

Data Access Committee for SG10K Pilot Dataset

Version 3.0

Date: 20 September 2021

Purpose

The National Precision Medicine Data Access Committee (NPM DAC) manages the access and usage of the SG10K_Pilot Dataset, deposited at the European Genome-phenome Archive (EGA), which is hosted by EMEL-EBI and CRG, under accession number EGAS00001003875. The SG10K_Pilot Investigators shall remain as the rightful owners of the datasets and reserve the rights to use the data for their own research and collaborations.

This document summarises the data access policies and procedures to enable researchers to access and use the genomic data generated in the SG10K_Pilot study.

Definitions

“NPCO” refers to the NPM Program Coordinating Office, A*STAR, which is the Program Office housed under Genome Institute of Singapore providing secretariat support to the DAC.

“CRG” refers to the Centre for Genomic Regulation

“Data Access Committee” (DAC) refers to the committee managing access to the available SG10K_Pilot Datasets

“EGA” European Genome-phenome Archive (www.ega-archive.org)

“EMBL-EBI” refers to the European Bioinformatics Institute

“NPM” refers to the Singapore National Precision Medicine Program

“PMSC” refers to the Precision Medicine Steering Committee

“SG10K_Pilot Dataset” refers to the *Cell* dataset (EGAD00001005337) publically available via European Genome-phenome Archive (or EGA)

“SG10K_Pilot Investigators” refer to the investigators involved in the generation of the SG10K_Pilot Datasets and the institutional PIs of each cohort contributing to the SG10K_Pilot Datasets

SG10K Pilot Dataset

SG10K_Pilot published in *Cell* and archived at the EGA (www.ega-archive.org) relates to whole genome sequencing data for 4,810 Singaporeans. All datasets have been pseudonymised, and so considered de-identified as per described in the paper.

IRB approval for use of the SG10K Pilot Dataset

IRB approvals are in place to enable the data to be used for biomedical research that is in the public interest. Applications will be judged against these existing IRB approvals to assess whether the proposed use is acceptable and permitted. In exceptional circumstances, a new IRB approval may be needed to enable the work proposed.

Data Access Policy

1) Data access requests should be submitted for review by the Data Access Committee, using the SG10K_Pilot Data Access Form.

2) The Committee secretariat will be responsible for correspondence with applicants, co-ordinating meetings, distributing applications for review in advance of meetings, keeping minutes on all discussions, recording outcomes, communicating outcomes to the applicant and liaising with appropriate Data Management Teams.

3) The Committee will maintain an active internal registry of data access requests and monitor progress. A register of approved applications and SG10K_Pilot primary publications will be made publicly available through the study website.

4) In the initial phase, there will be no data access fees. In the future, the PMSC may consider introducing reasonable data access fees to applicants (cost-recovery), to help ensure the sustainability of maintaining the data for use by external researchers.

5) The DAC recognizes that it is impossible for these policies to fully cover all possible publication scenarios and manuscripts. As such, the Committee reserves the right to modify these guidelines for publications on a case-by-case basis, based on the specifics of the proposed data access request.

Obligations of Data Access Requestors

1) Applicants must:

- Provide evidence that they are bona-fide researchers (including ORCID ID)
- Submit a study Data Access Request form (Annex A) and obtain DAC approval before using the data for the research proposed
- Ensure data are held securely, and used only by applicants listed on the proposal, or by the researchers employed in the teams that they lead at their respective host institutions.
- Use the data only for the experiment described in the application
- Submit an application to extend the use if necessary for further review by the DAC.
- Agree to make no attempt to re-identify individuals in the study
- Provide a half-page progress of the research project in a form as required by DAC 1.5 years from the date of data access.
- Destroy the data at the end of the study, or when asked to do so by the DAC.
- Acknowledge the contributions of SG10K_Pilot Dataset investigators in all grants and publications arising, using the approved text for methods and acknowledgments.
- Agree to any other conditions of use specified by the DAC

2) Applicants agree to refer to the “SG10K_Pilot” Dataset where these data have been used, and to indicate the data freeze used in their manuscript. Applicants also agree to include the SG10K_Pilot acknowledgment statement and reference our *Cell* paper [DOI: 10.1016/j.cell.2019.09.019] (illustrated below)

Acknowledgment statement

“We thank the “SG10K_Pilot Investigators” for providing the SG10K_Pilot data (EGAD00001005337) The data from the “SG10K_Pilot Study” reported here were obtained from EGA. This manuscript was not prepared in collaboration with the “SG10K_Pilot Study” and does not necessarily reflect the opinions or views of the “SG10K_Pilot Study”.

Publication

*“Wu D, Dou J, Chai X, Bellis C, Wilm A, Shih CC, Soon W, Bertin N, Lin CB, Khor CC, DeGiorgio M, Cheng S, Bao L, Karnani N, Hwang W, Davila S, Tan P, Shabbir A, Moh A, Tan EK, Foo JN, Goh LL, Leong KP, Foo R, Lam C, Richards AM, Cheng CY, Aung T, Wong TY, Ng HH, Liu JJ, Wang C and SG10K Consortium. **Large-scale whole-genome sequencing of three diverse Asian populations in Singapore.** *Cell*. 2019 Oct 17;179(3):736-749.e15”.*

3) In situations where non-compliance to these policies are discovered, the DAC reserves the right to block submission of manuscripts, and may consider contacting relevant journals to raise concerns.