EBioMedicine 5 (2016) 68–73

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

Research Paper

Blood Epigenetic Age may Predict Cancer Incidence and Mortality

Yinan Zheng [a,b,1], Brian T. Joyce [a,c,1], Elena Colicino [d], Lei Liu [a,e], Wei Zhang [a,e], Qi Dai [f], Martha J. Shrubsole [f], Warren A. Kibbe [g,h], Tao Gao [a], Zhou Zhang [a,i], Nadereh Jafari [j], Pantel Vokonas [k], Joel Schwartz [d,l], Andrea A. Baccarelli [d], Lifang Hou [a,e,*]

[a] Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA
[b] Institute for Public Health and Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA
[c] Division of Epidemiology and Biostatistics, School of Public Health, University of Illinois-Chicago, Chicago, IL 60613, USA
[d] Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA
[e] Robert H. Lurie Comprehensive Cancer Center, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA
[f] Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN 37203, USA
[g] Center for Biomedical Informatics and Information Technology, National Cancer Institute, 9609 Medical Center Drive, Rockville, MD 20850, USA
[h] Biomedical Informatics Center (NUBIC), Northwestern University Clinical and Translational Sciences Institute (NUCATS), Chicago, IL 60611, USA
[i] Driskill Graduate Program in Life Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA
[j] Genomics Core Facility, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA
[k] VA Boston Healthcare System and Boston University Schools of Public Health and Medicine, Boston, MA 02215, USA
[l] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

ARTICLE INFO

Article history:
Received 15 December 2015
Received in revised form 14 January 2016
Accepted 5 February 2016
Available online 8 February 2016

Keywords:
Epigenetic age
Cancer risk
DNA methylation

ABSTRACT

Biological measures of aging are important for understanding the health of an aging population, with epigenetics particularly promising. Previous studies found that tumor tissue is epigenetically older than its donors are chronologically. We examined whether blood Δ_{age} (the discrepancy between epigenetic and chronological ages) can predict cancer incidence or mortality, thus assessing its potential as a cancer biomarker. In a prospective cohort, Δ_{age} and its rate of change over time were calculated in 834 blood leukocyte samples collected from 442 participants free of cancer at blood draw. About 3–5 years before cancer onset or death, Δ_{age} was associated with cancer risks in a dose-responsive manner ($P = 0.02$) and a one-year increase in Δ_{age} was associated with cancer incidence (HR: 1.06, 95% CI: 1.02–1.10) and mortality (HR: 1.17, 95% CI: 1.07–1.28). Participants with smaller Δ_{age} and decelerated epigenetic aging over time had the lowest risks of cancer incidence ($P = 0.003$) and mortality ($P = 0.02$). Δ_{age} was associated with cancer incidence in a 'J-shaped' manner for subjects examined pre-2003, and with cancer mortality in a time-varying manner. We conclude that blood epigenetic age may mirror epigenetic abnormalities related to cancer development, potentially serving as a minimally invasive biomarker for cancer early detection.

1. Introduction

As the population continues to age in the coming decades, the need for biological measures of age and more precise screening tests for age-related diseases will become increasingly urgent. Genetic and epigenetic studies have added to the potential clinical utility of this subject area, with researchers identifying elements associated with age-related processes (e.g., telomeres) that nonetheless have unverified predictive power in human populations (Brooks-Wilson, 2013). Epigenetics serves as an intersection between genetic and environmental risk factors for aging processes and age-related diseases, holding great promise for constructing biological age measures that will provide clinical diagnostic tools for aging-related diseases such as cancer. Blood-based epigenetic markers are particularly well-suited to these purposes due to their minimally invasive method of collection and cost-effectiveness on the population scale.

Epigenetic age is a recently developed algorithm that uses DNA methylation measurements to describe biological age at the level of human tissues, cells, and organs (Hannum et al., 2013; Horvath, 2013; Weidner et al., 2014). Epigenetic age does not always parallel chronological age, particularly in tumor samples (Hannum et al., 2013; Horvath, 2013), a discrepancy we refer to here as Δ_{age}. Furthermore, since the methods for measuring epigenetic age incorporate loci in pathways related to both cancer development and aging in general (e.g., DNA damage, cellular proliferation, and oxidative stress)

* Corresponding author at: Department of Preventive Medicine, Northwestern University, 680 N. Lake Shore Drive, Suite 1400, Chicago, IL 60611, USA.
E-mail address: l-hou@northwestern.edu (L. Hou).
[1] These authors contributed equally.

http://dx.doi.org/10.1016/j.ebiom.2016.02.008

(Hannum et al., 2013; Horvath, 2013), it is highly possible that Δ_{age} can be a predictive biomarker for cancer risk, metastasis, and mortality in addition to serving as an indicator of aging. With further study and refinement, the concept of epigenetic age may also be useful for improving our understanding of mechanisms by which age and cancer are related. However, no longitudinal analysis has yet evaluated how blood epigenetic age changes over time prior to cancer diagnosis or cancer-related death, and whether blood Δ_{age} can predict future risk of cancer incidence and mortality.

Our objective is to assess whether white blood cell (WBC) Δ_{age} can predict cancer incidence and mortality, and to evaluate these predictions over time. We will achieve our goals by comparing multiple estimates of Δ_{age} obtained using blood DNA samples collected prior to cancer incidence and death in: 1) individuals who developed cancer relative to cancer-free individuals and 2) individuals who died of cancer relative to both cancer survivors and cancer-free individuals.

2. Materials and Methods

2.1. Study Design and Participants

The U.S. Department of Veterans Affairs' Normative Aging Study (NAS) is a longitudinal cohort established in 1963 (Bell et al., 1966). Between March 1999 and December 2013, 802 out of 829 (96.7%) of active participants agreed to donate blood, 686 of whom were randomly selected and profiled using the Illumina 450 K BeadChip array at up to three follow-up visits separated by median time intervals of 3.5 years (IQR 3.1–5.7) (Supplementary material, Fig. S1). We excluded 18 participants who were non-White or had missing information on race to minimize potential confounding effects of genetic ancestry, 182 participants who had been diagnosed with malignant cancer prior to the first blood draw, and 44 participants who had been diagnosed with neoplasms of uncertain behavior, leaving a total of 442 participants for analysis. Cancer diagnoses and comorbidities (prevalent diabetes, hypertension, coronary artery disease, and stroke) were obtained from questionnaires and confirmed via blinded medical record review. Official death certificates were reviewed by a physician, and all contributing causes coded by an experienced research nurse. Telomere length was measured using quantitative real-time polymerase chain reaction and reported in relative units, as previously described (Hou et al., 2015). Of the 442 participants free of cancer at first blood draw, 132 developed cancer during follow-up (38 prostate, 50 skin, 44 other) and 34 died from cancer (4 prostate, 2 skin, 28 other). The remaining 98 participants died from non-cancer causes or were still alive as of latest follow-up, and are defined as cancer survivors for our analysis (Supplementary material, Fig. S2). Median follow-up time (from first blood draw) to cancer incidence or censoring was 10.1 years (IQR 5.8–12.7) and 11.9 years (IQR 8.8–13.0) to cancer mortality or censoring. This study was approved by the Institutional Review Boards of all participating institutions, and all participants provided written consent.

2.2. Illumina Infinium HumanMethylation450 BeadChip DNA Methylation Profiling

DNA was extracted from the buffy coat using the QIAamp DNA Blood Kit (QIAGEN, Valencia, CA, USA). A total of 500 ng of DNA was used to perform bisulfite conversion using the EZ-96 DNA Methylation Kit (Zymo Research, Orange, CA, USA). To limit chip and plate effects a two stage, age-stratified algorithm was used to randomize samples and ensure similar age distributions across chips and plates. We randomized 12 samples (sampled across all age quartiles) per chip, and then randomized chips to plates. Quality control analysis was performed to remove samples where >1% of probes had a detection $P > 0.05$. The remaining samples were preprocessed using Illumina-type background correction without normalization as re-implemented in the Bioconductor *minfi* package and used to generate beta values (Aryee et al., 2014). The working set included all 485,512 CpG and CpH probes for epigenetic age estimation.

2.3. Epigenetic Age and Δ_{age} Calculation

We estimated epigenetic age as measured by Hannum's 71-CpG method using Horvath's online calculator (http://labs.genetics.ucla.edu/horvath/dnamage/), which incorporates both previously-published methods (Hannum et al., 2013; Horvath, 2013) based on prior work showing superior predictive accuracy in blood (Marioni et al., 2015). We selected Hannum's method because the model was trained using blood samples and it yielded accurate and longitudinally stable epigenetic age estimation using our blood-based methylation data (Supplementary material, Fig. S3). To remove potential confounding due to alterations in WBC composition with age (Marioni et al., 2015) and naïve T cell changes due to thymic involution (Taub and Longo, 2005), cell abundances were estimated via previously published methods (Horvath, 2013; Houseman et al., 2012). We then used the residuals from a linear regression of epigenetic age on chronological age and these cell type abundances to calculate Δ_{age} (with $\Delta_{age} > 0$ indicating epigenetic age older than chronological age), resulting in a construct less sensitive to normal age-related changes in WBC composition and immunosenescence (Supplementary material, Fig. S4).

2.4. Statistical Analysis

We performed descriptive analyses of participant characteristics (including epigenetic age and Δ_{age}) at first blood draw using linear mixed-effect models. We then expanded these descriptive analyses to assess dose–response via a multivariable linear mixed-effect model comparing Δ_{age} across samples (estimated by least squares means) in participants who died from cancer, cancer survivors, and cancer-free participants. Next, we used Random Matrix Theory (Plerou et al., 2002) available in R package *isva* (Teschendorff et al., 2011) to determine that there were two significant components of variation in our data. We made a heat map (Supplementary material, Fig. S5) of *P*-values matrix of associations between the significant components and: biological (Houseman's six blood cell type proportions; Houseman et al., 2012), epidemiological (the 11 covariates we considered) and technical factors (plate number, chip number, row number and column number of the chip). We found that the components were strongly associated with biological and technical factors, but not epidemiological factors. Therefore, all subsequent models adjusted for the top two principle components computed from the 450 K array data to account for batch effects and residual biological confounding. We included the third top PC to better capture latent population stratification of the data (Barfield et al., 2014).

Prior to running our survival analyses, we evaluated the proportional hazards assumption for both cancer incidence and mortality by plotting Schoenfeld residuals over time. Although the hazards for cancer incidence were proportional across all follow-up time, the mortality analysis violated the proportional hazards assumption with marginal statistical significance ($P = 0.04$; data available upon request). Therefore, our subsequent analyses examined Δ_{age} and cancer both across all follow-up visits and stratified by year of blood DNA collection. We selected a stratification point of January 1st, 2003 as 1) it separated two data collection 'cycles' in the NAS rotation protocol (Supplementary material, Fig. S1), and 2) the resulting subgroups were of similar size. Subsequent testing of the proportional hazards assumption within these strata revealed no violations for either cancer incidence or mortality.

We plotted KM curves, comparing participants epigenetically older than their chronological age ($\Delta_{age} > 0$) to those epigenetically younger than their chronological age ($\Delta_{age} \leq 0$) using log-rank tests. Next, we