The Xeva User’s Guide

**Arvind Mer**\(^1,2\) and **Benjamin Haibe-Kains**\(^1,2,3,4,5\)

\(^1\)Princess Margaret Cancer Centre, University Health Network, Toronto, Canada
\(^2\)Department of Medical Biophysics, University of Toronto, Toronto, Canada
\(^3\)Department of Computer Science, University of Toronto, Toronto, Canada
\(^4\)Vector Institute, Toronto, Ontario, Canada
\(^5\)Ontario Institute for Cancer Research, Toronto, Ontario, Canada

October 24, 2023

Contents

1 Introduction ................................................. 2
2 Installation and Settings .................................. 2
3 Definitions ..................................................... 2
4 Data Access ................................................... 3
5 Visualizing PDX Growth Curve ......................... 4
6 Replicate-based PDX experiments ...................... 7
7 PDX Model Drug Response ............................... 8
8 Gene-drug association .................................... 11
9 Creating new Xeva object ................................. 12
10 SessionInfo ............................................... 16
1 Introduction

The Xeva package provides efficient and powerful functions for patient-driven xenograft (PDX) based pharmacogenomic data analysis [1].

2 Installation and Settings

Xeva requires that several packages be installed. All dependencies are available from CRAN or Bioconductor:

```r
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install("Xeva", version = "3.8")
```

The package can also be installed directly from GitHub using devtools:

```r
#install devtools if required
devtools::install_github("bhklab/Xeva")
```

Load Xeva into your current workspace:

```r
library(Xeva)
```

Load the dataset you wish to analyze. For the sake of this tutorial, here we load the Novartis PDXE [2] breast cancer dataset as an example:

```r
data(brca)
print(brca)
```

```r
## XevaSet
## name: PDXE.BRCA
## Creation date: Fri Sep 14 11:41:33 2018
## Number of models: 849
## Number of drugs: 22
## Moleculer dataset: RNASeq, mutation, cnv
```

3 Definitions

Before we further dive into the analysis and visualization, it is important to understand the terminology used in the Xeva package. In a Xeva object, the experiment slot stores the data for each individual PDX/mouse. With the exception of tumor growth data (time vs. tumor volume), for each individual PDX/mouse, you can access metadata such as the patient’s age, sex, tissue histology, and passage information. All of this metadata is stored in the pdxModel class, where a unique ID called model.id is given to each PDX/mouse model. As for the tumor growth information, Xeva provides separate functions for retrieving and visualizing time vs. tumor volume data. We will see later how to get these data for an individual model.id, but first, let’s define some other terms that appear in the Xeva package.
A PDX experiment can be one of the two categories:

- **treatment** represents experiments in which the PDX receives some kind of drug (or drug combination)
- **control** represents experiments in which the PDX receives no drug

To see the effect of a drug, several replicate experiments are done for both the control and the treatment categories. In **Xeva**, a collection of PDX model.ids originating from the same patient is organized in **batches** (batch). A batch has two arms: **control** and **treatment**. This is illustrated in Figure 1.

**Figure 1**: A PDX experiment. The text under each of the PDX/mouse (ie. m1, m2, p1, etc.) denotes the model.id in **Xeva**. In this example, three PDXs are declared as control (m1, m2, and m3). Similarly, in the treatment arm, 3 PDXs are given the drug paclitaxel (p1, p2, and p3), 3 are given tamoxifen (t1, t2, and t3), and 3 are given binimetinib (b1, b2, b3). The PDXs in the control arm and one of the treatment arms together constitute a batch. For example, control arm models (m1, m2, and m3) and treatment arm models (t1, t2, and t3) together create a batch called batch-2.

A **Xeva** object binds together all individual experiments, batch information, and molecular data into one single class called **XevaSet**.

## 4 Data Access

As mentioned earlier, **Xeva** stores metadata for each individual PDX model. We can retrieve the meta-information about each PDX, such as number of models and tissue type, using:

```r
brca.mod <- modelInfo(brca)
dim(brca.mod)
## [1] 849 5
brca.mod[1:4,]
##            model.id tissue tissue.name patient.id     drug
## X.1004.BG98 X.1004.BG98 BRCA Breast Cancer X-1004 BGJ398
## X.1004.biib X.1004.biib BRCA Breast Cancer X-1004 binimetinib
## X.1004.BK20 X.1004.BK20 BRCA Breast Cancer X-1004 BKM120
## X.1004.BY19 X.1004.BY19 BRCA Breast Cancer X-1004 BYL719
```
The output shows that the brca dataset contains 849 PDX models. We can also see the time vs. tumor volume data for a model using:

```r
model.data <- getExperiment(brca, model.id = "X.1004.BG98")
head(model.data)
##   model.id drug.join.name time volume body.weight volume.normal
## 1     X.1004.BG98     BGJ398   0   199.7     28.2         0.0000000
## 2     X.1004.BG98     BGJ398   2   181.9     28.0       -0.0891337
## 3     X.1004.BG98     BGJ398   5   172.7     28.4       -0.1352028
## 4     X.1004.BG98     BGJ398   9   129.6     27.2       -0.3510265
## 5     X.1004.BG98     BGJ398  12    91.3     26.7       -0.5428142
## 6     X.1004.BG98     BGJ398  16   117.1     26.2       -0.4136204
```

Similarly, for **batch** names, we can obtain all predefined batch names using:

```r
batch.name <- batchInfo(brca)
batch.name[1:4]
## [1] "X-1004.BGJ398"  "X-1004.binimetinib"  "X-1004.BKM120"
## [4] "X-1004.BYL719"
```

The information about a **batch** can be shown using:

```r
batchInfo(brca, batch = "X-1004.binimetinib")
## $X-1004.binimetinib
##   name       control       treatment
## X-1004.binimetinib X.1004.uned X.1004.biib
```

Here, for the batch named **X-1004.binimetinib**, we can see that the control sample is **X.1004.uned** and the treatment sample is **X.1004.biib**.

### 5 Visualizing PDX Growth Curve

Xeva provides a function to plot time vs. tumor volume data for individual models as well as for individual batches. These data can be plotted by using the name of the batch:

```r
plotPDX(brca, batch = "X-4567.BKM120")
```

You can choose to see different aspects of this visualization. For example, we can plot normalized volume; we can also change the colors of the lines:

```r
plotPDX(brca, batch = "X-4567.BKM120", vol.normal = TRUE, control.col = "#a6611a",
        treatment.col = "#018571", major.line.size = 1, max.time = 40)
```

Data can also be visualized at the patient level by specifying **patient.id**:

```r
plotPDX(brca, patient.id="X-3078", drug="paclitaxel",control.name = "untreated")
```
Figure 2: Tumor growth curves for a batch of control and treated PDXs

Figure 3: Tumor growth curves for a batch of control and treated PDXs. Here, the volume is normalized and plots are truncated at 40 days
Figure 4: Tumor growth curves for a batch of control and treated PDXs generated using patient ID and drug name.
6 Replicate-based PDX experiments

Xeva can also handle replicate-based experiment design. The datasets included in the package also contain replicate-based PDX experiments. To plot replicate-based data:

```r
data("repdx")
plotPDX(repdx, vol.normal = TRUE, batch = "P1")
```

![Tumor growth curves for a batch of control and treated PDXs with replicates](image)

**Figure 5:** Tumor growth curves for a batch of control and treated PDXs with replicates

```r
plotPDX(repdx, batch = "P3", SE.plot = "errorbar")
plotPDX(repdx, batch = "P4", vol.normal = TRUE, SE.plot = "ribbon")
```
Xeva can effectively summarize PDX drug response data. Here we summarize the \texttt{mRECIST} values for the models in our dataset:

```r
brca_mr <- summarizeResponse(brca, response.measure = "mRECIST")
brca_mr[1:5, 1:4]
```
Waterfall plots are also commonly used to visualize PDX drug response data. Xeva provides a function to visualize and color waterfall plots:

```r
waterfall(brca, drug="binimetinib", res.measure="best.average.response")
```

![Waterfall plot for binimetinib drug response in PDXs](image)

**Figure 8:** Waterfall plot for binimetinib drug response in PDXs

It is useful to color the bars of your waterfall plot by genomic properties. Here we create a waterfall plot for drug BYL719 and color it based on the mutation status of the CDK13 gene. First, we extract the genomic data for the models. Then, we can plot the waterfall plots:

```r
mut <- summarizeMolecularProfiles(brca, drug = "BYL719", mDataType="mutation")
model.type <- Biobase::exprs(mut)["CDK13", ]
model.type[grepl("Mut", model.type)] <- "mutation"
model.type[grepl("Del", model.type)] <- "deletion"
model.type[grepl("Mut", model.type)] <- "wild type"
model.color <- list("mutation"="#b2182b", "wild type"="#878787")
waterfall(brca, drug="BYL719", res.measure="best.average.response", model.id=names(model.type), model.type= model.type, type.color = model.color)
```

In Xeva we have implemented difference matrix to compute PDX response. The Xeva function `response` provides a unified interface for this purpose. In the example below we compute the angle between treatment and control PDXs.
A function for linear mixed-effects model (lmm) has also been implemented.

```r
data("repdx")
response(repdx, batch="P1", res.measure="lmm")
## computing lmm for batch P1
## Linear mixed-effects model fit by REML
## Data: data
## Log-restricted-likelihood: 3.184927
## Fixed: log(volume) ~ time * exp.type
## (Intercept) time exp.typetreatment
## 5.32592390 0.03091321 -0.28718018
## time:exp.typetreatment
## -0.01983870
##
## Random effects:
## Formula: ~1 | model.id
## (Intercept) Residual
## StdDev: 0.3802407 0.1975011
##
## Number of Observations: 194
```
The Xeva User’s Guide

## Number of Groups: 12

### 8 Gene-drug association

The main aim of the pharmacogenomic experiments is to find biomarkers for drug response prediction. The Xeva package provides the `drugSensitivitySig` function to compute the univariate association between PDX’s molecular data (such as gene expression) and response to a drug (gene-drug association). In the example below, we are computing the association between gene expression (RNASeq) and slope of the PDXs for the drug ‘tamoxifen’ using linear regression (`lm`).

```r
data(brca)
drugSensitivitySig(object=brca, drug="tamoxifen", mDataType="RNASeq",
features=c(1,2,3,4,5), sensitivity.measure="slope", fit="lm")
```

```
## Running for drug tamoxifen
## feature drug estimate se n tstat fstat
## 1 PIK3R6 tamoxifen 0.12174756 0.1654268 38 0.73596007 0.541637222
## 2 FGFR1 tamoxifen -0.16019355 0.1645143 38 -0.97373653 0.948162830
## 3 NTRK2 tamoxifen 0.16560706 0.1643653 38 1.00755487 1.015166810
## 4 AKT1 tamoxifen 0.00996873 0.1666584 38 0.05981535 0.003577876
## 5 FGFR4 tamoxifen 0.07953256 0.1661387 38 0.47871178 0.229164971
## pvalue df fdr
## 1 0.4665236 36 0.7775393
## 2 0.3366852 36 0.7775393
## 3 0.3203927 36 0.7775393
## 4 0.9526335 36 0.7937981
## 5 0.6350385 36 0.7937981
```

In this above example we took only 5 features (genes), however this can be extended for any number of genes. For larger analyses, this function also provides out of box parallel computation. Users can set the number of cores using the parameter `nthread` in the function.

Users can choose different sensitivity measures of the PDX response for the association analysis by setting the parameter `sensitivity.measure`. For example, below we use `best.average.response` as the PDX’s response matrix in the association analysis:

```r
data(brca)
drugSensitivitySig(object=brca, drug="tamoxifen", mDataType="RNASeq",
features=c(1,2,3,4,5), sensitivity.measure="best.average.response", fit="lm")
```

```
## Running for drug tamoxifen
## feature drug estimate se n tstat fstat
## 1 PIK3R6 tamoxifen 0.10155551 0.1658050 38 0.6124998 0.3751559 0.5440570
## 2 FGFR1 tamoxifen 0.25347267 0.1612238 38 1.5721794 2.4717480 0.1246577
## 3 NTRK2 tamoxifen 0.15268488 0.1647125 38 0.9269782 0.8592885 0.3601112
## 4 AKT1 tamoxifen -0.19722241 0.1633931 38 -1.2070422 1.4569510 0.2352874
## 5 FGFR4 tamoxifen -0.06903067 0.1662691 38 -0.4151744 0.1723698 0.6804783
## pvalue df fdr
## 1 0.6800713
## 2 0.5882184
```
For the drug-gene association analysis, users can also choose a different method of association calculation (such as concordance index, Pearson or Spearman correlation) by setting the parameter `fit`.

```r
data(brca)
drugSensitivitySig(object=brca, drug="tamoxifen", mDataType="RNASeq",
                      features=c(1,2,3,4,5),
                      sensitivity.measure="best.average.response", fit="pearson")
```

## Running for drug tamoxifen

<table>
<thead>
<tr>
<th>feature</th>
<th>drug</th>
<th>cor</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIK3R6</td>
<td>tamoxifen</td>
<td>0.10155551</td>
<td>0.5440570</td>
</tr>
<tr>
<td>FGFR1</td>
<td>tamoxifen</td>
<td>0.25347267</td>
<td>0.1246577</td>
</tr>
<tr>
<td>NTRK2</td>
<td>tamoxifen</td>
<td>0.15268488</td>
<td>0.3601112</td>
</tr>
<tr>
<td>AKT1</td>
<td>tamoxifen</td>
<td>-0.19722241</td>
<td>0.2352874</td>
</tr>
<tr>
<td>FGFR4</td>
<td>tamoxifen</td>
<td>-0.06903067</td>
<td>0.6804783</td>
</tr>
</tbody>
</table>

9 Creating new Xeva object

New Xeva objects can be created using `createXevaSet`. The different components that are needed by the function is as follows:

- **model** A data.frame containing `model.id` and other relevant model information, such as tissue of origin, patient ID, and passage information. All PDXMI variables can be inserted into this data.frame. A required column in the data.frame is `model.id`. An example of the `model` can be found in the package as:

```r
model=read.csv(system.file("extdata", "model.csv", package = "Xeva"))
head(model)
```

<table>
<thead>
<tr>
<th>model.id</th>
<th>tissue</th>
<th>tissue.name</th>
<th>patient.id</th>
</tr>
</thead>
<tbody>
<tr>
<td>X.1004.biib</td>
<td>BRCA</td>
<td>Breast Cancer</td>
<td>X-1004</td>
</tr>
<tr>
<td>X.1008.biib</td>
<td>BRCA</td>
<td>Breast Cancer</td>
<td>X-1008</td>
</tr>
<tr>
<td>X.1286.biib</td>
<td>BRCA</td>
<td>Breast Cancer</td>
<td>X-1286</td>
</tr>
<tr>
<td>X.1004.uned</td>
<td>BRCA</td>
<td>Breast Cancer</td>
<td>X-1004</td>
</tr>
<tr>
<td>X.1008.uned</td>
<td>BRCA</td>
<td>Breast Cancer</td>
<td>X-1008</td>
</tr>
<tr>
<td>X.1286.uned</td>
<td>BRCA</td>
<td>Breast Cancer</td>
<td>X-1286</td>
</tr>
</tbody>
</table>

- **drug** A data.frame containing information about the drugs used in the experiments. The required column is `drug.id`. An example of the `drug` can be found in the package as:

```r
drug=read.csv(system.file("extdata", "drug.csv", package = "Xeva"))
head(drug)
```

<table>
<thead>
<tr>
<th>drug.id</th>
<th>standard.name</th>
<th>treatment.target</th>
<th>treatment.type</th>
</tr>
</thead>
<tbody>
<tr>
<td>binimetinib</td>
<td>binimetinib</td>
<td>MAP2K1,MAP2K2,MAPK</td>
<td>single</td>
</tr>
</tbody>
</table>
The Xeva User’s Guide

## 2 untreated untreated UNKNOWN,Untreated single

### experiment
A data.frame with time vs. tumor volume information. The required columns are `model.id`, `time`, `volume`, and `drug`. Other available information such as dose amount and mouse weight can be specified as columns of this data.frame. An example of the `experiment` can be found in the package as:

```r
experiment = read.csv(system.file("extdata", "experiments.csv", package="Xeva"))
head(experiment)
```

<table>
<thead>
<tr>
<th>#</th>
<th>model.id</th>
<th>time</th>
<th>volume</th>
<th>drug</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X.1004.biib</td>
<td>0</td>
<td>250.8</td>
<td>binimetinib</td>
</tr>
<tr>
<td>2</td>
<td>X.1004.biib</td>
<td>3</td>
<td>320.4</td>
<td>binimetinib</td>
</tr>
<tr>
<td>3</td>
<td>X.1004.biib</td>
<td>7</td>
<td>402.3</td>
<td>binimetinib</td>
</tr>
<tr>
<td>4</td>
<td>X.1004.biib</td>
<td>11</td>
<td>382.6</td>
<td>binimetinib</td>
</tr>
<tr>
<td>5</td>
<td>X.1004.biib</td>
<td>18</td>
<td>384.0</td>
<td>binimetinib</td>
</tr>
<tr>
<td>6</td>
<td>X.1004.biib</td>
<td>22</td>
<td>445.9</td>
<td>binimetinib</td>
</tr>
</tbody>
</table>

### expDesign
This is an R list object that contains the batch information. Each element of the list should have 3 components `batch.name`, `treatment` and `control`. The model.ids in the treatment and control must be present in `model` variable described earlier. An example of `expDesign` is:

```r
expDesign = readRDS(system.file("extdata", "batch_list.rds", package="Xeva"))
expDesign[[1]]
```

```r
## $batch.name
## [1] "X-1004.binimetinib"

## $treatment
## [1] "X.1004.biib"

## $control
## [1] "X.1004.uned"
```

### molecularProfiles
This list contains all the molecular data such as RNAseq or mutation information. Each element of this list should contain an `ExpressionSet` object. An example of such an object is:

```r
RNASeq = readRDS(system.file("extdata", "rnaseq.rds", package="Xeva"))
print(RNASeq)
```

```r
# ExpressionSet (storageMode: lockedEnvironment)
# assayData: 22 features, 2 samples
#   element names: exprs
# protocolData: none
# phenoData
# sampleNames: X-1004 X-1008
```
modToBiobaseMap A data.frame which contains mapping between the PDX model.id and the molecularProfiles sample names. It requires 3 variables: model.id, biobase.id, and mDataType. An example of modToBiobaseMap is:

```r
modToBiobaseMap = read.csv(system.file("extdata","modelToExpressionMap.csv", package = "Xeva"))
```

14

A new Xeva object can be created as:

```r
xeva.set = createXevaSet(name = "example xevaSet", model = model, drug = drug, experiment = experiment, expDesign = expDesign, molecularProfiles = list(RNASeq = RNASeq), modToBiobaseMap = modToBiobaseMap)
```

14
References


sessionInfo()

## R version 4.3.1 (2023-06-16)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 22.04.3 LTS
##
## Matrix products: default
## BLAS: /home/biocbuild/bbs-3.18-bioc/R/lib/libRblas.so
## LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.10.0
##
## locale:
## [1] LC_CTYPE=en_US.UTF-8 LC_NUMERIC=C
## [3] LC_TIME=en_GB LC_COLLATE=C
## [5] LC_MONETARY=en_US.UTF-8 LC_MESSAGES=en_US.UTF-8
## [7] LC_PAPER=en_US.UTF-8 LC_NAME=C
## [9] LC_ADDRESS=C LC_TELEPHONE=C
##
## attached base packages:
## [1] stats graphics grDevices utils datasets methods base
##
## other attached packages:
## [1] Biobase_2.62.0 BiocGenerics_0.48.0 Xeva_1.18.0
##
## loaded via a namespace (and not attached):
## [1] RCColorBrewer_1.1-3 Rmisc_1.5.1
## [3] jsonlite_1.8.7 shape_1.4.6
## [5] MultiAssayExperiment_1.28.0 magrittr_2.0.3
## [7] farver_2.1.1 rmarkdown_2.25
## [9] GlobalOptions_0.1.2 zlibbioc_1.48.0
## [11] vctrs_0.6.4 magicaxis_2.2.14
## [13] RCurl_1.98-1.12 htmltools_0.5.6.1
## [15] S4Vectors_1.2.0 plotrix_3.8-2
## [17] SparseArray_1.2.0 pracma_2.4.2
## [19] KernSmooth_2.23-22 htmlwidgets_1.6.2
## [21] plyr_1.8.9 igraph_1.5.1
## [23] mime_0.12 lifecycle_1.0.3
## [25] piano_2.18.0 iterators_1.0.14
## [27] pkgconfig_2.0.3 Matrix_1.6-1.1
## [29] R6_2.5.1 fastmap_1.1.1
## [31] GenomeInfoDbData_1.2.11 MatrixGenerics_1.14.0
## [33] shiny_1.7.5.1 clue_0.3-65
## [35] digest_0.6.33 colorspace_2.1-0
## [37] S4Vectors_0.40.0 GenomicRanges_1.54.0
## [39] SnowballC_0.7.1 CoreGx_2.6.0
## [41] labeling_0.4.3 fansi_1.0.5
The Xeva User's Guide

```r
## [43] abind_1.4-5        coop_0.6-3
## [45] compiler_4.3.1     withr_2.5.1
## [47] doParallel_1.0.17  downloader_0.4
## [49] marray_1.80.0      backports_1.4.1
## [51] BiocParallel_1.36.0 bench_1.1.3
## [53] PharmacoGx_3.6.0    highr_0.10
## [55] gplots_3.1.3        maps_3.4.1
## [57] MASS_7.3-60         DelayedArray_0.28.0
## [59] rjson_0.2.21        gtools_3.9.4
## [61] caTools_1.18.2      tools_4.3.1
## [63] NISTunits_1.0.1     httpuv_1.6.12
## [65] relations_0.6-13    glue_1.6.2
## [67] nlme_3.1-163        promises_1.2.1
## [69] grid_4.3.1          checkmate_2.2.0
## [71] cluster_2.1.4       reshape2_1.4.4
## [73] fgsea_1.28.0        generics_0.1.3
## [75] gttable_0.3.4       sm_2.2-5.7.1
## [77] data.table_1.14.8  utf8_1.2.4
## [79] XVector_0.42.0      RANN_2.6.1
## [81] foreach_1.5.2       pillar_1.9.0
## [83] stringr_1.5.0       limma_3.58.0
## [85] BumpyMatrix_1.10.0  later_1.3.1
## [87] circlize_0.4.15     dplyr_1.1.3
## [89] BBmisc_1.13         lattice_0.22-5
## [91] tidyselect_1.2.0    ComplexHeatmap_2.18.0
## [93] knitr_1.44           IRanges_2.36.0
## [95] SummarizedExperiment_1.32.0 stats4_4.3.1
## [97] xfun_0.40            shinydashboard_0.7.2
## [99] statmod_1.5.0       matrixStats_1.0.0
## [101] DT_0.30              visNetwork_2.1.2
## [103] stringi_1.7.12      yaml_2.3.7
## [105] boot_1.3-28.1       evaluate_0.22
## [107] codetools_0.2-19    lsa_0.73.3
## [109] tibble_3.2.1        BiocManager_1.30.22
## [111] cli_3.6.1            xtable_1.8-4
## [113] munsell_0.5.0       Rcpp_1.0.11
## [115] GenomeInfoDb_1.38.0 mapproj_1.2.11
## [117] png_0.1-8            parallel_4.3.1
## [119] sets_1.0-24          ellipsis_0.3.2
## [121] ggplot2_3.4.4        bitops_1.0-7
## [123] slam_0.1-50          celestial_1.4.6
## [125] scales_1.2.1         crayon_1.5.2
## [127] BiocStyle_2.30.0    GetoptLong_1.0.5
## [129] rlang_1.1.1           cowplot_1.1.1
## [131] fastmatch_1.1-4      shinyjs_2.1.0
```