

# NPM SG10K\_Health Data Access Policy

## Introduction

This document sets out the data access policies related to use of the SG10K\_Health Datasets generated within the NPM programme. As a community resource project, the policies aim to strike a balance between providing users of the resource timely access to the SG10K\_Health data, yet recognizing the legitimate interest of the SG10K\_health data producers to publish primary reports of their work.

The access guidelines in this document are specific to manuscripts intended for publications in journals or as internal white papers to guide policy decisions. They do not cover uses of SG10K\_Health in other transmissions, such as social media (blogs, tweets), conference abstracts and presentations, and other media mentions. For these latter avenues, presenters are to contact the NPM Programme Coordinating Office for guidance (contact\_npco@gis.a-star.edu.sg).

## Purpose

This document summarises the data access procedures to enable researchers from academic settings in Singapore to use the genomic data generated in Phase 1 of the National Precision Medicine (NPM) programme.

Requests for data access from industry and from researchers outside Singapore will be managed on a case-by-case basis. Please contact the NPM Programme Coordinating Office (NPCO, contact\_npco@gis.a-star.edu.sg) for more information.

## About SG10K\_Health Datasets

Phase 1 of the National Precision Medicine (NPM) Project will generate whole genome sequencing data and methylation data by microarray for ~10,000 Singaporeans. (Table 1). All datasets have been pseudonymised, and so are considered de-identified.

**Table 1.** NPM Phase I National Precision Dataset

<b><i>Clinical data</i></b>
<ul style="list-style-type: none"><li>• Year of birth</li></ul>
<ul style="list-style-type: none"><li>• Gender</li></ul>
<ul style="list-style-type: none"><li>• Ethnic group</li></ul>
<b><i>Genomic</i></b>
<ul style="list-style-type: none"><li>• Whole genome sequence data</li></ul>
<ul style="list-style-type: none"><li>• EPIC array methylation data</li></ul>
<b><i>Cohorts contributing (N, WGS Depth, design)</i></b>
<ul style="list-style-type: none"><li>• SEED (N=1534, 15x, population-based)</li></ul>
<ul style="list-style-type: none"><li>• TTSH (N=630, 15x, population-based)</li></ul>
<ul style="list-style-type: none"><li>• PRISM (N=1349, 30x, population-based)</li></ul>
<ul style="list-style-type: none"><li>• GUSTO-kids (N=1000, 30x, birth cohort)</li></ul>
<ul style="list-style-type: none"><li>• MEC (N=3000, 15x, population-based)</li></ul>
<ul style="list-style-type: none"><li>• HELIOS (N=2400, 15x, population-based)</li></ul>
<ul style="list-style-type: none"><li>• SSMP (N=96, 30x, population-based)</li></ul>

## IRB approval for use of the NPM datasets

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 NPM Data Access Form.
2. Data requestors should recognize that NPM leadership intends to author a series of primary manuscripts based on the SG10K\_Health Precision Medicine Dataset. Data requests overlapping these primary manuscripts will be referred to the NPM PI assigned to the primary manuscript.
3. In the initial stages of the NPM, the DAC will prioritise applications from researchers in Singapore and other key strategic collaborations as determined by NPM leadership.
4. Publications should include an appropriate number of NPM co-PIs involved in generation of the SG10K\_Health Precision Medicine dataset, so as to ensure appropriate data interpretation. The Committee reserves the right to decide on the suitability and number of NPM co-PIs, depending on the scale and complexity of the data request.
5. The Committee recognizes the importance of team science in community resources projects such as SG10K\_Health, and expects that most, if not all, primary publications will be co-led by multiple lead and senior authors.
6. The Committee will maintain an active internal registry of data access requests and monitor progress. A register of approved applications and NPM primary publications will be made publicly available through the study website.
7. 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.
8. As a general principle, the DAC will only allow data download of aggregated data.
9. 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

10. Applicants must:
  - a. Provide evidence that they are bona-fide researchers (including ORCID ID)
  - b. Submit a study Data Access Request form (Annex B) and obtain DAC approval before using the data for the research proposed
  - c. 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.
  - d. Use the data only for the experiment described in the application
  - e. Submit an application to extend the use if necessary for further review by the DAC.
  - f. Agree to make no attempt to re-identify individuals in the study
  - g. Provide an annual report on study progress
  - h. Destroy the data at the end of the study, or when asked to do so by the DAC.
  - i. Acknowledge the contributions of NPM and its investigators in all grants and publications arising, using the approved text for methods and acknowledgments.
  - j. Agree to any other conditions of use specified by the DAC or PMSC

11. Prior to journal submission, applicants agree to send a version of the manuscript to the DAC for approval. This includes deposition of manuscripts in pre-publication archives (e.g. BioRxiv).
12. If intellectual property is generated as a result of the data analysis, publication authors agree to file a Technology Disclosure using templates provided by A\*STAR and agree to work with A\*STAR on subsequent intellectual property protection activities.
13. Applicants agree to refer to the “SG10K\_Health” dataset where these data have been used, and to indicate the data freeze used in their manuscript. Applicants also agree to include the NPM funding acknowledgment statement (illustrated below). This will be updated at intervals, and applicants should obtain the current version from the DAC before submitting manuscripts for publication.

*“This study made use of data generated as part of the Singapore National Precision Medicine program funded by the Industry Alignment Fund (Pre-Positioning) (IAF-PP: H17/01/a0/007). This study made use of data / samples collected in the following cohorts in Singapore: (1) The Health for Life in Singapore (HELIOS) study at the Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore (supported by grants from a Strategic Initiative at Lee Kong Chian School of Medicine, the Singapore Ministry of Health under its Singapore Translational Research Investigator Award (NMRC/STaR/0028/2017) and the IAF-PP: H18/01/a0/016; (2) The Growing up in Singapore Towards Healthy Outcomes (GUSTO) study, which is jointly hosted by the National University Hospital (NUH), KK Women’s and Children’s Hospital (KKH), the National University of Singapore (NUS) and the Singapore Institute for Clinical Sciences (SICS), Agency for Science Technology and Research (A\*STAR) (supported by the Singapore National Research Foundation under its Translational and Clinical Research (TCR) Flagship Programme and administered by the Singapore Ministry of Health’s National Medical Research Council (NMRC), Singapore - NMRC/TCR/004-NUS/2008; (3) The Singapore Epidemiology of Eye Diseases (SEED) cohort at Singapore Eye Research Institute (SERI) (supported by NMRC/CIRG/1417/2015; NMRC/CIRG/1488/2018; NMRC/OFLCG/004/2018); (4) The Multi-Ethnic Cohort (MEC) cohort (supported by NMRC grant 0838/2004; BMRC grant 03/1/27/18/216; 05/1/21/19/425; 11/1/21/19/678, Ministry of Health, Singapore, National University of Singapore and National University Health System, Singapore); (5) The SingHealth Duke-NUS Institute of Precision Medicine (PRISM) cohort (supported by NMRC/CG/M006/2017\_NHCS; NMRC/STaR/0011/2012, NMRC/STaR/ 0026/2015, Lee Foundation and Tanoto Foundation); (6) The TTSH Personalised Medicine Normal Controls (TTSH) cohort funded (supported by NMRC/CG12AUG17 and CGAug16M012).*

*The views expressed are those of the author(s) are not necessarily those of the National Precision Medicine investigators, or institutional partners. We thank all investigators, staff members and study participants who made the National Precision Medicine Project possible.”*

*[This text is for illustration only – contact the DAC for the current version]*

14. For certain publications, data requestors will agree if directed by the DAC to include in the authorship list that includes the “SG10K\_Health Consortium” author block, and refer to an expanded list of co-authors in the paper, provided to the authors by the Committee secretariat.

15. For publications, data requestors agree to include as co-authors NPM investigators that have contributed to the publication, using guidelines for scientific authorship as described by the International Committee of Medical Journal Editors (ICMJE)<sup>1</sup>.
16. 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.
17. In situations where manuscript-related disagreements may arise, publication authors have the right to appeal to the DAC. DAC decisions related to appeals will be regarded as final.

---

<sup>1</sup> <http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html>

### Definitions

“NPCO” refers to the NPM Programme Coordinating Office, which is the Programme Office housed under Genome Institute of Singapore, A\*STAR, supervising the implementation of NPM Phase IA.

“Cohort PIs” refer to the institutional PIs of each cohort contributing to Precision Medicine datasets

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

“IAF-PP” refers to the Industry Alignment Fund Pre-positioning grant mechanism, which has funded the creation of the Precision Medicine Dataset

“NPM” refers to the Singapore National Precision Medicine Program

“Phase IA” refers to the NPM segment related to the production of genomic data related to SG10K\_Health

“Phase IB” refers to the linkage of SG10K\_Health individual identifiers to electronic medical record and hospital operational data.

“PMSC” refers to the Precision Medicine Steering Committee

“SG10K\_Health” refers specifically to the collection of genomic data derived from ~10,000 individuals from Singapore, that were apparently healthy at enrolment. At the time of writing, “SG10K\_Health” comprises samples from the PRISM, GUSTO, HELIOS, MEC, SEED and TTSH cohorts.

“SG10K\_Health Precision Medicine Dataset” refers to the final data set of genomic data obtained from the joint processing and variant calling of all SG10K\_Health samples and DNA methylation data.

Additional cohorts and samples will be added to the NPM dataset as the study evolves; some of these cohorts may be disease based. The Data access form makes clear which cohorts are population-based, and which are disease-based.

<b><u>NPM SG10K Health Data Access Form</u></b>		
<b>For official use</b>		
Request received date: Click or tap to enter a date.		Request number:
<b>Section A: Requestor details</b>		
A1. Full name:		A2. Designation:
A3. Department /organisation:		
A4. Email address:		A5. Phone number:
A6. ORCID:		A7. Date of request: Click or tap to enter a date.
<b>Section B: Details of request</b>		
B1. Title of research study:		
B2. Proposed start date: Click or tap to enter a date.		
B3. Proposed end date: Click or tap to enter a date.		
B4. Key co-investigators (suggest limit to 5-10)		
	Name and Email	Institution
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
<input type="checkbox"/> Yes. I confirmed that the co-investigators have read and approved the application.		
B5. Does this request relate to any previous data or samples that you have requested from DAC? <input type="checkbox"/> No, this is a first request/this has no connection with previous request(s). <input type="checkbox"/> Yes. <i>Please state request form reference number and specify how they are related:</i>		

B6. Is this request an amendment to a previously approved application by the DAC?

☐ Yes. *Please state request form reference number and specify the key amendments:*



**B7. Datasets requested (tick as appropriate)**

All data provided will be coded (de-identified) individual, record-level data, i.e. data containing surrogate identifiers in place of information that identifies the research participant.

Purpose of request:

☐ Non-Commercial Purposes (not Commercial Purposes, including internal, academic, research and development purposes).

☐ Commercial Purposes (purposes primarily intended to generate revenue and/or profit and the purpose of running a business).

Datasets	Tick as appropriate
Research Phenotype data	<input type="checkbox"/>
<b>WGS genomic data</b>	
Joint-called VCF (SG10K_Health_r5.3)	<input type="checkbox"/>
AC>4 phased joint-called VCF_r5.5 (for imputation only)	<input type="checkbox"/>
SG10K_Health structural variants_r1.4	<input type="checkbox"/>
<b>Epigenetic data</b>	
DNA methylation	<input type="checkbox"/>

Brief justification for the basis of selection. If a subset of the WGS or methylation data is requested (e.g. a specific genomic region), please provide details here.

**B8. Background (~500 words)****B9. Specific Aims (~200 words)****B10. Overview of study design and analysis plan (~500 words)****B11. Key references (up to 10)**

### Section C: Acknowledgment by applicant

I confirm that the information provided above is true and accurate, and I agree to comply with and be bound by NPM terms and conditions, including:

- *I recognize that NPM leadership intends to author a series of primary manuscripts based on the SG10K\_Health Precision Medicine Dataset, and that proposals should not overlap these primary manuscripts.*
- *I agree: i. to ensure data are held securely, used only by the approved, named personnel for the purposes described in the application, ii. to make no attempt to re-identify individuals in the study, iii. to provide an annual report on study progress, and iv. to destroy the data at the end of the study, or when asked to do so by the DAC.*
- *I agree to include NPM co-authors, or an "SG10K\_Health Consortium" author block in publications, as guided by the DAC.*
- *I agree to recognize the support of NPM funding in the Funding section of manuscripts using specific wording provided by the NPM team.*
- *I agree to utilize standardized terminologies and naming conventions (eg "SG10K\_Health") as specified by the DAC, and to reference specific NPM publications if directed. I also agree to indicate the SG10K\_Health data freeze used in their manuscript.*
- *I agree to submit a copy of the manuscript prior to submission for approval by the Data Access Committee. This will include evaluation for intellectual property.*
- *If I substantially change the proposed use of the data (eg change in analysis theme, change in lead or senior authors), I agree to submit a new data access form.*
- *I agree to have my data request listed on a public web-site.*
- *I agree to respect any specific requests made by the DAC in relation to this application.*

---

Signature of Requestor

Name of Requestor:

Designation:

Date:

---

Signature of Institution Representative

Name of Institution Representative:

Designation:

Date:

***In situations where non-compliance to these policies are discovered, the Committee reserves the right to block submission of manuscripts, and may consider contacting relevant journals to raise concerns. In situations where manuscript-related disagreements may arise, publication authors have the right to appeal to the Committee. Committee decisions related to appeals will be regarded as final.***

*Questions related to publications should be directed to the NPM Programme Coordinating Office, A\*STAR (contact\_npc@gis.a-star.edu.sg).*