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

Proposing Author:
Ryan Arathimos

Author's affiliation, phone, and e-mail address:

Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, SE5 8AF, UK
Tel: +44 (0)20 7848 0873 ryan.arathimos@kcl.ac.uk

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

Proposed co-authors:
Helen Fisher, Louise Arseneault, Avshalom Caspi, Jon Mill, Temi Moffitt, Andrea Danese, Ben Williams, Karen Sugden, Cathryn Lewis

Provisional Paper Title: *Methylome-wide association study of major depressive disorder (MDD), anxiety disorder and internalising factors*

Date: 25/06/21

Objective of the study and its significance:

Major Depressive Disorder (MDD) is a heterogeneous psychiatric disorder with a polygenic architecture (McIntosh et al., 2019). Recent genome-wide association studies (GWAS) have uncovered a large number of genetic risk variants, with polygenic risk scores (PRS) explaining 1.5-3.2% of the variation in MDD (Howard et al., 2019). However, PRS may not capture larger environmental contributions to MDD risk. DNA methylation has become increasingly important in the prediction of MDD, with methylation risk scores explaining ~1.75% of the variance in MDD, acting additively to PRS (Barbu et al., 2020). Similarly to MDD, GWAS of anxiety disorder have identified several risk variants, which however explain only part of the heritability (Levey et al., 2020). Recent methylome-wide association studies (MWAS) of depressive (Davies et al 2014, Jovanova et al., 2018; Starnawska et al., 2019) and severe anxiety symptoms (Emeny et al 2017) have identified a number of methylome-wide and suggestive cytosine-phosphate-guanine (CpG) sites, with several that are annotated to genes participating in brain-related phenotypes. However, additional studies are needed to replicate and extend these findings in MDD and anxiety disorders.

To address this, we propose to use a dual analytical approach, to conduct a cross-sectional methylome-wide association study of depression and of anxiety disorder, using DNA methylation ascertained in blood at age 18 from the Illumina 450K array, as well as a paired twin-discordant analysis (if we have > 100 discordant twin pairs).

Given the well documented substantial overlap and comorbidity between MDD and anxiety disorder (Mineka et al., 1998), we also plan to investigate the potential relationship between differential DNA methylation and an "internalising score'. This previously calculated dimensional score (constructed using

a 3-factor confirmatory factor analysis, Schaefer et al. 2018) will enable an investigation of the overlap between MDD and anxiety disorder (and some related disorders). Several conditions, including major depressive episode, generalised anxiety disorder, PTSD, eating pathology (i.e., eating disorders) loaded on the internalizing factor when it was calculated in the E-Risk sample.

As a secondary aim, we plan to contribute summary results from the MWAS of MDD to the Psychiatric Genomics Consortium's (PGC) ongoing meta-analysis of MDD MWAS. The PGC meta-analysis will leverage summary level data from individual MWAS in contributing cohorts to identify CpG sites associated with MDD (as a binary phenotype) in the largest MDD MWAS meta-analysis conducted to date. Eligible studies must have methylation from blood samples from either the Illumina 450k or EPIC arrays and have some measure of broad depression (self-reported diagnosis, electronic health records, symptoms questionnaire etc). Planned follow-up analyses of the PGC MDD meta-analysis will include Mendelian randomization and calculation of methylation-based risk scores for prediction.

Statistical analyses:

Current depression status (binary yes/no) will be based on participants who meet the criteria for a major depressive episode according to DSM4 (previously derived), using reports at age 18. Generalised anxiety disorder (also binary) will likewise be based on GAD DSM4 status, as reported at age 18.
Separate MWAS for each mental health measure will be conducted in R using a process based on the package limma and methylation M-values as the outcome. We will use a stepwise approach to covariate adjustment. We will adjust for multiple testing using both Bonferroni adjustment, as a conservative threshold, and the false discovery rate (FDR), as a threshold for more suggestive associations.

Several lifestyle factors have been shown to have large effects on DNA methylation and may therefore confound associations between mental health measures and DNA methylation. The most well studied of these factors, considered to have the largest effect on DNA methylation is smoking status (Zeilinger et al. 2013). Additionally, alcohol consumption (Liu et al 2018) and BMI (Dick et al. 2014) have also been shown to have substantial effects. To assess the degree to which these factors may confound associations we will perform stepwise adjustments for these as covariates in each MWAS. In the base model, we will include sex as a covariate, as well as adjustments for estimated cell-counts, technical batch and any cohort-related technical variables. We will also use the genetic data available for the subsample of individuals with DNA methylation data to calculate genetic principal components (PCs) that will capture ancestry and population structure in the sample, which we will then adjust for as covariates. We will also adjust for family-relatedness as a random effect given the relatedness structure of the sample. In the second model, we will adjust for BMI, alcohol consumption (drinks per week) and smoking (using either self-reported smoking or if missing in too many subjects, then AHRR CpG methylation as a proxy) additionally.
As an example, for MDD, the simplified formula for fully adjusted model will be;

DNA methylation ~ MDD + sex + BMI + alcohol consumption + smoking status + PCs + cell types, cluster = FamilyID

Separate MWAS models with then be run for GAD and internalizing dimension score.

Post-MWAS analyses will include pathway analysis (GO and KEGG pathway enrichment) and annotation of any CpGs that survive multiple testing adjustment, to genes/genomic regions. We will also conduct a differentially methylated region (DMR) analysis using dmrff (Suderman et al. 2018) to search for loci where methylation is associated with each trait across CpGs, even if no individual site is associated with the trait (following adjustment for multiple testing).

For the twin-discordant analysis, where there are 266 discordant twin pairs for MDD and 113 discordant twin pairs for GAD, the simplified formula for fully adjusted twin discordant model will be;

DNA methylation ~ discordant status (affected/unaffected) + sex + BMI + alcohol consumption + smoking status + PCs + cell types, cluster = FamilyID

Given that the internalising factor is a continuous score (unlike MDD and GAD which is dichotomous), we will select discordant twin pairs based on the twin difference score on this measure.

Post-MWAS analyses will include pathway analysis (GO and KEGG pathway enrichment) and annotation of any CpGs that survive multiple testing adjustment, to genes/genomic regions. We will also conduct a differentially methylated region (DMR) analysis using dmrff (Suderman et al. 2018) to search for loci where methylation is associated with each trait across CpGs, even if no individual site is associated with the trait (following adjustment for multiple testing).

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)

Age 18:
Covariates
BMIE18 (BMI - P18 – Elder)
BMIY18 (BMI - P18 – Younger)

ALCVOLE18 (Alcohol - num of drinks per week (past year) - P18 – Elder)
ALCVOLY18 (Alcohol - num of drinks per week (past year) - P18 – Younger)

TAGEE18 (Age at Interview - P18 – Elder (and Younger))
TAGEGE18 (Age at Interview (Grouped) - P18 – Elder (and Younger))

SMKPKYRE18 (Smoking - pack years, ages 12 to 18)
SMKCNUME18 (Smoking - current number of cigarettes)

Methylation data
Illumina EPIC DNA methylation data from peripheral blood at age-18 + related variables (probes, batch number, methylation array control probe principal components, chipID etc) for both elder and younger twin.

Genetic data
Genetic data (directly genotyped) from all twins with DNA methylation data, or genetic principal components (PCs) calculated on that subset.

Internalizing factor
INTCF_E Internalizing 3-factor, age 18

Depression
DXMDEE18 Major depressive episode, dsm4 - P18 - Elder

DXMDEY18 Major depressive episode, dsm4 - P18 - Younger

Anxiety

DXGADE18 Gen Anxiety Disorder, dsm4_based - P18 - Elder
DXGADY18 Gen Anxiety Disorder, dsm4_based - P18 - Younger

References cited:

Barbu, M.C., Shen, X., Walker, R.M. et al. Epigenetic prediction of major depressive disorder. Mol Psychiatry (2020). https://doi.org/10.1038/s41380-020-0808-3

Davies, M.N., Krause, L., Bell, J.T. et al. Hypermethylation in the ZBTB20 gene is associated with major depressive disorder. Genome Biol 15, R56 (2014). https://doi.org/10.1186/gb-2014-15-4-r56

Dick KJ, Nelson CP, Tsaprouni L, Sandling JK, Aïssi D, Wahl S, Meduri E, Morange PE, Gagnon F, Grallert H, Waldenberger M. DNA methylation and body-mass index: a genome-wide analysis. The Lancet. 2014 Jun 7;383(9933):1990-8.

Emeny RT, Baumert J, Zannas AS, Kunze S, Wahl S, Iurato S, Arloth J, Erhardt A, Balsevich G, Schmidt MV, Weber P, Kretschmer A, Pfeiffer L, Kruse J, Strauch K, Roden M, Herder C, Koenig W, Gieger C, Waldenberger M, Peters A, Binder EB, Ladwig KH. Anxiety Associated Increased CpG Methylation in the Promoter of Asb1: A Translational Approach Evidenced by Epidemiological and Clinical Studies and a Murine Model. Neuropsychopharmacology. 2018 Jan;43(2):342-353. doi: 10.1038/npp.2017.102. Epub 2017 May 25. PMID: 28540928; PMCID: PMC5729551.

Howard, D.M., Adams, M.J., Clarke, TK. et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat Neurosci 22, 343–352 (2019). https://doi.org/10.1038/s41593-018-0326-7

Jovanova O, Nedeljkovic I, Spieler D, Walker RM, Liu C, Luciano M, Bressler J, Brody J, Drake AJ, Evans KL, Gondalia R, Kunze S, Kuhnel B, Lahti J, Lemaitre RN, Marioni RE, Swenson B, Himali JJ, Wu H, Li Y, McRae AF, Russ TC, Stewart J, Wang Z, Zhang G, Ladwig KH, Uitterlinden AG, Guo X, Peters A, Räikkönen K, Starr JM, Waldenberger M, Wray NR, Whitsel EA, Sotoodehnia N, Seshadri S, Porteous DJ, van Meurs J, Mosley TH, McIntosh AM, Mendelson MM, Levy D, Hou L, Eriksson JG, Fornage M, Deary IJ, Baccarelli A, Tiemeier H, Amin N. DNA Methylation Signatures of Depressive Symptoms in Middle-aged and Elderly Persons: Meta-analysis of Multiethnic Epigenome-wide Studies. JAMA Psychiatry. 2018 Sep 1;75(9):949-959. doi: 10.1001/jamapsychiatry.2018.1725. Erratum in: JAMA Psychiatry. 2018 Nov 1;75(11):1206. PMID: 29998287

Levey DF, Gelernter J, Polimanti R, Zhou H, Cheng Z, Aslan M, Quaden R, Concato J, Radhakrishnan K, Bryois J, Sullivan PF. Reproducible genetic risk loci for anxiety: results from∼ 200,000 participants in the Million Veteran Program. American Journal of Psychiatry. 2020 Mar 1;177(3):223-32.

Liu C, Marioni RE, Hedman ÅK, Pfeiffer L, Tsai PC, Reynolds LM, Just AC, Duan Q, Boer CG, Tanaka T, Elks CE. A DNA methylation biomarker of alcohol consumption. Molecular psychiatry. 2018 Feb;23(2):422-33.

McIntosh AM, Sullivan PF, Lewis CM. Uncovering the Genetic Architecture of Major Depression. Neuron. 2019 Apr 3;102(1):91-103. doi: 10.1016/j.neuron.2019.03.022. PMID: 30946830

Mineka S, Watson D, Clark LA. Comorbidity of anxiety and unipolar mood disorders. Annual review of psychology. 1998 Feb;49(1):377-412.

Schaefer JD, Moffitt TE, Arseneault L, Danese A, Fisher HL, Houts R, Sheridan MA, Wertz J, Caspi A. Adolescent victimization and early-adult psychopathology: approaching causal inference using a longitudinal twin study to rule out noncausal explanations. Clinical Psychological Science. 2018

May;6(3):352-71.

Starnawska A, Tan Q, Soerensen M, McGue M, Mors O, Børglum AD, Christensen K, Nyegaard M, Christiansen L. Epigenome-wide association study of depression symptomatology in elderly monozygotic twins. Transl Psychiatry. 2019 Sep 2;9(1):214. doi: 10.1038/s41398-019-0548-9. PMID: 31477683

Suderman M, Staley JR, French R, Arathimos R, Simpkin A, Tilling K. dmrff: identifying differentially methylated regions efficiently with power and control. BioRxiv. 2018 Jan 1:508556.

Zeilinger S, Kühnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, Weidinger S, Lattka E, Adamski J, Peters A, Strauch K. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. PloS one. 2013 May 17;8(5):e63812.

Data Security Agreement

Provisional Paper Title	Methylome-wide association study of major depressive disorder (MDD) and anxiety disorder
Proposing Author	Ryan Arathimos & Chloe Wong
Today's Date	20/07/21

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

_RA_ 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/).

_RA_ My project has ethical approval from my institution.

_RA_ 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.

_RA_ 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.

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

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

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

_RA_ 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.

_RA_ 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.

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

_RA_ I will delete the data after the project is complete.

_ - _ **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)

_RA_ **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:  ....

CONCEPT PAPER RESPONSE FORM

A. To be completed by the proposing author

Proposing Author:

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

Provisional Paper Title: Methylome-wide association study of major depressive disorder (MDD) and anxiety disorder

Potential co-authors: Helen Fisher, Louise Arseneault, Avshalom Caspi, Jon Mill, Temi Moffitt, Andrea Danese, Ben Williams, Karen Sugden, Cathryn Lewis

Potential Journals:

Intended Submission Date (month/year): May 2022

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: .....