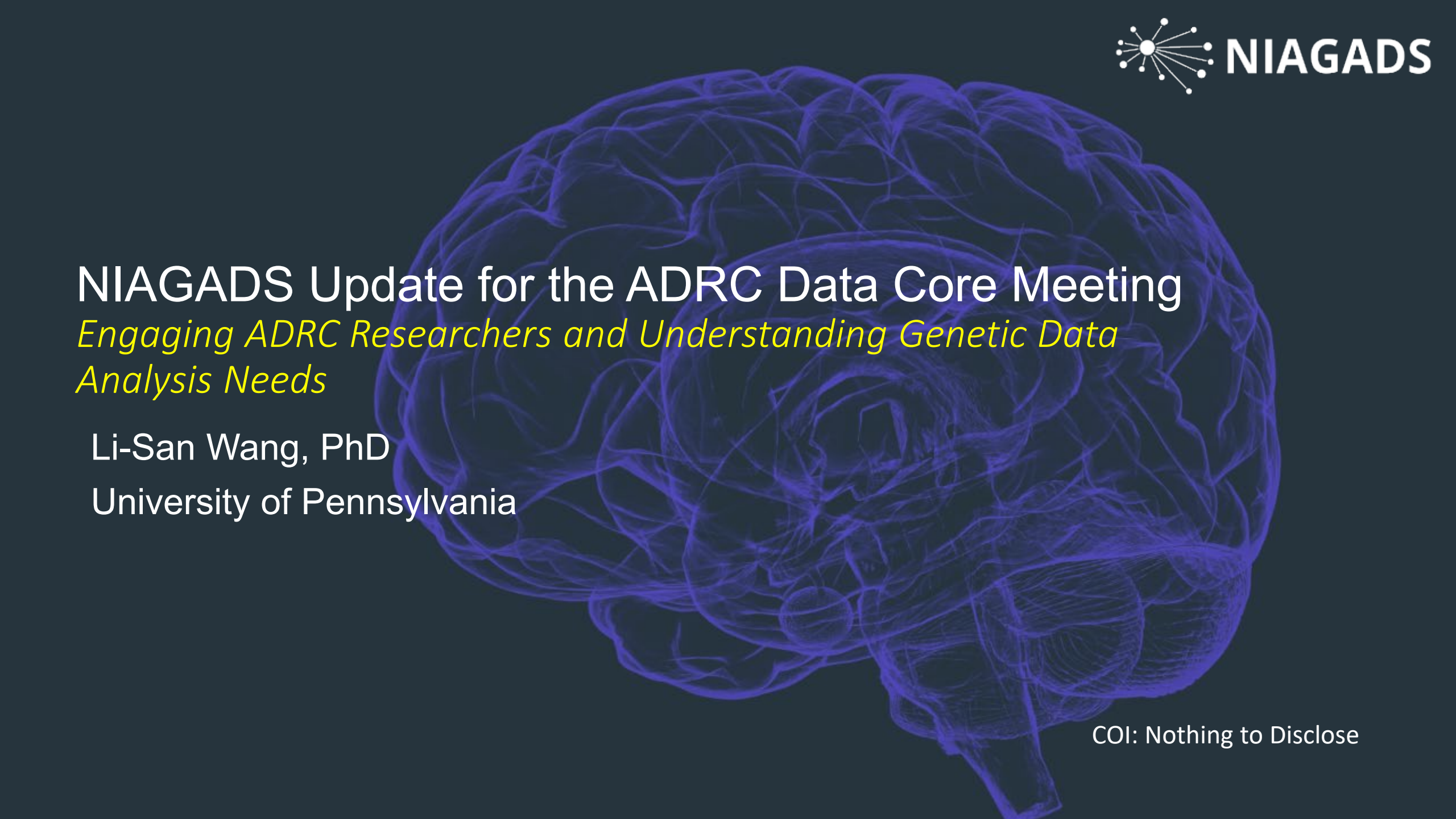


NIAGADS Update for the ADRC Data Core Meeting

Engaging ADRC Researchers and Understanding Genetic Data Analysis Needs

Li-San Wang, PhD

University of Pennsylvania

COI: Nothing to Disclose

RECAP: ADSP 2024 Releases

NIAGADS

Released in 2024

October

Additional files added for R2, R3, and R4 releases

Structural Variant joint genotype calls, multi-allelic / chrX quality control results

Coming Soon in 2024

November

R5 Whole Genome Sequencing for ~24K new samples

Individual sample raw data (CRAMs)

Joint genotype calls preview pVCF across all ~60k genomes released by ADSP

December

Additional cohorts from the **Phenotype Harmonization Consortium** (PHC) released

RECAP: Next ADSP Data Release (Pending Nov 2024)

ADSP R5 – Up to 60K Whole Genomes

ADSP Release	Primary Ancestry/Ethnicity	Cases	Controls	Other	Totals
GCAD Release 5	Black/African American	2,646	5,608	777	9,031
	Asian	1,273	4,083	5	5,361
	Latino/Hispanic	4,488	10,248	2,777	17,513
	Non-Hispanic White	9,124	14,459	6,067	29,650
	Total	**17,531**	**34,398**	**9,626**	**61,555**

ADRC samples in R5

12,629
genomes
20% of total ADSP WGS

4,590
Exomes (previously released)
22% of total ADSP WES

ADRCs can get WGS/WES data from their own samples from us!
Send an email to niagads@pennmedicine.upenn.edu to initiate.

RECAP: Next ADSP Data Release (Pending Dec 2024)

Additional Clinical Data from the Phenotype Harmonization Consortium (PHC)

Cohorts	Cognition	Fluid Biomarkers	Neuro-pathology	Vascular Risk Factors	Imaging - DTI	Imaging - FLAIR	Imaging - T1	Imaging - PET
ADNI			N/A					
WHICAP		N/A						N/A
EFIGA		N/A			N/A	N/A	N/A	N/A
NIA-AD FBS		N/A			N/A	N/A	N/A	N/A
WRAP			N/A			N/A		
ROS/MAP/MARS		N/A						N/A
ACT		N/A			N/A	N/A	N/A	N/A
NACC								
WashU (Knight ADRC)								
A4		N/A	N/A					
KBASE								
TARCC		N/A	N/A					
HABS-HD		N/A	N/A					
Miami Brain Bank	N/A	N/A		N/A	N/A	N/A	N/A	N/A
HIHG Brain Bank	N/A	N/A		N/A	N/A	N/A	N/A	N/A
Case Western	N/A	N/A		N/A	N/A	N/A	N/A	N/A

Coming in December 2024

Harmonization in Progress

Domain data not available

RECAP: ADGC GWAS Data Release

NIAGADS

Leads: Adam Naj & Amanda Kuzma

Batch 1:

- ADRC array datasets (15 datasets)
  - PHC assigned phenotypes
  - ADSP ID mapping

COMPLETE

Batch 2:

- GWAS datasets overlapping with PHC harmonized phenotypes (8 datasets)

IN PROGRESS

Batch 3:

- Other GWAS datasets already deposited into NIAGADS (8 datasets)

Batch 4:

- Other GWAS datasets not yet shared through NIAGADS (many)

Release Features

- All subject IDs standardized using updated schema
- PLINK files converted to forward strand and lifted to GRCh38
- Covariate files updated with consistent variables/formatting
- Standardized file naming schema

Challenges in Preparing Genetics Data for Researchers

Lessons learned from supporting the ADSP Artificial Intelligence and Machine Learning Consortium (AI-ML)

Analysis of imaging and genomic data

- Paul Thompson, **Ultrascale Machine Learning to Empower Discovery in Alzheimer's Disease Biobanks (AI4AD)**
- Degui Zhi, **Genetics of Deep-Learning-Derived Neuroimaging Endophenotypes for Alzheimer's Disease**

Analysis of gene expression data

- Towfigue Raj, **Learning the Regulatory Code of Alzheimer's Disease Genomes**
- Feixiong Cheng, **Alzheimer's MultiOme Data Repurposing: Artificial Intelligence, Network Medicine, Therapeutics Discovery**
- Wei Pan, **Causal and integrative deep learning for Alzheimer's disease genetics**
- Zhongming Zhao, **AIM-AI: an Actionable, Integrated and Multiscale genetic map of Alzheimer's disease via deep learning**

Developing and applying machine learning tools/platforms for AD research

- Olivier Lichtarge, **Cognitive Computing of Alzheimer's Disease Genes and Risk**
- Honghuang Lin, **Assessing Alzheimer Disease Risk and Heterogeneity Using Multimodal Machine Learning Approaches**
- Jason Moore, **Artificial Intelligence Strategies for Alzheimer's Disease Research**

https://adsp.niagads.org/funded-programs/artificial-intelligence-and-machine-learning/

What makes data AI/ML-ready?

What makes data AI/ML-ready?

Ask this question instead - **What issues will cause problems with AI / ML methods?**

What makes data AI/ML-ready?

Ask this question instead - **What issues will cause problems with AI / ML methods?**

Issues	Strategy
Data not properly integrated Lack of interoperability	• Ensure all data are harmonized for interoperability. E.g. annotations/WGS calls are processed for the same genome build

What makes data AI/ML-ready?

Ask this question instead - **What issues will cause problems with AI / ML methods?**

Issues	Strategy
Data not properly integrated Lack of interoperability	• Ensure all data are harmonized for interoperability. E.g. annotations/WGS calls are processed for the same genome build
Data not formatted for AI/ML algorithms File format inoperable	• Use community / industry standard file format • Detailed documentation

What makes data AI/ML-ready?

Ask this question instead - **What issues will cause problems with AI / ML methods?**

Issues	Strategy
Data not properly integrated Lack of interoperability	• Ensure all data are harmonized for interoperability. E.g. annotations/WGS calls are processed for the same genome build
Data not formatted for AI/ML algorithms File format inoperable	• Use community / industry standard file format • Detailed documentation
Data heterogeneity Different recruitment/DX criteria; different cognitive tests	• Document design of cohorts • Go beyond just linking variables - harmonize clinical data by domain experts (e.g. PHC) • Test robustness of harmonization extensively

What makes data AI/ML-ready?

Ask this question instead - **What issues will cause problems with AI / ML methods?**

Issues	Strategy
Data not properly integrated Lack of interoperability	• Ensure all data are harmonized for interoperability. E.g. annotations/WGS calls are processed for the same genome build
Data not formatted for AI/ML algorithms File format inoperable	• Use community / industry standard file format • Detailed documentation
Data heterogeneity Different recruitment/DX criteria; different cognitive tests	• Document design of cohorts • Go beyond just linking variables - harmonize clinical data by domain experts (e.g. PHC) • Test robustness of harmonization extensively
Sample bias Case/control imbalance; age mismatch	• Create virtual cohort for AI/ML analysis based on cohort characteristics; work with domain (AD and epi) experts

Example: Defining FunGen Data Standards

Project lead: Fanny Leung / UPenn

FunGen-AD ADSP Functional Genomics Consortium

adsp Alzheimer's Disease Sequencing Project

GCAD Genome Center for Alzheimer's Disease

SageBionetworks

Meta Data Schema Development

Define **meta-data** for describing various types of experimental data

Define **ontologies/open vocabularies** to be used for describing FunGen data attributes

Define **minimal information for standardization across data sources**: genome build, annotation, ontologies, file formats, etc.

Standardized Fields for Summary Statistics

hipFG[1]

- ✓ Variant position: chr, start, end
- ✓ Allele information: ref, alt, a1, a2
- ✓ Feature name (e.g. gene name, protein name)
- ✓ P-value and/or Q-value
- ✓ Effect size (Beta and Beta SE) or Spearman correlation p-value
- ✓ Detailed sample source, molecular trait and organism; provide protocol details if iPSCs
- ✓ Description of all the columns
- ✓ Software name and version used to perform the analyses

Cifello et al, *Bioinformatics* 2023

Harmonized Genomic Knowledge via NIAGADS Open Access Data and Annotation Resources

unrestricted data and annotations (variant, gene, sequence) available for immediate access

dss NIAGADS

Open Access Portal

direct data downloads for GWAS summary statistics (p-values) and **ADSP Variant Annotations**

https://dss.niagads.org

filer [1]

75,207

standardized, searchable functional genomics tracks

https://tf.lisanwanglab.org/FILER/

advp [2]

curated AD-GWAS catalog, documenting

>18k

published associations

https://advp.niagads.org

GenomicsDB [3]

AD-genetics knowledgebase and **Genome Browser**, containing

>260M

annotated variants

https://www.niagads.org/genomics

ASHG 2024 Platform Talk
Pavel Kuksa

1. Kuksa et al. 2022, *NAR Genomics and Bioinformatics,* 4(1): lqab123. 2. Kuksa et al. 2021, *Journal of Alzheimer's Disease*, 86(1): 461-477. 3. Greenfest-Allen et al. 20204, *Alzheimer's & Dementia,* 20(2):1123-1136

Visualize and Explore Data

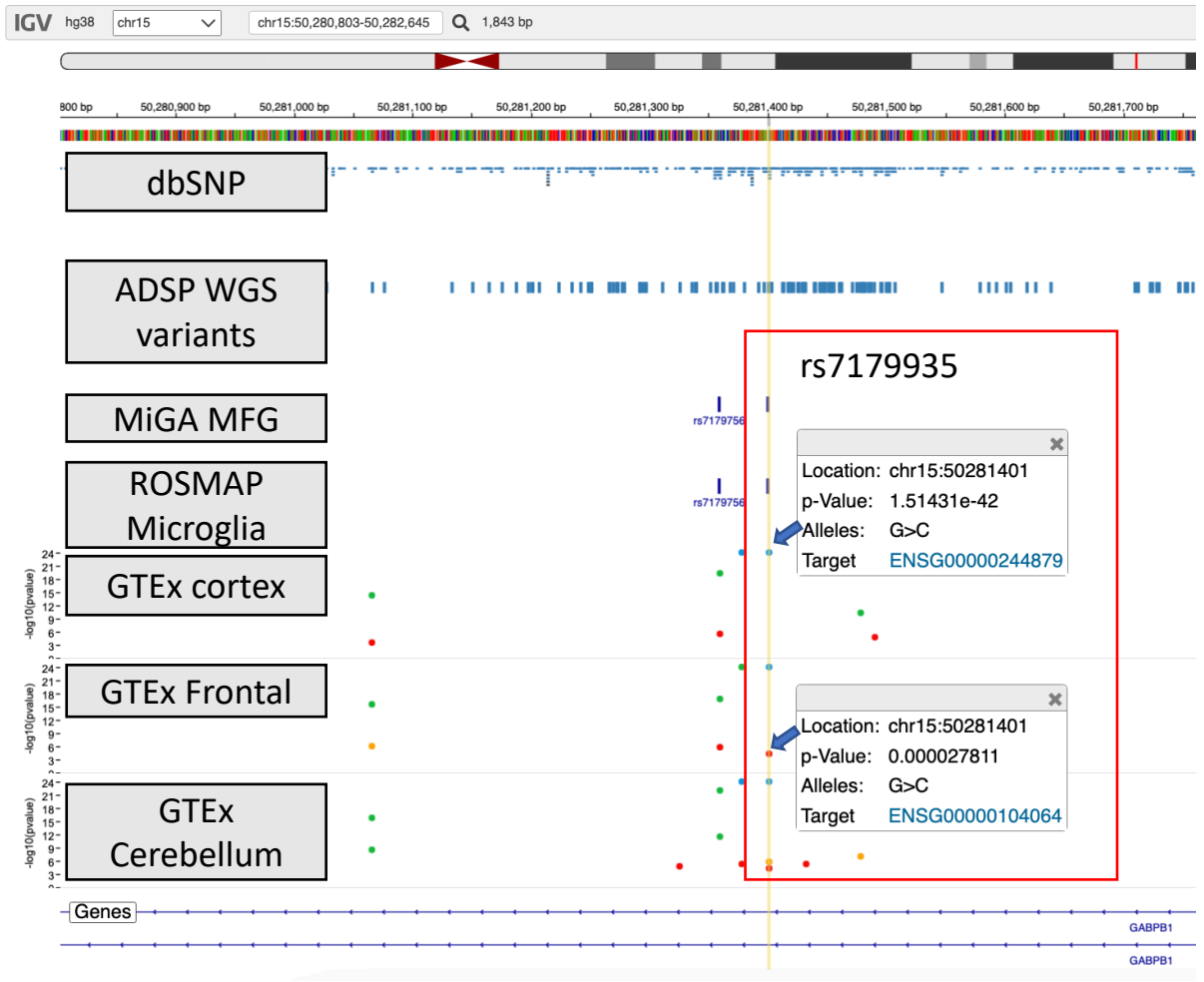
Lead: Emily Greenfest-Allen

FunGen-AD
ADSP Functional Genomics Consortium

NIAGADS

GenomicsDB

Visualize Across Tracks



xQTL Track

Track in FILER /GenomicsDB

Features

- Track selector
- Any QTL datasets
- FILER/user's own input tracks

https://www.niagads.org/genomics/app

1. Kuksa et al. *NAR Genom Bioinform* 2022. 2. Greenfest-Allen et al. *Alzheimers Dement* 2024.

NIAGADS Open Access
NIAGADS API*

Lead: Emily Greenfest-Allen

advp

Alzheimer's Disease Variant Portal

filer

Functional Genomics Repository

dss NIAGADS

Open Access Portal

GenomicsDB

Alzheimer's Genomics Database

API
NIAGADS

- Programmatic access
- Bulk, ranged, and genome wide queries
- Integration into analysis pipelines
- Better support ADSP activities and AI/ML pipelines

conforms to OpenAPI standards (& is FAIR)

* under development; beta-testing version available in May 2024

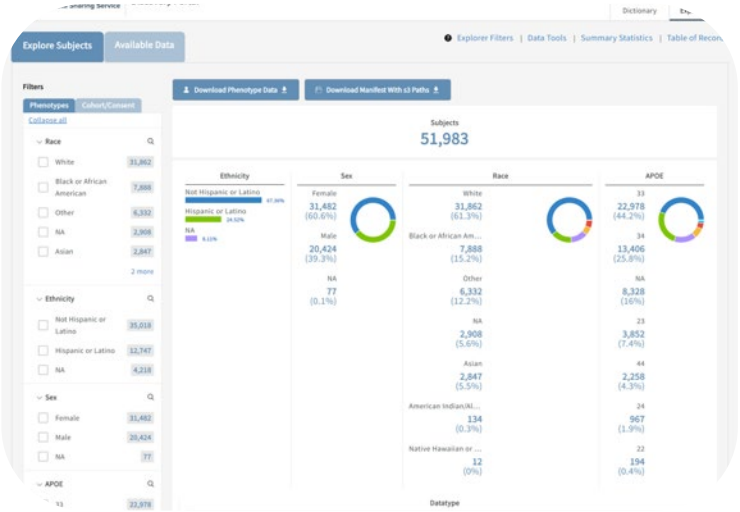
Supporting FAIR Principles

NIAGADS

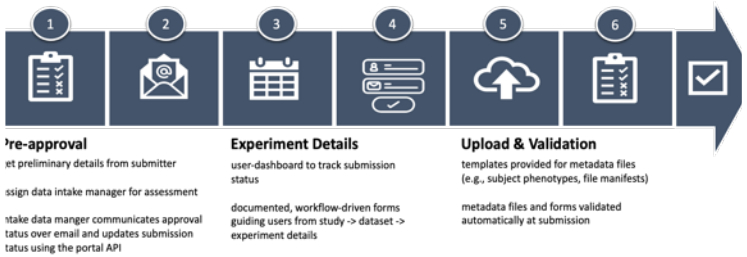
Initiatives	Findable	Accessible	Interoperable	Reusable
Gen3 Platform				
RAS Authentication				
Data Submission Portal				
DOI Implementation				
ADGC GWAS Standardization				
Outreach Initiatives				
NIAGADS Knowledgebases (Alzheimer's Genomics Database, FILER, ADVP, VariXam)				
NIAGADS Open Access				

Supporting FAIR Principles

NIAGADS



ADSP Data Discovery Platform Integration powered by **Gen3** Technology



Streamline data submission and validation using **DSS Data Submission Portal**

DataCite
FIND, ACCESS, AND REUSE DATA

Digital Object Identifiers (DOI) Enabled through **NIH DataCite Consortium** membership

Supporting Data Use

User Support

NIAGADS offers several support resources launched over the last year:

Office Hours

Held at least once a quarter

Video Tutorials

https://www.youtube.com/@niagads1814

NIAGADS Support Portal

All NIAGADS documentation in one place, new informational pages to address user questions, dedicated FAQ page

Ticket Support

Users can now submit a ticket to work on resolving technical issues with the NIAGADS team.

Awareness

Expanding NIAGADS reach through increased investigator awareness was accomplished through informational outreach sessions at AAIC and will continue at ASHG in 2024

Exhibit Hall Booth

Visit us at booth 1023 at ASHG!

Poster Presentations

ADRC workshop at AAIC in partnership with NACC, Sage, and NCRAD

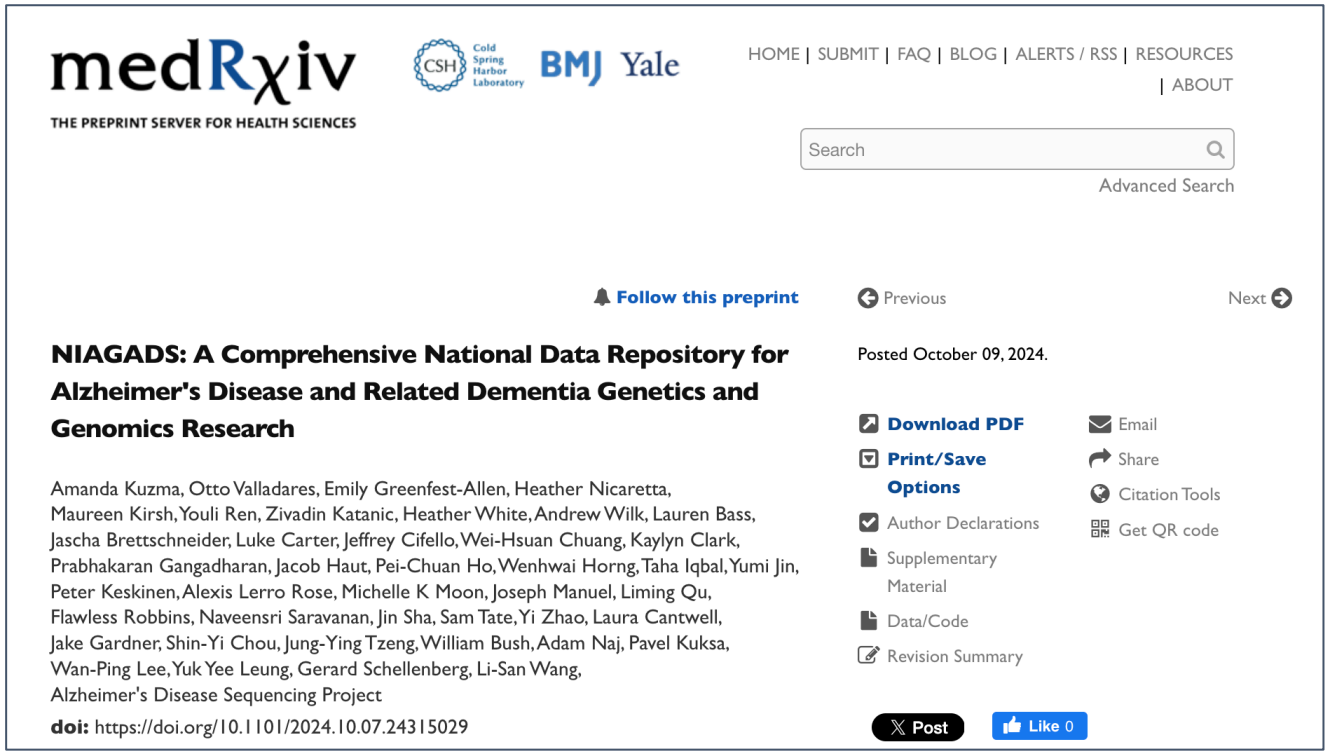
ASHG 2024 Activities

NIAGADS

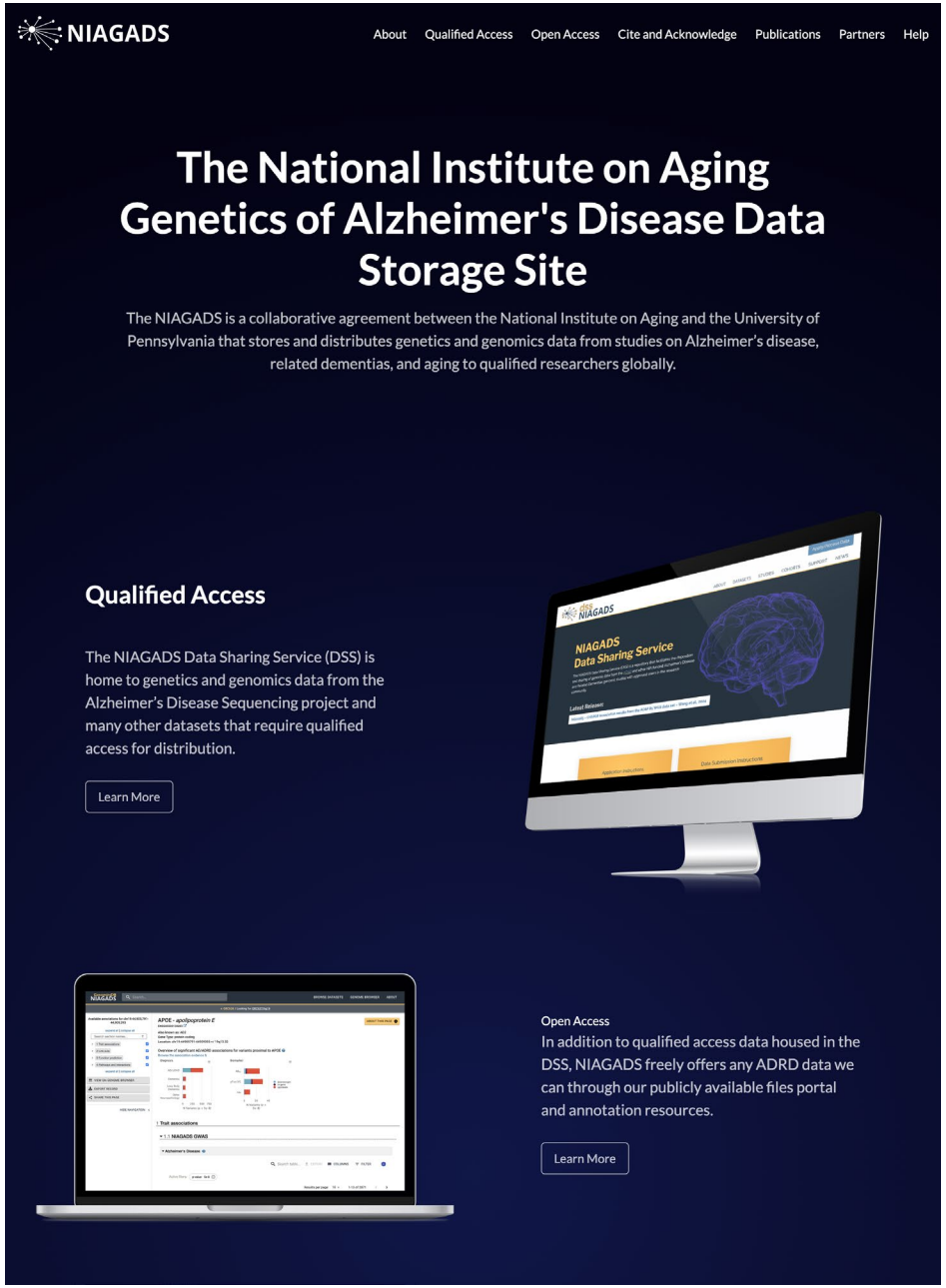
ASHG Exhibit Hall Booth

*Visit us at booth **1023D** at ASHG*

See you this November in Denver!



DOI: 10.1101/2024.10.07.24315029



We are forming a **Genetics Data Workgroup** within the ADRC Program

- NACC and NIAGADS are organizing the workgroup
- Currently in planning stage
- Primary goal: understand ADRC needs in accessing and using genetic data; inform and advise NIAGADS and NACC

Please contact **Sarah Biber** or **Li-San Wang** if you are interested in participating!!