

RImmPort: Quick Start Guide

Ravi Shankar

2016-01-22

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 regisiter 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)
```

```
## Loading required package: gsubfn
## Loading required package: proto
## Loading required package: RSQLite
```

```
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.

```
# 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.

```
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'

```
# 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)
```

```
## [1] "SDY139-DR16_Tab.zip" "SDY208-DR16_Tab.zip"
```

4.2.2 Build a local SQLite ImmPort database instance.

```
# 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)
```

```
## [1] "ImmPort.sqlite"
```

4.2.3 Connect to the ImmPort SQLite database

```
# 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

```
setImmPortDataSource(sqlite_conn)
```

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 Subject Visits 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 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 SPECIES STRAIN SBSTRAIN
## 1 SDY139 DM SUB118053 2 Months Unknown Mus musculus BALB/c NA
## 2 SDY139 DM SUB118054 2 Months Unknown Mus musculus BALB/c NA
## 3 SDY139 DM SUB118055 2 Months Unknown Mus musculus BALB/c NA
## 4 SDY139 DM SUB118056 2 Months Unknown Mus musculus BALB/c NA
## 5 SDY139 DM SUB118057 2 Months Unknown Mus musculus BALB/c NA
## 6 SDY139 DM SUB118058 2 Months Unknown Mus musculus BALB/c NA
## ARMCD ARM
## 1 ARM678 BALB/c
## 2 ARM678 BALB/c
## 3 ARM678 BALB/c
## 4 ARM678 BALB/c
## 5 ARM678 BALB/c
## 6 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 Inclusion Exclusion Criteria          TI  
## 12             Trial Summary          TS  
## 13          Subject Visits          SV  
## 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
```

```
?"Demographics Domain"
```

9 Get list of studies with specific domain data

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

```
# 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 specific domain data of one or more studies

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

```
# 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"
```

```
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
```

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

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

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

```
## STUDYID DOMAIN USUBJID ZBSEQ ZBTEST ZBCAT ZBMETHOD
## 1 SDY139 ZB SUB118106 1 Figure-1_FCM Cellular_Phenotype FCM
## 2 SDY139 ZB SUB118106 2 Figure-1_FCM Cellular_Phenotype FCM
## 3 SDY139 ZB SUB118106 3 Figure-1_FCM Cellular_Phenotype FCM
## 4 SDY139 ZB SUB118106 4 Figure-1_FCM Cellular_Phenotype FCM
## 5 SDY139 ZB SUB118106 5 Figure-1_FCM Cellular_Phenotype FCM
## 6 SDY139 ZB SUB118106 6 Figure-1_FCM Cellular_Phenotype FCM
## ZBPOPDEF
## 1 B220- live,doublet excluded,CD4+ CD44low,CCR7+
## 2 B220- live,doublet excluded,CD4+ CD44highCXCR5highPD1high,CCR7+
## 3 B220- live,doublet excluded,CD4+ CD44high,either CXCR5dim or PD1dim,CCR7+
## 4 B220- live,doublet excluded,CD4+ CD44highCXCR5highPD1high,CD62L+
## 5 B220- live,doublet excluded,CD4+ CD44low,CD62L+
## 6 B220- live,doublet excluded,CD4+ CD44high,either CXCR5dim or PD1dim,CD62L+
## ZBPOPNAME ZBORRES ZBORRESU ZBBASPOP ZBSPEC
## 1 CCR7+ CD4+CD44low 192974 events per experiment sample <NA> Cell
## 2 CCR7+ Tfh 1701 events per experiment sample <NA> Cell
## 3 CCR7+ Non-Tfh 43396 events per experiment sample <NA> Cell
## 4 CD62L+ Tfh 1386 events per experiment sample <NA> Cell
## 5 CD62L+ CD4+CD44low 166232 events per experiment sample <NA> Cell
## 6 CD62L+ Non-Tfh 38427 events per experiment sample <NA> Cell
## VISITNUM VISIT ZBELTM
## 1 3 Day 8 Sample collection, SL_Sant_Plos1_2012_d8 P8D
## 2 3 Day 8 Sample collection, SL_Sant_Plos1_2012_d8 P8D
## 3 3 Day 8 Sample collection, SL_Sant_Plos1_2012_d8 P8D
```

```
## 4      3 Day 8 Sample collection, SL_Sant_Plos1_2012_d8      P8D
## 5      3 Day 8 Sample collection, SL_Sant_Plos1_2012_d8      P8D
## 6      3 Day 8 Sample collection, SL_Sant_Plos1_2012_d8      P8D
##                                ZBTPTREF  ZBREFID                                ZBXFN
## 1 Time of initial vaccine administration BS645819  CCR7 PAN_Tube_009.fcs
## 2 Time of initial vaccine administration BS645819  CCR7 PAN_Tube_009.fcs
## 3 Time of initial vaccine administration BS645819  CCR7 PAN_Tube_009.fcs
## 4 Time of initial vaccine administration BS645819  CD62L Pan_Tube_009.fcs
## 5 Time of initial vaccine administration BS645819  CD62L Pan_Tube_009.fcs
## 6 Time of initial vaccine administration BS645819  CD62L Pan_Tube_009.fcs
```

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_1 <- 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_1) > 0)
  names(elispot_1)
```

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

```
head(elispot_1$zb_df)
```

```
##   STUDYID DOMAIN  USUBJID ZBSEQ          ZBTEST          ZBCAT
## 1  SDY139      ZB SUB118053 8657 Figure-4_ELISPOT Cellular_Quantification
## 2  SDY139      ZB SUB118053 8658 Figure-4_ELISPOT Cellular_Quantification
## 3  SDY139      ZB SUB118053 8659 Figure-4_ELISPOT Cellular_Quantification
## 4  SDY139      ZB SUB118053 8660 Figure-4_ELISPOT Cellular_Quantification
```

```
## 5 SDY139 ZB SUB118053 8661 Figure-4_ELISPOT Cellular_Quantification
## 6 SDY139 ZB SUB118053 8662 Figure-4_ELISPOT Cellular_Quantification
## ZBMETHOD ZBPOPDEF ZBPOPNAME ZBORRES ZBORRESU ZBBASPOP
## 1 ELISPOT IL-21 IL-21 5312.50 1000000 TfH cells TfH cells
## 2 ELISPOT IL-21 IL-21 8337.50 1000000 TfH cells TfH cells
## 3 ELISPOT IL-21 IL-21 11775.00 1000000 TfH cells TfH cells
## 4 ELISPOT IL-21 IL-21 3925.00 1000000 TfH cells TfH cells
## 5 ELISPOT IL-21 IL-21 1931.25 1000000 TfH cells TfH cells
## 6 ELISPOT IL-21 IL-21 600.00 1000000 Non-TfH cells Non-TfH cells
## ZBSPEC VISITNUM VISIT ZBELTM
## 1 Cell 3 Day 8 Sample collection, SL_Sant_Plos1_2012_d8 P8D
## 2 Cell 3 Day 8 Sample collection, SL_Sant_Plos1_2012_d8 P8D
## 3 Cell 3 Day 8 Sample collection, SL_Sant_Plos1_2012_d8 P8D
## 4 Cell 3 Day 8 Sample collection, SL_Sant_Plos1_2012_d8 P8D
## 5 Cell 3 Day 8 Sample collection, SL_Sant_Plos1_2012_d8 P8D
## 6 Cell 3 Day 8 Sample collection, SL_Sant_Plos1_2012_d8 P8D
## ZBPTREF ZBREFID ZBXFN
## 1 Time of initial vaccine administration BS645680 ELISPOT_Results.txt
## 2 Time of initial vaccine administration BS645680 ELISPOT_Results.txt
## 3 Time of initial vaccine administration BS645680 ELISPOT_Results.txt
## 4 Time of initial vaccine administration BS645680 ELISPOT_Results.txt
## 5 Time of initial vaccine administration BS645680 ELISPOT_Results.txt
## 6 Time of initial vaccine administration BS645679 ELISPOT_Results.txt
```

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")

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

14 Load the serialized 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")

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

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



```

# load the serialized data of study `SDY208`
study_id <- 'SDY208'
dm_1 <- loadSerializedStudyData(rds_dir, study_id, "Demographics")

##
## domain_file_path = /Users/user/Projects/RImmPort/inst/extdata/ImmPortStudies/Rds/SDY208/dm.rds
## suppdomain_file_path = /Users/user/Projects/RImmPort/inst/extdata/ImmPortStudies/Rds/SDY208/suppdm.

head(dm_1[[1]])

##   STUDYID DOMAIN   USUBJID AGE  AGEU   SEX   SPECIES STRAIN SBSTRAIN
## 1  SDY208     DM SUB120516  6 Weeks Female Mus musculus <NA>      NA
## 2  SDY208     DM SUB120517  6 Weeks Female Mus musculus <NA>      NA
## 3  SDY208     DM SUB120518  6 Weeks Female Mus musculus <NA>      NA
## 4  SDY208     DM SUB120519  6 Weeks Female Mus musculus <NA>      NA
## 5  SDY208     DM SUB120520  6 Weeks Female Mus musculus <NA>      NA
## 6  SDY208     DM SUB120521  6 Weeks Female Mus musculus <NA>      NA
##   ARMCD
## 1 ARM881
## 2 ARM882
## 3 ARM883
## 4 ARM884
## 5 ARM885
## 6 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

```