dbGaP Study Release Notes

Release Notes for NHGRI FUSION Islet, phs001188.v2.p1
"The FUSION Study - Islet Expression and Regulation"

For any questions or comments, please contact: dbgap-help@ncbi.nlm.nih.gov.

August	9, 2017	Version 1 Data set release
December	27, 2019	Version 2 Data set release
January	7, 2020	Version 2 ID list correction

2020-01-07
Version 2 ID list correction

31 sample IDs that were missing from the subject sample mapping and sample attributes datasets have been added back.

The highlighted counts have been added to the final counts for version 2:

Consent group 1 (c1): General Research Use (GRU)

Data	subjects	samples
Phenotype	106	391
RNA-Seq*	81	142
WGS*	12	17
miRNA-Seq*	68	68
Molecular_QTL	31	31
Seq_RNA_Expression	31	31
Array_DNA_Methylation	33	41
Array_SNP	103	103
Imputation_SNP	103	103
ATAC-seq*	13	13
Bisulfite-Seq*	12	12

For a description of the SAMPLE_USE terms, please see:
https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi
*These data are brokered through the Sequence Read Archive (SRA). Please see Authorized Access instructions below.

The 2 datasets affected:

pht	version	Dataset Name
6300	2	FUSION_Islet_Sample
6302	2	FUSION_Islet_Sample_Attributes

Authorized Access (Individual Level Data and SRA Data)

Individual level data and Sequence Read Archive (SRA) data are available for download through the dbGaP Authorized Access System:

- http://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- ftp://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001188/phs001188.v2.p1

2019-12-27

Version 2 Data set release for NHGRI FUSION Islet now available

This release includes updated phenotype tables and new molecular data added. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): General Research Use (GRU)

Data	subjects	samples
Phenotype	106	391
RNA-Seq*	81	142
WGS*	12	17
miRNA-Seq*	68	68
Array_DNA_Methylation	33	41
Array_SNP	103	103
Imputation_SNP	103	103
ATAC-seq*	13	13
Bisulfite-Seq*	12	12

For a description of the SAMPLE_USE terms, please see:
https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi
*These data are brokered through the Sequence Read Archive (SRA). Please see Authorized Access instructions below.

Study and Phenotype Updates

1. NHGRI FUSION Islet study version 1 phs001188.v1.p1 has been updated to version 2. The dbGaP accession for the current set of data is **phs001188.v2.p1**. The participant set number (p#) has not changed in version 2; no subjects have changed consent groups. New subjects have been added.

2. **Updated Datasets** (n=4 datasets)

pht	version	Dataset Name
6300	2	FUSION_Islet_Sample
6301	2	FUSION_Islet_Subject_Phenotypes
6302	2	FUSION_Islet_Sample_Attributes
6615	2	FUSION_Islet_Subject

3. **New Variables** (n=3 variables)

pht	pht version	Dataset Name	phv	Variable Name
6302	2	FUSION_Islet_Sample_Attributes	420835	FUSION_USE
6302	2	FUSION_Islet_Sample_Attributes	420836	USE_ME
6302	2	FUSION_Islet_Sample_Attributes	420837	REASON

4. **Retired Variables** (n=3 variables)

pht	Dataset Name	phv	version	Variable Name
6300	FUSION_Islet_Sample	290113	1	SAMPLE_USE
6301	FUSION_Islet_Subject_Phenotypes	290124	1	CauseOfDeath
6301	FUSION_Islet_Subject_Phenotypes	290125	1	CMV

Molecular Data Updates

See download components 'sample-info' and 'marker-info' for manifest of genotyped samples and files, and information about marker set used for genotyping (when available).

1. Array and imputed genotype calls are available in multi-sample VCF format ('genotype-calls-vcf').
2. phg001306.v1 beside array methylation data contains supplemental material from open chromatin experiment packed in phg001306.v1.NHGRI_FUSION_Islet_Methylation.sample-info.MULTI.tar

phg accession	Dataset name	Data type	Sample cnt	Subject cnt
phe000025.v1	NHGRI_FUSION_Islet	Seq_RNA_Expression	31	31
phe000025.v1	NHGRI_FUSION_Islet_eQTLs	Molecular_QTL	31	31
phg001304.v1	NHGRI_FUSION_Islet_ArrayGTs	Array_SNP	103	103
phg001305.v1	NHGRI_FUSION_Islet_Imputed	Imputation_SNP	103	103
phg001306.v1	NHGRI_FUSION_Islet_Methylation	Array_DNA_Methylation	41	33

Authorized Access (Individual Level Data and SRA Data)

Individual level data and Sequence Read Archive (SRA) data are available for download through the dbGaP Authorized Access System:

- http://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- ftp://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001188/phs001188.v2.p1

dbGaP Study Release Notes

Version 1 Data set release for NHGRI FUSION Islet now available

This release includes phenotype tables, SNP arrays, imputation data, molecular QTL, expression data, and RNA and chromatin sequences brokered through the Sequence Read Archive (SRA). Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): General Research Use (GRU)

	Phenotype	Array_SNP	Imputation_SNP _CNV	Molecular_QTL	Seq_DNA _Chromatin*	Seq_RNA*	Seq_RNA_Expression
subjects	33	33	33	31	2	31	31
samples	64	33	33	31	2	31	31

For a description of the SAMPLE_USE terms, please see:
https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi
*These data are brokered through the Sequence Read Archive (SRA). Please see Authorized Access instructions below.

Molecular Data

Two accessions containing molecular data, phg000918.v1 with array and imputed genotypes and phe000025.v1 with RNAseq expression and eQTLs are included in the release:

1. See corresponding download components 'sample-info' and 'marker-info' for manifests of the assayed samples and information about marker sets used for the assays.
2. Genotype calls are available in matrix format ('genotype-calls-vcf') as VCF file(s) with samples of same data sharing consent.
3. See phe000025.v1.NHGRI_FUSION_Islet.molecular-data-qc.MULTI.tar for details on RNAseq expression and eQTLs computations.

phg_name	data_type	markerset	sample_cnt	subject_cnt
phe000025.v1	NHGRI_FUSION_Islet	Seq_RNA_Expression	RNAseq_probe_set_grc37	31
phe000025.v1	NHGRI_FUSION_Islet_eQTLs	Molecular_QTL	RNAseq_probe_set_grc37	31
phg000918.v1	FUSION_Islet_Array	Array_SNP	HumanOmni2.5-4v1_H	33
phg000918.v1	FUSION_Islet_Imputation	Imputation_SNP_CNV	1000G_ref_panel	33

Authorized Access (Individual Level Data and SRA Data)

Individual level data and Sequence Read Archive (SRA) data are available for download through the dbGaP Authorized Access System:

- http://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- ftp://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001188/phs001188.v1.p1