**Background**

**Data Tables:**

<https://biodatacatalyst.nhlbi.nih.gov/resources/data/studies>

<https://biodatacatalyst.nhlbi.nih.gov/covid-19>

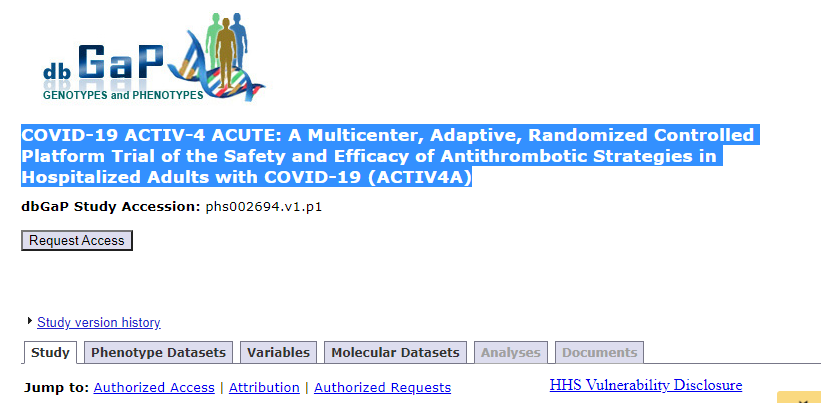
**Preparation Steps**

1. Get a document/spreadsheet/etc. of the new BioData Catalyst studies - this must be verified by DRMWG as released throughout the BioData Catalyst Ecosystem
2. Determine which of the studies is a COVID study or not
   1. COVID studies go to the [COVID-19 Studies](https://biodatacatalyst.nhlbi.nih.gov/covid-19) Table
   2. Other studies go to the [BioData Catalyst Studies](https://biodatacatalyst.nhlbi.nih.gov/resources/data/studies) Table
3. Locate your previous spreadsheets and make copies reflective of the current date

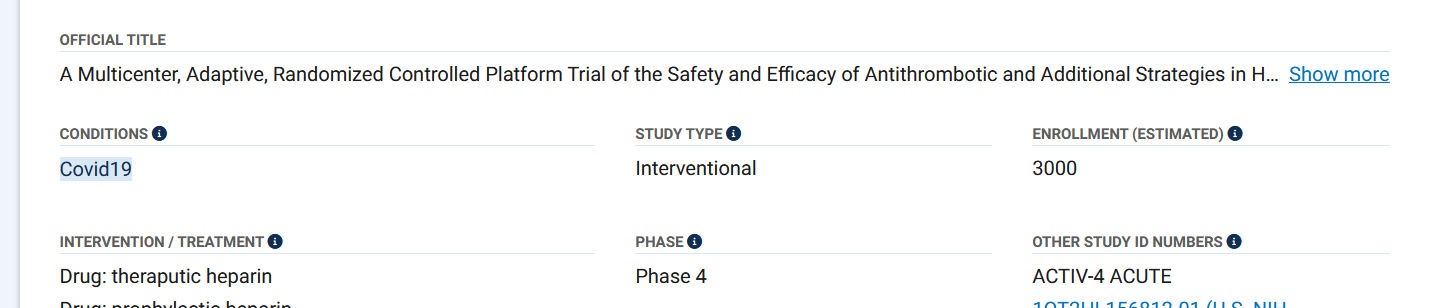
**Updating the COVID-19 Studies Table**

|  |  |
| --- | --- |
| **Term** | **Definition** |
| **Study Title** | Title of study, found on dbGaP study landing page |
| **Study Short Name** | Short name of study, if applicable. Found within parentheses in the title |
| **Study Description** | A short (2-4 sentence) paragraph describing the study, synthesized from the dbGaP study landing page |
| **Accession** | The dbGaP identifier for a study |
| **Study Type** | The category of study |
| **Link to Study Landing Page** | A url link to the study landing page |
| **Link to Organizational Network** | The organizational network is responsible for organizing or managing a study. This is a link to the home page, if applicable |
| **Responsible Party** | The name and institution of the one responsible for a given study |
| **Other Information** | If anyone from BDC3 suggests that additional contextualizing information for the study is necessary, that information goes in this field |

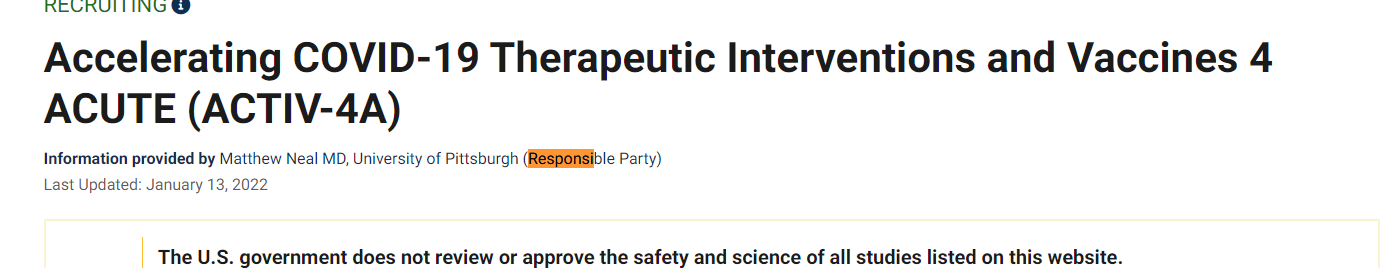
1. Navigate to the dbGaP page for the study (e.g. <https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs002694.v1.p1>)
2. Copy the Accession ID (e.g. phs002694.v1.p1) into the 'Accession' field
3. Copy the title at the top of the page and the short name in parenthesis. Enter these into 'Study Title' and 'Study Short Name' fields



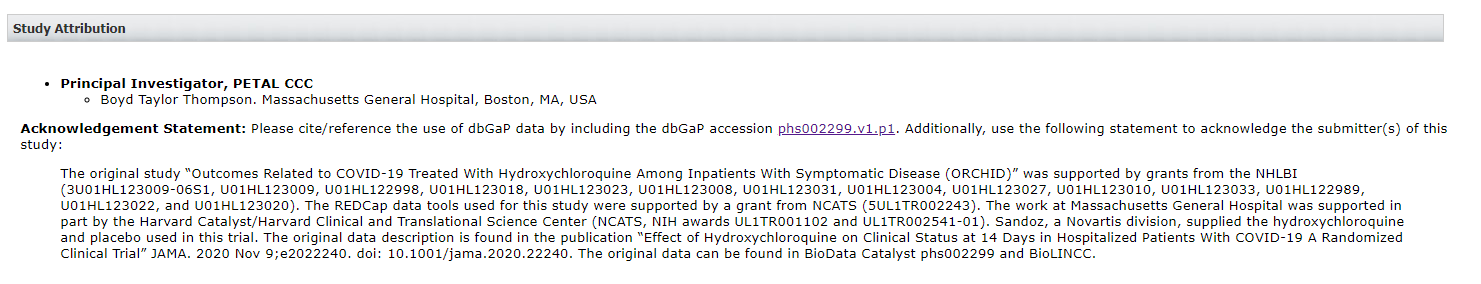
1. From the dbGaP landing page for this study, shorten the description to create a 2-3 sentence description of this study. Enter this into the 'Study Description' field.
2. Check if there is a link to clinicaltrials.gov by searching 'NCT'. If there is, copy this link into the 'Link to Study Landing Page'
3. Find the Study Type either on the dbGaP page or on the clinicaltrials.gov page (if applicable)



1. 'Responsible Person' can be found in the clinicaltrials.gov site, near the title or deeper into the body of the study. If none can be found, look for a SPONSOR and write that name.



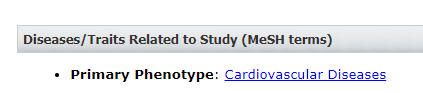
1. To find the network, google the study short-name + network. E.g., 'ACTIV-4A network' reveals that this study is part of the PETAL network.
2. 'Other Information' can be found under the Study Attribution tab on the dbGaP Landing Page:



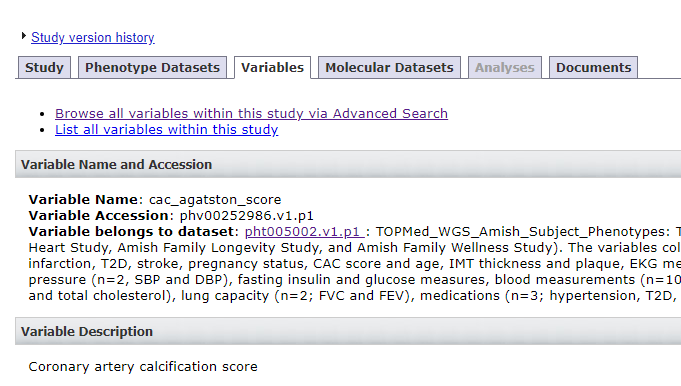
**Updating the BioData Catalyst Studies Table**

|  |  |
| --- | --- |
| **Term** | **Definition** |
| **Study Accession** | The dbGaP identifier for a study |
| **Study/Cohort Abbreviation** | Originally, the study abbreviation as taken from e.g. <https://topmed.nhlbi.nih.gov/topmed-whole-genome-sequencing-methods-freeze-8>. Generally it's what's found between the parenthesis in a study title |
| **Study Name (dbGaP link)** | Title of study, found on dbGaP study landing page, with corresponding link |
| **Study Description (Verbose)** | 1-2 paragraphs describing the study, synthesized from the dbGaP study landing page |
| **Study Description (Short)** | A short (2-4 sentence) paragraph describing the study, synthesized from the dbGaP study landing page |
| **Primary Research Focus** | The primary MeSH term as designated by dbGaP |
| **Total Number of Variables** | The total number of variables that the study contains |
| **dbGaP listed variable** | The variable listed on the 'Variables' tab in dbGaP |
| **Study Type** | The category of study |
| **Study-Reported Population(s)** | The populations reported by the study, typically corresponding to race or ethnicity |
| **Study Consent** | The full consent group(s) which define how the data for this study must be used. Semicolon-delimited. |
| **Consent Short** | Same as 'Study Consent', but only the abbreviations |
| **Type of Molecular Data Available** | The type of -omics data available (if any) for this study, e.g. RNA-Seq, SNPs, etc. |
| **Primary Data Dictionary Link** | The ID of and link to the data dictionary listed under the Variables tab and the header 'Variable belongs to dataset' |

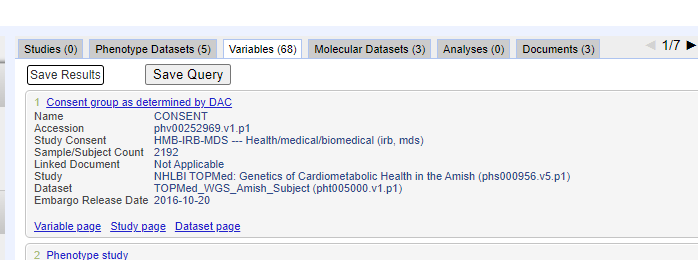
1. Navigate to the dbGaP landing page for this study (e.g. <https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000956.v4.p1>)
2. Copy the Accession ID (e.g. phs000956.v4.p1) into the 'Study Accession' field
3. Copy the title at the top of the page and the short name in parenthesis. Enter these into 'Study Name' and 'Study/Cohort Abbreviation' fields
4. From the dbGaP landing page for this study, shorten the description to create a 1-2 paragraph-long description of this study. Enter this into the 'Study Description (Verbose)' field.
5. Shorten the 'Study Description (Verbose)' field to 2-4 sentences and enter into the 'Study Description (Short)' field.
6. Scroll down to the 'Diseases/Traits Related to Study (MeSH terms)' section, and enter in the Primary phenotype into 'Primary Research Focus'



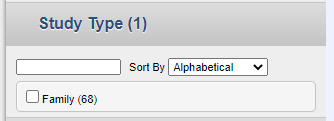
1. The dbGaP listed variable can be found on the variables tab of the dbGaP study landing page, under 'Variable Description'



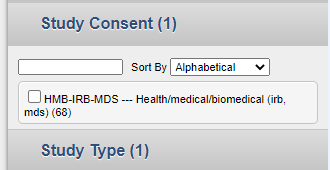
1. The primary data dictionary link can be found by copying what is in the 'Variable belongs to dataset' prompt, e.g. pht005002.v1.p1.
2. To find the total number of variables, navigate to the 'Variables' tab and then the hyperlink Browse all variables within this study via Advanced Search. You may have to navigate to the newest version of the study to get this information. Within the Advanced Search, get the number of variables from the 'Variables' tab



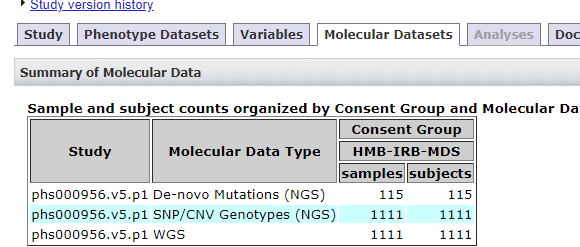
1. Also within the advanced search, record what study type it is under 'Study Type' on the left side panel.



1. Also within the advanced search, Record the study consent on the 'Study Consent' pane on the left side panel. For multiple consents, delimit using a semicolon (;)



1. Shorten the full consent to just the ID (e.g. HMB-IRB-MDS) for the 'Consent Short' field. For multiple consents, delimit using a semicolon (;)
2. Under the Molecular Datasets tab on the landing page, identify the type of molecular data that is in this study. Multiple types of Molecular Data are to be delimited using a semicolon (;)



1. Find population information on the 'Study' tab, the 'dbGaP estimated ancestry' link under the Study Description pane. Delimit all ancestry values with semicolons (;)

**Converting CSV to JSON**

1. Once the Excel spreadsheets are updated, convert them into to CSV files within Microsoft Excel.
2. Follow the instructions in the [github repo](https://github.com/jcheadle-rti/bdc3_data_table) to convert the CSV