COPDGene Clinical Data Exploration Tool

Quick Reference Guide

GETTING STARTED

COPDGene
COPD Genetic Epidemiology

NIH National Institutes of Health
Office of Strategic Coordination - The Common Fund

NIH National Heart, Lung, and Blood Institute

Data **C**ommons **P**ilot **P**hase **C**onsortium

Data **S**torage, **T**oolspace, **A**ccess and analytics for bi**G** data **E**mpowerment

THE FUNDAMENTALS

https://copdgene.hms.harvard.edu

AUTHORIZATION

First a DCPPC user must be authorized to access data from COPDGene. In order to become authorized you must complete and be up-to-date the NIH Data Commons Onboarding form located at the following location: https://tinyurl.com/dcppc-START. **Note this form has VERY recently changed and individuals will need to check and make sure they have answered all questions on the form. Including *Data Access* and *Privacy agreements* added at the beginning of July**.

You can check your approval status by checking the 'DARWhitelist' tab of the following spreadsheet: https://tinyurl.com/dcppc-DAR. An approved user will have an "Approved" response under the column labeled *Current DAR status of individual's access*. If you are not currently approved, this requires updating the NIH Data Commons onboarding form and having a 'Yes" response to the controlled access data question on the NIH Data Commons Onboarding form.

ACCESS

You may request access to the tool by contacting Jessica Lyons, Jessica_Lyons@hms.harvard.edu and providing your preferred user ID, either your eRA Commons, ORCID, Github or Google ID. Once access has been granted you will receive an email with a link to access the application and You will then be able to log-in and will be presented with the list of data integrated in the COPDGene i2b2/tranSMART application. Summary Statistics and Advanced Workflow options will now be available for access.

ORGANIZATION

STEP 1: *Explore ontologies by opening yellow folders to view data*

The home page is divided into two sections: the left side contains the search tree for registry data and the right side contains the cohort selection boxes (***Subset 1*** and ***Subset 2***).

COHORT (SUBSET) SELECTION

STEP 2: *Drag and drop criteria from left to select subset of individuals*

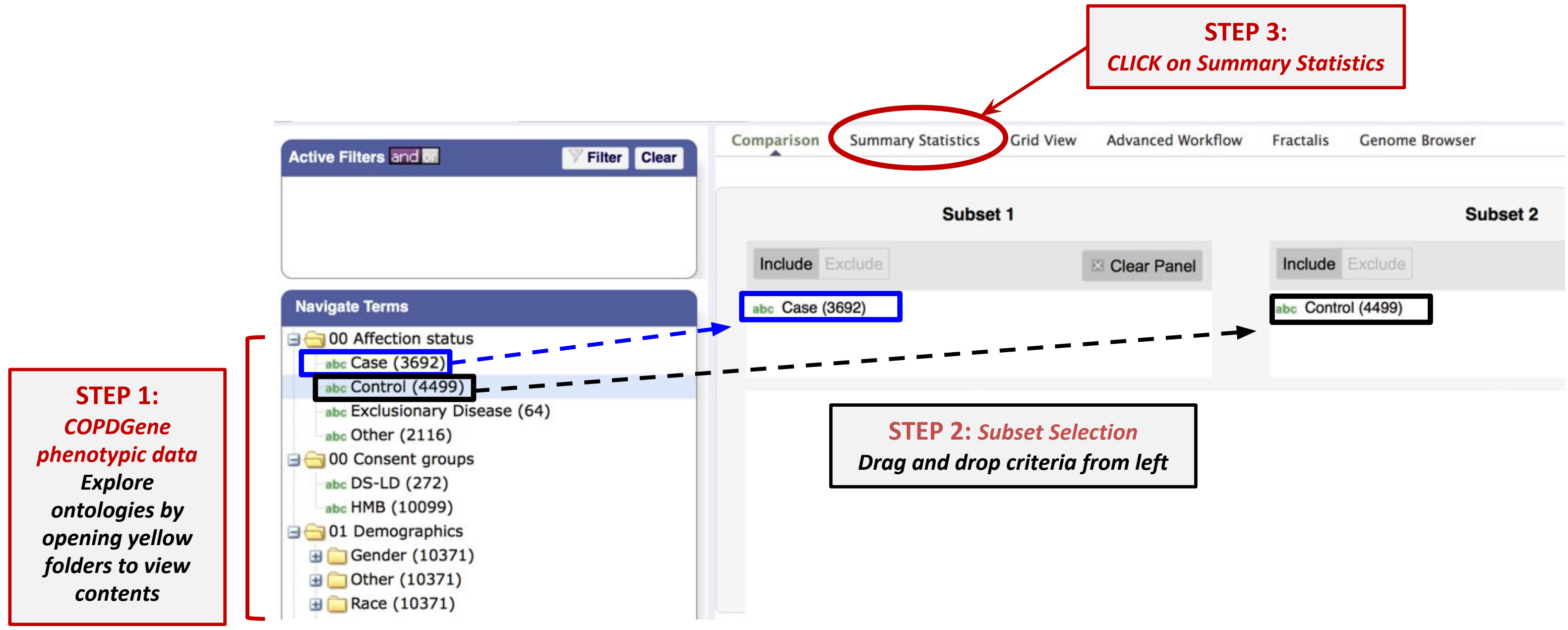
Subset selection criterion can be very simple or more comprehensive, using combinations of the Boolean logic 'and' (entries in stacked subset boxes), 'or' (entries in the same subset box), and 'not' (by clicking the **Exclude** option for the contents of a box).

USING STATISTICAL TOOLS TO QUERY DATA

STEP 3: *Generate Summary Statistics*

1) After selecting cohort(s), click on **Generate Summary Statistics**.
2) Subsets can be verified at the top of the **Summary Statistics** section. The i2b2/tranSMART application automatically generates a table with subject totals and statistical analysis by age, sex and race for each subset, if data are available.
3) Drag and drop any of the variables (from left side to anywhere on the right side of the home page) to generate statistical analysis based upon that variable.

COPDGENE i2b2/tranSMART HOME PAGE



STEP 3:
CLICK on Summary Statistics

STEP 1:
COPDGene phenotypic data
Explore ontologies by opening yellow folders to view contents

STEP 2: *Subset Selection*
Drag and drop criteria from left

USE CASE EXAMPLE

Data Exploration: IS THERE A DIFFERENCE IN RESTING OXYGEN SATURATION BETWEEN CASE AND CONTROL GROUPS?

STEP 1: ***Explore ontologies by opening yellow folders to view data or you can search for a term using the search box if you know key terms or the variable name and hit Enter.***

STEP 2: ***Drag and drop criteria from left to select subsets of individuals as shown on previous pane***

Subset 1 will be individuals who under affection status ***Case***. To select this cohort, follow this path in the folder, and drag and drop **Case** into the top box of Subset 1:

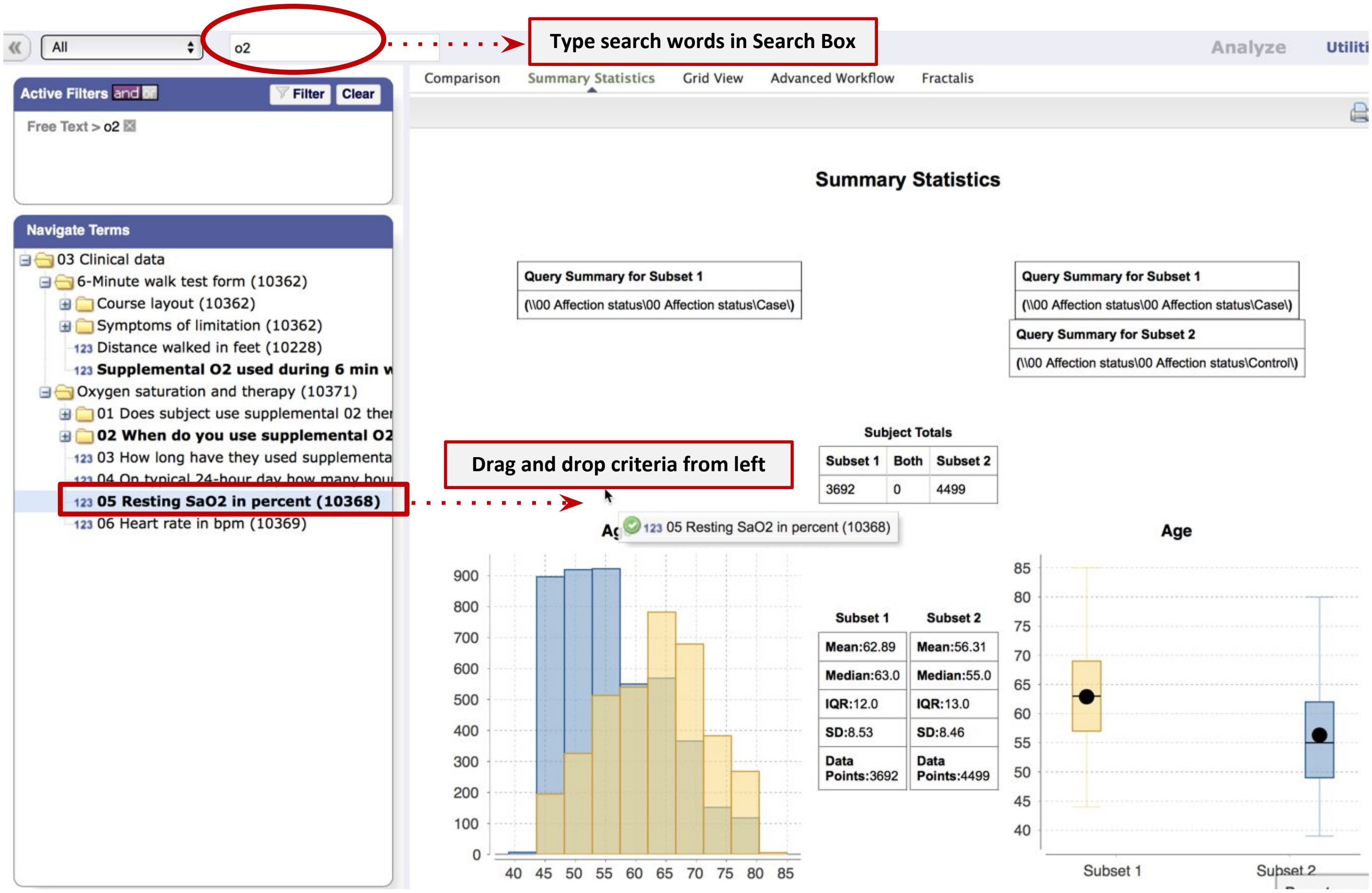
00 Affection status/Case

Subset 2 will be individuals who under affection status ***Control***. To select this cohort, follow this path in the folder, and drag and drop **Control** into the top box of Subset 2 (The values "Exclusionary Disease " and "Other" are not used in this case):

00 Affection status/Control

STEP 3: ***Click Summary Statistics.*** Below are the summary statistics comparing the Case and Control Subsets. **Then, drag and drop the variable Resting SaO2 in percent** (under the **Clinical Data** folder) **into the Summary Statistics side of the user interface.**

03 Clinical data/Oxygen saturation and therapy/05 Resting SaO2 in percent



All
o2
Type search words in Search Box
Analyze
Utiliti
Active Filters and or
Filter
Clear
Free Text > o2
Comparison
Summary Statistics
Grid View
Advanced Workflow
Fractalis
Navigate Terms
03 Clinical data
6-Minute walk test form (10362)
Course layout (10362)
Symptoms of limitation (10362)
123 Distance walked in feet (10228)
123 Supplemental O2 used during 6 min w
Oxygen saturation and therapy (10371)
01 Does subject use supplemental 02 ther
02 When do you use supplemental O2
123 03 How long have they used supplementa
123 04 On typical 24-hour day how many hou
123 05 Resting SaO2 in percent (10368)
123 06 Heart rate in bpm (10369)
Drag and drop criteria from left
123 05 Resting SaO2 in percent (10368)
Ag
Age
900
800
700
600
500
400
300
200
100
0
40
45
50
55
60
65
70
75
80
85
40
45
50
55
60
65
70
75
80
85
Subset 1
Subset 2

Summary Statistics

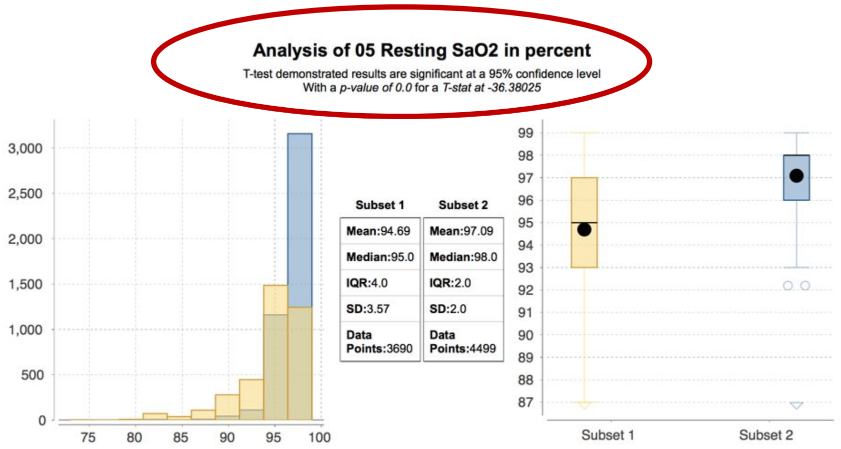
Query Summary for Subset 1
(\\00 Affection status\00 Affection status\Case\)

Query Summary for Subset 1
(\\00 Affection status\00 Affection status\Case\)
Query Summary for Subset 2
(\\00 Affection status\00 Affection status\Control\)

Subject Totals

Subset 1	Both	Subset 2
3692	0	4499

Subset 1	Subset 2
Mean:62.89	Mean:56.31
Median:63.0	Median:55.0
IQR:12.0	IQR:13.0
SD:8.53	SD:8.46
Data Points:3692	Data Points:4499



Analysis of 05 Resting SaO2 in percent

T-test demonstrated results are significant at a 95% confidence level
With a *p-value of 0.0* for a *T-stat at -36.38025*

Subset 1	Subset 2
Mean:94.69	Mean:97.09
Median:95.0	Median:98.0
IQR:4.0	IQR:2.0
SD:3.57	SD:2.0
Data Points:3690	Data Points:4499

A t-test is automatically performed, based on having a continuous variable. The results show the mean of 94.7% for the Cases compared to 97% of the controls of SaO2. The p-value in this case is 0.0.

For questions or issues with the COPDGene User Interface, please contact Jessica_Lyons@hms.harvard.edu.