A genetic model for in vivo proximity labelling of the mammalian secretome

Rui Yang1,2,†, Amanda S. Meyer1,2,†, Ilia A. Droujinine3,†, Namrata D. Udesti4, Yanhui Hu5, Jinjin Guo1,2, Jill A. McMahon1,2, Dominique K. Carey4, Charles Xu6, Qiao Fang6, Jihui Sha7, Shishang Qin8, David Rocco5, James Wohlschlegel7, Alice Y. Ting9,10, Steven A. Carr4, Norbert Perrimon5,11 and Andrew P. McMahon1,2

1Department of Stem Cell Biology and Regenerative Medicine, and 2Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research, University of Southern California, Los Angeles, CA, USA
3Department of Molecular Medicine, Scripps Research, La Jolla, CA, USA
4Broad Institute of Harvard and MIT, Cambridge, MA, USA
5Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA, USA
6Department of Molecular Genetics, University of Toronto, Toronto, ON Canada, M5S 3E1
7Department of Biological Chemistry, Geffen School of Medicine at UCLA, University of California, Los Angeles, Los Angeles, CA, USA
8BIOPIC, Beijing Advanced Innovation Center for Genomics, School of Life Sciences, Peking University, Beijing, People’s Republic of China
9Chan Zuckerberg Biohub, San Francisco, CA, USA
10Departments of Genetics, Biology, and Chemistry, Stanford University, Stanford, CA, USA
11Howard Hughes Medical Institute, Boston, MA, USA

†These authors contributed equally.

Organ functions are highly specialized and interdependent. Secreted factors regulate organ development and mediate homeostasis through serum trafficking and inter-organ communication. Enzyme-catalysed proximity labelling enables the identification of proteins within a specific cellular compartment. Here, we report a BirA*G3 mouse strain that enables CRE-dependent promiscuous biotinylation of proteins trafficking through the endoplasmic reticulum. When broadly activated throughout the mouse, widespread labelling of proteins was observed within the secretory pathway. Streptavidin affinity purification and peptide mapping by quantitative mass spectrometry (MS) proteomics revealed organ-specific secretory profiles and serum trafficking. As expected, secretory proteomes were highly enriched for signal peptide-containing proteins, highlighting both conventional and non-conventional secretory processes, and ectodomain shedding. Lower-abundance proteins with hormone-like properties were recovered and validated using orthogonal approaches. Hepatocyte-specific activation of BirA*G3 highlighted liver-specific biotinylated secretome profiles. The BirA*G3 mouse model demonstrates enhanced labelling efficiency and tissue specificity over viral transduction approaches and will facilitate a deeper understanding of secretory protein interplay in development, and in healthy and diseased adult states.

1. Introduction

Protein secretion plays a critical role in coordinating local and systemic cellular responses in development, homeostasis and disease [1–4]. Multi-organ failure suggests an aberrant organ–organ crosstalk resulting in linked organ pathology such as pulmonary–renal syndromes [5,6], with an increased risk of sequential organ failure and morbidity [7]. Secreted proteins may be identified using liquid chromatography-tandem mass spectrometry (LC-MS/MS) proteomics...

© 2022 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, provided the original author and source are credited.
of serum [8]. However, it is challenging to identify low abundance proteins and difficult to track the organs of origin and ultimate destination of protein interactions [9,10].

Analysis of cell secretomes has benefited from enzymecatalyzed proximity labelling approaches such as BioID [11] and TurBOX [12]. In these, the activity of a promiscuous biotin-ligase in the cellular secretory pathway biotinylates resident and secreted proteins, which can then be detected by affinity enrichment and quantitative MS [11,12]. These labelling methods allow sensitive and stable detection of endogenous secreted proteomes in live cells, including fly (Drosophila melanogaster) [12,13] and worm (Caenorhabditis elegans) models [12], mouse tumor transplants [13] and specific mammalian target tissues through viral directed gene delivery [14-16]. These studies have provided new insights into tissue secretomes and inter-organ communication [14-19].

To overcome the limitations of virus-mediated approaches for systematic temporal and spatial analysis of mammalian cell secretomes in vivo, we generated and validated a mouse model system. In the secretome reporter strain, DNA sequences encoding an endoplasmic reticulum directed promiscuous biotin ligase, BirA*G3 [12], were inserted into the ubiquitously expressed Rosa26 locus [13]. BirA*G3 is a pre-cursor to TurBOX, generated in the directed evolution of E. coli BirA, that has a higher affinity for biotin and may be able to catalyze biotinylation prior to addition of exogenous biotin [12]. Conditional (CRE recombinase- and exogenous biotin-dependent) BirA*G3 activity resulted in rapid biotinylation of proteins trafficking through the secretory pathway and permitted the analysis of cellular secretomes through streptavidin affinity purification and quantitative mass spectrometry proteomics [12,13].

2. Results

2.1. Activation of birA*G3 in Sox2-BirA*G3 mice

We previously reported a Cre-inducible BirA*G3 cassette, inserted in the Rosa26 (R26) ‘safe-harbour’ locus in mouse embryonic stem cells (mESCs) (figure 1a) [13]. Here, we derived mouse strains, from three independently targeted mESCs containing the Cre-inducible BirA*G3 cassette (A11, B1 and C2). As expected, no differences were observed comparing the three lines consequently; we refer to data as if from a single line and we chose the A11 line for more extensive characterization.

The modified Rosa26 locus was designed to drive ubiquitous GFP expression downstream of a CAGGS (β-actin/CMV) regulatory sequence. GFP transcription blocks downstream expression of a BirA*G3 cassette and mKate2 reporter (figure 1a) [13]. BirA*G3 encodes a promiscuously active biotin ligase selected by protein evolution [12] with a signal peptide and ER retention signal, designed to target and retain BirA*G3 within the cell’s secretory pathway (figure 1b). mKate2 encodes a monomeric, photostable, pH resistant, low toxicity, bright far-red fluorescent protein designed as a cellular indicator of cells producing BirA*G3 and activating biotinylation of secretory pathway proteins on biotin administration (figure 1b) [20-22]. CRE recombination at loxP sites flanking the GFP cassette enables the tissue and time-dependent expression of BirA*G3 and mKate2 [13].

To initially examine BirA*G3 activity body-wide to broadly calibrate the model, we crossed the R26 BirA*G3 mice to the Sox2-Cre strain which results in recombination and predicted activation of BirA*G3 and mKate2 in the epi-blast, and thereafter, in all cell types of the conceptus [23]. As a consequence, cells with the unrecombined BirA*G3 allele will be green (GFP+), while those undergoing CRE-mediated recombination in Sox2-Cre; BirA*G3 (abbreviated as Sox2-BirA*G3) mice will be red (mKate2+) and BirA*G3 positive (figure 1c).

Viability and fertility were normal in Sox2-BirA*G3 mice. As predicted, whole-mount images of selected organs from control mice (BirA*G3+/ unless otherwise stated) were GFP+/mKate2-, with variable GFP intensity across organs (figure 1d), while Sox2- BirA*G3 mice were GFP+/mKate2+, demonstrating excision of the GFP cassette and activation of mKate2 reporter (figure 1e). High-magnification confocal images showed that BirA*G3 extensively colocalized with the ER resident protein calnexin (figure 1f; electronic supplementary material, figure S1A; electronic supplementary material, figure S1D). In Sox2-BirA*G3 brain, mKate2 signal and BirA*G3 expression was present in neurons (MAP2 representative cortical neurons), astrocytes (GFAP+), and microglia (IBA1+) (figure 1i; electronic supplementary material, figure S2C). Taken together, the data indicate a broad cell and tissue distribution for BirA*G3 and mKate2, though levels vary significantly depending on the cell type.

To compare BirA*G3 and the mKate2 reporter with another CAGGS (β-actin/CMV) driven reporters targeted to a similar position with the same transcriptional orientation in the R26 locus, we crossed Sox2-Cre mice with the widely used R26 TdTomato reporter mouse strain [24]. We observed homogeneous production of TdTomato in Sox2-Cre; TdTomato (Sox2-TdTomato) tissues (electronic supplementary material, figure S5A,B) suggesting that uneven BirA*G3/mKate2 in Sox2-BirA*G3 mice is specific to the BirA*G3 targeted locus.

To specifically examine the patchy hepatocyte distribution, mKate2low and mKate2high hepatocytes were sorted from Sox2-BirA*G3 liver and analysed by qPCR (electronic supplementary material, figure S5C). No BirA*G3 mRNA expression was detected in control liver as expected (electronic supplementary material, figure S5D). In Sox2-BirA*G3 liver, BirA*G3 mRNA was present in both mKate2low and mKate2high cells, though BirA*G3 mRNA levels were also lower in mKate2low cells (electronic supplementary material, figure S5D). By contrast, albumin mRNA expression was comparable in mKate2low and mKate2high populations (electronic supplementary material, figure S5E). The lower BirA*G3 mRNA expression was consistent with lower BirA*G3 protein level in mKate2low hepatocytes (electronic supplementary material, figure S5F-H). Protein biotinylation is still observed in mKate2low cells, although at lower levels.
Figure 1. (Caption overleaf.)
Figure 1. (Overleaf) Generation and characterization of Sox2-BirA*G3 mice. (a) Schematic diagram shows CRE-mediated excision at loxP sites removes the GFP cassette resulting in production of BirA*G3-ER and mKate2 fluorescent protein. Adapted from Droujinine et al. [13]. (b) Schematic diagram shows that proteins that reside or travel through ER would be biotinylated by BirA*G3. (c) Schematic diagram of mouse mating to generate Sox2-BirA*G3 mice. All cells in control mice are expected to express GFP. All cells in Sox2-BirA*G3 mice are expected to express mKate2. Adapted from Droujinine et al. [13]. (d) Native GFP and native mKate2 fluorescence in whole-mount organs (scale bar: 2 mm) and tissue sections (scale bar: 50 µm) of control mice. S.I: small intestine. L.I: large intestine. (e) Native GFP and native mKate2 fluorescence in whole-mount organs (scale bar: 2 mm) and tissue sections (scale bar: 50 µm) of Sox2-BirA*G3 mice. (f) Immunofluorescence staining shows expression of native GFP, native mKate2, BirA*G3, and ER marker Calnexin in the cell culture of isolated mouse embryonic fibroblasts (MEFs) from Sox2-BirA*G3 and control mice. Scale bar: 25 µm. (g) Immunofluorescence staining shows expression of native GFP, native mKate2, BirA*G3, and UMOD, which marks the thick ascending limb of Henle’s loop (TALH), in the kidney sections of Sox2-BirA*G3 and control mice. Scale bar: 50 µm. (h) Immunofluorescence staining shows expression of native GFP, native mKate2, BirA*G3, and MAP2, a cortical neuron-specific protein, in the brain sections of Sox2-BirA*G3 and control mice. Scale bar: 50 µm. Unlike BirA*G3, mKate2 expression is not restricted to the ER, which may result in varied signal intensity due to differences in cell morphology.

2.2. Protein proximity labelling in Sox2-BirA*G3 mice

Next, we characterized biotinylation of proteins in Sox2-BirA*G3 and control mice. Mice were fed biotin in chow (2000 ppm biotin) for 5 days ad libitum. Brain samples were collected and analysed by western blotting to detect biotinylated proteins through streptavidin-IRDye 800CW binding. Prominent biotin/BirA*G3-dependent labelling was observed for a broad range of protein species (figure 2a). Additionally, we observed biotin/BirA*G3-independent labelling of two prominent proteins in control and experimental brain samples (figure 2a) which most likely represent cytoplasmic biotin conjugates with pyruvate carboxylase (approx. 130 kDa) and methylcrotonyl-CoA carboxylase/propionyl-CoA carboxylase (approx. 75 kDa) [14] (figure 2a; electronic supplementary material, figure S3A).

Examining a broad range of total protein lysates from a range of organs showed strong evidence for BirA*G3-dependent biotinylation of target tissues (figure 2b; electronic supplementary material, figure S3B). Notably, each tissue displayed a unique biotinylation pattern (figure 2b), suggesting diverse secretomes across tissues. Further, analysis of serum detected a robust BirA*G3-dependent biotinylated protein signature (figure 2c; electronic supplementary material, figure S3C). Comparable serum labelling was obtained through multiday (7 days) labelling with biotin addition to either water (5 mM) or chow (2000 ppm) (figure 2d; electronic supplementary material, figure S3D,E). Subcutaneous injection of biotin (180 mM, 100 µl) resulted in significant labelling of serum proteins within 1 hour of injection with an increased recovery of proteins up to 12 h (figure 2e; electronic supplementary material, figure S3F). These results highlight the rapid kinetics for secretion of biotinylated proteins though comparison with protein products observed with continuous 7-day labelling in chow demonstrates enhanced labelling of lower molecular weight protein species (figure 2e).

The accumulation of biotinylated proteins could also be visualized by streptavidin-dependent immunofluorescence directly in sections of multiple tissue types. Abundant biotinylated proteins were observed in Sox2-BirA*G3 muscle, heart, brain, compared to their control counterparts (figure 2f; electronic supplementary material, figure S4A,B), consistent with CRE-dependent biotinylation. For the liver and kidney, endogenous biotin stores result in equivalent strong streptavidin signals in both Sox2-BirA*G3 and control mice (electronic supplementary material, figure S4C,D), even though western analysis shows Sox2-BirA*G3-dependent labelling of proteins (figure 2h).

To look for detrimental effects of BirA*G3 in Sox2-BirA*G3 mice, we performed haematocrit and eosin staining and bulk mRNA-sequencing. No pathological changes were detected by histology comparing Sox2-BirA*G3 and control mice fed biotin chow for 7 days (electronic supplementary material, figure S7A). Principal component analysis (PCA) of mRNA-sequencing comparing liver, brain and kidney between control (n = 2 per tissue) and Sox2-BirA*G3 (n = 2 per tissue) mice showed a tight clustering by tissue (electronic supplementary material, figure S7B). Differential gene expression analysis of all Sox2-BirA*G3 samples compared to all control samples showed a total of only four differentially expressed genes (DEGs) (electronic supplementary material, figure S7C). Furthermore, analysis of ER stress, unfolded protein response (UPR), and cell death markers showed no difference (adj. p-value > 0.05) at the transcript level between Sox2-BirA*G3 and control samples (electronic supplementary material, figure S7D). Additionally, western blotting showed no indication of ER stress or UPR in Sox2-BirA*G3 and control livers compared to ER stress-induced (tunicamycin treated) mouse embryonic fibroblasts (MEFs; electronic supplementary material, figure S7E-J).

2.3. MS-based tissue proteomics of biotinylated proteins

As a prelude to MS analyses of each of the three tissues (liver, brain, and kidney), affinity purified biotinylated proteins were visualized by both streptavidin western analysis and silver
stain. The biotinylation banding patterns were comparable to those observed in total protein lysates of corresponding samples, suggesting efficient and unbiased enrichment of biotinylated proteins (S8A-D). Silver stain showed specific bands in Sox2-BirA*G3 samples, despite a strong background of non-specific binding of unlabelled proteins to streptavidin-conjugated beads in all three tissues (S8E-H).

The biotinylated proteomes of the three tissues were defined by quantitative TMT-based LC-MS/MS following streptavidin bead enrichment from Sox2-BirA*G3 and control liver, brain and kidney. Of the thousands of proteins detected and quantified in the Sox2-BirA*G3 tissues, several hundred proteins were found to be significantly enriched in liver (n = 189), brain (n = 200), kidney (n = 578) compared to their control counterparts (log2 fold change (FC) > 1.0 and adj. p-value < 0.05) (figure 3a,b; electronic supplementary material).
Figure 3. (Caption overleaf.)
Figure 3. (Overleaf) Identification of biotinylated proteins in Sox2-BirA*G3 liver, brain, and kidney tissues by mass spectrometry. (a) Representative schematic of TMT-based 6plex LC-MS/MS workflow for liver Sox2-BirA*G3 (n = 3) and control (n = 3) samples. The same 6plex LC-MS/MS design was used for brain and kidney for individual MS runs. (b) Volcano plots of proteins detected in liver of Sox2-BirA*G3 mice compared to control mice after streptavidin pulldown. Log2 FC were plotted on the x-axis and $-10 \log_{10} (p$ value) were plotted on the y-axis. Significantly enriched proteins (adj. p-value < 0.05 and log2 FC > 1.0) in Sox2-BirA*G3 A mice compared to control mice are shown in green. (c) Relative abundance (log2FC) of representative proteins in liver, brain, and kidney. ES are labelled for the abundant proteins in their corresponding tissues. (d) Venn diagram showed the overlap of enriched proteins (ES method) among liver, brain, and kidney. (e) Shared enriched proteins among three tissues (113 proteins) were predicted with SignalP/TMH (left) and analysed with DAVID analysis for cellular components annotation (right). Gene ratio indicates the percentage of genes annotated with the term over the total number of genes in the list. (f) Liver-specific enriched proteins (115 proteins) predicted with SignalP/TMH were analysed with clusterProfiler (3.16.1) EnrichGO analysis. Left: a pie chart displayed the distribution of liver-specific proteins with SignalP/TMH prediction. Right: dot plots displayed the functional categorization of liver-specific enriched based on EnrichGO annotation, and the number of each category is displayed based on biological process. Gene ratio indicate the percentage of genes annotated with the term over the total number of genes in the list.

material, figure S9A–C and data S1–S3). Principal component analysis (PCA) demonstrated tight grouping of biological replicates within the Sox2-BirA*G3 group, indicating similar proteomic profiles among these samples (electronic supplementary material, figure S10A–C). Relative abundance of representative signature proteins for each tissue is shown in figure 3c.

An enrichment score (ES) was generated as a measure of the abundance of proteins in Sox2-BirA*G3 replicate samples compared to control samples, as previously described [13,25]. Briefly, for 6plex TMT ratios of each tissue, TMT ratios were calculated by comparing each of the three Sox2-BirA*G3 over each of the three control samples, giving rise to nine different datasets of TMT ratios. Then, we determined the false positive rate (FPR) based on proteins that are retained or traffic through the ER (positive control) and proteins that do not (negative control) (electronic supplementary material, figure S10D,E; see section 4.16 below). The number of TMT ratios for each protein that pass the TMT ratio cutoff based on an FPR of 0.1 was defined as enrichment score (ES) (electronic supplementary material, figure S10D). A protein where all its ratios pass the cutoffs has an enrichment score (ES) of 9 (highest confidence), where a protein where none of its ratios pass the cutoffs has a score of 0 (lowest confidence). Assigning an ES ≥5 as the cutoff (electronic supplementary material, figure S10F) maximized the recovery of ER-targeted proteins with high specificity (electronic supplementary material, figure S10G–I).

In the conventional secretory pathway, proteins with either a signal peptide (SignalP) or transmembrane helix (TMH) travel through ER-Golgi apparatus. We used SignalP (v. 5.0) [26–28] and TMHMM (v. 2.0) [29–31] to predict the presence of SignalP and TMH on each tissue sample. Our analysis revealed that higher ES correlated with higher ratios of proteins with predicted SignalP/TMH (electronic supplementary material, figure S9D–F), suggesting that proteins with higher ES contain ‘hits’ of higher confidence. Comparing across liver, brain and kidney, 113 proteins were present in all three organ samples (figure 3d) and over 93% of these (106) showed a SignalP and/or TMH (figure 3c). Applying Gene Ontology (GO) functional category enrichment analysis to identify characteristic cellular component attributes of these tissue-shared proteins highlighted extra-cellular exosome and ER-Golgi terms in DAVID analysis (figure 3c; electronic supplementary material, figure S9G), consistent with ER localization of BirA*G3 and the expected labelling of proteins in the secretory pathway. These proteins are enriched in ER-Golgi related functions (electronic supplementary material, figure S9H).

TissueEnrich [32], a tool for tissue-specific gene enrichment, was applied to liver, brain, and kidney-specific proteins (electronic supplementary material, figure S11A–C). Notably, all organ samples showed a strong enrichment profile for the expected organ type. Significant enrichment of SignalP/TMH was also seen in tissue-specific hits (figure 3f; electronic supplementary material, figure S11D, E). For biological process analysis, significantly enriched terms in liver showed unique features of liver function, which include lipid catabolic process and blood coagulation (figure 3f). Similarly, significantly enriched terms in brain showed unique features of brain function, such as synapse organization and assembly (electronic supplementary material, figure S11F). However, kidney-specific functional annotation terms were not observed among top terms in the kidney data (electronic supplementary material, figure S11G). The large majority of kidney-enriched transporter and channel proteins are confined to short segments within the renal epithelium, organ-wide enrichment is likely to select for more secretory pathway proteins that are broadly distributed, or particularly abundant proteins such as UMOD, that are segmentally restricted (figure 3c; electronic supplementary material, figure S11E,G).

2.4. Analysis of the serum secretomes of labelled tissues

To examine biotinylated proteins secreted into the serum of female Sox2-BirA*G3 mice, serum from labelled mice was enriched using streptavidin-conjugated beads and bound fractions from secretomes of Sox2-BirA*G3 (n = 3) mice and associated controls (n = 3) were analysed by MS (6plex) (figure 4a; electronic supplementary material, figure S13A and data S4–S5). PCA demonstrated tight clustering of the Sox2-BirA*G3 group (electronic supplementary material, figure S12A,B). The affinity purified serum proteome identified by TMT-based MS has minimal background compared with tissue samples (electronic supplementary material, figure S13B). Accordingly, we used a $\log_{2} FC > 1.0$ and adj. p-value < 0.05 to define BirA*G3 serum enriched proteins over control serum to score enriched serum proteins from two independent MS runs of the same samples processed by two different research laboratories at different universities and run by the same MS group (plex1 and plex2).

The majority of serum proteins (70.6%) were identified in both MS runs (plexes) (electronic supplementary material, figure S12C), highlighting the reproducibility of this method. Most serum enriched proteins are annotated to
Figure 4. (Caption overleaf.)
contain a SignalP or TMH (64.8% of total and 76.6% of proteins shared between two plexes; electronic supplementary material, figure S13C) accounting for enriched cellular component terms for extracellular exosome/space/region (electronic supplementary material, figure S12D-E). As expected, biological process annotation for serum enriched proteins predicted with SignalP/TMH are enriched in immunity and complement-related terms (electronic supplementary material, figure S12F). To identify potential proteins secreted from tissues to serum, we compared the enriched proteins from serum with those from the three tissues and found 220 total overlapping proteins between serum and (i) liver, (ii) brain or (iii) kidney (figure 4b). A list of selective serum enriched proteins was shown in a heat map representing log2FC between Sox2-BirA*G3 and control sample, including classic hepatocyte secreted proteins, such as angiotensinogen (AGT) and albumin (ALB), brain cell marker, neuron neuropetide oxytocin (OXT), as well as kidney markers, uromodulin (UMOD) and insulin-like growth factor-binding protein-3 (IGFBP3) (figure 4c).

Out of the 220 tissue-serum shared proteins, 65.9% were annotated as secreted by Uniprot/NCBI (figure 4d), and the majority (85.9%) were predicted to have a SignalP or TMH (figure 4e). These tissue-serum shared proteins were then analysed for GO term enrichment analysis (figure 4f; electronic supplementary material, figure S13D,E). DAVID analysis demonstrated significant enrichment in extracellular annotations in the cellular component annotation (figure 4f). The biological process GO term for these tissue-shared proteins with predicted SignalP/TMH are enriched in coagulation and haemostasis (electronic supplementary material, figure S13E).

Ectodomain shedding is an important post-translational mechanism for regulating the function of cell surface proteins, which involves the proteolytic cleavage of transmembrane (TM) cell surface proteins, and release of circulating, soluble form [33,34]. Five well-documented TM proteins (LIFR, EGFR, VCAM1, Slc38a10 and PIGR) appeared in our serum proteomic datasets. For each, sequenced peptides exclusively mapped to the annotated extracellular domains, consistent with ectodomain shedding (figure 4g; electronic supplementary material, figure S13G–I). While ectodomain shedding has been reported for LIF receptor subunit α (LIFR) [35], polymeric immunoglobulin receptor (PIGR) [36], epidermal growth factor receptor (EGFR) [37], and vascular cell adhesion molecule 1 (VCAM1) [38], solute carrier family 38 member 10 (Slc38a10) have not been directly linked to shedding, though an extramembrane fragment of human Slc38a10 has been reported in plasma [39] (figure 4g,h; electronic supplementary material, figure S13F–H). Approximate protein cleavage sites can be assigned on the basis of the most N- or C-terminal peptides (figure 4g,h; electronic supplementary material, figure S13F–H). For example, the in vivo cleavage site for LIFR is predicted to be between the fifth fibronectin type 3 (FN3) domain and the TM domain, similar to a previous report [15]. Intriguingly, we also detected thioredoxin (TXN) and Annexin1 (ANXA1) in our dataset (figure 4c), each of which are reported to be released from the cell through unconventional protein secretion (UPS) [40–42], suggesting this method can give further insight into UPS, with potentially other candidates among the detected proteins lacking a predicted SignalP or TMH domain.

To confirm key hits in these secreted serum protein data, we performed western blot analysis of 11 proteins observed in the MS analyses of biotinylated serum proteins (electronic supplementary material, figure S13I,J). These validation analyses were carried out in both male and female mice (electronic supplementary material, figure S13 J). Consistent with the MS data, all 11 proteins were enriched in Sox2-BirA*G3 serum after affinity purification.

2.5. Biotin labelling of the hepatocyte secreome

We next applied and extended the approach to determine the feasibility of identifying a tissue cell type specific secreome. For this, we used Alb-Cre to enable hepatocyte-specific biotinylation of proteins in the liver [59]. Biotin chow was given to adult Alb-Cre; BirA*G3 (abbreviated as Alb-BirA*G3) and control mice for 5 days, and then well-perfused tissues were collected after exsanguination (electronic supplementary material, figure S14A). mKate2 was specifically detected in the Alb-BirA*G3 liver (figure 5a; electronic supplementary material, figures S15A and S14B). BirA*G3 showed a homogeneous distribution restricted to the hepatocytes of Alb-BirA*G3 liver (figure 5b). BirA*G3 activity in Alb-BirA*G3 liver enabled specific protein biotinylation in liver protein lysates to be compared with control mice (figure 5c). Alb-BirA*G3 serum samples showed strong, specific protein biotinylation relative to serum from biotin-administered controls lacking active BirA*G3 (figure 5d). The biotinylation banding patterns of affinity purified biotinylated proteins were comparable to those observed in total protein lysates of liver and serum samples (figure 5c,d; electronic supplementary material, figure S15B–F). Silver staining of affinity purified biotinylated proteins showed specific bands in Alb-BirA*G3 serum, with similar size distributions between sexes (figure 5e).
2.6. Identification of hepatocyte secretome

Streptavidin purified serum proteins of Alb-BirA*G3 (n = 4) and control (n = 4) mice were analysed by LC-MS/MS (8plex) to identify secreted hepatocyte proteins (electronic supplementary material, data S6). PCA demonstrated male (n = 2) and female (n = 2) Alb-BirA*G3 samples cluster together, indicating similar hepatocyte secretome between the two sexes (figure 5f). Approximately 80% of proteins (181/189) were specifically enriched (log2FC > 1.0 and adj. p-value < 0.05) in the Alb-BirA*G3 group (figure 5g). A heatmap showed very low protein abundance in the control
group, consistent with the absence of biotinylated proteins in the control serum samples (figure 6; figure 5e). To determine cell labeling specificity, TissueEnrich on serum enriched proteins in the Alb-BirA*G3 group showed a highly liver-specific organ profile using the mouse ENCODE dataset (figure 6b).

Over half (69.1%) of the proteins enriched in Alb-BirA*G3 serum were annotated as secreted and a majority (77.9%) of these proteins were predicted with SignalP/TMH (figure 6c,d). Functional annotation showed significant enrichment for ‘extracellular’ terms (electronic supplementary material, figure S16A-B). Biological process annotation for the subset of serum enriched proteins with predicted SignalP or TMH, specific to the Alb-BirA*G3 group demonstrated enriched terms for key liver protein functions: coagulation and haemostasis (figure 6e) [41].

Recently, other groups have used adeno-associated virus carrying ER-TurboID under the control of the hepatocyte-selective Tbg (thyroxine binding protein) promoter (AAV-Tbg-ER-TurboID) [15] and adenovirus carrying Sec61b (Protein transport protein Sec61 subunit beta)-TurboID [14] to label the liver secretome of mice. While these two previous datasets only share 5 proteins, both displayed a significant overlap (23/64 and 25/27, respectively) with our Alb-BirA*G3 dataset, highlighting the power of the transgenic strain (figure 6f).

Further, the 137 proteins uniquely detected in the Alb-BirA*G3 dataset displayed a strong correlation with liver-specific proteins and the majority (67.2%) was as predicted with SignalP/TMH (electronic supplementary material, figure S16C,D). Well-characterized hepatocyte secreted proteins, including APOE, AGT and IGFBP2, were only seen in figure S16C,D). Well-characterized hepatocyte secreted proteins, including APOE, AGT and IGFBP2, were only seen in the Alb-BirA*G3 dataset, highlighting an improved coverage of secreted liver proteins using this strategy.

Finally, we asked if secreted proteins can be labelled in the organ of origin and subsequently detected in a destination organ after affinity purification. The renin-angiotensin-aldosterone system regulates blood pressure in conjunction with the liver, kidney and lung [43]. In this pathway, angiotensinogen (AGT) is produced by the liver and sequentially converted by kidney and lung enzymes to vasoconstrictive angiotensin II (AGTII). Liver released AGT is reabsorbed from the renal filtrate by proximal tubule cells in the kidney [44]. Consistent with this view, AGT was detected in the liver, serum and kidney total protein from both Alb-BirA*G3 and control mice (figure 6g–i), but following streptavidin affinity purification, AGT was only detected in the liver, serum, and kidney from Alb-BirA*G3 mice, but not in control mice (figure 6g–i). Together, these data support the trafficking of AGT produced in liver through the blood stream with uptake in the kidney.

3. Discussion

The genetic system presented here allows for rapid (within 1 hour) and broad (representative cell types) in vivo biotinylation of proteins within the secretory pathway of the living mouse. Long-term studies of BirA*G3 production show no obvious pathology. Although we do observe cell-type variability in BirA*G3 levels, that may be countered using tissue or time-dependent CRE activator lines as we show here comparing Sox2-BirA*G3 with CAGG-BirA*G3.

TMT-based LC-MS/MS captured a number of well-characterized secreted proteins in Sox2-BirA*G3 serum, including proteins with hormone-like properties, known to circulate at µg ml⁻¹ (ADIPOQ [45]) and ng ml⁻¹ (ANGPTL [46], MST1 [47], MSTM [48], RETN [45], CXCL7 [49], IGF1 [50], FGFl5 [51] levels. Interestingly, these represent examples of proteins derived from multiple tissues: adipose tissue (ADIPOQ, Adipsin [52], and RETN), muscle (MSTM), immune cells (CXCL7), liver (IGF1) and intestine (FGF15). Furthermore, the approach provided insight into ectodomain shedding and UPS.

Three research groups have reported various applications of proximity ligation techniques to profile the mammalian secretome. Two groups [14,15] profiled the mouse hepatocyte secretome via viral delivery of TurboID. Viral vectors however present unavoidable problems, such as viral tropism, which limits the use in some tissues. For instance, kidney has been quite difficult to transduce with any viral vector currently available [53]. For brain, viruses need to be intracranially administered due to the blood-brain barrier, for higher transduction efficiency and brain-specific targeting and viral vectors have broad tropisms for neural and glial cell types [18,19]. Directly comparing the Alb-BirA*G3 mouse strain to the two viral studies profiling the hepatocyte secretome, our study showed an improved coverage for relevant hepatocyte proteins which may reflect the stable and efficient expression of BirA*G3, as well as other differences in the downstream analytical pipeline.

A third group [17] reported a transgenic mouse expressing BiolID, a less efficient biotin ligase (as shown in various studies [12,17,54]), in the endoplasmic reticulum. Characterization of the endothelial and muscle secretome showed few secreted proteins, probably due to the low activity of BiolID2 in ER lumen [12]. Furthermore, the kinetics of labelling is reported to be much faster with BirA*G3, suggesting the line reported here will be better suited to dynamic labelling studies [12].

Incomplete retention of TurboID-KDEL in the endoplasmic reticulum has been documented [14]. In our MS data, we identified tryptic peptides for BirA*G3 in both Sox2-BirA*G3 and Alb-BirA*G3 serum. However, using western blotting, we did not detect BirA*G3 in total serum proteins or affinity purified proteins in Sox2-BirA*G3 serum (figure 2c; electronic supplementary material, figure S8D). Biotinylation, with biotin ligase, is an ATP-dependent reaction, and ATP is normally restricted within the cell [55]. Biotin concentration within the cell and the concentration of proteins in the secretory pathway, probably also impact the efficiency of the enzymatic reaction. Together, these kinetic arguments make it unlikely there is a significant impact from low levels of circulating BirA*G3 on biotinylation of serum proteins. This is supported by the Alb-BirA*G3 data which demonstrated a robust biotin profile for specific liver-derived proteins with no labelling of abundant proteins secreted by other organs, identified in the Sox2-BirA*G3 study.

Recent studies have found that many cytosolic proteins lacking a SignalP/TMH (leaderless cargo) are released through a type III UPS mechanism [40]. ANXA1 is one such protein detected in our serum dataset (figure 4c). ANXA1 entry to the secretory pathway has been linked to the ER-Golgi intermediate compartment (ERGIC), a distinct compartment to the ER. Labelling in our data may reflect the transient movement of BirA*G3-ER into the ERGIC[40]. Recent studies show that approximately 13% of the human protein-coding genes encode for roughly 2640 secreted
Figure 6. Analysis of hepatocyte-secreted serum proteins in Alb-BirA*G3 mice. (a) Cluster heatmap of proteins enriched in Alb-BirA*G3 serum relative to controls. Log2 expression values are shown by colour and intensity of shading. Blue, low; red, high. Serum-enriched proteins in the Alb-BirA*G3 group are highlighted in black colour in the hierarchical clustering. (b) Enriched proteins in Alb-BirA*G3 serum were analysed with TissueEnrich to calculate tissue-specific gene enrichment. (c) Pie chart displayed the number of enriched proteins in Alb-BirA*G3 serum that are annotated as secreted by UniProt/NCBI. (d) Pie chart displayed the distribution of enriched proteins in Alb-BirA*G3 serum with predicted SignalP/TMH. (e) Enriched proteins in Alb-BirA*G3 serum predicted with SignalP/TMH were analysed with clusterProfiler (3.16.1) EnrichGO analysis for biological process. Gene ratio indicates the percentage of genes annotated with the term over the total number of genes in the list. (f) Area-proportional venn diagram showed the overlap among enriched proteins in Alb-BirA*G3 serum, AAV-Tgb-ER-TurboID plasma [15], and Sec61-TurboID plasma [14]. (g–i) AGT levels in total protein lysates (upper) and affinity purified biotinylated proteins (lower) from liver (f), serum (g) and kidney (h) in Alb-BirA*G3 mice compared to control mice. Input: 500 µg proteins. Bead lanes are affinity purification negative control without protein input. Each lane is a biological replicate from individual mice (n = 3 per genotype).
proteins, but relatively few have been functionally annotated and characterized [4]. Since many protein functions are conserved within mammals, we anticipate our platform will serve as a valuable resource for deciphering the mammalian secretome, under healthy or diseased conditions.

4. Material and methods

4.1. Animal studies

Institutional Animal Care and Use Committees (IACUC) at the University of Southern California reviewed and approved all animal work as performed in this study. All work adhered to institutional guidelines. mESCs (B6(Cg)-Tyr < c-2 J > /J, Stock No.: 00058, The Jackson Laboratory) [56] carrying the loxP-flanked BirA*G3 were aggregated with 8-cell C57BL/6J embryos to obtain chimeras. The resulting chimeric Sox2-Cre [43x372]mice were backcrossed to C57BL/6J for 13 generations) [57] females to remove attB-neoR-attP cassette. The resulting laboratory, backcrossed to C57BL/6J for 13 generations) [59] were used as described above.

4.2. In vivo assays

For all in vivo assays, tissues were collected as follows. Mice were euthanized at 8–32 weeks, blood was then collected from the inferior vena cava, followed by perfusion with 1 x cold DPBS (Dulbecco’s Phosphate Buffered Saline). The blood was allowed to clot at room temperature for 30 min and then spun down at 2000g for 15 min at 4°C. The serum was then collected and spun again at 2000g for 15 min at 4°C, then removed to a fresh tube and flash frozen (in liquid nitrogen) before being stored at −80°C until being used. After perfusion, tissues were collected and rinsed in 1 x cold DPBS before being minced with a razor blade and aliquoted into tubes. Tissues were then flash frozen and stored at −80°C until being used.

For the biotin administration method study, Sox2-BirA*G3 and control mice at 8–12 weeks were given biotin via water (n = 2) (5 mM, pH 7.4; Sigma B4639-5G), chow (n = 2) (2000 ppm; LabDiet, SWLP), or both (n = 2) for 7 days. After 7 days, serum was collected and stored as described above.

For the Cre dependence labelling study, Sox2-BirA*G3 and control mice at 8–12 weeks were given biotin chow (2000 ppm) or normal diet (LabDiet 5053) for 7 days. After 7 days, serum, liver, kidney, and brain were collected as described above.

For the BirA*G3 temporal labelling assay, Sox2-BirA*G3 (n = 3/timepoint) and control mice (n = 3) at 8–12 weeks were given biotin by subcutaneous injection (100 µl 180 mM sterile biotin water, pH 7.4) and water (5 mM, pH 7.4) after the injection until collection. Serum, liver, kidney and brain were collected from mice at 1, 2, 4, 6, 10 and 12 h post-injection. Serum was collected as described above.

For all other mouse studies, Sox2-BirA*G3 and control mice at 8–32 weeks were given biotin chow (2000 ppm) and control mice at 8–32 weeks were given rifamycin (2 mg/40 g body weight in corn oil; Sigma cat. T5748-1G) was administered by abdominal injection twice, 3 days apart prior to biotin studies.

4.3. In vitro assays

To generate endoplasmic reticulum stress positive controls 80% confluent mouse 3T3 cells (NIH/3T3 ATCC cat. CRL-1658) were treated with 5 µg ml−1 Tunicamycin (R&D Systems 590507) in DMSO for 5 h. Cells were then briefly rinsed with 1X DPBS and then scraped off the plate in fresh DPBS. Cells were pelleted by centrifugation at 300g for 10 min. Supernatant was then removed, and dry cell pellets were snap frozen on dry ice and then stored at −80°C until used. For high-resolution calnexin-BirA staining, mouse embryonic fibroblasts (MEFs) isolated from Sox2-BirA*G3 (n = 1) and control (Sox2-Cre) (n = 1) mice were cultured on coverslips coated with 0.1% gelatin. Cells at 70% confluence were briefly rinsed in 1X DPBS and then fixed in 4% PFA for 10 min followed by 3 washes in 1X DPBS. Coverslips with cells were stored at 4°C in 1X DPBS in the dark until used.

4.4. Haematoxylin and eosin staining

Tissues (n = 3 per genotype) were collected and fixed in 10% phosphate-buffered formalin overnight at 4°C. Paraffin sections were prepared using standard procedures. The tissue sections were deparaffinized by immersing in xylene and rehydrated through graded alcohol series, dyed with haematoxylin and eosin (H&E) and then rinsed with water. All slices were dyed with H&E and then rinsed with water. Each slide was dehydrated through graded alcohols. Tissue sections were finally soaked in xylene twice. The sections were examined under a light microscope for evaluation of pathological changes.

4.5. Frozen tissue preparation and sectioning

Briefly, tissues were harvested from DPBS-perfused mice and then fixed in 4% paraformaldehyde for 2 h at 4°C. Tissues (except brain: fixation overnight and 30% sucrose for 48 h) were then washed 3 times in 1X DPBS with calcium and magnesium before being incubated in 30% sucrose overnight. The following day, tissues were washed in OCT 3 times to remove excess sucrose and then embedded in OCT (VWR, 25608-930) and frozen in a dry-ice ethanol bath before being stored at −80°C. Tissue blocks were thawed to −20°C overnight and then cryosectioned at 10–16 µm at −20°C and placed on glass slides. Slides were then stored at −80°C until immunostained.

4.6. Immunofluorescent staining and confocal microscopy

Frozen sectioned tissues were thawed at room temperature for 10 min. Slides were then rinsed in 1X DPBS with calcium and magnesium for 10 min. Coverslips with cells were
removed from 4°C 1X DPBS and treated as slides. Slides were permeabilized in 0.25% Triton-X100 (Sigma X100–500ML) for 5 min, then incubated in blocking buffer (2.0% Sea Block (Thermo, 37527) + 0.125% Triton-X100 in 1× DPBS with calcium and magnesium) for 1 hour at room temperature. Slides were then incubated in primary antibody (electronic supplementary material, table S1) diluted in blocking buffer (2.0% Sea Block, H3570) in 1× DPBS with calcium and magnesium for 10 min at room temperature. Slides were then washed twice in 1× DPBS with calcium and magnesium for 5 min each. Slides were mounted in Immumount (Thermo, 9990402) and imaged at 40× or 63× using the Leica SP8 confocal microscope. Brain images are median intensity projections from z-stacks of 0.5–1.0 µm steps except the whole brain tiles scans. All images presented represent at least three images per tissue and 2 biological samples per whole brain tiles scans. All images presented represent at least three images per tissue and 2 biological samples per whole brain tiles scans. All images presented represent at least three images per tissue and 2 biological samples per whole brain tiles scans. All images presented represent at least three images per tissue and 2 biological samples per whole brain tiles scans.

4.7. Hepatocyte isolation and qPCR

We used the classic two-step collagenase perfusion technique to isolate primary mouse hepatocytes. Briefly, the Sox2- BirA*G3 and control livers were perfused by perfusion medium (GIBCO #17701-038) through Inferior vena cava. Then, the livers were dissociated by liver digested medium (GIBCO 17703-034), which is incubated 30–60 min in 37°C before use. mKate2low and mKate2high hepatocytes were sorted by flow cytometry based on mKate2 levels. Total RNA was extracted using RNeasy mini kit (Qiagen, 74104). cDNA was synthesized using SuperScript™ IV VILO™ Master Mix (Thermo #11766050). qPCR was performed with Luna Universal qPCR Master Mix Protocol (New England Biolab #M3003) on a Roche LightCycler 96 System. Primers used in RT-qPCR are listed as follows:

**BirA*G3:**
- Forward: CTCCCCGTGTTGACTCCTAC
- Reverse: CTCCCCGTGTTGACTCCTAC
- *Alb*:
- Forward: GTCTTAGTGAGGTGGAGCATGACAC
- Reverse: CTCCCCGTGTTGACTCCTAC
- *Gapdh*:
- Forward: CATGGGCTTCTCGGTTCCTCA
- Reverse: CCTGCTTACCACCCAACCTTCTTGAT

The delta-delta Ct method, also known as the 2ΔΔCt method, was used to calculate the relative fold gene expression.

4.8. RNA sequencing and analysis

Mouse samples were collected for RNA as described above. After collecting samples were immediately stored in RNALater (Sigma 90901-100ML) at 4°C overnight before being stored at −80°C until processing. Before RNA extraction samples were removed from RNALater and briefly rinsed in RNase free dH2O. Total RNA was then prepared from Sox2-BirA*G3 and control liver, kidney, and whole brain (n = 2 per genotype, all female) using RNeasy mini kit (Qiagen 74104) according to kit instructions with the following modifications. For whole brain, brains were homogenized in 1400 µl RLT buffer. Brain samples were then further diluted 1:25 in fresh RLT buffer (total final volume 350 µl) to avoid overloading the columns. Liver and kidney tissue samples were processed exactly following kit instructions. RNA integrity was determined using Agilent TapeStation 4200. Samples were then sequenced by Washington University in St. Louis School of Medicine Genome Technology Access Center (GTAC) using the following methods. Total RNA integrity was determined using Agilent BioAnalyzer. Library preparation was performed with 10 ng of total RNA with Bioanalyzer RIN score of greater than 8.0. ds-cDNA was prepared using the SMARTer Ultra Low RNA kit from Illumina Sequencing ( Takara-Clontech) per manufacturer’s protocol. cDNA was fragmented using a Covaris E220 sonicator using peak incident power 18, duty factor 20%, cycles per burst 50 for 120 s. cDNA was blunt ended, had A base added to the 3’ ends, and then had Illumina sequencing adapters ligated to the ends. Ligated fragments were then amplified for 12–15 cycles using primers incorporating unique dual index tags. Fragments were sequenced on an Illumina NovaSeq-6000 using paired end reads extending 150 bases. Raw paired end read files were then trimmed using Trimmomatic v. 0.38 and then aligned to GRCh39 using STAR v. 2.7.0e with default options. Read counts were quantified using RSubread_2.2.6 (featureCounts) without multimapping. Read counts were then analysed using DESeq2 v. 1.28.1 in R (v. 4.0.0) using standard approaches with cutoffs of log2 fold change > 2.0 and adjusted p-value < 0.05.

4.9. Protein lysate preparation

Protein lysates were prepared as described previously [13] with the following modifications. Homogenized in 500 µl RIPA complete lysis buffer (RIPA buffer (ThermoFisher, 89901)) with 1× complete EDTA-free protease inhibitor cocktail (Sigma, 11873580001), 1 mM benzamidine hydrochloride (VWR, TCBO013-100G), 4 µM pepstatin A (Sigma, EI10), 100 µM PMSF (Sigma, 11359061001) and bead homogenized using stainless steel beads (NextAdvance, SSD14B-RNA) for 5 min at setting 10, Bullet Blender Storm (NextAdvance, BT24M). Samples were then centrifuged at 14 000 g for 15 min at 4°C. Supernatants were transferred to protein LoBind (Eppendorf) tubes. Protein lysate concentrations were determined using Pierce BCA (Thermo cat. 23227) microplate assay per manufacturer’s instructions. Lysates were then stored at −80°C.

4.10. Streptavidin beads pulldowns

Streptavidin pulldowns were performed as described previously [13] with modifications. Streptavidin magnetic beads (Thermo cat. 88817) were resuspended in lysis buffer (above) by magnetic separation (BioRad 1614916). We tested a series of volumes of beads and washing conditions and found that 5 µl beads per 100 µg protein together with the following washing conditions are sufficient (results...
available upon request). Pulldown reactions were set up in 450 µl lysis buffer with 5 µl beads per 100 µg protein. Pulldowns were then incubated overnight at 4°C in a wheel rotator. The following day, pulldown reactions were washed 2× in lysis buffer, then 1× in 2 M Urea in 10 mM Tris, and finally 2× lysis buffer. After the final wash, lysis buffer was removed and beads were either boiled in 12 µl 1× loading buffer (Li-Cor 928-40004) with 1.43 M β-mercaptoethanol or resuspended in 100 µl lysis buffer and flash frozen, for western blotting and mass spectrometry respectively.

### 4.11. Silver stain analysis

Silver staining was done using Richard J. Simpson’s protocol from Cold Spring Harbor (CSH) or using the Silver Stain PlusTM kit (BioRad 1610449). Gels were fixed in a 50% methanol (VWR BDH1135-4LG), 5% glacial acetic acid (VWR 97064-482) solution, gently shaking at room temperature for 20 min. Gels were then incubated in 50% methanol for 10 min, gently shaking at room temperature, followed by a 10 minute incubation in dH2O. Gels were then soaked in 0.02% sodium thiosulfate (Sigma 72049) for 1 min and then in dH2O for 1 min, twice. Gels were then incubated in chilled 0.1% silver nitrate (Sigma 209139) for 20 min, gently shaking at 4°C in the dark. Gels were then rinsed twice in dH2O for 1 min each. Gels were developed in a 2% sodium carbonate (Sigma 222321) and 0.04% formaldehyde (Thermo 28906) until desired intensity was reached. Developing was stopped with a 5% glacial acetic solution and gels were stored in 1% glacial acetic acid until being discarded.

### 4.12. Fluorescent western blot analysis

Western blots were performed with standard protocols and the following modifications. Equal amounts of total protein lysate were loaded per sample per reaction with 1× Li-Cor loading buffer (Li-Cor, 928-40004) with 1.43 M β-mercaptoethanol. For streptavidin pulldowns, beads were resuspended in 12 µl 1× Li-Cor loading buffer (Li-Cor, 928-40004) with 1.43 M β-mercaptoethanol. All samples were then boiled at 95°C for 5 min to elute, then briefly spun down and kept on ice prior to loading. Total protein samples and pulldown elutes were loaded on 10% SDS acrylamide gels and ran in standard 1× SDS-Running buffer with Li-Cor 5 µl one-colour molecular marker (Li-Cor, 928-4000) at 60 V for 30 min, followed by 120 V for approximately 50 min or until loading dye ran off. Samples on the gel were transferred to methanol activated PVDF 0.45 µm membranes using BioRad’s wet tank mini-protein system for 1–3 h at 250–300 constant mA in a sample dependent context. After transfer, membranes were dried at 37°C for 5 min and then re-activated with methanol. Blots were stained with Li-Cor’s Revert-700 Total Protein Stain (Li-Cor, 926-11010) for normalization and imaged using a Li-Cor Odyssey Clx. Blots were then de-stained per kit instructions and put in block (Li-Cor Intercept block, 927-60001) for 1 h, room temperature, shaking. Blots were then transferred to primary antibody (electronic supplementary material, table S1) (block with 0.2% Tween20 overnight at 4°C, shaking. The following day, blots were washed four times in TBS-T for 5 min each at room temperature, shaking, and then incubated in secondary antibody (electronic supplementary material, table S2) in block with 0.2% Tween20 and 0.1% SDS, and/or streptavidin conjugate (1:5000; 680 or 800, Li-Cor, 926–68079, 926-32230) if visualizing biotinylated proteins, for 1 hour at room temperature, shaking. Blots were then washed twice with TBS-T for 5 min each, room temperature, shaking, followed by two 5-minute TBS washes at room temperature, shaking. Blots were imaged on a Li-Cor Odyssey Clx using Li-Cor’s ImageStudio (v. 5.2.5). After imaging blots were dried at 37°C for 5 min, then stored. For phosphorylated proteins (EIF2α), proteins were blotted for the phosphorylated state as described above. After imaging, phospho-blotos were stripped is 10 ml 1X Restore Fluorescent Western Blot stripping buffer (Thermo cat. 62300) for 20 min at room temperature, shaking. Blots were then briefly rinsed with dH2O twice, re-blocked for 30 min at room temperature shaking, before primary incubation with the total protein antibody and secondary as described above. All western blot images were exported from Li-Cor, pseudo-coloured and converted to RGB tiffs in Imagej (v. 1.51S) for figures. For specific proteins, bands were selected based on molecular weight from antibody manufacturer information and literature. Note that all western blot experiments were repeated at least twice with different biological samples and produced consistent results.

### 4.13. Fluorescent western blot quantification

Biotinylation levels and proteins of interest were quantified via western blot using Li-Cor’s fluorescent western blot ImageStudio (v. 5.2.5) and Emperia Studio (v. 1.3.0.83) analysis software and protocols. Total protein stain images of each blot were used to normalize biotinylation (streptavidin) or protein of interest signal intensity in RStudio (v. 1.3.959, R v. 4.0.0) by determining the lane normalization factor (Li-Cor protocol) for each blot per manufacturer’s instructions. ggplot2 (v. 3.3.5) and GraphPad Prism 9.0 were used to visualize normalized biotinylated protein signal.

### 4.14. Sox2-BirA*G3 analysis by MS (corresponding to figures 3–4 and electronic supplementary material, figures S9–S13)

After streptavidin beads pulldowns, Sox2-BirA*G3 and control samples were sent to The Broad Institute of Harvard and MIT for MS.

#### 4.14.1. On-bead digestion

Samples collected and enriched with streptavidin magnetic beads were washed twice with 200 µl of 50 mM Tris-HCl buffer (pH 7.5), transferred into new 1.5 ml Eppendorf tubes, and washed 2 more times with 200 µl of 50 mM Tris (pH 7.5) buffer. Samples were incubated in 0.4 µg trypsin in 80 µl of 2 M urea/50 mM Tris buffer with 1 mM DTT, for 1 h at room temperature while shaking at 1000 r.p.m. Following pre-digestion, 80 µl of each supernatant was transferred into new tubes. Beads were then incubated in 80 ul of the same digestion buffer for 30 min while shaking at 1000 r.p.m. Supernatant was transferred to the tube containing the previous elution. Beads were washed twice with 60 µl of 2 M urea/50 mM Tris buffer, and these washes were combined with the supernatant. The eluates were spun down at 5000g for 1 min and the supernatant was transferred to a new tube.
Samples were reduced with 4 mM DTT for 30 min at room temperature, with shaking. Following reduction, samples were alkylated with 10 mM iodoacetamide for 45 min in the dark at room temperature. An additional 0.5 µg of trypsin was added and samples were digested overnight at room temperature while shaking at 700 g. Following overnight digestion, samples were acidified (pH < 3) with neat formic acid (FA), to a final concentration of 1% FA. Samples were spun down and desalted on C18 StageTips as previously described56. Eluted peptides were dried to completion and stored at −80°C.

4.14.2. TMT labelling of peptides

Desalted peptides were labelled with TMT (6-plex) reagents (ThermoFisher Scientific). Peptides were resuspended in 80 µl of 50 mM HEPES and labelled with 20 ul 20 mg ml⁻¹ TMT6 reagents in ACN. Samples were incubated at RT for 1 h with shaking at 1000×r.p.m. TMT reaction was quenched with 4 µl of 5% hydroxyamine at room temperature for 15 min with shaking. TMT labelled samples were combined, dried to completion, reconstituted in 100 µl of 0.1% FA, and desalted on StageTips.

4.14.3. bRP stage tip fractionation of peptides

50% of the TMT labelled peptide sample was fractionated by basic reverse phase (bRP) fractionation. StageTips packed with 3 disks of SDB-RPS (Empore) material. StageTips were conditioned with 100 µl of 100% MeOH, followed by 100 µl 50% MeCN/0.1% FA and two washes with 100 µl 0.1% FA. Peptide samples were resuspended in 200 µl 1% FA (pH < 3) and loaded onto StageTips. 6 step-wise elutions were carried out in 100 µl 20 mM ammonium formate buffer with increasing concentration of 5%, 10%, 15%, 20%, 25% and 45% MeCN. Eluted fractions were dried to completion.

4.14.4. Liquid chromatography and mass spectrometry

Single-shot LC-MS/MS analyses were performed on 50% of each sample. The remaining 50% of each sample was fractionated using bRP StageTip fractionation. For single shot and all fractionated samples, desalted peptides were resuspended in 9 µl of 3% MeCN/0.1% FA and 4 µl was injected for serum samples, an Orbitrap Fusion Lumos Tridiss Mass Spectrometer (ThermoFisher Scientific) was used. For all other plexes, an Orbitrap Exploris 480 (ThermoFisher Scientific) was used. Mass spectrometers were coupled online to a Proxeon Easy-nLC 1200 (ThermoFisher Scientific) as previously described56. Briefly, 4 µl of each sample was loaded at onto a microcapillary column (360 µm outer diameter × 75 µm inner diameter) containing an integrated electrospray emitter tip (10 μm), packed to approximately 24 cm with ReproSil-Pur C18-AQ 1.9 µm beads (Dr Maisch GmbH) and heated to 50°C. bRP fractionated samples were analysed using a 110 min LC–MS. Mobile phase flow rate was 200 nl min⁻¹, comprises 3% acetonitrile/0.1% formic acid (Solvent A) and 90% acetonitrile /0.1% formic acid (Solvent B). The 110-min LC–MS/MS method used following gradient profile: (min.%B) 0 : 2; 1 : 6; 85 : 30; 94 : 60; 95 : 90; 100 : 90; 101 : 50; 110 : 50 (the last two steps at 500 nl min⁻¹ flow rate). Data acquisition was done in the data-dependent mode acquiring HCD MS/MS scans (r = 15 000) on the top 12 most abundant ions using an MS1 AGC target of 4 × 10⁵ and an MS2 AGC target of 5 × 10². The maximum ion time used for MS/MS scans was 120 ms; the HCD-normalized collision energy was set to 36 (Fusion Lumos) or 28 (Exploris 480); the dynamic exclusion time was set to 20 s, and the peptide match and isotope exclusion functions were enabled. Charge exclusion was enabled for charge states that were unassigned, 1 and >7.

4.15. MS data analysis

All protein trafficking MS data were analysed using Spectrum Mill software package v. 7.07 (proteomics.broadinstitute.org). Similar MS/MS spectra acquired on the same precursor m/z within ±60 s were merged. MS/MS spectra were excluded from searching if they were not within the precursor MH⁺ range of 600–6000 Da or if they failed the quality filter by not having a sequence tag length >0. MS/MS spectra were searched against a UniProt mouse database with a release date of December 28, 2017 containing 46 519 proteins and 264 common contaminants modified to include GFP, mKate2 and BirA*G3-ER. All spectra were allowed ±20 ppm mass tolerance for precursor and product ions, 40% minimum matched peak intensity, and ‘trypsin allow P’ enzyme specificity with up to 2 missed cleavages. The fixed modifications were carbamidomethylation at cysteine, and TMT6 at N-termini. The variable modifications used were oxidized methionine and N-terminal protein acetylation. Individual spectra were automatically designated as confidently assigned using the Spectrum Mill autovalification module. Specifically, a target-decoy-based false-discovery rate (FDR) scoring threshold criteria via a two-step auto threshold strategy at the spectral and protein levels was used. First, peptide mode was set to allow automatic variable range precursor mass filtering with score thresholds optimized to yield a spectral level FDR of <1.2%. A protein polishing autovalification was applied to further filter the peptide spectrum matches using a target protein level FDR threshold of 0. Following autovalification, a protein–protein comparison table was generated, which contained experimental over control TMT ratios. For all experiments, non-mouse contaminants and reverse hits were removed. Furthermore, the data were median normalized. For serum data, we performed a moderated T-test (limma R package v. 4.1) to identify proteins significantly enriched in the experimental conditions compared to controls. We corrected for multiple hypotheses (Benjamini–Hochberg procedure). Any protein with an adjusted p-value of less than 0.05 and a log2 fold change greater than 1 was considered statistically enriched. For tissue data, we used the ES method described below (MS hit analysis) to identify enriched proteins.

4.16. MS hit analysis

To identify enriched proteins from MS data, we established threshold TMT ratios for hit-calling using positive control (PC) and negative control (NC) protein lists. For the PC list, we used UniProt annotated secreted proteins, while the NC list was the UniProt overlapping list of transcription factors and nuclear proteins, and cytoskeletal genes. Note that the NC list was compared with secreted, receptors, ER proteins and overlapping genes were removed.
Proteins identified by MS were compared to the PC and NC lists and assigned to as being a PC or NC protein. For each experiment (liver, brain, and kidney), there were 9 TMT ratio comparisons: we calculate the TMT ratios of every Sox2-BirA*G3 sample over every control samples (Sox2-BirA*G3-1/control-1, Sox2-BirA*G3-1/control-2, Sox2-BirA*G3-1/control-3, Sox2-BirA*G3-2/control-1, Sox2-BirA*G3-2/control-2, Sox2-BirA*G3-3/control-1, Sox2-BirA*G3-3/control-2, and Sox2-BirA*G3-3/control-3). The false positive rate (FPR) is calculated using the following equation [25]:

\[
PFR = \frac{P(\text{TMT ratio} | \text{false positive})}{P(\text{TMT ratio} | \text{secreted positive})}
\]

The denominator is the conditional probability of finding a known secreted protein in this range, which is calculated as the percentage of proteins on the PC list in this range over all proteins identified on the PC list. The numerator is the conditional probability of finding a false positive protein in a particular protein in a specific TMT ratio range. The result calculated using this equation represents the percentage of false positive proteins in this TMT ratio range over the total false positive proteins identified. We plotted FPR over TMT ratio range (electronic supplementary material, figure S10E; other plots available upon request) and selected the TMT ratio cutoff based on an FPR of 0.1 (electronic supplementary material, figure S10D; other plots available upon request), which means that a protein is 10 times more likely to be a true secreted protein than a false positive. Enrichment score (ES) of a specific protein was defined as the number of TMT ratios that exceeds the TMT ratio cutoff. Thus, proteins with ES of 0 are background, proteins with ES of 1 are lower confidence hits, and proteins with ES of 9 are highest confidence hits.

4.17. Alb-BirA*G3 serum analysis by MS (corresponding to figures 5–6 and electronic supplementary material, figure S16)

After streptavidin beads pulldowns, Alb-BirA*G3 and control samples were sent to the UCLA proteomics core, Department of Biological Chemistry, Geffen School of Medicine at UCLA for MS.

4.17.1. Serum sample digestion

Streptavidin-bound proteins were reduced and alkylated on bead via sequential 20-minute incubations with 5 mM TCEP and 10 mM iodoacetamide at room temperature in the dark while being mixed at 1200 rpm in an Eppendorf thermomixer. Proteins were then digested by the addition of 0.1 µg Lys-C (FUJIFILM Wako Pure Chemical Corporation, 125-05061) and 0.8 µg Trypsin (Thermo Scientific, 90057) while shaking at 37°C overnight.

4.17.2. TMT labelling and CIF fractionation

The supernatant was transferred to new tubes and 8 µl of carboxylate-modified magnetic beads (CMMB, and also thermomixer. Proteins were then digested by the addition of 0.1 µg Lys-C (FUJIFILM Wako Pure Chemical Corporation, 125-05061) and 0.8 µg Trypsin (Thermo Scientific, 90057) while shaking at 37°C overnight.

Streptavidin-bound proteins were reduced and alkylated on bead via sequential 20-minute incubations with 5 mM TCEP and 10 mM iodoacetamide at room temperature in the dark while being mixed at 1200 rpm in an Eppendorf thermomixer. Proteins were then digested by the addition of 0.1 µg Lys-C (FUJIFILM Wako Pure Chemical Corporation, 125-05061) and 0.8 µg Trypsin (Thermo Scientific, 90057) while shaking at 37°C overnight.

Streptavidin-bound proteins were reduced and alkylated on bead via sequential 20-minute incubations with 5 mM TCEP and 10 mM iodoacetamide at room temperature in the dark while being mixed at 1200 rpm in an Eppendorf thermomixer. Proteins were then digested by the addition of 0.1 µg Lys-C (FUJIFILM Wako Pure Chemical Corporation, 125-05061) and 0.8 µg Trypsin (Thermo Scientific, 90057) while shaking at 37°C overnight.

4.18. Data analysis and statistics

Data were analysed using Microsoft Excel, R (ve 4.0.0 (2020-04-24), Platform: x86_64-appledarwin17.0 (64-bit); RStudio v. 1.3.959) and Python. For secretion annotations, proteins were annotated based on the subcellular localization data from UniProt and the cellular component data from National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov). Proteins in fasta formats were uploaded to SignalP5.0 (https://services.healthtech.dtu.dk/service.php?SignalP5.0) and TMHMM (v. 2.0) (https://services.healthtech.dtu.dk/service.php?TMHMM-2.0) for the prediction of N-terminal and C-terminal signal peptide, respectively. EnhancedVolcano was used to generate volcano plots based on log2FC value. We ran TissueEnrich (https://bioconductor.org/packages/TissueEnrich) [32] on a list of proteins to look for enrichment for tissue-specific genes using mouse ENCODE datasets. Gene ontology function annotation was performed on two platforms-DAVID (https://david.ncifcrf.gov) and EnrichGO in clusterProfiler (3.16.1). The top GO terms were visualized with dotplot in ggplot2. PCA was used to study the similarities between samples. The analysis was conducted without filtering any proteins. For ectodomain shedding analysis, we wrote a python program to map all the peptides identified in mass spec for each protein to their corresponding full-length protein. The reference sequences were annotated based on UniProt topology information. Domain information was based on SMART (http://smart.embl-heidelberg.de).
References

1. Mitchell KJ et al. 2001 Functional analysis of secreted and transmembrane proteins critical to mouse development. Nat. Genet. 28, 241–249. (doi:10.1038/90074)

2. Kotas ME, Medzhitov R. 2015 Homeostasis, inflammation, and disease susceptibility. Cell 160, 816–827. (doi:10.1016/j.cell.2015.02.010)

3. Ambazhagan R, Kavatharpu R, Mathur KP, Kottis T, Prakash H. 2021 Editorial: Systemic regulation of organ homeostasis and implications of hormones and immunity. Front. Endocrinol. (Lausanne) 12, 740835. (doi:10.3389/fendo.2021.740835)

4. Uhlen M et al. 2019 The human secretome. Sci. Signal. 12, eaaz0274. (doi:10.1126/sci.signal.aaz0274)

5. Basu RK, Wheeler DS. 2013 Kidney-lung cross-talk and acute kidney injury. Pediatr. Nephrol. 28, 2239–2248. (doi:10.1007/s00467-012-2386-3)

6. Husain-Syed F, Slutzky AS, Ronco C. 2016 Lung–kidney cross-talk in the critically ill patient. Am. J. Respir. Crit. Care Med. 194, 402–414. (doi:10.1164/rccm.201602-020CP)

7. Goury NM, Nikkats N. 2020 Multiple organ dysfunction syndrome. J. Intensive Care Med. 35, 1564–1575. (doi:10.1177/0885066619871452)

8. Gillette MA, Carr SA. 2013 Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. Nat. Methods, 10, 28–34. (doi:10.1038/nmeth.2309)

9. Pietrokovska M, Wlosowicz A, Gawin M, Wellał P. 2019 Ms-based proteomic analysis of serum and plasma: problem of high abundant components and light and shadows of albumin removal. Adv. Exp. Med. Biol. 1073, 57–76. (doi:10.1007/978-3-030-12298-0_3)

10. Keshishian H, Burgess MW, Specht H, Wallace L, Clauser KR, Gillette MA, Carr SA. 2017 Quantitative, multiplexed workflow for deep analysis of human blood plasma and biomarker discovery by mass spectrometry. Nat. Protoc. 12, 1683–1701. (doi:10.1038/nprot.2017.054)

11. Ruck JR, Kim DI, Raida M, Burke S. 2012 A promiscuous bortezomib ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J. Cell Biol. 196, 801–810. (doi:10.1083/jcb.201112098)

12. Branon TC, Bosch JA, Sanchez AD, Udeshi ND, Svitkina T, Carr SA, Feldman JS, Perrimon N, Ting AT. 2018 Efficient proximity labeling in living cells and organisms with TurboID. Nat. Biotechnol. 36, 880–887. (doi:10.1038/nbt.4201)

13. Droujinine IA et al. 2021 Proteomics of protein trafficking by in vivo tissue-specific labeling. Nat. Commun. 12, 2382. (doi:10.1038/s41467-021-22599-x)

14. Kim KE, Park I, Kim J, Kang MG, Choi WG, Shin H, Kim J-S, Rhee H-W, Suh JM. 2021 Dynamic tracking and identification of tissue-specific secretory proteins in the circulation of live mice. Nat. Commun. 12, 5204. (doi:10.1038/s41467-021-25546-y)

15. Wei W, Riley NM, Yang AC, Kim JT, Terrell SM, Li VL, Garcia-Contreras M, Bertozi CR, Long JZ. 2021 Cell type-selective secretome profiling in vivo. Nat. Chem. Biol. 17, 326–334. (doi:10.1038/s41589-020-0098-y)

16. Strack R. 2021 Revealing the secretome. Nat. Methods 18, 1273. (doi:10.1038/s41592-021-01320-2)

17. Liu J, Jiang PY, Piozoaizia M, Liu S, Finkel T. 2021 The secretome mouse provides a genetic platform to delineate tissue-specific in vivo secretion. Proc. Natl Acad. Sci. USA 118, e2003141118.

18. Ueza A, Kanak DJ, Bradshaw TW, Soderblom EJ, Catavero CM, Burette AC, Weinberg RJ, Soderling SH. 2016 Identification of an elaborate complex mediating postynaptic inhibition. Science 353, 1123–1129. (doi:10.1126/science.aag0821)

19. Spence ES, Dube S, Ueza A, Locke M, Soderblom EJ, Soderling SH. 2019 In vivo proximity proteomics of nascent synapses reveals a novel regulator of cytoskeleton-mediated synaptic maturation. Nat. Commun. 10, 386. (doi:10.1038/s41467-019-08288-w)

20. Zhou X et al. 2018 In vivo long-term investigation of tumor bearing mTag2 by an in-house fluorescence molecular imaging system. Biomed. Eng. Online. 17, 187. (doi:10.1186/s12938-018-0615-0)

21. Cranfill PJ et al. 2016 Quantitative assessment of fluorescent proteins. Nat. Methods 13, 557–562. (doi:10.1038/nmeth.3981)

22. Scherbo D et al. 2009 Far-red fluorescent tags for protein imaging in living tissues. Biochem. J. 418, 567–574. (doi:10.1042/BJ20081949)

23. Hashiy S, Lewis P, Perrin L, McMahon AP. 2002 Efficient gene modulation in mouse epiblast using a Sox2Cre transgenic mouse strain. Mech. Dev. 119(Suppl 1), S97–S101. (doi:10.1016/S0925-4773(03)00099-6)
