

300 Million and Counting:

How the AD Workbench Powers the World's Largest Protein Biomarker Discovery Effort



The [Global Neurodegeneration Proteomics Consortium](#) (GNPC) v1 was co-funded by Gates Ventures and Johnson & Johnson (J&J) to accelerate drug discovery and improve patient outcomes for the 57 million people worldwide suffering from neurodegenerative diseases. The GNPC brought together dementia researchers from dozens of top research institutions to build the largest protein biomarker discovery effort to date, including 300 million proteins from 40,000 biofluid sample analyses. To facilitate this collaboration, GNPC leveraged the AD Data Initiative's AD Workbench research environment to create, curate, and make available the GNPC v1 Harmonized Dataset (HDS) to qualified researchers around the world.

Why a Proteomics Dataset?

One of the most exciting frontiers in neurodegenerative research is the potential discovery of new biomarkers that can rapidly accelerate the diagnosis and treatment of neurodegenerative diseases such as Alzheimer's Disease (AD), Parkinson's Disease (PD), amyotrophic lateral sclerosis (ALS), and frontotemporal dementia (FTD). But, until recently, biomarker discovery has been slowed by barriers to large data analysis ranging from a lack of data harmonization to challenging data governance rules.

The GNPC – a consortium of two dozen top neurodegenerative research institutions – was created to systematically address these challenges, encourage collaboration, and propel neurodegenerative research forward. The goal was to bring together data and bio-samples from many diverse cohort studies into one large, harmonized proteomics database for the scientific community to generate actionable insights about neurodegenerative disease and transform patient trajectories.

The GNPC v1 HDS

Made available to GNPC members in June 2024 and publicly in June 2025, the v1 HDS includes 250 million unique protein measurements from 35,000 biofluid samples from some of the largest and most highly characterized clinical cohorts in the world. It is the largest biomarker discovery effort for neurodegenerative diseases to date, spanning disorders including AD, PD, ALS, and FTD.

Primary data was generated via the [SomaScan](#) 7k aptamer-based platform, which measures thousands of proteins across plasma, serum, and cerebrospinal fluid samples. With flexible funding support from GNPC partners, sample providers could select either cross-sectional or longitudinal samples to be prioritized. Samples were analyzed and the resulting data was quality controlled and matched to a clinical profile from the time of sample collection.

The GNPC v1 HDS , cont'd

Each contributing cohort was required to provide a minimal set of associated clinical metadata for each proteomic record which included diagnosis, disease stage, neuropsychological assessments, and demographics. These clinical data were harmonized across cohorts by a professional vendor in collaboration with the contributing cohorts and anonymized prior to sharing with the consortium members. Over 50 clinical features are represented in the v1 HDS.

35,000 anonymized proteomic samples analyzed in the v1 HDS	23+ data contributors from the US and EU signed DCAs	100+ concurrent GNPC researchers in the Common Workspace
---	---	---

Powered by the AD Workbench

GNPC worked closely with the AD Data Initiative to develop the HDS v1 to maximize collaboration, data availability, and adherence to all regulatory protocols—all while minimizing costs for users. For example, the AD Data Initiative created a common Data Contributor Agreement (DCA) for GNPC data providers, signed prior to any data transfer, that adheres to all international regulatory guidance on neurodegenerative data. The AD Data Initiative also ensured all data are harmonized at no cost to providers.

Most notably, the GNPC HDS v1 was curated, harmonized, and is now accessible on [AD Workbench](#), the AD Data Initiative's secure, cloud-based, collaborative environment for neurodegenerative research. Within the AD Workbench workspaces, users have access to multimodal analysis tools, and researchers may also bring in their own code, models, and approved external datasets.

Within the AD Workbench workspaces, users have access to multimodal analysis tools, and researchers may also bring in their own code, models, and approved external datasets.

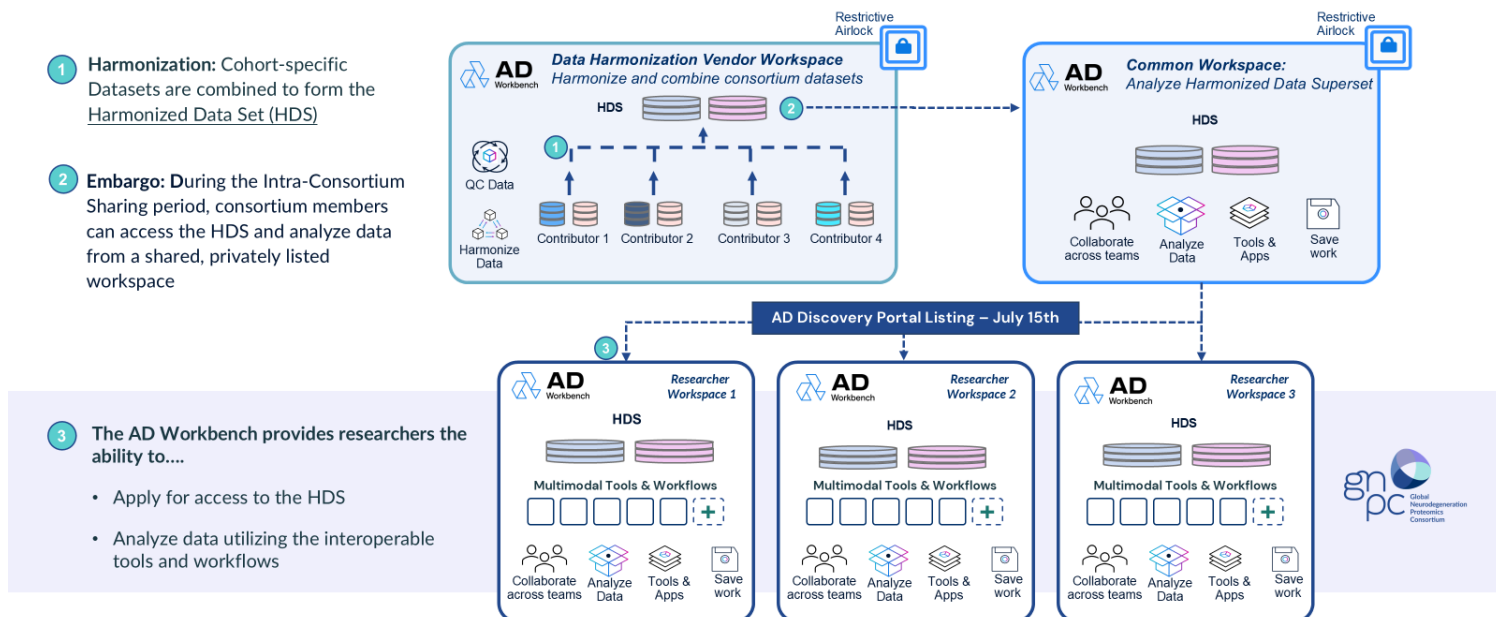
Analysis of the HDS v1 on AD Workbench means access to free compute and virtual machines for neuroscience researchers. It also allowed for a 1-year embargo period, so GNPC members could gain exclusive access without worrying about data security. As of summer 2025, the HDS is now broadly available and discoverable on the [AD Discovery Portal](#), the user-friendly, publicly accessible dataset catalog that helps researchers discover data available via AD Workbench and its interoperability partners. See **Figure 1** on the next page.

Explore the [HDS v1 on the Discovery Portal](#) and [sign up for AD Workbench](#) to enjoy a no-cost, secure, cloud-based data sharing and analytics environment.

Figure 1.

THE GNPC v1 HARMONIZED DATASET IS AVAILABLE ON THE AD DISCOVERY PORTAL, ENABLING PROTEOMIC DATA SHARING AT SCALE.

The three-stage flow shows (1) multi-cohort data harmonization in the harmonization vendor workspace, (2) intra-consortium sharing during the 12-month embargo period, and (3) researcher access to the HDS via individually permissioned AD Workbench workspaces.



GNPC v1, vMP, and Beyond

Upon public release of the HDS v1 data in June 2025, [GNPC published several papers in Nature Medicine](#) detailing the creation of the v1 dataset and the key findings from the one-year embargo period. Preliminary analysis yielded an extensive map of disease-specific differential abundance of plasma proteins across the pan-neurodegenerative disease dataset, as well as transdiagnostic proteomic signatures of clinical risk that can inform predictive models of disease. Researchers also identified a clear APOE ϵ 4 plasma proteomic signature robust to disease type and/or clinical status, and enhanced proteomic aging ‘clock’ measurements across disparate organ systems that are differentially affected by disease type.

GNPC is now working on regular updates to the HDS to keep augmenting its utility for neurodegenerative research. GNPC’s v1 HDS is [now available to researchers](#) through [AD Workbench](#) and approximately 600 users have applied to access it in their own private workspaces during the first 9 months of its release.

GNPC v1 funding partners



Johnson&Johnson

Learn more about [GNPC publications](#) in Nature Medicine detailing the creation of the v1 dataset and key findings from the one-year embargo period.

GNPC v1, vMP, and Beyond, cont'd

Building on the v1 foundation, the **GNPC vMultiplatform (vMP) expansion** – due to be released to consortium members in Summer 2026 and publicly in Summer 2027 – will significantly broaden the scope of the dataset and strengthen the infrastructure supporting it.

Some of the key priorities for this vMP 2026 expansion include:

- multiple proteomic platforms [SomaScan 7k/11k, Olink HT, and Alamar Biosciences (NULISA CNS & Inflammation)], dramatically increasing proteomic depth to ~17,000 independent protein measurements per biosample, and enabling cross-platform comparison studies;
- greater geographic diversity by adding studies based in Africa, Latin American, East Asia, and South Asia;
- interventional cohorts that supplement observational data with studies including behavioral and lifestyle modification and pharmacological interventions such as antibody therapies; and
- new data informatics capabilities, with automated data analysis pipelines via the [Global Research & Imaging Platform \(GRIP\)](#), digital twin modeling, and novel AI/ML model development within the AD Workbench environment.

The GNPC vMultiplatform (vMP) expansion – due to be released to consortium members in Summer 2026 and publicly in Summer 2027 – will significantly broaden the scope of the dataset and strengthen the infrastructure supporting it.

GNPC vMP funding partners



Alzheimer's
Drug Discovery
Foundation



Johnson & Johnson

ROBERTSON
FOUNDATION

sanofi



Shanahan
Family
Foundation

Anonymous donor

If you are interested in discussing whether the AD Data Initiative product suite might help strengthen your research initiatives, please contact us at info@alzheimersdata.org.