Aging Cell (2013) pp1–3 Doi: 10.1111/acel.12044

SHORT TAKE

Toward a gene expression biomarker set for human biological age

Alice C. Holly,*[1] David Melzer,*[2] Luke C. Pilling,[2] William Henley,[3] Dena G. Hernandez,[4] Andrew B. Singleton,[4] Stefania Bandinelli,[5] Jack M. Guralnik,[6] Luigi Ferrucci[7] and Lorna W. Harries[1]

[1]Institute of Biomedical and Clinical Science, Exeter Medical School, University of Exeter, Exeter, EX2 5DW, UK
[2]Epidemiology and Public Health, Exeter Medical School, University of Exeter, Exeter, EX2 5DW, UK
[3]Institute of Health Service Research, Exeter Medical School, University of Exeter, Exeter, EX2 4SG, UK
[4]Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
[5]Geriatric Unit, Azienda Sanitaria di Firenze, Florence, Italy
[6]Laboratory of Epidemiology, Demography and Biometry, National Institute on Aging, Bethesda, MD, USA
[7]National Institute on Aging, Baltimore, MD, USA

Summary

We have previously described a statistical model capable of distinguishing young (age <65 years) from old (age ≥75 years) individuals. Here we studied the performance of a modified model in three populations and determined whether individuals predicted to be biologically younger than their chronological age had biochemical and functional measures consistent with a younger biological age. Those with 'younger' gene expression patterns demonstrated higher muscle strength and serum albumin, and lower interleukin-6 and blood urea concentrations relative to 'biologically older' individuals (odds ratios 2.09, 1.64, 0.74, 0.74; $P = 2.4 \times 10^{-2}$, 3.5×10^{-4}, 1.8×10^{-2}, 1.5×10^{-2}, respectively). We conclude that our expression signature of age is robust across three populations and may have utility for estimation of biological age.

Key words: biological aging; mRNA expression; cell senescence; predictive model.

Although the importance of a healthy and active lifestyle in promoting good health during aging is well understood (van den Brandt, 2011), the physiological processes that influence biological rather than chronological aging are not clear. Identification of biomarkers for biological aging could provide a key insight into the heterogeneity of age-related decline in function, but have been an evasive goal. Recently, gene expression analyses have shown promise as a new tool to measure physiological age and identify a physical marker of aging (Weindruch *et al.*, 2002). We previously carried out a transcriptome-wide gene expression analysis in human peripheral blood leukocyte samples, to determine which transcripts were most associated with advancing age in the InCHIANTI study, a large well-characterized population-representative cohort aged 30–104 years (Ferrucci *et al.*, 2000). We found that large-scale differences in transcript expression levels occurred for only a small, focused set of genes, and that using the expression levels of only six of these (*LRRN3, CD27, GRAP, CCR6, VAMP5* and *CD248*), we could distinguish between younger (age <65 years) and older subjects (age ≥75 years) with high accuracy (Harries *et al.*, 2011).

In this new study, we aimed to validate a modified InCHIANTI marker set in two independent cohorts and test whether expression marker-based prediction of age in the original InCHIANTI data could identify people who were 'biologically younger' than their chronological age. To validate our marker set, we used blood leukocyte samples from the Exeter-10000 study (http://www.peninsulacrf.org/node/155) and isolated lymphocyte-derived Affymetrix expression array data from the San Antonio Family Heart Study (SAFHS) (Goring *et al.*, 2007). Cohort characteristics are given in supplementary table 1 online. Expression data on five of the six original transcripts were available in all three populations (*LRRN3, GRAP, CCR6, VAMP5,* and *CD27*). Three of these (*LRRN3*, *CCR6,* and *CD27*) were significantly associated with age by quantitative real-time PCR in $n = 95$ samples from the Exeter 10000 population and five were significantly associated with age ($P < 0.05$) in SAFHS data ($n = 1{,}238$) (supplementary table 2 online). Using the expression levels of these five transcripts in fitted multivariable logistic regression models distinguishing younger (age <65 years) from older (age ≥75 years) and subjects in the InCHIANTI and SAFHS cohorts produced Receiver Operating Characteristic (ROC) curves with Area Under the Curve (AUC) values of 94% and 87%, respectively (supplementary figure S1). We were unable to fit a ROC curve with these age cut-offs in the Exeter 10000 data, as only three individuals in this sample were aged ≥75 years, although a curve with cut-offs of ≥60 vs. <60 was feasible. To assess the validity of applying these amended cut-offs of ≥60 vs. <60 in the Exeter 10000 data, we first fitted ROC curves using these modified cut points in the two larger populations. We found that the choice of age cut-off had little effect, with ROC curves with Area Under the Curve (AUC) values of 92% and 82% being obtained in InCHIANTI and SAFHS cohorts, respectively (supplementary figure 2). Applying the modified cut-off in the Exeter 10000 data, we obtained an AUC value of 73% (supplementary figure 2). This suggests that our five-transcript expression panel is robust across populations, despite differences in cohort characteristics, white cell subsets, sample handing and analytical approach, and that the model is robust across two different age categories.

Correspondence
Professor David Melzer, Exeter Medical School, Barrack Rd, Exeter EX2 5DW, UK. Tel.: 01392 406753; fax: 01392 406767; e-mail: David.Melzer@pms.ac.uk

*These authors contributed equally to this publication.

Accepted for publication

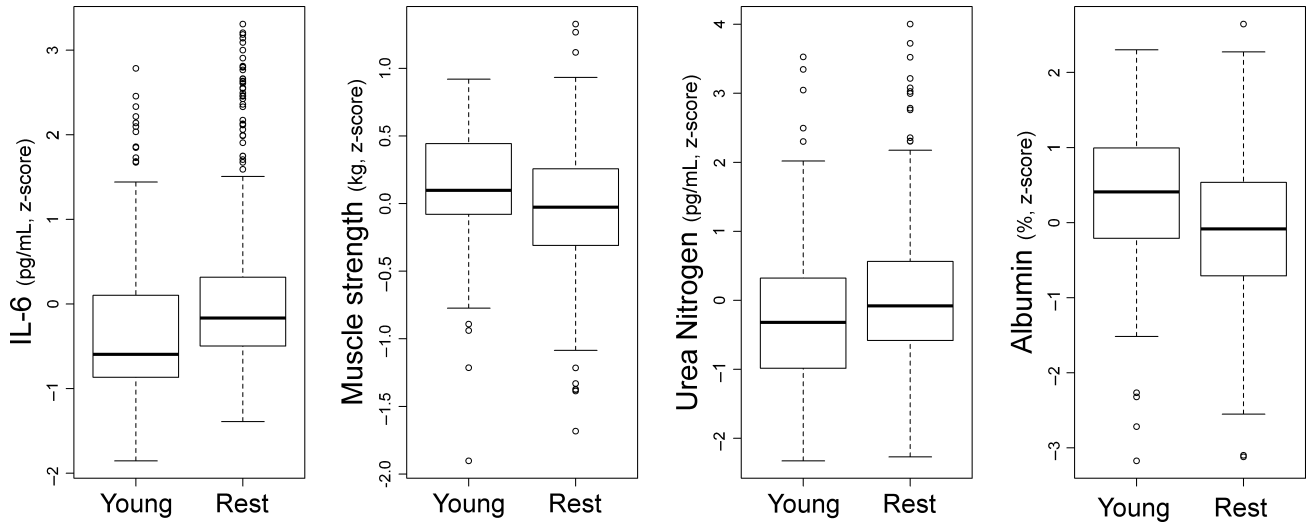

Fig. 1 Functional and physiological parameters in 'biologically younger' respondents. This figure shows the adjusted distribution of four aging biomarkers in respondents with five transcript regression-predicted ages >8.5 years younger than their chronological age, vs. the remainder of the study population.

To compare biological and chronological age in the InCHIANTI respondents, we used a linear regression model to estimate each individual's age based on the five-expression transcript intensities, with continuous age as the outcome. A total of 131 individuals were classified as 'biologically younger' if, in a regression model of the five transcripts, their predicted age was >8.5 years younger than their actual chronological age (corresponding to 20% of the sample). To explore associations with clinical markers of aging we included only those aged ≥ 65 years, as clinical aging markers such as the Short Physical Performance Battery (SPPB) are validated mainly in such older groups (Guralnik *et al.*, 1994; Freire *et al.*, 2012). This group was then analyzed for biomarkers of biological aging including C-reactive protein (CRP), Short Physical Performance Battery (SPPB) scores, muscle strength, interleukin-6 (IL-6) serum levels, systolic blood pressure, hematocrit, serum albumin, and blood urea level, which are known physiological markers of aging (Studenski *et al.*, 2003; Nakamura & Miyao, 2007).

We determined that four of these eight parameters (IL-6, muscle strength, blood urea, and serum albumin) displayed a statistically significant difference between the biologically young outliers ($n = 131$) and the rest of the cohort ≥ 65 years ($n = 406$) (Odds Ratios: 0.74, 2.09, 0.74, 1.64; *P*-values: 1.8×10^{-2}, 2.4×10^{-2}, 1.5×10^{-2}, 3.5×10^{-4}, respectively) (Fig. 1). Muscle strength decreases with age and predicts mobility disability in older adults (Bean *et al.*, 2003), accordingly muscle strength was significantly higher in the 'biologically young' group. IL-6 is an inflammatory marker (Cesari *et al.*, 2004) and serum concentrations rise with advancing age: significantly lower levels of IL-6 were found in the 'biological young' group. Individuals who were predicted to be younger than their actual age demonstrated higher serum albumin and lower blood urea nitrogen levels. Serum albumin decreases with age (Liu *et al.*, 2012) and is associated with increased morbidity, mortality, and disability in older people (Baumgartner *et al.*, 1996). In particular, low blood albumin is associated with reduced muscle mass and may be indicative of sarcopenia in elderly individuals (Visser *et al.*, 2005). Blood urea nitrogen levels increase with age (Aono *et al.*, 1994) and may represent reduced kidney function. The lower levels in 'biologically young' individuals may indicate better renal performance. Systolic blood pressure, SPPB scores, CRP, and hematocrit did not differ significantly between the 'biologically young' and the remaining population, but this is probably due to the increased variability of these parameters and a lack of statistical power. These data indicate that individuals who have gene expression signatures characteristic of a younger age group also have some biochemical or functional features consistent with a younger age and we conclude that our bi-class discriminative model may therefore be capable of estimating biological, rather than chronological age, and may hold utility in the future to predict individuals who may age badly.

Funding

This work was supported by the intramural program of the National institute on Aging, internal funds of the University of Exeter Medical School and the Exeter Peninsula NIHR Clinical Research Facility and the National Institute for Health Research (NIHR).

Acknowledgments

We thank the people and participants who contributed to the InCHIANTI, Exeter 10000 and SAFHS studies. William Henley was supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care (CLAHRC) for the South West Peninsula. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

References

Aono T, Matsubayashi K, Kawamoto A, Kimura S, Doi Y, Ozawa T (1994). [Normal ranges of blood urea nitrogen and serum creatinine levels in the community-

dwelling elderly subjects aged 70 years or over–correlation between age and renal function]. *Nihon Ronen Igakkai Zasshi* **31**, 232–236.
Baumgartner RN, Koehler KM, Romero L, Garry PJ (1996) Serum albumin is associated with skeletal muscle in elderly men and women. *Am. J. Clin. Nutr.* **64**, 552–558.
Bean JF, Leveille SG, Kiely DK, Bandinelli S, Guralnik JM, Ferrucci L (2003) A comparison of leg power and leg strength within the InCHIANTI study: which influences mobility more? *J. Gerontol. Ser. A: Biol. Sci. Med. Sci.* **58**, M728–M733.
van den Brandt PA (2011) The impact of a Mediterranean diet and healthy lifestyle on premature mortality in men and women. *Am. J. Clin. Nutr.* **94**, 913–920.
Cesari M, Penninx BWJH, Pahor M, Lauretani F, Corsi AM, Williams GR, Guralnik JM, Ferrucci L (2004) Inflammatory markers and physical performance in older persons: the InCHIANTI study. *J. Gerontol. Ser. A: Biol. Sci. Med. Sci.* **59**, M242–M248.
Ferrucci L, Bandinelli S, Benvenuti E, Di Iorio A, Macchi C, Harris TB, Guralnik JM (2000) Subsystems contributing to the decline in ability to walk: bridging the gap between epidemiology and geriatric practice in the InCHIANTI study. *J. Am. Geriatr. Soc.* **48**, 1618–1625.
Freire AN, Guerra RO, Alvarado B, Guralnik JM, Zunzunegui MV (2012) Validity and reliability of the short physical performance battery in two diverse older adult populations in Quebec and Brazil. *J. Aging Health.* **24**, 863–878.
Goring HH, Curran JE, Johnson MP, Dyer TD, Charlesworth J, Cole SA, Jowett JB, Abraham LJ, Rainwater DL, Comuzzie AG, Mahaney MC, Almasy L, MacCluer JW, Kissebah AH, Collier GR, Moses EK, Blangero J (2007) Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat. Genet.* **39**, 1208–1216.
Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, Scherr PA, Wallace RB (1994) A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J. Gerontol.* **49**, M85–M94.
Harries LW, Hernandez D, Henley W, Wood A, Holly AC, Bradley-Smith RM, Yaghootkar H, Dutta A, Murray A, Frayling TM, Guralnik JM, Bandinelli S, Singleton A, Ferrucci L, Melzer D (2011) Human aging is characterized by focused changes in gene expression and deregulation of alternative splicing. *Aging Cell* **10**, 868–878.
Liu M, Chan C-P, Yan BP, Zhang Q, Lam Y-Y, Li R-J, Sanderson JE, Coats AJS, Sun J-P, Yip GW-K, Yu C-M (2012) Albumin levels predict survival in patients with heart failure and preserved ejection fraction. *Eur. J. Heart Fail.* **14**, 39–44.
Nakamura E, Miyao K (2007) A method for identifying biomarkers of aging and constructing an index of biological age in humans. *J. Gerontol. A Biol. Sci. Med. Sci.* **62**, 1096–1105.
Studenski S, Perera S, Wallace D, Chandler JM, Duncan PW, Rooney E, Fox M, Guralnik JM (2003) Physical Performance Measures in the Clinical Setting. *J. Am. Geriatr. Soc.* **51**, 314–322.
Visser M, Kritchevsky SB, Newman AB, Goodpaster BH, Tylavsky FA, Nevitt MC, Harris TB (2005) Lower serum albumin concentration and change in muscle mass: the Health, Aging and Body Composition Study. *Am. J. Clin. Nutr.* **82**, 531–537.
Weindruch R, Kayo T, Lee C-K, Prolla TA (2002) Gene expression profiling of aging using DNA microarrays. *Mech. Ageing Dev.* **123**, 177–193.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

Fig. 1 ROC curves with cutoff points of <65 vs. ≥75. ROC curves with Area Under the Curve (AUC) for InCHIANTI and SAFHS populations with age cut off points of younger subjects <65 and older subjects ≥75. AUC values were 94% and 87% for the InCHIANTI (A) and SAFHS (B) populations respectively.

Fig. 2 ROC curves with Area Under the Curve (AUC) for InCHIANTI, SAFHS and Exeter 10000 populations with age cut off points of younger subjects <60 and older subjects ≥ 60. AUC values were 92%, 82% and 73% for the InCHIANTI (A), SAFHS (B) and Exeter 10000 (C) populations respectively.

Data S1 Experimental procedures.

Table S1 Characteristics of the InCHIANTI, SAHFS and Exeter 10000 cohorts. N refers to the number of individuals in the sample. % refers to the percentage that the figure represents of the total cohort. SD refers to the Standard deviation of the measurement for waist circumference. n/a = data not available.

Table S2 Genes included in the bi-class discriminant model of age in the InCHIANTI, SAFHS and Exeter 10000 cohorts. Genes included in the bi-class discriminant model are given below. The *q*-value refers to the significance adjusted for false discovery rate (FDR). 95% CI refers to the confidence intervals for the TLDA analysis. B = coefficient (change in standard deviations of expression per year of age).