RELEASE NOTES- ADSP Umbrella

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.