

RESEARCH ARTICLE

Open Access



Lung transcriptomic clock predicts premature aging in cigarette smoke-exposed mice

Mohamed-Amin Choukrallah*, Julia Hoeng, Manuel C. Peitsch and Florian Martin

Abstract

Background: Lung aging is characterized by a number of structural alterations including fibrosis, chronic inflammation and the alteration of inflammatory cell composition. Chronic exposure to cigarette smoke (CS) is known to induce similar alterations and may contribute to premature lung aging. Additionally, aging and CS exposure are associated with transcriptional alterations in the lung. The current work aims to explore the interaction between age- and CS-associated transcriptomic perturbations and develop a transcriptomic clock able to predict the biological age and the impact of external factors on lung aging.

Results: Our investigations revealed a substantial overlap between transcriptomic response to CS exposure and age-related transcriptomic alterations in the murine lung. Of particular interest is the strong upregulation of immunoglobulin genes with increased age and in response to CS exposure, indicating an important implication of B-cells in lung inflammation associated with aging and smoking. Furthermore, we used a machine learning approach based on Lasso regression to build a transcriptomic age model that can accurately predict chronological age in untreated mice and the deviations associated with certain exposures. Interestingly, CS-exposed-mice were predicted to be prematurely aged in contrast to mice exposed to fresh air or to heated tobacco products (HTPs). The accelerated aging rate associated with CS was reversed upon smoking cessation or switching to HTPs. Additionally, our model was able to predict premature aging associated with thoracic irradiation from an independent public dataset.

Conclusions: Aging and CS exposure share common transcriptional alteration patterns in the murine lung. The massive upregulation of B-cell restricted genes during these processes shed light on the contribution of cell composition and particularly immune cells to the measured transcriptomic signal. Through machine learning approach, we show that gene expression changes can be used to accurately monitor the biological age and the modulations associated with certain exposures. Our findings also suggest that the premature lung aging is reversible upon the reduction of harmful exposures.

Keywords: Age prediction, Machine learning, Transcriptomics, Cigarette smoke exposure, Lung inflammation

* Correspondence: MohamedAmin.Choukrallah@pmi.com
Philip Morris International R&D, Quai Jeanrenaud 5, 2003 Neuchâtel, Switzerland



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Aging of multi-cellular organisms is a complex biological process involving various molecular alterations such as epigenetic [1], genetic [2], and transcriptional changes [3–6]. These changes are likely modulated by various environmental factors that can either induce premature aging [7] or decelerate the natural aging rate [8]. The individual rates of aging greatly differ within a population, leading to the concept that the biological age can be monitored by a plethora of metrics, including omics-based measurements [9–11], providing robust indicators of healthy aging.

Modulation of transcriptional patterns is a central step in cellular response to numerous cues, including those related to aging. Gene regulation is one of the most investigated biological processes in relation to aging and a number of diseases in both humans and model organisms. Over the last decade, a massive number of transcriptomes have been generated, primarily by microarrays [12], in order to identify markers of aging and to understand the underlying mechanisms of aging processes. These studies revealed a number of common features for age-related transcriptomic changes, including the upregulation [13, 14] of genes involved in inflammation and stress response and the downregulation of genes associated with extracellular matrix constitution and metabolism [15, 16].

In the lung, biological aging correlates with anatomical and structural changes, such as fibrosis [17, 18] and immune dysregulation [19], mainly characterized by the alteration of inflammatory cell composition and chronic low-grade inflammation that are associated with progressive functional alterations. In rats, age-related lung fibrosis correlates with an increase in peri-bronchial collagen deposition and a decrease in matrix metalloproteinases (MMP) activity concomitant with an increase of tissue inhibitors of MMP (TIMP-1 and TIMP-2, 15). A number of external factors can induce aging-like structural and transcriptional alterations. For example, irradiated young mice exhibit lung transcriptomic profiles similar to unirradiated older mice [20]. Similarly, infection of young mice by respiratory syncytial virus induces transcriptomic alterations similar to those observed in uninfected old mice [16].

Chronic exposure to cigarette smoke (CS) contributes to many age-associated lung diseases, such as chronic obstructive pulmonary disease (COPD) and lung cancer [21]. Furthermore, aging and the response to CS exposure share common mechanisms involved in lung pathogenesis, such as impairment of proteostasis and autophagy leading to cellular senescence [22], suggesting that CS may contribute to premature lung aging.

Age- and CS-related transcriptomic alterations in the lung have been previously investigated; however, a detailed analysis of the overlap between these 2 transcriptomic patterns

from technically comparable datasets is missing. Here, we took advantage of highly standardized transcriptomics experiments where the impact of CS on the lung was investigated over periods of exposure ranging from 6 to 8 months [23–25]. Therefore, these datasets from the same experimental setup were analyzed to provide information about age- and CS-associated transcriptional changes.

The current work seeks to investigate the transcriptomic crosstalk between CS and aging in the murine lung, build a transcriptomic age predictor, and evaluate the effect of environmental factors on transcriptomic age. We hypothesized that deviations from the average aging rate may indicate perturbed aging processes.

We found a substantial overlap between age- and CS-regulated genes; among those, immunoglobulin genes were the most upregulated in response to both aging and CS exposure. Moreover, we derived an age prediction model from transcriptomes of sham (fresh air)-exposed mice. Strikingly, the majority of age predictor genes were deregulated by CS, further supporting a strong interaction between aging and CS exposure. In line with this observation, CS-exposed mice were predicted to be prematurely aged, whereas mice exposed to aerosols from heated tobacco products (HTPs) had age predictions similar to those of the sham-exposed mice. Smoking cessation or switching to HTPs reduced the predicted accelerated transcriptomic aging. Similarly, mice exposed to fibrogenic irradiation (public datasets [20]) were predicted to be prematurely aged in comparison with untreated mice, further supporting the robustness of our prediction model when applied to an independent set of data. Altogether, our results indicate that transcriptomic clocks are a valuable tool to monitor the impact of environmental exposures on biological aging.

Results

Datasets

To investigate the relationship between lung aging and CS exposure at the transcriptional level, we leveraged 3 lung transcriptomic datasets [23–25] generated from wild-type female C57BL/6 or apolipoprotein E-deficient (ApoE^{-/-}) transgenic mice exposed to CS (3R4F reference cigarette), fresh air (sham), or HTP aerosols over 6 to 8 months. The protocols also included smoking cessation and switching from CS to HTPs. Hereafter, these studies will be named E-MTAB-5281 [23], E-MTAB-5280 [24], and E-MTAB-7444 [25], as summarized in Supplementary Figure 1. All of the mice used in these studies were 2 months old at the beginning of the exposure.

Substantial overlap between CS- and age-related transcriptional alterations in the lung

Animals from the sham exposure groups were considered as representative of naturally aging mice and were

used to calculate age-dependent differential gene expression. Age-regulated genes have been identified by comparing gene expression levels between sham samples from different age groups in each study (supplementary Table 1). In total, 1325 age-regulated genes were identified (see Methods). Strikingly, 62% of these genes were also deregulated by CS when compared to the corresponding sham controls from the same time points (Fig. 1a). This association was further supported by the strong correlation between the first principal component loadings derived from gene expression regulation profiles in response to aging and CS (Fig. 1b). Comparing the maximum values of gene expression fold changes (FC) associated with age and in response to CS across the 3 analyzed datasets showed strong directional interactions between CS- and age-related transcriptional responses (Fig. 1c). Of note, comparing maximum gene expression variations does not imply a statistical significance, nor does it indicate stable directional variations. Nevertheless, we observed a strong overlap between statistically significant differentially expressed genes (colored green in Fig. 1c) and genes with high maximum FC. Of particular interest, the top 10 genes co-upregulated by age and CS are central components of the immune response, and most of them are immunoglobulin genes (B-cell receptors), suggesting a substantial contribution of B-cell lineage in immune response to aging and CS exposure.

Building a transcriptomic age predictor

It has been shown that DNA methylation levels of certain CpGs can be used to predict chronological age in

humans [9, 11, 26, 27] and mice [28–32]; predictions in mice are generally less accurate than in humans and the most accurate DNA methylation clocks might not be optimal for evaluating aging modulations [29]. Transcriptomic age predictors have also been developed for human blood samples but are less accurate than their epigenetic counterparts [10]. To our knowledge, no transcriptomic age predictor has been developed for mice.

To test whether gene expression levels in murine lungs can be used to estimate chronological age, we used sham transcriptomes to derive an age regression model by applying the least absolute shrinkage and selection operator (LASSO) regression (see Methods). The performance of this prediction approach was evaluated by cross-validation, where sham samples were randomly partitioned into training and validation sets (Fig. 2a) and “leave one study out” cross-validation (Fig. 2b), resulting in a mean average error in the validation sets of 0.83 and 1 month, respectively, largely outcompeting the majority of mouse epigenetic clocks that are usually based on hundreds of regressors (CpGs). The final predictor derived by using all sham samples from the 3 analyzed datasets resulted in a set of 57 predictor genes. In line with our previous observations, the majority (44 out of 57) of those age predictor genes were also deregulated in response to CS exposure (Fig. 2c). Age predictor genes include key markers of CS exposure, such as *Cyp1a1*, *Lcn2*, and *Mmp* genes, in addition to genes involved in immune response, such as immunoglobulin genes.

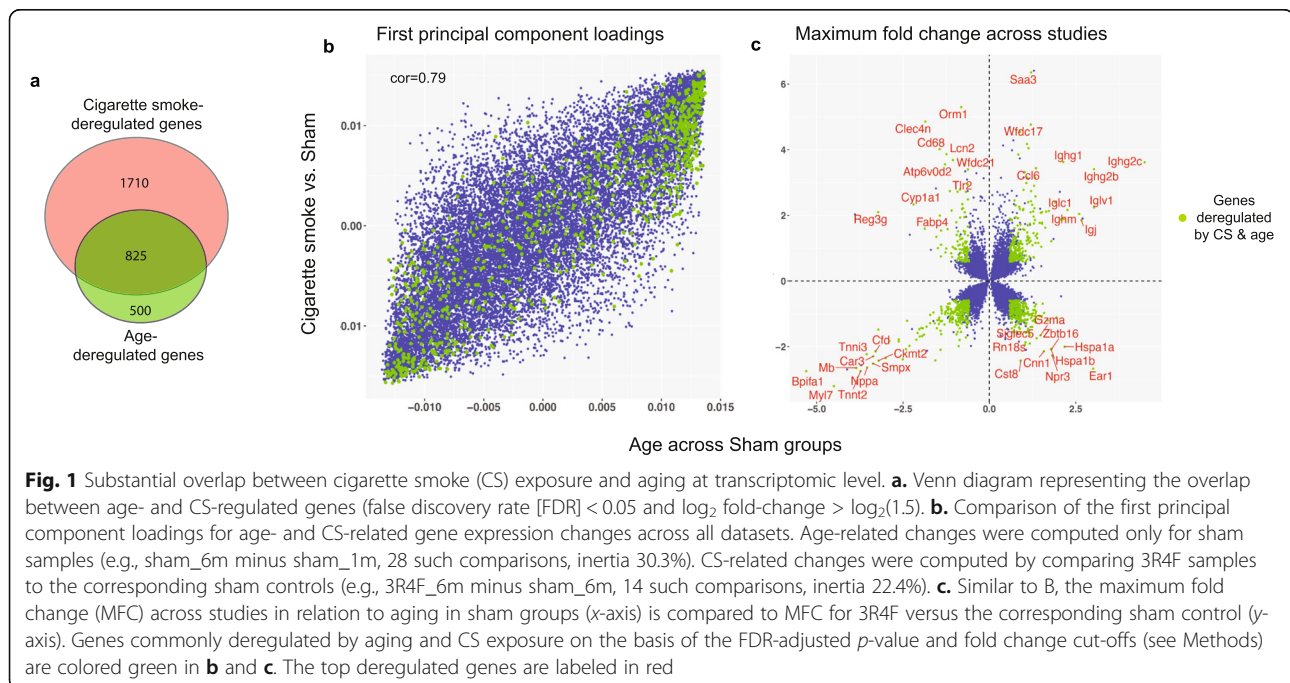


Fig. 1 Substantial overlap between cigarette smoke (CS) exposure and aging at transcriptomic level. **a.** Venn diagram representing the overlap between age- and CS-regulated genes (false discovery rate [FDR] < 0.05 and \log_2 fold-change > $\log_2(1.5)$). **b.** Comparison of the first principal component loadings for age- and CS-related gene expression changes across all datasets. Age-related changes were computed only for sham samples (e.g., sham_6m minus sham_1m, 28 such comparisons, inertia 30.3%). CS-related changes were computed by comparing 3R4F samples to the corresponding sham controls (e.g., 3R4F_6m minus sham_6m, 14 such comparisons, inertia 22.4%). **c.** Similar to B, the maximum fold change (MFC) across studies in relation to aging in sham groups (x-axis) is compared to MFC for 3R4F versus the corresponding sham control (y-axis). Genes commonly deregulated by aging and CS exposure on the basis of the FDR-adjusted p -value and fold change cut-offs (see Methods) are colored green in **b** and **c**. The top deregulated genes are labeled in red

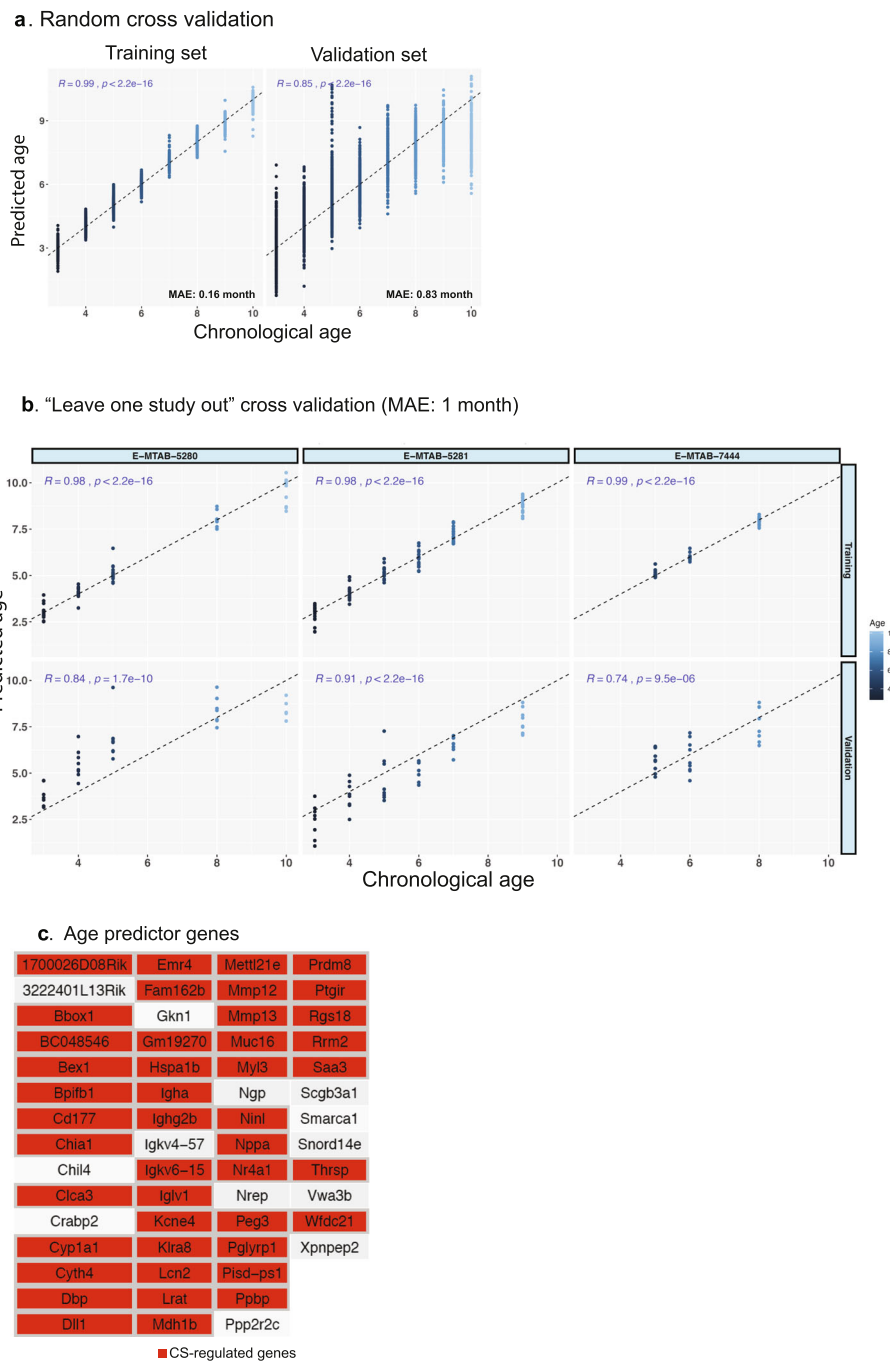


Fig. 2 Development of age-prediction model from lung transcriptomics data (E-MTAB-5281, E-MTAB-5280 and E-MTAB-7444) using LASSO regression. **a.** Random leave-out cross-validation over 100 runs. For every run, the 111 Sham samples from the 3 studies are randomly divided into training (75%) and validation sets (25%). **b.** Leave-one-study-out cross-validations. In each round, one study is excluded from the training but used for the validation. **c.** The final list of age predictor genes as selected by LASSO when applied to all of the 111 Sham transcriptomes. CS-regulated genes (see Methods) are colored red. MAE: mean average error

CS and thoracic irradiation accelerate lung transcriptional aging

Given the strong transcriptional overlap between CS and aging, we hypothesized that CS exposure may affect the transcriptomic age predicted by the model above. To

test this hypothesis, we applied the transcriptomic age predictor to CS-exposed samples. Although the model can accurately predict the age progression for those samples (Fig. 3a), predictions consistently showed a transcriptomic age higher than their chronological age (Fig.

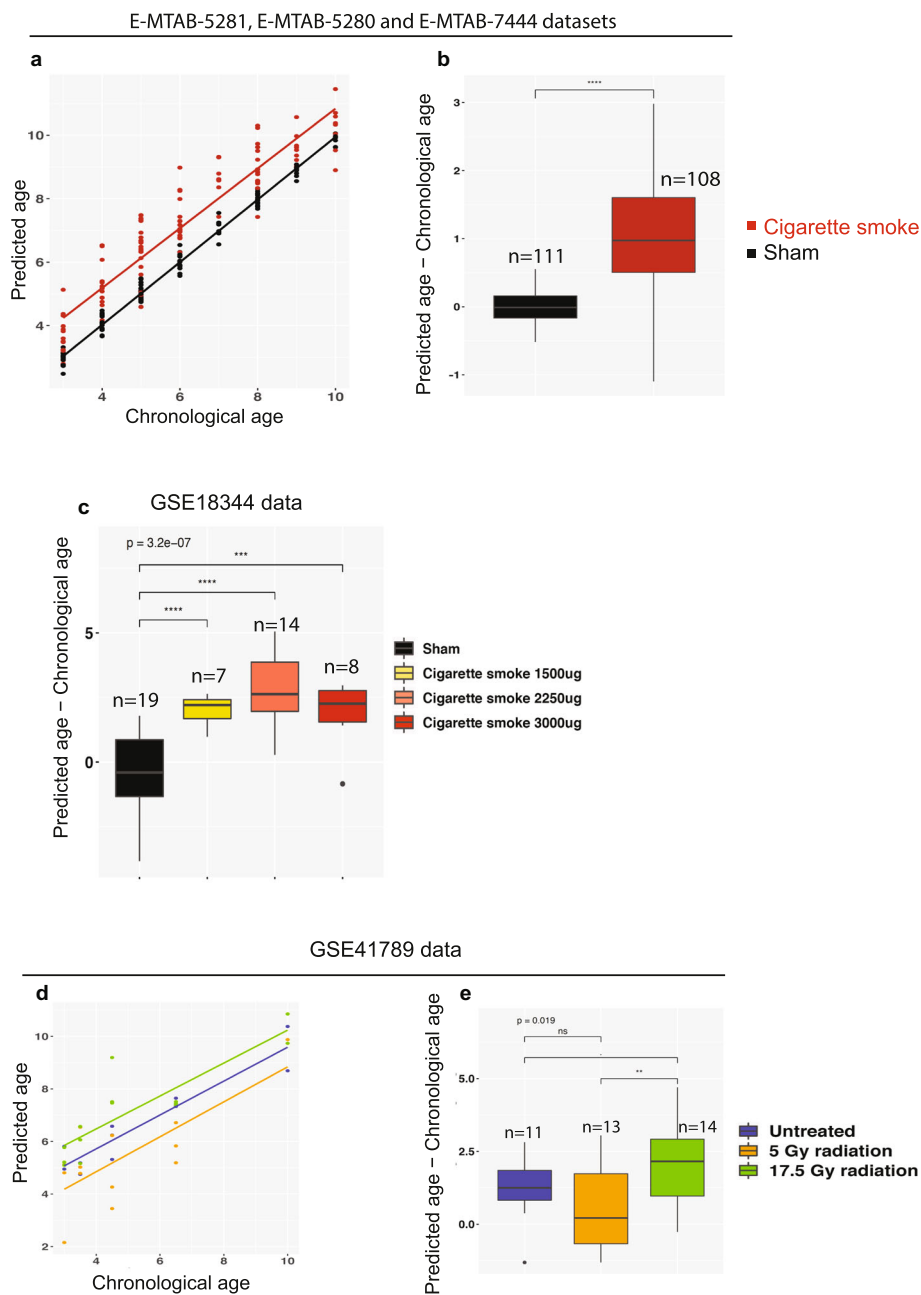


Fig. 3 CS-exposed mice and highly irradiated mice are predicted to be prematurely aged. **a.** Chronological age versus predicted age for sham- and CS-exposed mice from E-MTAB-5281, E-MTAB-5280 and E-MTAB-7444 datasets. **b.** Boxplots summarizing the residuals (predicted age minus chronological age) for the same datasets. **c.** Age predictions of samples from an independent public dataset (GSE18344) where CD-1 mice were exposed to different daily doses of CS over 5 months. **d** and **e.** Prediction from the model applied to lung transcriptomes from the public dataset GSE41789. Age predictions were computed for 3 groups of mice; untreated mice and mice irradiated to 5 Gy or 17.5 Gy. Only 17.5 Gy induced lung fibrosis, senescence, and marks of accelerated aging according to the original study. Significance levels for t-test are indicated for each comparison (**** $P < 0.0001$; *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$; $P < 0.1$; ns $P > 0.1$). The p-value from one-way anova test comparing all the groups is indicated

3a and b). This result suggests that CS may contribute to premature lung aging, which is further supported by the overexpression of a number of immunoglobulin genes in CS-exposed samples when compared to their sham counterparts (Supplementary Figure. 2).

To test the performance of our model on independent datasets, we used publicly available transcriptomic lung data generated from 12 weeks CD-1 mice exposed to different daily doses of cigarette smoke (2R4F) over 5 months or to sham as a control [33]. The age predictor

performed remarkably well in this dataset resulting in a clear distinction between fresh air and cigarette smoke-exposed samples (Fig. 3c). While the difference between chronological and predicted age was close to zero for sham samples, all CS-exposed samples were predicted to be prematurely aged (Fig. 3c).

To further evaluate the impact of other exogenous stimuli on the transcriptional age, we applied the age predictor to lung transcriptomes generated from mice exposed to different doses of thoracic irradiation (0 Gy, 5 Gy, and 17.5 Gy) [20]. The model accurately predicted age progression in the 3 experimental groups (Fig. 3c). Mice exposed to a fibrogenic dose of irradiation (17.5 Gy) were predicted to be prematurely aged in our model compared to non-fibrogenic treatments (Fig. 3d and e), in agreement with the original observations made by Citrin et al., where the 17.5 Gy irradiation dose, but not the 5 Gy dose, induced lung fibrosis, cellular senescence, and reduced survival in addition to transcriptional variations mimicking natural aging in irradiated young mice [20]. Surprisingly, the 5 Gy-exposed mice were, on average, predicted to be younger than the untreated mice, but with an increased variability in the prediction outputs.

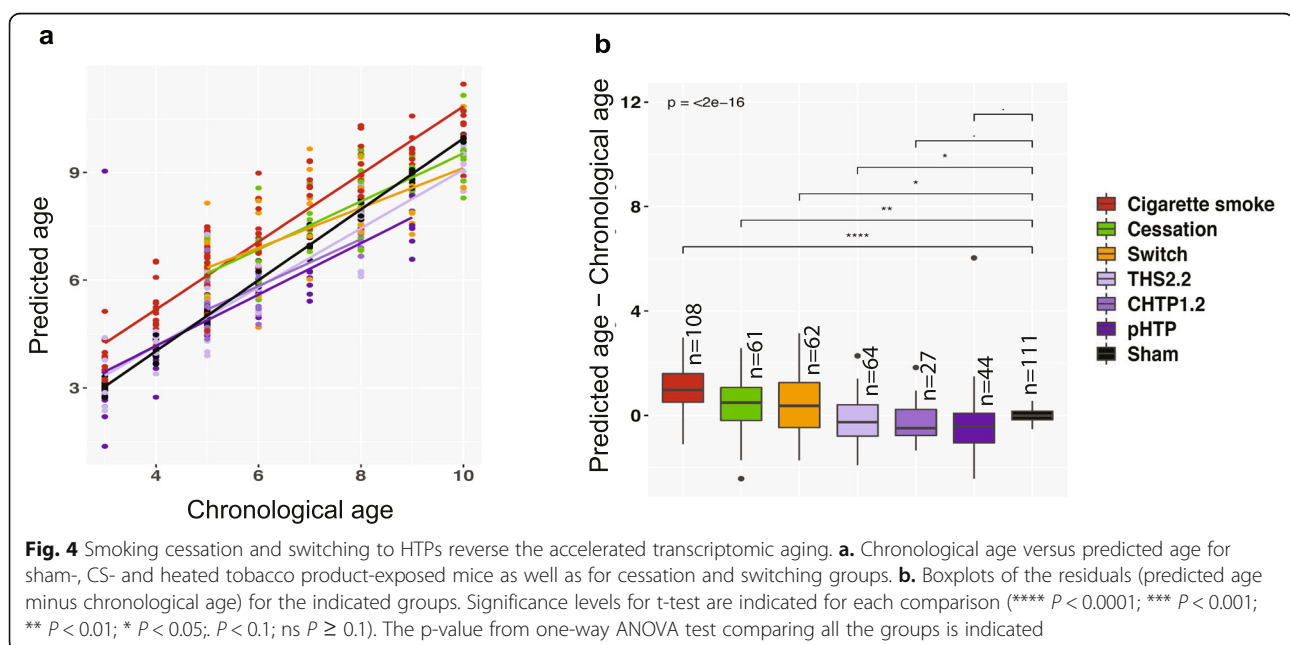
Smoking cessation and switching to HTPs reduce CS-associated accelerated transcriptional aging

We then evaluated the impact of smoking cessation or switching to potentially less harmful tobacco products on transcriptional premature aging. We applied our prediction model to transcriptomes generated from mice exposed to HTP aerosols under conditions matching exposure time and nicotine concentration of CS-exposed

groups as well as to transcriptomes from cessation and switching groups (Supplementary Figure 1). At early time points, the transcriptomic age for cessation and switching groups was close to that of mice continuously exposed to CS and rapidly decreased to reach the range of age predictions for sham-exposed mice (Fig. 4a). This suggests that the CS-associated premature aging can be attenuated upon smoking cessation or switching to reduced-exposure products. In contrast to CS, chronic exposure to HTP aerosols correlates with weak deviations of the predicted age from the chronological age (Fig. 4a and b).

Discussion

Modulation of gene expression is an important aspect of cellular adaptation to external factors, such as environmental exposures, but also to intrinsic physiological cues, including those associated with natural aging. The gene expression pattern measured at a given time reflects the cellular response to various concurrent stimuli, and this response is very likely to be affected by past events that might have produced stable genetic or epigenetic alterations. Thus, decoupling transcriptional changes associated with age from those associated with environmental exposures is challenging. Long-term exposure studies using animal models provide an ideal experimental setup to independently investigate transcriptional variations in response to different exposures as well as the variations associated with natural aging. Such investigations require large datasets generated with standardized procedures and covering a reasonable age range.



In the current work, we aimed to identify key transcriptomic markers for lung aging and investigate a potential association with CS exposure using three publicly available transcriptomic datasets generated from murine lung tissue exposed to CS and HTPs. Our investigations revealed a strong overlap between age- and CS-regulated genes in the lung. Age- and CS-associated transcriptional changes follow the same direction for some genes and adopt an opposite direction for others, demonstrating the complexity of transcriptional interactions between aging and exposure effects. One of the most striking observations was the strong upregulation of immunoglobulin genes, expression of which is exclusively restricted to B-cell lineage. This observation suggests that when exposed to CS or with increased age, the lung parenchyma is infiltrated by a significant number of B-cells. This observation is in line with previous studies reporting the presence of various B-cell structures in inflamed lungs. B-cell-rich lymphoid follicles were demonstrated in lung parenchyma of mice exposed to CS and in COPD patients with emphysema [34]. B-cells have also been reported to be present as part of the germinal centers within the parenchyma of inflamed mouse lungs and to produce local immunoglobulin E (IgE) isotype in response to antigenic challenge [35]. CS exposure has been shown to increase the volume of inducible bronchus-associated lymphoid tissue (iBALT) in aged mice in contrast to young mice [36] further supporting a higher susceptibility of aged tissues to environmentally induced inflammation. The iBALT volume seems to increase with age independently of CS exposure [36] suggesting that CS enhances a naturally occurring iBALT formation. Consistent with these results, the B-cell chemoattractant *Cxcl13*, known to be involved in lymphoid neogenesis in COPD [37] is upregulated with increased age and in response to CS exposure in our model (Supplementary Figure 2).

Further investigations of lymphoid infiltrates in aged lungs are required to better understand the contribution of B-cells to the age-associated inflammation. At the analytical level, the high expression of B-cell-specific genes in lung transcriptomes raises the question about the contribution of infiltrating immune cells to the measured gene expression or other molecular signals in the investigated tissues.

Over the last decade, a number of age predictors have been derived from methylomes [9, 27] transcriptomes [10, 38] and proteomes [39, 40]. These omics-based models sought to select a set of elements (e.g., CpGs, genes, proteins, etc.) that may predict chronological age and, eventually, the deviations associated with external factors. Although these reports provide valuable information about age-related molecular changes, they suffer from a number of limitations, including the interference

of environmental factors and technical variabilities due to different platforms.

In the current work, we built an age predictor using a technically homogenous set of transcriptomes generated from untreated mice to minimize exposure and technical variability biases. We applied the LASSO regression algorithm that selects a limited number of regressor variables (genes) required for an optimal association with the response variable (age). This approach resulted in a set of 57 genes that can predicted chronological age with an average error of 0.83 months in the validation set. Interestingly, the majority of the age predictor genes were deregulated by CS, further supporting a strong connection between CS exposure and lung aging.

The predicted transcriptomic age from healthy untreated mice showed a high correlation with chronological age and can therefore be considered as a surrogate of natural biological age. Deviations from this standard aging curve may reflect perturbations in the natural aging rate. A positive delta age (predicted age minus chronological age) can be indicative of premature aging, whereas a negative delta age may indicate decelerated aging. To test this hypothesis, we applied our age predictor to transcriptomes from CS, HTPs, cessation, and switching groups. CS-exposed mice showed a consistent positive delta age, suggesting premature lung aging, in contrast to HTP-exposed mice, that showed a weak deviation from chronological age. Interestingly, transcriptomic age for cessation and switching groups rapidly deviated from that of CS-exposed group to reach that of the sham group.

Our prediction model was able to accurately predict age progression from an independent set of lung transcriptomes generated from mice irradiated to 5 Gy or 17.5 Gy and their untreated counterparts. Although we observed a positive delta age for the 3 experimental groups, the fibrogenic irradiation dose (17.5 Gy) had the highest predicted age, in line with the original conclusions made by Citrin et al., where the 17.5 Gy irradiation induced fibrosis and age-related transcriptomic alterations. By contrast, the untreated and 5 Gy-irradiated mice showed a weak delta age.

Conclusions

The current work presents a comprehensive characterization of transcriptional interactions between cigarette smoke exposure and lung aging. We have provided a proof of concept that a machine learning approach applied to a technically homogenous set of transcriptomes can be used to model natural aging progression and detect deviations associated with environmental exposures. We also demonstrate the contribution of infiltrating immune cells to the measured gene expression in the investigated tissues. Additionally, we show

that reducing harmful exposures may attenuate premature aging of lung tissue.

Methods

Microarray data processing

Raw data files were processed with the custom Chip Description File environment Mouse4302_Mm_ENTREZG v16.0 [41] and normalized using frozen robust microarray analysis [42].

Identification of differentially expressed genes

Differentially expressed genes were identified based on the Benjamini–Hochberg FDR-adjusted p -values (cutoff 0.05) and fold changes (cutoff $\log_2(1.5)$) generated for the control-case contrasts with the *limma* R-package [43]. To define gene expression changes over age, expression values at a given time point were compared to those from all other time points within the same study for sham-exposed samples. For each contrast, the fold change is calculated as the expression value (\log_2 scale) in the oldest group minus the expression value in the youngest group (i.e., E-MTAB-5280 Sham_8m minus E-MTAB-5280 Sham_1m). Age-regulated genes were identified as genes meeting fold change and p -value criteria in at least one contrast. Given that 2 datasets used in this work are generated from ApoE^{-/-} mice, the ApoE gene was excluded from the entire analysis.

Construction of transcriptomic age predictor

A LASSO model (implemented in the R-package *glmnet* [44]) was used to model chronological age (response variable) as a function of gene expression values (predictor variables) from sham samples. Lasso is a regression method that aims at selecting a limited subset of predictor variables through regularization while minimizing the prediction errors [45]. During the training step, the LASSO algorithm selects a set of regressors (genes) that optimally predict the dependent variable (age) via several rounds of internal cross-validations. LASSO is known to exclude correlated and non-informative regressors, thereby reducing the size of the predictor set. The set of genes used as age predictors was selected based the regularization parameter, leading to the minimum mean squared error value upon internal 5-fold cross-validations. The final model consisted of a linear regression of age as a function of the expression values of predictor genes selected by LASSO.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12864-020-6712-z>.

Additional file 1: Figure S1. Study design for the 3 datasets used in the current work. Sham corresponds to fresh air and is used as the

exposure control. 3R4F corresponds to the standard reference cigarette, THS 2.2 and CHTP 1.2 correspond to candidate HTPs, and pHTP corresponds to prototype HTP. THS 2.2, Tobacco Heating System 2.2; CHTP 1.2, Carbon Heated Tobacco Product 1.2. The tables list the number of samples per experimental group; “m” stands for months of exposure. CS, cigarette smoke. ArrayExpress identifiers are indicated. All these studies used female mice.

Additional file 2: Figure S2. Mean gene expression values for *Ighg2b*, *Ighg1*, *Igha*, and *Cxcl13* genes to illustrate the expression profile of B-cell-associated genes during aging. “m” in x-axis labels stands for months and indicates the exposure time. All animals were 2 months old at the beginning of the exposure. Error bars represent the standard error of the mean. Significance levels for ANOVA test are indicated for each group (**** $P < 0.0001$; *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$; . $P < 0.1$; ns $P \geq 0.1$)

Additional file 3.

Abbreviations

ApoE^{-/-}: Apolipoprotein E-deficient; CHTP 1.2: Carbon Heated Tobacco Product 1.2; COPD: Chronic obstructive pulmonary disease; CS: Cigarette smoke; FC: Fold changes; FDR: False discovery rate; HTP: Heated tobacco product; iBALT: Inducible bronchus-associated lymphoid tissue; Ig: Immunoglobulin; LASSO: Least absolute shrinkage and selection operator; MFC: Maximum fold change; MMP: Matrix metalloproteinases; THS 2.2: Tobacco Heating System 2.2

Acknowledgements

Not applicable.

Authors' contributions

FM and MAC conceived the study and performed the computational work. FM supervised the computational framework and developed the prediction methodology. MAC wrote the manuscript with feedback from JH, MP, and FM. All authors discussed the results and contributed to the final manuscript. The author(s) read and approved the final manuscript.

Funding

Philip Morris International is the sole source of funding and sponsor of this work. The study design, data collection, analysis and interpretations presented in this article are solely those of the authors, and the article was prepared exclusively by the authors.

Availability of data and materials

The datasets analyzed during the current study are available in the ArrayExpress and GEO repositories: <https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-5281> <https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-5280> <https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-7444> <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE41789> <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE18344>

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

All authors are employees of Philip Morris International.

Received: 30 August 2019 Accepted: 31 March 2020

Published online: 09 April 2020

References

1. Pal S, Tyler JK. Epigenetics and aging. *Sci Adv* [Internet]. 2016;2(7) Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4966880/> [cited 2019 Feb 15].
2. Lodato MA, Rodin RE, Bohrsen CL, Coulter ME, Barton AR, Kwon M, et al. Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science*. 2018;359(6375):555–9.

3. Forster MJ, Lal H. Estimating age-related changes in psychomotor function: influence of practice and of level of caloric intake in different genotypes. *Neurobiol Aging*. 1999;20(2):167–76.
4. Gruber MP, Coldren CD, Woolum MD, Cosgrove GP, Zeng C, Barón AE, et al. Human lung project: evaluating variance of gene expression in the human lung. *Am J Respir Cell Mol Biol*. 2006;35(1):65–71.
5. Harris SE, Riggio V, Evenden L, Gilchrist T, McCafferty S, Murphy L, et al. Age-related gene expression changes, and transcriptome wide association study of physical and cognitive aging traits, in the Lothian birth cohort 1936. *Aging (Albany NY)*. 2017;9(12):2489–503.
6. Zahn JM, Poosala S, Owen AB, Ingram DK, Lustig A, Carter A, et al. AGEMAP: a gene expression database for aging in mice. *PLoS Genet*. 2007;3(11):e201.
7. Morita A. Tobacco smoke causes premature skin aging. *J Dermatol Sci*. 2007;48(3):169–75.
8. Weindruch R, Walford RL, Fligiel S, Guthrie D. The retardation of aging in mice by dietary restriction: longevity, cancer, immunity and lifetime energy intake. *J Nutr*. 1986;116(4):641–54.
9. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. 2013;14(10):R115.
10. Peters MJ, Joehanes R, Pilling LC, Schurmann C, Conneely KN, Powell J, et al. The transcriptional landscape of age in human peripheral blood. *Nat Commun*. 2015;6:8570.
11. Weidner CI, Lin Q, Koch CM, Eisele L, Beier F, Ziegler P, et al. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol*. 2014;15(2):R24.
12. Pan F, Chiu C-H, Pularupa S, Mehan MR, Nunez-Iglesias J, Zhang K, et al. Gene Aging Nexus: a web database and data mining platform for microarray data on aging. *Nucleic Acids Res*. 2007;35:D756–9.
13. de Magalhães JP, Curado J, Church GM. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics*. 2009;25(7):875–81.
14. Misra V, Lee H, Singh A, Huang K, Thimmulappa RK, Mitzner W, et al. Global expression profiles from C57BL/6J and DBA/2J mouse lungs to determine aging-related genes. *Physiol Genomics*. 2007;31(3):429–40.
15. Calabresi C, Arosio B, Galimberti L, Scanziani E, Bergottini R, Annoni G, et al. Natural aging, expression of fibrosis-related genes and collagen deposition in rat lung. *Exp Gerontol*. 2007;42(10):1003–11.
16. Pennings JLA, Mariman R, Hodemaekers HM, Reemers SSN, Janssen R, Guichelaar T. Transcriptomics in lung tissue upon respiratory syncytial virus infection reveals aging as important modulator of immune activation and matrix maintenance. *Sci Rep*. 2018;8(1):16653.
17. Raghu G, Weycker D, Edelsberg J, Bradford WZ, Oster G. Thannickal. *Am J Respir Crit Care Med*. 2006;174(7):810–6.
18. Thannickal VJ. Mechanistic Links between Aging and Lung Fibrosis. *Biogerontology* [Internet]. 2013;14(6) Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3852192/> [cited 2019 Feb 18].
19. Meyer KC, Ershler W, Rosenthal NS, Lu XG, Peterson K. Immune dysregulation in the aging human lung. *Am J Respir Crit Care Med*. 1996; 153(3):1072–9.
20. Citrin DE, Shankavaram U, Horton JA, Shield W, Zhao S, Asano H, et al. Role of type II Pneumocyte senescence in radiation-induced lung fibrosis. *J Natl Cancer Inst*. 2013;105(19):1474–84.
21. Madan R, Matalon S, Viviero M. Spectrum of smoking-related lung diseases: imaging review and update. *J Thorac Imaging*. 2016;31(2):78–91.
22. Vij N, Chandramani-Shivalingappa P, Van Westphal C, Hole R, Bodas M. Cigarette smoke-induced autophagy impairment accelerates lung aging, COPD-emphysema exacerbations and pathogenesis. *Am J Physiol Cell Physiol*. 2018;314(1):C73–87.
23. Phillips B, Veljkovic E, Peck MJ, Buettner A, Elamin A, Guedj E, et al. A 7-month cigarette smoke inhalation study in C57BL/6 mice demonstrates reduced lung inflammation and emphysema following smoking cessation or aerosol exposure from a prototypic modified risk tobacco product. *Food Chem Toxicol*. 2015;80:328–45.
24. Phillips B, Veljkovic E, Boue S, Schlage WK, Vuillaume G, Martin F, et al. An 8-month systems toxicology inhalation/cessation study in ApoE^{-/-} mice to investigate cardiovascular and respiratory exposure effects of a candidate modified risk tobacco product, THS 2.2, compared with conventional cigarettes. *Toxicol Sci*. 2016;149:411–32.
25. Phillips B, Szostak J, Titz B, Schlage WK, Guedj E, Leroy P, et al. A six-month systems toxicology inhalation/cessation study in ApoE^{-/-} mice to investigate cardiovascular and respiratory exposure effects of modified risk tobacco products, CHTP 1.2 and THS 2.2, compared with conventional cigarettes. *Food Chem Toxicol*. 2019;126:113–41.
26. Bocklandt S, Lin W, Sehl ME, Sánchez FJ, Sinsheimer JS, Horvath S, et al. Epigenetic Predictor of Age. *PLoS One* [Internet]. 2011;6(6) Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3120753/> [cited 2019 Mar 4].
27. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada S, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell*. 2013 Jan 24;49(2):359–67.
28. Multi-tissue DNA methylation age predictor in mouse. - PubMed - NCBI [Internet]. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/28399939> [cited 2019 Apr 11].
29. Thompson MJ, Chwiałkowska K, Rubbi L, Lusis AJ, Davis RC, Srivastava A, et al. A multi-tissue full lifespan epigenetic clock for mice. *Aging (Albany NY)*. 2018;10(10):2832–54.
30. Petkovich DA, Podolskiy DI, Lobanov AV, Lee S-G, Miller RA, Gladyshev VN. Using DNA Methylation Profiling to Evaluate Biological Age and Longevity Interventions. *Cell Metab*. 2017;25(4):954–60 e6.
31. Meer MV, Podolskiy DI, Tyshkovskiy A, Gladyshev VN. A whole lifespan mouse multi-tissue DNA methylation clock. *Elife*. 2018;14:7.
32. Epigenetic aging signatures in mice livers are slowed by dwarfism, calorie restriction and rapamycin treatment | *Genome Biology* | Full Text [Internet]. [cited 2019 Aug 12]. Available from: <https://genomebiology.biomedcentral.com/articles/> <https://doi.org/10.1186/s13059-017-1186-2>.
33. Gebel S, Diehl S, Pype J, Friedrichs B, Weiler H, Schüller J, et al. The transcriptome of Nrf2^{-/-} mice provides evidence for impaired cell cycle progression in the development of cigarette smoke-induced emphysematous changes. *Toxicol Sci*. 2010;115(1):238–52.
34. van der Strate BWA, Postma DS, Brandsma C-A, Melgert BN, Luinge MA, Geerlings M, et al. Cigarette smoke-induced emphysema: a role for the B cell? *Am J Respir Crit Care Med*. 2006;173(7):751–8.
35. Chvatchko Y, Kosco-Vilbois MH, Herren S, Lefort J, Bonnefoy JY. Germinal center formation and local immunoglobulin E (IgE) production in the lung after an airway antigenic challenge. *J Exp Med*. 1996;184(6):2353–60.
36. John-Schuster G, Günter S, Hager K, Conlon TM, Eickelberg O, Yildirim AO. Inflammation increases susceptibility to cigarette smoke-induced COPD. *Oncotarget*. 2016;7(21):30068–83.
37. Litsiou E, Semitekolou M, Galani IE, Morianos I, Tsoutsas A, Kara P, et al. CXCL13 production in B cells via toll-like receptor/lymphotoxin receptor signaling is involved in lymphoid neogenesis in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2013;187(11):1194–202.
38. Holly AC, Melzer D, Pilling LC, Henley W, Hernandez DG, Singleton AB, et al. Towards a gene expression biomarker set for human biological age. *Aging Cell*. 2013;12(2):324–6.
39. Krištić J, Vučković F, Menni C, Klarić L, Keser T, Beceheli I, et al. Glycans are a novel biomarker of chronological and biological ages. *J Gerontol A Biol Sci Med Sci*. 2014;69(7):779–89.
40. Menni C, Kastenmüller G, Petersen AK, Bell JT, Psatha M, Tsai P-C, et al. Metabolomic markers reveal novel pathways of ageing and early development in human populations. *Int J Epidemiol*. 2013;42(4):1111–9.
41. Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, et al. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res*. 2005;33(20):e175.
42. McCall MN, Bolstad BM, Irizarry RA. Frozen robust multichannel analysis (FRMA). *Biostatistics*. 2010 Apr;11(2):242–53.
43. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. 2004;3:Article3. <https://doi.org/10.2202/1544-6115.1027>.
44. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1–22.
45. Tibshirani R. Regression shrinkage and selection via the Lasso. *J R Stat Soc Ser B Methodol*. 1996;58(1):267–88.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.