

国家基因组科学数据中心人类遗传资源数据共享政策

Principle for the Access of Human Genetic Resource Data in NGDC

一、目的 Objective

伴随测序技术的不断发展，海量人类遗传资源测序数据产生。为了有效管理和保护我国人类遗传资源数据，促进数据有序共享与合理利用，国家基因组科学数据中心依据《科学数据管理办法》、《中华人民共和国人类遗传资源管理条例》等规定，制定本数据共享政策。

Rapid developing sequencing technology continuously promotes the huge production of the sequencing data on human genetic resources, which eagerly needs a data sharing policy to effectively manage and protect Chinese human genetics resources for the positively data sharing and rational utilization in China. In this background, the National Genomics Data Center (NGDC) made this policy according to the Method of Scientific Data Management and the Regulation on the Management of Human Genetic Resources of the People's Republic of China.

二、范围和适用性 Applicability

本政策主要适用于通过公共资金资助获取的“人类遗传资源信息”的收集和共享，具体包括测序数据信息及相关元数据信息。GSA-Human 数据库可承担人类遗传资源信息的集中管理和共享发布。GSA-Human 中数据共享分为“开放共享”和“受控共享”两种模式。“开放共享”的数据可以被研究人员直接下载使用，“受控共享”的数据需研究人员向数据递交者“申请”使用，并经审核通过后方可获得数据使用权限。

This policy is mainly applicable to the submission and access of human genetic resource information funded by public grants, including sequencing data information and related metadata information. The GSA-Human Database of NGDC is designated for the collective management and access of human genetic resource information. There

are two modes of data access in GSA-Human, namely Open Access and Controlled Access. The Open-Access data can be downloaded and used directly, while the Controlled-Access data can only be acquired after applying to the original data submitter and getting authorized.

三、 向 GSA-Human 提交数据 **Data Submission to GSA-Human**

3.1 数据提交者的权利 **Rights of data submitter**

1. 数据提交者对符合伦理原则采集的我国人类遗传资源原始测序数据和相关元数据信息有指定数据使用限制的权利。

The data submitter has the right to restrict the specific usage of the submitted raw sequencing data and related metadata of China's human genetic resources acquired in accordance with ethics.

2. 数据提交者在其数据集关联研究成果已发表或相关知识产权已经获得的情况下，仍可以设定数据集共享和释放的时间，具体释放条件由数据提交者所在数据管理委员会（DAC）协商讨论后确定。

The data submitter can set the time of dataset access and release, even if the related research result has been published or the intellectual property has been obtained. The detailed release conditions should be decided with consensus of the data administration committee (DAC) of the data submitter.

3. 数据提交者及其数据管理委员会（DAC）有“同意”或“拒绝”数据申请者对其数据使用请求的权利。

The data submitter and the DAC have the right to Agree or Disagree for the applicants of the data access and usage.

3.2 数据提交者的责任 **Responsibilities of data submitter**

1. 数据提交者应为其递交数据集的责任人，故应为研究组的负责人。数据集在提交至 GSA-Human 前应确保具备以下条件：（1）已经从数据集对应的研究对象处获得知情同意书（应当事先告知人类遗传资源提供者采集目的、采集用途、

对健康可能产生的影响、个人隐私保护措施及其享有的自愿参与和随时无条件退出的权利[1]), 并符合伦理原则, 通过相应的伦理审查; (2) 数据集的提交和共享已经获得研究对象的书面同意并获得提交者所在机构伦理审查委员会批准。

The data submitter is the person responsible for the dataset submitted, which should be the group leader. Before submitting to GSA-Human, the following conditions must be reached: (1) the informed consent has been achieved from corresponding research objectives (with inform of the collection purpose and usage, possible influence to health, privacy protection measure, and the right to participate voluntarily and quit unconditionally [1]), in accordance with ethics and has passed ethic review; (2) the dataset submission and access has obtained the written agreement of corresponding research objectives, and has been approved by the ethics review committee of the data submitter's organization.

2. 数据提交者提供的信息必须对人类遗传资源提供者的全部或部分个人信息进行替换或删除。

The information provided by the data submitter must replace or delete the complete or partial personal information of the human genetic resource provider(s).

3. 数据提交者必须按照 GSA-Human 在线系统要求, 填写所有必要的元数据信息并完成所有提交步骤, 确保填写信息的全面、完整、真实、准确。同时, 提供数据集以及任何必要的附属数据 (如数据集的质量控制信息等)。

The data submitter must fulfill all the necessary metadata information and complete all the submission steps online as required in the GSA-Human webpage, and make sure that the information is complete and accurate. The dataset and any necessary appendix (e.g. the quality control information) should be provided.

4. 提交至 GSA-Human 的数据默认为受控访问数据类型, 数据提交者需组织并成立数据管理委员会 (DAC), 并指定相应的联系人。数据管理委员会应及时对数据申请者发出的数据使用请求进行回复和处理。

The default mode of the data submitted to GSA-Human is Controlled Access. The data submitter should organize a DAC and designate corresponding person. The DAC should response the applications for the data access and usage in a timely manner.

5. 数据递交者在对外发布其数据集时，应根据《中华人民共和国人类遗传资源管理条例》的规定，获得相关数据集在“科学技术部政务服务平台”中“开放使用”类别的备案号。

For the release of the dataset, the data submitter should obtain the record number of the related dataset under the “Open Access” Category in the Administration Service Platform of the Ministry of Science and Technology of the People’s Republic of China, which is required by the Regulation on the Management of Human Genetic Resources of the People’s Republic of China.

6. 如果有发现相关数据集违反《科学数据管理办法》、《中华人民共和国人类遗传资源管理条例》等规定，或被发现是故意或过失包含缺陷或错误的数据，数据提交者应主动与国家基因组科学数据中心联系并暂停发布相关数据共享。数据提交者必须在一定时间内重新上传符合规定和修正后的数据。如果没有在规定时间内完成相应操作，国家基因组科学数据中心将取消原有数据编号。

In case that the related dataset is inconsistent with the legal regulations, or any data with intentionally or unintentionally error is found, the data submitter should contact NGDC immediately and pause the related data access and release. The data submitter must re-upload the data in accordance with the legal regulations or the corrected data in a certain time, otherwise NGDC will cancel the original data numbers.

7. 当已经提交的数据集对应的研究对象选择拒绝共享数据或中途选择退出所参加的研究项目时，数据提交者需及时通知国家基因组科学数据中心，同时，数据提交者必须及时处理数据，并避免数据的后续使用。

In case that the corresponding research objectives of the submitted dataset refuse to the data access or quit the ongoing research program, the data submitter should inform NGDC and handle the data in time, and stop the subsequent use of the data.

3.3 数据提交主要流程 Process of data submission

1. 数据提交者需确认上述提及的“数据提交者的责任”都可以正常履行。

The data submitter should make sure that the aforementioned “responsibilities of data submitter” are applicable.

2. 数据提交者应为该数据集所在项目的研究组负责人，数据提交者需先在 GSA-Human 注册账号并填写研究组负责人相关信息。

The data submitter should be the group leader of the related program, and should first register an account in GSA-Human and fulfill the related personal information.

3. 数据提交者创建项目，填写数据管理委员会相关信息并制定数据共享原则。

The data submitter create a program and fulfill the related information of DAC and the principle of data access.

4. 数据提交者创建个体、研究和序列等元数据信息并上传数据。

The data submitter create metadata information including individual, research and sequence, and upload the data.

5. 国家基因组科学数据中心对数据提交者提交元数据信息和数据进行审核。

NGDC verifies the submitted metadata information and data.

6. 数据审核不通过，数据提交者进行相应修改；数据审核通过，数据归档，发送数据编号给数据提交者。

If the verification is negative, the data submitter should make corresponding corrections and resubmit. If the verification is approved, the data will be archived and the data accession number will be sent to the data submitter.

四、 GSA-Human 数据下载与使用 Data Download and Usage

4.1 数据使用者的权利 Rights of data user

1. 数据使用者可以通过 GSA-Human 系统对所需要的数据发起应用请求。

The data user can make applications for the data needed through the GSA-Human system.

2. 数据使用者在获得数据集的使用许可后，可以在有限期限内完成数据下载。

The data user can download the data in certain time after obtaining the approval of accessing the dataset.

3. 数据使用者使用数据集研究结果获取的知识产权归数据使用者所有。

The data user has the intellectual property of the subsequent research result based on the dataset.

4. 数据使用者可以通过 GSA-Human 指定的 ftp 或 Aspera 等工具对获批数据进行下载。

The data user can download the approved data by tools like ftp or Aspera, which is designated by GSA-Human.

4.2 数据使用者的责任 Responsibilities of data user

1. 数据使用者需在数据申请承诺研究范围内使用下载的数据集，数据集仅限研究目的使用，申请获得的数据集仅限研究组长及其所在研究组使用和研究合作者共同使用，不可以将数据集随意分发他人使用，不可以利用数据集识别个体身份。

The data user should use the downloaded dataset in the promised research scope of application. The dataset is restricted to research purpose, and can only be used for the research group and its research collaborators. Dataset distribution to other person is not allowed and cannot be used to identify individuals.

2. 数据使用者必须自行对下载的数据集质量、内容和科学有效性进行评估。

The data user should verify the quality, content and scientific validity of the downloaded dataset by his/her own.

3. 数据使用者在使用从 GSA-Human 下载的数据集公开发表文章时，必须在文章中标明下载数据集的数据编号并说明数据可以从 GSA-Human 下载获得。

If the data user publish a paper using the downloaded dataset from GSA-Human, the data accession number must be indicated in the paper with the note that the data can be acquired from GSA-Human.

4. 数据使用者必须对使用数据负全部责任（包括对第三方的责任）。数据使用者应了解，在数据管理和处理中出现问题时，其职责将延伸至其所属组织或机构的负责人。

The data user must take full responsibility of the data usage (including third-party

responsibility). The data user should be aware that, in case there are questions of data management and process, the responsibility will be extended to the head of his/her organization/affiliation.

5. 数据使用者必须保障数据集的安全使用。如果发生数据泄露等安全事件, 数据使用者必须立即断开相关设备与网络的连接, 并将事件报告给国家基因组科学数据中心。数据使用者必须按照国家基因组科学数据中心的指示, 立即实施事故后措施。

The data user must guarantee the safe use of the dataset. In case there are incidents of data divulgence, the data user must cut off the network connection of related device immediately and inform NGDC, and then conduct remediation measures following the instructions of NGDC.

6. 如果使用者申请使用的数据集对应研究对象中途退出研究, 数据使用者应立即停止使用数据集, 并在此后也不得再使用此数据集。

In case that the corresponding research objectives of the downloaded dataset quit the ongoing research program, the data user should stop using the dataset immediately and should not use it hereafter.

7. 数据使用者应同意国家基因组科学数据中心在某些数据库统计信息中公开使用其信息。可公开的数据使用者信息主要包括所用数据的数据集 ID、数据使用者名称、数据使用者所在单位信息、数据使用期限、研究题目等。

The data user should be aware that NGDC may use some public information for database statistic purposes. The public information includes by not limited to the dataset ID, name of the data user and affiliation, period of data usage, and the research program.

4.3 数据申请与下载主要流程 Process of data application and download

1. 数据申请者需为研究组负责人, 在 GSA-Human 注册账号并填写研究组负责人相关信息。

The data user should be the group leader, and should first register an account in GSA-Human and fulfill the related personal information.

2. 登录注册账号, 填写数据下载申请。如果是来自不同单位的多个研究员

或教授开展协作研究，必须每个单位单独提出数据申请。

The data user logs on to the system and fulfill the data download application. If there are plural group leaders from different organizations for joint research, each group leader must apply respectively.

3. 数据申请通过系统提交后，GSA-Human 数据管理员将审核数据申请，通过审核的数据申请将递送至数据集所在数据管理委员会联系人。

The data application will be verified by the GSA-Human data manager, and will then be sent to the corresponding person of the DAC of the dataset.

4. 数据管理委员会联系人将组织委员对数据申请进行审核讨论，审核通过信息将通过系统传递给 GSA-Human 管理员和数据申请者。

The corresponding person of the DAC should organize the committee members to review the application, and then send the review result to the GSA-Human data manager and the data user.

5. GSA-Human 数据管理员在获得数据管理委员会的审核通过通知后，协助用户完成数据集及其相关元数据信息的下载。

After receiving the approval notice of DAC, the GSA-Human data manager should help the data user for the download of the dataset and related metadata information.

6. 数据下载期限一般为三个月，数据申请者应在下载期限内完成数据下载，逾期的下载将无法进行，必须重新提交数据申请。

The valid period of data download is three months in general. The data user should complete the data download within the valid period, otherwise the download will be unavailable and must be re-applied.

五、 该政策的修订流程 **Revision Process of This Principle**

5.1 修订建议 **Submission of revision advice**

为数据提交者、数据使用者更好的共享和合理有效利用 GSA-Human 数据，所有用户可以随时向国家基因组科学数据中心提出本政策的修订建议。具体信息可发送至 gsa@big.ac.cn。

To help the data submitters and users to access and use the GSA-Human data more conveniently and effectively, all users are welcome to submit revision advice of this principle to NGDC. Detailed information can be sent to gsa@big.ac.cn.

5.2 修订建议的审核 Review of revision advice

国家基因组科学数据中心 GSA-Human 工作组在收到政策修订建议后，会立即组织召开建议内容审核会议，并决定该建议是否应该被通过、被否决或应加以修改后通过。

After receiving the advice, the GSA-Human work group of NGDC will hold a review meeting and discuss the advice, and then make decision including accept, reject, or accept after revision.

5.3 发布和实施修订 Release and conduction of revision

一旦决定修订的细节，国家基因组科学数据中心 GSA-Human 工作组会在网站上及时公布，并尽快开始实施。修订后的准则可追溯适用于修订前已经获得批准提交数据或使用数据的单位或个人。

Once the revision details are approved, the GSA-Human work group will publish it on the NGDC website and conduct in a timely manner. The revised principle can be applicable to the organizations or individuals that had been approved for data submission or usage before the revision.

[1] 《中华人民共和国人类遗传资源管理条例》，2019年3月20日国务院第41次常务会议通过，自2019年7月1日起施行。

Regulation on the Management of Human Genetic Resources of the People's Republic of China. Approved by the 41st Standing Meeting of the State Council of PRC on March 20, 2019, and has been implemented from July 1, 2019.