MEDICAL GENOME REFERENCE BANK DATA ACCESS APPLICATION FORM

Medical Genome Reference Bank Data Access Application

Prepared by the MGRB Data Access Committee

Version 1.1 – 1st April 2017

SECTION 1 – BACKGROUND

1.1 The Medical Genome Reference Bank

The Medical Genome Reference Bank (MGRB) is comprised of whole genome sequencing (WGS) data and phenotypic information from (up to) 4,000 healthy Australians over 70 years of age. MGRB participants, consented through contributing studies, 45 and Up (Sax Institute, Sydney), and the ASPirin in Reducing Events in the Elderly (ASPREE) clinical trial (Monash University, Melbourne), are free from cardiovascular disease, degenerative neurological disorders and of a history of cancer at the time of consent into the study. The dataset can function as a powerful filter to distinguish between causal genetic and population-based genetic variation and be a resource to maximise the efficiency of genomic discovery in both the research and clinical setting.

WGS was performed on the Illumina HiSeq X-Ten platform at the Garvan Institute of Medical Research (Sydney, Australia) under clinically accredited conditions (ISO 15189). Data was aligned and variant files were generated using best practice (BWA, GATK) pipelines to make results comparable with other cohorts. Joint-called variants were loaded into an analytical framework (openCGA) to enable the interrogation of genomic variants using advanced filters based on basic clinical traits and genomic annotations. These curated data are openly accessible to the international research community through the **MGRB Data Portal**. Preliminary features of the data portal include a data Beacon, extensive variant annotation, complex population-based clustering queries, visualisation of variant data and analysis tools for assessing the genetic burden of individual variants and variant subsets. While basic demographic and phenotypic information are incorporated into the MGRB data portal, researchers are invited to complete a **Data Access Application** to gain access to comprehensive genotypic and clinical information, to support high-level integrative analysis. This application is based on the exacting standards of the European Genome-phenome Archive.

VCF, gVCF, BAM and FASTQ data are available. Information on the comprehensive clinical data that is available through ASPREE and 45 and UP (pending a successful application) can be found in the following documents:

ASPREE
ASPREE Protocol AUS Version# 9 Nov 2014

45 and Up
Download the Baseline Questionnaire for Women (PDF 537KB)
Download the Baseline Questionnaire for Men (PDF 537KB)

1.2 Data Access Policy

To maintain participant privacy and confidentiality, whilst maximising MGRB utility, we have deployed a tiered data management system that determines the depth of data that is made available to researchers (as summarised in the schematic below). This consists of 3 access tiers; Open access, Controlled access and Restricted access. Completion of the MGRB DAA is required for access to Controlled and Restricted access data. The MGRB Data Access Committee (DAC) will review completed applications within 6 weeks. Upon approval, Controlled access can then be immediately made available to the applicant. Should Restricted access data be requested, the application will be immediately forwarded to the governing body of the participating cohort, and will be considered in parallel with the MGRB DAC review. The cohort governing body may insist on an independent application process, the time-line for which dependent solely on the governing body – more information is provided on the respective websites for ASPREE and 45 and UP. A copy of the full MGRB Data Access Policy can be found here.

1.3 Data Access Policy Schematic

Tier	Open Access	Controlled Access	Restricted Access
Access	Institutional email address required for MGRB data-portal access (**not** required for Beacon) (www.sgc.garvan.org.au/mgrb)	Data Access Application (DAA) must be approved by the MGRB Data Access Committee (DAC)	DAA must be approved by the MGRB DAC **and** referred to the applicable cohort governing body for **further** approval
Clinical Data	Basic demographic data are provided - genomic queries can be filtered according to these fields	Basic demographic data and minimal clinical information (where available) are provided per individual record	Comprehensive clinical data that is potentially specific to a participating cohort is provided per individual record
Genomic Data	Beacon and preprocessed population-scale variant analysis	Individual record data provided – either processed (VCF/ gVCF format), or unprocessed (FASTQ or BAM format) (dependent on justification criteria being met)	

1.4 File Access

Once downloaded, minimum protection measures are required to protect all MGRB data. Data can be held in unencrypted files on an institutional compute system, with the equivalent of Unix user group read/write access for one or more appropriate groups but not Unix world read/write access behind a secure firewall. Laptops holding these data should have password protected logins and screenlocks (set to lock after 5 min of inactivity). If held on USB keys or other portable hard drives, the data must be encrypted.

SECTION 2 - DATA ACCESS APPLICATION

2.1 Project Overview

Level of data required **Controlled** ☐ **Restricted** ☐
Provide a title and 1-page overview of the proposed project. **Include background, rationale, methodological details and analytic approach. Please also include what level of information is being requested and how MGRB data will be used.** **(500 words or more)** Nb. Use <option/alt> + <return> for new line.
Title: **Overview:**

2.2 Genomic Data

Select the MGRB data format required for the proposed project (tick both if required)	
45 and Up ☐ **ASPREE** ☐	
Select the MGRB data format required for the proposed project	
VCF ☐ **gVCF** ☐ **BAM** ☐ **FAST Q** ☐	
Should BAM/ FASTQ data be requested – provide appropriate justification as to why this format (and not preprocessed data) is necessary for purpose of the proposed project	
Provide detail as to how MGRB data will be stored and what security measures are in place to ensure conformity with the conditions stipulated in Section 1.4)	

2.3 Clinical Information

Clinical data (Controlled Access) for 45 and UP and ASPREE are listed below – select the data required for the proposed study. (Should additional clinical information be required from either cohort, please state what data are required and justify the release of this data – release of this information is at the discretion of the external governing board for the cohort in question and maybe subject to an independent proposal)	
45 and UP	
Sex ☐ **Year of Birth** ☐ **Height** ☐ **Weight** ☐	
Additional clinical information required from 45 and Up and justification	
ASPREE	
Sex ☐ **Year of Birth** ☐ **Height** ☐ **Weight** ☐ **Systolic Blood Pressure** ☐ **Evidence of Macular Degeneration** ☐ **Abdominal Circumference** ☐ **Resting Glucose** ☐	
Additional clinical information required from ASPREE and justification	

SECTION 3 – TERMS AND CONDITIONS

These terms and conditions govern access to the Medical Genome Reference Bank dataset to which the User Institution has requested access (details of which are set out in Section 2.2 and 2.3). The User Institution agrees to be bound by these terms and conditions.

3.1 The User Institution agrees to only use these Data for the purpose of the Project (described in Section 2.1) and only for Research Purposes. The User Institution further agrees that it will only use these Data for Research Purposes, which are within the limitations of their existing ethical framework.

3.2 The User Institution agrees to preserve, at all times, the confidentiality of these Data. In particular, it undertakes not to use, or attempt to use these Data to compromise or otherwise infringe the confidentiality of information on Research Participants. Without prejudice to the generality of the foregoing, the User Institution agrees to use at least the measures set out in Section 1.4 to protect these Data.

3.3 The User Institution agrees to protect the confidentiality of Research Participants in any research papers or publications that they prepare by taking all reasonable care to limit the possibility of identification.

3.4 The User Institution agrees not to link or combine these Data to other information or archived data available in a way that could re-identify the Research Participants, even if access to that data has been formally granted to the User Institution or is freely available without restriction.

3.5 The User Institution agrees only to transfer or disclose these Data, in whole or part, or any material derived from these Data, to the Authorised Personnel (as defined in Section 5). Should the User Institution wish to share these Data with an External Collaborator, the External Collaborator must complete a separate application for access to these Data.

3.6 The User Institution agrees that the Data Producers, and all other parties involved in the creation, funding or protection of these Data: a) make no warranty or representation, express or implied as to the accuracy, quality or comprehensiveness of these Data; b) exclude to the fullest extent permitted by law all liability for actions, claims, proceedings, demands, losses (including but not limited to loss of profit), costs, awards damages and payments made by the Recipient that may arise (whether directly or indirectly) in any way whatsoever from the Recipient's use of these Data or from the unavailability of, or break in access to, these Data for whatever reason and; c) bear no responsibility for the further analysis or interpretation of these Data.

3.7 The User Institution agrees to follow the *Fort Lauderdale Guidelines* (*http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtd003207.pdf*) and the *Toronto Statement* (*http://www.nature.com/nature/journal/v461/n7261/full/461168a.html*). This includes but is not limited to recognising the contribution of the Data Producers and including a proper acknowledgement in all reports or publications resulting from the use of these Data.

3.8 The User Institution agrees to follow the *Publication Policy* described in Section 4

3.9 The User Institution agrees not to make intellectual property claims on these Data and not to use intellectual property protection in ways that would prevent or block access to, or use of, any element of these Data, or conclusion drawn directly from these Data.

3.10 The User Institution agrees to destroy/discard the Data held, once it is no longer used for the Project, unless obliged to retain the data for archival purposes in conformity with audit or legal requirements.

3.11 The User Institution will notify the MGRB Data Access Committee within 30 days of any changes or departures of Authorised Personnel.

3.12 The User Institution will notify the MGRB Data Access Committee as soon as it becomes aware of a breach of the terms or conditions of this agreement.

3.13 The MGRB Data Access Committee may terminate this agreement by written notice to the User Institution. If this agreement terminates for any reason, the User Institution will be required to destroy any Data held, including copies and backup copies. This clause does not prevent the User Institution from retaining these data for archival purpose in conformity with audit or legal requirements.

3.14 The User Institution accepts that it may be necessary for the Data Producers to alter the terms of this agreement from time to time. In the event that changes are required, the Data Producers or their appointed agent will contact the User Institution to inform it of the changes and the User Institution may elect to accept the changes or terminate the agreement.

3.15 If requested, the User Institution will allow data security and management documentation to be inspected to verify that it is complying with the terms of this agreement.

3.16 The User Institution agrees to distribute a copy of these terms to the Authorised Personnel. The User Institution will procure that the Authorised Personnel comply with the terms of this agreement.

3.17 This agreement (and any dispute, controversy, proceedings or claim of whatever nature arising out of this agreement or its formation) shall be construed, interpreted and governed by the laws of News South Wales, Australia.

SECTION 4 - PUBLICATION POLICY

The User Institution is required to include an appropriate acknowledgement (and authorship if deemed necessary) in any Publication that makes use of or reference to MGRB dataset in accordance with the following terms;

4.1 Open Access and/or Controlled Access data

Publications that make use of Open Access and/or Controlled Access data will be required to acknowledge the MGRB Collaborative in the acknowledgements section of the Publication. Should the User Institution wish to add MGRB Collaborative to the list of authors, the User Institution should share the Publication with the MGRB Data Access Committee for approval prior to submission. Manuscripts or presentations should be submitted to MGRB@garvan.org.au for review. Authors are also encouraged to recognise the contribution of the appropriate cohort convenors via the acknowledgements section in their Publication.

An example of a proper attribution is:
"The results <published or shown> here are in whole or part based upon data generated by the MGRB Collaborative: http://sgc.garvan.org.au/mgrb/initiatives.

4.2 Restricted Access data

Publications that make use of Restricted Access data should acknowledge the MGRB collaborative (as described in Section 4.1) and give authorship to contributing members of the MGRB collaborative in accordance with ICMJE guidelines. Publications that include Restricted Access MGRB data must be submitted to the MGRB Data Access Committee for approval prior to submission (or presentation). Manuscripts or presentations should be submitted to MGRB@garvan.org.au for review.

SECTION 5 - USER INFORMATION

5.1 All Individuals who the User Institution request to be named as registered users

Name of User	***Email***	***Job Title***	***User Institution***

5.2 Executed as an Agreement.

Both parties confirm that they have read, understood and accept the terms and conditions outlined in this Application.

Requesting USER INSTITUTION

SIGNED on behalf of **USER INSTITUTION**, by its duly authorised officer, in the presence of:

Date

Signature of witness

Signature of officer

Name

Name

Position

Position

SIGNED for **THE MGRB DATA ACCESS COMMITTEE**, by its duly authorised officer, in the presence of:

Date

Signature of witness

Signature of officer

Name

Name

Position

Position

Please return the completed Application to MGRB@garvan.org.au

SECTION 6 - DEFINITIONS

***Authorised Personnel**:* The individuals at the User Institution to whom the MGRB Data Access Committee grants access to the Data. This includes the User, the individuals listed in Section 5 and any other individuals for whom the User Institution subsequently requests access to the Data. Details of the initial Authorised Personnel are set out in Section 6.

***Data**:* The managed access datasets to which the User Institution has requested access.

***Data Producers**:* The MGRB collaborative that is responsible for the development, organisation, and oversight of these Data.

***External Collaborator**:* A collaborator of the User, working for an institution other than the User Institution.

***Project**:* The project for which the User Institution has requested access to these Data. A description of the Project is set out in Section 2.1.

***Publications**:* Includes, without limitation, articles published in print journals, electronic journals, reviews, books, posters and other written and verbal presentations of research.

***Research Participant**:* An individual whose data form part of these Data.

Research Purposes: Shall mean research that is seeking to advance the understanding of genetics and genomics, including the treatment of disorders, and work on statistical methods that may be applied to such research.

***User**:* The principal investigator for the Project.

***User Institution(s)**:* The Institution that has requested access to the Data.