KINGHORN CENTRE FOR CLINICAL GENOMICS | GARVAN INSTITUTE OF MEDICAL RESEARCH

Medical Genome Reference Bank – *Data Access Policy*

Version 1.03

Summary Statement

The Data Access Policy (DAP) described herein summarises the governance policy applied to individual and institutional access to the Medical Genome Reference Bank (MGRB).

The MGRB is comprised of whole genome sequencing (WGS) data and phenotypic information from 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. 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).

Curated data will be openly accessible to the international research community through an MGRB website. Preliminary features will include a Beacon, as defined by the Global Alliance for Genomics and Health, extensive variant annotation, complex queries (including genetic annotations, and genomic regions), visualisation of variant data (e.g. genome viewer/ gene networks) and ultimately, analysis tools for assessing the genetic burden of individual variants and variant subsets. While basic demographic and phenotypic information will be incorporated into the MGRB data portal, researchers are invited to apply for access to comprehensive genotypic and clinical information to support high-level integrative analysis.

To maintain participant privacy and confidentiality, whilst maximising MGRB utility, we have deployed a tiered data management system that determines the richness 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.

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. **Open Access Tier:**

- The Open Access data tier is comprised of the MGRB Beacon and Data Portal (www.sgc.garvan.org.au/mgrb)

- The Open Access tier **does not** require formal application.

- The MGRB data Beacon is publically accessible through the MGRB landing page (https://sgc.garvan.org.au/initiatives/mgrb).

- To access the Open Access MGRB Data Portal, users are required to register with an email address, primarily to record usage and distribute MGRB updates (opt-out available).

- Users are permitted to query the MGRB dataset using Gene Symbols, SNP IDs, as well as intra-chromosomal genomic ranges. Variants present in a pre-selected region will be presented both visually, and be downloadable (.csv format) as a summary statistical representation of the cohort (similar to the ExAC browser).

- Basic demographic filters, including gender, height, weight, year of birth, systolic blood pressure, abdominal circumference and resting glucose (where available), can be applied to refine queries. However, results will not be provided for groups of less than 50 individuals, where multiple filters have been applied.

2. Controlled Access Tier:

- The Controlled Access data tier permits researchers to access genomic data, and basic clinical information (as described in the Open Access tier) from a specified (or non-specified), non-identifiable population of MGRB individuals.

- The Controlled Access data tier requires the User to submit a Data Access Agreement (DAA) application, which is based broadly on the European Genome-phenome Archive (EGA) DAA template. Applicants must agree to all conditions described in the DAA.

- The MGRB Data Access Committee (DAC), which is made up of representatives of Garvan, Monash University and Sax Institute, periodically assess DAA applications.

- Genomic data will be provided as pre-processed VCF or gVCF files. FASTQ and BAM files will also be available should appropriate justification be provided. Data will be stored and securely downloaded from the National Computational Infrastructure (NCI).

3. Restricted Access Tier:

- The Restricted Access data tier is an extension of the Controlled Access tier, in that applicants must firstly submit a DAA application, which then may be recommended by the MGRB DAC for escalation to a Restricted Access application. Escalation to the Restricted Access tier will result if a researcher requests (and can justify) comprehensive participant clinical information.

- Should clinical information that is not provided in tier 1/2 be requested, approval would be at the discretion of the governing body of ASPREE and/or 45 and UP.

- The purpose of the MRGB DAC as an intermediary is to capture the volume of requests for Restricted Access usage, to assign tier of access required, and to monitor consumer satisfaction with this process.