From exons to export: How alternative splicing modulates the secretion specificity for large proteins

Inaugural-Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

Submitted to the Department of Biology, Chemistry, Pharmacy

of Freie Universität Berlin

by

Felix Ostwaldt

Berlin 2024

This work was carried out between March 2020 and February 2024 under the supervision of Prof. Dr. Florian Heyd at the Institute of Chemistry and Biochemistry, Freie Universität Berlin, Germany.

First reviewer:

Prof. Dr. Florian Heyd
RNA Biochemistry
Institute for Chemistry and Biochemistry
Freie Universität Berlin
Takustraße 6
14195 Berlin, Germany

Second reviewer:

Junior Prof. Dr. Francesca Bottanelli
Membrane Trafficking
Institute for Chemistry and Biochemistry
Freie Universität Berlin
Thielallee 63
14195 Berlin

Date of defense: 13.05.2024

Acknowledgements

Firstly, I want to thank my supervisor, Florian Heyd, for the opportunity to work on this fascinating project. You gave me the support to learn and room to grow as a scientist.

Secondly, I want to thank Francesca Bottanelli for acting as my second reviewer and for all the material and scientific support you gave me.

In the following, I want to thank further:

Alexander Stockhammer and Luis Wong, your know-how and cell lines are crucial for my work.

Marco Preussner for guiding my bioinformatical progress and being there for scientific support.

Bruna Los for your friendship. I am very grateful for how much you helped me in my beginning and throughout my work. I deeply miss my daily opportunities to annoy you.

All current and former members of the AG Heyd. You were the best emotional support I could hope for, this work would not be possible without you. Special thanks goes to: Benjamin Dimos-Röhl for successful scientific work together, Ann-Kathrin Emmerichs for always having an open ear for my sorrows, Stefan Meinke for you 2 µL pipette, Ioana Weber for incredible good scientific feedback and absurd internet references, Mateusz Drozdz for being the best roommate, Silvia Scalzitti for lighting up the mood in the big office and especially Elena Becker for being there for me in the time of writing and acting as an very helpful editor.

All my friends in Berlin, distributed in Germany and all over the world.

My family, Janne, Tin, Meike and Ulrike for always being there for me, helping me with everything and being a point of rest in turbulent times.

“Follow your passion, and life will reward you.”

-Uncle Iroh

Declaration of Independence

Herewith I certify that I have prepared and written my thesis independently and that I have not used any sources and aids other than those indicated by me.

I also declare that I have not submitted the dissertation in this or any other form to any other institution as a dissertation.

Table of Contents

I. Abstract

Compartmentalization is one of the decisive characteristics of cellular life, enabling the cell to build a complex network of enzymatic reactions and metabolic pathways. Drivers of this cellular management are organelles and the vesicles and tubular structures that connect them, all of which are separated from the cytoplasm by a lipid membrane. Intracellular transport, especially secretion to the extracellular space or to the cell membrane, is a highly regulated and organized process. In this work, we focused on the secretion of big protein cargo like chylomicrons and collagens. We aimed to analyze the role of alternative splicing as a regulation mechanism for the secretion of large cargo. Alternative splicing is a mechanism in gene expression where different combinations of exons are included or excluded from mRNA transcripts, leading to the production of multiple protein isoforms from a single gene.

We comprehensively analyzed RNA sequencing data from human tissues as well as B-lymphocytes. We found a previously uncharacterized exon in the gene SEC31A that codes for the outer layer of coat protein complex II (COPII), a complex responsible for the transport from the endoplasmic reticulum (ER) to the Golgi-apparatus. Using a correlation analysis with the inclusion levels of the exon and gene expression throughout the human tissue data, we found leads to the functionality and regulation of this alternatively spliced exon. We could show that the inclusion of the exon enhances the transport of lipids in polar differentiated Caco-2 cells. Mini gene experiments demonstrated that the inclusion is regulated by the splicing factor RBM47. Finally, AlphaFold structure prediction with the resulting alternative protein isoforms revealed a change in interaction with COPII partner SEC23A.

We also investigated a previously characterized alternative splice site in SEC31A, which leads to a shortened exon. This shortened exon decreased the secretion efficiency of glycosylphosphatidylinositol (GPI)-anchored and e-cadherin cargo in the RUSH assay but enhanced collagen secretion. In a fluorescent recovery after photobleaching (FRAP) experiment, we observed a reduced mobile fraction of COPII puncta with the shortened exon. Lastly, we analyzed RNA sequencing data of differentiating B-lymphocytes as a model for the regulatory effect of the transcriptome on intracellular trafficking and secretion. B-lymphocytes show rather extensive changes in gene expression during their change from memory cells, to plasmablasts, and plasma cells, but alternative splicing changes are most abundant in memory cells. We found multiple splice events, which are unique to memory cells, particularly in PICALM and KLHL12, which could potentially have a significant impact

on the function of the proteins.

In summary, we were able to showcase the mechanism by which the structure and function of COPII is altered via alternative splicing of SEC31A, and we demonstrated similar mechanisms in B-lymhocytes. With these results, we depended our knowledge of the role that alternative splicing has on secretory specificity and regulation of large cargo secretion.

II. Zusammenfassung

Die Kompartimentierung ist eines der entscheidenden Merkmale des zellulären Lebens und ermöglicht es der Zelle, ein komplexes Netzwerk enzymatischer Reaktionen und Stoffwechselwege aufzubauen. Treiber dieses zellulären Managements sind Organellen, Vesikel und die sie verbindenden Strukturen. Der intrazelluläre Transport, insbesondere die Sekretion in den extrazellulären Raum oder an die Zellmembran, ist ein stark regulierter und organisierter Prozess. In dieser Dissertation setzen wir uns mit der Sekretion großer Proteine wie Chylomikronen und Kollagene auseinander. Unser Ziel war es, die Rolle des alternativen Spleißens als Regulierungsmechanismus für die Sekretion großer Proteine zu analysieren. Alternatives Spleißen ist ein Mechanismus in der Genexpression, bei dem verschiedene Kombinationen von Exons in mRNA-Transkripte eingeschlossen oder ausgeschlossen werden, was zur Produktion mehrerer Proteinisoformen aus einem einzelnen Gen führt.

Wir haben RNA-Sequenzierungsdaten aus menschlichem Gewebe sowie B-Lymphozyten umfassend analysiert. Dabei wurde ein bisher nicht charakterisiertes Exon im Gen SEC31A gefunden, das für die äußere Schicht des *coat protein complex II* (COPII) kodiert, einem Komplex, der für den Transport vom endoplasmatischen Retikulum (ER) zum Golgi-Apparat verantwortlich ist. Mithilfe einer Korrelationsanalyse zwischen dem Exon und der Genexpression in den menschlichen Gewebedaten fanden wir Hinweise auf die Funktionalität und Regulierung dieses alternativ gespleißten Exons. Wir konnten zeigen, dass der Einschluss des Exons den Transport von Lipiden in polar differenzierten Caco-2-Zellen steigert. Mini-Gen-Experimente zeigten, dass der Einschluss durch den Spleißfaktor RBM47 reguliert wird. Schließlich zeigte die AlphaFold-Strukturvorhersage mit den resultierenden alternativen Proteinisoformen eine Veränderung in der Interaktion mit dem COPII-Partner SEC23A.

Wir haben eine zuvor charakterisierte alternative Spleißstelle in SEC31A untersucht, die zu einem verkürzten Exon führt. Dieses verkürzte Exon verringerte die Sekretionseffizienz von Glycosylphosphatidylinositol (GPI)-verankerter und E-Cadherin-Reporter im RUSH-Assay.

Die Effizienz der Kollagensekretion wurde jedoch erhöht. In einem *fluorescent recovery after photobleaching* (FRAP) Experiment beobachteten wir eine Verringerung der mobilen Fraktion von COPII puncta mit dem verkürzten Exon.

Schließlich analysierten wir RNA-Sequenzierungsdaten differenzierender B-Lymphozyten als Modell für die regulatorische Wirkung des Transkriptoms auf den intrazellulären Transport und die Sekretion. B-Lymphozyten zeigen während ihres Übergangs von Gedächtniszellen zu Plasmablasten und Plasmazellen ziemlich umfangreiche Veränderungen in der Genexpression Änderungen im alternativen Spleißen kommen jedoch am häufigsten in Gedächtniszellen vor. Wir fanden mehrere alternative Exons, die nur in Gedächtniszellen vorkommen, insbesondere in PICALM und KLHL12, und möglicherweise einen erheblichen Einfluss auf die Funktion der Proteine haben könnten. Zusammenfassend konnten wir einen Mechanismus aufzeigen, durch den die Struktur und Funktion von COPII durch alternatives Spleißen von SEC31A verändert wird, und wir konnten ähnliche Mechanismen in B-Lymphozyten nachweisen. Mit diesen Ergebnissen stützten wir unser Wissen über die Rolle, die alternatives Spleißen auf die Sekretionsspezifität und die Regulierung der Sekretion großer Proteine spielt.

1 Introduction

1.1 Precursor mRNA

The central dogma of molecular biology states, that the information for protein synthesis is stored in deoxyribonucleic acid (DNA) and mediated by ribonucleic acid (RNA) (Crick 1970). This simple and linear approach to the building blocks of life is challenged frequently, and an ever more complex picture of cellular life has been postulated by many(Song 2021; Chang and Qi 2023). The copy of the DNA is called a precursor-RNA, for protein coding genes precursor-messenger RNA (pre-mRNA). The nucleobase thymine, which is part of the DNA is replaced by uracil in the RNA, but otherwise pre-mRNA is an exact copy of the DNA. The pre-mRNA has to be further modified to fulfill its tasks. It is capped with a 7-methylguanosine at the 5' end, polyadenylation is added at the 3' end, and the introns are removed, a process which is called splicing. Only after these post-transcriptional modifications is the mature messenger RNA (mRNA) ready to be translated into proteins.

1.1.1 Splicing

Each human gene is composed of, on average, ~145 nucleotides long fragments called exons, which contain the nucleotides of the mature mRNA. Between these exons are located the on average 3400 nt long non-coding regions, which are called introns (Hnilicová and Staněk 2011). The human genome, thus, is made up in large parts of introns. Even more of

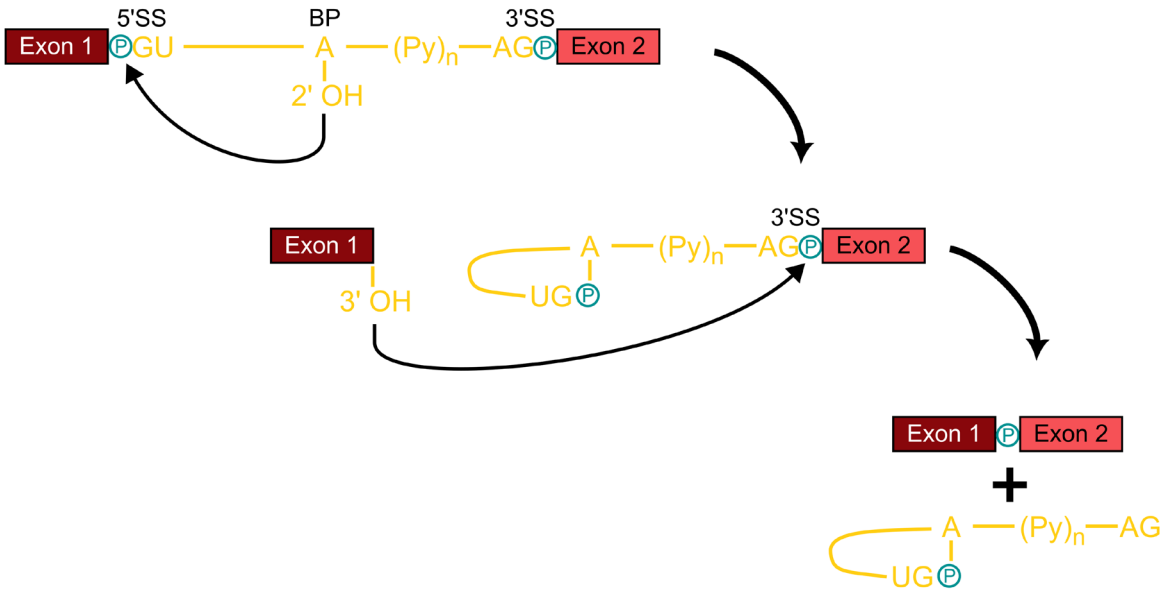

Figure 1-1: Simple scheme of the splicing reaction. The intron is removed by a sequence of two transesterifications, where at the first the 2'-hydroxy group of the branch point (BP) nucleophilic attack the phosphate of the 5' splice site (SS). The second reaction happens between the now freed 3'-hydroxy group of the free exon attacking the phosphate of the 3' splice site. Exons are shown as rectangles, introns and their compartments are drawn yellow.

the DNA is made up of non-transcribed areas. What previously was named junk DNA (Ohno and Yomo 1991) is in modern science debate for complex regulatory mechanisms and evolution (Ariel and Manavella 2021). The process of removing introns and joining the exons to together during transcription is called splicing.

Splicing is comprised of two transesterification reactions. They are mediated by the spliceosome, an enormous macromolecular machine which identifies the site of bond cleavage, forms substrate RNA, activates nucleophiles, and catalyzes the transesterification reactions. The first reaction occurs at the 5' splice site of the pre-mRNA, between the 5' exon and the intron. This site almost always contains the consensus sequence guanine-uracil (GU). The spliceosome activates an intronic adenosine so that the 2'-hydroxy group can act as a nucleophile for attacking the phosphor atom in the phosphodiester group at the intronic 5' splice site. This reaction is a bimolecular nucleophilic substitution (Moore and Sharp 1993)(Figure 1-1). The 5' exon is subsequently detached from the intron. The active adenosine is bound to a phosphodiester at the 2' and 5' sites, creating a lariat structure. This creates a point where the nucleotide chain is branched, hence the name branch point (BP) of the adenosine. The BP is usually followed by a poly-pyrimidine track. The free 3'-hydroxy group of the 5' exon carries out the second transesterification reaction. This hydroxy group nucleophilic attacks the phosphor atom of the phosphodiester bond of the 3' splice site (Figure 1-1). This 3' splice site usually contains an adenosine-guanine (AG) pair. The result is a new phosphodiester group between the 5' and 3' exon.

The spliceosome comprises several small nuclear ribonucleoproteins (snRNPs), non-snRNP proteins, and splicing factors. Two types of spliceosomes, the major and the minor, differ by their composition and target introns. The major spliceosome, or U2 type, removes over 99% of all introns and recognizes the 5' GU and 3' AG splice sites. It consists of the snRNPs U1, U2, U4, U5, and U6. The minor spliceosome, or U12 type, recognizes other canonical consensus sequences and consists of U11, U12, U4atac, U5, and U6atac (Turunen et al. 2013). The spliceosome is a single turnover enzyme complex assembled stepwise on every individual intron.

To start the splicing process, U1 binds at the 5' splice site, and the U2 auxiliary factors (U2AF) bind at the branch point, the polypyrimidine tract, and the 5' splice site, thus forming the E complex (Figure 1-2). The A complex is formed when U2 binds the BP. The binding of the preorganized tri-snRNP complex comprised of U4, U5, and U6 forms the B complex. What follows is an extensive rearrangement and activation of the B complex (Brow 2002). During this activation, U1 and U4 are discharged, and the conformation and composition are changed to the juxtaposition of the 5' splice site and the BP. That ultimately forms the C

complex. The 5' and 3' exons are brought into contact for ligation, which results in the spliced mRNA and an intron lariat. Other non-snRNPs are crucial for correct splicing, like the NineTeen complex (Hogg et al. 2010) or numerous DExD/H-box helicases (Staley and Guthrie 1998; Bortoli et al. 2021).

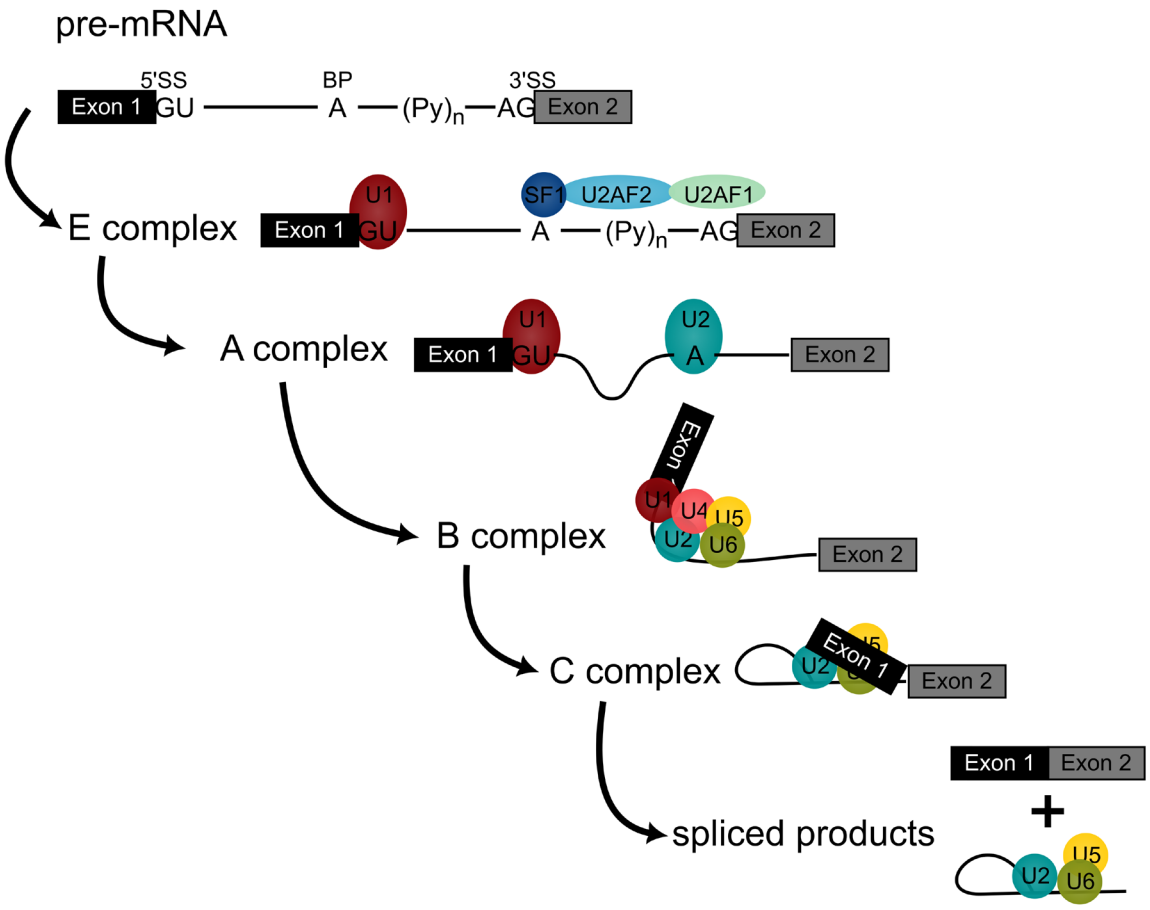

Figure 1-2: Detailed splicing mechanism with the main components of the spliceosome. Introns are depicted as rectangles and introns as lines with their parts in black. All the spliceosome ribonucleoproteins are shown in colorful circles.

1.1.2 Alternative splicing

The human genome has a relatively moderate number of less than 20,000 protein-coding genes (Amaral et al. 2023), while some animals and plants exceed this number (Pertea and Salzberg 2010). However, in humans, genes can be dynamically changed in a much more complex way, and one of these mechanisms that broaden the genetic spectrum is alternative splicing. While some introns connecting two constitutive exons on genomic level are always removed in healthy organisms during splicing in the same way, other introns lack reliable splice efficiency. Usually, this is due to change in the consensus sequences at the splice sites and subsequently an increased reliance on other additional factors for efficient splicing (Matlin et al. 2005). Gene isoforms resulting from alternative splicing have various effects on

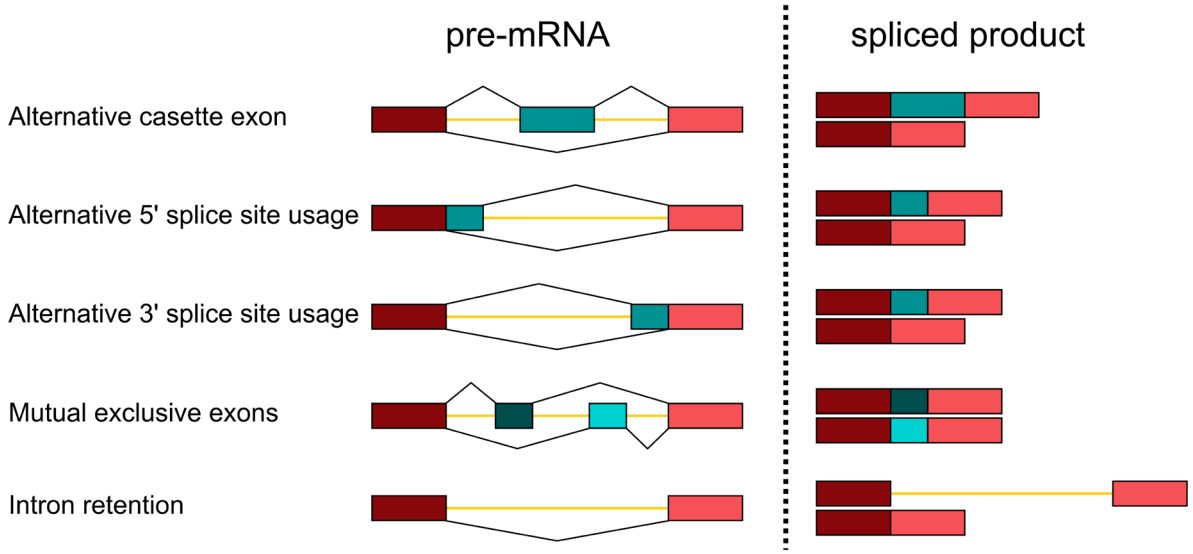

Figure 1-3: The main mechanics of alternative splicing. On the left is the different splicing mechanic are depicted with the flanking exons in shades of red, alternative exons in shades of green and the introns depicted as yellow lines. Black lines connecting the exons show the exon junctions. On the right depicted are the spliced products.

translation, mRNA stability, or on the translated protein. This expands the diversity of the human genome immensely, as roughly 90% of all genes have at least one alternatively spliced exon (Wang et al. 2008). Advanced mass spectrometry methods allow us to examine the effect of alternative splicing on the proteome, and showed that a majority of alternative splicing isoforms are actually translated into protein (Sinitcyn et al. 2023).

Alternative splicing comes in a few different forms as illustrated in Figure 1-3. The most common form in humans is the “alternative cassette exon”, where a whole exon is either skipped or included. A different mechanism of the “alternative 5’ or 3’ splice site”, where the spliceosome can choose a different splice site, causing the corresponding exon tt change its size. “Mutual exclusive exons” usually lie next to each other and cannot be included in the same mRNA. “Intron retention” is simply the lack of splicing that results in the retention of the whole intron between two exons (Figure 1-3). Different organisms can have different preferences for using particular alternative splicing mechanics.

The consequence of alternative splicing are manifold (Liu et al. 2017b). Alternatively spliced exons in protein-coding regions placed in organized areas can drastically change the structure of the protein, altering its functionality. Equally important but more subtle changes happen when alternative splicing controls the appearance of phosphorylation or glycosylation sites. Even changes in disorganized protein regions can significantly change the interactions with other proteins (Yang et al. 2016a). Alternative splicing can also include poison exons, which contain a premature-stop-codon. This leads to nonsense-mediated decay (NMD), which occurs during translation when the ribosome reaches a stop codon before all exon junction complexes are removed. The result is the degradation of the mRNA. Hence, this process is regarded as regulation of gene expression levels (Lareau et al. 2007). The

alternative splicing of non-coding RNA is also relevant for its regulatory mechanisms, for example the inclusion of alternative polyadenylation signals.

1.1.3 Regulation of alternative splicing

As mentioned, 5' and 3' splice sites and BPs are conserved through most species to ensure proper splicing. A distinct pattern can be detected when comparing the splice sites and BPs across species. In budding yeast, the conservation of these sites is stringent, while the more complex an organism is, the less conserved the splice sites are (Will and Lührmann 2011). Less conservation of splice sites can lead to less efficient splicing. This, in turn, increases the possibility of alternative splicing, as the influence of other binding factors become more relevant.

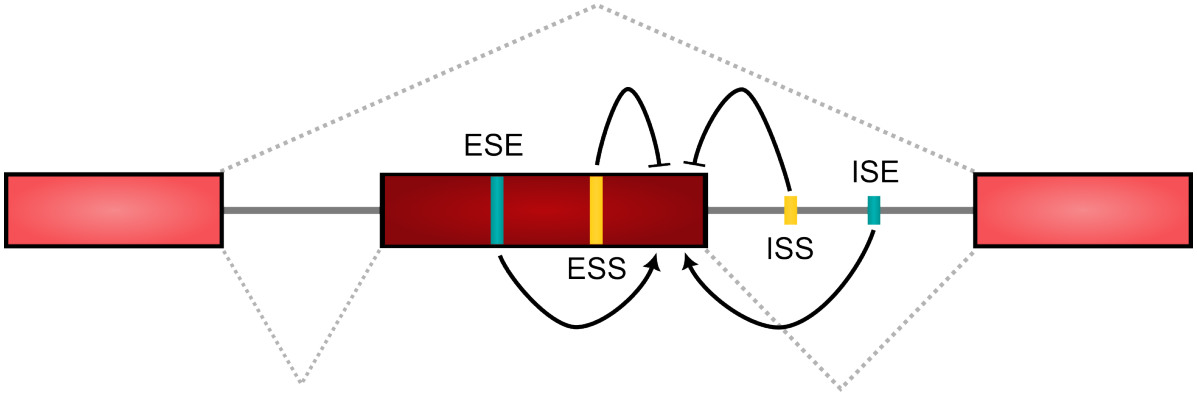

Figure 1-4: Regulation of alternative splicing. Shown is a classical alternative cassette exon, with the exonic splice enhancer (ESE) and silencer (ESS) shown as green and yellow markers in the exon, and the intronic splice enhancer (ISE) and silencer (ISS) as marked regions in the introns.

Non-spliceosomal RNA-binding proteins have long been known to regulate alternative splicing (Chen and Manley 2009). They are roughly divided into three categories: Heterogenous nuclear ribonucleoproteins (hnRNP) (Martinez-Contreras et al. 2007), the serine-arginine rich (SR) proteins (Long and Caceres 2009), and tissue-specific RNA-binding proteins like nova, PTB/hnRNPI (Darnell 2013), the RBFOX family (Lovci et al. 2013), or the musclebound/CELF family (Wang et al. 2012). SR proteins have been shown to generally activate splicing (Änkö 2014), while hnRNP instead represses splicing (Martinez-Contreras et al. 2007). The tissue-specific splicing factors are very dynamic and regulate in both directions. As an example, hnRNPL can act as a splicing repressor or enhancer, depending on the binding position relative to the alternative exon (Rossbach et al. 2014). In general, the effect of splicing regulators depends strongly on the region in which they bind. These regions can be found on exons or introns and are called exon/introns splice enhancers/silencers (Figure 1-4).

1.2 The basics of in silico analysis of transcriptome data

Transcriptomic analysis is a powerful tool in modern science. It revolves around examining the totality of RNAs present in a sample. This includes gathering information on all types of RNA like mRNA, small-noncoding RNA (snRNA), ribosomal RNA (rRNA), or others. The basis for transcriptome analysis is usually with RNA sequencing (RNA-seq). Sanger sequencing was very powerful for initial mass sequencing of DNA, as it sequences single nucleotide strands with a high precision per reaction. However, RNA-seq does not need these properties. Instead of high precision for every single nucleotide, the abundance of transcripts for differential gene expression and isoform analysis is necessary in RNA-seq. RNA-microarrays were the first accurate method for quantitative transcriptome analysis (Shi et al. 2006; Zilliox and Irizarry 2007). Their mechanism relies on hybridization of tagged complementary DNA (cDNA) fragments on microarrays with pre-determined transcript probes and measuring their abundance per well chemically of with fluorescence. This makes RNA-microarrays prone to deterministic biases and limiting on signal range. Next-generation sequencing was another big milestone. Its ability to give a multi-read deep sequence and ever-lower cost led to an overabundance of information waiting to be analyzed (Metzker 2010). The most common next-generation sequencing methods, like Illumina or Ion torrent, consist of a DNA fragmentation of approximately 150 nt for multilayered short-read sequencing and post-sequencing assembly of the short reads. To use next-generation sequencing for RNA, the RNA usually gets reverse-transcribed into cDNA beforehand. Third-generation techniques, like Pacific Bioscience or Oxford Nanopore, achieve much longer reads of 10k nt and higher. Since it doesn't need to be assembled, this can be used to sequence whole transcripts. By sequencing RNA directly, third-generation sequencing now offers the opportunity to sequence RNA without relying on cDNA, thereby circumventing the cDNA bias. Still, even with the most advanced sequencing techniques available today, a major problem for quantitative transcriptome analyses is the low read coverage.

1.2.1 Next generation sequencing

Next-generation sequencing, or second-generation sequencing, is the next step of progress developed from the previously used first-generation sequencing, such as Sanger sequencing. Massive sequencing of short reads amplified clonally gives it a significant advantage for later quantitative analyses (Tucker et al. 2009). Sequencing is performed on DNA molecules, so RNA must be reverse transcribed into cDNA to be properly sequenced. The source RNA should be of high quality, which can be determined by electrophoresis systems or standard sensitivity RNA chip. RNA preparations from mammalian cells typically contain up to 90% rRNA, which, in most cases, is not the focus of the examination. Several RNA extraction kits have been developed to remove rRNA from samples. Before the RNA is transcribed into

cDNA, it must be fragmented to the desired length. Prior to sequencing, oligonucleotide adapters are attached to the cDNA fragments (Head et al. 2014).

The actual sequencing process is divided into two steps. First, the fragments are clonally amplified, which is necessary to obtain a strong signal from multiple strands. Second, the individual nucleotides per fragment are detected. The main principle here is usually sequencing by synthesis. That means, that one complementary nucleotide is added to the clonal group of fragments and chemically blocked, to prevent rampant synthesis. After obtaining the information of the nucleotide via fluorescence or electrical signal, the blocking group is removed, and the mechanism is repeated. The output is millions of raw reads. Subsequent analysis steps need to be performed to obtain the desired results.

1.2.2 Gene expression and splicing analysis

The multi-parallel methods of next-generation sequencing supplies large amounts of data, which makes the performance of bioinformatical processes necessary. The sequencing machines themselves are undertaking the first steps. The electrical or fluorescence signals are translated into nucleotides, reads are trimmed and filtered. Sequencing information is gathered, along with quality scores, and stored in the FASTQ format. These are known as the “raw reads”, which are then subject to further processing.

FASTQ files contain unmapped reads with a quality score for each nucleotide. These reads

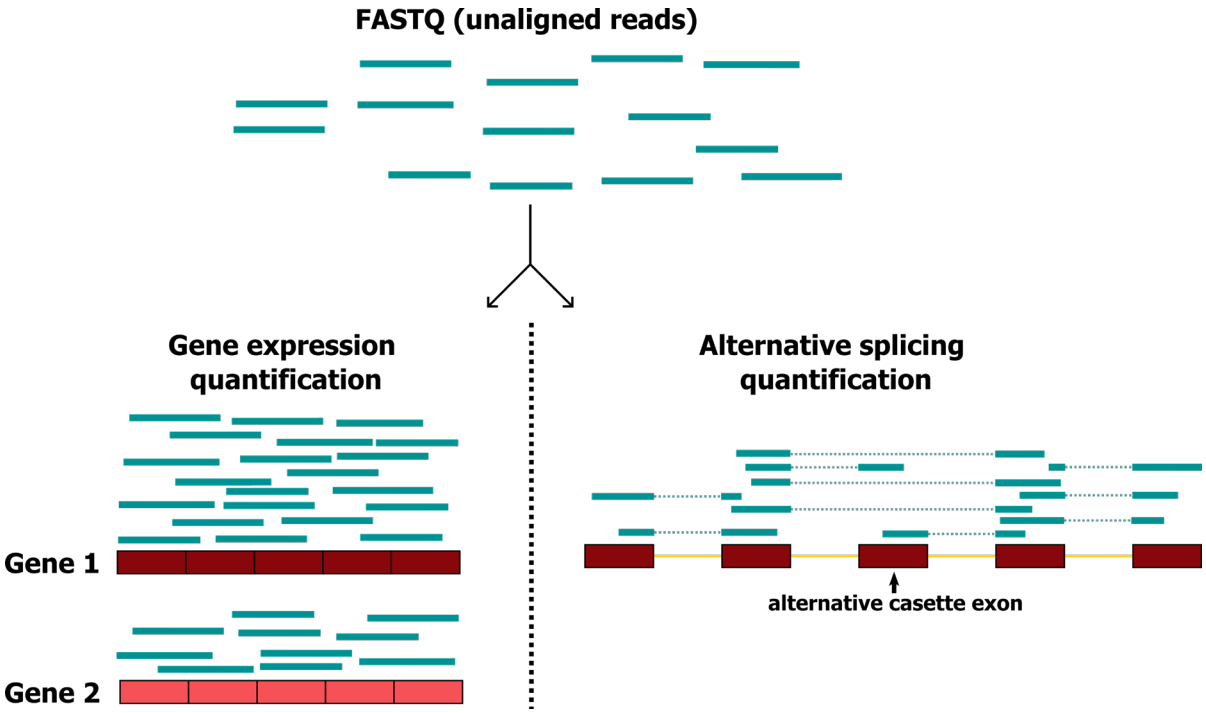

Figure 1-5: RNA-seq analysis methods. Raw, unaligned reads from FASTQ files can be aligned for gene expression analysis depicted on the left. For that the reads can be aligned on a transcriptome. Alternatively they can be aligned on a genome, depicted on the right, spanning over introns and generating the possibility of quantification of alternative splicing.

need to be aligned to a reference genome. FASTQ files from RNA samples do not match the genomic reference perfectly, as in RNA samples, introns are missing. One way to approach this situation is to use transcriptome references, where introns are already removed (Figure 1-5, left). Quality control parameters can also be applied to the mapping, with the percentage of mapped reads estimating overall sequencing quality and DNA contamination. Another parameter is the uniformity of aligned reads.

Gene expression analysis generally examines the abundance of mapped reads on each transcript. The simplest way is counting the raw reads on each transcript, done with HTSeq.count (Anders et al. 2015), or feature counts (Liao et al. 2014). These methods can give an in-depth view of the data without any normalization. However, they are not well suited to answer biological questions. To use the data for gene expression analysis, it must be normalized, as transcript length, total numbers of reads, and sequencing bias must be removed. Therefore, normalized concepts like Reads per kilobase of transcript per Million mapped reads (RPKM) were developed (Mortazavi et al. 2008). Fragments per kilobase of exon per million (FPKM) and transcript per million (TPM) are similar methods. These remove biases constructed by the feature-length and library-size effect. Correcting gene length is necessary for sample comparison of gene expression, which these methods excel at. However, differential gene expression between samples required the sequencing depth, that RPKM, FPKM, and TPM remove. These methods also usually show more bias and are therefore not well suited for differential gene expression. Additionally crucial for differential gene expression are changes in transcript length across samples or conditions (Trapnell et al. 2013), positional biases in coverage along the transcripts, average fragment size (Roberts et al. 2011), and GC content. The most commonly used, normalized methods for inter-sample gene expression analysis are TMM (Robinson and Oshlack 2010), DEseq (Anders and Huber 2010), PiossonSeq (Li et al. 2012), and UpperQuartile (Bullard et al. 2010). DEseq needs pre-normalized data for its quantification and statistics. Salmon is a tool, which aligns fast on transcriptome and normalizes the data in one convenient procedure (Patro et al. 2017).

Gene expression analysis is only one-half of the possible transcriptome analysis that can be performed on RNA-seq data. Another is splicing analysis, which examines the abundance of transcript isoforms caused by alternative splicing. The tools for these analyses can be categorized into two groups. The first group bases their splicing analysis on differential isoform expression analysis. CuffDiff estimates the isoform expression and then compares the different isoform usage (Trapnell et al. 2012). Another tool, rSEQDiff, simultaneously analyses differential isoform expression and differential gene expression (Shi and Jiang 2013). The second group of tools for splicing analysis is focused on single alternative splice events, regardless of whole isoform abundance. DEXseq (Anders et al. 2012) and DSGseq

(Wang et al. 2013) work on the principle of testing for significance between differences in read counts on exons and introns. rMATS quantifies event inclusion using a ratio of the percent-splice-in (PSI) with junction reads (Shen et al. 2014). A requirement of these tools is the alignment of the reads on a genome to span the reads over introns and provide critical information for the tool (Figure 1-5, right). Conversely, rDiff detects alternative splicing via read counts on alternative regions on the genes with or without the alternative splice event (Drewe et al. 2013). A lot of splicing analysis tools require exon junction reads for their quantification. As such, longer reads increase the probability of them spanning over introns. The shorter the average read length is, the fewer reads can be used for the splicing quantification (Xie et al. 2014). An improvement is the usage of multiple RNA-seq analysis tools, as it minimizes the occurrences of false positive results (Olofsson et al. 2023).

1.3 Secretion

A fantastic feat of cellular life is converting a plurality of lifeless chemical reactions into a complex biologically active construct. An essential requirement is the compartmentation of cells, so a plethora of enzymatic reactions can occur in their optimal chemical environment at the same time. A mosaic of organelles, vesicles, and tubes, separated by lipid membranes, is one building block of cellular life. This, however, requires extensive regulatory mechanisms for intracellular trafficking. Multicellular life additionally depends on the synergy of multiple specialized cell types, which depend on each other to work together and ensure the survival of the whole organism. Communication with other surrounding cells and exchanging nutrients and signal molecules is critical. The cell employs highly regulated mechanisms for the secretion and embedding on its outer cell membrane or into extracellular space.

Secretion of protein cargo is an essential aspect of the cellular system, as 11 to 15% of all proteins are secreted as soluble proteins (Uhlén et al. 2015; Uhlén et al. 2019), and another 10% are transported to the plasma membrane (Thul et al. 2017). But a much larger fraction of protein is secreted, as 30% of all mammalian protein mass alone is secreted collagen (Ricard-Blum 2011). The secretory pathway starts in the endoplasmic reticulum (ER). There, chaperones bind to the hydrophobic areas of the proteins and guide them into the correct folding by hydrolyzing adenosine triphosphate (ATP). Before the release from the ER, glycans can be covalently added to the proteins (Xu and Ng 2015). The proteins then exit the ER on a ribosome-free part of the ER with the help of the coat protein complex II (COPII). COPII mainly comprises SEC proteins and the small GTPase SAR1 (McCaughey and Stephens 2018). The COPII was examined in yeast as the transporter between the ER and the Golgi-apparatus (Golgi) (Barlowe et al. 1994). In the more complex human cells, COPII is responsible for the transport of proteins from the ER to the ER Golgi intermediate complex

(ERGIC) (Appenzeller-Herzog and Hauri 2006). The release of protein from the ER, mediated by COPII, is the first step in the secretion pathway and, for some cargo, a critical regulatory process.

The further transport to the Golgi from the ERGIC is realized by the coat protein complex I (COPI), as well as the transport from the Golgi to the ER, and intra-Golgi transport. COPI is associated with the Arf GTPase family (Taylor et al. 2023). The Golgi is crucial for protein modification and has been suggested to be central for regulating secretory sorting. In the Golgi glycosylation is progressively carried out (Ohtsubo and Marth 2006), and proteins are sorted to their final destination. These proteins can then take one of four routes. The first path is back to the ER via COPI, mediated by the lysine, aspartic acid, glutamic acid, and leucine (KDEL) motive (Munro and Pelham 1987). The second path is restraining Golgi residents like glycosyltransferases (Tu and Banfield 2010). The third path is the export out of the Golgi via clathrin-coated vesicles to the endosomal system (Chen et al. 2017). The fourth path is the transport from the Golgi to the plasma membrane for subsequent secretion into extracellular space or embedding into the membrane. No single mechanism was found to be responsible for the transport between the trans-Golgi network and the cell membrane; instead, multiple systems are believed to play a role. Among the candidates for molecules that mediate this process, are tubular structures, which are coated with Arf GTPases (Bottanelli et al. 2017), biosynthetic LAMP1 carriers (Chen et al. 2017), Rab6 associated carriers (Miserey-Lenkei et al. 2017), CtBP1-s/BARS carrier (Valente et al. 2012) or carriers of the TGN to the cell surface (CARTS) (Wakana et al. 2013).

1.3.1 COPII

As the first step in secretion, the assembly of COPII is a vital regulator of the export of proteins and, in some cases, controls the secretion efficiency. The COPII machinery can bend the ER membrane extensively into vesicles and tubule structures. It is a highly conserved mechanism, as all components of COPII were present in the last common eukaryotic ancestor (Schlacht and Dacks 2015). As a simple description, COPII consists of a cage comprising an inner and an outer layer, surrounding and bending the ER membrane. To better understand the regulatory effect of COPII, the mechanism of its assembly must be understood.

The COPII assembly starts with the 250 kDa protein SEC16. It is primarily disordered, serves as a scaffold for COPII budding and organizes and regulates its assembly (Bharucha et al. 2013). The central domain is tightly bound to the ER membrane, while the C-terminus binds other COPII components. Some regard the presence of SEC16 as the definition of the ER exit site (ERES). As the first step, SEC16 recruits SEC12 to the ER membrane (Montegna et al. 2012). SEC12 itself recruits the small GTPase SAR1, for which it acts as a guanine

nucleotide exchange factor (GEF) (Weissman et al. 2001) (Figure 1-6, step 1). SAR1, as one of the first COPII members to be discovered (Saito et al. 1998), is often regarded as an essential regulator of COPII-mediated secretion (van der Verren and Zanetti 2023). COPII dependent secretion continues when SAR1 is removed, but then secretion is highly altered, demonstrating the importance of regulation by SAR1 (Kasberg et al. 2023). Upon exchange of the GDP with GTP under SEC12 influence, the amphipathic helix of SAR1 is extended and inserted into the ER membrane (Huang et al. 2001; Bielli et al. 2005). This is usually regarded as the start of the bending of the ER membrane. The GTP inside of SAR1 is not hydrolyzed, as the GTPase activity of SAR1 alone is insufficient. For that, the GTPase-activating protein (GAP) SEC23, with SEC24, is recruited to SAR1 and enhances its GTPase activity (Saito et al. 1998) (Figure 1-6, step 2). SEC23 and SEC24 together make up the inner coat of the COPII. SEC24 is the cargo binder of the COPII and binds many different cargos directly when presented as transmembrane domains through the ER membrane or with the help of transmembrane adaptor proteins. With four isoforms, SEC24 is highly versatile and specific in cargo recognition and can be highly specific in which cargo is secreted. With an additional palette of adaptor proteins, COPII is able to very specifically sort cargo (Miller et al. 2002; Miller et al. 2003; Adolf et al. 2016). The cargo binding by SEC24 leads to an accumulation of proteins in the ER lumen at the side of the COPII. The increased GTPase activity of SAR1 recruits the SEC13-SEC31 heterodimer, which constructs the outer coat of COPII (Figure 1-6, step 3). This outer coat gives the COPII rigidity and increases the bending of the ER

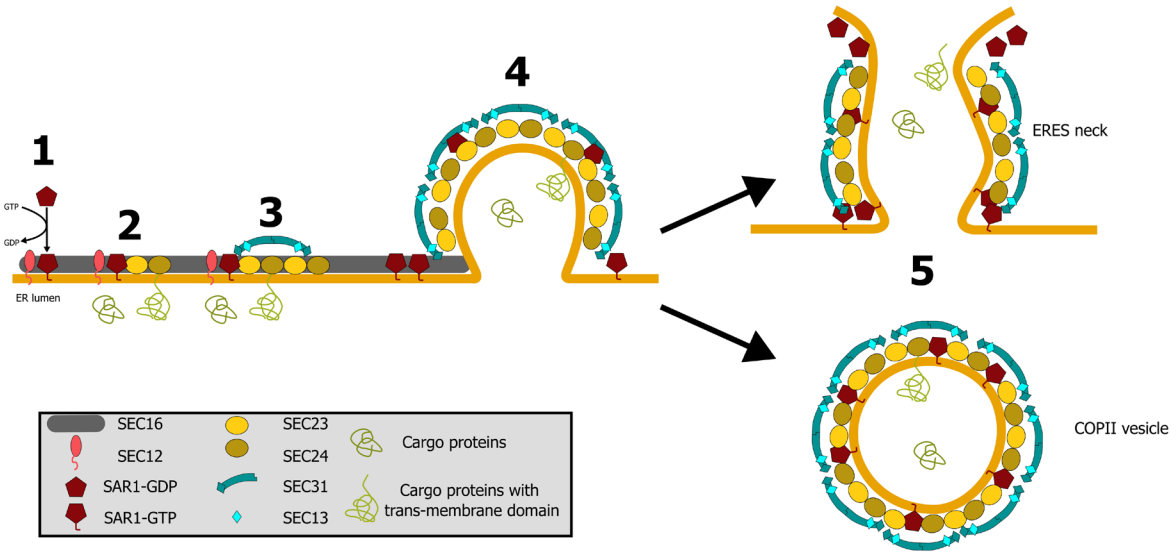

Figure 1-6: Schematic overview of COPII assembly as it bends and forms the ER membrane, depicted as orange line. Step 1: Recruiting of SEC12 and SAR1 at the ERES through SEC16 and GDP-GTP exchange at SAR1 mediated through SEC12. Step 2: Recruiting of the inner coat SEC23, interaction with SAR1 and SEC24 with binding of luminal cargo. Step 3: GTPase activity driven recruitment of outer coat SEC13-SEC31, enhancement of GTPase activity through SEC31. Step 4: Increased GTPase activity and recruitment of COPII compartments lead to extensive bending of ER membrane. Step 5: The result of the COPII machinery. It either creates a tubule structure that is necessary for secretion of cargo to the ERGIC, or budding of COPII vesicles. Adapted from (McCaughey and Stephens 2018)

membrane (Hutchings et al. 2018). The proline-rich C-terminus of SEC31 also interacts with the GTPase active pocket of SAR1 and increases the GTPase activity, subsequently acting as a GAP-helper (Bi et al. 2007; Hutchings et al. 2021). That increased GTPase activity further, accelerates the recruitment of the outer coat of COPII, thus increasing the rate of assembly at the curvature of the ER membrane (Figure 1-6, step 4). The established model of COPII transport suggests that the increased SAR1 activity would eventually lead to a vesicle fission and the released COPII-coated vesicles filled with cargo which are transported to the ERGIC (Peotter et al. 2019) (Figure 1-6, step 5). However, commonly secreted oversized cargos, like collagens and chylomicrons, are too large to fit into the 90 nm big COPII vesicles. Therefore, a lot of research efforts have been put into finding the mechanisms for the secretion of oversized cargo, which could not fit into classical COPII vesicles (Malhotra and Erlmann 2011). A recent study suggested that COPII does not actually form vesicles but instead opens a neck between the ER and the ERES, which subsequently transports proteins to the ERGIC via tubule structures. The authors claim that Cargo is transported via tubule structures mediated by COPI (Weigel et al. 2021). Regardless of the model, COPII seems to be an essential regulator of the early secretory pathway and an exciting object of further investigation.

1.3.2 COPII specializations for big cargos

COPII assembly and export from the ER are essential steps for secretion and are highly regulated and versatile processes. Most compartments can be modified or specialized for the numerous cargo proteins and their specific requirements. Huge cargo like pro-collagens and chylomicrons are key molecules that can be used to study specialized COPII, as they usually do not fit into the standard COPII vesicle.

One COPII component that has been shown to regulate the secretion of large cargo is SEC31. It is an essential interaction partner for SEC13, SEC24, and SEC23 and critical for the export from the ER (Salama et al. 1997). Many features of SEC31 can help with the secretion of special cargo and interactions with other regulators, for example, its C-terminus. This area is, among others, responsible for the interaction between SEC31 and SEC13. Its characteristics create a perfect balance between rigidity to bend the ER membrane and flexibility to create flexible coats that help secrete oversized cargo (Paraan et al. 2018). Calcium is known for having a significant influence on COPII-mediated ER export. It was reported that luminal calcium could regulate the flux of cargo through the apoptosis-linked gene 2 (ALG2) (Helm, JR et al. 2014) via interaction with SEC31 (Yamasaki et al. 2006; Shibata et al. 2007). Post-translational modification on SEC31 also plays a big role in regulating specialized secretion. The phosphorylation of SEC31 decreases the interaction with SEC23 and ER membrane, thus regulating secretion efficiency. CK2 was identified as a SEC31-specific kinase, thus

being essential for secretory management (Koreishi et al. 2013). Monoubiquitylation of SEC31 is another mechanism that changes the behavior of the COPII cage. It can increase the flexibility of the outer cage and is mediated by KLHL12 (Jin et al. 2012).

The interaction with other COPII constituents is crucial for the secretion of big cargo. A tight interaction between SEC31 and SEC23 is known to be beneficial for the efficiency of pro-collagen secretion. It has been suggested that the extended presence of COPII proteins at the ER membrane and their prolonged interaction helps support the secretion of big cargo (Kim et al. 2012). Mutations in SEC23 that impair this interaction leads to the autosomal recessive disorder Cranio-lenticulo-sutural dysplasia (Townley et al. 2008). SEC31 interacts with SEC23 and SAR1. There, it inserts a tryptophan, asparagine, aspartic acid (WND) motive from its proline-rich region into the active pocket of SAR1. This helps to shield it from solvents and pushes the His78 of SAR1 into the catalytic side (Bi et al. 2002; Bi et al. 2007). SAR1, as a central actor of COPII, can act as a regulator. Mutations in its isoforms, SAR1A and SAR1B, are known to be associated with lipid retention diseases (Jones et al. 2003; Levy et al. 2019); however, cells in culture show survival when either of the isoforms is impaired (Cutrona et al. 2013; Sané et al. 2017).

Another mechanism through which SAR1 GTPase activity is used for specialized COPII coats is through the binding of the proteins cutaneous T-cell lymphoma-associated antigen 5 (cTAGE5) and transport and Golgi organization 1 (TANGO1) to SEC12, which is the GEF of SAR1 (Saito et al. 2017). TANGO1 is a protein that closely interacts with the COPII machinery for the secretion of special cargo. It was found to be responsible for the secretion of pro-collagens (Wilson et al. 2011) and then later for other big cargos (Ríos-Barrera et al. 2017) like chylomicrons (Santos et al. 2016). The mechanism is mediated by the close binding of the binding of cTAGE5 (Saito et al. 2011) and TANGO1-like (TALI) (Santos et al. 2016). With the influence of TANGO1 on SEC12, it regulates the fundamentals of COPII (Saito et al. 2014). TANGO1 forms rings at the ERES, suggesting a stabilizing effect on COPII on the ER membrane and preventing early fissure, thereby increases the efficiency of secreting large cargo (Raote et al. 2017). A recent study showed that TANGO1 enables collagen secretion by creating inter-organelle tunnels between ER and ERGIC. TANGO1 is supposed to induce entropic forces with its intrinsically disordered regions that pushes pro-collagen towards ERGIC (Bunel et al. 2024). Generally, a functional COPII machinery is needed for the secretion of big cargos, as impairment of many components like SEC13 (Townley et al. 2008; Townley et al. 2012), SEC24D (Garbes et al. 2015; Ohisa et al. 2010), SEC23A (Boyadjiev et al. 2006), or SEC23B (Lang et al. 2006) can lead to severe disease phenotypes.

1.3.3 Secretion in EMT

Secretion is a universal mechanism in most human cell types. It can be part of the regulation of cell type transitions, as well as being regulated by different cellular processes. One of these processes the epithelial-mesenchymal transition (EMT), the reverse counterpart of the mesenchymal-epithelial transition (MET). As the names suggest, it is the reversible transition between epithelial and mesenchymal cells. During EMT, epithelial cells lose their junctions with neighboring cells and their apical-basal polarity, their cytoskeleton is reorganized, and they lose their cell shape; the gene expression and splicing are heavily altered, and they gain motility and invasiveness (Thiery and Sleeman 2006; Thiery et al. 2009). Downregulation of E-cadherin and upregulation of N-cadherin are key molecular aspects in EMT (Taube et al. 2010; Thomson et al. 2011). The upregulation of the production of extracellular matrix proteins, like collagens, and their secretion is a key aspect of the MET (Bryant and Mostov 2008). While EMT is reported to be a leading factor in wound healing and cancer progression (Thiery et al. 2009; Chapman 2011), MET is known to be responsible for the development of various tissues like kidneys and intestines (Chaffer et al. 2007). In general, constitutive rounds of EMT and MET are required in human development (Chaffer et al. 2007; Lim and Thiery 2012). As the EMT enhances mobility and metastatic behavior, it has been a target for cancer research (Hanahan and Weinberg 2011)

Various transcription factors (TF) are involved in EMT. Most of them activate each other, creating feedback loops that accelerate the transition. SNAIL1 expression is a central factor, which gets enhanced by TGFβ, WNT proteins, and NOTCH (Peinado et al. 2007). Also involved in EMT are ZEB1/2 and TWIST. In addition to that, alternative splicing, as well as gene expression, plays a pivotal role in EMT / MET. It was reported that RBPs KH domain-containing RNA-binding protein (QKI) and RNA binding protein fox-1 homolog 2 (RBFOX2) enhance splice events, which lead to EMT, while the RBPs epithelial splicing regulatory protein 1/2 (ESPR1/2) and RNA-binding motive 47 (RBM47) increase MET-specific splice events (Yang et al. 2016b).

COPII plays a dual role of being regulated by and regulating EMT, as its compartments are upregulated in *drosophila melanogaster* during the development of the salivary gland, a tubular organ responsible for high secretion. There, COPII regulates the secretion of large cargo (Abrams and Andrew 2005). From another perspective, COPII is also responsible for the proper mechanism of MET. It is known that COPII stabilizes the polarity of epidermal differentiation in *drosophila* (Norum et al. 2010) and is required for apical protein secretion (Grieder et al. 2008). The importance of MET has also been reported and validated in human cell cultures. Caco-2 cells with deficient SEC13 show defective polarity, adhesion, and differentiation (Townley et al. 2012).

1.3.4 Secretion in B-lymphocytes

Antibodies (ab) play a crucial role in the adaptive immune system for antigen recognition. They are produced by and secreted from antibody-secreting cells (ASC), a B-cell subpopulation. Naïve B-cells, after T-cell dependent activation, can differentiate into ASC and memory-B-cells (MBC) for direct and long-term unity against specific antigens (Radbruch et al. 2006).

The T-cell-dependent ASC differentiation happens in two steps. The first step is called the extrafollicular step. B-cells act as antigen-presenting cells and load fractions of antigens onto the major histocompatibility complex II. These in turn are recognized by T-helper cells and activate the B-cells. The B-cells undergo a class-switch recombination and differentiate into short-lived plasmablasts (PB) (MacLennan et al. 2003). These ASCs mediate a fast adaptive ab response to new antigens. PBs secrete more ab than B-cells, but less than fully developed plasma cells (PC). They divide rapidly and are still able to present antigens. The second step of ASC differentiation happens when activated B-cells enter the B-cell follicle. There, they build germinal centers with the help of specialized T follicular helper cells (Victora and Nussenzweig 2012). The germinal centers are transiently formed structures of extensive proliferation and selection (Shlomchik and Weisel 2012). The B-cells subsequently undergo somatic hypermutation in these germinal centers of the B-cell follicles. The results are long-lived PC with a high yield of ab with extensive affinity (Radbruch et al. 2006). In the same germinal centers, B-cells can also differentiate into MBCs (Kometani et al. 2013). These extremely long-lived cells do not secrete ab but expose ab on their surface and get activated after antigen binding, subsequently differentiate back to PB and PC (Nutt et al. 2015).

The ASC transcriptome is mainly influenced by the transcription factor PAX5 (Cobaleda et al. 2007), but other factors also regulate the differentiation. The unique challenge of ASC and especially PC is their high secretion load in combination with longevity. ASC adapt to these challenges through an enlarged ER and Golgi for extensive secretion (Nguyen et al. 2019; Tellier and Nutt 2019). Two additional TFs are of interest for B-cell differentiation in secretion. The TF IRF4 was shown to target Prdm1 and is essential for B-cell survival (Sciammas et al. 2006). Contrary to that, the TF BLIMP-1 is dispensable for survival but is required for proper ab secretion (Tellier et al. 2016). The ab get further processed and secreted there (Kirk et al. 2010).

1.4 Objectives

This thesis aims to find and characterize a connection between alternative splice events and the regulating of secretion specificity. In initial inspection of RNA-seq data, we discovered an alternative splicing event in SEC31A, which is highly tissue-specific spliced. This alternatively

spliced exon encodes peptides near the interaction side of SEC31A with SEC23 and SAR1. Based on these results, we propose the hypothesis that this splicing event could regulate the secretion of specific cargo. Additionally, we are investigating if the previously revealed alternative splice site in SEC31A (Neumann et al. 2019) could be responsible for secretory specificity. We are using microscopy techniques and a lipid permeability assay in Caco-2 cells to investigate our hypothesis. Finally, we are using in-silico methods to examine the change of splicing and gene expression in differentiating B-cells and the potential influence on intracellular trafficking and secretion.

2 Materials and Methods

2.1 Materials

2.1.1 Used data and gene ontology (GO) terms

Table 1: Summary of all used, publicly available RNA-seq data sets.

Accession number	Description in text	Description	Publication
PRJEB4337	human tissue data	Collection of tissue samples from healthy humans as frozen specimens, characterized by pathologists before RNA sequencing. Average spot length: 202, paired end. Sample numbers vary between 3 and 9 per tissue.	(Fagerberg et al. 2014)
PRJNA304418	H358 siRBM47 data	Human epithelial-like lung cancer cell line H358 was treated with either control siRNA or RBM47 siRNA, and subsequently, RNA was sequenced. Average spot length: 200, paired end. Sample number per treatment: 3	(Yang et al. 2016b)
PRJNA626563	B-cell data	B-cells, by healthy donors, were differentiated in vitro and RNA sequenced. Average spot length: 151, paired end. Sample number per treatment: 3	(Kassambara et al. 2021; Alaterre et al. 2021)

Table 2: Used gene ontology terms.

GO term number	GO term name
GO:0098826	endoplasmic reticulum tubular network membrane
GO:0031201	SNARE complex
GO:0061852	retrograde transporter complex, Golgi to ER
GO:0030127	COPII vesicle coat
GO:0012507	ER to Golgi transport vesicle membrane
GO:0030136	clathrin-coated vesicle
GO:0070971	endoplasmic reticulum exit site
GO:0033116	endoplasmic reticulum-Golgi intermediate compartment membrane
GO:0032588	trans-Golgi network membrane
GO:0042627	chylomicron
GO:0005581	collagen trimer
GO:0008380	RNA splicing
GO:0032527	protein exit from the endoplasmic reticulum
GO:0090114	COPII-coated vesicle budding

2.1.2 Buffers

5x TBE (Urea gels)

1.8 M	Tris base
1.8 M	Boric acid
440 mM	EDTA pH 8.0

50x TAE (Agarose gels)

2 M	Tris Base
5.7%	Acetic acid
400 mM	EDTA pH 8.0

5x Hybridization buffer (RT)

1.5 M	NaCl
50 mM	Tris pH 7.5
10 mM	EDTA

1.25x RT Mix

12.5 mM	DTT
12.5 mM	Tris pH 8.0
7.5 mM	$MgCl_2$

10x Taq buffer

0.5 M	KCl
0.1 M	Tris pH 8.3
15 mM	MgCl2
0.01 % (v/v)	Gelatin

2x Formamide loading dye

80%	Formamide
1 mM	EDTA
0.05% (v/v)	Bromophenol blue
0.05% (v/v)	Xylene cyanol

1x PNK buffer

70 mM	Tris-HCl, pH 8.3
10 mM	$MgCl_2$
5 mM	DTT, pH 7.6

10 X T4 DNA ligase buffer

500 mM	Tris-HCl, pH 7.5
100 mM	MgCl2
100 mM	DTT
10 mM	ATP
250 µg/mL	BSA

2.1.3 Commercial kits and reagents/buffers

2.1.3.1 Kits

Table 3: All commercially available kits used for this work.

Kit	Supplier
Direct-zol RNAMiniPrep	Zymo
NucleoSpin Gel and PCR Clean-up	Macherey-Nagel

2.1.3.2 Reagents/buffers

Table 4: All commercially available reagents and buffers used in this work.

Reagent/buffer	Supplier
RNA Tri-Flüssig Extraktion	Bio&Sell
ROTIfect Transfection Reagent	Carl Roth
Phusion® HF Buffer (5 X)	Mobidiag
Lipofectamine™ 2000 Transfection Reagent	Thermo Fisher Scientific

2.1.4 Antibodies

Table 5: All antibodies used in this work.

Antibody	Usage	Source organism	Supplier
GFP (A11122)	Immunofluorescence, primary antibody	Rabbit	Invitrogen
HA-tag (F-7, sc7392)	Immunofluorescence, primary antibody	Mouse	Santa Cruz
Collagen I (ab34710)	Immunofluorescence, primary antibody	Rabbit	Abcam
Alexa Fluor™ 488 – anti Rabbit	Immunofluorescence, secondary antibody	Goat	Invitrogen
Alexa Fluor™ 647 – anti Rabbit	Immunofluorescence, secondary antibody	Goat	Invitrogen
Alexa Fluor™ 568 – anti Mouse	Immunofluorescence, secondary antibody	Donkey	Invitrogen

2.1.5 Primers

Table 6: Summary of all used primers in this work.

Sequence	Description	Usage
TGTACGCAGCACAGCACC	SEC31A (human) Exon 24 forward	Radioactive PCR

TTCTTCTTTTTGGGTACTCTGTTCAAAG	SEC31A (human) Exon 25 reverse	Radioactive PCR
CCACCTTACCCACAGCCG	SEC31A (mouse) Exon 24 forward	Radioactive PCR
CTCTGTTCAGAGCTGGAGGA	SEC31A (mouse) Exon 25 reverse	Radioactive PCR
ATTCCTACTGCCGGGATCTA	RBM47 (human) forward	qPCR
TAGCCTGCGTATCCTCCATA	RBM47 (human) reverse	qPCR
CGTGGAAAGGGTCAAGAAGAT	RBM47 (mouse) forward	qPCR
TGTGCCGTTGAGGTTGTT	RBM47 (mouse) reverse	qPCR
CTGGACCAGCAGAACTCAGA	MIA2 (human)	qPCR
AGCAAGAGGTGGAGGAACAA	MIA2 (human)	qPCR
AAAACCCGGGTGTTAAGTAATTTTTACATTTTGTCTACAGAAAACCAGTCTATCCAAGACCAGGCACCTATGTTGGAAGGTGAATTGGTTGTAAAATATACCTTTGGAACTCGAGAAAA	Exon 24c short (S) forward	Minigene cloning
TTTTCTCGAGTTCCAAAGGTATATTTTACAACCAATTCACCTTCCAACATAGGTGCCTGGTCTTGGATAGACTGGTTTTCTGTAGACAAAATGTAAAAATTACTTAACACCCGGGTTTT	Exon 24c short (S) reverse	Minigene cloning
AAAACCCGGGtaaagtttagaatgtcttcagtaaacactgttaag	Exon 24c medium (M) forward	Minigene cloning
tgcatgtgaacttcaaagatgatgatatcCTCGAGAAAA	Exon 24c medium (M) reverse	Minigene cloning
AAAACCCGGGtagggaaacattgttattttaaatatgaag	Exon 24c large (L) forward	Minigene cloning
ctatgtagaaccagaagtttaagaatttgtCTCGAGAAAA	Exon 24c large (L) reverse	Minigene cloning
GCAAGCTGACCCTGAAGTTC	GFP forward	Radioactive PCR
GACGACGGCAACTACAAGAC	GFP reverse	Radioactive PCR

2.1.6 Plasmids

Table 7: All used Plasmids in this work.

Name	**Usage**	**Source**
Str-KDEL_SBP-EGFP-Ecadherin	RUSH assay	Addgene
Str-KDEL_SBP_EGFP_GPI	RUSH assay	Addgene
pEGFP-N1-Flag-humanRBM47	RBM47 overexpression	Addgene

pAC1801_pmax-dCasRx	Cas13 guided splicing silencing	Addgene
pXR003_CasRx gRNA cloning backbone	gRNA backbone for Cas13	Addgene
Minigene (GFP splicing reporter)	Minigene cloning	(Neumann et al. 2020)

2.1.7 siRNA and MO

Table 8. Sequences of used siRNA.

Gene Name	Sequence
RBM47	UCGACUACUAUAAGCGAUA
	AUAUGGAGCAGUUCGAACA
	CGGUUAUGGGCCAGAGAGU
	GUCCAAAGAGUGCCCAGUA
MIA2	CCGCCAGGACAAUCAUAUCCUGAUUTT
	GACGAGAUUCUAAUCUUUAUGGUUUGT
siAllstar (control)	UUCUCCGAACGUGUCACGU

Table 9: All used morphlino oligos in this work.

Target	Sequence
SEC31A exon 24c	TTACAACCAATTCACCTTCCAACAT
SEC31A exon 24b	CCCTTTCAAGCGCAGACCTGTTCTT

2.2 Bioinformatics

2.2.1 RNA sequencing analysis

Three data sets were used for this work: a human tissue data set (PRJEB4337), an H358 siRBM47 data set (PRJNA304418), and a B-cell data set (PRJNA626563) (Table 1). Raw reads were downloaded using the SRA toolkit. SRA files were converted to FASTQ files. These files were aligned with STAR (Dobin et al. 2013) to an index generated with the human genome GRCh38/hg38. The unsorted BAM file outputs were indexed and merged using SAMtools (Danecek et al. 2021). The BAM files were used to quantify alternative splicing with the help of rMATS (Shen et al. 2014) on standard settings. In the human tissue data set, where we did not have a control sample group to compare to, we let rMATS run each tissue sample group compared to the liver sample group. To extract quantitative percent spliced-in (PSI) values for each sample, we used Numpy (Harris et al. 2020) and Pandas (The pandas

development team 2023) in Python3 (van Rossum and Drake 2009). We used Salmon (Patro et al. 2017) to align FASTQ reads to the human genome and quantify the gene expression for gene expression analysis. Statistics for the gene expression were done using DEseq2 (Love et al. 2014). All figures were generated using Seaborn (Waskom 2021) and Matplotlib (Thomas A Caswell et al. 2023).

2.2.2 Correlation studies

We performed correlation analyses to better understand the interconnection of genes and splice events. We set one PSI value for each sample for each quantified splice event. Similarly, we generated one value of the normalized gene expression for each sample and each quantified gene. Each value represents the level of gene expression or exon inclusion. Regardless of the different levels of gene expression or exon inclusion, we could calculate correlation indices that show us the correlation of exon inclusion with gene expression. For that, the PSI values of an exon were plotted against the gene expression levels of all of the quantified genes corresponding to the sample. The Spearman or Pearson correlation was calculated between the exon inclusion and the gene expression. The result is a correlation index per gene, showing how much the gene expression correlates with the exon inclusion.

2.2.3 AlphaFold

To predict protein structures of protein isoforms produced by alternative splicing, we used AlphaFold 2 (Jumper et al. 2021). To deal with the considerable amount of computational power, we used the Curta system at the Zedat cluster at the Freie Universität Berlin (Bennett et al. 2020). The predictions for multiple proteins in complex were done with the model preset multimer, and single protein predictions were made with the preset monomer. To safe computational time, we used the reduced database preset. The single residue confidence pLDDT, the predicted aligned error (pae), and the confidence score were extracted using Numpy and saved as numpy arrays. AlphaFold structures in .pdb format were analyzed and prepared for the figures using Pymol (Schrödinger 2015).

2.3 Minigene cloning

To identify the direct involvement of RBM47 in regulating alternative exons, we used an adapted luciferase reporter previously used in our work group (Neumann et al. 2020). It is a GFP vector with an intron included and strong 5’ and 3’ splice sites, leading to constitutive splicing of the GFP transcript in mammalian cells. XmaI and XhoI were used to open the vector without destroying its intronic branchpoint. We amplified the exon with specific primers, including flanking intron regions in varying lengths, using genomic Hek cell DNA as a template, and introducing XmaI and XhoI restriction sites. After creating sticky ends, using

restriction digestion on the insert, we ligated them with the opened GFP intron vector. Positive clones were identified using PCR and Sanger Sequencing.

2.4 Cell culture

2.4.1 Cell lines and subculturing

HEK293, HeLa, and Caco-2 cells were subcultured in DMEM High Glucose (Biowest, France) media supplemented with 10% (v/v) fetal bovine serum (Bio&Sell, Lot: BS.67665481.5T) and 1% (v/v) penicillin/streptomycin (Invitrogen, USA). Cells were kept in humid conditions at 37°C and 5% CO_2. HEK293 and HeLa cells were split 1/10, and Caco-2 cells were split 1/5 every three to four days. For passaging cells in T75 flasks (Sarstedt), they were washed with Dulbecco's Phosphate Buffered Saline (PBS) (Biowest, France) and detached using Trypsine (GIPCO). After three to ten minutes of incubation at 37°C, detached cells were diluted in 9 mL DMEM containing fetal bovine serum. The medium was removed using 5 min centrifugation at 200 g. Cells were resuspended in a fresh medium, sown out in a new T75 flask, and used for further experiments. The cells were tested for *Mycoplasma* sp. contamination once a month.

2.4.2 Transfection

Plasmids and small interfering RNA (siRNA) were transfected in HEK and HeLa cells using RotiFect (Roth) in 12-well plates (Sartstedt). 2 µg plasmid or 20 µM/mL siRNA (Table 8) were mixed in 100 µL Optimem (GIPCO) per well. In parallel, 2 µL Rotifect was mixed in 100 µL Optimem and incubated at room temperature for 5 minutes. Plasmid or siRNA mix and Rotifect mix were mixed and incubated at room temperature for 20 minutes. The transfection mix was added to each well carefully.

Morpholino oligos (MO) transfection of HEK and HeLa cells was done following manufacturer's protocol using endoporter Caco-2 cells that were transfected via electroporation using the Amaxa Nucleofector II. $15x10^5$ cells were washed twice with Optimem. The cells were concentrated to a volume of 100 µL. 7.5 µL MO (10 mM) was added to the cells and the mix was filled into cuvettes. After slight mixing, the cells were electroporated using the Caco-2 setting (B-024).

The siRNA sequence against MIA2 was taken from a previously conducted study (Saito et al. 2011).

2.5 RNA

2.5.1 Extraction

RNA was extracted from cultured cells and mice tissues. The cells were washed once with cold PBS and then subjected to RNA extraction. The liver, kidney, brain, spleen, heart, and lung were carefully removed from the mice and immediately frozen in liquid nitrogen. From the tissues, approximately 5 mm in diameter sections were separated and homogenized in RNA extraction reagent. RNA was extracted with Direct-zol RNA miniprep (Zymo Research) or RNA Tri_Flüssig Extraktion (Bio&Sell) with a subsequent DNAse treatment. RNA extraction with the miniprep kit was done as described in the user manual. For RNA extraction with trizol, the medium of the cells was removed, and the cells were washed with PBS. Cells were resuspended in 500 µL trizol and supplemented with 100 µL Trichloromethane/Chloroform (Carl Roth). The mixture was shaken vigorously and let to rest for ten minutes at 20°C. After incubation, the mixture is centrifuged for ten minutes at 17,000 g at 4°C. The newly separated upper aqueous phase was transferred to a new reaction tube. 300 µL ice-cold 2-propanol was added to the aqueous layer and shaken vigorously. The mixture was again centrifuged at four 4 °C, with 17,000 g for 20 minutes. The liquid was removed, and the pellet was washed once with 70% (v/v) ethanol. The pellet was dried and resuspended with 16 µL RNAse/DNAse free water. To remove DNA contamination, 2 µL 10xDNAse buffer and 2 µL DNase were added and incubated at 37°C for 20 minutes. 180 µL water and 200 µL ROTI®Aqua-P/C/I (Carl Roth) was added and vigorously shaken. The mixture was centrifuged at 17,000 g at 4°C for 15 minutes, and the upper aqueous layer was taken to a new reaction tube containing 1 µL of glycogen, 20 µL 3 M sodium acetate, and 500 µL 100% (v/v) ethanol absolute. After mixing, the mixture was centrifuged for 10 minutes at 4°C at 17,000 g. The liquid was removed, and the pellet washed with 70% (v/v) ethanol. After a short drying, the pellet was resuspended in 12 µL RNAse/DNAse free water. The concentration was measured in a NanoDrop spectrometer at 260 nm and 280 nm wavelengths, and the concentration adjusted to 500 ng/µL.

2.5.2 Reverse transcription

cDNA was manufactured of the RNA using gene-specific reverse primers. 1000 ng of total RNA was mixed with the 5x Hybridization buffer and 0.1 µM/µL reverse primer mix. The primer mix was annealed to the template RNA using the RT hybridization program (Table 10). After annealing, prewarmed (43°C) 19.75 µL 1.25x RT reaction buffer and 50 Units MmuLV reverse transcriptase (Enzymatics) were added to the RT reactions inside the thermocycler, where the reaction was incubated (Table 11).

Table 10 Program of thermocycler for RT hybridization reaction.

Temperature / °C	Cycle	Time / s
90 … 70	1	20
69 … 51		30
50 … 44		40
43		∞

Table 11 Program of thermocycler for RT reaction.

Temperature / °C	Cycle	Time / min
43	1	30
94		5
4		∞

2.5.3 Radioactive RT-PCR and denaturing polyacrylamide gel electrophoresis

RT-PCR was done with radioactive [^{32}P] labeled forward primers. For this, 200 ng of forward primers, 10 µL 10x PNK buffer, 1 µL poly-nucleotide-kinase, 84.5 mL water, and 2.5 µL (2.5 µCl) γ-ATP (Hartmann Analytic) incubated for 45 minutes at 37°C. The reaction was mixed with 20 µL sodium-acetate (3 M, pH 5), 200 mL PCI, 1 µL glycogen, and 80 µL water to purify the primer. It was vigorously mixed and centrifuged for 3 minutes at 10,000 g. The top aqueous layer was transferred to a new reaction tube and supplemented with 100% (v/v) ethanol. After spinning for 10 minutes at 10,000 g, the liquid was discarded, and the pellet washed with 70% (v/v) ethanol. After the last centrifugation step, the pellet was briefly air-dried and resuspended in 80 µL water.

For the radioactive PCR, we used 5 µL of cDNA, 1.5 µL 10x Taq PCR buffer, 2.5 ng forward primer, 5 ng reverse primer, 10 µL water, 2.5 ng 32P-labelled forward primer, and 0.5 µL of Taq polymerase and incubate in the thermocycler (Table 12). The reaction solution was covered in a thin layer of mineral oil. PCR products were removed from the oil and diluted in 2x formamide dye. The products were separated using denaturing 5% polyacrylamide (PAGE) gels with 7 M Urea and 0.5x TBE buffer. The separation was done at 24 mA for 40 minutes. The gels were fixated in a 10% methanol, 10% acetic acid fixation buffer then vacuum-dried and visualized using the phosphorimager Typhoon 9200 (GE Healthcare) and quantified using image QantTL software (Cytiva).

Table 12 Program of thermocycler for radioactive PCR.

Temperature / °C	Cycle	Time / min
95	1	3:00
95	28	0:30
Primer specific (~60)		0:30
72		0:45
72	1	7:00
4	1	∞

2.5.4 RT-qPCR

Gene expression in total RNA was analyzed using RT-qPCR. Gene-specific primers were used in the RT, with an additional housekeeping gene for reference. The housekeeping gene for human cell culture samples was GAPDH, and for mouse samples it was Hprt. The RT-qPCR was performed in a 96-well format using a Blue S'Green qPCR kit (Biozym). 250 nM of gene-specific reverse and forward primer, 1x qPCR S'Green BlueMix, and diluted cDNA were mixed in a total volume of 20 µL. The RT-qPCR reaction was performed in Line-Gene 9600 Plus Real-time PCR System (Bioer Technology Co). The ΔCT was calculated using the CT of the housekeeping gene and of the gene of interest. Gene expression levels were depicted as relative to the housekeeping gene.

2.6 Microscopy

2.6.1 Quantitative RUSH assay

Retention using selective hooks (RUSH) assay was adopted after Boncompain et al. (Boncompain et al. 2012). The commercially available constructs with the glycosylphosphatidylinositol (GPI)-anchor Str-KDEL_SBP-EGFP-GPI (KDEL-GPI) and Str-KDEL_SBP-EGFP-Ecadherin (KDEL-E-cadherin) were bought from addgene. HEK cells had to be seeded on coverslips to be analyzed under the microscope, so coverslips (14 mm ø, VWR) were washed with ethanol and covered with poly-L-lysine (100 g/mL, Sigma Aldrich) for 2 h in the cell culture incubator at 37°C, 5% CO^2, and humid conditions. The poly-L-lysine was removed, and the coverslips were irradiated with UV light. In a 12-well plate, $0.8x10^5$ cells per well were seeded onto the coverslips and attached overnight. On the next day, the cells were transfected with Cas13 and gRNA plasmids, guiding the CAS to the center of the 5'splice site of exon 24 in SEC31A. After another night of incubation, the cells were transfected with the RUSH plasmids (KDEL-GPI or KDEL-Ecadherin). After 16 h of the transfection, 100 µL DMEM containing 10% fetal bovine serum, 57 µM Biotin (Sigma), and

10 µg/mL cycloheximide (Roth) were added to the wells and incubated for another 60 minutes. After that, the medium was removed, and the wells were washed with cold PBS. The coverslips were removed from the wells and placed onto parafilm in a humid, darkened chamber. Cells were fixated by incubating them for 15 minutes in 3.7% formaldehyde (Roth). After that, the cell preparation was washed 3 times with PBS and blocked for 60 minutes with 5% goat serum (Sigma) in PBS. After that, the preparation was incubated for 120 minutes with the primary rabbit anti-GFP antibody in a 1:250 dilution in the blocking solution. After washing with PBS 4 times, the preparation was incubated for 60 minutes with the secondary donkey anti-rabbit Alexa Fluor 647 (Invitrogen) antibody in a 1:250 dilution in blocking solution. The preparation was washed 4 times with PBS, sealed with the ProLong Gold DAPI mounting medium (Invitrogen) and put faced down on a coverslight. Preparations were analyzed under a confocal microscope SP8 (Laica). The colocalization of GFP and Alexa 647 signal was analyzed using the Python script of Alexander Neumann (Neumann et al. 2019).

2.6.2 Collagen secretion assay

Collagen secretion was analyzed by staining of extracellular collagen separately from total collagen. $0.5x10^5$ HeLa cells were seeded on coverslips in 12-well plates. One day after, they were transfected with Cas13 and gRNA directed to SEC31A exon 24 alternative 5' splice site. Two days after the transfection medium of the wells was removed, the coverslips were washed with cold PBS and placed on parafilm in a darkened, humid chamber. Cells were fixated with 3.7% formaldehyde for 10 minutes. After washing the preparations three times, they were blocked using 5% goat serum in PBS. After blocking, they were incubated with a primary rabbit anti-collagen antibody in a 1:500 dilution in blocking solution for 120 minutes. After washing with PBS three times, they were incubated with the donkey anti-rabbit Alexa Fluor 647 antibody in a 1:500 dilution. After washing three times, the preparations were permeabilized with 0.1% Triton-X (Roth) in PBS with the addition of 5% goat serum after the 45 minutes of incubation, and three times washing with PBS, the incubation with the primary antibody was repeated. The Alexa Fluor 488 (Invitrogen) was used as the secondary antibody in a 1:500 dilution in blocking buffer. Finally, the preparations were sealed with the ProLong gold DAPI mounting medium and placed on a coverslight. The preparations were analyzed under the confocal microscope. The colocalization of the red extracellular collagen and the total green collagen was quantified using a modified version of the RUSH quantification script (Neumann et al. 2019).

2.6.3 FRAP

Fluorescence recovery after photobleaching (FRAP) was done using HeLa SEC13-GFP tagged cells provided by Francesca Bottanelli's lab. $1x10^5$ cells were seeded in live-cell imaging chambers and, after one day, transfected with an MO targeting SEC31A exon 24

alternative 5' splice site (Neumann et al. 2019). After two days, the cells were placed under the Laica confocal microscope at 37°C in a 5% CO^2 environment. After 5 frames pre-bleach, solitary slow-moving SEC13 puncta were chosen to be bleached 3 times with a 60% argon laser. The signal was measured up to 50 s after bleaching. Each frame was individually analyzed and normalized to be presented as one experiment. In each frame, the fluorescent intensities of three areas were measured: the bleached puncta ($FRAP_{puncta}$), a background area outside any visible cell compartment, and a reference area with stable fluorescence intensity (ref). First, the background is subtracted from the $FRAP_{puncta}$ and the ref for each time point. To remove the error of the constant photobleaching that occurs automatically when measuring with a confocal microscope, we calculate the rate of photobleaching ***a*** with the time *t* and the fluorescence values at the beginning with ref_0 and at time *t* with ref_t.

$$\mathrm{ref}_t = \mathrm{ref}_0 \cdot e^{-at} \quad (1)$$

With that, we can remove the photobleaching effect of the $FRAP_{puncta}$. We had to normalize the data to compare multiple bleach experiments with different signal intensities. For that, we calculated the normalized signal at timepoint *t*, $NORM_t$ with the $FRAP_{puncta}$ signal at timepoint *t*, $FRAP_t$ the $FRAP_{puncta}$ signal at the timepoint of bleaching, $FRAP_{bleach}$ and the $FRAP_{puncta}$ signal before the bleaching, $FRAP_{prebleach}$. For $FRAP_{prebleach,}$ we used the average of the five signals we obtained before the bleaching.

$$\mathrm{NORM}_t = \frac{\mathrm{FRAP}_t - \mathrm{FRAP}_{\mathrm{bleach}}}{\mathrm{FRAP}_{\mathrm{preBleach}} - \mathrm{FRAP}_{\mathrm{bleach}}} \quad (2)$$

This normalized data we could fit on a function that describes the recovery of fluorescent signal, where A is the plateau of intensity, *t* corresponds to the time after photobleaching, and τ represents the bleaching rate, which needs to be fitted.

$$\mathrm{y(t)} = A \cdot (1 - e^{-\tau t}) \quad (3)$$

The recovery rate is the time it needs to reach half of the plateau recovery. It is calculated by the following calculation using the fitted τ.

$$t_{1/2} = \frac{\ln(0.5)}{-\tau} \quad (4)$$

2.7 Lipid transport assay

Caco-2 cells were transfected with MO, as described in chapter 2.4.2. Transfected cells were transferred into a 12-well falcon well plate with a semi-permeable inserted transwell PET membrane and a pore size of 1.0 µm (Falcon). After 12 h, the apical media was exchanged with DMEM supplemented with 20% (v/v) fetal bovine serum, and the basal chamber was filled with 1 mL DMEM supplemented with 10% (v/v) fetal bovine serum. The medium was changed every second day. After one week of differentiation, the medium in both chambers

was exchanged with FluoroBrite (Gibco) supplemented with L-Glutamin (Gibco). To stimulate lipid transport, an emulsion of 1 µM oleic acid (Carl Roth) and 2 mM Taurocholate (TCI) was heated to 37°C and stirred until the emulsion was clear. To this emulsion, 2.5 nM fluorescently labeled palmitic acid-NBD (Avanti) was added and transferred to the apical chamber of the transwells. After 24 hours, the apical and basal mediums were harvested, and the fluorescence was measured in the Line-Gene 9600 Plus Real-time PCR System with FAM detectors. Measurements were depicted quantified as the ratio between apical and basal fluorescence signals.

2.8 Statistical analysis

Statistical analysis done by the bioinformatic tools DEseq2 and rMATS was taken into account. All of our own statistical tests for experiments were performed in Python3 using Scipy (Virtanen et al. 2020) and Python.stat. First, the Shapiro–Wilk test tested each data population for normal distribution. When the null hypothesis was rejected ($p<0.05$), it was assumed that the population was not normally distributed, and the Mann–Whitney U test was performed to test statistical significance. When a normal distribution was available, Levene's test was performed to test the equality of variance. When the null hypothesis of equal variance was rejected ($p<0.05$), Welch's t-test was performed to test statistical significance. A Student´s t-test was performed to test statistical significance when the populations have equal variance and normal distribution.

3 Results

3.1 Alternative splicing of two exons in SEC31A

We focused on analyzing human tissue transcriptomic data to find alternative splice events that can regulate secretion. We worked with the human tissue transcriptomic dataset from the Science of Life laboratory in Stockholm, Sweden, with the Bioproject number PRJEB4337. The data set for our analyses was chosen for a sufficient read length of 200 bases paired-ends. Additionally, it has a good read coverage of samples per tissue (Fagerberg et al. 2014). DEseq was used to quantify the gene expression of the transcriptome data. A gene count value was calculated for each sample and each gene. These gene counts correspond to gene expression level and can be used to compare gene expression of single genes in different samples. A principal component analysis (PCA) benefits data sets with many samples. This way, we were able to show multi-dimensional samples as two-dimensional and compare each sample. When plotted, the similarity of each sample is shown by their distance from each other. In the PCA of the gene expression of the human tissues, it is clear that each tissue has

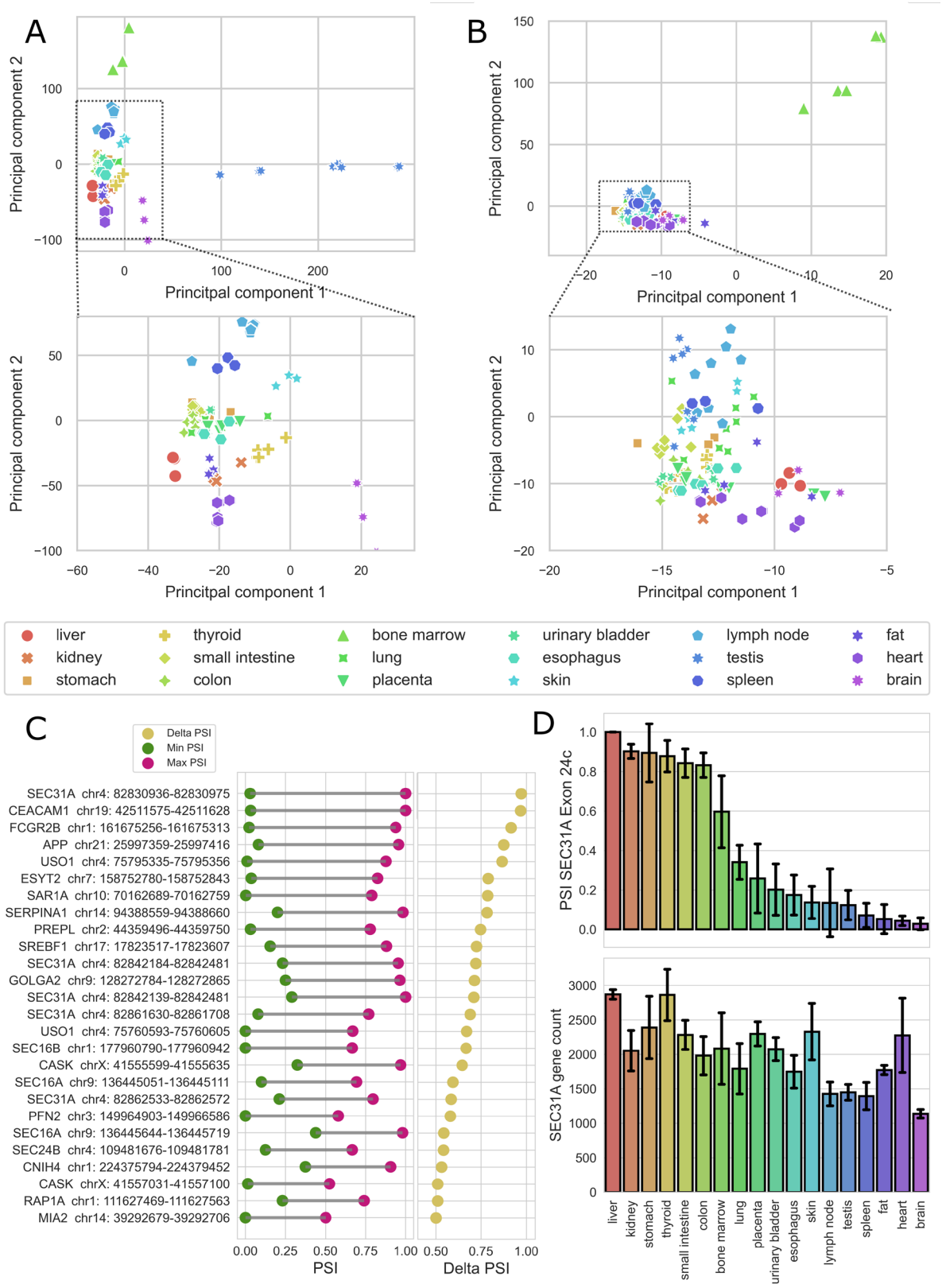

Figure 3-1 RNA-seq analysis of human tissue transcriptomic data. A) Global PCA plot of the gene expression of human tissues (top) and zoomed in plot to excluding bone marrow and lymph node samples for more detail (bottom). B) PCA plot of global splicing (top) and zoomed in version excluding bone marrow samples for more detail (bottom). C) Splice events in all secretory genes sorted by the delta PSI of the maximum and the minimum PSI value across all samples. The events were filtered by delta PSI > 0.5. D) Mean of the PSI values of the exon 24c in SEC31A (top) and the quantification of the gene expression by DEseq2 from SEC31A (bottom) as bar plot, with the standard deviation as error bars. (n > 2, n < 10).

a lot of similarities with the same or similar tissues (Figure 3-1A). Bone marrow and lymph node tissue samples are the most separated and distinct from the rest and are therefore excluded for a better view in a zoomed-in version of the PCA plot. There, the distinct clustering of each tissue becomes more apparent. Even subclusters are prevalent, as the colon, small intestine, and stomach samples also cluster together, exhibiting similar functions in the human body. Similarly, we analyzed alternative splicing using rMATS. There, the result is a PSI value for each quantified alternatively spliced event. This is the probability of the exon, alternative splice site, or intron to be spliced in. Observing the PCA of the splicing of the tissue samples, we saw a slightly different effect (Figure 3-1B). Bone marrow samples show in the splicing PCA a very distinct behavior, as well as in gene expression PCA. All other tissues show an equal behavior in the PCA. All samples seem to cluster more together and are closer to each other. Same-tissue samples still display similar localization but are also closer to other tissues. Similar tissues like the colon, small intestine, or stomach also show substantial similarity. This indicates that tissue-specific gene expression is more prevalent than splicing. However, splicing still shows a tissue dependency and gives reason to use this data set as the basis for tissue-specific splice analysis.

We are interested in splicing in secretory-related genes, so we filtered the splicing data with secretion-related GO terms. This list of secretory splice events must be sorted to identify the most compelling. To achieve this, we calculated the maximum and minimum PSI for each splice event, regardless of the sample. We calculated the delta PSI value with these extreme values sorted the splice events descending and filtered for a minimal delta PSI of 0.5 (Figure 3-1C). The result is a list of 26 secretion-related splice events, which are highly tissue-dependently spliced. This list highlights which aspect of secretion is potentially regulated by differences in alternative splicing in a tissue dependent manner. There are a few COPII complex members (SEC31A, SAR1A, SEC16B, SEC16A, SEC24B), cargo binder and adaptors (APP, USO1, ESYT2, CNIH4), and Golgi compartments (GOLGA2, PREPL). We also found the alternative 5’ splice site of exon 24 in SEC31A and the cassette exon in SEC16A, which were already characterized (Wilhelmi et al. 2016; Neumann et al. 2019). The most tissue-dependent spliced event is located in SEC31A, between exon 24 and 25, with the coordinates chr4:82830936-82830975 (hg38). This finding is striking, as SEC31A is an important part of the COPII complex and is known to regulate secretion (Barlowe et al. 1994; Antonny et al. 2001; Bi et al. 2007; Stagg et al. 2008). Also, there is already a splice event in SEC31A known, as mentioned previously, that can regulate the efflux of cargo (Neumann et al. 2019). While that alternative 5’ splice site of exon 24 was named exon 24b, we chose to name that new exon 24c as it is not well characterized and not annotated in GENCODE V44.

The splicing pattern of exon 24c is highly tissue-dependent (Figure 3-1D, top). In the liver with

a PSI of 1, exon 24c is always included. Also, the kidney, stomach, thyroid, small intestine, and colon tissue show high PSIs (0.8 to 0.9). It is noticeable that three of these tissues with high exon inclusion are gastrointestinal. The tissues of the esophagus, skin, lymph node, testis, spleen, fat, heart, and brain show all inclusion levels of under 0.2. These levels of high change of inclusion of exon 24c in these tissues are even more remarkable compared to the gene expression of SEC31A (Figure 3-1D, bottom). While it is clear that gene expression is also slightly tissue-dependent, it is succeeded by the extent of how much tissue dependence on splicing for exon 24c is far larger. These findings imply that tissue-dependent splicing of exon 24c in SEC31A may have a regulatory function in secretion involving gastrointestinal tissues.

To validate that this exon exists and is tissue-dependent spliced, we performed an RNA extraction of healthy wild-type mice organs and a radioactive splice-sensitive RT-PCR (Figure 3-2A). The gel and the quantification show that the liver and kidney, similar to the human tissue RNA-seq data, show high inclusion of exon 24c. Plotting the values of the human RNA-seq data and the mouse organ PCR together displays a high level of conservation (Figure 3-2B). Not only can we show that the prediction of exon 24c holds up, but also that it is evolutionary and highly conserved and adds to the emphasis that it does have a regulatory function.

The previously mentioned alternative 5' splice site 24b in SEC31A was partially described in our previous work (Neumann et al. 2019). This work discovered the exon in a genome-wide

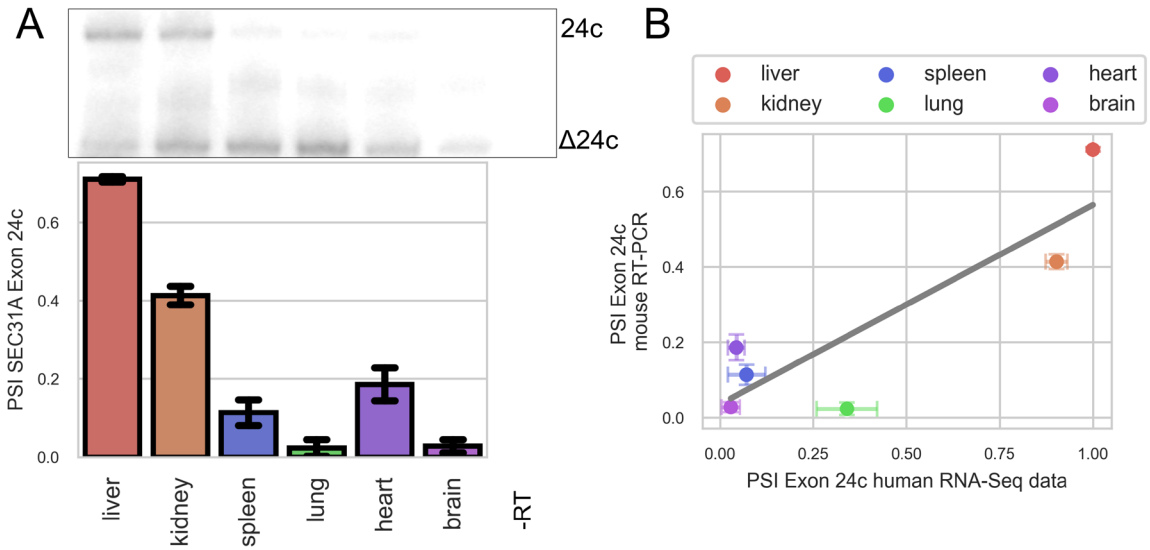

Figure 3-2: Wet lab validatoin of SEC31A exon 24c. A) Radioactive gel of the splice sensitive RT-PCR observing the region of SEC31A between exon 24 to 25 of mouse tissues, and the quantification. Depicted are the mean PSI as bar plot and the standart deviation as error bars (n = 3). B) Linear regression of the mice tissue PCR validation and the human tissue RNA-seq analysis. Depicted are the mean PSI of the of the gel quantification and standart deviation as y-error bars (n = 3) and the mean PSI of the RNA-seq quantification and its standart deviaiton as x-error bars (n > 2, n < 10).

search for alternative splicing that could regulate secretion. The splicing switch of exon 24b showed a regulatory effect on the secretion in the RUSH assay. When exon 24b is not included, the secretion of GPI reporter is slowed so that 20% less GPI is transported to the outer cell membrane after one hour. Investigating the tissue splice profile of exon 24b, we observe that it is similarly strong tissue-dependent spliced as exon 24c (Figure 3-3A). Bone marrow, is the tissue with the highest inclusion of exon 24b with a PSI of nearly 1, whereas the placenta shows the lowest inclusion with around 0.25. All in all, exon 24b is greatly tissue-

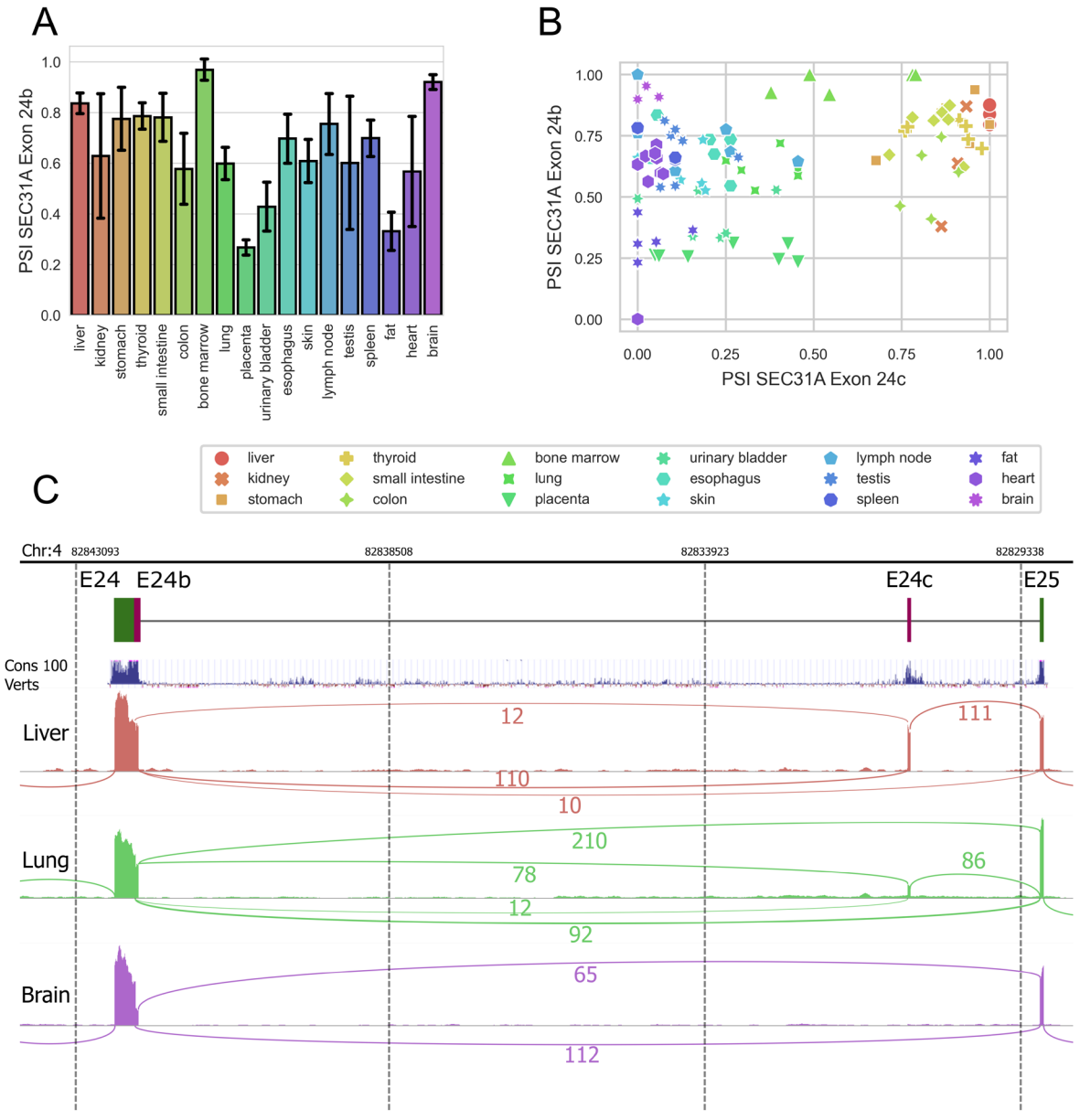

Figure 3-3: Tissue dependent RNAseq analysis introducing exon 24b in SEC31A. A) Mean PSI values of exon 24b in SEC31A in human tissue samples depicted as bar plot and the standart deviation as error bars (n > 2, n < 10). B) Scatterplot of PSI values of exon 24c against exon 24b in SEC31A in human tissues. C) Sashimi plots of SEC31A between exon 24 and exon 25 in different human tissue samples, with the conservation of 100 vertabrates taken from the USCS genome browser.

dependent alternatively spliced. Comparing the splice profile of exon 24b and exon 24c, it is clear that there is no direct linear relation between these two (Figure 3-3B). Less than half of the tissue samples include both exons, approximately half of them include mostly exon 24b but not 24c, around 10% include none of them, and only three samples include mostly exon 24c but not exon 24b. This shows that these two exons, regardless of their proximity to each other in the genome, are not co-spliced or mutually exclusively spliced and act relatively independently of each other. Exon 24b and 24c are, when included, concatenated in mRNA On a genomic level a 10,000 nucleotides long intron lies between them (Figure 3-3C). The conservation in vertebrates is most striking, as exon 24b is highly conserved, but exon 24c shows lower conservation. In contrast, the intronic region downstream of exon 24c shows higher than usual levels of conservation, which could point to the possibility of consensus sequences for RNA binding proteins (Figure 3-3C). Sashimi blots, like the one in Figure 9C, are a powerful tool for conveying information from raw data in a genomic context. Our samples show the individual inclusion of exons 24b and 24c (Figure 3-3C). There are junction reads from exon 24b skipping to exon 25 and exon 24c, as well as from exon 24b inclusion to exon 24c and exon 25b. These junction reads are also dynamically changed in different tissue samples.

3.1.1 Correlation analysis of the alternative exons reveals possible function

When dealing with new uncharacterized exons, the difficult question is, what, if any, molecular function does the alternative splicing have? One method is to calculate the correlation between the inclusion of the exon and the gene expression of all quantified genes. Correlation analysis is a powerful tool, but it is dangerous to misinterpret correlation as causation. What correlation shows instead is a quantification of similarity in the behavior of two different data pools. The context is essential to give the correlation analysis its validity. In our case, we compared the inclusion levels of the exons of SEC31A and the gene expression of all genes quantified in DEseq. Because we could not tell which distribution the expression of the genes and the inclusion of the exons have, we calculated both the Spearman and the Pearson correlation indices. The results are lists containing one Pearson and one Spearman correlation index for each gene, describing the correlation to the exons. A high correlation means this gene has similar gene expression levels compared to the exon's inclusion levels. The correlation also works when comparing genes, ranging from 0 to 10,000 gene count and exons ranging from 0 to 1 PSI. When plotting the Pearson and Spearman correlation, interesting genes that correlate strongly positively or negatively lie in the upper right or the lower left corner, respectively. To make the analysis more accessible, we colored specific groups of genes, the GO terms, to find valuable connections and genes. In general, exon 24b shows a lot of genes that correlate strongly negatively with the exon and fewer that correlate

positively with it (Figure 3-4A). Observing the gene groups of secretion and mRNA splicing, no genes are sticking out or are otherwise distinct to be of interest. We also analyzed specifically the genes related to chylomicrons and collagens for a possible secretory cargo

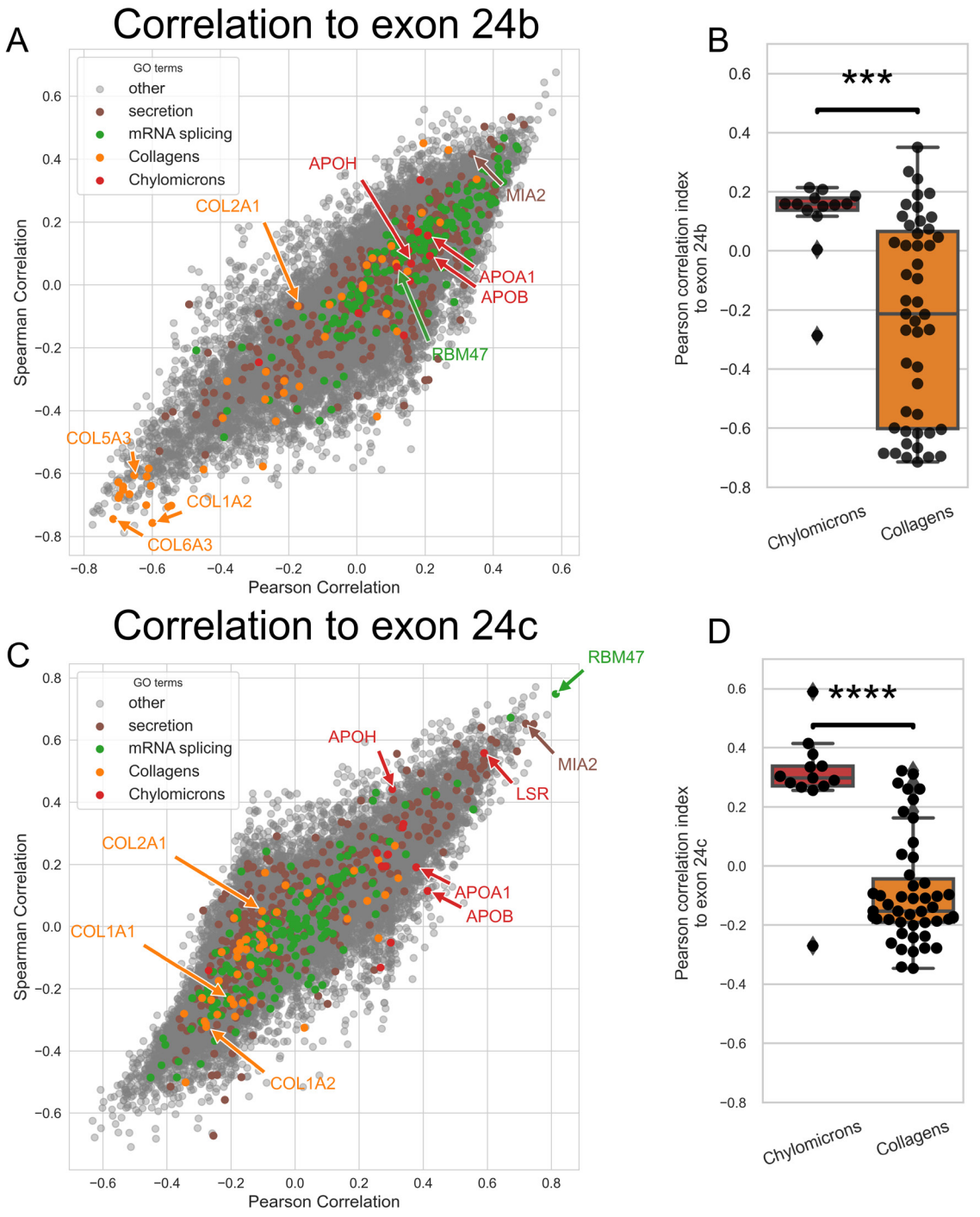

Figure 3-4: Correlation analysis with exon 24b and exon 24c. A) Spearman and Pearson correlation indices between exon 24b inclusion and the gene expression of all quantified genes. The colors highlight specific GP terms and in the plot some genes of interest are labelled. B) Pearson indices of genes related to apolipoporteins and collagens to exon 24b as box-whisker-plots. C) Correlation indices to exon 24c inclusion. Pearson indices to exon 24c. In the box-whisker-plots, the box shows the first and the third quartile, the line inside is the median and the whiskers depict the interquartile range. The data was not normally distributed, so the Mann-Whitney-U test was performed and significance is shown by asterisks: p**** < 0.0001.

selection. Whereas the group of chylomicrons showed an average Pearson index of around 0.18, the indices of the collagens are much more heterogeneous and have a significantly lower average correlation index (Figure 3-4B). The correlation analysis plot shows a cluster of collagen genes in the lower left corner. This means that the inclusion of exon 24b correlates negatively with the gene expression of a specific collagen group. In other words, collagen expression is enhanced when also exon 24b is excluded. Since exon 24b is located in a gene that codes for a secretion-regulating gene, we can assume that exon 24b exclusion could regulate collagen secretion.

For exon 24c, we found a slightly different picture in the correlation analysis (Figure 3-4C). Firstly, the Correlation analysis shows that more genes have a higher positive correlation with the inclusion of exon 24c compared to exon 24b. RBM47 is the gene with the strongest positive correlation. This splicing factor regulates alternative splicing in EMT in healthy and cancer cells (Rokavec et al. 2017; Kim et al. 2019). It is promising to find a splicing factor in the correlation analysis, as it is possible that RBM47 could regulate the splicing of exon 24c. Another promising gene that is highly correlated with exon 24c is melanoma inhibitory activity 2 (*MIA2)*. The *MIA2* gene codes for the cTAGE5 protein, which is substantial for the secretion of big cargo (Santos et al. 2016). When we assume that exon 24c is responsible for regulating the secretion of specific cargo, we can estimate with the correlation with *MIA2* that exon 24c probably regulates the secretion of big cargo. The correlation assay can also estimate this kind of cargo. Observing the collagen and the chylomicron genes again, we see that the correlation with the chylomicron genes is higher overall than the correlation with collagen genes (Figure 3-4D). That could mean that exon 24c is responsible for regulating the secretion of chylomicrons.

3.2 Different isoforms of SEC31A regulate the secretion of big cargo

3.2.1 Exon 24b splicing is helping collagen secretion, specifically

As our correlation indicates, skipping exon 24b could eventually lead to increased collagen secretion. To further analyze this, we established a method to modulate the inclusion of exon 24b and measure collagen secretion.

To modulate splicing in our model cells Hek and HeLa, we used dCasRx, a catalytic dead Cas13 that can bind to pre-mRNA (Konermann et al. 2018). With guide RNA, which is complementary to an alternative exon, and its size is usually enough to sterically block the spliceosome and inhibit inclusion (Figure 3-5A). Exon 24c is not included in Hek or HeLa cells in a substantial matter to be accounted for in this experiment. With the introduction of the dCasRx with guide RNA targeting exon 24b, we can see a robust 4-fold reduction of the inclusion of that exon (Figure 3-5B). To strengthen our data, we repeated the RUSH assay already done in our lab (Neumann et al. 2019) with the GPI reporter and added an E-cadherin reporter (Figure 3-5C-D). Cadherins are calcium-dependent cell adhesion proteins, and their

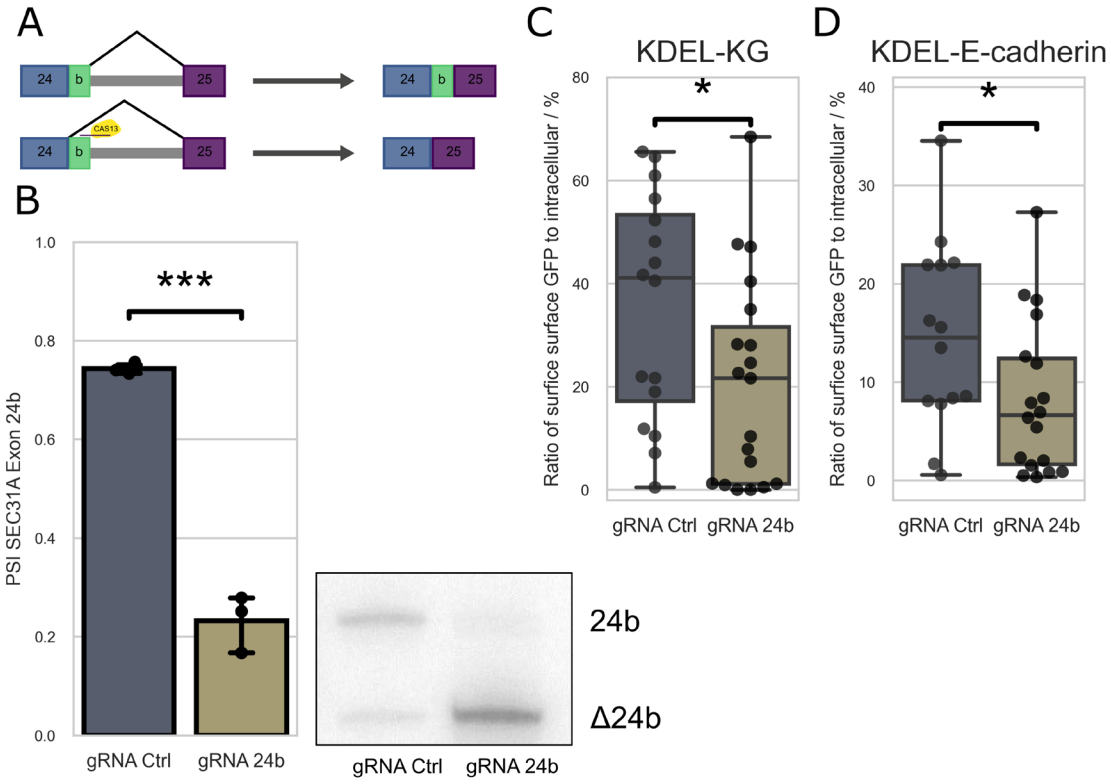

Figure 3-5 RUSH assay with modulation of SEC31A exon 24b splicing. (A) Schematic overview of the molecular mechanism of the dCasRx to repress exon 24b inclusion. B) Validation of modulation of exon 24b inclusion using dCasRx. The mean PSI was depicted as a bar plot, and the error bars show the standard deviation. C) Quantification of the RUSH assay using the Str-KDEL_SBP-EGFP-GPI plasmid, or the D) the Str-KDEL_SBP-EGFP-Ecadherin plasmid. The line in the box-whisker plot shows the median, the box the first and the third quartile, and the whiskers the interquartile range. Statistical significance was verified using the Students-T-test (B) or Mann-Whitney U test (C and D) and showed by asterisks: *p < 0.05. ***p < 0.001

transport to the cell membrane is tightly regulated (Liu et al. 2017a). The RUSH assay with the GPI reporter and the E-cadherin reporter showed a similar reduction in secretion under Cas13 treatment introducing exon 24b gRNA. General cargo secretion directed to the outer

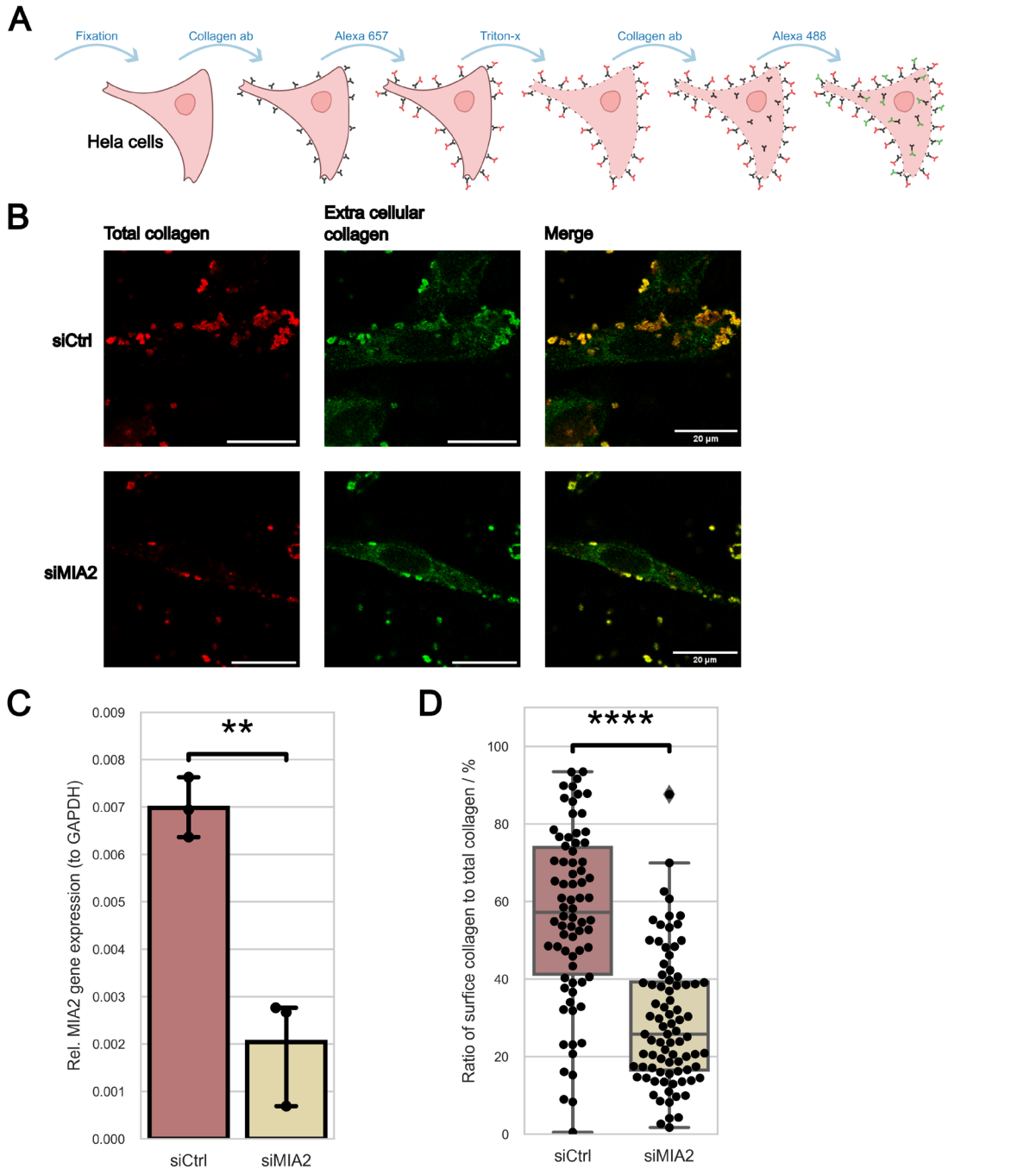

Figure 3-6:Establishing the double collagen staining in HeLa cells to quantify collagen secretion. A) Schematic overview of surface collagen staining. B) Exemplary images of surface collagen staining in HeLa cells treated with siRNA targeting MIA2. Depicted is the extra cellular collagen red, the total collagen green. C) Validation of the siRNA targeting the MIA2 gene in HeLa cells using RT-qPCR. The mean gene expression relative to the GAPDH is shown as bar plot, and the standard deviation is shown as error bars (n = 3) D). Quantification of surface collagen staining as surface collagen to total collagen ratio as box-whisker plot. The box edges depict the first and the third quartile, the line inside the box is the median and the whiskers show the interquartile range. Outlier are shown separately (n ~ 70). Statistical significance was validated using Students-T-test (C) or Mann-Whitney U test (D) and showed by asterisks: p** < 0.01, p**** < 0.0001.

cell membrane is reduced after the inhibition of exon 24b in SEC31A.

To examine if splicing could change the secretion of big cargo, we established an assay to detect the secretion of extracellular collagen. This method will stain the extracellular collagen using a collagen-specific primary antibody and an Alexa fluor647 secondary antibody. After permeabilization of the cells, another round of primary antibodies with an Alexa fluor488 secondary antibody is performed (Figure 3-6A). The result is a staining of extracellular collagen with Alexa fluor647, and a total collagen staining with Alexa fluor488. These two fluorescent tags can be measured separately. The ratio between the green signal colocalized with the red signal against only the green signal quantifies collagen secretion. We used a siRNA against MIA2 (Figure 3-6B). Notably, cells treated with MIA2 siRNA show more signal of intracellular collagen; in contrast, the scrambled Ctrl siRNA exhibits more extracellular collagen. After confirming the reduction of MIA2 transcripts using RT-qPCR (Figure 3-6C), we also quantified the images and found a matching apparent decrease in surface collagen (Figure 3-7D). This led us to the conclusion that this assay is suited to quantify collagen secretion after splicing modulation.

Combining the dCasRx method to modify splicing and the collagen staining to quantify collagen secretion, we analyzed collagen secretion with impaired exon 24b inclusion. We used an HA-tag antibody to stain the Ha-tagged dCasRx to ensure the quantification of only

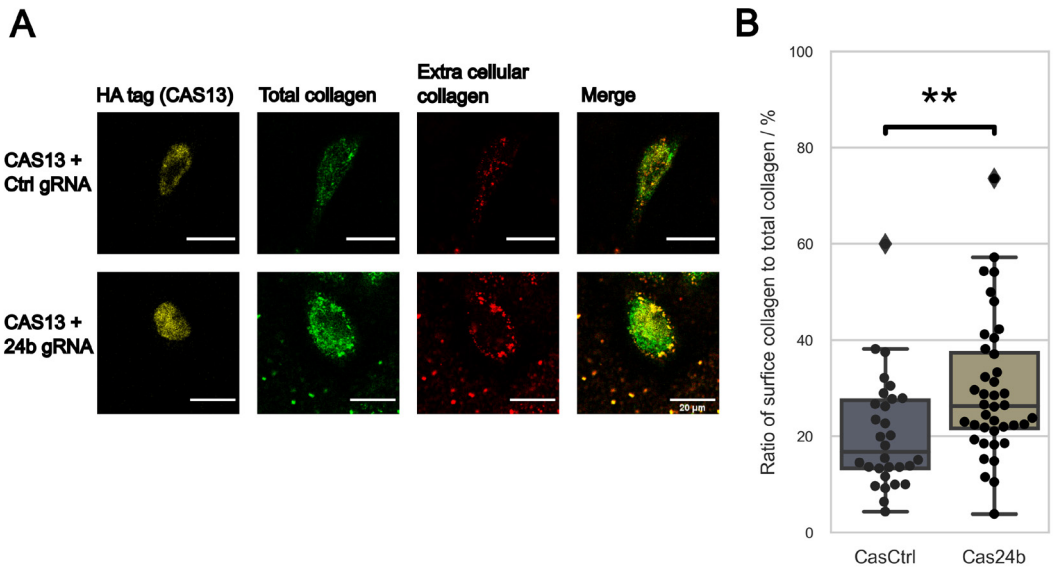

Figure 3-7: Surface Collagen assay with dCasRx targeting exon 24b in SEC31A. A) Exemplary images of HeLa cells with surface collagen staining. The HA-tag staining for the dCasRx was depicted in yellow, the total collagen in green, and the extracellular collagen in red B). Quantifying surface collagen staining as the ratio between surface and total collagen with dCasRx treatment against SEC31A exon 24b. The box in the box-whisker plot shows the first and the third quartile, the line inside is the median, and the whiskers show the interquartile range. Outliers are shown separately (n ~ 40). Because the data was not normally distributed, statistical significance was validated using the Mann-Whitney U test and showed by asterisks: p** < 0.01.

Cas-positive cells. Only in cells where we could detect the dCasRx in the nucleus did we quantify the collagen secretion (Figure 3-7A). We removed the HA channel in the quantification. Secretion of collagen with inhibited exon 24b inclusion is slightly but significantly increased. We observed around 10% more surface collagen in the samples with dCasRx plus exon 24b guide RNA than in the samples with the Ctrl guide RNA (Figure 3-7B). That means the alternative splicing of exon 24b, selectively regulates the secretion of specific cargo and can enhance collagen secretion when not included.

3.2.2 Lipid transport is regulated by exon 24c

Observations in the correlation assay done with exon 24c in SEC31A lead to the assumption that the alternative splicing of that exon could lead to the regulation of lipid transport (Figure 3-4D). The high correlation between exon 24c and MIA2 strengthens this hypothesis (Figure 3-4C, Figure 3-8A). Evidently, human tissue samples that highly express MIA2 also have a high inclusion of exon 24c.

To explore whether alternative splicing of exon 24c regulates the secretion of lipids, we performed a lipid transport assay in Caco-2 cells. Caco-2 cells are human intestinal cells from a colon adenocarcinoma isolated in 1977 (Fogh et al. 1977). These cells behave under normal conditions as cancer cells; when placed on a semi-permeable membrane, the cells begin to spontaneously differentiate into polarized endothelial-like cells (Natoli et al. 2012). This effect is successfully used as a model to simulate drug absorption through the intestines (Hubatsch et al. 2007), food availability (Glahn 2022), or directed lipid transport (Nauli and Whittimore 2015). We can use Caco-2 cells to measure lipid transport after modulating the splicing of exon 24c. For that, we use electroporation to transfect MO, reducing the inclusion of exon 24c. These cells are then placed in a trans-well, a semi-permeable barrier, where the cells can spontaneously differentiate. After one week of differentiation, we fed the cells with a medium supplemented with fluorescent-labeled palmitic acid (Figure 3-8B). These fluorescently labeled lipid acids can be measured, and the basal to apical signal can be put in ratio. We used splice site blocking MO to tune the splicing of exon 24c. These are modified oligonucleotides designed to hybridize to the splice site of the target exons in pre-mRNA and prohibit successful splicing (Figure 3-8C). To confirm if the MO is stable enough to block splicing after one week of differentiation, we validated the splicing of exon 24c in transfected and differentiated Caco-2 cells (Figure 3-8D). To our surprise, the inclusion of exon 24c drastically increases after one week of differentiation in Caco-2 cells. Just one week of differentiation is sufficient to increase the inclusion of exon 24c from 15% to almost 60%. The exon 24c inclusion could successfully be reduced by the MO, even in one-week differentiated Caco-2 cells. With this setup, we were able to investigate if the alternative splicing of exon

24c can regulate the secretion of lipids (Figure 3-8E). We observed a strong basal to apical ratio in fluorescence signal reduction when we treated the Caco-2 cell with the exon 24c splice site blocking MO compared to the control MO. Exon 24c in SEC31A, therefore, seems to be an additional regulator for lipid secretion and connects alternative splicing with the regulation of lipid secretion.

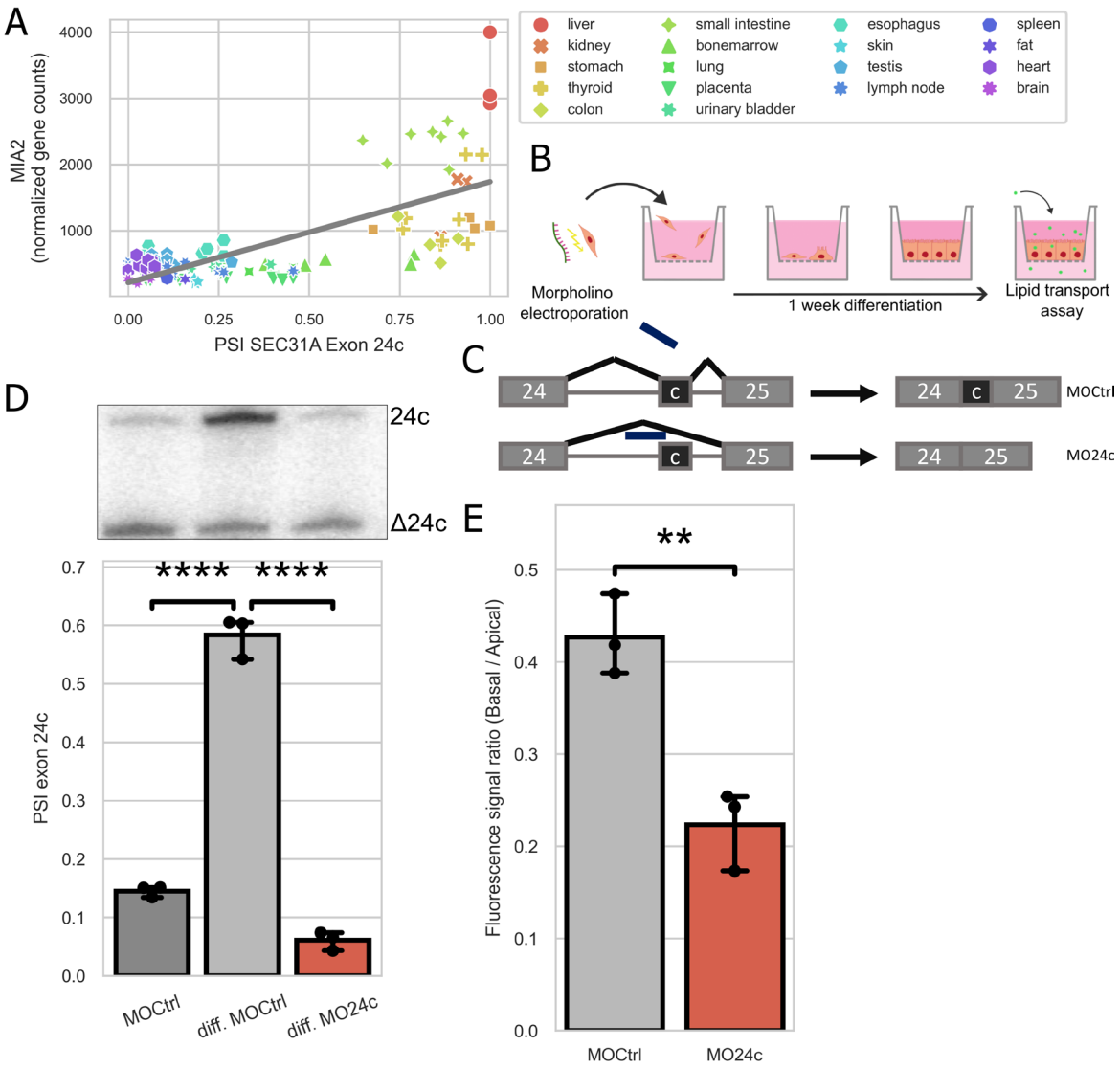

Figure 3-8: Lipid transport assay with modulated exon 24c inclusion. A) Linear regression with exon 24c inclusion and gene expression of MIA2. B) Schematic description of Lipid transport assay with Caco-2 cells. Caco-2 cells were electroporated with MO and seeded on a trans-well. After 1 week of differentiation with unsymmetric FBS concentration, fluorescently labelled lipids are fed to the cells. Fluorescent signal is measured in apical and basal media. C) Schematic of the molecular mechanistic of splice site blocking using MO. D) Validation of exon 24c in differentiated Caco-2 cells after MO transfection. Shown are the mean PSI as bar plot and the error bars depict the standard deviation (n = 3). E) Quantification of lipid-transport assay using the ratio of the fluorescent signal of the basal to apical medium. Depicted are the mean ratios as bar plot and the error bars show the standard deviation (n = 3) Statistical significance was validated using Students-t-test and showed by asterisks: $p^{**} < 0.01$, $p^{****} < 0.0001$.

3.3 Splicing switches regulate structural and functional properties of SEC31A

3.3.1 Stationary phase is reduced without exon 24b in FRAP experiments

We performed FRAP with modulated exon inclusion to get more insight into the molecular function of exon 24b. We were generously helped by Francesca Bottanelli at the Freie Universität Berlin, and her work group for sharing the SEC13-GFP tagged HeLa cell line, with which we could perform live-cell imaging and FRAP. SEC13 and SEC31 compose the outer cage of the COPII complex (Stagg et al. 2006). With SEC13-GFP, we have a direct marker for the COPII complex and an indirect marker for SEC31A. We hypothesize that splicing of SEC31A exon 24b influences the behavior of the COPII complex. In the FRAP assay, we chose stable COPII puncta, which we could observe for at least a minute in the same space. After two frames of bleaching with a 60% powered laser, we could detect a recovery of fluorescent signal from the puncta (Figure 3-9A). We can estimate differences in the two treatments by plotting the signals from the bleached puncta (Figure 3-9B). We observe that COPII puncta without exon 24b in SEC31A shows a lower recovery. To properly quantify the observation, we can calculate two properties of the recovery. The recovery rate is defined as the time at which the signal reaches the half maximum of recovery. Besides high variation, the two isoforms of SEC31A do not show significant differences (Figure 3-9C). That means the mobility of SEC13 inside COPII is not changed by the alternative splicing of SE31A exon 24b. The mobile fraction is defined as the maximum signal that the puncta approaches reach after bleaching. It shows which fraction of SEC13 in the COPII can move and lead to the recovery of the signal. We observe a reduction of around 20% mobile fraction in COPII when we inhibit the inclusion of exon 24b in SEC31A (Figure 3-9D). We validated the reduction of inclusion of exon 24c after MO treatment using splice-sensitive RT-PCR (Figure 3-9E).

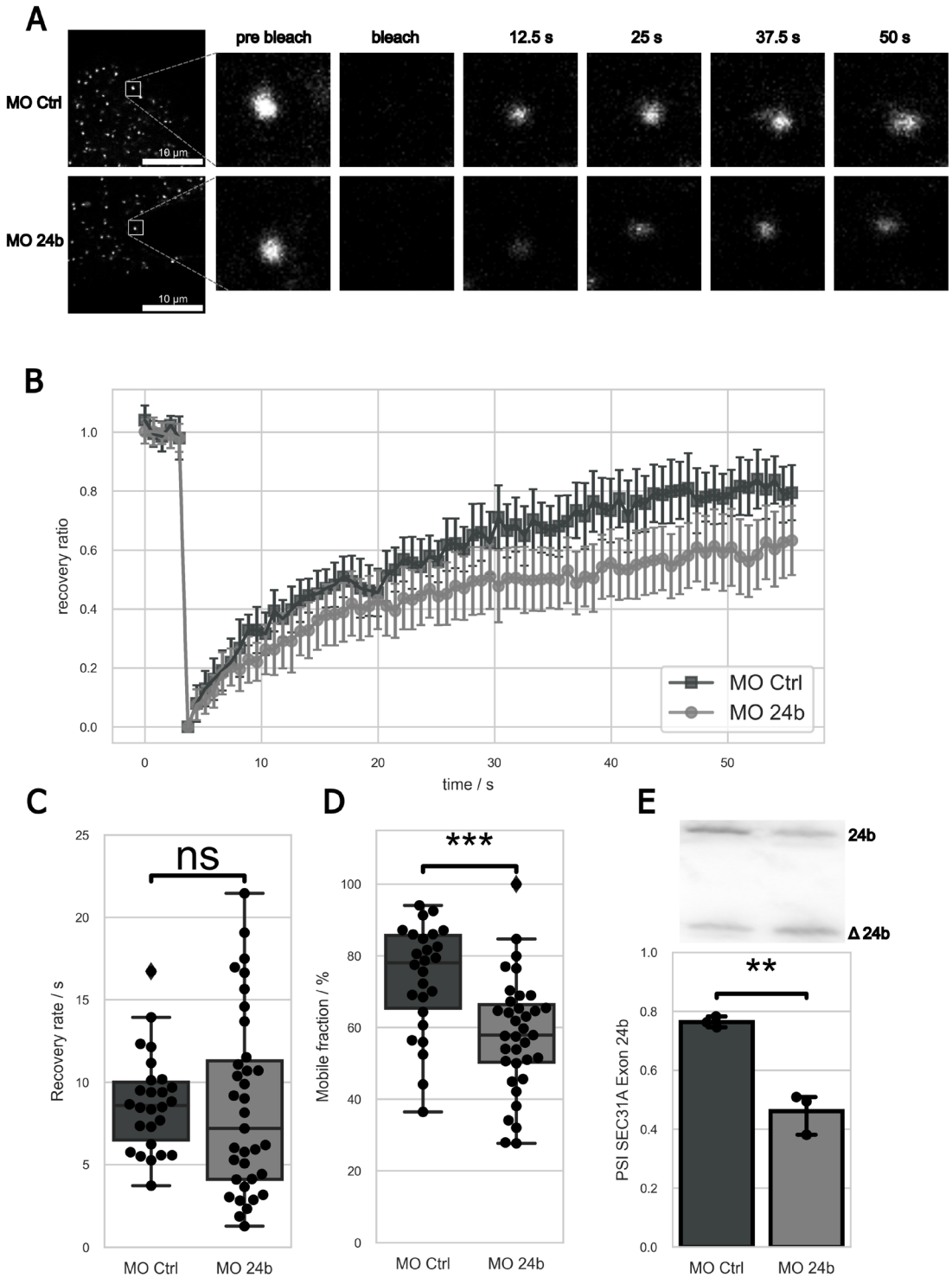

Figure 3-9: FRAP analysis on SEC13 puncta with exon 24b splicing. A) Representative array of pictures of the FRAP experiment from cells treated with MO against exon 24b and Ctrl MO. B) Relative signal recovery against time pre and post bleach from puncta with and without exon 24b. The dots represent the mean and the error bars the standard deviation at each time point (n = 30). C) Recovery of the FRAP as box-whisker plot D) The mobile fraction as plateau of the function fitted on the data from the FRAP as box-whisker plot. The box represents the first and third quartiles, the line inside the box is the median and the whiskers represent the interquartile range. E) Validation of the splice reduction by the MO. The bar of the bar plot represents the mean and the error bars the standard deviation (n = 3). The data in C) was not normally distributed, so the Mann-Whitney U test was performed. In D) and E) the data was normally distributed so the Student's-t-test was performed and significance was depicted with asterisks: $p > 0.05$ = ns, $p^{**} < 0.01$, $P^{***} < 0.001$

3.3.2 Alphafold prediction offers structural insight in exon 24c switch

In the previous chapters, we established the effect alternative splicing of SEC31A exon 24c has on lipid transport. But how does a small exon switch change the behavior of chylomicron secretion? The first detail to be examined is whether the exon codes for an amino acid sequence or if it has a regulatory function; for example, does it lead to NMD? If the inclusion of exon 24c leads to NMD, we will suspect that the splicing of exon 24c does influence the gene expression of SEC31A. We did not find evidence for that in the human tissue data (Figure 7D). Additionally, exon 24c is 39 nucleotides long, which means that it does not lead to a frameshift, and it also does not code for a premature stop codon. The 13 amino acids that the 39 nucleotides of exon 24c code for match a UniProt isoform with the identifier O94979-8 and the name isoform 8. This isoform is identical to isoform 1, except it has 13 additional amino acids at position 989. We, therefore, can extrapolate that the exon 24c in SEC31A is a coding exon. We then investigated how much the exon is conserved on the translational level. We compared the amino acid sequences of different higher animals in the region of exon 24c in SEC31A homologs and calculated the homology (Figure 3-10A). It is striking that the WND motive is highly conserved across all the compared organisms. For context, SEC23 is engaged as a GAP for SAR1, thus increasing its GTPase activity. SEC31 is a supportive GAP. This WND motive in SEC31A inserts itself into the catalytic site of SAR1 and interacts with SEC23. The WND motive helps the His79 of SAR1 to orient to the catalytic site and shield it from solvents. This mechanism increases the GTPase activity of SAR1 even more (Bi et al. 2002; Bi et al. 2007). So, it is not surprising that this motive is heavily conserved in most organisms that we compared. We detect slightly less conservation by inspecting the peptides for which exon 24c is coding. In particular, the conservation is relatively low for organisms not closely related to humans. However, organisms that are closely related to humans show higher conservation. That leads to the assumption that the peptides for which exon 24c is coding have a relatively conserved function.

Bi et al. describe the mechanism in a solved structure in which SEC23 and SE31 can enhance the GTPase activity in SAR1 (Bi et al. 2007). To achieve this, they solved the structure of yeast Sec23 and a truncated version of yeast Sar1 together with a 40 amino acid-long polypeptide of yeast SEC31A containing the WND motive. To enlighten our findings on a molecular level, we performed an AlphaFold prediction for the human form of SEC23A, SAR1A, and an active fragment of SEC31A spanning over the WND motive and containing the peptides of exon 24c (Figure 3-10B). The Local Distance Difference Test (lDDT) is performed to estimate a model's prediction quality and reliability. The lDDT validated the stereochemical plausibility of a predicted model for each residue (Mariani et al. 2013). AlphaFold calculates the plDDT for each prediction and ranks them according to the mean

pIDDT. The active fragments of the pIDDT from the SAR1A, SEC23A, and SEC31A are generally high, between a pIDDT of 90 and 100 almost everywhere (Figure 3-10C). SEC23A and SAR1A are highly structured proteins, except for the unstructured N-terminus in SAR1A and a larger flexible linker in SEC23A. The active fragment, however, shows a very low

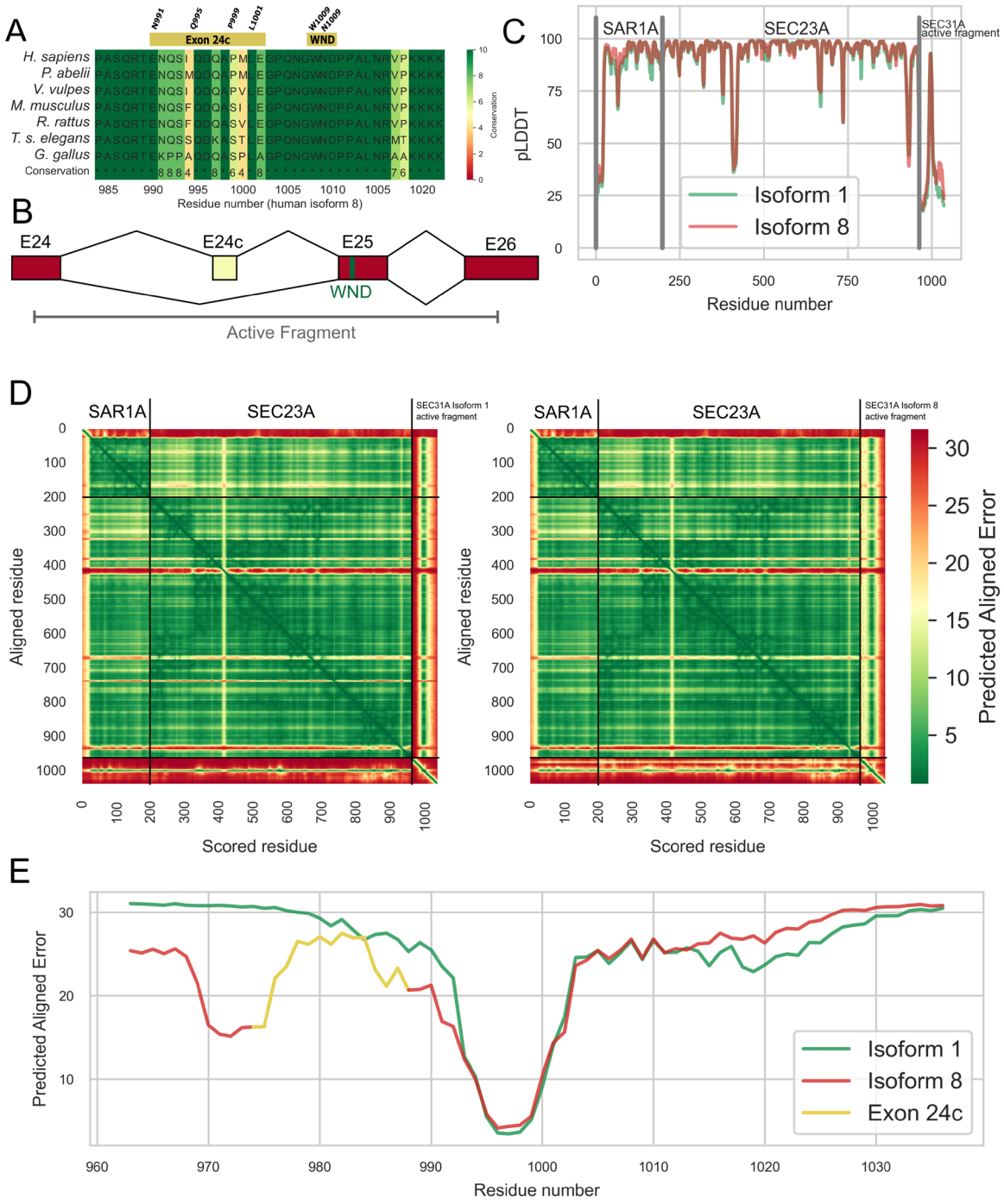

Figure 3-10: Estimating exon 24c function on protein level. A) Conservation of the region around the exon 24c coding region and the WND motive in SEC31A of multiple homologs. B) Schemata of active fragment used for the AlphaFold prediction and its location on the genome C) The pIDDT plot of the AlphaFold prediction of SEC23A, SAR1A and the active fragment of SEC31A isoform 1 and isoform 8. D) Two dimensional Predicted alignmend error plot of the multimeric complex including the active fragment of SEC31A isoform 1 (left) and isoform 8 (right). E) One dimensional plot showing the predicted alignment error of the complex and the active fragment isoforms at the scored residue 800 between the aligned residue 964 and 1036.

plDDT. This was expected, as this area is highly unorganized and part of the bigger disordered C-terminus of SEC31A. One high confidence peak lies in the middle of the active fragment and corresponds to the WND fragment. The interaction of the WND motive with SEC23A and SAR1A stabilizes these amino acids. This in turn enhances the confidence of folding in AphaFold. Isoform 1 and 8 of the active fragments only show minuscule differences in confidence.

Another indicator for the model prediction is the pae. This index calculates the error in the angstrom of the aligned residue for each scored residue. It shows, how confident AlphaFold is about the relative position of that residue about all other residues, regardless of the rate of plDDT. This creates a graph where a diagonal line shows 0 angstroms of error for each residue to itself. Each area that expresses low pae usually corresponds to compact domains or interactions between residues and residues/domain/protein. The page plot of SEC23A, SAR1A, and an active fragment of SEC31A shows many residues that have low error values to each other (Figure 3-10D). SEC23A and SAR1A are known to interact closely with each other. Together with the fact that they are highly organized proteins, it is not surprising that most residues in SEC23A and SAR1A have very few errors from each other. AlphaFold is highly confident in the relative position to each other of these proteins. Examining the active fragment of SEC31A, we see a similar effect to that of the plDDT. The highly disorganized active fragment has very little interaction with SEC23A or SAR1. Thus, it does show high pae. Of course, the WND motif is apparent as a peak in pae. The interaction with SAR1A and SEC23A does give AlphaFold a high confidence in the relative position of the WND motive. Additionally, the plot shows a slight difference between the active fragments of SEC31A isoform1 and isoform 8. Whereas isoform 1 merely shows a low level of pae at the area of the WND motif, isoform 8 shows an additional area of high confidence of relative position a few amino acids upstream of the active fragment. To get a closer look, we decided to filter the pae plot for the residues between 963 and 1030 aligned residue and do a cross section at the 800th scored residue (Figure 3-10E). Both isoforms show a reduced pae at the WND motive, but only isoform 8 has an area upstream that also shows a slightly reduced pae. The amino acids coded by exon 24c are near the drop in pae.

The plDDT and the pae plots both show high confidence in prediction and even show that AlphFold correctly recognized the WND motif in SEC31A and possibly inserted it in the correct position between SAR1A and SEC23A in the catalytical region. To validate the prediction of AlphaFold, we aligned the predicted multimeric structure from human SEC23A, SAR1A, and the active fragment of SEC31A with the solved structure of yeast (Bi et al. 2007) (Figure 3-11A). Overall, the predicted structure of the human multimer prediction and the yeast structure are highly similar. In both models, the tryptophan 1008 in the active fragment in

SEC31A interacts with the histidine 79 of SAR1A and directs it to the active fragment. Even though the asparagine 1009 in SEC31A has the same position in the prediction and the solved structure, the glutamine 714 in SEC23A has a 90° flip but is still in the correct position. With that, we can confidently say that the AlphaFold prediction does show a similar model of the interaction site of SAR1A, SEC23A, and the active fragment of SE31A with yeast protein. We

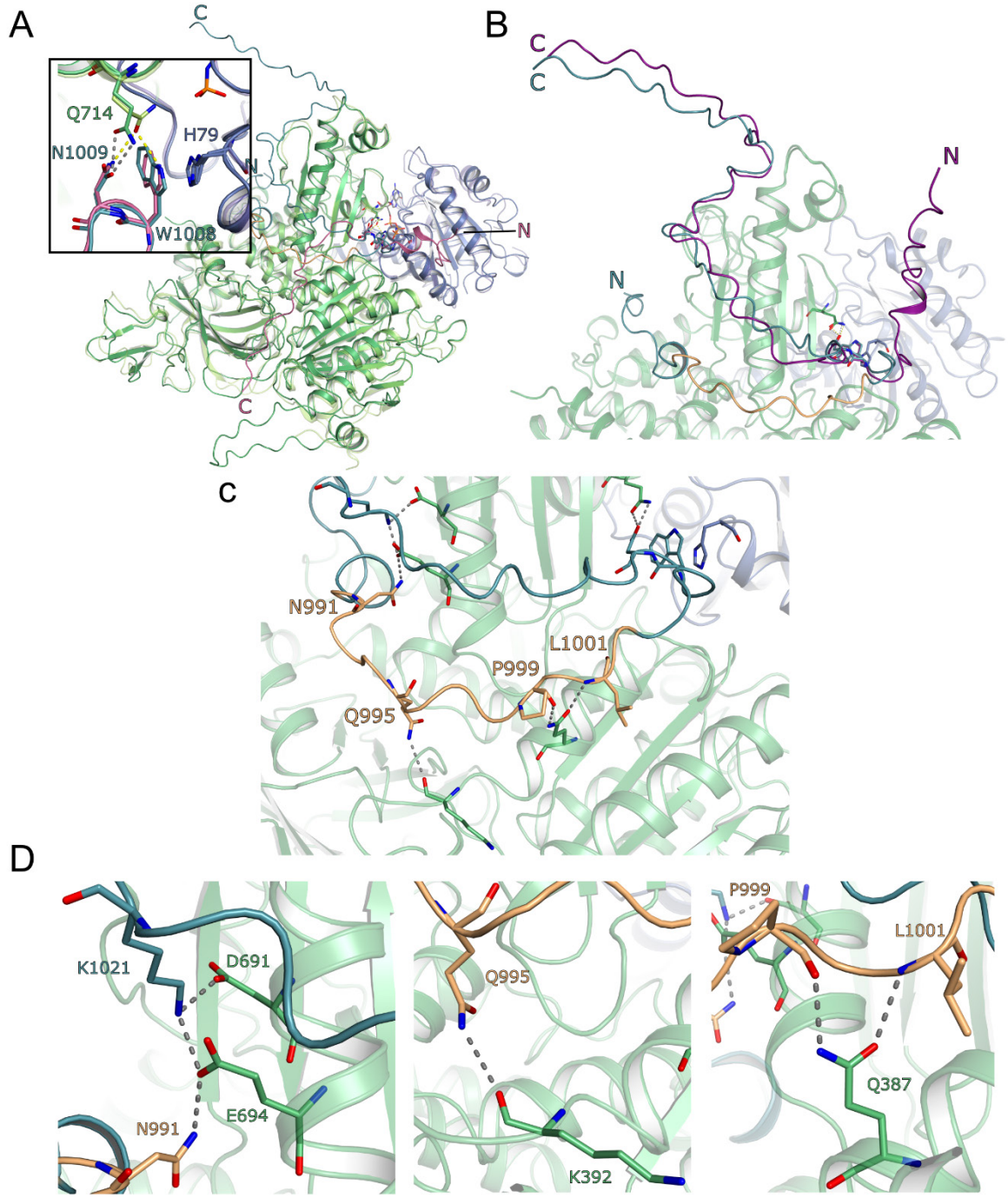

Figure 3-11: Structures of AlphaFold prediction of the multimer SEC23A, SAR1A and SEC31A isoform 1 and isoform 8 active fragment. A) Structural alignment of the multimer with the yeast structure (Bi *et al,* 2007) and a zoom to the active WND motive in SEC31A. B) Alignment of the multimer structures and the active fragment of SEC31A isoform 1 and 8, exon 24c encoded residues marked in yellow. C) Detailed depiction of the interaction site of the SEC31A isoform 8 interaction site and SEC23A. D) Zoom to the three interacting residue areas in the active fragment. (Figure done by Benjamin Dimos-Röhl)

see the same effect in all predicted structures with isoforms 1 and 8.

Comparing the multimer structure prediction of SAR1A, SEC23A, and the active fragment of SEC31A, we again see strong similarities between those two protein complexes. However, in two of the top five ranked structures within the SEC31A active fragment isoform 8, we observe that upstream of the WND motif at the N-terminus lies a region that heavily interacts with SEC23A (Figure 3-11B). This is striking, because none of the predicted structures with SEC31A isoform 1 show additional interaction with SEC23A. In these cases, the polypeptide dissociates away from the complex. The active fragment of SEC31A with isoform 8 seems to have residues that can interact with SEC23A, lacking in isoform 1. Upon close inspection, we find four additional polar contacts between the residues encoded in exon 24c and SEC23A (Figure 3-11C). In closer observation, we detect three interactions that stabilize the disorganized active fragment in that position (Figure 3-11D). First, the asparagine 991 in the active fragment has a polar contact with the glutamic acid 694 in SEC23A. Additionally, this interaction is stabilized by the lysine 1021 in the C-terminus of the active fragment and aspartic acid 691 of SEC23A. Another interaction site is the polar contact between the glutamine 995 in the active fragment and the lysine 392 of SEC23A. The last interaction sites are polar contacts between proline 999 and leucine 1001 in the active fragment and glutamine 387 in SEC23A. These potent interactions could lead to a stronger connection between SEC23A and SEC13A.

3.4 RBM47 is a potent splicing regulator for exon 24c in SEC31A

We were now able to examine the effect of alternative splicing of the exon 24c in SEC31A on lipid secretion and a possible structural model based on a multimer AlphaFold prediction. The last thing to in depth characterize exon 24c is to examine the regulation of its tissue dependent alternative splicing. As it is alternatively spliced in a tissue-dependent manner, we could start to search for a regulator in the correlation analysis (Figure 3-4C). We highlighted all genes in the GO term “mRNA splicing.” Many splicing genes show low correlation with exon 24c, ranging between -0.3 and 0.3 in both Pearson and Spearman correlation indices. But one

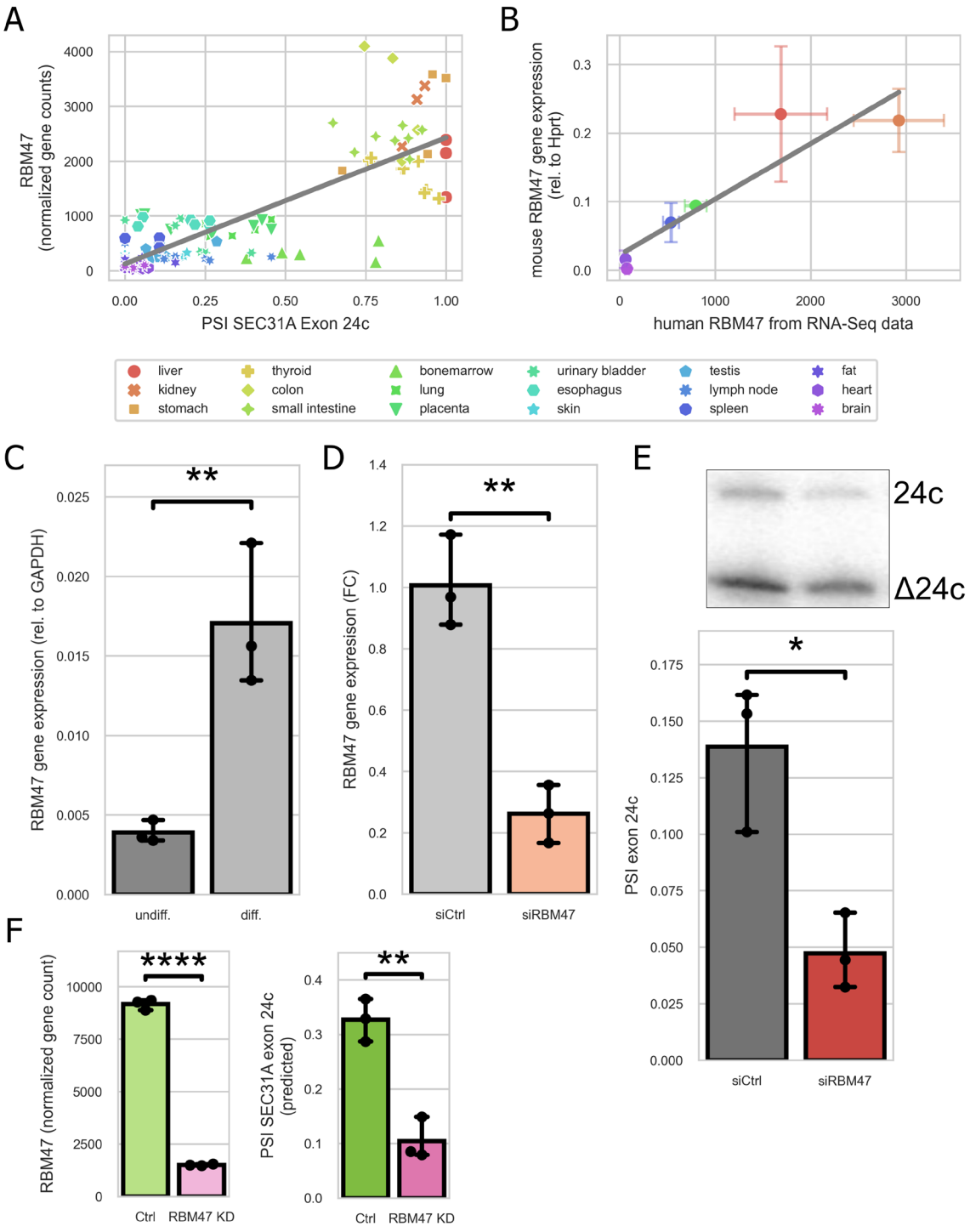

Figure 3-12: RBM47 gene expression correlates with exon 24c inclusion. A) Exon 24c inclusion and RBM47 gene expression in the human tissue RNAseq data as linear regression. B) Linear regression of human tissue RNAsew data and mouse tissue RT-qPCR. The points y-coordinate depict the mean RBM47 expression of mouse tissues relative to Hprt, and the y-error bars show its standard deviation (n = 3). The X-coordinate shows the mean RBM47 gene expression of human RNA-seq data and their standard deviation as x-error bars (n > 2, n < 10) C) RT-qPCR for RBM47 in Caco-2 cells before and after differentiation. D) Exon 24c inclusion in Caco-2 cells with siRBM47 treatment. E) RBM47 levels and exon 24c levels in a RNAseq data set of human lung fibroblasts treated with siRBM47. The mean of the data (C - E) is shown as bar plot, and their standard deviation as error bars. Statistical significance was validated with the Student-t-test and shows via asterisks: p* < 0.05, p** < 0.01, p**** < 0.0001.

gene from the GO term “mRNA splicing”, RBM47, correlates exceptionally with exon 24c and even shows with the Spearman and the Pearson correlation, which is also the highest correlation for genes overall, and that is RBM47 (Figure 3-12A). RBM47 is a known tissue-dependent splice regulator that activates alternative splice switches in the context of the EMT/MET (Yang et al. 2016b). Together with other splice regulators, it promotes the epithelial splice pattern. In this context, RBM47 is also involved in cancer progression as it is downregulated in the transition to mesenchymal cells (Sakurai et al. 2016; Rokavec et al. 2017; Kim et al. 2019).

To validate the tissue dependency of RBM47 gene expression, we performed RT-qPCR on the mouse tissues and compared it to the gene expression levels in human tissues in the RNAseq data (Figure 3-12B). The mouse tissues show highly similar gene expression levels for RBM47 and demonstrate the conservation across species for this splicing factor. We know that exon 24c inclusion increases in Caco-2 cells after one week of differentiation (Figure 3-8D). Therefore, similarly to as in the mice tissues, we examined the RBM47 gene expression in differentiated Caco-2 cells (Figure 3-12C). The RBM47 gene expression is greatly increased in differentiated Caco-2 cells, highlighting a correlation between the expression of the splicing regulator and the inclusion of exon 24c. When treated with siRNA against RBM47, we could observe a decrease of exon 24c in Caco-2 cells (Figure 3-12E). A similar experiment was already done on human lung fibroblasts, where they treated the cells with siRBM47, and sequenced the transcriptome (Yang et al. 2016b). After quantifying the reduction of RBM47 gene expression in the H358 siRBM47 data set, we could clearly show that exon 24c inclusion was reduced by reducing the gene expression of RBM47 (Figure 3-12F).

All these data indicate that the gene expression of RBM47 and the inclusion of exon 24c correlate on multiple different data sets and cell types. We performed a minigene experiment to validate further that RBM47 is directly involved in the regulation of the inclusion of exon 24c. First, we validated that the RBM47-induced inclusion of exon 24c also works in Hek and HeLa cells with a FLAG-RBM47 overexpression vector (Figure 3-13A). We then inserted the exon 24c and different versions of flanking introns in an adapted luciferase-assay construct we had already implemented in our work group (Neumann et al. 2020). This construct contains a GFP gene with an inserted intron. This intron contains a branchpoint and strong splice sites. We inserted the exon 24c in this intron without interfering with the branchpoint. To examine whether and where RBM47 would bind on SEC31A pre-mRNA, we inserted exon24c with the flanking intron of 99 nt (S), 202 nt (M), and 685 nt (L). We transfected the constructs in Hek cells with either the FLAG-RBM47 overexpression vector or an empty FLAG-Ctrl vector. After RNA extraction, we performed splice-sensitive RT-PCR with primers

in the GFP regions, thus flanking the intron and the inserted exon 24c (Figure 3-13B). We did not detect any meaningful splicing of exon 24c in the control minigene. We also could not detect any inclusion of the exon with the S construct, having just 99 nt. But in the M and L construct, with longer flanking intron sequences, we could observe a significant difference in the inclusion of exon 24c after RBM47 over expression. Additionally, there was no significant difference in exon 24c inclusion between the M and L construct. This shows that RBM47 requires the intronic region around exon 24c of the pre-mRNA of SEC31A to increase the inclusion of that exon. Our findings furthermore indicate that the 103 nt long region between the S and M construct is crucial for this regulation.

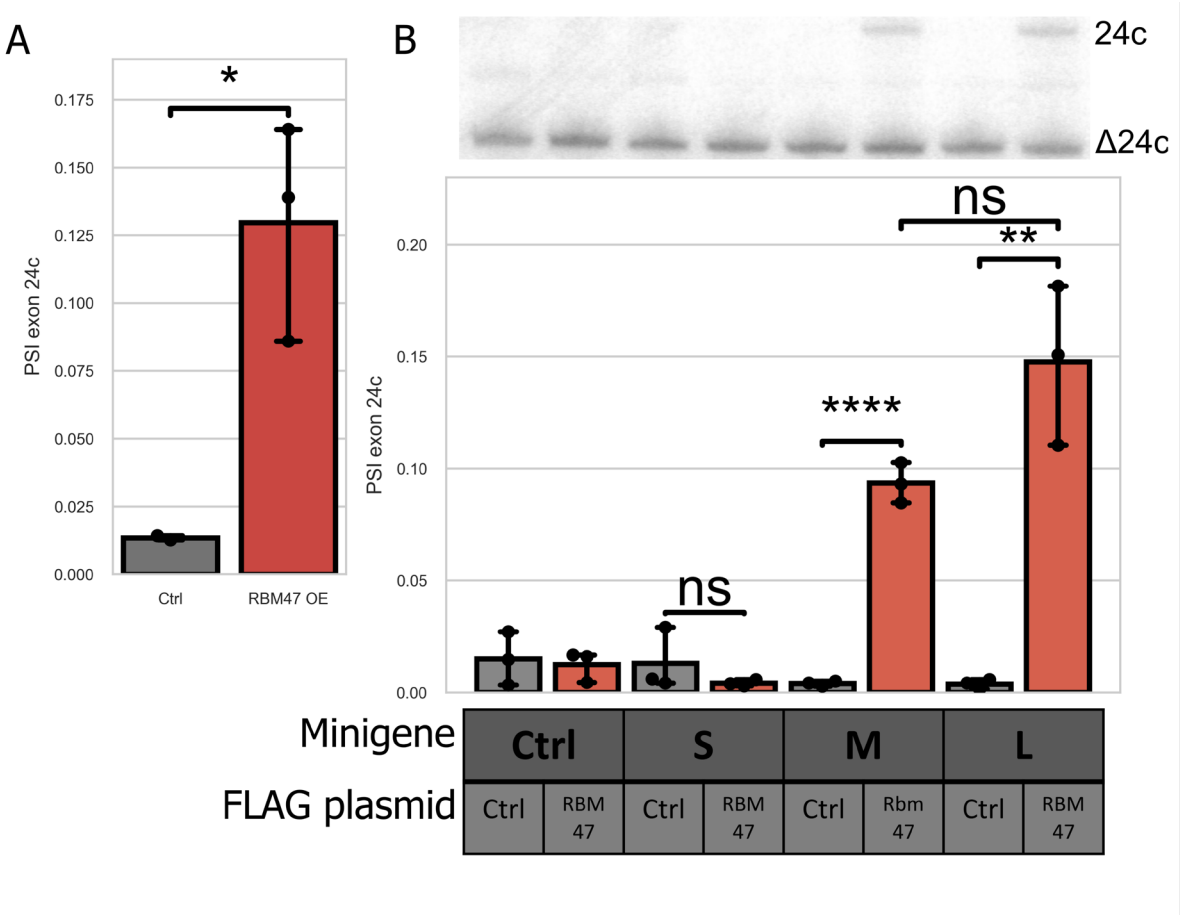

Figure 3-13: Minigene experiment show connection of RBM47 expression and exon 24c inclusion. A) Exon 24c inclusion in Hek and HeLa cells after transfection with a FLAG-RBM47 overexpression vector. B) Minigene experiment on Hek cells with three different minigene constructs, containing exon 24c and control exon. The mean PSI was depicted as bar plots, with their standard deviation as error bars (n= 3). Statistical significance was validated using Students-t-test and shown via asterisks: ns $p > 0.05$, p* < 0.05, p** < 0.01, p**** < 0.0001.

3.5 In silico study shows transcriptional changes in secretory genes during B Lymphocytes differentiation

The results of this chapter are part of an already published paper in the scope of this doctorate project in Frontiers of Immunology, "In silico analysis of alternative splicing events implicated in intracellular trafficking during B-lymphocyte differentiation" by Felix Ostwaldt, Bruna Los, and Florian Heyd (Ostwaldt et al. 2022).

3.5.1 Gene expression and splicing in secretory genes are heavily altered during B-cell differentiation

After identifying and thoroughly characterizing specific splice events in SEC31A, which regulate secretion, we wanted to conduct a comprehensive transcriptomic study regarding alternative splicing in secretion regulation. We chose to examine gene expression and alternative splicing in B lymphocytes. B lymphocytes are very suitable for studying the regulatory mechanisms for secretion as they undergo a heavy change in behavior during their life cycle. The primary function of PC and PB is to secrete antibodies, and therefore, their entire metabolism is changed to aid that function (Nguyen et al. 2019; Tellier and Nutt 2019). Furthermore, we analyze pre-plasmablasts (preBP) and MBC. PreBP is the intermediate form of PB that will develop into a total PB but still lacks the full capacities of antibody-secreting cells, and MBCs are long-lived cells that can become antibody-secreting cells. In their MBC form, they also lack the high efficiency that PCs possess to secrete. We used an openly available RNAseq data set consisting of MBC, preBP, BP, and PC, each as triplicates (GSE148924). This data set is exceptionally suitable for transcriptomic analysis as the samples show a high degree of similarity inside the sample group and are distinctive enough to show effects at the transcriptomic level (Kassambara et al. 2021; Alaterre et al. 2021).

We first examined the global gene expression of the B lymphocytes. Using DEseq2, we were able to quantify 15,754 genes in the data set. We filtered the genes for each comparison for a log2FC > 0.5, or < -0.5, and a Padj < 0.001 and visualized them in a clustering heatmap (Figure 3-14A). Overall, we detect a high similarity between PB and PC gene expression. Only 125 genes are differently expressed between these two sample groups. The preBP visibly

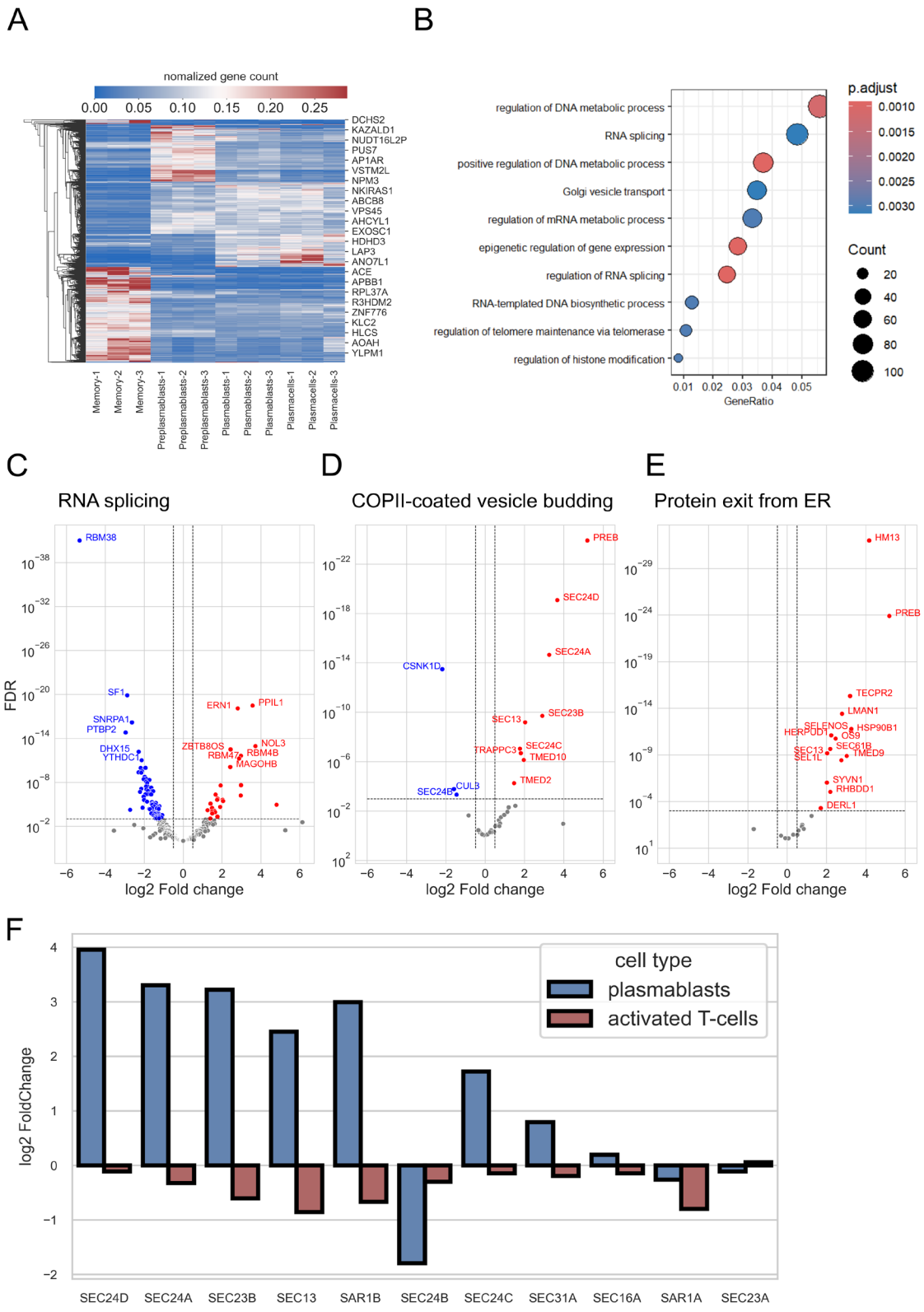

Figure 3-14: Gene expression analysis of B lymphocytes. A) Heatmap showing the DEseq gene count of all significantly changed genes in B lymphocyte differentiation. B) Gene ontology enrichment analysis of genes upregulated in plasmablasts in comparison to memory B cells. C) Volcano plot of genes in the GO term RNA splicing of the comparison pasmacells with memomry B cells. D) Volcano plot of the same comparison with gene of the GO term COPII-coated vesicle budding. E) Volcano plot of the genes of the GO term Protein exit from the ER. F) Comparison of the gene expression of the key genes of the COPII between activated T-cells and the comparison of memory B-cells to plasamacells. Original figures done by Bruna Los, adapted from (Ostwaldt et al. 2022)

showed similarities in gene expression compared to PB but not entirely, but we observed they show a large cluster of genes that are upregulated in prePB and have a lower expression in PB. 2383 genes are differently expressed between preBP and BP, and 3079 differed between preBP and PC. That mirrors the transition prePB goes through to eventually become PC. However, the outlier in the data set is the MBC. This group shows the opposite gene expression profile as the other three. 6362 genes are differently expressed between MBC and preBP, 6144 between MBC and PB, and 4929 between MBC and PC. We performed a GO term enrichment analysis in all the different stages of B lymphocytes. Interestingly, in the comparison between MBC and PB, we found that the GO terms “mRNA processing” (GO:0006397) and “RNA splicing” (GO:0008380) have a higher enrichment in PB, showing the possibility of alternative splicing in these stages of differentiation (Figure 3-14B). When comparing MBC and PC and observing only the gene in the “RNA splicing” GO term, we could see robust changes in gene expression in both directions, arguing for a regulation of MBC on a splicing level (Figure 3-14C). The “COPII-coated vesicle budding” GO term is interesting in the comparison between MBC and PC, as it shows that critical factors in the COPII complex are upregulated in PC, arguing for an increased secretory load and the need for adapting to that (Figure 3-14D). However, in this comparison, three genes are upregulated in MBC: SEC24B, CUL3, and CSNK1D. These could be MBC-specific regulations, changing the behavior of the secretory pathway. In the GO term, “protein exits from the ER”, only genes are upregulated in PC compared to MBC, showing a strong need for increased expression of secretory genes in PC.

To fully understand the mechanisms by which PC adapts to the higher secretory load, we compared the expression of the key COPII genes from not antibody-secreting preBP to antibody-secreting PC with the change of expression in activated T-cells (Figure 3-14F). Whereas the gene expression of T-cells mostly stays the same, and most of the regulation happens on the splicing level (Neumann et al. 2019), PC shows a substantial increase in the gene expression of the COPII genes. That shows that B and T lymphocytes practice different kinds of regulation of the secretory pathway.

We performed a splicing analysis with the B lymphocyte data using rMATS for a complete picture of the transcriptomic landscape. A total of 44,879 splice events were quantified, whereas 3179 of them were alternatively spliced in at least one comparison. We depicted all significant splice events (delta PSI > 0.4 or < -0.4, P-value < 0.001) for all samples as a clustering heatmap (Figure 3-15A). The significant splice events are 80.07% exon skipping (2683), 9.53% alternative 3’ splice site (234), 8.18% alternative 5’ splice site (209), and 2.16% intron retention (53). In strong contrast to the gene expression, alternative splicing of differentiating B lymphocytes is very similar in preBP, PB, and PC. However, the alternative

splicing of MBC shows vast differences from all other stages of differentiation. All significant alternative splice events are spliced differently in MBC. That is a strong indication that, while differentiation of B lymphocytes from preBP, over PB to PC, is mainly regulated on gene expression level, differentiation to MBC requires, in addition, the effects of alternative splicing.

Because we are interested in how splicing can regulate the secretory pathway, we filtered the splice events with the GO term “COPII-coated vesicle budding” (Figure 3-15B). We found seven of these splice events are primarily included in MBC, whereas 13 are included in PreBP, BP, and PC. Prominent members of the COPII complex, like SE31A, SEC16A, and

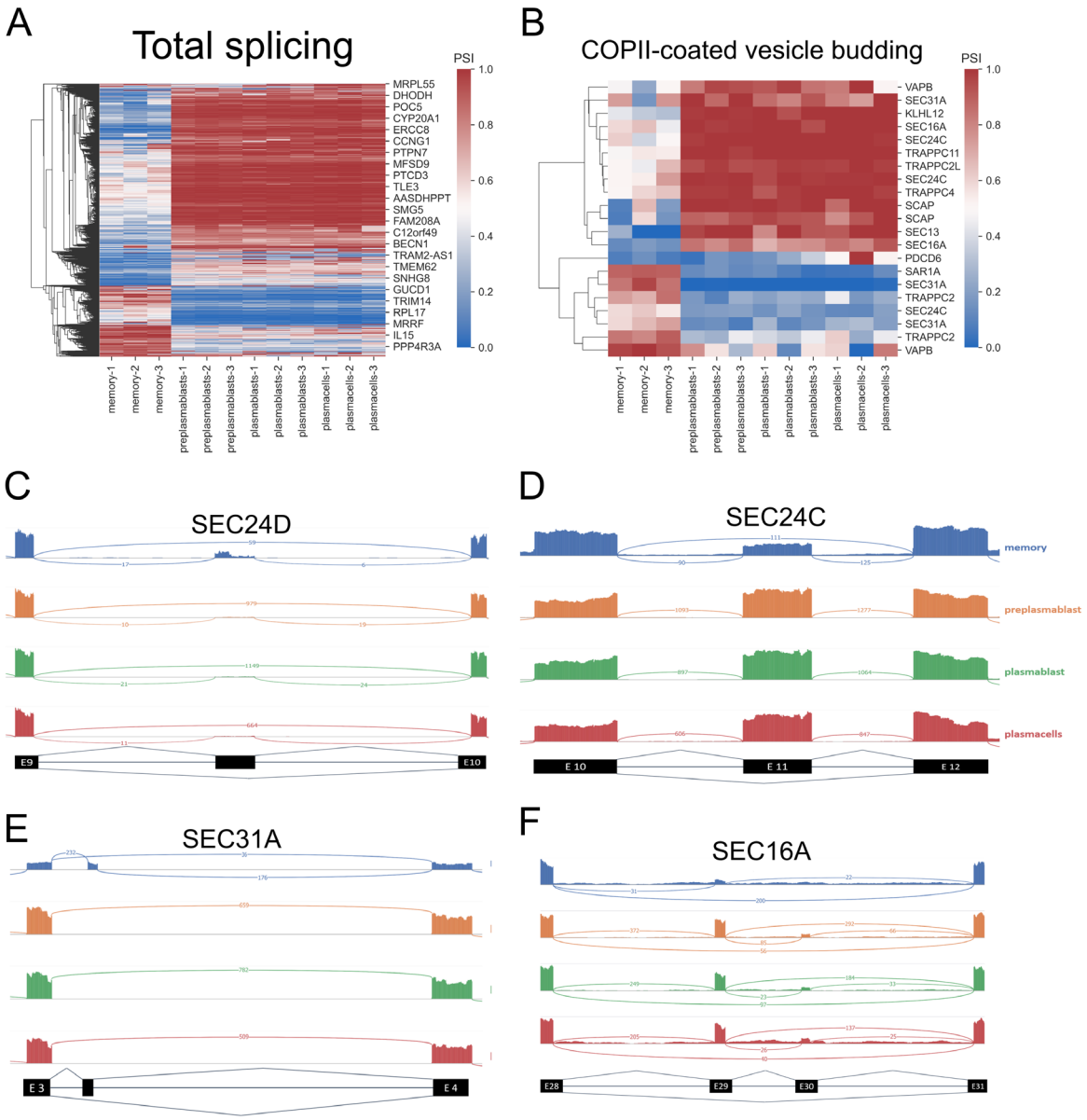

Figure 3-15: Splicing analysis of differentiating B lymphocytes. A) Heatmap of the mean PSI of all significantly changed splice events (cassette exons, alternative splice sites and intron retention) in the four stages of B lymphocyte differentiation (n = 3). B) Heatmap of significantly spliced alternative splice events in the GO term COPII-coated vesicle budding. C) Sashimi plots of the raw reads of the four stages of B lymphocyte differentiation for SEC24D, D) SEC24C, E) SEC31A, F) and SEC16A. Adapted from (Ostwaldt et al. 2022)

SEC24C, have multiple alternatively spliced exons to differentiate B lymphocytes. SEC24D has an alternative cassette exon, which has a slightly higher inclusion in MBC than in the other stages of B lymphocyte differentiation (Figure 3-15C). This exon between exons 9 and 10 includes a premature stop codon. The inclusion of this poison exon would trigger an NMD, leading to a lower abundance of SEC24D transcript (Kurosaki et al. 2019). This is in accordance with the lower gene expression level of SEC24D in MBC (Figure 3-14D). Like SEC24D SEC24C, it has an alternatively spliced exon 11 with a length of 116 nucleotides. Exclusion of this exon would lead to a frameshift and premature stop codon. In MBC, we observe a roughly 50% inclusion of exon 11 in SEC24C, while in the other stages of B lymphocyte differentiation, it is included by 100% (Figure 3-15D). SEC24C's function is to bind cargo with its IxM cargo binding motive. It helps secrete Q-SNARES like Syntaxin, GS27, and BET1 (Adolf et al. 2016). That could mean that MBC changes their specificity of secretion with the help of alternative splicing of SEC24C.

Our analysis suggests that these COPII components could be regulated through poison exons and NMD. Other COPII components show possible functional alternative spicing, leading to a functional protein isoform. MBC includes a 50-nucleotide-long exon after exon 3 in SEC31A (Figure 3-15E). In all other B lymphocyte differentiation stages, this exon is not included. This alternative exon is likely coupled to an alternative translation start and a shorter N-terminus of SEC31A. At the N-terminus of SEC31A lays the WD40 domain, which interacts with the second component of the outer layer of the COPII complex SEC13 ((Stagg et al. 2008)). The truncated WD40 domain in SEC31A could lead to a weaker interaction between SEC31A and SEC13 and, therefore, an impaired regulation of secretion through ubiquitylation.

In a previous work of our group, we showed that alternative splicing of the COPII scaffolding protein SEC16A controls the efficiency of splicing in activated T-cells (Wilhelmi et al. 2016). Exon 29, predominantly included in activated T-cells, could enhance the export efficiency. In B lymphocyte differentiation, exon 29 is primarily included in prePB, PB, and PC and, to a lesser amount, in MBC. This could mean that similar like T-cells after activation, B lymphocytes, after differentiation, regulate their secretion through alternative splicing of SEC16A exon 29.

3.5.2 KLHL12 and PICALM splicing show strong switches in B-cells.

We showed that components of the COPII complex regulate gene expression and alternative splicing in differentiating B lymphocytes. Similarly, other genes that are not directly related to the COPII family are regulated and also show effects on secretion. The KLHL12 shows an exciting splice pattern in B lymphocytes. KLHL12 is an adaptor of the E3 ubiquitin ligase

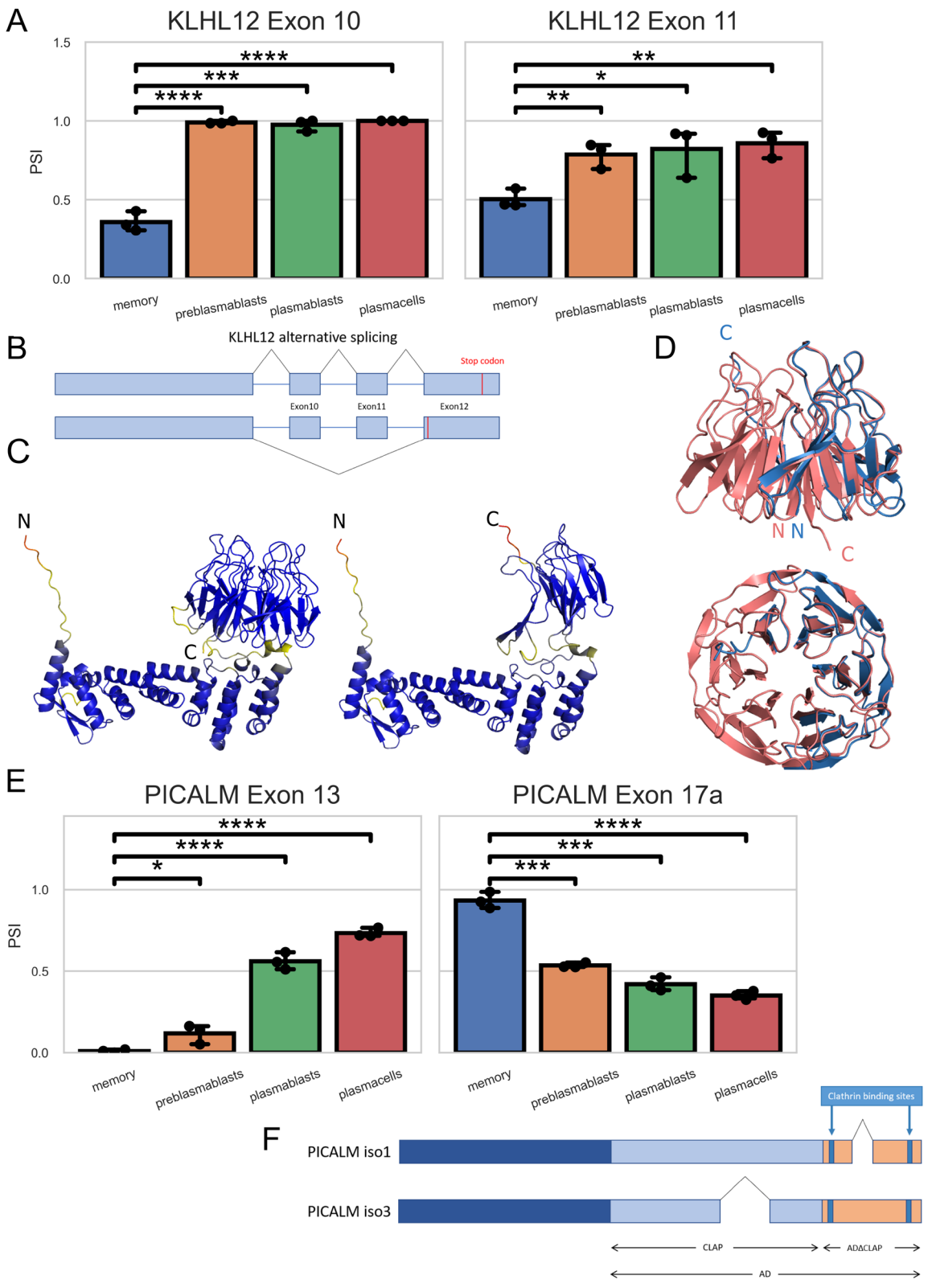

Figure 3-16: Alternative splicing of KLHL12 and PICALM in B lymphocyte differentiation. A) Mean PSI of the alternatively spliced exons 10 and 11 in KLHL12 depicted as bar plot with their standard deviation as error bars (n = 3). B) Schematic of the splice mechanics in KLHL12. C) Alphafold prediction of KLHL12 wild type (left) and with exon 10 and 11 skipping and subsequent shortened C-terminus (right). The color of the structure corresponds with the pae, where blue is a pae of 100 and red is 0. D) Alphafold prediction o the kelch-like domain of KLHL12 wild type (red) or skipped exon 10 and 11 form (blue) superimposed. E) Mean PSI of the alternatively pliced exons 13 and 17 in PICALM depicted as bar plot, with the error bars showing standard deviation. E) Overview of PICALM isoforms and the effect of the protein arrangement. Adapted from (Ostwaldt et al. 2022)

CUL3. The function of the $CUL3^{KLHL12}$ complex is to catalyze the monoubiquitylation of SEC31A and SEC31B together with the co-adapter subunit PEF1 (Akopian et al. 2022). KLHL12 is responsible for the interaction with SEC31 between the kelch domain of KLHL12 and the PGXPP motive in SEC31 (Chen et al. 2020b). The effect of ubiquitylation of SEC31 is an enhancement of collagen secretion (Jin et al. 2012). In the four stages of B lymphocyte differentiation, we detect two exons in KLHL12, which are alternatively spliced (Figure 3-16A). Exons 10 and 11 are primarily included in preBP, BP, and PC. In contrast, in MBC, the inclusion of exon 11 is reduced by around 30%, and that of exon 10 is by 60% less included compared to the other differentiation stages. KLHL12 has 12 coding exons, so exons 10 and 11 are near the 3' end of the RNA. Exon 10 has 99 nucleotides, so alternative splicing would not induce a frameshift. However, exon 11 has a length of 187 nucleotides, thus inducing a frameshift. This frameshift mediates a premature stop codon, yet it is still located in the last exon 12; thus, it is unable to induce NMD and produces a short C-terminus (Figure 3-16, Figure 3-15C). Based on these clues, we propose that the MBC-specific isoform of KLHL12 with exon 10 and 11 skippings could be translated into a protein. To estimate how the alternative splicing of KLHL12 would affect the protein, we performed an AlphaFold prediction of the standard isoform of KLHL12 and the delta exon 10 and 11 (Figure 3-16Figure 3-15C). The KLHL12 protein is highly ordered and produces an easy structure in the in-silico model. The exon 10 and 11 skipping isoform shows an impaired kelch domain. Three out of six kelch repeats are not included (Figure 3-16D). This will hinder KLHL12 from performing its adapter function for ubiquitylation of SEC31 by CUL3. MBCs are probably adapting to their low secretory life cycle with this alternative splicing switch.

The second highlight we observed in our data is the alternative splicing of the Phosphatidylinositol-binding clathrin assembly protein (PICALM). PICALM is responsible for the correct assembly of clathrin-coated vesicles for endocytosis (Meyerholz et al. 2005). It does this via two binding sites at its C-terminus, which binds clathrin triskelia and brings them close (Ye and Lafer 1995; Moshkanbaryans et al. 2016). With this interaction, PICALM can regulate the size of the clathrin-coated vesicles (Miller et al. 2015). In the B lymphocyte data, we found that exon 13 in PICALM included to almost 100% included in MBC and decreases its inclusion stepwise from prePB and PB to 40% inclusion in PC (Figure 3-16E). The opposite effect can be observed for exon 17a, which is not included in MBC, is slightly included to a small extent in preBP, and is included to around 60% in PB and 70% in PC. The length of both exons is divisible by 3, meaning alternative splicing does not induce a frameshift. Both isoforms can be found in Uniprot as isoform 1 and isoform 3, which is strong evidence that both splice isoforms get translated into protein. The peptides, for which both exons code, lie in intrinsically disordered regions. Performing an AlphaFold prediction is probably not

valuable for this case, as AlphaFolds limits lie specifically in intrinsically disordered regions. The protein isoform 1 with the exon 17a skipping has a shorter C-terminus. With this effect, the two clathrin binding sites lie closer together (Figure 3-16F). This could have an impact on the assembly function of the clathrin-coated vesicles and, therefore, could regulate the endocytosis in B lymphocytes.

3.5.3 NERF / ELF2 splicing could be a key regulator for secretory genes in B-cell differentiation

We noticed that in MBC, in the GO term “COPII-coated vesicles”, only three genes are upregulated compared to PC: SEC24B, CUL3, and CSNK1D show higher gene expression in MBC than in PC. As we could not find alternative splicing of poison exons to explain this phenomenon, we tried to find a common transcription regulator that could influence the gene expression of the secretory-related genes in MBC. For that, we performed a correlation analysis between the gene expression of SEC24C, CUL3, and CSNK1D and the gene expression of the TF. These factors were defined with the help of the GO terms “transcription factor binding” (GO:0008134) and “DNA-binding transcription factor activity, RNA polymerase II-specific” (GO:0000981). We performed the correlation analysis with the resulting 773 TFs and filtered all genes with a person correlation index of > 0.9 in one of the secretory genes upregulated in MBC (Figure 3-17A). From the resulting 236 genes, only 135 genes would correlate with SEC24B, CUL3, and CSNK1D. To narrow down the possible TF for MBC secretory genes, we performed the same correlation analysis with the human tissue data set we already used for our global transcriptome analysis. A total of 135 TFs were correlated with SEC24B, CUL3, and CSNK1D in the human tissue data set. The correlation indices of each gene with the three secretory genes were averaged and plotted against the mean p-value of the correlation. This analysis showed that NERF/ELF2 does correlate the strongest with our three secretory genes.

NERF/ELF2 is an Ets-related TF known to regulate many T- and B-cell-specific gene expression patterns (Oettgen et al. 1996). NERF/ELF2 gene expression is 4-fold times higher compared to the gene expression in prePB, PB, and PC (Figure 3-17B). In addition to regulating gene expression levels, NERF/ELF2 exhibits multiple splice isoforms. NERF-1a and NERF-1b have a substantially shorter N-terminus caused by an alternative promoter and translation start in exon 5(Figure 3-17C). NERF-2a and NERF-2b have an earlier translation, including an alternative promoter and a skipping of exon 5. The isoforms a and b differ from each other through an alternative splice site in exon 7; a isoform skips the alternative splice site, and b isoforms have them included. Examining the distribution of the isoforms, we detect that no b isoform is expressed in MBC (Figure 3-17D). NERF-1a makes up the majority of NERF/ELF2 transcript with 70%, while NERF-2a makes up around 25%. This distribution is

changed during the following stages of MBC differentiation. NERF-1a and NERF-1b have a roughly equal distribution of around 40% in NERF/ELF2 transcripts in prePB and PB.

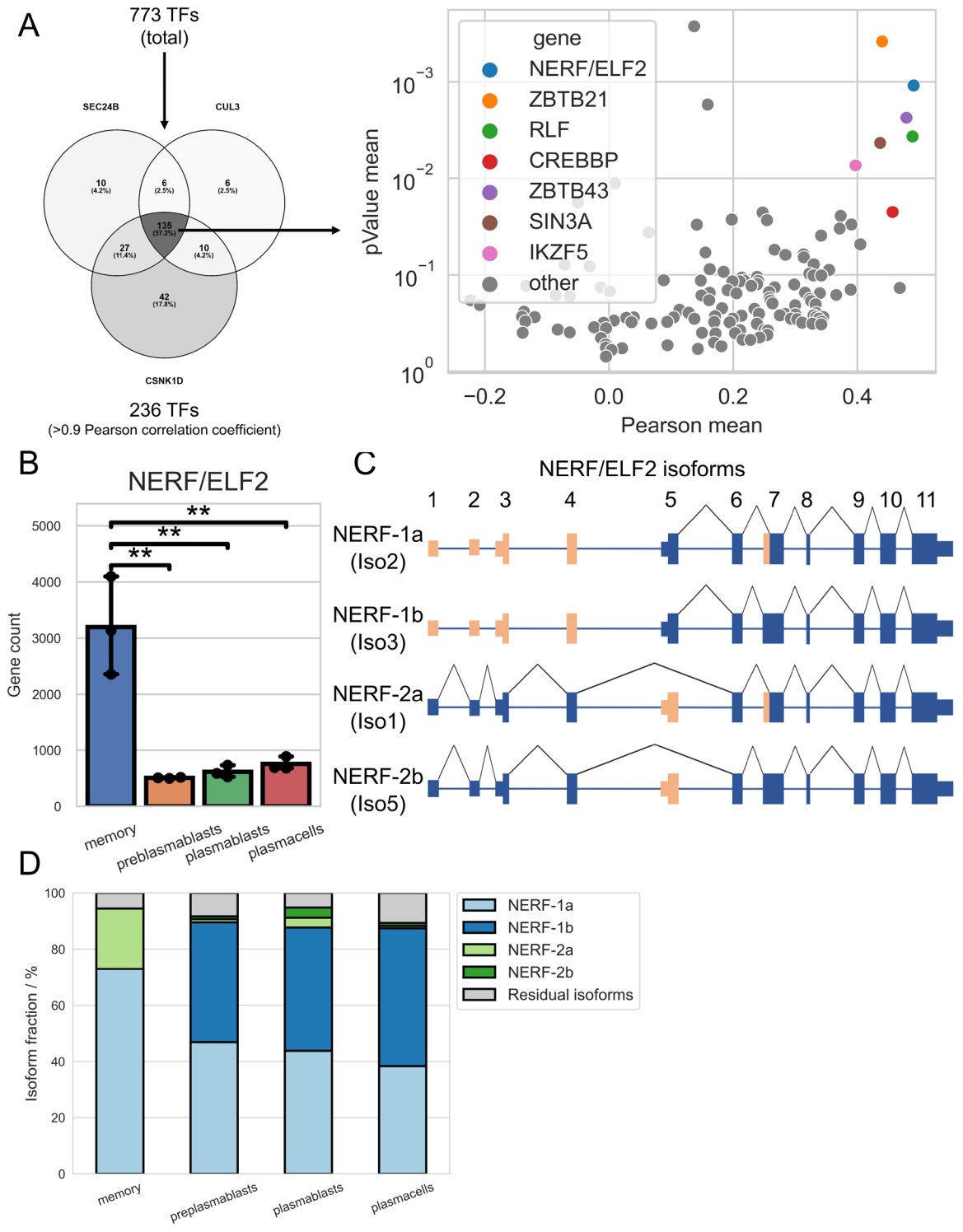

Figure 3-17: NERF2/ELF as possible transcription factor for B lymphocyte secretion. A) Through correlation analysis with the genes SEC24C, CUL3 and CSNK1D found 135 transcription factors, which where additionally correlated in the human tissue data set. B) NERF/ELF2 mean gene expression levels in B lymphocyte differentiation stages depicted as bar plot with their standard deviation as error bars (n = 3). C) Schematic overview of NERF/ELF2 isoforms taken from uniport. D) Distribution of mean NERF/ELF2 isoforms in the stages of B lymphocyte differentiation. Adapted from (Ostwaldt et al. 2022)

NERF-1b expression is increasing in PC to around 50%. This could mean that NERF/ELF2 could regulate the fate of B lymphocyte differentiation through gene expression and alternative splicing and specifically regulate the gene expression of secretory genes in MBC.

3.5.4 RBM47 gene expression and SEC31A splicing is changed with B lymphocyte differentiation

It was previously reported that RBM47 post-transcriptional regulation is crucial for regulating B lymphocytes (Wei et al. 2019). Subsequently, we examined the gene expression of RBM47 in the B lymphocyte differentiation data set. Gene expression of RBM47 quite drastically changes between MBC and PB, approximately 10 fold (Figure 3-18A). PB shows the highest extent of RBM47 gene expression of the four differentiation stages of B lymphocytes, followed by PC and prePB. Caused by the high variation of the gene expression levels of RBM47 in PC and PB, these differences are not significant. Following the RBM47 gene expression, we examined the alternative splicing of exons 24b and 24c in SEC31A (Figure 3-18B). Exon 24b inclusion is generally on a high level with a PSI of around 0.9 in B lymphocytes, except in MBC, where it shows a slight (but significant) reduction of approximately 10% compared to the inclusion in MBC. Contrary to that, the inclusion of exon 24c is generally on a low level

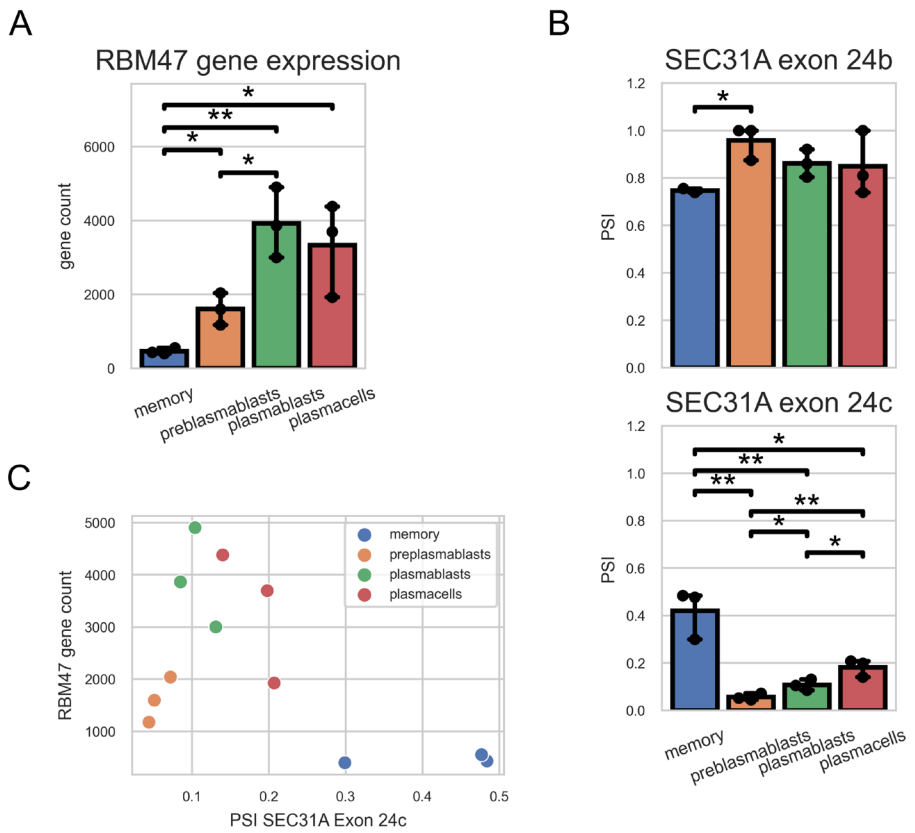

Figure 3-18: SEC31A splicing and RBM47 change with B lymphocyte differentiation. A) RBM47 gene expression levels in memory B cells, pre plasmablasts, plasmablasts and plasmacells. B) Splicing quantification of exon 24b (left) and exon 24c (right) of SEC31A in B lymphocyte differentiation. C) Exon 24c inclusion and RBM47 gene expression as scatterplot. The bar plots show the mean, and the error bars the standard deviation (n=3). Statistical significance validated using Students-t-test and depicted as asterisks: p* < 0.05, p** < 0.01.

and has its maximum in MBC with a PSI of 0.4. The lowest inclusion is observed in prePB with a PSI lower than 0.1 and step-wise increment over PB to PC. It seems to be regulated very tightly, as the inclusion levels of exon 24c have little variances and subsequently small significant changes. A low correlation is detected when examining the relation between RBM47 gene expression and exon 24c splicing (Figure 3-18C). Especially the high level of RBM47 gene expression in MBC seem to not enhance the inclusion of exon 24c. Additional factors have to be responsible for the MBC inclusion of exon 24c.

4 Discussion

The findings presented in this scientific work contribute to our understanding of how alternative splicing regulates the secretory pathway. We show how modifications caused by alternative splicing in key secretory genes changes their behavior and effect on secretion. We thoroughly used bioinformatic tools for transcriptomic analysis of gene expression and alternative splicing, which guided our experimental work. Computational analyses like AlphaFold helped explain observations we made in our experiments. We characterized exon 24c in SEC31A, consolidated the insight into the already described exon 24b in the same gene, and showed alternative splicing in differentiating B lymphocytes via in-silico analysis. With these results, we successfully connected alternative splicing with secretion and extended the regulatory relation between these two. Few but noteworthy publications examine the regulatory effects of specific splicing events on secretion. For example, an alternative translation initiation of the t-SNARE syntaxin 5 regulates its positional switch between the early Golgi and the ER (Hui et al. 1997). The alternative splicing of the NR1 gene controls the release of NMDARs from the ER and subsequent synaptic delivery (Mu et al. 2003). Activation-induced alternative splicing of Sec16 in T-cells controls the COPII mediated transport efficiency (Wilhelmi et al. 2016). A genome-wide examination of transcriptomic data revealed that alternative splicing impacts multiple stages of the secretory pathway (Neumann et al. 2019). Finally, two splicing isoforms of BMP2K, which both interact with SEC16, regulate the abundance and distribution of COPII assemblies, autophagy and erythroid differentiation (Cendrowski et al. 2020).

4.1 Tissue specific splicing analysis reveals new, and solidifies previously known splicing events

We used RNA sequencing data from a variety of human tissues (Fagerberg et al. 2014) to find relevant splicing events that could potentially change the secretory behavior of the cell. Distinct tissues usually show a diversity of functions. As such, secretion intensity, effectivity,

and specificity are diverse. So, in the case of searching for alternative splicing events, it is plausible to look at tissue data. By examining the rate of distinction the tissue samples show on gene expression and splicing level (Figure 7A and B), we detected elevated variance between the tissues in gene expression rather than splicing. Alternative splicing is known to be more species-dependent than tissue-dependent (Barbosa-Morais et al. 2012). The splicing patterns of liver cells in humans show more similarities with the splicing pattern in human kidney cells than with the liver cells of a mouse. Merkin and Russel et al. showed that, except for brain, muscle, and testis tissues, alternative splicing is more conserved in the same organism, as opposed to tissues (Merkin et al. 2012). Likewise, we detect a relatively distinct cluster of brain, heart, and testis tissues in the PCA plot for splicing but very similar splicing behavior for the rest of the tissues. Although alternative splicing differs more between various organisms, numerous alternative splicing events are indispensable for tissue development. For example, RBFOX2 regulates splicing events that lead to EMT (Murphy et al. 2018; Yang et al. 2016b). On the other hand, RBM47 is responsible for altering alternative splicing for a transition to epithelial from mesenchymal cells (Rokavec et al. 2017; Kim et al. 2019). These splicing factors also influence the severity of cancer development. Additionally, alternative splicing can change the development of adipose tissues (Vernia et al. 2016) and circadian rhythms in pancreatic cells (Marcheva et al. 2020). Despite these limitations for tissue-specific alternative splice analyses, we found multiple splicing events in secretory tissues (Figure 3-1C). We found several splicing events regulating secretion efficiency, that were already previously reported by others, which shows the effectiveness of our splicing analysis in tissue data for secretory genes.(Wilhelmi et al. 2016; Neumann et al. 2019). It was previously described that alternative splicing regulates secretion throughout the cells organelles (Neumann et al. 2019), which we see in tissue splicing analysis. Exon 24c has not been characterized previously, and it is the most profoundly changed splicing event in our human tissue data. Additional exons in SEC31A, exon 24, exon 13, and exon 14 could be relevant for splicing regulation.

The tissue specificity of exon 24c is highly conserved between the human RNA sequencing data and mouse RT-qPCR analysis (Figure 3-2). In 100 vertebrates (from the USCS genome browser), the conservation of exon 24c itself was shown to be slightly lower than constitutive exons (Figure 3-3C), but the downstream intronic region is more conserved than ordinary introns. Together with the conservation of RBM47 expression (Figure 3-12B) and the conservation of amino acids in closely related species (Figure 3-10A) we propose that tissue-specific alternative splicing of SEC31 exon 24c, its regulation and effect on the protein is highly conserved among higher vertebrates. As the amino acid sequence for exon 24c in chicken already shows major differences, and the SEC31A homologs in yeast completely

lack the exon, we suggest the conservation of exon 24c is particular to organisms closely related to humans. As mentioned, organism-specific alternative splicing is usually more prevalent than tissue-dependent alternative splicing. Examining an example with exon 24c, the direct opposite effect can be observed: alternative splicing of this exon is strongly tissue-specific and conserved across species. This is also in line with previously published data (Reyes et al. 2013), as it was stated that extremely tissue-specific exon usage is usually conserved between species. In addition, these splicing events frequently appear in disordered protein regions, thus increasing the regulation of protein-protein interactions across tissues.

The alternative splicing of exon 24b is slightly less tissue-specific and alternatively spliced compared to exon 24c. Exon 24b and 24c splice patterns do not correlate (Figure 3-3). These two exons lay sequentially next to each other on the mRNA level. On the pre-mRNA level, an 11,000-nucleotide-long intron is located between them. As they are alternatively spliced exons their inclusion depends on splice factors. We do not observe a close relationship between these exons, so we assume that the splice factors regulating those exons differ from each other. In accordance with that, the primary splicing regulator of exon 24c, RBM47, does not change the inclusion of exon 24b (Figure 3-12). We can identify three distinct populations of tissue samples: one where exon 24c and 24b are included, one where only exon 24b is included, and the last one where none of the two exons are included. Few samples show an exon 24c inclusion without exon 24b. This tells us that they are semi-independently regulated. The correlation analyses of the two exons also show a distinct group of genes with which they correlate more. Correlation analyses of gene expression data were primarily done for co-expression analyses to construct gene expression networks (Ruprecht et al. 2017; van Dam et al. 2018). These were used extensively in biological contexts, for example, addressing genes related to mouse weight (Ghazalpour et al. 2006), brain development (Oldham et al. 2008) and disease (Willsey et al. 2013; Boukas et al. 2019). Networks with isoform abundance and splicing with tissue-specific gene expression were already done genome-wide (Saha et al. 2017). We used this approach for a single splicing event, regarding total gene expression, without needing a total co-expression network. With a strong positive correlation, we could predict the function of exon 24c as a regulator of lipid transport. The strong negative correlation between the expression of some collagens and the inclusion of exon 24b also successfully led us to the regulation of collagen secretion by exon 24b alternative splicing. Regarding co-expression networks, it is known that higher-expressed genes generally correlate better with each other (Crow et al. 2016). This bias was previously addressed (Wang et al. 2022), but regarding our work it was not relevant, as gene expression and splicing had a different quantification. However, this improvement suggested in the work of Wang et al. could be used to enhance the power of the correlation analysis.

4.2 Correlation analysis with exon 24c inclusion unveils its function and regulation

Following the correlation analysis of exon 24c, we found evidence that including this exon enhances lipid transport in differentiated Caco-2 cells (Figure 3-8). However, the regulatory effect of exon 24c is not to be overestimated, as the impact of TANGO1 and cTAGE5 is much more crucial. Exon 24c is probably a supporting actor, adding complexity to the field of big cargo secretion. There is, of course, the slight possibility that the inclusion of exon 24c is not only responsible for secreting lipids. Other mechanisms of action could be the secretion of extracellular matrix proteins, which is vital for the differentiation of Caco-2 cells, or generally the enhancement of differentiation efficiency through enhanced ER export of differentiation regulators with exon 24c inclusion. One way to examine these would be to measure the transepithelial electrical resistance to examine the impermeability of the cell monolayer. The inclusion of the exon is highly increased after the differentiation of Caco-2 cells (Figure 3-8D) and nearly completely skipped in HEK cells (Figure 3-13A), which again indicates a highly specific function for the Caco-2 cells.

Progressing from the function of exon 24c, we were also able to get insight into the regulation of exon 24c through correlation studies. Utilizing additional knockdown data, gene expression validation in mouse tissue RNA, and knockdown experiments in Caco-2 cells, we showed that RBM47 gene expression correlates with inclusion of exon 24c (Figure 3-12). Finally, the minigene experiment showed that RBM47 is responsible for the inclusion of exon 24c and that the intron 100 nucleotides downstream of the exon a crucial for this regulation(Figure 3-13). The approach using gene expression data to find splicing regulators is straightforward but has its limits. We can only detect correlations if activation of the protein is also regulated by its gene expression. Regulators, which are translational or otherwise regulated, could not be detected here. A further correlation with mass spectrometry data could provide more accurate insights. In this case, it would be very plausible that RBM47 is not the only splicing regulator influencing exon 24c inclusion. Also, the presence of an antagonistic regulator to exon 24c is possible.

It was previously shown that RBM47 is affiliated with the EMT. TGF-β/SMAD downregulate RBM47 (Sakurai et al. 2016), and the IL-6/IL-6R/STAT3 pathways (Rokavec et al. 2017) in EMT and RBM47 depletion caused a reduction of pro-epithelial genes like CDH1, OCLN, CLDN1, and TJP1. TJP1 pre-mRNA is also a target of RBM47, as RBM47 regulates its alternative splicing and enhances an epithelial-specific isoform (Kim et al. 2019). RBM47, ESRP1 and ESRP2 were reported to regulate other epithelial-specific alternative splicing events (Yang et al. 2016b; Vanharanta et al. 2014). We do not have any information on the

composition of cell types in the tissues in the human tissue data set. Therefore, we cannot say if the inclusion of exon24c is specific for epithelial cells. Further analysis of single-cell data would be needed.

RBM47 expression is widely regarded as a marker for the severity of cancer or even as a tumor suppressor gene (Su et al. 2023). It was first connected to breast cancer, as it shows suppressing characteristics and is downregulated in basal and claudin-low breast cancer (Vanharanta et al. 2014). Also, brain metastasis was related to the inactive variant RBM47^{I281fs}. Subsequently, many other cancer studies showed the affiliation between RBM47 downregulation or mutation and cancer severity (Rokavec et al. 2017; Chen et al. 2020a; Li et al. 2020; Tirosh et al. 2019; Sakurai et al. 2016). RBM47 also regulates the functionality of immune cells. Interleukin 10 (IL-10) is vital for the immune suppressive function of regulatory B cells (Cho et al. 2017). It was reported that RBM47 increases the IL-10 production in B cells via binding at the AU-rich 3' untranslated region and stabilizes IL-10 mRNA (Wei et al. 2019). Also, it helps with the degradation of mitochondrial antiviral signaling protein and, consequently, interferon production in zebrafish (Lu et al. 2020). PBs also produce IL-10 as a regulatory mechanism in autoimmune inflammation (Matsumoto et al. 2014) or for autocrine regulation of B lymphocyte differentiation to IgM or IgG-secreting cells (Heine et al. 2014). As we described earlier, RBM47 does not seem to regulate exon 24c inclusion in a linear relation. Particularly in MBC, the RBM47 gene expression is the lowest under the four B-cell differentiation stages, and the exon 24c inclusion is the highest (Figure 3-18). Additional factors for exon 24c inclusion must be considered for other cell types that do not behave like epithelial cells or are otherwise related to the EMT/MET. The inclusion of exon 24c in MBC could be part of the tumor-suppressor activity, as it could help regulate cell surface markers and inhibit cancer immune escape. It could also change the efficiency and specificity of secretion in B lymphocytes, as it is dynamically changed in B lymphocyte differentiation (Figure 3-18B). How exon 24c inclusion could be an element of immune cell regulation and cancer development should be a topic of future studies. In our commonly used cancer cell systems (HeLa, HEK), exon 24c is not included, which supports the association between exon 24c skipping and cancer progression (Figure 3-12A).

4.3 Structure prediction of SEC23A-SAR1A-SEC31A interface enlightens the mechanism of COPII mediated export of large cargo

Using AlphaFold, we were able to illustrate how alternative splicing of exon 24c in SEC31A could regulate lipid transport. Exon 24c in SEC31A lies in a disorganized proline-rich region,

complicating the estimation of structural change. The WND motive is located five amino acids downstream of exon 24c. This is the active motive of SEC31A, which enhances the SAR1 GTPase activity by correcting the histidine 79 position in SAR1 and shielding it from solvents (Bi et al. 2007). Bi et al. solved a crystal structure of the yeast homologs of SAR1 and SEC23 and an active fragment of SEC31. This active fragment threads itself between SAR1 and SEC23 and subsequently enhances the GTPase activity of SAR1. Exon 24c would be situated precisely in the region of this active fragment, in the vicinity of the WND motive, if it would also be conserved in yeast. Applying AlphaFold on this multimer of human SAR1A, SEC23A, and the alternatively spliced active fragment of SEC31A, we anticipated a change in the particular association of the WND motive with SAR1A and SEC23A. But we could not observe a change in this interplay in the predicted structure of the exon 24c, including isoform 8. As we have no data showing a changed or unchanged GTPase activity of SAR1 with the SEC31A active fragment, we cannot rule out if the alternative splicing of SEC31A impacts the GTPase activity. We can state that the interaction between the active fragment of SEC31A and SEC23A in the AlphaFold prediction is increased by four amino acids (Figure 17C), and the orientation of the active pocket of SAR1A is unchanged. Through this interaction, we infer that exon 24c regulates the lipid secretion by alternative splicing of SEC31A exon 24c regulates the lipid secretion.

The established model of the secretion via COPII vesicles has been challenged through the advancement of microscope technologies. New data suggests that transport from the ER to the Golgi is facilitated by tubule structures extending from the ERES (Weigel et al. 2021). COPII opens the neck to the ERES, and COPI is responsible for forming the tubule structure of the ER. That does not mean that the results of the last 30 years of science done on COPII was done in vain, but it needs to be reevaluated with the new model. Most regulatory effects of the COPII elements that were previously reported are valid but not necessarily in the context of vesicles. COPII controls the stability of the ERES neck and regulates the cargo flux. With varying neck sizes, COPII probably changes the efficiency of the secretion of big cargo. With the increased interaction between SEC23A and SEC31A, enhanced by alternative splicing of exon 24c, neck stability could be improved and positively regulate the secretion of chylomicrons. With that, we can build the following model: Secretion starts with the scaffolding of SEC16 and building the ERES, which recruits SEC12. SEC12 acts as the GEF for SAR1; therefore, it recruits the GTPase to the ERES and facilitates the exchange of SAR1-bound GDP with GTP. That leads to the extension of the amphipathic helix of SAR1 and its insertion into the ER membrane. The membrane subsequently deforms and bends to shapes like vesicles or the ERES neck. This bending recruits the inner COPII shell, SEC23, and followed by SEC24, which starts recruiting cargo through direct or indirect binding.

SEC23 acts as the GAP for SAR1 and, as a result, enhances its GTPase activity. The complex of SEC23 and SAR1 then recruits SEC31, and SEC13, subsequently forming the outer layer of the COPII. The insertion of the WND of SEC31 in the active pocket of SAR1 additionally increases its GTPase activity and leads to the hydrolyzation of GTP. This induces the amphipathic helix of SAR1 to retrieve and dissociate from the ER membrane. This usually causes COPII vesicle fissure and the ERES neck to collapse. We propose that the increased interaction between SEC31A and SEC23A, caused by the alternative inclusion of exon 24c, could increase the ERES neck's stability and its long-term sustainability. That could cause big cargo such as chylomicrons to be secreted more efficiently. The interaction between SEC31 and SEC23 was examined previously to increase the secretion efficiency of big cargo (Townley et al. 2008). This could help explain our findings in a COPII vesicle model as well.

The mutation F382L in SEC23A causes Cranio-lenticulo-sutural dysplasia (Boyadjiev et al. 2006). The missense mutation M702V was also reported to have similar clinical manifestations (Boyadjiev et al. 2011). Cranio-lenticulo-sutural dyplasia is an extremely rare disease; in 2006, only two families were reported to show that phenotype. A recent study discovered another patient with similar symptoms, with a predicted E599K mutation in SEC23A (Cisarova et al. 2022). All those mutations in SEC23A have in common that they secure the position of the α-helix, where Q387 and K392 are positioned. AlphaFold predicts these residues to interact with the exon 24c encoded amino acids in SEC31A (Figure 3-11C, D). It is suggested that cranio-lenticulo-sutural dysplasia is caused by a weakening of the interaction between SEC23A and SEC31 (Townley et al. 2008). Patients with this autosomal recessive syndrome show large fontanels, Y-shape cataracts, skeletal defects, and other facial dysmorphisms. They might be caused by a deficiency of secretion, especially of collagen. Cultured fibroblasts of patients also show an enlarged ER (Boyadjiev et al. 2011). This can be directly connected to the regulatory effect of exon 24c alternative splicing, as it also probably increases the interaction between SEC23A and SEC31A. This gives credence to how the importance of this interaction is and what effect it can have on secretion.

4.4 Exon 24b exclusion enhances cargo secretion, while reducing overall secretion

Secretion of collagen is an intensively studied mechanism. Past models of the COPII dictate a vesicle size of 90 nm, which is much too small for pro-collagen trimers that are produced in the ER. Multiple studies have been conducted connecting collagen secretion to SEC23 (Boyadjiev et al. 2006; Townley et al. 2008; Kim et al. 2012), SAR1 (Saito et al. 2017; Bi et al. 2007; Kasberg et al. 2023) and TANGO1/cTAGE5 (Saito et al. 2011; Raote et al. 2017; Wilson et al. 2011; Bunel et al. 2024). These are all large mechanistic studies, which are

essential for our understanding of the secretion of collagen. Our work shows an association between alternative splicing and the secretion of collagen. This adds a much-nuanced fine-tuned regulation into this field of research.

For reducing of the inclusion of exon 24b, we used a catalytic dead form of CasRX (Konermann et al. 2018). Like in the previous study with MO (Neumann et al. 2019), we could sufficiently reduce the inclusion of exon 24b (Figure 3-5B). In the same way, we observe the general reduction of secretion efficiency (Figure 3-5C-D). But in contrast to the previous study, we could show that the alternative splicing of exon 24b regulates the efficiency of secretion of some proteins differently. Collagen secretion is enhanced with the absence of exon 24b (Figure 3-7B). That alternative spicing could regulate the specificity of cargo in secretion was not previously reported. The correlation analysis also indicated the connection of exon 24b induced collagen secretion, as a group of collagens negatively correlated with the inclusion of the exon. We adapted the assay we used to measure collagen secretion from a publication by Satio et al. (Saito et al. 2011). We also used the siRNA against MIA2 from this work. The raw fluorescent signal from the accumulated collagen was measured. With our adaptation of staining the extracellular collagen before staining the total collagen, we transformed this into a much more quantifiable method. MIA2 knockdown assured the validity of this assay. The FRAP does provide an explanation for mechanisms on a molecular level. The recovery rate is not changed, which leads to the assumption that enzymatic reactions are not altered by the skipping of exon 24b. However, the fraction of the mobile phase is significantly reduced by approximately 20%. The mobile phase is an indicator of unmovable molecules, in this case, SEC13-GFP, that do not get naturally exchanged after being photobleached. That usually means that the molecule of interest is bound to partners, scaffolding, or is otherwise immovable. In our approach, we measured the mobility of SEC13 and influenced the splicing of exon 24b in SEC31A. This means we can indirectly change the behavior of SEC13 through SEC31A. With this approach, we can at least say that alternative splicing indeed can change the mobility of the outer COPII coat, and we will eventually even be able to make predictions about the whole COPII function. In summary, we propose that alternative splicing of SEC31A exon 24b enhances the stability of COPII, demonstrated by the low mobile fraction in the FRAP experiment with reduced exon 24b inclusion. This could be an additional specification for the secretion of collagens, as seen with TANGO1/cTAGE5. Here, collagen secretion was presumed to be mediated through tubule structures held by TANGO5/cTAGE5 and, consequently, COPII. This effect could be enhanced when exon 24b is excluded, leading to a higher stability of COPII.

4.5 Gene expression analysis in B-cell differentiation reveals strong regulation of secretion and splicing

Many cellular processes in life influence the cell's behavior. We are foremost invested in how alternative splicing can regulate secretion specificity. T-cell-dependent B-cell activation is one cellular process that has huge implications for secretion. We used a publicly available RNA-seq data set that includes MBC, preBP, BP, and PC, and enabled us to investigate the relationship between the transcriptome and secretion (Kassambara et al. 2021). The ASC, which includes PB and PC, are highly specialized to secrete vast quantities of ab (Nutt et al. 2015). MBC, however, shows little secretory potential. ASC morphology, metabolism, and gene expression are adapted specifically for the secretion (Nguyen et al. 2019). They do that, by expanding their ER and Golgi, so high amounts of ab can be secreted (Kirk et al. 2010) Alternative splicing is known to affect B-cell differentiation (Grigoryev et al. 2009; Martinez and Lynch 2013).

In the data set we found changes in gene expression and alternative splicing, which could potentially have an influence on secretion and intracellular trafficking. Gene expression analysis shows distinct behavior of the four analyzed B-cell differentiation stages (Figure 3-14A). It is obvious that MBCs have a particular gene expression that differs strongly from the other differentiation stages. PB and PC are very similar in their gene expression, as they have very similar tasks in the cell, and prePB shows a gene expression pattern that matches its role as a precursor for the ASCs. It is very striking that most of the genes in the GO term “protein exit from ER” are significantly upregulated in PB, and none of them are upregulated in MBC (Figure 3-14E). Similar findings can be seen compared with activated T-cells (Figure 3-14F). T-cells and B-cells seem to handle the higher secretion load differently. Whereas T-cells do not increase the gene expression of COPII members, PB has a much higher gene expression change than MBC. These two lymphocyte species seem to engage different approaches when it comes to specialization into highly secreting cells. Many members of COPII are upregulated in PB when compared with MBC, including SEC24A, SEC24C, and SEC24B, among others. Only SEC24B is significantly upregulated in MBC. That is most peculiar, as SEC24 is the cargo binder of the COPII and actively controls the passing of specific cargos (Miller et al. 2002; Adolf et al. 2016; Wang et al. 2018). The differential secretion of SEC24 between MBC and PB could indicate that these genes are responsible for specificity in cargo secretion. One exciting facet of the gene expression is the occurrence of the GO terms “mRNA processing” and “RNA splicing” in a gene enrichment analysis between MBC and PB (Figure 3-14B, D). This is not surprising, as alternative splicing is known for being an essential regulator for B-cell differentiation (Kassambara et al. 2021), but it also enables us to continue with alternative splicing analysis in this direction confidently.

4.6 Alternative splicing of secretory genes in B-cell differentiation helps understanding adaptation of MBC

We found that alternative splicing is very distinct in MBC compared to the other three analyzed stages of B-cell differentiation (Figure 3-15A). Alternative splicing seems to be a driving force for MBC differentiation and shows little to no difference between prePB, PB, and PC. Not surprisingly, almost all found alternatively spliced events found have a distinct MBC isoform that differs from the other B-cell differentiation stages. As COPII is essential for the export of ab from the ER (Kirk et al. 2010) we decided to investigate the members of that complex. Some of them probably induce NMD through exon length that is not divisible by 3. This, in return, explains the reduced gene expression levels in MBC. SEC16 is also alternatively spliced in MBC in a way that was already described (Wilhelmi et al. 2016), which works similarly to activated T-cells.

Ultimately, we found two genes with alternative splice events, which were striking and worth a closer look. The first was KLHL12, with its impaired kelch domain caused by the skipping of exons 10 and 11 in MBC (Figure 3-16A-D). The kelch domain is responsible for the interaction with its reaction partners (Chen et al. 2020b). This isoform could, when translated into protein, have a robust regulatory function on secretion by lowering the effectiveness of monoubiquitylation of SEC31 (Akopian et al. 2022). The presence of the uniport peptide Q9H7R2, which shows half of the kelch domain, points to the possibility of protein-coding of the aforementioned isoforms. Ubiquitylation of SEC31 enables big cargo like collagens to be secreted, and as MBC does not need the capacity of high specialized secretion, this could be downregulated by the alternative splicing of KLHL12. The second gene with interesting alternative splice events is PICALM, which is linked to the endocytic intake in cells (Miller et al. 2015). The skipping of exons 13 and 17a was already described in a publication examining the relation of PICALM with Alzheimer's disease (Parikh et al. 2014). In the peptide that exon 13 encodes, a DPF motive can be found. This is one of four sites that contribute to the binding of PICALM to AP2 (Meyerholz et al. 2005). This binding was associated with the regulation of the size of clathrin-coated vesicles. The function of exon 17a was investigated, but no connection to AP2 or clathrin was found (Scotland et al. 2012). The resulting short C-terminus could be a decrease of clathrin triskelia, as binding sites are flanking exon 17a. Whereas secretion has one of the highest priorities in ASCs, it is possible that MBC relies more on endocytosis, as signals for eventual differentiation to PC could be presented to them at any time. The alternative splicing of PICALM could be a mechanism in which the clathrin-dependent endocytosis could be enhanced, which can be such an adaptation.

4.7 Potential transcription factor of B-cell differentiation is revealed by in silico study

The last investigation, which is needed, is to figure out which transcription factor could be involved in the secretory genes in the B-cell differentiation process. Many TF are known to be crucial for B-cell development, like Pax5, PU.1 and E2A (Hagman and Lukin 2006; Fischer et al. 2020), we were investigated, if an transcription factor could explain the gene expression of secretory genes in B-cell differentiation. Our study determined that the secretory genes SEC24b, CUL3, and CSNK1D are upregulated in MBC. With the correlation analysis, we found that NERF/ELF2 does correlate with all three genes in the B-cell differentiation data set and the tissue data set (Figure 3-17A). That was quite interesting, as NERF/ELF2 is in itself heavily spliced in the B-cell differentiation stages (Figure 3-17D). The splice isoforms are known to have vastly different responsibilities. The isoform NERF-2 is known for activation of target genes, and NERF-1 is a repressor (Cho et al. 2004). Furthermore, NERF-1 induces apoptosis and reduces overall cell proliferation, whereas NERF-2 generates the opposing effect (Guan et al. 2017). The same study reveals that the isoform ratio of NERF/ELF2 changes between CD4+ and CD8+ T-cells and B-cell maturation. It is apparent that MBCs have a much longer lifespan than the ASCs. The apoptosis-reducing isoform NERF-2 is in higher abundance in MBC than in ASC. It is that the NERF-2 isoform helps the long-lived MBC reduce apoptosis. Additionally, the general higher gene expression of NERF/ELF2 could be responsible for the gene expression of secretory genes like SEC24B, CUL3, and CSNK1D. These are all results of an in silico study without any wet lab evidence. The findings of gene expression, splicing switches, and NERF/ELF2 influence on gene B-cell differentiation need to be experimentally evaluated.

5 References

Abrams, Elliott W.; Andrew, Deborah J. (2005): CrebA regulates secretory activity in the Drosophila salivary gland and epidermis. In *Development* 132 (12), pp. 2743–2758. DOI: 10.1242/dev.01863.

Adolf, Frank; Rhiel, Manuel; Reckmann, Ingeborg; Wieland, Felix T. (2016): Sec24C/D-isoform-specific sorting of the preassembled ER-Golgi Q-SNARE complex. In *Molecular biology of the cell* 27 (17), pp. 2697–2707. DOI: 10.1091/mbc.E16-04-0229.

Akopian, David; McGourty, Colleen A.; Rapé, Michael (2022): Co-adaptor driven assembly of a CUL3 E3 ligase complex. In *Molecular Cell* 82 (3), 585-597.e11. DOI: 10.1016/j.molcel.2022.01.004.

Alaterre, Elina; Vikova, Veronika; Kassambara, Alboukadel; Bruyer, Angélique; Robert, Nicolas; Requirand, Guilhem et al. (2021): RNA-Sequencing-Based Transcriptomic Score with Prognostic and Theranostic Values in Multiple Myeloma. In *Journal of Personalized Medicine* 11 (10), p. 988. DOI: 10.3390/jpm11100988.

Amaral, Paulo; Carbonell-Sala, Silvia; La Vega, Francisco M. de; Faial, Tiago; Frankish, Adam; Gingeras, Thomas et al. (2023): The status of the human gene catalogue. In *Nature* 622 (7981), pp. 41–47. DOI: 10.1038/s41586-023-06490-x.

Anders, Simon; Huber, Wolfgang (2010): Differential expression analysis for sequence count data. In *Genome Biol* 11 (10), R106. DOI: 10.1186/gb-2010-11-10-r106.

Anders, Simon; Pyl, Paul Theodor; Huber, Wolfgang (2015): HTSeq--a Python framework to work with high-throughput sequencing data. In *Bioinformatics (Oxford, England)* 31 (2), pp. 166–169. DOI: 10.1093/bioinformatics/btu638.

Anders, Simon; Reyes, Alejandro; Huber, Wolfgang (2012): Detecting differential usage of exons from RNA-seq data. In *Genome Res.* 22 (10), pp. 2008–2017. DOI: 10.1101/gr.133744.111.

Änkö, Minna-Liisa (2014): Regulation of gene expression programmes by serine-arginine rich splicing factors. In *Seminars in Cell & Developmental Biology* 32, pp. 11–21. DOI: 10.1016/j.semcdb.2014.03.011.

Antonny, B.; Madden, D.; Hamamoto, S.; Orci, L.; Schekman, R. (2001): Dynamics of the COPII coat with GTP and stable analogues. In *Nature cell biology* 3 (6), pp. 531–537. DOI: 10.1038/35078500.

Appenzeller-Herzog, Christian; Hauri, Hans-Peter (2006): The ER-Golgi intermediate

compartment (ERGIC): in search of its identity and function. In *Journal of cell science* 119 (Pt 11), pp. 2173–2183. DOI: 10.1242/jcs.03019.

Ariel, Federico D.; Manavella, Pablo A. (2021): When junk DNA turns functional: transposon-derived non-coding RNAs in plants. In *Journal of experimental botany* 72 (11), pp. 4132–4143. DOI: 10.1093/jxb/erab073.

Barbosa-Morais, Nuno L.; Irimia, Manuel; Pan, Qun; Xiong, Hui Y.; Gueroussov, Serge; Lee, Leo J. et al. (2012): The evolutionary landscape of alternative splicing in vertebrate species. In *Science (New York, N.Y.)* 338 (6114), pp. 1587–1593. DOI: 10.1126/science.1230612.

Barlowe, C.; Orci, L.; Yeung, T.; Hosobuchi, M.; Hamamoto, S.; Salama, N. et al. (1994): COPII: a membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum. In *Cell* 77 (6), pp. 895–907. DOI: 10.1016/0092-8674(94)90138-4.

Bennett, Loris; Melchers, Bernd; Proppe, Boris (2020): Curta: A General-purpose High-Performance Computer at ZEDAT, Freie Universität Berlin. With assistance of Universitätsbibliothek der FU Berlin.

Bharucha, Nike; Liu, Yang; Papanikou, Effrosyni; McMahon, Conor; Esaki, Masatoshi; Jeffrey, Philip D. et al. (2013): Sec16 influences transitional ER sites by regulating rather than organizing COPII. In *Molecular biology of the cell* 24 (21), pp. 3406–3419. DOI: 10.1091/mbc.E13-04-0185.

Bi, Xiping; Corpina, Richard A.; Goldberg, Jonathan (2002): Structure of the Sec23/24-Sar1 pre-budding complex of the COPII vesicle coat. In *Nature* 419 (6904), pp. 271–277. DOI: 10.1038/nature01040.

Bi, Xiping; Mancias, Joseph D.; Goldberg, Jonathan (2007): Insights into COPII coat nucleation from the structure of Sec23.Sar1 complexed with the active fragment of Sec31. In *Developmental Cell* 13 (5), pp. 635–645. DOI: 10.1016/j.devcel.2007.10.006.

Bielli, Anna; Haney, Charles J.; Gabreski, Gavin; Watkins, Simon C.; Bannykh, Sergei I.; Aridor, Meir (2005): Regulation of Sar1 NH2 terminus by GTP binding and hydrolysis promotes membrane deformation to control COPII vesicle fission. In *J Cell Biol* 171 (6), pp. 919–924. DOI: 10.1083/jcb.200509095.

Boncompain, Gaelle; Divoux, Severine; Gareil, Nelly; Forges, Helene de; Lescure, Aurianne; Latreche, Lynda et al. (2012): Synchronization of secretory protein traffic in populations of cells. In *Nat Methods* 9 (5), pp. 493–498. DOI: 10.1038/nmeth.1928.

Bortoli, Francesca de; Espinosa, Sara; Zhao, Rui (2021): DEAH-Box RNA Helicases in Pre-mRNA Splicing. In *Trends in biochemical sciences* 46 (3), pp. 225–238. DOI:

10.1016/j.tibs.2020.10.006.

Bottanelli, Francesca; Kilian, Nicole; Ernst, Andreas M.; Rivera-Molina, Felix; Schroeder, Lena K.; Kromann, Emil B. et al. (2017): A novel physiological role for ARF1 in the formation of bidirectional tubules from the Golgi. In *Molecular biology of the cell* 28 (12), pp. 1676–1687. DOI: 10.1091/mbc.E16-12-0863.

Boukas, Leandros; Havrilla, James M.; Hickey, Peter F.; Quinlan, Aaron R.; Bjornsson, Hans T.; Hansen, Kasper D. (2019): Coexpression patterns define epigenetic regulators associated with neurological dysfunction. In *Genome research* 29 (4), pp. 532–542. DOI: 10.1101/gr.239442.118.

Boyadjiev, Simeon A.; Fromme, J. Christopher; Ben, Jin; Chong, Samuel S.; Nauta, Christopher; Hur, David J. et al. (2006): Cranio-lenticulo-sutural dysplasia is caused by a SEC23A mutation leading to abnormal endoplasmic-reticulum-to-Golgi trafficking. In *Nature genetics* 38 (10), pp. 1192–1197. DOI: 10.1038/ng1876.

Boyadjiev, Simeon A.; Kim, S-D; Hata, A.; Haldeman-Englert, C.; Zackai, E. H.; Naydenov, C. et al. (2011): Cranio-lenticulo-sutural dysplasia associated with defects in collagen secretion. In *Clinical genetics* 80 (2), pp. 169–176. DOI: 10.1111/j.1399-0004.2010.01550.x.

Brow, David A. (2002): Allosteric cascade of spliceosome activation. In *Annual review of genetics* 36, pp. 333–360. DOI: 10.1146/annurev.genet.36.043002.091635.

Bryant, David M.; Mostov, Keith E. (2008): From cells to organs: building polarized tissue. In *Nat Rev Mol Cell Biol* 9 (11), pp. 887–901. DOI: 10.1038/nrm2523.

Bullard, James H.; Purdom, Elizabeth; Hansen, Kasper D.; Dudoit, Sandrine (2010): Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. In *BMC Bioinformatics* 11 (1), p. 94. DOI: 10.1186/1471-2105-11-94.

Bunel, Louis; Pincet, Lancelot; Malhotra, Vivek; Raote, Ishier; Pincet, Frédéric (2024): A model for collagen secretion by intercompartmental continuities. In *Proceedings of the National Academy of Sciences of the United States of America* 121 (1), e2310404120. DOI: 10.1073/pnas.2310404120.

Cendrowski, Jaroslaw; Kaczmarek, Marta; Mazur, Michał; Kuzmicz-Kowalska, Katarzyna; Jastrzebski, Kamil; Brewinska-Olchowik, Marta et al. (2020): Splicing variation of BMP2K balances abundance of COPII assemblies and autophagic degradation in erythroid cells. In *eLife* 9. DOI: 10.7554/eLife.58504.

Chaffer, Christine L.; Thompson, Erik W.; Williams, Elizabeth D. (2007): Mesenchymal to epithelial transition in development and disease. In *Cells Tissues Organs* 185 (1-3), pp. 7–

19. DOI: 10.1159/000101298.

Chang, Howard Y.; Qi, Lei S. (2023): Reversing the Central Dogma: RNA-guided control of DNA in epigenetics and genome editing. In *Molecular Cell* 83 (3), pp. 442–451. DOI: 10.1016/j.molcel.2023.01.010.

Chen, Hao; Chong, Wei; Yang, Xiaorong; Zhang, Yuan; Sang, Shaowei; Li, Xiangchun; Lu, Ming (2020a): Age-related mutational signature negatively associated with immune activity and survival outcome in triple-negative breast cancer. In *Oncoimmunology* 9 (1), p. 1788252. DOI: 10.1080/2162402X.2020.1788252.

Chen, Mo; Manley, James L. (2009): Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. In *Nat Rev Mol Cell Biol* 10 (11), pp. 741–754. DOI: 10.1038/nrm2777.

Chen, Yu; Gershlick, David C.; Park, Sang Yoon; Bonifacino, Juan S. (2017): Segregation in the Golgi complex precedes export of endolysosomal proteins in distinct transport carriers. In *The Journal of cell biology* 216 (12), pp. 4141–4151. DOI: 10.1083/jcb.201707172.

Chen, Zhuoyao; Wasney, Gregory A.; Picaud, Sarah; Filippakopoulos, Panagis; Vedadi, Masoud; D'Angiolella, Vincenzo; Bullock, Alex N. (2020b): Identification of a PGXPP degron motif in dishevelled and structural basis for its binding to the E3 ligase KLHL12. In *Open biology* 10 (6), p. 200041. DOI: 10.1098/rsob.200041.

Cho, Je-Yoel; Akbarali, Yasmin; Zerbini, Luiz F.; Gu, Xuesong; Boltax, Jay; Wang, Yihong et al. (2004): Isoforms of the Ets transcription factor NERF/ELF-2 physically interact with AML1 and mediate opposing effects on AML1-mediated transcription of the B cell-specific blk gene. In *Journal of Biological Chemistry* 279 (19), pp. 19512–19522. DOI: 10.1074/jbc.M309074200.

Cho, Kyung-Ah; Lee, Jun-Kyu; Kim, Yu-Hee; Park, Minhwa; Woo, So-Youn; Ryu, Kyung-Ha (2017): Mesenchymal stem cells ameliorate B-cell-mediated immune responses and increase IL-10-expressing regulatory B cells in an EBI3-dependent manner. In *Cell Mol Immunol* 14 (11), pp. 895–908. DOI: 10.1038/cmi.2016.59.

Cisarova, Katarina; Garavelli, Livia; Caraffi, Stefano Giuseppe; Peluso, Francesca; Valeri, Lara; Gargano, Giancarlo et al. (2022): A monoallelic SEC23A variant E599K associated with cranio-lenticulo-sutural dysplasia. In *American journal of medical genetics. Part A* 188 (1), pp. 319–325. DOI: 10.1002/ajmg.a.62506.

Cobaleda, César; Schebesta, Alexandra; Delogu, Alessio; Busslinger, Meinrad (2007): Pax5: the guardian of B cell identity and function. In *Nature immunology* 8 (5), pp. 463–470. DOI: 10.1038/ni1454.

Crick, F. (1970): Central dogma of molecular biology. In *Nature* 227 (5258), pp. 561–563. DOI: 10.1038/227561a0.

Crow, Megan; Paul, Anirban; Ballouz, Sara; Huang, Z. Josh; Gillis, Jesse (2016): Exploiting single-cell expression to characterize co-expression replicability. In *Genome biology* 17 (1), p. 101. DOI: 10.1186/s13059-016-0964-6.

Cutrona, Meritxell B.; Beznoussenko, Galina V.; Fusella, Aurora; Martella, Oliviano; Moral, Pedro; Mironov, Alexander A. (2013): Silencing of mammalian Sar1 isoforms reveals COPII-independent protein sorting and transport. In *Traffic* 14 (6), pp. 691–708. DOI: 10.1111/tra.12060.

Danecek, Petr; Bonfield, James K.; Liddle, Jennifer; Marshall, John; Ohan, Valeriu; Pollard, Martin O. et al. (2021): Twelve years of SAMtools and BCFtools. In *GigaScience* 10 (2). DOI: 10.1093/gigascience/giab008.

Darnell, Robert B. (2013): RNA protein interaction in neurons. In *Annual review of neuroscience* 36, pp. 243–270. DOI: 10.1146/annurev-neuro-062912-114322.

Dobin, Alexander; Davis, Carrie A.; Schlesinger, Felix; Drenkow, Jorg; Zaleski, Chris; Jha, Sonali et al. (2013): STAR: ultrafast universal RNA-seq aligner. In *Bioinformatics (Oxford, England)* 29 (1), pp. 15–21. DOI: 10.1093/bioinformatics/bts635.

Drewe, Philipp; Stegle, Oliver; Hartmann, Lisa; Kahles, André; Bohnert, Regina; Wachter, Andreas et al. (2013): Accurate detection of differential RNA processing. In *Nucleic Acids Res* 41 (10), pp. 5189–5198. DOI: 10.1093/nar/gkt211.

Fagerberg, Linn; Hallström, Björn M.; Oksvold, Per; Kampf, Caroline; Djureinovic, Dijana; Odeberg, Jacob et al. (2014): Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. In *Molecular & cellular proteomics : MCP* 13 (2), pp. 397–406. DOI: 10.1074/mcp.M113.035600.

Fischer, Ute; Yang, Jun J.; Ikawa, Tomokatsu; Hein, Daniel; Vicente-Dueñas, Carolina; Borkhardt, Arndt; Sánchez-García, Isidro (2020): Cell Fate Decisions: The Role of Transcription Factors in Early B-cell Development and Leukemia. In *Blood Cancer Discov* 1 (3), pp. 224–233. DOI: 10.1158/2643-3230.BCD-20-0011.

Fogh, J.; Fogh, J. M.; Orfeo, T. (1977): One hundred and twenty-seven cultured human tumor cell lines producing tumors in nude mice. In *Journal of the National Cancer Institute* 59 (1), pp. 221–226. DOI: 10.1093/jnci/59.1.221.

Garbes, Lutz; Kim, Kyungho; Rieß, Angelika; Hoyer-Kuhn, Heike; Beleggia, Filippo; Bevot, Andrea et al. (2015): Mutations in SEC24D, encoding a component of the COPII machinery,

cause a syndromic form of osteogenesis imperfecta. In *American journal of human genetics* 96 (3), pp. 432–439. DOI: 10.1016/j.ajhg.2015.01.002.

Ghazalpour, Anatole; Doss, Sudheer; Zhang, Bin; Wang, Susanna; Plaisier, Christopher; Castellanos, Ruth et al. (2006): Integrating genetic and network analysis to characterize genes related to mouse weight. In *PLoS genetics* 2 (8), e130. DOI: 10.1371/journal.pgen.0020130.

Glahn, Raymond P. (2022): The Caco-2 Cell Bioassay for Measurement of Food Iron Bioavailability. In *JoVE (Journal of Visualized Experiments)* (182), e63859. DOI: 10.3791/63859.

Grieder, Nicole C.; Caussinus, Emmanuel; Parker, David S.; Cadigan, Kenneth; Affolter, Markus; Luschnig, Stefan (2008): gammaCOP is required for apical protein secretion and epithelial morphogenesis in Drosophila melanogaster. In *PloS one* 3 (9), e3241. DOI: 10.1371/journal.pone.0003241.

Grigoryev, Yevgeniy A.; Kurian, Sunil M.; Nakorchevskiy, Aleksey A.; Burke, John P.; Campbell, Daniel; Head, Steve R. et al. (2009): Genome-wide analysis of immune activation in human T and B cells reveals distinct classes of alternatively spliced genes. In *PloS one* 4 (11), e7906. DOI: 10.1371/journal.pone.0007906.

Guan, Fiona H. X.; Bailey, Charles G.; Metierre, Cynthia; O'Young, Patrick; Gao, Dadi; Khoo, Teh Liane et al. (2017): The antiproliferative ELF2 isoform, ELF2B, induces apoptosis in vitro and perturbs early lymphocytic development in vivo. In *Journal of hematology & oncology* 10 (1), p. 75. DOI: 10.1186/s13045-017-0446-7.

Hagman, James; Lukin, Kara (2006): Transcription factors drive B cell development. In *Current Opinion in Immunology* 18 (2), pp. 127–134. DOI: 10.1016/j.coi.2006.01.007.

Hanahan, Douglas; Weinberg, Robert A. (2011): Hallmarks of cancer: the next generation. In *Cell* 144 (5), pp. 646–674. DOI: 10.1016/j.cell.2011.02.013.

Harris, Charles R.; Millman, K. Jarrod; van der Walt, Stéfan J.; Gommers, Ralf; Virtanen, Pauli; Cournapeau, David et al. (2020): Array programming with NumPy. In *Nature* 585 (7825), pp. 357–362. DOI: 10.1038/s41586-020-2649-2.

Head, Steven R.; Komori, H. Kiyomi; LaMere, Sarah A.; Whisenant, Thomas; van Nieuwerburgh, Filip; Salomon, Daniel R.; Ordoukhanian, Phillip (2014): Library construction for next-generation sequencing: overviews and challenges. In *BioTechniques* 56 (2), 61-4, 66, 68, passim. DOI: 10.2144/000114133.

Heine, Guido; Drozdenko, Gennadiy; Grün, Joachim R.; Chang, Hyun-Dong; Radbruch,

Andreas; Worm, Margitta (2014): Autocrine IL-10 promotes human B-cell differentiation into IgM- or IgG-secreting plasmablasts. In *European journal of immunology* 44 (6), pp. 1615–1621. DOI: 10.1002/eji.201343822.

Hnilicová, Jarmila; Staněk, David (2011): Where splicing joins chromatin. In *Nucleus (Austin, Tex.)* 2 (3), pp. 182–188. DOI: 10.4161/nucl.2.3.15876.

Hogg, Rebecca; McGrail, Joanne C.; O'Keefe, Raymond T. (2010): The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing. In *Biochemical Society transactions* 38 (4), pp. 1110–1115. DOI: 10.1042/BST0381110.

Huang, M.; Weissman, J. T.; Beraud-Dufour, S.; Luan, P.; Wang, C.; Chen, W. et al. (2001): Crystal structure of Sar1-GDP at 1.7 A resolution and the role of the NH2 terminus in ER export. In *J Cell Biol* 155 (6), pp. 937–948. DOI: 10.1083/jcb.200106039.

Hubatsch, Ina; Ragnarsson, Eva G. E.; Artursson, Per (2007): Determination of drug permeability and prediction of drug absorption in Caco-2 monolayers. In *Nat Protoc* 2 (9), pp. 2111–2119. DOI: 10.1038/nprot.2007.303.

Hui, N.; Nakamura, N.; Sönnichsen, B.; Shima, D. T.; Nilsson, T.; Warren, G. (1997): An isoform of the Golgi t-SNARE, syntaxin 5, with an endoplasmic reticulum retrieval signal. In *Molecular biology of the cell* 8 (9), pp. 1777–1787. DOI: 10.1091/mbc.8.9.1777.

Hutchings, Joshua; Stancheva, Viktoriya; Miller, Elizabeth A.; Zanetti, Giulia (2018): Subtomogram averaging of COPII assemblies reveals how coat organization dictates membrane shape. In *Nature communications* 9 (1), p. 4154. DOI: 10.1038/s41467-018-06577-4.

Hutchings, Joshua; Stancheva, Viktoriya G.; Brown, Nick R.; Cheung, Alan C. M.; Miller, Elizabeth A.; Zanetti, Giulia (2021): Structure of the complete, membrane-assembled COPII coat reveals a complex interaction network. In *Nature communications* 12 (1), p. 2034. DOI: 10.1038/s41467-021-22110-6.

Jin, Lingyan; Pahuja, Kanika Bajaj; Wickliffe, Katherine E.; Gorur, Amita; Baumgärtel, Christine; Schekman, Randy; Rape, Michael (2012): Ubiquitin-dependent regulation of COPII coat size and function. In *Nature* 482 (7386), pp. 495–500. DOI: 10.1038/nature10822.

Jones, Bethan; Jones, Emma L.; Bonney, Stephanie A.; Patel, Hetal N.; Mensenkamp, Arjen R.; Eichenbaum-Voline, Sophie et al. (2003): Mutations in a Sar1 GTPase of COPII vesicles are associated with lipid absorption disorders. In *Nat Genet* 34 (1), pp. 29–31. DOI: 10.1038/ng1145.

Jumper, John; Evans, Richard; Pritzel, Alexander; Green, Tim; Figurnov, Michael; Ronneberger, Olaf et al. (2021): Highly accurate protein structure prediction with AlphaFold. In *Nature* 596 (7873), pp. 583–589. DOI: 10.1038/s41586-021-03819-2.

Kasberg, William; Luong, Peter; Hanna, Michael G.; Minushkin, Kayla; Tsao, Annabelle; Shankar, Raakhee et al. (2023): The Sar1 GTPase is dispensable for COPII-dependent cargo export from the ER. In *Cell reports* 42 (6), p. 112635. DOI: 10.1016/j.celrep.2023.112635.

Kassambara, Alboukadel; Herviou, Laurie; Ovejero, Sara; Jourdan, Michel; Thibaut, Coraline; Vikova, Veronika et al. (2021): RNA-sequencing data-driven dissection of human plasma cell differentiation reveals new potential transcription regulators. In *Leukemia* 35 (5), pp. 1451–1462. DOI: 10.1038/s41375-021-01234-0.

Kim, Sun-Don; Pahuja, Kanika Bajaj; Ravazzola, Mariella; Yoon, Joonsik; Boyadjiev, Simeon A.; Hammamoto, Susan et al. (2012): The corrected SEC23-SEC31 corrected interface plays critical role for export of procollagen from the endoplasmic reticulum. In *The Journal of biological chemistry* 287 (13), pp. 10134–10144. DOI: 10.1074/jbc.M111.283382.

Kim, Yong-Eun; Won, Minho; Lee, Sung-Gwon; Park, Chungoo; Song, Chang-Hwa; Kim, Kee K. (2019): RBM47-regulated alternative splicing of TJP1 promotes actin stress fiber assembly during epithelial-to-mesenchymal transition. In *Oncogene* 38 (38), pp. 6521–6536. DOI: 10.1038/s41388-019-0892-5.

Kirk, Semra J.; Cliff, Jacqueline M.; Thomas, J. Alero; Ward, Theresa H. (2010): Biogenesis of secretory organelles during B cell differentiation. In *Journal of leukocyte biology* 87 (2), pp. 245–255. DOI: 10.1189/jlb.1208774.

Kometani, Kohei; Nakagawa, Rinako; Shinnakasu, Ryo; Kaji, Tomohiro; Rybouchkin, Andrei; Moriyama, Saya et al. (2013): Repression of the transcription factor Bach2 contributes to predisposition of IgG1 memory B cells toward plasma cell differentiation. In *Immunity* 39 (1), pp. 136–147. DOI: 10.1016/j.immuni.2013.06.011.

Konermann, Silvana; Lotfy, Peter; Brideau, Nicholas J.; Oki, Jennifer; Shokhirev, Maxim N.; Hsu, Patrick D. (2018): Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. In *Cell* 173 (3), 665-676.e14. DOI: 10.1016/j.cell.2018.02.033.

Koreishi, M.; Yu, S.; Oda, M.; Honjo, Y.; Satoh, A. (2013): CK2 phosphorylates Sec31 and regulates ER-To-Golgi trafficking. In *PloS one* 8 (1). DOI: 10.1371/journal.pone.0054382.

Kurosaki, Tatsuaki; Popp, Maximilian W.; Maquat, Lynne E. (2019): Quality and quantity control of gene expression by nonsense-mediated mRNA decay. In *Nat Rev Mol Cell Biol* 20 (7), pp. 406–420. DOI: 10.1038/s41580-019-0126-2.

Lang, Michael R.; Lapierre, Lynne A.; Frotscher, Michael; Goldenring, James R.; Knapik, Ela W. (2006): Secretory COPII coat component Sec23a is essential for craniofacial chondrocyte maturation. In *Nature genetics* 38 (10), pp. 1198–1203. DOI: 10.1038/ng1880.

Lareau, Liana F.; Inada, Maki; Green, Richard E.; Wengrod, Jordan C.; Brenner, Steven E. (2007): Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. In *Nature* 446 (7138), pp. 926–929. DOI: 10.1038/nature05676.

Levy, Emile; Poinsot, Pierre; Spahis, Schohraya (2019): Chylomicron retention disease: genetics, biochemistry, and clinical spectrum. In *Current opinion in lipidology* 30 (2), pp. 134–139. DOI: 10.1097/MOL.0000000000000578.

Li, Jun; Witten, Daniela M.; Johnstone, Iain M.; Tibshirani, Robert (2012): Normalization, testing, and false discovery rate estimation for RNA-sequencing data. In *Biostatistics* 13 (3), pp. 523–538. DOI: 10.1093/biostatistics/kxr031.

Li, Ruilei; Li, Heng; Ge, Chunlei; Fu, Qiaofen; Li, Zhen; Jin, Yarong et al. (2020): Increased expression of the RNA-binding motif protein 47 predicts poor prognosis in non-small-cell lung cancer. In *Oncology letters* 19 (4), pp. 3111–3122. DOI: 10.3892/ol.2020.11417.

Liao, Yang; Smyth, Gordon K.; Shi, Wei (2014): featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. In *Bioinformatics (Oxford, England)* 30 (7), pp. 923–930. DOI: 10.1093/bioinformatics/btt656.

Lim, Jormay; Thiery, Jean Paul (2012): Epithelial-mesenchymal transitions: insights from development. In *Development* 139 (19), pp. 3471–3486. DOI: 10.1242/dev.071209.

Liu, Jie; Li, Juan; Li, Pingping; Wang, Yaochun; Liang, Zheyong; Jiang, Yina et al. (2017a): Loss of DLG5 promotes breast cancer malignancy by inhibiting the Hippo signaling pathway. In *Sci Rep* 7 (1), p. 42125. DOI: 10.1038/srep42125.

Liu, Yansheng; Gonzàlez-Porta, Mar; Santos, Sergio; Brazma, Alvis; Marioni, John C.; Aebersold, Ruedi et al. (2017b): Impact of Alternative Splicing on the Human Proteome. In *Cell reports* 20 (5), pp. 1229–1241. DOI: 10.1016/j.celrep.2017.07.025.

Long, Jennifer C.; Caceres, Javier F. (2009): The SR protein family of splicing factors: master regulators of gene expression. In *Biochem J* 417 (1), pp. 15–27. DOI: 10.1042/BJ20081501.

Lovci, Michael T.; Ghanem, Dana; Marr, Henry; Arnold, Justin; Gee, Sherry; Parra, Marilyn et al. (2013): Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. In *Nat Struct Mol Biol* 20 (12), pp. 1434–1442. DOI: 10.1038/nsmb.2699.

Love, Michael I.; Huber, Wolfgang; Anders, Simon (2014): Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. In *Genome biology* 15 (12), p. 550. DOI: 10.1186/s13059-014-0550-8.

Lu, Long-Feng; Zhang, Can; Zhou, Xiao-Yu; Li, Zhuo-Cong; Chen, Dan-Dan; Zhou, Yu et al. (2020): Zebrafish RBM47 Promotes Lysosome-Dependent Degradation of MAVS to Inhibit IFN Induction. In *J Immunol* 205 (7), pp. 1819–1829. DOI: 10.4049/jimmunol.1901387.

MacLennan, Ian C. M.; Toellner, Kai-Michael; Cunningham, Adam F.; Serre, Karine; Sze, Daniel M-Y; Zúñiga, Elina et al. (2003): Extrafollicular antibody responses. In *Immunological reviews* 194, pp. 8–18. DOI: 10.1034/j.1600-065x.2003.00058.x.

Malhotra, Vivek; Erlmann, Patrik (2011): Protein export at the ER: loading big collagens into COPII carriers. In *The EMBO journal* 30 (17), pp. 3475–3480. DOI: 10.1038/emboj.2011.255.

Marcheva, Biliana; Perelis, Mark; Weidemann, Benjamin J.; Taguchi, Akihiko; Lin, Haopeng; Omura, Chiaki et al. (2020): A role for alternative splicing in circadian control of exocytosis and glucose homeostasis. In *Genes & development* 34 (15-16), pp. 1089–1105. DOI: 10.1101/gad.338178.120.

Mariani, Valerio; Biasini, Marco; Barbato, Alessandro; Schwede, Torsten (2013): lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. In *Bioinformatics (Oxford, England)* 29 (21), pp. 2722–2728. DOI: 10.1093/bioinformatics/btt473.

Martinez, Nicole M.; Lynch, Kristen W. (2013): Control of alternative splicing in immune responses: many regulators, many predictions, much still to learn. In *Immunological reviews* 253 (1), pp. 216–236. DOI: 10.1111/imr.12047.

Martinez-Contreras, Rebeca; Cloutier, Philippe; Shkreta, Lulzim; Fisette, Jean-François; Revil, Timothée; Chabot, Benoit (2007): hnRNP proteins and splicing control. In *Advances in experimental medicine and biology* 623, pp. 123–147. DOI: 10.1007/978-0-387-77374-2_8.

Matlin, Arianne J.; Clark, Francis; Smith, Christopher W. J. (2005): Understanding alternative splicing: towards a cellular code. In *Nature reviews. Molecular cell biology* 6 (5), pp. 386–398. DOI: 10.1038/nrm1645.

McCaughey, Janine; Stephens, David J. (2018): COPII-dependent ER export in animal cells: adaptation and control for diverse cargo. In *Histochem Cell Biol* 150 (2), pp. 119–131. DOI: 10.1007/s00418-018-1689-2.

Merkin, Jason; Russell, Caitlin; Chen, Ping; Burge, Christopher B. (2012): Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. In *Science (New York, N.Y.)*

338 (6114), pp. 1593–1599. DOI: 10.1126/science.1228186.

Metzker, Michael L. (2010): Sequencing technologies - the next generation. In *Nat Rev Genet* 11 (1), pp. 31–46. DOI: 10.1038/nrg2626.

Meyerholz, Anika; Hinrichsen, Lars; Groos, Stephanie; Esk, Peter-Christopher; Brandes, Gudrun; Ungewickell, Ernst J. (2005): Effect of clathrin assembly lymphoid myeloid leukemia protein depletion on clathrin coat formation. In *Traffic (Copenhagen, Denmark)* 6 (12), pp. 1225–1234. DOI: 10.1111/j.1600-0854.2005.00355.x.

Miller, Elizabeth; Antonny, Bruno; Hamamoto, Susan; Schekman, Randy (2002): Cargo selection into COPII vesicles is driven by the Sec24p subunit. In *The EMBO journal* 21 (22), pp. 6105–6113. DOI: 10.1093/emboj/cdf605.

Miller, Elizabeth A.; Beilharz, Traude H.; Malkus, Per N.; Lee, Marcus C. S.; Hamamoto, Susan; Orci, Lelio; Schekman, Randy (2003): Multiple cargo binding sites on the COPII subunit Sec24p ensure capture of diverse membrane proteins into transport vesicles. In *Cell* 114 (4), pp. 497–509. DOI: 10.1016/S0092-8674(03)00609-3.

Miller, Sharon E.; Mathiasen, Signe; Bright, Nicholas A.; Pierre, Fabienne; Kelly, Bernard T.; Kladt, Nikolay et al. (2015): CALM regulates clathrin-coated vesicle size and maturation by directly sensing and driving membrane curvature. In *Developmental Cell* 33 (2), pp. 163–175. DOI: 10.1016/j.devcel.2015.03.002.

Miserey-Lenkei, Stéphanie; Bousquet, Hugo; Pylypenko, Olena; Bardin, Sabine; Dimitrov, Ariane; Bressanelli, Gaëlle et al. (2017): Coupling fission and exit of RAB6 vesicles at Golgi hotspots through kinesin-myosin interactions. In *Nat Commun* 8 (1), p. 1254. DOI: 10.1038/s41467-017-01266-0.

Montegna, Elisabeth A.; Bhave, Madhura; Liu, Yang; Bhattacharyya, Dibyendu; Glick, Benjamin S. (2012): Sec12 binds to Sec16 at transitional ER sites. In *PloS one* 7 (2), e31156. DOI: 10.1371/journal.pone.0031156.

Moore, M. J.; Sharp, P. A. (1993): Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. In *Nature* 365 (6444), pp. 364–368. DOI: 10.1038/365364a0.

Mortazavi, Ali; Williams, Brian A.; McCue, Kenneth; Schaeffer, Lorian; Wold, Barbara (2008): Mapping and quantifying mammalian transcriptomes by RNA-Seq. In *Nat Methods* 5 (7), pp. 621–628. DOI: 10.1038/nmeth.1226.

Moshkanbaryans, Lia; Xue, Jing; Wark, Jesse Ray; Robinson, Phillip James; Graham, Mark Evan (2016): A Novel Sequence in AP180 and CALM Promotes Efficient Clathrin Binding and

Assembly. In *PloS one* 11 (8), e0162050. DOI: 10.1371/journal.pone.0162050.

Mu, Yuanyue; Otsuka, Takeshi; Horton, April C.; Scott, Derek B.; Ehlers, Michael D. (2003): Activity-dependent mRNA splicing controls ER export and synaptic delivery of NMDA receptors. In *Neuron* 40 (3), pp. 581–594. DOI: 10.1016/s0896-6273(03)00676-7.

Munro, S.; Pelham, H. R. (1987): A C-terminal signal prevents secretion of luminal ER proteins. In *Cell* 48 (5), pp. 899–907. DOI: 10.1016/0092-8674(87)90086-9.

Murphy, Patrick A.; Butty, Vincent L.; Boutz, Paul L.; Begum, Shahinoor; Kimble, Amy L.; Sharp, Phillip A. et al. (2018): Alternative RNA splicing in the endothelium mediated in part by Rbfox2 regulates the arterial response to low flow. In *eLife Sciences Publications, Ltd*, 1/2/2018. Available online at https://elifesciences.org/articles/29494, checked on 7/27/2023.

Natoli, Manuela; Leoni, Bruno D.; D'Agnano, Igea; Zucco, Flavia; Felsani, Armando (2012): Good Caco-2 cell culture practices. In *Toxicology in vitro : an international journal published in association with BIBRA* 26 (8), pp. 1243–1246. DOI: 10.1016/j.tiv.2012.03.009.

Nauli, Andromeda M.; Whittimore, Judy D. (2015): Using Caco-2 Cells to Study Lipid Transport by the Intestine. In *Journal of visualized experiments : JoVE* (102), e53086. DOI: 10.3791/53086.

Neumann, Alexander; Meinke, Stefan; Goldammer, Gesine; Strauch, Miriam; Schubert, Daniel; Timmermann, Bernd et al. (2020): Alternative splicing coupled mRNA decay shapes the temperature-dependent transcriptome. In *EMBO reports* 21 (12), e51369. DOI: 10.15252/embr.202051369.

Neumann, Alexander; Schindler, Magdalena; Olofsson, Didrik; Wilhelmi, Ilka; Schürmann, Annette; Heyd, Florian (2019): Genome-wide identification of alternative splicing events that regulate protein transport across the secretory pathway. In *Journal of cell science* 132 (8). DOI: 10.1242/jcs.230201.

Nguyen, Doan C.; Joyner, Chester J.; Sanz, Iñaki; Lee, F. Eun-Hyung (2019): Factors Affecting Early Antibody Secreting Cell Maturation Into Long-Lived Plasma Cells. In *Frontiers in immunology* 10, p. 2138. DOI: 10.3389/fimmu.2019.02138.

Norum, Michaela; Tång, Erika; Chavoshi, Tina; Schwarz, Heinz; Linke, Dirk; Uv, Anne; Moussian, Bernard (2010): Trafficking through COPII stabilises cell polarity and drives secretion during Drosophila epidermal differentiation. In *PloS one* 5 (5), e10802. DOI: 10.1371/journal.pone.0010802.

Nutt, Stephen L.; Hodgkin, Philip D.; Tarlinton, David M.; Corcoran, Lynn M. (2015): The generation of antibody-secreting plasma cells. In *Nature reviews. Immunology* 15 (3),

pp. 160–171. DOI: 10.1038/nri3795.

Oettgen, P.; Akbarali, Y.; Boltax, J.; Best, J.; Kunsch, C.; Libermann, T. A. (1996): Characterization of NERF, a novel transcription factor related to the Ets factor ELF-1. In *Molecular and cellular biology* 16 (9), pp. 5091–5106. DOI: 10.1128/MCB.16.9.5091.

Ohisa, Satoshi; Inohaya, Keiji; Takano, Yoshiro; Kudo, Akira (2010): sec24d encoding a component of COPII is essential for vertebra formation, revealed by the analysis of the medaka mutant, vbi. In *Developmental biology* 342 (1), pp. 85–95. DOI: 10.1016/j.ydbio.2010.03.016.

Ohno, Susumu; Yomo, Tetsuya (1991): The grammatical rule for all DNA: Junk and coding sequences. In *Electrophoresis* 12 (2-3), pp. 103–108. DOI: 10.1002/elps.1150120203.

Oldham, Michael C.; Konopka, Genevieve; Iwamoto, Kazuya; Langfelder, Peter; Kato, Tadafumi; Horvath, Steve; Geschwind, Daniel H. (2008): Functional organization of the transcriptome in human brain. In *Nat Neurosci* 11 (11), pp. 1271–1282. DOI: 10.1038/nn.2207.

Olofsson, Didrik; Preußner, Marco; Kowar, Alexander; Heyd, Florian; Neumann, Alexander (2023): One pipeline to predict them all? On the prediction of alternative splicing from RNA-Seq data. In *Biochemical and Biophysical Research Communications* 653, pp. 31–37. DOI: 10.1016/j.bbrc.2023.02.053.

Ostwaldt, Felix; Los, Bruna; Heyd, Florian (2022): In silico analysis of alternative splicing events implicated in intracellular trafficking during B-lymphocyte differentiation. In *Frontiers in immunology* 13, p. 1030409. DOI: 10.3389/fimmu.2022.1030409.

Paraan, M.; Bhattacharya, N.; Uversky, V. N.; Stagg, S. M. (2018): Flexibility of the Sec13/31 cage is influenced by the Sec31 C-terminal disordered domain. In *Journal of structural biology* 204 (2). DOI: 10.1016/j.jsb.2018.08.016.

Parikh, Ishita; Fardo, David W.; Estus, Steven (2014): Genetics of PICALM expression and Alzheimer's disease. In *PloS one* 9 (3), e91242. DOI: 10.1371/journal.pone.0091242.

Patro, Rob; Duggal, Geet; Love, Michael I.; Irizarry, Rafael A.; Kingsford, Carl (2017): Salmon provides fast and bias-aware quantification of transcript expression. In *Nature methods* 14 (4), pp. 417–419. DOI: 10.1038/nmeth.4197.

Peinado, Héctor; Olmeda, David; Cano, Amparo (2007): Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? In *Nat Rev Cancer* 7 (6), pp. 415–428. DOI: 10.1038/nrc2131.

Peotter, Jennifer; Kasberg, William; Pustova, Iryna; Audhya, Anjon (2019): COPII-mediated

trafficking at the ER/ERGIC interface. In *Traffic* 20 (7), pp. 491–503. DOI: 10.1111/tra.12654.

Pertea, Mihaela; Salzberg, Steven L. (2010): Between a chicken and a grape: estimating the number of human genes. In *Genome biology* 11 (5), p. 206. DOI: 10.1186/gb-2010-11-5-206.

Radbruch, Andreas; Muehlinghaus, Gwendolin; Luger, Elke O.; Inamine, Ayako; Smith, Kenneth G. C.; Dörner, Thomas; Hiepe, Falk (2006): Competence and competition: the challenge of becoming a long-lived plasma cell. In *Nature reviews. Immunology* 6 (10), pp. 741–750. DOI: 10.1038/nri1886.

Raote, Ishier; Ortega Bellido, Maria; Pirozzi, Marinella; Zhang, Chong; Melville, David; Parashuraman, Seetharaman et al. (2017): TANGO1 assembles into rings around COPII coats at ER exit sites. In *The Journal of cell biology* 216 (4), pp. 901–909. DOI: 10.1083/jcb.201608080.

Reyes, Alejandro; Anders, Simon; Weatheritt, Robert J.; Gibson, Toby J.; Steinmetz, Lars M.; Huber, Wolfgang (2013): Drift and conservation of differential exon usage across tissues in primate species. In *Proceedings of the National Academy of Sciences of the United States of America* 110 (38), pp. 15377–15382. DOI: 10.1073/pnas.1307202110.

Ricard-Blum, Sylvie (2011): The collagen family. In *Cold Spring Harb Perspect Biol* 3 (1), a004978. DOI: 10.1101/cshperspect.a004978.

Ríos-Barrera, L. D.; Sigurbjörnsdóttir, S.; Baer, M.; Leptin, M. (2017): Dual function for Tango1 in secretion of bulky cargo and in ER-Golgi morphology. In *Proceedings of the National Academy of Sciences of the United States of America* 114 (48), E10389-E10398. DOI: 10.1073/pnas.1711408114.

Roberts, Adam; Trapnell, Cole; Donaghey, Julie; Rinn, John L.; Pachter, Lior (2011): Improving RNA-Seq expression estimates by correcting for fragment bias. In *Genome Biol* 12 (3), R22. DOI: 10.1186/gb-2011-12-3-r22.

Robinson, Mark D.; Oshlack, Alicia (2010): A scaling normalization method for differential expression analysis of RNA-seq data. In *Genome Biol* 11 (3), R25. DOI: 10.1186/gb-2010-11-3-r25.

Rokavec, Matjaz; Kaller, Markus; Horst, David; Hermeking, Heiko (2017): Pan-cancer EMT-signature identifies RBM47 down-regulation during colorectal cancer progression. In *Sci Rep* 7 (1), p. 4687. DOI: 10.1038/s41598-017-04234-2.

Rossbach, Oliver; Hung, Lee-Hsueh; Khrameeva, Ekaterina; Schreiner, Silke; König, Julian; Curk, Tomaž et al. (2014): Crosslinking-immunoprecipitation (iCLIP) analysis reveals global regulatory roles of hnRNP L. In *RNA Biology* 11 (2), pp. 146–155. DOI: 10.4161/rna.27991.

Ruprecht, Colin; Proost, Sebastian; Hernandez-Coronado, Marcela; Ortiz-Ramirez, Carlos; Lang, Daniel; Rensing, Stefan A. et al. (2017): Phylogenomic analysis of gene co-expression networks reveals the evolution of functional modules. In *The Plant journal : for cell and molecular biology* 90 (3), pp. 447–465. DOI: 10.1111/tpj.13502.

Saha, Ashis; Kim, Yungil; Gewirtz, Ariel D. H.; Jo, Brian; Gao, Chuan; McDowell, Ian C. et al. (2017): Co-expression networks reveal the tissue-specific regulation of transcription and splicing. In *Genome research* 27 (11), pp. 1843–1858. DOI: 10.1101/gr.216721.116.

Saito, Kota; Maeda, Miharu; Katada, Toshiaki (2017): Regulation of the Sar1 GTPase Cycle Is Necessary for Large Cargo Secretion from the Endoplasmic Reticulum. In *Frontiers in cell and developmental biology* 5, p. 75. DOI: 10.3389/fcell.2017.00075.

Saito, Kota; Yamashiro, Koh; Ichikawa, Yuki; Erlmann, Patrik; Kontani, Kenji; Malhotra, Vivek; Katada, Toshiaki (2011): cTAGE5 mediates collagen secretion through interaction with TANGO1 at endoplasmic reticulum exit sites. In *Molecular biology of the cell* 22 (13), pp. 2301–2308. DOI: 10.1091/mbc.E11-02-0143.

Saito, Kota; Yamashiro, Koh; Shimazu, Noriko; Tanabe, Tomoya; Kontani, Kenji; Katada, Toshiaki (2014): Concentration of Sec12 at ER exit sites via interaction with cTAGE5 is required for collagen export. In *The Journal of cell biology* 206 (6), pp. 751–762. DOI: 10.1083/jcb.201312062.

Saito, Y.; Kimura, K.; Oka, T.; Nakano, A. (1998): Activities of mutant Sar1 proteins in guanine nucleotide binding, GTP hydrolysis, and cell-free transport from the endoplasmic reticulum to the Golgi apparatus. In *Journal of biochemistry* 124 (4), pp. 816–823. DOI: 10.1093/oxfordjournals.jbchem.a022185.

Sakurai, T.; Isogaya, K.; Sakai, S.; Morikawa, M.; Morishita, Y.; Ehata, S. et al. (2016): RNA-binding motif protein 47 inhibits Nrf2 activity to suppress tumor growth in lung adenocarcinoma. In *Oncogene* 35 (38), pp. 5000–5009. DOI: 10.1038/onc.2016.35.

Salama, N. R.; Chuang, J. S.; Schekman, R. W. (1997): Sec31 encodes an essential component of the COPII coat required for transport vesicle budding from the endoplasmic reticulum. In *Molecular biology of the cell* 8 (2), pp. 205–217. DOI: 10.1091/mbc.8.2.205.

Sané, Alain Théophile; Seidman, Ernest; Peretti, Noel; Kleme, Marie Laure; Delvin, Edgard; Deslandres, Colette et al. (2017): Understanding Chylomicron Retention Disease Through Sar1b Gtpase Gene Disruption: Insight From Cell Culture. In *Arteriosclerosis, thrombosis, and vascular biology* 37 (12), pp. 2243–2251. DOI: 10.1161/ATVBAHA.117.310121.

Santos, António J. M.; Nogueira, Cristina; Ortega-Bellido, Maria; Malhotra, Vivek (2016): TANGO1 and Mia2/cTAGE5 (TALI) cooperate to export bulky pre-chylomicrons/VLDLs from

the endoplasmic reticulum. In *J Cell Biol* 213 (3), pp. 343–354. DOI: 10.1083/jcb.201603072.

Schlacht, Alexander; Dacks, Joel B. (2015): Unexpected ancient paralogs and an evolutionary model for the COPII coat complex. In *Genome biology and evolution* 7 (4), pp. 1098–1109. DOI: 10.1093/gbe/evv045.

Schrödinger, L. L.C. (2015): The PyMOL Molecular Graphics System, Version 1.8.

Sciammas, Roger; Shaffer, A. L.; Schatz, Jonathan H.; Zhao, Hong; Staudt, Louis M.; Singh, Harinder (2006): Graded expression of interferon regulatory factor-4 coordinates isotype switching with plasma cell differentiation. In *Immunity* 25 (2), pp. 225–236. DOI: 10.1016/j.immuni.2006.07.009.

Scotland, Paula B.; Heath, Jessica L.; Conway, Amanda E.; Porter, Natasha B.; Armstrong, Michael B.; Walker, Jennifer A. et al. (2012): The PICALM protein plays a key role in iron homeostasis and cell proliferation. In *PloS one* 7 (8), e44252. DOI: 10.1371/journal.pone.0044252.

Shen, Shihao; Park, Juw Won; Lu, Zhi-xiang; Lin, Lan; Henry, Michael D.; Wu, Ying Nian et al. (2014): rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. In *Proceedings of the National Academy of Sciences of the United States of America* 111 (51), E5593-601. DOI: 10.1073/pnas.1419161111.

Shi, Leming; Reid, Laura H.; Jones, Wendell D.; Shippy, Richard; Warrington, Janet A.; Baker, Shawn C. et al. (2006): The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. In *Nature biotechnology* 24 (9), pp. 1151–1161. DOI: 10.1038/nbt1239.

Shi, Yang; Jiang, Hui (2013): rSeqDiff: detecting differential isoform expression from RNA-Seq data using hierarchical likelihood ratio test. In *PloS one* 8 (11), e79448. DOI: 10.1371/journal.pone.0079448.

Shibata, Hideki; Suzuki, Hironori; Yoshida, Haruna; Maki, Masatoshi (2007): ALG-2 directly binds Sec31A and localizes at endoplasmic reticulum exit sites in a Ca2+-dependent manner. In *Biochemical and Biophysical Research Communications* 353 (3), pp. 756–763. DOI: 10.1016/j.bbrc.2006.12.101.

Shlomchik, Mark J.; Weisel, Florian (2012): Germinal center selection and the development of memory B and plasma cells. In *Immunological reviews* 247 (1), pp. 52–63. DOI: 10.1111/j.1600-065X.2012.01124.x.

Sinitcyn, Pavel; Richards, Alicia L.; Weatheritt, Robert J.; Brademan, Dain R.; Marx, Harald; Shishkova, Evgenia et al. (2023): Global detection of human variants and isoforms by deep

proteome sequencing. In *Nature biotechnology* 41 (12), pp. 1776–1786. DOI: 10.1038/s41587-023-01714-x.

Song, Yiyun (2021): Central dogma, redefined. In *Nat Chem Biol* 17 (8), p. 839. DOI: 10.1038/s41589-021-00850-2.

Stagg, Scott M.; Gürkan, Cemal; Fowler, Douglas M.; LaPointe, Paul; Foss, Ted R.; Potter, Clinton S. et al. (2006): Structure of the Sec13/31 COPII coat cage. In *Nature* 439 (7073), pp. 234–238. DOI: 10.1038/nature04339.

Stagg, Scott M.; LaPointe, Paul; Razvi, Abbas; Gürkan, Cemal; Potter, Clinton S.; Carragher, Bridget; Balch, William E. (2008): Structural basis for cargo regulation of COPII coat assembly. In *Cell* 134 (3), pp. 474–484. DOI: 10.1016/j.cell.2008.06.024.

Staley, J. P.; Guthrie, C. (1998): Mechanical devices of the spliceosome: motors, clocks, springs, and things. In *Cell* 92 (3), pp. 315–326. DOI: 10.1016/s0092-8674(00)80925-3.

Su, Qingfu; Pan, Zhenliang; Chen, Heyi; Chen, Jiabi; Zhang, Yanmei; Zhuang, Wei (2023): RBM47 restrains renal cell carcinoma progression and chemoresistance through interacting with lncRNA HOXB-AS1. In *Cell death discovery* 9 (1), p. 329. DOI: 10.1038/s41420-023-01623-7.

Taube, Joseph H.; Herschkowitz, Jason I.; Komurov, Kakajan; Zhou, Alicia Y.; Gupta, Supriya; Yang, Jing et al. (2010): Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. In *Proceedings of the National Academy of Sciences of the United States of America* 107 (35), pp. 15449–15454. DOI: 10.1073/pnas.1004900107.

Taylor, Rebecca J.; Tagiltsev, Grigory; Briggs, John A. G. (2023): The structure of COPI vesicles and regulation of vesicle turnover. In *FEBS Letters* 597 (6), pp. 819–835. DOI: 10.1002/1873-3468.14560.

Tellier, Julie; Nutt, Stephen L. (2019): Plasma cells: The programming of an antibody-secreting machine. In *European journal of immunology* 49 (1), pp. 30–37. DOI: 10.1002/eji.201847517.

Tellier, Julie; Shi, Wei; Minnich, Martina; Liao, Yang; Crawford, Simon; Smyth, Gordon K. et al. (2016): Blimp-1 controls plasma cell function through the regulation of immunoglobulin secretion and the unfolded protein response. In *Nat Immunol* 17 (3), pp. 323–330. DOI: 10.1038/ni.3348.

The pandas development team (2023): pandas-dev/pandas: Pandas: Zenodo.

Thiery, Jean Paul; Acloque, Hervé; Huang, Ruby Y. J.; Nieto, M. Angela (2009): Epithelial-

mesenchymal transitions in development and disease. In *Cell* 139 (5), pp. 871–890. DOI: 10.1016/j.cell.2009.11.007.

Thiery, Jean Paul; Sleeman, Jonathan P. (2006): Complex networks orchestrate epithelial-mesenchymal transitions. In *Nature reviews. Molecular cell biology* 7 (2), pp. 131–142. DOI: 10.1038/nrm1835.

Thomas A Caswell; Elliott Sales de Andrade; Antony Lee; Michael Droettboom; Tim Hoffmann; Jody Klymak et al. (2023): matplotlib/matplotlib: REL: v3.7.2: Zenodo.

Thomson, Stuart; Petti, Filippo; Sujka-Kwok, Izabela; Mercado, Peter; Bean, James; Monaghan, Melissa et al. (2011): A systems view of epithelial-mesenchymal transition signaling states. In *Clin Exp Metastasis* 28 (2), pp. 137–155. DOI: 10.1007/s10585-010-9367-3.

Thul, Peter J.; Åkesson, Lovisa; Wiking, Mikaela; Mahdessian, Diana; Geladaki, Aikaterini; Ait Blal, Hammou et al. (2017): A subcellular map of the human proteome. In *Science (New York, N.Y.)* 356 (6340). DOI: 10.1126/science.aal3321.

Tirosh, Amit; Mukherjee, Sanjit; Lack, Justin; Gara, Sudheer Kumar; Wang, Sophie; Quezado, Martha M. et al. (2019): Distinct genome-wide methylation patterns in sporadic and hereditary nonfunctioning pancreatic neuroendocrine tumors. In *Cancer* 125 (8), pp. 1247–1257. DOI: 10.1002/cncr.31930.

Townley, Anna K.; Feng, Yi; Schmidt, Katy; Carter, Deborah A.; Porter, Robert; Verkade, Paul; Stephens, David J. (2008): Efficient coupling of Sec23-Sec24 to Sec13-Sec31 drives COPII-dependent collagen secretion and is essential for normal craniofacial development. In *Journal of cell science* 121 (Pt 18), pp. 3025–3034. DOI: 10.1242/jcs.031070.

Townley, Anna K.; Schmidt, Katy; Hodgson, Lorna; Stephens, David J. (2012): Epithelial organization and cyst lumen expansion require efficient Sec13-Sec31-driven secretion. In *Journal of cell science* 125 (Pt 3), pp. 673–684. DOI: 10.1242/jcs.091355.

Trapnell, Cole; Hendrickson, David G.; Sauvageau, Martin; Goff, Loyal; Rinn, John L.; Pachter, Lior (2013): Differential analysis of gene regulation at transcript resolution with RNA-seq. In *Nat Biotechnol* 31 (1), pp. 46–53. DOI: 10.1038/nbt.2450.

Trapnell, Cole; Roberts, Adam; Goff, Loyal; Pertea, Geo; Kim, Daehwan; Kelley, David R. et al. (2012): Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. In *Nat Protoc* 7 (3), pp. 562–578. DOI: 10.1038/nprot.2012.016.

Tu, Linna; Banfield, David Karl (2010): Localization of Golgi-resident glycosyltransferases. In *Cell. Mol. Life Sci.* 67 (1), pp. 29–41. DOI: 10.1007/s00018-009-0126-z.

Tucker, Tracy; Marra, Marco; Friedman, Jan M. (2009): Massively parallel sequencing: the next big thing in genetic medicine. In *American journal of human genetics* 85 (2), pp. 142–154. DOI: 10.1016/j.ajhg.2009.06.022.

Turunen, Janne J.; Niemelä, Elina H.; Verma, Bhupendra; Frilander, Mikko J. (2013): The significant other: splicing by the minor spliceosome. In *Wiley interdisciplinary reviews. RNA* 4 (1), pp. 61–76. DOI: 10.1002/wrna.1141.

Uhlén, Mathias; Fagerberg, Linn; Hallström, Björn M.; Lindskog, Cecilia; Oksvold, Per; Mardinoglu, Adil et al. (2015): Proteomics. Tissue-based map of the human proteome. In *Science (New York, N.Y.)* 347 (6220), p. 1260419. DOI: 10.1126/science.1260419.

Uhlén, Mathias; Karlsson, Max J.; Hober, Andreas; Svensson, Anne-Sophie; Scheffel, Julia; Kotol, David et al. (2019): The human secretome. In *Science signaling* 12 (609). DOI: 10.1126/scisignal.aaz0274.

Valente, Carmen; Turacchio, Gabriele; Mariggiò, Stefania; Pagliuso, Alessandro; Gaibisso, Renato; Di Tullio, Giuseppe et al. (2012): A 14-3-3γ dimer-based scaffold bridges CtBP1-S/BARS to PI(4)KIIIβ to regulate post-Golgi carrier formation. In *Nat Cell Biol* 14 (4), pp. 343–354. DOI: 10.1038/ncb2445.

van Dam, Sipko; Võsa, Urmo; van der Graaf, Adriaan; Franke, Lude; Magalhães, João Pedro de (2018): Gene co-expression analysis for functional classification and gene-disease predictions. In *Briefings in bioinformatics* 19 (4), pp. 575–592. DOI: 10.1093/bib/bbw139.

van der Verren, Sander E.; Zanetti, Giulia (2023): The small GTPase Sar1, control centre of COPII trafficking. In *FEBS Letters* 597 (6), pp. 865–882. DOI: 10.1002/1873-3468.14595.

van Rossum, Guido; Drake, Fred L. (2009): Python 3 Reference Manual. Scotts Valley, CA: CreateSpace.

Vanharanta, Sakari; Marney, Christina B.; Shu, Weiping; Valiente, Manuel; Zou, Yilong; Mele, Aldo et al. (2014): Loss of the multifunctional RNA-binding protein RBM47 as a source of selectable metastatic traits in breast cancer. In *eLife Sciences Publications, Ltd*, 6/4/2014, checked on 1/10/2024.

Vernia, Santiago; Edwards, Yvonne Jk; Han, Myoung Sook; Cavanagh-kyros, Julie; Barrett, Tamera; Kim, Jason K.; Davis, Roger J. (2016): An alternative splicing program promotes adipose tissue thermogenesis. In *eLife* 5. DOI: 10.7554/eLife.17672.

Victora, Gabriel D.; Nussenzweig, Michel C. (2012): Germinal centers. In *Annual review of immunology* 30, pp. 429–457. DOI: 10.1146/annurev-immunol-020711-075032.

Virtanen, Pauli; Gommers, Ralf; Oliphant, Travis E.; Haberland, Matt; Reddy, Tyler;

Cournapeau, David et al. (2020): SciPy 1.0: fundamental algorithms for scientific computing in Python. In *Nature methods* 17 (3), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

Wakana, Yuichi; Villeneuve, Julien; van Galen, Josse; Cruz-Garcia, David; Tagaya, Mitsuo; Malhotra, Vivek (2013): Kinesin-5/Eg5 is important for transport of CARTS from the trans-Golgi network to the cell surface. In *The Journal of cell biology* 202 (2), pp. 241–250. DOI: 10.1083/jcb.201303163.

Wang, Bo; Joo, Joung Hyuck; Mount, Rebecca; Teubner, Brett J. W.; Krenzer, Alison; Ward, Amber L. et al. (2018): The COPII cargo adapter SEC24C is essential for neuronal homeostasis. In *The Journal of clinical investigation* 128 (8), pp. 3319–3332. DOI: 10.1172/JCI98194.

Wang, Eric T.; Cody, Neal A. L.; Jog, Sonali; Biancolella, Michela; Wang, Thomas T.; Treacy, Daniel J. et al. (2012): Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. In *Cell* 150 (4), pp. 710–724. DOI: 10.1016/j.cell.2012.06.041.

Wang, Eric T.; Sandberg, Rickard; Luo, Shujun; Khrebtukova, Irina; Zhang, Lu; Mayr, Christine et al. (2008): Alternative isoform regulation in human tissue transcriptomes. In *Nature* 456 (7221), pp. 470–476. DOI: 10.1038/nature07509.

Wang, Yi; Hicks, Stephanie C.; Hansen, Kasper D. (2022): Addressing the mean-correlation relationship in co-expression analysis. In *PLoS computational biology* 18 (3), e1009954. DOI: 10.1371/journal.pcbi.1009954.

Waskom, Michael (2021): seaborn: statistical data visualization.

Wei, Yinxiang; Zhang, Fanghui; Zhang, Yu; Wang, Xiaoqian; Xing, Chen; Guo, Jing et al. (2019): Post-transcriptional regulator Rbm47 elevates IL-10 production and promotes the immunosuppression of B cells. In *Cell Mol Immunol* 16 (6), pp. 580–589. DOI: 10.1038/s41423-018-0041-z.

Weigel, Aubrey V.; Chang, Chi-Lun; Shtengel, Gleb; Xu, C. Shan; Hoffman, David P.; Freeman, Melanie et al. (2021): ER-to-Golgi protein delivery through an interwoven, tubular network extending from ER. In *Cell* 184 (9), 2412-2429.e16. DOI: 10.1016/j.cell.2021.03.035.

Weissman, J. T.; Plutner, H.; Balch, W. E. (2001): The mammalian guanine nucleotide exchange factor mSec12 is essential for activation of the Sar1 GTPase directing endoplasmic reticulum export. In *Traffic (Copenhagen, Denmark)* 2 (7), pp. 465–475. DOI: 10.1034/j.1600-0854.2001.20704.x.

Wilhelmi, Ilka; Kanski, Regina; Neumann, Alexander; Herdt, Olga; Hoff, Florian; Jacob, Ralf

et al. (2016): Sec16 alternative splicing dynamically controls COPII transport efficiency. In *Nature communications* 7, p. 12347. DOI: 10.1038/ncomms12347.

Will, Cindy L.; Lührmann, Reinhard (2011): Spliceosome structure and function. In *Cold Spring Harb Perspect Biol* 3 (7), a003707. DOI: 10.1101/cshperspect.a003707.

Willsey, A. Jeremy; Sanders, Stephan J.; Li, Mingfeng; Dong, Shan; Tebbenkamp, Andrew T.; Muhle, Rebecca A. et al. (2013): Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. In *Cell* 155 (5), pp. 997–1007. DOI: 10.1016/j.cell.2013.10.020.

Wilson, Deanna G.; Phamluong, Khanhky; Li, Li; Sun, Mei; Cao, Tim C.; Liu, Peter S. et al. (2011): Global defects in collagen secretion in a Mia3/TANGO1 knockout mouse. In *The Journal of cell biology* 193 (5), pp. 935–951. DOI: 10.1083/jcb.201007162.

Xie, Yinlong; Wu, Gengxiong; Tang, Jingbo; Luo, Ruibang; Patterson, Jordan; Liu, Shanlin et al. (2014): SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. In *Bioinformatics (Oxford, England)* 30 (12), pp. 1660–1666. DOI: 10.1093/bioinformatics/btu077.

Xu, Chengchao; Ng, Davis T. W. (2015): Glycosylation-directed quality control of protein folding. In *Nat Rev Mol Cell Biol* 16 (12), pp. 742–752. DOI: 10.1038/nrm4073.

Yamasaki, Akinori; Tani, Katsuko; Yamamoto, Akitsugu; Kitamura, Naomi; Komada, Masayuki (2006): The Ca2+-binding protein ALG-2 is recruited to endoplasmic reticulum exit sites by Sec31A and stabilizes the localization of Sec31A. In *Molecular biology of the cell* 17 (11), pp. 4876–4887. DOI: 10.1091/mbc.e06-05-0444.

Yang, Xinping; Coulombe-Huntington, Jasmin; Kang, Shuli; Sheynkman, Gloria M.; Hao, Tong; Richardson, Aaron et al. (2016a): Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. In *Cell* 164 (4), pp. 805–817. DOI: 10.1016/j.cell.2016.01.029.

Yang, Yueqin; Park, Juw Won; Bebee, Thomas W.; Warzecha, Claude C.; Guo, Yang; Shang, Xuequn et al. (2016b): Determination of a Comprehensive Alternative Splicing Regulatory Network and Combinatorial Regulation by Key Factors during the Epithelial-to-Mesenchymal Transition. In *Molecular and cellular biology* 36 (11), pp. 1704–1719. DOI: 10.1128/MCB.00019-16.

Ye, W.; Lafer, E. M. (1995): Clathrin binding and assembly activities of expressed domains of the synapse-specific clathrin assembly protein AP-3. In *Journal of Biological Chemistry* 270 (18), pp. 10933–10939. DOI: 10.1074/jbc.270.18.10933.

Zilliox, Michael J.; Irizarry, Rafael A. (2007): A gene expression bar code for microarray data. In *Nat Methods* 4 (11), pp. 911–913. DOI: 10.1038/nmeth1102.

6 Appendix

6.1 Abbreviations

A	adenosine
ab	antibodies
ALG2	apoptosis-linked gene 2
ASC	antibody-secreting cells
ATP	adenosine triphosphate
BP	branch point
CARTS	carriers of the TGN to the cell surface
cDNA	complementary DNA
COPI	coat protein complex I
COPII	coat protein complex II
cTAGE5	cutaneous T-cell lymphoma-associated antigen 5
DNA	Deoxyribonucleic acid
EMT	epithelial-mesenchymal transition
ER	endoplasmic reticulum
ERES	ER exit site
ERGIC	ER Golgi intermediate complex
ESRP	Epithelial splicing regulatory protein
FPKM	fragments per kilobase of exon per million
FRAP	fluorescence recovery after photobleaching
G	guanine
GAP	GTPase-activating protein
GEF	guanine nucleotide exchange factor
GO	gene ontology
Golgi	Golgi-apparatus
GPI	glycosylphosphatidylinositol
hnRNP	heterogenous nuclear ribonucleoproteins
IL-10	Interleukin 10
KDEL	lysine, aspartic acid, glutamic acid, leucine
KDEL-E-cadherin	Str-KDEL_SBP-EGFP-Ecadherin
KDEL-GPI	Str-KDEL_SBP-EGFP-GPI
lDDT	local Distance Difference Test
MBC	memory-B-cells
MET	mesenchymal-epithelial transition
MO	morpholino oligos
mRNA	messenger RNA
NMD	nonsense-mediated decay
pae	predicted aligned error

PB	plasmablasts
PBS	Dulbecco's Phosphate Buffered Saline
PC	plasma cells
PCA	principal component analysis
PCR	Polymerase chain reaction
PICALM	Phosphatidylinositol-binding clathrin assembly protein
pre-mRNA	precursor messenger RNA
prePB	preplasmablasts
PSI	percent-splice-in
QKI	KH domain-containing RNA-binding protein
RBM47	RNA-binding motive 47
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RPKM	Reads per kilobase of transcript per Million mapped reads
rRNA	ribosomal RNA
RUSH	Retention using selective hooks
snRNA	small-noncoding RNA
snRNP	small nuclear ribonucleoproteins
SR	serine-arginine rich
SS	splice site
TALI	TANGO1-like
TANGO1	transport and Golgi organization 1
TCI	Taurocholate
TF	transcription factors
TPM	transcript per million
U	uracil
U2AF	U2 auxiliary factors
WND	tryptophan, asparagine, aspartic acid

6.2 List of Figures

6.3 List of Tables