

Chapitre 7 : joining tables

Armelle le Court

20/05/2021

1 : COMBINING TABLES

load the package

```
library("rWSBIM1207")
library("dplyr")
```

join the two tables

```
data(jdf)
jdf1 #with acesssion nbr + organelle + entry

## # A tibble: 25 x 3
##   uniprot organelle entry
##   <chr>   <chr>   <chr>
## 1 P26039 Actin cytoskeleton TLN1_MOUSE
## 2 Q99PL5 Endoplasmic reticulum/Golgi apparatus RRBP1_MOUSE
## 3 Q6PB66 Mitochondrion LPPRC_MOUSE
## 4 P11276 Extracellular matrix FINC_MOUSE
## 5 Q6PR54 Nucleus - Chromatin RIF1_MOUSE
## 6 Q05793 Extracellular matrix PGBM_MOUSE
## 7 P19096 Cytosol FAS_MOUSE
## 8 Q9JKF1 Plasma membrane IQGA1_MOUSE
## 9 Q9QZQ1-2 Plasma membrane AFAD_MOUSE
## 10 Q6NS46 Nucleus - Non-chromatin RRP5_MOUSE
## # ... with 15 more rows

jdf2 #with description of the gene + accession nbr + species the prot comes from

## # A tibble: 25 x 4
##   gene_name description uniprot organism
##   <chr>   <chr>   <chr>   <chr>
## 1 Iqgap1 Ras GTPase-activating-like protein IQGAP1 Q9JKF1 Mmus
## 2 Hspa5 78 kDa glucose-regulated protein P20029 Mmus
## 3 Pdcd11 Protein RRP5 homolog Q6NS46 Mmus
## 4 Tfrc Transferrin receptor protein 1 Q62351 Mmus
## 5 Hspd1 60 kDa heat shock protein, mitochondrial P63038 Mmus
## 6 Tln1 Talin-1 P26039 Mmus
## 7 Smc1a Structural maintenance of chromosomes protein 1A Q9CU62 Mmus
## 8 Lamc1 Laminin subunit gamma-1 P02468 Mmus
```

```
## 9 Hsp90b1 Endoplasmin P08113 Mmus
## 10 Mia3 Melanoma inhibitory activity protein 3 Q8BI84 Mmus
## # ... with 15 more rows
```

```
#to join those two tables we need "dplyr", it will be joined thanks to uniprot
full_join(jdf1,jdf2)
```

```
## Joining, by = "uniprot"
```

```
## # A tibble: 25 x 6
##   uniprot organelle      entry gene_name description      organism
##   <chr>   <chr>         <chr> <chr>    <chr>         <chr>
## 1 P26039 Actin cytoskele~ TLN1_~ Tln1      Talin-1          Mmus
## 2 Q99PL5 Endoplasmic ret~ RRBP1~ Rrbp1     Ribosome-binding protein~ Mmus
## 3 Q6PB66 Mitochondrion   LPPRC~ Lrprrc    Leucine-rich PPR motif-c~ Mmus
## 4 P11276 Extracellular m~ FINC_~ Fn1       Fibronectin      Mmus
## 5 Q6PR54 Nucleus - Chrom~ RIF1_~ Rif1      Telomere-associated prot~ Mmus
## 6 Q05793 Extracellular m~ PGBM_~ Hspg2     Basement membrane-specif~ Mmus
## 7 P19096 Cytosol         FAS_M~ Fasn      Fatty acid synthase     Mmus
## 8 Q9JKF1 Plasma membrane IQGA1~ Iqgap1    Ras GTPase-activating-li~ Mmus
## 9 Q9QZQ1-2 Plasma membrane AFAD_~ Mllt4     Isoform 1 of Afadin      Mmus
## 10 Q6NS46 Nucleus - Non-c~ RRP5_~ Pdcd11   Protein RRP5 homolog    Mmus
## # ... with 15 more rows
```

if different names

```
#in case none of the names match, we can set them manually using the by argument
#jdf1 and jdf3 have, for example, different names for uniprot
names(jdf3) #here it is "UniProt", and jdf1 is "uniprot"
```

```
## [1] "gene_name" "description" "UniProt" "organism"
full_join(jdf1, jdf3, by = c("uniprot" = "UniProt"))
```

```
## # A tibble: 25 x 6
##   uniprot organelle      entry gene_name description      organism
##   <chr>   <chr>         <chr> <chr>    <chr>         <chr>
## 1 P26039 Actin cytoskele~ TLN1_~ Tln1      Talin-1          Mmus
## 2 Q99PL5 Endoplasmic ret~ RRBP1~ Rrbp1     Ribosome-binding protein~ Mmus
## 3 Q6PB66 Mitochondrion   LPPRC~ Lrprrc    Leucine-rich PPR motif-c~ Mmus
## 4 P11276 Extracellular m~ FINC_~ Fn1       Fibronectin      Mmus
## 5 Q6PR54 Nucleus - Chrom~ RIF1_~ Rif1      Telomere-associated prot~ Mmus
## 6 Q05793 Extracellular m~ PGBM_~ Hspg2     Basement membrane-specif~ Mmus
## 7 P19096 Cytosol         FAS_M~ Fasn      Fatty acid synthase     Mmus
## 8 Q9JKF1 Plasma membrane IQGA1~ Iqgap1    Ras GTPase-activating-li~ Mmus
## 9 Q9QZQ1-2 Plasma membrane AFAD_~ Mllt4     Isoform 1 of Afadin      Mmus
## 10 Q6NS46 Nucleus - Non-c~ RRP5_~ Pdcd11   Protein RRP5 homolog    Mmus
## # ... with 15 more rows
```

QUESTION : Using the full_join function demonstrated above, join tables jdf4 and jdf5. What has happened for observations P26039 and P02468?

```
names(jdf4)
```

```
## [1] "uniprot" "organelle" "entry"
```

```
names(jdf5)
```

```
## [1] "gene_name" "description" "uniprot" "organism"
```

```
#they both have same name for "uniprot"
```

```
full_join(jdf4,jdf5)
```

```
## Joining, by = "uniprot"
```

```
## # A tibble: 14 x 6
```

	uniprot	organelle	entry	gene_name	description	organism
	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
## 1	P26039	Actin cytoskel~	TLN1_~	<NA>	<NA>	<NA>
## 2	Q99PL5	Endoplasmic re~	RRBP1~	<NA>	<NA>	<NA>
## 3	Q6PB66	Mitochondrion	LPPRC~	<NA>	<NA>	<NA>
## 4	P11276	Extracellular ~	FINC_~	<NA>	<NA>	<NA>
## 5	Q6PR54	Nucleus - Chro~	RIF1_~	<NA>	<NA>	<NA>
## 6	Q05793	Extracellular ~	PGBM_~	<NA>	<NA>	<NA>
## 7	P19096	Cytosol	FAS_M~	Fasn	Fatty acid synthase	Mmus
## 8	Q9JKF1	Plasma membrane	IQGA1~	<NA>	<NA>	<NA>
## 9	Q9QZQ1-2	Plasma membrane	AFAD_~	<NA>	<NA>	<NA>
## 10	Q6NS46	Nucleus - Non~	RRP5_~	<NA>	<NA>	<NA>
## 11	P02468	<NA>	<NA>	Lamc1	Laminin subunit gamma-1	Mmus
## 12	P08113	<NA>	<NA>	Hsp90b1	Endoplasmin	Mmus
## 13	Q8BI84	<NA>	<NA>	Mia3	Melanoma inhibitory activ~	Mmus
## 14	Q6P5D8	<NA>	<NA>	Smchd1	Structural maintenance of~	Mmus

```
#P02468 and P02468 are only present in jdf4 and jdf5 respectively and their respective values for the v
```

2 : DIFFERENT TYPES OF JOINS

An inner join keeps observations that are present in both tables. A left join keeps observations that are present in the left (first) table, dropping those that are only present in the other. A right join keeps observations that are present in the right (second) table, dropping those that are only present in the other. A full join keeps all observations.

QUESTIONS

```
#Join tables jdf4 and jdf5, keeping only observations in jdf4.
```

```
left_join(jdf4,jdf5)
```

```
## Joining, by = "uniprot"
```

```
## # A tibble: 10 x 6
```

	uniprot	organelle	entry	gene_name	description	organism
	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
## 1	P26039	Actin cytoskeleton	TLN1_MO~	<NA>	<NA>	<NA>
## 2	Q99PL5	Endoplasmic reticulum/G~	RRBP1_M~	<NA>	<NA>	<NA>
## 3	Q6PB66	Mitochondrion	LPPRC_M~	<NA>	<NA>	<NA>
## 4	P11276	Extracellular matrix	FINC_MO~	<NA>	<NA>	<NA>
## 5	Q6PR54	Nucleus - Chromatin	RIF1_MO~	<NA>	<NA>	<NA>
## 6	Q05793	Extracellular matrix	PGBM_MO~	<NA>	<NA>	<NA>
## 7	P19096	Cytosol	FAS_MOU~	Fasn	Fatty acid syn~	Mmus

```
## 8 Q9JKF1 Plasma membrane IQGA1_M~ <NA> <NA> <NA>
## 9 Q9QZQ1-2 Plasma membrane AFAD_MO~ <NA> <NA> <NA>
## 10 Q6NS46 Nucleus - Non-chromatin RRP5_MO~ <NA> <NA> <NA>

# Join tables jdf4 and jdf5, keeping only observations in jdf5.
right_join(jdf4,jdf5)

## Joining, by = "uniprot"

## # A tibble: 5 x 6
##   uniprot organelle entry gene_name description organism
##   <chr> <chr> <chr> <chr> <chr> <chr>
## 1 P19096 Cytosol FAS_MO~ Fasn Fatty acid synthase Mmus
## 2 P02468 <NA> <NA> Lamc1 Laminin subunit gamma-1 Mmus
## 3 P08113 <NA> <NA> Hsp90b1 Endoplasmin Mmus
## 4 Q8BI84 <NA> <NA> Mia3 Melanoma inhibitory activity pro~ Mmus
## 5 Q6P5D8 <NA> <NA> Smchd1 Structural maintenance of chromo~ Mmus

#Join tables jdf4 and jdf5, keeping only observations in jdf5.
full_join(jdf4,jdf5)
```

```
## Joining, by = "uniprot"

## # A tibble: 14 x 6
##   uniprot organelle entry gene_name description organism
##   <chr> <chr> <chr> <chr> <chr> <chr>
## 1 P26039 Actin cytoskel~ TLN1_~ <NA> <NA> <NA>
## 2 Q99PL5 Endoplasmic re~ RRBP1~ <NA> <NA> <NA>
## 3 Q6PB66 Mitochondrion LPPRC~ <NA> <NA> <NA>
## 4 P11276 Extracellular ~ FINC_~ <NA> <NA> <NA>
## 5 Q6PR54 Nucleus - Chro~ RIF1_~ <NA> <NA> <NA>
## 6 Q05793 Extracellular ~ PGBM_~ <NA> <NA> <NA>
## 7 P19096 Cytosol FAS_M~ Fasn Fatty acid synthase Mmus
## 8 Q9JKF1 Plasma membrane IQGA1~ <NA> <NA> <NA>
## 9 Q9QZQ1-2 Plasma membrane AFAD_~ <NA> <NA> <NA>
## 10 Q6NS46 Nucleus - Non~~ RRP5_~ <NA> <NA> <NA>
## 11 P02468 <NA> <NA> Lamc1 Laminin subunit gamma-1 Mmus
## 12 P08113 <NA> <NA> Hsp90b1 Endoplasmin Mmus
## 13 Q8BI84 <NA> <NA> Mia3 Melanoma inhibitory activ~ Mmus
## 14 Q6P5D8 <NA> <NA> Smchd1 Structural maintenance of~ Mmus
```

3 : MULTIPLES MATCHES

jdf6 -> the accession number Q99PL5 is repeated twice

```
jdf6

## # A tibble: 5 x 4
##   uniprot organelle entry isoform
##   <chr> <chr> <chr> <dbl>
## 1 P26039 Actin cytoskeleton TLN1_MOUSE 1
## 2 Q99PL5 Endoplasmic reticulum RRBP1_MOUSE 1
## 3 Q99PL5 Golgi apparatus RRBP1_MOUSE 2
## 4 Q6PB66 Mitochondrion LPPRC_MOUSE 1
## 5 P11276 Extracellular matrix FINC_MOUSE 1
```

#If we now want to join jdf6 and jdf2, the variables of the latter will be duplicated.
`full_join(jdf6,jdf2)`

```
## Joining, by = "uniprot"

## # A tibble: 26 x 7
##   uniprot organelle   entry   isoform gene_name description      organism
##   <chr>   <chr>       <chr>   <dbl> <chr>   <chr>       <chr>
## 1 P26039 Actin cytos~ TLN1_M~      1 Tln1     Talin-1      Mmus
## 2 Q99PL5 Endoplasmic~ RRB1_~      1 Rrbp1    Ribosome-binding pro~ Mmus
## 3 Q99PL5 Golgi appar~ RRB1_~      2 Rrbp1    Ribosome-binding pro~ Mmus
## 4 Q6PB66 Mitochondri~ LPPRC_~      1 Lrprrc   Leucine-rich PPR mot~ Mmus
## 5 P11276 Extracellul~ FINC_M~      1 Fn1      Fibronectin      Mmus
## 6 Q9JKF1 <NA>         <NA>         NA Iqgap1    Ras GTPase-activatin~ Mmus
## 7 P20029 <NA>         <NA>         NA Hspa5     78 kDa glucose-regul~ Mmus
## 8 Q6NS46 <NA>         <NA>         NA Pdc11     Protein RRP5 homolog Mmus
## 9 Q62351 <NA>         <NA>         NA Tfrc      Transferrin receptor~ Mmus
## 10 P63038 <NA>         <NA>         NA Hsp1      60 kDa heat shock pr~ Mmus
## # ... with 16 more rows
```

In the case above, repeating is useful, as it completes jdf6 with correct information from jdf2. One needs however to be careful when duplicated keys exist in both tables. Below, we create an inner join between jdf6 and jdf7, both having duplicated Q99PL5 entries.

```
inner_join(jdf6,jdf7)
```

```
## Joining, by = "uniprot"

## # A tibble: 4 x 9
##   uniprot organelle   entry   isoform gene_name description      organism isoform_num
##   <chr>   <chr>       <chr>   <dbl> <chr>   <chr>       <chr>       <dbl>
## 1 Q99PL5 Endoplasmic~ RRB1_~      1 Rrbp1    Ribosome-bi~ Mmus           1
## 2 Q99PL5 Endoplasmic~ RRB1_~      1 Rrbp1    Ribosome-bi~ Mmus           2
## 3 Q99PL5 Golgi appa~ RRB1_~      2 Rrbp1    Ribosome-bi~ Mmus           1
## 4 Q99PL5 Golgi appa~ RRB1_~      2 Rrbp1    Ribosome-bi~ Mmus           2
## # ... with 1 more variable: measure <dbl>
```

4 : MATCHING ACCROSS MULTIPLES KEYS

we see that the duplicated UniProt accession numbers in the jdf6 and jdf7 tables refer to different isoforms of the same RRB1 gene. To uniquely identify isoforms, we need to consider two keys, namely the UniProt accession number (named uniprot in both tables) as well as the isoform number, called isoform and isoform_num respectively.

```
inner_join(jdf6, jdf7, by = c("uniprot" = "uniprot", "isoform" = "isoform_num"))
```

```
## # A tibble: 2 x 8
##   uniprot organelle   entry   isoform gene_name description      organism measure
##   <chr>   <chr>       <chr>   <dbl> <chr>   <chr>       <chr>       <dbl>
## 1 Q99PL5 Endoplasmic~ RRB1_~      1 Rrbp1    Ribosome-bind~ Mmus          102
## 2 Q99PL5 Golgi appar~ RRB1_~      2 Rrbp1    Ribosome-bind~ Mmus           3
```

QUESTION

Can you think of another way to merge tables jdf6 and jdf7 using the two keys?

```
i <- grep("isoform", names(jdf7))
names(jdf7)[i] <- "isoform"
inner_join(jdf6, jdf7)

## Joining, by = c("uniprot", "isoform")

## # A tibble: 2 x 8
##   uniprot organelle   entry   isoform gene_name description   organism measure
##   <chr>   <chr>       <chr>   <dbl> <chr>      <chr>      <chr>      <dbl>
## 1 Q99PL5 Endoplasmic~ RRBP1_~     1 Rrbp1    Ribosome-bind~ Mmus        102
## 2 Q99PL5 Golgi appar~ RRBP1_~     2 Rrbp1    Ribosome-bind~ Mmus         3
```

5 ROW AND COLUMN BINDING

first we need to create the dataframes d1 and d2

```
x <- c(1,2,3)
y <- c(1,2,3)
d1 <- data.frame(x,y)

a <- c(4,5)
b <- c(4,5)
d2 <- data.frame(a,b)
```

the number of columns dont match so there is a error if we use `try(rbind(d1,d2))`

now we know we have to match those names before rbinding

```
names(d2) <- names(d1)
rbind(d1, d2)
```

```
##   x y
## 1 1 1
## 2 2 2
## 3 3 3
## 4 4 4
## 5 5 5
```

If we want to bind to dataframes along their columns

```
# cbind(d2, d3)
```

rbind dataframes with identical column names but different variables, which end up all being coerced into characters

```
d4 <- data.frame(x = letters[1:2], y = letters[1:2])
str(rbind(d1, d4))
```

```
## 'data.frame':    5 obs. of  2 variables:
## $ x: chr  "1" "2" "3" "a" ...
## $ y: chr  "1" "2" "3" "a" ...
```

6 : ADDITIONAL EXERCISES

1

```
#?
```

2

2.1. : load packages

```
#1st, we have to load the packages and datas
library(rWSBIM1207)
data("clinical2")
data("expression")
```

2.2. : Inspect the expression data. How many samples are recorded?

```
str(expression)
```

```
## tibble[,8] [570 x 8] (S3: tbl_df/tbl/data.frame)
## $ sampleID: Factor w/ 570 levels "TCGA-05-4244-01A",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ patient : chr [1:570] "TCGA-05-4244" "TCGA-05-4249" "TCGA-05-4250" "TCGA-05-4382" ...
## $ type     : chr [1:570] "tumor" "tumor" "tumor" "tumor" ...
## $ A1BG     : num [1:570] 26 120.1 50.9 145.9 127.4 ...
## $ A1CF     : num [1:570] 0 0.322 0 0 0 ...
## $ A2BP1    : num [1:570] 1.75 1.61 0 0 0 ...
## $ A2LD1    : num [1:570] 135.5 89.1 151.1 112.1 87.6 ...
## $ A2ML1    : num [1:570] 0.349 1.61 0 4.786 0 ...
```

```
expression %>%
  count(sampleID)
```

```
## # A tibble: 570 x 2
##   sampleID          n
##   <fct>          <int>
## 1 TCGA-05-4244-01A      1
## 2 TCGA-05-4249-01A      1
## 3 TCGA-05-4250-01A      1
## 4 TCGA-05-4382-01A      1
## 5 TCGA-05-4384-01A      1
```

```
## 6 TCGA-05-4389-01A      1
## 7 TCGA-05-4390-01A      1
## 8 TCGA-05-4395-01A      1
## 9 TCGA-05-4396-01A      1
## 10 TCGA-05-4397-01A     1
## # ... with 560 more rows
```

2.3. : Join the expression and clinical2 tables by the patient reference, using the left_join and the right_join functions. Why are the results different?

```
names(clinical2)
```

```
## [1] "patientID"      "gender"          "years_at_diagnosis"
```

```
names(expression)
```

```
## [1] "sampleID" "patient"  "type"      "A1BG"      "A1CF"      "A2BP1"      "A2LD1"
## [8] "A2ML1"
```

```
head(clinical2)
```

```
## # A tibble: 6 x 3
##   patientID  gender years_at_diagnosis
##   <chr>      <chr>          <dbl>
## 1 TCGA-05-4249 male           67.2
## 2 TCGA-05-4382 male           68.1
## 3 TCGA-05-4384 male           66.9
## 4 TCGA-05-4389 male           70.3
## 5 TCGA-05-4390 female          58.7
## 6 TCGA-05-4395 male           76.6
```

```
head(expression)
```

```
## # A tibble: 6 x 8
##   sampleID      patient      type  A1BG  A1CF A2BP1 A2LD1 A2ML1
##   <fct>      <chr>      <chr> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 TCGA-05-4244-01A TCGA-05-4244 tumor  26.0  0     1.75 136.  0.349
## 2 TCGA-05-4249-01A TCGA-05-4249 tumor  120.  0.322 1.61 89.1  1.61
## 3 TCGA-05-4250-01A TCGA-05-4250 tumor  50.9  0     0    151.  0
## 4 TCGA-05-4382-01A TCGA-05-4382 tumor  146.  0     0    112.  4.79
## 5 TCGA-05-4384-01A TCGA-05-4384 tumor  127.  0     0    87.6  0
## 6 TCGA-05-4389-01A TCGA-05-4389 tumor  67.1 36.2  0    112. 36.6
```

#we have to join by the patientID and patient names

```
full_join(clinical2,expression, by = c("patientID" = "patient"))
```

```
## # A tibble: 573 x 10
##   patientID  gender years_at_diagnosis sampleID  type  A1BG  A1CF A2BP1 A2LD1
##   <chr>      <chr>          <dbl> <fct>      <chr> <dbl> <dbl> <dbl> <dbl>
## 1 TCGA-05-4~ male           67.2 TCGA-05-- tumor  120.  0.322 1.61 89.1
## 2 TCGA-05-4~ male           68.1 TCGA-05-- tumor  146.  0     0    112.
## 3 TCGA-05-4~ male           66.9 TCGA-05-- tumor  127.  0     0    87.6
## 4 TCGA-05-4~ male           70.3 TCGA-05-- tumor  67.1 36.2  0    112.
## 5 TCGA-05-4~ female          58.7 TCGA-05-- tumor  165.  0    97.7  64.9
## 6 TCGA-05-4~ male           76.6 TCGA-05-- tumor  22.0  0     0    181.
## 7 TCGA-05-4~ male           77.0 TCGA-05-- tumor  17.4 15.9  0    134.
```



```
## 8 TCGA-05-4~ female          47.9 TCGA-05-- tumor 258.  0.218  0.435 149.
## 9 TCGA-05-4~ female          57.0 TCGA-05-- tumor  48.5  0      0      62.4
## 10 TCGA-05-4~ male           76.4 TCGA-05-- tumor  64.2  1.95  0      174.
## # ... with 563 more rows, and 1 more variable: A2ML1 <dbl>
```

2.4. : Join expression and clinical2 tables in order to create a table containing merged data exclusively for normal samples.

```
expression_norm <- expression %>%
  filter(type == "normal")

expression_norm
```

```
## # A tibble: 58 x 8
##   sampleID      patient    type    A1BG  A1CF A2BP1 A2LD1  A2ML1
##   <fct>         <chr>    <chr> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 TCGA-38-4625-11A TCGA-38-4625 normal  237.  0.375  1.12  91.5  0
## 2 TCGA-38-4626-11A TCGA-38-4626 normal  244.  0      0     59.9  0.365
## 3 TCGA-38-4627-11A TCGA-38-4627 normal  368.  0      0.539 111.  0
## 4 TCGA-38-4632-11A TCGA-38-4632 normal   95.4 0      0     154.  2.89
## 5 TCGA-44-2655-11A TCGA-44-2655 normal   86.1 0      0.308  57.1  0.616
## 6 TCGA-44-2657-11A TCGA-44-2657 normal  139.  0.569 0      68.1  50.1
## 7 TCGA-44-2661-11A TCGA-44-2661 normal   74.3 0      0.474  70.2  0
## 8 TCGA-44-2662-11A TCGA-44-2662 normal   51.3 0      0      60.3  0
## 9 TCGA-44-2665-11A TCGA-44-2665 normal  104.  0      1.39   81.6  1.05
## 10 TCGA-44-2668-11A TCGA-44-2668 normal   99.7 0      0      78.7  0.768
## # ... with 48 more rows
```

```
inner_join(clinical2, expression_norm, by = c("patientID" = "patient"))
```

```
## # A tibble: 58 x 10
##   patientID  gender years_at_diagnos~ sampleID  type    A1BG  A1CF A2BP1 A2LD1
##   <chr>      <chr>          <dbl> <fct>    <chr> <dbl> <dbl> <dbl> <dbl>
## 1 TCGA-38-46~ female          57.3 TCGA-38-4~ norm~  244.  0      0     59.9
## 2 TCGA-38-46~ female          64.4 TCGA-38-4~ norm~  368.  0      0.539 111.
## 3 TCGA-38-46~ male           42.2 TCGA-38-4~ norm~   95.4 0      0     154.
## 4 TCGA-44-26~ female          65.4 TCGA-44-2~ norm~   86.1 0      0.308  57.1
## 5 TCGA-44-26~ female          74.8 TCGA-44-2~ norm~  139.  0.569 0      68.1
## 6 TCGA-44-26~ female          69.4 TCGA-44-2~ norm~   74.3 0      0.474  70.2
## 7 TCGA-44-26~ male           65.8 TCGA-44-2~ norm~   51.3 0      0      60.3
## 8 TCGA-44-26~ female          55.8 TCGA-44-2~ norm~  104.  0      1.39   81.6
## 9 TCGA-44-33~ female          74.2 TCGA-44-3~ norm~  105.  0      0.592  74.2
## 10 TCGA-44-56~ female          61.6 TCGA-44-5~ norm~  101.  0      0.936  65.1
## # ... with 48 more rows, and 1 more variable: A2ML1 <dbl>
```