Alzheimer's Disease Data Initiative

Global Neurodegeneration Proteomics Consortium (GNPC) Data Challenge Rules & Guidelines

About GNPC

The Global Neurodegeneration Proteomics Consortium (GNPC) is a major neurodegenerative disease biomarker discovery effort. Launched in 2023, the consortium is uniting and expanding the available molecular "fingerprinting" data for thousands of patient samples across dozens of different dementia and population cohorts, including healthy aging, Alzheimer's disease (AD), Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS), and frontotemporal dementia (FTD).

By uniting diverse data and researchers working on neurodegenerative disorders globally, GNPC seeks to contribute to the advancement of the diagnosis and treatment of conditions impacting over 57 million people in need. To date, the GNPC has gathered over 40,000 patient samples from over 20 international research groups to create the V1 Harmonized Data Set (HDS), comprising nearly 300,000,000 unique protein measurements. The HDS spans multiple neurodegenerative disorders including Alzheimer's disease, Parkinson's disease, ALS and FTD. This represents the world's largest protein biomarker discovery effort for neurodegenerative diseases, including proteomics data from some of the largest and most highly characterized clinical cohorts in the world.

The following five studies are members of GNPC and have contributed samples and participant measures to the GNPC HDS. The cohort-specific data, including additional variables, are shared via AD Workbench and are included as part of the data challenge.

AnswerALS has revolutionized the study of ALS and neurological diseases by using patient-derived stem cells and comprehensive multi-omics data to recreate and analyze motor neurons affected by ALS. The program provides a globally shared, standardized repository of multi-layered datasets that accelerate discovery and sets a new model for collaborative neuroscience research.

The Global Alzheimer's Platform (GAP) study Bio-Hermes offers direct comparisons of data available from many leading AD diagnostic tests, including blood-based biomarkers, digital tests, and cognitive assessments. The study includes 24% of participants from traditionally underrepresented communities.

National Institutes of Aging (NIA)-funded Alzheimer's Disease Research Centers (ADRCs) act as resources in the United States for research on AD and related dementias, housed within major medical institutions across the United States. The Emory and University of Kansas ARDC datasets include clinical, cognitive and proteomic data on study participants enrolled at the ADRCs at these institutes.

The Mayo Clinic Study of Aging (MCSA) is a population-based study centered in Olmstead County, Minnesota. The dataset includes longitudinal information on study participants including blood-based biomarker measurements, outcomes from cognitive assessments, and select PET measures.

About Proteomics

Proteomics is the large-scale study of proteins, their identities, quantities, modifications, and interactions, typically using bottom-up mass spectrometry (MS) or affinity-based assays. In MS workflows, proteins are digested to peptides, spectra are matched to sequences, and peptide intensities are summarized to protein abundance, with care for false discovery rates, missing values, and post-translational modifications (PTMs). Affinity platforms instead quantify proteins by binding reagents; the SomaLogic SomaScan uses chemically modified DNA aptamers (SOMAmer® reagents) that bind target proteins and are read out by DNA quantification, enabling high throughput (~7,000 targets), small sample volumes, and broad dynamic range from serum or plasma.

SomaScan data are stored as relative fluorescence units (RFU), typically log2-transformed, with vendor normalization steps (e.g., hybridization controls, plate calibration) plus downstream modeling (e.g., linear models with age/sex covariates) much like RNA-seq differential analysis. In analysis[1], best practice is to pair rigorous QC[2] (scale factors, outlier checks), adjust for batch and biological covariates, correct for multiple testing, and, where possible, orthogonally validate key findings (e.g., by MS or immunoassay).

The Challenge

The AD Data Initiative is creating new possibilities for over 55 million people worldwide living with Alzheimer's or related dementia. To accelerate novel discoveries in ADRD, this challenge offers researchers to submit their projects utilizing the GNPC's HDS; optionally, participants can add AnswerALS, Bio-Hermes, Emory Goizueta ADRC, KU ADRC and/or MCSA datasets, all exclusively available via AD Workbench. When requesting these datasets for use in the challenge, please include a message in the request so the data approving teams can expedite the review.

Challenge participants must use the HDS that is accessed, with permission, via the AD Workbench. Participants may also use AnswerALS, Bio-Hermes, Emory Goizueta ADRC, KU ADRC, MCSA datasets, other permissioned AD Workbench datasets, or data from other sources as it relates to the research project. They must conduct their analyses on the AD Workbench.

The submitted project must address one or more of the following priority research questions that have been assembled by the AD Data Initiative coalition and other key stakeholders:

- How can protein changes over time help track aging and disease progression? Do protein changes help mark pre-clinical and post-diagnosis AD?
- Do changes in plasma markers influence or underpin speed of progression of disease? Are any markers promising as treatment targets?
- Can we develop prognostic markers of cognitive decline and neurodegeneration?
- How can we leverage biomarker data to classify subtypes of Alzheimer's disease?
- Can proteomics help identify novel ways to delay the onset of disease?
- How can we leverage proteomics for early diagnosis?

Don't see your favorite research question listed? Send us your suggestions by November 1, 2025. We will choose and publish up to two additional questions based upon user input by November 15, 2025.

Challenge entrants need not register for the challenge beforehand. To submit a solution to the challenge, please follow the Submission Guidelines below.

---

[1] R package to analyze SomaLogic SomaScan data on cran: SomaDataIO.
[2] Protocol outlining quality control and normalization of SomaScan assays.

Eligibility

Up to two challenge entries per question will be selected as winners. Due to United States financial restrictions, applicants cannot reside or work for an organization located in Afghanistan, Belarus, Central African Republic, China, Cuba, Democratic Republic of Korea, Democratic Republic of the Congo, Ethiopia, Iran, Iraq, Lebanon, Libya, Mali, Myanmar, Nicaragua, Russian Federation, Somalia, South Sudan, Sudan, Syria, Ukraine, Venezuela, Yemen, or Zimbabwe.

Members of the GNPC consortium, including all researchers who had access to the HDS during embargo, are not eligible for this challenge.

We believe high-quality science needs to include different perspectives, so we are building a diverse and inclusive environment. We encourage people from all cultural, geographical, and technical backgrounds to apply, including those who are from underrepresented backgrounds in their field.

Prizes

Up to sixteen cash prizes will be awarded, two per listed question and up to two per additional question, subject to the restrictions and conditions herein. For each question, the participant entry with the highest aggregate score across the categories will receive $10,000 USD and the entry with the second highest aggregate score will receive $5,000 USD. Upon selection, the winners are expected to additionally provide a draft manuscript and to share with AD Data Initiative all code that was used or developed for the Challenge. Publishing fees will be paid once a manuscript has been accepted by a mutually agreed journal, and code will be shared openly with the community.

Submission Guidelines

If participating in the challenge, submissions must be received by 15:00 PST (23:00 GMT) Thursday, January 15, 2026, using the application portal.

A complete submission will include:

- Profile details (as requested in the online form)
- Name of the AD Workbench workspace where the analysis was performed
- Name(s) of the datasets upon which the analysis was performed
- Which of the priority research questions was addressed
- Project abstract (maximum 200 words)
- Project summary including the following sections and any relevant tables and/or figures (maximum 2 pages, single spaced):
    - Introduction
    - Methods
    - Results
    - Conclusion
    - References (not counted in the 2-page limit)

If selected as a winner, participants must additionally provide a complete set of code used and/or developed for the project, properly documented and cleaned for ease of review, archived and easy to reference and a draft manuscript. Participants give AD Data Initiative permission to share the code via AD Workbench for researchers to copy, redistribute, or utilize under the MIT license.

Applicants can submit entries for up to two questions. If more than two are submitted, the first two entries will be considered, and the others will be removed. During the review period, applicants may be contacted if further clarification is needed.

Selection of Winners

A panel of up to three judges per question will evaluate the Submissions for these categories on a scale from one (low value) to five (highest value):

- Topic Responsiveness: How well does the submitted research address the chosen priority research question?
- Innovative Approach: Does the idea offer an unconventional, creative approach to addressing the chosen research question? Does it leverage existing methods in novel ways, such as integrating diverse data types (e.g., clinical, imaging, and genetic data), and providing interpretable outputs?
- Significance and Impact Potential: Does the proposed research have the potential to generate new hypotheses or challenge existing ones about the biology, detection, diagnosis or treatment of dementia? Does it align with the mission of the AD Data Initiative? Does the proposed research demonstrate a clear pathway to real-world applications, ensuring that the findings can be translated into clinical settings?

Privacy and Confidentiality

All submitted applications will be kept confidential, except: (1) as noted below; (2) as necessary for our evaluation; (3) to comply with any applicable laws; and (4) if the application is made public or available to others through no fault of AD Data Initiative. Unsuccessful applications will remain confidential. Successful applications may be shared publicly after reasonable notice to the applicant to enable the management of the program. Application materials will not be returned to applicants and will be destroyed. For further details about the AD Data Initiative's privacy policies and use of any personal data that may be submitted with your application, please refer to the AD Data Initiative's privacy policy.

Publicity and Marketing

Challenge entrants may be asked to collaborate with the AD Data Initiative on publicity and marketing related to their research, which may use information included in their application. Any information that will be publicly disseminated that specifically references the applicant or their work will be shared with the applicant prior to being shared with others.

No Third-Party Infringement

By applying, the applicant is representing to AD Data Initiative that they have the right to provide the information submitted and the materials do not infringe any third-party privacy or intellectual property rights. Applicants with questions concerning the contents of their application may contact AD Data Initiative at support@alzheimersdata.org.

How to Apply

Provide *all* submission materials by 15:00 PST (23:00 GMT) Thursday, January 15, 2026, using the application portal. Final decisions are expected to be made in March 2026.

Questions

Email questions to support@alzheimersdata.org.