Data Science project: thyroid specific antigens

Supervisors: Dr. Carl Herrmann, Dr. Maria Dinkelacker, Nils Mechtel

Camila Vacas, Line Weiß, Carina Keßler

19/7/2021

Introduction

Tissue restricted antigens (TRA) are those genes which are expressed in at most five out of fourty-five tissues, and additionally in the lymph system, e.g. in the medullary thymic epithelial cells (mTECs), where the negative selection of the T-cells takes place (Kyewske and Klein, 2006). A defective regulation of TRA expression in mTECs can lead to autoimmune diseases, caused by the presence of T-cells reacting to a self-antigen. In most patients of e.g. the Hashimoto's or Grave's disease, the most common autoimmune disorders concerning the thyroid, are antibodies present against the thyroglobulin or the thyroid peroxidase. Due to this negative selection, the immune system cannot react properly to tumor cells which express tumor associated antigens (TAA), like differentiation or onco-fetal antigens (Kyewske and Klein,2006). The tumor escapes the immune system and can grow. Radiation has a positive impact on tumor development as well, especially in thyroid cancer: the thyroid is higly sensitive to radiation exposure and it is likely to develop long-term effects. Though females tend to develope more often thyroid cancer than males, youth is a greater risk factor. (Albi, 2017)

Papillary thyroid cancer (PTC) represents 80 % of thyroid tumors. Usually, there is one gene mutated in the MAPK signal cascade, the most frequently affected genes are *braf* (point mutation, mostly V600E) and *ret* (RET/PTC rearrangement).

We were working with 14 samples of thyroid tissues. Five of them were extracted from patients, who were exposed to radiation and developed PTC, another four came from patients who were not exposed to radiation and developed sporadic PTC and the last five samples were from patients with healthy thyroid tissue. The gene expression analysis was performed using the gene chip hgu133plus2hsenstcdf.

Our first step of the quality control was the single chip control, and we decided to keep all of our chips. The normalization we did afterwards is necessary to reduce the impact of the usage of different amounts of fluorescent dye and it made the data comparable (Fig. 1). In the RNA degradation plots were some slopes deviating from the group pattern, which indicates that a lot of RNA was degrated. Even though there were some irregularities in this plot, we decided to reject none of our chips, because the rest of the quality control seemed to be fine.

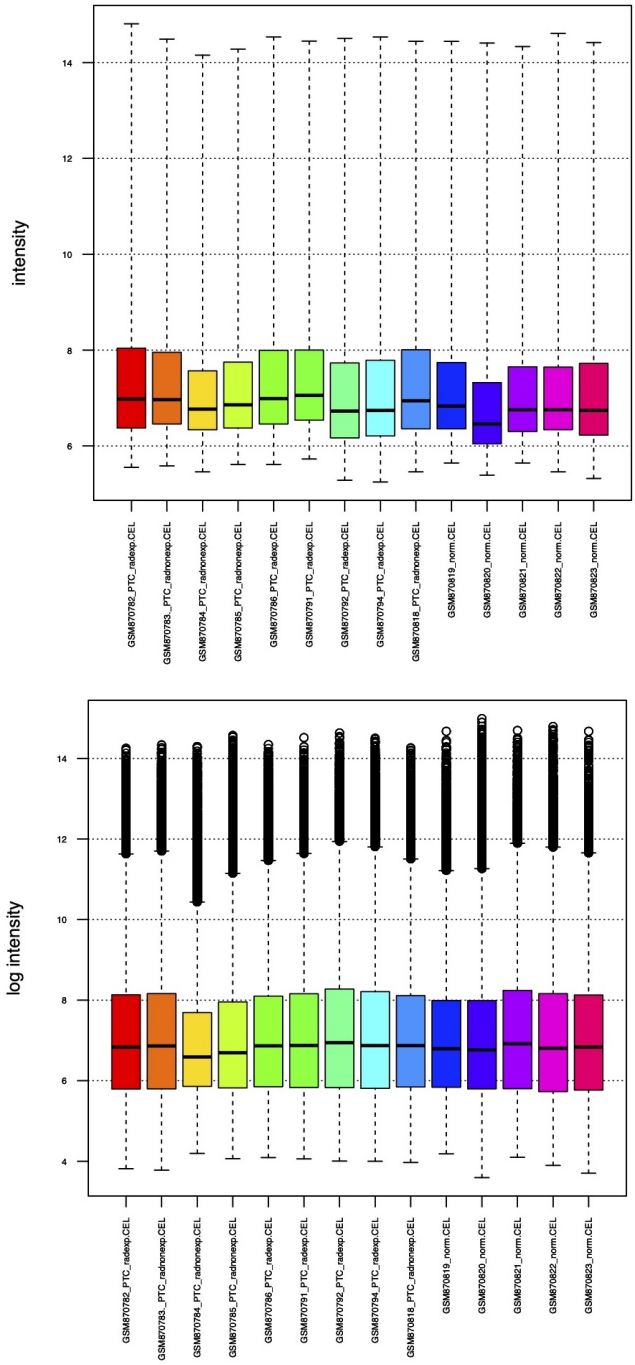

Figure 1: Gene expression in the thyroid tissue of 14 patients, who were either exposed to radion, not exposed to radiation or healthy, before and after normalization (GSE35570) (Handkiewicz-Junak et all., 2016). The boxplots show the distribution of the expression of all genes per chip before (top) and after (bottom) normalization. The goal is to bring the quantiles more or less into line, which makes the data comparable.

In the next steps we focused on the TRA's in the thyroid (our genes of interest), and extracted them by using five TRA lists (from human and mouse samples). We found 360 TRA's of thyroid tissue on our microarrays (see appendix) and created a matrix containig our genes of interest and their expression data. This matrix

was the basis for our following analyses.

We wanted to find out which of these thyroid specific genes are differentially expressed in low-dose radiation-induced PTC, sporadic PTC and healthy thyroid tissue and how the gene expression differs between the three groups.

Main Part

We started our project by taking a closer look at the data. We found out, that our genes of interest are spread over all human chromosomes, except for the Y chromosome (Fig. 2). Unfortunately we cannot say on which chromosomes the most thyroid specific TRA's can be found, due to the missing information of chromosome localisation of almost half of our genes.

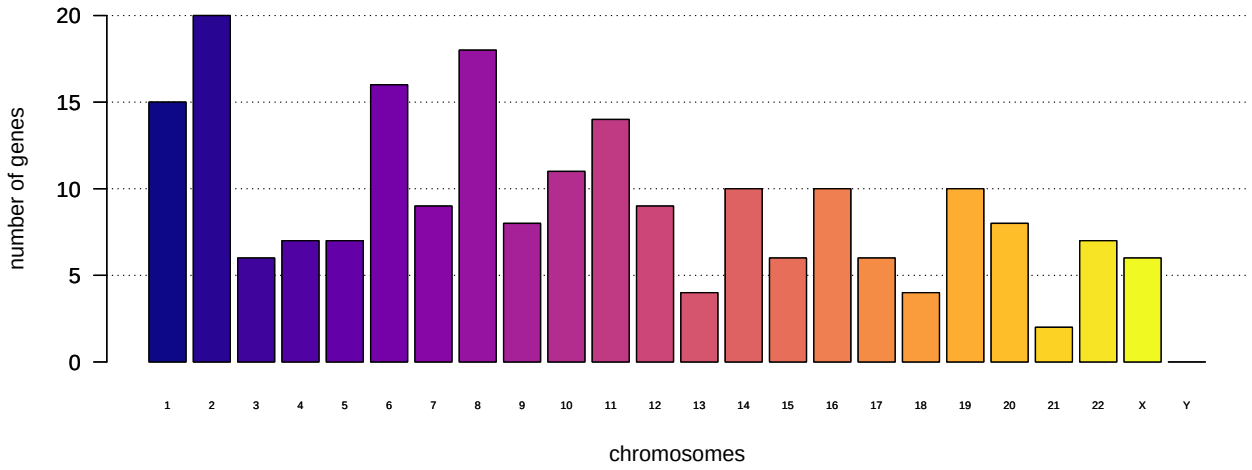

Figure 2: Distribution of thyroid specific genes over chromosomes. Only the genes with the information of chromosome localisation were taken into consideration. The TRA's of the thyroid are spread over almost all human chromosomes. Dinkelacker, 2019, PhD thesis, University of Heidelberg

We also created a heatmap to get an overview of the gene expression of the thyroid specific antigens (Fig. 3).

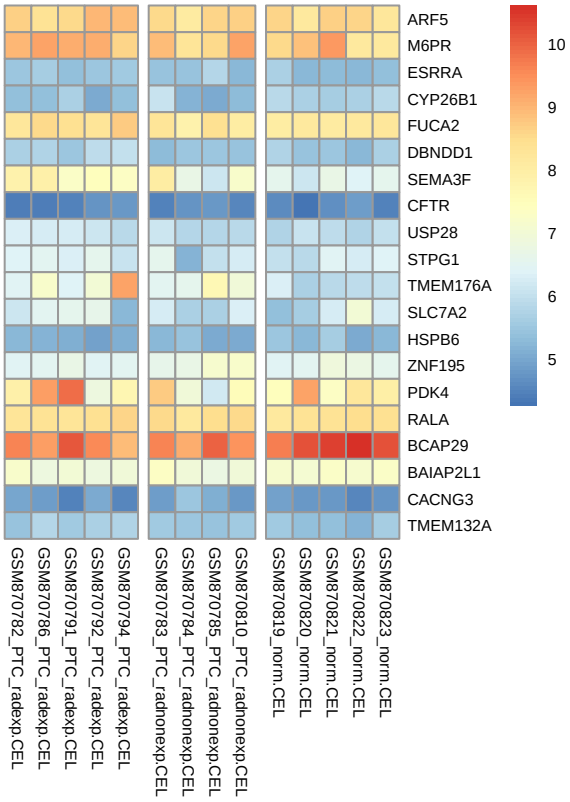

Figure 3: Heatmap of 20 thyroid TRA's. This heatmap illustrates the gene expression of the first 20 thyroid specific antigens in PTC and healthy thyroid tissue (GSE35570). (Handkiewicz-Junak et all., 2016)

We decided to perform a principal component analysis (PCA) to detect any patterns. We coloured the different spots according to the tissue sample the microarray was prepared with. The five healthy samples and the five radiation exposed samples formed both a clear individual cluster. Two of the samples of the sick patients, who were not exposed to radiation, formed a seperated cluster, while the other two seemed to be more similiar to the radiation exposed samples (Fig. 4).

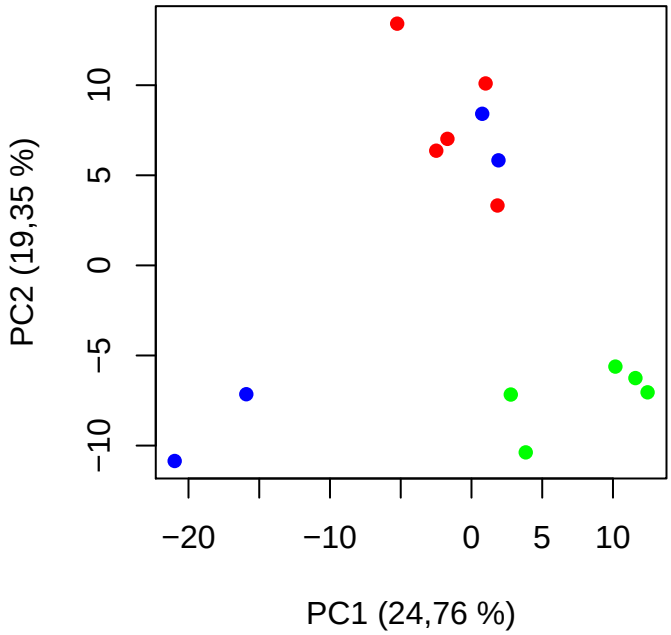

Figure 4: PCA of the 14 chips colored according to to tissue origin (red: radiation-exposed, blue: not-exposed, green: healthy). The healthy tissue samples form a clear cluster, while the sick samples do not.(Handkiewicz-Junak et all., 2016)

To find out, wether those were just outliers, we repeated the PCA with 30 chips, 10 chips in each group. This time it was clear, that the healthy samples formed one cluster and the sick samples another one (Fig. 5).

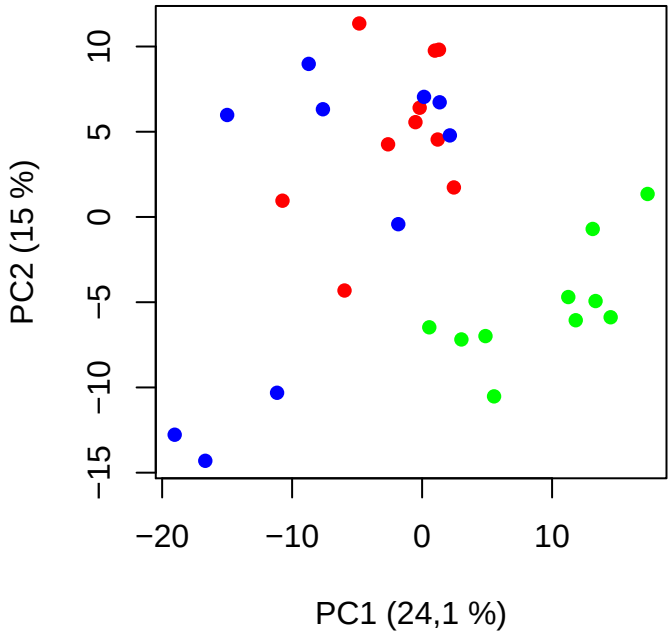

Figure 5: PCA of the 30 chips colored according to tissue origin (red: radiation-exposed, blue: not-exposed, green: healthy). The healthy tissue samples form a clear cluster.(Handkiewicz-Junak et all., 2016)

We wanted to prove this by performing k-means clustering and used the elbow method to determine the

best number of clusters, which was two (Fig. 6).

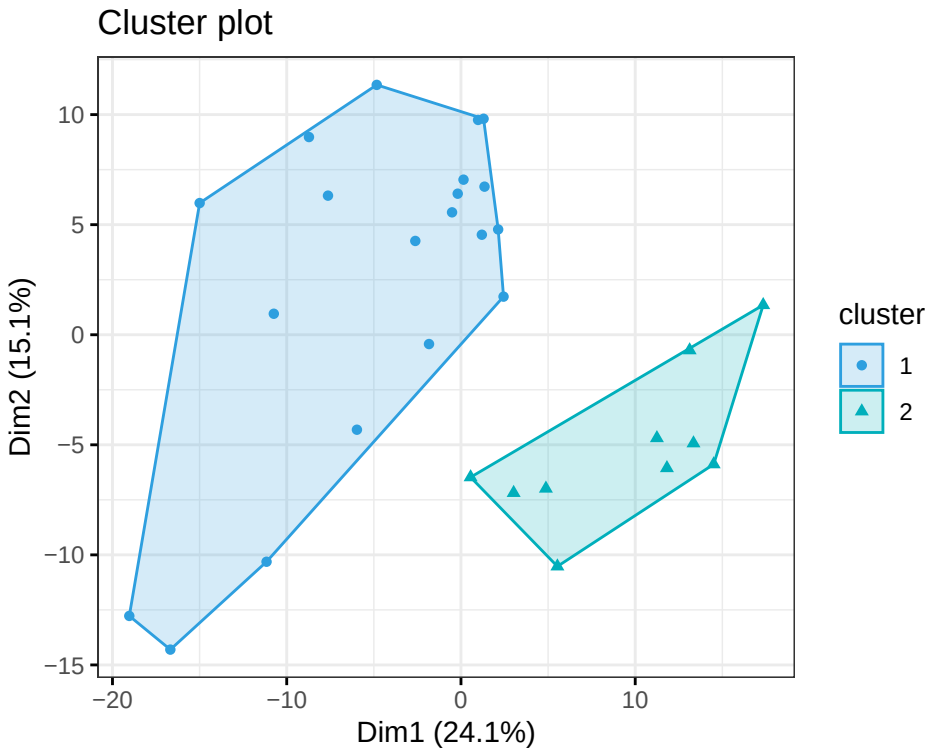

Figure 6: PCA of the 30 chips colored according to k-means clustering. As suspected, the healthy samples and the sick ones formed each one cluster.

To investigate, which of the 360 TRA's showed a significant change in gene expression upon the three different groups, we performed an ANOVA test for each gene. We found 153 significantly different expressed genes (see appendix).

```
# ANOVA Test for each gene
sig.genes.vector=c()

for (i in 2:dim(anovamatrix)[2]){
  if (summary.aov(aov(anovamatrix[,i]~ anovamatrix[,1]))[[1]][[1,5]] < 0.05){
    sig.genes.vector = rbind(sig.genes.vector, colnames(anovamatrix)[i])
  }

}
  # when the difference in gene expression for gene i between two groups
  # is higher than 5%, the corresponding gene name (=> colnames(anovamatrix)[i])
  # is added to a vector called sig.genes.vector
sig.genes.matrix= thyroid.expr.matrix.sub[sig.genes.vector,]
# the sig.genes.matrix contains the expression values of these genes.
```

The posthoc test of each of the significantly different expressed genes showed us how the according gene was expressed in the three groups and between which of the groups was a significant difference (Fig. 7). For that we used the TukeyHSD function to perform the tukey honestly significant difference posthoc test. It enabled us to compare the differences of the mean gene expression between our three groups. It helped us with the visualization and analysis of the ANOVA test as well.

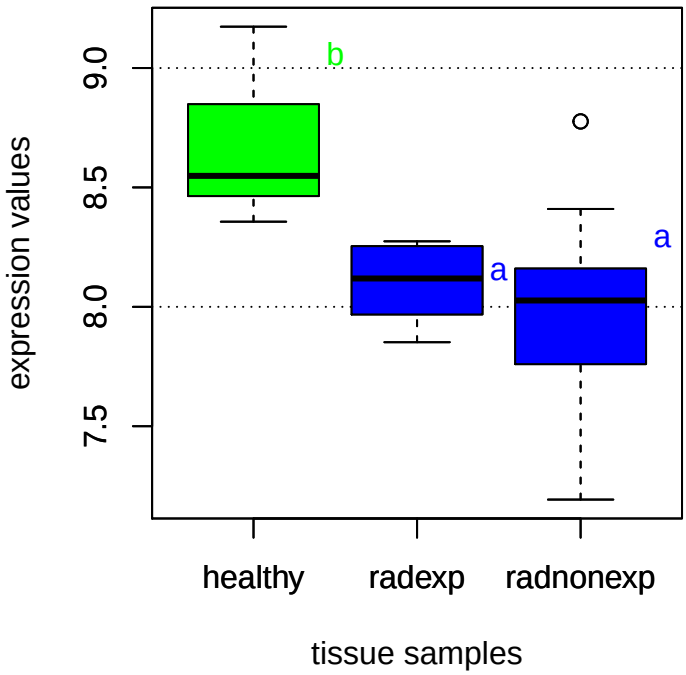

Figure 7: Example of a result of the posthoc test for the gene DCN. This gene is significantly different expressed in the healthy samples compared to the sick samples. DCN is higher expressed in healthy tissue than in cancer tissue.

Due to the results of the PCA, which showed us that the gene expression of radiation exposed and non-exposed tissue did not differ as much as we thought it would, we decided to concentrate ourselves on the cancer tissue further. We performed another PCA and a k-means clustering to investigate whether there is another pattern within these groups.

We found that the chips could not be seperated into two groups according to their radiation exposure. The elbow method, we used to define the perfect number of clusters, revealed, that the cancer tissues should be divided into three groups (Fig. 8). We assumed that these could be three different PTC subtypes.

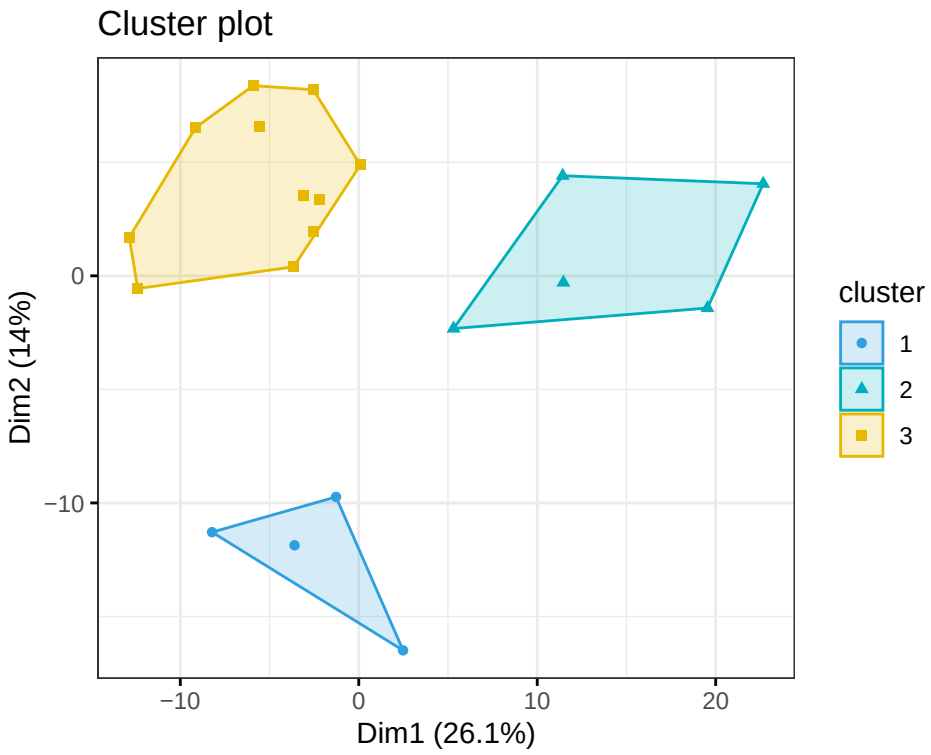

Figure 8: PCA of the 20 sick samples only, colored according to k-means clustering. The division of these 20 chips suggest that there are three PTC subtypes.

To look into this, we examined the gene expression pattern of these three cancer subtypes. We performed another ANOVA test, defining the genes, that are significantly different expressed in our subtypes. We found 134 genes (see appendix).

We wanted to design a regression model to predict the cancer subtype through the expression values of the significant genes. Since 134 variables are to many to work with, we performed another posthoc TukeyHSD test and selected the genes that are differentially expressed in all three subtypes (see appendix). With these eleven genes we trained our multinominal logistic regression model with 14 chips and tested it on another six chips.

```
##
## training.model subtype 1 subtype 2 subtype 3
##      subtype 1         3         0         0
##      subtype 2         0         4         0
##      subtype 3         0         0         7

##
## test.model  subtype 1 subtype 2 subtype 3
##   subtype 1         1         0         0
##   subtype 2         0         1         0
##   subtype 3         0         0         4
```

For the evaluation of the multinominal regression model we created the tables above. The columnames of the table stand for the actual subtypes the chips belong to and the rownames stand for the subtype the chips are assigned to by the regression model. The first table belongs to the assignment of the 14 chips we trained our regression model with and the second one to the assignment of the six chips we tested our model with. All chips were assigned properly.

Results

We identified 560 TRA's in the thyroid, but only 360 were represented on the microarray-type we used. Out of these 360 TRA's 153 were significantly different expressed between the three groups healthy, exposed and not-exposed to radiation. Through PCA we came across that there seemed to be no difference in the gene expression between the two carcinogenic tissue types. However, there was a clear difference between the gene expression in the healthy patients compared to the sick ones. We focused on the sick samples further and found three clear groups that differ in gene expression and suspected that we had discovered three subtypes of PTC. Moreover, we decided to create a regression model that would enable us to determine the subtype according to the gene expression. Therefore, we identified the 134 genes that were significantly different expressed between at least two of the subtypes using an ANOVA test. But since 134 genes were too many we narrowed them down to eleven genes, which were significantly different expressed between all three subtypes. Based on this eleven genes we were able to design a regression model with an accuracy of 100 %.

Discussion

We are aware that the number of chips we used is to little to get valid results. Although we could not find any clear differences between the radiation-exposed and not-exposed samples, an examination with more than the 30 chips, could come to another conclusion. Furthermore, we looked into the known variants of PTC and found out that there are three, that occur the most: the conventional variant, the follicular variant and the tall cell variant (LLyod et al., 2011). Even though there are many other less common subtypes as the three we mentioned before, we suspect that the three groups we discovered correspond to these variants. To prove this further analyses would be necessary. Additionally, the accuracy of the regression model seemed to high at first. But regarding the criteria for selecting the eleven genes the model is based on, it seems logical to get an accuracy of 100 %: All of our genes show a unique expression level in each subtype, so the model prediction has to be unambiguously.

During our work on this report we discovered that certain analyses are based on coincidence, which prevented us from recreating the same results. After looking into it, we found out that the expression values of our TRA's as well as the kmeans clustering results differed a little bit after every run of our code. We think that the change in the expression values is due to the normalization process, because it is a method based on coincidence, as well as the kmeans clustering method. This is why we decided to load those two variables into the RMarkdown, to present the results we found during our project. We are aware that this limits the validity of our results.

Literature

TRA datas:

GTEX 2015, Human genomics (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. GTEx Consortium. Science 348(6235),648-60.

Lattin, J. E., Schroder, K., Su, A. I., Walker, J. R., Zhang, J., Wiltshire, T., Saijo, K., Glass, C. K., Hume, D. A., Kellie, S., Sweet, M. J. (2008). Expression analysis of G Protein-Coupled Receptors in mouse macrophages. Immunome Res. 29,4-5.

Roth, R.B., Hevezi, P. Lee, J., Willhite, D., Lechner, S. M., Foster, A. C., Zlotnik, A. (2006). Gene expression analyses reveal molecular relationships among 20 regions of the human CNS. Neurogenetics 2,67-80.

Su, A. I., Cooke, M. P., Ching, K. A., Haka, Y., Walker, J. R., Wiltshire, T., Orth, A. P., Vega, R. G., Sapinoso, L. M., Moqrich, A., Patapoutian, A., Hampton, G. A., Schultz, P. G., Hogenesch, J. B. (2002). Large-scale analysis of the human and mouse transcriptomes. Proc Natl Acad Sci U S A 99(7), 4465-70.

Su, A. I., Wiltshire, T., Batalov, S., Lap, H., Ching, K. A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., Cooke, M. P., Walker, J. R., Hogenesch, J. B. (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A 101(16),6062-7.

Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinogl, A., Sivertsson, A., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I.M., Edlund, K., Lundberg, E., Navani, S., Szigyart, C.A.K., Odeberg, J., Djureinovic, D., Takanen, J.O., Hober, S., Alm, T., Edqvist, P.H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J.M., Hamsten, M., Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., Haijne, G., Nielsen, J. and Pontén, F. (2015). Proteomics. Tissue-based map of the human proteome. Science. 23;347(6220):1260419.

Others:

Albi, E., Cataldi, S., Lazzarini, A., Codini, M., Beccari, T., Ambesi-Impiombato., F. S., Curcio, F (2017). Radaition and Thyroid Cancer. Int. J. Mol. Sci. 18(5), 911

Dinkelacker, M. (2019). Chromosomal clustering of tissue restricted antigens, Dissertation, University Heidelberg, Germany.

Dinkelacker, M. (2007). A database of genes that are expressed in a tissue-restricted manner to analyse promiscous gene expression in medullary thymic epithelial cells. Diplomarbeit, Albert-Ludwigs-Universitaet, Freiburg, Germany.

Kyewski, B. and Klein, L. (2006). A Central Role for Central Tolerance. Annu. Rev. Immunol. 24, 571-606.

Lloyd, R. V., Buehler, D. and Khanfshar, E. (2011). Papillary Thyroid Carcinoma Variants. Head Neck Pathol. 5(1), 51-56.

Appendix

List of the 360 TRA's:

```
## [[1]]
##   [1] "ARF5"       "M6PR"       "ESRRA"      "CYP26B1"    "FUCA2"
##   [6] "DBNDD1"     "SEMA3F"     "CFTR"       "USP28"      "STPG1"
##  [11] "TMEM176A"   "SLC7A2"     "HSPB6"      "ZNF195"     "PDK4"
##  [16] "RALA"       "BCAP29"     "BAIAP2L1"   "CACNG3"     "TMEM132A"
##  [21] "DVL2"       "GGCT"       "RPAP3"      "IFRD1"      "UPP2"
##  [26] "PRSS21"     "HOXA11"     "PRR5"       "TRAPPC6A"   "SPATA20"
##  [31] "CEACAM7"    "CD79B"      "RHBDD2"     "TSR3"       "OSBPL7"
##  [36] "ARHGAP33"   "NDUFAB1"    "YBX2"       "PDK2"       "KRT33A"
##  [41] "IL32"       "TFAP2D"     "MAPK8IP2"   "PGLYRP1"    "CAMK1G"
##  [46] "CD9"        "CD74"       "RPS20"      "DYRK4"      "FAM76A"
##  [51] "CPA1"       "SYPL1"      "PLEKHG6"    "VPS13D"     "TSPAN9"
##  [56] "QPCTL"      "HIVEP2"     "PPP5C"      "NME1"       "UBR7"
##  [61] "MAP4K5"     "INMT"       "ERCC1"      "GPRC5A"     "HEBP1"
##  [66] "DNASE1L1"   "COX15"      "MS4A12"     "XYLT2"      "FTSJ1"
##  [71] "SCTR"       "RALBP1"     "PSD"        "SYT13"      "SNAI2"
##  [76] "VDAC3"      "SLC7A9"     "GABRA1"     "RTF2"       "RB1CC1"
##  [81] "AKAP11"     "PIGQ"       "CDH17"      "RNASET2"    "B4GALT7"
##  [86] "SLC39A9"    "FAM13B"     "RANBP3"     "CHPF2"      "GABARAPL2"
##  [91] "MYOC"       "FAM136A"    "NSMAF"      "OTC"        "TTC17"
##  [96] "HOXC8"      "MRI1"       "BOD1L1"     "TARBP1"     "RIPOR1"
## [101] "BEST2"      "RTN4R"      "PSMA4"      "RIPOR3"     "SPATA7"
## [106] "ZPBP"       "CTNS"       "LCP2"       "GOPC"       "DCN"
## [111] "TNFRSF17"   "MRPS10"     "GRN"        "THAP3"      "VAMP3"
## [116] "UTS2"       "RFC2"       "PPP1R3F"    "NEXMIF"     "ARHGEF5"
## [121] "NFE2L3"     "TNIP3"      "SIKE1"      "TLL1"       "CBLN4"
## [126] "CLDN11"     "FAM168A"    "RELT"       "CDC27"      "PKP2"
## [131] "IL1RAP"     "YIPF1"      "ADCK2"      "INPP4A"     "SLC2A3"
## [136] "ENPP2"      "CTDP1"      "STYK1"      "GNA11"      "CREB3L3"
## [141] "PIGV"       "HDAC7"      "MRPS35"     "BTN3A1"     "APPBP2"
## [146] "RPL18"      "CA11"       "ISOC2"      "EPN1"       "CD22"
## [151] "GSC2"       "ZFY"        "KCNQ1"      "TRPM5"      "TRIB2"
## [156] "OTUD5"      "GPKOW"      "AQR"        "KLK13"      "KLK14"
## [161] "ST6GALNAC1" "MBD3"       "LMCD1"      "FNDC8"      "CAMKK1"
## [166] "FAM50A"     "ACAP1"      "DERL2"      "KDM4B"      "PANX2"
## [171] "ICAM3"      "CTTNBP2"    "PRSS22"     "CEACAM1"    "CELSR3"
## [176] "REEP1"      "FSTL3"      "EPHA8"      "VASH1"      "LLGL2"
## [181] "KRT14"      "KRT20"      "FGF4"       "RGS11"      "MASP1"
## [186] "ST6GAL1"    "BPIFB2"     "REXO1"      "MKRN2"      "TNS1"
## [191] "NGFR"       "NTN1"       "TRO"        "MNT"        "AP3M2"
## [196] "LAPTM4A"    "ADAM7"      "PTPN18"     "DRD4"       "SCT"
## [201] "ALDH3A2"    "STK10"      "TBX21"      "MRPS34"     "TIGAR"
## [206] "MTMR7"      "FAM107B"    "TSPAN32"    "FAR2"       "TRAM2"
## [211] "CLDN18"     "ROPN1"      "NDUFB4"     "HEATR6"     "ERLEC1"
## [216] "TMEM131"    "ARPP21"     "TMEM38A"    "ZNF222"     "PSG6"
## [221] "ACTR6"      "NR1H4"      "FAP"        "NFKB2"      "MUSK"
## [226] "DMRT3"      "OSBPL6"     "CCN5"       "ANKRD16"    "GAL3ST2"
## [231] "IMPG2"      "PCDHB2"     "FAM135A"    "DLG3"       "PLOD1"
## [236] "TDRD3"      "OXCT1"      "ZNF324"     "ZNF416"     "FAM234B"
## [241] "PILRA"      "MOGAT2"     "DIMT1"      "NME8"       "MRPL28"
## [246] "HBQ1"       "RBM22"      "HSD17B2"    "MAP2"       "ZW10"
```

```
## [251] "PPP1R15A"   "TRIP6"      "CETP"       "TFAP2C"     "EPS8L1"
## [256] "NOL4L"      "PDRG1"      "PPP1R13B"   "TGM6"       "ESF1"
## [261] "OAS1"       "SIRT4"      "HAUS5"      "GMIP"       "LAG3"
## [266] "MLF2"       "OTUB2"      "TFAP4"      "PDCD7"      "SPG21"
## [271] "FLT3LG"     "GOLGA3"     "EFNB1"      "GLG1"       "PLEKHG2"
## [276] "DLL3"       "LAMB4"      "DLD"        "CMTM6"      "APOH"
## [281] "SPAG7"      "ESR1"       "RGS17"      "CMA1"       "HAUS4"
## [286] "OSGEP"      "SLC22A17"   "TGM1"       "TBX15"      "TEKT2"
## [291] "XYLB"       "OR1I1"      "ADH7"       "ADH1A"      "CDC6"
## [296] "CBX5"       "KLHL20"     "FMO2"       "DHPS"       "PSMD5"
## [301] "NANS"       "SEMA4G"     "TPSD1"      "MAPK13"     "BAG6"
## [306] "VARS1"      "PRRT1"      "LST1"       "ZNF184"     "VCL"
## [311] "SIRT1"      "TGFBR3"     "TMED1"      "ERMP1"      "MZF1"
## [316] "OCEL1"      "PSMD8"      "STX1B"      "BCL7C"      "ATP5F1D"
## [321] "AMELY"      "FGF22"      "SMIM24"     "IGFALS"     "MARCHF2"
## [326] "TIMM13"     "CDC34"      "MIF"        "TBC1D10A"   "SF3A1"
## [331] "RNF215"     "SNRPD3"     "TCN2"       "CRYBB3"     "MORC2"
## [336] "SLC25A1"    "PLA2G3"     "LGALS2"     "LGALS1"     "PIK3IP1"
## [341] "SRRD"       "PATZ1"      "SNU13"      "MICALL1"    "CENPM"
## [346] "KDELR3"     "DDX17"      "HSCB"       "CBY1"       "TOMM22"
## [351] "RSPH14"     "XBP1"       "RTCB"       "JOSD1"      "GTPBP1"
## [356] "DNAL4"      "C22orf31"   "MIOX"       "LMF2"       "RHBDD3"
```

List of the 153 significantly different expressed genes between radiation-exposed, not-exposed and healthy samples:

```
## [[1]]
##   [1] "CYP26B1"  "SEMA3F"   "USP28"    "TMEM176A" "BCAP29"   "TMEM132A"
##   [7] "GGCT"     "RPAP3"    "PRSS21"   "HOXA11"   "TRAPPC6A" "SPATA20"
##  [13] "CD79B"    "TSR3"     "PDK2"     "CD74"     "SYPL1"    "PPP5C"
##  [19] "UBR7"     "MAP4K5"   "INMT"     "HEBP1"    "COX15"    "SYT13"
##  [25] "VDAC3"    "RTF2"     "RB1CC1"   "RNASET2"  "SLC39A9"  "FAM13B"
##  [31] "CHPF2"    "MYOC"     "HOXC8"    "RIPOR1"   "RTN4R"    "SPATA7"
##  [37] "LCP2"     "GOPC"     "DCN"      "MRPS10"   "NEXMIF"   "ARHGEF5"
##  [43] "NFE2L3"   "RELT"     "CDC27"    "PKP2"     "IL1RAP"   "YIPF1"
##  [49] "INPP4A"   "SLC2A3"   "GNA11"    "PIGV"     "APPBP2"   "CA11"
##  [55] "ISOC2"    "CD22"     "GSC2"     "KCNQ1"    "TRIB2"    "GPKOW"
##  [61] "MBD3"     "ACAP1"    "DERL2"    "KDM4B"    "CTTNBP2"  "CELSR3"
##  [67] "FSTL3"    "VASH1"    "KRT14"    "KRT20"    "ST6GAL1"  "MKRN2"
##  [73] "TNS1"     "NGFR"     "TRO"      "MNT"      "PTPN18"   "STK10"
##  [79] "TBX21"    "TIGAR"    "TSPAN32"  "FAR2"     "TRAM2"    "ERLEC1"
##  [85] "TMEM131"  "ZNF222"   "PSG6"     "FAP"      "DMRT3"    "OSBPL6"
##  [91] "CCN5"     "ANKRD16"  "PCDHB2"   "OXCT1"    "ZNF416"   "FAM234B"
##  [97] "PILRA"    "NME8"     "MRPL28"   "RBM22"    "MAP2"     "ZW10"
## [103] "PPP1R15A" "PDRG1"    "PPP1R13B" "OAS1"     "SIRT4"    "SPG21"
## [109] "EFNB1"    "PLEKHG2"  "DLL3"     "DLD"      "CMTM6"    "APOH"
## [115] "SPAG7"    "RGS17"    "OSGEP"    "TGM1"     "TEKT2"    "XYLB"
## [121] "CDC6"     "CBX5"     "FMO2"     "TPSD1"    "MAPK13"   "VARS1"
## [127] "LST1"     "VCL"      "TGFBR3"   "ERMP1"    "MZF1"     "OCEL1"
## [133] "STX1B"    "ATP5F1D"  "SMIM24"   "TIMM13"   "SF3A1"    "SLC25A1"
## [139] "LGALS1"   "SRRD"     "PATZ1"    "CENPM"    "KDELR3"   "TOMM22"
## [145] "RSPH14"   "XBP1"     "DNAL4"    "MIOX"     "LMF2"     "RHBDD3"
```

List of the 134 significantly different expressed genes between at least two of the three cancer subtypes:

```
## [[1]]
##   [1] "ARF5"       "M6PR"       "SEMA3F"     "CFTR"       "USP28"
##   [6] "TMEM176A"   "CACNG3"     "RPAP3"      "PRR5"       "TRAPPC6A"
##  [11] "CD79B"      "RHBDD2"     "TSR3"       "PDK2"       "IL32"
##  [16] "TFAP2D"     "MAPK8IP2"   "PGLYRP1"    "CAMK1G"     "CD74"
##  [21] "SYPL1"      "HIVEP2"     "UBR7"       "MAP4K5"     "HEBP1"
##  [26] "DNASE1L1"   "XYLT2"      "RALBP1"     "GABRA1"     "RTF2"
##  [31] "AKAP11"     "PIGQ"       "B4GALT7"    "FAM13B"     "RANBP3"
##  [36] "CHPF2"      "FAM136A"    "TTC17"      "HOXC8"      "BOD1L1"
##  [41] "TARBP1"     "LCP2"       "GOPC"       "MRPS10"     "VAMP3"
##  [46] "NEXMIF"     "ARHGEF5"    "TLL1"       "FAM168A"    "CDC27"
##  [51] "INPP4A"     "SLC2A3"     "CTDP1"      "STYK1"      "GNA11"
##  [56] "HDAC7"      "BTN3A1"     "APPBP2"     "KCNQ1"      "TRIB2"
##  [61] "GPKOW"      "AQR"        "KLK14"      "ST6GALNAC1" "MBD3"
##  [66] "FNDC8"      "CAMKK1"     "FAM50A"     "ACAP1"      "PRSS22"
##  [71] "REEP1"      "FGF4"       "REXO1"      "MKRN2"      "MNT"
##  [76] "ADAM7"      "SCT"        "ALDH3A2"    "STK10"      "TBX21"
##  [81] "MTMR7"      "FAM107B"    "CLDN18"     "HEATR6"     "ZNF222"
##  [86] "FAP"        "MUSK"       "IMPG2"      "DLG3"       "PLOD1"
##  [91] "OXCT1"      "ZNF324"     "ZNF416"     "MOGAT2"     "MRPL28"
##  [96] "RBM22"      "TRIP6"      "CETP"       "PDRG1"      "TGM6"
## [101] "OAS1"       "GMIP"       "LAG3"       "PDCD7"      "FLT3LG"
## [106] "GOLGA3"     "EFNB1"      "GLG1"       "PLEKHG2"    "DLD"
## [111] "CMA1"       "OSGEP"      "SLC22A17"   "CDC6"       "KLHL20"
## [116] "NANS"       "TPSD1"      "BAG6"       "LST1"       "ZNF184"
## [121] "TMED1"      "MZF1"       "STX1B"      "ATP5F1D"    "AMELY"
## [126] "CDC34"      "SF3A1"      "MORC2"      "LGALS2"     "LGALS1"
## [131] "PATZ1"      "CBY1"       "LMF2"       "RHBDD3"
```

List of the eleven significantly different expressed genes between all three cancer subtypes:

```
## [[1]]
##  [1] "SYPL1"   "CHPF2"   "FAM136A" "HDAC7"   "TRIB2"   "STK10"   "TGM6"
##  [8] "GOLGA3"  "EFNB1"   "LGALS1"  "RHBDD3"
```