Population-based RNA profiling in Add Health finds social disparities in inflammatory and antiviral gene regulation to emerge by young adulthood

Steven W. Cole[a,b,c,1], Michael J. Shanahan[d,e], Lauren Gaydosh[f,g], and Kathleen Mullan Harris[h,i,1]

[a]School of Medicine, University of California, Los Angeles, CA 90095; [b]Norman Cousins Center for Psychoneuroimmunology, Semel Institute for Neuroscience & Human Behavior, University of California, Los Angeles, CA 90095; [c]Jonsson Comprehensive Cancer Center, University of California, Los Angeles, CA 90095; [d]Department of Sociology, University of Zürich, CH 8050 Zürich, Switzerland; [e]Jacobs Center for Productive Youth Development, University of Zürich, CH 8050 Zürich, Switzerland; [f]Center for Medicine, Health, and Society, Vanderbilt University, Nashville, TN 37235; [g]Public Policy Studies, Department of Political Science, Vanderbilt University, Nashville, TN 37235; [h]Department of Sociology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27516; and [i]Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27516

Contributed by Kathleen Mullan Harris, December 23, 2019 (sent for review December 17, 2018; reviewed by Elissa S. Epel, Stephen Manuck, and Colter Mitchell)

Health in later life varies significantly by individual demographic characteristics such as age, sex, and race/ethnicity, as well as by social factors including socioeconomic status and geographic region. This study examined whether sociodemographic variations in the immune and inflammatory molecular underpinnings of chronic disease might emerge decades earlier in young adulthood. Using data from 1,069 young adults from the National Longitudinal Study of Adolescent to Adult Health (Add Health)—the largest nationally representative and ethnically diverse sample with peripheral blood transcriptome profiles—we analyzed variation in the expression of genes involved in inflammation and type I interferon (IFN) response as a function of individual demographic factors, sociodemographic conditions, and biobehavioral factors (smoking, drinking, and body mass index). Differential gene expression was most pronounced by sex, race/ethnicity, and body mass index (BMI), but transcriptome correlates were identified for every demographic dimension analyzed. Inflammation-related gene expression showed the most pronounced variation as a function of biobehavioral factors (BMI and smoking) whereas type I IFN-related transcripts varied most strongly as a function of individual demographic characteristics (sex and race/ethnicity). Bioinformatic analyses of transcription factor and immune-cell activation based on transcriptome-wide empirical differences identified additional effects of family poverty and geographic region. These results identify pervasive sociodemographic differences in immune-cell gene regulation that emerge by young adulthood and may help explain social disparities in the development of chronic illness and premature mortality at older ages.

social genomics | biodemography | life-span development | social epidemiology | Add Health

Most chronic illnesses show marked demographic variations in prevalence and outcome, including cardiovascular (1), neoplastic (2), metabolic (3), and neurodegenerative diseases (4). These demographic disparities become increasingly prevalent in mid to later adulthood (5, 6), resulting in shorter life spans for men relative to women, for blacks and Hispanics relative to Asians and non-Hispanic whites, for the poor relative to the affluent, and for residents of the southern United States compared to other regions (7–9). However, the biological underpinnings of these late-life health disparities may emerge decades earlier in adolescence and young adulthood (9–15), well before such morbidities are commonly diagnosed. Most chronic diseases develop over the course of many years and are driven in part by the activity of disease-promoting molecular pathways involved in inflammation, metabolism, and immune function (16). Measurement of gene expression can provide insight into the molecular processes that underlie these sociodemographic gradients in health. However, little is known about sociodemographic variation in the molecular precursors of disease because population health studies have rarely surveyed the molecular characteristics of adolescents or young adults. Here we report results from a transcriptome profiling analysis of a large, nationally representative and ethnically diverse sample of young adults and find significant demographic variation in the molecular antecedents of chronic disease decades before those diseases typically manifest in late adulthood.

To determine whether demographic variations in gene regulation during young adulthood might contribute to social gradients in late-life disease risk, this study analyzed genome-wide transcriptional profiles in blood samples from a nationally representative sample of 1,126 young adults (mean age 37) participating in the National Longitudinal Study of Adolescent to Adult Health (Add Health) (17). Add Health is the largest, most comprehensive longitudinal study of adolescents ever undertaken, with national representation of all race, ethnic, immigrant, socioeconomic status, and geographic subgroups in the United States. Add Health used

Significance

Health in later life and longevity vary substantially across sociodemographic groups, but the biological mechanisms of these disparities remain poorly understood. We conducted a transcriptome profiling study of inflammatory and antiviral gene activity in a large, nationally representative and ethnically diverse sample of young adults and found that sociodemographic variations in the activity of these molecular pathways emerge by young adulthood—well before they manifest as late-life chronic illness. Inflammation related to biobehavioral factors (BMI, smoking), interferons related to individual characteristics (sex, race/ethnicity), and transcription factor and immune-cell activation showed additional links to social context (family poverty, geographic region). These data suggest that interventions early in life may address the predisease physiological disparities that manifest as late-life health disparities.

Author contributions: S.W.C., M.J.S., L.G., and K.M.H. designed research; S.W.C. and K.M.H. performed research; S.W.C. and L.G. analyzed data; and S.W.C., M.J.S., L.G., and K.M.H. wrote the paper.

Reviewers: E.S.E., University of California, San Francisco; S.M., University of Pittsburgh; and C.M., University of Michigan.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: steve.cole@ucla.edu or kathie_harris@unc.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1821367117/-/DCSupplemental.

probability population-representative sampling to enroll a nationwide sample of adolescents (grades 7 to 12) in 1994 to 1995 and has followed that cohort longitudinally since then (17). We analyzed gene expression profiles in whole-blood samples collected ∼22 y later during young adulthood to assess transcriptome variation as a function of individual demographic characteristics (age, sex, race/ethnicity), sociodemographic conditions (family poverty status, geographic region), and biobehavioral factors that might potentially be confounded with demographic characteristics [smoking, alcohol consumption, and body mass index (BMI)]. Our initial analysis focused on quantifying variations in health-relevant gene expression among young adults as a function of fundamental demographic, social, and behavioral factors known to define disparities in chronic disease. In addition to clarifying the molecular origins of late- life health disparities, this analysis provides an essential platform for more detailed analyses of specific risk factors in adolescence and young adulthood, as well as methodological guidance to avoid the risk of sociodemographic confounding in future genomic research.

Our analyses focused on two molecular pathways involved in the pathogenesis of multiple chronic diseases (16): 1) genes involved in inflammation and 2) genes involved in type I interferon (IFN) responses. These two gene sets represent functionally distinct immunoregulatory programs (18, 19) and were selected for analysis based on their well-established relationship to chronic disease and longevity, both as empirical predictors (16, 20–27) and as molecular mechanisms of disease (16, 28–33). Both gene sets are subject to physiological regulation by tissue injury and microbial stimuli as well as by the neural and endocrine systems (34).

Neural/endocrine regulation of gene expression has been hypothesized to constitute one pathway through which social environmental conditions might contribute to health disparities, for example, through stress-induced activation of the Conserved Transcriptional Response to Adversity (CTRA) RNA profile that involves up-regulation of inflammatory genes and a reciprocal down-regulation of type I IFN genes in the circulating leukocyte pool (35–37). Basic laboratory research has found the CTRA transcriptome shift to be mediated in part by sympathetic nervous system (SNS)-induced increases in hematopoietic output of myeloid lineage immune cells—monocytes, dendritic cells, and neutrophil granulocytes (38–40).

In addition to examining basic sociodemographic variations in inflammatory and type I IFN gene modules due to their established relevance for chronic disease, we also conducted analyses testing whether the more specific CTRA pattern (i.e., IFN – inflammation) and related neuroendocrine and cellular mechanisms might contribute to such demographic variations. As such, the present analysis quantified demographic variation in young adult blood-cell gene expression profiles using three complementary analytic approaches corresponding to three distinct levels of biological influence on gene expression (41): 1) analyzing expression of a-priori–defined sets of inflammatory and IFN indicator genes used in previous research (level 1) (42); 2) analyzing genome-wide empirical differences in RNA expression in terms of their coregulation by transcription factors involved in inflammatory, type I IFN, SNS, and neuroendocrine response (level 2) (34, 36); and 3) analyzing genome-wide empirical differences in RNA expression in terms of their coexpression in specific immune-cell subsets involved in inflammatory and IFN gene expression (particularly monocytes, dendritic cells, and neutrophils) (level 3) (38, 39, 43).

Results

We analyzed gene expression data from a nationally representative subsample (sample 1) of Add Health Wave V (2016 to 2017). Sample 1 is a random one-third subsample of the nationally representative Wave V (see *SI Appendix*, *Methods*, for details), so it, too, is nationally representative (44). Among the 1,126 participants with transcriptome profiles available in sample 1, 57 were missing data on one or more demographic or behavioral variables, leaving 1,069 individuals in the final analytic sample. Characteristics of the analytic sample are presented in Table 1 and are closely representative of the national Add Health cohort of young adults in sample 1 Wave V (any differences are <5%).

Transcriptome profiles were derived by sequencing whole-blood polyadenylated RNA and tested for quantitative variations as a function of: 1) individual demographic characteristics (age, sex, race/ethnicity); 2) social and geographic context (family poverty status, region of residence); and 3) biobehavioral factors that might potentially confound sociodemographic effects (smoking, alcohol consumption, BMI). Quantitative variations in gene expression were analyzed by linear statistical models that adjusted the estimated effect of each demographic characteristic for any correlated effects of other dimensions, as well as for technical covariates (sample RNA integrity, sample RNA profile quality, sample sequencing depth, and assay batch).

We assessed sociodemographic variations in inflammatory and type I IFN gene expression using three complementary levels of transcriptome analysis involving 1) a-priori-specified sets of inflammatory and IFN indictor genes used in previous research to capture broad variations in innate immune activity (42); 2) empirical differences in genome-wide transcriptional profiles analyzed in terms of their regulation by transcription factors involved in inflammation, type I IFN, and neuroendocrine responses (34, 36); and 3) empirical differences in genome-wide transcriptional profiles analyzed in terms of their cellular origins, focusing particularly on innate immune cells involved in inflammatory and IFN responses (monocytes, dendritic cells, and neutrophils) (38, 39, 43).

Level 1: A Priori Gene Composites. Analyses of prespecified composites of 19 representative proinflammatory genes (e.g., *IL1B*, *IL6*, COX2/*PTGS2*, *TNF*) and 32 IFN-related genes (e.g., *IFI*-, *OAS*-, and *MX*-family genes) identified significant sociodemographic variation in gene expression across the entire set of analyzed transcripts: $F(20, 1{,}012) = 5.98$, $P = 3.6 \times 10^{-15}$. Follow-up analyses of each gene set separately indicated significant sociodemographic variation in expression of the inflammatory gene composite: $F(10, 1{,}040) = 2.09$, $^{*}P = 0.0230$; expression of the type I IFN gene composite: $F(10, 1{,}040) = 7.13$, $^{*}P = 6.3 \times 10^{-11}$; and expression of the CTRA composite (inflammation – type I IFN): $F(10, 1{,}040) = 3.53$, $^{*}P = 1.3 \times 10^{-4}$ (the asterisk indicates a value significant after correction for hierarchical multiple testing at a false discovery rate of $q < 0.05$). As shown in Fig. 1, these effects were most pronounced for individual demographic characteristics [i.e., age, sex, and race/ethnicity; inflammatory: $F(6, 1{,}040) = 3.01$, $^{*}P = 6.3 \times 10^{-3}$; IFN: $F(6, 1{,}040) = 10.86$, $^{*}P = 9.5 \times 10^{-12}$; CTRA $F(6, 1{,}040) = 5.48$, $^{*}P = 1.3 \times 10^{-5}$]. Inflammatory gene expression was up-regulated in females relative to males and in blacks relative to non-Hispanic whites. Type I IFN gene expression was up-regulated even more strongly in females relative to males and in Asians and blacks relative to non-Hispanic whites. As a result of males' markedly lower type I IFN activity, the CTRA profile (inflammatory – type I IFN) was up-regulated in males compared to females; it was also down-regulated in Asians relative to non-Hispanic whites. Ancillary analyses also found expression of the inflammatory gene composite to vary as a function of biobehavioral characteristics [omnibus $F(4, 1{,}040) = 8.12$, $P = 1.9 \times 10^{-6}$], with effects driven predominately by BMI and smoking (Fig. 1). Broadly speaking, type I IFN gene expression varied most strongly as a function of individual demographic characteristics, whereas inflammatory gene expression varied most strongly as a function of biobehavioral factors. None of the three broad gene composites varied significantly as a function of family poverty or residential region [all $F(4, 1{,}040) < 1$, $P > 0.5$].

Table 1. Analytic sample characteristics (n = 1,069)

	Mean (SD) or %		
	Analytic sample (n = 1,069)	Wave V sample 1 (n = 3,872)	Difference P value
Age (y)	36.5 (1.9)	36.6 (1.8)	0.481*
Sex (female) (%)	54.2	50.1	0.057†
Race/ethnicity (%)			0.100‡
White (non-Hispanic)	69.2	65.8	
Black (non-Hispanic)	14.2	16.0	
Hispanic	9.2	10.9	
Asian	2.8	3.2	
Other	4.7	4.2	
Region (%)			0.001‡
Northeast	13.0	16.1	
Midwest	35.5	31.0	
South	39.9	38.7	
West	11.6	14.2	
Poverty (%)	16.3	20.4	0.054†
BMI (kg/m^2)	30.1 (7.8)	29.9 (7.5)	0.516*
Smoking history (%)	48.2	47.1	0.690†
Regular drinking (%)	4.2	5.3	0.162†
Binge drinking (ordinal 0 to 6)	1.2 (1.7)	1.1 (1.5)	0.340*

Note: Descriptive statistics are weighted; 43 in analytic sample are missing sample 1 weights.
*Two-tailed single-sample t test: RNA sample mean = Wave V sample 1 mean.
†Two-sided binomial test: RNA sample proportion = Wave V sample 1 proportion.
‡χ^2 test: RNA sample proportions = Wave V sample 1 proportions.

The a-priori–specified gene composites analyzed here were originally derived on theoretical grounds to capture broad variations in activity of the two major immunoregulatory gene modules involved in innate immunity (18, 42). However, such broad indices can obscure more nuanced and differentiated aspects of gene regulation that become apparent in empirical gene coregulation analyses. To map the coregulatory substructure of the overall 19-gene inflammatory composite, we conducted exploratory principal factor analysis (*SI Appendix*, Fig. S1*A* and Dataset S1) and identified seven coregulated gene modules, each of which was structured around distinct patterns of transcription factor activity (*SI Appendix*, Fig. S1*B*) [factor 1 (F1) = *JUNB*, *FOSL2*, *RELA*, *RELB*; F2 = *FOS*; F3 = *REL*, *NFKB1*, *NFKB2*; F4 = *FOSB*, *JUN*; F5 = *JUND*; F6 = *FOSL1*] and distinct effector molecules (F2 = IL8/*CXCL8*, COX2/*PTGS2*; F3 = *IL1B*; F4 = *TNF*; F5 = COX1/*PTGS1*; F6 = *IL1A*). Results also identified a single-gene module (F7) involving variation in *IL6* expression that was largely uncorrelated with the other inflammatory gene modules. All but one of the inflammatory gene modules (F6) showed significant demographic variation in activity (Fig. 2*A*), with the specific demographic correlates varying across modules.

Exploratory principal factor analysis of the type I IFN composite (*SI Appendix*, Fig. S1*A* and Dataset S1) also identified seven major coregulated gene modules that were again structured around distinct transcription factors (F1 = *IRF7*; F2 = *IRF2*; F3 = low *IRF8*) and associated with distinct effector molecules (*SI Appendix*, Fig. S1*C*). Two single-gene modules emerged (*IFNB1* and *IGLL1*). All but one of the type I IFN subcomponents (F5) showed significant demographic variation in activity (Fig. 2*B*) with the specific demographic correlates again varying across modules.

Levels 2 and 3: Empirical Transcriptome Variation. In addition to analyzing a-priori–defined sets of inflammatory and type I IFN indicator genes, we also quantified empirical variation in the genome-wide transcriptomic correlates of sociodemographic factors (Fig. 3*A*; *SI Appendix*, Fig. S2; and Dataset S2). Each sociodemographic parameter was associated with hundreds of genes showing >20% difference in expression across the observed range of variation (although the statistical significance of these individual transcript associations varied, with some dimensions such sex, race, and BMI showing large numbers of differentially expressed genes at a genome-wide false discovery rate of 5%, whereas others failed to yield any significant differences after correction for genome-wide multiple testing) (*SI Appendix*, Fig. S2 and Dataset S2).

Level 2: Transcription Factor Activity. To characterize the empirical transcriptomic correlates of sociodemographic factors in terms of their upstream gene-regulatory influences (41), we conducted promoter-based bioinformatics analyses of transcription factor-binding motif (TFBM) prevalence for a prespecified set of transcription factors involved in inflammation (NF-κB and AP-1), type I IFN response (IFN-stimulated response element; ISRE), and neuroendocrine activity [CREB, which mediates SNS-induced β-adrenergic signaling, and the glucocorticoid receptor (GR) which mediates cortisol signaling from the hypothalamus-pituitary-adrenal axis] (45). Results showed significant demographic variation in activity of each transcription factor (Fig. 3*B*), with particularly marked effects for the immunoregulatory transcription factors (NF-κB, AP-1, ISRE) and CREB.

Level 3: Cellular Origins. To characterize empirical transcriptomic correlates of sociodemographic variation in terms of their shared cellular origins (41), we conducted Transcript Origin Analyses (43) of the same sets of differentially expressed genes using reference data from previous genome-wide transcriptional profiling of isolated leukocyte subsets (Gene Expression Omnibus GSE101489) (46). Results indicated significant demographic variation in activity of each cell type (Fig. 3*C*), with effects particularly prevalent for the myeloid lineage immune cells involved in proinflammatory and type I IFN innate immune responses (i.e., monocytes, dendritic cells, and neutrophils).

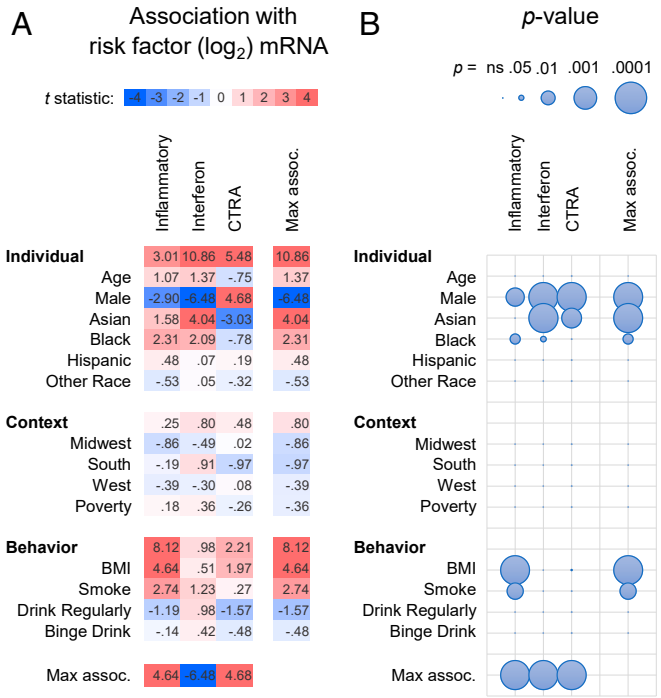

Fig. 1. Demographic variation in expression of inflammation- and type I IFN-related genes. Differential expression composites of 19 proinflammatory genes, 32 type I IFN-related genes, and their difference (i.e., CTRA profile) as a function of individual demographic characteristics, contextual characteristics, and biobehavioral factors. Estimates come from linear statistical modeling of $\log_2$ gene expression values from $n = 1{,}069$ study participants with adjustment for all other listed factors as well as assay technical covariates. Effects are expressed as (*A*) *t*-statistics (effect size/SE; red: up-regulated; blue: down-regulated) and as (*B*) statistical significance (symbol area proportional to $-\log_{10} p$). In *A*, rows with left-adjusted bold labels contain omnibus *F* statistics summarizing all parameters within the category (Individual, Context, or Behavior). Parameters represent effects of age (in years), sex (male relative to female), race/ethnicity categories (relative to non-Hispanic whites), US Census region (relative to Northeast region 1), poverty (relative to household income above poverty line), BMI (kg/m^2), history of regular smoking (relative to none), regular alcohol consumption (relative to none), and frequency of binge drinking (7-point ordinal scale). "Max assoc." indicates the maximum magnitude of association observed over all demographic dimensions or over all gene sets analyzed. *SI Appendix*, Table S1, contains the underlying numerical data for this figure.

Discussion

This population-representative transcriptome profiling study reveals significant demographic variations in the expression of inflammatory and type I IFN response genes that emerge by young adulthood and are thus active decades before chronic diseases commonly manifest in older age. Significant sociodemographic variations in gene expression appeared at every level of analysis examined, including prespecified sets of inflammatory and type I IFN indicator genes (level 1) and empirically mapped genome-wide transcriptional differences analyzed in terms of transcription factor coregulation (level 2) and coexpression in myeloid lineage immune cells (particularly monocytes, dendritic cells, and neutrophils) (level 3). As the largest social genomics study conducted to date, as well as the most demographically diverse sample so far analyzed, the unprecedented power available in this sample allowed for the detection of significant variations in gene regulation as a function of every sociodemographic factor analyzed, including individual demographic characteristics (age, sex, race/ethnicity) and social context (family poverty, region of residence). Molecular characteristics also varied as a function of biobehavioral factors (BMI, smoking, alcohol consumption) that vary across sociodemographic groups. However, all analyses controlled for biobehavioral factors and continued to identify significant molecular correlates of individual and contextual demographic features. These data establish a molecular framework for analyzing social disparities in late-life health and mortality in terms of sociodemographic variations in gene regulation that emerge decades earlier in young adulthood (1–6) and can thus exert a temporally extended impact on the molecular processes that culminate in late-life chronic disease.

Expression of inflammatory and type I IFN response genes varied significantly as a function of each of the sociodemographic factors examined, but the magnitude of such effects varied greatly across factors. Using an absolute effect-size reference point of 20% difference in RNA abundance, both racial and ethnic identity and BMI were associated with substantially more differentially expressed genes than were the other factors analyzed (Fig. 3*A*). These results are particularly relevant for understanding racial disparities in chronic disease risk, as blacks showed greater expression of both a broad composite of inflammatory genes and the more specific F5 inflammatory gene module (Fig. 2*A*, *JUND* transcription factor and COX1/*PTGS1* inflammatory mediator) as well as bioinformatic indications of NF-κB and myeloid lineage immune-cell activity (monocytes, neutrophils, and dendritic cells) (Fig. 3). Similar effects of race/ethnicity and BMI emerged when comparing the number of statistically significant transcript associations (*SI Appendix*, Fig. S2), although this metric also revealed substantial sex differences that manifest as quantitatively large differences in expression of a relatively small number of genes (compare sex differences in *SI Appendix*, Fig. S2, vs. Fig. 3*A*). Greater expression of the CTRA profile in males may shed light on the well-documented longevity disadvantage of males relative to females (47). The present findings are also broadly consistent with previous studies of older adults

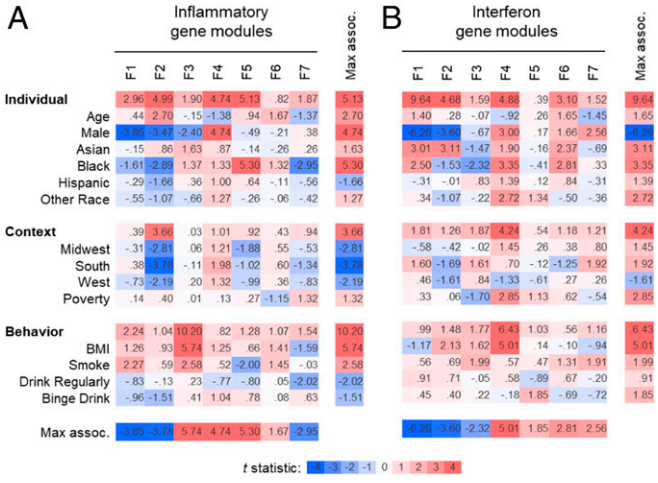

Fig. 2. Demographic variation in expression of coregulated modules of inflammatory genes (*A*) and type I IFN genes (*B*). Principal factor analysis empirically identified seven coregulated gene modules within both the overall inflammatory and type I IFN gene sets (*SI Appendix*, Fig. S1 and Dataset S1). Data show variations in expression of these coregulated gene modules as a function of individual demographic characteristics, contextual conditions, and behavioral factors. Estimates come from linear statistical modeling as in Fig. 1, with effects expressed as *t*-statistics (effect size/SE; red: up-regulated; blue: down-regulated) in subcategory rows with right-adjusted nonbold labels. Rows with left-adjusted bold labels contain omnibus *F* statistics summarizing all parameters in a given category of influence (Individual, Context, Behavior). "Max assoc." indicates the maximum magnitude of association observed over all demographic dimensions or over all gene sets analyzed. Dataset S1 contains underlying numerical data for this figure.

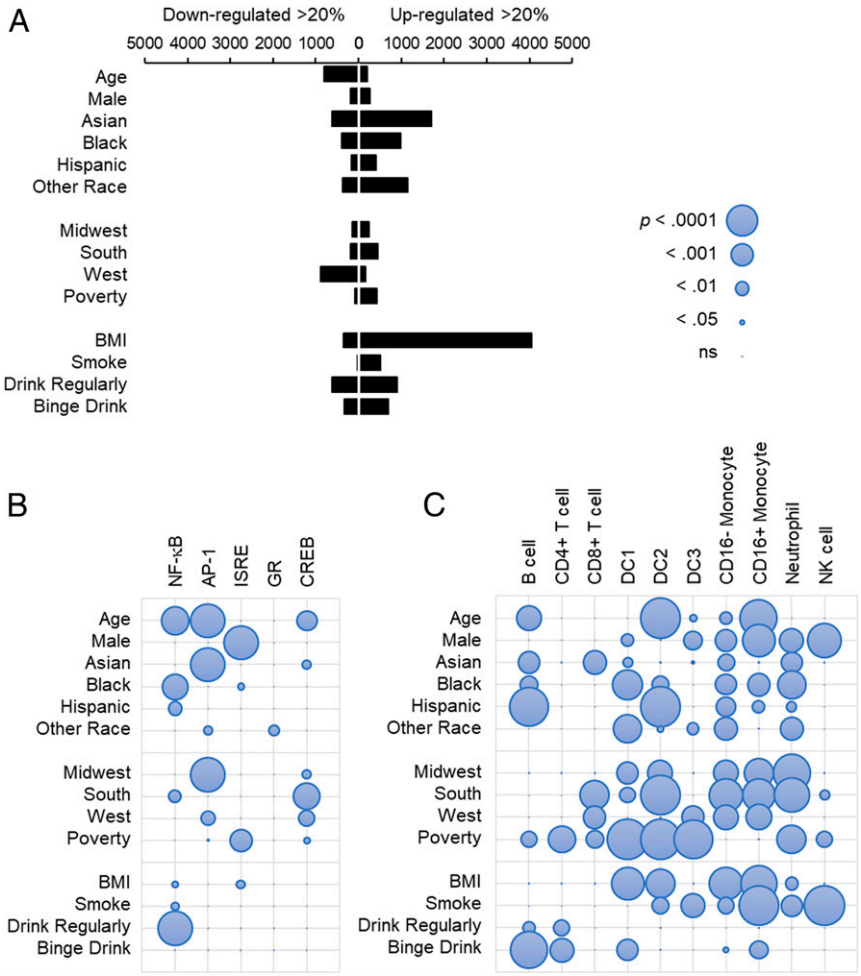

Fig. 3. Demographic variation in empirical gene expression and bioinformatic inferences of transcription factor activity and cellular activation. (*A*) Number of genes up- and down-regulated by >20% as a function of each sociodemographic parameter (Dataset S2 lists individual transcripts). (*B*) Bioinformatic analysis of promoter TFBM distributions for targeted proinflammatory transcription factors (NF-κB, AP-1), IFN response factors (ISRE), SNS response factors (CREB), and the GR for each set of differentially expressed genes. Symbol area is proportional to statistical significance ($-\log_{10} p$); see legend at the *Right* in *A*. (*C*) Bioinformatic analysis of shared cellular origins for each set of differentially expressed genes. Symbol area is proportional to the maximal statistical significance of results for up- vs. down-regulated genes ($-\log_{10} p$, as in legend at the *Right* in *A*).

that have documented marked differences in white-blood-cell gene expression as a function of sex (48, 49), race (50–52), and BMI (19, 53, 54) (admittedly in smaller and less representative convenience samples). It is difficult to quantitatively compare the demographic variations in gene expression observed here in young adults to the effects previously observed in older adults because no other large, population-representative study has so far reported any general demographic analysis of genome-wide transcriptional profiles. However, several population health studies are currently collecting transcriptome data from older adult samples, which should allow for life-span developmental comparisons in the future.

The biological character of the transcriptome variations observed here differed markedly across demographic dimensions, with individual demographic factors showing the most pronounced association with type I IFN activity and biobehavioral factors associating most strongly with inflammatory gene expression (Fig. 1). In contrast, social context (family poverty and geographic region) showed little association with broad composite measures of inflammatory or type I IFN activity. However, the lack of association with level 1 global composite measures masked significant regional differences in more specific measures of proinflammatory transcription factor activity (level 2) (Fig. 3*B*), monocyte and neutrophil activation (level 3) (Fig. 3*C*), and empirically coregulated subcomponents of a global inflammatory gene set (particularly F2 involving the transcription factor, *FOS*, and the inflammatory mediators COX2/*PTGS2* and IL8/*CXCL8*) (Fig. 2*A*). Elevated health risk in the southern United States in particular (8, 9) may relate to the observed upregulation of the F2 inflammatory gene module (*FOS*/*PTGS2*/*CXCL8*) and associated differences in activation of NF-κB, classical and nonclassical monocytes, and the CREB transcription factor involved in β-adrenergic signaling from the SNS (Fig. 3 *B* and *C*). Analyses of level 1 global composite scores also missed substantial family poverty-related differences in IFN response factor activity (level 2) (Fig. 3*B*), dendritic cell activation (level 3) (Fig. 3*C*), and the IFN coregulatory module F4 (*IFI27L1*, *IFI27L2*) (Fig. 2*B*).

Beyond this study's substantive implications for the early life biological development of sociodemographic disparities in late-life health, the pattern of results observed here may also have significant implications for analytic approaches in social genomics. Whereas level 1 analyses of a-priori–specified gene composites showed little effect of contextual variables such as residential region and family poverty, level 2 and 3 analyses that map empirical differences in gene expression and interpret them in terms of prespecified substantive hypotheses involving

transcription factor activity (level 2) and immune-cell mediators (level 3) identified multiple effects of region and poverty (compare level 1 contextual effects in Fig. 1*B* with those of level 2 and 3 analyses in Fig. 3 *B* and *C*). Similarly, decomposition of the global inflammatory and IFN gene composites into empirically coregulated gene modules also revealed significant associations that were missed in analyses of global composite scores (compare contextual effects in Fig. 1*A* with those in Fig. 2 *A* and *B*). The differential sensitivity of level 1 analyses (prespecified gene sets) and level 2 and 3 analyses (empirically identified gene sets interpreted in terms of prespecified hypotheses regarding their shared biological function) underscores the utility of deploying multilevel bioinformatic approaches to characterize transcriptomic diversity, rather than relying solely on prespecified gene composites to assess complex physiological processes. Previous analyses have noted the conceptual and statistical advantages of "abstractionist" bioinformatic analyses that treat empirical differences in genome-wide transcriptional profiles as input into higher-order bioinformatics analyses testing specific substantive hypotheses involving transcription factor activity and cellular differentiation (levels 2 and 3) (37, 55). The present findings are consistent with that perspective and lay the groundwork for more differentiated analyses of inflammatory biology in future research using both bioinformatic inferences of latent causal factors (i.e., transcription factors and cellular context) as well as empirically refined sets of indicator genes (e.g., the seven coregulatory modules empirically identified within each of the global indicator gene sets analyzed here).

These data also have more specific implications for the analysis of inflammation as a biological mechanism of sociodemographic differences. The present analyses identify seven distinct coregulated gene modules within the overall set of 19 general inflammatory indicator genes (*SI Appendix*, Fig. S1). These modules typically involved a transcription factor accompanied by a set of inflammatory effector molecules (i.e., cytokines, prostaglandin synthases, chemokines, and other innate immune response genes). This analysis also revealed that one of the most commonly measured proinflammatory cytokines, *IL6*, was largely uncorrelated with the activity of the other six inflammatory gene modules. The other six inflammatory gene modules also showed patterns of demographic variation that differed from those of *IL6*. These results suggest that *IL6* (and its downstream reporters such as *CRP*) should not be used as a summary measure of inflammatory activity; several other distinct proinflammatory gene modules also exist and drive the expression of empirically distinct inflammatory effector systems involving *IL1B*, *TNF*, IL8/*CXCL8*, and COX2/*PTGS2*.

This study has several strengths, most notably the application of genome-wide transcriptional profiling to a large, population-based sample in Add Health with national representation of all racial, ethnic, geographic, and income subgroups. However, these findings are also limited in several respects. These data come from a contemporary representative sample of community-dwelling young adults in the United States, and it is unclear whether similar patterns would hold for other groups that differ in age, health status, global region (particularly given the US health disadvantage in early and midlife) (9), or other factors. It will be important to replicate the present analyses in other samples to assess the generalizability of these findings to other age groups or global regions. Given the restricted age range in this sample, these data likely under-represent the total range of transcriptomic variation across the adult life span. These analyses document the presence of significant demographic differences in human genome function at the time of young adulthood, but it is possible that such differences emerge even earlier in development (e.g., adolescence, childhood, infancy, or in utero) (56–59). A critical topic for future population-based genomic research will be pushing back the etiological time line even further than achieved here to identify the specific developmental periods in which the demographic differences in gene expression first appear. These data come from an observational study, and it is unclear whether the observed associations represent causal effects of demographic or biobehavioral factors. Demographic variations may stem from genetic differences, socioenvironmental exposures and consequent neurobiological responses (e.g., SNS or hypothalamus-pituitary-adrenal axis activity), or differential physicochemical and microbial exposures. The quantitative relationship between the specific molecular differences observed here and subsequent disease/mortality risks is not yet known and remains to be defined in future research. However, the present study analyzed gene expression through the lens of two biological processes that have previously been shown to play a significant role in chronic disease risk and longevity: inflammation and type I IFN responses (16). This study was not designed to provide a comprehensive discovery-based analysis of transcriptomic differences, and other biological processes in addition to those analyzed here may also differ as a function of demographic characteristics. The gene coregulatory modules identified here are derived from observational data in a specific cohort, and the structure of those modules may differ in other populations; like the findings for cellular and transcription factor effects, these findings need to be replicated in future research. It is also possible that different effects would be observed in analyses using different representations of sociodemographic variation (e.g., more differentiated measure of socioeconomic status than the poverty classification used here, although we found no significant difference in expression of the overall inflammatory or type I IFN gene composites as a function of household income or educational attainment) (*SI Appendix*, Table S2). Finally, it is important to note that we analyzed individual sociodemographic and biobehavioral variables as distinct influences on gene expression, but some of these factors are empirically correlated in the social ecology (e.g., race and poverty; poverty, BMI, and smoking, etc.). As such, the present covariate-adjusted estimates will underestimate the magnitude of each factor's overall association with gene expression (i.e., unadjusted for other correlated risk factors).

Despite these limitations, this study provides a comprehensive map of the landscape of demographic variation in human gene expression in the contemporary US, and it identifies the emergence of marked differences in inflammatory and type I IFN gene expression by young adulthood. These data establish a molecular framework for understanding social disparities in late-life health and mortality in terms of disparities in gene regulation that emerge decades earlier (and may initially develop even earlier than observed here—in infancy, childhood, or adolescence). These findings also provide a framework for reducing health disparities by mitigating molecular risk gradients before they develop into overt disease. For example, the differential expression of inflammatory and type I IFN genes observed here could serve as outcome biomarkers to assess the impact of interventions that seek to mitigate health disparities by altering social contexts or family environments in early life (58, 60). Indeed, the identification of demographic gradients in inflammatory and antiviral gene regulation in young adults underscores the need to initiate social, behavioral, and policy interventions early in life in order to most effectively mitigate social disparities in disease risk that would otherwise become clinically evident only decades later in older adulthood (9, 12, 15, 61–63).

Methods

Sample and Survey Procedures. Data come from Add Health, a nationally representative study of US adolescents in grades 7 to 12 in 1994 to 1995 who have been followed into adulthood over five waves of data collection. We used data from sample 1 Wave V (2016 to 2017) that was collected when respondents were aged 32 to 42. Study design, interview procedures, and demographic and biobehavioral assessments have been previously described (13, 17). Participants provided written informed consent and all procedures were approved by the University of North Carolina School of Public Health

Institutional Review Board. Details on measurement and coding are provided in *SI Appendix*.

Blood Transcriptome Profiling. Venipuncture whole-blood samples were assayed by RNA sequencing using a 3′ messenger RNA counting assay (Lexogen QuantSeq 3′ FWD) on an Illumina HiSeq 4000 system following the manufacturers' standard protocols. The 65-base single-strand reads were mapped to the ENSEMBL hg38 human transcriptome to estimate gene-level transcript abundance using STAR. Transcript abundance values were normalized using 11 reference genes (64) and analyzed by linear statistical models relating $\log_2$-transcript abundance to individual demographic characteristics (age, sex, race/ethnicity), sociodemographic contextual characteristics (US region, family poverty status), biobehavioral factors (BMI, smoking, alcohol consumption), and technical covariates [sample RNA integrity number (RIN), assay plate, sequencing depth, and profile consistency with other samples].

Sociodemographic Variables and Technical Controls. Variables were coded as follows: age (continuous self-reported years); sex (self-reported biologically assigned male sex at birth, coded by an indicator relative to reference point female); race/ethnicity (self-identified Asian, non-Hispanic black, Hispanic, and other race/ethnicity, each coded by an indicator relative to reference point non-Hispanic white); US region (census regions 2 to 4: Midwest, South, and West, each coded by an indicator relative to reference point region 1, Northeast); family poverty status (self-reported household income less than or equal to 2015 US federal poverty level based on household size, coded by an indicator relative to nonpoverty status); BMI (continuous kg/m^2 derived from self-reported continuous height and weight); smoking history (self-reported ever smoked coded by an indicator relative to never smoked reference point); and alcohol consumption [represented as two variables: one "regular drinking" variable indicating whether participants self-reported drinking beer, wine, or liquor every day or almost every day, relative to less frequent drinking during the past 12 mo; and a second "binge drinking" ordinal variable reflecting days during the past 12 mo during which participants drank (female 4/male 5) or more drinks in a row, (coded none = 0, 1 to 2 d/y = 1, 3 to 12 d/y = 1 d/mo = 2, 2 to 3 d/mo = 3, 1 to 2 d/wk = 4, 3 to 5 d/wk = 5, every/almost every day = 6)]; assay batch (nominal indicators for plates 1 to 11 relative to reference point plate 12); sample RIN (continuous 0 to 10), total mapped reads per sample (continuous/10^6); read alignment rate (continuous percentage); and profile consistency (average Pearson *r* with 95 other samples).

Analytic Methods. Data analyses examined inflammatory and type I IFN gene regulation at three distinct levels of biological function: 1) expression of a-priori-defined sets of inflammatory and type I IFN indicator genes (42); 2) activity of transcription factors involved in mediating inflammatory, type I IFN, SNS, and neuroendocrine responses (34, 36); and 3) activation of specific immune-cell subsets involved in inflammatory and IFN gene expression, particularly monocytes, dendritic cells (DCs), and neutrophils (38, 39, 43).

For level 1 analyses, prespecified general inflammatory and IFN composite scores were computed by averaging standardized expression values for 19 genes involved in inflammation or for 32 genes involved in type I IFN responses (42). We also examined a previously derived CTRA indicator contrast score computed as the difference between inflammatory and type I IFN composites (inflammatory composite score − type I IFN composite score). Each of these molecular parameters was tested for differential expression as a function of individual demographic characteristics (age, sex, race/ethnicity), and contextual conditions (US region, family poverty status), with ancillary analyses examining potentially confounding effects of biobehavioral factors (BMI, smoking, alcohol consumption), while controlling for technical covariates as noted above. To avoid capitalizing on chance due to multiple testing, we followed standard biostatistical procedures by computing a single integrated omnibus hypothesis test of our primary hypothesis that there exists significant sociodemographic variation (either individual or contextual) in the expression of one or more of the examined gene sets (65–68). Contingent on a significant omnibus test of global sociodemographic variation in gene set expression, we conducted interpretive follow-up analyses testing for significant sociodemographic variation in expression of each gene composite in isolation [with a false discovery rate (69) correction for multiple testing]. For gene sets showing a significant omnibus test of global sociodemographic variation in activity, we presented the individual parameter estimates underlying that global result for descriptive/interpretive purposes and conducted follow-up nested aggregate hypothesis tests to assess the respective effects of individual vs. contextual demographic factors (again with a false discovery rate correction for multiple testing). Ancillary aggregate hypothesis tests also examine biobehavioral factors that might potentially confound sociodemographic effects. Throughout these analyses individual parameter estimates are presented for interpretive purposes only and do not serve as the analytic basis for primary substantive conclusions. To ensure that the a-priori-specified global inflammatory and type I IFN composite scores did not obscure the effects of more differentiated coregulated gene modules within each global set, we also conducted exploratory follow-up analyses of the analyzed gene sets to map their fine-grain coregulatory structure, using principal factor analysis (70) to identify sets of coregulated genes while accounting for residual sources of sampling variability (i.e., unique variance components).

For level 2 and 3 analyses, empirical variations in genome-wide transcriptional profiles were mapped by identifying all genes showing >20% differential expression as a function of a binary demographic indicator variable or a 4-SD difference in a continuous demographic variable (ranging from 2 SD below the mean to 2 SD above the mean). Gene-specific statistical significance was based on a 5% dependent false discovery rate allowing for potential correlation among genes (71). In level 2 analyses, transcription factor activity was assessed by TELiS bioinformatic analysis (45) of RefSeq core promoter DNA sequences for all genes showing a maximum-likelihood point estimate of >20% differential expression as a function of a target demographic variable. Genes were screened into TELiS analyses based on differential expression effect size because effect-size–screened gene lists have been found to be more replicable than those based on *p*- or *q*-value screening (42, 72–75). TELiS analyses used TRANSFAC position-specific weight matrices for NF-κB, AP-1, ISRE, CREB, and the GR (76), with detection by the TRANSFAC mat_sim information criterion and statistical significance assessed by bootstrap resampling of linear model residual vectors to account for correlation among genes (77). Level 3 analyses examined the relative contributions of 10 leukocyte subsets to the same set of differentially expressed genes using Transcript Origin Analysis (43) based on reference transcriptome profiles derived from isolated cell samples (Gene Expression Omnibus GSE101489) (46) and bootstrap analysis of statistical significance.

Additional analytic details are available in *SI Appendix*. Analyses were performed using SAS 9.4 software.

Data Availability. Add Health data are available at https://www.cpc.unc.edu/projects/addhealth/documentation/. SAS code used in these analyses is available upon request from the corresponding authors.

ACKNOWLEDGMENTS. This research was supported by NIH Grants R01-HD087061 (specifying the present analyses), P30-AG017265, R01-AG043404, and R01-AG033590; and by the Jacobs Center for Productive Youth Development (University of Zürich). This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris (University of North Carolina at Chapel Hill) and funded by Grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations (https://www.cpc.unc.edu/projects/addhealth/about/funders).

1. G. A. Mensah, A. H. Mokdad, E. S. Ford, K. J. Greenlund, J. B. Croft, State of disparities in cardiovascular health in the United States. *Circulation* **111**, 1233–1241 (2005).
2. E. Ward *et al*., Cancer disparities by race/ethnicity and socioeconomic status. *CA Cancer J. Clin.* **54**, 78–93 (2004).
3. E. K. Spanakis, S. H. Golden, Race/ethnic difference in diabetes and diabetic complications. *Curr. Diab. Rep.* **13**, 814–823 (2013).
4. L. M. Lines, J. M. Wiener, *Racial and Ethnic Disparities in Alzheimer's Disease: A Literature Review* (Office of The Assistant Secretary for Planning and Evaluation, Department of Health and Human Services, 2014).
5. NCHS, *Health, United States, 2015: With Special Feature on Racial and Ethnic Health Disparities* (National Center for Health Statistics, 2016).
6. M. S. Shiels *et al*., Trends in premature mortality in the USA by sex, race, and ethnicity from 1999 to 2014: An analysis of death certificate data. *Lancet* **389**, 1043–1054 (2017).
7. P. A. Braveman, S. Egerter, *Overcoming Obstacles to Health. Report from the Robert Wood Johnson Foundation to the Commission to Build a Healthier America* (Robert Wood Johonson Foundation, 2008).
8. C. J. Murray *et al*., Eight Americas: Investigating mortality disparities across races, counties, and race-counties in the United States. *PLoS Med.* **3**, e260 (2006).
9. I. O. M. National Research Council, *U.S. Health in International Perspective: Shorter Lives, Poorer Health* (National Academies Press, 2013).
10. G. E. Miller *et al*., Low early-life social class leaves a biological residue manifested by decreased glucocorticoid and increased proinflammatory signaling. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 14716–14721 (2009).
11. G. E. Miller, E. Chen, K. J. Parker, Psychological stress in childhood and susceptibility to the chronic diseases of aging: Moving toward a model of behavioral and biological mechanisms. *Psychol. Bull.* **137**, 959–997 (2011).

12. Y. C. Yang, K. Gerken, K. Schorpp, C. Boen, K. M. Harris, Early-life socioeconomic status and adult physiological functioning: A life course examination of biosocial mechanisms. *Biodemogr. Soc. Biol.* **63**, 87–103 (2017).
13. K. M. Harris, An integrative approach to health. *Demography* **47**, 1–22 (2010).
14. K. M. Harris, P. Gordon-Larsen, K. Chantala, J. R. Udry, Longitudinal trends in race/ethnic disparities in leading health indicators from adolescence to young adulthood. *Arch. Pediatr. Adolesc. Med.* **160**, 74–81 (2006).
15. Y. C. Yang *et al*., Social relationships and physiological determinants of longevity across the human life span. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 578–583 (2016).
16. C. E. Finch, *The Biology of Human Longevity: Inflammation, Nutrition, and Aging in the Evolution of Lifespans* (Academic Press, 2007).
17. K. M. Harris *et al*., Cohort profile: The national longitudinal study of adolescent to adult health (Add health). *Int. J. Epidemiol.* **48**, 1415–1415k (2019).
18. I. Amit *et al*., Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* **326**, 257–263 (2009).
19. M. Preininger *et al*., Blood-informative transcripts define nine common axes of peripheral blood gene expression. *PLoS Genet.* **9**, e1003362 (2013).
20. J. Jylhävä *et al*., Methylomic predictors demonstrate the role of NF-κB in old-age mortality and are unrelated to the aging-associated epigenetic drift. *Oncotarget* **7**, 19228–19241 (2016).
21. J. Jylhävä *et al*., Identification of a prognostic signature for old-age mortality by integrating genome-wide transcriptomic data with the conventional predictors: The vitality 90+ study. *BMC Med. Genomics* **7**, 54 (2014).
22. E. Kingwell *et al*., Multiple sclerosis: Effect of beta interferon treatment on survival. *Brain* **142**, 1324–1333 (2019).
23. W. M. Passtoors *et al*., Transcriptional profiling of human familial longevity indicates a role for ASF1A and IL7R. *PLoS One* **7**, e27759 (2012).
24. M. Puzianowska-Kuźnicka *et al*., Interleukin-6 and C-reactive protein, successful aging, and mortality: The PolSenior study. *Immun. Ageing* **13**, 21 (2016).
25. P. Sebastiani *et al*., Biomarker signatures of aging. *Aging Cell* **16**, 329–338 (2017).
26. D. Melzer *et al*., Gene expression biomarkers and longevity. *Annu. Rev. Gerontol. Geriatr.* **33**, 233–258 (2013).
27. S. Giovannini *et al*., Interleukin-6, C-reactive protein, and tumor necrosis factor-alpha as predictors of mortality in frail, community-living elderly individuals. *J. Am. Geriatr. Soc.* **59**, 1679–1685 (2011).
28. G. Gasparini, R. Longo, R. Sarmiento, A. Morabito, Inhibitors of cyclo-oxygenase 2: A new class of anticancer agents? *Lancet Oncol.* **4**, 605–615 (2003).
29. R. E. Harris, J. Beebe, G. A. Alshafie, Reduction in cancer risk by selective and non-selective cyclooxygenase-2 (COX-2) inhibitors. *J. Exp. Pharmacol.* **4**, 91–96 (2012).
30. A. Mantovani, Molecular pathways linking inflammation and cancer. *Curr. Mol. Med.* **10**, 369–373 (2010).
31. P. M. Ridker *et al*.; CANTOS Trial Group, Antiinflammatory therapy with canakinumab for atherosclerotic disease. *N. Engl. J. Med.* **377**, 1119–1131 (2017).
32. P. M. Ridker *et al*.; CANTOS Trial Group, Effect of interleukin-1β inhibition with canakinumab on incident lung cancer in patients with atherosclerosis: Exploratory results from a randomised, double-blind, placebo-controlled trial. *Lancet* **390**, 1833–1842 (2017).
33. F. K. Swirski, M. Nahrendorf, Cardioimmunology: The immune system in cardiac homeostasis and disease. *Nat. Rev. Immunol.* **18**, 733–744 (2018).
34. M. R. Irwin, S. W. Cole, Reciprocal regulation of the neural and innate immune systems. *Nat. Rev. Immunol.* **11**, 625–632 (2011).
35. S. W. Cole, Social regulation of human gene expression: Mechanisms and implications for public health. *Am. J. Public Health* **103** (suppl. 1), S84–S92 (2013).
36. S. W. Cole, Human social genomics. *PLoS Genet.* **10**, e1004601 (2014).
37. S. W. Cole, The conserved transcriptional response to adversity. *Curr. Opin. Behav. Sci.* **28**, 31–37 (2019).
38. N. D. Powell *et al*., Social stress up-regulates inflammatory gene expression in the leukocyte transcriptome via β-adrenergic induction of myelopoiesis. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 16574–16579 (2013).
39. T. Heidt *et al*., Chronic variable stress activates hematopoietic stem cells. *Nat. Med.* **20**, 754–758 (2014).
40. D. B. McKim *et al*., Social stress mobilizes hematopoietic stem cells to establish persistent splenic myelopoiesis. *Cell Rep.* **25**, 2552–2562.e3 (2018).
41. G. Gibson, *A Primer of Human Genetics* (Sinauer Associates, 2014).
42. B. L. Fredrickson *et al*., A functional genomic perspective on human well-being. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13684–13689 (2013).
43. S. W. Cole, L. C. Hawkley, J. M. Arevalo, J. T. Cacioppo, Transcript origin analysis identifies antigen-presenting cells as primary targets of socially regulated gene expression in leukocytes. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3080–3085 (2011).
44. R. M. Groves, F. Fowler, Jr, M. Couper, J. Lepkowski, E. Singer, and R. Tourangeau, *Survey Methodology* (John Wiley and Sons, ed. 2, 2009).
45. S. W. Cole, W. Yan, Z. Galic, J. Arevalo, J. A. Zack, Expression-based monitoring of transcription factor activity: The TELiS database. *Bioinformatics* **21**, 803–810 (2005).
46. D. S. Black, S. W. Cole, G. Christodoulou, J. C. Figueiredo, Genomic mechanisms of fatigue in survivors of colorectal cancer. *Cancer* **124**, 2637–2644 (2018).
47. B. B. Kalben, Why men die younger. *N. Am. Actuar. J.* **4**, 83–111 (2000).
48. A. R. Whitney *et al*., Individuality and variation in gene expression patterns in human blood. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1896–1901 (2003).
49. M. E. Levine, S. W. Cole, D. R. Weir, E. M. Crimmins, Childhood and later life stressors and increased inflammatory gene expression at older ages. *Soc. Sci. Med.* **130**, 16–22 (2015).
50. J. D. Storey *et al*., Gene-expression variation within and among human populations. *Am. J. Hum. Genet.* **80**, 502–509 (2007).
51. B. A. Kohrt *et al*., Psychological resilience and the gene regulatory impact of post-traumatic stress in Nepali child soldiers. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 8156–8161 (2016).
52. T. W. McDade *et al*., Genome-wide profiling of RNA from dried blood spots: Convergence with bioinformatic results derived from whole venous blood and peripheral blood mononuclear cells. *Biodemogr. Soc. Biol.* **62**, 182–197 (2016).
53. A. P. Nath, D. Arafat, G. Gibson, Using blood informative transcripts in geographical genomics: Impact of lifestyle on gene expression in fijians. *Front. Genet.* **3**, 243 (2012).
54. M. R. Mehl, C. L. Raison, T. W. W. Pace, J. M. G. Arevalo, S. W. Cole, Natural language indicators of differential gene regulation in the human immune system. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 12554–12559 (2017).
55. S. W. Cole, Elevating the perspective on human stress genomics. *Psychoneuroendocrinology* **35**, 955–962 (2010).
56. E. Chen *et al*., Genome-wide transcriptional profiling linked to social class in asthma. *Thorax* **64**, 38–43 (2009).
57. G. E. Miller *et al*., Maternal socioeconomic disadvantage is associated with transcriptional indications of greater immune activation and slower tissue maturation in placental biopsies and newborn cord blood. *Brain Behav. Immun.* **64**, 276–284 (2017).
58. G. E. Miller, G. H. Brody, T. Yu, E. Chen, A family-oriented psychosocial intervention reduces inflammation in low-SES African American youth. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 11287–11292 (2014).
59. S. W. Cole *et al*., Transcriptional modulation of the developing immune system by early life social adversity. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 20578–20583 (2012).
60. J. Ludwig *et al*., Neighborhoods, obesity, and diabetes: A randomized social experiment. *N. Engl. J. Med.* **365**, 1509–1519 (2011).
61. G. Conti *et al*., Primate evidence on the late health effects of early-life adversity. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 8866–8871 (2012).
62. F. Campbell *et al*., Early childhood investments substantially boost adult health. *Science* **343**, 1478–1485 (2014).
63. K. P. Jakubowski, J. M. Cundiff, K. A. Matthews, Cumulative childhood adversity and adult cardiometabolic disease: A meta-analysis. *Health Psychol.* **37**, 701–715 (2018).
64. E. Eisenberg, E. Y. Levanon, Human housekeeping genes, revisited. *Trends Genet.* **29**, 569–574 (2013).
65. J. Cao, S. Zhang, Multiple comparison procedures. *JAMA* **312**, 543–544 (2014).
66. A. D. Althouse, Adjust for multiple comparisons? It's not that simple. *Ann. Thorac. Surg.* **101**, 1644–1645 (2016).
67. R. Bender, S. Lange, Adjusting for multiple testing: When and how? *J. Clin. Epidemiol.* **54**, 343–349 (2001).
68. R. J. Feise, Do multiple outcome measures require p-value adjustment? *BMC Med. Res. Methodol.* **2**, 8 (2002).
69. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
70. S. A. Mulaik, *Foundations of Factor Analysis* (CRC Press, 2010).
71. Y. Benjamini, D. Yekateuli, The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29**, 1165–1188 (2001).
72. S. W. Cole, Z. Galic, J. A. Zack, Controlling false-negative errors in microarray differential expression analysis: A PRIM approach. *Bioinformatics* **19**, 1808–1816 (2003).
73. L. Shi *et al*., The balance of reproducibility, sensitivity, and specificity of lists of differentially expressed genes in microarray studies. *BMC Bioinformatics* **9** (suppl. 9), S10 (2008).
74. D. M. Witten, R. Tibshirani, "A comparison of fold-change and the t-statistic for microarray data analysis" (Stanford University Technical Report, 2007).
75. A. W. Norris, C. R. Kahn, Analysis of gene expression in pathophysiological states: Balancing false discovery and false negative rates. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 649–653 (2006).
76. E. Wingender, P. Dietze, H. Karas, R. Knüppel, TRANSFAC: A database on transcription factors and their DNA binding sites. *Nucleic Acids Res.* **24**, 238–241 (1996).
77. B. Efron, R. J. Tibshirani, *An Introduction to the Bootstrap* (Chapman & Hall, 1993).