Concept Paper Template

Provisional Paper Title: What is the significance of epigenetic discoveries for smoking in population health science?

Proposing Author: Karen Sugden/Dan Belsky/Eilis Hannon/Av Caspi

Author's Email: ac115@duke.edu

P.I. Sponsor: Av Caspi
(if the proposing author is a student or colleague of an original PI)

Today's Date: January 17, 2017

Please describe your proposal in 2-3 pages with sufficient detail for helpful review.

What is the significance of epigenetic discoveries for smoking in population health science?

Smoking tobacco is associated with pervasive alterations to the blood epigenome. Epigenome-wide association studies (EWAS) of DNA methylation have identified thousands of CpG sites that are differentially methylated in smokers compared to non-smokers. The significance of these discoveries for population health science remains uncertain.

The proposed paper will take initial EWAS discoveries forward. We will use data from two cohort studies to examine the implications of epigenetic discoveries about smoking for population health science.

We propose to use methylation data in Dunedin (at ages 26 and 38) and E-Risk (at phase 18). In both cohorts, we will use DNA methylation data generated using Illumina 450K arrays. (These data have been previously described; see Marzi et l., 2018.)

The proposed research/report will address four questions.

OUTLINE

The first part of the report would examine the association between exposure to tobacco smoking and DNA methylation.

1. **How are methylation polygenic scores differentially distributed in never-, former-, and current-smokers.**

Drawing on the EWAS report by Joehanes et al. (2016, Circulation Cardiovascular Genetics), we will create a polygenic methylation score (mDNA) which reflects all smoking-associated DNA methylation correlates across the genome. We calculate mDNA values by multiplying the % methylation measured at a CpG site with the effect-size estimated for that CpG in the Joehanes EWAS.

The point here is to document the reproducibility in Dunedin and in E-risk of the associations reported in previous EWAS.

Note: We may need to show EWAS results in Dunedin and in Erisk and to document how these overlap with Joehanes et al. This could be done in a supplement.

2. **Dunedin Study members' methylation polygenic scores tracked upwards with cumulative cigarette consumption through age 38 years.**

This analysis would focus on pack years, (a) to phase 38 years in Dunedin and (b) to age 18 years in E-risk.

The point of this analysis is to document a dose-response association between smoking and methylation.

The second part of the report will test the hypothesis that tobacco smoking has causal effects on DNA methylation, in three ways, (a) by comparing twins who are discordant in their smoking hisotyr, (b) by evaluating changes in DNA methylation over time in people who increase their smoking consumption, and (c) by evaluating changes in DNA methylation over time in people who quit smoking tobacco.

1. **Methylation polygenic score associations with smoking were not confounded by genetics.**

This analysis would be carried out in E-risk.

We need to show the twin-difference correlations, in both DZ and MZ twins, between smoking and DNA methylation**.**

2. **Dunedin Study members who smoked more between methylation measurements at ages 26 and 38 experienced an increase in their methylation polygenic scores.**

3. **Dunedin Study members who quit smoking between methylation measurements at ages 26 and 38 experienced a relative decline in their methylation polygenic scores.**

In addition, we can also extend the last analysis to focus on gene expression.

4. **Gene expression analysis at phase 38 in Dunedin.**

First, we can show that DNA methylation differences identified in epigenome-wide association studies of smoking and gene expression differences identified transcriptome-wide studies of smoking are correlated with one another in both smokers and non-smokers.

Second, we can report gene expression analysis in quitters. This analysis would parallel the quitter analysis of DNA methylation above (although it would be a cross-sectional gene expression analysis at phase 38, rather than focused on change).

The third part of the report will examine health damage associated with DNA methylation. We focus on two health outcomes: lung function (measure as FEV1/FVC) and periodontal disease

(gingival recession).

1. **Does the methylation polygenic score change between 26 and 38 years mediate smoking-associated damage to lungs.**

2. **Does the methylation polygenic score change between 26 and 38 years mediate smoking-associated damage to gums.**

Each analysis will include three steps: we will test (a) the association between change in smoking and change in lung function/recession; (b) the association between change in DNA methylation and change in lung function/recession, and (c) whether change in DNA methylation mediates the effect of change in smoking on chang in lung function/recession.

The fourth analysis evaluates whether cigarette smoking confounds associations between psychosocial risk factors and DNA methylation. Epigenetic modifications are thought to play a role in the biological embedding of psychosocial risks. But many psychosocial risk are also associated with smoking. What are the implications for epigenetic epidemiology and for the study of biological embedding?

1. ***Cigarette smoking confounds associations between psychosocial risk factors and DNA methylation.***

These analyses will be done in both Dunedin and E-risk. We need to narrow down the list of psychosocial risk factors.

The current possibilities are listed below. Note that our parallel measure in both cohorts is the CTQ:

E-risk:
Polyvictimization
CTQ
ACES (not yet constructed)

Dunedin:
ACES
CTQ
Maltreatment

NOTE: we need to decide which to focus on. Also, if we want we can relax the threshold for genome-wide significance, to get more hits. After all, the point here is to document confounding, and we can do that with a more liberal p value, if needed.

NOTE: I am inclined to use polyvictimization in e-risk, as we've published on that. In Dunedin, we can use ACEs. An alternative is to use the CTQ in both cohorts

The analysis would process in three steps: (a) document associations between psychosocial risk and DNA methylation; (b) document associations between psychosocial risk and tobacco smoking; (c) does tobacco smoking confound the associations between psychosocial risk and DNA

methylation.

The third step can involve 3 strategies: (a) we can control for pack years smoking; (b) we can control for mDNA, and (c) we can conduct an enrichment analysis, evaluating overlap between the probes that are associated with psychosocial risk and the probes in the mDNA score.

Data Security Agreement

Provisional Paper Title	**What is the significance of epigenetic discoveries for smoking in population health science?**
Proposing Author	Avshalom Caspi
Today's Date	January 17, 2017

Please keep one copy for your records and return one to the PI Sponsor

Please initial your agreement

X	I am current on Human Subjects Training (CITI (www.citiprogram.org) or equivalent)
X	My project is covered by Duke or Otago ethics committee OR I have /will obtain ethical approval from my home institution.
X	I will treat all data as "restricted" and store in a secure fashion. My computer or laptop is: a) encrypted (recommended programs are FileVault2 for Macs, and Bitlocker for Windows machines) b) password-protected c) configured to lock-out after 15 minutes of inactivity AND d) has an antivirus client installed as well as being patched regularly.
X	I will not "sync" the data to a mobile device.
X	In the event that my laptop with data on it is lost, stolen or hacked, I will immediately contact Professor Moffitt or Caspi. (919-684-6758, tem11@duke.edu, ac115@duke.edu)
X	I will not share the data with anyone, including my students or other collaborators not specifically listed on this concept paper.
X	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 Dunedin Study Members have not given informed consent for unrestricted open access, so we have a managed-access process. Speak to Terrie or Avshalom for strategies for achieving compliance with data-sharing policies of journals.*
X	I will delete all data files from my computer after the project is complete. Collaborators and trainees may not take a data file away from the office. The data remains the property of the Study and cannot be used for further analyses without an approved concept paper for new analyses.

Signature: _____**Avshalom Caspi**_____

CONCEPT PAPER RESPONSE FORM

A

Provisional Paper Title	**What is the significance of epigenetic discoveries for smoking in population health science?**
Proposing Author	Avshalom Caspi
Other Contributors	Karen Sugden, Ben Williams, David Corcoran, Joey Prinz, Temi Moffitt, Renate Houts Jon Mill, Eilis Hannon Richie Poulton, Jonathan Broadbent, Bob Hancox Louise Arseneault, Chloe Wong
Potential Journals	An epidemiology journal
Today's Date	January 17, 2017
Intended Submission Date	May 2018

Please keep one copy for your records and return one to the proposing author

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:

_____