Data Release Date: October 5, 2022, **Dataset Version:** ng00067.v9

Release Information:
This data release includes, 1) new whole-genome sequencing on 19,456 samples joint-genotype called with the R3 16,905 whole-genomes previously released, totaling 36,361 samples, 2) ADSP-PHC harmonized phenotypes for select cohorts in 3 domains: neuropathology, cognitive, and fluid biomarker.

1. **ADSP Release 4 (R4) Whole Genome Sequencing (WGS) Preview.** This ADSP release, containing 36,361 whole-genomes, is referred to as Release 4 (R4). It includes 1) sequencing read alignments in CRAM (compressed BAM) format for the newly sequenced 19,456 samples, (2) genomic Variant Call Format (gVCF) files generated by GATK4.1.1 on the newly sequenced 19,456 samples, and (3) project-level joint-genotype calls in VCF format (pVCF) across all samples generated by GATK4.1.1. The pVCF released here is provided as a preview to the formal ADSP quality control that will be released in a few months. Checks of the dataset are ongoing, and the released files may be subject to change in the full quality-controlled release.

   The R4 preview pVCF includes whole-genome data from 36,361 samples, including, 1,020 ADSP Family Discovery and Discovery Extension WGS samples, 2,959 ADSP Case Control Extension WGS samples, 809 ADNI-WGS-1 samples, 886 CurePSP and Tau Consortium PSP samples, 408 PSP UCLA, and 617 NIH, CurePSP and Tau Consortium PSP samples, 209 University of Pittsburg- Kamboh samples, 8,159 ADSP Follow-Up Study (FUS) 1 samples (FUS1 contains 885 ADSP FUS1 APOE Extremes samples, 2,771 ADSP FUS1 ADC Autopsy samples, 1,517 ADSP FUS1 PR1066 samples, 1,923 ADSP FUS1 ADGC African American Dataset release 1 samples, 757 ADSP FUS1 ADNI-WGS-2 samples, 92 ADSP FUS1 Miami HIHG Brain Bank samples, 214 ADSP FUS1 StEP-AD samples), 207 Cache County samples, 77 Knight ADRC-WGS samples, 91 FASe_families-WGS samples, 137 NACC-Genentech samples, 730 AMP-AD ROSMAP samples, 344 AMP-AD MSSM samples, 252 AMP-AD MAYO samples, 12,621 ADSP FUS2 samples (FUS2 contains 981 ADSP-FUS2-EOAD1 samples, 2,620 ADSP-FUS2-MHAS samples, 964 ADSP-FUS2-IBADAN samples, 166 ADSP-FUS2-CWAutopsy1 samples, 285 ADSP-FUS2-ADC-AA samples, 80 ADSP-FUS2-ADC-Amerindian samples, 1,081 ADSP-FUS2-EFIGA samples, 312 ADSP-FUS2-Miami-BrainBank samples, 738 ADSP-FUS2-PRADI samples, 97 ADSP-FUS2-CuAADI samples, 1,040 ADSP-FUS2-AmishPV samples, 99 ADSP-FUS2-StEPAD2 samples, 736 ADSP-FUS2-REAAADI samples, 171 ADSP-FUS2-CW-RapidDecline1 samples, 247 ADSP-FUS2-PeADI samples, 1,007 ADSP-FUS2-WRAP samples, 1,202 ADSP-FUS2-ADC-Hispanics samples, 777 ADSP-FUS2-NOMAS samples, 18 ADSP-FUS2-Replicate samples), 3,132 EOAD1 samples, 1,017 ADGC-TARCC samples, and 2,686 LASI-DAD samples.

2. **ADSP-PHC Harmonized Phenotypes**

   This is the first release of the ADSP-Phenotype Harmonization Consortium (ADSP-PHC) and includes a subset of harmonized, ADSP-sequenced data from the Adult Changes in Thought (ACT) Study, Alzheimer's Disease Neuroimaging Initiative (ADNI), the National Alzheimer's Coordinating Center (NACC), National Institute on Aging Late Onset of Alzheimer's Disease (NIA-LOAD), the Religious Orders Study, Memory and Aging Project at Rush, and Minority Aging Research Study (ROS/MAP - Rush/MARS), and the Memory and Aging Project at the Knight ADRC (Knight ADRC). ReadMe files, data dictionaries, and harmonized data files are available across the following domains for joint genomic analysis:

   - Cognitive Harmonization (ACT; ADNI; NACC; ROS, MAP - Rush, MARS)
   - Fluid Biomarker Harmonization (ADNI; NACC; Knight ADRC)
   - Neuropathology Harmonization (NACC; NIA-LOAD, ROS, MAP - Rush, MARS)

Study-Specific Information:

1. **Dropped Subject:** A subject in the R2 "20k" WES dataset is no longer consented for use and was removed from the newly released files: A-ADC-AD005087. Please remove this subject from all analyses. For a running list of any subjects/samples that have been dropped over time, please refer to the file, ng00067_subject_droplist_2022.08.18.xlsx.
2. **WGS Derived APOE Genotype**: A new variable called, 'APOE_WGS', has been added to the four phenotype files. This variable reports the APOE genotype derived from the R4 preview pVCF.
3. **File Corrections**: The R3 17k WGS biallelic pVCF README and Quick Start Guide were updated to remove text regarding subset-specific VFLAGS as all VFLAGS are based on the complete set of all samples in the R3 17k dataset.

File Manifest: https://st1.niagads.org/portal/download-public/NG00067.v9/fm

Subject Consents:
Sequenced subjects in this dataset belong to the following consent levels as indicated by the submitting study IRBs:

Consent Level*	# Subjects
DS-ADRDAGE-IRB-PUB	1865
DS-ADRD-IRB-PUB	3794
DS-ADRD-IRB-PUB-NPU	5122
DS-ADRDMEM-IRB-PUB-NPU	284
DS-AGEADLT-IRB-PUB	650

DS-AGEBRMEM-IRB-PUB-GSO	2620
DS-ND-IRB-PUB	1099
DS-ND-IRB-PUB-MDS	124
DS-ND-IRB-PUB-NPU	1587
DS-NEURO-IRB-PUB	477
DS-NEURO-IRB-PUB-NPU	424
GRU-IRB-PUB	34523
GRU-IRB-PUB-MDS	408
GRU-IRB-PUB-NPU	225
HMB-IRB-PUB	4167
HMB-IRB-PUB-GSO	747
HMB-IRB-PUB-MDS	1824
HMB-IRB-PUB-NPU	1787
HMB-IRB-PUB-NPU-MDS	274
HMB-IRB-PUB-NPU-MDS-GSO	483
Total	62484

*Consent level definitions can be found on the Data Use Limitations page.

There are two new consent levels added to this release. To obtain access to the data with new consent levels, you will need to revise/renew your current Data Access Request (DAR) and have the NADAC review your research use.

Dataset Accession Numbers Available in ng00067.v5:

Type	Description	Accession
Dataset	Alzheimer's Disease Sequencing Project Umbrella Study	ng00067
Study	Alzheimer's Disease Sequencing Project	sa000001
Study	Alzheimer's Disease Neuroimaging Initiative	sa000002
Study	Alzheimer's Disease Genetics Consortium: African Americans	sa000003
Study	The Familial Alzheimer Sequencing Project	sa000004
Study	Brkanac- Family-based genome scan for AAO of LOAD	sa000005
Study	HIHG Miami Families with AD	sa000006
Study	Washington Heights/Inwood Columbia Aging Project	sa000007
Study	Charles F. and Joanne Knight Alzheimer's Disease Research Center	sa000008
Study	Corticobasal degeneration Study	sa000009
Study	Progressive Supranuclear Palsy Study	sa000010
Study	Accelerating Medicines Partnership-Alzheimer's Disease (AMP-AD)	sa000011
Study	University of Pittsburg- Kamboh (UPitt)	sa000012
Study	NACC Genentech Study	sa000013

Study	Cache County Study	sa000014
Study	NIH, CurePSP and Tau Consortium PSP WGS	sa000015
Study	CurePSP and Tau Consortium PSP WGS	sa000016
Study	UCLA Progressive Supranuclear Palsy	sa000017
Study	The Diagnostic Assessment of Dementia for the Longitudinal Aging Study of India (LASI-DAD)	sa000019
Study	Dissecting the Genomic Etiology of non-Mendelian Early-Onset Alzheimer Disease (EOAD) and Related Phenotypes	sa000023
Sampleset	ADSP_Discovery WGS/WES	snd10000
Sampleset	ADSP_Extension WGS	snd10001
Sampleset	ADNI-WGS-1 WGS	snd10002
Sampleset	ADGC_AA WES	snd10003
Sampleset	FASe_Families WES	snd10004
Sampleset	Brkanac_Families WES	snd10005
Sampleset	Miami_Families WES	snd10006
Sampleset	WHICAP WES	snd10007
Sampleset	KnightADRC WES	snd10008
Sampleset	CBD WES	snd10009
Sampleset	PSP WES	snd10010
Sampleset	AMP-AD WGS	snd10011
Sampleset	UPITT-Kamboh1 WGS	snd10012
Sampleset	NACC-Genentech WGS	snd10013
Sampleset	CacheCounty	snd10014
Sampleset	PSP NIH-CurePSP-Tau WGS	snd10015
Sampleset	PSP CurePSP-Tau WGS	snd10016
Sampleset	PSP UCLA WGS	snd10017
Sampleset	FASe WGS	snd10018
Sampleset	KnightADRC WGS	snd10019
Sampleset	ADSP FUS1	snd10020
Sampleset	ADGC-TARCC WGS	snd10030
Sampleset	ADSP-FUS2 WGS	snd10031
Sampleset	EOAD1 WGS	snd10032
Sampleset	LASI-DAD WGS	snd10033
Fileset	R1 5K, R3 17K, and R4 36K WGS CRAMs/GATK gVCFs and VCF Structural Variant (SV) calls	fsa000001
Fileset	Phenotypes, Sample Manifest, Consent Files	fsa000002
Fileset	R1 5K WGS Project Level VCF	fsa000003
Fileset	R2 20K WES CRAMs/GATK gVCFs	fsa000004

Fileset	R2 20K WES Project Level VCF	fsa000005
Fileset	R3 17K WGS Project Level VCF	fsa000006
Fileset	ADSP R3 17K WGS BioGraph SV Calls	fsa000022
Fileset	GCAD R3 17K WGS GraphTyper SV calls	fsa000023
Fileset	R4 36K WGS Project Level VCF	fsa000026
Fileset	ADSP PHC Harmonized Phenotypes	fsa000027

Data Release Date: March 24, 2022, **Dataset Version:** ng00067.v8

Release Information:
This data release includes: 1) quality-controlled X-chromosome VCF on the R3 16,905 whole-genomes, 2) quality-controlled X-chromosome PAR VCF on the R3 16,905 whole-genomes, 3) GCAD R3 WGS GraphTyper SV Callset, 4) CHARGE R3 WGS BioGraph SV Callset, and 5) formatting and clarifying corrections to a selection of files.

1. **ADSP Release 3 (R3) WGS Quality-Controlled X-Chromosome pVCF.** This pVCF contains the QC-ed bi-allelic single nucleotide variant (SNV) and insertion-deletion variant (indel) genotypes from the X-chromosome from the ADSP "17k" WGS (R3) dataset.

2. **ADSP Release 3 (R3) WGS Quality-Controlled X-Chromosome Pseudoautosomal Region (PAR) pVCF.** This pVCF contains the QC-ed bi-allelic single nucleotide variant (SNV) and insertion-deletion variant (indel) genotypes from the X-chromosome PAR region from the ADSP "17k" WGS (R3) dataset.

3. **GCAD Release 3 (R3) WGS GraphTyper Structural Variant Callset.** This SV joint genotyping pVCF contains the merged Manta and Smoove callsets from the ADSP "17k" WGS (R3) dataset.

4. **CHARGE Release 3 (R3) WGS Biograph Structural Variant Callset.** This SV joint genotyping pVCF contains SVs called by Biograph on 16,841 samples in the ADSP "17k" WGS (R3) dataset.

5. **File Corrections.** Corrections were made to a group of phenotype, qc metrics, and companion files. Files that were corrected are:
    - ADSPCaseControlPhenotypes_DS_2021.02.19.v3_<CONSENT>.txt
      Corrected the formatting of a phenotype value (Age and Age_baseline for A-MIA-UM000474)
    - SampleManifest_DS_2021.02.19.v3_<CONSENT>.txt
      Standardized platform and sequencing center information
    - gcad.qc.r3.wgs.allchr.16905.GATK.2021.08.24.v2.biallelicvars.sample.info.<CONSENT.ids
      Standardized platform and sequencing center information
    - gcad.r3.wgs.16906.VCPA1.1.2020.05.26.v2.qcmetrics.<seq/CRAM/gVCF>.<CONSENT>.txt
      Standardized platform and sequencing center information
    - Phenotype file data dictionaries
      Included additional clarifying variable definitions for age values
    - The headers for 209 samples were corrected in the R3 WGS individual-level files (CRAMs,

gVCFs, and three sets of SV files). All samples are in fileset fsa000001 and sampleset snd10012.

File Manifest: https://st1.niagads.org/portal/download-public/NG00067.v8/fm

Subject Consents:
Sequenced subjects in this dataset belong to the following consent levels as indicated by the submitting study IRBs:

Consent Level*	# Subjects
DS-ADRDAGE-IRB-PUB	1786
DS-ADRD-IRB-PUB	2024
DS-ADRD-IRB-PUB-NPU	3260
DS-ADRDMEM-IRB-PUB-NPU	241
DS-AGEADLT-IRB-PUB	739
DS-ND-IRB-PUB	837
DS-ND-IRB-PUB-MDS	25
DS-ND-IRB-PUB-NPU	1942
DS-NEURO-IRB-PUB	135
DS-NEURO-IRB-PUB-NPU	403
GRU-IRB-PUB	18832
GRU-IRB-PUB-MDS	408
GRU-IRB-PUB-NPU	177
HMB-IRB-PUB	2630
HMB-IRB-PUB-GSO	808
HMB-IRB-PUB-MDS	1633
HMB-IRB-PUB-NPU	1254
HMB-IRB-PUB-NPU-MDS	274
Total	37409

*Consent level definitions can be found on the Data Use Limitations page.

Dataset Accession Numbers Available in ng00067.v5:

Type	Description	Accession
Dataset	Alzheimer's Disease Sequencing Project Umbrella Study	ng00067
Study	Alzheimer's Disease Sequencing Project	sa000001
Study	Alzheimer's Disease Neuroimaging Initiative	sa000002

Study	Alzheimer's Disease Genetics Consortium: African Americans	sa000003
Study	The Familial Alzheimer Sequencing Project	sa000004
Study	Brkanac- Family-based genome scan for AAO of LOAD	sa000005
Study	HIHG Miami Families with AD	sa000006
Study	Washington Heights/Inwood Columbia Aging Project	sa000007
Study	Charles F. and Joanne Knight Alzheimer's Disease Research Center	sa000008
Study	Corticobasal degeneration Study	sa000009
Study	Progressive Supranuclear Palsy Study	sa000010
Study	Accelerating Medicines Partnership-Alzheimer's Disease (AMP-AD)	sa000011
Study	University of Pittsburg- Kamboh (UPitt)	sa000012
Study	NACC Genentech Study	sa000013
Study	Cache County Study	sa000014
Study	NIH, CurePSP and Tau Consortium PSP WGS	sa000015
Study	CurePSP and Tau Consortium PSP WGS	sa000016
Study	UCLA Progressive Supranuclear Palsy	sa000017
Sampleset	ADSP_Discovery WGS/WES	snd10000
Sampleset	ADSP_Extension WGS	snd10001
Sampleset	ADNI-WGS-1 WGS	snd10002
Sampleset	ADGC_AA WES	snd10003
Sampleset	FASe_Families WES	snd10004
Sampleset	Brkanac_Families WES	snd10005
Sampleset	Miami_Families WES	snd10006
Sampleset	WHICAP WES	snd10007
Sampleset	KnightADRC WES	snd10008
Sampleset	CBD WES	snd10009
Sampleset	PSP WES	snd10010
Sampleset	AMP-AD WGS	snd10011
Sampleset	UPITT-Kamboh1 WGS	snd10012
Sampleset	NACC-Genentech WGS	snd10013
Sampleset	CacheCounty	snd10014
Sampleset	PSP NIH-CurePSP-Tau WGS	snd10015
Sampleset	PSP CurePSP-Tau WGS	snd10016
Sampleset	PSP UCLA WGS	snd10017
Sampleset	FASe WGS	snd10018
Sampleset	KnightADRC WGS	snd10019
Sampleset	ADSP FUS1	snd10020

Fileset	R1 5K and R3 17K WGS CRAMs/GATK gVCFs and VCF Structural Variant (SV) calls	fsa000001
Fileset	Phenotypes, Sample Manifest, Consent Files	fsa000002
Fileset	R1 5K WGS Project Level VCF	fsa000003
Fileset	R2 20K WES CRAMs/GATK gVCFs	fsa000004
Fileset	R2 20K WES Project Level VCF	fsa000005
Fileset	R3 17K WGS Project Level VCF	fsa000006
Fileset	ADSP R3 17K WGS BioGraph SV Calls	fsa000022
Fileset	GCAD R3 17K WGS GraphTyper SV calls	fsa000023

Data Release Date: October 27, 2021, **Dataset Version:** ng00067.v7

Release Information:
This data release includes, 1) quality-controlled project level VCF on 16,905 samples, 2) X-chromosome PAR data on the R2 20,503 whole-exomes, 3) individual-level VCF structural variant calls (manta, strelka, and smoove callers).

1. **ADSP Release 3 (R3) WGS Quality-Controlled Project Level VCF (pVCF).** This pVCF contains autosomal biallelic single nucleotide variant (SNV) and insertion-deletion variant (indel) genotypes on 16,905 samples that have undergone the ADSP quality-control procedures.

2. **ADSP Release 2 (R2) WES Quality-Controlled X-Chromosome Pseudoautosomal Region (PAR) pVCF.** This pVCF contains bi-allelic single nucleotide variant (SNV) and insertion-deletion variant (indel) genotypes from the X-chromosome PAR region from the ADSP "20k" WES (R2) dataset.

3. **ADSP Release 3 (R3) Individual-level VCF Structural Variant (SV) calls.** Structural variants called independently by Strelka, Smoove and Manta in VCF format for all the sequenced 16,905 samples.

Subject-Sample Mapping ID updates:
One sample was found to be labeled with the incorrect Subject ID. The sample and subject mappings were updated as described in the table below and phenotypes were added to the ADSPCaseControlPhenotypes_DS_2021.02.19.v2_<CONSENT>.txt file. Updated IDs can be found in ng00067_updated_subject-sample_mapping_2021.08.24.xlsx

Old Subject ID	Old Sample ID	New Subject ID	New Sample ID	Sample Set
A-ADC-AD011859	A-ADC-AD011859-BL-NCR-14AD73961	A-ADC-AD018987	A-ADC-AD011859-BL-NCR-14AD73961	snd10020

Study-Specific Information:
1. The sample A-ADC-AD005043-BR-NCR-09AD15795 was found to have poor sequencing quality and should be removed from all analyses. The data for this

sample was previously released in the R3 17k WGS preview pVCF and has not been removed from the previously released files, but has been removed from all newly released files. All previously released R3 files contain 16,906 samples while the newly released R3 files contain 16,905. See the Policy Update section below for additional details.

2. Corrected QC Metrics- Information on samples of borderline quality were missing in the previous release of the CRAM and gVCF QC metrics. Those columns were added back to include reasons why samples were of borderline quality and subsequently flagged.
3. The INFO field in preview pVCF README (gcad.preview.r3.wgs.16906.GATK.2020.05.26.v2.genotypes.README.docx) was updated with a note to inform users that the INFO fields were generated based on 17,683 samples, not the total number of samples included in the preview pVCFs. It also includes how users can recalculate those fields on their end if needed.
4. Phenotype Files:
    a. This release includes phenotype updates found during the R3 17k WGS IBD pairwise comparison review.
    b. The Duplicate_SUBJID column was corrected and updated to include subjects that were identified as being genetically identical to another subject through pairwise IBD with non-unique Subject IDs. In previous versions, some related pairs were identified instead of strictly duplicates, which was corrected.
    c. The definition of the 'AD' variable for case/control subjects was clarified in the file ADSPCaseControlPhenotypes_DD_2021.02.19.v2.xlsx.
    d. For more information about phenotype updates, please refer to the file README_PhenotypesADSPUmbrella_ng00067.v7.docx

Policy Update:
Subject/Sample Removal Policy: moving forward, any samples that have already been released to the public, but need to be dropped due to newly discovered issues with data quality, subject inaccuracy, or withdrawn consent will not be removed from previously released files. Users will be notified of samples to exclude and are expected to remove the data from further analyses. A file named ng00067_subject_droplist_2021.08.24.xlsx will identify any samples that need to be removed from analyses.

File Manifest: https://st1.niagads.org/portal/download-public/NG00067.v7/fm

Subject Consents:
Sequenced subjects in this dataset belong to the following consent levels as indicated by the submitting study IRBs:

Consent Level*	# Subjects
DS-ADRDAGE-IRB-PUB	1786
DS-ADRD-IRB-PUB	2024

DS-ADRD-IRB-PUB-NPU	3260
DS-ADRDMEM-IRB-PUB-NPU	241
DS-AGEADLT-IRB-PUB	739
DS-ND-IRB-PUB	837
DS-ND-IRB-PUB-MDS	25
DS-ND-IRB-PUB-NPU	1942
DS-NEURO-IRB-PUB	135
DS-NEURO-IRB-PUB-NPU	403
GRU-IRB-PUB	18832
GRU-IRB-PUB-MDS	408
GRU-IRB-PUB-NPU	177
HMB-IRB-PUB	2630
HMB-IRB-PUB-GSO	808
HMB-IRB-PUB-MDS	1633
HMB-IRB-PUB-NPU	1254
HMB-IRB-PUB-NPU-MDS	274
Total	37409

*Consent level definitions can be found on the Data Use Limitations page.

Dataset Accession Numbers Available in ng00067.v5:

Type	Description	Accession
Dataset	Alzheimer's Disease Sequencing Project Umbrella Study	ng00067
Study	Alzheimer's Disease Sequencing Project	sa000001
Study	Alzheimer's Disease Neuroimaging Initiative	sa000002
Study	Alzheimer's Disease Genetics Consortium: African Americans	sa000003
Study	The Familial Alzheimer Sequencing Project	sa000004
Study	Brkanac- Family-based genome scan for AAO of LOAD	sa000005
Study	HIHG Miami Families with AD	sa000006
Study	Washington Heights/Inwood Columbia Aging Project	sa000007
Study	Charles F. and Joanne Knight Alzheimer's Disease Research Center	sa000008
Study	Corticobasal degeneration Study	sa000009
Study	Progressive Supranuclear Palsy Study	sa000010
Study	Accelerating Medicines Partnership-Alzheimer's Disease (AMP-AD)	sa000011
Study	University of Pittsburg- Kamboh (UPitt)	sa000012
Study	NACC Genentech Study	sa000013
Study	Cache County Study	sa000014
Study	NIH, CurePSP and Tau Consortium PSP WGS	sa000015

Study	CurePSP and Tau Consortium PSP WGS	sa000016
Study	UCLA Progressive Supranuclear Palsy	sa000017
Sampleset	ADSP_Discovery WGS/WES	snd10000
Sampleset	ADSP_Extension WGS	snd10001
Sampleset	ADNI-WGS-1 WGS	snd10002
Sampleset	ADGC_AA WES	snd10003
Sampleset	FASe_Families WES	snd10004
Sampleset	Brkanac_Families WES	snd10005
Sampleset	Miami_Families WES	snd10006
Sampleset	WHICAP WES	snd10007
Sampleset	KnightADRC WES	snd10008
Sampleset	CBD WES	snd10009
Sampleset	PSP WES	snd10010
Sampleset	AMP-AD WGS	snd10011
Sampleset	UPITT-Kamboh1 WGS	snd10012
Sampleset	NACC-Genentech WGS	snd10013
Sampleset	CacheCounty	snd10014
Sampleset	PSP NIH-CurePSP-Tau WGS	snd10015
Sampleset	PSP CurePSP-Tau WGS	snd10016
Sampleset	PSP UCLA WGS	snd10017
Sampleset	FASe WGS	snd10018
Sampleset	KnightADRC WGS	snd10019
Sampleset	ADSP FUS1	snd10020
Fileset	R1 5K and R3 17K WGS CRAMs/GATK gVCFs and VCF Structural Variant (SV) calls	fsa000001
Fileset	Phenotypes, Sample Manifest, Consent Files	fsa000002
Fileset	R1 5K WGS Project Level VCF	fsa000003
Fileset	R2 20K WES CRAMs/GATK gVCFs	fsa000004
Fileset	R2 20K WES Project Level VCF	fsa000005
Fileset	R3 17K WGS Project Level VCF	fsa000006

Mid-Release Updates: August 4, 2021, **Dataset Version:** ng00067.v6

Release Information:
In the last release, ng00067.v6, the chr8 files of the R3 Whole Genome Sequencing Preview dataset were corrected and released. After releasing, it was discovered that the chr8 compact and compact_filtered files were not zipped correctly. The original files were generated using gzip instead of bgzip. During the conversion of gzip to bgzip to generate

the proper index file, the original vcf.gz files (in gzip format) were not overwritten since the suffix was the same (vcf.gz) for both gzip and bgzip files. Both the vcf.gz (in bgzip format) and their corresponding .tbi files have been corrected and re-released.

The four files replaced are:
gcad.preview.compact.r3.wgs.16906.GATK.2020.05.26.v2.biallelic.genotypes.chr8.ALL.vcf.gz
gcad.preview.compact.r3.wgs.16906.GATK.2020.05.26.v2.biallelic.genotypes.chr8.ALL.vcf.gz.tbi
gcad.preview.compact_filtered.r3.wgs.16906.GATK.2020.05.26.v2.biallelic.genotypes.chr8.ALL.vcf.gz
gcad.preview.compact_filtered.r3.wgs.16906.GATK.2020.05.26.v2.biallelic.genotypes.chr8.ALL.vcf.gz.tbi

All files mentioned above have been updated and replaced on the portal. The file names and the dataset accession version remain the same, but the release date for the four corrected files has been updated to August 5, 2021.

Data Release Date: July 6, 2021, **Dataset Version:** ng00067.v6

Release Information:
This release includes a file correction to the ADSP Release 3 (R3) Whole Genome Sequencing (WGS) Preview files released in ng00067.v5. It was discovered that there were missing chunks of variants within the chr8 preview files. After careful investigation, we found that the missing chunks were the tails of regions when chr8 was divided into several smaller regions to accelerate data processing.

To repair the missing regions, the chr8 files were reprocessed in two ways; first by breaking the regions up, and second, without breaking the regions up. This process helped to ensure the final variant numbers of the resulting files matched the original GATK VCF variant numbers.

Each chromosome was reviewed to ensure that there were not additional missing regions and we determined that the issue was isolated to the chromosome 8 files specifically.

Below is a summary of the variants missing from the biallelic chr8 files released in ng00067.v5:

Chr	Start	End	# of variants missed per region
chr8	3970783	14513863	1,293,700
chr8	65322114	72569316	552,876
chr8	106996453	116110896	689,367
chr8	123400090	130624762	565,083
chr8	137716386	145078573	630,014

# of variants missed in ng00067.v5 chr8 files	3,731,040

# of variants released in ng00067.v5 chr8 files	8,384,451
# of variants released in ng00067.v6 chr8 files	12,115,491

The corrected files are now ready for download via DSS. Below is a list of the file names to replace from ng00067.v5:

1. gcad.preview.r3.wgs.16906.GATK.2020.05.26.biallelic.genotypes.chr8.<CONSENT>.vcf.gz
2. gcad.preview.compact.r3.wgs.16906.GATK.2020.05.26.biallelic.genotypes.chr8.<CONSENT>.vcf.gz
3. gcad.preview.compact_filtered.r3.wgs.16906.GATK.2020.05.26.biallelic.genotypes.chr8.<CONSENT>.vcf.gz

File Manifest: https://st1.niagads.org/portal/download-public/NG00067.v6/fm

Subject Consents:
Sequenced subjects in this dataset belong to the following consent levels as indicated by the submitting study IRBs:

Consent Level*	# Subjects
DS-ADRDAGE-IRB-PUB	1786
DS-ADRD-IRB-PUB	2024
DS-ADRD-IRB-PUB-NPU	3260
DS-ADRDMEM-IRB-PUB-NPU	241
DS-AGEADLT-IRB-PUB	739
DS-ND-IRB-PUB	837
DS-ND-IRB-PUB-MDS	25
DS-ND-IRB-PUB-NPU	1942
DS-NEURO-IRB-PUB	135
DS-NEURO-IRB-PUB-NPU	403
GRU-IRB-PUB	18833
GRU-IRB-PUB-MDS	408
GRU-IRB-PUB-NPU	177
HMB-IRB-PUB	2630
HMB-IRB-PUB-GSO	808
HMB-IRB-PUB-MDS	1633
HMB-IRB-PUB-NPU	1254
HMB-IRB-PUB-NPU-MDS	274
Total	37409

*Consent level definitions can be found on the Data Use Limitations page.

Data Release Date: February 25, 2021, **Dataset Version:** ng00067.v5

Release Information:
This data release includes, 1) new whole-genome sequencing on 12,118 samples joint-genotype called with the R1 4,788 whole-genomes previously released, totaling 16,906 samples, 2) quality-controlled X-chromosome data on the R2 20,503 whole-exomes, and 3) updates to consent levels for samples previously released.

4. **ADSP Release 3 (R3) Whole Genome Sequencing (WGS) Preview.** This ADSP release, containing 16,906 whole-genomes, is referred to as Release 3 (R3). It includes 1) sequencing read alignments in CRAM (compressed BAM) format for the newly sequenced 12,118 samples, (2) genomic Variant Call Format (gVCF) files generated by GATK4.1.1 on all 16,906 samples, and (3) project-level joint-genotype calls in VCF format (pVCF) across all samples generated by GATK4.1.1. The pVCF released here is provided as a preview to the formal ADSP quality control that will be released in a few months. Checks of the dataset are ongoing and the released files may be subject to change in the full quality-controlled release.

   The R3 preview pVCF includes whole-genome data from 1,020 ADSP Family Discovery and Discovery Extension samples, 2,959 ADSP Case Control Extension samples, 809 ADNI-WGS-1 samples, 886 CurePSP and Tau Consortium PSP samples, 408 PSP UCLA samples, 617 NINDS, CurePSP and Tau Consortium PSP samples, 209 University of Pittsburgh- Kamboh samples, 207 Cache County samples, 77 Knight ADRC samples, 91 FASe_families samples, 137 NACC-Genentech samples, 730 AMP-AD ROSMAP samples, 344 AMP-AD MSSM samples, 252 AMP-AD MAYO samples, and 8,160 ADSP Follow-Up Study 1 samples (FUS1 contains 885 ADSP FUS1 APOE Extremes samples, 2,772 ADSP FUS1 ADC Autopsy samples, 1,517 ADSP FUS1 PR1066 samples, 1,923 ADSP FUS1 ADGCAA samples, 757 ADSP FUS1 ADNI-WGS-2 samples, 92 ADSP FUS1 Miami HIHG Brain Bank samples, and 214 ADSP FUS1 StEP-AD samples).

   NOTE: All gVCFs generated for the R1 5k dataset were processed using VCPA1.0 (GATK3.7) while the R3 17k gVCFs were processed using VCPA 1.1 (GATK4.1.1) to call variants. You will notice 2 sets of gVCFs for all samples in the R1 5k dataset. The VCPA1.0 generated gVCFs will only be available for download for another ~6 months before they get moved to deep archival storage. These archived files can be requested directly from NIAGADS.

5. **ADSP Release 2 (R2) Quality-Controlled X-Chromosome data.** This dataset contains bi-allelic single nucleotide variant (SNV) and insertion-deletion variant (indel) genotypes from the X-chromosome from the ADSP “20k” WES (R2) dataset.

6. **Updated Consent.** This release includes updated consent levels for 225 subjects that went from DS-NEURO-IRB-PUB to a more restrictive consent for non-profit use only, DS-NEURO-IRB-PUB-NPU. Another 35 samples had a consent level change to a less restrictive consent. Additionally, we were notified that 2 participants are no

longer consented for use. There was one subject in the R1 “5k” dataset (A-ADC-AD010043) as well as one subject in the R2 “20k” dataset (A-ADC-AD003737) that were removed from the files previously released. Please remove these two subjects from all analyses. All files previously released have been updated to reflect these changes and can be identified within the file manifest by looking for file_release_date = 02/25/2021. NOTE: the R1 and R2 file names still contain 4,789 and 20,504 samples respectively even though a sample was removed from each. The file name increased in version, but the number of samples in the file name stayed the same.

File Manifest: https://st1.niagads.org/portal/download-public/NG00067.v5/fm

Subject Consents:
Sequenced subjects in this dataset belong to the following consent levels as indicated by the submitting study IRBs:

Consent Level*	# Subjects
DS-ADRDAGE-IRB-PUB	1786
DS-ADRD-IRB-PUB	2024
DS-ADRD-IRB-PUB-NPU	3260
DS-ADRDMEM-IRB-PUB-NPU	241
DS-AGEADLT-IRB-PUB	739
DS-ND-IRB-PUB	837
DS-ND-IRB-PUB-MDS	25
DS-ND-IRB-PUB-NPU	1942
DS-NEURO-IRB-PUB	135
DS-NEURO-IRB-PUB-NPU	403
GRU-IRB-PUB	18833
GRU-IRB-PUB-MDS	408
GRU-IRB-PUB-NPU	177
HMB-IRB-PUB	2630
HMB-IRB-PUB-GSO	808
HMB-IRB-PUB-MDS	1633
HMB-IRB-PUB-NPU	1254
HMB-IRB-PUB-NPU-MDS	274
Total	37409

*Consent level definitions can be found on the Data Use Limitations page.

Dataset Accession Numbers Available in ng00067.v5:

Type	Description	Accession
Dataset	Alzheimer's Disease Sequencing Project Umbrella Study	ng00067

Study	Alzheimer's Disease Sequencing Project	sa000001
Study	Alzheimer's Disease Neuroimaging Initiative	sa000002
Study	Alzheimer's Disease Genetics Consortium: African Americans	sa000003
Study	The Familial Alzheimer Sequencing Project	sa000004
Study	Brkanac- Family-based genome scan for AAO of LOAD	sa000005
Study	HIHG Miami Families with AD	sa000006
Study	Washington Heights/Inwood Columbia Aging Project	sa000007
Study	Charles F. and Joanne Knight Alzheimer's Disease Research Center	sa000008
Study	Corticobasal degeneration Study	sa000009
Study	Progressive Supranuclear Palsy Study	sa000010
Study	Accelerating Medicines Partnership-Alzheimer's Disease (AMP-AD)	sa000011
Study	University of Pittsburg- Kamboh (UPitt)	sa000012
Study	NACC Genentech Study	sa000013
Study	Cache County Study	sa000014
Study	NIH, CurePSP and Tau Consortium PSP WGS	sa000015
Study	CurePSP and Tau Consortium PSP WGS	sa000016
Study	UCLA Progressive Supranuclear Palsy	sa000017
Sampleset	ADSP_Discovery WGS/WES	snd10000
Sampleset	ADSP_Extension WGS	snd10001
Sampleset	ADNI-WGS-1 WGS	snd10002
Sampleset	ADGC_AA WES	snd10003
Sampleset	FASe_Families WES	snd10004
Sampleset	Brkanac_Families WES	snd10005
Sampleset	Miami_Families WES	snd10006
Sampleset	WHICAP WES	snd10007
Sampleset	KnightADRC WES	snd10008
Sampleset	CBD WES	snd10009
Sampleset	PSP WES	snd10010
Sampleset	AMP-AD WGS	snd10011
Sampleset	UPITT-Kamboh1 WGS	snd10012
Sampleset	NACC-Genentech WGS	snd10013
Sampleset	CacheCounty	snd10014
Sampleset	PSP NIH-CurePSP-Tau WGS	snd10015
Sampleset	PSP CurePSP-Tau WGS	snd10016
Sampleset	PSP UCLA WGS	snd10017
Sampleset	FASe WGS	snd10018

Sampleset	KnightADRC WGS	snd10019
Sampleset	ADSP FUS1	snd10020
Fileset	R1 5K and R3 17K WGS CRAMs/GATK gVCFs	fsa000001
Fileset	Phenotypes, Sample Manifest, Consent Files	fsa000002
Fileset	R1 5K WGS Project Level VCF	fsa000003
Fileset	R2 20K WES CRAMs/GATK gVCFs	fsa000004
Fileset	R2 20K WES Project Level VCF	fsa000005
Fileset	R3 17K WGS Project Level VCF	fsa000006

Data Release Date: November 24, 2020, **Dataset Version:** ng00067.v4

Release Information:
This release includes updated consent for 104 subjects that went from GRU-IRB-PUB to GRU-IRB-PUB-NPU. Also, within the 5k WGS data, two corrections were made to the files previously released: 1.) the CRAM index file of ADNI_068_S_0473 was corrupted and has been replaced, and 2.) within the previously released gcad.wgs.4789.GATK.2018.07.30.qcmetrics.seq.CONSENT.txt file, the ADNI samples were listed as PCR-free, but they were all PCR-amplified. All files that have been updated can be located within the file manifest by looking for file_release_date = 11/24/2020.

File Manifest: https://st1.niagads.org/portal/download-public/NG00067.v4/fm

Subject Consents:
Sequenced subjects in this dataset belong to the following consent levels as indicated by the submitting study IRBs:

Consent Level*	# Subjects
DS-ADRDAGE-IRB-PUB	1046
DS-ADRD-IRB-PUB	1181
DS-ADRD-IRB-PUB-NPU	2276
DS-ADRDMEM-IRB-PUB-NPU	134
DS-AGEADLT-IRB-PUB	647
DS-ND-IRB-PUB	343
DS-ND-IRB-PUB-MDS	18
DS-ND-IRB-PUB-NPU	1091
DS-NEURO-IRB-PUB	352
GRU-IRB-PUB	11713
GRU-IRB-PUB-NPU	104
HMB-IRB-PUB	1375
HMB-IRB-PUB-GSO	745

HMB-IRB-PUB-MDS	1315
HMB-IRB-PUB-NPU	1254
HMB-IRB-PUB-NPU-MDS	274
Total	23868

*Consent level definitions can be found on the Data Use Limitations page.

Data Release Date: September 24, 2020, **Dataset Version:** ng00067.v3

Release Information:
This release includes whole-exome CRAMs, gVCFs, and phenotypes for 582 subjects that were previously not released due to missing consent as well as the joint genotype called project level VCF for 20,504 samples that has undergone quality control by the ADSP. Additionally, some phenotypes have been updated/corrected since the ng00067.v2 release.

Dataset Accession Numbers Available in ng00067.v3:

Type	Description	Accession
Dataset	Alzheimer's Disease Sequencing Project Umbrella Study	ng00067
Study	Alzheimer's Disease Sequencing Project	sa000001
Study	Alzheimer's Disease Neuroimaging Initiative	sa000002
Study	Alzheimer's Disease Genetics Consortium: African Americans	sa000003
Study	The Familial Alzheimer Sequencing Project	sa000004
Study	Brkanac- Family-based genome scan for AAO of LOAD	sa000005
Study	HIHG Miami Families with AD	sa000006
Study	Washington Heights/Inwood Columbia Aging Project	sa000007
Study	Charles F. and Joanne Knight Alzheimer's Disease Research Center	sa000008
Study	Corticobasal degeneration Study	sa000009
Study	Progressive Supranuclear Palsy Study	sa000010
Sampleset	ADSP_Discovery	snd10000
Sampleset	ADSP_Extension	snd10001
Sampleset	ADNI-WGS-1	snd10002
Sampleset	ADGC_AA	snd10003
Sampleset	FASe_Families	snd10004
Sampleset	Brkanac_Families	snd10005
Sampleset	Miami_Families	snd10006
Sampleset	WHICAP	snd10007
Sampleset	KnightADRC	snd10008
Sampleset	CBD	snd10009

Sampleset	PSP	snd10010
Fileset	R1 5K WGS CRAMs/GATK gVCFs	fsa000001
Fileset	Phenotypes, Sample Manifest, Consent Files	fsa000002
Fileset	R1 5K WGS Project Level VCF	fsa000003
Fileset	R2 20K WES CRAMs/GATK gVCFs	fsa000004
Fileset	R2 20K WES Project Level VCF	fsa000005

Subject-Sample Mapping ID updates:
Two samples were found to be labeled with the incorrect Subject IDs. The sample and subject mappings were updated as described in the table below and phenotypes were added to the ADSPCaseControlPhenotypes_DS_2020.06.26_ALL.xlsx file.

Old Subject ID	Old Sample ID	New Subject ID	New Sample ID	Sample Set
A-ADC-AD007656	A-ADC-AD010404-BL-NCR-12AD50881	A-ADC-AD010404	A-ADC-AD010404-BL-NCR-12AD50881	snd10003
A-ADC-AD007783	A-ADC-AD007783-BL-NCR-11AD44272	A-ADC-AD008405	A-ADC-AD007783-BL-NCR-11AD44272	snd10000

Study-Specific Information:

1. This version introduces new phenotype variables including baseline values (representing the phenotype values when the subject entered the ADSP study), update indicators, duplicate subject IDs, and a correction indicator to help show changes in phenotypes over time. New variables are listed below with a brief description and the phenotype file that the variable has been added to. Detailed information about the new variables can be found in the README provided with the phenotype files.
2. The Autopsy variable used in case/control and family-based phenotype files is under review. Since the current data may not be accurately represented, the Autopsy variable values have been changed to “NA” until they have been reevaluated. We will update the phenotypes with improved variables after review is complete.

File Manifest: https://st1.niagads.org/portal/download-public/NG00067.v3/fm

Subject Consents:
582 subjects that were previously not released due to consent issues have been added back into the dataset as they are now consented for use. All corresponding files have been updated with the added 582 subjects. There are now 146 subjects that are not consented and are not included in the dataset (listed here SubjectNotConsented_2020.06.26.xlsx). This only affects samples included in the R2 20K WES dataset, all samples in the R1 5k WGS dataset are consented.

Sequenced subjects in this dataset belong to the following consent levels as indicated by the submitting study IRBs:

Consent Level*	# Subjects
DS-ADRDAGE-IRB-PUB	1046
DS-ADRD-IRB-PUB	1181
DS-ADRD-IRB-PUB-NPU	2276
DS-ADRDMEM-IRB-PUB-NPU	134
DS-AGEADLT-IRB-PUB	647
DS-ND-IRB-PUB	343
DS-ND-IRB-PUB-MDS	18
DS-ND-IRB-PUB-NPU	1091
DS-NEURO-IRB-PUB	352
GRU-IRB-PUB	11817
HMB-IRB-PUB	1375
HMB-IRB-PUB-GSO	745
HMB-IRB-PUB-MDS	1315
HMB-IRB-PUB-NPU	1254
HMB-IRB-PUB-NPU-MDS	274
Total	23868

*Consent level definitions can be found on the Data Use Limitations page.

Data Release Date: February 19, 2020, **Dataset Version:** ng00067.v2

Release Information:
This release includes whole-exome CRAMs, gVCFs, and phenotypes from nine different studies, including 10,088 ADSP Discovery Case Control samples, 3,144 ADGC African American samples, 75 Brkanac Families samples, 346 Corticobasal degeneration samples, 3,861 Columbia WHICAP samples, 1,100 FASe Families samples, 650 Knight ADRC samples, 108 Miami Families samples, and 550 Progressive supranuclear palsy samples. In addition to the new whole-exome data, phenotypes and APOE genotypes were updated for the samples released in the r1 5k WGS dataset.

Accession Number Updates:
NIAGADS improved how accession numbers are assigned and used in our database. See below for the updates made to the previous release:

Type	Description	Old Accession	New Accession
Dataset	ADSP Umbrella	None	NG00067
Study	ADSP	NG00067	sa000001
Study	ADNI	NG00066	sa000002
Fileset	ADSP/ADNI Phenotypes/Pedigrees	dnd00001	fsa000002
Fileset	ADSP/ADNI Project Level pVCF	None	fsa000003
Fileset	ADSP/ADNI CRAMs/gVCFs	None	fsa000001

Dataset Accession Numbers Available in ng00067.v2:

Type	Description	Accession
Dataset	Alzheimer's Disease Sequencing Project Umbrella Study	ng00067
Study	Alzheimer's Disease Sequencing Project	sa000001
Study	Alzheimer's Disease Neuroimaging Initiative	sa000002
Study	Alzheimer's Disease Genetics Consortium: African Americans	sa000003
Study	The Familial Alzheimer Sequencing Project	sa000004
Study	Brkanac- Family-based genome scan for AAO of LOAD	sa000005
Study	HIHG Miami Families with AD	sa000006
Study	Washington Heights/Inwood Columbia Aging Project	sa000007
Study	Charles F. and Joanne Knight Alzheimer's Disease Research Center	sa000008
Study	Corticobasal degeneration Study	sa000009
Study	Progressive Supranuclear Palsy Study	sa000010
Sampleset	ADSP_Discovery	snd10000
Sampleset	ADSP_Extension	snd10001
Sampleset	ADNI-WGS-1	snd10002
Sampleset	ADGC_AA	snd10003
Sampleset	FASe_Families	snd10004
Sampleset	Brkanac_Families	snd10005
Sampleset	Miami_Families	snd10006
Sampleset	WHICAP	snd10007
Sampleset	KnightADRC	snd10008
Sampleset	CBD	snd10009
Sampleset	PSP	snd10010
Fileset	R1 5K WGS CRAMs/GATK gVCFs	fsa000001
Fileset	Phenotypes, Sample Manifest, Consent Files	fsa000002
Fileset	R1 5K WGS Project Level VCF	fsa000003
Fileset	R2 20K WES CRAMs/GATK gVCFs	fsa000004

Subject-Sample Mapping ID updates:

16 pairs of samples were found to be duplicates included in the WGS and WES datasets with different subject IDs. In order to account for this relationship, the WGS subject IDs were renamed with the WES subject IDs. The original sample IDs were retained as the final sample ID (see below):

Old Subject ID	Old Sample ID	New Subject ID	New Sample ID	Sample Set
A-ADC-AD010404	A-ADC-AD010404-SA-NCR-12AD49194	A-ADC-AD007656	A-ADC-AD010404-SA-NCR-12AD49194	snd10001

A-ACT-AC000175	A-ACT-AC000175-BL-NCR-12AD50669	A-ACT-AC003385	A-ACT-AC000175-BL-NCR-12AD50669	snd10001
A-ACT-AC000210	A-ACT-AC000210-BL-NCR-12AD50611	A-ACT-AC003408	A-ACT-AC000210-BL-NCR-12AD50611	snd10001
A-ACT-AC000412	A-ACT-AC000412-BL-NCR-12AD50653	A-ACT-AC003370	A-ACT-AC000412-BL-NCR-12AD50653	snd10001
A-ACT-AC000846	A-ACT-AC000846-BL-NCR-12AD50601	A-ACT-AC003402	A-ACT-AC000846-BL-NCR-12AD50601	snd10001
A-ACT-AC000942	A-ACT-AC000942-BL-NCR-12AD50651	A-ACT-AC003382	A-ACT-AC000942-BL-NCR-12AD50651	snd10001
A-ACT-AC001005	A-ACT-AC001005-BL-NCR-12AD50629	A-ACT-AC003405	A-ACT-AC001005-BL-NCR-12AD50629	snd10001
A-ACT-AC002132	A-ACT-AC002132-BL-NCR-12AD50655	A-ACT-AC003404	A-ACT-AC002132-BL-NCR-12AD50655	snd10001
A-ACT-AC002570	A-ACT-AC002570-BL-NCR-12AD50662	A-ACT-AC003399	A-ACT-AC002570-BL-NCR-12AD50662	snd10001
A-ACT-AC002636	A-ACT-AC002636-BL-NCR-12AD50661	A-ACT-AC003391	A-ACT-AC002636-BL-NCR-12AD50661	snd10001
A-ACT-AC002737	A-ACT-AC002737-BL-NCR-12AD50649	A-ACT-AC003411	A-ACT-AC002737-BL-NCR-12AD50649	snd10001
A-ACT-AC003443	A-ACT-AC003443-BL-NCR-11AD38117	A-ACT-AC002976	A-ACT-AC003443-BL-NCR-11AD38117	snd10001
A-ACT-AC003423	A-ACT-AC003423-BL-NCR-11AD38132	A-ACT-AC002991	A-ACT-AC003423-BL-NCR-11AD38132	snd10001
A-ACT-AC003431	A-ACT-AC003431-BL-NCR-11AD38163	A-ACT-AC003021	A-ACT-AC003431-BL-NCR-11AD38163	snd10001
A-ACT-AC003434	A-ACT-AC003434-BL-NCR-11AD38188	A-ACT-AC003046	A-ACT-AC003434-BL-NCR-11AD38188	snd10001
A-ACT-AC003435	A-ACT-AC003435-BL-NCR-11AD38198	A-ACT-AC003056	A-ACT-AC003435-BL-NCR-11AD38198	snd10001

Study-Specific Information:

I. In addition to the ADSP family-based study, there are an additional four studies containing families. They are from the ADGC_AA_WES (family and case/control), FASe_Families_WES, Brkanac_Families_WES, Miami_Families_WES studies. Where there were individuals from the same family sequenced across multiple studies, we tried to include connecting family members. There are 10 families where only one sample passed QC and made it into the final pVCF (table below). These 10 subjects were moved into the case/control phenotype file instead of the family based. There is one family included with only one sample sequenced: NI0002F, G-NIMH-NI000005, from Brkanac_Families_WES.

FamID	SUBJID	Dataset
LD0490F	A-LOAD-LD002591	FASe_Families
LD0534F	A-LOAD-LD002792	Brkanac_Families
LD1264F	A-LOAD-LD006382	Brkanac_Families
LD1517F	A-LOAD-LD010146	FASe_Families
LD1775F	A-LOAD-LD011502	Brkanac_Families
LD1808F	A-LOAD-LD011839	FASe_Families
UM0145F	A-MIA-UM000295	Miami_Families
UM0222F	A-MIA-UM000955	Miami_Families
NC0076F	A-NCRD-NC005044	FASe_Families
NC0181F	A-NCRD-NC011010	FASe_Families

II. Several ADSP Discovery WES samples were dropped after reprocessing on the vcpa1.1 pipeline in comparison to the data that was released in phs000572.v7 through dbGaP. phs000572.v7 contained 10,929 samples. 123 samples were dropped due to contamination (freemix >0.05) and 4 samples were dropped because they were unexpected duplicates between other non-AD samples in the r2 20k WES dataset. The dropped samples list and reasons can be found in the document gcad.r2.wes.19922.VCPA1.1.2019.11.01_dropped.xlsx.

III. Cohort specific information regarding how the phenotypes are reported can be found in the following document: https://dss.niagads.org/datasets/NG00067/adsp_phenotype_notes_bycohort.docx.

File Manifest: https://dss.niagads.org/dss_file-manifest_adspumbrella.xlsx

Subject Consents:
There are a few cohorts from which we do not have a GDS Institutional Certification form yet, so their data cannot be shared. There are currently 728 subjects where this documentation is missing. A list of these samples can be found in the file, SubjectNotConsented_2020_01.15.xlsx. This only affects samples included in the r2 WES dataset, all samples in the r1 5k WGS dataset are consented.

Consent levels were updated for 1,425 subjects released in the previous version from the 5k WGS dataset. If your Data Access Request is not approved for all consent levels, you will see each file broken out by consent level. To access all consent levels, you will need to revise your DAR for the NADAC to review. A list of file changes due to updated consent levels can be found in this spreadsheet: https://dss.niagads.org/dss_file-version-changes_adspumbrella_NG00067.v2.

Sequenced subjects in this dataset belong to the following consent levels as indicated by the submitting study IRBs:

Consent Level*	# Subjects
DS-ADRDAGE-IRB-PUB	1046
DS-ADRD-IRB-PUB	1180
DS-ADRD-IRB-PUB-NPU	2276
DS-ADRDMEM-IRB-PUB-NPU	134
DS-AGEADLT-IRB-PUB	647
DS-ND-IRB-PUB	343
DS-ND-IRB-PUB-MDS	18
DS-ND-IRB-PUB-NPU	1091
DS-NEURO-IRB-PUB	352
GRU-IRB-PUB	11235
HMB-IRB-PUB	1375

HMB-IRB-PUB-GSO	745
HMB-IRB-PUB-MDS	1315
HMB-IRB-PUB-NPU	1254
HMB-IRB-PUB-NPU-MDS	274
Total	23285

*Consent level definitions can be found on the Data Use Limitations page.

Data Release Date: October 30, 2018, **Dataset Version:** 2018.09.17

This release includes the ADSP quality control checked GATK joint called VCF containing all 4789 whole genomes released as part of 2018.07.30 (described below). Available file types include project level VCFs and quality control companion files.

File Manifest: https://dss.niagads.org/dss_file-manifest_adspumbrella.xlsx

Phenotype Updates:
Documentation about phenotypes released with the ADSP data mapped to hg37 in dbGaP are available on the dbGaP website in the ADSP entry Release Notes for phs000572.v7. Several updates have been made to the phenotypes since the last release in dbGaP, April 2017, and are described below:

1. Twelve sex mismatches found in the Discovery WES dataset; 8 samples updated, 4 dropped from study (this wes dataset has not been released through NIAGADS yet, it is currently only available through dbGaP, phs000572.v7):
   - C-ASPS-30075-BL-ASPS-801050- updated to male
   - C-ASPS-50003-BL-ASPS-199000- updated to male
   - C-ASPS-51466-BL-ASPS-36800- updated to male
   - C-RS-40001-BL-ERA-5663001- updated to male
   - C-ASPS-51642-BL-ASPS-63800- updated to female
   - A-RAS-RA000011-BL-UPN-27627- updated to female
   - C-ASPS-52021-BL-ASPS-127401- updated to female
   - C-ASPS-51379-BL-ASPS-38900- updated to female
   - A-LOAD-LD012112-BR-NCR-10AD24166- incorrect sex, sample dropped
   - C-RS-30149-BL-ERA-8643003- incorrect sex, sample dropped
   - C-RS-50723-BL-ERA-4525001- incorrect sex, sample dropped
   - A-MAP-MA000771-BR-RUS-003- incorrect sex, sample dropped

2. Several pedigree structural inconsistencies were found:
   - A-CUHS-CU000723 (Family CU0015F) is unrelated to the rest of the pedigree and was removed.
   - A-CUHS-CU000970 (Family CU0022F) was found to be unrelated to A-CUHS-CU000978 and a half-sib to A-CUHS-CU000971, A-CUHS-CU000973, and A-CUHS-CU000972. A dummy parent was created to replace A-CUHS-CU000978. Dummy parent ID = A-CUHS-CU009813.

- A-CUHS-CU001246 (Family CU0029F) is a spouse control and does not add any additional information. This subject along with all members within the branch have been removed from the pedigree.
- A-CUHS-CU003128 (Family CU0082F) was sequenced, but appears unrelated to the rest of the family. Sample has been dropped from the study.
- Many subjects in the CU sample set were reported as unaffected, but should have been marked as unknown AD status. These have been corrected within the phenotype file.
- A-CUHS-CU001552 (Family CU0036F) had the incorrect parents reported. The correct parents are A-CUHS-CU001556 and A-CUHS-CU001558.

3. Two samples are dropped from the hg38 version of the data as the BAM file was found to be discordant with the GWAS data. The two samples are:
   - A-MIA-UM001976-BL-MIA-20010205
   - A-MIA-UM000315-BL-MIA-19961724

4. In the 2018.07.30 release, it was noted that 23 ADNI IDs contained a lowercase 's' instead of an uppercase 'S'. All ADNI subjects should have an uppercase 'S', however due to time constraints we decided to change the phenotypes to lowercase 's' to match the sequencing IDs. In addition, all ADNI subjects in the phenotype files were missing the leading "ADNI_" in the ID and we have updated the IDs to contain this. These changes have been updated in this release, October 30, 2018. The updated files are labeled '2018.07.30.v2' in the filename.

Data Release Date: July 30, 2018, **Dataset Version:** 2018.07.30

This release includes whole-genome sequencing (WGS) data from the ADSP and ADNI studies. Available file types include CRAMs, GATK-called gVCFs, sequencing metrics, phenotypes, and pedigree structures for family based subjects.

These data were processed using the Genome Center for Alzheimer's Disease (GCAD) pipeline, VCPA1.0. All samples were mapped to Genome Reference Consortium Human Build 38 (GRCh38) and variant called using GATK.

Sample Set	Accession	Number of Subjects	Number of Samples
ADSP Discovery	snd10000	n = 574	n = 580
ADSP Extension	snd10001	n = 3367	n = 3400
ADNI-WGS-1	snd10002	n = 809	n = 809

The subjects belong to the following consent levels as indicated by the submitting study IRBs:

Consent Level*	# Samples
DS-ADRDAGE-IRB-PUB	214
DS-ADRD-IRB-PUB	98
DS-ADRDMEM-IRB-PUB-NPU	20
DS-AGEADLT-IRB-PUB	173
DS-AGEADLT-IRB-PUB-NPU	77
DS-AGEBRMEM-IRB-PUB-GSO	7
DS-DEMND-IRB-PUB	186
DS-DEMND-IRB-PUB-NPU	91
DS-ND-IRB-PUB	61
DS-ND-IRB-PUB-MDS	4
DS-ND-IRB-PUB-NPU	64
DS-NEURO-IRB-PUB	173
DS-NEURO-IRB-PUB-NPU	1
GRU-IRB-PUB	3110
GRU-IRB-PUB-NPU	36
HMB-IRB-PUB	250
HMB-IRB-PUB-GSO	102
HMB-IRB-PUB-NPU	122

*Consent level definitions can be found on the Data Use Limitations page.