GoDMC Code of Conduct version 2 August 2023

1. Goals, aims and objectives

Epigenetic epidemiology is now a well--established field of research offering a view on new strategies to predict, prevent and treat a wide spectrum of common complex diseases. As with classic genetic association studies, the rapid development of this field was aided by technological advances including Illumina's DNA methylation Bead Arrays. Numerous studies have been published linking epigenetic modifications, principally DNA methylation, to a range of genetic and environmental exposures (e.g. smoking, alcohol, nutrient) as well as health outcomes (e.g. cancer, diabetes, psychosis). In particular genetic studies of DNA methylation have proven to be a strong foundation for any research question in the field from insights in the processes regulating DNA methylation to causal inference. Therefore, the Genetics of DNA Methylation Consortium (GoDMC) was established: the largest and concerted effort to build a detailed catalog of genetic effects on DNA methylation to serve the field of epigenetic epidemiology. We used this catalog to study whether DNAm differences in blood are causally related to disease. We performed a bidirectional mendelian randomization study between DNAm and 116 traits and we found that DNAm and traits often share genetic factors but are often not on the causal path. This has several implications which are investigated in followup studies including evaluating whether DNAm is a target of natural selection to study its functionality, increasing the coverage of DNAm sites, investigating whether DNAm might only be causal in the disease relevant cell type or tissue.

The initial goal of GoDMC was to bring together researchers with an interest in studying the genetic basis of DNA methylation variation, consolidate as many resources and expertise as possible and thereby expedite this area of research. This collaborative effort resulted in a publication in Nature Genetics, and our database has been utilised in many other follow up publications. While completion of Phase One was a major achievement of its own, the analytical infrastructure we developed, and consortium of cohorts attracted interest for a number of follow on projects. Furthermore, the technologies have moved on since our initial pipeline was developed and we believe there is now scope to expand and update our mQTL catalogue. This has led to GoDMC Phase2 to reflect the evolution of our objectives.

The specific objectives for GoDMC Phase 2 are:

1. To encourage collaborations for rapid large--scale replication and meta--analyses following a conventional GWAS consortium structure.
2. To provide an online forum to publicise potential data sources and on--going studies.
3. To lead on the analysis of genetic variation against DNA methylation, generating the most comprehensive catalogue of DNA methylation quantitative trait loci (mQTL) and to make these available as a resource to the wider community. This includes analyses of different types of genetic variation and statistics derived from either DNA methylation or genetic variants.

4. To enable projects that require DNA methylation and/or genetic data with goals aligned to our own to take advantage of our consortium model.

This Code of Conduct outlines the principles of collaboration to which all consortium members should abide.

2. GoDMC Structure and Membership

Oversight of GoDMC will be provided by the Executive Committee comprising Prof. Caroline Relton, Dr. Jordana Bell, Prof. Bas Heijmans & Prof. Jonathan Mill. In addition to making the above commitments the Executive Committee will:

1. Promote GoDMC, its aims and objectives, and identify new opportunities for development.
2. Provide and maintain a forum to encourage and aid collaborations for replication/meta--analysis studies.
3. Promote and encourage best practice in data harmonisation and analysis methods developed through GoDMC.
4. Contribute to the development of good working relationships between collaborators by, for example, providing an appropriate code of conduct to which all members must abide.
5. Identify funding opportunities to aid replication and meta--analysis studies and to ensure appropriate use/distribution of such funds.
6. Meet regularly (twice a year) to discuss all matters relating to GoDMC and feedback any relevant information to all consortium members and support follow-on projects.
1.

The core group conducts coordinating tasks relevant for GoDMC and overseen by the Executive Committee. For example, they coordinate the implementation of new pipelines by the working groups, organize and coordinate working groups and ensure open science policies are followed. The core group comprises Prof Tom Gaunt, Prof Gibran Hemani, Dr. Eilis Hannon and Dr. Josine Min.

The GoDMC core group's primary tasks are:

2. To facilitate the science of the working groups
3. To identify new research opportunities
4. To coordinate the direction of the consortium

A Steering Committee will be established to:

1. Meet annually to offer comment on the portfolio of activities ongoing in GoDMC
2. Offer technical and ethical guidance on data use/access
3. Offer suggestions and highlight opportunities for new developments
4. Be willing to promote the consortium to new audiences as appropriate

In addition, the GoDMC coordinator organizes membership and communication with consortium members, developer group, core group, working groups , executive committee and steering committee.

Since phase 1, we have approved a number of follow on projects led by members of the consortium. This has led to the establishment of a Developer Group which consists of research

leads and their developers/analysts implementing the next phase of work packages.

For phase one, GoDMC included representatives from many research groups who conducted analysis and contributed summary statistics for analysis. These included summary statistics from both population and disease-specific cohorts capturing a range of ages. The inclusion criteria for GoDMC phase 2 have changed and therefore it is likely that the membership will similarly change. Each cohort has one or more PIs and analysts who are considered members of the consortium. The PI(s) is(are) responsible for approving participation and the analyst(s) responsible for performing analyses on the cohort level data locally. It is stressed that individual groups can participate in GoDMC to varying degrees, selecting which work packages they wish to contribute to. Our aim is to provide a user-friendly modular infrastructure so that minimal coding and time is required by analysts and which is user friendly for developers to contribute to. This is so that analyses can be planned into existing workloads as appropriate and reasonable timelines can be set to finalise the analyses. Similarly, not all projects arising through GoDMC will be applicable to every cohort. Hence, decisions regarding participation and authorship should be made on a project-by-project basis and this is implemented by the modular nature of the pipeline. As stated in our authorship guidelines, co-authors will be the ones that are doing the work or contribute data. Members do not, but can be involved in follow-on projects and working groups and will be part of the GoDMC banner included on the paper. To ensure transparency in the consortium, all available data sources, on--going research projects and code will be publicised on the GoDMC website (http://www.godmc.org.uk/).

In the unlikely event that disputes arise, the Executive Committee and Steering Group are on hand to help resolve such issues.

It is assumed that any data contributed to GoDMC 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.

3. GoDMC Phase 2 project areas and working groups

Six project areas have been set out and will be overseen by members of the Executive Committee. These project areas reflect the types of available datasets, analysis methods and data interpretation. To date projects have focused on some of the priority areas listed below. Going forward new projects would not necessarily need to align with these areas specifically.

1. SNPs and DNA methylation variation, deciphering *cis* vs. *trans* effects
2. Structural variants and DNA methylation variation, deciphering *cis* vs. *trans* effects
3. Tissue specificity and cell type specificity in DNA methylation
4. Variation in DNA methylation profiles across the lifecourse
5. Integrating methylation with other 'omics': evidence for shared genetic mechanisms
6. Using methylation quantitative trait loci as instrumental variables in Mendelian randomization approaches to identify causal mechanism of complex traits

A number of projects within these areas have already been proposed, details of which can be found on the GoDMC website www.godmc.org.uk/projects.html. In addition several working groups have been established including a Mendelian randomization working group and a EWAS of polygenic risk scores and GWAS of methylation derived scores working group. All cohorts meeting the entry criteria will be eligible and are encouraged to participate in these projects and working groups. It is hoped that these projects will be highly collaborative with joint contributions from all members.

4. Data sharing and confidentiality

The Executive Committee expects most collaborations 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.

However, in some instances sharing of individual level data (i.e. in the form of .idat files and phenotypic data) may be required. Each institution has its own policies for data sharing (i.e. data and material transfer agreements), which must be abided by. These policies will likely include guidelines regarding data transfer, confidentiality and intellectual property. It is the responsibility of the research--lead to identify and address these issues with interested GoDMC members. The Executive Committee encourages a responsible approach and commitment to respect the policies of individual institutions. GoDMC will not provide a means of data sharing and will not store any private data. It is felt that existing web-- based/university-based file transfer systems are adequate.

The summary statistics or partial derivatives will be analysed on secure servers at University of Bristol, King's College London, University of Exeter or Surfsara Netherlands or other secure servers approved by GoDMC. 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 or partial derivatives are used for meta-analyses
2. Meta-analysis results will be made publicly available
3. Cohort level summary statistics will be shared to collaborators of approved GoDMC projects

All GoDMC 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 GoDMC 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."

5. Data deposit in controlled-access repositories

Data deposits in open-- or controlled--access repositories are not required for participation in this consortium. However, **whenever it is compatible with consent etc., GoDMC members are strongly encouraged to upload their full dataset (including .idat files, genotypic data and sample annotations) into appropriate repositories. Alternatively, whenever it is compatible with consent de--identified genotype and methylation data should be uploaded at an appropriate time.** In addition to expediting progress via data sharing within the scientific community, this will prevent unnecessary delays to future GoDMC publications as many journals request that genomic data are made available in some way. Links to established repositories can be found on the GoDMC website www.godmc.org.uk/resources.html.

6. Software development and dissemination

GoDMC is committed to an open science ethos which includes making pipelines and all analytical code accessible and open through relevant code sharing platforms (e.g. Github) and with relevant open source licenses (e.g. MIT, GPL3).

Software may be developed privately amongst consortium analysts, but we will commit to making all relevant source code available no later than the time of manuscript submissions.

We expect all consortium members to adhere to the direction of GoDMC leads regarding the timing of when source code can be shared with members outside of the GoDMC consortium.

GoDMC members who contribute software or code will agree to the terms of the licenses for those specific code repositories.

7. Establishing projects and collaborations within GoDMC

The aim of GoDMC is to expedite research within this field by supporting large--scale replication and meta--analyses collaborations across the research community. Hence, GoDMC members are encouraged to establish and lead additional analysis projects; either genome--wide meta--analyses or targeted replication studies.

To ensure transparency and prevent duplication of research projects, all ongoing studies will be publicised on the GoDMC website www.godmc.org.uk/projects.html. This will also enable new consortium members to contribute to existing collaborations.

A project proposal form has been generated and will be provided by the GoDMC coordinator. Use of this form will ensure consistency across project plans and shared information. These forms should be filled out by the initiating research group (research--lead) and provide 1) background to the proposed project, 2) information on the primary analysis performed in the case of replication studies and 3) details of the proposed analysis pipeline and requested information. These details will enable potential collaborators to make informed decisions regarding their ability and/or wish to be involved. **Once collaborations have been established and analysis plans finalised, details of the project should be returned to the GoDMC coordinator for posting on the website.**

The Executive Committee do not wish to impose rules and regulations regarding future collaborations among consortium members. However, they do encourage that future projects, particularly large scale meta--analyses, utilise all available resources to their full potential by involving, where appropriate, all GoDMC cohorts interested in and able to contribute. Therefore projects will be reviewed by the Executive Committee and members of this committee are available to offer advice and guidance on research plans and collaborations if required. These points are in keeping with the aims and overarching goal of GoDMC.

8. Analysis plan and data generation

Primarily, GoDMC foresees members implementing a conventional GWAS consortium structure in which all collaborating groups run a standard analysis pipeline and feed back summary statistics to the research--lead for meta--analysis. Our aim is to provide an user-friendly pipeline so that minimal coding is required by analysts, which hopefully ensures that it is quick to execute, avoids duplication of efforts, increases reproducibility and reduces potential for error. It is the responsibility of the research--lead to provide a detailed analysis plan or github wiki page to members of the working group. This should include methods for pre--processing and statistical analysis. As a consensus regarding the optimum/appropriate pre--processing method for Illumina's Beadchip data is yet to be established within the research community, no single method will be championed or insisted upon by GoDMC at this time. However, some form of pre--processing and quality control is strongly encouraged. All GoDMC members are invited to make suggestions regarding what methods should be used. **Equally, in the interest of data harmonisation and consolidation, GoDMC members are encouraged to use meffil.**

9. Publication policy and dissemination of results

Publication of papers resulting from collaborations arising from GoDMC can take one of three forms. 1) Papers utilising data from all, or the majority, of cohorts contributing to GoDMC (i.e. in the case of large scale meta--analyses) can be published under a consortium byline with all contributing investigators listed as collaborators. 2) In addition to the consortium byline,

individual investigators providing a substantial contribution to the research can be named authors. In either case it is recommended that a footnote listing the name, affiliation and specific contributions of consortium members is provided (details can be obtained from the GoDMC coordinator). 3) In the case of smaller collaborations/projects all contributing investigators can be named authors. In this case, we ask authors to acknowledge the contribution of GoDMC in supporting and encouraging the collaboration.

Decisions regarding authorship, timing of data release and publication (both manuscript and conference proceedings) are the responsibility of the individual working groups and the research-lead to address fairly. Please note that individual journals have different policies regarding authorship, especially in the case of consortium bylines, and provide details on their websites which should be considered prior to manuscript submission.

In GoDMC we follow the widely accepted icmje criteria for authorship on scientific papers as described in our authorship policy. Following our open science ethos, manuscripts will be uploaded to preprint servers after submission and papers will be published in open-access journals.

Members of this consortium retain the capacity to publish data on their own datasets and to engage in any additional research and collaborations they choose. However, if members do engage in additional work that overlaps with an ongoing GoDMC project it is courteous to inform their GoDMC collaborators about this. Again, it is stressed that participation in GoDMC should occur at a time that is appropriate for the individual group.

Details of any publications (both manuscript and conference proceeding) arising from GoDMC, or regarding cohorts and data contributing towards GoDMC, can be forwarded to the coordinator for posting on the website.

10. Indication of agreement

You are to indicate your agreement with this Code of Conduct by signing the cohort sheet.