# Smurfness-based two-phase model of ageing helps deconvolve the ageing transcriptional signature

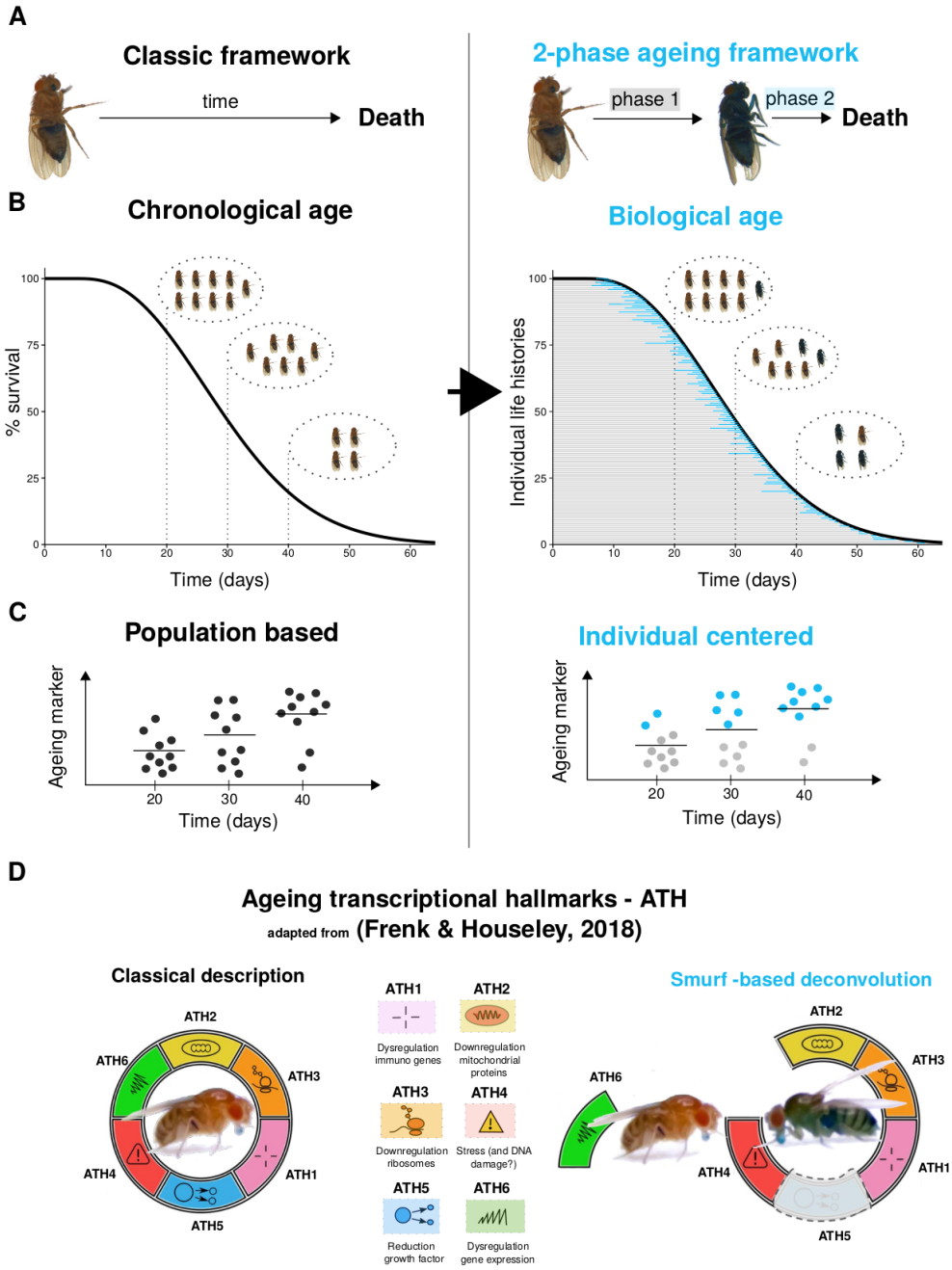
Flaminia Zane[1,2], Hayet Bouzid[1,2], Sofia Sosa Marmol[1], Savandara Besse[1], Julia Lisa Molina[2], Céline Cansell[3], Fanny Aprahamian[4,5], Sylvère Durand[4,5], Jessica Ayache[5], Christophe Antoniewski[2], Michael Rera[1]

Affiliations:
[1] Université Paris Cité, INSERM UMR U1284, 75004 Paris, France
[2] Sorbonne Université, Institut de Biologie Paris Seine, 75005, Paris, France
[3] Université Paris-Saclay, AgroParisTech, INRAE, UMR PNCA, 91120, Palaiseau, France
[4] Metabolomics and Cell Biology Platforms, UMS AMMICa, Institut Gustave Roussy, Villejuif 94805, France.
[5] Centre de Recherche des Cordeliers, Equipe labellisée par la Ligue contre le cancer, Université de Paris, Sorbonne Université, INSERM U1138, Institut Universitaire de France, Paris 75006, France.
[5] Université Paris Cité, Institut Jacques Monod, CNRS UMR 7592, 75013 Paris, France
corresponding author: michael.rera@cri-paris.org

## Abstract

Ageing is a common feature of living organisms, showing shared molecular features called hallmarks of ageing. Usually they are quantified in groups of individuals as a function of their chronological age (time passed since birth) and display continuous and progressive changes. Such approaches are based on the assumption that individuals taken at a given chronological age are biological replicates. However, even in genetically homogeneous and synchronised populations individuals do die at different chronological ages. This highlights the difference between chronological age and biological age, the latter being defined by the actual mortality risk of the organism, reflecting its physiology. The Smurf assay, previously described by Rera and colleagues, allows the identification of individuals at higher risk of death from natural causes amongst a population of a given chronological age. We found that the categorization of individuals as Smurf or non-Smurf, permits to distinguish transcriptional changes associated with either chronological or biological age. We show that transcriptional heterogeneity increases with chronological age, while four out of the six currently defined transcriptional hallmarks of ageing are associated with the biological age of individuals, i.e. their Smurf state. In conclusion, we demonstrate that studying properties of ageing by applying the Smurf classification allows us to differentiate the effect of time from the effect of a physiological response triggering an end-of-life switch (i.e. Smurf phase). More specifically, we show that the ability to isolate a pre-death phase of life *in vivo* enables us not only to study late life mechanisms preceding death, but also investigate early physiological changes triggering such phase. This allowed the identification of novel pro-longevity genetic interventions.
We anticipate that the use of the evolutionary conserved Smurf phenotype in ageing studies will allow significant advances in our comprehension of the underlying mechanisms of ageing.

# Graphical abstract




**The two-phase model of ageing allows to study separately the effect of chronological and physiological age. (A)** Classic approaches for studying ageing tend to consider it as a black box affecting all individuals progressively from birth to death. Instead, the Smurf phenotype shows that life can be divided into two consecutive phases separated by an abrupt transition. **(B)** All individuals undergo this transition at a different moment in their life, prior to death. This allows us to switch from population based approaches, comparing bulks of age-matched individuals through time, to individuals-centred approaches relying on direct access to their transition status. **(C)** Such paradigm shift shows that hallmarks of ageing long thought to progressively change with age are actually mostly affected in a growing proportion of Smurfs, allowing for the identification of the chain of events accompanying ageing and death from natural causes. **(D)** By studying the behaviour of the ageing transcriptome as a function of chronological age and Smurfness separately, we demonstrate that the progressively changing transcriptional ageing signature, as described in Frenk & Houseley (2018), is in fact the convolution changes accompanying chronological age signature (increased transcriptional noise) and changes associated with Smurfness (or biological age) signature (increased stress response and inflammation, decreased expression of ribosomal and mitochondrial genes). We also identified a hallmark partially associated with only old Smurfs (ATH5), suggesting that chronological age can affect, late in life, the Smurf response.

# Introduction

## Chronological and physiological ageing

In humans and multiple model organisms such as *Drosophila melanogaster*, ageing is commonly defined as a progressive decrease in functional efficiency associated with an age-related increasing vulnerability to death[1]. Even though the so-called hallmarks of ageing[1], or physiological changes linked with ageing, have been described, the argument about the origins and consequences of the process is still ongoing[2]. A major limitation of ageing studies is the lack of standard markers to discern the biological age - i.e. individuals' instant mortality risk - from chronological age - i.e. the time passed since birth. Even though the chronological age has been referred to as an "imperfect surrogate method to study the ageing process"[3], it is still the main marker employed in ageing studies with model organisms due to its simplicity. However, the ability to follow biological age *in vivo* would allow us to study ageing as it occurs in an organism, without the interindividual variability inherent to using chronological age as the main ageing marker.

In a given population, individuals can be of the same chronological age, i.e. born at the same time, but of different biological ages, i.e. showing different risks of mortality. In humans, the concept of frailty (defined as the combination of ageing markers, diseases, and other factors - such as fitness, nutritional status - that make an individual vulnerable and at a higher risk of death compared to his peers) was introduced as an estimation of an individual's biological status[4–6]; it was followed by the definition of frailty indexes, a fix set of biological parameters that can be used to predict the vulnerability of an individual and its risk of death independently of its chronological age[7–10].

In the last decade, ageing clocks have been developed with the aim of the biological age of an organism according to a set of molecular markers. Different epigenetic clocks have been proposed in humans and mice, all based on the changing 5-cytosine methylation of CpG sites (regions with repetition of cytosine and guanine)[3,11–13]. Even though performing well in humans and mammals, they do not apply to model organisms with no or low level of methylation, such as *Caenorabditis elegans* or *D. melanogaster,* which are however widely used in the field. Research has been conducted in order to identify an "universal" transcriptomic clock using *C. elegans*[14], with the recent publication of the BiT age clock[15], which was shown to apply also on human cells. In addition, plasma-based proteomic clocks have also been recently assessed in humans[16,17]. However, these tools do not allow for easy and non-invasive tracking of age, and might therefore not be suitable for biomarking age *in vivo* in a research context, especially using model organisms.

## The Smurf-based approach for studying ageing

The Smurf assay allows an *in vivo* assessment of increased intestinal permeability (IP) and was previously shown to be a powerful marker of biological age in *D. melanogaster*[18] , as well as other model organisms[19]. Flies with impaired IP are identified through co-ingestion of the non-toxic blue food dye FD&C #1. The dye, normally not absorbed by the digestive tract,

spreads throughout the body in flies with altered IP, turning them blue[18,20] (hence their name, Smurfs) (Fig. 1a). When ageing a population on standard food containing the above-mentioned dye, the proportion of Smurfs increases as a function of time[18], with all the flies undergoing a final Smurf transition prior to death[18,21]. Furthermore, Smurfs present a low and constant remaining life expectancy ($T_{50}$ estimated at ~ 2.04 days across different genetic backgrounds from the DGRP set[22]) independently of their chronological age when the transition occurs[18,21]. The survival dynamics of an ageing population can be decomposed into the joint dynamics of Smurf and non-Smurf sub-populations[21] (Fig. 1b). While the absolute number of Smurf flies peaks around the $T_{50}$, the proportion of Smurf flies ($n_{s(t)}/[n_s+n_{ns}]_{(t)}$) increases with age in a way that has been so far approximated as being linear[18]. Interestingly, Smurfs were shown to be the only individuals in the population carrying ageing markers such as increased transcription of inflammatory genes, decreased triglycerides and glycogen stores, impaired mobility[18] and reduced fertility[23]. The above-mentioned studies led to the hypothesis that markers, currently considered as progressively and continuously changing during ageing, might actually exhibit a biphasic behaviour accompanying the Smurf transition (two-phase model of ageing[21,24], Fig. 1c). Given the evolution of the two sub-populations, the time-progressive increase in mortality and ageing markers described population can then be re-interpreted as the time-progressive increase of Smurf individuals' proportion carrying these changes (Fig. 1d). The use of the Smurf phenotype could then allow us to switch from population-based to individual-based ageing studies, by identifying at each time-point individuals that are biologically older and at a higher risk of death than their age-matched peers (see graphical abstract).

## Smurf-based study of ageing

The near constant remaining life expectancy of Smurfs and the observed biphasic behaviour of the analysed ageing markers suggest the presence of a physiologically and molecularly stereotypical late phase of life, which might be characterised by common upstream regulators. In order to investigate our hypothesis, we proceeded by assessing the transcriptional changes occurring in flies as a function of both their Smurf status and chronological age. In the past years, conserved ageing transcriptional markers have been described[25], allowing for direct comparison with our results.

RNA-Sequencing (RNA-Seq) was performed on Smurf and non-Smurf individuals after total RNA extraction from the whole body of mated female flies. To consider age-related effects, samples were collected at 20, 30 and 40 days after eclosion, which in the used line (*Drs*-GFP) corresponds to approximately 90%, 50% and 10% survival (Fig. S1). In brief, flies were transferred on blue medium and Smurfs collected the next morning. The first collection corresponds to individuals that have been Smurfs for an unknown amount of time, the second for 5 hours maximum and the last for 24 hours maximum. This allows for the exploration of the Smurf transcriptome evolution post transition. We previously showed that all females turn Smurf prior to death in this genetic background[21] while it is debated in others[18,20,26] . In addition, the GFP signal was previously shown to be a good surrogate for Smurfness[18,23] (Fig. S2). Further details are presented in Materials and Methods.

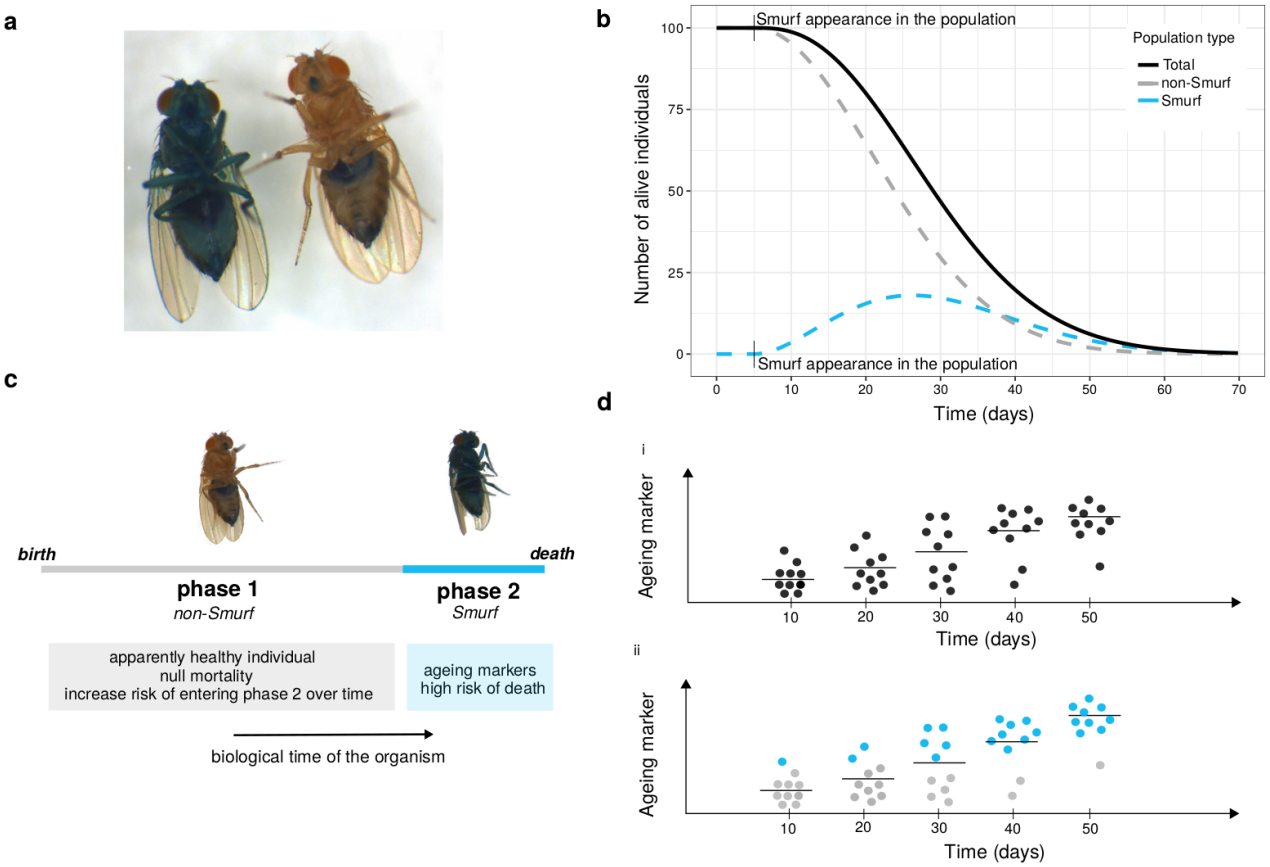



**Figure 1. Using the two-phase ageing model to deconvolve chronological and physiological age. a) Smurf phenotype.** Smurf and non-Smurf female flies (*Drs*-GFP line), 18 days old. After ingestion of FD&C blue#1-containing medium, non-Smurf flies present the dye restricted to the digestive tract; Smurf are entirely blue, with the exception of the wings and the eyes. **b) Dynamics of Smurf and non-Smurf subpopulations during ageing and how it affects random sampling. i) Population dynamics.** The survival curve of the population (in black) is the result of the sum of the non-Smurf (grey dashed curve) and Smurf (blue dashed curve) population dynamics. The first Smurf appears when the population is about to leave the survival plateau, which in this simulated case occurs at day 5. From that moment, while the non-Smurf population decreases up to the moment when no individuals are left, the Smurf population initially increases in absolute number (peaking around the $T_{50}$ of the population) and subsequently decreases as few individuals are left alive in the general population. Note that the proportion of Smurfs, computed at a given time point as the number of Smurfs over the total number of individuals alive, increases over time[18]. The survival curve and corresponding sub-populations dynamics are simulated from the following 2PAC model parameters: a = 0.0039 , b = -0.019, k = 0.1911. Those parameters were experimentally estimated from longevity of isogenic *Drs*-GFP mated female population[21]. **c) 2-phase model.** The life of the individual may be divided into 2-phases. In phase 1, or non-Smurf phase, the individual does not present ageing markers, showing in particular a null mortality. However, with time it experiences an increased probability of entering phase 2, or Smurf phase. In this last stage, the individual presents ageing markers and a high risk of death. Flies can enter phase 2 at a different chronological age, while spending on average the same amount of time in phase 2, independently of their age. **d) Evolution of an ageing marker in a population.** An ageing marker is a process (such as for instance expression of a gene or accumulation of a protein) which marks the state of ageing of an individual. By studying the evolution of an ageing marker by chronological age using random sampling, we will observe a progressive average increase in the population with time **(i)**. According to the Smurf model, the progressive increase of the marker at a population level would correspond to an increasing sampling bias for Smurfs with time. If we consider the curves in **(b)**, on day 10, when only a few Smurfs are present, the probability of sampling them is low, and a possible Smurf sample might result in an outlier compared to the population average. The increasing proportion of Smurf over time leads to an increasing representation of Smurfs in the samples, with the detection of the so-called ageing signal carried by such individuals.

# Results

## Smurfs have a stereotypical transcriptome

We first performed a Principal Component Analysis (PCA) to explore how our multiple samples did relate to each other. Results showed a prominent separation by the first component (45% of variance explained) of Smurf and non-Smurf samples, independently of their chronological age (Fig. 2a). This component is significantly associated with Smurfness ($R^2$ ANOVA = 0.604, p-value = $1.67e^{-07}$), while no significant correlation with age is found (p-value > 0.05). The second component (13%) segregates samples as a function of their age (Pearson ϱ = 0.717, p-value = $3.92e^{-06}$), with no significant association with Smurfness (p-value > 0.05). Overall, these results suggest that the transcriptome of our samples is shaped by both Smurfness and age, with Smurfness accounting for around three times as much changes as chronological age. However, we can notice the presence of three 40 days Smurfs samples out of six clustering with the age-matched non-Smurfs, a trend confirmed using independent tSNE (t-distributed stochastic neighbour embedding) and hierarchical clustering on sample-to-sample distance (Fig. S3 and S4). This indicates less transcriptomic differences between old Smurfs and non-Smurfs than in young ones. Given the noteworthy separation identified, we proceeded to quantify the differences between Smurfs and non-Smurfs, independently of their age, through differential gene expression analysis (DESeq2[27]). By comparing the 16 Smurf samples with the 16 non-Smurfs, we identified 3009 differentially expressed genes (DEGs), with a cutoff of 5% on the adjusted p-value (false discovery rate, FDR) (Fig 2b) (DESeq2 results in Supplementary File 1). Confirming the PCA result, these genes represent a Smurf-specific signature which is able to almost perfectly cluster the Smurfs samples together independently from their age (Fig 2c). Once again, the Smurfs do not visibly cluster according to age, while non-Smurf samples do. This could suggest an age-dependent behavior of these genes in the non-Smurfs. In addition, following a trend already visible in the PCA, 5-hours (5h) and 24-hours (24h) samples do not show a discernible segregation, suggesting that the Smurf signature is not majorly affected by the time passed from Smurf transition. DESeq2 results were validated by using the edgeR[28] pipeline, which identified 2609 DEGs, 90% of which are overlapping with the DESeq2 output and present a strong correlation (Pearson ϱ = 0.99) for $\log_2$FC estimation (Fig. S5).

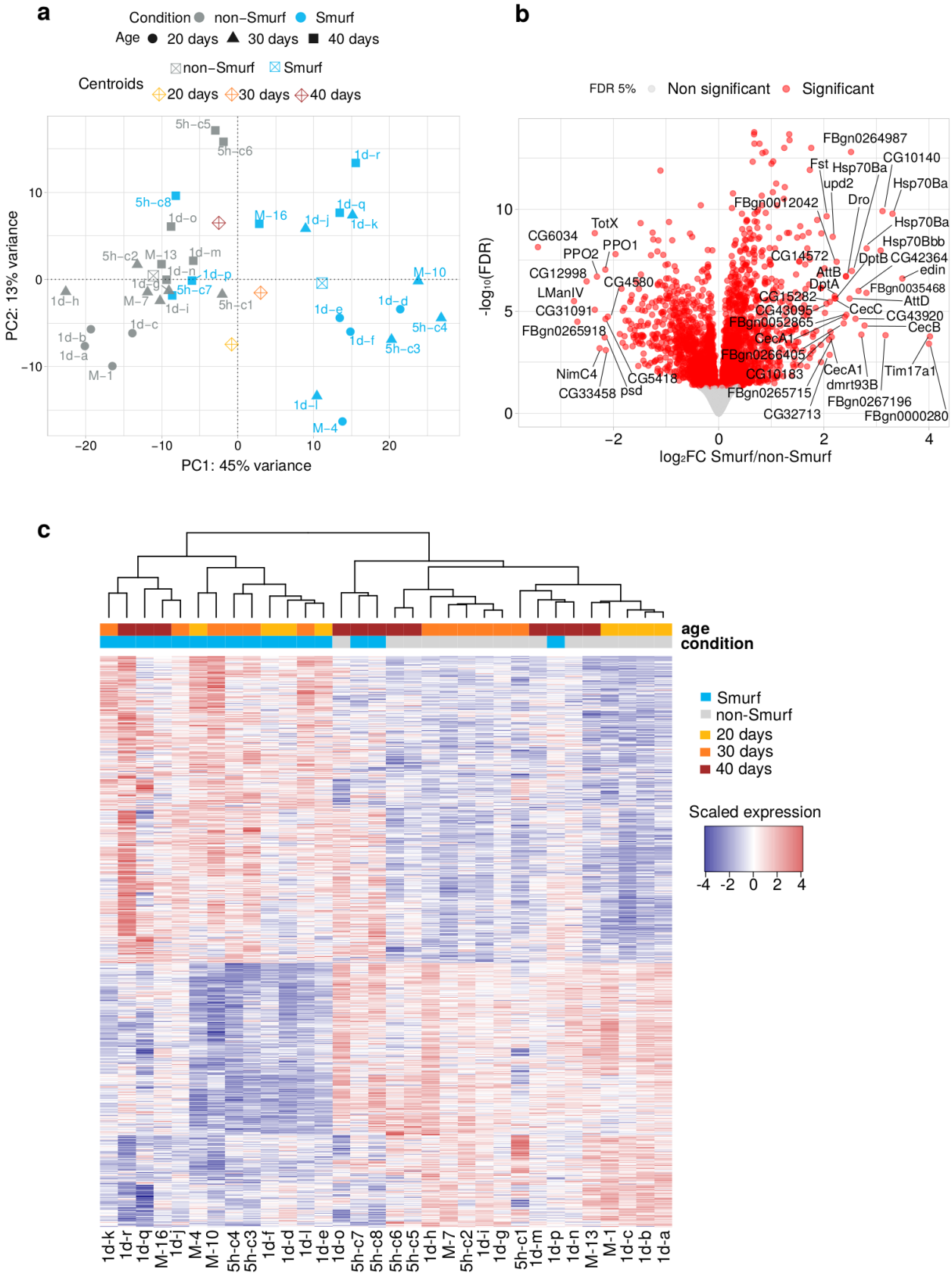

a
Condition non-Smurf Smurf
Age 20 days 30 days 40 days
Centroids non-Smurf Smurf
20 days 30 days 40 days
PC2: 13% variance
PC1: 45% variance
b
FDR 5% Non significant Significant
-log10(FDR)
log2FC Smurf/non-Smurf
FBgn0264987
CG10140
Fst
Hsp70Ba
upd2
Hsp70Ba
Dro
FBgn0012042
Hsp70Ba
Hsp70Bbb
TotX
CG6034
PPO1
PPO2
CG14572
DptB
CG42364
edin
AttB
DptA
CG12998
CG4580
FBgn0035468
LManIV
CG15282
AttD
CG31091
CecC
CG43920
CG43095
FBgn0052865
CecB
FBgn0265918
CecA1
FBgn0266405
NimC4
CG5418
CG10183
CecA1
Tim17a1
CG33458
psd
FBgn0265715
dmrt93B
FBgn0267196
CG32713
FBgn0000280
c
age
condition
Smurf
non-Smurf
20 days
30 days
40 days
Scaled expression
-4 -2 0 2 4
1d-k 1d-r 1d-q M-16 1d-j M-4 M-10 5h-c4 5h-c3 1d-f 1d-d 1d-l 1d-e 1d-o 5h-c7 5h-c8 5h-c6 5h-c5 1d-h M-7 5h-c2 1d-i 1d-g 5h-c1 1d-m 1d-p 1d-n M-13 M-1 1d-c 1d-b 1d-a

**Figure 2. Smurfness is associated with a characteristic transcriptome. a) Samples plotted in the space of the first two components after PCA analysis.** PCA performed on the 1000 top-variance genes of the dataset results in a striking separation of Smurf and non-Smurf samples, independently of the age. On the other hand, samples distribute according to age along the second component, with the presence of three Smurf "outliers". The PCA results suggest that Smurfness is the biological variable in our dataset with the biggest effect on the transcriptome (45% of variance explained by PC1), followed by age (13% of variance explained by PC2). Smurf samples are in blue, non-Smurf in grey. Shapes indicate the age as illustrated in the legend. Centroids coordinates for a specific group are the mean of the group coordinates. Each sample is associated with an acronym specifying the time after the transition (5h = 5 hours, 1d = 1 day and M = mixed - unknown time -) and a unique letter or number identifying the sample itself. b) **Volcano plot of the DEG analysis results.** The negative logarithm with base 10 of the adjusted p-value (FDR) is plotted as a function of the shrinked (DESeq2 apeglm method[29]) fold change (logarithm with base 2) of the Smurf/non-Smurf expression ratio for each given gene. The significant 3009 DEGs are represented in red. Upregulated Smurf genes (1618) plot on the right side of the graph, while downregulated genes (1391) on the left. Genes with a $\log_2$FC > |2| are labelled (gene symbols when available, annotation Flybase ID otherwise). Amongst the genes annotated as upregulated we can notice the presence of immune response genes (*Dro*, *AttB*, *AttC*, *DptA*, *DptB*, *CecA1*, *CecB*, *CecC*), confirming what already described in Smurfs[18]. c) **Smurf DEGs represent a Smurf specific signature.** Unsupervised hierarchical clustering on the samples by Smurf DEGs only almost perfectly divides Smurfs from non-Smurfs independently of their age, demonstrating that those genes are a Smurf specific signature. Non-Smurf samples tend to cluster by age, suggesting an age trend in the expression of Smurf DEGs in non-Smurf . The same three outliers of **(a)** are identified, indicating that those three samples indeed present a weaker expression pattern compared to the other Smurfs. Expression of genes in the heatmap is re-centered on the mean across samples, for easy visualisation of upregulated and downregulated genes (colour code in legend).

# Smurf transcriptome recapitulates the transcriptional signature of ageing

We used biological processes (BP) Gene ontology (GO)[30] as gene sets in Gene Set Enrichment Analysis (GSEA)[31] to characterise the Smurf signature. In order to fully examine the observed signal, we chose not to apply any filtering on the $\log_2$FC. Results are presented in Fig. 3 and Table S1. Genes upregulated in Smurfs are enriched in immune and stress response, as previously described[18]. The immune response is widely upregulated, with activation of both Toll (fungi and Gram-positive response)[32] and Immune deficiency (Imd, Gram-negative response)[33,34] pathways. Antimicrobial peptides (AMPs), which are inflammatory indicators in flies, are strongly upregulated (*CecA1*, *CecA2*, *CecB*, *CecC*, *DptA*, *Def*, *Dpt*, *Drs*, average $\log_2$FC = 2.33). Upstream of the AMPs transcription, the two transcription factors *Rel* (Imd pathway, $\log_2$FC = 0.61) and *dl* (Imd pathway, $\log_2$FC = 0.27), are also upregulated. Increased inflammation is listed amongst the hallmarks of ageing[1,25], and various transcriptomic studies in Drosophila[35–40] and other organisms[41–46] (including humans[47]) have demonstrated its increase with age. Chronic inflammation has previously been proposed as one of the drivers of the ageing process[48–50]. Many categories related to protein folding and unfolded protein response (UPR) are over represented in our dataset. Smurfs present a significant induction of 22% amongst the Drosophila chaperons and co-chaperons (Flybase[51] annotation, version FB2022_04), with a broad upregulation of the Hsp70 family (6 out of 7 genes detected are upregulated, average $\log_2$FC = 2.60). *Hsp70* genes have already been described as upregulated during ageing in Drosophila[38,52]. Upregulation of *Hsps* is also part of the stress response listed as an ageing hallmark[1,25].

A low but significant upregulation ($\log_2$FC = 0.21) of *Xbp1* transcription factor, involved in UPR in the endoplasmic reticulum (ER)[53], is also detected. *Ire1*, coding for the transmembrane protein responsible for *Xbp1* unconventional splicing under ER stress[53], is also slightly but significantly upregulated ($\log_2$FC 0.21) in Smurfs. Finally, we detect a significant upregulation of 51% of the

annotated cytosolic Glutathione S-transferases (Gst). The expression of this family of genes, involved in detoxification, have been shown to increase in rats with age[54]. A decrease of their enzymatic activity was also observed in humans with ageing[55]. Alterations in the expression of some Gsts result in pro-longevity effects in different organisms[56,57].

Downregulated genes show a broad enrichment in metabolism-related categories. Genes involved in fatty acid biosynthesis, such as *FASN1* ($log_2FC$ = -0.61) , *ACC* ($log_2FC$ = -0.31) and *eloF* ($log_2FC$ = -0.41) are significantly downregulated in Smurfs. These results provide a clue regarding the pathways involved in the decreased triglycerides content previously described in Smurfs[18]. Genes of the mitochondrial electron transport chain (ETC), also show a broad downregulation. In our dataset, 123 genes belonging to the categories deregulated by GSEA are detected (based on Flybase annotation). 38% of Complex I subunits are downregulated (average $log_2FC$ = -0.18), 33% of Complex II (average $log_2FC$ = -0.17 ), 29% of Complex III (average $log_2FC$ = -0.21), 19% of Complex IV (average $log_2FC$ = -0.18), 41% of Complex V (average $log_2FC$ = -0.19). We see *ND-20* (Complex I), *ND-SGDH* (Complex I), and *UQCR-14L* (Complex III) downregulation, all of which have been shown to positively modulate longevity when knocked-down[58]. Despite the minor fold changes, the ETC components' persistent downregulation may indicate that the aerobic metabolism they mediate is also downregulated. In addition, the upregulation of lactate dehydrogenase gene (*Ldh*) - $log_2FC$ 0.95 - could suggest a compensatory anaerobic metabolism replacing a probable dysfunction of the aerobic ETC path, or an altered pyruvate intake into the mitochondria. Adding evidence to this trend, *Idh3A*, *Idh3B*, *Mdh1*, *Mdh2* and *Fum1*, involved in the tricarboxylic acid (TCA) cycle are downregulated, with fold changes similar to the ones reported above.

Mitochondrial dysfunction, as well as downregulation of mitochondrial genes, have been described as hallmarks of ageing[1,25]. Downregulation of mitochondrial metabolic genes with ageing has been previously documented in Drosophila[59,35,38,60,61,36,62,63], worms[41,60,64] and mammals[43,46,61,65]. In addition, upregulation with ageing of *Ldh* is documented in flies head[66]; life-long overexpression of this genes in neurons negatively affects lifespan[66], suggesting that the overexpression late in life of *Ldh* is detrimental, or is a coping mechanism to a decreased mitochondrial function.

Genes involved in ecdysone biosynthesis, such as *sad*, *spo* and *phm* are also downregulated in Smurf flies. In adult females ecdysone is mostly produced in the ovaries[67]. However, its activity in adult flies has been reported on a broad range of tissues[68], not allowing for a specific interpretation of the signal on whole-body data. Recent findings show a role of ecdysone signalling in post-mating gut reshaping through intestinal stem cells[69], and further studies could be conducted to investigate the possible role of the hormone in ageing-driven gut changes in the flies. Genes involved in egg formation, as *Vm26Aa*, *Vm26Ab*, *Vml* and *psd* are downregulated ($log_2FC$ is respectively -2.67, -2.63, -2.51, -2.49). This result thus gives a molecular hint for explaining the previously reported decrease in fertility in Smurf females and males[23], and overlaps with a gene expression trend already described with ageing in Drosophila[35,40].

A few categories related to proteostasis are also present amongst the ones deregulated in Smurfs. The ribosome biogenesis category (GO:0042254) presents 23% of significant downregulation amongst the detected genes (44 downregulated genes out of 190 detected). As for the ETC, the genes' $log_2FCs$ are low (average ~ -0.14), but as they come in a unidirectional

deregulation, it could still be biologically relevant. We could not detect major changes in ribosomal particles. Downregulation of ribosomal-related genes is listed as one of the hallmarks of transcriptional ageing[25]. Regarding protein degradation, we detected the downregulation of 10 trypsin-like endopeptidases and 14 Jonah genes (serine endopeptidases family). Despite the fact that little is known about their function, *Jonah* genes have been shown to be expressed in larval and adult Drosophila gut[70]. No significant deregulation is nevertheless detected for proteasome subunits.

Finally, it is worth mentioning the deregulation of a few extracellular matrix (ECM) related genes. We can detect a downregulation of laminin (*LanB1* - log2FC -0.31 - and *LanB2l* - log2FC -0.35 -), a structural component of the nucleus whose mutations cause accelerated ageing in humans[71], and collagen (*Col4a1* - log2FC -0.89 - and *vkg* - log2FC -0.87 -). Although we cannot know whether the detected signal is specific to certain tissues, it has been shown that *Col4a1* mutants present premature loss of intestinal integrity and increased inflammation markers in the gut[72]. Interestingly, the two ECM reshaping metalloproteinases *Mmp1* and *Mmp2* are upregulated in Smurfs, indicating a possible remodelling of the ECM in this last phase of life.

Altogether, the results of our analysis point to a systemic state of stress and physiological impairment in Smurf individuals. As described above, the Smurf signal overlaps with numerous changes that were described so far as being ageing-related. Although this might not be surprising, since Smurfs are individuals about to die, it is important to highlight that the non-Smurfs, independently of their chronological age, do not carry such changes (Fig. 2c). Those changes appear therefore to be Smurf-specific. In order to check how the Smurf signal overlaps with what has been described so far as the “ageing transcriptional signature”, we mapped the hallmarks of transcriptional ageing described in[25] on the GSEA network (Fig. 3). As the authors do not provide a specific list of genes to compare with, but rather a description of the affected processes, we mapped the hallmarks on our signature when the biological categories described show an overlap.

As mentioned above, the dysregulation of immune genes (Ageing Transcriptional Hallmark 1 - ATH1) that has been reported with ageing in Drosophila and other organisms is here a feature of Smurfness. The mitochondria-related marker (ATH2) follows a similar behaviour. Regarding the protein synthesis marker (ATH3), grouping ribosomal proteins and ribosome biogenesis factors, we can only detect the latter in Smurfs. However, ribosomal genes are highly expressed, and changes in their absolute mRNAs levels might lead to small fold changes, missing significant threshold in RNA-Seq analysis. Nevertheless, the ribosomal proteins are downregulated in our proteomic analysis of Smurf flies, supporting our findings from the Smurf transcriptome. The stress response marker (ATH4), which is intertwined with the upregulation of inflammatory genes, maps to the *Hsp* signal from Smurfs. The DNA damage response (ATH4) is indicated with a question mark in Fig. 3 following what Frenk and colleagues reported, as conflicting data are present, preventing the univocal classification of this process. Indeed, cells appear to display reduced ability to cope with DNA damage with ageing and senescence instead[73,74]. The reduction in growth factor signalling (ATH5) marker refers to reduced cell growth and proliferation during ageing, mostly associated with downregulation in cell cycle and DNA replication genes. Such a signal is not detected in our analysis. The same applies to the mRNA pre-processing category (ATH6). The dysregulation in gene expression regards an

increased transcriptional heterogeneity with ageing (ATH6), a biological question that cannot be addressed through DEG analysis and that will be further discussed later in the section.
Overall, we described a Smurf-specific transcriptome, mostly independent of the samples' ages, that predicts four out of six hallmarks of transcriptional ageing (ATH1-4).
Deregulation at a transcriptional level might not necessarily translate to protein deregulation (protein concentration or activity). Therefore, to investigate if the Smurf were indeed experiencing biological alterations in the pathways detected through DEGs and GSEA analysis, we compared our results with proteomic and metabolomic data obtained from Smurf and non-Smurfs collected in the laboratory, using females from the same *Drs*-GFP line. Proteomic quantification was conducted at 90% and 10% survival, correspondingly to the RNA-Seq experimental design for young and old flies. Metabolomic analysis was instead performed on DrsGFP at 50% survival, corresponding to the intermediate time point of RNA-Seq sampling. More information about sampling and processing are reported in the Materials & Methods section. Enrichment analysis on significantly differentially represented proteins (ANOVA p-value $< 0.05$, for complete results see Supplementary File 2) confirms our findings at a transcriptional level, detecting downregulation of fatty acid catabolism, mitochondrial respiration and ribosomal proteins (Fig. S6). Response to stress (including genes such as *cact, Hsp70* and *Cat*) is detected amongst the upregulated processes, in line with what described in our transcriptome study. On the other hand, upregulated GO BP categories display a signal coming from IMP (Inosine monophosphate nucleotide, intermediate step in purine biosynthesis), which has not been detected at a transcriptomic level.
Quantitative enrichment analysis on metabolites concentrations in Smurfs and non-Smurfs (Supplementary file 3) confirms the molecular separation between the two phases (Fig. S7) and the metabolic transcriptional signature observed. We detected deregulation of fatty acid biosynthesis and degradation pathways (KEGG[75]), as well as pyruvate metabolism (which includes metabolites from the TCA cycle) (Table S2). We mapped the Smurf-relative metabolites quantification and the differentially expressed genes on the KEGG pathways retrieved by the metabolomic enrichment analysis, confirming the downregulation of fatty acid biosynthesis at the transcriptional level by a decrease of the final fatty acid products in Smurfs (palmitic acid - $\log_2FC = -1.37$ - and myristic acid - $\log_2FC = -1.69$ - , Fig. S8). Regarding glucose metabolism, the overexpression of *Ldh* is confirmed by a significant (Wilcoxon test, p-value $< 0.05$) lactic acid increase in Smurfs ($\log_2FC = 0.90$) (Fig. S9). The TCA cycle displays a significant general decrease at a transcriptomic level, and a general impairment at a metabolomic level (though the only metabolite significant to Wilcoxon test is succinate, $\log_2FC = 1.28$) (Fig. S10).
These results suggest that the deregulation observed in the gene expression of Smurfs is not only transcriptional, and is probably instead part of major changes affecting the individual at different levels, with a functional impact.

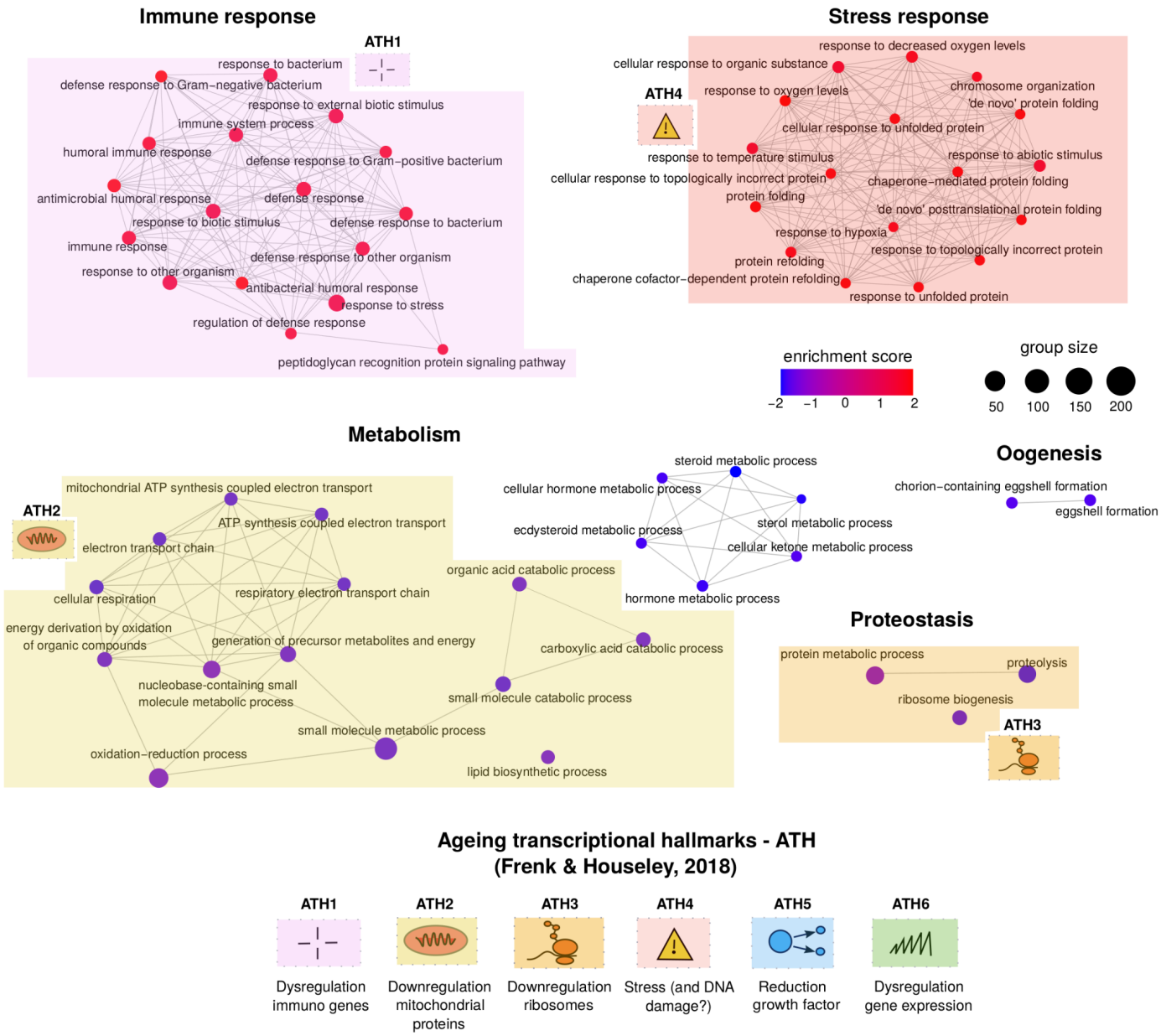



**Figure 3. GSEA analysis (GO biological process categories) on Smurf specific genes.** GSEA results are represented as a network, where nodes are significantly enriched categories (deregulation colour code as in legend) and edges are connected categories with overlapping genes. From the 59 significant categories, we identified and manually annotated five "hubs": immune response, stress response, metabolism, proteostasis and oogenesis. Hallmarks of transcriptional ageing, as enunciated in Frenk & Houseley[25], are represented at the bottom of the figure. The hallmarks present in the Smurf specific signature (ATH1-4) are mapped close to the related categories. Overall, in the Smurfs specific genes we detect four hallmarks of transcriptional ageing.

# Removing the Smurf-specific signature unveils the transcriptional effects of chronological age

In the previous section, we showed how the Smurf status correlates with ageing-associated transcriptional changes. In order to confirm the Smurf-specificity of the detected signature, we removed Smurf samples from the study and compared the non-Smurfs using chronological age as the only ageing marker. In contrast to the 3009 DEGs found in the Smurf/non-Smurf comparison, 526 DEGs were found when comparing the 40-day non-Smurf (10% survival of the population) to the 20-day non-Smurf (90% survival) (DESeq2 results in Supplementary File 3). It is interesting to note that 59% of these genes are Smurf DEGs. 22 GO BP deregulated categories were found by GSEA analysis using the same settings as those previously utilised for Smurfs (Fig. 4a and Table S3). Older non-Smurfs exhibit a heightened immunological response (ATH1) and reduced transcription of genes important for oogenesis. The transcriptional decrease of oogenesis gene might imply a decrease in fertility, which has also been observed in ageing[35,40] and in Smurfs[23]. We also compared young non-Smurfs to mid-life non-Smurfs (50% survival), detecting only 57 DEGs. As already suggested by the PCA, no major differences are present between young and mid-life non-Smurfs.

In order to further investigate the presence of ageing transcriptional markers in old non-Smurfs, we manually mapped the genes differentially expressed in Smurfs and in old non-Smurfs to processes that are described as being associated with ageing[25]. We either used the Flybase annotation or the GO BP categories for the mapping (depending on the most complete annotation for the specific process). We added the insulin-like receptor signalling pathway (IIS, Flybase annotation), as its modulation is well described as positively affecting ageing in multiple organisms[76–79].

The genes that are known as being downregulated with ageing, are actually down regulated mostly in Smurfs (Fig 4b, point i), with little to no effect associated with chronological age. In the case of some stress response genes (chaperones, GSTs and inflammatory genes), the non-Smurfs detect more genes than in the categories mentioned above (Fig. 4b, point ii). However, they overlap with the Smurf DEGs, suggesting that the non-Smurfs samples do not add information on these pathways. More specifically, the biggest overlap is observed for the immune response pathways (as expected from the GSEA results). Out of the overlapping genes (20), 50% are AMPs, produced downstream the pathway (average $\log_2$FC ~ 2.91, while for the same genes the average Smurf $\log_2$FC is ~ 2.35). We do not find significant deregulation of the *dl* transcription factor (Smurf significant $\log_2$FC = 0.27), while *rel* is upregulated ($\log_2$FC = 0.42, while for the Smurf we detected a $\log_2$FC of 0.61). These results suggest that the immune response is active in the old non-Smurf, even if to a lower extent than the Smurfs for the number of deregulated genes.

Regarding the genes mapping to the insulin-like receptor signalling (IIS) pathway (Fig 4b, point iii), we do not detect any deregulation in the non-Smurfs, with IIS core components being affected only in Smurfs. More specifically, while no significant change is detected for the *Ilp* genes (insulin-like peptides activating the pathway), we find low but significant upregulation of the downstream genes *Inr* (receptor, $\log_2$FC = 0.42), *chico* (first kinase of the cascade, $\log_2$FC 0.23) and the kinase *Akt1* ($\log_2$FC = 0.18). *Inr* and *chico* are well-described longevity genes in Drosophila, positively affecting ageing when negatively modulated[77,78]. No significant changes

are detected for the Drosophila mTOR genes *Tor* and *raptor,* nor *foxo*. However, we find significant upregulation of *Thor*, coding for the homologous mammalian translation initiation factor 4E-BP. *Thor* is a *foxo* target, and its upregulation (together with the one of *InR*, another foxo target) confirms what was already described at the protein level in Smurfs[18], and might suggest a possible increased activity of *foxo* due to IIS repression, as occurs with ageing.

Unfortunately, no annotated database of ageing hallmarks-associated genes exists for *Drosophila*. However, the Ageing Atlas[80] contains annotations for 500 human (and mouse) genes as being ageing-associated markers. In order to validate our mapping, we retrieved the human orthologs of the DEGs for both Smurf and non-Smurf and searched our lists with the annotated one (Table S4 and S5).

All the genes annotated are present in our dataset. 26.8% of the annotated genes are present in the Smurf list (134 human genes out of 500, corresponding to 121 Drosophila genes), while only 4% is present in the old non-Smurf (25 human genes out of 500, corresponding to 24 Drosophila genes) (Fig. 4c). 22 out of the 25 detected in the old non-Smurf are overlapping with the Smurf list. Interestingly, the mitochondrial dysfunction related hallmark is detected only in the Smurf list, confirming its specificity to the Smurf state.

Over the past 40 years, numerous genes have been shown as modulating ageing when artificially deregulated - i.e. longevity genes. Given the fact that the Smurf transcriptome recapitulates most of the ageing transcriptional signature independently of the age, while the non-Smurf one does not, we were interested in exploring how our list of DEGs were overlapping with the described “longevity genes”. We extracted the Drosophila longevity genes annotated in GenAge[81] and intersect them with the Smurf and non-Smurf DEGs. Out of the 201 annotated genes, 188 are present in our dataset. The 13 not found genes are distributed the following way: four mRNA not detected in our dataset (*CG11165*, *PolG2*, *Hsp22*, *Prx5*), five miRNA (let-7, mir-125, mir-14, mir-184, mir277, mir-34) and two human gene variants artificially expressed in Drosophila (ΔOTC and UCP2). Smurfs DEGs allow the detection of 37% of longevity annotated genes, while the old non-Smurf DEGs detect only 6% (Fig. 4d and Tables S6 and S7). Furthermore, all the longevity genes present in the non-Smurf DEGs are also present in the Smurf DEGs. These results suggest that the previously identified longevity genes might target processes involved in Smurf transition, possibly extending lifespan by delaying the entrance in such phase. These results confirm that Smurfness is better at describing so-called ageing-associated changes than chronological age.

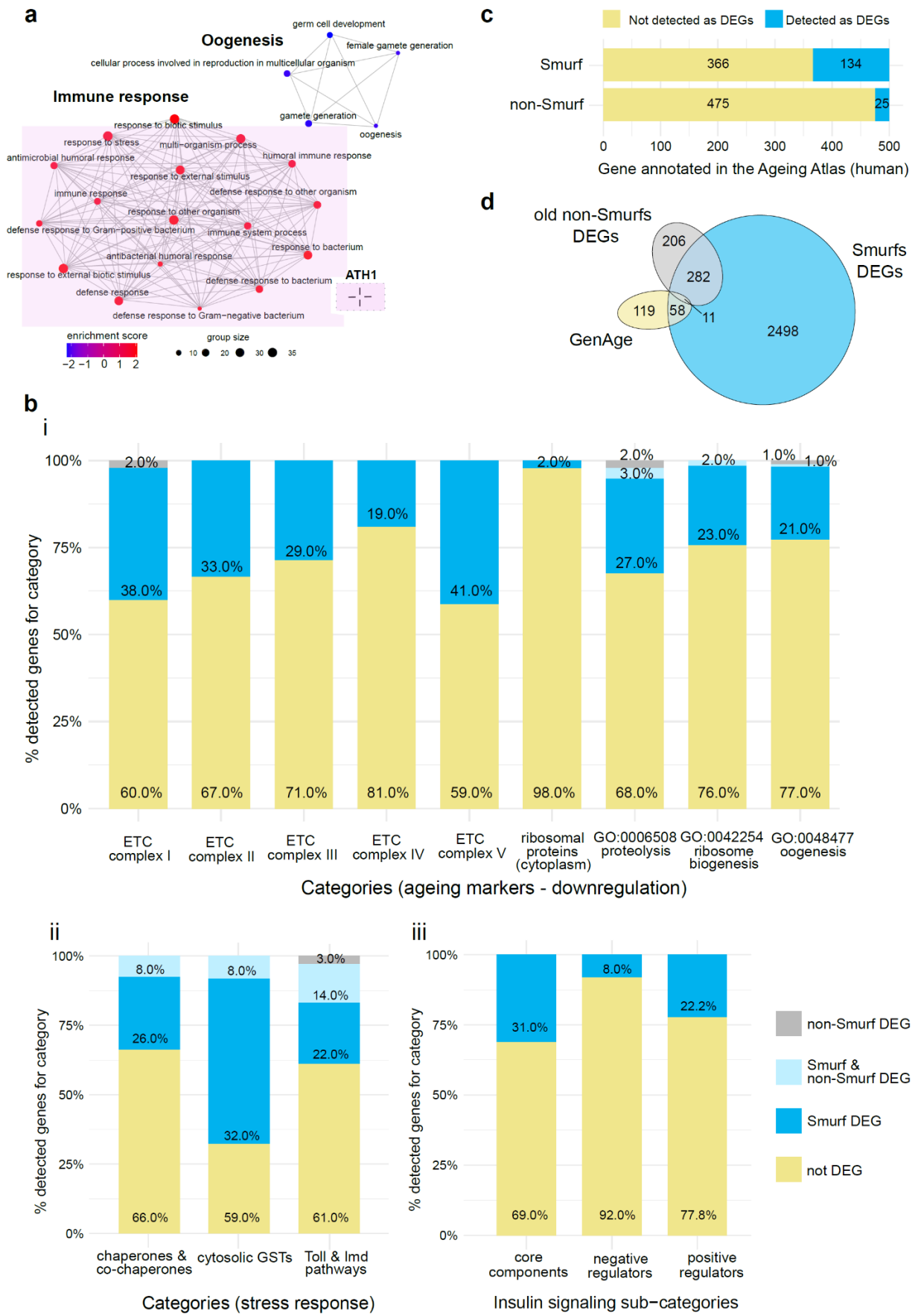

a
Oogenesis
germ cell development
female gamete generation
cellular process involved in reproduction in multicellular organism
gamete generation
oogenesis
Immune response
response to biotic stimulus
response to stress
multi−organism process
antimicrobial humoral response
humoral immune response
response to external stimulus
immune response
defense response to other organism
response to other organism
defense response to Gram−positive bacterium
immune system process
response to bacterium
antibacterial humoral response
response to external biotic stimulus
defense response to bacterium
defense response
defense response to Gram−negative bacterium
ATH1
enrichment score
−2 −1 0 1 2
group size
10 20 30 35
c
Not detected as DEGs
Detected as DEGs
Smurf
366
134
non-Smurf
475
25
0 100 200 300 400 500
Gene annotated in the Ageing Atlas (human)
d
old non-Smurfs DEGs
Smurfs DEGs
GenAge
206
282
119
58
11
2498
b
i
% detected genes for category
100%
75%
50%
25%
0%
2.0%
38.0%
60.0%
33.0%
67.0%
29.0%
71.0%
19.0%
81.0%
41.0%
59.0%
2.0%
98.0%
2.0%
3.0%
27.0%
68.0%
2.0%
23.0%
76.0%
1.0%
1.0%
21.0%
77.0%
ETC complex I
ETC complex II
ETC complex III
ETC complex IV
ETC complex V
ribosomal proteins (cytoplasm)
GO:0006508 proteolysis
GO:0042254 ribosome biogenesis
GO:0048477 oogenesis
Categories (ageing markers - downregulation)
ii
% detected genes for category
8.0%
26.0%
66.0%
8.0%
32.0%
59.0%
3.0%
14.0%
22.0%
61.0%
chaperones & co-chaperones
cytosolic GSTs
Toll & Imd pathways
Categories (stress response)
iii
% detected genes for category
31.0%
69.0%
8.0%
92.0%
22.2%
77.8%
core components
negative regulators
positive regulators
Insulin signaling sub−categories
non-Smurf DEG
Smurf & non-Smurf DEG
Smurf DEG
not DEG

**Figure 4. Smurfness is a better predictor of transcriptional ageing markers than chronological age. a) GSEA analysis (GO BP categories) on old non-Smurf specific genes.** Results are represented as in Fig. 3 . GSEA analysis identifies 22 deregulated GO BP categories, related to immune response (upregulation, in red) and oogenesis (downregulation in blue). The analysis carried on chronological age can therefore detect only one hallmark of transcriptional ageing[25] (ATH1, for representation of transcriptional hallmarks, see Fig. 3). **b) Manual mapping of Smurf and old non-Smurf DEGs on ageing processes.** For each process, the histograms represent the percentage of genes mapping to it but not detected as DEGs in our analysis (yellow), detected as Smurf DEGs (blue), detected as both Smurf and non-Smurf DEGs (light blue), or only detected in the old non-Smurf DEGs (grey). When not stated otherwise, the gene lists are retrieved from Flybase. Genes downregulated with ageing **(i)** are mostly detected only in Smurfs, with the exception of structural ribosomal proteins, whose downregulation is not significant in Smurfs. For the processes upregulated with ageing **(ii)**, the Smurf samples do retrieve more information than the non-Smurfs, with the last however carrying more signal than in the case of the downregulated genes, especially for the immune response (as already hinted by **(a)**). Similarly, the IIS pathway displays deregulation in the Smurfs, while no gene is detected as deregulated when looking only at chronological age **(iii)**. **c) Mapping of Smurf and non-Smurf DEGs to human ageing-related genes (annotated in the Ageing Atlas).** The Ageing Atlas annotates 500 human ageing-related genes. All of those have orthologs in Drosophila and are present in our dataset. By studying the Smurf phenotype, we can detect 134 genes out of the annotated 500. The number of detected genes drops to 25 when using chronological age only as ageing marker. **d) Longevity genes and Smurfness.** The venn diagram shows the overlap between the annotated longevity genes in drosophila (GenAge), the Smurf DEGs and the non-Smurf DEGs. While Smurf-centered analysis retrieves ~37% of the longevity genes, the non-Smurf centered analysis only retrieves ~6%, not adding information to what was already detected by the Smurf analysis.

# Old Smurfs carry additional age-related changes

In order to explore a possible effect of age on the Smurfs, we compared the Smurf sample along chronological age through DEG analysis, as previously done with the non-Smurfs. Similarly to what occurred in the case of the non-Smurfs, the 20 and 30 days sample appear very similar, with only 4 DEGs detected (FDR cut-off at 5%). However, in contrast to the poorer signal (526 against 3009 of the Smurf signature) detected in the old non-Smurfs, the 40 days Smurfs present 2320 DEGs when compared to the 20 days Smurf (1385 upregulated and 935 downregulated) (DESeq2 results in Supplementary File 5). GSEA identified 125 deregulated GO BP categories. Interestingly, 115 are downregulated and only 10 are upregulated. This suggests that the downregulated genes, even if in minority, present a coherent deregulation, while the upregulated genes might affect genes more randomly. Results are presented in Fig. 5 and Table S8. The majority of the detected categories are associated with RNA processing (such as splicing factors), transcription, chromatin organisation, DNA replication and repair. The dysregulation of RNA processing and splicing has been reported as being associated with ageing and age-related diseases[82,25,83,84], with decreased expression of splicing factors observed also in healthy human blood samples with age[85]. The chromatin organisation-related categories are also of interest, alteration in epigenetic-mediated gene regulation has been described with ageing[1,86–88]. In the case of old Smurfs, we find downregulation of genes involved in histone methylation (*trr, Cfp1, Dpy-30L1, Smyd5, NSD, CoRest, Lpt*, average $\log_2$FC ~ -0.26), amongst which genes of the Polycomb Repressive Complex 2 (*esc, E(z), Su(z)12l,* average $\log_2$FC ~ -0.24). We also detect the downregulation of the histone deacetylase *HDAC1* ($\log_2$FC = -0.18) and genes involved in histone acetylation (as *CG12316, Ing3, Ing5, Taf1, Atac3, Brd8, Spt20, mof,* average $\log_2$FC ~ -0.30). Even though the downregulation is weak ($0 < |\log_2 FC| < 1$ in all cases), the signal shows that chromatin-related gene regulation is broadly affected in old Smurfs. This would require further confirmation with appropriate epigenomic studies, meanwhile our proteome analysis shows a significant decrease of H3.3B ($\log_2$FC = -0.43) and H4 ($\log_2$FC = -0.54) in Smurfs. In particular, those results are interesting given the histone regulation

alteration documented with ageing[89,90], that gave rise to the ageing theory of “loss of heterochromatin”[91]. Deregulation in gene expression through chromatin alteration is listed amongst the “dysregulation of gene expression” ageing transcriptional hallmark (ATH6). Another interesting signal is the DNA repair nodes (GO:0006302 double-strand break repair, GO:0006281 DNA repair), where we retrieve 12% of the detected genes as significantly downregulated (average $log_2FC$ = -0.24). As previously mentioned, loss in physiological capacity to repair DNA has been reported with ageing, is described as one of the hallmarks of ageing[1], and is proposed as one of its leading causes[92]. We also retrieved nodes associated with downregulation of genes involved in cell cycle (as cyclins), or their regulators (as *E2f2*, $log_2FC$ ~ -0.17). Genes involved in spindle organisation during mitosis are also found downregulated (as *Mtor* - $log_2FC$ ~ -0.28- and *Chro* - $log_2FC$ ~ -0.19- ) suggesting a broad dysregulation of cell proliferation processes.

Overall, the old Smurf signature carries therefore the two hallmarks of transcriptional ageing that we did not detect in the Smurf specific signature, the dysregulation of RNA processing (ATH6) and the reduction in growth factor signals (decreased expression of cell cycle genes, ATH5). In addition, they present decreased expression of DNA repair genes. It is important to mention that we do not find Smurf-related categories in the GSEA output, confirming that young Smurf and old Smurfs indeed do carry the same Smurf signature illustrated in Fig. 3. However, analysis shows that the old Smurfs carry additional transcriptional changes, which mostly relate to transcription and DNA regulation, and might be specific to the final stage of life of old individuals.

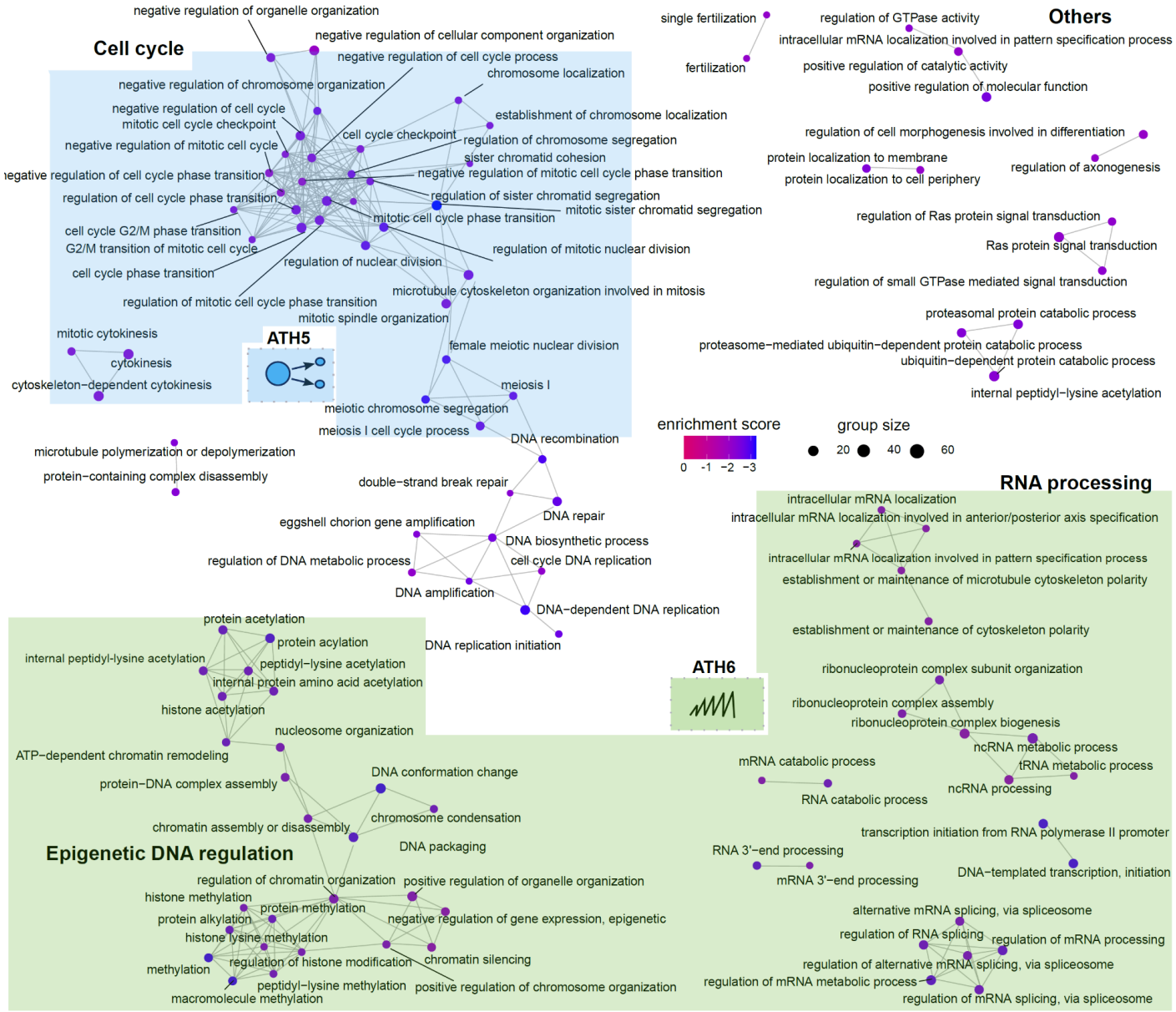



**Figure 5. Old Smurfs carry ageing-related signal amongst downregulated genes.** Results of the GSEA analysis are represented as in Fig 3. Only downregulated nodes presenting at least one interconnection are represented here. Complete list of deregulated categories can be found in Supplementary Table S8. GSEA analysis identifies 115 downregulated GO BP categories, which are mostly related to DNA regulation, RNA processing and cell cycle regulation. A few nodes are associated with DNA repair. Interestingly, the signal carried by the old Smurfs maps (at least partially) to the "dysregulation in gene expression" (in green, ATH6) and the "reduction in growth factors" (ATH5) transcriptional ageing markers that were not detected in the Smurf specific signature. In addition, the DNA damage nodes show downregulation of genes involved in DNA repair, which has also been discussed as an ageing marker. Interestingly, there are no hubs in the network overlapping with the Smurf specific signature of Fig.3, proving that the core Smurf signal is not affected by chronological age. However, the old Smurfs do carry an additional signature compared to their younger counterparts, hinting for the existence of a "chronological-age burden" that might enhance the probability of entering the Smurf pre-death phase, without however being necessary or sufficient for it.

## Chronological age dependent signature

In the previous paragraph we have shown how most of the transcriptional alterations described as age-related are actually restricted to Smurfness, with only a small part of the signal retrieved in old non-Smurfs when compared to young ones (Fig. 4). We hypothesised that additional light (but relevant) age-related non-Smurfs changes might be present and missed by the DESeq2 approach. We therefore decided to further investigate if some genes show an increasing (or decreasing) trend with chronological age in the non-Smurfs. As the evolution with time of the non-Smurf subpopulation has been so far successfully approximated by a linear model[18,21], we fitted the following model for each genes over the three time points available in our dataset:

$$g = \beta_0 + \beta_1 t + \varepsilon$$

where $g$ is the expression of a gene, $t$ the time (20, 30, 40 days) and $\varepsilon$ the noise - i.e. the proportion of $g$ that cannot be explained by time. If time does not affect the expression of the given gene, we expect the coefficient $\beta_1$ to not be significant. We identified 1012 genes whose $\beta_1$ coefficient was significant to the F-test (out of the 15364 tested). In order to select the genes for which time explains most of the observed trend, we further selected the ones presenting an $R^2$ greater than 0.5, obtaining 301 genes (207 increasing their expression with time, 94 decreasing) (Table S9). Interestingly, the unbalanced trend confirmed what was seen with differential expression, where we detected 79% of upregulated genes amongst the significant ones (415 out of 526). Such imbalance is not present in the Smurf DEGS, where only 54% of the genes are upregulated (1618 out of 3009). In other words, while in Smurfs we detect a balance between significantly upregulated and downregulated genes, chronological age in non-Smurf appears to have a larger effect on gene upregulation. Around half of the genes detected (51.6 %) also belong to the Smurf DEGs, which confirms what was observed before, i.e. the effect of age in the non-Smurf at least partially overlaps with the effect of Smurfness. As we had already investigated the Smurf specific process, we filtered out those genes and focused on the 146 remaining (93 with positive slope, 53 negative). Interestingly, none of those genes are annotated as longevity genes (GenAge database) and no enrichment in GO categories was found (GOrilla enrichment[93]). In order to focus our attention on genes showing the strongest trends, we focused on the first and fourth quartiles of the slope values distribution. Results are represented in Fig. 6a. Amongst the genes showing positive trend we have three genes involved in immune response (*PGRP-SA, CG1572, Tep2*) and 5 genes predicted to have peptidase activity (*CG11951*, *CG30091*, *CG30082*, *CG30197*, *l(2)k05911*). Amongst the genes presenting a negative trend, there are three muscle-specific genes (*Scp1*, *TpnC41C*, *TpnC25D*) and two genes part of the RNA polymerase II complex (*Rpb12* and *l(2)37Cg*). Overall, it is interesting to mention that ~19% of the selected genes are reported in Flybase as expressed in adult heads or neurons, while the other tissues do not show a particular enrichment. There are also genes reported as being expressed in testis or involved in spermatogenesis (such as *CG13946* and *aly*). Their detection in our female samples raises the question of their possible role outside of the male gonad. Overall, our analysis shows that age-associated changes occurring in non-Smurf partially recapitulate those observed in Smurfs, with a weaker signal. The results suggest that the old non-Smurf signal is mostly a pre-Smurf signal, detected as the 40 days

non-Smurfs are sampled at 10% survival. At that time-point, most of the flies in our population are close to death - i.e. about to become Smurf. Our hypothesis is supported by the fact that the number of detected DEGs between age-matched Smurf and non-Smurfs decreases with chronological age (Fig. S11).

Ageing has been reported as increasing the gene expression heterogeneity in a variety of organisms, tissues and cell types[94–101] (ATH6). Nevertheless, such an effect is still debated, as some studies do not point in this direction[35,102,103]. We wanted to investigate whether this signal, which could not be addressed by the analysis conducted so far, was a Smurf or chronological-age marker. In that matter, we computed the relative standard deviation of all genes for each group (Smurfness and age). We then plotted the distributions of the relative standard deviation (RSD) across groups and compared them using Kolmogrov-Smirnov (KS) statistic (Fig. 6b). All genes are affected equally, independently of their expression levels (Figure S12). Age appears to have the largest influence on transcriptional noise. In both Smurfs and non-Smurfs the distribution at 20 and 40 days are significantly different to the KS statistic (Smurf $D_{20\text{-}40} = 0.27$, non-Smurf $D_{20\text{-}40} = 0.23$, where D refers to the maximal vertical distance between the cumulative distributions compared; p-value $< 10^{-16}$). The peak of the RSD distribution shifts towards the right with age (1.93-fold increase for the Smurfs, and 1.84-fold for the non-Smurfs), suggesting that gene expression increases in heterogeneity with chronological age. We also compared the age-matched distributions (Smurf/non-Smurf, $D_{20} = 0.06$, $D_{30} = 0.10$, $D_{40} = 0.05$ ; p-value $< 10^{-16}$ , p-value $< 10^{-16}$ , p-value $< 10^{-15}$ ). However, the estimated distance between age-matched Smur and non-Smurf distributions is lower than the one detected across ages within each group.

In conclusion, the analysis presented so far allows for a reinterpretation of the ageing transcriptional signature in the light of the Smurf classifier. Our results show that four out of six transcriptional ageing markers (ATH1-4) are specific to the Smurf phenotype, independently of their chronological age (Fig. 3). On the other hand, the alternation in chromatin-related genes and mRNA processing, as well as cell cycle genes (together with a weaker DNA repair signal) are carried by the old Smurfs (ATH 5-6) (Fig. 5). We could not identify biological processes strictly related to the old non-Smurfs compared to their young counterparts (Fig.4). However, the increased heterogeneity in gene expression (ATH6) shows a similar pattern for Smurfs and non-Smurfs (Fig. 6b), suggesting that it is a chronological age-related process. Therefore, the Smurf classification allows us to deconvolve and study the ageing process across two dimensions, Smurfness (a condition of high risk of death or advanced biological age) and chronological age.

To summarise and easily visualise how biological processes are affected by these two dimensions, we computed the correlation in the expression of each gene with the Smurf status (independently of the chronological age) and with the chronological age (independently of the Smurf status). Genes affected mostly by Smurfness should present a high correlation with the latter, while genes most affected by time should present a higher correlation with chronological age than Smurfness. Genes affected by both (as the old non-Smurf DEGs) should map to a zone of high correlation with both dimensions. We mapped all the Drosophila KEGG pathways into these two-dimensional maps, and tested for a significant difference in the KEGG pathway expression compared to the background using a two-dimension implementation of the KS statistic (Fasano-Franceschini test[104]). After extracting the 133 annotated KEGG pathways, we

selected those where at least 10 genes present in our dataset are mapped, for a final number of 112. We performed the Fasano-Franceschini test, and corrected for the p-value using the FDR method. We retrieved 86 significant pathways. We used average gene expression correlation as an indicative statistic to discern the pathways correlating with Smurfness from the ones correlating with chronological age. We finally obtain 48 correlating with Smurfness (Table S10) and 38 correlating with chronological age (Table S11). Fig. 6c groups pathways linked to the ageing signature previously described and their distribution. As expected, the Toll and Imd pathway mostly displays positive correlation with Smurfness; the ETC (oxidative phosphorylation pathway) and fatty acid degradation and elongation mostly negatively correlates with Smurfness, while showing a lower correlation with age. Interestingly, transcription-related pathways (such as spliceosome and basal transcription factors, in figure) and DNA amplification and repair pathways (DNA replication and Nucleotide excision repair, in figure) distribute in a zone of higher negative correlation to chronological age compared to Smurfness. It is worth noting that the correlation is here computed starting from all samples (Smurf and non-Smurfs); this can lead to speculation that those pathways, initially retrieved by DEG analysis as old Smurf specific (Fig. 5), might present a downregulation also in the old non-Smurfs, later exacerbated in the Smurfs. This downregulation might be however too weak to be detected by DESeq2 with the available statistical power (same reasoning applying to the linear regression, as many of the detected genes were Smurf DEGs). Finally, the proteasome and ribosome biogenesis show a similar correlation with chronological age and Smurfness, suggesting that they might be affected by both.

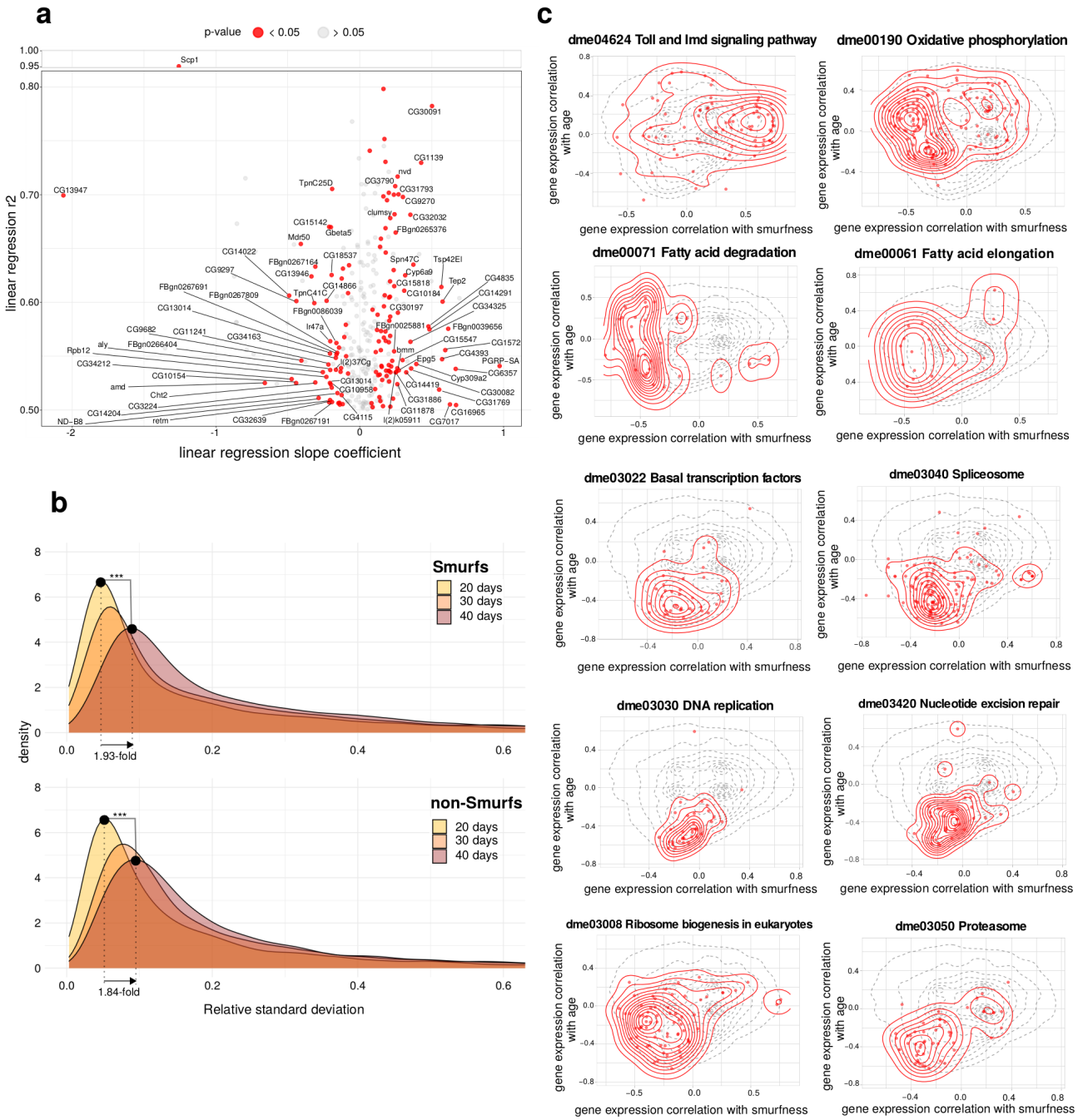



**Figure 6. Chronological age and Smurfness effects on the transcriptome. a) Linear regression of gene expression in non-Smurfs over time.** The $r^2$ of the applied linear model is plotted as a function of the slope coefficient. Only genes non differentially expressed in Smurfs are plotted, in order to focus on a possible specific non-Smurf signal. Genes presenting a significant slope are plotted in red. Genes belonging to the first and fourth quartile of the slope distribution are labelled with gene symbols (or Flybase ID if gene symbols could not be retrieved). **b) Chronological age effect on transcriptional heterogeneity.** The RSD densities are plotted for the different ages group (Smurf and non-Smurf). The tail of the distribution is cut at RSD = 0.6 for illustration purposes. Smurfs and non-Smurfs present a similar behaviour, with the peak of the distribution showing a almost 2-fold increase from 20 days to 40 days ($peak_{S20}$ = 0.046 , $peak_{S40}$ = 0.089, $peak_{NS20}$ = 0.051, $peak_{NS40}$ = 0.094, where peak refers to the RSD value associated to the peak of the distribution), showing the effect of chronological age on transcriptional noise. ***p-value < $10^{-16}$ (KS statistic). **c) Effect of Smurfness and chronological age on biological pathways.** Smurfness (marker of the organism's biological time) and chronological age (marker of the organism's chronological time) both affect the biology of the individual. By computing the correlation of gene expression with smurfness (independently of the chronological age) and with time (independently of the smurf status) and plotting the genes in the space of those two dimensions, we can visualise which biological processes are

affected by smurfness, by age, or by both. Here we show how some of the pathways above discussed are distributing over age and smurfness. Dotted line in the background corresponds to the density of all the genes analysed. Red points and density correspond to the genes mapping to the pathway (KEGG database) of interest. The significance of the difference in distribution among pathway and background was assessed using the Fasano-Franceschini test (FDR adjusted p-value). In order: Toll and Imd pathways ($r_{smurf}$ = 0.248 , $r_{age}$ = 0.080 , p-value = 5.2e-06), oxidative phosphorylation (ETC genes, $r_{smurf}$ = -0.217 , $r_{age}$ = 0.088, p-value = 4.5e-15), fatty acid degradation ($r_{smurf}$ = -0.388 , $r_{age}$ = -0.063, p-value = 4.3e-09) and fatty acid elongation ( $r_{smurf}$ = -0.255 , $r_{age}$ = -0.031, p-value = 3.8e-03) are mostly correlating with smurfness; spliceosome ($r_{smurf}$ = - 0.124 , $r_{age}$ = -0.288, p-value = 1.5e-17), basal transcription factors ($r_{smurf}$ = -0.096 , $r_{age}$ = -0.318 , p-value = 3.1e-08 ), DNA replication ($r_{smurf}$ = -0.070, $r_{age}$ = -0.393 , p-value = 2.2e-09) and repair (Nucleotide excision repair, $r_{smurf}$ = -0.073, $r_{age}$ = -0.338, p-value = 1.2e-10) are mostly correlating with age; Ribosome biogenesis ($r_{smurf}$ = -0.203, $r_{age}$ = -0.159, p-value = 4.0e-10) and proteasome ($r_{smurf}$ = -0.166, $r_{age}$ = -0.276 , p-value = 3.5e-09) apper to occupy a zone of similar correlation with both Smurfness and age (with the peak of the density for the ribosomial pathway occupying a zone of high correlation with Smurfness, as expected given the results obtain in our analysis -Fig. 3 and Fig. 4-).

# Using Smurfness to identify new genes involved in the control of longevity

The Smurf-associated transcriptome gathers most of the so far described age-related changes, as well as the deregulation of a non-negligible number of “longevity genes”. We therefore decided to investigate whether, amongst the Smurf DEGs, we could have identified new longevity genes.

First, we decided to focus our attention on transcription factors (TFs), as their altered level could be upstream of the transcriptional signature of Smurfs. We identified 102 TFs changing their expression in Smurfs (77 upregulated, 25 downregulated, Table S12) out of the 629 annotated in Flybase. In order to reduce the possible functional redundancy in the obtained list, we used i-cisTartget[105,106] to predict putative upstream regulators of the Smurfs TFs. We selected the hits presenting a score > 4 (a score of 3 is recommended as the minimum threshold). Second, in order to not restrict our selection criteria only to TFs, we applied the same i-cisTarget algorithm to genes showing high $\log_2$FC (> |2|). Results are shown in Table S13. Out of the genes retrieved, we select 17 TFs of interest to test (Table 1), filtering the ones obtaining the best i-cisTarget scores or high deregulation in our dataset.

| Gene symbol | Selection method | Deregulation |
|---|---|---|
| *Adf1* | i-cisTarget | putative regulator TFs up in Smurf |
| *Aef1* | i-cisTarget | putative regulator TFs up in Smurf |
| *CG4360* | i-cisTarget | putative regulator TFs up in Smurf |
| *FoxP* | DESeq2 & i-cisTarget | up in Smurf & putative regulator TFs up in Smurf |
| *Hsf* | i-cisTarget | putative regulator genes up in Smurf |
| *Trl* | i-cisTarget | putative regulator TFs up in Smurf |
| *dmrt93B* | DESeq2 | up in Smurf |
| *Ets21C* | DESeq2 | up in Smurf |
| *Hey* | DESeq2 | up in Smurf |
| *kay* | DESeq2 | up in Smurf |
| *Mef2* | DESeq2 & i-cisTarget | up in Smurf & putative regulator TFs up in Smurf |
| *rib* | DESeq2 | up in Smurf |
| *Ets96B* | DESeq2 | down in Smurf |
| *GATAd* | i-cisTarget | putative regulator TFs down in Smurf |
| *GATAe* | i-cisTarget | putative regulator TFs down in Smurf |
| *NF-YB* | DESeq2 & i-cisTarget | up in Smurf & putative regulator TFs down in Smurf |
| *srp* | i-cisTarget | putative regulator TFs down in Smurf |

**Table 1. List of TFs select for experimental validation.** 17 TFs were selected for experimental testing: 8 were found in the i-cisTarget analysis, 3 in both DESeq2 and i-cisTarget analysis and 6 in the DESeq2 analysis (and chosen for their strong deregulation). The gene symbol, selection method and deregulation kind is reported.

In order to monitor the possible impact of modulating these genes on longevity , we proceeded with their knock-down (KD) and/or overexpression (OX) through GeneSwitch[107,108] (GS). Such technique, widely used in Drosophila, potentially allows tuning the KD or OX spatially and temporally (by controlling the activation of the system with the RU486 inducer) the manipulation of the genes of interest. It is particularly convenient in the case of longevity experiments as it allows to manipulate, control and treat populations of the same genetic background. As our

results are obtained from whole body data, we used the ubiquitous daughterless-GS (*da*GS) driver. When possible (i.e. transgenic line available) we performed both KD and OX during the adulthood of the fly (i.e. after eclosion) or during its whole life (development and adulthood), in order to account for possible effect on longevity due to a developmental activity of the gene. We followed the experimental scheme presented in Fig. 7a. Experiments were performed on mated females, using five different concentrations of RU486 (0 µg/mL -control-, 10 µg/mL, 50 µg/mL, 100 µg/mL, 200 µg/mL) to explore a broad range of inducing conditions as in[109]. During development, we lowered the concentrations by a factor of 10, in order to keep the same fold differences in the doses used while avoiding possible toxic effects, as suggested by[107] and done in[110].

The 48 longevity experiments performed are summarised in Fig. S13 and Table S14. While most of the genes displayed no effect, the genes selected with i-cisTarget as regulators of TFs and genes up in Smurfs displayed multiple positive hits (Fig. S13). More specifically, four TFs presented a positive effect on mean lifespan (ML) when knocked-down with one or more concentrations of RU486 during adulthood (Thritorax-like (*Trl)* + 9.53% , Adh transcription factor 1 (*Adf1*) +7.57%, *CG4360* +7.28%, *Ets96B* +6.61%) and one when overexpressed during adulthood and development (Heat shock factor (*Hsf)* +10.33%). However, a second round of validation did not confirm the effect observed for *Ets96B* and *Hsf*, while it confirmed the extension of ML for the whole body KD of *Trl* KD, *Adf1* KD, *CG4360* KD (Fig. 7b, point i). During this validation experiment, the effect of CG4360 was detected, contrary to the first screening, only for the “development & adulthood” setting. We performed the experiment a third time (Fig. S14) , validating the ML extension obtained for the RU10 condition (adulthood & development) in the second experiment. We also confirmed the expression of our transgenes through qPCR for each used line (Fig. S15), and validated that the RU486 itself has no effect on lifespan (Fig. S16).

To investigate how the KD of the genes affects the evolution of the Smurf subpopulation, we periodically recorded the proportion of Smurfs in the control and treated populations (Fig. 7b, point ii). We analysed the data by fitting two different linear regression models. First, in order to test the effect of time on the two populations independently, we fitted the following model:

$$Smurfs \sim time + \varepsilon \qquad (1)$$

where *Smurfs* is the proportion of Smurfs recorded, *time* the time interval over which we are fitting the model and $\varepsilon$ the residual.

Secondly, in order to investigate the difference between the two populations, we fitted the following model:

$$Smurfs \sim time + dose + time * dose + \varepsilon \qquad (2)$$

where *Smurfs* is the proportion of Smurfs recorded, *time* the time interval over which we are fitting the model, *dose* is the RU486 and $\varepsilon$ the residual. The final term (*time * dose*) consists in the interaction between the time and the dose of drug, and its significance (F-statistic) indicates that the drug concentration (correlating in this case with the ML increase) has an effect on the Smurf proportion as a function of time. The results (Fig. 7b, point ii) confirm the significance of

the Smurf proportion increase with time, and prove a significant effect of the drug dose on the time-dependence of Smurfs' proportion. More specifically, in all cases a ML increase is associated with a slower increase in the Smurf proportion. This does not occur when the two populations present non-significantly different lifespan (Fig. S17). These results suggest that the KD of the studied genes increases the mean lifespan by extending the non-Smurf phase and delaying the Smurf transition.

After validating the longevity effect on females, we wanted to test whether it was conserved in males. We therefore followed the same procedure as in Fig. 7a, by selecting males instead of females after the mating time. No significant effect was detected on longevity (Fig. S18), but for a negative effect (-8.1%) for Trl KD. However, even though statistically significant (log-rank test, p-value = $1.1e^{-04}$), the lifespan extension appears not to be relevant when considering the whole longevity curves.

The fact that we do not detect an increase in lifespan in males does not invalidate the results obtained on females. Indeed, the gene expression changes that we characterised in this work were based on mated female flies. However, those results highlights the issue of physiological sexual dimorphism of Drosophila longevity, which has been recently more investigated [111–113].

It is of interest that the three genes identified have all been detected as regulators of TFs upregulated in Smurfs, while the TFs downregulated in Smurfs and their putative regulators do not show significant effect. In order to investigate the possible interactions amongst the three, we referred to STRING database[114]. Fig. 6c reports the interaction network of *Adf1*. Interestingly, amongst others, we can retrieve *Trl*, *CG4360* and *Aef1* (Adult enhancer factor 1). The latter was also amongst the i-cisTarget results, and was tested in the first screening (Fig. S19). Its KD led to a negative effect on longevity correlating with the inducer dose, suggesting a possible effect of the gene on longevity dependent on the RNAi strength. However, we could not test the OX due to line unavailability. This gene might be a putative interesting target for further studies. Its annotated interaction with *Aef1* comes from a study[115] showing how *Adf1*, *Aef1* and *Trl* are putative co-actors of FOXO (binding sites are found enriched in FOXO binding regions). Interestingly, the interaction between *Adf1* and *Trl* has been experimentally demonstrated in a whole body study[116]. In addition, both the proteins have been shown to be required for the expression of the ATP synthase α-subunit[117]. However, the two genes have mostly been studied in different contexts. The GAF protein coded by *Trl* works as an antirepressor, inducing gene expression by promoting open chromatin configuration[118]. Its action has been documented as influencing disparate biological processes, such as dosage compensation in males[119], activation of genes of Hsp family[120,121], and both oogenesis and spermatogenesis[122–124].

*Adf1* is instead mostly known for its role in the nervous system. Its mutant *nalyot* presents defects in long-term memory[125]. *Adf1* expression has been confirmed in the motor neurons of the thoracic abdominal ganglion in the adult[126], regulating the dendritic branching through *Fas2* expression (also present in the STRING network). However, in none of the experiments did we observe striking differences in locomotor activity compared to other lines.

*CG4360* is zinc finger TFs[127] predicted to have repressor activity. Flybase reports it as an orthologous of the human zinc TFs ZSCAN10, ZNF157 and ZNF182, which have been poorly characterised. We decided to dub this gene *Sag1* for *Smurf-Associated Gene 1*. A blastn[128] on the sequence of *Sag1* against Drosophila genome reveals 70% identity with *Aef1*, which also

belongs to the zinc finger family. The edge in the STRING network presents an edge between *Sag1* and *Adf1*, which corresponds to interaction of the homologous genes in *C. elegans* [129].
Concerning the remaining nodes of the network, they show weaker evidence. CG11275 and CG5292 have been shown to interact with Adf1 on two-yeast hybrid assay on the FlyBI project (https://flybi.hms.harvard.edu/). *Adf1* and *stwl* are co-cited in a few publications due to their structural similarity (N-terminal MADF domain)[130–132]. The same occurs with *Dip3*[133] and *jigr1*[134,135]. Finally, as previously mentioned, *Adh* is present in the network as *Adf1* has been shown to bind its promoter and has been demonstrated to bind in vitro[136].
Our experimental results show that we can identify new genes affecting longevity through investigation of the Smurf phenotype. Interestingly, all the three genes detected affect longevity by delaying the Smurf transition, and correspondingly increasing the time the individuals spend as non-Smurfs - i.e. apparently healthy. Those results suggest that these genes are somehow acting during the non-Smurf phase, possibly modulating early steps of ageing. Further exploring the role of this process in the phase 1 (non-Smurf) - phase 2 (Smurf) transition might be a relevant target for future longevity studies aiming at identifying pro-longevity genes.

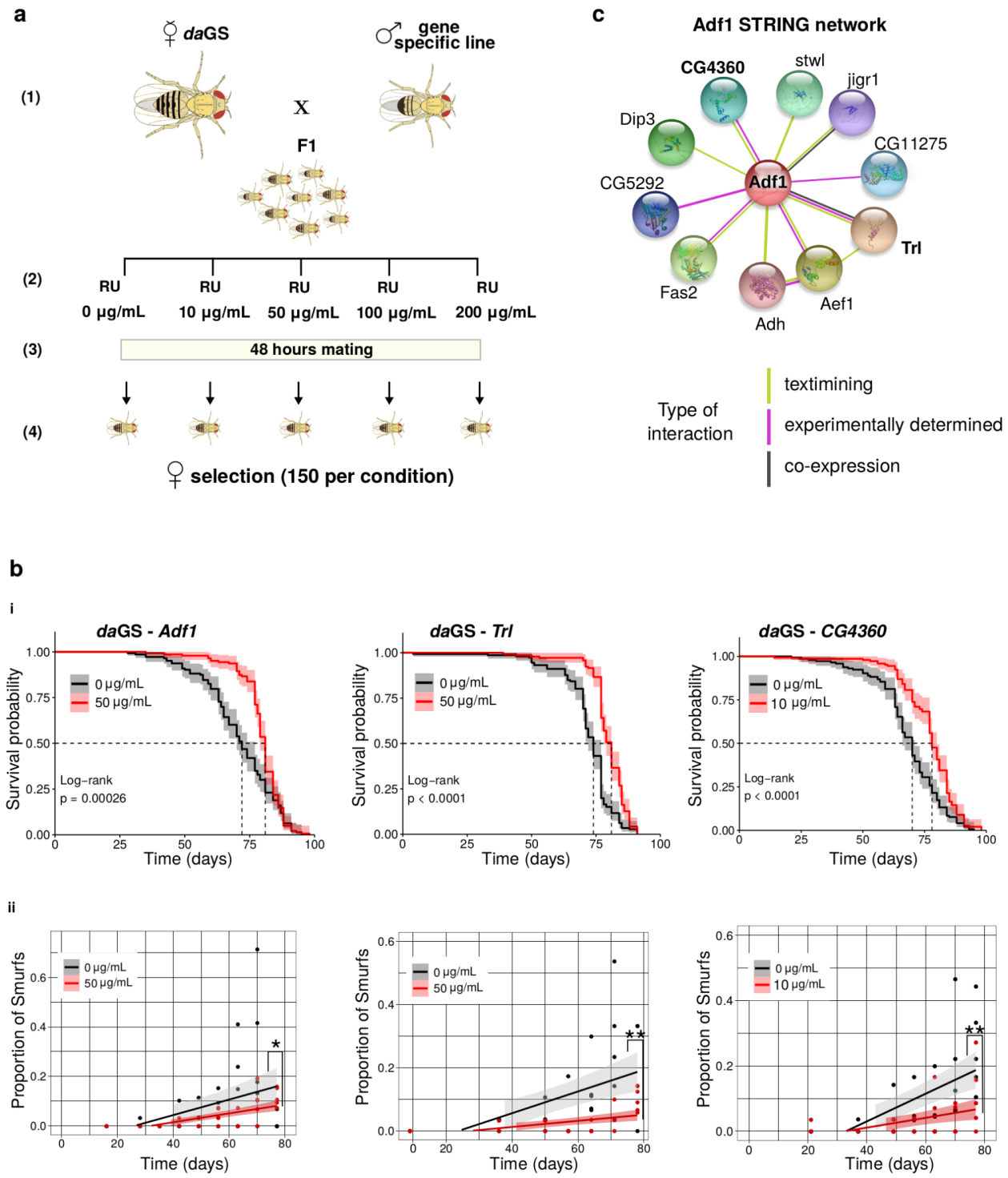



**Figure 7. Identification of new longevity genes through the study of the Smurf phenotype. a) Gene expression alteration through GeneSwitch (GS), followed protocol.** In order to perform KD and/or overexpression of the target gene in the whole body of Drosophila, we crossed virgins females of the ubiquitous *daughterless*-GS (*da*GS) driver with males carrying the genetic construct to target the gene of interest (Step 1). The F1 was developing either on food without the activator drug RU486 (experiment with gene alteration only during adulthood), either with food presenting the following RU486 gradient: 0 µg/mL -control-, 1 µg/mL, 5 µg/mL, 10 µg/mL, 20 µg/mL (experiment with gene alteration during the whole life). At the moment of eclosion, flies are transferred

in food with the following RU486 gradient: 0 µg/mL -control-, 10 µg/mL, 50 µg/mL, 100 µg/mL, 200 µg/mL. Flies are randomly distributed if not developed on drug, otherwise they are distributed according to the drug gradient (Step 2). Flies are left mating for 48 hours (Step 3), and subsequently 150 females per concentration (divided on 5 vials/30 females each) are randomly selected for the longevity experiment (Step 4). **b) Effect of *Adf1*, *Trl* and *CG4360* KD on longevity and on the Smurf dynamics in the population. (i)** The KD of *Adf1* (+11.8% , $ML_{RU0}$ = 71.0 , $ML_{RU50}$ = 79.5) and *Trl* (+10.5% , $ML_{RU0}$ = 72.2 , $ML_{RU50}$ = 79.8) in the whole body during adulthood significantly extend lifespan , as well as for the KD during the whole life of *CG4360* (+12.4% , $ML_{RU0}$ = 68.5, $ML_{RU10}$ = 77.0). P-values of the log-rank test are reported in picture. **(ii)** The proportion of Smurfs for the control and treat population of (i), plotted as a function of time are reported in the same order as in (i). Proportion of Smurf is computed as the number of Smurfs over the total amount of alive flies (Smurfs + non-Smurfs) at the given time point. Data are analyzed through the linear models reported in the text, given the linear increase displayed by the Smurf proportion in a population with time[18,21]. In all cases, the populations show a significant increase with time of the Smurf proportion (F-statistic) (*Adf1*: $slope_{RU0}$ = 0.0055 , $p\text{-value}_{RU0}$ = 4.72e-03, $slope_{RU50}$ = 0.0018 , $p\text{-value}_{RU50}$ = 4.17e-07; *Trl*: $slope_{RU0}$ = 0.0044, $p\text{-value}_{RU0}$ = 6.53e-04, $slope_{RU50}$ = 0.0009, $p\text{-value}_{RU50}$ = 5.39e-04; *CG4360*: $slope_{RU0}$ = 0.0042 , $p\text{-value}_{RU0}$ = 6.58e-05, $slope_{RU10}$ = 0.0015 , $p\text{-value}_{RU10}$ = 6.51e-03). Furthermore, the slope of the control population is significantly different from the one of the treated (F-statistic), which displays a slower increase in the Smurf proportion with time. P-value indicated in figure : * < 0.05; ** < 0.01. **(c) Adf1 interaction network from STRING database.** The three TFs identified as new longevity genes have been retrieved from i-cisTarget as putative regulators of upregulated Smurf TFs. The annotated interactions in the STRING database show how those genes have been already described together. Adf1 and Trl displayed stronger evidence (text mining, co-expression and proved interaction in Drosophila *in vitro*), while the evidence for CG4360 and Adf1 interaction comes from text mining and interaction shown between homologous in *C. elegans).* We decided to assign to CG4360 the gene name of *Sag1* (Smurf Associated Gene 1) given its potential involvement in the Smurf phase.

# Discussion

The characterization of age-associated gene expression changes based on the comparison of individuals taken at different chronological ages has given us, in the past 40 years, an overview of ageing as made of continuous and progressive changes. Nevertheless, this approach does not take into consideration the heterogeneity of life expectancies amongst a population. Indeed, even in highly genetically homogeneous and synchronised populations, such as the Drosophila ones we work with in the laboratory, individuals can die as early as 10 days old when their siblings can outlive them by more than 50 days. To account for that variability in human populations, the notion of frailty was introduced in 1979[137]. The Smurf phenotype, first described in *Drosophila melanogaster*, was shown in that organism - later extended to *C. elegans* and *Danio rerio* - to allow for the identification of individuals at higher risk of impending death than the rest of the population it belongs to, a sort of direct *in vivo* measurement of frailty. Our previous work using Drosophila suggested that in addition, these individuals are uniquely showing certain major hallmarks of ageing, such as systemic inflammation and loss of energy stores, that led us to propose a two-phase model of ageing.

In the present study, we used the Smurf phenotype as a tool to investigate what part of the ageing Drosophila transcriptome can be explained by time - individuals' chronological age - and what part can be explained by Smurfness - individuals' biological age. Previous ageing transcriptome studies in Drosophila have been conducted as a function of chronological age, therefore mixing the effect of chronological and biological age on the characterised transcriptional signature. The characterization of ageing flies' transcriptome as a function of both chronological age and Smurfness helped show that Smurf flies have a stereotypical transcriptional profile independently of their chronological age (Fig. 2a and 2c), with deregulated gene ontologies reminiscent of four out the six (ATH 1-4) literature-defined transcriptional ageing markers (Fig. 3). On the other hand, these so-called hallmarks of ageing are mostly absent from non-Smurfs as they get older (Fig. 4). The transcriptional signature of ageing being mostly

Smurf-specific confirms the ability of Smurfness to be an in *vivo* marker of biological age. One fifth of the transcriptome is affected by the Smurf phase in a time-dependent growing proportion of individuals; when simply considering samples from different chronological age the ageing signal obtained is thus a convolution of the changes occurring in Smurfs and those occurring in non-Smurfs as a function of time. By allowing us to remove this strong Smurf signal, we could access and identify the effect of time alone on the transcriptome of both non-Smurfs and Smurfs separately. Time seems to mostly increase transcriptional heterogeneity (ATH6) - by almost doubling the relative standard deviation of genes across biological replicates in both Smurfs and non-Smurfs (Fig. 6b). On Smurfs only, it appears that chronological age significantly decreases the expression of epigenetic regulators, DNA repair genes, RNA processing and cell cycle genes (ATH5-6) (Fig. 5). However, when testing the correlation of such pathways with chronological age independently from the Smurf status, they display a negative trend (Fig. 6c). We therefore speculate that such changes are already present to a weaker extent in old non-Smurfs to be subsequently exacerbated by the Smurf transition.

The Smurf phenotype is assessed through loss of controlled intestinal permeability which has been previously hypothesised to be caused by decreased and/or mislocated junction proteins[138,139]. Here we detect a strong upregulation of JAK/STAT activator ligand *upd2* ($\log_2$FC = 2.27) - chronic activation of this pathway is involved in the gut dysplasia observed with ageing in flies[140], triggering tissue regeneration in case of damage[141] - as well as a significant upregulation of midgut septate junctions components *Ssk* ($\log_2$FC = 0.50), *mesh* ($\log_2$FC = 0.68) and *Tsp2A* ($\log_2$FC = 0.52), suggesting a possible transcriptional compensatory response to decreased protein levels in the intestine or simply that these genes are expressed differently in the rest of the body. Indeed, it would be naive to consider the signal we observe at the whole body level as being strongly affected by gut-specific changes. In addition it has been shown that different tissues in Drosophila present different ageing-associated changes[36,82] and tissue-specific studies will be required to better deconvolve our signature and understand how different tissues are affected by the Smurf transition.

The Smurf phenotype was previously assessed in both males and females in Drosophila as well as in zebrafish. In both organisms, the phenotype showed sexual dimorphism, making Smurfs easier to observe in female than in male flies, while it is the opposite in the zebrafish[19]. Because of this, most of the Smurf characterization has been so far conducted using females. However, sex-specific gut differences during ageing have been described in Drosophila[111], as well as differences in the feeding behaviour or Smurf life expectancy (unpublished data). We therefore cannot exclude that the ageing-related Smurf phenotype assessed through impairment of intestinal permeability presents molecular changes specific to either males or females. Such sex-specificity could explain the results we obtained regarding the pro-longevity genetic interventions we identified (*Adf1*, *Trl, CG4360*), similarly to what already documented[111,113]. Further studies will be needed to validate the existence of the Smurf transcriptional signature in Drosophila males and look for the differences with the female one presented here. In addition, our study is centred on the transcriptional response associated with time and/or Smurfness. Even if it proved sufficient to show that a Smurf-specific transcriptional signal can be identified and used to deconvolve the “ageing transcriptome”, this approach becomes limitative when our aim is to fully understand the biology behind the Smurf transition and ageing. Our preliminary analysis on proteomic and metabolomic data confirmed that the transcriptional signature of

Smurfness we have identified is functional. Nevertheless, the combined analysis of genetic, transcriptional, translational and metabolic layers is required in order to build a multi-layer interactome framework able to infer core elements possibly triggering that transition.

The results presented here demonstrate that the Smurf phenotype is able to predict ageing markers at the gene expression level better than chronological age does. In other words, Smurf is a valid tool for identifying biological age *in vivo* in Drosophila. Secondly, our results question the way ageing has been studied and defined until now. The so-called hallmarks of ageing seem to be mostly specific to the Smurf phase, an end-of-life phenotype, although they have previously been identified as being progressively increasing with age. As initially hypothesised, the progressive increase in markers of ageing with age at the population level appears to be due to a progressive increase in the probability of sampling Smurf individuals with age. Hence, our results imply that the hallmarks of ageing are actually markers of a late stage of life, in which the individual is irremediably committed to death within a defined time frame.

Our study points towards a biphasic behaviour of ageing markers, as already hypothesised in [18,21]. In this framework, the only marker of ageing that behaves as canonically defined is the increase in transcriptional noise with age. Our work thus allows us to propose a model in which transcriptional noise increases with chronological age, enhancing the risk of triggering a dramatic transcriptional response preceding death from natural causes (Smurf phase). The stereotypical nature of this phase as shown by our different -omics approaches and the dynamics of this death, seemingly following a purely random decay[21], seem to indicate that the Smurf phase is mostly a non-plastic phase of ageing. Should we target, in the future, genes affected by the Smurf transition, or prior, in order to improve health and life spans? The genetic interventions we conducted here on transcription factors identified as putative key nodes for the Smurf transition (*Adf1*, *Trl* and *CG4360*) showed a moderate but significant modulation of longevity by increasing the amount of time an individual spends as a non-Smurf, supporting our hypothesis that the Smurf phase is mostly non-plastic. Taken together, these results question the very nature of ageing. Is it programmed[142–145] , quasi[146]- or not programmed at all? The two-phase model of ageing based on the very simple Smurf assay opens new possibilities for addressing this question, by allowing to separate a first phase that could be programmed by the genetic interaction network of a given organism, accumulating noise with time, followed by a second phase, quasi-programmed, the “shadow of actual programs”[146] accompanying death in a stereotypical manner. Reconstructing the multi-layer interactome of individuals through time and as a function of their Smurf status might be a stepping stone towards answering these questions and understanding when does ageing start. More immediately, our results question what is presently defined as ageing, if most of its molecular characteristics are actually present only in the very last phase of life that we call Smurf phase.

In the light of the results presented here we recommend the Smurf phenotype to be taken into account in Drosophila ageing studies. The absence of Smurf classification in the experimental design would indeed result in a non-negligent confounding factor altering the interpretability of the results. Since it is evolutionarily conserved, we also encourage researchers in the ageing field to take into consideration the Smurf phenotype whatever their model organism of interest, and investigate the behaviour of ageing markers as a function of the Smurf condition. The Smurf phenotype could possibly become a standard parameter to take into account in ageing research, not only as a measurement of intestinal permeability but as a marker for end-of-life.

# Supplementary materials

All codes and associated processed data are available at https://github.com/MichaelRera/SmurfsTrsc
Raw RNAseq and proteomics data are available at
Supplementary figures are available at Supplementary_figures.pdf
Supplementary tables are available at Supplementary_tables.pdf

Supplementary files:
Supplementary file 1: SupFile1_res_DESeq2_Smurf_nonSmurf.xlsx, DESeq2 results Smurf vs non-Smurf;
Supplementary file 2: SupFile2_Proteomic_results_Smurf_nonSmurf.xlsx, proteome analysis results;
Supplementary file 3: SupFile3_Metabolomic_data_Smurf_nonSmurf.xlsx, metabolomic processed data, used as an input for MetaboAnalyst;
Supplementary file 4: SupFile4_res_DESeq2_40daysNS_20daysNS.xlsx, DESeq2 results 40 days non-Smurf vs 20 days non-Smurf;
Supplementary file 5: SupFile5_res_DESeq2_40daysS_20daysS.xlsx, DESeq2 results 40 days Smurf vs 20 days Smurf.

# Material and methods

**RNA-seq: experimental design.** A synchronous isogenic population of *drosomycin*-GFP (*Drs*-GFP) Drosophila line was used for the RNA-sequencing experiment. For the longevity recording, flies were transferred on fresh food and deaths scored on alternative days. Flies were sampled for the sequencing experiment at day 20 (80% survival), day 30 (50% survival) and day 40 (10% survival). Every sample is a mixture of 8 flies. The sampling protocol for Smurfs and age-matched non-Smurfs is the following: all flies - the ones used for longevity and the ones used for sampling - are transferred on blue food overnight; at 9 a.m. 1 Smurf sample and age-matched non-Smurf are collected (Mixed samples), and all the remaining Smurfs are discharged; five hours later, 2 Smurf and non-Smurf samples are collected (5 hours Smurfs), and all the remaining Smurfs are discharged; twenty-four hours later, 3 Smurf and non-Smurf samples are collected. Note that at 90% no 5 hours Smurfs could be collected due to the low probability of flies turning Smurf at this age. After sampling, flies were immediately frozen in liquid $N_2$ and stored at -80°C up to RNA extraction.

**RNA-seq: pre-processing.** Sequencing was externalised to Intragen. Library preparation was done using 'TruSeq Stranded mRNA Sample Prep Illumina' kit and conducted on HiSeq4000 Illumina sequencer (paired-end sequencing). Data preprocessing was performed on Galaxy[147] server. Quality control was performed using FastQC[148], and resulted in no reads filtering. Reads were aligned with Hisat2[149] on the reference *D. melanogaster* genome BDGP6.95. Reads count was performed with featureCounts[150], resulting in a raw counts matrix of 15364 genes.

**RNA-seq: analysis.** Unless stated otherwise, all analysis were performed on R 3.5.3 and plots generated with ggplot2 3.3.5. PCA was performed using package DESeq2 1.22.2. Association of components with

Smurfness and age was computed using the functions PCA and dimdesc from FactoMineR 2.4. tSNE was performed on package Rtsne 0.15. Sample-to-sample distance heatmap was computed using function dist from stats 3.5.3, and plotted using heatmap 1.0.12. The main DEGs analysis was performed on DESeq2 1.22.2, while validation analysis on edgeR 3.24.3. Enrichmend analysis was performed with the Bioconductor package clusterProfiler 3.10.1, which calls fgsea 1.8.0; analysis was ran with the following parameters: nPerm = 15000, minGSSize = 10, maxGSSize = 600. Enrichment plot was generated with the function emmaplot from the same package. Venn diagram (Fig. 4C) was generated using eulerr Rshiny app.

**Proteomic data collection and analysis.** *Drs*GFP Smurfs (8 hours) and non-Smurfs were sampled at 80 and 10% survival in quadruplicates of 10 females. Flies were quickly homogenised in 96µL NU-PAGE 1X sample buffer containing antiproteases and quickly spun to precipitate debris. 40µL of samples were then loaded on a NU-PAGE 10% Bis-Tris gel prior to being sent for label free proteomics quantification.

**Metabolomic data collection and analysis.** *Drs*GFP Smurfs and non-Smurfs were sampled at 50% survival. Each sample corresponds to a mixture of 20/30 individuals, for a total of 7 Smurf and 7 non-Smurf samples. Drosophila were weighted to reach around 30 mg in a 2 mL-homogenizer tube with ceramic beads (Hard Tissue Homogenizing CK28, 2.8 mm zirconium oxide beads; Precellys, Bertin Technologies, France). Then, 1 mL of ice-cold $CH_3OH$/water (9/1, -20°C, with internal standards) was added to the homogenizer tube. Samples were homogenised (3 cycles of 20 s/ 5000 rpm; Precellys 24, Bertin Technologies) and homogenates were then centrifuged (10 min at 15000 g, 4°C). Supernatants were collected and several fractions were split to be analysed by different Liquid and Gaz chromatography coupled with mass spectrometers (LC/MS and GC/MS)[151] .Widely targeted analysis by GC-MS/MS was performed on a coupling 7890A gas chromatography (Agilent Technologies) Triple Quadrupole 7000C (Agilent Technologies) and was previously described in[152]. Polyamines, nucleotides, cofactors, bile acids and short chain fatty acids analyses were performed by LC-MS/MS with a 1260 UHPLC (Ultra-High Performance Liquid Chromatography) (Agilent Technologies) coupled to a QQQ 6410 (Agilent Technologies) and were previously described in[152]. Pseudo-targeted analysis by UHPLC-HRAM (Ultra-High Performance Liquid Chromatography – High Resolution Accurate Mass) was performed on a U3000 (Dionex) / Orbitrap q-Exactive (Thermo) coupling, previously described in[152,153]. All targeted treated data were merged and cleaned with a dedicated R (version 4.0) package (@Github/Kroemerlab/GRMeta). 202 metabolites were detected. All the analysis presented (fold change estimation, Wilcoxon test and quantitative enrichment analysis) were done using MetaboAnlyst[156]. One Smurf sample was removed from the analysis as generated starting from 8 individuals only, resulting in a total N of 7 non-Smurfs and 6 Smurfs. Samples were normalised by weight. Gene expression and metabolites representation KEGG maps were generated using pathview 1.2[154] (R package).

**Longevity experiments.** All the flies are kept in closed vials in incubators at controlled temperature, humidity and 12 hours light cycle. Experiments are carried at 26°C. Longevity experiments were run on the following food composition: 5.14% (w/v) yeast, 2.91\% (w/v) corn, 4.28% (w/v) sugar, 0.57% (w/v) agar and Methyl 4-hydroxybenzoate (Moldex) at a final concentration of 5.3 g/L to prevent fungi contamination. Just after eclosion, flies are collected in tubes with food and RU486 (Fig. 6a). Males and females are left together to mate for 48 hours. After that time, males or females (depending on the experiment) are sorted in a number of 30 per vial, with 5 vials for each RU concentration (total N per concentration is 150). Flies are transferred to new vials with fresh food and scored three times per week (Monday, Wednesday, Friday). An exception are the first two weeks of the experiment, when females undergo an additional transfer on Saturday or Sunday due to the fertilised eggs altering the food composition. The food is prepared the day before the scoring (1.25 mL per vial) and stored at room temperature.

**Lines used.** *daGS* driver (provided by Tricoire laboratory, Université de Paris). Bloomington stock (with associated targeted gene if GS): *Drs*-GFP 55707, *dmrt93B*, 27657; *Ets21C*, 39069; *Hey*, 41650; *kay*, 27722; *Mef2*, 28699; *rib*, 50682; *Ets96B*, 31935; *GATAd*, 34625; *GATAe*, 33748; *srp*, 28606; *NF-yB*, 57254; *Aef1*, 80390; *CG4360*, 51813; *FoxP*, 26774; *Hsf*: 41581; *Trl* 41582. FlyORF stock (with associated targeted genes): *NF-yB*, F001895; *CG4360*, F000063; *dmrt93B*, F000445; *Ets96B*, F000142; *Ets21C*, F000624; *srp*, F000720; *GATAd*, F000714; *Hsf*, F000699. VRDC stock (with associated gene): *Adf1*, 4278.

**Smurf assay recording.** Flies were transferred to food containing the blue dye FD&C #1 at 2.5\% (w/v) 24 hours prior to Smurfs counting. The dye is added as the last component in the food preparation, and dissolved in it. At the moment of the counting, flies were transferred back on normal food. All the flies are therefore spending the same amount of time on blue food, in order not to introduce bias in the counts. Note that with the following method we are not having information about the time at which the Smurfs are becoming such. However, as the Smurfs spend on average the same amount of time in this phase[21], recording the presence of a “mixed” Smurf population provides a good estimation of their appearance in the population. Smurf counting was performed every two weeks while the population was in the survival plateau, and every week once it exited it.

**RNA extraction and qPCR quantification.** Extraction of RNA was performed using the Trizol protocol as in[155], adapted to the amount of tissue used. Each sample corresponds to a mixture of 3 flies for the RT-qPCR experiments and 8 flies for the RNA-Seq. For the RT-qPCRs, RNA was retro-transcribed using the Applied Biosystems cDNA Reverse Transcription Kit. RT-qPCR was subsequently performed using the Applied Biosystem PowerTrack SYBR Master Mix on Biorad CFX 96. Primers were designed on Benchling. *Adf1* Fw: ACAGCCCTTCAACGGCA, *Adf1* Rw: CGGCTCGTAGAAGTATGGCT; *CG4360* Fw: CAGCAGAGCACCCTTACCAA, *CG4360* Rw: GGAGCGGGCATTGAGTGAT; *Trl* Fw: TCCTATCCACGCCAAAGGCAAA, Trl Rw: TAGCAAATGGGGCAAGTAGCAGG; *Act* Fw: CCATCAGCCAGCAGTCGTCTA, *Act* Rw: ACCAGAGCAGCAACTTCTTCG.

# Acknowledgements

We thank Camille Garcia from the proteomic platform of Institut Jacques Monod (ProtéoSeine) for producing and pre-processing the proteomics data presented in the manuscript. We thank Bastian Greshake Tzovaras for his helpful comments on the manuscript.

# Contributions

M.R. conceived the presented idea and model. F.Z and M.R. conceived, planned and performed the analysis and experiments as well as wrote the manuscript. H.B. performed the RT-qPCR experiments and analysed the results. S.S.M. and J.L.M. participated in the longevity experiments. S.B. provided technical support for the analysis. C.C., F.A. and S.D. performed and analysed the metabolomics experiments. J.A. and M.R. performed and analysed the proteomics experiments. C.A. helped with the RNAseq analysis. All authors discussed the results and contributed to the final manuscript.

# Funding

Michael Rera is funded by the CNRS, Flaminia Zane is funded by Sorbonne Université Interdisciplinary research PhD grant. This project was funded by the ANR ADAGIO (ANR-20-CE44-0010). Thanks to the Bettencourt Schueller Foundation long term partnership, this work was partly supported by the CRI Core Research Fellowship to Michael Rera.

# References


1. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The Hallmarks of Aging. *Cell* **153**, 1194–1217 (2013).
2. Cohen, A. A. *et al.* Lack of consensus on an aging biology paradigm? A global survey reveals an agreement to disagree, and the need for an interdisciplinary framework. *Mech. Ageing Dev.* **191**, 111316 (2020).
3. Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* **19**, 371–384 (2018).
4. Rockwood, K., Fox, R. A., Stolee, P., Robertson, D. & Beattie, B. L. Frailty in elderly people: an evolving concept. *CMAJ Can. Med. Assoc. J.* **150**, 489–495 (1994).
5. Rockwood, K. *et al.* A brief clinical instrument to classify frailty in elderly people. *Lancet Lond. Engl.* **353**, 205–206 (1999).
6. Fried, L. P. *et al.* Frailty in older adults: evidence for a phenotype. *J. Gerontol. A. Biol. Sci. Med. Sci.* **56**, M146-156 (2001).
7. Mitnitski, A. B., Mogilner, A. J. & Rockwood, K. Accumulation of Deficits as a Proxy Measure of Aging. *Sci. World J.* **1**, 323–336 (2001).
8. Rockwood, K. & Mitnitski, A. Frailty in relation to the accumulation of deficits. *J. Gerontol. A. Biol. Sci. Med. Sci.* **62**, 722–727 (2007).
9. Rockwood, K., Andrew, M. & Mitnitski, A. A comparison of two approaches to measuring frailty in elderly people. *J. Gerontol. A. Biol. Sci. Med. Sci.* **62**, 738–743 (2007).
10. Kim, S., Welsh, D. A., Cherry, K. E., Myers, L. & Jazwinski, S. M. Association of healthy aging with parental longevity. *Age* **35**, 1975–1982 (2013).
11. Bocklandt, S. *et al.* Epigenetic Predictor of Age. *PLoS ONE* **6**, e14821 (2011).
12. Hannum, G. *et al.* Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Mol. Cell* **49**, 359–367 (2013).
13. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115 (2013).
14. Tarkhov, A. E. *et al.* A universal transcriptomic signature of age reveals the temporal scaling of Caenorhabditis elegans aging trajectories. *Sci. Rep.* **9**, 1–18 (2019).
15. Meyer, D. H. & Schumacher, B. BiT age: A transcriptome-based aging clock near the theoretical limit of accuracy. *Aging Cell* **20**, e13320 (2021).
16. Tanaka, T. *et al.* Plasma proteomic signature of age in healthy humans. *Aging Cell* **17**, e12799 (2018).
17. Lehallier, B. *et al.* Undulating changes in human plasma proteome profiles across the lifespan. *Nat. Med.* **25**, 1843–1850 (2019).
18. Rera, M., Clark, R. I. & Walker, D. W. Intestinal barrier dysfunction links metabolic and inflammatory markers of aging to death in Drosophila. *Proc. Natl. Acad. Sci.* **109**, 21528–21533 (2012).
19. Dambroise, E. *et al.* Two phases of aging separated by the Smurf transition as a public path to death. *Sci. Rep.* **6**, (2016).
20. Martins, R. R., McCracken, A. W., Simons, M. J. P., Henriques, C. M. & Rera, M. How to Catch a Smurf? – Ageing and Beyond… In vivo Assessment of Intestinal Permeability in Multiple Model Organisms. *Bio-Protoc.* **8**, e2722 (2018).
21. Tricoire, H. & Rera, M. A New, Discontinuous 2 Phases of Aging Model: Lessons from Drosophila melanogaster. *PLOS ONE* **10**, e0141920 (2015).
22. Mackay, T. F. *et al.* The Drosophila melanogaster Genetic Reference Panel. *Nature* **482**, 173–8 (2012).
23. Rera, M., Vallot, C. & Lefrançois, C. The Smurf transition: new insights on ageing from

end-of-life studies in animal models. *Curr. Opin. Oncol.* **30**, 38–44 (2018).
24. Clark, R. I. *et al.* Distinct Shifts in Microbiota Composition during Drosophila Aging Impair Intestinal Function and Drive Mortality. *Cell Rep.* **12**, 1656–1667 (2015).
25. Frenk, S. & Houseley, J. Gene expression hallmarks of cellular ageing. *Biogerontology* **19**, 547–566 (2018).
26. Bitner, K., Shahrestani, P., Pardue, E. & Mueller, L. D. Predicting death by the loss of intestinal function. *PLOS ONE* **15**, e0230970 (2020).
27. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
28. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
29. Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinforma. Oxf. Engl.* **35**, 2084–2092 (2019).
30. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
31. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545–15550 (2005).
32. Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J. M. & Hoffmann, J. A. The dorsoventral regulatory gene cassette spätzle/Toll/cactus controls the potent antifungal response in Drosophila adults. *Cell* **86**, 973–983 (1996).
33. Lemaitre, B. *et al.* A recessive mutation, immune deficiency (imd), defines two distinct control pathways in the Drosophila host defense. *Proc. Natl. Acad. Sci.* **92**, 9465–9469 (1995).
34. Dushay, M. S., Asling, B. & Hultmark, D. Origins of immunity: Relish, a compound Rel-like gene in the antibacterial defense of Drosophila. *Proc. Natl. Acad. Sci.* **93**, 10343–10347 (1996).
35. Pletcher, S. D. *et al.* Genome-Wide Transcript Profiles in Aging and Calorically Restricted Drosophila melanogaster. *Curr. Biol.* **12**, 712–723 (2002).
36. Girardot, F., Lasbleiz, C., Monnier, V. & Tricoire, H. Specific age related signatures in Drosophila body parts transcriptome. *BMC Genomics* **7**, 69 (2006).
37. Zhan, M. *et al.* Temporal and spatial transcriptional profiles of aging in Drosophila melanogaster. *Genome Res.* **17**, 1236–1243 (2007).
38. Landis, G. N. *et al.* Similar gene expression patterns characterize aging and oxidative stress in Drosophila melanogaster. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 7663–7668 (2004).
39. Moskalev, A. A. *et al.* Transcriptome Analysis of Long-lived Drosophila melanogaster E ( z ) Mutants Sheds Light on the Molecular Mechanisms of Longevity. *Sci. Rep.* **9**, 1–11 (2019).
40. Bordet, G., Lodhi, N., Kossenkov, A. & Tulin, A. Age-Related Changes of Gene Expression Profiles in Drosophila. *Genes* **12**, 1982 (2021).
41. Wang, X. *et al.* Ageing induces tissue-specific transcriptomic changes in Caenorhabditis elegans. *EMBO J.* **41**, e109633 (2022).
42. Lee, C.-K., Weindruch, R. & Prolla, T. A. Gene-expression profile of the ageing brain in mice. *Nat. Genet.* **25**, 294–297 (2000).
43. de Magalhães, J. P., Curado, J. & Church, G. M. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinforma. Oxf. Engl.* **25**, 875–881 (2009).
44. Benayoun, B. A. *et al.* Remodeling of epigenome and transcriptome landscapes with aging in mice reveals widespread induction of inflammatory responses. *Genome Res.* **29**, 697–709 (2019).

45. Kazakevych, J., Stoyanova, E., Liebert, A. & Varga-Weisz, P. Transcriptome analysis identifies a robust gene expression program in the mouse intestinal epithelium on aging. *Sci. Rep.* **9**, 1–8 (2019).
46. Palmer, D., Fabris, F., Doherty, A., Freitas, A. A. & de Magalhães, J. P. Ageing transcriptome meta-analysis reveals similarities and differences between key mammalian tissues. *Aging* **13**, 3313–3341 (2021).
47. Furman, D. *et al.* Expression of specific inflammasome gene modules stratifies older individuals into two extreme clinical and immunological states. *Nat. Med.* **23**, 174–184 (2017).
48. Franceschi, C. *et al.* Inflamm-aging. An evolutionary perspective on immunosenescence. *Ann. N. Y. Acad. Sci.* **908**, 244–254 (2000).
49. Franceschi, C. & Campisi, J. Chronic Inflammation (Inflammaging) and Its Potential Contribution to Age-Associated Diseases. *J. Gerontol. Ser. A* **69**, S4–S9 (2014).
50. Ferrucci, L. & Fabbri, E. Inflammageing: chronic inflammation in ageing, cardiovascular disease, and frailty. *Nat. Rev. Cardiol.* **15**, 505–522 (2018).
51. Larkin, A. *et al.* FlyBase: updates to the Drosophila melanogaster knowledge base. *Nucleic Acids Res.* **49**, D899–D907 (2021).
52. Yang, J. & Tower, J. Expression of hsp22 and hsp70 Transgenes Is Partially Predictive of Drosophila Survival Under Normal and Stress Conditions. *J. Gerontol. A. Biol. Sci. Med. Sci.* **64A**, 828–838 (2009).
53. Yoshida, H., Matsui, T., Yamamoto, A., Okada, T. & Mori, K. XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. *Cell* **107**, 881–891 (2001).
54. Xu, S., Hou, D., Liu, J. & Ji, L. Age-associated changes in GSH S-transferase gene/proteins in livers of rats. *Redox Rep. Commun. Free Radic. Res.* **23**, 213–218 (2018).
55. Maurya, P. K. & Rizvi, S. I. Age-Dependent Changes in Glutathione-S-Transferase: Correlation with Total Plasma Antioxidant Potential and Red Cell Intracellular Glutathione. *Indian J. Clin. Biochem.* **25**, 398–400 (2010).
56. Simonsen, A. *et al.* Promoting basal levels of autophagy in the nervous system enhances longevity and oxidant resistance in adult Drosophila. *Autophagy* **4**, 176–184 (2008).
57. Singh, S. P. *et al.* Disruption of the mGsta4 Gene Increases Life Span of C57BL Mice. *J. Gerontol. A. Biol. Sci. Med. Sci.* **65A**, 14–23 (2010).
58. Copeland, J. M. *et al.* Extension of Drosophila Life Span by RNAi of the Mitochondrial Respiratory Chain. *Curr. Biol.* **19**, 1591–1598 (2009).
59. Zou, S., Meadows, S., Sharp, L., Jan, L. Y. & Jan, Y. N. Genome-wide study of aging and oxidative stress response in Drosophila melanogaster. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 13726–13731 (2000).
60. McCarroll, S. A. *et al.* Comparing genomic expression patterns across species identifies shared transcriptional profile in aging. *Nat. Genet.* **36**, 197–204 (2004).
61. Zahn, J. M. *et al.* Transcriptional Profiling of Aging in Human Muscle Reveals a Common Aging Signature. *PLOS Genet.* **2**, e115 (2006).
62. Curtis, C. *et al.* Transcriptional profiling of MnSOD-mediated lifespan extension in Drosophila reveals a species-general network of aging and metabolic genes. *Genome Biol.* **8**, R262 (2007).
63. Cannon, L. *et al.* Expression patterns of cardiac aging in Drosophila. *Aging Cell* **16**, 82–92 (2017).
64. Ma, X. *et al.* Analysis of C. elegans muscle transcriptome using trans-splicing-based RNA tagging (SRT). *Nucleic Acids Res.* **44**, e156 (2016).
65. Peters, M. J. *et al.* The transcriptional landscape of age in human peripheral blood. *Nat. Commun.* **6**, 8570 (2015).
66. Long, D. M. *et al.* Lactate dehydrogenase expression modulates longevity and

neurodegeneration in Drosophila melanogaster. *Aging* **12**, (2020).
67. Niwa, R. & Niwa, Y. S. Enzymes for ecdysteroid biosynthesis: their biological functions in insects and beyond. *Biosci. Biotechnol. Biochem.* **78**, 1283–1292 (2014).
68. Schwedes, C. C. & Carney, G. E. Ecdysone signaling in adult Drosophila melanogaster. *J. Insect Physiol.* **58**, 293–302 (2012).
69. Zipper, L., Jassmann, D., Burgmer, S., Görlich, B. & Reiff, T. Ecdysone steroid hormone remote controls intestinal stem cell fate decisions via the PPARγ-homolog Eip75B in Drosophila. *eLife* **9**, e55795 (2020).
70. Carlson, J. R. & Hogness, D. S. Developmental and functional analysis of Jonah gene expression. *Dev. Biol.* **108**, 355–368 (1985).
71. Eriksson, M. *et al.* Recurrent de novo point mutations in lamin A cause Hutchinson–Gilford progeria syndrome. *Nature* **423**, 293–298 (2003).
72. Kiss, M. *et al.* Drosophila type IV collagen mutation associates with immune system activation and intestinal dysfunction. *Matrix Biol. J. Int. Soc. Matrix Biol.* **49**, 120–131 (2016).
73. Niedernhofer, L. J. *et al.* Nuclear Genomic Instability and Aging. *Annu. Rev. Biochem.* **87**, 295–322 (2018).
74. Collin, G., Huna, A., Warnier, M., Flaman, J.-M. & Bernard, D. Transcriptional repression of DNA repair genes is a hallmark and a cause of cellular senescence. *Cell Death Dis.* **9**, 1–14 (2018).
75. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
76. Friedman, D. B. & Johnson, T. E. A mutation in the age-1 gene in Caenorhabditis elegans lengthens life and reduces hermaphrodite fertility. *Genetics* **118**, 75–86 (1988).
77. Tatar, M. *et al.* A mutant Drosophila insulin receptor homolog that extends life-span and impairs neuroendocrine function. *Science* **292**, 107–110 (2001).
78. Clancy, D. J. *et al.* Extension of life-span by loss of CHICO, a Drosophila insulin receptor substrate protein. *Science* **292**, 104–106 (2001).
79. Coschigano, K. T. *et al.* Deletion, but not antagonism, of the mouse growth hormone receptor results in severely decreased body weights, insulin, and insulin-like growth factor I levels and increased life span. *Endocrinology* **144**, 3799–3810 (2003).
80. Aging Atlas Consortium. Aging Atlas: a multi-omics database for aging biology. *Nucleic Acids Res.* **49**, D825–D830 (2021).
81. Tacutu, R. *et al.* Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Res.* **46**, D1083–D1090 (2018).
82. Stegeman, R. & Weake, V. M. Transcriptional signatures of aging. *J. Mol. Biol.* **429**, 2427–2437 (2017).
83. Latorre, E. & Harries, Lorna. W. Splicing regulatory factors, ageing and age-related disease. *Ageing Res. Rev.* **36**, 165–170 (2017).
84. Bhadra, M., Howell, P., Dutta, S., Heintz, C. & Mair, W. B. Alternative splicing in aging and longevity. *Hum. Genet.* **139**, 357–369 (2020).
85. Holly, A. C. *et al.* Changes in splicing factor expression are associated with advancing age in man. *Mech. Ageing Dev.* **134**, 356–366 (2013).
86. Pagiatakis, C., Musolino, E., Gornati, R., Bernardini, G. & Papait, R. Epigenetics of aging and disease: a brief overview. *Aging Clin. Exp. Res.* **33**, 737–745 (2021).
87. Saul, D. & Kosinsky, R. L. Epigenetics of Aging and Aging-Associated Diseases. *Int. J. Mol. Sci.* **22**, 401 (2021).
88. Halim, M. A. *et al.* Ageing, Drosophila melanogaster and Epigenetics. *Malays. J. Med. Sci. MJMS* **27**, 7–19 (2020).
89. Maleszewska, M., Mawer, J. S. P. & Tessarz, P. Histone Modifications in Ageing and Lifespan Regulation. *Curr. Mol. Biol. Rep.* **2**, 26–35 (2016).
90. Lee, J.-H., Kim, E. W., Croteau, D. L. & Bohr, V. A. Heterochromatin: an epigenetic point

of view in aging. *Exp. Mol. Med.* **52**, 1466–1474 (2020).
91. Villeponteau, B. The heterochromatin loss model of aging. *Exp. Gerontol.* **32**, 383–394 (1997).
92. Schumacher, B., Pothof, J., Vijg, J. & Hoeijmakers, J. H. J. The central role of DNA damage in the ageing process. *Nature* **592**, 695–703 (2021).
93. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).
94. Bahar, R. *et al.* Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature* **441**, 1011 (2006).
95. Perez-Gomez, A., Buxbaum, J. N. & Petrascheck, M. The Aging Transcriptome: Read Between the Lines. *Curr. Opin. Neurobiol.* **63**, 170–175 (2020).
96. Somel, M., Khaitovich, P., Bahn, S., Pääbo, S. & Lachmann, M. Gene expression becomes heterogeneous with age. *Curr. Biol.* **16**, R359–R360 (2006).
97. Enge, M. *et al.* Single-Cell Analysis of Human Pancreas Reveals Transcriptional Signatures of Aging and Somatic Mutation Patterns. *Cell* **171**, 321-330.e14 (2017).
98. Kedlian, V. R., Donertas, H. M. & Thornton, J. M. The widespread increase in inter-individual variability of gene expression in the human brain with age. *Aging* **11**, 2253–2280 (2019).
99. Işıldak, U., Somel, M., Thornton, J. M. & Dönertaş, H. M. Temporal changes in the gene expression heterogeneity during brain development and aging. *Sci. Rep.* **10**, 4080 (2020).
100. Brinkmeyer-Langford, C. L., Guan, J., Ji, G. & Cai, J. J. Aging Shapes the Population-Mean and -Dispersion of Gene Expression in Human Brains. *Front. Aging Neurosci.* **8**, (2016).
101. Martinez-Jimenez, C. P. *et al.* Aging increases cell-to-cell transcriptional variability upon immune stimulation. *Science* **355**, 1433–1436 (2017).
102. Rogina, B., Vaupel, J. W., Partridge, L. & Helfand, S. L. Regulation of gene expression is preserved in aging Drosophila melanogaster. *Curr. Biol.* **8**, 475–478 (1998).
103. Ximerakis, M. *et al.* Single-cell transcriptomic profiling of the aging mouse brain. *Nat. Neurosci.* **22**, 1696–1708 (2019).
104. Fasano, G. & Franceschini, A. A multidimensional version of the Kolmogorov-Smirnov test. *Mon. Not. R. Astron. Soc.* **225**, 155–170 (1987).
105. Herrmann, C., Van de Sande, B., Potier, D. & Aerts, S. i-cisTarget: an integrative genomics method for the prediction of regulatory features and cis-regulatory modules. *Nucleic Acids Res.* **40**, e114 (2012).
106. Imrichová, H., Hulselmans, G., Kalender Atak, Z., Potier, D. & Aerts, S. i-cisTarget 2015 update: generalized cis-regulatory enrichment analysis in human, mouse and fly. *Nucleic Acids Res.* **43**, W57–W64 (2015).
107. Osterwalder, T., Yoon, K. S., White, B. H. & Keshishian, H. A conditional tissue-specific transgene expression system using inducible GAL4. *Proc. Natl. Acad. Sci.* **98**, 12596–12601 (2001).
108. Roman, G., Endo, K., Zong, L. & Davis, R. L. P{Switch}, a system for spatial and temporal control of gene expression in Drosophila melanogaster. *Proc. Natl. Acad. Sci.* **98**, 12602–12607 (2001).
109. Tricoire, H. *et al.* The steroid hormone receptor EcR finely modulates Drosophila lifespan during adulthood in a sex-specific manner. *Mech. Ageing Dev.* **130**, 547–552 (2009).
110. Rera, M., Monnier, V. & Tricoire, H. Mitochondrial electron transport chain dysfunction during development does not extend lifespan in Drosophila melanogaster. *Mech. Ageing Dev.* **131**, 156–164 (2010).
111. Regan, J. C. *et al.* Sex difference in pathology of the ageing gut mediates the greater response of female lifespan to dietary restriction. *eLife* **5**, e10956.

112. Belmonte, R. L., Corbally, M.-K., Duneau, D. F. & Regan, J. C. Sexual Dimorphisms in Innate Immunity and Responses to Infection in Drosophila melanogaster. *Front. Immunol.* **10**, (2020).
113. Garratt, M. Why do sexes differ in lifespan extension? Sex-specific pathways of aging and underlying mechanisms for dimorphic responses. *Nutr. Healthy Aging* **5**, 247–259 (2020).
114. Szklarczyk, D. *et al.* The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612 (2021).
115. Birnbaum, A., Wu, X., Tatar, M., Liu, N. & Bai, H. Age-Dependent Changes in Transcription Factor FOXO Targeting in Female Drosophila. *Front. Genet.* **10**, (2019).
116. Lomaev, D. *et al.* The GAGA factor regulatory network: Identification of GAGA factor associated proteins. *PLOS ONE* **12**, e0173602 (2017).
117. Talamillo, A. *et al.* Expression of the Drosophila melanogaster ATP synthase α subunit gene is regulated by a transcriptional element containing GAF and Adf-1 binding sites. *Eur. J. Biochem.* **271**, 4003–4013 (2004).
118. Farkas, G. *et al.* The Trithorax-like gene encodes the Drosophila GAGA factor. *Nature* **371**, 806–808 (1994).
119. Greenberg, A. J., Yanowitz, J. L. & Schedl, P. The Drosophila GAGA factor is required for dosage compensation in males and for the formation of the male-specific-lethal complex chromatin entry site at 12DE. *Genetics* **166**, 279–289 (2004).
120. Weber, J. A., Taxman, D. J., Lu, Q. & Gilmour, D. S. Molecular architecture of the hsp70 promoter after deletion of the TATA box or the upstream regulation region. *Mol. Cell. Biol.* **17**, 3799–3808 (1997).
121. Leibovitch, B. A. *et al.* GAGA factor and the TFIID complex collaborate in generating an open chromatin structure at the Drosophila melanogaster hsp26 promoter. *Mol. Cell. Biol.* **22**, 6148–6157 (2002).
122. Fedorova, E. V. *et al.* GAGA protein is required for multiple aspects of Drosophila oogenesis and female fertility. *genesis* **57**, e23269 (2019).
123. Dorogova, N. V., Fedorova, E. V., Bolobolova, E. Us., Ogienko, A. A. & Baricheva, E. M. GAGA protein is essential for male germ cell development in Drosophila. *genesis* **52**, 738–751 (2014).
124. Dorogova, N. V., Khrushcheva, A. S., Fedorova, E. V., Ogienko, A. A. & Baricheva, E. M. [Role of GAGA Factor in Drosophila Primordial Germ Cell Migration and Gonad Development]. *Ontogenez* **47**, 40–48 (2016).
125. DeZazzo, J. *et al.* nalyot, a Mutation of the Drosophila Myb-Related Adf1 Transcription Factor, Disrupts Synapse Formation and Olfactory Memory. *Neuron* **27**, 145–158 (2000).
126. Timmerman, C. *et al.* The Drosophila Transcription Factor Adf-1 (nalyot) Regulates Dendrite Growth by Controlling FasII and Staufen Expression Downstream of CaMKII and Neural Activity. *J. Neurosci.* **33**, 11916–11931 (2013).
127. Schertel, C. *et al.* A large-scale, in vivo transcription factor screen defines bivalent chromatin as a key property of regulatory factors mediating Drosophila wing development. *Genome Res.* **25**, 514–523 (2015).
128. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
129. Reece-Hoyes, J. S. *et al.* Extensive Rewiring and Complex Evolutionary Dynamics in a C. elegans Multiparameter Transcription Factor Network. *Mol. Cell* **51**, 116–127 (2013).
130. Clark, K. A. & McKearin, D. M. The Drosophila stonewall gene encodes a putative transcription factor essential for germ cell development. *Dev. Camb. Engl.* **122**, 937–950 (1996).
131. Azzam, G. & Liu, J.-L. Only One Isoform of Drosophila melanogaster CTP Synthase Forms the Cytoophidium. *PLoS Genet.* **9**, e1003256 (2013).

132. Hull, R. *et al.* The Drosophila Retinoblastoma Binding Protein 6 Family Member Has Two Isoforms and Is Potentially Involved in Embryonic Patterning. *Int. J. Mol. Sci.* **16**, 10242–10266 (2015).
133. Bhaskar, V. & Courey, A. J. The MADF-BESS domain factor Dip3 potentiates synergistic activation by Dorsal and Twist. *Gene* **299**, 173–184 (2002).
134. Fernando, C., Audibert, A., Simon, F., Tazi, J. & Juge, F. A Role for the Serine/Arginine-Rich (SR) Protein B52/SRSF6 in Cell Growth and Myc Expression in Drosophila. *Genetics* **199**, 1201–1211 (2015).
135. Ellis, K., Friedman, C. & Yedvobnick, B. Drosophila domino Exhibits Genetic Interactions with a Wide Spectrum of Chromatin Protein-Encoding Loci. *PLoS ONE* **10**, e0142635 (2015).
136. Heberlein, U., England, B. & Tjian, R. Characterization of Drosophila transcription factors that activate the tandem promoters of the alcohol dehydrogenase gene. *Cell* **41**, 965–977 (1985).
137. Vaupel, J. W., Manton, K. G. & Stallard, E. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**, 439–454 (1979).
138. Dornan, A. J., Halberg, K. A., Beuter, L.-K., Davies, S.-A. & Dow, J. A. T. *The septate junction protein Snakeskin is critical for epithelial barrier function and tissue homeostasis in the Malpighian tubules of adult Drosophila*. http://biorxiv.org/lookup/doi/10.1101/2020.12.14.422678 (2020) doi:10.1101/2020.12.14.422678.
139. Salazar, A. M. *et al.* Intestinal Snakeskin Limits Microbial Dysbiosis during Aging and Promotes Longevity. *iScience* **9**, 229–243 (2018).
140. Herrera, S. C. & Bach, E. A. JAK/STAT signaling in stem cells and regeneration: from Drosophila to vertebrates. *Development* **146**, dev167643 (2019).
141. Li, H., Qi, Y. & Jasper, H. Preventing Age-Related Decline of Gut Compartmentalization Limits Microbiota Dysbiosis and Extends Lifespan. *Cell Host Microbe* **19**, 240–253 (2016).
142. Kirkwood, T. B. & Melov, S. On the programmed/non-programmed nature of ageing within the life history. *Curr Biol* **21**, R701-7 (2011).
143. Kowald, A. & Kirkwood, T. B. L. Can aging be programmed? A critical literature review. *Aging Cell* **15**, 986–998 (2016).
144. Bredesen, D. E. Rebuttal to Austad: 'Is aging programmed? *Aging Cell* **3**, 261–262 (2004).
145. Austad, S. N. Is aging programed? *Aging Cell* **3**, 249–251 (2004).
146. Blagosklonny, M. V. Aging is not programmed: Genetic pseudo-program is a shadow of developmental growth. *Cell Cycle* **12**, 3736–3742 (2013).
147. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* **46**, W537–W544 (2018).
148. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
149. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
150. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinforma. Oxf. Engl.* **30**, 923–930 (2014).
151. Grajeda-Iglesias, C. *et al.* Oral administration of Akkermansia muciniphila elevates systemic antiaging and anticancer metabolites. *Aging* **13**, 6375–6405 (2021).
152. Durand, S. *et al.* Chapter 11 - The intracellular metabolome of starving cells. in *Methods in Cell Biology* (eds. Kepp, O. & Galluzzi, L.) vol. 164 137–156 (Academic Press, 2021).
153. Abdellatif, M. *et al.* Nicotinamide for the treatment of heart failure with preserved ejection fraction. *Sci. Transl. Med.* **13**, eabd7064 (2021).
154. Luo, W. & Brouwer, C. Pathview: an R/Bioconductor package for pathway-based data

integration and visualization. *Bioinformatics* **29**, 1830–1831 (2013).
155. Rio, D. C., Ares, M., Hannon, G. J. & Nilsen, T. W. Purification of RNA using TRIzol (TRI reagent). *Cold Spring Harb. Protoc.* **2010**, pdb.prot5439 (2010).
156. Xia, J. & Wishart, D. S. Metabolomic Data Processing, Analysis, and Interpretation Using MetaboAnalyst. *Curr. Protoc. Bioinforma.* **34**, 14.10.1-14.10.48 (2011).