Validation of the tool

Etienne Rolland

20/04/2020

Contents

Checking the corpus of validation

```
library(segmenteR)
library(udpipe)
```

```
setwd("~/Bureau")
validation_set<-list.files("Validation Set/")
```

```
articles<-c("Ansari, M 2018.pdf",
"Bai, K J et al 2018.pdf",
"Blum, J L et al 2014.pdf",
"Chen et al 2008.pdf",
"Jamshidzadeh, A et al 2015.pdf",
"Li, C et al 2009.pdf",
"Li, J et al 2013.pdf",
"Li, J G et al 2009.pdf",
"Li, Y et al 2010.pdf",
"Liu, Y et al 2014.pdf",
"Love, S A et al 2012.pdf",
"Maurizi, L et al 2015.pdf",
"Nešković, O et al 2013.pdf",
"Neun, B W et al 2018.pdf",
"Reddy, A et al 2010.pdf",
"Richter, AW et al 1983.pdf",
"Sadaf, A et al 2012.pdf",
"Tang, J et al 2009.pdf",
"Vicente, S et al 2017.pdf",
"Wang, X et al 2010.pdf",
"Xiaoli, F et al 2017.pdf",
"Yan, M et al 2011.pdf",
"Yu, T et al 2011.pdf",
```

```
"Zook, J M et al 2011.pdf",
"Abbasalipourkabir R et al 2015.pdf",
"Abe, S et al 2012.pdf",
"Adedara, I A et al 2018.pdf",
"Aijie, C et al 2017.pdf",
"Amrollahi-Sharifabadi et al. 2018.pdf",
"Aragao-Santiago, L 2016.pdf",
"AshaRani et al. 2009.pdf",
"Bendre, V et al 2011.pdf",
"Cabral, H et al 2005.pdf")
```

```
table(articles %in% validation_set)
```

```
##
## FALSE  TRUE
##     3    30
```

```
table(validation_set %in% articles)
```

```
##
## FALSE  TRUE
##    18    30
```

```
articles[which(articles %in% validation_set)]
```

```
##  [1] "Ansari, M 2018.pdf"                 "Bai, K J et al 2018.pdf"
##  [3] "Blum, J L et al 2014.pdf"           "Chen et al 2008.pdf"
##  [5] "Jamshidzadeh, A et al 2015.pdf"     "Li, C et al 2009.pdf"
##  [7] "Li, J et al 2013.pdf"               "Li, J G et al 2009.pdf"
##  [9] "Li, Y et al 2010.pdf"               "Liu, Y et al 2014.pdf"
## [11] "Love, S A et al 2012.pdf"           "Maurizi, L et al 2015.pdf"
## [13] "Nešković, O et al 2013.pdf"         "Neun, B W et al 2018.pdf"
## [15] "Reddy, A et al 2010.pdf"            "Richter, AW et al 1983.pdf"
## [17] "Sadaf, A et al 2012.pdf"            "Tang, J et al 2009.pdf"
## [19] "Vicente, S et al 2017.pdf"          "Wang, X et al 2010.pdf"
## [21] "Xiaoli, F et al 2017.pdf"           "Yan, M et al 2011.pdf"
## [23] "Yu, T et al 2011.pdf"               "Zook, J M et al 2011.pdf"
## [25] "Abbasalipourkabir R et al 2015.pdf" "Abe, S et al 2012.pdf"
## [27] "Adedara, I A et al 2018.pdf"        "Aijie, C et al 2017.pdf"
## [29] "AshaRani et al. 2009.pdf"           "Bendre, V et al 2011.pdf"
```

```
articles[which(!articles %in% validation_set)]
```

```
## [1] "Amrollahi-Sharifabadi et al. 2018.pdf"
## [2] "Aragao-Santiago, L 2016.pdf"
## [3] "Cabral, H et al 2005.pdf"
```

The three missing articles, [1] “Amrollahi-Sharifabadi et al. 2018.pdf” “Aragao-Santiago, L 2016.pdf” “Cabral, H et al 2005.pdf”
Has been downloaded in the 20 April of 2021. They were absent from the repository in pdf.

This may explain why the double box plot only contains 28 points, where the manual ranking contain 33 articles. 3 were missing.

```
list.files("Missing_validation_set/")
```

```
## [1] "Amrollahi-Sharifabadi et al. 2018.pdf"
## [2] "Aragao-Santiago, L 2016.pdf"
```

```
## [3] "Cabral, H et al 2005.pdf"
```

```
validation_set<-list.files("Validation_set_combined//")
articles[which(!articles %in% validation_set)]
```

```
## character(0)
```

```
table(articles %in% validation_set)
```

```
##
## TRUE
##   33
```

Third folder to managed more easily the manual reading of the article in pdf

```
articles_to_clean <- list.files("Validation_tool_folder/")
for (article in articles_to_clean) {
  print(article)
  file.remove(paste0("Validation_tool_folder/", article))
}
```

```
## [1] "Abbasalipourkabir R et al 2015.pdf"
## [1] "Abbasalipourkabir R et al 2015.pdf.rds"
## [1] "Abe, S et al 2012.pdf"
## [1] "Abe, S et al 2012.pdf.rds"
## [1] "Adedara, I A et al 2018.pdf"
## [1] "Adedara, I A et al 2018.pdf.rds"
## [1] "Aijie, C et al 2017.pdf"
## [1] "Aijie, C et al 2017.pdf.rds"
## [1] "Amrollahi-Sharifabadi et al. 2018.pdf"
## [1] "Amrollahi-Sharifabadi et al. 2018.pdf.rds"
## [1] "Ansari, M 2018.pdf"
## [1] "Ansari, M 2018.pdf.rds"
## [1] "Aragao-Santiago, L 2016.pdf"
## [1] "Aragao-Santiago, L 2016.pdf.rds"
## [1] "AshaRani et al. 2009.pdf"
## [1] "AshaRani et al. 2009.pdf.rds"
## [1] "Bai, K J et al 2018.pdf"
## [1] "Bai, K J et al 2018.pdf.rds"
## [1] "Bendre, V et al 2011.pdf"
## [1] "Bendre, V et al 2011.pdf.rds"
## [1] "Blum, J L et al 2014.pdf"
## [1] "Blum, J L et al 2014.pdf.rds"
## [1] "Cabral, H et al 2005.pdf"
## [1] "Cabral, H et al 2005.pdf.rds"
## [1] "Chen et al 2008.pdf"
## [1] "Chen et al 2008.pdf.rds"
## [1] "Jamshidzadeh, A et al 2015.pdf"
## [1] "Jamshidzadeh, A et al 2015.pdf.rds"
## [1] "Li, C et al 2009.pdf"
## [1] "Li, C et al 2009.pdf.rds"
## [1] "Li, J et al 2013.pdf"
## [1] "Li, J et al 2013.pdf.rds"
## [1] "Li, J G et al 2009.pdf"
## [1] "Li, J G et al 2009.pdf.rds"
```

```
## [1] "Li, Y et al 2010.pdf"
## [1] "Li, Y et al 2010.pdf.rds"
## [1] "Liu, Y et al 2014.pdf"
## [1] "Liu, Y et al 2014.pdf.rds"
## [1] "Love, S A et al 2012.pdf"
## [1] "Love, S A et al 2012.pdf.rds"
## [1] "Maurizi, L et al 2015.pdf"
## [1] "Maurizi, L et al 2015.pdf.rds"
## [1] "Nešković, O et al 2013.pdf"
## [1] "Nešković, O et al 2013.pdf.rds"
## [1] "Neun, B W et al 2018.pdf"
## [1] "Neun, B W et al 2018.pdf.rds"
## [1] "Reddy, A et al 2010.pdf"
## [1] "Reddy, A et al 2010.pdf.rds"
## [1] "Richter, AW et al 1983.pdf"
## [1] "Richter, AW et al 1983.pdf.rds"
## [1] "Sadaf, A et al 2012.pdf"
## [1] "Sadaf, A et al 2012.pdf.rds"
## [1] "segmenteR_Abbasalipourkabir R et al 2015.pdf.rds"
## [1] "segmenteR_Abe, S et al 2012.pdf.rds"
## [1] "segmenteR_Adedara, I A et al 2018.pdf.rds"
## [1] "segmenteR_Aijie, C et al 2017.pdf.rds"
## [1] "segmenteR_Amrollahi-Sharifabadi et al. 2018.pdf.rds"
## [1] "segmenteR_Ansari, M 2018.pdf.rds"
## [1] "segmenteR_Aragao-Santiago, L 2016.pdf.rds"
## [1] "segmenteR_AshaRani et al. 2009.pdf.rds"
## [1] "segmenteR_Bai, K J et al 2018.pdf.rds"
## [1] "segmenteR_Bendre, V et al 2011.pdf.rds"
## [1] "segmenteR_Blum, J L et al 2014.pdf.rds"
## [1] "segmenteR_Cabral, H et al 2005.pdf.rds"
## [1] "segmenteR_Chen et al 2008.pdf.rds"
## [1] "segmenteR_Jamshidzadeh, A et al 2015.pdf.rds"
## [1] "segmenteR_Li, C et al 2009.pdf.rds"
## [1] "segmenteR_Li, J et al 2013.pdf.rds"
## [1] "segmenteR_Li, J G et al 2009.pdf.rds"
## [1] "segmenteR_Li, Y et al 2010.pdf.rds"
## [1] "segmenteR_Liu, Y et al 2014.pdf.rds"
## [1] "segmenteR_Love, S A et al 2012.pdf.rds"
## [1] "segmenteR_Maurizi, L et al 2015.pdf.rds"
## [1] "segmenteR_Nešković, O et al 2013.pdf.rds"
## [1] "segmenteR_Neun, B W et al 2018.pdf.rds"
## [1] "segmenteR_Reddy, A et al 2010.pdf.rds"
## [1] "segmenteR_Richter, AW et al 1983.pdf.rds"
## [1] "segmenteR_Sadaf, A et al 2012.pdf.rds"
## [1] "segmenteR_Tang, J et al 2009.pdf.rds"
## [1] "segmenteR_Vicente, S et al 2017.pdf.rds"
## [1] "segmenteR_Wang, X et al 2010.pdf.rds"
## [1] "segmenteR_Xiaoli, F et al 2017.pdf.rds"
## [1] "segmenteR_Yan, M et al 2011.pdf.rds"
## [1] "segmenteR_Yu, T et al 2011.pdf.rds"
## [1] "segmenteR_Zook, J M et al 2011.pdf.rds"
## [1] "Tang, J et al 2009.pdf"
## [1] "Tang, J et al 2009.pdf.rds"
## [1] "Vicente, S et al 2017.pdf"
```

```
## [1] "Vicente, S et al 2017.pdf.rds"
## [1] "Wang, X et al 2010.pdf"
## [1] "Wang, X et al 2010.pdf.rds"
## [1] "Xiaoli, F et al 2017.pdf"
## [1] "Xiaoli, F et al 2017.pdf.rds"
## [1] "Yan, M et al 2011.pdf"
## [1] "Yan, M et al 2011.pdf.rds"
## [1] "Yu, T et al 2011.pdf"
## [1] "Yu, T et al 2011.pdf.rds"
## [1] "Zook, J M et al 2011.pdf"
## [1] "Zook, J M et al 2011.pdf.rds"
```

```
print("After cleaning :")
```

```
## [1] "After cleaning :"
```

```
list.files("Validation_tool_folder/")
```

```
## character(0)
```

```
#Validation_tool_folder
length(articles)
```

```
## [1] 33
```

```
for (article in articles) {
  print(article)
  file.copy(paste0("Validation_set_combined/", article), paste0("Validation_tool_folder/", article))
}
```

```
## [1] "Ansari, M 2018.pdf"
## [1] "Bai, K J et al 2018.pdf"
## [1] "Blum, J L et al 2014.pdf"
## [1] "Chen et al 2008.pdf"
## [1] "Jamshidzadeh, A et al 2015.pdf"
## [1] "Li, C et al 2009.pdf"
## [1] "Li, J et al 2013.pdf"
## [1] "Li, J G et al 2009.pdf"
## [1] "Li, Y et al 2010.pdf"
## [1] "Liu, Y et al 2014.pdf"
## [1] "Love, S A et al 2012.pdf"
## [1] "Maurizi, L et al 2015.pdf"
## [1] "Nešković, O et al 2013.pdf"
## [1] "Neun, B W et al 2018.pdf"
## [1] "Reddy, A et al 2010.pdf"
## [1] "Richter, AW et al 1983.pdf"
## [1] "Sadaf, A et al 2012.pdf"
## [1] "Tang, J et al 2009.pdf"
## [1] "Vicente, S et al 2017.pdf"
## [1] "Wang, X et al 2010.pdf"
## [1] "Xiaoli, F et al 2017.pdf"
## [1] "Yan, M et al 2011.pdf"
## [1] "Yu, T et al 2011.pdf"
## [1] "Zook, J M et al 2011.pdf"
## [1] "Abbasalipourkabir R et al 2015.pdf"
## [1] "Abe, S et al 2012.pdf"
## [1] "Adedara, I A et al 2018.pdf"
```

```
## [1] "Aijie, C et al 2017.pdf"
## [1] "Amrollahi-Sharifabadi et al. 2018.pdf"
## [1] "Aragao-Santiago, L 2016.pdf"
## [1] "AshaRani et al. 2009.pdf"
## [1] "Bendre, V et al 2011.pdf"
## [1] "Cabral, H et al 2005.pdf"
```

```
length(list.files("Validation_tool_folder/"))
```

```
## [1] 33
```

```
list.files("Validation_tool_folder/")
```

```
##  [1] "Abbasalipourkabir R et al 2015.pdf"
##  [2] "Abe, S et al 2012.pdf"
##  [3] "Adedara, I A et al 2018.pdf"
##  [4] "Aijie, C et al 2017.pdf"
##  [5] "Amrollahi-Sharifabadi et al. 2018.pdf"
##  [6] "Ansari, M 2018.pdf"
##  [7] "Aragao-Santiago, L 2016.pdf"
##  [8] "AshaRani et al. 2009.pdf"
##  [9] "Bai, K J et al 2018.pdf"
## [10] "Bendre, V et al 2011.pdf"
## [11] "Blum, J L et al 2014.pdf"
## [12] "Cabral, H et al 2005.pdf"
## [13] "Chen et al 2008.pdf"
## [14] "Jamshidzadeh, A et al 2015.pdf"
## [15] "Li, C et al 2009.pdf"
## [16] "Li, J et al 2013.pdf"
## [17] "Li, J G et al 2009.pdf"
## [18] "Li, Y et al 2010.pdf"
## [19] "Liu, Y et al 2014.pdf"
## [20] "Love, S A et al 2012.pdf"
## [21] "Maurizi, L et al 2015.pdf"
## [22] "Nešković, O et al 2013.pdf"
## [23] "Neun, B W et al 2018.pdf"
## [24] "Reddy, A et al 2010.pdf"
## [25] "Richter, AW et al 1983.pdf"
## [26] "Sadaf, A et al 2012.pdf"
## [27] "Tang, J et al 2009.pdf"
## [28] "Vicente, S et al 2017.pdf"
## [29] "Wang, X et al 2010.pdf"
## [30] "Xiaoli, F et al 2017.pdf"
## [31] "Yan, M et al 2011.pdf"
## [32] "Yu, T et al 2011.pdf"
## [33] "Zook, J M et al 2011.pdf"
```

Extraction of the materials and methods and results sections for comparison

Automatic extraction via segmenteR

```
## udpipe got the model for annotation
dl <- udpipe::udpipe_download_model("english-gum")
```

```
## Downloading udpipe model from https://raw.githubusercontent.com/jwijffels/udpipe.models.ud.2.5/maste
##  - This model has been trained on version 2.5 of data from https://universaldependencies.org
##  - The model is distributed under the CC-BY-SA-NC license: https://creativecommons.org/licenses/by-n
##  - Visit https://github.com/jwijffels/udpipe.models.ud.2.5 for model license details.
##  - For a list of all models and their licenses (most models you can download with this package have
## Downloading finished, model stored at '/home/erolland/Bureau/english-gum-ud-2.5-191206.udpipe'
```

```
str(dl)
```

```
## 'data.frame':    1 obs. of  5 variables:
##  $ language        : chr "english-gum"
##  $ file_model      : chr "/home/erolland/Bureau/english-gum-ud-2.5-191206.udpipe"
##  $ url             : chr "https://raw.githubusercontent.com/jwijffels/udpipe.models.ud.2.5/master/in
##  $ download_failed : logi FALSE
##  $ download_message: chr "OK"
```

```
model <- udpipe_load_model(file = dl$file_model)
model
```

```
## $file
## [1] "/home/erolland/Bureau/english-gum-ud-2.5-191206.udpipe"
##
## $model
## <pointer: 0x55c0f030f7c0>
##
## attr(,"class")
## [1] "udpipe_model"
```

```
library(stringr)
```

```
for (article in articles) {
  #do segmentation
  #from manual extraction :
  #path = paste0("Validation_tool_folder/", pdf_name)
  print(article)
  path = paste0("Validation_tool_folder/", article)
  section_aliases <- c("material", "method", "experimental", "experiment", "methodology")

  #model definition can be skipped, the function can download it automatically
  material_and_methods <- data.frame()
  material_and_methods <- try(segmenteR::extract_section_from_pdf(pdf_name=path,
                                                                  udpipe_model=model,
                                                                  section_aliases=section_aliases))
  saveRDS(material_and_methods, file=paste0("Validation_tool_folder/segmenteR_", article, ".rds"))

}
```

```
## [1] "Ansari, M 2018.pdf"
## [1] "Bai, K J et al 2018.pdf"
## [1] "Blum, J L et al 2014.pdf"
## [1] "Chen et al 2008.pdf"
## [1] "Jamshidzadeh, A et al 2015.pdf"
## [1] "Li, C et al 2009.pdf"
## [1] "Li, J et al 2013.pdf"
```

```
## ad_hoc_reorder(section_title_df) has been called
```

```
## [1] "Li, J G et al 2009.pdf"
## [1] "Li, Y et al 2010.pdf"
## [1] "Liu, Y et al 2014.pdf"
## [1] "Love, S A et al 2012.pdf"
## [1] "Maurizi, L et al 2015.pdf"
## [1] "Nešković, O et al 2013.pdf"
## [1] "Neun, B W et al 2018.pdf"
## [1] "Reddy, A et al 2010.pdf"
## [1] "Richter, AW et al 1983.pdf"
## Error in positions_sections_df$occurrences[[i + 1]] : indice hors limites
## [1] "Sadaf, A et al 2012.pdf"
## [1] "Tang, J et al 2009.pdf"
## [1] "Vicente, S et al 2017.pdf"
## [1] "Wang, X et al 2010.pdf"
## [1] "Xiaoli, F et al 2017.pdf"
## [1] "Yan, M et al 2011.pdf"
## [1] "Yu, T et al 2011.pdf"
## [1] "Zook, J M et al 2011.pdf"
```

```
## Warning in if (assumed_title_df$Font == font_section) {: la condition a une
## longueur > 1 et seul le premier élément est utilisé
```

```
## [1] "Abbasalipourkabir R et al 2015.pdf"
## [1] "Abe, S et al 2012.pdf"
## [1] "Adedara, I A et al 2018.pdf"
## [1] "Aijie, C et al 2017.pdf"
## [1] "Amrollahi-Sharifabadi et al. 2018.pdf"
```

```
## PDF error (6624): Mismatched EMC operator
```

```
## PDF error: Mismatched EMC operator
```

```
## [1] "Aragao-Santiago, L 2016.pdf"
## [1] "AshaRani et al. 2009.pdf"
## [1] "Bendre, V et al 2011.pdf"
```

```
## Clean_title_journal has been called
```

```
## Error in positions_sections_df$occurrences[[i + 1]] : indice hors limites
## [1] "Cabral, H et al 2005.pdf"
```

Manual extraction

Functions

```
pdf_name<-"Cabral, H et al 2005.pdf"
```

```
#function to open the article passed as input and preprocess
open_pdf_and_annotate <- function(pdf_name, model) {

  #open the pdf so the user can read it

  path = paste0("Validation_tool_folder/", pdf_name)
  fs::file_show(path)

  remove_bibliography <- TRUE
```

```
  txt_pdf <- tabulizer::extract_text(path) # read the text from the pdf
  txt_pdf <- segmenteR::preprocess_article_txt(txt_pdf)
  conllu_df <- segmenteR::annotate_txt_pdf(txt_pdf, udpipe_model=model ) # create the dataframe for NLP

  return(conllu_df)
}
```

```
conllu_df<-open_pdf_and_annotate(pdf_name, model)
```

```
head(conllu_df)
```

```
##   doc_id paragraph_id sentence_id
## 1   doc1            1           1
## 2   doc1            1           1
## 3   doc1            1           2
## 4   doc1            1           2
## 5   doc1            1           3
## 6   doc1            1           3
##
## 1
## 2
## 3
## 4
## 5 com/locate/jconrelJournal of Controlled ReleasPreparation and biological properties of dichloro(1,
## 6 com/locate/jconrelJournal of Controlled ReleasPreparation and biological properties of dichloro(1,
##   token_id                     token                      lemma  upos xpos
## 1        1                       www                        www  INTJ   UH
## 2        2                         .                          . PUNCT    .
## 3        1                  elsevier                   elsevier  NOUN   NN
## 4        2                         .                          . PUNCT    .
## 5        1 com/locate/jconrelJournal com/locate/jconrelJournal   ADJ   JJ
## 6        2                        of                         of   ADP   IN
##         feats head_token_id dep_rel deps          misc
## 1        <NA>             0    root <NA> SpaceAfter=No
## 2        <NA>             1   punct <NA>          <NA>
## 3 Number=Sing             0    root <NA> SpaceAfter=No
## 4        <NA>             1   punct <NA>          <NA>
## 5  Degree=Pos            19   nsubj <NA>          <NA>
## 6        <NA>             3    case <NA>          <NA>
```

```
word <- "Materials"

display_occurences <- function(word, conllu_df) {
  idx<-which(conllu_df$token==word)
  print(idx)
  conllu_df[idx,]$sentence
}

display_occurences(word, conllu_df)
```

```
## [1]   44 1504 1511
```

```
## [1] "aDepartment of Materials Science and Engineering, Graduate School of Engineering, The University
metal complex formation of DACHPt with poly(ethylene glycol)-poly(glutamic acid) block copolymer [PEG-
P(Glu)] in distilled water."
```

```
## [2] "2. Materials and methods 2."
## [3] "1. Materials h-Benzyl l-glutamate was bought from Sigma Chemical (St. Louis, MO)."
```

```
display_occurences("570", conllu_df)
```

```
## [1] 2628
```

```
## [1] "Cell viability was measured by the formed formazan absorbance at 570 nm."
```

```
cut_conllu_df <- function(idx1, idx2, conllu_df) {

  section <- conllu_df[idx1:idx2,]
  print(unique(section$sentence))
  return(section)

}
```

```
res<-cut_conllu_df(1504, 2628, conllu_df)
```

```
##  [1] "2. Materials and methods 2."
##  [2] "1. Materials h-Benzyl l-glutamate was bought from Sigma Chemical (St. Louis, MO)."
##  [3] "Bis(trichloromethyl)carbonate (triphosgene) was purchased from Tokyo Kasei Kogyo (Tokyo, Japan)
##  [4] "Dichloro(1, 2-diamminocyclohexane)platinum(II) (DACHPt) and AgNO 3 were purchased from Aldrich
##  [5] "a-Methoxy-N-aminopoly(ethylene glycol) (CH 3 O-PEG-NH 2; Mw=12, 000) was purchased from Nippon
##  [6] "Murine colon adenocarcinoma 26 (C-26) cells were kindly supplied from the National Cancer Cente
##  [7] "C-26 cells were maintained in RPMI 1640 medium (Sigma Chemical) containing 10% fetal bovine se
P(Glu)"
##  [8] "PEG-P(Glu) block copolymers were synthesized according to the previously described synthetic me
P(Asp) [ 9 ] with a slight modification."
##  [9] "Briefly, the N-carboxy anhydride of g-benzyl lglutamate was synthesized by the Fuchs-
Farthing method, using triphosgene."
## [10] "Then, N-carboxy anhydride of g-benzyl l-glutamate was polymerized in DMF, initiated by the prim
PEG-NH 2, to obtain PEG-poly(g-benzyl lglutamate) (PEG-PBLG) block copolymer."
## [11] "A narrow distribution (Mw/Mn: 1. 06) of PEG-PBLG was confirmed by gel permeation chromatography
## [12] "The polymerization degree of PBLG was determined to be 46 by comparing the proton ratios of me
## [13] "3 ppm) in 1 H-NMR measurement [JEOL EX 270 (JEOL, Tokyo, Japan) solvent:"
## [14] "DMSO-d 6; temperature: 80 8 C]."
## [15] "The deprotection of the benzyl group of PEG-PBLG was carried out by mixing with 0. 5 N NaOH at
P(Glu)."
## [16] "2. 3."
## [17] "Preparation of DACHPt-loaded micelles DACHPt (5 mM) was suspended in distilled water and mixed
## [18] "The solution was kept in dark at 25 8 C for 24 h. AgCl precipitates were found after reaction."
## [19] "Next, the mixture was centrifuged at 3000 rpm for 10 min to eliminate the AgCl precipitates."
## [20] "Afterward, the supernatant was purified by passage through a 0. 22-Am filter."
## [21] "PEG-P(Glu) was added to the solution of DACHPt aqueous complex (polymer concentration: 2. 1 mg/
## [22] "The prepared micelles were purified by ultrafiltration [molecular weight cutoff size (MWCO): 10
## [23] "The size distribution of DACHPt-loaded micelles was evaluated by dynamic light scattering (DLS)
## [24] "/ Journal of Controlled Release 101 (2005) 223-232 227 Japan)."
## [25] "The Pt content in the micelles was determined by ion coupled plasma-mass spectrometry (ICP-MS,
## [26] "2. 4. Drug release in phosphate buffered saline"
## [27] "The release of the platinum from the micelles in phosphate buffered saline at 37 8 C was evalua
## [28] "Briefly, a micellar solution of known platinum drug concentration was placed inside a dialysis
## [29] "The released"
## [30] "Pt outside of the dialysis bag was sampled at defined time periods and measured by ICP-MS. 2. 5
## [31] "The kinetic stability of DACHPt-loaded micelles in phosphate buffered saline at 37 8 C was eval
## [32] "Changes in the scattering light intensity (i. e., the Rayleigh ratio at 908 of the scattering a
```

```
## [33] "In this analysis, a decrease in the scattering light intensity is due to a decrease in the appa
## [34] "2. 6."
## [35] "Platinum concentration in plasma and platinum accumulation in tumor CDF 1 mice (female, n=6) we
## [36] "After 14 days, oxaliplatin and DACHPt-loaded micelles were administered i. v. at a dose of 100
## [37] "The mice were sacrificed after defined time periods (1, 5, 9, 27, 48, and 72 h)."
## [38] "The tumor was excised, and the blood was collected from the inferior vena cava, heparinized, an
## [39] "The plasma and the tumor were decomposed on heating in nitric acid, evaporated to dryness, and
## [40] "The Pt concentration in the solution was measured by ICP-MS. 2. 7. In vitro cytotoxicity"
## [41] "Fifty percent growth inhibitory concentration (IC 50) of free oxaliplatin and DACHPt-loaded mic
## [42] "C-26 cells (5000 cells) were cultured in RPMI 1640 containing 10% fetal bovine serum in a 96-we
## [43] "Then, the cells were exposed to free oxaliplatin or DACHPt-loaded micelles for 48 or 72 h, and
## [44] "Cell viability was measured by the formed formazan absorbance at 570 nm."
```

```
head(res)
```

```
##      doc_id paragraph_id sentence_id
## 1504   doc1            1          66
## 1505   doc1            1          66
## 1506   doc1            1          66
## 1507   doc1            1          66
## 1508   doc1            1          66
## 1509   doc1            1          67
##                                                                      sentence
## 1504                                                2. Materials and methods 2.
## 1505                                                2. Materials and methods 2.
## 1506                                                2. Materials and methods 2.
## 1507                                                2. Materials and methods 2.
## 1508                                                2. Materials and methods 2.
## 1509 1. Materials h-Benzyl l-glutamate was bought from Sigma Chemical (St. Louis, MO).
##      token_id     token    lemma  upos xpos        feats head_token_id dep_rel
## 1504        3 Materials material  NOUN  NNS  Number=Plur             6   nsubj
## 1505        4       and      and CCONJ   CC         <NA>             5      cc
## 1506        5   methods   method  NOUN  NNS  Number=Plur             3    conj
## 1507        6         2        2   NUM   CD NumType=Card             0    root
## 1508        7         .        . PUNCT    .         <NA>             6   punct
## 1509        1         1        1   NUM   CD NumType=Card             5  nummod
##      deps          misc
## 1504 <NA>          <NA>
## 1505 <NA>          <NA>
## 1506 <NA>          <NA>
## 1507 <NA> SpaceAfter=No
## 1508 <NA>          <NA>
## 1509 <NA> SpaceAfter=No
```

```
#save RDS file
pdf_name
```

```
## [1] "Cabral, H et al 2005.pdf"
```

```
saveRDS(res, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

Extraction of the materials and methods section

One block for each article :

```
pdf_name <- "Ansari, M 2018.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)
```

```
word_start_mm <- "Materials"
word_end_mm <- "CA"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 846
```

```
## [1] "2 Materials and methods Copper sulphate (CuSO 4), polyvinylpyrrolidone (PVP), sodium borohydrid
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1] 1852
```

```
## [1] "Statistical significance (5%) was evaluated by one-way analysis of variance followed by Tukey t
```

```
mm_section<-cut_conllu_df(846, 1852, conllu_df)
```

```
##  [1] "2 Materials and methods Copper sulphate (CuSO 4), polyvinylpyrrolidone (PVP), sodium borohydri
##  [2] "All chemicals used were of analytical grade."
##  [3] "DG roots were collected, chopped and ground coarsely using mechanical pulveriser."
##  [4] "Soxhlet extraction was carried out to obtain a crude aqueous extract."
##  [5] "The obtained sample was freeze dried (lyophilised) for further use."
##  [6] "2. 1 Synthesis of CuO NPs CuO NPs (CuNP"
##  [7] "DG) were prepared using CuSO 4 (10 mM) as a precursor and aqueous DG root extract (25 mg/ml) a
##  [8] "The colour change from brown to black was observed confirming the formation of CuO NPs."
##  [9] "Thus, obtained NPs were dried and stored [ 10]."
## [10] "In the synthesis of chemical CuO NPs, NaBH 4 (100 mM) was used as a reducing agent with PVP (2
## [11] "The mixture was centrifuged to yield black coloured NPs and then dried for further use."
## [12] "2. 2 Characterisation of CuO NPs"
## [13] "The obtained CuO NPs were dried, characterised using UV-visible spectroscopy, X-ray diffractio
## [14] "The IET Nanobiotechnol., 2018, Vol. 12 Iss. 5, pp."
## [15] "669-672 © The Institution of Engineering and Technology 2018 669 absorption spectrum of the Cu
800 nm, using UV-visible spectrophotometer (Perkin-Elmer Lambda 2)."
## [16] "The crystallinity of synthesised CuO NPs was examined using X-ray diffractometry (Ultima IV, R
## [17] "Diffraction pattern was recorded in a wide range of Bragg angles ( ) using optimum operating pa
## [18] "5405 Å)."
## [19] "To identify the functional groups interacting with CuO NPs, FTIR spectrum was recorded using t
4000 cm-1 at a resolution of 4 cm-1. 2. 3 In-vivo toxicity study Male Wistar rats (260 ± 15 g, 8-
10 weeks) were procured from the Central Animal Facility, SASTRA University, India (213/ SASTRA/RPP/IAE
## [20] "Rats were housed in a humidified condition with 12 h dark/12 h light and were given free acces
## [21] "The rats were then randomly grouped into control and treated, among which treated groups were
## [22] "Chem and CuNP"
## [23] "DG and kept under observation for any behavioural changes for 2 weeks."
## [24] "On the 14 th day, urine and serum were collected and analysed for functional enzymes like alan
## [25] "Ltd, Gujarat, INDIA) (Table 1)."
## [26] "Later, rats were dissected under anaesthesia [sodium thiopentone (i. p.)-80 mg/kg b. wt.] to e
## [27] "The kidney was weighed and cut into two parts wherein one part was stored at -80 C for biochemi
## [28] "2. 4 Biochemical analysis Renal tissue was analysed for thiobarbituric acid reactive substance
## [29] "Chem and CuNP DG NPs."
## [30] "The kidney epithelial cells (LLC PK 1; NCCS, Pune) were cultured in a 48-well plate (0. 2 × 10
## [31] "Once cells reached its 60% confluence, CuO NPs were added and incubated for 24 h."
## [32] "Then the media was assayed for LDH, a typical cellular damage marker."
## [33] "In contrast to the existing hypothesis, ‘green route is safer than the chemical route’, our re
## [34] "2. 6"
## [35] "Isolated mitochondria toxicity"
## [36] "In continuation of cellular toxicity analysis, we designed an experiment to determine the toxi
```

```
## [37] "Briefly, mitochondria were isolated from renal tissue and incubated with varying concentration
## [38] "DG) for 2 h (Table 2)."
## [39] "After incubation, samples were assayed for (i)"
## [40] "SDH activity - to assess the functional status of mitochondria [ 16 ] and (ii) calcium-induced
to determine the integrity of mitochondria [ 17 ]. 2. 7 Statistical analysis Values were expressed as t
## [41] "Statistical significance (5%) was evaluated by one-way analysis of variance followed by Tukey
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Bai, K J et al 2018.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "Methods"
word_end_mm <- "05"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 1639
```

```
## [1] "Methods NiONP preparation and reagent sources NiONPs were obtained from Nanostructured & Amorph
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1] 3658 3893 4244 4266 4299 4319 4381 4723
```

```
## [1] "The criteria used for selection of differentially expressed proteins were previously reported,
## [2] "The level of significance was set to P b 0. 05."
## [3] "LDH significantly increased with NiONP exposure at 20  g in BALF (P b 0. 05), whereas pulmonary
## [4] "LDH significantly increased with NiONP exposure at 20  g in BALF (P b 0. 05), whereas pulmonary
## [5] "We also observed that the inflammatory IL-6 significantly increased with NiONP exposure at 10 a
## [6] "We also observed that the inflammatory IL-6 significantly increased with NiONP exposure at 10 a
## [7] "In terms of lung tissues (adjusted to the total protein), oxidative stress-related 8-OHdGand ap
## [8] "We observed that 8-OHdG in lung tissues (adjusted to the total protein) significantly increased
```

```
mm_section<-cut_conllu_df(1639, 3893, conllu_df)
```

```
##  [1] "Methods NiONP preparation and reagent sources NiONPs were obtained from Nanostructured & Amorp
##  [2] "NiONPs were thoroughly dispersed in phosphate-buffered saline (PBS) supplemented with 5% fetal
##  [3] "The medium was prepared at a final concentration of 20  g/ml after 30 min of sonication for phy
##  [4] "All of the other reagents were obtained from Sigma (St. Louis, MO, USA), unless otherwise expli
##  [5] "Field emission-scanning electron microscopy (FE-SEM) and energy-dispersive x-ray (EDX) microana
##  [6] "The physicochemistry of NiONPs (raw powder and prepared in 5% FBS-containing medium) was inves
##  [7] "16"
##  [8] "Briefly, NiONPswere allowed to adhere to 12-mm carbon adhesive tabs on 13-mm aluminum SEM stubs
##  [9] "The stubs were subsequently coated with platinum (Pt) to an average thickness of 10 nm by a sp
## [10] "FE-SEM images were obtained at an accelerating voltage of 15 kV with ×105 resolution."
## [11] "The EDX Genesis Microanalysis"
## [12] "System was used to determine elements of NiONPs."
## [13] "Specific surface area and endotoxin"
## [14] "The specific surface area of NiONPs was measured by nitrogen adsorption at−196 °C using a Tris
Emmett-Teller (BET) method."
## [15] "The amount of endotoxin in the particles was determined using a Pierce LAL Chromogenic Endotox
## [16] "Hydrodynamic diameter and zeta potential"
## [17] "To determine the physical characteristics of NiONPs when suspended in a 5% FBS-containing solu
## [18] "NiONPs were intratracheally instilled in mice at various concentrations (0, 10, 20, 50, and 10
## [19] "SPECT was used to evaluate inflammatory responses in the lungs at 24 h of exposure."
## [20] "Chest CT was conducted to observe abnormities on days 0, 1, 7, and 28."
```

```
## [21] "On day 1 (24 h) and 29, mice were euthanized (n = 6 per group)."
## [22] "2331 K.-J."
## [23] "Bai et al / Nanomedicine: Nanotechnology, Biology, and Medicine 14 (2018) 2329-
2339 diameters and zeta potential were determined with a Malvern Zetasizer Nano-ZS (Malvern Instruments
## [24] "Animals Six-week-old female BALB/c mice were obtained from BiOLASCO (Taipei, Taiwan)."
## [25] "Mice were maintained at 22 ± 2 °C and 55% ± 10% relative humidity with a 12:12-h light/dark cy
## [26] "Animal experiments were approved and performed in compliance with the animal and ethics review
## [27] "Experimental design"
## [28] "The experimental design is shown in Figure 1."
## [29] "Mice were randomly divided into five groups for exposure to NiONPs at various mass concentrati
## [30] "On day 0 before pulmonary exposure, a mouse in each group was randomly selected for chest compu
## [31] "On day 0, mice received an intratracheal (IT) instillation of 0, 10, 20, 50, or 100  g NiONPs i
## [32] "On day 1, one mouse was randomly selected from each group for the single-photon emission CT (SP
## [33] "The mouse used for the SPECT analysis (n = 1 per group) was not used for any further experimen
## [34] "On day 1, six mice per group were randomly euthanized to determine the acute pulmonary respons
## [35] "On day 29, themice were euthanized (n = 6 per group), and BALF and organs were collected."
## [36] "For BALF sample collection, 1 ml of PBS was used to lavage the lungs followed by centrifugatio
## [37] "Lung tissues were snap-frozen or fixed in 4% (m/v) paraformaldehyde in PBS for the histological
## [38] "The doses of exposure applied in the present study are relevant to human exposure scenarios ac
## [39] "17"
## [40] "The mass concentrations applied in the present study were relevant to occupational exposure sc
## [41] "For acute health effects in mice, we selected 20  g as the highest concentration for IT adminis
## [42] "SPECT analysis"
## [43] "SPECT was conducted with a preclinical tri-modality imaging system (FLEX Triumph, Gamma Medica-
## [44] "21"
## [45] "We injected 0. 5 mCi (37 MBq) of 67 Ga-citrate in 0. 1 mL of a PBS solution into each mouse thr
## [46] "SPECT images were taken at 24 h post-injection of gallium citrate."
## [47] "Those SPECT images were coregistered to relevant CT images and resliced as a 512 × 512 × 512 ma
## [48] "To evaluate the inflammation status, both the right and left lungs were divided into three regi
## [49] "Scoring procedures were conducted in a blinded fashion by the chest physicians KYC and KYL who
## [50] "Total scores were then obtained from the summation of six regions in the lungs."
## [51] "More information is provided in the supplementary materials."
## [52] "CT analysis CTwas performedwith a preclinical tri-modality imaging system (FLEX Triumph, Gamma
## [53] "21 Abnormalities was evaluated on the CT images in a blinded fashion by the chest physicians KY
## [54] "Inflammation, lactic dehydrogenase (LDH), and total protein inBALF Interleukin (IL)-6, IL-10, t
## [55] "Average diameter (nm)a Purity (wt%)a Endotoxin (EU/ml) BET-specific surface area (m 2/g)"
## [56] "DLS hydrodynamic diameter (nm)"
## [57] "Zeta potential (mV) NiONP 20 99. 9 b0."
## [58] "1 7. 2 685. 7 -31."
## [59] "3 a Provided by the manufacturer."
## [60] "Figure 2."
## [61] "FE-SEM and EDX analyses of raw powdered nickel oxide nanoparticles (NiONPs) and suspended NiONP
## [62] "There was no significant change in the morphology when NiONPs were suspended in FBS-containing
## [63] "2332 K.-J. Bai et al / Nanomedicine: Nanotechnology, Biology, and Medicine 14 (2018) 2329-
2339 manufacturer's instructions."
## [64] "Complexes of beads and proteins labeled with phycoerythrin antibodies were acquired with a BD I
## [65] "LDH and total proteinwere determinedwith anLDHcytotoxicity assay kit (Cayman, AnnArbor, MI, USA
## [66] "Enzyme-linked immunosorbent assay (ELISA) of lung tissues"
## [67] "For protein isolation, lung tissueswerehomogenized inRIPAbuffer (Sigma) with Complete  protease
## [68] "An ELISA was used to determine 8-hydroxy-2 -deoxyguanosine (8-OHdG; Cayman) and caspase-3 activ
## [69] "Levels of these markers were adjusted to the total lung-isolated proteins."
## [70] "Protein digestion and iTRAQ labeling"
## [71] "In the present study, lung-isolated proteins within the same group (n = 6) at 20  g of exposure
## [72] "24-h 20  g NiONPs (20  g) and 28-day control Figure 3. Acute effects of nickel oxide nanopartic
```

```
## [73] "H&E staining of lung sections of mice with inflammatory infiltration (×400) for NiONP exposure
## [74] "Cytotoxic lactic dehydrogenase (LDH), total protein, inflammatory interleukin (IL)-6, IL-10, tu
## [75] "Significant increases in LDH, total protein, and IL-6 and a decrease in IL-10 in BALF were obse
## [76] "* P b 0. 05."
## [77] "2333 K.-J. Bai et al / Nanomedicine: Nanotechnology, Biology, and Medicine 14 (2018) 2329-
2339(0  g) vs."
## [78] "28-day 20  g NiONPs (20  g)."
## [79] "Lung-isolated protein samples were depleted of albumin and immunoglobulin G (IgG) with a Qprote
## [80] "Protein digestion and iTRAQ labeling were previously reported."
## [81] "22 More information is provided in the supplementary materials."
## [82] "Liquid chromatography-tandem mass spectroscopy (LC-MS/MS) iTRAQ-labeled protein samples were an
## [83] "23 Briefly, samples were analyzed using a Q Exactive mass spectrometer (Thermo Fisher Scientifi
## [84] "The criteria used for selection of differentially expressed proteins were previously reported,
## [85] "77."
## [86] "Protein functional analysis"
## [87] "The identified upregulated and downregulated proteins in the lungs for the 24-h and 28-day grou
## [88] "abcc. ncifcrf."
## [89] "gov/).25, 26"
## [90] "For the DAVID analysis, an enhanced score  1. 3, set as the threshold, was considered to be si
## [91] "27 Histology"
## [92] "Lungs were fixed in 10% neutral buffered formalin, embedded in paraffin, sectioned, and stained
## [93] "A histopathologist (CHL) examined the histological samples in a blinded fashion under light mic
## [94] "Data analysis Data are presented as the mean ± standard deviation (SD), and all experiments wer
## [95] "A one-way analysis of variance (ANOVA) with Tukey's post-hoc test was used for comparisons amon
## [96] "The statistical analyses were performed using GraphPad vers. 5 (CA, USA) for Microsoft Windows
## [97] "The level of significance was set to P b 0. 05."
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Blum, J L et al 2014.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "Materials"
word_end_mm <- "05"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 1674
```

```
## [1] "Materials and methods Experimental design Mice-About 12-to-14-week-old male CD 1 mice (Charles
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1]  3478 10012 10028 10310 10428 10537 10649 10744 10912
```

```
## [1] "Treatment means were considered significantly different at p  0. 05."
## [2] "*Significant (p<0. 05) loss of body weight compared to control (A) or significant (p<0. 05) char
## [3] "*Significant (p<0. 05) loss of body weight compared to control (A) or significant (p<0. 05) char
## [4] "*p<0. 05."
## [5] "*p<0. 05."
## [6] "*p<0. 05."
## [7] "*p<0. 05."
## [8] "*p<0. 05."
## [9] "Bars with different letters are significantly different (*p<0. 05) using Fisher's LSD."
```

```
mm_section<-cut_conllu_df(1674, 3478, conllu_df)
```

```
##  [1] "Materials and methods Experimental design Mice-About 12-to-14-week-old male CD 1 mice (Charles
##  [2] "PJ Murphy, Montville, NJ) at the New York University, Department of Environmental Medicine Anir
##  [3] "Mice were maintained at 22°C and 55% humidity on a 12-hour light/dark cycle and provided both w
##  [4] "Page 3 Inhal Toxicol."
##  [5] "Author manuscript; available in PMC 2014"
##  [6] "June 02. N IH -P"
##  [7] "A A uthor M anuscript N IH -P"
##  [8] "A A uthor M anuscript IN) ad libitum."
##  [9] "All animal studies were performed in accordance with the NYU Medical Center Institutional Anima
## [10] "CdO NP generation and exposure-Mice were acclimated to the nose-only exposure tubes for three d
## [11] "Following acclimation, animals were exposed daily for three hours/day for seven consecutive day
## [12] "Briefly, CdO NPs were generated in a Palas (Model GFG-1000) arc furnace (Karlsruhe, Germany) at
## [13] "995% pure, ESPI, Ashland, OR) as electrodes."
## [14] "The generated NPs were carried by 3. 5 L/min"
## [15] "Ar flow and mixed with 23."
## [16] "5 L/min particle-free air."
## [17] "Cooled (using ice packs) carrier gas was mixed with 1 L/min oxygen (Ultrapure, AGI, NJ) resulti
## [18] "Size distribution and number concentration of freshlygenerated CdO NPs were determined in real-
## [19] "Nanoparticles were collected on a pre-weighed Teflon filter (37 mm, 0. 2  m pore size) and the
## [20] "For the studies here, experimental animals were exposed by inhalation to an average concentrati
## [21] "3 ± 0. 1 nm) and were polydispersed with a size geometric standard deviation of 1."
## [22] "5."
## [23] "Nanoparticle size (measured at a high-voltage frequency of 40 Hz) was measured using an atomic
## [24] "The current permissible exposure limits for cadmium in the workplace is a time-weighted average
## [25] "Assuming an average human worker inhales 0. 5 m 3 of air per hour, exposure at the regulatory l
## [26] "The average mouse respiration rate is 220 breaths/min with 0. 18 mL per breath (Braun et al., 1
## [27] "Under these experimental conditions, the mice received ~1. 7  g CdO NPs per day which is below
## [28] "Following the final exposure, mice were either euthanized after 24 h or allowed to recover for
## [29] "Equal numbers of mice (n = 5-7 per time point/treatment group) were exposed either to filtered
## [30] "Both treatment groups were euthanized at the appropriate time postexposure using SleepAway (0.
## [31] "Bronchoalveolar lavage Bronchoalveolar lavage fluid (BALF) was collected following euthanasia l
## [32] "To maximize the collection of cells and lavagable proteins, lungs were lavaged three times in s
## [33] "Blum et al."
## [34] "Page 4 Inhal Toxicol."
## [35] "June 02."
## [36] "N IH -P"
## [37] "A A uthor M anuscript volume of 1 mL."
## [38] "BALF samples were recovered (~1 mL), stored on ice prior to centrifugation, and were centrifuge
15 min at 6400 rpm (Capsule TOMY HF-120 microcentrifuge, Peninsula Laboratories, Inc., Belmont, CA)."
## [39] "The supernatants were aliquoted and used immediately to measure lacatate dehydrogenase (LDH) ac
## [40] "The recovered cell pellets were resuspended in 0. 4 mL of DPBS and the total cell number and vi
## [41] "To determine cell profiles by differential counts, 100  L of the cell suspension were spread on
## [42] "Cell sub-population profiles were determined by counting 100 white blood cells per slide (two s
## [43] "Only macrophages and neutrophils were counted, as they were the only white blood cell types obs
## [44] "After lavage, lungs were fixed by instilling 1 mL of the 10% buffered formalin."
## [45] "Fixed lungs were then embedded in paraffin and sections cut and stained with H&E."
## [46] "Slides were analyzed for any major changes by a pathologist blinded to the treatment groups."
## [47] "LDH activity Lactate dehydrogenase activity was measured in BALF samples within 1 hour followin
## [48] "Standards were prepared from L-LDH from bovine heart (Sigma-Aldrich, Allentown, PA) in concentr
## [49] "LDH activity was calculated from a standard curve and presented as LDH activity units."
## [50] "Lavagable protein Total protein concentration was measured using a Biorad DC Protein assay (Bio
1000  g/mL bovine serum albumin as the standard."
## [51] "Gelatin zymography Matrix metalloproteinase (MMP) activity was measured using gelatin zymograph
## [52] "The low protein concentration of BALF measured from control mice necessitated the need for acet
```

```
## [53] "Precipitation was performed by adding five sample volumes of cold acetone to each sample, incul
## [54] "The supernatant was removed and pellet allowed to air dry prior to the addition of loading buf
## [55] "To normalize MMP activity between replicate gels, each gel contained a reference sample consis
## [56] "After de-staining, zymograms were dried on a gel dryer and imaged using a digital imager (Alpha
## [57] "Quantification was based on differences in band integrated density values (IDV) with the backg
## [58] "Page 5 Inhal Toxicol."
## [59] "A A uthor M anuscript Quantification of cytokines in BALF Lavage samples were thawed slowly on
## [60] "Differential blood counts"
## [61] "Blood was collected from the abdominal aorta in a heparinized syringe following euthanasia."
## [62] "At the time of collection, blood was spread on duplicate glass slides and allowed to dry overni
## [63] "White blood cells were identified and counted (100 cells/slide on duplicate slides from each mo
## [64] "Plasma samples, collected by centrifugation using a protocol similar to that used for BALF samp
## [65] "Phagocyte uptake assay Blood collected at the time of euthanasia from the abdominal aorta was p
## [66] "The following day, blood samples were incubated with pHrodo Escherichia coli BioParticles (Phag
## [67] "Samples were analyzed using a 633 nm argon-ion laser (MACSQuant Analyzer, Miltenyi Biotec Inc.
## [68] "Phagocytic uptake was measured as a percentage of total counted cells using a plot of forward s
## [69] "Statistical analysis"
## [70] "A one-way analysis of variance (ANOVA) followed by Fisher's LSD post hoc testing (when appropri
## [71] "Treatment means were considered significantly different at p  0. 05."
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Chen et al 2008.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "Materials"
word_end_mm <- "conditions"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 952
```

```
## [1] "Materials and Methods Nanoparticles."
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1]  395 1557 2381 2414 2752 2893
```

```
## [1] "Inhalation of SiO 2 nanoparticles under identical conditions caused pulmonary and cardiovascula
## [2] "Generation conditions for"
## [3] "SIEMN is not only a novel facility in which animals can inhale manufactured nanoparticles under
## [4] "Using SIEMN, young, adult, and old rats inhaled SiO 2 nanoparticles under identical conditions."
## [5] "The results imply that old rats suffered a more severe inflammatory response than young and adul
## [6] "This clearly demonstrated that, under identical exposure conditions, manufactured"
```

```
mm_section<-cut_conllu_df(952, 2414, conllu_df)
```

```
##  [1] "Materials and Methods Nanoparticles."
##  [2] "Amorphous SiO 2 nanoparticles (purity >99. 9%) were purchased from Jiangsu Haitai Nano Material
##  [3] "Before animal experiments, aerosol concentration in the particle exposure system was measured b
##  [4] "Initially, the distribution of particle size was measured by high-resolution atomic force micro
##  [5] "The SiO 2 nanoparticle sample was scanned by AFM."
##  [6] "The diameter in three directions was measured for each nanoparticle, and mean values of the geo
##  [7] "SiO 2 nanoparticles selected randomly calculated (Figure S 1 in Supporting Information)."
##  [8] "Their arithmetic mean *"
##  [9] "Corresponding author address:"
## [10] "Laboratory for Bio-Environmental Effects of Nanomaterials and Nanosafety, Chinese Academy of Sc
```

```
## [11] "cn (Y. Z.), xinggm@ ihep."
## [12] "ac."
## [13] "cn (G. X.). † Chinese Academy of Sciences and National Center for Nanosciences and Technology
## [14] "| The Graduate University of Chinese Academy of Science. § Third Hospital of Peking University
## [15] "Environ."
## [16] "Sci."
## [17] "Technol."
## [18] "2008, 42, 8985-8992 10."
## [19] "1021/es 800975 u CCC: $40. 75  2008 American Chemical Society VOL. 42, NO."
## [20] "23, 2008 / ENVIRONMENTAL SCIENCE & TECHNOLOGY 9 8985 Published on Web 11/04/2008 D ow nl oa de
## [21] "Fe br ua ry 7 , 2 01 9 at 1 6: 11 :4 2 (U T C )."
## [22] "Se e ht tp s: //p ub s. ac s. or g/ sh ar in gg ui de lin es f or o pt io ns o n ho w to le gi
## [23] "The average size of the SiO 2 nanoparticle was 37. 9 ( 3. 3 nm; the specific surface area was
## [24] "83 × 105 cm 2/g, and the particle number was 1. 52 × 1010 µg-1."
## [25] "SiO 2 nanoparticles were dispersed into medium by ultrasonic waves in sample preparation for A
## [26] "Thus, results obtained from AFM measurement (Table S 1, Supporting Information) are close to t
## [27] "SiO 2 nanoparticles but do not reflect the actual size of SiO 2 nanoparticles as an aerosol in
## [28] "Nanoparticles present as aerosol in the exposure chamber may change their size from their orig
## [29] "We therefore did an online measurement of SiO 2 aerosol generated in the exposure chamber, whi
## [30] "Generation conditions for"
## [31] "SiO 2 aerosol were identical to those in the animal experiments."
## [32] "We measured the background inside the exposure chamber before generation FIGURE 1. Pulmonary da
## [33] "Panels a, b, c, and d display changes in bronchoalveolar lavage parameters as a function of ag
## [34] "A and B denote the Duncan class of Duncan's multiple-range test."
## [35] "Panels e-j show the pathologic changes in lung tissues of exposed rats in the old (e), young (
## [36] "Panels h-j are their corresponding controls (magnification, × 100)."
## [37] "B indicates bronchia, and IC denotes inflammatory cells. n ) 6 in each group, identical for al
## [38] "8986 9 ENVIRONMENTAL SCIENCE & TECHNOLOGY / VOL. 42, NO."
## [39] "23, 2008 of SiO 2 aerosol."
## [40] "SiO 2 aerosol was generated and particle sizes (sample data) were measured online."
## [41] "Final results were obtained from subtracting background data from sample data (Figure S 2, Supp
## [42] "Online measurement by Scanning Mobility Particle Sizers gave a mean size of SiO 2 aerosol in t
## [43] "This was attributed to self-aggregation of"
## [44] "SiO 2 nanoparticles in the aerosol state and was also demonstrated from the size distribution
## [45] "More details for the size distribution measurement of SiO 2 aerosol in the exposure chamber ar
## [46] "Methods."
## [47] "Animal experiments were carried out in compliance with guidelines set by the local ethics comm
## [48] "Details regarding animals, the particle exposure system, and online measurement of aerosol con
## [49] "In brief, male Sprague-Dawley rats aged 3 weeks (65 g), 8 weeks (265 g), and 20 months (670 g)
## [50] "They inhaled air containing"
## [51] "SiO 2 nanoparticles for four weeks."
## [52] "Grouping of the experimental animals was based on a two-factor cross-classification model: age
## [53] "The age factor involved three levels (young/ adult/old) and the factor of administration involv
## [54] "Rats in groups labeled YE (young exposure), AE (adult exposure), and OE (old exposure) inhaled
## [55] "SiO 2 nanoparticles for 40 min per day."
## [56] "Rats in corresponding control groups labeled YC (young control), AE (adult control), and OE (o
## [57] "A concentration of SiO 2 nanoparticles of 24."
## [58] "1 mg/m 3 was chosen."
## [59] "This concentration is approximately equivalent to 10× the particulate concentration in sand-du
## [60] "In the workplace or for users of nanoparticle-related products, the concentration in a local a
## [61] "The conversion quantity of animal-inhaled SiO 2 nanoparticles was designed to be approximately
## [62] "At the end of administration, analysis of bronchoalveolar lavage (BAL) fluid, electrocardiogra
## [63] "Mimicking natural (physiologic) inhalation is important for risk evaluation of respirable nano
## [64] "There are few nanotoxicity studies with natural inhalation exposure in animals because of the
```

```
## [65] "We therefore designed a novel nanoparticles exposure system: system of inhalation exposure for
## [66] "SIEMN is a sealed Plexiglas exposure chamber specifically for nanoparticle inhalation in animal
## [67] "A detailed description of SIEMN is given in Supporting Information (Figures S 3 and S 2 and Tab
## [68] "SIEMN is not only a novel facility in which animals can inhale manufactured nanoparticles under
## [69] "Using SIEMN, young, adult, and old rats inhaled SiO 2 nanoparticles under identical conditions
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Jamshidzadeh, A et al 2015.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "MATERIALS"
word_end_mm <- "05"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 918
```

```
## [1] "MATERIALS AND METHODS Chemicals Gentamicin sulfate was obtained from Acros (Fair Lawn, NJ)."
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1] 1541 1632 1648 1731 1747
```

```
## [1] "Differences between experimental groups were considered significant when P < 0. 05. 0 0. 1 0. 2
## [2] "* indicates a significant difference as compared with control animals (P < 0. 05). a indicates s
## [3] "* indicates a significant difference as compared with control animals (P < 0. 05). a indicates s
## [4] "* indicates a significant difference as compared with control animals (P < 0. 05). a indicates s
## [5] "* indicates a significant difference as compared with control animals (P < 0. 05). a indicates s
```

```
mm_section<-cut_conllu_df(918, 1541, conllu_df)
```

```
##  [1] "MATERIALS AND METHODS Chemicals Gentamicin sulfate was obtained from Acros (Fair Lawn, NJ)."
##  [2] "Sodium alginate was purchased from Sigma Aldrich (St. Louis, MO)."
##  [3] "Formaldehyde was obtained from Merck (Dardamstd, Germany)."
##  [4] "Kits for evaluating plasma markers of kidney functionality (BUN and creatinine) were prepared i
##  [5] "All salts for buffer solutions were of the highest grade commercially available."
##  [6] "Animals Male Swiss albino mice (25-35 g) were obtained from Shiraz University of Medical Scienc
##  [7] "Mice were housed in cages on wood bedding at a temperature of 25 ± 3°C."
##  [8] "Animals had free access to food and water."
##  [9] "The animals received humane care and use and were handled according to the animal handling prot
## [10] "Mice were randomly divided equally into four groups of six animals."
## [11] "The treatments were as follows: (A) control (vehicle-treated mice, 0. 5 mL of normal saline, i
## [12] "J Biochem Molecular Toxicology DOI 10."
## [13] "1002/jbt Volume 29, Number 2, 2015 GENTAMICIN NANOPARTICLES NEPHROTOXICITY 59 Plasma Biochemica
## [14] "Then, the kidneys were removed."
## [15] "Plasma samples were used to measure creatinine and BUN as the standard parameters of evaluating
## [16] "Creatinine and BUN were assessed by commercial kits available (Pars Azmun, Tehran, Iran)."
## [17] "Kidney Histopathological Evaluation"
## [18] "For histopathological evaluation, samples of kidney were fixed in formalin (10%)."
## [19] "Paraffin-embedded sections of tissue were prepared and stained with hematoxylin and eosin befor
## [20] "The ionic gelation between gentamicin sulfate and sodium alginate was used to design desired na
## [21] "An aqueous solution of alginate was prepared and while being mixed gentamicin sulfate solution
## [22] "The preparation method was optimized with respect to the drug volume, polymer volume (constant
## [23] "Then the effect of each group of variables on particle size distribution and release profiles w
## [24] "To optimize the particle size distribution and to get optimum average particle size, an orthogo
## [25] "The average particle size of nanoparticle in the optimized formulation was 80 nm."
```

```
## [26] "Gentamicin was assayed in pharmaceutical samples using a HPLC method precolumn derivatization w
## [27] "Loading efficiency of the nanoparticles in the optimized formulation was 85%."
## [28] "The physical stability of the particles at room temperate and in refrigerated samples during th
## [29] "Statistical Analysis"
## [30] "Results are shown as mean ± SE for at least six animals."
## [31] "Comparisons between multiple groups were made by a one-way analysis of variance followed by Tur
## [32] "Differences between experimental groups were considered significant when P < 0. 05. 0 0. 1 0. 1
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Li, C et al 2009.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "Materials"
word_end_mm <- "t-test"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 3161
```

```
## [1] "Materials and methods Animal handling Animal experiments were conducted in the animal facility a
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1] 5965
```

```
## [1] "All data were shown as mean+S. E. M. and statistical analyses were conducted using the student t
```

```
mm_section<-cut_conllu_df(3161, 5965, conllu_df)
```

```
##   [1] "Materials and methods Animal handling Animal experiments were conducted in the animal facility
##   [2] "Six- to 10-week-old male BALB/c mice were used (Vital River, Beijing)."
##   [3] "They were caged in a specific pathogen-free facility as groups of five or less and fed ad libi
##   [4] "Euthanasia was performed with pentobarbital sodium."
##   [5] "Cells, PAMAM dendrimer nanoparticles and antibodies"
##   [6] "The human lung adenocarcinoma"
##   [7] "A 549 cell line was purchased from ATCC, and cultured in F-12/HAM's (Hyclone) medium supplemen
##   [8] "PAMAM dendrimers G 1, G 2, G 3, G 3. 5, G 4, G 4. 5, G 5, G 5. 5, G 6, G 7, G 7. 5 and G 8 wer
##   [9] "The primary antibodies used in the analysis, anti-mTOR, anti-phospho-mTOR (Ser 2481), anti-AKT
##  [10] "Anti-TSC 2 and anti-ATG 6 antibodies were purchased from Santa Cruz Biotechnology."
##  [11] "Anti-b-actin antibody was purchased from Sigma-Aldrich."
##  [12] "Horseradish peroxidase-conjugated secondary antibodies and western blotting luminal reagents w
##  [13] "CellTiter 96 AQueous One Solution Cell Proliferation"
##  [14] "Assay kit was purchased from Promega Corporation."
##  [15] "Caspase-3 fluorescence determination kit was purchased from Beijing Baosai Biotech Limited Cor
##  [16] "Handling of nanoparticles PAMAM dendrimers G 1, G 2, G 3, G 3. 5, G 4, G 4. 5, G 5, G 5. 5, G
##  [17] "They were air Figure 4 Autophagy is a general mechanism in toxic PAMAM nanoparticles-induced c
##  [18] "(A) LC 3 immunoblotting of lysates of A 549 cells after control, G 5. 5 (100 mg/ml), G 4 (100
##  [19] "Band density was calculated using AlphaEaseFC software."
##  [20] "(C) MTT assay of A 549 cells after treatment for 24 h with control, G 5. 5 (100 mg/ml), G 4 (1
##  [21] "The concentration of 3 MA is 3 mM in control, G 5. 5, G 4, G 5, G 6 and G 7 groups, but 1 mM i
##  [22] "**P , 0. 01 and *P , 0. 05."
##  [23] "PAMAM nanoparticles promote acute lung injury Journal of Molecular Cell Biology j 41 D ow nloa
##  [24] "oup."
##  [25] "com /jm cb/article-abstract/1/1/37/873399 by EC JR C Ispra site user on 14 February 2019"
##  [26] "Figure 5 PAMAM"
##  [27] "G 3 induces autophagy in A 549 cells through the AKT-TSC-mTOR pathway. (A) Western blot analys
```

```
##  [28] "(C) Western blot analysis of PAMAM dendrimer (100 mg/ml)-treated and control cells probed wit
##  [29] "PAMAM dendrimer treatment was for 24 h. (D) Relative ratio of phospho-S 6 band density to tha
##  [30] "(E)"
##  [31] "Confocal images of A 549 cells, transfected with control siRNA or TSC 2 siRNA, then followed
##  [32] "G 3 treatment for 24 h. (G) MTT assay of A 549 cells transfected with control siRNA or TSC 2 s
##  [33] "Relative ratio of phospho-Akt band density to that of Akt in control, PAMAM G 5. 5 (100 mg/ml
##  [34] "Band density was calculated using AlphaEaseFC software. (J) Schematic representation of the si
##  [35] "42 j Journal of Molecular Cell Biology Li et al."
##  [36] "D ow nloaded from https://academ ic. oup."
##  [37] "com /jm cb/article-abstract/1/1/37/873399 by EC JR C Ispra site user on 14 February 2019 drie
##  [38] "As vehicle, PBS was then added to dissolve the nanoparticles."
##  [39] "MTT assay"
##  [40] "A 549 cells were seeded in 96-well plate at 1 105/ml."
##  [41] "PAMAM G 1, G 2, G 3, G 3. 5, G 4, G 4. 5, G 5, G 5. 5, G 6, G 7, G 7. 5 and G 8, or an equal v
##  [42] "In the PAMAMs plus 3-MA group, 3-MA was added 1 h before PAMAMs."
##  [43] "Each group had triplicate wells."
##  [44] "After 24 h, 20 ml of CellTiter 96 AQueous One Solution Cell Proliferation"
##  [45] "Assay solution was added to each well, and incubated for another 2 h."
##  [46] "Absorbance was then recorded at 490 nm."
##  [47] "DNA extraction"
##  [48] "A 549 cells were seeded in 6-cm plate at 1 106/ml."
##  [49] "DMSO 6%, PAMAM G 3 (100 mg/ml) or an equal volume of control was added to the plate on the ne
##  [50] "Cells were harvested and DNA was extracted as previously described."
##  [51] "Caspase-3 activity determination"
##  [52] "DMSO 6%, PAMAM G 3 (100 mg/ml), PAMAM G 5. 5 (100 mg/ml) or an equal volume of control was ad
##  [53] "After 24 h, cells were lysed and caspase-3 activity was determined Figure 6"
##  [54] "Autophagy plays a critical role in PAMAM G 3-induced acute lung injury in mice. (A) HE staini
##  [55] "(B)"
##  [56] "Wet/dry ratios of lung tissues of Balb/c mice after intratracheal administration of control v
##  [57] "PAMAM dendrimer treatment was for 16 h. **P , 0. 01 and *P , 0. 05."
##  [58] "(C) Changes of lung elastance in Balb/c mice after intratracheal administration of control ve
##  [59] "(D)"
##  [60] "Survival rate of Balb/c mice after intratracheal administration of control vehicle, PAMAM G 5
##  [61] "PAMAM nanoparticles promote acute lung injury Journal of Molecular Cell Biology j 43 D ow nloa
##  [62] "com /jm cb/article-abstract/1/1/37/873399 by EC JR C Ispra site user on 14 February 2019 using
##  [63] "Western blotting A 549 cells were seeded at 1 105/ml in 12-well plate."
##  [64] "PAMAM"
##  [65] "G 3 (100 mg/ml), PAMAM G 5. 5 (100 mg/ml) or an equal volume of control was added to the wells
##  [66] "After 24 h, cells were lysed in lysis buffer, denatured at 978 C for 10 min and subjected to w
##  [67] "Transmission electron microscopy"
##  [68] "A 549 cells were seeded in 6-well plate at 2 105/ml."
##  [69] "PAMAM G 3 (100 mg/ml), PAMAM G 5."
##  [70] "5 (100 mg/ml) or an equal volume of control was added to the wells in the next day."
##  [71] "In the PAMAM G 3 plus 3-MA group, 3-MA (10 mM) was added 1 h before PAMAM G 3."
##  [72] "After 24 h, cells were trypsin-digested and centrifuged at 800 g for 5 min."
##  [73] "The supernatant was discarded and the cells were fixed with 2. 5% glutaraldehyde in 0. 1 M so
##  [74] "The samples were then fixed in 1% OsO 4 for 1 h and dehydrated by increasing concentrations o
##  [75] "Ultra-thin sections were obtained and stained with uranyl acetate and lead citrate."
##  [76] "A cell showing two or more autophagosomes was defined to be an autophagypositive cell."
##  [77] "LC 3-EGFP counting"
##  [78] "A 549 cells were seeded on coverslips in 24-well plate."
##  [79] "One day later, LC 3-EGFP plasmid was transfected."
##  [80] "Forty-eight hours after transfection, PAMAM G 3 (30 mg/ml), PAMAM G 5. 5 (30 mg/ml) or an equa
##  [81] "In the PAMAM G 3 plus 3 MA group, 3 MA (10 mM) was added 1 h before PAMAM G 3. After 24 h, EG
```

```
##  [82] "Images were captured under the 100X oil objective (Plan-Apo 1. 4) with the confocal acquisitio
##  [83] "A cell containing 10 or more EGFP dots was defined to be an LC 3-positive cell."
##  [84] "LC 3-EGFP counting in TSC 2 siRNA-treated A 549 cells"
##  [85] "A 549 cells were seeded in 24-well plate the day before being transfected with siRNA against T
##  [86] "Twenty-four hours after TSC 2 transfection, cells were transfected with LC 3-EGFP plasmid."
##  [87] "Another 36 h later, the effect of the siRNA was determined by western blot with anti-TSC 2 ant
##  [88] "In parallel, cells were treated with PAMAM G 3 (30 mg/ml)."
##  [89] "The accumulation of EGFP-LC 3 was determined by Leica Confocal Microscope as described above."
##  [90] "MTT assay in TSC 2 siRNA-treated cells A 549 cells were seeded in 24-well plate the day before
##  [91] "Another 48 h later, the effect of the siRNA was determined by western blot with anti-TSC 2 ant
##  [92] "In parallel, 24 h after transfection, cells were trypsin-digested and seeded on 96-well plate
##  [93] "G 3 (100 mg/ml) was added to the TSC 2 siRNA and control siRNA group in the next day, and the
##  [94] "MTT assay in ATG 6 siRNA-treated cells A 549 cells were seeded in 24-well plate."
##  [95] "Twenty-four hours later, cells were transfected with siRNA against ATG 6 (100 mM, Santa Cruz B
##  [96] "Another 48 h later, the effect of the siRNA was determined by western blot with anti-ATG 6 ant
##  [97] "In parallel, 24 h after transfection, cells were trypsin-digested and seeded on 96-well plates
##  [98] "G 3 (100 mg/ml) was added to the ATG 6 siRNA and control siRNA group in the next day, and the
##  [99] "Mice lung tissue histopathological examination"
## [100] "Four hours after intratracheal administration of control, PAMAM G 5. 5 (50 mg/kg) or PAMAM G 3
## [101] "Lungs were fixed in formalin for 48 h and then embedded in paraffin."
## [102] "Ultra-thin sections were obtained and stained with hematoxylin-eosin."
## [103] "Each slide was independently examined by three different pathologists."
## [104] "Mice lung wet/dry ratio assay The Balb/c mice were randomly grouped."
## [105] "After anesthesia by intraperitoneal injection with 1% pentobarbital sodium solution, they were
## [106] "In the 3 MA-only group, 3 MA (15 mg/kg) was injected intraperitoneally."
## [107] "In the rescue group, 1 h after intraperitoneal injection with 3 MA (15 mg/kg), PAMAM G 3 (50 m
## [108] "After spontaneous breathing for 16 h, mice were sacrificed and the lungs were assessed for the
## [109] "To obtain the dry weight, the lungs of mice were dried in an oven at 558 C for 24 h."
## [110] "Assay for mice lung elastance changes"
## [111] "The Balb/c mice were randomly grouped."
## [112] "After anesthesia by intraperitoneal injection with 1% pentobarbital sodium solution, they were
## [113] "In the 3-methyladenine (3 MA) rescue group, 3 MA (15 mg/ kg) was injected intraperitoneally be
## [114] "G 3 (50 mg/kg) was administered intratracheally."
## [115] "Then elastance was tested by BUXCO pulmonary function testing (PFT) every 30 min during the sp
## [116] "44 j Journal of Molecular Cell Biology Li et al."
## [117] "Mice survival rate assay The Balb/c mice were randomly grouped."
## [118] "After anesthesia by intraperitoneal injection with 1% pentobarbital sodium solution they were
## [119] "3 MA (15 mg/kg) was injected twice intraperitoneally, 12 h and 1 h, respectively, before intra
## [120] "The survival/death status of mice was recorded every 1 h for a total of 24 h."
## [121] "The data were analysed by SPSS software."
## [122] "Statistical analyses"
## [123] "All data were shown as mean+S. E. M. and statistical analyses were conducted using the student
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Li, J et al 2013.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "2."
word_end_mm <- "05"

display_occurences(word_start_mm, conllu_df)
```

```
##  [1]  1044  1151  1538  1651  2232  2236  2647  3492  3504  3509  3579  3707
## [13]  3812  3840  3938  3944  3951  9097 10507
```

```
##  [1] "2. Materials and methods 2."
##  [2] "2. 2. Animal handling and sample collection"
##  [3] "2. 4. Blood biochemical analysis Clinical chemistry analysis of serum was carried out by standa
##  [4] "2. 5. Sample preparation and 1 H NMR spectroscopic analysis Samples of plasma (255 µl) were mi
##  [5] "5 µs and an acquisition time of 2. 66 s. 2. 6."
##  [6] "5 µs and an acquisition time of 2. 66 s. 2. 6."
##  [7] "Here, a correlation coefficient (determined by the Pearson's product-moment correlation coeffi
##  [8] "36 c 54. 10 ± 2. 10d 47."
##  [9] "33 ± 3. 37 56. 63 ± 2. 88 58. 47 ± 2. 36 54."
## [10] "33 ± 3. 37 56. 63 ± 2. 88 58. 47 ± 2. 36 54."
## [11] "87 ± 2. 10 25."
## [12] "40 ± 27. 68 e 42. 50 ± 2. 07 50. 22 ± 6. 85 d 43."
## [13] "4. 07 ± 0. 49 3. 31 ± 0. 69 d 3. 20 ± 0. 53 d GGT (U/L) 0. 08 ± 0. 04 2. 00 ± 0."
## [14] "4. 38 ± 2. 33 e ALP (U/L)"
## [15] "1. 86 ± 0. 34 2. 57 ± 0. 35 e 2. 32 ± 0. 19 d 2. 32 ± 0. 14 2. 74 ± 0. 20e"
## [16] "1. 86 ± 0. 34 2. 57 ± 0. 35 e 2. 32 ± 0. 19 d 2. 32 ± 0. 14 2. 74 ± 0. 20e"
## [17] "1. 86 ± 0. 34 2. 57 ± 0. 35 e 2. 32 ± 0. 19 d 2. 32 ± 0. 14 2. 74 ± 0. 20e"
## [18] "93 0. 80 - a Abbreviations in tables explained in the notes of figure 2. pyruvate."
## [19] "84 - 0. 80 a Abbreviations in tables explained in the notes of figure 2. 11 Nanotechnology 24
```

```
display_occurences(word_end_mm, conllu_df)
```

```
##  [1] 2645 2693 3156 3179 3228 3239 3638 3995 4026 4179 4501 4585 4900 5070 5476
## [16] 5499 5621
```

```
##  [1] "Here, a correlation coefficient (determined by the Pearson's product-moment correlation coeffi
##  [2] "The criterion for statistical significance was p < 0. 05."
##  [3] "H, high-dose group; 6, 6 h post-dose; 48, 48 h post-dose. * Significant differences between tw
##  [4] "of spleen/BD, which increased to 114% of control (P < 0. 05), while the ratios of kidney/BD, li
##  [5] "of spleen/BD, which increased to 114% of control (P < 0. 05), while the ratios of kidney/BD, li
##  [6] "of spleen/BD, which increased to 114% of control (P < 0. 05), while the ratios of kidney/BD, li
##  [7] "39 ± 0. 10 1. 25 ± 0. 06 0. 99 ± 0. 04 e 0. 96 ± 0. 05 e Ibil (µmol/L)"
##  [8] "e 0. 58 ± 0. 05 0. 68 ± 0. 07 0. 81 ± 0. 06"
##  [9] "71 ± 0. 14 LDL-C (mmol/L) 0. 15 ± 0. 05 0. 29 ± 0. 11 d 0. 21 ± 0. 04 d 0. 19 ± 0. 02 0. 37 ± 0
## [10] "3. 82 ± 1. 56 6. 26 ± 0. 88 d 14. 09 ± 3. 67 e 5. 80 ± 1. 08 5. 05 ± 0."
## [11] "d Significant differences between two groups as judged by Student's t-test using SPSS with P <
## [12] "* Significant differences between two groups as judged by Student's t-test (*P < 0. 05; **P < 0
## [13] "8-fold at 48 h pd, respectively, while ALP elevations were moderate, to 152% (P < 0. 001, low-d
## [14] "The ratios of AST/ALT in both the low- and high-dose groups show significant decrease (P < 0. 0
## [15] "As shown in table 1, exposure of rats to Mn-NPs resulted in an increase of creatinine (Cn) to h
## [16] "As shown in table 1, exposure of rats to Mn-NPs resulted in an increase of creatinine (Cn) to h
## [17] "At 48 h pd, no apparent change was measured in the levels of Cn, BUN and uric acid (UA) and the
```

```
mm_section<-cut_conllu_df(1044, 2693, conllu_df)
```

```
##  [1] "2. Materials and methods 2."
##  [2] "1. Synthesis and characterization of MnO nanoparticles Size-homogeneous and monodisperse 10 nm
##  [3] "MnO nanoparticles were synthesized by high-temperature thermal decomposition of manganese oleat
##  [4] "org/Nano/24/455102/mmedia)."
##  [5] "Mn-NPs were freshly well dispersed by ultrasound in saline solution before use."
##  [6] "Prior to animal experiments, the Mn concentration of the nanoparticles in phosphate buffer solu
atomic emission spectrometry (ICP-AES)."
##  [7] "2. 2. Animal handling and sample collection"
##  [8] "All animal experimental protocols complied with the local guidelines for animal use and care, a
##  [9] "A total of 43 nine-week-old male SD rats (243 ± 13 g) were used in our study."
## [10] "The environment conditions were set at 21-26 ◦C with a relative humidity of 50 ± 10%, and a 12/
```

```
## [11] "Food and tap water were provided ad libitum, and body weights were recorded daily."
## [12] "After one week of acclimatization, a single dose of Mn-NPs in saline was administrated intraver
## [13] "A control group (n = 13) was treated with saline only."
## [14] "Individual urine samples were collected in ice-cooled vessels containing 1% sodium azide (0. 1
## [15] "Animals were sacrificed by exsanguination under isoflurane anesthesia at time points 6 (six ra
## [16] "The blood sample was divided into two aliquots, one serum for biochemical analysis and the oth
## [17] "After weighing, brain, kidney, liver, lung and spleen tissue were excised in triplicate: one bo
## [18] "These samples were stored at-80 ◦C until used."
## [19] "2. 3. Histopathology"
## [20] "The randomly selected samples of kidney, liver, lung and spleen tissue from Mn-NP-treated and c
## [21] "After dehydrating, the biopsies embedded in wax were sectioned at 5 µm, and stained with hemato
## [22] "2. 4. Blood biochemical analysis Clinical chemistry analysis of serum was carried out by standa
## [23] "The biochemical parameters included glucose, indirect bilirubin, total bile acid, UA, BUN, crea
## [24] "All parameters are expressed as mean± standard deviation (SD)."
## [25] "2. 5. Sample preparation and 1 H NMR spectroscopic analysis Samples of plasma (255 µl) were mix
## [26] "After centrifugation at 10 000 g 2 Nanotechnology 24 (2013) 455102 J Li et al at 4 ◦C for 10 mi
## [27] "1 H NMR spectra of these samples were acquired using a 500 MHz Varian spectrometer at 296 K. St
## [28] "1 D 1 H spectra were acquired with a water-suppressed CPMG pulse sequence."
## [29] "For each sample, 64 FIDs were collected into 20 K data points over a spectral width of 10 000"
## [30] "Hz with a relaxation delay of 0."
## [31] "23 ms."
## [32] "Samples of urine (455 µl) were mixed with 55 µl of deuterated phosphate buffer solution (NaH 2
## [33] "The mixture was left to stand for 10 min and centrifuged at 10 000 g at 4 ◦C for 10 min to remo
## [34] "500 µl of the supernatants were transferred into a 5 mm NMR tube and analyzed by NMR spectrosco
## [35] "1 H NMR spectra of these samples were acquired using the standard NOESYPR 1 D pulse sequence on
## [36] "For each sample, 128"
## [37] "FIDs were collected into 40 K data points over a spectral width of 10 000"
## [38] "Hz with a relaxation delay of 0. 23 ms."
## [39] "The polar metabolites in the rat tissue were extracted according to the protocol established by
## [40] "In brief, pre-weighed brain, kidney, liver, lung, or spleen samples (100 mg per sample) were ho
## [41] "The homogenates were transferred into a 2. 5 ml tube, and combined with 400 µl of CHCl 3 and 20
## [42] "After 10 min partitioning on ice, the samples were centrifuged for 5 min (10 000 g, 4 ◦C)."
## [43] "The upper supernatants were transferred into 1. 5 ml tubes, and lyophilized to remove CH 3 OH a
## [44] "The extracts were reconstituted in 0. 5 ml D 2 O containing 1 mM TSP, then transferred into 5 m
## [45] "1 H NMR spectra of these samples were acquired on a Bruker-AV 600 spectrometer at 296 K. Standa
## [46] "For each sample, 64 free induction decays (FIDs) were collected into 64 k data points over a sp
## [47] "Hz with a relaxation delay of 6."
## [48] "5 µs and an acquisition time of 2. 66 s. 2. 6."
## [49] "Spectral processing and pattern recognition"
## [50] "The FIDs were multiplied by an exponential function corresponding to a 1 Hz line-broadening fac
## [51] "The acquired NMR spectra were manually phased and baseline-corrected using Bruker TOPSPIN 3. 0
## [52] "The peaks observed were assigned on the basis of their chemical shifts and signal multiplicity
## [53] "Each 1 H NMR spectrum of plasma, urine, or aqueous tissue extract was segmented into regions of
## [54] "The segments of  6. 0-5. 5 and  5. 20-4. 29 in the plasma spectra (  9."
## [55] "0-0. 5), and of  6. 02-5."
## [56] "45 and  5. 35-4."
## [57] "24 in the urine spectra (  9."
## [58] "5-0. 5) were removed to exclude the urea signal and the uncertainty of residual water signal."
## [59] "In the case of the chloroform/methanol tissue extract, the regions of  5. 226-
4."
## [60] "675 and  3. 40-3."
## [61] "31 in the spectra were excluded to remove variation in residual water and methanol signal."
## [62] "Normalization was applied to the total sum of integrated data from each sample, which makes the
## [63] "The mean centered data were analyzed by principal component analysis (PCA) to reveal trends, hi
```

```
## [64] "0 (Umetrics, Sweden)."
## [65] "Then, the orthogonal projection to latent structure with discriminant analysis (OPLS-DA) metho
## [66] "Score plots provide the most efficient 2 D representation of the information contained in the
## [67] "Here, a correlation coefficient (determined by the Pearson's product-moment correlation coeffi
## [68] "The data were expressed as mean± SD."
## [69] "Inter-group variations were assessed by one-way analysis of variance (ANOVA) followed by a pos
## [70] "The criterion for statistical significance was p < 0. 05."
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

Following article have a material and method section starting by EXPERIMENTAL.

```
pdf_name <- "Li, J G et al 2009.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "EXPERIMENTAL"
word_end_mm <- "microscopy"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 493
```

```
## [1] "EXPERIMENTAL DETAILS 2. 1. MWCNTs and Animals The MWCNTs were obtained from Shenzhen Nanotech P
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1] 1212
```

```
## [1] "Sections were stained with hematoxylin and eosin, and then examined by light microscopy."
```

```
mm_section<-cut_conllu_df(493, 1212, conllu_df)
```

```
##  [1] "EXPERIMENTAL DETAILS 2. 1. MWCNTs and Animals The MWCNTs were obtained from Shenzhen Nanotech
##  [2] "2 wt%, average special surface area of 280 m 2/g, and amorphous carbon <3%."
##  [3] "The pristine MWCNTs were used directly in this experiment without any physicochemical processes
##  [4] "The experimental animals, female Kunming mice with weight of 30 g, were purchased from Shangha
##  [5] "All mice were housed in clean cages in a ventilated animal room."
##  [6] "Room temperature was maintained at 18-20 C, relative humidity at 55± 10%, and illumination on
##  [7] "The mice were supplied with sterilized food and pure water, and they were allowed to acclimate
##  [8] "The study was carried out in compliance with the national regulations related to the conduct o
##  [9] "2. 2. Inhalation Equipment and Exposure"
## [10] "The inhalation equipment was mainly made up of powder generator, first depositor, second depos
## [11] "Nanotechnol."
## [12] "2009, Vol. 9, No. 2 1533-4880/2009/9/1384/004 doi:10."
## [13] "1166/jnn."
## [14] "2009. C 162 Delivered by Ingenta to: Purdue University Libraries IP: 146."
## [15] "185."
## [16] "202."
## [17] "43 On: Tue, 07 Jun 2016 01:12:05 Copyright: American Scientific Publishers R E S E A R C H A R
## [18] "The Pulmonary Toxicity of Multi-Wall Carbon Nanotubes in Mice 30 and 60 Days After Inhalation
## [19] "Eighteen mice were exposed to MWCNTs aerosol in inhalation chamber, once in very two days, 6 h
## [20] "According to exposure period, the mice were divided into 2 groups, nine in each group, and den
## [21] "Their actual exposure time was 15 days in 30-day group and 30 days in 60-day group."
## [22] "The inhalation experiment was conducted in ventilated hoods at room temperature (18-
20 C)."
## [23] "In addition, nine mice housed in cages of the same animal room were denominated as control gro
## [24] "2."
## [25] "3. Biochemical Indices in Bronchoalveolar Lavage Fluid (BALF)"
```

```
## [26] "After inhalation exposure, the mice in control and experimental groups were anaesthetized by i
## [27] "Six mice were randomly chosen from each group, then whose thoraxes were opened and bronchoalve
## [28] "About 1. 5 ml BALF was pooled, and stored at 4 C. The BALF were centrifuged (400 g, 10 min, 4 
## [29] "Total protein, alkaline phosphatase (ALP), acid phosphatase (ACP) and lactate dehydrogenase (L
## [30] "2. 4."
## [31] "Statistical Analysis Statistical analysis was performed using one-way analysis of variance (AN
## [32] "For analysis, each of the experimental values was compared with that in control group."
## [33] "Data were expressed as means ± SD."
## [34] "Significance was judged at P < 0 05 and P < 0 01 probability level."
## [35] "2. 5."
## [36] "Pathological Examination"
## [37] "The remained 3 mice in each group were intraperitoneally anesthetized with 0. 3 ml 0. 5% pento
## [38] "The formalin-fixed lungs were embedded in paraffin, thin-sectioned, and mounted on glass micro
## [39] "Sections were stained with hematoxylin and eosin, and then examined by light microscopy."
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Li, Y et al 2010.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "MATERIALS"
word_end_mm <- "microscopy"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 895
```

```
## [1] "MATERIALS AND METHODS 2. 1. TiO 2 NPs"
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1] 1475
```

```
## [1] "Sections were stained with hematoxylin-eosin and examined by light microscopy."
```

```
mm_section<-cut_conllu_df(895, 1475, conllu_df)
```

```
##  [1] "MATERIALS AND METHODS 2. 1. TiO 2 NPs"
##  [2] "The synthesis and characterization of 3 nm TiO 2 NPs in this study were performed as described
##  [3] "11 2. 2. Animals"
##  [4] "The animals used in this work were 7-week-old male Kunming mice (The Slack Experimental Animal
##  [5] "The animal facility was maintained under a 12-h light-dark cycle at a temperature of 21-
23 C, and a relative humidity of 40-60%."
##  [6] "All mice were supplied with sterilized food and water."
##  [7] "They were acclimated for 7 days prior to instillation."
##  [8] "The average animal weight was 28 g at the beginning of the study."
##  [9] "The animal program is fully accredited by the national related regulations."
## [10] "2. 3. Intratracheal Instillation The TiO 2 NPs, in stable colloidal state, were diluted into 1
## [11] "The suspension was ultrasonicated for 15 min before intratracheal instillation."
## [12] "Thirty-nine mice were randomly divided into three groups (control group, Millipore water group
## [13] "All the mice were intraperitoneally anesthetized with 0. 2 ml 0. 5% pentobarbital sodium solut
## [14] "Then the mice were intratracheally instilled with 0. 1 ml Millipore ultrapure water and 0. 1 m
## [15] "TiO 2 NPs suspension, respectively."
## [16] "The procedure was repeated once a week for consecutive four weeks."
## [17] "2. 4. Evaluation of Biochemical Parameters in BALF"
## [18] "Five mice randomly chosen from each group were anaesthetized by intraperitoneal injection of 0
## [19] "Bronchoalveolar lavage was performed on them by cannulating the trachea and lung lavaging two 
```

```
## [20] "The BALF were collected and centrifuged (400× g, 10 min, 4 C), and biochemical parameters (tota
## [21] "Five mice randomly selected from other eight mice in each group were anaesthetized with 0. 3 ml
## [22] "Brain lavage was performed as follow: the right atrium of mice was incised and lavage fluid (3
## [23] "The perfusion pressure was about 100 mmHg."
## [24] "After perfusion, brains were excised and homogenated."
## [25] "The brain homogenates were centrifuged (10000× g, 10 min, 4 C), and then superoxide anion, hydr
## [26] "2. 6."
## [27] "Histopathological Examination"
## [28] "The lungs, brains, livers, kidneys, and spleens of the remained mice in each group were excised
## [29] "Sections were stained with hematoxylin-eosin and examined by light microscopy."
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Liu, Y et al 2014.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "Materials"
word_end_mm <- "significant"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 1314
```

```
## [1] "2. Materials and methods 2."
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1]  671 3507 3601 5362 5568 5589 6514 8577 8840
```

```
## [1] "Although information about the toxicity of IONPs continues to increase, a significant knowledge
## [2] "The result was judged positive if a statistically significant increase in the frequency of MNPCE
## [3] "A value of P < 0. 05 was considered statistically significant."
## [4] "The results were judged positive if reproducible and significant increases in the frequencies of
## [5] "The result was judged positive if a statistically significant increase in the frequency of MNPCE
## [6] "The result was judged positive if a statistically significant increase in the frequency of MNPCE
## [7] "For example, Fe and Fe 2 O 3 nanoparticles are known to be toxic and can cause significant DNA d
## [8] "For example, Saxena et al [ 49 ] showed that acid-functionalized SWCNTs caused markedly signifi
## [9] "They found that the highly negatively charged QDs with carboxyl groups induced DNA damage after
```

```
mm_section<-cut_conllu_df(1314, 3601, conllu_df)
```

```
##  [1] "2. Materials and methods 2."
##  [2] "1."
##  [3] "IONPs"
##  [4] "IONPs with neutral non-functional PEG coating and different particle size (10 nm, SMG-10; 30 nm
##  [5] "Their characteristics as reported by the manufacturer are:"
##  [6] "SEI-10, solution in DI water, 5-10 pH stability, 1 mg ml-1 (Fe); SMG-10 and SMG 30, solution in
10 pH stability, 1 mg ml-1 (Fe)."
##  [7] "The morphology and particle size of IONPs were observed on a Philips"
##  [8] "CM-120 transmission electron microscope (TEM) operating at an acceleration voltage of 80 kV."
##  [9] "IONPs were homogeneously dispersed in water, and one drop of the suspension was deposited on a
## [10] "The particle size distribution and zeta potential of IONPs were measured by dynamic light scatt
## [11] "2. 2. Sample preparation The SMG-10, SMG-30 and SEI-10 were aqueous solution."
## [12] "In each experiment, they were diluted in sterile physiological saline."
## [13] "2. 3. Preparation of mammalian liver S 9 fraction"
## [14] "The S 9 fraction was prepared from the livers of male Sprague-Dawley rats treated with the comb
## [15] "The S 9 metabolic activator was prepared immediately before use, by adding: 0. 1 ml of S 9 frac
```

```
## [16] "Thus, 1 ml of S 9 mix contained 0. 1 ml S 9 (10%), 0. 2 M of phosphate buffer (pH 7. 4), 8  M c
## [17] "The S 9 mixture was kept on ice during testing."
## [18] "2. 4."
## [19] "Ames test"
## [20] "The experiments were conducted according to the OECD Guideline for Testing of Chemicals 471 'Ba
## [21] "typhimurium TA 98 (hisG 3052 rfa ΔuvrB pKM 101), S. typhimurium TA 100 (hisG 46 rfa ΔuvrB pKM 1
## [22] "Typhimurium TA 102 (hisG 428 rfa pKM 101 pAQ 1), were supplied by Molecular Toxicology."
## [23] "The strain genotypes were confirmed by testing the presence of specific genetic markers and phe
## [24] "The tester strains used in each experiment were freshly prepared by preculturing them into the
16 h) at 37 °C."
## [25] "2 Nanotechnology 25 (2014) 425101"
## [26] "Y Liu et al"
## [27] "Top agar was prepared by dissolving 6 g of agar (Kelong Chemical, Chengdu, China) and 5 g of Na
15 min at 121 °C using an autoclave."
## [28] "For the culture of S. typhimurium, a mixed aqueous solution of L-Hsitidine (0. 5 mM) and Dbioti
## [29] "Plate incorporation assay was applied to detect reverse mutation in the presence and absence of
## [30] "For the direct method (no metabolic activation), 0. 5 ml of purified water and 0. 1 ml of each
## [31] "For the metabolic activation method, 0. 5 ml of the S 9 mix and 0. 1 ml of each bacterial suspe
## [32] "The mixtures were poured into plates and incubated at 37 °C for 48 h."
## [33] "The number of revertant colonies was counted using an automatic colony counter (Hangzhou Shines
## [34] "The maximum concentration of test substances were 1 mgml-1, the volume of administration was 10
## [35] "Based on our preliminary data, 100  g per plate was set as the highest concentration for SEI-10
## [36] "Five-fold intervals were set for the descending dose levels."
## [37] "The toxicity of the test materials was evaluated as reduction in the number of revertant colon
## [38] "A positive response in the test was defined as an increase (at least two-fold above the control
## [39] "In vitro mammalian chromosome aberration test"
## [40] "The experiments were conducted according to the OECD Guideline for Testing of Chemicals 473, 'I
## [41] "Chinese hamster lung (CHL) cells were obtained from Shanghai Institutes for Biological Sciences
## [42] "The CHL cells were cultured in MEM medium (Hyclone) containing 10% (v/v) heat-inactivated Calf
3 days."
## [43] "Serum was obtained from GIBCO and antibiotics were from North China Pharmaceutical Group Corpor
## [44] "The vehicle control used was physiological saline."
## [45] "The positive control substances used were 0. 25  g ml-1 mitomycin C (MMC) for the exposure with
## [46] "The test was performed in the presence and absence of metabolic activation (S 9 mix)."
## [47] "The final concentration of S 9 in the treatment medium was 10% (v/v)."
## [48] "The study consisted of short-term exposure (4 h, with and without metabolic activation) and con
## [49] "In both cases, about 5 × 104 cells were seeded on each tissue culture flasks (50 ml) and incuba
48 h."
## [50] "For short-term exposure, the test substances were administrated for 4 h followed by a recovery
## [51] "For the metabolic activation, the cells were treated with the S 9 mix (0. 5 ml) together with t
## [52] "For continuous exposure, the test substances were kept in culture for 24 h."
## [53] "All the cells were sampled at 24 h after the beginning of treatment."
## [54] "At 4 h before sampling, they were treated with colchicine (Sigma; 0. 4  g ml-1 final concentrat
## [55] "Two hundred metaphases per dose were examined for structural aberrations and polyploidy."
## [56] "A cell growth inhibition test was also concomitantly conducted under the same conditions."
## [57] "The maximum concentration of test substances were 1 mg ml-1, the volume of administration was 0
## [58] "In the preliminary test, SMG-10 and SMG-30 administered at the maximum final concentration of 1
## [59] "In the presence of cytotoxicity, the top concentration selected should be one that caused 50% c
## [60] "According to the results of the preliminary test, the concentrations of SEI-10 selected for the
## [61] "The concentrations of SMG-10 and SMG-30 selected for the main test were 2. 5, 5, 10, 20  g ml-1
## [62] "Duplicate plates were examined at each dose."
## [63] "The clastogenic potential was judged as negative, suspicious and positive if the incidence of c
10% and >10%, respectively [ 17 ]. 2. 6."
## [64] "In vivo mammalian erythrocyte micronucleus test"
```

```
## [65] "The experiments were conducted according to the OECD Guideline for Testing of Chemicals 474 'Ma
## [66] "Adult male NIH mice (5-6 weeks old, 25-30 g body weight) obtained from the breeding center of (
## [67] "The mice were housed in clean polypropylene cages and maintained in an air-conditioned conventi
70% relative humidity and 12 h light/dark cycle (light on at 07:30)."
## [68] "The animals were provided with commercial rat pellet (Shanghai SLAC Laboratory Animal) and tap
## [69] "After 5 days acclimatization, the mice were randomly assigned to the control and treatment grou
## [70] "Y Liu et al control (physiological saline), positive control (CPA, 40 mg kg-1), SEI-10, SMG-10
## [71] "Animals were administered with negative control (physiological saline), SEI-10, SMG-10 and SMG-
## [72] "The injected volume was 10 ml kg-1 body weight."
## [73] "After dosing, the mice were examined regularly for mortality and clinical signs of toxicity unt
## [74] "The femurs of mice in all groups were removed 24 h after the final administration, and the bone
## [75] "The specimens were air-dried, fixed with methanol and stained with Giemsa (1 ml of the dye was
## [76] "The animals received humane care and the study protocol was approved by Institutional Animal Ca
## [77] "A preliminary test was carried out for maximal tolerance in two male via intravenous injection
## [78] "The maximum concentrations of test substances were 1 mg ml-1."
## [79] "The volume of administration was 10 ml kg-1; hence, the highest dose corresponded to 10 mg kg-1
## [80] "No death was found and no abnormalities were detected in mice treated with SMG-10 and SMG-30 at
10 mg kg-1."
## [81] "However, SEI-10 caused animal death at 2. 5 mg kg-1 or higher doses."
## [82] "Therefore, in the in vivo micronucleus test, the doses set for SMG-10 and SMG-30 were 2. 5 mg k
## [83] "The frequency of micronucleated polychromatic erythrocytes (MNPCE) was calculated on the basis
## [84] "Furthermore, the proportion of PCE among total erythrocytes (TE) was determined by counting 200
## [85] "The result was judged positive if a statistically significant increase in the frequency of MNPC
## [86] "2. 7. Statistical analysis"
## [87] "All values were expressed as mean ± SD in Ames test and micronucleus test."
## [88] "Results were tested for significance with one-way analysis of variance (ANOVA)."
## [89] "The percentage of cells with aberrations was presented with contingency table and evaluated usi
## [90] "A value of P < 0. 05 was considered statistically significant."
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

Love, S A et al 2012.pdf : no real material and method section

```
pdf_name <- "Love, S A et al 2012.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "Nanoparticle"
word_end_mm <- "USA"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 1272 2626 3028 3385 3973 5327 7570
```

```
## [1] "n Nanoparticle synthesis"
## [2] "Nanoparticle Diameter (by TEM, nm) LSPR (lmax, nm) z potential (mV) Seed Au(+) Au(-) 26. 5 ± 4.
## [3] "n Nanoparticle synthesis & characterization Spherical Au nanoparticles of approximately 30-nm di
## [4] "Exposure time (h) 100 80 60 40 20 0 100 80 60 40 20 0 100 80 60 40 20 0 24 48 72 24 48 72 24 48
## [5] "Nanoparticle exposure Figure 3. Rate of reactive oxygen species generation was not significantly
## [6] "Time (s) [Nanoparticle] (μg/ml) 0 50 100 0 60 120 180 240 4 3 2 1 0 5 15 25 50 15 μg/ml Seed Au
## [7] "Nanoparticle size and surface properties determine the protein corona with possible implications
```

```
display_occurences(word_end_mm, conllu_df)
```

```
##  [1]  766 1221 1247 1268 1347 1412 1674 1726 1774 1801 1843 2296 2542 2569 2769
## [16] 3021 6420 7593
```

```
##  [1] "1 Department of Chemistry, University of Minnesota, Kolthoff & Smith Halls, 207 Pleasant Stree
```

```
##  [2] "Unless otherwise stated, chemicals utilized were purchased from Sigma Aldrich (WI, USA) and use
##  [3] "All blood was drawn and collected in heparin sodium vacuum tubes at Memorial Blood Center (MN,
##  [4] "Water used in synthesis was purified using a Milli-Q® Integral Water Purification System (Milli
##  [5] "First, the particles designated 'seed Au' were prepared in the following manner: 240 µl of appr
##  [6] "First, the particles designated 'seed Au' were prepared in the following manner: 240 µl of appr
##  [7] "The synthesized nanoparticles were characterized by transmission electron microscopy (TEM) afte
##  [8] "Size was determined using ImageJ  software (NIH, MD, USA) based on the average of at least 200
##  [9] "Hydrodynamic diameter, both in the original solution (water) and 10% bovine calf serum, was mea
## [10] "Extinction measurements were conducted using an Ocean Optics® USB 2000 spectrometer (FL, USA) a
## [11] "All nanoparticle concentrations were assessed by inductively coupled plasma atomic emission spe
## [12] "Briefly, heparinized blood was layered over Ficoll Plaque Plus© (GE Healthcare, NJ, USA) and al
## [13] "Emission readings (l excitation = 485 nm, l emission = 528 nm) were taken on a Biotek Synergy 2
## [14] "Linear regression analysis (Microsoft Excel®, WA, USA) was then performed to determine the rate
## [15] "Review AuthorseaRch aRticle Love, Thompson & Haynes n Platelet aggregation Platelets were ise
## [16] "All trials were allowed to proceed for 5 min or until 100% aggregation had been achieved; for a
## [17] "Lux Research, Inc., NY, USA (2008)."
## [18] "USA 105(38), 14265-14270 (2008)."
```

```
mm_section<-cut_conllu_df(1272, 3021, conllu_df)
```

```
##  [1] "n Nanoparticle synthesis"
##  [2] "All nanoparticles were synthesized using previously published methods [ 2 ]."
##  [3] "All glassware used was cleaned with aqua regia (3:1 concentrated HCl:HNO 3 ) and thoroughly rin
##  [4] "First, the particles designated 'seed Au' were prepared in the following manner: 240 µl of appr
##  [5] "Second, the particles designated 'Au(+)' and 'Au(-)' were synthesized by adjusting the seed Au
##  [6] "This solution was then stirred overnight (~12-18 h) and centrifuged at 28, 627 × g for 15 min."
##  [7] "After centrifugation, the nanoparticles were washed with water and 100 µl of either 10 mM 11-me
##  [8] "After addition of 11-mercaptoundecanoic acid, the solution was again allowed to stir overnight
18 h); however, 1 h after the addition of 11-mercaptoundecylamine, 100 µl of 1 M"
##  [9] "HCl was added per ml of solution and then allowed to stir overnight."
## [10] "To purify the nanoparticles after synthesis and ligand exchange, both nanoparticle solutions we
## [11] "The final pellet was resuspended in H 2 O and the nanoparticles were characterized using a vari
## [12] "n Characterization"
## [13] "The synthesized nanoparticles were characterized by transmission electron microscopy (TEM) afte
## [14] "futuremedicine."
## [15] "com 1357 future science group Development of screening assays with charged Au NPs for NP toxici
## [16] "Size was determined using ImageJ  software (NIH, MD, USA) based on the average of at least 200
## [17] "Hydrodynamic diameter, both in the original solution (water) and 10% bovine calf serum, was mea
## [18] "Extinction measurements were conducted using an Ocean Optics® USB 2000 spectrometer (FL, USA) a
## [19] "All nanoparticle concentrations were assessed by inductively coupled plasma atomic emission spe
## [20] "This dosing range was chosen to represent the range from lower/unintentional (i. e., release in
## [21] "The exposure time frames were chosen to focus on short-term (i. e., 5 min, depending on cell li
## [22] "All nanoparticles were stored in the dark or under minimal light exposure prior to and during e
## [23] "n Hemolysis"
## [24] "RBCs were isolated according to a modified version of the method employed by Lin and Haynes [ 1
## [25] "Briefly, a 5 ml blood sample was mixed with 10 ml of phosphate-buffered saline (PBS) lacking ca
## [26] "The supernatant was drawn off and five subsequent washes were carried out before diluting the s
## [27] "The RBCs were then diluted 1:5 into:"
## [28] "PBSd (negative control); water (positive control); or a nanoparticle solution (in PBSd) at vary
## [29] "The negative control was composed of PBSd with water at the same volume as the largest volume e
## [30] "The samples were left at room temperature in the dark for 24 h and then centrifuged at 10, 016
## [31] "After centrifuging, the supernatant was transferred to a 96-well plate and the absorbance at 57
## [32] "Additional assessments were made over 72 h of exposure, at 24-h intervals."
## [33] "Percent hemolysis was calculated using Equation 1: (% )Hemolysis of positive control Abs Abs"
## [34] "Abs"
```

```
## [35] "Abs 100 pos neg sample neg = - -c m (1) ROS generation Neutrophils were isolated from human bl
## [36] "[ 27 ]."
## [37] "Briefly, heparinized blood was layered over Ficoll Plaque Plus© (GE Healthcare, NJ, USA) and a
## [38] "The leukocyte-rich supernatant was then drawn off, layered over Ficoll Plaque Plus, and centri
## [39] "The neutrophil-containing pellet was then washed three-times with PBS containing glucose but l
## [40] "ROS generation was then measured using a DCFDA-based spectroflurometric assay."
## [41] "For this assay, 80 µl of the neutrophil solution was transferred to a 96-well plate where it w
## [42] "The negative control was prepared in the same manner but replacing the nanoparticle solution w
## [43] "Two different positive controls were prepared: one exposed to UV light (at 365 nm with an irra
## [44] "5 J m-2) for the duration of the exposure, and the other was exposed to 1 unit/ml of thrombin
## [45] "Each condition was carried out in eight replicates and was performed with at least three indep
## [46] "Emission readings (l excitation = 485 nm, l emission = 528 nm) were taken on a Biotek Synergy
## [47] "Linear regression analysis (Microsoft Excel®, WA, USA) was then performed to determine the rat
## [48] "These rates were then compared using the Student's t-test to assess significant differences fr
## [49] "Nanomedicine (2012) 7(9)1358 future science group Seed Au(+) Au(-)"
## [50] "Nanoparticle Diameter (by TEM, nm) LSPR (lmax, nm) z potential (mV) Seed Au(+) Au(-) 26. 5 ± 4
## [51] "2 ± 4. 7 26. 9 ± 7. 6 525."
## [52] "4 529. 6 529. 6 -41. 0 ± 1. 2 41. 9 ± 1. 3 -39."
## [53] "4 ± 1. 8 20 nm 20 nm 20 nm Figure 1."
## [54] "Representative transmission electron microscopy images of seed, Au(+) and Au(-) nanoparticles
## [55] "LSPR: Localized surface plasmon resonance; TEM: Transmission electron microscopy."
## [56] "Review AuthorseseaRch aRticle Love, Thompson & Haynes n Platelet aggregation Platelets were is
## [57] "Briefly, whole blood was centrifuged for 15 min at 100 × g. The platelet-rich plasma (PRP) was
## [58] "The PRP needed for individual trials was placed in a heating block warmed to a temperature of
## [59] "A small volume of PRP was then centrifuged at 2400 × g for 20 min to obtain platelet-poor plas
## [60] "After centrifugation, the PPP was warmed to 37°C and placed in the blank channel of the aggreg
## [61] "Once the temperature of the PRP for each trial had reached 37°C, it was placed into the aggreg
## [62] "The following trials were then performed while stirring vigorously (1000 rpm): 100 µl of PBSg
## [63] "All trials were allowed to proceed for 5 min or until 100% aggregation had been achieved; for
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Maurizi, L et al 2015.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "MATERIALS"
word_end_mm <- "t-test"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 1311
```

```
## [1] "MATERIALS AND METHODS Materials Reagents Iron (II) chloride tetrahydrate (FeCl 2 · 4 H 2 O), ir
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1] 3877
```

```
## [1] "All statistics were performed using the GraphPad Prism software and an unpaired t-test."
```

```
mm_section<-cut_conllu_df(1311, 3877, conllu_df)
```

```
##   [1] "MATERIALS AND METHODS Materials Reagents Iron (II) chloride tetrahydrate (FeCl 2 · 4 H 2 O),
##   [2] "THF was distilled over sodium metal and benzophenone just before use."
##   [3] "3-(4, 5-Dimethylthiazol-2-yl)-2, 5 diphenyltetrazolium bromide (MTT) reagent and Prussian blu
##   [4] "Control Iron Oxide Suspension"
##   [5] "Used as an MRI Contrast Agent Resovist® from Schering is a negatively-charged carboxydextran
```

```
##   [6] "This commercially available MRI T 2 contrast agent was used as a reference for this study. 8 S
##   [7] "The naked SPION suspension formed was then washed and peptized several times with 1 M HNO 3 so
##   [8] "The suspension was then dialyzed at pH 3 with a HNO 3 solution for 2 days, and was subsequentl
##   [9] "The particle concentration was 23 mg ·mL−1."
##  [10] "X-ray Diffraction (XRD) patterns of the nanoparticles were characteristic of the magnetite spi
##  [11] "PEG 2 k-Si was obtained by reacting ICPTS with PEG 2 k 43 in anyhdrous THF at 60 C under nitro
##  [12] "The PEG 2 k-Si was then precipitated in hexane."
##  [13] "Pegylated SPIONs were obtained as follows: 0. 13 mmol of PEG 2 k-Si was dissolved in 40 mL etl
45 One hundred milligrams of SPIONs were added and the reaction proceeded for 48 h under N 2 flow."
##  [14] "The suspension was then dialyzed with water for one week."
##  [15] "The resultant nanoparticles were called PEG 2 k-SPIONs."
##  [16] "PEG 2 k-SPIONs were also grafted with DMSA."
##  [17] "40 Pegylated particles were obtained as described above, except for the addition of 0. 0439 mm
##  [18] "The modified SPIONs were then sedimented 3 times at pH 3 in the presence of a magnetic field,
##  [19] "J. Biomed."
##  [20] "Nanotechnol. 11, 126-136, 2015 127"
##  [21] "Delivered by Publishing Technology to: SUNY"
##  [22] "Upstate Medical University IP: 213."
##  [23] "199."
##  [24] "30."
##  [25] "167 On: Wed, 14 Oct 2015 16:06:03 Copyright: American Scientific Publishers"
##  [26] "Influence of Surface Charge and Polymer Coating on Internalization and Biodistribution of PEG
##  [27] "Figure 1. (a) Representation of the three types of synthesized SPION (shown in orange)."
##  [28] "From left to right: naked SPION, SPION coated with PEG 2 k (shown in grey), and SPION coated w
##  [29] "Naked SPIONs, PEG 2 k-SPIONs and DMSA+ PEG 2 kSPIONs are schematically represented in Figure 1
##  [30] "Cell Culture Macrophage cell types from mice with leukemia (RAW 264. 7) and HepG 2 hepatocytes
##  [31] "Animal Study 3 month old Swiss Albino mice (30 g) were purchased from the animal care unit of
##  [32] "Physicochemical Characterization"
##  [33] "The zeta potential ( ) and the hydrodynamic size (dH ) of the nanoparticles were measured with
##  [34] "The nanoparticle suspensions were diluted to approximately 200 g ·mL−1 in 10−2 M NaCl solution
##  [35] "The NaCl stock solution had previously been filtered through a 0. 8 m filter."
##  [36] "The potentials of functionalized and naked SPION suspensions were measured from pH 3 to pH 12
##  [37] "HCl and 0. 1 M NaOH, respectively."
##  [38] "TEM observations were performed with a JEOL JEM 2100 instrument operating at 200 kV."
##  [39] "The copper grids were dipped in a dilute suspension of samples and left to dry completely befo
##  [40] "The mean size of the crystallites, dTEM, was obtained from at least one hundred particles."
##  [41] "Specific surface area measurement of naked SPIONs was performed using a 5 isotherm point BET
Emett-Teller) method with a BELSORP-mini apparatus with N 2 gas adsorption."
##  [42] "Thermogravimetric analyses (TGA) were performed with a symmetrical thermobalance (SETARAM TAG
##  [43] "The heating rate was 2 C ·min−1 up to 800 C under N 2/O 2 (0. 12/0."
##  [44] "04 L ·min−1)."
##  [45] "Sample weights were around 5 to 10 mg."
##  [46] "Any weight loss was assumed to be due to polymer degradation."
##  [47] "The amount of DMSA on the particles was measured with Ellman's reagent assay titration."
##  [48] "40 46"
##  [49] "The particle relaxivities (r 2 in mM −1 Fe · s−1) were calculated using the following method."
##  [50] "47 Water proton transverse relaxation time (T 2) measurements at 3. 0 T were carried out at 29
##  [51] "Relaxation time measurements were performed on test tubes containing SPIONs, with or without s
##  [52] "For T 2 determination, a multi-echo spin-echo (SE) pulse sequence was used, with TR (repletion
##  [53] "Images were acquired at 32 echoes, from TE = 8 2 ms to 275."
##  [54] "2 ms, with an 8. 2 ms interval."
##  [55] "Image analysis was performed using ImageJ (Image Analysis Software developed by the NIH, USA)
##  [56] "The signal decay curve was fitted to the equation S TE = A exp −TE/T 2 using a non-linear func
##  [57] "The relaxivity, r 2 (in mM −1 Fe · s−1), was then determined by fitting the relaxation rate (1
```

```
##  [58] "Biological Observations Cell Viability"
##  [59] "The MTT assay was used to measure the cytotoxicity of SPIONs in two different cell lines (RAW
##  [60] "This measured mitochondrial metabolic activity, which converted the MTT reagent from a yellow
##  [61] "Formazan crystals were then dissolved in an organic solvent and quantified by measurement of
##  [62] "To determine cell viability, the cells were seeded in a 96-well plate and incubated at 37 C in
##  [63] "Cells were then incubated in the presence of the different types of SPIONs or Resovist® (at tw
##  [64] "These two concentrations were chosen on the basis of the dose injected in the in vivo study."
##  [65] "Briefly, 5. 6 gFe · g−1 mouse were injected (30 g mice, 128 J. Biomed."
##  [66] "Nanotechnol."
##  [67] "11, 126-136, 2015"
##  [68] "Maurizi et al."
##  [69] "Influence of Surface Charge and Polymer Coating on Internalization and Biodistribution of PEG
##  [70] "In 96-well plates, the working volume was 100 L so we selected 8. 4 gFe/well (about 25 gFe · c
##  [71] "This concentration was then doubled for high-dose cytotoxicity assessment."
##  [72] "Previous studies have employed these doses to investigate both MRI biodistribution and cytotox
##  [73] "Then 100 L MTT (2 mg ·mL−1) was added in Puck G+ medium (126."
##  [74] "6 mM NaCl, 3. 8 mM KCl, 0. 11 mM CaCl 2 ·2 H 2 O, 0. 62 mM MgSO 4 ·7"
##  [75] "H 2 O, 0. 86 mM"
##  [76] "Na 2 HPO 4 ·2 H 2 O, 0. 61 mM KH 2 PO 4, 14."
##  [77] "3 mM NaHCO 3, and 1. 1 g ·L−1 glucose, pH 7. 4) for 1 h at 37 C."
##  [78] "Puck G+ corresponded to basic osmotic medium and was used to avoid interference during the MTT
##  [79] "Excess MTT was then removed and 200 L of isopropanol with HCl (0. 1 M) was incubated with the
##  [80] "The absorbance in each well was then read in a spectrophotometer (Wallac Victor 3, Perkin Elme
##  [81] "Average absorbance (from n= 4) was normalized to the control."
##  [82] "The viability (in %) was then calculated by absorbancetest/absorbancecontrol×100."
##  [83] "SPION Interactions with Macrophages and Hepatocytes Optical microscopy: RAW 264. 7 and HepG 2
##  [84] "The cells were then washed 2 times with PBS and fixed with 4% paraformaldehyde, 0. 5% glutaral
##  [85] "Then the cells were fixed in sodium borohydride (25 mg ·mL−1 in PBS) for 10 min, washed once w
##  [86] "HCl for 10 min, and washed with water."
##  [87] "Cells were then stained with 0. 2% paroraxilin for 1 min, before washing consecutively with wa
##  [88] "The dissolved iron from the SPIONs reacted with potassium ferrocyanide to give a blue product
##  [89] "The cells were then mounted on glass slides using Eukitt® (Sigma) and observed by optical mic
##  [90] "RAW 264. 7 and HepG 2 cells incubated with complete media were observed as control."
##  [91] "Transmission Electron Microscopy: RAW 264. 7 cells were incubated with Resovist®, PEG 2 k-SPIO
##  [92] "Cells were then washed 3 times with PBS and fixed at room temperature with 4% paraformaldehyde
##  [93] "The cells were then washed 4 times in phosphate buffer and then one last time overnight at 4 C
##  [94] "Cells were then embedded in epoxy resin (EMBed-812) overnight at 37 C, followed by 48 h at 60
##  [95] "After cutting, the slices were stained with uranyl acetate and lead citrate solution for conti
##  [96] "Mice MRI Analyses"
##  [97] "Mice were injected with 100 L of physiological saline (NaCl at 0."
##  [98] "15 M) (n = 6), Resovist® (n = 6), PEG 2 k-SPIONs (n= 6), or DMSA+ PEG 2 k-SPIONs (n= 4)."
##  [99] "SPION suspensions were diluted in physiological saline to a concentration of 1. 68 mgFe ·mL−1
## [100] "To reach a concentration of 5. 6 mgFe · kg−1 mouse, the injected dose was around 168 gFe per m
## [101] "This experiment was conducted in accordance with French ethical guidelines for in vivo studies
## [102] "Injected mice were then analyzed by 3. 0 T MRI at 1, 3, and 24 h after injection."
## [103] "Gradient echo T *2 weighted was used for acquisition with echo-time TE = 2. 9 ms, repetition t
## [104] "On coronal images, the average contrast of the liver was calculated using Image J and reported
## [105] "Half of each mouse population was sacrificed 3 h after injection for organ analysis and the ot
## [106] "The livers and hearts of the sacrificed mice were harvested and sliced."
## [107] "These organs were then treated with Prussian blue to reveal the presence of iron from dissolve
## [108] "Optical microscopy was then performed on the slices and the images obtained from test and cont
## [109] "Statistics"
## [110] "All statistics were performed using the GraphPad Prism software and an unpaired t-test."
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

Nešković, O et al 2013.pdf : book chapter

```
pdf_name <- "Nešković, O et al 2013.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "Materials"
word_end_mm <- "ImageJ."

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 1231
```

```
## [1] "2 Materials 2. 1 Laboratory Equipment 2. 2 Solutions 318 Olivera Nešković et al."
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1] 3005
```

```
## [1] "At least 200 cells should be analyzed to evaluate the number of g -H 2 AX positive foci by using
```

```
mm_section<-cut_conllu_df(1231, 3005, conllu_df)
```

```
##   [1] "2 Materials 2. 1 Laboratory Equipment 2. 2 Solutions 318 Olivera Nešković et al."
##   [2] "4. Fixative solution:"
##   [3] "In 3× volume of methanol (CH 3 OH), add 1× volume of acetic acid (CH 3 COOH)."
##   [4] "5."
##   [5] "SØRENSEN'S phosphate buffer: (Stock solutions: (a) 0. 2 M Na"
##   [6] "2 HPO 4 and (b) 0."
##   [7] "2 M NaH 2 PO 4 )"
##   [8] "To prepare 100 ml of working buffer (0. 1 M, pH 6. 8), mix 24. 5 ml of (a) with 25. 5 ml of (
##   [9] "Dilute to 100 ml with ddH 2 O. 6. 10%"
##  [10] "Giemsa in SØRENSEN'S buffer:"
##  [11] "To prepare 100 ml, add 10 ml Giemsa stain to 90 ml of SØRENSEN'S buffer."
##  [12] "7. 1%"
##  [13] "HCl ethanol solution:"
##  [14] "To prepare 1 l, add 10 ml 37% HCl to 990 ml 95% ethanol. 8. 90% ethanol solution:"
##  [15] "To prepare 100 ml, add 10 ml ddH 2 O to 90 ml 100% ethanol. 9. 70% ethanol solution:"
##  [16] "To prepare 100 ml, add 30 ml ddH 2 O to 70 ml 100% ethanol."
##  [17] "1. Human dermal fi broblasts HDMEC (PromoCell GmbH, Heidelberg, Germany)."
##  [18] "2. T 25 tissue culture fl asks."
##  [19] "3."
##  [20] "Dulbecco's modi fi ed Eagle's medium (DMEM)."
##  [21] "4. Fetal bovine serum."
##  [22] "5. 0. 05% tripsin-EDTA. 6."
##  [23] "Polyprep glass slides. 7."
##  [24] "Disposable Petri dishes (100 mm). 8."
##  [25] "Lithium heparin BD Vacutainer ® tubes with BD Hemogard ® closure (Becton-Dickinson, Franklin
##  [26] "1. 1× phosphate-buffered saline (PBS):"
##  [27] "To prepare 1 l, add 8 g NaCl, 0. 2 g KCl, 1. 44 g Na 2 HPO 4 , and 0. 24 g KH 2 PO 4 to 800 m
##  [28] "Adjust pH to 7. 4. Adjust volume to 1 l with deionized H 2 O. Sterilize by autoclaving and st
##  [29] "Prepare fresh ( see Note 3 )."
##  [30] "3. 0. 2% Triton-X: Add 200 m l Triton-X (Sigma-Aldrich Co., Steinheim, Germany) to 100 ml dei
##  [31] "Store at 4°C."
##  [32] "4. 1× TBS-T (Tris-buffered saline Tween-20) buffer:"
```

```
##  [33] "To prepare 1 l, add 8. 8 g NaCl, 0. 2 g KCl, 3 g Tris-base, and 500 m l Tween 20-
800 ml ddH 2 O. Adjust pH to 7."
##  [34] "4. Adjust volume to 1 l with ddH 2 O."
##  [35] "Sterilize by autoclaving."
##  [36] "Store at 4°C ( see Note 4 )."
##  [37] "5. 0. 5% BSA: Dissolve 0. 05 g bovine serum albumin (BSA, Sigma-Aldrich) into 10 ml deionized
##  [38] "2. 3 Cell Culture Components 2. 4 Immuno fl uorescence ( g -H 2 AX Assay) Components 319 Geno
##  [39] "Primary antibody: Anti-phospho-H 2 AX (Ser 139) mouse monoclonal antibody (Upstate Cell Signal
##  [40] "Antimouse fl uorescein isothiocyanate (FITC) antibody, dilute in 0. 5% BSA (v/v = 1:400)."
##  [41] "8. 4, 6-Diamidino-2-phenylindole (DAPI)-containing antifade solution (Vector Laboratories Inc
##  [42] "Coverslips (22 × 50 × 0. 13 mm)."
##  [43] "10."
##  [44] "Clear, non fl uorescent nail varnish."
##  [45] "1. Collect fresh blood by venepuncture in 6 ml lithium heparin BD Vacutainer ® tubes with BD
##  [46] "Store blood at 4°C prior to procedure."
##  [47] "2. Aliquot PB-max karyotyping medium (Invitrogen-Gibco, Paisley, UK), 4. 5 ml in each 10 ml s
##  [48] "Add 0. 5 ml of whole blood."
##  [49] "Close the cap of the test tube and put into the incubator."
##  [50] "3. Keep cell culture in an incubator at 37°C."
##  [51] "One hour after the stimulation of cells, add the agent of interest to cultures: different volu
##  [52] "1. Add 0. 1 ml of the cytochalasin B solution after 44 h of culture, and then incubate for ne
##  [53] "Remove the supernatant."
##  [54] "4. Resuspend pellet on vortex mixer and add pre-warmed hypotonic solution."
##  [55] "5. Keep in water bath for 5 min at 37°C. 6."
##  [56] "Centrifuge at 500 × g for 10 min. 7. Remove supernatant up to 1 ml and add fi xative solution
##  [57] "Repeat steps 6 and 7 until the suspension is clear."
##  [58] "After last centrifugation, aspirate up to 0. 5 ml."
##  [59] "Resuspend pellet with Pasteur pipette and prepare slides ( 37 ) ."
##  [60] "3 Methods 3. 1 Cell CultureHuman Blood Cells 3. 2 Micronuclei Preparation: Micronucleus Assay
##  [61] "1."
##  [62] "Degrease slides with detergent, wash thoroughly with distilled water, and keep over night in
##  [63] "Prior to use, wash slides with distilled water and bi-distilled water."
##  [64] "2. Onto clean, dry slides, put three drops of the cell suspension."
##  [65] "3. Air-dry the slides."
##  [66] "4. Stain slides (in staining jar) with 10% Giemsa in SØRENSEN's buffer (pH 6. 8) for 10 min."
##  [67] "Score at least 1, 000 binuclear (BN) cells per sample."
##  [68] "Analyze slides with a microscope using magni fi cation 400× or 1, 000× when necessary."
##  [69] "2. Score a minimum of 1, 000 binucleated cells to evaluate the percentage of cells with one,
##  [70] "Calculate a cytokinesis-block proliferation index (CBPI) as follows:"
##  [71] "CBPI = MI + 2 MII + 3(MIII + MIV)/N, where MI-MIV represents the number of cells with one to
90% con fl uence, remove growth medium from the fl ask by aspiration."
##  [72] "Incubate the fl ask with 1 ml 0. 05% Trypsin-EDTA at 37°C for 4-7 min."
##  [73] "Examine the fl asks microscopically to make sure the cells begin to round."
##  [74] "The cells should detach from the fl ask surface after 7 min ( see Note 5 )."
##  [75] "4. Tighten cap and lightly tap the side of the fl ask to lift the remaining cells from the fl
##  [76] "Wash the sides of the fl ask with growth medium to inactivate the trypsin ( see Note 6 )."
##  [77] "Gently mix cells and medium."
##  [78] "Pipette the cell suspension up and down so as to obtain a suspension of individual cells."
##  [79] "Distribute 1 ml aliquots of the cell suspension to polyprep slides ( see Note 7 )."
##  [80] "2. Transfer polyprep slides into Petri dishes ( see Note 8 )."
##  [81] "3. Add appropriate dose of carbon nanotubes (0. 5-30 m l per/ml) into cell suspension seeded
##  [82] "Close the dish and incubate at 37°C in a humid atmosphere for the next 24 h ( see Note 9 )."
##  [83] "4. Wash slides in PBS for 5 min."
##  [84] "3. 3 Preparing Slides for Micronucleus Assay 3. 4 Slide Scoring for Micronucleus Assay 3. 5 C
```

```
##  [85] "Immuno fl uorescence ( g -H 2 AX Assay) 321 Genotoxic Assessment of Carbon Nanotubes 5. Fix c
##  [86] "Permeabilize cells in 0. 2% Triton-X at 4°C for 10 min. 7. Block reaction by transferring sli
##  [87] "While blocking, prepare primary antibody. 9."
##  [88] "Aspirate blocking solution and apply 100 m l diluted primary antibody to each slide."
##  [89] "Incubate for 1 h in a light-tight damp container."
##  [90] "Wash slides three times in TBS-T for 3 min each."
##  [91] "11."
##  [92] "Incubate slides with 100 m l secondary anti-goat antibody conjugated with Fluorescein isothio
##  [93] "12."
##  [94] "13."
##  [95] "Fix cells with 70, 90, and 100% ethanol, 5 min each, and airdry in the dark."
##  [96] "14."
##  [97] "Counter stain cells with 15 m l 4 ¢ ,6 ¢ -diamidino-2-phenylindole (DAPI)-containing antifade
##  [98] "Apply nail varnish to seal the samples."
##  [99] "For the best results, examine specimens immediately."
## [100] "For long-term storage (several weeks), store slides at 4°C protected from light."
## [101] "15."
## [102] "At least 200 cells should be analyzed to evaluate the number of g -H 2 AX positive foci by us
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

Neun, B W et al 2018.pdf : book chapter

```
pdf_name <- "Neun, B W et al 2018.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "Materials"
word_end_mm <- "Doxil"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 1283
```

```
## [1] "2 Materials"
```

```
display_occurences(word_end_mm, conllu_df)
```

```
##  [1]  658  872 1418 1456 1518 2095 2110 2969 2975 3022 3056 3095 3125 3570 3585
## [16] 3630 4907
```

```
##  [1] "For example, liposomal formulations of doxorubicin (Doxil) and amphotericin (Ambisome), when a
##  [2] "In the case of Doxil, the combination of three factors, including particle shape, PEG coating,
##  [3] "Plasma samples treated with negative control, positive (cobra venom factor), and Doxil were an
##  [4] "The mean responses from positive control (PC) and Doxil samples were divided by the mean respo
##  [5] "Doxil (Doxorubicin HCl, liposome, injection)."
##  [6] "PEGylated liposomal doxorubicin (Doxil) can also be used as a nanoparticle-relevant positive c
##  [7] "Doxil is doxorubicin formulated in nanoliposomes."
##  [8] "If a nanoparticle under study generated a positive response in any of the EIA assays, compare
##  [9] "Doxil is used in the clinic and is Box 1 Example calculation of nanoparticle concentrations fo
## [10] "Using Doxil helps to interpret results of this in vitro study for a test nanoparticle."
## [11] "If the degree of activation observed for the test nanoparticle is equal to or greater than tha
## [12] "If the degree of activation is lower than that of Doxil, complement activation should be consi
## [13] "If the degree of activation is lower than that of Doxil, complement activation should be consi
## [14] "4 Analysis of complement activation by Doxil in individual and pooled plasma."
## [15] "Plasma from ten donors was used to analyze Doxil."
## [16] "Shows results for controls and Doxil in individual plasma from 10 donors."
## [17] "Liposome-induced complement activation and related cardiopulmonary distress in pigs: factors p
```

```
mm_section<-cut_conllu_df(1283, 3125, conllu_df)
```

```
##   [1] "2 Materials"
##   [2] "1."
##   [3] "Sterile Ca 2+/Mg 2+-free phosphate buffered saline (PBS)."
##   [4] "2. Cobra venom factor as the positive control."
##   [5] "3. Veronal buffer."
##   [6] "4."
##   [7] "Pooled human plasma, anti-coagulated with sodium citrate."
##   [8] "5. iC 3 b EIA kit (e. g., MicroVue by Quidel Corp.)."
##   [9] "6. 1. 0 N HCl, as stop solution."
##  [10] "Stop solution is provided with each kit, but can also be prepared separately."
##  [11] "Dilute stock hydrochloric acid to a final concentration of 1. 0 N. Filter and store and room
##  [12] "Fig."
##  [13] "2 Comparison between C 3 a and iC 3 b EIAs."
##  [14] "Plasma samples treated with negative control, positive (cobra venom factor), and Doxil were ar
##  [15] "Arg and iC 3 b EIA."
##  [16] "Each sample was analyzed in duplicate (%CV < 10)."
##  [17] "The mean responses from positive control (PC) and Doxil samples were divided by the mean respo
##  [18] "The background in the C 3 a EIA was higher."
##  [19] "Due to the differences in the background, the S/N ratio of iC 3 b assay was higher 152 Barry W
##  [20] "Doxil (Doxorubicin HCl, liposome, injection)."
##  [21] "This is a prescription medication available from a licensed pharmacy."
##  [22] "It may not be available to some research laboratories."
##  [23] "8. Cremophor. 9. Taxol (Paclitaxel in Cremophor EL)."
##  [24] "This drug may not be available to some research laboratories."
##  [25] "10."
##  [26] "Multichannel (8-12 channel) pipettor for 50-300  L volumes."
##  [27] "3 Comparison between western blotting and EIA of iC 3 b."
##  [28] "The samples of plasma treated with negative control, positive control, and nanoparticle were a
##  [29] "Each bar shows mean and SD of duplicate response."
##  [30] "(a) Cremophor-EL was not analyzed in western blot assay because the presence of oil affects pr
##  [31] "Western blot results were analyzed by the NIH ImageJ software and shown is the peak intensity
##  [32] "(b)"
##  [33] "The scale was changed to show levels of iC 3 b in Cremophor-EL and nanoparticle-treated sample
##  [34] "The level of iC 3 b in the positive control sample (CVF) is 871 11  g/mL."
##  [35] "NC: negative control, NP: nanoparticle, CVF: cobra venom factor Effects on Complement 153 11."
##  [36] "Reagent reservoirs."
##  [37] "12."
##  [38] "ELISA plate reader capable of operating at 405 nm. 13."
##  [39] "See Note 1. 3 Methods 3. 1 Preparation of Controls and Plasma 1. Prepare Cobra Venom Factor (C
##  [40] "CVF is supplied as a frozen solution."
##  [41] "Thaw this stock, prepare singleuse aliquots and store them at a nominal temperature of 80 C as
##  [42] "Avoid repeated freeze/thaw cycles."
##  [43] "After thawing single-use aliquot and using it in the assay, discard any leftover material."
##  [44] "For this experiment, use 30  L (1. 1-50 U) of CVF solution."
##  [45] "This control activates the complement system through the alternative pathway (see Note 2)."
##  [46] "2. Prepare Positive Control 2, a nanoparticle-relevant control."
##  [47] "Cremophor-EL is an excipient commonly used in the pharmaceutical industry to dissolve hydropho
##  [48] "Cremophor-EL is a nanosized micelle, which is known to induce CARPA syndrome [ 5 ], and there
##  [49] "The following procedure can be used to prepare Cremophor-EL with the composition similar to th
##  [50] "(polyoxyethylated castor oil) and 49. 7% (v/v) dehydrated alcohol, USP and 2 mg of citric acid
##  [51] "Store at room temperature."
##  [52] "To prepare CremophorEL mix commercial Cremophor 1:1 with ethanol containing 2 mg/mL of citric
```

```
##  [53] "Cremophor-EL-formulated Paclitaxel (Taxol) can be used as an alternative for this nanoparticle
##  [54] "It is supplied at a stock concentration of 6 mg/mL of paclitaxel."
##  [55] "When used in this assay, the final concentration of Paclitaxel is 2 mg/mL."
##  [56] "Store at 2-8 C."
##  [57] "PEGylated liposomal doxorubicin (Doxil) can also be used as a nanoparticle-relevant positive c
##  [58] "Doxil is doxorubicin formulated in nanoliposomes."
##  [59] "It is available through the pharmacy as 20 mg of Doxorubicin HCl in 10 mL of vehicle."
##  [60] "Store at 2-8 C. 3. Prepare the Inhibition/Enhancement Control."
##  [61] "Use the positive control sample after incubation."
##  [62] "Prior to loading this sample onto the ELISA plate, add nanoparticles at the same final concent
##  [63] "For example, one can 154 Barry W. Neun et al."
##  [64] "mix 20  L of the positive control sample and 10  L of the test nanoparticle."
##  [65] "The test result for this sample needs to be adjusted by the dilution factor 1. 5 prior to comp
##  [66] "If the test results do not differ more than 25% of each other, the test nanoparticle at the gi
##  [67] "4. Prepare the Negative Control."
##  [68] "Sterile Ca 2+/Mg 2+-free PBS is used as a negative control."
##  [69] "Store at room temperature for up to 6 months."
##  [70] "5. Prepare the Vehicle Control, which is relevant to the given nanoparticle."
##  [71] "When nanoparticles are formulated in a vehicle other than saline or PBS, the vehicle sample sh
##  [72] "This control is specific to each given nanoparticle sample."
##  [73] "It should be prepared to match the formulation buffer of the nanoparticle by both the composit
##  [74] "6."
##  [75] "To prepare plasma, the blood is spun down in a centrifuge for 10 min at 2500 g (see Note 3)."
##  [76] "3. 2 Preparation of Nanoparticle Samples"
##  [77] "This assay requires 400  L of nanoparticles in PBS at a concentration three times higher than
##  [78] "The concentration is selected based on the plasma concentration of the nanoparticle at the int
##  [79] "For the purpose of this protocol, this concentration is called the "theoretical plasma concent
##  [80] "Considerations for estimating theoretical plasma concentration were reviewed elsewhere [ 21 ]
##  [81] "When the intended therapeutic concentration is unknown, the highest final concentration is 1 m
##  [82] "For example, if the final theoretical plasma concentration to be tested is 0. 2 mg/mL, then a
##  [83] "When 0. 1 mL of each of these samples is added to the test tube and mixed with 0. 1 mL of plas
##  [84] "In a microcentrifuge tube, combine equal volumes (100  L of each) of veronal buffer, human pla
##  [85] "Prepare two replicates of each sample."
##  [86] "Effects on Complement 155 2. Vortex tubes to mix all reaction components, spin briefly in a mi
##  [87] "3. Prepare 100  L aliquots and either use in EIA immediately or freeze at 20 C for later analy
##  [88] "Follow the manufacturer's instruction to reconstitute complement standard, buffers, and contro
##  [89] "5. Dilute plasma samples prepared in step 3 in complement specimen diluent reagent (provided w
##  [90] "Use the following dilution guide for each individual assay (see Note 4): iC 3 b-
1:1500 for positive control sample; 1:75 for negative control, and other test samples."
##  [91] "C 4 d-1:30 for all samples."
##  [92] "Bb-1:75 for all samples. 6."
##  [93] "Follow the manufacturer's instruction for plate loading volumes, incubation time, and plate wa
##  [94] "3. 4 Calculations and Results Interpretation"
##  [95] "Do not forget to use the appropriate dilution factor for control and study samples."
##  [96] "Compare determined amount of complement components between positive control or study samples w
##  [97] "An increase in the complement component species two fold or higher above the background (negat
##  [98] "If a nanoparticle under study generated a positive response in any of the EIA assays, compare
##  [99] "Doxil is used in the clinic and is Box 1 Example calculation of nanoparticle concentrations fo
## [100] "Reproduced with permission from ref."
## [101] "21 156 Barry W. Neun et al."
## [102] "known to induce complement activation-related hypersensitivity reactions in sensitive patients
## [103] "Using Doxil helps to interpret results of this in vitro study for a test nanoparticle."
## [104] "If the degree of activation observed for the test nanoparticle is equal to or greater than tha
## [105] "If the degree of activation is lower than that of Doxil, complement activation should be consi
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Neun, B W et al 2018.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "Materials"
word_end_mm <- "Doxil"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 1283
```

```
## [1] "2 Materials"
```

```
display_occurences(word_end_mm, conllu_df)
```

```
##  [1]  658  872 1418 1456 1518 2095 2110 2969 2975 3022 3056 3095 3125 3570 3585
## [16] 3630 4907
```

```
##  [1] "For example, liposomal formulations of doxorubicin (Doxil) and amphotericin (Ambisome), when ac
##  [2] "In the case of Doxil, the combination of three factors, including particle shape, PEG coating,
##  [3] "Plasma samples treated with negative control, positive (cobra venom factor), and Doxil were ana
##  [4] "The mean responses from positive control (PC) and Doxil samples were divided by the mean respor
##  [5] "Doxil (Doxorubicin HCl, liposome, injection)."
##  [6] "PEGylated liposomal doxorubicin (Doxil) can also be used as a nanoparticle-relevant positive co
##  [7] "Doxil is doxorubicin formulated in nanoliposomes."
##  [8] "If a nanoparticle under study generated a positive response in any of the EIA assays, compare t
##  [9] "Doxil is used in the clinic and is Box 1 Example calculation of nanoparticle concentrations for
## [10] "Using Doxil helps to interpret results of this in vitro study for a test nanoparticle."
## [11] "If the degree of activation observed for the test nanoparticle is equal to or greater than that
## [12] "If the degree of activation is lower than that of Doxil, complement activation should be consid
## [13] "If the degree of activation is lower than that of Doxil, complement activation should be consid
## [14] "4 Analysis of complement activation by Doxil in individual and pooled plasma."
## [15] "Plasma from ten donors was used to analyze Doxil."
## [16] "Shows results for controls and Doxil in individual plasma from 10 donors."
## [17] "Liposome-induced complement activation and related cardiopulmonary distress in pigs: factors pr
```

```
mm_section<-cut_conllu_df(1283, 3125, conllu_df)
```

```
##   [1] "2 Materials"
##   [2] "1."
##   [3] "Sterile Ca 2+/Mg 2+-free phosphate buffered saline (PBS)."
##   [4] "2. Cobra venom factor as the positive control."
##   [5] "3. Veronal buffer."
##   [6] "4."
##   [7] "Pooled human plasma, anti-coagulated with sodium citrate."
##   [8] "5. iC 3 b EIA kit (e. g., MicroVue by Quidel Corp.)."
##   [9] "6. 1. 0 N HCl, as stop solution."
##  [10] "Stop solution is provided with each kit, but can also be prepared separately."
##  [11] "Dilute stock hydrochloric acid to a final concentration of 1. 0 N. Filter and store and room t
##  [12] "Fig."
##  [13] "2 Comparison between C 3 a and iC 3 b EIAs."
##  [14] "Plasma samples treated with negative control, positive (cobra venom factor), and Doxil were ar
##  [15] "Arg and iC 3 b EIA."
##  [16] "Each sample was analyzed in duplicate (%CV < 10)."
##  [17] "The mean responses from positive control (PC) and Doxil samples were divided by the mean respo
```

```
##  [18] "The background in the C 3 a EIA was higher."
##  [19] "Due to the differences in the background, the S/N ratio of iC 3 b assay was higher 152 Barry W
##  [20] "Doxil (Doxorubicin HCl, liposome, injection)."
##  [21] "This is a prescription medication available from a licensed pharmacy."
##  [22] "It may not be available to some research laboratories."
##  [23] "8. Cremophor. 9. Taxol (Paclitaxel in Cremophor EL)."
##  [24] "This drug may not be available to some research laboratories."
##  [25] "10."
##  [26] "Multichannel (8-12 channel) pipettor for 50-300  L volumes."
##  [27] "3 Comparison between western blotting and EIA of iC 3 b."
##  [28] "The samples of plasma treated with negative control, positive control, and nanoparticle were a
##  [29] "Each bar shows mean and SD of duplicate response."
##  [30] "(a) Cremophor-EL was not analyzed in western blot assay because the presence of oil affects pr
##  [31] "Western blot results were analyzed by the NIH ImageJ software and shown is the peak intensity
##  [32] "(b)"
##  [33] "The scale was changed to show levels of iC 3 b in Cremophor-EL and nanoparticle-treated sample
##  [34] "The level of iC 3 b in the positive control sample (CVF) is 871 11  g/mL."
##  [35] "NC: negative control, NP: nanoparticle, CVF: cobra venom factor Effects on Complement 153 11."
##  [36] "Reagent reservoirs."
##  [37] "12."
##  [38] "ELISA plate reader capable of operating at 405 nm. 13."
##  [39] "See Note 1. 3 Methods 3. 1 Preparation of Controls and Plasma 1. Prepare Cobra Venom Factor (C
##  [40] "CVF is supplied as a frozen solution."
##  [41] "Thaw this stock, prepare singleuse aliquots and store them at a nominal temperature of 80 C as
##  [42] "Avoid repeated freeze/thaw cycles."
##  [43] "After thawing single-use aliquot and using it in the assay, discard any leftover material."
##  [44] "For this experiment, use 30  L (1. 1-50 U) of CVF solution."
##  [45] "This control activates the complement system through the alternative pathway (see Note 2)."
##  [46] "2. Prepare Positive Control 2, a nanoparticle-relevant control."
##  [47] "Cremophor-EL is an excipient commonly used in the pharmaceutical industry to dissolve hydropho
##  [48] "Cremophor-EL is a nanosized micelle, which is known to induce CARPA syndrome [ 5 ], and there
##  [49] "The following procedure can be used to prepare Cremophor-EL with the composition similar to th
##  [50] "(polyoxyethylated castor oil) and 49. 7% (v/v) dehydrated alcohol, USP and 2 mg of citric acid
##  [51] "Store at room temperature."
##  [52] "To prepare CremophorEL mix commercial Cremophor 1:1 with ethanol containing 2 mg/mL of citric
##  [53] "Cremophor-EL-formulated Paclitaxel (Taxol) can be used as an alternative for this nanoparticle
##  [54] "It is supplied at a stock concentration of 6 mg/mL of paclitaxel."
##  [55] "When used in this assay, the final concentration of Paclitaxel is 2 mg/mL."
##  [56] "Store at 2-8 C."
##  [57] "PEGylated liposomal doxorubicin (Doxil) can also be used as a nanoparticle-relevant positive c
##  [58] "Doxil is doxorubicin formulated in nanoliposomes."
##  [59] "It is available through the pharmacy as 20 mg of Doxorubicin HCl in 10 mL of vehicle."
##  [60] "Store at 2-8 C. 3. Prepare the Inhibition/Enhancement Control."
##  [61] "Use the positive control sample after incubation."
##  [62] "Prior to loading this sample onto the ELISA plate, add nanoparticles at the same final concent
##  [63] "For example, one can 154 Barry W. Neun et al."
##  [64] "mix 20  L of the positive control sample and 10  L of the test nanoparticle."
##  [65] "The test result for this sample needs to be adjusted by the dilution factor 1. 5 prior to comp
##  [66] "If the test results do not differ more than 25% of each other, the test nanoparticle at the gi
##  [67] "4. Prepare the Negative Control."
##  [68] "Sterile Ca 2+/Mg 2+-free PBS is used as a negative control."
##  [69] "Store at room temperature for up to 6 months."
##  [70] "5. Prepare the Vehicle Control, which is relevant to the given nanoparticle."
##  [71] "When nanoparticles are formulated in a vehicle other than saline or PBS, the vehicle sample sh
```

```
##  [72] "This control is specific to each given nanoparticle sample."
##  [73] "It should be prepared to match the formulation buffer of the nanoparticle by both the composi
##  [74] "6."
##  [75] "To prepare plasma, the blood is spun down in a centrifuge for 10 min at 2500 g (see Note 3)."
##  [76] "3. 2 Preparation of Nanoparticle Samples"
##  [77] "This assay requires 400  L of nanoparticles in PBS at a concentration three times higher than
##  [78] "The concentration is selected based on the plasma concentration of the nanoparticle at the in
##  [79] "For the purpose of this protocol, this concentration is called the "theoretical plasma concen
##  [80] "Considerations for estimating theoretical plasma concentration were reviewed elsewhere [ 21 ]
##  [81] "When the intended therapeutic concentration is unknown, the highest final concentration is 1 m
##  [82] "For example, if the final theoretical plasma concentration to be tested is 0. 2 mg/mL, then a
##  [83] "When 0. 1 mL of each of these samples is added to the test tube and mixed with 0. 1 mL of pla
##  [84] "In a microcentrifuge tube, combine equal volumes (100  L of each) of veronal buffer, human pla
##  [85] "Prepare two replicates of each sample."
##  [86] "Effects on Complement 155 2. Vortex tubes to mix all reaction components, spin briefly in a mi
##  [87] "3. Prepare 100  L aliquots and either use in EIA immediately or freeze at 20 C for later analy
##  [88] "Follow the manufacturer's instruction to reconstitute complement standard, buffers, and contro
##  [89] "5. Dilute plasma samples prepared in step 3 in complement specimen diluent reagent (provided w
##  [90] "Use the following dilution guide for each individual assay (see Note 4): iC 3 b-
1:1500 for positive control sample; 1:75 for negative control, and other test samples."
##  [91] "C 4 d-1:30 for all samples."
##  [92] "Bb-1:75 for all samples. 6."
##  [93] "Follow the manufacturer's instruction for plate loading volumes, incubation time, and plate wa
##  [94] "3. 4 Calculations and Results Interpretation"
##  [95] "Do not forget to use the appropriate dilution factor for control and study samples."
##  [96] "Compare determined amount of complement components between positive control or study samples w
##  [97] "An increase in the complement component species two fold or higher above the background (negat
##  [98] "If a nanoparticle under study generated a positive response in any of the EIA assays, compare
##  [99] "Doxil is used in the clinic and is Box 1 Example calculation of nanoparticle concentrations fo
## [100] "Reproduced with permission from ref."
## [101] "21 156 Barry W. Neun et al."
## [102] "known to induce complement activation-related hypersensitivity reactions in sensitive patients
## [103] "Using Doxil helps to interpret results of this in vitro study for a test nanoparticle."
## [104] "If the degree of activation observed for the test nanoparticle is equal to or greater than tha
## [105] "If the degree of activation is lower than that of Doxil, complement activation should be consi
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Reddy, A et al 2010.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "Materials"
word_end_mm <- "significant"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 1395
```

```
## [1] "Materials and methods .1."
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1]  959 1219 2922 3193 3464 3677 4121 4706 4854
```

```
## [1] "Rats exposed by inhalation exposure to 13 C-labeled ultrafine carbon particles demonstrated tha
## [2] "Kagan et al. (2006) also demonstrated that SWCNT esulted in a significant loss of intracellular
## [3] "Results with p < 0. 05 ere considered to be statistically significant. . Results .1."
```

```
## [4] "Exposure of MWCNT (10-100 g/ml) to HEK 293 cells for 48 h resulted in significant (p < 0. 001) I
## [5] "Similar to quartz, exposure of MWCNT (10-100 g/ml) for 8 h caused a significant dose-dependant i
## [6] "As showed in Fig. 6, exposure of both nanoparicles to HEK 293 cells resulted in significant (p <
## [7] "Exposure to MWCNT for 48 h caused a significant time-dependant increase in IL-8 release from HEK
## [8] "The ecent study by Oberdorster (2004) indicated that nanomaterils (Fullerenes C 60) induced oxid
## [9] "They reported that exposure of Ag nanoparticles to RL 3 A rat liver cells for 24 h resulted in c
50 g/ml."
```

```
mm_section<-cut_conllu_df(1395, 2922, conllu_df)
```

```
##  [1] "Materials and methods .1."
##  [2] "Chemicals Fetal bovine serum (FBS), Dulbecco's modified eagle's medium (DMEM), peniillin, amphot
##  [3] "3-(4, 5-Dimethylthiazol-2-yl)-2, 5-diphenyltetrazolium bromide (MTT), 2 hiobarbituric acid (TBA
##  [4] "Louis, MO, USA)."
##  [5] "The glutathione, nterleukin-8 (IL-8), lactate dehydrogenase (LDH) assay kits were purchased fro
##  [6] ".2. Multi wall carbon nanoparticles-characterization Multi walled carbon nanotubes produced by
##  [7] "However, these anotubes exist primarily as agglomerated "ropes" of nanotubes."
##  [8] "In dry powder orm, each particle system was analyzed for various physico-chemical properties."
##  [9] "ize and crystallinity were determined by dynamic light scattering spectroscopy DLS) (Berne and
## [10] "The surface area of the carbon nanoparticles were easured by the Brunauer-Emmett-Teller method
## [11] "All the nanoparticles ere suspending in phosphate buffer saline (PBS) to get the stock concentr
## [12] "The final concentrations were made n the cultured media (DMEM) without serum for the uniform di
## [13] "Cell culture"
## [14] "The human embryonic kidney cell line (HEK 293) was purchased from cell bank f national centre i
## [15] "Cells were cultured in a full DMEM edium containing 10% FBS, 100 U/ml penicillin and 100 g/ml s
16 2. 4. Exposure of MWCNT"
## [16] "The stock suspensions of carbon nanoparticles were freshly diluted to different concentrations
## [17] "After cells had attached for 12 h in the full medium, the medium was replaced with low serum DM
## [18] "Freshly dispersed particle suspensions were immediately applied to the cells and allowed to inc
## [19] "Cells free of carbon nanoparticles were used as control cells throughout each assay."
## [20] "Tests for LDH release, cytotoxicity and cell viability assay (MTT), cytokine production (IL-8)
## [21] "2. 5."
## [22] "Assessment of cytotoxicity Mitochondrial function and cell viability were measured by the MTT a
## [23] "HEK 293 cells were plated into a 96-well plate at a density of 1. 0 × 104 cells/well."
## [24] "Cells were grown overnight in the full medium and then switched to the low serum media followed
## [25] "After 48 h of treatment with different concentrations of nanoparticles, the cells were incubate
## [26] "The medium was then removed and 100 l of DMSO was added into each well to dissolve formazan cry
## [27] "After thoroughly mixing, the plate was read at 570 nm for optical density that is directly corr
## [28] "Cell death rate was calculated from the relative absorbance at 570 nm and expressed as the perc
## [29] "The concentrations of carbon nanotubes used in this cytotoxicity assay were 3-
300 g/ml."
## [30] "2. 6."
## [31] "LDH release Cells were seeded in 24-well plates, exposed to increasing concentrations of partic
100 g/ml)."
## [32] "After 48 h of incubation, the plate was centrifuged at 1900 rpm for 4 min."
## [33] "The media were transferred into a fresh 24-well plate and analyzed for LDH release as describe
## [34] "Each experiment was done in triplicate."
## [35] "Cytotoxicity is expressed relative to the basal LDH release by untreated control cells."
## [36] "2. 7."
## [37] "Production of IL-8 Cells (2 × 104 cells/ml) were grown in 24-well plates and incubated with di
100 g/ml) nanoparticles for 48 h."
## [38] "The supernatants were collected, centrifuged to remove any remaining nanoparticles."
## [39] "Concentrations of the pro-inflammatory cytokine, interleukine-8 (IL-8), were determined by huma
## [40] "Cells incubated without nanoparticles were used as a control."
## [41] "The absorbance was measured at 450 nm and quantified with a microplate reader."
```

```
## [42] "2. 8."
## [43] "Quantification of intracellular GSH levels"
## [44] "Cellular levels of reduced GSH were determined using the GSH-400 colorimetric assay kit."
## [45] "The method is based on a chemical reaction between GSH and 5, 5 -dithiobis (2-nitrobenzoic acid
## [46] "Thus GSH concentration in a sample solution can be determined by the measurement at 412 nm abso
## [47] "HEK 293 cells were plated into a 24-well plate at a density of 2 × 104 cells/ml."
## [48] "After 48 h exposure to carbon nanoparticles, the cells were washed twice in ice-cold PBS and th
## [49] "The cell homogenate was centrifuged at 3000×g at 4 ∘C for 10 min."
## [50] "The assay was performed on 200 l centrifugation supernatants according to manufacturer's protoc
visible spectrophotometer (Elico, India)."
## [51] "Protein content was determined for the same cell homogenate."
## [52] "GSH level was calculated and expressed as the percentage of control."
## [53] "2. 9."
## [54] "Estimation of lipid peroxidation"
## [55] "The MDA content, a measure of lipid peroxidation, was assayed in the form of thiobarbituric aci
## [56] "HEK 293 cells were plated into a 24-well plate at a density of 1 × 105 cells/well."
## [57] "After 48 h exposure to MWCNT (10-100 g/ml), the cells were washed with ice-cold PBS and homoger
## [58] "The cell homogenates were used in the TBARS assay."
## [59] "Briefly, 100 l cell homogenates were mixed with 1 ml of 0. 67% TBA, 1. 5 ml 20% trichloroaceti
## [60] "The mixtures were incubated in a boiling water bath for 20 min."
## [61] "After cooling to room temperature, the reaction mixture was centrifuged at 4000×g for 10 min an
visible spectrophotometer."
## [62] "The concentrations of TBARS were calculated using tetraethoxypropane as A. R. N. Reddy et al. /
16 13 Table 1 Characterization of multi wall carbon nanotubes."
## [63] "Carbon nanotubes Method of production Size (nm) Crystallinity Surface area (m 2/g) ARC Electric
150 Hexagonal 197 CVD Chemical vapor deposition (CVD) 60-80 Cubic 252 (SEM) a T 2 ( 2 S c w w 3 3 o n g
## [64] "IL-8 serves as aFig."
## [65] "1. Scanning electron micrographs reference standard."
## [66] "The quantities of TBARS were presented as the percentage of BARS production over the control."
## [67] ".10. Protein assay"
## [68] "The total protein concentration was measured by the Bradford method Bradford, 1976) using bovi
## [69] ".11."
## [70] "Statistical analysis"
## [71] "All the experimental values were expressed as mean ± standard deviation (SD)."
## [72] "tatistical analysis was performed for the experiments conducted in at least tripliate using one
## [73] "Student's t-test as used to compare the means of MWCNT 1 with MWCNT 2."
## [74] "Results with p < 0. 05 ere considered to be statistically significant. . Results .1."
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

"Richter, AW et al 1983.pdf" : it is a scan. 1983.

```
pdf_name <- "Richter, AW et al 1983.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "Materials"
word_end_mm <- "significant"

display_occurences(word_start_mm, conllu_df)
```

```
## integer(0)
```

```
## character(0)
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## integer(0)
```

```
## character(0)
```

```
mm_section<-cut_conllu_df(1395, 2922, conllu_df)
```

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

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Sadaf, A et al 2012.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "MATERIALS"
word_end_mm <- "05"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 889
```

```
## [1] "2. MATERIALS AND METHODS 2. 1."
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1] 1746 2155 2356
```

```
## [1] "An asterisk indicates the WBCs were significantly higher than that in control group (P < 0 05).
## [2] "Statistical differences were evaluated using the t test and considered significance at P < 0 05
## [3] "An asterisk indicates the level of BUN and CREA were significantly higher than that of control
```

```
mm_section<-cut_conllu_df(889, 2155, conllu_df)
```

```
##  [1] "2. MATERIALS AND METHODS 2. 1."
##  [2] "Materials"
##  [3] "Water for all reactions, solutions preparation was ultrapure water (18 M · cm−1) double distil
##  [4] "All chemicals mentioned in the current investigations were used as received; cadmium perchlora
Aldrich."
##  [5] "2. 2."
##  [6] "Synthesis of CdTe QDs and CdTe@SiO 2 NPs CdTe QDs and CdTe@SiO 2 NPs were prepared as describe
##  [7] "35 Briefly, the CdTe QDs were synthesized by dissolving 1."
##  [8] "262 g of Cd(ClO 4 2 · 6 H 2 O (3. 01 mmol) into 160 mL of water, and then adding 0."
##  [9] "371 g of TGA (3. 91 mmol) in continuous stirring."
## [10] "The H 2 Te gas was added with stirring followed by drop wise addition of 13 mL of 0. 5 M H 2 S
## [11] "The resulting QDs were stabilized by adding 0. 371 g of TGA (3. 91 mmol) in continuous stirrin
## [12] "TGA capped CdTe QDs were modified with silica shell according using reverse microemulsion meth
## [13] "Finally prepared CdTe@SiO 2 were washed with PBS (pH 7. 4), isolated, and then saved for usage
## [14] "2. 3. Photophysical Characterization Prepared CdTe@SiO 2 NPs were characterized using JEOL JSM
## [15] "2. 4."
## [16] "Animal Studies"
## [17] "The study complied with standards for the care and use of laboratory animals (Laboratory Anima
## [18] "Thirty female BALB/c mice (20± 1 g, 6-7 weeks old) were obtained from the Animal Center of Sou
## [19] "Each group of mice was separately housed in positive pressure (25-30 C, 50-
60% relative humidity, 12/12 hours light/dark cycle) and fed a commercial diet while water was availabl
## [20] "To study the toxicity of the NPs in vivo, the mice in group 1 and 2 were injected with CdTe QD
## [21] "The QDs, NPs and normal saline were sterilized before injection."
## [22] "(E) (A) Fig. 1. (A) Photoluminescence spectra of CdTe Qds and CdTe@SiO 2 NPs. (B) TEM images o
## [23] "The white circle in each picture indicates a single CdTe QDs."
## [24] "The scanning transmission electron microscope (SEM) image of the silica coated CdTe QDs (D)."
## [25] "TEM images of CdTe quantum dots (E)."
## [26] "8288 J. Nanosci."
```

```
## [27] "Nanotechnol."
## [28] "12, 8287-8292, 2012"
## [29] "Delivered by Publishing Technology to: Chinese University of Hong Kong IP: 117."
## [30] "255."
## [31] "235."
## [32] "34 On: Tue, 12 Jan 2016 03:26:53 Copyright: American Scientific Publishers R E S E A R C H A R
## [33] "Toxicity Evaluation of Hydrophilic CdTe Quantum"
## [34] "Dots and CdTe@SiO 2 Nanoparticles in Mice WBC 0 5 10 15 20 25 0 Days after injection"
## [35] "W B C 10 9 / L CdTe QDs"
## [36] "CdTe@SiO 2 NPs PBS 7 15 30 RBC 0 2 4 6 8 10 12 Days after injection CdTe QDs"
## [37] "CdTe@SiO 2 NPs PBS R B C (1 01 2 / L ) 0 7 15 30 PLT 0 150 300 450 600 750 900 1050 Days after
## [38] "Pl at el et c ou nt s (1 01 2 / L ) CdTe QDs"
## [39] "CdTe@SiO 2 NPs PBS 0 7 15 30 Fig. 2."
## [40] "Blood cell counts of mice."
## [41] "White blood cells (WBCs), and red blood cells (RBCs)"
## [42] "Platelet counts (PLT) of experimental mice were measured at 0, 7, 15, 30 days after intravenous
## [43] "An asterisk indicates the WBCs were significantly higher than that in control group (P < 0 05)
## [44] "Data are shown as mean±S. D. for three mice per group."
## [45] "While no significant difference was noted in PLT and RBCs counts of the treated and control gr
## [46] "2. 5. Blood Analysis"
## [47] "Urine and blood samples were collected at the time intervals of 0, 7, 15, 30 days post inocula
## [48] "Venous blood samples were collected in evacuated tubes containing EDTA, sodium citrate anticoag
## [49] "CdTe@SiO 2 NPs PBS 7 15 30 ALT 0 20 40 60 CdTe QDs"
## [50] "CdTe@SiO 2 NPs PBS C on c. ( IU /L ) 0 Days after injection 7 15 30 Fig. 3. The plasma levels
## [51] "Data are shown as mean±S. D."
## [52] "for three mice per group."
## [53] "using a standard sphenoid vein blood puncture from each group (n= 3) and were analyzed for star
## [54] "Liver function was evaluated with serum levels of alanine amino transferase (ALT), aspartate ar
## [55] "Nephrotoxicity was determined by blood urea nitrogen (BUN) and creatinine (CREA)."
## [56] "These parameters were assayed using a Hitachi 7600 Automatic Biochemical Autoanalyzer."
## [57] "27 29 2. 6."
## [58] "Histopathology of Tissues"
## [59] "The histopathology of the tissues was performed as described previously, 37 with few modificati
## [60] "The mice were killed at 30 days post treatment and their organs such as liver, spleen, kidneys
## [61] "Haematoxylin and eosin-stained histological sections of the fixed organs were observed with lig
## [62] "2. 7. Statistical Analysis"
## [63] "The results were presented as mean ± SD."
## [64] "Statistical differences were evaluated using the t test and considered significance at P < 0 05
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Tang, J et al 2009.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "MATERIALS"
word_end_mm <- "significant"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 637
```

```
## [1] "2. MATERIALS AND METHODS 2. 1."
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1] 1354 1491 2692 2842 3651
```

```
## [1] "P < 0. 05 was considered statistically significant."
## [2] "There were no significant differences between the SNP group and SMP group (P > 0. 05) in the si
## [3] "There was no significant difference between the SMPs group and the control group (P > 0. 05)."
## [4] "Figure 7 shows the silver content in blood with time, and shows a significant difference in the
## [5] "Figure 7 shows a significant difference in the silver content in blood between the SNPs group a
```

```
mm_section<-cut_conllu_df(637, 1354, conllu_df)
```

```
##  [1] "2. MATERIALS AND METHODS 2. 1."
##  [2] "Experimental Materials SNPs and silver microparticles (SMPs), both pure silver, were purchased
##  [3] "DMEM (Dulbecco's modified eagle medium, Gibco, Beijing, China)."
##  [4] "2. 2."
##  [5] "Characteristics of SNPs and SMPs"
##  [6] "After high pressure and high temperature sterilization, SNPs were diluted in the DMEM at a rat
##  [7] "The size and morphology of the 4924 J. Nanosci."
##  [8] "Nanotechnol."
##  [9] "2009, Vol. 9, No. 8 1533-4880/2009/9/4924/009 doi:10."
## [10] "1166/jnn."
## [11] "2009."
## [12] "1269"
## [13] "Delivered by Publishing Technology to: Rice University IP: 96."
## [14] "57."
## [15] "51."
## [16] "250 On: Fri, 11 Dec 2015 11:36:25 Copyright: American Scientific Publishers RE S E A R C H A R
## [17] "Distribution, Translocation and Accumulation of Silver Nanoparticles in Rats SNPs were observe
## [18] "EM 208 S transmission electron microscope (TEM; Royal Dutch Philips"
## [19] "Electronics Ltd., Eindhoven, Netherlands)."
## [20] "After high pressure and high temperature sterilization, the size and morphology of SMPs was de
## [21] "2. 3. In Vivo Trial After high pressure and high temperature sterilization, the weighed SNPs a
## [22] "Ninety"
## [23] "Wistar female rats (120 g±5 g) were randomly divided into three groups: control group, SNPs gr
## [24] "Each group was treated with its corresponding suspension by a subcutaneous injection, at a dos
## [25] "Five animals were collected after 2, 4, 8, 12, 18, and 24 weeks, and faeces and urine were har
## [26] "In addition, 3 ml femoral artery blood was sampled, and then animals were sacrificed to obtain
## [27] "Samples from brains, liver, spleen, lung and kidney were processed for electron microscopy by
## [28] "Then thin sections were obtained for the observation of cellular ultrastructure with a TEM, th
## [29] "Organs, blood and excrement were soaked in digestion solution, consisting of high purity nitri
## [30] "They were then heated at 60 C and diluted to a metered volume using distilled water."
## [31] "ICP-MS was used to determine the content of silver in the metered solution."
## [32] "The total mass of silver in each organ was then determined ( g)."
## [33] "Results are expressed as mean ± standard deviation."
## [34] "2. 4."
## [35] "Statistical Analysis Results were processed with SPSS 12. 0 software (SPSS Inc., Shanghai, Chi
## [36] "Limit states design (LSD) was used to discern multiple comparisons of equalizing values."
## [37] "P < 0. 05 was considered statistically significant."
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Vicente, S et al 2017.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "METHODS"
word_end_mm <- "USA"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 1235
```

```
## [1] "2 | METHODS 2. 1 | Materials: Particles, chemicals and antibodies LUDOX® amorphous non-porous si
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1] 1266 1458 1516 1540 2179 2601
```

```
## [1] "2 | METHODS 2. 1 | Materials: Particles, chemicals and antibodies LUDOX® amorphous non-porous si
## [2] "ActinRed®, Lysotracker® Red DND-99, propidium iodide and Hoechst 33342 were purchased from Molec
## [3] "Anti-Rab 7 (D 95 F 2) XP® rabbit mAb, anti-caveolin-1 (D 46 G 3) XP  rabbit mAb, anti-rabbit and
## [4] "Anti-LAMP 1 (H 4 B 4) mouse mAb were from Santa Cruz Biotechnology Inc. (Dallas, Texas, USA)."
## [5] "To measure the level of lactate dehydrogenase (LDH) released from damaged cells, the CytoTox 96®
## [6] "Post-acquisition treatment of images was performed using ImageJ 1. 48 (NIH, Bethesda, Maryland,
```

```
mm_section<-cut_conllu_df(1235, 2601, conllu_df)
```

```
##  [1] "2 | METHODS 2. 1 | Materials: Particles, chemicals and antibodies LUDOX® amorphous non-porous s
##  [2] "SiNP 70, SMP 200 and SMP 500 amorphous non-porous silica particles were obtained from Postnova
##  [3] "A summary of the characteristics provided by the manufacturers is shown in Table 1."
##  [4] "Fluorescein-labelled SiNP 70, SMP 200 andTABLE 1 Characteristics of colloidal silica obtained i
##  [5] "N/A: Not available SiNP 20 SiNP 70 SMP 200 SMP 500 Manufacturer Sigma-Aldrich Postnova analytic
258 43 15 6"
##  [6] "Polydispersity N/A < 0. 2 < 0. 2 < 0. 2 SMP 500 with the same characteristics were obtained fro
##  [7] "Paraformaldehyde (PFA) 16% aqueous solution was obtained from Alfa-Aesar GmbH (Karlsruhe, Germa
##  [8] "Bovine serum albumin (BSA) and Triton X-100 were purchased from Sigma-Aldrich."
##  [9] "Gelatin was obtained from BioRad (Marnes-La-Coquette, France)."
## [10] "ActinRed®, Lysotracker® Red DND-99, propidium iodide and Hoechst 33342 were purchased from Molec
## [11] "Anti-Rab 7 (D 95 F 2) XP® rabbit mAb, anti-caveolin-1 (D 46 G 3) XP  rabbit mAb, anti-rabbit an
## [12] "Anti-LAMP 1 (H 4 B 4) mouse mAb were from Santa Cruz Biotechnology Inc. (Dallas, Texas, USA)."
## [13] "2. 2 | Particle size measurements and stability in culture media Dynamic light scattering (DLS)
## [14] "Three measurements were made for each sample using Zetasizer (Malvern, Worcestershire, UK) at 1
## [15] "Before particle analysis, the DLS profiles of cell culture media were acquired for the presence
## [16] "2. 3 | Cells and culture conditions"
## [17] "Human dermal fibroblasts (HDF) isolated from adult skin were from Gibco® (Life Technologies Co
## [18] "To culture these cells, Medium 106 supplemented with LSGS 50X (low serum growth supplement) and
## [19] "K 17 keratinocytes (K 17) are immortalized cells produced from normal adult keratinocytes kindl
## [20] "The complete culture medium was prepared using EpiLife® supplemented with HKGS 100X (human kera
## [21] "Cells were maintained at 37°C in an atmosphere of 5% CO 2. 2. 4 | In vitro toxicological analys
## [22] "SiPs were diluted in complete culture medium."
## [23] "The concentrations ranged from 10 to 200  g/ml."
## [24] "SiPs were then incubated with the cells for 24 h. 2. 4. 1 | WST-1"
## [25] "To evaluate the metabolic activity of the cells, the cell proliferation reagent WST-1 was used
## [26] "The quantification of soluble formazan dye produced by metabolically active cells was measured
## [27] "The cell viability rate was calculated by considering the signal of non-treated cells as the 10
## [28] "1398 VICENTE ET AL."
## [29] "2. 4. 2 | Propidium iodide staining and cell counting Propidium iodide (PI) can stain the nucle
## [30] "After 24 h of incubation with SiPs, 5  g/ml of PI diluted in PBS were added and incubated for 2
## [31] "Non-treated cells and positive controls, consisting of cells incubated with Triton X-100 0. 025
## [32] "After treatment with PI, cells were washed with PBS and fixed using a solution of PFA 4% (in PB
## [33] "Then, all the nuclei in the sample were stained with Hoechst 33342 at 0. 1 mg/ml for 10 min at
## [34] "For cell counting, a wide field fluorescence microscope (Zeiss Axio Observer Z 1, Jena, Germany
## [35] "Images acquisition and post- acquisition analysis were performed using Morphostrider software
## [36] "16 fields (840."
## [37] "4 × 640."
## [38] "3  m) per well were acquired using an objective 10× and the adequate filter cubes for Hoechst a
```

```
## [39] "In the post-analysis process, both Hoechst- and PI-labelled nuclei were counted, and the ratio
## [40] "2. 4. 3 | LDH release"
## [41] "To measure the level of lactate dehydrogenase (LDH) released from damaged cells, the CytoTox 9
## [42] "Non-treated cells and positive controls (Triton X-100 0. 025%; 30 min) were also included."
## [43] "After treatment with SiPs for 24 h, the level of LDH in the culture medium of treated cells an
## [44] "The levels of LDH of the positive control were considered as 100%.2."
## [45] "4. 4 | Caspase 3/7 activity measurement"
## [46] "To measure the levels of caspase 3/7, the Apo-ONE Homogeneous Caspase-3/7 Assay (Promega) was u
## [47] "This method allows the quantification of the caspase 3/7 by the transformation of the kit reag
## [48] "The analysis was performed according to the manufacturer's protocol, and the measurement of th
## [49] "Non-treated cells were used as negative control (non-apoptotic cells).2."
## [50] "5 | Internalization and intracellular localization of SiPs K 17 keratinocytes were plated in 8
## [51] "For HDF, poly-D-lysine-coated 8-well ibidi -slides were used at 12, 000 cells/well."
## [52] "Fluorescein-labelled SiNP 70, SMP 200 and SMP 500 at 10 g/ml were incubated for either 2 or 24
## [53] "Cells were then washed with PBS, fixed with PFA 4% (15 min, RT) and permeabilized with Triton
## [54] "Samples were blocked and primary antibodies were incubated overnight at 4°C using adequate dilu
## [55] "After extensive washing, secondary antibodies at 1:1000 were incubated for 1 h at RT."
## [56] "Cells were then washed with PBS and nuclei stained with Hoechst 33342 at 0. 1 mg/ml (10 min, R
## [57] "To stain lysosomes, SiPs were removed and Lysotracker® Red DND-99 diluted in culture medium at
## [58] "Lysotracker® was replaced with fresh medium."
## [59] "Imaging was performed in living cells."
## [60] "Images were obtained using a Carl Zeiss Confocal Laser Scanning Microscope (LSM 510; Jena, Ger
## [61] "4 oil DIC objective."
## [62] "Post-acquisition treatment of images was performed using ImageJ 1. 48 (NIH, Bethesda, Maryland
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Wang, X et al 2010.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "MATERIALS"
word_end_mm <- "05"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 1412
```

```
## [1] "2. MATERIALS AND METHODS 2. 1."
```

```
display_occurences(word_end_mm, conllu_df)
```

```
##  [1] 2622 2907 2976 3382 3401 4114 4127 4394 4727 4740 5466 5666 5687 5995
```

```
##  [1] "Statistical significance was considered at p < 0 05."
##  [2] "There was no difference on body weight gaining between all exposed groups and the blank contro
##  [3] "Shown in Table I, lung indices of tau-MWNTs of all dose levels at 1 d post-exposure were signi
##  [4] "0004 0. 0073±0. 0022 aGroups statistically different from the PBS control, P < 0 05, bGroups o
##  [5] "0004 0. 0073±0. 0022 aGroups statistically different from the PBS control, P < 0 05, bGroups o
##  [6] "5400a b aGroups statistically different from the PBS control, P < 0 05, bGroups statistically
##  [7] "5400a b aGroups statistically different from the PBS control, P < 0 05, bGroups statistically
##  [8] "9200 aGroups statistically different from the PBS control, P < 0 05, bGroups statistically dif
##  [9] "6200a b aGroups statistically different from the PBS control, P < 0 05, bGroups statistically
## [10] "6200a b aGroups statistically different from the PBS control, P < 0 05, bGroups statistically
## [11] "Dose-dependent effect was explicitly seen at 1 d, 1 mg/kg group presented the highest level, 1
## [12] "The silicon dioxide induced much higher response of GSH to mouse lungs than the two CNTs sampl
## [13] "In groups of raw MWNTs, MDA of all dose groups went up significantly (p < 0 05) and peaked at
## [14] "The hydroxyproline levels of raw MWNTs (except for 0. 125 mg/kg) peaked at 7 d significantly (
```

```
mm_section<-cut_conllu_df(1412, 2622, conllu_df)
```

```
##  [1] "2. MATERIALS AND METHODS 2. 1."
##  [2] "Materials and Characterization"
##  [3] "The raw MWNTs, produced by chemical vapor deposition method (CVD) with a diameter of 10-
20 nm and a purity of >95% were purchased from Shenzhen Nanoharbor, China."
##  [4] "The morphological characterization of length and diameter of the purified MWNTs were made by ti
##  [5] "MWNTs were further characterized in our laboratory by thermogravimetric analysis (TGA) and ICP-
##  [6] "7 TauMWNTs were synthesized in our laboratory and further characterized by a variety of methods
##  [7] "Information on synthesis and more detailed characterization were stated in our previous work."
##  [8] "13 Silicon dioxide was used as a positive control."
##  [9] "It was obtained from the National Institute for Occupational Health and Poison Control, Chinese
## [10] "Silicon dioxide has a purity of 99%."
## [11] "More than 95% of the crystalline silica particles were less than 5 m in diameter."
## [12] "The surface area for silicon dioxide is 21."
## [13] "26 m 2/g according to the BET measurement."
## [14] "2. 2."
## [15] "Preparation of MWNTs Suspension"
## [16] "MWNTs and tau-MWNTs were freshly suspended in phosphate buffered saline (PBS) with a Dounce hor
## [17] "A stable suspension of MWNTs and tau-MWNTs were obtained in this way and used immediately. 2. 3
## [18] "They were randomly grouped (5 mice/group) and kept in a conventional animal facility and housed
## [19] "They were cared for and used humanely according to Animal Care and Use Program Guidelines of Pe
## [20] "The experimental protocol was approved by the local ethical committee for animals in research."
## [21] "J. Nanosci."
## [22] "Nanotechnol."
## [23] "10, 8516-8526, 2010 8517 Delivered by Ingenta to: Purdue University Libraries IP: 91."
## [24] "200. 82."
## [25] "192 On: Thu, 09 Jun 2016 06:43:32 Copyright: American Scientific Publishers R E S E A R C H A I
## [26] "Exposure"
## [27] "After being anesthetized with ether for 20 30 sec in a small chamber, individual mouse was secu
## [28] "The trachea was exposed by a 0. 5 cm incision on the ventral neck skin, and a small hole was ma
## [29] "25, 0. 5 or 1 mg/kg tau-MWNTs, raw MWNTs, 1 mg/kg silicon dioxide and PBS (negative control) we
## [30] "The neck incision was then sutured, swabbed with iodine."
## [31] "The mice were recovered active within 3 min after removal from anesthetic."
## [32] "The incision healed within one day, and the mice were observed daily until their scheduled term
## [33] "There were another 10 mice taken as blank control that was instilled nothing."
## [34] "Mice were sacrificed respectively after exposure to CNTs for 1, 7, 14 and 28 days."
## [35] "The blood samples and the organs including lungs, liver, kidneys, heart and spleen were collect
## [36] "2. 5."
## [37] "Biochemical Assay"
## [38] "The blood samples were collected by picking off the eyeballs."
## [39] "Serum was obtained after centrifugation of the collected blood samples at 4000 rpm for 10 min."
## [40] "All biochemical assays were performed using a Hitachi 7170A clinical automatic chemistry analyz
## [41] "Activity of LDH, ALP and ACE were measured using the commercial k ts (Bühlmann Laboratories, Su
## [42] "2. 6."
## [43] "Histopathological Study"
## [44] "The portions of lungs, liver, kidneys, heart and spleen samples for microscopic examination wer
## [45] "Sections were stained with hematoxylin-eosin and examined by light microscopy."
## [46] "2. 7. Ultrastructural Analysis of the Lungs Representative portions of the lungs were prefixed
## [47] "After being washed, they were postfixed with 1% osmium tetraoxide at 4 C for 3 hours and washed
## [48] "Following dehydration, resin embedding, sectioning to 70 nm thick and staining with uranyl acet
## [49] "2. 8. Preparation of Lung Homogenates A piece of each lung sample (0. 5-1 g) was taken for test
## [50] "The tissues were minced and washed several times by cold physiological saline (0. 9% NaCl), and
## [51] "At last, the homogenates were centrifuged at 2000 rpm for 10 min to obtain the lung supernatant
```

```
## [52] "Protein concentrations of lung supernatant were determined according to the Bradford's method,
## [53] "14 BSA and coomassie brilliant blue G-250 was purchased from Sigma."
## [54] "2. 9."
## [55] "Determination of GSH, TSH and MDA"
## [56] "The levels of GSH and TSH in lung supernatant were examined using the modified Ellman method b
## [57] "15"
## [58] "The levels of the TSH were measured at 412 nm based on the reduction of 5, 5 -dithiobis 2-nitro
## [59] "The levels of lipid peroxidation were reflected by the value of MDA, which was determined using
## [60] "16 2. 10."
## [61] "Determination of Hydroxyproline"
## [62] "The levels of tissue hydroxyproline (Hyp) were examined using the modified method of Edwards a
## [63] "17 2. 11."
## [64] "Statistics"
## [65] "Each of the experimental data was presented as mean± S. D. and compared to the blank control f
## [66] "A one-way analysis of variance (ANOVA) was calculated."
## [67] "When the F test from ANOVA was significant, least significant difference (LSD) t test was used
## [68] "Statistical significance was considered at p < 0 05."
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Xiaoli, F et al 2017.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "Materials"
word_end_mm <- "significant"

display_occurences(word_start_mm, conllu_df)
```

```
## [1]    53  1698 10838
```

```
## [1] "Materials & methods:"
## [2] "comfuture science group Prenatal exposure to nano ZnO in rats: neurotoxicity & postnatal impair
## [3] "Materials & methods •"
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1]  4366  4788  4848  5779  5936  7842  7932  8899 10922
```

```
## [1] "A p-value of <0. 05 was considered to be statistically significant."
## [2] "Regarding the coefficients of the lung, no significant change was found in the different groups
## [3] "Distribution pattern of zinc content Repeated exposure to ZnO NPs for 18 days in pregnant femal
## [4] "Prenatal exposure to ZnO NPs caused subtle but significant changes in these genes in brains of
## [5] "Furthermore, on training days 1-3, there was no significant difference (p > 0. 05) in escape la
## [6] "However, the authors found no significant differences in zinc content in the brain."
## [7] "Interestingly, the zinc concentrations in the blood of weaning pups and dams showed a minimal bu
## [8] "The results of 8-OhdG immunohistochemistry (Figure 8) also revealed that NP exposure would lead
## [9] "Prenatal exposure to ZnO NPs was found to result in a significant increase in the zinc concentra
```

```
mm_section<-cut_conllu_df(53, 4366, conllu_df)
```

```
##   [1] "Materials & methods:"
##   [2] "Pregnant Sprague-Dawley rats were exposed to ZnO nanoparticles (NPs) by gavage."
##   [3] "Toxicity was assessed including zinc biodistribution, cerebral histopathology, antioxidant sta
##   [4] "Results: A significantly elevated concentration of zinc was detected in offspring brains."
##   [5] "Transmission electron microscope observations showed abnormal neuron ultrastructures."
##   [6] "Histopathologic changes such as decreased proliferation and higher apoptotic death were obser
##   [7] "An obvious imbalanced antioxidant status occurred in brains."
```

```
##   [8] "Adult experimental offspring exhibited impaired learning and memory behavior in the Morris wa
##   [9] "Conclusion:"
##  [10] "These adverse effects on offspring brain may cause impaired learning and memory capabilities
##  [11] "First draft submitted: 17 November 2016; Accepted for publication: 20 January 2017; Published
##  [12] "neurotoxicity • oxidative stress • prenatal exposure •"
##  [13] "ZnO nanoparticles"
##  [14] "Along with the rapid development of nanotechnology, nanomaterials (NMs) are rapidly becoming
##  [15] "NMs are defined as materials composed of unbound particles or particles with at least one ext
##  [16] "Such particles normally possess typical nanostructure-dependent properties, including chemica
##  [17] "Thus, the toxicological properties of engineered nanoparticles (NPs) may greatly differ from
##  [18] "However, it should be noted that NPs are not inherently benign, and they may influence biolog
##  [19] "Consequently, the vast and increasing array of engineered NPs entering the environment and so
##  [20] "Zinc oxide NPs are among the NMs most widely incorporated into market goods."
##  [21] "Due to their unique photocatalytic, electronic, optical and dermatological properties, and hig
##  [22] "In recent years, more biomedical applications have considered ZnO NPs due to their antibacter
11 ]."
##  [23] "However, the increasing applications of ZnO NPs also increase the opportunity for human conta
##  [24] "Unfortunately, the safety evaluation of ZnO NPs currently lags far behind their emerging deve
##  [25] "Prenatal exposure to nanosized zinc oxide in rats: neurotoxicity and postnatal impaired learn
##  [26] ",1 1 Nanfang Hospital, Southern Medical University, Guangzhou 510515, China 2 School & Hospit
##  [27] "cn"
##  [28] "For reprint orders, please contact: reprints@futuremedicine."
##  [29] "com 10."
##  [30] "2217/nnm-2016-0397 Nanomedicine (Lond.)"
##  [31] "(Epub ahead of print) future science group Research Article Xiaoli, Junrong, Xuan et al."
##  [32] "After exposure, NPs can transfer throughout the body, deposit in target organs and further ca
##  [33] "Importantly, recent in vivo studies have demonstrated that ZnO NPs may transport across the b
brain barrier (BBB) into the brain [ 17 ] or even translocate through the olfactory nerve [ 18 ]."
##  [34] "As interest in their potential adverse effects increases, the toxic effects of ZnO NPs have b
23 ]."
##  [35] "At present, ZnO NPs are considered to be highly toxic, with the lowest LD 50 among the variou
##  [36] "Many recent reports have indicated that the mechanism of nanoZnO toxicity involves the induct
##  [37] "Although the literature on their biological effects is rapidly increasing, the knowledge on t
##  [38] "Moreover, it should be noted that the brain may be more vulnerable to NMs when exposures to s
##  [39] "Compared with adult brains, fetal brains are more likely influenced by blood-borne substances
##  [40] "In an animal study performed by Yamashita et al."
##  [41] "[ 29 ], pregnant BALB/c mice were injected with nanosilica particles, nano-TiO 2 and fullere
##  [42] "The author found that nanosilica and nano-TiO 2 accumulated in the placenta and fetal brain,
##  [43] "This finding was consistent with other reports, which also indicated that NMs may cross the p
##  [44] "In fact, reproductive and developmental toxicity has been integrated into the NM research str
##  [45] "However, compared with other toxicity studies, very little is known regarding the developmenta
##  [46] "In a recent study performed by Hong et al."
##  [47] "[ 33 ], the author found that oral administration of 400 mg/kg/day ZnO NPs in pregnant rats w
##  [48] "Similar toxicity was also reported in the embryos and larvae of zebra fish when exposed to na
##  [49] "However, the potential adverse effects on postnatal development and behavior in adulthood wer
##  [50] "In another in vivo study, subcutaneous administration of ZnO NPs to the pregnant ICR (Institu
##  [51] "One major limitation of this literature is that the toxicity mechanism was not involved in th
##  [52] "Considering the current limited understanding on the developmental neurotoxicity of ZnO NPs,
##  [53] "Compared with inhalation or skin exposure, the oral intake of NPs in food-related products ha
##  [54] "To date, people possess a higher chance of being exposed to ZnO NPs than other NPs in food in
##  [55] "For this reason, the present study exposed pregnant Sprague-Dawley (SD) rats to fully charact
##  [56] "There is a possibility in ZnO NPs of gaining a direct entry into the body or entering the GI
##  [57] "The doses at which humans may get exposed to these NPs at the above-mentioned scenario are ye
##  [58] "So we choose to work with the toxicological standard recommended by KFDA (Korea Food and Drug
```

```
##  [59] "Furthermore, in our recent toxicology review of ZnO NPs in 2016"
##  [60] "[ 40], we summarized the presence of ZnO through the gastrointestinal absorption following ora
5 g/kg BW in some studies."
##  [61] "The highest dose of 5 g/kg BW was selected on the basis of the evaluation of Scientific Commit
##  [62] "The dose level used in the current study can be used to assess the health risks from exposure
##  [63] "In this study, the potential neurotoxic effects on the offspring of the female rats were asses
##  [64] "Furthermore, a neurobehavioral test, the Morris water maze (MWM), was performed in 8-week-old
##  [65] "We hope that the present study will contribute to a better understanding of the nano-related i
##  [66] "10."
##  [67] "2217/nnm-2016-0397 www."
##  [68] "futuremedicine."
##  [69] "comfuture science group Prenatal exposure to nano ZnO in rats: neurotoxicity & postnatal impai
##  [70] "The ZnO NPs used in this study were a type of nanopowder and purchased from Sigma-Aldrich (USA
##  [71] "The primary particle size and morphology were determined using a transmission electron microsc
##  [72] "The hydrodynamic size and  potential of NPs in distilled water (DW) were determined using a M
##  [73] "Additionally, x-ray-diffraction patterns were obtained at room temperature using a RAX-10 diff
##  [74] "The specific surface area was measured through N 2 adsorption-desorption Brunauer-
Emmett-Teller adsorption analysis on a Micromeritics ASAP 2010M+ C instrument (Micromeritics Co., USA)."
##  [75] "Finally, endotoxin contamination levels were measured using an E-Toxate Kit (Sigma-Aldrich, US
##  [76] "Suspensions of ZnO NPs were prepared using saline containing 0. 05% v/v Tween-80."
##  [77] "The suspensions was sonicated for 30 min using a sonicator and mechanically vibrated for 5 min
##  [78] "Durability of ZnO NPs in acidic fluid To evaluate the durability of ZnO NPs in biological cond
##  [79] "AGF was prepared according to a previously described method [ 42 ], and it contained 0. 2% sod
##  [80] "Suspending agent (saline with 0. 05% Tween 80) was chosen as the control solution."
##  [81] "NPs at 5 mg/ml were incubated in solutions for 15, 25, 45 and 60 min, and 2, 4, 6, 12 and 24 h
##  [82] "After incubation, the nano-ZnO suspensions were centrifuged at 15, 000 ×g for 30 min, then the
##  [83] "Inductively coupled plasma-mass spectrometry (ICP-MS; Thermo Fisher Scientific, USA) was used
##  [84] "Animals & treatment Pregnant SD rats (220 ± 30 g) at GD 1 were purchased from the Animal Cente
##  [85] "The animals were housed in stainless steel cages in a ventilated animal room under 12 h of lig
65% relative humidity."
##  [86] "Sterilized food and DW were provided ad libitum."
##  [87] "All procedures used in the experiment were performed in compliance with the regulations and gu
##  [88] "The approval number provided by the ethical committee was NFYY-2014-058."
##  [89] "The day after arrival (GD 2), the animals were weighed and randomly divided into two groups of
##  [90] "The ZnO NP suspension and the vehicle were administered to the pregnant rats by intragastric g
19)."
##  [91] "The mortality and clinical behavior were observed every day during the 18 days."
##  [92] "On the last day of exposure (GD 19), the female rats were housed alone and monitored for birth
##  [93] "The first day of delivery was considered to be postnatal day (PND) 0 for the pups."
##  [94] "On PND 2, pups from eight litters (3-4 pups per litter) were selected and sacrificed by decapi
##  [95] "Different tissues were collected and weighed to calculate tissue coefficients."
##  [96] "In addition, the brain samples were used for histopathological and ultrastructural examination
##  [97] "At weaning (PND 21), one pup per litter (n = 8) was randomly selected, anesthetized with 40 mg
##  [98] "Blood (stabilized in 0. 17 mol/l K 2 EDTA) was thereafter quickly collected via the abdominal
##  [99] "Brain tissues from weaned offspring (n = 5, one pup per litter) were also collected, immediate
## [100] "Finally, one male and one female pup from each litter (n = 6) were randomly selected for behav
## [101] "A schematic representation of the animal protocol was provided in Figure 1."
## [102] "Coefficients of tissues"
## [103] "On PND 2, the male pups were counted and removed from each litter (3-4 per litter, n = 8), wei
## [104] "The hearts, livers, spleens, lungs, kidneys and brains were dissected and immediately rinsed i
## [105] "After weighing the bodies and tissues, the coefficients of different tissues to BW were calcul
## [106] "ICP-MS elemental analysis"
## [107] "To evaluate tissue biodistribution, samples from the heart, liver, spleen, lung, kidney and bi
## [108] "The blood samples were collected from male weaned pups (PND 21) and the dams (n = 8)."
```

```
## [109] "The zinc concentrations were quantified by ICP-MS."
## [110] "Briefly, collected tissues (200 mg per sample) and blood (0. 1 ml per sample) were predigested
## [111] "After add 10."
## [112] "(Epub ahead of print)"
## [113] "Figure 1. Schematic illustrations of experiments on the neurotoxicity of rat offspring after p
## [114] "Pregnant rats were given nano-ZnO by gavage from GDs 2-19 once daily."
## [115] "The 2-day-old pups were sacrificed, and organs and blood were collected."
## [116] "The blood of dams and weaned pups was also collected on PND 21."
## [117] "The offspring were subjected to a Morris water maze to test their learning and memory ability
## [118] "GD: Gestational day; PND: Postnatal day."
## [119] "Gestational days (GDs) 2-19"
## [120] "The females were given ZnO NPs in oral administration once daily (n = 10)"
## [121] "Heart, liver, spleen, lung, kidney, brain and blood were collected in newborns (n = 8)"
## [122] "Brains were collected in pups (n = 5); blood was collected in dams and pups (n = 8)"
## [123] "The offspring were subjected to the Morris water maze (n = 7 for males and females) n = 8-
12 per littler"
## [124] "The rst day of delivery was considered as postnatal day (PND) 0 PND 2 PND 21 PND 60 future sci
## [125] "ing 0. 5 ml of hydrogen peroxide (H 2 O 2 ), the mixed solutions were heated at 160°C using a
## [126] "The solutions were heated again at 120°C until the remaining nitric acid was evaporated almost
## [127] "The resulting solutions were finally diluted to 2 ml with 1% nitric acid + 0. 1%"
## [128] "Triton-100 and analyzed using ICP-MS."
## [129] "Histopathological examination Brain histological analysis was performed in the male pups (n =
## [130] "Then, the samples were dehydrated in a series of ethanol and xylene solutions, embedded in pa
## [131] "Representative photos were captured using a charge-coupled device digital camera fixed to a li
## [132] "Furthermore, the presence of apoptotic and proliferating cells in brain sections was analyzed
## [133] "The oxidative damage to nucleic acids of brain cells was analyzed by 8-hydroxy-2-deoxyguanosin
## [134] "Photographs of positive staining (brown) were captured at randomly selected fields both in the
## [135] "The integrated optical density value of each image was analyzed using image proplus 6. 0 softw
## [136] "Regions without yellow ( or brown) staining served as background."
## [137] "All images were captured with uniform threshold and intensity settings."
## [138] "TEM analysis"
## [139] "Brains were quickly removed from 2-day-old pups from five litters (n = 5) and fixed in 3% glut
## [140] "3 cacodylate buffer (pH 7."
## [141] "2-7. 4) and postfixed in 1% osmium tetroxide for 1 h."
## [142] "The samples were then dehydrated using a series of ethanol concentrations and embedded in Epon
## [143] "Finally, an H-7500 TEM (Hitachi, Japan) was used to visualize the ultrathin sections (500 nm)
## [144] "10. 2217/nnm-2016-0397 www. futuremedicine."
## [145] "comfuture science group Prenatal exposure to nano ZnO in rats: neurotoxicity & postnatal impai
## [146] "The frozen brain samples collected from pups in PND 2 were homogenized in phosphate-buffered s
## [147] "The homogenate (1:9 w/v) was centrifuged at 2500 × g r. p. m. for 10 min at 4°C."
## [148] "After centrifugation, the supernatant was collected."
## [149] "O 2 - in the brain tissue was measured using dichlorofluorescein-diacetate."
## [150] "Briefly, fresh 5% homogenate was added to a 96-well plate (190  l per well), and then 10  l of
## [151] "The plate was incubated at 37°C for 30 min in the dark."
## [152] "The fluorescence intensity was measured using a microplate reader (Molecular Device, USA)."
## [153] "Oxidative stress markers (superoxide dismutase [SOD], malondialdehyde [MDA] and glutathione pe
## [154] "SOD activity was determined using the xanthine oxidase method."
## [155] "One unit of SOD activity is the amount of SOD necessary to cause 50% inhibition of the product
## [156] "Lipid peroxidation of the brain was determined as the concentration of MDA, which could react
## [157] "GSH-PX activities were determined following the changes in the velocity of the catalytic react
## [158] "Protein content was measured using Coomassie blue staining."
## [159] "Total RNA extraction & real-time PCR Brain samples were obtained from newborns (PND 2) and off
## [160] "Frozen tissues were homogenized and extracted using TRIzol reagent (Gibco, USA) according to t
## [161] "The extracted RNA was further purified using Qiagen  RNeasy Mini columns to remove any genomic
```

```
## [162] "The total RNA was measured at 260 and 280 nm using an M 5 spectrophotometer (Molecular Devices
## [163] "The purity of the RNA sample was measured as the 260/280 nm ratio with expected values betwee
## [164] "Realtime PCR (RT-PCR) was conducted using a commercial kit (SYBR Premix Ex Taq II, TaKaRa) an
## [165] "The expression level of each target gene was normalized to its Actb mRNA content."
## [166] "The primers for PT-PCR are shown in Table 1. MWM test The MWM consisted of a large circular p
## [167] "The sequence of primers used in the RT-PCR analysis."
## [168] "Gene name Description Primer sequence Primer size (bp) Actb Actin, beta-F GAGAGGGAAATCGTGCGTGA
## [169] "1 Superoxide dismutase 1-F GGTCCACGAGAAACAAGATGA Superoxide dismutase 1-R CAATCCCAATCACACCACAA
## [170] "Fmo 2 Flavin containing monooxygenase 2-F CCTGGAAGACTCGCTTGTTA Flavin containing monooxygenase
## [171] "Thioredoxin interacting protein-F TGCCTCTCTGCTTGAACTCC Thioredoxin interacting protein-R CCTAC
## [172] "Alox 12 b Arachidonate 12-lipoxygenase, 12 R type-F ACTCTTCGCTGTCTGGCTTC Arachidonate 12-lipox
## [173] "ture, 26°C), made opaque with added black ink."
## [174] "The pool was divided into four quadrants: northeast (NE), southeast, southwest and northwest.
## [175] "A circular platform (diameter, 10 cm) was placed in the middle of the NE quadrant and 2 cm bel
## [176] "Visual cues were placed on the walls surrounding the pool."
## [177] "Adult rats (PND 60) received four trials a day from four starting positions with a 5-min inter
## [178] "During each trial, the rats were required to find the hidden platform and spend 15 s on the pl
## [179] "If they failed to find the hidden platform within 90 s, the rats were guided and placed on the
## [180] "Two parameters, including the number of crossings of exact position of the former platform and
## [181] "On days 7 and 8 (reacquisition phase), reverse platform training was instituted, and the hidde
## [182] "Reversal trials were performed as described in the acquisition phase."
## [183] "Throughout the training and testing, the latency to reach the platform, swim paths and the rel
## [184] "Statistical analysis"
## [185] "All results are expressed as the mean ± standard error of the mean."
## [186] "Analysis of variance with repeated measures was applied in the MWM test."
## [187] "A randomized block design (one-way analysis of variance), taking into account litter effects,
## [188] "Unpaired Student's t-tests were applied in other comparisons between two groups."
## [189] "All analyses were performed in SPSS 19. 0 software."
## [190] "A p-value of <0. 05 was considered to be statistically significant."
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Yan, M et al 2011.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "Materials"
word_end_mm <- "test"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 1713
```

```
## [1] "Materials and methods .1."
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1] 4188
```

```
## [1] "Data were presented as ean ± SD and evaluated for statistical significance by one- or two-way an
```

```
mm_section<-cut_conllu_df(1713, 4188, conllu_df)
```

```
##   [1] "Materials and methods .1."
##   [2] "Chemicals and materialsCadmium chloride (CdCl 2, 99. 0%), tellurium powder (Te, 99."
##   [3] "5%) were from lfa Aesar (London, UK)."
##   [4] "Sodium borohydride (96%), hydrochloric acid (HCl) and odium hydroxide (NaOH) were from Sinopha
##   [5] "All chemicals were used without further purification."
```

```
##   [6] "Solutions were repared using Milli-Q water (Millipore) as the solvent."
##   [7] "MSA was from Sigmaldrich Crop. (St. Louis, USA)."
##   [8] "Water-soluble (carboxyl coated) CdTe quantum 282 (2011) 94-103 95 dots were purchased from Zh
##   [9] "Medium 200 (M 200), low serum growth supplement (LSGS), fetal bovine serum (FBS), calcein-AM,
##  [10] ",3, 3 -tetraethylbenzimidazolylcarbocyanine iodide (JC-1) were from Invitrogen Corp. (Carlsbad
##  [11] "The antibodies including anti-human Bcl-2, antihuman Bax, anti-human cytochrome c, anti-human
##  [12] "The reactive oxygen species (ROS) assay kit (containing 2 "
##  [13] ",7 -dichlorfluorescein-diacetate, DCFHDA, and positive control agent Rosup) as well as N-acety
##  [14] "The 3, 3, 5, 5 tetramethylbenzidine (TMB) Liquid Substrate kit was from Amresco, Inc. (Solon,
##  [15] "The secondary antibodies and other biological reagents were from SigmaAldrich Crop. (St. Louis
##  [16] "2. 2. Preparation and characterization of quantum dots"
##  [17] "The CdTe QDs were synthesized based on the reaction of CdCl 2 and NaHTe solution with MSA as s
##  [18] "First, Cd 2+-MSA precursor solution was prepared by dissolving CdCl 2 (0. 184 g) and MSA (0."
##  [19] "183 g) in 60 mL deionized ultrapure water."
##  [20] "The pH of the solution was adjusted to 6."
##  [21] "5 using 1 M NaOH solution."
##  [22] "This mixture was degassed and kept stirring at room temperature."
##  [23] "Then, NaHTe solution (4 mL, 0. 05 mM), which was prepared by dissolving"
##  [24] "Te powder in NaBH 4 solution, was injected into the precursor solution."
##  [25] "The reaction between cadmium ions and NaHTe took place immediately after the injection of NaH
##  [26] "The molar ratio of Cd 2+:NaHTe:MSA was 1:0."
##  [27] "2:1. 2. Next, the CdTe precursor crude solution was refluxed under open-air condition at 100
##  [28] "The as-prepared QDs were precipitated with an equivalent amount of 2-propanol, and then resusp
##  [29] "The pellet of purified QDs was dried overnight at room temperature in vacuum, and the final pr
##  [30] "Stock solutions of QDs for our following biological experiments were prepared by dissolving Cd
##  [31] "The high resolution transmission electron microscopy (HRTEM) image was taken by a JEM 2010 mic
##  [32] "The absorption spectrum and emission spectrum were recorded with a UV 2550 scanning spectrophe
##  [33] "To measure their photostability in physiological environment and weak acidic environment, our
##  [34] "Then the QD solutions were added into the microplate and their emission peak intensities were
##  [35] "To determine the composition of MSA-capped CdTe QDs, the concentration of mercaptosuccinate ca
##  [36] "The concentration of cadmium from QDs was measured by graphite furnace atomic absorption spect
##  [37] "To measure the free cadmium ions released from CdTe QDs, 2 mL CdTe QDs stock solution (1 mg/ml
##  [38] "Ltd.) and then placed in polypropylene beakers containing 500 mL ultra pure water."
##  [39] "The dialysis experiment lasted for 30 h, and 2 mL dialysis solution was taken at the 6 th, 18
##  [40] "Finally, the collected samples were analyzed using GFAA."
##  [41] "All the measurements were performed at room temperature and in triplicate."
##  [42] "2. 3. Cell culture and treatments Primary HUVECs were obtained from human umbilical cord veins
##  [43] "In brief, the vein was cannulated and filled with 0. 1% type I collagenase."
##  [44] "After incubation (37 ∘C, 15 min), the obtained pellets were centrifugated and resuspened in M
##  [45] "The cells were used at passage 2-4 in our study."
##  [46] "After reaching 90% confluence, M 200 with LSGS was removed and then HUVECs were cultured with
##  [47] "Untreated control cells were incubated with medium (also containing 2% FBS) alone."
##  [48] "2. 4."
##  [49] "Cell viability assessment"
##  [50] "The viability of HUVECs treated with CdTe QDs was estimated with both 3-(4, 5 dimethylthiazoly
##  [51] "For the MTT assay, after reaching 90% confluence, culture medium was removed and then HUVECs
100 g/mL for 24 h. Then, the medium was removed; 150 L MTT solutions (0. 5 mg/mL) was added to each well
##  [52] "The amount of viable cells in each well was determined by the absorbance of solubilized formaz
##  [53] "The optical density (O. D.) was measured at 96 M. Yan et al."
##  [54] "/ Toxicology 282 (2011) 94-103"
##  [55] "F t 4 e m i C 2 w k I F r c w 1 w t 2 T c w t fl Fig. 2. MSA-capped CdTe QDs characterizations
##  [56] "1. MSA-capped CdTe QDs characterizations. (A) HRTEM of CdTe QDs; (B) absorpion and PL spectrum
##  [57] "Bar = 5 nm. 90 nm with a microplate reader (Versa Max, Molecular Devices)."
##  [58] "For using a calcein ster green fluorescence assay, cells were washed with PBS twice after the
```

```
##  [59] "The fluorescence mages of cells were obtained with a laser scanning confocal microscope (LSM 5
##  [60] ".5. Detection of cellular apoptosis Exposed membrane phosphatidylserine (PS) in apoptotic cell
##  [61] "n brief, after 90% confluence, HUVECs were cultured with M 200 (containing 2% BS) in the prese
##  [62] "1-10 g/mL CdTe QDs for 24 h."
##  [63] "The treated cells were inse twice with PBS, harvested by trypsinization, and labeled with 5 g
##  [64] "The fluorescence of Annexin-V-positive HUVECs as measured with FACScalibur flow cytometer (Be
##  [65] "For statistical significance, at least 10, 000 cells ere analyzed in each sample."
##  [66] "The involvement of ROS in CdTe QDs-induced apoptosis was also examined by his FCM assay."
##  [67] "HUVECs were pre-incubated with a ROS scavenger NAC (1 mM) for h; and then with or without a QD
##  [68] "Then, after treatment, he apoptotic HUVECs were also detected on the FL-1 detector of the FACS
##  [69] "with different pH values were measured at 405 nm excitation and 599 nm emission every 20 min i
##  [70] "2. 6."
##  [71] "Evaluation of intracellular ROS and mitochondrial membrane potential ( m)"
##  [72] "DCFH-DA was used for ROS detection."
##  [73] "DCFH-DA is transported across the cell membrane and cleaved by nonspecific esterases to form D
##  [74] "HUVECs were rinsed twice with PBS, trypsinized and stained with DCFH-DA (5 M) after 12 or 24 h
##  [75] "The green fluorescence of DCF was detected on the FL 1 detectors in a FACScalibur flow cytomet
##  [76] "HUVECs stimulated with Rosup (50 g/mL) for 30 min were taken as positive controls of ROS gener
##  [77] "To explore the effect of NAC on potential ROS elevation which caused by CdTe QDs, HUVECs were
##  [78] "The DCF fluorescence intensity was also measured with flow cytometry."
##  [79] "For quantitative analysis of m , QDs-treated HUVECs were also rinsed twice with PBS, trypsiniz
##  [80] "JC-1 is a lipophilic probe which potential-dependently accumulated in mitochondria and its flu
##  [81] "The fluorescence of J-monomers was quantified on FL 1 detector of FACScalibur flow cytometer."
##  [82] "The negative control for flow cytometer compensation was prepared by addition of m disrupter m
##  [83] "To further clarify the role of QDs-caused ROS in induction of mitochondrial depolarization, HU
##  [84] "Subsequently, the cells were rinsed, trypsinized, loaded with JC-1 and analyzed by flow cytome
##  [85] "For statistical significance, at least 10, 000 cells were analyzed in each sample. M. Yan et a
103 97"
##  [86] "F with fl inatio 2 M s w G 2 t 1 w i I d m 2 l t l r l e d w w c o k 2 m o ig."
##  [87] "3. CdTe QDs caused decline of cell viability in HUVECs."
##  [88] "HUVECs were treated uorescence staining; and (B) MTT assay."
##  [89] "Data represent mean ± SD of four determ .7. Visualization of mitochondria structure"
##  [90] "To examine mitochondrial structure in HUVECs treated with QDs for 24 h, itoTracker Red FM (0.
##  [91] "After a hort incubation period of 15 min, the fluorescence images of mitochondria ere obtained
##  [92] ".8."
##  [93] "Cytochrome c immunostaining in QDs-treated HUVECs"
##  [94] "The release of cytochrome c from mitochondria was detected using immunosaining analysis."
##  [95] "Briefly, HUVECs were washed twice with PBS and loaded with 00 nM MitoTracker Red FM for 20 min
##  [96] "Then, cells ere fixed with 3. 7% paraformaldehyde and permeabilized for immunostainng."
##  [97] "Mouse polyclonal anti-cytochrome c (1:200) and FITC-conjugated mouse gG (1:500) were used."
##  [98] "The translocation of cytochrome c from mitochonria to cytoplasm was observed by a Carl-Zeiss D
##  [99] "Immunoblot analysis of cellular expression of apoptotic proteins HUVECs exposed to CdTe QDs fo
## [100] "After emoving cellular debris by centrifugation (13, 000 rpm, 10 min, 4 ∘C), the cell ysates w
PAGE) and subsequently transferred onto polyvinylidene ifluoride (PVDF) membranes (Millipore)."
## [101] "Blocking of unspecific binding sites ith nonfat dry milk (5% in Tris-buffered saline containi
## [102] "Statistical analysis Experiments were performed at least three times."
## [103] "Data were presented as ean ± SD and evaluated for statistical significance by one- or two-way
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Yu, T et al 2011.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "METHODS"
word_end_mm <- "software"
```

```
display_occurences(word_start_mm, conllu_df)
```

```
## [1] 7273
```

```
## [1] "METHODS Synthesis of Nonporous and Mesoporous"
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1] 9242
```

```
## [1] "The LC 10 values in the hemolysis assay were determined by using ED 50plus v 1. 0 software."
```

```
mm_section<-cut_conllu_df(7273, 9242, conllu_df)
```

```
##   [1] "METHODS Synthesis of Nonporous and Mesoporous"
##   [2] "SiO 2."
##   [3] "Nonporous silica nanoparticles (Stöber) were produced using the modified Stö- ber method."
##   [4] "23 Water (34. 82 mL), 3. 25 mL of ammonium hydroxide (29. 7%), and 100 mL of ethanol were mix
##   [5] "Tetraethyl orthosilicate (6. 20 mL) was added at an injection rate of 5 mL/min upon stirring a
##   [6] "The reaction was conducted for 1 h, and the product was washed twice by ethanol and stored in
##   [7] "Mesoporous"
##   [8] "SiO 2 of different shapes were synthesized through a one-step condensation under dilute silica
##   [9] "24 29 Generally, cetyltrimethylammonium bromide was dissolved in aqueous medium with mild heat
##  [10] "After the solution was cooled to room temperature (22 C), aqueous ammonia was introduced and t
##  [11] "TEOS was added at the rate of 5 mL/min while the stirring continued."
##  [12] "Themixture was further stirred for 4 h, and the product was autoclaved at 100 C for 24 h. 30,
##  [13] "As-synthesized nanoparticles were suspended in ethanolic HCl (1. 5 mL of HCl in 150 mL of etha
##  [14] "The complete removal of CTAB was confirmed by Fourier transform infrared (FT-IR) spectroscopy
##  [15] "Surface Functionalization."
##  [16] "To modify the surface of"
##  [17] "SiO 2 with primary amine functionalities, 33 100 mg of SiO 2 was resuspended in 100 mL of anhy
##  [18] "The mixture was stirred at room temperature for 20 h. Aminemodified"
##  [19] "SiO 2 were collected by centrifugation and washed extensively with ethanol and water."
##  [20] "SiO 2 were stored in ethanol at 4 C and transferred to water immediately before use."
##  [21] "Nanoparticle Characterization."
##  [22] "Transmission electron microscopy images were taken with a Philips"
##  [23] "Tecnai microscope operating at 120 kV."
##  [24] "FT-IR spectra were recorded on a Varian Cary FT-IR 1000 spectrometer using KBr pellets."
##  [25] "X-ray diffraction patterns of SiO 2 were analyzed on a Philips"
##  [26] "PANalytical X0Pert X-ray diffractometer (Spectris, England) using Cu Ka radiation (  = 0. 1542
##  [27] "The XRD spectra were recorded in the 2  range of 2 10with a step size of 0. 02 in a 2  scatter
##  [28] "The slit sizes and specimen length were also adjusted for divergence slit, antiscattered slit
##  [29] "Nitrogen adsorption desorption isothermmeasurementswere completed on aMicromeritics ASAP 2010
##  [30] "The SiO 2 were dried at 100 C overnight before analysis."
##  [31] "The Brunauer Emmett Teller specific surface areas were calculated by using adsorption data at
##  [32] "The external surface areas of mesoporous"
##  [33] "SiO 2 were calculated from the t plots of their N 2 adsorption isotherms."
##  [34] "32 Pore volume and pore size distributionswere obtained froman adsorption branch by using the
##  [35] "19, 21 Acute Cytotoxicity Assay."
##  [36] "The acute toxicity effect of SiO 2 was determined by the WST-8 assay on A 549 cells or RAW 264
##  [37] "Cells from passages 5 through 20 were used with medium changing once every three days."
##  [38] "A 549 cells or RAW 264. 7 macrophages were seeded at 8000 cells/well or 16 000 cells/well in a
##  [39] "SiO 2 at incremental concentrationof 100, 250, or 500 g/mLwere added to cells."
##  [40] "Supernatants TABLE 5. Summary of LC 10 Values of SiO 2 in Human Erythrocytesa LC 10 values ( g
##  [41] "SiO 2 and that of amine-modified counterparts (p < 0. 001)."
##  [42] "A RTIC LE YU ET AL . VOL."
```

```
##  [43] "5 ’ NO. 7 ’ 5717-5728 ’ 2011 www."
##  [44] "acsnano."
##  [45] "org 5727 from nanoparticle stock solutions and respective growth medium served as controls."
##  [46] "Post 24 h, old medium was aspirated, and cells werewashed three timeswith PBS."
##  [47] "A 100  L amount of complete medium containing 10% (v/v) Cell Counting Kit-8 (Dojindo, Rockvill
##  [48] "The absorbance of the plate was recorded at 450 nm on a UV/vis reader with a reference wavele
##  [49] "Proliferation Inhibition Assay."
##  [50] "The cytotoxicity of SiO 2 was evaluated by the WST-8 viability assay on A 549 cells or RAW 264
##  [51] "Initially, A 549 or RAW cells were seeded at 2000 or 4000 cells/well in a 96-well plate and al
##  [52] "Then 10, 50, 100, 250, 500, or 1000  g/mL of bare"
##  [53] "SiO 2 or amine-modified"
##  [54] "SiO 2 was added into the 96-well plate in triplicates."
##  [55] "Supernatants from nanoparticle stock solutions and respective growth medium served as controls
##  [56] "Post 72 h, old medium was aspirated, and the following steps were the same as the procedures i
##  [57] "Plasma Membrane Integrity Assay."
##  [58] "Determination of propidium iodide uptake was used to assess the integrity of the plasma membra
##  [59] "A 549 cells or RAW cells were seeded at 8 104 or 1. 6 105 cells/well on a 12-well plate in tri
##  [60] "After 24 h, selected nanoparticles were added into each well at the concentration of 250  g/mL
##  [61] "Twenty-four hours later, cells and medium from each well were collected into a 5 mL flow cytom
##  [62] "The cell suspension was centrifuged at 1000 rpm for 5 min, the supernatant was decanted, and t
##  [63] "A 5  L portion of propidium iodide solution (50  g/mL in water) was added to each tube."
##  [64] "The tube was gently vortexed and incubated for 15 min at room temperature in the dark."
##  [65] "Then 400  L of PBSwas added into each tube, and the samples were analyzed by flow cytometry (F
##  [66] "Quantitation of Cellular Association."
##  [67] "Cellular association of SiO 2 was evaluated on A 549 and RAW 264. 7 cells."
##  [68] "A 549 or RAW cells were seeded at 8 104 or 1. 6 105 cells/well on a 12-well plate in triplicat
##  [69] "Cells were incubated with 100  g/mL of SiO 2 for 24 h (37 C, 5% CO 2)."
##  [70] "After cell/particle incubation, the old medium was aspirated and the cells were washed three t
##  [71] "Then the cells were treated with 0. 5 mL of 0. 1% (v/v) Triton X-100 solution in water for 15
##  [72] "After that, the cell lysate was collected into a centrifuge tube, and the wells were further w
##  [73] "The wash was also collected into the same centrifuge tube."
##  [74] "Aliquots of 100  L of cell lysate were used for protein content measurement by BCA assay (Ther
##  [75] "The concentration of silicon in the cell lysate was measured by direct Si measurement using in
##  [76] "The amount of cellular-associated"
##  [77] "SiO 2 post 1 h incubation at 4 or 37 C was also measured."
##  [78] "RAW cells were seeded at 3. 2 105 cells/well on a 12-well plate in triplicate and incubated fo
##  [79] "After that, cells were preconditioned to 4 C by incubating at 4 C for a brief period of 10 min
##  [80] "Then silica nanoparticles were added to the cells at a concentration of 100  g/mL, and the cel
##  [81] "The treatment that followed was the same as mentioned above."
##  [82] "To make sure that relative cell viability in the 4 C treated plate was not dramatically influe
##  [83] "Hemolysis."
##  [84] "Heparin-stabilized human blood was freshly collected according to an approved University of Ut
##  [85] "36, 43 A 4 mL sample of whole blood was added to 8 mL of Dulbecco's phosphate-buffered saline
##  [86] "The RBCs were further washed five times with sterile D-PBS solution."
##  [87] "Following the last wash, the RBCs were diluted to 40 mL of D-PBS."
##  [88] "Then 0. 2 mL of the diluted RBC suspension was added to 0. 8 mL of the silica nanoparticle sus
##  [89] "All samples were prepared in triplicate, and the suspension was briefly vortexed before leavin
##  [90] "After that, the mixture was briefly vortexed again and centrifuged at 10016 g for 3 min."
##  [91] "A 100  L amount of supernatant from the sample tube was transferred to a 96-well plate."
##  [92] "The absorbance value of hemoglobin at 577 nm was measured with the reference wavelength of 655
##  [93] "A 0. 2 mL amount of diluted RBC suspension incubated with 0. 8 mL of D-PBS and 0. 8 mL of wate
##  [94] "The percent of hemolysis was calculated as follows:"
##  [95] "Hemolysis % = [(sample absorbance negative control)/ (positive control negative control)] 100%
##  [96] "Statistical Analysis."
```

```
##  [97] "The difference between multiple groups was analyzed by one-way ANOVA."
##  [98] "The Tukey post test was used where a difference was detected."
##  [99] "For two-group comparison, Student's t-test was used."
## [100] "The difference between two groups was considered significant when p < 0. 05."
## [101] "The LC 10 values in the hemolysis assay were determined by using ED 50plus v 1. 0 software."
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Zook, J M et al 2011.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "Materials"
word_end_mm <- "centrifugation"

display_occurences(word_start_mm, conllu_df)
```

```
## [1]  1739  1742 11340 11634
```

```
## [1] "Materials and methods Materials Certain commercial equipment, instruments or materials are iden
## [2] "Materials and methods Materials Certain commercial equipment, instruments or materials are iden
## [3] "Adv Materials 21(16):1549-1559."
## [4] "J Materials Chem 20(3):512-518."
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1] 1366 4766 4790
```

```
## [1] "Furthermore, the simplest fractionation techniques (e. g., centrifugation or filtration) have g
## [2] "The absorbance of the supernatant of each AgNP without blood was also measured after centrifuga
## [3] "The absorbance of the supernatant of each AgNP without blood was also measured after centrifuga
```

```
mm_section<-cut_conllu_df(1739, 4790, conllu_df)
```

```
##   [1] "Materials and methods Materials Certain commercial equipment, instruments or materials are id
##   [2] "Such 518 J. M. Zook et al."
##   [3] "identification does not imply recommendation or endorsement by the National Institute of Stan
##   [4] "Dulbecco's Modified Eagle Medium (DMEM) with 4. 5 g/L glucose and sodium pyruvate but without
##   [5] "Bovine serum albumin (BSA) was from Sigma Aldrich (St Louis, MO, USA) (‡ 96%, essentially fat
##   [6] "The antibiotics streptomycin and penicillin were from Invitrogen and were added at 100 mg/mL a
##   [7] "DMEM + 4% BSA by weight was filtered through a 0. 2 mm polypropylene filter to remove large pi
##   [8] "Citrate-stabilized 30 nm gold nanoparticles (AuNPs) were the standard reference material 8012
##   [9] "Citrate-stabilized silver nanoparticles (~ 23 nm intensity-weighted mean diameter in water by
##  [10] "2010)."
##  [11] "The silver nanoparticles were concentrated to ~ 1. 1 mg/mL by stirred cell ultrafiltration thr
##  [12] "For the hemolysis experiments, horse blood anticoagulated with heparin was obtained from Hemos
##  [13] "Cerium oxide (CeO 2) NPs with a nominal size of 64 nm specified by the manufacturer were creat
##  [14] "Aminated polystyrene NPs with a nominal size of 57 nm were created by Bangs Laboratories, Inc
##  [15] "Nanoparticle dispersion"
##  [16] "NP agglomerates are prepared with the following two procedures as depicted in Figure 1. In the
##  [17] "In this case, NPs are stabilized by BSA before they agglomerate to large sizes, with faster mi
##  [18] "In the second procedure, Figure 1 b, the NP solution is pipetted into DMEM without BSA while v
##  [19] "In this case, the high ionic strength causes the NPs to agglomerate quickly, and they are allo
##  [20] "After the NPs agglomerate to the desired size, BSA is added to coat the NP agglomerates and pr
##  [21] "Specifically, as described in Figure 1 a, to obtain the '0 s' agglomerates (singly dispersed o
##  [22] "Next, 0. 2 mLNP solution was added with a pipette to this solution while the vortexer was set
##  [23] "To make the 1 s and 5 s agglomerates, as described in Figure 1 b, 0. 8 mL DMEM was added to a
##  [24] "The tube was held in place on the vortexer in a foam holder, the vortexer was turned on at 115
```

```
##  [25] "1. Inject NP solution 2. Wait for agglomeration 3."
##  [26] "Inject DMEM/BSA DMEM + 2 % BSA Vortex Vortex Vortex DMEM DMEM/ NPs a b Figure 1."
##  [27] "Methods used to produce different agglomeration states of nanoparticles in this work."
##  [28] "(a) To minimize agglomeration, a welldispersed aqueous NP solution was pipetted into the cell
##  [29] "(b)"
##  [30] "To achieve different mean agglomerate sizes, a well-dispersed aqueous NP solution was first pi
##  [31] "Then, the particles were allowed to agglomerate over time to the desired size before adding 2%
##  [32] "Stable nanoparticle agglomerates 519 centrifuge tube."
##  [33] "The tube was then vortexed while adding 0. 2 mL NP solution for about 2 s. After allowing the
##  [34] "For the polystyrene NPs, either the original 10 mg/mL solution of NPs in water was pipetted in
##  [35] "During stability measurements, the samples were kept in the 15 mL centrifuge tubes in a dark 3
##  [36] "Spectrophotometry"
##  [37] "Spectrophotometric absorbance measurements between 300 nm and 1100 nm were made using a Hewlet
##  [38] "The absorbance of the 4. 8 mg/mL AuNPs and 6. 7 mg/mL AgNPs was measured directly in a 1 cm cu
##  [39] "The 110 mg/mL AgNP solution was diluted 20 with deionizedwater and the absorbance wasmeasured
##  [40] "Dynamic light scattering (DLS)"
##  [41] "DLS was performed using a Brookhaven Instruments (Holtsville, NY, USA) ZetaPALS with a 660 nm
##  [42] "Samples were measured for 100 s. The autocorrelation curve was exported for further analysis i
##  [43] "Because the 4% BSA in DMEM solution was filtered through a 0. 2 mm filter, all large agglomera
##  [44] "Simultaneously fitting two exponential decay curves (A 1 e -t/T 1 + A 2 e -t/T 2) was found to
##  [45] "94 ms, A 2 = 3. 495 106, and T 2 = 101 ms."
##  [46] "Since autocorrelation curves are additive, the curve for BSA can theoretically simply be subtr
##  [47] "However, the absolute intensity of the scattered light varies from run to run on many DLS inst
##  [48] "Therefore, the weighting of the BSA autocorrelation function was used as a fitting parameter
##  [49] "The first and second cumulants were fit to the NP part of the curve to provide a better fit to
##  [50] "This resulted in four simultaneous fitting parameters (B 1, B 2, T 3, and b) in Equation 1. B
##  [51] "The first cumulant of the NP scattering T 3 was used to calculate the intensity-weighted mean
##  [52] "( )2 where kB is Boltzmann's constant, T is the temperature in K, h is the viscosity, q = (2 p
##  [53] "The effect of BSA scattering on the measured size is most significant for small, unagglomerate
##  [54] "For example, 30 nm AuNPs dispersed at 4. 8 mg/mL in 2% BSA without agglomeration have a measur
##  [55] "5 ± 1. 0 nm with the BSA scattering included, but the diameter is 28."
##  [56] "2 ± 1. 2 nm after subtracting the scattering due to BSA."
##  [57] "However, when the same NPs are allowed to agglomerate even slightly to an average size of 45.
##  [58] "4 ± 1. 4 nm."
##  [59] "Similarly, for the smallest agglomerates of ~ 23 nm AgNPs dispersed at 6. 7 mg/ mL directly in
##  [60] "2 ± 2. 6 nmwith the BSA scattering included, but the diameter is 36. 7 ± 1. 7 nm after subtrac
##  [61] "For the same NPs dispersed at a higher concentration of 110 mg/mL with slightly larger agglome
##  [62] "For even larger agglomerates of AuNPs and AgNPs, the diameter is generally affected less than
##  [63] "Previous work has shown that rotation and vibration of the fractal-shaped agglomerates of NPs
##  [64] "This factor was found to be universally dependent on qR, where q is the scattering vector defi
##  [65] "1990)."
##  [66] "We fit a function to the published curve, finding 520 J. M. Zook et al. exp{0. 867/[ 1+4."
##  [67] "108*(qR)-2. 1 ]} provided a good fit for qR < 12."
##  [68] "Therefore, all measured effective hydrodynamic diameters were multiplied by this factor."
##  [69] "Although this universal curve is not applicable to single particles or very small agglomerates
##  [70] "The scattering vector q is a function of the angle at which scattering is measured, so rotatio
##  [71] "For the instrument used in this work, which measures scattering at 90 at a wavelength of 660 n
##  [72] "Calculating the size distribution from DLCA theory The compensated intensity-weighted hydrodyn
##  [73] "DLCA assumes the aggregation/agglomeration kinetics are such that particles permanently stick
##  [74] "Adding a high ionic strength solution sufficiently lowers the electrostatic repulsion between
##  [75] "When assuming DLCA, the Smoluchowski agglomeration rate equations predict that the particle ma
##  [76] "1990): N M N M M M( ) ( ) ( )= - -0 2 11 1 3 where N(M) is the number of agglomerates consisti
##  [77] "The number of particles in an agglomerate M can be converted into the geometric agglomerate di
##  [78] "The geometric diameter is directly related to the hydrodynamic diameter by the equation dh = (
```

```
##  [79] "The cluster number distribution described by Equation"
##  [80] "3 can be converted to the more common mass-weighted size distribution in several steps."
##  [81] "First, M can be calculated from the compensated intensity-weighted hydrodynamic diameter calcu
##  [82] "AFM methods and image analysis AFM Images were collected on a Dimension 3100 with a Nanoscope
##  [83] "Methodology followed the Nanotechnology Characterization Laboratory Assay Cascade (NIST-NCL P
##  [84] "Sample prep was carried out in a BSL-2 cabinet."
##  [85] "Approximately 20 mL of solution was incubated for 8 min (110 mg/mL AgNPs) or 30 min (6. 7 mg/
##  [86] "The 'Analyze Particles' method in ImageJ (http:// rsbweb. nih."
##  [87] "gov/ij/) was used to count semiautomatically the total number of particles in ten 20 20 mm ima
##  [88] "The images were first converted to black and white by setting the threshold at a level (65 ou
##  [89] "The threshold was kept identical for all images to maintain consistency."
##  [90] "After the images were converted to black and white, the 'Analyze Particles' method in ImageJ
##  [91] "Stable nanoparticle agglomerates 521 Hemolysis"
##  [92] "The hemolysis experiments were based on the ASTM protocol 'E 2524 - 08: Standard"
##  [93] "Test Method for Analysis of Hemolytic Properties of Nanoparticles', modified to produce multi
##  [94] "Instead of diluting the NP solution with Dulbecco's phosphate buffered saline as described in
##  [95] "Dilution series of AgNPs were created by combining the appropriate amounts of DMEM + 2% BSA w
##  [96] "As a control, the blood was added to DMEM + 2% BSA without AgNPs."
##  [97] "The dispersed AgNPs were stored at room temperature at 110 mg/mL for 3 h prior to incubation
##  [98] "Then the AgNPs were incubated with the blood for 3 h at 37 C while continuously rotating the
##  [99] "After centrifuging at 800 g to remove the intact red blood cells, the supernatant was centrif
## [100] "The absorbance of the supernatant of each AgNP without blood was also measured after centrifug
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Abbasalipourkabir R et al 2015.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "Materials"
word_end_mm <- "less"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 1061 1068
```

```
## [1] "2. Materials and methods 2. 1."
## [2] "Materials Zinc oxide nanopowder was purchased from Iranian Nanomaterials Pioneers"
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1] 2015 2339
```

```
## [1] "The results are expressed as mean ± standard deviation and analyzed using SPSS statistical softw
## [2] "The significant difference was considered to be p < 0. 05 or less."
```

```
mm_section<-cut_conllu_df(1061, 2339, conllu_df)
```

```
##  [1] "2. Materials and methods 2. 1."
##  [2] "Materials Zinc oxide nanopowder was purchased from Iranian Nanomaterials Pioneers"
##  [3] "Company, NANOSANY (Mashhad, Iran)."
##  [4] "Its characteristics are presented in Table 1 and transmission electron micrograph (TEM) and cry
##  [5] "1 and 2, respectively."
##  [6] "2. 2. Preparation of nanoparticle suspension"
##  [7] "A stock suspension was prepared by suspending ZnO nanoparticles in bi-distilled water."
##  [8] "2. 3."
##  [9] "Experimental design and procedure"
## [10] "Thirty adult male Wistar rats aged 6 e 8 weeks, weighing 180e 200 gwere purchased from Pasteur
```

```
## [11] "The animals were housed two rats per plastic cage and allowed to acclimatize under standard co
## [12] "The rats were given free access to distilled water and commercialized food throughout the expe
## [13] "The animals were divided into five groups of six animals each."
## [14] "Groups one to four received a dose of 50, 100, 150 and 200 mg zinc oxide nanoparticles (nZnO)/
## [15] "The rats were injected intraperitoneally daily for ten days."
## [16] "The control group was injected with bi-distilled water and the effect of shock injection is th
## [17] "At the end of the study, the rats were scarified and blood samples were collected from healthy
## [18] "The blood samples were allowed to clot at room temperature and centrifuged at 1000 g for 10 mi
## [19] "The rats were also studied for histological examination of liver and renal tissue and sperm an
## [20] "The experimental procedurewas approved at the Faculty of Medicine at Hamadan University of Med
## [21] "Certificate of analysis Content ZnO% Cu 99% 3 ppm Particles size nm Specific surface area m 2/
## [22] "2. 4. Measurement of superoxide dismutase activity Superoxide dismutase (SOD) catalyzes the br
## [23] "The method was carried out following the procedure of Marklund andMarklund (1974)."
## [24] "In this method, one unit of SOD activity is defined as the amount of enzyme required to inhibi
## [25] "Autoxidation of pyrogallol was quantified at 420 nm for 3 min."
## [26] "Therefore, the rate of decreased optical density between the 1 st and 3 rd min was expressed a
## [27] "2. 5."
## [28] "Measurement of glutathione peroxidase activity Glutathione peroxidase (GPX) activity was deter
## [29] "According to the study of Paglia and Valentine (1967), GPX converts H 2 O 2 to H 2 O and catal
## [30] "The optical density of the final mixture was read at 340 nm in UV/Vis Spectrophotometer (Unico
## [31] "Decrease of OD was expressed as enzyme activity, U/L."
## [32] "2. 6."
## [33] "Measurement of MDA, TAC and TOS Total antioxidant capacity (TAC) in serum samples was assessed
## [34] "Maolnedialdehyde (MDA) as a lipid peroxidation index was determined using fluorometric thiobar
## [35] "The oxidation of ferrous ion to ferric ion accompanied with a number of oxidant species in aci
## [36] "The ferric ion was determined by xylenol orange (Erel, 2005).Mn Cd"
## [37] "Pb 5 ppm 9 ppm 9 ppm Crystal phase Crystal Morphology"
## [38] "True density g/cm"
## [39] "3 Single Nearly spherical 5. 606"
## [40] "Fig. 2."
## [41] "X-ray diffraction (XRD) of zinc oxide nanoparticles from Iranian Nanomaterials Pioneers"
## [42] "Company, NANOSANY."
## [43] "R."
## [44] "Abbasalipourkabir et al. / Food and Chemical Toxicology 84 (2015) 154 e 1601562. 7. Analyzing
## [45] "Sperm analysis"
## [46] "At the end of the study after anesthetizing the animals and removing the epididymis, sperm ana
## [47] "The sperm count was assessed using a haemocytometer."
## [48] "Also, the qualitative and quantitative morphology and motility of sperms were observed with a
## [49] "Then, the percentage of live sperms to the total sperms was evaluated with a magnification of
## [50] "The dead sperms were observed in reddish color."
## [51] "The results are expressed as mean ± standard deviation and analyzed using SPSS statistical sof
## [52] "Histopathological study"
## [53] "At the end of the study, the rats were sacrificed and examined for tissue abnormalities."
## [54] "Samples of liver and kidney from all groups were immediately fixed in 10% formalin overnight,
## [55] "54 29. 49 ± 11."
## [56] "85 40. 00 ± 7 GPX (U/L)"
## [57] "1. 76 ± 0. 91 1. 08 ± 0. 49 2. 21 ± 0 TAC (mmol/ml) 1. 24 ± 0. 36 1. 06 ± 0. 20 1. 06 ± 0 MDA
## [58] "*p < 0. 05;"
## [59] "DP < 0. 01; #p < 0. 001."
## [60] "S capacity; MDA ¼ Malondialdehyde; TOS ¼ Total oxidant status. a Comparing with Control. b Com
## [61] "The tissue sections were viewed under a light microscope (Nikon Y-S 100, German)."
## [62] "2. 10."
## [63] "Statistical analysis"
## [64] "The data were expressed as mean ± standard deviation."
```

```
## [65] "For statistical analysis, the experimental values were compared to their corresponding control
## [66] "One-way analysis of variance (ANOVA) in SPSS software (Version 16. 0) was used to illustrate t
## [67] "The significant difference was considered to be p < 0. 05 or less."
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Abe, S et al 2012.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "EXPERIMENTAL"
word_end_mm <- "Ref"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 1127
```

```
## [1] "2. EXPERIMENTAL DETAILS 2. 1. Materials and Syntheses"
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1] 1701
```

```
## [1] "Details of the procedure were described in Ref. [ 26 ]. 3. RESULTS AND DISCUSSION 3. 1."
```

```
mm_section<-cut_conllu_df(1127, 1701, conllu_df)
```

```
##  [1] "2. EXPERIMENTAL DETAILS 2. 1. Materials and Syntheses"
##  [2] "In this study, we prepared two types of water-dispersible carbon nanotubes."
##  [3] "The MWCNTs used in this study were obtained from NanoLab (Brighton, MA, USA)."
##  [4] "They were heated at 500 C and washed with HCl for purification."
##  [5] "To prepare a water-dispersible carbon nanotube, a polycarboxylation reaction of MWCNTs was und
##  [6] "The MWCNTs were dispersed in o-dichlorobenzene with sonication."
##  [7] "A large excess of succinic acid peroxide was added to the solution, and the reaction mixtures
##  [8] "To obtain other water-dispersible CNTs, the purified CNTs were dispersed in nitric acid and re
##  [9] "After each modification reaction, they were filtered and then dried in vacuo."
## [10] "The details are described in Refs. [ 23 f and 26 ]."
## [11] "To identify the modification reaction processing, we measured the Raman and infrared spectra o
## [12] "The morphologies of the obtained CNTs were observed by scanning electron microscopy (SEM: Hita
## [13] "Nanoscope IIIa)."
## [14] "2. 2."
## [15] "Biodistribution Male mice (Jcl: ICR), 8-12 weeks old, were obtained from Nippon CLEA Japan, In
## [16] "The mice were randomly divided into groups (five mice per group)."
## [17] "We injected 0. 5 mL of the modified CNTs in aqueous dispersion (0. 5 mg/mL) into the tail vein
## [18] "After the waterdispersible CNTs were administered, the body weight of the mice was monitored f
## [19] "CNT-administered mice were sacrificed at 1 day, 1 week, and 4 weeks post-injection."
## [20] "The lung, liver, spleen and kidney were excised and subjected to transmission electron microsc
## [21] "All operations on animals were in accord with the institutional animal use and care regulation
## [22] "Cytocompatibility"
## [23] "To estimate the cytotoxicity of water-dispersible CNTs, we exposed these particles to human he
## [24] "Hc cells were seeded on 12-well plates (Hydro Cell, Wako, Osaka, Japan) at a density of approx
## [25] "After 1 day incubation, the culture medium was aspirated and another fresh medium including th
## [26] "The concentrations of the particles were adjusted to 0, 0. 01, 0. 1, 1 and 10 ppm, respectivel
## [27] "We also used THP-1 as a representative of monocytes to compare with cell functions of Hc."
## [28] "During incubation, TNF- in the supernatant of cell culture medium at 1 day after the incubatio
## [29] "Details of the procedure were described in Ref. [ 26 ]. 3. RESULTS AND DISCUSSION 3. 1."
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Adedara, I A et al 2018.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "Materials"
word_end_mm <- "significant"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 1055
```

```
## [1] "2. Materials and methods 2."
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1]  342 2583 2649 2825 3127 4238 4304
```

```
## [1] "Results indicated that exposure to MWCNTs caused liver damage evidenced by significant elevatio
## [2] "Values of P< 0. 05 were considered significant."
## [3] "Administration of carboxylated MWCNTs caused a dose-dependent significant decrease in the SOD a
## [4] "Further, there was a dose-dependent significant elevation in the hepatic level of MDA, a biomar
## [5] "Moreover, there was a significant, dose-dependent increase in serum levels of pro-inflammatory
## [6] "The present investigation demonstrated that short-term exposure to MWCNTs treatment elicited he
## [7] "Further, the present study demonstrated that administration of purified MWNCNTs to rats caused
```

```
mm_section<-cut_conllu_df(1055, 2583, conllu_df)
```

```
##  [1] "2. Materials and methods 2."
##  [2] "1."
##  [3] "Chemicals Carboxylated multi-walled carbon nanotubes (MWCNTs) from NanoLab Inc. (Newton, MA) w
##  [4] "Research Laboratory, NIH-RCMI Center for Environmental Health, CSET, Jackson State University,
##  [5] "1-chloro-2, 4-dinitrobenzene (CDNB), 50, 50-dithiobis-2-nitrobenzoic acid (DTNB), epinephrine,
##  [6] "All other chemicals and reagents were of analytical grade and were procured from Sigma Chemica
##  [7] "2."
##  [8] "2. Physicochemical characterization of functionalized multiwalled carbon nanotubes Multi-walle
##  [9] "The MWCNTs were subsequently heated under argon (2 l/min) to 2000 C at the rate of 10 C/min to
## [10] "Purified MWCNTs [purity> 95% by Thermogravimetry analysis (TGA)] were then subjected to a refl
## [11] "The resulting functionalized carboxylated nanotubes have 2 e 7% COOH by weight."
## [12] "The morphology and size of these pure MWCNTs were read by Transmission Electron Microscope (TE
## [13] "Prior to visualizing the samples with TEM, MWCNTs were directly placed on a TEM grid and allow
## [14] "Further, the surface areas were assessed by the isothermal gas adsorption BET method [ 13 ] us
## [15] "To characterize our system, MWCNTs were dispersed in 1% tween-80 and sterile saline using phys
## [16] "Eventually, the purified long MWCNTs had a diameter of 11."
## [17] "5 nm, length of 12 mm and specific surface area of 42 m 2/g, respectively."
## [18] "2. 3. Animal model and care Fifty pubertal male Wistar rats (8 weeks old) weighing between 160
## [19] "The rats were housed in plastic cages placed in a well-ventilated vivarium, under standard lab
## [20] "Animal care and experimental etiquettes were executed according to the approved guidelines set
## [21] "Use of Laboratory Animals' prepared by the National Academy of Science (NAS) and published by
## [22] "2. 4. Experimental design Following one week of acclimatization, the rats were assigned into f
## [23] "Pure carboxylated MWCNTs suspension was administered intraperitoneally to the rats at the dose
## [24] "0mg/kg for 5 consecutive days."
## [25] "Control rats were administered saline plus 1% tween-80 in the same manner as in the treatment
## [26] "The doses of carboxylated MWCNTs used in the present study were chosen based on the previously
## [27] "Twenty-four hours following the final treatment, the blood was collected from retroorbital ven
## [28] "Subsequently, serum samples were obtained by centrifugation of the clotted blood at 3000 g for
## [29] "2. 5. Assessment of oxidative stress biomarkers"
## [30] "Liver samples from control and MWCNTs-treated rats were homogenized in 50mM TriseHCl buffer (p
```

```
## [31] "The homogenate was subsequently centrifuged at 12, 000 g for 15 min at 4 C and the supernatant
## [32] "Hepatic protein concentration was determined according to the method described by Bradford [ 14
## [33] "Superoxide dismutase (SOD) activity was assayed according to the method described by Misra and
## [34] "Level of reduced glutathione (GSH) was assayed at 412 nm according to the method described by .
## [35] "[ 17 ]."
## [36] "Glutathione peroxidase (GPx) activity was assayed according to the method described by Rotruck
## [37] "[ 18 ]."
## [38] "Glutathione-S-transferase (GST) activity was assayed according on the method described by Habig
## [39] "[ 19 ]."
## [40] "Hydrogen peroxide level was assayed according to the method described byWolff [ 20]."
## [41] "Malondialdehyde (MDA) level, an index of lipid peroxidation (LPO), was assayed according to the
## [42] "[ 21 ] with slight modification."
## [43] "All assays with the exception of SOD and CAT activities were analyzed using a SpectraMax plate
## [44] "2. 6."
## [45] "Measurement of liver function and pro-inflammatory biomarkers Serum activities of alanine amino
## [46] "A. Adedara et al. / Biochemical and Biophysical Research Communications 503 (2018) 3167 e 3173
## [47] "The serum concentrations of pro-inflammatory cytokines interleukin-1 b (IL-1 b), interleukin-6
## [48] "2. 7. In vivo micronucleus (genotoxicity) assay Genotoxicity assay was performed according to
## [49] "Briefly, a small opening to the marrow was created at the proximal ends of the femurs with a pa
## [50] "Subsequently, the femur was immersed in fetal calf serum and the bone marrow carefully released
## [51] "The bonemarrow suspensionwas then smeared on the slide, fixed in methanol for 3 e 5 min and all
## [52] "The slides were thereafter stained with May-Gruenwald followed by 5% diluted"
## [53] "Giemsa solution for 30min."
## [54] "Further, the slides were rinsed in phosphate buffer followed by distilled water and air dried."
## [55] "Subsequently, the dried stained slides were mounted in DPX with coverslips and examined under
## [56] "The incidence of micronucleated polychromatic erythrocytes (MNPCE) was recorded using a tally
## [57] "The level of hepatic COX-2 and iNOS expression was assessed in a 5 mm thick sections of formali
## [58] "Following deparaffinization of the sections in xylene and rehydration with graded alcohol, anti
## [59] "The liver sections were subsequently blocked in goat serum followed by an overnight incubation
## [60] "The slides were subsequently washed with tris buffer saline and then incubated with horse-radis
## [61] "Immune complexes were visualized using 0. 05% 3, 3-diaminobenzidene and the slides examined und
## [62] "The images were captured using a digital camera (Leica Biosystems, UK) attached to the microsco
## [63] "The quantitative assessment of COX-2 and iNOS protein expression was performed by counting 10 r
## [64] "The levels of expression of the proteins were expressed as percentage of the total cells counte
## [65] "2. 9."
## [66] "Statistical analyses"
## [67] "Statistical analyses were done using one-way analysis of variance (ANOVA) to compare the experi
## [68] "Values of P< 0. 05 were considered significant."
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Aijie, C et al 2017.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "Material"
word_end_mm <- "significant"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 1280
```

```
## [1] "Material & methods Characterization of ZnO & TiO 2 NPs Powderform ZnO and TiO 2 NPs were obtaine
```

```
display_occurences(word_end_mm, conllu_df)
```

```
##  [1]  3030  4167  7894  8030  8076  8151  8865  9651 10257 10267
```

```
##  [1] "A pvalue of <0. 05 was considered statistically significant."
##  [2] "The physicochemical properties of the ZnO and TiO 2 NPs are summarized in Table 3. Biodistribu
##  [3] "Compared with the control group, the ZnO NP and TiO 2 NP groups exhibited no significant diffe
##  [4] "Moreover, compared with the control group, the rats exposed to ZnO NPs on training days 1, 2, 3
##  [5] "During the first day of reacquisition training (day 7), rats in the ZnO NPs and TiO 2 NPs group
##  [6] "Furthermore, rats in the TiO 2 NPs group travelled less distance in the target quadrant than th
##  [7] "However, in our study, there was no significant Zn or Ti accumulation in blood and other tissue
##  [8] "There was no significant difference between ZnO NPs and TiO 2 NPs in immunohistochemical analys
##  [9] "2 NPs showed significant defi cits in behavioral performance."
## [10] "There is no significant difference between ZnO NP and TiO 2 NP in neuro toxicity by tongue inst
```

```
mm_section<-cut_conllu_df(1280, 3030, conllu_df)
```

```
##  [1] "Material & methods Characterization of ZnO & TiO 2 NPs Powderform ZnO and TiO 2 NPs were obtain
##  [2] "The primary particle sizes and morphology were measured using transmission electron microscopy
##  [3] "The com positions of the NPs were measured by energy disper sive spectroscopy (SwiftED 3000, To
##  [4] "The hydrodynamic size and zeta potentials of the NPs in dis tilled water (DW) were determined t
##  [5] "Additionally, the spe cific surface areas of the NPs were measured through N 2 adsorption-
desorption Brunauer-Emmett-Teller adsorption analysis on a Micromeritics ASAP 2010M+ C instrument (Micro
##  [6] "Levels of endotoxin contamination were measured using an EToxate Kit (SigmaAldrich, MO, USA)."
##  [7] "Animals & experimental design Healthy Wistar male rats, 4-6 weeks old and 130-
150 g in body weight (BW), were purchased from the Animal Center of Southern Medical University (Guangzh
##  [8] "The animals were placed in a sterile room (23 ± 1°C room temperature, 60 ± 10% rel ative humidi
##  [9] "All procedures used in this experiment were approved by the National Animal Ethics Committee of
## [10] "The approval number provided by the Southern Medical University ethical committee was 2015021."
## [11] "The rats were randomly allocated into three groups with comparable weights: the control, ZnO NP
## [12] "ZnO NPs and TiO"
## [13] "2 NPs were dis persed in DW, sonicated for 30 min and vortexed for 3 min before the tongue inst
## [14] "In the studies of ZnO NPs and TiO 2 NPs toxicity, oral administration is one of the most common
## [15] "We referred to the doses of oral administration in other animal models [ 36-
38 ], and finally choose a relatively low dose (50 mg/kg)."
## [16] "Rats were weighed and anesthetized by an intraperitoneal injection of 1% pentobarbital sodium
## [17] "The rats were held in lateral position, and the tongue was pulled out from the corner of the mo
## [18] "Then, a microsyringe with a suspension of 50 mg/ml ZnO NPs and TiO 2 NPs (50 mg/kg BW per mouse
## [19] "The instillation procedure lasted about an hour."
## [20] "After instillation, the tongue was rinsed with DW repeatedly until no suspension residue remain
## [21] "The control group was instilled with an equal amount of DW."
## [22] "This anesthetization and instillation procedure was performed every other day for 30 days."
## [23] "The animals were sacrificed at 24 h (30 days) after the last exposure, and the tissues, includi
## [24] "Nanoparticles Taste bud Chorda tympani Facial nerve Nucleus of solitary tract Cerebral cortex C
## [25] "Rats were evaluated in the Morris water maze (MWM) test after exposure (n = 6)."
## [26] "A total of 90 rats were used in the study."
## [27] "Determination of Zn content & Ti content"
## [28] "The frozen brain, nerves (CT and glossopharyngeal nerve), heart, liver, spleen, kidneys, lung a
## [29] "We used calibration standards of 0. 05, 0. 1, 0. 2, 0. 4, 0. 6 and 0. 8 mg/l Zn and Ti to valid
## [30] "Transmission electron microscopy Small blocks ( 1 mm 3) of brain tissue (n = 6) were fixed for
## [31] "Then, the samples were dehydrated through an ethanol gradient and embedded in Epon resin."
## [32] "Finally, ultrathin tissue sections (500 nm) were mounted on grids."
## [33] "The brain tissue of the cerebral cor tex is taken from the gustatory cortex."
## [34] "Ultrastructural examination and photography were performed with an H7500 transmission electron
## [35] "Oxidative stress in the CNS"
## [36] "Six brain tissues of each group were weighed, chopped and put into a centrifuge tube."
## [37] "The homogenate (1:9 w/v) was centrifuged at 2500 r/min for 10 min at 4°C."
## [38] "After centrifugation, the supernatant was diluted in phosphatebuffered saline according to rec
## [39] "Oxidative stress markers (SOD, MDA, GSH, GSHPx and GSH/GSSG) were identified using a commercial
```

```
## [40] "The protein content was measured using Coomassie Blue staining."
## [41] "Gene expression by quantitative real-time PCR"
## [42] "The brain samples were removed and maintained at 80°C after the rats were sacrificed."
## [43] "The frozen tissues were homogenized and extracted using TRIzol reagent (Gibco, MA, USA) in acc
## [44] "Synthesized complemen tary DNA was prepared using the PrimeScript  RT reagent kit (TaKaRa, Shig
## [45] "qRTPCR was conducted using a commercial kit (SYBR Premix Ex Taq II, TaKaRa, Shiga, Japan) and a
## [46] "The primer sequences for oxidative stressrelated genes (7dehydrocholesterol reductase [Dhcr 7 ]
## [47] "After exposure for 30 days, brain tissues (n = 6) were excised, washed with 0. 9% cold saline a
## [48] "Briefly, the samples were dehydrated in ethanol and xylene solutions, embedded 2456 Nanomedicin
## [49] "(2017) 12(20) future science group Research Article Aijie, Huimin, Jia et al."
## [50] "in paraffin blocks and sliced at a thickness of 4  m using a microtome."
## [51] "After hematoxylineosin stain ing, pathological changes were observed with a digital camera into
## [52] "Immunohistochemistry method Immunohistochemistry was performed to identify the specific neurote
## [53] "The presence of prolif erating cells and apoptotic cells in brain sections was analyzed using I
## [54] "The oxidative damage to nucleic acids of brain cells was analyzed by 8hydroxy2deoxyguano sine
## [55] "The activation of astrocytes was detected by glial fibrillary acidic protein (GFAP; CST, MA, US
## [56] "Photographs of positive stain ing (brown) were captured at randomly selected fields in both the
## [57] "To evaluate the impact of ZnO NPs and TiO 2 NPs on spatial learning and memory, male rats (n =
## [58] "The maze, filled with water (2. 0 m diameter, 0. 8 m high), was divided into four quadrants wit
## [59] "Visible cues were placed around the water maze."
## [60] "During the 5day training sessions, rats underwent four trials in different quadrants."
## [61] "They were required to find the escape platform in 90 s."
## [62] "On the sixth day (probe trial), rats were placed at the opposite side to swim for 60 s."
## [63] "On the seventh and eighth days, reverse escape platform training (reacquisition trials) was per
## [64] "Through out the training and testing, parameters in the MWM such as the latency to reach the pl
## [65] "Statistical analysis"
## [66] "All results are expressed as mean ± SEM."
## [67] "An analysis of variance with repeated measures was applied in the MWM test, and unpaired Studen
## [68] "All ana lyzes were performed in SPSS 19. 0 software."
## [69] "A pvalue of <0. 05 was considered statistically significant."
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Amrollahi-Sharifabadi et al. 2018.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "Materials"
word_end_mm <- "USA"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 1790 8160
```

```
## [1] "Materials and methods gO and characterization GO (CAS"
## [2] "Graphene-Based Materials for Biosensors: A Review."
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1] 2051 2492 3112
```

```
## [1] "The Brunauer-Emmett-Teller (BET) surface area of GO was acquired through a NOVA 2000e surface ar
## [2] "The first one was put into BD Vacutainer® blood collection tubes containing K 3 -EDTA (Beckton &
## [3] "The data analysis was performed by GraphPad Prism Version 5. 04 (GraphPad Software, Inc., La Jol
```

```
mm_section<-cut_conllu_df(1790, 3112, conllu_df)
```

```
##  [1] "Materials and methods gO and characterization GO (CAS"
```

```
##  [2] "No of 7732-18-5; carbon purity of 99%) was obtained from the US Research Nanomaterials, Inc. (
##  [3] "Characterization of GO was done using X-ray diffraction (XRD) analysis, Raman spectroscopy, fi
##  [4] "XRD measurement was performed on a Philips X'Pert MPD Diffractometer (UK)"
##  [5] "International Journal of Nanomedicine 2018:13 submit your manuscript | www."
##  [6] "dovepress."
##  [7] "com Dovepress"
##  [8] "Dovepress 4759"
##  [9] "In vivo toxicity of graphene oxide nanoplatelets using Cu K  radiation ( =1."
## [10] "5418 A°) at the voltage of 40 kV and the current of 15 mA by scanning in a 2  range from 0° to
## [11] "The Raman spectroscopy spectrum was recorded by an Almega Thermo Nicolet Dispersive Raman Spec
3, 000 cm-1."
## [12] "The surface morphology of GO was inspected utilizing FESEM (Sigma, Carl Zeiss, Germany) equipp
Thornley detector (SE 2) at the accelerating voltage of 15 kV and working distance of 8. 7 mm."
## [13] "TEM (Hitachi S 4160, Tokyo, Japan) was used at the accelerating voltage of 30 kV to study the
## [14] "The Brunauer-Emmett-Teller (BET) surface area of GO was acquired through a NOVA 2000e surface
## [15] "In vivo animal study Twenty adult male Wistar rats (8-12 weeks age, body weight of 150-
200 g) were obtained from Pasture Institute (Tehran, Iran)."
## [16] "The animals were housed in polyethylene cages and maintained under the controlled temperature
## [17] "They were allowed ad libitum access to the standard pellet and tap water."
## [18] "Animals received humane care, and all the experiments were performed in accordance with the Gu
## [19] "6. 17)."
## [20] "Twenty rats were randomly divided into four groups of five animals per group (Table 1)."
## [21] "Group I was assigned as a vehicle control and received normal saline 0. 9% by intraperitoneal
## [22] "Groups II-IV were received four doses of GO at 50, 150, or 500 mg/kg every 48 hours during 1 w
## [23] "The selected dose levels of GO were based on our range-finding study."
## [24] "Moreover, our dose selection was in agreement with a recently published study that conducted a
## [25] "30"
## [26] "Two weeks after the last dosing, the animals euthanized and blood and organ samples harvested
## [27] "The weights of animals were determined before dosing animals with"
## [28] "GO and at the final of the experimental period."
## [29] "Preparation of gO"
## [30] "For each animal, appropriate milligrams of GO were weighed and transferred into the sterile tu
## [31] "Then, the suspensions of nanoparticles were prepared by vortexing the tubes for 10 minutes."
## [32] "GO was homogenously dispersed in normal saline 0. 9% solution without producing any agglomerat
## [33] "All doses of homogenous"
## [34] "GO suspensions were prepared freshly and warmed at 37°C each time before the injection."
## [35] "hematological and biochemical measurements"
## [36] "Rats were anesthetized with diethyl-ether, and blood samples were collected through the cardia
## [37] "The first one was put into BD Vacutainer® blood collection tubes containing K 3 -EDTA (Beckton
## [38] "Serum samples were prepared by centrifugation at 3, 000 g for 15 minutes."
## [39] "Hematological parameters in the whole blood included hemoglobin (Hgb), haematocrit (Hct), whit
## [40] "These parameters were analyzed by a hematology autoanalyzer Sysmex Kx-21 n (Sysmex Corporation
## [41] "The biochemical parameters in the serum samples included blood urea nitrogen (BUN), creatinine
## [42] "Group Substance Number of animals Number of i. p."
## [43] "injections"
## [44] "Interval between each injection Interval between the last injection and sampling I c (Ns) 5 4
## [45] "II gONs in Ns (50 mg/kg) 5 4 48 hours 2 weeks III gONs in Ns (150 mg/kg) 5 4 48 hours 2 weeks
## [46] "Ns, normal saline; i. p., intraperitoneal."
## [47] "Dovepress 4760 amrollahi-sharifabadi et al aminotransferase (ALT), alkaline phosphatase (ALP),
## [48] "They analyzed by a biochemical autoanalyzer BT 3000 (BioTechnica, Italy) with standard diagnos
## [49] "Serum levels of creatine kinase-MB (CK-MB) were measured on the basis of chemiluminescent immu
## [50] "Germany)."
## [51] "Serum electrolytes included sodium (Na+), potassium (K+), and calcium (Ca 2+) were determined
## [52] "histopathological examinations"
```

```
## [53] "The animals were killed by cervical dislocation after blood collection and immediately underwe
## [54] "The internal organs of animals, including the liver, kidney, spleen, lung, intestine, brain, a
## [55] "Organs were dehydrated in a graded series of ethanol and xylene and embedded in paraffin and s
## [56] "Then, the blocks were deparaffinized and stained with hematoxylin and eosin (H&E)."
## [57] "Microscopic evaluations were performed by using a bright-field microscope (Nikon Eclipse 80i, .
## [58] "Data analysis"
## [59] "The results are presented as means and SD."
## [60] "The one-way ANOVA test with post hoc Bonferroni analysis was used for comparing the groups."
## [61] "A P-value, 0. 05 was considered as statistically significant."
## [62] "The data analysis was performed by GraphPad Prism Version 5. 04 (GraphPad Software, Inc., La J
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Aragao-Santiago, L 2016.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "Methods"
word_end_mm <- "lower"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 1113
```

```
## [1] "Methods Nanoparticle preparation PLGA NPs were prepared by the emulsion/evaporation method, as p
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1] 3548 6876
```

```
## [1] "Data analysis was considered significant at p 50. 05 or lower."
## [2] "Despite the non-biodegradability of PS NPs, the lower inflammatory response in comparison to ana
```

```
mm_section<-cut_conllu_df(1113, 3548, conllu_df)
```

```
##  [1] "Methods Nanoparticle preparation PLGA NPs were prepared by the emulsion/evaporation method, as
##  [2] "Hervé Hillaireau, Faculté de Pharmacie, Institut Galien Paris-Sud, Université de Paris-Sud 11,
##  [3] "Tel: +33 1 46 83 54 27."
##  [4] "Fax: +33 1 46 61 93 34."
##  [5] "E-mail: herve."
##  [6] "hillaireau@u-psud."
##  [7] "fr D ow nl oa de d by [ C ol or ad o St at e U ni ve rs ity ] at 0 0: 32 2 1 N ov em be r 20 15
##  [8] "Neutral (PLGA/PVA) and positive (PLGA/CS) NPs were prepared by dissolving 100 mg of PLGA (75:25
89% hydrolyzed, 30-70 kDa, Sigma Aldrich, Lyon, France) for PLGA/PVA or with 0. 6 mg mL 1 chitosan (Prot
90% deacetylation, 50- 150 kDa, NovaMatrix , FMC BioPolymer, Drammen, Norway) and 5 mg mL 1 PVA (Mowiol
##  [9] "The mixture was kept on ice and sonicated (Branson Digital Sonifier) at 80 W for 1 or 2 min, re
## [10] "For the preparation of negatively charged (PLGA/PF 68) NPs, 50 mg of PLGA were dissolved in 5 m
## [11] "This mixture was then vortexed for 1 min and kept on ice during a 2 min homogenization at 10 00
## [12] "KG, Staufen, Germany) equipped with an 18 G stainless steel dispersing tool."
## [13] "For all preparations, the organic phase was evaporated at room temperature overnight under magn
## [14] "The purification of PLGA/PVA and PLGA/CS (removal of stabilizers) NPs was done by ultracentrifu
## [15] "PLGA/PF 68 NPs were purified by dialysis during 24 h against MilliQ water using a cellulose est
## [16] "For in vivo imaging studies, PLGA NPs covalently labeled with DY-700, a nearinfrared dye, were
## [17] "The same protocol was followed with the formulation reduced to one quarter (25 mg of PLGA in te
## [18] "TiO 2 suspensions were prepared by dispersion of anatase (spheres, 525 nm) or rutile (rods, 10
## [19] "TiO 2 nanopowders (Sigma-Aldrich, Lyon, France) in MilliQ water."
## [20] "Briefly, a 1 mg mL 1 suspension of TiO 2 nanopowder was vortexed and had its pH adjusted to 11
## [21] "KG, Staufen, Germany)."
```

```
## [22] "Polystyrene NPs (PS) are commercially available as an aqueous suspension of 0. 20 lm NPs (Poly
## [23] "All NPs were tested as endotoxin-free at the concentrations they were given to animals."
## [24] "Nanoparticle characterization Nanoparticle average diameter size and polydispersity index were
## [25] "The zeta potential was determined using the same instrument after dilution in 1 mM NaCl."
## [26] "All measurements were done in triplicate."
## [27] "The morphology of NPs in aqueous suspension was studied by transmission electron microscopy (T
## [28] "A 5 mL of NP aqueous suspensions was deposited onto copper grids covered with formvar film (40
## [29] "PLGA samples were then stained using 2% phosphotungstic acid (PLGA/PVA, PLGA/PF 68) or 2% urany
## [30] "TiO 2 samples were used without staining."
## [31] "The excess liquid was blotted off using filter paper, and the grids were dried before observati
## [32] "Images were acquired using a post-column high-resolution (11 megapixels) high-speed camera (SC
## [33] "Animals Toxicity of NPs was investigated in Balb/cJ mice (4 weeks-old, n¼ 4-
6/group) purchased from Janvier laboratories (France)."
## [34] "Food and water were administered ad libitum."
## [35] "Mice were housed in plastic cages under controlled environmental conditions (temperature 19-
21 C, humidity 40-70%, lights on 07:00 h to 19:00 h)."
## [36] "Mice were acclimated to these conditions for 7 days before experiments started."
## [37] "The rules from the local ethics committee (Comité d'éthique pour l'expérimentation animale, pr
## [38] "Biodistribution of NPs was followed in NMRI nude mice (4 weeks old) purchased from Janvier labo
## [39] "Mice were submitted to the above cited conditions in addition to being housed under germ-free
## [40] "Pulmonary delivery of aerosolized NPs Nebulization was performed as previously described (Bivas
## [41] "Balb/c mice were exposed to 3% isoflurane (CSP, Cournon-d'Auvergne, France) and, once a deep s
## [42] "TiO 2 or rutile"
## [43] "TiO 2 NPs in 5% glucose (final pH 6-7)."
## [44] "For the nebulization, mice were suspended by the upper teeth at a 45 angle, in supine position
## [45] "The mouth was opened, the tongue was displaced with the help of a forceps and a mouse intubati
## [46] "Once a clear view of the trachea was obtained, with the visualization of the vocal chords, the
## [47] "The tip was withdrawn and the mouse was taken off the support and allowed to recover under visu
## [48] "One set of mice (n¼ 4-6/group) was euthanized 24 or 48 h following a single nebulization and t
## [49] "Another set of mice (n¼ 4-6/group) was used for lung excision and fixation for histological ana
## [50] "Before conducting the experiments, the efficacy of the nebulization procedure was evaluated by
## [51] "Animals were sacrificed immediately after nebulization and lung distribution was visualized in
## [52] "Bronchoalveolar lavage BAL collection was adapted from previously described techniques (Maxeino
## [53] "Segel et al., 2005)."
## [54] "Briefly, the trachea was surgically exposed and a small incision was made between tracheal ring
## [55] "Lungs were instilled successively with isotonic ice-cold saline that 2 L. Aragao-Santiago et al
## [56] "Nanotoxicology, Early Online : 1-11 D ow nl oa de d by [ C ol or ad o St at e U ni ve rs ity ]
## [57] "The first two lavages (300 and 400 lL) were collected in a 1. 5 mL tube and the subsequent (500
## [58] "After centrifugation at 400 g for 15 min at 4 C, the supernatant from the first two lavages was
## [59] "BALF differential cell count and imaging Differential cell count was performed after a cell ali
## [60] "Cytoslides were fixed and thereafter stained with May- Grünwald-Giemsa prior to manual cell cou
## [61] "Macrophages, neutrophils, eosinophils, basophils and lymphocytes were identified by their chara
## [62] "Light microscopic images of stained cells were taken with Leitz Diaplan microscope (Leica Micro
## [63] "Total protein quantification Total protein content in BALF was measured spectrophotometrically
## [64] "Manufacturer's protocol for Bio-Rad Protein Assay (Bio-Rad Laboratories GmbH, Munich, Germany)
## [65] "Cytokine measurements Cell-free BALF was used for cytokine analysis using BD Cytometric Bead A
C motif) ligand 2 (CCL 2), were measured."
## [66] "Samples were acquired on BD Accuri C 6 flow cytometer (BD Biosciences, Le Pont de Claix, Franc
## [67] "The predicted assay detection limit is 5 pg mL 1 for IL-6, 17. 5 mg mL 1 for IL-10, 52. 7 mg m
## [68] "Histopathological evaluation"
## [69] "At 24 h after nebulization, lung tissue was dissected, rinsed from blood with PBS and fixed in
## [70] "Paraffin-embedded lungs were 5 lm sectioned and stained with hematoxylin-eosin (H-
E)."
## [71] "Slices were evaluated by an anatomopathologist using light microscopy."
```

```
## [72] "In vivo distribution of PLGA NPs NMRI nude mice (6-8 weeks old, Janvier, France; n¼ 2-
3/group) were anesthetized with a medetomidine (200 lg kg of BW 1, Dormitor , Orion Pharma, Espoo, Finla
## [73] "kg of BW 1, Imalgéne, Merial, France) mixture in phosphate buffered saline and intratracheally
## [74] "Mice were sacrificed by cardiac exanguination under anesthesia 1."
## [75] "5-2% mixture of isoflurane anesthesia (CSP, Cournon-d'Auvergne, France) at 2, 5, 8, 24 and 48 h
## [76] "Tissues were illuminated by 675 mm light-emitting diodes equipped with interference filters and
## [77] "Sciences Inc., Hopkinton, MA)."
## [78] "Fluorescence images were acquired during 10 s and image analysis was performed using the Wasabi
## [79] "Semi-quantitative data were obtained by drawing regions of interest (ROI) around each organ."
## [80] "Fluorescent images were acquired by a back-thinned CCD camera at 80 C (ORCAII-BT-512 G, Hamamat
## [81] "Statistical analysis Statistical differences were tested by one-way analysis of variance (ANOVA
## [82] "Two-way ANOVA was applied for tissue fluorescence data, followed by the Bonferroni post-hoc tes
## [83] "For cytokine measurements, the concentration in samples below the detection limit was set to 0
## [84] "Data analysis was considered significant at p 50. 05 or lower."
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "AshaRani et al. 2009.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "MATERIALS"
word_end_mm <- "Internet"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 7422
```

```
## [1] "MATERIALS AND METHODS"
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1] 10247
```

```
## [1] "This material is available free of charge via the Internet at http://pubs."
```

```
mm_section<-cut_conllu_df(7422, 10247, conllu_df)
```

```
##   [1] "MATERIALS AND METHODS"
##   [2] "The particle synthesis was carried out using the standard procedure through reduction of silve
##   [3] "All experiments were done in a clean atmosphere to eliminate the chances of endotoxin contamin
##   [4] "All chemicals used for nanoparticle synthesis were purchased from Sigma-Aldrich."
##   [5] "Preparation of Starch-Capped Ag-np."
##   [6] "Starch-coated silver nanoparticles were synthesized by a method reported by Raveendran et al."
##   [7] "19"
##   [8] "The choice of capping agent was done based on the stability of nanoparticles in cell culture m
##   [9] "Starch-capped nanoparticles showed lesser degree of agglomeration even at high concentrations
##  [10] "Furthermore, the choice of capping agent is important since the properties of nanoparticles ca
##  [11] "The distribution of nanoparticles in the body is strongly influenced by its surface characteri
##  [12] "The hydrophilic nature of starch as compared to organic polymers could enhance the water dispe
##  [13] "Moreover, using starch as the capping agent removed the need of other organic solvents, or cap
##  [14] "Additionally, our experiments showed that starch controls were not cytotoxic to the cells unde
##  [15] "Briefly, soluble starch from potatoes (0."
##  [16] "28 g) was dissolved in 10 mL of boiling ultrapure water and filtered using a 0. 2 m syringe fi
##  [17] "Silver nanoparticles were synthesized by reducing silver nitrate solution"
##  [18] "(1 mM), using sodium borohydride (0. 03 g) followed by the addition of the filtered starch sol
##  [19] "The color of the solution changed to dark brown with time, indicating nanoparticle formation,
##  [20] "The nanoparticle suspension was centrifuged at 18 000 rpm for 1 h to pellet nanoparticles."
```

```
##  [21] "The pellets were further washed in ultrapure water to remove traces of unbound starch."
##  [22] "The dry pellet obtained after the lyophilization of the centrifuged nanoparticles was dissolv
##  [23] "The size of the nanoparticles was determined by TEM (Figure 1 A) analysis and ultraviolet (UV
##  [24] "A size distribution histogram was extracted from Figure 1 A using Gatan digital micrograph so
##  [25] "Electron micrographs of Ag-np are included in the Figure S 1 in the Supporting Information."
##  [26] "Cell Culture and Nanoparticle Treatment."
##  [27] "Cell lines were purchased from commercial sources, IMR-90: Coriell Cell Repositories, USA; U 1
##  [28] "Human glioblastoma cells (U 251) were maintained in Dulbecco's modified eagles medium (DMEM, S
##  [29] "Louis, MO) supplemented with 10% fetal bovine serum (FBS, GIBCO, Invitrogen, Grand Island, NY
##  [30] "Normal human fibroblasts (IMR-90, at passage 20 3) were maintained in modified eagles medium w
##  [31] "Cells were maintained in a 5% CO 2 incubator at 37 °C."
##  [32] "Stock solutions of nanoparticles (5 mg/mL) were prepared in sterile distilled water and dilute
##  [33] "Appropriate concentrations of Ag-np stock solution were added to the cultures to obtain respec
##  [34] "CK 40) to detect morphological changes and photographed using an Olympus C 7070WZ camera."
##  [35] "Transmission Electron Microscopy (TEM) of Ag-np Treated Cells."
##  [36] "Ultrathin sections of the cells were analyzed using TEM to reveal the distribution of nanopart
##  [37] "Briefly, the cells (1. 5 106 cells) were treated with Ag-starch nanoparticles (net concentrati
##  [38] "At the end of the incubation period, culture flasks were washed many times with phosphate buff
##  [39] "Cells were trypsinized and washed 4 5 times in phosphate buffer and fixed in 2. 5% glutaraldeh
##  [40] "Post-fixation staining was done using 1% osmium tetroxide for 1 h at room temperature."
##  [41] "Cells were washed well and dehydrated in alcohol (40, 50, 70, 80, 90, 95, and 100% ethanol) an
##  [42] "acsnano."
##  [43] "org VOL."
##  [44] "3  NO."
##  [45] "2  279-290  2009 287 ity resin (1:1), for 18 h."
##  [46] "Cells were further treated with pure resin for 24 h and embedded in beem capsules containing p
##  [47] "Resin blocks were hardened at 70 °C for 2 days."
##  [48] "Ultrathin sections (70 nm) were cut using Reichert Jung Ultracut."
##  [49] "The sections were stained with 1% lead citrate and 0. 5% uranyl acetate and analyzed under JEO
##  [50] "The presence of nanoparticles was confirmed by the electron dispersive X-ray analysis (EDX, Fi
##  [51] "Scanning Transmission Electron Microscopy (STEM)."
##  [52] "Scanning transmission electron microscopy (STEM) and the elemental mapping of cell sections we
##  [53] "The sections prepared for the TEM analysis were employed for STEM study."
##  [54] "The instrument performed elemental mapping by labeling silver as red dots in the image capture
##  [55] "The analysis was done using JEOL JSM 6701"
##  [56] "F at an accelerating voltage of 25 kV."
##  [57] "Cell Viability Assay."
##  [58] "The viability of Ag-np treated cells was measured using Cell-Titer glow luminescent cell viabi
##  [59] "This assay is a homogeneous method for determining the number of viable, metabolically active
##  [60] "The procedure involves addition of an equal volume of reagent to the medium in test wells, whi
##  [61] "The reagent contains detergents to break the cell membrane, causing ATP release in the medium
##  [62] "The assay is based on the conversion of beetle luciferin to oxyluciferin by a recombinant luci
##  [63] "The observed luminescence is proportional to the quantity of ATP in cells."
##  [64] "The experiments were performed in white opaque walled 96-well plates (Corning, Costar, NY)."
##  [65] "Additional controls were included in the test to rule out autoluminescence and quenching by si
##  [66] "For the ATP assay, 1 104 cells per well were plated and treated with different concentrations
##  [67] "The dependence of toxicity on purity of nanoparticles was studied using the supernatants from
##  [68] "The supernatant (50 mL) obtained after removing the nanoparticle pellet was concentrated by ly
##  [69] "Different volumes of the stock solution (0, 5, 10, 15, 20, and 25 L) were dispensed in to 100
##  [70] "Cell titer blue cell viability assay (Promega, Madison, WI) is a fluorimetric measurement of t
##  [71] "The mitochondrial and microsomal enzymes reduce resazurin in the reagent to resorufin, which a
##  [72] "Cells were seeded at a density of 1 104 cells per well, in black opaque walled 96-well plates
##  [73] "A time-dependent study was conducted employing different incubation period (24, 48, and 72 h)
##  [74] "The experiments were carried out as per supplier's instructions."
```

```
##  [75] "Cell Cycle Analysis."
##  [76] "Cell cycle analysis was carried out by staining the DNA with propidium iodide (PI) followed by
##  [77] "Approximately 4 105"
##  [78] "U 251 cells and 8 105 IMR 90 cells were placed in 100 mm tissue culture dish (Falcon, Franklin
##  [79] "Following the Ag-np treatments for 48 h (concentrations employed were similar as in viability
##  [80] "Cells were washed in 1 X phosphate buffered saline (PBS, 1 st Base, Singapore) trypsinized, ha
##  [81] "The pellet was washed in PBS, fixed in ice-cold ethanol (70%), and stored at 20 °C."
##  [82] "Before flow cytometry analysis, cells were washed in PBS and stained with propidium iodide (PI
##  [83] "Flow cytometry analysis was performed using Epics Altra (Beckman and Coulter) at an excitation
##  [84] "Data collected for 2 104 cells was analyzed using WinMDI 2. 8 software."
##  [85] "53 Annexin-V Staining."
##  [86] "Annexin-V staining was performed to differentiate apoptosis from necrotic cell death induced b
##  [87] "Annexin-V has a high affinity for phosphotidyl serine, which is translocated from the inner to
##  [88] "Its conjugation with the fluorescent probe FITC facilitates measurement by flow cytometric ana
##  [89] "Use of propidium iodide (PI) staining helps distinguish between apoptosis and necrosis due to
##  [90] "Cell number, concentrations, and culture conditions were similar to cell cycle analysis."
##  [91] "Treated cells were harvested and washed twice in PBS."
##  [92] "The staining was carried out as per manufacturer's instruction (annexin-V FITC apoptosis detec
##  [93] "Louis, MO)."
##  [94] "Data analyses were done using WinMDI software."
##  [95] "Detection of Reactive Oxygen Species (ROS) Production."
##  [96] "The generation of hydrogen peroxide and superoxide radical was monitored by employing 2=,7=- d
##  [97] "DCF-DA is nonfluorescent unless oxidized by the intracellular ROS."
##  [98] "Dihydroethidium is blue fluorescent in the reduced form, which upon oxidation by superoxide ra
##  [99] "Dose- and time-dependent measurements of the generation of reactive oxygen species were done b
## [100] "Hydrogen peroxide treated cells (0. 09% H 2 O 2) were used as positive control for DCF-DA anal
## [101] "DDC is a strong inhibitor of superoxide dismutase activity in cells."
## [102] "Cells were then washed twice in serum-free medium and analyzed using Epics Altra flow cytomete
## [103] "The concentrations were chosen based on the viability data."
## [104] "For each sample, 1 104 cells were collected (Epics Altra, Beckman Coulter), and data were anal
## [105] "Cytokinesis-Blocked Micronucleus Assay (CBMN)."
## [106] "Cytokinesis-blocked micronucleus assay (CBMN) measures the chromosomal breakage that occurs du
## [107] "42 Cell density was similar to cell cycle analysis."
## [108] "The cells were treated with two different concentrations of Ag-np (100 and 200 g) for 48 h fol
## [109] "The analysis was performed according to a reported procedure."
## [110] "56"
## [111] "Cells were harvested and treated with ice cold KCl and centrifuged immediately."
## [112] "The pellet was fixed in Carnoy's fixative (3:1 methanol/acetic acid), and a few drops of forma
## [113] "The cells were aged for at least 4 days at 4 °C, streaked on clean glass slides, and dried."
## [114] "The slides were then stained with acridine orange (30 g/mL), which differentially stains the n
## [115] "57"
## [116] "One thousand binucleated cells were scored, and the number of micronuclei was recorded."
## [117] "The IMR 90 cells had approximately 700 binucleated cells."
## [118] "Alkaline Single-Cell Gel Electrophoresis (Comet Assay)."
## [119] "Alkaline singlecell gel electrophoresis (Comet assay) detects DNA damage through electrophores
## [120] "Treated cells were harvested and washed twice in PBS before resuspending in Hank's balance sal
## [121] "The cells were embedded in 0. 8% low melting agarose (Pronadisa, Spain) on comet slides (Trevi
## [122] "Cells were then subjected to denaturation in alkaline buffer (0. 3 M NaCl, 1 mM EDTA) for 40 m
## [123] "Electrophoresis was performed at 25 V and 300 mA for 20 min."
## [124] "The slides were immersed in neutralization buffer (0. 5 M Tris-HCl, pH 7. 5) for 15 min follow
## [125] "The slides were airdried and stained with SYBR green dye."
## [126] "The tail moments of the nuclei were measured as a function of DNA damage."
## [127] "Analysis was done using comet imager v 1. 2 software (Metasystems GmbH, Altlussheim, Germany)
## [128] "A RT IC LE VOL."
```

```
## [129] "2   ASHARANI ET AL. www."
## [130] "org 288 Statistical analyses of the values for all experiments are expressed as mean standard
## [131] "The data were analyzed using Student's t test (Microsoft Excel, Microscoft Corporation, USA) w
## [132] "Acknowledgment."
## [133] "This work was supported by the Office of Life Sciences (OLS) at the National University of Sin
## [134] "We acknowledge facilities support by the NUS-Nanoscience and Nanotechnology Initiative (NUSNNI
## [135] "The authors thank L. V. Bindhu and S. Shubhada for their help with the manuscript."
## [136] "Supporting Information Available:"
## [137] "Additional details of experiments and results are included."
## [138] "This material is available free of charge via the Internet at http://pubs."
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

```
pdf_name <- "Bendre, V et al 2011.pdf"
conllu_df<-open_pdf_and_annotate(pdf_name, model)

word_start_mm <- "MATERIALS"
word_end_mm <- "required"

display_occurences(word_start_mm, conllu_df)
```

```
## [1] 574
```

```
## [1] "2. MATERIALS AND METHODS"
```

```
display_occurences(word_end_mm, conllu_df)
```

```
## [1] 744 980
```

```
## [1] "The only information required to be input is the temperature of the liquid under J. Biomed."
## [2] "Once analysed, the sample is simply withdrawn from the unit for re-use, if required."
```

```
mm_section<-cut_conllu_df(574, 980, conllu_df)
```

```
##  [1] "2. MATERIALS AND METHODS"
##  [2] "In practice the NTA technique requires a 70 l sample of liquid containing particles at a concen
##  [3] "40 mW at = 635 nm) is passed."
##  [4] "Particles within the path of the specially configured beam are observed via a microscopebased s
##  [5] "4 The motion of the particles in the field of view (approx."
##  [6] "100× 100 m) is recorded (at 30 fps) and the subsequent video analysed."
##  [7] "Each and every particle visible in the image is individually but simultaneously tracked from fi
##  [8] "Results are displayed as an equivalent hydrodynamic diameter particle distribution, calculated
##  [9] "The only information required to be input is the temperature of the liquid under J. Biomed."
## [10] "Nanotechnol."
## [11] "2011, Vol. 7, No. 1 1550-7033/2011/7/195/002 doi:10."
## [12] "1166/jbn."
## [13] "2011."
## [14] "1265 195"
## [15] "Delivered by Ingenta to: Main CID is 80004805 (JPP)"
## [16] "IP: 178."
## [17] "57. 68."
## [18] "194 On: Mon, 20 Jun 2016 06:17:18 Copyright: American Scientific Publishers R E S E A R C H A I
## [19] "(a) (b) (c) (d) Fig. 1. A mixture of 200 nm and 300 nm particles; (a) still image, overlaid wi
## [20] "analysis and the viscosity (at that temperature) of the solvent in which the nanoparticles are
## [21] "Otherwise the technique is one of the few analytical techniques which is absolute and therefore
## [22] "Notably, because the instrument visualizes particles on an individual basis, particle number co
## [23] "Once analysed, the sample is simply withdrawn from the unit for re-use, if required."
```

```
saveRDS(mm_section, file=paste0("Validation_tool_folder/", pdf_name, ".rds"))
```

Benchmach and performance of the tool

```
all_sections <-list.files("Validation_tool_folder/", pattern = ".rds")
all_sections
```

```
##  [1] "Abbasalipourkabir R et al 2015.pdf.rds"
##  [2] "Abe, S et al 2012.pdf.rds"
##  [3] "Adedara, I A et al 2018.pdf.rds"
##  [4] "Aijie, C et al 2017.pdf.rds"
##  [5] "Amrollahi-Sharifabadi et al. 2018.pdf.rds"
##  [6] "Ansari, M 2018.pdf.rds"
##  [7] "Aragao-Santiago, L 2016.pdf.rds"
##  [8] "AshaRani et al. 2009.pdf.rds"
##  [9] "Bai, K J et al 2018.pdf.rds"
## [10] "Bendre, V et al 2011.pdf.rds"
## [11] "Blum, J L et al 2014.pdf.rds"
## [12] "Cabral, H et al 2005.pdf.rds"
## [13] "Chen et al 2008.pdf.rds"
## [14] "Jamshidzadeh, A et al 2015.pdf.rds"
## [15] "Li, C et al 2009.pdf.rds"
## [16] "Li, J et al 2013.pdf.rds"
## [17] "Li, J G et al 2009.pdf.rds"
## [18] "Li, Y et al 2010.pdf.rds"
## [19] "Liu, Y et al 2014.pdf.rds"
## [20] "Love, S A et al 2012.pdf.rds"
## [21] "Maurizi, L et al 2015.pdf.rds"
## [22] "Nešković, O et al 2013.pdf.rds"
## [23] "Neun, B W et al 2018.pdf.rds"
## [24] "Reddy, A et al 2010.pdf.rds"
## [25] "Richter, AW et al 1983.pdf.rds"
## [26] "Sadaf, A et al 2012.pdf.rds"
## [27] "segmenteR_Abbasalipourkabir R et al 2015.pdf.rds"
## [28] "segmenteR_Abe, S et al 2012.pdf.rds"
## [29] "segmenteR_Adedara, I A et al 2018.pdf.rds"
## [30] "segmenteR_Aijie, C et al 2017.pdf.rds"
## [31] "segmenteR_Amrollahi-Sharifabadi et al. 2018.pdf.rds"
## [32] "segmenteR_Ansari, M 2018.pdf.rds"
## [33] "segmenteR_Aragao-Santiago, L 2016.pdf.rds"
## [34] "segmenteR_AshaRani et al. 2009.pdf.rds"
## [35] "segmenteR_Bai, K J et al 2018.pdf.rds"
## [36] "segmenteR_Bendre, V et al 2011.pdf.rds"
## [37] "segmenteR_Blum, J L et al 2014.pdf.rds"
## [38] "segmenteR_Cabral, H et al 2005.pdf.rds"
## [39] "segmenteR_Chen et al 2008.pdf.rds"
## [40] "segmenteR_Jamshidzadeh, A et al 2015.pdf.rds"
## [41] "segmenteR_Li, C et al 2009.pdf.rds"
## [42] "segmenteR_Li, J et al 2013.pdf.rds"
## [43] "segmenteR_Li, J G et al 2009.pdf.rds"
## [44] "segmenteR_Li, Y et al 2010.pdf.rds"
## [45] "segmenteR_Liu, Y et al 2014.pdf.rds"
```

```
## [46] "segmenteR_Love, S A et al 2012.pdf.rds"
## [47] "segmenteR_Maurizi, L et al 2015.pdf.rds"
## [48] "segmenteR_Nešković, O et al 2013.pdf.rds"
## [49] "segmenteR_Neun, B W et al 2018.pdf.rds"
## [50] "segmenteR_Reddy, A et al 2010.pdf.rds"
## [51] "segmenteR_Richter, AW et al 1983.pdf.rds"
## [52] "segmenteR_Sadaf, A et al 2012.pdf.rds"
## [53] "segmenteR_Tang, J et al 2009.pdf.rds"
## [54] "segmenteR_Vicente, S et al 2017.pdf.rds"
## [55] "segmenteR_Wang, X et al 2010.pdf.rds"
## [56] "segmenteR_Xiaoli, F et al 2017.pdf.rds"
## [57] "segmenteR_Yan, M et al 2011.pdf.rds"
## [58] "segmenteR_Yu, T et al 2011.pdf.rds"
## [59] "segmenteR_Zook, J M et al 2011.pdf.rds"
## [60] "Tang, J et al 2009.pdf.rds"
## [61] "Vicente, S et al 2017.pdf.rds"
## [62] "Wang, X et al 2010.pdf.rds"
## [63] "Xiaoli, F et al 2017.pdf.rds"
## [64] "Yan, M et al 2011.pdf.rds"
## [65] "Yu, T et al 2011.pdf.rds"
## [66] "Zook, J M et al 2011.pdf.rds"
```

```
automatic_extraction<-list.files("Validation_tool_folder/", pattern = "segmenteR")
automatic_extraction
```

```
##  [1] "segmenteR_Abbasalipourkabir R et al 2015.pdf.rds"
##  [2] "segmenteR_Abe, S et al 2012.pdf.rds"
##  [3] "segmenteR_Adedara, I A et al 2018.pdf.rds"
##  [4] "segmenteR_Aijie, C et al 2017.pdf.rds"
##  [5] "segmenteR_Amrollahi-Sharifabadi et al. 2018.pdf.rds"
##  [6] "segmenteR_Ansari, M 2018.pdf.rds"
##  [7] "segmenteR_Aragao-Santiago, L 2016.pdf.rds"
##  [8] "segmenteR_AshaRani et al. 2009.pdf.rds"
##  [9] "segmenteR_Bai, K J et al 2018.pdf.rds"
## [10] "segmenteR_Bendre, V et al 2011.pdf.rds"
## [11] "segmenteR_Blum, J L et al 2014.pdf.rds"
## [12] "segmenteR_Cabral, H et al 2005.pdf.rds"
## [13] "segmenteR_Chen et al 2008.pdf.rds"
## [14] "segmenteR_Jamshidzadeh, A et al 2015.pdf.rds"
## [15] "segmenteR_Li, C et al 2009.pdf.rds"
## [16] "segmenteR_Li, J et al 2013.pdf.rds"
## [17] "segmenteR_Li, J G et al 2009.pdf.rds"
## [18] "segmenteR_Li, Y et al 2010.pdf.rds"
## [19] "segmenteR_Liu, Y et al 2014.pdf.rds"
## [20] "segmenteR_Love, S A et al 2012.pdf.rds"
## [21] "segmenteR_Maurizi, L et al 2015.pdf.rds"
## [22] "segmenteR_Nešković, O et al 2013.pdf.rds"
## [23] "segmenteR_Neun, B W et al 2018.pdf.rds"
## [24] "segmenteR_Reddy, A et al 2010.pdf.rds"
## [25] "segmenteR_Richter, AW et al 1983.pdf.rds"
## [26] "segmenteR_Sadaf, A et al 2012.pdf.rds"
## [27] "segmenteR_Tang, J et al 2009.pdf.rds"
## [28] "segmenteR_Vicente, S et al 2017.pdf.rds"
## [29] "segmenteR_Wang, X et al 2010.pdf.rds"
## [30] "segmenteR_Xiaoli, F et al 2017.pdf.rds"
```

```
## [31] "segmenteR_Yan, M et al 2011.pdf.rds"
## [32] "segmenteR_Yu, T et al 2011.pdf.rds"
## [33] "segmenteR_Zook, J M et al 2011.pdf.rds"
```

```
manual_extraction<-all_sections[!all_sections %in% automatic_extraction]
```

```
benchmarck_df <- data.frame(manual_extraction, automatic_extraction)
benchmarck_df
```

```
##                            manual_extraction
## 1     Abbasalipourkabir R et al 2015.pdf.rds
## 2                  Abe, S et al 2012.pdf.rds
## 3             Adedara, I A et al 2018.pdf.rds
## 4                Aijie, C et al 2017.pdf.rds
## 5  Amrollahi-Sharifabadi et al. 2018.pdf.rds
## 6                      Ansari, M 2018.pdf.rds
## 7             Aragao-Santiago, L 2016.pdf.rds
## 8                AshaRani et al. 2009.pdf.rds
## 9                 Bai, K J et al 2018.pdf.rds
## 10               Bendre, V et al 2011.pdf.rds
## 11               Blum, J L et al 2014.pdf.rds
## 12               Cabral, H et al 2005.pdf.rds
## 13                    Chen et al 2008.pdf.rds
## 14        Jamshidzadeh, A et al 2015.pdf.rds
## 15                   Li, C et al 2009.pdf.rds
## 16                   Li, J et al 2013.pdf.rds
## 17                 Li, J G et al 2009.pdf.rds
## 18                   Li, Y et al 2010.pdf.rds
## 19                  Liu, Y et al 2014.pdf.rds
## 20               Love, S A et al 2012.pdf.rds
## 21              Maurizi, L et al 2015.pdf.rds
## 22             Nešković, O et al 2013.pdf.rds
## 23               Neun, B W et al 2018.pdf.rds
## 24                Reddy, A et al 2010.pdf.rds
## 25             Richter, AW et al 1983.pdf.rds
## 26                Sadaf, A et al 2012.pdf.rds
## 27                 Tang, J et al 2009.pdf.rds
## 28              Vicente, S et al 2017.pdf.rds
## 29                 Wang, X et al 2010.pdf.rds
## 30              Xiaoli, F et al 2017.pdf.rds
## 31                  Yan, M et al 2011.pdf.rds
## 32                   Yu, T et al 2011.pdf.rds
## 33              Zook, J M et al 2011.pdf.rds
##                                  automatic_extraction
## 1     segmenteR_Abbasalipourkabir R et al 2015.pdf.rds
## 2                  segmenteR_Abe, S et al 2012.pdf.rds
## 3             segmenteR_Adedara, I A et al 2018.pdf.rds
## 4                segmenteR_Aijie, C et al 2017.pdf.rds
## 5  segmenteR_Amrollahi-Sharifabadi et al. 2018.pdf.rds
## 6                      segmenteR_Ansari, M 2018.pdf.rds
## 7             segmenteR_Aragao-Santiago, L 2016.pdf.rds
## 8                segmenteR_AshaRani et al. 2009.pdf.rds
## 9                 segmenteR_Bai, K J et al 2018.pdf.rds
## 10               segmenteR_Bendre, V et al 2011.pdf.rds
## 11               segmenteR_Blum, J L et al 2014.pdf.rds
```

```
## 12              segmenteR_Cabral, H et al 2005.pdf.rds
## 13                   segmenteR_Chen et al 2008.pdf.rds
## 14        segmenteR_Jamshidzadeh, A et al 2015.pdf.rds
## 15                  segmenteR_Li, C et al 2009.pdf.rds
## 16                  segmenteR_Li, J et al 2013.pdf.rds
## 17                segmenteR_Li, J G et al 2009.pdf.rds
## 18                  segmenteR_Li, Y et al 2010.pdf.rds
## 19                 segmenteR_Liu, Y et al 2014.pdf.rds
## 20              segmenteR_Love, S A et al 2012.pdf.rds
## 21             segmenteR_Maurizi, L et al 2015.pdf.rds
## 22            segmenteR_Nešković, O et al 2013.pdf.rds
## 23              segmenteR_Neun, B W et al 2018.pdf.rds
## 24              segmenteR_Reddy, A et al 2010.pdf.rds
## 25            segmenteR_Richter, AW et al 1983.pdf.rds
## 26               segmenteR_Sadaf, A et al 2012.pdf.rds
## 27               segmenteR_Tang, J et al 2009.pdf.rds
## 28             segmenteR_Vicente, S et al 2017.pdf.rds
## 29                segmenteR_Wang, X et al 2010.pdf.rds
## 30              segmenteR_Xiaoli, F et al 2017.pdf.rds
## 31                 segmenteR_Yan, M et al 2011.pdf.rds
## 32                  segmenteR_Yu, T et al 2011.pdf.rds
## 33              segmenteR_Zook, J M et al 2011.pdf.rds
```

Function from this package [this package]https://rdrr.io/cran/qlcMatrix/man/sim.strings.html). More easy to compute than creating a dtm on the whole corpus (i.e., all the material and sections.)

```
#function to got the cosine similarity
library(qlcMatrix)
```

```
## Loading required package: Matrix
```

```
## Loading required package: slam
```

```
## Loading required package: sparsesvd
```

```
Manual<-readRDS("Validation_tool_folder/Abbasalipourkabir R et al 2015.pdf.rds")
Auto<-readRDS("Validation_tool_folder/segmenteR_Abbasalipourkabir R et al 2015.pdf.rds")
```

```
Manual_sentences<-unique(Manual$sentence)
Manual_sentences<-paste0(Manual_sentences, collapse = " ")
Manual_sentences
```

```
## [1] "2. Materials and methods 2. 1. Materials Zinc oxide nanopowder was purchased from Iranian Nanoma
```

```
Auto_sentences<-unique(Auto$sentence)
Auto_sentences<-paste0(Auto_sentences, collapse = " ")
Auto_sentences
```

```
## [1] "2. Materials and methods 2. 1. Materials Zinc oxide nanopowder was purchased from Iranian Nanoma
```

```
qlcMatrix::sim.strings(Manual_sentences, Auto_sentences)
```

```
## [1] 0.9999945
```

```
got_cosine_similiarty <- function(benchmarck_df_row) {

  Manual<-readRDS(paste0("Validation_tool_folder/", benchmarck_df_row$manual_extraction))
  Auto<-readRDS(paste0("Validation_tool_folder/", benchmarck_df_row$automatic_extraction))
```

```
  Manual_sentences<-unique(Manual$sentence)
  Manual_sentences<-paste0(Manual_sentences, collapse = " ")

  Auto_sentences<-unique(Auto$sentence)
  Auto_sentences<-paste0(Auto_sentences, collapse = " ")

  cosine_similarity <- qlcMatrix::sim.strings(Manual_sentences, Auto_sentences)
  return(cosine_similarity)
}
```

```
benchmarck_df$cosine_similarity <- 0
head(benchmarck_df)
```

```
##                           manual_extraction
## 1    Abbasalipourkabir R et al 2015.pdf.rds
## 2                 Abe, S et al 2012.pdf.rds
## 3           Adedara, I A et al 2018.pdf.rds
## 4               Aijie, C et al 2017.pdf.rds
## 5 Amrollahi-Sharifabadi et al. 2018.pdf.rds
## 6                   Ansari, M 2018.pdf.rds
##                                 automatic_extraction cosine_similarity
## 1    segmenteR_Abbasalipourkabir R et al 2015.pdf.rds                 0
## 2                 segmenteR_Abe, S et al 2012.pdf.rds                 0
## 3           segmenteR_Adedara, I A et al 2018.pdf.rds                 0
## 4               segmenteR_Aijie, C et al 2017.pdf.rds                 0
## 5 segmenteR_Amrollahi-Sharifabadi et al. 2018.pdf.rds                 0
## 6                   segmenteR_Ansari, M 2018.pdf.rds                 0
```

```
for (i in 1:33) {
  print(i)
  if (i == 10 |i == 25) {
    print(benchmarck_df[i,]$manual_extraction)
    print("Error")
    #to remove the error message in the benchmarck df due to the failure of try()
    benchmarck_df[i,]$cosine_similarity <- 0
    next

  }
  cosine_sim <-try(got_cosine_similiarty(benchmarck_df[i,]))
  benchmarck_df[i,]$cosine_similarity <- cosine_sim
}
```

```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
## [1] 9
## [1] 10
## [1] "Bendre, V et al 2011.pdf.rds"
## [1] "Error"
## [1] 11
```

```
## [1] 12
## [1] 13
## [1] 14
## [1] 15
## [1] 16
## [1] 17
## [1] 18
## [1] 19
## [1] 20
## [1] 21
## [1] 22
## [1] 23
## [1] 24
## [1] 25
## [1] "Richter, AW et al 1983.pdf.rds"
## [1] "Error"
## [1] 26
## [1] 27
## [1] 28
## [1] 29
## [1] 30
## [1] 31
## [1] 32
## [1] 33
```

```
benchmarck_df
```

```
##                            manual_extraction
## 1     Abbasalipourkabir R et al 2015.pdf.rds
## 2                   Abe, S et al 2012.pdf.rds
## 3             Adedara, I A et al 2018.pdf.rds
## 4                 Aijie, C et al 2017.pdf.rds
## 5  Amrollahi-Sharifabadi et al. 2018.pdf.rds
## 6                       Ansari, M 2018.pdf.rds
## 7             Aragao-Santiago, L 2016.pdf.rds
## 8                AshaRani et al. 2009.pdf.rds
## 9                 Bai, K J et al 2018.pdf.rds
## 10               Bendre, V et al 2011.pdf.rds
## 11               Blum, J L et al 2014.pdf.rds
## 12               Cabral, H et al 2005.pdf.rds
## 13                    Chen et al 2008.pdf.rds
## 14        Jamshidzadeh, A et al 2015.pdf.rds
## 15                  Li, C et al 2009.pdf.rds
## 16                  Li, J et al 2013.pdf.rds
## 17                Li, J G et al 2009.pdf.rds
## 18                  Li, Y et al 2010.pdf.rds
## 19                 Liu, Y et al 2014.pdf.rds
## 20              Love, S A et al 2012.pdf.rds
## 21             Maurizi, L et al 2015.pdf.rds
## 22            Nešković, O et al 2013.pdf.rds
## 23              Neun, B W et al 2018.pdf.rds
## 24                Reddy, A et al 2010.pdf.rds
## 25            Richter, AW et al 1983.pdf.rds
## 26               Sadaf, A et al 2012.pdf.rds
## 27                Tang, J et al 2009.pdf.rds
```

```
## 28             Vicente, S et al 2017.pdf.rds
## 29                Wang, X et al 2010.pdf.rds
## 30              Xiaoli, F et al 2017.pdf.rds
## 31                 Yan, M et al 2011.pdf.rds
## 32                  Yu, T et al 2011.pdf.rds
## 33              Zook, J M et al 2011.pdf.rds
##                                 automatic_extraction cosine_similarity
## 1     segmenteR_Abbasalipourkabir R et al 2015.pdf.rds         0.9999945
## 2                   segmenteR_Abe, S et al 2012.pdf.rds         1.0000000
## 3             segmenteR_Adedara, I A et al 2018.pdf.rds         0.9998625
## 4                segmenteR_Aijie, C et al 2017.pdf.rds         1.0000000
## 5  segmenteR_Amrollahi-Sharifabadi et al. 2018.pdf.rds         1.0000000
## 6                     segmenteR_Ansari, M 2018.pdf.rds         0.9999604
## 7             segmenteR_Aragao-Santiago, L 2016.pdf.rds         0.9947144
## 8                segmenteR_AshaRani et al. 2009.pdf.rds         0.9999569
## 9                 segmenteR_Bai, K J et al 2018.pdf.rds         1.0000000
## 10               segmenteR_Bendre, V et al 2011.pdf.rds         0.0000000
## 11               segmenteR_Blum, J L et al 2014.pdf.rds         1.0000000
## 12               segmenteR_Cabral, H et al 2005.pdf.rds         0.9999179
## 13                   segmenteR_Chen et al 2008.pdf.rds         1.0000000
## 14        segmenteR_Jamshidzadeh, A et al 2015.pdf.rds         0.9957538
## 15                  segmenteR_Li, C et al 2009.pdf.rds         1.0000000
## 16                  segmenteR_Li, J et al 2013.pdf.rds         0.9999685
## 17                segmenteR_Li, J G et al 2009.pdf.rds         0.9999832
## 18                  segmenteR_Li, Y et al 2010.pdf.rds         0.9980745
## 19                 segmenteR_Liu, Y et al 2014.pdf.rds         0.9998103
## 20               segmenteR_Love, S A et al 2012.pdf.rds         0.9994587
## 21              segmenteR_Maurizi, L et al 2015.pdf.rds         1.0000000
## 22             segmenteR_Nešković, O et al 2013.pdf.rds         0.9990717
## 23               segmenteR_Neun, B W et al 2018.pdf.rds         0.9917729
## 24               segmenteR_Reddy, A et al 2010.pdf.rds         0.9998762
## 25             segmenteR_Richter, AW et al 1983.pdf.rds         0.0000000
## 26                segmenteR_Sadaf, A et al 2012.pdf.rds         0.9971360
## 27                segmenteR_Tang, J et al 2009.pdf.rds         0.9957357
## 28              segmenteR_Vicente, S et al 2017.pdf.rds         0.7193796
## 29                segmenteR_Wang, X et al 2010.pdf.rds         0.9999475
## 30              segmenteR_Xiaoli, F et al 2017.pdf.rds         0.7341130
## 31                 segmenteR_Yan, M et al 2011.pdf.rds         0.9998822
## 32                  segmenteR_Yu, T et al 2011.pdf.rds         0.9993641
## 33              segmenteR_Zook, J M et al 2011.pdf.rds         1.0000000
```

What about the two articles on which it fails ?

Running the code line by line, on Bendre, V et al 2011.pdf, it appears that clean_title_journal(pdf_name, section_title_df) remove all section title because the article only does 2 pages.

Richter, AW et al 1983.pdf is a pdf, but a pdf of a scan, and has no information regarding or, well, text.