Provisional Paper Title:
Multi-cohort Epigenome-wide Association Study of Aggressive Behavior

Proposing Authors:
From Duke University:
Terrie Moffitt, Karen Sugden, Avshalom Caspi,

From Free University Amsterdam:
Jenny van Dongen, j.van.dongen@vu.nl,
Boomsma, D.I. di.boomsma@vu.nl
Vrije Universiteit, Dept Biological Psychology, Netherlands Twin Register
Van der Boechorststraat 1, 1081 BT, Amsterdam, The Netherlands
Room 2B41
phone: 31-20-5988787 / fax: 31-20-5988832
email: di.boomsma@vu.nl
twitter: @NTRscience
url: www.tweelingenregister.org

Author's Email: tem11@duke.edu,

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

Today's Date: 3 February, 2017

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

Objective of the study:

NOTE: Data analysis for this project will take place at Duke University.

As part of the ACTION project on aggression (http://www.action-euproject.eu/), colleagues at the Free University Amsterdam previously conducted an epigenome-wide association study (EWAS) of aggressive behavior in adults based on whole blood Illumina 450k methylation data collected by the Netherlands Twin Register (NTR). This study found suggestive evidence for associations, with p-values just below the genome-wide significance threshold. They now propose to perform an epigenome-wide association study (EWAS) meta-analysis of aggressive behavior, by recruiting multiple cohorts to take part. They would like to ask each cohort to perform the EWAS analysis on their cohort and to provide the results (summary statistics) for a meta-analysis. In addition to the EWAS approach, a sub-analysis will aim to test the relationship between aggressive behavior and the epigenetic clock (DNA methylation age, DNAmAge). This team hypothesizes that because aggressive behavior is associated with adverse life conditions in general, a higher level of aggressive behavior is associated with accelerated epigenetic ageing. Moreover, because participating cohorts include samples with childhood and adult aggression measures and vary with respect to the moment of DNA collection from just after birth till adulthood, the meta-analysis will be able to examine the association between methylation and aggression across the lifespan.
This analysis plan aims to coordinate the analysis to be performed by the analyst of each cohort,

and to harmonize the output file format (summary statistics). Each cohort team uses the same scripts.

Timeline

Cohorts will send EWAS results to the Amsterdam team in spring 2017. Depending on the meta-analysis results, there is a plan to perform secondary analyses for top hits, including analyses to test if DNA methylation level is associated with gene expression level (RNA-seq and/or Affymetrix U219 array). If cohorts have gene expression data and would like to contribute to the secondary analysis stage, this will be possible.

At Duke, the first thing we propose to do is undertake an EWAS of an appropriate aggression phenotype in the Dunedin Study (by appropriate we mean the best measure, with the best distribution, and one that is age-matched to the blood collection). This will be the MPQ aggression score at Phase 26.

The second thing we propose to do is undertake an EWAS of appropriate an aggression phenotype in the E-risk Study. This will be self-reported conduct problems at Phase 18.

Data analysis methods:

2.1 - EWAS

Model:

For all methylation sites, please run 2 models:

1. Methylation ~ Aggression + Age + Sex + WBC percentages + Technical + Cohort Specific
2. Methylation ~ Aggression + Age + Sex + WBC percentages + Technical + Cohort Specific + BMI + Smoking

Note: If your cohort includes related individuals (e.g. family members, twins), please apply a statistical approach that takes the clustering of data into account (e.g. gee or linear mixed models).

Covariates

Age= Age when DNA sample was collected

WBC= White blood cell percentage in the same blood sample from which DNA was extracted. If you did not measure white blood percentages in the same sample as used for the DNA methylation measurement, please estimate WBC percentages using a prediction method (e.g. Houseman's reference based method). For computational reasons, please do not include multiple WBC that are highly correlated with each other or that show very little variation between people in your cohort. For example, in the NTR, we use the following WBC as covariates: monocyte percentage, eosinophil percentage, neutrophil percentage.

Technical covariates + Cohort Specific covariates Please correct for technical (batch) covariates and other cohort-specific covariates as you deem necessary. For example, at the NTR, we include 450k array row and either sample plate or (carefully chosen) principal

components from the methylation data. If your cohort includes multiple population ancestries, please take this into account as you think is most appropriate for your cohort (possible strategies include exclusion of small groups of 'ethnic outliers', running the analysis separately by ethnicity, inclusion of a covariate denoting ethnicity, or inclusion of principle components based on genotype data).
BMI= Body mass index
Smoking= Smoking status at the moment of blood sampling, 3 levels: 0=never smoked, 1=former smoker, 2=current smoker.

Software and example R-script
Feel free to use your own analysis pipeline or preferred software to run the models outlined above. Example R-code with instructions is provided in:
Example LM script_012016.r [suited for cohorts that include unrelated subjects]
Example gee script_012016.r [suited for cohorts that include related subjects (family members)]

2. Analysis
2.2 – DNAm age acceleration and aggression

Estimating DNAm age
DNAm age acceleration can be easily calculated with the convenient online tool from Steve Horvath: (http://labs.genetics.ucla.edu/horvath/dnamage/). Please follow the instructions on the website.
In short: ***2.2 – DNAm age acceleration and aggression***

Estimating DNAm age
DNAm age acceleration can be easily calculated with the convenient online tool from Steve Horvath: (http://labs.genetics.ucla.edu/horvath/dnamage/). Please follow the instructions on the website.
In short:
* Upload raw (un-normalized) methylation beta-values (you may preselect probes or include all genome-wide data)
* Remove the bad quality samples you also removed from your EWAS
* Include at least the probes listed in "datMiniAnnotiation.csv " (downloadable from the website)
* Upload a sample annotation file, with the column names "SampleID", "Age", "Female", "Tissue"
* Column "Female" should contain 1 for females, and 0 for males.
* Crucial: Methylation file and sample annotation file must be in the same order
* For very large cohorts, it is recommended to upload your data in batches of up to around 1000 samples (and to apply preselection of probes)
* Max file size that can be uploaded is 800 Mb
* Crucial: select the options **Normalize Data** and **Advanced Analysis of Blood**
* Crucial: **do NOT use** the option Fast Imputation

* Hint: First test if your input files are in the correct format by running the calculator without the option normalize data and without the option advanced analysis of blood, which should give you the output very quickly. If you do not receive an email with results within one hour, the format of your input files is not correct.

Variables
Please collect the following variables from the DNAm age calculator output file:
* AgeAccelerationResidual
* AHOAdjCellCounts
* AAHAAdjCellCounts

Model
Please run 3 models:
1. AgeAccelerationResidual ~ Aggression + Sex + BMI + Smoking + Cohort Specific
2. AHOAdjCellCounts ~ Aggression + Sex + BMI + Smoking + Cohort Specific
3. AAHAAdjCellCounts ~ Aggression + Sex + BMI + Smoking + Cohort Specific

Note: If your cohort includes related individuals (e.g. family members, twins), please apply a statistical approach that takes the clustering of data into account (e.g. gee or linear mixed models).

Variables needed at which ages:

Dunedin Study: Genome-wide methylation from Phase 26 blood.
MPQ Aggression at Phase 26.

E-Risk: Dunedin Study: Genome-wide methylation from Phase 18 blood.
Self-reported conduct problems at Phase 18.

Covariates for both cohorts:
Sex,
Age at blood draw
WBC= White blood cell percentage in the same blood sample
BMI
Smoking status at time of blood draw

Significance: A better understanding of the genetics of aggression.

Data Security Agreement

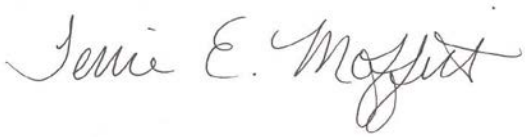
A

Provisional Paper Title	**Multi-cohort Epigenome-wide Association Study of Aggressive Behavior**
Proposing Authors	Moffitt, Sugden, Caspi
Today's Date	3 February 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 IRB OR I have /will obtain IRB approval from my home institution.
x	I will treat all data as "restricted" and store in a secure fashion.
x	I will not share the data with anyone, including 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 cannot be shared because the Study Members have not given informed consent for unrestricted open access. Speak to Terrie or Avshalom for strategies for dealing with data sharing requests from Journals.*
x	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
x	I will submit analysis scripts and new variable documentation to project data manager after manuscript gets accepted for publication.
x	I will return all data files to the Data Manager after the project is complete. Collaborators and graduates of DPPP may not take a data file away from the DPPP office. The data remains the property of the Study and cannot be used for further analyses without express, written permission.



Signature:

CONCEPT PAPER RESPONSE FORM
A

Provisional Paper Title	**Multi-cohort Epigenome-wide Association Study of Aggressive Behavior**
Proposing Authors	Moffitt, Sugden, Caspi from Duke
Other Contributors	Van Dongen, Boomsma from Free University Amsterdam Richie Poulton from Otago Univ., Dunedin Study director. Louise Arseneault from King's College, E-Risk director. Jon Mill from Exeter University, EWAS leader for E-Risk.
Potential Journals	To be decided by the Amsterdam meta-analysis team
Today's Date	3 Feb 2017
Intended Submission Date	During 2017

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

B. To be completed by each potential co-author and returned to the proposing author:

	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:

_____