RESEARCH

The poly-omics of ageing through individual-based metabolic modelling

Elisabeth Yaneske[*] and Claudio Angione

[*]Correspondence:
e.yaneske@tees.ac.uk
Department of Computer Science
and Information Systems, Teesside
University, Borough Road,
Middlesbrough, UK
Full list of author information is
available at the end of the article

Abstract

Background. Ageing can be classified in two different ways, chronological ageing and biological ageing. While chronological age is a measure of the time that has passed since birth, biological (also known as transcriptomic) ageing is defined by how time and the environment affect an individual in comparison to other individuals of the same chronological age. Recent research studies have shown that transcriptomic age is associated with certain genes, and that each of those genes has an effect size. Using these effect sizes we can calculate the transcriptomic age of an individual from their age-associated gene expression levels. The limitation of this approach is that it does not consider how these changes in gene expression affect the metabolism of individuals and hence their observable cellular phenotype.

Results. We propose a method based on poly-omic constraint-based models and machine learning in order to further the understanding of transcriptomic ageing. We use normalised CD4 T-cell gene expression data from peripheral blood mononuclear cells in 499 healthy individuals to create individual metabolic models. These models are then combined with a transcriptomic age predictor and chronological age to provide new insights into the differences between transcriptomic and chronological ageing. As a result, we propose a novel metabolic age predictor.

Conclusions. We show that our poly-omic predictors provide a more detailed analysis of transcriptomic ageing compared to gene-based approaches, and represent a basis for furthering our knowledge of the ageing mechanisms in human cells.

Keywords: Ageing; Biological age; Metabolic age; Metabolic modelling; Flux balance analysis; Poly-omics; Machine learning; CD4 T-cells

Introduction

Ageing is a complex process characterised by phenotypes such as greying hair and wrinkles, as well as age-associated diseases such as cancer, osteoarthritis and cardiovascular disease. Phenotypes of ageing and age-associated diseases can be linked to age-associated changes in metabolic subsystems [1–3]. Identifying these metabolic links has recently led to the discovery of age-associated biomarkers [4,5].

There are many different theories of the underlying mechanisms of ageing, including the mitochondrial theory of ageing, accumulation of metabolic by-products and dysregulation of regulatory pathways. The mitochondrion is the primary or-

ganelle responsible for metabolic cellular respiration; it takes in oxygen and nutrients and converts them into energy in the form of adenosine triphosphate (ATP). The mitochondrial theory of ageing states that oxidative damage caused by reactive oxygen species (ROS) produced by the mitochondria contributes to ageing by causing damage to mitochondrial DNA, lipids and proteins, which ultimately leads to cell death [6, 7]. Mitochondrial dysfunction and oxidative damage have been linked to age-associated neurodegenerative disorders such as Alzheimer's disease, Parkinson's disease and Huntington's disease [8, 9], as well as to the pathogenesis of cancer [10, 11].

Metabolism is increasingly being considered as a driver, rather than a marker, of the ageing process [12]. Three examples of metabolic by-products linked with ageing are amyloid proteins, advanced glycation end-products (AGEs) and lipofuscin. Accumulation of amyloid proteins in the central nervous system is associated with neurodegenerative disease in ageing [13,14]; for instance, β-amyloid plaques in brain tissue are linked with the pathogenesis of Alzheimer's disease [15,16]. AGEs can be ingested in foods or formed in the body by non-enzymatic glycation of lipids, nucleic acids and proteins, and their accumulation is thought to contribute to the ageing process [17,18]. AGEs are formed when foods are processed at high temperatures such as deep-frying, grilling and roasting. They can increase oxidative stress, upregulate inflammation, and form cross-links with proteins [19], which cause impaired elasticity to blood vessels, therefore leading to poor heart health [20,21]. Upregulation of inflammation caused by AGEs has been also linked to cancer [22–24]. Eating raw foods or foods cooked at lower temperatures can help to reduce dietary intake of AGEs [17]. Lipofuscin is a non-degradable metabolic by-product that builds up in lysosomes with time, and has been associated with age-related cellular degeneration [25], particularly macular degeneration [26].

One important example of the dysregulation of regulatory pathways as we age is chronic inflammation. The term 'inflammaging' was proposed by Franceschi et al. [27] to describe the imbalance between pro- and anti-inflammatory networks, which contributes to the chronic diseases of ageing. The function of the immune system declines as we age, leading to increased susceptibility to infectious diseases such as influenza [28], as well as decreased response to vaccinations against them [29]. This decline in function has been reported in CD4 T-cells [30,31], which are used in this study, along with changes in the ageing transcriptome [32].

Age can be defined as chronological or transcriptomic/biological. Chronological age is a measure of the time that has passed since our birth, whereas transcriptomic age represents the difference in how time and the environment have affected the cells and organs of our body as compared to others of the same chronological age. Our transcriptomic age can therefore be older or younger than our chronological age. Until now, transcriptomic age has been calculated for individuals using transcriptomic-only data [33].

The limitation of this approach is that it does not take into account how age-associated gene expression affects the metabolism within cells and thus their observable cellular phenotype. This paper aims to improve on the current understanding of ageing (based on transcriptomics data alone) by modelling how age-associated gene expression changes metabolic processes, therefore enabling the identification of metabolic age predictors, selected using machine learning techniques.

Metabolic models have proven to be valuable computational tools to study metabolism, as they allow predicting phenotypes from genotypes. By modelling most of the known biochemistry of a cell, they allow achieving a mechanistic understanding of the genotype-phenotype relationship. Coupled with tools for integration of omics data, metabolic models have been successfully exploited in a wide range of applications in health and disease, including personalised, condition- and tissue-specific cancer modelling [34–37].

Gene expression data and other types of omics-derived data can be used to constrain metabolic models for phenotype prediction [38]. The process of linking metabolic networks to phenotypes enables a better prediction of cellular phenotype compared to predictions from gene expression alone [39]. Exploiting this idea to generate a poly-omic model of ageing, we first generate individual-based genome-scale metabolic models and the associated fluxomic profiles. Specifically, we use CD4 T-cell transcriptomics data to modify a constraint-based metabolic model and achieve the predicted flux distributions (fluxomic profiles) for each individual in the cohort. Then, we adapt machine learning techniques in order to investigate metabolic changes linked to the chronological age of the individuals. We compare transcriptomic- and fluxomic-based clustering with chronological age and find that metabolic models are a better predictor of chronological age [40].

Our poly-omic pipeline also enables us to identify metabolic biomarkers of ageing, which are validated by recent literature, and to obtain metabolic age predictors. As a result, we build a metabolic age predictor capable of calculating the metabolic ages of individuals. Although a small number of metabolomics biomarkers have been proposed [41], to our knowledge this is the first time genome-scale predictors have been identified. We conclude that moving towards a poly-omic understanding of biological ageing can help provide a more accurate prediction of biological age, therefore leading to more targeted therapies for ageing individuals in a variety of environmental and physiological conditions.

Methods

The poly-omic ageing pipeline

Our pipeline starts from a meta-analysis of CD4 T-cell data containing the gene expression levels from human peripheral blood mononuclear cells and the chronological ages of 499 healthy individuals in the Boston area, comprised of 294 females and 205 males [42]. As the CD4 T-cell expression data was profiled on Affymetrix Human Gene 1.0 ST microarrays, it was first normalised using RMA (Robust Multi-array Average) [43]. In absence of a control profile, the gene expression values were then divided by the mean value for their associated probe. Using this normalised gene expression data, the transcriptomic ages for all 499 individuals were calculated as described in the next subsection (formulas (1) and (2)).

Having obtained the transcriptomic ages, we then used individuals' transcriptomic data to generate their personalised CD4 T-cell metabolic models. These were created using constraint based modelling of the CD4 T-cell [44] augmented with transcriptomics through GEMsplice [45], by setting individual constraints on the CD4 model (see the following subsections on constraint-based modelling for details on how the mapping was achieved). On the personalised models, we finally adapted

a set of statistical and machine learning methods based on clustering, PCA analysis and elastic net regression to identify metabolic predictors of ageing. Our pipeline thus enabled us to progress to a poly-omic understanding of ageing in human cells (Figure 1; see also the following subsections for details on the modelling approach adopted in this manuscript).

Transcriptomic age predictor

The transcriptomic age of an individual within a sample can be calculated by first obtaining their transcriptomic predictor, Z, using the gene expression levels of 1497 age-associated genes (i.e. those found to be differentially expressed with chronological age). This is achieved through a linear combination of the expression levels, where coefficients are their associated effect sizes [33]:

$$Z = \sum_i b_i x_i, \tag{1}$$

where x_i is the gene expression level of the ith probe, and b_i is the effect size for the ith probe. Effect sizes were associated with individual genes, whereas the original data contained gene expression data associated with probes. Therefore, where a probe was mapped to more than one age-associated gene, the effect sizes for those genes were averaged to give an overall average effect size b_i for that probe.

The transcriptomic predictor for each individual is then scaled using the mean and standard deviation of the chronological ages, and the mean and standard deviation of the transcriptomic predictors from all the individuals in the sample [33]. This allows defining the *transcriptomic age* of an individual:

$$SZ = \mu_{age} + (Z - \mu_Z)\frac{\sigma_{age}}{\sigma_Z}, \tag{2}$$

where μ_{age} and σ_{age} are the mean and the standard deviation of the chronological age across all the individuals within the sample, while μ_Z and σ_Z are the mean and the standard deviation of the predictor Z across all the individuals in the sample.

Constraint-based modelling to generate individual-based metabolic models

Metabolic models can be analysed using constraint-based modelling and flux balance analysis (FBA, the most widely-used technique to simulate metabolic models at steady state [46]), to enable predictions of the distribution of reaction flux rates in the cell. Given the matrix S of all known metabolic biochemical reactions and their stoichiometry, and given the vector v of reaction flux rates in a given growth or physiological condition, the steady-state condition is set by the constraint $Sv = 0$. Additional constraints are added on lower and upper bounds of v ($v^{\min}$ and $v^{\max}$). Constraints are included according to the growth or physiological condition that is simulated; these can also be set taking into account multiple omics data (e.g. transcriptomics data as used in our pipeline) [47]. Further constraints can include codon usage [48], splice-isoforms [45, 49], and can be analysed using pathway-oriented approaches [50, 51]. The metabolic network is then solved by maximising one or more cellular objectives (usually the biomass and energy-related or application-specific production of metabolites). For a comprehensive introduction to constraint-based

metabolic modelling and its poly-omic extensions, the reader is referred to the reviews by Palsson and Vijayakumar et al. [52, 53].

As omics data to constrain the model, here we use transcriptomic data from each individual to generate personalised metabolic models. Through GEMsplice [45], we modify the upper- and lower- limits of reactions as a function of the expression levels of the genes involved in the reaction. More specifically, for each individual, to predict the cellular flux distribution (fluxomic profile) when multiple objectives have to be taken into account, we use the following bilevel linear program:

$$
\begin{aligned}
& \max g^{\mathsf{T}} v \\
& \text{such that} \qquad \max f^{\mathsf{T}} v, \quad Sv = 0, \\
& \qquad v^{\min} \varphi(\Theta) \leq v \leq v^{\max} \varphi(\Theta).
\end{aligned}
\tag{3}
$$

The vectors f and g are weights to select (or combine) the objectives to be maximised from the vector v. The vector Θ represents the expression of a biochemical reaction, defined from the individual-based expression levels of its genes with a rule involving the max and min operators, depending on the type of enzyme (single gene, isozyme, or enzymatic complex). The function φ, which acts on Θ, converts the reaction expression values into coefficients for the bounds of reactions activated by those genes [54]. Here we set the primary objective f as biomass and the secondary objective g as ATP maintenance. Simulations were performed in Matlab.

Cluster analysis

Cluster analysis was used in order to group individual response according to both their transcriptomic and fluxomic profiles, and visualise them with chronological age. We compared both agglomerative hierarchical clustering (AHC) and k-means clustering using a novel application of the silhouette method. The silhouette method calculates a value which is a measure of the similarity of the values within a cluster (cohesion) and the dissimilarity of the values within that cluster to other clusters (separation). The silhouette calculation gives a value between -1 and 1. Silhouette values close to 1 are desirable as they indicate a cluster has high cohesion and high separation; if most values are close to 1 then the number of clusters is a good representation of the data.

Here we use the silhouette value to measure the cohesion and separation of the clustering of individuals by chronological age [55]. We define the silhouette value of an individual within a cluster as:

$$
s(c) = \frac{l(c) - a(c)}{max(a(c), l(c))},
\tag{4}
$$

where s is the silhouette value ($-1 \leqslant s(c) \leqslant 1$), c is the chronological age of the individual, a is the average dissimilarity of c to the other ages in the same cluster and l is the lowest average dissimilarity of c to any other age in a different cluster.

Our motivation for using the silhouette method was twofold. Firstly, we wanted to statistically compare the silhouette values for AHC and k-means to see which method performed better at clustering the data with chronological age. Secondly, we wanted to statistically compare whether transcriptomic-based or fluxomic-based clusters of individuals were consistent with chronological age.

Principal component analysis

Multidimensional data such as fluxomic datasets can be visualised using Principal Component Analysis (PCA). PCA can reduce multidimensional datasets to as few as two or three latent dimensions (components), which allows inference of variables causing the largest variations in the data. Here we use PCA to identify the fluxes accounting for the greatest variation between individuals in different age groups according to chronological age. In our PCA analysis, the fluxomic data was split according to three chronological age groups: 21 and under (112 individuals: 64 female and 48 male), 22 to 49 (360 individuals: 219 female and 141 male), and 50 and over (27 individuals: 11 female and 16 male). The analysis was performed in R and visualised using FactoMineR [56].

The CD4 T-cell model contains 4229 flux variables (metabolic reactions). PCA analysis gives the contribution of each variable (reaction) to the variability of each component. The proportion of variability accounted for by a component is defined numerically by its eigenvalue. The total contribution T_v of a given variable across the components can be calculated by determining its overall weighted sum as follows:

$$T_v = \sum_{i=1}^{n} E_i V_i, \tag{5}$$

where E_i is the eigenvalue for the principal component i, V_i is the variable (reaction) contribution to the principal component i, and n is the number of components chosen to represent the data.

Variables (reactions) can be mapped to a subsystem (metabolic pathway), which contains a number of reactions that are interlinked to perform a cellular metabolic function. The CD4 T-cell model contains 95 pathways, each of which corresponds to a number of reactions and their flux values. A flux value for each pathway was calculated as the mean of its reaction flux values, and PCA was also performed on the pathway flux rates obtained for each individual. Similarly, the total contribution T_s of a given pathway across the components can be found using:

$$T_s = \sum_{i=1}^{n} E_i S_i, \tag{6}$$

where E_i is the eigenvalue for the principal component i, S_i is the subsystem (pathway) contribution to the principal component i, and n is the number of components.

Elastic net regression

We use elastic net regression to identify metabolic predictors of chronological age and their effect sizes. Elastic net regression is a linear hybrid of the L_2 penalty of ridge regression [57] and the L_1 penalty of lasso regression [58]. For α between 0 and 1, where 0 is ridge regression and 1 is lasso regression, and a strictly non-negative λ, elastic net is defined as [59]:

$$(\hat{\beta}, \hat{\beta}_0) = \underset{\beta, \beta_0}{\operatorname{argmin}} \left(\frac{1}{2N} \sum_{i=1}^{N} (y_i - \beta_0 - x_i^T \beta)^2 + \lambda P_\alpha(\beta) \right), \tag{7}$$

where

$$P_\alpha(\beta) = \frac{(1-\alpha)}{2}\|\beta\|_2^2 + \alpha\|\beta\|_1 = \sum_{j=1}^{p}\left(\frac{(1-\alpha)}{2}\beta_j^2 + \alpha|\beta_j|\right), \tag{8}$$

N is the number of individuals, y_i is the chronological age of individual i, x_i is a $p \times 1$ vector of p metabolic pathway fluxes at individual i, α is set to 0.5 to achieve a balance between L_1 and L_2 norms, λ is a positive regularization parameter, β_0 is a scalar parameter, and β is a $p \times 1$ vector of effect sizes on chronological age (regression coefficients), where p is the number of metabolic pathways. We also performed an equivalent analysis directly on reaction fluxes (Additional file 3).

Elastic net overcomes some of the limitations of using the lasso method alone [59]. When analysing high dimensional data, such as the individual by reaction data (499×4229), where the number of predictors is greater than the number of observations, the lasso method can only select at most the same number of variables as observations. However, omic data tends by its nature to often be highly correlated, which means there is high correlation between regression predictors. Where there is a group of highly correlated predictors, the lasso method will only select one variable from the group. The advantage of the elastic net method in our context is that while the L_1 part of the penalty generates a sparse model, the quadratic part of the penalty (L_2 regularisation $\|\beta\|^2$), taken from ridge regression, allows the number of selected variables to be greater than the number of observations, and allows groups of strongly correlated variables to be selected.

Results and Discussion

Clustering shows best dataset for prediction of chronological age

From the plot of average age-based silhouette values (Figure 2A), optimal cluster numbers were chosen from the point closest to 1 at which there is an 'elbow bend' in the curve, indicating a drop in the amount of variance explained by the clusters after this point [60]. For the transcriptomic data seven clusters were chosen, while for fluxomic data six clusters were chosen. The pairwise distance of the chosen cluster numbers and the results of clustering for both transcriptomic and fluxomic data with chronological and transcriptomic age are shown in Figures 2E, 2C, 2D, and 2B respectively. We selected k-means as a clustering algorithm because it performed consistently better than hierarchical clustering both in the transcriptomic (Kolmogorov-Smirnov test statistic = 0.66, p-value = $2.91 \cdot 10^{-6}$) and in the fluxomic-based clustering (Kolmogorov-Smirnov test statistic = 0.8, p-value = $5.59 \cdot 10^{-9}$).

Remarkably, with both methods, and considering a variable number of clusters between 2 and 30 (Figure 2A and Additional file 1), fluxomic-based clustering consistently outperformed transcriptomic-based clustering in terms of age-based average silhouette values (Kolmogorov-Smirnov test statistic = 0.38, p-value = 0.022 for k-means; Kolmogorov-Smirnov test statistic = 0.62, p-value = $1.15 \cdot 10^{-5}$ for hierarchical). We also analysed the pattern of silhouette values based on deviations from linearity. In general, we found that drops in silhouette values corresponded to smaller clusters being merged into much larger, less distinct clusters, or cluster boundaries changing such that there was a loss of intra-cluster cohesion and

inter-cluster separation (more details can be found in Additional file 5). Our results therefore suggest that, compared to gene expression values, individual-based poly-omic models and their predicted flux rates are a better predictor of chronological age.

Principal component analysis identifies predictors of ageing

The number of components to retain in the analysis was determined by their eigen-values and the total contribution to variance explained by the components. Three criteria were applied to the data: (i) according to Kaiser's criterion [61] only those components with eigenvalues greater than 1 should be retained; (ii) the overall con-tribution to variance of the retained components should be 50% or greater; and (iii) for an $n \times m$ matrix, if the data were randomly distributed the expected contribu-tion to variance of the eigenvalue for each axis would be $100/(n-1)$ % in terms of rows [62]. Therefore, any axis with a contribution to variance larger than this proportion should be retained as "significant". The threshold percentage variable contributions to variance for each group used in the analysis is shown in Table 1. The resulting number of components retained for analysis for each group is shown in Table 2.

Using only the significant components, the overall contribution of each reaction to component variability was calculated using (5) for each of the three age groups, 21- (21 and under), 22-49 and 50+ (50 and over). The contributions of each of the reactions to the different age groups were then compared by calculating the difference between them. The differences between the reaction contributions for (i) 21- and 22-49, (ii) 21- and 50+, and (iii) 22-49 and 50+ were obtained in order to determine the reactions that vary the most with age (see Additional file 2).

In the results of the analysis of the differences in overall contribution of the 95 pathways, four pathways appeared in the top 20 of all the age group compar-isons: CoA synthesis, vitamin D metabolism, hyaluronan metabolism and pyruvate metabolism. All four of these pathways also decreased in their contribution with age. Pyruvate and CoA are both key components of the citric acid cycle, which is es-sential in energy production in the mitochondria. Reduced stamina observed in the ageing population is thought to be related to impairment of mitochondrial energy production [63, 64]. Hypovitaminosis D in the ageing population is a major cause of impaired bone formation and mineralisation (osteoporosis) [65, 66]. Hyaluronan or Hyaluronic acid (HA) has a high capacity to bind and retain water molecules and is found in high levels in the extracellular matrix of skin where it regulates skin moisture. Reduced levels of HA are associated with the loss of moisture in ageing skin [67, 68]. Changes in HA size contribute to age-related impairment of wound healing in skin [69, 70] and to the viscoelasticity of synovial fluid, which can contribute to osteoarthritis, a common disease of ageing [71–73].

In the 21- to 22-49 group, the overall difference in contribution values for all but one of the top 20 pathways increased from the 21- group to the 22-49 group. Inter-estingly, CoA synthesis increased but CoA catabolism decreased, suggesting higher CoA levels due to increased synthesis and decreased degradation. Conversely, for the 21- group and 50+ group, the overall contribution values for all but one of the top 20 pathways decreased. Only Vitamin B2 metabolism increased. This is consis-tent with recent studies showing that the activity of vitamin B2 metabolism does

not decrease until after the age of 50 [74]. All of the top 20 pathways contribution values decreased for the 50+ group compared to the 22-49 group.

Squalene and cholesterol synthesis, vitamin A metabolism and glycine, serine, alanine and threonine metabolism overall contributions all decrease with age. Sebum, produced by the sebaceous glands, contains both cholesterol and squalene. Squalene is correlated with α-tocopherol (vitamin E) levels on the surface of the skin. α-tocopherol is the main antioxidant on the skin [75] and its decrease with ageing may contribute towards the signs of ageing skin [76, 77].

The reactions that make up the squalene and cholesterol synthesis pathway are part of the mevalonate pathway, which produces the precursors of all the steroid hormones, heme cholesterol, coenzyme Q_{10}, and vitamin K [78]. The decrease in plasma levels of high density lipoprotein (HDL) cholesterol is correlated with a higher risk of atherosclerosis [79]. Low vitamin K has been associated with osteoarthritis and impaired cognitive function in older adults [80, 81]. The synthesis of mitochondrial coenzyme Q_{10} can decrease with age [82]; this constitutes an important factor for health, as coenzyme Q_{10} is an antioxidant that protects against diseases that involve oxidative stress, such as cardiovascular and neurogenerative diseases [83–88]. Furthermore, the synthesis of heme, the major functional form of iron in the body, decreases with age [89, 90] and has been linked with neurodegenerative disorders such as Alzheimer's disease [91]. Steroid hormones are also known to decline with age [92, 93]; decreasing sex steroid hormone deficiency in oestrogens and androgens contributes to ageing skin [94, 95] and increased risk of cardiovascular disease in women [93, 96].

Glycine, serine, alanine and threonine are all non-essential amino acids. Decreasing glycine levels have been linked with age-associated oxidative stress [97], while age-associated decrease in serine metabolism has been linked with impaired memory function in the brain [98, 99]. Alanine metabolism is associated with the liver, and alanine transaminase has been suggested as a biomarker for ageing [100]. Vitamin A cannot be produced by the body and is therefore obtained through diet; a decrease in the active form of vitamin A in the body, retinol, is thought to be linked with age-associated slowing in visual dark-adaption [101, 102].

Figures 3A-C show the factor maps for the top 20 pathways contributing to variance for each age group. For the 21- group, the overall contribution of components 1 and 2 to variation is 15.2%. For the 22-49 group, it is 12.4% and for the 50+ group it is 24%. If the variance was evenly distributed across all pathways, then the expected average variation would be $100/95 \% = 1.05\%$. Interestingly, the variance explained by the first two components in the 50+ group is higher compared to the other two groups, suggesting that less latent pathways, but with an increasingly strong role, characterise the ageing phenotype.

As a result, the biplot (Figure 4D) of individuals grouped by age and the top 10 pathways contributing to overall variance of components 1 and 2 shows differentiation between the three age groups, with the biggest differentiation shown in the axis of the 50+ group compared to the other two. The 50+ age group appears to lie along the glyoxylate and dicarboxylate metabolism axis. To identify the amount of intercorrelation in the pathway flux data, correlation plots were created for the pathways with the top 20 overall difference in contribution in each group, 21- and

22-49, 21- and 50+, and 22-49 and 50+ (Figures 4A-C). The plots show the large amount of intercorrelation between pathways.

Exchange/demand reaction and oxidative phosphorylation remain the high contributors to variance for all three age groups. The most notable rise in contribution to variation with age is vitamin C metabolism, which moves from rank 18 in the 21- to rank 2 in the 50+ group. Changes in oxidative stress are known to affect the levels of vitamin C in the body [103]. Furthermore, oxidative damage to the genome caused by ROS (reactive oxygen species) is thought to be one of the causes of ageing [104]. Interestingly, ROS detoxification is one of the top 10 contributors in the 21- and 22-49 groups but does not appear in the top 20 contributors in the 50+ group, suggesting a possible link with ageing, oxidative stress and the changing variation contributed by vitamin C metabolism. Glyoxylate and dicarboxylate metabolism also has a large rise in contribution to variance with age. Glyoxylate and dicarboxylate metabolism is strictly linked with glycine, serine and threonine metabolism, pyruvate metabolism, ascorbate metabolism (a mineral salt of vitamin C), all of which our results have identified as linked to ageing.

Elastic net regression identifies metabolic-age predictors

The analysis was performed using tenfold cross validation for both metabolic reactions and metabolic pathways. The results of the metabolic pathways analysis are reported here, while the results for the metabolic reactions can be found in Additional file 3. Elastic net regression of the 95 metabolic pathways returned 100 possible λ values with their associated effect sizes. The value $\lambda = 1.57$ was chosen as this had the lowest mean squared error value. This gave metabolic effect sizes for three pathways: butanoate metabolism, pyrimidine synthesis and beta-alanine metabolism (Table 3).

Both butanoate and beta-alanine metabolism have been linked to sarcopenia, deterioration of skeletal muscle, with age [105]. Butanoate and beta-alanine metabolism are facilitated by the enzyme aldehyde dehydrogenase. One of the substrates of this enzyme is nicotinamide adenine dinucleotide (NAD). Reversal of NAD loss as we age is currently undergoing human trials following successful age reversal of mitochondrial function in the skeletal muscle cells of mice [106]. Mitochondrial dysfunction, which is linked to ageing, also leads to a reduction of pyrimidine synthesis [107, 108].

In order to calculate the biological/metabolic ages of the individuals, equations (1) and (2) were modified to use the metabolic effect sizes. We therefore define the following metabolic predictor:

$$M = \sum_i b_i f_i, \tag{9}$$

where M is the metabolic predictor, f_i is the flux value of the ith pathway, and b_i is the effect size for the ith pathway. The *metabolic age* was then defined as:

$$SM = \mu_{age} + (M - \mu_M)\frac{\sigma_{age}}{\sigma_M}, \tag{10}$$

where μ_{age} and σ_{age} are the mean and the standard deviation of the chronological age, while μ_M and σ_M are the mean and the standard deviation of the metabolic predictor M.

The metabolic ages showed correlation (p-value $= 4.7 \cdot 10^{-4}$) with chronological age. Both the metabolic ages and the chronological ages of all 499 individuals can be found in Additional file 3, which also includes the results of the regression performed directly on reactions. This is a promising starting point for a metabolic age predictor, which can be further refined using more samples to improve predictor accuracy.

Conclusion

While chronological age gives an accurate measurement of the time since an individual's birth, biological age – also called transcriptomic age – gives a more accurate representation of the relative health of an individual compared to others of the same age. Where an individual has a biological age greater than their chronological age, they are ageing more quickly than their peers, and therefore have decreased life expectancy. Finding predictors of biological ageing and measuring their presentation in individuals can allow targeted and personalised interventions, both medication and lifestyle-based, to improve health and life expectancy.

A number of age-associated genes along with their effect sizes have previously been identified in the literature and used to calculate the biological age of individuals [33]. The limitation of this approach is that it does not take into account the metabolic effects of those gene expression values on the cellular phenotype. For example, a gene that has been found to be differentially expressed with age may have little or no effect on cellular metabolism. Here we have achieved the first steps towards a metabolic age predictor to overcome the limitations of previous transcriptomic-only approaches.

Using a genome-scale metabolic model of CD4 T-cells combined with transcriptomic data, we were able to obtain individual-specific metabolic models and generate fluxomic data. The subsequent stage of our analysis used a novel application of the silhouette method and clustering techniques, from which we identified that fluxomic data clusters better with chronological age, therefore suggesting that metabolic models are a better predictor of the chronological age of an individual. Applying PCA analysis and elastic net regression enabled us to identify potential metabolic predictors of ageing. Many of these predictors have also been identified in the literature as linked to the ageing process, validating the reliability of the method. Finally, elastic net regression produced metabolic age predictors and their effect sizes, from which we calculated the metabolic age of each individual.

Our next step will be to further refine the metabolic age predictors obtained from elastic net regression with more data from individuals across the different age ranges. Future studies could also use other age predictors, such as epigenetic clocks [109] and telomere length [110] as variables to investigate how they correlate with our metabolic age predictor. Furthermore, although proteomics data has some limitations, including providing less coverage than transcriptomic data [111, 112], it can be included as a complementary method to improve the accuracy of the results. We will also investigate whether there are differences in the metabolic age

predictors based on different phenotypes/attributes of individuals such as gender. This will allow further refinement of predictors based on individual data, and will suggest more personalised interventions to reduce metabolic age and improve life expectancy.

Competing interests
The authors declare that they have no competing interests.

Author's contributions
CA conceived and coordinated the study. EY and CA developed the methods. EY wrote the code and performed the analysis. EY and CA wrote the manuscript. Both authors read and approved the final manuscript.

Availability
All flux data and results are available as additional files. Source code and normalised transcriptomic data is available on request for academic use.

References
1. Barzilai, N., Huffman, D.M., Muzumdar, R.H., Bartke, A.: The critical role of metabolic pathways in aging. Diabetes **61**(6), 1315–1322 (2012)
2. Newgard, C.B., Pessin, J.E.: Recent progress in metabolic signaling pathways regulating aging and life span. Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences **69**(Suppl_1), 21–27 (2014)
3. Wallace, D.C.: A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. Annu. Rev. Genet. **39**, 359–407 (2005)
4. Amadoz, A., Sebastian-Leon, P., Vidal, E., Salavert, F., Dopazo, J.: Using activation status of signaling pathways as mechanism-based biomarkers to predict drug sensitivity. Scientific reports **5**, 18494 (2015)
5. Houtkooper, R.H., Argmann, C., Houten, S.M., Cantó, C., Jeninga, E.H., Andreux, P.A., Thomas, C., Doenlen, R., Schoonjans, K., Auwerx, J.: The metabolic footprint of aging in mice. Scientific reports **1**, 134 (2011)
6. Miquel, J., Economos, A., Fleming, J., Johnson, J.: Mitochondrial role in cell aging. Experimental gerontology **15**(6), 575–591 (1980)
7. Kauppila, T.E., Kauppila, J.H., Larsson, N.-G.: Mammalian mitochondria and aging: an update. Cell metabolism **25**(1), 57–71 (2017)
8. Lin, M.T., Beal, M.F.: Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. nature **443**(7113), 787 (2006)
9. Langley, M., Ghosh, A., Charli, A., Sarkar, S., Ay, M., Luo, J., Zielonka, J., Brenza, T., Bennett, B., Jin, H., et al.: Mito-apocynin prevents mitochondrial dysfunction, microglial activation, oxidative damage, and progressive neurodegeneration in mitopark transgenic mice. Antioxidants & Redox Signaling (2017)
10. Fulda, S., Galluzzi, L., Kroemer, G.: Targeting mitochondria for cancer therapy. Nature reviews. Drug discovery **9**(6), 447 (2010)
11. Green, D.R.: Cancer and apoptosis: Who is built to last? Cancer cell **31**(1), 2–4 (2017)
12. Feng, Z., Hanson, R.W., Berger, N.A., Trubitsyn, A.: Reprogramming of energy metabolism as a driver of aging. Oncotarget **7**(13), 15410 (2016)
13. Pike, C.J., Walencewicz, A.J., Glabe, C.G., Cotman, C.W.: In vitro aging of ß-amyloid protein causes peptide aggregation and neurotoxicity. Brain research **563**(1), 311–314 (1991)
14. Jack, C.R., Wiste, H.J., Weigand, S.D., Therneau, T.M., Knopman, D.S., Lowe, V., Vemuri, P., Mielke, M.M., Roberts, R.O., Machulda, M.M., et al.: Age-specific and sex-specific prevalence of cerebral β-amyloidosis, tauopathy, and neurodegeneration in cognitively unimpaired individuals aged 50–95 years: a cross-sectional study. The Lancet Neurology **16**(6), 435–444 (2017)
15. Hardy, J., Selkoe, D.J.: The amyloid hypothesis of alzheimer's disease: progress and problems on the road to therapeutics. science **297**(5580), 353–356 (2002)
16. Vemuri, P., Knopman, D.S., Lesnick, T.G., Przybelski, S.A., Mielke, M.M., Graff-Radford, J., Murray, M.E., Roberts, R.O., Vassilaki, M., Lowe, V.J., et al.: Evaluation of amyloid protective factors and alzheimer disease neurodegeneration protective factors in elderly individuals. JAMA neurology (2017)
17. Semba, R.D., Nicklett, E.J., Ferrucci, L.: Does accumulation of advanced glycation end products contribute to the aging phenotype? Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences **65**(9), 963–975 (2010)
18. Yu, T., Lin, L.: Advanced glycation end products accumulated with aging inhibit colonic smooth muscle contraction by upregulation of bk channel î'1-subunit. Gastroenterology **152**(5), 911 (2017)
19. Uribarri, J., Cai, W., Peppa, M., Goodman, S., Ferrucci, L., Striker, G., Vlassara, H.: Circulating glycotoxins and dietary advanced glycation endproducts: two links to inflammatory response, oxidative stress, and aging. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences **62**(4), 427–433 (2007)
20. Aronson, D.: Cross-linking of glycated collagen in the pathogenesis of arterial and myocardial stiffening of aging and diabetes. Journal of hypertension **21**(1), 3–12 (2003)
21. Wannamethee, S.G., Welsh, P., Papacosta, O., Ellins, E.A., Halcox, J.P., Whincup, P.H., Sattar, N.: Circulating soluble receptor for advanced glycation end product: Cross-sectional associations with cardiac markers and subclinical vascular disease in older men with and without diabetes. Atherosclerosis **264**, 36–43 (2017)
22. Riehl, A., Németh, J., Angel, P., Hess, J.: The receptor rage: Bridging inflammation and cancer. Cell Communication and Signaling **7**(1), 12 (2009)

23. Sims, G.P., Rowe, D.C., Rietdijk, S.T., Herbst, R., Coyle, A.J.: Hmgb1 and rage in inflammation and cancer. Annual review of immunology **28**, 367–388 (2009)

24. Turner, D.: Chapter one-the role of advanced glycation end-products in cancer disparity. Advances in cancer research **133**, 1–22 (2017)

25. Brunk, U.T., Terman, A.: Lipofuscin: mechanisms of age-related accumulation and influence on cell function. Free Radical Biology and Medicine **33**(5), 611–619 (2002)

26. Hyttinen, J.M., Błasiak, J., Niittykoski, M., Kinnunen, K., Kauppinen, A., Salminen, A., Kaarniranta, K.: Dna damage response and autophagy in the degeneration of retinal pigment epithelial cells–implications for age-related macular degeneration (amd). Ageing Research Reviews (2017)

27. Franceschi, C., Capri, M., Monti, D., Giunta, S., Olivieri, F., Sevini, F., Panourgia, M.P., Invidia, L., Celani, L., Scurti, M., et al.: Inflammaging and anti-inflammaging: a systemic perspective on aging and longevity emerged from studies in humans. Mechanisms of ageing and development **128**(1), 92–105 (2007)

28. Blomberg, B.B., Frasca, D.: Quantity, not quality, of antibody response decreased in the elderly. The Journal of clinical investigation **121**(8), 2981–2983 (2011)

29. Targonski, P.V., Jacobson, R.M., Poland, G.A.: Immunosenescence: role and measurement in influenza vaccine response among the elderly. Vaccine **25**(16), 3066–3069 (2007)

30. Marco, M.-G., Rebeca, A.-A.: When aging reaches CD4+ T-Cells: Phenotypic and functional changes. Frontiers in Immunology **4**, 107 (2013)

31. Lefebvre, J.S., Haynes, L.: Aging of the CD4 T cell compartment. Open Longevity Science **6**, 83 (2012)

32. Zhao, M., Qin, J., Yin, H., Tan, Y., Liao, W., Liu, Q., Luo, S., He, M., Liang, G., Shi, Y., et al.: Distinct epigenomes in CD4+ T cells of newborns, middle-ages and centenarians. Scientific Reports **6**, 38411 (2016)

33. Peters, M.J., Joehanes, R., Pilling, L.C., Schurmann, C., Conneely, K.N., Powell, J., Reinmaa, E., Sutphin, G.L., Zhernakova, A., Schramm, K., et al.: The transcriptional landscape of age in human peripheral blood. Nature communications **6** (2015)

34. Cook, D.J., Nielsen, J.: Genome-scale metabolic models applied to human health and disease. Wiley Interdisciplinary Reviews: Systems Biology and Medicine **9**(6) (2017)

35. Eyassu, F., Angione, C.: Modelling pyruvate dehydrogenase under hypoxia and its role in cancer metabolism. Royal Society open science **4**(10), 170360 (2017)

36. Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhori, G., Benfeitas, R., Arif, M., Liu, Z., Edfors, F., et al.: A pathology atlas of the human cancer transcriptome. Science **357**(6352), 2507 (2017)

37. Zielinski, D.C., Jamshidi, N., Corbett, A.J., Bordbar, A., Thomas, A., Palsson, B.O.: Systems biology analysis of drivers underlying hallmarks of cancer cell metabolism. Scientific reports **7**, 41241 (2017)

38. Angione, C., Conway, M., Lió, P.: Multiplex methods provide effective integration of multi-omic data in genome-scale models. BMC bioinformatics **17**(4), 83 (2016)

39. Samal, S.S., Radulescu, O., Weber, A., Fröhlich, H.: Linking metabolic network features to phenotypes using sparse group lasso. Bioinformatics (2017)

40. Yaneske, E., Angione, C.: A data-and model-driven analysis reveals the multi-omic landscape of ageing. In: International Conference on Bioinformatics and Biomedical Engineering, pp. 145–154 (2017). Springer

41. Jylhävä, J., Pedersen, N.L., Hägg, S.: Biological age predictors. EBioMedicine (2017)

42. Raj, T., Rothamel, K., Mostafavi, S., Ye, C., Lee, M.N., Replogle, J.M., Feng, T., Lee, M., Asinovski, N., Frohlich, I., et al.: Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. Science **344**(6183), 519–523 (2014)

43. Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., Speed, T.P.: Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics **4**(2), 249–264 (2003)

44. Han, F., Li, G., Dai, S., Huang, J.: Genome-wide metabolic model to improve understanding of CD4+ T cell metabolism, immunometabolism and application in drug design. Molecular BioSystems **12**(2), 431–443 (2016)

45. Angione, C.: Integrating splice-isoform expression into genome-scale models characterizes breast cancer metabolism. Bioinformatics, 562 (2017)

46. Orth, J.D., Thiele, I., Palsson, B.Ø.: What is flux balance analysis? Nature biotechnology **28**(3), 245–248 (2010)

47. Opdam, S., Richelle, A., Kellman, B., Li, S., Zielinski, D.C., Lewis, N.E.: A systematic evaluation of methods for tailoring genome-scale metabolic models. Cell Systems **4**(3), 318–329 (2017)

48. Kashaf, S.S., Angione, C., Lió, P.: Making life difficult for clostridium difficile: augmenting the pathogen's metabolic model with transcriptomic and codon usage data for better therapeutic target characterization. BMC systems biology **11**(1), 25 (2017)

49. Ryu, J.Y., Kim, H.U., Lee, S.Y.: Framework and resource for more than 11,000 gene-transcript-protein-reaction associations in human metabolism. Proceedings of the National Academy of Sciences, 201713050 (2017)

50. O'Brien, E.J., Monk, J.M., Palsson, B.O.: Using genome-scale models to predict biological capabilities. Cell **161**(5), 971–987 (2015)

51. Angione, C., Pratanwanich, N., Lió, P.: A hybrid of metabolic flux analysis and bayesian factor modeling for multiomic temporal pathway activation. ACS synthetic biology **4**(8), 880–889 (2015)

52. Palsson, B..: Systems Biology: Constraint-based Reconstruction and Analysis, 2nd edn. Cambridge University Press, Cambridge (2015)

53. Vijayakumar, S., Conway, M., Lió, P., Angione, C.: Seeing the wood for the trees: a forest of methods for optimization and omic-network integration in metabolic modelling. Briefings in Bioinformatics (2017)

54. Angione, C., Lió, P.: Predictive analytics of environmental adaptability in multi-omic network models. Scientific reports **5** (2015)

55. Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics **20**, 53–65 (1987)

56. Lê, S., Josse, J., Husson, F., et al.: FactoMineR: an R package for multivariate analysis. Journal of statistical

software **25**(1), 1–18 (2008)

57. Hoerl, A., Kennard, R.: Ridge regression, in 'Encyclopedia of Statistical Sciences', Vol. 8. Wiley, New York (1988)

58. Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 267–288 (1996)

59. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **67**(2), 301–320 (2005)

60. Madhulatha, T.S.: An overview on clustering methods. arXiv preprint arXiv:1205.1117 (2012)

61. Kaiser, H.F.: The application of electronic computers to factor analysis. Educational and psychological measurement **20**(1), 141–151 (1960)

62. Bendixen, M.T.: Compositional perceptual mapping using chi-squared trees analysis and correspondence analysis. Journal of Marketing Management **11**(6), 571–581 (1995)

63. Yarian, C.S., Toroser, D., Sohal, R.S.: Aconitase is the main functional target of aging in the citric acid cycle of kidney mitochondria from mice. Mechanisms of ageing and development **127**(1), 79–84 (2006)

64. Tepp, K., Puurand, M., Timohhina, N., Adamson, J., Klepinin, A., Truu, L., Shevchuk, I., Chekulayev, V., Kaambre, T.: Changes in the mitochondrial function and in the efficiency of energy transfer pathways during cardiomyocyte aging. Molecular and Cellular Biochemistry, 1–18 (2017)

65. Lanske, B., Razzaque, M.S.: Vitamin D and aging: old concepts and new insights. The Journal of nutritional biochemistry **18**(12), 771–777 (2007)

66. Hin, H., Tomson, J., Newman, C., Kurien, R., Lay, M., Cox, J., Sayer, J., Hill, M., Emberson, J., Armitage, J., et al.: Optimum dose of vitamin D for disease prevention in older people: BEST-D trial of vitamin D in primary care. Osteoporosis International **28**(3), 841–851 (2017)

67. Papakonstantinou, E., Roth, M., Karakiulakis, G.: Hyaluronic acid: A key molecule in skin aging. Dermato-endocrinology **4**(3), 253–258 (2012)

68. Li, W.-H., Wong, H.-K., Serrano, J., Randhawa, M., Kaur, S., Southall, M.D., Parsa, R.: Topical stabilized retinol treatment induces the expression of HAS genes and HA production in human skin in vitro and in vivo. Archives of dermatological research **309**(4), 275–283 (2017)

69. Reed, M.J., Damodarasamy, M., Chan, C.K., Johnson, M.N., Wight, T.N., Vernon, R.B.: Cleavage of hyaluronan is impaired in aged dermal wounds. Matrix Biology **32**(1), 45–51 (2013)

70. Damodarasamy, M., Johnson, R.S., Bentov, I., MacCoss, M.J., Vernon, R.B., Reed, M.J.: Hyaluronan enhances wound repair and increases collagen iii in aged dermal wounds. Wound repair and regeneration **22**(4), 521–526 (2014)

71. Arrich, J., Piribauer, F., Mad, P., Schmid, D., Klaushofer, K., Müllner, M.: Intra-articular hyaluronic acid for the treatment of osteoarthritis of the knee: systematic review and meta-analysis. Canadian Medical Association Journal **172**(8), 1039–1043 (2005)

72. Moreland, L.W.: Intra-articular hyaluronan (hyaluronic acid) and hylans for the treatment of osteoarthritis: mechanisms of action. Arthritis Res Ther **5**(2), 54 (2003)

73. Martel-Pelletier, J., Raynauld, J.-P., Mineau, F., Abram, F., Paiement, P., Delorme, P., Pelletier, J.-P.: Levels of serum biomarkers from a two-year multicentre trial are associated with treatment response on knee osteoarthritis cartilage loss as assessed by magnetic resonance imaging: an exploratory study. Arthritis research & therapy **19**(1), 169 (2017)

74. Russell, R.M.: The aging process as a modifier of metabolism. The American journal of clinical nutrition **72**(2), 529–532 (2000)

75. Thiele, J., Schroeter, C., Hsieh, S., Podda, M., Packer, L.: The antioxidant network of the stratum corneum. In: Oxidants and Antioxidants in Cutaneous Biology vol. 29, pp. 26–42 (2001). Karger Publishers

76. Miettinen, H.E., Rönö, K., Koivusalo, S., Stach-Lempinen, B., Pöyhönen-Alho, M., Eriksson, J.G., Hiltunen, T.P., Gylling, H.: Elevated serum squalene and cholesterol synthesis markers in pregnant obese women with gestational diabetes mellitus. Journal of lipid research **55**(12), 2644–2654 (2014)

77. Iizaka, S., Nagata, S., Sanada, H.: Nutritional status and habitual dietary intake are associated with frail skin conditions in community-dwelling older people. The journal of nutrition, health & aging **21**(2), 137–146 (2017)

78. Holstein, S.A., Hohl, R.J.: Isoprenoids: remarkable diversity of form and function. Lipids **39**(4), 293–309 (2004)

79. Steffen, B.T., Bielinski, S.J., Decker, P.A., Berardi, C., Larson, N.B., Pankow, J.S., Michos, E.D., Hanson, N.Q., Herrington, D.M., Tsai, M.Y.: Low HDL cholesterol and particle concentrations are associated with greater levels of endothelial activation markers in multi-ethnic study of atherosclerosis participants. Journal of Clinical Lipidology (2017)

80. Shea, M.K., Kritchevsky, S.B., Hsu, F.-C., Nevitt, M., Booth, S.L., Kwoh, C.K., McAlindon, T.E., Vermeer, C., Drummen, N., Harris, T.B., et al.: The association between vitamin K status and knee osteoarthritis features in older adults: The health, aging and body composition study. Osteoarthritis and cartilage **23**(3), 370–378 (2015)

81. Presse, N., Belleville, S., Gaudreau, P., Greenwood, C.E., Kergoat, M.-J., Morais, J.A., Payette, H., Shatenstein, B., Ferland, G.: Vitamin K status and cognitive function in healthy older adults. Neurobiology of aging **34**(12), 2777–2783 (2013)

82. Littarru, G.P., Langsjoen, P.: Coenzyme Q10 and statins: biochemical and clinical implications. Mitochondrion **7**, 168–174 (2007)

83. Rosenfeldt, F.L., Pepe, S., Linnane, A., Nagley, P., Rowland, M., Ou, R., Marasco, S., Lyon, W., Esmore, D.: Coenzyme Q10 protects the aging heart against stress. Annals of the New York Academy of Sciences **959**(1), 355–359 (2002)

84. Navas, P., Villalba, J.M., de Cabo, R.: The importance of plasma membrane coenzyme Q in aging and stress responses. Mitochondrion **7**, 34–40 (2007)

85. Ochoa, J.J., Quiles, J.L., Huertas, J.R., Mataix, J.: Coenzyme Q10 protects from aging-related oxidative

stress and improves mitochondrial function in heart of rats fed a polyunsaturated fatty acid (pufa)-rich diet. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences **60**(8), 970–975 (2005)

86. Mancuso, M., Orsucci, D., Volpi, L., Calsolaro, V., Siciliano, G.: Coenzyme Q10 in neuromuscular and neurodegenerative disorders. Current drug targets **11**(1), 111–121 (2010)

87. Luo, M., Yang, X., Hu, J., Ruan, X., Mu, F., Fu, Y.: The synthesis of coenzyme Q10. Current Organic Chemistry **21**(6), 489–502 (2017)

88. Paroha, S., Chandel, A.K.S., Dubey, R.D.: Nanosystems for drug delivery of coenzyme Q10. Environmental Chemistry Letters, 1–7 (2017)

89. Atamna, H., Liu, J., Ames, B.N.: Heme deficiency selectively interrupts assembly of mitochondrial complex IV in human fibroblasts relevance to aging. Journal of Biological Chemistry **276**(51), 48410–48416 (2001)

90. Pabis, K., ScheiberMojdehkar, B., Valencak, T., Nowikovsky, K.: Altered iron homeostasis in mouse models of aging. Experimental Gerontology **94**, 118 (2017)

91. Sung, H.Y., Choi, B.-O., Jeong, J.H., Kong, K.A., Hwang, J., Ahn, J.-H.: Amyloid beta-mediated hypomethylation of heme oxygenase 1 correlates with cognitive impairment in alzheimers disease. PloS one **11**(4), 0153156 (2016)

92. Walther, A., Philipp, M., Lozza, N., Ehlert, U.: The rate of change in declining steroid hormones: a new parameter of healthy aging in men? Oncotarget **7**(38), 60844 (2016)

93. Knowlton, A., Lee, A.: Estrogen and the cardiovascular system. Pharmacology & therapeutics **135**(1), 54–70 (2012)

94. Wilkinson, H.N., Hardman, M.J.: The role of estrogen in cutaneous ageing and repair. Maturitas **103**, 60–64 (2017)

95. Luebberding, S., Krueger, N., Kerscher, M.: Age-related changes in male skin: quantitative evaluation of one hundred and fifty male subjects. Skin pharmacology and physiology **27**(1), 9–17 (2014)

96. Sun, C., Simon, S.I., Foster, G.A., Radecke, C.E., Hwang, H.V., Zhang, X., Hammock, B.D., Chiamvimonvat, N., Knowlton, A.A.: 11, 12-epoxyecosatrienoic acids mitigate endothelial dysfunction associated with estrogen loss and aging: Role of membrane depolarization. Journal of molecular and cellular cardiology **94**, 180–188 (2016)

97. Sekhar, R.V., Patel, S.G., Guthikonda, A.P., Reid, M., Balasubramanyam, A., Taffet, G.E., Jahoor, F.: Deficient synthesis of glutathione underlies oxidative stress in aging and can be corrected by dietary cysteine and glycine supplementation. The American journal of clinical nutrition **94**(3), 847–853 (2011)

98. Billard, J.-M.: d-serine in the aging hippocampus. Journal of pharmaceutical and biomedical analysis **116**, 18–24 (2015)

99. Mothet, J., Rouaud, E., Sinet, P.-M., Potier, B., Jouvenceau, A., Dutar, P., Videau, C., Epelbaum, J., Billard, J.-M.: A critical role for the glial-derived neuromodulator d-serine in the age-related deficits of cellular mechanisms of learning and memory. Aging cell **5**(3), 267–274 (2006)

100. Le Couteur, D.G., Blyth, F.M., Creasey, H.M., Handelsman, D.J., Naganathan, V., Sambrook, P.N., Seibel, M.J., Waite, L.M., Cumming, R.G.: The association of alanine transaminase with aging, frailty, and mortality. Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences **65**(7), 712–717 (2010)

101. Cubizolle, A., Guillou, L., Mollereau, B., Hamel, C.P., Brabet, P.: Fatty acid transport protein 1 regulates retinoid metabolism and photoreceptor development in mouse retina. PloS one **12**(7), 0180148 (2017)

102. Tahir, H.J., Rodrigo-Diaz, E., Parry, N.R., Kelly, J.M., Carden, D., Murray, I.J.: Slowed dark adaptation in older eyes; effect of location. Experimental eye research **155**, 47–53 (2017)

103. Michels, A.J., Hagen, T.M., Frei, B.: Human genetic variation influences vitamin c homeostasis by altering vitamin C transport and antioxidant enzyme function. Annual review of nutrition **33**, 45–70 (2013)

104. Pinto, M., Pickrell, A.M., Wang, X., Bacman, S.R., Yu, A., Hida, A., Dillon, L.M., Morton, P.D., Malek, T.R., Williams, S.L., *et al.*: Transient mitochondrial dna double strand breaks in mice cause accelerated aging phenotypes in a ROS-dependent but p53/p21-independent manner. Cell Death & Differentiation **24**(2), 288–299 (2017)

105. Holeček, M.: Beta-hydroxy-beta-methylbutyrate supplementation and skeletal muscle in healthy and muscle-wasting conditions. Journal of Cachexia, Sarcopenia and Muscle (2017)

106. Gomes, A.P., Price, N.L., Ling, A.J., Moslehi, J.J., Montgomery, M.K., Rajman, L., White, J.P., Teodoro, J.S., Wrann, C.D., Hubbard, B.P., *et al.*: Declining NAD+ induces a pseudohypoxic state disrupting nuclear-mitochondrial communication during aging. Cell **155**(7), 1624–1638 (2013)

107. Desler, C., Lykke, A., Rasmussen, L.J.: The effect of mitochondrial dysfunction on cytosolic nucleotide metabolism. Journal of nucleic acids **2010** (2010)

108. Desler, C., Lillenes, M., Tønjum, T., Rasmussen, L.: The role of mitochondrial dysfunction in the progression of alzheimer's disease. Current medicinal chemistry (2017)

109. Mitteldorf, J.: How does the body know how old it is? introducing the epigenetic clock hypothesis. Biochemistry (Moscow) **78**(9), 1048–1053 (2013)

110. Benetos, A., Okuda, K., Lajemi, M., Kimura, M., Thomas, F., Skurnick, J., Labat, C., Bean, K., Aviv, A.: Telomere length as an indicator of biological aging. Hypertension **37**(2), 381–385 (2001)

111. Schubert, O.T., Röst, H.L., Collins, B.C., Rosenberger, G., Aebersold, R.: Quantitative proteomics: challenges and opportunities in basic and applied research. Nature Protocols **12**(7), 1289–1294 (2017)

112. Di Meo, A., Pasic, M.D., Yousef, G.M.: Proteomics and peptidomics: moving toward precision medicine in urological malignancies. Oncotarget **7**(32), 52460 (2016)

Figures

Table 1 Eigenvalue threshold % variance values for PCA. Significant components had eigenvalues with an individual contribution to variance greater than the threshold percentage variance value. The threshold percentage variances for each of the different groups of data are shown along with the details of how that value was obtained.

Group	Individuals	Average	Threshold % Variance
All individuals	499	100/498	0.2
21-	112	100/111	0.9
22-49	360	100/359	0.28
50+	27	100/26	3.85

Table 2 Number of components retained in PCA. For each group, the number of significant components are shown along with their cumulative contribution to variance.

Group	Components Retained	% Variance
All individuals	151	74.88
21-	39	63.42
22-49	112	71.79
50+	10	55.14

Additional Files

Additional file 1 — Silhouette values raw data

Average chronological age-based silhouette values for Hierarchical Transcriptomic Clusters (AHC T), Hierarchical Fluxomic Clusters (AHC F), k-means Transcriptomic Clusters (k-means T) and k-means Fluxomic Clusters (k-means F).

Additional file 2 — Supplementary PCA analysis raw data

Overall reaction contributions for each age group, differences between the overall reaction contributions for each age group, overall pathway contributions for each age group, differences between the overall pathway contributions for each age group, and absolute differences between the overall pathway contributions for each age group. More details about the reactions can be found in additional file 4.

Additional file 3 — Metabolic ages

Metabolic ages calculated using pathway flux values, metabolic ages calculated using reaction flux data including predictors, effect sizes, λ values, MSE and p-value.

Additional file 4 — Fluxomic data

The fluxomic data generated using the normalised gene expression data and the GEMsplice toolbox. Full details from the CD4 model of the reactions, reaction names and their formulas. The raw data used for this project can be found under GEO Accession Number GSE56033 http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56033

Additional file 5 — Plots of k-means clusters

Contains plots of clustered data used for comparison to analyse the cause of drops in the silhouette values. In the fluxomic k-means data we analysed the drop in silhouette values between clusters 16 and 17 and between clusters 18 and 19. We found that the cohesion of the clusters represented by the numbers 5, 9 and 12 in the plot of 16 clusters was lost in the plot of 17 clusters . Similarly, the cohesion of the clusters represented by the numbers 5, 6, 15 and 17 in the plot of 17 clusters plot were lost in the plot of 18 clusters.

In the transcriptomic k-means data we analysed the drop in silhouette values between clusters 10 and 11, 14 and 15 and 17 and 18. In the plot of 10 clusters the clusters represented by the numbers 2 and 4 lose their distinction in the plot of 11 clusters as they are merged into one large cluster. Similarly the clusters represented by the numbers 10 and 11 in the plot of 14 clusters are merged into a less distinct single cluster in the plot of 15 clusters. Additionally, the clusters represented by the numbers 2, 9 show better cohesion in the plot of 14 clusters. In the comparison of the plots of 17 and 18 clusters the cluster represented by the number 17 is particularly better defined in the plot of 17. Clusters 7, 15 and 16 are also better defined in the plot of 17 clusters.

Table 3 Metabolic effect sizes. The effect sizes on age of the three metabolic pathways selected by elastic net regression.

Subsystem	Effect size
Butanoate metabolism	-0.004462535
Pyrimidine synthesis	-2.18609E-05
beta-Alanine metabolism	-0.002443117

Data
Gene expression data from CD4 T-cells, chronological age and transcriptomic age from 499 individuals.

Transcriptomic Layer Chronological Ageing   **Transcriptomic Layer Transcriptomic Ageing**

Transcriptomic ageing
Recent research has identified age-associated genes allowing mapping of chronological and transcriptomic ageing in the transcriptomic layer.

Genome-scale poly-omic modelling
The transcriptomic data is fed into a model of the CD4 cell and mapped to the cell metabolism.

Poly-omic ageing map
A pipeline of machine learning methods is applied to the metabolic models, showing that fluxomic data clusters better with chronological age than transcriptomic data, and identifying metabolic predictors of individual ageing phenotypes.

Functional Map of Chronological Ageing

Enrichment
Low ○ ● ● High

Functional Map of Transcriptomic Ageing

Enrichment
Low ○ ● ● High

Individual-specific metabolic models
Running the genome-scale model on the gene expression data generates individual/sample specific metabolic models.

Figure 1 Poly-omic ageing pipeline. We start with the transcriptomic data and chronological ages from the CD4 T-cells of 499 individuals. We use the chronological data and corresponding age-associated transcriptomic predictors to obtain the effect of both chronological and transcriptomic ageing on the transcriptomic layer. We then combine with the functional biological network data determined by the metabolism and poly-omic model to obtain individual-based metabolic models and their fluxomic profiles. Finally, we adapt machine learning techniques to show that fluxomic data clusters better with chronological age than transcriptomic data, and to identify metabolic predictors of ageing (the poly-omic ageing map).
Definition of key terms. Transcriptomic – gene expression data represented by a measurement of the mRNA transcripts within a cell. **Fluxomic** – fluxomic data refers to reaction flux rates, namely the value for the rate of metabolite conversion, measured in millimoles per hour per grams of dry weight, for each reaction or collection of related reactions (subsystems) within a cell. **Poly-omic** – the integration of more than one type of 'omic' data e.g. transcriptomic and fluxomic data.

Figure 2 k-means clustering with age. We propose and investigate an average age-based silhouette value (A). This is calculated using the chronological age, and clustering data from both hierarchical and k-means clustering. The silhouette values are calculated by averaging all the individuals' silhouette scores, for a number of clusters ranging from 2 to 30. k-means clustering performs consistently better than hierarchical clustering. In both types of clustering, fluxomic data clusters better with chronological age than transcriptomic data. The pairwise distance of clusters with chronological age is visualised in scatter plots for both transcriptomic (E) and fluxomic (C) data. Clusters are annotated with different shapes, while age is shown with colour. Individual clusters are plotted against transcriptomic and chronological age for both transcriptomic (D) and fluxomic (B) data. Note that since transcriptomic age was calculated from transcriptomic data, we would expect to see more distinction in transcriptomic age between clusters for the transcriptomic data.
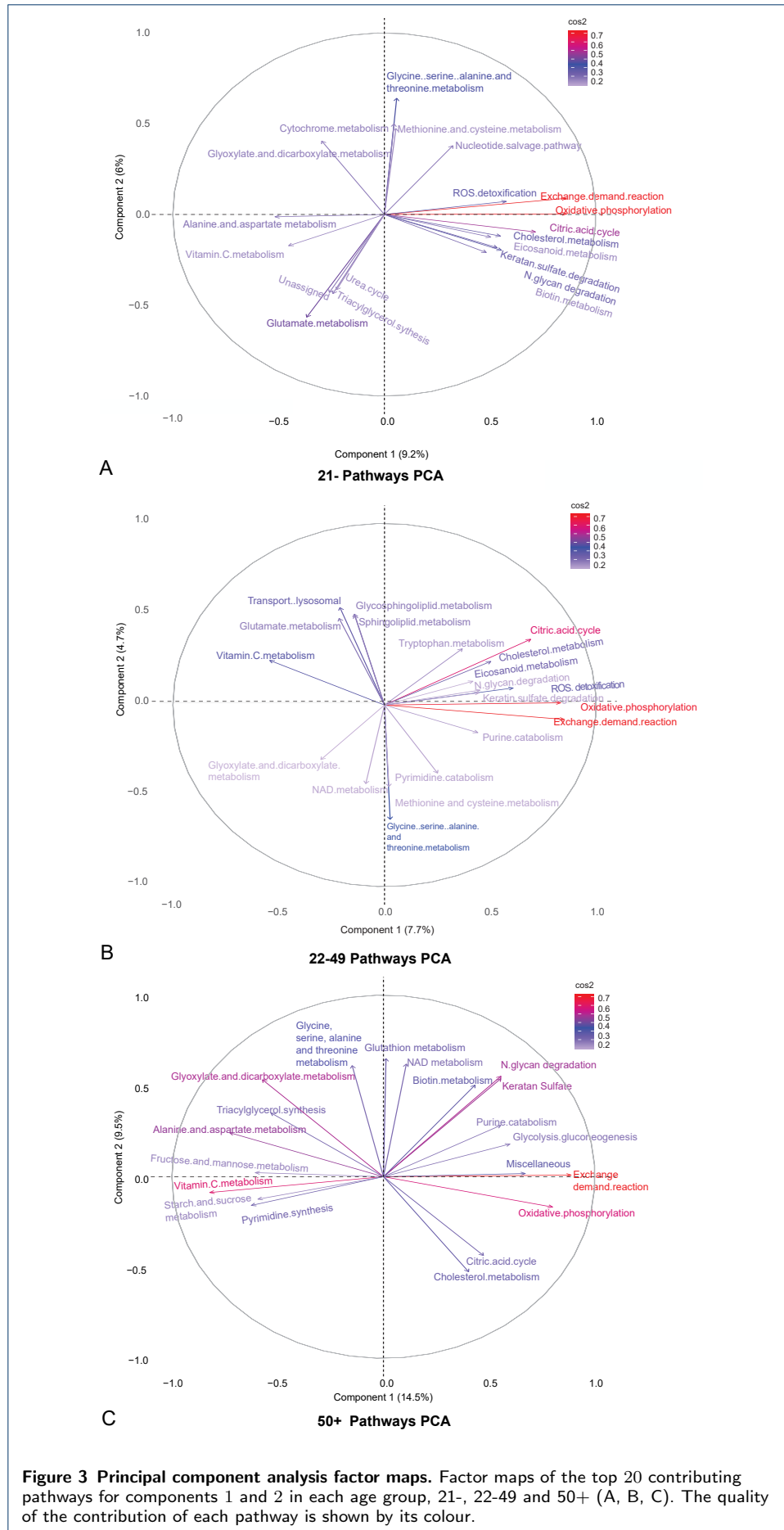
Figure 3 Principal component analysis factor maps. Factor maps of the top 20 contributing pathways for components 1 and 2 in each age group, 21-, 22-49 and 50+ (A, B, C). The quality of the contribution of each pathway is shown by its colour.
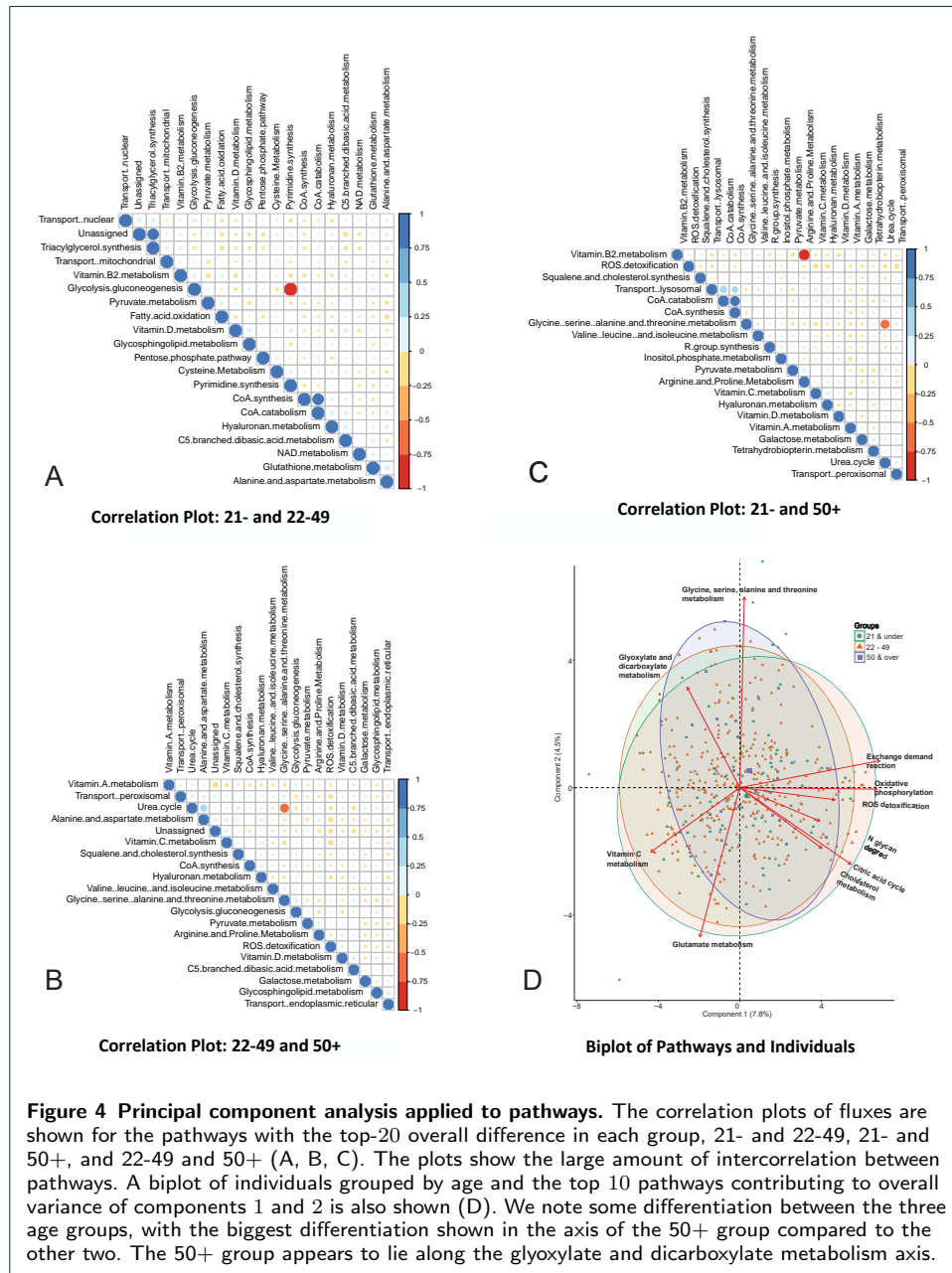
Figure 4 Principal component analysis applied to pathways. The correlation plots of fluxes are shown for the pathways with the top-20 overall difference in each group, 21- and 22-49, 21- and 50+, and 22-49 and 50+ (A, B, C). The plots show the large amount of intercorrelation between pathways. A biplot of individuals grouped by age and the top 10 pathways contributing to overall variance of components 1 and 2 is also shown (D). We note some differentiation between the three age groups, with the biggest differentiation shown in the axis of the 50+ group compared to the other two. The 50+ group appears to lie along the glyoxylate and dicarboxylate metabolism axis.