Answer ALS: Data Sharing Policy

Answer ALS: Data Sharing Policy

General Principles

Answer ALS (AALS) is a project dedicated to developing and implementing a unified strategy to identify new therapeutic targets for Amyotrophic Lateral Sclerosis (ALS). We achieve change through unifying our community toward agreed-upon goals in research, science, technology and education. Implicit in the establishment of the Answer ALS data and cell repositories, is the view that scientific progress in this area will be greatly enhanced if the data produced by these studies are readily available to all investigators in the research community.

Collection and Submission of Data

Clinical Data:
NeuroBANK will provide Answer ALS with a copy of de-identified participant data once per month for use and distribution through the Answer ALS data portal. Answer ALS will validate and curate the data before each new release to ensure no identifiable information is present and that no errors occurred during the copy process.

OMIC data:
OMIC data generation labs will provide QC'd data, levels 1-4, levels dependent on assay type, to Answer ALS through encrypted channels. These data sets will placed in "Staging" for internal groups to QC and curate before releasing to the public. Once QC'd, the data will be pushed to production in Azure.

Data Storage

Clinical Data:
Master data will be stored within the NeuroBANK system with a limited set of de-identified copy data transferred and stored on Azure.

OMIC Data:
Master data is kept with each OMICs facility for levels 1, 2, 3 and 4 as applicable by assay. A copy is transferred to Azure through a secure file transfer system.

Data Access, Availability, and Use

OMIC dataset access will be provided to research investigators who, along with their institutions, have certified their agreement with the expectations and terms detailed below. It is the intent of the Answer ALS Consortium that approved users of AALS-provided datasets recognize any restrictions on data use delineated within the original informed consent agreements of contributing studies, as identified by the submitting institutions and stated on the AALS and other study websites. Minimal demographic data will be available through the data portal and acknowledgement of each data generation site will be present. In addition, a de-identified metadata set will be available to researchers that register with the data portal.

Data Standards

To ensure that AALS data are useful and interpretable, the Answer ALS Consortium is committed to developing and applying appropriate metadata standards, reagents registries etc. Whenever feasible, these standards and registries are based on existing or emerging community standards (e.g. LINCS Project or other NIH Data Projects). Refer to Answer ALS Data Level Definitions on the Answer ALS Neuromine site.

Data QA/QC Process

AALS Institutes are committed to data quality and reproducibility, particularly for types of data for which agreed upon standards do not yet exist. The Answer ALS Consortium therefore offer no warranty associated with the data, we seek indemnification from its use and that the interpretation that investigators assign to their work are solely their responsibility.

Data Availability

Consortium Data Availability. All Answer ALS data, defined as all project generated data, including genomic, transcriptomic, epigenomic, proteomic and clinical data, should be made available to all AALS consortium investigators as soon as practical.

Collaboration Study Availability. All Answer ALS data, defined as all project generated data, including genomic, transcriptomic, epigenomic, proteomic and clinical data, should be made available to all AALS collaborative institutions after all data has been QC'd and validated.

Public Data Availability. All Answer ALS data will be made available to registered users once data has been appropriately QC'd and curated. This accessibility will be through a Data Access Request (DAR) form on the Neuromine data portal directly.

a. Level 1 immutable whole genome sequencing data will be released upon user registration via a special request by emailing a request to terri@answerals.org.

b. Any user requesting access to controlled data will be required to fill out a request form for Data Access Request (DAR), which is granted by an AALS Data Access Committee (DAC). DACs review and approve or disapprove all requests based on whether the request conforms to the specification within the program specific requirements or procedures (if any). Once approved by the Data Access Committee (DAC), access will be granted to the data requestor.

c. Controlled data will not be shared with 3rd parties without review and permission by the DAC.

Computational Workflows. To make our data analysis transparent, reproducible and accessible to the scientific community computational workflows will be available publicly through an open platform such as GitHub Platform. If they are not available, the researcher can contact Answer ALS to ask for specific pipelines.

Terms of Access

The Requester agrees that if access is approved, the Principal Investigator and designated individuals named in the Data Access Request (DAR) shall become Approved Users of the requested dataset(s). Research use will occur solely in connection with the research project described in the DAR, which includes a description of the research objectives and design. New uses of these data outside those described in the DAR will require submission of a new DAR; modifications to the research project will require submission of an amendment to this application.

The Requester and all Approved Users may use the dataset(s) only in accordance with the parameters approved for the appropriate research use, and any limitations on such use, of the dataset(s) and as required by law.

Research access to the requested dataset(s) is granted for a period of four (4) years as defined below.

Contributing Consortium Investigators to the Answer ALS data base, or their direct collaborators at the same institution, who provided the data or samples used to generate an Answer ALS dataset and who have appropriate IRB approval, if applicable, for broader use of the data are exempt from the limitation on the scope of the research use as defined in the DAR.

The Requester agrees through the submission of the DAR that the PI named in the DAR has reviewed and understands the principles for responsible research use and data handling of the AALS datasets as defined.

Long Term Data Access – Any Requester needing access longer than four (4) years from access approval must send a request to data@answerals.org with an explanation of extended research proposal. If the proposal differs from the original submission, a new DAR will need to be placed for the new project.

Whenever a recognized long-term public repository exists for data generated by AALS consortium (e.g dbGap, GEO, Chorus or other locations), these repositories are used in addition to any AALS-specific repository. Specific details including locations of the data, access to it, and any “requested” data use restrictions will be listed on the Neuromine website.

Data Use

In general, all investigators are encouraged, to publish results as soon as practical based on AALS data. These results may include, but would not be limited to, integrating AALS data with data from other sources. AALS data are released with the sole restriction that they must be correctly cited so that others can establish provenance and access the original data; the correct citation will be released with each data set and will comprise either a AALS ID reference or a unique AALS identifier. Investigators are encouraged to inform the AALS Project of ongoing integrative analyses in order to minimize duplicative and competitive analyses groups.

Investigators and institutions seeking controlled data from the AALS data repository will be expected to meet data security measures and will be asked to submit a DAR, including a Data Use Agreement (DUA), that is co-signed by the investigator and the designated Institutional Official(s). Data Access Requests should include a brief description of the proposed research use of the requested AALS dataset(s). Within a Data Access Agreement, investigators will agree, among other things, to:

- Use the data only for the approved research;
- Protect data confidentiality;
- Follow appropriate data security protections;
- Follow all applicable laws, regulations and local institutional policies and procedures for handling AALS data;
- Not attempt to identify individual participants from whom data within a dataset were obtained;
- Not sell any of the data elements from datasets obtained from the AALS data repository;
- Not share with individuals other than those listed in the request any of the data elements from datasets obtained from the AALS data repository;
- Agree to the listing of a summary of approved research uses within the AALS data repository along with his or her name and organizational affiliation;
- Agree to report, in real time, violations of the AALS policy to Answer ALS;
- Acknowledge the AALS policy with regard to publication and intellectual property.

Answer ALS: Data Sharing Policy

Data Access Committees or their designees will review requests for access to determine whether the proposed use of the dataset is scientifically and ethically appropriate and does not conflict with constraints or informed consent limitations identified by the institutions that submitted the dataset to the AALS data repository. In the event that requests raise concerns related to privacy and confidentiality, risks to populations or groups, or other concerns, the DAC will consult with other experts as appropriate.

Biosample Use

Biosamples collected and used to generate AALS data, were obtained through patient consent. Cell lines generated from these samples are available to researchers through the and Cedars-Sinai Biomanufacturing Center and biosamples are available through the Northeast Amyotrophic Lateral Sclerosis (NEALS) Sample Repository. The general principles for sharing these resources are listed below:

- Biosamples collected and cell lines generated for the AALS study from consented participants and any metadata associated with those samples should be made immediately available to all AALS consortium investigators.
- Biosamples and Cell lines generated for the AALS study will be made available to the scientific community through the Cedars-Sinai Biomanufacturing Center.

Data Sharing

Answer ALS has been funded to create a controlled publicly available resource that makes it possible for authorized users to make new discoveries as rapidly as possible. We therefore agree to comply with the Fort Lauderdale Principles with respect to the dissemination and unrestricted re-use of AALS data and signatures. The Fort Lauderdale Agreement is a public declaration of scientists and professionals supporting the free and unrestricted use of genome sequencing data by the scientific community before that data is used for publication. The Agreement is considered to be a founding document of open-access data publishing and has been adopted by funding agencies including the National Institutes of Health.

The Requester and Approved Users agree to retain control over the data and further agree not to distribute data obtained through an approved Data Access Request to any entity or individual not covered in the submitted Data Access Request. If Approved Users are provided access to datasets for inter-institutional collaborative research Data Access Request, and all members of the collaboration are also Approved Users through their home institution(s), data obtained through a Data Access Request may be securely transmitted within the collaborative group. All data security practices, and other terms of use defined in Answer ALS Data Sharing policies should follow guidelines as described in the *Answer ALS Data Security Policy Principles* listed in the definitions portion of this document, for both the raw and the derived data, including any transmission of the data.

Answer ALS: Data Sharing Policy

The Requester and Approved Users acknowledge responsibility for ensuring the review and agreement to the terms and the appropriate research use of AALS data by research staff associated with any approved project, subject to applicable laws and regulations. AALS datasets obtained through this Data Access Request, in whole or in part, may not be sold to any individual or organization or party at any point in time for any purpose without explicit permission from the Answer ALS Executive Director in writing.

Approved Users agree that if they change institutions during the access period, they will submit a new Data Access Request in which the new institution agrees to the Answer ALS data use policy before data access resumes. Any versions of data stored at the prior institution for the approved use will be destroyed and documented per this policy. However, if advance written notice and approval by the Answer ALS Data Access Committee is obtained to transfer responsibility for the approved research project to another Approved User within the same institution the data may not need to be destroyed.

Data Release and/or Publication

Data Publication Policy

1. **Original Data vs. Derivatives:** You may not publish or claim the original data. However, you're encouraged to publish derivatives or findings resulting from the data.
2. **Notification of Publications:** If you produce publications using AALS samples and datasets, please inform us. We would like to cite your work on other data repository websites.
3. **Acknowledgments and Citations:** In your publications and presentations involving Answer ALS Data or Study Materials Results, kindly acknowledge the Answer ALS program, our funders, and the cohorts who contributed data. This recognition is vital for tracking research outputs from our data—a critical measure for our sponsors.

Below are the general Answer ALS Publications Policy guidelines:

- You are not permitted to publish or claim the original data. Derivatives or findings from the data, however, can be published.
- Whenever you use Answer ALS Data or Study Materials in manuscripts and presentations, ensure you acknowledge the Answer ALS program, its funders, and the contributing cohorts with the appropriate language provided below:

Answer ALS Data Use Acknowledgement:
"Data used in the preparation of this article were obtained from the ANSWER ALS Data Portal **(AALS-01184).** For up-to-date information on the study, visit https://dataportal.answerals.org.

Answer ALS: Data Sharing Policy

To Cite Answer ALS Description Paper:
Baxi, E. G. et al. Answer ALS, a large-scale resource for sporadic and familial ALS combining clinical and multi-omics data from induced pluripotent cell lines. Nat Neurosci 1-12 (2022) doi:10.1038/s41593-021-01006-0.

Answer ALS Cohort Acknowledgements:
"Clinical data and biosamples used in the preparation of this article were obtained from the Answer ALS Foundation Program, 'Answer ALS'. For up-to-date information on the program, visit https://www.answerals.org.

Data Release

Data will be released on periodically determined basis until all data has been generated.

Data Centers will release primary and processed data on a periodic basis. AALS data will be released as soon as logistically possible after QA/QC has been completed. The concept of data levels is explained on the Neuromine website under Data Information tab and follows principles used in other projects like the NIH LINCS Project (Library of Integrated Network-based Cellular Signatures). For some data types, data levels have already been established and quality assessment metrics exist. Data Centers will therefore conform to these standards.

All AALS data-types will have a suitable hierarchy defined from raw data, to processed data, to signatures. In this "data-level" hierarchy raw data is data level 1 and signatures the highest data level (typically data level 4). However, the number of data levels differs for different data types and the definitions are in active development. These data levels will be made available at the time the data are posted.

Data Destruction

The media on which the Protected health information (PHI) or other Answer ALS Controlled information is stored or recorded is to be destroyed in one of the following ways:

- Paper, film, or other hard copy media have been shredded or destroyed such that the PHI or other controlled information cannot be read or otherwise cannot be reconstructed. Redaction is specifically excluded as a means of data destruction.
- Electronic media have been cleared, purged, or destroyed consistent with NIST Special Publication 800-88, Guidelines for Media Sanitization such that the PHI and/or other controlled information cannot be retrieved. Actions of destruction should be logged in a "destruction log" for accountability.

Answer ALS: Data Sharing Policy

IP Policy

Achieving maximum public benefit is the goal of data distribution through the Answer ALS data repositories. The AALS Program believes that these data should be considered as pre-competitive and requests Approved Users not to make IP claims derived directly from the Answer ALS dataset(s). However, the AALS Program also recognizes the importance of the subsequent development of IP on downstream discoveries, especially in therapeutics, which will be necessary to support full investment in products to benefit the public. The AALS Program requests that the Requestor and all Approved Users acknowledge that they will not claim IP on any of the datasets made available through the data portal. However, it does support IP development based on derivative discoveries made from these data.

It is expected that these AALS-provided data, and conclusions derived from the data, will remain freely available without the requirement for licensing. The AALS Consortium encourages the broad use of Answer ALS datasets and a responsible approach to the management of intellectual property derived from downstream discoveries.

Answer ALS: Data Sharing Policy

Definitions

Answer ALS Data
All data and metadata generated by the Answer ALS consortium. This includes but not limited to raw binary data generated from assays, processed data and clinical and experimental metadata.

Answer ALS Consortium
Answer ALS, a group of organizations (“Consortium”) organized to provide individualized ALS Therapeutic Initiatives (refer to “Master Agreement” for further details):

- Cedars-Sinai Medical Center (“CSMC”)
- Emory University (“Emory”)
- Massachusetts General Hospital/Harvard University (“MGH”)
- Massachusetts Institute of Technology (“MIT”)
- The Ohio State University (“OSU”)
- University of California Irvine (“UCI”)
- Gladstone Institute
- Washington University
- Texas Neurology (“TexasNeuro”)
- Northwestern University (“Northwestern”)
- Johns Hopkins University

Controlled Data
Controlled data includes low-level genomic sequencing data, germline variants, raw RNA-Seq, raw ATAC-Seq, all processed data derived from these data and clinical data elements. Access to controlled data is granted by program-specific Data Access Committee.

Data Access Request (DAR)
A request from potential user of the generated data to the AALS Consortium. This is contained in the Data Access Agreement and a PDF of the request can be sent by email to xxxx@answerals.org.

Data Access Agreement (DAA)
A document that states the terms of access such as research use, institutional and PI recognition of appropriate handling of datasets (data security), IP rights and publication policy. This is a signed agreement between AALS, PI and their institution.

Answer ALS: Data Sharing Policy

Data Security Policy Principles

The general guidelines outlined below aim to assist organizations in creating and executing a robust security strategy. Organizations managing private medical and genetic information should, at the very least:

- Focus on earning the trust of those involved by putting their needs at the forefront when evaluating and tackling data security concerns. The individuals involved are the cornerstone of all research initiatives.
- Acknowledge the fast-paced changes in the realms of security, healthcare, and technology. Therefore, organizations should embed security into their organizational culture and services, ensuring the adaptability and upgradability of security measures.
- Aim to maintain the reliability of the data so that it can be depended upon by everyone involved, including researchers, physicians, and other scientific professionals.
- Assess critical vulnerabilities and create plans to manage these risks, without hindering the progress of scientific research.
- Offer all involved parties in the institution straightforward guidelines and an open outline of security protocols.
- Implement protective measures to safeguard data, without using these as grounds to restrict individuals from accessing their own information or limiting its proper use in research.
- Exercise due diligence in minimizing the risk of exposing personal information, while keeping everyone informed of any security incidents to sustain long-term trust.
- Openly discuss challenges and triumphs in the realm of security to promote collective learning among organizations.

De-identification

De-identification is the removal of identifying information (such as name, date of birth, address, social security number) from a dataset so that information is not directly or indirectly linked with specific individuals. De-identification is a significant technical control that Answer ALS organizations should employ where appropriate, that can help protect the privacy of a participant. However, no de-identification process guarantees that individuals can never be re-identified. Therefore, Answer ALS organizations should not rely on de-identification alone as a security control or as a privacy protecting technique.

Answer ALS: Data Sharing Policy

Data Access Committee (DAC)
A body of one or more named individuals whom are responsible for data release to external requestors based on consent and/or National Research Ethics terms. A DAC is typically formed from the same organization that collected the samples and generated any associated analyses.

Data In-Motion
Data is “in-motion” when it is transferred through communications media, ranging from physical cables and cords to wireless communications like Bluetooth and Wi-Fi. Information in the process of being exchanged between internet users is considered “in-motion.”

Data At-Rest
Data is “at rest” when it is not being read, written to, or sent over a network by a user, and is stored physically in a digital form like a hard drive.

Encryption
A cryptographic process of encoding data in a way that unauthorized parties cannot access the data.

Encryption Key
An encryption key is used in an encryption process to render data unreadable. To access the data, a user needs the encryption key to decrypt the data to make it readable again. Keys can be password-based, biometric-based, or randomly generated (e.g., certificates).

Process Data
Data that is processed and generated from raw/controlled data. Usually some kind of cleaning, transformation are performed to convert the raw data into a format that can be analyzed, visualized. File formats include but not limited to CVS, TXT, XLSX, TSV.

The Fort Lauderdale Agreement
The Fort Lauderdale Agreement is a public declaration of scientists and professionals supporting the free and unrestricted use of genome sequencing data by the scientific community before that data is used for publication. Along with the Bermuda Principles, the Agreement is considered to be a founding document of open-access data publishing and has been adopted by funding agencies including the National Institutes of Health.