusting for sex. Because TFA values are approximately normally distributed, a linear regression is justified. In cases where we were interested in correlating gene expression with

age, we used Spearman correlations. Age association analyses were only performed for tissues and cell types that had sufficient representation across at least three to four age groups (at least 300 cells in total with balanced representation across time points). When performing the age association analysis for the 328 hematopoietic SCIRA TF regulons, we generally only report results that passed a stringent Bonferroni threshold ( $0.05/328 = 1.5 \times 10^{-4}$ ). For DOROTHEA regulons, because these were generally analyzed in a meta-analysis context, we relaxed the significance thresholds using Benjamin–Hochberg FDR  $< 0.05$ .

**M1/M2 polarization analysis in lung alveolar macrophages.** Lung alveolar macrophages were abundantly profiled ( $n=1,261$ ) to allow testing of a shift in M1/M2 macrophage polarization with age. After processing the data and performing dimensional reduction and visualization with t-SNE, we observed that cells from mouse ID 19 (aged 30 months) did not cluster with the 30-month-old cells from the other three mice (mouse IDs 20, 21 and 22), and were paradoxically closer to the young cells from other mice. To exclude a potential batch effect, we thus removed the cells from mouse ID 19, leaving 1,124 cells: 517 (1 month), 184 (3 months), 193 (18 months), 91 (21 months) and 139 (30 months). To annotate these 1,124 lung alveolar macrophages into M1/M2 subtypes, we first identified five robust murine M1 markers (Cd80, Cd86, Fpr2, Tlr2 and Cd40) and five robust M2 markers (Egr2, Myc, Arg1, Mrc1 and Cd163) from the literature<sup>81</sup>. In an initial annotation, we declared cells as M1 if they coexpressed at least two of the five M1 markers, and similarly for M2. Cells annotated to both M1 and M2 subtypes were reassigned an undetermined category alongside all other cells not annotated to either M1 or M2, resulting in 308 M1 cells, 195 M2 cells and 621 UD cells. We reasoned that UD cells clustering predominantly with either M1 or M2 cells could be reassigned to M1/M2 subtypes. To this end, we developed an iterative algorithm that reassigns the status of UD cells to either M1 or M2, depending on their relative proportions among the neighbors of a given UD cell. In more detail, we used the cell graph as inferred using Seurat, and a multinomial test with a P-value threshold  $< 0.05$  threshold to identify the UD cells whose polarization status could be reassigned to either M1 or M2 status. We also required the absolute difference between the proportion of M1 and M2 neighbors of a given UD cell to be larger than 0.2. This procedure was iterated 20 times, but numbers already converged after 7 iterations, resulting in 464 M1 cells, 214 M2 cells and 446 UD cells.

We then tested for differential M1/M2 proportions between young (1 and 3 months) and old (21 and 30 months) mice using a chi-squared test. In a separate analysis, we used the Milo algorithm and associated miloR 1.0.0 package<sup>82</sup> to test for differential abundance taking the variability of different but identically aged mice into account. Briefly, we used miloR to derive spatial FDRs of cellular neighborhoods, reflecting differential abundance of cellular neighborhoods in relation to age. Cellular neighborhoods were annotated to their main cell type and their corresponding log fold changes showing differential abundance displayed in a Beeswarm plot as implemented in miloR.

**Analysis of DNA methylation dataset of purified monocytes.** This Illumina 450k DNAm dataset derives from the MESA study and consists of 1,202 purified monocyte samples<sup>39</sup>. Raw IDAT files were downloaded from the GEO under accession number [GSE56046](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56046), and processed using the minfi Bioconductor package<sup>82</sup>, without background correction and using Illumina's definition of beta value. Background correction increases technical variance despite removing bias in the U and M channels, and therefore is not advisable in this context. Probes with detection P values above 0.05 were assigned missing values. We only allowed probes to contain at most 1% missing values. Remaining missing values were imputed with k-nearest neighbors ( $k=5$ )<sup>83</sup>. To correct for type II probe bias, we used our BMIQ normalization procedure<sup>84</sup>. Inter-sample variation was assessed using singular value decomposition, where we correlated inferred components of variation to technical factors, including beadchip, race, gender and cohort of origin. Race, gender and cohort of origin were confounders. When deriving statistics of association with age, we correlated (linear regression) DNAm levels to age while adjusting for race, gender and cohort of origin. In this regression analysis, we removed three samples that had unique race, gender and cohort values. The regression analysis was performed on a DNAm (beta-valued) data matrix of 482,091 probes and 1,199 samples.

**Analysis of TCGA lung adenocarcinoma RNA-seq dataset.** The LUAD RNA-seq dataset from the TCGA<sup>45</sup> was normalized as described by us previously<sup>85</sup>. Here we focused our analysis on the 45 normal adjacent tumor pairs. To estimate TFA of the macrophage-specific TFs, we ran a linear regression of the normalized RNA-seq profile of a sample against the corresponding binding profile of the TF, with positive gene targets encoded as +1, no targets as 0, and negative regulations as -1. The TFA is the estimated t-statistic of the regression. When performing differential TFA analysis between normal adjacent and tumor tissue using a linear regression model, we adjusted for macrophage expression using LYZ and CD14 as markers, that is, the expression profiles of these genes were included as covariates. Kaplan–Meier survival analysis was performed using the survival 3.2.13 R package.

**Validation analyses.** In the case of Rfx5, a TF known to activate MHC class II genes, we validated the increased regulatory activity with age by studying the age-

related expression patterns of genes involved in the APM. Specifically, we obtained a list of 16 APM-related genes (*B2m*, *Tap1*, *Tapbpl*, *Calr*, *Psmb9*, *Psmb10*, *Erap1*, *Pdia3*, *Nlrc5*, *Psme1*, *Psme2*, *Psme3*, *Citta*, *Hsp90ab1*, *Hsp90aa1* and *Hsp90b1*) from work by Thompson et al.<sup>54</sup>. We then calculated Spearman correlations of these genes with age in each of 75 cell types from 11 different tissues. We used two-tailed *t*-tests to assess significance of increased/decreased expression trends with age in each of these 75 cell types. We also calculated the average Spearman correlation over the 16 APM genes for each of these cell types and asked if this correlated with the statistic of age association of Rfx5's TFA. Importantly, there is zero overlap between the ten Rfx5 regulon genes (*Cd74*, *Col1a2*, *H2-DMa*, *H2-DMb2*, *H2-DMb1*, *H2-Oa*, *H2-Ob*, *H2-Ab1*, *H2-Ea* and *H2-Eb2*) and the 16 APM genes, rendering any putative correlation between Rfx5 TFA and APM gene expression nontrivial. In the case of the DOROTHEA predictions in the skin TMS scRNA-seq dataset, we validated these findings by direct application of the DOROTHEA + VIPER algorithm in an independent human skin aging scRNA-seq dataset from Solé-Baldo et al.<sup>55</sup>. Finally, to validate predictions made for Nfkb1, we were able to obtain corresponding ChIP-seq data in a purified cell population (*Cd4<sup>+</sup>* T cells) from the ChIP-seq atlas<sup>22</sup>. The ChIP-seq peak profiles have been annotated to gene targets using three window sizes (1 kb, 5 kb and 10 kb). For each of these three gene-target lists, we removed those that were already present in the Nfkb1 DOROTHEA regulon. We note that the overlapping proportion was low, resulting in 73 (1 kb), 126 (5 kb) and 181 (10 kb) ChIP-seq targets that were not part of the Nfkb1 DOROTHEA regulon. For these ChIP-seq targets, we computed SCCs between their gene expression and age in the T cell populations of tissues where we had predicted an increased Nfkb1 regulatory activity with age. To assess validation strength, we tested if the SCCs of the ChIP-seq targets were significantly higher than zero.

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

## Data availability

Data analyzed in this manuscript is publicly available from the GEO ([www.ncbi.nlm.nih.gov/geo/](http://www.ncbi.nlm.nih.gov/geo/)) under accession numbers GSE56046 (DNAm data) and GSE130973 (skin aging scRNA-seq dataset), from EBI ArrayExpress (<https://www.ebi.ac.uk/arrayexpress>) under accession numbers E-MTAB-6149 and E-MTAB-6653 (lung cancer scRNA-seq dataset). FACS-sorted expression data from purified blood cell subtypes are available from <https://haemosphere.org/datasets/show>. The lung cancer bulk RNA-seq dataset is available from the TCGA data portal (<http://tcgaportal.org/>). The TMS scRNA-seq data are available from <https://doi.org/10.6084/m9.figshare.8273102.v2>. The regulons for the 328 hematopoietic TFs and their cell-specific lineage information are provided in Supplementary Data 1. The DOROTHEA regulons are available from <https://saezlab.github.io/dorothea>. ChIP-seq data were downloaded from the ChIP-seq Atlas (<https://chip-atlas.org/>). Processed data are available from the corresponding author upon reasonable request. Source data are provided with this paper.

## Code availability

SCIRA functions for estimating TFA are available from the scira R package (<http://github.com/aet21/scira/>). We also provide an R markdown file and associated data objects from the figshare repository ([https://figshare.com/articles/software/R-markdown\\_file\\_and\\_data\\_objects\\_for\\_estimating\\_TFA\\_in\\_TMS\\_lung-tissue\\_scRNA-seq\\_dataset/17167085](https://figshare.com/articles/software/R-markdown_file_and_data_objects_for_estimating_TFA_in_TMS_lung-tissue_scRNA-seq_dataset/17167085)) that illustrate in a few examples how to estimate TFA and how to correlate it with age in the TMS dataset.

Received: 30 June 2021; Accepted: 3 May 2022;

Published online: 6 June 2022

## References

- Lopez-Otin, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194–1217 (2013).
- Donertas, H. M., Fabian, D. K., Valenzuela, M. F., Partridge, L. & Thornton, J. M. Common genetic associations between age-related diseases. *Nat Aging* **1**, 400–412 (2021).
- Fabian, D. K., Fuentealba, M., Donertas, H. M., Partridge, L. & Thornton, J. M. Functional conservation in genes and pathways linking ageing and immunity. *Immun. Ageing* **18**, 23 (2021).
- Vijg, J. & Kennedy, B. K. The essence of aging. *Gerontology* **62**, 381–385 (2016).
- Brunauer, R., Alavez, S. & Kennedy, B. K. Stem cell models: a guide to understand and mitigate aging? *Gerontology* **63**, 84–90 (2017).
- Graf, T. & Enver, T. Forcing cells to change lineages. *Nature* **462**, 587–594 (2009).
- Yamanaka, S. & Blau, H. M. Nuclear reprogramming to a pluripotent state by three approaches. *Nature* **465**, 704–712 (2010).
- Zhou, X., Sen, I., Lin, X. X. & Riedel, C. G. Regulation of age-related decline by transcription factors and their crosstalk with the epigenome. *Curr. Genomics* **19**, 464–482 (2018).
- Palmer, D., Fabris, F., Doherty, A., Freitas, A. A. & de Magalhaes, J. P. Ageing transcriptome meta-analysis reveals similarities and differences between key mammalian tissues. *Aging* **13**, 3313–3341 (2021).
- Tabula Muris. C. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature* **583**, 590–595 (2020).
- Martinez-Jimenez, C. P. et al. Ageing increases cell-to-cell transcriptional variability upon immune stimulation. *Science* **355**, 1433–1436 (2017).
- Levy, O. et al. Age-related loss of gene-to-gene transcriptional coordination among single cells. *Nat Metab.* **2**, 1305–1315 (2020).
- Teschendorff, A. E. & Wang, N. Improved detection of tumor suppressor events in single-cell RNA-seq data. *NPJ Genom. Med.* **5**, 43 (2020).
- Holland, C. H. et al. Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data. *Genome Biol.* **21**, 36 (2020).
- Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
- Basso, K. et al. Reverse engineering of regulatory networks in human B cells. *Nat. Genet.* **37**, 382–390 (2005).
- Margolin, A. A. et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7**, S7 (2006).
- Chen, Y., Widschwendter, M. & Teschendorff, A. E. Systems-epigenomics inference of transcription factor activity implicates aryl-hydrocarbon-receptor inactivation as a key event in lung cancer development. *Genome Biol.* **18**, 236 (2017).
- Garcia-Alonso, L., Holland, C. H., Ibrahim, M. M., Turei, D. & Saez-Rodriguez, J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.* **29**, 1363–1375 (2019).
- de Graaf, C. A. et al. Haemopedia: an expression atlas of murine hematopoietic cells. *Stem Cell Reports* **7**, 571–582 (2016).
- Kanamori, M. et al. A genome-wide and nonredundant mouse transcription factor database. *Biochem. Biophys. Res. Commun.* **322**, 787–793 (2004).
- Oki, S. et al. ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO Rep.* **19**, e46255 (2018).
- Jojic, V. et al. Identification of transcriptional regulators in the mouse immune system. *Nat. Immunol.* **14**, 633–643 (2013).
- Heng, T. S., Painter, M. W. & Immunological Genome Project, C. The Immunological Genome Project: networks of gene expression in immune cells. *Nat. Immunol.* **9**, 1091–1094 (2008).
- Tabula Muris, C. et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018).
- Setty, M. et al. Characterization of cell fate probabilities in single-cell data with Palantir. *Nat. Biotechnol.* **37**, 451–460 (2019).
- Elyahu, Y. et al. Aging promotes reorganization of the CD4 T cell landscape toward extreme regulatory and effector phenotypes. *Sci. Adv.* **5**, eaaw8330 (2019).
- Jonkman, T. H. et al. Functional genomics analysis identifies T and NK cell activation as a driver of epigenetic clock progression. *Genome Biol.* **23**, 24 (2022).
- Moskowitz, D. M. et al. Epigenomics of human CD8 T cell differentiation and aging. *Sci. Immunol.* **2**, eaag0191 (2017).
- Hu, B. et al. Distinct age-related epigenetic signatures in CD4 and CD8 T Cells. *Front. Immunol.* **11**, 585168 (2020).
- Hu, B. et al. Transcription factor networks in aged naive CD4 T cells bias lineage differentiation. *Aging Cell* **18**, e12957 (2019).
- Gustafson, C. E., Cavanagh, M. M., Jin, J., Weyand, C. M. & Goronzy, J. J. Functional pathways regulated by microRNA networks in CD8 T cell aging. *Aging Cell* **18**, e12879 (2019).
- Kurachi, M. et al. The transcription factor BATF operates as an essential differentiation checkpoint in early effector CD8<sup>+</sup> T cells. *Nat. Immunol.* **15**, 373–383 (2014).
- Quigley, M. et al. Transcriptional analysis of HIV-specific CD8<sup>+</sup> T cells shows that PD-1 inhibits T cell function by upregulating BATF. *Nat. Med.* **16**, 1147–1151 (2010).
- Sokalski, K. M. et al. Deletion of genes encoding PU.1 and Spi-B in B cells impairs differentiation and induces pre-B cell acute lymphoblastic leukemia. *Blood* **118**, 2801–2808 (2011).
- Butcher, S., Chahel, H. & Lord, J. M. Review article: ageing and the neutrophil: no appetite for killing? *Immunology* **100**, 411–416 (2000).
- Feinberg, M. W. et al. The Kruppel-like factor KLF4 is a critical regulator of monocyte differentiation. *EMBO J.* **26**, 4138–4148 (2007).
- Alder, J. K. et al. Kruppel-like factor 4 is essential for inflammatory monocyte differentiation *in vivo*. *J. Immunol.* **180**, 5645–5652 (2008).
- Reynolds, L. M. et al. Age-related variations in the methylome associated with gene expression in human monocytes and T cells. *Nat. Commun.* **5**, 5366 (2014).
- Feinberg, M. W. et al. Kruppel-like factor 4 is a mediator of proinflammatory signaling in macrophages. *J. Biol. Chem.* **280**, 38247–38258 (2005).

41. Liao, X. et al. Kruppel-like factor 4 regulates macrophage polarization. *J. Clin. Invest.* **121**, 2736–2749 (2011).
42. Luo, X. et al. Kruppel-like factor 4 is a regulator of proinflammatory signaling in fibroblast-like synoviocytes through increased IL-6 expression. *Mediators Inflamm.* **2016**, 1062586 (2016).
43. Dann, E., Henderson, N. C., Teichmann, S. A., Morgan, M. D. & Marioni, J. C. Differential abundance testing on single-cell data using *k*-nearest neighbor graphs. *Nat. Biotechnol.* **40**, 245–253 (2021).
44. Lambrechts, D. et al. Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat. Med.* **24**, 1277–1289 (2018).
45. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).
46. Kriegsman, B. A. et al. Frequent loss of IRF2 in cancers leads to immune evasion through decreased MHC class I antigen presentation and increased PD-L1 expression. *J. Immunol.* **203**, 1999–2010 (2019).
47. Alvarez, M. J. et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* **48**, 838–847 (2016).
48. Jo, A. et al. PARIS farnesylation prevents neurodegeneration in models of Parkinson's disease. *Sci. Transl. Med.* **13**, eaax8891 (2021).
49. Podolsky, M. J. et al. Age-dependent regulation of cell-mediated collagen turnover. *JCI Insight* **5**, e137519 (2020).
50. Xia, J. et al. SIRT1 deacetylates RFX5 and antagonizes repression of collagen type I (COL1A2) transcription in smooth muscle cells. *Biochem. Biophys. Res. Commun.* **428**, 264–270 (2012).
51. Donato, A. J., Morgan, R. G., Walker, A. E. & Lesniewski, L. A. Cellular and molecular biology of aging endothelial cells. *J. Mol. Cellular Cardiol.* **89**, 122–135 (2015).
52. Wang, Y. et al. Global transcriptomic changes occur in aged mouse podocytes. *Kidney Int.* **98**, 1160–1173 (2020).
53. O'Brien, Z. K., Van Nostrand, E. L., Higgins, J. P. & Kim, S. K. The inflammatory transcription factors NFκB, STAT1 and STAT3 drive age-associated transcriptional changes in the human kidney. *PLoS Genet.* **11**, e1005734 (2015).
54. Thompson, J. C. et al. Gene signature of antigen processing and presentation machinery predicts response to checkpoint blockade in non-small cell lung cancer and melanoma. *J. Immunother. Cancer* **8**, e000974 (2020).
55. Solé-Boldo, L. et al. Single-cell transcriptomes of the human skin reveal age-related loss of fibroblast priming. *Commun. Biol.* **3**, 188 (2020).
56. Li, Z. & Rasmussen, L. J. TIP60 in aging and neurodegeneration. *Ageing Res. Rev.* **64**, 101195 (2020).
57. Schaum, N. et al. Ageing hallmarks exhibit organ-specific temporal signatures. *Nature* **583**, 596–602 (2020).
58. Krausgruber, T. et al. IRF5 promotes inflammatory macrophage polarization and T<sub>H</sub>1–T<sub>H</sub>17 responses. *Nat. Immunol.* **12**, 231–238 (2011).
59. Mantovani, A. & Longo, D. L. Macrophage checkpoint blockade in cancer back to the future. *N. Engl. J. Med.* **379**, 1777–1779 (2018).
60. Galdiero, M. R. et al. Tumor-associated macrophages and neutrophils in cancer. *Immunobiology* **218**, 1402–1410 (2013).
61. Porta, C. et al. Macrophages in cancer and infectious diseases: the 'good' and the 'bad'. *Immunotherapy* **3**, 1185–1202 (2011).
62. Chen, H. H. et al. IRF2BP2 reduces macrophage inflammation and susceptibility to atherosclerosis. *Circ. Res.* **117**, 671–683 (2015).
63. Frieler, R. A. et al. Myeloid-specific deletion of the mineralocorticoid receptor reduces infarct volume and alters inflammation during cerebral ischemia. *Stroke* **42**, 179–185 (2011).
64. Cruz, S. A. et al. Loss of IRF2BP2 in microglia increases inflammation and functional deficits after focal ischemic brain injury. *Front. Cell Neurosci.* **11**, 201 (2017).
65. Xu, L. S. et al. ETV6–RUNX1 interacts with a region in SPIB intron 1 to regulate gene expression in pre-B cell acute lymphoblastic leukemia. *Exp. Hematol.* **73**, 50–63 (2019).
66. Cartwright, T., Perkins, N. D. & Wilson, C. L. NFKB1: a suppressor of inflammation, ageing and cancer. *FEBS J.* **283**, 1812–1822 (2016).
67. Concetti, J. & Wilson, C. L. NFKB1 and cancer: friend or foe? *Cells* **7**, 133 (2018).
68. Acosta-Rodriguez, V. A., Rijo-Ferreira, F., Green, C. B. & Takahashi, J. S. Importance of circadian timing for aging and longevity. *Nat. Commun.* **12**, 2862 (2021).
69. Tonsfeldt, K. J. et al. The contribution of the circadian gene *Bmal1* to female fertility and the generation of the preovulatory luteinizing hormone surge. *J. Endocr. Soc.* **3**, 716–733 (2019).
70. Rakshit, K. & Matveyenko, A. V. Induction of core circadian clock transcription factor *Bmal1* enhances beta cell function and protects against obesity-induced glucose intolerance. *Diabetes* **70**, 143–154 (2021).
71. Breen, D. P. et al. Sleep and circadian rhythm regulation in early Parkinson disease. *JAMA Neurol.* **71**, 589–595 (2014).
72. Hulme, B. et al. Epigenetic regulation of BMAL1 with sleep disturbances and Alzheimer's disease. *J. Alzheimers Dis.* **77**, 1783–1792 (2020).
73. Ehlen, J. C. et al. *Bmal1* function in skeletal muscle regulates sleep. *eLife* **6**, e26557 (2017).
74. Oh, G. et al. Circadian oscillations of cytosine modification in humans contribute to epigenetic variability, aging, and complex disease. *Genome Biol.* **20**, 2 (2019).
75. Oh, E. S. & Petronis, A. Origins of human disease: the chrono-epigenetic perspective. *Nat. Rev. Genet.* **22**, 533–546 (2021).
76. Babagana, M. et al. Hedgehog dysregulation contributes to tissue-specific inflamming of resident macrophages. *Aging* **13**, 19207–19229 (2021).
77. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115 (2013).
78. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, 3 (2004).
79. Zhuang, J., Widschwendter, M. & Teschendorff, A. E. A comparison of feature selection and classification methods in DNA methylation studies using the Illumina Infinium platform. *BMC Bioinformatics* **13**, 59 (2012).
80. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
81. Jablonski, K. A. et al. Novel markers to delineate murine M1 and M2 macrophages. *PLoS ONE* **10**, e0145342 (2015).
82. Aryee, M. J. et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).
83. Troyanskaya, O. et al. Missing value estimation methods for DNA microarrays. *Bioinformatics* **17**, 520–525 (2001).
84. Teschendorff, A. E. et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450k DNA methylation data. *Bioinformatics* **29**, 189–196 (2013).
85. Yang, Z., Jones, A., Widschwendter, M. & Teschendorff, A. E. An integrative pan-cancer-wide analysis of epigenetic enzymes reveals universal patterns of epigenomic deregulation in cancer. *Genome Biol.* **16**, 140 (2015).

## Acknowledgements

This work was supported by National Science Foundation of China grants (31771464, 32170652 and 31970632). We thank the TMS consortium for making their data open access and to everyone else who shares data in public repositories.

## Author contributions

A.E.T. conceived the study, performed the statistical analyses and wrote the manuscript. A.K.M., X.H. and T.Z. helped with analyses.

## Competing interests

The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. The authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s43587-022-00233-9>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s43587-022-00233-9>.

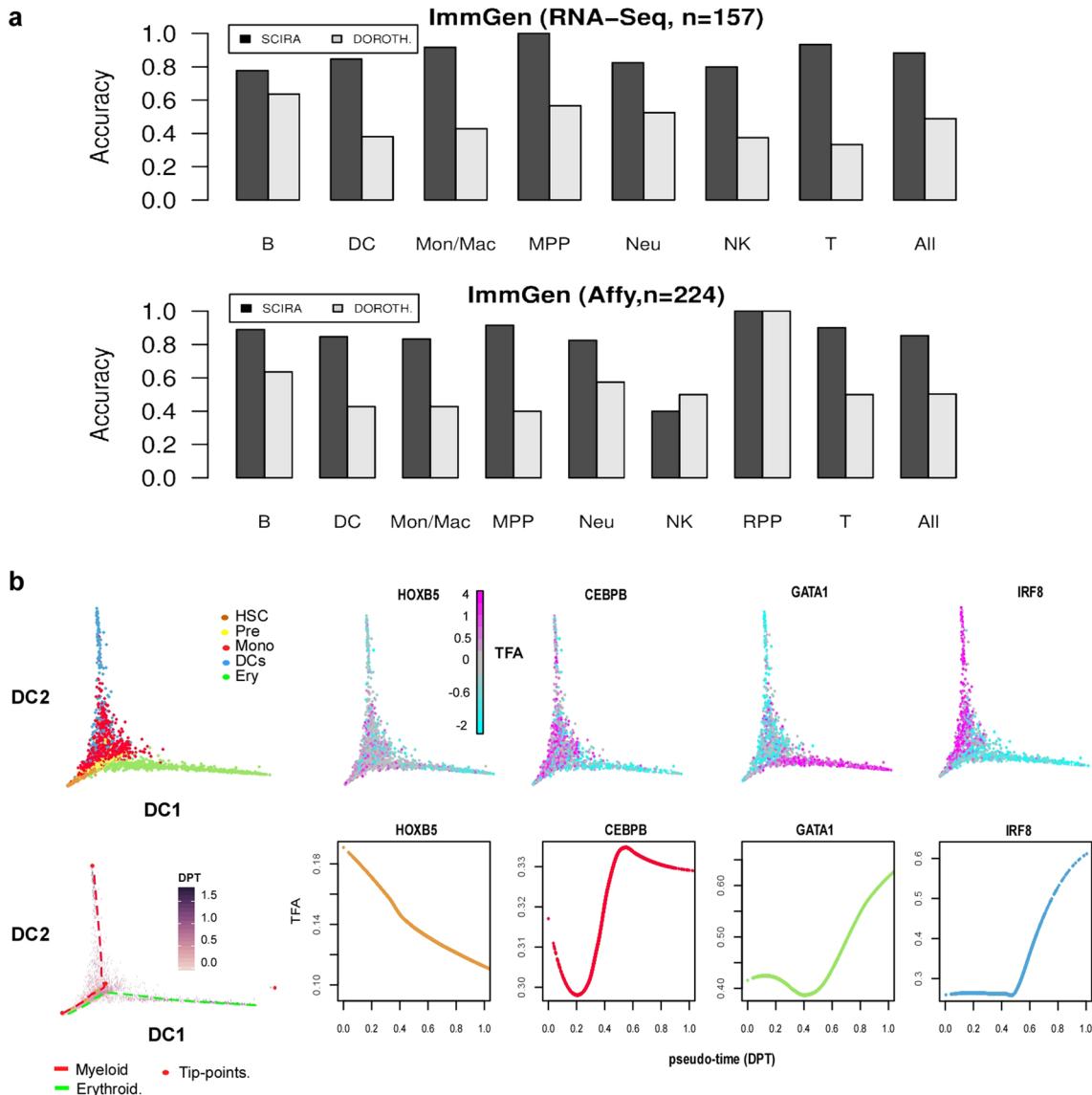
**Correspondence and requests for materials** should be addressed to Andrew E. Teschendorff.

**Peer review information** *Nature Aging* thanks Amir Bashan, Joao Pedro de Magalhaes and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

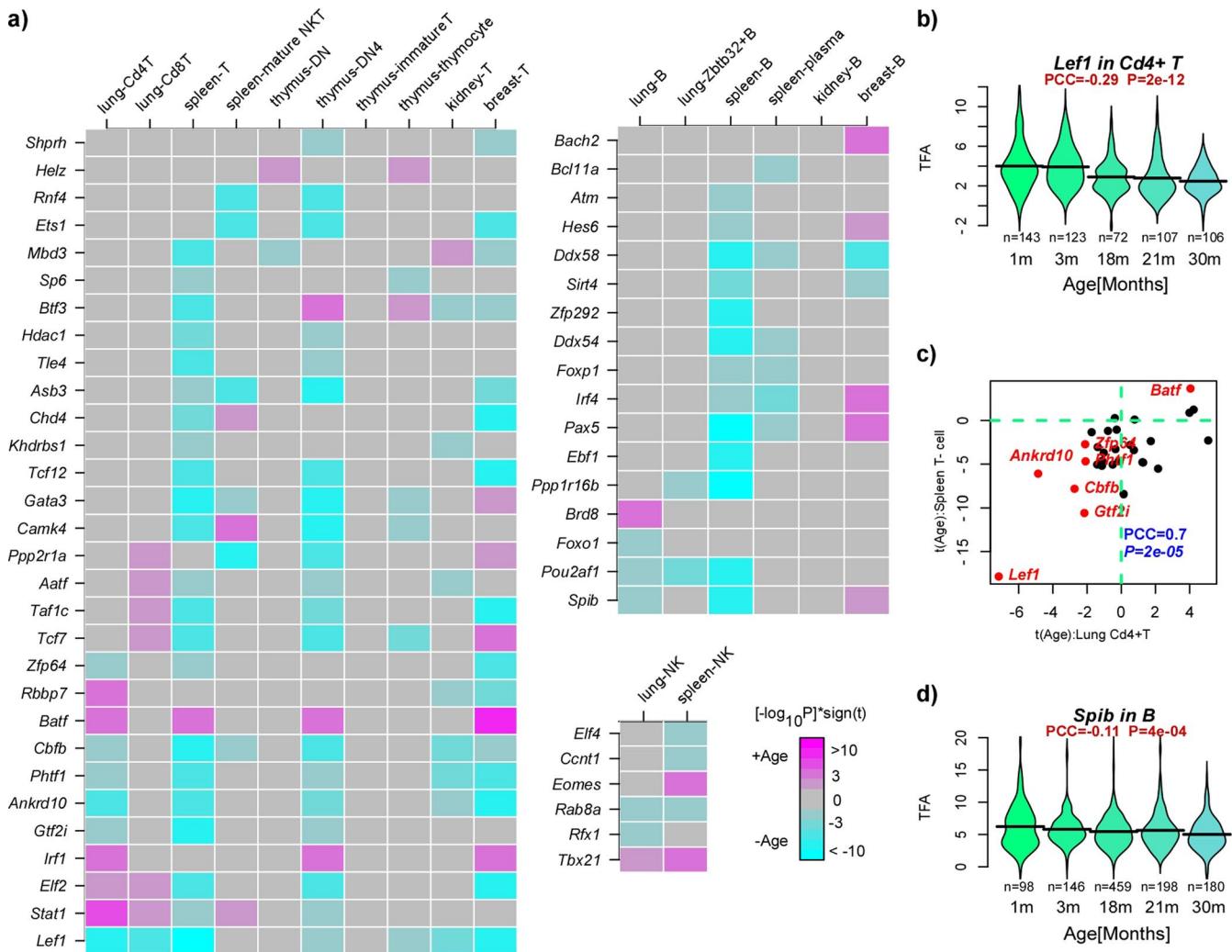
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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

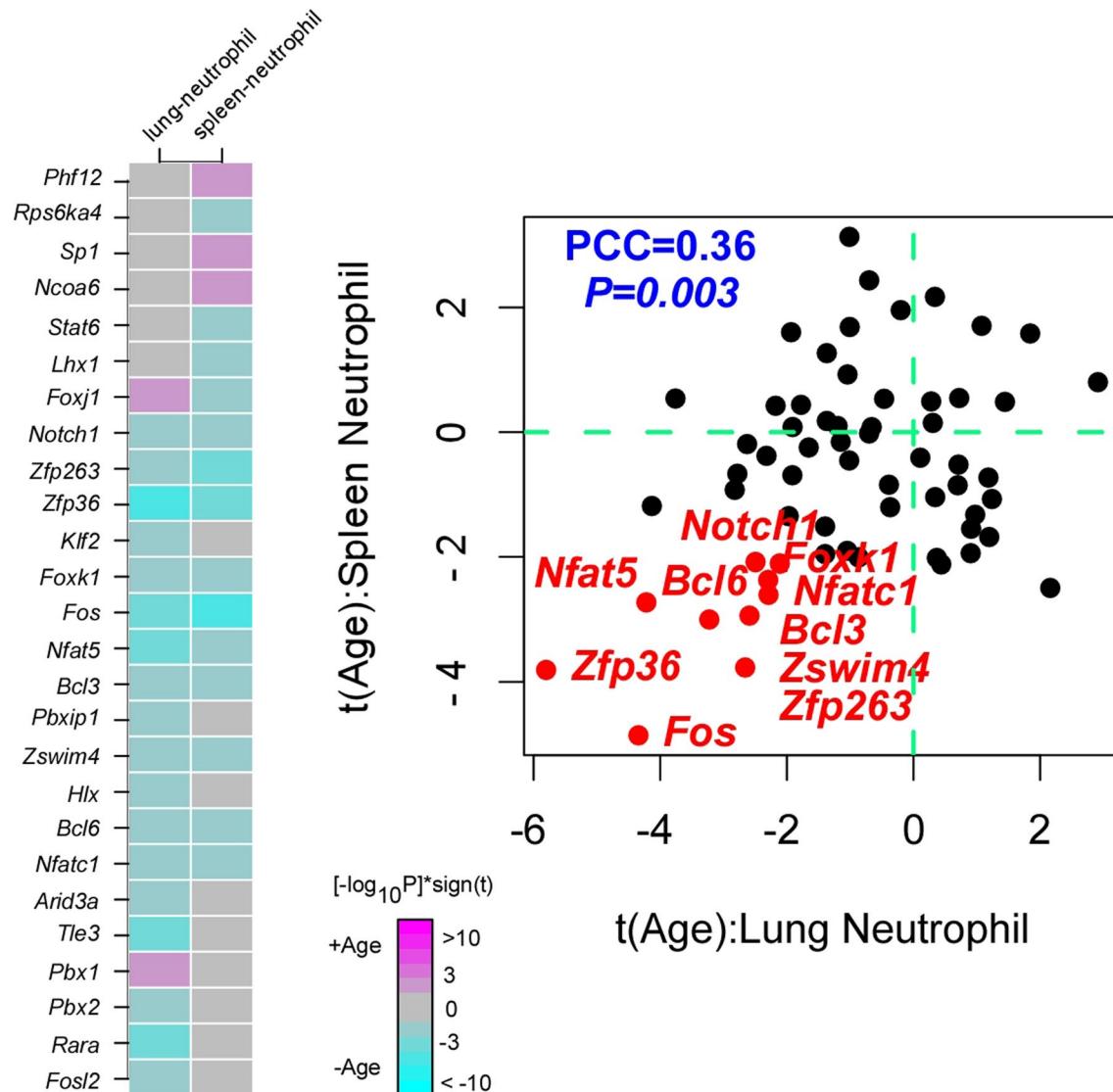
© The Author(s), under exclusive licence to Springer Nature America, Inc. 2022


**Extended Data Fig. 1 | Comparison of validation accuracies between SCIRA and Dorothea and validation of mouse hematopoietic TF-regulons. a)**

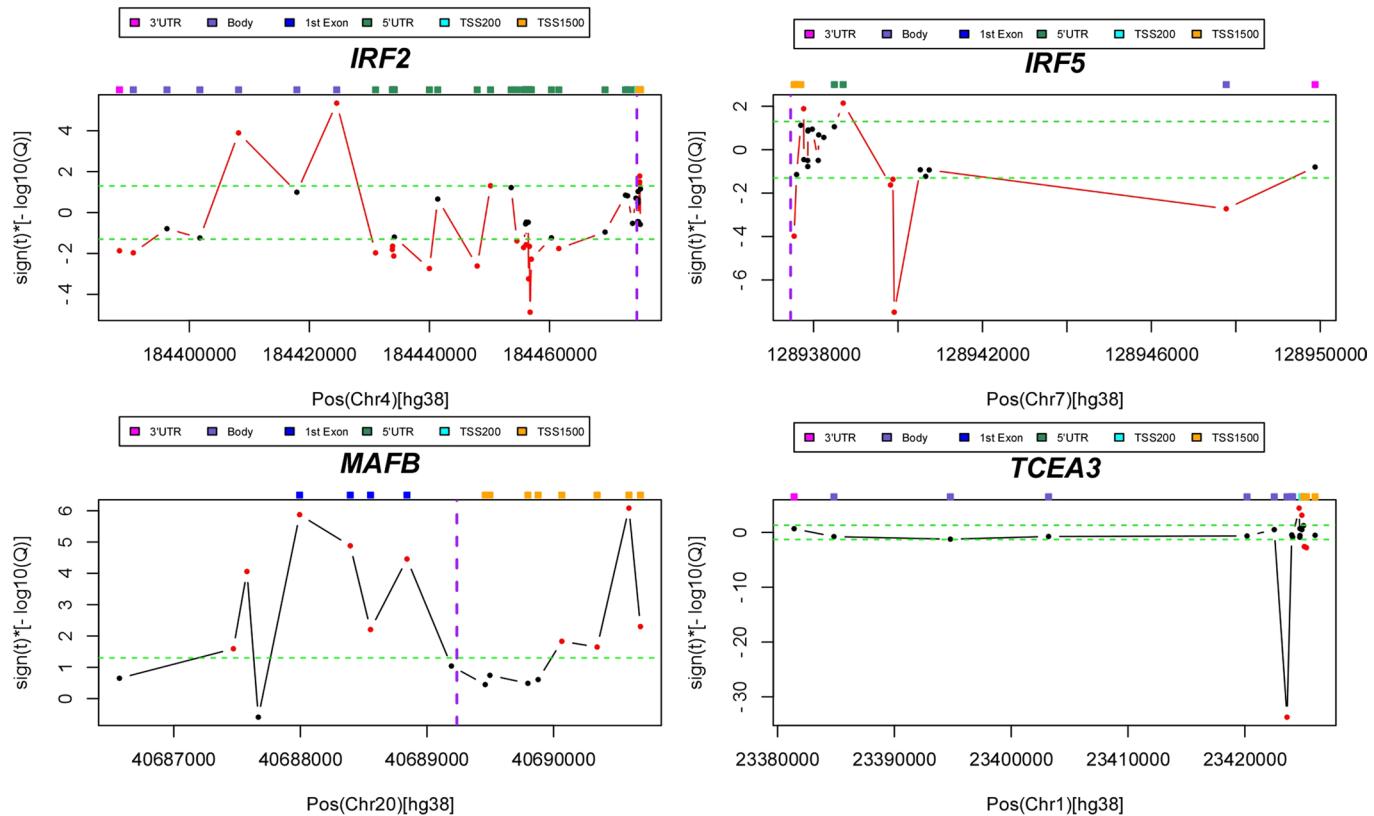
Barplots compare validation accuracies of estimated regulatory activities for blood-cell-lineage specific TF-regulons, obtained using the TF-regulons from SCIRA or those from the DOROTHEA database, as indicated. Top-panel is for mouse RNA-Seq ImmGen dataset which profiled purified blood cell subtypes from the various lineages as shown. Lower-panel is for the mouse ImmGen2 dataset. B = B-cell-lineage, DC = dendritic cell, Mon/Mac = Monocyte/Macrophage lineage, MPP = multipotent progenitor, Neu = Neutrophil, NK = Natural Killer, RPP = Restricted potency progenitor, T = T-cell lineage, All = accuracy obtained over all lineages. **b)** Diffusion maps on the left display diffusion components DC1 & DC2, of the transcription factor activity (TFA) matrix defined over the hematopoietic TFs and 4142 human bone marrow cells from Setty et al, with cells colored by cell-type as annotated in Setty et al (HSC = hematopoietic stem cell, Pre = precursor/multipotent progenitor, Mono = monocytes, DC = dendritic cells, Ery = Erythroid cells, top panel), and with cells colored by estimated diffusion pseudotime (DPT, lower panel). TFA was estimated using our SCIRA-derived TF-regulons. Tip-points are marked in red, displaying the starting branch point in the lower left corner, and the two terminal branch points representing the endpoints of the erythroid (green dashed line) and myeloid (red dashed line) branches. The diffusion maps to the right have cells colored by the estimated TFA of TFs that are known to be specific for HSCs (HOXB5), monocytes (CEBPB), erythrocytes (GATA1) and dendritic cells (IRF8) (top-panel). Lower panel displays the TFA of these same TFs as a function of diffusion pseudotime along their respective branches.



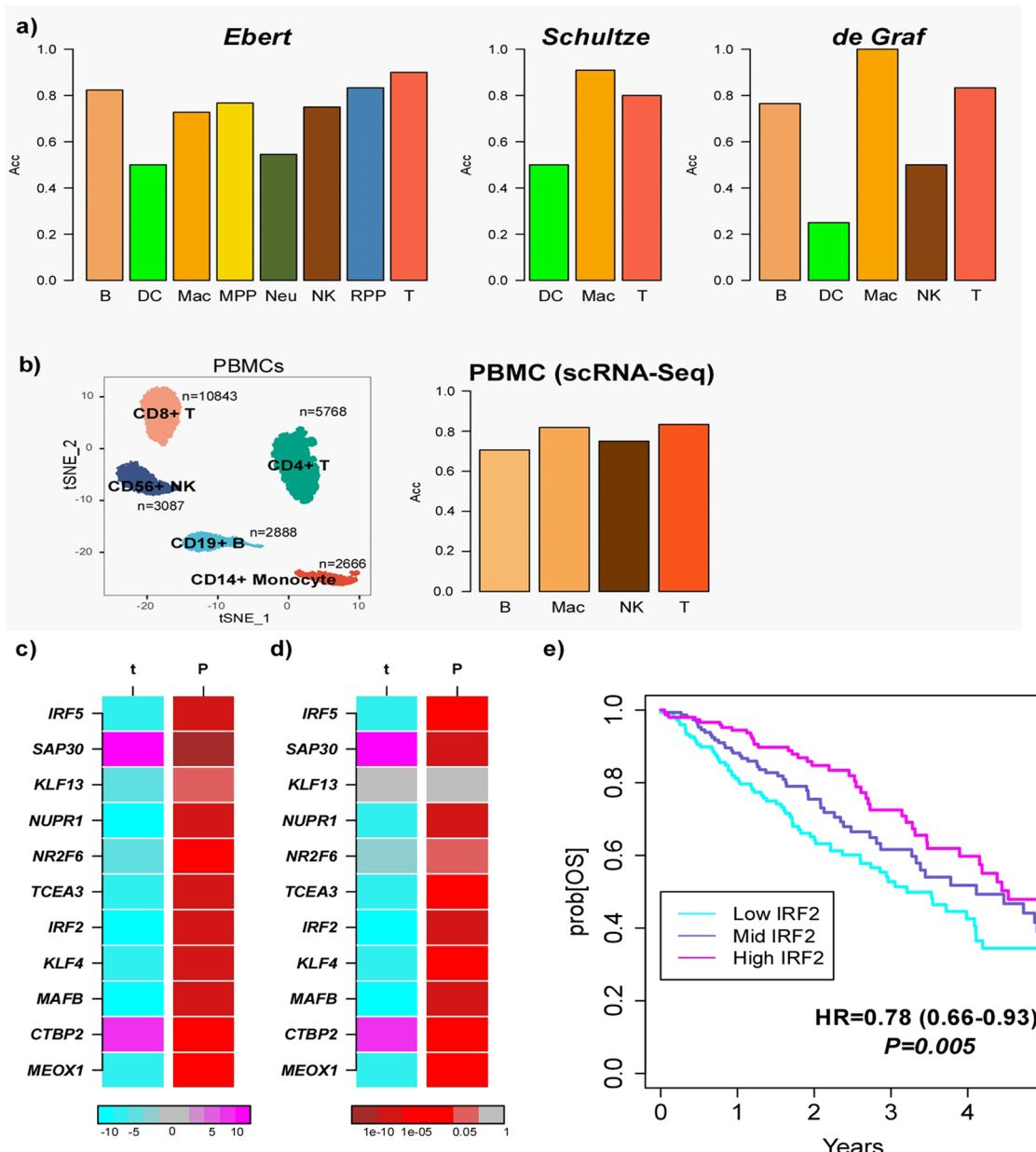
**Extended Data Fig. 2 | Cross-tissue comparison of age-associated TFA patterns lymphoid cells.** **a)** Heatmap panels display the age-associated patterns of differentiation activity (TFA) for lineage-specific TFs in different tissues and in corresponding cell-types of that lineage. Clockwise, panels represent the patterns for T-cell, B-cell and NK lineages. Each heatmap displays the signed statistical significance level as indicated, where the P-value is derived from the t-statistic of a regression of TFA against age, adjusting for sex. **b)** Violin plot displaying estimated TFA vs age-group for *Lef1* in lung Cd4+ T-cells. Pearson Correlation Coefficient (PCC) and associated P-value are shown. **c)** Scatterplot of the t-statistics of association of TFA with age for T-cell specific TFs as derived in CD4+ T-cells from lung (x-axis) vs. T-cells from spleen (y-axis). TFs significant in both tissues are colored. Pearson Correlation Coefficient and P-value between the two tissues is given. **d)** Violin plot displaying estimated TFA vs age-group for *Spib* in lung B-cells. Pearson Correlation Coefficient (PCC) and associated P-value are shown.



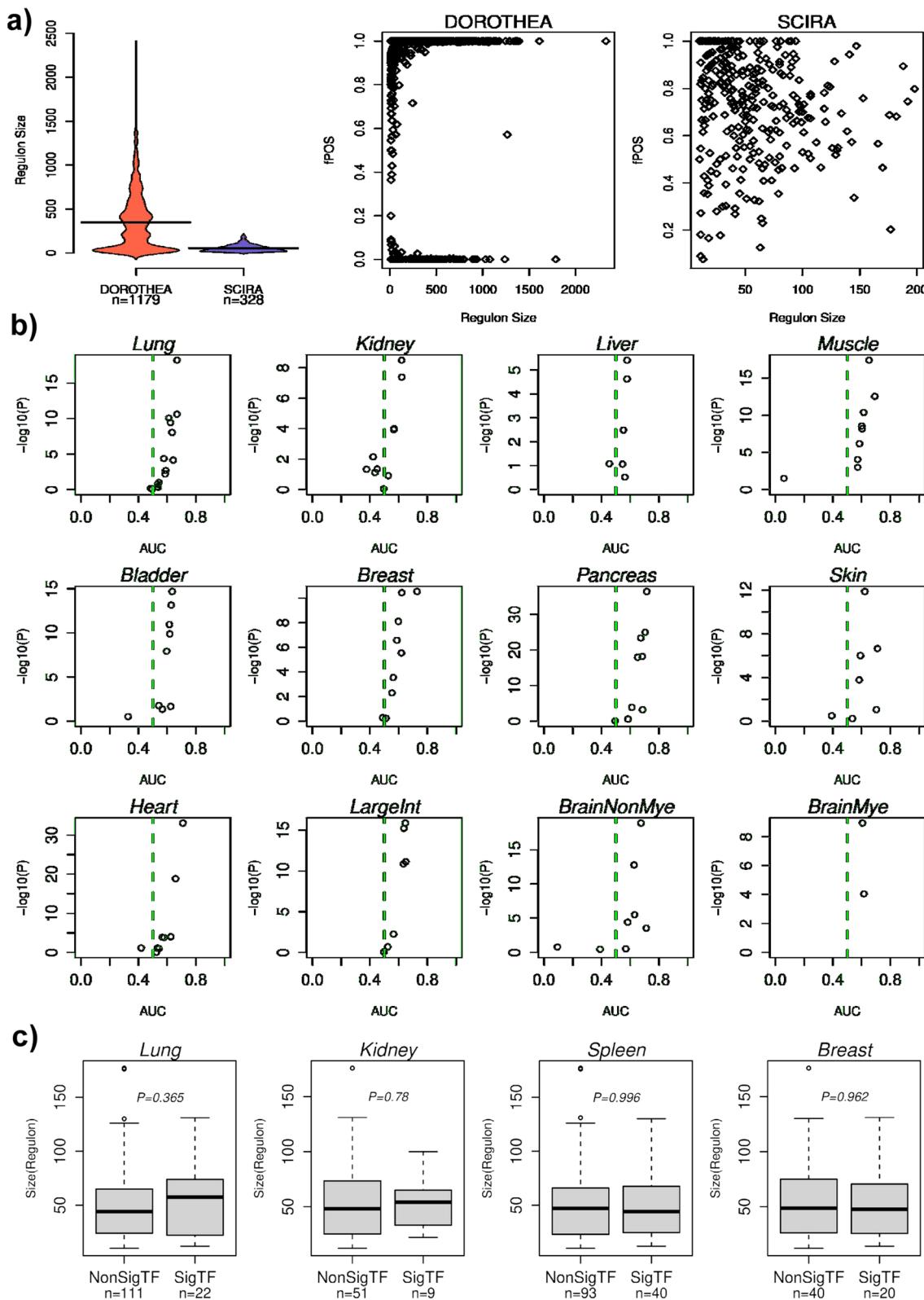
**Extended Data Fig. 3 | Cross-tissue comparison of age-associated TFA patterns neutrophils.** Heatmap displays age-associated differentiation activity (TFA) changes for neutrophils from lung and spleen, as inferred from the Tabula Muris Senis dataset. Scatterplot displays the corresponding t-statistics of association of TFA with age. TFs colored in red are significant in both tissues. Pearson Correlation Coefficient (PCC) and associated two-sided P-value between lung and spleen is given.



**Extended Data Fig. 4 | Age-associated DNA methylation at macrophage-specific TFs in purified monocyte samples.** For 4 macrophage-specific TFs that exhibit dysregulation of regulatory activity with age in monocytes/macrophages, as inferred from the TMS 10X scRNA-seq dataset, we display the signed statistical significance (y-axis) of age-association of Illumina 450k DNA methylation probes as derived from the 1200 purified monocyte sample set from Reynolds et al. In the y-axis we display the sign of the age-associated t-statistic times the negative  $\log_{10}[Q\text{-value}]$ , where Q-value is the FDR-estimate. X-axis labels the genomic position of the probes. Probes in red are significantly associated with age ( $Q < 0.05$ ).

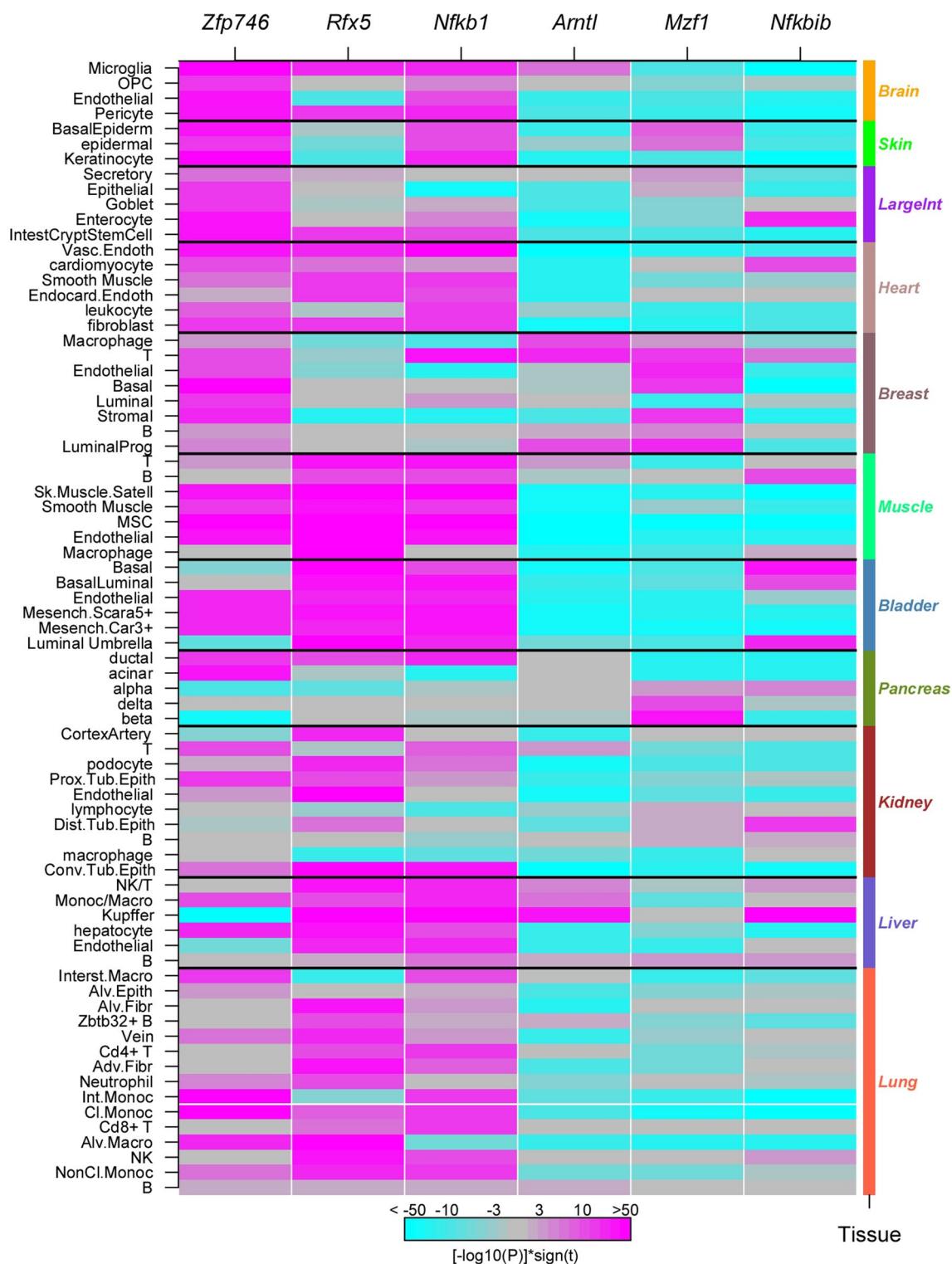


**Extended Data Fig. 5 | Validation of hematopoietic TF regulons in human datasets and TFA patterns of macrophage-specific TFs in lung adenocarcinoma.** **a-b) Validation of hematopoietic TF regulons in human datasets.** **a)** Barplots displaying the validation accuracy of blood cell-type specific SCIRa TF-regulons in 3 independent human FACS mRNA expression datasets (Ebert, Schultze & de Graf) from the Haemopedia resource. Accuracy was estimated as the fraction of cell-type specific TFs whose regulons predicted a significantly higher TFA in the corresponding blood cell types compared to all other cell-types, as assessed using a one-sided t-test ( $P < 0.05$ ). Only cell-types with at least 5 samples were included in this analysis. Total number of cell-sorted samples per dataset were:  $n = 211$  (Ebert),  $n = 384$  (Schultze),  $n = 42$  (de Graf). **b)** Left panel: tSNE-diagram of a 10X scRNA-Seq peripheral blood mononuclear cell (PBMCs) dataset from Zheng et al. Right panel: Barplot displaying the validation accuracy of blood cell type specific TF-regulons derived from cell-sorted bulk expression data from Haemopedia in the scRNA-Seq PBMC data from Zheng et al. **c-e) Pattern of TFA of macrophage-specific TFs in lung adenocarcinoma TCGA set.** **c)** Color bars displaying the t-statistics (t) and P-values (P) between the TFA of 11 macrophage-specific TFs and normal-cancer status using only paired samples ( $n = 45$  pairs). Thus, cyan indicates lower TFA in tumor vs normal. **d)** As a), but now adjusted for macrophage marker (LYZ & CD14) expression. **e)** Kaplan Meier overall survival curves for all primary LUAD samples, with samples stratified into low, middle and high tertiles according to IRF2 TFA activity. Hazard Ratio and chi-square test P-value derive from a Cox-regression of IRF2 TFA (treated as continuous variable).

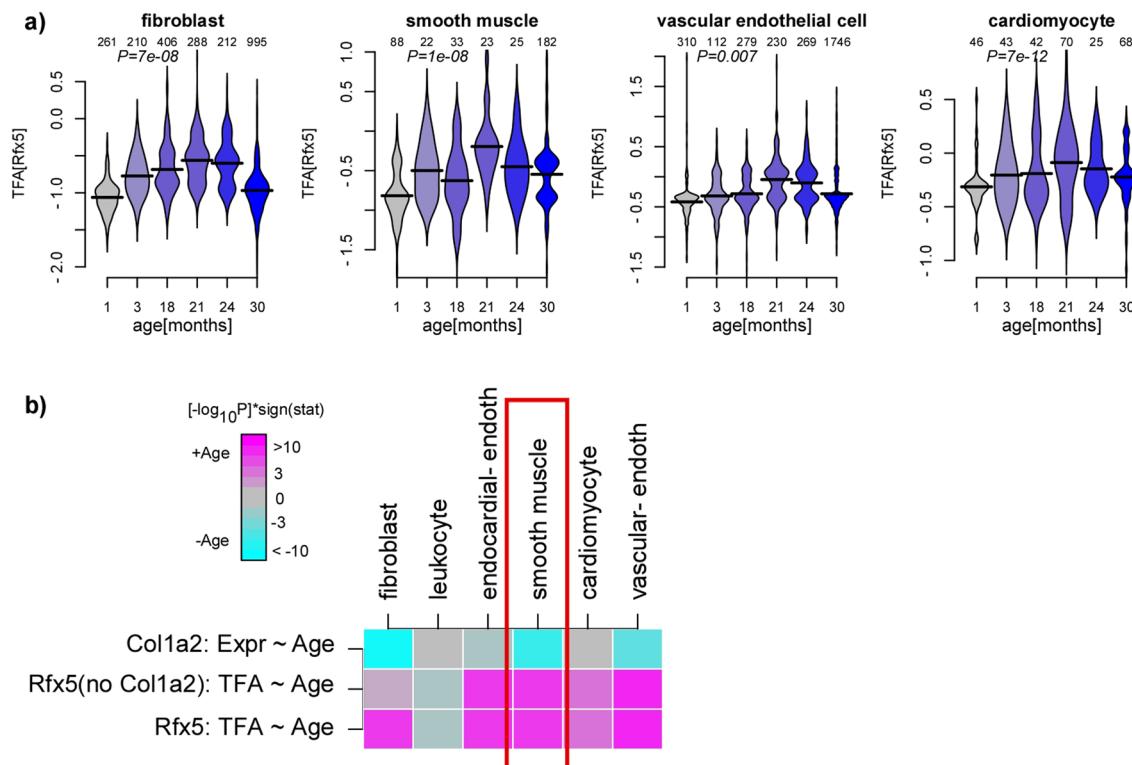


Extended Data Fig. 6 | See next page for caption.

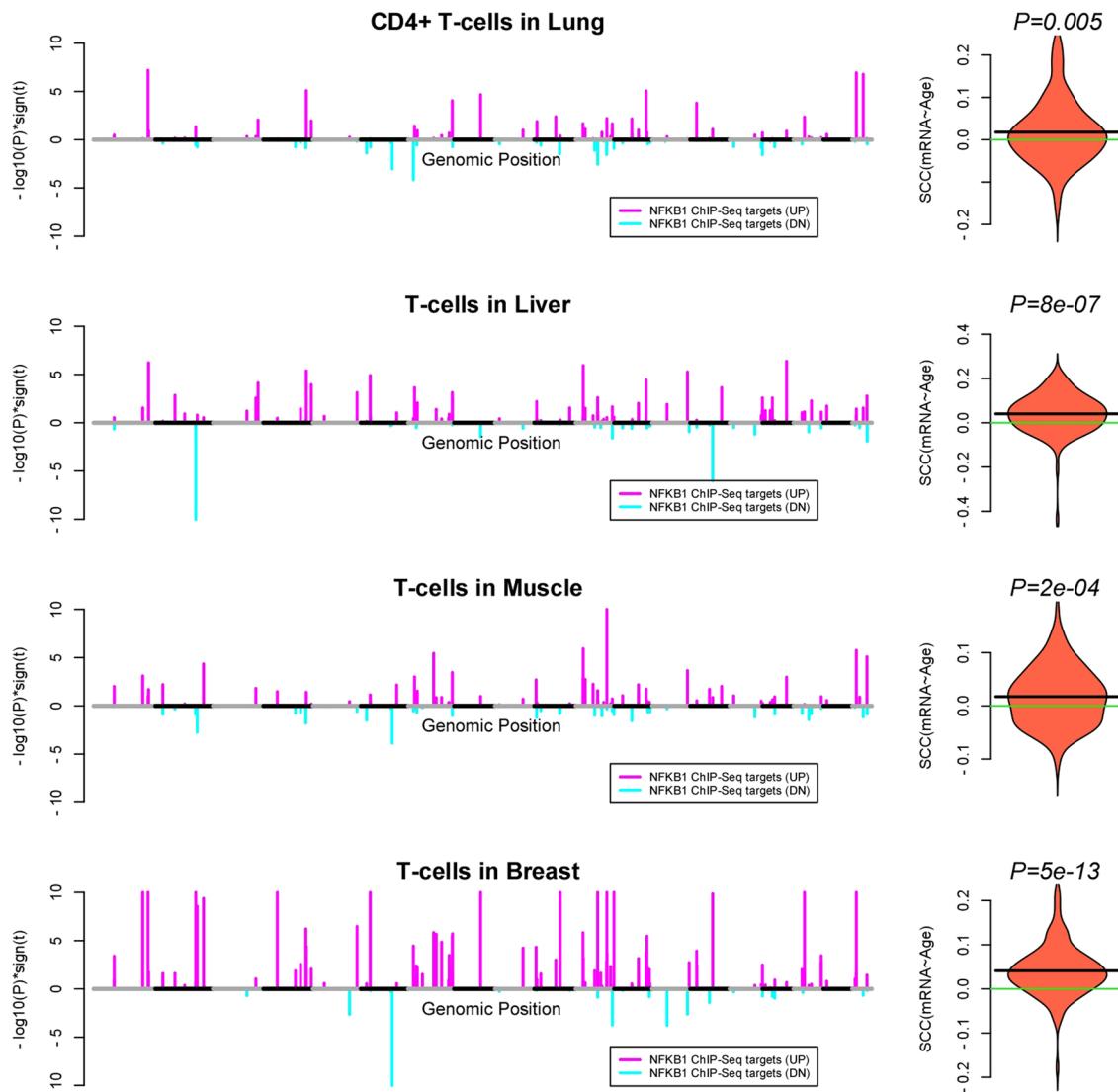
**Extended Data Fig. 6 | Regulon size comparison between DOROTHEA and SCIRA and their influence on age-associations.** **a)** Violin plots compare the regulon sizes of the DOROTHEA regulons vs the hematopoietic TF regulons built with SCIRA. The number of TF-regulons in each group is displayed below. Scatterplots to the right plot the regulon size (x-axis) vs. the fraction of regulon genes where the interaction between TF and gene is positive (fPOS, y-axis). **b)** Panels display scatterplots of statistical significance ( $-\log_{10}P$ , y-axis) vs AUC (x-axis) derived from a two-sided Wilcoxon-test comparing the regulon size of DOROTHEA TF-regulons that were significantly associated with age (Linear model  $P < 0.001$ ) vs the size of TF-regulons that were not associated with age, as assessed in cell-types within the given tissue. Each datapoint represents one distinct cell-type within the tissue. The green dashed-line indicates the line  $AUC = 0.5$ . Points to the right ( $AUC > 0.5$ ) are cell-types for which the regulon sizes of age-associated TF-regulons were higher compared to non age-associated ones, with the y-axis indicating statistical significance level. **c)** Boxplots display the SCIRA TF regulon size vs a binary variable indicating whether the TFA derived from the TF-regulon was significantly associated with age (SigTF) or not (NonSigTF), as assessed using a  $P < 0.001$  threshold. The number of TF-regulons in each category is given below each boxplot. The P-value comparing the regulon sizes between the significant and non-significant regulons is indicated within each panel and is derived from a two-tailed Wilcoxon rank sum test. The bar within each boxplot represents the median, the box itself the interquartile range (IQR), and whiskers extend to 1.5 times the IQR.



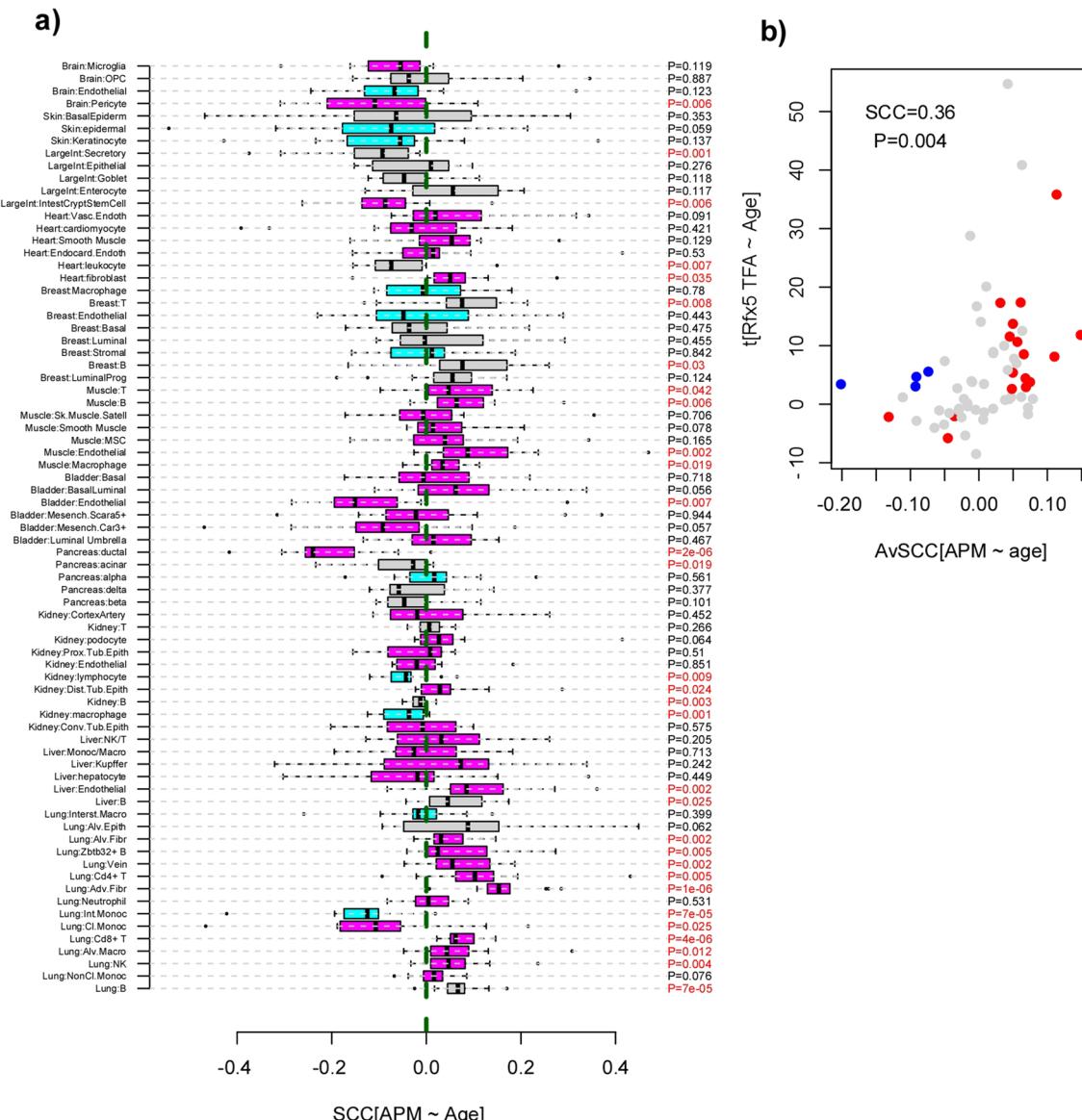
**Extended Data Fig. 7 | Age-associated changes of TFA for selected TFs across 75 cell-types.** Heatmap displays the signed significance of the age-associated t-statistics of TFA with age for 6 selected TFs across 75 cell-types and 11 tissue types as indicated. The 6 TFs were selected on the basis of displaying highly significant skews towards either increased or decreased regulatory activity with age.



**Extended Data Fig. 8 | Rfx5 activity changes in heart cells and inverse association with Col1a2 expression.** **a)** Violin plots display the regulatory activity (TFA) of Rfx5 against age [months] for 4 cell-types in heart, as indicated. The number of cells at each timepoint is indicated above violin plot. P-value shown is from a linear regression of TFA against age, adjusting for sex. **b)** Heatmap of significance statistics for Rfx5 TFA against age (including Col1a2 in its regulon), Rfx5 TFA against age not including Col1a2 in its regulon ('no Col1a2') and for Col1a2 expression against age.



**Extended Data Fig. 9 | Validation of Nfkbt increased activity in T-cells using CD4 + T-cell NFKB1 ChIP-Seq targets (5 kb).** Left panels: Genome plots display the signed statistical significance of NFKB1 target expression with age, as assessed in T-cells from the given tissue-type. Gene targets within 5 kb of the NFKB1 ChIP-Seq peak are shown with the color indicating increased (magenta) or decreased (cyan) mRNA expression with age. Alternating grey/black bars indicate successive mouse chromosomes. Right panels: Violin plots of the corresponding Spearman Rank Correlation Coefficients (SCC), with the P-value derived from a one-tailed t-test, testing for the null that the average SCC = 0. The number of ChIP-Seq targets is 126. Similar results are seen when defining NFKB1 ChIP-Seq targets as having a peak within 10 kb and 1 kb of the gene.



**Extended Data Fig. 10 | Association of antigen processing machinery (APM) and Rfx5 regulatory activity with age.** **a)** Boxplots display the Spearman rank correlation coefficient (SCC) between the gene-expression of APM genes ( $n=16$ ) with age, stratified according to tissue and cell-type. Boxes shown in magenta (cyan) indicate cell-types where Rfx5 activity increased (decreased) significantly with age ( $P < 0.05$ ). P-values above boxplots derive from a two-tailed t-test, testing if the average SCC values are significantly different from zero. P-values in red are those with  $P < 0.05$ . Number of datapoints in each boxplot is 16. The bar within each boxplot represents the median, the box itself the interquartile range (IQR), and whiskers extend to 1.5 times the IQR. **b)** A scatterplot of the average SCC with age (x-axis) against the t-statistic of association between Rfx5 regulatory activity and age (y-axis), with each datapoint representing one tissue-cell-type pair (as in a)). Datapoints highlighted in red indicate consistent significant associations, those in blue indicate inconsistent associations. Overall Spearman Correlation Coefficient (SCC) and two-sided P-value are given.

## 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	No particular code was used to collect data. Data was downloaded from the websites using accession numbers as given in the data-availability statement.
Data analysis	The main R-packages and version numbers used in this work include: scira_0.9.1 ( <a href="https://github.com/aet21/scira">https://github.com/aet21/scira</a> ), limma_3.40.6 ( <a href="http://bioconductor.org/packages/release/bioc/html/limma">http://bioconductor.org/packages/release/bioc/html/limma</a> ), corpcor_1.6.9 ( <a href="https://cran.r-project.org/web/packages/corpcor">https://cran.r-project.org/web/packages/corpcor</a> ), survival_3.2.13 ( <a href="https://cran.r-project.org/web/packages/survival/index.html">https://cran.r-project.org/web/packages/survival/index.html</a> ), viper_1.18.1 ( <a href="http://bioconductor.org/packages/release/bioc/html/viper">http://bioconductor.org/packages/release/bioc/html/viper</a> ), Seurat_4.0.3 ( <a href="http://bioconductor.org/packages/release/bioc/html/Seurat">http://bioconductor.org/packages/release/bioc/html/Seurat</a> ) and miloR_1.0.0 ( <a href="http://bioconductor.org/packages/release/bioc/html/miloR">http://bioconductor.org/packages/release/bioc/html/miloR</a> ). SCIRA functions for deriving regulons and estimating TFA are available from the scira R-package, <a href="http://github.com/aet21/scira">http://github.com/aet21/scira</a> . We also provide an R-markdown file and associated data objects from the figshare repository ( <a href="https://figshare.com/articles/software/R-markdown_file_and_data_objects_for_estimating_TFA_in_TMS_lung-tissue_scRNA-Seq_dataset/17167085">https://figshare.com/articles/software/R-markdown_file_and_data_objects_for_estimating_TFA_in_TMS_lung-tissue_scRNA-Seq_dataset/17167085</a> ), that illustrate in a few examples how to estimate TFA and how to correlate it with age in the TMS dataset.

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](#)

Data analyzed in this manuscript is publicly available from GEO ([www.ncbi.nlm.nih.gov/geo/](http://www.ncbi.nlm.nih.gov/geo/)) under accession numbers GSE56046 (DNA methylation data) and GSE130973 (skin-aging scRNA-Seq dataset), from EBI ArrayExpress (<https://www.ebi.ac.uk/arrayexpress>) under accession numbers E-MTAB-6149 and E-MTAB-6653 (lung cancer scRNA-Seq dataset). FACS-sorted expression data from purified blood cell subtypes is available from <https://haemosphere.org/datasets/show>. The lung cancer bulk RNA-Seq dataset is available from the TCGA data portal (<http://tcgaportal.org>). The TMS scRNA-Seq data is available from <https://doi.org/10.6084/m9.figshare.8273102.v2>. The regulons for the 328 hematopoietic TFs and their cell-specific lineage information is provided in Supplementary Data File 1. The DOROTHEA regulons are available from <https://saezlab.github.io/dorothea>. ChIP-Seq data was downloaded from the ChIP-Seq Atlas: <https://chip-atlas.org>

## Field-specific reporting

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

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

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://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	When deriving SCIRA regulons from the FACS-purified expression datasets, as well as when performing their validations in the bulk and TMS scRNA-Seq dataset, we were always adequately powered. In the TMS dataset, we excluded tissues and cell-types with measurements across less than 4 age timepoints and for fewer than 300 cells in total, with the exception of breast where thousands of cells were available across 3 age-groups.
Data exclusions	In general, no data was excluded. As mentioned above, in the TMS dataset, we excluded tissues and cell-types with insufficient cell-numbers across timepoints, to ensure adequate power to detect age-associated changes.
Replication	SCIRA-derived regulons were validated in two independent FACS-purified datasets, in 4 separate scRNA-Seq datasets, and using ChIP-Seq data from the ChIP-Seq Atlas. When reporting findings in the TMS-dataset, we focused on findings that were replicated across different tissue-and cell-types, e.g. endothelial cells from different tissue-types.
Randomization	Not applicable, as we only analyzed publicly available data.
Blinding	Not applicable, as we only analyzed publicly available data. Blinding can't be done in the context of our study because we need the age information to find age-associated TFA patterns.

## Reporting for specific materials, systems and methods

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

### Materials & experimental systems

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

### Methods

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