



Invited Commentary

Invited Commentary: Integrating Genomics and Social Epidemiology—Analysis of Late-Life Low Socioeconomic Status and the Conserved Transcriptional Response to Adversity

Daniel W. Belsky* and Noah Snyder-Mackler

* Correspondence to Daniel W. Belsky, 2020 W. Main St, Suite 201, Durham, NC 27708 (e-mail: dbelsky@duke.edu).

Initially submitted February 6, 2017; accepted for publication February 16, 2017.

Socially disadvantaged children face increased morbidity and mortality as they age. Understanding mechanisms through which social disadvantage becomes biologically embedded and devising measurements that can track this embedding are critical priorities for research to address social gradients in health. The analysis by Levine et al. (*Am J Epidemiol.* 2017;000(0):000–000) of genome-wide gene expression in a subsample of US Health and Retirement Study participants suggests important new directions for the field. Specifically, findings suggest promise in integrating gene expression data into population studies and provide further evidence for the conserved transcriptional response to adversity as a marker of biological embedding of social disadvantage. The study also highlights methodological issues related to the analysis of gene expression data and social gradients in health and a need to examine the conserved transcriptional response to adversity alongside other proposed measurements of biological embedding. Looking to the future, advances in genome science are opening new opportunities for socio-genomic epidemiology.

aging; gene expression; genomics; social epidemiology

Abbreviations: CTRA, conserved transcriptional response to adversity; HRS, Health and Retirement Study; NF- κ B, nuclear factor κ -light-chain-enhancer of activated B cells; SES, socioeconomic status.

Socially disadvantaged children face excess morbidity and mortality as they age (1, 2). This burden arises at least partly from social exposures “getting under the skin” to influence pathways of disease and disability (3, 4). Advances in genome science offer new opportunities to understand and quantify mechanisms of this biological embedding. New assays allow direct measurement of genomic changes that are hypothesized to mediate effects of social disadvantage on health and aging, such as shortening of telomeres (5–7) and differential methylation of DNA (8–10). In this issue of the *Journal*, Levine et al. (11) focus their attention on another aspect of gene regulation: gene expression. They report preliminary evidence that older adults with lower socioeconomic status (SES) exhibit a pattern of increased inflammatory and decreased antiviral peripheral-blood gene expression known as “conserved transcriptional response to adversity” (CTRA). Their study highlights challenges and opportunities for the integration of genomics into social epidemiology.

The CTRA is characterized by differential expression of genes related to immune function (12). Broadly, these changes have been interpreted to indicate a down-regulation of antiviral defenses and a corresponding up-regulation of proinflammatory wound-healing and antibacterial defenses (13, 14). CTRA-related changes, in theory, facilitate an adaptive response to acute threats that could result in injury. However, chronic CTRA activation could leave an organism vulnerable to viral infection and subject to collateral damage from systemic inflammation, ultimately leading to disease, disability, and death.

Levine et al. used microarrays to measure genome-wide gene expression in peripheral blood RNA from 120 participants in the US Health and Retirement Study (HRS). They linked gene expression data with longitudinal data on HRS participants’ education and household income. They classified individuals as having low SES if they had lived in households meeting US Census poverty criteria in at least 1 of 5 previous HRS assessments or if they lacked a high school

education and their household income fell below 200% of the poverty threshold in at least 1 of 5 previous HRS assessments. The authors used their sociogenomic database to test the hypothesis that older adults with low SES would exhibit stronger CTRA-related patterns of gene expression as compared with higher-SES adults.

The authors conducted 2 analyses. In one analysis, they used a traditional, hypothesis-driven approach to test whether low SES was associated with expression levels of 53 CTRA candidate genes. The authors composed a summary score of CTRA gene expression by applying weightings to array-measured expression levels of the candidate genes. Weights of -1 were applied to genes for which the CTRA model hypothesizes down-regulation in response to stress (antiviral-response genes) and weights of $+1$ were applied to genes for which the CTRA model hypothesizes up-regulation in response to stress (proinflammatory genes). Weighted gene expression values were then summed to form a CTRA gene expression score for each study participant. Analysis showed that HRS participants with low SES had higher CTRA expression scores than did HRS participants with higher SES.

In their second analysis, the authors conducted hypothesis-free genome-wide screening of the roughly 35,000 transcripts assayed for association with SES. They selected transcripts that exhibited an at least 1.2-fold difference in expression between the low-SES group and those with higher SES. Differentially expressed transcripts were localized to genes, although this procedure was not described. The analysis identified 141 unique genes as being differentially expressed in the low- versus higher-SES groups, a little less than 1% of the genes assayed. The authors then conducted bioinformatic analysis of the identified genes. They scanned the DNA sequences “upstream” of transcription start sites (called “promoters”) for putative binding sites of 4 CTRA-related transcription factors. Transcription factors are proteins that bind to a DNA sequence and increase or decrease expression of nearby genes (although regulation of more distal genes has also been observed (15)). The goal of transcription factor analysis was to identify genomic mechanisms that might mediate SES effects on gene expression. The set of 141 SES-associated genes was enriched for binding sites of 3 of the 4 transcription factors implicated in CTRA gene regulation, including the proinflammatory nuclear factor κ -light-chain-enhancer of activated B cells (NF- κ B) family of transcription factors. NF- κ B is directly regulated by the glucocorticoid receptor (16), a key regulator of the biological response to stress. Thus, findings suggest that stress-induced proinflammatory gene expression is one mechanism linking socioeconomic disadvantage with increased inflammation and related morbidity.

The work of Levine et al. complements a growing literature identifying dysregulated immune function as a key mediator of social gradients in health. Social disadvantage in childhood and adulthood predict increased systemic inflammation and chronic infection (17–20). The present study by Levine et al., combined with others (21, 22), suggests differential regulation of gene expression as a potential mechanism in this social gradient. These human studies are bolstered by experiments conducted with nonhuman primates (23, 24). Researchers, including one of the authors of this commentary, have conducted experiments to manipulate dominance rank, the primate equivalent of social

status, in captive rhesus monkeys through a procedure of staged introduction to new social groups. This design rules out confounding by genetic or personal history factors. It also allows for tests of reversibility. Similar to the low-SES HRS participants studied by Levine et al., the low-status monkeys showed increased proinflammatory gene expression and evidence for a mediating role of NF- κ B. Critically, just as experimental assignment to low social status induced these changes, experimental reassignment to higher social status reversed them, suggesting potential for interventions with humans to reverse biological embedding of social stress. One difference between human studies and the primate experiments is that the primate experiments identified large fractions of the genome as being differentially expressed across the social gradient (15%–20% compared with the \sim 1% reported by Levine et al.). This finding suggests that larger-scale human studies or, possibly, studies of natural experiments (25) could uncover further genomic correlates of social-stress exposure.

The sociogenomic analysis by Levine et al. highlights both the promise and the challenges facing the field. Below, we briefly discuss 5 issues related to the integration of genomics in general and the CTRA specifically into social epidemiology.

Drinking from the firehose with a straw

A challenge in analysis of genomic data is the very large number of potential hypotheses. Genomic assays typically comprise tens or even hundreds of thousands of measurements. An established approach is hypothesis-free discovery analysis, essentially testing of all hypotheses possible and correcting for the number of tests (26). The study by Levine et al. illustrates an alternative approach. The authors formulated hypotheses and then used a genomic database to conduct targeted tests. In their analysis of CTRA expression score (what Levine et al. call contrast score), they drew measurements of 53 candidate genes identified in previous studies from their whole-genome database and used expression levels to compose a single measure for use in a single hypothesis test. This method mirrors “genetic risk score” analysis commonly applied to DNA sequence data (27). In their bioinformatics analysis, the authors integrated hypothesis-driven and hypothesis-free approaches, first conducting a hypothesis-free search and then feeding results from that search into a hypothesis-driven analysis of a targeted set of transcription factors. Both methods provide models for future studies of whole-genome databases that lack adequate power for hypothesis-free discovery research.

Cell type heterogeneity: Confounder? Phenotype?

Levine et al. analyzed gene expression levels in samples of whole blood, which included a mixture of different cell types (e.g., T cells, neutrophils, and others). Cell type heterogeneity between samples can confound genomic analysis because different cell types have characteristic methylation and expression profiles (28, 29). To isolate effects on expression within cell types, researchers often include statistical covariates for counts or proportions of different cell types within samples. Levine et al. lacked information on cell counts or proportions. Instead, following established practice (30), they used their expression data to impute this information based on reference data sets.

They then included these imputed values as covariates in their analysis. This approach attempts to estimate average differences in expression across the social gradient as if cell type composition of blood samples is the same for everyone.

Cell type heterogeneity is also a plausible phenotype for analysis of biological embedding of social stress. For example, in the primate studies (23, 24), experimentally induced changes in social status caused not only within-cell-population changes in gene expression but also changes in relative proportions of several cell populations. One implication is that SES-related differences in the relative proportions of different cell types could be a phenotype of biological embedding. More consideration of this possibility is warranted. Levine et al. helpfully conducted their analysis with and without statistical adjustment for imputed cell type composition in the blood samples. Estimates were similar in both analyses.

Future genomic analyses of biological embedding must carefully address the issue of cell composition in samples. Ignoring heterogeneity of cell types across samples risks mistaking genomic phenotypes that reflect differences in cell populations for something else. On the other hand, applying methods to exclude cell type heterogeneity across blood samples without first examining it risks throwing out the proverbial baby with the bathwater.

Chronic disease status: a collider in studies of biological embedding of social stress?

Collider bias arises when statistical adjustment is made for a variable that is caused by both exposure and outcome (see Greenland et al. (31)). In the study by Levine et al., social disadvantage (the exposure) and the CTRA (the outcome) are conceptualized as causes of chronic disease. This conceptual model produces the causal diagram $E \rightarrow C \leftarrow D$, where E is the exposure (here, low SES), C is the collider (here, chronic disease), and D is the outcome (here, the CTRA). In this model, the association between SES and the CTRA will be biased by adjustment for chronic disease status (32). As with cell type heterogeneity, Levine et al. helpfully report results with and without adjustment for chronic disease status. Similarity between the 2 sets of results indicates that bias is small in this case. Future studies should nevertheless consider this issue when selecting covariates in models testing associations between social exposures and genomic phenotypes.

Translating the CTRA into a tool for epidemiology

The gene expression CTRA score reported in the article by Levine et al. represents a potentially reproducible index of the CTRA that can be implemented in other studies. The 53-gene set that composes the score aims to track the multiple changes in immune function that are observed in socially stressed humans and animals and which, in turn, predict increased risk for infection and chronic disease (33). As a multimarker measurement of a multidimensional, cumulative process of biological embedding, the CTRA is substantially aligned with allostatic load (34) and the emerging literature on so-called biological aging (35–37). Studies are needed that can compare these different approaches to quantifying biological embedding. Among the questions to answer are: “Do the different

measures measure the same thing?” “Can they be quantified with comparable acuity at different stages in the life course?” and “To what extent can they be modified by either social or clinical intervention?”

Charting a course for sociogenomic epidemiology

Looking beyond the CTRA, large-scale discovery studies are needed to better evaluate the range of genes and gene networks that are differentially regulated across social gradients. In designing these studies, investigators will face tradeoffs between more precise operationalizations of social exposures and sample size. An example from genome-wide association studies may be instructive. In genome-wide association studies to identify genetic correlates of individual differences in human behavior, favoring sample size has paid off. Genome-wide association studies of educational attainment, a very coarse behavioral phenotype, conducted in very large samples has generated replicable discoveries that predict a wide range of human characteristics and outcomes, ranging from early cognitive development to intergenerational socioeconomic mobility (38–40). In parallel, because non-DNA sequence elements of the genome can be changed by the social environment, there may also be a role for carefully conducted experiments. Studies that randomly assign social exposures and conduct repeated assays of genomic phenotypes before and after may offer a complementary path to discovery.

ACKNOWLEDGMENTS

Author affiliations: Department of Population Health Sciences, Department of Medicine, School of Medicine, Duke University, Durham, North Carolina (Daniel W. Belsky); Division of Geriatrics, Department of Medicine, School of Medicine, Duke University, Durham, North Carolina (Daniel W. Belsky); Center for the Study of Aging and Human Development, Duke University, Durham, North Carolina (Daniel W. Belsky, Noah Snyder-Mackler); Duke Population Research Institute, Durham, North Carolina (Daniel W. Belsky); and Department of Evolutionary Anthropology, Duke University, Durham, North Carolina (Noah Snyder-Mackler).

D.W.B. acknowledges support from the National Institute on Aging (grants R01AG032282, P30AG028716, and P30AG034424) and from the Jacobs Foundation. N.S.-M. is supported by the National Institute on Aging (grant K99AG051764).

Conflict of interest: none declared.

REFERENCES

1. Adler NE, Rehkopf DH. US disparities in health: descriptions, causes, and mechanisms. *Annu Rev Public Health*. 2008;29(1): 235–252.
2. Miller GE, Chen E, Parker KJ. Psychological stress in childhood and susceptibility to the chronic diseases of aging: moving toward a model of behavioral and biological mechanisms. *Psychol Bull*. 2011;137(6):959–997.

3. Danese A, McEwen BS. Adverse childhood experiences, allostasis, allostatic load, and age-related disease. *Physiol Behav.* 2012;106(1):29–39.
4. Hertzman C. Putting the concept of biological embedding in historical perspective. *Proc Natl Acad Sci USA.* 2012; 109(suppl 2):17160–17167.
5. Needham BL, Rehkopf D, Adler N, et al. Leukocyte telomere length and mortality in the National Health and Nutrition Examination Survey, 1999–2002. *Epidemiology.* 2015;26(4): 528–535.
6. Geronimus AT, Hicken MT, Pearson JA, et al. Do US black women experience stress-related accelerated biological aging? A novel theory and first population-based test of black-white differences in telomere length. *Hum Nat.* 2010;21(1):19–38.
7. Shalev I, Entringer S, Wadhwa PD, et al. Stress and telomere biology: a lifespan perspective. *Psychoneuroendocrinology.* 2013;38(9):1835–1842.
8. Heijmans BT, Tobi EW, Stein AD, et al. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci USA.* 2008;105(44): 17046–17049.
9. Oberlander TF, Weinberg J, Papsdorf M, et al. Prenatal exposure to maternal depression, neonatal methylation of human glucocorticoid receptor gene (*NR3C1*) and infant cortisol stress responses. *Epigenetics.* 2008;3(2):97–106.
10. Johnstone SE, Baylin SB. Stress and the epigenetic landscape: a link to the pathobiology of human diseases? *Nat Rev Genet.* 2010;11(11):806–812.
11. Levine ME, Crimmins EM, Weir DR, et al. Contemporaneous social environment and the architecture of late-life gene expression profiles. *Am J Epidemiol.* 2017;000(00):000–000.
12. Cole SW. Human social genomics. *PLoS Genet.* 2014;10(8): e1004601.
13. Irwin MR, Cole SW. Reciprocal regulation of the neural and innate immune systems. *Nat Rev Immunol.* 2011;11(9): 625–632.
14. Cole SW. Social regulation of human gene expression: mechanisms and implications for public health. *Am J Public Health.* 2013;103(suppl 1):S84–S92.
15. Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. *Cell.* 2011;144(3):327–339.
16. Almawi WY, Melemedjian OK. Negative regulation of nuclear factor-kappaB activation and function by glucocorticoids. *J Mol Endocrinol.* 2002;28(2):69–78.
17. Danese A, Moffitt TE, Harrington H, et al. Adverse childhood experiences and adult risk factors for age-related disease: depression, inflammation, and clustering of metabolic risk markers. *Arch Pediatr Adolesc Med.* 2009;163(12): 1135–1143.
18. Zajacova A, Dowd JB, Aiello AE. Socioeconomic and race/ethnic patterns in persistent infection burden among US adults. *J Gerontol A Biol Sci Med Sci.* 2009;64(2):272–279.
19. Fagundes CP, Glaser R, Malarkey WB, et al. Childhood adversity and herpesvirus latency in breast cancer survivors. *Health Psychol.* 2013;32(3):337–344.
20. Yang YC, Boen C, Gerken K, et al. Social relationships and physiological determinants of longevity across the human life span. *Proc Natl Acad Sci USA.* 2016;113(3):578–583.
21. Miller GE, Chen E, Fok AK, et al. Low early-life social class leaves a biological residue manifested by decreased glucocorticoid and increased proinflammatory signaling. *Proc Natl Acad Sci USA.* 2009;106(34):14716–14721.
22. Powell ND, Sloan EK, Bailey MT, et al. Social stress up-regulates inflammatory gene expression in the leukocyte transcriptome via β -adrenergic induction of myelopoiesis. *Proc Natl Acad Sci USA.* 2013;110(41):16574–16579.
23. Snyder-Mackler N, Sanz J, Kohn JN, et al. Social status alters immune regulation and response to infection in macaques. *Science.* 2016;354(6315):1041–1045.
24. Tung J, Barreiro LB, Johnson ZP, et al. Social environment is associated with gene regulatory variation in the rhesus macaque immune system. *Proc Natl Acad Sci USA.* 2012; 109(17):6490–6495.
25. Costello EJ, Erkanli A, Copeland W, et al. Association of family income supplements in adolescence with development of psychiatric and substance use disorders in adulthood among an American Indian population. *JAMA.* 2010;303(19):1954–1960.
26. Pearson TA, Manolio TA. How to interpret a genome-wide association study. *JAMA.* 2008;299(11):1335–1344.
27. Belsky DW, Israel S. Integrating genetics and social science: genetic risk scores. *Biodemography Soc Biol.* 2014;60(2):137–155.
28. Watkins NA, Gusnanto A, de Bono B, et al. A HaemAtlas: characterizing gene expression in differentiated human blood cells. *Blood.* 2009;113(19):e1–e9.
29. Reinius LE, Acevedo N, Joerink M, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One.* 2012; 7(7):e41361.
30. Birney E, Smith GD, Grealley JM. Epigenome-wide association studies and the interpretation of disease -omics. *PLoS Genet.* 2016;12(6):e1006105.
31. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology.* 1999;10(1):37–48.
32. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology.* 2003;14(3):300–306.
33. Glaser R, Kiecolt-Glaser JK. Stress-induced immune dysfunction: implications for health. *Nat Rev Immunol.* 2005; 5(3):243–251.
34. Seeman TE, McEwen BS, Rowe JW, et al. Allostatic load as a marker of cumulative biological risk: MacArthur studies of successful aging. *Proc Natl Acad Sci USA.* 2001;98(8): 4770–4775.
35. Levine ME, Crimmins EM. Evidence of accelerated aging among African Americans and its implications for mortality. *Soc Sci Med.* 2014;118:27–32.
36. Simpkin AJ, Hemani G, Suderman M, et al. Prenatal and early life influences on epigenetic age in children: a study of mother-offspring pairs from two cohort studies. *Hum Mol Genet.* 2016; 25(1):191–201.
37. Moffitt TE, Belsky DW, Danese A, et al. The longitudinal study of aging in human young adults: knowledge gaps and research agenda. *J Gerontol A Biol Sci Med Sci.* 2017;72(2): 210–215.
38. Rietveld CA, Medland SE, Derringer J, et al. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science.* 2013;340(6139):1467–1471.
39. Belsky DW, Moffitt TE, Corcoran DL, et al. The genetics of success: how single-nucleotide polymorphisms associated with educational attainment relate to life-course development. *Psychol Sci.* 2016;27(7):957–972.
40. Okbay A, Beauchamp JP, Fontana MA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature.* 2016;533(7604):539–542.