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]Bio-infromatics, Hanze Groningen

2025-06-09

Abstract

Liver transplantation is a critical treatment for liver failure, but donor shortages have led to the use of expanded criteria donor (ECD) livers, which carry higher risks of complications. Assessing biliary viability, with the purpose of assessing donor liver quality remains challenging. This article explores bile- and liver-based biomarkers associated with biliary viability using a multi-omics approach. Proteomic analysis of samples collected during normothermic machine perfusion (NMP) identified 44 differentially expressed proteins between high- and low-viability groups. Key proteins, including MMP7, MUC1, and COPZ1, showed significant abundance changes, suggesting roles in bile duct function and liver health. Machine learning analysis highlighted SND1 and SNRPD1 as strong classifiers (AUROC > 0.9), indicating their potential as biomarkers. Transcriptomic data did not reveal significant differences, possibly due to technical limitations. These findings contribute to a better understanding of biliary viability and propose candidate biomarkers for improving donor liver evaluation. Further validation could enhance transplant outcomes by refining organ selection.

Introduction

Liver transplantation is the first treatment option for patients with liver dysfunctionality (C. Shen et al. (2023)), e.g. in patients with liver cirrhosis or acute liver failure. There is a shortage of donor livers that meet required standards (Arulraj and Neuberger (2011)), so now expanded criteria donors (ECD)(Feng and Lai (2014)) 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 for liver transplantation can provide a solution for the shortage of liver donors, however, there are also risks associated with ECD livers, such as increased chances of graft failure and death especially in recipients < 60 years old and patients with diabetes (Van Ittersum et al. (2016)). ECD livers are also more susceptible to biliary complications and ischemia-reperfusion injury, which is damage that occurs when an organ lacking oxygen gets reperfused with oxygen again (C. Shen et al. (2023)).

Certain techniques, such as normothermic machine perfusion (NMP) (Hosgood and Nicholson (2011)) can reduce the risk of post-transplant complications. NMP attempts to restore and 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 (C. Shen et al. (2023)). NMP is considered a better alternative to static cold storage (SCS)(Lee and Mangino (2009)), as it reduces the risk of ischemia-reperfusion injury, early allograft dysfunction (EAD), and biliary complications, improving general graft viability (C. Shen et al. (2023)). NMP also allows evaluation of the donor liver pre-transplant, further reducing transplantation risks (Thorne et al. (2023)).

Thorne et al. (2023) describe how NMP is used for assessing organ viability before liver transplantation. Here, the general viability of the organ is divided into two distinct measurements. First, is hepatocellular viability (Groen et al. (2024)), 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 by analyzing the composition of the bile, including glucose levels and pH (Groen et al. (2024)), but as of today, there is no golden standard (Dingfelder et al. (2022)), 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 (Thorne et al. (2023)).

To gain a more comprehensive understanding of the molecular changes that contribute to biliary viability, a multi-omics approach (Hasin, Seldin, and Lusis (2017)) 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. Biomarkers include various biomolecules that can serve as identifiers for a condition, such as genes, transcripts and proteins (Rosati et al. (2024)). 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)(Peng et al. (2024)), and other relevant statistical and bioinformatics analyses to identify differentiating genes and proteins.

We are expecting to find our results comparable to Thorne et al. (2023), given the shared focus on biliary viability assesments. 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 differentially abundant proteins 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

Different proteomics and transcriptomic methods were used to analyse and possibly discover biomarkers. The used methods and analyses are presented in the following method section.

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. The mean percentage of missingness for the features is 42% for the 30 min samples and 56% for the 150 min samples. 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 at least one of the biliary viability groups were removed. The data is normalized across samples using median centering (Dubois et al. (2022)), this method is used to center the data over the median of the distribution of each sample. The features were then normalized by 'variance stabilizing transformation' using 'normalize_vsn' of the DEP R-package (v.1.26.0) (Zhang et al. (2018)), this function uses a robust variant of the maximum-likelihood estimator (Huber et al. (2002)). Missing values in the data were identified to mostly consist of 'missing not at random' values (MNAR)(Heymans and Twisk (2022)), making the missing data left-censored. Thus, imputation was performed using the 'MinDet' function of DEP

(v.1.26.0) (Zhang et al. (2018)), 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)(Jolliffe (2002)) 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 (Fawcett (2006)) (Bradley (1997)) from the resulting confusion matrix. Statistical significance was assessed using Fisher's exact test (Fisher (1932)) 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 proteins are down/upregulated, a pathway analysis was performed. The results were visualized using e.g. volcano plots and pathview. Further literature research was performed for all DEP's with a logFC > 2 or < -2.

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. Cross validation was used, which assessed the generalisation ability of the model and to prevent overfitting ("Cross-Validation" (2019)) , specifically cross validation with the leave-one-out method. This method used each sample once as a test set while the remaining samples made up the training dataset (developers (2024)). The model generated from this was analysed, and this was done using plots that showed what parameters were important for the model according to weights and variable importance predictions. The components that this model generated were also analysed and plotted with a scatter plot.

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)).

Statistical methods

Differential enrichment analysis (DEA) (Peng et al. (2024)) is a method for detecting differences in protein abundance across groups, where the data gets grouped based on certain phenotypes (e.g. disease vs healthy). Resulting from the DEA are differentially expressed proteins (DEPs), which are biomarkers and can be used for constructing phenotype-specific abundance profiles. DEA in proteomics is performed on quantified, normalized, and imputed data, using one of the available tools. The selected tool, limma (Ritchie et al. (2015)) is not designed specifically for proteomics but has proven to perform well in any setting with quantified data and is a popular option for proteomics (Peng et al. (2024)). limma uses linear fitting onto each of the features in order to find differences/correlations between conditions. To improve estimates further statistical methods can be applied to the results from limma, in this article, the Empirical Bayes method (Smyth (2004)) is applied for smoothing estimates of discrete probabilities, and lastly, a false discovery rate (FDR) correction (Aggarwal and Yadav (2016)) is applied. The FDR correction is based on Benjamini and Hochberg (1995) and corrects p-values for the expected proportion of the features that are falsely found to be statistically significant, as a result of multiple testing.

PCA is used on high dimensional data. It is a method of reducing the dimensions of data. This is useful for our data since we are looking at the expression of > 2000 different genes. A PCA works by calculating a covariance matrix, linear correlations, between all of the different variables. One value in this matrix is the linear correlation between 2 variables, or genes in our case. Eigenvectors will be calculated from the variance matrix, these are the directions where the variance of the data is maximized. These values can be used to calculate the loadings, or 'strength', of the variables. These loadings, every component has loadings for all the variables, can then be projected onto the data. (Jolliffe (2002))

Partial Least Squares Discriminant Analysis (PLS-DA) is a supervised multivariate dimensionality-reduction tool employed to identify complex relationships within high-dimensional datasets. As a variant of Partial Least Squares (PLS) regression, PLS-DA directly utilizes class information to derive components that optimally discriminate between predefined groups (Kucheryavskiy (2020)) . This method was used to identify biomarkers in the present study. For model validation, cross-validation is crucial to assess the generalization ability of the model and to prevent overfitting (developers (2024)) . Model quality is quantified by metrics such as R2Y (explained variance in Y) and Q2 (predictive variance). A permutation test is performed to confirm statistical significance and check for overfitting. Variable Importance in Projection (VIP) scores identify the X-variables most influential in group discrimination, typically with a threshold of VIP > 1.0. ("Cross-Validation" (2019))

Results

The results of the analyses and methods (mentioned in Materials and Methods section) to discover biomarkers will be presented in the following sections.

DEA

The DEA on the proteomics samples in total evaluated 2255 features for the 30-min samples and 1610 features for the 150-min samples. For the 30-min samples significantly enriched expressed proteins were found. The 150-min samples contained a total of 44 differentially abundant proteins (Fig.1), details on e.g. p-values and logFC per differentially abundant protein can be found in Table 1. The DEA on the transcriptomics data data revealed no significant changes in the abundance of proteins between a high and low biliary viability
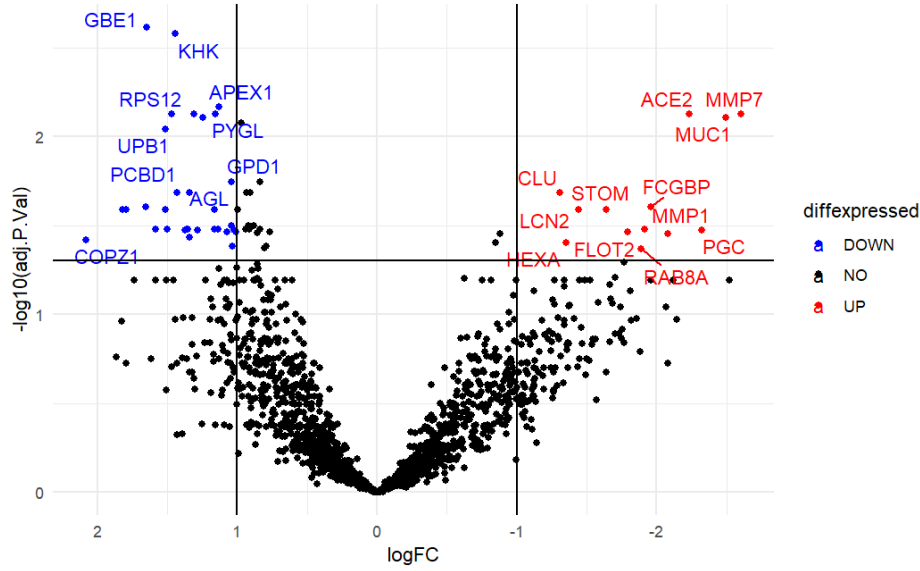
Figure 1: Volcano plot showing all significantly upregulated- and downregulated proteins in the comparison of high biliary viability against low biliary viability in the samples taken during NMP after 150 minutes. Proteins were deemed as significantly upregulated if the adjusted p-value was ≤ 0.05 and the $\log_2$ fold change was ≥ 1.0. Proteins 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 less spread throughout the second component. Samples with a high biliary viability show to have a 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. The first and second component explain 16.4% and 10.5% of the total variance, respectively.

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. The first and second component explain 20.8% and 11.6% of the total variance, respectively.
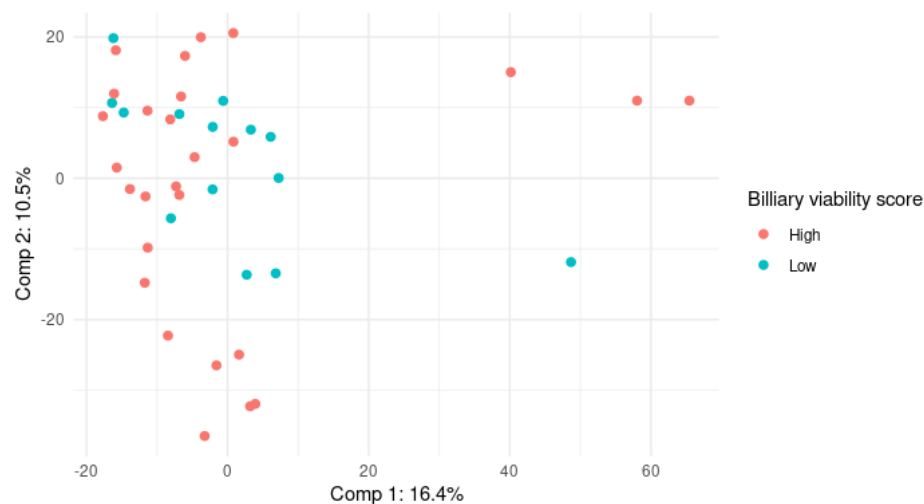
Figure 2: PCA performed on proteomic samples 30 minutes after starting NMP. Shows how the samples are spread over component 1 and component 2. Biliary viability is displayed with color.
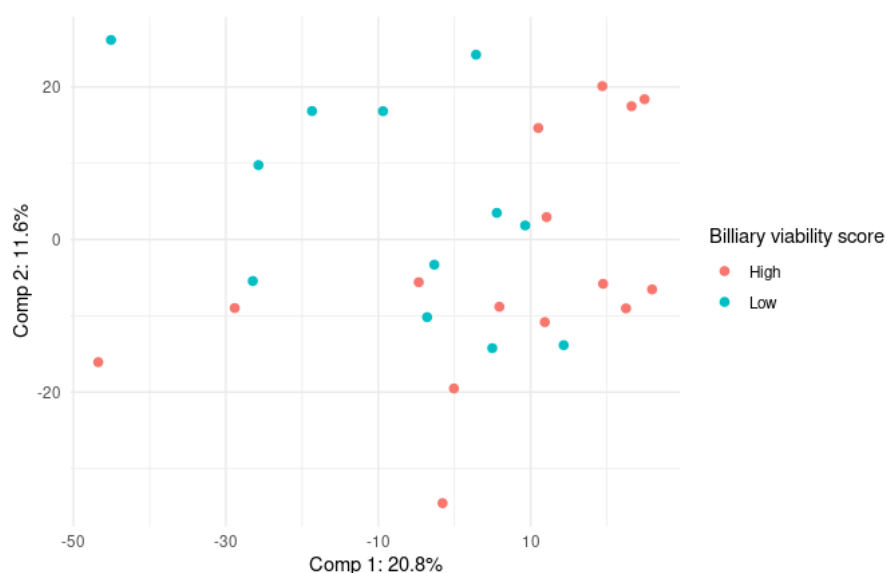

Figure 3: PCA performed on proteomic samples 150 minutes after starting NMP. Shows how the samples are spread over component 1 and component 2. Biliary viability is displayed with color.

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 the loadings of the most important component. This displays all proteins that have a contribution above 8. Figure 5 shows the comparison between the 1st and 3rd component. The 1st component contains 82.43% of the information of the orignial data. Component 3 contains 2.06% of the information of the original data. This comparison was the only one that showed 2 groups, this divide is shown by the dotted line.
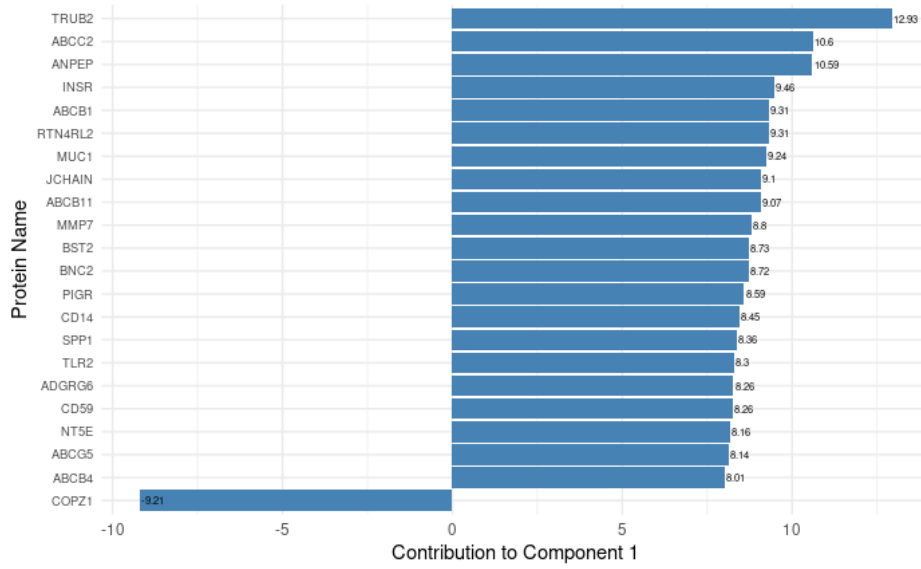
6

Figure 4: Contribution of proteins to Component 1 in the PLS-DA model. The bar plot displays the loadings of the individual proteins on the first compnent. The proteins were filtered with a minimum of contribution above 8. This was done to show a clear overview of the most influential proteins. Proteins with larger absolute loadings are more influential in defining the variation captured by Component 1 and thus contribute significantly to the group separation observed in the PLS-DA score plot.
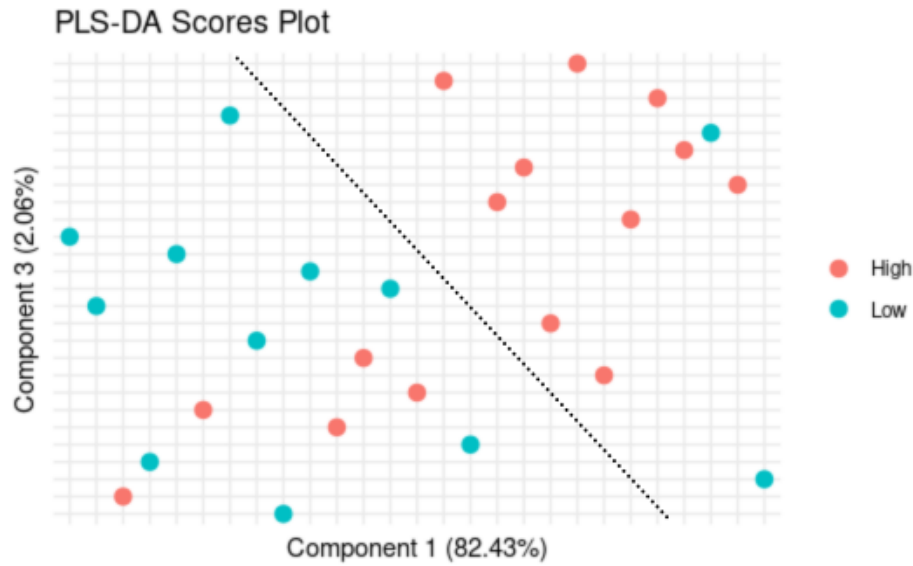

Figure 5: PLS-DA Score plot showing component 1 – component 3, performed on the proteomic samples at 150 min after starting NMP. Showing how the samples are laid out over components 1 and 3. Viability is shown with colour. A Dotted line has been drawn to show the 2 groups that were formed.

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/Conclusion

Protein abundances were compared in a DEA between two groups and two timepoints, with low against high biliary viability at timepoint 30 minutes and at timepoint 150 minutes. Resulting from this was the identification of 44 statistically significant and differentially enriched potential biomarker proteins ($p-adjust \leq 0.05$ & $log2FC \geq 1.0 \vee log2FC \leq -1.0$) (Table 1). These biomarkers have a logFC range of -2.60 to 2.08 for the comparison of high biliary viability against a reference of low biliary viability. This means any protein with a positive logFC is more abundant in high biliary viability livers, and a negative logFC means the protein is more abundant in low biliary viability livers. A further literature search was performed for all found proteins with a $logFC \geq 2.0 \vee logFC \leq -2.0$, corresponding to a fold change of 4. This results in 5 proteins that are more abundant in low biliary viability livers, including: MMP7 (logFC: -2.60, p-adj: 7.49e-03), MUC1 (logFC: -2.49, p-adj: 7.81e-03), PGC (logFC: -2.32, p-adj: 3.35e-02), ACE2 (logFC: -2.23, p-adj: 7.49e-03) and ABCG5 (logFC: -2.08, p-adj: 3.53e-02), and 1 protein that is more abundant in high biliary viability livers, which is COPZ1 (logFC: 2.08, p-adj: 3.83e-02).

MMP7 (matrix metallopeptidase 7) is part of a family of enzymes responsible for the breakdown of extracellular matrix proteins (Shantha Kumara et al. (2021)). MMP7 has been found to contribute to the pathogenesis of bile duct epithelial injury (Allam et al. (2024)) and was also highlighted in our PLS-DA. MUC1 (Mucin 1) is a mucin family member that is associated with cell signaling and proliferation pathways and expression of MUC1 is predominantly found in cholangiocytes (Thorne et al. (2023)). MUC1 function differs in normal cells as opposed to cancer cells, in normal cells, it acts as a lubricant but in cancer cells, overexpression of MUC1 can occur, leading to aberrant glycosylation (W. Chen et al. (2021)). MUC1 is also expressed in the liver and it has been found to be overexpressed during the carcinogenesis of hepatobiliary cancers (Kasprzak and Adamek (2019)). MUC1 was found in our PLS-DA analysis, here it was found to be a significant variable that allows for differentiating between biliary viability groups. PGC (Progastricsin) in enriched abundance plays a role in tumorigenesis of hepatocellular carcinoma (H. Chen et al. (2019)). ACE2 (Angiotensin-converting enzyme 2) 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 (ATP-binding cassette sub-family G member 5) is needed for transporting cholesterols into bile, Yu et al. (2002) found that expression of ABCG5 increased biliary cholesterol concentrations. Increased cholesterol concentrations can lead to cholesterol supersaturation, which can cause the formation of gallstones (Dikkers and Tietge (2010)). Noticeably all of the 5 proteins with increased abundance in low liver viability, can be linked to (liver) disease,

this signifies that they are interesting candidates for further investigation as potential biomarkers.

COPZ1 (COPI Coat Complex Subunit Zeta 1) 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.)). COPZ1 was found to be overexpressed in high biliary viability livers, but interestingly this COPZ1 was also found overexpressed in multiple types of cancer, including cholangiocarcinoma and liver hepatocellular cancer (Hong et al. (2023a)). This protein is a prognostic marker in liver hepatocellular carcinoma and plays an essential role in maintaining the survival of a variety of tumor types (Hong et al. (2023b)). COPZ1 was also identified as an important variable for differentiating between the biliary viability groups in our PLS-DA analysis.

Analysing the loadings of the PLS-DA (Figure 4) led to the identification of two other significant proteins with liver functions that were not found in the DEA, which include ABCC2 and ABCB11. ABCC2, also known as the Multidrug Resistance Protein 2, is involved in the transport of substances out of cells. It plays a crucial role in drug disposition, by mediating ATP-dependent efflux of a wide range of compounds, including drug metabolites, from cells into bile or urine. (MedlinePlus Genetics (2018)). ABCC2's transport of bilirubin out of the liver cells and into the bile is particularly relevant. Bilirubin is a substance that is produced during the breakdown of old red blood cells and is frequently used to check the health of a patient's liver; therefore, the identification of ABCC2 as an important potential biomarker for biliary viability. (MedlinePlus (2025)). ABCB11, also known as the Bile Salt Export Pump (BSEP). BSEP is responsible for the ATP-dependent transport of primary hydrophobic bile salts, such as taurine and glycine-conjugated cholic acid, from hepatocytes into the bile (UniProt Consortium (2025)). This process is fundamental for hepatic bile acid homeostasis and, consequently, for lipid homeostasis through the regulation of biliary lipid secretion. Dysregulation or dysfunction of BSEP can lead to severe cholestatic liver injury (Byrne et al. (2002)), as observed in conditions like Progressive Familial Intrahepatic Cholestasis Type 2 (PFIC2), where biallelic mutations in ABCB11 cause an accumulation of hepatic bile acids and subsequent hepatotoxicity, ultimately leading to liver failure (Gruget, Reddy, and Moore (2025)). Given BSEP's critical role in bile acid transport, it's an important protein for evaluating biliary viability and, therefore a significant biomarker.

Some of our findings of our DEA and PLS-DA overlap with what Thorne et al. (2023) have found. Thorne et al. (2023) found three to show significant differences between biliary viability groups; MUC5B, MUC1, and LCN2. We have been able to reinforce two of these findings, MUC1 (logFC: -2.49, p-adj: 7.81e-03) and LCN2 (p-adjust: 4.38x10-4, logFC: 1.44). LCN2 (Lipocalin-2) has been found to contribute to liver fibrosis and hypertension of the portal vein (J. Chen et al. (2020)). MUC5B was not found to show significant differences between biliary viability groups in our research.

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. These issues could explain why we didn't find any significant changes between high and low variability.

In Figure 5, two groups are shown divided by the dotted line. However, two low-biliary-viability livers appeared to be misclassified as having 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 analysing these livers could provide valuable insights into the mechanisms that differentiate between the high and low classifications. By identifying the key variables that cause the most separation, a deeper understanding can be gained of what molecular drivers distinguish these groups and facilitate a more accurate assessment of biliary viability.

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 principal component analyses revealed no clear seperation between low and high biliary viability. This could be a result of the chosen imputation method, which could have caused a lower variance between the high and low biliary viability samples.

The analysis found several proteins with statistically significant categorization performance based on p-values ≤ 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 categorization, 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 on 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 highlights 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

In this article we have presented novel proteins that could be used as biomarkers in the differentiation between ECD-donor livers with high- and low biliary viability, which brings us closer to being able to develop biomarker profiles for biliary viability levels allowing for more structural analysis of biliary viability in ECD-donor livers, potentially enlarging the donor pool. Further research is needed to validate the relationships between these proteins of interest and biliary viability.

Appendix

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

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

References

Aggarwal, Sandeep, and Anil K. Yadav. 2016. "False Discovery Rate Estimation in Proteomics." In *Statistical Analysis in Proteomics*, edited by Klaus Jung, 1362:95–108. Methods in Molecular Biology. New York, NY: Humana Press. https://doi.org/10.1007/978-1-4939-3106-4_7.

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.

Arulraj, Rajesh, and James Neuberger. 2011. "Liver Transplantation: Filling the Gap Between Supply and Demand." *Clinical Medicine* 11 (2): 194–98. https://doi.org/10.7861/clinmedicine.11-2-194.

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Methodological)* 57 (1): 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Bradley, Andrew P. 1997. "The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms." *Pattern Recognition* 30 (7): 1145–59.

Byrne, Jane A., Sandra S. Strautnieks, Giorgina Mieli-Vergani, Christopher F. Higgins, Kenneth J. Linton, and Richard J. Thompson. 2002. "The Human Bile Salt Export Pump: Characterization of Substrate Specificity and Identification of Inhibitors." *Gastroenterology* 123 (5): 1649–58. https://doi.org/10.1053/gast.2002.36591.

Chen, H., H. Zhu, X. Yu, X. Shi, E. Bilegsaikhan, H. Guo, R. Huang, T. Liu, X. Shen, and J. Zhu. 2019. "Overexpressed Pepsinogen c Is Associated with Poor Prognosis in Human Hepatocellular Carcinoma: A Tissue Microarray Study." *Cancer Management and Research* 11: 2927–34. https://doi.org/10.2147/cmar.s192241.

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.

"Cross-Validation." 2019. In *Encyclopedia of Bioinformatics and Computational Biology*. Elsevier.

developers, Scikit-learn. 2024. "Sklearn.model_selection.LeaveOneOut." https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.LeaveOneOut.html.

Dikkers, Arne, and Uwe J. F. Tietge. 2010. "Biliary Cholesterol Secretion: More Than a Simple ABC." *World Journal of Gastroenterology* 16 (47): 5936–45. https://doi.org/10.3748/wjg.v16.i47.5936.

Dingfelder, Julia, Lukas Rauter, Georg A. Berlakovich, and Daniel Kollmann. 2022. "Biliary Viability Assessment and Treatment Options of Biliary Injury During Normothermic Liver Perfusion—a Systematic Review." *Transplant International* 35: 10398. https://doi.org/10.3389/ti.2022.10398.

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.

Fawcett, Tom. 2006. "An Introduction to ROC Analysis." *Pattern Recognition Letters* 27 (8): 861–74.

Feng, Shan, and Jennifer C. Lai. 2014. "Expanded Criteria Donors." *Clinics in Liver Disease* 18 (3): 633–49. https://doi.org/10.1016/j.cld.2014.05.005.

Fisher, Ronald A. 1932. *Statistical Methods for Research Workers*. 4th ed. Oliver; Boyd.

Groen, Pieter C., O. Bote Van Leeuwen, Jeroen De Jonge, and Robert J. Porte. 2024. "Viability Assessment of the Liver During Ex-Situ Machine Perfusion Prior to Transplantation." *Current Opinion in Organ Transplantation* 29 (4): 239–47. https://doi.org/10.1097/mot.0000000000001152.

Gruget, Camille, Bhargav G. Reddy, and James M. Moore. 2025. "A Structural and Mechanistic Model for BSEP Dysfunction in PFIC2 Cholestatic Disease." *Communications Biology* 8: 531. https://doi.org/10.

1038/s42003-025-07908-0.

Hasin, Yehonatan, Marcus Seldin, and Aldons Lusis. 2017. "Multi-Omics Approaches to Disease." *Genome Biology* 18: 83. https://doi.org/10.1186/s13059-017-1215-1.

Heymans, M. W., and J. W. Twisk. 2022. "Handling Missing Data in Clinical Research." *Journal of Clinical Epidemiology* 151: 185–88. https://doi.org/10.1016/j.jclinepi.2022.08.016.

"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. 2023b. "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.

———. 2023a. "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.

Hosgood, Sarah A., and Michael L. Nicholson. 2011. "First in Man Renal Transplantation After Ex Vivo Normothermic Perfusion." *Transplantation* 92 (7): 735–38. https://doi.org/10.1097/tp.0b013e31822d4e04.

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.

Jolliffe, I. T. 2002. *Principal Component Analysis.* 2nd ed. Springer Series in Statistics. New York: Springer.

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, Sergey V. 2020. *Mdatools: R Package for Multivariate Data Analysis.* https://cran.r-project.org/web/packages/mdatools/index.html.

Lee, Chien Y., and Martin J. Mangino. 2009. "Preservation Methods for Kidney and Liver." *Organogenesis* 5 (3): 105–12. https://doi.org/10.4161/org.5.3.9582.

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.

MedlinePlus. 2025. "Bilirubin Blood Test." U.S. National Library of Medicine; https://medlineplus.gov/lab-tests/bilirubin-blood-test/.

MedlinePlus Genetics. 2018. "ABCC2 Gene — ATP-Binding Cassette Subfamily c Member 2." U.S. National Library of Medicine; https://medlineplus.gov/genetics/gene/abcc2/#conditions.

Peng, Hao, Haoyang Wang, Wei Kong, Jian Li, and Wilson Wen Bin Goh. 2024. "Optimizing Differential Expression Analysis for Proteomics Data via High-Performing Rules and Ensemble Inference." *Nature Communications* 15 (1). https://doi.org/10.1038/s41467-024-47899-w.

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.

Rosati, Daniela, Marco Palmieri, Gianni Brunelli, Andrea Morrione, Francesco Iannelli, Elisa Frullanti, and Anna Giordano. 2024. "Differential Gene Expression Analysis Pipelines and Bioinformatic Tools for the Identification of Specific Biomarkers: A Review." *Computational and Structural Biotechnology Journal*

23: 1154–68. https://doi.org/10.1016/j.csbj.2024.02.018.

Shantha Kumara, H., H. Miyagaki, S. A. Herath, E. Pettke, X. Yan, V. Cekic, and R. L. Whelan. 2021. "Plasma MMP-2 and MMP-7 Levels Are Elevated First Month After Surgery and May Promote Growth of Residual Metastases." *World Journal of Gastrointestinal Oncology* 13 (8): 879–92. https://doi.org/10.4251/wjgo.v13.i8.879.

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.

Smyth, Gordon K. 2004. "Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments." *Statistical Applications in Genetics and Molecular Biology* 3 (1): 1–25. https://doi.org/10.2202/1544-6115.1027.

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.

UniProt Consortium. 2025. "UniProtKB Entry for O95342 (ABCB11, Bile Salt Export Pump)." UniProt; https://www.uniprot.org/uniprotkb/O95342/entry.

Van Ittersum, Frank J., A. C. Hemke, F. W. Dekker, L. B. Hilbrands, M. H. L. Christiaans, J. I. Roodnat, A. J. Hoitsma, and M. Van Diepen. 2016. "Increased Risk of Graft Failure and Mortality in Dutch Recipients Receiving an Expanded Criteria Donor Kidney Transplant." *Transplant International* 30 (1): 14–28. https://doi.org/10.1111/tri.12863.

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.