Logo, company name

Description automatically generatedProject AHTA

Boris Vandemoortele

Tristan Vanneste

21/11/2020

# Abstract

The combination of multiple single experiments has the potential to reveal new insights and mechanics in complex diseases. Therefore, we performed an omics-based study to further analyse breast cancer. We analysed differences in gene expression between healthy and tumor tissue with microarrays and RNA sequencing (RNAseq) and linked this to the epigenomic differences (more specifically DNA methylation) with genome wide methylation profiling. Additionally, we analysed the ER cistrome to gain insight into endocrine treatment unresponsiveness in hormone receptor positive (HR+)breast cancers.

## Introduction

Among women, breast cancer is the most common cause of cancer related death worldwide in both developing and developed countries 1. 1.15 million breast cancer cases are diagnosed each year worldwide, making breast cancer one of the most prevalent cancers in the world today 2. Every year, over 411,000 deaths result from breast cancer, this accounts for over 1.6% of all female deaths worldwide. Breast cancer incidence is increasing in developing countries due to increased life expectancy and adaptation of the western lifestyle. Despite the common misconception that the majority of breast cancers are occurring in wealthy countries, the most breast cancer related deaths in fact occur in developing rather than developed countries. *Ferlay et al* estimated that the global incidence and mortality could even increase in the future 3. Finding an efficient and cost-effective treatment for breast cancer is thus an urgent unmet medical need.

As in most cancers, early detection and an accurate diagnosis are paramount for disease outcome. Early detection improves prognosis greatly, as physicians estimate that 70-80% of patients with early stage, non-metastatic tumors are curable. In Belgium, women between 50 and 70 years old can get a free mammography every two years to screen for breast cancer. If they belong to a risk group, they are allowed to do this yearly and starting from a younger age. Also, a genetic test for the *BRCA1* and *BRCA2* genes, of which inherited mutations are strong indicators of breast and ovarian cancer risk 4, is free for risk groups (addendum, figure x). Regular screening is however more difficult in developing countries, and cancer is often only detected in a very late and metastatic stage in these countries. The five year survival rate drops to 26% for breast cancers that developed to the metastatic stage 5. An exact mechanism that drives metastasis is yet to be unravelled, but some important new insights have been gained recently 6. Next to the occurrence of metastasis, also the cancer subtype is important for disease prognosis. Breast cancer is classically categorized into tree major subtypes. These subtypes are based on the presence of two molecular markers, being estrogen or progesterone receptor (ER or PR), which are both hormone receptors (HR), and human epidermal growth factor 2 (ERBB2). 70% of patients is HR+/ERBB2-, 15-20% is HR-/ERBB2+ and 15% is triple negative 7. The median survival rate is lowest for metastatic triple negative breast cancer.

Distinct cancer subtypes require distinct treatment methods. Patients presenting with HR positive tumors receive endocrine therapy. Tamoxifen, a synthetic ER inhibitor, is today’s gold standard of selective ER modulators (SERMs). It’s mechanism of action is a competitive inhibition of binding of estrogen to the ER 8. Another possibility is the administration of aromatase inhibitors such as *Astranazole.* They decrease circulation estrogen levels by inhibiting the conversion of androgens to estrogen by the aromatase enzyme. However, this kind of treatment is only applicable to post-menopausal women 9. Unresponsiveness to endocrine therapy has been described but the underlying mechanisms are not completely understood yet. We found that the ER cistrome differs genome wide between samples obtained from tumors that were responsive and non-responsive. Further research into this changing cistrome might help us understand endocrine therapy resistance in the future. ERBB2+ breast cancers are usually treated with ERBB2-targetted antibodies in combination with chemotherapy. The triple negative subtype, which is associated with the highest mortality, is difficult to treat and the only FDA approved treatment is chemotherapy 10.

In conclusion, breast cancer is a disease far from perfectly understood and needs further research to clarify the exact mechanisms behind the disease. Since cancers are known to completely change the genomic, transcriptomic and epigenomics landscape of cells, is an omics-based approach trivial to better understand the cancer in case. We, therefore, aim to identify differences in gene expression between healthy cells and breast cancer tissue, by analysing a microarray experiment and an RNA sequencing experiment. As transformation of healthy cells to cancer cells is often accompanied by epigenetic changes, a genome wide methylation profiling will be performed as well. If possible, differential methylation of genes and gene promotors will be linked to differential gene expression. Additionally, the ER cistrome was analysed in order to gain insight into endocrine treatment unresponsiveness in HR+ breast cancers.

# Materials and methods

## Microarray transcription profiling

The microarray transcription profiling data was obtained from ArrayExpress 11 under the ArrayExpress experiment identifier E-GEOD-1585212 in the raw format. 43 paired samples were collected from both tumor and normal tissues from multi-ethnic breast cancer patients. In total 86 samples were analysed for gene expression by using Affymetrix genechip U133A. Explorative Quality Control (QC) was performed on the normal data and the log-transformed data with the *arrayQualityMetrics* package. Background correction as well as quantile normalization were performed on the raw data with Robust Multi-array Average (RMA). A design matrix was made for the disease with a blocking factor for patients. Finally, differential expression (DE) was analysed using *limma* 13.  Adjusted p-values were computed, according to the Benjamini-Hochberg procedure, with a threshold of 0.05 for marking significant probes. Specific gene annotations were obtained using the BioMart query function.

## Methylation profiling by array

Genome-wide profiling of DNA methylation in 4 pairs of matched tumor tissue and normal breast tissue were obtained from the Gene Expression Omnibus (GEO) 14 data repository under the identifier GSE101443 15. These 8 samples were measured with an Illumina HumanMethylation450 BeadChip. Probes for which called p-values were insufficient were filtered out together with NA values. An initial analysis was performed on the average methylation percentages for tumor tissue vs control tissue with a t-test. The raw data was normalized using the *dasen* function in R. MethyLumiM objects were made from the normalized data and used for further analysis. Differential methylation (DM) analysis was also performed with *limma* and the adjusted p-values were computed according to the Benjamini-Hochberg procedure with an alpha threshold of 0.05 for significant probes. The design matrix was constructed with the factor condition and a blocking design for the factor patient. Finally, the annotation was obtained using the package *ChAMPdata* in R. The significant genes, with an more lax cut-off (alpha = 0.10) were analysed using Enrichr 16.

## RNAseq

Datesets were obtained from the *European Nucleotide Archive (ENA)* (Leinonen *et al.*, 2011)  under project ID PRJNA142887 17. This dataset RNA comprises sequencing results from both normal breast cells (HMEC) and breast cancer cells (HCC1954). Quality control was performed using fastqc 18 and overrepresented sequences were removed using Trimmomatic’s 19 paired end Illuminaclip function. Seed mismatches were set to two, palindrome clip threshold was 30 and simple clip threshold was 10. The human hg38 reference genome and genome annotation were downloaded from the ensemble genome browser and a pseudotranscriptome was built using Kallisto 20, after which a genome index was built using standard settings. Reads were quantified without bootstrapping. The obtained transcript abundance files were then pasted together to make further processing in R possible. Counts per transcript were summarized to counts per gene and then normalized using the TMM method in *EdgeR* 21. Since no biological replicates were present, we used a user defined dispersion value of 0.4. to identify DE genes. Threshold values for DE genes were set at an absolute log fold change (LFC) > 2 and false discovery rate (FDR) < 0.05. Finally, the Enrichr web application was used for gene ontology analysis.

## ChIPseq

Datasets were obtained from the *ENA* under project ID PRJNA175144 22. Only datasets from ER samples were downloaded. Fastqc was used to perform QC, after which overrepresented sequences were removed using Trimmomatic’s illuminaclip. A genome index was built using Bowtie2 23 and the same human genome as described in the RNAseq experiment. Aligning reads to the hg38 reference genome was performed with default settings. A small output formatting step was required to make this data compatible with our further analysis in HOMER (addendum, figure x line y). Tag directories were made and peaks were called using HOMER 24 and the inbuilt hg38 reference genome. We used a minimum fold enrichment over input of 8 for peak calling. Only uniquely aligned reads were preserved. Peaks from replicates were merged so that all peaks appearing in at least one replicate were retained and the overlap between genes associated to those peaks was visualized by the VIB Venn diagram tool. Gene ontology enrichment of these genes was performed in Enrichr and enriched motifs in these peaks were identified using standard settings in HOMER.

# Results

## Microarray transcription profiling

After loading the data from ArrayExpress we performed multiple QC on the raw data, log transformed data and normalized data. We used the RMA expression measure to perform background correction and quantile normalization. We found 5333 probes to be statistically significant after correcting for multiple testing with the Benjamini-Hochberg procedure (α < 0.05) (figure 1).

Chart, scatter chart

Description automatically generated

**Figure 1. Volcano plot for differential expression analysis of tumor vs normal tissue.** Red: probes that are statistically significant for DE (FDR < 0.05) with a blocking design for patients (n = 43) using limma. We found 5333 significant probes.

We have to be careful with interpreting that many statistically significant probes since we do not expect that many significant results. If we look at statistically significant probes and their LFCs, we see that most LFCs are rather low. We choose to filter out those probes that had an absolute LFC < 1 to only keep biologically relevant probes. After filtering for biologically relevant probes and probes with gene annotation we finally found 39 significantly DE genes (FDR <0.05 and absolute value (LFC) > 1.0).

## Methylation profiling by array

After filtering out probes that had too little counts and removing NA values, we found 0 probes to be significant at an alpha level of 0.05 (Figure 3). For further analysis we used an alpha level cut off from 0.10 to compare our results with amongst the different methods.

A picture containing chart

Description automatically generated

**Figure 3. Volcano plot for differential methylation analysis of tumor vs normal tissue.** No statistically significant probes were found using limma at the alpha level of 0.05 with the adjusted p-values computed according to the Benjamini-Hochberg procedure, with a blocking design for patients (n = 4).

We found 18540 probes to be statistically significant at the FDR level of 0.10 using *limma.* Since we are only interested in the most biologically relevant probes, we filtered out the probes that have an absolute (LFC) < 2. From the 18540 significant probes, 933 probes had an absolute LFC larger than 2. Only 622 of these 933 probes had gene annotation. 481 genes were found that have a positive LFC and 141 genes have a negative LFC. Note that a positive LFC matches with a higher methylation in tumor tissues compared to the normal tissue. A total amount of 231 genes were found when selecting for probes linked to promotor or exon 1 regions (the features ‘TSS’ and ‘1stExon’).

We further looked at the functions of these 622 significantly DM genes using Enrichr. In Enrichr, we found some interesting results regarding the molecular function of these DM genes (Table 2). We found that the 250 DM genes were significantly enriched for genes associated with DNA binding and hormone activity. Since there are 182 significant genes with a positive LFC and only 68 with a negative LFC, we can carefully conclude that the higher methylation in tumor samples (positive LFC) might lead to different hormone activity and DNA binding. The significant enrichment in genes with a hormone activity function could be explained by the fact that breast cancer is often driven by estrogen signaling 25. The increased methylation in tumor cells might lead to differential gene expression, which in its turn can influence estrogen signalling

## HCC1954 cells change their expression pattern to mediate cell survival.

Finding differentially expressed genes between healthy HMEC cells and cancerous HCC1954 cells was difficult due to the fact that no biological replicates were available in this study. Results should thus be critically evaluated before making any conclusions. When using a fixed estimated dispersion value of 0.4, 239 genes were found to be significantly upregulated and 184 genes were downregulated (FDR < 0.05 and abs(logFC) > 2). We then examined which cellular processes and functions these genes were associated to using Enrichr. Table 1 contains an overview of some selected hits that we believe to be interesting.

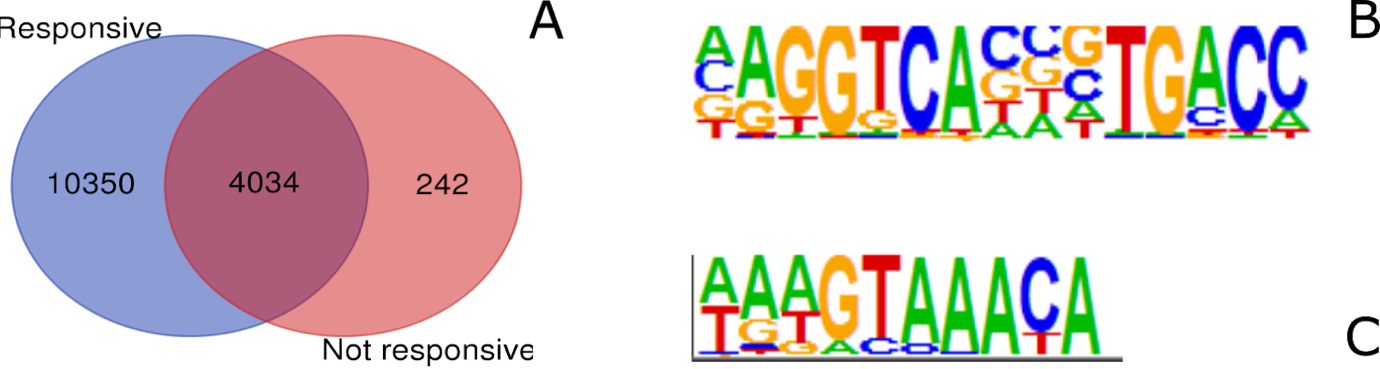
**Table 3: GO analysis of differentially expressed genes in HCC1954 cells vs HMEC cells.** 239 and 184 genes were up- and downregulated respectively in HCC1954 cells compared to HMEC cells (n=1, FDR < 0.05 and absolute values (LFC) > 2). Shown are the interesting hits of GO analysis in Enrichr with respective genes belonging to the given categories and associated p-values of enrichment (test).

|  |  |  |
| --- | --- | --- |
| GO category | associated genes | p-value |
| Upregulated genes | | |
| Glucuronidation | UGT1A10, UGT1A1, UGT1A5, UGT1A4, UGT1A3, UGT1A7 | 0.0000005154 |
| Positive regulation of monocyte chemotaxis | CXCL10, CCL5, CXCL17, S100A7 | 0.00004188 |
| Downregulated genes | | |
| Cytokines and inflammatory response | IL1A, CSF2, CXCL1, TNF | 0.00008847 |
| TNF signalling | CSF2, VEGFC, CXCL1, CXCL3, TNF | 0.003507 |
| Estrogen signalling pathway | KRT27, KRT16, KRT14, CALML3, HBEGF | 0.008786 |
| Regulation of cell proliferation | FGFBP1, TNFRSF6B, IRS1, EIF5A2, PINX1, VEGFC, IRS2, CXCL1, CXCL3, EREG, TNFRSF10D, FOSL1, IL1A, CCND2, IGFBP6, SOX7, HBEGF | 0.00005000 |

We believe our results show that HCC1954 cells are reprogrammed towards a more proliferating and immunosuppressive state, which is in line with cancer progression. TNF signalling for example, which is classically known as a potent mediator of cell death in cancer 26, was found to be downregulated in HCC1954 cells. We also found that other inflammatory response genes were significantly enriched in the list of downregulated genes (table 1). The reduced inflammatory signalling we observe might protect cancer cells from the immune system. However, we also found that upregulated genes were enriched for positive regulation of chemotaxis. The role of inflammation in cancer is not thus straightforward, since it can both positively and negatively affect tumour growth. Finally, the list of downregulated genes was enriched for genes associated to the regulation of cell proliferation. We thus show that HCC1954 cells maximize their survival chances by increasing cell proliferation and limiting inflammation.

## ER signalling seems to be intact in the highly diminished ER cistrome of aromatase treatment unresponsive tumors

A common cause of breast cancer is the uncontrolled proliferation of estrogen receptor-positive breast cells. Aromatase inhibitors, which block the conversion of androgen into estrogen performed by the enzyme aromatase (ref zoeken), aim to stop this process. Our dataset contained 2 groups of patients (n=5), one of which was responsive to aromatase inhibitor treatment and one who did not respond. After ChIPseq targeted against the ER, peaks were called using HOMER. After merging peaks within each group, we found 37372 peaks in the group that responded to aromatase treatment. Only 6854 peaks were found in the other group, which indicates that the ER has a severely compromised DNA binding capacity in this group. To further explore this phenomenon, we annotated peaks to the nearest gene and studied the compositions of these gene sets (figure 1, A). 4034 genes were associated with peaks in both groups, while 10350 genes were only found in the responsive group and 242 genes were only found in the non-responsive group. GO analysis revealed that genes involved in the nuclear receptor transcription pathway were enriched (p-value 0.003347, Bioplanet 2019) in the gene subset that was only observed in samples from non-responsive patients. This observation reflects the mode of action of aromatase inhibitors, namely blocking ER signalling, as ER signalling is still functional in these samples.



**Figure 4: ChIP sequencing reveals differences in ER DNA binding pattern between aromatase inhibitor responsive and non-responsive tumors.** A: Venn diagram showing the overlap between genes associated to peaks found in aromatase inhibitor responsive vs non-responsive tumors. Peaks were called by HOMER using a minimal fold enrichment over background threshold of 8 and peaks occurring in at least one replicate were merged within each group (n = 5). 4034 were found in both groups, while only 242 genes were uniquely found in the group non-responsive to aromatase inhibitor treatment. 10350 genes were associated to peaks that were only found in the group responsive to aromatase treatment, B: ERE motif identified by HOMER motif analysis, C: Fox1a motif identified by HOMER, this motif is unexpectedly the top hit in both conditions.

Motif analysis revealed a higher enrichment of ER elements (EREs) (figure 1, B) in peaks found in responsive samples compared to in peaks found in non-responsive samples, with 14,50% and 9,44% of target sequences containing the motif respectively. This supports our previous observations that ER DNA binding might be impaired in non-responsive samples. Strangely, the top motifs that are enriched in both conditions seem to be Fox-like (figure 1, C) motifs, while no major similarity between Fox motifs and ERE motifs seems to be present.

ChIP sequencing directed against the ER thus revealed a decreased ER DNA binding capacity in tumor samples that did not respond to aromatase treatment compared to samples that did, which is reflected by a lower number of peaks and a lower enrichment of ERE motifs in these peaks. Genes associated to peaks that are solely found in non-responsive samples are related to ER signalling, which suggests that ER signalling is still functional even after inhibition of estrogenic production. A mutation in the aromatase enzyme might cause this (voor in de discussie).

# Discussion

# References

1. Anderson, B. O. *et al.* Guideline implementation for breast healthcare in low-income and middle-income countries: Overview of the breast health global initiative Global Summit 2007. in *Cancer* (2008). doi:10.1002/cncr.23844.

2. Parkin, D. M., Bray, F., Ferlay, J. & Pisani, P. Global Cancer Statistics, 2002. *CA. Cancer J. Clin.* (2005) doi:10.3322/canjclin.55.2.74.

3. Ferlay J, Bray F, Pisani P, et al. Cancer Incidence, Mortality and Prevalence Worldwide. *GLOBOCAN* (2002).

4. Fackenthal, J. D. & Olopade, O. I. Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations. *Nature Reviews Cancer* (2007) doi:10.1038/nrc2054.

5. Peart, O. Metastatic breast cancer. *Radiol. Technol.* (2017).

6. Krøigård, A. B. *et al.* Identification of metastasis driver genes by massive parallel sequencing of successive steps of breast cancer progression. *PLoS One* (2018) doi:10.1371/journal.pone.0189887.

7. Waks, A. G. & Winer, E. P. Breast Cancer Treatment: A Review. *JAMA - Journal of the American Medical Association* (2019) doi:10.1001/jama.2018.19323.

8. Nazarali, S. A. & Narod, S. A. Tamoxifen for women at high risk of breast cancer. *Breast Cancer: Targets and Therapy* (2014) doi:10.2147/BCTT.S43763.

9. Schneider, R. E., Barakat, A., Pippen, J. & Osborne, C. Aromatase inhibitors in the treatment of breast cancer in post-menopausal female patients: An update. *Breast Cancer Targets Ther.* (2011) doi:10.2147/BCTT.S22905.

10. Collignon, J., Lousberg, L., Schroeder, H. & Jerusalem, G. Triple-negative breast cancer: Treatment challenges and solutions. *Breast Cancer: Targets and Therapy* (2016) doi:10.2147/BCTT.S69488.

11. Athar, A. *et al.* ArrayExpress update - From bulk to single-cell expression data. *Nucleic Acids Res.* (2019) doi:10.1093/nar/gky964.

12. Pau Ni, I. B. *et al.* Gene expression patterns distinguish breast carcinomas from normal breast tissues: The Malaysian context. *Pathol. Res. Pract.* (2010) doi:10.1016/j.prp.2009.11.006.

13. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* (2015) doi:10.1093/nar/gkv007.

14. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* (2002) doi:10.1093/nar/30.1.207.

15. Collignon, E. *et al.* Immunity drives TET1 regulation in cancer through NF-kB. *Sci. Adv.* (2018) doi:10.1126/sciadv.aap7309.

16. Chen, E. Y. *et al.* Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* (2013) doi:10.1186/1471-2105-14-128.

17. Hon, G. C. *et al.* Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res.* (2012) doi:10.1101/gr.125872.111.

18. Andrews, S. FASTQC A Quality Control tool for High Throughput Sequence Data. *Babraham Inst.* (2015).

19. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* (2014) doi:10.1093/bioinformatics/btu170.

20. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* (2016) doi:10.1038/nbt.3519.

21. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* (2009) doi:10.1093/bioinformatics/btp616.

22. Jansen, M. P. H. M. *et al.* Hallmarks of aromatase inhibitor drug resistance revealed by epigenetic profiling in breast cancer. *Cancer Res.* (2013) doi:10.1158/0008-5472.CAN-13-0704.

23. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* (2012) doi:10.1038/nmeth.1923.

24. Benner, C., Heinz, S. & Glass, C. K. HOMER - Software for motif discovery and next generation sequencing analysis. *Http://Homer.Ucsd.Edu/* (2017).

25. Xue, M. *et al.* Regulation of estrogen signaling and breast cancer proliferation by an ubiquitin ligase TRIM56. *Oncogenesis* (2019) doi:10.1038/s41389-019-0139-x.

26. O’Malley, W. E., Achinstein, B. & Shear, M. J. Action of bacterial polysaccharide on tumors. ii. damage of sarcoma 37 by serum of mice treated with serratia marcescens polysaccharide, and induced tolerance. *J. Natl. Cancer Inst.* (1962) doi:10.1093/jnci/29.6.1169.

27. Cai, X. *et al.* Inflammatory factor TNF-α promotes the growth of breast cancer via the positive feedback loop of TNFR1/NF-κB (and/or p38)/p-STAT3/HBXIP/TNFR1. *Oncotarget* (2017) doi:10.18632/oncotarget.16873.

28. Multhoff, G., Molls, M. & Radons, J. Chronic inflammation in cancer development. *Frontiers in Immunology* (2012) doi:10.3389/fimmu.2011.00098.

# Addendum

## Introduction

Afbeelding van de flowchart die toont wie in aanmerking komt voor BRCA test

Chart, histogram

Description automatically generated

## Methylation profiling by array

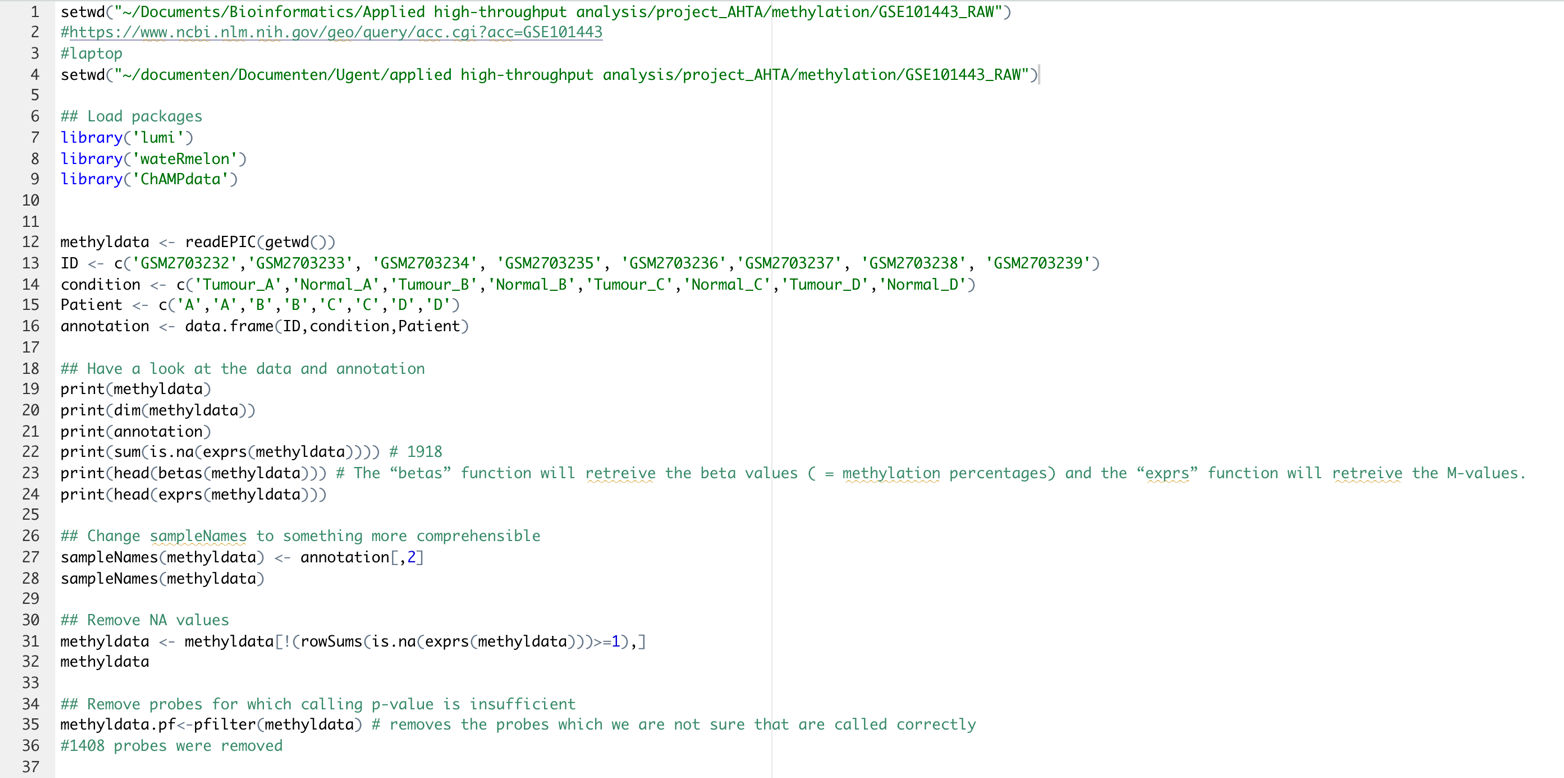


Figure 5. Loading in the data and filtering op probes which are not called correctly.

Graphical user interface, text, application

Description automatically generated

Figure 6. Comparing the average methylation between controls and tumor samples.

Chart

Description automatically generated

Figure 7. A boxplot of the Betas for each individual sample.

Chart, box and whisker chart

Description automatically generated

Figure 8. A boxplot of the average Betas for the tumor and normal tissue.

Text

Description automatically generated

Figure 9. Result of the t-test for the average Betas for tumor and normal tissue.

Graphical user interface, text

Description automatically generated

Figure 10. Normalization of the data.

Chart, histogram

Description automatically generated

Figure 11. A density plot of the Betas before and after normalization.

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generatedChart, histogram

Description automatically generated

Figure 9. A density plot of the M-values before normalization

Figure 10. A density plot of the M-values after normalization

Graphical user interface, text, application

Description automatically generated

Figure 11. Differentially methylation analysis with LIMMA.

A picture containing chart

Description automatically generated

Figure 12 A volcano plot of the differential methylation analysis with LIMMA.

Chart, scatter chart

Description automatically generated

Figure 13. MA plot of the differential methylation analysis with LIMMA.

Graphical user interface, text, application, email

Description automatically generated

Figure 14. Loading and linking of the annotation file.

Graphical user interface, text, application

Description automatically generated

Figure 15. Interpretation of the results.

Differentially methylated genes analysis in Enrichr:

<https://maayanlab.cloud/Enrichr/enrich?dataset=1e302a0a8bb41eddaf3245ad02633a95>

## RNA sequencing

Afbeelding met tekst

Automatisch gegenereerde beschrijving

**Figure 16: shell script to download data, perform quality control and adapter trimming**

Afbeelding met monitor, binnen, scherm, schermafbeelding

Automatisch gegenereerde beschrijving

**Figure 17: shell script to quantify reads using Kallisto**

Afbeelding met tekst

Automatisch gegenereerde beschrijving

Afbeelding met tekst

Automatisch gegenereerde beschrijving

Figure 18: R script for detection of DE genes after transcript quantification by Kallisto

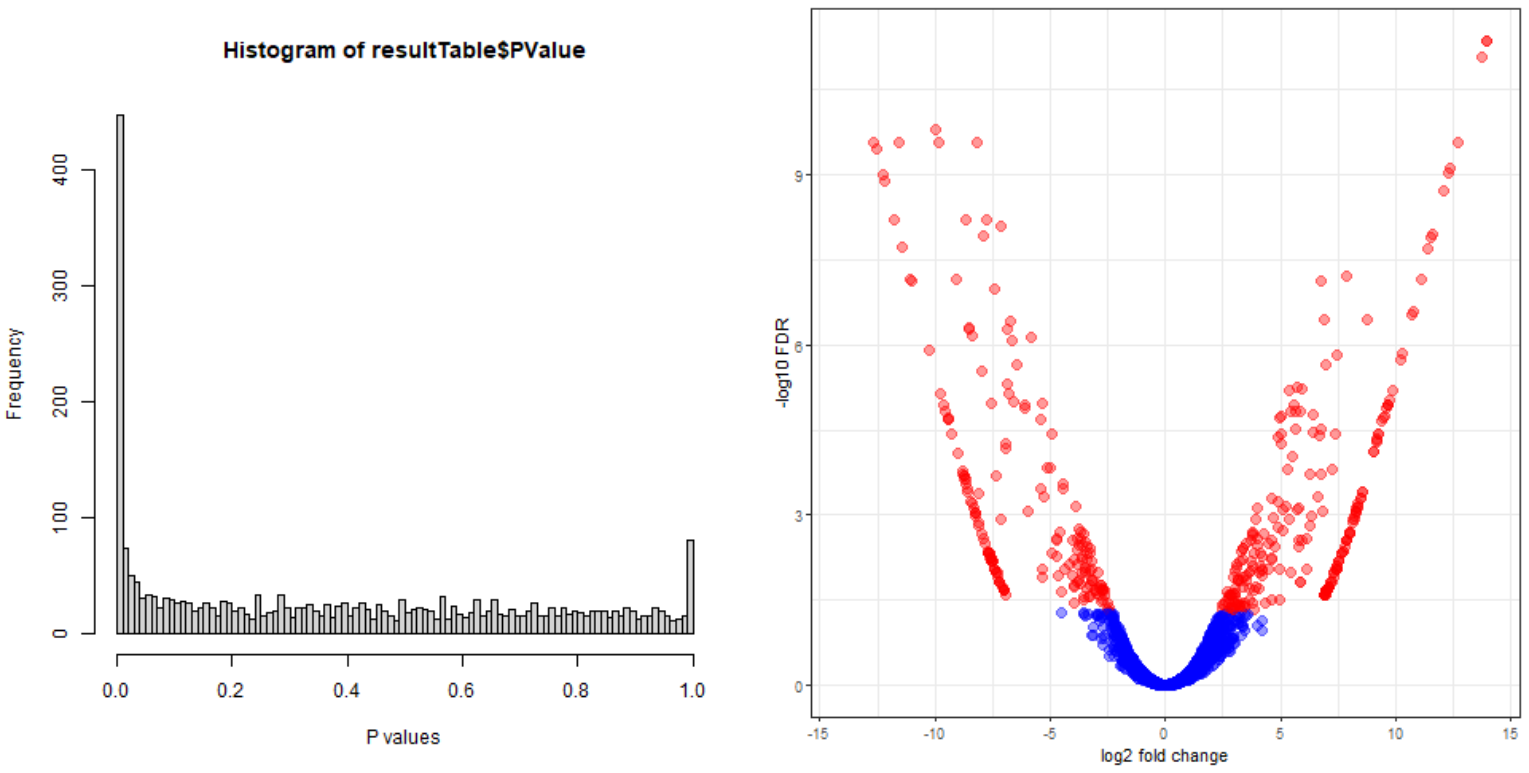


Figure 19: histogram of p-values and volcano plot after EdgeR analysis

RNA sequencing gene ontology analysis in Enrichr:

Upregulated genes: <https://maayanlab.cloud/Enrichr/enrich?dataset=07b6432de934c1db58623be24d1b62b5>

Downregulated genes: <https://maayanlab.cloud/Enrichr/enrich?dataset=11be9f939ee477c5c2eedf19064f2f52>

## ChIP sequencing

Afbeelding met tekst

Automatisch gegenereerde beschrijving

**Figure 19: shell script to download data, perform quality control and trimming**

Afbeelding met tekst

Automatisch gegenereerde beschrijvingAfbeelding met tekst

Automatisch gegenereerde beschrijving

**Figure 20: shell script for mapping to the hg38 reference genome, do some output formatting on bam files and rename files**

Afbeelding met tekst

Automatisch gegenereerde beschrijving

**Figure 21: shell script for peak calling using HOMER**

Afbeelding met tekst

Automatisch gegenereerde beschrijving

**Figure 22: shell script for HOMER motif finding**

Gene ontology analysis of genes associated to peaks:

Genes associated to peaks from the aromatase inhibitor unresponsive group: <https://maayanlab.cloud/Enrichr/enrich?dataset=381b05eb88816ca70959a5f8e8c8aa2b>

Genes associated to peaks in both groups: <https://maayanlab.cloud/Enrichr/enrich?dataset=67c4fb04642336c4ad2e68b80ea4805f>