INCLUDE
Data Coordinating Center

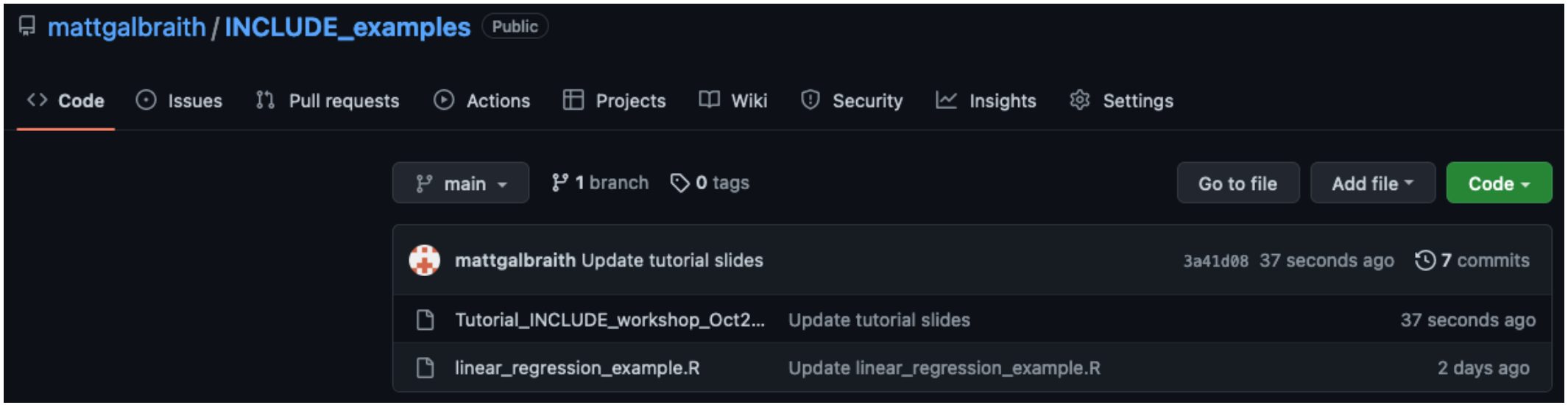
Analysis of INCLUDE data using CAVATICA

Matthew Galbraith

INCLUDE Workshop October 2022

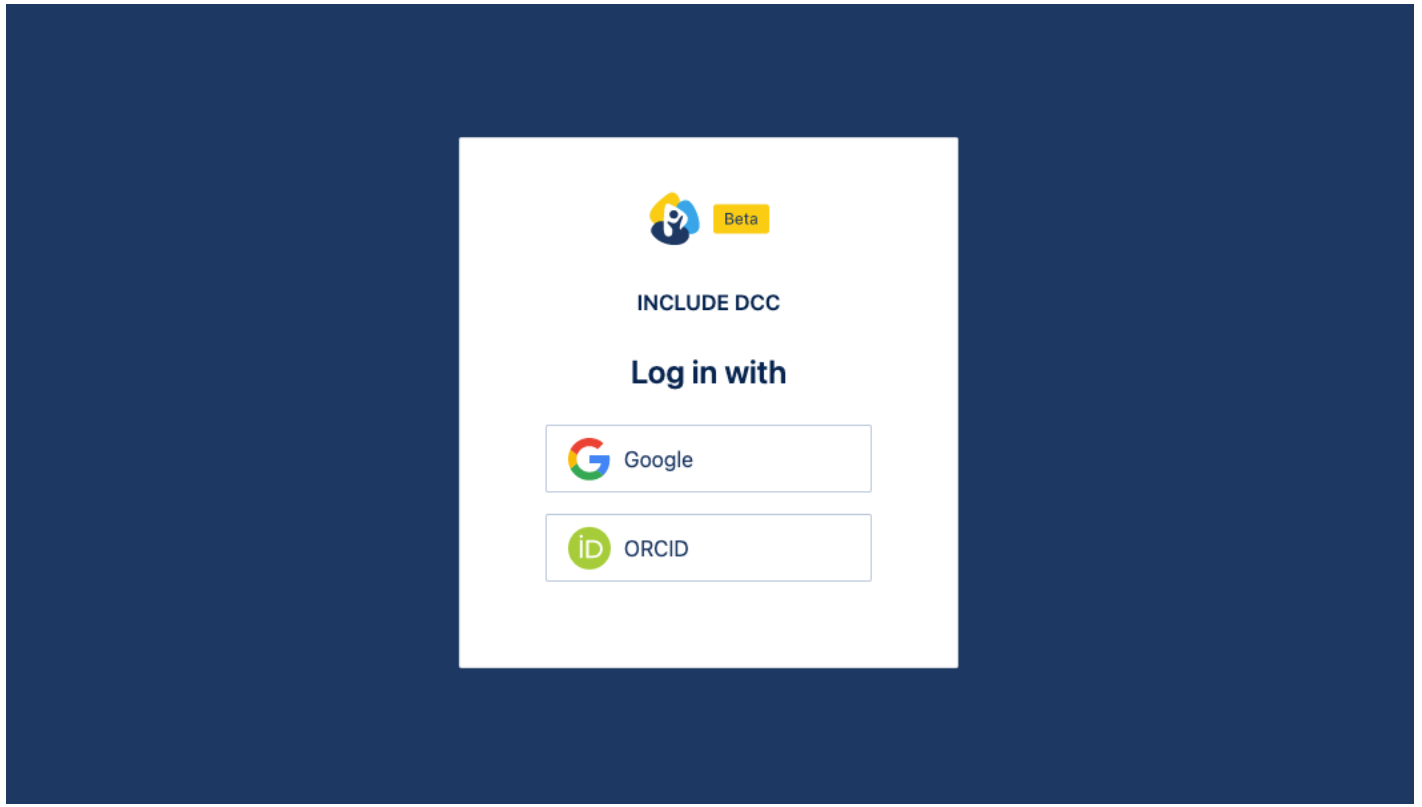
Linda Crnic Institute for Down Syndrome
UNIVERSITY OF COLORADO **ANSCHUTZ MEDICAL CAMPUS**

Tutorial slides and R code available at:
https://github.com/mattgalbraith/INCLUDE_examples/

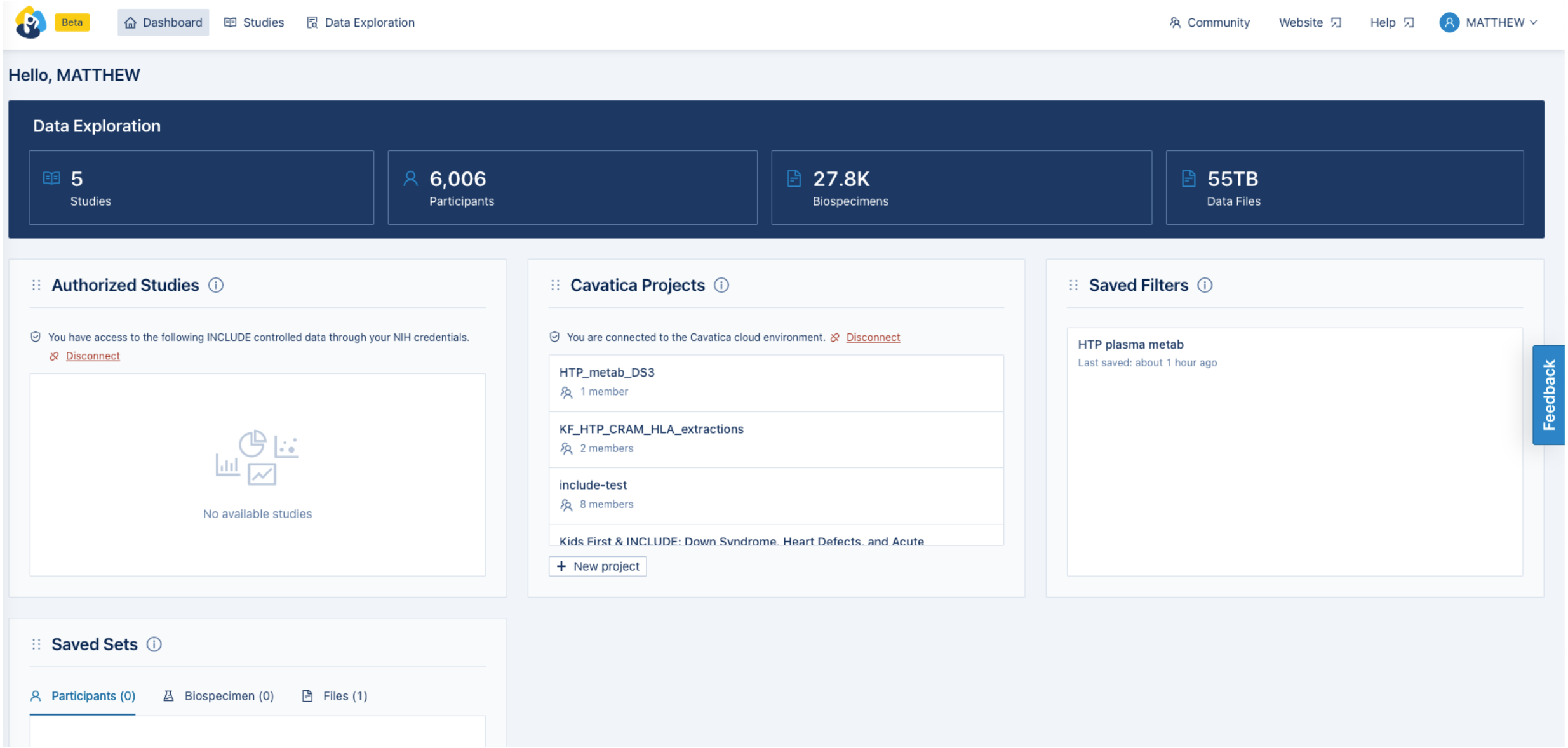


INCLUDE Data Hub

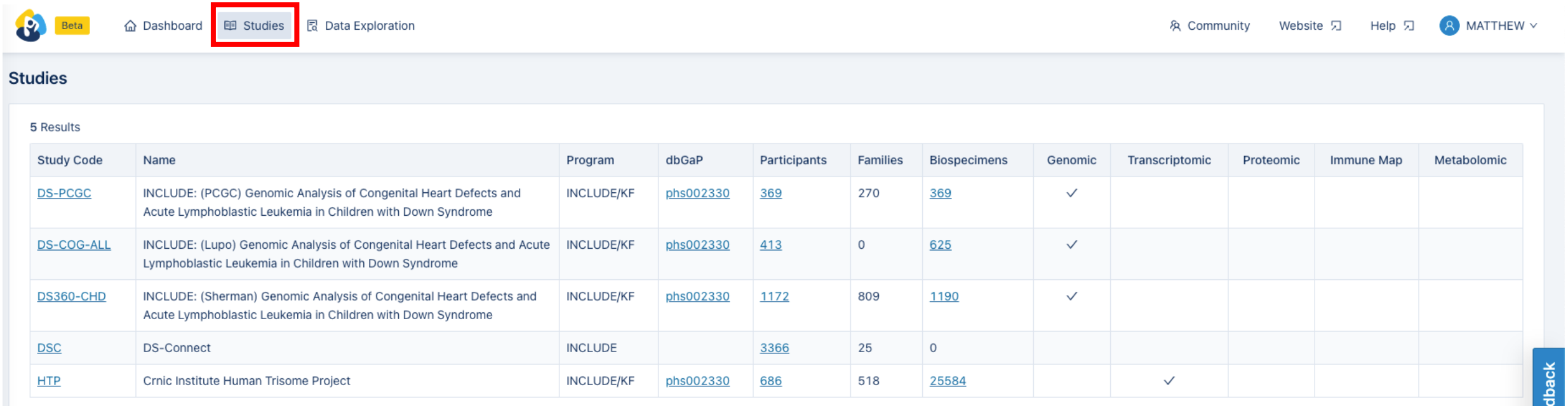
https://portal.includedcc.org/



INCLUDE Data Hub: Dashboard



INCLUDE Data Hub: Studies



INCLUDE Data Hub: Studies

Crnic Institute Human Trisome Project®

LINDA CRNIC INSTITUTE
HUMAN TRISOME PROJECT™
GLOBAL DOWN SYNDROME FOUNDATION

Visit Website

Introduction

The Crnic Institute Human Trisome Project® (HTP) is an in-depth study of people with Down syndrome using the latest technologies in precision medicine. The goal of the HTP is to enable advanced therapeutic approaches to enhance the quality of life and extend the lifespan of those with trisomy 21 through the study of the co-occurring conditions of Down syndrome.

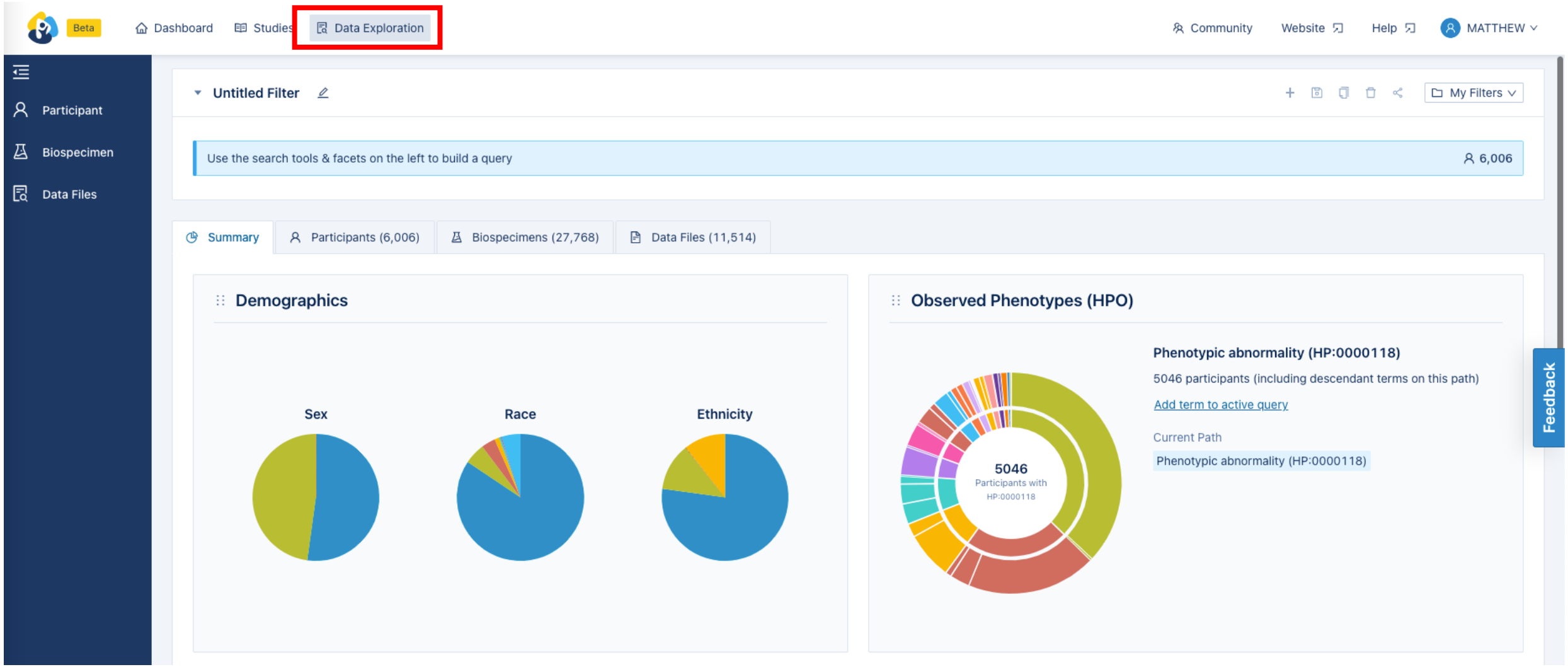
Principal Investigators

Espinosa, Joaquin M

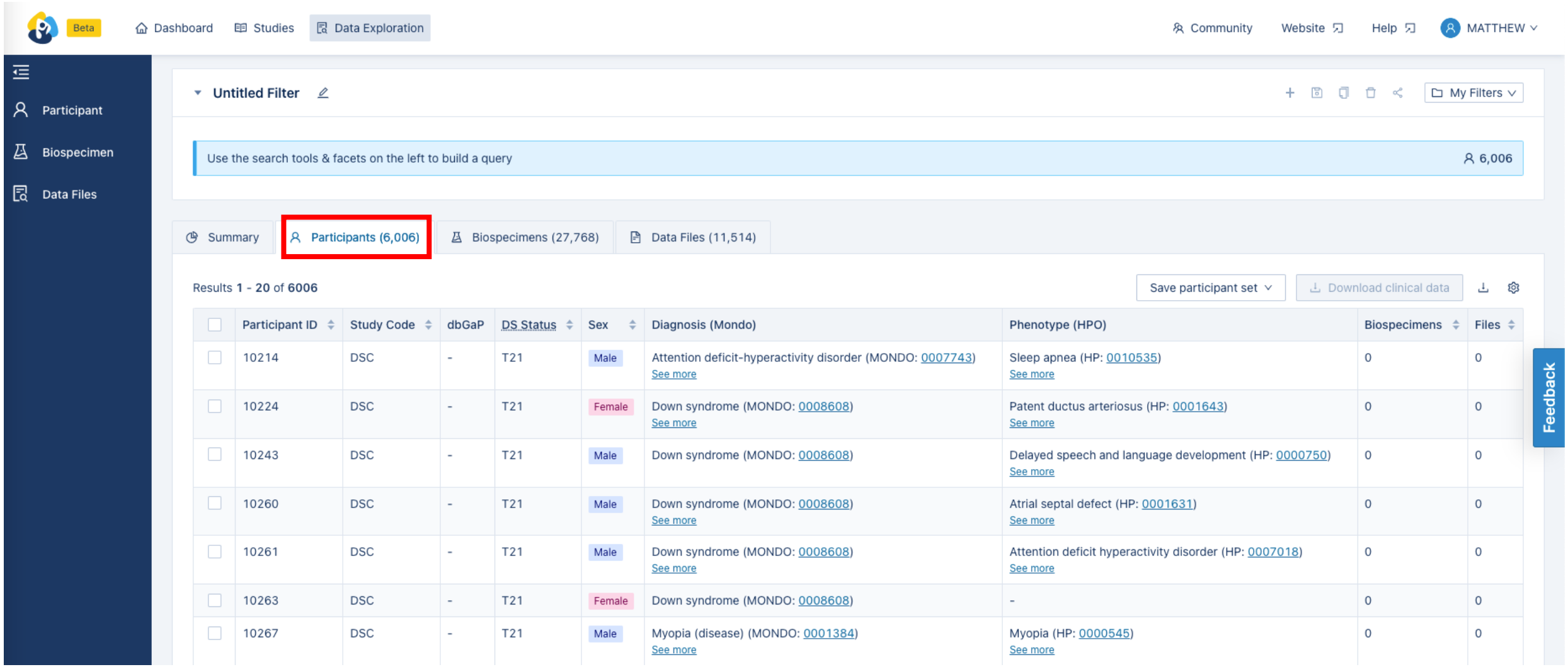
About

The Human Trisome Project (HTP) leverages a multidisciplinary team of biomedical researchers, clinicians and data scientists located across multiple departments, divisions, institutes and centers at the University of Colorado who work together toward a single goal: to decipher why people with trisomy 21 have a different disease spectrum, being predisposed to some medical conditions while being protected from others.

INCLUDE Data Hub: Data Exploration - Summary

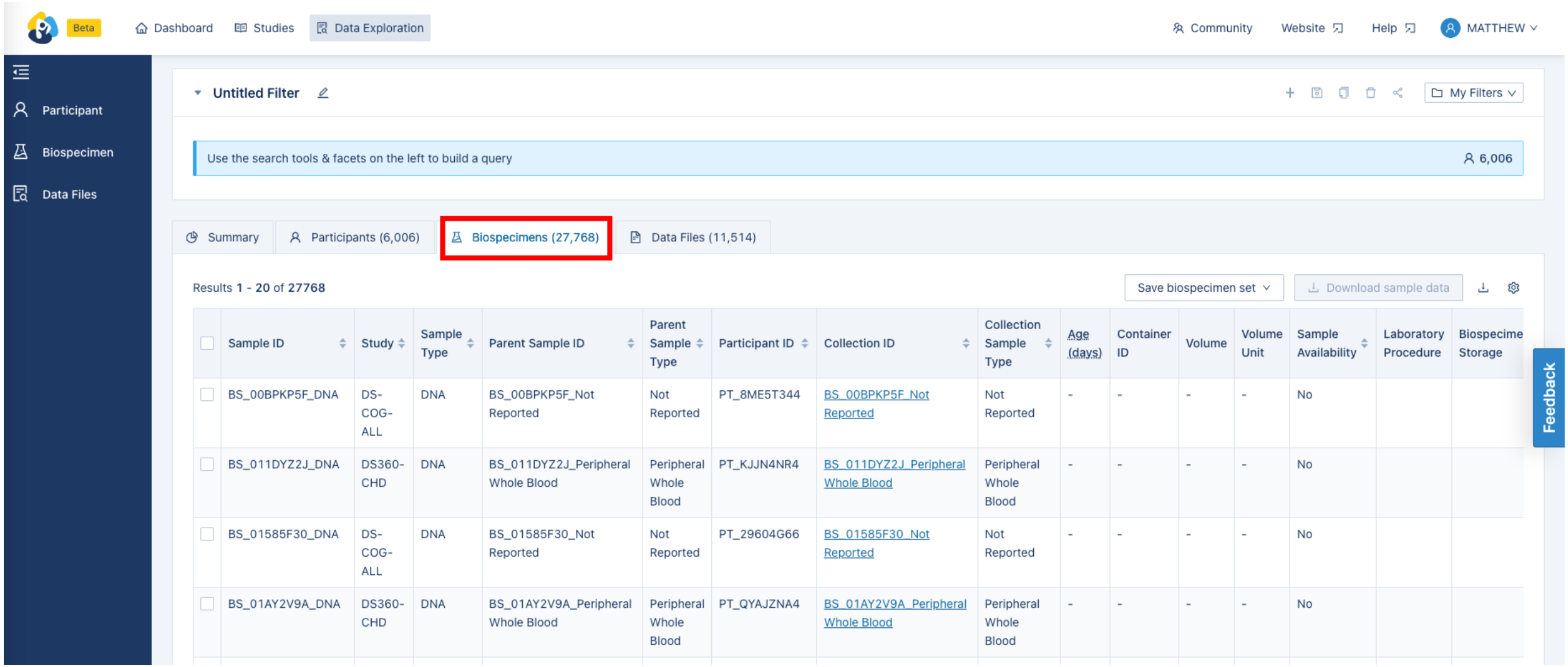


INCLUDE Data Hub: Participants



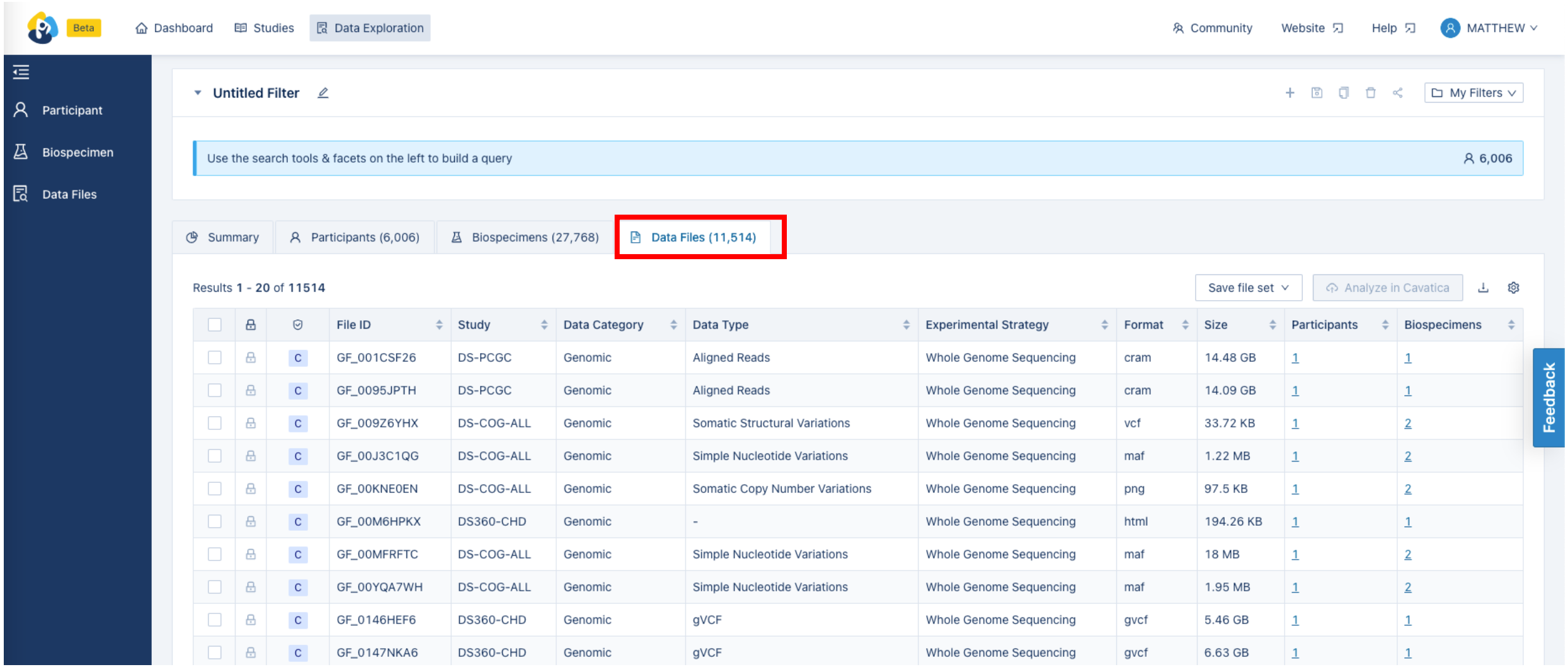
Participant ID	Study Code	dbGaP	DS Status	Sex	Diagnosis (Mondo)	Phenotype (HPO)	Biospecimens	Files
10214	DSC	-	T21	Male	Attention deficit-hyperactivity disorder (MONDO: 0007743) See more	Sleep apnea (HP: 0010535) See more	0	0
10224	DSC	-	T21	Female	Down syndrome (MONDO: 0008608) See more	Patent ductus arteriosus (HP: 0001643) See more	0	0
10243	DSC	-	T21	Male	Down syndrome (MONDO: 0008608)	Delayed speech and language development (HP: 0000750) See more	0	0
10260	DSC	-	T21	Male	Down syndrome (MONDO: 0008608) See more	Atrial septal defect (HP: 0001631) See more	0	0
10261	DSC	-	T21	Male	Down syndrome (MONDO: 0008608) See more	Attention deficit hyperactivity disorder (HP: 0007018) See more	0	0
10263	DSC	-	T21	Female	Down syndrome (MONDO: 0008608)	-	0	0
10267	DSC	-	T21	Male	Myopia (disease) (MONDO: 0001384) See more	Myopia (HP: 0000545) See more	0	0

INCLUDE Data Hub: Biospecimens



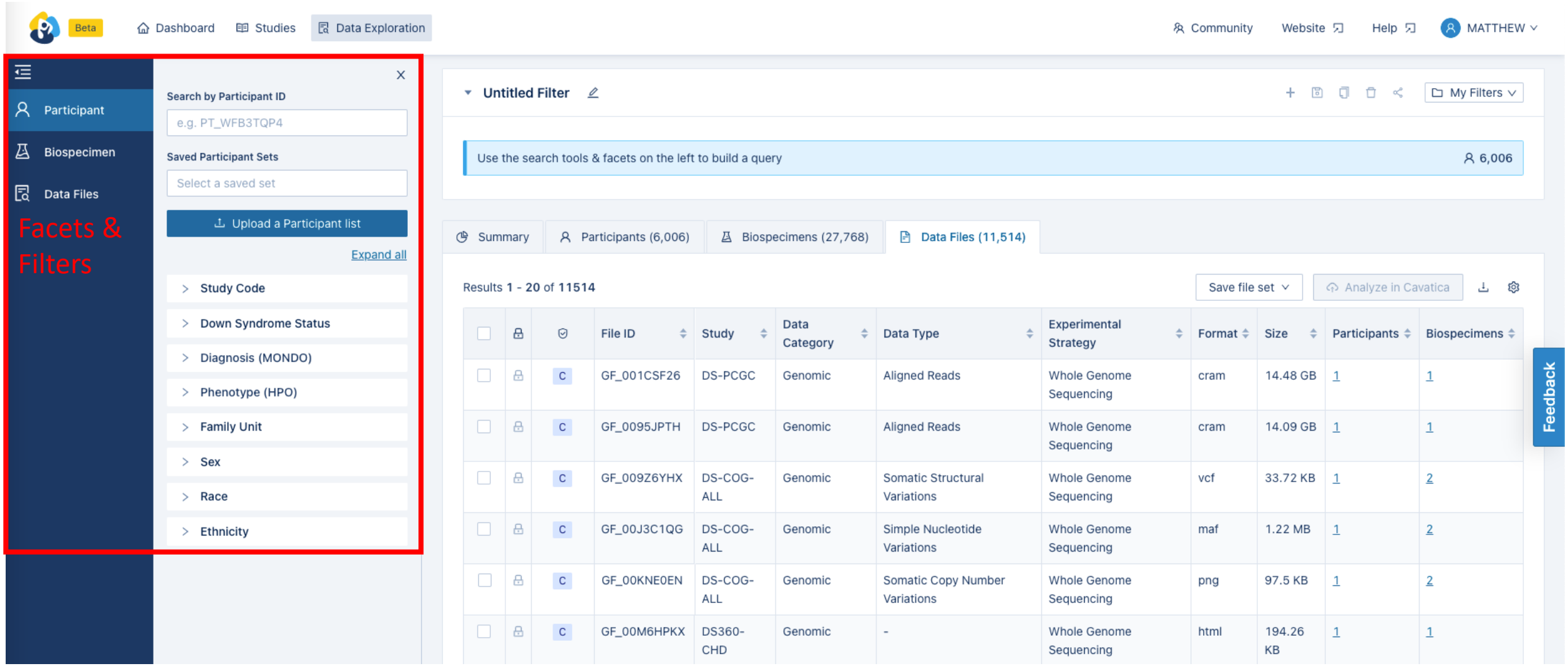
Sample ID	Study	Sample Type	Parent Sample ID	Parent Sample Type	Participant ID	Collection ID	Collection Sample Type	Age (days)	Container ID	Volume	Volume Unit	Sample Availability	Laboratory Procedure	Biospecime Storage
BS_00BPKP5F_DNA	DS-COG-ALL	DNA	BS_00BPKP5F_Not Reported	Not Reported	PT_8ME5T344	BS_00BPKP5F_Not Reported	Not Reported	-	-	-	-	No		
BS_011DYZ2J_DNA	DS360-CHD	DNA	BS_011DYZ2J_Peripheral Whole Blood	Peripheral Whole Blood	PT_KJJN4NR4	BS_011DYZ2J_Peripheral Whole Blood	Peripheral Whole Blood	-	-	-	-	No		
BS_01585F30_DNA	DS-COG-ALL	DNA	BS_01585F30_Not Reported	Not Reported	PT_29604G66	BS_01585F30_Not Reported	Not Reported	-	-	-	-	No		
BS_01AY2V9A_DNA	DS360-CHD	DNA	BS_01AY2V9A_Peripheral Whole Blood	Peripheral Whole Blood	PT_QYAJZNA4	BS_01AY2V9A_Peripheral Whole Blood	Peripheral Whole Blood	-	-	-	-	No		

INCLUDE Data Hub: Data Files



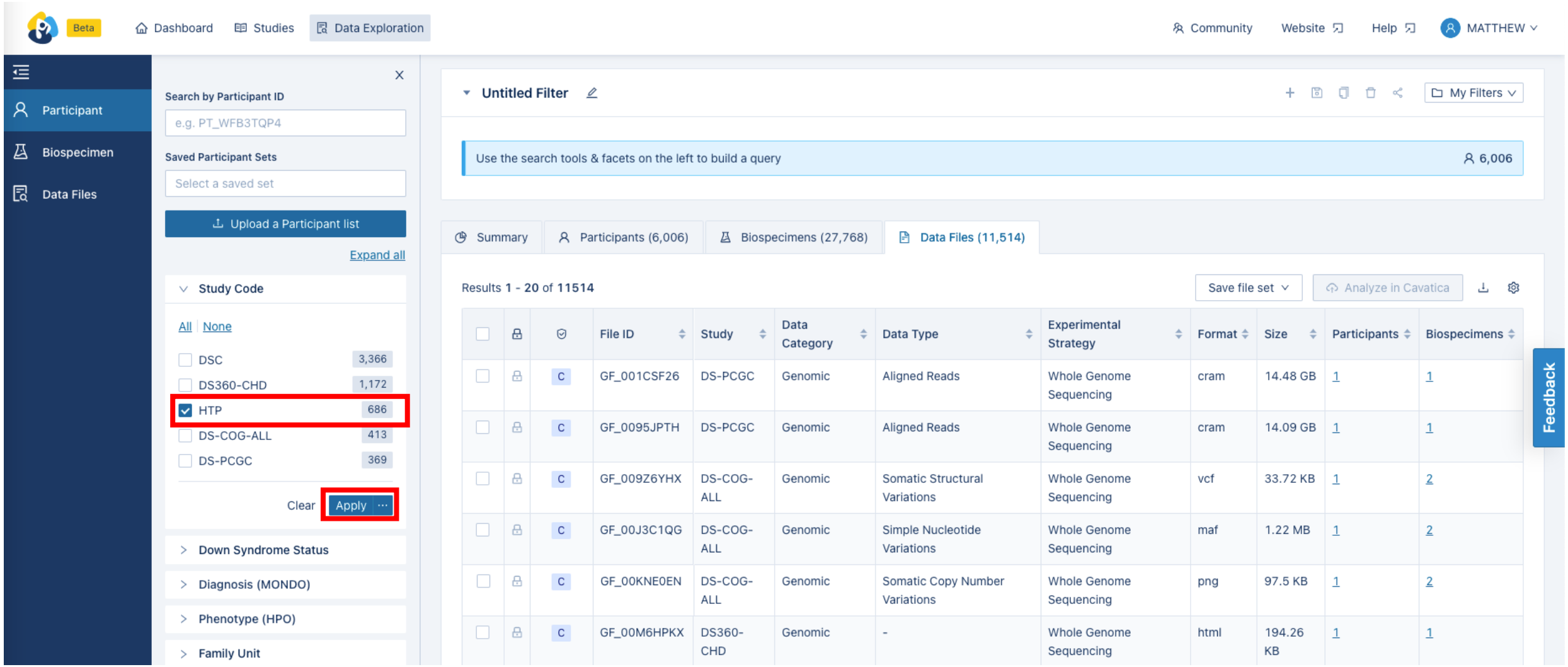
File ID	Study	Data Category	Data Type	Experimental Strategy	Format	Size	Participants	Biospecimens
GF_001CSF26	DS-PCGC	Genomic	Aligned Reads	Whole Genome Sequencing	cram	14.48 GB	1	1
GF_0095JPTH	DS-PCGC	Genomic	Aligned Reads	Whole Genome Sequencing	cram	14.09 GB	1	1
GF_009Z6YHX	DS-COG-ALL	Genomic	Somatic Structural Variations	Whole Genome Sequencing	vcf	33.72 KB	1	2
GF_00J3C1QG	DS-COG-ALL	Genomic	Simple Nucleotide Variations	Whole Genome Sequencing	maf	1.22 MB	1	2
GF_00KNE0EN	DS-COG-ALL	Genomic	Somatic Copy Number Variations	Whole Genome Sequencing	png	97.5 KB	1	2
GF_00M6HPKX	DS360-CHD	Genomic	-	Whole Genome Sequencing	html	194.26 KB	1	1
GF_00MFRFTC	DS-COG-ALL	Genomic	Simple Nucleotide Variations	Whole Genome Sequencing	maf	18 MB	1	2
GF_00YQA7WH	DS-COG-ALL	Genomic	Simple Nucleotide Variations	Whole Genome Sequencing	maf	1.95 MB	1	2
GF_0146HEF6	DS360-CHD	Genomic	gVCF	Whole Genome Sequencing	gvcf	5.46 GB	1	1
GF_0147NKA6	DS360-CHD	Genomic	gVCF	Whole Genome Sequencing	gvcf	6.63 GB	1	1

INCLUDE Data Hub: Filtering



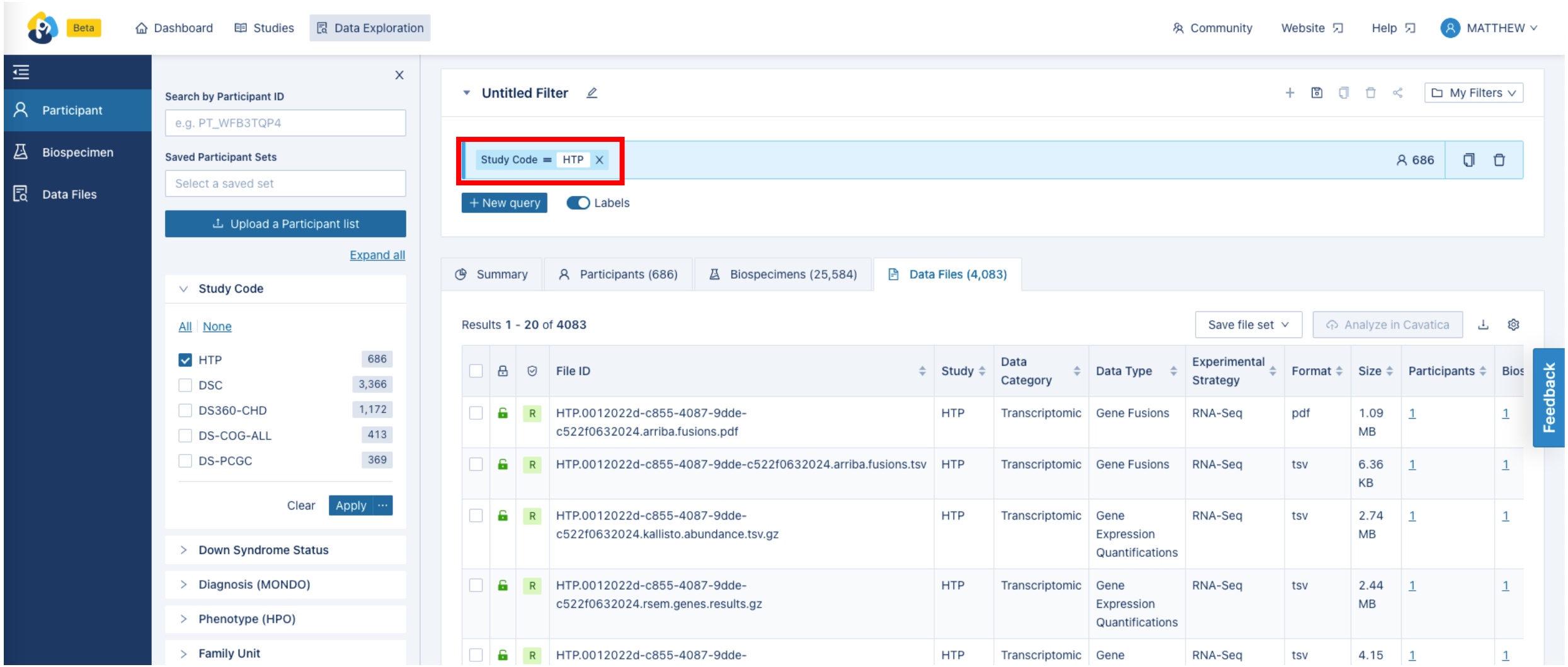
File ID	Study	Data Category	Data Type	Experimental Strategy	Format	Size	Participants	Biospecimens
GF_001CSF26	DS-PCGC	Genomic	Aligned Reads	Whole Genome Sequencing	cram	14.48 GB	1	1
GF_0095JPTH	DS-PCGC	Genomic	Aligned Reads	Whole Genome Sequencing	cram	14.09 GB	1	1
GF_009Z6YHX	DS-COG-ALL	Genomic	Somatic Structural Variations	Whole Genome Sequencing	vcf	33.72 KB	1	2
GF_00J3C1QG	DS-COG-ALL	Genomic	Simple Nucleotide Variations	Whole Genome Sequencing	maf	1.22 MB	1	2
GF_00KNE0EN	DS-COG-ALL	Genomic	Somatic Copy Number Variations	Whole Genome Sequencing	png	97.5 KB	1	2
GF_00M6HPKX	DS360-CHD	Genomic	-	Whole Genome Sequencing	html	194.26 KB	1	1

INCLUDE Data Hub: Filtering by Study Code

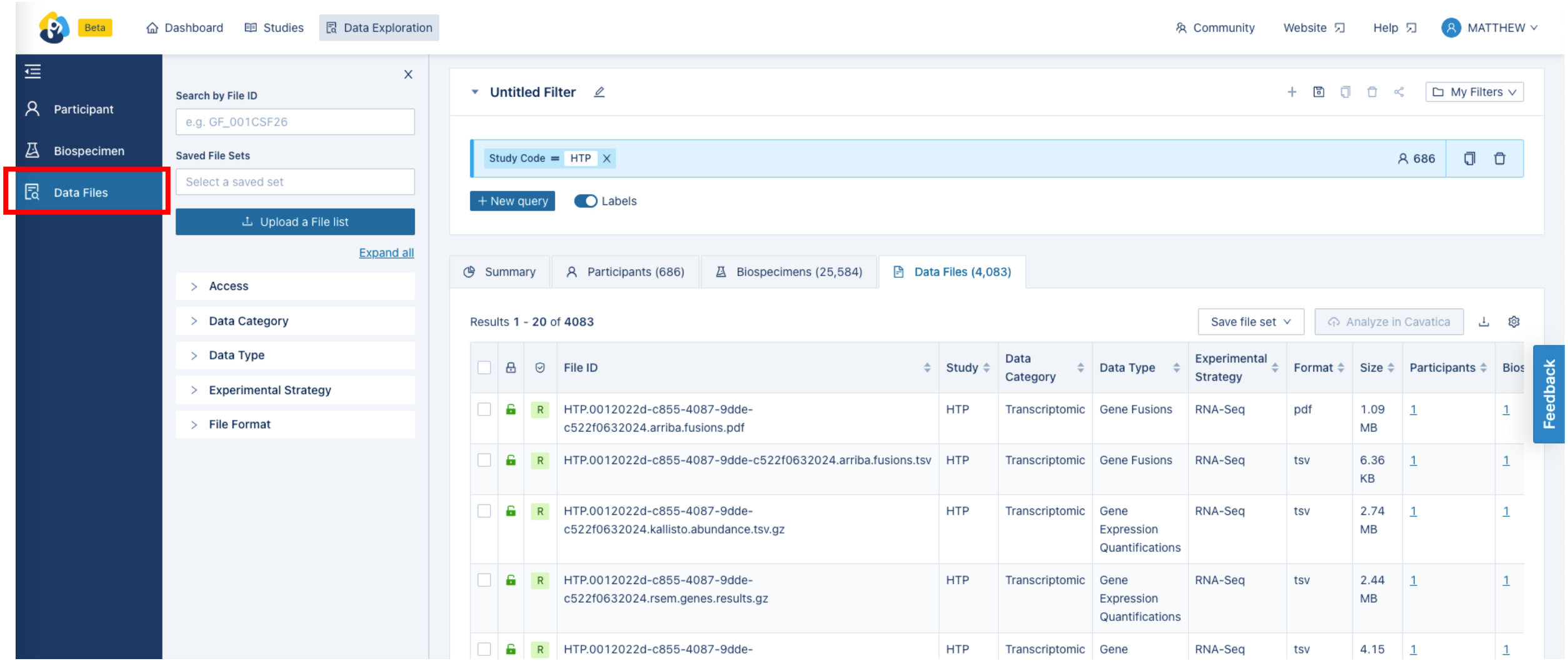


File ID	Study	Data Category	Data Type	Experimental Strategy	Format	Size	Participants	Biospecimens
GF_001CSF26	DS-PCGC	Genomic	Aligned Reads	Whole Genome Sequencing	cram	14.48 GB	1	1
GF_0095JPTH	DS-PCGC	Genomic	Aligned Reads	Whole Genome Sequencing	cram	14.09 GB	1	1
GF_009Z6YHX	DS-COG-ALL	Genomic	Somatic Structural Variations	Whole Genome Sequencing	vcf	33.72 KB	1	2
GF_00J3C1QG	DS-COG-ALL	Genomic	Simple Nucleotide Variations	Whole Genome Sequencing	maf	1.22 MB	1	2
GF_00KNE0EN	DS-COG-ALL	Genomic	Somatic Copy Number Variations	Whole Genome Sequencing	png	97.5 KB	1	2
GF_00M6HPKX	DS360-CHD	Genomic	-	Whole Genome Sequencing	html	194.26 KB	1	1

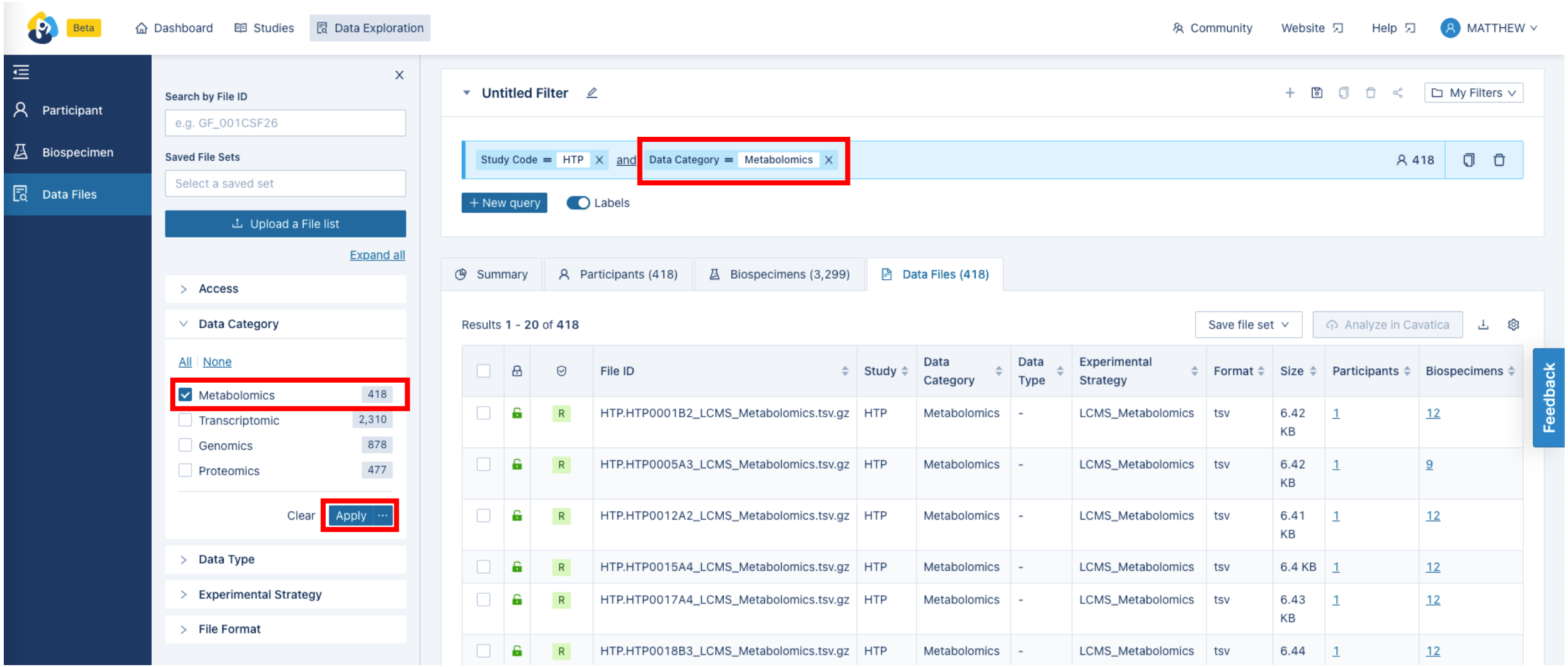
INCLUDE Data Hub: Filtering by Study Code



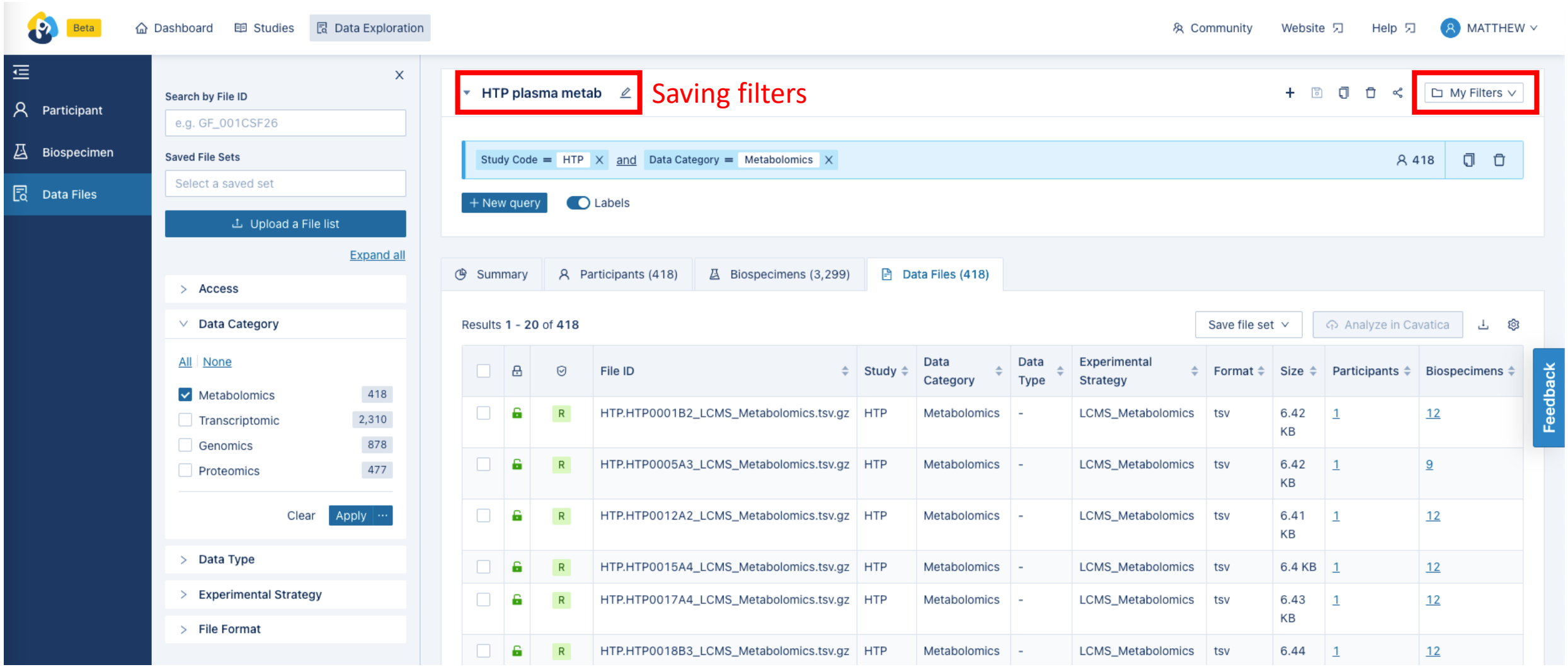
INCLUDE Data Hub: Filtering Data Files



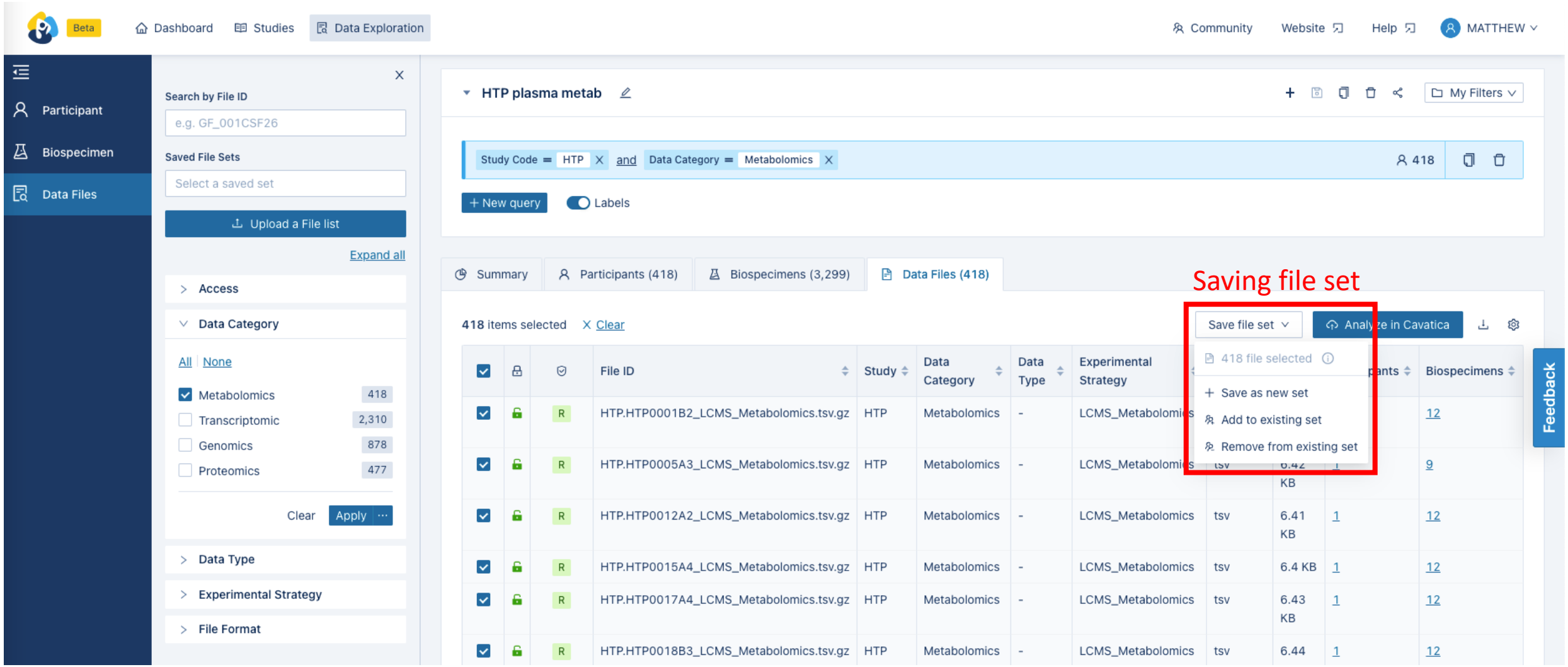
INCLUDE Data Hub: Filtering Data Files



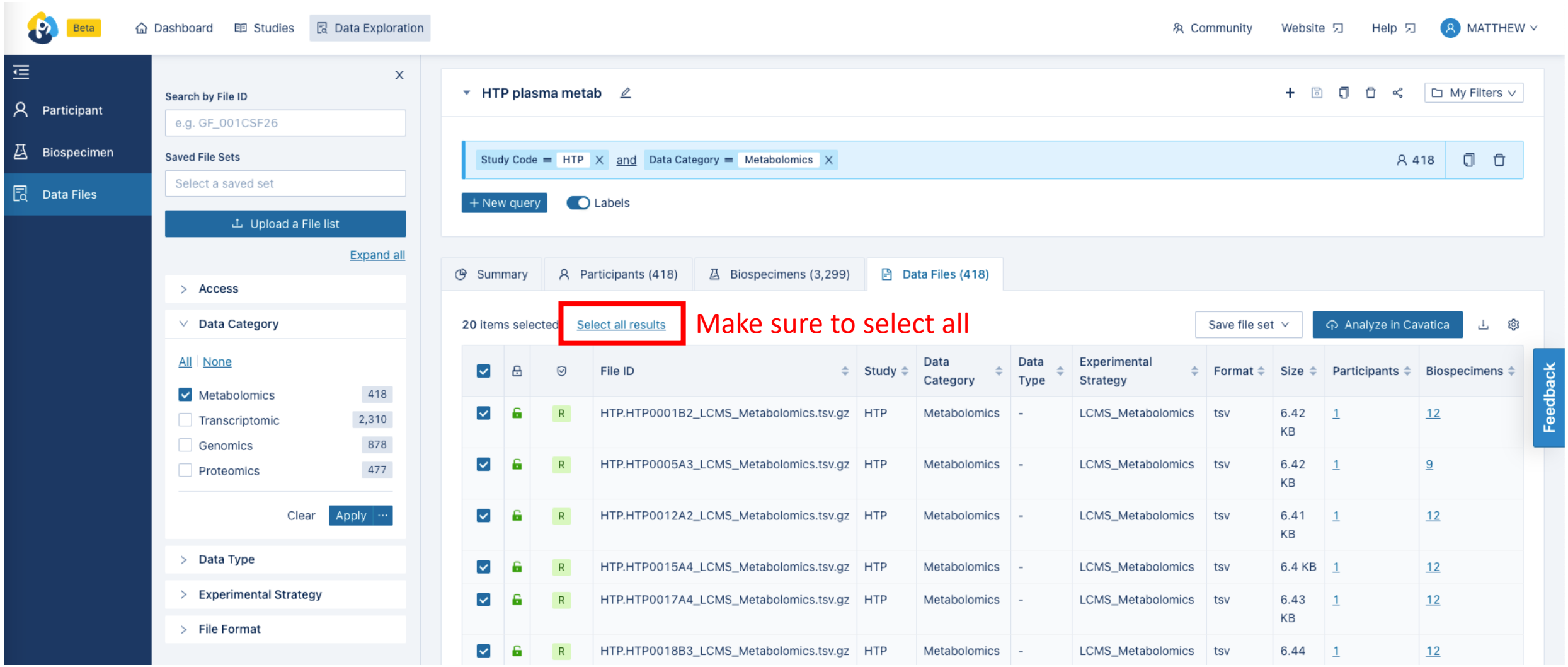
INCLUDE Data Hub: Saving a filter



INCLUDE Data Hub: Saving a file set

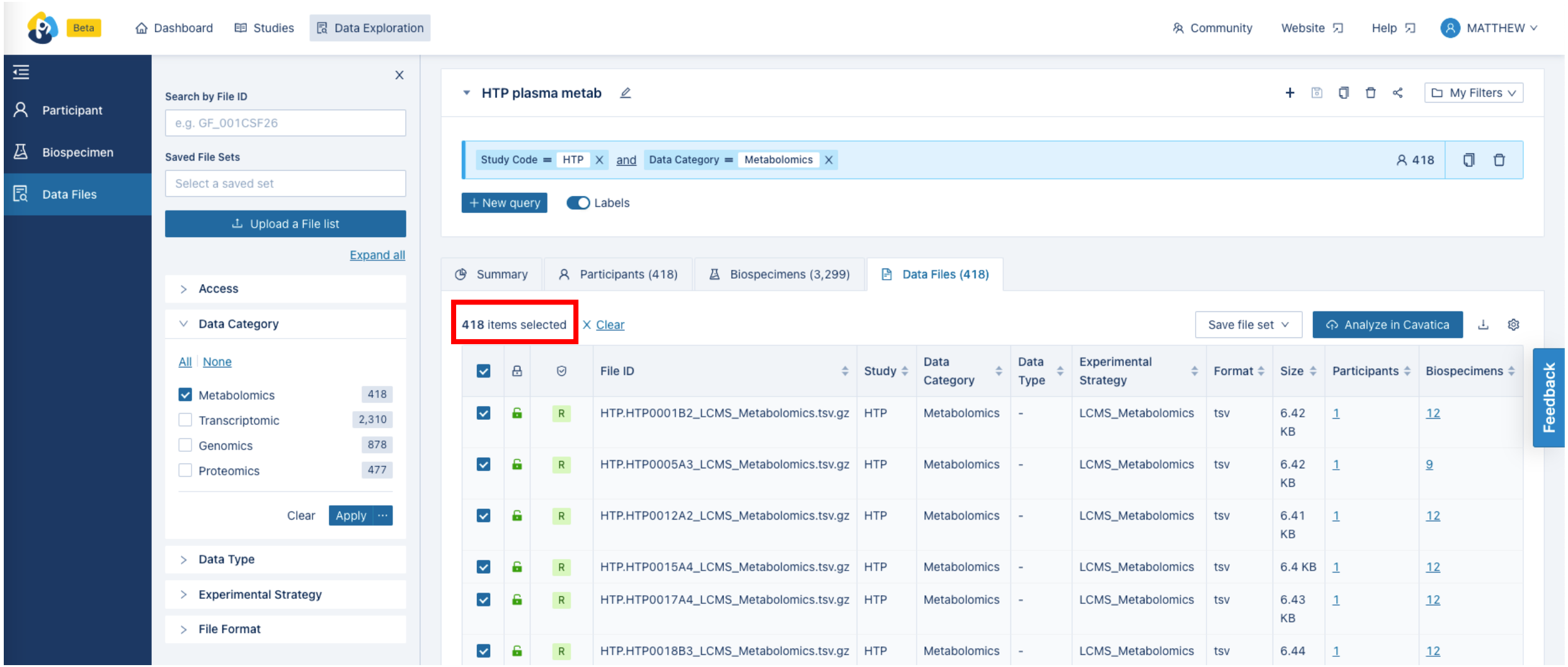


INCLUDE Data Hub: Selecting files

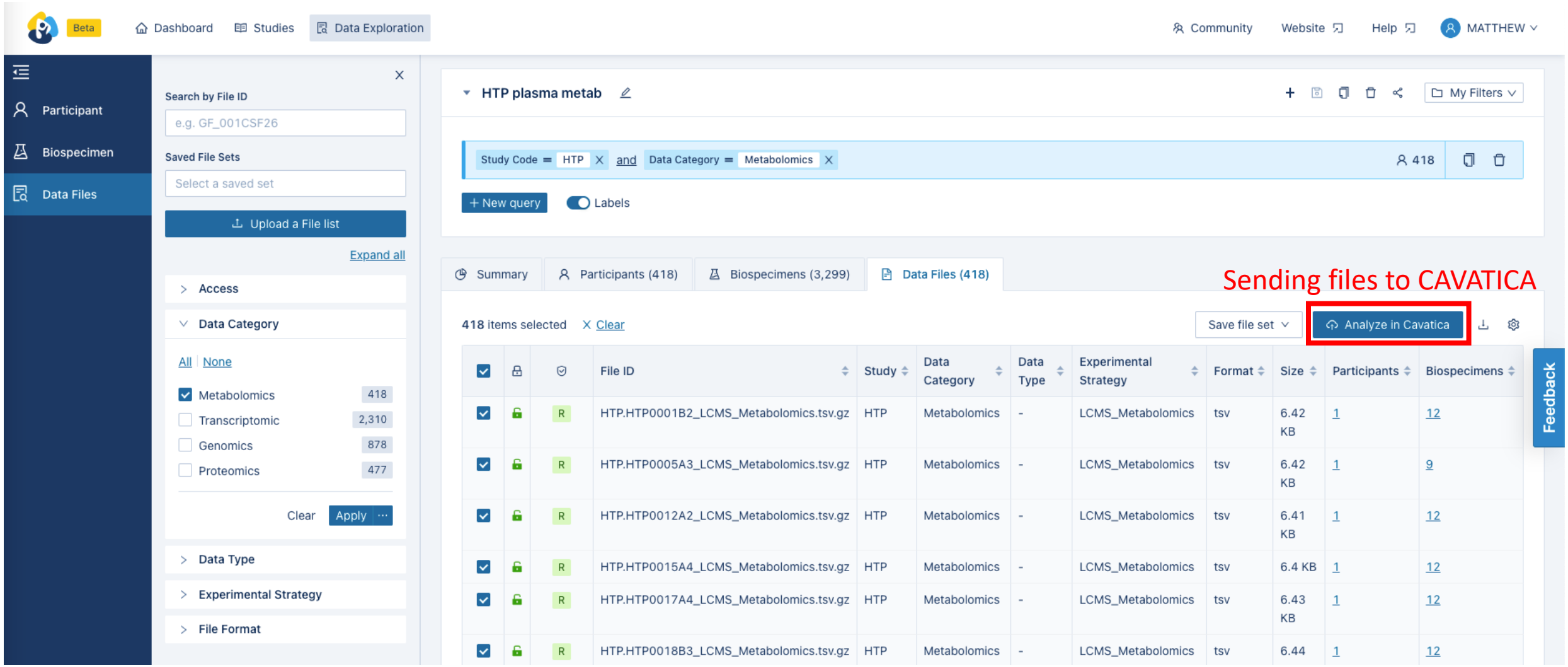


File ID	Study	Data Category	Data Type	Experimental Strategy	Format	Size	Participants	Biospecimens
HTP.HTP0001B2_LCMS_Metabolomics.tsv.gz	HTP	Metabolomics	-	LCMS_Metabolomics	tsv	6.42 KB	1	12
HTP.HTP0005A3_LCMS_Metabolomics.tsv.gz	HTP	Metabolomics	-	LCMS_Metabolomics	tsv	6.42 KB	1	9
HTP.HTP0012A2_LCMS_Metabolomics.tsv.gz	HTP	Metabolomics	-	LCMS_Metabolomics	tsv	6.41 KB	1	12
HTP.HTP0015A4_LCMS_Metabolomics.tsv.gz	HTP	Metabolomics	-	LCMS_Metabolomics	tsv	6.4 KB	1	12
HTP.HTP0017A4_LCMS_Metabolomics.tsv.gz	HTP	Metabolomics	-	LCMS_Metabolomics	tsv	6.43 KB	1	12
HTP.HTP0018B3_LCMS_Metabolomics.tsv.gz	HTP	Metabolomics	-	LCMS_Metabolomics	tsv	6.44	1	12

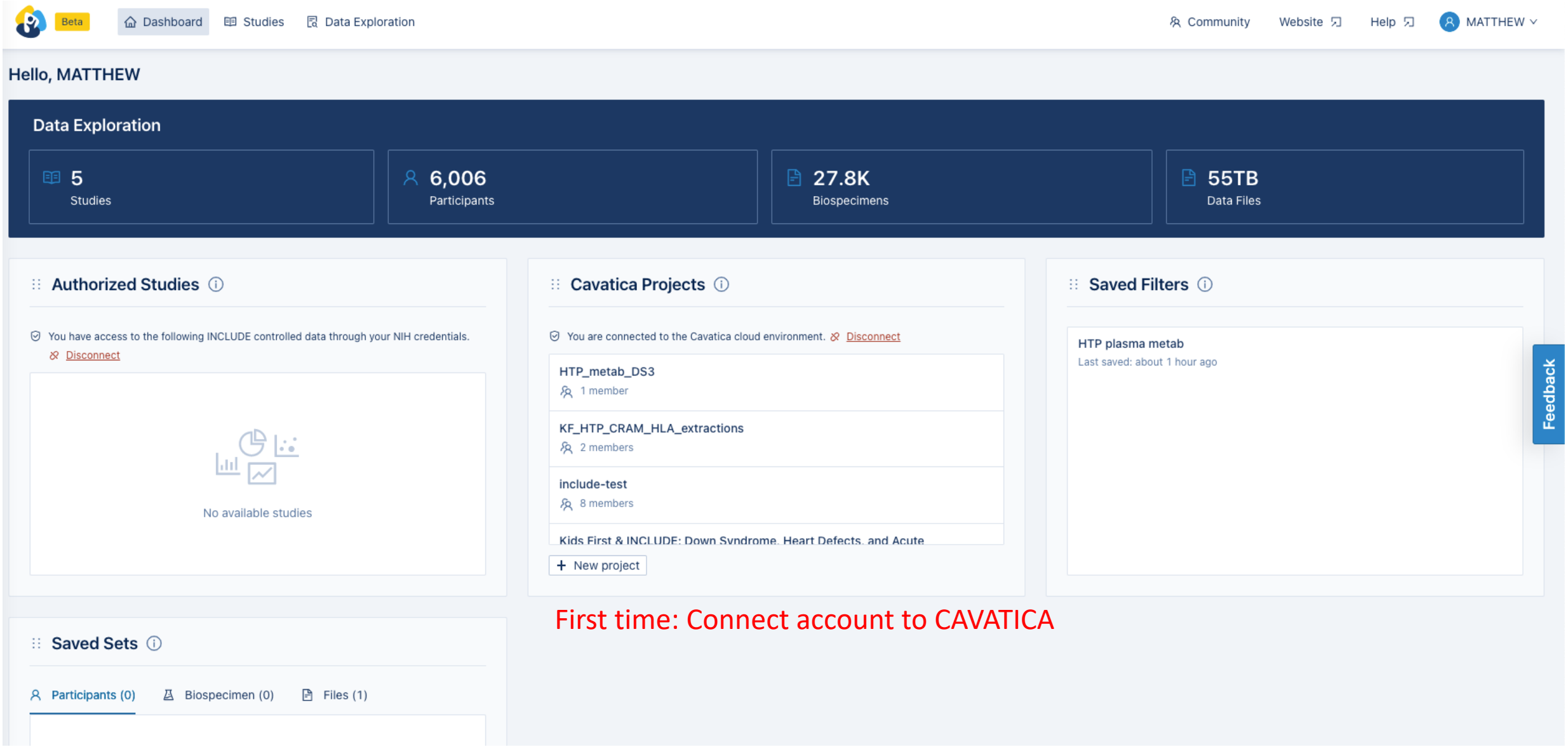
INCLUDE Data Hub: Selecting files



INCLUDE Data Hub: Sending files to CAVATICA for analysis



INCLUDE Data Hub: Sending files to CAVATICA for analysis

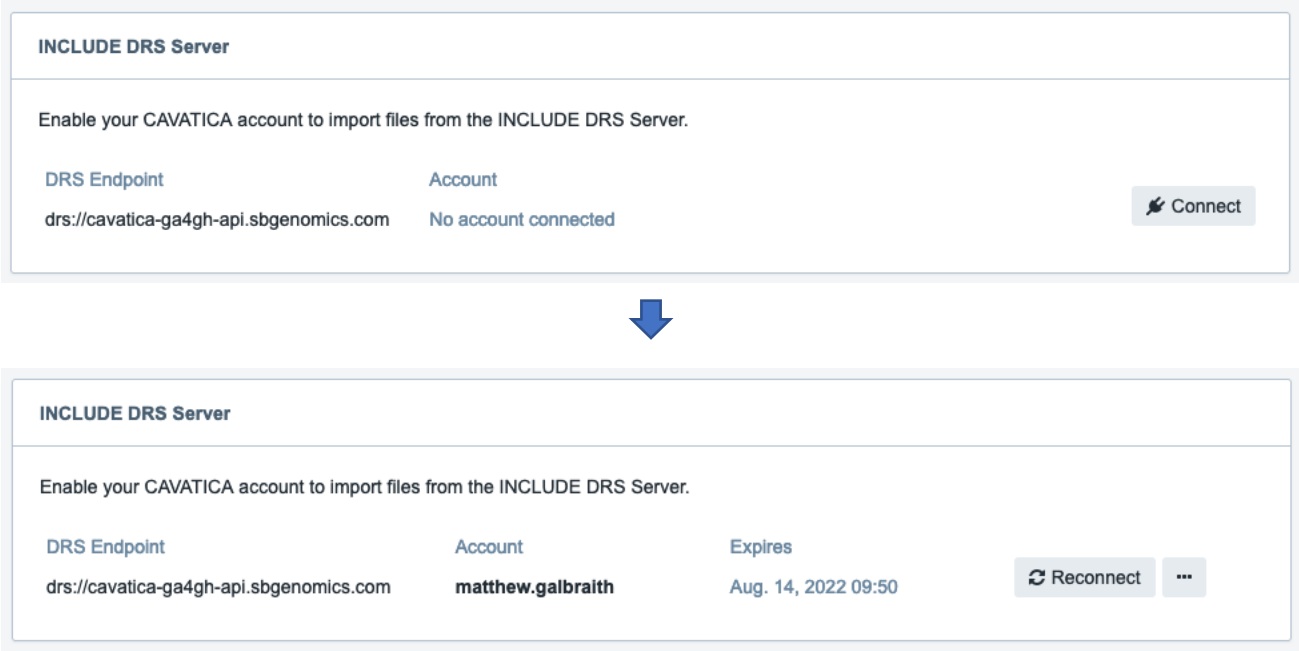


First time: Connect account to CAVATICA

CAVATICA: Sending files to CAVATICA for analysis

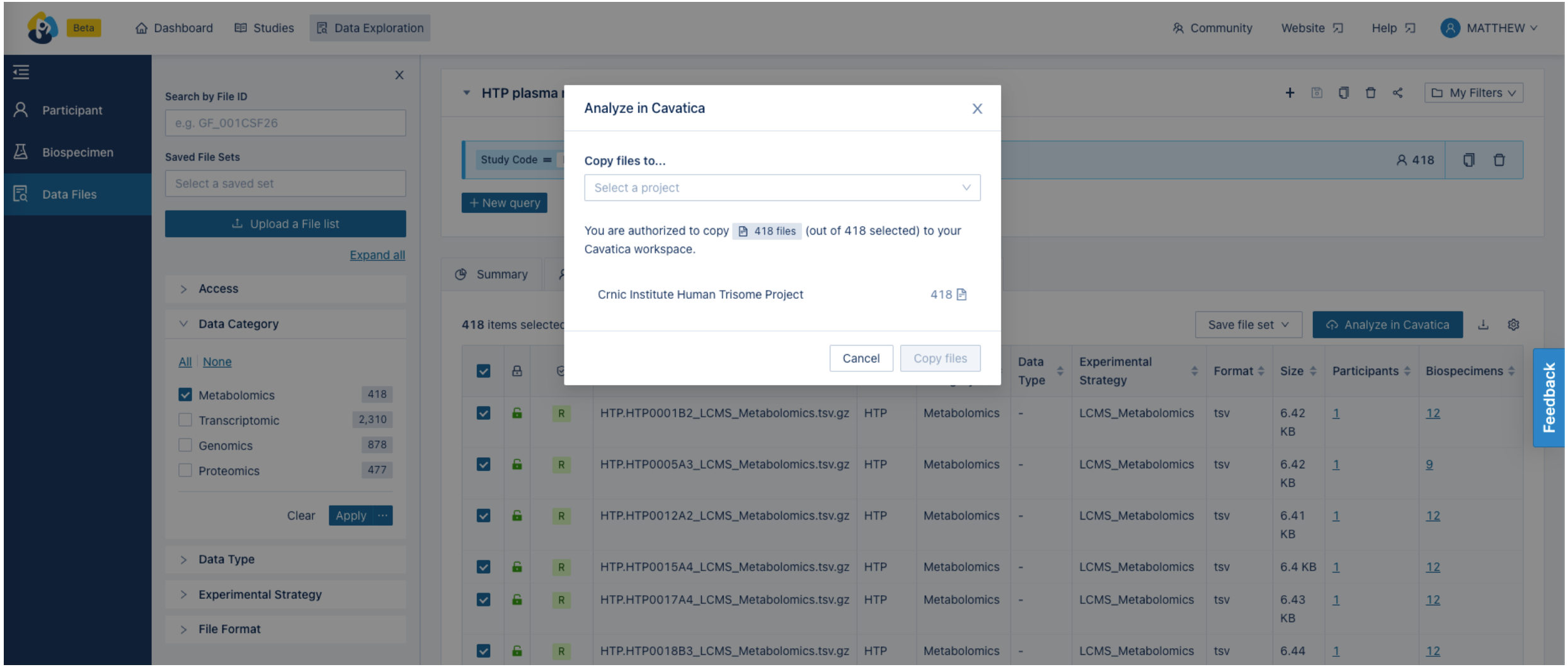
Ensure the INCLUDE DRS Server is enabled before trying to send files!

Account Settings --> Datasets

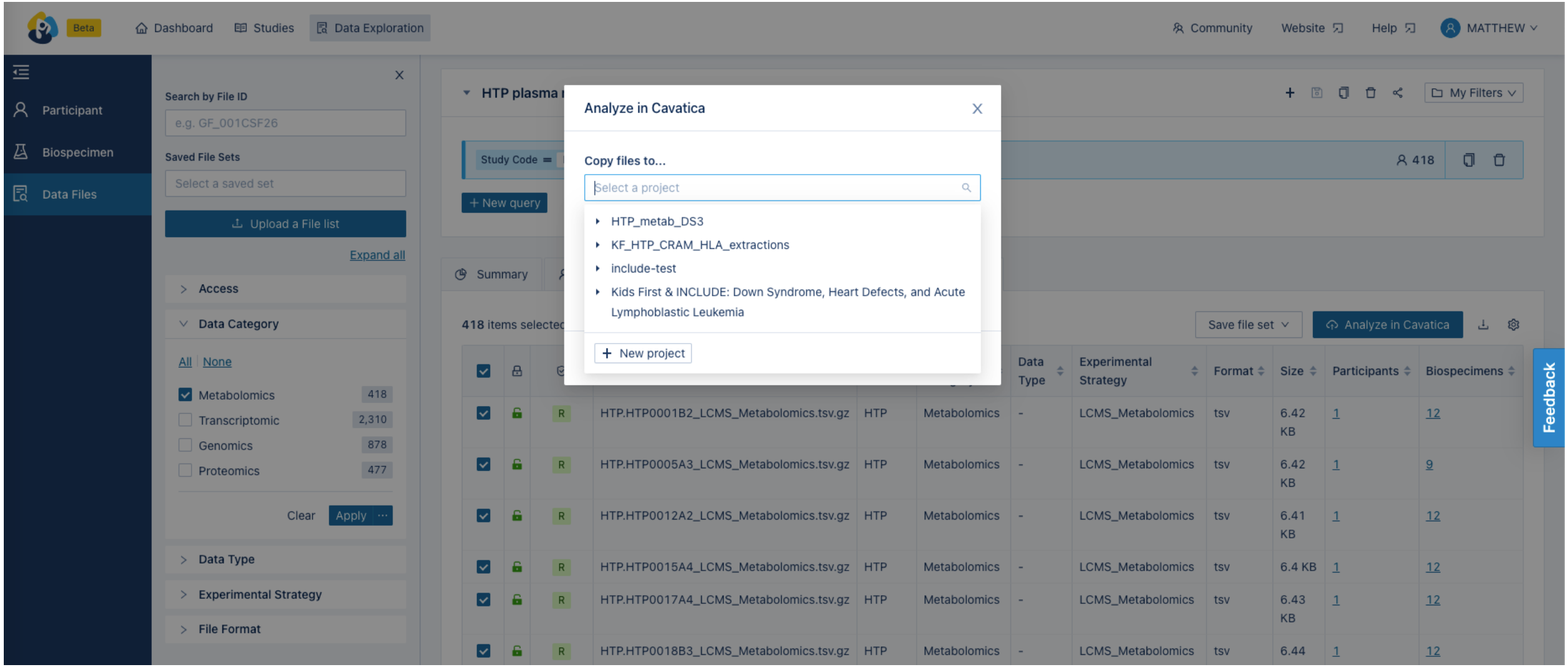


You may also need to apply for access to certain data files (eg dbGAP)
see https://help.includedcc.org/docs/applying-for-access

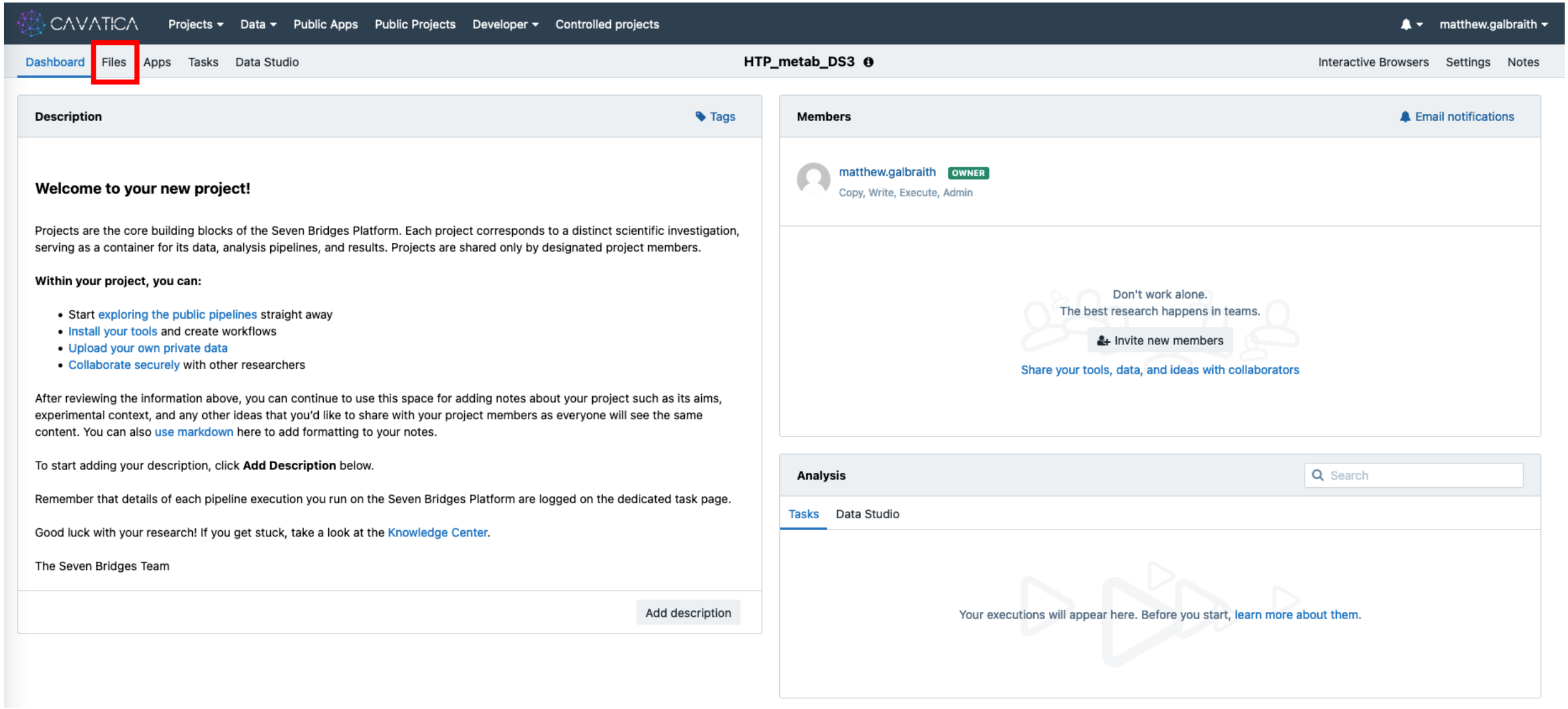
INCLUDE Data Hub: Sending files to CAVATICA for analysis



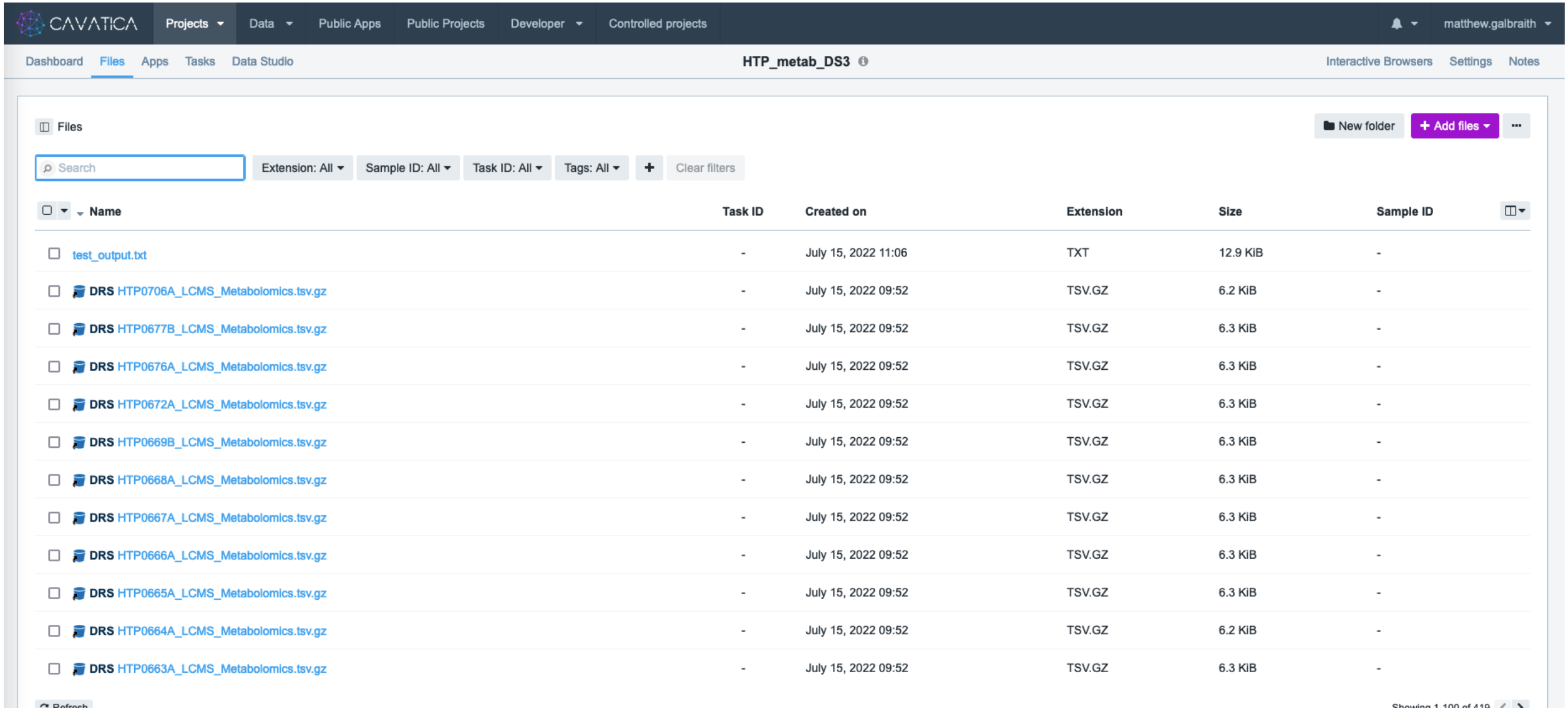
INCLUDE Data Hub: Sending files to CAVATICA for analysis



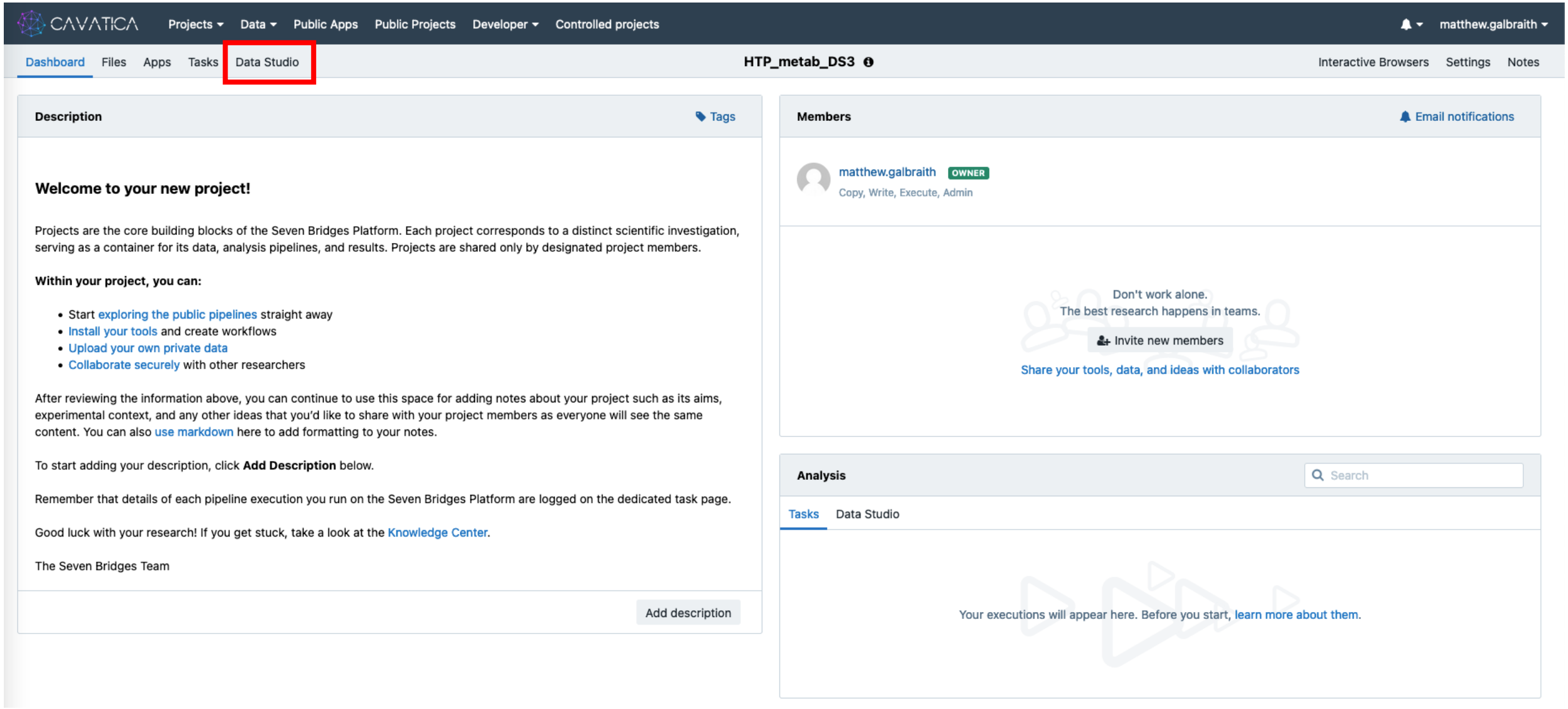
CAVATICA: Project dashboard



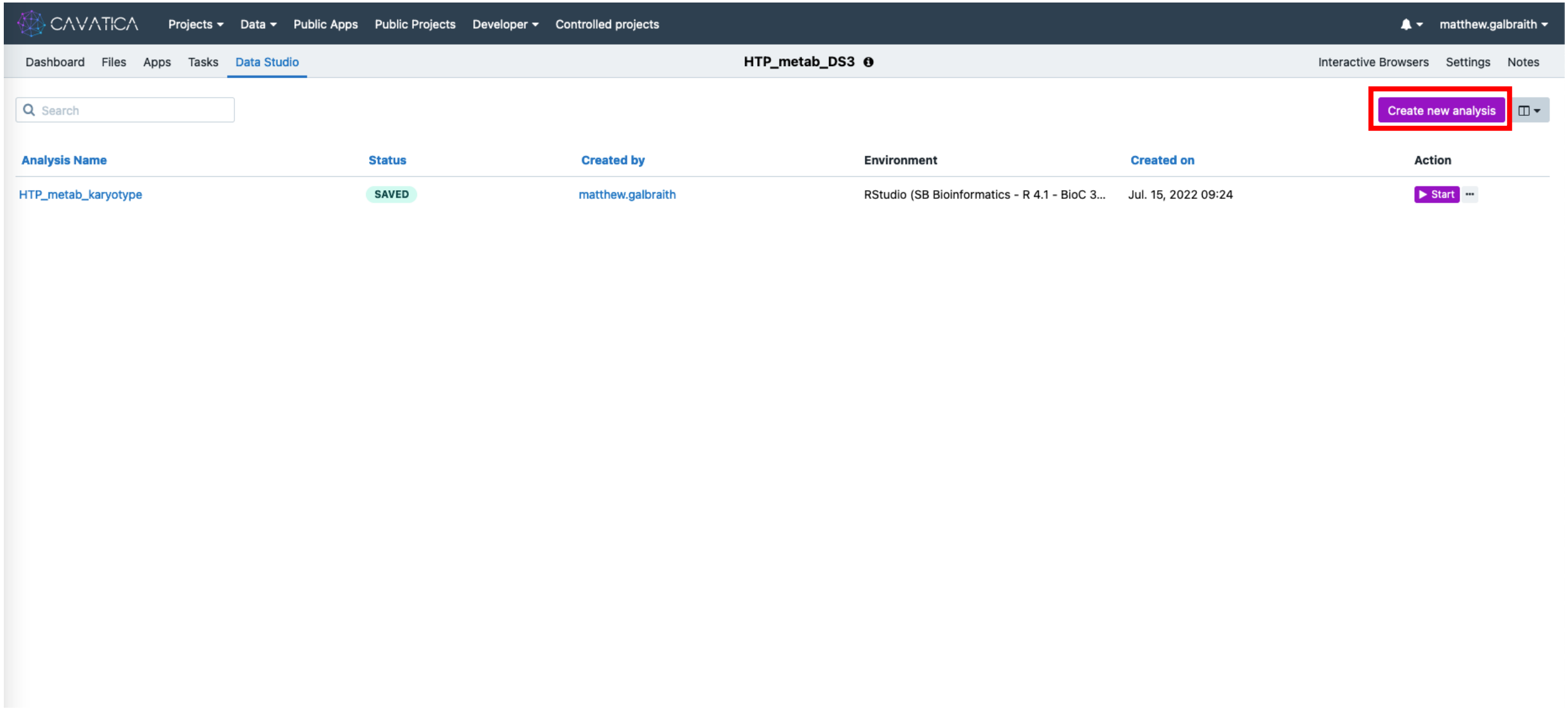
CAVATICA: Files



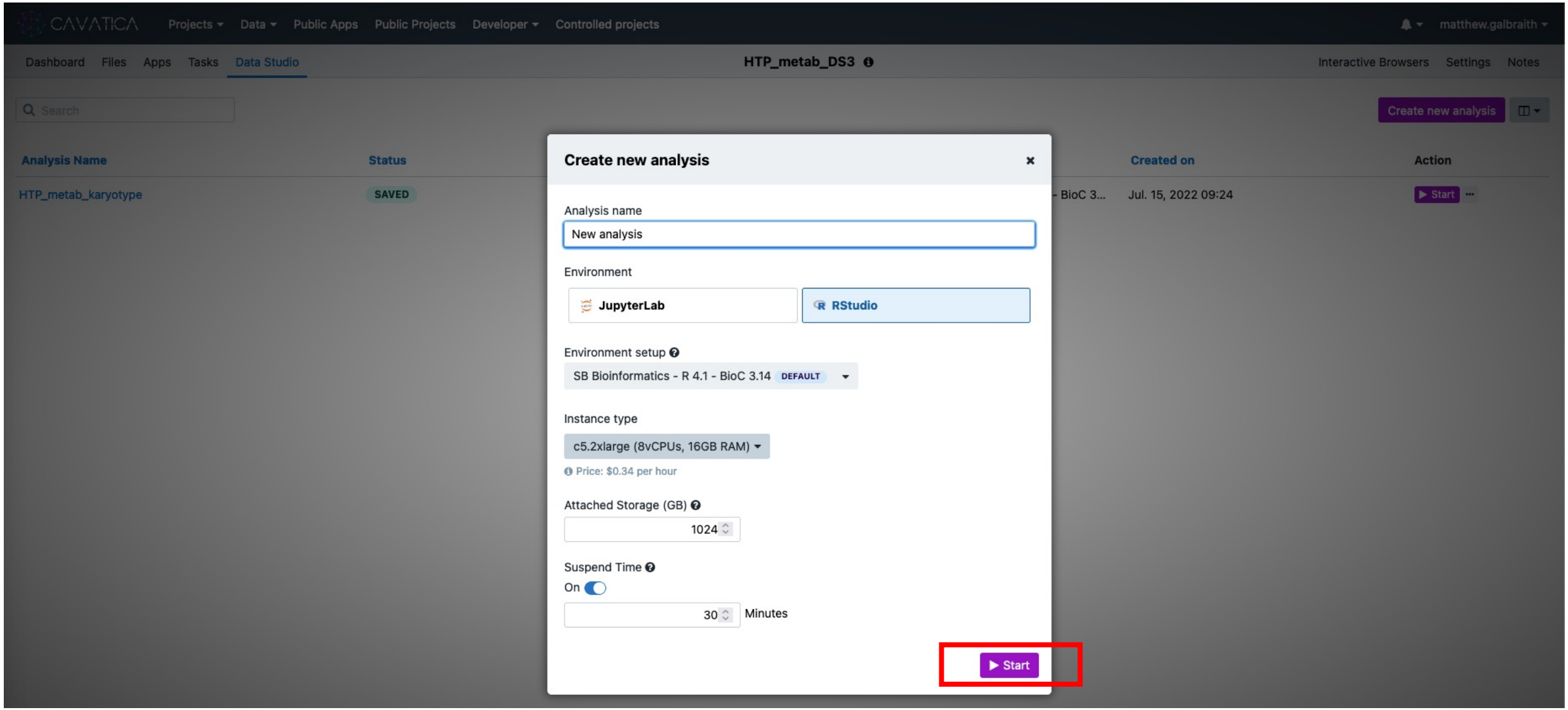
CAVATICA: Data Studio (interactive analysis)



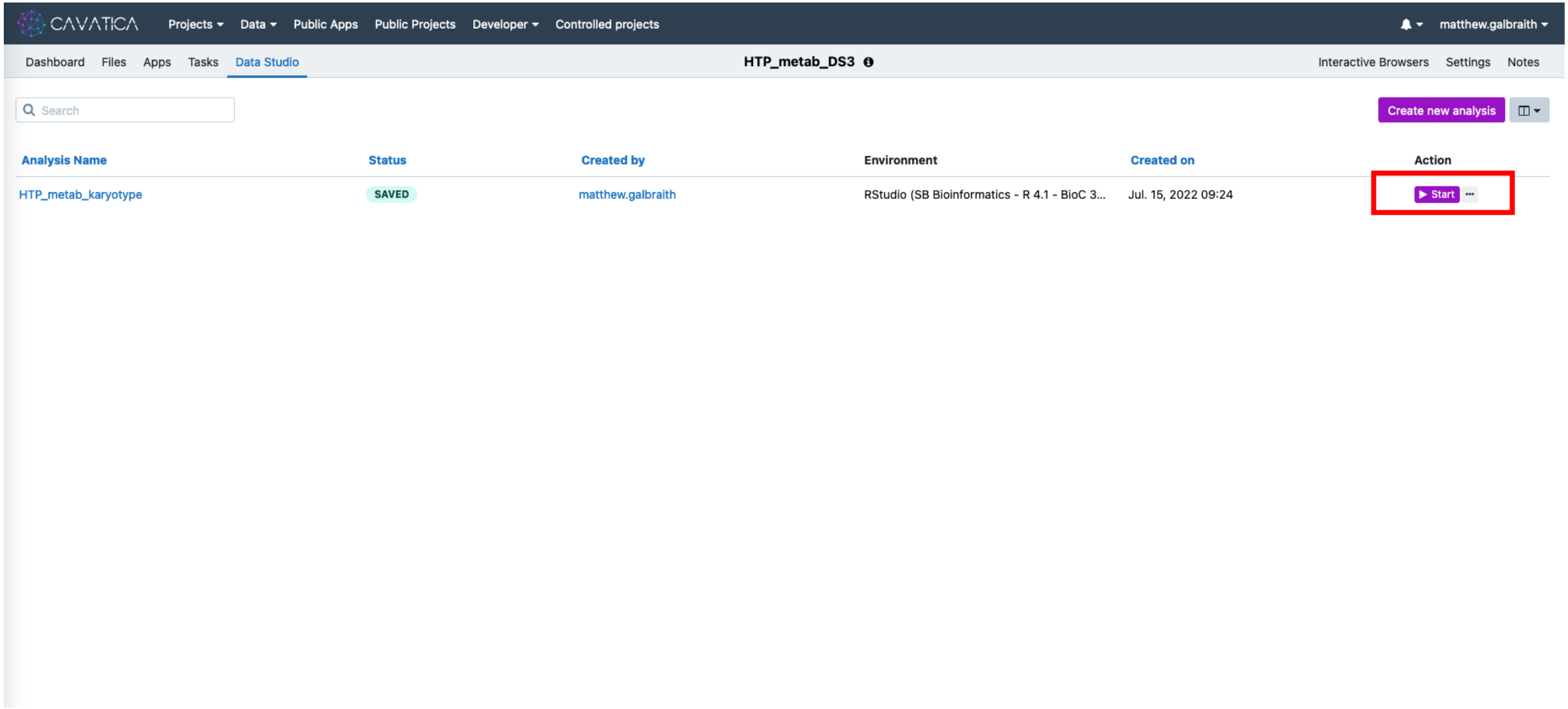
CAVATICA: Data Studio – new analysis



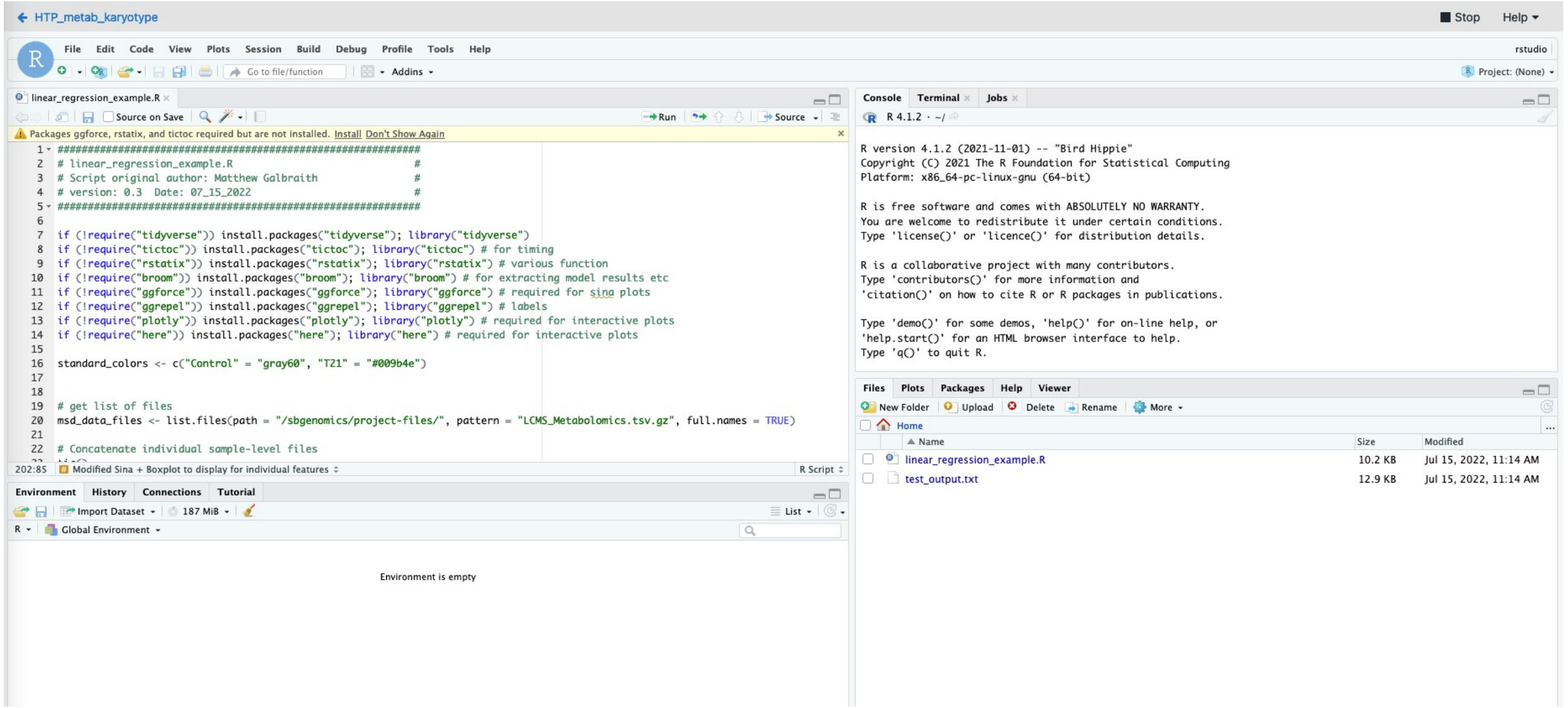
CAVATICA: Data Studio – new analysis



CAVATICA: Data Studio – existing analysis

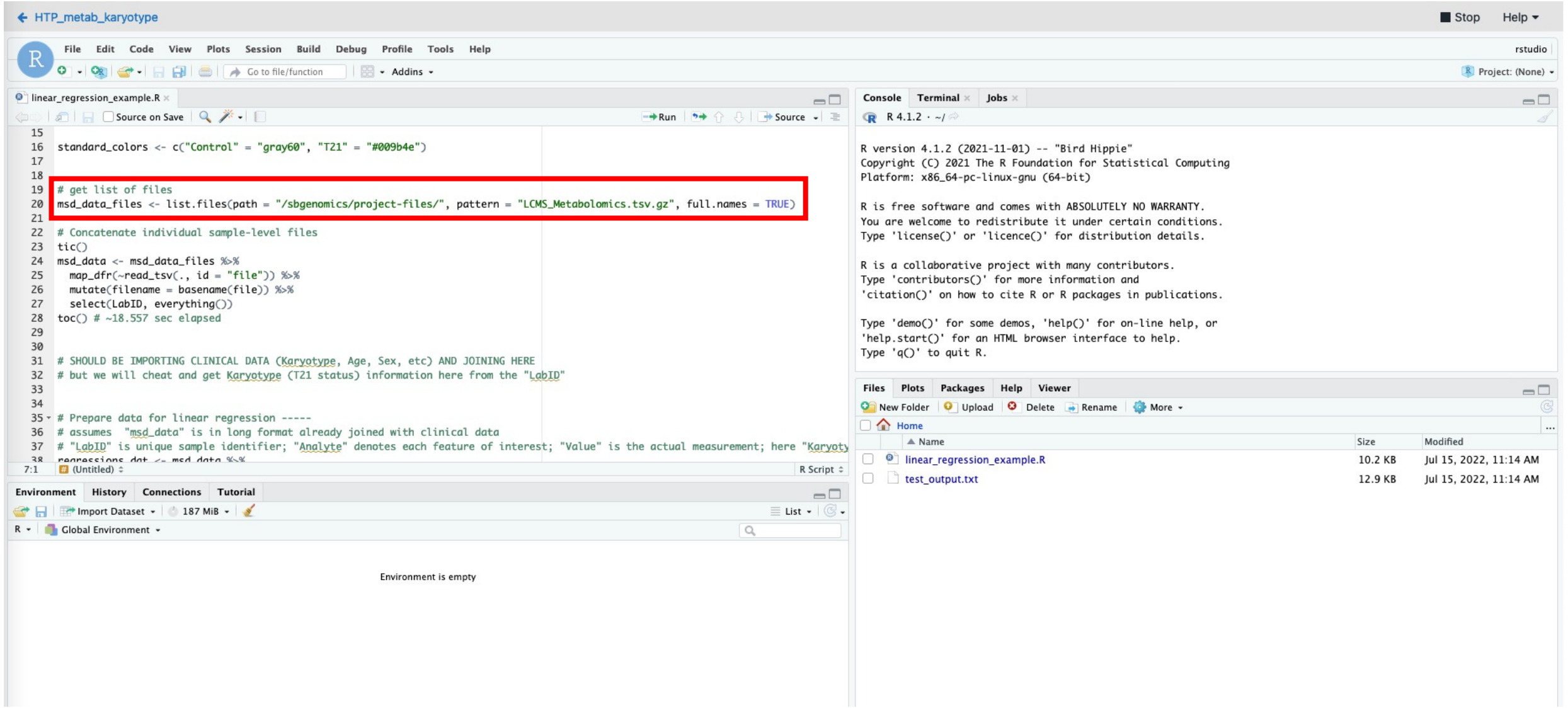


CAVATICA: Data Studio – Rstudio session



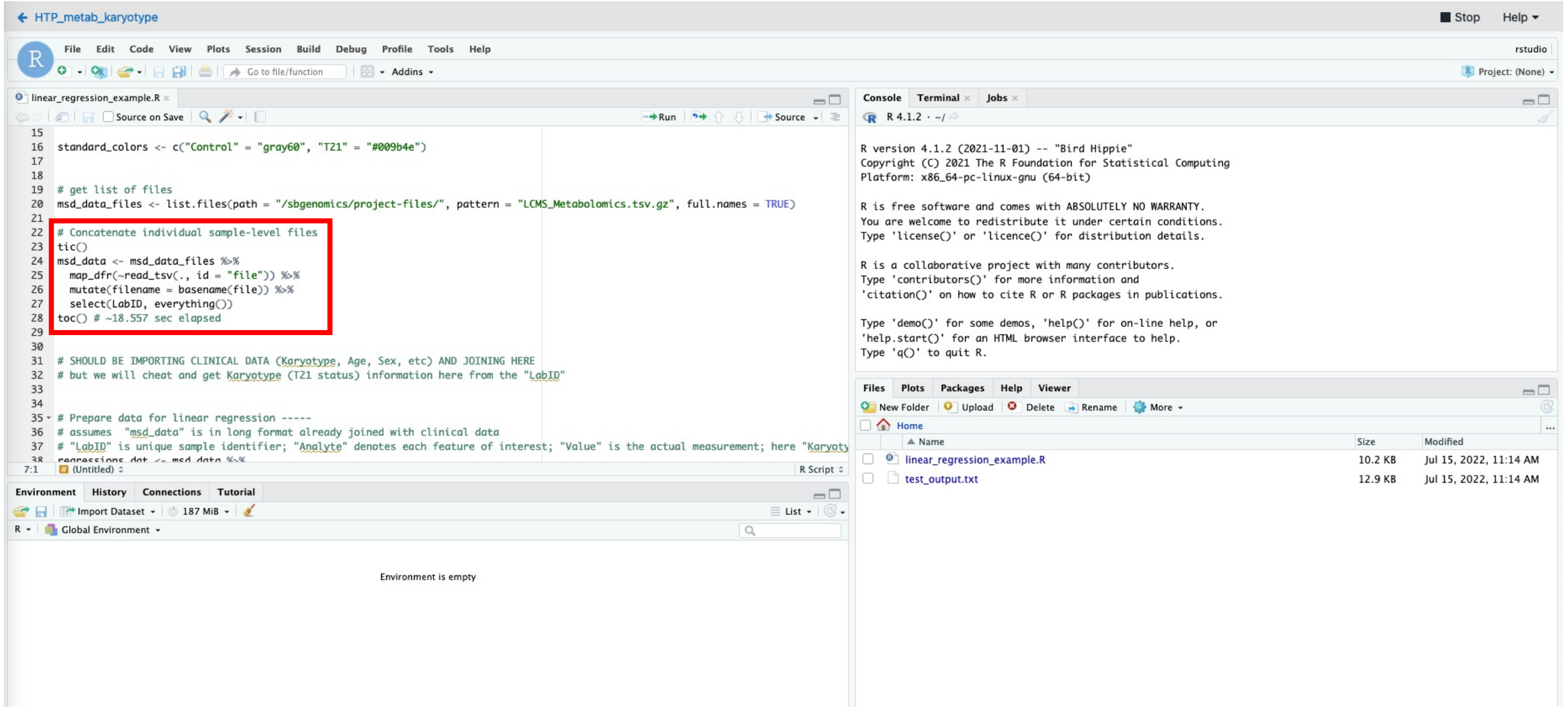
Find code at: https://github.com/mattgalbraith/INCLUDE_examples/blob/main/linear_regression_example.R

CAVATICA: Data Studio – input files location

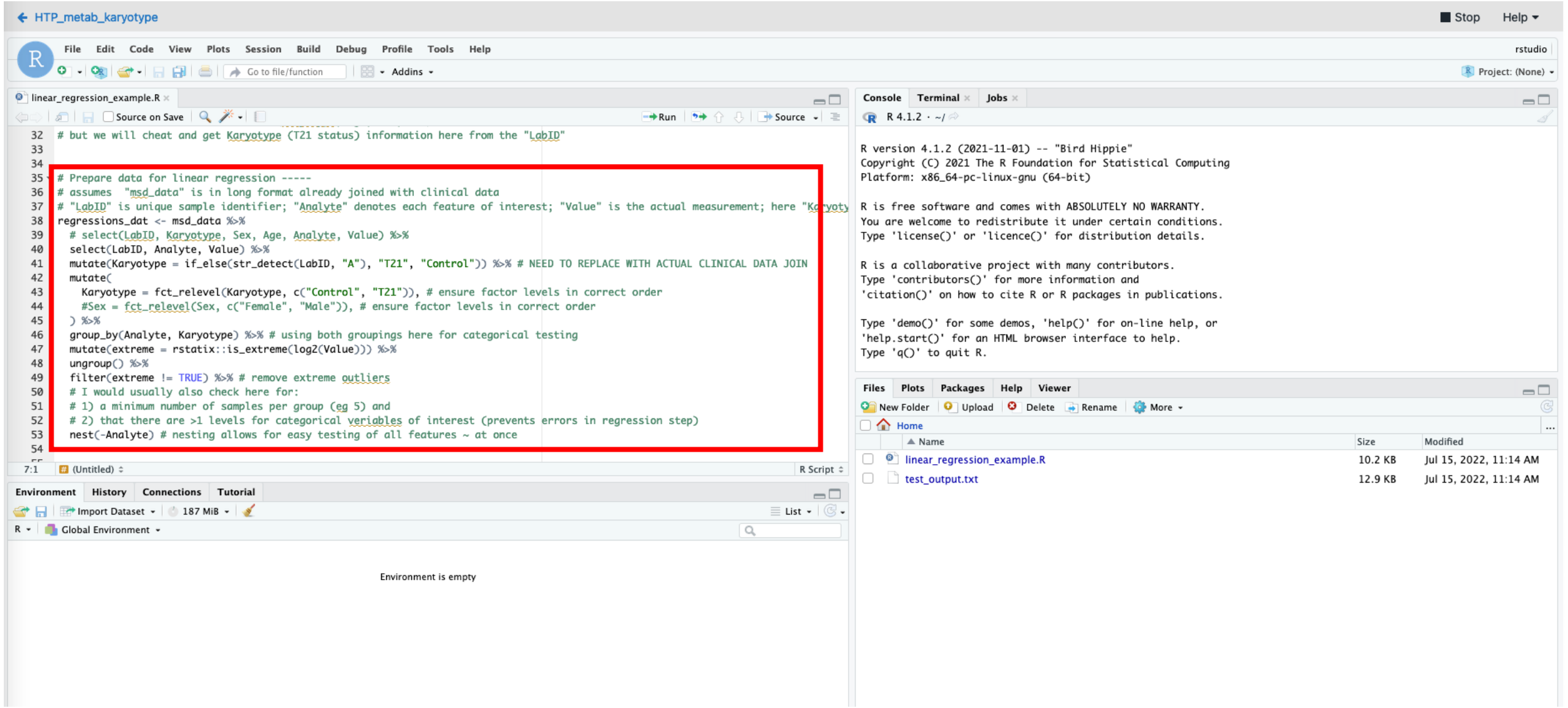


Files delivered from INCLUDE data hub found here: `/sbgenomics/project-files/` (read-only from Data Studio)

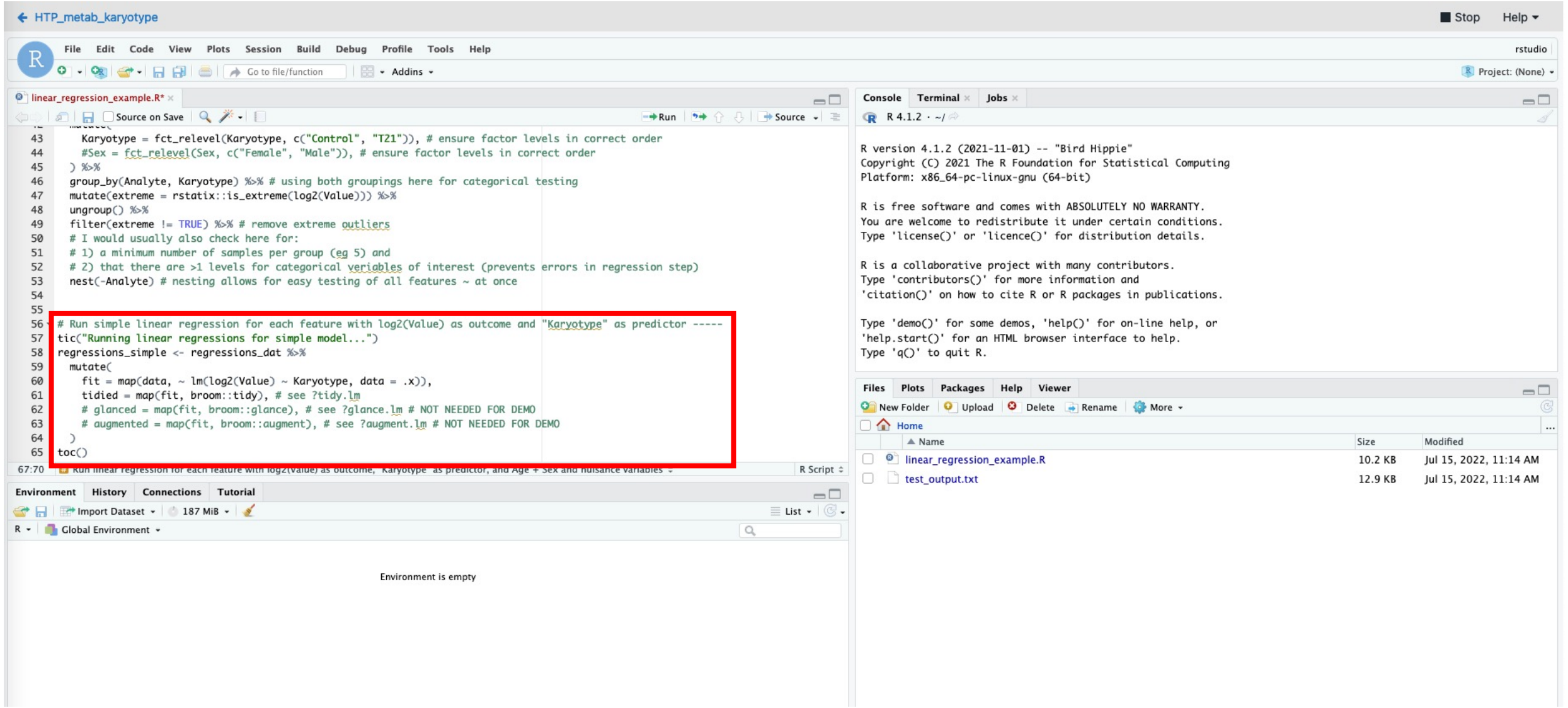
CAVATICA: files from INCLUDE data hub need to be combined



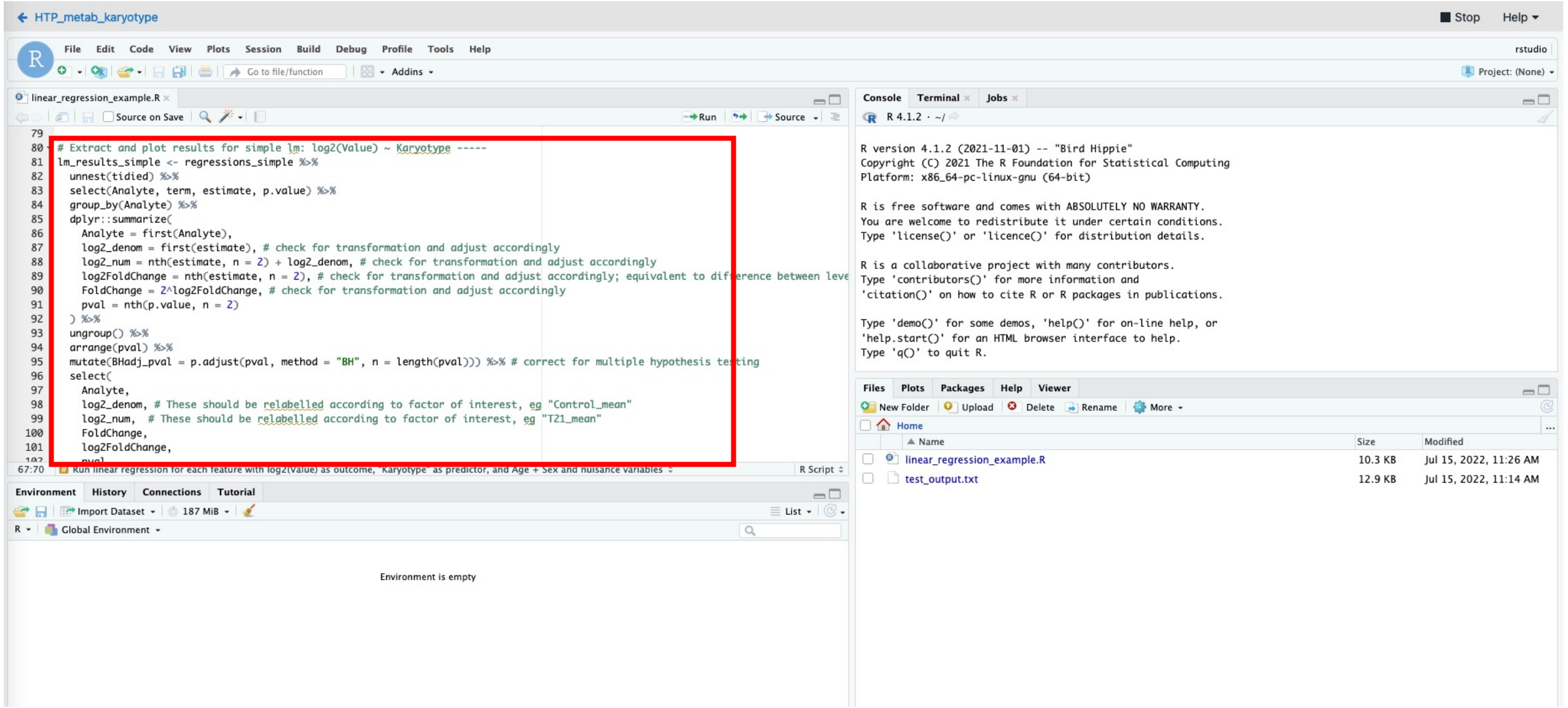
CAVATICA: example linear regression analysis with HTP metabolomics



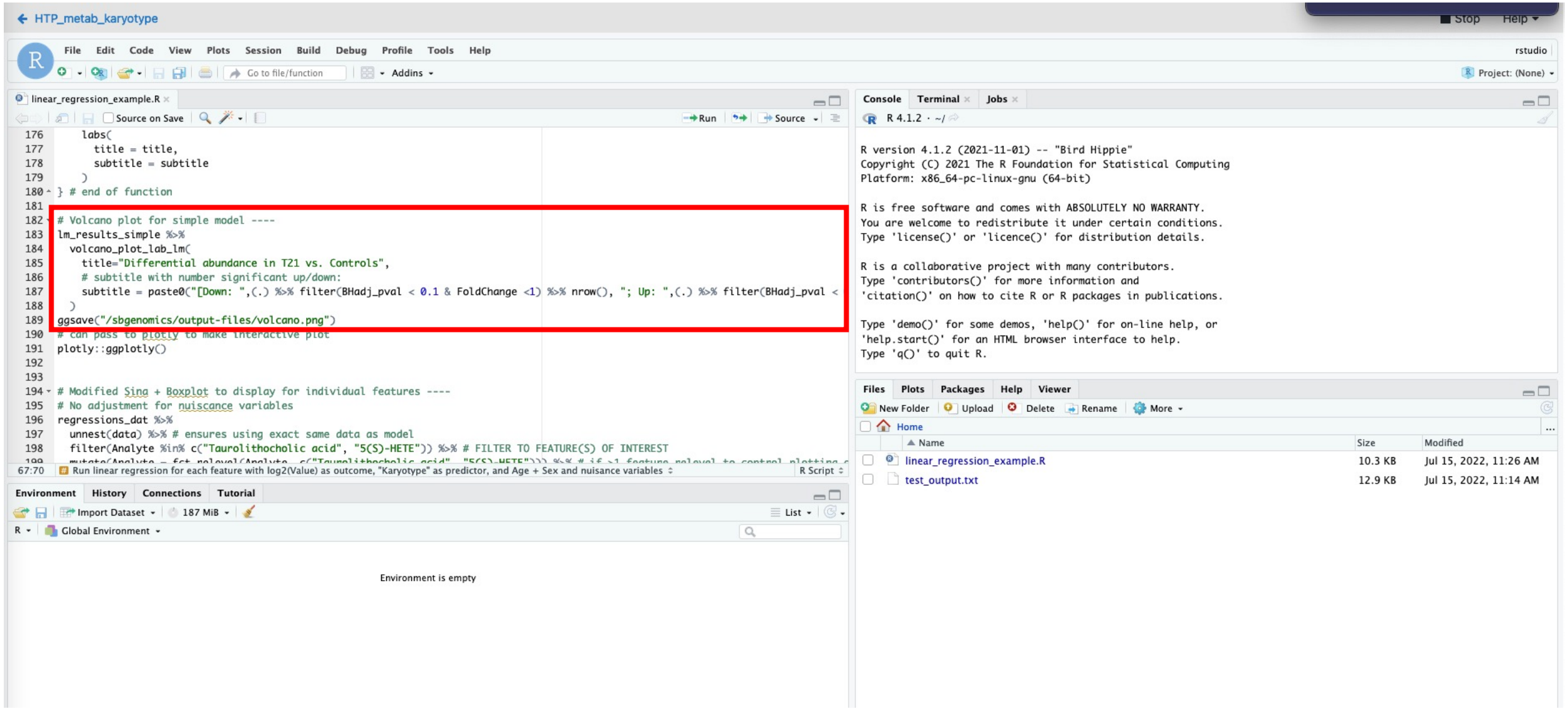
CAVATICA: example linear regression analysis with HTP metabolomics



CAVATICA: example linear regression analysis with HTP metabolomics



CAVATICA: example linear regression analysis with HTP metabolomics



To be visible in "Files" tab, output files must be saved here: `/sbgenomics/output-files/`

CAVATICA: example linear regression analysis with HTP metabolomics