Molecular complexity of the major urinary protein system of the Norway rat, *Rattus norvegicus*

Guadalupe Gómez-Baena[1], Stuart D. Armstrong[1], Josiah O. Halstead[2], Mark Prescott[1], Sarah A. Roberts[2], Lynn McLean[1], Jonathan M. Mudge[3], Jane L. Hurst[2], Robert J. Beynon[1¶]

[1]Centre for Proteome Research, Institute of Integrative Biology, University of Liverpool, Crown Street, L697ZB, Liverpool, United Kingdom

[2]Mammalian Behaviour and Evolution Group, University of Liverpool, Leahurst Campus, Neston, United Kingdom

[3]EMBL-EBI, Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, United Kingdom

¶Corresponding author: Robert J. Beynon (r.beynon@liverpool.ac.uk)

Keywords: *Rattus norvegicus*, α2u-globulins, MUPs, phosphorylation

Short title: Urinary MUPs in rats

ABSTRACT

Major urinary proteins (MUP) are the major component of the urinary protein fraction in house mice (*Mus* spp.) and rats (*Rattus* spp.). The structure, polymorphism and functions of these lipocalins have been well described in the western European house mouse (*Mus musculus domesticus*), clarifying their role in semiochemical communication. The complexity of these roles in the mouse raises the question of similar functions in other rodents, including the Norway rat, *Rattus norvegicu*s. Norway rats express MUPs in urine but information about specific MUP isoform sequences and functions is limited. In this study, we present a detailed molecular characterization of the MUP proteoforms expressed in the urine of two laboratory strains, Wistar Han and Brown Norway, and wild caught animals, using a combination of manual gene annotation, intact protein mass spectrometry and bottom-up mass spectrometry-based proteomic approaches. Detailed sequencing of the proteins reveals a less complex pattern of primary sequence polymorphism than the mouse. However, unlike the mouse, rat MUPs exhibit added complexity in the form of post-translational modifications including phosphorylation and exoproteolytic trimming of specific isoforms. The possibility that urinary MUPs may have different roles in rat chemical communication than those they play in the house mouse is also discussed.

INTRODUCTION

Physiological production of substantial protein in the urine is well known in both rats and house mice [1, 2]. The protein fraction is dominated by 18-19 kDa, eight stranded beta-barrel lipocalins known as major urinary proteins (MUPs, also named as α2u-globulins when first identified in rats [2, 3]). Urinary MUPs are a heterogeneous mixture of multiple isoforms that are very similar in mass and isoelectric point [4-6]. The functions of MUPs have largely been studied in the western European house mouse (*Mus musculus domesticus*) where they play critical roles in olfactory communication. First, they act as carriers for low molecular weight pheromones and other constituents, delaying their release from urinary scent marks [7-9]. MUP polymorphism also provides an identity signal for individual and kin recognition [10-14] and may play a role in species recognition [6]. Finally, MUPs act as pheromones in their own right [14-17]. In particular, darcin (MUP20, MGI nomenclature; http://www.informatics.jax.org/) has a number of unique properties, including a highly specialized role as a male sex pheromone that also induces competition between males. This protein binds most strongly the abundant volatile male pheromone (S)-2-(sec-butyl)-4,5-dihydrothiazole in mouse urine [16-19]. The pheromonal properties of darcin are retained in the recombinant protein showing that it acts as a pheromone in the absence of bound ligands [16, 17].

The structure and functions of MUPs in the house mouse are well established and serve to emphasise the significantly lower degree of understanding of the MUP system in rats, which differ in social organization from house mice [20]. Evidence is emerging that rat MUPs are likely to be important in male sexual and/or competitive communication, with urinary MUP output appearing around male puberty and increasing with a surge in testosterone levels [21, 22]. Male rats that are preferred by females express a greater amount of urinary MUP, and female rats are attracted to spend time near the high molecular weight fraction of male urine that contains rat MUPs and other urinary proteins [23]. Females also spend longer sniffing glass rods painted with castrated male urine if recombinant MUPs are added to the urine at normal physiological concentration [22]. Exposure to recombinant MUPs stimulates increased expression of the immediate early gene *c-fos* in the accessory olfactory bulbs of females and in brain areas known to be involved in pheromone-induced sexual behaviours [22]. However, the specific functions played by rat MUPs in sexual and/or intrasexual competitive communication have yet to be addressed, and it is not known whether different MUP isoforms play different roles as among mice.

While urinary MUP expression has been well characterized in mice and the urinary protein pattern can largely be reconciled with genome-level evidence [24-26], comparable information about the isoforms of MUPs expressed in rats is limited. There are no studies that have provided a deep analysis of the MUP protein complement in rat urine, a necessary step that precedes characterization of the function of the individual isoforms in communication. Although the rat genome sequence was first published in 2004 [27],

77 gene annotation has lagged behind that of the mouse genome and it is more difficult to connect proteins
78 observed in rat urine to the cognate coding sequences predicted from the genome sequence. Furthermore,
79 phenotyping of individual urinary MUPs isoforms in rats has previously been based largely on 1D/2D-SDS-
80 PAGE, or isoelectric focusing (with or without prior purification) [2, 3, 21, 22, 28-32]. Neither PAGE nor
81 isoelectric focusing alone provides adequate resolution for the highly heterogeneous mixture of MUPs
82 isoforms. By contrast, intact mass analysis by electrospray ionization (ESI-MS), complemented with mass
83 spectrometry based protein sequencing, has proved a valuable tool for the characterization of the urinary
84 MUP profiles in different species and strains [4-6, 25]. There have been some studies that make use of
85 analytical mass spectrometry to study rat urinary proteome but none of these have addressed the issue of
86 complexity and isoform phenotyping.
87
88 To gain a similar level of understanding on rat MUPs that exists for the house mouse, we have performed a
89 manual annotation of the MUP genome cluster in the latest assemblies of the rat genome and completed
90 phenotypic urinary profiling of male and female individuals from the laboratory strains Brown Norway and
91 Wistar Han, and from some wild caught individuals. This strategy has allowed the detailed characterization
92 of individual isoforms, reconciling genomic information and protein data, and has provided new insight into
93 the post-translational modifications undergone by the rat MUP family, including phosphorylation and
94 exoproteolysis.
95

Results and Discussion

MUP genome cluster analysis

The first iteration of the sequenced rat genome was published in 2004 [33]. Since then, several assemblies have been released. We performed manual annotation of the rat MUP cluster on rat chromosome 5 using the genome assemblies RGSC_3.4 (v4) (December 2004), Rnor_5.0 (v5) (March 2012) and Rnor_6.0 (v6) (July 2014) from the Rat Genome Sequencing Consortium (Figure 1). Manual annotation revealed ten genes in v4 (numbered to maintain the nomenclature utilized by Logan and colleagues [26]) and eight in v5 and v6 (named A-H), along with several pseudo-genes (twelve in v4 and ten in v5 and v6).

The annotation at v4 accords well with that previously reported [26] (Figure 1) but with the addition of a single protein coding gene (gene H). Six out the 10 protein coding loci have transcriptional evidence, although only three in hepatic tissues (genes 1, 10, 13, yielding mature predicted masses of 18340, 18716 and 18728 Da respectively).

[Insert Figure 1 here]

In v5 and v6, several genes (genes 1, 9, 10, 12 and 13 from v4) are removed compared to the v4 annotation, including two genes for which protein-level evidence had been obtained in urine; gene 13 (18728 Da) and gene 1 (18340 Da) [22, 34-36]. The fact that these putative genes (genes 1, 9, 10, 12 and 13 from v4) have transcriptional support indicates that these are genuine protein coding genes. Assembly v5 and v6 define a duplication of genes 3 and 4 that previously had single instances in v4. Additionally, genes F and G in v5 and v6 are incompletely covered in the genome sequence. Ultimately, the incomplete nature of the genome sequence across the rat MUP cluster compromises the ability to produce fully comprehensive gene annotations at this time. By contrast with the genome project of the C57BL/6 mouse, which utilized a hierarchical mapping and clone-based sequencing strategy, the rat genome sequences were generated almost entirely through whole-genome shotgun sequencing. We may anticipate that the highly duplicative nature of the MUP locus presents particular challenges to this strategy, including the assembly of DNA sequences into a correct genomic region. In fact, we cannot assume that the v5/v6 assemblies are necessarily of better quality than v4 across this particular locus, and indeed our findings below illustrate that v4 contains genuine gene features that were lost during subsequent reassemblies.

Protein analysis reveals sexual dimorphism in urinary MUP expression

Urinary MUPs are synthesized in the liver, secreted into the bloodstream and passed through the glomerular filter before being released in the urine (for a review [37]). Hepatic expression of MUPs is under

131 sex and growth hormone control in both the mouse and the rat [38-40]. However, a striking difference
132 between mouse and rat is the much more pronounced sexual dimorphism in expression of urinary MUPs in
133 the rat. Whilst female mice have urinary MUP output that is approximately a third to a quarter that of males
134 on average [8], female rats express virtually no MUPs in the liver [41-44] and as a consequence, no MUPs
135 are apparent in urine [29, 30, 45].
136
137 The overall workflow for MUP characterization is summarized in Supplementary Figure 1. First, we
138 measured the protein concentration in urine from male and female adults of two laboratory strains (Wistar
139 Han and Brown Norway) as well as from some wild caught individuals. To correct for variation in urine
140 dilution, protein output was normalized to urinary creatinine, a non-enzymatic by-product of muscle
141 metabolism [46]. This confirmed that males had significantly higher protein output (Figure 2A). For Brown
142 Norway males, the level was almost three times higher than that in females; for Wistar Han males it was
143 almost five times higher and among wild males total urinary protein output was double that of females.
144
145
146 [Insert Figure 2 here]
147
148
149 SDS-PAGE analysis of urine (Figure 2B) revealed a strongly expressed band at 18-19 kDa, present in all male
150 samples and representing approximately 70 – 80 % of the total urinary protein. By contrast, a much fainter
151 band was evident at a similar position in female urine. In-gel digestion, followed by PMF and tandem mass
152 spectrometry, confirmed the presence of MUP peptides in the male expressed band, however, in-gel
153 digestion revealed the identity of protein in the female band as rat urinary protein 2 (RUP-2, Uniprot KB
154 P81828). The absence of peptides from MUPs in the female band is in good agreement with previous
155 studies showing the scarcity of MUPs in female urine [29, 30, 45]. Additionally, in males, other proteins,
156 including prostatic steroid binding protein (PsBpc2, Uniprot KB P02781) and the serine protease inhibitor
157 A3K (SPI-A3, Uniprot KB P05545), were identified in other bands (Figure 2B). The urine of both sexes
158 contained albumin, immunoglobulins and rat urinary protein 1 (RUP-1, Uniprot KB P81827, not to be
159 confused with MUPs). Both laboratory rat strains and wild rats exhibited a low level of albuminuria.
160
161 Phenotypic profiling to evaluate MUP heterogeneity and polymorphism
162 In laboratory mouse strains, the pattern of MUP expression is limited by inbreeding and consequent
163 homozygosity at the *Mup* locus, but in wild-caught mice, heterogeneity is much more pronounced, both
164 between mouse populations and between individuals of the same population [5, 6, 25, 47]. The highly
165 polymorphic combinatorial nature of wild mouse urinary MUPs is the basis for individual recognition [10, 11,

166 13], driving assessment of genetic heterozygosity [48] and avoidance of inbreeding [12, 49]. It was of
167 interest therefore to explore the heterogeneity in rat urinary MUPs. We have previously used ESI-MS to
168 profile the isoforms of the MUPs secreted in mouse urine [6, 13, 25]. MUPs yield strong signals on ESI-MS
169 and the intact masses can be determined to within ± 1 Da, permitting matching to predicted mature protein
170 masses from genomic or cDNA sequences [25]. The masses obtained by ESI-MS correspond to the neutral
171 average mass of the mature form of the protein, after the removal of the predicted signal peptide [50], and
172 subtraction of 2 Da for the formation of a single disulphide bond, based on homology with known MUP
173 structures [19, 36]. ESI-MS also allows semi-quantitative assessment of the relative amounts of each isoform
174 [51].
175
176 In ESI-MS analysis of male urine, MUPs dominated the deconvoluted mass spectra (Figures 3 and 4) and
177 several proteins and multiple discrete masses in the 18-19 kDa mass range were evident. All laboratory male
178 rats, irrespective of strain, expressed proteins of masses 18712 Da, 18728 Da, and 18826 Da, the protein at
179 18728 being the most intense in all instances. Additionally, we identified strain-specific proteins at
180 18340 Da, 18420 Da and 18670 Da in Brown Norway rats, whereas proteins at 18553 and 18633 Da were
181 exclusive to Wistar Han rats, the pattern being very stable within individuals of the same strain
182 (Supplementary Figure 2).
183
184
185 [Insert Figure 3 here]
186
187
188 In wild caught animals, urinary MUPs at 18728 Da and 18712 Da were dominant in 8 out of 9 individuals
189 examined (Figure 4A). Only one individual was distinct in having a dominant peak at 18715 Da. Less
190 abundant protein peaks were present, most prominently at 18340 Da and 18420 Da in six of the wild caught
191 individuals and further minor peaks were observed at 18471 Da and 18694 Da. Thus, the pattern for wild
192 individuals matched more closely that of the Brown Norway strain (Figure 3).
193
194
195 [Insert Figure 4 here]
196
197
198 Although MUP profiles differed between the two laboratory strains and, as expected, within each strain the
199 pattern was rather stable, the low degree of polymorphism among wild caught individuals (Figure 4A) was
200 unanticipated. Compared with previous observations of house mice [5, 6, 13, 25, 47], there was significantly

201 less polymorphism in protein isoforms as evidenced by the ESI-MS pattern of wild rats. To explore this in
202 more detail, the same samples of wild rat urine were resolved by isoelectric focusing (IEF) (Figure 4B) to
203 separate proteins by net charge – since MUPs were the predominant bands, they would be most prominent
204 bands after isoelectric focusing. The protein banding patterns of nine individual male wild rats were similar
205 and most of the urine samples resolved to three major and a few low intensity discrete bands. This was
206 consistent with previous IEF studies of laboratory rat MUPs [22, 30, 52] but with fewer bands than recorded
207 for house mice [10].
208
209 Roberts et al. [13] showed that mice are sensitive to changes in the relative ratios of MUP isoforms in urine.
210 In rats, despite the absence of qualitative polymorphism between urine samples, the relative amount of
211 each isoform differed between individuals. We quantified the relative abundance of each MUP mass from
212 the peak area of the ESI-MS deconvoluted spectra, and calculated the correlation between the amounts of
213 each protein, per individual (Figure 5). While laboratory strains showed high correlations between the
214 relative amounts of the isoforms, this was not the case for samples derived from wild caught rats. The two
215 protein masses that correlated in intensity most strongly among the wild caught individuals was the pair at
216 18340 Da and 18420 Da.
217
218
219 [Insert Figure 5 here]
220
221
222 By contrast, ESI-MS of female rat urine (Supplementary Figure 3), showed two clusters of protein masses of
223 around 11 kDa, and no mass peaks in the expected range of MUPs (18-19 kDa). We have not investigated
224 these 11 kDa proteins further, but they are likely to be RUPs (rat urinary proteins). As anticipated, ESI-MS
225 provides further confirmation of the lack of MUP expression in female rats.
226
227 Characterization of the MUP proteoforms secreted in rat urine
228 To provide further MUP characterization, native gel electrophoresis and strong anion exchange
229 fractionation were used to resolve the MUP mixture into discrete proteins to sequence by PMF and LC-
230 MS/MS. For this purpose, we created an in-house database containing the sequences of the mature forms
231 of MUPs derived from the gene annotation and transcript sequences published to date (Figure 1, Table 1),
232 combined with all the protein entries in the Uniprot database for *Rattus norvegicus*. Supplementary Figures
233 4-8 provide the results of the different experimental approaches for the predicted proteins in Table 1.
234 Supplementary Figure 9 shows a comparison of the protein sequences of the predicted rat MUP isoforms
235 highlighting unique peptides for each isoform. From this detailed analysis, we could compile the evidence

236 for each of the predicted proteins in the rat gene assembly.

237

MUP gene	RGD genome Annotation	Predicted mature mass (Da)	UniParc	Uniprot accession	Uniprot names	Transcript supportive information
MUP 1	v4	18340	UPI000017083A	Q78E14	Obp3 protein; Rat salivary gland (alpha)2(mu)globulin, type 1	Liver (BC086942) Salivary gland (X14552)
	v4	18553	UPI00000E8911	Q63213	Alpha-2u globulin; PGCL4	Submaxillary gland (J00738) Preputial gland (AB039825)
MUP 2	v4/v5/v6	18642	UPI00000E7901	Q9JJI3	Alpha-2u globulin; PGCL3	Preputial gland (AB039824)
MUP 3	v4 and 2 loci in v5/v6	18670	UPI00000E7381	Q9JJH9	Alpha-2u globulin; Protein Mup4; PGCL8	Preputial gland (AB039829)
MUP 4	v4 and 2 loci in v5/v6	18909	UPI0000506C83	A0A096MK41	Uncharacterized protein	
MUP 9	v4	19010				Preputial gland (AB039827, sequence conflict K79 to E79, Uniparc UPI00000E7BD9; Q9JJI1)
MUP10	v4	18716	UPI00000E7542	Q8K1Q6	Alpha-2u globulin; PGCL2	Liver (BC086943) Preputial gland (AB039823; sequence conflict in signal peptide, UniParc UPI00000E8694, Q9JJI4)
MUP 12	v4	19021	UPI00000E5BE6	Q9JJI2	Alpha-2u globulin; PGCL5	Preputial gland (AB039826)
MUP 13	v4	18728	UPI000000086C	P02761	Major urinary protein (MUP_RAT); PGCL1; Allergen rat n1; Alpha-2u globulin PGCL1	Liver (M26835); Liver (M26837); Preputial gland (AB039822); Liver (BC088109); Liver (BC098654); Spleen (BC105816); U31287; Liver (V01220); Liver (J00737)
MUP 15	v4 (full) in v5/v6 (fragments)	18712				
MUP H	v4/v5/v6	18772	UPI00005035E7	MOR620	Major urinary protein like	
Transcript AB039828	No annotation	18822	UPI00000E6465	Q9JJI0	Alpha-2u globulin; PGCL7	preputial gland
Transcript M26836	No annotation	18726	UPI00000E6420	Q63024	Rat alpha-2u-globulin (L type)	Liver
Transcript M26838	No annotation	18712	UPI00000E6E81	Q63025	Rat alpha-2u-globulin (S type)	Liver

238

239 **Table 1| Current knowledge of rat MUP genes, transcript expression and**
240 **protein products.** This table shows an updated compilation of data from three releases of
241 the rat genome sequence (RGSC_3.4 (v4) (December 2004), Rnor_5.0 (v5) (March 2012) and
242 Rnor_6.0 (v6) (July 2014) from the Rat Genome Sequencing Consortium) using the
243 annotations compiled in the Rat Genome Database (http://rgd.mcw.edu/) and Uniprot
244 (http://www.uniprot.org/). Where possible, data relating predicted mature protein product is
245 cross-correlated with experimental data that confirm true protein products.

246

247 In female rat samples, shotgun 'bottom-up' proteomics allowed the identification of MUPs at very low
248 levels, in good agreement with previous papers establishing the presence of trace levels of MUPs in female
249 urine [29, 30]. However, the very low abundance of these proteins meant that few peptides were observed
250 and the resulting protein coverage did not allow confident assignment to any of the predicted proteins. By
251 contrast, the same approach revealed the MUP isoform composition of male samples. Below, we discuss the
252 protein-level evidence for each of the genes and include information from transcripts published to date
253 (Table 1), focusing predominantly on genome assembly v4 for these assignments. For these analyses, we
254 have retained the rat MUP numbering scheme first proposed by Logan et al [26] although this scheme also
255 numbered the pseudogenes in the same sequence, in genome order. This numbering is now referenced in
256 other studies [22, 53]. Indeed, a logical nomenclature based on gene order is impossible until a fully
257 assembled and annotated analysis of this region of the rat genome is available.

258

259 *Mup1 gene:* Manual annotation of the genome assembly v4 predicts a protein of mature mass 18340 Da
260 although, as previously mentioned, this gene was omitted from later assemblies (Figure 1). There are

261 multiple transcriptional support data for this gene from liver (BC086942) [54], which is the primary source
262 of urinary MUPs, and also salivary gland [55], with the cDNA predicting 18340 Da as mature protein mass.
263 This MUP has also been referred to as OBP3 [22, 56], despite its high similarity to other MUPs and much
264 lower similarity to rat OBP1 (28%) or OBP2 (18%). It is now clear that the gene encoding this protein is part
265 of the MUP gene cluster. Further, unlike nasal MUPs in mice, which seem to be tissue specific, it is possible
266 that the same MUP could play a dual role in odour reception and scent signalling, as it is expressed at high
267 level in both the nose and urine. ESI-MS intact mass phenotyping showed the mass of 18340 Da in Brown
268 Norway and in wild individuals. More detailed molecular analysis allowed assignment of this mass to the
269 protein predicted by the *mup1* gene. Native gel electrophoresis followed by PMF provided good coverage of
270 the protein (Supplementary Figure 5, band E and Supplementary Figure 6, band F). Besides, fractionation of
271 the urine followed by proteolytic digestion of the protein and tandem MS of the peptides provided
272 confident identification of the MUP1 protein (Supplementary Figure 8). There is also transcriptional support
273 for the *mup1* gene from preputial gland (PGCL4) [57] and submaxillary gland (J00738; UniProt Q63213_RAT)
274 [58]. However, after detailed examination of these sequences we conclude that they predict a protein
275 identical to MUP1 except for two additional amino acids (-RG) at the C-terminus. This longer form predicts a
276 mature mass of 18553 Da, a mass that we observed in the intact mass profile of Wistar Han males and
277 occasionally in wild animals. Our analysis (Supplementary Figure 4, band E) allowed the assignment of the
278 18553 Da mass to the protein predicted by these transcripts, even though a gene designation may not have
279 been possible because the Brown Norway strain used for the rat genome analysis does not express this
280 mass.
281
282 We observed two protein peaks, of masses 18420 Da (Brown Norway and wild) and 18633 Da (Wistar Han)
283 that could not be predicted from any of the genes described in any annotation of the MUP gene cluster, nor
284 could these masses be generated by exopeptidase trimming of any known MUP sequence. Notably, these
285 masses both differed from predicted masses 18340 Da and 18553 Da by 80 Da, a mass shift that might have
286 been a consequence of multiple primary sequence changes but which was also consistent with the addition
287 of a single phosphate group to a side chain residue. When urinary proteins were fractionated to resolve
288 additional variants, the pairs at 18340/18420 Da, and 18553/18633 Da, eluted very closely in the
289 chromatogram (Supplementary Figures 7 and 8), although the heavier protein was slightly more anionic in
290 both cases, consistent with phosphorylation. Proteomic analysis of the chromatographic fractions
291 containing proteins at 18633 Da and 18420 Da yielded extensive coverage for the protein sequences of that
292 corresponded to the MUPs of masses 18553 Da and 18340 Da respectively, indicating a strong primary
293 sequence relationship between the 80 Da separated proteins. To explore this further, LC-MS/MS peptide
294 data from each protein fraction were analysed using Peaks software (Bioinformatics solutions Inc.) to search
295 for post-translational modifications. For both protein fractions, the top-scoring endopeptidase Lys-C

296 peptides revealed convincing evidence for phosphorylation of a serine residue at position 4 in the mature
297 sequence of the protein (Figure 6). The proteins of masses 18553 Da and 18340 Da both share the same N-
298 terminal sequence (Supplementary Figure 9). Manual annotation of the product ion spectra from either the
299 Lys-C cleaved ($[M+2H]^{2+}$, **m/z**=844.86) or tryptic N-terminal peptide ($[M+2H]^{2+}$, **m/z**=474.18^{2+}) revealed
300 high quality coverage and unambiguous identification of a phosphorylation event at Ser_4 (Figure 6). The
301 phosphorylated forms were also resolved and identified from native gel electrophoresis and PMF (the
302 unmodified N-terminal Lys-C peptide corresponding to m/z=1608 Da ($[M+H]^+$) and the phosphorylated
303 version corresponding to m/z=1688 Da ($[M+H]^+$)) (Supplementary Figures 4, 5 and 6).
304
305
306 [Insert Figure 6 here]
307
308
309 Although phosphorylation of extracellular proteins is not well studied, the Ser_4 residue sits within a
310 consensus sequence motif (SxE) (Supplementary Figure 9) for phosphorylation by FAM20C kinase, the
311 enzyme responsible for the phosphorylation of most secreted proteins in humans [59]. The rat genome
312 contains an ortholog gene of human FAM20C kinase on chromosome 12 (RGD:1311980) and other
313 members of the rat MUP family contain this phosphorylation motif (Supplementary Figure 9), invoking the
314 possibility of phosphorylation in other isoforms, although we have no evidence so far that other isoforms
315 are phosphorylated to the same extent as MUP1.
316
317 There is a report of phosphorylation of MUPs in *Rattus rattus*, specifically from the preputial gland [60].
318 Phosphorylation of MUPs was considered based on spot distribution on 2D gels, however no other evidence
319 was provided and the proposed site, at Ser51 does not sit within the consensus sequence of FAM20C
320 kinase. Therefore, this putative phosphorylation site requires further validation. Phosphorylation
321 significantly influences ligand binding affinities of porcine OBP [61] suggesting that this may have an
322 influence on both the signature of urinary volatiles bound and released by MUPs and the capture of odours
323 in the nose, but further studies are needed to understand the significance of this modification.
324
325 *Mup2 gene:* The gene encoding this protein predicts a mature mass of 18642 Da. There is transcriptional
326 support for this protein sequence from rat preputial gland (PGCL3, [43]). We found no evidence for this
327 mass in intact mass profiles of either intact urine or after ion exchange fractionation. However, shotgun
328 proteomics gave us high protein coverage including peptides unique to MUP2: 80% protein coverage in
329 both Wistar Han and wild individuals; and 60% protein coverage in Brown Norway. Therefore, we
330 hypothesized three possibilities for the absence of the 18642 Da mass in the intact mass profile: the protein

331 is only present in small amounts, the protein is phosphorylated, or the protein is trimmed at the N-terminal.
332 Regarding phosphorylation, although this protein contains a serine residue at mature sequence position 4, it
333 does not contain the sequence motif for FAM20C kinase and, as anticipated, there was no evidence for
334 phosphorylation in the N-terminal peptide in shotgun proteomics analysis. A common feature in all the
335 identifications of this protein, regardless of individual donor, was the incomplete coverage of the N-terminal
336 part of the sequence, which might suggest trimming of the N terminus to remove between 7 and 10 amino
337 acids. However, we were unable to identify an intact mass peak matching a trimming event either.
338
339 *Mup3 gene:* The predicted protein mass for *mup3* gene product is 18670 Da. Again, there is transcriptional
340 support from rat preputial gland (PGCL8 [43]). PMF after native electrophoresis allowed the identification of
341 this protein in Wistar Han (Supplementary Figure 4, band B), but not in Brown Norway or wild individuals.
342 Further, shotgun proteomics provided 66% sequence coverage and 5 unique peptides on average. However,
343 no evidence of this mass was obtained in the intact mass profiling, suggesting that the protein is expressed
344 at low levels only. The intact mass profile of Brown Norway and wild individuals showed a peak at 18670 Da,
345 however, protein fractionation and further analysis demonstrated that this mass does not correspond to
346 MUP3 but likely to a trimming of the 18728 Da form at the C-terminal (-G) (discussed below).
347
348 *Mup4 gene:* The predicted protein mass for *mup4 gene* product is 18909 Da. There is no transcriptional
349 support for this gene, and we could find no evidence for a gene product in urine in any of the individuals.
350 However, some MUPs are not expressed in urine, and we cannot exclude the possibility of expression in a
351 tissue other than liver, the likely source of urinary MUPs.
352
353 *Mup9 gene:* For this sequence, SignalP [62] predicts a signal peptide two amino acids shorter than that
354 commonly observed in MUPs (17 instead 19 amino acids). Hence this isoform is two amino acids longer than
355 the rest of the isoforms at the N-terminus (Supplementary Figure 9) and the predicted mass of the mature
356 protein is 19010 Da. No peak at that mass was found in any of the samples that were analysed. By contrast,
357 Wistar Han and some wild individuals demonstrated a mass peak at 18745 Da, which matches the predicted
358 protein mass of the *mup9* mature gene product after removal of the usual 19 amino acid signal peptide.
359 Furthermore, while no evidence was found for the predicted N-terminal peptide corresponding to the long
360 form (**HE**EEASFER-), the N-terminal peptide corresponding to the 'short form' (EEASFER-) was readily
361 identified by PMF and LC-MS/MS after native PAGE and in-gel digestion of the corresponding band
362 (Supplementary Figure 4, band A). Therefore, we venture that in this instance, the prediction of the signal
363 peptide cleavage is incorrect and that in common with other MUPs, this protein loses a signal peptide of 19
364 amino acids and has the commonly seen N-terminal sequence of GluGlu. Further, a minor sequence conflict
365 arose at Lys_{81} in the annotated sequence of the *mup9* gene from genome assembly v4, to Glu_{81} suggested

366 by the transcript AB039827 (PGCL6 [43]). We confirmed by in-gel digestion after native PAGE that Wistar
367 Han males possess the Glu_{81} form. Additional evidence is provided by shotgun proteomics, yielding good
368 coverage for this protein in Wistar Han and some wild individuals and equally confirming the Glu_{81} form.
369 There was no evidence for the expression of this protein in the Brown Norway strain.
370
371 *Mup10 gene:* The predicted protein mass for *mup10 gene* product is 18716 Da, supported by transcriptional
372 information from rat liver and preputial gland (PGCL2 [43]). Intact mass analysis, in-gel digestion after native
373 PAGE and SAX fractionation approaches all provided conclusive evidence for the *mup10* predicted protein
374 sequence in both laboratory strains and wild individuals (Supplementary Figure 4).
375
376 *Mup12 gene:* The predicted protein mass for *mup7 gene* product is 19021 Da, for which there is
377 transcriptional support from rat preputial gland (PGCL5 [43]). However, we did not find confident evidence
378 for the expression of this protein in any of the urine samples analysed.
379
380 *Mup13 gene:* The predicted protein mass for *mup13 gene* product is 18728 Da with multiple transcripts
381 supporting the expression of this gene (Table 1). For both laboratory strains, and for wild caught individuals,
382 we obtained confident identification of the protein predicted by the *mup13* gene. This mass is the most
383 intense in the ESI-MS profile (Figures 3 and 4) and the most intense in native or IEF electrophoresis
384 (Supplementary Figure 4 and Figure 4).
385
386 Some of the observed masses in the ESI-MS profile are consistent with an N-terminally processed protein of
387 18728 Da. For example, the mass at 18470 Da, observed in the ESI-MS protein profile in some individuals, is
388 consistent with trimming of the N-terminal amino acids from the 18728 Da isoform. SAX fractionation
389 revealed three masses in the flow through volume (18470, 18399 and 18312 Da) that can be explained by
390 the trimming of the N-terminal amino acids from the 18728 Da form (EE-, EEA- and EEAS-, respectively). The
391 trimming of these amino acids means that the pI becomes close to 6 for all three proteins, which is the pH
392 at which the chromatography is performed and explains their appearance in the flow through (the net
393 charge of the proteins is zero under these conditions, preventing the protein from binding to the column).
394 Native electrophoresis allowed the isolation and sequencing of the protein corresponding to the predicted
395 mass 18470 Da, confirming the trimming of the N-terminus (Supplementary Figure 4-6, band A). For this
396 protein, we identified the N-terminal Lys-C cleaved peptide, corresponding to the removal of EE-, at 1217.58
397 Da by both PMF and LC-MS/MS. Another example of trimming is the protein explaining the mass 18670 Da,
398 seen specifically in the ESI-MS from Brown Norway rats (Supplementary Figure 5, band B). We also found
399 evidence for a C-terminal trimming of the 18728 Da protein (loss of a glycine residue) that would explain the
400 mass 18670 Da (within 1 Da instrument error). Although the MUPs in rodent urine are generally

401 proteolytically-resistant, rat urine contains proteases that could attack the termini of the protein (such as
402 meprin and neprilysin [63], Gómez-Baena et al, in preparation), although further experiments are needed to
403 explore the extent of processing in urine and the biological significance thereof.
404
405 In one wild-caught individual, the mass at 18728 Da was less intense in the ESI-MS profile (Figure 4A,
406 individual 6L), although the band in the range of the 18728 Da in the IEF profile (Figure 4B) was strongly
407 stained for this sample. PMF of this IEF band allowed the sequencing of a new MUP sharing the sequence of
408 the 18728 form but with one amino acid change, from Thr to Ser in position 154, which corresponds with a
409 mass shift of -14 Da explaining the mass of 18714 Da in the intact mass profile for this individual. This
410 mutation was also confirmed by MS/MS data using Peaks PTM predictor.
411
412 *Mup15 gene:* The predicted protein mass for *mup15 gene* product is 18712 Da. There is no transcriptional
413 support for expression and no protein of this mass was apparent in rat urine. Although the intact mass
414 profile shows a peak at 18712 Da, further analysis showed that this mass does not correspond to the
415 predicted MUP15 protein sequence, but the sequence of the transcript M26838 (discussed below).
416
417 *MupH gene:* We refer to this as *mupH* to reflect the fact that it was only identified in later genome
418 assemblies and was thus labelled. The predicted protein mass for *mupH gene* product is 18772 Da. There is
419 no transcriptional support for expression of this gene and there was no evidence for the expression of this
420 protein in any of the samples analysed. It is not yet clear whether this is a true protein coding gene.
421
422 AB039828 transcript: This transcript was isolated from preputial gland (PGCL7 [43]) and would have a
423 predicted mass for the mature protein of 18822 Da. Although we were not able to identify the mass 18822
424 Da in the ESI-MS profile, the mass of 18694 Da, seen in some wild males, matches the cleavage of a single E
425 from the N-terminal of the 18822 Da MUP (18693.4 Da). However, we could obtain no data to support this
426 possibility.
427
428 M26836 transcript: This transcript was isolated from liver [54]. The predicted mass for the protein is 18726
429 Da. However, we could find no evidence for expression of this protein in urine.
430
431 M26838 transcript: This transcript was also isolated from liver [54]. The predicted mass for the protein is
432 18712 Da. In the ESI-MS profile of most of the males a mass at 18712 Da was observable. Native gel
433 electrophoresis followed by PMF and LC-MS/MS allowed confident identification of the protein predicted by
434 the M26838 transcript in wild individuals (Supplementary Figure 4). This protein is one of the most intense
435 bands in the native gels and is likely to be a highly expressed MUP in wild individuals, while our results

436 suggest a lesser expression in Wistar Han and Brown Norway strains.

437

438 ## Analysis of the protein products of the gene cluster

439 We provide a detailed analysis of the isoforms of the major urinary protein system expressed in the urine of
440 *Rattus norvegicus*. We characterized the urinary MUPs from two of the most widely used laboratory strains,
441 Wistar Han and Brown Norway, as well as wild caught individuals. We provide evidence at the protein level
442 for several proteins predicted by the genome assembly suggesting strain-specific expression. There are two
443 levels of variance: at the gene and allele level and at the post-translational level. The entire panoply of the
444 urinary protein products, and their post-translational space, is summarized in Figure 7. Most of the residues
445 that differ between MUP10, MUP13 and M26838 are not shared with other rat MUPs (E_4D in MUP10, D_{15}A
446 in MUP13, L_{118}A in MUP10 and M26838, R_{158}H in M26838), although one variant (D_{29}N) is shared across
447 MUPs 10, 13, 4 and H. The degree of similarity between these three rat MUPs is similar to that between a
448 set of highly similar MUPs in house mice encoded by approximately 15 genes in the central region of the
449 mouse MUP cluster [25]. In mice, these highly similar MUPs provide the basis for an individuality signal in
450 urine scent marks [10, 13, 14], with each individual expressing a fixed signature of these MUPs.
451 Combinatorial polymorphism arises both from variation in MUP sequences (involving a limited set of
452 variable sites) and differential transcription of Mup genes [25]. Mice are able to discriminate different MUP
453 signatures, both through V2Rs in the vomeronasal organ that detect MUPs directly [14] and through
454 differences in the signature of ligands bound and released by MUPs [13]. Although rats express fewer MUPs
455 in urine than mice, combinatorial polymorphism in the relative amounts of each MUP could still provide
456 considerable capacity to encode individual differences. Consistent differences in MUP signatures between
457 strains suggest a high degree of genetic determination, but studies have not yet addressed how stable MUP
458 profiles are in rats, or the sensitivity of rats to discriminate these relatively small differences between rat
459 MUP isoforms or their relative ratios. The molecular characterization of the MUP proteoforms expressed by
460 rats presented here now provides the opportunity for such detailed studies to be carried out, an essential
461 next step to understand the functions of MUPs in rat scent signals. Understanding whether some MUPs, or
462 the extent of post-translational modifications, are particularly sensitive to the hormonal and/or behavioural
463 status of individuals could also provide very useful insight into potential functions.

464

465

466 [Insert Figure 7 here]

467

468

469 Most strikingly, our analysis further revealed the complexity of the post-translational modifications that are
470 applied to rat MUPs, including phosphorylation of MUP1 and protein trimming of MUP8 (summarized in

471 Figure 7). Neither of these modifications is evident in the best studied MUP system in the mouse, *Mus*
472 *musculus domesticus* and it is possible that the rat relies on post-translational modification to elicit further
473 variance in semiochemical properties, but confirmation of this must await functional bioassay in behavioural
474 tests. Additionally, our study emphasizes the need for detailed protein analysis to identify individual
475 proteoforms prior to functional characterization.
476

477 MATERIALS AND METHODS

478 *Animals and urine collection:* Laboratory rat urine donors were 11 Wistar Han® outbred rats and 11 Brown
479 Norway BN/RijHsd inbred rats obtained from Harlan UK (now Envigo) at 3 weeks of age. Animals were
480 housed in GPR2 cages (56 x 38 x 25 cm, North Kent Plastics, UK) on Corn Cob Absorb 10/14 substrate (IPS
481 Product Supplies Ltd, London). Water and food (lab diet 5002, Purina Mills) were given *ad libitum.* All rats
482 were provided with paper wool nesting material, cardboard houses and plastic tubes (8 cm diameter) for
483 home cage enrichment. Urine samples were obtained from adult rats aged 3 to 9 months. For urine
484 collection from laboratory strains, individual rats were placed in a clean empty wire-floored polypropylene
485 RC2R cage (56×38×22cm) without food or water. The cages were suspended over trays (checked every 30
486 min) into which the urine could collect. After 2-4 h rats were returned to their home cage. Adult wild rat
487 samples were provided by the former Central Science Laboratory of Defra (now part of the Animal and Plant
488 Health Agency, UK) from rats that were trapped on farms within 15 miles of the Central Science Laboratory
489 (Sand Hutton, North Yorkshire). Wild-caught animals were individually housed in suspended wire cages with
490 free access to food and water. Urine samples were collected overnight on a clean waxed paper sheet in the
491 tray under the cage. All samples were aspirated by pipette, avoiding feces and food fragments, and stored
492 at -20 °C until use.

493

494 *Protein and creatinine concentration assays*: Protein concentration was determined using the Coomassie
495 Protein Plus assay kit (Thermo Scientific). Urinary creatinine was quantified using a creatinine assay kit
496 (Sigma-Aldrich).

497

498 *Electrospray ionization mass spectrometry (ESI-MS) of intact proteins*: Urine samples were diluted in 0.1%
499 (v/v) formic acid and centrifuged at 13,000 g for 10 min. All analyses were performed on a Synapt G2 mass
500 spectrometer (Waters Corporation), fitted with an API source. Samples were desalted and concentrated on
501 a C4 reverse phase trap (Thermo Scientific) and protein was eluted at a flow rate of 10 µL/min using three
502 repeated 0–100 % (v/v) acetronitile (ACN) gradients. Data was collected between 800 and 3500 Th (m/z),
503 processed and transformed to a neutral average mass using MaxEnt 1 (Maximum Entropy Software, Waters
504 Corporation). The instrument was calibrated using a 2 pmol injection of myoglobin from equine heart
505 (Sigma-Aldrich; M1882).

506

507 *Polyacrylamide gel electrophoresis (PAGE):* SDS-PAGE was performed as described by Laemmli
508 [64]. Samples were resuspended in 2x SDS sample buffer (125 mM Tris-HCl; 140 mM SDS; 20%
509 (v/v) glycerol; 200 mM DTT and 30 mM bromophenol blue) and heated at 95 °C for 5 min before
510 loading onto gel. Electrophoresis was set at a 200 V constant potential for 45 min through a 4 %
511 (w/v) stacking gel followed by a 15 % (w/v) resolving polyacrylamide gel. PAGE under native

512 conditions was performed following the same protocol but in the absence of SDS and DTT during
513 the process. Electrophoresis was set at a 200 V constant potential for 60 min. Protein bands were
514 visualized with Coomassie Brilliant Blue stain (Sigma-Aldrich).
515
516 *Isoelectric focusing (IEF):* IEF was performed using a Multiphor flatbed system (Amersham Biosciences)
517 using an Immobiline dry-plate gel, pH range 4-7 (GE Healhcare Life Sciences) and cooled to 10 °C. Urine
518 samples were concentrated and desalted using Vivaspin centrifugal concentrators (3 kDa MWCO,
519 Vivascience). Urine samples were diluted to 1 mg/mL with deionized water and 5 µL was applied to sample
520 strips placed on the gel. Samples were loaded into the gel at 200 V, 5 mA and 15 W for 200 V· h. The sample
521 strips were removed and the gel was run at 3500 V, 5 mA and 15 W for 14.8 kV· h. After fixation with 20 %
522 TCA (v/v), the gel was stained with Coomassie Brilliant Blue.
523
524 *Strong anion exchange chromatography (SAX):* Urine was desalted using Zeba columns (Pierce, 0.5 mL) and
525 then filtered through a 0.45 µm Millipore filter prior to injection. Proteins in rat urine were separated in
526 different fractions by high resolution strong anion exchange on an AKTA instrument equipped with a
527 Resource Q column (GE Life Sciences, V= 1 mL). The column was equilibrated with MES buffer (50 mM, pH
528 6), and bound proteins were eluted using a linear gradient of 0 to 1 M NaCl over 20 column volumes, with a
529 flow rate of 2 mL/min. Fractions were manually collected and analysed individually.
530
531 *In gel digestion*: Gel plugs were removed from the gel using a Pasteur glass pipette, placed into low binding
532 tubes and then destained using 50 µL of 50 mM ammonium bicarbonate/50 % (v/v) ACN for 30 min at 37 °C.
533 The plugs were then incubated with 10 mM dithiothreitol (DTT) for 60 min at 60 °C. The DTT was then
534 discarded and 55 mM iodoacetamide (IAM) stock solution was added to each tube and incubated for 45 min
535 at room temperature in the dark. After discarding the IAM, the plugs were washed twice using 50 mM
536 ammonium bicarbonate/50 % (v/v) ACN. The plugs were then dehydrated by adding 10 µL of 100 % ACN.
537 Sequencing grade endoproteinase Lys C (Wako) (diluted in 25 mM Tris-HCl, 1 mM EDTA, pH 8.5) was then
538 added and the digests incubated overnight at 37 °C. The reaction was stopped by adding formic acid
539 solution to a 1 % final concentration (v/v).
540
541 *Peptide mass fingerprinting (PMF)*: Peptide mixtures from the proteolytic digestion reactions were analysed
542 on a Bruker UltraFlex matrix-assisted laser-desorption ionization–time of flight-mass spectrometer (MALDI–
543 TOF) (Bruker Daltonics), operated in the reflectron mode with positive ion detection, or a MALDI Synapt G2
544 Si (Waters Corporation). Samples were mixed 1:1 (v/v) with a 10 mg/mL solution of α-cyano-4-
545 hydroxycinnamic acid in 60 % ACN/0.2 % TFA (v/v), before being spotted onto the MALDI target and air-
546 dried. Spectra were acquired at 35-40 % laser energy with 500-2000 laser shots per spectrum. Spectra were

547 gathered between m/z 900 and 4500. External mass calibration was performed using a mixture of des-Arg
548 bradykinin (904.47 Da), neurotensin (1672.92 Da), ACTH (corticotrophin, 2465.2 Da) and oxidized insulin â
549 chain (3495.9 Da) (2.4, 2.4, 2.6 and 30 pmol/µL, respectively) in 50 % ACN/0.1 % TFA (v/v).
550
551 *In solution digestion*: Liquid samples were denatured with RapiGest (Waters Corporation) and alkylated,
552 prior to digestion with trypsin or endopeptidase Lys C. To stop the proteolytic reaction and to inactivate and
553 precipitate the detergent, TFA (final concentration 0.5 % (v/v)) was added, followed by incubation for 45
554 min at 37 °C. To remove all insoluble material, samples were centrifuged twice at 13,000 g for 15 min [65].
555
556 *Liquid chromatography-tandem mass spectrometry analysis:* LC-MS/MS analysis was performed using a
557 QExactive instrument (Thermo Scientific) coupled to an Ultimate 3000 LC nano system (Thermo Scientific).
558 Protein digests were resolved on an Easy-spray PepMap RSLC C18 column over a linear gradient from 3 to
559 40% (v/v) ACN in 0.1% v/v formic acid. The QExactive instrument was operated in data dependent
560 acquisition mode. Full scan MS spectra (m/z 300-2000) were acquired at 70,000 resolution and the ten most
561 intense multiply charged ions (charge ≥ 2) were sequentially isolated and fragmented by high energy
562 collisional dissociation (HCD) at 30% standardized collision energy. Fragments ions were detected at 35,000
563 resolution and dynamic exclusion was set at 20 s. Proteome Discoverer (Thermo Scientific) version 1.4 was
564 used to generate peak lists using default parameters and Mascot version 2.4 (Matrix Science) to identify
565 peptides and proteins, using a database containing all the entries annotated for *Rattus norvegicus* in
566 Uniprot (www.uniprot.org) (updated on 20170605) and the sequences of the mature MUP proteins,
567 applying a FDR < 1 %. Either trypsin or Lys C was selected as the specific enzyme, allowing one missed
568 cleavage. MS/MS data were also analysed using Peaks Studio 8.0 (Bioinformatics solutions Inc.) to identify
569 post-translational modifications. All raw mass spectrometry files will be made immediately available upon
570 request.
571
572 *Data analysis:* Data were visualised and analysed using Aabel (Gigawiz software, http://www.gigawiz.com/)
573 and R (v.3.2) (http://www.R-project.org/). Protein maps were generated using PeptideMapper [66].
574

575 Legends to Figures

576 **Figure 1| The MUP gene cluster of the rat.**

577 Three iterations of the rat MUP gene cluster have been produced in different genome assemblies. Besides,
578 we show annotation reported by Logan et al [26] on v4. The gene identities between the assemblies are
579 indicated by grey lines. Protein coding genes are green arrows and pseudogenes are blue boxes. In green,
580 above each protein coding gene is the predicted mature mass for the protein in Da (corrected for signal
581 peptide cleavage and a single disulfide bond formation). In red, transcriptional supporting information
582 already available in the literature [43, 54, 58, 67], with the tissue of origin in brackets (pp: preputial gland).

583

584 **Figure 2| Protein expression in the urine of male and female rats.**

585 Urine samples were recovered from male and female rats of two different laboratory strains (Wistar Han,
586 WH; Brown Norway, BN) and wild caught individuals. A: Protein output was expressed as mg protein/mg
587 creatinine to correct for urine dilution. B: Urine samples were also analysed by SDS-PAGE. Proteins
588 identified by in-gel digestion followed by PMF and tandem mass spectrometry are labelled and described in
589 the text.

590

591 **Figure 3 | Intact mass protein profiling of male rat urine.**

592 Urine samples from Wistar Han (Panel A) or Brown Norway (Panel B) male rats were analysed by ESI-MS to
593 obtain the profile of protein masses, here focused on 18,000 Da to 19,000 Da. Each spectrum is an average
594 spectrum from 10 individual animal/urine replicates. Full data are presented in Supplementary Figure 2.

595

596 **Figure 4 | Intact mass protein profiling in wild male rat urine.**

597 Urine samples from male wild caught rats were analysed by ESI-MS intact protein mass profiling (Panel A)
598 and by isoelectric focusing (Panel B).

599

600 **Figure 5| Spearman correlation coefficient analysis of intact mass areas.**

601 The relative amount of each isoform was quantified based on the peak area of the deconvoluted spectra.
602 Spearman correlation coefficient was calculated between the amounts of each protein in different
603 individuals. High correlation was found in laboratory strains whereas weak correlation was found in wild
604 individuals, suggesting the possibility of quantitative polymorphism and a higher degree of variance than in
605 the laboratory strains.

606

607 **Figure 6 | Evidence of phosphorylation of specific rat MUPs.**

608 Manual annotation of the N-terminal peptide of MUP1 showing evidence for phosphorylation of a serine
609 residue at position 4.

610

611 Figure 7 | Summary of phenotypic profiling of rat urinary MUPs.

612 The phenotypic analysis of urinary MUPs is mapped to the *mup* region of rat genome version 4. Above the
613 gene annotation we include transcriptional evidence from other studies, highlighting tissue of origin. Below
614 the gene annotation, we summarise the findings in this paper. For each protein, we report the evidence for
615 a mature gene product by intact mass analysis (adjacent to the predicted mass in orange) and from bottom
616 up peptide analysis (adjacent to the sequence data) for each of the three groups of animals tested: Wistar
617 Han (WH), Brown Norway (BN) and wild-caught (Wild) individuals. A green circle defines confident protein-
618 level evidence, a red circle denotes the absence of evidence for this particular gene product.

619

620 Supplementary Figure 1| Workflow followed to analyze rat urine samples.

621

622 Supplementary Figure 2| ESI-MS intact mass deconvoluted spectra from individual Wistar Han and
623 Brown Norway males.

624

625 Supplementary Figure 3| ESI-MS analysis of female rat urine.

626 Deconvolution of the spectrum estimates two masses of about 11 kDa (11065 and 11450 Da) likely
627 corresponding to the rat urinary proteins 1 and 2.

628

629 Supplementary Figure 4| Sequencing of MUP isoforms from Wistar Han males by native
630 electrophoresis of urine followed by in-gel LysC digestion and analysis by PMF and LC-MS/MS.

631 Peptide maps show sequence coverage. Red boxes show unique peptides for the isoform and blue boxes
632 show common peptides to several MUP isoforms.

633

634 Supplementary Figure 5| Sequencing of MUP isoforms from Brown Norway males by native
635 electrophoresis of urine followed by in-gel LysC digestion and analysis by PMF and LC-MS/MS.

636 Peptide maps show sequence coverage. Red boxes show unique peptides for the isoform and blue boxes
637 show common peptides to several MUP isoforms.

638

639 Supplementary Figure 6| Sequencing of MUP isoforms from wild caught males by native
640 electrophoresis of urine followed by in-gel LysC digestion and analysis by PMF and LC-MS/MS.

641 Peptide maps show sequence coverage. Red boxes show unique peptides for the isoform and blue boxes
642 show common peptides to several MUP isoforms.

643

644 Supplementary Figure 7| Sequencing of MUP isoforms from Wistar Han males by ion exchange
645 chromatography fractionation followed by LysC digestion and analysis by LC-MS/MS.

646 ESI-MS intact mass deconvoluted spectra from individual Wistar Han are shown for each fraction.

647

648 Supplementary Figure 8| Sequencing of MUP isoforms from Brown Norway males by ion exchange
649 chromatography fractionation followed by LysC digestion and analysis by LC-MS/MS.
650 ESI-MS intact mass deconvoluted spectra from individual Brown Norway are shown for each fraction.

651

652 Supplementary Figure 9| Network representation of a comparison of the expected trypsin peptides
653 from protein sequences of the predicted rat MUP isoforms, highlighting unique peptides for each isoform.
654 Peptide mapper [66] was used to perform in-silico digestion of protein sequences and network was built
655 using Cytoscape [68].

656

Acknowledgements

This work is supported by the Biotechnology and Biological Sciences Research (BBSRC) programme BB/J002631/1. We are grateful to Dr. A D MacNicoll for supplying the wild rat samples. The authors gratefully acknowledge instrumental support from Dr. Philip Brownbridge (Centre of Proteome Research), Dr. Richard Humphries, Amanda J. Davidson, John Waters, Rachel Spencer and Joshua Beeston (Mammalian Behaviour and Evolution Group).

665 References

666 1. Finlayson JS, Potter M, Runner CR. Electrophoretic Variation and Sex Dimorphism of the Major Urinary
667 Protein Complex in Inbred Mice: A New Genetic Marker. Journal of the National Cancer Institute. 1963;31:91-
668 107. Epub 1963/07/01. PubMed PMID: 14043041.
669 2. Roy AK, Neuhaus OW, Gardner E. Studies on Rat Urinary Proteins. Federation proceedings.
670 1965;24(2p1):507-&. PubMed PMID: ISI:A19656257802090.
671 3. Finlayson JS, Morris HP. Molecular Size of Rat Urinary Protein. Proceedings of the Society for
672 Experimental Biology and Medicine Society for Experimental Biology and Medicine. 1965;119:663-6. Epub
673 1965/07/01. PubMed PMID: 14328970.
674 4. Robertson DH, Cox KA, Gaskell SJ, Evershed RP, Beynon RJ. Molecular heterogeneity in the Major Urinary
675 Proteins of the house mouse Mus musculus. Biochem J. 1996;316 (Pt 1):265-72. Epub 1996/05/15. PubMed
676 PMID: 8645216; PubMed Central PMCID: PMC1217333.
677 5. Robertson DH, Hurst JL, Bolgar MS, Gaskell SJ, Beynon RJ. Molecular heterogeneity of urinary proteins in
678 wild house mouse populations. Rapid communications in mass spectrometry : RCM. 1997;11(7):786-90. Epub
679 1997/01/01. doi: 10.1002/(SICI)1097-0231(19970422)11:7<786::AID-RCM876>3.0.CO;2-8. PubMed PMID:
680 9161047.
681 6. Hurst JL, Beynon RJ, Armstrong SD, Davidson AJ, Roberts SA, Gómez-Baena G, et al. Molecular
682 heterogeneity in major urinary proteins of Mus musculus subspecies: potential candidates involved in speciation.
683 Scientific reports. 2017;7:44992. doi: 10.1038/srep44992. PubMed PMID: 28337988; PubMed Central PMCID:
684 PMCPMC5364487.
685 7. Hurst JL, Robertson DHL, Tolladay U, Beynon RJ. Proteins in urine scent marks of male house mice
686 extend the longevity of olfactory signals. Animal behaviour. 1998;55(5):1289-97. Epub 1998/12/16. PubMed
687 PMID: 9632512.
688 8. Beynon RJ, Hurst JL. Urinary proteins and the modulation of chemical scents in mice and rats. Peptides.
689 2004;25(9):1553-63. Epub 2004/09/18. doi: 10.1016/j.peptides.2003.12.025. PubMed PMID: 15374657.
690 9. Kwak J, Strasser E, Luzynski K, Thoss M, Penn DJ. Are MUPs a Toxic Waste Disposal System? PLoS One.
691 2016;11(3):e0151474. doi: 10.1371/journal.pone.0151474. PubMed PMID: 26966901; PubMed Central PMCID:
692 PMCPMC4788440.
693 10. Hurst JL, Payne CE, Nevison CM, Marie AD, Humphries RE, Robertson DH, et al. Individual recognition in
694 mice mediated by major urinary proteins. Nature. 2001;414(6864):631-4. Epub 2001/12/12. doi:
695 10.1038/414631a. PubMed PMID: 11740558.
696 11. Cheetham SA, Thom MD, Jury F, Ollier WE, Beynon RJ, Hurst JL. The genetic basis of individual-
697 recognition signals in the mouse. Current biology : CB. 2007;17(20):1771-7. Epub 2007/10/24. doi:
698 10.1016/j.cub.2007.10.007. PubMed PMID: 17949982.
699 12. Green JP, Holmes AM, Davidson AJ, Paterson S, Stockley P, Beynon RJ, et al. The Genetic Basis of Kin
700 Recognition in a Cooperatively Breeding Mammal. Curr Biol. 2015;25(20):2631-41. doi:
701 10.1016/j.cub.2015.08.045. PubMed PMID: 26412134.
702 13. Roberts SA, Prescott MC, Davidson AJ, McLean L, Beynon RJ, Hurst JL. Individual odour signatures that
703 mice learn are shaped by involatile major urinary proteins (MUPs). BMC biology. 2018;16(1):48. doi:
704 10.1186/s12915-018-0512-9. PubMed PMID: 29703213; PubMed Central PMCID: PMCPMC5921788.
705 14. Kaur AW, Ackels T, Kuo TH, Cichy A, Dey S, Hays C, et al. Murine pheromone proteins constitute a
706 context-dependent combinatorial code governing multiple social behaviors. Cell. 2014;157(3):676-88. doi:
707 10.1016/j.cell.2014.02.025. PubMed PMID: 24766811.
708 15. Chamero P, Marton TF, Logan DW, Flanagan K, Cruz JR, Saghatelian A, et al. Identification of protein
709 pheromones that promote aggressive behaviour. Nature. 2007;450(7171):899-902. Epub 2007/12/08. doi:
710 10.1038/nature05997. PubMed PMID: 18064011.
711 16. Roberts SA, Simpson DM, Armstrong SD, Davidson AJ, Robertson DH, McLean L, et al. Darcin: a male
712 pheromone that stimulates female memory and sexual attraction to an individual male's odour. BMC biology.
713 2010;8:75. Epub 2010/06/08. doi: 10.1186/1741-7007-8-75. PubMed PMID: 20525243; PubMed Central PMCID:
714 PMC2890510.

715 17. Roberts SA, Davidson AJ, McLean L, Beynon RJ, Hurst JL. Pheromonal induction of spatial learning in
716 mice. Science. 2012;338(6113):1462-5. Epub 2012/12/15. doi: 10.1126/science.1225638. PubMed PMID:
717 23239735.
718 18. Armstrong SD, Robertson DH, Cheetham SA, Hurst JL, Beynon RJ. Structural and functional differences in
719 isoforms of mouse major urinary proteins: a male-specific protein that preferentially binds a male pheromone.
720 Biochem J. 2005;391(Pt 2):343-50. Epub 2005/06/07. doi: 10.1042/BJ20050404. PubMed PMID: 15934926;
721 PubMed Central PMCID: PMC1276933.
722 19. Phelan MM, McLean L, Armstrong SD, Hurst JL, Beynon RJ, Lian LY. The structure, stability and
723 pheromone binding of the male mouse protein sex pheromone darcin. PLoS One. 2014;9(10):e108415. doi:
724 10.1371/journal.pone.0108415. PubMed PMID: 25279835.
725 20. Berdoy M, Drickamer LC. Comparative social organization and life history of Rattus and Mus. Rodent
726 societies: An ecological and evolutionary perspective Chicago, Illinois: University of Chicago Press; 2007. p. 380-
727 92.
728 21. Vettorazzi A, Wait R, Nagy J, Monreal JI, Mantle P. Changes in male rat urinary protein profile during
729 puberty: a pilot study. BMC Res Notes. 2013;6(1):232. Epub 2013/06/19. doi: 10.1186/1756-0500-6-232.
730 PubMed PMID: 23767887; PubMed Central PMCID: PMC3751546.
731 22. Guo X, Guo H, Zhao L, Zhang YH, Zhang JX. Two predominant MUPs, OBP3 and MUP13, are male
732 pheromones in rats. Front Zool. 2018;15:6. doi: 10.1186/s12983-018-0254-0. PubMed PMID: 29483934; PubMed
733 Central PMCID: PMCPMC5824612.
734 23. Kumar V, Vasudevan A, Soh LJ, Le Min C, Vyas A, Zewail-Foote M, et al. Sexual attractiveness in male rats
735 is associated with greater concentration of major urinary proteins. Biology of reproduction. 2014;91(6):150. doi:
736 10.1095/biolreprod.114.117903. PubMed PMID: 25359898.
737 24. Mudge JM, Armstrong SD, McLaren K, Beynon RJ, Hurst JL, Nicholson C, et al. Dynamic instability of the
738 major urinary protein gene family revealed by genomic and phenotypic comparisons between C57 and 129 strain
739 mice. Genome Biol. 2008;9(5):R91. Epub 2008/05/30. doi: 10.1186/gb-2008-9-5-r91. PubMed PMID: 18507838;
740 PubMed Central PMCID: PMC2441477.
741 25. Sheehan MJ, Lee V, Corbett-Detig R, Bi K, Beynon RJ, Hurst JL, et al. Selection on Coding and Regulatory
742 Variation Maintains Individuality in Major Urinary Protein Scent Marks in Wild Mice. Plos Genet.
743 2016;12(3):e1005891. doi: 10.1371/journal.pgen.1005891. PubMed PMID: 26938775; PubMed Central PMCID:
744 PMCPMC4777540.
745 26. Logan DW, Marton TF, Stowers L. Species specificity in major urinary proteins by parallel evolution. PLoS
746 One. 2008;3(9):e3280. Epub 2008/09/26. doi: 10.1371/journal.pone.0003280. PubMed PMID: 18815613;
747 PubMed Central PMCID: PMC2533699.
748 27. Hancock JM. A bigger mouse? The rat genome unveiled. BioEssays : news and reviews in molecular,
749 cellular and developmental biology. 2004;26(10):1039-42. Epub 2004/09/24. doi: 10.1002/bies.20121. PubMed
750 PMID: 15382132.
751 28. Roy AK, Neuhaus OW. Identification of rat urinary proteins by zone and immunoelectrophoresis.
752 Proceedings of the Society for Experimental Biology and Medicine Society for Experimental Biology and Medicine.
753 1966;121(3):894-9. Epub 1966/03/01. PubMed PMID: 4160706.
754 29. Kondo Y, Yamada J. Male urinary protein-1 (Mup-1) expression in the female rat. The Journal of heredity.
755 1983;74(4):280-2. Epub 1983/07/01. PubMed PMID: 6886376.
756 30. Vandoren G, Mertens B, Heyns W, Van Baelen H, Rombauts W, Verhoeven G. Different forms of alpha
757 2u-globulin in male and female rat urine. Eur J Biochem. 1983;134(1):175-81. Epub 1983/07/15. PubMed PMID:
758 6190651.
759 31. Aksu S, Tanrikulu F. Differentiation of protein species of alpha-2u-globulin according to database entries:
760 A half-theoretical approach. J Proteomics. 2016;134:186-92. doi: 10.1016/j.jprot.2015.12.024. PubMed PMID:
761 26746007.
762 32. Lee RS, Monigatti F, Lutchman M, Patterson T, Budnik B, Steen JA, et al. Temporal variations of the
763 postnatal rat urinary proteome as a reflection of systemic maturation. Proteomics. 2008;8(5):1097-112. Epub
764 2008/03/08. doi: 10.1002/pmic.200700701. PubMed PMID: 18324733.
765 33. Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, et al. Genome sequence of
766 the Brown Norway rat yields insights into mammalian evolution. Nature. 2004;428(6982):493-521. doi:
767 10.1038/nature02426. PubMed PMID: 15057822.

768 34. Bayard C, Holmquist L, Vesterberg O. Purification and identification of allergenic alpha (2u)-globulin
769 species of rat urine. Biochimica et biophysica acta. 1996;1290(2):129-34. Epub 1996/06/04. PubMed PMID:
770 8645715.
771 35. Bocskei Z, Groom CR, Flower DR, Wright CE, Phillips SE, Cavaggioni A, et al. Pheromone binding to two
772 rodent urinary proteins revealed by X-ray crystallography. Nature. 1992;360(6400):186-8. Epub 1992/11/12. doi:
773 10.1038/360186a0. PubMed PMID: 1279439.
774 36. Chaudhuri BN, Kleywegt GJ, Bjorkman J, Lehman-McKeeman LD, Oliver JD, Jones TA. The structures of
775 alpha 2u-globulin and its complex with a hyaline droplet inducer. Acta crystallographica Section D, Biological
776 crystallography. 1999;55(Pt 4):753-62. Epub 1999/03/25. PubMed PMID: 10089305.
777 37. Gómez-Baena G, Armstrong SD, Phelan MM, Hurst JL, Beynon RJ. The major urinary protein system in
778 the rat. Biochem Soc T. 2014;42:886-92. doi: 10.1042/Bst20140083. PubMed PMID: WOS:000340329200027.
779 38. Knopf JL, Gallagher JF, Held WA. Differential, multihormonal regulation of the mouse major urinary
780 protein gene family in the liver. Mol Cell Biol. 1983;3(12):2232-40. Epub 1983/12/01. PubMed PMID: 6656765;
781 PubMed Central PMCID: PMC370094.
782 39. Kuhn NJ, Woodworth-Gutai M, Gross KW, Held WA. Subfamilies of the mouse major urinary protein
783 (MUP) multi-gene family: sequence analysis of cDNA clones and differential regulation in the liver. Nucleic Acids
784 Res. 1984;12(15):6073-90. PubMed PMID: 6548015; PubMed Central PMCID: PMCPMC320058.
785 40. Kulkarni AB, Gubits RM, Feigelson P. Developmental and hormonal regulation of alpha 2u-globulin gene
786 transcription. Proc Natl Acad Sci U S A. 1985;82(9):2579-82. Epub 1985/05/01. PubMed PMID: 2581250; PubMed
787 Central PMCID: PMC397607.
788 41. MacInnes JI, Nozik ES, Kurtz DT. Tissue-specific expression of the rat alpha 2u globulin gene family. Mol
789 Cell Biol. 1986;6(10):3563-7. Epub 1986/10/01. PubMed PMID: 2432391; PubMed Central PMCID: PMC367109.
790 42. Murty CV, Mancini MA, Chatterjee B, Roy AK. Changes in transcriptional activity and matrix association
791 of alpha 2u-globulin gene family in the rat liver during maturation and aging. Biochim Biophys Acta.
792 1988;949(1):27-34. Epub 1988/01/25. PubMed PMID: 2446666.
793 43. Saito K, Nishikawa J, Imagawa M, Nishihara T, Matsuo M. Molecular evidence of complex tissue- and
794 sex-specific mRNA expression of the rat alpha(2u)-globulin multigene family. Biochem Biophys Res Commun.
795 2000;272(2):337-44. Epub 2000/06/02. doi: 10.1006/bbrc.2000.2694. PubMed PMID: 10833415.
796 44. Elliott BM, Ramasamy R, Stonard MD, Spragg SP. Electrophoretic variants of alpha 2u-globulin in the
797 livers of adult male rats: a possible polymorphism. Biochim Biophys Acta. 1986;870(1):135-40. Epub 1986/03/07.
798 PubMed PMID: 2418881.
799 45. Wait R, Gianazza E, Eberini I, Sironi L, Dunn MJ, Gemeiner M, et al. Proteins of rat serum, urine, and
800 cerebrospinal fluid: VI. Further protein identifications and interstrain comparison. Electrophoresis.
801 2001;22(14):3043-52. Epub 2001/09/22. doi: 10.1002/1522-2683(200108)22:14<3043::AID-ELPS3043>3.0.CO;2-
802 M. PubMed PMID: 11565799.
803 46. Payne CE, Malone N, Humphries RE, Bradbook C, Veggerby C, Beynon RJ, et al. Heterogeneity of major
804 urinary proteins in the house mice: population and sex differences. In: Marchelewska-Koj A, Muller-Schwarze D,
805 Lepri J, editors. Chemical Signals in Vertebrates. New York: Plenum Press; 2001. p. 233-40.
806 47. Beynon RJ, Veggerby C, Payne CE, Robertson DH, Gaskell SJ, Humphries RE, et al. Polymorphism in major
807 urinary proteins: molecular heterogeneity in a wild mouse population. Journal of chemical ecology.
808 2002;28(7):1429-46. Epub 2002/08/30. PubMed PMID: 12199505.
809 48. Thom MD, Stockley P, Jury F, Ollier WE, Beynon RJ, Hurst JL. The direct assessment of genetic
810 heterozygosity through scent in the mouse. Current biology : CB. 2008;18(8):619-23. Epub 2008/04/22. doi:
811 10.1016/j.cub.2008.03.056. PubMed PMID: 18424142.
812 49. Sherborne AL, Thom MD, Paterson S, Jury F, Ollier WE, Stockley P, et al. The genetic basis of inbreeding
813 avoidance in house mice. Current biology : CB. 2007;17(23):2061-6. Epub 2007/11/13. doi:
814 10.1016/j.cub.2007.10.041. PubMed PMID: 17997307; PubMed Central PMCID: PMC2148465.
815 50. Drickamer K, Kwoh TJ, Kurtz DT. Amino acid sequence of the precursor of rat liver alpha 2 micro-
816 globulin. The Journal of biological chemistry. 1981;256(8):3634-6. Epub 1981/04/25. PubMed PMID: 6163771.
817 51. Beynon RJ, Armstrong SD, Claydon AJ, Davidson AJ, Eyers CE, Langridge JI, et al. Mass spectrometry for
818 structural analysis and quantification of the Major Urinary Proteins of the house mouse. Int J Mass Spectrom.
819 2015;391:146-56. doi: 10.1016/j.ijms.2015.07.026. PubMed PMID: WOS:000367124800018.

820 52. Mertens B, Verhoeven G. Influence of neonatal androgenization on the expression of alpha 2u-globulin
821 in rat liver and submaxillary gland. J Steroid Biochem. 1985;23(5A):557-65. Epub 1985/11/01. PubMed PMID:
822 2417039.
823 53. Papes F, Logan DW, Stowers L. The vomeronasal organ mediates interspecies defensive behaviors
824 through detection of protein pheromone homologs. Cell. 2010;141(4):692-703. Epub 2010/05/19. doi:
825 10.1016/j.cell.2010.03.037. PubMed PMID: 20478258; PubMed Central PMCID: PMC2873972.
826 54. Ichiyoshi Y, Endo H, Yamamoto M. Length polymorphism in the 3' noncoding region of rat hepatic alpha
827 2u-globulin mRNAs. Biochimica et biophysica acta. 1987;910(1):43-51. Epub 1987/10/09. PubMed PMID:
828 2443176.
829 55. Gao F, Endo H, Yamamoto M. Length heterogeneity in rat salivary gland alpha 2 mu globulin mRNAs:
830 multiple splice-acceptors and polyadenylation sites. Nucleic Acids Res. 1989;17(12):4629-36. Epub 1989/06/26.
831 PubMed PMID: 2473438; PubMed Central PMCID: PMC318020.
832 56. Lobel D, Strotmann J, Jacob M, Breer H. Identification of a third rat odorant-binding protein (OBP3).
833 Chemical senses. 2001;26(6):673-80. Epub 2001/07/28. PubMed PMID: 11473933.
834 57. Saito H, Chi Q, Zhuang H, Matsunami H, Mainland JD. Odor coding by a Mammalian receptor repertoire.
835 Science signaling. 2009;2(60):ra9. Epub 2009/03/06. doi: 10.1126/scisignal.2000016. PubMed PMID: 19261596;
836 PubMed Central PMCID: PMC2774247.
837 58. Laperche Y, Lynch KR, Dolan KP, Feigelson P. Tissue-specific control of alpha 2u globulin gene
838 expression: constitutive synthesis in the submaxillary gland. Cell. 1983;32(2):453-60. Epub 1983/02/01. PubMed
839 PMID: 6186396.
840 59. Tagliabracci VS, Wiley SE, Guo X, Kinch LN, Durrant E, Wen J, et al. A Single Kinase Generates the
841 Majority of the Secreted Phosphoproteome. Cell. 2015;161(7):1619-32. doi: 10.1016/j.cell.2015.05.028. PubMed
842 PMID: 26091039; PubMed Central PMCID: PMCPMC4963185.
843 60. Rajkumar R, Ilayaraja R, Alagendran S, Archunan G, Maralidharan AR, Huang PH, et al. Characterization
844 of rat odorant binding protein variants and its post-translational modifications (PTMs):LC-MS/MS analyses of
845 protein Eluted from 2D-Polyacrylamide gel electrophoresis. Proteomics and Bioinformatics. 2011;4(10):210-7.
846 61. Brimau F, Cornard JP, Le Danvic C, Lagant P, Vergoten G, Grebert D, et al. Binding specificity of
847 recombinant odorant-binding protein isoforms is driven by phosphorylation. Journal of chemical ecology.
848 2010;36(8):801-13. doi: 10.1007/s10886-010-9820-4. PubMed PMID: 20589419.
849 62. Nielsen H. Predicting Secretory Proteins with SignalP. Methods Mol Biol. 2017;1611:59-73. doi:
850 10.1007/978-1-4939-7015-5_6. PubMed PMID: 28451972.
851 63. Beynon RJ, Oliver S, Robertson DH. Characterization of the soluble, secreted form of urinary meprin.
852 Biochem J. 1996;315 (Pt 2):461-5. PubMed PMID: 8615815; PubMed Central PMCID: PMCPMC1217218.
853 64. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4.
854 Nature. 1970;227(5259):680-5. PubMed PMID: 5432063.
855 65. Hammond DE, Claydon AJ, Simpson DM, Edward D, Stockley P, Hurst JL, et al. Proteome Dynamics:
856 Tissue Variation in the Kinetics of Proteostasis in Intact Animals. Mol Cell Proteomics. 2016;15(4):1204-19. doi:
857 10.1074/mcp.M115.053488. PubMed PMID: 26839000; PubMed Central PMCID: PMCPMC4824850.
858 66. Beynon RJ. A simple tool for drawing proteolytic peptide maps. Bioinformatics. 2005;21(5):674-5.
859 PubMed PMID: 15539446.
860 67. Strausberg RL, Feingold EA, Grouse LH, Derge JG, Klausner RD, Collins FS, et al. Generation and initial
861 analysis of more than 15,000 full-length human and mouse cDNA sequences. Proc Natl Acad Sci U S A.
862 2002;99(26):16899-903. doi: 10.1073/pnas.242603899. PubMed PMID: 12477932; PubMed Central PMCID:
863 PMCPMC139241.
864 68. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment
865 for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498-504. doi:
866 10.1101/gr.1239303. PubMed PMID: 14597658; PubMed Central PMCID: PMCPMC403769.
867

Figure 1

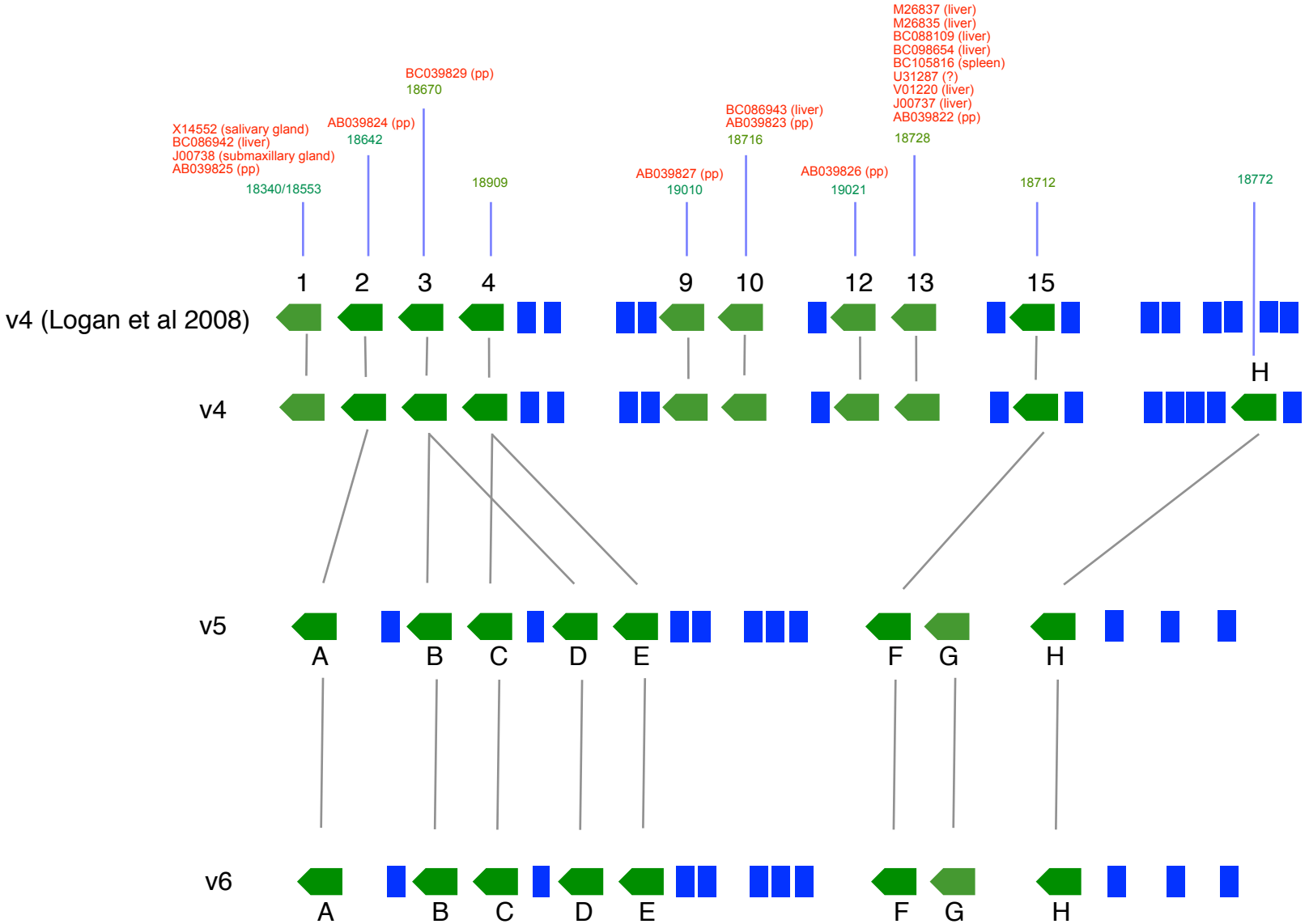
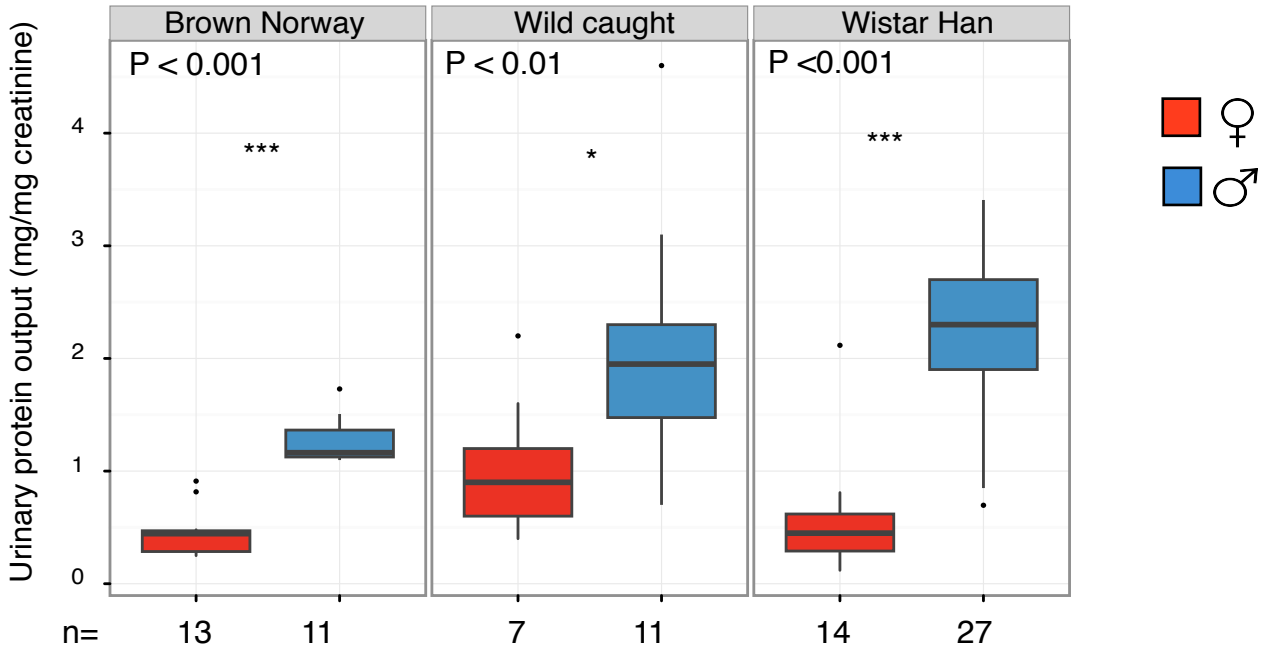

X14552 (salivary gland)
BC086942 (liver)
J00738 (submaxillary gland)
AB039825 (pp)
18340/18553
AB039824 (pp)
18642
BC039829 (pp)
18670
18909
AB039827 (pp)
19010
BC086943 (liver)
AB039823 (pp)
18716
AB039826 (pp)
19021
M26837 (liver)
M26835 (liver)
BC088109 (liver)
BC098654 (liver)
BC105816 (spleen)
U31287 (?)
V01220 (liver)
J00737 (liver)
AB039822 (pp)
18728
18712
18772
1 2 3 4 9 10 12 13 15
v4 (Logan et al 2008)
v4
H
v5
A B C D E F G H
v6
A B C D E F G H

Figure 2

A


Brown Norway
Wild caught
Wistar Han
P < 0.001
P < 0.01
P <0.001
***
*
***
Urinary protein output (mg/mg creatinine)
0
1
2
3
4
n=
13
11
7
11
14
27
♀
♂

B

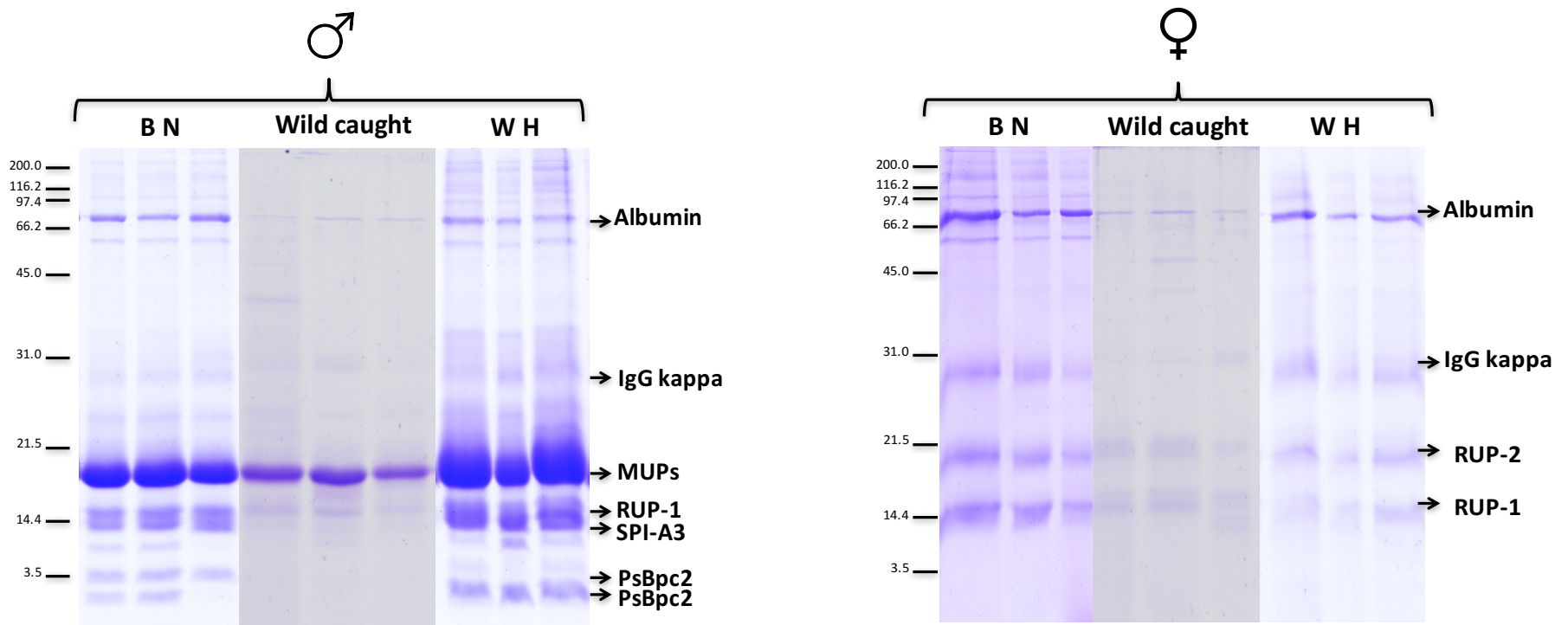

♂
B N
Wild caught
W H
200.0
116.2
97.4
66.2
45.0
31.0
21.5
14.4
3.5
Albumin
IgG kappa
MUPs
RUP-1
SPI-A3
PsBpc2
PsBpc2
♀
B N
Wild caught
W H
200.0
116.2
97.4
66.2
45.0
31.0
21.5
14.4
3.5
Albumin
IgG kappa
RUP-2
RUP-1

Figure 3

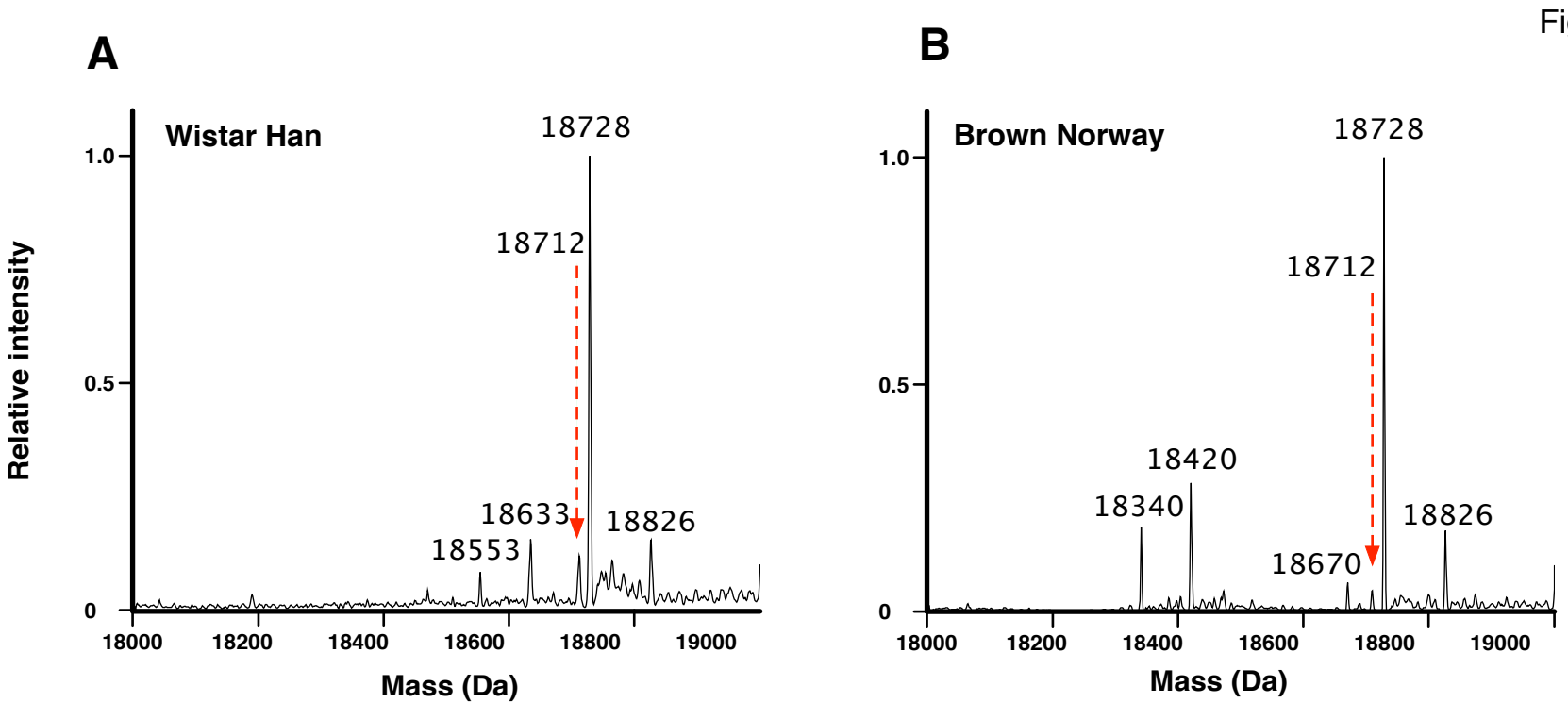

A
Wistar Han
18728
18712
18633
18553
18826
Relative intensity
1.0
0.5
0
18000
18200
18400
18600
18800
19000
Mass (Da)
B
Brown Norway
18728
18712
18420
18340
18670
18826
1.0
0.5
0
18000
18200
18400
18600
18800
19000
Mass (Da)

Figure 4

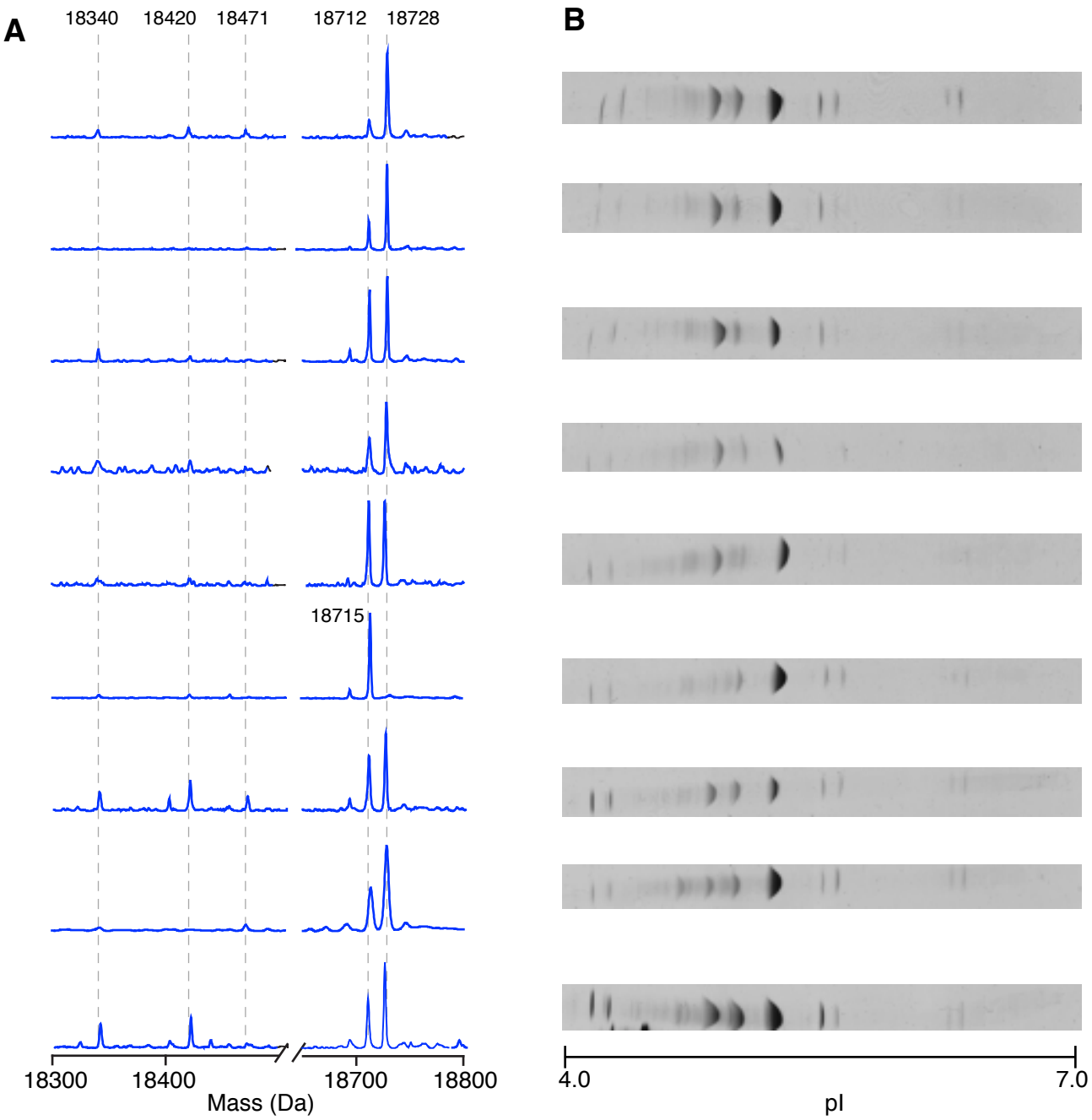

A
18340
18420
18471
18712
18728
18715
18300
18400
18700
18800
Mass (Da)
B
4.0
7.0
pI

Figure 5

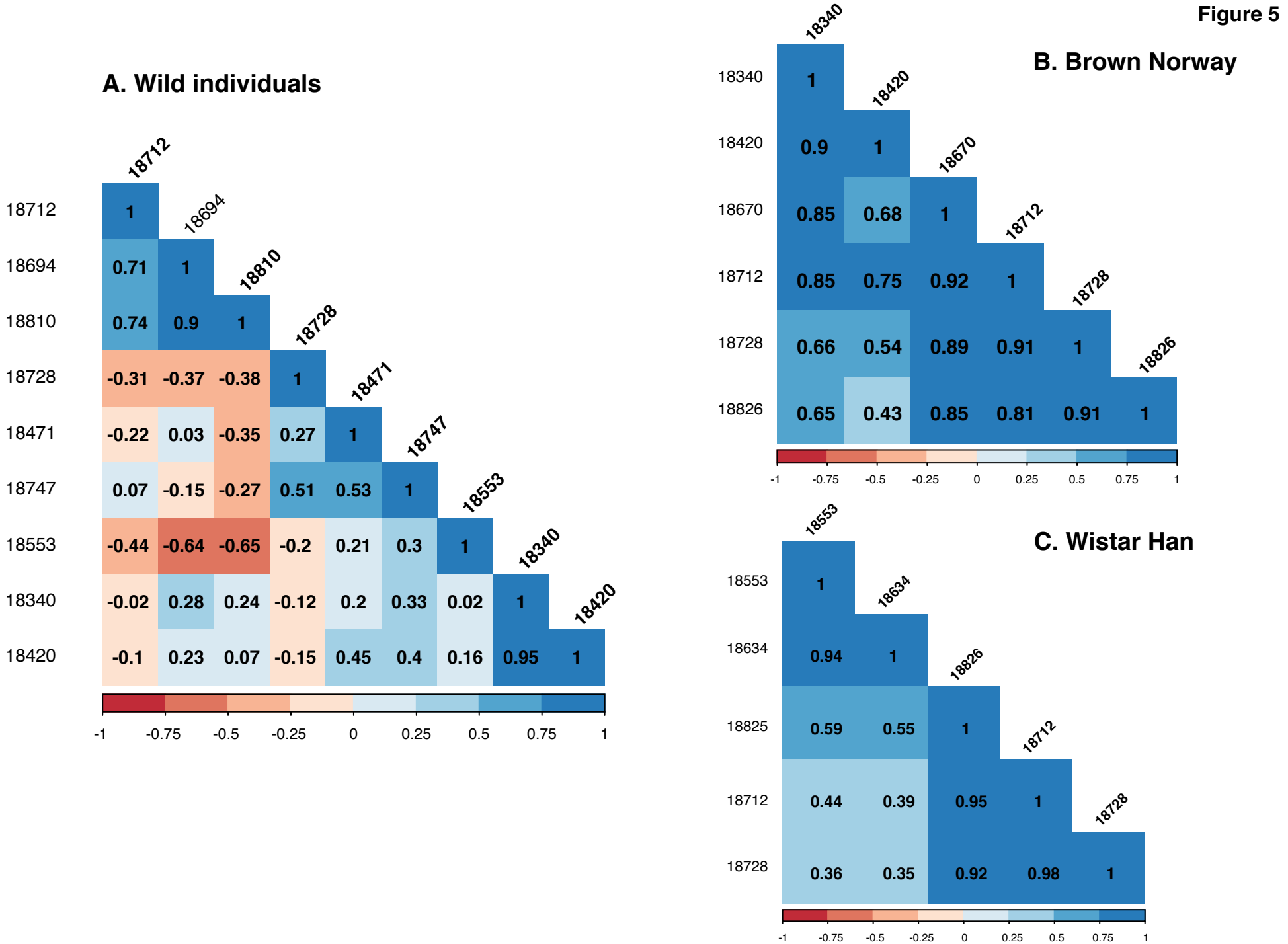
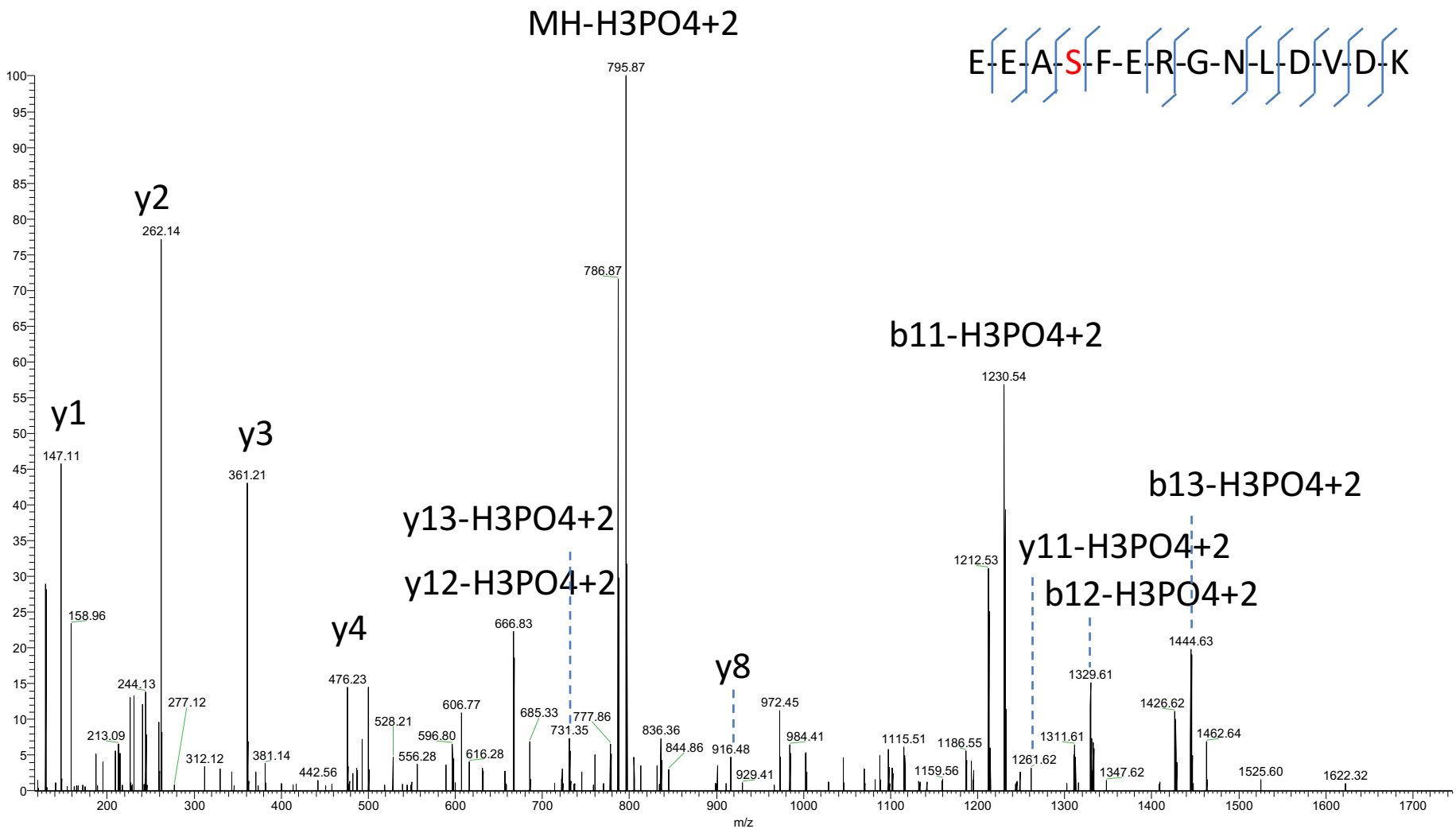

A. Wild individuals
18712
18694
18810
18728
18471
18747
18553
18340
18420
1
0.71 1
0.74 0.9 1
-0.31 -0.37 -0.38 1
-0.22 0.03 -0.35 0.27 1
0.07 -0.15 -0.27 0.51 0.53 1
-0.44 -0.64 -0.65 -0.2 0.21 0.3 1
-0.02 0.28 0.24 -0.12 0.2 0.33 0.02 1
-0.1 0.23 0.07 -0.15 0.45 0.4 0.16 0.95 1
-1 -0.75 -0.5 -0.25 0 0.25 0.5 0.75 1
B. Brown Norway
18340
18420
18670
18712
18728
18826
1
0.9 1
0.85 0.68 1
0.85 0.75 0.92 1
0.66 0.54 0.89 0.91 1
0.65 0.43 0.85 0.81 0.91 1
-1 -0.75 -0.5 -0.25 0 0.25 0.5 0.75 1
C. Wistar Han
18553
18634
18825
18712
18728
1
0.94 1
0.59 0.55 1
0.44 0.39 0.95 1
0.36 0.35 0.92 0.98 1
-1 -0.75 -0.5 -0.25 0 0.25 0.5 0.75 1

Figure 6


MH-H3PO4+2
E-E-A-S-F-E-R-G-N-L-D-V-D-K
y2
y1
y3
y4
y13-H3PO4+2
y12-H3PO4+2
y8
b11-H3PO4+2
b13-H3PO4+2
y11-H3PO4+2
b12-H3PO4+2
795.87
786.87
262.14
147.11
361.21
158.96
244.13
277.12
213.09
312.12
381.14
442.56
476.23
528.21
556.28
596.80
606.77
616.28
666.83
685.33
731.35
777.86
836.36
844.86
916.48
929.41
972.45
984.41
1115.51
1159.56
1186.55
1212.53
1230.54
1261.62
1311.61
1329.61
1347.62
1426.62
1444.63
1462.64
1525.60
1622.32
m/z

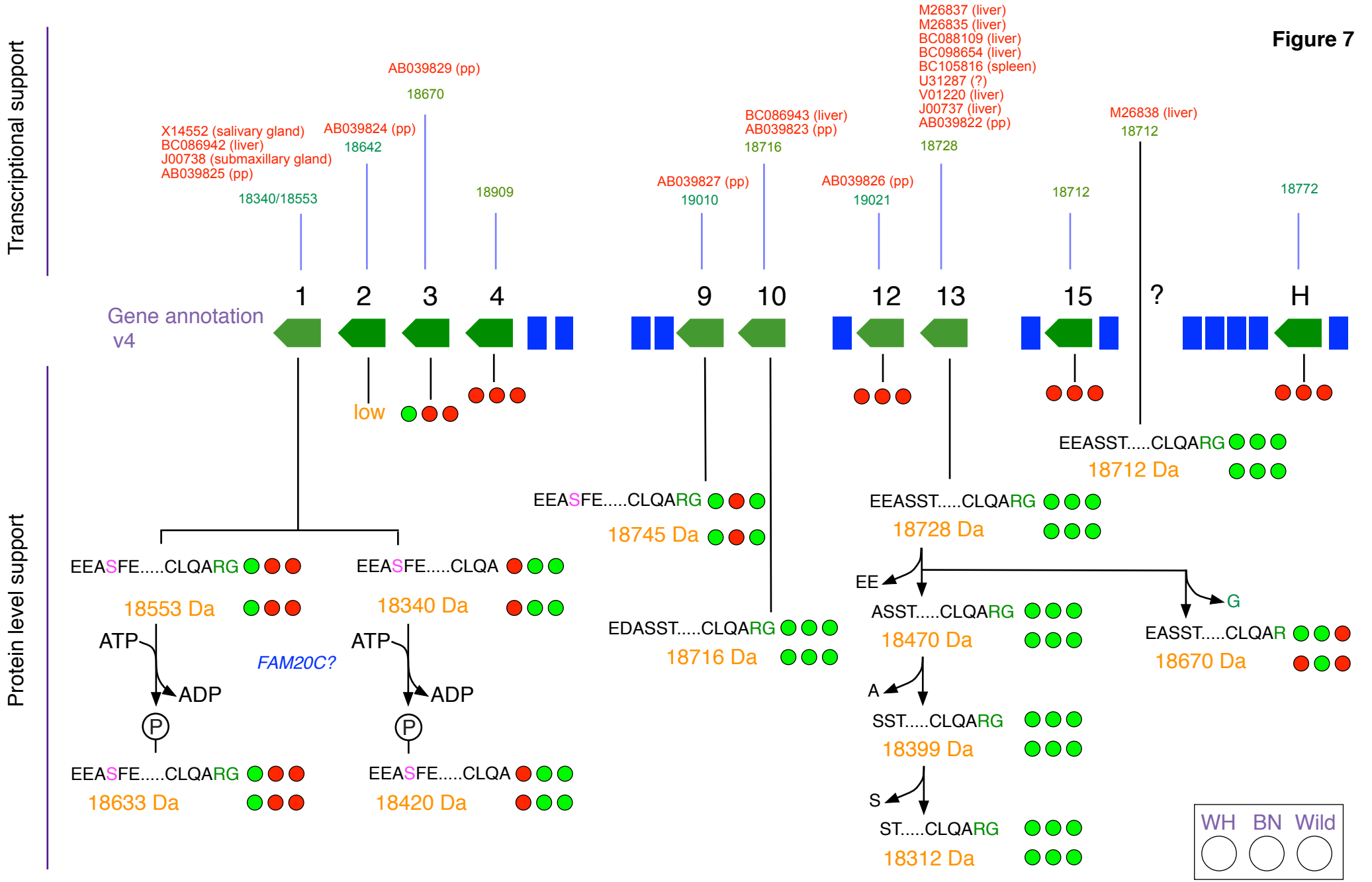


Figure 7
Transcriptional support
Protein level support
X14552 (salivary gland)
BC086942 (liver)
J00738 (submaxillary gland)
AB039825 (pp)
18340/18553
AB039824 (pp)
18642
AB039829 (pp)
18670
18909
AB039827 (pp)
19010
BC086943 (liver)
AB039823 (pp)
18716
AB039826 (pp)
19021
M26837 (liver)
M26835 (liver)
BC088109 (liver)
BC098654 (liver)
BC105816 (spleen)
U31287 (?)
V01220 (liver)
J00737 (liver)
AB039822 (pp)
18728
18712
M26838 (liver)
18712
18772
Gene annotation v4
1
2
3
4
9
10
12
13
15
?
H
low
EEASFE.....CLQARG
18553 Da
EEASFE.....CLQA
18340 Da
ATP
ADP
FAM20C?
P
EEASFE.....CLQARG
18633 Da
EEASFE.....CLQA
18420 Da
EEASFE.....CLQARG
18745 Da
EDASST.....CLQARG
18716 Da
EEASST.....CLQARG
18728 Da
EE
ASST.....CLQARG
18470 Da
A
SST.....CLQARG
18399 Da
S
ST.....CLQARG
18312 Da
G
EASST.....CLQAR
18670 Da
EEASST.....CLQARG
18712 Da
WH
BN
Wild

MUP gene	RGD genome Annotation	Predicted mature mass (Da)	UniParc	Uniprot accession	Uniprot names	Transcript supportive information
MUP 1	v4	18340	UPI000017083A	Q78E14	Obp3 protein; Rat salivary gland (alpha)2(mu)globulin, type 1	Liver (BC086942) Salivary gland (X14552)
	v4	18553	UPI00000E8911	Q63213	Alpha-2u globulin; PGCL4	Submaxillary gland (J00738) Preputial gland (AB039825)
MUP 2	v4/v5/v6	18642	UPI00000E7901	Q9JJI3	Alpha-2u globulin; PGCL3	Preputial gland (AB039824)
MUP 3	v4 and 2 loci in v5/v6	18670	UPI00000E7381	Q9JJH9	Alpha-2u globulin; Protein Mup4; PGCL8	Preputial gland (AB039829)
MUP 4	v4 and 2 loci in v5/v6	18909	UPI0000506C83	A0A096MK41	Uncharacterized protein	
MUP 9	v4	19010				Preputial gland (AB039827, sequence conflict K79 to E79, Uniparc UPI00000E7BD9; Q9JJI1)
MUP10	v4	18716	UPI00000E7542	Q8K1Q6	Alpha-2u globulin; PGCL2	Liver (BC086943) Preputial gland (AB039823; sequence conflict in signal peptide, UniParc UPI00000E8694, Q9JJI4)
MUP 12	v4	19021	UPI00000E5BE6	Q9JJI2	Alpha-2u globulin; PGCL5	Preputial gland (AB039826)
MUP 13	v4	18728	UPI000000086C	P02761	Major urinary protein (MUP_RAT); PGCL1; Allergen rat n1; Alpha-2u globulin PGCL1	Liver (M26835); Liver (M26837); Preputial gland (AB039822); Liver (BC088109); Liver (BC098654); Spleen (BC105816); U31287; Liver (V01220); Liver (J00737)
MUP 15	v4 (full) in v5/v6 (fragments)	18712				
MUP H	v4/v5/v6	18772	UPI00005035E7	MOR620	Major urinary protein like	
Transcript AB039828	No annotation	18822	UPI00000E6465	Q9JJI0	Alpha-2u globulin; PGCL7	preputial gland
Transcript M26836	No annotation	18726	UPI00000E6420	Q63024	Rat alpha-2u-globulin (L type)	Liver
Transcript M26838	No annotation	18712	UPI00000E6E81	Q63025	Rat alpha-2u-globulin (S type)	Liver