

**Application of BioID**  
**to *in vitro* organelle and *in vivo* cell-type-specific proteomics**

Dissertation  
zur  
Erlangung des Doktorgrades (Dr. rer. nat.)  
der  
Mathematisch-Naturwissenschaftlichen Fakultät  
der  
Rheinische Friedrich-Wilhelms-Universität Bonn

vorgelegt von  
**Shiva Ahmadi**  
aus  
Karaj, Iran

Bonn, 2020

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen  
Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Volkmar Gieselmann

2. Gutachter: Prof. Dr. Christoph Thiele

Tag der Promotion: 08.06.2020

Erscheinungsjahr: 2020

## Table of Content

Table of Content.....	i
1. Abstract .....	1
2. Introduction .....	3
2.1. Techniques for cell-type-specific proteomics .....	3
2.1.1. Single cell proteomics.....	3
2.1.2. Fluorescent activated cell sorting.....	3
2.1.3. Laser capture microdissection.....	4
2.1.4. Induced pluripotent stem cells .....	4
2.1.5. <i>In vivo</i> approaches; <i>Cre/loxP</i> -dependent, tetracycline inducible transgenic system .....	4
2.1.6. <i>In vivo</i> approaches; <i>in vivo</i> labeling.....	6
2.2. Techniques for Organelle Specific Proteomics.....	8
2.3. Proximity Labeling using BioID and APEX.....	8
2.4. Enrichment based on Avidin-biotin interaction .....	10
2.5. Mass spectrometry-based proteomics.....	10
2.6. Quantitative Mass Spectrometry.....	12
2.7. Data analysis of biotin enriched proteins after quantitative mass spectrometry .....	14
2.8. Aims of the study.....	14
3. Materials.....	15
3.1. Chemicals/Reagents .....	15
3.2. Enzymes .....	16
3.3. Cell lines.....	17
3.4. Mice.....	17
3.5. Primers .....	17
3.6. Primary Antibodies.....	19
3.7. Secondary Antibodies and Chemicals .....	20
3.8. Instruments and consumables .....	20
3.9. Software and online tools .....	22
3.10. Ready to use reaction kits.....	22
3.11. Drugs and Treatments and Mouse Food .....	23
4. Methods.....	24
4.1. Cell culture techniques .....	24
4.1.1. Cell culture and labeling of proteins using SILAC amino acids .....	24
4.1.2. Transfection and Biotinylation .....	24
4.2. Molecular biology techniques.....	24
4.2.1. Polymerase Chain Reaction (PCR).....	24
4.2.6. Genomic DNA Extraction from tail tips and genotyping PCR .....	25
4.2.2. Digestion of DNA using restriction endonucleases .....	26

4.2.3. Agarose gel electrophoresis .....	26
4.2.4. DNA extraction and purification from agarose gels .....	26
4.2.5. Ligation and transformation of <i>E. coli</i> .....	27
4.3. Protein techniques and mass spectrometry sample preparation .....	28
4.3.1. Determination of protein concentrations.....	28
4.3.2. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) .....	28
4.3.3. Staining of polyacrylamide gels.....	28
4.3.4. Western blot.....	28
4.3.5. Immunofluorescence staining of cells.....	29
4.3.6. Immunofluorescence staining of tissues .....	29
4.3.7. Protein precipitation.....	29
4.3.9. Affinity enrichment of biotinylated proteins from cells ( <i>in vitro</i> ).....	30
4.3.10. Affinity enrichment of biotinylated proteins from tissue samples ( <i>in vivo</i> ) .....	30
4.3.11. OASIS desalting .....	31
4.3.12. Quantitative peptide assay .....	31
4.3.13. RapiGest Digestion.....	31
4.3.14. Labeling of peptides using dimethyl labeling reagents .....	32
4.3.15. Labeling of peptides using TMT 6plex and 10plex .....	32
4.1.16. OFFGEL fractionation.....	32
4.3.17. High-pH reversed-phase peptide fractionation (Hp-RP).....	33
4.1.18. Peptide cleanup with StageTips .....	34
4.4. Mass Spectrometry .....	34
4.4.1. Nano-UHPLC-MS/MS-Orbitrap-Velos (Orbitrap-Velos).....	34
4.4.2. Nano-UHPLC-MS/MS-Orbitrap-Lumos (Orbitrap-Lumos) .....	35
4.5. Data analysis, bioinformatics and statistical analysis .....	35
4.5.1. Data Analysis using MaxQuant .....	35
4.5.2. Data Analysis using Proteome Discoverer.....	35
4.5.3. Perseus analysis for <i>in vitro</i> experiments (SILAC labeled samples).....	36
4.5.4. Perseus analysis for <i>in vivo</i> experiments (Dimethyl and TMT labeled samples) .....	36
4.6. Mice techniques.....	38
4.6.1. Treatment of mice.....	38
4.6.2. Blood glucose measurements.....	38
4.6.3. Dissection of tissues .....	38
5. Results .....	40
5.1. Establishment of a statistical approach for cut-off determination in BioID-based enrichment methods and its implementation in organelle proteomics.....	40
5.1.1. Experimental determination of a cut-off for the acceptance of biotinylated peptides .....	40
5.1.2. Analysis of organelle-specific proteomes including one control sample .....	44
5.1.3. Effect of biotinylated proteins in the binding pattern of non-specific proteins .....	52

5.1.4. Analysis of organelle-specific proteomes including two control samples .....	59
5.2. <i>In vivo</i> cell-type-specific analysis in BioID mice models.....	65
5.2.1. Generation of BioID mice.....	65
5.2.2. Functionality and tissue expression of BioID mice.....	67
5.2.3. MS-based enrichment analysis .....	73
5.2.4. Enrichment for proteins specific to beta cells of pancreas.....	83
5.2.5. Effect of Harmine and STZ on beta-cell Proteome .....	91
6. Discussion .....	98
6.1. Establishment of a statistical approach for cut-off determination in BioID-based enrichment methods and its implementation in organelle proteomics.....	98
6.1.1. Experimental determination of a cut-off for the acceptance of biotinylated peptides .....	98
6.1.3. Effect of biotinylated proteins in the binding pattern of non-specific proteins .....	102
6.1.4. Analysis of organelle-specific proteomes including two control samples .....	103
6.1.5. Comparison between organelle proteomics approaches.....	105
6.2. <i>In vivo</i> cell-type-specific analysis in BioID mice models.....	105
6.2.1. Tissue expression of BioID mice (Rosa26 vs COL1A1) .....	105
6.2.2. “Biotinylation cut-off” and “Drug-effect cut-off” for interpretation of data.....	107
6.2.3. Method Development .....	108
6.2.4. Tissue proteomics including two control samples .....	110
6.2.5. Duration for Doxycycline administration .....	111
6.2.6. Beta Cell proteomics.....	111
6.2.7. Effect of Harmine on beta cell proteome .....	112
6.2.8. Effect of STZ on beta cell proteome .....	113
6.2.9. Possible mistakes in data analysis.....	114
6.2.10. Comparison of BioID cell-type-specific approach with other in-vivo labeling approaches.....	115
7. Conclusion.....	118
8. List of abbreviations .....	119
9. List of figures .....	122
10. List of tables .....	124
11. List of annex tables.....	125
12. References .....	126
13. Appendix .....	132
14. Acknowledgement .....	141

## 1. Abstract

A genetically modified biotin ligase BirA (Arg118Gly), termed BioID (BirA\*), can be induced to add a reactive form of biotin to proximate proteins in the cytosol and nucleus. The biotinylated proteins can be purified and identified by mass spectrometry (MS). In this project, we combined the BioID proximity biotinylation approach with MS-based quantitative proteomics for mapping organelle specific (*in vitro*) and cell-type-specific (*in vivo*) proteomes.

Using a stable isotope labeling by/with amino acids (SILAC) labeled Human embryonic kidney (HEK 293T) cell line transfected with BioID, we evaluated the effect of biotinylated proteins on the binding pattern of background/non-specific proteins to streptavidin beads. By combining biotinylated and two control samples in different ways and evaluating protein intensities we realized that non-specific proteins have a low chance for binding the beads when only one control sample exists and therefore their intensities are below MS detection limits. However as soon as a second control sample is added, the chance of non-specific proteins in binding the beads and being identified by MS increases. To distinguish between biotinylated and non-specifically binding proteins, using the two control samples we defined a statistical cut-off including 95% of the proteins in control samples. This cut-off was found to be almost identical with the cut-off determined by “slide approach” based on Gene Ontology (GO) analysis (Branon et al., 2018); this proved the reliability of our cut-off determination method. Organelle proteomics was performed using double control strategy (in a 3plex SILAC approach) and data validated by comparison to the literature. We found 97% of the Cytosol/nucleus proteins, 77% and 53% of nuclear and mitochondrial proteins previously linked to these organelles.

After successful implementation of BioID technique in cell culture and enrichment for organelle proteomes, we expanded the application of the technique to *in vivo* cell-type-specific analysis. It is reported that certain diseases only affect a subpopulation of a specific cell type in a tissue. The ability to examine cell-type-specific protein expression is critical for investigating the molecular pathology of such diseases. To this aim, under the control of Col1a1 and Rosa26 loci, two BioID mouse lines were generated. Universal expression of BioID was achieved by crossing these mice with lines constitutively expressing rtTA and Cre recombinase in all tissues. From these animals, twelve tissues were analyzed by Western blot confirming the expression of, and biotinylation by, BioID. Selected tissues were investigated by immunofluorescence and mass spectrometry. Using a dimethyl labeling-based strategy, on-bead digestion protocol was optimized improving the identification and enrichment rates by ~5-folds. The best method was applied on 3 tissue samples to enrich for biotinylated proteins. Biotinylation degrees were calculated as 46%, 87%, and 92% for lung, pancreas, and liver, respectively. Using Ins1Cre mouse strain, BioID was expressed exclusively in pancreatic beta cells, on-bead digestion performed and biotinylated and control samples labeled with different tandem mass tags (TMT). An *in vivo* beta-cell proteome was generated including a total number of 4133 proteins, 2992 significantly enriched (biotinylated); 96% of enriched proteins were previously identified by large scale islet or beta-cell omics studies. The dataset was depleted for markers of other pancreatic cell types. Using TMT strategy, the effect of pro- and anti-diabetic drugs Streptozotocin (STZ) and Harmine, on the proteome of beta cells was evaluated. To validate the data, the regulated proteins were uploaded in GO platform and searched for enriched Reactome pathways and compared to previously published literature on mechanisms involved by Harmine and STZ. Harmine, known as an antidiabetic and anticancer drug, resulted in downregulation of 2518 proteins and upregulation of 97 proteins; more than 60 pathways enriched in this dataset were found previously connected to

## Abstract

the effects of this drug. STZ, a diabetes-inducing drug, resulted in downregulation of 1274 proteins and upregulation of 6 proteins. Out of the enriched pathways, 86 were found previously linked to diabetes and STZ.

## 2. Introduction

The function of proteins in cells is to catalyze, conduct, and control most processes at specific time points and locations. Cells are made up of several different cell types and subcellular compartments. Each cell type and subcellular compartment has unique characteristics and distinct populations of proteins. The specific properties of organelles and cell-types are crucial for normal cellular function and play important roles in development, differentiation, mitosis, and various forms of stress response (Jean Beltran, Mathias, and Cristea 2016). Mis-localization of proteins have often been associated with cellular dysfunction and disease including disorders of the lysosomes, mitochondria, and peroxisomes, endosomes, and endoplasmic reticulum (ER) (Jean Beltran, Mathias, and Cristea 2016). Also, many diseases are caused by the malfunction of distinct cell types or a sub-population of a specific cell type in a tissue. For example, in diabetes, the beta cells of the pancreas are affected while acinar cells, as well as alpha and delta cells of islets, are not involved in development of diabetes (Da Silva Xavier 2018). The ability to examine cell-type-specific and organelle-specific protein expression is critical for investigating the molecular pathology of such diseases and is central for understanding protein functions, interactions, and cellular mechanisms. This suggests the need for techniques for reliable organelle and cell-type-specific assignment of proteins.

### 2.1. Techniques for cell-type-specific proteomics

To date, the most common approach for tissue proteomics is the homogenization of whole tissue or further culture of individual cell types *in vitro*. Determining the proteomes of specific cells is a particular challenge since cell types form a network with dense chemical, electrical, and physical inter-connectivity. By separating them on a culture dish, cells may lose their physical network, cellular structures, and connectivity; the resulting cells adapt proteomes which do not resemble the *in vivo* physiological conditions (Krogager et al. 2018)

While some cell types comprise a large population of cells in a particular tissue, the ratio of some other cell populations is very small. For example, 98% of pancreas cells are exocrine acinar cells whereas beta cells account for 1-2% of pancreatic cells. With whole tissue proteomics techniques, it's almost impossible to study the alterations of beta cell proteome as a consequence of diabetes.

#### 2.1.1. Single cell proteomics

A solution to this would be to perform proteomics at the single-cell level. Nowadays, massive data are generated from single-cell genomics and transcriptomics analysis. However, single-cell proteomics has lagged behind the transcriptomics counterpart due to substantial technical challenges. One major problem is the low abundance of proteins due to sample loss during the upstream experimental steps or low amount of protein in the sample which is below the threshold of detection for mass spectrometry (MS) analysis. Moreover, many of the single-cell isolation techniques are incapable of retraining cellular structure and integrity which can influence and alter the results (Schoof et al. 2019).

#### 2.1.2. Fluorescent activated cell sorting

Fluorescent activated cell sorting (FACs) is another method for the isolation of different cell populations using antibodies or fluorescent reporters under control of cell-type-specific promoters as well as enrichment using cell-type-specific antibodies (Holt and Olsen 2016). The major drawback of this technique is that the efficiency

## Introduction

of isolation of the sparse population of cells from small tissue regions is very low and requires the pooling of tissue samples.

### **2.1.3. Laser capture microdissection**

Another approach is laser capture microdissection (LCM) which relies on isolation of a subpopulation of cells using laser and their visualization using microscopy. However, due to the time consuming and low yield nature of this technique, the subsequent analyses of the proteins need very sensitive mass spectrometers, especially if the expression of proteins is low at input level (Schwarz 2015).

### **2.1.4. Induced pluripotent stem cells**

Another cell-type-specific isolation method, especially for human tissue samples, is based on the development of cell types from induced pluripotent stem cells (iPSCs). iPSCs can be derived from easily accessible cell types such as blood cells or skin fibroblast of human patients and based on their intrinsic self-renewal properties, differentiate into nearly any cell types of the human body (Shi et al. 2017). While iPSC models have shown significant promise for moving closer to the human model system and increasing the protein yield, they are still *in vitro* experiments. This means that cell type connectivity, as observed in a real *in vivo* model, cannot be maintained, which results in perturbed protein expression.

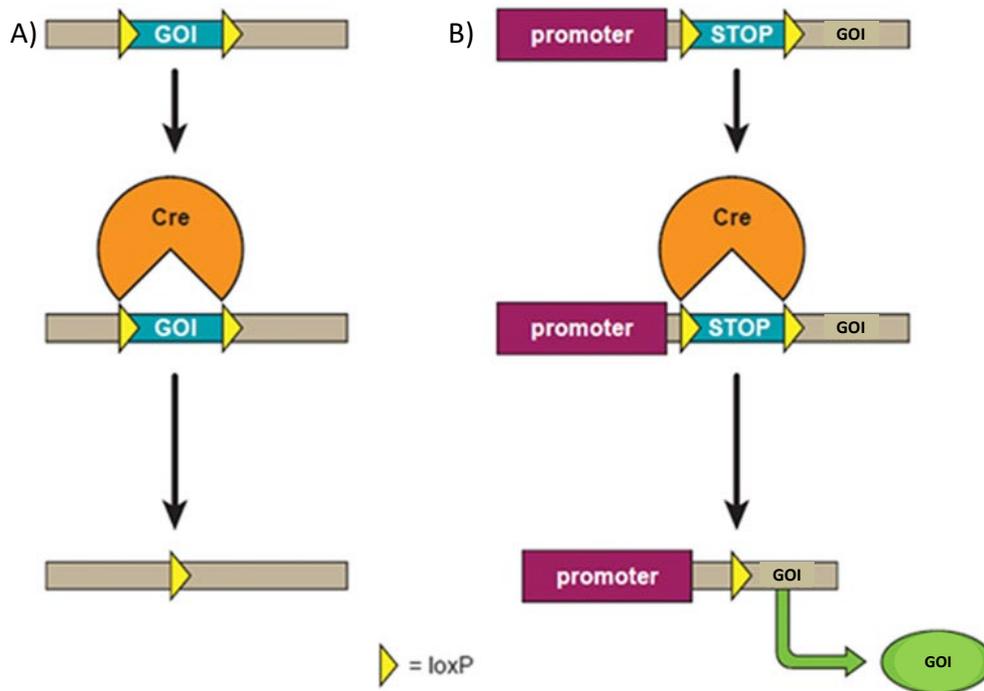
### **2.1.5. *In vivo* approaches; Cre/loxP-dependent, tetracycline inducible transgenic system**

The classical *in vivo* cell-type-specific proteomics relies on transgenic mouse lines with cell-type-specific promoters to over-express reporter proteins in the cell type of interest. One of these tools has been developed by taking advantage of a bacterial artificial chromosome (BAC) to express a green fluorescent marker in specific cell lines (Gong et al. 2003). Cre mice (harbouring the Cre/loxP genetic system) in combination with tetracycline (TET)-inducible lines (employing Tet-off/Tet-on genetic systems) enable spatial and temporal regulation of transgene expression *in vivo* and present optimal tools for microscopic visualization of intracellular fine structural details of specific cell types in disease conditions (C. Chen et al. 2017). However, due to complexity of tissue proteome and limited effect size of small sub-populations of cells (e.g. in case of the beta cells), the MS-based investigation of proteome alteration in the specific cell type is only possible if this tool is coupled with other cell-type-specific isolation methods (R. S. Wilson and Nairn 2018).

#### **2.1.5.1. Cre/loxP genetic system**

In the Cre/LoxP system, Cre recombinase, a tyrosine recombinase enzyme derived from the P1 bacteriophage, enables site-specific recombination events between two DNA recognition sites, the LoxP sites. Cre/loxP system is one of the most powerful genetic systems allowing spatiotemporal control of gene expression in mouse models. It is very useful for excision of a gene due to irreversible nature of this action. This system has been adopted for Cre-dependant sequence knockout or expression. With the two loxP sites in the same orientation, the sequence in the middle is excised upon crossing the mouse harbouring loxP sites with a Cre mouse. For gene knockout, the gene is flanked by two loxP sites and is knocked out upon Cre activation. For activation of the gene expression, a “loxP-stop cassette-loxP” is placed upstream of the gene. In absence of Cre, the stop cassette prevents the translational expression of the gene and as soon as Cre is present, the stop cassette is excised and the gene expression proceeds (Figure 2-1; (Nagy 2000)).

## Introduction



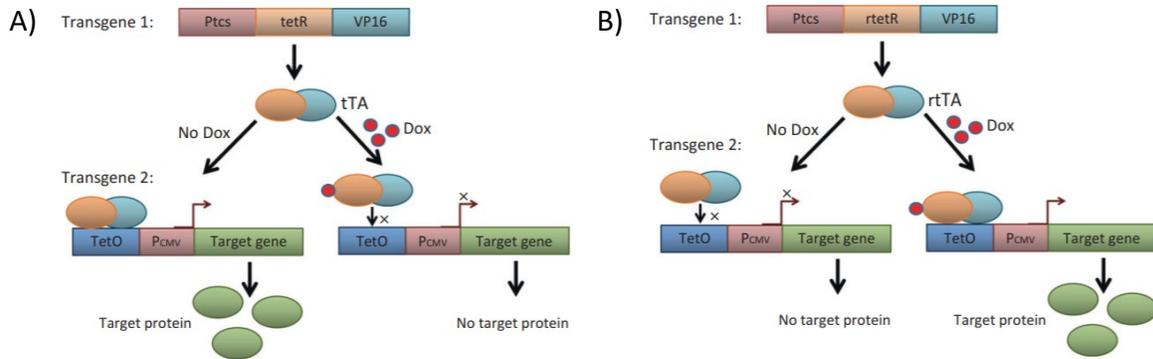
**Figure 2-1. Cre/lox system.**

A) Cre/loxP system can be used to remove the gene of interest (GOI). B) Cre recombinase can be used to activate the expression of a transgene by removing the stop cassette flanked by loxP sites (Adapted from Cox et al., 2012).

### 2.1.5.2. Tet-off/Tet-on genetic systems

Tet-off and Tet-on expression systems are two transgenic systems in which expression of the gene of interest is dependent on the activity of an inducible transcriptional activator. A Tet response element (TRE) is composed of 7 repeats of a 19 nucleotide tetracycline operator (tetO) sequence and is recognized by the tetracycline repressor (tetR). A tetracycline-controlled transactivator (tTA) is created by fusing tetR with the C-terminal domain of VP16 (virion protein 16), an essential transcriptional activation domain from HSV (herpes simplex virus). In the Tet-off system, the expression of the gene is stopped in the presence of Tetracycline (Tc)/Doxycycline (Dox). Upon removal of Tc/Dox, the tTA binds the TRE sequences and gene expression is activated. In the presence of Tc/Dox, it binds the TetR and the expression remains inactive. Tet-on system is complementary to the Tet-off system and is based on a reverse tetracycline-controlled transactivator, rtTA. rtTA is a fusion protein comprised of TetR repressor and the VP16 domain. rtTA can recognize the TRE sequence in the presence of Tc or Dox (Figure 2-2; (Yamamoto, Hen, and Dauer 2001)). The Tet-on and Tet-off systems can be used to generate inducible gene expression animal models which can serve as valuable tools for the investigation of human diseases (Zhou et al. 2009).

## Introduction



**Figure 2-2. Tet-on and Tet-off systems.**

A) Tet-off system; upon depletion of Tc/Dox, tTA binds the TetO sequence and induces the expression of the transgene. B) Tet-on system; upon administration of Tc/Dox, rtTA recognizes the TetO sequence and results in expression of the gene of interest (Adapted from Sakai 2014). Dox: Doxycycline; Tc: Tetracycline; TetO: Tetracycline operator; Ptcs: cell-type-specific promoter.

### 2.1.6. *In vivo* approaches; *in vivo* labeling

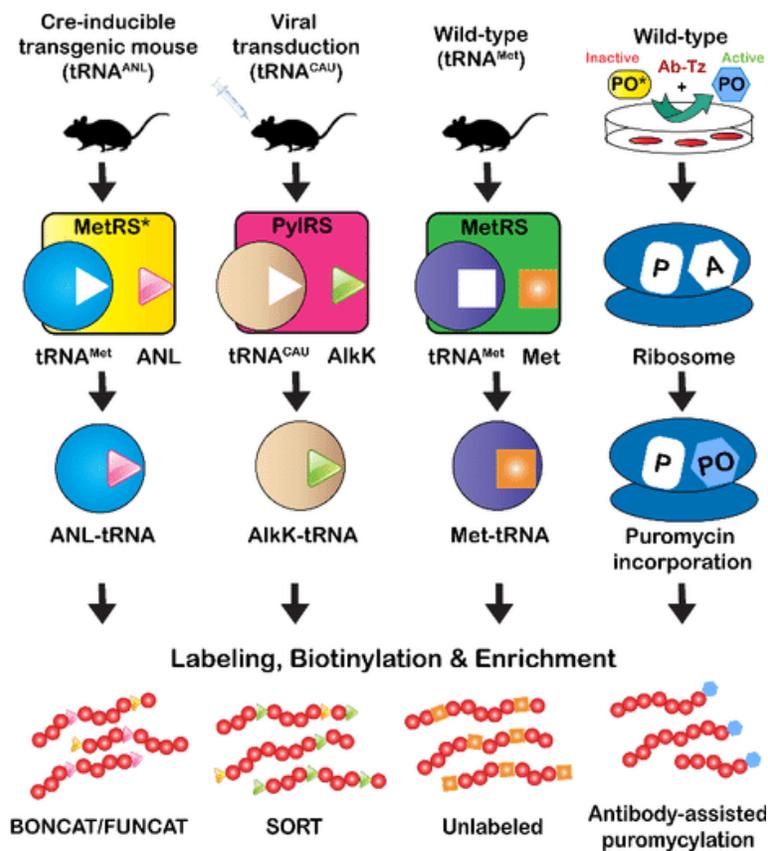
The most recent *in vivo* approaches for cell-type-specific proteomics are based on proteome labeling. These approaches take advantage of a cell's protein synthesis machinery and enable incorporation of a noncanonical amino acid into the proteome of interest. Recently, this method has transitioned to cell-type-specific labeling of proteomes through bioorthogonal non-canonical amino acids azidonorleucine (ANL) or azidohomoalanine (AHA) that are used as substitutes for methionine in protein translation. ANL and AHA have exposed azide groups, which upon integration of ANL or AHA into proteins can conjugate to a biotin or fluorescent tag by an azide-alkyne cycloaddition reaction, referred to as click reaction and imaged or isolated by affinity purification. Application of ANL or AHA is combined with the expression of a mutant methionyl-tRNA synthetase (MetRS) activating these non-canonical amino acids resulting in their coupling to the correspondent tRNA<sup>met</sup> due to an enlarged substrate binding pocket (Müller et al. 2015). The technique is called bioorthogonal non-canonical amino acid tagging (BONCAT) and has successfully been applied to label nascent proteomes of *Escherichia Coli* (*E. coli*), *Caenorhabditis elegans* (*C. elegans*), *Drosophila melanogaster* (Müller et al. 2015) and has most recently been used in a mouse model (Alvarez-Castelao et al. 2017, 2019). To enable *in vivo* cell specific nascent proteome labeling, a floxed-STOP cassette was placed upstream of GFP-2A-mutant MetRS for conditional expression under a CAG enhancer-promoter inside the Rosa26 locus. For cell-type-specific labeling, this mouse has to be crossed with different Cre lines. For protein purification, the labeled proteins are clicked with a cleavable biotin-alkyn tag, and enriched by affinity purification and selectively eluted by the cleavage and reduction of the alkyne (Alvarez-Castelao et al. 2017, 2019).

Another recent method for cell-type-specific analysis is called stochastic orthogonal recoding of translation (SORT) (Krogager et al. 2018). It is also based on metabolic labeling of proteins using non-canonical amino acids containing bioorthogonal N-(propargyloxycarbonyl-L-Lysin, AikK which is activated by orthogonal pyrrolysyl-tRNA<sup>xxx</sup> pair (where xxx indicates the anticodon sequence and the pair is derived from *Methanosarcina* species). This pair recognizes the AikK and decodes a specific sense codon in competition with the endogenous aminoacyl-tRNA synthetase-tRNA<sup>xxx</sup> pair, resulting in tagging of the proteome with AikK. The pyrrolysyl-tRNA<sup>xxx</sup> is transferred to the mouse via viral transduction, and in the presence of AikK labeling of the proteome is achieved. The tagged proteins can later be labeled or purified by click chemistry (Krogager et al. 2018).

## Introduction

The challenges with BONCAT and SORT techniques include variability in MetRS expression levels depending on the cell-type-specific promoter used (BONCAT) and transduction efficiency and accuracy of injection (SORT). Low incorporation rates and artificial nature of amino acids being introduced into the proteome of a living organism, and hydrolysis of AHA over time are other problems with these techniques (Elliott et al. 2014; R. S. Wilson and Nairn 2018). These reactions are based on extensive chemical modifications which occur in case of SORT and BONCAT during 7 and 21 days of supplementation of non-canonical amino acids in water to mice. These un-natural amino acids can alter the physiological conditions and introduce variation and bias to the biological system (Alvarez-Castelao et al. 2017; Krogager et al. 2018).

Another labeling approach uses puromycin analogue tags and takes advantage of the cell's native protein synthesis machinery, the ribosome. Puromycin binds the acceptor (A) site of the ribosome and is incorporated into the nascent polypeptide chain before inhibition of the protein synthesis (Du et al. 2017). Although puromycin is less harmful than artificial amino acids for labeling, to date the technique has only been demonstrated in cell culture.



**Figure 2-3. Cell-type-specific labeling approaches.**

Cell type specific labeling is based on incorporation of a non-canonical amino acid or puromycin into the target proteome. The non-canonical amino acids are activated through cell-type-specific tRNA synthase expression accomplished either by genetic engineering of a Cre mouse (BONCAT/FUNCAT) or through viral transduction (SORT). The artificial amino acids are further biotinylated and enriched prior to mass spectrometry analysis (BONCAT/FUNCAT). The other labeling technique is based on introduction of a cell-type specific enzyme tagged antibody (Ab-Tz) and an inactive puromycin analog which gets activated upon reaction with Tz and labels nascent polypeptides which can be puromycylated and identified by mass spectrometry (Adapted from Wilson and Nairn, 2018).

## 2.2. Techniques for Organelle Specific Proteomics

The well-established approaches for isolation of organelles include differential/density gradient centrifugation (Breuza et al. 2004; Knoblach et al. 2003; Mears et al. 2004; Pisitkun, Shen, and Knepper 2004; Yan, Aebersold, and Raines 2009) and free-flow electrophoresis (Prokisch et al. 2004; Zischka et al. 2006). These techniques are typically prone to contamination by other organelles since they overlap in the physicochemical properties used for enrichment of the organelle of interest. Methods like protein correlation profiling (PCP) and digital signature (used for generating the localization of organelle proteins by isotope tagging (LOPIT) platform) have improved the quality of organelle assignments by data analysis but did not improve the fraction approach as such (Andersen et al. 2003; Dunkley et al. 2004). In LOPIT approach the distribution across multiple density gradient is measured for markers of different organelles and subcellular localization is assigned by comparing profiles of other proteins to those of specific markers (Geladaki et al. 2019).

Another technique for organelle purification is immunopurification using antibodies (Satori et al. 2013). A limiting factor for this approach is availability of antibodies against proteins localized on the surface of organelles. Another method that applies flow cytometer to detect and sort organelles based on fluorescence characteristics is fluorescence-activated organelle sorting (FAOS). This method is limited to availability of organelle specific fluorescent probes including fluorescently labeled proteins, antibodies, and chemical reagents (Satori et al. 2013; R. B. Wilson and Murphy 1989) Optical tweezers (OT) and Laser capture microdissection (LCM) techniques are based on using laser beam on a target object to manipulate/capture it and have been applied to isolation of organelles (Pflugrad et al. 2011).

The analysis of subcellular proteomes by targeted labeling of an organelle *in vitro* is another approach which utilizes proximity labeling enzymes that generate reactive radicals to covalently tag neighboring proteins with biotin. The main enzymes used for proximity labeling are proximity-dependent biotin identification (BioID) and engineered ascorbate peroxidase (APEX). They present an advantage over classical approaches, as almost exclusively proteins in the organelle are labeled and enriched and contaminations from other organelles are reduced.

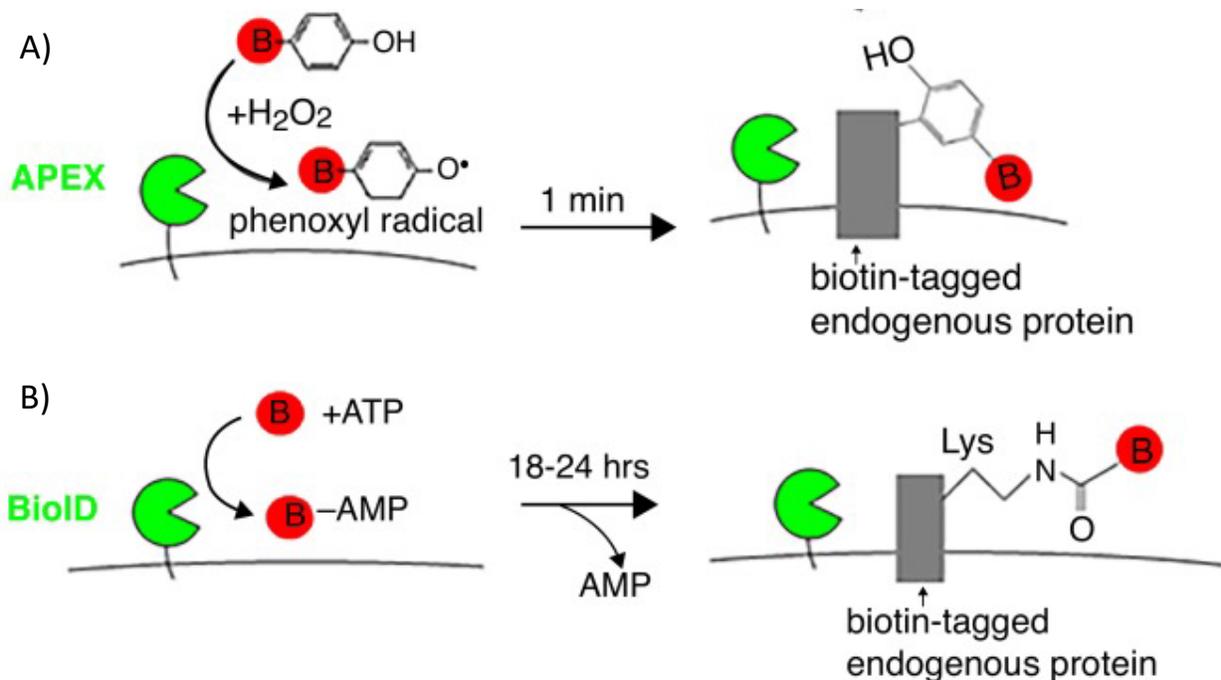
## 2.3. Proximity Labeling using BioID and APEX

Proximity labeling techniques rely on the expression of a biotinylating enzyme fused to a protein or subcellular signal sequence and identification of the interaction partners or proteins localized in a subcellular compartment. One of these methods, APEX, uses an engineered ascorbate peroxidase fusion protein for biotinylation. APEX fused to mitochondrial targeting sequences has been applied to map the proteome of the mitochondrial matrix and intermembrane space in mammalian cells (Hung et al. 2014). APEX has also been used in *C. elegans* to identify tissue-specific and subcellular localized proteomes (Reinke et al. 2017). It also has been used to characterize the proteome of primary cilia (Mick et al. 2015), Golgi in yeast (Hwang and Espenshade 2016), and ER-plasma membrane junctions in living cells (Jing et al. 2015) and in *Drosophila melanogaster* for mapping of mitochondrial proteins (C.-L. Chen et al. 2015). In a recent study APEX2, the recent generation of APEX, was used in dissected rat cortical neuron cultures to map proteins at inhibitory and excitatory synaptic clefts (Cijssouw et al. 2018). APEX in the one min tagging time generates short-lived free radicals such as phenolic aryl azide derivatives or tyramide derivatives which in presence of H<sub>2</sub>O<sub>2</sub>, covalently link to amino acids. These free radicals have limited permeability and high toxicity to living cells and therefore APEX is not the optimal method for live-animal proteome labeling (Branon et al. 2018).

## Introduction

BioID or BirA\* is a mutant form of BirA (BirAR118G); BirA is a 35-kD DNA-binding biotin protein ligase in *E. coli*, which specifically biotinylates a subunit of acetyl-CoA carboxylase or a similar synthetic peptide sequence termed Avi-tag (Chapman-Smith et al. 2001; Knowles 1989). BirA has been employed for the enrichment of membrane proteins, the affinity purification of newly synthesized proteins, and screening of DNA-protein interactions (Van Hoang et al. 2003; Landgraf et al. 2015; Shoaib et al. 2013). Additionally, this approach has been used for the cell-type-specific isolation of organelles from whole tissues (Chitikova and Steiner 2016).

BioID enables on the other hand biotinylation of proteins non-specifically and in a proximity-dependent fashion in the cell (Roux et al. 2012). BioID is fused to the protein of interest or organelle-targeting signal sequence and expressed in cells. BioID has a slow tagging time of 15-24 h. It catalyzes the conversion of biotin to biotinoyl-5'-adenosine monophosphate (AMP). This reactive form of biotin then covalently binds the lysine residues of vicinity proteins, in a radius of 10 nm (Trinkle-Mulcahy 2019).



**Figure 2-4. Proximity labeling methods.**

Proximity labeling is divided into two main categories based on the enzyme used to do the catalysis: Peroxidase based (APEX) and biotin ligase (BioID) based proximity labeling. A) APEX is an engineered ascorbate peroxidase which oxidizes biotin phenol into phenoxyl radical in presence of H<sub>2</sub>O<sub>2</sub> reacting with nearby proteins at electron rich side chains. B) BioID requires biotin and ATP as substrates to generate a reactive form of biotin which tags the lysine residues of proximal proteins (Adapted from Han et al., 2018).

BioID fused to proteins has been used in several studies for characterization of protein-protein interactions in different compartments/cell types such as the nuclear envelope (Roux et al. 2012), the nuclear pore complex (D. I. Kim et al. 2014), centrosomes (Firat-Karalar et al. 2014), Trypanosoma bilobe components and flagella (Morriswood et al. 2013), and cell junctions and caveolae (Mendoza-Topaz et al. 2018).

To enhance the selective targeting of biotin ligase a smaller version of BioID, named BioID2 has been developed which requires less biotin (D. I. Kim et al. 2016). Recently, two new-generation tools, namely TurboID (TbID) and miniTurbo (mTb) with higher catalytic efficiency and shorter proximity labeling time

## Introduction

(~10 min) have been created and used in mitochondrial, nuclear, and ER proteome mapping. While original BioID captures protein associations over time and reports a history of events, the new generations enable an analysis of events occurring over short time periods (Branon et al. 2018).

Biotin is a vitamin and has minimal adverse effects on complex living organisms (Koutsikos, Agroyannis, and Tzanatos-Exarchou 1990). BirA and BioID have also been applied to *in vivo* systems. Through the cell-type-specific expression of either BirA or a tagged version of the protein of interest, organelles from distinct cell populations in a complex tissue can be isolated. Siska et al. generated a mouse strain that ubiquitously expresses BirA from the Rosa26 promoter; upon crossing with any transgenic mouse expressing an Avi-tagged protein, BirA specifically adds biotin to the protein of interest (Driegen et al. 2005). BioID has also been exploited in complex *in vivo* systems to study candidate Myc protein interactors in xenografted tumors (Dingar et al. 2015) and to study synaptic complexes (Spence et al. 2019; Uezu and Soderling 2019) in mice. TurboID has been applied to living organisms including yeast, *Drosophila melanogaster*, and *C. elegans* (Branon et al. 2018).

After supplementation with biotin, the vicinity proteins biotinylated by APEX or BioID get biotinylated and can be enriched by avidin resins. The mechanism of biotin enrichment is based on the extraordinary interaction between biotin and avidin.

### 2.4. Enrichment based on Avidin-biotin interaction

Avidin and streptavidin, as well as their available analogues, Neutravidin and CaptAvidin, are tetrameric proteins which bind up to four biotin molecules with a very high affinity ( $K_a 10^{15} M^{-1}$ ) (Figure 5-1; (Livnah et al. 1993)). For more than 50 years, this extraordinary strong interaction has been exploited for the purification and detection of a variety of molecules such as DNA, RNA, and proteins (Ren et al. 2015).

To date, several avidin-coupled secondary detection reagents have been established (e.g. Horseradish peroxidase (HRP) and fluorophores) and are used routinely in affinity cytochemistry, cell cytometry, and blotting technologies (Rao, Anderson, and Bachas 1997; Rösli et al. 2008). The biotin-avidin interaction is also extensively utilized in a variety of affinity chromatography approaches, e.g. in form of solid supports comprising agarose/sepharose beads (Hirsch et al. 2002), as well as resins (Rösli et al. 2008), chips (Hutsell et al. 2010), or plates (Kay, Thai, and Volgina 2009) coated with immobilized avidin or its derivatives. The application of these complexes has been extended to include other fields in molecular biotechnology such as targeted drug delivery (Lesch et al. 2010), avidin-based sensors (Fang et al. 2008), cross-linking (Moody et al. 2015), live-cell imaging (J. Zhang et al. 2002), and a variety of proteomics applications (Dundas, Demonte, and Park 2013; Wilchek and Bayer 1988). In proteomics, the avidin-biotin interaction is frequently used for the enrichment of protein sub-populations followed by mass spectrometric analysis (Schiapparelli et al. 2014).

### 2.5. Mass spectrometry-based proteomics

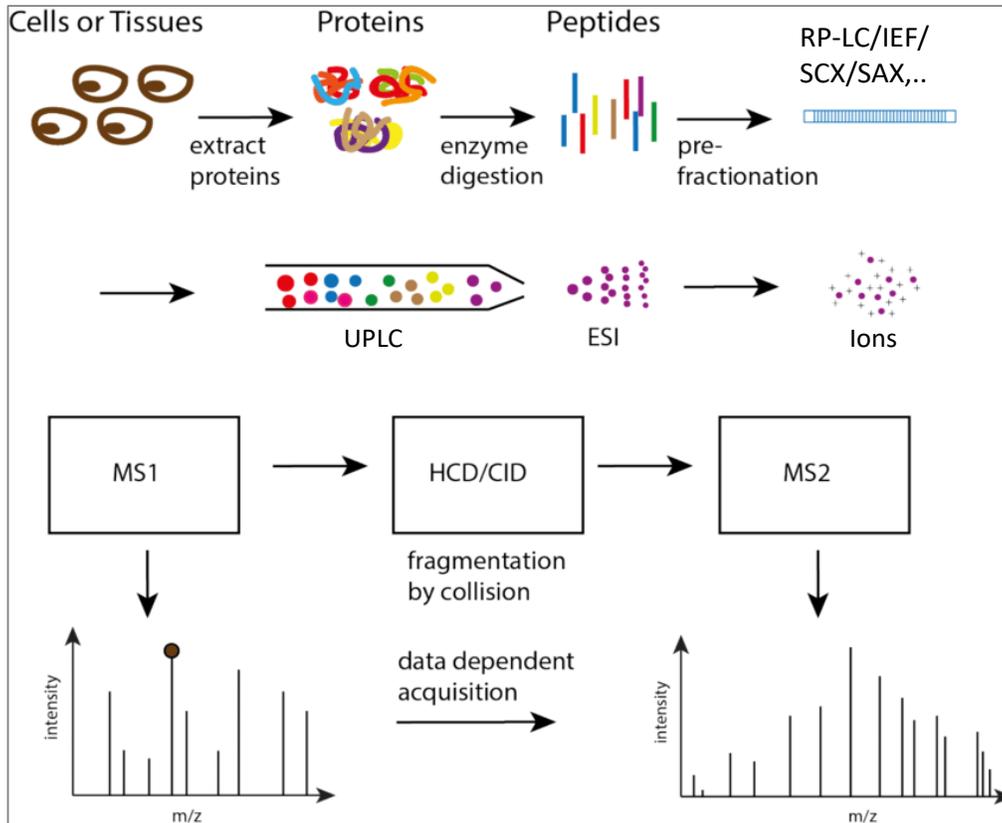
Among the proteomics methods dealing with complex protein samples, mass spectrometry (MS) has established itself as the dominant technique for interpretation of the complex proteins samples. MS-based proteomics has been enabled by the availability of genomic databases and development of protein ionization techniques (Han, Aslanian, and Yates 2008).

MS-based proteomics is a technique for separation of electrically charged ions originating from proteins in the gas phase and their injection and acceleration in an electric or magnetic field. The main methods for ionization of the proteins are electrospray ionization (ESI; (Fenn et al. 1989)) and matrix-assisted laser

## Introduction

desorption/ionization (MALDI; (Karas and Krüger 2003)). In ESI the proteins are in solution and in MALDI, they are embedded within a solid matrix and ions are generated by laser light. While ESI produces multiply charged ions allowing better fragmentation of high mass proteins and peptides, MALDI mostly generates singly charged molecules and is preferred for quick identification of the molecules (Trimpin et al. 2010).

The two main proteomics approaches used for characterization of the proteins are "top-down" and "bottom-up" strategies. While top-down is based on ionization of intact proteins, in bottom-up approach proteins are first enzymatically digested into peptides using proteases prior to introduction to the mass spectrometer (Y. Zhang et al. 2013).



**Figure 2-5. Bottom up mass spectrometry-based proteomics workflow**

In LC-MS-based bottom up proteomics, proteins extracted from cells or tissues are first subjected to enzymatic cleavage, possibly fractionated, and applied on a liquid chromatography (LC) device connected to the mass spectrometer. Peptides eluting from the LC system are ionized using the electrospray generated by electrospray ionization (ESI) technique. The precursor ions for the abundant peaks are fragmented by (higher energy) collision induced dissociation (HCD/CID) followed by recording the  $m/z$  ratio of each fragment (MS/MS or MS2). RP-LC: reversed-phase liquid chromatography; IEF: isoelectric focusing; SCX: strong cation exchange chromatography; SAX: strong anion exchange chromatography; UPLC: ultra performance liquid chromatography (Adapted from Zhu, 2018).

In a common bottom-up workflow, after tryptic digestion of proteins, the resulting peptides are first ionized in an ion source and transferred to the gas phase and transported to a mass analyzer. Mass analyzer applies the electric or magnetic field to separate ions of different mass-to-charge ( $m/z$ ) values. Certain precursor ions are selected and fragmented into product ions by a collision technique such as collision-induced dissociation (CID), higher-energy collisional dissociation (HCD) or electron transfer dissociation (ETD). The MS/MS

## Introduction

spectra resulted from the fragmentation of the peptides in the mass spectrometer are finally compared to theoretical spectra of in-silico digest of a protein database to identify the amino acid sequence of the peptides and extrapolate the identities of the proteins in the sample (Aebersold and Mann 2016). Proximity labeling approaches are normally coupled with quantitative mass spectrometry to allow measuring the changes in abundance of biotinylated proteins after enrichment with avidin beads.

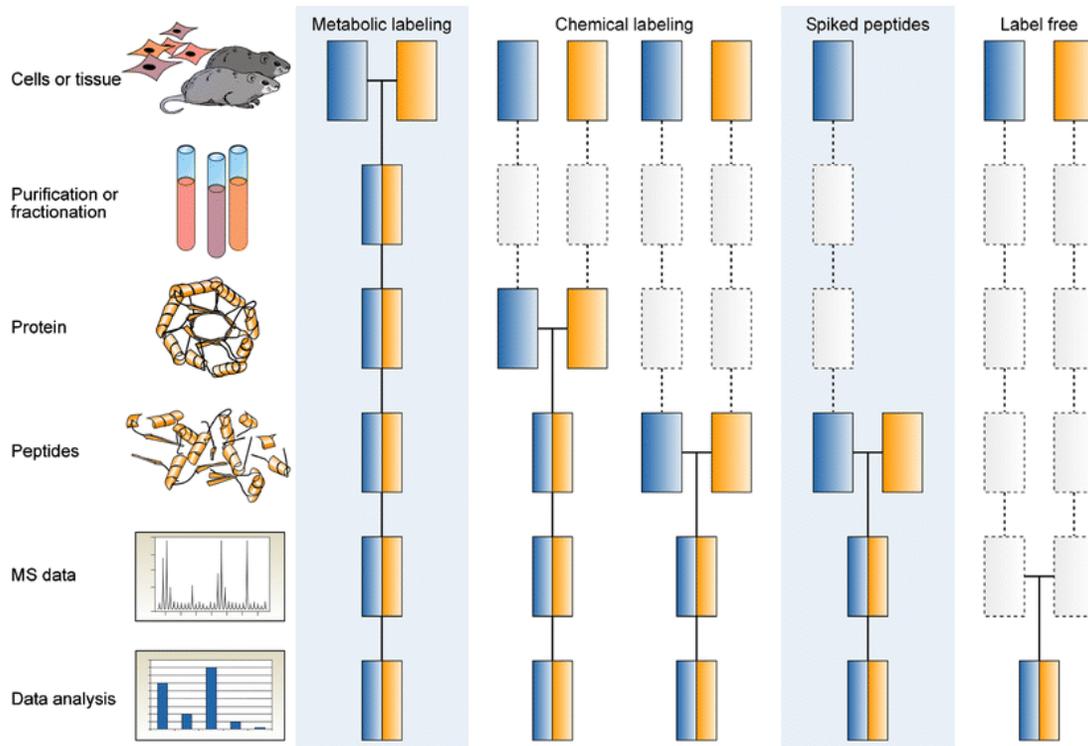
### 2.6. Quantitative Mass Spectrometry

To measure the changes in protein expression or abundance, quantitative strategies can be used. Using these approaches, the abundance of a certain protein in different samples (e.g. biotinylated vs control) can be quantified and compared. Quantitative strategies include two major approaches: label-based and label-free quantification (Bantscheff et al. 2007).

Label based quantification is based on the incorporation of labels into the protein/peptide samples. This is possible by several methods such as stable isotopes ( $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ , and  $^{18}\text{O}$ ) used in metabolic labeling (e.g. stable isotope labeling of amino acids in cell culture (SILAC)), chemical labeling (e.g. dimethyl labeling, tandem mass tag (TMT), isotope-coded affinity tags (ICAT), and isobaric tags for relative and absolute quantitation (iTRAQ)), enzymatic labeling ( $^{18}\text{O}$ -Labeling), and by use of spiked-in peptides (e.g. quantification concatamers (QconCATs) and absolute quantification (AQUA) tags) (Bantscheff et al. 2007). Peptides containing different isotopes have different masses allowing them to be distinguished in the mass spectrometer. SILAC is based on metabolic incorporation of a “light” or “heavy” form of an amino acid (e.g., labeled with deuterium,  $^{13}\text{C}$ , or  $^{15}\text{N}$ ) into all newly synthesized proteins resulting in a mass shift of the corresponding peptides which can be detected by the mass spectrometer (Ong and Mann 2005). SILAC has been widely used in cell culture and model organisms such as mice, rats, worms (Gouw, Krijgsveld, and Heck 2010) and zebrafish (Konzer et al. 2013). SILAC is known to be very accurate and precise for quantification and as samples are labeled and mixed early in the experimental workflow, the technical variability across the samples is minimal. However, it is not applicable to non-dividing cells and is very expensive to be used in living animals (Lau et al. 2014). Chemical labeling is an alternative method, which is based on adding chemical tags encoded with different masses on reactive groups of proteins and peptides, such as primary amines. iTRAQ and TMT reagents have identical overall mass but vary in terms of the distribution of heavy isotopes around their structure which create reporter ions upon fragmentation and enable quantification at MS2 level (Bantscheff et al. 2007). While multiplexing provides sampling depth, quantitative accuracy is reported to be lower with isobaric tags in complex mixtures due to co-elution and co-fragmentation of interfering precursor ions (Bantscheff et al. 2007). Stable isotope dimethyl labeling is another chemical labeling approach to tag primary amines (i.e., the N-terminus and the  $\epsilon$ -amino group of lysine) in proteins or peptides through reductive amination and quantification at MS1 level (Boersema et al. 2009). In comparison to multiplexed chemical labeling, dimethyl chemicals are cost-effective, but this approach does not allow more than 5 samples to be co-analyzed (Wu et al. 2014). iTRAQ, TMT, and dimethyl labels are mostly used at peptide level and in comparison with protein level labeling, the samples are mixed at later stages of the workflow and therefore, are more prone to technical variation. They are, however, more cost-effective than labeling the whole animal and allow effective labeling of tissue samples after digestion of the organs (Bantscheff et al. 2007). Another powerful approach to determine absolute protein/peptide amount is spiking isotopic-labeled AQUA and QconCAT internal standards with a known concentration into the sample (Brun et al. 2007). With prior information about the amount of the internal standard in the sample, comparing the intensities of the standard

## Introduction

versus endogenous peptide, one can determine the absolute amount of the endogenous peptide in that sample. So far the scale of these techniques ranges from only a few peptides of interest (for AQUA) to 4000 of proteins (for QconCATs) and therefore they not yet best suited for large scale proteomics studies (Bantscheff et al. 2007; Brownridge et al. 2011).



**Figure 2-6. Stages of incorporation of stable isotope labels in quantitative proteomics experiments.**

Blue and orange boxes represent experimental conditions to be compared (like experiment and control). Horizontal lines indicate the step at which samples are combined and the dashed lines show the points at which technical variations can be introduced. A) metabolic labeling strategies have the benefit of early stage combination and reduction of technical variations and undesired losses. B) chemical labeling allow samples to be combined at the peptide or protein level C) in spike-in methods synthetic peptides are used to obtain absolute quantification of specific peptides/proteins. D) in label free experiments, each sample is measured in a separate mass spectrometry (MS) measurement and compared (adopted from Bantscheff et al., 2007).

In the second approach, label-free quantification, the quantitative information is inferred with two strategies: (1) spectral counting that implies a counting and a comparison of the number of fragment-ion spectra (MS/MS) acquired for peptides of a given protein; (2) measurement of chromatographic peak areas (also termed mass spectrometric signal intensities) of peptide precursor ions. This approach does not acquire extra experimental steps but is prone to technical variation among MS measurements which can affect the results (Elias and Gygi 2010; Zhu, Smith, and Huang 2010).

After quantitative mass spectrometry, the data need to be analysed by proper statistical approaches for correct interpretation of the results. This is of particular importance for the analysis of biotin enriched proteomes as the enrichment of biotinylated proteins by widely used approach (on-bead digestion) results in a pool of unmodified proteins of two origins. Biotinylated proteins bind the streptavidin-coated beads with high affinity and a large number of proteins bind non-specifically to these affinity resins. Therefore, proper data analysis

## Introduction

approaches are required to distinguish between truly biotinylated proteins and those binding nonspecifically to the beads.

### **2.7. Data analysis of biotin enriched proteins after quantitative mass spectrometry**

Normally the experimental setting for biotin enrichment analysis is as follows: the biotinylated proteins and the control sample are labeled with different chemical/metabolic labeling reagents, combined and analyzed. Intensities of precursor ions with different masses (SILAC/Dimethyl labeling) or of different reporter ions (TMT/iTRAQ labeling) for biotinylated and control samples allow relative comparison between them after biotin-avidin enrichment. In order to distinguish between the biotinylated proteins and the background population several approaches exist. (1) Subtraction of the background protein IDs from the list of enriched proteins; this approach has the risk of removing proteins in the control list which truly reside in the region of interest and bind the avidin resin (Andersen and Mann 2006). (2) Application of an arbitrary cut-off; this is a widely applied method to separate two populations in a list based on biotin/control ratios. However, since the cut-off has no statistical basis this method is prone to acceptance of non-biotinylated proteins as well as the removal of the true positive hits from the protein list (M. S. Kim, Zhong, and Pandey 2016). (3) Cut-off based on GO analysis or “slide approach”; this approach is recently used in studies which use APEX/BioID for mapping of organelle proteomes. Out of the proteins’ list being studied, based on GO analysis and information about the localization of proteins, two lists are created: one list that contains true positives (expected in the studied organelle) and one list of true negative proteins (not belonging to the reference organelle). In other words, these lists are created based on prior knowledge on the probability of presence/absence of proteins in the organelle of interest (Branon et al. 2018). A biotin/control cut-off is slid stepwise on the ratio distribution of these two populations until the majority of proteins known to be part of the population of interest are above the cut-off. This is, however, only applicable to compartments and structures for which prior experimental data is available.

### **2.8. Aims of the study**

The goal of this study is to develop an approach to generate a cut off with a statistical basis to distinguish biotinylated proteins from those binding non-specifically to the beads. We aim to define a cut-off on 95<sup>th</sup> percentile of control/control ratios using untreated HEK 293T cells as part of a null experiment. We will next apply it to proteomes of different cellular compartments that are enriched by organelle-targeting BioID constructs. To verify the credibility of this cut-off we will compare it to that determined by slide approach and based on previous knowledge of the expected proteins in these organelle datasets.

Besides application of BioID to *in vitro* analysis, BioID is planned to be used *in vivo*, incorporated into the genome of two mouse models. We aim to evaluate the functionality of these genetic systems and examine the expression and biotinylation in different tissues of these mice. We also aim to direct the expression of BioID to specific cell types of these mice and perform cell-type-specific proteomics for the first time using BioID.

### 3. Materials

#### 3.1. Chemicals/Reagents

All mass spectrometry-related chemicals were HPLC or ULC/MS grade.

**Table 3-1. list of chemicals/reagents**

Chemicals/Reagent	Provider
2-Mercaptoethanol	Sigma-Aldrich (Munich, Germany)
Acetic acid, glacial	Biosolve (Valkenswaard, Netherlands)
Acetonitrile	Biosolve (Valkenswaard, Netherlands)
Acrylamide	Sigma-Aldrich (Munich, Germany)
Ammonium bicarbonate	AppliChem GmbH (Darmstadt, Germany)
Ammonium hydroxide	Sigma-Aldrich (Munich, Germany)
Ammonium persulfate (APS)	Sigma-Aldrich (Munich, Germany)
Bovine serum albumin	Sigma-Aldrich (Munich, Germany)
Bromophenol Blue	Sigma-Aldrich (Munich, Germany)
Calcium chloride	AppliChem GmbH (Darmstadt, Germany)
Clarity™ Western ecl substrate	Bio-Rad Laboratories, Inc. (Hercules, CA, USA)
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich (Munich, Germany)
Dipotassium phosphate	AppliChem GmbH (Darmstadt, Germany)
Dithiothreitol	Sigma-Aldrich (Munich, Germany)
DMEM for SILAC	Thermo Fisher Scientific (Bremen, Germany)
Doxycycline Hyclate	Sigma-Aldrich (Munich, Germany)
Dulbecco's modified eagle's medium/F12	Invitrogen Gibco (Paisley, United Kingdom)
Dulbecco's modified eagle's medium –glutamine – pyruvate	Invitrogen Gibco (Paisley, United Kingdom)
Dulbecco's phosphate-buffered saline (DPBS)	Invitrogen Gibco (Paisley, United Kingdom)
Ethanol	AppliChem GmbH (Darmstadt, Germany)
Ethylenediaminetetraacetic acid (EDTA)	Sigma-Aldrich (Munich, Germany)
Fetal calf serum	Invitrogen Gibco (Paisley, United Kingdom)
Fluoromount-G™, with DAPI	Thermo Fisher Scientific (Bremen, Germany)
Formic acid	Biosolve (Valkenswaard, Netherlands)
Glycerol	Merck KGaA (Darmstadt, Germany)
Glycolic acid	Sigma-Aldrich (Munich, Germany)
HEPES (2,4 Hydroxyethyl piperazineethanesulfonic acid)	Merck KGaA(Darmstadt, Germany)
Iodoacetamide	Thermo Fisher Scientific (Bremen, Germany)
L-ARGININE: HCL (1-13C, 99%)	Cambridge Isotope Laboratories, Inc. (Tewksbury, MA, USA)
L-ARGININE: HCL ( <sup>13</sup> C <sub>6</sub> , 99%)	Cambridge Isotope Laboratories, Inc. (Tewksbury, MA, USA)
L-ARGININE: HCL ( <sup>13</sup> C <sub>6</sub> , 99%; <sup>15</sup> N <sub>4</sub> , 99%)	Cambridge Isotope Laboratories, Inc. (Tewksbury, MA, USA)
L-glutamine	Sigma-Aldrich (Munich, Germany)
L-LYSINE: 2HCL (1-13C, 99%)	Cambridge Isotope Laboratories, Inc. (Tewksbury, MA, USA)
L-LYSINE: 2HCL (4,4,5,5-D <sub>4</sub> , 96-98%)	Cambridge Isotope Laboratories, Inc. (Tewksbury, MA, USA)
L-LYSINE: 2HCL ( <sup>13</sup> C <sub>6</sub> , 99%; <sup>15</sup> N <sub>2</sub> , 99%)	Cambridge Isotope Laboratories, Inc. (Tewksbury, MA, USA)
Magnesium chloride	Sigma-Aldrich (Munich, Germany)

## Materials

<b>Chemicals/Reagent</b>	<b>Provider</b>
Methanol	Merck KGaA (Darmstadt, Germany)
Mineral oil	GE Healthcare BioSciences (Uppsala, Sweden)
Monopotassium phosphate	AppliChem GmbH (Darmstadt, Germany)
Nonidet P-40	Fluka Chemie Ag (Buchs, Switzerland)
Tissue-Tek OCT compound	Sakura Finetek (Torrance, CA, USA)
PageBlue™ protein staining solution	Thermo Fisher Scientific (Bremen, Germany)
Paraformaldehyde	Sigma-Aldrich (Munich, Germany)
Potassium chloride	AppliChem GmbH (Darmstadt, Germany)
Roche cOMplete Protease Inhibitor Cocktail	F. Hoffmann-La Roche AG (Basel, Switzerland)
Sodium dodecyl sulfate (SDS)	Bio-Rad Laboratories, Inc. (Hercules, CA, USA)
Sodium chloride	Sigma-Aldrich (Munich, Germany)
Sodium fluoride	AppliChem GmbH (Darmstadt, Germany)
Sodium hydrogen phosphate	Merck KGaA (Darmstadt, Germany)
Sodium hydroxide	Carl Roth GmbH (Karlsruhe, Germany)
Sodium orthovanadate	AppliChem GmbH (Darmstadt, Germany)
Sodium pyrophosphate	Sigma-Aldrich (Munich, Germany)
Sodium selenite	Sigma-Aldrich (Munich, Germany)
Streptavidin sepharose high-performance affinity resin	GE Healthcare BioSciences (Uppsala, Sweden)
Streptomycin	Invitrogen Gibco (Paisley, United Kingdom)
Tetramethylethylenediamine (TEMED)	Carl Roth GmbH (Karlsruhe, Germany)
Trifluoroacetic acid	Biosolve (Valkenswaard, Netherlands)
Tris	AppliChem GmbH (Darmstadt, Germany)
TX100	Fluka Chemie Ag (Buchs, Switzerland)
Trypan blue solution 0.04%	Sigma-Aldrich (Munich, Germany)
TurboFect transfection reagent	Thermo Fisher Scientific (Bremen, Germany)
Urea	Sigma-Aldrich (Munich, Germany)
Water	Biosolve (Valkenswaard, Netherlands)
Biotin	Sigma-Aldrich (Munich, Germany)
Tween20	Carl Roth GmbH (Karlsruhe, Germany)
Immobiline DryStrip Gels	GE Healthcare BioSciences (Uppsala, Sweden)
IPG-Buffer, pH 3-10	GE Healthcare BioSciences (Uppsala, Sweden)
Dimethyl labeling materials	Cambridge Isotope Laboratories, Inc. (Tewksbury, MA, USA)
Sodium deoxycholate	Sigma-Aldrich (Munich, Germany)
Thiourea	Sigma-Aldrich (Munich, Germany)
DC™ Protein Assay	Bio-Rad Laboratories, Inc. (Hercules, CA, USA)

## 3.2. Enzymes

**Table 3-2. list of Enzymes**

<b>Enzyme</b>	<b>Provider</b>
Conventional restriction enzymes	Thermo Fisher Scientific (Bremen, Germany)
FastDigest restriction enzymes	Thermo Fisher Scientific (Bremen, Germany)
Phusion high-fidelity DNA polymerase	Thermo Fisher Scientific (Bremen, Germany)
Sequencing grade modified trypsin, porcine	Promega (Madison, WI, USA)
Mass Spec grade rLysC	Promega (Madison, WI, USA)
T4 DNA ligase	Thermo Fisher Scientific (Bremen, Germany)

## Materials

Enzyme	Provider
Taq DNA polymerase, recombinant	Thermo Fisher Scientific (Bremen, Germany)

### 3.3. Cell lines

**Table 3-3. List of cell lines**

Cell line/Strain	Species	Explanation
XL-1 blue	<i>E. Coli.</i>	Competent cells for cloning
HEK 293T	Human	Derivative of HEK 293T cells; stably transfected with SV40 Large T-antigen
NIH/3T3	Mouse	A continuous cell line of high contact-inhibition established from NIH Swiss mouse embryo culture

### 3.4. Mice

The C57BL6 wild type mice used in this study were obtained from in-house animal facility. The experimental mice were housed in accordance with guidelines of local and state authorities regarding animal welfare

**Table 3-4. List of mice**

Mouse	Original Publication	Background Strain	Producer or Provider	Animal application Nr.
Rosa26-CAG-BioID	(Madisen et al. 2010)	CD1	Generated in collaboration with the Institute for physiology 1; University of Bonn, Germany	Az 84-02.04.2014.A069
Col1a1-BioID	(Beard et al. 2006)	C57Bl6	Generated in collaboration with the Institute for developmental pathology; University of Bonn, Germany	Az 84-02.04.2014.A069
Rosa26-Stop-tTA	(L. Wang et al. 2008)	C57Bl6	Jackson Labs Stock No: 008600	Az 84-02.04.2014.A069
Rosa26-rtTA	(Yu et al. 2005)	C57Bl6	Jackson Labs Stock No: 016999	Az 84-02.04.2014.A069
Ins1-Cre	(Thorens et al. 2015)	C57Bl6	Jackson Labs Stock No: 026801	Az 84-02.04.2017.A346
CMV-Cre	(Schwenk, Baron, and Rajewsky 1995)	C57Bl6	Jackson Labs Stock No : 006054, provided by Deutsches Zentrum für Neurodegenerative Erkrankungen e. V. (DZNE), Bonn, Germany	Az 84-02.04.2014.A069

### 3.5. Primers

**Table 3-5. List of primers**

REV: reverse; FWD: forward; bp: base pair; MUT: mutant; WT: wild type

Primer	Application	Gene/Mouse	Sequence	Size
BioID FWD	Genotyping	Col1a1 BioID	ACACCGGGACCGATCCAGCCTCC	600 bp

## Materials

BioID REV	Genotyping	Col1a1 BioID	CACCAGGATGCCAGCC	
Primer	Application	Gene/Mouse	Sequence	Size
Col1 mut FWD	Genotyping	Col1a1 BioID	GCAGAAGCGCGGCCGTCTGG	MUT = ~500 bp WT = 331 bp
Col1 wt FWD	Genotyping	Col1a1 BioID	GCACAGCATTGCGGACATGC	
Col1 common REV	Genotyping	Col1a1 BioID	CCCTCCATGTGTGACCAAGG	
R26 rttWT Rev	Genotyping	Rosa26 rtTA	GGAGCGGGAGAAATGGATATG	MUT = ~350 bp WT = 600 bp
R26 rtt common FWD	Genotyping	Rosa26 rtTA	AAAGTCGCTCTGAGTTGTTAT	
R26 rtt mt REV	Genotyping	Rosa26 rtTA	GCGAAGAGTTTGTCTCAACC	
R26 tTA common FWD	Genotyping	Rosa26 tTA	AAG GGA GCT GCA GTG GAG TA	MUT = ~240bp WT = 300bp
R26 tTA wt REV	Genotyping	Rosa26 tTA	CCGAAAATCTGTGGGAAGTC	
R26 tTA mut REV	Genotyping	Rosa26 tTA	TCA TCA AGG AAA CCC TGG AC	
Ins1 Cre wt FWD	Genotyping	Ins1 Cre	GTC AAA CAG CAT CTT TGT GGT C	MUT = 675 bp WT = 488 bp
Ins1 Cre mt FWD	Genotyping	Ins1 Cre	GCT GGA AGA TGG CGA TTA GC	
Ins1 Cre common REV	Genotyping	Ins1 Cre	GGA AGC AGA ATT CCA GAT ACT TG	
CMV-Cre Forward	Genotyping	CMV-Cre	GCG GTC TGG CAG TAA AAA CTA TC	MUT = ~100 bp, Internal positive control = 324 bp
CMV-Cre Reverse	Genotyping	CMV-Cre	GTG AAA CAG CAT TGC TGT CAC TT	
Internal Positive Control Forward	Genotyping	CMV-Cre	CTA GGC CAC AGA ATT GAA AGA TCT	
Internal Positive Control Reverse	Genotyping	CMV-Cre	GTA GGT GGA AAT TCT AGC ATC ATC C	
CAG Rosa fwd wt	Genotyping	Rosa26 CAG BioID	AAGGGAGCTGCAGTGGAGTA	MUT = 195 bp, WT = 297 bp
CAG Rosa rev	Genotyping	Rosa26 CAG BioID	CCGAAAATCTGTGGGAAGTC	

## Materials

wt				
CAG WPRE rev mt	Genotyping	Rosa26 CAG BioID	GGCATTAAAGCAGCGTATCC	
CAG BioID fwd mt	Genotyping	Rosa26 CAG BioID	AGGACGGAATCATCAAGCC	
<b>Primer</b>	<b>Application</b>	<b>Gene/Mouse</b>	<b>Sequence</b>	<b>Size</b>
NheI BioID FWD	Cloning	BioID	AAAGCTAGCATGGAACAAAACTCATCTCAGAAG	996 bp
XmaI BioID Rev	Cloning	BioID	TTTCCCGGGCTACTTCTCTGCGCTTCTCAGG	
EcoNI BioID FWD	Cloning	BioID-3xNLS	AAACCTNNNNNAGGCTGGCATCAACCTGGACC	1072 bp
3xNLS ECoRV REV	Cloning	BioID-3xNLS	CTCGAGTTAAACCTTCTCTTCTTCTTAGGGTCAAC CTTCTCTTCTTCTTAGGGTCAACCT TCTCTTCTTCTTAGG CTTCTCTGCGCTTCTCAGG	
NheI SU9 FWD	Cloning	SU9	AAAGCTAGCATG GCCTCCACTCGTGTC	210 bp
NheI SU9 REV	Cloning	SU9	AAAGCTAGCGACCGGTGAGTAGGCGC	

## 3.6. Primary Antibodies

**Table 3-6. List of antibodies**

IF: immunofluorescence; WB: Western blotting.

Target name	Manufacturer	Host species	Product No.	Application	Dilution
Myc	Abcam (Cambridge, United Kingdom)	Rabbit	ab9106	WB and IF	WB(1:5000), IF (1:400)
Beta-Actin	Sigma-Aldrich (Munich, Germany)	Mouse	A5316	WB	1:5000
alpha-Tubulin	Rockland Immunochemicals Inc. (Limerick, PA, USA)	Rabbit	600-401- 880	WB	1:2000
GAPDH	Santa Cruze Biotechnology (Heidelberg, Germany)	Rabbit	FI-335	WB	1:1000
alpha-Tubulin	Abcam (Cambridge, United Kingdom)	Mouse	ab195887	IF	1:400
ATP synthase beta monoclonal	Thermo Fisher Scientific (Bremen, Germany)	Mouse	A21351	IF	1:500
Glucagon	Abcam (Cambridge, United Kingdom)	Mouse	ab10988	IF	1:400
Insulin	Dako (Hamburg, Germany)	Guinea Pig	IR002	IF	1:300

## Materials

Target name	Manufacturer	Host species	Product No.	Application	Dilution
Insulin	Abcam (Cambridge, United Kingdom)	Rabbit	ab6995	IF	1:400
Ki67	Thermo Fisher Scientific (Bremen, Germany)	Rabbit	RM9106	IF	1:200

### 3.7. Secondary Antibodies and Chemicals

**Table 3-7. List of secondary antibodies and chemicals**

IF: immunofluorescence; WB: Western blotting.

Antibody Name	Manufacturer	species	Product No.	Application	Dilution
HRP-goat anti-rabbit	Dianova (Hamburg, Germany)	Goat	111-035-003	WB	1:5000
HRP-goat anti-mouse	Dianova (Hamburg, Germany)	Goat	115-035-044	WB	1:5000
Goat IgG anti-rabbit IgG (H+L)-Cy3	Dianova (Hamburg, Germany)	Goat	111-165-144	IF	1:400
MFP488 goat anti-mouse IgG (H+L)	MoBiTec (Goettingen, Germany)	Goat	MFP-A1029	IF	1:400
Cy2-donkey anti-rabbit antibody	Jackson ImmunoResearch (West Grove, PA)	Donkey	711-225-152	IF	1:400
Cy3-goat anti-mouse antibody	Jackson ImmunoResearch (West Grove, PA)	Goat	115-165-146	IF	1:400
Alexa Flour 647 donkey anti-guinea pig antibody	Jackson ImmunoResearch (West Grove, PA)	Guinea pig	706-605-148	IF	1:400
Streptavidin 647	Thermo Fisher Scientific (Bremen, Germany)	-	S21374	IF	1:400
Alexa Fluor 647 streptavidin	Biolegend	-	405237	IF	1:400
Alexa Fluor 421 streptavidin	Biolegend	-	405225	IF	1:400
Streptavidin Protein-HRP conjugate	Thermo Fisher Scientific (Bremen, Germany)	-	21126	WB	1:10000

### 3.8. Instruments and consumables

**Table 3-8. List of instruments and consumables**

Equipment and consumables	Provider
37°C Heraeus incubator	Thermo Fisher Scientific (Bremen, Germany)
Analytical balance CP 124-OCE	Sartorius AG (Göttingen, Germany)
Axio Vert.A1 microscope	Carl Zeiss AG (Oberkochen, Germany)
Axiovert 200M microscope	Carl Zeiss AG (Oberkochen, Germany)
Axygen pipette tips maximum recovery	Corning Inc. (Corning, NY, USA)
Axygen reaction tubes maximum recovery (1.5 ml)	Corning GmbH (Kaiserslautern, Germany)
Biometra T3 thermal cycler	Analytik Jena AG (Jena, Germany)

## Materials

<b>Equipment and consumables</b>	<b>Provider</b>
Cell culture dishes	Sarstedt AG & Co. (Nümbrecht, Germany)
Cell culture well-plates	Corning Inc. (Corning, NY, USA)
Cell scraper	Sarstedt AG & Co. (Nümbrecht, Germany)
Centrifuge 5415D	Eppendorf AG (Hamburg, Germany)
Centrifuge 5424R	Eppendorf AG (Hamburg, Germany)
Centrifuge 5810R	Eppendorf AG (Hamburg, Germany)
Centrifuge Labofuge 400	Thermo Fisher Scientific (Bremen, Germany)
Centrifuge MIKRO 200	Andreas Hettich GmbH & Co. (Tuttlingen, Germany)
CO <sub>2</sub> Water Jacketed Incubator	Thermo Fisher Scientific (Bremen, Germany)
Conical tubes (15 and 50 ml)	Sarstedt AG & Co. (Nümbrecht, Germany)
Corning 40 µm Cell Strainer	Corning Inc. (Corning, NY, USA)
Cover slides, 12 mm	VWR (Leicestershire, UK)
Cryogenic vials	Thermo Fisher Scientific (Bremen, Germany)
Cryostat Leica CM3050S	Leica Microsystems, Wetzlar, Germany
Dry block heater	STAR LAB (Hamburg, Germany)
Empore C18 Disks	3M Corporation (St. Paul, MN, USA)
Freezer -20°C	AEG AG (Berlin, Germany)
Freezer -80°C Ultra low	Sanyo Scientific (Osaka, Japan)
FUSION SOLO 4M System	VilberLourmat (Eberhardzell, Germany)
Galaxy MiniStar microcentrifuge	VWR (Darmstadt, Germany)
Glassware (beakers, bottles, cylinders)	Carl Roth GmbH and Co. KG (Karlsruhe, Germany)
Heating and magnetic stirrer ARE	Velp Scientifica (Usmate, Italy)
Heating Block	Gebr. Liebisch GmbH & Co. KG (Bielefeld, Germany)
Ice machine	Ziegler Ice Machines Ltd. (Isernhagen, Germany)
Infinite 200 PRO NanoQuant Microplate Readers	Tecan (Männedorf, Switzerland)
Incubator, INCU-Line	VWR (Darmstadt, Germany)
Inverted Research Microscope ECLIPSE Ti	Nikon (Tokyo, Japan)
Labtherm Heater	Gebr. Liebisch GmbH & Co. KG (Bielefeld, Germany)
Laminar air-flow device CA R 6	Clean Air (Minneapolis, MN, USA)
Liquid chromatograph EASY-nLC 1000	Thermo Fisher Scientific (Bremen, Germany)
Mass spectrometer LTQ Orbitrap-Velos	Thermo Fisher Scientific (Bremen, Germany)
Mass spectrometer Orbitrap Fusion Lumos	Thermo Fisher Scientific (Bremen, Germany)
nano-UHPLC-Dionex Ultimate 3000 system	Thermo Fisher Scientific (Bremen, Germany)
Nitrocellulose Membrane filters (0.2 and 0.45 µm)	Sarstedt AG & Co. (Nümbrecht, Germany)
Micropipettes	Eppendorf AG (Hamburg, Germany)
Microplate reader GENios	Tecan (Männedorf, Switzerland)
Mini incubator INCU-line	VWR (Darmstadt, Germany)
Mini-PROTEAN Tetra Cell Casting Stand	Bio-Rad Laboratories, Inc. (Hercules, CA, USA)
Multi-Rotator Multi RS-60	BioSan (Riga, Latvia)
Neubauer Hemocytometer (Counting chamber)	Paul Marienfeld GmbH & Co. (Lauda-Königshofen, Germany)
OASIS HLB 10 mg cartridge	Waters Corporation (Milford, MA, USA)
Omnifix syringe 50 ml	B. Braun Melsungen AG (Melsungen, Germany)
P-2000 Laser-Based Micropipette Puller	Sutter Instruments (Novato, CA, USA)
Pasteur Pipettes 230 mm	BRAND GMBH & Co. (Wertheim, Germany)
PC 4400 scale	METTLER TOLEDO (Greifensee, Switzerland)
PerfectBlue™ Semi-Dry Elektrobloetter	VWR (Darmstadt, Germany)

## Materials

Equipment and consumables	Provider
pH-Meter, Calimatic 761	Knick Elektronische Messgeräte GmbH & Co. (Berlin, Germany)
Pipette tips	Greiner Bio-One GmbH (Frickenhausen, Germany)
Polymax1040 orbital shaker	Heidolph Instruments (Schwabach, Germany)
Reaction tubes 1.5 and 2 ml	Sarstedt AG & Co. (Nümbrecht, Germany)
Refrigerator 4°C	AEG AG (Berlin, Germany)
ReproSil-Pur 120 C18 AQ	Dr. Maisch GmbH (Ammerbuch-Entringen, Germany)
Rotamax 120 orbital shaker	Heidolph Instruments (Schwabach, Germany)
Rotamix RM1	ELMI Ltd. (Riga, Latvia)
RS-TR05 roller mixer	Carl Roth GmbH (Karlsruhe, Germany)
Sachtopore NP 5 µm 300 Å TiO <sub>2</sub> bulk material	Sachtleben Chemie GmbH (Duisburg, Germany)
Water Bath	GFL Gesellschaft für Labortechnik mbH (Burgwedel, Germany)
Specimen slide	Engelbrecht Medizin- und Labortechnik GmbH (Erdermünde, Germany)
Syringe needle, BD Microlane	Becton, Dickinson and Company (Franklin Lakes, NJ, USA)
Thermomixer Comfort	Eppendorf AG (Hamburg, Germany)
Branson Ultrasonics™ Sonifier S-250A with 2 mm tip	Branson Ultrasonics (Danbury, CT, USA)
Ultrasonic waterbath 2510	Branson Ultrasonics (Danbury, CT, USA)
Ultimate 3000 HPLC system	Thermo Fisher Scientific (Bremen, Germany)
Vacuum centrifuge RVC 2-18	Martin Christ Gefriertrocknungsanlagen GmbH (Osterode am Harz, Germany)
Vortex, UNIMAG ZX3	UniEquip Laborgerätebau- und Vertriebs GmbH (Leipzig, Germany)
G3100A OFFGEL Fractionator	Agilent Technologies (Ratingen, Germany)

## 3.9. Software and online tools

**Table 3-9. List of software and online tools**

Software or tool	Provider or Website
AxioVision SE64 Rel.4.9.1	Carl Zeiss AG (Oberkochen, Germany)
Biodocanalyze 2.1 gel documentation software	Analytik Jena
FusionCapt Advance Solo 4 16.15	VilberLourmat (Eberhardzell, Germany)
GraphPad Prism 6.07	GraphPad Software Inc. (San Diego, CA, USA)
MaxQuant 1.6.5.0	Max Planck Institute of Biochemistry (Planegg, Germany)
Office Professional Plus 2010	Microsoft Corporation (Redmond, WA, USA)
Thermo Xcalibur 2.2	Thermo Fisher Scientific (Bremen, Germany)
Perseus version 1.6.0.7 and 1.6.10.43	Max Planck Institute of Biochemistry (Planegg, Germany)
BioRender	BioRender.com

## 3.10. Ready to use reaction kits

**Table 3-10. List of ready to use reaction kits**

Reaction Kit Name	Provider
DC™ Protein Assay	Bio-Rad Laboratories, Inc. (Hercules, CA, USA)

## Materials

Reaction Kit Name	Provider
PureLink® Gel Extraction Kit	Invitrogen Gibco (Paisley, United Kingdom)
QIAquick PCR purification kit	Qiagen Inc. (Hilden, Germany)
PureLink® HiPure Plasmid Midiprep Kit	Invitrogen Gibco (Paisley, United Kingdom)
QIAprep Spin Miniprep Kit	Qiagen Inc. (Hilden, Germany)
Pierce Quantitative Fluorescent Peptide Assay	Thermo Fisher Scientific (Bremen, Germany)
Ultra-Sensitive Rat Insulin ELISA Kit	Cristal Chem (Zaandam, Netherlands)

### 3.11. Drugs and Treatments and Mouse Food

**Table 3-11. List of drugs and treatments**

Drug/Treatment	USAGE	Provider
Streptozotocin	Drug	Sigma-Aldrich (Munich, Germany)
Harmine Hydrochloride	Drug	Santa Cruz (Heidelberg, Germany)
Rod 16-A	Normal Mouse food	LASvendi (Soest, Germany)
R/M-H low phytoestrogen 625 mg/kg Doxycyclin (720 mg/kg doxycycline hyclate)	Treatment	Ssniff (Soest, Germany)
Doxycycline Hyclate	Treatment	Sigma-Aldrich (Munich, Germany)
Biotin	Treatment	Sigma-Aldrich (Munich, Germany)

## 4. Methods

### 4.1. Cell culture techniques

#### 4.1.1. Cell culture and labeling of proteins using SILAC amino acids

All cell culture experiments were carried out under the laminar flow hood in sterile conditions. Sterile media, buffers, and solutions were pre-heated at 37°C before use. The cells were incubated in the cell culture incubators with 5% CO<sub>2</sub> at 37°C. HEK 293T and NIH/3T3 cells were cultured in complete growth medium (90% DMEM, 10% FCS, 2 mM L-glutamine, 100 U/ml penicillin, 100 µg/ml streptomycin) in the incubator. Upon reaching confluence, cells were washed with 1 x with PBS and removed from the plate by adding a sufficient amount of 0.05% trypsin (~1 ml to 100 mm and 2 ml to 150 mm dishes) and incubated at 37°C for 3-5 min. Detached cells were mechanically dissociated into single cells by trituration, harvested, and centrifuged at 1,000 g for 5 min. A proportion of the cell pellet resuspended in 1 ml of growth media was then transferred to a new dish containing fresh media. For sustaining culture in 100 mm dishes, HEK 293T cells were split 1:10. For culture in 24-well plates, NIH/3T3 cells were split 1:20 and 1 ml of the cell suspension was added per well. For SILAC labeling, HEK 293T cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) high glucose (4.5 g/l) for SILAC (deficient in arginine and lysine), supplemented with 10% dialyzed FCS, L-glutamine (2 mM), penicillin (100 U/ml), and streptomycin (0.1 mg/ml), and respective amino acids for at least 5 passages. Dependent on the SILAC channel, the medium was supplemented either with light (Arg0, Lys0), medium (Arg <sup>13</sup>C<sub>6</sub>, Lys D<sub>4</sub>), or heavy (Arg <sup>13</sup>C<sub>6</sub><sup>15</sup>N<sub>4</sub>, Lys <sup>13</sup>C<sub>6</sub><sup>15</sup>N<sub>2</sub>) amino acids to a final concentration of 42 mg/l (0.8 mM) and 73 mg/l (0.4 mM) for Arg and Lys, respectively.

#### 4.1.2. Transfection and Biotinylation

The procedure is explained for cells grown in a 6-well/100 mm plate. NIH/3T3 or HEK 293T cells were seeded 1 day prior to transfection at a density to reach 70-90% confluency at the time of transfection. From the DNA sample, 2/10 µg was added to 400/1000 µl of serum-free DMEM and mixed well. TurboFect transfection reagent was vortexed briefly and 3/20 µl was added to the diluted DNA. The tube containing DNA and TurboFect was vortexed immediately and incubated at RT for 20 min. The mixture of transfection reagent and DNA was pipetted dropwise into each well of a 6-well plate/100 mm plate containing 4/10 ml growth media and plates incubated at 37°C. The media was changed with fresh media 6 h after transfection. 24-48 h post-transfection the media was replaced with medium supplemented with 50 µM biotin and incubated at 37°C for 24 h. Cells were harvested by scraping in ice-cold 1 x PBS, pelleted at 1,500 g, and stored until further use at -80°C.

### 4.2. Molecular biology techniques

#### 4.2.1. Polymerase Chain Reaction (PCR)

PCR reactions were carried out using either Phusion High Fidelity DNA Polymerase or Taq DNA polymerase. Pipetting instructions for the reactions mediated by Phusion and Taq DNA polymerase are provided in tables 4.1. and 4.2., respectively. The cycling program for PCR reaction is shown in table 4.3.

## Methods

**Table 4-12. Pipetting instructions for PCR mediated by Phusion DNA polymerase**

\*5x GC Phusion buffer was used for amplification of GC rich templates.

Constituent	Amount	Final Concentration
cDNA/plasmid	10% of cDNA/10 ng reaction mix	1-500 ng
5 x Phusion HF Buffer*	4 $\mu$ l	1 x
2 mM dNTPs	2 $\mu$ l	200 $\mu$ M each
10 $\mu$ M Forward Primer	1 $\mu$ l	0.5 $\mu$ M
10 $\mu$ M Reverse Primer	1 $\mu$ l	0.5 $\mu$ M
2 U/ $\mu$ l Phusion DNA Polymerase	0.2 $\mu$ l	0.02 U/ $\mu$ l
H <sub>2</sub> O	Up to 20 $\mu$ l	-

**Table 4-13. Pipetting instructions for PCR mediated by Taq DNA polymerase**

Constituent	Amount	Final Concentration
Genomic DNA	2% of Genomic DNA	1-500 ng
10 x Taq Buffer	2 $\mu$ l	1 x
2 mM dNTPs	2 $\mu$ l	2 mM each
5 U/ $\mu$ l Taq DNA Polymerase	0.5 $\mu$ l	0.05 U/ $\mu$ l
DMSO	0.6 $\mu$ l	3%
50 mM MgCl <sub>2</sub>	0.75 $\mu$ l	1.5 mM
10 $\mu$ M Forward Primer	1 $\mu$ l	0.5 $\mu$ M
10 $\mu$ M Reverse Primer	1 $\mu$ l	0.5 $\mu$ M
Water	Up to 25 $\mu$ l	-

**Table 4-14. PCR cycling program**

Step	Phusion DNA polymerase		Taq DNA polymerase	
	Temp. (°C)	Time (Sec)	Temp. (°C)	Time (Sec)
Initial Denaturation	98	30	94	180
25-35 Cycles				
Annealing*	98	5-10	94	30
Extension	T <sub>m</sub> + 3	10-30	T <sub>m</sub> - 5	30
	72	15-30 per kb	15-30 per kb	60 per kb
Final Extension	72	300-600	72	600
Hold	4	-	4	-

### 4.2.6. Genomic DNA Extraction from tail tips and genotyping PCR

Tail tips from mice (age > 3weeks) were cut (1 mm) and lysed in 750  $\mu$ l of genomic DNA extraction lysis buffer (including 50 mM Tris-HCl, pH 8; 100 mM EDTA; 100 mM NaCl; 1% SDS; 0.5 mg/ml Proteinase K)

## Methods

and incubated at 55°C overnight while shaking at 400 rpm. 250 µl 6 M NaCl was added, mixed by inversion, and centrifuged at full speed at RT for 20 min. 750 µl of the middle phase were transferred to a new tube containing 500 µl isopropanol, mixed by inversion, and centrifuged at full speed for 30 min. The supernatant was removed and the pellet was washed with 1.5 ml 70% ethanol, centrifuged at full speed, and ethanol was discarded. The pellet was dried at 37°C for 1 h and then incubated with 100 µl MilliQ water at 37°C, 400 rpm to dissolve. DNA extract was kept at 4°C for short-term and at -20°C for long-term storage. Genotyping PCR was performed using the primers specific for the mutant and wild type alleles of genes and Taq DNA polymerase. The annealing temperature and extension times for PCR are listed in table 4.4. Detailed information on genotyping primers is provided in material section 3.5.

**Table 4-15. Annealing temperature, extension time, and primers used for genotyping PCR**

Gene name	Mouse	Annealing temperature (°C)	Extension time (sec)	primers
BioID	Coll1a1-BioID	60.6	120	BioID FWD, BioID REV
Coll1a1	Coll1a1-BioID	55	90	Coll1 mut FWD, Coll1 wt FWD, Coll1 common REV
tta (wt/mut)	Rosa26 tTA	60.6	120	R26 tTA common FWD, R26 tTA wt REV, R26 tTA mut REV
rtta (wt/mut)	Rosa26 rtTA	60.6	120	R26 rtta common FWD, R26 rttWT Rev, R26 rtta common FWD
Ins1	Ins1-Cre	60	120	Ins1 Cre wt FWD, Ins1 Cre mt FWD, Ins1 Cre common REV
CAG	Rosa26-CAG-BioID	60	60	CAG Rosa fwd wt, CAG Rosa rev wt, CAG WPRE rev mt, CAG BioID fwd mt
CMV	CMV-Cre	60	120	CMV-Cre Forward, CMV-Cre Reverse, Internal Positive Control Forward, Internal Positive Control Reverse

### 4.2.2. Digestion of DNA using restriction endonucleases

Digestion was performed at 37 °C for 1 h using one or two appropriate FastDigest enzymes (1 Unit/µg DNA) and 1 x final concentration of the corresponding buffer. Subsequently, digested DNA was separated on an agarose gel to confirm the size of digested products or to purify the DNA fragment from the gel.

### 4.2.3. Agarose gel electrophoresis

50 x Tris-boric acid-EDTA (TBE: 2.5M Tris-HCl, 100 mM EDTA, and 1 M NaC<sub>2</sub>H<sub>3</sub>O<sub>2</sub>; pH 8) was diluted with water to 1 x final concentration. One or two per cent gels were prepared according to the size of the expected DNA fragments. For the preparation of the agarose gels, an appropriate amount of agarose powder was dissolved in 1 x TBE and allowed to boil for 2 min. 800 ng/ml (final concentration) ethidium bromide was used to visualize DNA in the gels. Gels were loaded with DNA in loading buffer (0.25 mM Bromophenol blue, 30% Glycerol, 10 mM Tris-HCl pH 7) and electrophoresis was performed in 1 x TBE-buffer applying 80-120 V in a Peqlab horizontal electrophoresis casting system. The gels were illuminated by BioDocAnalyze (BDA) digital at 366 nm. The gel imaging was carried out using Biodocanalyze 2.1 gel documentation software.

### 4.2.4. DNA extraction and purification from agarose gels

The PCR products and digested DNA fragments were excised from the agarose gel using low-intensity UV-light to visualize ethidium bromide-stained DNA and purified using buffers and spin columns from QIAquick PCR purification kit. One volume of the DNA-containing agarose gel was dissolved in three volumes binding

## Methods

buffer after a 15-min incubation time and the mixture was transferred into a fresh spin column. The column was centrifuged at 13,000 g for 1 min and the flowthrough was collected and reloaded onto the column twice. Finally, for the elution of nucleotides 35 µl of ddH<sub>2</sub>O was added into the column and after 1 min incubation at RT, dissolved DNA was collected by centrifugation at 13,000 g for 1 min.

### 4.2.5. Ligation and transformation of *E. coli*

DNA fragments, e.g. generated by PCR, and target vectors were digested with the restriction endonucleases generating compatible ends. The amount of insert was calculated using the following formula which is based on an insert/vector molar ratio of 5/1.

$$\text{Amount of insert (ng)} = \frac{\text{amount of vector (ng)} \times \text{insert (bp)} \times 5}{\text{vector (bp)}}$$

20 ng of vector was incubated for 2 h at RT with a calculated amount of insert in presence of one unit of T4 DNA-ligase. For transformation, 100 µl chemical competent cells (*E. coli* XL-1 Blue) were incubated with 10 µl of the ligation mixture or approximately 10 ng of each vector for 30 min on ice (still partially frozen). After a heat shock (45 sec at 42°C) and incubation on ice for 2 min, 900 µl of pre-warmed lysogeny broth (LB) medium (1% NaCl, 0.5 - 1% yeast extract, 1% Trypton, pH 7.2, autoclaved) was added to the tube followed by incubation at 37°C and 225 rpm for 1 h. 10-100 µL of the mixture was distributed on LB-plates containing ampicillin (100 µg/ml) for selection of positive clones. The plates were incubated overnight (16-18 h) at 37°C, colonies were picked and inoculated into 5 ml of LB-medium containing ampicillin (100 µg/ml) in glass tubes and incubated at 37°C in an orbital shaker at 200 rpm overnight. Then, the plasmid of interest was isolated from bacteria using Qiagen MiniPrep kit according to the manufacturer's protocol.

The bacterial suspension was centrifuged for 5 min and the supernatant discarded. The cell pellet was resuspended in 250 µl resuspension buffer containing RNase A and transferred to a fresh tube. Cell membranes were disrupted and the plasmid DNA was denaturated by addition of 250 µl alkaline lysis buffer containing SDS. After 5 min, the reaction was stopped and proteins were precipitated by 350 µl of the acidic neutralization buffer and inversion of the tube several times. Cells were then centrifuged for 10 min to remove precipitated compounds and to transfer the supernatant onto the QIAprep spin column. Subsequently, samples were centrifuged, washed with 750 µl of the provided washing buffers, before centrifuging again at maximum speed to remove any solution. 50 µl ddH<sub>2</sub>O were added onto the column and the plasmid was eluted by centrifugation for 1 min at maximum speed. The correct sequence of the plasmid was confirmed by sequencing (GATC Biotech AG, Konstanz, Germany) using the sequencing primers. The plasmids with verified sequence were cultured in 100 ml LB medium containing ampicillin (100 µg/ml) and isolated by midiprep was performed using the PureLink® HiPure Plasmid Midiprep Kit and according to manufacturer's protocol. The bacterial suspension was centrifuged at 4,000 g and resuspended in resuspension buffer. The columns were equilibrated with the equilibration buffer containing RNase A. Cell membranes were disrupted by the addition of lysis buffer for 5 min and the reaction was stopped using the precipitation buffer. After several inversions, a precipitate was formed and the solution was decanted into the equilibrated column, followed by two washing steps using the kit's washing buffer. The plasmid DNA was eluted using 5 ml elution buffer, followed by the precipitation of DNA using isopropanol. After 1 min incubation, the suspension was centrifuged for 30 min at 4°C and 12,000 g. The supernatant was removed and pellet dried at RT and resuspended in 100 µl ddH<sub>2</sub>O. The

## Methods

concentration and purity of the resuspended plasmid DNA was further determined using the NanoDrop spectrophotometer and the plasmid stored at -20°C.

### 4.3. Protein techniques and mass spectrometry sample preparation

#### 4.3.1. Determination of protein concentrations

Detergent-compatible (DC) assay, a modified form of Lowry assay, was performed according to the manufacturer's protocol in 96-well plates. 1:10, 20, and 40 dilutions of samples were prepared. 5 µl of several dilutions of bovine serum albumin (BSA) standards (0, 0.0156, 0.0312, 0.0625, 0.125, 0.25, 0.5, and 1 mg/ml) as well as the samples were loaded into a microtiter plate. 25 µl of working reagent A' (prepared by addition of 20 µL of reagent S per ml of reagent A) and 200 µl of reagent B were added sequentially to each well. After incubation for 15 min, the absorbance was measured at 750 nm using a Tecan microplate reader. Protein amounts in samples were determined based on the standard curve.

#### 4.3.2. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Using the following reagents, the resolving and stacking gels were cast on the Mini-PROTEAN Tetra Cell Casting Stand:

**Table 4-16. Constituents of resolving and stacking gels for SDS gel electrophoresis**

Constituent	Resolving gel		Stacking gel	
	Volume (ml)	Final conc.	Volume (ml)	Final conc.
Water	4.79	-	3.02	-
40% acrylamide	2.5	10%	0.625	5%
1.5 M Tris-HCL, pH 6.8	-	-	1.25	375 mM
1.5 M Tris-HCL, pH 8.8	2.5	375 mM	-	-
10% SDS	0.1	0.1%	0.05	0.1%
10% APS	0.1	0.1%	0.05	0.1%
TEMED	0.01	0.1%	0.005	0.1%

Sufficient amount of the loading buffer (50 mM Tris-HCl pH 6.8, 100 mM DTT, 2% SDS, 0.1% Bromophenol blue, 10% glycerol) was added to each sample (1 x) and all samples were boiled at 95°C for 10 min and loaded along with 3-4 µL of prestained protein ladder on SDS gel. Electrophoresis was performed at 120 V in BioRad Mini-Protean Tetra Cell devices filled with 1 x SDS PAGE running buffer until the loading front reached the bottom of the gel. 1 x SDS running buffer was prepared from the 10x stock buffer containing 250 mM Tris-HCl pH 8.6, 2 M glycine, and 0.1% SDS.

#### 4.3.3. Staining of polyacrylamide gels

After SDS-PAGE, gels were washed (3 x 10 min) with water and stained with PageBlue protein staining solution (Coomassie) overnight under gentle agitation. After staining, gels were washed several times with water until background staining was minimal.

#### 4.3.4. Western blot

SDS-PAGE was performed using 10% gels at 120 V for 1.5 h and the proteins were transferred to nitrocellulose membranes using a PEQlab Perfect Blue Semi-Dry Electro Blotter system at 150 mA/7 x 8 cm

## Methods

gel according to the manufacturer instructions. The layers of the sandwich consisted of (anode/-) 3 layers of Whatman papers, nitrocellulose membrane, the SDS-PAGE gel, 3 layers of Whatman paper (catode/+) all presoaked in blotting buffer (39 mM Glycine, 0.037% (w/v) SDS, 20% (v/v) Methanol, 48 mM Tris-HCl pH 6.8) The arrangement was finally covered with the blotter's cathode lid. The membranes were blocked for 1 h with either 5% nonfat dry milk in 1 x TBS (100 mM Tris-HCl pH 8, 500 mM NaCl) containing 0.05% Tween 20 (TBS-T), or 3% BSA in 1 x PBS containing 0.4% TX100 (PBS-T) at RT, washed 3 times with 1 x TBS-T/PBS-T, and incubated with the primary antibody in antibody dilution buffer (the respective blocking solution diluted 1:10) overnight at 4°C. The membranes were then washed 3 times with 1 x TBS-T/PBS-T and the secondary antibody was applied in the respective antibody dilution buffer for 1 h at RT. Protein signals were detected using Bio-rad's enhanced chemiluminescence (ECL) kit and visualized with a FUSION SOLO 4M System.

### **4.3.5. Immunofluorescence staining of cells**

NIH/3T3 transfected with the different organelle-targeting BioID constructs were permeabilized using 0.2% v/v TX100 (in 1 x PBS) for 5 min at RT, washed three times with 1 x PBS, and blocked using 3% w/v BSA (in 1 x PBS). Samples were then incubated with the respective primary antibodies overnight at 4°C. Next, slides were washed three times with 1 x PBS and incubated for 1 h in the dark with secondary antibodies. Slides were washed three times 1 x with PBS, 1 x with water, and mounted with fluoromount-G mounting solution containing DAPI for nuclear staining on specimen slides. After incubation overnight at RT, the slides were stored for later use at 4°C in the dark. Images were taken using the Axiovert 100 M microscope, equipped with an AxioCamHR camera. All pictures were taken using the Plan Apochromat 63x/1.40 Oil DIC and the filter sets 38, 43, 49, and 50 from Zeiss.

### **4.3.6. Immunofluorescence staining of tissues**

Tissue samples were fixed overnight in 4% PFA, washed the next day using 1 x PBS and stored in 20% sucrose, 1 x PBS. Samples were frozen and tissues were embedded in OCT compound from Tissue-Tek and cryosectioned (10 µm) using a cryostat from Leica. Sections were permeabilized using 0.2% v/v TX100 (in 1 x PBS) for 10 min at RT, washed three times with 1 x PBS, and blocked using 5% normal donkey serum (in 1 x PBS). Samples were then incubated with the respective primary antibodies for 2 h at RT. Next, slides were washed three times with 1 x PBS and incubated for 1.5 h in the dark with secondary antibodies. Finally, the slides were washed three times with 1 x PBS and mounted with fluoromout mounting solution on specimen slides. After incubation at RT overnight, the slides were stored for later use at 4°C in the dark. Images were taken using an Eclipse confocal microscope from Nikon. These experiments were performed by Dr. Kenichi Kimura from the institute of Physiology I, university of Bonn.

### **4.3.7. Protein precipitation**

1 ml of ice-cold chloroform-methanol (2:1, v:v) was added to 200 µl of protein sample containing maximum 2 mg protein, thoroughly vortexed for 30 sec and placed on ice for 1 h. The samples were centrifuged at 20,000 g for 30 min at 4°C, the aqueous phase was carefully removed and the protein pellet was washed with 1 ml of ice-cold methanol by vigorous vortexing followed by centrifugation at 20,000 g, 30 min, 4°C. The liquid was discarded and the pellet was dried by air and reconstituted in 1% SDS in 100 mM HEPES pH 7.4.

### **4.3.8. Cell and Tissue Lysis**

## Methods

For lysis of cell pellets, RIPA (8 times the pellet volume) containing 50 mM Tris-HCl pH 7.4, 500 mM NaCl, 2% TX100, 0.4% SDS, 5 mM EDTA, 1 mM DTT, 1 x protease inhibitor (cOmplete™ Protease Inhibitor Cocktail) was added to the cell pellets on ice. Samples were sonicated for 40 sec using Ultrasonics Sonifier with 2 mm tip up to 1 min with 60% amplitude and incubated for 1 h on a tumbling shaker at 12 rpm, 4°C. Subsequently, the lysate solutions were centrifuged at 20,000 g for 30 min at 4°C.

For tissue homogenization, tissue was cut into smaller pieces using a scalpel and either of the following was performed: (1) the ice-cold RIPA lysis buffer was added to the tissue at an 8:1 ratio (lysis buffer: tissue (v:w)). The tissue was sonicated for 1 min using Ultrasonics Sonifier with 2 mm tip up to 1 min with 60% amplitude and placed on rotator for 30 min at 4°C at 12 rpm followed by centrifugation at 20,000 g at 4°C. (2) 4% SDS in 100 mM HEPES pH 7.4 was used as lysis buffer and added to tissues in 4:1 ratio (v:w), samples were boiled at 95°C for 15 min, sonicated for 1 min with 60% amplitude using a 2 mm tip and boiled for another 15 min and centrifuged for 30 min at 20,000 g at RT.

In all procedures, supernatants were finally transferred to the new tubes for protein assay and further experiments. The homogenates were stored at – 80 °C until further processed.

### **4.3.9. Affinity enrichment of biotinylated proteins from cells (*in vitro*)**

For avidin affinity purification, 75 µl of Streptavidin Sepharose High Performance affinity bead slurry were used per 1 mg of cell lysate (slurry contains 80% bead and 20% ethanol). Beads were washed three times with lysis buffer and incubated with the sample overnight on a tumbling shaker at 12 rpm, 4°C. Between individual steps, a centrifugation step at 3,000 g for 1 min at RT was performed to settle the beads and remove the wash buffer. Streptavidin beads were washed using 1 ml of the following wash buffers for 10 min, 800 rpm, RT: 2% SDS in 50 mM HEPES pH 7.5 (twice); 0.1% sodium deoxycholate, 1% TX100, 500 mM NaCl, 1 mM EDTA, 50 mM HEPES pH 7.5 (once); 250 mM LiCl, 0.5% NP-40, 0.5% sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8.1 (once); 50 mM NaCl, 50 mM Tris-HCl pH 7.5 (twice). Subsequently, samples were reduced and alkylated on the beads (300 µl, 10 mM DTT, 100 mM HEPES, 56°C, 30 min and 25 mM acrylamide, RT, 30 min, 10 mM DTT, RT, 10 min) and digested overnight using 2.5 µg trypsin/mg input protein diluted in final 300 µl of 100 mM HEPES pH 8 at 37 °C overnight while shaking at 800 rpm. Subsequently, peptides were extracted two times from the beads with 300 µl 5% ACN, 0.1% formic acid (FA).

### **4.3.10. Affinity enrichment of biotinylated proteins from tissue samples (*in vivo*)**

For avidin affinity purification, 75 µl of Streptavidin Sepharose High Performance affinity bead slurry (slurry contains 80% bead and 20% ethanol) was used per 1 mg of tissue lysate. The beads were washed three times with lysis buffer and incubated with the sample overnight on a tumbling shaker at 12 rpm, 4°C. Between individual steps, a centrifugation step at 3,000 g for 1 min at RT was performed to settle the beads and remove the wash buffer. The methods compared were all based on on-bead digestion followed by either TX100/NP-40 wash or urea wash described in the following sections. The other differences between the protocols included the choices of lysis buffer (RIPA or 4% SDS) and inclusion or exclusion of a precipitation step before enrichment of biotinylated proteins summarized in Figure 5-41.

#### **4.3.10.1. On-bead digestion followed by SDS, TX100, and NP-40 wash**

Streptavidin beads were washed using 1 ml of the following buffers for 10 min, on a tumbling shaker at 20 rpm, RT: 2% SDS in water (twice); 0.1% sodium deoxycholate, 1% TX100, 500 mM NaCl, 1 mM EDTA, 50 mM HEPES, pH 7.5 (once); 250 mM LiCl, 0.5% NP-40, 0.5% sodium deoxycholate, 1 mM EDTA, 10 mM

## Methods

Tris-HCl, pH 8.1 (once); 50 mM NaCl, 50 mM Tris-HCl pH 7.4 (twice). Subsequently, samples were reduced and alkylated on the beads (300  $\mu$ l, 10 mM DTT, 100 mM HEPES pH 8, 56°C, 30 min and 25 mM acrylamide, RT, 30 min followed by 10 mM DTT, RT, 10 min) and digested using 2.5  $\mu$ g trypsin/mg input protein diluted in final 300  $\mu$ l of 100 mM HEPES pH 8) at 37 °C overnight while shaking at 800 rpm. Subsequently, peptides were extracted two times from the beads with 300  $\mu$ l 5% ACN, 0.1% FA. Samples were desalted by OASIS cartridges and reconstituted in 5% ACN, 100 mM HEPES pH 8.

### **4.3.10.2. On-bead digestion followed by urea wash**

Streptavidin beads were washed using 1 ml of the following buffers for 10 min, 20 rpm, RT: 0.5% SDS (once), 4 M urea in 50 mM Triethylammonium bicarbonate (TEAB) (5 times). Subsequently, samples were reduced and alkylated on the beads (300  $\mu$ l, 10 mM DTT in 300  $\mu$ l 4 M Urea in 50 mM TEAB, RT, 60 min and 25 mM acrylamide, RT, 30 min, followed by 10 mM DTT, RT, 10 min) and digested 8 h using 0.8  $\mu$ g/mg input protein lysC diluted in final 300  $\mu$ l of 4 M Urea in 50 mM TEAB at 37 °C while shaking on an orbital shaker at 20 rpm. 4 M urea was diluted 1:3 using 50 mM TEAB containing 2.5  $\mu$ g trypsin/mg input protein and samples were incubated at 37°C while rotating on an orbital shaker at 20 rpm for 12 h. Peptides were extracted two times from the beads with 300  $\mu$ l 5% ACN, 0.1% FA. Samples were OASIS desalted and reconstituted in 5% ACN, 100 mM HEPES pH 8.

### **4.3.11. OASIS desalting**

Samples were desalted using OASIS HLB 1 ml (10 mg) cartridges according to the following procedure: the cartridge was washed with 4  $\times$  1000  $\mu$ l 70% acetonitrile (ACN) in 0.5% acetic acid (AcOH) and equilibrated with 4  $\times$  1000  $\mu$ l of 0.5% AcOH. The total volume of the digest was passed through the cartridge 3 times followed by 7 times wash with 1000  $\mu$ l 0.5% AcOH. Elution was performed sequentially with 500  $\mu$ l of 30% ACN, 0.5% AcOH, 300  $\mu$ l of 50% ACN, 0.5% AcOH, and 300  $\mu$ l of 70% ACN, 0.5% AcOH (each passed through the cartridge three times). Elution fractions were combined and dried in a vacuum centrifuge at 60°C.

### **4.3.12. Quantitative peptide assay**

First, the standard dilution series was prepared using the peptide digest standard from peptide assay kit to include 1000, 500, 250, 125, 62.5, 31.3, 15.6, 7.8, and 0  $\mu$ g/ml concentrations in a peptide mixture. 10  $\mu$ l of each standard, as well as sample replicate (diluted based on expected sample amount), were pipetted into a well of a black microplate. 70  $\mu$ l of fluorometric peptide assay buffer and 20  $\mu$ l of fluorometric peptide assay reagent were subsequently added to each well. After 5 min of incubation at RT, the fluorescence was measured using Ex/Em at 390 nm/475 nm using infinite 200 pro Tecan microplate reader. The standard curve was used to determine the peptide concentration for each sample.

### **4.3.13. RapiGest Digestion**

To 100  $\mu$ g of a protein sample, 10  $\mu$ l of 1% RapiGest (w/v) in 0.1 M HEPES pH 7.6 was added and sample was incubated at 37°C for 45 min. Another 10  $\mu$ l of 0.1 M HEPES pH 7.6 was added to reduce the rapigest concentration to 0.5%. Samples were reduced (5 mM DTT, 56 °C, 30 min), alkylated (20 mM Acrylamide, RT, 50 min) and the reaction quenched with DTT (5 mM DTT, RT, 10 min). 1  $\mu$ g trypsin was added and the volume was adjusted to 100  $\mu$ l with 0.1 M HEPES (final protein concentration: 1  $\mu$ g/ $\mu$ l). Samples were incubated at 37°C for overnight digestion. At this step peptide quantification and TMT labeling of the samples was performed, samples mixed and RapiGest was precipitated. For precipitation of RapiGest samples were acidified with trifluoroacetic acid (TFA) to a final concentration of 1%. Samples were placed at 37°C for 45

## Methods

min while shaking at 600 rpm and centrifuged at 20,000 g for 10 min at RT. The supernatant was carefully transferred to a new tube.

### 4.3.14. Labeling of peptides using dimethyl labeling reagents

The ACN concentration was reduced to less than 1% by successive addition of HPLC-water and vacuum centrifugation. In-solution isotope dimethyl labeling was performed according to the protocol developed by Boersema et al. (Boersema et al. 2009). In short, to 100 µg (in 100 µl of ~1% ACN, 100 mM HEPES pH 7.6) of peptides to be labeled as light, medium, and heavy, 16 µl of 4% (v/v) 2H-formaldehyde (CH<sub>2</sub>O), 2D-formaldehyde (CD<sub>2</sub>O), and 13C,2D-Formaldehyde (13CD<sub>2</sub>O) were added, respectively. The peptide and corresponding label reagent were shortly mixed and centrifuged. 16 µl of 0.6 M sodium cyanoborohydride (NaBH<sub>3</sub>CN) were added to light and intermediate and 16 µl of 0.6 M deuterated sodium cyanoborohydride (NaBD<sub>3</sub>CN) were added to heavy samples. After 1 h of incubation at RT while mixing at 700 rpm, the labeling was quenched by 64 µl of 1% (v/v) ammonia solution (NH<sub>4</sub>OH) and samples were acidified using 32 µl FA. Labeled samples were pooled and desalted by OASIS cartridges.

### 4.3.15. Labeling of peptides using TMT 6plex and 10plex

For labeling, peptides were resuspended in 100 mM HEPES pH 8 and peptide assay was performed. The volume of peptide solution containing 100 µg of peptides was adjusted to 106 µl using 5% ACN, 100 mM HEPES pH 8. TMT label reagents were equilibrated to RT, dissolved in 36 µl 100% ACN for 5 min with occasional vortexing. Vials were then briefly centrifuged and 106 µl of sample (~1 µg/µl) were added to each tube. The labeling reaction was performed at RT for 90 min and quenched by adding 8 µl 5% hydroxylamine in 0.1 M HEPES pH 8.

6 or 9 differentially labeled samples were combined and reduced to ~200 µl in a vacuum centrifuge at 60 °C. Finally, the concentration of ACN was minimized to ensure binding of peptides to the reversed-phase material in OASIS columns. Therefore, 200 µl HPLC-water was added and samples were thoroughly mixed and again concentrated to 200 µl in the vacuum centrifuge. The last steps were repeated several times until the final concentration of ACN in the sample dropped below 1%.

### 4.1.16. OFFGEL fractionation

Peptides were fractionated on immobilized pH gradient (IPG) gel based on their isoelectric points into 12 fractions using a 3100 OFFGEL fractionator. First, the IPG gel strip (pH 3 - 10, 13 cm) was assembled into the tray and fixed with a 12 well plastic frame. Rehydration of the strip was performed using 20 µl per well of the IPG buffer (pH 3-10) for at least 30 min while not exceeding 2 h. The dried samples were resolubilized with 800 µl of IPG buffer for 15 min, the volume was increased to 1.8 ml and an equal amount of the peptide mixture was pipetted into each well of the tray (150 µl/well). A high voltage was applied to the ends of the gel strips, accomplishing the separation of peptides into 12 fractions overnight. Detailed instrumental settings are provided in Table 4.6. The next day, each fraction was transferred into a new microtube. 150 µl 0.1% TFA was added into every well of the tray, incubated for 15 min and extracts were finally combined with the corresponding fractions and dried in a vacuum centrifuge at 60°C.

**Table 4-17. OFFGEL program for 12 fractions**

	Power (KVh)	Voltage (V)	Current (µA)	Power (mW)	Time (h)
Focusing	20	4500	50	200	100
Hold	-	500	20	50	-

#### 4.3.17. High-pH reversed-phase peptide fractionation (Hp-RP)

Peptides were separated via basic reversed-phase offline fractionation using an Ultimate 3000 HPLC system equipped with a binary pump, column oven, autosampler and UV-detector. The separation was achieved using an XBridge Shield RP<sub>18</sub> main column (1.0 x 100 mm, 3.5  $\mu$ m, Waters) in combination with a Gemini C<sub>18</sub> SecurityGuard column (4.0 x 2.0 mm, Phenomenex). 20 mM ammonium formate pH 10 was used as buffer A and pure acetonitrile as buffer B. For the separation, desalted and dried peptides were first dissolved in buffer A, followed by injection into the offline HPLC system. Peptides were separated at a flow rate of 100  $\mu$ l/min using a non-linear gradient of 5% to 40% B in 91 min (Table 4.7). Peptide elution was monitored at a wavelength of 220 nm. Fractions were collected into a 96 well plate at a 2 min interval (200  $\mu$ l per fraction) using a Gilson FC203B fraction collector for the entirety of the separation. Afterwards, fractions were pooled via concatenation (Table 4.8 and Table 4.9) to either 12 or 24 fractions and finally dried in a vacuum centrifuge at 60 °C. This experiment was performed by Dr. Robert Hardt.

**Table 4-18. Hp-RP non-linear gradient.**

Retention time (min)	%B
0	5
15	5
18	9
22	10
26	12
31	13
35	14
40	16
45	17
50	18
56	20
61	21
67	23
74	25
80	27
88	29
100	33
106	40
107	85
112	85
113	5
120	5

**Table 4-19. Concatenation schema for 12 fractions of Hp-RP**

Combined Fraction	1	2	...	11	12
Original fractions from	A11 A12	- B1	- ...	- B10	- B11

## Methods

96 well plate	B12	C1	...	C10	C11
	C12	D1	...	D10	D11
	D12	E1	...	E10	E11
	-	-	-	-	E12

**Table 4-20. Concatenation schema for 24 fractions of Hp-RP**

Combined Fraction	1	2	...	23	24
Original fractions from 96 well plate	A12	-	-	-	-
	B1	B2	...	C11	C12
	D1	D2	...	E11	E12
	-	-	-	-	F1
	-	-	-	-	F2
	-	-	-	-	F3

### 4.1.18. Peptide cleanup with StageTips

Six small discs of C<sub>18</sub> material were pressed into 200 µl pipette tips and the tips were placed into miniprep columns with pierced holes (used as holders) that were put in 2 ml microtubes. C<sub>18</sub> material was activated by 20 µl methanol and with 20 µl of buffer B (80% ACN, 0.5% AcOH) and then washed with 20 µl of buffer A (0.5% AcOH in HPLC-water). For OFFGEL fractionated samples, peptides in 300 µl 0.1% TFA were directly applied on stage tips and centrifuged 2 min at 2,000 g. Other peptide samples were dissolved in 20 µl of 5% ACN, 5% FA, sonicated for 5 min and volume adjusted to 100 µl by HPLC-water and centrifuged 2 min at 2,000 g. Stage tips were washed with 100 µl of buffer A. Peptides were eluted by addition of 20 µl buffer B (2x) into fresh tubes. The eluate was dried in the vacuum centrifuge at 60°C and resuspended in 20 µl of buffer 5% ACN, 5% FA (Rappsilber, Ishihama, and Mann 2003). Samples were sonicated in an ultrasonic water bath for 5 min and centrifuged for 15 min at 20,000 g.

## 4.4. Mass Spectrometry

### 4.4.1. Nano-UHPLC-MS/MS-Orbitrap-Velos (Orbitrap-Velos)

Peptides were reconstituted in 5% ACN, 5% FA. 5 µl of peptide solution were containing 5 µg peptide was injected to the analytical column (60 µm outer diameter (OD), 100 µm inner diameter (ID) fused silica capillary, produced in house using a Sutter P2000 laser puller device and packed with 5 µm particles [ReproSil C<sub>18</sub>-AQ, Dr. Maisch GmbH, Ammerbuch-Entringen, Germany]) using an EASY-nLC 1000 UHPLC system at a flow rate of 1 µl/min. Peptides were separated with a 60 min linear gradient of 100% solvent A (5% DMSO, 0.1% FA) to 65% A/35% B (94.9% ACN, 5% DMSO, 0.1% FA) at a flow rate of 400 nl/min. Peptides eluting from the column were ionized by electrospray ionization using 1.5 kV at positive mode and infused into an LTQ Orbitrap-Velos mass spectrometer. Survey scans were performed at a resolution of 60,000 (m/z 400-1200) in the Orbitrap part of the instrument followed by fragmentation of the 10 most intense precursor ions in the ion trap (normalized energy 35%). The repeat count was set to 1 and the dynamic exclusion to 60 s.

### 4.4.2. Nano-UHPLC-MS/MS-Orbitrap-Lumos (Orbitrap-Lumos)

Peptides were reconstituted in 5% ACN, 5% FA and analyzed using a nano-UHPLC-Dionex Ultimate 3000 system connected to an Orbitrap Fusion Lumos mass spectrometer. 1 to 4  $\mu\text{g}$  of the sample were loaded directly on a 50 cm reversed-phase analytical column at a flow rate of 600 nl/min using 100% solvent A (0.1% FA in HPLC-water). The columns were produced in-house as follows: spray tips were generated from 360  $\mu\text{m}$  OD, 100  $\mu\text{m}$  ID fused silica capillaries using a P-2000 laser puller and packed with 1.9  $\mu\text{m}$  Reprosil AQ C<sub>18</sub> particles (Dr. Maisch, Ammerbuch-Entringen, Germany). Peptide separation was performed with 60 min linear gradients from 5% - 35% solvent B (90% ACN, 0.1% FA) at a flow rate of 300 nl/min. To obtain accurate TMT-quantifications mass spectra were acquired using a synchronous precursor selection (SPS)-based MS3 method (McAlister et al. 2014). For this, MS1 spectra were acquired in the orbitrap mass analyzer from m/z 375-1500 at a resolution of 60 K. The automatic gain control (AGC) was set to standard and the maximum ion injection time to auto. Precursor ions (charge states 2+ to 7+) above an intensity of  $5 \times 10^3$  were isolated in the quadrupole (1.0 m/z isolation width) and fragmented using higher collision dissociation (HCD) fragmentation with normalized collision energy (NCE) of 35 % and dynamic exclusion set to 30 seconds. MS2 scans were acquired in the iontrap analyzer using the rapid scan mode (AGC: standard, max. inject time: Auto). From the fragment spectra, 8 ions in the mass range of 400 to 2000 m/z were co-isolated by SPS (2.0 m/z isolation width), further fragmented by HCD (65% NCE) and measured in the Orbitrap mass analyzer (resolution: 30K, mass range: 100 – 1000 m/z, AGC: 200%, max. inject time: Auto) to obtain MS3 TMT reporter ion spectra.

### 4.5. Data analysis, bioinformatics and statistical analysis

#### 4.5.1. Data Analysis using MaxQuant

MS raw files were analyzed using MaxQuant 1.6.5.0 (Cox and Mann, 2008) against Uniprot/Trembl human (74416 entries, released on 06/2019). Enzyme specificity was set to trypsin, up to two missed cleavages were allowed, and initial mass tolerance of  $\pm 20$  ppm and a final mass tolerance of  $\pm 9$  ppm for precursor masses and  $\pm 0.5$  Da for the fragment ions defined. Propionamide (cysteine) was used as fixed modification; oxidation (methionine) and acetylation (protein N-termini) were set as variable modifications. For protein ratio calculation of SILAC labeled samples heavy/medium arginine (Arg 10/6) and heavy/medium lysin (Lys 8/4) were defined as labels. To enable protein ratio calculation for dimethyl labeled samples, heavy/medium/light lysin (dimethyl Lysin 8/4/0) and heavy/medium/light n-terminal (dimethyl-n term 8/4/0) were defined as labels. The false discovery rate (FDR) was set to 1% at the protein level.

#### 4.5.2. Data Analysis using Proteome Discoverer

For the analysis of TMT labeled samples measured by the SPS-MS3 acquisition method, Proteome Discoverer™ software (Thermo Scientific, version 2.3) in combination with an in-house Mascot server (Matrix Sciences, version 2.6.1) was used. Raw files for the fractions of each TMT experiment were pooled in one search and analyzed with processing and consensus workflows. In the processing workflow the spectrum selector node was used with the default settings except for the scan event filter which was set to let pass only MS2-spectra. For Mascot the following search parameters were applied: Protease: Trypsin/P; Precursor mass tolerance: 10 ppm; fragment mass tolerance: 0.5 Da; maximum missed cleavage sites: 2; fixed modifications: propionamide (C); dynamic modifications: oxidation (M), acetylation (Protein N-term) additionally for the labeling experiment: TMT10/6plex modifications. Searches were conducted against Uniprot/Trembl mouse

## Methods

(74141 entries, released on 6/2019) and a common contaminant database (cRAP, <https://www.thegpm.org/crap/>). The false discovery rate calculation on PSM level was performed using the Percolator node with following decoy database search parameters: Target FDR (Strict): 0.01; Target FDR (Relaxed): 0.05; Validation based on: q-value. To gain more identifications, spectra not passing the strict FDR of 0.01 were re-searched by Mascot using slightly modified parameters (Protease: semiTrypsin; max missed cleavages: 1; dynamic modifications: oxidation (M), acetylation (N-term protein), propionamide (C), TMT (K, N-terminus)) and also validated by Percolator. Based on the TMT 10/6-plex derived reporter ion intensities at the MS3 level, relative quantification was performed using the “Reporter Ions Quantifier” node. Afterwards, the spectrum identifications and quantifications from the processing workflow were further processed in the consensus workflow. The peptide spectrum matches (PSM) were first grouped by the PSM Grouper (Site Probability Threshold: 75) and then passed on to the “Peptide Validator” and “Peptide and Protein Filter” node, where the identifications were scored and filtered using default settings (PSM & peptide FDR: 0.01). Afterwards, peptides were grouped via the “Peptide Isoform Grouper” node, protein scores were calculated by the “Protein Scorer” node and protein groups were constructed by the “Protein Grouping” node using the principle of strict parsimony. The “Peptide in Protein Annotation” and “Modification Sites” were also applied using default settings. Protein FDRs were calculated by the “Protein FDR Validator” node. The “Protein Marker” node was also utilized (default parameters). Quantifications were processed in the “Reporter Ions Quantifier” node with the following parameters. Peptides to Use: Unique; Consider Protein Groups: True; Used Shared Quan Results: True; Reporter Abundance Based on: Automatic; Co-Isolation Threshold: 60%; Average Reporter S/N: 10; SPS Mass Matches: 65%; Normalization: None; Scaling: None; Protein Ratio Calculation: Protein Abundance Based; Max. Allowed Fold Change: 100; Imputation: None.

### **4.5.3. Perseus analysis for *in vitro* experiments (SILAC labeled samples)**

H/L (and M/L) ratios provided by MaxQuant were inverted. In 3plex studies for determination of a statistical cut-off the H/M (control/control) normalized ratios were used. The list was filtered for proteins with all 3 quantified H/M normalized ratios and the cut-off ratio at 95<sup>th</sup> percentile of the distribution (95% cut-off) was determined. Then L/M normalized ratios were loaded, contamination (CON), reverse (REV), and the hits only identified by site (SITE) were removed and ratios were  $\log_2$  transformed. Proteins were accepted only if all 3 ratios were quantified. After a one-sample t-test, the list was filtered for significant proteins with  $p$ -values lower than 0.05 and the 95% cut-off calculated was applied. Proteins with  $p$ -values < 0.05 (significant) and  $\log_2$  L/M  $\geq$  95% cut-off (enriched) were defined as biotinylated.

### **4.5.4. Perseus analysis for *in vivo* experiments (Dimethyl and TMT labeled samples)**

#### **Perseus Analysis for 2plex and 3plex dimethyl labeled samples (Liver, Lung, and Pancreas)**

##### **2plex: H: biotinylated, L: control**

H and L intensities were loaded into Perseus software and contamination (CON), reverse (REV), and the hits only identified by site (SITE) were filtered out and intensities  $\log_2$  transformed. The missing values in the control channel (L) were replaced with small random values by the application of the imputation action. Proteins with less than 3 valid intensities in the H channel (and 2 for the pancreas) were removed and  $\log_2$  H/L ratios were generated. A one-sample t-test was performed and the  $\log_2$  cut-off of 1 was applied. Proteins with  $p$ -values < 0.05 (significant) and  $\log_2$  H/L  $\geq$  1 (enriched) were defined as biotinylated.

##### **3plex: H: biotinylated, L and M: control**

## Methods

H and L and M intensities were loaded into Perseus software and contamination (CON), reverse (REV), and the hits only identified by site (SITE) were filtered out and intensities  $\log_2$  transformed. To normalize the data and correct for pipetting errors the distance of the mean of M/L for each replicate was found and subtracted from  $\log_2$  M/L and  $\log_2$  H/L ratios of that replicate. 95% cut-off was identified for each  $\log_2$  M/L ratio separately and the mean was calculated to serve as the cut-off for biotinylation. A one-sample t-test was performed and the  $\log_2$  cut-off was applied. Proteins with  $p$ -values  $< 0.05$  (significant) and  $\log_2$  H/L  $\geq$  95% cut-off (enriched) were defined as biotinylated.

### **Perseus Analysis for 6plex TMT samples (Control (C) 1, 2, 3: 126, 127, 128; Biotinylated (B) 1, 2, 3: 129, 130, 131)**

To identify the beta-cell proteome, the grouped protein abundances (average of all 12 fractions) of all TMT channels were loaded into Perseus and  $\log_2$  transformed. Proteins with less than 3 values in biotin channels (129, 130, and 131) were filtered out and missing values in light channels (126, 127, and 128) were replaced with small values using the imputation function. Control/control and biotin/control ratios were generated (C2/C1, C3/C2, C3/C1 and B2/C1, B3/B2, B3/B1). The mean of the C2/C1 ratio was calculated and subtracted from C2/C1 and B2/C1 ratios as a means of normalization. The same was applied to other ratios. 95% cut-off was calculated for each C/C ratio; the average was used as the  $\log_2$  95% ratio cut-off for biotinylation. A one-sample t-test was performed and the 95%  $\log_2$  cut-off was applied. Proteins with  $p$ -values  $< 0.05$  (significant) and  $\log_2$  biotin/control  $\geq$  95% cut-off (enriched) were defined as biotinylated.

To identify the regulated proteins in whole proteome dataset the grouped protein abundances (average of all 24 fractions) of all TMT channels were loaded into Perseus and  $\log_2$  transformed.  $\log_2$  biotin/control ratios were generated (B3/C3, B2/C2, B1/C1) and proteins with less than three  $\log_2$  biotin/control ratios were removed. A one-sample t-test was performed and cut-offs of 1 and -1 were applied to find the significant up- and down-regulated proteins, respectively.

### **Perseus Analysis for 9plex TMT samples (Control (C) 1, 2, 3: 127C, 127N, 128C; Biotinylated (B) 1, 2, 3: 128N, 129C, 129N; Biotinylated + Harmine/STZ (H/S) 1, 2, 3: 130C, 130N, 131)**

The grouped protein abundances (average of all 12 fractions) of all TMT channels were loaded into Perseus and  $\log_2$  transformed. To identify the biotinylated proteins (beta-cell proteome), C and B intensities were used. Proteins with less than 3 values in biotin (B) channels (128N, 129C, 129N) were filtered out and missing values in control (C) channels (127C, 127N, 128C) were replaced with small values using imputation function.  $\log_2$  Control/control and  $\log_2$  biotin/control ratios were built (C2/C1, C3/C2, C3/C1 and B2/C1, B3/B2, B3/B1). Mean of the  $\log_2$  C2/C1 was calculated and subtracted from  $\log_2$  C2/C1 and  $\log_2$  B2/C1 ratios as a means of normalization. The same was applied to other ratios. 95% cut-off was calculated for each  $\log_2$  C/C ratio; the average was used as the  $\log_2$  95% ratio cut-off for biotinylation. Next, the biotin (B) and Harmine or STZ (H or S) intensities were combined to create following  $\log_2$  ratios: B2/B1, B3/B2, B3/B1 and “H2 or S2/B1”, “H3 or S3/B2”, “H3 or S3/B1” for biotinylated proteins. The mean of  $\log_2$  B2/B1 ratios was identified and subtracted from  $\log_2$  B2/B1 and  $\log_2$  “H2 or S2/B1” to normalize the data. The same was performed for the other two  $\log_2$  B/B ratios. 95% cut-off was determined for each  $\log_2$  B/B and averaged to find the drug-effect cut-off. One sample t-test was performed and proteins with  $p$ -values  $< 0.05$  and cut-off  $\leq$  “S or H/B” ratios  $\leq$  -cut-off were defined as perturbed by the drug.

## 4.6. Mice techniques

Animals were kept with food and water *ad libitum*. All experiments were performed in accordance with the German 'Tierschutzgesetz' under animal application numbers Az 84-02.04.2014.A069 and Az 84-02.04.2017.A346.

### 4.6.1. Treatment of mice

#### 4.6.1.1. Doxycyclin and Biotin treatment

Doxycycline (Dox) and biotin were administered to rtTA-Coll1a1-BioID mice in the drinking water supplemented with 5% sucrose for 7 days at a concentration of 2 mg/ml and 0.22 mg/ml, respectively. The Dox containing drinking water was kept in black bottles to prevent light-induced degradation and was replaced every 2–3 days. In case of treatment only with biotin (for Rosa-CAG-BioID-CMV Cre and Ins1Cre-Coll1a1-BioID-tTA mice), biotin was dissolved in the water at a concentration of 0.22 mg/ml and given to mice for a duration of 7 days.

#### 4.6.1.2. Streptozotocin (STZ) treatment

Streptozotocin was administered to Ins1Cre-Coll1a1-BioID-tTA mice in 40 mg/kg/day doses by intraperitoneal (IP) injection for 5 days. 4 h prior to STZ treatment, all food was removed from the cage. Water was provided as normal. Immediately prior to injection, STZ was dissolved in 50 mM sodium citrate buffer (pH 4.5) to a final concentration of 4 mg/ml and 10 µl of the stock/g body weight was injected within 5 min of being dissolved intraperitoneally.

#### 4.6.1.3. Harmine treatment

Harmine was administered to Ins1Cre-coll1a1-BioID-tTA mice through IP injection at a concentration of 10 mg/kg harmine HCl/injection for 7 successive days. 1 mg/ml harmine hydrochloride was prepared by dissolving Harmine in normal saline solution (NaCl 0.9%) with sonication and heating for 30 min at 40 °C and 10 µl Harmine solution was injected per g of body weight.

### 4.6.2. Blood glucose measurements

Glucose concentrations were measured twice at all time points before and after 4 h of daily fasting using a drop of blood collected from the tail tip on Accu-check.

### 4.6.3. Dissection of tissues

For isolation of all organs except pancreas from Rosa-CAG-BioID-CMV and Coll1a1-BioID-rtTA Cre mice, animals were anaesthetized by intraperitoneal injection of 100 mg/kg Ketamin und 10 mg/kg Rompum and perfused by injection of 1 x PBS into the left ventricle after cutting the right atrium. With a longitudinal skin and muscle incision, the abdomen was opened from the diaphragm to the throat. The diaphragm was cut free from ribs and thorax was opened on the left and right sides of the sternum. A syringe was inserted from the tip of the left ventricle and the right atrium of the beating heart was punctured by scissors. PBS was slowly injected through the circulatory system to fully perfuse the heart and replace the blood with PBS. Brain, lungs, liver, spleen, kidneys, brain, sciatic nerve, muscle, heart, and eye were isolated and cut to pieces and transferred to tubes and immediately shock frozen on dry ice. For isolation of pancreas from Rosa-CAG-BioID-CMV and Coll1a1-BioID-rtTA Cre mice, animals were perfused by direct injection of 10 ml of PBS into the left ventricle after cutting the right atrium. For isolation of pancreas from Ins1Cre coll1a1 BioID tTA mice, mice were not

## Methods

perfused. Pancreas was isolated from all these mice (in less than 3 min after starting the perfusion and less than 30 seconds after cutting the diaphragm) and frozen on dry ice. From certain tissue samples, a fraction was transferred to 4% PFA for fixation (for microscopy) and the rest of the tissues were stored at  $-80^{\circ}\text{C}$  until further processing (for Western blotting and mass spectrometry).

## 5. Results

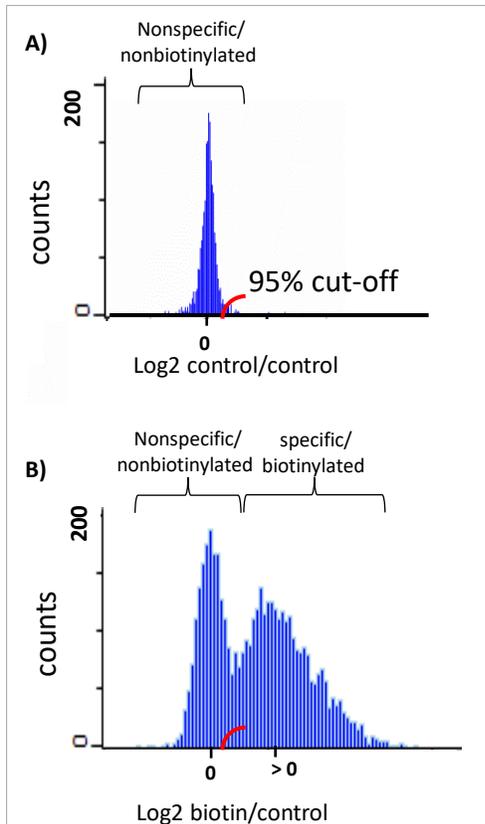
### 5.1. Establishment of a statistical approach for cut-off determination in BioID-based enrichment methods and its implementation in organelle proteomics

#### 5.1.1. Experimental determination of a cut-off for the acceptance of biotinylated peptides

It is common practice in affinity enrichment experiments, to distinguish proteins binding specifically to the beads from such non-specifically enriched by stable isotope labeling approaches. In such experiments, the fold-change cut-off to accept a protein as specifically enriched is often defined in an arbitrary way (e.g.  $> 1$ -fold or  $> 2$ -fold (Alvarez-Castelao et al. 2017; Krogager et al. 2018)) or based on previous knowledge by definition of the cut-off in a way to include a known population of proteins in the significantly enriched population (slide approach; (Branon et al. 2018)). Assignment of a fold change arbitrary cut-off may result either in underestimation of cut-offs and therefore the rejection of truly biotinylated proteins, or overestimation and in turn acceptance of non-biotinylated proteins as true positives. Slide method is the most reliable cut-off determination approach in the field. However, it is dependent on the prior knowledge on the localization of the proteins present in the list and could not be used for enrichment studies with no previous experimental data available.

In a biotinylation experiment normally biotinylated and control samples are differentially labeled and mixed. The proteins binding to the beads in the biotinylated sample are expected to be a mixture of truly biotinylated (with higher intensities) and non-specific binding proteins (with lower intensities). In such an enrichment experiment with two populations of biotinylated and control samples, differentially labeled with SILAC amino acids, the non-specific proteins in both biotinylated and control samples are expected to form a distribution of SILAC ratios around  $\log_2$  ratio 0 (as proteins in both groups have similar intensities) and a second population should acquire higher ratios due to enrichment and higher intensities in biotin channel (expected to cluster around  $\log_2$  ratios  $> 0$ ) (Figure 5-1 B). We hypothesized, given that background proteins form a normal distribution, it should be possible to determine a cut-off based on experimental data by analyzing the distribution of non-specific binding proteins in an individual null experiment (Figure 5-1 A) and apply it on biotinylation experiments (Figure 5-1 B). In biological experiments (and other fields of natural sciences) the investigation of the null distribution is a common approach to distinguish signal from noise and interpret the treated proteome in a correct way (Bickel 2011; Ping et al. 2013). In general, biological data are assumed to be normally distributed and form a Gaussian distribution. In this case, the statistical inferences from the data are assumed to be with a high probability correct and reliable (Bickel 2011). One of the most commonly used statistical definitions for interpretation of data is the interval containing 95% of the reference population (Fay and Gerow 2013). We proposed such a one-tailed interval (as biotinylation is a unidirectional effect) in a null experiment should contain only the non-biotinylated proteins with 95% probability allowing us to distinguish between biotinylated and non-biotinylated proteins.

## Results



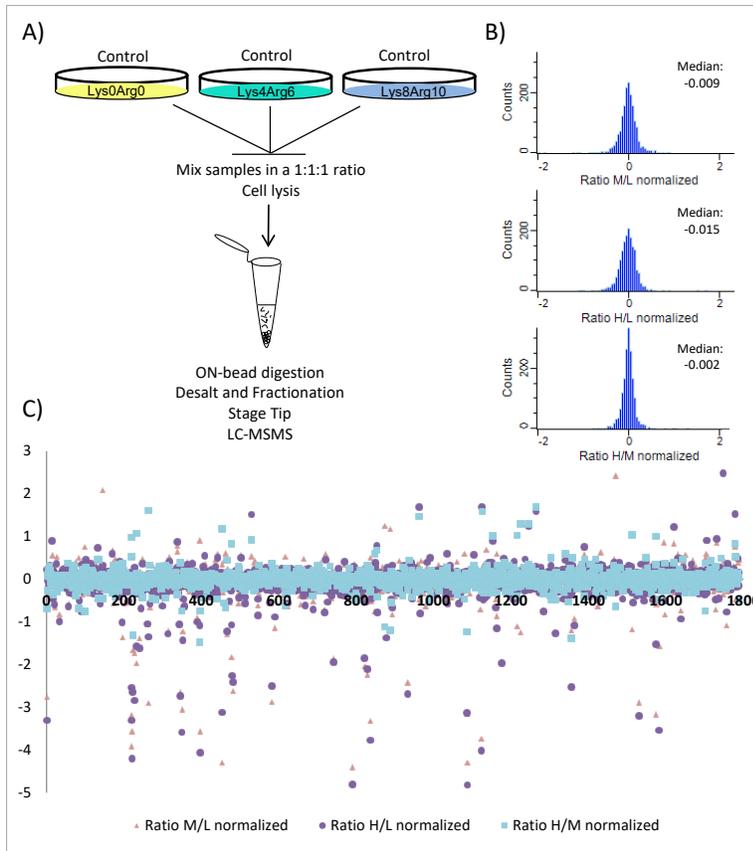
### Figure 5-1. 95% Cut-off determination strategy.

These two histograms (A and B) are drawn based on hypothetical assumptions to explain our strategy and not based on experimental data. We hypothesized that it should be possible to find a 95% cut-off based on the distribution of non-biotinylated proteins as part of a null experiment (A) and apply it on a biotinylation experiment (B) to distinguish biotinylated proteins from those binding non-specifically to streptavidin beads. The logic behind this hypothesis is that non-specific binding proteins are present in both biotinylated and non-biotinylated samples and are expected to bind the streptavidin beads with the same strength in both populations resulting in a cluster of ratios around log<sub>2</sub> biotin/control of 0 (non-specific cluster in B). It should be in theory possible to find a 95% cut-off (which includes 95% of these non-specific proteins) in a null experiment (A) and apply it on the second cluster with biotin/control > 0 (in B). To find this cut-off we planned to perform a separate null experiment (A) including only proteins which are not biotinylated and postulated a cut-off including 95% of the proteins in this list should be applicable to any biotinylation experiments (e.g. B) to differentiate between biotinylated and non-specific proteins.

In order to determine a cut-off for biotinylation, we performed a null experiment employing non-biotinylated light, medium, and heavy SILAC labeled HEK 293T cells. We combined 1 mg whole-cell lysates from each SILAC labeled sample and incubated the samples with streptavidin beads, followed by washing of the beads, reduction, alkylation and digestion of the proteins. Peptides were fractionated by OFFGEL, desalted by OASIS cartridges, and analyzed using LC-MSMS (Orbitrap-Velos- 60 min gradient) in 5 independent biological replicates (Figure 5-2 A). The normalized SILAC ratios calculated by MaxQuant (Annex Table 5-1) were analyzed in Perseus (1.6.5.0). For the proteins with intensities in all three channels, the normalized SILAC ratios were log<sub>2</sub> transformed and the distribution of the ratios plotted (Figure 5-2 B and C).

As expected, since all samples contained proteins with similar affinity to the beads (low affinity for background proteins and higher affinity for endogenously biotinylated carboxylases), the intensities of a majority of proteins were similar in all three channels and therefore SILAC ratios for the majority of the proteins were normally distributed around 1 (log<sub>2</sub> = 0) (Figure 5-2 B and C). The naturally biotinylated carboxylases (pyruvate carboxylase, PC; methylcrotonoyl-CoA carboxylase subunit alpha, MCC1; acetyl-CoA carboxylase 1, ACACA; acetyl-CoA carboxylase 1, ACACB; propionyl-CoA carboxylase alpha chain, PCCA; (Roux et al. 2018)) were found in high intensities in all three SILAC channels. Ratios for the majority of proteins were close to the median values of the dataset (-0.009, -0.015, and -0.002 for log<sub>2</sub> of M/L, H/L, and H/M normalized ratios, respectively (Figure 5-2 B).

## Results

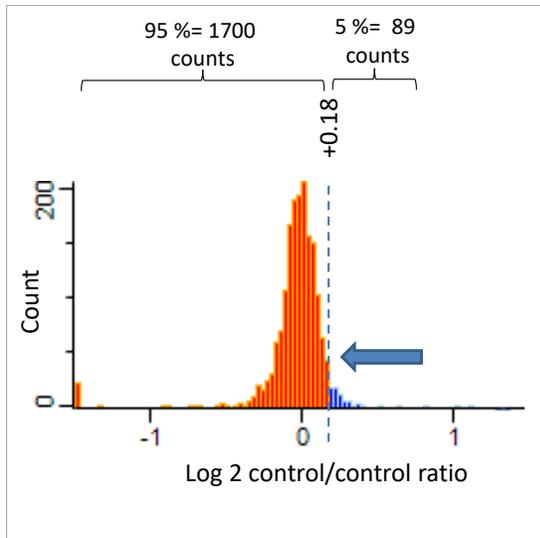


**Figure 5-2. Schematic workflow and ratio distribution for the null experiment.**

A) For the null (background) experiment HEK 293T cells were labeled with light, medium, and heavy SILAC amino acids, lysed and mixed 1:1:1 and on-bead digested. Samples were OASIS desalted, fractionated using OFFGEL and stage-tipped before LC-MSMS analysis. Raw files were analyzed with MaxQuant/Perseus. Log<sub>2</sub> normalized SILAC ratios quantified by MaxQuant were plotted. B) Histogram for log<sub>2</sub> normalized SILAC ratios shows distribution of ratios around 0 that is the log<sub>2</sub> value for 1. The medians for log<sub>2</sub> normalized M/L, H/L, and H/M ratios were -0.009, -0.015, and -0.002, respectively. C) Scatter plot of the log<sub>2</sub> normalized ratios in the null experiment. Experiments' ID numbers: SA3122, SA3133, and SA504.

Accordingly, we then determined the confidence limit of the log<sub>2</sub> fold-change for the inclusion of 95% of the population identified in this study and defined this value as significance cut-off for specific binding. For this, we used the mean ratio of the final M/L, H/L, and H/M (all control/control) normalized ratios provided by MaxQuant. To define the 95% percentile of the ratios, we divided the number of proteins with quantified values by 20 to find the number of proteins that count for 5% of the list. In this study, the protein list is comprised of 1,789 proteins and 5% contains 89 proteins. Considering that biotinylation only results in an increase of the intensities and is a unidirectional effect, we arranged the ratios from smallest to the largest values and moved from the largest towards the smallest ratio 89 values, until the 5% of proteins was reached; this value was defined as the 95% ratio cut-off for biotinylation. This cut-off was the log<sub>2</sub> ratio +0.18 and 95% of the SILAC ratios fall below this value (Figure 5-3; Annex Table 5-2).

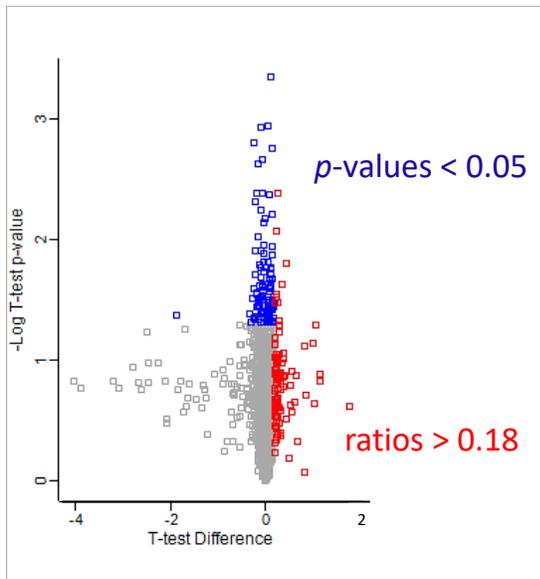
## Results



**Figure 5-3. Determination of the 95% cut-off for biotinylation.**

As part of a background (null experiment) HEK 293T cells were labeled with light, medium, and heavy SILAC amino acids, lysed and mixed 1:1:1 and on-bead digested. Samples were OASIS desalted, fractionated using OFFGEL and stage-tipped before LC-MSMS analysis. Raw files were analyzed with MaxQuant/Perseus. To define the 95% cut-off for biotinylation, the mean of the normalized  $\log_2$  transformed SILAC ratios calculated by MaxQuant for proteins in the list was determined. The mean values were arranged from smallest to the biggest values. We then moved from biggest number towards the smallest until 5% of the proteins were counted and the ratio at 5% threshold was defined as 95% cut-off value, as 95% of the proteins have ratios smaller than this value. This value was +0.18. Since biotinylation results in increment in intensities this cut-off is a unidirectional positive value. Experiments' ID numbers: SA3122, SA3133, and SA504.

We then performed a one-sample t-test and filtered for proteins with  $p$ -values  $< 0.05$ . Only 6% of the proteins identified were significantly changed between the SILAC channels (105 out of total 1789) and the normalized ratios of 91% of them (96/105) were below the cut-off ratio determined for enrichment (Figure 5-4). As proteins binding specifically in affinity enrichment experiments can by definition only increase in their ratio, proteins below the cut-off can be neglected reducing the number of potential false positives to 9/1789 (0.05%).



**Figure 5-4. Potential false positive hits in the null experiment.**

As part of a background (null experiment) HEK 293T cells were labeled with light, medium, and heavy SILAC amino acids, lysed and mixed 1:1:1 and on-bead digested. Samples were OASIS desalted, fractionated using OFFGEL and stage-tipped before LC-MSMS analysis. Raw files were analyzed with MaxQuant/Perseus.  $\log_2$  normalized SILAC ratios quantified by MaxQuant were used to find a cut-off which includes 95% of the proteins in this study ( $\log_2 = +0.18$ ). A one sample t-test was performed and scatter plot created. Out of the 1789 proteins binding non-specifically to the beads from 5 null experiments, only 105 hits have  $p$ -values lower than 0.05 (in blue). Out of these proteins only 9 proteins have normalized SILAC ratios higher than the 95% cut-off (the red proteins beside the blue proteins). This means the number of significant proteins with ratios higher than cut-off (potential false positives) is 9 out of 1789 (0.05%). Experiments' ID numbers: SA3122, SA3133, and SA504.

To conclude, the background proteins formed a normal distribution which enabled us to find a 95% cut-off for this population ( $\log_2 = +0.18$ ). As background proteins are expected to show a similar distribution pattern in biotinylation studies, we hypothesize that this 95% cut-off can be applied on biotinylation studies to distinguish non-specific binding proteins from biotinylated proteins; any protein with biotin/control ratios above this cut-off and with  $p$ -values  $< 0.05$  will be accepted as biotinylated.

## **5.1.2. Analysis of organelle-specific proteomes including one control sample**

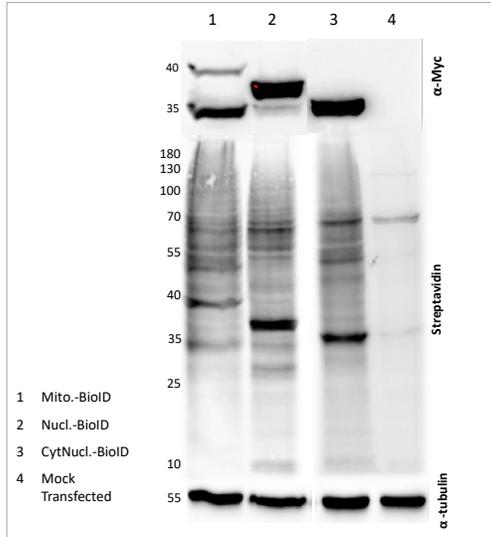
### **5.1.2.1. Generation of BioID organelle constructs and confirmation of organelle targeting**

Next, we planned to apply our 95% cut-off to the analysis of three different subcellular proteomes. To perform the organelle proteomics analysis, we first created organelle targeting BioID constructs and evaluated their correct localization. We aimed to direct myc-tagged BioID to the nucleus (Nucl.), mitochondria (Mito.), and the cytoplasm and nucleus (CytoNucl.) by the generation of organelle-specific BioID fusion constructs. The BioID gene was amplified by PCR (template: Addgene Plasmid #35700) and inserted into a pFIV3.2RSV-IRES-Puromycin vector (C. S. Stein and Davidson 2002). For organelle-specific constructs, localization signal sequences were either included in the BioID primer (nuclear signal sequence) or inserted as part of a second cloning step (mitochondrial localization) into the pFIV3.2CMV-IRES-Puromycin vector (C. S. Stein and Davidson 2002). The localization sequences were as follows: for nucleus PPKKRRKVDPPKKRQKVDPPKKRQK at the protein C-terminus (Kalderon et al. 1984) and for mitochondrial matrix SU9 at the protein C-terminus (John et al. 2002). BioID without signal sequences in pFIV3.2CMV-IRES-Puromycin vector was used as the cytoplasmic-nuclear (CytoNucl.) construct. pcDNA3-EGFP (Addgene Plasmid #13031) was used for mock transfection.

The expression and biotinylation by BioID constructs were evaluated by Western blot and microscopy analysis using HEK 293T and NIH/3T3 cells, respectively. Cells were transiently transfected by BioID constructs and after 24 h biotin (50  $\mu$ M, final concentration) was added. After another 24 h cells were fixed or harvested and anti-myc and streptavidin were used to assess the expression of BioID and biotinylation, respectively.

In Western blot experiment, the bands corresponding to myc-tagged BioID constructs had the anticipated overall size of BioID plus the respective signal sequence fused to it (Figure 5-5, myc blot). SU9 is reported to cleave off from the protein upon import into mitochondria (John et al. 2002); this is the reason for observing two bands for mitochondrial BioID (Figure 5-5; myc blot, lane 1). Proteins biotinylated by different BioID constructs (Figure 5-5; streptavidin blot) show dissimilar patterns and much stronger signal than the mock-transfected control sample with very few weak bands. Differences in the pattern of biotinylated proteins in streptavidin blots suggest the presence of different proteins and the few weak bands in the mock-transfected lane possibly belong to naturally biotinylated carboxylases (Figure 5-5; streptavidin blot). The results obtained from the Western blot experiment were in line with our expectations.

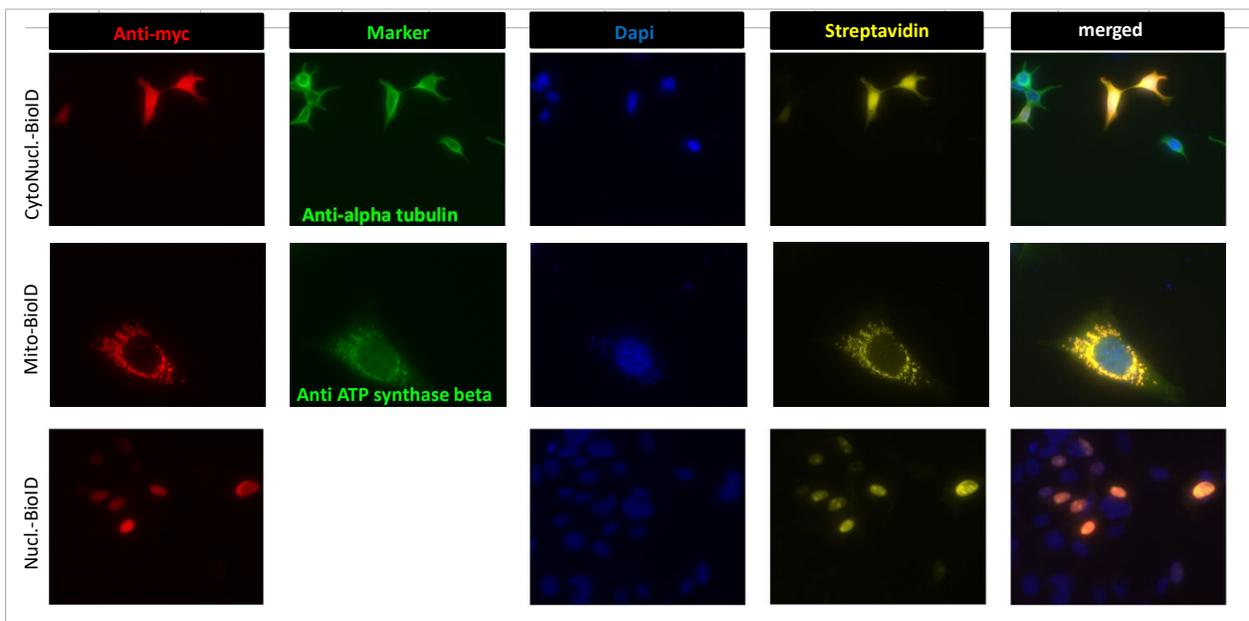
## Results



**Figure 5-5. Confirmation of BioID expression and biotinylation by Western blot.**

HEK 293T cells were transiently transfected with myc-tagged BioID organelle constructs, biotinylated (50  $\mu$ M - 24 h) and lysed. 75  $\mu$ g from each sample was loaded on SDS gel and analyzed by western blot. Anti-myc (Abcam, ab9106, 1: 3,000; blocked with 5% milk in TBS-T) and HRP-conjugated streptavidin (21126, ThermoFischer scientific: 1: 10,000; blocked with 3.5% BSA in PBS-T) were used to detect the expression of BioID and biotinylation by BioID, respectively. Anti-beta-3 Tubulin Monoclonal Antibody (MA119187, Dianova, 1: 5,000; blocked with 5% milk in TBS-T) was used as loading control. The difference in size of the BioID constructs is due to the signal sequence fused to it.

Next, we confirmed the correct subcellular localization and biotinylation by immunofluorescence microscopy (Figure 5-6) using anti-myc (BioID expression), streptavidin (biotinylation), DAPI (nuclear marker), and the other markers for the organelle of interest (anti-alpha tubulin for cytoplasm and anti-ATP synthase beta for mitochondria). We set two controls for each co-localization experiment, using the markers of other organelles to further approve the specificity of each organelle construct (Appendix Figures 1 to 3). For all organelle constructs, BioID was expressed and biotinylated the proteins within the isolated organelle of interest (Figure 5-6).



## Results

### **Figure 5-6. Confirmation of correct localization of BioID organelle constructs by microscopy.**

NIH/3T3 cells were transfected with myc-tagged CytoNucl.-BioID., Mito.-BioID, and Nucl.-BioID constructs, treated with biotin (50  $\mu$ M - 24 h), fixed and stained with anti-myc antibody (ab9106, Abcam; 1: 600- showing BioID expression), anti-alpha tubulin antibody [DM1A] (Alexa Fluor® 488) (ab195887, Abcam, 1: 400- Cytosolic marker), anti-ATP synthase beta monoclonal antibody [3D5AB1] (A21351, ThermoFischer scientific, 1: 500- Mitochondrial marker); Alexa Fluor 647 streptavidin (S21374, ThermoFischer scientific, 1: 400) was used to detect biotinylated proteins. Secondary antibodies were: Goat IgG anti-rabbit IgG (H+L)-Cy3 (111-165-144, Dianova, 1: 400), and MFP488 goat anti-mouse IgG (H+L) (MoBiTec, MFP-A1029, 1: 400). Coverslips were mounted with fluoromount-G mounting solution containing DAPI for nuclear staining, used also as nuclear marker (ThermoFischer scientific, Bremen, Germany). A similar increase of brightness was applied to all figures showing DAPI staining to help for visibility. The merged figures show the expression and biotinylation by BioID in the organelle of interest based on its co-localization with the respective organelle marker. Experiment's ID number: SA4178.

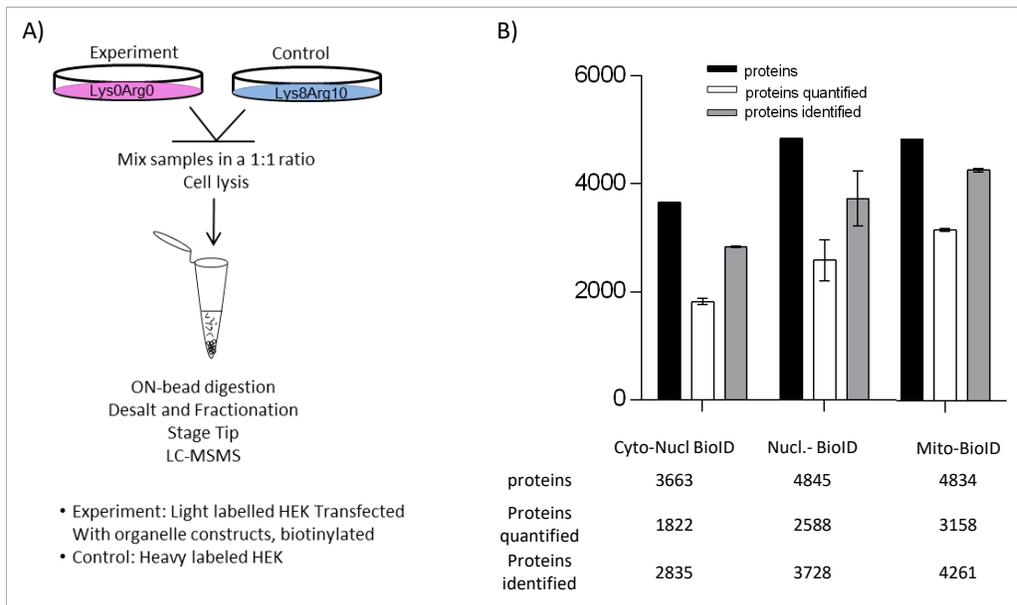
### **5.1.2.2. Organelle proteomics**

We then performed for each organelle sample on-bead digestion and LC-MSMS analysis in three biological replicates (Figure 5-7 A). For these experiments, HEK 293T light cells were transiently transfected with the different BioID constructs. 48 h post-transfection, cells were supplemented with 50  $\mu$ M biotin (final concentration), incubated for 24 h, and harvested. Heavy labeled untransfected HEK 293T cells were used as control. Equal amounts of experiment and control samples (1 mg from each) were combined and incubated with streptavidin beads, followed by washing of the beads, reduction, alkylation and digestion of the proteins. Peptides were fractionated by OFFGEL, desalted by OASIS cartridges, and analyzed using LC-MSMS (Orbitrap-Velos- 60 min gradient) (Figure 5-7 A).

MaxQuant searches were performed on pooled raw files for the three replicates of each individual experiment (Annex Table 5-3). After inversion of the ratios provided by MaxQuant (since the control is heavy labeled L/H ratios or biotin/control are needed), we applied the 95% cut-off ratio determined in the null experiment to differentiate between the background and biotinylated proteins. Accordingly, we considered proteins with  $\log_2$  normalized L/H ratios  $\geq 0.18$  as biotinylated.

We compared the organelle-specific proteomes with regard to total number of proteins, proteins quantified (proteins for which ratios could be determined), and identified in biotin channel (proteins with intensities  $> 0$  in light channel). The reason for evaluating the number of proteins identified in biotin channel (and not in control channel) is that this reflects the number of proteins which can possibly obtain biotin/control ratios. Since non-specific proteins bind the streptavidin beads weakly, it is likely that their intensities are below detection threshold of MS and they are not identified. It is however possible to change the intensities for proteins in control channel from 0 (or NaN : not a number) to a small value to enable generation of biotin/control ratio. We therefore were more interested in number of proteins identified in biotin channel. For all organelle-specific datasets, we achieved roughly similar numbers for total proteins, identified in biotin channel as well as quantified proteins, with the exception of the CytoNuclear samples which resulted in ~70% of the other organelle-specific proteomes (Figure 5-7 B). In all methods, the total number of proteins identified in biotinylated sample (with intensities in the light channel) was higher than proteins quantified (with intensities in both channels creating the L/H ratio). This means that the number of proteins with intensities  $> 0$  in light (biotin) channel was higher than those in heavy (control) channel. This can be either due to their absence in the control channel and no binding/very loose binding to beads and therefore acquiring intensities lower than the mass spectrometric detection threshold. This however could be anticipated as the intensities of background (non-specific) proteins are expected to be lower than those binding specifically to the beads and enriched.

## Results

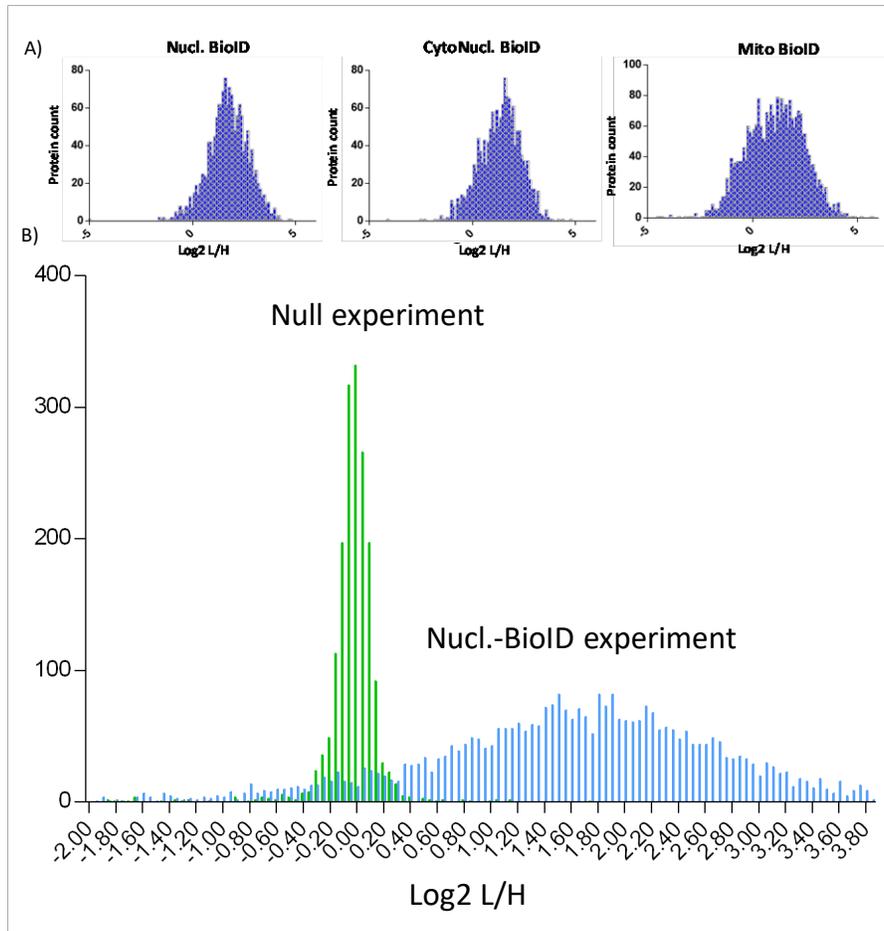


**Figure 5-7. Identification and quantification efficiencies by BioID organelle constructs (including one control sample)**

A) For organelle proteomics experiments including one control sample, HEK 293T light labeled cells were transiently transfected with BioID organelle constructs and biotinylated (50  $\mu$ M - 24 h), lysed, and mixed 1:1 with HEK 293T heavy labeled control cell lysate followed by on-bead digestion. Samples were OASIS desalted, fractionated using OFFGEL and stage-tipped before LC-MSMS analysis. Raw files were analyzed with MaxQuant/Perseus. B) Bar graphs show the number of total proteins, proteins with quantified L/H ratios, and those identified based on the intensities in light (biotin) channel in three replicates of each organelle proteomics experiment. Experiment's ID number: SA4160.

Dependent on the subcellular localization of BioID we observed different distributions of  $\log_2$  L/H ratios (Figure 5-8 A) which differed markedly from the background (null) study (Figure 5-2 B and 5-3) and was against our expectations (Figure 5-1 B). A large population of proteins included in the 95% interval of the background population shifted to a cluster with  $\log_2$  L/H ratios around 2 to 3, irrespective of the subcellular localization of BioID. In general, this trend was more pronounced in CytoNucl.-BioID and Nucl.-BioID datasets. Mito.-BioID showed a population of proteins with ratios around 0 although not clearly separated from the rest of the proteins (Figure 5-8 A). The overlay of the  $\log_2$  ratio distribution for proteins in the null experiment (in green) and those in nuclear datasets (in blue) in Figure 5-8 B shows the altered behavior of ratios in the nuclear dataset and shift for the majority of proteins towards values greater than 0. This can be due to lower intensities of non-specific (non-nuclear) proteins in the control channel (H) of nuclear dataset which results in large biotin/control values. To see if this is the case, we looked at the ratio distribution of nuclear and non-nuclear proteins in the null experiment and Nucl.-BioID datasets.

## Results



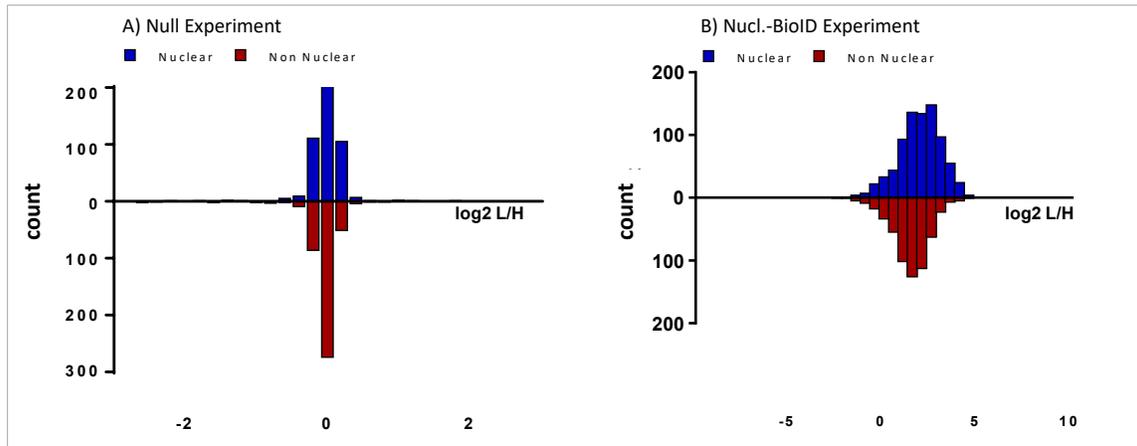
**Figure 5-8. Distribution of  $\log_2$  L/H ratios in organelle proteomics datasets and overlay of  $\log_2$  L/H ratio distributions in background (null) and nuclear BioID datasets.**

For organelle proteomics experiments including one control sample (A and blue histogram in B), HEK 293T light labeled cells were transiently transfected with BioID organelle constructs biotinylated (50  $\mu$ M - 24 h), lysed, and mixed 1:1 with HEK 293T heavy labeled control cell lysate followed by on-bead digestion. For the null experiment (green histogram in B), HEK 293T cells were labeled with light, medium, and heavy SILAC amino acids, lysed and mixed 1:1:1 and on-bead digested. All samples were OASIS desalted, fractionated using OFFGEL and stage-tipped before LC-MSMS analysis. Raw files were analyzed with MaxQuant/Perseus. A)  $\log_2$  normalized L/H (biotin/control) ratios were plotted as histogram for each organelle dataset. B) The mean of SILAC  $\log_2$  ratios in background were merged with L/H  $\log_2$  ratios in nuclear BioID dataset in an overlay histogram. Green: distribution of  $\log_2$  normalized L/H ratios quantified in the null experiment. Blue: distribution of  $\log_2$  normalized L/H ratios quantified in Nucl.-BioID dataset. Experiment's ID number: SA4160.

We listed the ratios determined in the background experiment and the dataset with nuclear BioID expression (Nucl.-BioID) for proteins found in both analyses. We performed GO analysis on the list of common proteins and created two lists with nuclear and non-nuclear annotations and compared the ratios for these proteins between the datasets. The  $\log_2$  L/H ratios for non-nuclear (in blue) and nuclear (in red) proteins were plotted in two separate histograms corresponding to null experiment (Figure 5-9 A) and Nucl.-BioID experiment (Figure 5-9 B). While in the null experiment the  $\log_2$  ratios for both groups are clustered around 0 (Figure 5-9 A), in the nuclear dataset with respect to bins with the highest count of proteins (peak of the histogram), the nuclear proteins show only a slight shift in distribution (mean L/H: 1.9) in comparison to non-nuclear proteins (mean L/H: 1.4) (Figure 5-9 B). We concluded that  $\log_2$  biotin/control ratios for both nuclear and non-nuclear proteins increase in Nucl.-BioID dataset. This is expected from the nuclear proteins as they are biotinylated and

## Results

enriched in the sample which results in higher intensities in biotin channel. However, non-biotinylated proteins are expected to behave similarly in both null and Nucl.-BioID experiments and form a normal distribution around  $\log_2 0$ . As these ratios are generated from the abundance of proteins in light (biotin) and heavy (control) channels, to elucidate what occurs to background proteins, we next focused on SILAC intensities instead of final ratios for the non-nuclear proteins in both null and Nucl.-BioID datasets.



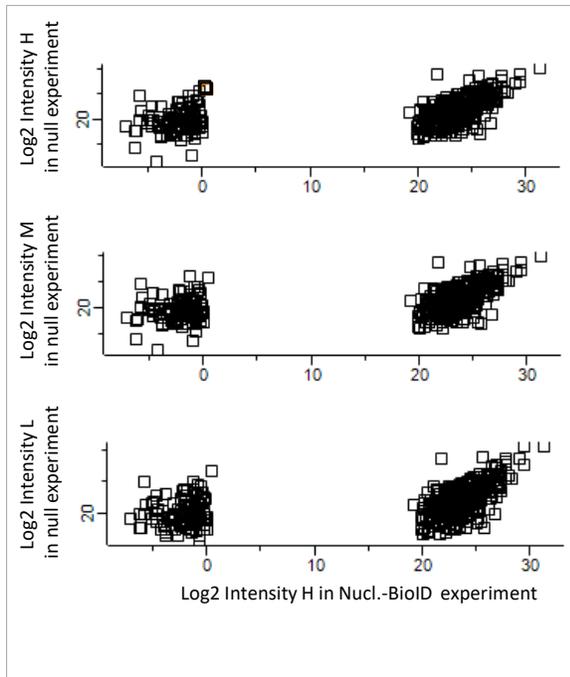
**Figure 5-9. Ratio distribution for nuclear and non-nuclear proteins in Nucl.-BioID and null experiment datasets.**

For the null experiment (A), HEK 293T cells were labeled with light, medium, and heavy SILAC amino acids, lysed and mixed 1:1:1 and on-bead digested. For Nucl.-BioID experiment (B) HEK 293T light labeled cells were transiently transfected with Nucl.-BioID construct, biotinylated (50  $\mu$ M - 24 h), lysed, and mixed 1:1 with HEK 293T heavy labeled control cell lysate followed by on-bead digestion. All samples (in A and B) were OASIS desalted, fractionated using OFFGEL and stage-tipped before LC-MSMS analysis. Raw files were analyzed with MaxQuant/Perseus.

Common proteins between Nucl.-BioID and null datasets were found and distributed into two lists of nuclear (blue) and non-nuclear (red) proteins, based on GO analysis. A) Histogram for  $\log_2$  normalized L/H SILAC ratios for nuclear and non-nuclear proteins quantified in background (null) experiment. B) Histogram for  $\log_2$  normalized L/H SILAC ratios for nuclear and non-nuclear proteins quantified in Nucl.-BioID experiment. Experiments' ID numbers: SA3122, SA3133, SA504, and SA4160.

We wondered if background (non-nuclear) proteins shared between the null experiment and nuclear datasets have different intensities which result in such a shift in their L/H ratios. We plotted the intensities of these proteins in the null experiment (in each of the heavy, medium and light channels) versus their intensities in control (heavy) channel of the Nucl.-BioID experiment (Figure 5-10). We detected a population with similar intensities in both studies (right side) and a second population with much lower intensities in Nucl.-BioID dataset (left side).

## Results



**Figure 5-10. Comparison between intensities of background proteins in Nucl.-BioID vs null experiment datasets.**

For the null experiment HEK 293T cells were labeled with light, medium, and heavy SILAC amino acids, lysed and mixed 1:1:1 and on-bead digested. For Nucl.-BioID experiment HEK 293T light labeled cells were transiently transfected with Nucl.-BioID construct, biotinylated (50  $\mu$ M - 24 h), lysed, and mixed 1:1 with HEK 293T heavy labeled control cell lysate followed by on-bead digestion. All samples were OASIS desalted, fractionated using OFFGEL and stage-tipped before LC-MSMS analysis. Raw files were analyzed with MaxQuant/Perseus.

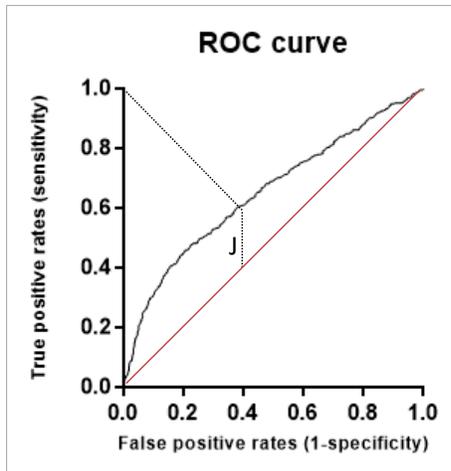
For the shared non-nuclear proteins (based on GO analysis), intensities of control samples in Nucl.-BioID experiment (heavy channel-X axis) and null experiment (light, medium and heavy channels-Y axis) were  $\log_2$  transformed and plotted which formed two clusters of proteins 1) with similar intensities in both experiment (right side) and 2) with much lower intensities in Nucl.-BioID dataset (left side). Experiments' ID numbers: SA3122, SA3133, SA504, and SA4160.

The distribution of non-nuclear proteins in Nucl.-BioID dataset suggests that part of these proteins acquire lower intensities when incubated on beads with biotinylated proteins in the same tube (as part of the on-bead digestion protocol) which probably results in bigger L/H values for the majority of the proteins. This could be due to biotinylated proteins competing with background proteins for binding the beads.

At this point, we performed the “slide approach” (Branon et al. 2018) to see where a cut-off based on GO analysis would fall and how different it would be compared to our null experiment cut-off. In this approach, the cut-off is adapted stepwise until the majority of proteins known to be part of the population of interest are above the cut-off (Branon et al. 2018). In each dataset, two populations of true positives and false positives are defined based on GO analysis and their biotin/control SILAC ratios are plotted in the receiver operating characteristic (ROC) curve which displays the full picture of trade-off between the sensitivity (true positive rate/TPR) and (1- specificity) (false positive rate/FPR) across a series of cut-off points. To find the optimal threshold point a criterion called “Youden index” can be used which gives equal weight to sensitivity and specificity and is defined as the furthest point on the line of equality on ROC curve. It maximizes the difference between TPR and FPR and yields in best correct classification of the data. It is calculated as “sensitivity + specificity – 1” (Lijana Zaletel-Kragelj 2010).

Using the L/H  $\log_2$  ratios of proteins in the two lists (of nuclear and non-nuclear proteins) we created a roc (receiver operating characteristic) curve which depicts the trade-off between TPR and FPR (Figure 5-11) and calculated the Youden index for each suggested cut-off. The maximum Youden index proposed, suggesting the optimum cut-off, was the  $\log_2$  ratio +2.013 that is very different from our cut-off from the null experiment, +0.18.

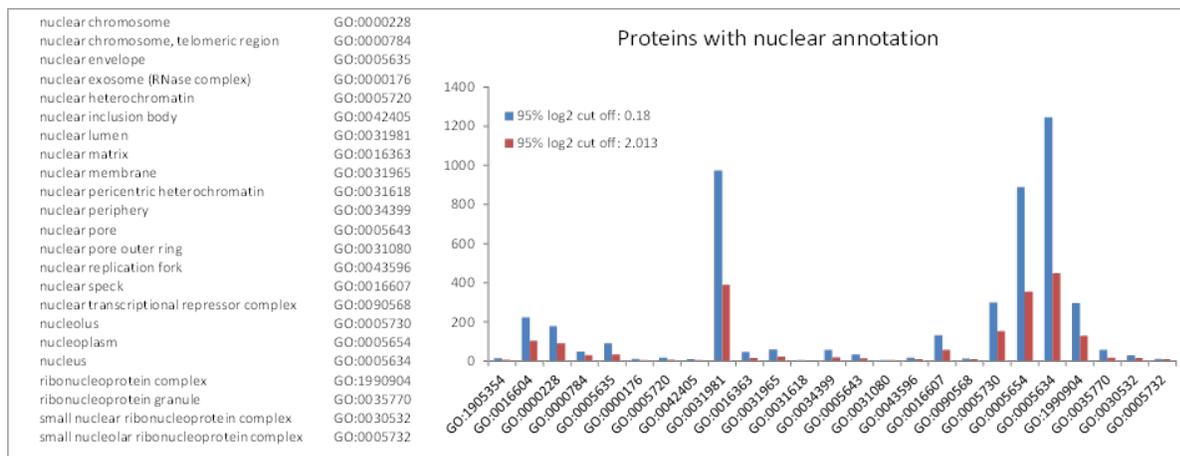
## Results



**Figure 5-11. Youden index and ROC curve analysis on Nucl.-BioID dataset.**

For Nucl.-BioID experiment HEK 293T light labeled cells were transiently transfected with Nucl.-BioID construct, biotinylated (50  $\mu$ M - 24 h), lysed, and mixed 1: 1 with HEK 293T heavy labeled control cell lysate followed by on-bead digestion. Samples were OASIS desalted, fractionated using OFFGEL and stage-tipped before LC-MSMS analysis. Raw files were analyzed with MaxQuant/Perseus. To find a cut-off for biotinylation, the “slide approach” was used. Two lists of true positives (nuclear) and false positive (non-nuclear) proteins were created based on GO analysis and their L/H ratios used to create the ROC curve. Youden index was calculated for all the points of a ROC curve and the maximum youden index suggested +2.013 as the optimum cut-off for biotinylation which keeps the best balance between sensitivity (true positive rate) and 1-specificity (false positive rate).

We applied 0.18 and 2.013 as cut-offs for biotinylation to  $\log_2$  L/H values in Nucl.-BioID dataset and 2,162 and 617 proteins passed the threshold, respectively. GO enrichment (Panther GO-slim cellular component) analysis showed for both lists, nuclear part (GO: 0044428) and nucleus (GO: 0005634) as the top categories. For the proteins above the 0.18 cut-off, the FDR  $p$ -values for the nuclear part and nucleus categories were  $2.86E^{-49}$  and  $2.21E^{-43}$  (with 424 proteins in nuclear part and 222 in nucleus categories). For the slide cut-off of 2.013, the FDR  $p$ -values were  $2.12E^{-48}$  and  $2.80E^{-45}$  for the two categories and the number of proteins in nuclear part and in nucleus categories were 118 and 183, respectively. GO cellular component analysis was also performed and the categories with nuclear annotation were filtered and compared between the two lists (Figure 5-12).



**Figure 5-12. Comparison between the number of nuclear proteins passing null experiment cut-off and the “slide approach” cut-off.**

For Nucl.-BioID experiment HEK 293T light labeled cells were transiently transfected with Nucl.-BioID construct biotinylated (50  $\mu$ M - 24 h), lysed, and mixed 1:1 with HEK 293T heavy labeled control cell lysate followed by on-bead digestion. Samples were OASIS desalted, fractionated using OFFGEL and stage-tipped before LC-MSMS analysis. Raw files were analyzed with MaxQuant/Perseus.

Two approaches were taken to determine a cut-off for biotinylation for Nucl.-BioID dataset: 1) Null experiment approach (cut-off: +0.18): 95% cut-off was found for the mean of  $\log_2$  normalized L/M, H/M and H/L values with a normal ratio distribution. For this experiment untransfected cells labeled with light, medium and heavy SILAC amino acids were lysed, combined, on-bead digested and analyzed by LC-MSMS. 2) Slide approach (cut-off: +2.013): this was determined using ROC curve and Youden index analysis based on ratio distribution of two lists with nuclear and non-nuclear proteins. The proteins passing each cut-off were subjected to GO enrichment cellular component analysis and the numbers of proteins in each nucleus-linked GO term (left side) were plotted. These are the proteins which passed the null experiment cut-off (in red) and slide approach cut-off (in blue).

## Results

The results indicate that the smaller cut-off finds more nuclear proteins as well as a higher number of non-specific hits. In general, based on GO analysis both cut-offs result in the identification of a similar percentage of proteins with nuclear annotation in comparison to all proteins passing the cut-off.

To conclude, against our expectations of a bimodal ratio distribution with two clusters for non-specific and biotinylated proteins (explained in Figure 5-1), all datasets showed a unimodal distribution of L/H SILAC ratios clustered around 2-3 while a small population had ratios around 0. The latter population was smaller in nuclear and cytonuclear datasets. This was surprising to us, as we expected a subpopulation of proteins behaving similar to the background study which could allow us to use our null experiment cut-off on an enrichment study. This promoted the hypothesis that when biotinylated and control proteins are mixed and incubated on beads, even though the protein-bead ratio is kept constant, biotinylated proteins compete for binding to the beads and replace the background population increasing L/H ratios. Based on the distribution of nuclear and non-nuclear proteins (Figure 5-9) (1) it is clear that it is not possible to simply apply the cut-off from the null experiment to another dataset and (2) the slide approach cut-off appears to be more reasonable and realistic. Determination of this cut-off, however, is dependent on prior knowledge about localization of the proteins in the list.

### **5.1.3. Effect of biotinylated proteins in the binding pattern of non-specific proteins**

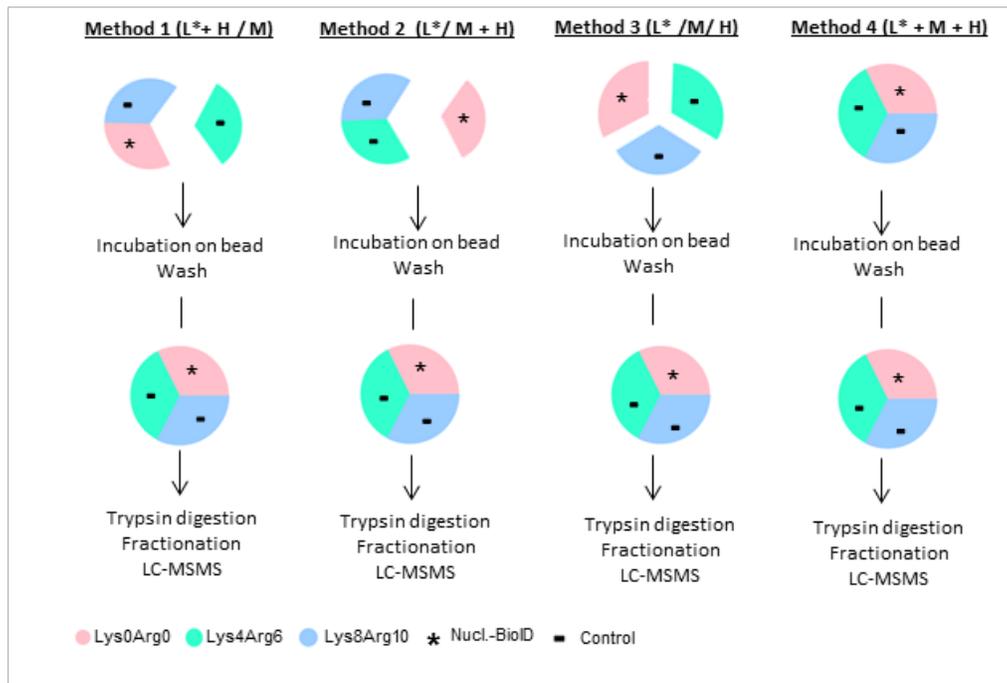
We further attempted to find an unbiased approach to set a cut-off for biotinylation for which no previous knowledge on localization of proteins is required. We suggested it should be possible to define a cut-off within the study instead of performing separate background and enrichment analyses by including a second control sample. This also helps to evaluate the influence of biotinylated proteins on binding behaviour of non-specific background (and vice versa). To these aims, we included two control samples (instead of one) beside one biotinylated sample and mixed the samples at different steps of on-bead digestion protocol. The distribution of control/control ratios (created from the two control samples in each study) should allow determining a 95% cut-off which may have the credibility to be applied on biotin/control ratios (in the same study) to distinguish biotinylated proteins from those binding the beads non-specifically.

We transiently transfected light labeled HEK 293T cells with nuclear BioID (Nucl.-BioID) construct which showed the largest cut-off shift among the datasets in organelle proteomics experiments and used medium and heavy labeled cells as controls. Cells were biotinylated 48 h after transfection, for 24 h (50  $\mu$ M biotin) and harvested. Heavy and medium SILAC labeled HEK 293T cells served as controls. Using 1 mg starting material from each condition, we combined the biotinylated and the 2 control samples 1:1:1 in different steps of the on-bead digestion protocol and adjusted the bead amount accordingly. Biotinylated proteins in light and heavy control samples were mixed before incubation with beads in method 1 and further mixed with medium control at digestion level (L\*+ H/M). In method 2 (L\* / M + H) medium and heavy control samples were mixed before incubation with beads and biotinylated samples were added on the top, right before trypsin digestion. In method 3 (L\* / M / H) and 4 (L\* + M + H) samples were mixed before digestion and before incubation on bead, respectively (Figure 5-13). Peptides were fractionated by OFFGEL, desalted by OASIS cartridges, and analyzed using LC-MSMS (Orbitrap-Velos- 60 min gradient). MaxQuant searches were performed on pooled raw files for the three replicates of each individual experiment (Annex Table 5-5).

We aimed to evaluate the intensities of biotinylated and control samples and to assess potential effects of biotinylated proteins on such binding non-specifically and vice versa when they are incubated together on the beads, or separately. In method 1, by comparing the intensities of heavy (incubated with biotinylated proteins) and medium (incubated separately) control samples, we can evaluate the effect of biotinylated proteins on

## Results

binding pattern of control samples. The medium control sample in Method 4 can be compared to that in method 1, to show the effect of biotinylated proteins on binding pattern of non-biotinylated proteins when a second control sample is added to the beads. Method 2 shows how control samples potentially influence each other when incubated together on the beads. Method 3 can show the intensities proteins acquire when treated in separate tubes, and also reflect the technical variation across the control samples when processed individually.



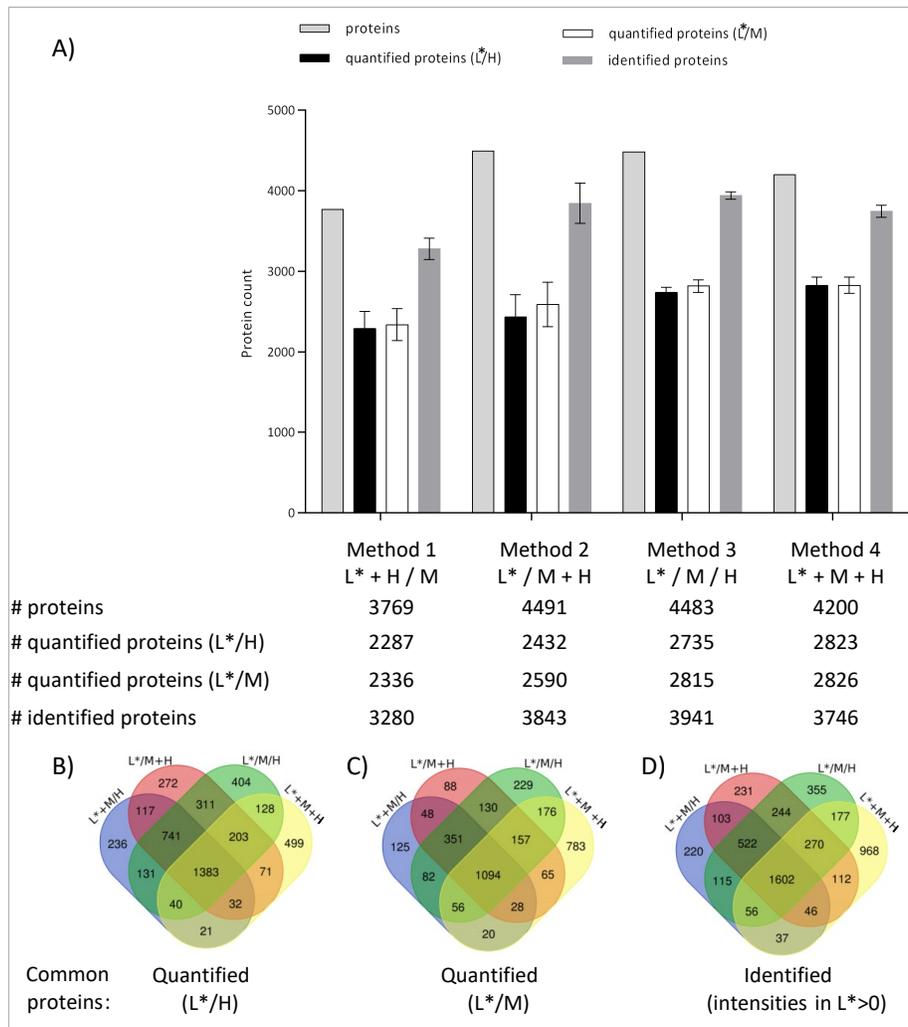
**Figure 5-13. Schematics of four different strategies for combination of biotinylated and control proteins.**

Light labeled HEK 293T cells were transiently transfected with Nucl.-BioID construct and biotinylated (50  $\mu$ M - 24 h); medium and heavy labeled cells served as controls. Biotinylated and the 2 control samples were combined 1:1:1 in different steps of the on-bead digestion experiment. Method 1 (L\* + H / M): Biotinylated proteins in light were mixed before incubation on bead with heavy control sample and further mixed with medium control before trypsin digestion. In method 2 (L\* / M + H) medium and heavy control samples were mixed before incubation with beads and mixed with biotinylated samples before trypsin digestion. In method 3 (L\* / M / H) and 4 (L\* + M + H) samples were combined before trypsin digestion and before incubation with beads, respectively. Samples were OASIS desalted, fractionated using OFFGEL and stage-tipped before LC-MSMS analysis. Raw files were analyzed with MaxQuant/Perseus. Experiment's ID number: SA590.

### 5.1.3.1. Method comparison and cut-off determination

We first compared the methods with regard to identification of biotinylated proteins (with intensities in the light channel), quantification, and reproducibility. While method 1 resulted on average in lower numbers of proteins identified, the other three approaches performed similarly and combination method 4 yielded with a quantification (number of proteins with quantified ratios/total number of proteins) and identification of proteins in biotin channel (number of proteins with intensities in light channel/total number of proteins) rates of 67% and 78% the best results (Figure 5-14).

## Results



**Figure 5-14. Identification and quantification efficiencies for each combination method.**

Light labeled HEK 293T cells were transiently transfected with Nucl.-BioID construct and biotinylated (50  $\mu$ M - 24 h); medium and heavy labeled cells served as controls. Biotinylated and the 2 control samples were combined 1:1:1 in different steps of the on-bead digestion experiment. Method 1 (L\* + H / M): Biotinylated proteins in light were mixed before incubation on bead with heavy control sample and further mixed with medium control before trypsin digestion. In method 2 (L\* / M + H) medium and heavy control samples were mixed before incubation with beads and mixed with biotinylated samples before trypsin digestion. In method 3 (L\* / M / H) and 4 (L\* + M + H) samples were combined before trypsin digestion and before incubation on bead, respectively. Samples were OASIS desalted, fractionated using OFFGEL and stage-tipped before LC-MSMS analysis. Raw files were analyzed with MaxQuant/Perseus. A) The number of total proteins as well as average of proteins with quantified L\*/H and L\*/M ratios and identified based on intensities in biotin (light) channel in the three replicates are shown in the bar graph. Venn diagrams show the number proteins shared among the methods with B) quantified L\*/H values in all 3 replicates of each method C) quantified L\*/M values in all 3 replicates of each method D) identified based with intensities in biotin (light) channel in all 3 replicates of each method. Experiment's ID number: SA590.

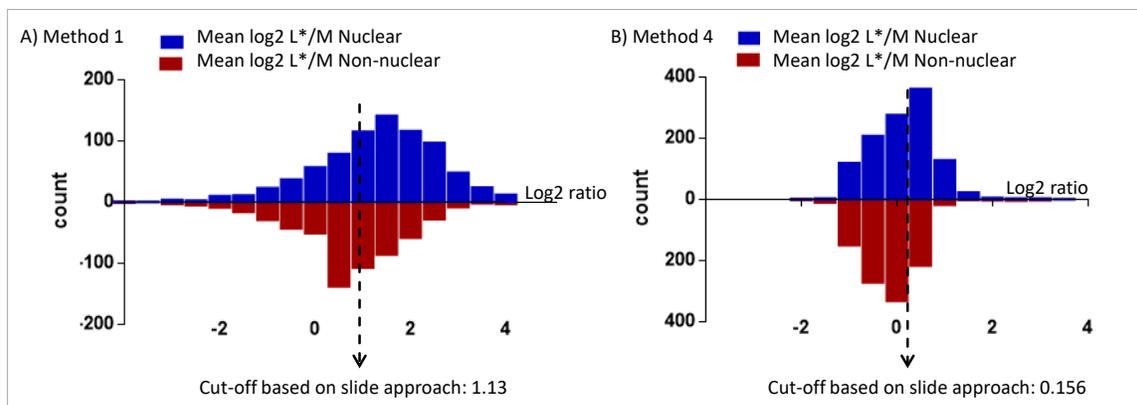
We next compared the lists of shared identified and quantified proteins (found in all replicates of each method) among the four methods. Combination of proteins before incubation with streptavidin beads (L\* + M +H/ method 4) was found to be 3 to 4 fold more effective in quantification (Figure 5-14 B and C) and 3 to 8 fold better in the identification of unique proteins in light channel (Figure 5-14 D) in comparison to other methods.

## Results

This is probably due to less technical variation introduced to biotinylated and control samples in each replicate of method 4, as they are mixed at the very beginning of the protocol which in turn contributes to similar biotin/control ratios among the replicates. We also determined the Pearson correlation coefficient for each method to investigate the reproducibility and dependence within the replicates of each experiment. Method 4 acquired the highest correlation coefficient (0.933) across the methods; this means it is the most reproducible method between the biological replicates (Appendix Figure 4). We concluded that in terms of identification, quantification, and reproducibility, method 4 outperforms the other methods.

Next, in order to find a statistical cut-off based on the control channels we analyzed the proteins lists in Perseus by loading the H/M (control/control) normalized ratios. We filtered for the proteins quantified in all replicates, calculated the mean of the H/M ratios of the three replicates and determined the cut-off ratio covering 95% of the dataset (Detailed information on calculation of 95% cut-off for biotinylation is provided at section 5.1.1.). The cut-offs determined were as follows: 1.07, 0.63, 0.31, and 0.14 for methods 1, 2, 3, and 4, respectively.

To further evaluate the validity of our cut-offs, we compared them with the cut-offs proposed by slide approach for two of our experiments. We focused on these two methods as they are the approaches that best reflect the inter-effects of biotinylated and control samples (described in section 5.1.3.2). We created two lists of nuclear and non-nuclear proteins out of the proteins lists of method 1 ( $L^* + H / M$ ) and 4 ( $L^* + H + M$ ) based on GO ontology annotations and determined the maximum Youden index based on ROC curve analysis. Interestingly the cut-offs suggested for greatest Youden index were  $\log_2$  ratios 1.13 and 0.156 for methods 1 and 4 that are very similar to the 95% cut-offs we determined for these methods (1.07 and 0.14, respectively) (Figure 5-15); this further validated our cut-off determination strategy.



**Figure 5-15. “Slide approach” cut-off for biotinylation in combination methods 1 and 4.**

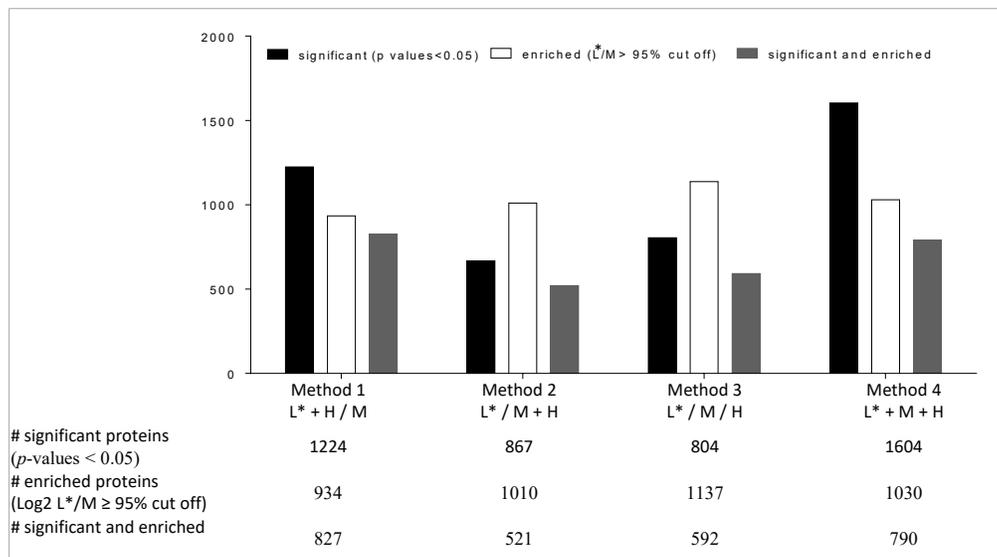
Light labeled HEK 293T cells were transiently transfected with Nucl.-BioID construct and biotinylated (50  $\mu$ M - 24 h); medium and heavy labeled cells served as controls. Biotinylated and the 2 control samples were combined 1:1:1 in different steps of the on-bead digestion experiment. Method 1 ( $L^* + H / M$ ): Biotinylated proteins in light were mixed before incubation on bead with heavy control sample and further mixed with medium control before trypsin digestion. In method 4 ( $L^* + M + H$ ) samples were all combined before incubation on bead. Samples were OASIS desalted, fractionated using OFFGEL and stage-tipped before LC-MSMS analysis. Raw files were analyzed with MaxQuant/Perseus.

A and B) Two lists of true positives (nuclear/blue) and false positive (non-nuclear/red) proteins were created based on GO analysis and their  $L^*/M$  ratios used to create the ROC curve. Youden index was calculated for all the points of a ROC curve and the maximum youden index suggested +1.13 and +0.156 as the optimum cut-offs for biotinylation for methods 1 and 4, respectively. Experiment’s ID number: SA590.

## Results

To find the biotinylated proteins in each list, we only used L\*/M (biotin/control) ratios (and not the L\*/H ratios) as for the cut-off ratios the H/M (control/control) values had been used. We used the same denominator for both ratios (M) to be able to correlate the control/control and biotin/control distributions and apply the 95% cut-off determined from control/control ratios on biotin/control ratios.

We loaded the list of L\*/M (biotin/control) normalized ratios; filtered out the potential contaminants, log<sub>2</sub> transformed the accepted ratios and kept the proteins only if all 3 ratios were quantified. After a one-sample t-test, we filtered for the significant proteins with *p*-values lower than 0.05 and applied the 95% cut-off calculated for the respective method (Annex Table 5-6). Method 4 yielded the highest number of significant proteins (in black) with *p*-values < 0.05 and method 3 resulted in the best enrichment efficiency (in white; proteins with ratios above biotinylation cut-off) and the number of biotinylated and significant proteins (in grey) were found higher in methods 1 and 4 in comparison to the other methods (Figure 5-16) (For detailed explanation on data analyses see 4.5.3).



**Figure 5-16. Efficiency of different combination methods for purification of biotinylated proteins.**

Light labeled HEK 293T cells were transiently transfected with Nucl.-BioID construct and biotinylated (50 μM - 24 h); medium and heavy labeled cells served as controls. Biotinylated and the 2 control samples were combined 1:1:1 in different steps of the on-bead digestion experiment. Method 1 (L\*+ H/M): Biotinylated proteins in light were mixed before incubation on bead with heavy control sample and further mixed with medium control before trypsin digestion. In method 2 (L\* / M + H) medium and heavy control samples were mixed before incubation with beads and mixed with biotinylated samples before trypsin digestion. In method 3 (L\* / M / H) and 4 (L\* + M + H) samples were combined before trypsin digestion and before incubation on bead, respectively. Samples were OASIS desalted, fractionated using OFFGEL and stage-tipped before LC-MSMS analysis. Raw files were analyzed with MaxQuant/Perseus. Proteins with all three L\*/M ratios quantified were kept and a one sample t-test was performed. The 95% cut-off was calculated based on log<sub>2</sub> H/M ratios and applied on log<sub>2</sub> L\*/M ratios. Bar graphs show the significant proteins with *p*-values lower than 0.05 (black), enriched proteins with L\*/M ratios higher than the 95% cut-off value (white) and significant and enriched proteins with *p*-values < 0.05 and ratios > 95% cut-off for biotinylation (grey). Experiment's ID number: SA590.

To compare the enrichment potencies of different methods, we detected the localization of the proteins passing the cut-off and significant (with *p*-values < 0.05) in each method. We compared the significant and enriched proteins in our lists with proteins annotated as nuclear in GO, Cell ATLAS, and literature (Branon et al. 2018) for each of the methods. Proteins with nuclear annotation comprised 85%, 80%, 81%, and 78% of the datasets

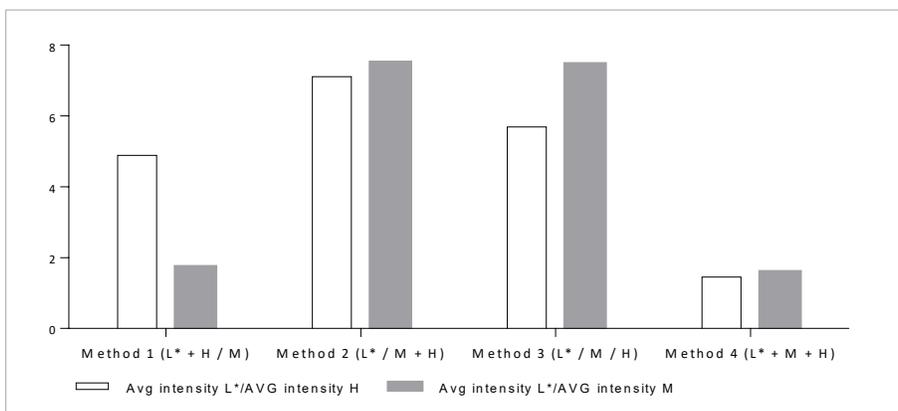
## Results

for method 1 to 4, with 719, 422, 484, and 619 proteins, respectively. This means that the efficiency of all combination methods for enrichment of truly biotinylated proteins is similar.

The similarity of the 95% cut-off determined using the two control samples and the “Slide approach” cut-off validated the credibility of our statistical approach. While the percentage of proteins defined as biotinylated was slightly higher in method 1, method 4 (incubation of all samples with beads) has shown to be the most effective approach in other respects (quantification, identification, reproducibility, and significance).

### 5.1.3.2. Intensity analysis to find the inter-effect of biotinylated and control proteins

To assess potential effects of biotinylated proteins on such binding non-specifically and vice versa we plotted the average intensities of SILAC channels in all methods and compared them with emphasis on methods 1 and 4, as they can show the effect of biotinylated and control samples on each other. In method 1, the biotinylated sample is in the same tube with the heavy control samples until digestion. This resembles the mixing of organelle biotinylated samples with control sample before binding to beads (See Figure 5-7). The mixture of heavy and light samples meets the medium labeled control sample at digestion step; medium control sample has been processed in a separate tube. In method 4 however, both controls are combined with biotinylated proteins before incubation on beads. We determined the average intensity in each SILAC channel and plotted the average intensity of light (L\*) channel over the heavy (H) and the medium (M) channels (Figure 5-17).



**Figure 5-17. Intensity analysis for SILAC channels in different combination methods.**

Light labeled HEK 293T cells were transiently transfected with Nucl.-BioID construct and biotinylated (50  $\mu$ M - 24 h); medium and heavy labeled cells served as controls. Biotinylated and the 2 control samples were combined 1:1:1 in different steps of the on-bead digestion experiment. Method 1 (L\* + H / M): Biotinylated proteins in light were mixed before incubation on bead with heavy control sample and further mixed with medium control before trypsin digestion. In method 2 (L\* / M + H) medium and heavy control samples were mixed before incubation with beads and mixed with biotinylated samples before trypsin digestion. In method 3 (L\* / M / H) and 4 (L\* + M + H) samples were combined before trypsin digestion and before incubation on bead, respectively. Samples were OASIS desalted, fractionated using OFFGEL and stage-tipped before LC-MSMS analysis. Raw files were analyzed with MaxQuant/Perseus.

Average intensities for light (biotin), medium (control), and heavy (control) channels were determined and the two biotin/control intensity ratios were plotted for each method. Experiment's ID number: SA590.

The results suggest a possible reason for the high biotin/control values in organelle proteomics datasets with one control channel. For method 1 the average intensity of the light (biotin) SILAC channel is 4.5 fold higher than the average intensity of the heavy control proteins; however, the ratio is less than two for L\*/M, where medium labeled proteins have been incubated and washed on beads in a separate tube (Figure 5-17). In method 4 however, both average intensity ratios of L\*/M and L\*/H are similar (less than 2). The only difference here is that the second control (in the medium channel) has been mixed with the other two samples before incubation

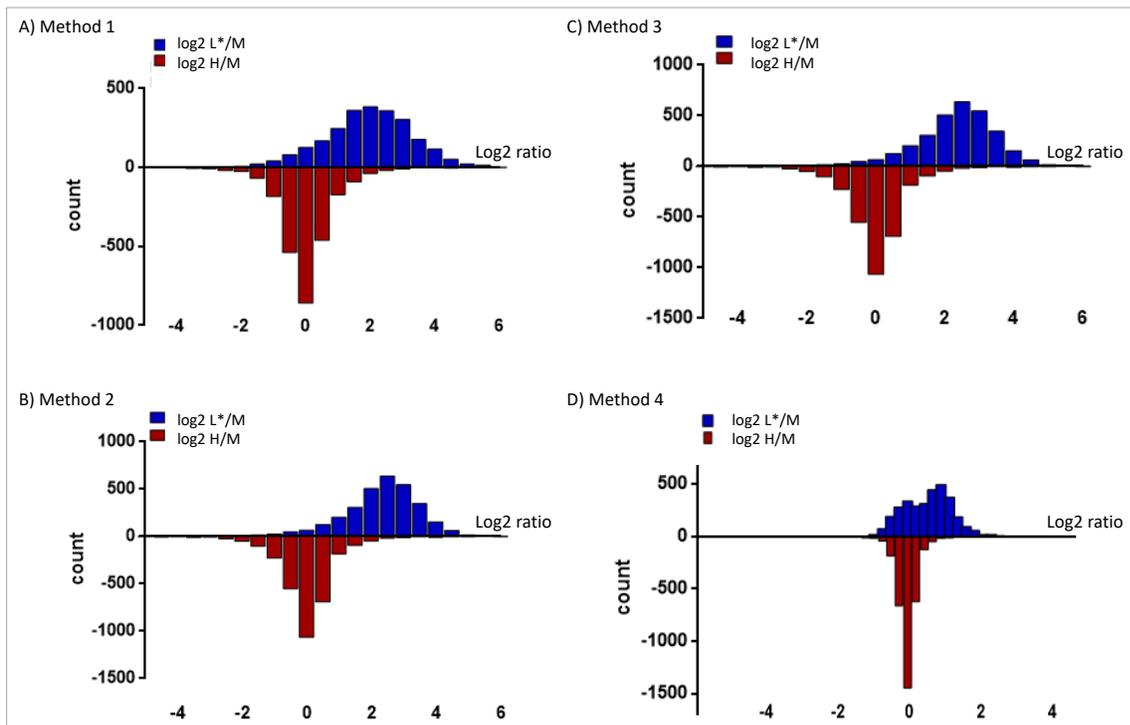
## Results

with beads and not afterwards. This seemingly has resulted in similar intensities for proteins in both control channels and consequently similar L\*/M and L\*/H average intensity ratios. One explanation would be that in 3plex studies (including 2 control sample), the background proteins present in three populations have a higher chance to bind the beads and be detected by mass spectrometer and create the null distribution. Incubation of control proteins in a separate tube (experiment 2; L\*/M+H) and each of the samples separately (Experiment 3; L\* / M / H), result in higher L\*/M and L\*/H values. However, in comparison to method 4, these two ratios are not as similar and related to each other.

### 5.1.3.3. The best method for combination of biotinylated and control samples

To find the best combination of samples resulting in efficient enrichment and finding the most reliable cut-off we monitored the distribution of  $\log_2$  biotin/control (L\*/M) and  $\log_2$  control/control (H/M) ratios in each method. We determined the average of all biotin/control ratios and used the values to create a histogram for each dataset. Since the two control channels were included, we expected a distinct separation of populations originating by the ratios of the two control channels (H/M) versus such determined by the biotinylated and the control channel (L\*/M): The proteins in control channels bind the beads with lower strength and therefore acquire lower intensities which form a control/control ratio distribution around 1, or  $\log_2$  ratio of 0. However, the biotinylated proteins with higher affinity to the beads, are enriched and obtain higher intensities which result in formation of a cluster with biotin/control ratios  $> 1$  (or  $\log_2$  ratios  $> 0$ )

Combination of all three samples before incubation with the streptavidin beads (method 4) resulted in a clear separation of the dataset into two populations: the background cluster with L\*/M ratios around 0 which overlays perfectly with H/M background ratios and a second cluster for biotinylated proteins with  $\log_2$  L\*/M ratios around 0.9. In all other three methods, a larger population of proteins acquires high biotin/control values resulting in a more uniform distribution of the biotin/control and control/control ratios (Figure 5-18).



## Results

### Figure 5-18. Distribution of $\log_2$ L\*/M and H/M ratios values in different combination methods.

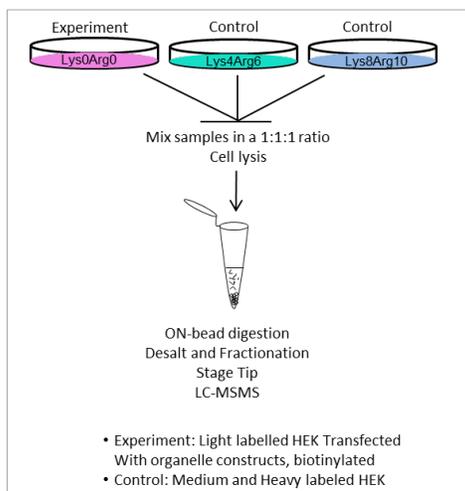
Light labeled HEK 293T cells were transiently transfected with Nucl.-BioID construct and biotinylated (50  $\mu$ M - 24 h); medium and heavy labeled cells served as controls. Biotinylated and the 2 control samples were combined 1:1:1 in different steps of the on-bead digestion experiment. Method 1 (L\* + H/M): Biotinylated proteins in light were mixed before incubation on bead with heavy control sample and further mixed with medium control before trypsin digestion. In method 2 (L\* / M + H) medium and heavy control samples were mixed before incubation with beads and mixed with biotinylated samples before trypsin digestion. In method 3 (L\* / M / H) and 4 (L\* + M + H) samples were combined before trypsin digestion and before incubation on bead, respectively. Samples were OASIS desalted, fractionated using OFFGEL and stage-tipped before LC-MSMS analysis. Raw files were analyzed with MaxQuant/Perseus. A to D)  $\log_2$  L\*/M (blue) and H/M (red) ratio distribution were plotted as mirror histograms for methods 1 to 4. Experiment's ID number: SA590.

Method 4 (samples mixed before incubation with beads) was the only method which created a bimodal histogram for L\*/M (biotin/control) ratios that clearly divided the proteins into two populations. This was similar to our expectations of biotin/control ratios explained in section 5.1.1. Therefore, it should be possible to find a 95% cut-off using the control/control ratios (normally distributed) within the study which overlaps the first cluster of biotin/control ratios (including non-biotinylated proteins) and apply it on the second cluster (including biotinylated proteins). The unique distribution pattern of ratios in method 4 could be due to the inter-effects of biotinylated and non-specific binding proteins when combined from the beginning of the protocol.

Method 4 also resulted in the highest number of identified proteins in biotin channel (with intensities > 0 in L\* channel), quantified proteins (with quantified L\*/M and L\*/H ratios), and significant proteins ( $p$ -values < 0.05), and the strongest correlation among the individual replicates (0.93). The 95% cut-off determined for this approach also matched the slide approach cut-off. We, therefore, decided to perform all subsequent experiments with this combination strategy.

### 5.1.4. Analysis of organelle-specific proteomes including two control samples

The 95% cut-off determined from the null experiment (section 5.1.1) was not credible to separate biotinylated proteins from such binding non-specifically to the beads in organelle proteomics experiments including one control sample (2plex studies-section 5.1.2). We showed that by the inclusion of a second control in the experimental setting it is possible to determine a cut-off within the study and apply it on biotin/control ratios to detect the biotinylated proteins. We repeated the organelle proteomics in a 3plex setting and combined the biotinylated proteins (light) and the non-biotinylated control samples (medium and heavy) in a 1:1:1 ratio (1 mg from each) before incubation with beads (Figure 5-19; based on combination method 4 in section 5.1.3).



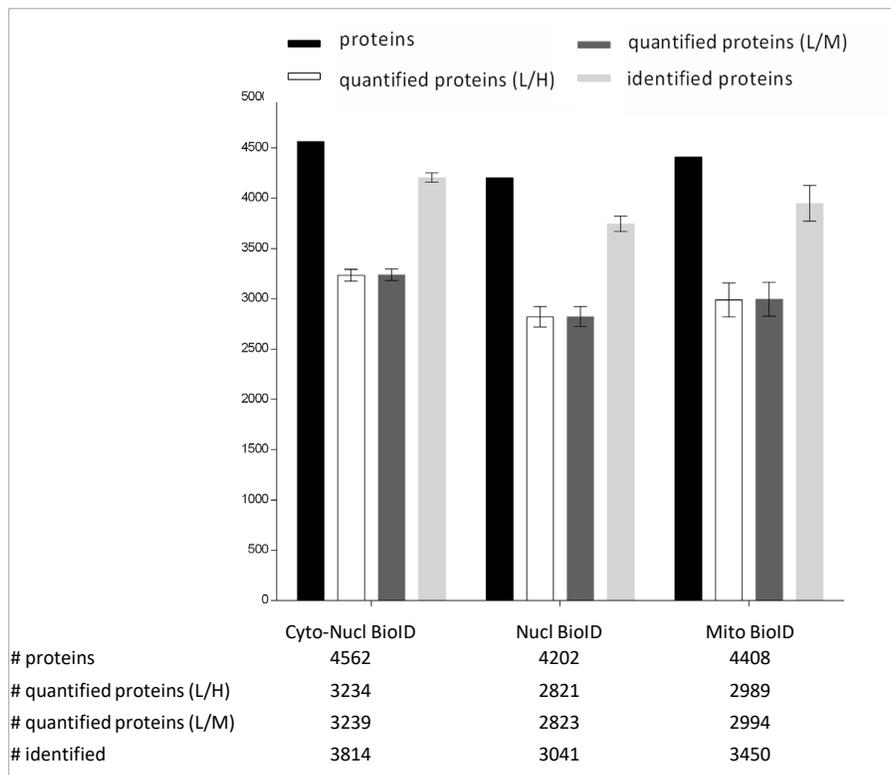
### Figure 5-19. Organelle proteomics (including two control samples)

HEK 293T light labeled cells were transiently transfected with BioID organelle constructs biotinylated (50  $\mu$ M - 24 h), and lysed, and mixed 1:1:1 with HEK 293T medium and heavy labeled control cell lysate followed by on-bead digestion. Samples were OASIS desalted, fractionated using OFFGEL and stage-tipped before LC-MSMS analysis. Raw files were analyzed with MaxQuant/Perseus. Experiment's ID number: SA5107.

## Results

HEK 293T light cells were transiently transfected with the different BioID constructs. 48 h post-transfection, cells were supplemented with 50  $\mu$ M biotin (final concentration), incubated for 24 h, and harvested. Heavy and medium labeled untransfected HEK 293T cells were used as controls. Equal amounts of experiment and control samples (1 mg from each) were combined and incubated with streptavidin beads, followed by washing of the beads, reduction, alkylation and digestion of the proteins. Peptides were fractionated by OFFGEL, desalted by OASIS cartridges, and analyzed using LC-MSMS (Orbitrap-Velos- 60 min gradient) (Figure 5-19) MaxQuant searches were performed on pooled raw files for the three replicates of each individual experiment (Annex Table 5-7).

The total number of proteins, proteins quantified based on L/H and L/M, and identified in biotin channel (based on intensities above 0 in the light channel) in all replicates were comparable in all experiments (Figure 5-20) The number of proteins quantified based on L/M and L/H channels were similar in each of the experiments.



**Figure 5-20. Identification and quantification efficiency for each organelle BioID dataset (including two control samples)**

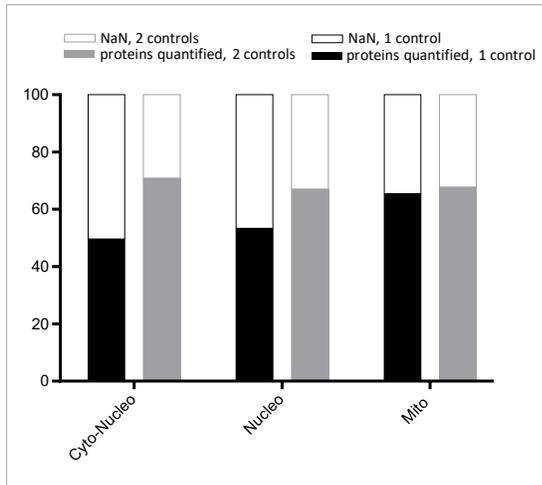
HEK 293T light labeled cells were transiently transfected with BioID organelle constructs biotinylated (50  $\mu$ M - 24 h), lysed, and mixed 1:1:1 with HEK 293T medium and heavy labeled control cell lysates followed by on-bead digestion. Samples were OASIS desalted, fractionated using OFFGEL and stage-tipped before LC-MSMS analysis. Raw files were analyzed with MaxQuant/Perseus.

The number of total proteins as well as the proteins with quantified L/H and L/M ratios and identified based on intensities in biotin (light) channel in the three replicates are shown in the bar graph. Experiment's ID number: SA5107.

In comparison to organelle proteomics with one control channel quantification rates were slightly improved for Mito.-BioID (from 65% to 68%) dataset and significantly increased for CytoNucl.-BioID (from 50 to 71%) and

## Results

Nucl.-BioID (from 53 to 67%) datasets (Figure 5-21). Improvement in quantification, in general, can be due to higher abundance of background proteins (in 2 samples instead of one) which in turn allows their quantification by the algorithm. The reason for the more pronounced increase in case of the nuclear and CytoNuclear datasets could be the proteins from these two organelles represent the majority of non-specifically binding proteins and that the current approach allows for more of them to be quantified as twice the amount of background proteins are used.

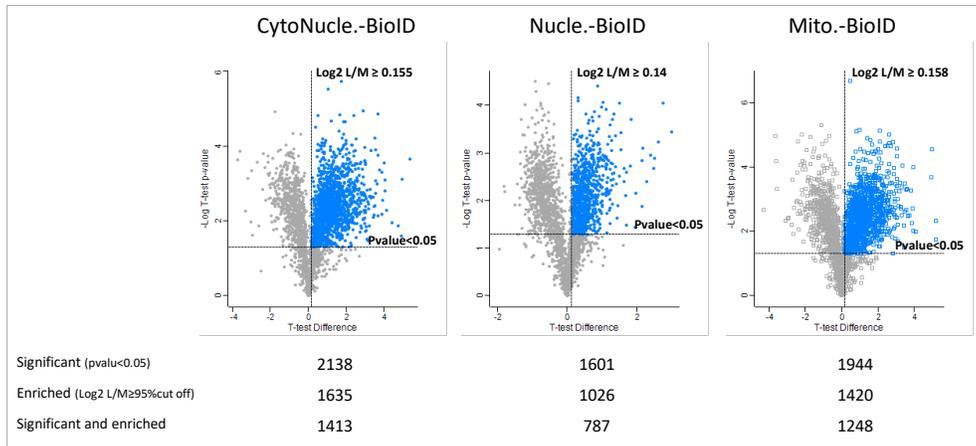


**Figure 5-21. Comparison of quantification rates between organelle proteomics experiments including one and two control populations.**

In the approach including one control sample, HEK 293T light labeled cells were transiently transfected with BioID organelle constructs biotinylated (50  $\mu$ M - 24 h) and lysed and mixed 1:1 heavy labeled control cell lysate followed by on-bead digestion. In the approach including two control samples, a medium labeled control cell lysate was added to the other two samples. Samples were OASIS desalted, fractionated using OFFGEL and stage-tipped, measured by LC-MSMS and analyzed by MaxQuant/Perseus. Bar graph shows the percentage of proteins with quantified L/H ratios in organelle experiments over the total number of proteins for single control approach (black) vs double control approach (grey). NaN: Not a number.

Next, in order to find a statistical cut-off based on the control channels we analyzed the proteins lists in Perseus by loading the H/M (control/control) normalized ratios. We filtered for the proteins quantified in all replicates, calculated the mean of the H/M ratios of the three replicates and determined the cut-off ratio containing 95% of the proteins in control samples (Detailed information on calculation of 95% cut-off for biotinylation can be found at section 5.1.1.). For CytoNucl., Nucl., and Mito.-BioID datasets ratios were 0.155, 0.141, and 0.158, respectively. We applied these ratio cut-offs on  $\log_2$  biotin/control ratios. To define the biotinylated proteome in each organelle dataset, we loaded the L/M (biotin/control) ratios in Perseus, removed the potential contaminants,  $\log_2$  transformed the accepted ratios and kept the proteins only if all 3 L/M ratios were quantified. After a one-sample t-test, we filtered for the significant proteins with  $p$ -values lower than 0.05 and applied the 95% cut-off calculated for the respective organelle dataset (Figure 5-22; Annex Table 5-8). The proteins with  $p$ -values < 0.05 (significant) and  $\log_2$  L/M ratios > 95% cut-off (enriched) were assigned to biotinylation.

## Results



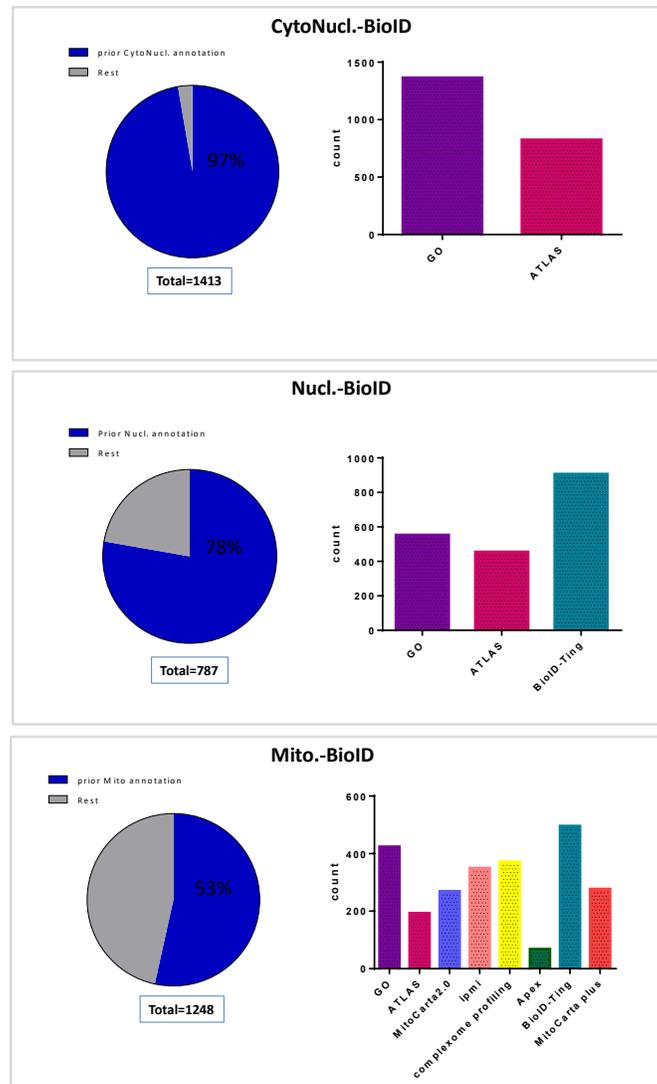
**Figure 5-22. Efficiency of different BioID constructs for purification of biotinylated proteins (including two control samples)**

HEK 293T light labeled cells were transiently transfected with BioID organelle constructs biotinylated (50  $\mu$ M - 24 h), lysed, and mixed 1:1:1 with HEK 293T medium and heavy labeled control cell lysates followed by on-bead digestion. Samples were OASIS desalted, fractionated using OFFGEL and stage-tipped before LC-MSMS analysis. Raw files were analyzed with MaxQuant/Perseus. Proteins with 3 quantified L/M ratios were subjected to biotinylation analysis. A one sample t-test was performed and proteins with normalized L/M (biotin/control) log<sub>2</sub> ratios higher than the 95% cut-off (enriched) and *p*-values < 0.05 (significant) were visualized in volcano plot in blue and defined as biotinylated. Experiment's ID number: SA5107.

To validate the correct localization of the biotinylated proteins, the proteins enriched (with L/M ratios greater than the 95% cut-off) and significant (with *p*-values < 0.05) were compared with GO (geneontology.org), cell ATLAS (www.proteinatlas.org; (Regev et al. 2017)), and other literature data lists especially the tables provided by Ting lab (TbID, miniID, and BioID, (Branon et al. 2018)) annotated with respective organelles. The CytoNucl.-BioID dataset with 97% of the proteins previously annotated as cytoplasmic or nuclear had the highest coverage by literature. 78% of the proteins in Nuclear dataset were previously identified in GO, Cell ATLAS, and nuclear datasets provided by Ting lab. The number of proteins enriched in nuclear dataset was quite small (787) which could be due to low transfection efficiency in this sample in comparison to the other samples.

Mitochondrial BioID list was compared with GO, Cell ATLAS, MitoCarta and MitoCarta plus (Calvo, Clauser, and Mootha 2016), IPMI (Smith and Robinson 2016), complexome profiling (Wessels et al. 2013), APEX (Rhee et al. 2013), and Ting lab (Branon et al. 2018) datasets and in total 53% of the proteins were already identified by literature ((Branon et al. 2018); Figure 5-23).

## Results



**Figure 5-23. Organelle proteomes vs literature.**

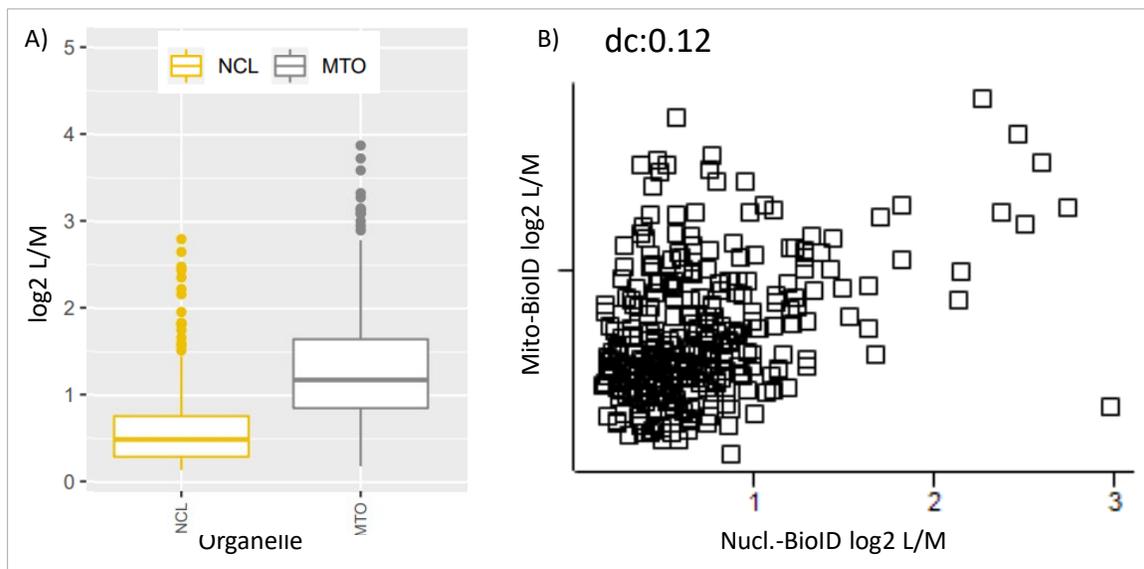
HEK 293T light labeled cells were transiently transfected with BioID organelle constructs biotinylated (50  $\mu$ M - 24 h), lysed, and mixed 1:1:1 with HEK 293T medium and heavy labeled control cell lysates followed by on-bead digestion. Samples were OASIS desalted, fractionated using OFFGEL and stage-tipped before LC-MSMS analysis. Raw files were analyzed with MaxQuant/Perseus. Proteins with 3 quantified L/M ratios were subjected to biotinylation analysis. A one sample t-test was performed and proteins with normalized L/M (biotin/control)  $\log_2$  ratios higher than the 95% cut-off (enriched) and  $p$ -values  $< 0.05$  (significant) were called biotinylated.

The list of biotinylated proteins for each organelle dataset was compared to GO, Cell ATLAS, and literature. Pie charts show the proteins previously linked to each organelle in blue and those not previously identified in grey. Bar graphs show the number of proteins in our datasets previously linked to the organelle of interest for each individual database/publication. Experiment's ID number: SA5107.

It is reported that organelles share proteomes; more than half of the cellular proteins have been linked to multiple organelles (Geladaki et al. 2019; Thul et al. 2017). These proteins however may have different roles and abundance in each organelle. We compared proteins with  $p$ -values  $< 0.05$  which passed the biotinylation cut-off across the datasets (Figure 5-24). As the CytoNuclear dataset shares proteins with the nuclear dataset, we kept it out for these analyses. We identified 295 proteins shared between Mito.- and Nucl.-BioID datasets which showed different ratio distribution. Most of the nuclear proteins acquired lower  $\log_2$  ratios and formed a unimodal histogram. Mitochondrial proteins obtained  $\log_2$  ratios higher than the other dataset in the shape of a

## Results

unimodal histogram (Figure 5-24 A). The majority of nuclear proteins acquired smaller biotin/control ratios which can be due to their lower abundance in the nucleus in comparison to the mitochondria. Based on other publications, proteins associated with different cellular compartments show distinct fold change distribution due to their different abundances and functions (Parca et al. 2018). These differences in distribution are the basis for creating particular profiles and allocating proteins with unknown location to organelles (Geladaki et al. 2019). We finally performed distance correlation analysis to see how different the ratios for the enriched proteins shared between the datasets are. Small values for distance correlation reflect a high variation and independence of datasets. This value was 0.12 for Mito. vs Nucl.-BioID (Figure 5-24 B) which shows a very low correlation between these two datasets. Low values for distance correlation reflect the different ratio distribution of the proteins common among the datasets that are probably due to their different abundances and roles in each organelle.



**Figure 5-24. Distribution of  $\log_2 L/M$  ratios and distance analysis for common biotinylated proteins in nuclear and mitochondrial datasets**

HEK 293T light labeled cells were transiently transfected with BioID organelle constructs biotinylated (50  $\mu$ M - 24 h), and lysed, and mixed 1:1:1 with HEK 293T medium and heavy labeled control cell lysate followed by on-bead digestion. Samples were OASIS desalted, fractionated using OFFGEL and stage-tipped before LC-MSMS analysis. Raw files were analyzed with MaxQuant/Perseus. Proteins with 3 quantified L/M ratios were subjected to biotinylation analysis. T-test was performed and proteins with normalized L/M  $\log_2$  ratios higher than the cut-off calculated based on H/M ratios (enriched) and  $p$ -values < 0.05 (significant) were called biotinylated.

A) The list of biotinylated proteins shared among nuclear and mitochondrial datasets were compared for  $\log_2 L/M$  ratio distribution and visualized as box (left side). MTO: mitochondrial; NCL: nuclear. B) The list of biotinylated proteins shared among nuclear and mitochondrial datasets were compared for distance correlation among their  $\log_2 L/M$  ratios using Perseus software. Experiment's ID number: SA5107.

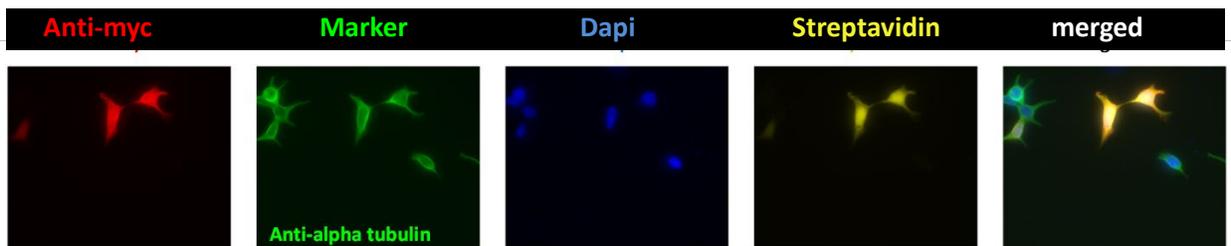
We showed the possibility of applying our newly established setting (biotinylated+control1+control2 mixed before incubation with beads) for the enrichment of organelle-specific biotinylated proteins and confirmed the correct localization of the enriched proteomes with literature/datasets lists of proteins previously annotated as belonging to these organelles.

## 5.2. *In vivo* cell-type-specific analysis in BioID mice models

Besides implementation of BioID technique in cell culture, establishment of cut-off determination method, and enrichment for organelle proteomes, we expanded the application of the technique to *in vivo* cell-type-specific labeling. The *in vivo* project aims to address certain diseases that involve a specific cell type within a tissue (like beta cells in case of diabetes) and therefore indicate the need for techniques for reliable cell-type-specific assignment of proteins to investigate the molecular pathology of such diseases.

### 5.2.1. Generation of BioID mice

To allow for cell-type-specific proteome analyses, in collaboration with groups of Professor Schorle (Institute for Developmental Pathology; University of Bonn) and Professor Fleischmann (Institute for Physiology 1; University of Bonn) two transgenic mouse lines were generated which allow for the regulation of myc-BioID expression through the Cre-loxP system under *Coll1a1* (mouse 1) and *Rosa26* (mouse 2) loci, respectively. In cells expressing BioID, it locates to the cytosolic and nuclear compartment (Figure 5-25) and should, therefore, allow for biotinylation of all proteins exposed to either of them.

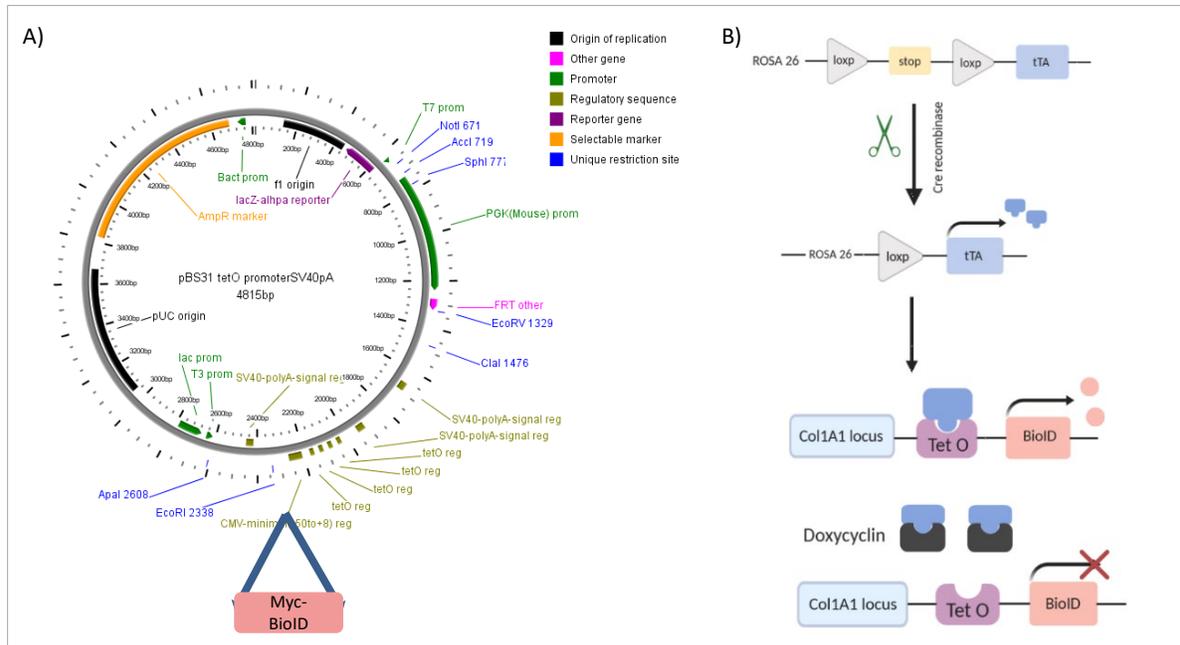


**Figure 5-25. Localization of BioID in cytosol and nucleus.**

NIH/3T3 cells were transfected with myc-tagged BioID construct, treated with biotin (50  $\mu$ M - 24 h), fixed and stained with anti-alpha tubulin antibody [DM1A] (ab195887, Abcam, 1: 400), anti-myc (Abcam, ab9106, 1: 400), and Alexa Fluor 647 streptavidin (S21374, ThermoFischer scientific, 1: 400) and mounted with fluoromout-G mounting solution containing DAPI for nuclear staining. The merged figure shows the expression and biotinylation by BioID in cytosol and nucleus of the cell. Experiment's ID number: SA4178.

For mouse 1, generated in collaboration with the group of professor Schorle, using flp-in recombinase, BioID was placed at the *Coll1a1* locus of KH2ES cells downstream of a TetO sequence (Figure 5-26). Correct genomic insertion was confirmed using Southern blot and doxycycline (Dox)-inducible BioID expression as well as biotinylation of cellular proteins were verified by Western blot (Appendix Figure 5). In these mice, cell-type-specificity can be controlled by breeding with loxP-STOP rtTA and tTA mice and expression of BioID can be induced by administration or withdrawal of doxycycline/tetracycline, allowing for spatiotemporal control of protein expression.

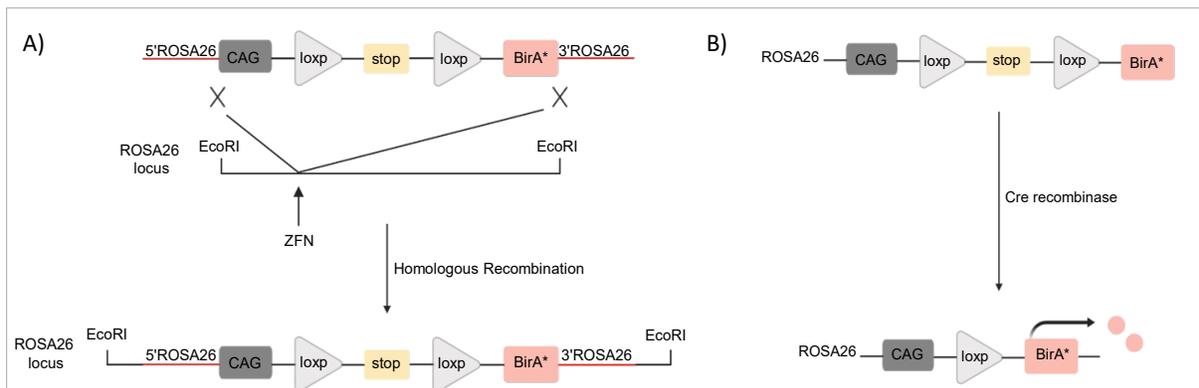
## Results



**Figure 5-26. Generation and mechanism of action in Col1a1-BioID mouse (mouse 1)**

A) BioID gene was amplified by PCR (template: Add gene Plasmid #35700) and cloned into pBS31 TetO promoter SV40pA plasmid using EcoRI site for ligation. B) Using the Flp-in system BioID was placed in the Col1A1 locus of KH2ES cells downstream of a TetO promoter. In Col1a1-BioID mouse, Tet operator is located upstream of BioID and induces expression of the gene upon crossing with two other mouse strains: LoxP-Stop-loxP-tTA and a Cre mouse. Upon presence of Cre recombinase, the stop cassette upstream of tetracycline transactivator protein (tTA) is removed and tTA binds the TetO sequence and drives the expression of BioID.

For mouse 2, generated in collaboration with the group of professor Fleischmann, BioID was inserted in the Rosa26 locus downstream of a CAG promoter. Protein expression is blocked by a loxP site flanked STOP cassette which is removed in a cell-type-specific manner by breeding with the respective Cre-lines (Figure 5-27).



**Figure 5-27. Generation and mechanism of action in Rosa26-CAG-BioID mouse (mouse 2)**

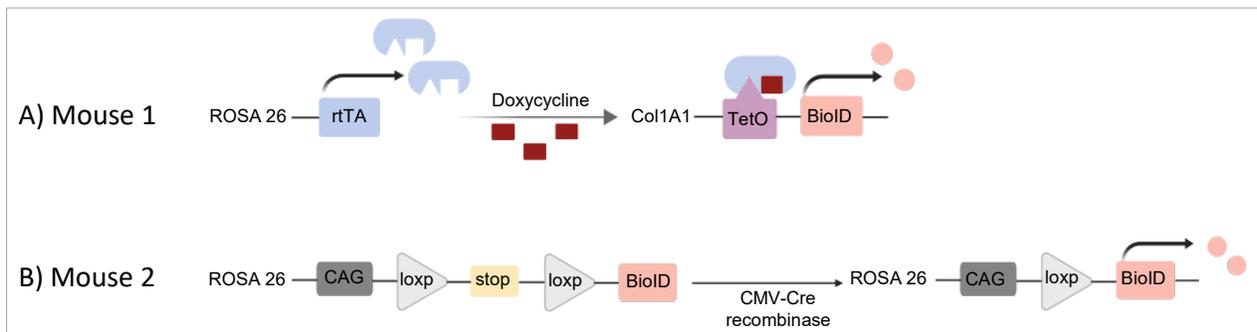
A) Myc-tagged BioID was cloned into the Ai6 Rosa26 vector (template: Add gene Plasmid #22798) to replace zsgreen. Co-transfection of G4 mouse ES cells with the targeting vector and zinc finger nucleases resulted in the integration of Myc tagged-BioID within the genomic Rosa26 locus by homologous recombination. B) loxP flanked stop cassette inhibits transgene expression. The stop cassette upstream of BioID is removed upon crossing with a Cre line and BioID starts catalyzing the biotinylation of proteins in the presence of biotin.

## Results

Insertion of the transgene, as well as the functionality of the construct for selected ES cell clones, was confirmed by PCR (Appendix Figure 6). The ES cells were also transfected with cDNA expressing Cre recombinase and analyzed by q-PCR and immunofluorescence to examine the transgene insertion (Appendix Figure 7). For the generation of transgenic mice, murine blastocysts were injected with cells of a positive clone and the resulting chimaeras were crossed with wild type mice to generate Rosa26-CAG-BioID mouse line on a mixed background.

### 5.2.2. Functionality and tissue expression of BioID mice

F1 generations of these mice were tested by PCR and bred into a C57Bl6 background. Initially, for investigation of BioID expression and biotinylation, we crossed Colla1-BioID/mouse 1 with a Rosa26-rtTA (resulting in Colla1-BioID-rtTA) and Rosa26-CAG-BioID/mouse 2 with CMV-Cre (resulting in CAG-BioID-CMV) lines. In mouse 1, rtTA binds the TetO promoter (located upstream of the transgene) and starts the BioID expression in all tissues and cell lines where Rosa26 and Colla1 loci are active (Figure 5-28 A). In mouse 2, Cre recombinase removes the STOP cassette upstream of BioID and results in BioID expression in all tissues and cell lines where Rosa26 and CMV are active (Figure 5-28 B). In other words, the expression level in different cell populations depends on the overall transcriptional activity of the loci/promoter under which the transgenes are located. We chose these two mice (Rosa26-rtTA and CMV-Cre) to investigate the expression pattern of BioID in different tissues due to the fact that Rosa26 and CMV are known as ubiquitous promoters allowing generalized expression in the mouse. We, therefore, can link the weak/strong BioID expression in any cell types of these mice to the potency of each BioID promoter to drive expression rather than to the activity of the promoters upstream of rtTA/Cre.



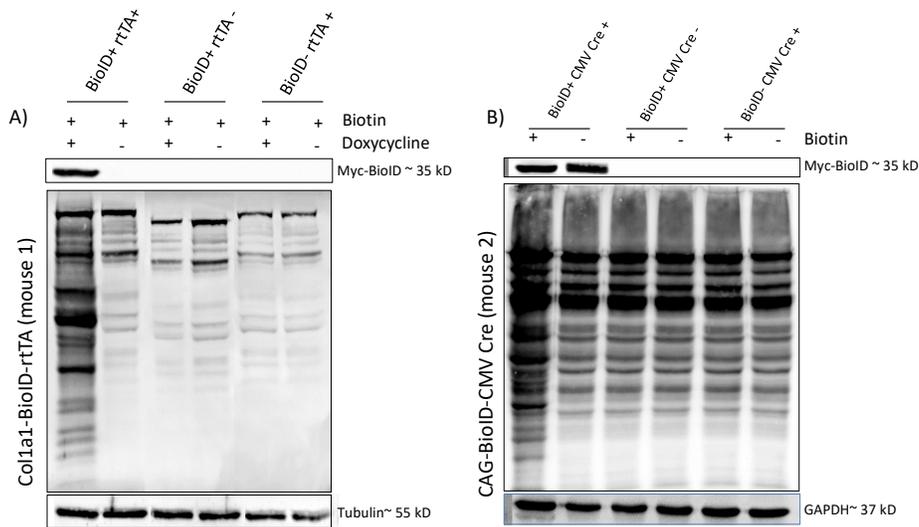
**Figure 5-28. Generation of mice with ubiquitous BioID expression.**

A) Upon crossing Mouse 1 (Colla1-BioID) with a Rosa26-rtTA, in access to Dox, rtTA is expressed and binds the TetO sequence and induces the expression of BioID. B) Upon crossing mouse 2 (Rosa26-CAG-BioID) with CMV-Cre, Cre recombinase excises the STOP cassette and induces expression of BioID.

The positive mice (BioID<sup>+</sup>rtTA<sup>+</sup> and BioID<sup>+</sup>CMV Cre<sup>+</sup>) used for the next experiments were heterozygote for both genes (BioID and Cre/rtTA). To evaluate the functionality of the genetic system, upon administration of biotin (mouse 1 and 2; 0.22 mg/ml) and Dox (mouse 1; 2 mg/ml) in the drinking water for seven days to mice with different genotypes, we compared biotinylation levels for liver tissue obtained from 8-week-old Colla1-BioID-rtTA and CAG-BioID-CMV Cre mice. The expression of BioID was only observed in tissues from mice harboring both genes (BioID and CMV/rtTA) and upon administration of Dox in case of mouse 1. Biotinylation by BioID was only seen in tissues from mice with the correct genotype upon administration of biotin (/and Dox in case of mouse 1 required to induce gene expression).

## Results

CAG-BioID-CMV Cre mice express BioID throughout their whole life span and the mouse food contains biotin as it is an essential vitamin (400 µg/kg, Rod 16-A, LASvendi); we wondered if this can result in undesired biotinylation. We did not observe any differences in global biotinylation levels between mice expressing BioID and control mice (Figure 5-29 B; from left, lane 2 vs lanes 3 to 6). As the expression is induced only in mice with correct genotype (and upon administration of Dox in case of mouse 1) we concluded that these genetic systems are functional, inducible, and non-leaky (Figure 5-29). Furthermore, since no undesired biotinylation occurs in mice with lifetime BioID expression (mouse 2) we concluded that effective biotinylation by BioID is dependent on administration of external biotin (0.22 mg/ml in water). All BioID mice appeared healthy and we did not detect any differences in behavior, weight gain, or reproduction between BioID and control mice.

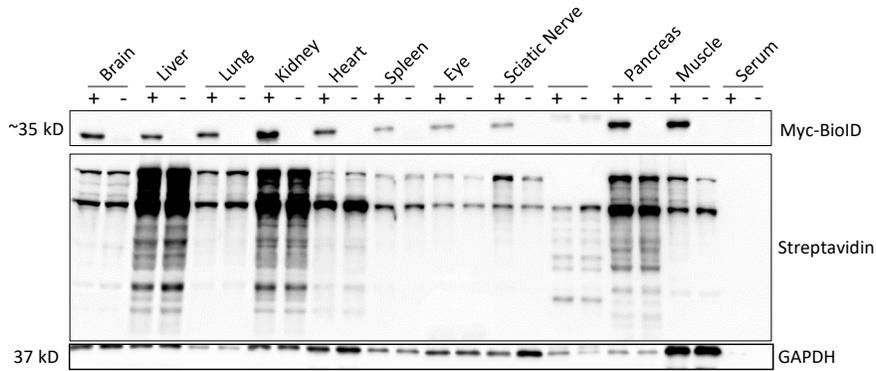


**Figure 5-29. Evaluation of functionality and leakiness of the genetic system-Western blot.**

Dox (mouse 1; 2 mg/ml) and biotin (mouse 1 and 2; 0.22 mg/ml) in water were administered to mice with different genotypes for 7 days. Mice were sacrificed, liver isolated, lysed, and 75 µg from each sample loaded on SDS gel and analyzed by Western blotting. Anti-myc (Abcam, ab9106, 1: 3,000; blocked with 5% milk in TBS-T) and HRP-conjugated streptavidin (21126, ThermoFischer scientific: 1: 10,000; blocked with 3.5% BSA in PBS-T) were used to detect the expression of BioID and biotinylation by BioID. Anti-beta-3 Tubulin Monoclonal Antibody (MA119187, Dianova, 1: 5,000; blocked with 5% milk in TBS-T) and anti-GAPDH antibody (Fl-335, Santa Cruz, 1: 1,000; blocked with 3.5% BSA in PBS-T) were used as loading controls. +: treated; -: not treated.

We further tested other tissues of mouse 2 (with continuous BioID expression), to ensure biotin in mouse food does not lead to unwanted biotinylation in any of them. In order to investigate this, we compared the level of biotinylation in 12 tissues isolated from an 8-week-old mouse 2 vs a control mouse when no biotin in water was administered; biotinylation level was equal in both mice which means the biotin in food does not lead to detectable labeling of protein in any organs (Figure 5-30).

## Results

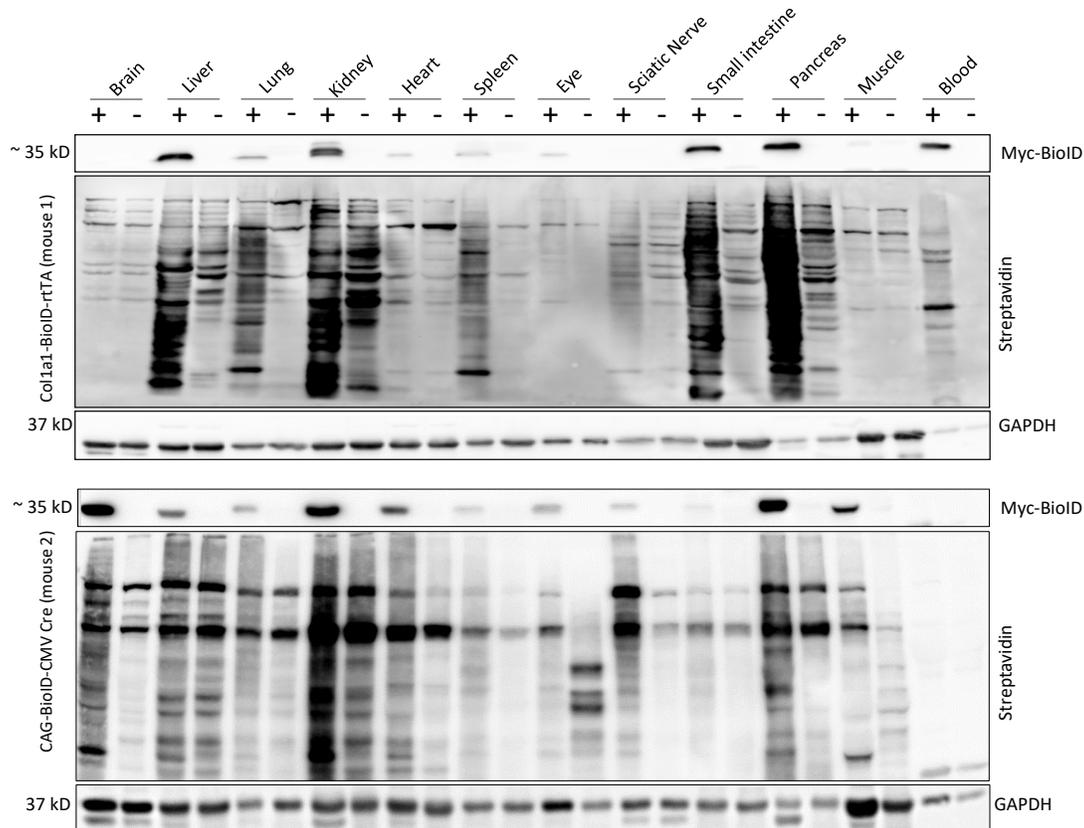


**Figure 5-30. Evaluation of internal biotinylation in different organs of Rosa26-CAG-BioID mouse (mouse 2) - Western blot.**

A CAG-BioID-CMV Cre mouse (mouse 2) and a negative control mouse (lacking one of the genes) were sacrificed, tissues isolated, lysed, 50  $\mu$ g from each sample loaded on SDS gel and analyzed by Western blotting. Anti-myc (Abcam, ab9106, 1: 3,000; blocked with 5% milk in TBS-T) and HRP-conjugated streptavidin (21126, ThermoFischer scientific: 1: 10,000; blocked with 3.5% BSA in PBS-T) were used to detect the expression of BioID and biotinylation by BioID. Anti-GAPDH antibody (FI-335, Santa Cruz, 1: 1,000; blocked with 3.5% BSA in PBS-T) was used as loading control. +: CAG-BioID-CMV Cre mouse; -: control mouse.

We were interested in levels of BioID expression in different tissues. We evaluated the expression of BioID and biotinylation in different tissues of mouse 1 and mouse 2 upon administration of Dox (mouse 1; 2 mg/ml) and biotin (mouse 1 and 2; 0.22 mg/ml) in the drinking water for seven days by Western blot (Figure5-31). The expression and biotinylation by BioID were evaluated by anti-myc antibody and streptavidin, respectively.

## Results

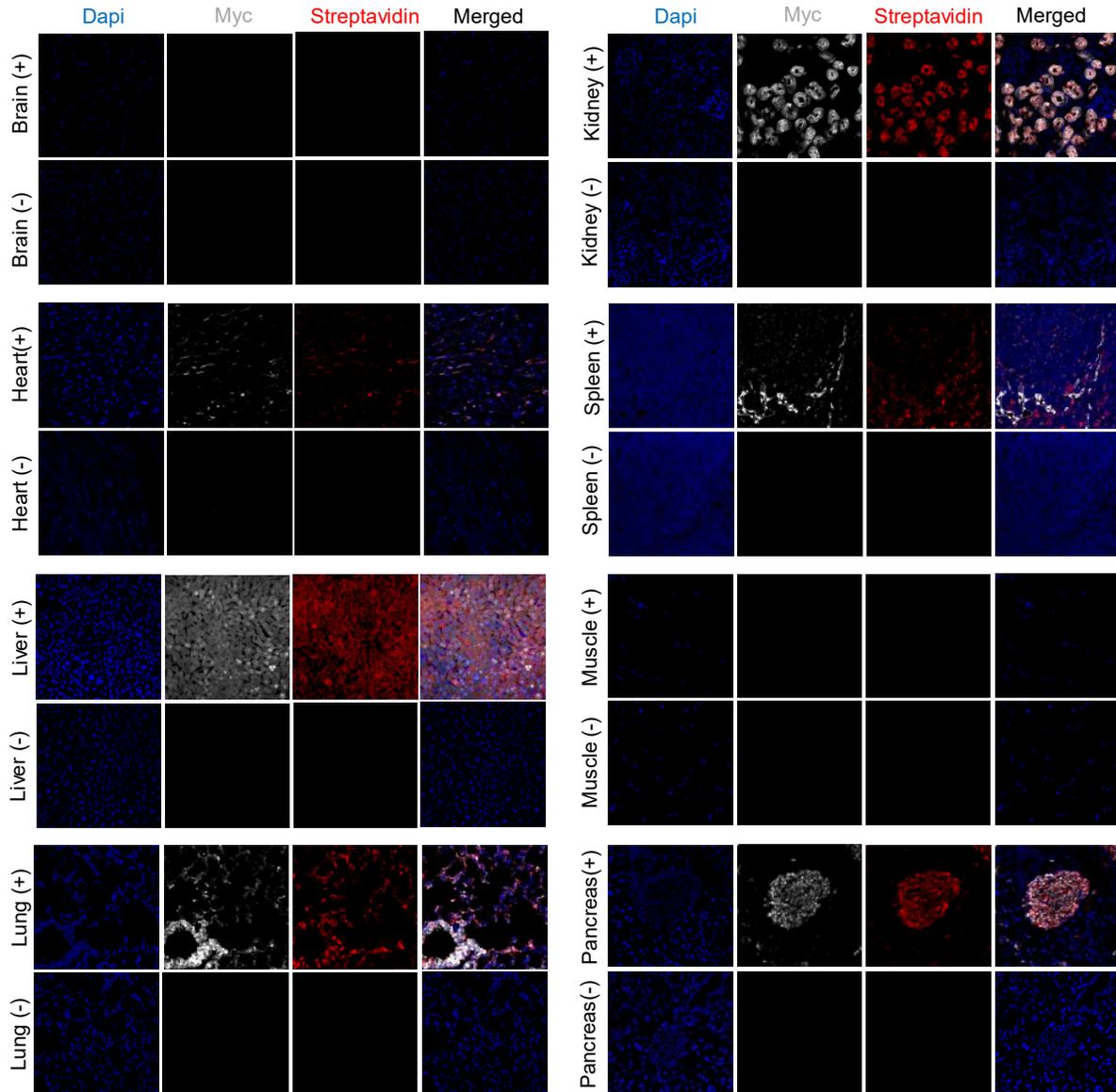


**Figure 5-31. BioID expression and biotinylation in Tissues-Western blot.**

Dox (mouse 1; 2 mg/ml) and biotin (mouse 1 and 2; 0.22 mg/ml) in water were administered to mice for 7 days. Mice were sacrificed, tissues isolated, lysed, 75  $\mu$ g from each sample loaded on SDS gel and analyzed by Western blotting. Anti-myc (Abcam, ab9106, 1: 3,000; blocked with 5% milk in TBS-T) and HRP-conjugated streptavidin (21126, ThermoFischer scientific: 1: 10,000; blocked with 3.5% BSA in PBS-T) were used to detect the expression of BioID and biotinylation by BioID. Anti-GAPDH antibody (F1-335, Santa Cruz, 1: 1,000; blocked with 3.5% BSA in PBS-T) was used as loading control. +: Col1a1-BioID-rtta mouse or CAG-BioID-CMV Cre mouse; -: control mouse.

For both mice, we observed BioID in several tissues with varying strength. The levels of protein expression correlated with the observed degree of biotinylation. Kidney showed high expression and biotinylation in both mouse lines while small intestine/liver and heart/muscle were more pronounced in mouse 1 and 2, respectively. We were only able to detect BioID in the blood of mouse 1, while for sciatic nerve/ brain we observed expression solely in mouse 2. We selected 8 tissues for further analysis by immunofluorescence staining (Figure 5-32: mouse 1 and 5-33: mouse 2) to detect the cell populations expressing BioID. The expression and biotinylation by BioID were evaluated by anti-myc (grey) antibody and streptavidin (red), respectively.

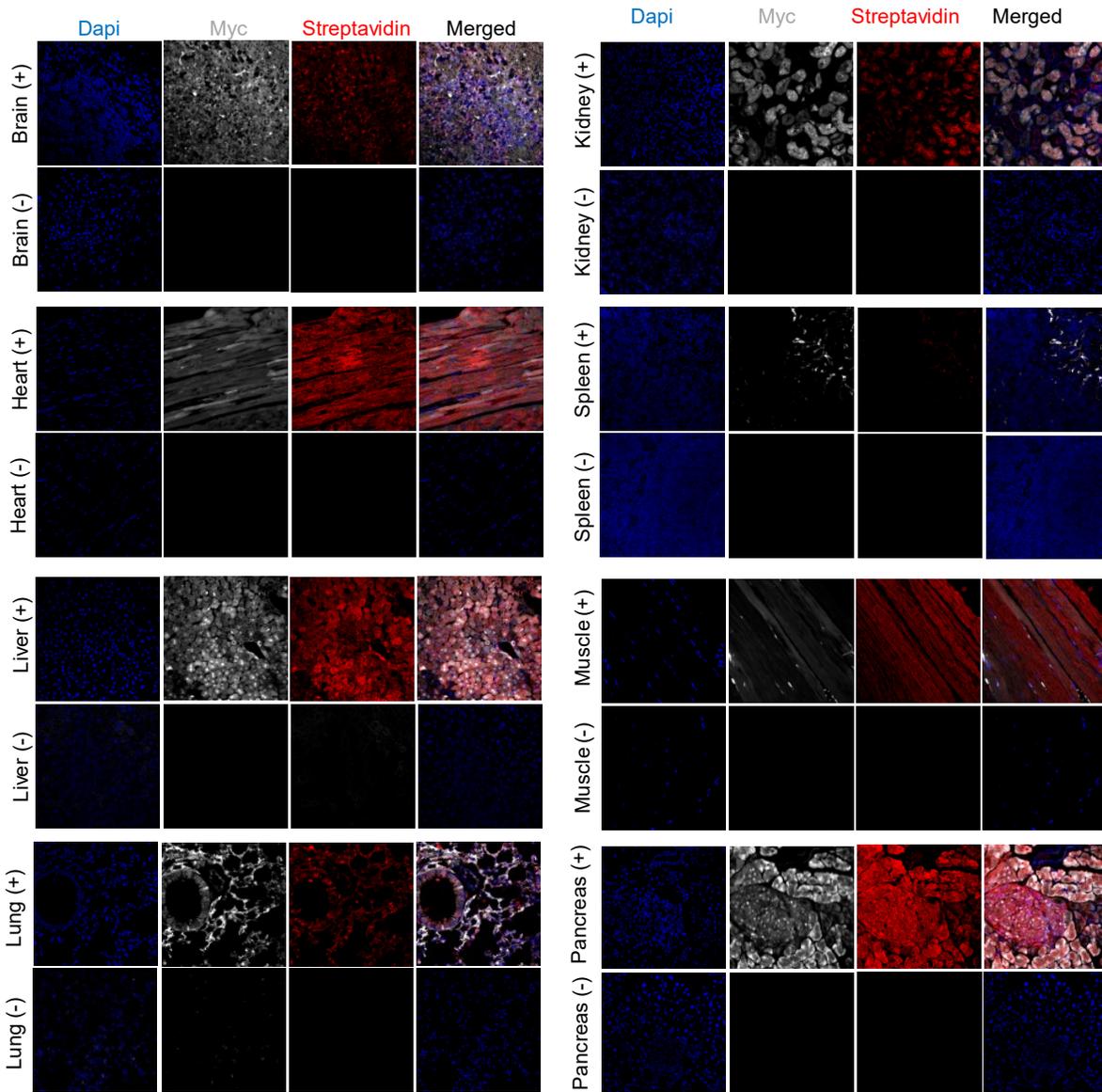
## Results



**Figure 5-32. BioID expression and biotinylation in mouse 1 tissues - Immunofluorescence.**

Dox (2 mg/ml) and biotin (0.22 mg/ml) in water were administered to a Coll1a1-BioID-rtTA mouse (and a control mouse) for 7 days. Mice were sacrificed; tissue samples from both mice were fixed overnight in 4% PFA and cryosectioned. Goat anti-myc antibody (ab9106, Abcam, 1: 400) was used to evaluate BioID expression. Alexa Fluor 647 streptavidin (405237, Biolegend, 1: 400) was used to detect biotinylation. Nucleus was stained with DAPI (ThermoFischer scientific, Bremen, Germany). The merged figure shows the expression and biotinylation by BioID in each tissue sample. +: Coll1a1-BioID-rtTA mouse; -: control mouse. Staining and microscopy were performed by Dr. Kenichi Kimura from the Institute of Physiology 1, University of Bonn.

## Results



**Figure 5-33. BioID expression and biotinylation in mouse 2 tissues - Immunofluorescence.**

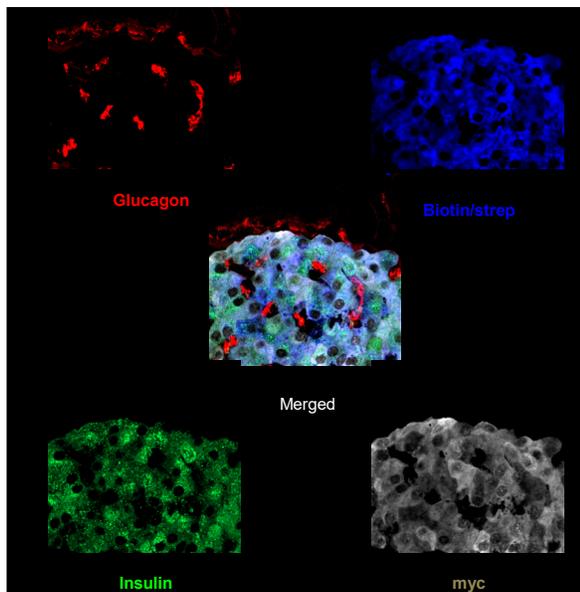
Biotin (0.22 mg/ml) was administered to a CAG-BioID-CMV Cre mouse (and a control mouse) for 7 days. Mice were sacrificed; tissue samples from were fixed overnight in 4% PFA and cryosectioned. Goat anti-myc antibody (ab9106, Abcam, 1: 400) was used to evaluate BioID expression. Alexa Fluor 647 streptavidin (405237, Biolegend, 1: 400) was used to show biotinylation. Nucleus was stained with DAPI (ThermoFischer scientific, Bremen, Germany). The merged figure shows the expression and biotinylation by BioID in each tissue sample. +: CAG-BioID-CMV mouse; -: control mouse. Staining and microscopy were performed by Dr. Kenichi Kimura from the Institute of Physiology 1, University of Bonn.

Both mice show a low and patchy expression profile of BioID in the spleen. In mouse 2, BioID expression (and biotinylation) can be observed in most of the cell types of pancreas, brain, muscle, lung, liver, kidney and heart. Mouse 1 shows high expression in liver and kidney, patchy expression in heart and lung and no expression in muscle and brain. In general, based on Western blot and microscopy results, in comparison to mouse 1, BioID expression and biotinylation were more pronounced in mouse 2 and detected in more cell populations. This is

## Results

due to the higher efficiency of Rosa26 (in combination with CAG-enhancer) in comparison to Colla1 locus in driving transgene expression.

We also observed perfect co-localization of myc-BioID and biotinylation confirming that BioID system is precise and specific; the reactive biotin is only contained in the cells of interest and is added to vicinity proteins residing in the same region. In mouse 1, we found an interesting pattern of BioID expression and biotinylation in pancreas mainly restricted to islets of Langerhans. This encouraged us to check the precise localization of BioID with higher magnification using anti-insulin (beta cell marker) and anti-glucagon (alpha cell marker) antibodies to assess if the expression is limited to beta cells or can be seen in alpha cells as well. Interestingly, the expression and biotinylation were restricted to beta cells of pancreatic islets and again perfect co-localization of myc and streptavidin signals proved the specificity of the system in proximity labeling (Figure 5-34).



**Figure 5-34. BioID expression and biotinylation in pancreatic islets of mouse 1.**

Dox (2 mg/ml) and biotin (0.22 mg/ml) in water were administered to a Colla1-BioID-rtTA (mouse 1) mouse for 7 days. Mouse was sacrificed and pancreas sample was fixed overnight in 4% PFA and cryosectioned. For Immunofluorescence analysis the following primary antibodies were used: goat anti-myc antibody (ab9106, Abcam, 1: 400), rabbit anti-insulin antibody [K36aC10] (ab6995, 1: 400), guinea pig anti-insulin antibody (IR002, DAKO), mouse anti-glucagon antibody [K79bB10] (ab10988, 1: 400). Alexa Fluor 421 streptavidin (405225, Biolegend, 1: 400) was used to show biotinylation. Staining and microscopy were performed by Dr. Kenichi Kimura from the Institute of Physiology 1, University of Bonn.

We showed that both mice are functional, inducible, and dependent on external biotin administration. Expression and biotinylation by BioID is more pronounced in mouse 2 (under Rosa26 locus) than in mouse 1 (under Colla1 locus). As both BioID mice show expression and biotinylation within the boundaries of the same cell with no leakage to neighboring cells, we concluded that they have the potency to be used as efficient models for cell-type-specific labeling followed by MS analysis. In case of the mouse 1, predominant expression of BioID in beta cells of pancreatic islets proposes this mouse as an interesting tool for beta-cell-specific proteomics.

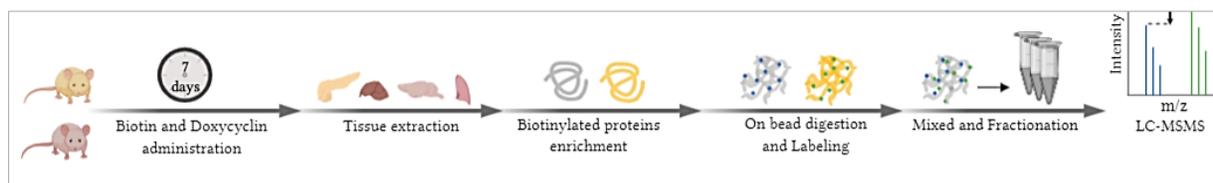
### 5.2.3. MS-based enrichment analysis

Based on the immunofluorescence analysis (Figure 5-33 and 5-34) we proposed mouse 1 as a potential tool for cell-type-specific proteomics in beta cells. However, since beta cells comprise approximately 1–2% of the total pancreatic tissue, we first focused on evaluation and improvement of our enrichment method towards higher sensitivity, credibility, and reproducibility for capturing labeled proteins from such low-abundant cells. We also focused on establishment of methods to define two cut-offs for cell-type-specific data analysis. Cell-type-specific enrichment analysis will include 3 variable factors (control/ biotinylated/ biotinylated and drug administered). For the analysis of such data two cut-offs are required: one to discriminate between biotinylated

## Results

and drug-affected proteomes (drug-effect cut-off) and one to distinguish biotinylated proteome from non-specific binding proteins (biotinylation cut-off).

In order to be able to quantify changes due to specific enrichment and establish approaches for data analysis (biotinylation and drug-effect cut-offs) we performed dimethyl labeling experiments on Coll1a1-BioID-rtTA mice (mouse 1) (Figure 5-35). We put our main emphasis on the pancreas from Coll1a1-BioID-rtTA (mouse 1) to evaluate the capacities of this mouse to be used as a model in beta cell studies. Dimethyl labeling approach is fast, straightforward, and inexpensive; therefore, we used it for enrichment analysis and method development. The positive mice (BioID<sup>+</sup>rtTA<sup>+</sup>) used for the next experiments were heterozygote for both genes (BioID and rtTA). The control mice were negative for either or BioID or rtTA genes. As we showed earlier both these mice are non-functional and do not express BioID. The exact genotype is therefore not reported for each individual experiment.



**Figure 5-35. Schematic of *in-vivo* proteomics workflow using dimethyl for quantification**

Dox (2 mg/ml) and biotin (0.22 mg/ml) in water were administered to Coll1a1-BioID-rtTA (mouse 1) and control mice for seven days. Mice were sacrificed and tissues isolated. Equal amounts (2 mg) of proteins were incubated with beads, washed, reduced, alkylated, and digested. Biotinylated and control peptide samples were labeled with heavy and light dimethyl labeling reagents and combined, OASIS desalted, fractionated using OFFGEL and stage-tipped before LC-MS/MS analysis. Raw files were analyzed with MaxQuant/Perseus. H/L (biotin/control) dimethyl ratios were analyzed to identify the biotinylated proteome.

### 5.2.3.1. Experimental determination of drug-effect cut-off

A drug-effect cut-off is determined based on the technical variability of biotinylated datasets originating from the same biological sample. Prior to the cell-type-specific experiments which include application of drugs on mice, it is important to establish a method to set up this experimental parameter. In the setting of experiments involving application of drugs both cell-type-specific mice will receive biotin, but only one of them is treated with the drug. We aim to find a cut-off which takes into account the alterations of biotin/biotin ratios due to technical issues and can be used to distinguish biotin enriched proteins from those biotinylated as well as altered/perturbed as a result of the drug administration.

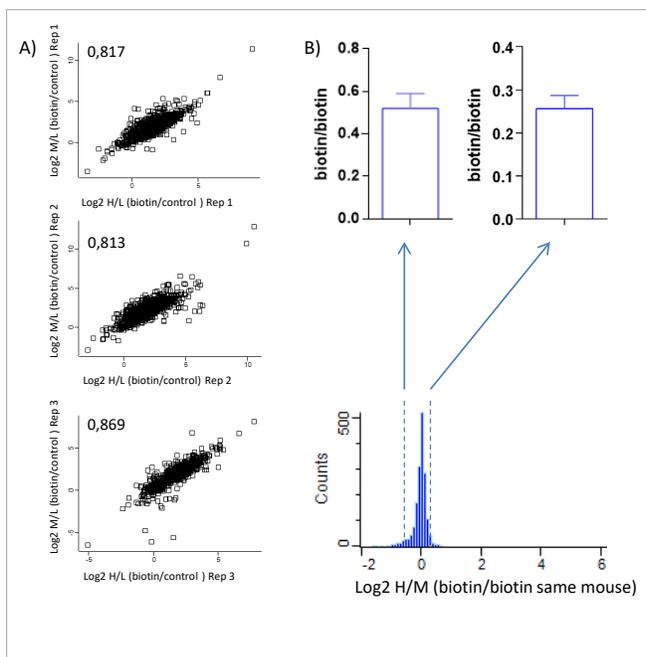
95% cut-off is a generally accepted interval for interpretation of biological data (Fay and Gerow 2013). We hypothesized that two 95% cut-offs on both tails of biotin/biotin ratio distribution would include most of the biotinylated proteins and in case of presence of a drug, any ratio values above and below these up and down intervals can be assigned to (with 95% probability) up- and down-regulation as the result of the drug, respectively.

To this aim, we performed a set of experiments to assess the technical variation between two differentially labeled proteomes originating from the same biotinylated biological sample that were processed similarly in two separate tubes. Three Coll1a1-BioID-rtTA mice (heterozygote for BioID and rtTA) and three control mice (lacking one of the genes) were administered biotin (0.22 mg/ml) and Dox (2 mg/ml) for seven days and sacrificed. Pancreas was isolated and 2 mg from each sample were incubated with streptavidin beads, followed by washing of the beads, reduction, alkylation and digestion of the proteins (section 4.3.10.1.). In each replicate, the same biotinylated sample received heavy and medium dimethyl labeling materials and control sample was labeled with light dimethyl labeling materials. Samples were combined, OASIS desalted,

## Results

fractionated by OFFGEL and analyzed using LC-MSMS (Orbitrap-Velos- 60 min gradient). Raw files were analyzed with MaxQuant (Annex Table 5-9)

The two biotin/control (H/L and M/L) ratios for each replicate showed acceptable Pearson correlations (above 0.8) which means the intensities in both biotin channels of the same animal have been similar resulting in comparable ratios (Figure 5-36 A). We evaluated the H/M (biotin/biotin) ratios in separate replicates and determined the two 95% intervals at both tails of the ratio distribution. The 95% cut-off for  $\log_2$  biotin/biotin ratios were 0.25, 0.23, and 0.29 (mean = 0.25) for upper/right tail and -0.55, -0.57, and -0.44 (mean = -0.52) for the lower/left tail of the three different biotinylated animals. This means in case of a second stimulus (like a drug) in the third dimethyl channel, the effect is only meaningful if  $0.25 \leq \log_2$  biotin+drug/biotin ratios  $\leq -0.52$  (Figure 5-36 B; Annex Table 5-10).



**Figure 5-36. Cut-off for drug effect.**

Dox (2 mg/ml) and biotin (0.22 mg/ml) in water were administered to three Coll1a1-BioID-rtTA (mouse 1) and three control mice for seven days. Mice were sacrificed and pancreas isolated. Equal amounts (2 mg) of proteins were incubated with beads, washed, reduced, alkylated, and digested. Same biotinylated pancreas sample was processed in two separate tubes and labeled with medium and heavy dimethyl reagents, respectively. Control mice samples were labeled with light dimethyl labeling materials and combined with medium and heavy biotinylated samples. Samples were OASIS desalted, fractionated using OFFGEL and stage-tipped before LC-MSMS analysis. Rawfiles were analysed with MaxQuant/Perseus. A) Correlation coefficient between M/L and H/L ratios in each replicate was determined. B) Two 95% cut-off for H/M ratios in each replicate were determined as 0.25 and -0.52 for upper and lower tails and were set as the cut-offs for drug-effect. Experiment's ID number: SA4119-3.

This cut-off enables differentiating biotinylated and biotinylated/drug-affected proteins when dimethyl is applied for labeling of the peptides. In case another labeling material is utilized, a similar approach needs to be taken to define the drug-effect cut-off.

### 5.2.3.2. Experimental determination of a cut-off for biotinylation and tissue proteomics including one control sample

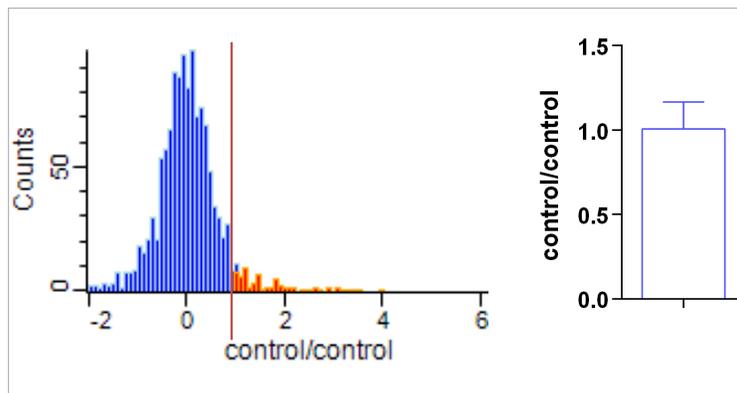
To find a cut-off for biotinylation, similar to *in vitro* organelle-specific proteomics and in parallel to those experiments, we performed on-bead digestion on non-biotinylated pancreas samples as part of a null experiment. The logic behind this experiment is similar to the *in vitro* null experiment and is explained in detail in section 5.1.1. As proteins bind the beads non-specifically they are expected to obtain low intensities and form a distribution of control/control ratios around 1 (or its  $\log_2$ , 0). 95% confidence interval is a commonly accepted cut-off for interpretation of normally distributed biological data (Fay and Gerow 2013). We hypothesized if the ratio distribution is normal, it should be possible to find a cut-off which contains 95% of the

## Results

non-specifically binding proteins and apply it on biotin/control ratios in biotinylation experiments. Since the intensities of biotinylated proteins increase in the sample upon enrichment and biotinylation is a unidirectional effect, the cut-off is defined only on the upper/right tail of the ratio distribution.

For this experiment three control mice (lacking one of the rtTA or BioID genes) were sacrificed, pancreas was isolated and 2 mg from each control sample were incubated with streptavidin beads, followed by washing of the beads, reduction, alkylation and digestion of the proteins (section 4.3.10.1.). Peptides were labeled with light, medium and heavy dimethyl labeling materials and combined, OASIS desalted, fractionated by OFFGEL and analyzed using LC-MSMS (Orbitrap-Velos- 60 min gradient). Raw files were analyzed with MaxQuant (Annex Table 5-11).

We defined the 95% ratio cut-off in control sample using the mean of the three  $\log_2$  of H/L, M/L, and H/M ratios (Annex Table 5-12) the cut-off for biotinylation (Figure 5-37); this was  $\log_2$  ratio +1 that is much greater than the *in vitro* null experiment cut-off ( $\log_2$  ratio +0.18). The difference between *in vitro* and *in vivo* analysis is the step at which samples are combined; normally *in vitro* samples incorporate the labels at the protein level and are combined before incubation with beads (section 5.1.1) resulting in less variation across the samples.



**Figure 5-37. Cut-off determination for biotinylation using control pancreas samples (null experiment)**

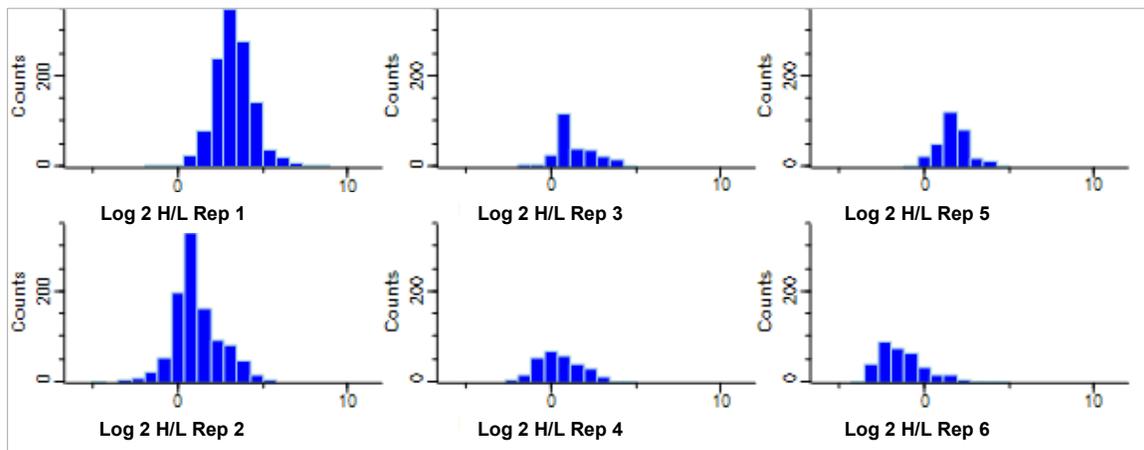
Nine control mice were sacrificed and pancreas isolated. Equal amounts (2 mg) of proteins were incubated with beads, washed, reduced, alkylated, and digested. Control peptide samples were labeled with heavy and light dimethyl labeling reagents and combined, OASIS desalted, fractionated using OFFGEL and stage-tipped before LC-MSMS analysis. Raw files were analyzed with MaxQuant/Perseus. Histogram shows the distribution of the mean of all  $\log_2$  H/L, H/M, and M/L ratios. Dimethyl ratios were used to find a 95% cut-off for biotinylation. The mean of 95% cut-off for the three  $\log_2$  H/L, H/M and M/L ratios was 1. Experiment's ID number: SA4119-2.

Using the 95% cut-off obtained from the null experiment, we next performed enrichment analysis on three organs: liver, lung, and pancreas. As the cut-off value obtained from pancreas null experiments is equal to the arbitrary cut-off widely used in the field ( $\log_2 = 1$ ) we applied it on lung and liver datasets as well to get an approximate evaluation of the enrichment efficiencies in these tissues. Enrichment of liver and lung samples was performed in three replicates and pancreas in 6 replicates. Coll1a1-BioID-rtTA (heterozygote for BioID and rtTA) and control mice (lacking one of the genes) were administered biotin (0.22 mg/ml) and Dox (2 mg/ml) in water for seven days and sacrificed. Pancreas was isolated and 2 mg from each sample were incubated with streptavidin beads, followed by washing of the beads, reduction, alkylation and digestion of the proteins (section 4.3.10.1.). Biotinylated and control mice organs were labeled with heavy and light dimethyl labeling reagents, respectively, combined, OASIS desalted, fractionated by OFFGEL and analyzed using LC-MSMS (Orbitrap-Velos- 60 min gradient) and MaxQuant (Annex Table 5-13)

Due to low data quality in 4 of the pancreas datasets in terms of identification, quantification, and reproducibility (Figure 5-38), we used only two replicates for data analysis. Replicates one and two

## Results

outperformed the other replicates in terms of quantification (Figure 5-38); in general, the quality of pancreas data was lower than the other tissue datasets.



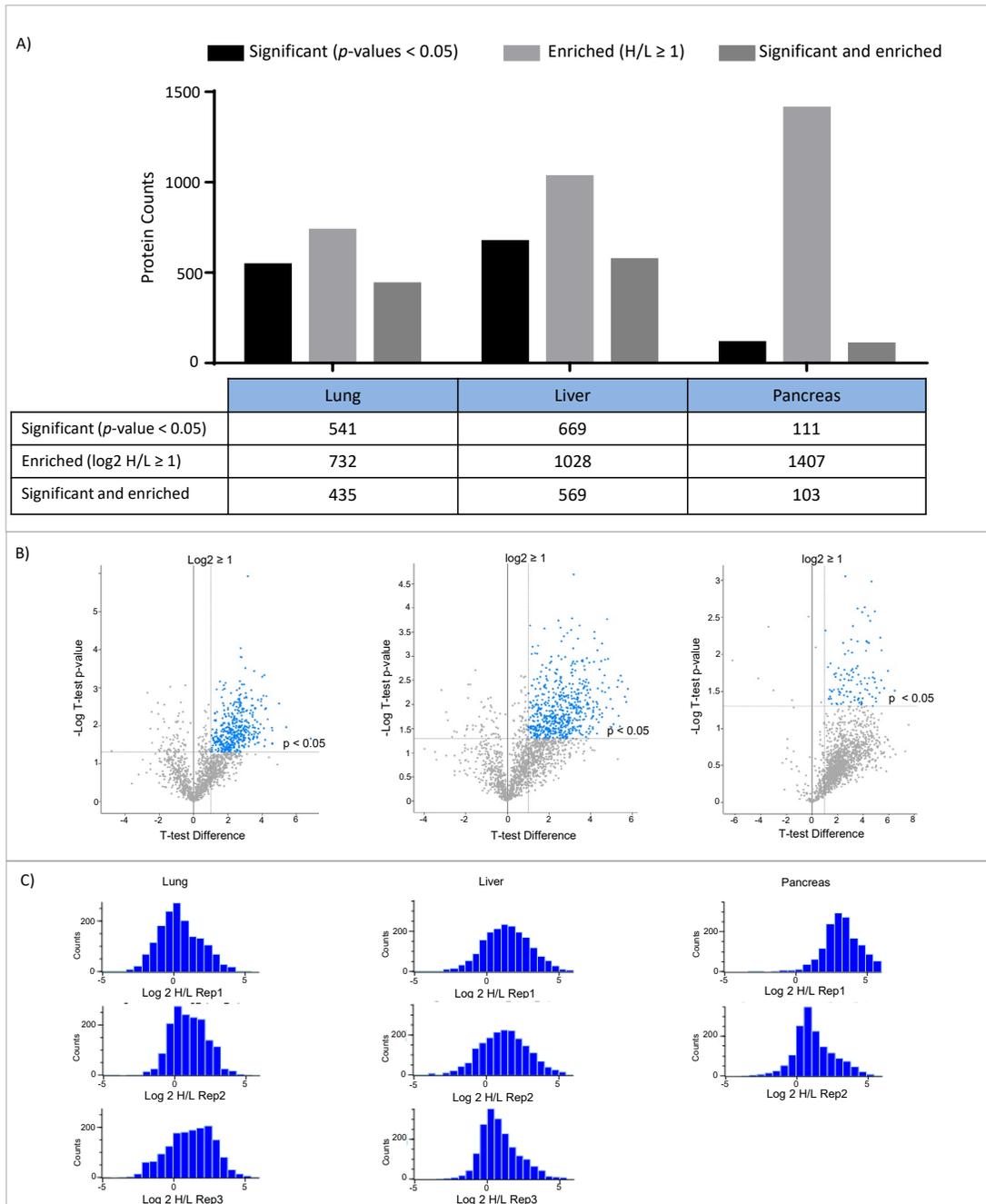
**Figure 5-38. Quality of the 6 pancreas datasets (including one control sample)**

Dox (2 mg/ml) and biotin (0.22 mg/ml) were administered to six Col1a1-BioID-rtTA (mouse 1) and six control mice for seven days. Mice were sacrificed and pancreas isolated. Equal amounts (2 mg) of proteins were incubated with beads, washed, reduced, alkylated, and digested. Biotinylated and control peptide samples were labeled with heavy and light dimethyl labeling reagents and combined, OASIS desalted, fractionated using OFFGEL and stage-tipped before LC-MSMS analysis. Raw files were analyzed with MaxQuant/Perseus. Histograms show the  $\log_2$  H/L (biotin/control) ratio distribution of 6 pancreas datasets. The only 2 datasets with strong correlation are replicates 1 and 2. Experiments' ID numbers: SA4119-1, SA3232, SA440.

For data analysis, H and L intensities were used to create the H/L (biotin/control) ratios. Since the intensities of the background (control) proteins are lower than biotinylated proteins they have a lower chance of being triggered for quantification. The absence of the intensities in control channels can lead to failure in the identification of biotinylated proteins as no biotin/control ratio can be generated. To allow quantification of all enriched proteins, we replaced the missing values in control channels (L) with a small value using the imputation action in Perseus. Imputation assigns random numbers to missing values that are drawn from a normal distribution created from the intensities in the same channel (Tyanova et al. 2016). After a one-sample t-test, the  $\log_2$  fold change cut-off of 1 was used to separate the biotinylated proteins from background (For details on data analysis see 4.5.4.). Proteins with biotin/control ratios higher than the 95% cut-off are defined as enriched and those with  $p$ -values  $< 0.05$  as significant (Annex Table 5-14). Proteins which pass both these criteria (enriched and significant) are assigned as biotinylated. For lung, liver, and pancreas we identified 2691, 2951, and 1608 proteins, and 435 (16%), 569 (19%), 103 (0.06%) had  $p$ -values  $< 0.05$  and  $\log_2$  of biotin/control ratios  $> 1$ , respectively (Figure 5-39 A and B).

The H/L ratio distribution pattern was dissimilar in different tissue samples and even in individual replicates of each experiment (Figure 5-39 C); therefore, a 95% cut-off from a separate null experiment (Figure 5-37) does not have the credibility to be applied on other datasets to find the biotinylated proteins. This was one of the conclusions in the *in vitro* experiments as well. However, since the *in vitro* and *in vivo* experiments were performed in parallel we were not aware of this fact at that point.

## Results



**Figure 5-39. Enrichment analysis and ratio distribution in tissue datasets (including one control sample)**

Dox (2 mg/ml) and biotin (0.22 mg/ml) in water were administered to Coll1a1-BioID-rtTA (mouse 1) and control mice for seven days. Mice were sacrificed and liver, lung, and pancreas were isolated. Equal amounts (2 mg) of proteins were incubated with beads, washed, reduced, alkylated, and digested. Biotinylated and control peptide samples were labeled with heavy and light dimethyl labeling reagents and combined, OASIS desalted, fractionated using OFFGEL and stage-tipped before LC-MS/MS analysis. Raw files were analyzed with MaxQuant/Perseus. H/L (biotin/control) dimethyl ratios were analyzed to identify the biotinylated proteome.

A and B) A one sample t-test was performed and proteins with normalized H/L (biotin/control)  $\log_2$  ratios higher than 1 were called as enriched, those with  $p$ -values < 0.05 were defined as significant, and those enriched and significant were assigned as biotinylated. Bar graphs (A) shows these three groups. Volcano plot (B) shows the biotinylated proteins (significant and enriched) in blue. C) Histograms show the  $\log_2$  H/L (biotin/control) ratio distribution for separate replicates of each tissue proteomics dataset. Rep: replicate. Experiments' ID numbers: SA4119-1, SA440, SA4110.

## Results

In general, the recovery of biotinylated proteins was not satisfactory in any of the experiments. We therefore concluded that the enrichment protocol needs to be improved for detection of biotinylated proteins in tissue samples and especially for the pancreas.

### 5.2.3.3. Method Development

In order to improve the identification and quantification efficiencies as well as reproducibility, especially for pancreas samples, we used biotinylated pancreas samples and compared 5 methods in 2 biological replicates. The difference between these five methods was A. the choice of lysis buffer (Ripa vs 4% SDS) B. the set of wash buffers (2% SDS, 1% TX100, 0.5% NP-40 vs 4 M urea) C. inclusion or absence of a precipitation step before incubation of samples with beads (Figure 5-40). The protocol that has been used in previous experiments used Ripa lysis buffer, contained 2% SDS, 1% TX100, 0.5% NP-40 in the wash buffer, and included no precipitation step (method 1-Figure 5-40).

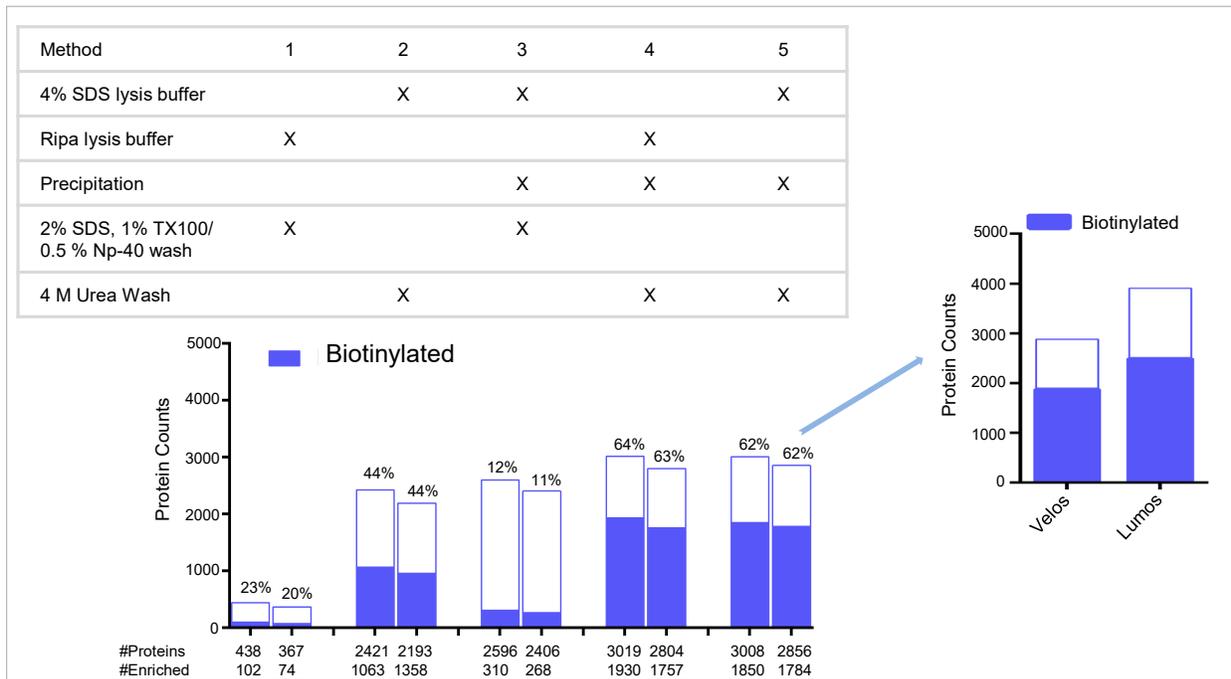
To increase the efficiency of on-bead enrichment methods we tried to exclude poly ethylene glycol (PEG) contamination from the samples. PEG-based molecules are severe contaminations for mass spectrometry as they prevent the peptides from binding the column by concentrating and saturating it. We, therefore, compared the PEG-based wash buffers (containing 1% TX100 and 0.5% NP-40) with urea wash buffers. Another idea to further clean the samples was to replace the Ripa lysis buffer containing 2% TX100, with 4% SDS and further precipitate the protein lysate and resolubilize it in 1% SDS before incubation with beads. The reason for considering precipitation of proteins is to remove the undesired agents (e.g. lipids and nucleic acids) which may interfere with protein assay.

For these experiments, two Col1a1-BioID-rtTA (heterozygote for BioID and rtTA) and two control mice (lacking one of the genes) were administered biotin (0.22 mg/ml) and Dox (2 mg/ml) in water for seven days and sacrificed. We performed on-bead digestion using 2 mg input from each biotinylated/control pancreas sample and based on the differences listed in Figure 5-40. After digestion, biotinylated and control peptide samples were labeled with medium and light dimethyl labeling materials, respectively, combined, OASIS desalted, fractionated by OFFGEL and analyzed using LC-MSMS (Orbitrap-Velos- 60 min gradient). Raw files were analyzed with MaxQuant (Annex Table 5-15). In the absence of heavy label, MaxQuant assigns the medium to heavy and creates H/L ratios. We analyzed the H/L ratios and defined the proteins with  $\log_2$  H/L (biotin/control) values higher than 1 as enriched.

As expected, urea washed and 4% SDS lysed samples resulted in very clean mass spectra. The best identification rates were obtained from methods 4 and 5, where lysis of tissues (in Ripa and 4% SDS, respectively) was followed by precipitation of proteins and dissolving the pellet in 1% SDS. (Figure 5-40; Annex Table 5-16). These methods increased the identification efficiency for pancreatic samples from ~400 proteins for the method we used so far (method 1 -Figure 5-40), to ~3000 (for methods 4 and 5 - Figure 5-40); we further improved this number to ~4000 by quantifying the samples prepared with method 5, using Orbitrap-Lumos. Enrichment efficiency was also improved from 11% for the least efficient method to ~63% for the methods of choice. The identification and enrichment rates were similarly efficient for samples lysed in Ripa buffer and 4% SDS, both precipitated and washed with 4 M urea (methods 4 and 5, respectively). As Ripa contains TX100 and is known to be sensitive to temperatures higher than 4 degrees (that means in each freeze-thaw cycle a fraction of protein is lost), we continued with 4% SDS as the lysis buffer. We also included a

## Results

precipitation step followed by resolubilization of proteins in 1% SDS before incubation with beads, and washed the beads with 4 M urea in the next enrichment experiments.



**Figure 5-40. Comparison between enrichment methods (method development)**

Dox (2 mg/ml) and biotin (0.22 mg/ml) in water were administered to two Coll1a1-BioID-rfTA (mouse 1) and two control mice for seven days, mice were sacrificed and pancreas isolated followed by on-bead digestion using equal amounts of proteins with 5 different approaches varying in type of lyses buffer, wash buffer, and inclusion or absence of precipitation step after lysis. Biotinylated and control mice organs were labeled with medium and light dimethyl labeling materials and combined, OASIS desalted, fractionated by OFFGEL and analyzed by Orbitrap-Velos MS. Raw files were analyzed with MaxQuant/Perseus. Number of proteins with normalized H/L (biotin/control)  $\log_2$  ratios higher than 1 (enriched; blue) were compared with total number of proteins. The samples processed with most effective method were remeasured on Orbitrap-Lumos and identification/enrichment rates were compared. Experiment's ID number: SA585.

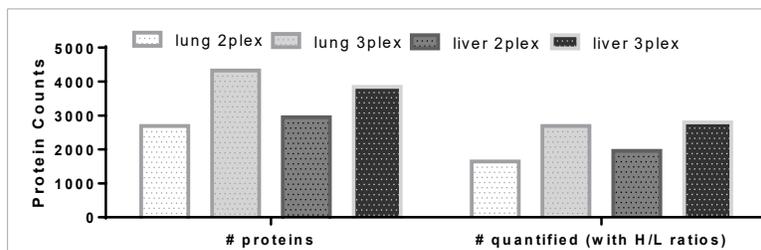
### 5.2.3.4. Tissue proteomics including two control samples

Tissue proteomics with one control sample (section 5.2.3.2) resulted in low-quality data for which the assessment and data analysis was not possible. One obstacle was poor identification, quantification, and recovery rates; to improve the quality of datasets we performed method development analysis which increased these rates significantly. The other problem was the diverse distribution pattern of dimethyl ratios in different tissues and different replicates of each experiment. Therefore, applying an arbitrary cut-off or the cut-off from a null pancreas study on other tissue datasets is prone to assigning wrong proteins as biotinylated or background. Considering all sources of technical and biological variations which resulted in dissimilar ratio distributions for the different replicates of the same tissue sample and to reduce the variation and experimental steps (by excluding a separate null experiment) we considered including two control samples in each experiment. We hypothesized it should be possible to find the cut-off for biotinylation within the study using the two control samples, instead of performing a separate null experiment.

We repeated the tissue proteomics using the best method (4% SDS lysis buffer and urea wash) and included 2 control samples in each experiment. Three Coll1a1-BioID-rfTA (heterozygote for BioID and rfTA) and six

## Results

control mice (lacking one of the genes) were administered biotin (0.22 mg/ml) and Dox (2 mg/ml) in water for seven days and sacrificed. Pancreas, liver and lung were isolated, lysed with 4% SDS and precipitated. 2 mg from each sample resolubilized in final 0.5% SDS were incubated with streptavidin beads, followed by washing of the beads with 4 M urea, reduction, alkylation, and digestion of the proteins (4.3.10.2.). Biotinylated mice tissue samples were labeled with heavy dimethyl labeling materials and control tissue received light and medium labels. Peptide samples were combined, OASIS desalted, fractionated by OFFGEL, and analyzed using LC-MSMS (Orbitrap-Velos- 60 min gradient) (Annex Table 5-17). Due to technical issues with Orbitrap-Velos, pancreas samples were analyzed by Orbitrap-Lumos (60 min gradient). Raw files were analyzed with MaxQuant. We used intensities and generated the ratios in Perseus software. In comparison to the 2plex experiments, identification and quantification rates were improved ~1.5 fold (Figure 5-41). We excluded the pancreas datasets from comparison as 2plex and 3plex raw files were analyzed by different instruments.



**Figure 5-41. Comparison between tissue datasets including one and two control samples (2plex and 3plex) in terms of identification and quantification.**

Dox (2 mg/ml) and biotin (0.22 mg/ml) were administered to Coll1a1-BioID-rtTA (mouse 1) and control mice for seven days, mice were sacrificed, and lung and liver isolated. In 2plex approach on-bead digestion was performed using one control sample and the least efficient protocol (method 1) in method development study.

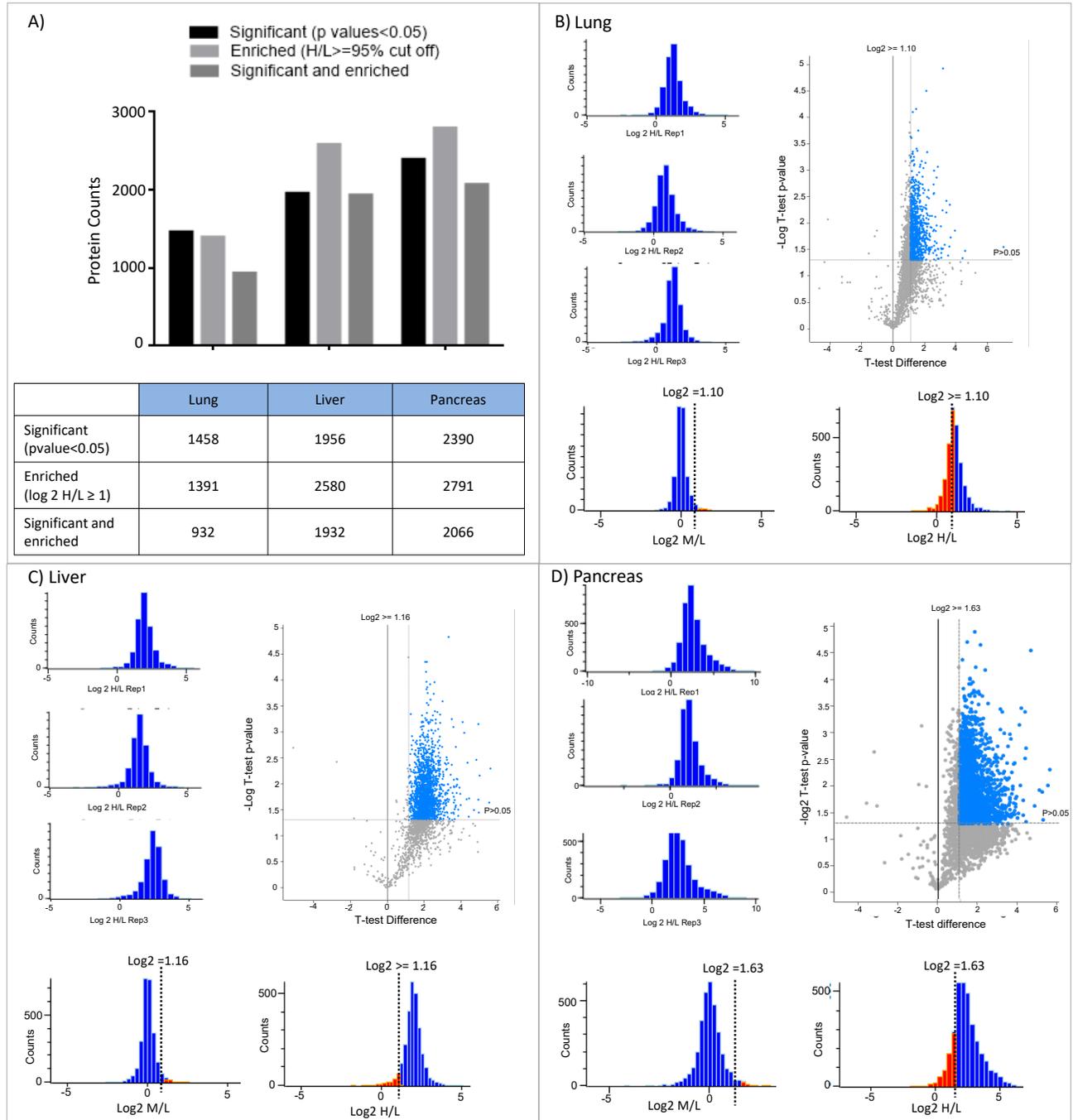
In 3plex approach two control samples were included and the method for enrichment was the best method in method development study (method 5). In both studies biotinylated tissues were labeled with heavy dimethyl labeling materials. The control sample received the light label in 2plex. In 3plex study the two control samples were labeled with light and medium dimethyl reagents. Samples were combined, OASIS desalted, fractionated by OFFGEL and analyzed by Orbitrap-Velos MS. Number of total proteins identified in each study as well as the number of proteins with quantified H/L ratios are compared in 2plex and 3plex approaches in lung and liver datasets. Experiments' ID numbers: SA4110 and SA5115.

Data were mean normalized and 95% cut-off was determined for each M/L (control/control) ratio separately and the mean was calculated to serve as the cut-off for biotinylation; the  $\log_2$  95% cut-offs were 1.10, 1.16, and 1.63 for lung, liver, and pancreas datasets, respectively. The ratio cut-offs for biotinylation were different across the organs. This suggests that non-biotinylated samples originating from different tissues have dissimilar dynamics and bind the beads with variable strength and pattern resulting in a different distribution of ratios and unique cut-off values.

To identify the biotinylated proteins, we generated the H/L (biotin/control) ratios out of the intensities. To allow quantification of all enriched proteins, we replaced the missing values in control channels (L) with a small value using the imputation action in Perseus. After a one-sample t-test, the  $\log_2$  fold change cut-offs determined for each tissue dataset were used to define biotinylated proteome. Proteins with biotin/control ratios higher than the 95% cut-off are defined as enriched and those with  $p$ -values  $< 0.05$  as significant. Proteins which pass both these criteria (enriched and significant) are assigned to biotinylation (for more detail on data analysis see 4.5.4; Annex Table 5-18; Figure 5-42 A)

## Results

Liver, with most of the cells biotinylated (Figure 5-32), showed the highest biotinylation level among the examined tissue samples. 92% of the proteins with at least 3 quantified biotin/control ratios (2,580 out of 2,798) acquired ratio values higher than the cut-off (Figure 5-42 C), 69% significant. In case of the lung with a patchy expression of BioID (Figure 5-32), the efficiency was 46% for the proteins with at least 3 quantified biotin/control ratios (1,391 out of 3,021) (Figure 5-42 D), 31% significant. In pancreas dataset, in total 5036 proteins were identified and 3462 acquired 3 quantified biotin/control ratios. Out of these 2791 (80%) obtained ratio values higher than the cut-off, 59% significant (Figure 5-42 B).



## Results

### Figure 5-42. Enrichment analysis and ratio distribution for tissue samples (including two control samples)

Dox (2 mg/ml) and biotin (0.22 mg/ml) were administered to Coll1a1-BioID-rtTA (mouse 1) and control mice for seven days. Mice were sacrificed and liver, lung, and pancreas were isolated followed by on-bead digestion (4% SDS lysis buffer and urea wash) using equal amounts of proteins. Biotinylated organs were labeled with heavy dimethyl labeling materials. Control sample received the light and medium labels and combined with biotinylated sample, OASIS desalted, fractionated by OFFGEL and analyzed by Orbitrap-Velos MS for lung and liver and Orbitrap-Lumos for pancreas. Raw files were analyzed with MaxQuant/Perseus. Using the M/L values a 95% cut-off for biotinylation was determined in each study. T-test was performed proteins with normalized H/L (biotin/control)  $\log_2$  ratios higher 95% cut-off were called as enriched, those with  $p$ -values  $< 0.05$  were defined as significant and those enriched and significant were assigned as biotinylated. Bar graphs (A) shows these three groups. B, C, D represent the data for lung, liver and pancreas: Histograms show the  $\log_2$  H/L (biotin/control) ratio distribution for separate replicates of each tissue proteomics dataset. Volcano plots show the biotinylated proteins (significant and enriched) in blue. The two histograms at the bottom of each panel represent the distribution  $\log_2$  M/L (control/control) ratios (left) and  $\log_2$  H/L (biotin/control) ratios (right). These show the proportion of proteins with  $\log_2$  H/L ratios above the cut-off for biotinylation. Rep: replicate. Experiment's ID number: SA5115.

The number of significant proteins ( $p$ -values  $< 0.05$ ), enriched ( $\log_2$  H/L  $\geq$  95% cut-off), as well as significant and enriched (Figure 5-42 A) were  $\sim 2$  to 3 fold higher in comparison to the 2plex study (Figure 5-39 A). The pattern of H/L ratios on histogram was also more similar which reflects the better reproducibility of the 3plex vs 2plex using the new enrichment method (Figure 5-42 B and C). We found a strong correlation (above 90%) both between independent control proteomes and between biotinylated samples (Appendix Figure 8). These are due to (1) method development and paying attention to different sources of technical variation; (2) inclusion of a second control channel which increases the chance for non-specific binding proteins to trigger the mass spectrometer and get identified and quantified.

We could not further validate the method, as no information on expected protein populations in these datasets is available; however, since the quality of data and enrichment efficiencies were improved significantly, we continued with this enrichment method to cell-type-specific analysis.

### 5.2.4. Enrichment for proteins specific to beta cells of pancreas

#### 5.2.4.1. Duration for Doxycycline administration and tTA zygosity

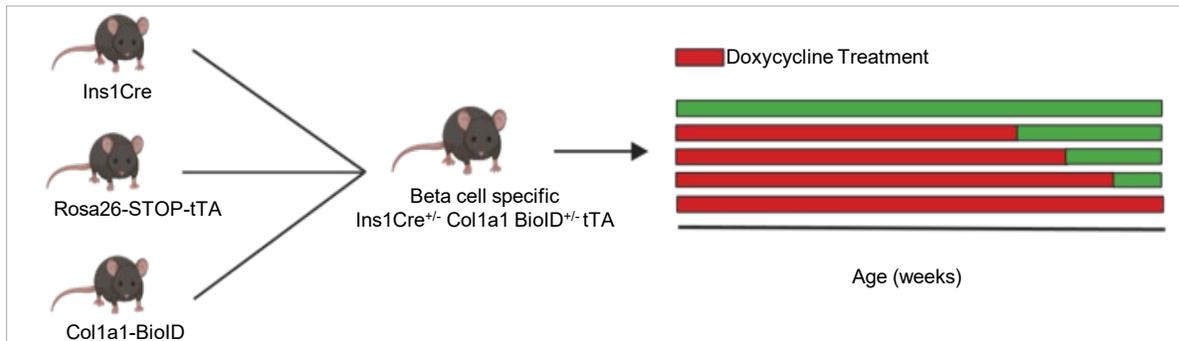
BioID expression in pancreas showed a distinctive islet-specific pattern with beta-cell-specific expression and biotinylation by BioID (Figure 5-34). However, Coll1a1-BioID-rtTA mouse alone without a beta cell specific Cre line could not secure beta-cell specific enrichment due to two reasons: (1) BioID may have an expression in any other islet cell types whose markers were not included in our co-staining immunofluorescence staining experiments (e.g. pancreatic polypeptide (PP), delta or epsilon cells) (2) Even if the expression is purely beta-cell-specific, the inclusion of any tissues connected to the pancreas as well as the blood (with high BioID expression; see Figure 5-31) during isolation can result in the wrong annotation of proteins as beta-cell-specific. To make sure the expression is limited to beta cells of pancreas we created Ins1 Cre-inducible Tet-off BioID mice. The tetracycline-inducible system has the benefit of allowing expression of BioID in a time-controlled manner.

We obtained Ins1Cre mice and bred them with Coll1a1-BioID and Rosa26-STOP-tTA mice to generate Ins1Cre<sup>+/-</sup> Coll1a1 BioID<sup>+/-</sup> tTA offspring. In these mice, expression of BioID is controlled by Tet-off system and initiates upon withdrawal of Dox (For more information on mechanism of action see Figure 5-26 B).

We first aimed to use the temporal control of gene expression by feeding the mice with Dox food for several weeks which ceases the gene expression, followed by induction and initiation of BioID expression by replacing

## Results

the Dox food with normal food. We tried to titrate the Dox withdrawal duration to reach the highest and most reproducible gene expression by comparing 3 different time durations of Dox withdrawal: 3 weeks, 2 weeks, 1 week for mice homozygote and heterozygote for tTA gene (Figure 5-43). Mice received 625 mg/kg Dox in food before it was withdrawn.



**Figure 5-43. Schematic for generation of beta cell specific mice and Dox titration experiment.**

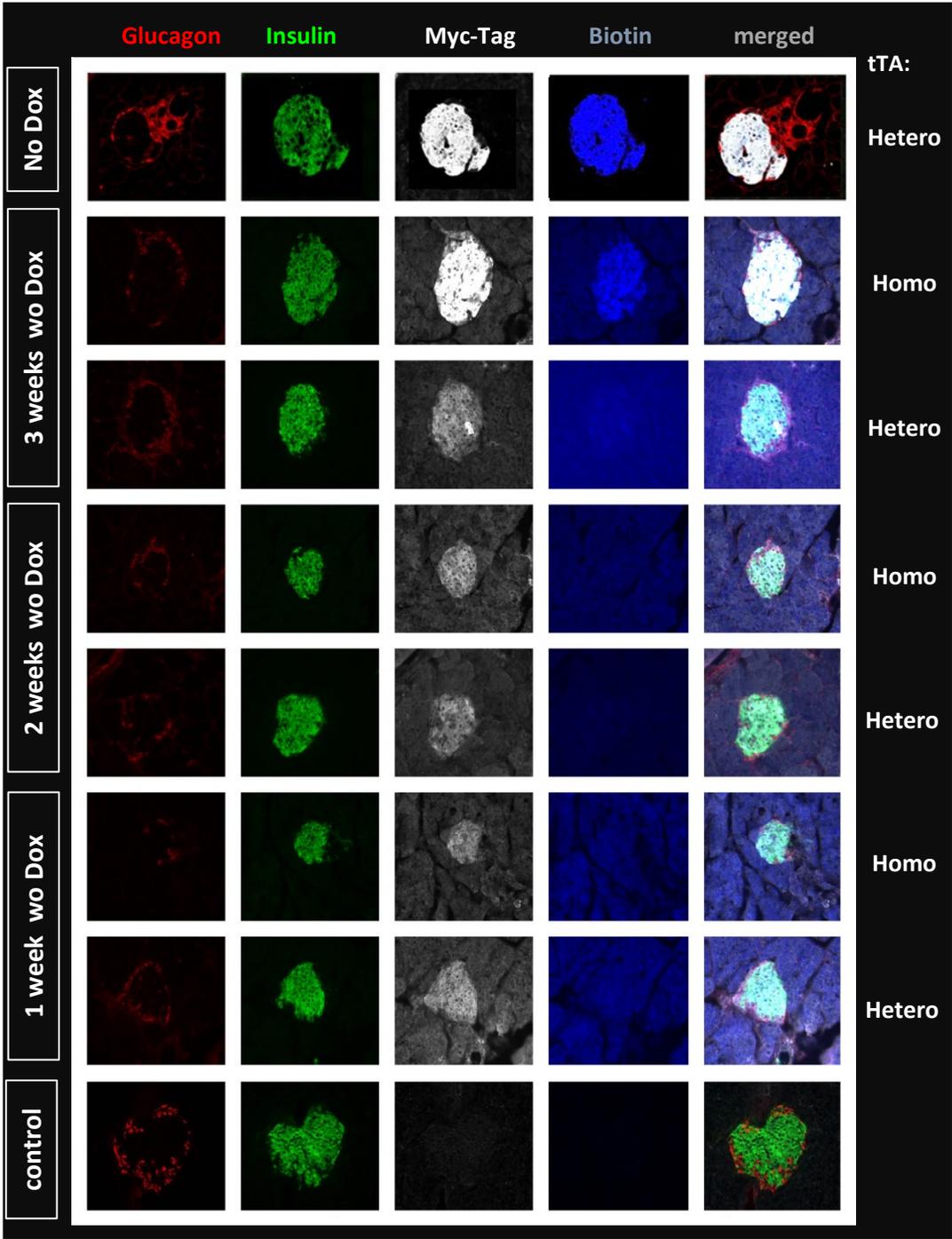
Ins1 Cre, Rosa26-STOP-TTA, and Col1a1-BioID mice were crossed to generate Ins1Cre Col1a1 BioID tTA offspring. Mice were compared for duration of Dox withdrawal resulting in highest expression of BioID in 3 groups (each containing 4 animals): 3, 2, and 1 weeks without Dox. An Ins1Cre<sup>-/-</sup> Col1a1 BioID<sup>+/-</sup> tTA<sup>+/-</sup> mouse was used as positive control which received no Dox. The negative control mouse lacked one of the genes and received no Dox. All mice received biotin (0.22 mg/ml) in water for 1 week before sacrificed.

An Ins1Cre<sup>+/-</sup> Col1a1 BioID<sup>+/-</sup> tTA<sup>+/-</sup> mouse received no Dox for 10 weeks and was used as a positive control. The negative control mouse (lacking one of the genes) received no Dox. All mice received biotin (0.22 mg/ml) in water for one week and were sacrificed. The pancreas tissues were isolated and fixed for immunofluorescence analysis. We compared expression and biotinylation by BioID using anti-myc and streptavidin, respectively (Figure 5-44). Insulin and glucagon, markers of beta and alpha cells, were used to ensure cell-type-specific expression and biotinylation by BioID.

In all mice, expression and biotinylation were only observed in the insulin producing beta cells of the pancreas and not in alpha cells. The mice with no Dox treatment for three (tTA<sup>+/-</sup>), two (tTA<sup>-/+</sup> and tTA<sup>+/+</sup>), and one weeks (tTA<sup>-/+</sup> and tTA<sup>+/+</sup>) showed similar poor BioID expression, whereas the mouse homozygote for tTA with no Dox for 3 weeks showed higher expression of BioID. Biotinylation was only observed in mice with no Dox for three weeks and was stronger in the mouse homozygote for tTA. The highest BioID expression and biotinylation was however recorded from the mouse with no Dox treatment and heterozygote for the tTA gene. As the mouse with no Dox treatment showed the least background in the streptavidin channel and highest BioID expression compared to the other mice, we finally performed another microscopy experiment to compare two groups (4 animals/group) of tTA homozygote mice: (1) with no Dox treatment (2) 3 weeks without Dox. BioID expression (the signal in myc channel) normalized on insulin expression (the signal in insulin channel) in 4 islets of each mouse was used as the criteria to evaluate BioID expression in these two groups. BioID expression was significantly higher in mice not treated with Dox (intensity in myc channel /intensity in insulin channel: 1.92 for no Dox and 0.69 for 3 weeks without Dox; *p*-value: 0.015; Figure 5-45).

Although we intended to use the advantage of temporal control of gene expression, as it did not result in maximum expression, we decided to eliminate it from our setting. To ensure highest expression and maximum biotinylation we decided to continue with tTA homozygous mice and no Dox treatment for enrichment analysis.

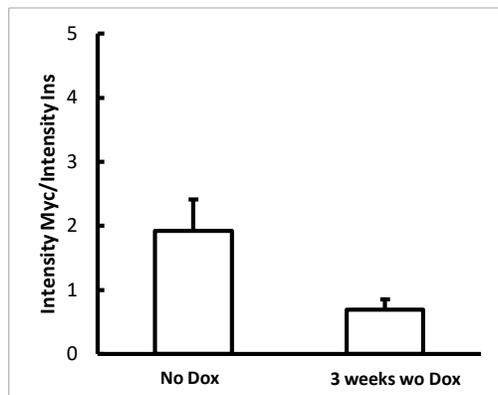
Results



## Results

### Figure 5-44. Dox duration titration experiment.

Ins1Cre<sup>+/-</sup> Colla1 BioID<sup>+/-</sup> tTA mice (2 heterozygotes and 2 homozygotes for tTA in each group), were compared for optimum duration of Dox withdrawal resulting in highest expression of BioID in 3 groups (each containing 4 animals): 3, 2, and 1 weeks without Dox. An Ins1Cre<sup>+/-</sup> Colla1 BioID<sup>+/-</sup> tTA<sup>+/-</sup> mouse was used as positive control which received no Dox. The negative control mouse lacked one of the genes and received no Dox. All mice received biotin (0.22 mg/ml) in water for 1 week before sacrificed. The pancreas samples were fixed overnight in 4% PFA and cryosectioned. For immunofluorescence analysis the following primary antibodies were used: goat anti-myc antibody (ab9106, Abcam: 1: 400), rabbit anti-insulin antibody [K36aC10] (ab6995, 1: 400), guinea pig anti-insulin antibody (IR002, DAKO), mouse anti-glucagon antibody [K79bB10] (ab10988, 1: 400). Alexa Fluor 421 (405225, Biolegend, 1: 400) streptavidin was used to show biotinylation. The exposure time has been constant for each channel. Staining and microscopy were performed by Dr. Kenichi Kimura and Martin Breitbach from the Institute of Physiology, University of Bonn.



### Figure 5-45. Dox duration titration experiment: no Dox vs 3 weeks without Dox

Ins1Cre<sup>+/-</sup> Colla1 BioID<sup>+/-</sup> tTA<sup>+/+</sup> mice were compared for two durations of Dox withdrawal resulting in highest expression of BioID (each containing 4 animals): No Dox and 3 weeks without Dox. All mice received biotin (0.22 mg/ml) for 1 week before sacrificed and analyzed by microscopy. Ratio intensity of myc (showing BioID expression over insulin (marker for beta cells) was used to compare the expression between the two groups after microscopy experiment. Intensity in myc channel /intensity in insulin channel: for no Dox: 1.92 and for 3 weeks without Dox: 0.69; t-test *p*-value: 0.015; n=16. wo: without; Ins: insulin. Analyses were performed by Dr. Kenichi Kimura from the Institute of Physiology, University of Bonn.

#### 5.2.4.2. Identification of beta cell proteome

In proteomics experiments on mice expressing BioID in beta cells of pancreas, we switched from dimethyl labeling and OFFGEL fractionation, and Orbitrap-Velos mass spectrometer to TMT labeling and High-pH reversed-phase peptide fractionation (Hp-RP) and Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Orbitrap-Lumos). For each of these decisions, we had a reason. TMT reagents make it possible to multiplex the analysis, decrease the instrument time and include all sample types in one experiment minimizing the technical variation. Unlike Orbitrap-Velos, Orbitrap-Lumos MS system is capable of synchronous precursor selection (SPS) multi-notch MS3 quantitative analyses which improve experimental throughput and eliminate the issues TMT ratio distortions caused by interfering ions and therefore provides better quantitative performance across more sample sets (Ting et al. 2011). Therefore, for analysis TMT labeled samples, Orbitrap-Lumos is the ideal instrument.

It is reported that in comparison to OFFGEL, Hp-RP fractionation results in more peptides and proteins. It also leads to higher resolution (the ability to limit a specific peptide to one particular fraction) and more uniform

## Results

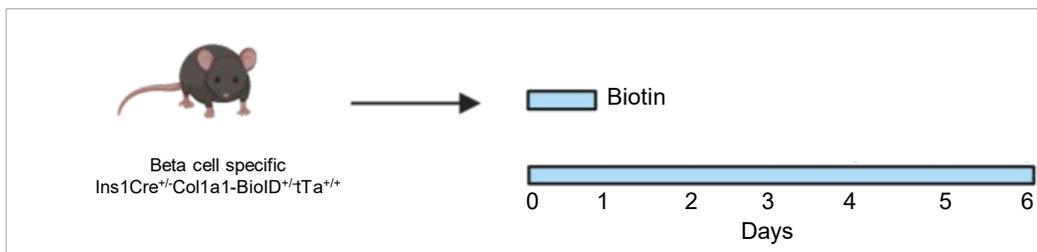
distribution of total and unique peptides for Hp-RP across all the collected fractions which suggest it as a better method for fractionation (D. R. Stein et al. 2013). Comparison experiments were performed to make certain of similar identification, quantification, and enrichment rates are obtained using these three techniques.

We performed the first cell-type-specific proteomics study using mice expressing BioID in beta cells of the pancreas and evaluated the potency of our method to enrich for biotinylated proteins in these cells accounting for only 1-2% of pancreatic cells. Since  $Ins1Cre^{+/-} Col1a1\ BioID^{+/-} tTA^{+/+}$  mice express BioID through their entire life span, in an individual experiment on the same mice tissues, we evaluated the expression of proteins at whole proteome level (using non-enriched samples) and compared it to protein expressions in control mice to ensure that BioID expression and biotinylation does not adversely affect the protein expression at the global level.

For both experiments the same tissue samples were used: Three  $Ins1Cre^{+/-} col1a1BioID^{+/-}tTA^{+/+}$  mice as well as three control mice (lacking one of the genes) were treated with biotin (0.22 mg/ml) in water for 7 days, sacrificed and pancreas was isolated (Figure 5-46). These mice were processed with two different protocols: once set of samples were enriched (on-bead digested) and once set were not enriched (whole proteome).

To enrich for beta cell specific proteins, we used 3 mg of each sample and purified the biotinylated proteins using our optimized on-bead digestion protocol. Pancreas samples were lysed with 4% SDS and precipitated. 3 mg from each sample resolubilized in final 0.5% SDS were incubated with streptavidin beads, followed by washing of the beads with 4 M urea, reduction, alkylation, and digestion of the proteins (section 5.2.3.3). In order to see if long term BioID expression results in any alterations in global protein expression in  $Ins1Cre^{+/-} col1a1BioID^{+/-}tTA^{+/+}$  mice, we digested the whole proteome of pancreas samples (40  $\mu$ g of each sample) using RapiGest in-solution digestion method.

After digestion, the whole proteome and enriched samples were labeled with TMT 6plex labels (control samples: 126, 127, 128 TMT tags, biotinylated samples: 129, 130, 131 TMT tags), mixed and desalted using OASIS columns. The whole proteome and biotin enriched samples were fractionated using Hp-RP into 24 and 12 fractions, respectively (performed by Dr. Robert Hardt) and analyzed using Orbitrap-Lumos. Raw files from each of the two experiments were pooled and searched by Mascot against UniProt/Trembl mouse database and the protein result file was analyzed using Perseus.



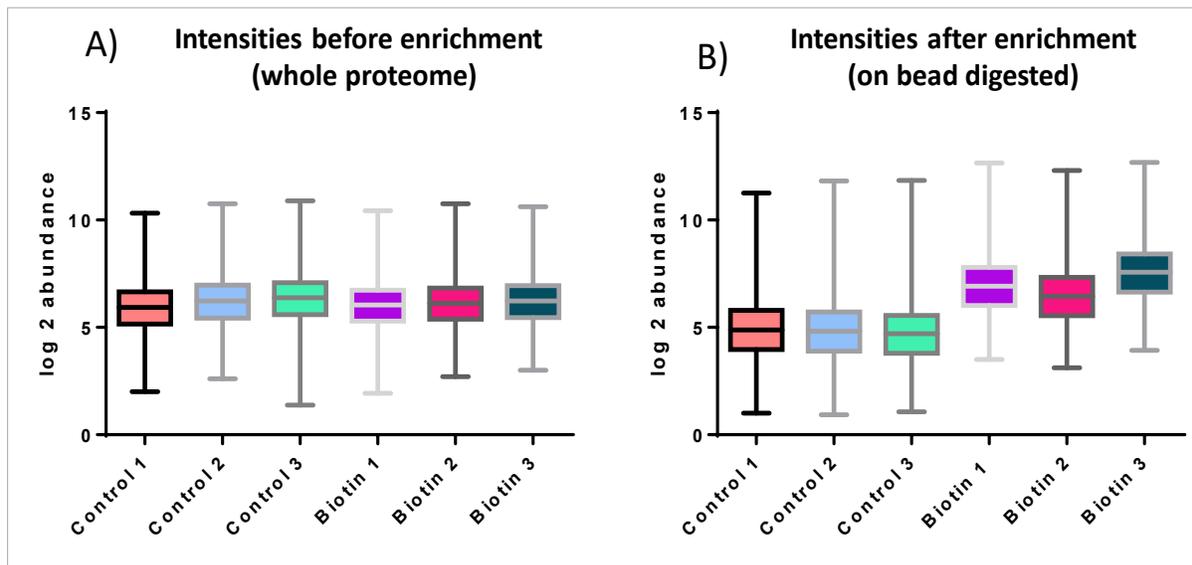
**Figure 5-46. Workflow for treatment of mice for beta cell proteomics**

3  $Ins1Cre^{+/-} Col1a1\ BioID^{+/-}tTA^{+/+}$  mice as well as 3 control mice (lacking one of the genes) were given biotin (0.22 mg/ml) in water for 7 days. At the end the animals were sacrificed and pancreas was isolated and further processed with enrichment and whole proteome analysis.

In the whole proteome analysis, we identified 51,543 peptide groups and 6,483 protein groups (20,675 proteins) (Annex Table 5-19). Enrichment of biotinylated proteins resulted in the identification of 27,275 peptide groups and 4,133 protein groups (13,148 proteins) (Annex Table 5-21.). Protein files contain isomers and proteins with sequence homology which are then categorized into protein groups.

## Results

We first monitored the intensities of biotinylated and control samples in the enriched and whole proteome datasets. For both experiments, intensities of proteins in different TMT channels were transformed into  $\log_2$  values and the average intensity (among the fractions) was calculated for each protein in each TMT channel. In the whole proteome study, control and biotinylated pancreas samples showed similar global protein intensities which means BioID expression does not lead to overall changes in protein expression in *Ins1Cre<sup>+/-</sup> Coll1a1 BioID<sup>+/-</sup> tTA<sup>+/+</sup>* mice (Figure 5-47 A). However, after enrichment for biotinylated proteins, TMT channels for biotinylated samples acquired higher intensities than the control channels; this is due to higher abundance for the proteins which bind the beads specifically and get enriched and therefore their intensities increase (Figure 5-47 B).



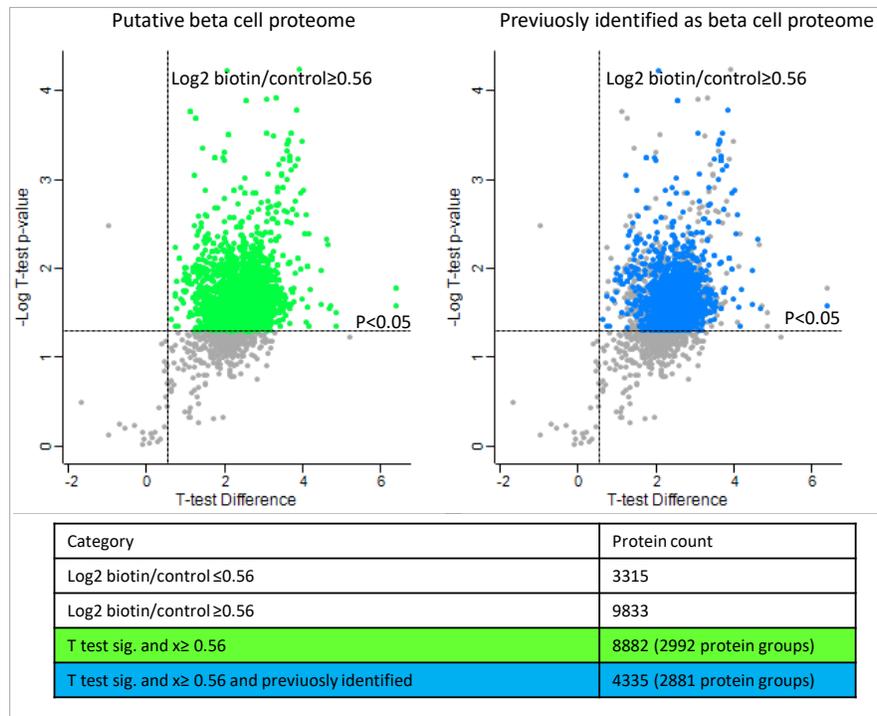
**Figure 5-47. comparison between intensities of TMT channels in whole proteome and beta cell enriched proteome**

3 *Ins1Cre<sup>+/-</sup> Coll1a1 BioID<sup>+/-</sup> tTA<sup>+/+</sup>* mice as well as 3 control mice (lacking one of the genes) were treated with biotin (0.22 mg/ml) in water for 7 days and sacrificed. Pancreas samples were lysed and processed with two approaches: whole proteome (in-solution digest/ 40  $\mu$ g input) and enrichment (on-bead digestion/ 3 mg input). Biotinylated and control samples were labeled with different TMT labels, mixed, and fractionated using Hp-RP into 24 and 12 fractions for whole proteome and enriched samples, respectively (performed by Dr. Robert Hardt). Samples were analyzed on Orbitrap-Lumos. Raw files were analyzed by Mascot/Perseus. Box plots show the intensities ( $\log_2$  abundance) of proteins in each biotin or control TMT channel in whole proteome and enriched datasets. Experiment's ID number: SA5156.

After evaluating the intensities at a global level in the whole proteome dataset, we also investigated individual proteins for possible regulation as an effect of BioID expression. For this, the grouped abundances of different TMT channels (3 biotins and 3 control) were transformed into  $\log_2$  values and  $\log_2$  biotin/control ratios were created. To find the regulated proteins a one-sample t-test was performed and proteins with  $\log_2$  biotin/control ratios  $\geq 1$  and  $\leq -1$  (shown in red in Figure 5-48; Annex Table 5-20) were labeled as regulated/perturbed. We detected 5 proteins downregulated (Heterogeneous nuclear ribonucleoprotein A; Cellular nucleic acid-binding protein; Selenoprotein; Golgin subfamily A member 4; Ubiquitin carboxyl-terminal hydrolase 10) and two upregulated (Zinc finger and BTB domain-containing protein and Isoform 2 of Zinc finger and BTB domain-containing protein). We could not link these proteins to BioID expression. As only very few proteins are



## Results



**Figure 5-49. Enrichment analysis on beta cell proteome and comparison to literature**

3 *Ins1Cre<sup>+/+</sup> Coll1a1 BioID<sup>+/+</sup> tTA<sup>+/+</sup>* mice as well as 3 control mice (lacking one of the genes) were administered biotin (0.22 mg/ml) in water for 7 days and sacrificed. Pancreas samples were lysed and processed with on-bead digestion using the same amount (3 mg) of proteins. Biotinylated and control samples were labeled with different TMT labels, mixed, and fractionated using Hp-RP into 12 fractions (performed by Dr. Robert Hardt) and analyzed on an Orbitrap-Lumos. Raw files were analyzed by Mascot/Perseus. Control/control ratios were used to find a 95% cut-off. A one-sample t-test was performed and proteins with normalized biotin/control log<sub>2</sub> ratios higher 95% cut-off (enriched) and *p*-values < 0.05 (significant) were visualized in volcano plot (in green) and defined as biotinylated (putative beta cell proteome). Biotinylated proteins were compared to proteins/genes/transcripts assigned to beta cells and islets in literature. Proteins previously identified in the other studies are shown in blue. Experiment's ID number: SA5156.

To additionally validate the cell-type-specificity of the beta-cell proteome the list of biotinylated beta-cell-specific proteins we focused on cell-type-specific markers of other islet cell types defined by previous studies. Using the IDs of the enriched proteome, GO enrichment analyses were performed to find the enriched and depleted biological processes. 30 biological processes related to insulin and beta cells were found enriched including insulin receptor recycling (FDR P: 5E<sup>-3</sup>) and regulation of insulin secretion (FDR P: 2.8 E<sup>-2</sup>). Our dataset was significantly de-enriched for an alpha, acinar, duct and delta cells (FDR P: 1) of the pancreas and enriched for beta cells (FDR P < 0.05). We did not recognize any of the alpha cell markers (glucagon, MAFB, NKX2.2, ARX, and PAX6) in our list of the enriched proteome, however, glucagon was found in whole proteome dataset. Glucagon secretion, regulation of glucagon secretion, and response to glucagon processes as well as acinar cell differentiation and regulation of acinar cell proliferation were all de-enriched (FDR P: 1). Our enriched dataset was depleted for processes involved in duct cell development and its known markers SOX9 and FOXA2 were missing in the list of enriched proteins. However, FOXA2 was detected in whole proteome dataset. Somatostatin the marker of delta cells was also missing in our list of the enriched proteome but was detected in whole proteome dataset. The absence of other cell-type-specific markers can be either due

## Results

to their low intensities and inability to trigger the MS for the detection or their absence in the enriched proteome sample.

Based on the results of whole proteome study, we could prove that BioID expression does not lead to changes in protein expression at global levels. We could successfully implement our enrichment method for the identification of beta cell proteins and recover the biotinylated proteins in such a scarce cell population. This is partly due to extensive method development which has led to the removal of many of the non-specific binding proteins and achieving high enrichment efficiencies. Also our data matched very well with published literature and was found depleted for markers of other islet's cell types which further confirms the specificity of biotinylation and capacities of our enrichment/data analysis approaches in capturing/assigning the truly biotinylated proteins.

### 5.2.5. Effect of Harmine and STZ on beta-cell Proteome

After successful implementation of our strategy for the enrichment of beta cell specific proteins, we assessed beta cell proteome's alterations as a result of administration two drugs: Streptozotocin (STZ) and Harmine. Harmine is a natural indole alkaloid and is mainly used as an anti-cancer and anti-tumour therapeutic agent (Li et al. 2015). Harmine has also been shown to trigger the proliferation of beta cells in both murine and human beta cells and to increase the mass of functional beta cells (Belgardt and Lammert 2016). Streptozotocin, (STZ) on the other hand, is a diabetes-inducing agent. It is an alkylating substance that targets beta cells and damages them resulting in hypoinsulinemia and hyperglycemia. It is used to model Type 1 diabetes phenotypes in mice (Graham et al. 2011). To investigate the alterations of the beta-cell proteome as a result of Harmine and STZ administration in two individual experiments we treated a group of mice with Harmine+biotin or STZ+biotin and compared them with biotinylated and control groups. For these experiments, only male animals were used because it has been shown that female animals have a higher tolerance to beta cell proteome alterations against external stimuli (Gannon et al. 2018).

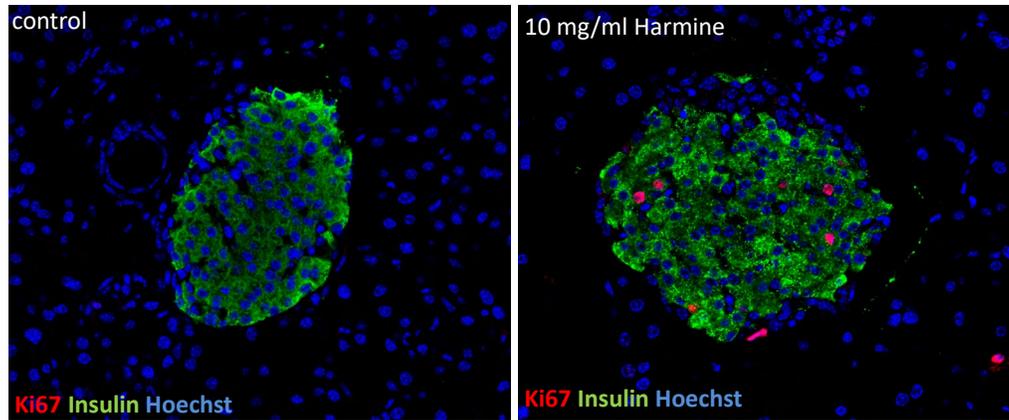
#### 5.2.5.1. Harmine treatment

We first performed a Harmine dose test to find the best dose to stimulate proliferation of beta cells. We tested four different doses of Harmine: 5, 10, 15, and 20 mg/kg on 4 wild type (C57Bl6) animals per dose and stained the islets with ki67 as the marker for proliferation, Hoechst for nucleus, and insulin for beta cells. The proportion of Ki67 positive cells/ insulin positive cells was calculated for 10 slides per mouse and the mean is listed in Table 6.2. Proliferation was seen in very few cells but the effect was greater in the group treated with 10 mg/kg than the other groups. 1.8% of beta cells showed proliferation signals as a result of Harmine treatment and this ratio was 0.6% in the control group. The highest percentage (Ki67 positive/Insulin positive cells) Wang reported with 10 mg/kg Harmine was 1.2%, while in the control group this ratio has been 0.2%. Harmine is known to cause an only modest increase in proliferation in ~1–1.5% of the beta cells engaged in the cell cycle (P. Wang et al. 2015). As 10 mg/kg resulted in the highest proliferation and our results matched the published data, we chose this dose of harmine for the main experiments. An example of ki67 staining for 10 mg/kg dose is shown in Figure 5-50.

Table 5-22. Harmine doses and percent of cells with proliferation signal.

Dose	5 mg/kg	10 mg/kg	15 mg/kg	20 mg/kg	Control
% Ki67+/Ins1+ cells	0.8	1.8	0.5	0.8	0.6

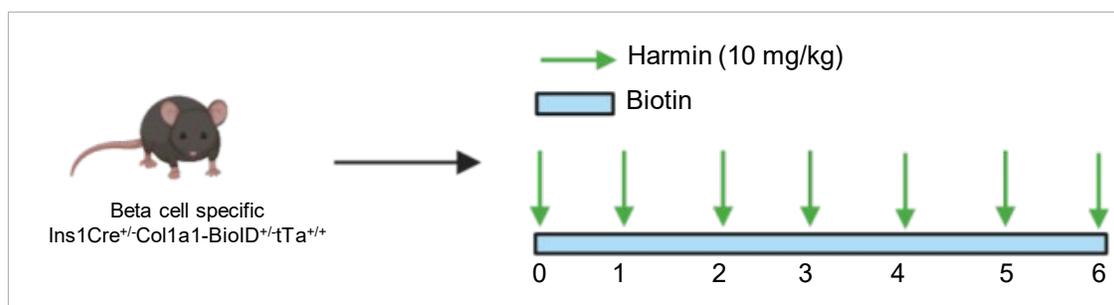
## Results



**Figure 5-50. Harmine dose test evaluation using Ki67 staining.**

Four different doses of Harmine: 5, 10, 15, and 20 mg/kg on 4 wild type animals per dose were compared for proliferation signals induced by Harmine using ki67 staining. 4 groups of wild type 10 week mice received biotin (0.22 mg/ml) for 1 week and were sacrificed and pancreas samples were fixed overnight in 4% PFA and cryosectioned. For immunofluorescence analysis the following primary antibodies were used: guinea pig anti-insulin antibody (IR002, DAKO), rabbit anti-Ki67 antibody [SP6] (RM9106, ThermoFisher scientific, 1: 200). Nuclei were stained with Hoechst 33342 (Sigma-Aldrich, Munich, Germany). This figure is an example showing the Ki67 staining. Staining and microscopy was performed by Dr. Kenichi Kimura from the Institute of Physiology, University of Bonn.

We used 3  $Ins1Cre^{+/-} coll1a1BioID^{+/-} tTA^{+/+}$  mice for drug treatment and injected them intraperitoneally with 10 mg/kg Harmine in normal saline solution for 7 successive days. These mice received biotin (0.22 mg/ml) in water during the 7 days of Harmine administration. (Figure 5-51) Another group of 3  $Ins1Cre^{+/-} coll1a1BioID^{+/-} tTA^{+/+}$  mice and 3 control mice (lacking one of the genes) were administered biotin (0.22 mg/ml) in water and received injections with saline solution for seven days and sacrificed. They were used to identify the drug-effect cut-off and the 95% cut-off for biotinylation, respectively All mice were 10-week-old at the beginning of the treatment.



**Figure 5-51. Workflow for treatment of mice with Harmine**

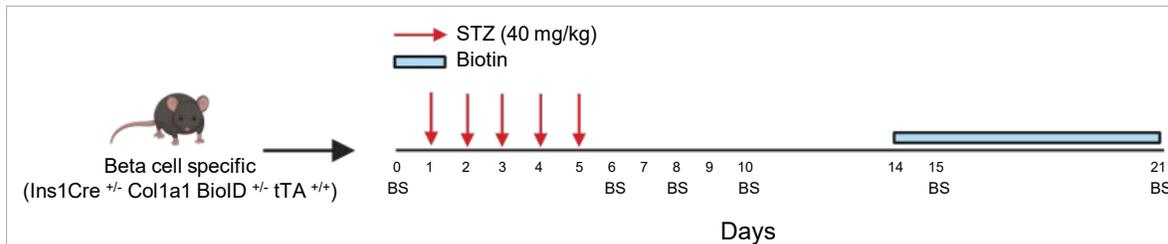
3  $Ins1Cre^{+/-} Coll1a1BioID^{+/-} tTA^{+/+}$  mice were treated with biotin in water, injected with 10 mg/kg/day Harmine in saline solution on 7 consecutive days and sacrificed. Another group of 3  $Ins1Cre^{+/-} coll1a1BioID^{+/-} tTA^{+/+}$  mice and 3 control mice (lacking one of the genes) received biotin and injected with saline solution for seven days and sacrificed (not shown in the picture). Pancreas was isolated and further processed with on-bead digestion.

### 5.2.5.2. STZ treatment and Diabetes Evaluation

For biotin labeling and analysis of the beta cell proteome under cell stress, 3 male  $Ins1Cre^{+/-} coll1a1BioID^{+/-} tTA^{+/+}$  10-week-old animals were injected with Streptozotocin (STZ) on 5 consecutive days. For this purpose,

## Results

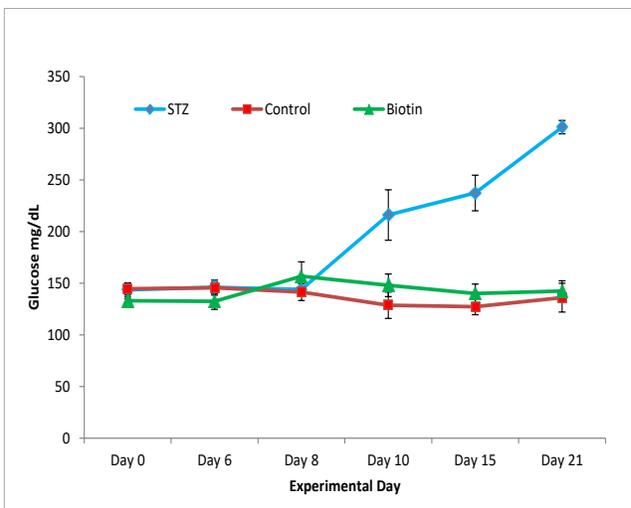
we first weighed the animals and 40 mg/kg/day of STZ was injected in 50 mM sodium citrate solution. From day 14 of the test, the animals were administered 0.22 mg/ml of biotin in the drinking water to initiate biotinylation of beta-cell proteins. Furthermore, 3 *Ins1Cre<sup>+/-</sup> Col1a1BioID<sup>+/-</sup>tTA<sup>+/+</sup>* mice, as well as 3 male control mice (lacking one of the genes) all 10-week-old, were treated similarly except for STZ injection which was replaced with sodium citrate solution; they were used to identify the drug-effect cut-off and the 95% cut-off for biotinylation, respectively. From the day before the first injection, as well as 1, 3, 5, 10 and 16 days after the last STZ injection blood sugar was measured, for which a small amount of blood (< 10  $\mu$ l) was taken. (Figure 5-52)



**Figure 5-52. Workflow for treatment of mice with STZ**

3 *Ins1Cre<sup>+/-</sup> col1a1BioID<sup>+/-</sup>tTA<sup>+/+</sup>* 10-week-old animals were injected with STZ (40 mg/kg/day) on 5 consecutive days. From day 14 of the test, the animals were given biotin (0.22 mg/ml) in water for 7 days. 3 *Ins1Cre<sup>+/-</sup> col1a1BioID<sup>+/-</sup>tTA<sup>+/-</sup>* mice as well as 3 control mice (lacking one of the genes) all 10-week-old were treated similarly except for STZ injection which was replaced with sodium citrate solution. At the end of the 21-day experiment, the animals were sacrificed and pancreas was isolated and further processed with on-bead digestion. The day before the first injection, as well as 1, 3, 5, 10 and 16 days after the last STZ injection blood sugar was measured. STZ: Streptozotocin; BS: blood sugar.

The increase in blood glucose concentration in STZ group over time confirmed the successful induction of hyperglycaemia (~ 300 mg/dL; Figure 5-53; (Furman 2015)).



**Figure 5-53. Blood glucose concentration in STZ, Biotin, and control groups.**

3 *Ins1Cre<sup>+/-</sup> col1a1BioID<sup>+/-</sup>tTA<sup>+/+</sup>* 10-week-old animals (STZ: blue) were injected with STZ (40 mg/kg/day) on 5 consecutive days. From day 14 of the test, the animals were given biotin (0.22 mg/ml) in water for 7 days. 3 *Ins1Cre<sup>+/-</sup> col1a1BioID<sup>+/-</sup>tTA<sup>+/-</sup>* mice (Biotin: green) as well as 3 control mice (lacking one of the genes; Control: red) all 10-week-old were treated similarly except for STZ injection which was replaced with sodium citrate solution. The day before the first injection, as well as 1, 3, 5, 10 and 16 days after the last STZ injection blood sugar (BS) was measured and plotted.

### 5.2.5.3. Harmine and STZ on-bead digestion and data analysis

Pancreas samples were lysed with 4% SDS and precipitated. 3 mg from each sample resuspended in final 0.5% SDS were incubated with streptavidin beads, followed by washing of the beads with 4 M urea, reduction, alkylation, and digestion of the proteins. Similar to beta cell enrichment analysis we used TMT for labeling the

## Results

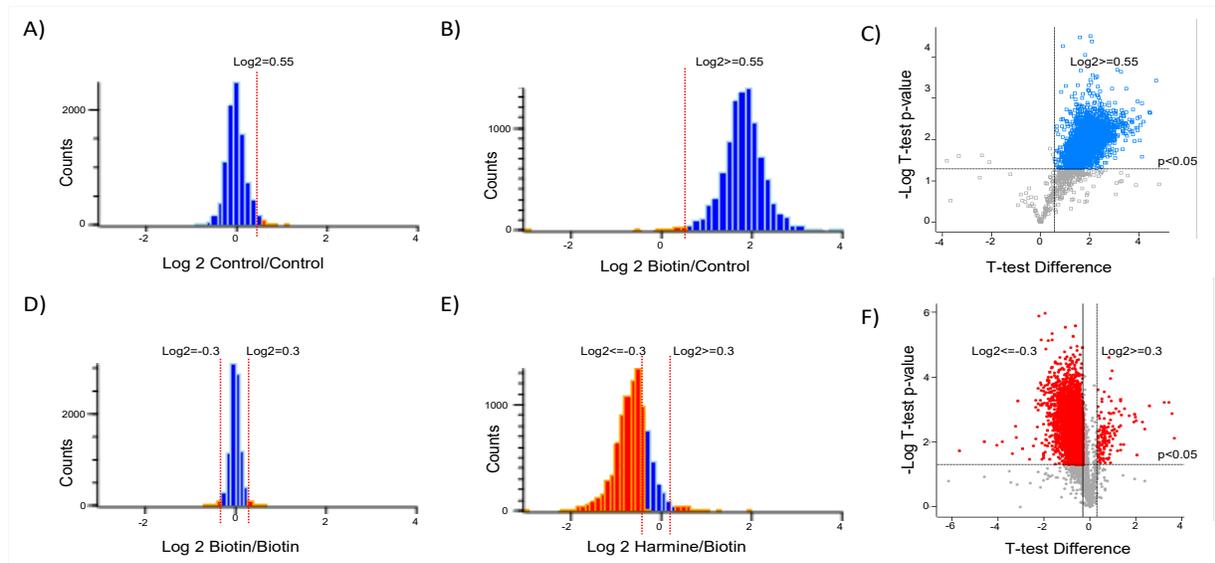
samples in these two experiments. 9plex TMT labeling allows multiplexing three biological replicates for each sample type (control/biotinylated/biotinylated and drug injected) and determination of biotinylation and drug effect cut-offs as part of a single experiment.

Drug+biotin (Harmine -H or STZ -S) treated mice received 131, 130C, and 130N TMT tags, while biotinylated tissue samples (Biotin -B) received 129C, 129N, and 128C, and control samples (Control -C) were labeled with 128N, 127N, and 127C labels. Assuming that TMT ratios form a normal distribution, we defined two parameters for the analysis of data: (1) biotinylation cut-off and (2) drug-effect cut-off. Datasets were analyzed in 2 rounds. First, the 95% cut-off for biotinylation was calculated from control/control ratios and applied on biotin/control ratios to identify the biotinylated proteins in each dataset. To detect the variation among the biotinylated samples we defined the 95% ratio drug-effect cut-off based on the distribution of B/B ratios. This cut-off reflects the alterations of protein intensities in biotinylated samples as a result of biological and technical variations. We use this cut-off to evaluate the effect of the drug; the alterations in protein expression can be assigned to the effect of drug if (Drug+biotin/biotin) is above or below this cut-off.

The grouped protein abundances (average of all 12 fractions) of all TMT channels were loaded into Perseus and  $\log_2$  transformed. To identify the biotinylated proteins (the beta-cell proteome), C and B intensities were used. Proteins with less than 3 values in biotin channels were filtered out and missing values in light channels were replaced with small values using imputation function. Control/control and biotin/control ratios were generated (C2/C1, C3/C2, C3/C1 and B2/C1, B3/C2, B3/C1) and normalized using the mean of each C/C ratio list. 95% cut-off was calculated for each  $\log_2$  C/C ratio; the average was 0.55 for Harmine dataset (Figure 5-54 A and B and C) and 0.44 for STZ dataset (Figure 5-55 A and B and C) that were used as the  $\log_2$  95% ratio cut-off for biotinylation.

To identify the proteins with altered intensities as an effect of the drug, we used the list of biotinylated proteins identified in previous step and combined the biotin (B) and Harmine or STZ (H or S) intensities to create following ratios: B2/B1, B3/B2, B3/B1 and “H2 or S2”/B1, “H3 or S3”/B2, “H3 or S3”/B1. Ratios were normalized using the mean of B/B ratio lists. 95% cut-off was determined for each  $\log_2$  B/B and averaged to find the  $\pm 95\%$   $\log_2$  ratio cut-off for drug effect. As biotin/biotin ratios formed a normal distribution and the same intensities have been used as denominator and numerator to generate the TMT ratios we used the upper tail’s 95% cut-off on both tails. This cut-off was 0.3 for Harmine (Figure 5-54 D and E and F) dataset and 0.31 for STZ dataset (Figure 5-55 D and E and F). A one sample T-test was performed for S/B or H/B ratios and rows were filtered to include the proteins with ratios above and below the drug-effect cut-off that were considered as altered by the drug effect (Figure 5-54 F and 5-55 F) (For more detailed information on data analysis see 4.5.4.).

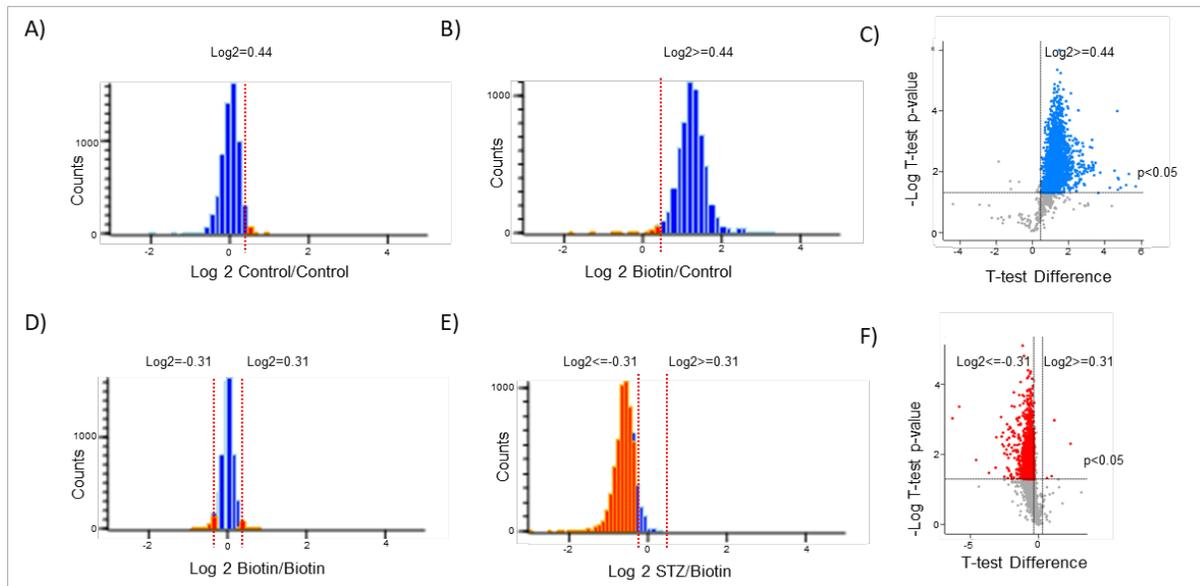
## Results



**Figure 5-54. Data analysis for identification of Harmine-perturbed proteome.**

3 *Ins1Cre<sup>+/-</sup> colla1BioID<sup>+/-</sup>tTA<sup>+/+</sup>* mice were treated with biotin (0.22 mg/ml) in water, injected with 10mg/Kg/day Harmine in saline solution for 7 days and sacrificed. Another group of 3 *Ins1Cre<sup>+/-</sup> colla1BioID<sup>+/-</sup>tTA<sup>+/+</sup>* mice and 3 control mice (lacking one of the genes) received biotin (0.22 mg/ml) and got injected with saline solution for seven days and sacrificed. Pancreas was isolated and further processed with on-bead digestion with similar protein amounts (3 mg). The 3 control, 3 biotinylated and 3 Harmine injected biotinylated animals were labeled with 9 TMT labels and mixed. Samples were fractionated using Hp-RP into 12 fractions (performed by Dr. Robert Hardt) and analyzed using an Orbitrap-Lumos. Raw files were analyzed by Mascot/Perseus. Control/control ratios were used to find A) 95% cut-off for biotinylation B) 95% cut-off for biotinylation was applied on proteins with 3 quantified biotin/control ratios and C) proteins with log<sub>2</sub> biotin/control ratios above the cut-off and *p*-values < 0.05 (as shown in volcano plot in blue) were defined as biotinylated. D) Biotin/biotin ratios were used to find a cut-off for drug effect. E) The drug-effect cut-off was applied on biotinylated proteins with 3 quantified Harmine/biotin ratios (in red). And F) proteins with ratios above and below the drug effect cut-off and with *p*-values < 0.05 were defined as up and downregulated as an effect of Harmine. Experiment's ID number: SA5150.

## Results



**Figure 5-55. Data analysis for identification of STZ-perturbed proteome.**

3 *Ins1Cre<sup>+/-</sup> colla1BioID<sup>+/-</sup>tTA<sup>+/+</sup>* 10-week-old animals were injected with STZ (40 mg/Kg/day) on 5 consecutive days. From day 14 of the experiment, the animals were given biotin (0.22 mg/ml) in the drinking water for 7 days. 3 *Ins1Cre<sup>+/-</sup> colla1BioID<sup>+/-</sup>tTA<sup>+/-</sup>* mice as well as 3 control mice (lacking one of the genes) all 10-week-old were treated similarly except for STZ injection which was replaced with sodium citrate solution. At the end of the 21-day experiment, the animals were sacrificed and pancreas was isolated and further processed with on-bead digestion with similar protein amounts (3 mg). The 3 control, 3 biotinylated and 3 STZ injected biotinylated animals were labeled with 9 TMT labels and mixed. Samples were fractionated using Hp-RP into 12 fractions (performed by Dr. Robert Hardt) and analyzed using an Orbitrap-Lumos. Raw files were analyzed by Mascot/Perseus. Control/control ratios were used to find A) 95% cut-off for biotinylation B) 95% cut-off for biotinylation was applied on proteins with 3 quantified biotin/control ratios and C) proteins with  $\log_2$  biotin/control ratios above the cut-off and  $p$ -values  $< 0.05$  (as shown in volcano plot in blue) were defined as biotinylated. D) Biotin/biotin ratios were used to find a cut-off for drug effect. E) The drug effect cut-off was applied on biotinylated proteins with 3 quantified STZ/biotin ratios (in red). And F) proteins with ratios above and below the cut-off with  $p$ -values  $< 0.05$  were defined as up and downregulated as an effect of STZ. Experiment's ID number: SA5150.

### 5.2.5.4. Effect of Harmine on beta-cell proteome alterations

After identification of beta-cell proteome, we investigated the effect of Harmine, on the proteome of beta cells. In total 4,611 protein groups (13,689 proteins) were identified in Harmine dataset (Annex Table 5-23). 2,518 protein groups (7,112 proteins) were found as downregulated and 97 proteins groups (252 proteins) were upregulated (Annex Table 5-24). Harmine has been reported to be an anti-cancer/ anti-diabetic drug. So far, no large scale omics analysis has been performed on the effects of Harmine; however, several groups have evaluated the effect of Harmine on biological pathways in different cell line and animal models. To verify the validity of data, the regulated proteins were loaded into Gene Ontology platform and the enriched Reactome pathways were compared with literature. Upregulated proteins could not be matched with any published data. The first two pathways were involved in biotin metabolism and transfer which could be an artefact due to external biotin administration for seven days. However, more than 60 pathways enriched based on the list of downregulated proteins in Harmine datasets were previously published by other groups are listed in (Annex Table 5-25). These pathways included induction of cell cycle arrest, inhibition of inflammation, reduction of lipid and proteins oxidation, inactivation of ERKs, inhibition of ERK/MAPK targets, etc.

### **5.2.5.5. Effect of STZ on beta cell proteome alterations**

In STZ dataset in total 2,742 protein groups (8,252 proteins) were identified (Annex Table 5-26). After applying biotinylation and drug-effect cut-offs, 1,274 groups (3,563 proteins) were found downregulated and only 4 protein groups (6 groups) upregulated (Annex Table 5-27). To validate the data, the downregulated proteins were first compared with previously published proteome and genome analysis. So far very few middle-scale omics analyses have been performed to evaluate the effects of STZ on biological systems. From overall published 168 genes and proteins, we found 47 proteins (Tonne et al. 2013; Tsai et al. 2013). The list of regulated proteins was uploaded in GO platform and searched for enriched Reactome pathways. The enriched pathways were compared to published literature on mechanisms involved in the induction and progression of diabetes type 1. Out of the enriched pathways in our dataset, we found 86 previously linked to diabetes and STZ (Annex Table 5-28). Among these pathways, mitochondrial dysfunction, cell cycle arrest, production of reactive nitrogen species (RNS), pathways related to DNA damage among, insulin receptor recycling and insulin processing can be mentioned.

## 6. Discussion

### 6.1. Establishment of a statistical approach for cut-off determination in BioID-based enrichment methods and its implementation in organelle proteomics

#### 6.1.1. Experimental determination of a cut-off for the acceptance of biotinylated peptides

While very efficient capture of biotinylated proteins is facilitated by the extraordinary stability of the avidin-biotin interaction, their release from avidin-beads is difficult and remains a major challenge. Over the last years, four major approaches have been employed in the majority of studies dealing with the mass spectrometry based characterization of biotin tagged proteins enriched by avidin residues: (1) on-bead digestion of the avidin bead/biotinylated protein complexes followed by extraction and analysis of the proteolytic peptides (Fukuyama et al. 2012); (2) elution of biotinylated proteins from avidin beads followed by in-gel digestion (Cheah and Yamada 2017; Rösli et al. 2008). (3) enrichment of biotinylated peptides from proteolytic digests by avidin (Direct Detection of Biotin-containing Tags; DiDBiT; (Schiapparelli et al. 2014)); (4) enrichment of biotinylated peptides from complex peptide mixtures using anti-biotin antibodies (Udeschi et al. 2017). Schiapparelli and his team detected from 6 mg of starting material with both on-bead digestion and DiDBiT methods (1 and 3) only around 1000 proteins (Schiapparelli et al. 2014). Udeschi and his group detected 526 and 671 APEX-labeled mitochondrial matrix proteins from 2 mg of HEK lysate, with elution and anti-biotin enrichment methods (2 and 3), respectively (Udeschi et al. 2017). Normally for large scale analysis, on-bead digestion is the method of choice (Branon et al. 2018; Sears, May, and Roux 2019) and therefore we applied it for our enrichment analysis and identified ~ 4000 proteins from 1 mg of biotinylated proteins.

While enrichment for biotinylated peptides (methods 3 and 4) allows their direct identification by the search engine through setting biotinylation as a modification, in the first two methods, on-bead digestion and elution, biotin-modified fragments of the proteins remain attached to avidin beads after trypsinization. These two methods result in the identification of a pool of unmodified peptides of two origins. The two populations are (1) biotinylated proteins which bind the streptavidin-coated beads with high affinity and (2) a large number of proteins binding non-specifically to these affinity resins. To distinguish between these two populations, these enrichment techniques are normally combined with quantitative (e.g. SILAC- or TMT-based) mass spectrometry and then the biotin/control ratios are analyzed with two major approaches. (1) A fold-change cut-off is set in an arbitrary way (e.g.  $\log_2$  biotin/control  $> 0$  or  $> 1$ ; (Alvarez-Castelao et al. 2017; Krogager et al. 2018)) to accept a protein as specifically enriched. An arbitrary cut-off, however, has no statistical basis and may result in the wrong annotation of proteins as biotinylated or non-specific. (2) “slide approach” is used which defines a cut-off in a way to include most of the true positive proteins in the significantly enriched population (Branon et al. 2018). Another step that is taken for removing the non-specific proteins before applying the “slide approach” is the subtraction of the background from the list of biotinylated proteins. In enzymatic tagging studies, normally 3plex SILAC is used where one sample expresses APEX/BioID in the region of interest (for example in mitochondria), one sample is non-biotinylated, and the third sample expresses APEX/BioID outside of the region of interest (for example in cytosol). The proteins biotinylated outside of the region of interest (cytosolic) are then subtracted from the reference list (mitochondrial proteome) and then the “slide approach” is used to find a cut-off and differentiate between biotinylated and non-biotinylated (non-

## Discussion

specific) proteins (Hung et al. 2016). Thul and his team mapped the entire human proteome by immunohistochemistry and immunocytochemistry in CELL ATLAS project (Thul et al. 2017) and revealed that approximately half of the proteins localize to multiple organelles. Also, localization of organelle proteins by isotope tagging (LOPIT) studies have shown that more than half of the proteins are present in multiple compartments of the cell (Geladaki et al. 2019). Hence, background subtraction has the risk of losing the biotinylated proteins residing in both organelles. “Slide approach” in general is the most reliable cut-off determination approach and state of the art in the field (Branon et al. 2018). However, it is dependent on the prior knowledge about localization of the proteins in the list and cannot be used for enrichment studies where no previous experimental data is available. We hypothesized it should be possible to establish an unbiased method to define a cut-off based on the ratio distribution of non-specific proteins as part of a null experiment. A dominant approach for statistical data analysis and to correlate a feature to a population (e.g. presence of an effect of interest or regulation by a treatment) assigns a null hypothesis to each feature, thereby yielding useful statistical information that can be applied for the analysis of affected populations. This method rests on the assumption that unaffected population has a distribution and if this distribution is uniform, it is possible to infer that statistical parameters such as  $p$ -values or cut-off intervals determined for this population are with a high probability correct and reliable (Bickel 2011). The null experiment has been used in different fields of biology and other natural sciences to monitor the distribution of non-specific/non-treated population. Ping et al. performed a null experiment on SILAC labeled cells to determine a cut-off to differentiate signal from noise in their subsequent analysis. They applied this cut-off on IPTG induced/non-induced SILAC ratios to analyze the data (Ping et al. 2013). Statistical values inferred from null-distribution have been applied to modeling the data of unaffected genes, non-associated single nucleotide polymorphism, and other features in genome-screening studies (Bickel 2011). Schwartzman and his group have used null distributions in the context of neuroimaging to allow correct inference from data (Schwartzman et al. 2009). Bansal et al., applied the null experiments to chromatin immunoprecipitation followed by sequencing of protein-bound DNA fragments (ChIPSeq) method to differentiate between random binding and specific DNA fragments that are bound by specific proteins *in vivo* (Bansal et al. 2015). Null distribution analysis has also been addressed in other fields of natural sciences such as ecology (Veech 2012) and biogeography (Cayuela, Gotelli, and Colwell 2015).

We performed a null experiment containing identical non-biotinylated populations labeled with different SILAC labels to monitor the distribution of non-specifically binding proteins. Also, we defined the 95<sup>th</sup> percentile of the control/control ratios as a cut-off for biotinylation and used it for interpretation of biotinylation studies. This is to my knowledge the first application of null experiment in proximity labeling studies and no one has before examined the credibility of a 95% confidence interval of a null distribution to be used as a cut-off for biotinylation.

We applied the 95% cut-off determined from a separate null-experiment on biotin/control ratios obtained from 2plex SILAC experiments (2plex organelle datasets). However, it was revealed that the distribution of non-specific proteins in biotinylation experiment can change due to the inter-effects of biotinylated and control samples. The control samples used for determination of cut-off (in the null experiment) had no physical exposure to biotinylated proteins and therefore, were not affected by these proteins. Thus, we concluded that a cut-off from an individual null experiment is not applicable to a separate biotinylation study. As a measure to evaluate the credibility of this cut-off, we also compared the null experiment’s cut-off with that of the “slide approach” that was determined based on GO enrichment analysis. These two cut-offs were very different values which confirmed that the null experiment cut-off is not reliable. We, therefore, designed a 3-plex SILAC

## Discussion

experiment including two control samples and one biotinylated sample (nuclear BioID) which were combined 1:1:1 and incubated together with the beads. In this setting, the two control samples create a null distribution which is still influenced by the presence of biotinylated proteins. We defined a 95% cut-off on control/control ratio distribution and compared it to a cut-off determined by “slide approach” and the values were almost identical. We also confirmed the localization of the enriched proteins by comparing them to literature, CELL ATLAS ((Thul et al. 2017); [www.proteinatlas.org](http://www.proteinatlas.org)), and GO (Geneontology.org) terms and found above 80% of the proteins linked to the nucleus. The equality of our method’s 95% cut-off to the slide cut-off and the good match with literature further confirmed the validity of our enrichment analysis and the potency of our cut-off determination approach to be implemented in any enrichment experiments. This method is very useful especially for the interpretation of data and identifying the true positive hits in studies that are performed for the first time and no prior information on the identity and localization of the examined proteins is available. For instance, this approach can be applied on studies that utilize BioID/APEX for characterization of the interacting proteins of a fused-protein or proteins residing in a specific cell type/organelle of a not well-known organism, for which no prior information may exist.

### **6.1.2. Analysis of organelle-specific proteomes including one control sample.**

BioID constructs fused to different organelle targeting signals were used to identify the proteins located in nucleus, mitochondria, and cytosol as part of a 2plex SILAC experiment. Light/heavy (biotin/control) SILAC ratios were analyzed using the 95% cut-off determined from the null experiment to identify the biotinylated proteins. We expected to see a bimodal ratio distribution with distinguishable clusters: one for biotinylated proteins and one for non-specific proteins that fits the normal distribution of the null experiment. However, this was not observed in all the 2plex datasets. Mitochondrial dataset showed a group of proteins with  $\log_2$  ratios around 0. This population was almost missing in cytonuclear and nuclear proteomes.

We compared our observations with the results of other groups who had utilized APEX and BioID for affinity enrichment of biotinylated proteins in organelles. In most of these studies “slide approach” has been used to identify the biotinylated proteins. For the characterization of slide cut-off, the distribution of specific (true positives) and non-specific (false positives) proteins are shown as mirror histograms which reveal the range of biotin/control ratios in different proteomes ((Branon et al. 2018); Figure 6-1). BioID and APEX data showed uniform ratio distributions with no separate clusters for biotinylated and control proteins; the same trend was observed in our 2plex histograms.

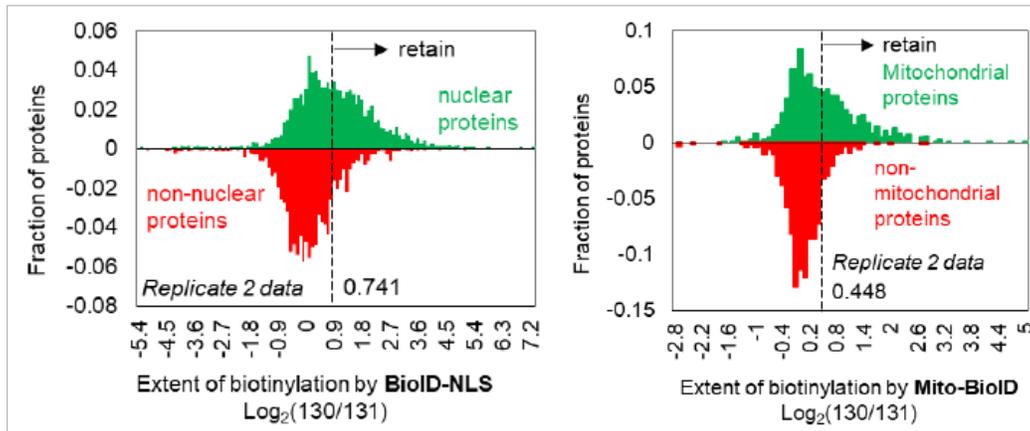
$\log_2$  biotin/control ratios for mitochondrial matrix proteins of HEK293 cells identified by APEX and BioID are ranged between -1 to 3 (Rhee et al. 2013; Branon et al. 2018). Our mitochondrial proteome showed a wider range between -2 and 5  $\log_2$  ratios. In all these three datasets, a group of proteins form a cluster around 0 that is similar to the null distribution. We, therefore, assumed it could be possible to apply the null experiment cut-off for the analysis of mitochondrial dataset, as it contained a cluster of nonspecific proteins around  $\log_2$  ratio 0. However, as nuclear and cytonuclear datasets were not showing this population and we aimed to find a cut-off that is applicable to all studies we avoided analysing these datasets with the null experiment cut-off.

We next compared the distribution of nuclear dataset with that provided by Branon. In Branon’s dataset, the nuclear proteins acquire higher biotin/control ratios than the mitochondrial proteins; a similar trend is observed in our datasets. The histogram for nuclear proteins provided by Branon has ranged between -1 to  $\sim 4.5$   $\log_2$  ratios while our proteome is distributed between -1 and 4  $\log_2$  ratios. The higher ratios for nuclear (and cytoNuclear) datasets in comparison to mitochondrial dataset might be due to the fact that highly abundant

## Discussion

proteins from these two organelles represent the majority of non-specific binding proteins and that the current strategy allows for more of them to be quantified as twice the amount of background proteins are used.

In the nuclear datasets from Branon's lab, a large number of proteins obtain  $\log_2$  ratios around 0, which is not similar to our observations for nuclear (and cytonuclear) datasets; in our datasets, 95% of proteins obtained ratios above 0 (Branon et al. 2018).



**Figure 6-1. Distribution of nuclear and mitochondrial matrix proteomes identified by BioID.**

Histograms show distribution of nuclear and mitochondrial matrix proteomes obtained via BioID-catalyzed labeling as well as non-nuclear and non-mitochondrial populations used for ROC analysis and cut-off determination (Adapted from Branon et al., 2018)

One reason for this discrepancy could be that the ratios provided by Branon's group are normalized based on biotin ligase expression in each dataset (Branon et al. 2018) and our values provided by MaxQuant normalized ratios based on the median of all individual peptide ratios. The histograms depict the normalized ratios and we cannot predict if both dataset were normalized with a similar approach how different the distribution ranges could be. Another reason for this could be the different materials used for the enrichment of biotinylated proteins. We used streptavidin beads from GE healthcare and Branon used streptavidin magnetic beads from Pierce (Branon et al. 2018). The matrix of these beads may have different affinities to biotinylated and non-specifically binding proteins resulting in different arrangements for proteins on the matrix of the bead and subsequently varying distribution of biotin/control ratios. Another assumption could be that in our experiments due to wrong readings of protein assay, higher amounts of biotinylated proteins in comparison to the control sample have been used, resulting in a shift of biotin/control ratios towards greater values. However, as the biotin/control ratios in all the three replicates of nuclear and cytonuclear BioID experiment are clustered around the same value, the probability of this hypothesis is low as it is very unlikely that the protein assay readings are wrong for all the three replicates of individual experiments.

To further evaluate the reason for high biotin/control ratios in our nuclear (as well as cytonuclear) datasets we compared the ratio distribution of non-nuclear proteins in the null experiment and in the nuclear dataset. The expected distribution of non-nuclear proteins (around  $\log_2$  ratio 0) as seen in the null experiment was not observed in the nuclear dataset; non-nuclear proteins were clustered around  $\log_2$  ratio 1.4. A possible reason for this would be that non-specific binding proteins in the nuclear proteome bind to the biotinylated proteins resulting in a shift for non-specific proteins towards larger biotin/control values and mixing with biotinylated proteins.

## Discussion

By assessing the intensities of non-specific (non-nuclear) proteins in the nuclear dataset, we detected two populations of proteins: A group of the non-nuclear proteins had equal intensities to those in the null experiment and another group acquired 20 times lower intensities. This promotes another hypothesis that when biotinylated and control proteins are mixed and incubated on beads, even though the protein-bead ratio is kept constant, biotinylated proteins compete for binding to the beads and replace the background population increasing the biotin/control ratios.

### 6.1.3. Effect of biotinylated proteins in the binding pattern of non-specific proteins

To discover which of the hypotheses about the effect of biotinylated proteins on such non-specifically binding is credible, we performed another experiment using nuclear BioID (which showed the highest shift in biotin/control values- in light channel). We included two control samples (in medium and heavy channels) and mixed the samples at different steps of on-bead digestion protocol. To evaluate the effect of biotinylated proteins on background proteins, we monitored the average intensities of biotin and control SILAC channels in methods 1 and 4. Method 1 is similar to the 2plex organelle experiments as biotinylated and the H control samples are mixed from the beginning (before incubation with beads) and meet the M control before digestion. In method 4 however, both controls are combined with biotinylated proteins before incubation on beads. In method 1, the average intensity of light SILAC channel (biotinylated proteins) was found to be much higher than the average intensity of H control proteins (combined from the beginning with biotinylated proteins); “the average intensity in biotin channel/ average intensity in H control” was  $\sim 4.5$ . However, “average intensity in biotin channel/ average intensity in M control” (that was incubated in a separate tube) was much smaller ( $< 2$ ). This is in line with the hypothesis that biotinylated proteins compete with background proteins to bind the beads (although the protein-bead ratio is constant) when they are in the same tube; in method 1 the biotinylated proteins ( $L^*$ ) bind stronger the empty spaces on bead surface and obtain higher intensities while non-specific binders in H channel, as well as those in  $L^*$  channel, have a lower chance to bind the beads and acquire lower intensities than their counterparts in M control channel. This is due to better access of M control proteins to the beads and no competence with biotinylated samples. As a result,  $L^*/H$  ratios shift to greater values. If this theory is true, the non-specific proteins in the biotinylated sample as well as those in H channel acquire low intensities and therefore many of them are not detected by MS. However, interestingly as soon as another group of non-specific proteins is added during incubation with beads (experiment 4;  $L^* + M + H$ ), the capacity and chance for non-specific proteins to bind the beads is increased and M and H controls (as well as the non-specific proteins in biotinylated sample) find the chance to bind the beads with similar strength. As a result, non-specific proteins in biotinylated samples ( $L^*$ ) and non-specific proteins in control samples (M and H) acquire similar intensities and create a cluster of  $\log_2$  biotin/control ( $\log_2 L^*/M$  and  $L^*/H$  ratios) around 0. Another cluster is formed by biotinylated proteins with higher intensities (in  $L^*$ ) over non-specific proteins in control samples (M and H) that is distributed around  $\log_2$  biotin/control ratios  $> 0$ . This could be the reason for the creation of a bimodal biotin/control ratio distribution pattern for this method, as described previously. We, therefore, think that competition for binding to beads is much alleviated in case of the 4<sup>th</sup> method (where all samples are mixed before incubation with beads) and therefore the biotin/control ratios are not as shifted as in the 1<sup>st</sup> method and organelle studies.

We also proposed another hypothesis for shifted biotin/control ratios after observing the ratio distribution of nuclear (true positive hits) and non-nuclear (false positive) proteins in nuclear 2plex dataset. We suggested this could be due to binding of non-specific proteins to specific proteins in the biotinylated population and an

## Discussion

increase in their intensities resulting in a mixture of proteins with high biotin/control ratios. To see if a similar trend is seen in the current study we assessed the ratio distribution for false and true positive proteins based on GO analysis for biotin/control ( $L^*/M$ ) ratios in method 1 ( $L^* + H / M$ ) and method 4 ( $L^* + H + M$ ) as part of the “slide approach” and observed the same trend. Nuclear and non-nuclear proteins were clustered around two ratios that were only slightly different from each other (the gap was bigger in case of method 1) and the proteins passing the slide cut off were a mixture of nuclear (majority) and non-nuclear (minority) proteins. This validates the hypothesis of non-specific binding of non-nuclear proteins to truly biotinylated ones. We also compared the proteins passing the 95% cut off in method 1 and method 4 with CELL ATLAS (Thul et al. 2017); [www.proteinatlas.org](http://www.proteinatlas.org), and GO (Geneontology.org) and the nuclear list provided by Branon (Branon et al. 2018). We found 85% and 77% of the proteins with prior nuclear annotation for methods 1 and 4, respectively. This means that the other 15%- 23% non-specific proteins in these lists may have acquired high biotin/control ratios due to non-specific binding to biotinylated proteins in the biotin ( $L^*$ ) channel. Considering all these observations, we believe method 4 minimizes the competition effect however still does not result in a pure enrichment for biotinylated proteins due to unavoidable mix of non-specific and specific proteins. The mirror histograms provided for cut-off determination in other BioID and APEX studies (Branon et al. 2018; Hung et al. 2016; Rhee et al. 2013) also confirm that the list of proteins assigned to each organelle includes a population of non-specific proteins (Figure 6-1). This means that the same trend can be seen in all approaches which involve enrichment of biotinylated proteins (e.g. proximity labeling) and pure enrichment cannot be achieved.

### 6.1.4. Analysis of organelle-specific proteomes including two control samples

We used the 3plex approach (containing one biotinylated and two control samples; differentially labeled with SILAC approach) to identify the proteomes of different cellular compartments labeled by organelle targeting BioID constructs. Although the total numbers of identified proteins were comparable, the numbers of biotinylated proteins (with  $p$ -values  $< 0.05$ ) were smaller than those identified in 2plex studies. This is probably due to using an invalid cut-off to find the enriched proteins in the 2plex studies which was previously defined in a separate null experiment and excluded the inter-effects of biotinylated and control samples. In 3plex studies, the cut-off was defined based on the null distribution of the two control samples within the study and matched the “slide approach” cut-off. We are therefore more confident that the proteins identified and assigned to organelles in the 3plex study contain more true positives. We compared the data with literature and found 97%, 77%, and 53% of the cytonuclear, nuclear, and mitochondrial proteins previously linked to these organelles by literature/databases (GO ([geneontology.org](http://geneontology.org)); (Branon et al. 2018; Calvo, Clauser, and Mootha 2016; Rhee et al. 2013; Smith and Robinson 2016; Thul et al. 2017; Wessels et al. 2013)). Branon et al. used nuclear BioID and detected 1512 nuclear proteins, 79% previously assigned to the nucleus. We found 787 proteins 77% linked by previous studies to the nucleus (Branon et al. 2018). The number of enriched proteins in nuclear dataset was also smaller than those enriched by our other organelle-targeting BioID constructs. One possible reason for this could be low transfection efficiency for nuclear-BioID sample in comparison to the other samples. The 3plex organelle proteomics was performed using the same biotinylated proteins applied to the 2plex study, except for the nuclear BioID, that was repeated from the transfection step again. We did not evaluate the transfection efficiency for this construct. Another reason could be that Branon used stably transfected cells and we performed transient transfection. Stable transfection results in a similar expression of BioID in all cells within the population used for enrichment and subsequently better enrichment efficiencies.

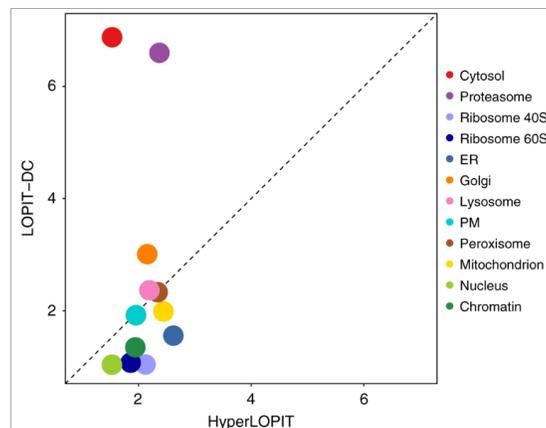
## Discussion

Also, differences in input amounts (Branon used 3mg and we used 1mg per condition), labeling material, and enrichment protocols could count for discrepancies in enrichment efficiencies.

For the mitochondrial BioID expression, Branon used mitochondrial matrix targeting sequence from COX4 (Lemire et al. 1989) and identified 362 proteins, 70% previously assigned as mitochondrial. We used the mitochondrial matrix targeting sequence SU9 (John et al. 2002), which resulted in identification 1248 proteins, 53% (660 proteins) with prior mitochondrial annotation. Rhee et al found 464 proteins in the matrix of HEK293 cells with APEX proximity labeling approach (Rhee et al. 2013). In total the mitochondrial proteome has been estimated to contain around 1000-1500 proteins (Thul et al. 2017). CELL ATLAS and MitoCarta 2. link in total 1098 (524 in multiple locations) and 1158 proteins to human cell's mitochondria. Mitochondrial matrix includes the mitochondrial DNA and the majority of the mitochondrial proteins (Nunnari and Suomalainen 2012). LOPIT approach in combination with differential ultracentrifugation assigned 464 proteins to the whole mitochondrion (Geladaki et al. 2019). We, therefore, think the performance of our approach for identification of mitochondrial proteins is comparable to the state of the art strategies. Our list may contain unknown mitochondrial proteins which need antibody-based tests to validate their localization.

We also identified 295 proteins shared among the nuclear and mitochondrial datasets which showed different and distinct ratio distribution across datasets. It is reported that proteins associated with different cellular organelles show a distinct distribution of fold change as a result of underlying differences in cell morphology or their roles in different compartments (Parca et al. 2018). Similarly, LOPIT approach is based on the measurement of differential distribution of organelle markers across density gradient fractions. This method assigns proteins to organelles based on proteins distribution (intensity in each fraction) across multiple gradient fractions and by comparing protein profiles to those of organelle markers using statistical and machine learning approaches. Although in LOPIT approach distribution of marker proteins is detected in density gradient fractions, it demonstrates that proteins specific to each organelle show distinct and differential distributions and abundance profile that allows assigning other proteins with a similar profile to the same organelle (Geladaki et al. 2019). These reports confirm that differences in ratio distribution of proteins from different organelles are expected and even useful for further categorization of proteins with unknown localization.

Distance analysis for these common proteins showed very low correlation between Mito.-BioID and Nucl.-BioID common biotinylated proteins. Also, PCA analysis performed by Geladaki et al. in LOPIT approach (Figure 6-2), confirms the distance of mitochondrial from nuclear proteins (Geladaki et al. 2019).



**Figure 6-2. Distance correlation across organelles analyzed by LOPIT approach.**

LOPIT in combination with differential ultracentrifugation (LOPIT-DC) and in combination with isotope tagging (hyperLOPIT) provide high-resolution maps of protein subcellular localization. The minimum distance for each of the 12 subcellular compartments by these two techniques are shown in the figure. (Adapted from Geladaki et al., 2019)

### **6.1.5. Comparison between organelle proteomics approaches**

The classical technique for isolation of organelles relies on cell fractionation and differential/density gradient centrifugation (Breuza et al. 2004; Knoblach et al. 2003; Mears et al. 2004; Pisitkun, Shen, and Knepper 2004; Yan, Aebersold, and Raines 2009). This technique is prone to cross-contamination from other organelles and only the compartments that are accessible can be purified (Agrawal et al. 2011). Proximity labeling using BioID and APEX has improved these techniques by selective and specific labeling of the proteins in the organelle of interest while avoiding the difficult separation techniques and reducing the contamination from other organelles. As proteins are not isolated in their native state, harsh and stringent wash conditions can be applied to reduce the number of false positive hits after lysis (Mair et al. 2019). Proximity labeling techniques exclude the fractionation steps for purification of the proteins and do not require complicated statistical analysis for interpretation of data (Mair et al. 2019). These techniques, however, provide subcellular snapshots of discrete cellular compartments that cannot be easily integrated to examine proteins with multiple localizations. LOPIT technique, on the other hand, enables protein distribution across multiple fractions and reveals possible multiple locations for proteins (Geladaki et al. 2019). LOPIT still involves density gradient centrifugation and relies on statistical and machine learning approaches which need knowledge and expertise in these fields. I, therefore, believe the results of these techniques in combination with antibody-based localization studies (Thul et al. 2017) can be used as complementary to each other to enable more reliable assignment of proteins to subcellular compartments in the cell.

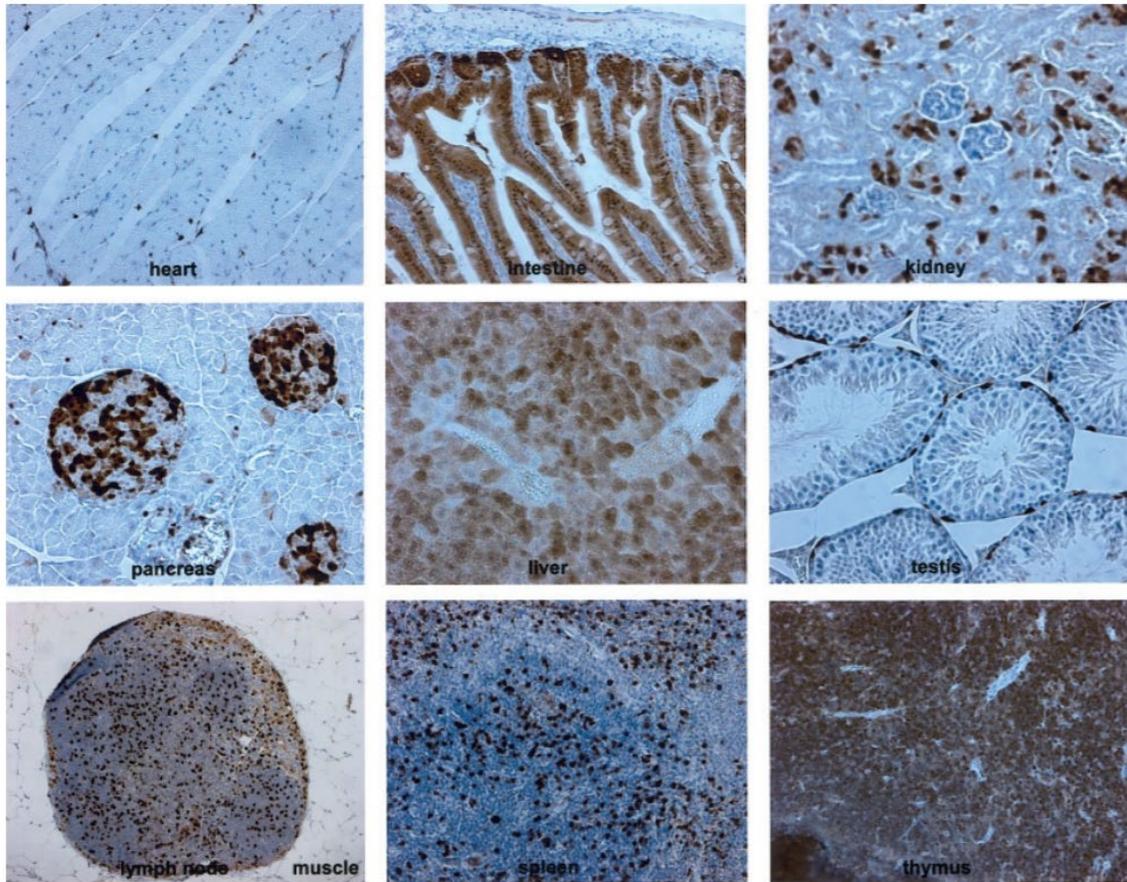
## **6.2. *In vivo* cell-type-specific analysis in BioID mice models**

Besides implementation of BioID method in organelle proteomics, we applied it to cell-type-specific analysis in mouse models. To this aim two mice harbouring BioID gene under *Col1a1* (mouse 1) and *Rosa 26* (mouse 2) were created which in combination with Cre lines enable cell-type-specific biotinylation and proteome identification.

### **6.2.1. Tissue expression of BioID mice (Rosa26 vs COL1A1)**

Upon expressing BioID under ubiquitous promoters, expression and biotinylation by this enzyme were evaluated in different tissue types. BioID expression and biotinylation were more significant and spread in various cell populations of mouse 2 (Rosa26-CAG-BioID) and less pronounced in mouse 1 (*Col1a1*-BioID). This is due to the fact that *Rosa26* is a ubiquitous and efficient locus that in combination with the CAG enhancer-promoter drives strong expression of BioID in diverse tissue and cell populations. *Col1a1* promoter, on the other hand, is a less versatile locus with lower expression potencies than *Rosa26*. *Col1a1* mouse was originally created by the group of professor Jänisch (Beard et al. 2006) and was reported to drive a patchy expression of the transgene in the lung and heart and no expression in brain and skeletal muscle (Figure 6-3). We detected a similar expression pattern in these tissues.

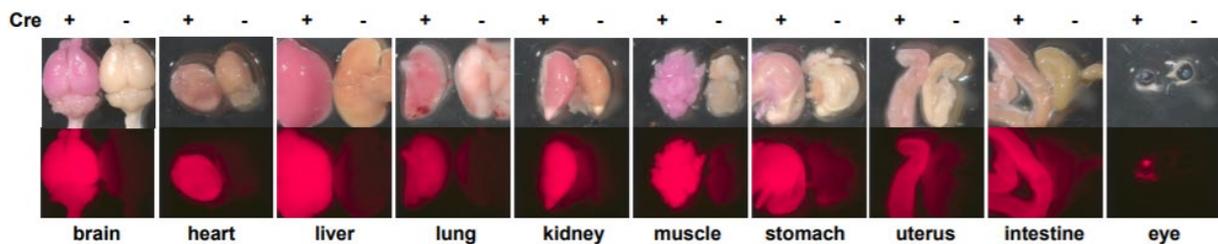
## Discussion



**Figure 6-3. Induction of EGFP expression in Col1a1-EGFP mice organs.**

Paraffin sections of representative tissues were immunostained with anti-EGFP antibody (brown signal) and counterstained with hematoxylin (Adapted from Beard et al., 2006).

Rosa-CAG-BioID was generated based on the publication by Madisen (Madisen et al. 2010) in which transgene was shown to be expressed in all major brain regions and cell populations as well as in a wide range of tissues (Figure 6-4).



**Figure 6-4. Comparison between EIIa-Cre/Rosa-CAG-Ai14 mouse (left, Cre+) with that in tissues from a control Rosa-CAG-Ai14 mouse (right, Cre-).**

EIIa-Cre (adenovirus EIIa promoter-driven Cre) induces recombination during early embryonic development and tdTomato expression is turned on throughout the body. In the lower panels, all EIIa-Cre/Ai14 tissues showed strong red fluorescence and no fluorescence was observed in Ai14 tissues (Adapted from Madisen et al., 2010).

We also detected a system-wide BioID expression and biotinylation in Rosa26-CAG-BioID mouse organs except for the spleen with patchy and low signals for which no data was provided by Madisen.

### 6.2.2. “Biotinylation cut-off” and “Drug-effect cut-off” for interpretation of data

We intended to perform cell-type-specific analysis on mice with 3 variable factors (control/ biotinylated/ biotinylated and drug administered). To analyze such data two cut-offs are required: one to distinguish biotinylated proteome from non-specific binding proteins (biotinylation cut-off) and one to discriminate between biotinylated and drug-affected proteomes (drug-effect cut-off). In general, the optimal situation would be to find both biotinylation and drug-effect cut-offs as part of a single experiment and not from individual studies. However, the 3plex dimethyl reagents do not allow the inclusion of more than 3 samples. To determine both cut-offs in the same study multiplex labeling is required. We quantified the tissue samples from mice with ubiquitous BioID expression (for method development) with dimethyl method as it is non-expensive and robust and for the cell-type-specific analyses, we considered multiplexing using TMT tags.

For determination of the 95% cut-off for biotinylation, similar to *in vitro* analysis, first, a separate null experiment on pancreatic tissue samples was performed. The 95% cut-off was determined as  $\log_2$  control/control ratio of +1; this was much greater than the *in vitro* null experiment cut-off ( $\log_2$  ratio +0.18). The non-specific proteins in tissue samples were distributed over a wide range of  $\log_2$  ratios (between -2 to 2) while the ratios for non-specific binding proteins in HEK 293T cells ranged between  $\log_2$  values -0.5 and 0.5. Normally *in vitro* samples can incorporate the labels at the protein level resulting in less variation across the samples. For *in vivo* samples it is too expensive to label the whole living organism (Gouw, Krijgsveld, and Heck 2010). Moreover, whole organism labeling only allows tagging the lysine residues and at the end, only two states (2plex) can be investigated (Gouw, Krijgsveld, and Heck 2010). For these reasons tissue samples are normally labeled at the peptide level which means protein samples are processed in separate tubes resulting in higher technical variation across the individual samples and a wider range of null distribution.

Similar to *in vitro* analysis, due to variations in distribution of non-specific proteins in biotinylation experiments, it was revealed to us that a 95% cut-off from a separate null experiment is not valid and applicable on a separate biotinylation study. As these experiments (*in vitro* and *in vivo*) were performed at the same time, we were not aware of this fact earlier. We, therefore, included besides biotinylated sample(s), two control samples for dimethyl studies (tissue proteomics) and three control samples in TMT studies (cell-type-specific proteomics) to enable calculation of biotinylation cut-off within the study based on the distribution of control/control ratios.

Drug-effect cut-off reflects the technical variation between two differentially labeled proteomes originating from the same biotinylated biological sample. Despite the biotinylation cut-off, this is a bidirectional cut-off as the drug causes up- and down-regulation of proteins. If the distribution of biotin/biotin ratios is not normal, it is important to define two individual 95% cut-offs on both tails of the histogram, as we did for dimethyl analysis. However, for TMT labeled samples, as the intensity of the same TMT labeled sample can be used as both numerator and denominator of a biotin/biotin ratio, the histograms created from the mean of biotin/biotin ratios are rather normally distributed and determination of 95% cut-off on one tail is probably sufficient. To ensure preciseness of the cut-off, I can propose defining as many biotin/biotin ratios as possible (6 ratios for the three biotinylated samples) which in turn results in higher uniformity of the final (mean) distribution of biotin/biotin ratios. Another suggestion would be to define two drug-effect cut-offs on both tails of the biotin/biotin distribution, especially if the data distribution is not normal.

### 6.2.3. Method Development

We realized after performing tissue proteomics that our enrichment protocol does not yield acceptable identification numbers and enrichment rates. To improve it, we performed extensive method development analysis with a focus on the pancreas which showed the worst performance across the tested tissue samples.

Efficient enrichment of biotinylated proteins using streptavidin beads and removal of non-specific binding proteins requires incorporating stringent wash steps using solutions containing poly ethylene glycol (PEG)-based detergents. PEGs share certain properties with peptides and can therefore be problematic when analyzing samples of biological origin. They can, for example, interfere with sample preparation or liquid chromatography–tandem mass spectrometry (LC–MS/MS) analysis due to competitive binding to C18 reversed phase chromatographic material (Ahmadi and Winter 2018). In the mass spectrometer, similar to peptides, PEGs and other polymers can result in multiply charged ions which can trigger fragmentation in the data-dependent acquisition (DDA) mode. Consequently, if present in high abundance, they may suppress peptide ionization or interfere with peptide detection, if the mass spectrometer selects the polymer peak instead of the peptide signal for fragmentation. To investigate the effect of PEG molecules on identification of peptides we developed a strategy to identify PEG and PEG-based detergents using Mascot, Protein Pilot ([www.sciex.com](http://www.sciex.com)), and MaxQuant search engines and adapted them for the automated identification of PEG by definition of an artificial amino acid in combination with a custom-made database. This allowed evaluating the potential influence of PEGs on the numbers of identified peptides in PEG-contaminated samples (Ahmadi and Winter 2018). For PEG-contaminated peptide samples, we performed single searches against PEG database and Uniprot as well as combined searches to allow the algorithm to assign a given MSMS spectrum to both peptides and polymers. The simultaneous searching did not alter the results of peptide identification. This indicates that most likely no peptide derived MSMS spectra were wrongly annotated as polymers (Ahmadi and Winter 2018).

PEG-based molecules are also severe contaminations for mass spectrometry as they collect and concentrate on the column and prevent the peptides from binding it and this becomes a more important issue when the proteome from certain scarce cell types across the tissue (although enriched) is to be identified. To increase the efficiency of on-bead enrichment methods we replaced PEG-based wash buffers with urea wash buffers as well as Ripa lysis buffer (containing 2% Triton X-100) with 4% SDS. We also included a precipitation step of the protein lysate followed by resolubilization of the pellet in 1% SDS before incubation with beads. This had two benefits: (1) PEG contamination was significantly reduced in spectra. (2) Protein assay was found to be more reproducible and reliable for samples precipitated and dissolved in 1% SDS; this is probably due to complexity of the tissue sample in Ripa buffer which may contain ingredients non-compatible with protein assay. As soon as the sample is precipitated and washed with methanol, many of the undesirable interfering agents such as lipids and nucleic acids are removed (Feist and Hummon 2015).

Besides the main steps explained for method development, we focused on minor sources of variation contributing to the dissimilar distribution of  $\log_2$  dimethyl ratios among the replicates of tissue samples and improved them while doing the method development experiments. Variations could have biological and technical sources: (1) biological variation: the expression of BioID can vary across the mice with the same genotype depending on promoter expression and uptake of biotin and doxycycline by each mouse (water consumption). To improve the latter we kept mice in separate cages to prevent their possible competition for water; (2) Technical variation: variabilities introduced during different steps of the protocol can impact reproducibility. Sources of technical variation are (A) Isolation speed; especially in case of the pancreas, it is

## Discussion

very important to isolate the organ in less than 30 seconds after mice are sacrificed and snap freeze it (Clardy et al. 2015) and so we did. (B) Variations in BioID expression in different parts of the tissue (C. Liu 2013); we used to isolate part of the tissue and not the whole tissue. It can also contain contaminations from other organs especially in case of the pancreas which is very difficult to separate from connective tissues and the fat bound to it (Clardy et al. 2015). In the case of the non-uniform pattern of expression or presence of contaminating organs (counting for protein concentration in protein assay), the biotinylation yield can vary from replicate to replicate. We tried to always isolate similar parts of the pancreatic tissue by repeating a certain pattern of isolation. (C) Incomplete lysis of the organ; if the extracted organ is not completely lysed, part of the proteome including biotinylated proteins is lost which may result in non-reproducible enrichment (Shao et al. 2016). For 2plex samples, we used 8 x volume RIPA buffer which not always allowed complete lysis of the whole organ. We ensured in method development analysis complete lysis of the organ. (D) Sample loss in each freeze-thaw cycle: RIPA buffer has to be used ice-cold and is very sensitive to temperature (Winter and Steen 2011) (Winter and Steen 2011), and it could be that parts of the biotinylated proteins were lost during freeze thaws. To avoid this we kept samples in aliquots but freeze-thaws are not always avoidable. (E) Bead loss: this can easily happen by pipetting the beads out during wash and elution steps, as beads are very light and unsettled. To avoid this we waited 1 minute for beads to settle down and used small pipette tips on top of larger tips to avoid losses. (F) Insufficient or harsh wash: any differences between the wash conditions such as speed, time, and volume, can result in better/worse enrichment efficiency of individual samples as more/fewer non-biotinylated proteins are detached from beads, respectively. Also, high-speed rotations can destroy the beads (Robert 2007). We, therefore, kept all these items constant in method development experiments for all the samples. (G) Insufficient exposure to trypsin: we realized by the time that it is not sufficient to keep the beads with the enzyme in thermomixer, as the beads settle and lower layer has no exposure to the enzyme. To solve this we used benchtop orbital shakers. (H) Non-specific peaks from avidin: another major technical problem with these samples and their analysis is the presence of large amounts of streptavidin peptides which may avoid the low signal peptides to be identified by MS analysis. One solution to this would be to set up the method in a way to specifically search for reporter ions of avidin peptides and exclude them from the MS fragmentation list. However, the suppression effect induced by the presence of major species cannot be completely excluded (Hodge et al. 2013). Another solution is chemical modification of the arginine and lysine residues of avidin to prevent their cleavage by trypsin (Barshop et al. 2019). As a result of paying attention to all these issues the identification efficiency for pancreatic samples was increased around 8 fold and the enrichment by 5 fold for the best enrichment method (4% SDS, 4M urea wash).

We applied this method for enrichment of biotinylated proteins in beta cells and identified a total number of 4133 protein groups. There are only two studies that have performed cell-type-specific analysis *in vivo*. Krogager and Alvarez-Castelao performed *in vivo* cell-type-specific labeling using noncanonical amino acid tagging (SORT and BONCAT techniques) and enriched for the proteins in striatal neurons and excitatory neurons of the mouse brain, respectively (Alvarez-Castelao et al. 2017; Krogager et al. 2018). Brain and pancreas are known to contain a comparable number of proteins (out of 19670 all human proteins 16227 and 14490 are detected in human brain and pancreas, respectively; (Uhlén et al. 2015)). We, therefore, compared the number of proteins identified in these studies to our detection rate as a measure to evaluate the identification efficiency of our optimized method. Krogager et al. identified a total number of 2119 proteins in the enriched proteome (before applying the cut off) from striatal neurons in the brain (Krogager et al. 2018). Alvarez-Castelao and his team found in total 2384 proteins (1110 unique) in hippocampal excitatory neurons of

## Discussion

the brain (Alvarez-Castelao et al. 2017). Our approach outperformed these two methods in terms of identification.

We also compared the enrichment efficiencies for recovery of labeled proteins across these studies. Krogager set the  $\log_2$  ratio of 1 as a cut-off and identified 1604 proteins as enriched in striatal neurons (Krogager et al. 2018). By applying this arbitrary cut-off on our dataset, we could find 2976 protein as enriched that also obtained  $p$ -values  $< 0.05$  and were therefore significant. With  $\log_2$  ratio 0 as a cut-off for biotinylation, Alvarez-Castelao and his team assigned 1247 proteins as enriched in hippocampal excitatory neurons (Alvarez-Castelao et al. 2017). We applied this arbitrary cut-off to our dataset and detected 2991 proteins as enriched that were also significant ( $p$  values  $< 0.05$ ). Based on these comparisons, I believe our enrichment technique is more efficient in the purification of the labeled proteome. This means that by stringent washes and efficient enrichment the numbers of enriched proteins, as well as their proportion with the total number of identified proteins, are improved.

As mentioned, we optimized our pancreas isolation protocol to avoid protein loss and performed mass spectrometry to identify and analyse the whole proteome of the pancreas (before enrichment) which resulted in the identification of 6,483 proteins on Orbitrap-Lumos from three biotinylated and 3 control samples. To evaluate the efficiency of our pancreas isolation method, we compared our identification with that of other groups who have performed whole proteome analysis on pancreas tissue. Hu and his group identified 302 proteins from the whole proteome of human pancreatic tissue using matrix-assisted laser desorption/ionization-time of flight-mass spectrometry (MALDI-TOF-MS) (Hu et al. 2004). Liu et al, detected 5368 proteins from a pool of 5 healthy and 5 diabetic patient whole proteome pancreas samples using a Q-Exactive mass spectrometer (C. W. Liu, Atkinson, and Zhang 2016). Burch et al identified 1,167 proteins from the combined MS runs from whole proteome analysis of pancreas tissues from 5 donors with diabetic and non-diabetic conditions using a Q-Exactive mass spectrometer (Burch et al. 2015). Based on these comparisons, we can infer that the established isolation techniques were effective and capable of preventing loss of proteins due to autodigestion of tissue by pancreatic enzymes.

### **6.2.4. Tissue proteomics including two control samples**

We enriched the biotinylated proteomes of lung, liver, and pancreas from mice with ubiquitous BioID expression as part of dimethyl 3plex studies (including 2 control samples). The efficiency of enrichment in lung and liver MS data correlated with the expression and biotinylation pattern previously observed by immunofluorescence analysis. In liver (with 92% enrichment efficiency), most of the cells were biotinylated while in lung (with 46% enrichment efficiency) a patchy expression and biotinylation pattern was observed. The enrichment efficiency for pancreas was however much better than the expected proportion of biotinylated cells overall pancreatic cells. Mouse 1 showed a distinctive pattern of BioID expression restricted to islets of Langerhans. Islets constitute ~2% of pancreas volume (Jansson et al. 2016) and the efficiency of enrichment for this tissue was around 85% which is possibly a result of extensive method development and optimization of isolation techniques exclusively for the pancreas. This means by improvements in isolation and enrichment techniques a large number of biotinylated proteins are captured and by optimized wash steps many of the non-specific binding proteins are removed and these result in higher proportion of enriched proteome over the total number of identified proteins.

### 6.2.5. Duration for Doxycycline administration

We aimed to use Tet-off inducible system in combination with cell-type-specific analysis to allow temporal control over BioID expression. We designed an experiment to find the duration of time required without Dox to reach maximum BioID expression. Tet inducible systems suffer from some drawbacks (Rennel and Gerwins 2002) and one problem which was observed in our experimental mice is that tTA is only poorly activated and requires more days without Dox for maximal induction of tTA expression and subsequently, maximum expression of BioID. Moreover, the induction levels for tTA can vary in different mice with the same genotype (Dow et al. 2014). Although we aimed to use the advantage of temporal control of gene expression through Tet-off system, to ensure achieving high and reproducible expression of BioID, we removed this parameter from the experimental setting. Using temporal control over BioID expression requires (1) optimization of Dox treatment to reach reproducible and high gene expression and (2) establishment of a method to normalize for the amount of biotinylated proteins at the input level, so that we can incubate the same amount of biotinylated proteins with beads rather than the same concentration of total proteins.

### 6.2.6. Beta Cell proteomics

Application of BioID to beta cell proteomics resulted in the identification of 2992 protein groups (8882 unique proteins). This is to my knowledge the most comprehensive catalogue of beta cell proteins and the first beta cell *in vivo* proteomics analysis performed so far. The other groups isolate the pancreas and cultivate the beta cells *in vitro* which does not resemble the *in vivo* physiological conditions. Some of them have mapped the proteome of pancreatic islets with different cell-type-specific approaches. Three groups have used laser capture microdissection (LCM) in combination with mass spectrometry to analyze differential protein expression in human islets obtained under different conditions and pathological states and have identified 300, 1104 and 2032 unique proteins from human pancreatic tissue samples (Nishida et al. 2014; L. Zhang et al. 2017; Nyalwidhe et al. 2017). Waanders et al. performed mass spectrometry on 2000-4000 single handpicked islets, as well as on pooled pancreatic islets and identified 2013 and 6873 unique proteins, respectively (Waanders et al. 2009). Sacco et al. performed mass spectrometry on whole pancreatic islets from C57Bl/6 mice cultured *in vitro* and identified 7,800 unique proteins (Sacco et al. 2016).

A few groups have performed beta cell proteomics to identify the proteins residing in this cell type. Sacco et al performed mass spectrometry on pancreatic beta cell line Min6 and identified 8360 unique proteins (Sacco et al. 2016). Brackeva and his group utilized FACs for purification and enrichment of the human and rat endocrine alpha and beta cells of the pancreas which yielded in the identification of 400-500 proteins per cell type using mass spectrometry (Brackeva et al. 2015). Vethe et al. identified 1874 proteins from S7 human iPSC derived insulin expressing beta cells using MS analysis (Vethe et al. 2017). Stützer and his group have provided the list of beta cell surface proteins based on previous small scale studies (Stützer, Esterházy, and Stoffel 2012) which contained 3693 proteins. Other studies have provided gene ((Martens et al. 2011), Eurodia; containing 7803 genes) and transcriptome (Kutla et al., 2009; containing 9850 transcripts) lists for beta cells. The lists of the proteins provided by Sacco and Waanders contain isoforms of similar proteins and proteins with sequence homology and therefore are comparable to our protein list (before being grouped) containing 8882 proteins. We, therefore, believe our identification rate is comparable to the other studies. However, since in our study labeling has been performed *in vivo* and the other beta cell proteomes are identified from *in vitro* studies, we believe our enriched proteome creates a more credible beta cell dataset in comparison to so far published data. We compared our protein list with gene, transcripts and proteins lists of beta cells/islets and

## Discussion

96% of the protein groups in our list were previously identified by other publications which is a good match to literature.

### 6.2.7. Effect of Harmine on beta cell proteome

We assessed the effect of Harmine on beta cell proteome and performed Reactome analysis on the list of downregulated proteins and compared the enriched pathways to literature (Annex Table 5-29). Harmine an indole alkaloid long known as an anti-tumor/ anti-cancer drug (Li et al. 2015) and recently its anti-diabetic effects have been studied (Belgardt and Lammert 2016). It is known to induce proliferation in ~2% of beta cells through inhibition of dual-specificity tyrosine-regulated kinase-1a (DYRK1A) and regulation by activated T cells (NFAT) family of transcription factors as likely mediators of human beta cell proliferation and differentiation) (P. Wang et al. 2015). We did not detect Dyrk1 and NFATs but MAO-A (enzyme monoamine oxidase A) which is another family member of Dyrk proteins and was found downregulated in our dataset. It is reported that DYRK1A can induce cell cycle exit in neural progenitors by different ways; promoting the nuclear export and degradation of the cell cycle activator Cyclin D1 (Yabut, Domogauer, and D'Arcangelo 2010), inducing the expression of the Cdk inhibitors p27 KIP1 and p21 CIP1, and promoting the stabilization of p27 KIP1 protein (Najas et al. 2015); we found SCF (Skp2)-mediated degradation of p27/p21 among the list of enriched pathways in our dataset. Although we had seen ki67 proliferation signals from 10 mg/kg Harmine treatment in our dose-finding experiment, the other proteins known to be involved in the proliferation of beta cells were not identified in Harmine dataset. This is probably due to their low abundance in the sample as harmine is known to cause an only modest increase in proliferation in ~1–1.5% of the beta cells engaged in the cell cycle (P. Wang et al. 2015). The highest percentage (Ki67+/Insulin+ cells) Wang reported with Harmine was 1.2%, while in the control group this ratio has been 0.2%. We applied the same dose of Harmine and the proportion of Ki67 positive cells over insulin positive cells were 1.8% for Harmine group and 0.6% for the control group. This means a maximum 1.2% of the beta cells are involved in proliferation which themselves count for 1-2% of pancreatic cells. We think that Harmine has an effect on beta cells but since only 1-2 % of these cells proliferate, the expression of the proteins linked to proliferation changes very slightly which is not possible to capture by mass spectrometric analysis. A solution for this would be to cross BioID mice with Cre lines directed to actively dividing cells such as Ki67<sup>cre/ERT2</sup> (Jax 029803) and label only for the proteins involved in proliferation.

On the other hand, we could identify around 60 enriched Reactome pathways in which Harmine is indicated as an effective medication for the treatment of cancer, tumors, as well as depression. Reus et al reported that reactive oxygen species (ROS) is found increased in the plasma of patients with bipolar disorder and major depression and may play an important role in the physiopathology of neurological and psychiatric diseases. ROS can cause cell damage, lipid peroxidation, and DNA modification. They evaluated the effect of administration of Harmine on these pathways in rat brain. Harmine in all doses (5-15 mg/kg) decreased the lipid and protein oxidation, induced DNA damage, and promoted antioxidant activities in prefrontal cortex and hippocampus of rat's brain (Réus et al. 2010). We found categories related to beta-oxidation among the enriched pathways meaning that the proteins involved in this mechanism are downregulated. Many categories related to DNA damage are also among the enriched pathways. Attenuation of oxidative stress (by Harmine) can result in a decrease in “Mitochondrial fatty acid beta-oxidation of saturated fatty acids” pathway (Douglas et al. 2016); we also found this category among the enriched (downregulated) mechanisms.

## Discussion

Harmine is reported to possess both anti-tumor and anti-nociceptive effects and to induce cell cycle arrest and mitochondrial pathway-mediated cellular apoptosis in SW620 cells via inhibition of the Akt and ERK signaling pathways (J. Liu et al. 2016); Liu demonstrates that harmine induces cell death and growth inhibition by alterations in cell cycle distribution. We found proteins involved in 21 cell cycle pathways downregulated as a possible result of cell arrest. Harmine also induces mitochondrial related cellular apoptosis by decreasing mitochondrial transmembrane potential and modulating the expression of Bcl-2 family proteins. We found three categories related to induction of apoptosis among the enriched categories. Harmine also decreased the levels of p-Akt and p-ERK in these cells. “Inactivation of ERKs” was one of the enriched mechanisms in our dataset. In other publications Harmine is reported to suppress the proliferation and migration of human ovarian cancer cells through inhibition of ERK/CREB pathway (X. F. Zhang et al. 2016) and to suppresses Ras–MAPK pathway (Ji et al. 2019); ERK/MAPK targets, MAPK6/MAPK4 signaling, ERKs, MAPK family signaling cascades, and MAPK1/MAPK3 signaling pathways were all among the significantly enriched Reactome pathways showing downregulation of their protein members. Nakagawa found out that Harmine results in mitochondrial dysfunction and biotransformation on isolated rat hepatocytes (Nakagawa et al. 2010) and similarly we found mitochondrial biogenesis proteins significantly downregulated in our list. In another study, Harmine was shown to be an inflammatory inhibitor through the suppression of NF- $\kappa$ B signaling in mouse macrophage RAW 264.7 cells and to decrease serum TNF- $\alpha$ , interleukin-1 $\beta$  (IL-1 $\beta$ ) and IL-6 levels (X. Liu et al. 2017). We also found TNFR2 non-canonical NF- $\kappa$ B pathway, interleukin-1 family signaling, interleukin-17 signaling, and cytokine signaling in immune system pathways all significantly enriched for the list of downregulated proteins.

In all abovementioned studies, Harmine has been shown to act as a potential therapeutic drug for cancer and tumor treatment. These findings support the fidelity of the cell-type-specific labeling of proteins affected by Harmine and potency of our enrichment method to capture the effect. In general, Harmine administration resulted in downregulation of 2518 protein groups (7111 proteins). Only 97 protein groups (252 proteins) were significantly upregulated among which none had a known link to Harmine treatment. The main upregulated proteins were involved in biotin transfer and metabolism which is probably due to their involvement in biotinylation.

### 6.2.8. Effect of STZ on beta cell proteome

Streptozotocin (STZ) is a glucosamine-nitrosourea compound with DNA alkylating properties widely used to induce hyperglycemia by specifically destroying the insulin-producing  $\beta$ -cells of the islets of Langerhans in experimental models of type I diabetes (Vinerean et al. 2011). We used this drug to evaluate the alterations of the beta cell proteome as a result of diabetes. 1274 groups (3563 proteins) were found downregulated and only 6 groups (6 proteins) were upregulated. So far very few middle-scale omics analyses have been performed to evaluate the effects of STZ on cells. Tonne and coworkers performed global gene expression profiling experiments and provided a list of 30 up-regulated and 50 down-regulated genes of pancreatic islets in mice during streptozotocin-induced  $\beta$ -cell damage and we found 6 proteins common with the list of downregulated genes. Tsai published a list of altered proteins in streptozotocin-induced diabetic rat kidney using mass spectrometry (Tsai et al. 2013) which contained 14 proteins and 4 were found in our dataset. Another group induced diabetes in rats by STZ and performed analysis of differentially expressed microRNAs and mRNAs related to mitochondrial dysfunction in dorsal root ganglia as a result of diabetes. They listed 175 and 104

## Discussion

genes as up and downregulated, respectively and 37 of the downregulated proteins were detected in our list of downregulated proteins (Guo et al. 2018).

The list of downregulated proteins was next submitted in GO enrichment analysis website for enriched Reactome pathways and compared to the literature for pathways involved in (STZ mediated) induction of diabetes and hyperglycemia; 86 Reactome pathways significantly enriched were linked to findings of other groups (Annex Table 5-30). STZ is a structural analogue of glucose and n-acetyl glucosamine and is delivered to pancreatic beta cells via GLUT 2 transporter and causes beta cell destruction by DNA fragmentation. One of the pathways associated with cell death is oxidative stress and free radical generation which are also reported to be induced in diabetic patients as a result of hyperglycemia (Volpe et al. 2018). In this stress condition, insulin signaling is impaired, resulting in the insulin resistance of the cell (Rains and Jain 2011). We could detect many of these mechanisms among the enriched pathways: reactive oxidative species (ROS) and reactive nitrogen species (RNS) production in phagocytes as well as insulin receptor recycling and insulin processing, signaling by insulin receptor, insulin receptor signalling cascade, regulation of insulin-like growth factor (igf) transport and uptake by insulin-like growth factor binding proteins (igfbps).

Another mechanism that results in cell death is the action of STZ as a nitric oxide donor in pancreatic cells which inhibits aconitase activity and leads to DNA damage (Eleazu et al. 2013). Also, hyperglycemia triggers DNA damage and p53 activation, which ultimately results in beta cell death and late-onset diabetes (Kung and Murphy 2016). We found production of reactive nitrogen species (RNS) as well as 5 pathways related to DNA damage among the enriched categories. In diabetes, antigen-presenting cells process certain beta cell proteins which act as autoantigens. Altered function of antigen-presenting cells in type 1 diabetes has been reported by previous studies (Creusot, Postigo-Fernandez, and Teteloshvili 2018); we found six of the pathways related to antigen presentation enriched in our list of proteins with altered expression.

Several other mechanisms previously studied including mitochondrial dysfunction, cell cycle arrest, inflammatory processes, NFAT activation, etc were characterized in our list of enriched pathways. Connection of the enriched proteins affected by STZ with previously published data was another proof that this mouse model can be an effective tool to capture the effect of external stimuli on a specific cell type which itself counts for only 1-2% of the cells in the pancreas.

### 6.2.9. Possible mistakes in data analysis

Plenty of proteins were detected as downregulated as a result of STZ and Harmine administration. It is very unlikely that a major population of the proteome is perturbed as a result of administration of a drug. It may be accepted in the case of STZ, as it leads to beta cell dysfunction and apoptosis and subsequently a possible downregulation in large scales; however, this is not expected from Harmine to result in such a proteome wide alteration in expression. It can be that due to flaws in data analysis, our assumptions about the population of proteins affected by the drugs are incorrect. It is possible that fix drug-effect cut-offs are not applicable for proper interpretation of the data.

“Harmine or STZ”/biotin log<sub>2</sub> ratios reflect the effect of biotinylation as well as drug administration. Biotin/biotin log<sub>2</sub> ratios on the other hand, reflect the alterations of proteome as a result of biotinylation. One possible approach to perform data analysis in a different way would be to overlay the distribution of biotinylated (biotin/biotin log<sub>2</sub> ratios) and drug affected (“Harmine or STZ”/biotin log<sub>2</sub> ratios) proteomes on each other (by superimposing medians of both histograms) and exclude the biotinylation effect by subtracting the overlapping area from the “Harmine or STZ”/biotin log<sub>2</sub> ratios distribution. The data points outside of the

## Discussion

biotin/biotin ratio distribution (on right and left sides of this histogram), thus present only in “Harmine or STZ”/biotin log<sub>2</sub> ratio distribution, can be accepted as up- and down-regulated as a result of the drug administration. However, as this requires sophisticated bioinformatics analysis to accurately overlay the histograms and subtract the proteomes, we were not yet able to examine this strategy.

After data analysis, we performed GO-based Reactome pathway enrichment analysis. We compared the down-regulated proteomes with all biological pathways in the cell to identify the enriched pathways. Another approach that may allow making more reliable assumptions about the enriched mechanisms as a result of the drug administration is to introduce a background proteome to GO-based Reactome pathways enrichment analysis instead of comparing our altered proteome to all the mechanisms in the cell (default background of Reactome analysis). This background proteome can be the list of biotinylated proteins; this helps subtracting the effect of biotinylation from drug-effect and draw more realistic conclusions about the enriched/perturbed pathways as a result of drug administration.

Due to time constraints, these two analytical approaches were not examined during the course of current research, but will be evaluated in near future.

### **6.2.10. Comparison of BioID cell-type-specific approach with other in-vivo labeling approaches**

We showed the first application of non-specific BioID (labeling all proteins in cytosol and nucleus) in mice for cell-type-specific proteomics. So far the specific type (BirA; labeling specifically the AviTag peptide) has been used in a mouse model that ubiquitously expresses BirA from the Rosa26 promoter. Upon crossing with any transgenic mouse expressing an Avi-tagged protein, BirA specifically adds biotin to the protein of interest (Driegen et al. 2005). The non-specific BioID has also been exploited in two other studies to identify protein-protein interactions surrounding c-MYC oncoprotein in tumour xenografts of mice (Dingar et al. 2015) and to analyze synaptic complexes in the mouse brain (Uezu et al. 2016). In both studies, BioID has been utilized to discover interaction partners of a fusion protein. Here we reported the adaptation of BioID method to the enrichment of all the proteins localized in cytoplasm and nucleus of a specific cell type in mice.

Two other groups have shown successful cell-type-specific labeling and enrichment using the unnatural amino acids methionine surrogate azidonorleucine (ANL) and N $\epsilon$ - (propargyloxycarbonyl)-l-lysine (AlkK). These amino acids are recognized with a mutant tRNA synthetase and incorporated into nascent polypeptides which can be tagged with click reaction and affinity purified (Alvarez-Castelao et al. 2017; Krogager et al. 2018). In these approaches, the mice are treated with unnatural chemicals ANL (60 mM) and AlkK (30 mg/ml) for 21 and 7 days in drinking water, respectively which can possibly introduce variation and bias to the biological system (Alvarez-Castelao et al. 2017; Krogager et al. 2018). In our approach instead of artificial substances, biotin is administered to animals (0.22 mg/ml in water for 7 days) which is a natural vitamin required for fatty acid biosynthesis. Biotin is not known to be toxic to cells even at high doses. Koutsikos et al., supplied biotin at doses up to 5 mg/day through intraperitoneal injection for two years to control biotin dosage in people with biotin metabolism disorder. They observed no toxicity or adverse effects on bioavailability associated with high doses of biotin in these patients (Koutsikos, Agroyannis, and Tzanatos-Exarchou 1990). While biotin itself does not harm the organism, addition of biotin to all proteins in cytosol and nucleus (biotinylation) in an unspecific manner may impact proteins, structure, enzymatic activity, or half-life. We did not detect any morphological differences between the biotinylated and control tissue samples in immunofluorescence experiments. We also verified by microscopy and Western blot experiments that detectable biotinylation occurs

## Discussion

only after the administration of external biotin and not during the life of animals by consumption of the biotin present in mouse food. This is probably due to activity level of BioID that is not sufficiently high to enable non-specific biotinylation of proteins using the biotin in food. It can be inferred that biotinylation is only significant in access to external biotin in water and for one week which is probably not leading to structural changes in proteins. We tried to investigate if one week of biotinylation does lead to alterations in protein expression by comparing abundances of the proteins between biotinylated and control mice tissues at whole proteome level (before enrichment) and detected less than 10 regulated proteins. Based on these findings, we believe that labeling of proteins with biotin is advantageous over biorthogonal labeling as it reduces the likelihood of unwanted side-effects.

Another advantage of using biotin is its natural delivery routes to tissues of interest. Biotin at low concentrations (less than 5  $\mu\text{M}$ ) is delivered to cells primarily by  $\text{Na}^+$ -dependent multivitamin transporter (SMVT1), which is expressed ubiquitously in various tissues. In higher concentrations (above 25  $\mu\text{M}$ ), biotin can diffuse passively across the membrane (Polyak 2015). Based on the availability of these entry routes, biotin delivery to tissue is possible through animal water or food. Our results also confirmed the successful delivery of biotin to the examined tissues. However, this method of transport is a source of variation in labeling among the individual animals as food/water intake may vary animal to animal. This is also a problem associated with SORT and BONCAT methods as they provide amino acids in the drinking water of mice (Krogager et al. 2018; Alvarez-Castelao et al. 2017).

In SORT method using AlkK as the unnatural amino acid, adeno associated virus was used for introduction of the mutated tRNA synthetase and therefore the system reproducibility is dependent on transduction efficiency and accuracy of injection (Krogager et al. 2018). Our genetic system should be more reproducible as the enzyme labeling the proteins is incorporated into the genome of mice and has shown comparable expression in mice with similar genotypes. However, to ensure comparable expression and reproducible results, similar genotype, age, and even sex of the mice should be used for the experiments which can be difficult and time-consuming to generate.

While SORT and BONCAT approaches are restricted to labeling of newly synthesized proteins (Krogager et al. 2018; Alvarez-Castelao et al. 2017), BioID approach allows tagging of all the proteins (not only the newly synthesized) in the cytoplasm and nucleus of the cells. However, BioID may not be as accessible in the secretory pathway, thus restricting its labeling to these two compartments. To enable labeling of proteins by BioID in specific organelles, BioID needs to be transported into cellular compartments e.g. through viral transduction; our current transgenic mice models are not applicable to such studies.

Another limitation of BioID approach is the relatively long incubation times required for efficient labeling; this does not permit the capture of proteome snapshots (as possible by APEX or TurboID) but rather provides a history of proteins associations over time. The recent generations of BioID, TurboID has been applied to living organisms including yeast, *Drosophila*, and *C. elegans*. (Branon et al. 2018). While original BioID captures protein associations over time and reports a history of events, the TurboID enable an analysis of events occurring over short time periods of 10 minutes. It is questionable, however, how far the fast biotinylation is compatible with mice. Branon and his group report a decrease in fly viability and size when TurboID is expressed constitutively and no exogenous biotin is supplied. They hypothesize that in this case TurboID consumes all the biotin and results in a biotin starvation condition. This suggests that the expression of TurboID in mice can lead to a similar issue and since the control of mouse protein expression is also a quite slow process the classical BioID is probably a more compatible tool for cell-type-specific analysis in mice.

## Discussion

This group also reports in many contexts a small degree of labeling before the external biotin is supplied for cells/organisms expressing TurboID. This again suggests that TurboID is very active and utilizes the biotin present in media/food (Branon et al. 2018). Biotin is an important vitamin for complex animals and its deficiency can lead to physiological perturbations in mice (Subramanya et al. 2011). We and Branon both show in our results that background biotinylation for BioID is not detected (Branon et al. 2018). This further confirms that BioID with lower catalytic activity is more applicable to complex animal models than TurboID. Branon and his team were not able to see any activity of the classical BioID in yeast, flies, or worms and they suggest the body temperature of these organisms (30, 25, and 20 °C, respectively), are below the optimum temperature for BioID activity that is 37 °C. TurboID however, was evolved in yeast (30°C) and was shown to be compatible with flies and worms as well (Branon et al. 2018). We, therefore, can speculate that TurboID is not active in mice and only BioID can be applied to mice.

We showed successful application of BioID to *in vivo* cell-type-specific labeling analysis. We could enrich and identify the proteome of beta cells of pancreas which constitute 1-2% of the pancreas volume. For these analyses the Col1a1-BioID mouse was used which in comparison to Rosa-CAG-BioID mouse showed a weaker expression profile in tissues. I believe the application of the strategy can be extended to specific cell types of other tissues, specifically by adoption of the other mouse which acquires a stronger and more efficient promoter that in combination with CAG enhancer can probably result in more consistent expression in cell types of interest. The Rosa-CAG-BioID mouse is currently in breedings with cre lines specific for different cell types of the brain as part of another project and is aimed to be applied to gaucher disease cell-type-specific analysis.

## 7. Conclusion

In the first part of the thesis, we were able to establish an unbiased approach for determination of a cut-off to distinguish between biotinylated and non-specific proteins in biotin enrichment analysis. We determined this cut-off based on the ratio distribution of control samples (null distribution). Our results showed that a cut-off determined from a separate null experiment is not applicable to individual biotin enrichment analysis. For proper cut-off determination analysis, two control samples have to be mixed with the biotinylated sample before incubation on beads. This can result in the creation of a bimodal histogram with two clusters of non-specific proteins (log<sub>2</sub> ratios around 0) and biotinylated proteins (log<sub>2</sub> ratios > 0). It is then possible to define a cut-off on 95<sup>th</sup> percentile of the control/control ratios and apply it on biotin/control ratios. We successfully applied this method on the analysis of proteomes labeled by BioID organelle-targeting constructs using the 3plex SILAC approach.

In the second part, the application of BioID technique was successfully expanded to *in vivo* cell-type-specific analyses. Two mice with BioID inserted in Rosa26 and Col1a1 loci were generated. Rosa26-BioID was shown to be a more versatile tool with BioID expression and biotinylation in a wider range of cell populations. We however decided to continue with Col1a1-BioID for beta-cell-specific analysis as the expression and biotinylation in this mouse were restricted to beta cells of pancreatic islets. Before beta-cell-specific analysis, we performed method development and could improve both identification and enrichment rates by ~8 and ~5 fold, respectively. We next enriched for the biotinylated proteins in tissues of Col1a1 and used two control samples for determination of biotinylation cut-off. We think the inclusion of a second control sample has the benefit (besides allowing determination of a cut-off) of increasing the chance for nonspecific binding proteins (with lower affinity to beads) to be identified which in turn contributes to better quantification and enrichment rates for biotinylated proteins. We were able for the first time to apply BioID to cell-type-specific *in vivo* labeling of beta cells using Ins1Cre Col1a1 tTA mice and enrich the proteome of this specific cell type. The high rate of matching between our beta cell proteome and literature, as well as depletion of the dataset of the proteins specific to other pancreatic cell types (based on GO analysis), confirmed cell-type-specific labeling and potency of our enrichment/cut-off determination techniques for capturing the biotinylation and detecting the true positive proteins. We also investigated the effect of pro- and anti-diabetic drugs Streptozotocin and Harmine and found enriched mechanisms previously connected to the impact of these two drugs by other groups. We believe these 2 BioID model organisms can be of great use for cell-type-specific proteomics analysis especially to monitor the effect of diseases which involve only specific cell types of tissues.

## 8. List of abbreviations

°C	Degree Celsius
Con	Contamination
REV	Reverse hits
SITE	The hits only identified by site
ACN	Acetonitrile
Ab	antibody
AcOH	Acetic acid
AHA	azidohomoalanine
AlkK	N $\epsilon$ - (propargyloxycarbonyl)-l-lysine
AMP	Adenosine Monophosphate
ANL	Bioorthogonal non-canonical amino acids azidonorleucine
APEX	Ascorbate peroxidase
APS	Ammonium persulfate
AQUA	Absolute quantification
Arg	Arginine
ATP	Adenosine triphosphate
BAC	Bacterial Artificial Chromosome
BONCAT	Bioorthogonal non-canonical amino acid tagging
bp	Base pair
BRP	Basic reversed-phase
BSA	Bovine serum albumin
CID	Collision-induced dissociation
CMV	Cytomegalovirus
Col1a1	Collagen, type I, alpha 1
Conc.	Concentration
Cre	Cre recombinase
Cyto.Nucl.	Cytosol Nucleus
DAPI	4',6-diamidin-2-phenylindol
DDA	Data-dependent acquisition
ddH <sub>2</sub> O	Double distilled water
DMEM	Dulbecco's modified eagle medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
Dox	Doxycycline
DPBS	Dulbecco's phosphate buffered saline
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmatic reticulum
ERLIC	Electrostatic repulsion-hydrophilic interaction chromatography
ESI	Electrospray ionization
EtOH	Ethanol
FA	Formic acid
FACs	Fluorescence activated cell sorting
FCS	Fetal calf serum
FDR	False discovery rate
FPR	False postive rate
FUNCAT	Fluorescent noncanonical amino acid tagging
g	Gram
GO	Gene ontology
h	Hour
HBSS	Hank's balanced salt solution

## List of abbreviations

HCD	Higher-energy collisional dissociation
HEK 293T	Human embryonic kidney cells T cells
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High pressure liquid chromatography
HRP	Horseradish peroxidase
ICAT	Isotope coded affinity tag
ICC	Immunocytochemistry
ID	Inner diameter
IF	Immunofluorescence
IPG	Immobilized pH gradient
iPSC	Nonviral induced pluripotent stem cell
iTRAQ	Isobaric tags for relative and absolute quantitation
kDa	Kilodalton
LCM	Laser capture microdissection
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
LOPIT	Localization of organelle proteins by isotope tagging
LTQ	Linear trap quadrupole
Lys	Lysin
M	Molar
m/z	Mass-to-charge ratio
MALDI	Matrix-assisted laser desorption/ionization
MeOH	Methanol
mg	Milligram
min	Minute
Mito	Mitochondria
ml	Milliliter
mM	Millimolar
mm	Millimeter
mRNA	Messenger RNA
MS	Mass spectrometry
mTb	MiniTurbo
MS	Mass spectrometry
NA	Not applicable
NaCl	Sodium chloride
Nano-LC	Nanoscale liquid chromatography
NHS	N-hydroxysuccinimide
nl	Nanoliter
nM	Nanomolar
Nucl.	Nucleus
OCT	Optical Coherence Tomography
OD	Outer Diameter
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEG	Poly Ethylen Glycol
PFA	Paraformaldehyde
PPI	Protein-protein interaction
PTM	Post-translational modification
QConCAT	Quantification concatamer
rb	Rabbit
Ripa	Radioimmunoprecipitation assay buffer
RNA	Ribonucleic acid
RP	Reversed-phase
RPM	Revolutions per minute
RSV	Respiratory syncytial virus
RT	Room temperature
rtTA	Reverse tetracycline-controlled transactivator

## List of abbreviations

SDS	Sodium dodecyl sulfate
SILAC	Stable Isotope Labeling by Amino acids in Cell Culture
SORT	Stochastic orthogonal recoding of translation
StageTips	Stop-and-go-extraction tips
STP	Stop
STZ	Streptozotocin
SU9	Subunit 9 of the oligomycin-sensitive mitochondrial ATPase
TBS	Tris-buffered saline
Tc	Tetracycline
TEAB	Triethylammonium bicarbonate
TEMED	Tetramethylethylenediamine
Temp.	Temperature
TetO	Tet Operator
TFA	Trifluoroacetic acid
TurboID	TbID
T <sub>m</sub>	Melting temperature
TMT	Tandem mass tag
TOF	Time-of-Flight
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
Tris-HCl	Tris(hydroxymethyl)-aminomethane hydrochloride
TPR	True Positive Rate
tRNA	Transfer RNA
tTA	Tetracycline-controlled transactivator
TX100	Triton X-100
U	Unit
UPLC	Ultra-performance liquid chromatography
V	Volume
V	Voltage
VP16	Virion protein 16
WB	Western blot
WT	Wild type
x g	Earth's gravitational acceleration
β-ME	2-Sulfanylethan-1-ol
μg	Microgram
μl	Microliter
μM	Micromolar

## 9. List of figures

Figure 2-1. Cre/lox system. ....	5
Figure 2-2. Tet-on and Tet-off systems. ....	6
Figure 2-3. Cell-type-specific labeling approaches. ....	7
Figure 2-4. Proximity labeling methods. ....	9
Figure 2-5. Bottom up mass spectrometry-based proteomics workflow.....	11
Figure 2-6. Stages of incorporation of stable isotope labels in quantitative proteomics experiments.....	13
Figure 5-1. 95% Cut-off determination strategy.....	41
Figure 5-2. Schematic workflow and ratio distribution for the null experiment.....	42
Figure 5-3. Determination of the 95% cut-off for biotinylation. ....	43
Figure 5-4. Potential false positive hits in the null experiment.....	43
Figure 5-5. Confirmation of BioID expression and biotinylation by Western blot.....	45
Figure 5-6. Confirmation of correct localization of BioID organelle constructs by microscopy. ....	46
Figure 5-7. Identification and quantification efficiencies by BioID organelle constructs (including one control sample).....	47
Figure 5-8. Distribution of $\log_2$ L/H ratios in organelle proteomics datasets and overlay of $\log_2$ L/H ratio distributions in background (null) and nuclear BioID datasets. ....	48
Figure 5-9. Ratio distribution for nuclear and non-nuclear proteins in Nucl.-BioID and null experiment datasets.....	49
Figure 5-10. Comparison between intensities of background proteins in Nucl.-BioID vs null experiment datasets.....	50
Figure 5-11. Youden index and ROC curve analysis on Nucl.-BioID dataset.....	51
Figure 5-12. Comparison between the number of nuclear proteins passing null experiment cut-off and the “slide approach” cut-off.....	51
Figure 5-13. Schematics of four different strategies for combination of biotinylated and control proteins.....	53
Figure 5-14. Identification and quantification efficiencies for each combination method.....	54
Figure 5-15. “Slide approach” cut-off for biotinylation in combination methods 1 and 4.....	55
Figure 5-16. Efficiency of different combination methods for purification of biotinylated proteins. ....	56
Figure 5-17. Intensity analysis for SILAC channels in different combination methods. ....	57
Figure 5-18. Distribution of $\log_2$ L*/M and H/M ratios values in different combination methods.....	59
Figure 5-19. Organelle proteomics (including two control samples).....	59
Figure 5-20. Identification and quantification efficiency for each organelle BioID dataset (including two control samples).....	60
Figure 5-21. Comparison of quantification rates between organelle proteomics experiments including one and two control populations.....	61
Figure 5-22. Efficiency of different BioID constructs for purification of biotinylated proteins (including two control samples)....	62
Figure 5-23. Organelle proteomes vs literature. ....	63
Figure 5-24. Distribution of $\log_2$ L/M ratios and distance analysis for common biotinylated proteins in nuclear and mitochondrial datasets.....	64
Figure 5-25. Localization of BioID in cytosol and nucleus.....	65
Figure 5-26. Generation and mechanism of action in Col1a1-BioID mouse (mouse 1).....	66
Figure 5-27. Generation and mechanism of action in Rosa26-CAG-BioID mouse (mouse 2).....	66
Figure 5-28. Generation of mice with ubiquitous BioID expression. ....	67
Figure 5-29. Evaluation of functionality and leakiness of the genetic system-Western blot. ....	68
Figure 5-30. Evaluation of internal biotinylation in different organs of Rosa26-CAG-BioID mouse (mouse 2) -Western blot.....	69
Figure 5-31. BioID expression and biotinylation in Tissues-Western blot. ....	70
Figure 5-32. BioID expression and biotinylation in mouse 1 tissues - Immunofluorescence. ....	71
Figure 5-33. BioID expression and biotinylation in mouse 2 tissues - Immunofluorescence. ....	72
Figure 5-34. BioID expression and biotinylation in pancreatic islets of mouse 1. ....	73
Figure 5-35. Schematic of <i>in-vivo</i> proteomics workflow using dimethyl for quantification. ....	74
Figure 5-36. Cut-off for drug effect.....	75
Figure 5-37. Cut-off determination for biotinylation using control pancreas samples (null experiment).....	76
Figure 5-38. Quality of the 6 pancreas datasets (including one control sample).....	77
Figure 5-39. Enrichment analysis and ratio distribution in tissue datasets (including one control sample).....	78
Figure 5-40. Comparison between enrichment methods (method development).....	80

Figure 5-41. Comparison between tissue datasets including one and two control samples (2plex and 3plex) in terms of identification and quantification.....	81
Figure 5-42. Enrichment analysis and ratio distribution for tissue samples (including two control samples).....	83
Figure 5-43. Schematic for generation of beta cell specific mice and Dox titration experiment. ....	84
Figure 5-44. Dox duration titration experiment.....	86
Figure 5-45. Dox duration titration experiment: no Dox vs 3 weeks without Dox.....	86
Figure 5-46. Workflow for treatment of mice for beta cell proteomics.....	87
Figure 5-47. comparison between intensities of TMT channels in whole proteome and beta cell enriched proteome.....	88
Figure 5-48. Perturbed proteins in the whole proteome experiment.....	89
Figure 5-49. Enrichment analysis on beta cell proteome and comparison to literature.....	90
Figure 5-50. Harmine dose test evaluation using Ki67 staining. ....	92
Figure 5-51. Workflow for treatment of mice with Harmine.....	92
Figure 5-52. Workflow for treatment of mice with STZ.....	93
Figure 5-53. Blood glucose concentration in STZ, Biotin, and control groups. ....	93
Figure 5-54. Data analysis for identification of Harmine-perturbed proteome.....	95
Figure 5-55. Data analysis for identification of STZ-perturbed proteome.....	96
Figure 1-1. Distribution of nuclear and mitochondrial matrix proteomes identified by BioID.....	101
Figure 1-2. Distance correlation across organelles analyzed by LOPIT approach. ....	104
Figure 1-3. Induction of EGFP expression in Coll1a1-EGFP mice organs. ....	106
Figure 1-4. Comparison between EIIa-Cre/Rosa-CAG-Ai14 mouse (left, Cre+) with that in tissues from a control Rosa-CAG-Ai14 mouse (right, Cre-). ....	106

## 10. List of tables

Table 3-1. list of chemicals/reagents .....	15
Table 3-2. list of Enzymes.....	16
Table 3-3. List of cell lines.....	17
Table 3-4. List of mice .....	17
Table 3-5. List of primers.....	17
Table 3-6. List of antibodies.....	19
Table 3-7. List of secondary antibodies and chemicals .....	20
Table 3-8. List of instruments and consumables .....	20
Table 3-9. List of software and online tools .....	22
Table 3-10. List of ready to use reaction kits .....	22
Table 3-11. List of drugs and treatments .....	23
Table 4-12. Pipetting instructions for PCR mediated by Phusion DNA polymerase.....	25
Table 4-13. Pipetting instructions for PCR mediated by Taq DNA polymerase .....	25
Table 4-14. PCR cycling program.....	25
Table 4-15. Annealing temperature, extension time, and primers used for genotyping PCR .....	26
Table 4-16. Constituents of resolving and stacking gels for SDS gel electrophoresis .....	28
Table 4-17. OFFGEL program for 12 fractions.....	32
Table 4-18. Hp-RP non-linear gradient. ....	33
Table 4-19. Concatenation schema for 12 fractions of Hp-RP .....	33
Table 4-20. Concatenation schema for 24 fractions of Hp-RP .....	34
Table 5-21. Comparison of large scale beta-cell/islet omics data available in the literature and our beta-cell proteome .....	89
Table 5-22. Harmine doses and percent of cells with proliferation signal.....	91

## 11. List of annex tables

Annex Table 5-1	Null experiment (in vitro) - MaxQuant proteinGroups
Annex Table 5-2	Null experiment (in vitro)-Perseus Analysis (biotinylation cut-off determination)
Annex Table 5-3	Organelle proteomics including one control sample- MaxQuant proteinGroups
Annex Table 5-4	Organelle proteomics including one control sample- Perseus Analysis
Annex Table 5-5	Different combination methods - MaxQuant proteinGroups
Annex Table 5-6	Different combination methods - Perseus Analysis
Annex Table 5-7	Organelle proteomics including two control sample- MaxQuant proteinGroups
Annex Table 5-8	Organelle proteomics including two control sample- Perseus Analysis
Annex Table 5-9	Drug-effect cut-off determination experiment- MaxQuant proteinGroups
Annex Table 5-10	Drug-effect cut-off determination experiment- Perseus Analysis
Annex Table 5-11	Null experiment (in vivo)- MaxQuant proteinGroups
Annex Table 5-12	Null experiment (in vivo)-Perseus Analysis (biotinylation cut-off determination)
Annex Table 5-13	Tissue proteomics including one control sample - MaxQuant proteinGroups
Annex Table 5-14	Tissue proteomics including one control sample- Perseus Analysis
Annex Table 5-15	Method development - MaxQuant proteinGroups
Annex Table 5-16	Method development - Data analysis
Annex Table 5-17	Tissue proteomics including two control sample - MaxQuant proteinGroups
Annex Table 5-18	Tissue proteomics including two control sample- Perseus Analysis
Annex Table 5-19	Pancreas whole proteome - PD proteins
Annex Table 5-20	Pancreas whole proteome- Perseus Analysis
Annex Table 5-21	Beta cell proteome- PD proteins
Annex Table 5-22	Beta cell proteome- Perseus Analysis
Annex Table 5-23	Harmine proteome - PD proteins
Annex Table 5-24	Harmine proteome - Perseus Analysis
Annex Table 5-25	Harmine Reactome pathways
Annex Table 5-26	Streptozotocin proteome - PD proteins
Annex Table 5-27	Streptozotocin proteome- Perseus Analysis
Annex Table 5-28	Streptozotocin Reactome pathways

## 12. References

- Aebersold, Ruedi, and Matthias Mann. 2016. "Mass-Spectrometric Exploration of Proteome Structure and Function." *Nature*.
- Agrawal, Ganesh Kumar et al. 2011. "Plant Organelle Proteomics: Collaborating for Optimal Cell Function." *Mass Spectrometry Reviews*.
- Ahmadi, Shiva, and Dominic Winter. 2018. "Identification of Poly(Ethylene Glycol) and Poly(Ethylene Glycol)-Based Detergents Using Peptide Search Engines." *Analytical Chemistry* 90(11): 6594–6600.
- Alvarez-Castelao, Beatriz et al. 2017. "Cell-Type-Specific Metabolic Labeling of Nascent Proteomes in Vivo." *Nature Biotechnology* 35(12): 1196–1201. <http://dx.doi.org/10.1038/nbt.4016>.
- Alvarez-Castelao, Beatriz, Christoph T. Schanzenbächer, Julian D. Langer, and Erin M. Schuman. 2019. "Cell-Type-Specific Metabolic Labeling, Detection and Identification of Nascent Proteomes in Vivo." *Nature Protocols* 14(2): 556–75.
- Andersen, Jens S. et al. 2003. "Proteomic Characterization of the Human Centrosome by Protein Correlation Profiling." *Nature* 426(6966): 570–74.
- Andersen, Jens S., and Matthias Mann. 2006. "Organellar Proteomics: Turning Inventories into Insights." *EMBO Reports* 7(9): 874–79.
- Bansal, Mukesh et al. 2015. "Direct ChIP-Seq Significance Analysis Improves Target Prediction." *BMC Genomics*.
- Bantscheff, Marcus et al. 2007. "Quantitative Mass Spectrometry in Proteomics: A Critical Review." *Analytical and Bioanalytical Chemistry* 389(4): 1017–31.
- Barshop, William D. et al. 2019. "Chemical Derivatization of Affinity Matrices Provides Protection from Tryptic Proteolysis." *Journal of Proteome Research*.
- Beard, Caroline et al. 2006. "Efficient Method to Generate Single-Copy Transgenic Mice by Site-Specific Integration in Embryonic Stem Cells." *Genesis (New York, N.Y. : 2000)* 44(1): 23–28. <http://www.ncbi.nlm.nih.gov/pubmed/16400644> (November 3, 2014).
- Belgardt, Bengt Frederik, and Eckhard Lammert. 2016. "DYRK1A: A Promising Drug Target for Islet Transplant-Based Diabetes Therapies." *Diabetes*.
- Bickel, David R. 2011. "Estimating the Null Distribution to Adjust Observed Confidence Levels for Genome-Scale Screening." *Biometrics*.
- Boersema, Paul J. et al. 2009. "Multiplex Peptide Stable Isotope Dimethyl Labeling for Quantitative Proteomics." *Nature Protocols* 4(4): 484–94.
- Brackeva, B., G. Kramer, J. P.C. Vissers, and G. A. Martens. 2015. "Quantitative Proteomics of Rat and Human Pancreatic Beta Cells." *Data in Brief* 3: 234–39. <http://dx.doi.org/10.1016/j.dib.2015.02.019>.
- Branon, Tess C et al. 2018. "Efficient Proximity Labeling in Living Cells and Organisms with TurboID HHS Public Access Author Manuscript." *Nat Biotechnol* 36(9): 880–87. [http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms).
- Breuzer, Lionel et al. 2004. "Proteomics of Endoplasmic Reticulum-Golgi Intermediate Compartment (ERGIC) Membranes from Brefeldin A-Treated HepG2 Cells Identifies ERGIC-32, a New Cycling Protein That Interacts with Human Erv46." *Journal of Biological Chemistry* 279(45): 47242–53.
- Brownridge, Philip et al. 2011. "Global Absolute Quantification of a Proteome: Challenges in the Deployment of a QconCAT Strategy." *Proteomics* 11(15): 2957–70.
- Brun, Virginie et al. 2007. "Isotope-Labeled Protein Standards: Toward Absolute Quantitative Proteomics." *Molecular and Cellular Proteomics* 6(12): 2139–49.
- Burch, Tanya C. et al. 2015. "Proteomic Analysis of Disease Stratified Human Pancreas Tissue Indicates Unique Signature of Type 1 Diabetes." *PLoS ONE*.
- Calvo, Sarah E., Karl R. Clauser, and Vamsi K. Mootha. 2016. "MitoCarta2.0: An Updated Inventory of Mammalian Mitochondrial Proteins." *Nucleic Acids Research*.
- Cayuela, Luis, Nicholas J. Gotelli, and Robert K. Colwell. 2015. "Ecological and Biogeographic Null Hypotheses for Comparing Rarefaction Curves." *Ecological Monographs*.
- Chapman-Smith, a et al. 2001. "The C-Terminal Domain of Biotin Protein Ligase from E. Coli Is Required for Catalytic Activity." *Protein science : a publication of the Protein Society* 10(12): 2608–17.
- Cheah, Joleen S., and Soichiro Yamada. 2017. "A Simple Elution Strategy for Biotinylated Proteins Bound to Streptavidin Conjugated Beads Using Excess Biotin and Heat." *Biochemical and Biophysical Research Communications*.
- Chen, Chiao-Lin et al. 2015. "Proteomic Mapping in Live Drosophila Tissues Using an Engineered Ascorbate Peroxidase." *Proceedings of the National Academy of Sciences of the United States of America* 112(39): 1–6. <http://www.ncbi.nlm.nih.gov/pubmed/26362788>.
- Chen, Chunguang et al. 2017. "Human Beta Cell Mass and Function in Diabetes: Recent Advances in Knowledge and Technologies to Understand Disease Pathogenesis." *Molecular Metabolism* 6(9): 943–57. <http://dx.doi.org/10.1016/j.molmet.2017.06.019>.
- Chitikova, Zhanna, and Florian A. Steiner. 2016. "Cell Type-Specific Epigenome Profiling Using Affinity-Purified Nuclei." *Genesis* 54(4): 160–69.
- Cijsouw, Tony et al. 2018. "Mapping the Proteome of the Synaptic Cleft through Proximity Labeling Reveals New Cleft Proteins." *Proteomes* 6(4).
- Clardy, Susan M. et al. 2015. "Rapid, High Efficiency Isolation of Pancreatic  $\beta$ -Cells." *Scientific Reports*.

## References

- Cox BC et al. 2012. "Conditional gene expression in the mouse inner ear using cre-loxP." *JARO - J Assoc Res Otolaryngol*. doi:10.1007/s10162-012-0324-5
- Creusot, Rémi J., Jorge Postigo-Fernandez, and Nato Teteloshvili. 2018. "Altered Function of Antigen-Presenting Cells in Type 1 Diabetes: A Challenge for Antigen-Specific Immunotherapy?" *Diabetes*.
- Dingar, Dharmendra et al. 2015. "BioID Identifies Novel C-MYC Interacting Partners in Cultured Cells and Xenograft Tumors." *Journal of Proteomics* 118: 95–111. <http://dx.doi.org/10.1016/j.jprot.2014.09.029>.
- Douglas, Donna N. et al. 2016. "Oxidative Stress Attenuates Lipid Synthesis and Increases Mitochondrial Fatty Acid Oxidation in Hepatoma Cells Infected with Hepatitis C Virus." *Journal of Biological Chemistry*.
- Dow, Lukas E. et al. 2014. "Conditional Reverse Tet-Transactivator Mouse Strains for the Efficient Induction of Tre-Regulated Transgenes in Mice." *PLoS ONE*.
- Driegen, Siska et al. 2005. "A Generic Tool for Biotinylation of Tagged Proteins in Transgenic Mice." *Transgenic Research*.
- Du, Shubo et al. 2017. "Cell Type-Selective Imaging and Profiling of Newly Synthesized Proteomes by Using Puromycin Analogues." *Chemical Communications*.
- Dundas, Christopher M., Daniel Demonte, and Sheldon Park. 2013. "Streptavidin-Biotin Technology: Improvements and Innovations in Chemical and Biological Applications." *Applied Microbiology and Biotechnology* 97(21): 9343–53.
- Dunkley, T. P. J. et al. 2004. "Localization of Organelle Proteins by Isotope Tagging (LOPIT)." *Molecular & Cellular Proteomics* 3(11): 1128–34. <http://www.mcponline.org/lookup/doi/10.1074/mcp.T400009-MCP200>.
- Eleazu, Chinedum O., Kate C. Eleazu, Sonia Chukwuma, and Udeme N. Essien. 2013. "Review of the Mechanism of Cell Death Resulting from Streptozotocin Challenge in Experimental Animals, Its Practical Use and Potential Risk to Humans." *Journal of Diabetes and Metabolic Disorders*.
- Elias, Joshua E, and Steven P Gygi. 2010. "Target-Decoy Search Strategy for Mass." *Bioinformatics book from Anton* 604(2): 55–71.
- Elliott, Thomas S. et al. 2014. "Proteome Labeling and Protein Identification in Specific Tissues and at Specific Developmental Stages in an Animal." *Nature Biotechnology* 32(5): 465–72.
- Fang, Lanyun, Zhaozi Lü, Hui Wei, and Erkang Wang. 2008. "Quantitative Electrochemiluminescence Detection of Proteins: Avidin-Based Sensor and Tris(2,2'-Bipyridine) Ruthenium(II) Label." *Biosensors and Bioelectronics* 23(11): 1645–51.
- Fay, David S., and Ken Gerow. 2013. "A Biologist's Guide to Statistical Thinking and Analysis." *WormBook: the online review of C. elegans biology*.
- Feist, Peter, and Amanda B. Hummon. 2015. "Proteomic Challenges: Sample Preparation Techniques for Microgram-Quantity Protein Analysis from Biological Samples." *International Journal of Molecular Sciences* 16(2): 3537–63.
- Fenn, John B. et al. 1989. "Electrospray Ionization for Mass Spectrometry of Large Biomolecules." *Science*.
- Firat-Karalar, Elif Nur, Navin Rauniyar, John R. Yates, and Tim Stearns. 2014. "Proximity Interactions among Centrosome Components Identify Regulators of Centriole Duplication." *Current Biology* 24(6): 664–70.
- Fukuyama, Hidehiro et al. 2012. "On-Bead Tryptic Proteolysis: An Attractive Procedure for LC-MS/MS Analysis of the Drosophila Caspase 8 Protein Complex during Immune Response against Bacteria." *Journal of Proteomics* 75(15): 4610–19.
- Furman, Brian L. 2015. "Streptozotocin-Induced Diabetic Models in Mice and Rats." *Current Protocols in Pharmacology*.
- Gannon, Maureen, Rohit N. Kulkarni, Hubert M. Tse, and Franck Mauvais-Jarvis. 2018. "Sex Differences Underlying Pancreatic Islet Biology and Its Dysfunction." *Molecular Metabolism*.
- Geladaki, Aikaterini et al. 2019. "Combining LOPIT with Differential Ultracentrifugation for High-Resolution Spatial Proteomics." *Nature Communications*.
- Gong, Shiaoqing et al. 2003. "A Gene Expression Atlas of the Central Nervous System Based on Bacterial Artificial Chromosomes." *Nature*.
- Gouw, Joost W., Jeroen Krijgsveld, and Albert J.R. Heck. 2010. "Quantitative Proteomics by Metabolic Labeling of Model Organisms." *Molecular and Cellular Proteomics*.
- Graham, Melanie L. et al. 2011. "The Streptozotocin-Induced Diabetic Nude Mouse Model: Differences between Animals from Different Sources." *Comparative Medicine*.
- Guo, Guojun et al. 2018. "Comprehensive Analysis of Differentially Expressed MicroRNAs and MRNAs in Dorsal Root Ganglia from Streptozotocin-Induced Diabetic Rats." *PLoS ONE*.
- Han, Xuemei, Aaron Aslanian, and John R. Yates. 2008. "Mass Spectrometry for Proteomics." *Current Opinion in Chemical Biology*.
- Hirsch, James D. et al. 2002. "Easily Reversible Desthiobiotin Binding to Streptavidin, Avidin, and Other Biotin-Binding Proteins: Uses for Protein Labeling, Detection, and Isolation." *Analytical Biochemistry* 308(2): 343–57.
- Van Hoang, M. et al. 2003. "Quantitative Proteomics Employing Primary Amine Affinity Tags." *Journal of Biomolecular Techniques* 14(3): 216–23.
- Hodge, Kelly, Sara Ten Have, Luke Hutton, and Angus I. Lamond. 2013. "Cleaning up the Masses: Exclusion Lists to Reduce Contamination with HPLC-MS/MS." *Journal of Proteomics*.
- Holt, Leanne Melissa, and Michelle Lynne Olsen. 2016. "Novel Applications of Magnetic Cell Sorting to Analyze Cell-Type Specific Gene and Protein Expression in the Central Nervous System." *PLoS ONE* 11(2): 1–18.
- Hu, Liping et al. 2004. "Two-Dimensional Protein Database of Human Pancreas." *Electrophoresis*.
- Hung, Victoria et al. 2014. "Proteomic Mapping of the Human Mitochondrial Intermembrane Space in Live Cells via Ratiometric APEX Tagging." *Molecular Cell* 55(2): 332–41.
- Hung Victoria et al. 2016. "Spatially Resolved Proteomic Mapping in Living Cells with the Engineered Peroxidase APEX2." *Nature Protocols*.

## References

- Hutsell, Stephanie Q. et al. 2010. "High-Affinity Immobilization of Proteins Using Biotin- and GST-Based Coupling Strategies." *Methods in molecular biology (Clifton, N.J.)* 627: 75–90.
- Hwang, Jiwon, and Peter J. Espenshade. 2016. "Proximity-Dependent Biotin Labelling in Yeast Using the Engineered Ascorbate Peroxidase APEX2." *Biochemical Journal*.
- Jansson, Leif et al. 2016. "Pancreatic Islet Blood Flow and Its Measurement." *Uppsala Journal of Medical Sciences*.
- Jean Beltran, Pierre M., Rommel A. Mathias, and Ileana M. Cristea. 2016. "A Portrait of the Human Organelle Proteome In Space and Time during Cytomegalovirus Infection." *Cell Systems*.
- Ji, Jiaojiao et al. 2019. "Harmine Suppresses Hyper-Activated Ras-MAPK Pathway by Selectively Targeting Oncogenic Mutated Ras/Raf in *Caenorhabditis Elegans*." *Cancer Cell International*.
- Jing, Ji et al. 2015. "Proteomic Mapping of ER-PM Junctions Identifies STIMATE as a Regulator of Ca<sup>2+</sup> Influx." *Nature Cell Biology*.
- John, George Babu, Rana Anjum, Ashok Khar, and Ramakrishnan Nagaraj. 2002. "Subcellular Localization and Physiological Consequences of Introducing a Mitochondrial Matrix Targeting Signal Sequence in Bax and Its Mutants." *Experimental Cell Research* 278(2): 198–208.
- Kalderon, Daniel, Bruce L. Roberts, William D. Richardson, and Alan E. Smith. 1984. "A Short Amino Acid Sequence Able to Specify Nuclear Location." *Cell* 39(3 PART 2): 499–509.
- Karas, Michael, and Ralf Krüger. 2003. "Ion Formation in MALDI: The Cluster Ionization Mechanism." *Chemical Reviews*.
- Kay, Brian K., Sang Thai, and Veronica V. Volgina. 2009. "High-Throughput Biotinylation of Proteins." *Methods in Molecular Biology* 498: 185–98.
- Kim, D. I. et al. 2014. "Probing Nuclear Pore Complex Architecture with Proximity-Dependent Biotinylation." *Proceedings of the National Academy of Sciences* 111(24): E2453–61. <http://www.pnas.org/cgi/doi/10.1073/pnas.1406459111>.
- Kim D. I. et al. 2016. "An Improved Smaller Biotin Ligase for BioID Proximity Labeling." *Molecular Biology of the Cell* 27(8): 1188–96. <http://www.molbiolcell.org/cgi/doi/10.1091/mbc.E15-12-0844>.
- Kim, Min Sik, Jun Zhong, and Akhilesh Pandey. 2016. "Common Errors in Mass Spectrometry-Based Analysis of Post-Translational Modifications." *Proteomics*.
- Knoblauch, Barbara et al. 2003. "ERp19 and ERp46, New Members of the Thioredoxin Family of Endoplasmic Reticulum Proteins." *Molecular & Cellular Proteomics* 2(10): 1104–19. <http://www.mcponline.org/lookup/doi/10.1074/mcp.M300053-MCP200>.
- Knowles, Jeremy R. 1989. "BIOTIN-DEPENDENT ENZYMES." : 195–221.
- Konzer, Anne et al. 2013. "Stable Isotope Labeling in Zebrafish Allows in Vivo Monitoring of Cardiac Morphogenesis." *Molecular and Cellular Proteomics*.
- Koutsikos, D., B. Agroyannis, and H. Tzanatos-Exarchou. 1990. "Biotin for Diabetic Peripheral Neuropathy." *Biomedicine and Pharmacotherapy*.
- Krogager, Toke P. et al. 2018. "Labeling and Identifying Cell-Specific Proteomes in the Mouse Brain." *Nature Biotechnology* 36(2): 156–59.
- Kung, Che Pei, and Maureen E. Murphy. 2016. "The Role of the P53 Tumor Suppressor in Metabolism and Diabetes." *Journal of Endocrinology*.
- Kutlu, Burak et al. 2009. "Detailed Transcriptome Atlas of the Pancreatic Beta Cell." *BMC Medical Genomics*.
- Landgraf, Peter, Elmer R. Antileo, Erin M. Schuman, and Daniela C. Dieterich. 2015. "BONCAT: Metabolic Labeling, Click Chemistry, and Affinity Purification of Newly Synthesized Proteomes." In *Site-Specific Protein Labeling: Methods and Protocols*, , 199–215.
- Lau, Ho Tak, Hyong Won Suh, Martin Golkowski, and Shao En Ong. 2014. "Comparing SILAC- and Stable Isotope Dimethyl-Labeling Approaches for Quantitative Proteomics." *Journal of Proteome Research*.
- Lemire, B. D., C. Fankhauser, A. Baker, and G. Schatz. 1989. "The Mitochondrial Targeting Function of Randomly Generated Peptide Sequences Correlates with Predicted Helical Amphiphilicity." *Journal of Biological Chemistry*.
- Lesch, Hanna P, Minna U Kaikkonen, Jere T Pikkarainen, and Seppo Ylä-Herttuala. 2010. "Avidin-Biotin Technology in Targeted Therapy." *Expert Opinion on Drug Delivery* 7(5): 551–64. <http://www.tandfonline.com/doi/full/10.1517/17425241003677749>.
- Li, Siwen et al. 2015. "Novel Harmine Derivatives for Tumor Targeted Therapy." *Oncotarget*.
- Liechti, Robin et al. 2010. "EuroDia: A Beta-Cell Gene Expression Resource." *Database : the journal of biological databases and curation*.
- Lijana Zaletel-Kragelj, Jadranka Božikov. 2010. "Test Validity Measures and Receiver Operating Characteristic (ROC) Analysis." In *Methods and Tools in Public Health*.
- Liu, Chengyu. 2013. "Strategies for Designing Transgenic DNA Constructs." *Methods in Molecular Biology*.
- Liu, Chih Wei, Mark A. Atkinson, and Qibin Zhang. 2016. "Type 1 Diabetes Cadaveric Human Pancreata Exhibit a Unique Exocrine Tissue Proteomic Profile." *Proteomics*.
- Liu, Jiming et al. 2016. "Harmine Induces Cell Cycle Arrest and Mitochondrial Pathway-Mediated Cellular Apoptosis in SW620 Cells via Inhibition of the Akt and ERK Signaling Pathways." *Oncology Reports*.
- Liu, Xin et al. 2017. "Harmine Is an Inflammatory Inhibitor through the Suppression of NF-KB Signaling." *Biochemical and Biophysical Research Communications*.
- Livnah, O., E. A. Bayer, M. Wilchek, and J. L. Sussman. 1993. "Three-Dimensional Structures of Avidin and the Avidin-Biotin Complex." *Proceedings of the National Academy of Sciences of the United States of America* 90(11): 5076–80.
- Madisen, Linda et al. 2010. "A Robust and High-Throughput Cre Reporting and Characterization System for the Whole Mouse Brain." *Nature Neuroscience*.

## References

- Mair, Andrea et al. 2019. "Proximity Labeling of Protein Complexes and Cell Type Specific Organellar Proteomes in Arabidopsis Enabled by TurboID." *eLife*.
- Martens, Geert A. et al. 2011. "Clusters of Conserved Beta Cell Marker Genes for Assessment of Beta Cell Phenotype." *PLoS ONE*.
- McAlister, Graeme C. et al. 2014. "MultiNotch MS3 Enables Accurate, Sensitive, and Multiplexed Detection of Differential Expression across Cancer Cell Line Proteomes." *Analytical Chemistry*.
- Mears, Rainy et al. 2004. "Proteomic Analysis of Melanoma-Derived Exosomes by Two-Dimensional Polyacrylamide Gel Electrophoresis and Mass Spectrometry." *Proteomics* 4(12): 4019–31.
- Mendoza-Topaz, C., I. Yeow, K. Riento, and B. J. Nichols. 2018. "BioID Identifies Proteins Involved in the Cell Biology of Caveolae." *PLoS ONE* 13(12): 1–22.
- Mick, David U. et al. 2015. "Proteomics of Primary Cilia by Proximity Labeling." *Developmental Cell* 35(4): 497–512.
- Moody, Paul R. et al. 2015. "Receptor Crosslinking: A General Method to Trigger Internalization and Lysosomal Targeting of Therapeutic Receptor:Ligand Complexes." *Molecular Therapy* 23(12): 1888–98.
- Morriswood, Brooke et al. 2013. "Novel Bilobe Components in Trypanosoma Brucei Identified Using Proximity-Dependent Biotinylation." *Eukaryotic Cell* 12(2): 356–67.
- Müller, Anke et al. 2015. "Monitoring Astrocytic Proteome Dynamics by Cell Type-Specific Protein Labeling." *PLoS ONE* 10(12): 1–21.
- Nagy, Andras. 2000. "Cre Recombinase: The Universal Reagent for Genome Tailoring." *Genesis*.
- Najas, Sònia et al. 2015. "DYRK1A-Mediated Cyclin D1 Degradation in Neural Stem Cells Contributes to the Neurogenic Cortical Defects in Down Syndrome." *EBioMedicine*.
- Nakagawa, Yoshio et al. 2010. "Mitochondrial Dysfunction and Biotransformation of  $\beta$ -Carboline Alkaloids, Harmine and Harmaline, on Isolated Rat Hepatocytes." *Chemico-Biological Interactions*.
- Nishida, Yoriko, Kaoru Aida, Makoto Kihara, and Tetsuro Kobayashi. 2014. "Antibody-Validated Proteins in Inflamed Islets of Fulminant Type 1 Diabetes Profiled by Laser-Capture Microdissection Followed by Mass Spectrometry." *PLoS ONE*.
- Nunnari, Jodi, and Anu Suomalainen. 2012. "Mitochondria: In Sickness and in Health." *Cell*.
- Nyalwidhe, Julius O. et al. 2017. "Comparative Quantitative Proteomic Analysis of Disease Stratified Laser Captured Microdissected Human Islets Identifies Proteins and Pathways Potentially Related to Type 1 Diabetes." *PLoS ONE* 12(9): 1–20.
- Ong, Shao En, and Matthias Mann. 2005. "Mass Spectrometry–Based Proteomics Turns Quantitative." *Nature Chemical Biology* 1(5): 252–62.
- Parca, Luca, Martin Beck, Peer Bork, and Alessandro Ori. 2018. "Quantifying Compartment-associated Variations of Protein Abundance in Proteomics Data." *Molecular Systems Biology*.
- Pflugradt, René et al. 2011. "A Novel and Effective Separation Method for Single Mitochondria Analysis." *Mitochondrion*.
- Ping, Lingyan et al. 2013. "Quantitative Proteomics Reveals Significant Changes in Cell Shape and an Energy Shift after IPTG Induction via an Optimized SILAC Approach for Escherichia Coli." *Journal of Proteome Research*.
- Pisitkun, T., R.-F. Shen, and M. A. Knepper. 2004. "Identification and Proteomic Profiling of Exosomes in Human Urine." *Proc. Natl. Acad. Sci. USA* 101(36): 13368–73.  
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=516573&tool=pmcentrez&rendertype=abstract>.
- Polyak, Steven W. 2015. "Mechanisms of Biotin Transport." *Biochemistry & Analytical Biochemistry*.
- Prokisch, H et al. 2004. "Integrative Analysis of the Mitochondrial Proteome in Yeast." *PLoS Biol* 2(6): e160.  
<http://www.ncbi.nlm.nih.gov/pubmed/15208715>  
<http://www.plosbiology.org/article/doi/10.1371/journal.pbio.0020160>&representation=PDF.
- Rains, Justin L., and Sushil K. Jain. 2011. "Oxidative Stress, Insulin Signaling, and Diabetes." *Free Radical Biology and Medicine*.
- Rao, Srivatsa V., Kimberly W. Anderson, and Leonidas G. Bachas. 1997. "Determination of the Extent of Protein Biotinylation by Fluorescence Binding Assay." *Bioconjugate Chemistry* 8(1): 94–98.
- Rappsilber, Juri, Yasushi Ishihama, and Matthias Mann. 2003. "Stop And Go Extraction Tips for Matrix-Assisted Laser Desorption/Ionization, Nano-electrospray, and LC/MS Sample Pretreatment in Proteomics." *Analytical Chemistry* 75(3): 663–70.
- Regev, Aviv et al. 2017. "The Human Cell Atlas." *eLife*.
- Reinke, Aaron W., Raymond Mak, Emily R. Troemel, and Eric J. Ben. 2017. "In Vivo Mapping of Tissue- and Subcellular-Specific Proteomes in Caenorhabditis Elegans." *Science Advances* 3(5): 1–12.
- Ren, Wen Xiu et al. 2015. "Recent Development of Biotin Conjugation in Biological Imaging, Sensing, and Target Delivery." *Chemical Communications* 51(52): 10403–18.
- Rennel, Emma, and Pär Gerwins. 2002. "How to Make Tetracycline-Regulated Transgene Expression Go on and Off." *Analytical Biochemistry*.
- Réus, Gislaine Z. et al. 2010. "Harmine and Imipramine Promote Antioxidant Activities in Prefrontal Cortex and Hippocampus." *Oxidative Medicine and Cellular Longevity*.
- Rhee, Hyun Woo et al. 2013. "Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging." *Science*.
- Robert, McMahon J. 2007. Avidin-Biotin Interactions *Avidin-Biotin Interactions*.
- Rösli, Christoph, Jascha-N Rybak, Dario Neri, and Giuliano Elia. 2008. 418 Methods in molecular biology (Clifton, N.J.) *Quantitative Recovery of Biotinylated Proteins from Streptavidin-Based Affinity Chromatography Resins*.  
<http://www.ncbi.nlm.nih.gov/pubmed/18287652>.

## References

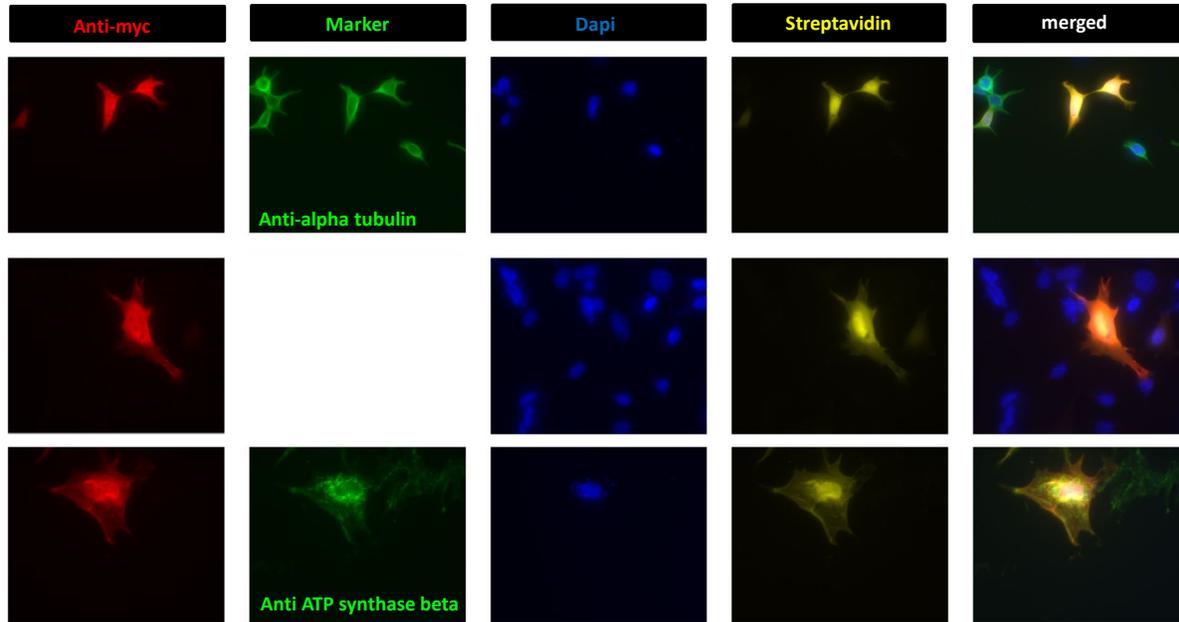
- Roux, Kyle J., Dae In Kim, Brian Burke, and Danielle G. May. 2018. "BioID: A Screen for Protein-Protein Interactions." *Current protocols in protein science*.
- Roux, Kyle J., Dae In Kim, Manfred Raida, and Brian Burke. 2012. "A Promiscuous Biotin Ligase Fusion Protein Identifies Proximal and Interacting Proteins in Mammalian Cells." *Journal of Cell Biology* 196(6): 801–10. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3308701&tool=pmcentrez&rendertype=abstract> (July 11, 2014).
- Sacco, Francesca et al. 2016. "Glucose-Regulated and Drug-Perturbed Phosphoproteome Reveals Molecular Mechanisms Controlling Insulin Secretion." *Nature Communications* 7(May).
- Sakai Norio 2014. "Principles for the use of in vivo transgene techniques: Overview and an introductory practical guide for the selection of tetracycline-controlled transgenic mice" *Methods Mol Biol*. doi:10.1007/978-1-4939-0404-4\_4
- Satori, Chad P. et al. 2013. "Bioanalysis of Eukaryotic Organelles." *Chemical Reviews* 113(4): 2733–2811.
- Schiapparelli, Lucio Matias et al. 2014. "Direct Detection of Biotinylated Proteins by Mass Spectrometry." *Journal of Proteome Research* 13(9): 3966–78.
- Schoof, Erwin M. et al. 2019. "A Quantitative Single-Cell Proteomics Approach to Characterize an Acute Myeloid Leukemia Hierarchy." *bioRxiv*.
- Schwartzman, Armin et al. 2009. "Empirical Null and False Discovery Rate Analysis in Neuroimaging." *NeuroImage*.
- Schwarz, Jaelyn M. 2015. "Using Fluorescence Activated Cell Sorting to Examine Cell-Type-Specific Gene Expression in Rat Brain Tissue." *Journal of Visualized Experiments* 2015(99): 1–12.
- Schwenk, Frieder, Udo Baron, and Klaus Rajewsky. 1995. "A Cre-Transgenic Mouse Strain for the Ubiquitous Deletion of LoxP-Flanked Gene Segments Including Deletion in Germ Cells." *Nucleic Acids Research*.
- Sears, Rhiannon M., Danielle G. May, and Kyle J. Roux. 2019. "BioID as a Tool for Protein-Proximity Labeling in Living Cells." In *Methods in Molecular Biology*.
- Shao, Shiyong et al. 2016. "Reproducible Tissue Homogenization and Protein Extraction for Quantitative Proteomics Using Micropestle-Assisted Pressure-Cycling Technology." *Journal of Proteome Research*.
- Shi, Yanhong, Haruhisa Inoue, Joseph C. Wu, and Shinya Yamanaka. 2017. "Induced Pluripotent Stem Cell Technology: A Decade of Progress." *Nature Reviews Drug Discovery*.
- Shoib, Muhammad et al. 2013. "PUB-NChIP-"in Vivo Biotinylation" Approach to Study Chromatin in Proximity to a Protein of Interest." *Genome Research* 23(2): 331–40.
- Da Silva Xavier, Gabriela. 2018. "The Cells of the Islets of Langerhans." *Journal of Clinical Medicine* 7(3): 54.
- Smith, Anthony C., and Alan J. Robinson. 2016. "MitoMiner v3.1, an Update on the Mitochondrial Proteomics Database." *Nucleic Acids Research*.
- Spence, Erin F. et al. 2019. "In Vivo Proximity Proteomics of Nascent Synapses Reveals a Novel Regulator of Cytoskeleton-Mediated Synaptic Maturation." *Nature Communications* 10(1): 1–16. <http://dx.doi.org/10.1038/s41467-019-08288-w>.
- Stein, Colleen S., and Beverly L. Davidson. 2002. "Gene Transfer to the Brain Using Feline Immunodeficiency Virus-Based Lentivirus Vectors." *Methods in Enzymology*.
- Stein, Derek R. et al. 2013. "High PH Reversed-Phase Chromatography as a Superior Fractionation Scheme Compared to off-Gel Isoelectric Focusing for Complex Proteome Analysis." *Proteomics*.
- Stützer, I., D. Esterházy, and M. Stoffel. 2012. "The Pancreatic Beta Cell Surface Proteome." *Diabetologia*.
- Subramanya, Sandeep B. et al. 2011. "Inhibition of Intestinal Biotin Absorption by Chronic Alcohol Feeding: Cellular and Molecular Mechanisms." *American Journal of Physiology - Gastrointestinal and Liver Physiology*.
- Thorens, Bernard et al. 2015. "Ins1 Cre Knock-in Mice for Beta Cell-Specific Gene Recombination." *Diabetologia*.
- Thul, Peter J. et al. 2017. "A Subcellular Map of the Human Proteome." *Science*.
- Ting, Lily, Ramin Rad, Steven P. Gygi, and Wilhelm Haas. 2011. "MS3 Eliminates Ratio Distortion in Isobaric Multiplexed Quantitative Proteomics." *Nature Methods*.
- Tonne, Jason M. et al. 2013. "Global Gene Expression Profiling of Pancreatic Islets in Mice during Streptozotocin-Induced  $\beta$ -Cell Damage and Pancreatic Glp-1 Gene Therapy." *DMM Disease Models and Mechanisms*.
- Trimpin, Sarah, Ellen D. Inutan, Thushani N. Herath, and Charles N. McEwen. 2010. "Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Method for Selectively Producing Either Singly or Multiply Charged Molecular Ions." *Analytical Chemistry*.
- Trinkle-Mulcahy, Laura. 2019. "Recent Advances in Proximity-Based Labeling Methods for Interactome Mapping [Version 1; Referees: 2 Approved]." *F1000Research* 8(0).
- Tsai, Pei Yun et al. 2013. "Proteome Analysis of Altered Proteins in Streptozotocin-Induced Diabetic Rat Kidney Using the Fluorogenic Derivatization-Liquid Chromatography-Tandem Mass Spectrometry Method." *Biomedical Chromatography*.
- Tyanova, Stefa et al. 2016. "The Perseus Computational Platform for Comprehensive Analysis of (Prote)Omics Data." *Nature Methods*.
- Udeshi, Namrata D. et al. 2017. "Antibodies to Biotin Enable Large-Scale Detection of Biotinylation Sites on Proteins." *Nature Methods* 14(12): 1167–70.
- Uezu, Akiyoshi et al. 2016. "Identification of an Elaborate Complex Mediating Postsynaptic Inhibition." *Science* 353(6304): 1123–29.
- Uezu, Akiyoshi, and Scott Soderling. 2019. "Identifying Synaptic Proteins by in Vivo BioID from Mouse Brain." In *Methods in Molecular Biology*.
- Uhlén, Mathias et al. 2015. "Tissue-Based Map of the Human Proteome." *Science*.
- Veech, Joseph A. 2012. "Significance Testing in Ecological Null Models." *Theoretical Ecology*.
- Vethe, Heidrun et al. 2017. "Probing the Missing Mature  $\beta$ -Cell Proteomic Landscape in Differentiating Patient iPSC-Derived

## References

- Cells." *Scientific Reports* 7(1): 1–14.
- Vinerean, Horatiu V., Lawrence S. Gazda, Richard D. Hall, and Barry H. Smith. 2011. "Streptozotocin Is Responsible for the Induction and Progression of Renal Tumorigenesis in Diabetic Wistar-Furth Rats Treated with Insulin or Transplanted with Agarose Encapsulated Porcine Islets." *Islets*.
- Volpe, Caroline Maria Oliveira, Pedro Henrique Villar-Delfino, Paula Martins Ferreira Dos Anjos, and José Augusto Nogueira-Machado. 2018. "Cellular Death, Reactive Oxygen Species (ROS) and Diabetic Complications Review-Article." *Cell Death and Disease*.
- Waanders, Leonie F. et al. 2009. "Quantitative Proteomic Analysis of Single Pancreatic Islets." *Proceedings of the National Academy of Sciences of the United States of America* 106(45): 18902–7.
- Wang, Lijun et al. 2008. "Restricted Expression of Mutant SOD1 in Spinal Motor Neurons and Interneurons Induces Motor Neuron Pathology." *Neurobiology of Disease*.
- Wang, Peng et al. 2015. "Induction of Human Pancreatic Beta Cell Replication by Inhibitors of Dual Specificity Tyrosine Regulated Kinase." *Nat Med*.
- Wessels, Hans J.C.T. et al. 2013. "Analysis of 953 Human Proteins from a Mitochondrial HEK293 Fraction by Complexome Profiling." *PLoS ONE*.
- Wilchek, Meir, and Edward A. Bayer. 1988. "The Avidin-Biotin Complex in Bioanalytical Applications." *Analytical Biochemistry* 171(1): 1–32.
- Wilson, Rashaun S., and Angus C. Nairn. 2018. "Cell-Type-Specific Proteomics: A Neuroscience Perspective." *Proteomes* 6(4): 1–23.
- Wilson, Russell B., and Robert F. Murphy. 1989. "Flow-Cytometric Analysis of Endocytic Compartments." *Methods in Cell Biology*.
- Winter, Dominic, and Hanno Steen. 2011. "Optimization of Cell Lysis and Protein Digestion Protocols for the Analysis of HeLa S3 Cells by LC-MS/MS." *Proteomics*.
- Wu, Yue et al. 2014. "Five-Plex Isotope Dimethyl Labeling for Quantitative Proteomics." *Chemical Communications* 50(14): 1708–10.
- Yabut, Odessa, Jason Domogauer, and Gabriella D'Arcangelo. 2010. "Dyrk1A Overexpression Inhibits Proliferation and Induces Premature Neuronal Differentiation of Neural Progenitor Cells." *Journal of Neuroscience*.
- Yamamoto, Ai, René Hen, and William T. Dauer. 2001. "The On and Offs of Inducible Transgenic Technology: A Review." *Neurobiology of Disease*.
- Yan, Wei, Ruedi Aebersold, and Elaine W. Raines. 2009. "Evolution of Organelle-Associated Protein Profiling." *Journal of Proteomics* 72(1): 4–11.
- Yu, Hsiao Man Ivy et al. 2005. "Development of a Unique System for Spatiotemporal and Lineage-Specific Gene Expression in Mice." *Proceedings of the National Academy of Sciences of the United States of America*.
- Zhang, Jin, Robert E. Campbell, Alice Y. Ting, and Roger Y. Tsien. 2002. "Creating New Fluorescent Probes for Cell Biology." *Nature Reviews Molecular Cell Biology* 3(12): 906–18.
- Zhang, Lina et al. 2017. "Proteomic Profiling of Human Islets Collected from Frozen Pancreata Using Laser Capture Microdissection." *Journal of Proteomics*.
- Zhang, Xiao Fei et al. 2016. "Synthesis and Mechanisms of Action of Novel Harmine Derivatives as Potential Antitumor Agents." *Scientific Reports*.
- Zhang, Yaoyang et al. 2013. "Protein Analysis by Shotgun/Bottom-up Proteomics." *Chemical Reviews*.
- Zhou, Hongxia et al. 2009. "Developing TTA Transgenic Rats for Inducible and Reversible Gene Expression." *International Journal of Biological Sciences*.
- Zhu, Wenhong, Jeffrey W. Smith, and Chun Ming Huang. 2010. "Mass Spectrometry-Based Label-Free Quantitative Proteomics." *Journal of Biomedicine and Biotechnology*.
- Zischka, Hans et al. 2006. "Differential Analysis of *Saccharomyces Cerevisiae* Mitochondria by Free Flow Electrophoresis." *Molecular & cellular proteomics : MCP* 5(11): 2185–2200.

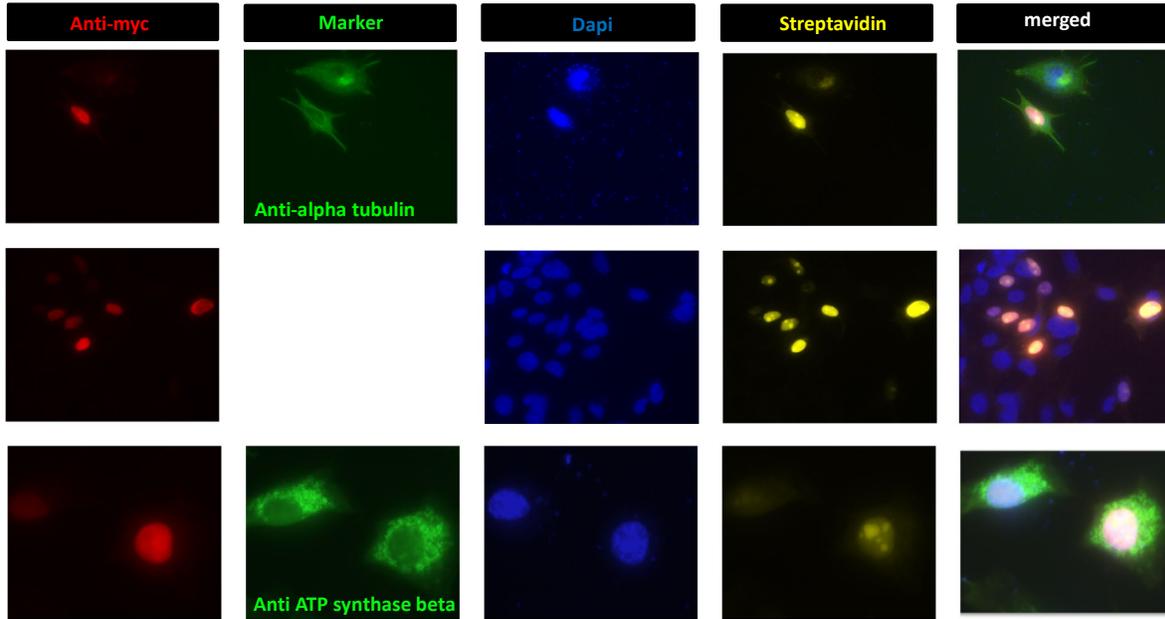
## 13. Appendix

## Cytosolic and Nuclear BioID (CytoNucl.-BioID)

**Appendix Figure 1. Confirmation of correct localization of CytoNucl.-BioID by microscopy.**

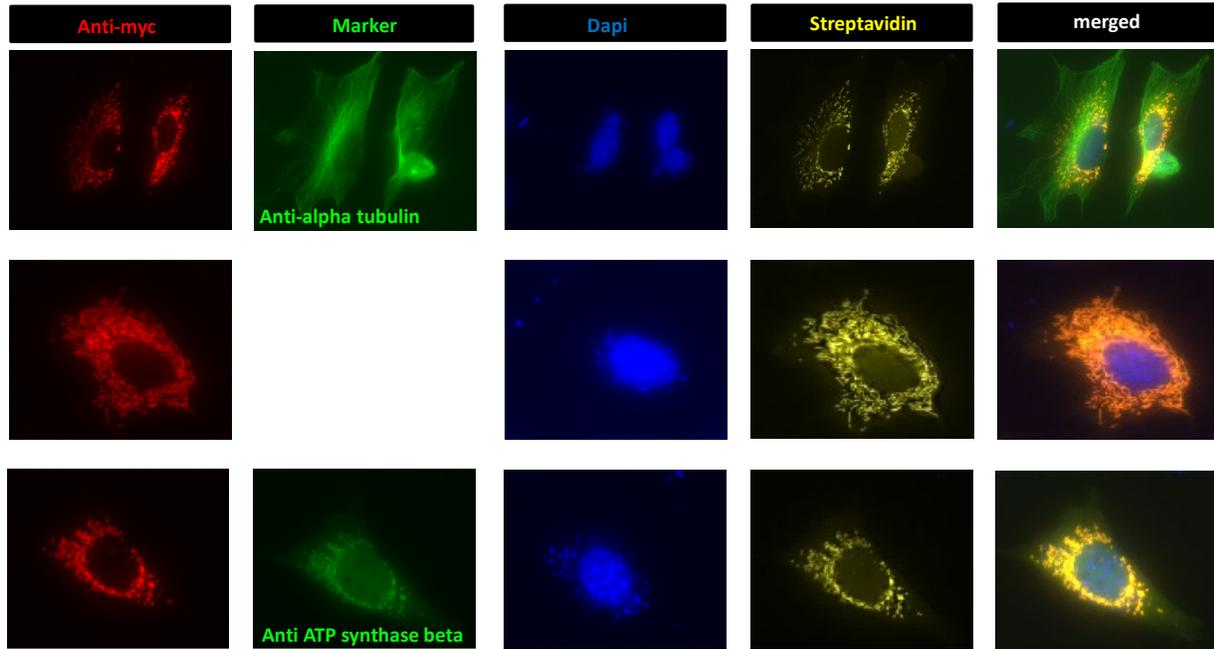
Streptavidin, anti-myc antibody and organelle specific antibodies were used to show localization of CytoNucl.-BioID expression and biotinylation. NIH/3T3 cells were transfected with myc-tagged CytoNucl.-BioID construct, treated with biotin (50  $\mu$ M - 24 h), fixed and stained with anti-myc antibody (ab9106, Abcam: 1: 600- showing BioID expression), anti-alpha tubulin antibody [DM1A] (Alexa Fluor® 488) (ab195887, Abcam, 1: 400- Cytosolic marker), anti-ATP synthase beta monoclonal antibody [3D5AB1] (A21351, ThermoFischer scientific, 1: 500- Mitochondrial marker); Alexa Fluor 647 streptavidin (S21374, ThermoFischer scientific, 1: 400) was used to detect biotinylated proteins. Secondary antibodies were: Goat IgG anti-rabbit IgG (H+L)-Cy3 (111-165-144, Dianova, 1: 400), and MFP488 goat anti-mouse IgG (H+L) (MoBiTec, MFP-A1029, 1: 400). Coverslips were mounted with fluoromout-G mounting solution containing DAPI for nuclear staining, used also as nuclear marker (ThermoFischer scientific, Bremen, Germany). A similar increase of brightness was applied to all figures showing DAPI staining to help for visibility.

## Nuclear BioID (Nucl.-BioID)

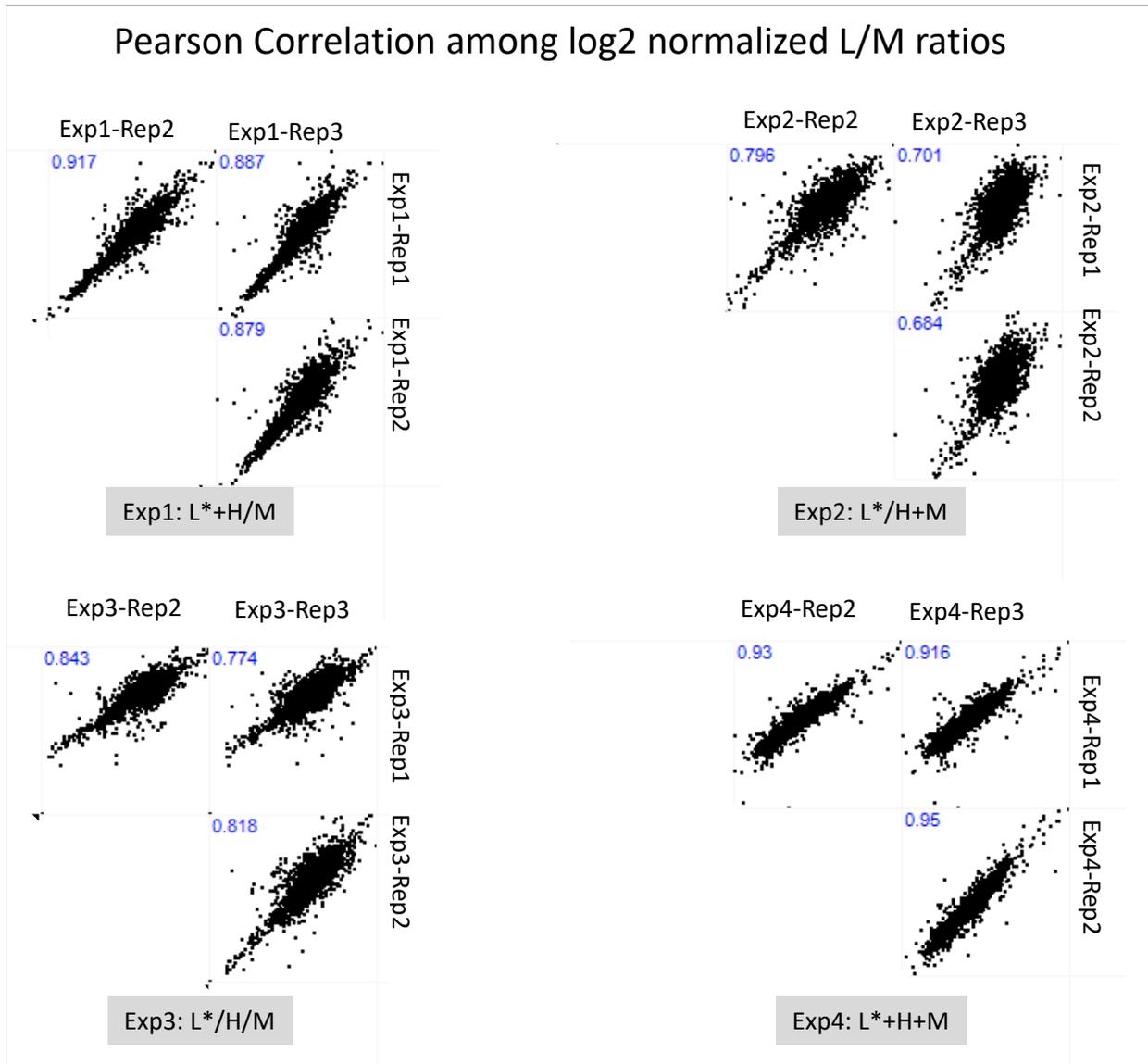
**Appendix Figure 2. Confirmation of correct localization of Nucl.-BioID by microscopy.**

Streptavidin, anti-myc antibody and organelle specific antibodies were used to show localization of Nucl.-BioID expression and biotinylation. NIH/3T3 cells were transfected with myc-tagged Nucl.-BioID construct, treated with biotin (50  $\mu$ M - 24 h), fixed and stained with anti-myc antibody (ab9106, Abcam: 1: 600- showing BioID expression), anti-alpha tubulin antibody [DM1A] (Alexa Fluor® 488) (ab195887, Abcam, 1: 400- Cytosolic marker), anti-ATP synthase beta monoclonal antibody [3D5AB1] (A21351, ThermoFischer scientific, 1: 500- Mitochondrial marker); Alexa Fluor 647 streptavidin (S21374, ThermoFischer scientific, 1: 400) was used to detect biotinylated proteins. Secondary antibodies were: Goat IgG anti-rabbit IgG (H+L)-Cy3 (111-165-144, Dianova, 1: 400), and MFP488 goat anti-mouse IgG (H+L) (MoBiTec, MFP-A1029, 1: 400). Coverslips were mounted with fluoromout-G mounting solution containing DAPI for nuclear staining, used also as nuclear marker (ThermoFischer scientific, Bremen, Germany). A similar increase of brightness was applied to all figures showing DAPI as well as streptavidin stainings to help for visibility.

## Mitochondrial BioID (Mito-BioID)

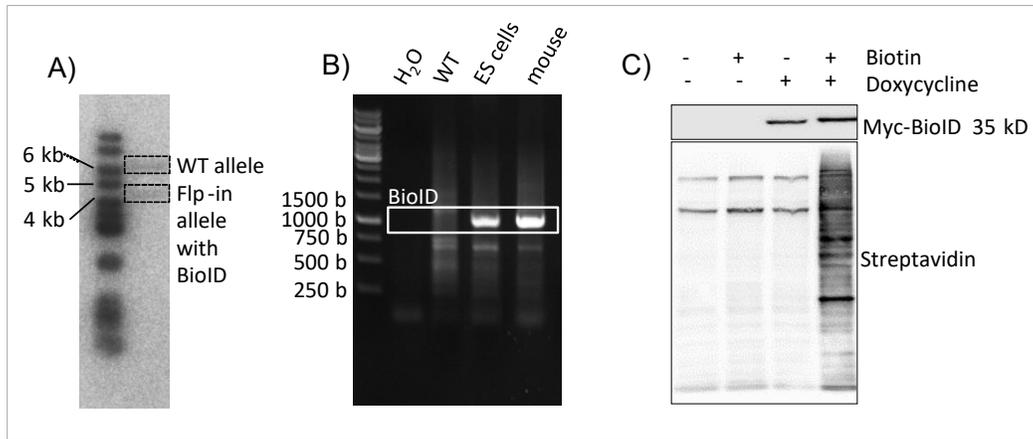
**Appendix Figure 3. Confirmation of correct localization of Mito-BioID by microscopy.**

Streptavidin, anti-myc antibody and organelle specific antibodies were used to show localization of Mito-BioID expression and biotinylation. NIH/3T3 cells were transfected with myc-tagged Mito-BioID construct, treated with biotin (50  $\mu$ M - 24 h), fixed and stained with anti-myc antibody (ab9106, Abcam: 1: 600- showing BioID expression), anti-alpha tubulin antibody [DM1A] (Alexa Fluor® 488) (ab195887, Abcam, 1: 400- Cytosolic marker), anti-ATP synthase beta monoclonal antibody [3D5AB1] (A21351, ThermoFischer scientific, 1: 500- Mitochondrial marker); Alexa Fluor 647 streptavidin (S21374, ThermoFischer scientific, 1: 400) was used to detect biotinylated proteins. Secondary antibodies were: Goat IgG anti-rabbit IgG (H+L)-Cy3 (111-165-144, Dianova, 1: 400), and MFP488 goat anti-mouse IgG (H+L) (MoBiTec, MFP-A1029, 1: 400). Coverslips were mounted with fluoromout-G mounting solution containing DAPI for nuclear staining, used also as nuclear marker (ThermoFischer scientific, Bremen, Germany). A similar increase of brightness was applied to all figures showing DAPI as well as streptavidin stainings to help for visibility.



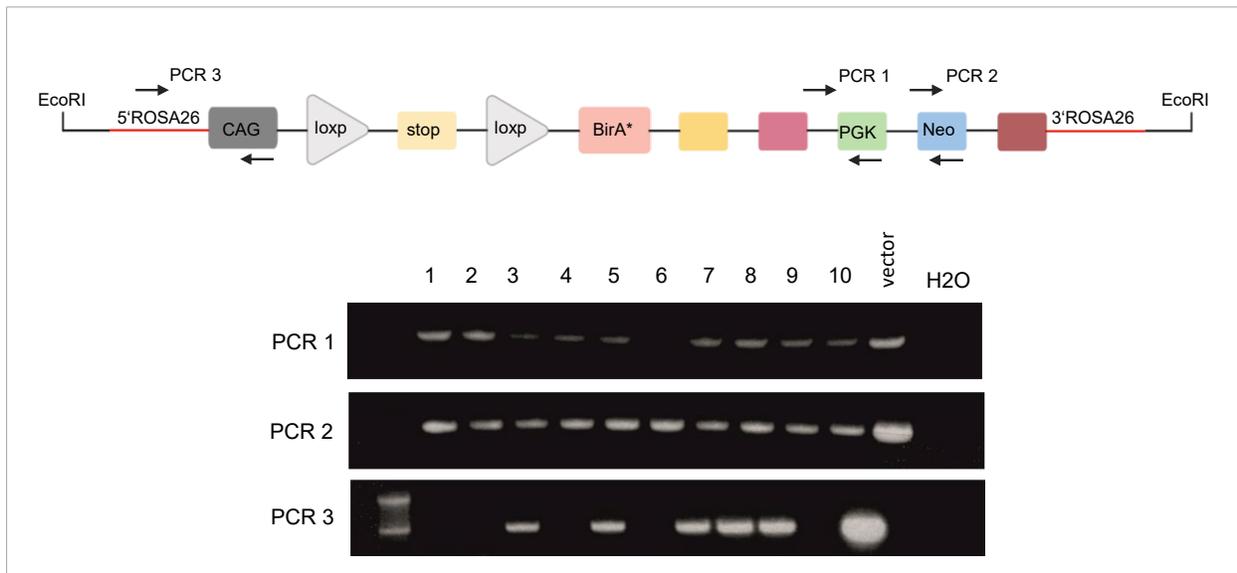
**Appendix Figure 4. Correlation of log<sub>2</sub>-transformed L/M ratios between biological replicates of combination experiments 1 to 4.**

Light labeled HEK 293T cells were transiently transfected with Nucl.-BioID construct and biotinylated; medium and heavy labeled cells served as controls. Biotinylated and the 2 control samples were combined 1:1:1 in different steps of the on-bead digestion experiment. Method/Exp 1 (L\*+ H / M): Biotinylated proteins in light were mixed before incubation on bead with heavy control sample and further mixed with medium control before trypsin digestion. In method/Exp 2 (L\* / M + H) medium and heavy control samples were mixed before incubation with beads and mixed with biotinylated samples before trypsin digestion. In method/Exp 3 (L\* /M/ H) and 4/Exp (L\* + M + H) samples were combined before trypsin digestion and before incubation on bead, respectively. All samples were measured by MS and analyzed by MaxQuant/Perseus. Log<sub>2</sub> transformed L\*/M ratios of individual replicates were plotted against each other and Pearson Correlation was determined.



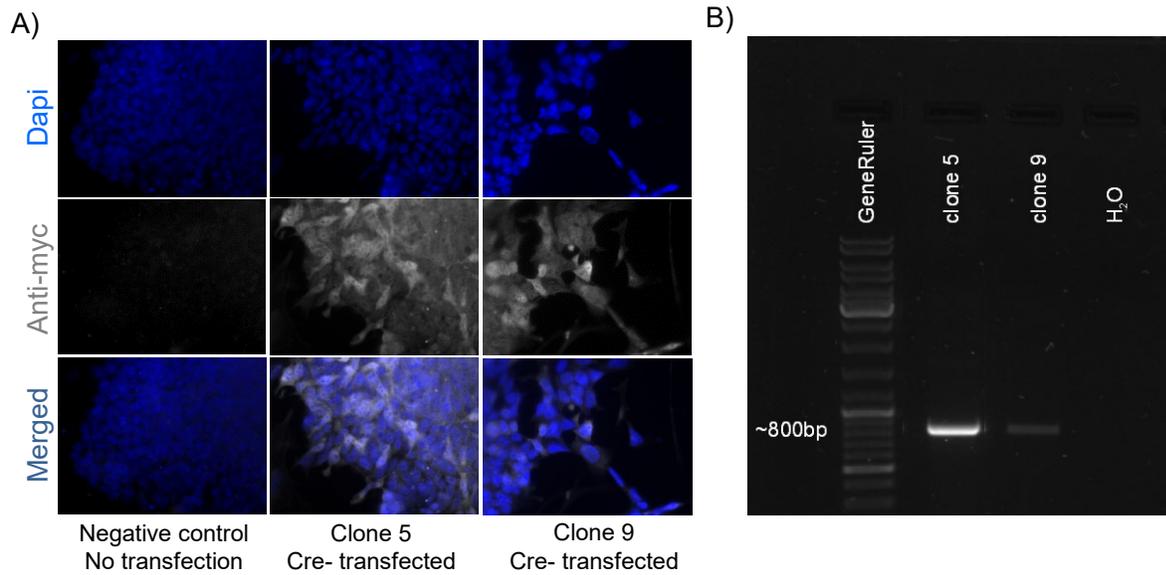
**Appendix Figure 5. Confirmation of transgene insertion in ES cells by southern blot, PCR, and western blot for Coll1a1-BioID mouse (mouse 1)**

a) Southern blot of KH2 ES-cells after insertion of BioID in the Coll1a1 locus by Flp recombination; b) PCR for BioID of KH2-ES cell lysates and the ISBmouse obtained in the F1 generation; c) verification of BioID expression by anti-Myc Western blot of KH2-ES cells treated with doxycycline / biotin; d) confirmation of biotinylation in KH2 ES-cell expressing BioID by Western blot with streptavidin coupled HRP after doxycyclin / biotin treatment. The experiments were performed and the data generated by Dr. Dominic Winter, Institute for Biochemistry and Molecular Biology, University of Bonn.



**Appendix Figure 6. Confirmation of transgene insertion by PCR for Rosa-CAG-BioID mouse (mouse 2).**

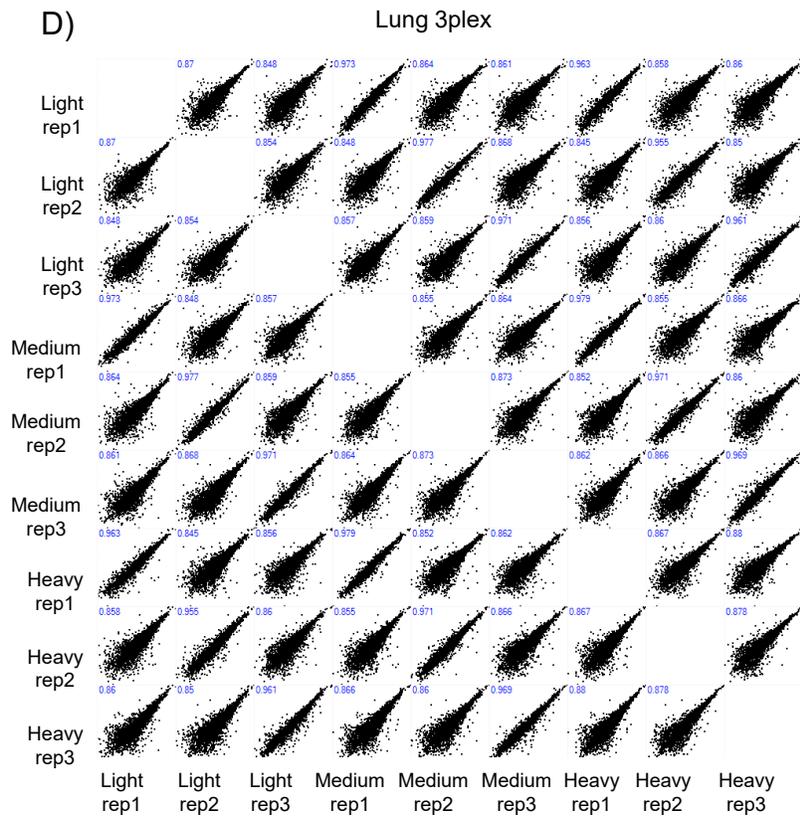
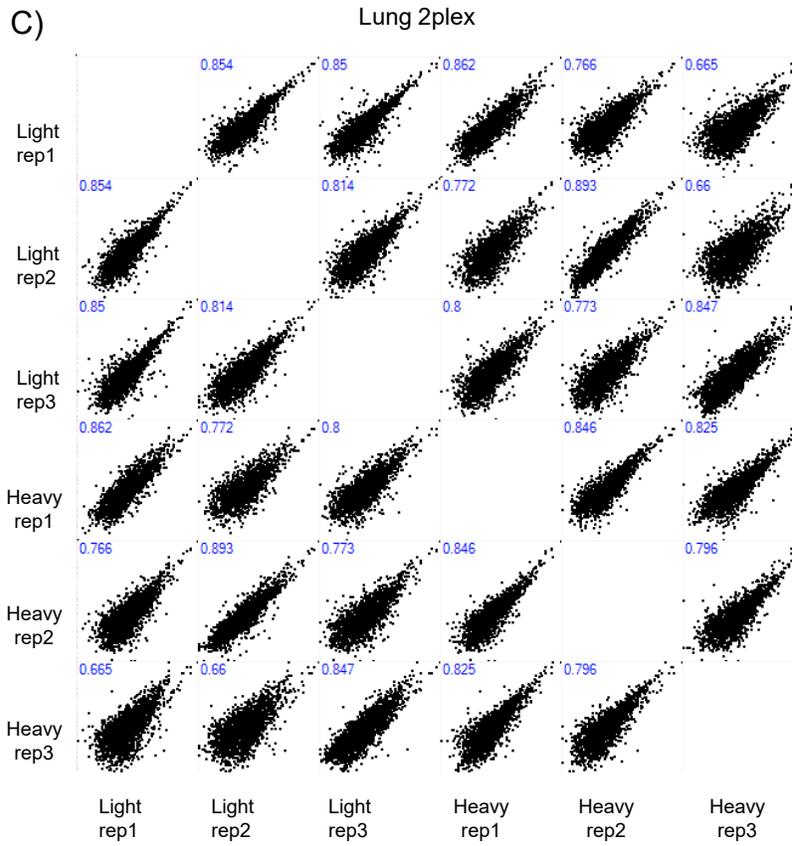
10 Neomycin resistant ES cell clones were analyzed by PCR. All clones were positive for Neomycin (PCR1). Except clone 6, all other clones were positive for PGK (PCR2) and Rosa CAG PCR was positive for clones 3, 5, 7, 8, 9 (PCR3). ES cell clone numbers 5 and 9 were further analyzed. The experiments were performed and the data generated by Dr. Caroline Geisen, Institute of Physiology 1, University of Bonn.



**Appendix Figure 7. ES cell transfection with cre recombinase for confirmation of transgene insertion for Rosa-CAG-BioID mouse (mouse2)**

Cells of clone 5 and 9 were transfected with a Cre expressing plasmid and RNA was isolated. A) ES cell clones were stained with anti-myc (Abcam, ab9106, 1: 400) and dapi to evaluate the expression of BioID. B) RT-PCR was performed using a primer pair detecting the BioID-transcript; forward primer was located at Exon 1 of the CAG promoter and the reverse primer inside the BioID sequences. The experiments were performed and data generated by Dr. Caroline Geisen, Institute of Physiology 1, University of Bonn.





**Appendix Figure 8. Correlation of log<sub>2</sub>-transformed dimethyl channel intensities between biological replicates of tissue datasets including one and two control samples (2plex and 3plex).**

In 2plex approach Biotin (0.22mg/ml) and Dox (2 mg/ml) were administered in water to BioID<sup>+</sup>rtTA<sup>+</sup> and control mice for seven days, mice were sacrificed and pancreas isolated followed by on-bead digestion (the old approach-method1 in method development study) using equal amounts of proteins. In 3plex approach another control mouse was inculcated and treated similarly. The method for enrichment of 3plex samples was the best method in method development study (method 5). Biotinylated organs were labeled with heavy dimethyl labeling materials. A control sample received the light label in 2plex and 3plex and the second control sample was labeled with medium label in 3plex approach. Samples were combined, OASIS desalted, fractionated by OFFGEL. All samples were measured by Orbitrap-Velos and analyzed by MaxQuant/Perseus. Log<sub>2</sub> transformed intensities of individual replicates were plotted against each other and Pearson Correlation was determined for A) 2plex-liver B) 3plex-liver C) 2plex-lung D) 3plex-Lung. The plots were generated by Dr. Robert Hardt.

## 14. Acknowledgement

I would like to thank my direct supervisor Dr. Dominic Winter for his scientific advices, novel ideas, patient guidance, and great support he has provided me during my doctoral studies. He was always willing to answer my questions even during his busy schedule.

I also would like to express my gratitude to Prof. Dr. Gieselmann for his support and guidance regarding my projects and accepting to be my first referee in my defense committee. Moreover, I am thankful to Prof. Thiele, Prof. Häberlein, and Prof. Kroupa who readily agreed to be my second referee, fachnahe, and fachfremde referees, respectively.

I would like to thank all the members of IBMB especially Elham Pourbarkhordariesfandabadi, Dr. Alireza Dehghani, Dr. Robert Hardt, and our technical assistant Norbert Rösel for their everyday availability, guidance, and support. In addition, I am grateful to my colleague Ulf Einsfelder for his scientific advices on my project and proof reading of my thesis.

I must also express my gratitude to my lovely parents and my friends Hassan Solhjou, Shahrbanoo Aboutalebi, and Tina Ketabi who experienced all the up and downs of my research and walked the path beside me with inexhaustible supply of love, support, encouragement, and patience.

And a great thanks to all IBMB members. I had the best colleagues one could wish for; supportive, helpful, and warm-hearted. We enjoyed countless moments together. I truly enjoyed my life in IBMB institute and will miss all my colleagues and friends wherever I go, but the sweet memories we shared will stay with me forever.

Thank you all!