

# Estudio poblacional de incidencia del cáncer de pulmón

Implicación de KRAS

*Pere Pavón*

*4 de junio de 2018*

## Contents

<b>Estadística global de incidencia de Cáncer con genes mutados en vía MAPK/ERK.</b>	<b>2</b>
Obtención de los datos estadísticos . . . . .	2
Estadística global de incidencia de Cáncer en el mundo y en Europa. . . . .	2
Preparación de los datos para las gráficas . . . . .	2
Gráficos comparativos estadísticos . . . . .	3
Incidencia del cáncer en la población mundial por Sexo . . . . .	3
Incidencia del cáncer en el continente europeo separado por Sexo . . . . .	4
Ranking de tipos de cáncer con mayor incidencia a nivel mundial . . . . .	4
Ranking de tipos de cáncer con mayor incidencia en el continente europeo . . . . .	5
Estadística de incidencia de Cáncer de pulmón en el mundo. . . . .	6
Preparación de los datos para las gráficas . . . . .	6
Gráficos comparativos estadísticos . . . . .	6
Incidencia del cáncer de pulmón en la población mundial por Sexo . . . . .	7
Incidencia de la mortalidad del cáncer de pulmón en el mundo separado por Sexo . . . . .	7
Comparativa de nuevos casos y de mortalidad del cáncer de pulmón en el mundo . . . . .	8
Comparativa de nuevos casos y de mortalidad del cáncer de pulmón en el mundo . . . . .	9
Estadística global de incidencia de Cáncer con genes mutados en vía MAPK/ERK. . . . .	9
Preparación y exploración de los datos . . . . .	9
Gráficos comparativos estadísticos . . . . .	12
Distribución de tipos de cáncer con mutaciones en genes de la vía MAPK/ERK . . . . .	12
Distribución de genes con mutaciones en la vía MAPK/ERK . . . . .	12
Cuadros de distribución de los tipos de cancer más frecuente con frecuencia de mutación de los genes . . . . .	13
<b>Anexo de código R realizado para la generación de las tablas y gráficas</b>	<b>17</b>
Estadística global de incidencia de Cáncer con genes mutados en vía MAPK/ERK. . . . .	18
Obtención de los datos estadísticos . . . . .	18
Estadística global de incidencia de Cáncer en el mundo y en Europa. . . . .	18
Preparación de los datos para las gráficas . . . . .	19
Gráficos comparativos estadísticos . . . . .	19
Incidencia del cáncer en la población mundial por Sexo . . . . .	19
Incidencia del cáncer en el continente europeo separado por Sexo . . . . .	20
Ranking de tipos de cáncer con mayor incidencia a nivel mundial . . . . .	20
Ranking de tipos de cáncer con mayor incidencia en el continente europeo . . . . .	21
Estadística de incidencia de Cáncer de pulmón en el mundo. . . . .	22
Preparación de los datos para las gráficas . . . . .	22
Gráficos comparativos estadísticos . . . . .	22
Incidencia del cáncer de pulmón en la población mundial por Sexo . . . . .	22
Incidencia de la mortalidad del cáncer de pulmón en el mundo separado por Sexo . . . . .	23
Comparativa de nuevos casos y de mortalidad del cáncer de pulmón en el mundo . . . . .	23
Comparativa de nuevos casos y de mortalidad del cáncer de pulmón en el mundo . . . . .	24
Estadística global de incidencia de Cáncer con genes mutados en vía MAPK/ERK. . . . .	25
Preparación y exploración de los datos . . . . .	26

Graficos comparativos estadísticos . . . . .	27
Distribución de tipos de cáncer con mutaciones en genes de la vía MAPK/ERK . . . .	27
Distribución de genes con mutaciones en la vía MAPK/ERK . . . . .	27
Cuadros de distribución de los tipos de cancer más frecuente con frecuencia de mutación de los genes . . . . .	28
<b>Referencias</b>	<b>29</b>

# Estadística global de incidencia de Cáncer con genes mutados en vía MAPK/ERK.

Los datos sobre la incidencia y mortalidad de cáncer a nivel mundial se han obtenido de IARC (2018) en consulta realizada en abril del 2018 y están basados a datos del año 2012. (URL:[http://globocan.iarc.fr/Pages/fact\\_sheets\\_population.aspx](http://globocan.iarc.fr/Pages/fact_sheets_population.aspx))

## Obtención de los datos estadísticos

Para realizar el estudio, se han cargado los datos desde los ficheros que están en el directorio `./Datos`, en los siguientes ficheros:

- **1-2 EstadisMundial-EuropaxTipoCancer.csv.** Contiene datos estadísticos de incidencia, mortalidad y prevalencia del cáncer a nivel mundial y del continente europeo (IARC).
- **3 EstadisMundialPulmonxArea.csv.** Contiene datos estadísticos de incidencia, mortalidad y prevalencia del cáncer de pulmón a nivel mundial, por diferentes subclasificaciones (IARC).
- **GenesMutadosMAPK-ERK\_DATA.csv.** Contiene datos de muestras de tejidos de cáncer con mutaciones en diferentes genes y en diferentes tipos de cáncer (COSMIC v84).

## Estadística global de incidencia de Cáncer en el mundo y en Europa.

En la tabla 1 se muestran los totales de incidencia, mortalidad y prevalencia en 5 años del cáncer a nivel mundial y de Europa. Los valores son en miles (x1000). ASR (age-standardised rate) es la tasa estandarizada por edad, una medida resumida de la tasa que tendría una población si tuviese una estructura de edad estándar, para evitar posibles diferencias entre poblaciones respecto a la edad. La tasa estandarizada por edad se calcula utilizando 10 grupos de edad (0-14; 15-39; 40-44; 45-49; 50-54; 55-59; 60-64; 65-69; 70-74; 75+).

Table 1: Distribución de Cáncer en el mundo por Sexo

Área	Sexo	Tipos	Incidencia		Mortalidad		Prevalencia
			Casos	ASR	Casos	ASR	5Años
Mundial	Hombre	All cancers excl. non-melanoma skin cancer	7410	204.9	4653	126.3	15296
Mundial	Mujer	All cancers excl. non-melanoma skin cancer	6658	165.2	3548	82.9	17159
Europa	Hombre	All cancers excl. non-melanoma skin cancer	1814	296.3	977	147.6	4508
Europa	Mujer	All cancers excl. non-melanoma skin cancer	1606	225.5	779	87.6	4571

## Preparación de los datos para las gráficas

Para poder efectuar las diferentes gráficas comparativas y tablas de datos, se van a preparar los datos cargados en los tres data frame de carga inicial, aplicando los filtros requeridos en función de las necesidades para cada tipo de gráfico.

```
# Mundial total
glMT <- subset(dataEst1, Estadistica == "Mundial" & Acum == "Total" & Sexo != "Ambos")
# Transforación de columnas a filas para la representación gráfica
glMT <- data.frame(TipoD=substr(colnames(glMT[c(5,8,11)]),1,
                             nchar(colnames(glMT[c(5,8,11)]))-1),
                  Sexo=rep(glMT$Sexo,each=3), Valor=matrix(t(glMT[,c(5,8,11)])))
# Cálculo porcentual por indicador
```

```

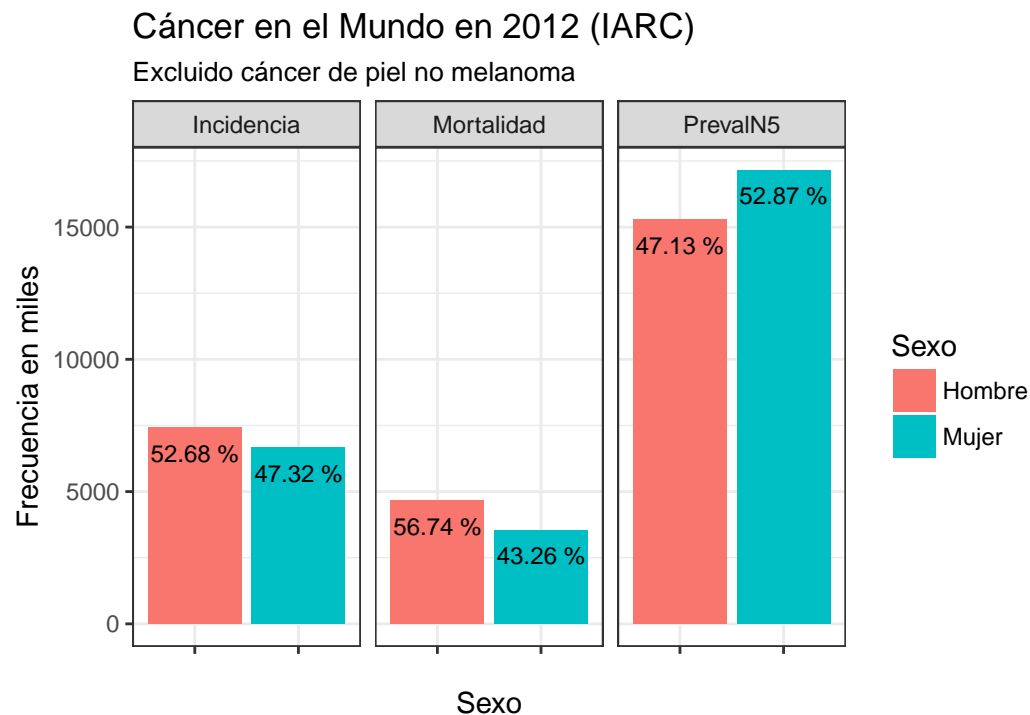
glMT <-glMT %>% # agrupamos por indicador
  group_by(TipoD) %>% # mutate para crear variable
  mutate(SexoProp = Valor/sum(Valor))
#
# Mundial por tipo de cáncer
glMc <- subset(dataEst1, Estadistica == "Mundial" & Acum == "TipoCancer" & Sexo == "Ambos")
#
# Europa total
glET <- subset(dataEst1, Estadistica == "Europa" & Acum == "Total" & Sexo != "Ambos")
# Transformación de columnas a filas para la representación gráfica
glET <-data.frame(TipoD=substr(colnames(glET[c(5,8,11)]),1,
                             nchar(colnames(glET[c(5,8,11)]))-1),
                  Sexo=rep(glET$Sexo,each=3), Valor=matrix(t(glET[,c(5,8,11)])))
# Cálculo porcentual por indicador
glET <-glET %>% # agrupamos por indicador
  group_by(TipoD) %>% # mutate para crear variable
  mutate(SexoProp = Valor/sum(Valor))
#
# Europa por tipo de cáncer
glEc <- subset(dataEst1, Estadistica == "Europa" & Acum == "TipoCancer" & Sexo == "Ambos")

```

## Gráficos comparativos estadísticos

### Incidencia del cáncer en la población mundial por Sexo

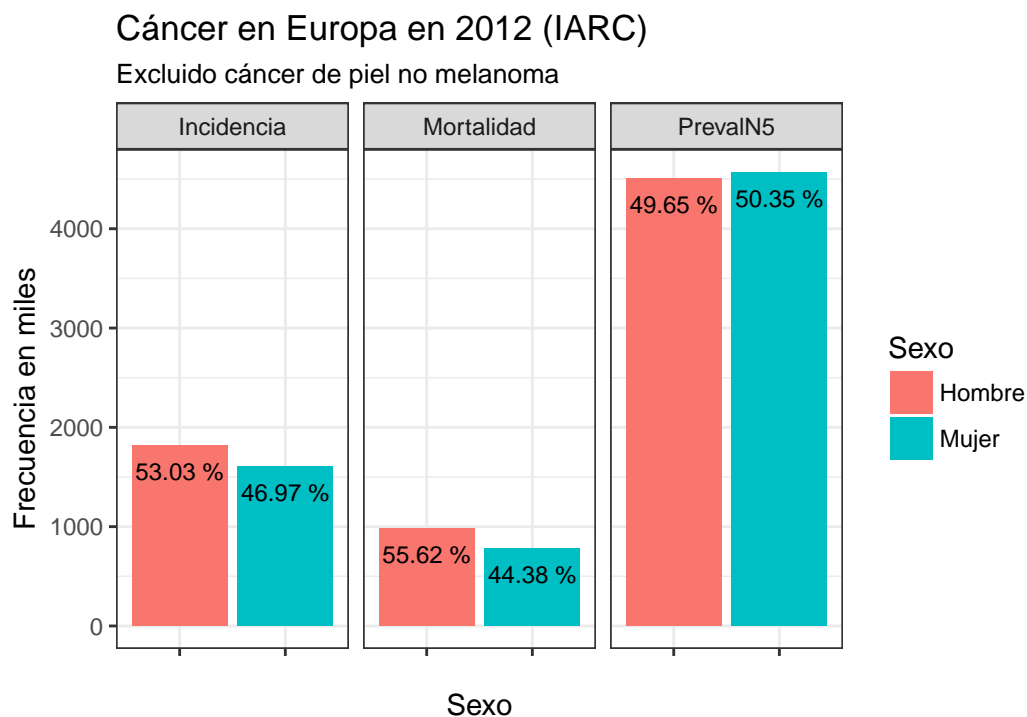
El valor mostrado en el gráfico Gr.1 representa la incidencia de cáncer a nivel mundial y no incluye el cáncer de piel no melanoma. Se presenta la proporción de incidencia por sexo, entre hombres y mujeres.



Gráfica Gr.1

### Incidencia del cáncer en el continente europeo separado por Sexo

El valor mostrado en el gráfico Gr.2 presenta la incidencia de cáncer en el continente Europeo. Al igual que en el Gr.1, no incluye el cáncer de piel no melanoma. También se presenta la proporción por sexo, entre hombres y mujeres.



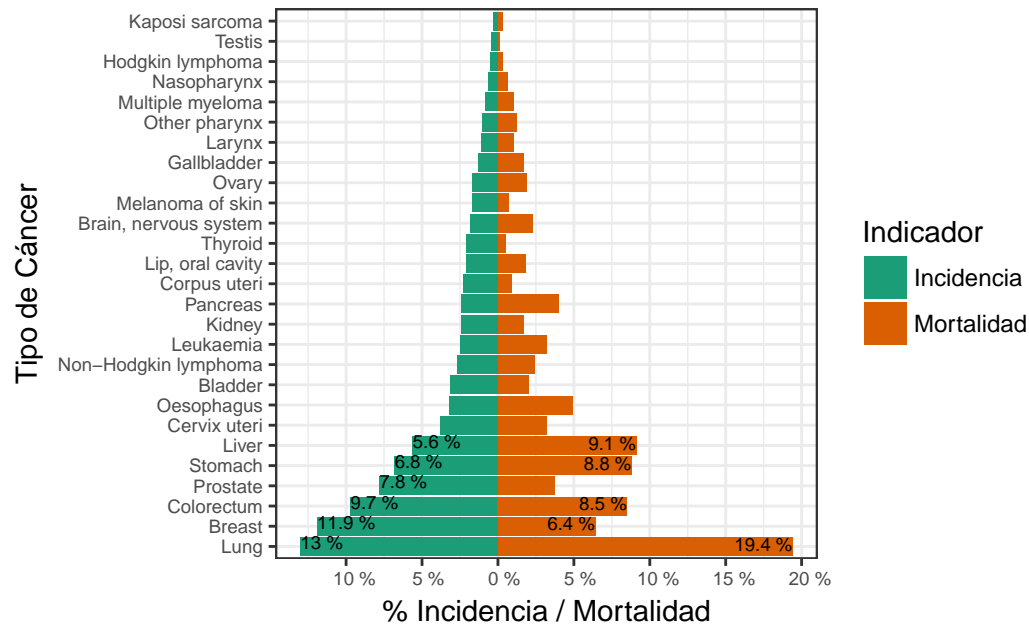
Gráfica Gr.2

### Ranking de tipos de cáncer con mayor incidencia a nivel mundial

En el gráfico Gr.3 se presentan los 20 tipos de cáncer con mayor incidencia sobre la población mundial, comparando los indicadores de incidencia, mortalidad y prevalencia en 5 años.

## Principales cánceres en el mundo en 2012 (IARC)

Excluido cáncer de piel no melanoma



Gráfica Gr.3

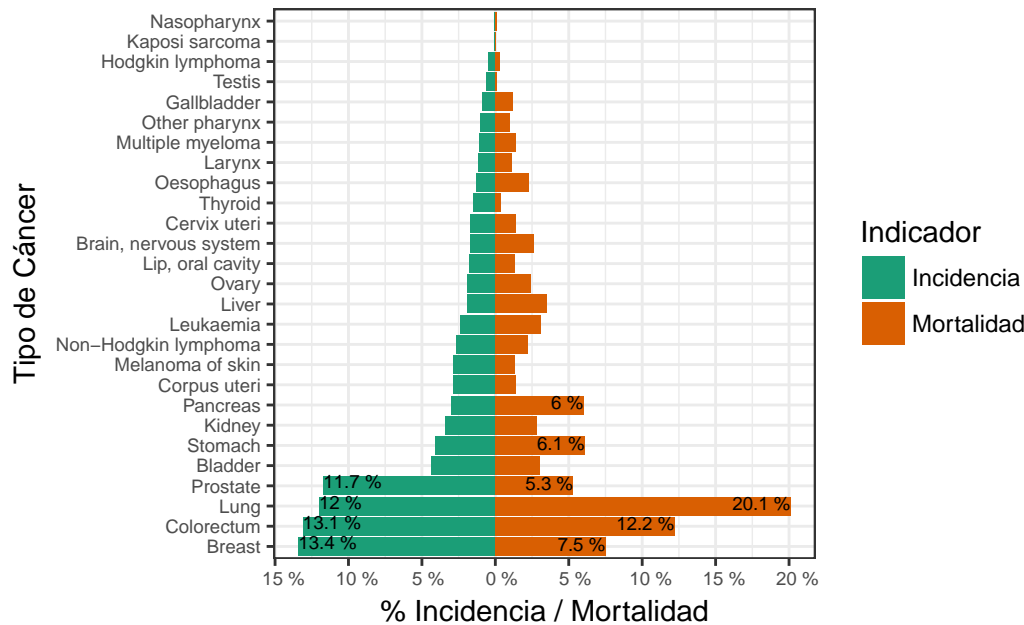
En la gráfica se indican los valores de los porcentajes superiores al 5 % de incidencia en la población.

### Ranking de tipos de cáncer con mayor incidencia en el continente europeo

En el gráfico Gr.4 se presentan los 20 tipos de cáncer con mayor incidencia en el continente europeo, comparando los indicadores de incidencia, mortalidad y prevalencia en 5 años.

## Principales cánceres en Europa en 2012 (IARC)

Excluido cáncer de piel no melanoma



Gráfica Gr.4

## Estadística de incidencia de Cáncer de pulmón en el mundo.

Centrando el foco de atención en el cáncer de pulmón, en este apartado se van a ver los datos más relevantes de la incidencia de este cáncer a nivel mundial según la distribución de países realizada por la Organización Mundial de la Salud (WHO) en base a la de recogida de datos y también según el nivel de desarrollo de los países.

### Preparación de los datos para las gráficas

En base a las necesidades de las gráficas a realizar, se van a preparar los datos aplicando los filtros requeridos para tratar solo los datos requeridos.

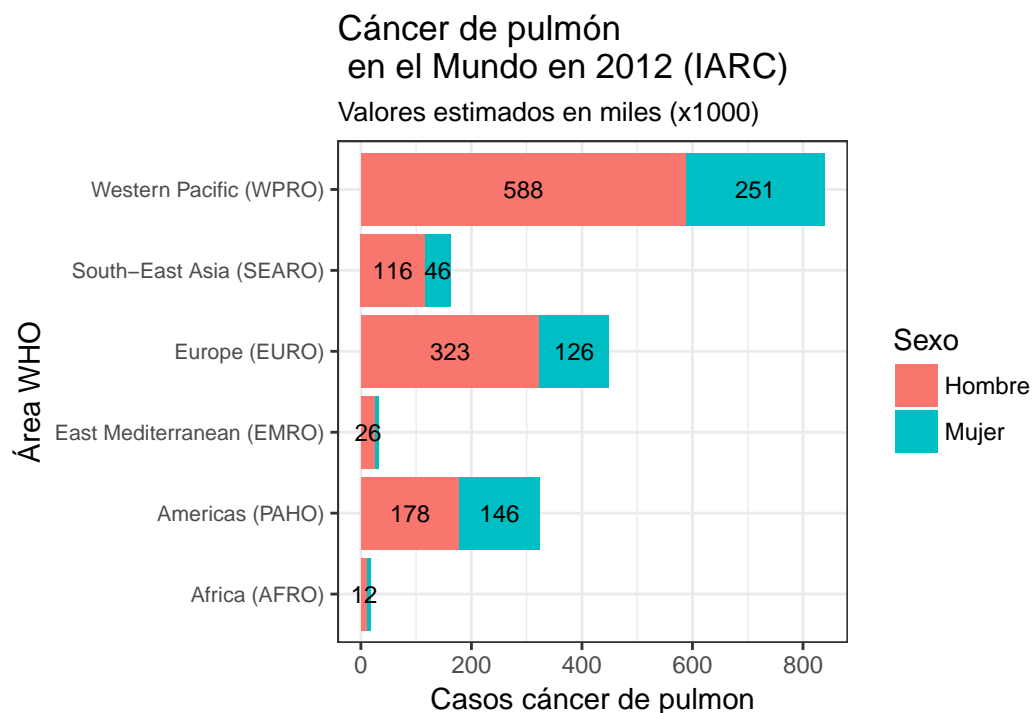
```
# Cancer de pulmón según división mundial de WHO x Sexo
glPwho <- subset(dataEst2, Tipo == "WHO" & Sexo != "Ambos")
# Cancer de pulmón según división mundial de WHO ambos sexos
glPwho2 <- subset(dataEst2, Tipo == "WHO" & Sexo == "Ambos")
# Cancer de pulmón según Nivel de desarrollo x Sexo
glPniv <- subset(dataEst2, Tipo == "Nivel" & Sexo != "Ambos")
# Cancer de pulmón según Nivel de desarrollo ambos sexos
glPniv2 <- subset(dataEst2, Tipo == "Nivel" & Sexo == "Ambos")
```

### Gráficos comparativos estadísticos

Las áreas de agrupación de países presentadas, hacen referencia a las agrupaciones confeccionadas por la OMS (WHO) según las diferentes recogidas de datos que realizan. Son valores en miles de casos (x1000).

### Incidencia del cáncer de pulmón en la población mundial por Sexo

El valor mostrado en el gráfico Gr5 de incidencia de cáncer de pulmón a nivel mundial, no incluye el cáncer de piel no melanoma. Se presenta la proporción de nuevos casos que se producen al año entre hombres y mujeres.



Gráfica Gr.5

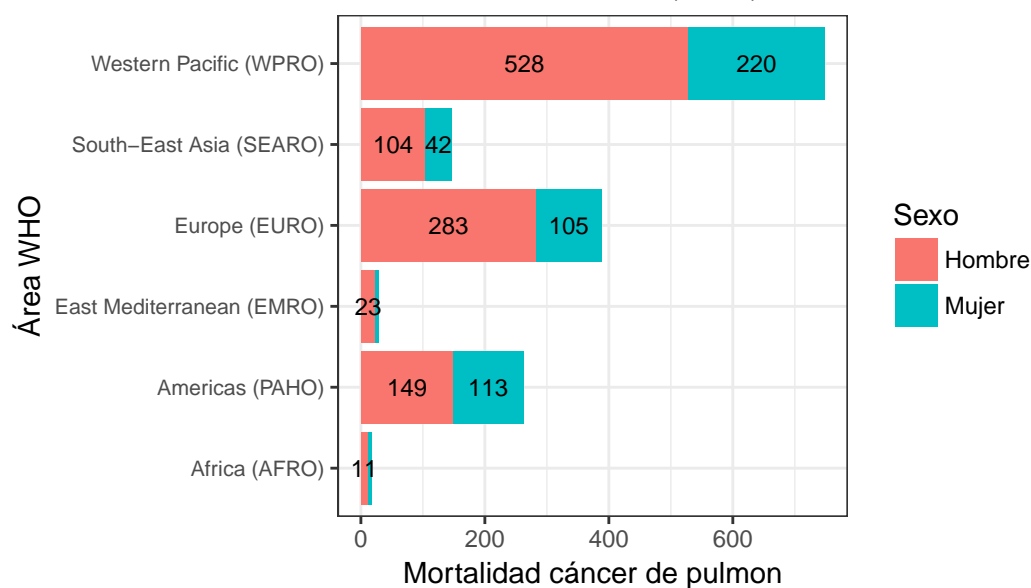
### Incidencia de la mortalidad del cáncer de pulmón en el mundo separado por Sexo

En el gráfico Gr6 se presentan los datos de mortalidad del cáncer de pulmón en el mundo, también según la proporción por sexo. Los valores presentados son en miles de casos (x1000).



## Mortalidad por cáncer de pulmón en el Mundo en 2012 (IARC)

Valores estimados en miles (x1000)



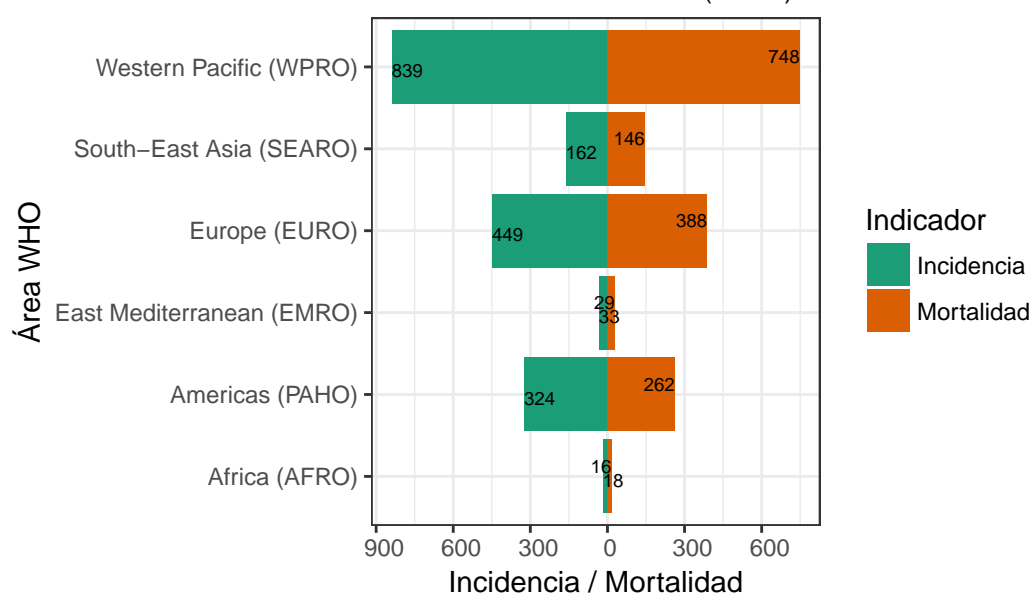
Gráfica Gr.6

## Comparativa de nuevos casos y de mortalidad del cáncer de pulmón en el mundo

El gráfico Gr7 se presenta una comparativa de los nuevos casos de cáncer de pulmón versus a la mortalidad que ocasionan, separado por áreas de la OMS. Son valores en miles de casos (x1000).

## Cáncer de pulmón en el Mundo en 2012 (IARC)

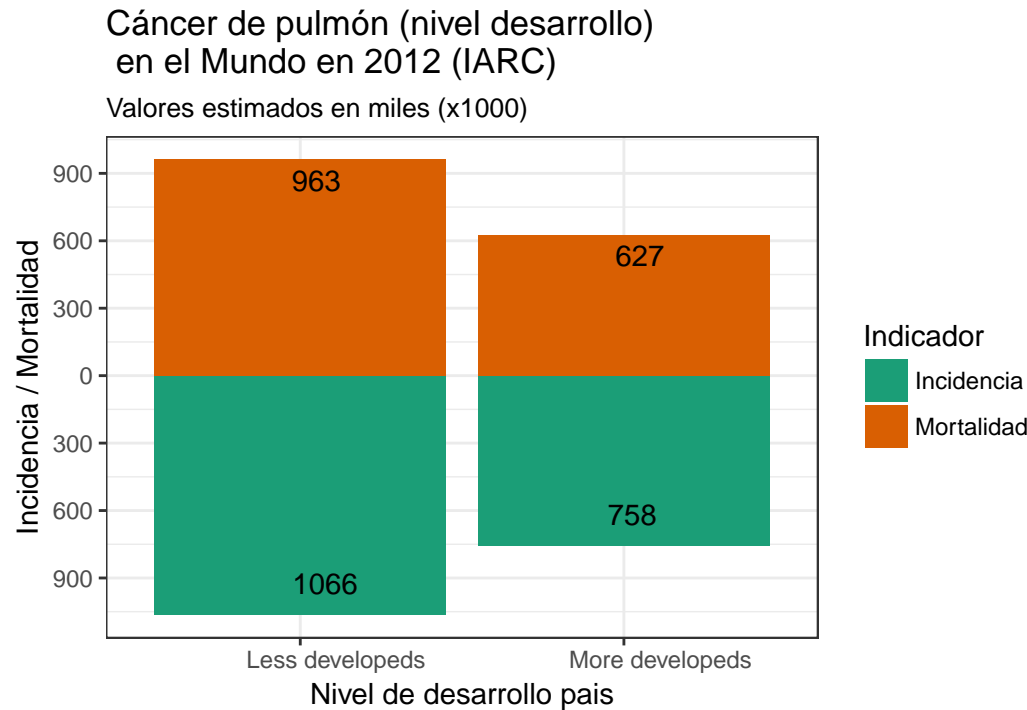
Valores estimados en miles (x1000)



Gráfica Gr.7

### Comparativa de nuevos casos y de mortalidad del cáncer de pulmón en el mundo

En el gráfico Gr8 se presenta una comparativa de los nuevos casos de cáncer de pulmón versus a la mortalidad que ocasionan, en función del nivel de desarrollo de los países, diferenciando entre nivel alto y nivel bajo de desarrollo. Son valores en miles de casos (x1000).



Gráfica Gr.8

### Estadística global de incidencia de Cáncer con genes mutados en vía MAPK/ERK.

Para realizar el estudio estadístico sobre la incidencia de los cánceres relacionados con los genes mutados en la vía MAPK/ERK, se han obtenido datos de muestras de tejidos con cáncer, registradas en v84 (2018).

Para determinar las muestras de interés para el estudio, se han seleccionado 12 los genes de mayor interés en la vía MAPK/ERK (ARAF, BRAF, EGFR, HRAS, KRAS, MAP2K1, MAP2K2, MAPK1, MAPK3, NRAS, PDGFRA, RAF1). La base total de registros obtenidas de v84 (2018) contiene 136999 registros de muestras de tejido de diferentes tipos de cáncer con genes mutados, en total hay registrados 38 tipos de cáncer diferentes.

### Preparación y exploración de los datos

A partir de los datos cargados de las muestras en el data frame de trabajo inicial, se preparan diferentes data frame con la estructura y campos requeridos para su presentación gráfica. En primer lugar, se muestran los diferentes tipos de cáncer y genes mutados en su totalidad. Posteriormente, para generar las siguientes gráficas comparativas que se muestran, se han seleccionado los genes mutados que tienen un porcentaje de frecuencia de mutación superior al 1%. Estos genes mutados son: BRAF, EGFR, HRAS, KRAS, NRAS,

PDGFRA. En referencia a los diferentes cáncer que presentan mutaciones en algunos de estos genes, también se han seleccionado aquellos que presentaban en total por encima del 1% de mutaciones.

```
#
# Data frame con datos de muestras por cáncer (Primary.Tissue)
tbe1.1 <- as.data.frame(table(dataEst3$Primary.Tissue))
colnames(tbe1.1) <- c("Primary.Tissue", "Freq")
# Proporciones
tbe1.1p <- as.data.frame(round(prop.table(table(dataEst3$Primary.Tissue))*100,2))
colnames(tbe1.1p) <- c("Primary.Tissue", "Freq")
#
# Data frame con datos de muestras por Gen mutado (Gene.Name)
tbe1.2 <- as.data.frame(table(dataEst3$Gene.Name))
colnames(tbe1.2) <- c("Gene.Name", "Freq")
# Proporciones
tbe1.2p <- as.data.frame(round(prop.table(table(dataEst3$Gene.Name))*100,2))
colnames(tbe1.2p) <- c("Gene.Name", "Freq")
#t1.1 <- t1.1[order(t1.1$Freq),]
#
# Data frame con datos de muestras por cáncer (Primary.Tissue) y Gen mutado (Gene.Name)
# (solo los significativos)
#
tbe1 <- as.data.frame(table(dataEst3$Primary.Tissue, dataEst3$Gene.Name))
colnames(tbe1) <- c("Primary.Tissue", "Gene.Name", "Freq")
#Selección de Cánceres y Genes más significativos
CanSel <- as.character(tbe1.1$Primary.Tissue[
  which(tbe1.1p$Freq>1&tbe1.1p$Primary.Tissue!="NS")])
CanSel2 <- as.numeric(apply(as.data.frame(CanSel), 1,
  function(par1) {which(tbe1$Primary.Tissue==par1)}))
tbe2 <- tbe1[CanSel2,]
GenSel <- as.character(tbe1.2$Gene.Name[which(tbe1.2p$Freq>1)])
GenSel2 <- as.numeric(apply(as.data.frame(GenSel), 1,
  function(par1) {which(tbe2$Gene.Name==par1)}))
tbe2 <- tbe2[GenSel2,]
PorcG <- as.numeric(apply(as.data.frame(tbe2[, 2:3]), 1,
  function(x){round(as.numeric(x[2])/tbe1.2[tbe1.2==x[1], 2]*100, 2)}))
tbe2 <- data.frame(tbe2, PorcG)
PorcC <- as.numeric(apply(as.data.frame(tbe2[, c(1, 3)]), 1,
  function(x){round(as.numeric(x[2])/tbe1.1[tbe1.1==x[1], 2]*100, 2)}))
tbe2 <- data.frame(tbe2, PorcC)
# Selección de KRAS mutado y cancer de pulmón
#
tbe3 <- subset(dataEst3, Gene.Name=="KRAS"&Primary.Tissue=="Lung")
# Reclasificación sobre los 3 tipos de cáncer NSCLC más relevantes y resto como otros.
tbe3$subtipo <- apply(data.frame(subset(dataEst3,
  Gene.Name=="KRAS"&Primary.Tissue=="Lung")[, 7]), 1,
  function(x) {if (x != 'Adenocarcinoma' & x != 'Squamous cell carcinoma'
    & x != 'Non small cell carcinoma') {"Other"} else {x}}})
```

En la tabla 2 se pueden ver los porcentajes de frecuencia de incidencia de los diferentes Tipos de cáncer que tienen algún gen mutado de los que participan en la vía de señalización de MAPK/ERK, antes de proceder a la selección de los más significativos.

En la tabla 3 se pueden ver los porcentajes de frecuencia de mutación de los genes en los anteriores tipos de cáncer, antes de la selección de los más significativos.

Table 2: Distribución frecuencias de Tipos de cáncer

Tipo Cáncer	Frecuencia	% Frecuencia
Adrenal gland	71	0.05
Autonomic ganglia	27	0.02
Biliary tract	905	0.66
Bone	106	0.08
Breast	392	0.29
Central nervous system	1092	0.80
Cervix	137	0.10
Endometrium	859	0.63
Eye	160	0.12
Gastrointestinal tract (site indeterminate)	76	0.06
Genital tract	38	0.03
Haematopoietic and lymphoid	5115	3.73
Kidney	146	0.11
Large intestine	37000	27.01
Liver	293	0.21
Lung	35088	25.61
Meninges	19	0.01
NS	2759	2.01
Oesophagus	163	0.12
Ovary	1299	0.95
Pancreas	6710	4.90
Parathyroid	2	0.00
Penis	27	0.02
Peritoneum	197	0.14
Pituitary	114	0.08
Pleura	29	0.02
Prostate	297	0.22
Salivary gland	133	0.10
Skin	14702	10.73
Small intestine	319	0.23
Soft tissue	1762	1.29
Stomach	583	0.43
Testis	52	0.04
Thymus	32	0.02
Thyroid	25326	18.49
Upper aerodigestive tract	461	0.34
Urinary tract	487	0.36
Vulva	21	0.02

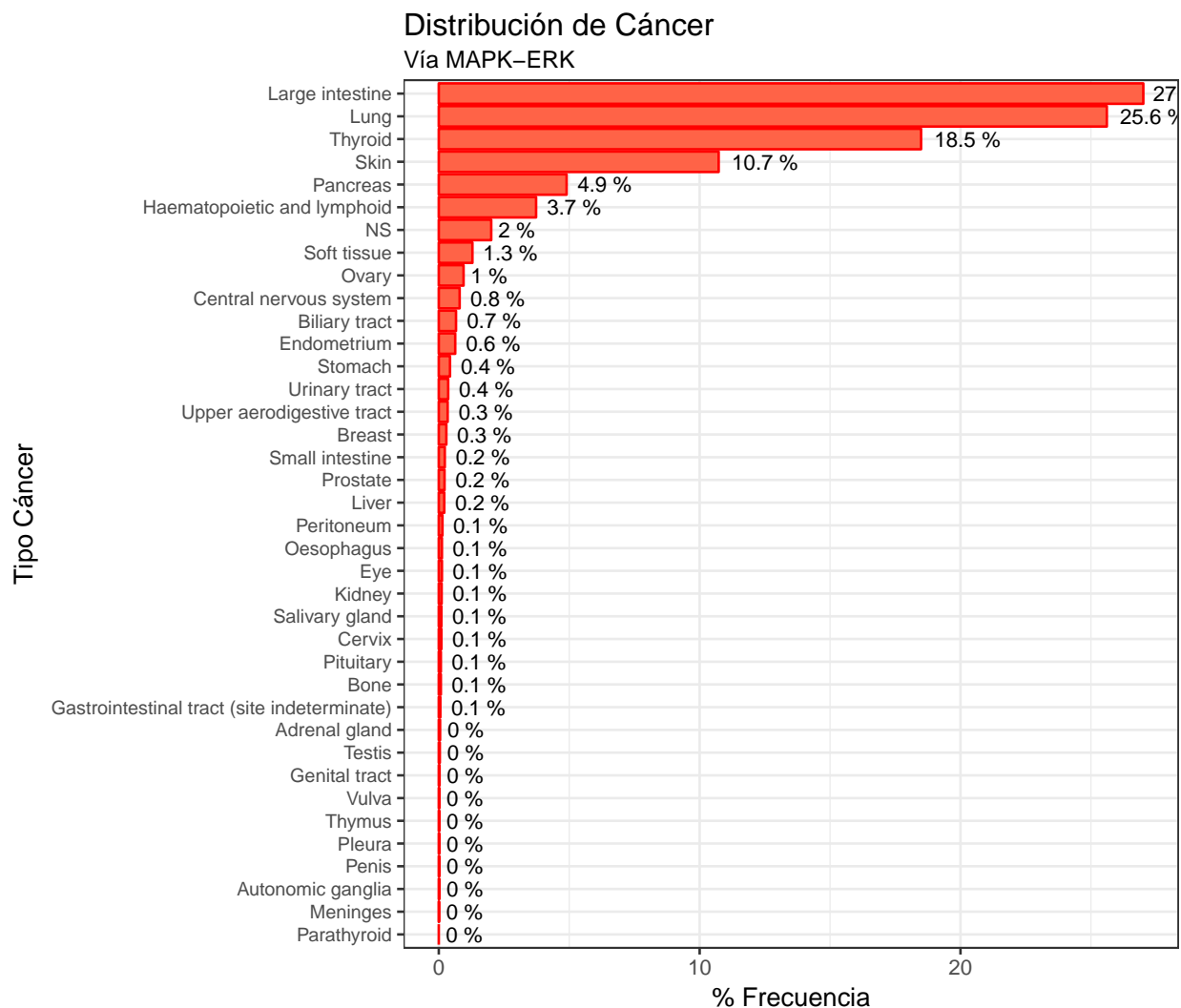
Table 3: Distribución frecuencias de Genes mutados

Gen mutado	Frecuencia	% Frecuencia
ARAF	214	0.16
BRAF	52249	38.14
EGFR	28939	21.12
HRAS	1933	1.41
KRAS	43640	31.85
MAP2K1	407	0.30
MAP2K2	289	0.21
MAPK1	132	0.10
MAPK3	88	0.06
NRAS	6607	4.82
PDGFRA	2279	1.66
RAF1	222	0.16

## Graficos comparativos estadísticos

### Distribución de tipos de cáncer con mutaciones en genes de la vía MAPK/ERK

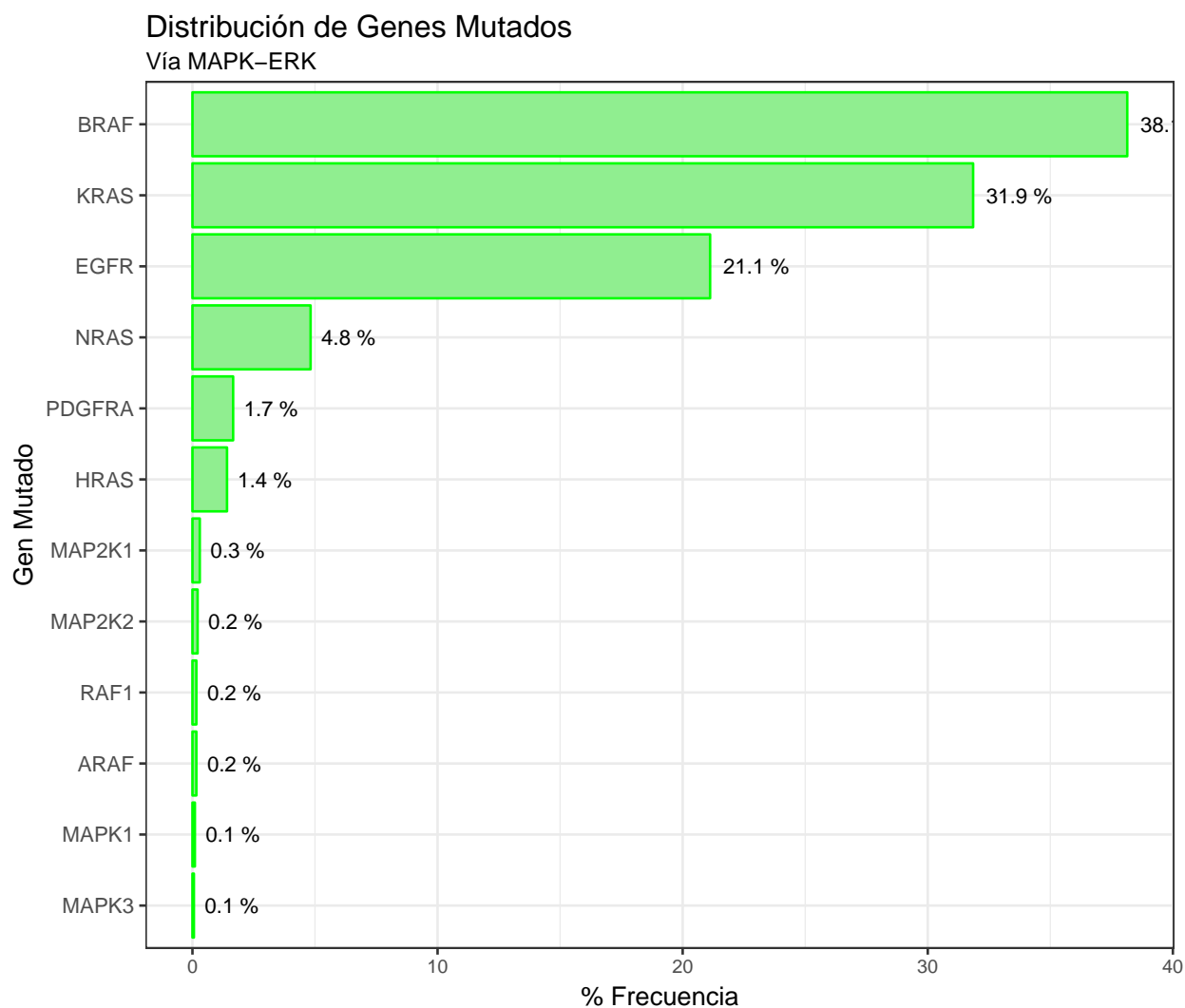
En la gráfica Gr9 se puede ver el ranking de los cánceres más frecuente con genes mutados en la vía MAPK/ERK. Se puede observar que los 6 primeros son los más frecuentes y entre ellos, en segundo lugar, se puede observar el cáncer de pulmón.



Gráfica Gr.9

### Distribución de genes con mutaciones en la vía MAPK/ERK

En la gráfica Gr10 se presenta la distribución de mutación de los diferentes genes estudiados. KRAS aparece en la segunda posición con el 31.9% de mutaciones.



*Gráfica Gr.10*

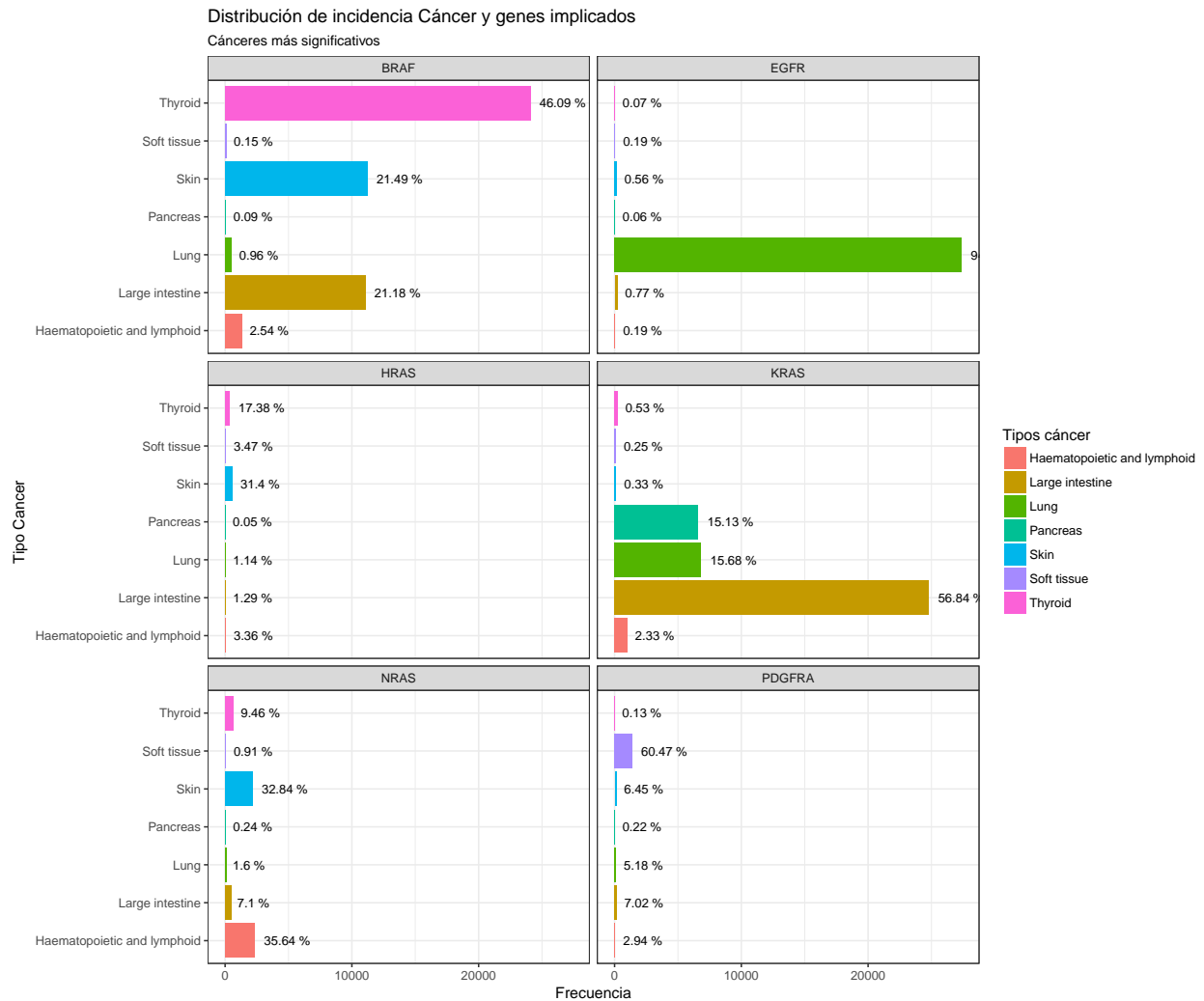
### **Cuadros de distribución de los tipos de cancer más frecuente con frecuencia de mutación de los genes**

Tal como se ha indicado en el apartado de preparación de datos, se ha efectuado un filtraje de los tipos de cáncer cuya frecuencia es superior al 1% en las muestras analizadas. Las gráficas que se muestran en adelante hace referencia a estos tipos de cáncer. En la tabla 3 se pueden ver los datos numéricos de la selección de cánceres y de los genes mutados implicados.

En la gráfica Gr11 se pueden observar los genes mutados por cada tipo de cáncer, junto a la frecuencia de cada uno de ellos. Se puede comprobar que KRAS aparece con una mayor frecuencia en el cáncer de Colón, de pulmón y de páncreas.

Table 4: Distribución frecuencias de Tipos de cáncer y Genes mutados

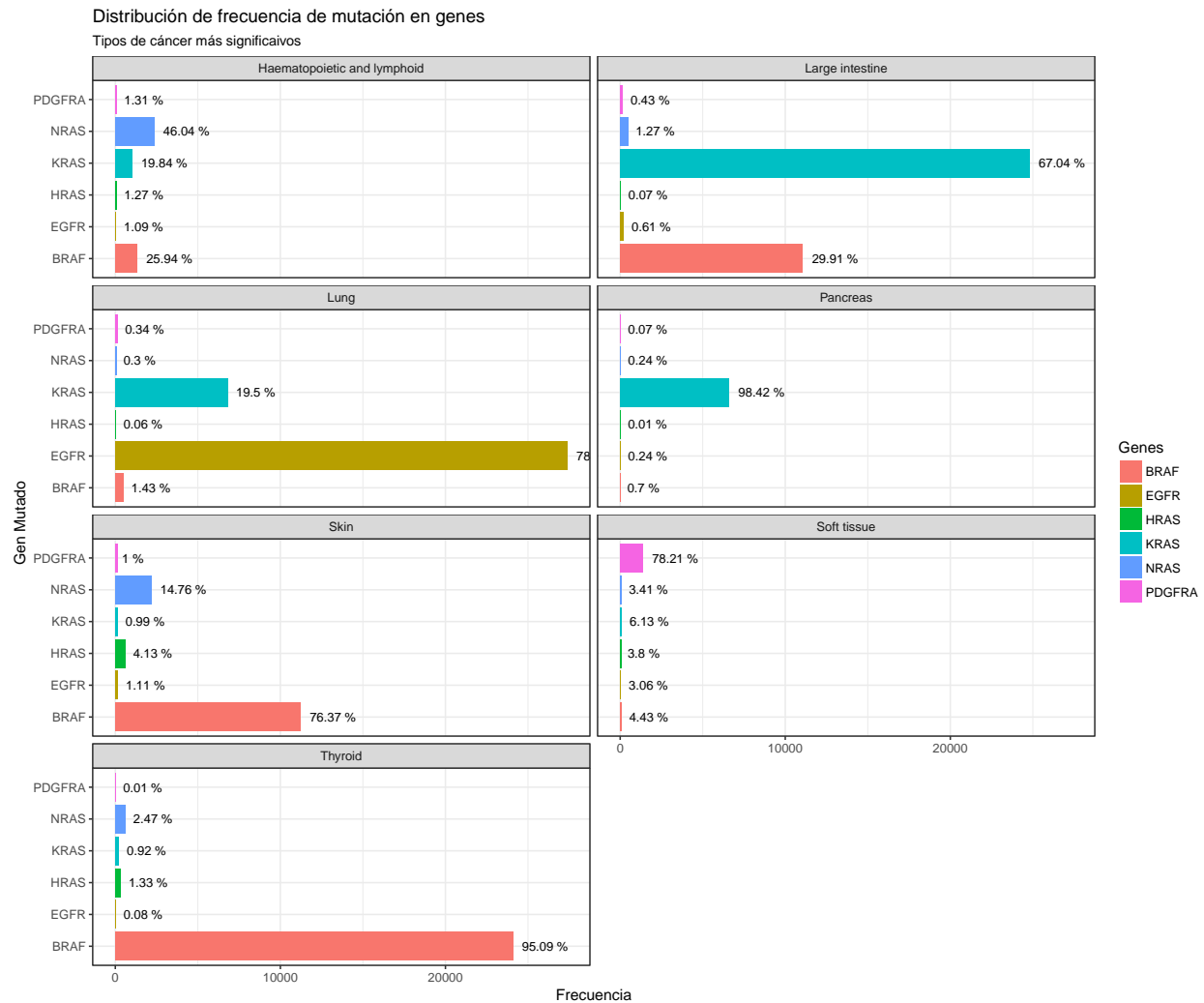
Tipo Cáncer	Gen Mutado	Frecuencia	% Gen	% Cancer
Haematopoietic and lymphoid	BRAF	1327	2.54	25.94
Large intestine	BRAF	11066	21.18	29.91
Lung	BRAF	502	0.96	1.43
Pancreas	BRAF	47	0.09	0.70
Skin	BRAF	11228	21.49	76.37
Soft tissue	BRAF	78	0.15	4.43
Thyroid	BRAF	24082	46.09	95.09
Haematopoietic and lymphoid	EGFR	56	0.19	1.09
Large intestine	EGFR	224	0.77	0.61
Lung	EGFR	27384	94.63	78.04
Pancreas	EGFR	16	0.06	0.24
Skin	EGFR	163	0.56	1.11
Soft tissue	EGFR	54	0.19	3.06
Thyroid	EGFR	20	0.07	0.08
Haematopoietic and lymphoid	HRAS	65	3.36	1.27
Large intestine	HRAS	25	1.29	0.07
Lung	HRAS	22	1.14	0.06
Pancreas	HRAS	1	0.05	0.01
Skin	HRAS	607	31.40	4.13
Soft tissue	HRAS	67	3.47	3.80
Thyroid	HRAS	336	17.38	1.33
Haematopoietic and lymphoid	KRAS	1015	2.33	19.84
Large intestine	KRAS	24804	56.84	67.04
Lung	KRAS	6843	15.68	19.50
Pancreas	KRAS	6604	15.13	98.42
Skin	KRAS	145	0.33	0.99
Soft tissue	KRAS	108	0.25	6.13
Thyroid	KRAS	232	0.53	0.92
Haematopoietic and lymphoid	NRAS	2355	35.64	46.04
Large intestine	NRAS	469	7.10	1.27
Lung	NRAS	106	1.60	0.30
Pancreas	NRAS	16	0.24	0.24
Skin	NRAS	2170	32.84	14.76
Soft tissue	NRAS	60	0.91	3.41
Thyroid	NRAS	625	9.46	2.47
Haematopoietic and lymphoid	PDGFRA	67	2.94	1.31
Large intestine	PDGFRA	160	7.02	0.43
Lung	PDGFRA	118	5.18	0.34
Pancreas	PDGFRA	5	0.22	0.07
Skin	PDGFRA	147	6.45	1.00
Soft tissue	PDGFRA	1378	60.47	78.21
Thyroid	PDGFRA	3	0.13	0.01



*Gráfica Gr.11*

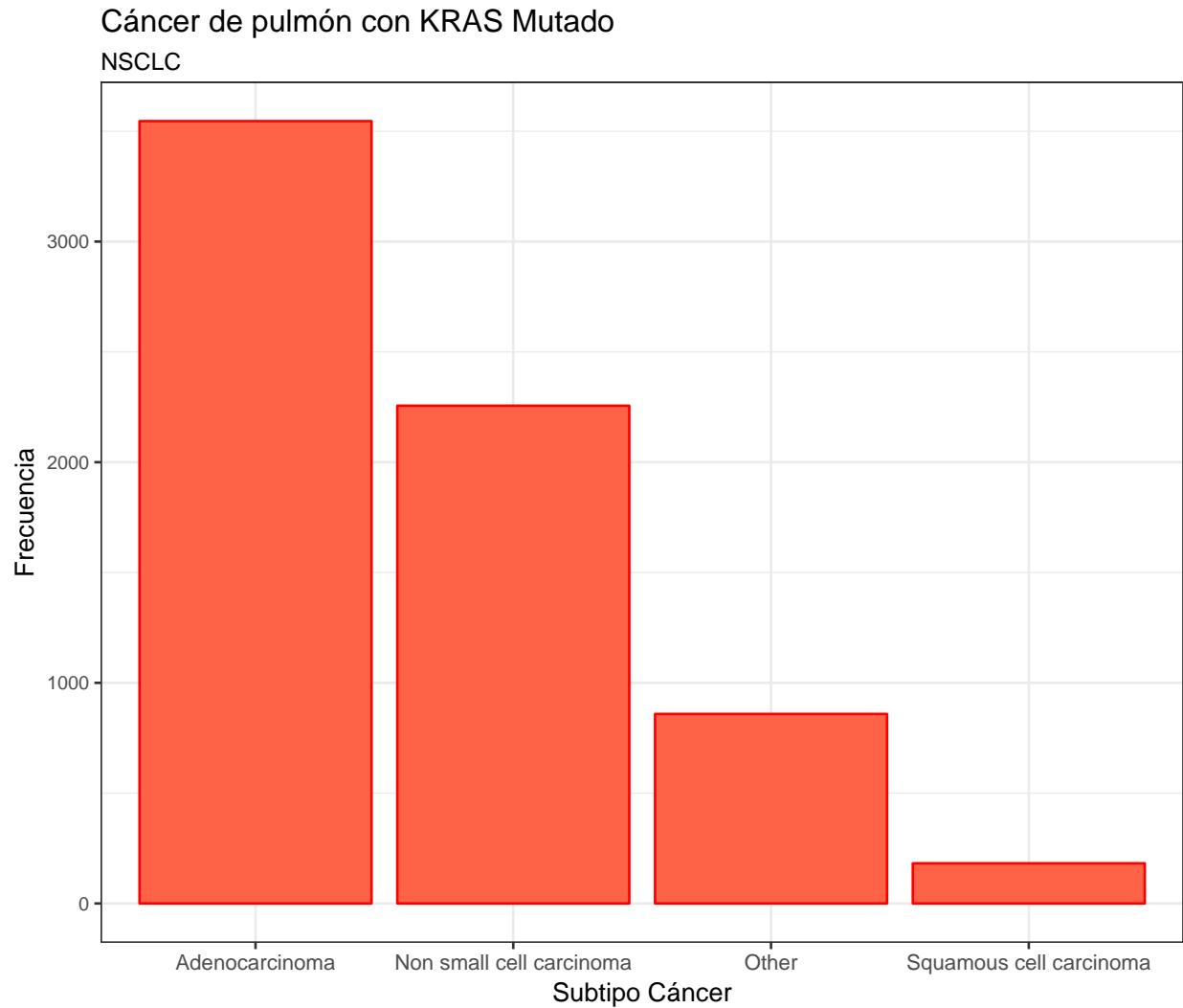
En la gráfica Gr12 se presenta la misma información que en la Gr11, pero cambiando el sentido de la observación entre los tipos de cáncer y los genes mutados. En esta visión se puede observar de forma rápida que en el cáncer de pulmón, los genes que más frecuentemente se encuentran mutados son el EGFR y el KRAS se pueden observar los genes mutados por cada tipo de cáncer, junto a la frecuencia de cada uno de ellos. Se puede comprobar que KRAS aparece con una mayor frecuencia en el cáncer de Colón, de pulmón y de páncreas.





Gráfica Gr.12

Centrando la atención en cáncer de pulmón de células no pequeñas (NSCLC), en la gráfica Gr.13 se pueden ver las frecuencias de los subtipos más relevantes.



Gráfica Gr.13

## Anexo de código R realizado para la generación de las tablas y gráficas

```
folder.data <- "./Datos"
folder.graf <- "./Datos/Graf"
file_data <- "1-2 EstadisMundial-EuropaxTipoCancer.csv"
file_data2 <- "3 EstadisMundialPulmonxArea.csv"
file_data3 <- "GenesMutadosMAPK-ERK_DATA.csv"

packages <- c("knitr","kableExtra","ggplot2","car", "caret",
              "plyr","dplyr","reshape2","magick")

check.packages <- is.element(packages, installed.packages()[, 1])==FALSE
packages.to.install <- packages[check.packages]
if (length(packages.to.install!=0)) {
```

```
install.packages(packages.to.install,
                 repos = "https://cran.rediris.es/", dep = TRUE)
}

# Librerías globales informe
stopifnot(require(knitr))
stopifnot(require(kableExtra))
stopifnot(require(ggplot2))
# Evaluación de los modelos
stopifnot(require(caret))
stopifnot(require(car))
stopifnot(require(plyr))
stopifnot(require(dplyr))
stopifnot(require(reshape2))
stopifnot(require(magick))
```

## Estadística global de incidencia de Cáncer con genes mutados en vía MAPK/ERK.

### Obtención de los datos estadísticos

```
# Carga de los fcheros en data frames.
#
# Estadística global del cáncer (Mundial y Europa)
dataEst1 <- read.csv2(file.path(folder.data,file_data),
                     stringsAsFactors = TRUE, sep = ";", header = TRUE, dec=".")
# Estadística del cáncer de pulmón (Mundial y Europa)
dataEst2 <- read.csv2(file.path(folder.data,file_data2),
                     stringsAsFactors = TRUE, sep = ";", header = TRUE, dec=".")
# Datos de muestras de cáncer con mutación de genes MAPK-ERK
dataEst3 <- read.csv2(file.path(folder.data,file_data3),
                     stringsAsFactors = TRUE, sep = ";", header = TRUE, dec=",")
#
```

## Estadística global de incidencia de Cáncer en el mundo y en Europa.

```
# Tabla 1 Datos globales de cancer en el mundo por sex
#
aux <- subset(dataEst1,Acum=="Total" & Sexo!="Ambos")[,c(1,3,4,5,7,8,10,11)]
aux[,c(4,6,8)] <- round(aux[,c(4,6,8)]/1000,0)
kable(aux, row.names=FALSE, format = "latex", booktabs = T,
      col.names= c("Área","Sexo", "Tipos", "Casos", "ASR","Casos","ASR", "5Años"),
      align= "cclrrrrr", caption = "Distribución de Cáncer en el mundo por Sexo") %>%
add_header_above(c(" ", " ", " ", " ", "Incidencia" = 2, "Mortalidad" = 2, "Prevalencia")) %>%
kable_styling(latex_options = c("striped", "scale_down","hold_position"))
# kable_as_image(paste(folder.graf,sep="", "Tabla1"))
```

## Preparación de los datos para las gráficas

```
# Mundial total
glMT <- subset(dataEst1, Estadistica == "Mundial" & Acum == "Total" & Sexo != "Ambos")
# Transforación de columnas a filas para la representación gráfica
glMT <- data.frame(TipoD=substr(colnames(glMT[c(5,8,11)]),1,
                             nchar(colnames(glMT[c(5,8,11)]))-1),
                  Sexo=rep(glMT$Sexo,each=3), Valor=matrix(t(glMT[,c(5,8,11)])))
# Cálculo porcentual por indicador
glMT <- glMT %>% # agrupamos por indicador
  group_by(TipoD) %>% # mutate para crear variable
  mutate(SexoProp = Valor/sum(Valor))
#
# Mundial por tipo de cáncer
glMc <- subset(dataEst1, Estadistica == "Mundial" & Acum == "TipoCancer" & Sexo == "Ambos")
#
# Europa total
glET <- subset(dataEst1, Estadistica == "Europa" & Acum == "Total" & Sexo != "Ambos")
# Transforación de columnas a filas para la representación gráfica
glET <- data.frame(TipoD=substr(colnames(glET[c(5,8,11)]),1,
                             nchar(colnames(glET[c(5,8,11)]))-1),
                  Sexo=rep(glET$Sexo,each=3), Valor=matrix(t(glET[,c(5,8,11)])))
# Cálculo porcentual por indicador
glET <- glET %>% # agrupamos por indicador
  group_by(TipoD) %>% # mutate para crear variable
  mutate(SexoProp = Valor/sum(Valor))
#
# Europa por tipo de cáncer
glEc <- subset(dataEst1, Estadistica == "Europa" & Acum == "TipoCancer" & Sexo == "Ambos")
```

## Gráficos comparativos estadísticos

### Incidencia del cáncer en la población mundial por Sexo

Gráfica Gr.1

```
# Gráfico Global 1
# Incidencia, mortalidas y Prevalencia (5años) distribuido por sexo
# a nivel mundial
#
file <- "grGl1.png"
grGl1 <- ggplot(glMT, aes(x = Sexo, y=round(Valor/1000,0), fill=Sexo)) +
  geom_bar(stat="identity" ) + facet_wrap(~ TipoD, nrow = 1) +
  theme_bw() +
  geom_text(aes(x = Sexo, y=round(Valor/1000,0),
                label = paste(round(SexoProp*100,2),"%"),
                position = position_dodge(width = 0.9), size=3,
                vjust=2, hjust=0.5 ,col="black") +
  theme(axis.text.x = element_text(hjust=1, vjust = 0.5, colour = "white"))+
  ylab("Frecuencia en miles") +
  ggtitle("Cáncer en el Mundo en 2012 (IARC) ",
          subtitle = "Excluido cáncer de piel no melanoma")
```

```
grGl1
ggsave(filename = file, path = folder.graf, width = 6, height = 4.5)
```

## Incidencia del cáncer en el continente europeo separado por Sexo

Gráfica Gr.2

```
# Gráfico Global 2
# Incidencia, mortalidad y Prevalencia (5años) distribuido por sexo
# en el continente europeo
#
file <- "grGl2.png"
grGl2 <- ggplot(glET, aes(x = Sexo, y=round(Valor/1000,0), fill=Sexo)) +
  geom_bar(stat="identity" ) + facet_wrap(~ TipoD, nrow = 1) +
  theme_bw() +
  geom_text(aes(x = Sexo, y=round(Valor/1000,0),
    label = paste(round(SexoProp*100,2),"%"),
    position = position_dodge(width = 0.9),
    size=3, vjust=2, hjust=0.5 ,col="black") +
  theme(axis.text.x = element_text(hjust=1, vjust = 0.5, colour = "white"))+
  ylab("Frecuencia en miles") +
  ggtitle("Cáncer en Europa en 2012 (IARC) ",
    subtitle = "Excluido cáncer de piel no melanoma")
grGl2
ggsave(filename = file, path = folder.graf, width = 6, height = 4.5)
```

## Ranking de tipos de cáncer con mayor incidencia a nivel mundial

Gráfica Gr.3

```
# Gráfico Global 3
# Incidencia y mortalidad por tipos de cáncer según porcentajes
# a nivel mundial
#
# Preparación Datos a mostrar
# Selección de porcentajes de incidencia y mortalidad
file <- "grGl3.png"
glMc2 <- glMc[,c(4,6,9)]
# Ponemos la incidencia en signo contrario para la gráfica
glMc2$IncidenciaP <- glMc2$IncidenciaP*-1
colnames(glMc2) <- c("Cancer", "Incidencia", "Mortalidad")
glMc2 <- melt(glMc2, value.name= 'value',
  variable.name = 'Indicador', id.vars = 'Cancer')
# Preparación de etiquetas para mostrar los porcentajes superiores al 5%
glMc2$labelv <- apply(data.frame(glMc2$value), 1,
  function(x) {if (x > 5) {paste(abs(x),sep=" ", "%")} else {NA}})
glMc2$labelv2 <- apply(data.frame(glMc2$value), 1,
  function(x) {if (x < -5) {paste(abs(x),sep=" ", "%")} else {NA}})
# Gráfica
#
grGl3 <- ggplot(glMc2, aes(x = reorder(Cancer,value), y=value, fill=Indicador)) +
  geom_bar(data= subset(glMc2,Indicador == "Incidencia"),stat="identity" ) +
  geom_bar(data= subset(glMc2,Indicador == "Mortalidad"),stat="identity" ) +
  theme_bw() +
```

```

scale_fill_brewer(palette = "Dark2")+
coord_flip()+
scale_y_continuous(breaks = seq(-20,20,5),
                    labels = paste(as.character(c(20,15,10,5,0,5,10,15,20)), "%") )+
geom_text(aes(x = reorder(Cancer,value), y=value, label = labelv),
           position = position_dodge(width = 0.5), size=2.5,
           col="black", vjust=0.5, hjust=1 ) +
geom_text(aes(x = reorder(Cancer,value), y=value, label = labelv2),
           position = position_dodge(width = 0.5), size=2.5,
           vjust=0, hjust=0 ,col="black") +
theme(axis.text.x = element_text(hjust=0.3, vjust = 0.5, size=7))+
theme(axis.text.y = element_text(hjust=1, vjust = 0.5, size=7))+
ylab("% Incidencia / Mortalidad") +
xlab("Tipo de Cáncer") +
ggtitle("Principales cánceres en el mundo en 2012 (IARC) ",
        subtitle = "Excluido cáncer de piel no melanoma")
grG13
ggsave(filename = file, path = folder.graf, width = 6, height = 4.5)

```

## Ranking de tipos de cáncer con mayor incidencia en el continente europeo

### Gráfica Gr.4

```

# Gráfico Global 4
# Incidencia y mortalidad por tipos de cáncer según porcentajes
#
# Preparación Datos a mostrar
# Selección de porcentajes de incidencia y mortalidad
file <- "grG14.png"
glEc2 <- glEc[,c(4,6,9)]
# Ponemos la incidencia en signo contrario para la gráfica
glEc2$IncidenciaP <- glEc2$IncidenciaP*-1
# glEc2$MortalidadP <- glEc2$MortalidadP*-1
colnames(glEc2) <- c("Cancer", "Incidencia", "Mortalidad")
glEc2 <- melt(glEc2, value.name= 'value',
              variable.name = 'Indicador', id.vars = 'Cancer')
# Preparación de etiquetas para mostrar los porcentajes superiores al 5%
glEc2$labelv <- apply(data.frame(glEc2$value), 1,
                      function(x) {if (x > 5) {paste(abs(x),sep=" ", "%") } else {NA}})
glEc2$labelv2 <- apply(data.frame(glEc2$value), 1,
                       function(x) {if (x < -5) {paste(abs(x),sep=" ", "%") } else {NA}})
# Gráfica
grG14 <- ggplot(glEc2, aes(x = reorder(Cancer,value), y=value, fill=Indicador)) +
  geom_bar(data= subset(glEc2,Indicador == "Incidencia"),stat="identity" ) +
  geom_bar(data= subset(glEc2,Indicador == "Mortalidad"),stat="identity" ) +
  theme_bw() +
  scale_fill_brewer(palette = "Dark2")+
  coord_flip()+
  scale_y_continuous(breaks = seq(-15,25,5),
                    labels = paste(as.character(c(15,10,5,0,5,10,15,20,25)), "%") )+
  geom_text(aes(x = reorder(Cancer,-value), y=value, label = labelv),
            position = position_dodge(width = 0.5), size=2.5,
            col="black", vjust=0.5, hjust=1 ) +

```

```

geom_text(aes(x = reorder(Cancer,value), y=value, label = labelv2),
          position = position_dodge(width = 0.5), size=2.5,
          vjust=0, hjust=0 ,col="black") +
theme(axis.text.x = element_text(hjust=0.3, vjust = 0.5, size=7))+
theme(axis.text.y = element_text(hjust=1, vjust = 0.5, size=7))+
xlab("Tipo de Cáncer") +
ylab("% Incidencia / Mortalidad") +
ggtitle("Principales cánceres en Europa en 2012 (IARC) ",
        subtitle = "Excluido cáncer de piel no melanoma")

grG14
ggsave(filename = file, path = folder.graf, width = 6, height = 4.5)

```

## Estadística de incidencia de Cáncer de pulmón en el mundo.

### Preparación de los datos para las gráficas

```

# Cancer de pulmón según división mundial de WHO x Sexo
glPwho <- subset(dataEst2, Tipo == "WHO" & Sexo != "Ambos")
# Cancer de pulmón según división mundial de WHO ambos sexos
glPwho2 <- subset(dataEst2, Tipo == "WHO" & Sexo == "Ambos")
# Cancer de pulmón según Nivel de desarrollo x Sexo
glPniv <- subset(dataEst2, Tipo == "Nivel"& Sexo != "Ambos")
# Cancer de pulmón según Nivel de desarrollo ambos sexos
glPniv2 <- subset(dataEst2, Tipo == "Nivel"& Sexo == "Ambos")

```

### Gráficos comparativos estadísticos

#### Incidencia del cáncer de pulmón en la población mundial por Sexo

Gráfica Gr.5

```

# Gráfico Global 5
# Casos de cáncer de pulmón por área de WHO en el mundo x Sexo
#
file <- "grG15.png"
grG15 <- ggplot(glPwho, aes(x=Area,y=IncidenciaN)) +
  geom_bar(stat="identity", aes(fill = Sexo),
          position = position_stack(reverse = TRUE))+
  coord_flip()+
  theme(legend.position = "top")+
  geom_text(aes(label = IncidenciaN), position = position_stack(vjust = 0.5),
          size=3, check_overlap = TRUE)+
  theme_bw() +
  theme(axis.text.y = element_text(hjust=1, vjust = 0.5, size = 8))+
  xlab("Área WHO") +
  ylab("Casos cáncer de pulmon") +
  ggtitle("Cáncer de pulmón \n en el Mundo en 2012 (IARC) ",
          subtitle = "Valores estimados en miles (x1000)")

grG15
ggsave(filename = file, path = folder.graf, width = 6, height = 4.5)

```

## Incidencia de la mortalidad del cáncer de pulmón en el mundo separado por Sexo

Gráfica Gr.6

```
# Gráfico Global 6
# Muertes por cáncer de pulmón por área de WHO en el mundo x Sexo
#
file <- "grGl6.png"
grGl6 <- ggplot(glpwho, aes(x=Area,y=MortalidadN)) +
  geom_bar(stat="identity", aes(fill = Sexo),position = position_stack(reverse = TRUE))+
  coord_flip()+
  theme(legend.position = "top")+
  geom_text(aes(label = MortalidadN), position = position_stack(vjust = 0.5),
            size=3, check_overlap = TRUE)+
  theme_bw() +
  theme(axis.text.y = element_text(hjust=1, vjust = 0.5, size = 8))+
  xlab("Área WHO") +
  ylab("Mortalidad cáncer de pulmon") +
  ggtitle("Mortalidad por cáncer de pulmón \n en el Mundo en 2012 (IARC) ",
          subtitle = "Valores estimados en miles (x1000)")
grGl6
ggsave(filename = file, path = folder.graf, width = 6, height = 4.5)
```

## Comparativa de nuevos casos y de mortalidad del cáncer de pulmón en el mundo

Gráfica Gr.7

```
# Gráfico Global 7
# Incidencia y mortalidad según región WHO
#
# Preparación Datos a mostrar
# Selección de incidencia y mortalidad
file <- "grGl7.png"
glpwho3 <- glpwho2[,3:5]
# Ponemos la incidencia en signo contrario para la gráfica
glpwho3$IncidenciaN <- glpwho2$IncidenciaN*-1
colnames(glpwho3) <- c("Area", "Incidencia", "Mortalidad")
glpwho3 <- melt(glpwho3, value.name= 'value',
               variable.name = 'Indicador', id.vars = 'Area')
# Preparación de etiquetas para mostrar los porcentajes superiores al 5%
glpwho3$labelv <- apply(data.frame(glpwho3$value), 1,
                       function(x) {if (x > 0) {x} else {NA}})
glpwho3$labelv2 <- apply(data.frame(glpwho3$value), 1,
                        function(x) {if (x < 0) {abs(x)} else {NA}})
# Gráfica
grGl7 <- ggplot(glpwho3, aes(x = Area, y=value, fill=Indicador)) +
  geom_bar(data= subset(glpwho3,Indicador == "Incidencia"),stat="identity" ) +
  geom_bar(data= subset(glpwho3,Indicador == "Mortalidad"),stat="identity" ) +
  theme_bw() +
  scale_fill_brewer(palette = "Dark2")+
  coord_flip()+
  scale_y_continuous(breaks = seq(-900,900,300),
                     labels = as.character(c(900,600,300,0,300,600,900)) )+
  geom_text(aes(x = Area, y=value, label = labelv),
            position = position_dodge(width = 0.5), size=2.5,
```



```

        col="black", vjust=0.5, hjust=1 ) +
geom_text(aes(x = Area, y=value, label = labelv2),
          position = position_dodge(width = 0.5), size=2.5,
          vjust=0, hjust=0 ,col="black") +
theme(axis.text.x = element_text(hjust=0.3, vjust = 0.5))+
xlab("Área WHO") +
ylab("Incidencia / Mortalidad") +
ggtitle("Cáncer de pulmón \n en el Mundo en 2012 (IARC) ",
        subtitle = "Valores estimados en miles (x1000)")

grGl7
ggsave(filename = file, path = folder.graf, width = 6, height = 4.5)

```

## Comparativa de nuevos casos y de mortalidad del cáncer de pulmón en el mundo

Gráfica Gr.8

```

# Gráfico Global 8
# Incidencia y mortalidad según nivel de desarrollo
#
# Preparación Datos a mostrar
# Selección de incidencia y mortalidad
file <- "grGl8.png"
glPniv3 <- glPniv2[,3:5]
# Ponemos la incidencia en signo contrario para la gráfica
glPniv3$IncidenciaN <- glPniv2$IncidenciaN*-1
colnames(glPniv3) <- c("Nivel", "Incidencia", "Mortalidad")
glPniv3 <- melt(glPniv3, value.name= 'value',
               variable.name = 'Indicador', id.vars = 'Nivel')
# Preparación de etiquetas para mostrar los porcentajes superiores al 5%
glPniv3$labelv <- apply(data.frame(glPniv3$value), 1,
                       function(x) {if (x > 0) {x} else {NA}})
glPniv3$labelv2 <- apply(data.frame(glPniv3$value), 1,
                        function(x) {if (x < 0) {abs(x)} else {NA}})

# Gráfica
grGl8 <- ggplot(glPniv3, aes(x = Nivel, y=value, fill=Indicador)) +
  geom_bar(data= subset(glPniv3,Indicador == "Incidencia"),stat="identity" ) +
  geom_bar(data= subset(glPniv3,Indicador == "Mortalidad"),stat="identity" ) +
  theme_bw() +
  scale_fill_brewer(palette = "Dark2")+
  # coord_flip()+
  scale_y_continuous(breaks = seq(-900,900,300),
                    labels = as.character(c(900,600,300,0,300,600,900)) )+
  geom_text(aes(x = Nivel, y=value, label = labelv),
            position = position_dodge(width = 0.5),
            col="black", vjust=1.5, hjust=1 ) +
  geom_text(aes(x = Nivel, y=value, label = labelv2),
            position = position_dodge(width = 0.5),
            vjust=-1, hjust=-0.5 ,col="black") +
  theme(axis.text.x = element_text(hjust=0.3, vjust = 0.5))+
  xlab("Nivel de desarrollo pais") +
  ylab("Incidencia / Mortalidad") +
  ggtitle("Cáncer de pulmón (nivel desarrollo) \n en el Mundo en 2012 (IARC) ",

```

```

        subtitle = "Valores estimados en miles (x1000)")

grG18
ggsave(filename = file, path = folder.graf, width = 6, height = 4.5)

```

Estadística global de incidencia de Cáncer con genes mutados en vía MAPK/ERK.

```

#
# Data frame con datos de muestras por cáncer (Primary.Tissue)
tbe1.1 <- as.data.frame(table(dataEst3$Primary.Tissue))
colnames(tbe1.1) <- c("Primary.Tissue", "Freq")
# Proporciones
tbe1.1p <- as.data.frame(round(prop.table(table(dataEst3$Primary.Tissue))*100,2))
colnames(tbe1.1p) <- c("Primary.Tissue", "Freq")
#
# Data frame con datos de muestras por Gen mutado (Gene.Name)
tbe1.2 <- as.data.frame(table(dataEst3$Gene.Name))
colnames(tbe1.2) <- c("Gene.Name", "Freq")
# Proporciones
tbe1.2p <- as.data.frame(round(prop.table(table(dataEst3$Gene.Name))*100,2))
colnames(tbe1.2p) <- c("Gene.Name", "Freq")
#t1.1 <- t1.1[order(t1.1$Freq),]
#
# Data frame con datos de muestras por cáncer (Primary.Tissue) y Gen mutado (Gene.Name)
# (solo los significativos) %Frecuencia > 1
#
tbe1 <- as.data.frame(table(dataEst3$Primary.Tissue,dataEst3$Gene.Name))
colnames(tbe1) <- c("Primary.Tissue", "Gene.Name", "Freq")
#Selección de Cánceres y Genes más significativos
CanSel <- as.character(tbe1.1$Primary.Tissue[which(tbe1.1p$Freq>1&tbe1.1p$Primary.Tissue!="NS")])
CanSel2 <- as.numeric(apply(as.data.frame(CanSel),1,
                           function(par1) {which(tbe1$Primary.Tissue==par1)}))
tbe2 <- tbe1[CanSel2,]
GenSel <- as.character(tbe1.2$Gene.Name[which(tbe1.2p$Freq>1)])
GenSel2 <- as.numeric(apply(as.data.frame(GenSel),1,
                           function(par1) {which(tbe2$Gene.Name==par1)}))
tbe2 <- tbe2[GenSel2,]
PorcG <- as.numeric(apply(as.data.frame(tbe2[,2:3]),1,
                         function(x){round(as.numeric(x[2])/tbe1.2[tbe1.2==x[1],2]*100,2)}))
tbe2 <- data.frame(tbe2, PorcG)
PorcC <- as.numeric(apply(as.data.frame(tbe2[,c(1,3)]),1,
                         function(x){round(as.numeric(x[2])/tbe1.1[tbe1.1==x[1],2]*100,2)}))
tbe2 <- data.frame(tbe2, PorcC)
# Selección de KRAS mutado y cancer de pulmón
#
tbe3 <- subset(dataEst3, Gene.Name=="KRAS"&Primary.Tissue=="Lung")
# Reclasificación sobre los 3 tipos de cáncer NSCLC más relevantes y resto como otros.
tbe3$subtipo <- apply(data.frame(subset(dataEst3,
                                       Gene.Name=="KRAS"&Primary.Tissue=="Lung")[,7]), 1,
                     function(x) {if (x != 'Adenocarcinoma' & x!= 'Squamous cell carcinoma'
                                       & x != 'Non small cell carcinoma') {"Other"} else {x}})

```

## Preparación y exploración de los datos

```
#
# Data frame con datos de muestras por cáncer (Primary.Tissue)
tbe1.1 <- as.data.frame(table(dataEst3$Primary.Tissue))
colnames(tbe1.1) <- c("Primary.Tissue", "Freq")
# Proporciones
tbe1.1p <- as.data.frame(round(prop.table(table(dataEst3$Primary.Tissue))*100,2))
colnames(tbe1.1p) <- c("Primary.Tissue", "Freq")
#
# Data frame con datos de muestras por Gen mutado (Gene.Name)
tbe1.2 <- as.data.frame(table(dataEst3$Gene.Name))
colnames(tbe1.2) <- c("Gene.Name", "Freq")
# Proporciones
tbe1.2p <- as.data.frame(round(prop.table(table(dataEst3$Gene.Name))*100,2))
colnames(tbe1.2p) <- c("Gene.Name", "Freq")
#t1.1 <- t1.1[order(t1.1$Freq),]
#
# Data frame con datos de muestras por cáncer (Primary.Tissue) y Gen mutado (Gene.Name)
# (solo los significativos)
#
tbe1 <- as.data.frame(table(dataEst3$Primary.Tissue, dataEst3$Gene.Name))
colnames(tbe1) <- c("Primary.Tissue", "Gene.Name", "Freq")
#Selección de Cánceres y Genes más significativos
CanSel <- as.character(tbe1.1$Primary.Tissue[
  which(tbe1.1p$Freq>1&tbe1.1p$Primary.Tissue!="NS")])
CanSel2 <- as.numeric(apply(as.data.frame(CanSel), 1,
  function(par1) {which(tbe1$Primary.Tissue==par1)}))
tbe2 <- tbe1[CanSel2,]
GenSel <- as.character(tbe1.2$Gene.Name[which(tbe1.2p$Freq>1)])
GenSel2 <- as.numeric(apply(as.data.frame(GenSel), 1,
  function(par1) {which(tbe2$Gene.Name==par1)}))
tbe2 <- tbe2[GenSel2,]
PorcG <- as.numeric(apply(as.data.frame(tbe2[,2:3]), 1,
  function(x){round(as.numeric(x[2])/tbe1.2[tbe1.2==x[1],2]*100,2)}))
tbe2 <- data.frame(tbe2, PorcG)
PorcC <- as.numeric(apply(as.data.frame(tbe2[,c(1,3)]), 1,
  function(x){round(as.numeric(x[2])/tbe1.1[tbe1.1==x[1],2]*100,2)}))
tbe2 <- data.frame(tbe2, PorcC)
# Selección de KRAS mutado y cancer de pulmón
#
tbe3 <- subset(dataEst3, Gene.Name=="KRAS"&Primary.Tissue=="Lung")
# Reclasificación sobre los 3 tipos de cáncer NSCLC más relevantes y resto como otros.
tbe3$subtipo <- apply(data.frame(subset(dataEst3,
  Gene.Name=="KRAS"&Primary.Tissue=="Lung")[,7]), 1,
  function(x) {if (x != 'Adenocarcinoma' & x != 'Squamous cell carcinoma'
    & x != 'Non small cell carcinoma') {"Other"} else {x}}})

# Tabla 2 Frecuencia en porcentajes de los tipos de cáncer con mutación de genes en la vía
#
file = "tabla2"
kable(data.frame(tbe1.1, tbe1.1p$Freq), format = "latex", booktabs = T,
  col.names = c("Tipo Cáncer", "Frecuencia", "% Frecuencia"),
  align = "lrr", caption = "Distribución frecuencias de Tipos de cáncer") %>%
```

```

# kable_styling(latex_options = c("striped", "hold_position")) %>%
kable_styling(font_size = 7 )
# kable_as_image(filename = paste(folder.graf, "/", sep="",
#                                 file), file_format = "png", density = 500)

# Tabla 3 Frecuencia en porcentajes de los genes mutados en la vía
#
file = "tabla3"
kable(data.frame(tbe1.2, tbe1.2p$Freq), format = "latex", booktabs = T,
      col.names = c("Gen mutado", "Frecuencia", "% Frecuencia"),
      align = "lrr", caption = "Distribución frecuencias de Genes mutados") %>%
# kable_styling(latex_options = c("striped", "hold_position")) %>%
kable_styling(font_size = 7 )
# kable_as_image(filename = paste(folder.graf, "/", sep="", file),
#                 file_format = "png", density = 500)

```

## Graficos comparativos estadísticos

### Distribución de tipos de cáncer con mutaciones en genes de la vía MAPK/ERK

Gráfica Gr.9

```

# Gráfico Global 9
# Ranking de tipos de cancer con genes mutados en la vía MAPK/ERK
#
file <- "grtb9.png"
#
gr9 <- ggplot(tbe1.1p, aes(x=reorder(Primary.Tissue, Freq), y=Freq)) +
  geom_bar(stat="identity", color="red", fill="tomato") +
  theme_bw() +
  coord_flip() +
  theme(axis.text.x = element_text(size = 9, hjust=0.5, vjust = 0)) +
  theme(axis.text.y = element_text(size = 8, hjust=1, vjust = 0.5)) +
  xlab("Tipo Cáncer") +
  ylab("% Frecuencia") +
  scale_fill_manual(values = c(2,3)) +
  ggtitle("Distribución de Cáncer ", subtitle = "Vía MAPK-ERK") +
  geom_text(aes(y = Freq, label = paste(round(Freq,1), "%")),
            position = position_dodge(width = 0.9), size=3,
            vjust=0.5, hjust=-0.2 , col="black")

gr9
ggsave(filename = file, path = folder.graf)

```

### Distribución de genes con mutaciones en la vía MAPK/ERK

Gráfica Gr.10

```

# Gráfico Global 10
# Ranking de genes mutados en la vía MAPK/ERK según frecuencia de mutación
#
file <- "grtb10.png"
#
gr10 <- ggplot(tbe1.2p, aes(x=reorder(Gene.Name, Freq), y=Freq)) +
  geom_bar(stat="identity", color="green", fill="lightgreen") +

```

```

theme_bw() +
coord_flip() +
theme(axis.text.x = element_text(size = 8, hjust=0.5, vjust = 0.5),
legend.text = element_text(size = 6)) +
xlab("Gen Mutado") +
ylab("% Frecuencia") +
scale_fill_manual(values = c(2,3)) +
ggtitle("Distribución de Genes Mutados ",
        subtitle = "Vía MAPK-ERK") +
geom_text(aes(y = Freq,label = paste(round(Freq,1),"%")),
          position = position_dodge(width = 0.9),
          size=3, vjust=0.5, hjust=-0.2 ,col="black")
gr10
ggsave(filename = file, path = folder.graf)

```

Cuadros de distribución de los tipos de cancer más frecuente con frecuencia de mutación de los genes

```

# Tabla 4 Frecuencia de los tipos de cáncer y genes mutados más significativos
#
file = "tabla4"
kable(tbe2[,1:5],row.names=FALSE,format = "latex", booktabs = T,
      col.names= c("Tipo Cáncer", "Gen Mutado", "Frecuencia", "% Gen", "% Cancer"),
      align= "lrrcc", caption = "Distribución frecuencias
de Tipos de cáncer y Genes mutados") %>%
# kable_styling(latex_options = c("striped","hold_position")) %>%
kable_styling(font_size = 7 )
# kable_as_image(filename = paste(folder.graf,"/",sep="",file),
#               file_format = "png",density = 500)

```

Gráfica Gr.11

```

# Gráfico Global 11
# Cuadro de distribución de frecuencia de cácer y frecuencia de mutación de genes en cada tipo
#
file <- "grtb11.png"
#
gr11 <- ggplot(tbe2, aes(x=Primary.Tissue,y=Freq,
                        colours=seq(1,length(Primary.Tissue)),
                        fill=Primary.Tissue)) +
  geom_bar(stat="identity") +
  coord_flip() +
  theme_bw() +
  facet_wrap(~ Gene.Name, nrow = 3) +
  guides(fill=guide_legend(title="Tipos cáncer")) +
  xlab("Tipo Cancer") +
  ylab("Frecuencia") +
  ggtitle("Distribución de incidencia Cáncer y genes implicados ",
          subtitle = "Cánceres más significativos") +
  geom_text(aes(y = Freq, label = paste(PorcG,"%")),
            position = position_dodge(width = 0.9),
            size=3, vjust=0.5, hjust=-0.2 ,col="black")
gr11
ggsave(filename = file, path = folder.graf)

```

### Gráfica Gr.12

```
# Gráfico Global 12
# Cuadro de distribución de frecuencia de cáncer y frecuencia de mutación de genes en cada tipo
#
file <- "grG112.png"
#
gr12 <- ggplot(tbe2, aes(x=Gene.Name,y=Freq,
                        colours=seq(1,length(Gene.Name)), fill=Gene.Name)) +
  geom_bar(stat="identity") +
  coord_flip() +
  theme_bw() +
  facet_wrap(~ Primary.Tissue, nrow = 4) +
  guides(fill=guide_legend(title="Genes")) +
  xlab("Gen Mutado") +
  ylab("Frecuencia") +
  ggtitle("Distribución de frecuencia de mutación en genes ",
          subtitle = "Tipos de cáncer más significativos") +
  geom_text(aes(y = Freq, label = paste(PorcC,"%")),
            position = position_dodge(width = 0.9), size=3,
            vjust=0.5, hjust=-0.2 ,col="black")
gr12
ggsave(filename = file, path = folder.graf)
```

### Gráfica Gr.13

```
# Gráfico Global 13
# Ranking de sutipos de NSCLC con KRAS mutado
#
file <- "grtb13.png"
#
gr13 <- ggplot(tbe3, aes(x=subtipo)) +
  # geom_bar() +
  geom_bar(color="red", fill="tomato") +
  theme_bw() +
  # coord_flip() +
  theme(axis.text.x = element_text(size = 9, hjust=0.5, vjust = 0)) +
  theme(axis.text.y = element_text(size = 8, hjust=1, vjust = 0.5)) +
  xlab("Subtipo Cáncer") +
  ylab("Frecuencia") +
  scale_fill_manual(values = c(2,3)) +
  ggtitle("Cáncer de pulmón con KRAS Mutado ", subtitle = "NSCLC")
  # geom_text(aes(y = Freq, label = paste(round(Freq,1,"%")),
  #           position = position_dodge(width = 0.9), size=3,
  #           vjust=0.5, hjust=-0.2 ,col="black")
gr13
ggsave(filename = file, path = folder.graf)
```

## Referencias

Dragulescu, Adrian A. 2014. *Advanced Graphics and Image-Processing in R*. <https://cran.r-project.org/package=xlsx>.

Fox, John, and Sanford Weisberg. 2011. *An R Companion to Applied Regression*. Second. Thousand Oaks

- CA: Sage. <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>.
- IARC, International Agency for Research on Cancer - WHO. 2018. “GLOBOCAN Project.” <http://globocan.iarc.fr/Default.aspx>.
- Kuhn, Max. 2018. *Misc Functions for Training and Plotting Classification and Regression Models*. <https://cran.r-project.org/package=caretl>.
- v84, COSMIC. 2018. “Cataloge of Somatic Mutations in Cancer.” <https://cancer.sanger.ac.uk/cosmic>.
- Wickham, Hadley. 2007. “Reshaping Data with the reshape Package.” *Journal of Statistical Software* 21 (12): 1–20. <http://www.jstatsoft.org/v21/i12/>.
- . 2009. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <http://ggplot2.org>.
- . 2011. “The Split-Apply-Combine Strategy for Data Analysis.” *Journal of Statistical Software* 40 (1): 1–29. <http://www.jstatsoft.org/v40/i01/>.
- Xie, Yihui. 2018. *Knitr: A General-Purpose Package for Dynamic Report Generation in R*. <https://yihui.name/knitr/>.
- Zhu, Hao. 2018. *Construct Complex Table with 'Kable' and Pipe Syntax*. <https://cran.r-project.org/package=kableExtra>.