Integrated view and comparative analysis of baseline protein expression in mouse and rat tissues

Shengbo Wang[1†], David García-Seisdedos[1,2†], Ananth Prakash[1,2†*], Deepti Jaiswal Kundu[1], Andrew Collins[3], Nancy George[1], Silvie Fexova[1], Pablo Moreno[1], Irene Papatheodorou[1,2], Andrew R. Jones[3], Juan Antonio Vizcaíno[1,2 *]

[1] European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD. United Kingdom.

[2] Open Targets, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD. United Kingdom.

[3] Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool L69 7ZB, United Kingdom.

*Corresponding authors.

†All three authors have contributed equally and they wish to be considered as joint first authors.

Dr. Ananth Prakash. European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK. Phone: + 44 (0) 1223 492610. Email: ananth@ebi.ac.uk

Dr. Juan Antonio Vizcaíno. European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK. Phone: + 44 (0) 1223 492686. Email: juan@ebi.ac.uk.

Keywords

Mass spectrometry, quantitative proteomics, comparative protein expression, public data re-use, rat proteome, mouse proteome, PRIDE

Abstract

The increasingly large amount of public proteomics data enables, among other applications, the combined analyses of datasets to create comparative protein expression maps covering different organisms and different biological conditions. Here we have reanalysed public proteomics datasets from mouse and rat tissues (14 and 9 datasets, respectively), to assess baseline protein abundance. Overall, the aggregated dataset contains 23 individual datasets, which have a total of 211 samples coming from 34 different tissues across 14 organs, comprising 9 mouse and 3 rat strains, respectively.

We compared protein expression between the different organs, including the distribution of proteins across them. We also performed gene ontology and pathway enrichment analyses to identify organ-specific enriched biological processes and pathways. As a key point we carried out a comparative analysis of protein expression between mouse, rat and human tissues. We observed a high level of correlation of protein expression among orthologs between all three species in brain, kidney, heart and liver samples. Finally, it should be noted that protein expression results have been integrated into the resource Expression Atlas for widespread dissemination.

54 **Author summary**

55

56 We have reanalysed 23 baseline public proteomics datasets stored in the PRIDE database
57 belonging to various mouse and rat tissues and organs. Overall, the aggregated dataset
58 contained 211 samples, coming from 34 different tissues across 14 organs, comprising 9
59 mouse and 3 rat strains, respectively. We analysed the distribution of protein expression
60 across organs in both species. Then we performed gene ontology and pathway enrichment
61 analyses to identify organ-specific enriched biological processes and pathways. We also
62 carried out a comparative analysis of baseline protein expression between mouse, rat and
63 human tissues, observing a high level of expression correlation among orthologs in all three
64 species, in brain, kidney, heart and liver samples. To disseminate these findings, we have
65 integrated the protein expression results into the resource Expression Atlas.

1. Introduction

High-throughput mass spectrometry (MS)-based proteomics approaches have matured significantly in recent years, becoming an essential tool in biological research [1]. This has been the consequence of very significant technical improvements in MS instrumentation, chromatography, automation in sample preparation and computational analyses, among other areas. The most used MS-based experimental approach is Data Dependent Acquisition (DDA) bottom-up proteomics. Among the main quantitative proteomics DDA techniques, label-free approaches remain very popular, although labelled-approaches, especially those techniques based on the isotopic labelling of peptides (MS^2 labelling), such as iTRAQ (Isobaric tag for relative and absolute quantitation) and TMT (Tandem Mass Tagging), are increasingly popular as well.

Following the steps initiated by genomics and transcriptomics, open data practices in the field have become embedded and commonplace in proteomics in recent years. In this context, datasets are now commonly available in the public domain to support the claims published in the corresponding manuscripts. The PRIDE database [2], located at the European Bioinformatics Institute (EBI), is currently the largest resource worldwide for public proteomics data deposition. PRIDE is also one of the founding members of the global ProteomeXchange consortium [3], involving five other resources, namely PeptideAtlas, MassIVE, iProX, jPOST and PanoramaPublic. ProteomeXchange has standardised data submission and dissemination of public proteomics data worldwide.

As a consequence, there is an unprecedented availability of data in the public domain, which is triggering multiple applications [4], including the joint reanalysis of datasets (so-called

91 meta-analysis studies) [5-7]. Indeed, public proteomics datasets can be systematically
92 reanalysed and integrated e.g., to confirm the results reported in the original publications,
93 potentially in a more robust manner since evidence can be strengthened if it is found
94 consistently across different datasets. Potentially, new insights different to the aims of the
95 original studies can also be obtained by re-analysing the datasets using different strategies. It
96 is the so-called re-purposing of public datasets [8], including for instance proteogenomics
97 studies, for genome annotation purposes [9-12].
98
99 In this context of reuse of public proteomics data, PRIDE has started to work on developing
100 data dissemination and integration pipelines into popular added-value resources at the EBI.
101 This is perceived as a more sustainable approach in the medium-long term than setting up
102 new bioinformatics resources. One of them is Expression Atlas [13], a resource that has
103 enabled over the years easy access to gene expression data across species, tissues, cells,
104 experimental conditions and diseases. Only recently, protein expression information coming
105 from reanalysed datasets has been integrated in the 'bulk' section of Expression Atlas. As a
106 result, proteomics expression data can be integrated with transcriptomics information, mostly
107 coming from RNA-Seq experiments. So far, we have performed two main meta-analysis
108 studies involving the reanalysis and integration of: (i) 11 public quantitative datasets coming
109 from cell lines and human tumour samples [14]; and (ii) 25 human baseline datasets coming
110 from 32 different organs [15].
111
112 The next logical step is to perform an analogous study of baseline protein expression in two
113 of the main model organisms: *Mus musculus* and *Rattus norvegicus*. To date, there are only a
114 small number of bioinformatics resources providing access to reanalysed MS-based
115 quantitative proteomics datasets, and even fewer if one considers only mouse and rat data. In

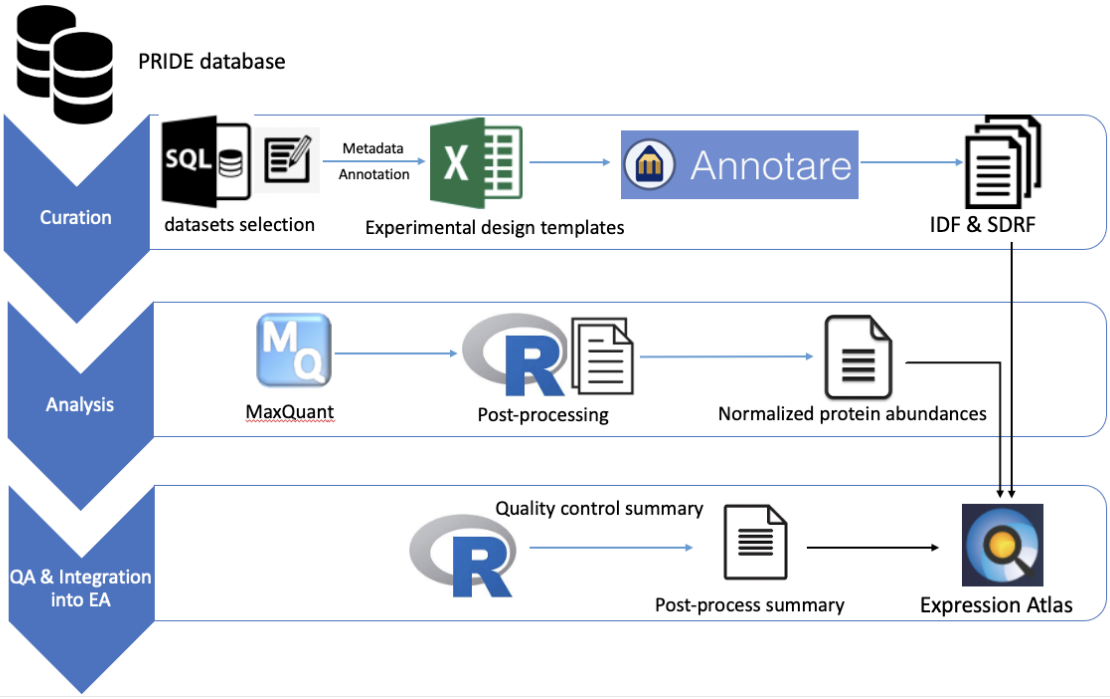
this context, at the end of 2020, ProteomicsDB [16] released a first version of the mouse proteome, based on the re-analysis of five datasets. To the best of our knowledge, there is no such public resource for rat data yet anywhere. Additionally, although antibody-based human protein expression information is provided *via* the Human Protein Atlas [17] , their efforts are focused on human protein expression. Here, we report the reanalysis and integration of 23 public mouse (14 datasets) and rat (9 datasets) label-free datasets, and the incorporation of the results into the resource Expression Atlas as baseline studies. Additionally, we report a comparative analysis of protein expression across mouse, rat and human (in this case using the results reported at [15] using the same methodology).

126 **2. Result**

127

128 **2.1. Baseline proteomics datasets**

129 Overall, we quantified protein expression from 34 healthy tissues in 14 organs coming from
130 23 datasets. The analyses covered a total of 1,173 MS runs from 211 samples that were
131 annotated as healthy/control/non-treated samples thus representing baseline protein
132 expression. Non-control/disease samples associated with these datasets were also analysed
133 but are not discussed here. Normalised protein abundances values (as ppb, parts per billion)
134 from both control/healthy/non-treated and disease/treated tissue samples are available to view
135 as heatmaps in Expression Atlas. The protein abundances along with sample annotations,
136 sample quality assessment summary and experimental parameter inputs for MaxQuant can be
137 downloaded from Expression Atlas as text files. Data selection and reanalysis protocol is
138 shown in Fig. 1, the total number of proteins and peptides identified in these datasets are
139 shown in Table 1.



140

141 **Figure 1.** An overview of the study design and reanalysis pipeline. QA: Quality assessment.

Expression Atlas accession numbers	PRIDE dataset identifiers	Tissues	Organs	Species	Strain	Number of MS runs	Number of samples	Number of protein groups†	Number of peptides †	Number of unique peptides†	Number of unique genes mapped†
E-PROT-7§	PXD000867 [18]	Liver	Liver	*Mus musculus*	C57BL/6J	24	4	12,792	246,738	167,725	9,715
E-PROT-10§	PXD000288 [19]	Triceps muscles	Triceps Muscles	*Mus musculus*	C57BL/6	36	3	10,870	189,553	126,670	6,421
E-PROT-16	PXD003155 [20]	Cerebellum, Liver	Brain, Liver	*Mus musculus*	C57BL/6	24	12	4,508	59,696	45,728	3,797
E-PROT-74	PXD004612 [21]	Achilles and Plantaris tendon	Tendon	*Mus musculus*	C57BL/6	8	8	457	6,643	3,271	273
E-PROT-75	PXD005230 [22]	Hippocampus, Cerebellum, Cortex	Brain	*Mus musculus*	C57BL/10J	72	36	7,663	63,479	41,683	6,037
E-PROT-76	PXD009909 [23]	Retina	Eye	*Mus musculus*	ND4 Swiss Webster	12	1	5,002	29,454	24,961	3,686
E-PROT-77	PXD012307 [24]	Lung	Lung	*Mus musculus*	C57BL/6	32	2	6,809	106,391	73,950	5,795
E-PROT-78	PXD009639 [25]	Lens	Eye	*Mus musculus*	CD1	10	1	4,519	20,779	18,006	3,064
E-PROT-79	PXD019394 [26]	Heart, Kidney, Liver, Lung, Brain, Spleen, Testis, Pancreas	Heart, Kidney, Liver, Lung, Brain, Spleen, Testis, Pancreas	*Mus musculus*	Swiss-Webster	96	8	9,853	141,506	105,701	8,185
E-PROT-81	PXD012636 [27]	Left atrium, Left ventricle, Right atrium, Right ventricle	Heart	*Mus musculus*	C57BL/6	120	4	7,772	146,966	99,577	6,435
E-PROT-82	PXD019431 [28]	Articular cartilage	Articular cartilage	*Mus musculus*	BALB\ c	72	6	1,815	17,695	15,191	1,518
E-PROT-83	PXD022614 [29]	Brain	Brain	*Mus musculus*	C57BL/6J: Rj	120	6	6,645	97,443	69,884	5,673

					C57BL/6JR ccHsd						
E-PROT-84	PXD004496 [30]	Hippocampus	Brain	*Mus musculus*	C57BL/6J	204	17	4,192	37,363	30,100	3,424
E-PROT-85	PXD008736 [31]	Right atrium, Sinus node	Heart	*Mus musculus*	C57BL/6J	143	6	7,906	144,926	94,379	6,554
E-PROT-86§	PXD012677 [32]	Amygdala	Brain	*Rattus norvegicus*	Sprague Dawley	3	3	1,872	15,326	12,367	1,382
E-PROT-87§	PXD006692 [33]	Lung	Lung	*Rattus norvegicus*	Sprague Dawley	10	10	2,079	14,440	11,696	1,398
E-PROT-88§	PXD016793 [34]	Liver	Liver	*Rattus norvegicus*	Sprague Dawley	8	8	4,787	57,998	46,411	3,743
E-PROT-89§	PXD004364 [35]	Testis	Testis	*Rattus norvegicus*	Sprague Dawley	3	3	2,351	15,880	13,674	1,601
E-PROT-91	PXD001839 [36]	Left ventricle	Heart	*Rattus norvegicus*	F344/BN	12	12	1,345	10,310	8,804	925
E-PROT-92§	PXD013543 [37]	Left ventricle	Heart	*Rattus norvegicus*	Wistar	8	8	1,858	17,303	13,622	1,340
E-PROT-93	PXD016958 [38]	First segment of proximal tubule, second segment of proximal tubule, third segment of proximal tubule, medullary thick ascending limb, cortical thick ascending limb, distal convoluted tubule, connecting tubule, cortical collecting duct, outer medullary collecting duct, inner medullary collecting duct	Kidney	*Rattus norvegicus*	Sprague Dawley	132	32	7,846	103,886	83,662	6,130

E-PROT-94	PXD003375 [39]	Caudal and rostral segments of spinal cord	Spinal cord	*Rattus norvegicus*	Wistar	21	18	2,477	29,213	22,025	1,926
E-PROT-95§	PXD015928 [40]	Tendon	Tendon	*Rattus norvegicus*	Wistar	3	3	199	1,253	1,063	101
TOTAL	23 datasets (Mouse: 14, Rat: 9)	34 tissues (Mouse: 21, Rat: 18)	14 organs (Mouse: 12, Rat: 8)			1,173 MS runs (Mouse: 973, Rat: 200)	211 samples (Mouse: 114, Rat: 97)				

Table 1. List of mouse and rat proteomics datasets that were re-analysed. §Only normal/untreated samples within this dataset are reported in this study. However, results from both normal and disease samples are available in Expression Atlas. † Numbers after post-processing.

146

147 **2.2. Protein coverage across organs and datasets**

148 One of the main aims of this study was to study protein expression across various organs. To
149 enable simplicity of comparison [15] we first grouped 34 different tissues into 14 distinct
150 organs, as discussed in 'Methods'. We defined 'tissue' as a distinct functional or structural
151 region within an 'organ'. We estimated the number of 'canonical proteins' identified across
152 organs by first mapping all members of each protein group to their respective parent genes.
153 We defined the parent gene as equivalent to the UniProt 'canonical protein' and we will
154 denote the term 'protein abundance' to mean 'canonical protein abundance' from here on in
155 the manuscript.

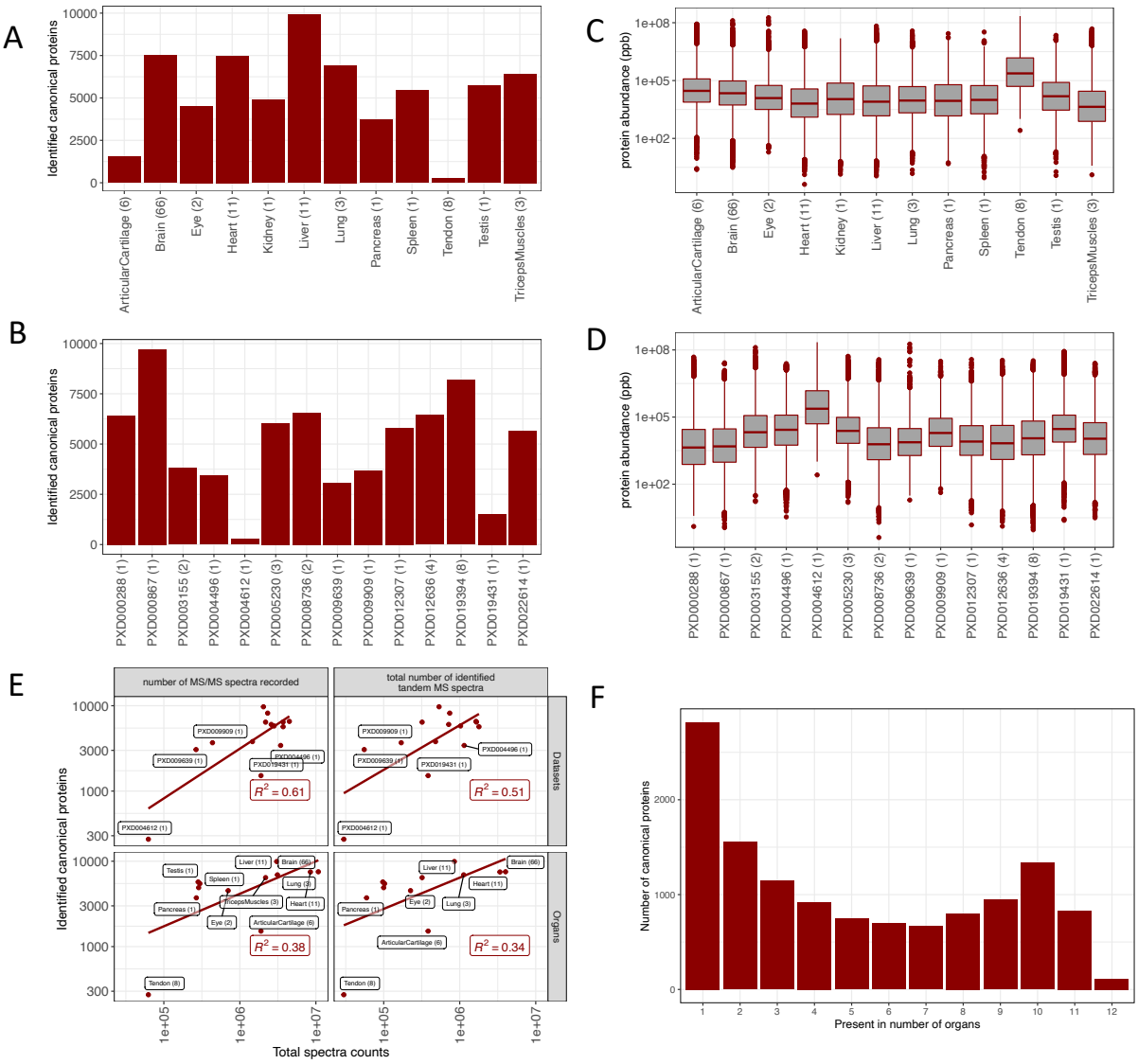
156

157 **2.2.1. Mouse proteome**

158 A total of 21,274 protein groups were identified from mouse datasets, among which 8,176
159 protein groups (38.4%) were uniquely present in only one organ and 70 protein groups
160 (0.3%) were ubiquitously observed (see the full list in Supplementary file 2). Mouse protein
161 groups were mapped to 12,570 genes (canonical proteins) (Supplementary file 3). We
162 detected the largest number of canonical proteins in samples from liver (9,920, 78.9% of the
163 total) and the lowest numbers in samples from tendon (273, 2.2%) and articular cartilage
164 (1,519, 12.1%) (Fig. 2A). The lower number of proteins identified in tendon could be
165 attributed to the smallest sample size (only one sample out of 114, 0.9%). Also, tendon is a
166 relatively hypocellular tissue, which has a low protein turnover rate. Dataset PXD000867,
167 containing mouse liver samples, had the highest number of canonical proteins detected
168 (9,715, 77.3%), while the least number of proteins were detected in dataset PXD004612
169 (tendon, 273, 2.2%), as highlighted above (Fig. 2B).

170

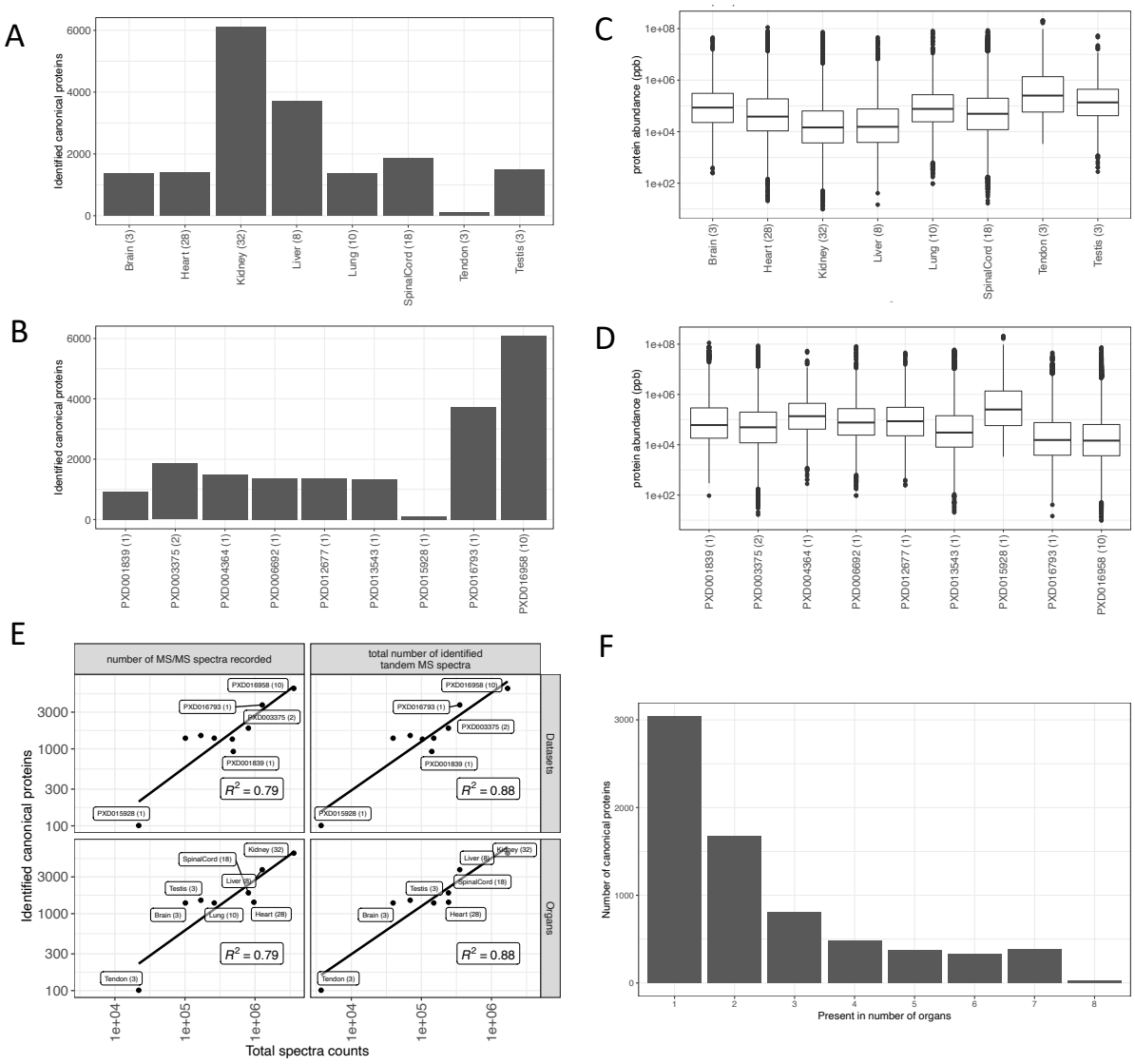
We studied the normalised protein abundance distribution in organs (Fig. 2C) and found that all organs, except tendon, had similar median abundances. However, one cannot attribute biological meaning to these observations, since the method of normalisation by definition fixes each sample to have the same “total abundance”, which then gets shared out amongst all proteins. The normalised protein abundance distribution in datasets indicated a higher than median abundances detected in datasets PXD004612 (tendon) and PXD003164 (testis) (Fig. 2D). A linear relation was observed with the number of canonical proteins detected in datasets and organs, when compared to the relative amount of their spectral data (Fig. 2E). We found a large number of proteins uniquely present in one organ (Fig. 2F).



182 **Figure 2.** (A) Number of canonical proteins identified across different mouse organs. The
183 number within the parenthesis indicates the number of samples. (B) Canonical proteins
184 identified across different datasets. The number within the parenthesis indicate the number of
185 unique tissues in the dataset. (C) Range of normalised iBAQ protein abundances across
186 different organs. The number within the parenthesis indicates the number of samples. (D)
187 Range of normalised iBAQ protein abundances across different datasets. The number within
188 parenthesis indicate the number of unique tissues in the dataset. (E) Comparison of total
189 spectral data with the number of canonical proteins identified in each dataset and organ. (F)
190 Distribution of canonical proteins identified across organs.
191
192 **2.2.2. Rat proteome**

193 A total of 7,769 protein groups were identified across 8 different rat organs among which
194 3,649 (46.9%) protein groups were unique to one specific organ while 13 (0.16%) protein
195 groups were present among all organs (see full list in Supplementary file 2). The protein
196 groups were mapped to 7,116 genes (canonical proteins) (Supplementary file 3). The highest
197 number of canonical proteins (6,106, 85.1%) was found in rat kidney samples. The lowest
198 number of canonical proteins (101, 1.4%) was found in samples from tendon, as shown in
199 Fig. 3A. The largest number of canonical proteins identified in kidney is because of the
200 relatively large number of samples (32 samples), when compared to other organs. However, it
201 is interesting to note that large number of canonical proteins were detected in liver samples,
202 which relatively has fewer number of samples, when compared to number of samples in heart
203 or spinal cord.
204
205 Datasets PXD016958 and PXD016793 consisted entirely of kidney and liver samples,
206 respectively, and as mentioned above had the largest number of canonical proteins identified

207 (Fig. 3B). The normalised protein abundances were similar among the various organs and
208 datasets (Fig. 3C, D). We also observed a linear relation between the number of canonical
209 proteins identified and the MS spectra identified (Fig. 3E). As seen in mouse datasets, we
210 also observed a large number of proteins uniquely present in one organ (Fig. 3F).
211



212

213 **Figure 3.** (A) Number of canonical proteins identified across different rat organs. The
214 number within the parenthesis indicates the number of samples. (B) Canonical proteins
215 identified across different datasets. The number within the parenthesis indicate the number of
216 unique tissues in the dataset. (C) Range of normalised iBAQ protein abundances across

217 different organs. The number within the parenthesis indicates the number of samples. (D)
218 Range of normalised iBAQ protein abundances across different datasets. The number within
219 parenthesis indicate the number of unique tissues in the dataset. (E) Comparison of total
220 spectral data with the number of canonical proteins identified in each dataset and organ. (F)
221 Distribution of canonical proteins identified across organs.
222
223 **2.3. Protein abundance comparison across organs**
224 Next, we studied how protein abundances compared across different datasets and organs. The
225 presence of batch effects between datasets makes this type of comparisons challenging. To
226 aid comparison of protein abundances between datasets we transformed the normalised iBAQ
227 intensities into ranked bins as explained in 'Methods', i.e., proteins in bin 5 are highly
228 abundant while proteins in bin 1 are expressed in lower abundances.
229
230 **2.3.1. Mouse proteome**
231 We found that 1,086 (8.6%) proteins were found with their highest expression in at least 3
232 organs, with a median bin value greater than 4 (Supplementary file 3). On the other end of the
233 scale, 138 (1.1%) canonical proteins were found with their lowest expression in at least 3
234 organs, with a median bin value less than 2. The bin transformed abundances in all organs are
235 provided in Supplementary file 3.
236
237 To compare protein expression across all organs in mouse, we calculated pairwise Pearson
238 correlation coefficients (rp) across 117 samples (Fig. 4A). We observed a good correlation of
239 protein expression within the brain (median rp = 0.55) and heart (median rp = 0.82) samples.
240 We performed PCA on all samples from mouse datasets for testing the effectiveness of bin
241 transformation method in reducing batch effects. Fig. 4B shows the clustering of samples

from various organs of mouse. We observed samples from the same organ clustered together. For example, we can observe that brain samples all clustered together in one group, although they come from different datasets (Fig. 4C). However, we also observed that samples from other organs such as liver, did not cluster according to their organ types but clustered together with the dataset they were part of, indicating the presence of residual batch effects.

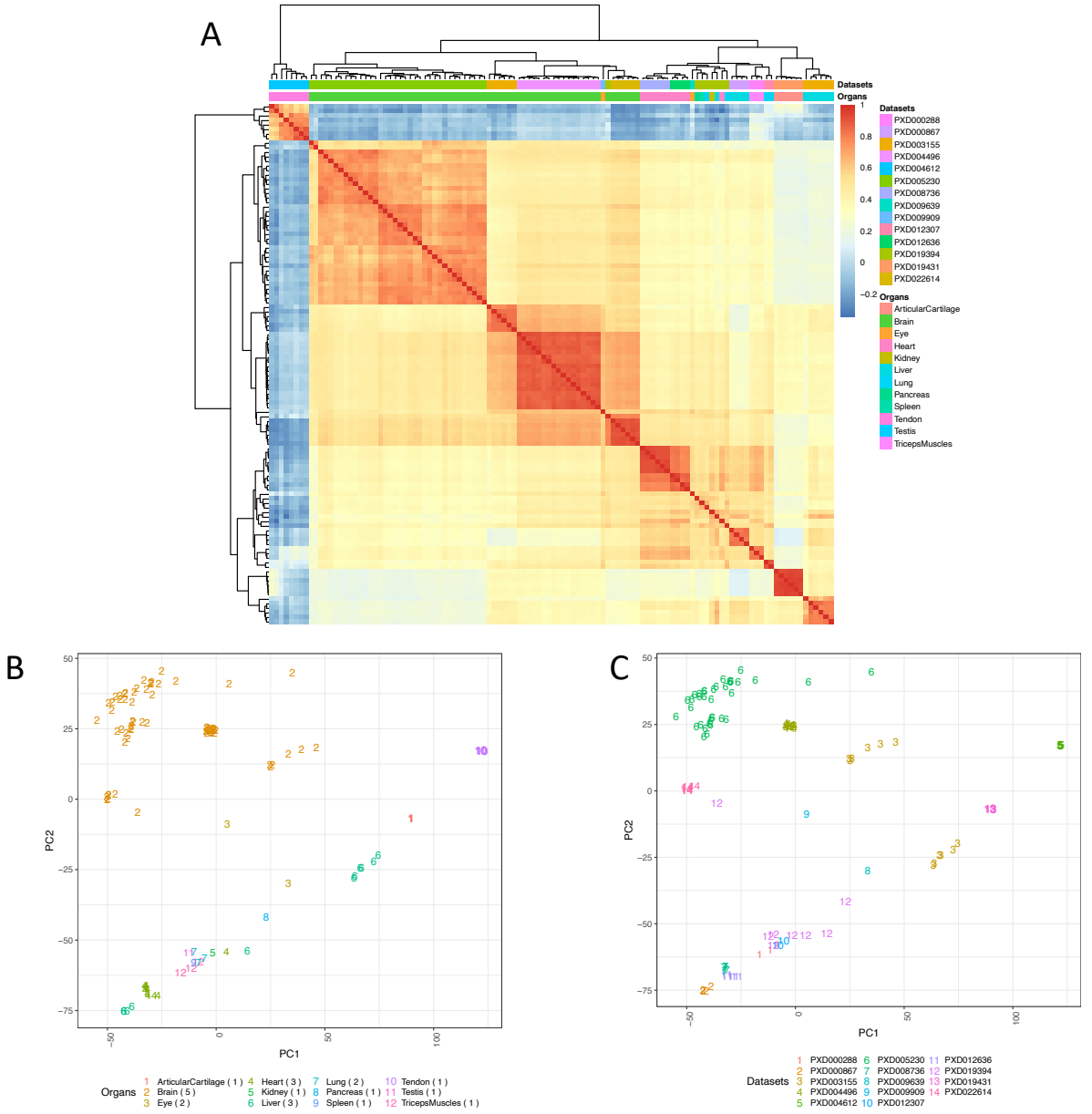


Figure 4. (A) Heatmap of pairwise Pearson correlation coefficients across all mouse samples. The colour represents the correlation coefficient and was calculated using the bin transformed

251 iBAQ values. The samples were hierarchically clustered on columns and rows using
252 Euclidean distances. (B) PCA of all samples coloured by the organ types. (C) PCA of all
253 samples coloured by their respective dataset identifiers. The numbers in parenthesis indicate
254 the number of datasets for each organ. Binned values of canonical proteins quantified in at
255 least 50% of the samples were used to perform the PCA.
256
257 **2.3.2. Rat proteome**
258 Similarly, in rat, on one hand, 311 (4.3%) proteins were found with highest expression in at
259 least 3 organs with a median bin value greater than 4. On the other hand, 27 (0.37%)
260 canonical proteins were found with their lowest expression in at least 3 organs, with a median
261 bin value of less than 2. The bin transformed abundances in all organs are provided in
262 Supplementary file 3.
263
264 The samples from rat datasets were also better correlated among each other (Fig. 5A) than in
265 the case of mouse. We observed a good correlation of protein expression within samples from
266 liver (median Pearson's correlation rp = 0.92), lung (median rp = 0.84), spinal cord (median
267 rp = 0.80), heart (median rp = 0.84) and brain (median rp = 0.93). We also observed a good
268 clustering in the PCA of samples coming from the same organ type (Fig. 5B). Kidney, lung,
269 spinal cord and heart samples all clustered together according to their organ type. Fig. 5C
270 shows the samples based on the dataset they are part of. However, most organ samples are
271 part of individual datasets except samples from heart, which came from two datasets
272 (PXD001839 and PXD013543). Fig. 5C shows that the heart samples clustered into two
273 groups, wherein each cluster included samples from a different dataset.

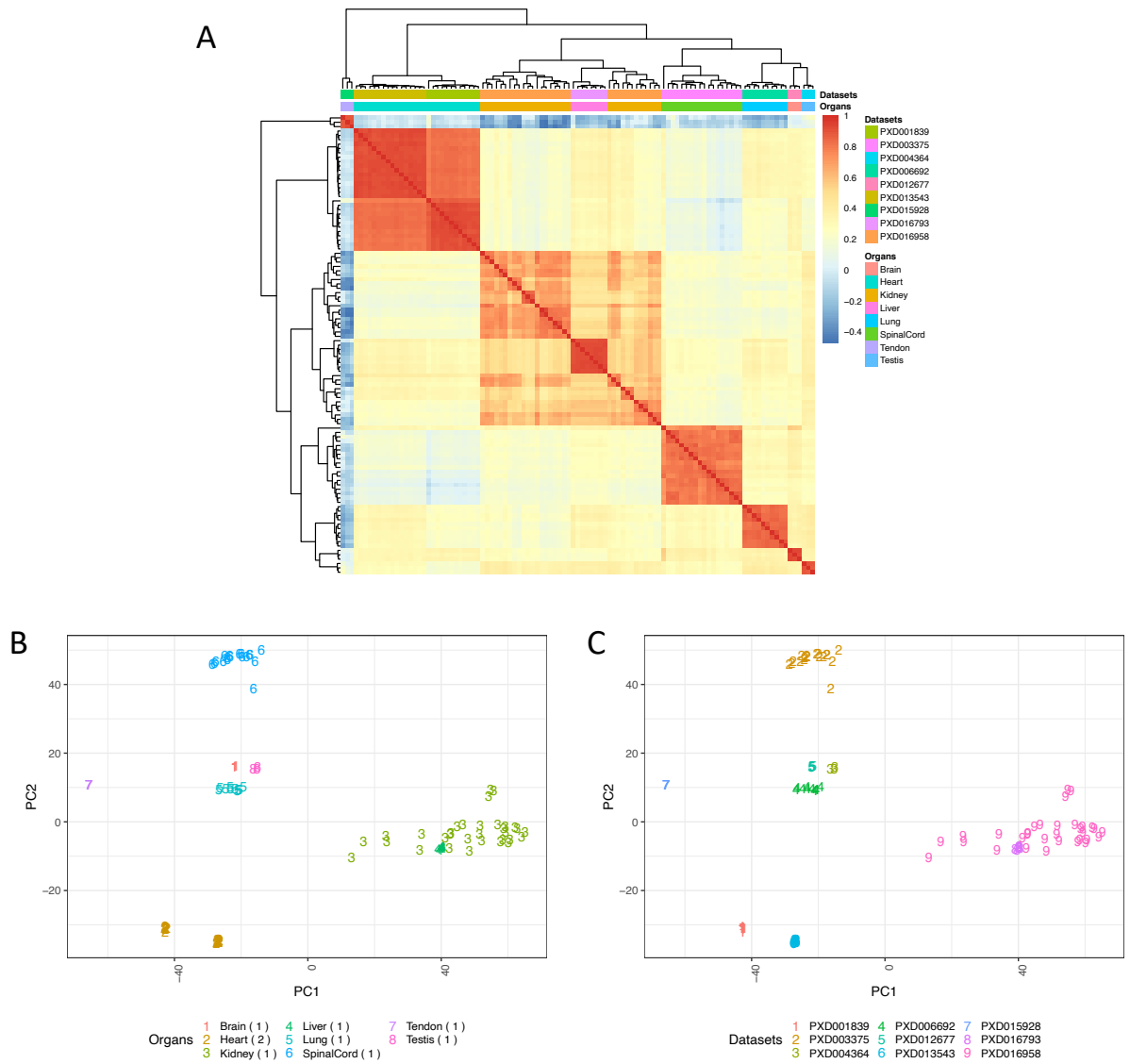
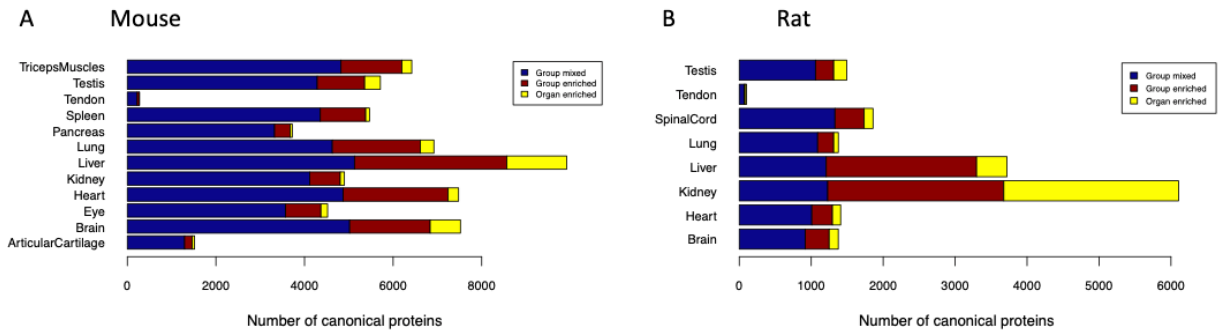


Figure 5. (A) Heatmap of pairwise Pearson correlation coefficients across all rat samples. The colour represents the correlation coefficient and was calculated using the bin transformed iBAQ values. The samples were hierarchically clustered on columns and rows using Euclidean distances. (B) PCA of all samples coloured by the organ types. (C) PCA of all samples coloured by their respective dataset identifiers. The numbers in parenthesis indicate the number of datasets for each organ. Binned values of canonical proteins quantified in at least 50% of the samples were used to perform the PCA.

2.4. The organ elevated proteome and the over-representative biological processes

284 Based on their expression, canonical proteins were classified into three different groups based
285 on their organ specificity: “mixed”, “group-enriched” and “organ-enriched” (see
286 Supplementary file 4). We considered over-expressed canonical proteins in each organ as
287 those which were in “group-enriched” and “organ-enriched”. The analysis showed that on
288 average, 20.8% and 26.0% of the total elevated canonical proteins were organ group-specific
289 in mouse and rat, respectively (Fig. 6). In addition, 4.3% and 14.2% were unique organ-
290 enriched in mouse and rat, respectively. The highest ratio of organ-enriched in mouse was
291 found in liver (13.6%), whereas it was found in kidney in rat (39.8%).

292 We then performed a gene ontology (GO) enrichment analysis of those proteins that were
293 'organ-enriched' and group-enriched' using GO terms associated with biological processes.
294 We found 1,036 GO terms to be statistically significant in all organs, as seen in
295 Supplementary file 5. The most significant GO terms for each organ are shown in Table 2.
296



297

298 **Figure 6.** Organ specificity of canonical proteins in (A) mouse and (B) rat.

299

Organ	Species	GO ID	Description	adjusted p-value
Articular cartilage	*Mus musculus*	GO:0030198 GO:0043062 GO:0045229	Extracellular matrix organization Extracellular structure organization External encapsulating structure organization	$8.94*10^{-38}$ $8.94*10^{-38}$ $8.94*10^{-38}$
Brain	*Mus musculus*	GO:0050804 GO:0099177 GO:0050808	Modulation of chemical synaptic transmission Regulation of trans-synaptic signalling Synapse organization	$7.03*10^{-65}$ $7.03*10^{-65}$ $1.41*10^{-48}$

Heart	*Mus musculus*	GO:0060047 GO:0008016 GO:0060537	Heart contraction Regulation of heart contraction Muscle tissue development	$7.10*10^{-11}$ $4.43*10^{-10}$ $6.16*10^{-10}$
Kidney	*Mus musculus*	GO:0015711 GO:0044282 GO:0016054	Organic anion transport Small molecule catabolic process Organic acid catabolic process	$4.59*10^{-19}$ $4.91*10^{-15}$ $6.25*10^{-15}$
Eye	*Mus musculus*	GO:0007601 GO:0001654 GO:0099504	Visual perception Eye development Synaptic vesicle cycle	$7.54*10^{-50}$ $5.31*10^{-31}$ $8.36*10^{-18}$
Liver	*Mus musculus*	GO:0016569 GO:0016570 GO:0019369	Covalent chromatin modification Histone modification Arachidonic acid metabolic process	$6.26*10^{-10}$ $1.71*10^{-08}$ $1.71*10^{-08}$
Lung	*Mus musculus*	GO:0120031 GO:0030031 GO:0044782	Plasma membrane bounded cell projection assembly Cell projection assembly Cilium organization	$3.61*10^{-14}$ $3.61*10^{-14}$ $9.83*10^{-14}$
Pancreas	*Mus musculus*	GO:0007586 GO:0032328	Digestion Alanine transport	0.005 0.018
Spleen	*Mus musculus*	GO:0046649 GO:0050776 GO:0045087	Lymphocyte activation Regulation of immune response Innate immune response	$4.12*10^{-22}$ $2.00*10^{-20}$ $2.23*10^{-20}$
Tendon	*Mus musculus*	GO:0003012 GO:0050879 GO:0050881	Muscle system process Multicellular organismal movement Musculoskeletal movement	$1.46*10^{-25}$ $3.14*10^{-19}$ $1.46*10^{-25}$
Testis	*Mus musculus*	GO:0048232 GO:0003341 GO:0044782	Male gamete generation Cilium movement Cilium organization	$8.75*10^{-49}$ $3.04*10^{-38}$ $6.78*10^{-37}$
Triceps muscles	*Mus musculus*	GO:0061061 GO:0055002 GO:0003009	Muscle structure development Striated muscle cell development Skeletal muscle contraction	$1.56*10^{-14}$ $2.41*10^{-14}$ $3.53*10^{-14}$
Brain	*Rattus norvegicus*	GO:0099537 GO:0007268 GO:0098916	Trans-synaptic signalling Chemical synaptic transmission Anterograde trans-synaptic signalling	$1.79*10^{-60}$ $1.79*10^{-60}$ $1.79*10^{-60}$
Heart	*Rattus norvegicus*	GO:0061061 GO:0003012 GO:0055001	Muscle structure development Muscle system process Muscle cell development	$2.94*10^{-17}$ $6.30*10^{-16}$ $4.00*10^{-15}$
Kidney	*Rattus norvegicus*	GO:0006396 GO:0045944 GO:0006260	RNA processing positive regulation of transcription by RNA polymerase II DNA replication	$6.19*10^{-13}$ $7.29*10^{-06}$ $1.74*10^{-05}$
Liver	*Rattus norvegicus*	GO:0008202 GO:0016054 GO:0032787	Steroid metabolic process Organic acid catabolic process Monocarboxylic acid metabolic process	$2.74*10^{-10}$ $1.61*10^{-09}$ $1.64*10^{-09}$
Lung	*Rattus norvegicus*	GO:0031589 GO:0009617 GO:0030036	Cell-substrate adhesion Response to bacterium Actin cytoskeleton organization	$7.62*10^{-08}$ $7.62*10^{-08}$ $1.40*10^{-07}$
Spinal cord	*Rattus norvegicus*	GO:0061564 GO:0099537 GO:0007268	Axon development Trans-synaptic signalling Chemical synaptic transmission	$4.26*10^{-18}$ $5.93*10^{-16}$ $5.93*10^{-16}$
Tendon	*Rattus norvegicus*	GO:0030199 GO:0061448 GO:0001501	Collagen fibril organization Connective tissue development Skeletal system development	$1.23*10^{-13}$ $2.31*10^{-09}$ $3.39*10^{-09}$

Testis	*Rattus norvegicus*	GO:0019953 GO:0051704 GO:0007018	Sexual reproduction Multi-organism process Microtubule-based movement	$3.98*10^{-24}$ $1.61*10^{-18}$ $4.00*10^{-12}$

300

301 **Table 2.** Analysis of the top three GO terms for each organ in mouse and rat using the
302 elevated organ-specific canonical proteins and group-specific as described in the 'Methods'
303 section.

304

305 **2.5. Protein abundances across orthologs in three species**

306 We had previously analysed 25 label-free proteomics datasets from healthy human samples to
307 assess baseline protein abundances in 14 organs following the same analytical methodology
308 [15]. We compared the expression of canonical proteins identified in all three organisms (rat,
309 mouse and human). Overall, 13,248 detected human genes (corresponding to the canonical
310 proteins) were compared with 12,570 genes detected in mouse and 7,116 genes detected in
311 rat. The number of orthologous mappings (i.e., "one-to-one" mappings, see 'Methods')
312 between rat, mouse and human genes are listed in table 3. We only considered one-to-one
313 mapped orthologues for the comparison of protein abundances.

314

Species	**Identified genes**	**Orthologs of human genes identified in [15]**	**Percentage of genes with different mapping against identified human genes**				
			one-to-one	**one-to-many**	**many-to-many**	**many-to-one**	**not mapped**
Mus musculus	12,570	10,601	80.4%	1.9%	0.56%	1.46%	15.7%
Rattus norvegicus	7,116	6,058	82.0%	2.2%	0.70%	0.25%	14.9%

315

316 **Table 3.** Homologs identified in mouse and rat datasets when compared with the background
317 list of genes identified in human datasets (Supplementary file 2 in [15]).

318 Among human and mouse orthologues we observed relatively high levels of correlation of
319 protein abundances in brain (R^2 = 0.61), heart (R^2 = 0.65) and liver (R^2 = 0.56) (Fig. 7A).
320 Human and rat orthologs showed also relatively high levels of correlation in brain (R^2 =
321 0.62), kidney (R^2 = 0.53) and liver (R^2 = 0.56), but almost no correlation in lung (R^2 = 0.12)
322 and testis (R^2 = 0.18) (Fig. 7B). Between mouse and rat orthologs, the correlation of protein
323 abundances was higher in liver (R^2 = 0.65), kidney (R^2 = 0.54) and brain (R^2 = 0.57) samples,
324 when compared to the samples coming from the rest of the organs (Fig. 7C). Fig. 7D shows
325 an illustration of some example comparisons of individual orthologs using binned protein
326 abundances.
327
328 For the same corresponding subsets, we also investigated the correlation of protein
329 expression between various organs within each organism. We observed that in general the
330 correlation of protein expression was slightly lower between organs within the same species,
331 when compared to a higher correlation, which was observed among orthologs
332 (Supplementary file 6). The found lower correlation of protein expression between different
333 organs was more apparent in mouse and rat.
334
335 Among the orthologs expressed in all organs in all three species, 747 (12.3%) orthologs were
336 detected with a median bin expression value of more than 4, i.e., proteins that appear to have
337 conserved high expression in all organs and all tissues. Additionally, 13 (0.2%) orthologs
338 were found with a median bin expression value less than 2 in all organs, although, it is harder
339 to detect consistently proteins with low abundances across all organs. A full list of the binned
340 protein abundances of orthologs is available in Supplementary file 7. The illustration of all
341 binned protein abundances across the three species is shown in Supplementary file 8.
342

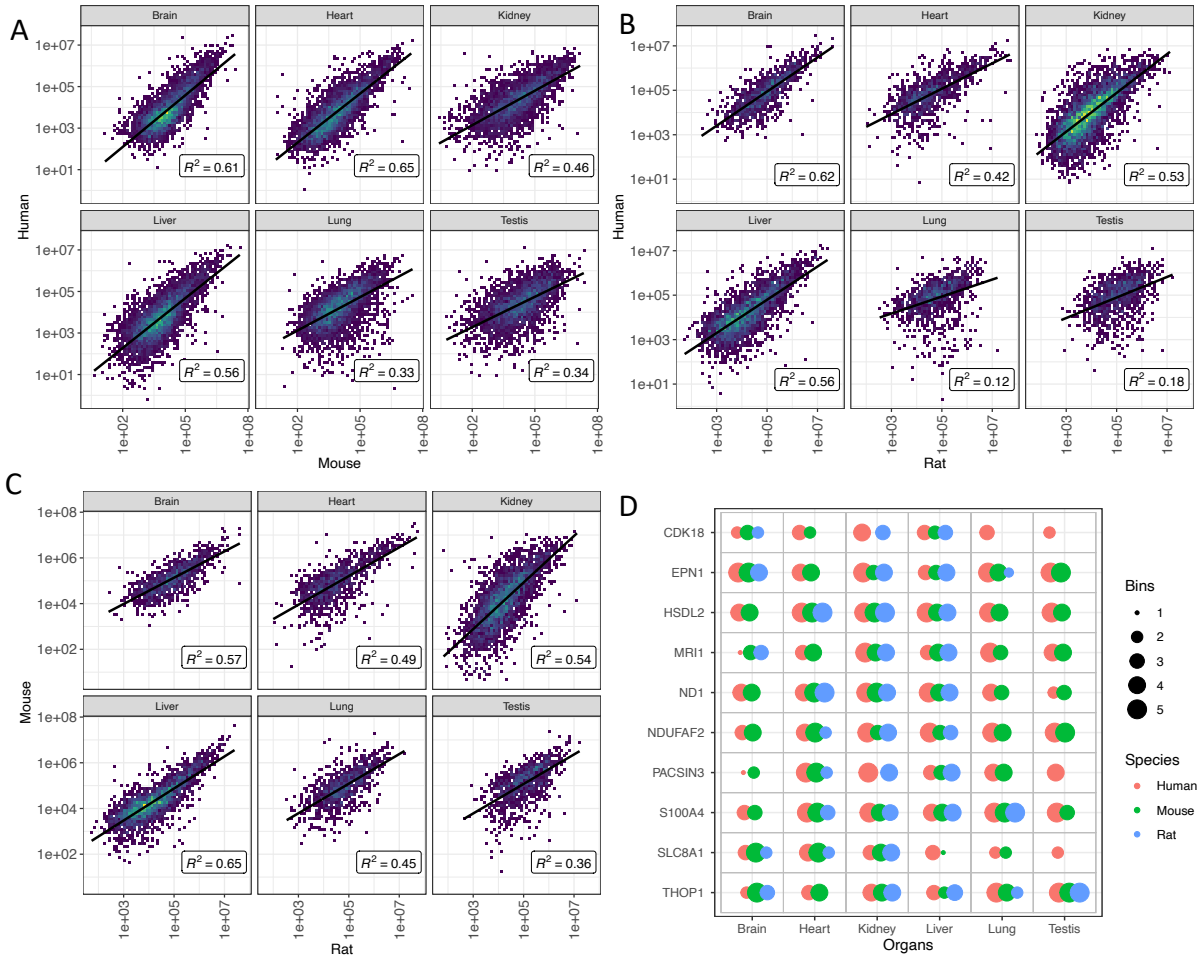
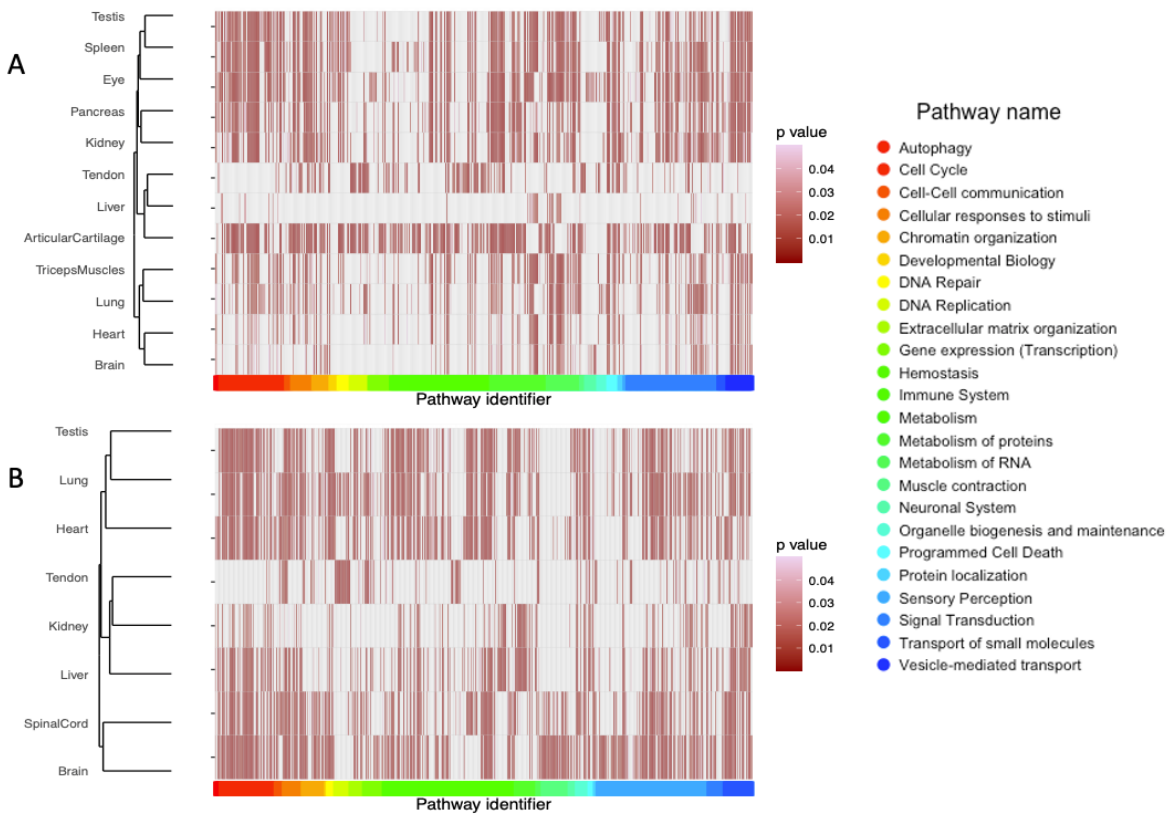


Figure 7. Comparison of protein abundances (in ppb) between one-to-one mapped orthologs of mouse, rat and human in various organs. (**A**) Pairwise correlation using normalised protein abundances of human and mouse orthologues. (**B**) Human and rat orthologs. (**C**) Mouse and rat orthologs. (**D**) As an example, the comparisons of binned protein expression of ten randomly sampled orthologs are shown. All data corresponding to all cases (as reported in panel D) are available in Supplementary file 7 and their corresponding illustration of binned values is available in Supplementary file 8. Orthologs in (D) are shown using their human Gene Symbol.

2.6. Pathway enrichment analysis

Based on the ortholog protein expression analysis described above, we mapped canonical proteins from mouse and rat to the corresponding ortholog human proteins, which were

356 subsequently subjected to pathway-enrichment analysis using Reactome (Fig. 8). After
357 filtering out the disease and statistically insignificant pathways, there were 2,990 pathways
358 found in all the organs of mouse and 2,162 pathways in all the organs of rat. In mouse
359 samples, the largest number of pathways (367) were found in articular cartilage, and the
360 lowest number of pathways was from liver (44). We also observed that Neuronal System-
361 related pathways are predominantly present in the brain and eye, which is consistent with
362 expectations. In rat samples, brain included the largest number of pathways (387), while the
363 lowest number of pathways was found in tendon, with 117.
364



365

366 **Figure 8.** Pathway analysis of the canonical proteins, showing the statistically significant
367 representative pathways (p-value < 0.05) in (A) mouse and (B) rat organs.
368
369

370 **3. Discussion**

371 We have previously reported two meta-analysis studies involving the reanalysis and
372 integration in Expression Atlas of public quantitative datasets coming from cell lines and
373 human tumour samples [14], and from human baseline tissues [15], respectively. In this
374 study, we reanalysed mouse and rat baseline proteomics datasets representing protein
375 expression across 34 healthy tissues and 14 organs. We have used the same methodology as
376 in the study involving baseline human tissues, which enabled a comparison of protein
377 expression levels across the three species. Our main overall aim was to provide a system-
378 wide baseline protein expression catalogue across various tissues and organs of mouse and rat
379 and to offer a reference for future related studies.

380

381 We analysed each dataset separately using the same software (MaxQuant) and the same
382 search protein sequence database. The disadvantage of this approach is that the False
383 Discovery Rate (FDR) statistical thresholds are applied at a dataset level and not to all
384 datasets together as a whole. However, as reported before [15], using a dataset per dataset
385 analysis approach is in our view the only sustainable manner to reanalyse and integrate
386 quantitative proteomics datasets, at least at present. As also reported in previous studies, one
387 of the major bottlenecks was the curation of dataset metadata, consisting of mapping files to
388 samples and biological conditions. Very recently, the MAGE-TAB-Proteomics format has
389 been developed and formalised to enable the reporting of the experimental design in
390 proteomics experience, including the relationship between samples and raw files, which is
391 recorded in the SDRF-Proteomics section of the file [41]. Submission of the SDRF-
392 Proteomics files to PRIDE is now supported. The more well-annotated datasets in the public
393 domain, the easier data re-use activities will become.

394

395 The generated baseline protein expression data can be used with different purposes such as
396 the generation of protein co-expression networks and/or the inference of protein complexes.
397 For the latter application, expression data can be alone or potentially refining predictions
398 obtained using different methods such as the recently developed AlphaFold-based protein
399 complexes predictions [42]. Mouse and rat are widely used species in the context of drug
400 discovery, the latter especially, to undertake regulatory pre-clinical safety studies. Therefore,
401 it is important to know quantitative protein expression distribution in these species in
402 different tissues [43] to assist in the selection of species for such studies and also for the
403 interpretation of the final results.
404
405 In addition to the analyses reported, it would also be possible to perform correlation studies
406 between gene and protein expression information. However, we did not find any relevant
407 public datasets in the context of this manuscript where the same samples were analysed by
408 both techniques, which is the optimal way to perform these studies. Future directions in
409 analogous studies will involve: (i) additional baseline protein expression studies of other
410 species, including other model organisms or other species of economic importance; (ii) the
411 inclusion of differential datasets (e.g. using TMT and/or iTRAQ); and (iii) include relevant
412 proteomics expression data coming from the reanalysis of Data Independent Acquisition
413 (DIA) datasets [44].
414
415 As mentioned above, we performed a comparative analysis of baseline protein expression
416 across human, mouse and rat. It was possible to perform this analysis for six common organs
417 (brain, heart, kidney, liver, lung and testis). Ortholog expression across species is useful to
418 infer protein function across experimentally studied proteins. This is particularly useful as
419 evolutionarily closely related species are likely to conserve protein function. We could not

find in the literature an analogous comparative study performed at the protein level. However, protein expression from closely related orthologs across tissues or organs has been compared at the transcriptomics level, providing a complete picture of gene expression. In this context, many studies have compared gene-expression in mouse, rat and human orthologues and found that orthologues had generally a highly correlated expression tissue distribution profile in baseline conditions [45-49]. Gene expression levels among orthologs were found to be highly similar in muscle and heart tissues, liver and nervous system and less similar in epithelial cells, reproductive systems, bone and endocrine organs [47]. Studies have also shown that variability of gene expression between homologous tissues/organs in closely related species can be much lower than the variability between unrelated tissues within the same organism [45, 46], in agreement with the results reported here at the protein level.

In conclusion we here present a meta-analysis study of public mouse and rat baseline proteomics datasets from PRIDE. We demonstrate its feasibility, perform a comparative analysis across the three species and show the main current challenges. Finally, the data is made available *via* Expression Atlas. Whereas there are several analogous studies performed at the gene expression level for mouse and rat tissues, this is the first of this kind at the protein expression level.

4. Materials and Methods

4.1. Datasets

As of May 2021, there were 2,060 mouse (*Mus musculus*) and 339 rat (*Rattus norvegicus*) MS proteomics datasets publicly available in the PRIDE database (https://www.ebi.ac.uk/pride/). Datasets were manually selected based on the selection criteria described previously [15]. Briefly, we selected datasets where baseline expression experiments were performed on (i) label-free samples from tissues not enriched for post-translational modifications; (ii) Thermo Fisher Scientific instruments such as LTQ Orbitrap, LTQ Orbitrap Elite, LTQ Orbitrap Velos, LTQ Orbitrap XL ETD, LTQ-Orbitrap XL ETD, Orbitrap Fusion and Q-Exactive, since they represent a large proportion of datasets in PRIDE and to avoid heterogeneity introduced by data from other vendor instruments; (iii) had suitable sample metadata available in the original publication or it was possible to obtain it by contacting the authors; and (iv) our previous experience in the team of some datasets deposited in PRIDE, which were discarded because they were not considered to be useful (data not shown). Overall, 14 mouse and 9 rat datasets were selected from all mouse and rat datasets for further analysis. Table 1 lists the selected datasets. The 23 datasets contained a total of 211 samples from 34 different tissues across 14 organs (meaning groups of related tissues, more details below), comprising 9 different mouse and 3 rat strains, respectively. The sample and experimental metadata were manually curated using the information provided in the respective publications or by contacting the original authors/submitters. Annotare [50] was used for annotating the metadata and stored using the Investigation Description Format (IDF) and Sample-Data Relationship Format (SDRF) file formats [41], which are required for integration in Expression Atlas. An overview of the experimental design including experimental factors, protocols, publication information and contact

464 information are present in the IDF file, and the SDRF includes sample metadata describing
465 the relationship between the various sample characteristics and the data files contained in the
466 dataset.
467
468 **4.2. Proteomics raw data processing**
469 All datasets were analysed with MaxQuant (version 1.6.3.4) [51, 52] on a Linux high-
470 performance computing cluster for peptide/protein identification and protein quantification.
471 Input parameters for each dataset, such as MS1 and MS2 tolerances, digesting enzymes, fixed
472 and variable modifications, were set as described in their respective publications, with two
473 missed cleavage sites. The FDR at the PSM (peptide spectrum match) and protein levels were
474 set to 1%. The MaxQuant parameters were otherwise set to default values: the maximum
475 number of modifications per peptide was 5, the minimum peptide length was 7, the maximum
476 peptide mass was set to 4,600 Da, and for the matches between runs the minimum match time
477 window was set to 0.7 seconds and the minimum retention time alignment window was set to
478 20 seconds. The MaxQuant parameter files are available for downloading from Expression
479 Atlas. The *Mus musculus* UniProt Reference proteome release-2021_04 (including isoforms,
480 63,656 sequences) and *Rattus norvegicus* UniProt Reference proteome release-2021_04
481 (including isoforms, 31,562 sequences) were used as the target sequence databases for mouse
482 and rat datasets, respectively. The built-in contaminant database within MaxQuant was used
483 and a decoy database was generated by MaxQuant by reversing the input database sequences
484 after the respective enzymatic digestion. The datasets were run separately in multi-threaded
485 mode.
486
487 **4.3. Post-processing**

488 The post-processing of results from MaxQuant is explained in detail in [15]. In brief, the
489 protein groups labelled as potential contaminants, decoys and those with less than 2 PSMs
490 were removed. Protein intensities in each sample were normalised by scaling the iBAQ
491 intensity values to the total amount of signal in each MS run and converted to parts per
492 billion (ppb).

493 $$ppb_iBAQ_i = \left({iBAQ_i} \Big/ {\sum_{i=1}^{n} iBAQ_i} \right) x\ 1{,}000{,}000{,}000$$

494 The 'majority protein identifiers' within each protein group were mapped to their Ensembl
495 gene identifiers/annotations using the Bioconductor package 'mygene'. For downstream
496 analysis only protein groups whose isoforms mapped to a single unique Ensembl gene ID
497 were considered. Protein groups that mapped to more than one Ensembl gene ID are provided
498 in Supplementary file 1. The protein intensity values from different protein groups with the
499 same Ensembl gene ID were aggregated as median values. The parent genes to which the
500 different protein groups were mapped to are equivalent to 'canonical proteins' in UniProt
501 (https://www.uniprot.org/help/canonical_and_isoforms) and therefore the term protein
502 abundance is used to describe the protein abundance of the canonical protein throughout the
503 manuscript.
504

505 **4.4. Integration into Expression Atlas**

506 The calculated canonical protein abundances (mapped to genes), together with the validated
507 SDRF files, summary files detailing the quality of post-processing and the input MaxQuant
508 parameter files (mqpar.xml) were integrated into Expression Atlas
509 (https://www.ebi.ac.uk/gxa/home) as proteomics baseline experiments (E-PROT identifiers
510 are available in Table 1).
511

512 **4.5. Protein abundance comparison across datasets**

513 To compare protein abundances, the normalised protein abundances (in ppb) from each group
514 of tissues in a dataset were converted into ranked bins. In this study, 'tissue' is defined as a
515 distinct functional or structural region within an 'organ'. For example, hippocampus,
516 cerebellum and cortex are defined as 'tissues' that are part of the brain (organ) and similarly
517 sinus node, left atria, left ventricle, right atria, right ventricle are defined as 'tissues' in heart
518 (organ). Protein abundances were transformed into bins by first grouping MS runs from each
519 tissue within a dataset as a batch. The normalised protein abundances (ppb) for each MS run
520 within a batch were sorted from lowest to highest abundance and ranked into 5 bins. Proteins
521 whose ppb abundances are ranked in the lowest bin (bin 1) represent lowest abundance and
522 correspondingly proteins within bin 5 are of highest abundance in their respective tissue.
523 When merging tissues into organs, median bin values were used.

524 Proteins that were detected in at least 50% of the samples were selected for PCA (Principal
525 Component Analysis) and was performed using R (The R Stats package) [53] using binned
526 abundance values. For generating heatmaps, a Pearson correlation coefficient (rp) for all
527 samples was calculated on pairwise complete observations of bin transformed values.
528 Missing values were marked as NA (not available). For each organ a median rp was
529 calculated from all pairwise rp values of their respective samples. Samples were
530 hierarchically clustered on columns and rows using Euclidean distances.

531 **4.6. Organ-specific expression profile analysis**

532 For comparison across organs, the tissues were aggregated into organs and their median bin
533 values were considered. As described previously [15] the classification scheme done by
534 Uhlén *et al.* [17] was modified to classify the proteins into one of the three categories: (1)
535 "Organ-enriched": present in one unique organ with bin values 2-fold higher than the mean
536 bin value across all organs; (2) "Group enriched": present in at least 7 organs in mouse or in

537 at least 4 organs in rat, with bin values 2-fold higher than the mean bin value across all
538 organs; and (3) “Mixed”: the remaining canonical proteins that are not part of the above two
539 categories.
540
541 Enriched gene ontology (GO) terms analysis was carried out through over-representation test
542 described previously [15], it was combined with “Organ-enriched” and “Group enriched”
543 mapped gene lists for each organ. In addition, Reactome [54] pathway analysis was
544 performed using mapped gene lists and running pathway-topology and over-representation
545 analysis, as reported previously [15].

546 **4.7. Comparison of protein expression across species**

547 The g:Orth Orthology search function in the g:Profiler suite of programs [55] was used for
548 translating gene identifiers between organisms. Since a custom list of gene identifiers could
549 not be used as the background search set, the mouse and rat genes were first mapped against
550 the background Ensembl database. The resulting list of mouse and rat genes mapped to
551 human orthologs were then filtered so that they only included parent gene identifiers of the
552 protein groups from mouse and rat organs identified in this study and the parent genes of
553 human organs described in our previous study (Supplementary file 2 in [15]), respectively.
554
555 The orthologs were grouped into various categories denoting the resulting mapping between
556 identifiers: “one-to-one”, “one-to-many”, “many-to-one”, “many-to-many”, and “no
557 mappings” between gene identifiers. Only “one-to-one” mapped ortholog identifiers were
558 used to compare protein intensities between mouse, rat and human organs. The normalised
559 ppb protein abundances of the one-to-one mapped orthologues in 6 organs (brain, heart,
560 kidney, liver, lung and testis), that were studied across all three organisms were used to assess

561 the pairwise correlation of protein abundances. The linear regression was calculated using the
562 linear fit 'lm' method in R.
563

564 **Data availability**

565 Expression Atlas E-PROT identifiers and PRIDE original dataset identifiers are included in
566 Table 1.

567 **Acknowledgements**

568 First of all, we would like to thank all data submitters who made their datasets available via
569 PRIDE and ProteomeXchange. We would also like to thank Andrew Leach and the rest of the
570 team involved in the Open Targets target safety project, for helpful discussions. This work
571 has been funded by Open Targets (project OTAR-02-068), Wellcome Trust [grant number
572 208391/Z/17/Z], BBSRC [BB/T019670/1 and BB/T019557/1] and EMBL core funding.

573 **Authors' contributions**

574 SW, DGS, AP, DJK selected and curated the datasets. SW, AP, DGS performed analyses. AC
575 and ARJ helped in the interpretation of results and designed approach for data normalisation.
576 NG, SF, PM, and IP helped in the integration of the results into Expression Atlas. SW, AP,
577 DGS, JAV wrote the manuscript. JAV, ARJ and IP obtained the funding for performing the
578 study. All authors have read and approved the manuscript.
579

580 **Supporting information**

581 **Supplementary file 1:** Protein groups from all datasets that are mapped to more than one
582 Ensembl Gene ID.

583 **Supplementary file 2:** Median protein abundances (in ppb) for each protein group across
584 various tissue samples in each organ.

585 **Supplementary file 3:** Median binned protein abundances across various tissue samples in
586 each organ of mouse and rat.

587 **Supplementary file 4:** Organ distribution of canonical proteins in mouse and rat.

588 **Supplementary file 5:** Gene Ontology enrichment analysis of 'organ-enriched' and 'group-
589 enriched' proteins.

590 **Supplementary file 6:** Correlation of protein expression between organs within each
591 organism.

592 **Supplementary file 7:** Binned protein abundances of one-to-one mapped orthologs across all
593 organs studied.

594 **Supplementary file 8:** Figure illustrating binned protein abundances of all one-to-one
595 mapped orthologs across six common organs in mouse, rat and human.

596

597 **Abbreviations**

598 DDA: Data Dependent Acquisition
599 DIA: Data Independent Acquisition
600 FOT: Fraction of Total
601 GO: Gene Ontology
602 iBAQ: intensity-based absolute quantification
603 iTRAQ: Isobaric tag for relative and absolute quantitation
604 IDF: Investigation Description Format

MS: Mass Spectrometry

ppb: Parts per billion

PCA: Principal Component Analysis

rp: Pearson correlation coefficient

SDRF: Sample and Data Relationship Format

TMT: Tandem Mass Tagging

611 **References**
612
613 1. Aebersold R, Mann M. Mass-spectrometric exploration of proteome structure and
614 function. Nature. 2016;537(7620):347-55. Epub 2016/09/16. doi: 10.1038/nature19949.
615 PubMed PMID: 27629641.
616
617 2. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ,
618 et al. The PRIDE database and related tools and resources in 2019: improving support for
619 quantification data. Nucleic Acids Res. 2019;47(D1):D442-D50. Epub 2018/11/06. doi:
620 10.1093/nar/gky1106. PubMed PMID: 30395289; PubMed Central PMCID:
621 PMCPMC6323896.
622
623 3. Vizcaino JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Rios D, et al.
624 ProteomeXchange provides globally coordinated proteomics data submission and
625 dissemination. Nat Biotechnol. 2014;32(3):223-6. Epub 2014/04/15. doi:
626 10.1038/nbt.2839. PubMed PMID: 24727771; PubMed Central PMCID:
627 PMCPMC3986813.
628
629 4. Martens L, Vizcaino JA. A Golden Age for Working with Public Proteomics Data.
630 Trends Biochem Sci. 2017;42(5):333-41. Epub 2017/01/26. doi:
631 10.1016/j.tibs.2017.01.001. PubMed PMID: 28118949; PubMed Central PMCID:
632 PMCPMC5414595.
633
634 5. Romanov N, Kuhn M, Aebersold R, Ori A, Beck M, Bork P. Disentangling Genetic
635 and Environmental Effects on the Proteotypes of Individuals. Cell. 2019;177(5):1308-18
636 e10. Epub 2019/04/30. doi: 10.1016/j.cell.2019.03.015. PubMed PMID: 31031010;
637 PubMed Central PMCID: PMCPMC6988111.
638
639 6. Skinnider MA, Foster LJ. Meta-analysis defines principles for the design and analysis
640 of co-fractionation mass spectrometry experiments. Nat Methods. 2021;18(7):806-15.
641 Epub 2021/07/03. doi: 10.1038/s41592-021-01194-4. PubMed PMID: 34211188.
642
643 7. Ochoa D, Jarnuczak AF, Vieitez C, Gehre M, Soucheray M, Mateus A, et al. The
644 functional landscape of the human phosphoproteome. Nat Biotechnol. 2020;38(3):365-73.
645 Epub 2019/12/11. doi: 10.1038/s41587-019-0344-3. PubMed PMID: 31819260; PubMed
646 Central PMCID: PMCPMC7100915.
647
648 8. Vaudel M, Verheggen K, Csordas A, Raeder H, Berven FS, Martens L, et al.
649 Exploring the potential of public proteomics data. Proteomics. 2016;16(2):214-25. Epub
650 2015/10/10. doi: 10.1002/pmic.201500295. PubMed PMID: 26449181; PubMed Central
651 PMCID: PMCPMC4738454.
652
653 9. Kumar D, Yadav AK, Jia X, Mulvenna J, Dash D. Integrated Transcriptomic-
654 Proteomic Analysis Using a Proteogenomic Workflow Refines Rat Genome Annotation.
655 Mol Cell Proteomics. 2016;15(1):329-39. Epub 2015/11/13. doi:
656 10.1074/mcp.M114.047126. PubMed PMID: 26560066; PubMed Central PMCID:
657 PMCPMC4762527.
658
659 10. Brunet MA, Brunelle M, Lucier JF, Delcourt V, Levesque M, Grenier F, et al.
660 OpenProt: a more comprehensive guide to explore eukaryotic coding potential and

661 proteomes. Nucleic Acids Res. 2019;47(D1):D403-D10. Epub 2018/10/10. doi:
662 10.1093/nar/gky936. PubMed PMID: 30299502; PubMed Central PMCID:
663 PMCPMC6323990.
664
665 11. Levitsky LI, Kliuchnikova AA, Kuznetsova KG, Karpov DS, Ivanov MV, Pyatnitskiy
666 MA, et al. Adenosine-to-Inosine RNA Editing in Mouse and Human Brain Proteomes.
667 Proteomics. 2019;19(23):e1900195. Epub 2019/10/03. doi: 10.1002/pmic.201900195.
668 PubMed PMID: 31576663.
669
670 12. Li H, Zhou R, Xu S, Chen X, Hong Y, Lu Q, et al. Improving Gene Annotation of the
671 Peanut Genome by Integrated Proteogenomics Workflow. J Proteome Res.
672 2020;19(6):2226-35. Epub 2020/05/06. doi: 10.1021/acs.jproteome.9b00723. PubMed
673 PMID: 32367721.
674
675 13. Moreno P, Fexova S, George N, Manning JR, Miao Z, Mohammed S, et al.
676 Expression Atlas update: gene and protein expression in multiple species. Nucleic Acids
677 Res. 2021. Epub 2021/12/02. doi: 10.1093/nar/gkab1030. PubMed PMID: 34850121.
678
679 14. Jarnuczak AF, Najgebauer H, Barzine M, Kundu DJ, Ghavidel F, Perez-Riverol Y, et
680 al. An integrated landscape of protein expression in human cancer. Sci Data.
681 2021;8(1):115. Epub 2021/04/25. doi: 10.1038/s41597-021-00890-2. PubMed PMID:
682 33893311; PubMed Central PMCID: PMCPMC8065022.
683
684 15. Prakash A, García-Seisdedos D, Wang S, Kundu DJ, Collins A, George N, et al. An
685 integrated view of baseline protein expression in human tissues. bioRxiv.
686 2021:2021.09.10.459811. doi: 10.1101/2021.09.10.459811.
687
688 16. Samaras P, Schmidt T, Frejno M, Gessulat S, Reinecke M, Jarzab A, et al.
689 ProteomicsDB: a multi-omics and multi-organism resource for life science research.
690 Nucleic Acids Res. 2020;48(D1):D1153-D63. Epub 2019/10/31. doi: 10.1093/nar/gkz974.
691 PubMed PMID: 31665479; PubMed Central PMCID: PMCPMC7145565.
692
693 17. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al.
694 Proteomics. Tissue-based map of the human proteome. Science. 2015;347(6220):1260419.
695 Epub 2015/01/24. doi: 10.1126/science.1260419. PubMed PMID: 25613900.
696
697 18. Azimifar SB, Nagaraj N, Cox J, Mann M. Cell-type-resolved quantitative proteomics
698 of murine liver. Cell Metab. 2014;20(6):1076-87. Epub 2014/12/04. doi:
699 10.1016/j.cmet.2014.11.002. PubMed PMID: 25470552.
700
701 19. Deshmukh AS, Murgia M, Nagaraj N, Treebak JT, Cox J, Mann M. Deep proteomics
702 of mouse skeletal muscle enables quantitation of protein isoforms, metabolic pathways,
703 and transcription factors. Mol Cell Proteomics. 2015;14(4):841-53. Epub 2015/01/27. doi:
704 10.1074/mcp.M114.044222. PubMed PMID: 25616865; PubMed Central PMCID:
705 PMCPMC4390264.
706
707 20. Meierhofer D, Halbach M, Sen NE, Gispert S, Auburger G. Ataxin-2 (Atxn2)-Knock-
708 Out Mice Show Branched Chain Amino Acids and Fatty Acids Pathway Alterations. Mol
709 Cell Proteomics. 2016;15(5):1728-39. Epub 2016/02/07. doi: 10.1074/mcp.M115.056770.
710 PubMed PMID: 26850065; PubMed Central PMCID: PMCPMC4858951.

711
712 21. Sarver DC, Kharaz YA, Sugg KB, Gumucio JP, Comerford E, Mendias CL. Sex
713 differences in tendon structure and function. J Orthop Res. 2017;35(10):2117-26. Epub
714 2017/01/11. doi: 10.1002/jor.23516. PubMed PMID: 28071813; PubMed Central PMCID:
715 PMCPMC5503813.
716
717 22. Duda P, Wojcicka O, Wisniewski JR, Rakus D. Global quantitative TPA-based
718 proteomics of mouse brain structures reveals significant alterations in expression of
719 proteins involved in neuronal plasticity during aging. Aging (Albany NY).
720 2018;10(7):1682-97. Epub 2018/07/22. doi: 10.18632/aging.101501. PubMed PMID:
721 30026405; PubMed Central PMCID: PMCPMC6075443.
722
723 23. Harman JC, Guidry JJ, Gidday JM. Comprehensive characterization of the adult ND4
724 Swiss Webster mouse retina: Using discovery-based mass spectrometry to decipher the
725 total proteome and phosphoproteome. Mol Vis. 2018;24:875-89. Epub 2019/02/05.
726 PubMed PMID: 30713425; PubMed Central PMCID: PMCPMC6334985.
727
728 24. Angelidis I, Simon LM, Fernandez IE, Strunz M, Mayr CH, Greiffo FR, et al. An
729 atlas of the aging lung mapped by single cell transcriptomics and deep tissue proteomics.
730 Nat Commun. 2019;10(1):963. Epub 2019/03/01. doi: 10.1038/s41467-019-08831-9.
731 PubMed PMID: 30814501; PubMed Central PMCID: PMCPMC6393476.
732
733 25. Zhao Y, Wilmarth PA, Cheng C, Limi S, Fowler VM, Zheng D, et al. Proteome-
734 transcriptome analysis and proteome remodeling in mouse lens epithelium and fibers. Exp
735 Eye Res. 2019;179:32-46. Epub 2018/10/26. doi: 10.1016/j.exer.2018.10.011. PubMed
736 PMID: 30359574; PubMed Central PMCID: PMCPMC6360118.
737
738 26. Huttlin EL, Jedrychowski MP, Elias JE, Goswami T, Rad R, Beausoleil SA, et al. A
739 tissue-specific atlas of mouse protein phosphorylation and expression. Cell.
740 2010;143(7):1174-89. Epub 2010/12/25. doi: 10.1016/j.cell.2010.12.001. PubMed PMID:
741 21183079.
742
743 27. Linscheid N, Santos A, Poulsen PC, Mills RW, Calloe K, Leurs U, et al. Quantitative
744 proteome comparison of human hearts with those of model organisms. PLoS Biol.
745 2021;19(4):e3001144. Epub 2021/04/20. doi: 10.1371/journal.pbio.3001144. PubMed
746 PMID: 33872299; PubMed Central PMCID: PMCPMC8084454.
747
748 28. Dudek M, Angelucci C, Pathiranage D, Wang P, Mallikarjun V, Lawless C, et al.
749 Circadian time series proteomics reveals daily dynamics in cartilage physiology.
750 Osteoarthritis Cartilage. 2021;29(5):739-49. Epub 2021/02/22. doi:
751 10.1016/j.joca.2021.02.008. PubMed PMID: 33610821; PubMed Central PMCID:
752 PMCPMC8113022.
753
754 29. Schroeder S, Hofer SJ, Zimmermann A, Pechlaner R, Dammbrueck C, Pendl T, et al.
755 Dietary spermidine improves cognitive function. Cell Rep. 2021;35(2):108985. Epub
756 2021/04/15. doi: 10.1016/j.celrep.2021.108985. PubMed PMID: 33852843.
757
758 30. Bundy JL, Vied C, Nowakowski RS. Sex differences in the molecular signature of the
759 developing mouse hippocampus. BMC Genomics. 2017;18(1):237. Epub 2017/03/18. doi:

10.1186/s12864-017-3608-7. PubMed PMID: 28302071; PubMed Central PMCID: PMCPMC5356301.

31. Linscheid N, Logantha S, Poulsen PC, Zhang S, Schrolkamp M, Egerod KL, et al. Quantitative proteomics and single-nucleus transcriptomics of the sinus node elucidates the foundation of cardiac pacemaking. Nat Commun. 2019;10(1):2889. Epub 2019/06/30. doi: 10.1038/s41467-019-10709-9. PubMed PMID: 31253831; PubMed Central PMCID: PMCPMC6599035.

32. Alugubelly N, Mohammed AN, Edelmann MJ, Nanduri B, Sayed M, Park JW, et al. Adolescent rat social play: Amygdalar proteomic and transcriptomic data. Data Brief. 2019;27:104589. Epub 2019/11/02. doi: 10.1016/j.dib.2019.104589. PubMed PMID: 31673590; PubMed Central PMCID: PMCPMC6817652.

33. Roffia V, De Palma A, Lonati C, Di Silvestre D, Rossi R, Mantero M, et al. Proteome Investigation of Rat Lungs subjected to Ex Vivo Perfusion (EVLP). Molecules. 2018;23(12). Epub 2018/11/24. doi: 10.3390/molecules23123061. PubMed PMID: 30467300; PubMed Central PMCID: PMCPMC6321151.

34. Bernier M, Harney D, Koay YC, Diaz A, Singh A, Wahl D, et al. Elucidating the mechanisms by which disulfiram protects against obesity and metabolic syndrome. NPJ Aging Mech Dis. 2020;6:8. Epub 2020/07/28. doi: 10.1038/s41514-020-0046-6. PubMed PMID: 32714562; PubMed Central PMCID: PMCPMC7374720.

35. Huang Q, Luo L, Alamdar A, Zhang J, Liu L, Tian M, et al. Integrated proteomics and metabolomics analysis of rat testis: Mechanism of arsenic-induced male reproductive toxicity. Sci Rep. 2016;6:32518. Epub 2016/09/03. doi: 10.1038/srep32518. PubMed PMID: 27585557; PubMed Central PMCID: PMCPMC5009432.

36. Kaushik G, Spenlehauer A, Sessions AO, Trujillo AS, Fuhrmann A, Fu Z, et al. Vinculin network-mediated cytoskeletal remodeling regulates contractile function in the aging heart. Sci Transl Med. 2015;7(292):292ra99. Epub 2015/06/19. doi: 10.1126/scitranslmed.aaa5843. PubMed PMID: 26084806; PubMed Central PMCID: PMCPMC4507505.

37. Vileigas DF, Harman VM, Freire PP, Marciano CLC, Sant'Ana PG, de Souza SLB, et al. Landscape of heart proteome changes in a diet-induced obesity model. Sci Rep. 2019;9(1):18050. Epub 2019/12/04. doi: 10.1038/s41598-019-54522-2. PubMed PMID: 31792287; PubMed Central PMCID: PMCPMC6888820.

38. Limbutara K, Chou CL, Knepper MA. Quantitative Proteomics of All 14 Renal Tubule Segments in Rat. J Am Soc Nephrol. 2020;31(6):1255-66. Epub 2020/05/03. doi: 10.1681/ASN.2020010071. PubMed PMID: 32358040; PubMed Central PMCID: PMCPMC7269347.

39. Devaux S, Cizkova D, Quanico J, Franck J, Nataf S, Pays L, et al. Proteomic Analysis of the Spatio-temporal Based Molecular Kinetics of Acute Spinal Cord Injury Identifies a Time- and Segment-specific Window for Effective Tissue Repair. Mol Cell Proteomics. 2016;15(8):2641-70. Epub 2016/06/03. doi: 10.1074/mcp.M115.057794. PubMed PMID: 27250205; PubMed Central PMCID: PMCPMC4974342.

40. Choi H, Simpson D, Wang D, Prescott M, Pitsillides AA, Dudhia J, et al. Heterogeneity of proteome dynamics between connective tissue phases of adult tendon. Elife. 2020;9. Epub 2020/05/13. doi: 10.7554/eLife.55262. PubMed PMID: 32393437; PubMed Central PMCID: PMCPMC7217697.

41. Dai C, Fullgrabe A, Pfeuffer J, Solovyeva EM, Deng J, Moreno P, et al. A proteomics sample metadata representation for multiomics integration and big data analysis. Nat Commun. 2021;12(1):5854. Epub 2021/10/08. doi: 10.1038/s41467-021-26111-3. PubMed PMID: 34615866; PubMed Central PMCID: PMCPMC8494749.

42. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer. bioRxiv. 2021:2021.10.04.463034. doi: 10.1101/2021.10.04.463034.

43. Bregman CL, Adler RR, Morton DG, Regan KS, Yano BL, Society of Toxicologic P. Recommended tissue list for histopathologic examination in repeat-dose toxicity and carcinogenicity studies: a proposal of the Society of Toxicologic Pathology (STP). Toxicol Pathol. 2003;31(2):252-3. Epub 2003/04/17. doi: 10.1080/01926230390183751. PubMed PMID: 12696587.

44. Walzer M, García-Seisdedos D, Prakash A, Brack P, Crowther P, Graham RL, et al. Implementing the re-use of public DIA proteomics datasets: from the PRIDE database to Expression Atlas. bioRxiv. 2021:2021.06.08.447493. doi: 10.1101/2021.06.08.447493.

45. Sollner JF, Leparc G, Hildebrandt T, Klein H, Thomas L, Stupka E, et al. An RNA-Seq atlas of gene expression in mouse and rat normal tissues. Sci Data. 2017;4:170185. Epub 2017/12/13. doi: 10.1038/sdata.2017.185. PubMed PMID: 29231921; PubMed Central PMCID: PMCPMC5726313.

46. Sudmant PH, Alexis MS, Burge CB. Meta-analysis of RNA-seq expression data across species, tissues and studies. Genome Biol. 2015;16:287. Epub 2015/12/24. doi: 10.1186/s13059-015-0853-4. PubMed PMID: 26694591; PubMed Central PMCID: PMCPMC4699362.

47. Zheng-Bradley X, Rung J, Parkinson H, Brazma A. Large scale comparison of global gene expression patterns in human and mouse. Genome Biol. 2010;11(12):R124. Epub 2010/12/25. doi: 10.1186/gb-2010-11-12-r124. PubMed PMID: 21182765; PubMed Central PMCID: PMCPMC3046484.

48. Liao BY, Zhang J. Evolutionary conservation of expression profiles between human and mouse orthologous genes. Mol Biol Evol. 2006;23(3):530-40. Epub 2005/11/11. doi: 10.1093/molbev/msj054. PubMed PMID: 16280543.

49. Prasad A, Kumar SS, Dessimoz C, Bleuler S, Laule O, Hruz T, et al. Global regulatory architecture of human, mouse and rat tissue transcriptomes. BMC Genomics. 2013;14:716. Epub 2013/10/22. doi: 10.1186/1471-2164-14-716. PubMed PMID: 24138449; PubMed Central PMCID: PMCPMC4008137.

50. Athar A, Fullgrabe A, George N, Iqbal H, Huerta L, Ali A, et al. ArrayExpress update - from bulk to single-cell expression data. Nucleic Acids Res. 2019;47(D1):D711-D5. Epub 2018/10/26. doi: 10.1093/nar/gky964. PubMed PMID: 30357387; PubMed Central PMCID: PMCPMC6323929.

51. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008;26(12):1367-72. Epub 2008/11/26. doi: 10.1038/nbt.1511. PubMed PMID: 19029910.

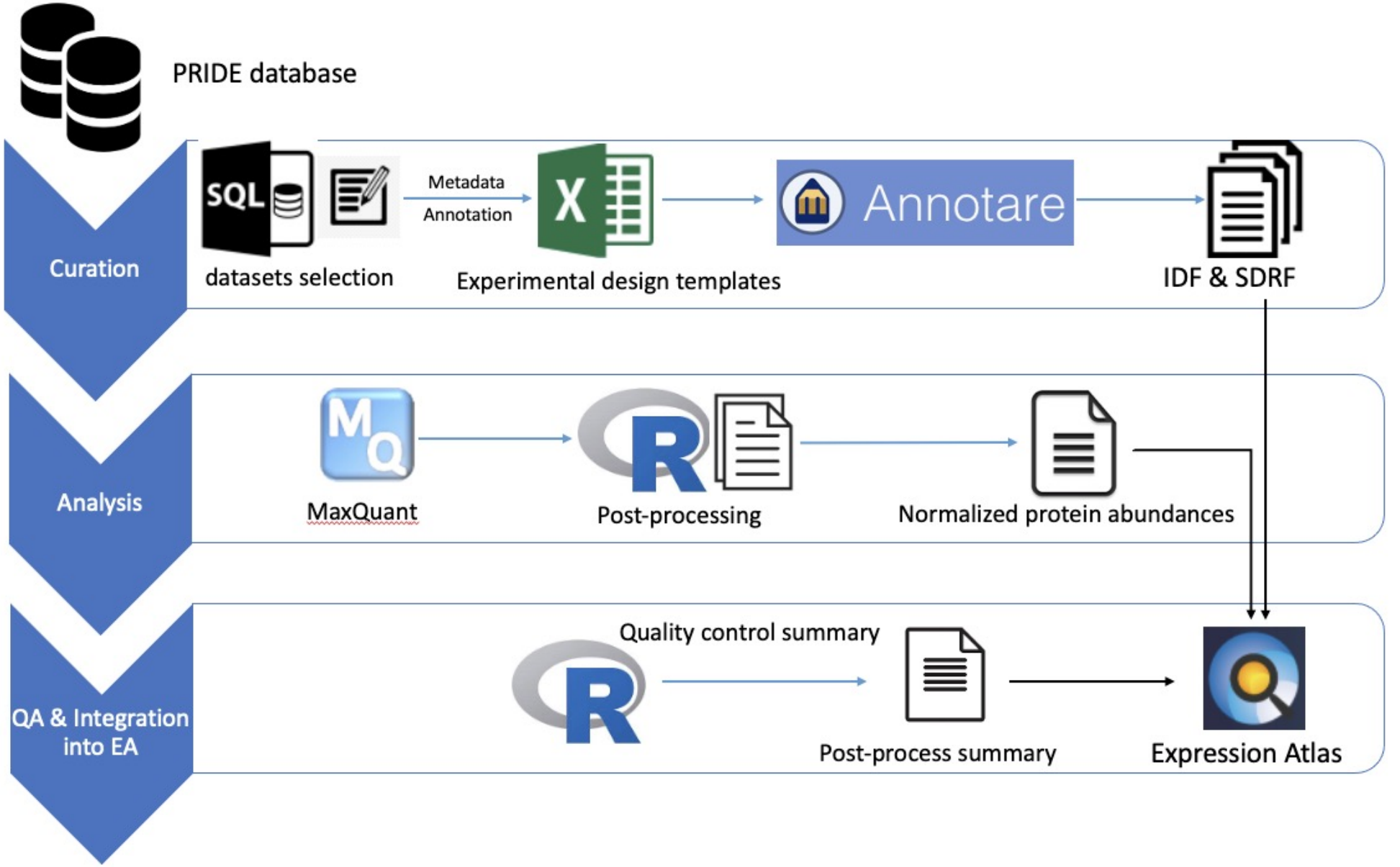
52. Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat Protoc. 2016;11(12):2301-19. Epub 2016/11/04. doi: 10.1038/nprot.2016.136. PubMed PMID: 27809316.

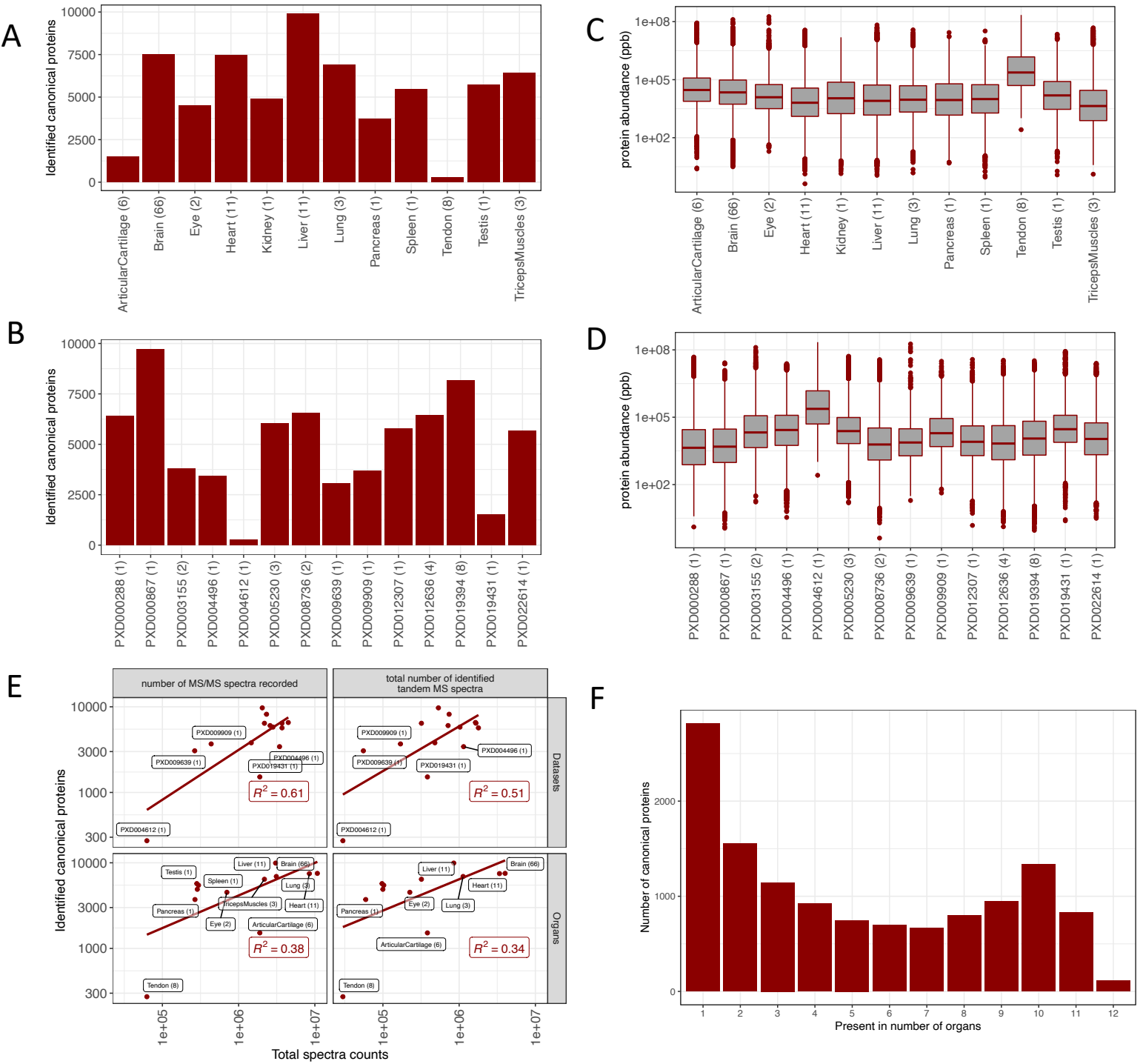
53. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://wwwR-projectorg/. 2013.

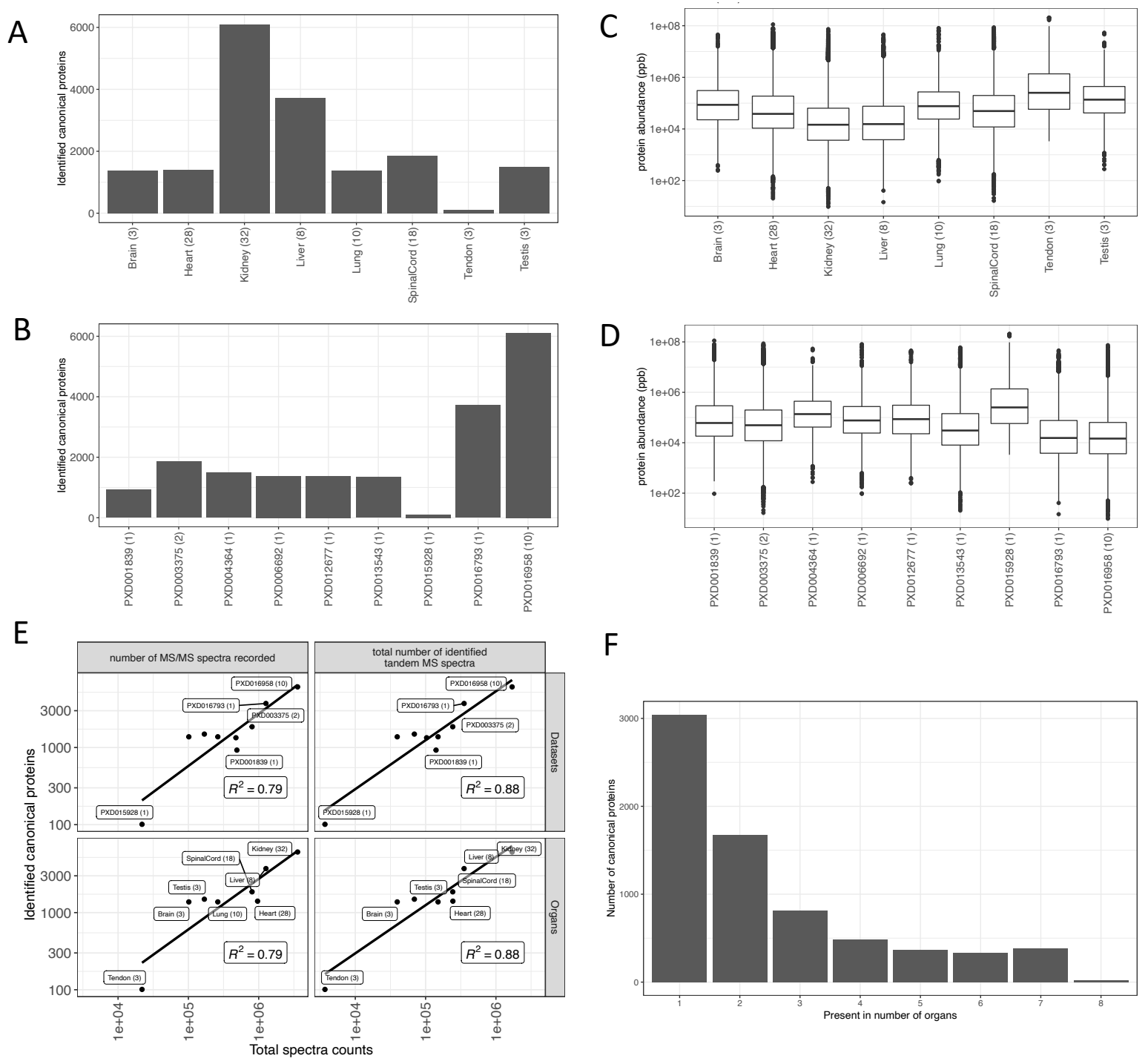
54. Wu G, Haw R. Functional Interaction Network Construction and Analysis for Disease Discovery. Methods Mol Biol. 2017;1558:235-53. Epub 2017/02/06. doi: 10.1007/978-1-4939-6783-4_11. PubMed PMID: 28150241.

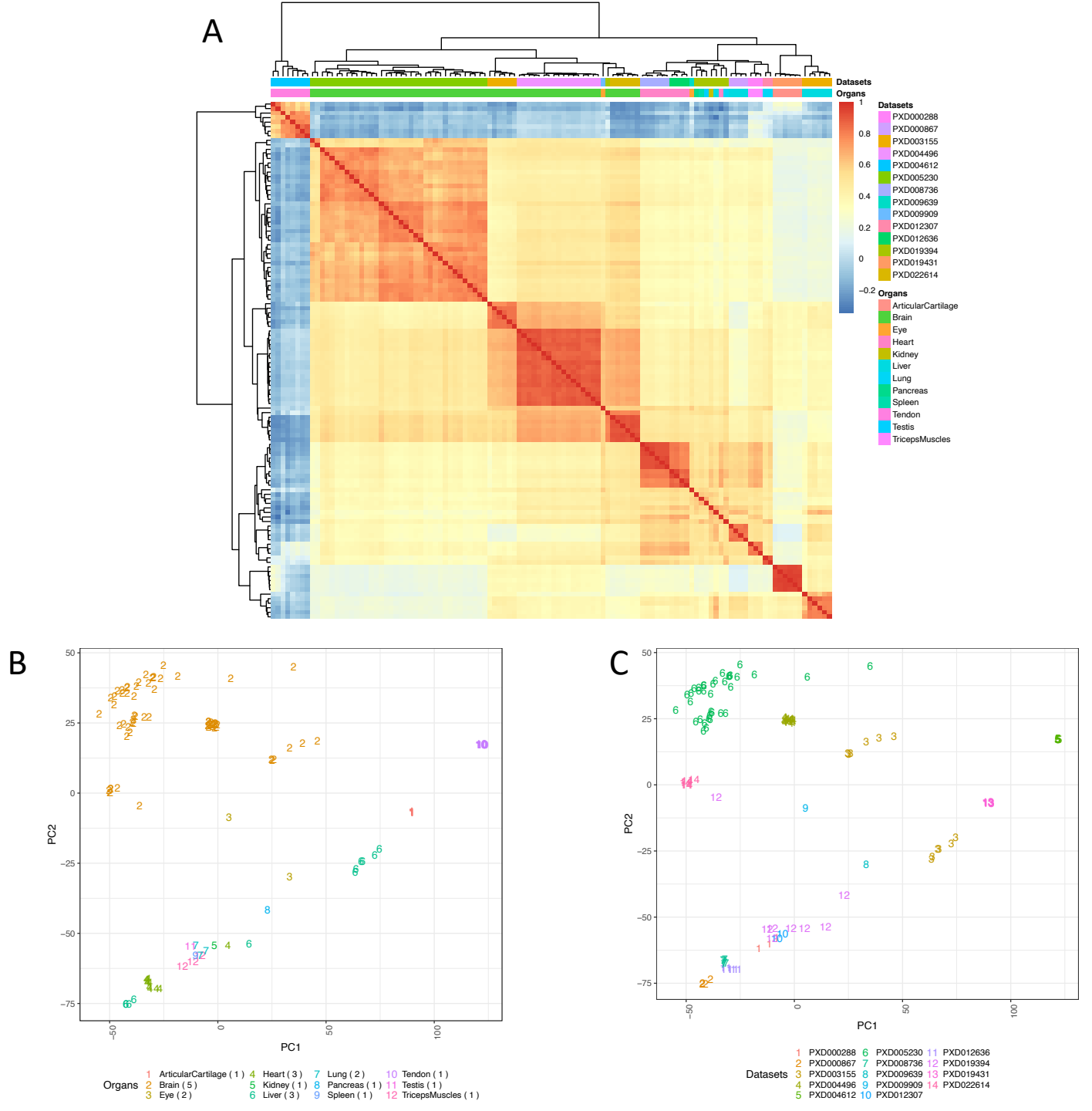
55. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. 2019;47(W1):W191-W8. Epub 2019/05/09. doi: 10.1093/nar/gkz369. PubMed PMID: 31066453; PubMed Central PMCID: PMCPMC6602461.

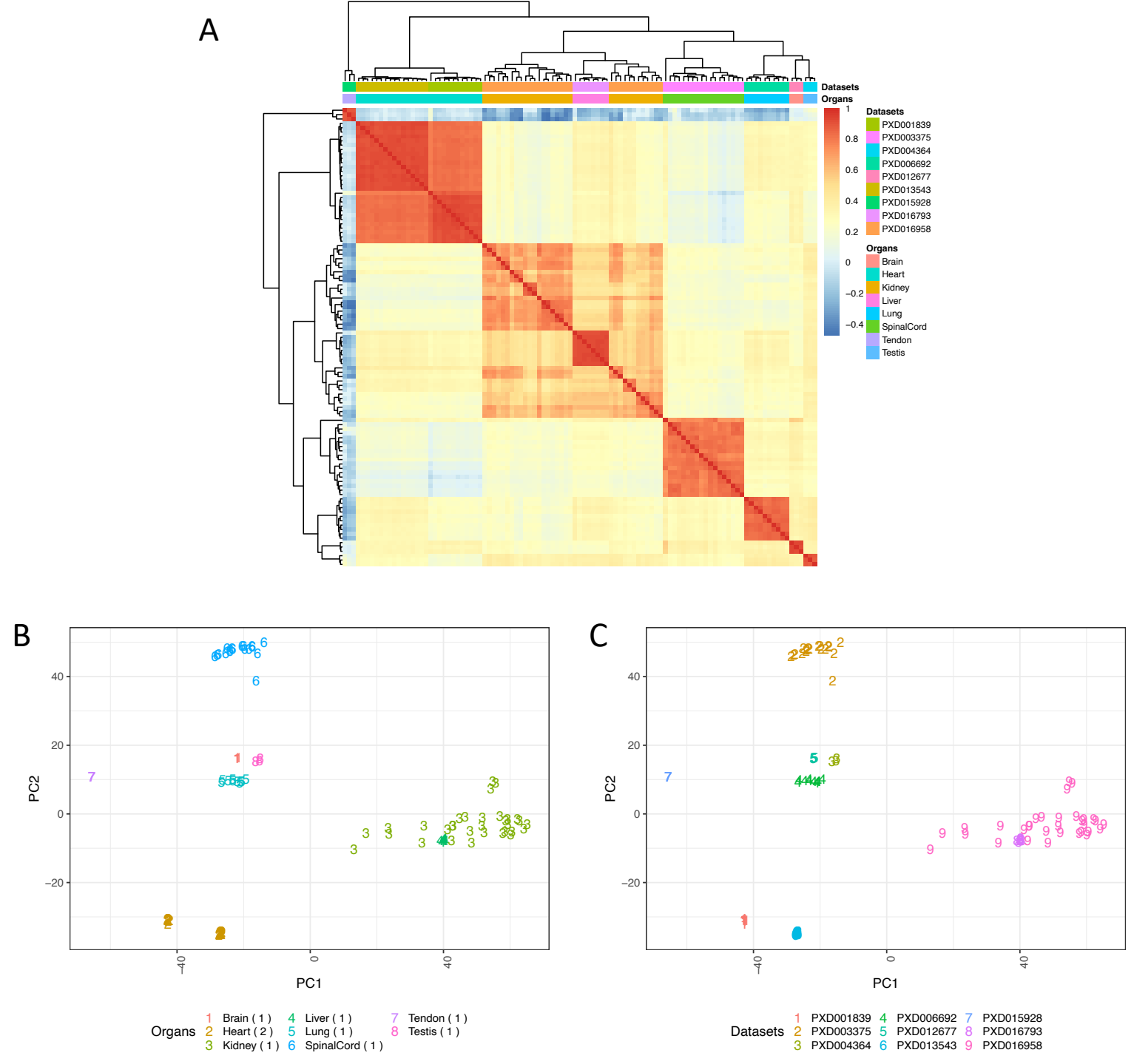
885

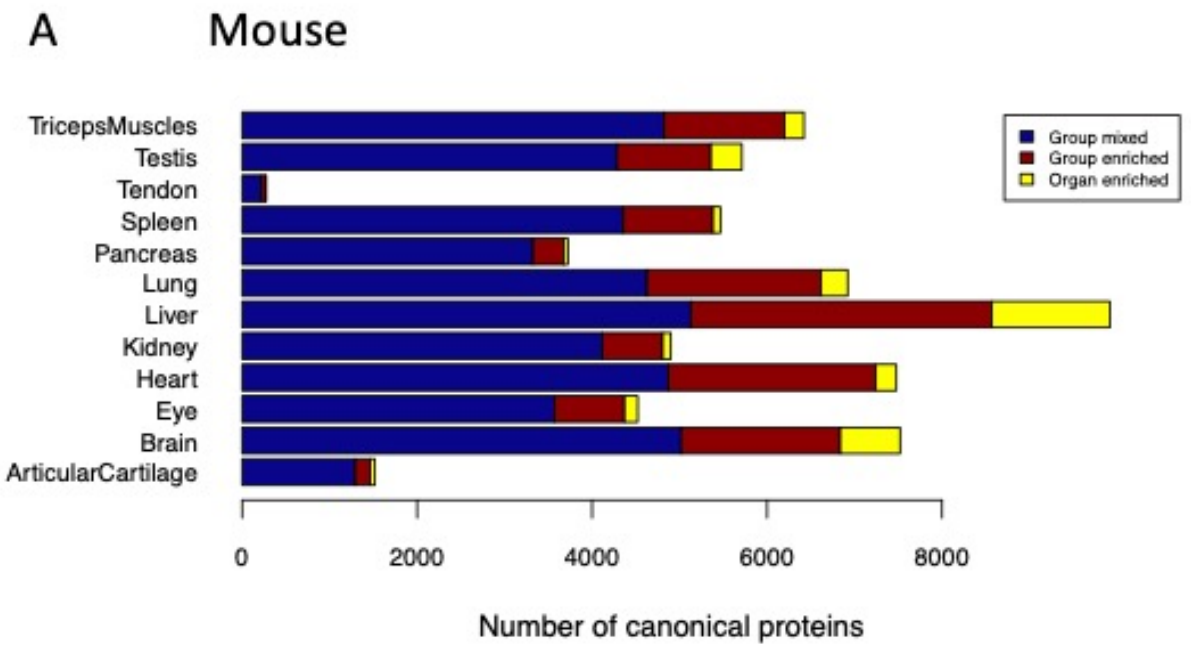

PRIDE database
Curation
SQL
datasets selection
Metadata
Annotation
X
Experimental design templates
Annotare
IDF & SDRF
Analysis
MaxQuant
R
Post-processing
Normalized protein abundances
QA & Integration into EA
Quality control summary
R
Post-process summary
Expression Atlas

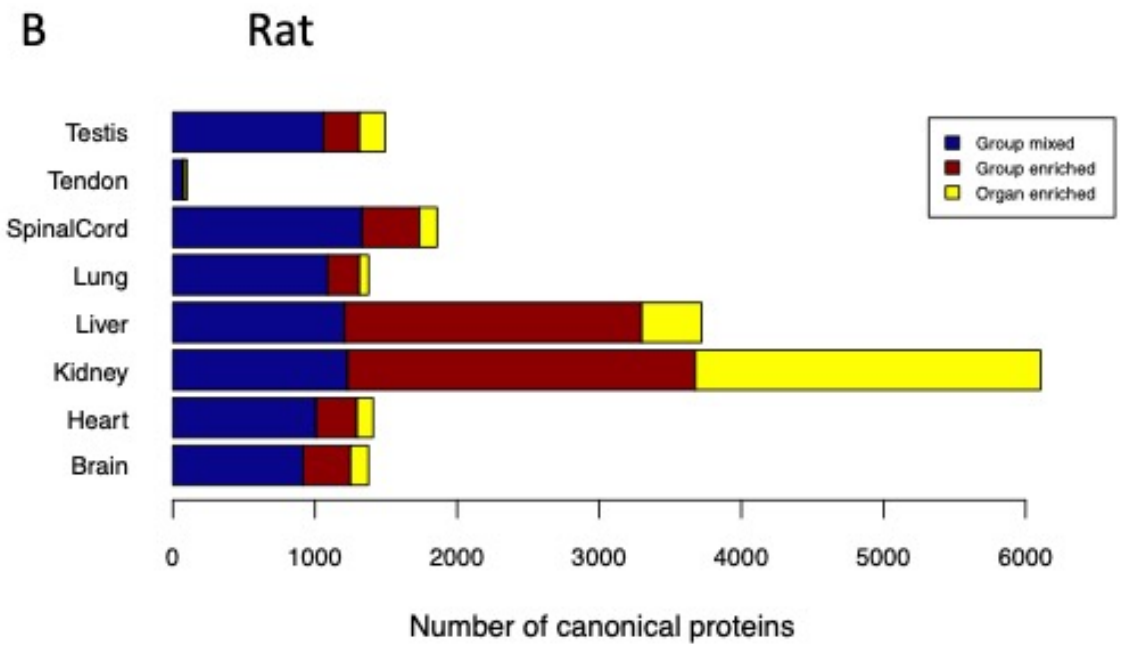

A
Identified canonical proteins
10000
7500
5000
2500
0
ArticularCartilage (6)
Brain (66)
Eye (2)
Heart (11)
Kidney (1)
Liver (11)
Lung (3)
Pancreas (1)
Spleen (1)
Tendon (8)
Testis (1)
TricepsMuscles (3)
C
protein abundance (ppb)
1e+08
1e+05
1e+02
B
PXD000288 (1)
PXD000867 (1)
PXD003155 (2)
PXD004496 (1)
PXD004612 (1)
PXD005230 (3)
PXD008736 (2)
PXD009639 (1)
PXD009909 (1)
PXD012307 (1)
PXD012636 (4)
PXD019394 (8)
PXD019431 (1)
PXD022614 (1)
D
E
number of MS/MS spectra recorded
total number of identified tandem MS spectra
Datasets
Organs
$R^2 = 0.61$
$R^2 = 0.51$
$R^2 = 0.38$
$R^2 = 0.34$
Total spectra counts
F
Number of canonical proteins
Present in number of organs

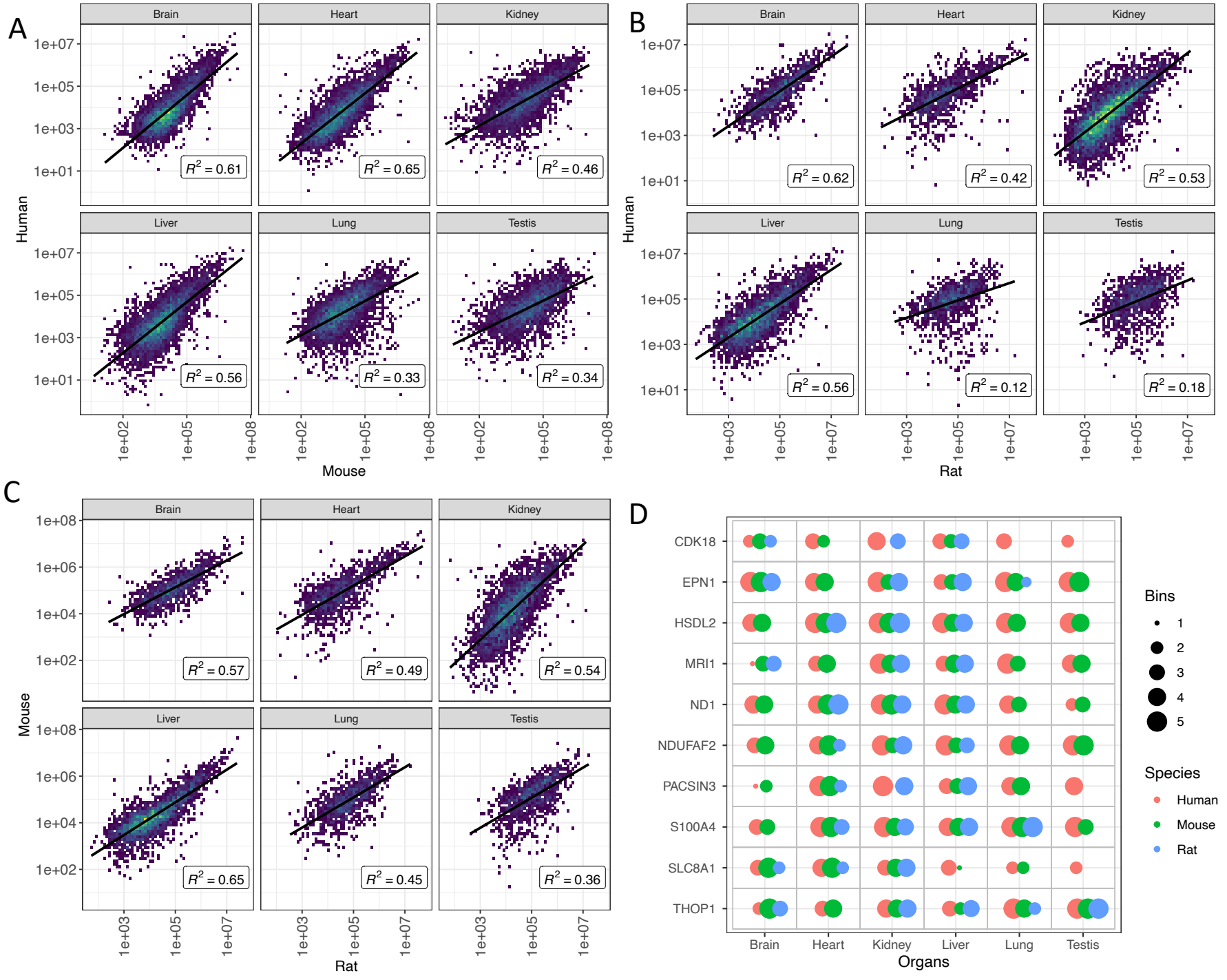

A
Identified canonical proteins
6000
4000
2000
0
Brain (3)
Heart (28)
Kidney (32)
Liver (8)
Lung (10)
SpinalCord (18)
Tendon (3)
Testis (3)
C
protein abundance (ppb)
1e+08
1e+06
1e+04
1e+02
Brain (3)
Heart (28)
Kidney (32)
Liver (8)
Lung (10)
SpinalCord (18)
Tendon (3)
Testis (3)
B
Identified canonical proteins
6000
4000
2000
0
PXD001839 (1)
PXD003375 (2)
PXD004364 (1)
PXD006692 (1)
PXD012677 (1)
PXD013543 (1)
PXD015928 (1)
PXD016793 (1)
PXD016958 (10)
D
protein abundance (ppb)
1e+08
1e+06
1e+04
1e+02
PXD001839 (1)
PXD003375 (2)
PXD004364 (1)
PXD006692 (1)
PXD012677 (1)
PXD013543 (1)
PXD015928 (1)
PXD016793 (1)
PXD016958 (10)
E
number of MS/MS spectra recorded
total number of identified tandem MS spectra
Datasets
Organs
Identified canonical proteins
3000
1000
300
100
PXD016958 (10)
PXD016793 (1)
PXD003375 (2)
PXD001839 (1)
PXD015928 (1)
$R^2 = 0.79$
$R^2 = 0.88$
Kidney (32)
SpinalCord (18)
Liver (8)
Testis (3)
Brain (3)
Lung (10)
Heart (28)
Tendon (3)
$R^2 = 0.79$
$R^2 = 0.88$
1e+04
1e+05
1e+06
Total spectra counts
F
Number of canonical proteins
3000
2000
1000
0
1
2
3
4
5
6
7
8
Present in number of organs

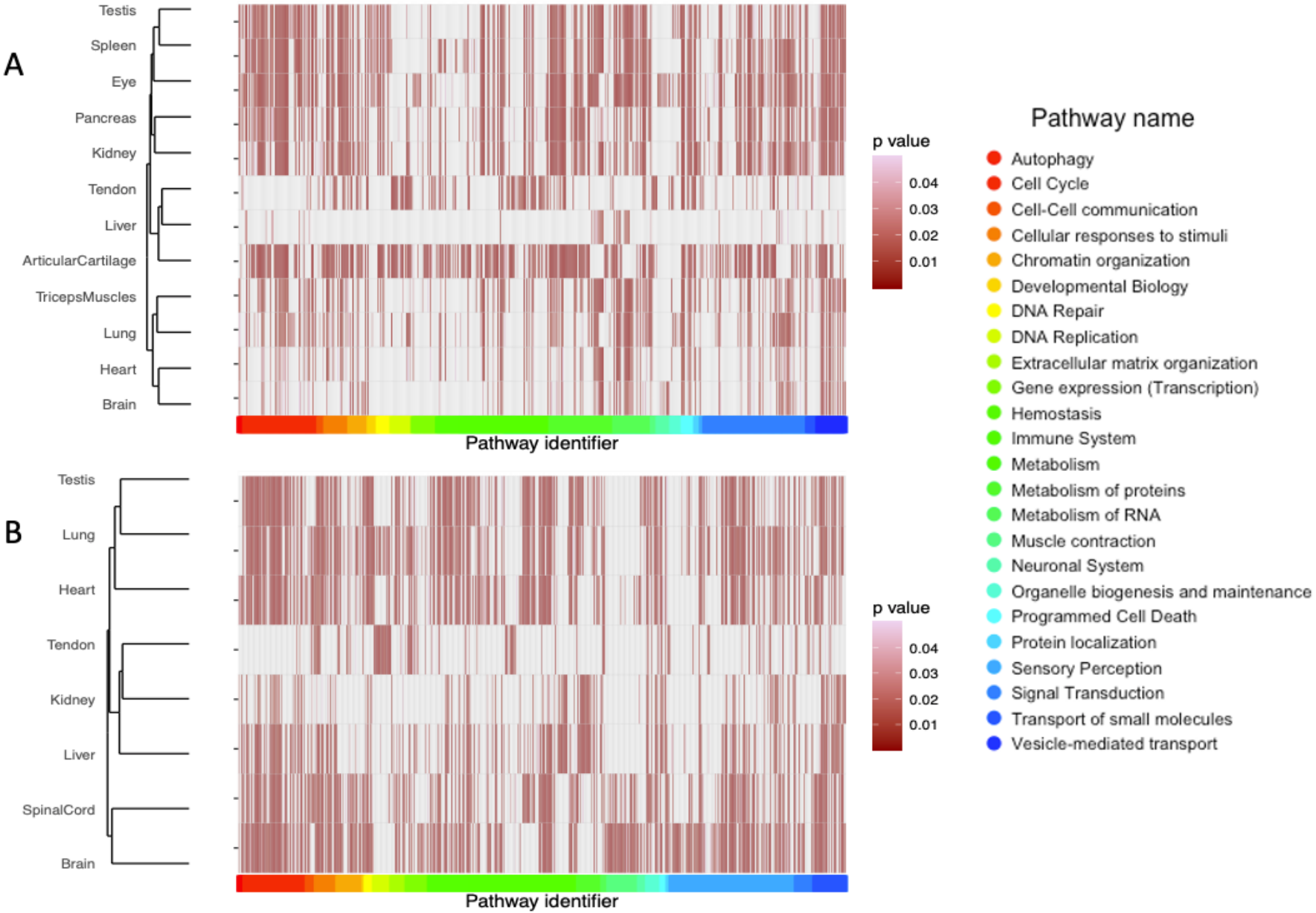

A
Datasets
Organs
1
0.8
0.6
0.4
0.2
0
−0.2
Datasets
PXD000288
PXD000867
PXD003155
PXD004496
PXD004612
PXD005230
PXD008736
PXD009639
PXD009909
PXD012307
PXD012636
PXD019394
PXD019431
PXD022614
Organs
ArticularCartilage
Brain
Eye
Heart
Kidney
Liver
Lung
Pancreas
Spleen
Tendon
Testis
TricepsMuscles
B
PC2
PC1
Organs
1 ArticularCartilage (1)
2 Brain (5)
3 Eye (2)
4 Heart (3)
5 Kidney (1)
6 Liver (3)
7 Lung (2)
8 Pancreas (1)
9 Spleen (1)
10 Tendon (1)
11 Testis (1)
12 TricepsMuscles (1)
C
PC2
PC1
Datasets
1 PXD000288
2 PXD000867
3 PXD003155
4 PXD004496
5 PXD004612
6 PXD005230
7 PXD008736
8 PXD009639
9 PXD009909
10 PXD012307
11 PXD012636
12 PXD019394
13 PXD019431
14 PXD022614


A
Datasets
Organs
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
Datasets
PXD001839
PXD003375
PXD004364
PXD006692
PXD012677
PXD013543
PXD015928
PXD016793
PXD016958
Organs
Brain
Heart
Kidney
Liver
Lung
SpinalCord
Tendon
Testis
B
PC2
PC1
40
20
0
−20
−40
0
40
Organs
1 Brain (1)
2 Heart (2)
3 Kidney (1)
4 Liver (1)
5 Lung (1)
6 SpinalCord (1)
7 Tendon (1)
8 Testis (1)
C
PC2
PC1
40
20
0
−20
−40
0
40
Datasets
1 PXD001839
2 PXD003375
3 PXD004364
4 PXD006692
5 PXD012677
6 PXD013543
7 PXD015928
8 PXD016793
9 PXD016958


A
Mouse
TricepsMuscles
Testis
Tendon
Spleen
Pancreas
Lung
Liver
Kidney
Heart
Eye
Brain
ArticularCartilage
Group mixed
Group enriched
Organ enriched
0
2000
4000
6000
8000
Number of canonical proteins


B
Rat
Testis
Tendon
SpinalCord
Lung
Liver
Kidney
Heart
Brain
Group mixed
Group enriched
Organ enriched
0
1000
2000
3000
4000
5000
6000
Number of canonical proteins


A
Brain
Heart
Kidney
Liver
Lung
Testis
Human
Mouse
$R^2 = 0.61$
$R^2 = 0.65$
$R^2 = 0.46$
$R^2 = 0.56$
$R^2 = 0.33$
$R^2 = 0.34$
1e+01
1e+03
1e+05
1e+07
1e+02
1e+05
1e+08
B
Rat
$R^2 = 0.62$
$R^2 = 0.42$
$R^2 = 0.53$
$R^2 = 0.56$
$R^2 = 0.12$
$R^2 = 0.18$
1e+03
1e+05
1e+07
C
$R^2 = 0.57$
$R^2 = 0.49$
$R^2 = 0.54$
$R^2 = 0.65$
$R^2 = 0.45$
$R^2 = 0.36$
1e+02
1e+04
1e+06
1e+08
D
CDK18
EPN1
HSDL2
MRI1
ND1
NDUFAF2
PACSIN3
S100A4
SLC8A1
THOP1
Organs
Bins
1
2
3
4
5
Species
Human
Mouse
Rat


A
Testis
Spleen
Eye
Pancreas
Kidney
Tendon
Liver
ArticularCartilage
TricepsMuscles
Lung
Heart
Brain
Pathway identifier
p value
0.04
0.03
0.02
0.01
B
Testis
Lung
Heart
Tendon
Kidney
Liver
SpinalCord
Brain
Pathway identifier
p value
0.04
0.03
0.02
0.01
Pathway name
Autophagy
Cell Cycle
Cell-Cell communication
Cellular responses to stimuli
Chromatin organization
Developmental Biology
DNA Repair
DNA Replication
Extracellular matrix organization
Gene expression (Transcription)
Hemostasis
Immune System
Metabolism
Metabolism of proteins
Metabolism of RNA
Muscle contraction
Neuronal System
Organelle biogenesis and maintenance
Programmed Cell Death
Protein localization
Sensory Perception
Signal Transduction
Transport of small molecules
Vesicle-mediated transport