ENVIRONMENTAL-RISK (E-RISK) LONGITUDINAL TWIN STUDY
CONCEPT PAPER FORM

Proposing Author: Aaron Reuben

Author's affiliation, phone, and e-mail: Duke University, 919-684-6684, aaron.reuben@duke.edu

Sponsoring Investigator: Avshalom Caspi and Terrie Moffitt

Proposed co-authors: Karen Sugden, David Corcoran, Joey Prinz, Ben Williams, Line Rasmussen, Eilis Hannon, Jon Mill, Chloe Wong, Helen Fisher, Andrea Danese, Louise Arseneault, Tim Matthews, Candice Odgers

Provisional Paper Title: The epigenetics of neighborhood deprivation

Date: 12/14/18

Objective of the study and its significance:

People living in socioeconomically deprived neighborhoods tend to display poorer self-rated health, worse mental health, more chronic disease, and greater pre-mature mortality than individuals living in less-deprived neighborhoods. These differences are not fully explained by the socioeconomic status of individual residents, or by the selection of individuals with worse health into poorer neighborhoods (Diez Roux & Mair, 2010). Numerous aspects of a neighborhood's physical and social environment could contribute to disparities in long-term health outcomes, including differential access to healthy food, exposure to pollutants, access to natural or recreational spaces, walkability, social cohesion, health behavior norms, physical safety, violence, and quality of housing and shelter (Olden, Olden, & Lin, 2015).

It has recently been proposed that epigenetic alterations that influence the expression of genes (e.g. DNA methylation) could be one additional or mediating mechanism by which neighborhoods influence long-term health trajectories (Olden et al., 2015). A number of studies have already demonstrated that individual-level socioeconomic status may alter DNA methylation patterns (Borghol et al., 2012; Lam et al., 2012; McGuinness et al., 2012). Could community-level socioeconomic factors do the same? An answer to that question would expand our fundamental understanding of human epigenetics, inform preventative disease research, and

potentially lay the groundwork for new polices to improve public health. Few studies have yet investigated the connection between socioeconomic, physical, and social characteristics of neighborhoods and their residents' epigenetic profiles. One study has reported differential methylation of genes related to stress reactivity and inflammation among residents of neighborhoods with varying levels of safety, social cohesion, and socioeconomic disadvantage (Smith et al., 2017), and another has reported higher methylation of a gene related to cancer risk, MEG3, in children raised in disadvantaged neighborhoods (King, Kane, Scarbrough, Hoyo, & Murphy, 2016).

This study seeks to replicate and expand upon these initial, early findings of associations between neighborhood characteristics and DNA methylation profiles, using the E-Risk cohort and high-quality measures of E-Risk twins' neighborhoods' built and social environments. The hypothesis to be tested would be that Study members living in more physically and socially deprived neighborhoods will show differential methylation relative to their peers from less-deprived neighborhoods. Because neighborhood characteristics have been most robustly associated with differential rates of smoking, obesity, stress and inflammation, and schizophrenia, we anticipate that methylation differences will be most pronounced in genes related to these outcomes and health-risk behaviors (Keita et al., 2014; Meijer, Röhl, Bloomfield, & Grittner, 2012; Truong & Ma, 2006; Yen, Michael, & Perdue, 2009).

Statistical analyses:

Our investigation of the associations between neighborhood characteristics and E-risk Study participants' methylation profiles will proceed with two arms: one that is hypothesis-free and one that is hypothesis-driven.

Arm 1 (hypothesis-free):

To examine the relationship between neighborhood characteristics and resident methylation profiles unbiased by prior evidence, we will run an epigenome-wide association study (EWAS) using Generalized Estimating Equations to test the association between participants' neighborhood "Ecorisk" status (primary predictor) across childhood and DNA-methylation status (primary outcome) across the genome (i.e., on all 450k probes on the Illumina array) in young adulthood.

- Statistical models will first adjust for sex, estimated white blood cell type proportions, and principal components (PCs) of measurement error in a baseline model.

- Additional models will then add adjustment for BMI, smoking, and a measure of neighborhood urbanicity.
- All EWAS tests will be corrected for multiple testing using Bonferroni correction and will account for familial clustering in the sample.
- Secondary follow-up tests will look for specificity in the relationship between neighborhood characteristics and DNA methylation rates by using individual components of the Ecorisk variable (secondary predictors: "deprived," "dirty," "dangerous," "disorganized") in the Generalized Estimating Equations.

If we identify CpG probes significantly associated with the Ecorisk variable, we will then:

1. Evaluate the genetic and environmental architecture of these CpG probes using the classical twin design applied to the MZ and DZ participants in the E-Risk Study.
2. Follow up on the CpG probes' functional potential by studying gene expression correlates in an independent sample: The Dunedin Study. We will correlate DNA methylation levels at each of the identified CpG probes with transcriptome-wide gene expression levels and identify significant methylation-gene expression correlations. We will then explore potential downstream biological implications by performing pathway analysis of the genes indexed by these potential co-regulatory relationships.

Arm 2 (hypothesis-driven):

To examine the relationship between neighborhood characteristics and resident methylation profiles in a manner that is informed by prior evidence, we will conduct two separate analytic steps using hypotheses derived from existing literature on neighborhood associations with specific health outcomes.

First, we will run a series of individual Generalized Estimating Equations with a candidate gene approach. Here we will test associations between the Ecorisk variable and probes annotated to genes that have been implicated in stress reactivity and inflammation. We choose probes annotated to these genes because they have been studied in the most detailed report about neighborhood deprivation and DNA methylation (Smith et al., 2017). These tests will follow the same modeling approach pursued in Arm 1, except that multiple testing will be corrected for on a gene-wide basis.

Second, we will run a series of Generalized Estimating Equations with polyepigenetic scores. We will create polyepigenetic scores made up of probes which index the putative DNA

methylation correlates of inflammation, obesity, smoking, and schizophrenia. We choose to create these polyepigenetic scores because their associated phenotypes have been studied in past reports about neighborhood deprivation and health (Broyles et al., 2012; Cohen, Sonderman, Mumma, Signorello, & Blot, 2011; Diez Roux & Mair, 2010; Fleischer et al., 2015; Keita et al., 2014; Krabbendam & van Os, 2005; Morris, Manley, & Ham, 2018; Putrik et al., 2015; Sariaslan et al., 2016; Sheehan, Cantu, Powers, Margerison-Zilko, & Cubbin, 2017; Steptoe & Feldman, 2001; Truong & Ma, 2006). These tests will follow the same modeling approach pursued in Arm 1, with baseline models adjusted for sex and additional models adjusted for child phenotypes relevant to each polyepigenetic score.

Variables Needed at Which Ages (names and labels):

Study: E-risk

Predictors:
ECORISK at all ages – ecorisk variable
Individual indicators of ecorisk at all ages - Deprived, Dirty, Dangerous, Disorganized

Outcomes:
DNA methylation at age 18:
- 450k probes
- Genes implicated in stress reactivity: NR3C1, FKBP5, BDNF, AVP, CRHR1, SLC6A4, OXTR
- Genes implicated in inflammation: CD1D, CCL1, F8, IL8, KLRG1, LTA4H, NLRP12, PYDC1, SLAMF7, TLR1, TLR3.5
- Polyepigenetic scores for: inflammation, obesity, smoking, and schizophrenia

Potentially relevant child phenotypes:
Inflammation:
- CRPmgL18 – CRP at 18

Obesity:
- BMI – BMI at age 18
- OVERWEIGHT – overweight or obese at age 18

Smoking:
- SMKPKYR – smoking, pack years (age12 to 18)

Schizophrenia
- PBF – general psychopathology at age 18
- THGBF – thought disorders at age 18

Neighborhood covariates
- SCLLUrban – indicator of urbanicity of neighborhood at age 12

References cited:

Borghol, N., Suderman, M., McArdle, W., Racine, A., Hallett, M., Pembrey, M., … Szyf, M. (2012). Associations with early-life socio-economic position in adult DNA methylation. *International Journal of Epidemiology*, *41*(1), 62–74. https://doi.org/10.1093/ije/dyr147

Broyles, S. T., Staiano, A. E., Drazba, K. T., Gupta, A. K., Sothern, M., & Katzmarzyk, P. T. (2012). Elevated C-reactive protein in children from risky neighborhoods: evidence for a stress pathway linking neighborhoods and inflammation in children. *PloS One*, *7*(9), e45419. https://doi.org/10.1371/journal.pone.0045419

Cohen, S. S., Sonderman, J. S., Mumma, M. T., Signorello, L. B., & Blot, W. J. (2011). Individual and neighborhood-level socioeconomic characteristics in relation to smoking prevalence among black and white adults in the Southeastern United States: a cross-sectional study. *BMC Public Health*, *11*, 877. https://doi.org/10.1186/1471-2458-11-877

Diez Roux, A. V., & Mair, C. (2010). Neighborhoods and health. *Annals of the New York Academy of Sciences*, *1186*, 125–145. https://doi.org/10.1111/j.1749-6632.2009.05333.x

Fleischer, N. L., Thrasher, J. F., Juárez, B. S. de M., Reynales-Shigematsu, L. M., Arillo-Santillán, E., Osman, A., … Fong, G. T. (2015). Neighbourhood deprivation and smoking and quit behaviour among smokers in Mexico: findings from the ITC Mexico Survey. *Tobacco Control*, *24*(Suppl 3), iii56–iii63. https://doi.org/10.1136/tobaccocontrol-2013-051495

Keita, A. D., Judd, S. E., Howard, V. J., Carson, A. P., Ard, J. D., & Fernandez, J. R. (2014). Associations of neighborhood area level deprivation with the metabolic syndrome and inflammation among middle- and older- age adults. *BMC Public Health*, *14*(1), 1319. https://doi.org/10.1186/1471-2458-14-1319

King, K. E., Kane, J. B., Scarbrough, P., Hoyo, C., & Murphy, S. K. (2016). Neighborhood and family environment of expectant mothers may influence prenatal programming of adult cancer risk: discussion and an illustrative biomarker example. *Biodemography and Social Biology*, *62*(1), 87–104. https://doi.org/10.1080/19485565.2015.1126501

Krabbendam, L., & van Os, J. (2005). Schizophrenia and urbanicity: A major environmental influence—conditional on genetic risk. *Schizophrenia Bulletin*, *31*(4), 795–799. https://doi.org/10.1093/schbul/sbi060

Lam, L. L., Emberly, E., Fraser, H. B., Neumann, S. M., Chen, E., Miller, G. E., & Kobor, M. S. (2012). Factors underlying variable DNA methylation in a human community cohort. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(Suppl 2), 17253–17260. https://doi.org/10.1073/pnas.1121249109

McGuinness, D., McGlynn, L. M., Johnson, P. C. D., MacIntyre, A., Batty, G. D., Burns, H., … Shiels, P. G. (2012). Socio-economic status is associated with epigenetic differences in the pSoBid cohort. *International Journal of Epidemiology*, *41*(1), 151–160. https://doi.org/10.1093/ije/dyr215

Meijer, M., Röhl, J., Bloomfield, K., & Grittner, U. (2012). Do neighborhoods affect individual mortality? A systematic review and meta-analysis of multilevel studies. *Social Science & Medicine*, *74*(8), 1204–1212. https://doi.org/10.1016/j.socscimed.2011.11.034

Morris, T., Manley, D., & Ham, M. V. (2018). Context or composition: How does neighbourhood deprivation impact upon adolescent smoking behaviour? *PLOS ONE*, *13*(2), e0192566. https://doi.org/10.1371/journal.pone.0192566

Olden, K., Olden, H. A., & Lin, Y.-S. (2015). The role of the epigenome in translating neighborhood disadvantage into health disparities. *Current Environmental Health Reports*, *2*(2), 163–170. https://doi.org/10.1007/s40572-015-0048-x

Putrik, P., van Amelsvoort, L., De Vries, N. K., Mujakovic, S., Kunst, A. E., van Oers, H., … Kant, Ij. (2015). Neighborhood environment Is associated with overweight and obesity, particularly in older residents: Results from cross-sectional study in Dutch Municipality. *Journal of Urban Health : Bulletin of the New York Academy of Medicine*, *92*(6), 1038–1051. https://doi.org/10.1007/s11524-015-9991-y

Sariaslan, A., Fazel, S., D'Onofrio, B. M., Långström, N., Larsson, H., Bergen, S. E., … Lichtenstein, P. (2016). Schizophrenia and subsequent neighborhood deprivation: revisiting the social drift hypothesis using population, twin and molecular genetic data. *Translational Psychiatry*, *6*(5), e796. https://doi.org/10.1038/tp.2016.62

Sheehan, C. M., Cantu, P., Powers, D., Margerison-Zilko, C., & Cubbin, C. (2017). Long-term nighborhood poverty trajectories and obesity in a sample of California mothers. *Health & Place*, *46*, 49–57. https://doi.org/10.1016/j.healthplace.2017.04.010

Smith, J. A., Zhao, W., Wang, X., Ratliff, S. M., Mukherjee, B., Kardia, S. L. R., … Needham, B. L. (2017). Neighborhood characteristics influence DNA methylation of genes involved in stress response and inflammation: The Multi-Ethnic Study of Atherosclerosis. *Epigenetics*, *12*(8), 662–673. https://doi.org/10.1080/15592294.2017.1341026

Steptoe, A., & Feldman, P. J. (2001). Neighborhood problems as sources of chronic stress: development of a measure of neighborhood problems, and associations with socioeconomic status and health. *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine*, *23*(3), 177–185. https://doi.org/10.1207/S15324796ABM2303_5

Truong, K. D., & Ma, S. (2006). A systematic review of relations between neighborhoods and mental health. *The Journal of Mental Health Policy and Economics*, *9*(3), 137–154.

Yen, I. H., Michael, Y. L., & Perdue, L. (2009). Neighborhood environment in studies of health of older adults: a systematic review. *American Journal of Preventive Medicine*, *37*(5), 455–463. https://doi.org/10.1016/j.amepre.2009.06.022

Data Security Agreement

Provisional Paper Title	The epigenetics of neighborhood deprivation
Proposing Author	Aaron Reuben
Today's Date	12/14/18

Please keep one copy for your records
(Please initial your agreement)

_ASR___ I am familiar with the King's College London research ethics guidelines (https://www.kcl.ac.uk/innovation/research/support/ethics/about/index.aspx) and the MRC good research practice guidelines (https://www.mrc.ac.uk/research/policies-and-guidance-for-researchers/good-research-practice/).

_ ASR ___ My project has ethical approval from my institution.

_ ASR ___ I am familiar with the EU General Data Protection Regulation (https://mrc.ukri.org/documents/pdf/gdpr-guidance-note-3-consent-in-research-and-confidentiality/), and will use the data in a manner compliant with its requirements.

_ ASR ___ My computer is (a) encrypted at the hard drive level, (b) password-protected, (c) configured to lock after 15 minutes of inactivity, AND (d) has an antivirus client which is updated regularly.

_ ASR ___ I will treat all data as "restricted" and store in a secure fashion.

_ ASR ___ I will not share the data with anyone, including students or other collaborators not specifically listed on this concept paper.

_ ASR ___I will not merge data from different files or sources, except where approval has been given by the PI.

_ ASR ___ I will not post data online or submit the data file to a journal for them to post. Some journals are now requesting the data file as part of the manuscript submission process. The E-Risk Study cannot be shared because the Study Members have not given informed consent for unrestricted open access. Speak to the study PI for strategies for dealing with data sharing requests from Journals.

_ ASR ___ Before submitting my paper to a journal, I will submit my draft manuscript and scripts for data checking, and my draft manuscript for co-author mock review, allowing three weeks.

_ ASR ___ I will submit analysis scripts and new variable documentation to project data manager after the manuscript gets accepted for publication.

_ ASR ___ I will delete the data after the project is complete.

_ ASR ___ **For projects using location data:** I will ensure geographical location information, including postcodes or geographical coordinates for the E-Risk study member's homes or schools, is <u>never</u> combined or stored with any other E-Risk data (family or twin-level data)

_ ASR ___ **For projects using genomic data:** I will only use the SNP and/or 450K data in conjunction with the phenotypes that have been approved for use in this project at the concept paper stage.

Signature: ...../s/ Aaron Reuben.....

CONCEPT PAPER RESPONSE FORM

A. To be completed by the proposing author

Proposing Author: Aaron Reuben

☒ I have read the E-Risk data-sharing policy guidelines and agree to follow them

Provisional Paper Title: The epigenetics of neighborhood deprivation

Potential co-authors: Karen Sugden, David Corcoran, Joey Prinz, Ben Williams, Line Rasmussen, Eilis Hannon, Jon Mill, Chloe Wong, Helen Fisher, Andrea Danese, Louise Arseneault, Tim Matthews, Candice Odgers

Potential Journals: Epigenetics, Environmental Health Perspectives, International Journal of Epigenetics, Social Science & Medicine, Health & Place

Intended Submission Date (month/year): 09/2019

Please keep one copy for your records and return one to Louise (louise.arseneault@kcl.ac.uk)

B. To be completed by potential co-authors:

☐ Approved ☐ Not Approved ☐ Let's discuss, I have concerns

Comments:

Please check your contribution(s) for authorship:

☐ Conceptualizing and designing the longitudinal study

☐ Conceptualizing and collecting one or more variables

☐ Data collection

☐ Conceptualizing and designing this specific paper project

☐ Statistical analyses

☐ Writing

☐ Reviewing manuscript drafts

☐ Final approval before submission for publication

☐ Acknowledgment only, I will not be a co-author

Signature: .....