

# Case study 2

Оценка мутационного профиля пациентов с  
раком молочной железы

# Дано

- Данные о наличии мутаций в генах BRCA1, BRCA2, TP53, PTEN – лист Mutations
- Клинические данные пациентов (рак молочной железы) – лист Clinical data

# Цель анализа

- Найти наличие мутационной со-встречаемости данных генов между собой
- Рассчитать оценку зависимости продолжительности жизни от наличия мутаций в данных генах

# 1. Импорт данных и объединение

```
library(readxl)
```

```
library(dplyr)
```

Функции импорта и объединения таблиц

## 2. Обработка

- Получить столбцы в бинарном виде – «1»/«0»
  - Функция `ifelse()`
  - Функция `mutate_at()`
- «Почистить» названия столбцов
  - `library(janitor)`
  - Функция `clean_names()`

## 2. Обработка

- Функция `ifelse()`

`ifelse(условие, вариант_TRUE, вариант_FALSE)`

`ifelse(vector > 0, "> 0", "<= 0")`

- Функция `mutate_at()`

`mutate_at(c(вектор столбцов), ~функция_для_столбцов)`

`dataframe %>%`

`mutate_at(c(col1, col2, col3), ~ifelse(. > 5, 0, 1))`

## 2. Обработка

- `library(janitor)`
- Функция `clean_names()`

«Чистит» название столбцов:

- Убирает заглавные буквы
- Пробелы заменяет на нижние подчеркивания

# 3. Со-встречаемость мутаций

## 1. Найти наличие мутационной со-встречаемости данных генов между собой (с помощью теста Фишера по парам генов)

- library(rstatix)
  - combn()
  - table()
  - fisher\_test()
- library(stringr)
  - str\_c()
  - str\_to\_upper()
- library(purrr)
  - map()
  - bind\_rows()



### 3. Со-встречаемость мутаций

- `combn()`
  - `combn(vector, 2)`
- `table`
  - Таблица сопряженности
- `fisher_test()`
  - Таблица теста Фишера с уровнем значимости

### 3. Со-встречаемость мутаций

- `str_c()`
  - Объединение вектора
- `str_to_upper()`
  - Верхний регистр

### 3. Со-встречаемость мутаций

- `map()`
  - Применение функции на столбцах таблицы
- `bind_rows()`
  - Объединение в таблицу

## 4. Выживаемость

### 2. Рассчитать оценку зависимости продолжительности жизни от наличия мутаций в данных генах (с помощью оценки Каплана-Мейера)

- `library(survival)`
- `library(ggfortify)`
- `library(survminer)`

`km_surv_fit()` – расчёт кривой выживаемости

`surv_pvalue()`\$pval – получение уровня значимости

`ggsurvplot()` – график кривой выживаемости

## 4. Выживаемость

Оценка Каплана-Мейера

<https://www.machinelearningmastery.ru/kaplan-meier-curves-c5768e349479/>

На вход: группа пациентов, время, статус (0/1).

## 4. Выживаемость

TP53 - breast cancer (Overall survival)

