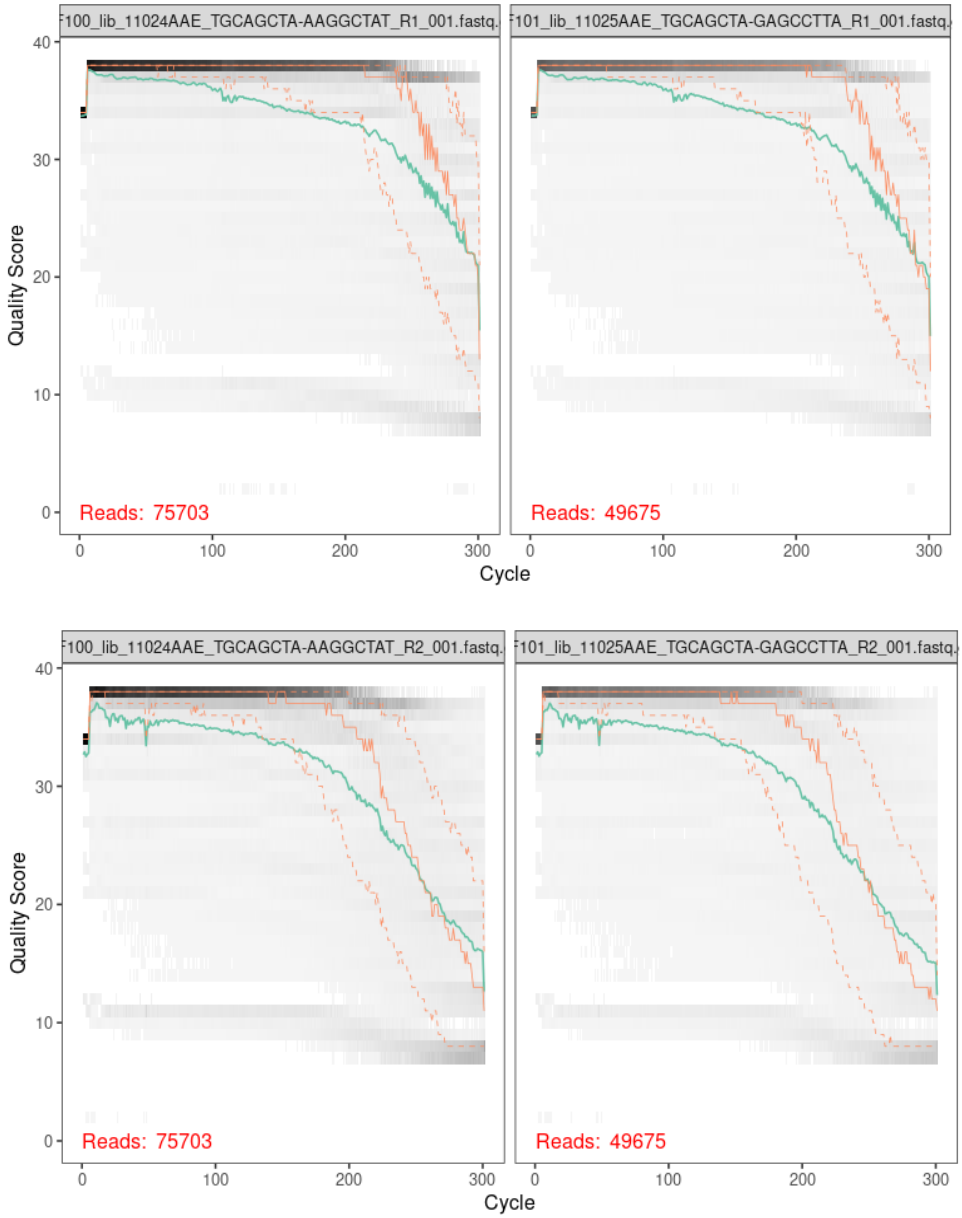
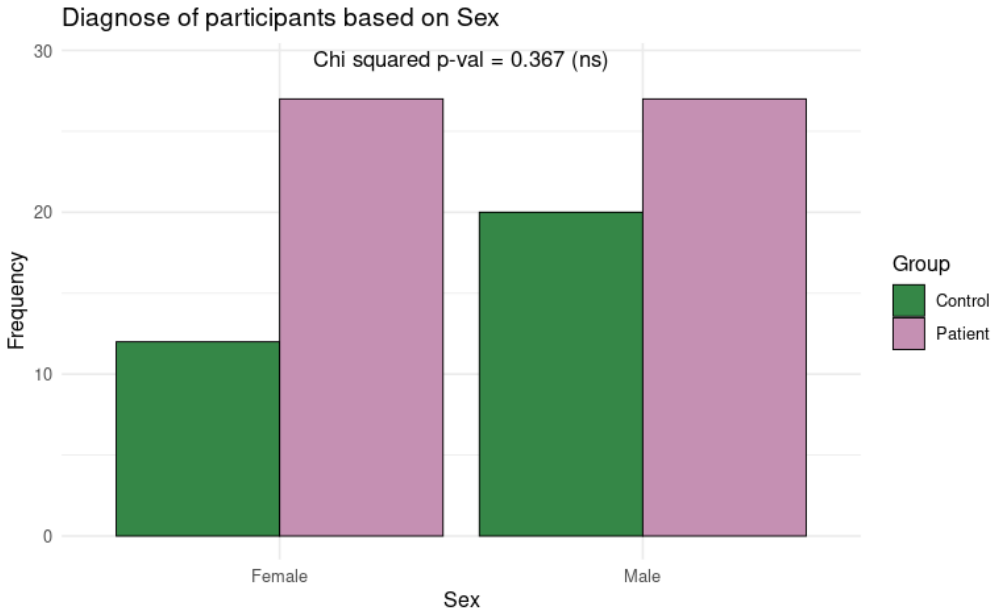
SUPPLEMENTARY MATERIAL

Supplementary Figure 1:



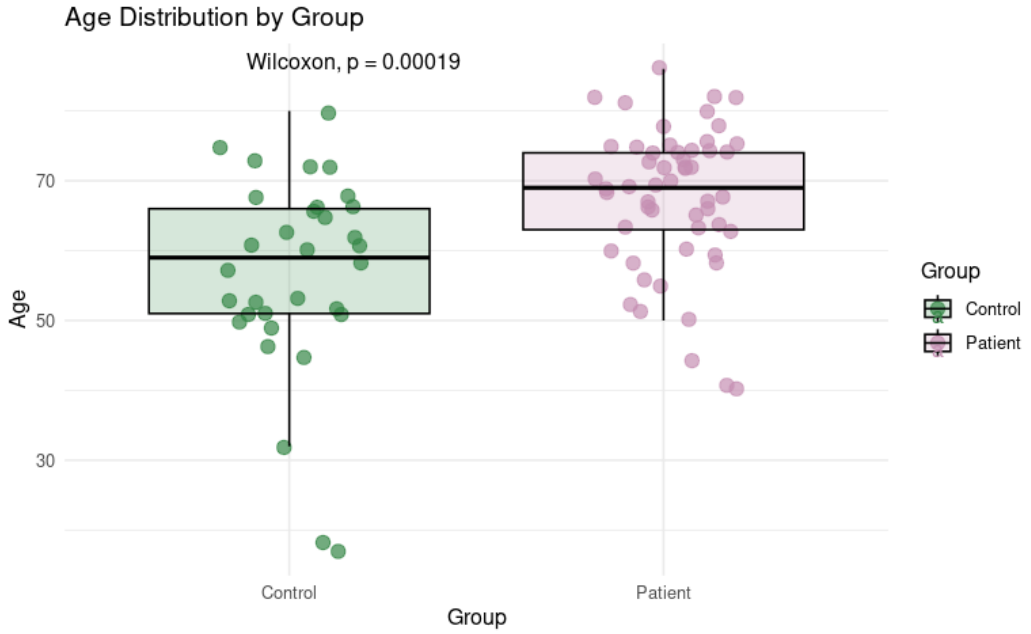
Supplementary Figure 1: Read quality profiles for 16S rRNA raw reads from samples F100 and F101. On top, the quality profiles of the forward reads, at the bottom, the quality profiles of the reverse reads from samples F100 and F101. The X-axis represents each base position, and the Y-axis represents the quality score. The median quality score at each position is shown by the green line, and the quartiles of the quality score distribution by the orange lines.

Supplementary Figure 2:



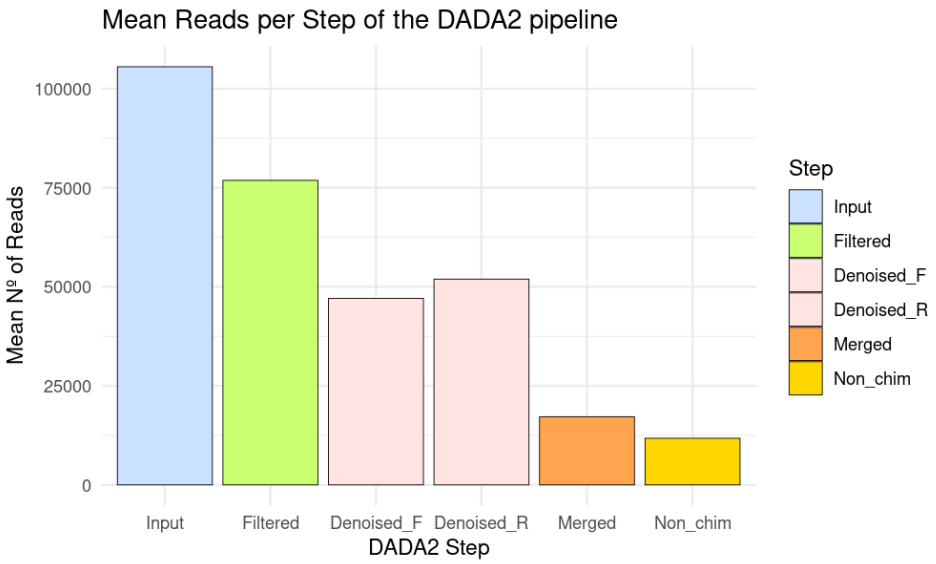
Supplementary Figure 2: Bar plot with the frequency of male and female subjects based on their diagnose group. Controls are displayed in green and pancreatic cancer patients in purple. The Chi squared test was performed to evaluate statistical differences in the sex of the patients based on their group.

Supplementary Figure 3:



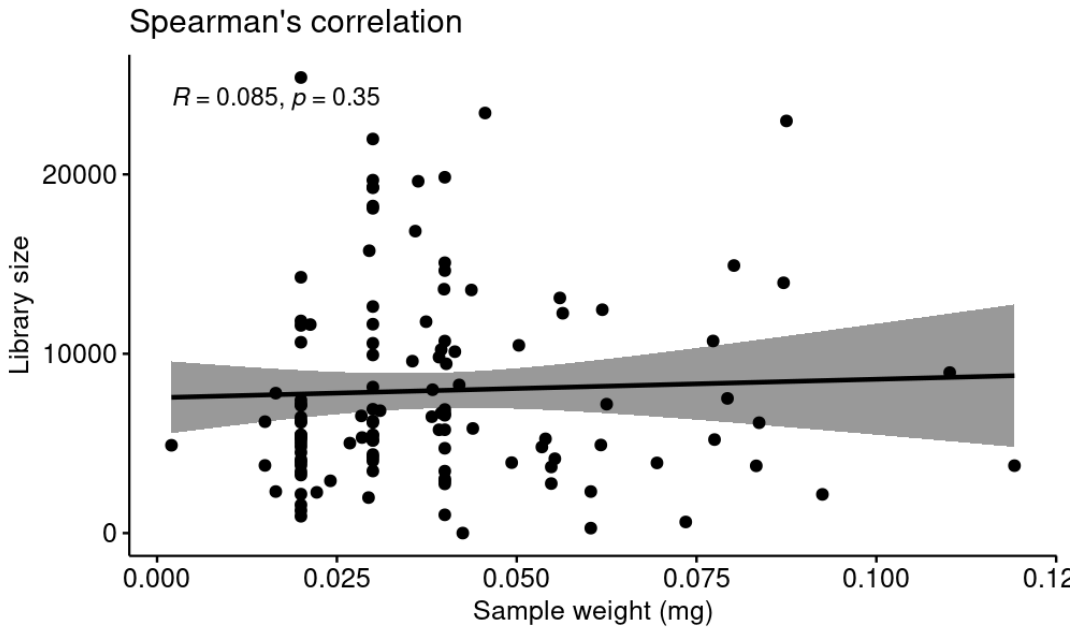
Supplementary Figure 3: Boxplot with the age distribution of the participants based on their group. Controls are displayed in green and pancreatic patients in purple. The Wilcoxon test for non-parametric data distributions was performed to evaluate statistical differences.

Supplementary Figure 4:



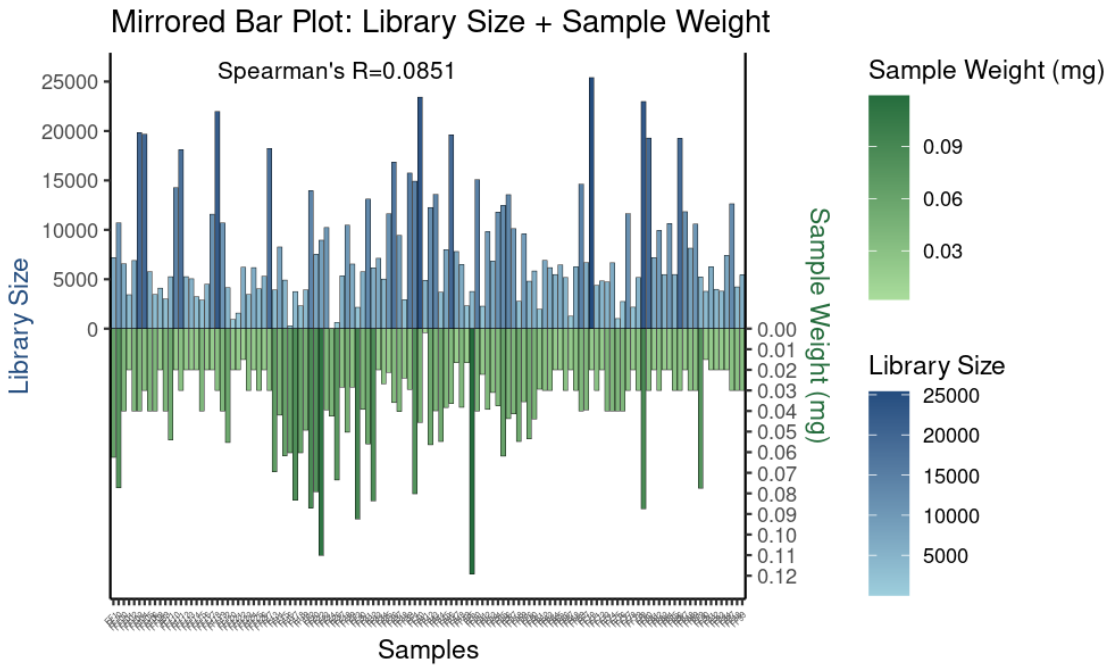
Supplementary Figure 4: Bar plot with the mean count of sequences across all samples through the DADA2 pipeline steps.

Supplementary Figure 5:



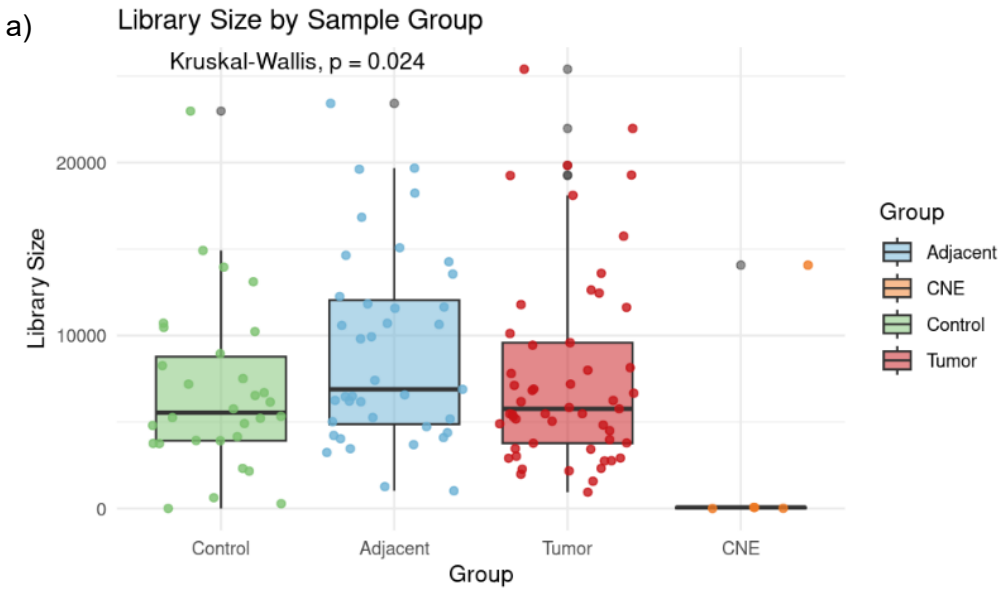
Supplementary Figure 5: Spearman's correlation plot of the library size of the samples in the phyloseq object, against the sample weight of each of the samples (mg) before 16S rRNA sequencing. The Spearman's correlation test was performed over the two variables to prove significance.

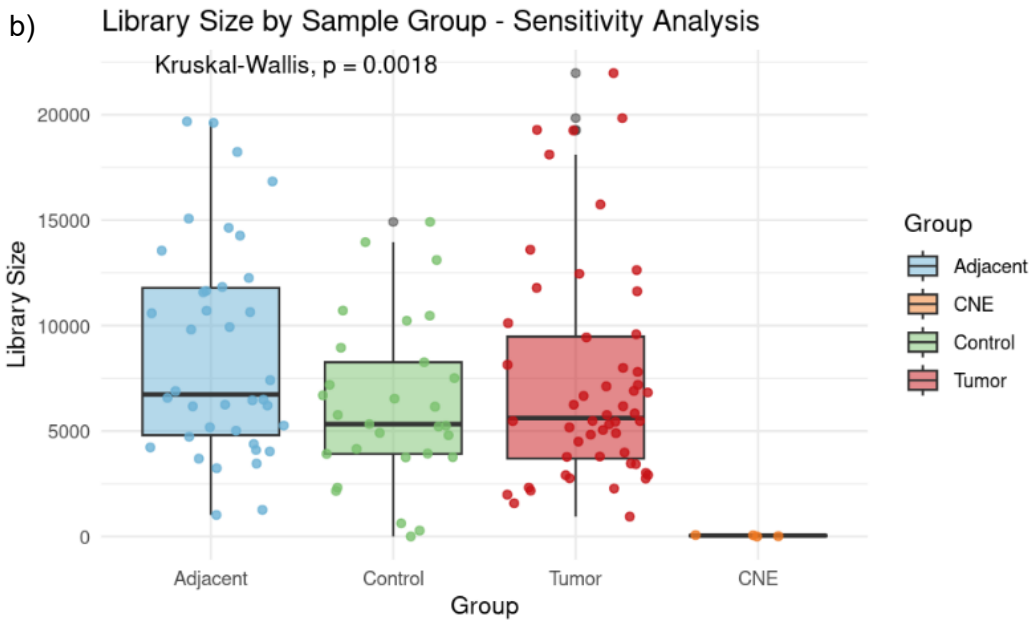
Supplementary Figure 6:



Supplementary Figure 6: Mirror plot with the library size of the samples on the ASV table of the phyloseq object (in blue), against the sample weight of each of the samples before 16S rRNA sequencing (in green). The Spearman's correlation test was performed over the two variables to prove no significant correlation.

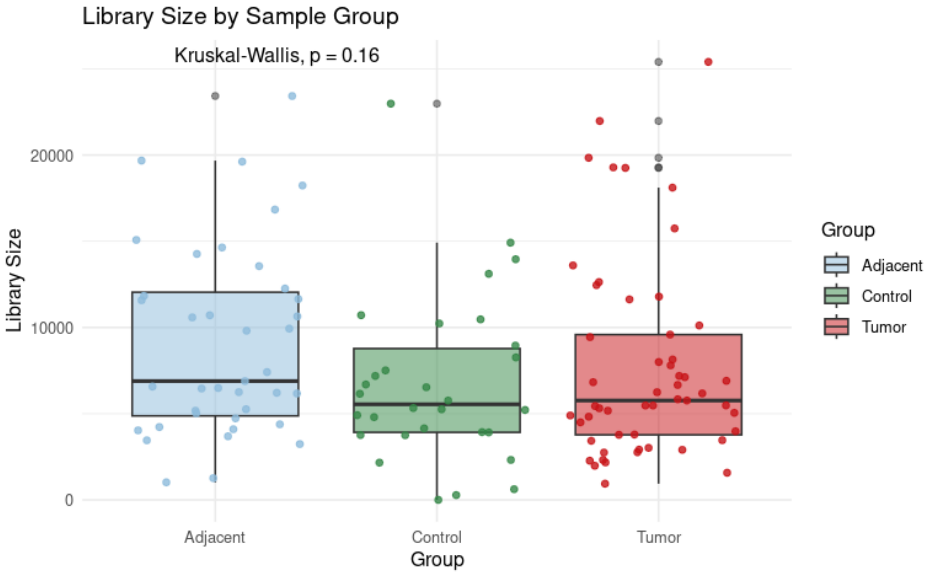
Supplementary Figure 7:





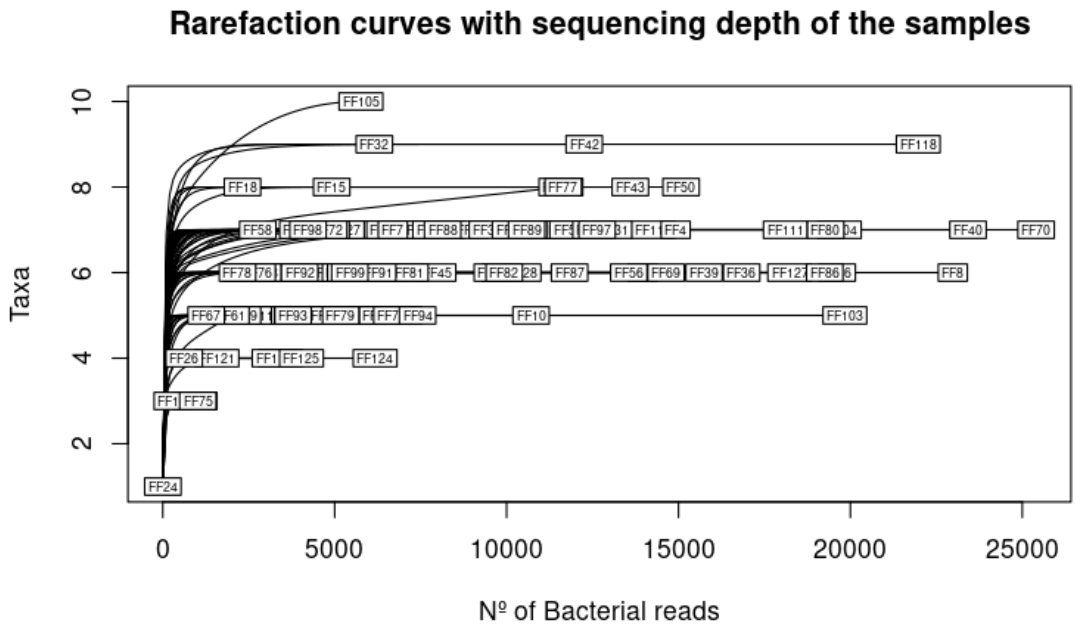
Supplementary Figure 7: Boxplots with the distribution of the library size across groups of samples: Control (green), Adjacent (blue), Tumor (Red) and Negative Experimental Controls (CNE, orange). The points in the distribution represent each one of the samples. The non-parametric Kruskal-Wallis test was used to evaluate significance. a) LS distributions by sample group. b) LS distributions by sample group after the sensitivity test was performed.

Supplementary Figure 8:



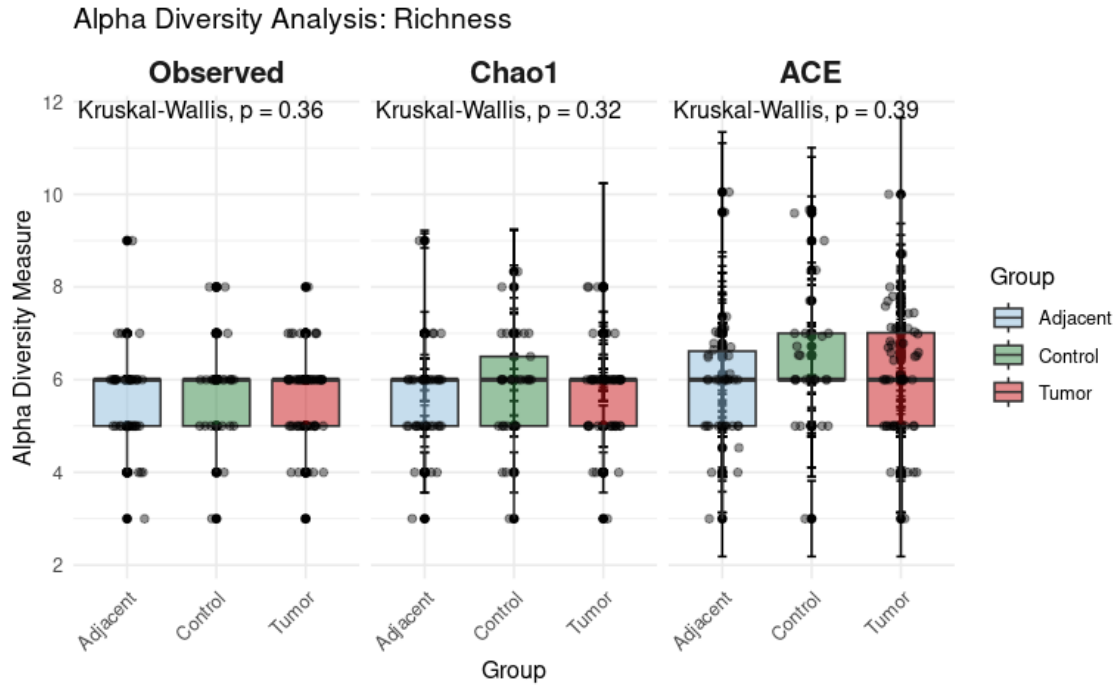
Supplementary Figure 8: Boxplots with the distribution of the library size across groups of samples: Control (green), Adjacent (blue) and Tumor (Red). Each dot represents a sample. The non-parametric Kruskal-Wallis test was used to evaluate significant differences.

Supplementary Figure 9:



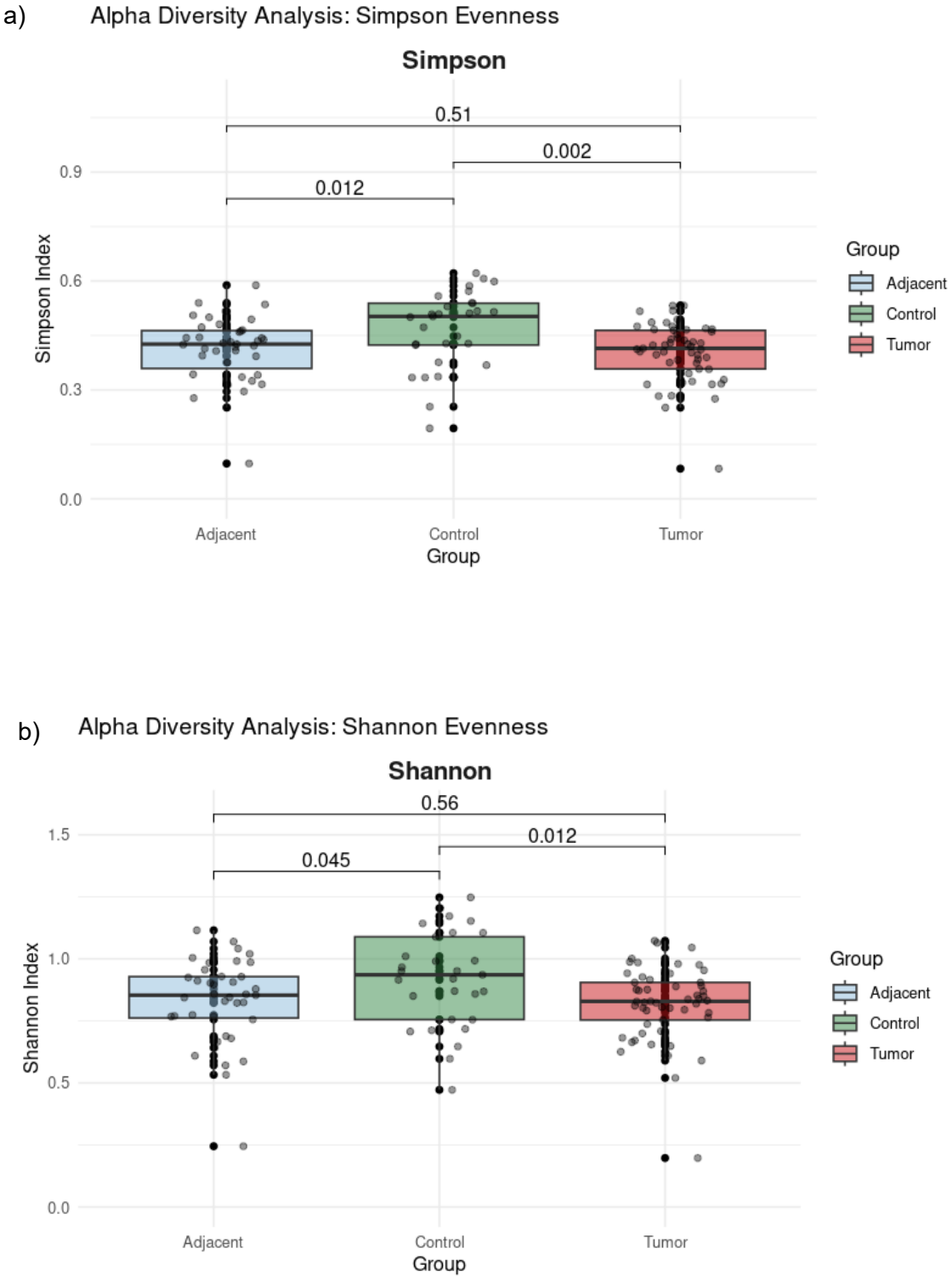
Supplementary Figure 9: Rarefaction curves of the LS of each of the samples and the identified taxonomies in each of the samples. The Y axis reflects the number of different taxa identified in each sample. The X axis provides the number of reads in each of the samples.

Supplementary Figure 10:



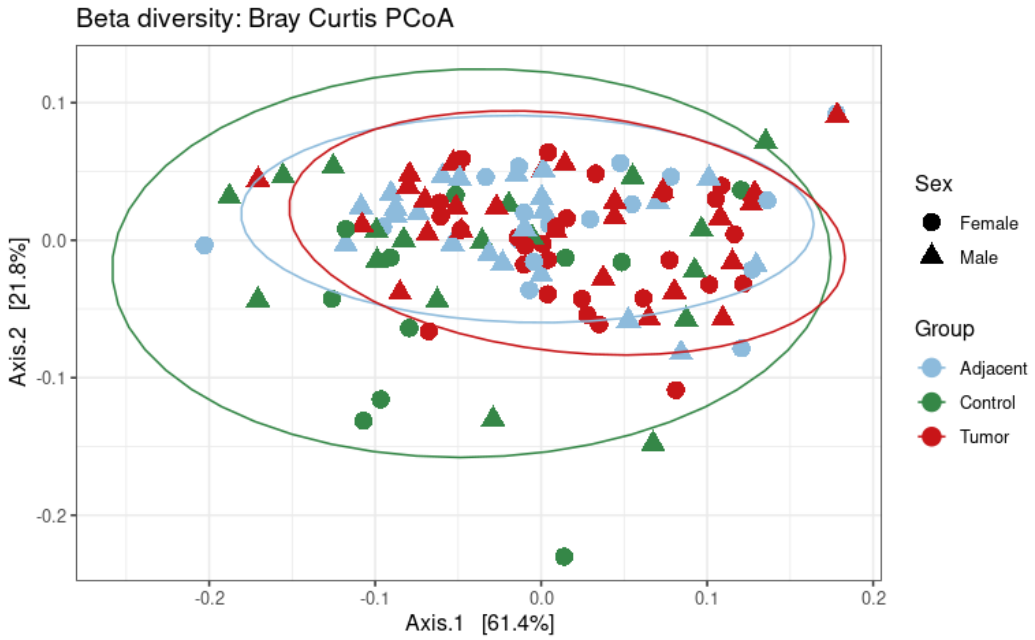
Supplementary Figure 10: Boxplots with the distribution of alpha diversity richness indexes among types of samples: adjacent (blue), control (green), and tumor (red). Kruskal Wallis test provides statistical significance p-values over comparison. On the right, the Observed richness values, in the middle, Chao1 index values, and on the left, ACE index values.

Supplementary Figure 11:



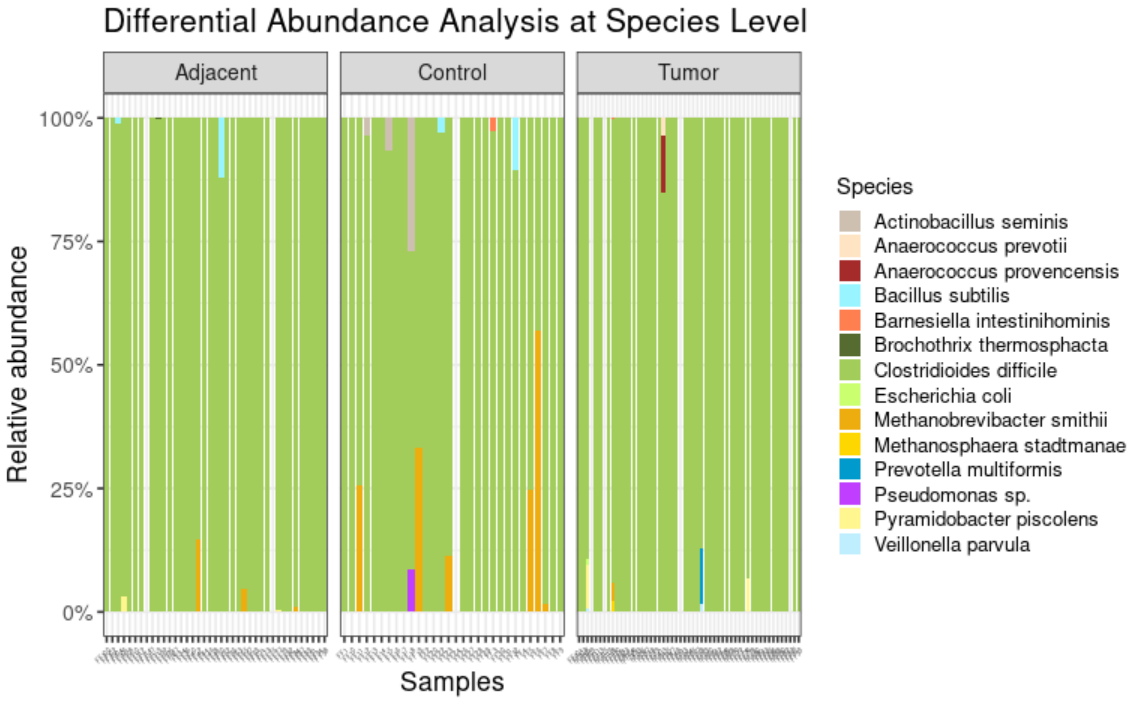
Supplementary Figure 11: Boxplots with the distribution of alpha diversity evenness indexes among types of samples: adjacent (blue), control (green), and tumor (red). Kruskal Wallis test provides statistical significance p-values after multiple comparisons. a) Simpson evenness index analysis. b) Shannon evenness index analysis.

Supplementary Figure 12:



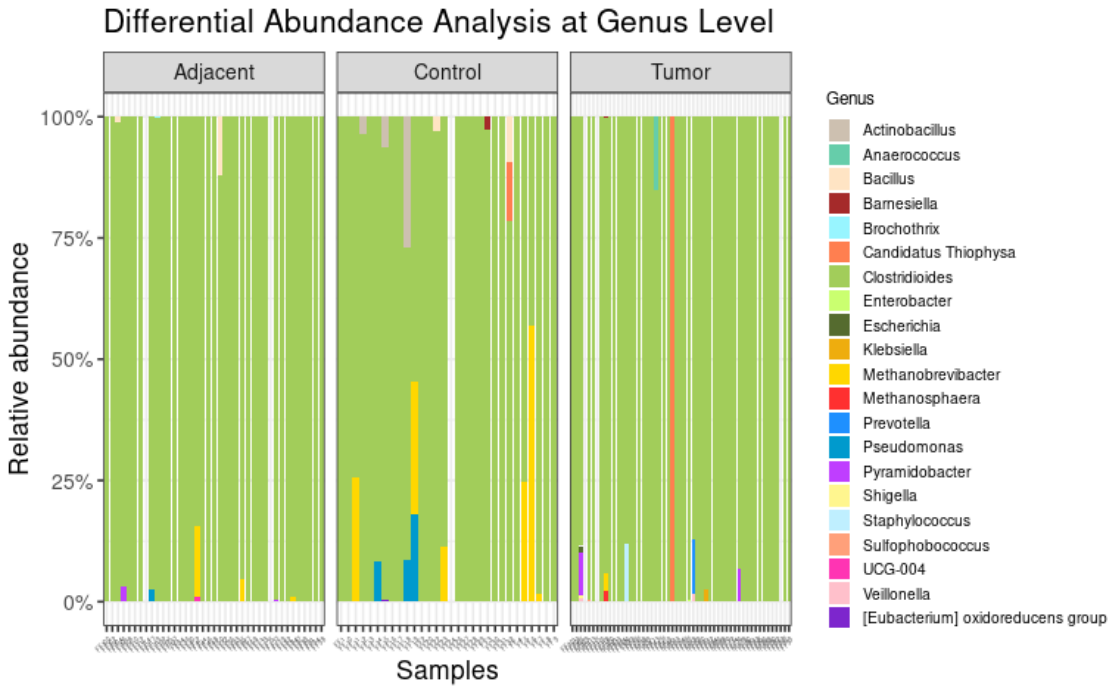
Supplementary Figure 12: Principal Coordinates Analysis (PCoA) with Bray Curtis distances to evaluate the beta diversity (overall differences in taxonomic composition) between samples. Dimension 1 gathers 61.4% of the variability, and Dimension 2 gathers 21.8% of the variability. Samples are coloured by sample type: adjacent (blue), control (green), and tumor (red). Samples are shaped according to sex.

Supplementary Figure 13:



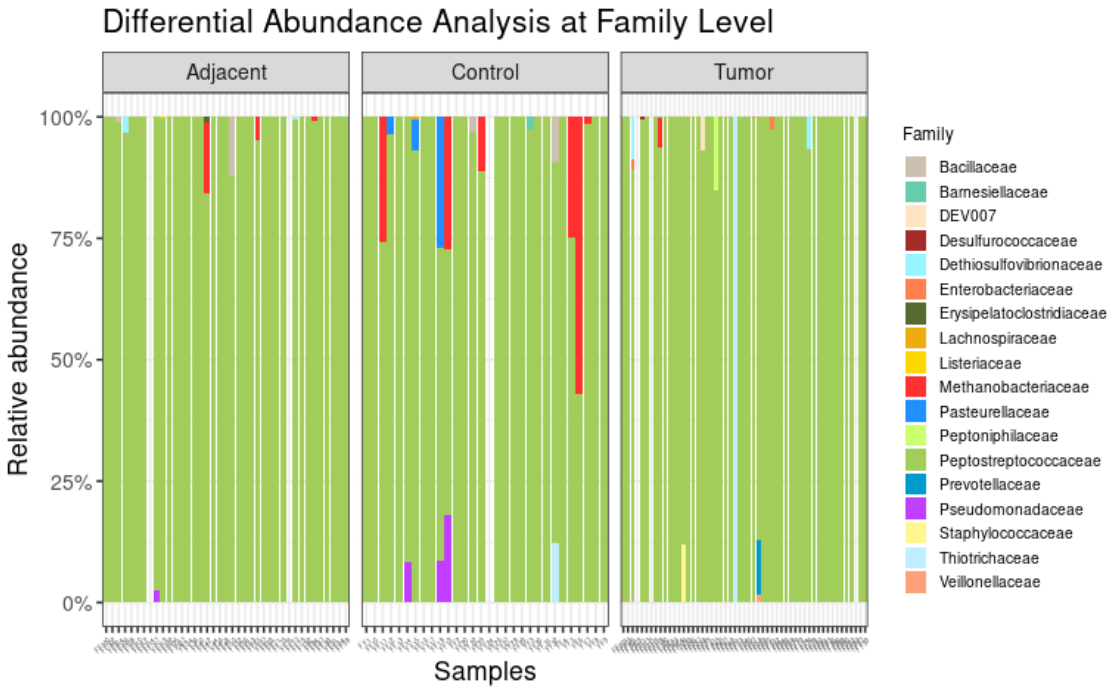
Supplementary Figure 13: Differential Abundance Analysis of the relative abundance of taxonomic composition across samples at species level. Samples are organized by group.

Supplementary Figure 14:



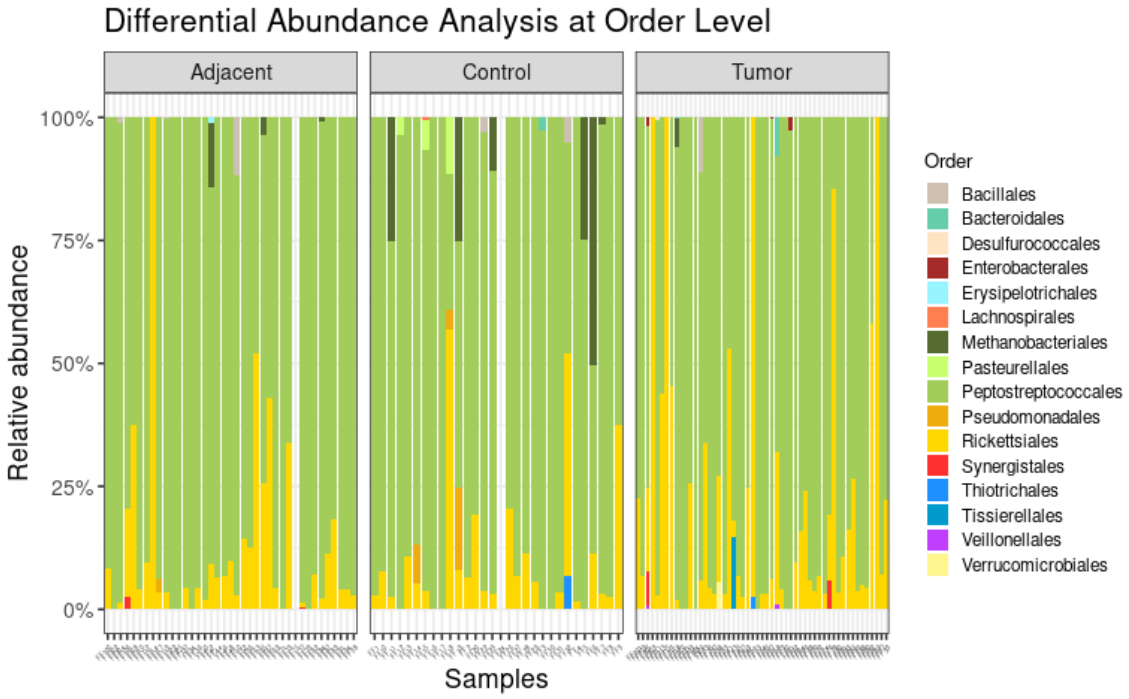
Supplementary Figure 14: Differential Abundance Analysis of the relative abundance of taxonomic composition across samples at genus rank. Samples are organized by group.

Supplementary Figure 15:



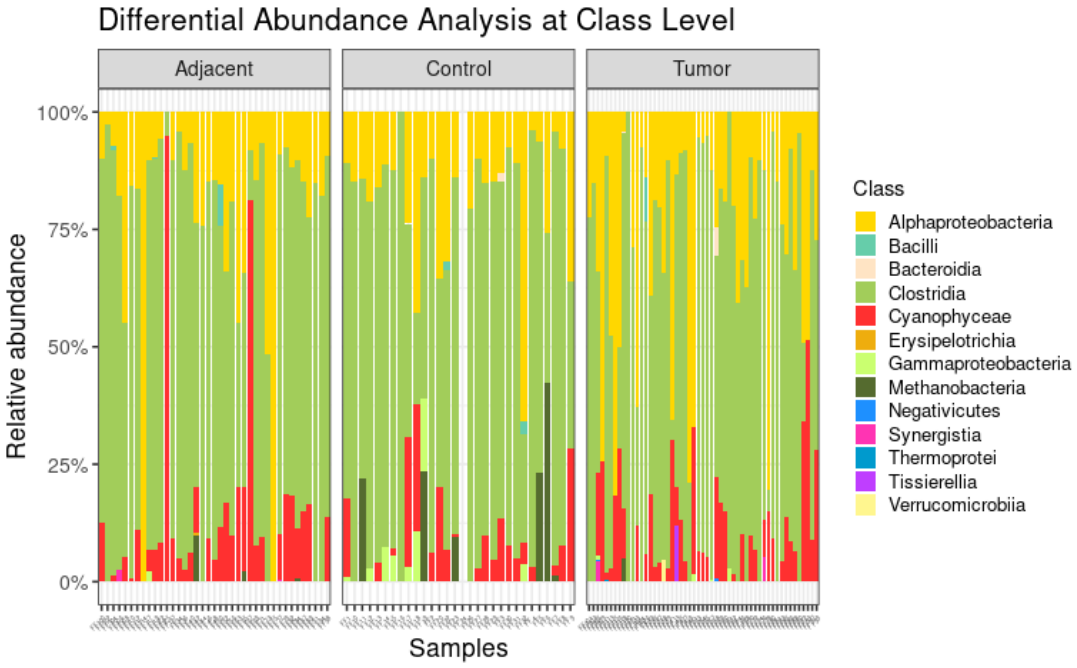
Supplementary Figure 15: Differential Abundance Analysis of the relative abundance of taxonomic composition across samples at family rank. Samples are organized by group.

Supplementary Figure 16:



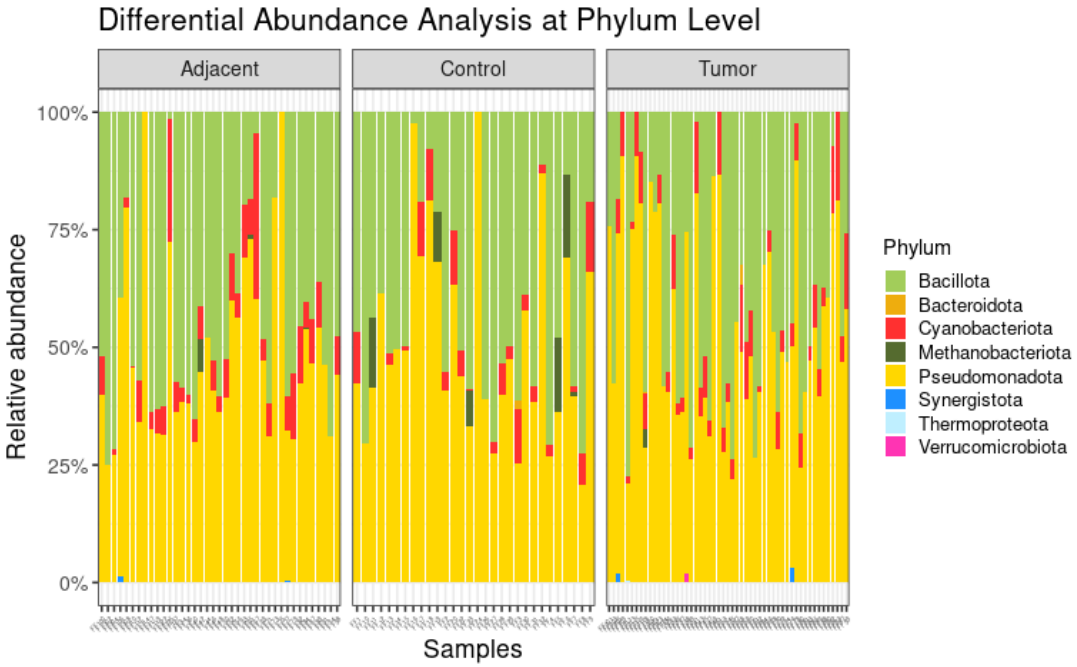
Supplementary Figure 16: Differential Abundance Analysis of the relative abundance of taxonomic composition of the samples at order rank. Samples are organized by group.

Supplementary Figure 17:



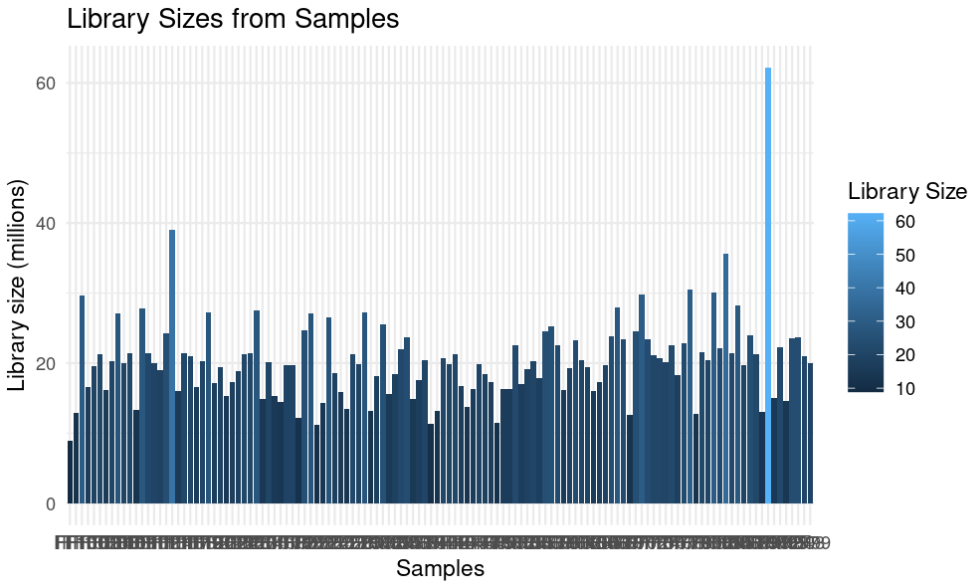
Supplementary Figure 17: Differential Abundance Analysis of the relative abundance of taxonomic composition of the samples at class rank. Samples are organized by group.

Supplementary Figure 18:



Supplementary Figure 18: Differential Abundance Analysis of the relative abundance of taxonomic composition of the samples at phylum rank. Samples are organized by group.

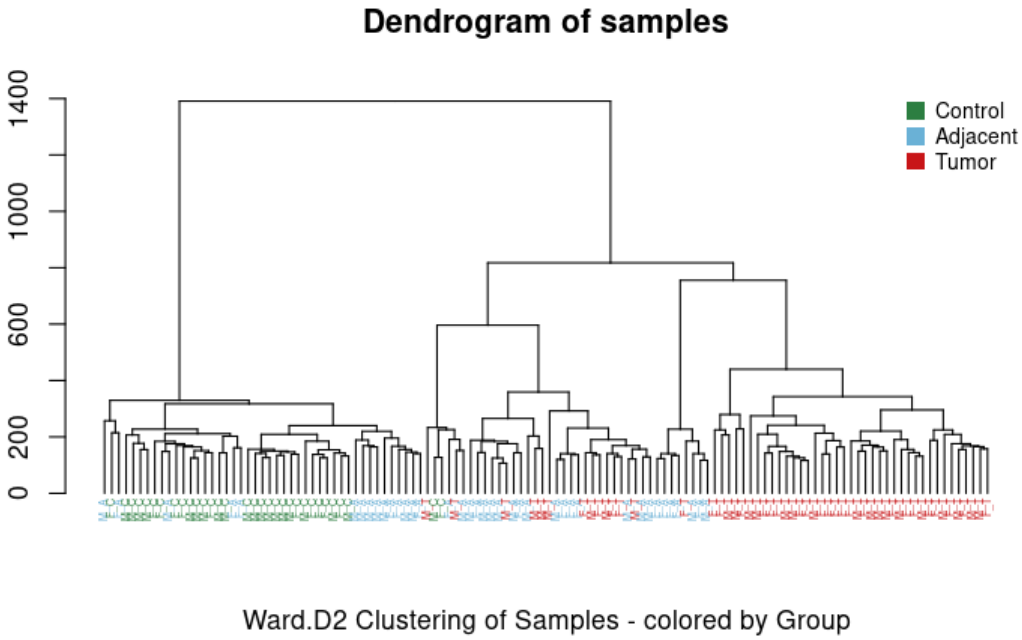
Supplementary Figure 19:

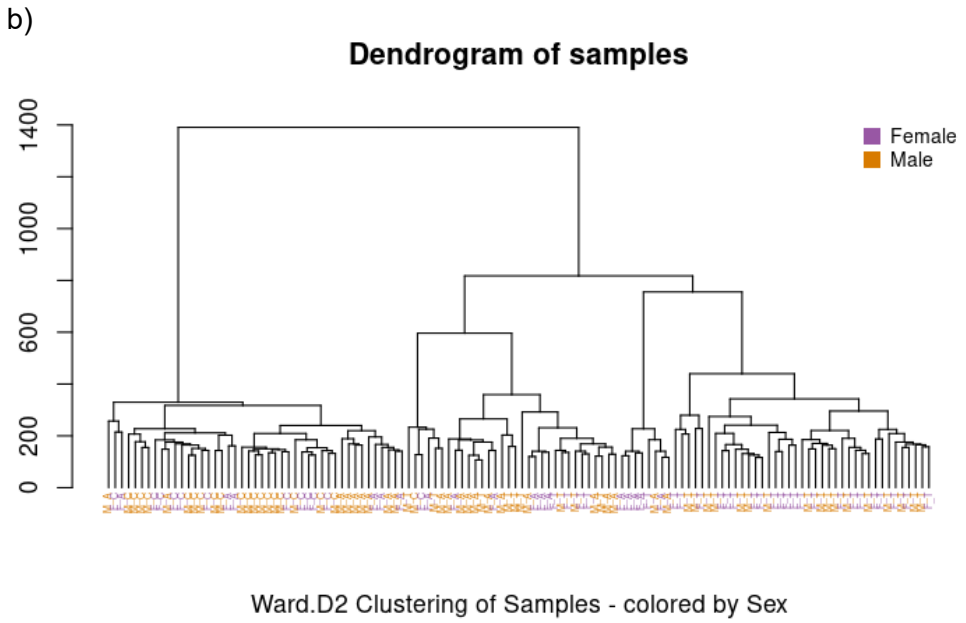


Supplementary Figure 19: Bar plot representing the library size (million counts) across all 124 samples of the study. The colors of the bars range from dark blue (lowest size) to light blue (biggest library size).

Supplementary Figure 20:

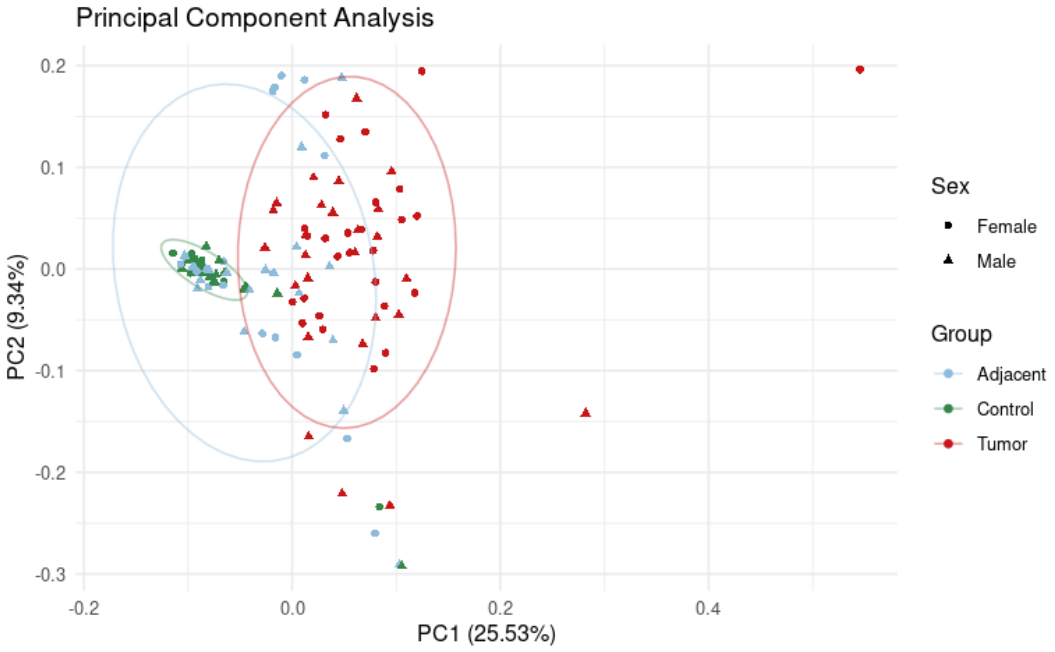
a)





Supplementary Figure 20: Dendrogram with the hierarchical clustering of the samples using the Ward.D2 method, using the gene expression data. Samples have been labelled according to their sex ("M" for male and "F" for female) and group ("C" for Control, "A" for adjacent or "T" for tumor). a) Samples are color coded based on the group. b) Samples are color coded based on the sex.

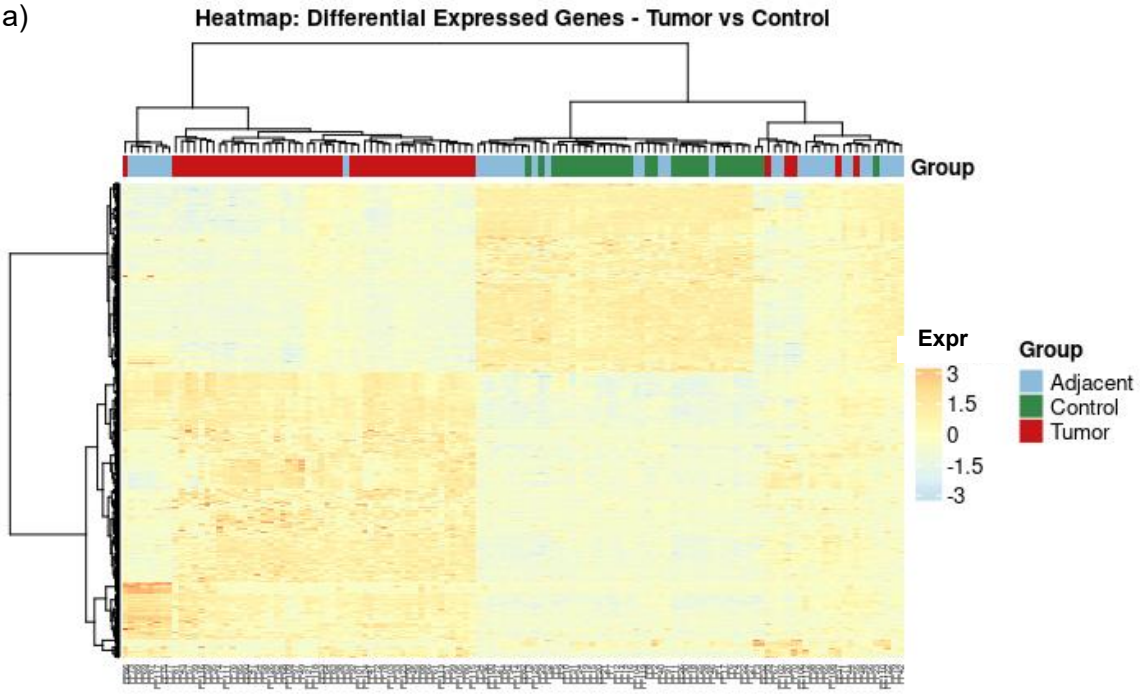
Supplementary Figure 21:



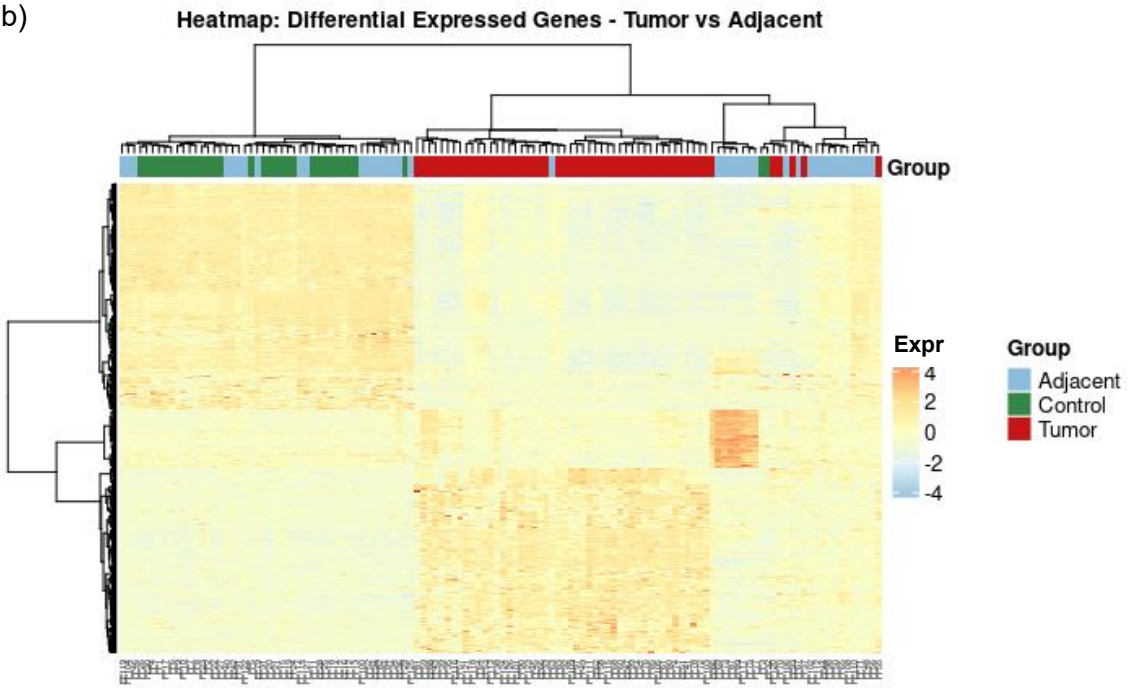
Supplementary Figure 21: PCA with gene expression data. Samples are colored based on their group and shaped based on their sex. PC1 = 25.53% and PC2 = 9.34%.

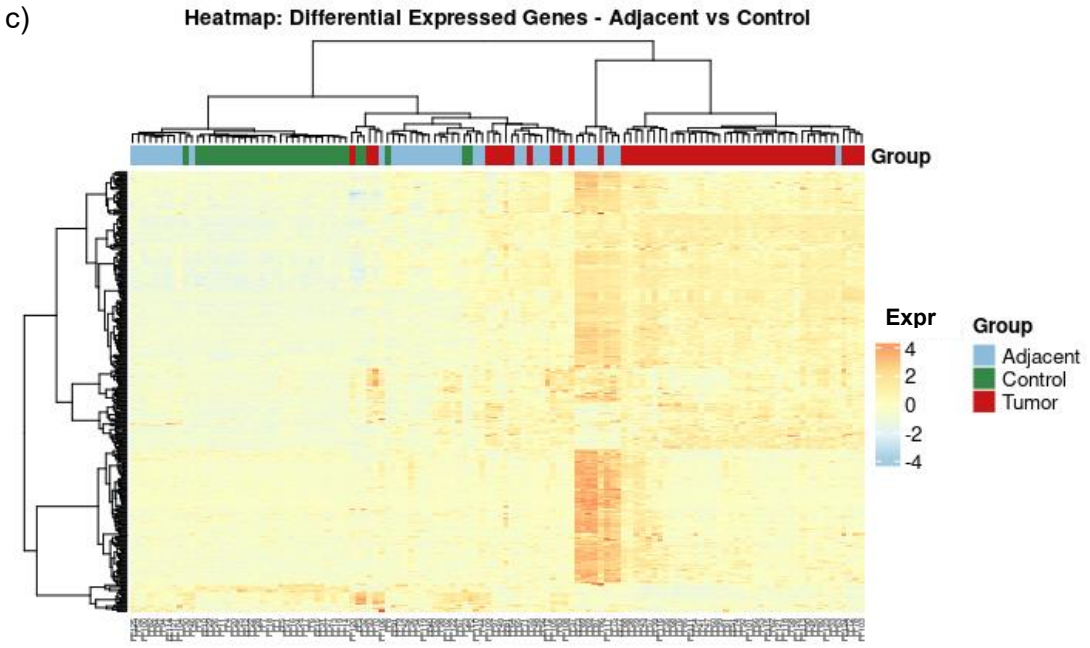
Supplementary Figure 22:

a)



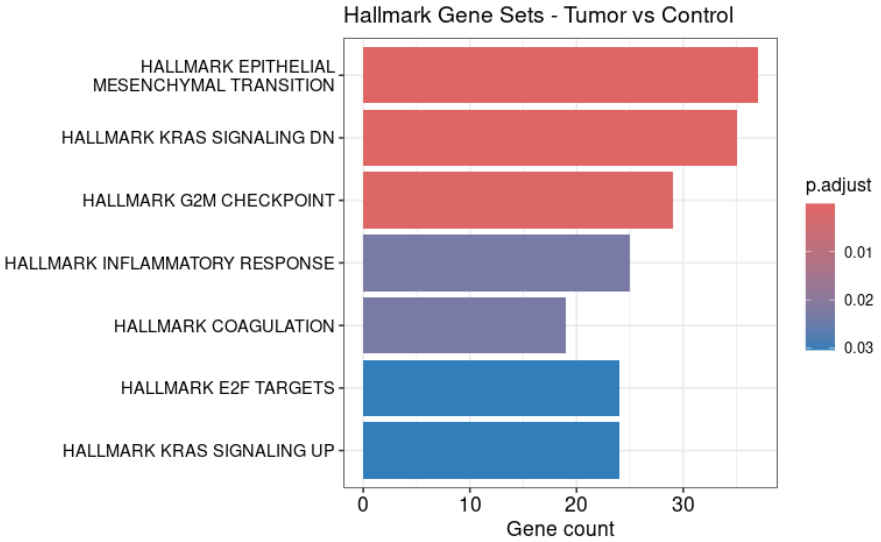
b)





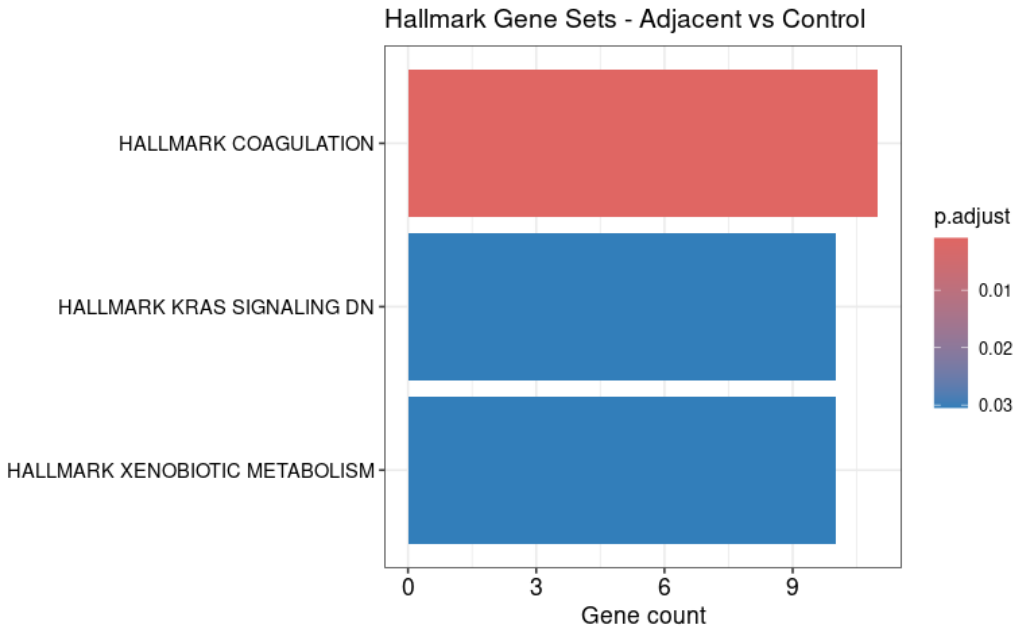
Supplementary Figure 22: Heatmaps with the gene expression patterns of the differential expressed genes across samples though all the pairwise comparisons. Rows represent genes, and columns correspond to pancreatic samples, clustered by Ward.D2 method and color coded by sample group. Gene expression levels are represented by a gradual change of color: ranging from blue (down-regulation) to red (up-regulation), based on the Log2FC value of the comparison. a) Differential expression patterns of DEGs comparing Tumor vs Control samples. b) Differential expression patterns of DEGs comparing Tumor vs Adjacent samples. c) Differential expression patterns of DEGs comparing Adjacent vs Control.

Supplementary Figure 23:



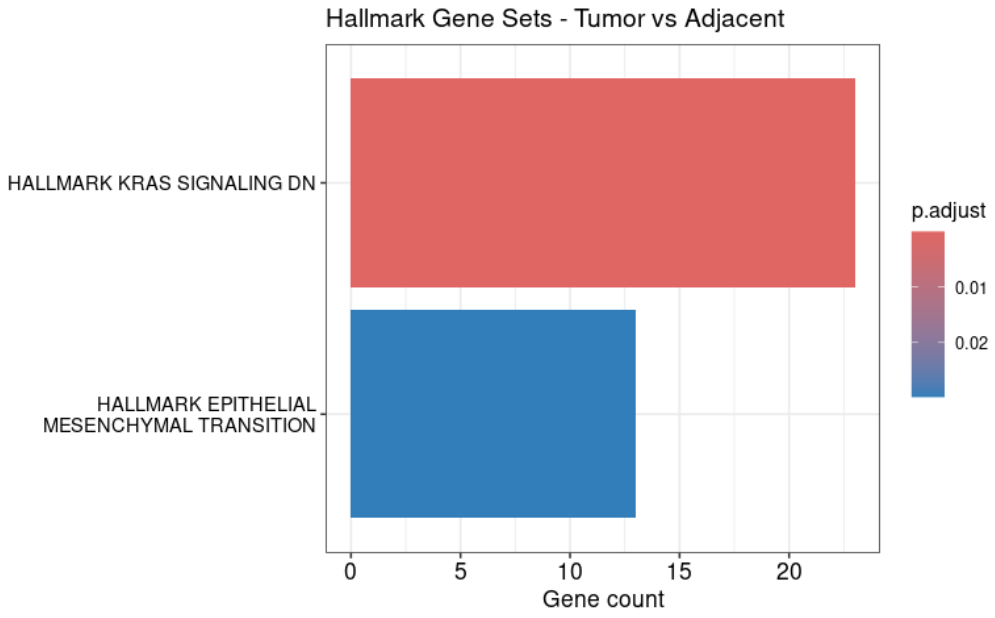
Supplementary Figure 23: Bar plot representing the count of differential expressed genes that are associated with the significantly associated Hallmark gene sets in Tumor vs Control comparison: epithelial mesenchymal transition (with 37 associated DEGs), the KRAS down-signalling (35 DEGs), G2M checkpoint (29 DEGs), inflammatory response (25 DEGs), coagulation (19 DEGs), E2F targets (24 DEGs), and KRAS up-signalling (24 DEGs). There is a color code raging from red to blue, representing the significance by which these hallmark gene sets are associated with the DEGs. All the adjusted p-values shown in the bar plot are significant.

Supplementary Figure 24:



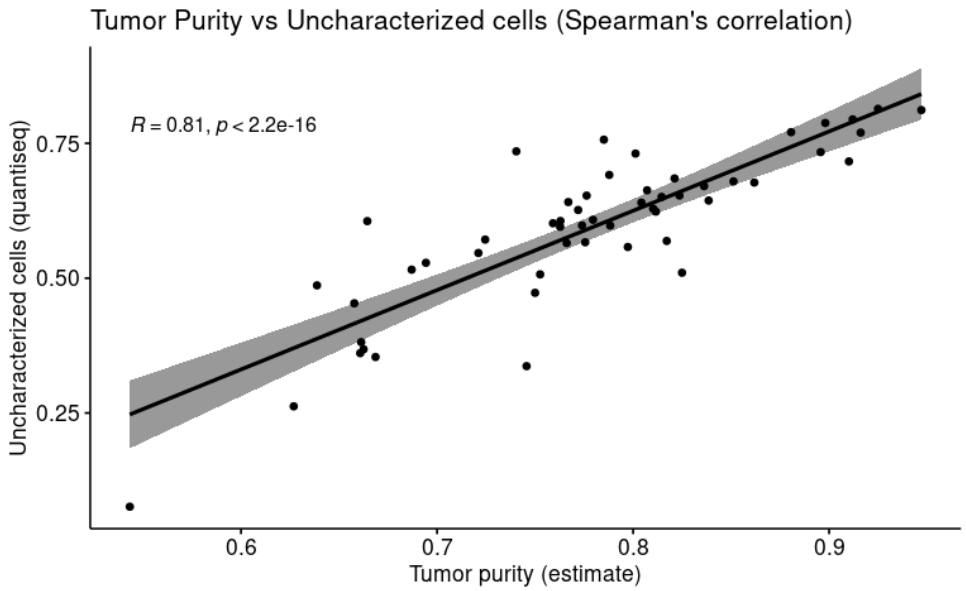
Supplementary Figure 24: Bar plot representing the count of differential expressed genes that are associated with the significantly associated Hallmark gene sets in Adjacent vs Control comparison: hallmark for coagulation (11 DEGs), the KRAS down-signalling (10 DEGs), and xenobiotic metabolism (10 DEGs). There is a color code raging from red to blue, representing the significance by which these hallmark gene sets are associated with the DEGs. All the adjusted p-values shown in the bar plot are significant.

Supplementary Figure 25:



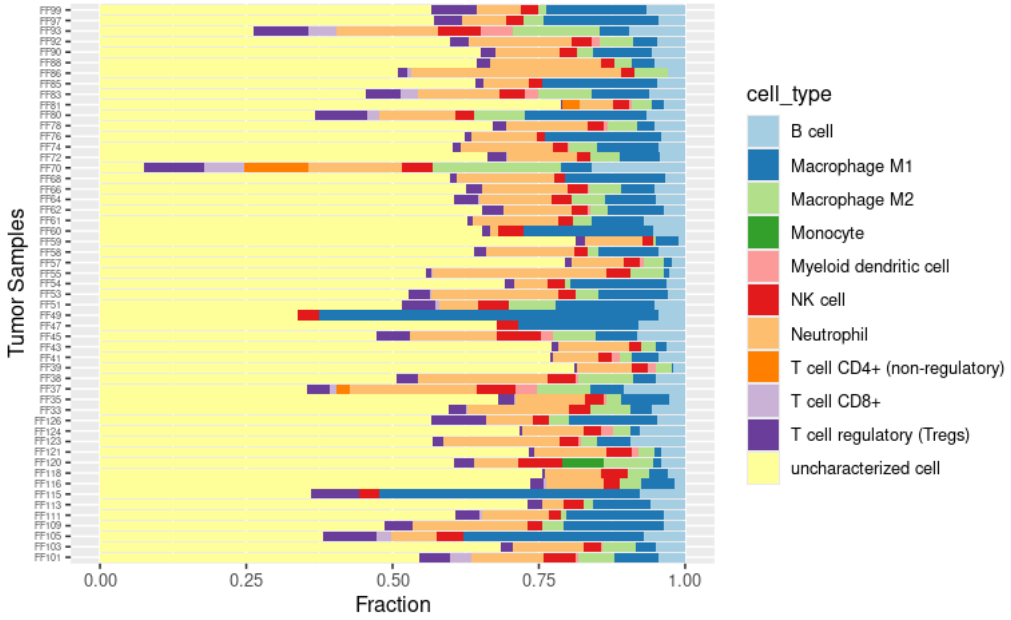
Supplementary Figure 25: Bar plot representing the count of differential expressed genes that are associated with the significantly associated Hallmark gene sets in Tumor vs Adjacent comparison: hallmark for KRAS down-signalling (23 DEGs), and epithelial mesenchymal transition (13 DEGs). There is a color code raging from red to blue, representing the significance by which these hallmark gene sets are associated with the DEGs. All the adjusted p-values shown in the bar plot are significant.

Supplementary Figure 26:



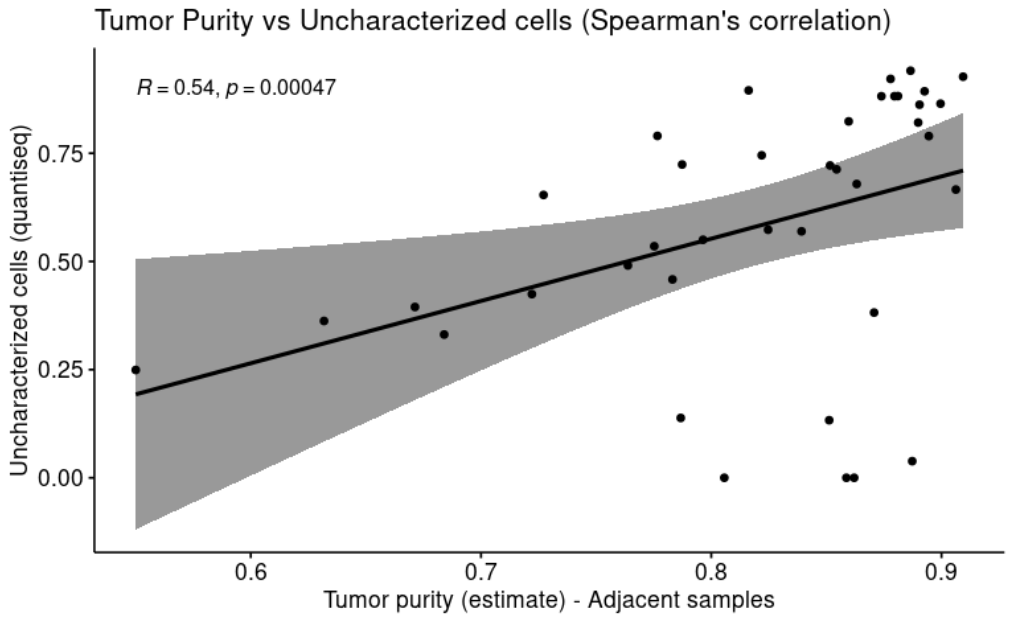
Supplementary Figure 26: Spearman's correlation plot of the percentage of tumor purity ("estimate" *deconvolute* parameter), against the percentage of uncharacterized cells ("quantiseq" *deconvolute* parameter) in the deconvolution analysis of the Tumor Samples. The Spearman's correlation test was performed to prove significance (Rho and p-value provided).

Supplementary Figure 27:



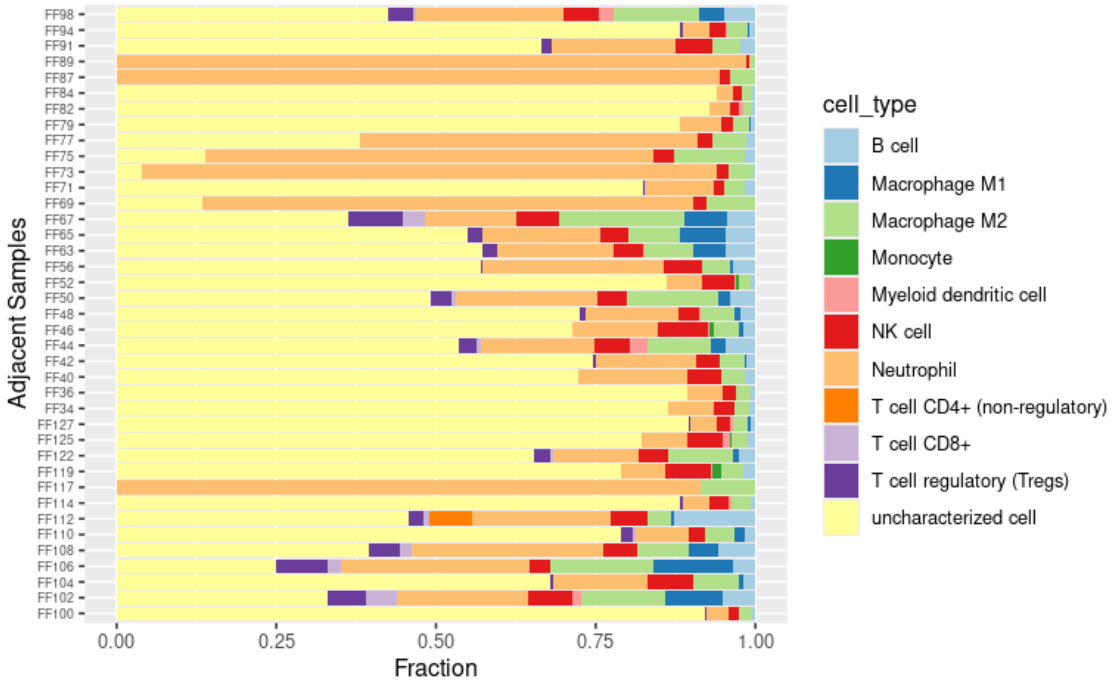
Supplementary Figure 27: Bar plot with the proportions of immune profile cells, and the proportion of uncharacterized cells of the tumor samples, result of the deconvolution analysis.

Supplementary Figure 28:



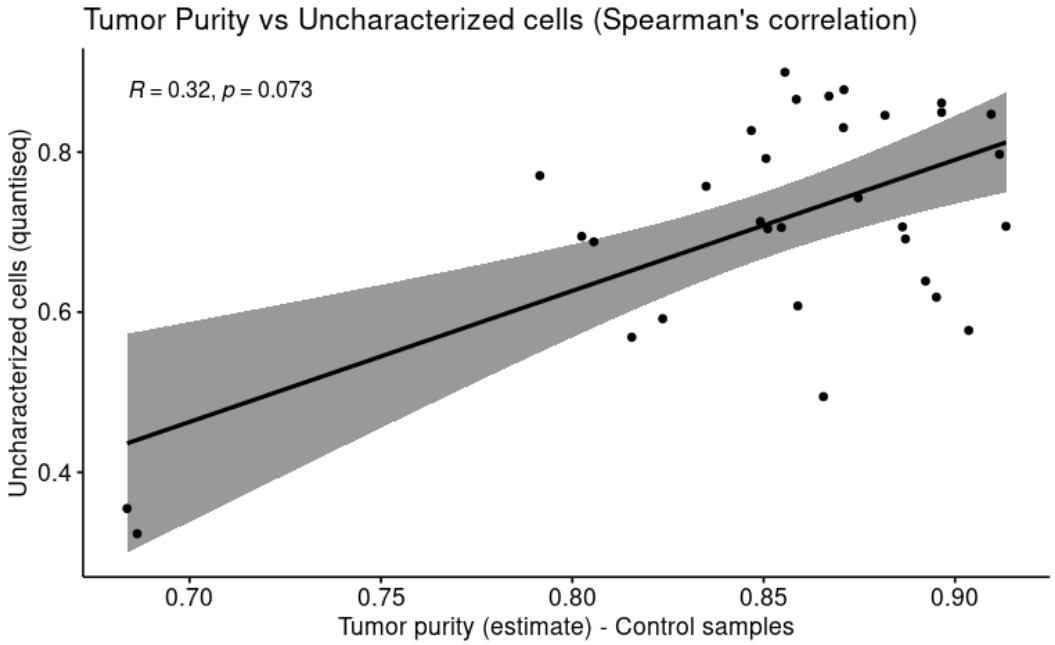
Supplementary Figure 28: Spearman's correlation plot of the percentage of tumor purity ("estimate" *deconvolute* parameter), against the percentage of uncharacterized cells ("quantiseq" *deconvolute* parameter) in the deconvolution analysis of the Adjacent Samples. The Spearman's correlation test was performed to prove significance (Rho and p-value provided).

Supplementary Figure 29:



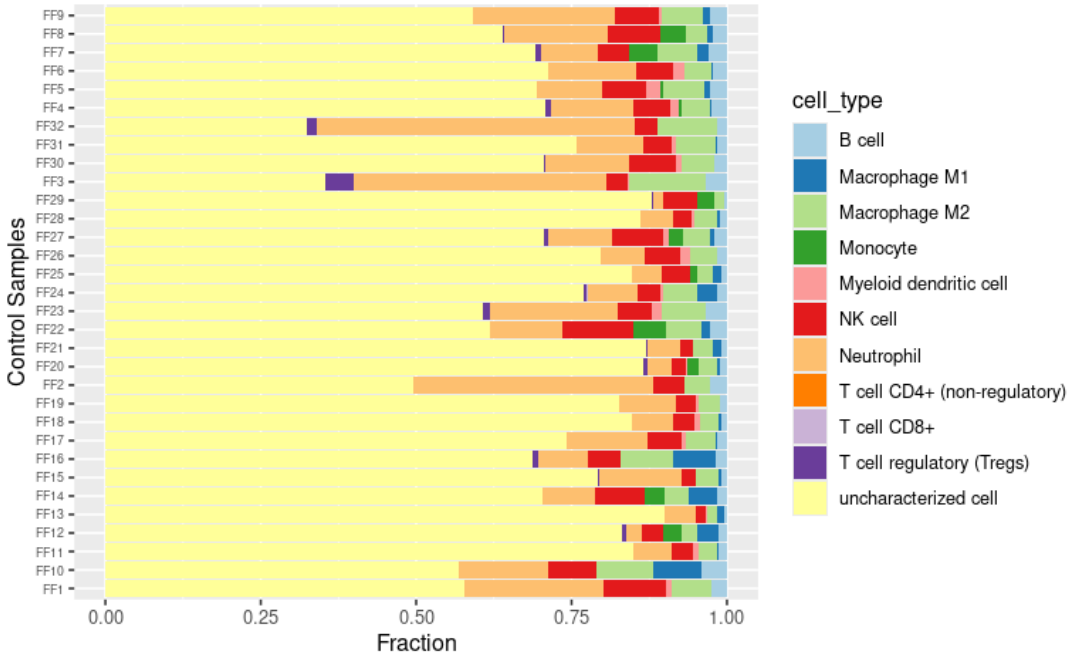
Supplementary Figure 29: Bar plot with the proportions of immune profile cells, and the proportion of uncharacterized cells of the adjacent samples, result of the deconvolution analysis.

Supplementary Figure 30:



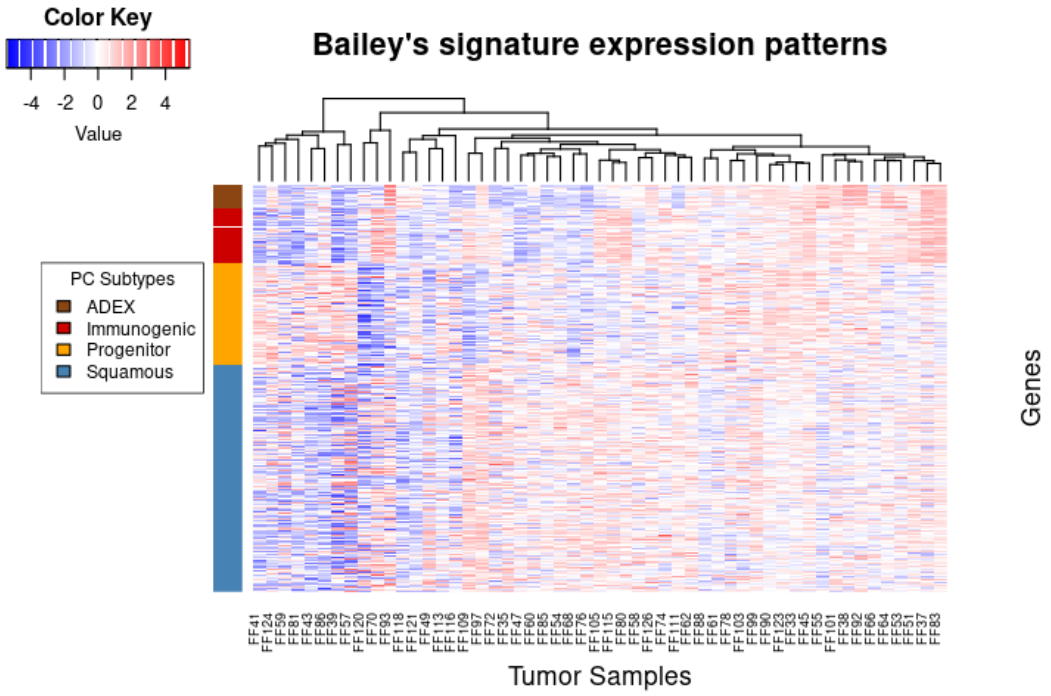
Supplementary Figure 30: Spearman's correlation plot of the percentage of tumor purity ("estimate" *deconvolute* parameter), against the percentage of uncharacterized cells ("quantiseq" *deconvolute* parameter) in the deconvolution analysis of the Control Samples. The Spearman's correlation test was performed to prove significance (Rho and p-value provided).

Supplementary Figure 31:



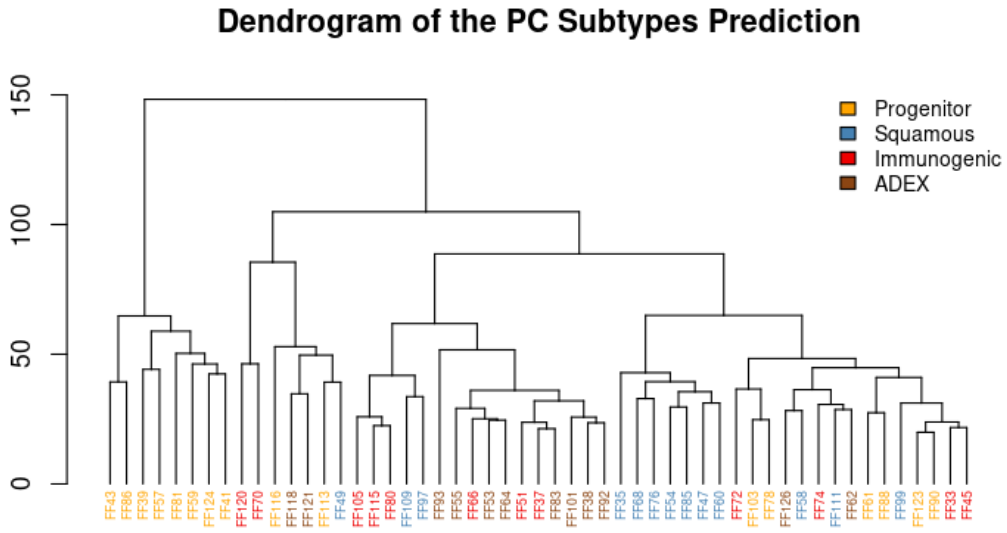
Supplementary Figure 31: Bar plot with the proportions of immune profile cells, and the proportion of uncharacterized cells of the control samples, result of the deconvolution analysis.

Supplementary Figure 32:



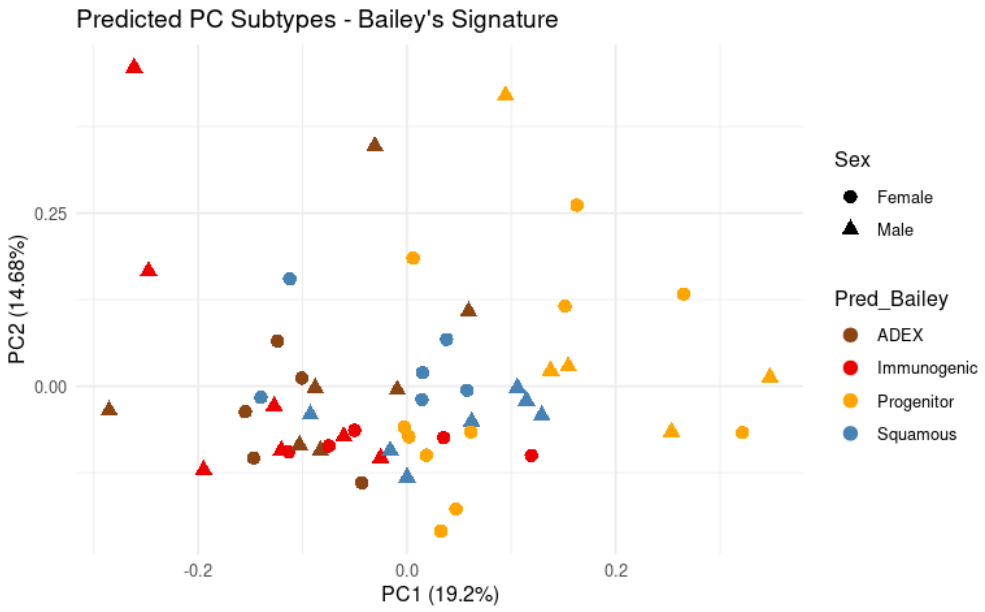
Supplementary Figure 32: Heatmap with the gene expression patterns of the Bailey's signature genes, ordered by PC subtypes (ADEX, Immunogenic, Progenitor and Squamous), across tumoral samples of the study. Samples are clustered by Ward D2 method. Gene expression levels are represented by a gradual change of color: ranging from blue (down-regulation) to red (up-regulation).

Supplementary Figure 33:



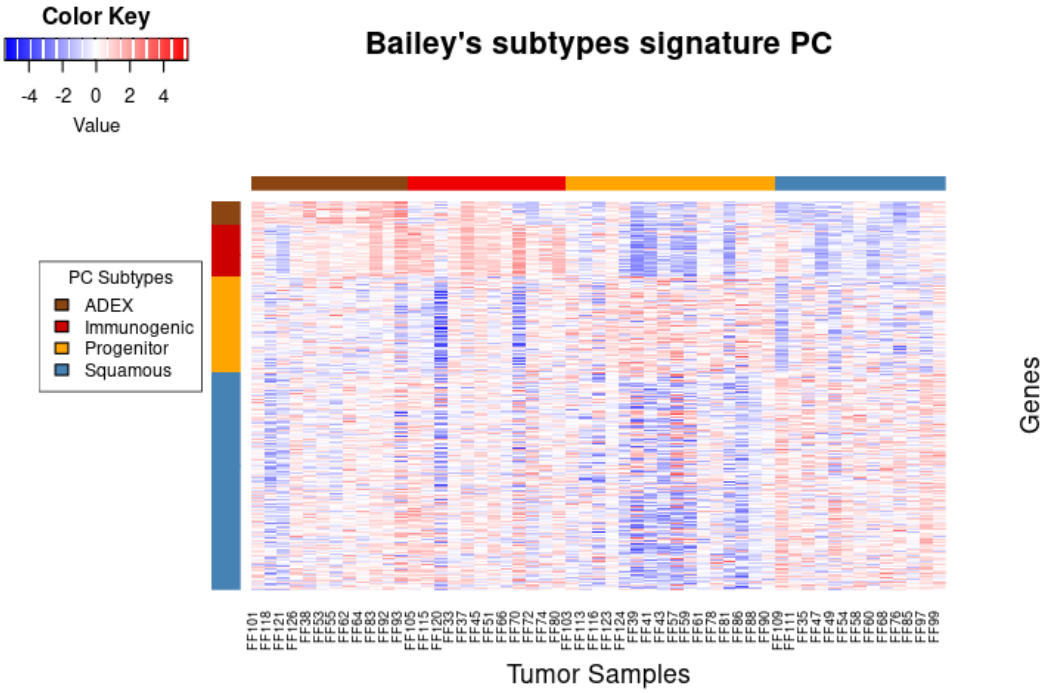
Supplementary Figure 33: Dendrogram with the hierarchical clustering of the tumoral samples using the Ward.D2 method, using the gene expression data of the genes that characterize the Bailey's PC subtyping signature. Samples have been color coded according to the assigned prediction of PC subtype.

Supplementary Figure 34:



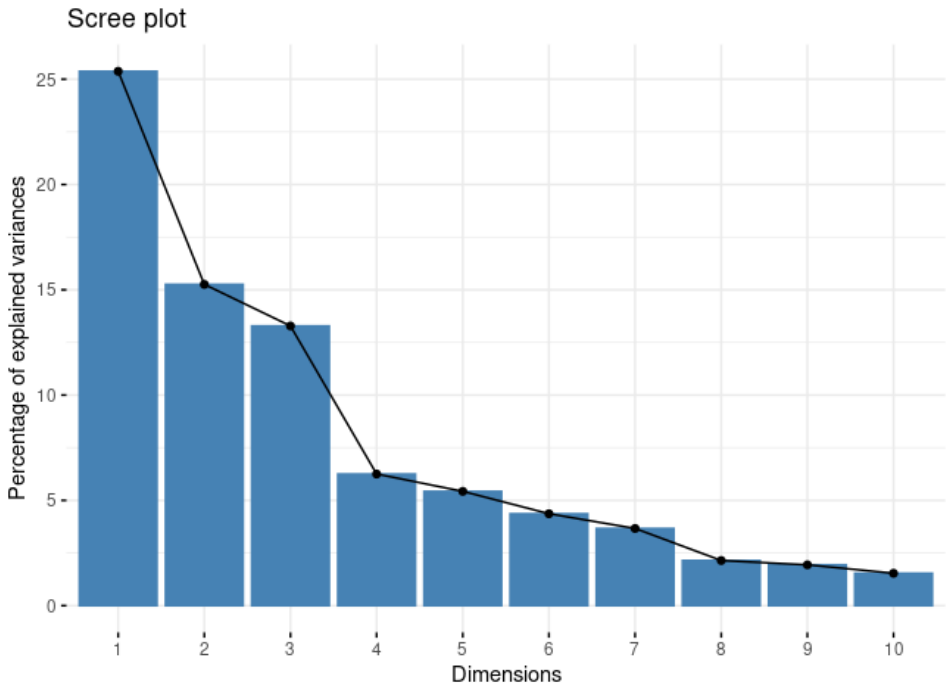
Supplementary Figure 34: Principal Component Analysis of the tumor samples, built from normalized gene expression data of all the genes. Samples are color coded based on their predicted PC subtype, and shaped based on the sex. PC1 accounts for 19.2% of variability, and PC2 accounts for 14.68% of variability.

Supplementary Figure 35:



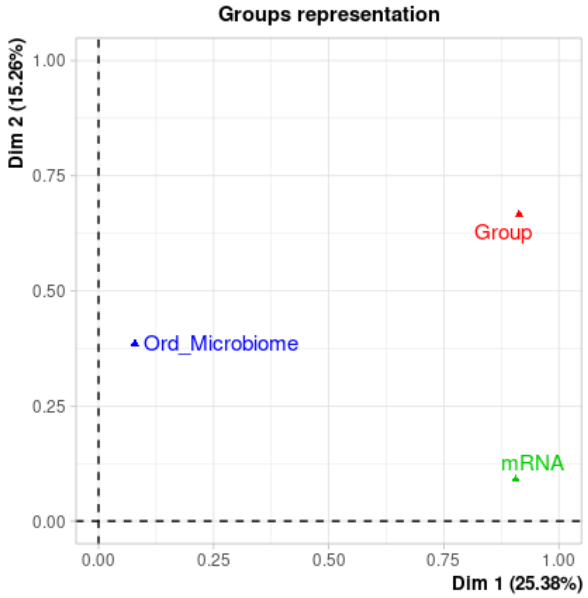
Supplementary Figure 35: Heatmap with the gene expression patterns of the Bailey's signature genes, ordered by PC subtypes (ADEX, Immunogenic, Progenitor and Squamous), across tumoral samples of the study. Samples are ordered based on their predicted PC subtype. Gene expression levels are represented by a gradual change of color: ranging from blue (down-regulation) to red (up-regulation).

Supplementary Figure 36:



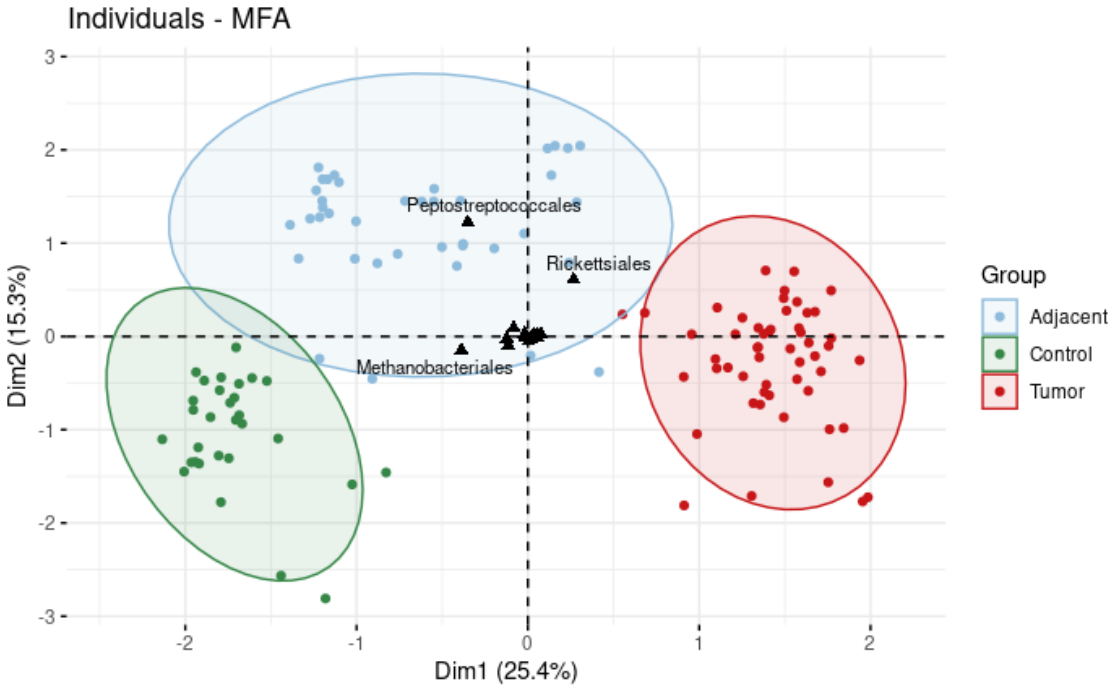
Supplementary Figure 36: Scree plot with the percentage of explained variation by each one of the dimensions derived from the Multiple Factor Analysis (MFA). Dimension 1 accounts for 25.38%, Dimension 2 accounts for 15.26% and Dimension 3 accounts for 13.3% of variances.

Supplementary Figure 37:



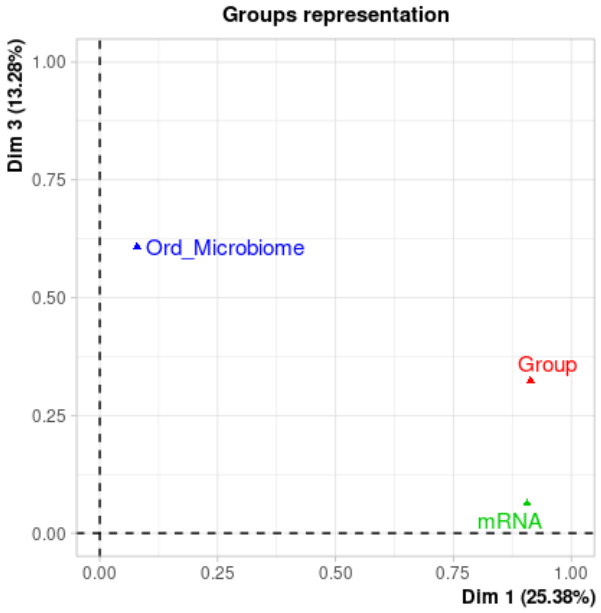
Supplementary Figure 37: Groups representation of the Multiple Factor Analysis (MFA). MFA integrates pancreatic associated microbiota at Order taxonomic level (Ord_Microbiota), with gene expression data (mRNA), and the Group of samples. Dimension 1 accounts for 25.38% and Dimension 2 accounts for 15.26% of variability.

Supplementary Figure 38:



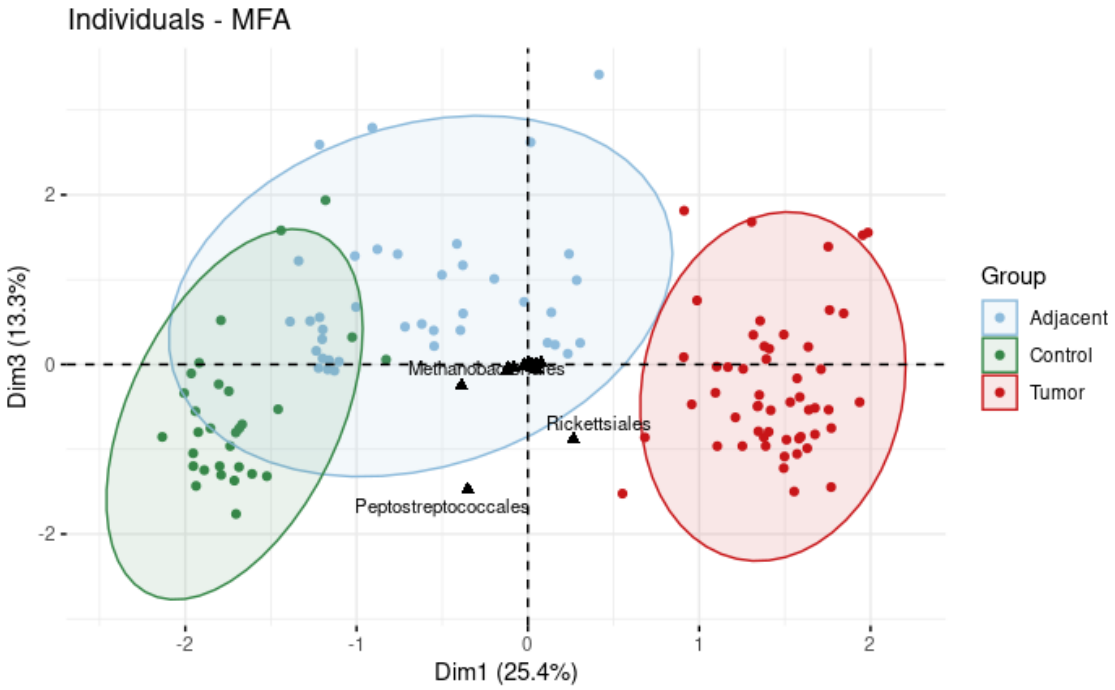
Supplementary Figure 38: Individuals representation in the Multiple Factor Analysis (MFA). Individual samples are represented as dots, and color coded based on their group of sample. OTUs at order level are represented as black triangles. Dimension 1 accounts for 25.38% and Dimension 2 accounts for 15.26% of variability.

Supplementary Figure 39:



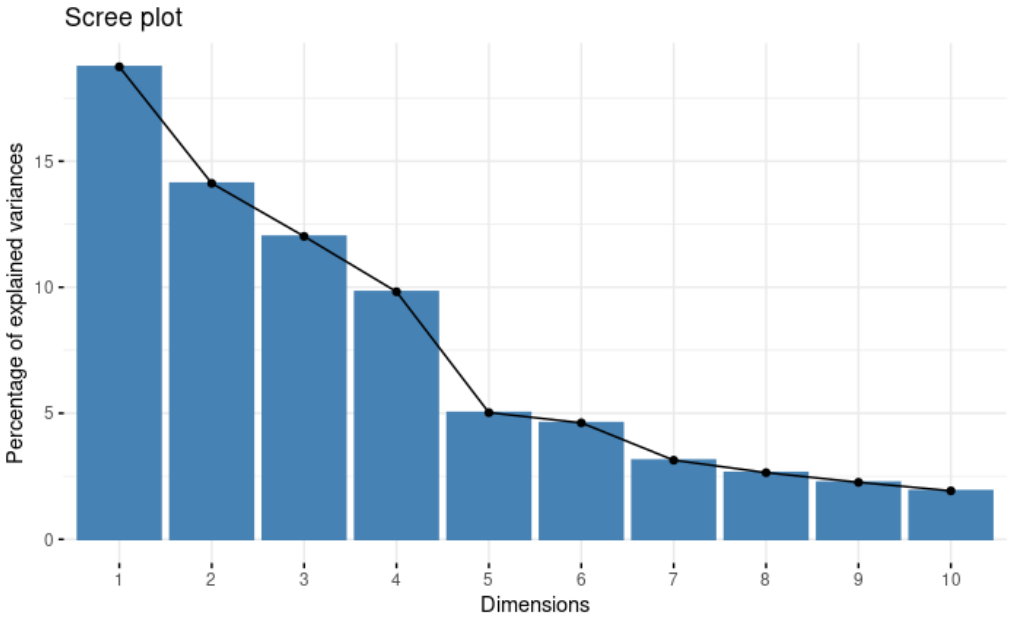
Supplementary Figure 39: Groups representation of the Multiple Factor Analysis (MFA). MFA integrates pancreatic associated microbiota at Order taxonomic level (Ord_Microbiome), with gene expression data (mRNA), and the Group of samples. Dimension 1 accounts for 25.38% and Dimension 3 accounts for 13.26% of variability.

Supplementary Figure 40:



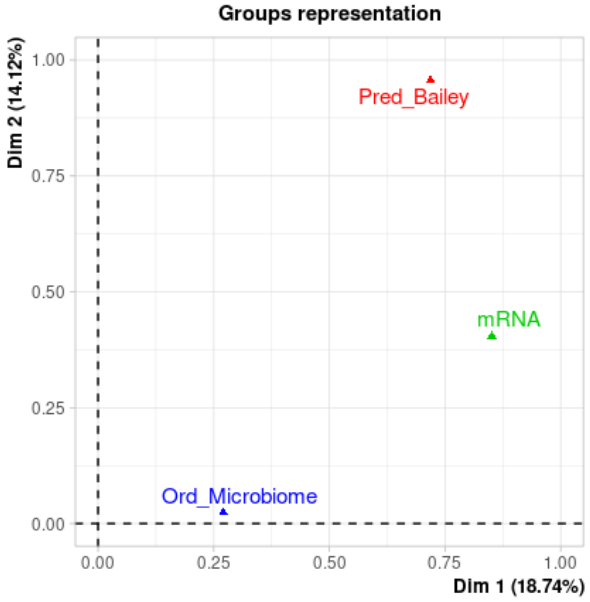
Supplementary Figure 40: Individuals representation in the Multiple Factor Analysis (MFA). Individual samples are represented as dots, and color coded based on their group of sample. OTUs at order level are represented as black triangles. Dimension 1 accounts for 25.38% and Dimension 3 accounts for 13.26% of variability.

Supplementary Figure 41:



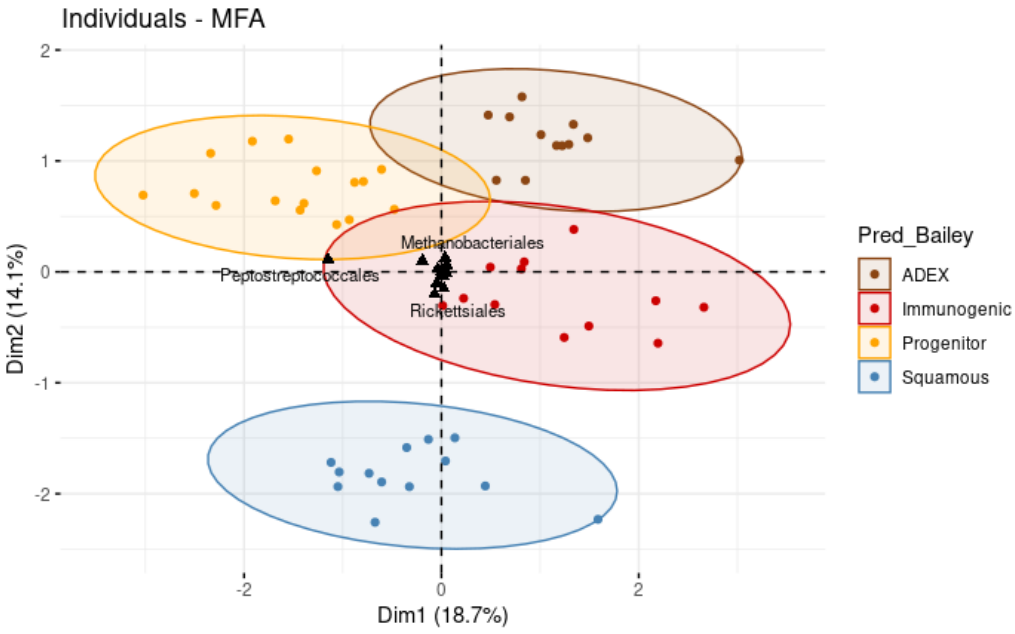
Supplementary Figure 41: Scree plot with the percentage of explained variation by each one of the dimensions derived from the Multiple Factor Analysis (MFA). Dimension 1 accounts for 18.74%, Dimension 2 accounts for 14.12% and Dimension 3 accounts for 12.01% of variances.

Supplementary Figure 42:



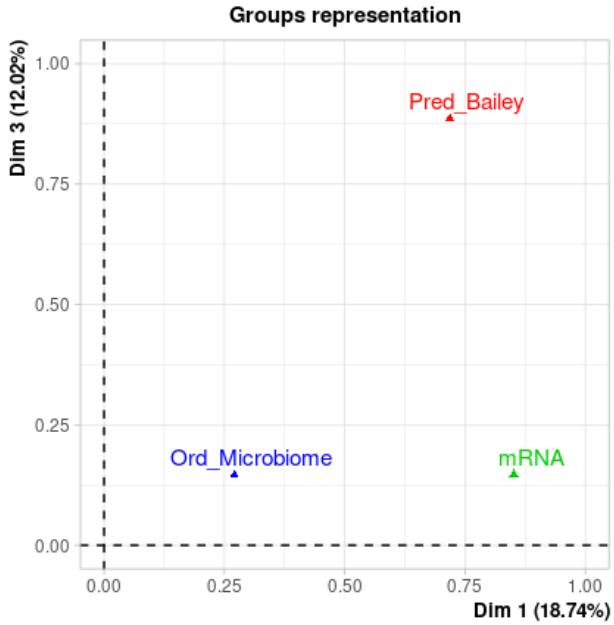
Supplementary Figure 42: Groups representation of the Multiple Factor Analysis (MFA). MFA integrates pancreatic associated microbiota at Order taxonomic level (Ord_Microbiota), with gene expression data (mRNA), and the Prediction of PC subtype on Bailey's signature (Pred_Bailey). Dimension 1 accounts for 18.74% and Dimension 2 accounts for 14.12% of variability.

Supplementary Figure 43:



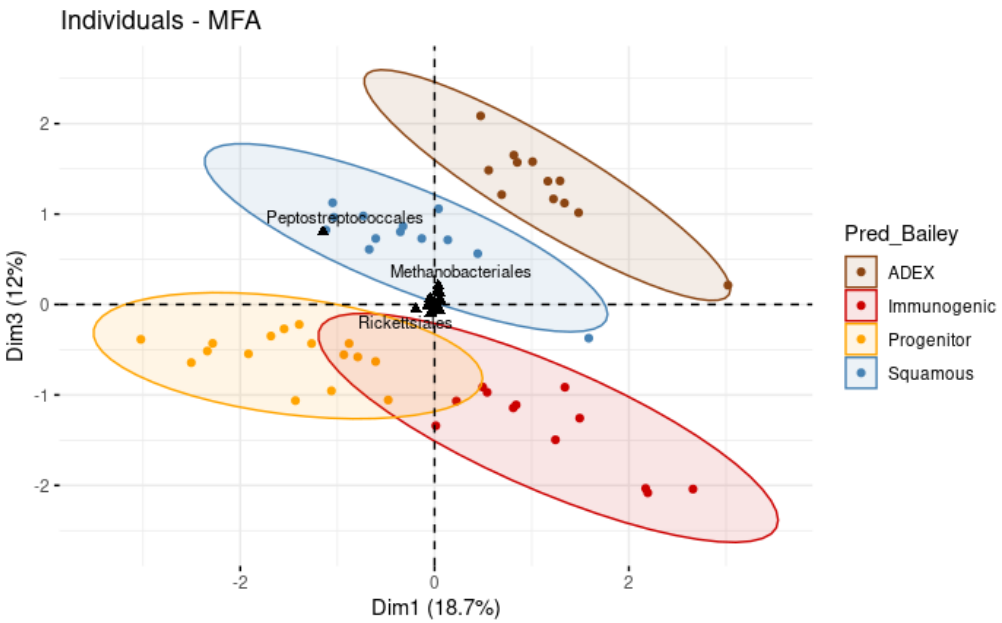
Supplementary Figure 43: Individuals representation in the Multiple Factor Analysis (MFA). Individual samples are represented as dots, and color coded based on their predicted PC subtype (Pred_Bailey). OTUs at order level are represented as black triangles. Dimension 1 accounts for 18.7% and Dimension 2 accounts for 14.1% of variability.

Supplementary Figure 44:



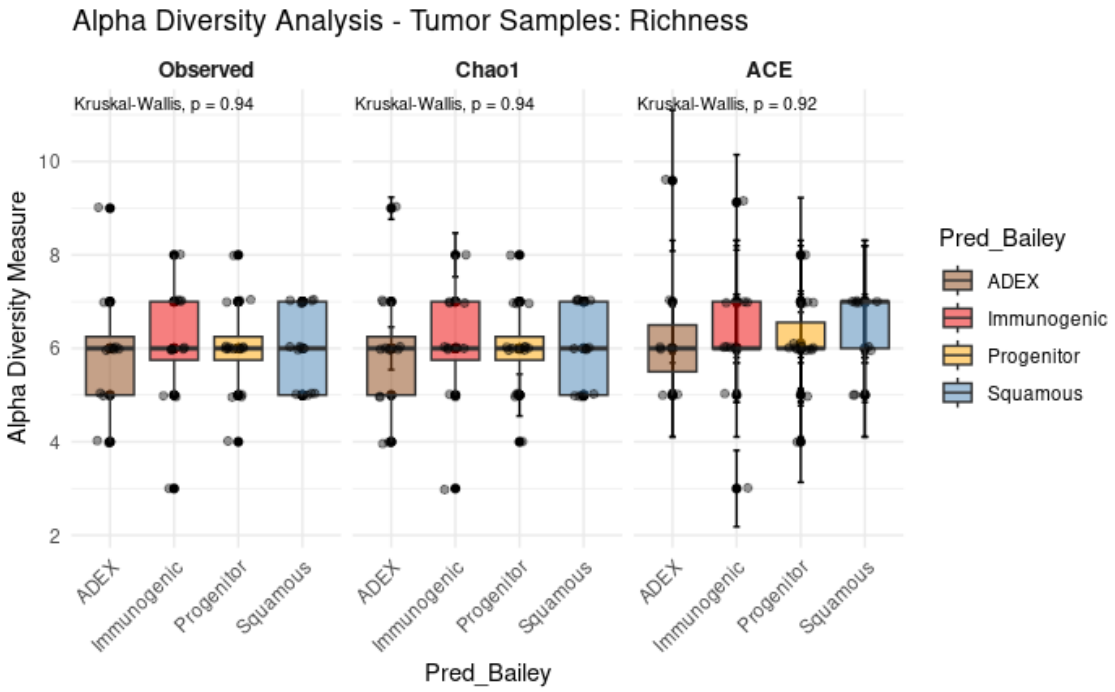
Supplementary Figure 44: Groups representation of the Multiple Factor Analysis (MFA). MFA integrates pancreatic associated microbiota at Order taxonomic level (Ord_Microbiota), with gene expression data (mRNA), and the prediction of the PC subtypes (Pred_Bailey). Dimension 1 accounts for 18.74% and Dimension 3 accounts for 12.02% of variability.

Supplementary Figure 45:



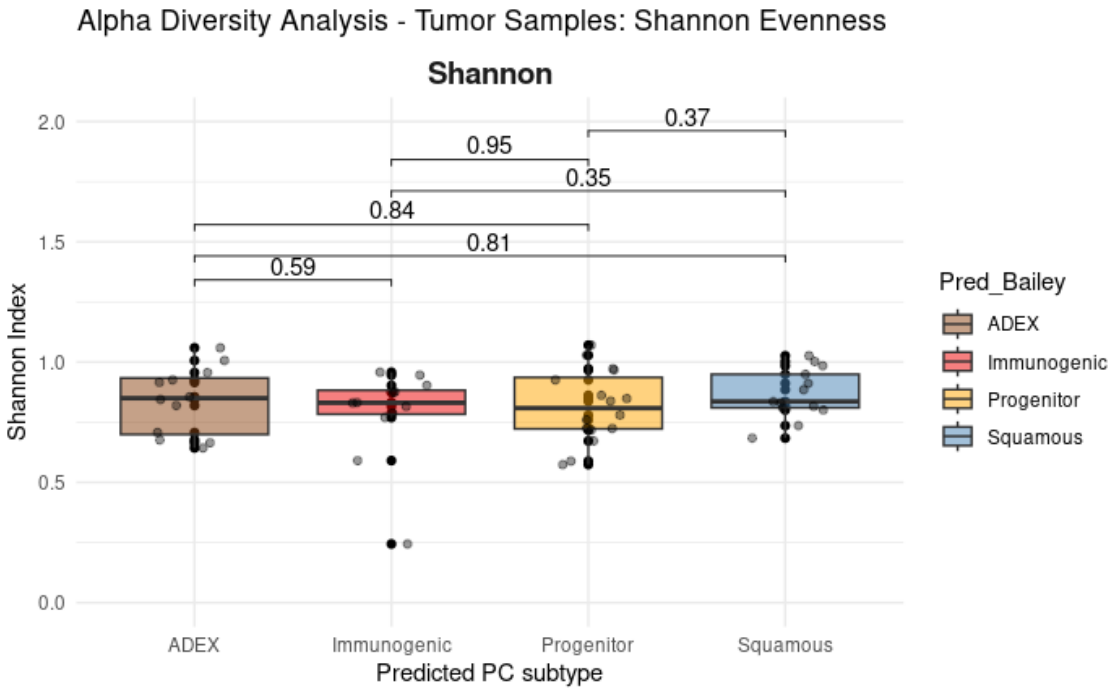
Supplementary Figure 45: Individuals representation in the Multiple Factor Analysis (MFA). Individual samples are represented as dots, and color coded based on their predicted PC subtype (Pred_Bailey). OTUs at order level are represented as black triangles. Dimension 1 accounts for 18.7% and Dimension 3 accounts for 12.01% of variability.

Supplementary Figure 46:



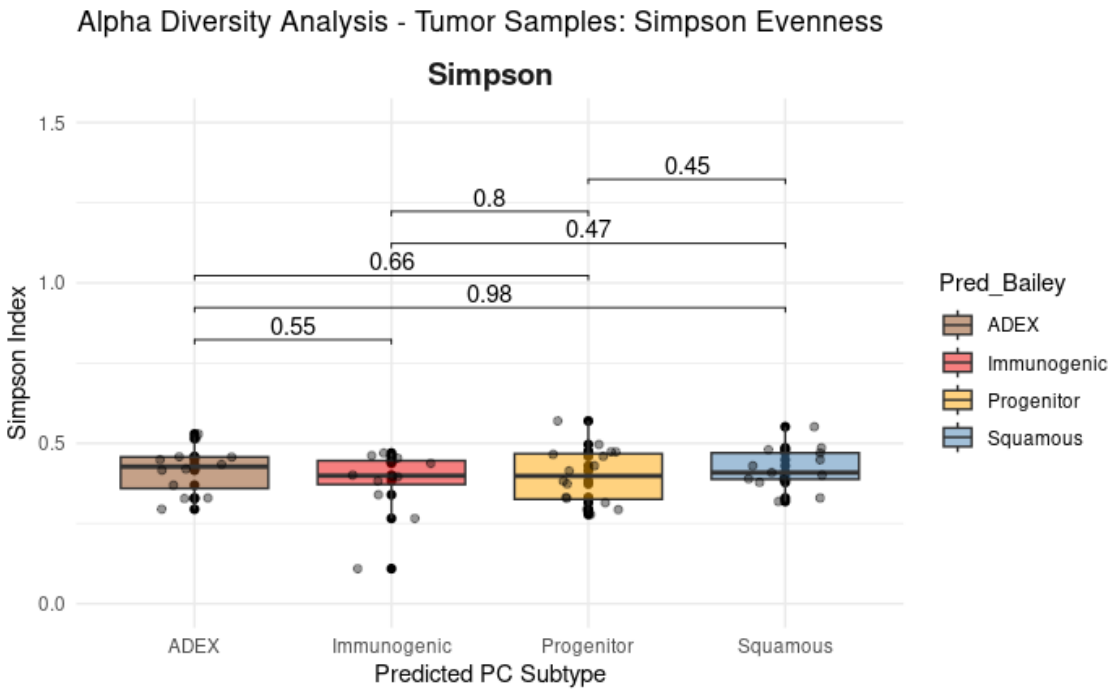
Supplementary Figure 46: Boxplots with the distribution of alpha diversity richness indexes among predicted PC subtypes. Kruskal Wallis test provides statistical significance p-values over comparison. On the right, the Observed richness values, in the middle, Chao1 index values, and on the left, ACE index values

Supplementary Figure 47:



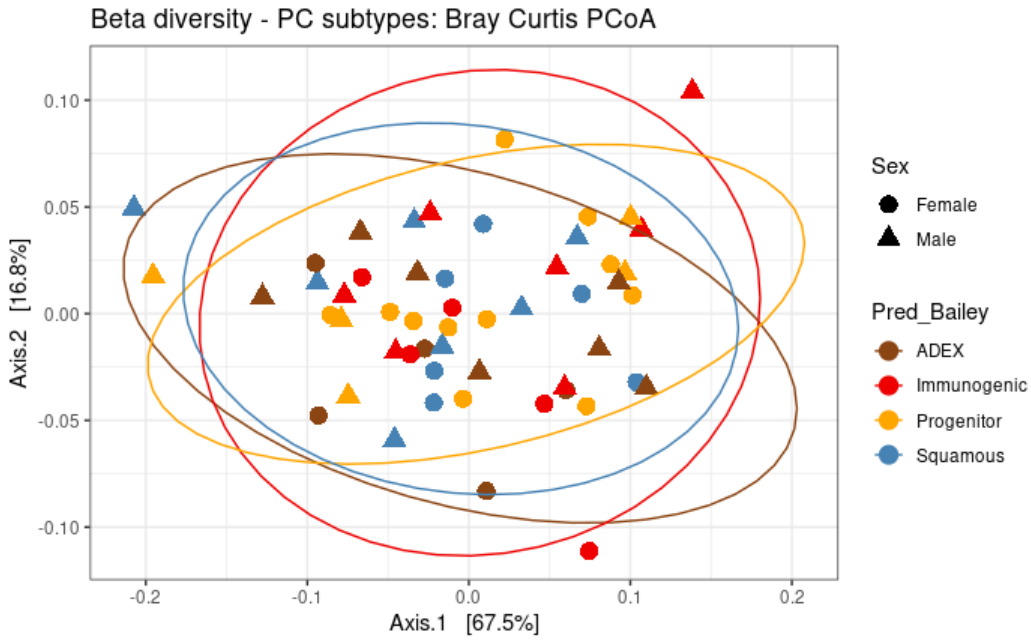
Supplementary Figure 47: Boxplots with the distribution of alpha diversity evenness Shannon index among predicted PC subtypes. Kruskal Wallis test provides statistical significance p-values after multiple comparisons. a) Simpson evenness index analysis. b) Shannon evenness index analysis.

Supplementary Figure 48:



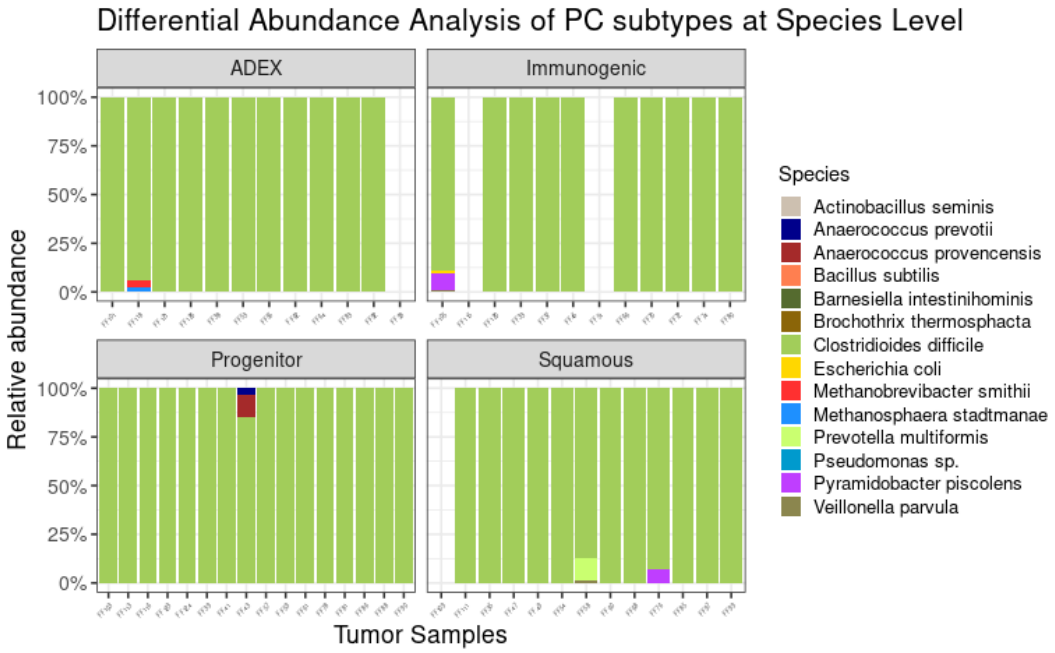
Supplementary Figure 48: Boxplots with the distribution of alpha diversity evenness Simpson index among predicted PC subtypes. Kruskal Wallis test provides statistical significance p-values after multiple comparisons. a) Simpson evenness index analysis. b) Shannon evenness index analysis.

Supplementary Figure 49:



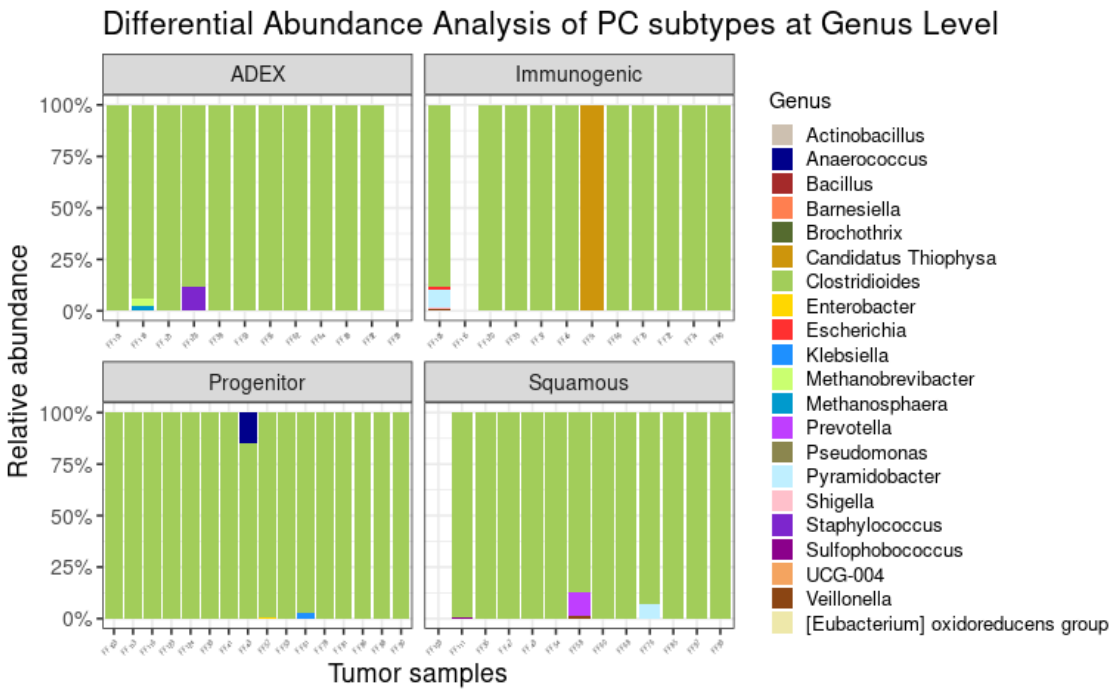
Supplementary Figure 49: Principal Coordinates Analysis (PCoA) with Bray Curtis distances to evaluate the beta diversity (overall differences in taxonomic composition) between predicted PC subtypes. Dimension 1 gathers 67.5% of the variability, and Dimension 2 gathers 16.8% of the variability. Samples are colored based on predicted PC subtype, and shaped according to sex.

Supplementary Figure 50:



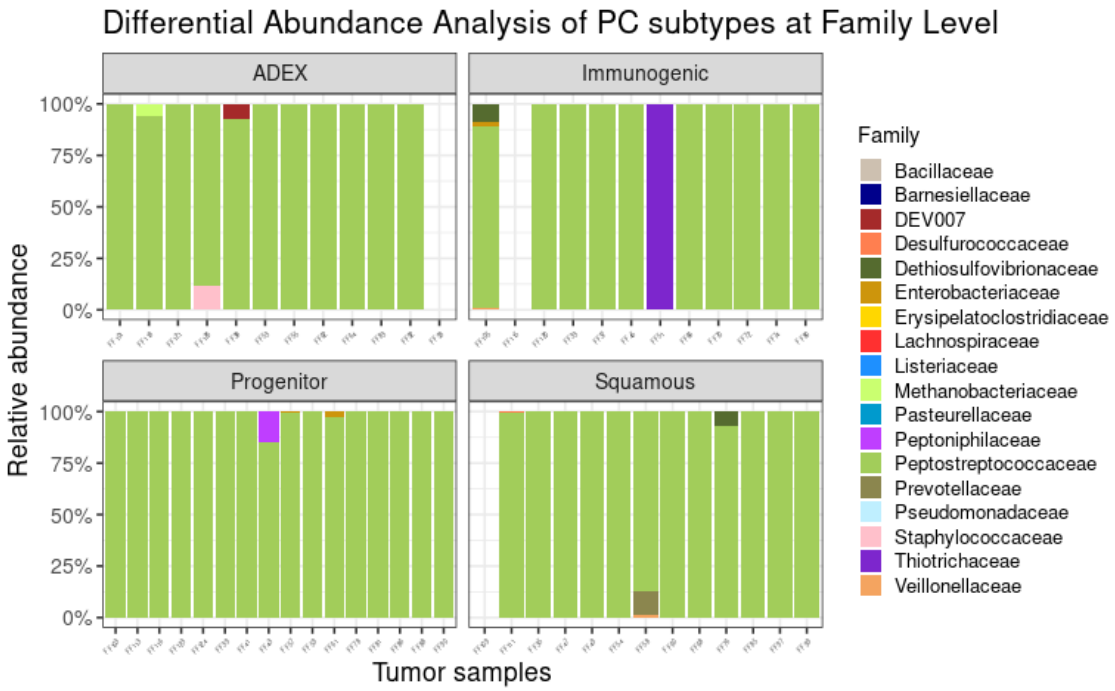
Supplementary Figure 50 Differential Abundance Analysis of the relative abundance of taxonomic composition of the tumoral samples at species rank. Samples are organized by predicted PC subtypes.

Supplementary Figure 51:



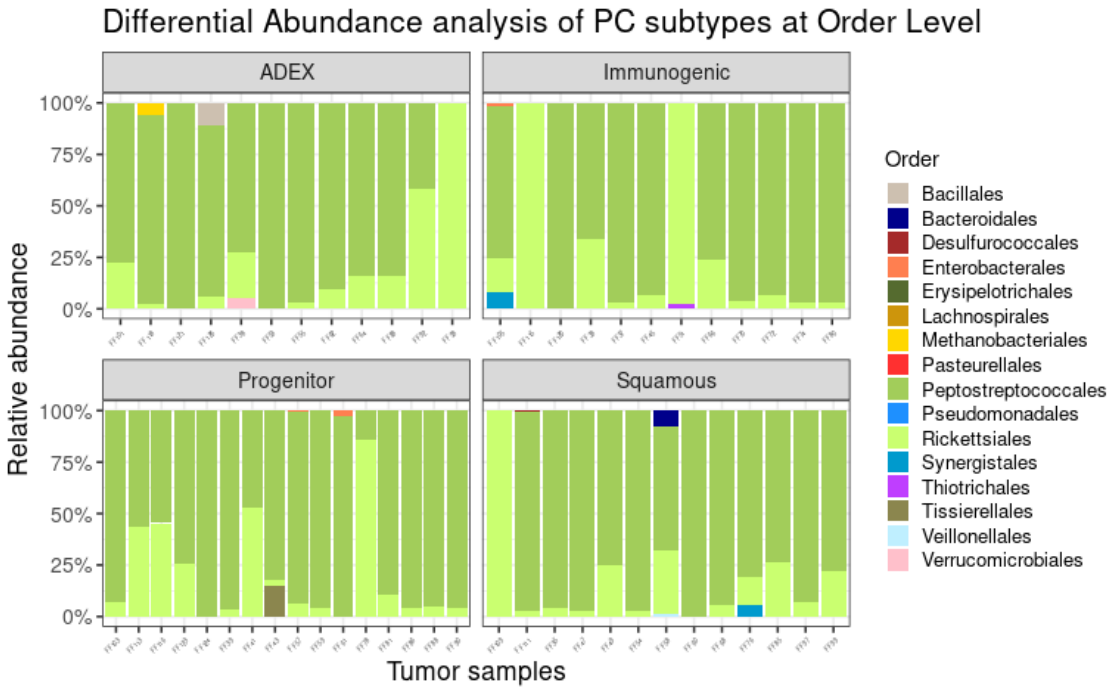
Supplementary Figure 51: Differential Abundance Analysis of the relative abundance of taxonomic composition of the tumoral samples at genus rank. Samples are organized by predicted PC subtypes.

Supplementary Figure 52:



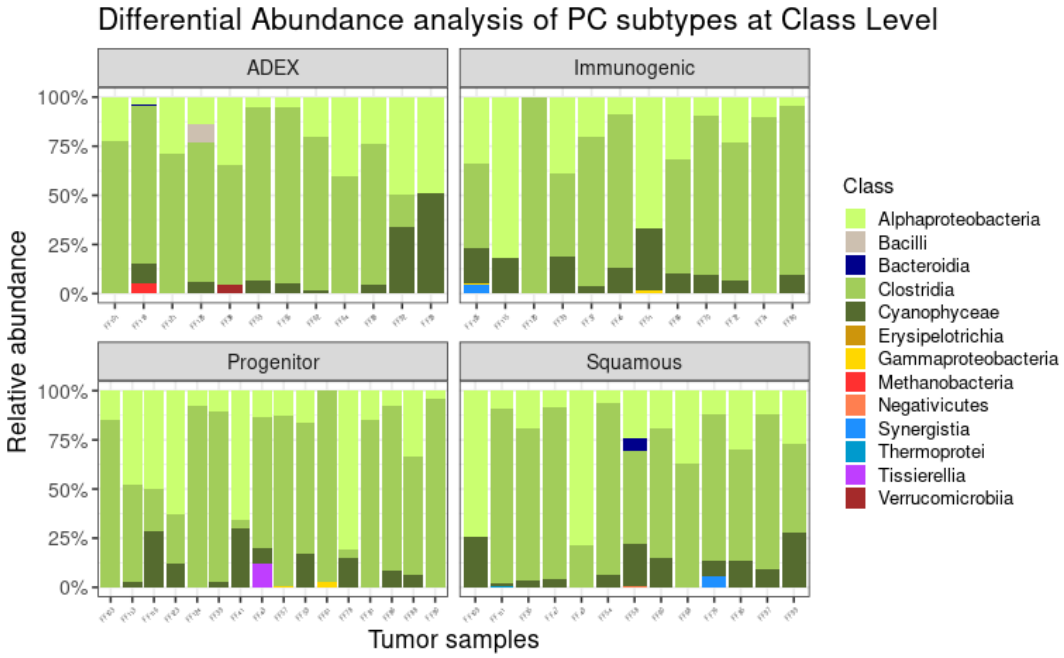
Supplementary Figure 52: Differential Abundance Analysis of the relative abundance of taxonomic composition of the tumoral samples at family rank. Samples are organized by predicted PC subtypes.

Supplementary Figure 53:



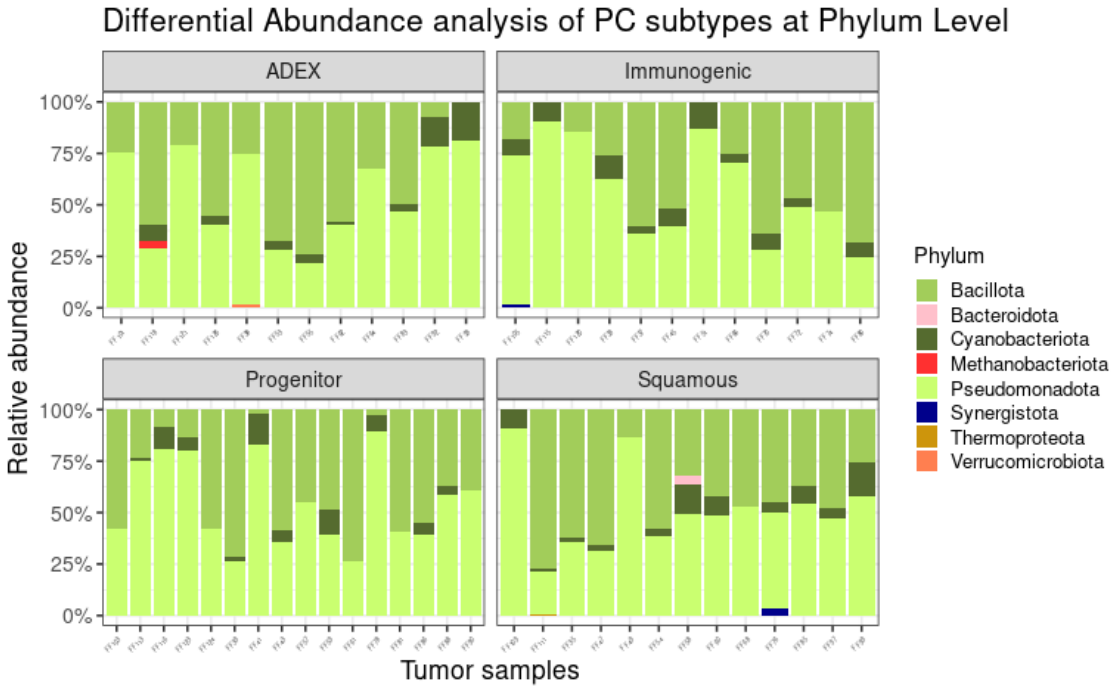
Supplementary Figure 53: Differential Abundance Analysis of the relative abundance of taxonomic composition of the tumoral samples at order rank. Samples are organized by predicted PC subtypes.

Supplementary Figure 54:



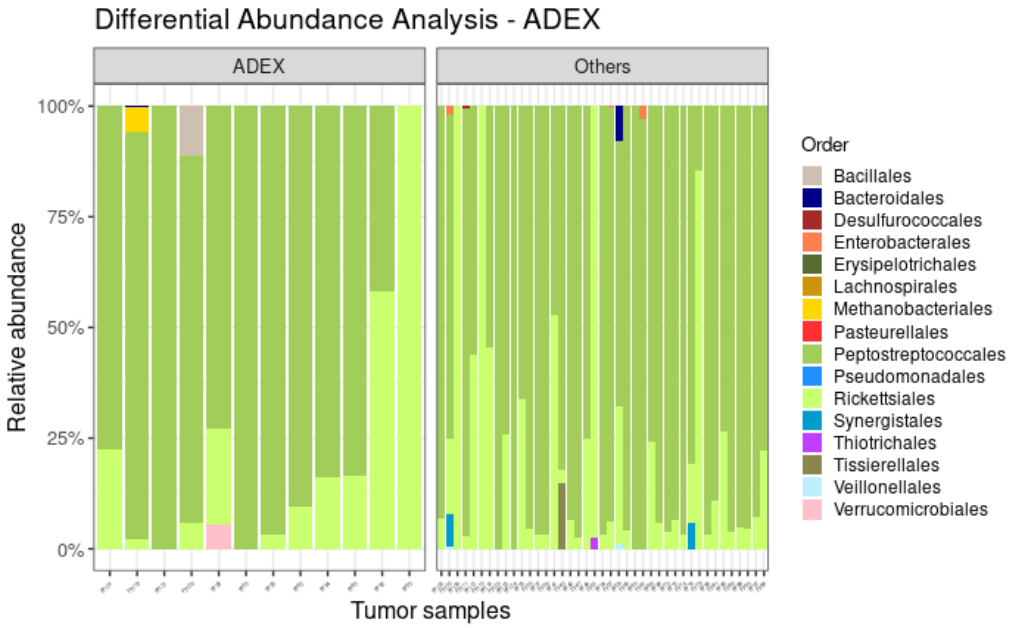
Supplementary Figure 54: Differential Abundance Analysis of the relative abundance of taxonomic composition of the tumoral samples at class rank. Samples are organized by predicted PC subtypes.

Supplementary Figure 55:



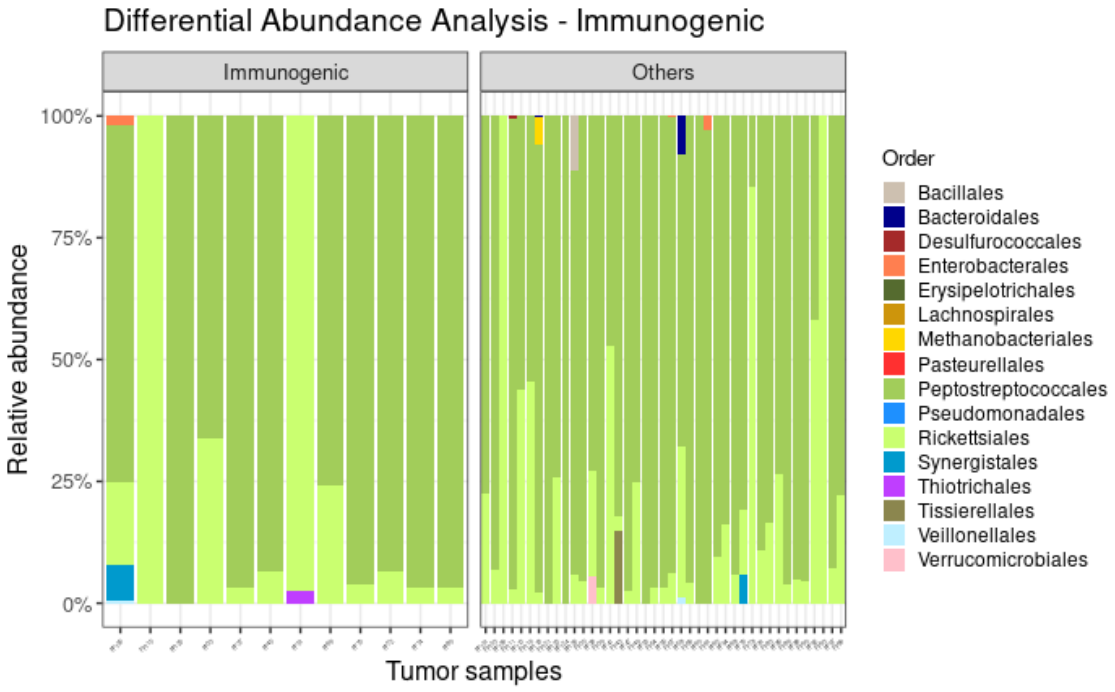
Supplementary Figure 55: Differential Abundance Analysis of the relative abundance of taxonomic composition of the tumoral samples at phylum rank. Samples are organized by predicted PC subtypes.

Supplementary Figure 56:



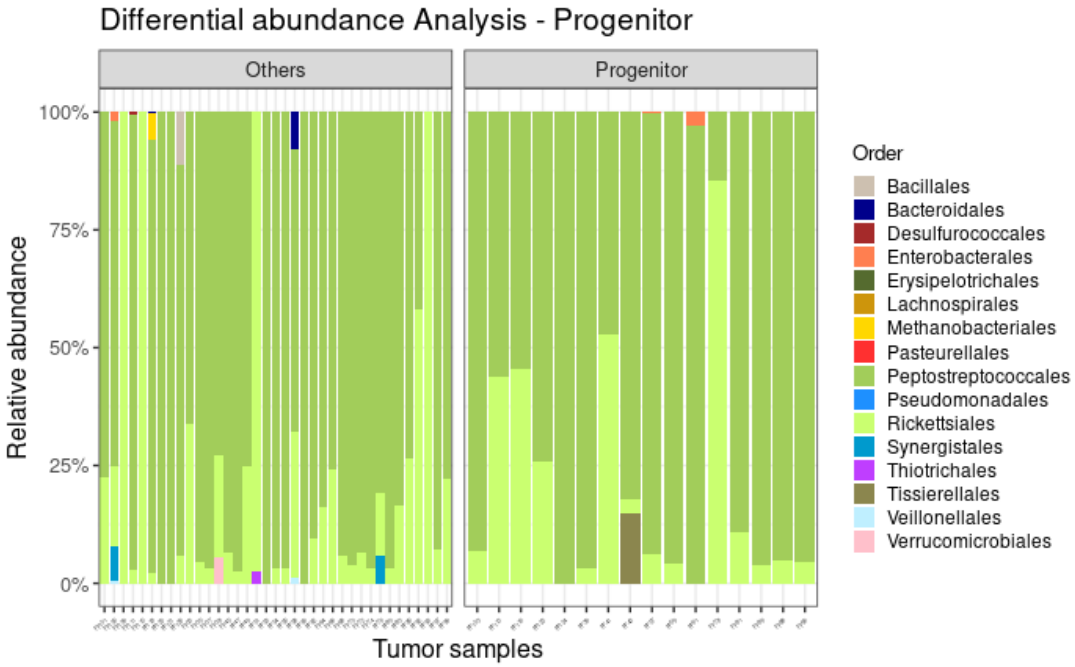
Supplementary Figure 56: Differential Abundance Analysis of the relative abundance of taxonomic composition of the tumoral samples at order rank. Pairwise comparison: ADEX subtype against the remaining tumoral samples.

Supplementary Figure 57:



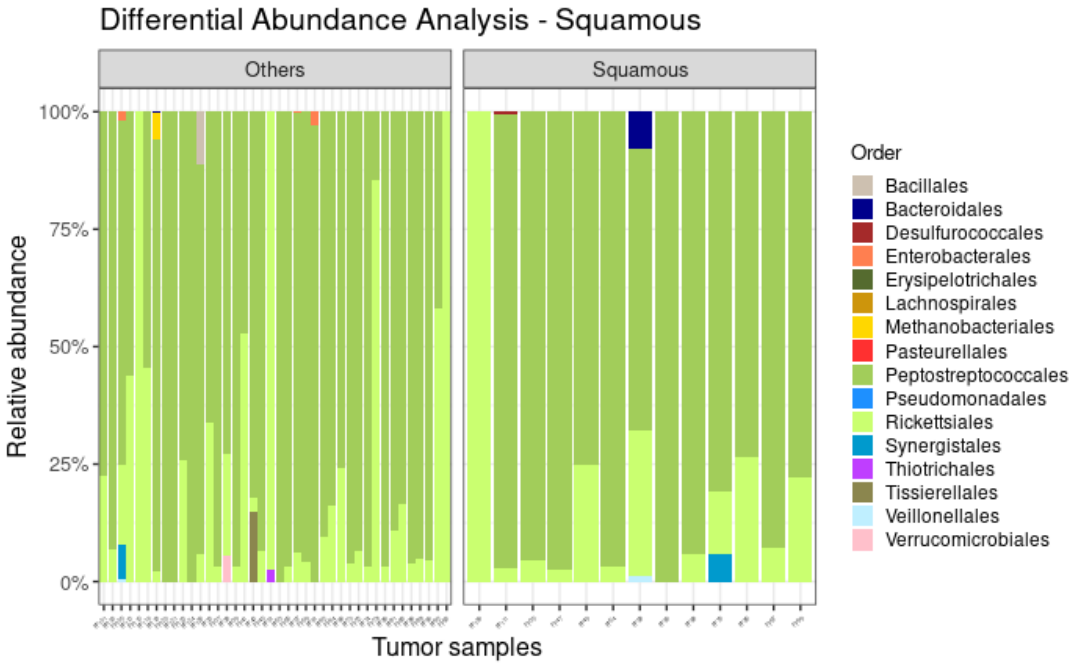
Supplementary Figure 57: Differential Abundance Analysis of the relative abundance of taxonomic composition of the tumoral samples at order rank. Pairwise comparison: immunogenic subtype against the remaining tumoral samples.

Supplementary Figure 58:



Supplementary Figure 58: Differential Abundance Analysis of the relative abundance of taxonomic composition of the tumoral samples at order rank. Pairwise comparison: progenitor subtype against the remaining tumoral samples.

Supplementary Figure 59:



Supplementary Figure 59: Differential Abundance Analysis of the relative abundance of taxonomic composition of the tumoral samples at order rank. Pairwise comparison: squamous subtype against the remaining tumoral samples.