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

Proposing Author:
Radhika Kandaswamy

Author's affiliation, phone, and e-mail address:
X5271: SGDP Centre, IoPPN, King's College London; radhika.kandaswamy@kcl.ac.uk

Sponsoring Investigator (if the proposing author is a student, a post-doc or a colleague):
Chloe Wong and Helen Fisher

Proposed co-authors:
Louise Arseneault, Avshalom Caspi, Jon Mill, Temi Moffitt, Eilis Hannon, Andrea Danese, Candice Odgers, Ben Williams, Karen Sugden, Joe Burrage, Georgina Mansell

Provisional Paper Title: Epigenetic trajectories of biological response to adolescent psychosocial stress

Date: 15th February 2018

Objective of the study and its significance:

Stress is a normal and adaptive biological and psychological response that has developed throughout evolution to maximise an individual's chances of survival when confronted with a stressor [1]. However, prolonged activation of the stress response system, in particular during key developmental periods such as early childhood development and puberty, not only have acute disruptive impacts on the persons' health and mental development but are also strongly linked with a range of stress-related diseases and cognitive impairments that persist well into adulthood [2]. The main aim of this project is to elucidate how psychosocial stress during adolescence gets under the skin to leave lasting biological imprints. One possibility is that environmental conditions could affect or interact with genes through epigenetic mechanisms, specifically DNA modifications.

Epigenetic processes are dynamic and can fluctuate across the lifespan in response to genetic and environmental influences in particular during key developmental milestones such as early childhood and adolescence [3]. There is an emerging body of evidence showing that individuals exposed to severe psychosocial stressors have different epigenetic fingerprints compared to individuals exposed to no/minimal stressful life events [4, 5]. Limited conclusions can be drawn from previous findings as these studies relied on adult retrospective reports of stress or trauma, unusual clinical groups (suicide victims, institutionalised children), and small samples [6-8]. However, a recent comprehensive epigenome-wide study of polyvictimization across childhood and adolescence using the complete E-Risk dataset reported very few significant associations with DNA methylation at age-18 in peripheral blood, and these were potentially confounded by tobacco smoking and/or did not survive co-twin control tests [9].

The current study proposes a powerful and sensitive design to ascertain a 'purer' impact of psychosocial stress on the epigenome by combining the unique discordant MZ twin design with a longitudinal approach [10, 11]. By comparing the epigenomic fingerprints before and after an MZ twin pair have differing exposures to psychosocial stressors, we can pinpoint more precisely the epigenomic region(s) that is/are influenced by the stressor and rule out those that may be attributable to potential confounders such as genetic variation, age, sex, and shared environmental exposures. We will assess DNA methylation profiles using buccal DNA collected at three different time points, ages 5,10 and 18, at >850,000 CpG sites in the human methylome using the Illumina Infinium EPIC array compared to ~480,000 probes assessed only in age-18 blood samples presented in Marzi et al [9]. In addition, we will compare age-18 DNA methylation from both buccal and blood DNA allowing us to interrogate the tissue-specific nature of the DNA

methylome and the impact of psychosocial stress on multiple tissue sources. Another advantage of this comparative analysis will be in determining whether these 'signatures' can be detected using minimally invasive procedures.

Statistical analyses:

In this study, we selected 168 MZ twin pairs including 89 pairs concordant for no severe stress during childhood but discordant for stress during adolescence, 23 pairs discordant for stress during childhood, 28 pairs concordant for no severe stress during childhood but concordant for exposure to stress during adolescence, and 28 pairs with no exposure to stress during childhood or adolescence. The major stressors considered in the selection criteria during childhood included domestic violence, frequent bullying, physical abuse, sexual abuse, emotional abuse or neglect, and physical neglect; and in adolescence included crime victimisation, maltreatment, neglect, family violence, peer/sibling victimisation, sexual victimisation, and cyber-victimisation.

DNA methylation has been quantified from buccal samples in the E-Risk study at ages 5, 10 and 18 and also from whole blood samples at age-18 using the Illumina Infinium EPIC array, giving quantitative data for >850,000 CpG sites across the genome. A stringent quality control (QC) analysis was performed to check the quality of the data and remove poor quality samples from further analyses. The steps included in the QC investigated multiple aspects including intensity (biggest indicator of sample quality), bisulphite conversion, reported vs predicted sex, genotype match (based on 59 probes on the EPIC array), genetic correlation with the co-twin and other data-points (advantage of having multiple data-points from the same individual and the co-twin), and finally filtering out probes and samples that do not meet the QC threshold. Furthermore, blood cell counts will be quantified for the age-18 blood samples using the estimateCellCounts() function from the minfi package, and the methylation data will be adjusted for cell type composition and batch using linear regression in the relevant analysis. All analyses will be performed for individual CpG sites (controlled for multiple-testing using the False Discovery Rate (FDR)), and for average DNA methylation across specific annotated features (e.g. gene promoters, CpG islands, gene bodies, intergenic regions).

Statistical analyses will be performed using R.

We propose to investigate the association between psychosocial stress exposure during adolescence and DNA methylation and its trajectory by comparing DNA methylomic data in twins discordant for severe stress exposure and those concordant for no severe stress throughout childhood to adolescence. The analyses will be controlled for smoking, given its important role as a confounder in methylation data, particularly in the context of mental health outcomes. Specifically, recorded smoking status will be compared to smoking status derived from measured DNA methylation at known smoking-related CpG sites. Linear regression will be used to adjust the data for the effects of smoking (pack years) and further analyses will be run using the adjusted data.

1. Specifically, we will use a multilevel modelling approach to investigate the effect of stress (exposure) during adolescence in the discordant twins on DNA methylation (outcome) in age-18 buccal DNA compared to ages 5 and 10 buccal DNA. Specifically, the model will be fitted across all the age-18 data points and then a model will be fit across all time points (ages 18, 10 and 5) for the significant probes with an interaction between the stress exposure and time/age.

Furthermore, an additional interaction model will be also be fitted to assess any common or tissue-specific effects at age-18 across multiple tissues (buccal and blood).

2. The longitudinal epigenetic trajectories associated with exposure to adolescent psychosocial stress will be interrogated using a linear mixed model. The range of adolescent stressors assessed within this cohort will allow us to also explore whether exposure to specific types of stress or combinations of stressors have distinct epigenetic signatures. We will also explore the interaction between time and exposure on the DNA methylomic profile of twins discordant for stress at different time points and twins concordant for no exposure to stress and compare the methylomic profiles/trajectories using the interaction model. Specifically, we will investigate the effect of the age variable and the interaction between age and exposure allowing us to see if there

is a relationship between outcome and age at exposure.

3. To explore the functional organisation of DNA methylation and how this is altered with early psychosocial stress exposure, we will employ weighted gene co-methylation network analysis to identify modules of co-methylated features via unsupervised hierarchical clustering on the basis of high topological overlap [12].

Variables Needed at Which Ages (names and labels):

Study:
FAMILYID (ID Family)
ATWINID (ID Twin 1)
BTWINID (ID Twin 2)
SAMPSEX (Sex of twins)
ZYGOSITY (Zygosity of twins)
RORDERP5 (Random order variable)
SESWQ35 (Social class composite)
SETHNIC Ethnicity of Twins

Age 12:
EX_SVE12 Exposed to severe victimization (0/1), 5-12, E-Twin
EX_SVY12 Exposed to severe victimization (0/1), 5-12, Y-Twin

Age 18:
SMKPKYRE18 (Smoking - pack years, ages 12 to 18)
SMKCNUME18 (Smoking - current number of cigarettes - *Number of cigarettes smoked per day at age 18 or age 19 smoking level if that age at interview)*

Illumina EPIC DNA methylation data from buccal DNA at ages 5,10,and 18 + related variables (probes, batch number, chipID etc) for both elder and younger twin
Illumina EPIC DNA methylation data from peripheral blood at age-18 + related variables (probes, batch number, chipID etc) for both elder and younger twin

POLYVCTZCE18 Poly-victimisation 4 cat (0,1,2,3+) - P18 - Elder
POLYVCTZCY18 Poly-victimisation 4 cat (0,1,2,3+) - P18 - Younger

VCTZCONE18 Conventional victimisation severity - P18 - Elder
VCTZCONY18 Conventional victimisation severity - P18 - Younger
VCTZMALE18 Maltreatment victimisation severity - P18 - Elder
VCTZMALY18 Maltreatment victimisation severity - P18 - Younger
VCTZPERE18 Peer victimisation severity - P18 - Elder
VCTZPERY18 Peer victimisation severity - P18 - Younger
VCTZSEXE18 Sexual victimisation severity - P18 - Elder
VCTZSEXY18 Sexual victimisation severity - P18 - Younger
VCTZFAME18 Family victimisation severity - P18 - Elder
VCTZFAMY18 Family victimisation severity - P18 - Younger
VCTZINTE18 Internet victimisation severity - P18 - Elder
VCTZINTY18 Internet victimisation severity - P18 - Younger
VCTZNEGE18 Neglect victimisation severity - P18 - Elder
VCTZNEGY18 Neglect victimisation severity - P18 – Younger

CTQCTOTE18 CTQ combined - types of abuse or neglect at mod/severe level (0-5) - P18 - Elder
CTQCTOTY18 CTQ combined - types of abuse or neglect at mod/severe level (0-5) - P18 - Younger

References cited:
1. Gunnar, M. and K. Quevedo, The neurobiology of stress and development. Annu Rev Psychol, 2007. 58: p. 145-73.
2. Shonkoff, J.P., et al., The lifelong effects of early childhood adversity and toxic stress. Pediatrics, 2012. 129(1): p. e232-46.
3. Dolinoy, D.C., J.R. Weidman, and R.L. Jirtle, Epigenetic gene regulation: linking early developmental environment to adult disease. Reprod Toxicol, 2007. 23(3): p. 297-307.
4. McGowan, P.O., et al., Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. Nat Neurosci, 2009. 12(3): p. 342-8.
5. Uddin, M., et al., Epigenetic and immune function profiles associated with posttraumatic stress disorder. Proc Natl Acad Sci U S A, 2010. 107(20): p. 9470-5.
6. Labonte, B., et al., Genome-wide epigenetic regulation by early-life trauma. Archives of General Psychiatry, 2012. 69: p. 722-731.
7. Naumova, O.Y., et al., Differential patterns of whole-genome DNA methylation in institutionalized children and children raised by their biological parents. Developmental Psychopathology, 2012. 24: p.143-155.
8. Mehta, D., et al., Childhood maltreatment is associated with distinct genomic and epigenetic profiles in posttraumatic stress disorder. Proc Natl Acad Sci U S A, 2013. 110(20): p. 8302-7.
9. Marzi et al., Analysis of DNA Methylation in Young People: Limited Evidence for an Association Between Victimization Stress and Epigenetic Variation in Blood Am J Psychiatry, 2018 [Epub ahead of print]
10. Bell, J.T. and T.D. Spector, A twin approach to unraveling epigenetics. Trends Genet, 2011. 27(3): p.116-25.
11. Rakyan, V.K., et al., Epigenome-wide association studies for common human diseases. Nat Rev Genet, 2011. 12(8): p. 529-41.
12. Zhang, B. and S. Horvath, A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol, 2005. 4: p. Article17

Data Security Agreement

Provisional Paper Title	Epigenetic trajectories of biological response to adolescent psychosocial stress
Proposing Author	Radhika Kandaswamy
Today's Date	15.02.2018

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

RK 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/)

RK My project has ethical approval from my institution.

RK 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.

RK I will treat all data as "restricted" and store in a secure fashion.

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

RK I will not merge data from different files or sources, except where explicit approval has been given by the PI.

RK 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.

RK 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.

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

RK **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 never combined or stored with any other E-Risk data (family or twin-level data)

RK **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: ..Radhika Kandaswamy.....