Assessment and evaluation of heterogeneity in data from immune infiltration spatial niches in lung cancer

Bachelor's Program in Mathematics

Mahta Keivani Najafabadi

August 2022

Abstract

The protein biomarker expressions in three types of sampled immune INFILTration spatial niches in lung cancer tissue were measured using the new technology Digital Spatial Profiler (DSP). The three types of immune INFILTration that were observed in lung tumors were STROMA identified as immune cells separate from tumor cells, Tertiary lymphoid structures (TLS) identified as dense structures of organized immune cells and finally Infiltraterate where immune cells dispersed among and in direct contact with tumor cells (INFILT). The pairwise inter and intra-patient correlation between the protein biomarkers were evaluated using the Bland- Altman methods for Calculating correlation coefficients with repeated observations. The result showed that the absolute value of the inter- patient correlation levels were higher for sample type INFILT compared to STROMA while the absolute value of the intra- patient correlation levels were slightly higher between the biomarkers of the sample type STROMA. In order to investigate added value of sampling multiple regions from individual tumors, the intra- patient heterogeneity of the protein markers in the three different spatial niches were evaluated. To this end, three different estimators were used: standard deviation, median absolute deviation and range. After comparing the results, it was concluded that standard deviation was the preferred method. Since it is applied on the complete set of available data and captures the behavior of the tail of the data which is desirable for our purpose. The mean squared error in the ANOVA table, with the patients identity as the independent variable and marker values for each sample type as the dependent variable was calculated as a measure of heterogeneity of the markers within patients.

Popular science summary

In cancer treatment, drugs that modulate the immune system to more effectively combat the tumor, known as immunotherapy, have become increasingly important. The form and level of immune infiltration in tumors vary substantially within and between tumors. A new technology make it possible to select certain areas on the tumor tissue samples and measure the amount of certain proteins in the chosen areas. Analysis of the protein expressions is then the key to identify proteins that could assist us in tumor subtyping, determination of prognosis and optimal therapeutic strategy for individual patients. In this study, Three types of immune infiltration spatial niches in lung cancer tissue have been sampled and more than 40 proteins have been measured in the chosen areas of the samples.

We started our analysis of the protein expressions by examining their distribution in the three sampled regions. This approach could be helpful in pinpointing potential targets associated with the presence of particular immune niches.

Next, we examined the linear relationship between the measured protein pairs both between and within patients in the three sampled regions. We had to take into consideration the fact that some samples came from the same patients and therefore could not be assumed to be independent samples. Establishing the strength and direction of this relationship is in particular important for variable selection in future model building.

Since the samples come from different types of tissue structures, variation in the measured protein expressions was expected. The challenge was how to assess the level of variation in the measurements to potentially make use

of, for instance, in predicting the outcome of a certain treatment. Different methods were used to estimates of the variability in protein values within patients.

The result suggested an overall reduction in the magnitude of the within-patient pairwise linear relationship of the proteins compared to the linear relationship observed within patients. This could potentially indicate that the relationship varies in different subgroups of the patients and needs to be further investigated with more samples.

The result of the estimated variation in protein expressions, suggested that level of variability varied between spatial immune infiltration niches and between the different biomarkers. This would highlights the added value in sampling multiple regions of different types from the same tumor. However, similarities between the results obtained using different methods could potentially be worrisome. Further examination of the result indicated that some of the similar patterns could be, for instance, associated with the vulnerability of the estimators to deviation of the distribution of the data from a normal or symmetric distribution. The measurement of the variability and its potential predictive value could be further studied in the future by employing estimators that are less sensitive to the shape of the distribution of the data.

Acknowledgements

I would like to express my gratitude to my co-supervisor Anna Sandström Gerdtsson for providing the data as well as for her invaluable guidance and constant support throughout this work. I would also like to thank my co-supervisor Magnus Wiktorsson for his many thoughtful insights, suggestions and feedback.

Contents

Chapter 1

Introduction

1.1 Background

Analysis of protein expression in tumor tissue is important to identify biomarkers for subtyping tumors, establish prognosis and suggest optimal therapeutic strategy for each patient. With immunotherapy, cancer can be treated with drugs that modulate the immune system to more efficiently combat the tumor.

The level and type of immune infiltration differs substantially between and within tumors. Tumor tissue have previously been assessed with low plex imaging (one biomarker at the time) or high plex methods where the spatial tissue context is not retained and the tumor heterogeneity is not captured. With recently established spatial omics technologies, we can now combine staining and tissue imaging with multiplex analysis of specific regions of the tissue. In this study, three types of immune infiltration spatial niches in lung cancer tissue have been sampled: infiltrating immune cells in direct or close contact with tumor cells (INFILT), immune cells present in the surrounding stroma and not in direct contact with tumor cells (STROMA) and immune cells present in tertiary lymphoid structures (TLS), which are dense regions of immune cells located outside the tumors. TLS are only present in a subset of patients and their role in tumor progression is still largely unknown. Figure

1.1 shows a visualization of these three regions in a single biopsy.

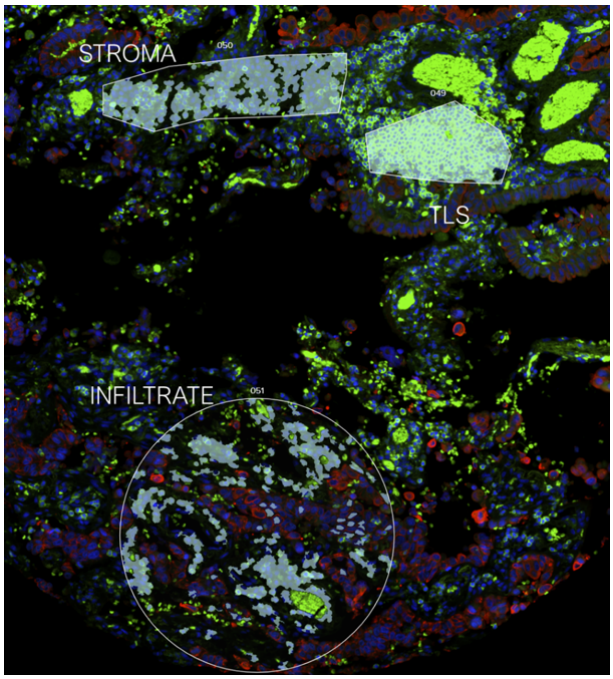


Figure 1.1: Three types of immune structures sampled from one biopsy, Red: tumor cells, green :immune cells, blue: cell nuclei

1.2 Data

Using the Digital Spatial Profiling technology (DSP), tumor tissue biopsies were stained with antibodies specific for immune- and tumor protein biomarkers. Three antibodies were coupled to fluorophores and used to visualize in a scanner (i) cell nucleus, (ii) tumor cells and (iii) immune cells.. The scan images were used to select regions of interest (ROIs) in the samples 1.1. The remaining antibodies were coupled to barcodes specific for each antibody. The barcodes were cleaved off the antibodies by directing UV-light at each ROI. For each ROI, we get a direct quantitation of how many barcodes (antibodies) that were present, which is in proportion to the level of the protein biomarker.

The data set consisted of 186 samples from three types of immune infiltration spatial niches in lung cancer tissue from 33 patients. Out of these 186 samples, 126 samples were of type STROMA from 30 patients, 41 samples of type INFILT from 14 patients and 19 samples of type TLS from 10 patients.

For the purpose of assessment and evaluation of heterogeneity within patients, we only looked at a subset of patients with at least three ROIs of the same sample type. Out of the 33 patients in the study, there were 23 patients with at least three samples of the type STROMA and a total number of 115 samples, 8 patients with at least three samples of the type INFILT and a combined number of 30 samples and finally 2 patients with more than two samples of type TLS and a total of 9 samples. It's worth noting that there were no patients in the cohort who satisfied this condition with respect to all three regions.

In order to calculate the pairwise correlation of biomarkers within patients, a subset of patients with at least two ROIs of the same sample type were used. Out of the 33 patients in the study, there were 27 patients with at least two samples of the type STROMA and a total number of 123 samples, 13 patients with at least two samples of the type INFILT and a combined number of 40 samples and finally 3 patients with more than one sample of type TLS and a total of 13 samples.

Fourty-four protein markers were measured per ROI (Appendix A). Normalization of data had been done prior to the start of the thesis project, by scaling to positive control proteins, which inherently normalizes differences in cell numbers, sampled area and non-specific antibody binding (background noise) (see Appendix B). Data was log2 transformed after normalization. Figure 1.2 shows the boxplot of the marker values from the three ROI type, which demonstrates that there is substantial patient-dependency in biomarker expression also after data normalization.

1.3 Objectives

The following questions were defined to be addressed during the course of this project:

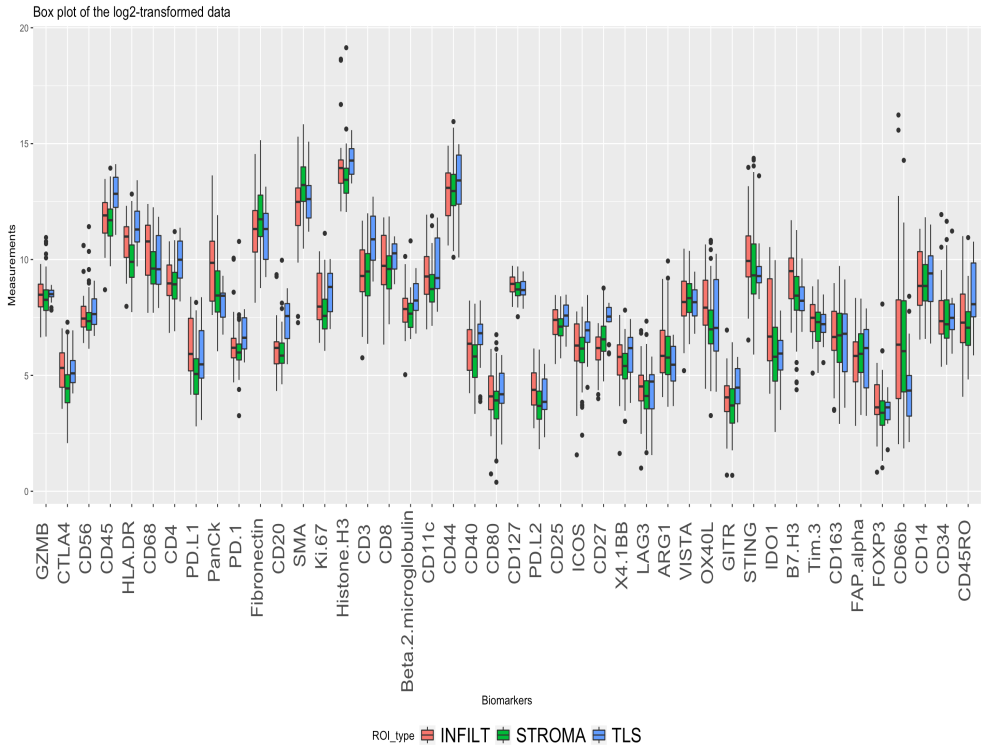


Figure 1.2: Boxplot of log2-transformed marker values

What is the underlying probability density of the measured biomarker values for each ROI-type across all patients? Density estimates offer an easily comprehensible presentation of the data for examination of the properties and features of the data [15] and application of appropriate tools for statistical analysis of the data.

What is the level of intra- and inter- patient pairwise correlation between the biomarkers for each sample type? The intra-patient pairwise correlations means the within single patient correlation between the markers while the intra-patient correlation refers to the between-patients correlation between the markers. Establishing the degree of pairwise linear relationship between the variables is in particular important for variable selection in future model building.

How beneficial/necessary is sampling multiple regions from individual tumors? This question was concerned with finding a suitable measure of heterogeneity in digital spatial profiling (DSP) data of immune infiltration in lung tumors and examining the added value of sampling multiple regions from individual tumors by characterizing heterogeneity in the data.

Chapter 2

Methods

2.1 Kernel Density Estimate

Given a sample of independent, identically distributed observations from a continuous univariate distribution, the kernel estimator is a nonparametric approach to estimating the density. It is obtained as the sum of individual kernels that are placed at each observation. The shape of the individual kernels is determined by the kernel function and their width is controlled by the smoothing parameter or bandwidth. Choosing a larger value for the smoothing parameter creates a smoother density that may hide the important details for too large values of the bandwidth while a smaller smoothing parameter creates more details that turn into spikes at the observations as the bandwidth approaches zero [15].

A useful method for choosing the bandwidth and shape of the kernel is to plot out and examine several density curves with different smoothing parameters and kernels to get more insight into the data and subsequently choose a density estimate that is subjectively more satisfactory given one's prior ideas about the underlying density of the sample. It's in particular useful to present the data in several plots with varying smoothness that may lead to different explanations when the purpose of the density estimation plots is to investigate possible models and hypothesis [15].

2.2 Correlation Levels

2.2.1 Pearson correlation

Pearson product-moment coefficient r is used as a measure of the degree and direction of the linear relationship between the protein pairs [11]

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are sample vectors of length n and $\bar{x}$ and $\bar{y}$ represent the corresponding sample means. The coefficient r tells us about the magnitude as well as the direction of such relationship. This analysis does not take into consideration the fact that for each sample type, there are different number of ROIs from individual patients.

2.2.2 Intra-Patient correlation

As mentioned earlier, the data set consists of different number of ROIs of the same sample type from each patient. Therefore, it may be misleading to calculate the pairwise correlation coefficients of the biomarkers for each sample type as if they were independent random samples. Bland and Altman introduced a technique [3] that can be used to calculate the pairwise intra-patient (within-patient) correlation coefficients when there are more than one ROIs of the same sample type from the patients.

The Bland- Altman approach means that in order to assess the intra-patient correlation between two biomarkers, we need to remove the inter- patient (between-patients) differences and look at the intra- patient variability. The one-way analysis of covariance (one-way ANCOVA) is used to partition the total intra- patient variation in a biomarker into two components: the variation that can be explained by the linear relationship between the two biomarkers and the unexplained variability. One-way ANCOVA can be viewed as an extension of one- way analysis of variance (one-way ANOVA) and the main ideas related to ANCOVA that are utilized in Bland and Altman analysis

can be demonstrated with an example [16] [1]. Suppose that we wish to assess the relative effectiveness of p different treatments on score of a test by conducting a p-group experiment. N participants are randomly assigned to the p treatments and the test scores are collected after the completion of the treatment process. In the framework of one-way ANOVA, the observed test scores can be written in the following fixed-effect model form:

$$y_{ij} = \mu + a_j + e_{ij}$$

Where y_{ij} is the test score from the ith participant in treatment j, μ is the grand mean, a_j is the effect of treatment j and e_{ij} is the random error associated with y_{ij}. The total variability in the test scores, defined as the total sum of squares, is split into two components: the variation between the groups and the variation within the groups:

$$SS_{Total} = SS_{between} + SS_{Within}$$

The total sum of squares is calculated as the sum of the squared differences of the test scores from the grand mean. The total sum of square is divided into the within treatment sum of squares which is given by the of sum of the squared deviation of the test scores from the group mean and the between treatments sum of squares which is given by the difference between the total and within treatment sum of squares.

Lets suppose that we're aware of the fact that the test score is correlated with patient age (the covariate) and that there are considerable age differences between patients within the treatment groups. Thus, part of the variation within treatment in the ANOVA model can be explained and accounted for by the linear relationship between test score and age. In this situation one-way ANCOVA is used to evaluate the relative effectiveness of the treatments while controlling the effect of the covariate (age). This is done by separating the relationship between the test score and age from the effect of treatments [10]. Assuming the equality of the slopes of the regression between the test score (the dependent variable) and covariate within the groups, parallel lines

$$Y_j = \bar{y_j} + b * (x - \bar{x_j})$$

are fitted for each treatment group, where Y_j is the test score predicted by the regression line in group j, $\bar{x_j}$ is the mean of covariate values in group j

and b is the common slope given by [1] :

$$b = \frac{\sum_j S_{xyj}}{\sum_j S_{xxj}}$$

where $S_{xxj} = \sum_i^{n_j} (x_{ij} - \bar{x_j})^2$ is the sum of squared deviations of the covariate values about the treatment mean in group j, $S_{xyj} = \sum_i^{n_j} (x_{ij} - \bar{x_j})(y_{ij} - \bar{y_j})$ represents the sum of the products of deviations of the test score (dependent variable) and the covariate about their corresponding jth treatment mean and n_j stands for the number of participants in treatment group j.
In the one-way ANCOVA model, the observed test scores can be written in the following fixed-effect model from [10]:

$$y_{ij} = \mu + a_j + b(x_{ij} - \bar{x}) + e_{ij}$$

where y_{ij} is the test score of ith participant in group j, μ is the grand mean of test scores, a_j is the jth treatment effect, b is the common regression slope, x_{ij} is the covariate value for ith participant in group j and $\bar{x}$ is the grand mean of covariate values. The total sum of squares in ANCOVA contains variability in the test scores due to different sources: variation due to differences between treatments that is independent of the covariate, variability that can be predicted from the linear relationship between the test score and the covariate and finally the unexplained variation (i.e. error). The Ancova summary table contains the sum of squares due to treatment differences independent from the covariate, the within treatment sum of squares resulting from the relationship between the test score and the covariate. And finally subtracting the within treatment sum of squares predictable from the covariate (within group regression sum of squares) from the total within group sum of squares, we obtain the within group residual sum of squares (error sum of squares). The main difference between ANOVA and ANCOVA is that the error term in ANOVA is based on the variation of the test scores around the treatment mean in each group while in ANCOVA the unexplained variation is calculated based on the deviation of the test scores from the regression line in each group [16].

Following Bland-Altman's technique, the proportion of the variation in one biomarker that is predictable from the variation in the other biomarker is given by the ratio between the within patient regression sum of squares and the total within patient sum of squares. The square root of this value, which

is always between 0 and 1, gives us the magnitude of the intra-patient correlation coefficient (ie. the linear relationship between the two biomarkers) and the sign of the correlation coefficient is determined by the overall regression slope. To compute the correlation coefficient of two markers using this technique, we can use multiple regression where we may choose one of the markers to be the outcome variable and the other marker and the patient's identity to be the predictor variables. [3].

It should be noted that although homogeneity of the regression slopes is an assumption in the ANCOVA model, it is not an assumption in the technique that we use in order to calculate the correlation coefficients. As mentioned before, the within patient residual sum of squares is the sum of the squared deviations of the observations in each group from the regression line. An increase in the heterogeneity of the slopes will therefore result in a larger intra- patient (within-patient) residual sum of squares and hence a smaller correlation coefficient in absolute value. A close to zero correlation coefficient indicates that there is a large variation in the linear relationship between the two biomarkers across the patients [2]. The magnitude of the correlation coefficient can also be affected by the presence of outliers as they can lead to heterogeneity of regression slope.

To assess the pairwise intra-patient correlation coefficient of the biomarkers, only patients with at least two ROIs of the same sample type were included in the calculations.

2.2.3 Inter-Patient Correlation

Taking into account the fact that the data set contains more than one sample of the same sample type from some patients, the pairwise inter-patient (i.e. between- patient) correlation coefficients of the biomarkers can be calculated as the correlation between the patient means [4]. This approach can be viewed as an assessment of the linear association between the averages of a pair of biomarkers in patients.

The number of ROIs from each sample type vary a lot among the patients (between 1 to 6 ROIs for sample types INFILT and TLS and from 1 to 12 ROIs for sample type STROMA). Therefore the weighted correlation coeffi-

cients are given by the formula:

$$r = \frac{\sum_{i=1}^{N} w_i(\bar{x}_i\text{-}m_{\bar{x}})(\bar{y}_i\text{-}m_{\bar{y}})}{\sqrt{\sum_{1=1}^{N} w_i(\bar{x}_i\text{-}m_{\bar{x}})^2 \sum_{i=1}^{N} w_i(\bar{y}_i\text{-}m_{\bar{y}_i})^2}}$$

where the weight w_i denotes the number of ROIs for patient i for $i = 1, ..., N$, $\bar{x}_i$ and $\bar{y}_i$ represent the biomarker means for patient i and $m_{\bar{x}} = \frac{\sum_{i=1}^{N} w_i \bar{x}_i}{\sum_{i=1}^{N} w_i}$, $m_{\bar{y}} = \frac{\sum_{i=1}^{N} w_i \bar{y}_i}{\sum_{i=1}^{N} w_i}$ stand for the weighted means.

2.3 Heterogeneity

2.3.1 Evaluating intra-patient variability in markers

Scale estimators are used to measure the amount of variation in a sample. A nonnegative-valued function of the sample such that it is location invariant and scale equivariant qualifies as a scale estimator [8]. Standard deviation (SD), median absolute deviation from the sample median (MAD) and range are three such estimators. Standard deviation is the most widely used estimator of dispersion of the observations and is the minimum variance unbiased estimator of the variance for Gaussian data

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

where $\boldsymbol{x}$ is a vector of observations of size n and $\bar{x}$ represents the sample mean.The sample standard deviation is not a robust estimator as its value is greatly influenced by the outlying values in the sample. The median absolute deviation from the sample median is a robust estimator of scale defined as

$$MAD = C.\underset{i}{median} \mid x_i - M \mid$$

where M is the sample median and C a scale factor equal to 1.4826 for normally distributed X_i to ensure that MAD is a consistent estimator of the standard deviation.

Range defined as the difference between the maximum and minimum value of the sample is another scale estimator that is sensitive to outliers. The value of the range is determined by the two extreme values in the sample and as such, the specific values of the rest of the sample between those two extremes do not affect its value.

In order to assess the intra-patient (i.e. within-patient) variation of the marker expressions for the specific ROI types, patients with at least three observations in the corresponding region were selected. The amount of variability was subsequently calculated separately for each marker as the SD , MAD and range of the biomarker measurements for each patient.

2.3.2 A measure of intra-patient marker heterogeneity

As explained earlier in 2.2, the total variability is partitioned into variability within groups and variability between groups in the framework of one-way ANOVA

$$\sum_{i}^{N}(Y_{ij}-\bar{Y})^2=\sum_{j=1}^{p}n_j(\bar{Y_i}-\bar{Y})^2+\sum_{j=1}^{p}(n_j-1)s_j^2$$

where s_j^2 is the scores estimated variance in the jth group.

Dividing the sum of squares in the formula above by their corresponding degrees of freedom results in three variances [16]: one related to the total variability , one associated with the variation between the groups $MSB = \frac{SS_{between}}{p-1}$ and finally the one associated with the variability within groups $MSE = \frac{SS_{within}}{N-p}$. One-way ANOVA assumes that the populations all have the same variance. The mean squared error (MSE) is the pooled variance that produces an unbiased estimate of the variance within each group.

The mean squared error in the ANOVA table, with the patients identity as the independent variable (i.e. the groups) and biomarker values for each sample type as the dependent variable is used as a measure of heterogeneity of the markers within patients [12].

It should be noted that only patients with more than three ROIs of the same

sample type were included in the calculations concerning the evaluation and measurement of heterogeneity within patients.

Chapter 3

Result

3.1 Density Estimation

The kernel density estimates were plotted separately for each biomarker and ROI-type across all patients with varying smoothing parameter and a Gaussian kernel. Figure 3.1 and Figure 3.2 show two sets of examples of such estimates for two biomarkers. Figure 3.1 shows that the estimated density curves with various smoothing parameters suggest different explanations of the data. Either a bimodal or multimodal distribution in all three sampling regions or approximately normal curves. In the case of CD27, Figure 3.2 indicates that although the distribution in INFILT and STROMA regions may roughly be described by a normal curve, the distribution in TLS region is better described by a bimodal curve indicating a combination of two populations. Drawing any conclusions about the population based on the density estimations, one must also recognize the limitations imposed by the small number of ROIs of type TLS and the high patient dependency demonstrated in Figure 3.3 and figure 3.4.
The density plots of CD3 and CD27 (Figure 3.1 and Figure 3.2), also reveal higher levels of e.g. T-cells in TLS compared to STROMA and INFILT, and a general higher level of CD27 positive cells. CD27 is a co-stimulatory marker and a potential target for immunotherapy. This type of results indicates that we can use this approach to identify tentative targets in relation to

presence of specific immune niches. The high patient dependency that needs to be taken into account can be addressed by using linear mixed models to establish biomarker signatures while allowing for patient dependency [5]

The estimated kernel density plots suggested that the underlying distribution of most biomarker values are approximately normal. In order to further assess the normality of the marker values, we examined the normal Q-Q plots.

3.1.1 Normality assessment

the data had been apriori log2 transformed. Protein expression is heavily skewed in a linear scale. Log transformation makes the data symmetrical/normally distributed, which is appropriate when comparing expression ratios between different groups of samples. Normal Q-Q plots were used in order to asses the normality of the log2-transformed data. Since protein expression is heavily skewed in a linear scale, log transformation makes the distribution of the data closer to symmetrical (normal). Looking at each biomarker across all three sample types and within each region separately indicated that most samples were reasonably normal. Marker CD66b was one of the biomarkers that showed deviation from normality in all three regions (Figure 3.5).

3.2 Correlation levels

3.2.1 Pearson correlation coefficients

The between-patient pairwise scatterplots of the proteins ,across all sample types as well as within individual sample types, suggest that there is a linear relationship between the biomarker pairs and that the direction of this relationship is positive in a majority of biomarker pairs. Although the strength varies markedly among them. Figure 3.6 shows exemplary inter-patient pairwise scatterplots of six biomarkers across all sample types as well as each individual ROI type separately.

The inter-patient Pearson correlation coefficients of the biomarkers across all ROIs are presented in Figure 3.7a. Figures 3.7b, 3.7c and 3.7d show the pairwise correlations of the proteins for each ROI type. The boxplot in Figure 3.8 demonstrates a comparison between the absolute values of the correlation coefficients. These results show that the correlation level is highest among the biomarkers for ROI type TLS and lowest for ROI type STROMA. The absolute value of the correlation levels seem to have an approximately symmetric distribution in all three sample types. Highest interquartile range of the values is observed in sample type TLS, while the lowest interquartile range of the values is associated with sample type STROMA. It should be noted that this analysis is hampered by lower number of patients in the TLS group.

3.2.2 Weighted inter-patient correlation coefficients

The results of the computation of the weighted correlation coefficients of the biomarker pairs for each sample type as well as across all types are presented in figure 3.9. The boxplot in Figure 3.10 demonstrates a comparison between the magnitude of the correlation coefficients in the three groups. The boxplot shows that the median of the magnitude of inter-patient (i.e. between-patient) weighted correlation coefficients of biomarkers is highest in sample type TLS and lowest in sample type STROMA. The distribution of the values appear to be almost symmetric in all three sample types. The interquartile range of the magnitude of correlations is highest in sample types INFILT and TLS. The lower number of patients in TLS group should again be taken into account when drawing conclusions based on this result.

3.2.3 Inra-patient correlation coefficients

The results of the calculations of the intra-patient (i.e. within-patients) correlation coefficients, across all sample types and for the individual sample types, are shown in the correlation matrices in figure 3.11. The boxplot in figure 3.12 provides a visual comparison of the absolute values of the within-patient correlation coefficients for the three sample types. The boxplot shows that the median of the magnitude of intra-patient correlations is only slightly higher in sample type STROMA than sample type INFILT while the highest median of the correlation magnitudes is associated with sample type TLS. The range and the interquartile range of the values appear to be approximately equal for INFILT and STROMA groups while both range and interquartile range are higher for sample type TLS compared to the other two groups. Comparing this result to those obtained in 3.2.2 shows that the range of the magnitude of the inter-patient correlations is larger compared to the magnitude of the inter-patient correlations in sample types INFILT and STROMA. The median of the the magnitude of the inter-patient correlations are also higher than the magnitude of the intra-patient correlations in all three sample groups.

3.3 Heterogeneity

3.3.1 Assessing intra-patient biomarker variability

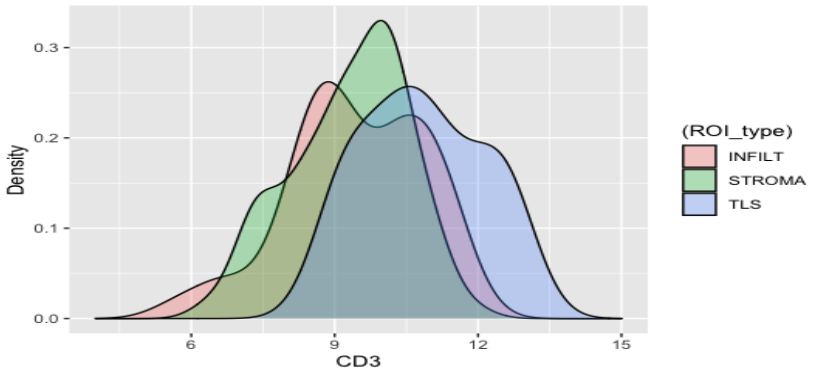
The boxplots of the variability estimates for each sample type, obtained by computing the standard deviation of the biomarker values calculated individually for patients with at least three ROIs of the same sample type, are presented in figure 3.13. The figure shows that markers CD66b, VISTA, SMA and FOXP3 in sample type INFILT and markers CD66b and STING in sample type STROMA have a higher interquartile range compared to the other biomarkers. The highest median values belong to biomarker CD66b in INFILT group and to CD66b and STING in STROMA group. In the sample type TLS, the largest median value is observed for marker PD.L1 and largest interquartile range for Fibronectin.

Figure 3.14 shows the box plots of the variability measures, estimated by the median absolute deviation from the sample median (MAD) of the biomarker values for individual patients for each sample type. The figure shows that the variability in the heterogeneity measures assessed by the interquartile range of the value is highest for markers CD66b, VISTA and FOXP3 in INFILT group and for CD66b and STING in STROMA group. The largest median values of the heterogeneity measures are associated with marker CD66b in sample type INFILT and CD66b and STING in sample type STROMA. With respect to the sample type TLS, the largest median value is observed for marker PD.L1 and largest interquartile range for CD163.

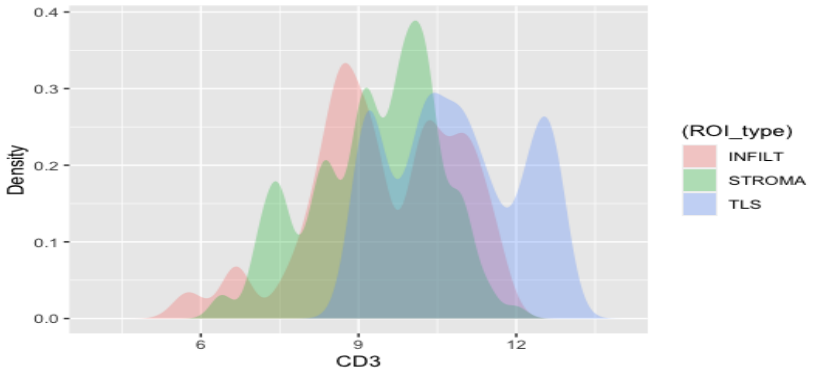
The variability in the biomarker values, estimated by the range of the biomarker measurements for individual patients and for each sample type are shown in figure 3.15. Markers FOXP3 and SMA have the largest interquartile range in sample type INFILT and marker CD66b in sample type STROMA. The highest median values belong to CD66b in both sample types INFILT and STROMA. With respect to the sample type TLS, the largest median value is observed for marker PD.L1 and largest interquartile range for Fibronectin.

3.3.2 Intra-patient heterogeneity measure

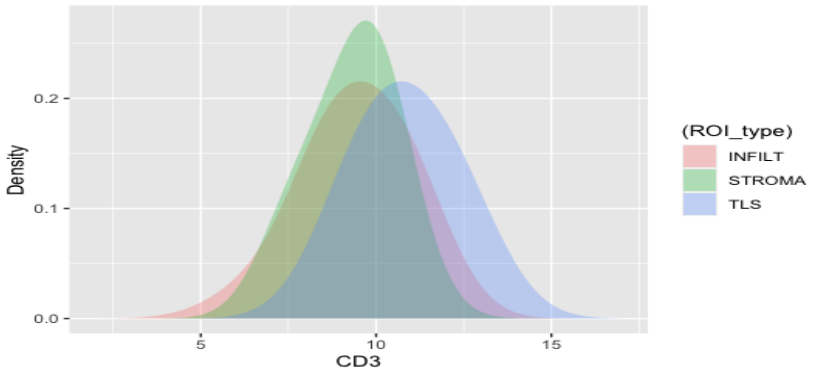
The mean square errors (MSEs) as a measure of within- patient heterogeneity were calculated separately for each marker and each sample type. The result of the computations are shown in Figure 3.16 and Appendix C. The highest values in the Figure 3.16 are associated with marker CD66b in sample types INFILT and STROMA. In sample type INFILT, the highest values belong to markers CD66b and SMA and in STROMA group to markers CD66b and STING. The lowest values are observed in TLS and the largest values in INFILT group.



(a) CD3 Kernel Density Estimation (Default)

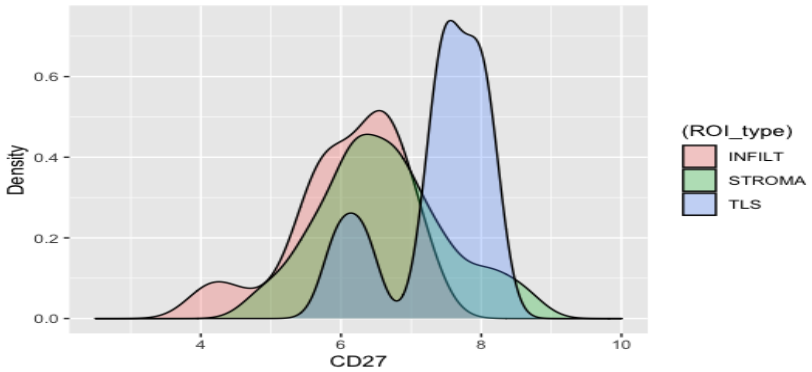


(b) CD3 Kernel Density Estimation (with decreased smoothness)

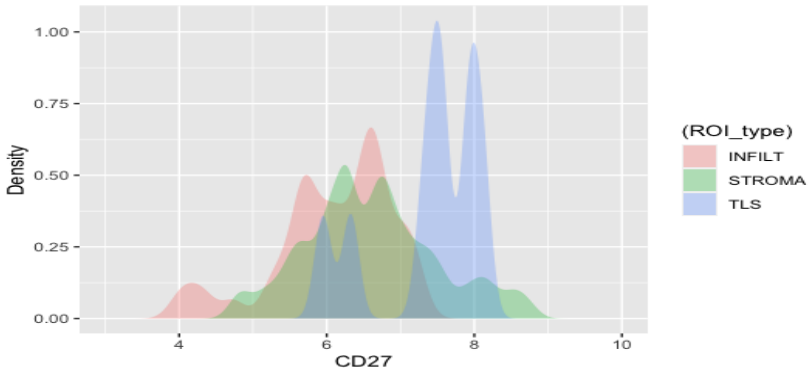


(c) CD3 Kernel Density Estimation (with increased smoothness)

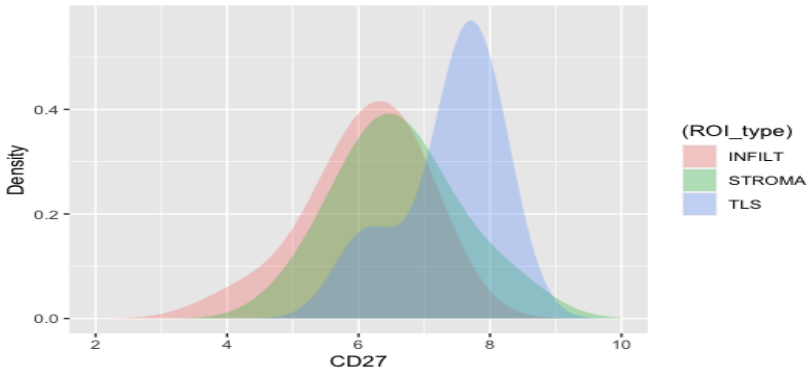
Figure 3.1: CD3 Kernel Density Estimations



(a) CD27 Kernel Density Estimation (Default)



(b) CD27 Kernel Density Estimation (with decreased smoothness)



(c) CD27 Kernel Density Estimation (with increased smoothness)

Figure 3.2: CD27 Kernel Density Estimations

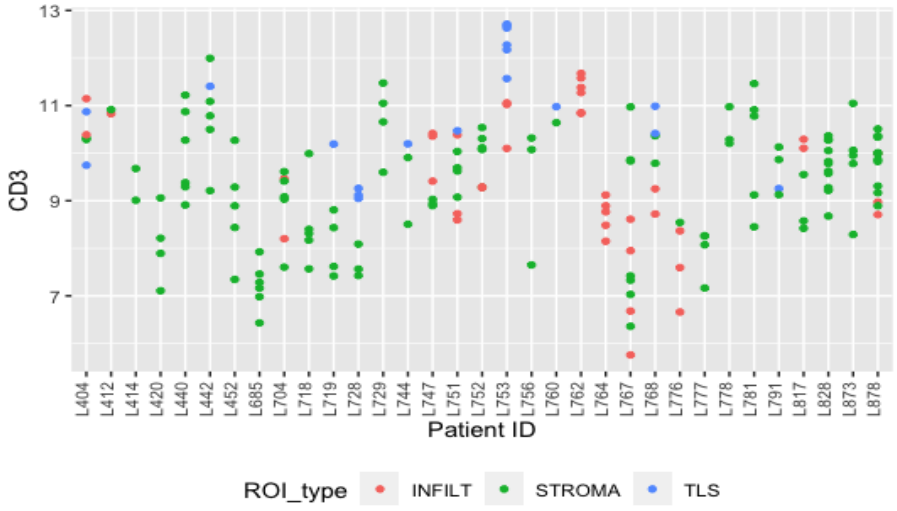


Figure 3.3: CD3 Scatterplot across all Patients

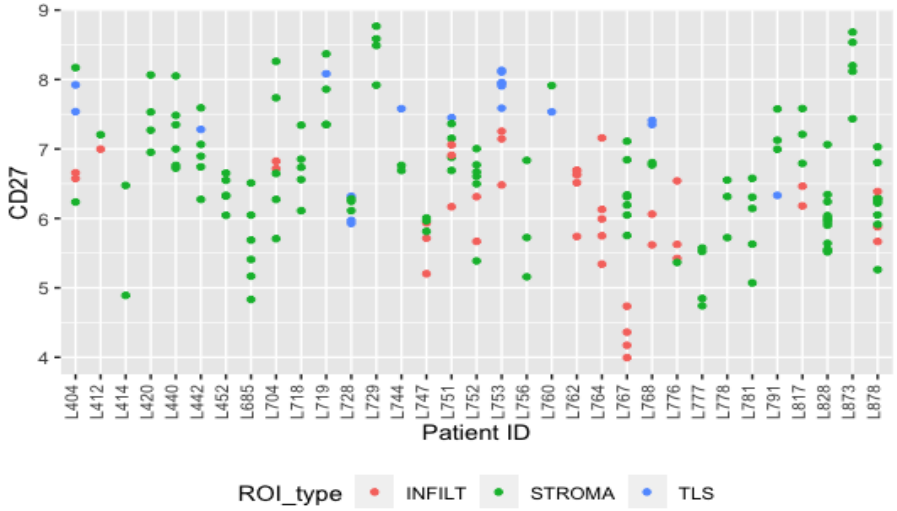
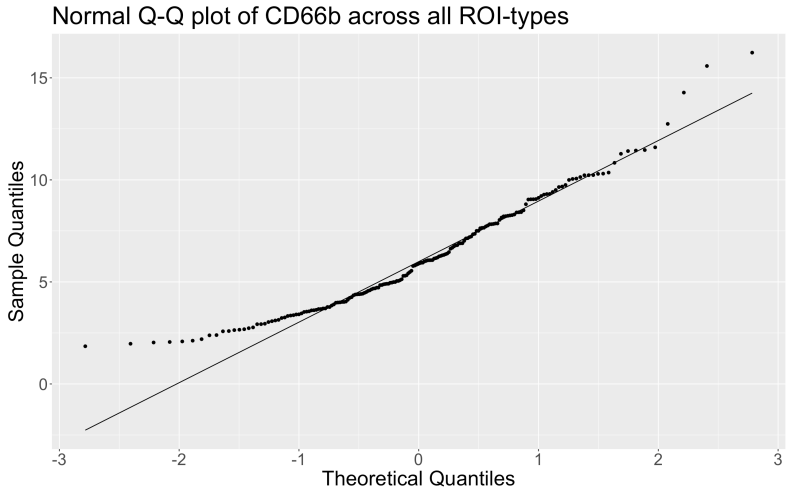
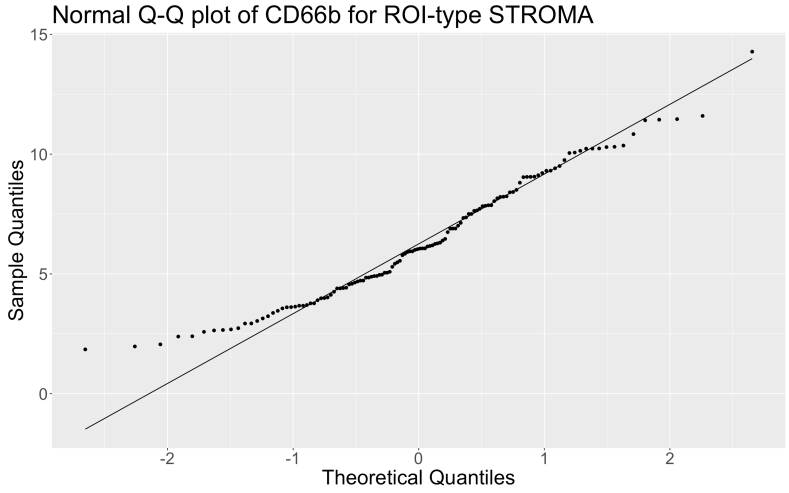


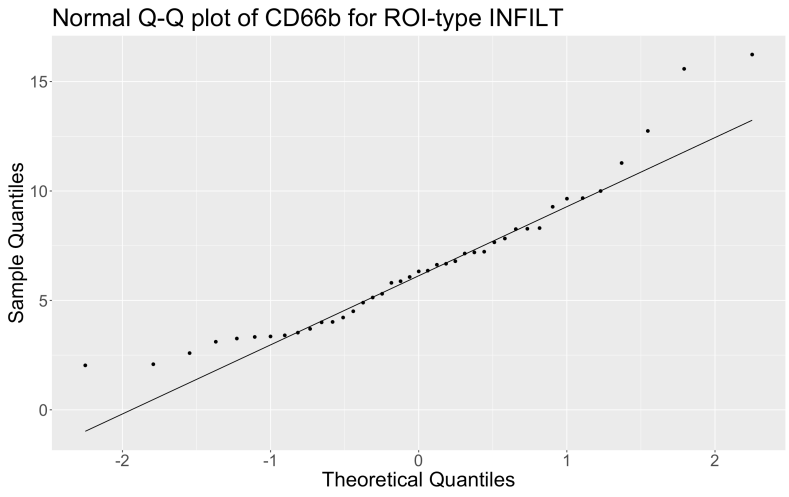
Figure 3.4: CD27 Scatterplot across all Patients



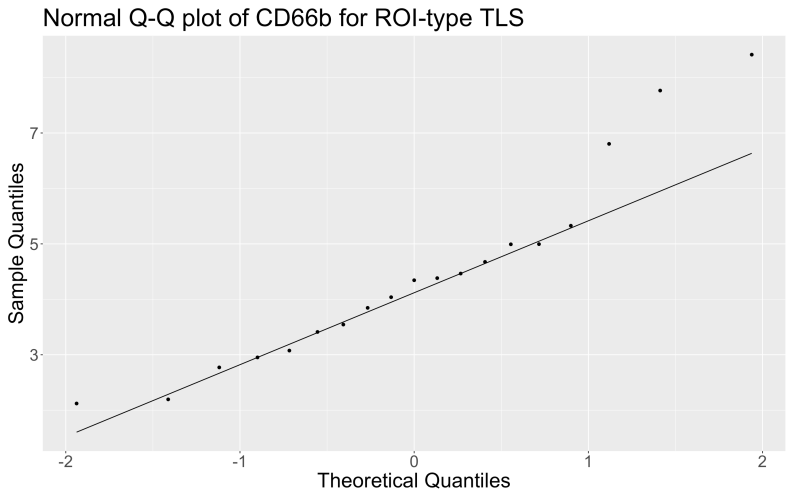
(a) Across all ROI-types



(b) ROI-type STROMA

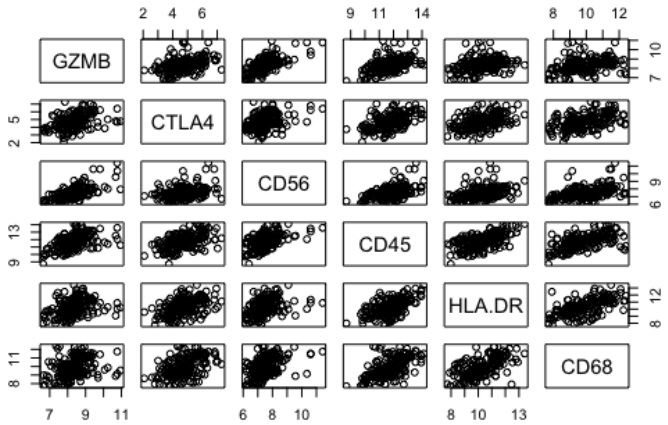


(c) ROI-type INFILT

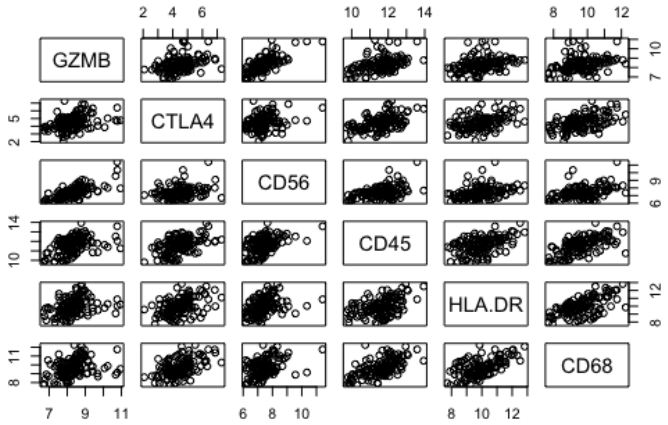


(d) ROI-type TLS

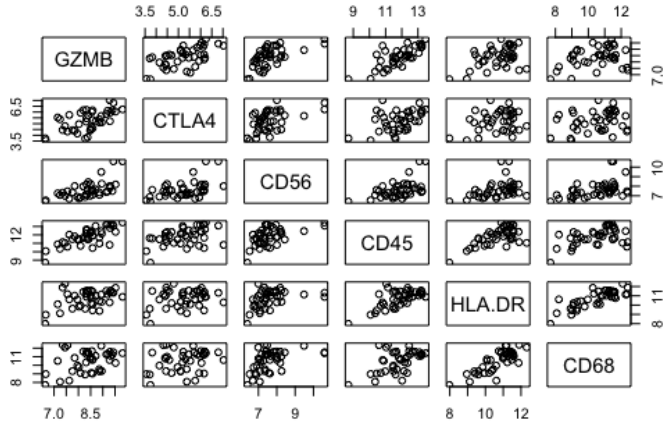
Figure 3.5: Normal Q-Q plots of CD66b: Marker CD66b is one of the markers that shows deviation from normality in all three regions.



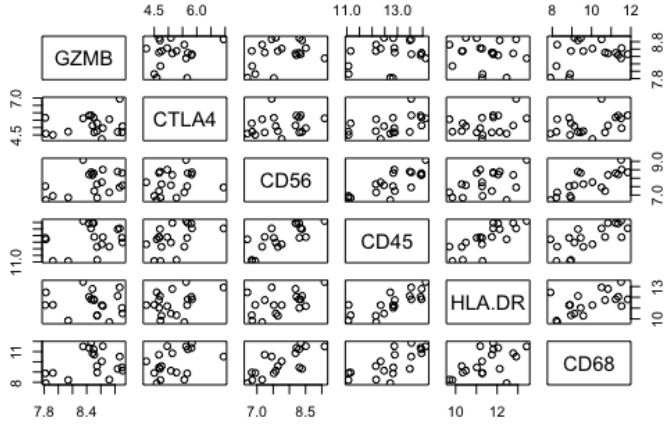
(a) Across all sample types



(b) Sample type STROMA

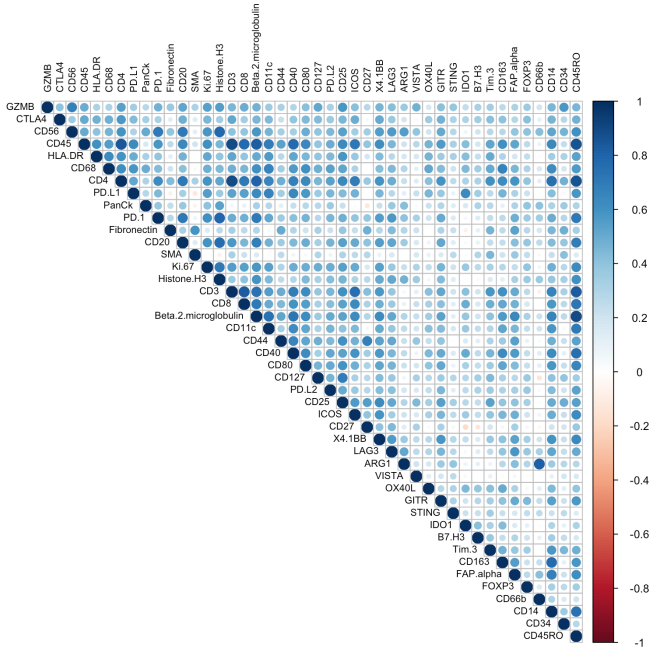


(c) Sample type INFILT

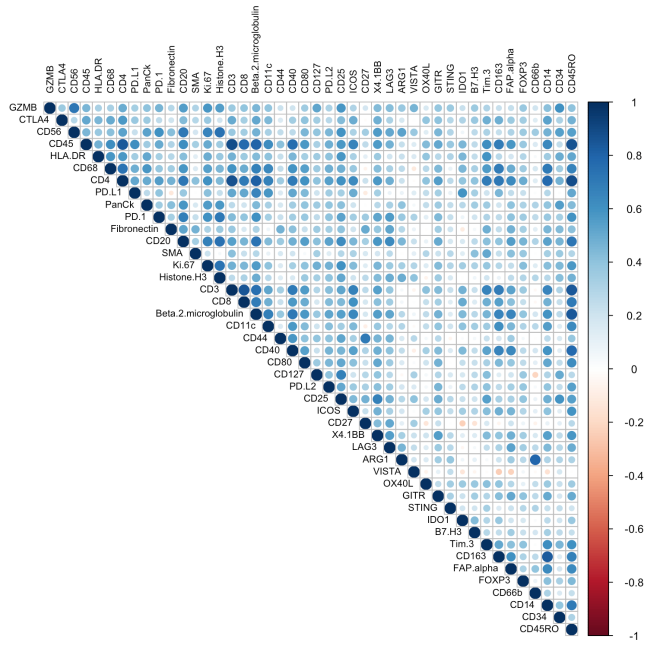


(d) Sample type TLS

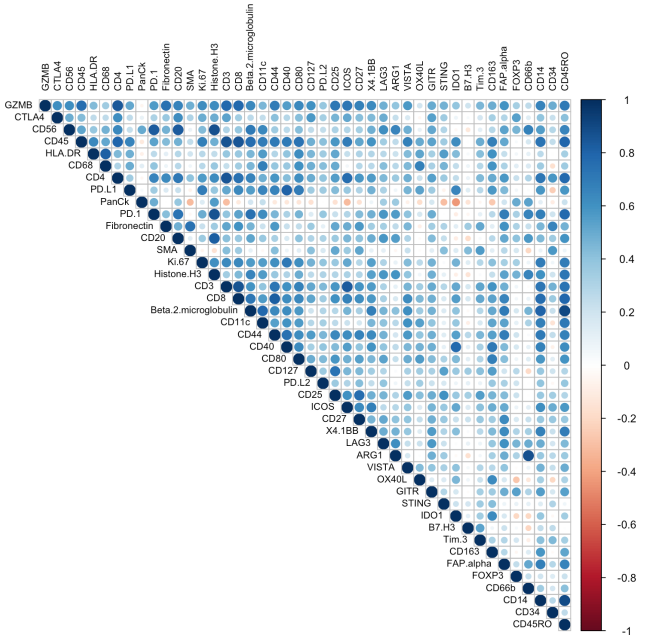
Figure 3.6: Inter-patient pairwise scatterplots of six biomarkers



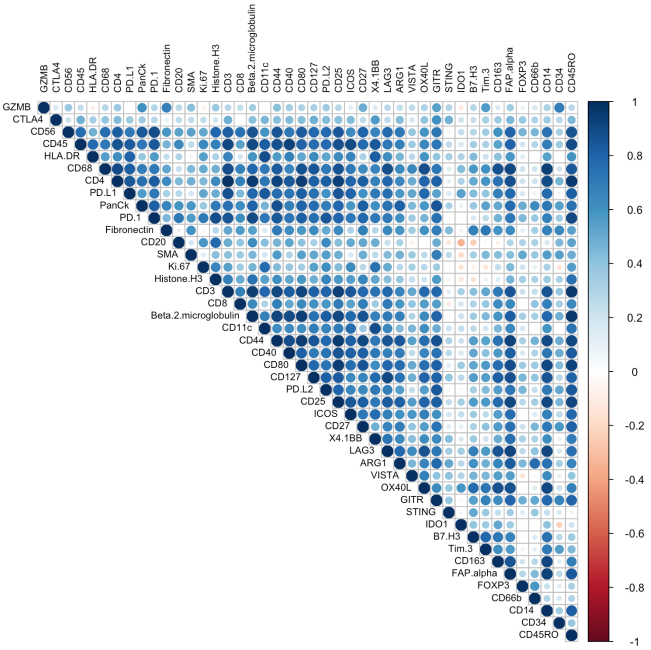
(a) Upper triangle of the Pearson correlation matrix across all sample types



(b) Upper triangle of the Pearson correlation matrix for sample type STROMA



(c) Upper triangle of the Pearson correlation matrix for sample type INFILT



(d) Upper triangle of the Pearson correlation matrix for sample type TLS

Figure 3.7: —Upper triangles of inter-patient Pearson correlation matrices of biomarkers

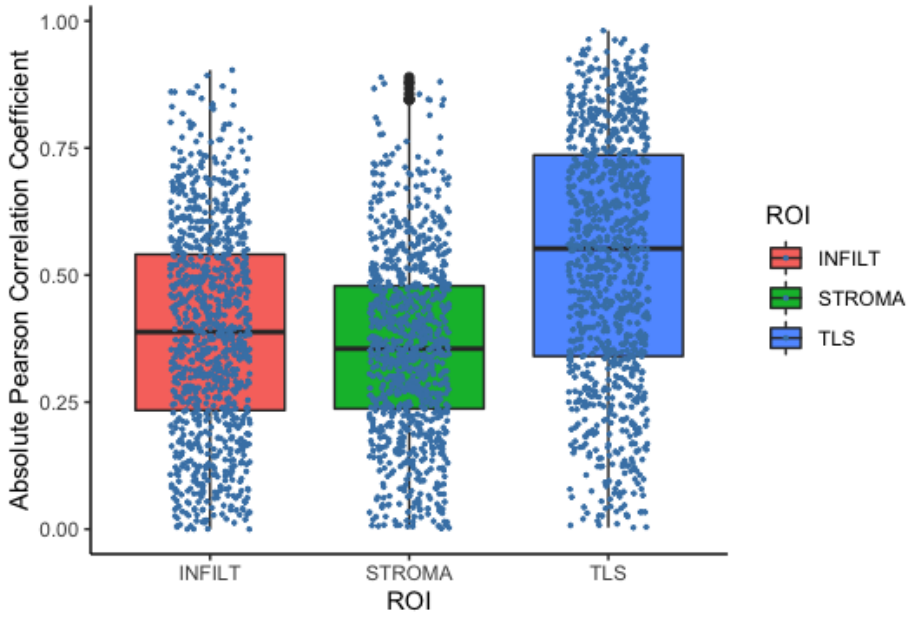
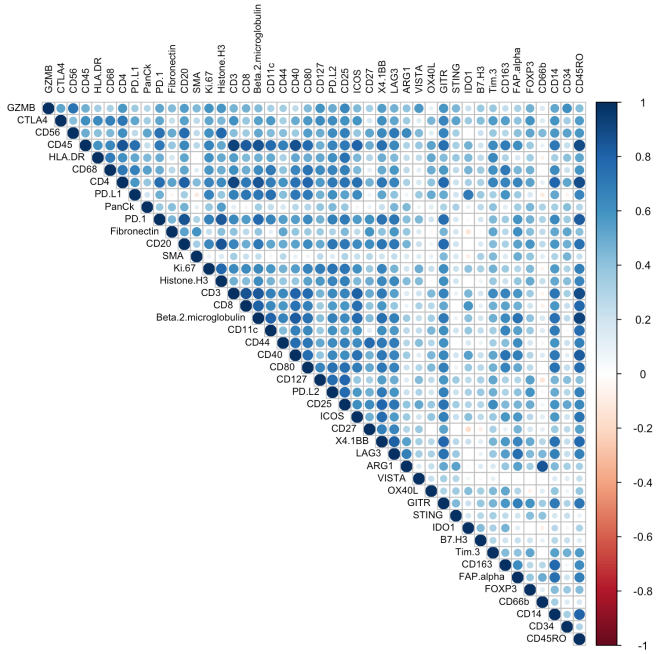
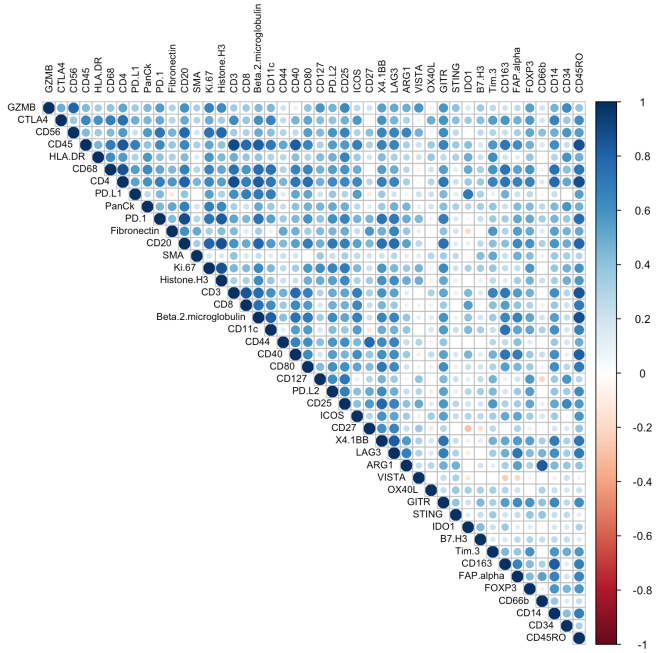


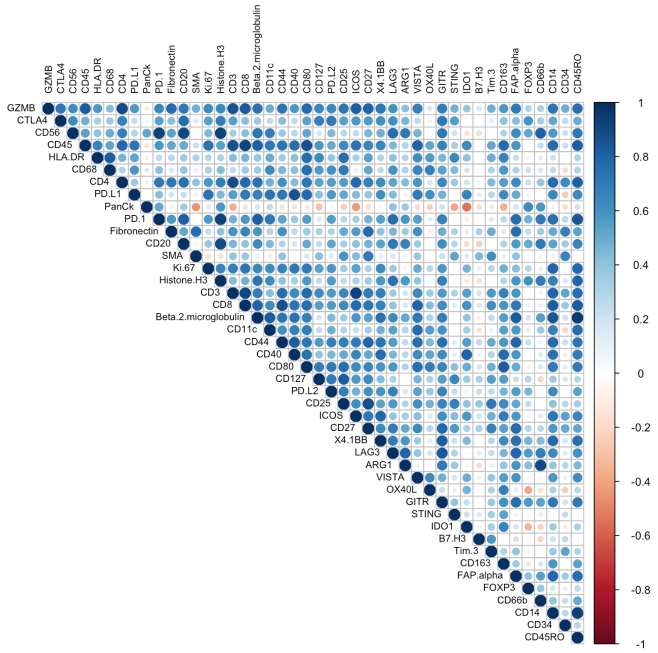
Figure 3.8: Boxplot of Absolute Pearson Correlation coefficients for STROMA, INFILT and TLS sample types: The figure indicates that the strength of biomarker correlations is highest in sample type TLS and lowest in STROMA. The absolute value of the correlation levels seem to have an approximately symmetric distribution in all three sample types. Highest interquartile range of the values is observed in sample type TLS, while the lowest interquartile range of the values is associated with sample type STROMA.



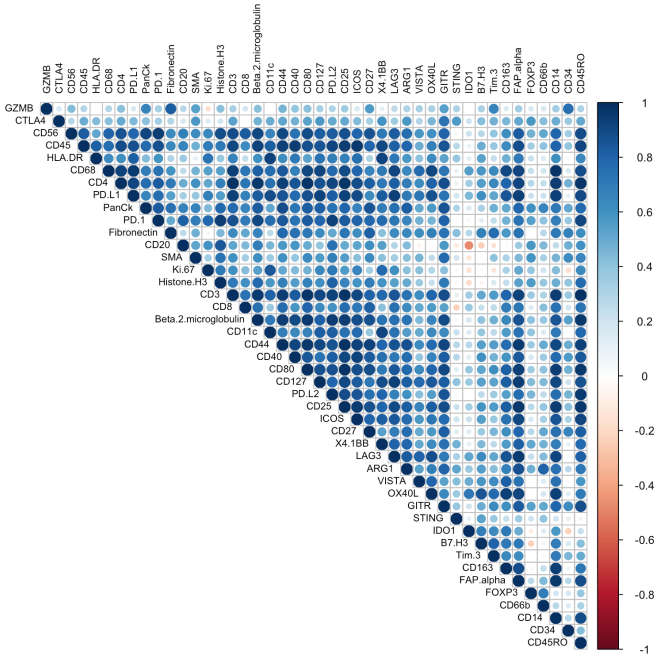
(a) Upper triangle of weighted inter-patient correlation matrix across all sample types



(b) Upper triangle of weighted inter-patient correlation matrix for sample type STROMA



(c) Upper triangle of weighted inter-patient correlation matrix for sample type INFILT



(d) Upper triangle of weighted inter-patient correlation matrix for sample type TLS

Figure 3.9: —Upper triangles of weighted inter-patient correlation matrices of biomarkers

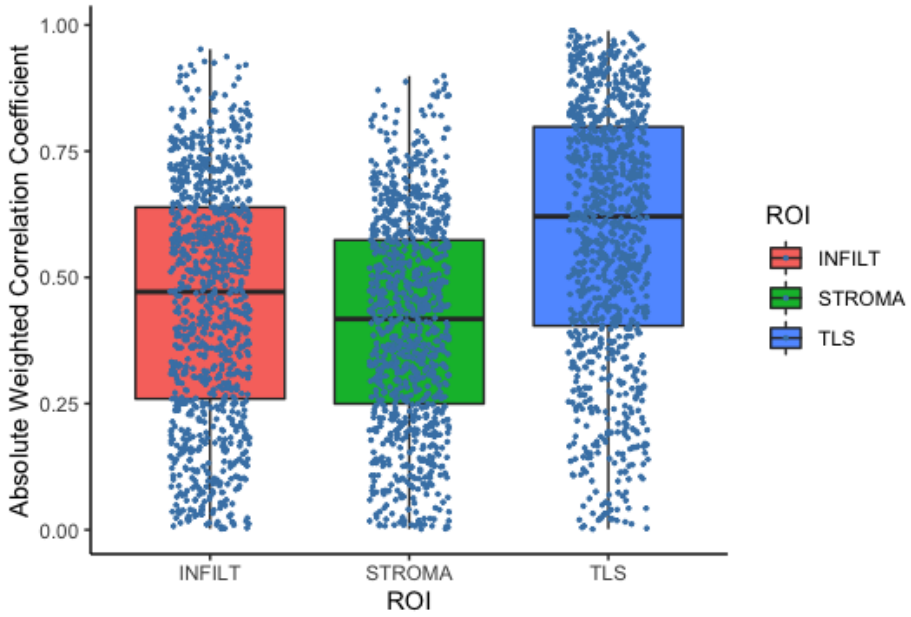
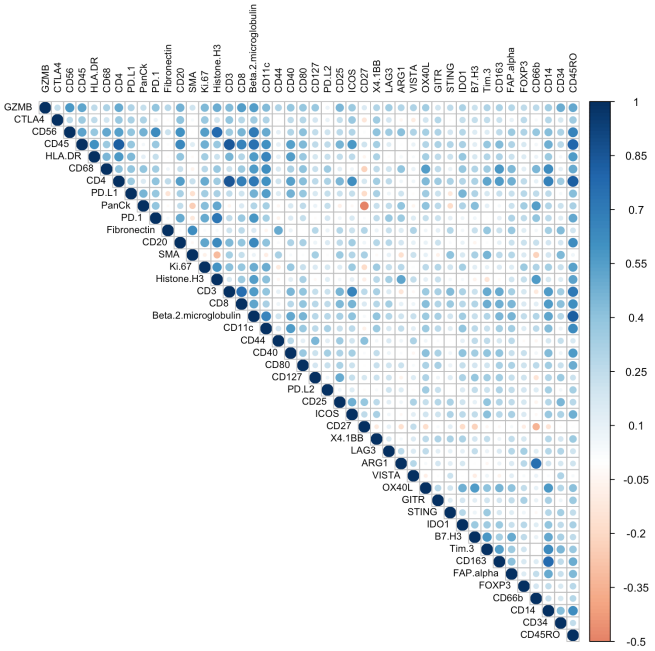
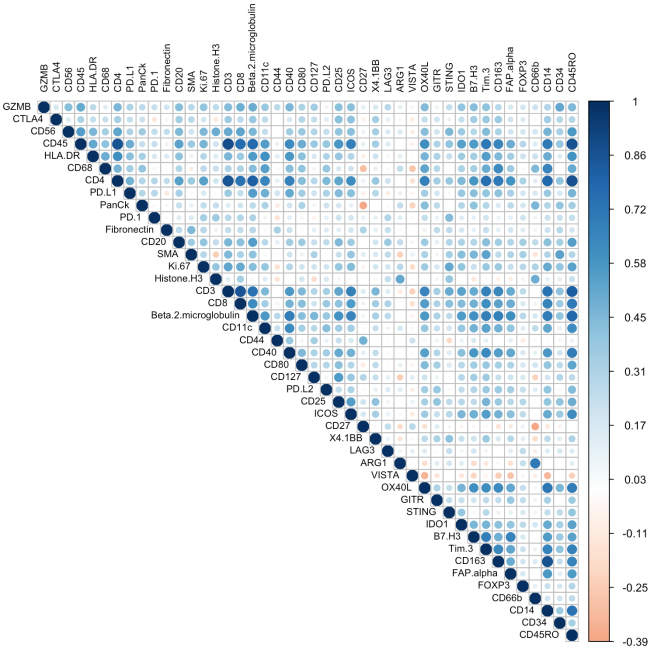


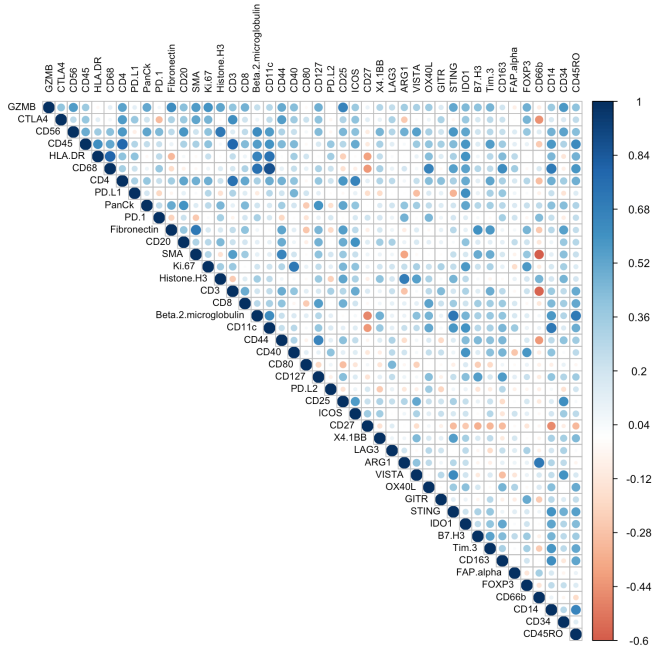
Figure 3.10: Boxplot of absolute weighted correlation coefficients for three sample types: The boxplot indicates that the median of the magnitude of inter-patient weighted correlations of biomarkers is highest in sample type TLS and lowest in sample type STROMA. The distribution of the values seem to be almost symmetric in all three sample types. The interquartile range of the magnitude of correlations is highest in sample types INFILT and TLS.



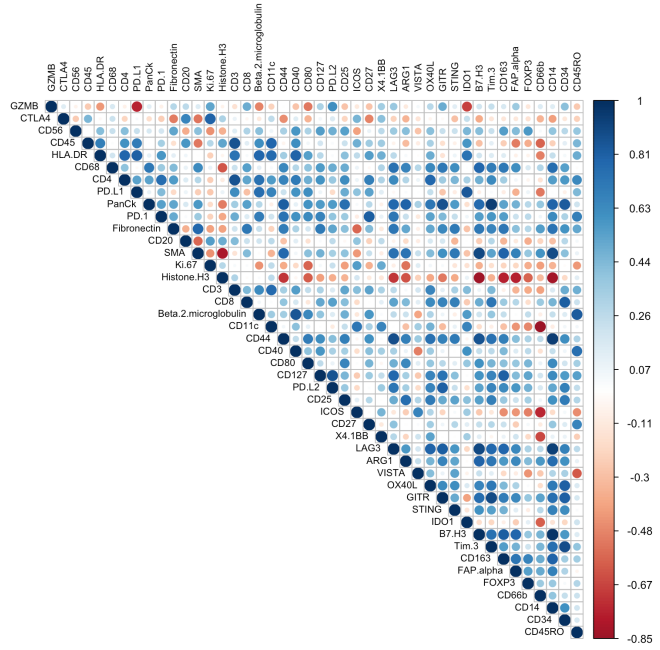
(a) Upper triangle of intra-patient correlation matrix across all sample types



(b) Upper triangle of intra-patient correlation matrix for sample type STROMA



(c) Upper triangle of intra-patient correlation matrix for sample type INFILT



(d) Upper triangle of intra-patient correlation matrix for sample type TLS

Figure 3.11: Upper triangle of intra-patient correlation matrices of biomarkers

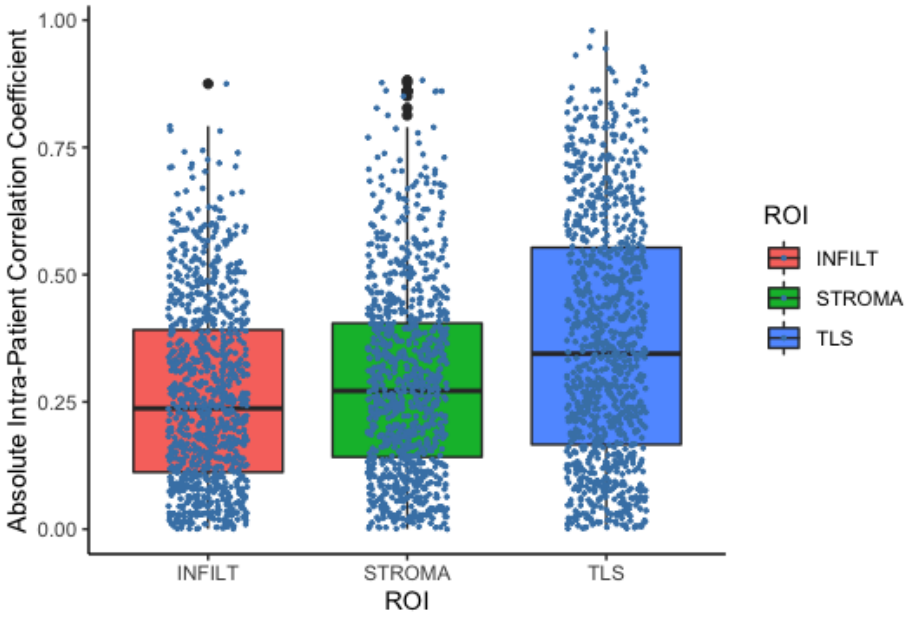


Figure 3.12: Boxplot of absolute intra-patient biomarker correlation coefficients for the three sample types: The boxplot suggests that the median of the magnitude of intra-patient correlations is only slightly higher in sample type STROMA than sample type INFILT while the highest median of the correlation magnitudes is associated with sample type TLS. The range and the interquartile range of the values appear to be approximately equal for INFILT and STROMA groups while both range and interquartile range are higher for sample type TLS compared to the other two groups.

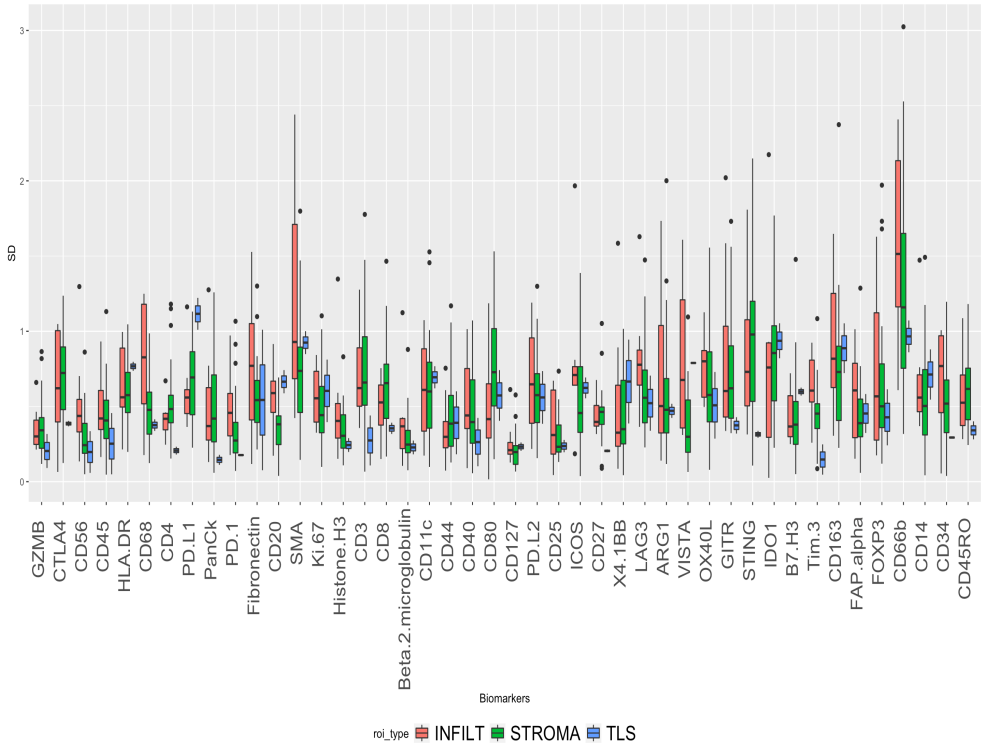


Figure 3.13: Intra-patient biomarker heterogeneity measured as the SD of marker values for individual patients: Markers CD66b, VISTA, SMA and FOXP3 have the largest interquartile range in INFILT group and markers CD66b and STING in sample type STROMA. The highest median values belongs to CD66b in INFILT and to CD66b and STING in sample type STROMA. In the sample type TLS, the largest median value is observed for marker PD.L1 and largest interquartile range for Fibronectin.

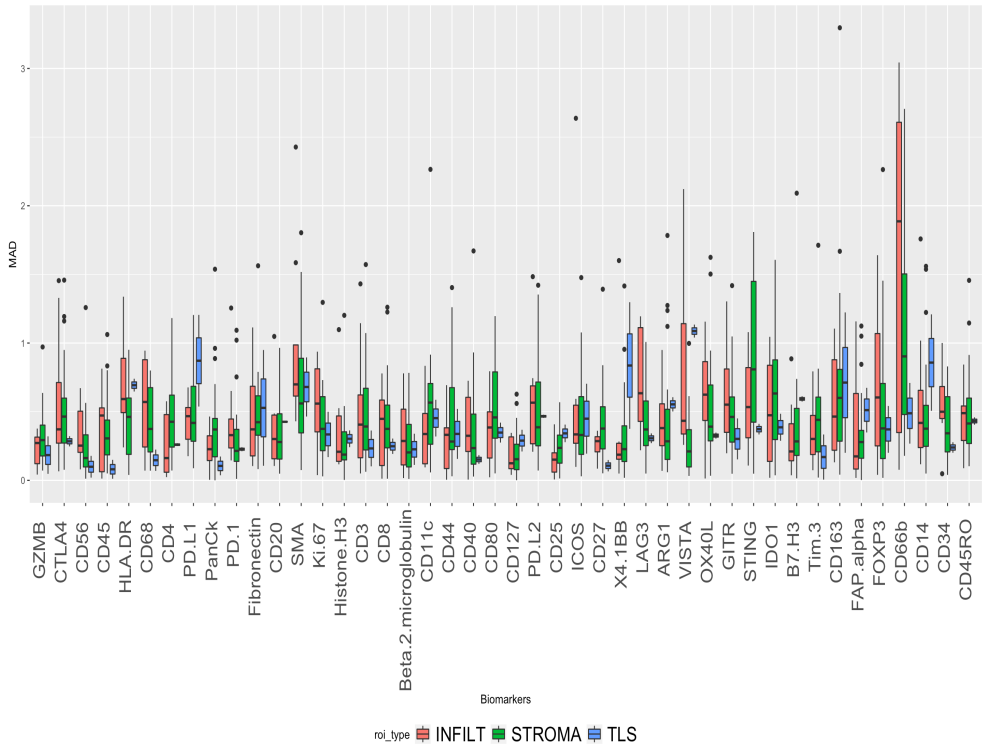


Figure 3.14: Intra-patient biomarker heterogeneity measured as the MAD of marker values for individual patients: Markers CD66b, VISTA and FOXP3 have the largest interquartile range in INFILT group and markers CD66b and STING in sample type STROMA. The highest median values belongs to CD66b in INFILT and to CD66b and STING in sample type STROMA. With respect to the sample type TLS, the largest median value is observed for marker PD.L1 and largest interquartile range for CD163.

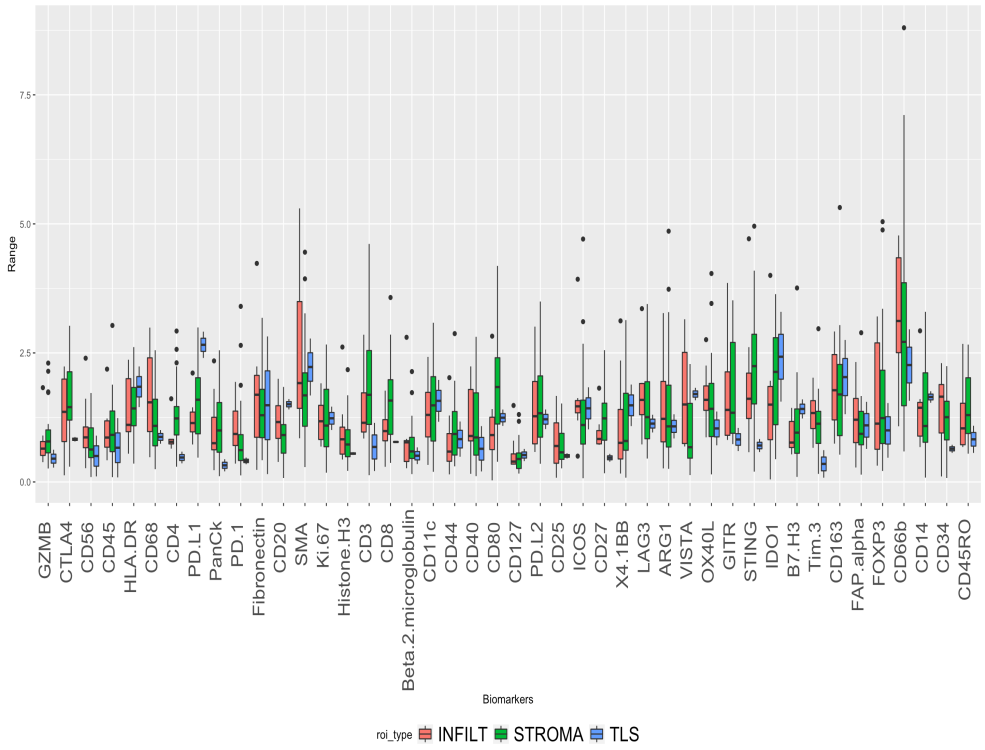


Figure 3.15: Intra-patient biomarker heterogeneity measured as the range of marker values for individual patients: Markers FOXP3 and SMA have the largest interquartile range in INFILT group and marker CD66b in sample type STROMA. The highest median values belongs to CD66b in INFILT and in sample type STROMA. With respect to the sample type TLS, the largest median value is observed for marker PD.L1 and largest interquartile range for Fibronectin.

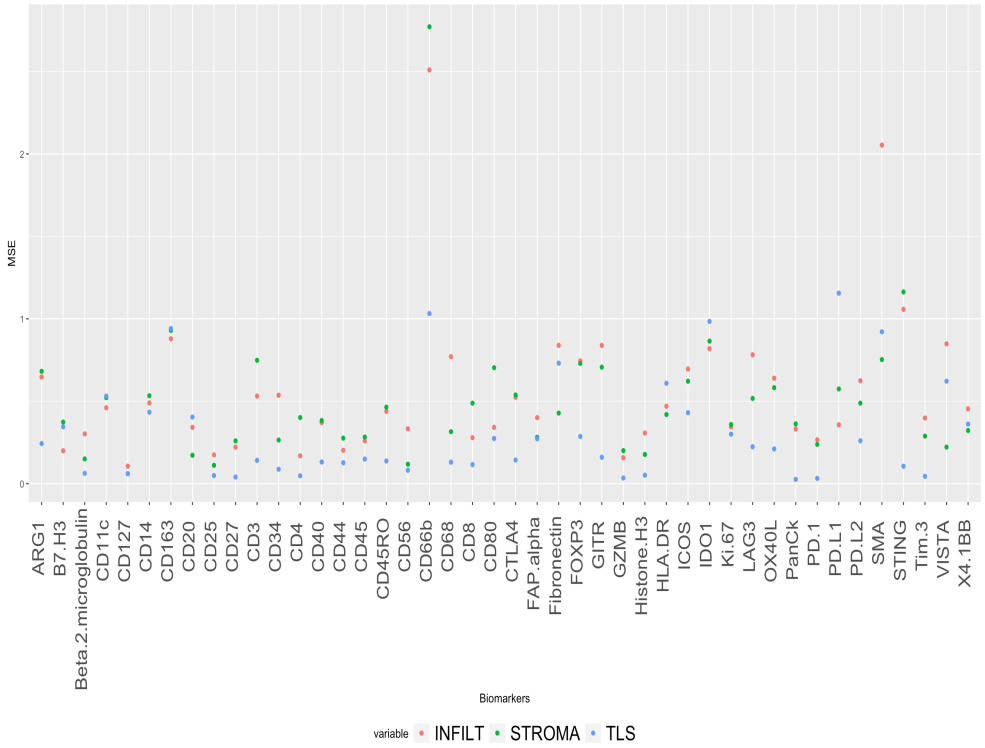


Figure 3.16: Intra-patient heterogeneity measured as MSE: The highest values in the plot are associated with marker CD66b in sample types INFILT and STROMA. In sample type INFILT, the highest values belong to markers CD66b and SMA and in STROMA group to markers CD66b and STING. The lowest values are observed in TLS and the largest values in INFILT group.

Chapter 4

Discussion and future work

Comparison the results of the inter-patient (i.e. between- patient) correlation coefficients to the intra-patient (within-patient) correlations coefficients in 3.2 showed that while the median of absolute value of the inter-patient correlation coefficients is higher among the biomarkers for ROI type INFILT compared to ROI type STROMA, the intra-patient median of the absolute value of correlation coefficients is higher for sample type STROMA than for sample type INFILT. In both case the median of the magnitude of correlation coefficients in sample type TLS was the largest compared to the other two sample types. We can also observe an overall reduction in the strength of the pairwise intra-patient correlations between the markers compared to the inter-patient correlations in 3.2. This difference may be due to the fact that we are dealing with heterogeneous subsamples. This point can be demonstrated with an example [7] [9]. Assessing the relationship between height and weight, using combined data from male and female participants, shows a strong linear relationship between the two variables. The strength of this relationship is substantially reduced if the correlation coefficient is calculated separately for the men and women in the sample. This result is explained by the fact that men are on average taller and heavier than women. By pooling the observations from two group with different means and different variation about the mean, the relationship between the two variables is enhanced compared to the within groups estimates due to the presence of a hidden variable which is sex in this example. In our analyses, differences in

age and stage of the decease between the patients for example, may give rise to heterogeneous subsamples. Calculating the correlation coefficients based on the pooled data may consequently lead to misleading results due to the presence of those neglected hidden variables. This may still be the case even if the correlation between two variables is the same within the subsamples [7]. The impact of the subgroups can be further examined in the future with a larger data set.

Although a thorough investigation of the differences between the strength of the pairwise correlation of the biomarkers was outside the scope and aim of this project, the following points should be noted. Firstly, using significance and insignificance as a criterion for comparison of effect estimates is inappropriate [6]. The fact that the correlation between a pair of biomarkers may be statistically insignificant in one sample type while it is significant with respect to another sample type, does not in itself imply that the correlations are different. Furthermore, the number of samples from sample type TLS was small compared to ROI types INFILT and STROMA and there are three patients with at least two samples of type TLS compared to 27 and 13 patients in STROMA group and INFILT group respectively. This suggests that the result obtained for sample type TLS is highly patient dependent and should be further investigated in future studies based on a larger sample size. Also, when investigating the difference between the pairwise correlation coefficients in sample types STROMA and INFILT, it must be noted that some patients have samples from both sample types and therefore it can be misleading to view the samples and the correlation coefficients as independent.

The intra-patient variability in protein markers was assessed using three different estimators, standard deviation, median absolute deviation from median and range. Figures 3.13, 3.14 and 3.15 all indicate that the level of heterogeneity varies between spatial immune infiltration niches and between the different biomarkers. This highlights the added value in sampling multiple regions of different types from the same tumor. Range seems to be a wasteful approach as it restricts the data to its maximum and minimum values. One of the purposes of the efforts to quantify biomarker heterogeneity is to explore the predictive value of variation such as predicting clinical outcome [see 12]. Although the robustness of MAD in the presence of outlying values is often viewed as an advantage, reducing the influence of the outliers and downweighting of the tail is likely not beneficial for our purpose as the

tails contain most of the information that is of interest to us. This of course does not imply that higher heterogeneity is associated with higher predictive power. It should also be noted that some of the patients have only 3 samples of a specific sample type. For a sample of size 3 it is impossible to have a outlier-robust scale estimator as we can not at the same time protect against both the estimator becoming arbitrarily large (explosion) and becoming arbitrarily close to zero (implosion) [14]. Rousseeuw & Verboven (2002) [14] suggest to use average distance to the median if implosion is to be avoided and to use MAD if we are protecting against explosion. For small samples of size $n \geq 4$, they recommend to use MAD and more advanced robust scale estimators that use MAD and median in their construction.

There are similarities between the result obtained by using standard deviation and median absolute deviation presented in Figures 3.13 and 3.14. However, the similarities can be viewed as alarming. For instance high levels of heterogeneity and variability in heterogeneity are associated with marker CD66b in sample type INFILT. Although it is not possible to estimate the density of the biomarkers at the individual patient level, an examination of the kernel density estimate for the data across all patients shows a skewed estimated density for this biomarker in sample type INFILT (Appendix D). For sample type STROMA, high levels of heterogeneity and variability in heterogeneity measures are associated with markers CD66b and STING according to both MAD and SD estimates. The estimated kernel density plots and normal Q-Q plots (see Appendix D and 3.5b) show skewed estimated distributions of these biomarkers and suggest deviation from normality. As noted in 3.1.1, biomarker CD66b was one of the markers that in all sample types showed most deviation from normality compared to other markers. Using MAD as well as sample SD as a measure of variability is suitable for symmetric distributions in the view of the fact that in their calculation, negative and positive deviations from a central position (median vs. mean) are treated as equally important. Therefore, although MAD is a more robust estimator of scale in the presence of outliers compared to SD, it is not an appropriate measure of scale when the distribution is highly skewed [13]. It is therefore possible that part of the results obtained here are due to use of scale estimators that are sensitive to the skewness of the distribution. Rousseeuw & Croux (1993) [13] propose alternatives to the MAD that are not sensitive to asymmetric distributions. In future studies, comparing the result obtained here to those obtained by using such robust scale estimators that are not

sensitive to asymmetric distributions can be enlightening to determine the degree to which the result is affected by the skewness of the distributions.

In conclusion, this work is a preliminary analysis of this type of data. The predictive value of heterogeneity measures discussed in this work can be further investigated in the future studies and their predictive power can be compared.

Appendix A

Biomarkers

GZMB	CTLA4	CD56	CD45	HLA.DR
CD68	CD4	PD.L1	PanCk	PD.1
Fibronectin	CD20	SMA	Ki.67	Histone.H3
CD3	CD8	Beta.2.microglobulin	CD11c	CD44
CD40	CD80	CD127	PD.L2	CD25
ICOS	CD27	X4.1BB	LAG3	ARG1
VISTA	OX40L	GITR	STING	IDO1
B7.H3	Tim.3	CD163	FAP.alpha	FOXP3
CD66b	CD14	CD34	CD45RO	

Table A.1: Protein biomarkers measured with the DSP

Appendix B

Normalization of data

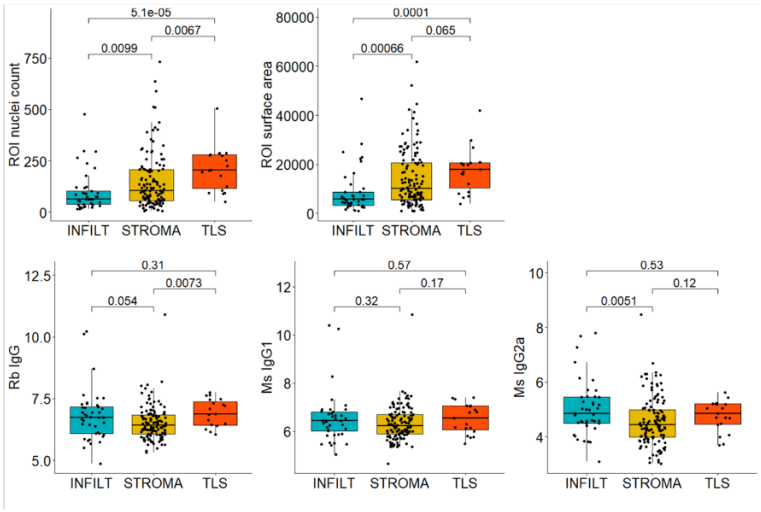


Figure B.1: Normalization of data with respect to differences in cell numbers, sampled area and background noise: The figures shows that significant differences in number of cells and areas sampled are not reflected in the normalized data, as shown for the three negative control antibodies

Appendix C

Biomarker intra-patient heterogeneity measures (MSE)

Biomarker	INFILT	STROMA	TLS
GZMB	0.157	0.201	0.035
CTLA4	0.523	0.538	0.143
CD56	0.333	0.118	0.082
CD45	0.260	0.283	0.149
HLA.DR	0.470	0.420	0.609
CD68	0.771	0.315	0.131
CD4	0.169	0.401	0.048
PD.L1	0.357	0.575	1.155
PanCk	0.332	0.363	0.026
PD.1	0.266	0.238	0.032
Fibronectin	0.839	0.428	0.731
CD20	0.342	0.172	0.404
SMA	2.054	0.753	0.921
Ki.67	0.345	0.359	0.300
Histone.H3	0.307	0.177	0.052
CD3	0.531	0.748	0.142
CD8	0.279	0.488	0.116
Beta.2.microglobulin	0.302	0.150	0.063

CD11c	0.460	0.522	0.531
CD44	0.203	0.276	0.127
CD40	0.371	0.383	0.131
CD80	0.342	0.703	0.274
CD127	0.106	0.061	0.061
PD.L2	0.624	0.488	0.260
CD25	0.175	0.111	0.049
ICOS	0.696	0.621	0.431
CD27	0.222	0.259	0.041
X4.1BB	0.454	0.322	0.362
LAG3	0.782	0.517	0.224
ARG1	0.647	0.682	0.243
VISTA	0.848	0.222	0.621
OX40L	0.639	0.582	0.211
GITR	0.839	0.707	0.161
STING	1.057	1.163	0.106
IDO1	0.819	0.865	0.984
B7.H3	0.200	0.373	0.346
Tim.3	0.398	0.288	0.045
CD163	0.879	0.929	0.941
FAP.alpha	0.401	0.282	0.273
FOXP3	0.744	0.729	0.286
CD66b	2.509	2.771	1.032
CD14	0.490	0.533	0.434
CD34	0.536	0.265	0.088
CD45RO	0.438	0.464	0.138

Table C.1: MSE as a measure of intra-patient heterogeneity calculated for each biomarker separately for each sample type

Appendix D

CD66b & Sting: Skewness and deviation from normality

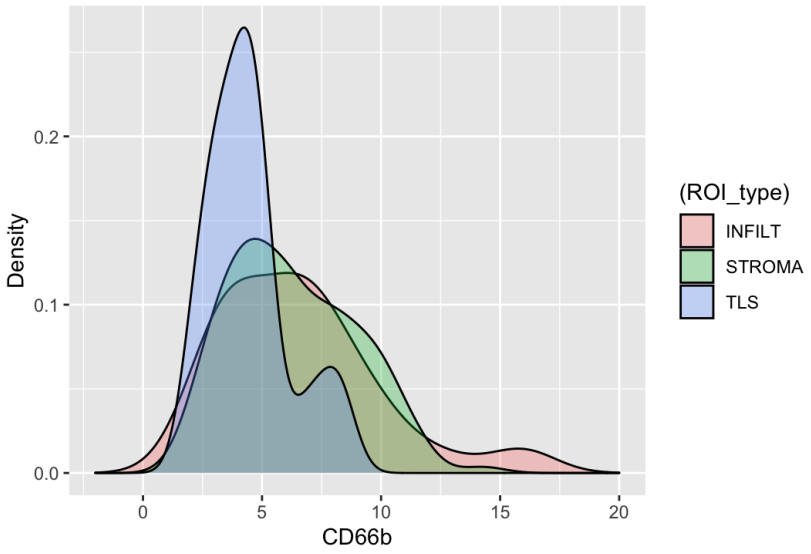


Figure D.1: CD66b Kernel Density Estimation

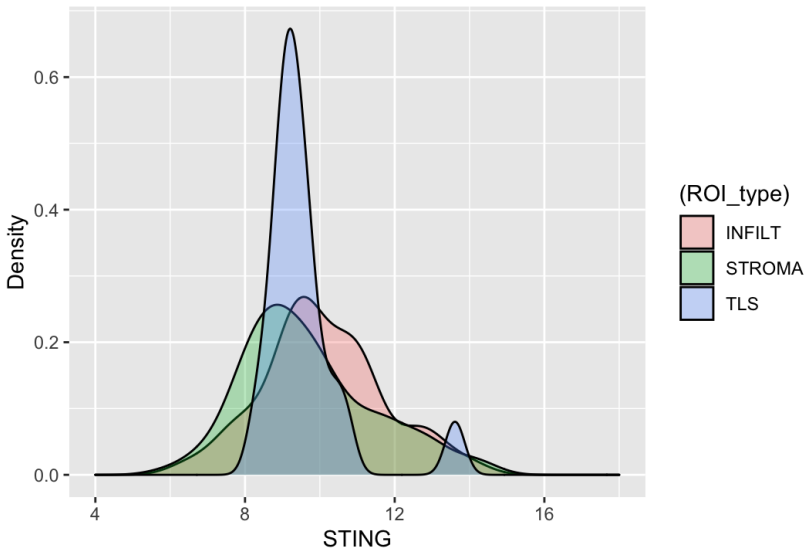


Figure D.2: STING Kernel Density Estimation

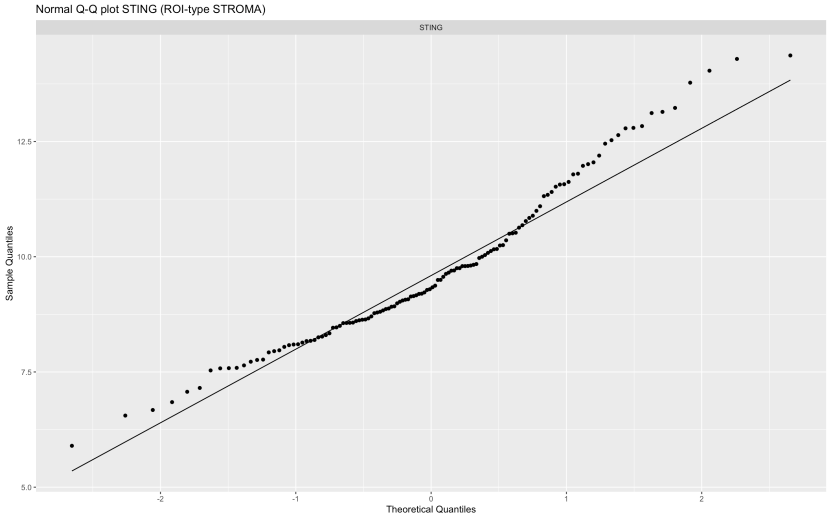


Figure D.3: Normal Q-Q plot of STING (sample type STROMA)

Bibliography

[1] B. Armitage, G. Berry, and J.N.S. Matthews. *Statistical Methods in Medical Research*. 4th ed. Wiley-Blackwell, 2001.

[2] J. Z. Bakdash and L. M. Marusich. "Repeated measures correlation". In: *Frontiers in Psychology: Quantitative Psychology and Measurement* 8.456 (2017).

[3] J. M. Bland and D. G. Altman. "Calculating Correlation Coefficients With Repeated Observations: Part 1: Correlation Within Subjects". In: *British Medical Journal* 310.6977 (1995), p. 446.

[4] J. M. Bland and D. G. Altman. "Calculating Correlation Coefficients With Repeated Observations: Part 2: Correlation Between Subjects". In: *British Medical Journal* 310.6980 (1995), p. 633.

[5] Lan H. Diep. "Variable selection for generalized linear mixed model by L1 penalization for predicting clinical parameters of ovarian cancer". Bachelor's Thesis. Lund: Lund Univesity, 2021.

[6] Andrew Gelman and Hal Stern. "The Difference Between "Significant" and "Not Significant" is not Itself Statistically Significant". In: *The American Statistician* 60.4 (November 2006).

[7] Uwe Hassler and Thorsten Thadewald. "Nonsensical and Biased Correlation Due to Pooling Heterogeneous Samples". In: *Journal of the Royal Statistical Society. Series D (The Statistician)* 52.3 (2003), pp. 367–379.

[8] David C. Hoaglin, Frederick Mosteller, and John W. Tukey. *Understanding Robust and Exploratory Data Analysis*. 1st ed. Wiley, 1983.

[9] David C. Howell. *Statistical Methods for Psychology*. 8th ed. Cengage Learning, 2012.

[10] Bradley E. Huitema. *The Analysis of Covariance and Alternatives: Statistical Methods for Experiments, Quasi-Experiments, and Single-Case Studies*. 2nd ed. Wiley, 2011.

[11] Roger E. Kirk. *Statistics: An Introduction*. 5th ed. Cengage Learning, 2007.

[12] K.L. McNamara et al. "Spatial proteomic characterization of HER2-positive breast tumors through neoadjuvant therapy predicts response". In: *Nat. Cancer* 2 (2021), pp. 400–413.

[13] Peter J. Rousseeuw and Christophe Croux. "Alternatives to the Median Absolute Deviation". In: *Journal of the American Statistical Association* 88.424 (1993).

[14] Peter J. Rousseeuw and Sabine Verboven. "Robust estimation in very small samples". In: *Computational Statistics & Data Analysis* 40 (2002), pp. 741–758.

[15] B.W. Silverman. *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall, 1986.

[16] B. G. Tabachnick and L. S. Fidell. *Using Multivariate Statistics*. 5th ed. Pearson Education, 2007.