GoDMC Code of Conduct

1. Introduction

Epigenetic epidemiology is now a well-established field of research offering the promise of prediction, prevention and treatment to 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 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). However, many studies have been hindered by small sample sizes and there are concerns that this field will suffer the same pitfalls as early genome-wide association studies (GWAS). The Genetics of DNA Methylation Consortium (GoDMC) was established with this concern in mind.

The overarching goal of the GoDMC is 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. In order to achieve this goal the following aims and objectives have been set out:

a. To encourage collaborations for rapid large-scale replication and meta-analyses following a conventional GWAS consortium structure.
b. To provide an online forum to publicise potential data sources and on-going studies.
c. To prioritise the analysis of Illumina HM450K BeadChip and GWAS data in the first instance (and other bisulphite modified DNA methylation data in due course).
d. To include all tissue types and disease/phenotypic subgroups.

As individual GWAS of DNA methylation will be context specific (e.g. tissue, age, disease status etc.), this creates many possibilities in terms of questions that can be addressed and multiple opportunities for wide ranging replication and meta-analysis collaborations between members of the research community. Most importantly, the success of this approach in achieving similar goals have been realised repeatedly in the numerous consortia arising from GWAS.

An Executive Committee (comprising Prof. Caroline Relton, Dr. Jordana Bell, Dr. Christoph Bock, Dr. Bas Heijmans & Prof. Jonathan Mill) and Steering Group (comprising Professors Andrea Baccarelli, Stephan Beck, George Davey Smith, Mark McCarthy, Peter Visscher and Dr. Nicole Soranzo) have been established to oversee the development and maintenance of GoDMC.

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

2. GoDMC Membership

At the time of this writing, representatives from 27 research groups had expressed interest in forming the consortium. These include a collection of population and disease specific cohorts capturing a range of ages and ethnic backgrounds.

It is intended that all GoDMC members will be equally involved with no single group or individual dominating. Individual investigators are encouraged to establish and lead specific analysis projects. This is likely to reflect the subgroup for which they intend to contribute data.

It is stressed that individual groups should participate in GoDMC at a time that is appropriate for their own research activities. Similarly, not all projects arising through GoDMC will be applicable to every cohort. Hence, decisions regarding participation should be made on a project-by-project basis.

To ensure transparency in the consortium, all available data sources and on-going research projects 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. Your Indication of Agreement to this code of conduct (see below) will be taken as confirmation of this.

3. GoDMC project areas

Five 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.

i. SNPs and DNA methylation variation, deciphering cis vs. trans effects
ii. Structural variants and DNA methylation variation, deciphering cis vs. trans effects
iii. Tissue specificity and ethnic differences in DNA methylation
iv. Variation in DNA methylation profiles across the lifecourse
v. Integrating methylation with other 'omics': evidence for shared genetic mechanisms

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. All cohorts meeting the entry criteria will be eligible and are encouraged to participate in these projects. It is hoped that these projects will be highly collaborative with joint contributions from all.

4. 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 co-ordinator. 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 co-ordinator 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 contributing GoDMC cohorts. Equally, projects will not be reviewed by, nor approved by, the Executive and Steering Committees. However, members of these committees 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.

5. 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. It is the responsibility of the research-lead to provide a detailed analysis plan to members of the working group. This should include methods for pre-processing and statistical analysis. Distribution of programming scripts may be advantageous.

As a consensus regarding the optimum/appropriate pre-processing method for HM450K 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 encouraged. Hence, details regarding currently available methods will be provided on the GoDMC website www.godmc.org.uk/resources.html with links to relevant websites. All GoDMC members are invited to make suggestions regarding what methods should be posted on the GoDMC website. **Equally, in the interest of data harmonisation and consolidation, GoDMC members are strongly encouraged to make their programming scripts publically available.** Such scripts can be posted on the GoDMC website if members wish.

6. 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 co-ordinator). 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.

Here are some links to websites with more details regarding authorship:

http://www.nlm.nih.gov/pubs/factsheets/authorship.html

http://www.pnas.org/content/101/29/10495.full

http://www.nature.com/ng/journal/v41/n4/full/ng0409-383.html

Members of this consortium retain the capacity to publish data on their own samples 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 co-ordinator for posting on the website.

7. 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 to enable replication and meta-analyses. This format overcomes the need to share individual level 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 encourage 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 individual level data. It is felt that existing web-based/university-based file transfer systems are adequate.

Finally, 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.

8. 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, 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.

9. Indication of agreement

You are to indicate your agreement with this Code of Conduct via email to the GoDMC co-ordinator.