# TCGA RNA-seq data analysis in breast invasive carcinoma

Gilabert-Navarro, JF. Madsen-Choppi, LPN. Garcia-Serrano, A.

May 5, 2016

**Abstract**

In this work we bla bla..

# Introduction

Breast cancer is the most common malignant cancer affecting women and is the second leading cause of cancer death worldwide[2]. This disease has more than 1,300,000 cases and 450,000 death each year around the world[3]. This disease is widely heterogeneous, having a large and diverse set of molecular, histological and clinical behaviours depending of the tumour[2]. In addition, the response to specific treatments it is also very different between patients. For this reason, breast cancer was been classified in different subtypes in order to achieve a better understanding of these disease. Traditionally, the classification has been based on clinicopathological features such as tumor type and size, lymph node status and histological grade[2]. Actually, nowadays this disease is an entity difficult to classify due to the wide range of classifiers that we can take into account: histological, immunopathological, transcriptional, genomic, miRNA-based, epigenetic, microenvironmental, macroenvironmental, longitudinal and other classifiers[4]. However, we have an actual classification based on simple molecular characteristics[3]

- **Estrogen receptor (ER) positive:** The most numerous and diverse, with several genomic tests to assist in predicting outcomes for ER+ patients receiving endocrine therapy.

- **HER2 or ERBB2 amplified:** Great clinical success because of effective therapeutic targeting of HER2, which has led to intense efforts to characterize other DNA copy number aberrations.

- **Triple negative:** Lacking expression of ER, progesterone receptor (PR) and HER2. It is also known as basal-like breast cancers, are a group with only chemotherapy options, and have an increased incidence in patients with germline BRCA1 mutations or of African ancestry.

The more frequently mutated genes in breast cancer are BRCA1, BRCA2, PALB2, ATM, TP53, PTEN, PIK3CA, AKT1, GATA3, CDH1, RB1, MLL3, MAP3K1, CDKN1B, between others[1]. If we focus in their functionalities, these genes are linked with DNA repair, control of cell cycle, apoptosis, cell proliferation and gorwth. BRCA1 and BRCA2 mutations are the most common cause of hereditary breast cancer[1]. In addition, women with these mutations also have higher risk of developing other cancers, mainly ovarian cancer[1]. In case of BRCA1 mutations, the risk compared with the population is about 60% meanwhile in BRCA2 mutations is about 45%[1].

Anyway, as we describe above, each tumour has a high specific profile with a lot of different variables difficulting the establishment of simple classifiers. In the last decades, molecular knowledge advances have allowed to initialize personalized medicine. In this way, we can use targeted drugs to very specific tumour types with a high percentage of effectiveness. The main problem of this personalized medicine is the very reduced number of tumours in which we can observe a remission. This is due to the very high specificity of the treatments, useful only for a tumour with a concrete molecular characteristics.

In this way, new technologies focused not only in mRNA expression profiling, DNA copy number analysis and massively parallel sequencing but also in detecting abnormalities in DNA methylation, miRNA and protein expression provides a wider range of information[3]. Therefore, we can use all these tools in order to get a deeper understanding about tumor molecular mechanisms resulting in advances towards personalized medicine.

# Objective

Our main goal is to compare breast samples with and without cancer in order to perform a differential mRNA expression analysis. This data could help us, in the future, to provide a better understanding that will lead to better management of patients.

# Materials and methods

- Define our subset

- Batch effect identification

- Quality assessment and normalization

- Differential expression analysis (F-Test, package SVA)

# Results

In this third section we will show how to make a table:

| Gene | Expression |
|------|------------|
| FOXP2 | overexpressed |
| HOXA | underexpressed |
| BRCA1 | overexpressed |
| p53 | overexpressed |

# Discussion

Here we show how to put some literal text, that is, text without any formatting beyond what we give as text and spaces. Because it uses a monospaced font (i.e., courier or the like), it is useful for showing source code or text output from some program:

```
> library(RColorBrewer)
> pms <- pm(spikein133)
> mms <- mm(spikein133)
> colors <- brewer.pal(8, "Dark2")
> selected_probeset <- colnames(pData(spikein133))[1]
> pns <- probeNames(spikein133)
> indices <- (1:length(pns))[pns==selected_probeset]
> nsamples <- length(sampleNames(spikein133))
> matplot(t(pms[indices, 1:nsamples]), pch="P", log="y", type="b", lty=1,
  xlab="samples", ylab=expression(log[2]~Intensity), col=colors)
> matplot(t(mms[indices, 1:nsamples]), pch="M", log="y", type="b", lty=3,
  add=TRUE, col=colors)
```

However, when you'll use Sweave to build a vignette this formatting will be done automatically and you'll only need to insert the direct R code.

# Conclusions

Not all these LaTeX chunks are mandatory, you can add and remove what you want.

# References

[1] American Cancer Society. `http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-risk-factors`, 2016.

[2] Rosa M. Advances in the Molecular Analysis of Breast Cancer: Pathway Toward Personalized Medicine. *Cancer Control*, 22:211–9, April 2015.

[3] Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*, 490:61–70, October 2012.

[4] Bertos NR and Park M. Breast cancer - one term, many entities? *J Clin Invest*, 121:3789–96, October 2011.

# Appendix