

Data Exploration and Monotonic Trend Testing

NMTox is an R package created to perform data exploration and to identify monotonic trend in the dose-response relationship of nanomaterial toxicity. It consists of functions that can be used to generate frequency tables and graphical displays and to perform inference by testing the monotonic trend. This document shows how NMTox can be used to perform data exploration and inference.

As an illustration on how to use the package, data of genetic toxicity in vitro (`gendata`) and specific data of Multi-walled carbon nanotubes nanomaterial are provided as supplementary documents. These data are available in the eNanoMapper database, as parts of the data of NanoinformaTIX project. The data can be read into R as follows

```
gendata<-readxl::read_xlsx("...\\gen invitro.xlsx")
nm400<-readxl::read_xlsx("...\\nm400.xlsx")
```

Example of `gendata` data:

```
head(gendata)

## # A tibble: 6 x 17
##   name    publicname supplier experimentID method studyprovider endpoint value
##   <chr>    <chr>      <chr>       <chr>      <chr>      <chr>      <dbl>
## 1 control CONTROL    NanoGeno~ d0714825-23~ MICRO~ IPL          MNBNC F~  0.006
## 2 control CONTROL    NanoGeno~ d0714825-23~ MICRO~ IPL          MNBNC F~  0.008
## 3 control CONTROL    NanoGeno~ 4097f223-de~ COMET     UAB          DNA STR~  6.88
## 4 control CONTROL    NanoGeno~ 4097f223-de~ COMET     UAB          DNA STR~  7.36
## 5 control CONTROL    NanoGeno~ da92beeb-b7~ COMET     IPH          DNA STR~ 11.2
## 6 control CONTROL    NanoGeno~ da92beeb-b7~ COMET     IPH          DNA STR~ 23.8
## # ... with 9 more variables: unit <chr>, celltype <chr>, treatment <chr>,
## #   exptimeunit <chr>, exptime <chr>, concentration_unit <chr>,
## #   concentration <chr>, concentration_ml_unit <chr>, concentration_ml <dbl>
```

Case study: Genetic toxicity in vitro

Function `Frtab` can be used to explore variables in the `gen invitro` dataset. By specifying `opt="list"` and `x="name"`, the list of the unique values of variable `name` will be given.

```
Frtab(data=gendata, x="name", opt="list")

## # A tibble: 22 x 1
##   name
##   <chr>
## 1 control
## 2 untreated
## 3 medium
## 4 medium + BSA
## 5 Control
## 6 NM-103 (Titanium Dioxide)
```

```

## 7 NM-400 (Multi-walled carbon nanotubes)
## 8 NM-401 (Multi-walled carbon nanotubes)
## 9 NM-402 (Multi-walled carbon nanotubes)
## 10 MWCNT (Mitsui)
## # ... with 12 more rows

```

It can be seen that there are 22 different names identified in `geninvitro` data. They consist of controls and nanomaterial observations. Using the same function, `Frtab`, the toxicity endpoint(s) for each nanomaterial can be obtained. The number of observations for each unique values of the variable can be shown by specifying `opt="ls.obs"`.

```

Frtab(data=geninvitro, x="endpoint", cat="name", opt="ls.obs")

## # A tibble: 39 x 3
##   name      endpoint     n
##   <chr>    <chr>       <int>
## 1 control  DNA STRAND BREAKS 306
## 2 control  MNBNC FREQUENCY 112
## 3 Control  DNA STRAND BREAKS 70
## 4 Control  MNBNC FREQUENCY 38
## 5 medium   DNA STRAND BREAKS 21
## 6 medium + BSA DNA STRAND BREAKS 25
## 7 MWCNT (Cheap Tubes) DNA STRAND BREAKS 157
## 8 MWCNT (Cheap Tubes) MNBNC FREQUENCY 34
## 9 MWCNT (Mitsui)    DNA STRAND BREAKS 157
## 10 MWCNT (Mitsui)   MNBNC FREQUENCY 31
## # ... with 29 more rows

```

To see the units of the concentration in the data, variable `concentration_unit` can be added in the `x` as follows:

```

Frtab(data=gendata, x=c("endpoint", "concentration_unit"), cat="name", opt="ls.obs")

## # A tibble: 64 x 4
##   name      endpoint      concentration_unit     n
##   <chr>    <chr>        <chr>           <int>
## 1 control  DNA STRAND BREAKS ug/cm2          302
## 2 control  DNA STRAND BREAKS <NA>            4
## 3 control  MNBNC FREQUENCY ug/cm2          99
## 4 control  MNBNC FREQUENCY <NA>           13
## 5 Control  DNA STRAND BREAKS ug/cm2          66
## 6 Control  DNA STRAND BREAKS <NA>            4
## 7 Control  MNBNC FREQUENCY ug/cm2          36
## 8 Control  MNBNC FREQUENCY <NA>            2
## 9 medium   DNA STRAND BREAKS ug/cm2          21
## 10 medium + BSA DNA STRAND BREAKS ug/cm2         25
## # ... with 54 more rows

```

Suppose if we are interested in a specific nanomaterial, such as `NM-400 (Multi-walled carbon nanotubes)`, `val="NM-400 (Multi-walled carbon nanotubes)"` can be specified in the function `Frtab` and the endpoints measured for this nanomaterial (and the number of observations) are shown as follows:

```

Frtab(gendata, cat="name", val="NM-400 (Multi-walled carbon nanotubes)",
x="endpoint", opt="ls.obs")

```

```
## # A tibble: 2 x 2
##   endpoint      n
##   <chr>     <int>
## 1 DNA STRAND BREAKS    162
## 2 MNBNC FREQUENCY      23
```

This package allows the exploration and analysis of all nanomaterials in the dataset simultaneously. Since the functions in this package identify different nanomaterials according to the names of the nanomaterials, control observations named as `control`, `Control`, `medium`, `medium + BSA` and `untreated` in `geninvitro` must be linked to the non-control observations belonging to the same experiment. This can be done by first creating a separate dataset (`controldata`) for these control observations using function `SubsetData` by specifying "name" in `x` and the names of the controls in `x.cat`. Another dataset, `invitrodata`, is created by excluding these control observations with different names (by adding `include=FALSE`). If control observations are not named differently (as appear in the `geninvitro`), this process can be ignored.

Example:

```
controldata<-SubsetData(data=gendata, x="name", x.cat=c("control", "Control",
"medium", "medium + BSA", "untreated"))
invitrodata<-SubsetData(data=gendata, x="name", x.cat=c("control", "Control",
"medium", "medium + BSA", "untreated"), include=FALSE)
```

Since the data consists of several endpoints and units of the dose, certain toxicity endpoint and unit of measurement can be selected. For a specific endpoint "DNA STRAND BREAKS" and a specific concentration unit "ug/cm²", scatter plots of the dose-response relationship can be generated as follows:

```
nmplot(data.nm=invitrodata, data.control=controldata, id="experimentID", nano="name",
response="value", dose="concentration", end="endpoint", end.cat="DNA STRAND BREAKS",
unit="concentration_unit", unit.cat="ug/cm2", type="dose", control.opt="same",
ncol = 3, nrow = 2)
```

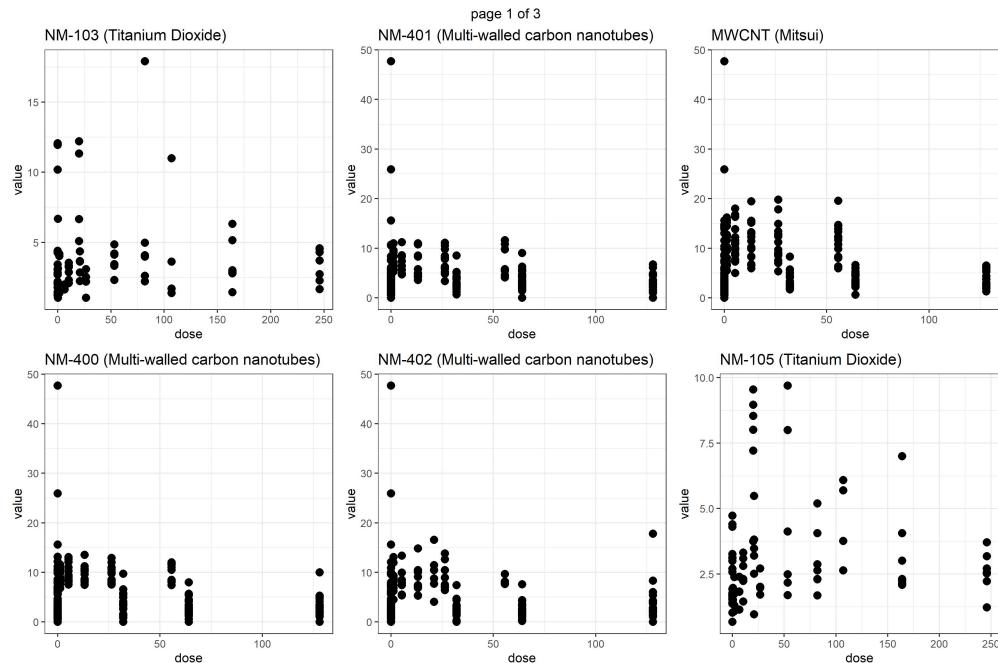


Figure 1: Dose-response plot (1-6)

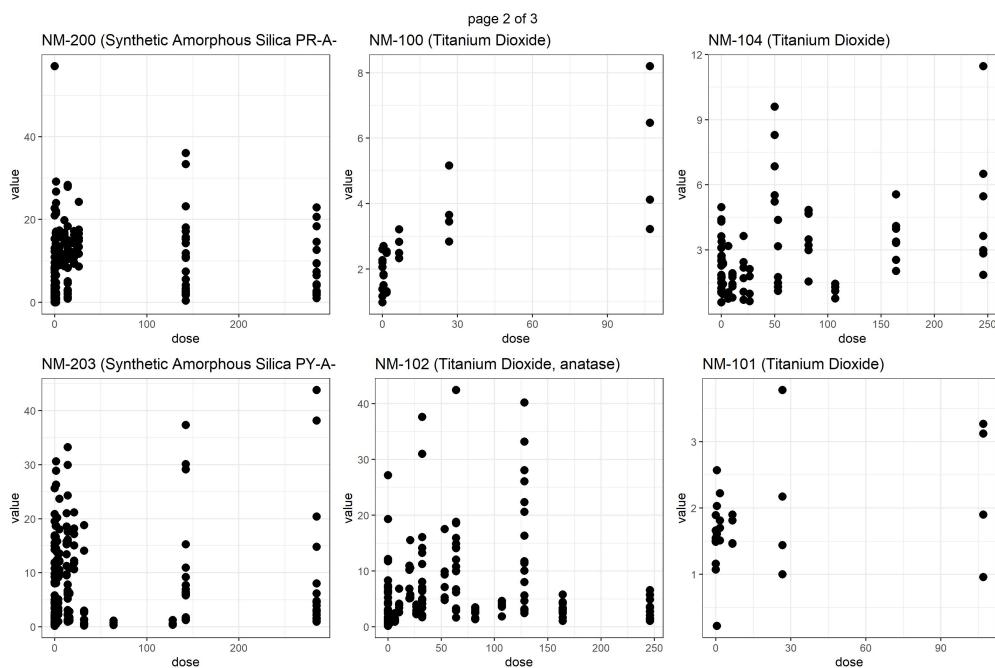


Figure 2: Dose-response plot (7-12)

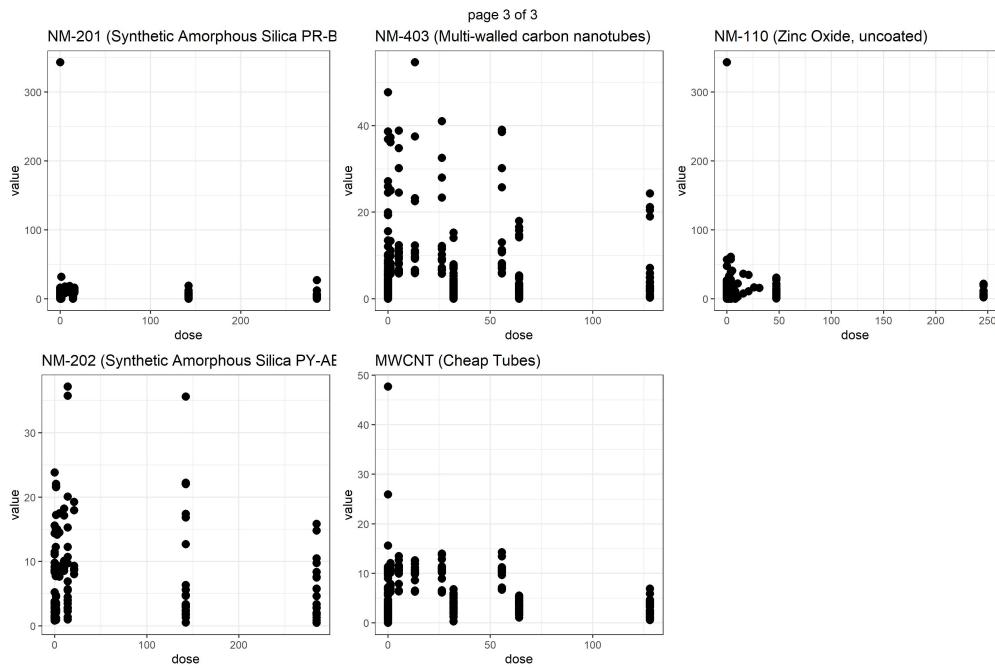


Figure 3: Dose-response plot (13-17)

To select a specific endpoint, such as DNA strand breaks, `end="endpoint"` and `end.cat="DNA STRAND BREAKS"` can be added, while a particular unit of concentration, $\mu\text{g}/\text{cm}^2$, can be specified through `unit="concentration_unit"` and `unit.cat="ug/cm2"`. The experiment identifier, `experimentID` is used to identify observations that belong to the same experiment. In the `control.opt`, by specifying `same`, only control doses in $\mu\text{g}/\text{cm}^2$ are included.

Since a data can contain observations in different units of measurement, it is possible to include controls that are measured in a different unit. Suppose if we are interested in unit ug/cm², to see how many controls are available if all units of measurement of control are considered and if only unit ug/cm² is included, function FrCtrl can be used:

```
control.list <- FrCtrl(data.nm=invitrodata, data.control=controldata, nano="name",
end="endpoint", end.cat="DNA STRAND BREAKS", id="experimentID", dose="concentration",
unit="concentration_unit", unit.cat="ug/cm2")

head(control.list)

##                                     Nanomaterial Freq(Dose=0.0).same Freq(obs).same
## 1           MWCNT (Cheap Tubes)          86            243
## 2           MWCNT (Mitsui)              90            247
## 3       NM-100 (Titanium Dioxide)        8             28
## 4       NM-101 (Titanium Dioxide)        8             28
## 5 NM-102 (Titanium Dioxide, anatase)    42            169
## 6       NM-103 (Titanium Dioxide)        20            83
##   Freq(Dose=0.0).all Freq(obs).all
## 1           86            243
## 2           90            247
## 3           8             28
## 4           8             28
## 5           42            169
## 6           20            83
```

The output table consists of column `Freq(Dose=0.0).same` , `Freq(obs).same`, `Freq(Dose=0.0).all` and `Freq(obs).all`, which contains the frequency of controls in ug/cm²; the number of observations with dose measured in ug/cm²; the frequency of control values with dose measured in ug/cm² and in other units of measurement; and the number of observations with dose measured in ug/cm² and control dose measured in any units of measurement, respectively. It seems that in this case there are no additional controls with different units that can be added to any of the nanomaterials.

It is also possible to explore the dose-response relationship through scatter plot according to another variable, for example, cell types, by adding `x.cat="celltype"`:

```
nmpplot.cat(data.nm=invitrodata, data.control=controldata, id="experimentID", nano="name",
response="value", dose="concentration", end="endpoint", end.cat="DNA STRAND BREAKS",
unit="concentration_unit", unit.cat="ug/cm2", x.cat="celltype", type="dose",
control.opt="same", nrow=1, ncol=1)
```

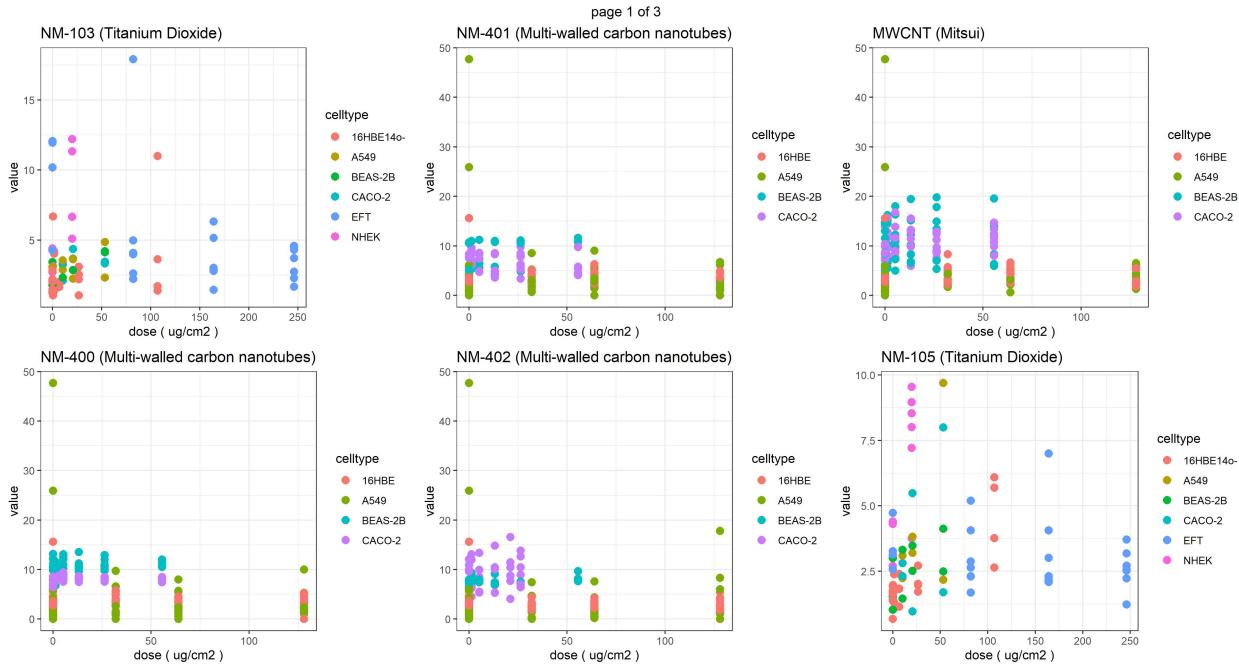


Figure 4: Dose-response plot by the cell type (plot 1-6)

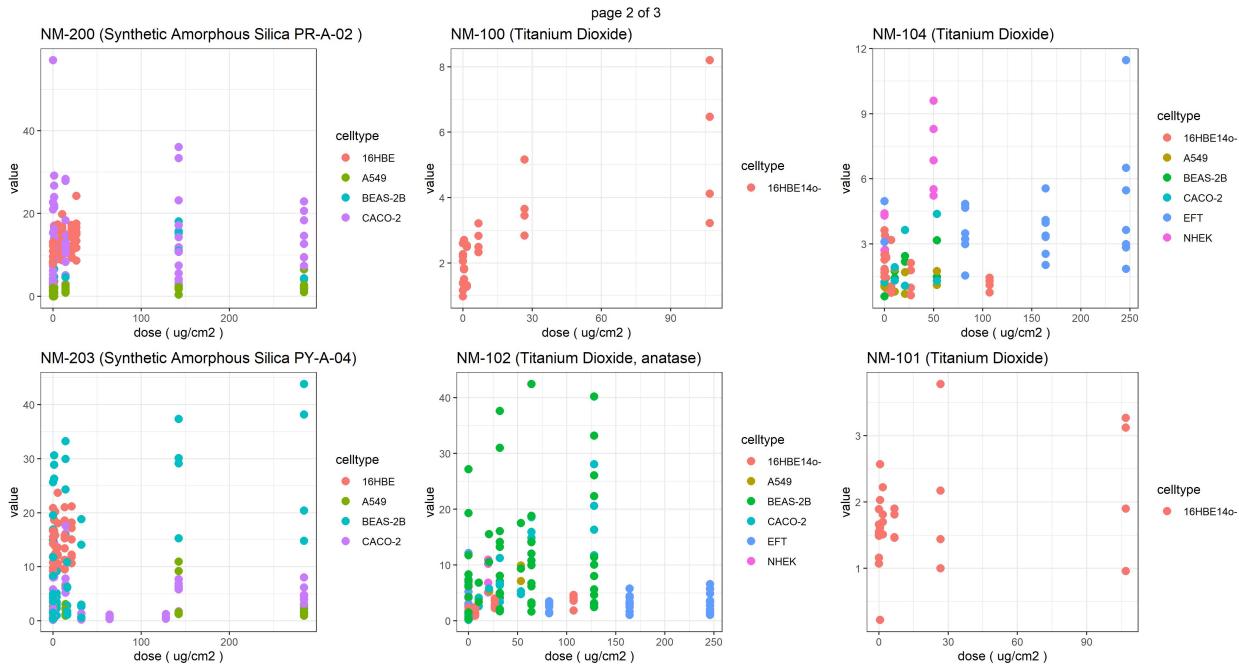


Figure 5: Dose-response plot by the cell type (plot 7-12)

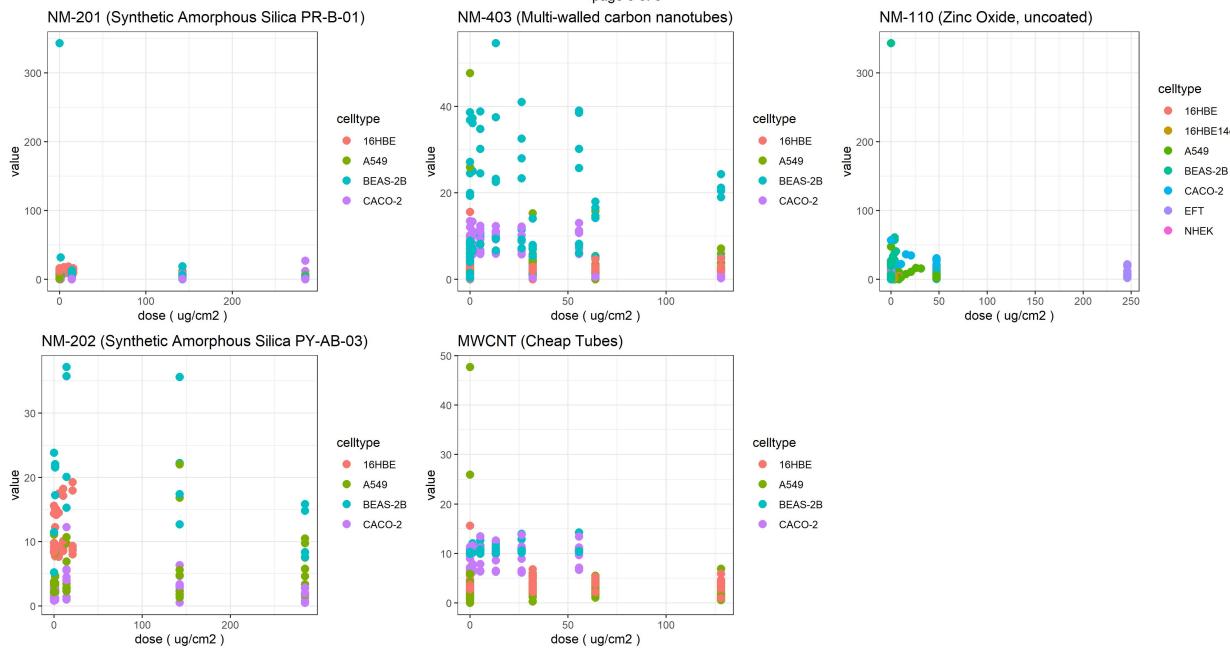


Figure 6: Dose-response plot by the cell type (plot 13-17)

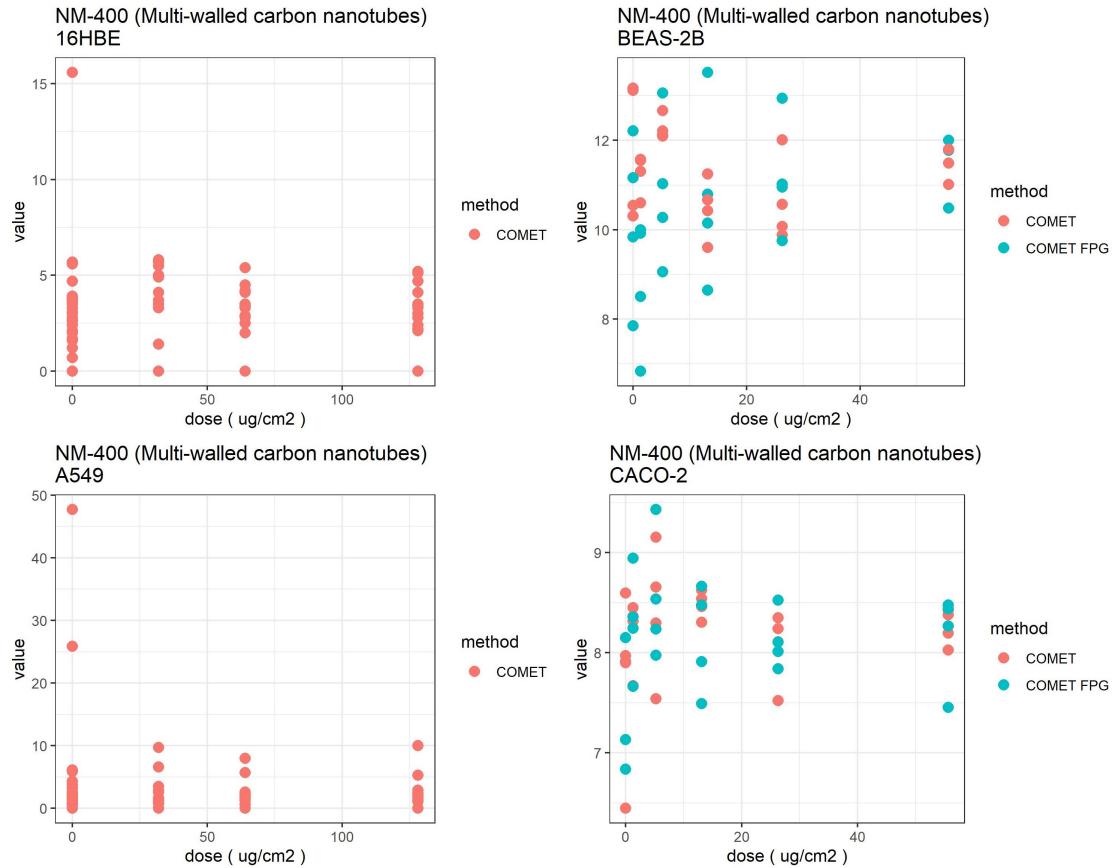


Figure 7: Dose-response plot for each cell type by the method (NM-400 (Multi-walled carbon nanotubes))

Figure 4, 5 and 6 show the scatter plots of the nanomaterials, with different colours represent different cell types. Other variables that may have an effect on the dose-response relationship can be explored in the same way. Scatter plots can also be generated for each cell type, by the method used in the experiment using function `nmpplot.ncat` as follows:

```
nmpplot.ncat(data.nm=invitrodata, data.control=controldata, id="experimentID",
  nano="name", response="value", dose="concentration", end="endpoint",
  end.cat="DNA STRAND BREAKS", unit="concentration_unit", unit.cat="ug/cm2",
  cat="celltype", x.cat="method", type="dose", control.opt="same", nrow=2, ncol=3)
```

Figure 7 shows the scatter plots of the dose-response relationship of NM-400 (Multi-walled carbon nanotubes), for each cell types (`cat="celltype"`) and differentiated by the method used in the experiment (`x.cat="method"`).

Isotonic means can also be plotted together with the data points using function `Isoplot` as follows:

```
Isoplot(data.nm=invitrodata, data.control=controldata, id="experimentID", nano="name",
  response="value", dose="concentration", end="endpoint", end.cat="DNA STRAND BREAKS",
  unit="concentration_unit", unit.cat="ug/cm2", dose.type="dose", control.opt="same",
  ncol=3, nrow=3)
```

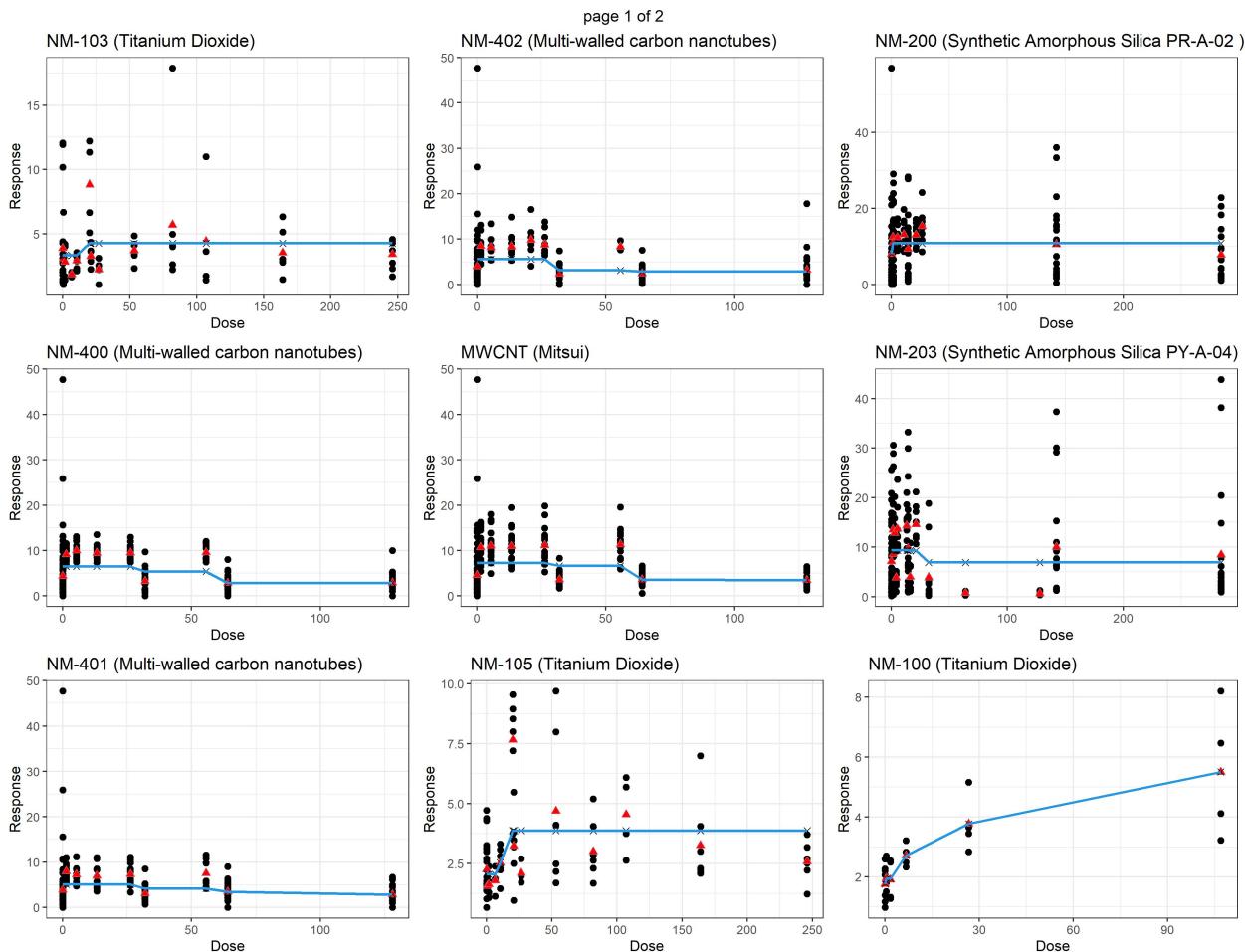


Figure 8: Isotonic means plot (plot 1-9)

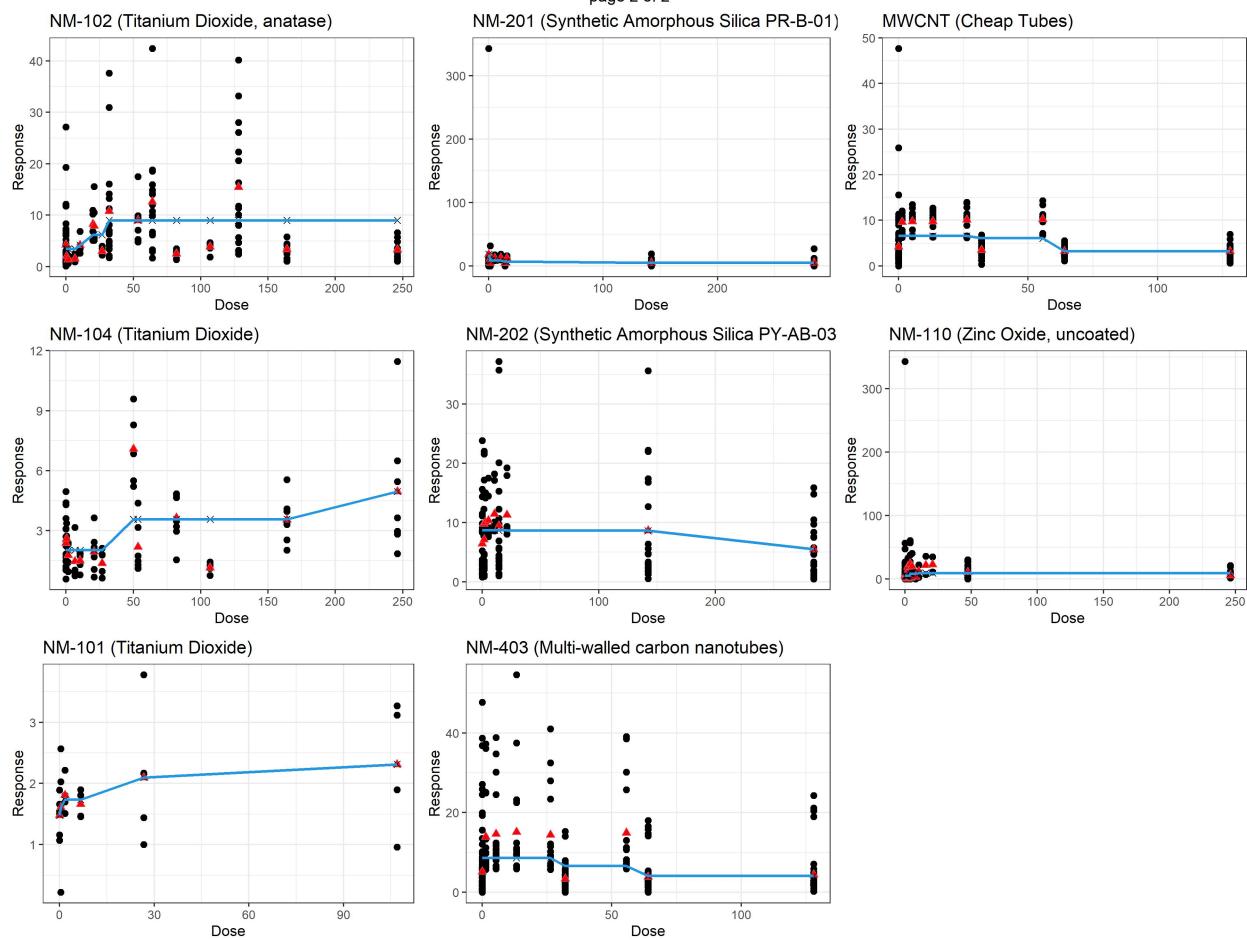


Figure 9: Isotonic means plot (plot 10-17)

Figure 8 and 9 show the plots of the isotonic means (\times), sample means (\circ), data points (\bullet) and isotonic regression curve (blue line). Suppose if we want to generate the isotonic means plots for each cell type, `vars="celltype"` can be added in the function `Isoplot`.

Analysis of the monotonic trend in the dose-response relationship can be performed using function `Isotest`. The test statistics used in the analysis can be specified in `stat` and the method used to adjust for the multiplicity can be specified in `method`. The p-values are calculated based on permutation with `niter` number of permutations.

```
test1<-Isotest(data.nm=invitrodata, data.control=controldata, id="experimentID",
nano="name", response="value", dose="concentration", end="endpoint",
end.cat="DNA STRAND BREAKS", unit="concentration_unit", unit.cat="ug/cm2",
stat="E2", niter=1000, method="BH", control.opt="same", set.seed=1234)

head(test1)

##                                     tstat.up: E2 tstat.dn: E2 pvalue.up
## NM-103 (Titanium Dioxide)      0.026      0.003      0.357
## NM-400 (Multi-walled carbon nanotubes) 0.032      0.091      0.019
## NM-401 (Multi-walled carbon nanotubes) 0.013      0.036      0.168
## NM-402 (Multi-walled carbon nanotubes) 0.009      0.074      0.283
```

```

## MWCNT (Mitsui)          0.059      0.074      0.002
## NM-105 (Titanium Dioxide) 0.191      0.005      0.002
##                                     pvalue.dn adj.p.up: BH adj.p.bn: BH
## NM-103 (Titanium Dioxide)    0.735      0.405      0.892
## NM-400 (Multi-walled carbon nanotubes) 0       0.046      0
## NM-401 (Multi-walled carbon nanotubes) 0.004      0.26       0.011
## NM-402 (Multi-walled carbon nanotubes) 0       0.344      0
## MWCNT (Mitsui)            0       0.007      0
## NM-105 (Titanium Dioxide) 0.683      0.007      0.892
##                                     direction
## NM-103 (Titanium Dioxide) "u"
## NM-400 (Multi-walled carbon nanotubes) "d"
## NM-401 (Multi-walled carbon nanotubes) "d"
## NM-402 (Multi-walled carbon nanotubes) "d"
## MWCNT (Mitsui)           "d"
## NM-105 (Titanium Dioxide) "u"

```

The result show the value of the test statistics (`tstat.up`: E2), the p-values (`pvalue.up`) and the adjusted p-values (`adj.p.up`: BH) with increasing alternative; the test statistics (`tstat.bn`: E2), the p-values (`pvalue.bn`) and the adjusted p-values (`adj.p.bn`: BH) with decreasing alternative; and the more likely direction of the monotonic trend (according to the likelihood).

Suppose if we want to perform the test for each cell type, `vars=celltype` can be added in the function `Isotest`. By adding `vars`, the data for each nanomaterial will be split according to the cell type prior to the testing.

Adjusted p-values and the unadjusted p-values (for both directions of the trend) can be plotted together as follows:

```

adjPlot(data.nm=invitrodata, data.control=controldata, id="experimentID", nano="name",
response="value", dose="concentration", end="endpoint", end.cat="DNA STRAND BREAKS",
unit="concentration_unit", unit.cat="ug/cm2", stat="E2", niter=1000, method="BH",
control.opt="same", set.seed = 1234, FDR=0.05)

```

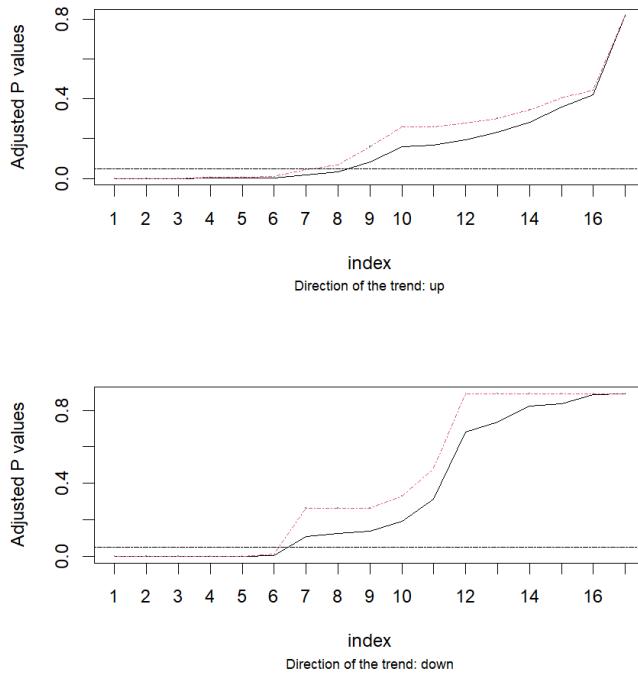


Figure 10: Plot of adjusted p-values

Figure 10 shows the plot of the raw p-values (black line) and the adjusted p-values (red dashed line).

Case study: NM-400 (Multi-walled carbon nanotubes nanomaterial)

Another example is nm400 dataset, which contains the data of NM-400 (Multi-walled carbon nanotubes nanomaterial) study (and control values) as follows:

```
head(nm400)

## # A tibble: 6 x 17
##   name    publicname supplier experimentID method studyprovider endpoint value
##   <chr>   <chr>     <chr>      <chr>     <chr>      <chr>     <dbl>
## 1 control CONTROL NanoGeno~ 9354ddc4-cf6~ COMET    NRCWE     DNA STR~  5.6
## 2 control CONTROL NanoGeno~ 9354ddc4-cf6~ COMET    NRCWE     DNA STR~  2.4
## 3 control CONTROL NanoGeno~ 9354ddc4-cf6~ COMET    NRCWE     DNA STR~  2.8
## 4 control CONTROL NanoGeno~ cbf58b83-d0d~ COMET    NRCWE     DNA STR~  3.8
## 5 control CONTROL NanoGeno~ cbf58b83-d0d~ COMET    NRCWE     DNA STR~  3
## 6 control CONTROL NanoGeno~ cbf58b83-d0d~ COMET    NRCWE     DNA STR~  2.6
## # ... with 9 more variables: unit <chr>, celltype <chr>, treatment <chr>,
## #   exptimeunit <chr>, exptime <chr>, concentration_unit <chr>,
## #   concentration <chr>, concentration_ml_unit <chr>, concentration_ml <dbl>
```

```
Frtab(data=nm400, x="endpoint", cat="name", opt="ls.obs")

## # A tibble: 2 x 3
##   name               endpoint       n
##   <chr>              <chr>     <dbl>
## 1 control            NRCWE     12
## 2 NM-400 (Multi-w~  NM-400     12
```

```

## <chr>                               <chr>          <int>
## 1 control                           DNA STRAND BREAKS  93
## 2 NM-400 (Multi-walled carbon nanotubes) DNA STRAND BREAKS 160

```

with concentration unit

```

Frtab(data=nm400, x="concentration_unit", opt="list")

## # A tibble: 1 x 1
##   concentration_unit
##   <chr>
##   1 ug/cm2

```

Since some control observations in this dataset are named as `control`, we separate the control observations as follows:

```

controldata2<-SubsetData(data=nm400, x="name", x.cat="control")

invitrodata2<-SubsetData(data=nm400, x="name", x.cat="control", include=FALSE)

```

Scatter plot of the dose and response for data of a specific nanomaterial such as `nm400` can be generated as follows:

```

nmplot.cat(data.nm=invitrodata2, data.control=controldata2, id="experimentID",
nano="name", response="value", dose="concentration", end="endpoint",
end.cat="DNA STRAND BREAKS", unit="concentration_unit", unit.cat="ug/cm2",
type="dose", control.opt="same", nrow=1, ncol=1, x.cat="celltype")

```

and the isotonic means can be plotted together with the data points using function `Isoplot` as follows:

```

Isoplot(data.nm=invitrodata2, data.control=controldata2, id="experimentID",
nano="name", response="value", dose="concentration", end="endpoint",
end.cat="DNA STRAND BREAKS", unit="concentration_unit", unit.cat="ug/cm2",
dose.type="dose", control.opt="same")

```

Figure 11 shows the dose-response plot of NM-400 (left) and with the isotonic means (right).

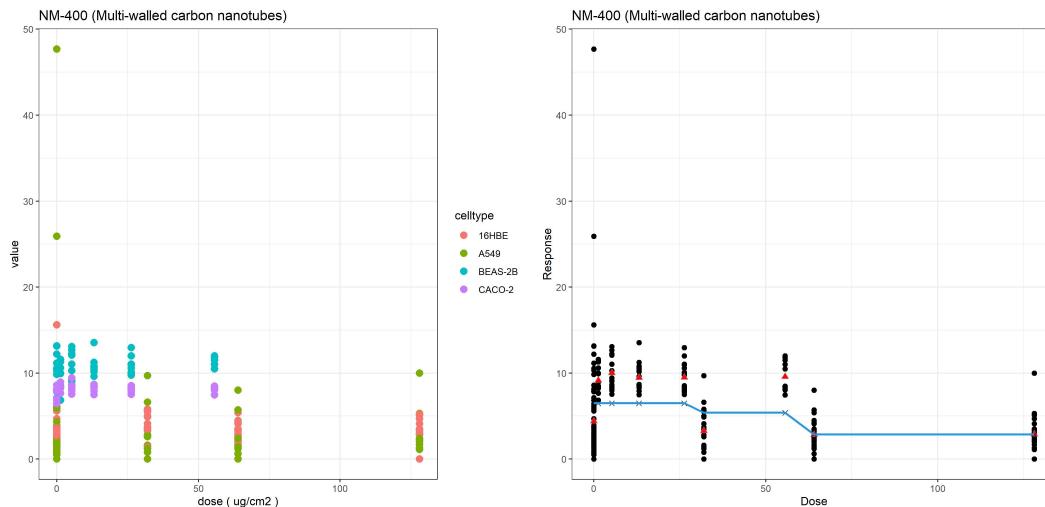


Figure 11: Dose-response plot of NM-400

A log(dose) plot can also be generated, by specifying `type=log` in the function `nmplot.cat`:

```
nmplot.cat(data.nm=invitrodata2, data.control=controldata2, id="experimentID",
nano="name", response="value", dose="concentration", end="endpoint",
end.cat="DNA STRAND BREAKS", unit="concentration_unit", unit.cat="ug/cm2",
type="log", control.opt="same", nrow=1, ncol=1, x.cat = "celltype")
```

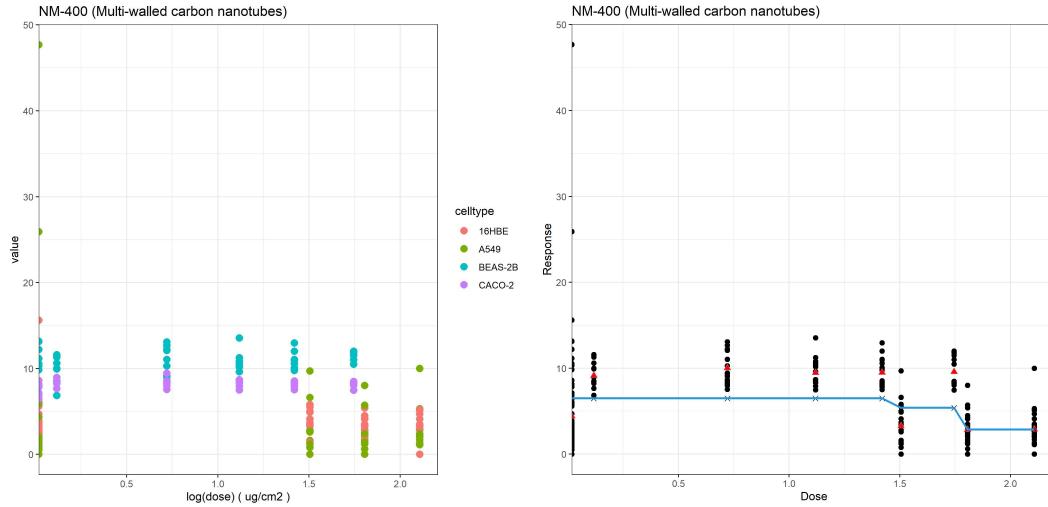


Figure 12: Dose-response plot of NM-400 (log(dose))

The significance of the monotonic trend can be investigated using `Isotest` as follows:

```
Isotest(data.nm=invitrodata2, data.control=controldata2, id="experimentID", nano="name",
response="value", dose="concentration", end="endpoint", end.cat="DNA STRAND BREAKS",
unit="concentration_unit", unit.cat="ug/cm2", stat="E2", niter=1000, control.opt="same",
set.seed=1234)

##                                     tstat.up: E2 tstat.dn: E2 pvalue.up
## NM-400 (Multi-walled carbon nanotubes) 0.032      0.091      0.019
##                                         pvalue.dn direction
## NM-400 (Multi-walled carbon nanotubes) 0           "d"
```

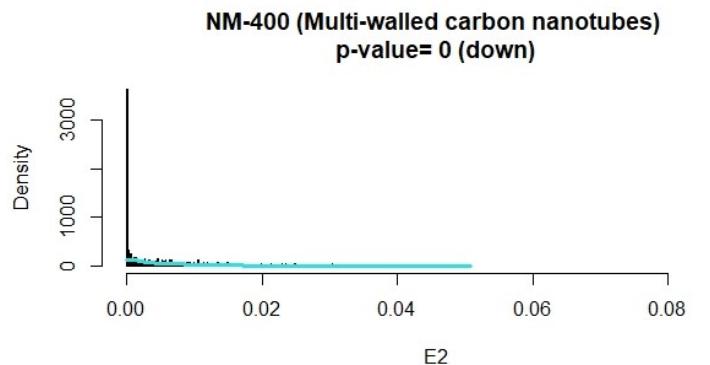


Figure 13: Plot of the null distribution and the observed test statistics

Suppose if we want to use a different test statistics, such as Williams test, `stat="Williams"` can be specified.

```
Isotest(data.nm=invitrodata2, data.control=controldata2, id="experimentID", nano="name",
response="value", dose="concentration", end="endpoint", end.cat="DNA STRAND BREAKS",
unit="concentration_unit", unit.cat="ug/cm2", stat="Williams", niter=1000,
control.opt="same", set.seed=1234)

##                                     tstat.up: Williams tstat.dn: Williams
## NM-400 (Multi-walled carbon nanotubes) 2.049           -1.759
##                                     pvalue.up pvalue.dn direction
## NM-400 (Multi-walled carbon nanotubes) 0.04      0.034      "d"
```