COMETS Analytics Tutorial

CAT Work group

2018-09-22

Contents

Introduction

COMETS, the **CO**nsortium of **MET**abolomics **S**tudies, has a major goal to produce jointly coordinated, multi-cohort, high-impact publications devoted to advancing the methods and scientific understanding of the human metabolome and its relationship to disease etiology and prognosis. Since its inception in 2014, COMETS has grown to include more than 50 cohorts worldwide.

Comets Analytics serves as the infrastructure to facilitate and coordinate data analysis efforts. This tutorial gives an overview and detailed instructions on the use of COMETS Analytics in support of the COMETS Consortium.

Prepared by the COMETS Tutorial Group:

Ella Temprosa

Steve Moore

Oana Zeleznik

Laura Trijsburg

Rachel Kelly

Errika Loftfield

Kathleen McClain

Kaitlyn Mazzili

Overview

Each cycle of analyses consists of 5 steps described below to bundle multiple projects to minimize burden on cohort analytic resources.

This tutorial serves as a guide through this cycle. You can register for access to Comets Analytics using the instructions in Chapter 2.1 Registration For a quick walk-through the cycle, you can use the sample input file and follow the directions in Chapter 1 [Get Started with Sample File]. Once you are ready to conduct your cohort-specific analyses, see the Chapter **??** [Conduct Cohort Data Analyses]. For more technical details, refer to the Chapter 2 Manual and [FAQ] for anwers to commonly asked questions.

Chapter 1

Get Started with the Sample File

This chapter guides you through the features of COMETS Analytics using the sample file to provide an introduction to the flow of analyses.

1.1 Sample Input File

There is a sample file located on the website that will serve as your data input template. After logging in, navigate to the 'Correlate' tab. On the lower left side, there is text stating 'Download Sample Input'. Click this text, and the sample file will download. This template was designed to conduct analyses of age and metabolite, and BMI and metabolite correlations. It illustrates how your input data should be formatted, and includes practice data for >600 metabolites and key covariates such as age and gender. Using this sample file as your input data will allow you to practice using the app functions and familiarize yourself with its output. You may want to begin by downloading and then inputting this sample file to the website to familiarize yourself with how the data analysis works.

1.2 Quick Analysis

To upload your input file to the COMETS Analytics website, please follow these steps:

1. Select the 'Correlate' tab.
2. Specify your cohort from the dropdown menu.
3. Choose your input data file, formatted as described above, using the 'Choose File' button.
4. Once you have uploaded your file the 'Check Integrity' button will activate, use this button to check the integrity of your data input file. The integrity check results are displayed in the right panel. The integrity check results contain a summary of the input data, including the number of measured metabolites (as defined by the user in the *Metabolites* sheet), the number of cohort subjects and covariates (as defined by the user in the *SubjectData* sheet), and the number of subjects by the number of metabolites measured (as defined by the user in the *SubjectMetabolites* sheet). Additionally, a harmonization summary is displayed that includes number of harmonized and non-harmonized metabolites as well as number of metabolites with variance equal to zero or with more than 25% of values represented by minimum values. Finally, histograms illustrating the distributions of the variance and the number of missing values are plotted (figures not shown here). Integrity checks are run to ensure the appropriate data are presented for analysis. If integrity checks fail, please email Kaitlyn Mazilli (kaitlyn.mazzilli@nih.gov). If all integrity checks are passed, as indicated in a green banner above the summary output, please click 'Download Results' in the right corner and email the file to comets.analytics@gmail.com which is managed by Nathan Appel at IMS.

Chapter 2

Manual

Important features and technical details are described in this chapter.

2.1 Registration

Users are required to LOG IN using an existing Facebook or Google account, or can sign up for a login using the "SIGN UP" button. Your gmail account through your university may be used for authentication.

<span class='t
Login**</span
active where yo
the 3rd party l
account. <span
class='textb'>*
Sign-Up** </s
this tab to regi
account. <span
class='textc'>*
login**
facebook or goo
login using you
from your acco
class='textd'>*
password**</s
your password.
required to pro
associated with
<span class='te
Proceed**</sp
web app by clic
orange button.

2.2 Data Preparation

Standard input using excel, a widely accessible format, is required for COMETS Analytics and can be created from various data file formats. Data Integrity checks for input errors provide meaningful and actionable messages to users for analyses to proceed. See chapter **??** [Step 1 Data Preparation]

2.2.1 Create Input

2.3 Harmonization details

The names of metabolites from each cohort has been mapped or harmonized to a common name. This important stage facilitates comparison of metabolites across different studies that use different platforms and/or naming conventions.

2.3.1 Preharmonization

Prior to conducting your cohort analyses, xxx.

2.4 Integrity checks

Prior to running the models for analyses, CA conducts multi-level checks to ensure data and models are appropriate for analyses.

2.4.1 Data

2.4.2 Models

2.5 Correlation Analyses

2.5.1 Details

2.5.2 Output