



OPEN

## Variation in mouse chemical signals is genetically controlled and environmentally modulated

Romana Stopková<sup>1</sup>, Tereza Matějková<sup>1</sup>, Alica Dodoková<sup>1</sup>, Pavel Talacko<sup>1</sup>, Petr Zacek<sup>1</sup>, Radislav Sedlacek<sup>2</sup>, Jaroslav Piálek<sup>3</sup> & Pavel Stopka<sup>1</sup>✉

In most mammals and particularly in mice, chemical communication relies on the detection of ethologically relevant fitness-related cues from other individuals. In mice, urine is the primary source of these signals, so we employed proteomics and metabolomics to identify key components of chemical signalling. We show that there is a correspondence between urinary volatiles and proteins in the representation of genetic background, sex and environment in two house mouse subspecies *Mus musculus musculus* and *M. m. domesticus*. We found that environment has a strong influence upon proteomic and metabolomic variation and that volatile mixtures better represent males while females have surprisingly more sex-biased proteins. Using machine learning and combined-omics techniques, we identified mixtures of metabolites and proteins that are associated with biological features.

All living organisms are constantly confronted with chemical signals from their environment and from other individuals. In mice, these signals often act upon innate<sup>1</sup> or learned<sup>2</sup> representations in the brain and yield behavioural responses that promote survival and fitness. For example, a male mouse will likely produce signals to advertise his fitness that would lead to avoidance behaviour in other males and to sexual attraction in females<sup>3–5</sup>. Some of these signals are species-specific or sub-specific, and used for peer-recognition<sup>6,7</sup>. Moreover, any individual, regardless of sex, will follow a cue that represents a favourite food item or avoid a cue indicating predators<sup>8</sup>. Most behaviour studies focus on the effect of a single or few compounds and proteins as signalling molecules. However, animals and their surrounding environments are more complex and instead of single or several studied compounds, most organisms also including bacteria<sup>9</sup> and plants<sup>10</sup> produce n-dimensional arrays of compounds. It is often the composition of these bouquets that induce behaviour and physiological responses in the receivers<sup>11</sup>. To make this puzzle even more complex, the response to the same cue might vary according to environmental factors. Thus, we asked whether biological traits such as sex and genetic background of an individual are manifested by proteomes or metabolomes and to what extent are these two sets linked or even correlated. This is important because sexual signals are known to prime sexually dimorphic circuits and striking sexually dimorphic sensory representations in the accessory olfactory bulb<sup>12</sup> and the medial amygdala<sup>13</sup>, but a comprehensive view on chemical signals that may trigger these representations has been lacking. In general, we were interested in how sexuality is displayed in an organism for which the fitness-related olfactory cues are more important than the visual ones.

Mouse urine contains large quantities and a variety of molecules that serve as olfactory signals. They are detectable by chemosensory receptors of the major olfactory epithelia and/or vomeronasal organ (VNO)<sup>14–22</sup>. These signals yield diverse physiological responses in the receiver<sup>13,23–31</sup> also when stimulated by selected non-volatile major urinary proteins (MUPs)<sup>32,33</sup>, short peptides<sup>34,35</sup>, and/or volatile organic compounds (VOCs)<sup>36–38</sup>. In mice, VOCs were considered as potent signals detectable by olfactory tissues<sup>18,39</sup> whilst MUPs were mostly considered as transporters of these signals in their eight-stranded beta-barrels<sup>37,40–47</sup> and thus shape the individual odour signatures<sup>48</sup>. However, various authors demonstrated that particular MUPs represent a signal on their own detectable by VNO<sup>32,49–51</sup> and that some of these molecules including male-biased MUP20 (known as darcin) elicit complex innate behaviours including aggression<sup>33</sup>, mate recognition<sup>52</sup> and learning<sup>32</sup>. However, since almost all previous studies concentrated only on MUPs, there is not yet any study showing the whole spectrum of urine proteins and volatiles that may also be involved in chemical communication, especially in wild living rodents.

<sup>1</sup>Department of Zoology, Faculty of Science, BIOCEV, Charles University, Vestec, Prague, Czech Republic. <sup>2</sup>Czech Centre for Phenogenomics, Institute of Molecular Genetics of the Czech Academy of Sciences, Vestec, Czech Republic. <sup>3</sup>Research Facility Studenec, Institute of Vertebrate Biology, Czech Academy of Sciences, Brno, Czech Republic. ✉email: pstopka@natur.cuni.cz

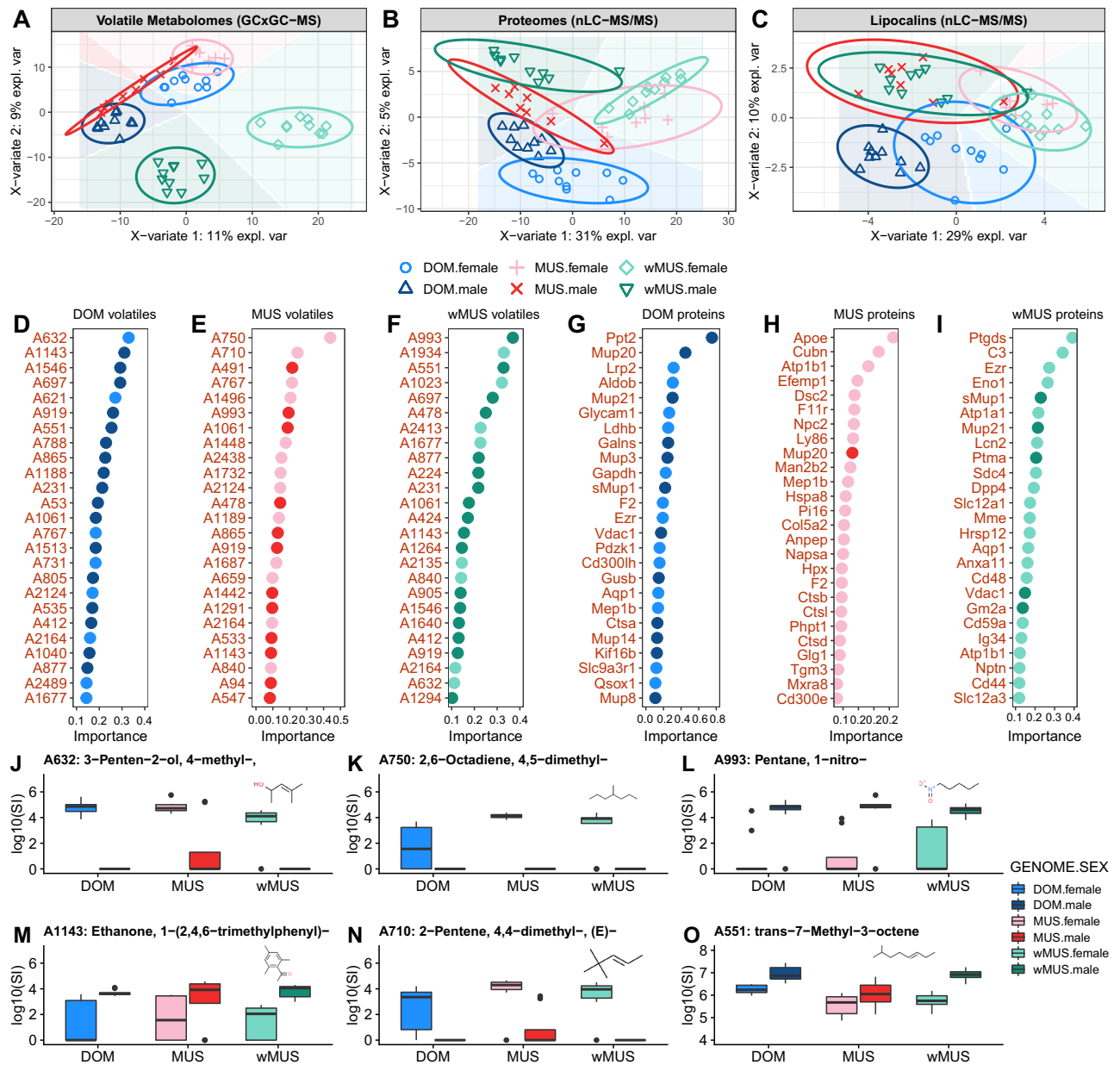
Owing to their importance for male sexual signalling, MUPs are highly abundant and male-biased in the mouse urine<sup>53–55</sup> where they protect and transport small volatile compounds in their eight-stranded beta barrel<sup>40,56</sup> and also delay their release<sup>57</sup>. Interestingly, MUP patterns and the level of sexual dimorphism are sub-species specific<sup>7,58,59</sup> making MUPs also important as candidate molecules in sub-species recognition<sup>60–62</sup>. Though females have less MUPs than males, these proteins are also involved in signalling of female sexual status because their concentration in the urine<sup>29</sup> and vaginal secretions<sup>63</sup> changes throughout the estrous cycle reaching the maxima during estrus. Similarly, social status affects the production of MUPs. This has been shown in wild derived *M. m. musculus* mice under laboratory conditions<sup>28</sup> and in seminatural enclosures, where males doubled the excretion of MUPs after acquiring a territory and became socially dominant<sup>64</sup>. MUP quantity constitutes of up to 85% (or even more) of all proteins in the urine<sup>65</sup>, and thus these proteins may have distracted the attention from several hundreds of other important proteins that are involved in various homeostatic, metabolic and signalling functions.

To explore whether there are subspecific and sex-specific differences in the house mouse, we collected urine samples from several wild-derived strains representing two subspecies, *M. m. musculus* (MUS, 7 strains) and *M. m. domesticus* (DOM, 9 strains). Importantly, both groups of strains were kept in the same breeding facility. We aimed to detect major components of chemical signalling on the two levels of resolution – metabolomic and proteomic. Thus, we generated volatile metabolomes with head-space two-dimensional comprehensive gas chromatography and mass spectrometry and aliquots of the same samples were used in parallel for the analysis of proteomes with nLC-MS/MS. Subspecific and sex-specific differences served as a proxy for evolutionary changes due to natural or sexual selection. However, resulting profiles may as well be influenced by the natural environment, so we surveyed additional samples from wild caught *M. m. musculus* (wMUS) mice. Taken together, we analysed complete proteomes and metabolomes from the three mouse groups of each sex to provide new insights into the mouse chemical signalling.

## Results

**Data manipulation.** The proteomic dataset contains a total of 958 protein identifications generated from 10 µl of urine of each sample and the resulting expression matrix was LFQ normalized (label-free quantification<sup>66</sup>). The same amount of urine (10 µl) was used to extract volatiles and the resulting data table contained a total of 2701 identifications based on unique masses and retention times. First, we reduced our datasets for singletons and doubletons such that only those molecules that were produced by at least three individuals from the same group (e.g. DOM males) were passed to further analyses. Final proteomic dataset thus contains 416 protein identifications. We did the same filtering for volatiles, however, volatile metabolomes are sensitive to false positives because the same molecules may naturally occur in animals but also in the air. To reduce the effect of false positives we removed all the molecules (i.e. rows) that were present only in blanks (i.e. samples of air from the labs where samples were processed). In the remaining set we have detected bimodal distribution, resulting from the mixture of two normal distributions. For these distributions overlap (see Methods), we calculated posterior p-value from normal-mixed models and if the p-value of belonging to blanks and samples was  $p < 0.05$  (i.e. corresponding  $FD < 7.1$ ) given rows were removed. This process corresponds to the Identity likelihood  $IL < 0.9$  (see methods). IL is a useful tool to visualize whether given volatiles are likely to be characteristic of the studied groups. This dataset finally contained a total of 875 molecules and the whole set was quantile normalized. More than 54% of metabolome components in this study are structurally related aliphatic aldehydes and alcohols. The length of the carbon backbone of these molecules are typically C6–C8. The most abundant molecule is 2-hexenal (33,8%). 2-hexenal is part of so called „green odour“ (GO), which is a mixture of eight aliphatic C6 aldehydes and C6 alcohols responsible for the smell of young leaves or freshly cut grass<sup>1</sup>. Several studies show high olfactory sensitivity of some mammals to GO<sup>2</sup>, including humans<sup>3</sup> and indicate antidepressant<sup>4</sup> and anxiolytic<sup>2</sup> effects on mice.

**Sources of variation: sex, subspecies, or environment?** To explore potential sources of variation in our data, we used Sparse Partial Least Squares Discriminant Analysis (sPLS-DA) for the fact that it has satisfying predictive performances in large datasets. In all the three comparisons in Fig. 1A–C, sex was the strongest factor influencing metabolome and proteome separation within the three mouse groups. We show in Fig. 1A that metabolomes of MUS and DOM of each sex are separated from each other. Since the MUS and DOM mice were kept for generations in the same conditions, this clear distinction between male vs. female metabolomes is most likely driven by genetic divergence between these two subspecies. However, so is separated wMUS of each sex from both the lab-bred MUS and DOM groups. This shows that urine metabolome of mouse is differentiated by subspecies, sex and environment. Thus, we predict that each component of the external environment such as food, microbiota<sup>67</sup>, plants, and naturally occurring air-borne molecules may have direct influence upon metabolomic profiles and metabolite processing in an individual. Proteomic data in Fig. 1B show that sex is again the major driver of separation (31% of explained variation, x-variate 1). Due to the genomic differences, separation of MUS males from DOM males is clear and this is also true for females. However, similarly as in metabolomes, wMUS males (wMUS.male vs Other(s), Comp 2 (y-axis): Area Under Curve—AUC = 0.9761,  $p = 2.798e-06$ ) and wMUS females (wMUS.female vs Other(s) AUC = 0.9543,  $p = 7.779e-06$ ) are again separated (i.e. near-perfect discrimination) from all other groups, though in this scenario, wMUS and MUS are closer to each other than to DOM. This is reasonable evidence that all three factors (sex, group and environment) has an effect upon differentiation, though environment has a lower influence than in metabolomes. Next, we extracted only lipocalins with eight-stranded beta barrels and calycins with ten-stranded beta barrels from the whole dataset for their known transporting functions in chemical communication, reviewed in<sup>68</sup>. Here (Fig. 1C), MUS and wMUS males overlap and so do MUS and wMUS females. However, DOM females are less separated from DOM males



**Figure 1.** Proteomes and metabolomes are under genomic and environmental control. Discriminant analysis sPLS-DA revealed strong influence of environment upon metabolomes (A) and proteomes (B) in that the wild-derived mice wMUS are clearly separated from laboratory-bred MUS and DOM groups. However, lipocalins (C) are under genomic control, which is demonstrated by DOM males and females being separated from overlapping MUS and wMUS within each sex. Background prediction (polygons) is based on the Maximum distance method. To detect which are the proteins and volatiles that best represent each sex (D), we show with random forest for classification the top thirty sex-determining molecules that are ranked by the importance based on permuted Out of bag (OOB) scores (D–I). Top two examples of important volatiles and in each strain are in (J–O). Chemical structures are free downloads from <https://www.chemspider.com>. Colour code: darker colours are males lighter colours are females throughout.

and reach the lowest AUC scores (DOM.female vs Other(s), Comp 1 (x-axis) – AUC=0.5630,  $p=5.350e-01$ ), but they are well separated from wMUS and MUS. This evidence also corroborates our previously reported finding that the level of sexual dimorphism in the expression of MUPs is higher in MUS than in DOM<sup>7</sup>. The major conclusion here is that lipocalin and calycin variation is driven by genetic differences (DOM vs. MUS) and not by environment (MUS vs. wMUS) while whole proteomes and metabolomes are environmentally modulated.

To detect which of these proteins and volatiles best predict the differences due to sex, we used random forest for classification (Fig. 1D–I). In Fig. 1J–O, we are showing examples of the most important compounds that best separate sex that have been detected with random forest. When looking at proteins, MUP20 (known as darcin) well represents sex in MUS (9th position) and DOM (2nd), but not in wMUS. For example, the 2nd position

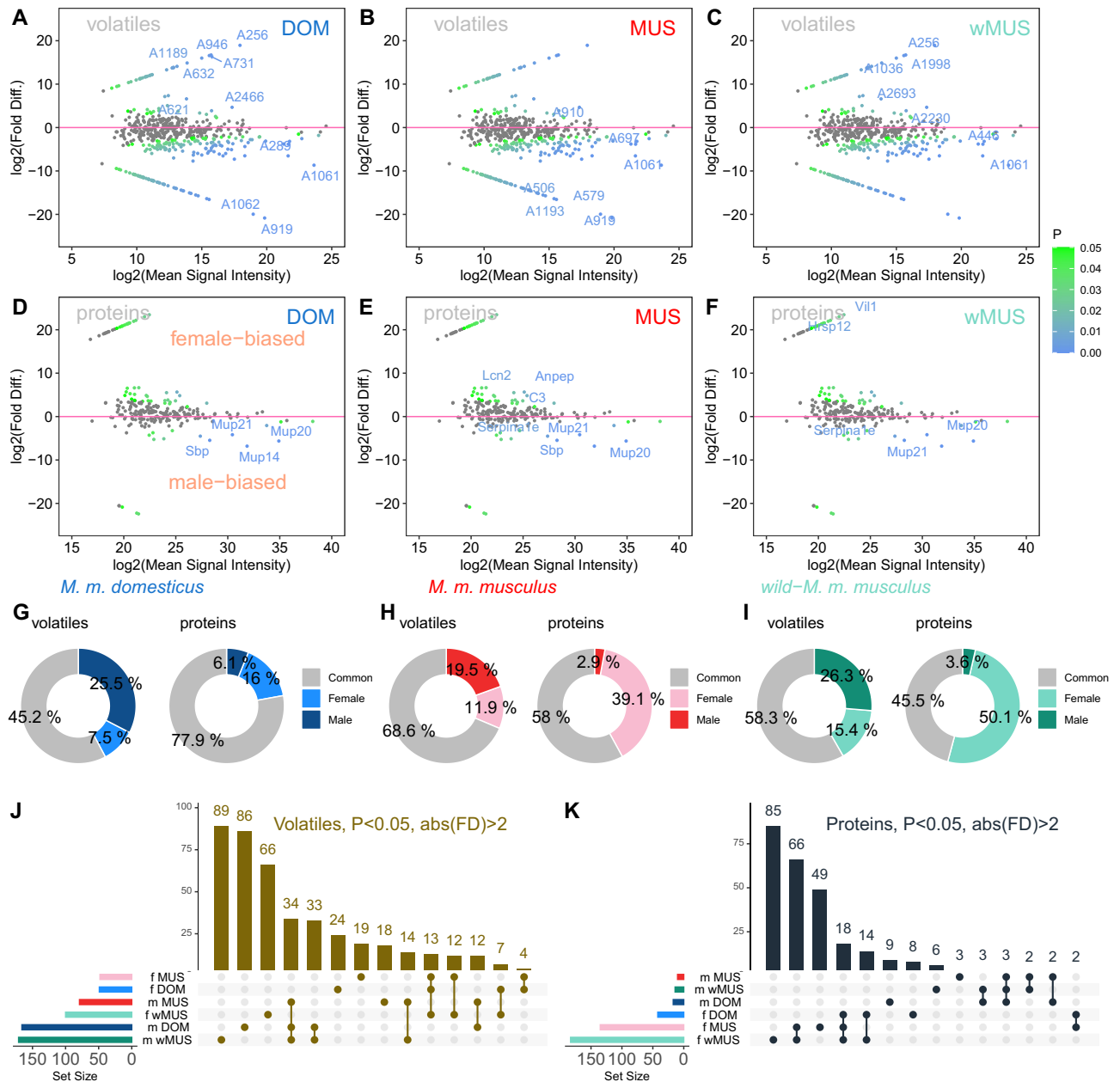
occupies the female-biased protein C3 (Power Law Global Error Model—PLGEM,  $FD_{F-M} = 20.6$ ,  $P_{wMUS} = 0.02$ ) that plays central roles in innate immunity and ENO1 on the 4th position ( $FD_{F-M} = 5.9$ ,  $P_{wMUS} = 0.003$ ) stimulates immunoglobulin production (extracted from UNIPROT functions, <https://www.uniprot.org/>). This discrepancy is likely caused by the fact that wMUS are direct offspring from wild-caught mice while MUS and DOM were bred and kept for generations in the same facility. The wild environment is immunologically more challenging than standard conditions of the mouse facility and thus wild mothers presumably transferred to their offspring their microbiota and an ‘immune memory’ which is influenced by the host microbiota<sup>67,69</sup>. This is also corroborated by the highly enriched and significant GO terms ( $FDR < 0.01$ ) in wMUS (unlike MUS and DOM), which is dominated by the immune system process (proteins: CD48, C3, CD59A, CD44, SDC4, LCN2, DPP4, EZR). Furthermore, differences in the composition of microbial communities between lab-bred DOM and MUS, and wild mice have been previously studied in the same facility using similar mice and design<sup>70</sup>. They found that laboratory DOM and MUS have similar microbiota and that both are different from wMUS and wDOM. Their finding suggests that diverging microbial communities may contribute to proteomic and metabolic variation.

**Levels of sexual dimorphism.** In mice, mate selection does not rely only on females but both sexes are to some extent ‘choosy’<sup>71,72</sup>, so we asked an important question: how many proteins and volatiles are sexually dimorphic and do both subspecies use the same system of chemical signalling? To test the hypothesis that males and females have different urine chemical profiles and that the two subspecies might have evolved different systems of signalling, we used PLGEM models of differential expression to extract levels of sexual dimorphisms, and in combination with deep learning we aimed to identify most important (representative) molecules in the clouds of significant sex-biased data for each of the three mouse groups. In Fig. 2A–F, we show MA plots where volatiles (A–C) and proteins (D–F) are plotted as fold differences (FD) against the mean signal intensities and in  $\log_2$  scale, Fig. 2A–F. Important proteins and volatiles that have been identified with Random forest and those that are ranked as top ~25 most important molecules (importance > 0.1) are labelled with gene or compound names. Reassuring message here is that most of the top molecules that were identified with deep learning were corroborated using the analysis of differential expression (e.g. MUP20 in DOM and MUS, MUP21 in wMUS).

The most interesting result that we report here is that males produce larger variety of volatiles whilst females produce larger variety of proteins, while the total protein volume is substantially enriched in males. This divergence, which is consistent across the three studied mouse groups, is unlikely to have emerged by chance (Fisher’s exact test on counts:  $P_{DOM} = 2.248e-11$ ,  $P_{MUS} < 2.2e-16$ ,  $P_{wMUS} < 2.2e-16$ ), Fig. 2G–I. In Fig. 2J–K we demonstrate using the intersection plots that sex identity is manifested by different molecules in each mouse group, or that those that are shared by all males (34 volatiles, 3 proteins) or by all females (13 volatiles, 18 proteins) are less common. This means that male and female odour spaces are dominated by molecules that have strain-biased expressions while those that are stereotypically produced in males or females regardless of the mouse origin are less common. Prime example of such shared molecule is MUP20 that is significantly male-biased in all the three studied mouse groups. Furthermore, MUP20 is not uniquely expressed only by males but is male-biased in all the three studied mouse groups. This means that previously reported pheromonal activities of darcin are driven by the expression differences rather than by sex-specific (unique) expressions.

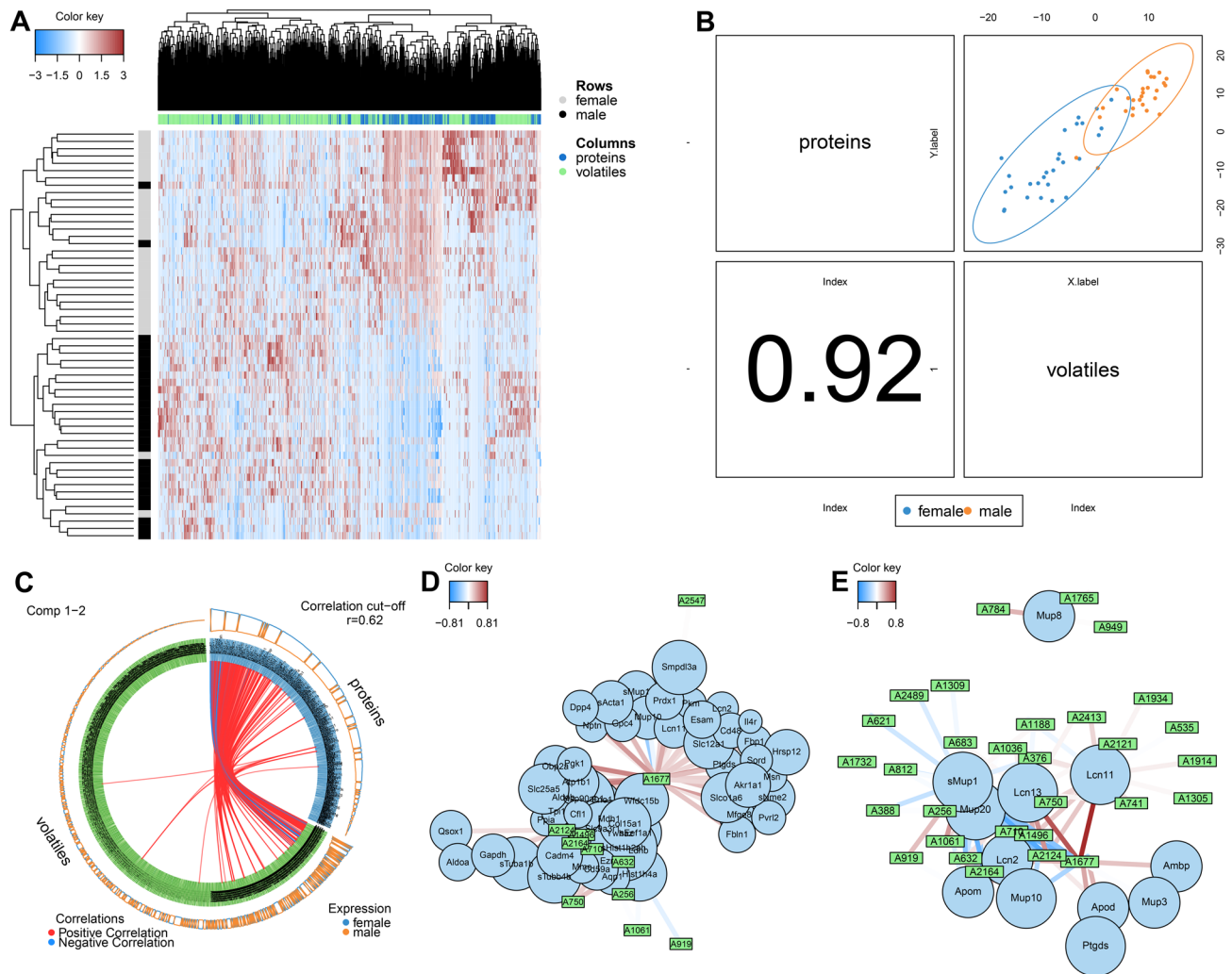
**Integration of proteomes and metabolomes revealed new putative semio-chemicals.** In order to disentangle how proteomes and metabolomes interact, we used the discriminant analysis on blocks (i.e. block of proteomes and block of volatiles). First, we asked whether males and females from all the three groups have common and observable features that define maleness and femaleness. In Fig. 3A, we clearly see the blocks of proteins typical for females while blocks dominated by volatiles better represent males. We used the Area Under the Curve (AUC) to provide evidence that the discrimination is perfect in both dimensions (AUC1 vs. AUC2) in proteins and even better in volatiles (proteins: AUC1 = 0.9413,  $p = 1.429e-08$ , AUC2 = 0.9452,  $p = 1.071e-08$ ; volatiles: AUC1 = 0.9796,  $p = 7.208e-10$ , AUC2 = 0.9821,  $p = 5.857e-10$ ). A global overview of the correlation structure at the component level in Fig. 3B revealed strong correlation between proteomic and metabolomic data ( $r = 0.92$ ), which offers an interpretation that both types of molecules represent sexuality in combinations. Figure 3C shows that the distribution of correlation coefficients ( $r > 0.65$ ) is not random and that – based on the circular histograms – the most abundant proteins positively correlate with the most abundant volatiles. Thus, we extracted a network in Fig. 3D that represents the best correlations ( $r > 0.62$ ) between proteins and volatiles.

The highest connectivity of a volatile to proteins and which has been identified as important by random forest (Fig. 1D,F) is typical for (A1677), which is 2(5H)-furanone, 5,5-dimethyl- (i.e. 5,5-dimethylfuran-2(5H)-one). It belongs to organic compounds known as butenolides. In our network, this sex-unbiased compound correlates with female-biased LCN2 and LCN11 and with male-biased MUP1 and MUP10. This connection is interesting because 2(5H)-furanone is a quorum-sensing molecule produced by fungi and bacteria to regulate bacterial growth and, for example, bovine OBPs scavenge this compound to prevent the bacterial growth thus turning to pathogens<sup>73</sup>. At the same time, LCN2 prevents bacterial growth by scavenging iron chelating bacterial siderophores in mice and humans<sup>74</sup>. Another important molecule (DOM in Fig. 1d, MUS in Fig. 1e) is A2124, which is (1,2,3,5,8,8a)-hexahydro-naphthalene, also known as dysoxylonene, a very hydrophobic molecule that belongs to sesquiterpenoids and in urine likely needs a protein transporter to enter aqueous environment. In our network, this compound also correlates with the female-biased protein LCN2, it is abundant in females of DOM, wMUS and MUS. This analysis also revealed high correlations between MUP20 and A919, which is 2-acetyl-3-thiazoline. This compound is highly similar to 2-s-butylthiazole, a natural ligand of darcin. However, in our data 2-acetyl-3-thiazoline better reveals maleness than 2-s-butylthiazole because it is unique in DOM and MUS males (~20 fold difference,  $p < 0.0001$ ) and significantly male-biased in wMUS (~eightfold difference,  $p < 0.0001$ ). Moreover, the structures of 2-acetyl-3-thiazoline and 2-s-butylthiazole are so similar that they might



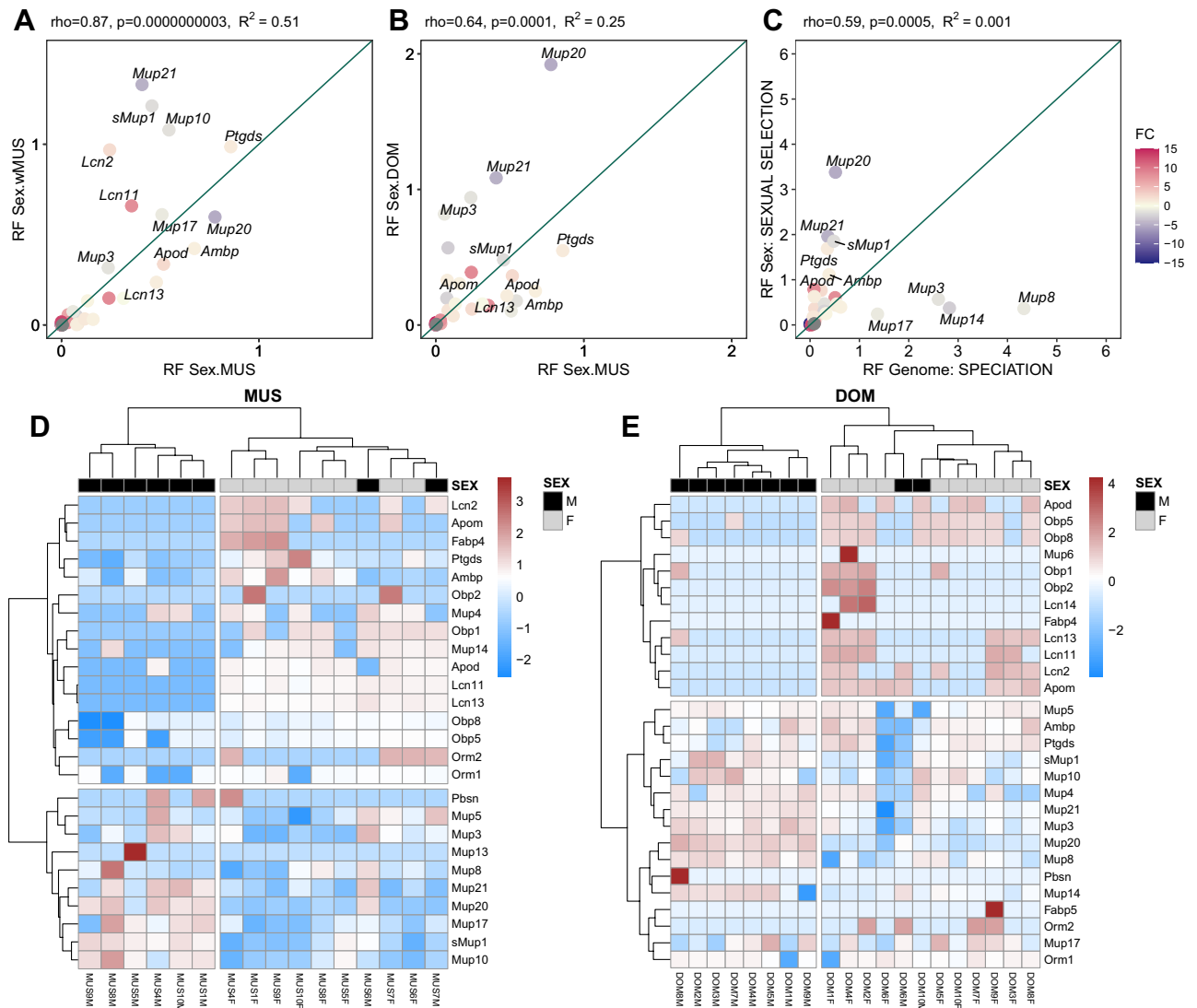
**Figure 2.** Sexually dimorphic molecules maintain sex- and strain-specific odour space. Differentially abundant volatiles (A–C) and proteins (D–F,  $p < 0.05$ ,  $\text{abs}(\text{FD}) > 2$ ) are scaled from green to blue but only top ten proteins and volatiles identified with ‘random forest’ as important are labelled with gene names or compound numbers. Above  $y = 0$  are the female biased molecules while the male-biased are below the red line ( $y = 0$ ). Next comparison involved significant sex-biased volatiles and proteins with  $p < 0.05$  and  $\text{abs}(\text{FD}) > 2$ . In all three comparisons (G–I), males have more sex-biased volatiles while females have more sex-biased proteins. Though this pattern is significant in all the three groups, each group reveals sexuality by different volatiles and proteins (intersection plots in J–K). Abbreviations:  $\text{abs}()$  means an absolute value of; FD stands for fold difference.

have the same transporter darcin (MUP20). To obtain a further insight onto potential relationships between volatiles and proteins, we performed sPLS-DA on blocks of a complete volatile set and lipocalins without other proteins. The relationships described above are again supported in Fig. 3E but we also found some new and interesting associations. The prime example is MUP8 correlating with A784 which is 2-Methyl-1-nonene-3-yne. This compound has a plant/food origin and has high antimicrobial activity<sup>75</sup>. It is elevated in DOM males and slightly less in DOM females ( $\text{FD} = 2$ ,  $P = 0.08$  thus NS) whereas only a few MUS and wMUS individuals had this compound. The correlation between MUP8 and methyl-1-nonene-3-yne across all individuals and mouse groups ( $r = 0.62$ ) is significant ( $P < 0.05$ ). Although this approach yields interesting interactions between proteins and their potential ligands it is necessary to perform further binding experiments which are, however, beyond the scope of this study.



**Figure 3.** Integration of proteomes and metabolomes with correlation analysis revealed new potential interactions. Clustered Image Map (A) shows that correlated proteins and metabolites best explain sex differences across all individuals regardless strain. Males have more correlated volatiles while females have more correlated proteins. There is a positive correlation between proteins and volatiles at the component level (correlation = 0.92,  $p < 0.05$ ) (B). This multi-omics molecular signature is mainly caused by the correlations between highly abundant proteins and volatiles (C), see the circular histograms. Stringent network analysis (correlation > 0.6) shows potential interactions between proteins and volatiles (D) and between (only lipocalins and volatiles).

**Lipocalin code.** In this comparison, we performed random forest on a subset of proteins from the lipocalin family. First, we plotted random forest (RF) out-of-bag importance of individual proteins for sex separation in wMUS versus MUS (Fig. 4A) and found that Spearman rank correlation is high ( $\rho = 0.86$ ) and significant ( $p = 3e-10$ ;  $R^2 = 0.51$ ). This is because they are genetically more alike and thus the same proteins are characteristic of sex separation. Similarly, we plotted the RF importance in DOM vs. MUS (Fig. 4B). As expected, correlation between DOM and MUS was lower ( $\rho = 0.64$ ;  $p = 0.0001$ ;  $R^2 = 0.25$ ) than between MUS and wMUS, because of the genetic dissimilarity between the two subspecies. Plotting the RF importance for sex separation against the RF importance for subspecific separation (MUS, DOM) revealed very low correlation ( $\rho = 0.59$ ;  $p = 0.0005$ ;  $R^2 = 0.001$ ), Fig. 4C. This diverging pattern provides evidence for the specialized nature of lipocalins where for example MUP20 and MUP21 reveal sex identity in all the studied groups while MUP14 and MUP8 abundances display subspecific status (see also Fig. 4D-E). Overall, MUP20 (darcin) is also the main driver of sexual dimorphism in our complete proteomic datasets. In heatmaps (Fig. 4D-E) only MUS and DOM are compared, because they were bred in the same facility. Here we demonstrate using sPLS-DA scores that there are more lipocalins in the urine than previously reported and that their variation is high. Hierarchical clustering corroborated that sex is a good though not an absolute predictor of lipocalin variation.



**Figure 4.** Combinatorial lipocalin code. Random forest (RF) plots of importance of particular proteins for sex separation: (A) wMUS against MUS; (B) DOM against MUS; in (C), the RF importance for separation of sexes is plotted against subspecies separation. Individual points are scaled from blue (male-biased) to red (female-biased). (D–E) Hierarchical clustering reveals the combinatorial nature of lipocalin abundances. In heatmaps, we demonstrate the relative contribution of proteins to sex separation using sPLS-DA (x-axis – individual names, y-axis – lipocalin gene names).

## Discussion

What is the most prominent feature of mouse chemical signals in the urine? Is it the volatiles or proteins? In contrast to other studies, we used higher volumes of urine (10  $\mu$ l) for our label-free mass spectrometry directly from the samples and avoided using IPG strips and gels. Many gel-based MUP studies employed isoelectric focusing of proteins on IPG strips or gels with the range of isoelectric points pI 3.9–5.1 which only detect MUPs and not most other lipocalins with higher pI such as OBPs and LCNs that are more elevated in females. This approach often led to a focus on male MUPs as the main drivers of sexual dimorphisms, however, they are over-represented in contrast to hundreds of other proteins that are present in the urine of females. We used the two house mouse subspecies DOM and MUS, which in Europe form a narrow hybrid zone running from Norway to the Black Sea<sup>76,77</sup>. We used sex-differences in the production of compounds as a proxy for sexual selection while subspecies differences were a proxy for speciation. We had yet another set of wMUS animals that served to find out whether the combination of environment and hygienic status (wild vs. lab-bred) influences urinary profiles.

For the first time, we show that female mice have a variety of female-biased lipocalins in their urine and that some of these lipocalins were previously detected only in mucosal secretions of eyes<sup>78</sup>, nose<sup>79–82</sup>, oral cavity<sup>83</sup>, and vagina<sup>63,84</sup>. MUPs are highly expressed in the liver<sup>85</sup> and it has been repeatedly demonstrated that they are excreted to the urine and deposited as urine marks thus slowly releasing their ligands (VOCs). Mouse OBPs are not expressed in the liver<sup>86</sup>, however they are abundant in vagina where they are co-expressed with other detected proteins including LCN11, LCN2, MUP9, darcin, and other lipocalin members. They are upregulated

during estrus and metestrus and they drop to lower levels in proestrus<sup>63</sup>. Thus, it is likely that female glands and reproductive organs produce some of the proteins detected in their urine, which reflects their reproductive state. This corroborates our previous studies which demonstrated that the abundance of female MUPs in the urine correlates with the estrous cycle in the laboratory<sup>29</sup> and wMUS<sup>28</sup> mice, reviewed in<sup>68</sup>.

Taking a broader view, we touched a fundamental question in biology, namely how is sexuality signalled<sup>87</sup> in animals that primarily depend on olfactory cues<sup>88,89</sup> and whether a single pheromone or a mixture of compounds may potentially serve to prime social and reproductive behaviours of the receiver. In mammals, sexuality is often maintained by sexual dimorphisms where some components evolved to display sexual traits that are specifically processed in the brain<sup>12,13,90</sup> while others are the consequences of sex-biased metabolic processing<sup>36</sup> and immune defence<sup>91,92</sup>. Not all proteins and compounds that are sex-specific are involved in chemical signalling. But from our data and other studies we can see that it is likely that a small effect of many molecules, rather than a strong effect of few, is characteristic for mouse chemical signals, similarly as in mole rat perioral secretions<sup>93</sup>. In our data, sexuality is well displayed via lipocalins (e.g. MUPs, OBPs, LCNs) that are known for their roles in chemical communication and by several volatiles that have been studied in many laboratories (e.g. SBT, farnesenes, pyrazines etc.) and species, for example the mole rats<sup>93</sup>. However, the majority of proteins and volatiles in our data have not been previously studied in the context of sexual signalling. Of course, it would be best to test each of the detected compounds in individual behavioural setups, but this is practically impossible. Another option, presented in this paper is a pre-selection based on the searches of compounds that have correlated patterns and thus may have the potential to represent biological features such as sex, sub-species and hygienic status of an individual. If a volatile is too hydrophobic, it needs a protein transporter that can help the ligand to enter aqueous environment (i.e., urine). Thus, it is reasonable to expect that proteins and volatiles will to some extent be correlated and this is exactly what our study shows. Volatiles are correlated with proteins but only few are organized in larger networks of proteins and their potential ligands. When these correlations are extracted, we can see putative relationships between combinations of proteins and ligands such as MUP20 and 2-acetyl-3-thiazoline as well as the new putative pairs (e.g. LCN11 and 2(5H)-furanone). We do realize that correlation is not the same as causation, but this approach may lead to new set of hypotheses based on the complexity of the molecular profiles indicated by this study.

To conclude, we used deep learning and data integration to identify in the urine metabolome and proteome of mice, molecules that are sex- and subspecies- specific, and are likely involved in chemical signalling. We have also shown for the first time that sexuality is displayed by at least 26 different lipocalins and calycins (12–16 are female biased) and not just by male-biased MUPs. However, the number of shared peptides in this group of proteins exposes a need for absolute quantification of these proteins, based on unbiased methods. Furthermore, striking differences in the abundance of sex-biased molecules between DOM and MUS revealed that there was a strong selection on systems of sexual signalling during the speciation of DOM and MUS mice.

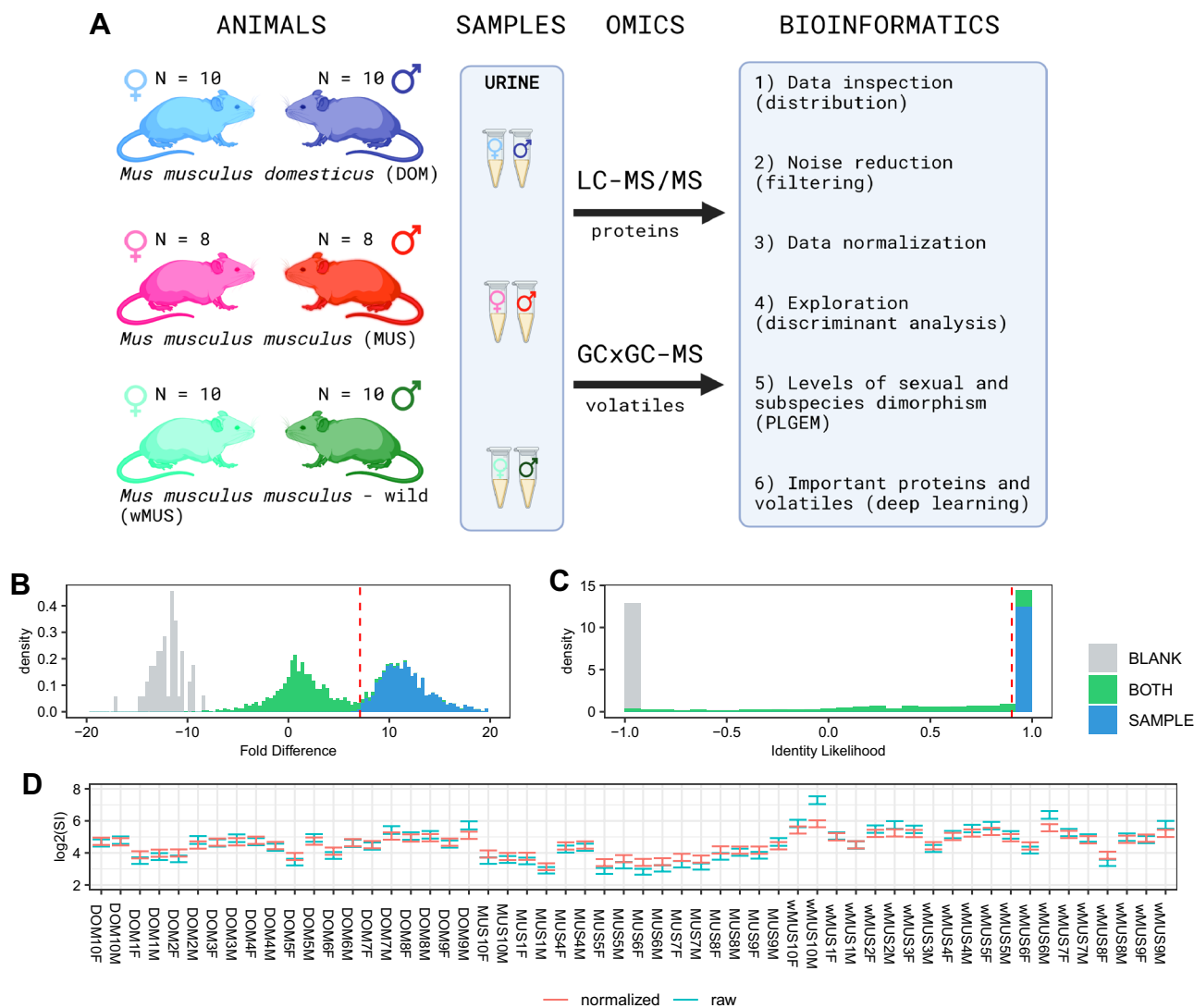
## Materials and methods

**Ethical standards.** All animal procedures were carried out in strict accordance with the law of the Czech Republic paragraph 17 no. 246/1992. Handling of wild MUS was approved by the local ethics committee of the Faculty of Science, Charles University in accordance with accreditation no. 27335/2013-17214. Wild-derived mice were kept in the breeding facility of the Institute of Vertebrate Biology in Studenec (authorized by the Ministry of Agriculture 61974/2017-MZE-17214). The following strains represented MUS (Armenia: MAM, Czechia: MCZ and PWK, Georgia: MGA, Poland: MPB, Bulgaria: SOK and SVEN) and DOM (Algeria: BZO, Austria: BING and URG, Cyprus: DCA and DCP, Denmark: DDO, Italy: DJO, Portugal: SAGR and SOAL) (for details see<sup>94–96</sup>, and <https://housemice.cz/en/strains/>). This study was performed and reported in accordance with ARRIVE guidelines (<https://arriveguidelines.org>).

**Subjects, housing conditions and experimental design.** In this experiment (Fig. 5A), we used a total of 56 mice from the three groups: wild *Mus musculus musculus* (wMUS: direct offspring from wild-caught mice, 10 pairs), laboratory *M. m. musculus* (MUS: generations G3-G60 of lab-bred mice, 8 pairs) and laboratory *M. m. domesticus* (DOM: generations G1-G67 of lab-bred mice, 10 pairs). All individuals had the same diet (ST1 pellets, Velaz, Prague, Czech Republic) and water ad libitum and were kept under stable conditions (i.e. 14:10 h, D:N, temperature  $t = 23$  °C). We collected their urine four to six times to overcome potential differences in dilutions. Final volume of urine was between 25 and 100  $\mu$ l per mouse and all samples were frozen ( $-80$  °C) before further analyses. These samples were measured with GCxGC-MS/MS and in parallel additional 10  $\mu$ l of samples were used for the nLC-MS/MS analyses of proteins (see below). None of the animals was euthanized or sacrificed during the sampling. We only manipulated individuals such that they urinated to a collection tube and then they were returned to their cage.

**Detection of volatile metabolites.** Urine volatiles were sampled using Headspace Solid Phase Micro Extraction (HS SPME) on fiber (DVB/CAR/PDMS\_grey; Supelco, USA). Samples were incubated for 5 min at 55 °C prior to the extraction. The extraction was carried out for 10 min. The volatiles were analyzed using two-dimensional comprehensive gas chromatography with mass detection (GCxGC-MS; Pegasus 4D, Leco Corporation, USA). A combination of mid-polar and non-polar separation columns was used for the separation: Primary column: SLB-IL60 (30 m  $\times$  0.25 mm, SigmaAldrich, USA); Secondary column Rxi-5sil MS (1.4 m  $\times$  0.25 mm, Restek, Australia). Other parameters were set as follows: inlet temperature 270 °C, spitless mode, constant He flow 1 ml/min, modulation time 4 s (hot pulse 0.6 s), modulation temperature offset with respect to the secondary oven 15 °C. The temperature program applied on the primary oven: 50 °C (hold 1 min), then increase (10 °C/min) to 320 °C (hold 3 min). The temperature offset applied on the secondary column was +5 °C. Transferline





**Figure 5.** Experiment design, filtering, and normalization. We used mouse urine repeatedly sampled from individuals of each sex from the three groups – laboratory-bred DOM and MUS, and wild wMUS (A). We focused on the analysis of their urine proteins and volatiles. We excluded volatiles that occurred only in blanks (B, grey bars), whilst those that occurred in blanks and samples (green) were selected based on the posterior p-values of the mixed-normal model (see methods). Those that had  $p < 0.05$  (i.e. corresponding FD  $< 7.1$ ) of belonging to blanks and samples were removed (red line); this corresponds to the Identity likelihood IL  $< 0.9$  (C). Remaining compounds ( $N = 875$ ) were considered as relevant because they only occurred in samples or in significantly higher quantities in samples than in blanks (FD  $> 7.1$ ). Quantile normalisation yielded reasonably low variation in signal intensities (SI) between samples (D). FD stands for fold difference. Tubes, mouse pics and chemical structures (A) were created by the authors with BioRender.com.

temperature was held on 250 °C. The mass detector was equipped with an EI ion source and TOF analyzer enabling unite mass resolution. The scanned mass range was 30–500 m/z. The ion source chamber was held on 250 °C. LECO's ChromaTOF v4.5 was employed to control the instrument and for data processing. Selected compounds were identified by matching their mass spectra with a library of mass spectra (NIST MS 2.2, USA).

**Analysis of volatile metabolites.** We generated histograms of data distribution and removed all the rows with compounds that occurred only in blanks and not in samples. The resulting distribution is bi-modal with compounds that occurred only in samples and in samples and blanks, Fig. 5B. To decide which compounds are biologically relevant, we used 'mixtools'<sup>97</sup> routine which calculates the posterior probability for the identity to either of the two peaks within the mixture of two overlapping normal distributions. We excluded all the compounds, which had the identity to blanks and samples with  $p < 0.05$ , Fig. 5C. To visualize the likelihood of identity to either of the two peaks, we used a simple index of identity LI = (sample - blank)/(sample + blank) ranging from -1 to 1 so all the remaining 'biologically relevant' compounds had LI  $> 0.9$  (Fig. 5C). Next, we used a normalization based upon quantiles, which normalizes a matrix of peak areas (i.e. intensities) with the function `normalize.quantiles` of the 'preprocessCore' package in R software<sup>98</sup>, visualised in Fig. 5D. To extract p-values of

differentially abundant compounds, we used the Power Law Global Error Model – PLGEM<sup>99</sup> similarly as in the analysis of proteomes (see below).

**Protein Digestion and nLC-MS/MS Analysis.** All protein samples were cold-acetone precipitated and centrifuged at 14,000 rcf for 10 min at 0 °C. This was followed by a re-suspension of dried pellets in the digestion buffer (1% SDC, 100 mM TEAB – pH = 8.5). The protein concentration of each lysate was determined using the BCA assay kit (Fisher Scientific). Cysteines in 20 µg of proteins were reduced with a final concentration of 5 mM TCEP (60° C for 60 min) and blocked with 10mM MMTS (i.e. S-methyl methanethiosulfonate, 10 min room temperature). Samples were cleaved with trypsin (1 µg of trypsin per sample) in 37 °C overnight. Peptides were desalted on a Michrom C18 column. Nano Reversed phase columns were used (EASY-Spray column, 50 cm × 75 µm ID, PepMap C18, 2 µm particles, 100 Å pore size). Eluting peptide cations were converted to gas-phase ions by electrospray ionization and analysed on a Thermo Orbitrap Fusion (Q-OT-qIT, Thermo) with the same parameters as described in<sup>78,79,83</sup>.

**Proteomic analysis.** LC–MS data were pre-processed with MaxQuant software (version 1.6.34)<sup>66</sup>. The false discovery rate (FDR) was set to 1% for both proteins and peptides and we specified a minimum peptide length of seven amino acids. The Andromeda search engine was used for the MS/MS spectra mapping against our modified Uniprot *Mus musculus* database (downloaded in June, 2015), containing 44,900 entries. We modified our databases such that all MUP and OBP sequences were removed and instead of them we have added a complete list of MUPs from Ensembl database, and OBPs from NCBI (sensu—citation<sup>86</sup>). Next, we added some Tremble sequences that were missing in Uniprot, for example KLKs, BPLs, SPINKs, SCGB/ABPs, and LCNs. We provide this dataset in FASTA format as Supplementary dataset 1. Enzyme specificity was set as C-terminal to Arg and Lys, also allowing cleavage at proline bonds<sup>100</sup> and a maximum of two missed cleavages. Dithiomethylation of cysteine was selected as fixed modification and N-terminal protein acetylation and methionine oxidation as variable modifications. The ‘match between runs’ feature of MaxQuant was used to transfer identifications to other LC–MS/MS runs based on their masses and retention time (maximum deviation 0.7 min). Quantifications were performed using the label-free algorithms<sup>66</sup> with a combination of unique and razor peptides. All subsequent analyses were performed in R software<sup>98</sup>. To check that the data distribution conforms to the same type of distribution after normalization, we used ‘mixtools’<sup>97</sup>. Second, we used the Power Law Global Error Model—PLGEM<sup>99</sup> to detect differentially expressed / abundant proteins using the functions `plgem.fit` and `plgem.stn`<sup>97</sup>. To detect the importance of significant proteins in separation between males and females we used Random Forest for Classification<sup>101</sup> within the R software<sup>98</sup>. All plots and figures were generated in R using `ggplot2`<sup>102</sup>. R software is distributed under the terms of the GNU General Public License. Copies of both versions 2 and 3 of the license can be found at <https://www.R-project.org/Licenses/>. Original and LC–MS/MS and GCxGC/MS tables are provided in Supplementary dataset 2.

## Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD037086 and <https://doi.org/10.6019/PXD037086>. Resulting tables are available as supplementary data. Metabolomics data have been deposited to the EMBL-EBI MetaboLights database (<https://doi.org/10.1093/nar/gkz1019>, PMID: 31691833) with the identifier MTBLS7422. The complete dataset can be accessed here <https://www.ebi.ac.uk/metabolights/MTBLS7422>.

Received: 14 September 2022; Accepted: 18 May 2023

Published online: 26 May 2023

## References

1. Yoon, H., Enquist, L. W. & Dulac, C. Olfactory inputs to hypothalamic neurons controlling reproduction and fertility. *Cell* **123**, 669–682 (2006).
2. Marom, K. *et al.* The vomeronasal system can learn novel stimulus response pairings. *Cell Rep.* **27**, 676–684.e676. <https://doi.org/10.1016/j.celrep.2019.03.042> (2019).
3. Zala, S. M., Potts, W. K. & Penn, D. J. Scent-marking displays provide honest signals of health and infection. *Behav. Ecol.* **15**, 338–344. <https://doi.org/10.1093/beheco/arih022> (2004).
4. Zala, S. M., Bilak, A., Perkins, M., Potts, W. K. & Penn, D. J. Female house mice initially shun infected males, but do not avoid mating with them. *Behav. Ecol. Sociobiol.* **69**, 715–722. <https://doi.org/10.1007/s00265-015-1884-2> (2015).
5. Mucignat-Caretta, C., Cavaggoni, A. & Caretta, A. Male urinary chemosignals differentially affect aggressive behavior in male mice. *J. Chem. Ecol.* **30**, 777–791 (2004).
6. Mucignat-Caretta, C. *et al.* Urinary volatile molecules vary in males of the 2 European subspecies of the house mouse and their hybrids. *Chem. Senses* **35**, 647–654. <https://doi.org/10.1093/chemse/bjq049> (2010).
7. Stopková, R., Stopka, P., Janotová, K. & Jedelsky, P. L. Species-specific expression of major urinary proteins in the house mice (*Mus musculus musculus* and *Mus musculus domesticus*). *J. Chem. Ecol.* **33**, 861–869 (2007).
8. Pérez-Gómez, A. *et al.* Innate predator odor aversion driven by parallel olfactory subsystems that converge in the ventromedial hypothalamus. *Curr. Biol.* **25**, 1340–1346. <https://doi.org/10.1016/j.cub.2015.03.026> (2015).
9. Yang, J. *et al.* Landscapes of bacterial and metabolic signatures and their interaction in major depressive disorders. *Sci. Adv.* **6**, eaba8555. <https://doi.org/10.1126/sciadv.aba8555> (2020).
10. Ninkovic, V., Markovic, D. & Rensing, M. Plant volatiles as cues and signals in plant communication. *Plant Cell Environ.* <https://doi.org/10.1111/pce.13910> (2020).
11. Manoel, D. *et al.* Deconstructing the mouse olfactory percept through an ethological atlas. *Curr. Biol.* <https://doi.org/10.1016/j.cub.2021.04.020> (2021).
12. Bansal, R. *et al.* Do all mice smell the same? Chemosensory cues from inbred and wild mouse strains elicit stereotypic sensory representations in the accessory olfactory bulb. *BMC Biol.* **19**, 133. <https://doi.org/10.1186/s12915-021-01064-7> (2021).

13. Bergan, J. F., Ben-Shaul, Y. & Dulac, C. Sex-specific processing of social cues in the medial amygdala. *Elife* **3**, e02743. <https://doi.org/10.7554/eLife.02743> (2014).
14. Nagel, M. *et al.* A systematic comparison of semiochemical signaling in the accessory olfactory system of wild and lab strain mice. *Chem. Senses* **43**, E31–E31 (2018).
15. Spehr, M. *et al.* Parallel processing of social signals by the mammalian main and accessory olfactory systems. *Cell. Mol. life Sci.* **63**, 1476–1484 (2006).
16. van der Linden, C., Jakob, S., Gupta, P., Dulac, C. & Santoro, S. W. Sex separation induces differences in the olfactory sensory receptor repertoires of male and female mice. *Nat. Commun.* **9**, 5081. <https://doi.org/10.1038/s41467-018-07120-1> (2018).
17. Santoro, S. W. & Jakob, S. Gene expression profiling of the olfactory tissues of sex-separated and sex-combined female and male mice. *Sci. Data* **5**, 180260. <https://doi.org/10.1038/sdata.2018.260> (2018).
18. Moss, R. L. *et al.* Urine-derived compound evokes membrane responses in mouse vomeronasal receptor neurons. *J. Neurophysiol.* **77**, 2856–2862 (1997).
19. Leinders-Zufall, T. *et al.* Ultrasensitive pheromone detection by mammalian vomeronasal neurons. *Nature* **405**, 792–796 (2000).
20. Ibarra-Soria, X., Levitin, M. O. & Logan, D. W. The genomic basis of vomeronasal-mediated behaviour. *Mamm. Genome* **25**, 75–86. <https://doi.org/10.1007/s00335-013-9463-1> (2014).
21. Wynn, E. H., Sánchez-Andrade, G., Carss, K. J. & Logan, D. W. Genomic variation in the vomeronasal receptor gene repertoires of inbred mice. *BMC Genomics* **13**, 415. <https://doi.org/10.1186/1471-2164-13-415> (2012).
22. Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. *Cell* **65**, 175–187. [https://doi.org/10.1016/0092-8674\(91\)90418-x](https://doi.org/10.1016/0092-8674(91)90418-x) (1991).
23. Whitten, W. K., Bronson, F. H. & Greenstein, J. A. estrus-inducing pheromone of male mice: Transport by movement of air. *Science* **161**, 584–585 (1968).
24. Whitten, W. K. Modification of the oestrous cycle of the mouse by external stimuli associated with the male. Changes in the oestrous cycle determined by vaginal smears. *J. Endocrinol.* **17**, 307–313 (1958).
25. Novotny, M. V., Ma, W., Wiesler, D. & Zidek, L. Positive identification of the puberty-accelerating pheromone of the house mouse: The volatile ligands associating with the major urinary protein. *Proc. R. Soc. Lond. B.* **266**, 2017–2022 (1999).
26. Jemiolo, B. & Novotny, M. V. Inhibition of sexual maturation in juvenile female and male mice by a chemosignal of female origin. *Physiol. Behav.* **55**, 519–522 (1994).
27. Jemiolo, B., Harvey, S. & Novotny, M. Promotion of the whitten effect in female mice by synthetic analogs of male urinary constituents. *PNAS* **83**, 4576–4579 (1986).
28. Janotova, K. & Stopka, P. The level of major urinary proteins is socially regulated in wild *Mus musculus musculus*. *J. Chem. Ecol.* **37**, 647–656. <https://doi.org/10.1007/s10886-011-9966-8> (2011).
29. Stopka, P., Janotova, K. & Heyrovsky, D. The advertisement role of major urinary proteins in mice. *Physiol. Behav.* **91**, 667–670 (2007).
30. Kahan, A. & Ben-Shaul, Y. extracting behaviorally relevant traits from natural stimuli: benefits of combinatorial representations at the accessory olfactory bulb. *PLoS Comput. Biol.* **12**, e1004798. <https://doi.org/10.1371/journal.pcbi.1004798> (2016).
31. Rusu, A. S., Krackow, S., Jedelsky, P. L., Stopka, P. & Konig, B. A qualitative investigation of major urinary proteins in relation to the onset of aggressive behavior and dispersive motivation in male wild house mice (*Mus musculus domesticus*). *J. Ethol.* **26**, 127–135 (2008).
32. Roberts, S. A., Davidson, A. J., McLean, L., Beynon, R. J. & Hurst, J. L. Pheromonal induction of spatial learning in mice. *Science* **338**, 1462–1465. <https://doi.org/10.1126/science.1225638> (2012).
33. Chamero, P. *et al.* Identification of protein pheromones that promote aggressive behaviour. *Nature* **450**, 899–903 (2007).
34. Sturm, T. *et al.* Mouse urinary peptides provide a molecular basis for genotype discrimination by nasal sensory neurons. *Nat. Commun.* **4**, 1616. <https://doi.org/10.1038/ncomms2610> (2013).
35. Leinders-Zufall, T. *et al.* MHC class I peptides as chemosensory signals in the vomeronasal organ. *Science* **306**, 1003–1037 (2004).
36. Kwak, J. *et al.* Changes in volatile compounds of mouse urine as it ages: Their interactions with water and urinary proteins. *Physiol. Behav.* **120**, 211–219. <https://doi.org/10.1016/j.physbeh.2013.08.011> (2013).
37. Kwak, J. *et al.* Differential binding between volatile ligands and major urinary proteins due to genetic variation in mice. *Physiol. Behav.* **107**, 112–120. <https://doi.org/10.1016/j.physbeh.2012.06.008> (2012).
38. Novotny, M. V. *et al.* A unique urinary constituent, 6-hydroxy-6-methyl-3-heptanone, is a pheromone that accelerates puberty in female mice. *Chem. Biol.* **6**, 377–383 (1999).
39. Novotny, M. V. Pheromones, binding proteins and receptor responses in rodents. *Biochem. Soc.* **31**, 117–122 (2003).
40. Zidek, L. *et al.* NMR mapping of the recombinant mouse major urinary protein i binding site occupied by the pheromone 2-sec-Butyl-4,5-dihydrothiazole. *Biochemistry* **38**, 9850–9861 (1999).
41. Sharrow, S. D., Vaughn, J. L., Židek, L., Novotny, M. V. & Stone, M. J. Pheromone binding by polymorphic mouse major urinary proteins. *Protein Sci.* **11**, 2247–2256 (2002).
42. Phelan, M. M., McLean, L., Hurst, J. L., Beynon, R. J. & Lian, L. Y. Comparative study of the molecular variation between “central” and “peripheral” MUPs and significance for behavioural signalling. *Biochem. Soc. Trans.* **42**, 866–872. <https://doi.org/10.1042/BST20140082> (2014).
43. Phelan, M. M. *et al.* The structure, stability and pheromone binding of the male mouse protein sex pheromone darcin. *PLoS One* **9**, e108415. <https://doi.org/10.1371/journal.pone.0108415> (2014).
44. Robertson, D., Hurst, J., Hubbard, S., Gaskell, S. J. & Beynon, R. Ligands of urinary lipocalins from the mouse: Uptake of environmentally derived chemicals. *J. Chem. Ecol.* **24**, 1127–1140 (1998).
45. Janotová, K. & Stopka, P. Mechanisms of chemical communication: The role of major urinary proteins. *Folia Zool.* **58**, 41–55 (2009).
46. Macek, P., Novak, P., Krizova, H., Zidek, L. & Sklenar, V. Molecular dynamics study of major urinary protein-pheromone interactions: A structural model for ligand-induced flexibility increase. *FEBS Lett.* **580**, 682–684 (2006).
47. Timm, D. E., Baker, L. J., Mueller, H., Zidek, L. & Novotny, M. V. Structural basis of pheromone binding to mouse major urinary protein (MUP-I). *Protein Sci.* **10**, 997–1004 (2001).
48. Roberts, S. A. *et al.* Individual odour signatures that mice learn are shaped by involatile major urinary proteins (MUPs). *BMC Biol.* **16**, 48. <https://doi.org/10.1186/s12915-018-0512-9> (2018).
49. Roberts, S. A. *et al.* Darcin: A male pheromone that stimulates female memory and sexual attraction to an individual male's odour. *BMC Biol.* <https://doi.org/10.1186/1741-7007-1188-1175> (2010).
50. Kaur, A. W. *et al.* Murine pheromone proteins constitute a context-dependent combinatorial code governing multiple social behaviors. *Cell* **157**, 676–688. <https://doi.org/10.1016/j.cell.2014.02.025> (2014).
51. Papes, F., Logan, D. W. & Stowers, L. The vomeronasal organ mediates interspecies defensive behaviors through detection of protein pheromone homologs. *Cell* **141**, 692–703 (2010).
52. Demir, E. *et al.* The pheromone darcin drives a circuit for innate and reinforced behaviours. *Nature* <https://doi.org/10.1038/s41586-020-1967-8> (2020).
53. Johnson, D., Al-Shawi, R. & Bishop, J. O. Sexual dimorphism and growth hormone induction of murine phero-binding proteins. *J. Mol. Endocrinol.* **14**, 21–34 (1995).

54. Clissold, P. M., Hainey, S. & Bishop, J. O. Messenger RNAs coding for mouse urinary proteins are differentially induced by testosterone. *Biochem. Genet.* **22**, 379–387 (1984).
55. Shaw, P. H., Held, W. A. & Hastie, N. D. The gene family for major urinary proteins: Expression in several secretory tissues of the mouse. *Cell* **32**, 755–761 (1983).
56. Zidek, L., Novotny, M. V. & Stone, M. J. Increased protein backbone conformational entropy upon hydrophobic ligand binding. *Nat. Struct. Biol.* **6**, 1118–1121 (1999).
57. Hurst, J. L., Robertson, D. H. L., Tolladay, U. & Beynon, R. J. Proteins in urine scent marks of male house mice extend the longevity of olfactory signals. *Anim. Behav.* **55**, 1289–1297 (1998).
58. Sheehan, M. J., Campbell, P. & Miller, C. H. Evolutionary patterns of major urinary protein scent signals in house mice and relatives. *Mol. Ecol.* **28**, 3587–3601. <https://doi.org/10.1111/mec.15155> (2019).
59. Hurst, J. L. *et al.* Molecular heterogeneity in major urinary proteins of *Mus musculus* subspecies: potential candidates involved in speciation. *Sci. Rep.* **7**, 44992. <https://doi.org/10.1038/srep44992> (2017).
60. Smadja, C. & Ganem, G. Divergence of odorant signals within and between the two European subspecies of the house mouse. *Behav. Ecol.* **19**, 223–230 (2008).
61. Smadja, C. & Ganem, G. Subspecies recognition in the house mouse: A study of two populations from the border of a hybrid zone. *Behav. Ecol.* **13**, 312–320 (2002).
62. Bimová, B., Albrecht, T., Macholán, M. & Piálek, J. Signalling components of mate recognition system in the house mouse. *Behav. Proc.* **80**, 20–27 (2009).
63. Cerna, M., Kuntova, B., Talacko, P., Stopkova, R. & Stopka, P. Differential regulation of vaginal lipocalins (OBP, MUP) during the estrous cycle of the house mouse. *Sci Rep* **7**, 11674. <https://doi.org/10.1038/s41598-017-12021-2> (2017).
64. Thoss, M. *et al.* Regulation of volatile and non-volatile pheromone attractants depends upon male social status. *Sci. Rep.* **9**, 489. <https://doi.org/10.1038/s41598-018-36887-y> (2019).
65. Enk, V. M. *et al.* Regulation of highly homologous major urinary proteins in house mice quantified with label-free proteomic methods. *Mol. Biosyst.* **12**, 3005–3016. <https://doi.org/10.1039/c6mb00278a> (2016).
66. Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell Proteomics* **13**, 2513–2526. <https://doi.org/10.1074/mcp.M1113.031591> (2014).
67. Levy, M., Blacher, E. & Elinav, E. Microbiome, metabolites and host immunity. *Curr. Opin. Microbiol.* **35**, 8–15. <https://doi.org/10.1016/j.mib.2016.10.003> (2017).
68. Stopková, R., Otčenášková, T., Matějková, T., Kuntová, B. & Stopka, P. Biological roles of lipocalins in chemical communication, reproduction, and regulation of microbiota. *Front. Physiol.* <https://doi.org/10.3389/fphys.2021.740006> (2021).
69. Moudra, A. *et al.* Phenotypic and clonal stability of antigen-inexperienced memory-like T cells across the genetic background, hygienic status, and aging. *J. Immunol.* **206**, 2109–2121. <https://doi.org/10.4049/jimmunol.2001028> (2021).
70. Kreisinger, J., Čížková, D., Vohánka, J. & Piálek, J. Gastrointestinal microbiota of wild and inbred individuals of two house mouse subspecies assessed using high-throughput parallel pyrosequencing. *Mol. Ecol.* **23**, 5048–5060. <https://doi.org/10.1111/mec.12909> (2014).
71. Gowaty, P. A., Drickamer, L. C. & Schmid-Holmes, S. Male house mice produce fewer offspring with lower viability and poorer performance when mated with females they do not prefer. *Anim. Behav.* **65**, 95–103 (2003).
72. Bimova, B. V. *et al.* Reinforcement selection acting on the European house mouse hybrid zone. *Mol Ecol* **20**, 2403–2424. <https://doi.org/10.1111/j.1365-294X.2011.05106.x> (2011).
73. Bianchi, F. *et al.* Vertebrate odorant binding proteins as antimicrobial humoral components of innate immunity for pathogenic microorganisms. *PLoS ONE* **14**, e0213545. <https://doi.org/10.1371/journal.pone.0213545> (2019).
74. Fluckinger, M., Haas, H., Merschak, P., Glasgow, B. J. & Redl, B. Human tear lipocalin exhibits antimicrobial activity by scavenging microbial siderophores. *Antimicrob. Agents Chemother.* **48**, 3367–3372. <https://doi.org/10.1128/AAC.48.9.3367-3372.2004> (2004).
75. Witkowska-Banaszczak, E. & Długaszewska, J. Essential oils and hydrophilic extracts from the leaves and flowers of *Succisa pratensis* Moench. and their biological activity. *J. Pharmacy Pharmacol.* **69**, 1531–1539. <https://doi.org/10.1111/jphp.12784> (2017).
76. Boursot, P., Auffray, J. C., Britton-Davidian, J. & Bonhomme, F. The evolution of house mice. *Rev. Ecol. Syst.* **24**, 119–152 (1993).
77. Đureje, L., Macholán, M., Baird, S. J. & Piálek, J. The mouse hybrid zone in Central Europe: From morphology to molecules. *Folia Zool.* **61**, 308–318 (2012).
78. Stopkova, R., Klempt, P., Kuntova, B. & Stopka, P. On the tear proteome of the house mouse (*Mus musculus musculus*) in relation to chemical signalling. *PeerJ* **6**, e3541. <https://doi.org/10.7717/peerj.3541> (2017).
79. Kuntova, B., Stopkova, R. & Stopka, P. Transcriptomic and proteomic profiling revealed high proportions of odorant binding and antimicrobial defense proteins in olfactory tissues of the house mouse. *Front. Genet* **9**, 26. <https://doi.org/10.3389/fgene.2018.00026> (2018).
80. Ibarra-Soria, X., Levitin, M. O., Saraiva, L. R. & Logan, D. W. The olfactory transcriptomes of mice. *Plos Genetics* **10**, e1004593. <https://doi.org/10.1371/journal.pgen.1004593> (2014).
81. Strotmann, J. & Breer, H. Internalization of odorant-binding proteins into the mouse olfactory epithelium. *Histochem. Cell Biol.* **136**, 357–369. <https://doi.org/10.1007/s00418-011-0850-y> (2011).
82. Trinh, K. & Storm, D. R. Detection of odorants through the main olfactory epithelium and vomeronasal organ of mice. *Nutr. Rev.* **62**, S189–S192 (2004).
83. Stopka, P. *et al.* On the saliva proteome of the Eastern European house mouse (*Mus musculus musculus*) focusing on sexual signalling and immunity. *Sci Rep* **6**, 32481. <https://doi.org/10.1038/srep32481> (2016).
84. Yip, K. S., Suvorov, A., Connerney, J., Lodato, N. J. & Waxman, D. J. Changes in mouse uterine transcriptome in estrus and proestrus. *Biol Reprod* **89**, 13. <https://doi.org/10.1095/biolreprod.112.107334> (2013).
85. Shahan, K., Denaro, M., Gilmartin, M., Shi, Y. & Derman, E. Expression of six mouse major urinary protein genes in the mammary, parotid, sublingual, submaxillary, and lachrymal glands and in the liver. *Mol. Cell. Biol.* **7**, 1947–1954 (1987).
86. Stopková, R. *et al.* Mouse lipocalins (MUP, OBP, LCN) are co-expressed in tissues involved in chemical communication. *Front. Ecol. Evol.* <https://doi.org/10.3389/fevo.2016.00047> (2016).
87. Garcia, M. & Ravnigani, A. Acoustic allometry and vocal learning in mammals. *Biol. Lett.* **16**, 20200081. <https://doi.org/10.1098/rsbl.2020.0081> (2020).
88. Martínez-Ricos, J., Augustin-Pavon, C., Lanuza, E. & Martínez-García, F. Role of the vomeronasal system in intersexual attraction in female mice. *Neuroscience* **153**, 383–395 (2008).
89. Martínez-Ricos, J., Agustín-Pavón, C., Lanuza, E. & Martínez-García, F. Intraspecific communication through chemical signals in female mice: reinforcing properties of involatile male sexual pheromones. *Chem. Senses* **32**, 139–148 (2007).
90. Stowers, L. & Logan, D. W. Sexual dimorphism in olfactory signaling. *Curr. Opin. Neurobiol.* **20**, 770–775. <https://doi.org/10.1016/j.conb.2010.08.015> (2010).
91. Klein, S. L. The effects of hormones on sex differences in infection: From genes to behavior. *Neurosci. Biobehav. Rev.* **24**, 627–638 (2000).
92. Kadel, S. & Kovats, S. Sex hormones regulate innate immune cells and promote sex differences in respiratory virus infection. *Front. Immunol.* **9**, 1653. <https://doi.org/10.3389/fimmu.2018.01653> (2018).

93. Caspar, K. R. *et al.* Perioral secretions enable complex social signaling in African mole-rats (genus *Fukomys*). *Sci. Rep.* **12**, 22366. <https://doi.org/10.1038/s41598-022-26351-3> (2022).
94. Gregorová, S. & Forejt, J. PWD/Ph and PWK/Ph inbred mouse strains of *Mus m. musculus* subspecies: A valuable resource of phenotypic variations and genomic polymorphisms. *Folia Biologica (Praha)* **46**, 31–41 (2000).
95. Chang, P. L. *et al.* Whole exome sequencing of wild-derived inbred strains of mice improves power to link phenotype and genotype. *Mamm. Genome* **28**, 416–425. <https://doi.org/10.1007/s00335-017-9704-9> (2017).
96. Piálek, J. *et al.* Development of unique house mouse resources suitable for evolutionary studies of speciation. *J. Hered.* **99**, 34–44. <https://doi.org/10.1093/jhered/esm083> (2007).
97. Gentleman, R. C. *et al.* Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80. <https://doi.org/10.1186/gb-2004-5-10-r80> (2004).
98. Crawley, M. J. *The R Book* (Wiley Publishing, 2007).
99. Pavelka, N. *et al.* A power law global error model for the identification of differentially expressed genes in microarray data. *BMC Bioinform.* **5**, 203. <https://doi.org/10.1186/1471-2105-5-203> (2004).
100. Rodriguez, J., Gupta, N., Smith, R. D. & Pevzner, P. A. Does trypsin cut before proline?. *J. Proteome Res* **7**, 300–305. <https://doi.org/10.1021/pr0705035> (2008).
101. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32. <https://doi.org/10.1023/A:1010933404324> (2001).
102. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, Chem, 2016).

## Acknowledgements

PS was supported by MICOBION project funded from EU H2020 (No 810224). TM was funded by the Grant Agency of the Charles University, Prague (GAUK; No. 1191419). Housing of mice in Studenec was supported by the CAS within the program of the Strategy AV 21 and by CSF grant 16-23773S. Project utilized Proteomic Core Facility, BIOCEV, Faculty of Science, Charles University, Prague (supported by OP VaVpI CZ.1.05/1.1.00/02.0109) for the mass spectrometric measurements. The Czech Centre for Phenogenomics supported by the Czech Academy of Sciences RVO 68378050 and by the project LM2018126 Czech Centre for Phenogenomics provided by MEYS and CZ.02.1.01/0.0/0.0/16\_013/0001789 Upgrade of the Czech Centre for Phenogenomics: developing towards translation research by MEYS and ESIF. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We are very grateful to prof. Miloš Macholán for his comments on the first draft of this manuscript.

## Author contributions

R.S., P.S. and A.D. wrote the first draft of the manuscript, J.P. prepared mouse models MUS and DOM in the facility Studenec, and T.M. collected the samples of wMUS. R.Se. helped with the experimental design and manuscript writing. P.Ž. run GCxGC-MS instruments, P.T. run nLC-MS/MS and both helped to prepare final datasets. All authors participated in writing and have reviewed the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-35450-8>.

**Correspondence** and requests for materials should be addressed to P.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023