OPEN BIOLOGY

royalsocietypublishing.org/journal/rsob

Research

Cite this article: Yang R *et al*. 2022 A genetic model for *in vivo* proximity labelling of the mammalian secretome. *Open Biol.* **12**: 220149. https://doi.org/10.1098/rsob.220149

Received: 17 May 2022
Accepted: 20 July 2022

Subject Area:
biotechnology/genetics/cellular biology

Keywords:
proximity-labelling, BirA, TurboID, secretome, inter-organ communication, serum proteins

Author for correspondence:
Andrew P. McMahon
e-mail: amcmahon@med.usc.edu

†These authors contributed equally.

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.6125239.

THE ROYAL SOCIETY PUBLISHING

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

Rui Yang[1,2,†], Amanda S. Meyer[1,2,†], Ilia A. Droujinine[3,†], Namrata D. Udeshi[4], Yanhui Hu[5], Jinjin Guo[1,2], Jill A. McMahon[1,2], Dominique K. Carey[4], Charles Xu[4], Qiao Fang[6], Jihui Sha[7], Shishang Qin[8], David Rocco[5], James Wohlschlegel[7], Alice Y. Ting[9,10], Steven A. Carr[4], Norbert Perrimon[5,11] and Andrew P. McMahon[1,2]

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

RY, 0000-0003-2291-4914; ASM, 0000-0003-1022-6137; IAD, 0000-0003-1311-4392; APM, 0000-0002-3779-1729

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

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 enzyme-catalyzed proximity labelling approaches such as BioID [11] and TurboID [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 proteome 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 insight into tissue secretomes and inter-organ communication [14–19].

To overcome the limitations of viral-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 precursor to TurboID, 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 embryo stem cells (mESCs) (figure 1*a*) [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 1*a*) [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 1*b*). 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 1*b*) [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 epiblast, 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 1*c*).

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 1*d*), while *Sox2-BirA*G3* mice were GFP-/mKate2+, demonstrating excision of the GFP cassette and activation of mKate2 reporter (figure 1*e*). High-magnification confocal images showed that BirA*G3 extensively colocalized with the ER resident protein calnexin (figure 1*f*; electronic supplementary material, figure S1A), consistent with BirA*G3 localization to the ER. Focusing on the kidney, liver and brain, we next sought to determine the distribution of BirA*G3 using representative cell markers, highlighting key cell populations.

In the *Sox2-BirA*G3* kidney, mKate2 signal and BirA*G3 were present throughout the nephron and collecting system highlighted by co-localization with different cell markers (figure 1*g*; electronic supplementary material, figures S1B, C and S2A,B). In *Sox2-BirA*G3* liver, mKate2 signal and BirA*G3 were present in hepatocytes (Albumin+), though the distribution was patchy (figure 1*h*) and neighbouring cholangiocytes (CK19+) showed much lower levels of reporter and BirA*G3 (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 1*i*; electronic supplementary material, figure S1E-F; 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, mKate2^{low} and mKate2^{high} 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 mKate2^{low} and mKate2^{high} cells, though *BirA*G3* mRNA levels were also lower in mKate2^{low} cells (electronic supplementary material, figure S5D). By contrast, albumin mRNA expression was comparable in mKate2^{low} and mKate2^{high} populations (electronic supplementary material, figure S5E). The lower *BirA*G3* mRNA expression was consistent with lower BirA*G3 protein level in mKate2^{low} hepatocytes (electronic supplementary material, figure S5F–H). Protein biotinylation is still observed in mKate2^{low} cells, although at lower levels

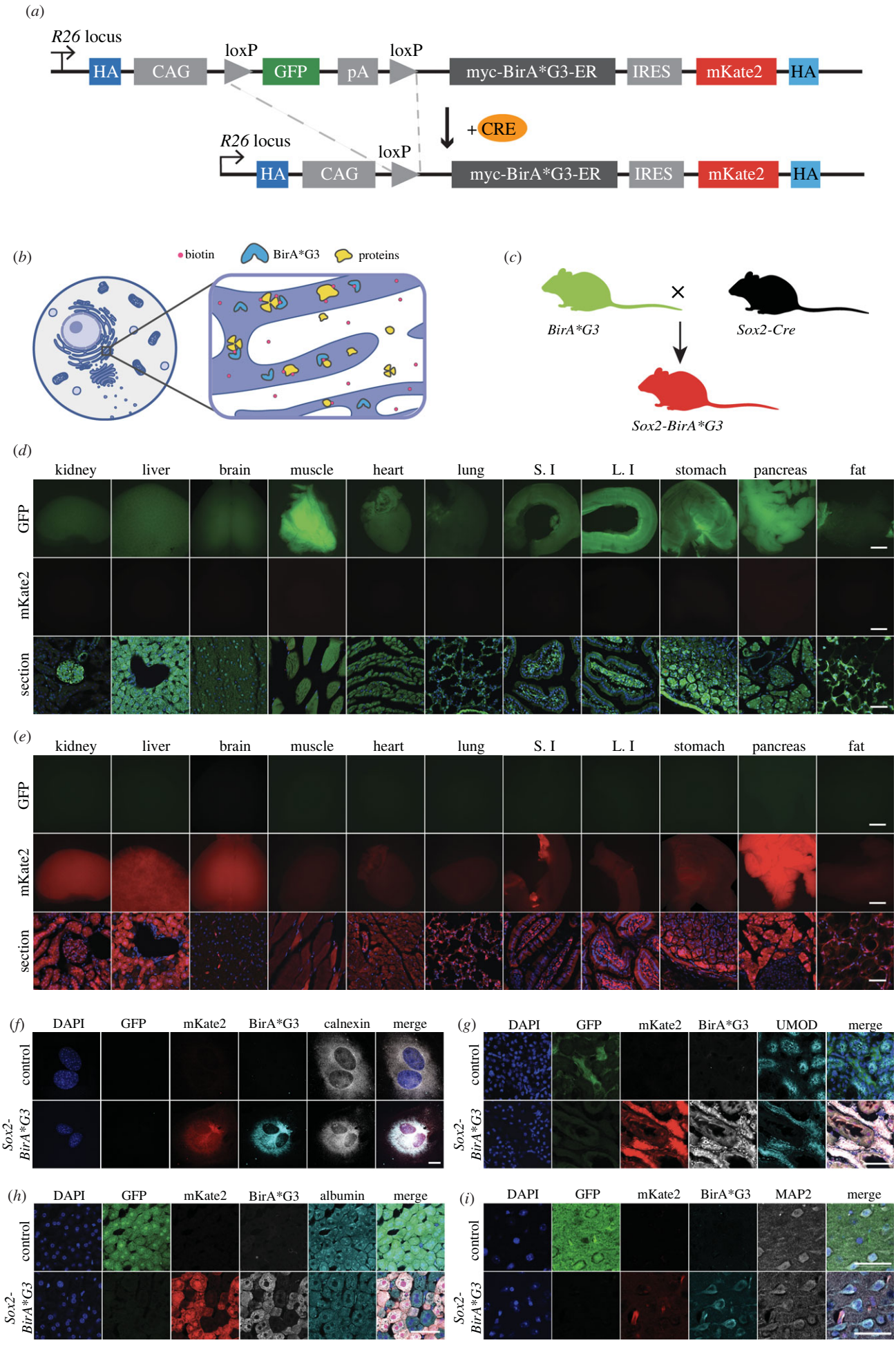

Figure 1. (*Caption overleaf.*)

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 fibroblast (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 Albumin, a hepatocyte marker, in the liver sections of *Sox2-BirA*G3* and control mice. Scale bar: 50 µm. (*i*) 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.

(electronic supplementary material, figure S5F–H), Thus, a mosaic reduction in transcriptional activity of the BirA*G3 locus likely underlies variable levels of mKate2 and BirA*G3 in hepatocytes.

To determine whether downregulation relates to the duration of allele activation, we examined BirA*G3 in *Sox2-BirA*G3* and control pups 10 days after birth. Interestingly, at this early time point, BirA*G3 and mKate2 show a relatively homogeneous distribution in hepatocytes (electronic supplementary material, figure S6A,B). While the molecular underpinnings for this observation are unclear, most cell/tissue types were not affected and inducible CRE strains can potentially overcome possible selection for allele silencing. Indeed, we observed homogeneous BirA*G3 in GFP- hepatocytes of adult *CAGGCre-ER^{TM}; BirA*G3* (*CAGG-BirA*G3*) mice following tamoxifen injection to induce broad, mosaic excision of the GFP cassette in the adult animal (electronic supplementary material, figure S6C-D).

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 2*a*). Additionally, we observed biotin/BirA*G3 independent-labelling of two prominent proteins in control and experimental brain samples (figure 2*a*) 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 2*a*; 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 2*b*; electronic supplementary material, figure S3B). Notably, each tissue displayed a unique biotinylation pattern (figure 2*b*), suggesting diverse secretomes across tissues. Further, analysis of serum detected a robust BirA*G3-dependent biotinylated protein signature (figure 2*c*; 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 2*d*; 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 2*e*; 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 2*e*).

The accumulation of biotinylated proteins could also be visualized by streptavidin-dependent immunofluorescence directly in sections of selective tissues. Abundant biotinylated proteins were observed in *Sox2-BirA*G3* muscle, heart, brain, compared to their control counterparts (figure 2*f*; 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 2*b*).

To look for detrimental effects of BirA*G3 in *Sox2*-BirA*G3 mice, we performed haematoxylin 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

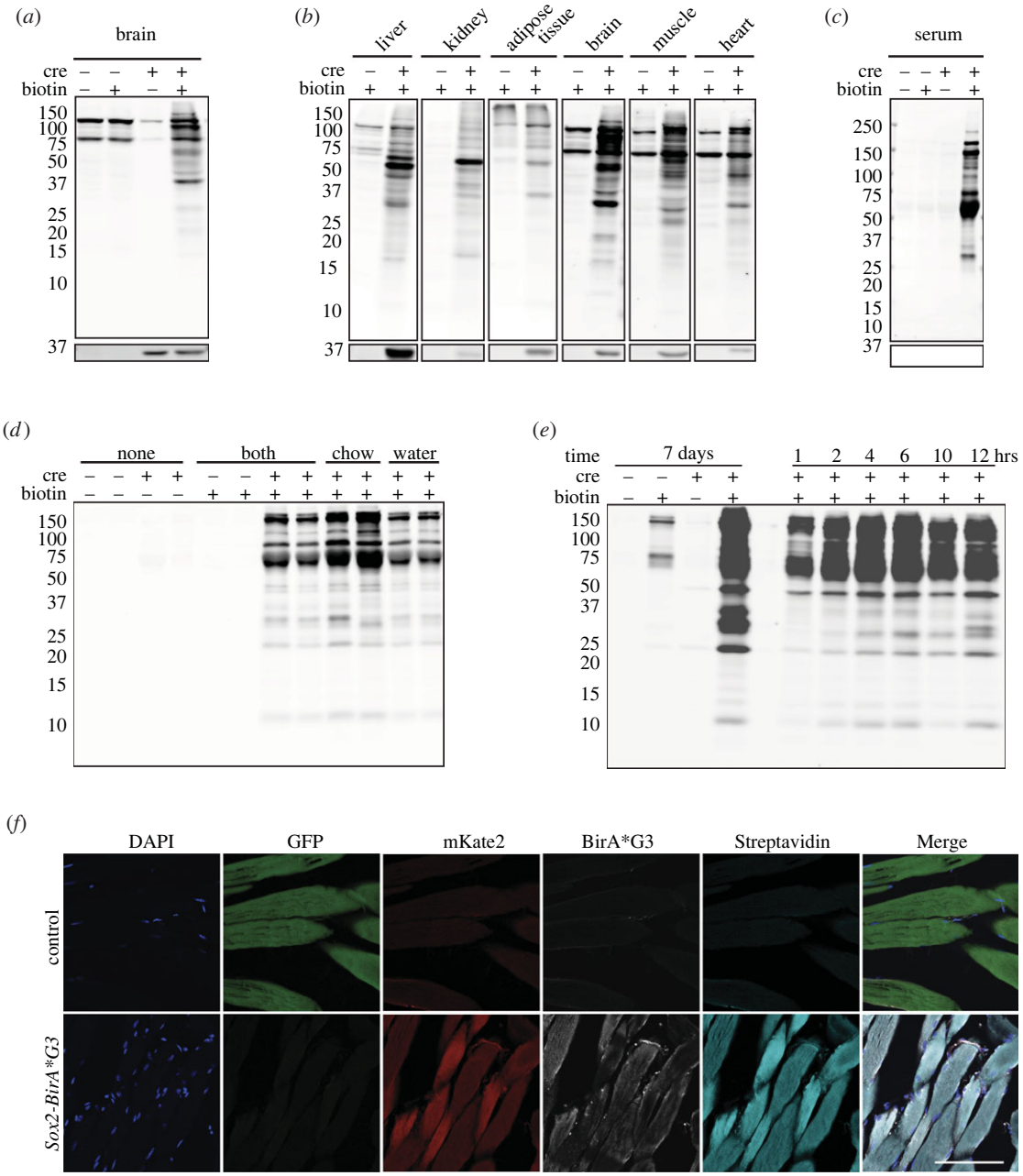

Figure 2. Analysis of biotinylated proteins in *Sox2-BirA*G3* and control mice. (*a*) Western blotting of protein lysates from brain in *Sox2-BirA*G3* mice (CRE+) compared to control mice (CRE−) with or without biotin chow administration for 5 days. Upper: streptavidin labelling. Lower: BirA*G3 (approx. 35 kDa). Each lane is a biological replicate from individual mice ($n = 2$ per genotype). (*b*) Western blotting of protein lysates from selective tissues in *Sox2-BirA*G3* mice compared to control mice. Note: due to varied streptavidin intensity by tissue, brain, heart, and muscle streptavidin westerns are shown with increased signal intensity. Upper: streptavidin labelling. Lower: BirA*G3 (approx. 35 kDa). Each lane is a biological replicate from individual mice ($n = 1$ per genotype). (*c*) Western blotting of serum total protein in *Sox2-BirA*G3* mice compared to control mice (CRE−) with or without biotin chow administration for 5 days. Upper: streptavidin labelling. Lower: BirA*G3 (approx. 35 kDa). Each lane is a biological replicate from individual mice ($n = 2$ per genotype). (*d*) Streptavidin labelling of biotinylated proteins in total serum from *Sox2-BirA*G3* and control mice administered with regular chow, biotin chow, biotin water or biotin chow and water for 7 days. Each lane is a biological replicate from individual mice ($n = 2$/biotin treatment). (*e*) Streptavidin labelling of affinity purified biotinylated proteins in serum from *Sox2-BirA*G3* and control mice given biotin chow for 7 days (left) or given biotin by subcutaneous injection and water (5 mM, pH 7.4) after the injection until collection (right). (*f*) Immunofluorescence images of native GFP, native mKate2, BirA*G3 staining, streptavidin staining in cryo-sectioned muscle tissues from *Sox2-BirA*G3* and control mice. Scale bar, 50 µm.

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 electronic supplementary material, figure 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 (electronic supplementary material, figure 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 ($\log_2$ fold change (FC) > 1.0 and adj. p-value < 0.05) (figure 3*a*,*b*; electronic supplementary

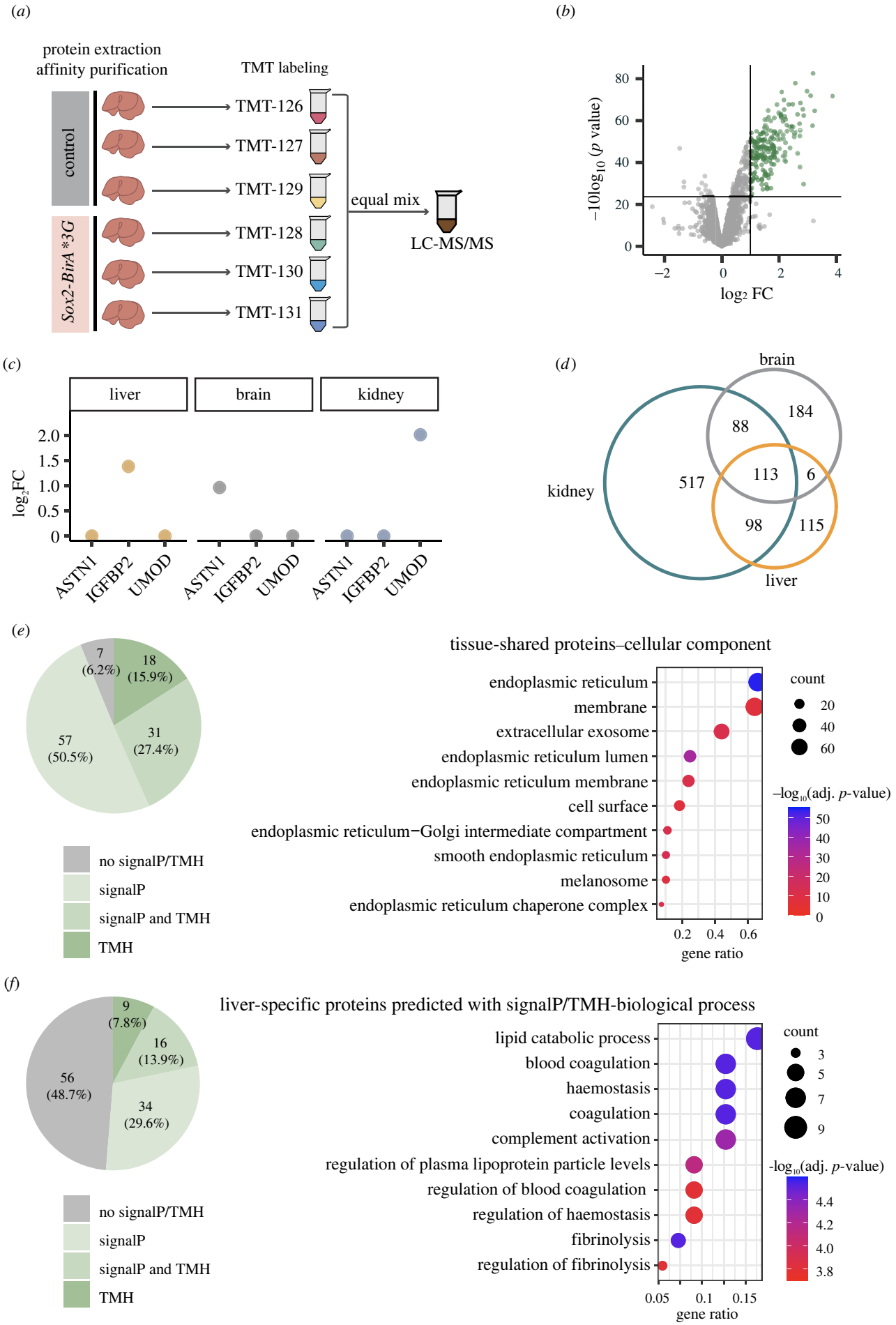

Figure 3. (*Caption overleaf.*)

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. Log_2 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 $\log_2\text{FC} > 1.0$) in *Sox2-BirA*G3 A* mice compared to control mice are shown in green. (*c*) Relative abundance ($\log_2\text{FC}$) 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 indicates 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 in figure 3*c*.

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 S10E). 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 $\text{ES} \geq 5$ 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 3*d*) and over 93% of these (106) showed a SignalP and/or TMH (figure 3*e*). Applying Gene Ontology (GO) functional category enrichment analysis to identify characteristic cellular component attributes of these tissue-shared proteins highlighted extracellular exosome and ER-Golgi terms in DAVID analysis (figure 3*e*; 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 3*f*; 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 3*f*). 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 3*c*; 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 4*a*; 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 $\log_2\text{FC} > 1.0$ 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

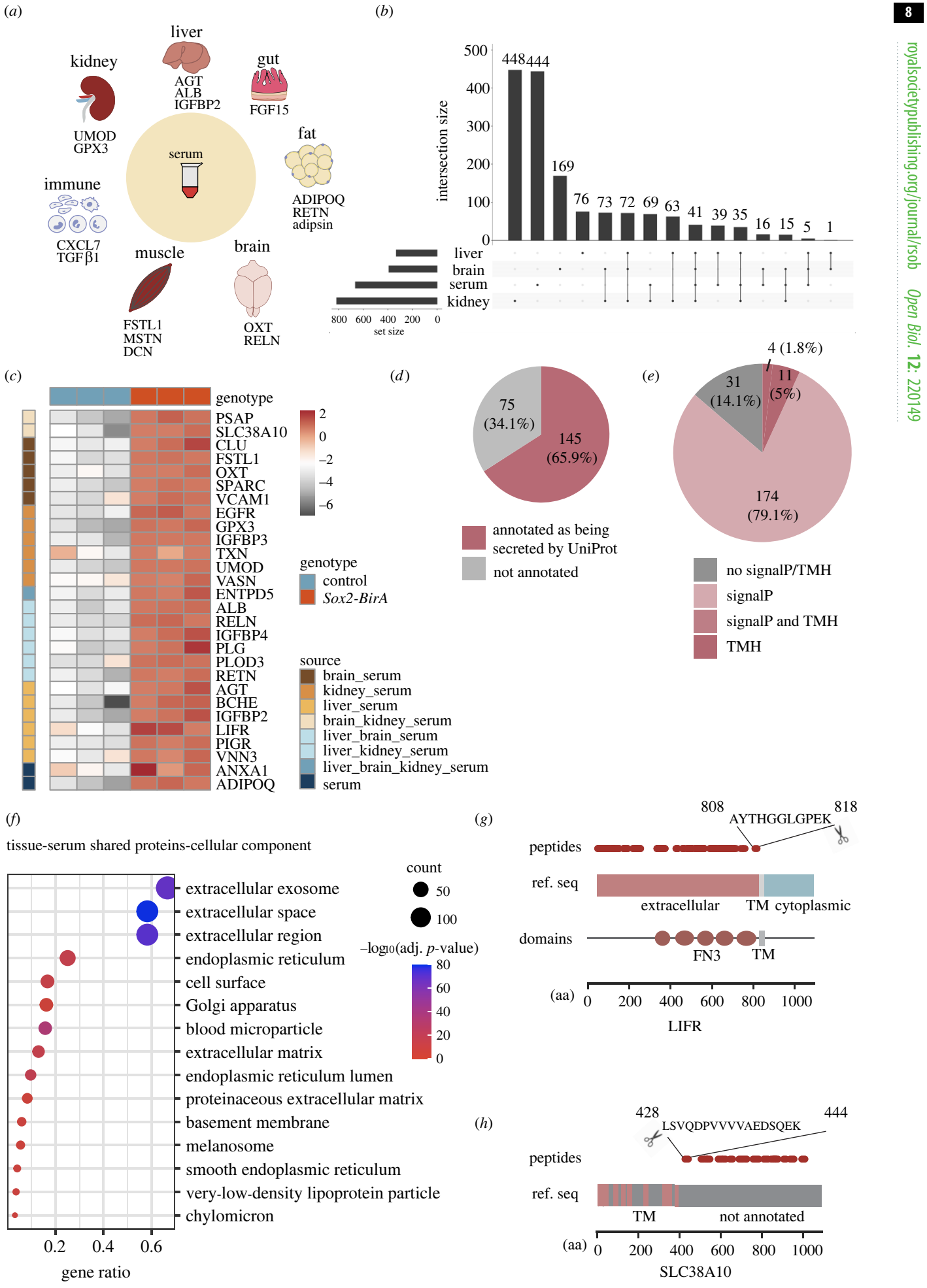

Figure 4. (*Caption overleaf.*)

Figure 4. (*Overleaf*.) Analysis of biotinylated proteins secreted to peripheral blood. (*a*) Schematic of tissue secreted proteins identified by LC-MS/MS (6plex; *Sox2-BirA*G3* $n = 3$, control $n = 3$) from affinity purified serum from *Sox2-BirA*G3* mice. (*b*) Upset plot showed the overlap of enriched proteins among serum, liver, brain and kidney. (*c*) Heatmap of representative proteins enriched in *Sox2-BirA*G3* serum relative to controls. Log_2 expression values are shown by colour and intensity of shading. Grey, low; red, high. (*d*) Pie chart displayed the number of shared enriched proteins between serum and three tissues (220 proteins) that are annotated as secreted by UniProt/NCBI. (*e*) Pie chart displayed the distribution of shared enriched proteins between serum and three tissues (220 proteins) with predicted SignalP/TMH. (*f*) Shared enriched proteins between serum and three tissues were analysed with DAVID analysis for cellular component annotation. Gene ratio indicates the percentage of genes annotated with the term over the total number of genes in the list. (*g*) Schematic of detected peptides for LIFR mapped onto its respective reference sequences with SMART protein domain annotation. Reference sequence is annotated with extracellular, transmembrane (TM) and cytoplasmic based on UniProt topology information. Amino acid sequences of the most C-terminal peptide are labelled. FN3: fibronectin type 3. (*h*) Schematic of detected peptides for Slc38a10 mapped onto its respective reference sequences. Reference sequence is annotated with TM based on UniProt topology information. Amino acid sequences of the most N-terminal peptide are labelled.

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 4*b*). A list of selective serum enriched proteins was shown in a heat map representing log_2FC between *Sox2-BirA*G3* and control sample, including classic hepatocyte secreted proteins, such as angiotensinogen (AGT) and albumin (ALB), brain cell marker, neuron neuropeptide oxytocin (OXT), as well as kidney markers, uromodulin (UMOD) and insulin-like growth factor-binding protein-3 (IGFBP3) (figure 4*c*).

Out of the 220 tissue-serum shared proteins, 65.9% were annotated as secreted by UniProt/NCBI (figure 4*d*), and the majority (85.9%) were predicted to have a SignalP or TMH (figure 4*e*). These tissue-serum shared proteins were then analysed for GO term enrichment analysis (figure 4*f*; electronic supplementary material, figure S13D,E). DAVID analysis demonstrated significant enrichment in 'extracellular' annotations in the cellular component annotation (figure 4*f*). 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 4*g*; 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 4*g*,*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 4*g*,*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 4*c*), 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 secretome

We next applied and extended the approach to determine the feasibility of identifying a tissue cell type specific secretome. 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 5*a*; electronic supplementary material, figures S15A and S14B). BirA*G3 showed a homogeneous distribution restricted to the hepatocytes of *Alb-BirA*G3* liver (figure 5*b*). BirA*G3 activity in *Alb-BirA*G3* liver enabled specific protein biotinylation in liver protein lysates to be compared with control mice (figure 5*c*). *Alb-BirA*G3* serum samples showed strong, specific protein biotinylation relative to serum from biotin-administered controls lacking active BirA*G3 (figure 5*d*). The biotinylation banding patterns of affinity purified biotinylated proteins were comparable to those observed in total protein lysates of liver and serum samples (figure 5*c*,*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 5*e*).

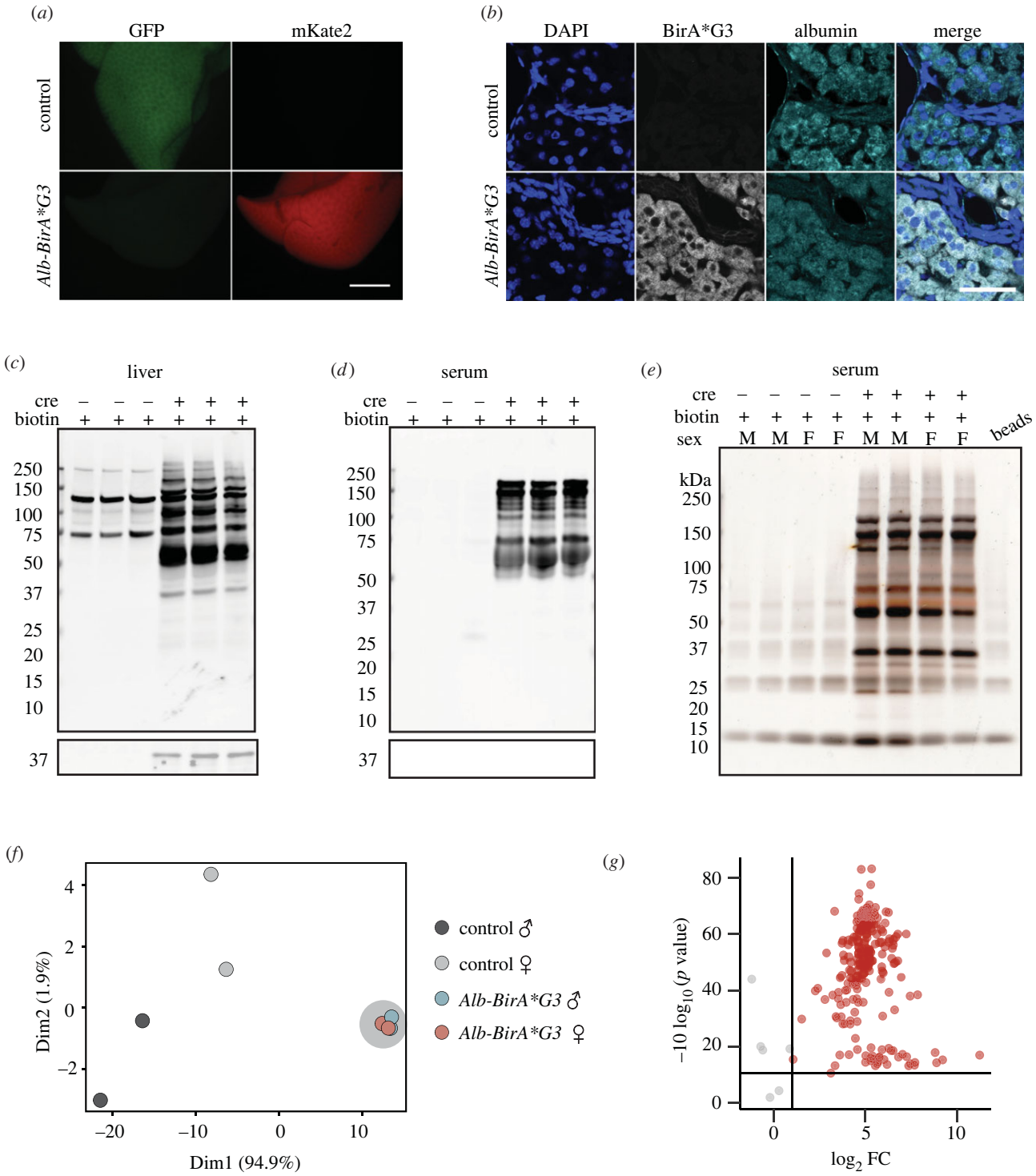

Figure 5. Generation and characterization of *Alb-BirA*G3* mice. (*a*) Whole mount images of native GFP and mKate2 fluorescence in livers of *Alb-BirA*G3* and control mice. Scale bar: 3 mm. (*b*) Immunofluorescence staining shows expression of BirA*G3 and hepatocyte marker Albumin in the liver sections from *Alb-BirA*G3* and control mice. Scale bar: 50 µm. (*c,d*) Western blotting of total protein from (*c*) liver and (*d*) serum in *Alb-BirA*G3* mice compared to control mice. Upper: Streptavidin labelling. Lower: BirA*G3 (approx. 35 kDa). Each lane is a biological replicate from individual mice ($n = 3$ per genotype). (*e*) Silver stain of affinity purified biotinylated proteins from serum in *Alb-BirA*G3* mice compared to control mice. Each lane is a biological replicate from individual mice ($n = 4$ per genotype). (*f*) PCA of streptavidin-purified serum proteins from *Alb-BirA*G3* and control mice. Each dot represents a sample, which is coloured by the annotation of mouse genotype and sex. *Alb-BirA*G3* samples are manually circled by grey shadow. (*g*) Volcano plots of proteins detected in serum of *Alb-BirA*G3* mice compared to control mice after streptavidin pulldown in mass spectrometry. $\log_2$ 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 $\log_2$FC > 1) in *Alb-BirA*G3* mice compared to control mice are shown in red.

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 5*f*). Approximately 80% of proteins (181/189) were specifically enriched ($\log_2$FC > 1.0 and adj. *p*-value < 0.05) in the *Alb-BirA*G3* group (figure 5*g*). 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*a*; figure 5*e*). 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 6*b*).

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 6*c*,*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 6*e*) [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 6*f*). 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 the *Alb-BirA*G3* unique 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 6*g–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 6*g–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], MSTN [48], RETN [45], CXCL7 [49], IGF1 [50], FGF15 [51]) levels. Interestingly, these represent examples of proteins derived from multiple tissues: adipose tissue (ADIPOQ, Adipsin [52], and RETN), muscle (MSTN), 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 BioID, 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 BioID2 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 2*c*; 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 cargoes) 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

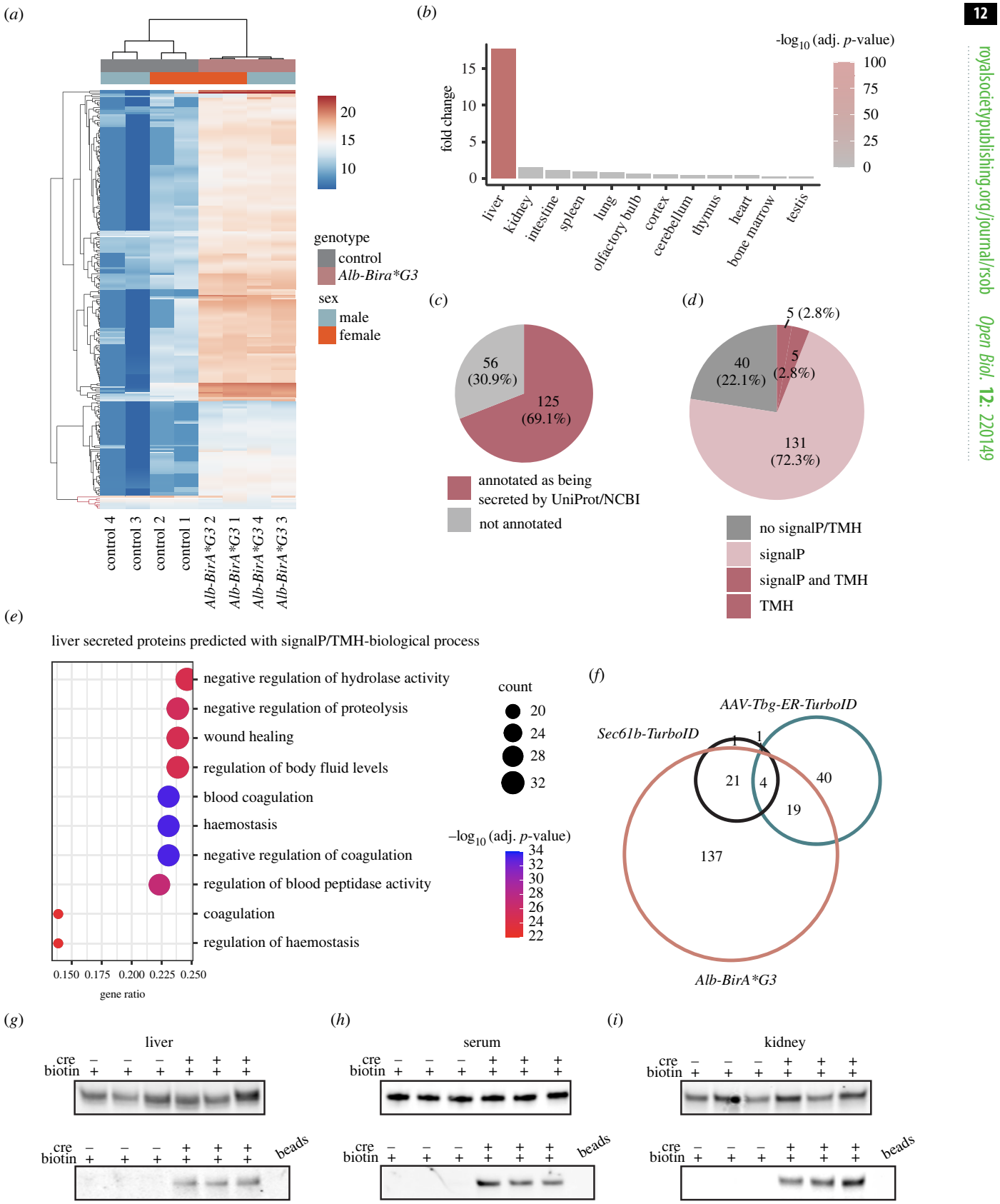

Figure 6. Analysis of hepatocyte-secreted serum proteins in *Alb-BirA*G3* mice. (*a*) Cluster heatmap of proteins enriched in *Alb-BirA*G3* serum relative to controls. Log_2 expression values are shown by colour and intensity of shading. Blue, low; red, high. Serum-enriched proteins in the *Alb-BirA*G3* 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.: 000058, The Jackson Laboratory) [56] carrying the loxP-flanked *BirA*G3* were aggregated with 8-cell C57bl/6 J embryos to obtain chimeras. The resulting chimeric mice were bred with R26PhiC31 (B6.129S4-*Gt(ROSA)26-Sor*$^{tm3(phiC31*)Sor}$/J, Stock No.: 007743, The Jackson Laboratory, backcrossed to C57bl/6 J for 13 generations) [57] females to remove attB-neoR-attP cassette. The resulting mice carry the *BirA*G3* allele, namely *BirA*G3* mice. The *Sox2-Cre* mice (B6.Cg-*Edil3*$^{Tg(Sox2-cre)1Amc}$/J, stock no.: 008454, The Jackson Laboratory) [23], the *TdTomato* mice (B6.Cg-*Gt(ROSA)*$^{26Sortm14(CAG-tdTomato)Hze}$/J, Stock No.: 007914, The Jackson Laboratory) [24], the *CAGGCre-ER*^{TM} mice(FVB.Cg-$^{Tg(CAG-cre/Esr1*)5Amc}$/J, stock no.: 017595, The Jackson Laboratory) [58], and the *Alb-Cre* mice (B6.Cg-*Speer6-ps1*$^{Tg(Alb-cre)21Mgn}$/J, stock no.: 003574) [59] were used as described previously.

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 × cold DPBS (Dulbecco's Phosphate Buffered Saline). The blood was allowed to clot at room temperature for 30 min and then spun down at 2000*g* for 15 min at 4°C. The serum was collected and spun again at 2000*g* 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 × 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 biotin and 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, brain, and kidney 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) for 5 days. Tissues were then collected as described above.

For ERT2 mice, tamoxifen (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 ug ml^{-1} Tunicamycin (R&Dsystems 590507) in DMSO for 5 h. Cells were then briefly rinsed with 1X DPBS and then scrapped off the plate in fresh DPBS. Cells were pelleted by centrifugation at 300*g* 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 1× DPBS and then fixed in 4% PFA for 10 min followed by 3 washes in 1× DPBS. Coverslips with cells were stored at 4°C in 1× 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 1× 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 1× 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, overnight at 4°C. The following day, primary antibodies were removed, and slides were washed in blocking buffer four times for 5 min each. Slides were then incubated in secondary antibody diluted in blocking buffer for 1 hour at room temperature. Secondary antibody (electronic supplementary material, table S2) was removed, and slides were washed in blocking buffer four times for 5 min each. Slides were then incubated in 1 mg ml^{-1} Hoechst 33342 (Thermo, 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 Immu-Mount (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 genotype. Whole tissue section scans were imaged using the Zeiss AxioScan Z1 Slide Scanner at 20× to generate high-resolution tiled images of tissue sections.

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. mKate2^{low} and mKate^{high} 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: CTCCCCGTGGTTGACTCTAC
Reverse: CTCCCCGTGGTTGACTCTAC
Alb:
Forward: GTCTTAGTGAGGTGGAGCATGACAC
Reverse: GCAAGTCTCAGCAACAGGGATACAG
Gapdh:
Forward: CATGGCCTTCCGTGTTCCTA
Reverse: CCTGCTTCACCACCTTCTTGAT

The delta-delta Ct method, also known as the $2^{-\Delta\Delta Ct}$ 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 dH_2O. 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-Clonetech) 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 first trimmed using Trimmomatic v. 0.38 and then aligned to GRCm39 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 $\log_2$ 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, TCB0013-100G), 4 µM pepstatin A (Sigma, EI10), 100 µM PMSF (Sigma, 11359061001)) and bead homogenized using stainless steel beads (NextAdvance, SSB14B-RNA) for 5 min at setting 10, Bullet Blender Storm (NextAdvance, BT24M). Samples were then centrifuged at 14 000g 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 dH_2O. Gels were then soaked in 0.02% sodium thiosulfate (Sigma 72049) for 1 min and then in dH_2O for 1 min, twice. Gels were then incubated in chilled 0.1% silver nitrate (Sigma 209139) for 20 min, gently shaking at 4C in the dark. Gels were then rinsed twice in dH_2O 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, 928–40000) 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-protean 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-blots 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 dH_2O 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 5000*g* 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% hydroxylamine 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 Tribrid 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 the 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 ($r = 15\,000$) after each MS1 scan ($r = 60\,000$) on the top 12 most abundant ions using an MS1 AGC target of 4×105 and an MS2 AGC target of 5×10^4. 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 autovalidation 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 autovalidation was applied to further filter the peptide spectrum matches using a target protein level FDR threshold of 0. Following autovalidation, 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*-2/control-3, *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]:

$$\mathrm{FPR} = \frac{P(\text{TMT ratio}|\ \text{false positive})}{P(\text{TMT ratio}|\ \text{secreted positive})}.$$

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 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 an 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 widely known as SP3 [60]) was added to each sample. 100% acetonitrile was added to each sample to increase the final acetonitrile concentration to >95% and induce peptide binding to CMMB. CMMB were then washed 3 times with 100% acetonitrile and then resuspended with TMT labelling buffer. 25 µg of each sample was labelled using TMT10 plex Isobaric Labels (Thermo Fisher Scientific) and the resulting 8 labelled samples were pooled. The pooled sample was fractionated by CMMB-based Isopropanol Gradient Peptide Fractionation (CIF) method [61] into 3 fractions before MS analysis.

4.17.3. LC-MS acquisition and analysis

Fractionated samples were separated on a 75 uM ID × 25 cm C18 column packed with 1.9 µm C18 particles (Dr Maisch GmbH) using a 140-minute gradient of increasing acetonitrile and eluted directly into a Thermo Orbitrap Fusion Lumos mass spectrometer where MS spectra were acquired using SPS-MS3.

Protein identification was performed using MaxQuant v. 1.6.17.0. The complete Uniprot mouse proteome reference database (UP000000589) was searched for matching MS/MS spectra. Searches were performed using a 20 ppm precursor ion tolerance. TMT10plex was set as a static modification on lysine and peptide N terminal. Carbamidomethylation of cysteine was set as static modification, while oxidation of methionine residues and N-terminal protein acetylation were set as variable modifications. LysC and Trypsin were selected as enzyme specificity with maximum of two missed cleavages allowed. 1% false discovery rate was used as a filter at both protein and PSM levels.

Statistical analysis was conducted with the MSstatsTMT Bioconductor package. The abundance of proteins missing from one condition but found in more than 2 biological replicates of the other condition for any given comparison were estimated by imputing intensity values from the lowest observed MS1-intensity across samples and *p*-values were randomly assigned to those between 0.05 and 0.01 for illustration purposes.

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?SignalP-5.0) and TMHMM (v. 2.0) (https://services.healthtech.dtu.dk/service.php?TMHMM-2.0) for the prediction of SignalP and transmembrane helix, separately. EnhancedVolcano was used to generate volcano plots based on $\log_2$FC and *p* 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).

Ethics. 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.

Data accessibility. The original mass spectra and the protein sequence databases used for searches have been deposited in the public proteomics repository MassIVE (http://massive.ucsd.edu) and are accessible at ftp://MSV000088848@massive.ucsd.edu. RNA sequencing data is available under BioProject PRJNA808087 from NCBI SRA at https://www.ncbi.nlm.nih.gov/sra. The following public databases were used: Uniprot (https://www.uniprot.org), mouse (https://www.uniprot.org/proteomes/UP000000589), SignalP 5.0 (https://services.healthtech.dtu.dk/service.php?SignalP-5.0), TMHMM 2.0 (https://services.healthtech.dtu.dk/service.php?TMHMM-2.0). All analysis code is available at https://github.com/asmeyer/A-genetic-model-for-in-vivo-proximity-labeling-of-the-mammalian-secretome.git. Corresponding authors will provide original data upon request. Source data are provided with this paper.

The data are provided in electronic supplementary material [62].

Authors' contributions. R.Y.: conceptualization, data curation, investigation, software, validation, visualization, writing—original draft; A.S.M.: conceptualization, data curation, investigation, software, validation, visualization, writing—original draft; I.A.D.: conceptualization, data curation, software, writing—original draft, writing—review and editing; N.D.U.: conceptualization, data curation, formal analysis, methodology, software; Y.H.: methodology, software, visualization; J.G.: data curation, validation, visualization; J.A.M.: conceptualization, data curation, project administration, visualization; D.K.C.: data curation, methodology, software, visualization; C.X.: data curation, software, visualization; Q.F.: software, visualization; J.S.: data curation, methodology, software; S.Q.: software, visualization; D.R.: data curation, visualization; J.W.: data curation, methodology, writing—review and editing; A.Y.T.: conceptualization, methodology, writing—review and editing; S.A.C.: methodology, software, writing—review and editing; N.P.: conceptualization, project administration, writing—review and editing; A.P.M.: conceptualization, methodology, project administration, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. A.P.M. is a scientific adviser for Novartis, TRESTLE BioTherapeutics, eGENESIS and IVIVA Medical. S.A.C. is a member of the scientific advisory boards of Kymera, PTM BioLabs, Seer and PrognomIQ. A.Y.T. declares intellectual property interests around proximity labelling technology.

Funding. This work was supported by a NIH Transformative R01 grant 5R01DK121409 to A.P.M., A.Y.T., S.A.C. and N.P., and HHMI funding to N.P. I.A.D. acknowledges support from the Ellen Browning Scripps Foundation.

Acknowledgements. We thank the Janelia Gene Targeting and Transgenics Facility (Director: Caiying Guo) for the BirA*G3 mouse generation. We thank Peter Calabrese and Ivy Xiong from University of Southern California (USC), Pierre Michel Jean Beltran from Broad Institute, Xianwen Ren from Peking University, Kevin Blighe from Rancho BioSciences, and Geetu Tuteja from Iowa State University for help with mass spectrometry data analysis. We thank Jeffrey Boyd and Bernadette Masinsin from the Flow Cytometry Facility of Eli and Edythe Broad CIRM Center for Regenerative Medicine and Stem Cell Research at USC for assisting with hepatocyte sorting. We thank Seth Ruffins from Optical Imaging Facility and Riana Parvez of Eli and Edythe Broad CIRM Center for Regenerative Medicine and Stem Cell Research at USC for assisting with fluorescence imaging. We thank Genome Technology Access Center from Washington University for performing bulk RNA-seq.

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. Anbazhagan R, Kavarthapu R, Mathur PP, Kostic TS, 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. Uhlén M *et al.* 2019 The human secretome. *Sci. Signal.* **12**, eaaz0274. (doi:10.1126/scisignal.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, Slutsky 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.201602-0420CP)
7. Gourd NM, Nikitas 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. Pietrowska M, Wlosowicz A, Gawin M, Widlak P. 2019 Ms-based proteomic analysis of serum and plasma: problem of high abundant components and lights 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. Roux KJ, Kim DI, Raida M, Burke B. 2012 A promiscuous biotin 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, Svinkina T, Carr SA, Feldman JL, Perrimon N, Ting AY. 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, Bertozzi CR, Long JZ. 2021 Cell type-selective secretome profiling in vivo. *Nat. Chem. Biol.* **17**, 326–334. (doi:10.1038/s41589-020-00698-y)
16. Strack R. 2021 Revealing the secretome. *Nat. Methods* **18**, 1273. (doi:10.1038/s41592-021-01320-2)
17. Liu J, Jang JY, Pirooznia 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**, e2005134118.
18. Uezu A, Kanak DJ, Bradshaw TW, Soderblom EJ, Catavero CM, Burette AC, Weinberg RJ, Soderling SH. 2016 Identification of an elaborate complex mediating postsynaptic inhibition. *Science* **353**, 1123–1129. (doi:10.1126/science.aag0821)
19. Spence EF, Dube S, Uezu 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 K *et al.* 2018 In vivo long-term investigation of tumor bearing mKate2 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.3891)
22. Shcherbo D *et al.* 2009 Far-red fluorescent tags for protein imaging in living tissues. *Biochem. J.* **418**, 567–574. (doi:10.1042/BJ20081949)
23. Hayashi S, Lewis P, Pevny 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)

24. Madisen L *et al.* 2010 A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140. (doi:10.1038/nn.2467)
25. Chen CL, Hu Y, Udeshi ND, Lau TY, Wirtz-Peitz F, He L, Ting AY, Carr SA, Perrimon N. 2015 Proteomic mapping in live *Drosophila* tissues using an engineered ascorbate peroxidase. *Proc. Natl Acad. Sci. USA* **112**, 12 093–12 098. (doi:10.1073/pnas.1515623112)
26. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, Von Heijne G, Nielsen H. 2019 SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol.* **37**, 420–423. (doi:10.1038/s41587-019-0036-z)
27. Nielsen H. 2017 Predicting secretory proteins with SignalP. *Methods Mol. Biol.* **1611**, 59–73. (doi:10.1007/978-1-4939-7015-5_6)
28. Nielsen H, Engelbrecht J, Brunak S, von Heijne G. 1997 Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1–6. (doi:10.1093/protein/10.1.1)
29. Möller S, Croning MD, Apweiler R. 2001 Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **17**, 646–653. (doi:10.1093/bioinformatics/17.7.646)
30. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001 Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580. (doi:10.1006/jmbi.2000.4315)
31. Sonnhammer EL, von Heijne G, Krogh A. 1998 A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **6**, 175–182.
32. Jain A, Tuteja G. 2019 TissueEnrich: Tissue-specific gene enrichment analysis. *Bioinformatics* **35**, 1966–1967. (doi:10.1093/bioinformatics/bty890)
33. Hooper NM, Karran EH, Turner AJ. 1997 Membrane protein secretases. *Biochem. J.* **321**, 265–279. (doi:10.1042/bj3210265)
34. Lichtenthaler SF, Lemberg MK, Fluhrer R. 2018 Proteolytic ectodomain shedding of membrane proteins in mammals-hardware, concepts, and recent developments. *EMBO J.* **37**, e99456. (doi:10.15252/embj.201899456)
35. Althoff K, Müllberg J, Aasland D, Voltz N, Kallen K, Grötzinger J, Rose-John S. 2001 Recognition sequences and structural elements contribute to shedding susceptibility of membrane proteins. *Biochem J.* **353**, 663–672. (doi:10.1042/bj3530663)
36. Hamburger AE, West AP, Bjorkman PJ. 2004 Crystal structure of a polymeric immunoglobulin binding fragment of the human polymeric immunoglobulin receptor. *Structure* **12**, 1925–1935. (doi:10.1016/j.str.2004.09.006)
37. Sanderson MP, Keller S, Alonso A, Riedle S, Dempsey PJ, Altevogt P. 2008 Generation of novel, secreted epidermal growth factor receptor (EGFR/ErbB1) isoforms via metalloprotease-dependent ectodomain shedding and exosome secretion. *J. Cell Biochem.* **103**, 1783–1797. (doi:10.1002/jcb.21569)
38. Garton KJ, Gough PJ, Philalay J, Wille PT, Blobel CP, Whitehead RH, Dempsey PJ, Raines EW. 2003 Stimulated shedding of vascular cell adhesion molecule 1 (VCAM-1) is mediated by tumor necrosis factor-alpha-converting enzyme (ADAM 17). *J. Biol. Chem.* **278**, 37 459–37 464. (doi:10.1074/jbc.M305877200)
39. Hashimoto M, Yamazaki A, Ohno A, Kimura T, Winblad B, Tjernberg LO. 2019 A fragment of S38AA is a novel plasma biomarker of Alzheimer's disease. *J Alzheimers Dis.* **71**, 1163–1174. (doi:10.3233/JAD-190700)
40. Zhang M *et al.* 2020 A translocation pathway for vesicle-mediated unconventional protein secretion. *Cell.* **181**, 637–52.e15. (doi:10.1016/j.cell.2020.03.031)
41. Rubartelli A, Bajetto A, Allavena G, Wollman E, Sitia R. 1992 Secretion of thioredoxin by normal and neoplastic cells through a leaderless secretory pathway. *J. Biol. Chem.* **267**, 24 161–24 164. (doi:10.1016/S0021-9258(18)35742-9)
42. Nüchel J *et al.* 2018 TGFB1 is secreted through an unconventional pathway dependent on the autophagic machinery and cytoskeletal regulators. *Autophagy* **14**, 465–486. (doi:10.1080/15548627.2017.1422850)
43. Lavoie JL, Sigmund CD. 2003 Minireview: overview of the renin-angiotensin system—an endocrine and paracrine system. *Endocrinology* **144**, 2179–2183. (doi:10.1210/en.2003-0150)
44. Matsusaka T, Niimura F, Shimizu A, Pastan I, Saito A, Kobori H, Nishiyama A, Ichikawa I. 2012 Liver angiotensinogen is the primary source of renal angiotensin II. *J. Am. Soc. Nephrol.* **23**, 1181–1189. (doi:10.1681/ASN.2011121159)
45. Catta-Preta M, Martins MA, Cunha Brunini TM, Mendes-Ribeiro AC, Mandarim-de-Lacerda CA, Aguila MB. 2012 Modulation of cytokines, resistin, and distribution of adipose tissue in C57BL/6 mice by different high-fat diets. *Nutrition.* **28**, 212–219. (doi:10.1016/j.nut.2011.05.011)
46. Shimamura M *et al.* 2007 Angiopoietin-like protein3 regulates plasma HDL cholesterol through suppression of endothelial lipase. *Arterioscler. Thromb. Vasc. Biol.* **27**, 366–372. (doi:10.1161/01.ATV.0000252827.51626.89)
47. Leonard EJ, Skeel AH. 1978 Isolation of macrophage stimulating protein (MSP) from human serum. *Exp. Cell Res.* **114**, 117–126. (doi:10.1016/0014-4827(78)90043-5)
48. Burch PM *et al.* 2017 Reduced serum myostatin concentrations associated with genetic muscle disease progression. *J. Neurol.* **264**, 541–553. (doi:10.1007/s00415-016-8379-6)
49. Spaks A. 2017 Role of CXC group chemokines in lung cancer development and progression. *J. Thorac. Dis.* **9**(Suppl 3), S164–SS71. (doi:10.21037/jtd.2017.03.61)
50. Elis S, Courtland HW, Wu Y, Rosen CJ, Sun H, Jepsen KJ, Majeska RJ, Yakar S. 2010 Elevated serum levels of IGF-1 are sufficient to establish normal body size and skeletal properties even in the absence of tissue IGF-1. *J. Bone Miner. Res.* **25**, 1257–1266. (doi:10.1002/jbmr.20)
51. Katafuchi T, Esterházy D, Lemoff A, Ding X, Sondhi V, Kliewer SA, Mirzaei H, Mangelsdorf D. 2015 Detection of FGF15 in plasma by stable isotope standards and capture by anti-peptide antibodies and targeted mass spectrometry. *Cell Metab.* **21**, 898–904. (doi:10.1016/j.cmet.2015.05.004)
52. Cook KS, Min HY, Johnson D, Chaplinsky RJ, Flier JS, Hunt CR, Spiegelman BM. 1987 Adipsin: a circulating serine protease homolog secreted by adipose tissue and sciatic nerve. *Science* **237**, 402–405. (doi:10.1126/science.3299705)
53. Favre D, Ferry N, Moullier P. 2000 Critical aspects of viral vectors for gene transfer into the kidney. *J. Am. Soc. Nephrol.* **11**(Suppl 16), S149–S153. (doi:10.1681/ASN.V11suppl_2s149)
54. Kim DI, Jensen SC, Noble KA, Kc B, Roux KH, Motamedchaboki K, Zheng Y. 2016 An improved smaller biotin ligase for BioID proximity labeling. *Mol. Biol. Cell.* **27**, 1188–1196. (doi:10.1091/mbc.E15-12-0844)
55. Stachon P *et al.* 2016 Extracellular ATP induces vascular inflammation and atherosclerosis via purinergic receptor Y2 in mice. *Arterioscler. Thromb. Vasc. Biol.* **36**, 1577–1586. (doi:10.1161/ATVBAHA.115.307397)
56. Hughes ED *et al.* 2007 Genetic variation in C57BL/6 ES cell lines and genetic instability in the Bruce4 C57BL/6 ES cell line. *Mamm Genome.* **18**, 549–558. (doi:10.1007/s00335-007-9054-0)
57. Raymond CS, Soriano P. 2007 High-efficiency FLP and PhiC31 site-specific recombination in mammalian cells. *PLoS ONE* **2**, e162. (doi:10.1371/journal.pone.0000162)
58. Hayashi S, McMahon AP. 2002 Efficient recombination in diverse tissues by a tamoxifen-inducible form of Cre: a tool for temporally regulated gene activation/inactivation in the mouse. *Dev. Biol.* **244**, 305–318. (doi:10.1006/dbio.2002.0597)
59. Postic C *et al.* 1999 Dual roles for glucokinase in glucose homeostasis as determined by liver and pancreatic beta cell-specific gene knock-outs using Cre recombinase. *J. Biol. Chem.* **274**, 305–315. (doi:10.1074/jbc.274.1.305)
60. Hughes CS, Foehr S, Garfield DA, Furlong EE, Steinmetz LM, Krijgsveld J. 2014 Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol. Syst. Biol.* **10**, 757. (doi:10.15252/msb.20145625)
61. Deng W, Sha J, Plath K, Wohlschlegel JA. 2021 Carboxylate-modified magnetic bead (CMMB)-based isopropanol gradient peptide fractionation (CIF) enables rapid and robust off-line peptide mixture fractionation in bottom-up proteomics. *Mol. Cell. Proteomics.* **20**, 100039. (doi:10.1074/mcp.RA120.002411)
62. Yang R *et al*. 2022 Data from: A genetic model for *in vivo* proximity labeling of the mammalian secretome. Figshare. (doi:10.6084/m9.figshare.c.6125239)