In our upcoming project we are going to analyze the expression of tissue restricted antigens, short TRAs, in pancreatic cells.

We want to concentrate our analysis on two main questions.

First, we want to find out if we can identify TRAs that are expressed in pancreatic ductal adenocarcinoma cells, short PDAC, which are not expressed in healthy ß-cells.

And secondly, we want to try to identify the fate of pancreatic cells, which means cell proliferation or apoptosis, based on their TRA expression.

I would like to start with a few epidemiological facts about pancreatic cancer. PDAC is the most common type of this cancer.

The median survival after the diagnosis is about 4.6 months. And only around 7 % of all patients still live 5 years after the diagnosis.

There are two main reasons for these unusually low numbers. On the one hand, pancreatic cancer most of the time is diagnosed in advanced stages. This is because it is often asymptomatic or has un-specific symptoms. Additionally, at the moment there are no sufficiently specific tumor markers known.

On the other hand, the treatment options are very limited. The cancer is often resistant to chemo-, radio- and molecularly targeted therapy and a surgically resection in later stages is usually not possible due to metastases.

This is why, today, in western societies pancreatic cancer is the 4th leading cause of death by cancer and it is predicted that until 2030 it may become the 2nd.

So it is clearly visible that it is necessary to develop new diagnostic and treatment options.

A promising approach are tissue restricted antigens, which play a role in establishing the central tolerance of the immune system. Developing T-cells get exposed to TRAs in the thymus and if they bind to these self-antigens, they get an apoptosis signal.

A malfunction of this negative selection process can lead to autoimmune diseases, like Diabetes Type 1.Here autoreactive T-cells attack β-cells inside the pancreas, which, among chronic inflammation, leads to ß-cell apoptosis.

TRAs are already and could be applicated in varying ways in other cancer types.The TRA Kallkrein-3 for example can be used as a tumor marker, to detect prostate cancer.Another application is in Immunotherapy, in which the immune system is stimulated to destroy the tumor itself. In some types of breast cancer testis-antigens are expressed which could be a target.

Or they could be used in classic cancer therapy, by binding antibodies with radiation treatment or chemotherapy, so the treatment is specifically targeted to the tumor. Which is now the standard treatment for Non-hodgens lymphoma.

Selina
We have two different types of data to analyze.
First ones are the TRA genes. We were given several tables where genes are described that are specific for certain tissues. Here (in the presentation) you can see two examples of the distribution of those TRAs between the tissue types. One is for humans and the other one is for mouse TRAs. After analyzing all the lists, we found that in total there are 330 human pancreas specific TRAs, and 59 mouse pancreas TRAs.

(slide change)
For the second data type we will be looking at Microarrays. They provide the necessary information about the gene expression in human pancreas cells. Over 10'000 genes were analyzed, but only 250 of those were part of the human pancreas TRAs we found earlier.
(slide change)
We have microarrays for two different pancreas cell types: β cells and pancreatic cancer cells
For each of those cell types there is a control and treatment group. The β cells were treated with the cytokines interleukin-1β and interferon-γ. The reason for that is to simulate the actions of the immune system in diabetes type 1.
The researchers treated the cells for different time periods, ranging from 0 to 96 hours.
On the other hand we have pancreatic cancer cells that were transfected with small interfering RNA. For the control group, unspecified siRNA was used, and for the actual treatment group they used siRNA targeted against the TBL1X gene, which encodes a transcriptional cofactor.
(slide change)
Here in this heatmap we plotted the expression values for each microarray and it clearly shows a successful knockout of the TBL1X gene.
By doing that the proliferation of the pancreatic cancer cells was reduced, as shown in the corresponding paper.
Between this heatmap and the actual raw data, several processing steps were done.
(slide change)
The most important one is the normalization of the expression values.
The first boxplot shows the distribution of all gene expressions for each microarray from the β cell data set. The distributions are not similar to each other regarding the quantiles, which makes it hard to compare the expression values. That is why we used the given function of the programming language R to normalize the values. The results are shown in the second boxplot. This data is now much better for further analysis.
(slide change)
However we have also found some outliers, which we are not sure about. That is why we are still in the process of reevaluating the quality of our microarrays, by using scatterplots.

slide change

Bianca (first part)

After we have gotten an overview of our topic and our data sets, we will take you through the steps we intend to take to answer our two main questions.

We drew up a timeline ranging from the day of our project proposal– to the 7th of July – by which we intend to finish our project. We chose the 7th of July to allow us a time buffer before our final presentation at the end of July, if anything should go wrong.

This leaves us with 8 weeks to answer our questions.

To achieve our goal, we structured our project into smaller pieces: Our Milestones.

We have already reached our first Milestone a few weeks ago. We cleaned our data – mostly, made sure gene symbols were assigned correctly to our Microarray data sets. We also created a new table, where all genes with low variance have been removed.

Overview Data:

After the Data Cleanup our next step is to take a look at the 3 Datasets individually and get a first impression through descriptive statistics and obtain some additional information with clustering.

In this context we also want to take a closer look at the expression of pancreas specific antigens in the additional Breast cancer dataset.Can an atypical gene expression be found across the Breast Cancer subtypes?

In our analysis we later want to compare healthy ß cells against pancreatic cancer cells. So first we need to determine whether the time factor is significant in ß cells.

Now moving on to our second milestone (**slide change**):

Second milestone:

Here we'd like to address the overall question if a possible biomarker or drug target can be identified which would ideally be highly selective to Pancreatic cancer but not ß cells.

For that we concentrate on the two control sets and analyze their TRA gene expression. First we'd like to start with an initial clustering after having run a PCA in order to get an initial look at the key differences between those two control sets.

After that we can look more closely at the TRA gene expression. Here we want to identify highly expressed genes in both cell types. After that we can determine the differentially expressed genes - so using an statistical Test to calculate the significance among the expression differences for each TRA gene.

(**Slide change**)

Having identified these three groups we can compare them and look for overlap between them, as shown in the diagram. Focusing on TRA genes that are highly expressed in Pancreatic cancer but not in Diabetes.

Those genes can then be looked at more closely concerning different aspects like their function and location in a cell by checking against existing literature. Doing so we can hopefully answer the question if a good selective gene for Pancreatic cancer can be identified.

slide change

Bianca (second part)

After reaching our second milestone, we will begin working on our third task.

We want to compare the levels of gene expression of beta-, diabetes- and pancreatic cancer cells. The result of this second main question should be very interesting: beta-cells have a highly increased cell death in diabetes, while pancreatic cancer cells are known for their high proliferation rates. Therefore, our goal for our third milestone is to identify TRAs that are linked to similarities or differences in the fate of pancreatic cells.

We will start by comparing beta-cells to diabetes cells, to research the effect cytokines have on beta-cells through clustering. Hopefully, we will be able to answer the question how TRA expression changes during treatment.

Following we will divide ourselves into two groups. One of which will compare just identified, differentially expressed TRAs to the gene expression of TRAs in pancreatic cancer cells. By

doing so, we will be able to visualize differences of TRA expression in pancreatic cells and thus answering our second main question.

The second group will run a linear regression model over our data set of beta-cells treated with cytokines using the temporal component.

If we have time left, we would also like to compare pancreatic cancer cells with their knockout counterparts to determine differences in TRA expression.

In the last week we plan to finalize our project: summarize our answers, choose meaningful graphics and bring all of this together in our final RMarkdown file.

As I already mentioned, should anything go wrong or take longer than previously anticipated, we have a buffer to ensure the success of our project.

Next, we want to give a short overview of what we have already accomplished.

For our first milestone we conducted a Principal Component Analysis as well as kmeans clustering. Here our results for the control and treatment group of pancreatic cancer cells are shown.

On the left you can see the PCA. The six dots represent our microarrays, which have been arranged according to their differential gene expression. The x-axis represents principal Component 1, making up about 30% of the variance, while the y-axis contains PC2 with about 20%. PC1 also seems to contain the main difference of control and treatment group, distinguished here by different colors.

Just as expected, we can see through kmeans clustering on the right that the control and treatment group have been sorted into two different clusters.

slide 14:

In context of the second milestone we looked at the TRA gene expression across the two control sets:
Here the goal was to identify highly expressed TRA genes in the pancreatic cells:
This boxplot shows the expression levels of the highest expressed genes in the PDAC control dataset. Those were determined using a threshold of 1.8 x the median value.

These genes can then be compared to the differential expressed genes, which we selected using a wilcoxon rank sum test.

page 15:

To visualize the differential expressed genes in pancreatic cancer cells and beta cells, we used a volcano plot. The y axis contains the p value of the conducted wilcoxon rank sum test between the two groups. Genes that are significant are above the red line.

On the x axis is the log2 foldchange value based on the ratio between the groups. Positive values in yellow mean that genes are higher expressed in beta cells compared to pancreatic cancer cells and negative values in blue mean that gene expression is higher in pancreatic cancer cells.

page 16: In this plot we compare the gene expression of treated vs. untreated beta cells. A significantly upregulated gene in the treated group is CXCL11. We were excited to find out that it is a chemokine known to be induced through IFN-y which regulates activated T-cell migration.

We are looking forward to sharing our progress and final results in our next presentation.

Thank you for your attention.