BIODEMOGRAPHY AND SOCIAL BIOLOGY
2021, VOL. 66, NOS. 3–4, 236–249
https://doi.org/10.1080/19485565.2021.1983761

Routledge
Taylor & Francis Group

Demographic correlates of inflammatory and antiviral gene expression in the study of Midlife in the United States (MIDUS)

Frank D. Mann[a], Robert F. Krueger[b], Sean Clouston[a], and Steven Cole[c]

[a]Program in Public Health and the Department of Family, Population, and Preventive Medicine, Stony Brook University, Stony Brook, New York, USA; [b]Department of Psychology, University of Minnesota, Minneapolis, Minnesota, USA; [c]Department of Psychiatry & Biobehavioral Sciences and Medicine, University of California, Los Angeles, California, USA

ABSTRACT

The present study examined the demographic correlates of gene expression in a sample of adults (n = 543) from the Study of Midlife in the United States (MIDUS). Inflammatory and antiviral gene sets were operationalized using *a priori* composite scores and empirically derived co-regulatory gene sets. For both composite scores and co-regulatory gene sets, White/European Americans showed lower while Black/African Americans showed higher expression of genes involved in interferon responses and antibody synthesis. The effects of chronological age on gene expression varied by sex, such that pro-inflammatory gene expression increased with age more rapidly for females than males. The difference between the average expression of inflammatory and antiviral genes also increased with age for females but not males. Results shed light on differential gene expression as a potential physiological correlate for race/ethnicity, age, and sex-related health disparities in adulthood.

Introduction

The prevalence and sequalae of physical and mental health problems are often heterogenous across demographic strata and tend to change with age. Demographic differences (including differences by age, race/ethnicity, and sex) are well documented for physical health problems, including cardiovascular disease (Mensah et al. 2005), adiposity (Wang and Beydoun 2007), and diabetes (Kautzky-Willer, Harreiter, and Pacini 2016; Spanakis and Golden 2013). DSM-defined psychiatric disorders also vary across lifespan development, as well as sex and race/ethnicity (American Psychiatric Association 2013). Sex differences in mortality due to parasitic and infectious disease vary across the lifespan, with higher rates for males peaking around midlife (Owens 2002). Moreover, sex differences in chronic disease incidence persist in late life and culminate in an average 6–8 year greater life expectancy for females relative to males (Austad and Fischer 2016). Mechanisms for these inequalities are not well understood but may include differences in parasitic exposures (Owens 2002), the impact of testosterone, estradiol, and progesterone on immunological response (Foo et al. 2017; Klein 2004), as well as social, medical, and economic causes (Phelan and Link 2015). These inequalities may also grow over time as exposures are

CONTACT Frank D. Mann frank.mann@stonybrookmedicine.edu Program in Public Health and the Department of Family, Population, and Preventive Medicine, Stony Brook University, 101 Nicolls Road, Stony Brook, NY 11794

amplified by new exposures (LeBrón et al. 2019). Nevertheless, the large range of outcomes over which age, sex, and racial/ethnic differences emerge is perplexing – how might we expect the same disparities to be sensitive to exposures consistently emerging across a range of disease outcomes?

Trying to understand how social factors might cause the body to "break down" over time, researchers are increasingly turning to novel measures, such as genetic expression, to identify potentially actionable cellular changes that mediate the effects of these exposures. Notably, with respect to race/ethnicity, Black Americans have the highest all-cause mortality rates compared to any other racial/ethnic group (Williams and Mohammed 2009). Consistent with the view that molecular changes in genome function might be a crucial mediator of social phenomena, a recent study found sex, age, and race/ethnic differences in the expression of inflammatory and antiviral genes using data from the National Longitudinal Study of Adolescent Health (n = 1069; Cole et al. 2020). In that study, genes related to type-I interferon response were upregulated for Black participants relative to non-Hispanic White, and higher levels of inflammatory gene expression were observed in females relative to males. Supporting the view that these disparities emerged via cellular mechanisms rather than the effects of disease, these results were identified in the Add Health study, which surveyed gene expression in early adulthood (at approximately 30 years of age), before the presence of disease outcomes are routinely recognized.

One potential mechanism that may translate distal risk factors into proximal molecular determinants of health is a gene regulatory pathway known as the conserved transcriptional response to adversity (CTRA; Cole 2019). The CTRA describes a coordinated suite of nervous system-activated inflammatory gene expression, with a simultaneous downregulation of antiviral genes involved in Type I interferon responses and antibody synthesis (Cole 2013). Hypothesized to have emerged over the course of evolution, this genomic stress response may have been adaptive under ancestral conditions, enabling the immune system to combat bacterial infection and expedite recovery from physical injury (Cole 2019). However, increased inflammatory gene expression and the CTRA have been found to be associated with social isolation and other adverse life circumstances in modern populations (Cole et al. 2015), even after controlling for indicators of physical health and health-risk behaviors, suggesting that the CTRA may be activated by stressors that cause subjective stress and social impairment. To the extent that demographic factors are associated with differential rates of exposure to social stressors, as well as pathogens and risk of physical injury, the CTRA may serve as a physiological mechanism that helps explain why demographic factors become biologically imbedded to influence social gradients in health.

Although longitudinal studies will ultimately be needed to map the role of stress-regulated gene expression dynamics as a mediator of demographic disparities in health, as well as highly controlled experimental work, an important first step is to identify the major gene expression correlates of health-relevant demographic factors. As noted above, this work has already identified differences between racial/ethnic groups and by sex. However, these results also identified age-based differences, raising the possibility that some demographic differences might potentially involve differential rates of aging. While studies often focus on the marginal effects of age and sex, the trend toward growing differences between sexes in mortality later in life highlights the possibility that the molecular mechanisms may be moderated by chronological age. Yet, to date, no studies have examined whether CTRA mechanisms are moderated by age. Consequently, the goals of the current study are twofold.

First, we plan to replicate prior work showing that age, sex, and race/ethnicity are associated with inflammatory, antiviral, and CTRA gene expression. Second, leveraging the heterogeneity of age in the present study, we interrogate whether sex differences in gene expression vary with age. Notably, the present study addresses the question of demographic differences in gene expression at a time in lifespan development when chronic diseases often begin to manifest clinically, and in a sample with greater age variation than available in the previous Add Health population-based transcriptome analysis.

Based on recent findings (Cole et al. 2020), we expect greater antiviral expression among Black participants, relative to other races/ethnicities. We also expect that females will have higher inflammatory expression than males, and antiviral and inflammatory expression will increase with age. How age-related differences in expression may differ for females and males is less clear. Although inflammatory expression is higher for females in early adulthood before the emergence of many age-related diseases (Cole et al. 2020), inflammatory expression may increase more rapidly with age for males, dovetailing with higher incidence of chronic disease and shorter life expectancy. Similarly, greater antiviral expression may be expected among males as they grow older, relative to females, because males tend to have greater parasitic exposure over the life course and higher mortality due to parasitic and infectious disease. On the other hand, the immunomodulatory effects of testosterone (Foo et al. 2017) may accumulate over time and, in turn, decrease antiviral expression in males as they grow older. In sum, there are reasons to expect that age-related differences in gene expression may differ for females and males, but contradictory mechanistic hypotheses render the expected pattern of differences difficult to predict.

Method

Sample

The present study analyzed data from the biomarker subsample from the Refresher Cohort of the Study of Midlife in the United States (MIDUS; n = 863). Information regarding participant recruitment, study design, and data collection can be found elsewhere (Ryff and Krueger 2018). The present study utilized data from participants from the Biomarker sample, who have data available for gene expression composite scores (n = 543). Sample characteristics are reported in Table 1. The age of participants spanned 27 to 76 years (mean = 52 years). Approximately half of the sample identified as female (49.53% male). The majority of the sample identified as White (~72%), ~18% of the sample identified as Black and ~10% identified as another race/ethnicity (see Table 1). The most common level of educational attainment among participants was a bachelor's degree (~25%).

Measures

Demographic, Behavioral & Physical Health Factors

Age was measured at the time of data collection. Participants reported sex (female or male), level of education, and race/ethnicity. Health-risk behavior was measured using participants' self-reports of smoking behavior and alcohol consumption. Participants responded to "Have you ever smoked cigarettes regularly – that is, at least a few cigarettes every day?"

Table 1. Descriptive statistics for demographic factors & health indicators.

MIDUS Refresher Biomarker Project	Full Sample (n = 863)	Analytic Sample Combined (n = 543)	Analytic Sample Female (n = 274)	Analytic Sample Male (n = 269)	p-value Sample Diff.	p-value Sex Diff.
Age	50.84 (13.41)	52.00 (13.26)	50.55 (12.50)	53.48 (13.87)	.001	.005
BMI	29.24 (7.22)	29.02 (7.12)	29.47 (8.51)	28.55 (5.29)	.228	.781
Chronic Conditions	2.80 (2.94)	2.78 (2.77)	3.19 (2.94)	2.36 (2.53)	.851	<.001
Race/Ethnicity					.067	<.001
White/European American	606 (70.22)	395 (72.74)	181 (66.06)	214 (79.55)		
Black/African American	161 (18.66)	86 (15.84)	60 (21.90)	26 (9.67)		
Native American	22 (2.55)	17 (3.13)	12 (4.38)	5 (1.86)		
Asian American	12 (1.39)	8 (1.47)	3 (1.09)	5 (1.86)		
Pacific Islander American	2 (0.23)	1 (0.18)	0 (0.00)	1 (0.37)		
Another Race/Ethnicity	55 (6.37)	34 (6.26)	17 (6.20)	17 (6.32)		
Don't Know or Refused	5 (0.58)	2 (0.37)	1 (0.36)	1 (0.37)		
Level of Education					.923	.139
Eighth Grade/Junior High	5 (0.58)	3 (0.55)	3 (1.10)	0 (0.00)		
Some High School	30 (3.48)	18 (3.32)	10 (3.66)	8 (2.97)		
General Education Diploma	12 (1.39)	8 (1.48)	6 (2.60)	2 (0.74)		
High School Diploma	102 (11.83)	65 (11.99)	33 (12.09)	32 (11.90)		
1–2 years college, No Degree	132 (15.31)	81 (14.95)	40 (14.65)	41 (15.24)		
>2 years college, No Degree	39 (4.52)	25 (4.61)	13 (4.76)	12 (4.46)		
Associate/Technical Degree	92 (10.67)	55 (10.15)	28 (10.26)	27 (10.04)		
Bachelor's Degree	211 (24.48)	137 (25.28)	72 (26.37)	65 (24.16)		
Some Graduate School	28 (3.25)	17 (3.14)	8 (2.93)	9 (3.35)		
Master's Degree	165 (19.14)	109 (20.11)	52 (19.05)	57 (21.19)		
J.D., M.D., Ph.D.	46 (5.34)	24 (4.43)	8 (2.93)	16 (5.95)		
Don't know or Refused	1 (0.12)	1 (0.18)	1 (0.36)	0 (0.00)		
Alcohol Consumption					.671	<.001
Never	258 (29.90)	169 (31.12)	110 (40.15)	59 (21.93)		
< 1 day a week	230 (26.65)	139 (25.60)	73 (26.64)	66 (24.54)		
1–2 days a week	154 (17.84)	95 (17.50)	44 (16.06)	51 (18.96)		
3–4 days a week	111 (12.86)	69 (12.71)	20 (7.30)	49 (18.22)		
5–6 days a week	58 (6.72)	37 (6.81)	18 (6.57)	19 (7.06)		
Everyday	52 (6.03)	34 (6.26)	9 (3.28)	25 (9.29)		
History of Smoking					.268	.602
Yes	342 (39.63)	207 (38.12)	101 (36.86)	106 (39.41)		
No	521 (60.37)	336 (61.88)	173 (63.14)	163 (60.59)		

Means and standard deviations or frequencies and percentages are reported. Before testing for sample and sex differences, two models were compared to determine whether level of education and alcohol consumption should be treated as nominal (multi-nominal) or ordinal (multi-ordinal). Information criteria (AIC & BIC) indicated that the multi-ordinal models were preferred, such that that both variables should be rank-ordered. *P*-values are reported for Mann-Whitney U tests with continuity correction for ordinal/interval variables and Pearson's $\chi 2$ or Fisher's Exact test for nominal variables.

(Yes = 38%, No = 62%). Participants also reported "During the past month, how often did you drink any alcoholic beverages, on the average?" (57% = Never or less than one day a week). Indicators of physical health included body mass index (median BMI = 27.69) and self-reports of chronic conditions that participants had been diagnosed with or treated for in the past 12 months. The MIDUS administrative staff created a composite score of the total number of chronic conditions reported by participants (~21% reported no conditions). Detailed descriptive statistics for continuous variables are reported in supplemental materials (Table S1).

Gene Expression

Participants in the MIDUS biomarker project had blood samples collected during a laboratory visit. After collection, samples were shipped to the University of Wisconsin for storage at a central biospecimen repository. Peripheral blood mononuclear cells (PBMCs) were isolated by standard Ficol density gradient centrifugation and frozen at −70°C for several years prior to RNA extraction. PBMC samples were then thawed and total RNA was extracted using standard protocols (Qiagen RNeasy). Extracted RNA was checked for suitable integrity (RNA integrity number > 3) and yield (> 50 ng), and subject to transcriptome profiling using a high efficiency mRNA-targeted sequencing approach (Lexogen QuantSeq 3′ FWD cDNA library synthesis with sequencing on an Illumina HiSeq 4000 instrument), which targeted >10 million 65-bp single stranded sequencing reads per sample (achieved mean = 14.5 million). Quality-annotated FASTQ sequence reads were mapped to the ENSEMBL hg38 reference human transcriptome using the STAR aligner (mean 97.2% of reads successfully mapped) and quantified as gene transcripts per million mapped reads (with values < 1 transcript per million floored at that value to suppress spurious measurement variability and allow log transformation). RNA sequencing was conducted in batches comprised of a 96-well plate, which was indicated by a nominal variable and included as a technical covariate in analyses that tested associations with demographic factors. Two sets of genes were selected *a priori* based on their previous use in research: 19 inflammatory genes (*IL1A, IL1B, IL6, IL8, TNF, PTGS1, PTGS2, FOS, FOSB, FOSL1, FOSL2, JUN, JUNB, JUND, NFKB1, NFKB2, REL, RELA*, & *RELB*) and 32 antiviral genes involved in Type I interferon responses and antibody production (*GBP1, IFI16, IFI27, IFI27L1- 2, IFI30, IFI35, IFI44, IFI44L, IFI6, IFIH1, IFIT1-3, IFIT5, IFIT1L, IFITM1-3, IFITM4P, IFITM5, IFNB1, IRF2, IRF7-8, MX1- 2, OAS1-3, OASL, IGJ, IGLL1*, & *IGLL3*).

A priori gene composite scores were calculated because CTRA gene expression involves the increased expression of inflammatory genes and decreased expression of genes involved in interferon responses and antibody production. Consequently, composite scores were calculated to capture the average expression of inflammatory genes, the average expression of antiviral genes, and the difference between the average expression of inflammatory and antiviral genes (i.e., a CTRA composite). As the variance and average of RNA expression are heterogenous across different genes, to prevent arithmetic means from being predominately weighted by a small number of genes, expression values were log2 transformed and standardized before calculating mean composite scores.

Data Analytic Procedures

Data were analyzed using R Studio version 1.2.5003 and Mplus (Muthén and Muthen 2017). After calculating descriptive statistics, a series of ordinary least squares regressions were used to estimate associations between demographic factors (age, sex, race/ethnicity, and level of education) and *a priori* gene expression composite scores, controlling for dummy-coded batch plates, RNA transcripts indicting the relative prevalence of T lymphocytes, B lymphocytes, NK cells, and monocytes. Alcohol use, history of smoking, BMI, and number of chronic conditions were also included as covariates. The Benjamini-Hochberg procedure was then used to correct for multiple independent variables within each regression model (Benjamini and Hochberg 1995). For multiple

regressions, age was mean-centered, and level of education was standardized (M = 0, SD = 1). Sex (male = −0.5, female = 0.5), Black (No = 0, Yes = 1) and other races/ethnicities (No = 0, Yes = 1) were included as predictors (with White/European Americans as the reference group). In addition, a product term between mean-centered age and effects-coded sex was included to test whether age-related differences in gene expression vary as a function of sex.

The distributions of gene expression composite scores approximated normality, but deviations from normality were nonetheless observed. Therefore, sensitivity analyses were performed whereby multiple linear regressions were estimated using maximum likelihood with robust standard errors (MLR; Muthén and Muthen 2017), and the results of hypothesis tests remained unchanged, as did the size and precision of estimated associations (see supplement). Note, including number of chronic conditions as a covariate may constitute a statistical over-control, as health conditions are one outcome that gene expression is hypothesized to mediate, but results also remained unchanged after excluding this potentially over-controlling covariate (Table S2).

Although composite scores provide a simple *a priori* method for operationalizing the CTRA, different genes may exhibit a correlated pattern of expression that is not fully captured by composite scores. Therefore, to identify empirically derived co-regulatory gene sets, an exploratory factor analysis (EFA) of inflammatory and antiviral gene expression was conducted in conjunction with a parallel analysis using 1000 random draws (Montanelli and Humphreys 1976). Before conducting EFA, quantile-normalized gene expression values were log2-transformed and standardized within gene. RNA sequencing can yield an excess of zeros. Therefore, prior to conducting EFA, the distributions of gene transcripts were screened, and descriptive statistics were calculated to exclude data from any gene with an average expression level < .5 log2 transcripts per million mapped reads (TPM- i.e., the native value of the normalized expression data), which resulted in the deletion of data from 9 genes (*CXCL8, IL1A, FOSL1, IL6, IFI27, IFITM4P, IFITM5, IFNB1, IGLL1*).

EFA models were estimated using Mplus version 8.1 (Muthén and Muthen 2017). One through seven factor solutions were estimated, and the best-fitting solution was determined by consulting model fit statistics (Table S5). We also compared the scree plots from the EFA and parallel analysis (Figure S1). After determining the number of co-regulatory factors needed to best characterize correlated expression across multiple transcripts, associations with study variables were estimated using a series of exploratory structural equation models (ESEMs; Asparouhov and Muthén 2009), whereby demographic factors, BMI, alcohol-use, and history of smoking were specified as exogenous predictors of co-regulatory factors, while simultaneously regressing co-regulatory factors on technical covariates. To identify co-regulatory factors in ESEMs, estimated factor loadings from the best fitting EFA model were included as fixed parameters to ensure that latent dependent variables were apposite to the co-regulatory factors extracted from the best fitting EFA model. Finally, because gene allele frequencies vary across world populations (Jakobsson et al. 2008; Wang, Zöllner, and Rosenberg 2012), additional sensitivity analyses were performed to evaluate whether findings were biased by the presence of genetic heterogeneity by estimating multiple regressions and EFA models after excluding observations from all nonwhite participants (see supplemental material).

Results

Gene Expression Composite Scores

Table 2 reports multiple regression coefficients, standard errors, *p*-values, and FDR-adjusted *p*-values (Benjamini and Hochberg 1995). In the text below, we report point estimates and 95% confidence intervals for noteworthy findings. Compared to other races/ethnicities, Black/African Americans showed greater antiviral gene expression (b = .201 [.104 to .297], p < .001, FDR-adjusted p < .001). Ancillary analyses indicated that White/European Americans showed lower antiviral gene expression compared to other races/ethnicities (b = −.127 [−.206 to −.050], p = .001, FDR-adjusted p = .005). The inflammatory gene expression composite scores were not significantly associated with race/ethnicity (ps > .10; Table 2). Thus, driven by the differential expression of antiviral genes (which represents a component of the overall CTRA profile), CTRA composite scores were lower for Black participants (b = −.154 [−.257 to −0.052], p = .003, FDR-adjusted p = .010).

Table 2. Multiple linear regressions of gene composite scores on technical covariate, demographic factors, and indicators of behavioral and physical health estimated using maximum likelihood with robust standard errors.

	Inflammatory Composite				Antiviral Composite				CTRA Composite			
	b	*SE*	*p*	*FDR*	*b*	*SE*	*p*	*FDR*	*b*	*SE*	*p*	*FDR*
Technical Covariates												
Batch Plate 2	−.169	.078	.032	.117	−.164	.126	.193	.314	−.005	.133	.971	.971
Batch Plate 3	**−.146**	**.040**	**<.001**	**.003**	**.327**	**.064**	**<.001**	**<.001**	**−.474**	**.068**	**<.001**	**<.001**
Batch Plate 4	**−.133**	**.040**	**.001**	**.006**	**.283**	**.064**	**<.001**	**<.001**	**−.416**	**.068**	**<.001**	**<.001**
Batch Plate 5	.037	.041	.364	.557	**.291**	**.065**	**<.001**	**<.001**	**−.254**	**.069**	**<.001**	**<.001**
Batch Plate 6	−.075	.043	.082	.193	**.356**	**.069**	**<.001**	**<.001**	**−.431**	**.073**	**<.001**	**<.001**
Batch Plate 7	−.068	.042	.105	.228	−.035	.068	.602	.745	−.033	.072	.645	.762
Batch Plate 8	**.239**	**.042**	**<.001**	**<.001**	**.188**	**.067**	**.006**	**.014**	.051	.071	.475	.686
CD3E	−.004	.023	.844	.896	−.048	.037	.189	.314	.044	.039	.261	.520
CD3E	.023	.017	.191	.327	**.085**	**.028**	**.002**	**.006**	**−.063**	**.029**	**.032**	**.084**
CD4	**.061**	**.020**	**.003**	**.014**	**.080**	**.033**	**.014**	**.034**	−.019	.034	.582	.735
CD8A	**.024**	**.012**	**.038**	**.122**	**.045**	**.019**	**.016**	**.035**	−.021	.020	.293	.520
CD14	.170	.014	<.001	<.001	**.111**	**.023**	**<.001**	**<.001**	**.059**	**.024**	**.016**	**.046**
CD19	.008	.010	.402	.580	−.023	.016	.139	.258	.031	.017	.059	.139
FCGR3A	.019	.011	.076	.193	**.086**	**.017**	**<.001**	**<.001**	**−.067**	**.019**	**<.001**	**.002**
RNCAM1	.007	.010	.485	.664	−.014	.016	.384	.588	.021	.017	.218	.473
Demographic Variables												
Age	**.002**	**.001**	**.048**	**.138**	.001	.002	.578	.745	.001	.002	.523	.716
Sex (Female)	−.012	.021	.554	.720	.007	.034	.844	.878	−.019	.036	.594	.735
Age × Sex	**.004**	**.002**	**.008**	**.030**	**−.005**	**.002**	**.046**	**.092**	**.009**	**.003**	**.001**	**.002**
African American	.046	.031	.135	.270	**.201**	**.049**	**<.001**	**<.001**	**−.155**	**.052**	**.003**	**.010**
Other Race/Ethnicity	−.008	.033	.806	.896	.040	.053	.457	.660	−.048	.056	.397	.607
Level of Education	.002	.011	.877	.896	.001	.018	.966	.966	.001	.019	.959	.971
Health Indicators												
Alcohol Consumption	−.009	.007	.201	.327	.003	.011	.762	.852	−.013	.012	.300	.520
History of Smoking	−.010	.022	.642	.791	.025	.036	.492	.673	−.035	.038	.356	.579
Body Mass Index	.000	.002	.896	.896	−.001	.002	.786	.852	.001	.003	.739	.836
Chronic Conditions	−.002	.004	.670	.791	−.002	.006	.733	.852	.000	.007	.943	.971

b = unstandardized multiple regression coefficient. *SE* = standard error. *p* = *p*-value for multiple regression coefficient. FDR = *p*-value for multiple regression coefficient adjusted for false discovery rate using the Benjamini-Hochberg procedure. Statistically significant (p < .05) effects are printed in bold font.

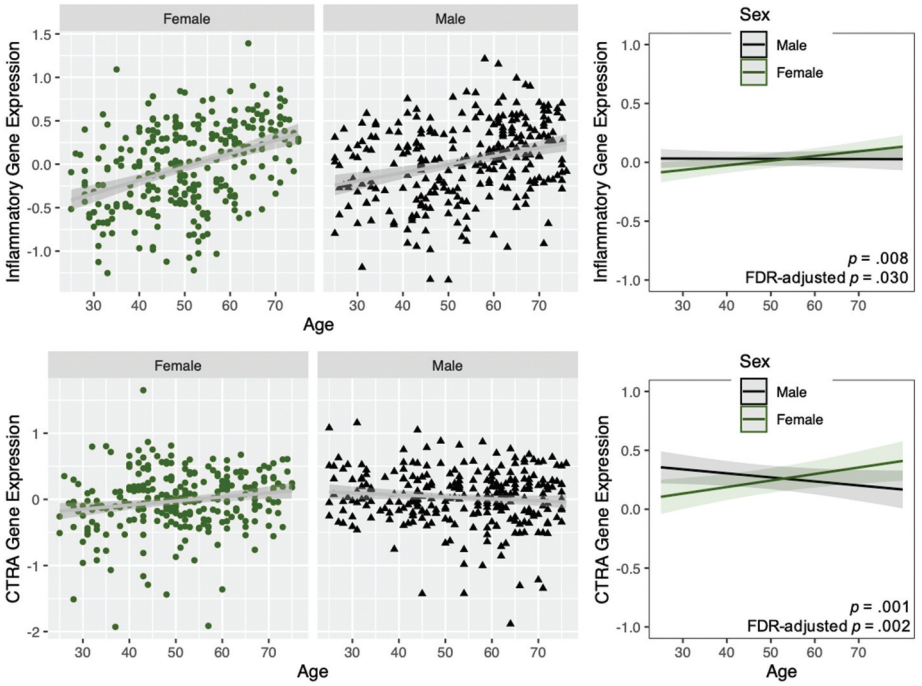

Figure 1. Age and sex differences in gene expression composite scores. **A**ge trends in inflammatory and CTRA gene expression composite scores are depicted before (left and middle panels) and after (right panel) adjusting for the effects of study variables. Individual observations are plotted using green circles and black triangles for females and males, respectively. The right panels depict age × sex interactions from multiple linear regressions predicting gene expression composite scores. p = probability of the observed interaction effect if the null hypothesis is true. Shaded regions around linear trends depict 95% confidence intervals.

Controlling for all study variables, the association between inflammatory expression and age was moderated by sex (b = .004 [.001 to .007], p = .008, FDR-adjusted p = .005). Depicted on the left and middle panels of Figure 1, the unadjusted association between age and inflammatory expression was positive for both females (r = .396 [.291, .491]) and males (r = .294 [.181, .400]). Holding other study variables constant, the estimated marginal means of age-trends from the multiple linear regression indicated that inflammatory gene expression was predicted to increase with age more rapidly for females (b = .004 [.001, .006]), compared to males (b = .000 [−.002, .002]). Similarly, the association between CTRA composite scores (i.e., inflammation – antiviral) and age was significantly moderated by sex (b = .009 [.004 to .014], p < .001, FDR-adjusted p = .001). The unadjusted association between CTRA expression and age was positive for females (r = .170 [.053, .283], but not for males (r = −.095 [−.213, .024]). Holding other study variables constant, CTRA gene expression was predicted to increase with age for females (b = .005 [.001, .010]) but not males (b = −.003 [−.007, .001]).

The interaction between age and sex was marginally significant for antiviral expression after adjusting for multiple predictors (b = −.005 [.000 to −.010], p = .046, FDR-adjusted p = .092). The unadjusted association between age and antiviral expression was positive for

both females (r = .200 [.084, .311]) and males (r = .329 [.219, .432]). However, after adjusting for the effects of other study variables, estimated linear trends indicated that antiviral gene expression increased with age for males (b = .003 [.000, .007]) but not females (b = −.002 [−.006, .002]).

Co-Regulatory Gene Sets

Factor loadings, standard errors, and coefficients of determination for gene transcripts are reported in Table 3. The eigenvalues from the EFA and parallel analysis, both average and 95th percentile, indicated that a five-factor solution was preferred to alternative solutions.

Table 3. Estimates for empirically identified co-regulatory gene sets.

	F1		F2		F3		F4		F5		
	λ	*SE*	λ	*SE*	λ	*SE*	λ	*SE*	λ	*SE*	R^2
Inflammatory Genes											
FOS	**.72**	.04	−.05	.04	.01	.03	−.02	.03	.20	.06	**.61**
FOSB	**.51**	.05	.07	.05	−.10	.05	−.18	.05	.02	.04	**.29**
FOSL2	**.35**	.04	**.68**	.04	−.01	.03	.04	.03	−.02	.03	**.75**
IL1B	**.30**	.05	.12	.06	.13	.05	**.28**	.05	.05	.05	**.36**
JUN	.16	.05	**.44**	.06	−.10	.05	.15	.05	.18	.06	**.38**
JUNB	**.52**	.04	**.58**	.05	.04	.03	−.01	.02	.03	.03	**.87**
JUND	.05	.03	**.66**	.08	−.01	.03	−.15	.04	.42	.06	**.79**
NFKB1	.00	.03	**.36**	.08	−.05	.04	**.34**	.05	.34	.06	**.49**
NFKB2	−.01	.05	**.39**	.06	.11	.05	**−.31**	.05	.01	.05	**.21**
PTGS1	−.11	.05	.28	.07	.13	.05	−.01	.04	.28	.07	**.23**
PTGS2	**.51**	.05	.13	.06	.04	.04	.12	.05	−.14	.05	**.34**
REL	−.01	.04	.30	.07	−.09	.05	**.42**	.05	−.27	.07	**.29**
RELA	−.16	.04	**.92**	.04	−.08	.04	.00	.03	.08	.06	**.77**
RELB	−.02	.04	**.76**	.03	.02	.04	−.10	.04	−.05	.06	**.53**
TNF	−.02	.04	**.55**	.05	.09	.05	.00	.04	.05	.05	**.36**
Antiviral Genes											
GBP1	−.04	.04	.23	.06	.25	.06	**.46**	.05	−.21	.05	**.49**
IFI16	.09	.04	−.11	.08	.08	.05	**.85**	.04	.06	.04	**.80**
IFI27L1	−.03	.05	**.33**	.06	.06	.05	−.04	.05	.23	.06	**.22**
IFI27L2	.05	.04	**.62**	.06	.07	.04	.08	.04	.20	.05	**.61**
IFI30	**.25**	.05	−.01	.04	.07	.05	.08	.05	**−.45**	.05	**.22**
IFI35	.13	.04	**.43**	.05	**.38**	.04	−.04	.04	.05	.05	**.54**
IFI44	.03	.02	**−.51**	.04	**.63**	.04	**.33**	.04	.07	.04	**.66**
IFI44L	−.06	.03	−.18	.05	**.94**	.03	−.03	.03	.00	.03	**.74**
IFI6	.06	.03	**.38**	.05	**.68**	.03	−.11	.03	.07	.03	**.81**
IFIH1	−.04	.03	−.27	.07	.18	.06	**.73**	.04	−.03	.03	**.64**
IFIT1	.03	.03	−.05	.04	**.78**	.03	.00	.03	.07	.04	**.63**
IFIT2	.09	.04	−.05	.04	**.46**	.05	**.46**	.04	−.01	.03	**.64**
IFIT3	.08	.03	.21	.05	**.72**	.03	.01	.03	−.03	.03	**.72**
IFIT5	−.11	.04	.09	.07	**.28**	.05	**.32**	.05	**.28**	.06	**.41**
IFITM1	.04	.04	.43	.08	.10	.04	.05	.04	.39	.07	**.55**
IFITM2	−.02	.03	**.91**	.03	.10	.04	.04	.03	−.02	.04	**.90**
IFITM3	.10	.04	**.57**	.04	**.43**	.04	−.03	.03	−.13	.05	**.68**
IRF2	.07	.05	.17	.07	−.03	.04	**.56**	.04	.00	.04	**.39**
IRF7	.08	.04	**.39**	.07	**.36**	.04	.00	.03	**.31**	.05	**.66**
IRF8	**.25**	.05	.26	.09	.01	.03	**.25**	.05	**.42**	.05	**.62**
JCHAIN	−.01	.04	.30	.07	.09	.05	.09	.05	.32	.06	**.33**
MX1	−.18	.04	.00	.02	**.84**	.03	.07	.04	−.09	.04	**.70**
MX2	−.03	.04	.07	.05	**.53**	.05	**.31**	.04	.04	.04	**.58**
OAS1	.18	.04	.40	.05	**.36**	.04	**.23**	.04	−.01	.04	**.70**
OAS2	−.05	.03	**.53**	.05	**.44**	.04	.12	.04	.09	.04	**.77**
OAS3	.02	.03	.12	.04	**.64**	.04	**.21**	.04	−.08	.04	**.65**
OASL	−.09	.04	**.25**	.05	**.43**	.05	.16	.05	−.05	.05	**.39**

F1 – F5 = Empirically-identified co-regulatory gene sets (i.e., exploratory latent factors). λ = Geomin rotated loading. *SE* = standard error. R^2 = coefficient of determination. Estimates printed in bold are statistically significant at $p < .001$

Moreover, *RMSEA* = .048 and *CFI* = .930 evinced good model fit for the five-factor solution, and a similar solution emerged when observations from nonwhites were excluded from analyses. Factor loadings that were statistically significant ($p < .001$) and greater than |.30| are emphasized in the interpretation of results. The first and third co-regulatory factors largely captured up-regulation of inflammatory genes (e.g., *FOS, FOSB, FOSL2, IL1B, JUNB*, & *PTGS2*) and up-regulation of antiviral genes (e.g., *IFI35, IFI44, IFI44L, IFI6, IFIT1, IFIT2, IFIT3, IFIT5 IFITM3, IRF7, MX1, MX2, OAS1, OAS2, OAS3*, & *OASL*), respectively. More nuanced patterns of expression emerged for the remaining co-regulatory factors (Table 3). Together, the five co-regulatory factors explained 21% to 90% of the variation in the analyzed gene transcripts (*p*s < .001).

The demographic correlates of empirically derived co-regulatory gene sets were similar to the demographic correlates of gene expression composite scores. For example, Black/African American was positively associated with the co-regulatory gene set characterized by upregulation of antiviral genes ($b = .371$ [.153 to .588], $p = .001$, FDR-adjusted $p = .012$), mirroring the racial/ethnic difference observed for the antiviral gene expression composite scores. The association between age and the co-regulatory gene set characterized by upregulation of inflammatory genes was also marginally moderated by sex ($b = .008$ [.000, .018], $p = .045$, FDR-adjusted $p = .108$). The results of ESEMs are reported in supplemental materials.

Discussion

There is increasing interest in using gene expression as a method for understanding how demographic disparities in health arise over the life course. This study of community-dwelling middle-aged US adults identified demographic differences in white blood cell expression of two key health-relevant sets of genes: those involved in inflammation, and those involved in innate antiviral responses. As in previous studies (Cole et al. 2020; Kohrt et al. 2016; McDade et al. 2016), the current study observed marked race-related differences in expression of innate antiviral genes and marked sex-related differences in the expression of inflammatory genes that varied in size across adulthood. The sample analyzed here involved a substantially greater range of ages than available in previous population health transcriptome profiling studies, allowing us to confirm a general age-related trend in the expression of both inflammatory and antiviral genes. The greater age variation in this sample also allowed us to test whether sex differences in gene expression might vary across age-groups, and potentially contribute to the greater risk of late-life disease and mortality in males.

This analysis examined inflammatory and antiviral gene sets because these specific aspects of white blood cell gene expression are subject to regulation by stress-related neuroendocrine signaling and may thus help to answer key questions in the public health literature (Williams, Lawrence, and Davis 2019) regarding the biological pathways through which health disparities arise (Spence et al. 2016). The present study operationalized differential gene expression using both *a priori* composite scores and *a posteriori* co-regulatory gene sets. Notably, irrespective of operationalization and across a variety of sensitivity analyses, age, sex, and racial/ethnic differences in gene expression emerged. The most consistent finding was that antiviral genes were expressed at relatively higher levels in Black adults, and at relatively lower levels for White adults.

A number of age-by-sex interactions also emerged. In contrast with recent findings from a study of young adults only, whereby higher levels of inflammatory gene expression were observed in females during *early* adulthood (Cole et al. 2020), the present study found that inflammatory expression was more strongly associated with age in females, relative to males. This finding is in the opposite direction as expected, given the higher incidence of all-cause mortality and shorter life expectancy for males after midlife. However, in this sample, females reported a greater number of chronic conditions than males (Table 1), which may contribute to the greater age-related inflammatory expression observed in females. The age-related trend toward greater antiviral expression for males in later adulthood is consistent with sex differences in mortality due to parasitic and infectious disease, though this trend was no longer statistically significant after accounting for multiple comparisons. As the current study tested for demographic differences in a relatively young subfield and was the first to test for sex differences in age-trends, future studies should interrogate the replicability and generalizability of these findings, especially given plausible countervailing hypotheses, high variability of gene expression, and the number of associations that were tested in the current study. Sex differences in linear age-trends were also relatively small in magnitude, and the lower bound of the 95% confidence intervals for these interactions approached zero. Moreover, sex differences in age trends appeared only in adjusted analyses and were not evident in the absence of covariate adjustment (Figure 1). Collectively, these conditional results underscore the need for future studies to assess the replicability and generalizability of sex differences in gene expression across lifespan development. On the other hand, there was a comparatively low probability of the racial/ethnic differences observed in the present study, assuming a null hypothesis of no difference, and the observed effects were consistent with those previously reported in a large sample of young adults (Cole et al. 2020), suggesting that racial/ethnic differences in antiviral expression may be robust across different stages of the life course.

After accounting for technical covariates and demographic differences, educational attainment and indicators of behavioral and physical health were not significantly associated with gene expression composite scores, despite other studies that have found otherwise (e.g., Cole et al. 2020. This may be explained, in part, by the fact that participants in the current study tended to be well-educated and were relatively healthy. Although the subsample of participants with gene expression data did not differ systematically on study variables from other participants in the biomarker Refresher cohort, other studies have found that the MIDUS biomarker subsample differs from the larger MIDUS cohort and is not fully representative of the U.S. (Dienberg Love et al. 2010). Future studies should further probe the sample characteristics, study designs, measures, specific health outcomes, and physiological mechanisms that modulate the relation between gene expression and physical health.

Limitations

The present study implemented a non-experimental, cross-sectional design. Therefore, causal inference is not warranted. Moreover, there are likely many down-stream phenotypes, both biological and behavioral, which link differential gene expression to physical and mental health outcomes. Future studies may benefit from implementing longitudinal designs that may track potential biological, behavioral, and social factors that mediate associations with demographic factors that were documented in the present study. In addition, sex/gender was only reported once in the current study but may change over

time, and sex was not distinguished from gender identity. Future studies may also benefit from asking respondents to report their gender at more than one point in time (Hanes and Clouston 2020) or differentiate biological sex as determined by genotype from gender identity.

The present study also included few participants that identified as a race or ethnicity other than White or Black. Similar to other genetically informative studies (Martin et al. 2019), future studies should include larger samples of other racial and ethnic groups to provide a more comprehensive assessment of biological processes underlying minority health disparities. Finally, the extent to which genetic variation contributes to demographic differences in antiviral expression should be interrogated in future studies that capture variation in both DNA and RNA in a racially/ethnically diverse sample. From the current study alone, it is unclear whether racial/ethnic differences in antiviral expression are due to differences in social exposures, microbial exposures, or genetic differences. Despite these limitations, the present study provides an important step toward understanding the biological mechanisms that may contribute to health disparities across demographic strata in adulthood.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

The MIDUS study has been funded by the following: John D. and Catherine T. MacArthur Foundation Research Network; National Institute on Aging (P01-AG020166); National institute on Aging (U19-AG051426). These analyses were also partially supported by a data analysis grant from the National Institute on Aging (R01 AG058595).

References

American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders (DSM-5®)*. Arlington, VA. American Psychiatric Pub.

Asparouhov, T., and B. Muthén. 2009. Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal* 16 (3):397–438. doi:10.1080/10705510903008204.

Austad, S. N., and K. E. Fischer. 2016. Sex differences in lifespan. *Cell Metabolism* 23 (6):1022–33. doi:10.1016/j.cmet.2016.05.019.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 57 (1):289–300.

Cole, S. W. 2013. Social regulation of human gene expression: Mechanisms and implications for public health. *American Journal of Public Health* 103 (S1):S84–S92. doi:10.2105/AJPH.2012.301183.

Cole, S. W. 2019. The conserved transcriptional response to adversity. *Current Opinion in Behavioral Sciences* 28:31–37. doi:10.1016/j.cobeha.2019.01.008.

Cole, S. W., M. E. Levine, J. M. G. Arevalo, J. Ma, D. R. Weir, and E. M. Crimmins. 2015. Loneliness, eudaimonia, and the human conserved transcriptional response to adversity. *Psychoneuroendocrinology* 62:11–17. doi:10.1016/j.psyneuen.2015.07.001.

Cole, S. W., M. J. Shanahan, L. Gaydosh, and K. M. Harris. 2020. Population-based RNA profiling in Add Health finds social disparities in inflammatory and antiviral gene regulation to emerge by young adulthood. *Proceedings of the National Academy of Sciences* 117 (9):4601–08. doi:10.1073/pnas.1821367117.

Dienberg Love, G., T. E. Seeman, M. Weinstein, and C. D. Ryff. 2010. Bioindicators in the MIDUS national study: Protocol, measures, sample, and comparative context. *Journal of Aging and Health* 22 (8):1059–80. doi:10.1177/0898264310374355.

Foo, Y. Z., S. Nakagawa, G. Rhodes, and L. W. Simmons. 2017. The effects of sex hormones on immune function: a meta-analysis. Biological Reviews, 92(1), 551–571.

Hanes, D. W., and S. A. P. Clouston. 2020. Ask again: Including gender identity in longitudinal studies of aging. *The Gerontologist*. doi:10.1093/geront/gnaa107.

Jakobsson, M., S.W. Scholz, P. Scheet, J. R. Gibbs, J. M. VanLiere, H. C. Fung, and A. B. Singleton. 2008. Genotype, haplotype and copy-number variation in worldwide human populations. Nature, 451(7181), 998–1003

Kautzky-Willer, A., J. Harreiter, and G. Pacini. 2016. Sex and gender differences in risk, pathophysiology and complications of type 2 diabetes mellitus. *Endocrine Reviews* 37 (3):278–316.

Klein, S. L. 2004. Hormonal and immunological mechanisms mediating sex differences in parasite infection. *Parasite Immunology* 26 (6–7):247–64. doi:10.1111/j.0141-9838.2004.00710.x.

Kohrt, B. A., C. M. Worthman, R. P. Adhikari, N. P. Luitel, J. M. Arevalo, J. Ma, H. McCreath, T. E. Seeman, E. M. Crimmins, and S. W. Cole. 2016. Psychological resilience and the gene regulatory impact of posttraumatic stress in Nepali child soldiers. *Proceedings of the National Academy of Sciences* 113 (29):8156–61. doi:10.1073/pnas.1601301113.

LeBrón, A. M., A. J. Schulz, G. B. Mentz, B. A. Israel, and C. A. Stokes. 2019. Social relationships, neighbourhood poverty and cumulative biological risk: Findings from a multi-racial US urban community. *Journal of Biosocial Science* 51 (6):799–816. doi:10.1017/S002193201900004X.

Martin, A. R., M. Kanai, Y. Kamatani, Y. Okada, B. M. Neale, and M. J. Daly. 2019. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics* 51 (4):584–91. doi:10.1038/s41588-019-0379-x.

McDade, T. W., M. Ross K, L, R. Fried, J. M. Arevalo, J. Ma, G. E. Miller, and S. W. Cole. 2016. Genome-wide profiling of RNA from dried blood spots: Convergence with bioinformatic results derived from whole venous blood and peripheral blood mononuclear cells. *Biodemography and Social Biology* 62 (2):182–97. doi:10.1080/19485565.2016.1185600.

Mensah, G. A., A. H. Mokdad, E. S. Ford, K. J. Greenlund, andJ. B. Croft. 2005. State of disparities in cardiovascular health in the United States. Circulation, 111(10), 1233–1241.

Montanelli, R. G., and L. G. Humphreys. 1976. Latent roots of random data correlation matrices with squared multiple correlations on the diagonal: A Monte Carlo study. *Psychometrika* 41 (3):341–48. doi:10.1007/BF02293559.

Muthén, L. K., and B. Muthen. 2017. *Mplus user's guide: Statistical analysis with latent variables, user's guide*. Muthén & Muthén. Los Angeles,CA.

Owens, I. P. F. 2002. Sex differences in mortality rate. *Science* 297 (5589):2008–09. doi:10.1126/science.1076813.

Phelan, J. C., and B. G. Link. 2015. Is racism a fundamental cause of inequalities in health? *Annual Review of Sociology* 41:311–30. doi:10.1146/annurev-soc-073014-112305.

Ryff, C. D., and R. F. Krueger, eds. 2018. *The Oxford handbook of integrative health science*. Oxford University Press. New York, NY, USA.

Spanakis, E. K., and S. H. Golden. 2013. Race/ethnic difference in diabetes and diabetic complications. *Current Diabetes Reports* 13 (6):814–23. doi:10.1007/s11892-013-0421-9.

Spence, N. D., S. Wells, K. Graham, and J. George. 2016. Racial discrimination, cultural resilience, and stress. *Canadian Journal of Psychiatry. Revue Canadienne De Psychiatrie* 61 (5):298–307. doi:10.1177/0706743716638653.

Wang, C., S. Zöllner, and N. A. Rosenberg. 2012. A quantitative comparison of the similarity between genes and geography in worldwide human populations. *PLoS Genetics* 8 (8):e1002886. doi:10.1371/journal.pgen.1002886.

Wang, Y., and M. A. Beydoun. 2007. The obesity epidemic in the United States—gender, age, socioeconomic, racial/ ethnic,and geographic characteristics: A systematic review and meta-regression analysis. *Epidemiologic Reviews* 29 (1):6–28. doi:10.1093/epirev/mxm007.

Williams, D. R., J. A. Lawrence, and B. A. Davis. 2019. Racism and health: Evidence and needed research. *Annual Review of Public Health* 40:105–25. doi:10.1146/annurev-publhealth-040218-043750.

Williams, D. R., and S. A. Mohammed. 2009. Discrimination and racial disparities in health: Evidence and needed research. *Journal of Behavioral Medicine* 32 (1):20–47. doi:10.1007/s10865-008-9185-0.