Using multi-omics to identify bile- and liver- biomarkers that correlate with biliary viability of ECD livers

Jarno Duiker [1], Ramon Reilman [1], Tai Vo [1], Yamila Timmer [1]
[1]Hanze Groningen

2025-06-04

Abstract

Introduction

Liver transplantation is the first treatment option for patients with liver dysfunctionality, e.g. in patients with liver cirrhosis or acute liver failure. There is a shortage of donor livers that meet required standards, so now extended criteria donors (ECD) are increasingly used, these consist of livers that do not meet the normally required standards and are often extracted from cardiac death donors (DCD) but also include e.g. elderly donors. Using ECD can provide a solution for the shortage of liver donors, however, there are also risks to these donor livers, such as the increased risk of primary organ dysfunction, which is a quick onset dysfunctionality of the graft post-transplant. ECD livers are also more susceptible to biliary complications and to ischemia-reperfusion injury, which is damage that occurs when an organ with a lack of oxygen gets reperfused with oxygen again (C. Shen et al. (2023)). Certain techniques, such as normothermic machine perfusion (NMP) can reduce the risk of post-transplant complications. NMP attempts to restore and/or maintain cell metabolism by simulating the natural environment of the liver, by maintaining a temperature of 37 degrees Celsius and providing the organ with the required oxygen and nutrition. NMP is considered a better alternative to static cold storage (SCS), as it reduces the risk of ischemia-reperfusion injury, early allograft dysfunction (EAD), and biliary complications, improving general graft viability. In various research, it has been shown that (prolonged) NMP can restore liver viability, which allows the transplantation of livers that have initially been deemed as not eligible for transplanting. NMP also allows evaluation of the donor liver pre-transplant, further reducing transplantation risks (C. Shen et al. (2023)).

Thorne et al. (2023) describe how NMP is used for assessing organ viability before transplantation of the liver. Here, the general viability of the organ is divided into two distinct measurements. First, hepatocellular viability, which focuses on the metabolic functionality of the parenchyma, or the 'functional tissue' of the liver. The second measurement focuses on biliary viability, which focuses on the state of the cholangiocytes, the epithelial cells of the bile ducts. Biliary viability is measured based on physiological factors such as bile pH and glucose reabsorption but as of today, there is no golden standard, and measurement criteria differ between centers. Biliary viability, unlike hepatocellular viability, is not yet adopted in all clinical trials that work with ECD donor livers; however, it has been shown that the use of both hepatocellular as well as biliary viability in the selection of ECD donors has led to a lower incidence of post-transplant complications affecting cholangiocytes, also known as cholangiopathies.

To gain a more comprehensive understanding of the molecular changes that contribute to biliary viability, a multi-omics approach is employed. Multi-omics integrates data from various biological levels, such as genomics, transcriptomics, metabolomics, and proteomics, to provide a view of complex biological systems, cellular responses, and disease progression. Specifically, this paper leverages a multi-omics strategy to identify bile and liver-based biomarkers that correlate with biliary viability. This is achieved by integrating and

analyzing data derived from various omics techniques, including principal component analysis (PCA), partial least squares discriminant analysis (PLS-DA), differential expression analysis (DEA), and other relevant statistical and bioinformatics analyses to identify differentiating genes and proteins.

We are looking to find similar results as Thorne et al. (2023), to confirm that our results are comparable. Since they accessed the viability based partly on 'metabolic functionality' we expect to see upregulations of genes that are part of metabolic pathways in the liver, in livers with high biliary viability. Such as the glycogen production pathway. We also expect to find DEGs with functions related to bile production and/or epithelial tissue, as the measured high/low biliary viabilities are based on factors such as bile pH and capacity of epithelial bile duct cells, also known as cholangiocytes.

Methods

Proteomics

For the proteomics analysis, samples collected by Thorne et al. (2023) were used. These samples consist of raw mass spectrometry data obtained from bile samples of donor livers. Samples were taken at different time points during NMP, at 30 minutes, and 150 minutes, and only for livers that ended up being transplanted at the end time point. The proteomics analysis was performed using only the 30- and 150-minute samples, as many low biliary viability samples did not end up being transplanted. In total, there are 43 samples for 150 min and 44 samples for 30 min. These samples have previously undergone protein identification and quantification. Samples containing more than 70% missing values were removed, which led to the removal of 2 samples for 150 min (n = 41) and 1 sample for 30 min (n = 43). Features missing in more than 4 samples in one or both biliary viability groups were removed. The data is normalized across samples using median centering, this method is used to center the data over the median of the distribution of each sample (Dubois et al. (2022)). The mean percentage of missingness for the features is 59% for the 30 min samples and 45% for the 150 min samples. The features were then normalized by 'variance stabilizing transformation' using vsn (v.3.72.0), this package uses a robust variant of the maximum-likelihood estimator (Huber et al. (2002)). Missing values in the data were identified to mostly consist of MNAR values (missing not at random), making the missing data left-censored. Thus, imputation was performed using DEP (v.1.26.0) (Zhang et al. (2018)) with the function "MinDet", which uses a deterministic minimal value approach. Here all the missing values are replaced with a minimal observed value per sample. Principal component analyses (PCA) were performed on the normalized and imputed data using the mdatools (v 0.14.2) package in R (Kucheryavskiy (2020)).

Potential biomarker protein evaluation was done using a confusion matrix framework to assess their viability in categorizing livers into high and low biliary viability. The tested thresholds were set in intervals of 0.01 in the range of the expression value of the corresponding potential biomarker protein, and predictive performance was done by calculating accuracy, precision, and AUC-ROC from the resulting confusion matrix. Statistical significance was assessed using Fisher's exact test to confirm whether classification outperformed chance.

Using limma (v.3.60.6) (Ritchie et al. (2015)) a DEA was performed on all features that passed the filtering steps. In this DEA, two distinct groups are compared to each other, the comparison consists of the samples with high biliary viability against samples with low biliary viability. First, a linear model is fitted to each protein. The Empirical Bayes method is applied for smoothing estimates of discrete probabilities and lastly, a false discovery rate correction is applied. After determining which genes are down/upregulated, a pathway analysis was performed. The results were visualized using e.g. volcano plots and pathview. The supervised discrimination method Partial Least Squares Discriminant Analysis (PLS-DA), based on PLS regression, was applied using the mdatools (version 0.14.2) R package (Kucheryavskiy (2020)). Normalized, imputed proteomics data was used as input with samples classified into high biliary viability and low biliary viability classes. The cross validation used was - full (leave one out). The model generated from this was analysed with different plots that showed its parameters and explained what proteins are important and what proteins are not.

Transcriptomics

Liver tissue biopsies, stored in the cold, were taken right before the start of the NMP. RNA from these samples was isolated using the Qiagen RNeasy Lipid Tissue Mini Kit. RNA was sequenced using the Illumina NextSeq 500. The resulting reads were trimmed with TrimGalore (v.0.6.7) (Krueger et al. (2023)), and deduplicated with SeqKit (v.2.4.0) (W. Shen, Sipos, and Zhao (2024)). They were then mapped with Hisat2 (v2.2.1) ("HISAT2" (n.d.)) to the human genome h38. The counts were extracted from the resulting files using featureCounts (v.2.0.3) (Liao, Smyth, and Shi (2014)). MultiQC (v 1.14) (Ewels et al. (2016)) was used to show a comprehensive overview of the quality of multiple FASTQ samples. This generated an HTML report that summarized the results from various QC reports. Showing the combined results across all samples. This report shows if there are potential issues such as adapter contamination or high repeat rates among other possibilities.

RNA-sequencing data was aligned to the GRCh38 ("Homo Sapiens Genome Assembly GRCh38.p14" (n.d.)) using HISAT2 (v 2.2.1) (Kim et al. (2019)). HISAT2 was used to build a splice-aware genome index, using the hisat2-build command with default parameters. This build genome index was then used to align the RNA-sequencing FASTQ files. This was done with certain parameters, -U (meaning unpaired FASTQ files), and using the –phred33 score read. The output format from this aligning was a SAM file. All genes with a variance of 0 were removed from the data, this is done to speed up the following analyses and tools. The data was log2 transformed and scaled using R for a PCA. A differential gene expression analysis was performed on the raw count data using the R-package DESeq2 (v.1.38.3) (Love, Huber, and Anders (2014)).

Results

DEA

The DEA in total evaluated 2255 features for the 30-minute samples and 1610 features for the 150-min samples. The DEA using the 30-minute samples did not yield any significantly differently expressed genes. The 150-min samples contained a total of 44 DEGs (Fig.1), details on e.g. p-values and logFC per differentially expressed gene can be found in Table 1.

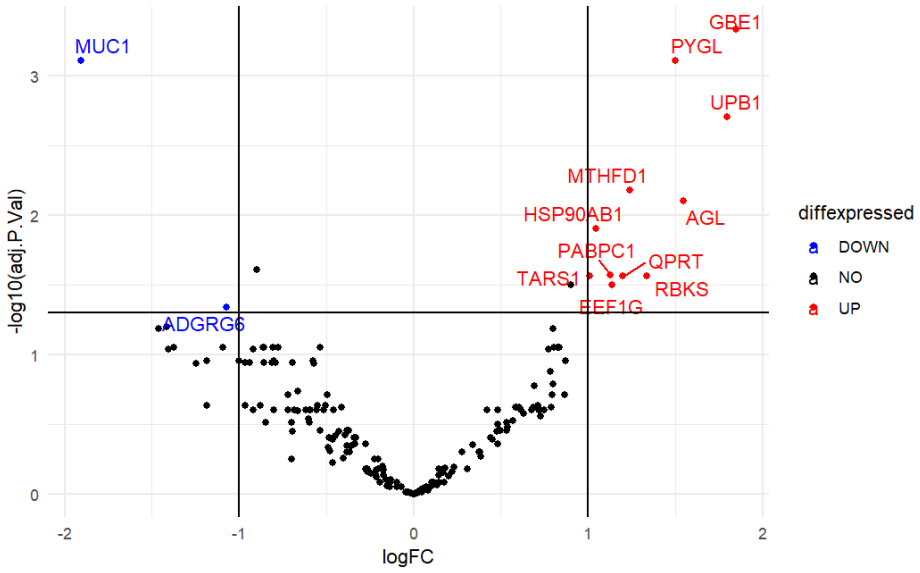

Figure 1: Volcano plot showing all significantly upregulated- and downregulated genes in the comparison of high biliary viability against low biliary viability in the samples taken during NMP after 150 minutes. Genes were deemed as significantly upregulated if the adjusted p-value was ≤ 0.05 and the $\log_2$ fold change was ≥ 1.0. Genes were deemed as significantly downregulated if the adjusted p-value was ≤ 0.05 and the $\log_2$ fold change was ≤ -1.0.

PCA

Figure 2 shows that the group with low biliary viability is a bit less spread throughout the second component. Samples with a high biliary viability show to have a bit higher spread over the second component compared to the samples with low viability. 3 samples with a high biliary viability and 1 sample with a low biliary viability cause a high variance in the first component. A perfect linear separation of the 2 groups is not possible.

From the PCA that was performed on the 150-minute samples (Figure 3) it becomes clear that there appears to be more of a separation between the 2 groups, compared to Figure 2. Still, there does not appear to be any clear clustering between the groups, only with a couple of samples.

Differential gene expression analyses on RNA data revealed 0 significant changes in the expression of genes between a high and low biliary viability

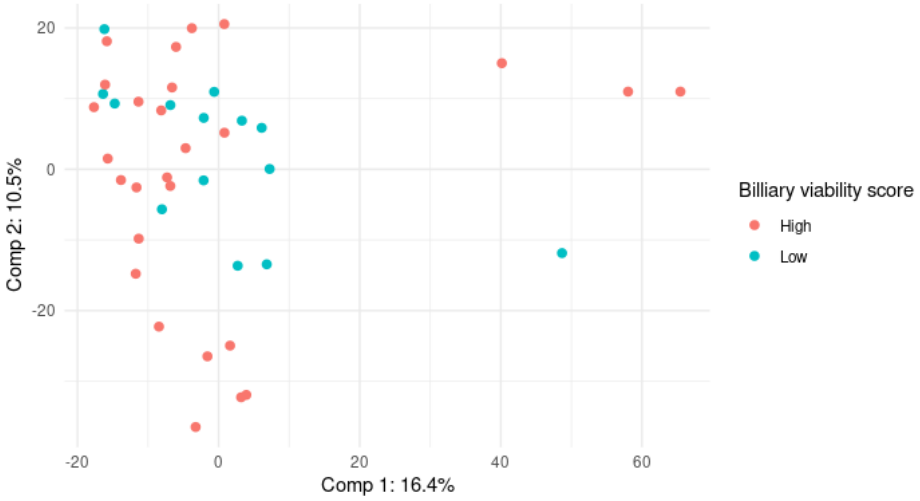

Figure 2: PCA performed on proteomic samples 30 minutes after starting NMP. Shows how the samples are spread over component 1 and component 2. Biliary viability is displayed with color.

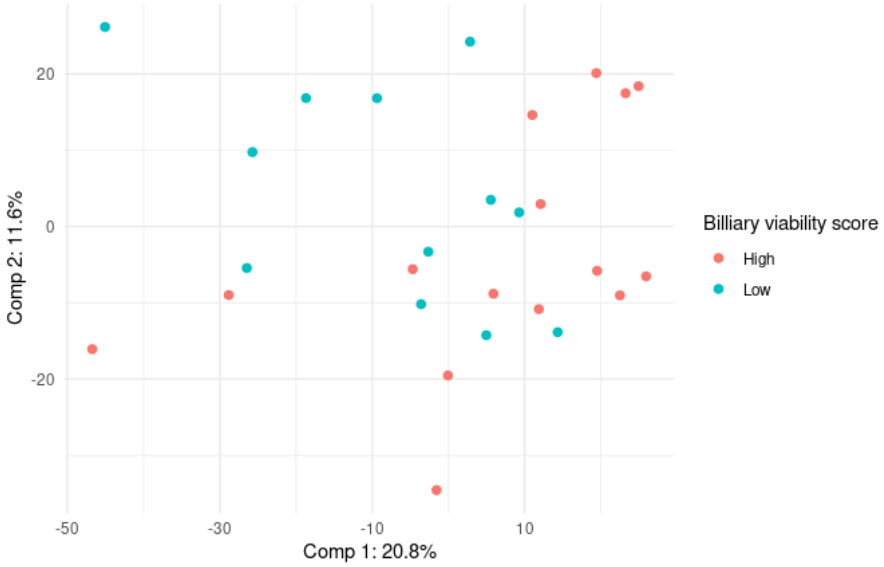

Figure 3: PCA performed on proteomic samples 150 minutes after starting NMP. Shows how the samples are spread over component 1 and component 2. Biliary viability is displayed with color.

PLS-DA

PLS-DA was performed on the normalized, imputed proteomics data. This data was then split into two groups, high biliary viability and low biliary viability. Figure 4 shows a comparison between the 3rd component, which contains 4.62% of the info, and Component 1, which contains 21.9% of the info. The mentioned info is the amount of each component that relates to the original data file. The reason for this comparison is that component 1 - component 2 and component 2 - component 3 showed nothing significant

and no groups; therefore, there were no clear results from these comparisons. In figure 5 a clear divide is visible in the middle.

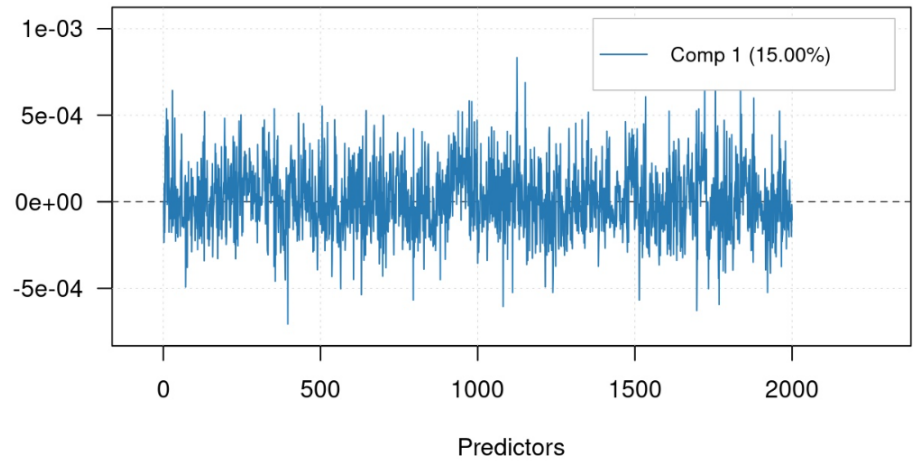

Figure 4: Weights of each protein, the weight means how much influence it has on each of the components. The names have been hidden for readability. But each spike shows an important protein.

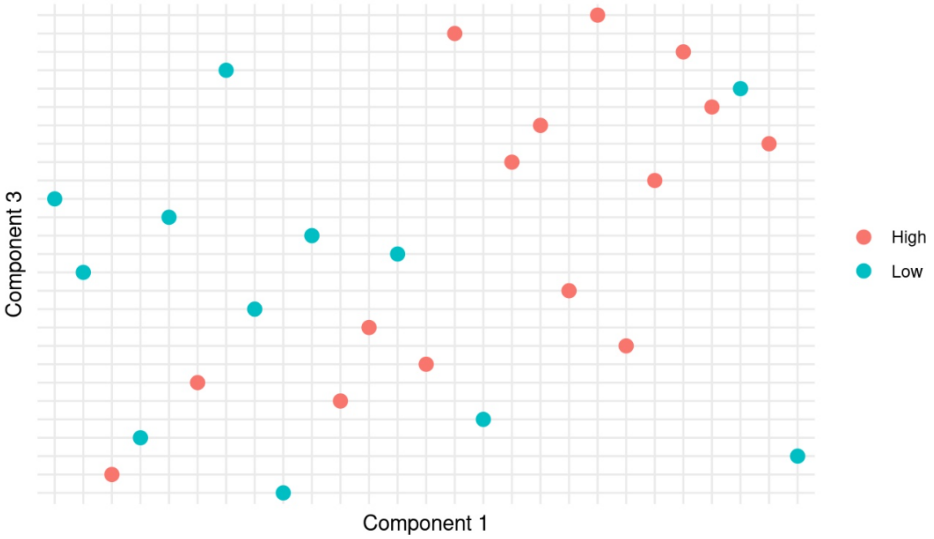

Figure 5: PLS-DA Score plot showing component 1 – component 3, performed on the proteomic samples at 150 min after starting NMP. Showing how the samples are laid out over components 1 and 3. Viability is shown with colour.

Confusion Matrix

To identify proteins and thresholds with potential as biomarkers, we filtered the dataset using a p-value cut-off of 0.05. This evaluation was performed on the 150-minute time point dataset, as no proteins-threshold combination met this minimum p-value requirement at the 30-minute time point. From the filtered 150-minute dataset, 14 unique thresholds were identified where proteins demonstrated statistically significant p-values (< 0.05)(Table 2). These protein-thresholds combinations correspond to optimal expression cut-offs used to classify samples based on protein abundance.

Among the protein-thresholds combinations meeting the significance criteria, SND1 and SNRPD1 consistently appeared at multiple thresholds. Both proteins demonstrated high classification metrics across these thresholds, indicating significant categorization performance ability. Specifically, SND1 achieved the highest area under the receiver operating characteristic curve (AUROC) of 0.95 indicating a high ability to distinguish between sample classes. It was observed at two thresholds: 13.89 and 13.79, with precision values of 0.67 and 0.75, and accuracy values of 0.85 and 0.89, respectively. The specificity for both thresholds was perfect (1.0), with zero false positive rates. Similarly, SNRPD1 was found at three thresholds (13.72, 13.62, and 13.52), reaching a maximum AUROC of 0.933. Precision ranged from 0.75 to 0.83, while accuracy improved from 0.89 to 0.93 across these thresholds. Specificity remained at 1.0 with zero false positive rates. The protein HSD17B13 was observed with a statistically significant p-value and showed moderate performance metrics, including an AUROC of 0.775, precision of 0.67, and accuracy of 0.85.

In contrast are the proteins ACE2, MMP7, and PUF60. Although they met the p-value criteria, they demonstrated notably lower classification performance. These proteins showed AUROC values below 0.7 (approximately 0.65–0.66), despite high precision values (0.93). However, their specificity was significantly lower (0.44), resulting in high false positive rates (approximately 0.56).

Discussion

Two out of the three genes with the highest average AUC for the comparison of ECD donors across biliary viability groups (high/low) as found by Thorne et al. (2023) are also found in this analysis, which are the genes MUC1 and FCGBP. The third gene with the highest average AUC, MUC5B, was not found significantly differentially expressed in our research. Thorne et al. (2023) also found that MUC5B, MUC1, and LCN2 are the only genes that show significant differences in expression between biliary viability groups, regardless of histological BDI, we did not compare histological BDI and focused on biliary viability, but we also found LCN2 to be a statistically significant DEG.

MUC1, FCGBP and LCN2 were significantly upregulated (p-adjust: 7.81×10-3, logFC: 2.49) (p-adjust:0.025, logFC: 1.96)(p-adjust: 4.38x10-4, logFC: 1.44) in ECD livers with high biliary viability, confirming what Thorne et al. (2023) found. MUC1 is associated with cell signalling and proliferation pathways and expression of MUC1 was predominantly found in cholangiocytes (Thorne et al. (2023)). FCGBP (IgGFc-binding protein) has been linked to contribute to wound healing in order to maintain epithelial barrier function (gorman2023iggfcbp). LCN2 has been found to contribute to liver fibrosis and hypertension of the portal vein (J. Chen et al. (2020)).

MUC1 was also found in our PLS-DA analysis, here it was found to be an significant variable that allows for differentiating between biliary viability groups. MUC1, the transmembrane glycoprotein mucin 1, is a mucin family member that has different functions in normal and cancer cells. This is caused by its structure and biochemical properties. Mucin1 can act as a lubricant, moisturizer, and physical barrier in normal cells. In cancer cells, mucin 1 gets overexpressed and often faces aberrant glycosylation (W. Chen et al. (2021)). MUC1 is also expressed in the liver and is key for maintaining cellular functions, specifically for the epithelial surfaces (Kasprzak and Adamek (2019)). This gene could help in the regeneration of liver cells and thereby increase liver viability.

Other than 3 DEGs that were found by Thorne et al. (2023), 41 novel DEGs were found (Table 1), having a logFC range from 2.60 to -2.08 for the comparison of low biliary viability against high biliary viability. Literature research was performed for all DEG's with a logFC > 2 or < -2. MMP7 (logFC: 2.60) was found to contribute to the pathogenesis of bile duct epithelial injury (Allam et al. (2024)), which does not seem very logical as samples with high biliary viability should have a better state of the epithelial tissue than the low biliary viability samples, but the DEA points towards genes expressed in high samples that have been tied to tissue damage. PGC (logFC: 2.32) is a known activator of bile acid biosynthesis (shin2003cyp7a1) and adds to the profile of healthy bile ducts. ACE2 (logFC:2.23) is a receptor that is involved in allowing viral entry into host cells, including members of the SARS-CoV group (Pirola and Sookoian (2020)). ABCG5 (logFC: 2.08) is needed for transporting cholesterols into bile, Yu et al. (2002) found that expression of ABCG5 increased biliary cholesterol concentrations. COPZ1 (logFC: -2.08) encodes for a subunit of the cytoplasmic

coatamer protein complex, this complex is involved in protein transport and autophagy ("COPZ1 COPI Coat Complex Subunit Zeta 1 [Homo Sapiens (Human)]" (n.d.)).

In Figure 4, two low biliary viability livers seemed to be misclassified as high biliary viability; this possibly suggests that these livers could have unique characteristics. This group formation could also imply that proteins driving the separation are biologically relevant, and the analysis of these livers could provide valuable insight into the mechanisms differentiating between the high and low classifications. By identifying the important variables that cause the most separation, there is a deeper understanding to be gained of what molecular drivers distinguish these groups and allow for better assessment of biliary viability.

Our PLS-DA analysis identified COPZ1 as an important variable differentiating between the biliary viability groups. COPZ1 was also found to be downregulated in high biliary viability livers. COPZ1 encodes for a subunit that makes up the cytoplasmic coatomer protein, complex. The coatomer protein complex is involved in autophagy and intracellular protein trafficking ("COPZ1 COPI Coat Complex Subunit Zeta 1 [Homo Sapiens (Human)]" (n.d.)). The primary transport it facilitates is from the Golgi complex to the endoplasmic reticulum, also known as retrograde transport. This is important to maintain homeostasis by removing aging proteins and organelles for degradation and recycling of components. (Kucheryavskiy (2020)) This protein is a prognostic marker in liver hepatocellular carcinoma. Other info about this protein is the essential role it plays in maintaining the survival of some types of tumors (Hong et al. (2023)).

The matrix metalloproteinase MMP7 was highlighted in our PLS-DA. MMP7 is part of a family of enzymes responsible for the breakdown of extracellular matrix expressed, the blood flow through the liver could be messed up and the blood vessels could be worn down, causing a lower biliary viability (Lambert et al. (2005)). SLC6A19, encoding for the system B transmembrane protein, was also identified in our analysis. This protein actively transports neutral amino acids across the apical membrane of epithelial cells ("SLC6A19 Solute Carrier Family 6 Member 19 [Homo Sapiens (Human)]" (n.d.)). This gene and protein don't seem to have any relation to biliary viability and mainly focus on transporting amino acids in organs such as the kidneys and intestine (Bröer (2008)).

The DGE analysis on the transcriptomics data showed no significant differences between the high and low biliary viability. This could be due to issues with the used data. We were able to download the RNA data in 2 different file formats: paired FASTQ files and BAM files. It appeared, after running feature counts on the BAM files, that these were not aligned to a reference genome. This was very unusual, BAM-files are supposed to be mapped to a reference genome. So, the FASTQ files were used to extract the counts. A new issue showed up, the alignment rate of these FASTQ files was very low (30%). Checking the quality revealed that, at random, either the reverse or forward file for every sample contained reads with a length of 6 base pairs. These reads were not used by the aligner, which caused it to have a very low alignment rate. Seeking guidance from our teachers revealed that no one could figure out how this happened. These issues could explain why we didn't find any significant changes between high and low variability.

MinDet was chosen as the imputation method as it is a deterministic minimal value approach, meaning all missing values are replaced with the lowest observed value per sample. This method was chosen as it ensures reproducibility because it is not a stochastic method and it is also fitting for our data, which is left-censored with many missing not-at-random (MNAR) values. MNAR values are missing values that are caused by low peptide intensity signals, meaning the intensities are below the instrumental detection limit (Shi et al. (2025)). However, the data also, to a lesser extent, contains missing completely at random (MCAR) values, which occur independently of peptide abundance. MAR values can occur at any intensity, unlike MNAR values and thus imputing these values using the lowest found value is not optimal. However, as there were a lot more MNAR values, it was still decided to use MinDet. Still, it should be noted that this imputation method can skew results to the left, towards lower values, which can affect outcomes.

The analysis found several proteins with statistically significant categorization performance based on p-values less than 0.05, evaluated across several expression thresholds. Proteins such as SND1, SNRPD1, COPG1, and COPG2 consistently demonstrated strong classification performance, with perfect specificity and zero false positive rates. SND1 achieved the highest AUROC value, while SNRPD1 reached a close maximum AUROC value. These high AUROC values, combined with the respective high precision up to and almost maximal accuracy, indicate very high classification performance. These metrics are particularly important

in biomedical applications where high specificity and low false positive rates are critical to avoid incorrect classification because of the consequences of failure. In contrast, ACE2, MMP7, and PUF60 also met the statistical significance threshold but displayed lower AUROC values. Specifically, all below 0.7, or the line of adequacy. Despite this, they exhibited high precision and moderate accuracy. However, their specificity was low and the false positive rate high. This combination suggests a potential imbalance in classification outcomes. While high precision and accuracy might indicate some use in categorisation, the low AUROC implies that these proteins do not consistently perform well across the range of decision thresholds. AUROC remains a more comprehensive measure of model performance, capturing both sensitivity and specificity over all possible thresholds. Therefore, low AUROC values point to limited generalizability and shed doubt the reliability of these proteins as classifiers in larger datasets and or other, despite their statistical significance. The appearance of multiple significant thresholds for proteins like SND1 and SNRPD1 further reinforces their potential. Their ability to classify consistently across varying cut off values highlights a robustness that is desirable in practical applications, where expression levels can vary across individuals and experimental conditions.

These results emphasize the importance of evaluating multiple performance metrics when assessing biomarker candidates. While statistical significance is a necessary starting point, it is not sufficient on its own. Metrics such as AUROC, specificity, and false positive rate provide crucial context for determining the practical utility of a given protein in classification tasks. The combination of high AUROC, consistent performance across thresholds, and perfect specificity makes proteins like SND1 and SNRPD1 strong candidates for further investigation

Appendix

	logFC	Average Expression	t	p-value	adjusted p-value	B	up/down regulated
MMP7	-2.60	14.34	-4.61	3.39e-05	7.49e-03	2.25	DOWN
MUC1	-2.49	14.79	-4.50	4.85e-05	7.81e-03	1.92	DOWN
PGC	-2.32	13.38	-3.55	9.20e-04	3.35e-02	-0.76	DOWN
ACE2	-2.23	13.81	-4.58	3.72e-05	7.49e-03	2.17	DOWN
ABCG5	-2.08	13.00	-3.47	1.16e-03	3.53e-02	-0.97	DOWN
FCGBP	-1.96	14.47	-3.91	3.09e-04	2.49e-02	0.23	DOWN
MMP1	-1.91	13.56	-3.58	8.59e-04	3.34e-02	-0.69	DOWN
RAB8A	-1.89	13.06	-3.36	1.62e-03	4.28e-02	-1.27	DOWN
FLOT2	-1.79	14.22	-3.50	1.06e-03	3.45e-02	-0.89	DOWN
STOM	-1.64	16.43	-3.83	3.97e-04	2.58e-02	0.01	DOWN
LCN2	-1.44	16.58	-3.80	4.38e-04	2.58e-02	-0.08	DOWN
HEXA	-1.35	13.91	-3.41	1.40e-03	3.96e-02	-1.13	DOWN
CLU	-1.31	18.68	-3.99	2.44e-04	2.08e-02	0.45	DOWN
HNRNPA2B1	1.01	15.96	3.52	1.02e-03	3.45e-02	-0.85	UP
ALDH1L1	1.02	17.16	3.54	9.47e-04	3.35e-02	-0.78	UP
RGN	1.03	15.51	3.38	1.52e-03	4.17e-02	-1.21	UP
GPD1	1.04	14.60	4.15	1.47e-04	1.80e-02	0.91	UP
PAH	1.04	15.87	3.68	6.21e-04	3.19e-02	-0.40	UP
COPB1	1.07	13.34	3.50	1.07e-03	3.45e-02	-0.89	UP
APEX1	1.13	13.42	4.91	1.27e-05	6.81e-03	3.16	UP
NUDC	1.14	12.14	3.63	7.42e-04	3.33e-02	-0.56	UP
PYGL	1.15	17.30	4.68	2.68e-05	7.49e-03	2.47	UP
OPLAH	1.16	13.38	3.82	4.18e-04	2.58e-02	-0.04	UP
GALM	1.16	14.48	3.60	7.96e-04	3.34e-02	-0.62	UP
UROC1	1.25	15.37	4.51	4.77e-05	7.81e-03	1.94	UP
GLUL	1.28	12.87	3.56	9.05e-04	3.35e-02	-0.74	UP
RPL10A	1.31	15.54	4.69	2.63e-05	7.49e-03	2.48	UP
AGL	1.34	16.40	4.04	2.12e-04	2.08e-02	0.58	UP
ADSS1	1.34	11.39	3.45	1.24e-03	3.70e-02	-1.03	UP
KARS1	1.35	11.83	3.65	6.82e-04	3.33e-02	-0.48	UP
SRPRA	1.36	11.89	3.59	8.23e-04	3.34e-02	-0.65	UP
NACA	1.37	12.64	3.54	9.57e-04	3.35e-02	-0.79	UP
PCBD1	1.43	15.24	3.99	2.45e-04	2.08e-02	0.45	UP
KHK	1.44	15.78	5.32	3.30e-06	2.66e-03	4.39	UP
RPS12	1.47	13.92	4.60	3.51e-05	7.49e-03	2.22	UP
GYS2	1.50	11.27	3.57	8.71e-04	3.34e-02	-0.71	UP
UPB1	1.51	16.22	4.40	6.79e-05	9.11e-03	1.62	UP
ARCN1	1.51	13.86	3.78	4.63e-04	2.58e-02	-0.13	UP
METAP2	1.58	11.92	3.58	8.54e-04	3.34e-02	-0.69	UP
GBE1	1.65	15.44	5.55	1.50e-06	2.45e-03	5.11	UP
RPL12	1.65	15.84	3.90	3.25e-04	2.49e-02	0.19	UP
RBKS	1.79	12.20	3.82	4.16e-04	2.58e-02	-0.04	UP
SRSF2	1.82	12.64	3.78	4.65e-04	2.58e-02	-0.14	UP
COPZ1	2.08	11.94	3.43	1.31e-03	3.83e-02	-1.07	UP

Table 1. All found differentially expressed genes in the comparison between high- and low billiary viability livers after 150 minutes of NMP, with information on certain statistics such as p-values.

	protein	threshold	precision	accuracy	specificity	fpr	p_value	auroc	neg_expr_cor
1	COPG1	13.935	0.667	0.852	1.000	0.000	0.0002230	0.886	TRUE
2	COPG2	12.612	0.667	0.852	1.000	0.000	0.0002230	0.903	TRUE
3	COPG2	12.512	0.667	0.852	1.000	0.000	0.0002230	0.903	TRUE
4	HSD17B13	11.076	0.667	0.852	1.000	0.000	0.0002230	0.775	TRUE
5	SND1	13.894	0.667	0.852	1.000	0.000	0.0002230	0.950	TRUE
6	SND1	13.794	0.750	0.889	1.000	0.000	0.0000469	0.950	TRUE
7	SNRPD1	13.720	0.750	0.889	1.000	0.000	0.0000469	0.933	TRUE
8	SNRPD1	13.620	0.750	0.889	1.000	0.000	0.0000469	0.933	TRUE
9	SNRPD1	13.520	0.833	0.926	1.000	0.000	0.0000078	0.933	TRUE
1736	ACE2	12.743	0.933	0.750	0.444	0.556	0.0474300	0.663	FALSE
1737	ACE2	12.643	0.933	0.750	0.444	0.556	0.0474300	0.663	FALSE
63381	MMP7	14.343	0.933	0.750	0.444	0.556	0.0474300	0.663	FALSE
63382	MMP7	14.243	0.933	0.750	0.444	0.556	0.0474300	0.663	FALSE
82048	PUF60	9.088	0.933	0.750	0.444	0.556	0.0474300	0.648	FALSE

Table 2. Identified significant protein-threshold combinations in the 150-minute time point dataset.

References

Allam, A. A., M. A. Khedr, S. S. Elkholy, et al. 2024. "Bile Duct Matrix Metalloproteinase-7 Expression: A New Modality for Diagnosis of Biliary Atresia." *Egypt Liver Journal* 14: 17. https://doi.org/10.1186/s43066-024-00320-z.

Bröer, S. 2008. "Amino Acid Transport Across Mammalian Intestinal and Renal Epithelia." *Physiological Reviews* 88 (1): 249–86. https://doi.org/10.1152/physrev.00018.2006.

Chen, J., J. Argemi, G. Odena, M. Xu, Y. Cai, V. Massey, A. Parrish, et al. 2020. "Hepatic Lipocalin 2 Promotes Liver Fibrosis and Portal Hypertension." *Scientific Reports* 10 (1). https://doi.org/10.1038/s41598-020-72172-7.

Chen, W., Z. Zhang, S. Zhang, P. Zhu, J. K. Ko, and K. K. Yung. 2021. "MUC1: Structure, Function, and Clinic Application in Epithelial Cancers." *International Journal of Molecular Sciences* 22 (12): 6567. https://doi.org/10.3390/ijms22126567.

"COPZ1 COPI Coat Complex Subunit Zeta 1 [Homo Sapiens (Human)]." n.d. https://www.ncbi.nlm.nih.gov/gene/22818.

Dubois, E., A. N. Galindo, L. Dayon, and O. Cominetti. 2022. "Assessing Normalization Methods in Mass Spectrometry-Based Proteome Profiling of Clinical Samples." *Biosystems* 215–216: 104661. https://doi.org/10.1016/j.biosystems.2022.104661.

Ewels, P., M. Magnusson, S. Lundin, and M. Käller. 2016. "MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report." *Bioinformatics* 32 (19): 3047–48. https://doi.org/10.1093/bioinformatics/btw354.

"HISAT2." n.d. http://DaehwanKimLab.github.io/hisat2/.

"Homo Sapiens Genome Assembly GRCh38.p14." n.d. https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.40/.

Hong, Y., Z. Xia, Y. Sun, Y. Lan, T. Di, J. Yang, J. Sun, M. Qiu, Q. Luo, and D. Yang. 2023. "A Comprehensive Pan-Cancer Analysis of the Regulation and Prognostic Effect of Coat Complex Subunit Zeta 1." *Genes* 14 (4): 889. https://doi.org/10.3390/genes14040889.

Huber, W., A. von Heydebreck, H. Sueltmann, A. Poustka, and M. Vingron. 2002. "Variance Stabilization Applied to Microarray Data Calibration and to the Quantification of Differential Expression." *Bioinformatics* 18 Suppl. 1: S96–104.

Kasprzak, A., and A. Adamek. 2019. "Mucins: The Old, the New and the Promising Factors in Hepatobiliary Carcinogenesis." *International Journal of Molecular Sciences* 20 (6): 1288. https://doi.org/10.3390/ijms20061288.

Kim, D., J. M. Paggi, C. Park, C. Bennett, and S. L. Salzberg. 2019. "Graph-Based Genome Alignment and Genotyping with HISAT2 and HISAT-Genotype." *Nature Biotechnology* 37 (8): 907–15. https:

//doi.org/10.1038/s41587-019-0201-4.

Krueger, F., F. James, P. Ewels, E. Afyounian, M. Weinstein, B. Schuster-Boeckler, G. Hulselmans, and sclamons. 2023. "FelixKrueger/TrimGalore: V0.6.10 - Add Default Decompression Path." Zenodo. https://doi.org/10.5281/zenodo.7598955.

Kucheryavskiy, S. 2020. "Mdatools – r Package for Chemometrics." *Chemometrics and Intelligent Laboratory Systems* 198: 103937. https://doi.org/10.1016/j.chemolab.2020.103937.

Lambert, D. W., M. Yarski, F. J. Warner, P. Thornhill, E. T. Parkin, A. I. Smith, N. M. Hooper, and A. J. Turner. 2005. "Tumor Necrosis Factor-a Convertase (ADAM17) Mediates Regulated Ectodomain Shedding of the Severe-Acute Respiratory Syndrome-Coronavirus (SARS-CoV) Receptor, Angiotensin-Converting Enzyme-2 (ACE2)." *Journal of Biological Chemistry* 280 (34): 30113–19. https://doi.org/10.1074/jbc.m505111200.

Liao, Y., G. K. Smyth, and W. Shi. 2014. "featureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features." *Bioinformatics* 30 (7): 923–30. https://doi.org/10.1093/bioinformatics/btt656.

Love, M. I., W. Huber, and S. Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15 (12): 550. https://doi.org/10.1186/s13059-014-0550-8.

Pirola, C. J., and S. Sookoian. 2020. "COVID-19 and ACE2 in the Liver and Gastrointestinal Tract: Putative Biological Explanations of Sexual Dimorphism." *Gastroenterology* 159 (4): 1620–21. https://doi.org/10.1053/j.gastro.2020.04.050.

Ritchie, M. E., B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth. 2015. "Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies." *Nucleic Acids Research* 43 (7): e47. https://doi.org/10.1093/nar/gkv007.

Shen, C., H. Cheng, T. Zong, and H. Zhu. 2023. "The Role of Normothermic Machine Perfusion (NMP) in the Preservation of Ex-Vivo Liver Before Transplantation: A Review." *Frontiers in Bioengineering And Biotechnology* 11. https://doi.org/10.3389/fbioe.2023.1072937.

Shen, W., B. Sipos, and L. Zhao. 2024. "SeqKit2: A Swiss Army Knife for Sequence and Alignment Processing." *iMeta* 3 (3): e191. https://doi.org/10.1002/imt2.191.

Shi, Y., H. Zhong, J. C. Rogalski, and L. J. Foster. 2025. "Optimizing Imputation Strategies for Mass Spectrometry-Based Proteomics Considering Intensity and Missing Value Rates." *Computational and Structural Biotechnology Journal*. https://doi.org/10.1016/j.csbj.2025.04.041.

"SLC6A19 Solute Carrier Family 6 Member 19 [Homo Sapiens (Human)]." n.d. https://www.ncbi.nlm.nih.gov/gene/340024.

Thorne, A. M., J. C. Wolters, B. Lascaris, S. B. Bodewes, V. A. Lantinga, O. B. Van Leeuwen, I. E. M. De Jong, et al. 2023. "Bile Proteome Reveals Biliary Regeneration During Normothermic Preservation of Human Donor Livers." *Nature Communications* 14 (1). https://doi.org/10.1038/s41467-023-43368-y.

Yu, L., J. Li-Hawkins, R. E. Hammer, K. E. Berge, J. D. Horton, J. C. Cohen, and H. H. Hobbs. 2002. "Overexpression of ABCG5 and ABCG8 Promotes Biliary Cholesterol Secretion and Reduces Fractional Absorption of Dietary Cholesterol." *Journal of Clinical Investigation* 110 (5): 671–80. https://doi.org/10.1172/jci16001.

Zhang, X., A. Smits, G. van Tilburg, H. Ovaa, W. Huber, and M. Vermeulen. 2018. "Proteome-Wide Identification of Ubiquitin Interactions Using UbIA-MS." *Nature Protocols* 13: 530–50.