RImmPort: Quick Start Guide

Ravi Shankar

2025-04-15

---

1 Introduction

ImmPort study data is available for download in two formats currently: MySQL and TSV (Tab) formats. The RImmPort workflow is as follows: 1) MySQL formatted study data: User downloads one or more studies in MySQL zip files. Unzips the files. Loads local database instance. Connects to the database. Sets the ImmPort data source to the connection handle. Invokes RImmPort functions. 2) Tab: User downloads one or more studies in Tab format. Passes the folder where the zip files are located to an RImmPort function that builds SQLite database. Connects to the database. Sets the ImmPort data source to the connection handle. Invokes RImmPort functions.

User downloads study data of interest from the ImmPort website (http://www.immport.org) **. Depending on the file format MySQL or Tab the data is loaded into a local MySQL and SQLite database respectively. The user installs the RImmPort package, loads the RImmPort library, connects to the ImmPort database, and calls RImmPort methods to load study data from the database into R. Please refer to RImmPort_Article.pdf for a detailed discussion on RImmPort.

** User need to regsiter to the ImmPort website for downloading the datasets.

2 Initial Steps

- Download MySQL or Tab formatted data of studies of interest from the ImmPort website
- If working with MySQL-format, load the data in to a local MySQL database
- Install and load RImmPort package, and other required packages.

3 Load the RImmPort library

```
library(RImmPort)
library(DBI)
library(sqldf)
library(plyr)
```

4 Setup ImmPort data source that all RImmPort functions will use

4.1 Option 1: ImmPort MySQL database

4.1.1 Download zip files of ImmPort study data in MySQL format. e.g.'SDY139' and 'SDY208'

4.1.2 Load the data into a local MySQL database

4.1.3 Connect to the ImmPort MySQL database.

```r
# provide appropriate connection parameters
mysql_conn <- dbConnect(MySQL(), user="username", password="password",
                        dbname="database",host="host")
```

4.1.4 Set the data source as the ImmPort MySQL database.

```r
setImmPortDataSource(mysql_conn)
```

4.2 Option 2: ImmPort SQLite database

4.2.1 Download zip files of ImmPort data, in Tab format. e.g.'SDY139' and 'SDY208'

```r
# get the directory where ImmPort sample data is stored in the directory structure of RImmPort package
studies_dir <- system.file("extdata", "ImmPortStudies", package = "RImmPort")

# set tab_dir to the folder where the zip files are located
tab_dir <- file.path(studies_dir, "Tab")
list.files(tab_dir)
```

4.2.2 Build a local SQLite ImmPort database instance.

```r
# set db_dir to the folder where the database file 'ImmPort.sqlite' should be stored
db_dir <- file.path(studies_dir, "Db")

# build a new ImmPort SQLite database with the data in the downloaded zip files
buildNewSqliteDb(tab_dir, db_dir)

list.files(db_dir)
```

4.2.3 Connect to the ImmPort SQLite database

```r
# get the directory of a sample SQLite database that has been bundled into the RImmPort package
db_dir <- system.file("extdata", "ImmPortStudies", "Db", package = "RImmPort")

# connect to the private instance of the ImmPort database
sqlite_conn <- dbConnect(SQLite(), dbname=file.path(db_dir, "ImmPort.sqlite"))
```

4.2.4 Set the data source to the ImmPort SQLite DB

```r
setImmPortDataSource(sqlite_conn)
```

```
## [1] 1
```

5 NOTE: In rest of the script, all RImmPort functions will use the SQLite ImmPort database as the data source.

6 Get all study ids

```
getListOfStudies()
```

```
## [1] "SDY139" "SDY208"
```

7 Get all data of a specific study

The `getStudyFromDatabase` queries the ImmPort database for the entire dataset of a specific study, and instantiates the `Study` reference class with that data.

```
?Study
```

```
# load all the data of study: `SDY139`
study_id <- 'SDY139'
sdy139 <- getStudy(study_id)
```

```
## loading Study ID =  SDY139
## loading Demographics data....done
## loading Concomitant Medications data....done
## loading Exposure data....done
## loading Substance Use data....done
## loading Adverse Events data....done
## loading Protocol Deviations data....done
## loading Medical History data....done
## loading Associated Persons Medical History data....done
## loading Laboratory Test Results data....done
## loading Physical Examination data....done
## loading Vital Signs data....done
## loading Questionnaires data....done
## loading Findings About data....done
## loading Skin Response data....done
## loading Genetics Findings data....loading HLA Typing Results data....done
## loading Array Results data....done
## done
## loading Protein Quantification data....loading ELISA Results data....done
## loading MBAA Results data....done
## done
## loading Cellular Quantification data....loading FCS Results data....done
## loading ELISPOT Results data....done
## done
## loading Nucleic Acid Quantification data....loading PCR Results data....done
## done
## loading Titer Assay Results data....loading HAI Assay Results data....done
## loading Neut. Ab Titer Results data....done
## done
## loading TrialArms data....done
## loading Trial Visits data....done
## loading  TrialInclusionExclusionCriteria data....done
## loading TrialSummary data.... SDY139  done
```

```
## done loading Study ID =  SDY139
# access Demographics data of SDY139
dm_df <- sdy139$special_purpose$dm_l$dm_df
head(dm_df)

##   STUDYID DOMAIN  USUBJID AGE   AGEU     SEX RACE ETHNIC      SPECIES STRAIN
## 1  SDY139     DM SUB118053  2 Months Unknown <NA>   <NA> Mus musculus BALB/c
## 2  SDY139     DM SUB118054  2 Months Unknown <NA>   <NA> Mus musculus BALB/c
## 3  SDY139     DM SUB118055  2 Months Unknown <NA>   <NA> Mus musculus BALB/c
## 4  SDY139     DM SUB118056  2 Months Unknown <NA>   <NA> Mus musculus BALB/c
## 5  SDY139     DM SUB118057  2 Months Unknown <NA>   <NA> Mus musculus BALB/c
## 6  SDY139     DM SUB118058  2 Months Unknown <NA>   <NA> Mus musculus BALB/c
##   SBSTRAIN  ARMCD    ARM
## 1     <NA> ARM678 BALB/c
## 2     <NA> ARM678 BALB/c
## 3     <NA> ARM678 BALB/c
## 4     <NA> ARM678 BALB/c
## 5     <NA> ARM678 BALB/c
## 6     <NA> ARM678 BALB/c
# access Concomitant Medications data of SDY139
cm_df <- sdy139$interventions$cm_l$cm_df
head(cm_df)

## NULL
# get Trial Title from Trial Summary
ts_df <- sdy139$trial_design$ts_l$ts_df
title <- ts_df$TSVAL[ts_df$TSPARMCD== "TITLE"]
title

## [1] "The peptide specificity of the endogenous T follicular helper cell repertoire generated after p
```

8   Get the list of Domain names.

Note that some RImmPort functions take a domain name as input.

```
# get the list of names of all supported Domains
getListOfDomains()

##                               Domain Name Domain Code
## 1                           Adverse Events          AE
## 2                   Concomitant Medications          CM
## 3                             Demographics          DM
## 4                                 Exposure          EX
## 5                          Medical History          MH
## 6   Associated Persons Medical History        APMH
## 7                    Laboratory Test Results          LB
## 8                      Physical Examination          PE
## 9                       Protocol Deviations          DV
## 10                              Trial Arms          TA
## 11                             Trial Visits          TV
## 12  Trial Inclusion Exclusion Criteria          TI
## 13                           Trial Summary          TS
## 14                            Substance Use          SU
## 15                             Vital Signs          VS
```

```
## 16            Questionnaires         QS
## 17            Findings About         FA
## 18             Skin Response         SR
## 19           Genetics Findings       PF
## 20         Protein Quantification    ZA
## 21         Cellular Quantification   ZB
## 22      Nucleic Acid Quantification  ZC
## 23           Titer Assay Results     ZD
```
```r
?"Demographics Domain"
```

9 Get list of studies with specifc domain data

The Domain name should be exact to what is found in the list of Domain names.

```r
# get list of studies with Cellular Quantification data
domain_name <- "Cellular Quantification"
study_ids_l <- getStudiesWithSpecificDomainData(domain_name)
study_ids_l
```

```
## [1] "SDY139" "SDY208"
```

10 Get specifc domain data of one or more studies

The Domain name should be exact to what is found in the list of Domain names.

```r
# get Cellular Quantification data of studies `SDY139` and `SDY208`

# get domain code of Cellular Quantification domain
domain_name <- "Cellular Quantification"
getDomainCode(domain_name)
```

```
## [1] "ZB"
```

```r
study_ids <- c("SDY139", "SDY208")
domain_name <- "Cellular Quantification"
zb_l <- getDomainDataOfStudies(domain_name, study_ids)
```

```
## loading Cellular Quantification data....loading FCS Results data....done
## loading ELISPOT Results data....done
## done
## loading Cellular Quantification data....loading FCS Results data....done
## loading ELISPOT Results data....done
## done
```

```r
if (length(zb_l) > 0)
  names(zb_l)
```

```
## [1] "zb_df"      "suppzb_df"
```

```r
head(zb_l$zb_df)
```

```
##   STUDYID DOMAIN   USUBJID ZBSEQ      ZBTEST                  ZBCAT
## 1  SDY139     ZB SUB118078     1 Figure-7_FCM Cellular Quantification
## 2  SDY139     ZB SUB118078     2 Figure-7_FCM Cellular Quantification
## 3  SDY139     ZB SUB118078     3 Figure-7_FCM Cellular Quantification
## 4  SDY139     ZB SUB118078     4 Figure-7_FCM Cellular Quantification
```

```
## 5  SDY139     ZB SUB118078     5 Figure-7_FCM Cellular Quantification
## 6  SDY139     ZB SUB118078     6 Figure-7_FCM Cellular Quantification
##          ZBMETHOD ZBPOPDEF ZBPOPNAM ZBORRES ZBORRESU ZBBASPOP ZBSPEC VISITNUM
## 1 Flow Cytometry                                              Cell        1
## 2 Flow Cytometry                                              Cell        1
## 3 Flow Cytometry                                              Cell        1
## 4 Flow Cytometry                                              Cell        1
## 5 Flow Cytometry                                              Cell        1
## 6 Flow Cytometry                                              Cell        1
##                                                      VISIT ZBELTM
## 1 Day 0 Protein/peptide inoculation, SL_Sant_Plos1_2012_d0    POD
## 2 Day 0 Protein/peptide inoculation, SL_Sant_Plos1_2012_d0    POD
## 3 Day 0 Protein/peptide inoculation, SL_Sant_Plos1_2012_d0    POD
## 4 Day 0 Protein/peptide inoculation, SL_Sant_Plos1_2012_d0    POD
## 5 Day 0 Protein/peptide inoculation, SL_Sant_Plos1_2012_d0    POD
## 6 Day 0 Protein/peptide inoculation, SL_Sant_Plos1_2012_d0    POD
##                              ZBTPTREF  ZBREFID                        ZBXFN
## 1 Time of initial vaccine administration ES662746   Tfh_Tfh CLN D0-1.297191.fcs
## 2 Time of initial vaccine administration ES662746   Tfh_Tfh CLN D0-1.297192.txt
## 3 Time of initial vaccine administration ES662766 Tfh 1_Tfh EAR D0-1.297261.fcs
## 4 Time of initial vaccine administration ES662766 Tfh 1_Tfh EAR D0-1.297262.txt
## 5 Time of initial vaccine administration ES662786   Tfh_Tfh ILN D0-1.297339.fcs
## 6 Time of initial vaccine administration ES662786   Tfh_Tfh ILN D0-1.297340.txt
```

11   Get the list of assay types from ImmPort studies

```
getListOfAssayTypes()
```

```
## [1] "ELISA"      "ELISPOT"     "Array"       "PCR"
## [5] "HLA Typing"  "MBAA"        "HAI"         "Neut Ab Titer"
## [9] "Flow"
```

12   Get specific assay data of one or more Immport studies

The assay type should be exact to what is found in the list of supported assay types.

```
# get 'ELISPOT' data of study `SDY139`
assay_type <- "ELISPOT"
study_id = "SDY139"
elispot_l <- getAssayDataOfStudies(study_id, assay_type)
```

```
## loading Protein Quantification data....done
## loading Cellular Quantification data....loading ELISPOT Results data....done
## done
## loading Nucleic Acid Quantification data....done
## loading Titer Assay Results data....done
## loading Genetics Findings data....done
```

```
if (length(elispot_l) > 0)
  names(elispot_l)
```

```
## [1] "zb_df"      "suppzb_df"
```

```
head(elispot_l$zb_df)
```

```
##   STUDYID DOMAIN   USUBJID ZBSEQ           ZBTEST                     ZBCAT
## 1  SDY139     ZB SUB118053  8675 Figure-4_ELISPOT Cellular Quantification
## 2  SDY139     ZB SUB118053  8658 Figure-4_ELISPOT Cellular Quantification
## 3  SDY139     ZB SUB118053  8673 Figure-4_ELISPOT Cellular Quantification
## 4  SDY139     ZB SUB118053  8660 Figure-4_ELISPOT Cellular Quantification
## 5  SDY139     ZB SUB118053  8662 Figure-4_ELISPOT Cellular Quantification
## 6  SDY139     ZB SUB118053  8663 Figure-4_ELISPOT Cellular Quantification
##   ZBMETHOD ZBPOPDEF ZBPOPNAM  ZBORRES ZBORRESU ZBBASPOP ZBSPEC VISITNUM
## 1  ELISPOT      IL-2     IL-2 622.8571  1000000             Cell        3
## 2  ELISPOT     IL-21    IL-21   8337.5  1000000             Cell        3
## 3  ELISPOT      IL-2     IL-2 1048.571  1000000             Cell        3
## 4  ELISPOT     IL-21    IL-21   3925.0  1000000             Cell        3
## 5  ELISPOT     IL-21    IL-21    600.0  1000000             Cell        3
## 6  ELISPOT     IL-21    IL-21 798.5714  1000000             Cell        3
##                                          VISIT ZBELTM
## 1 Day 8 Sample collection, SL_Sant_Plos1_2012_d8    P8D
## 2 Day 8 Sample collection, SL_Sant_Plos1_2012_d8    P8D
## 3 Day 8 Sample collection, SL_Sant_Plos1_2012_d8    P8D
## 4 Day 8 Sample collection, SL_Sant_Plos1_2012_d8    P8D
## 5 Day 8 Sample collection, SL_Sant_Plos1_2012_d8    P8D
## 6 Day 8 Sample collection, SL_Sant_Plos1_2012_d8    P8D
##                                ZBTPTREF  ZBREFID ZBXFN
## 1 Time of initial vaccine administration ES661770
## 2 Time of initial vaccine administration ES661753
## 3 Time of initial vaccine administration ES661768
## 4 Time of initial vaccine administration ES661755
## 5 Time of initial vaccine administration ES661757
## 6 Time of initial vaccine administration ES661758
```

13   Serialize RImmPort-formatted study data as .rds files

```
# serialize all of the data of studies `SDY139` and `SDY208'
study_ids <- c('SDY139', 'SDY208')

# the folder where the .rds files will be stored
rds_dir <- file.path(studies_dir, "Rds")

serialzeStudyData(study_ids, rds_dir)
list.files(rds_dir)
```

14   Load the serialzed data (.rds) files of a specific domain of a study from the directory where the files are located

```
# get the directory where ImmPort sample data is stored in the directory structure of RImmPort package
studies_dir <- system.file("extdata", "ImmPortStudies", package = "RImmPort")

# the folder where the .rds files will be stored
rds_dir <- file.path(studies_dir, "Rds")
```

```r
# list the studies that have been serialized
list.files(rds_dir)
```

```
## [1] "SDY139" "SDY208"
```

```r
# load the serialized data of study `SDY208`
study_id <- 'SDY208'
dm_l <- loadSerializedStudyData(rds_dir, study_id, "Demographics")
```

```
##
## domain_file_path =  /private/tmp/Rtmp4KRKOV/Rinstc3b7803adb/RImmPort/extdata/ImmPortStudies/Rds/SDY2(
## suppdomain_file_path =  /private/tmp/Rtmp4KRKOV/Rinstc3b7803adb/RImmPort/extdata/ImmPortStudies/Rds/S
```

```r
head(dm_l[[1]])
```

```
##    STUDYID DOMAIN   USUBJID AGE   AGEU    SEX RACE ETHNIC    SPECIES STRAIN
## 1  SDY208      DM SUB120516   6 Weeks Female <NA>   <NA> Mus musculus   <NA>
## 2  SDY208      DM SUB120517   6 Weeks Female <NA>   <NA> Mus musculus   <NA>
## 3  SDY208      DM SUB120518   6 Weeks Female <NA>   <NA> Mus musculus   <NA>
## 4  SDY208      DM SUB120519   6 Weeks Female <NA>   <NA> Mus musculus   <NA>
## 5  SDY208      DM SUB120520   6 Weeks Female <NA>   <NA> Mus musculus   <NA>
## 6  SDY208      DM SUB120521   6 Weeks Female <NA>   <NA> Mus musculus   <NA>
##   SBSTRAIN  ARMCD
## 1     <NA> ARM881
## 2     <NA> ARM882
## 3     <NA> ARM883
## 4     <NA> ARM884
## 5     <NA> ARM885
## 6     <NA> ARM886
##
## 1                                   Microneedle vaccination- 5 ug inactivated A/Califor
## 2                                  Subcutaneous vaccination-  5 ug inactivated A/Califor
## 3                                                  Uncoated microneedle vacci
## 4  Microneedle vaccination- 5 ug inactivated A/California/04/09 virus, Challenged: 10x LD50 A/Califor
## 5 Subcutaneous vaccination- 5 ug inactivated A/California/04/09 virus, Challenged: 10x LD50 A/Califor
## 6                      Uncoated microneedle vaccination-  Placebo, Challenged: 10x LD50 A/Califor
```