Lifecourse GWAS group – Principles of collaboration

This document outlines the principles of collaboration for the Lifecourse GWAS to which all consortium members should abide.

Version 1 – July 2024

1. Goals, aims and objectives

Genome-wide association studies have considered a wide range of traits and have identified many loci associated with those traits. Many GWAS now include hundreds of thousands (or millions) of individuals and identify large numbers of loci associated with the trait under consideration. However, these GWAS will usually group individuals across a, potentially wide, range of ages. This approach maximises the power to identify relevant loci but imposes the assumption that the per-allele effect of a SNP on the trait of interest is constant across all ages. However recently there has been growing interest in whether there may be age specific effects of some genetic variants on some traits. Additionally Mendelian Randomization methods have been extended recently to consider the estimation and interpretation of age varying effects, where either or both of the genetic variant – exposure trait effect of the exposure – outcome effect may vary across an individual's lifetime.

The goal of the lifecourse GWAS group is to estimate age specific associations between genetic variation and a set of traits across as wide a range of ages as possible. Individuals included in the study will be grouped based on age at the time of phenotype measurement at seven predefined life stage categories. We will also perform a more granular analysis grouping based on age, at one year intervals up to age 20 years and then at 5 year intervals for the rest of the lifecourse. By considering each age group separately we are avoiding imposing any structure on the time varying nature of the effect of genetic variation.

Summary statistics from the meta-analysis of each age group / trait combination analysed will be made freely available online.

2. Structure and Membership

The structure of the group will include an Executive committee, a core group of researchers and an extended group of researchers.

a. *Executive committee.* The executive committee membership is still to be confirmed. This group will provide high level oversight on the direction and management of the project. They will meet with the core group every 6 months to discuss the project and will additionally provide advice to the core group on an adhoc basis where needed.

b. *Core group.* The core group includes Prof. Gibran Hemani, Dr. Eleanor Sanderson, Dr Grace Power, Dr Genevieve Leyden and Dr David Carslake. This group will have

primary responsibility for developing and implementing the project. This will include developing the analysis pipelines, analysis of data available within the MRC Integrative Epidemiology Unit, and subsequent analysis of the summary statistics generated. This group will meet weekly/fortnightly depending on the stage of the project.

c. *Extended group.* The extended group includes the PI's, researchers and analysts from studies contributing data. The individuals in this group will be responsible for running the pipelines on their study data and for ensuring effective communication with the core group should any issues be identified. The core group will arrange meetings with the extended group at key points in the project and will maintain communication with individual researchers/studies as needed throughout the project.

3. Data sharing and confidentiality

It is assumed that any data contributed to the Lifecourse GWAS group activities carries appropriate ethical approval and consent. We can provide support to gain additional cohort approval. Your Indication of Agreement to this code of conduct (see below) will be taken as confirmation of this.

We expect most collaborations in the Lifecourse GWAS consortium will follow a conventional GWAS consortia model whereby all contributing groups perform specific analyses on their own data and return summary statistics or partial derivatives to enable replication and meta--analyses. This format overcomes the need to share individual level or private data, which may be prohibited due to consent issues, ethical review committee ban or by national law, for example. The summary statistics will be analysed on secure servers at University of Bristol or other secure servers approved by the Lifecourse GWAS consortium.

By participating in the consortium (i.e. uploading results from our pipeline for meta-analysis) you are agreeing to our open science ethos. This means that:

1. Cohort summary statistics provided will be used for meta-analyses in age specific groups as defined above.
2. Meta-analysis results will be made publicly available. **Note; some ages may contain data from a single cohort/study and in this case the results made available will be the summary statistics for that cohort/study for that age.**

All Lifecourse GWAS consortium members must be committed to protect the confidentiality of results and joint research activities. For instance, data and results should not be shared outside of the working group without prior permission; results from any downstream replication or functional experiments of loci identified in a meta--analysis should not be published in advance of the agreed--upon primary meta-analysis publication; equally, any finding or conclusion arising from any aspect of the Lifecourse GWAS consortium should be acknowledged as doing so.

Finally, you agree to not use the data or methods to make claims about racial superiority. You agree to strictly adhere to the American Society of Human Genetics (ASHG) position statement, "ASHG Denounces Attempts to Link Genetics and Racial Supremacy" and the International Genetic Epidemiology Society "Statement on Racism and Genetic Epidemiology". You agree to adhere to the principles articulated in the final two sections articulated by the ASHG position on "Advancing Diverse Participation in Research with Special Consideration for Vulnerable Populations", namely, "In the Conduct of Research with Vulnerable Populations, Researchers Must Address Concerns that Participation May Lead to Group Harm" and "The Benefits of Research Participation Are Profound, Yet the Potential Danger that Unethical Application of Genetics Might Stigmatize, Discriminate against, or Persecute Vulnerable Populations Persists."

4. Analysis plans and data generation

For the primary Lifecourse GWAS consortium project, members will implement a conventional GWAS consortium structure in which all collaborating groups run a standardised analysis pipeline and feed back summary statistics to the Core group for meta--analysis. Our aim is to provide a user-friendly pipeline so that minimal coding is required by analysts, which hopefully reduces burden on cohort analysts, avoids duplication of efforts, increases reproducibility and reduces potential for error.

It is the responsibility of the Core group to provide a detailed analysis plan or github wiki page alongside the pipeline to individual studies and analysts. This will include methods for pre--processing and statistical analysis as well as the data structure for the full dataset required to implement the age specific analyses.

5. Collaboration within the Lifecourse GWAS consortium

Our aim is to build an infrastructure of research activities around lifecourse genetic epidemiology, and we strongly encourage consortium members to propose, establish and lead additional projects. These projects can be for example new analytical approaches on the data already being generated, or the proposal of additional phenotypic areas for analysis.

New or existing members wishing to propose a new project should do so by completing the project proposal form available on the website (TBA) and submitting the proposal to the core research group. Proposals will undergo a basic evaluation by the executive committee and core research group for practical feasibility and scientific relevance. All accepted project proposals will be published on the website for transparency. For all new project proposals approved, the project lead will be responsible for establishing a dedicated working group and for executing the study, with subsequent papers appropriately attributing the Lifecourse GWAS consortium.

6. Publication and dissemination of results

The primary output from the Lifecourse GWAS consortium will be a paper summarizing the meta-analysed GWAS results from the initial set of traits considered. Authorship for this paper will be determined as outlined below including all studies contributing data to any of the traits included in that paper. Any follow up papers/secondary papers that publish novel GWAS results from the Lifecourse GWAS consortium (for example; a paper focusing on additional traits not included initially or resulting from a collaboration proposal) will apply the same authorship policy including all studies that have contributed novel summary statistics **included in that paper and not previously published elsewhere by the consortium**.

Defining authorship

For each paper, PIs determine the authors from their group using the widely accepted criteria for authorship on scientific papers (ICMJE criteria). These criteria are:

The ICMJE recommends that authorship be based on the following 4 criteria:

1. Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND
2. Drafting the work or revising it critically for important intellectual content; AND
3. Final approval of the version to be published; AND
4. Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Note that criterion (1) includes those responsible for study collection. Note that criteria (2,3,4) are also required for authorship. The number of authors for each cohort should be reasonable and not excessive. There should be no gratuitous authorship (e.g., a head of department who was not directly involved in the study).

Authorship for main papers from the Lifecourse GWAS consortium

Each study contributing GWAS data can name up to **4** authors who satisfy the criteria defined above (including 1 analyst and 1 PI). This can of course be increased if there is a clear case for including extra authors for a particular study.

Authors will be ordered based on contributions, which will be determined by assigning authors to each of the following tiers:

i. Central analysis and writing group - analysis
ii. Study-specific lead analyst
iii. Study-specific sample collection and data generation (phenotyping/genotyping/supporting analysis)
iv. Study PIs + study design
v. Central analysis and writing group - supervisory

The author list concludes with “Lifecourse GWAS consortium”. This term has no attached author or affiliation.

The studies are ordered according to their total contributing sample size across all age groups, from largest to smallest (tier ii) or smallest to largest (tiers iii, iv). Where there are multiple authors from one study in the same tier, authors are put in alphabetical order.

Authorship for papers resulting from collaboration proposals which involve novel GWAS analysis

For papers that arise from collaboration proposals, for example to consider a particular set of traits not considered initially, a working group will be generated for each project. The working groups will include relevant individuals from the core working group in addition to individuals from the group proposing the collaboration and any other individuals contributing to the overall analysis and writeup of the project. The authorship of the resulting paper will then take the same form as for the main paper with the project working group replacing the central analysis and writing group for i and v. All studies contributing novel GWAS summary statistics for at least one phenotype at one age point will be eligible for inclusion as authors.

Studies using only pre-released summary statistics

These are studies arising from external proposals. These are studies that can be led by external PIs (e.g. not a member of the Lifecourse GWAS group). These data consist of summary stats generated from the age specific GWAS meta-analyses including SNPs with alleles, betas, standard errors, p.values, N, info, MAF etc. They contain no individual or identifying data. These data will be publicly available after publication of the main paper (or secondary paper for additional GWAS analysis). For papers arising from external proposals that use pre-released summary statistics we require that "the Lifecourse GWAS consortium" is included in the authorship list. After release of these data, studies should cite the relevant Lifecourse GWAS paper and no authorship is required.

Author responsibilities

Large consortia have typically a large number of authors. For maximal progress, it is imperative that authors are highly responsive and responsible. They need to all queries promptly and provide accurate and up-to-date information. This is particularly the case for Nature family journals where even trivial author list changes after initial submission require explicit approval by all other authors.

Therefore, for the efficient management of the Lifecourse GWAs consortium the following steps are necessary

1. Lifecourse GWAS PIs will respond promptly to all requests from a writing team (within a few days).
2. Lifecourse GWAS PIs will promptly supply lists of authors from their studies upon request

3. Lifecourse GWAS PIs will ensure that authors meet authorship criteria (see above)
4. Lifecourse GWAS PIs will be responsible for ensuring that all names, initials, degrees, affiliation and addresses are complete, accurate, and consistent
5. For each Lifecourse GWAS paper, there will be an “author freeze” date. On this date, the author and affiliation list will be deemed final –no additions or changes after this date. The Lifecourse GWAS group will not request changes to any paper to correct errors that are the responsibility of a PI.

Software development for consortium analyses

We will recognise contributions to GWAS pipelines and post-GWAS analyses through GitHub contributions. Contributions of scripts or code to these initiatives will be under the licenses chosen by the core research group / executive committee and will most likely be MIT or GPLv3 licenses. We encourage project leads of additional proposals to also make software publicly available through open-source licenses.