Dog Olfactory Gene Expression Profiling Using Samples Derived from Nasal Epithelium Brushing

Francis Galibert (Francis.galibert@univ-rennes1.fr)
Universite de Rennes 1
Naoual Azzouzi Sadequi
IGDR: Institut de Genetique et Developpement de Rennes
Anne-Sophie Guillory
IGDR: Institut de Genetique et Developpement de Rennes
Gilles Chaudieu
IGDR: Institut de Genetique et Developpement de Rennes

Research Article

Keywords:

Posted Date: January 21st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1231706/v1

Abstract

Dogs have an exquisite sense of olfaction. In many instances this ability has been put to use by humans in a wide range of important situations. It is also thought that some breeds have better senses of smell than others. Dogs can detect many components at a very limited concentration in air. To achieve such high levels of detection, the dog olfactory system is both complex and highly developed. The dog genome encodes a large number of olfactory receptor (OR) genes. However, it still remains unclear as to what extent are all of these OR genes expressed?

To address this, a nasal brushing method was developed to recover dog nasal epithelium samples from which total RNA could be extracted to prepare high quality cDNA libraries. After capture by hybridization with a large set of oligonucleotides, the level of expression of each transcript was measured after deep sequencing by next generation sequencing (NGS). The reproducibility of the sampling approach was checked by analyzing several samples from the same animals (up to 6, 3 per each naris). The quality of the capture was also checked by analyzing two DNA libraries, which offered the advantage over RNA libraries by having an equal presence of each gene. Finally, we compared this brushing method on live animal to a biopsy approach applied to two terminally ill dogs, euthanized following consent from the owner.

Comparison of the levels of expression of each transcript indicate that the ratios of expression between the most and the least expressed OR in each sample are > 10,000 (paralog variation) and that a number of OR genes are not expressed.

The method developed here will allow us to address whether variations observed for any of the OR transcriptomes relate to dog life experiences and whether any differences observed between samples are dog-specific or breed-specific.

Lay Summary Abstract

All animals living in the wild depend very much on olfaction to find food, sexual partners and escape to predators. Dogs as a descendant of wolves have inherited this ability, but in addition many dog breeds were subjected to strong selection practice aimed at increasing their ability to perform a number of tasks highly dependent upon olfaction.

Dogs like most of the other mammals are equipped with a large number of olfactory receptors displayed at the surface of the olfactory sensory epithelium. So far, only one study has compared the level of expression of these OR dog genes sampled from three dead mixed breed dogs to that of a coupled of mammals.

In this paper we developed a non-aggressive method able to recover olfactory sensory neuron samples from alive animals. Thus, this method would allow to analyze dog OR RNA profiling of many dogs and dog breeds and their variation, if any, following life experiences or raining, which is not possible nowadays with the current method of sampling made on animals dead for any reasons."

Introduction

Olfactory receptors were discovered 30 years ago by Buck and Axel, who identified a novel sub-family of GPCR (G protein coupled receptor) in rat olfactory epithelium (1). The importance of this discovery, which transformed the whole field of olfaction, was recognized by their award of the Nobel Prize for physiology or medicine in 2004. For all

wild animals, olfaction is a vital function. It participates critically in the foraging of food, to the selection of sexual partners, prevention to danger and to escape from predators. For humans, even if these functions are fulfilled differently, olfaction still remains an important function and people suffering from anosmia or even hyposmia are very much at a disadvantage or in pain.

Since, the discovery of several OR gene transcripts in rat olfactory epithelium, many animal olfactory gene repertoires have been identified through full genome DNA sequencing, showing that these genes represent the largest gene family with several hundred members scattered across many chromosomes (2–7).

However, the number of studies relating to gene transcription and expression studies in olfactory tissues is limited to decipherer the number of expressed genes (8–13) These studies showed that 90% of human and rat and 70% of dog OR genes are expressed although at very different levels with no or limited differences between males and females. They also showed that transcription is not limited to intact OR genes but that some pseudogenes are also expressed (14, 15). As long and widely recognized, olfaction is a very important and well-developed function in dogs. However, what makes some dogs so good at finding hidden objects, drugs or explosives or even at detecting that its owner is developing a melanoma? (16, 17). Moreover, many dog breeds, like hunting dogs, have been derived with a particular attention toward this function. It is thus important to know if any differences exist regarding OR expression between dog breeds and also between dogs and other mammals. With these questions in mind, as soon as the dog genome sequence was available (18), we determined the dog OR repertoire (15, 19) and the genetic polymorphism of several OR genes within a cohort of 48 dogs representing six different breeds known for their different olfactory capabilities (20, 21). The number of OR genes and the extent of their genetic diversity are important parameters in determining the olfactory capabilities of a mammal. However, variation in ORs expression levels as well as of the proteins implicated in the odorant transduction signal toward the brain represent major aspects of understanding and explaining individual differences. Analysis of the dog olfactory epithelium poses difficult problems; these include ethical and painless tissue sampling, anatomical issues given the large size of the olfactory epithelium which can be up to 200 cm^2 for an adult German Shepherd and thus be challenging to recover in its entirety (22).

Two ways of accessing the olfactory epithelium exist. The first one is following euthanasia of ill and incurable dogs or from deceased dogs following an accident. Sampling nasal epithelium following euthanasia, in principle, gives access to the whole olfactory epithelium (OE). However, relying on euthanasia only, presents limitations not just in the number of samples but also to the variety of dog breeds investigated; and even more importantly the circumstances relating each situation. In other words, restriction of nasal epithelium sampling to just instances of euthanasia only, will seriously restrict research aimed at analyzing the full variation of olfactory transcriptomes in response to a given situation such as breed, training and the age of the animals. A second option is by the gentle scraping of the nasal epithelium with the aid of a nasopharyngeal swab, similar to those used at the hospitals in Otorhinolaryngology (ORL) services. In an effort to develop an ethical and minimally invasive general method of sampling which can be utilized in many circumstances, we investigated a brush sampling approach and compared its utility.

Material And Methods

Olfactory epithelium samples

All samples of nasal epithelium were collected at the Clinique Vétérinaire (Pole Santé Chanturgue-63100 Clermont-Ferrand –France) under general anesthesia (Ketamine-Imalgene 1000®and xylazine - Rompun® - aa, 0,1 mL/kg IV, Isoflurane-O2) performed for chirurgical purposes by gentle brushing made with endocervical DOC cytobrushes (Medispo.com). Two of these samples, one from a Bichon and a second from a Golden Retriever were used for total mRNA sequencing. We collected also several samples from four dogs, a Belgian Shepherd, a West Highland White Terrier, a Whippet and a Labrador Retriever. All animals were being anesthetized for programmed surgical interventions and samples collected with owner consent and ethic committee approval. After recovery of the nasal epithelium, brushes were immediately placed into a tube containing 1.5 ml of RNA later solution and sent to the laboratory where they were stored at – 80C° until subsequent nucleic acid extraction.

In addition, with owner consent several biopsies were taken from two euthanized dogs, a Cane Corso and a Golden Retriever that were at the terminal phase of lung cancer. These biopsies were taken in order to sample the olfactory epithelium at different locations. These samples were processed as all the other samples described above.

Nucleic acid isolation

Total RNA was extracted and purified with the Nucleospin RNA kit (Macherey Nagel). Following titration with a Nanodrop spectrometer, the quality and purity of the RNA samples were assessed with a BioAnalyzer (Agilent 2100), and only the RNA samples with a RIN score of ≥ to 8 were kept for analysis. DNA extraction from a Golden Retriever and Cane Corso samples were made with the Nucleospin tissue kit.

Sample processing

The Bichon and Golden RNA samples were sent to the French Genomic Platform (Centre Inra Toulouse- Midi Pyrénées) for library construction and NGS analysis. Whereas the extracted RNA from the other samples were sent to Integragen.

Library construction

cDNA libraries were constructed as follows: Starting with 1µg of total RNA, polyA RNA molecules were purified using poly-T oligo attached to magnetic beads. The attached RNA molecules were then fragmented using divalent cations under elevated temperature to obtain approximately 300nt pieces, followed by library construction and addition of Illumina adapters. For each of the two DNA libraries, 600 ng of genomic DNA was fragmented by sonication and purified to yield fragments of 150-200 bp. Paired-end adaptor oligonucleotides from the NEB kit were then ligated followed by 8 PCR cycles.

Hybridization capture

Following amplification of the cDNA and DNA libraries, 120ng of each of them were hybridized to the SureSelect oligoprobe capture set made of 45,142 overlapping biotinylated oligonucleotides (120 mers) covering the complete OR open reading frames. Oligonucleotides were designed and synthesized by Agilent (23) from the sequences of 999 OR genes and pseudogenes and 19 control gene sequences identified in Canfam3.1 (24).

Sequence data analysis

Sequencing was made on a HIseq 4000 sequencer (Illumina, 2x125nt) using the v4 chemistry/HBS Hiseq kit. One line was used for each sample to produce up to $300x10^6$ reads per sample. Image analysis and base calling were performed using Illumina Real-Time Analysis software version 2.7.3 with default parameters. Raw sequence data produced by the Genotoul platform and Integragen were sent to the laboratory on floppy disks for processing and

analysis. The two sequence extremities were first trimmed to remove the remaining primer sequences and any bases with poor quality base calling often present at the extremities of the reads with Cutadapt (25). The trimmed sequences were then aligned with STAR.v2 through Galaxy (Sigenae) (25). The resulting BAM files were analyzed with Samtool, Bedtool, Cufflink and Stringtie (26) on the Toulouse Genocluster and with the Geneious suite (27). Statistical analysis of the data and Heatmap constructions based on the FPKM values were made with the Manhattan and Ward method using R language by 'in house' written lines of commands (28).

Results

OSN transcriptome analyzes

Two samples of nasal epithelium tissues were obtained as described above. One sample was from a male Bichon, the second sample was from a female Golden Retriever, both aged of 8 years. The two samples were processed as described in the methods section. As we anticipated that the olfactory neurons might be contaminated by other cells, thus reducing the level of neuron-specific transcripts and the OR transcripts, each sample was deeply sequenced and up to 300 million reads obtained, maximizing the chances of capturing transcript differences from poorly expressed genes. A number of reads to be compared to the 60 million reads previously reported for murine neurons (10). As shown in Table 1, approximately 90% of the reads could be mapped at unique positions. This high percentage of mapping resulted from the good quality of the RNA, libraries and the sequencing itself. Given this, we therefore believe the two figures 0.27 and 0.31, representing the percent of mismatches between the sequence reads aligned onto the reference genome (CanFam3.1), are indicative and representative of any slight polymorphism differences

Analysis of the sequence data with the Geneious suite (27) allowed us to identify many genes and to calculate their respective FPKM values, i.e. the number of reads corresponding to each transcript, a metrics defining the abundance of all transcripts (Additional data files 1a, b) and OR transcripts (Additional data file 2). For these two samples, despite the very deep sequencing strategy only 14% and 16% only of the OR gene transcripts were detected with an FPKM > 0.1, corresponding to 112 Bichon and 104 Golden Retriever OR genes respectively, percentages to be compared to 90% for the OR murine and human repertoires and the 70% for canine repertoire (10, 13).

As summarized in Table 2, 88% and 90% of all the annotated genes (ENSEMBL.org) and 62 and 56% of the non-annotated genes are expressed at a detectable level. This high percentage of expressed genes is probably due to the composition of the samples made of several cellular cell types. Based on their respective FPKM values the 10 most expressed identified genes are listed in Table 3. A comparison of the dog gene expression ranks with their murine orthologs (10), indicates strong differences. None of the highly expressed dog genes was found to be strongly expressed in the murine tissue. SCGB1A1, the most highly expressed gene in the Bichon sample, is not even detected in a murine sample; VMO1 at the second position in the two dog samples ranks at position 1498 in the mouse sample and the same applies to TAGNL2 at position 6750. Similarly, the olfactory major protein (OMP), a protein characteristic of the olfactory tissues, and the 3rd most expressed transcript in murine OSN, ranks at positions 2502 and 1796 in the dog samples (Table 4). The Gα sub unit of the G(olf) protein encoded by *GNAL*, a key protein in the transduction pathway, ranks at position 9 in the mouse sample and 10680 and 8767 in the dog samples. These large differences in gene ranking observed between the dog and murine samples, strongly suggests that in the dog samples the olfactory sensory neurons (OSN) were heavily contaminated by adjacent cellular types,

such as the supporting cells diluting the expression of the OSN genes and comparatively increasing that of non-OSN genes (29, 30).

OR Targeted transcriptome analysis

Reliability and reproducibility of the approach

In this series of experiments, we analyzed 14 samples from four dogs only, i.e. two to six per dog, to appreciate the reproducibility of the samplings, as well as the difference that could exist between the right and left nostrils. As shown in Additional data file 3, a strong correlation does exist between the different FPKM values obtained for the same dog, either if one compares the right and left nostril samples or the different samples from the same nostril. In the heat map presented in Fig. 2 hierarchical clustering is seen for the four Labrador Retriever samples, for the two West Highland White Terrier samples and for four out of the five Whippet samples. A discrepancy was seen for the two Belgian Shepherd samples which are not grouped, suggesting a sampling or a sequencing problem. Alternatively, a particular physiological condition could have induced a different transcription profile of one of the two nostrils of this dog (31). Nevertheless, the coherent grouping of a large majority of the samples indicated good reproducibility of the sampling itself and the good quality of the sequencing. Moreover, the grouping of the different samples of each dog indicates, within the limit of the experimental procedure, the absence of difference between the right and left nostrils.

Hybridization capture effect

To appreciate to which extent, the hybridization capture may have biased the FPKM values of the RNA transcripts, we captured by hybridization with the same set of oligonucleotides, the OR gene sequences of two DNA libraries, made from two dog samples, a Cane Corso and a Golden Retriever. As each gene is present in two copies in any sample, the FPKM values of each pair of ortholog genes (i.e. the same gene in the Cane Corso and the Golden Retriever) should be identical or similar. The FPKM values obtained for the different OR genes are shown in Additional data file 4. As shown in the graphs, the reproducibility of the capture is excellent, the blue and red dots corresponding to a pair of orthologs are confounded or very close for a large majority of them with a correlation of 0.972 as calculated with the Pearson test.

However, not all OR genes of the same sample are captured with the same efficiency (32) (Additional data file 4): CfOR1812 and CfOR0039p genes, for example, were very efficiently captured leading to 1363 and 1357 FPKM values respectively. In contrast, the capture of CfOR0183 and CfOR0783p genes leads to much lower FPKM values 83.4 and 95.5 respectively. Since each gene is present in two copies, these differences in FPKM values between different paralog OR genes are mostly due to differences in oligonucleotide hybridization efficiency, which are dependent upon the sequences themselves (32). Given these results, to obtain a more realistic view of the OR transcriptome, the FPKM values of the RNA transcripts were normalized (Additional data file 5). This was done by conditioning each crude RNA FPKM value (Additional data file 3) by a factor corresponding to the ratio of the FPKM values of this gene, as obtained by the analysis of the two DNA gene libraries.

One of the first observations of the data summarized in Additional data file 5, is the large extended range of expression regardless of the samples used. Values ranged from above several thousand FPKM for the most expressed OR down to 0.1 for the least expressed. These data indicate that a variable, but large, proportion of the OR genes are not even detectable, having (if transcribed) an FPKM value below 0.1. If one concentrates on the OR genes having an FPKM value $\geq$ 1% of the most expressed OR gene of the sample, then these results are even more

surprising. About 30 genes are above this limit, as already observed for the Bichon and Golden Retriever of which their data were not normalized and for which we obtained 39 and 37 OR genes above this limit.

To address the issue of whether the normalization made was appropriate and correct we compared the list of 20 most expressed Bichon and Golden OR genes (Additional data files 1a and b) with that the 20 most expressed OR genes of the West Highland White terrier, Whippet and Labrador (Additional data file 5) and noticed that 9 out of these 20 genes are present in all samples, indicating that the correction made was correct.

A further issue to resolves was whether the rather low proportion of OR being expressed reflected the truth or was a consequence of a capture effect and/or of the variable number of sequencing reads per sample.

To address this, we plotted the DNA FPKM values of the Cane Corso and Golden Retriever samples in blue and the FPKM values of the corresponding RNA in red (Additional data files 6 a and b). As shown in the graphs, no correlation existed between the DNA and RNA FPKM values of any genes: e.g. Golden Retriever CfOR 12F06 or CfOR 0268 genes. Furthermore, a large number of genes were not transcribed at a detectable level, whereas their cognate genes were well captured by the same set of oligonucleotides. Thus, the fact that a large number of OR transcripts (660 Golden Retriever and 742 Cane Corso) was not detected cannot be due to the failure of the hybridization capture. Whereas the number of reads could have impacted the number of expressed OR genes, the comparison of the ratios and the plots detailed in Fig. 3 (a-d), indicated no correlation between the FPKM values and the number of expressed OR genes. Similarly, the number of OR genes having an FPKM value ≥ 0.1 in the different Whippet samples is not affected by the FPKM value of the most highly expressed OR. Around 370 OR were detected in any Whippet samples whereas the highest FPKM values varied between1,3249 for sample L1 and 21,545 for sample L2 (Additional data file 5).

Spatial segregation of OR expression

Spatial segregation of OR gene expression within the olfactory epithelium has already been documented although this remained to what extent a similar situation exists in dogs. Thus, we considered whether the limited number of OR genes number expressed in our samples was a consequence or not of this spatial segregation (19). To approach this question, four and six biopsy samples were taken from two dogs, a Cane Corso and a Golden Retriever respectively (Figure 4). In Additional data file 7 are given the FPKM values of the different OR transcripts. These data files show that up to 414 and 512 OR genes are not expressed in the Cane Corso and Golden Retriever samples respectively.

Discussion

The dog genome contains a large number of OR genes (4), able to explain in part, the great variety of volatile components a dog can recognize and how in a complex environment recognize an odor to which it has been trained. To be effective the OR genes have to be transcribed and expressed. Of these two aspects, we presently know very little. Several transcriptome studies have analyzed the spectrum of OR expressed in humans, rats, and mouse (8–12) but until recently very few studies have been done with dogs (13). A major reason has been the difficulties and problems associated with collecting appropriate samples.

As shown with the Bichon and the Golden retriever samples, a gentle brushing of the nasal epithelium allows us to recover sufficient olfactory neurons to extract their total RNA content and to perform a transcriptomic analysis. However, the relative quantification of the specific transcripts of the olfactory neurons such as those of the OMP or

the Gα subunit of the G(olf) protein, indicates that these two samples were heavily contaminated by other cells such as the supporting cells (29, 30). The main consequence of this heavy contamination is a dilution of the OSN transcripts. Thus, the OMP mRNA transcript, the 3rd most expressed transcript in murine OSN (10), ranks at positions 2502 and 1796 in the dog samples and many of the OR transcripts are even not detected in spite of a very deep sequencing analysis.

This contamination problem of the dog samples obtained by brushing limits their transcriptome analysis. First, it decreases the accuracy with which the FPKM values are obtained, second, it limits the number of detected genes and third this approach is expensive in terms of sequencing cost. To bypass this problem, two possibilities were envisaged. The first one relates to the purification of the OSN fraction by immunoprecipitation with a specific antibody. The second strategy, which we preferred for its easiest setting and the lack of appropriate antibody, relayed on hybridization and capture of the sequences of interest with a set of oligonucleotides.

Thus, in order to circumvent this problem, in another series of analyses, the cDNA libraries before sequencing were subjected to a capture of the entire sequences of all the canine OR genes by hybridization with a large panel of long and overlapping oligonucleotides. In order to check the reproducibility and efficiency of the capture itself, we also prepared two DNA libraries from samples taken from a Cane Corso and a Golden Retriever. Using the same set of oligonucleotides, we captured the OR gene nucleotide sequences and sequenced them. As each gene is present in two copies, we anticipated that (a) the FPKM values of each gene should be similar or identical in the two libraries and (b) the comparison of the FPKM of the different OR genes within each library will reveal the efficiency of the capture method itself.

In Additional Data file 4, we plotted the calculated FPKM values, in blue for the Cane Corso sample and in red for the Golden Retriever. As shown in this file, most of the blue and red dots are confounded or very close, indicating a good reproducibility of the capture itself, as supported by the Pearson correlation test.

However as seen in Additional Data file 4, comparison of the FPKM of the different OR gene within each of the two sample indicates a large difference, confirming the previous observation that the efficiency of hybridization capture is in part sequence dependent (32). This dependency does not prevent comparison between samples and the tight clustering (11 out of 14 samples) of the captured sequences based on their FPKM values indicates a good reproducibility of the entire process i.e. from the recovery of the samples up to the sequencing itself (Fig. 2). To the limit of the techniques used, this grouping also indicated no major differences between the level of expression of the mRNA extracted from left and right naris of the same animal.

However, in order to compare the transcription profile of the different paralog OR transcript within a sample a correction factor for crude FPKM values should be applied. Since each gene is present twice in any genome, the FPKM ratio of the two OR genes (Additional data file 4) of a pair is a consequence of the efficacy difference of the capture itself for each of these two genes. As explained in the result section (Additional data file 5), a correction was applied by affecting each crude RNA FPKM value by a factor corresponding to the ratio of the FPKM values of this gene to the most expressed gene of the sample calculated from the DNA gene libraries.

Previous studies have been made regarding the transcriptome analysis of several mammal olfactory tissues (8–13) These studies have shown that nearly all OR genes would be expressed at a detectable level. In contrast to this, in our study, it appears that far less dog OR dog gene transcripts are detected. Interestingly, no more than 30 genes reach an FPKM value ≥ 1% of the OR the most expressed in the sample (Additional data file 5). As shown in Fig. 3, we observed no correlation between the highest FPKM values in any sample and the number of expressed OR

genes in the corresponding sample and in all samples, we observed a very large range of expression as much as 10,000 times.

The absence of correlation between the FPKM values of the DNA genes and the RNA transcripts, indicates the low number of expressed genes is not a consequence of a failure of the hybridization capture but might correspond to a characteristic of the canine RNA olfactory profile, in at least of the samples analyzed (Additional data file 6),). Given this we considered whether the relative low number of OR gene being expressed could be due to the sampling itself, as a consequence of a strong regionalization of the expression of the different OR genes all along the canine olfactory epithelium, as previously observed with rat (19). To tackle this question, we prepared several samples representing different site locations of a Cane Corso and Golden Retriever olfactory epithelium from dogs compassionately euthanized for highly advanced cancer. As shown in the data presented in Additional data file 7, although the RNA profiles of the different biopsies are not strictly identical in either of the two animals, very importantly, up to 512 and 414 OR genes (i.e. 56 and 46% of the whole set of OR genes) are not detectable whatever the samples and their localization in the OE and up to 40% are silent when one combines the data of the 10 biopsy samples. At present we have no explanation regarding the much larger number of expressed OR genes found by Saraiva et al who reported that only 14% of OR genes were not detected (13). This could be due to breed differences sampled, (in the case of Saraiva et al it was a mixed breed). Nevertheless, it is important to keep in mind, that the absolute number of observed expressed OR genes, is probably less important in characterizing the RNA profiling of any species than the range to which the genes are expressed. It is unclear what meaning in biological terms such low expression of genes may represent. Whatever the issue of the absolute number of dog OR genes expressed, it is very important to consider the observation that the ratio of expression of the human and mouse OR genes RNA is much lower (8, 12) to that found by Saraiva (13) and from our study reported here.

Conclusion

The focus of this study was aimed at evaluating a non-invasive sampling approach which could be ethically and practically used in research applications study the olfactome of dogs across a wide range of different variables including breed diversity, age, behavioral conditioning and environmental situations.

Our approach, was inspired by human otorhinolaryngology routine practice and the need to establish a veterinarian led procedure which is easy to establish and does not inflict pain or impact on animal wellbeing. The data we present here support that this is a valid and robust qualitative and quantitative approach to investigating expression profiles and its variation. The different samples obtained by nasal brushing and biopsy are highly similar to each other. Up to approximately half of dog OR genes appear to be silent or not detectable although this needs to be established in sufficient sample sizes across a wide range of breeds. There is also a significant number of genes where their expression is either very low or exceptionally high. The absolute number of genes expressed genes may not necessarily represent an important biological parameter given the low level of expression of some genes. A concerted international effort is now required to investigate and characterize the dog olfactome and determine is impact on canine health and welfare.

Abbreviations

OR olfactory receptor

NGS next generation sequencing

GPCR G protein coupled receptor

OE olfactory epithelium

Bp base pair

FPKM fragment per kilobase million

OMP olfactory major protein

OSN olfactory sensory neuron

Declarations

Ethics approval and consent to participate

Samples were collected by DV G. Chaudieu at the Clinique Vétérinaire (Pole Santé Chanturgue-63100 Clermont-Ferrand –France) before surgical operation. The work was approved by the CNRS ethical board, France (35-238-13) for UMR6290 and made in compliance with file R.214- 89 (CRPM). The Dr.Vet asked the owner whether he would accept or not that few cells would be recovered with a cytobrush within the nasal cavity of this/her dog during the anesthesia which was done for the purpose of a surgical intervention. Only when the owner agreed cells were obtained and sent to the laboratory for RNA extraction.

DATA AVAILABILITY

All data will be freely available upon publication. Any further request could be obtained from the corresponding author galibert@univ-rennes1.fr.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

FUNDING

Funding were obtained from the CNRS (Centre National de la Recherche Scientifique) and the University of Rennes (UR1), France, through the annual recurrent budget. No grant was solicited neither obtained for this work.

Author's contribution

NA extracted the nucleic acids, analyzed the data, made the tables, figures and Additional data files and participated in the writing of the paper, GC provided the samples, A.S.G provided the biopsy samples and FG conceived the study, analyzed the data, wrote the paper and got the funding. All authors read and participated to its final edition.

ACKNOWLEDGEMENT

We are grateful to the Genotoul bioinformatics platform Toulouse Midi-Pyrenees (Bioinfo Genotoul) for providing computing resources and to Clemence Genthon and Olivier Bouchez from the Genomic Plateform INRA Auzeville.

References

1 Buck L, Axel R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. Cell. 1991; 5;65(1):175-87.

2 Glusman G,Yanai I, Rubin L, Lancet D. The complete human olfactory subgenome. Genome Res. 2001;11(5):685-702

3 Godfrey PA, Malnic B, Buck L. The mouse olfactory receptor gene family Proc.Natl.Acad.Sci.USA. 2004; 101 (7):2156-6

4 Quignon P, Giraud M, Rimbault M, Lavigne P, Tacher S, Morin E, Rtout E; Valin AS, Lindblad-Toh K, Nicolas J, Galibert F. The dog and rat olfactory receptor repertoires.Genome Biol. 2005; 6 (10):R83

5 Montague MJ, Gang L, Gandolfi B, Khan R, Aken BL Searle SMJ, Minx P et al. Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. Proc.Natl.Acad.Sci. USA. 2014;111(48):17230-35

6 Niimura Y,Matsui A, Touhara K. Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. Benome Res. 2014; 24(9) 1485-96.

7 Azzouzi N, Barloy-Hubler F, Galibert. Inventory of the cichlid olfactory receptor gene repertoires: identification of olfactory genes with more than one coding exon. BMC Genomics. 2014 Jul 11;15(1):586. doi: 10.1186/1471-2164-15-586.PMID: 25015101

8 Rimbault M, Robin S, Vaysse A, Galibert F. RNA profiles of rat olfactory epithelia: individual and age-related variations. BMC Genomics. 2009 Dec 2;10:572. doi: 10.1186/1471-2164-10-572.

9 Ibarra-Soria X, Levitin MO, Saraiva LR, and Logan DW. The Olfactory Transcriptomes of Mice. PLoS Genet. 2014 Sep; 10(9): e1004593.Published online 2014 Sep 4. doi: 10.1371/journal.pgen.1004593

10 Kanageswaran N, Demond M, Nagel M, Schreiner, Sabrina Baumgart B.S.P et al. Deep Sequencing of the Murine Olfactory Receptor Neuron Transcriptome. PLoS One. 2015; 10(1): e0113170. PMCID: PMC4295871 Published online 2015 Jan 15. doi: 10.1371/journal.pone.0113170

11 Wang Z, Zhou Y, Luo Y, Zhang J, Zhai Y,Yang D, Zhang Z, Li Y, Storm DR et al. Gene Expression Profiles of Main Olfactory Epithelium in Adenylyl Cyclase 3 Knockout Mice. Int J Mol Sci. 2015 Dec; 16(12): 28320–28333. PMCID: PMC4691054 Published online 2015 Nov 30. doi: 10.3390/ijms161226107

12 Olender T, Keydar I, Pinto JM, Tatarskyy P, Alkelai A, Chien MS et al. The human olfactory transcriptome. BMC Genomics. 2016; 17: 619. PMCID: PMC4982115 Published online 2016 Aug 11. doi: 10.1186/s12864-016-2960-3

13 Saraiva LR, Riveros-McKay F, Mezzavilla M, Abou-Moussa EH,. Arayata CJ. et al. transcriptomic atlas of mammalian olfactory mucosae reveals an evolutionary influence on food odor detection in humans *Sci. Adv.* 2019; 5: eaax0396 31 July 2019

14 Crowe ML,Perry BN Connerton JF. Olfactory Receptor-Encoding Genes and Pseudogenes Are Expressed in Humans. Gene1996 Mar 9;168 (2), 247-249

15 Prieto-Godino LL, Rytz R Bargeton B, Abun L, Arguello R, Peraro MD, Benton R. Olfactory receptor pseudo-pseudogenes. *Nature* volume 539, pages93–97(2016).

16 Campbell LF, et Farmery L, Creighton-George SM, Farrant PBJ Canine Olfactory Detection of Malignant Melanoma BMJ Case Rep 2013 Oct 14. DOI: 10.1136/bcr-2013-008566

17 Willis CM, Britton LE, Swindells MA, Jones EM, Kemp AE, Muirhead NL et al. Invasive melanoma in vivo can be distinguished from basal cell carcinoma, benign naevi and healthy skin by canine olfaction: a proof of principle study of differential volatile organic compound emission. Br J Dermatol. 2016 Nov;175(5):1020-1029. doi: 10.1111/bjd.14887. Epub 2016 Sep 11.

18 Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, et al. Nature. 2005 Dec 8;438(7069):803-19. Genome sequence, comparative analysis and haplotype structure of the domestic dog.

19 Quignon P, Kikness E,Cadieu E, Touleimat N, Guyon R, Renier C et al. Comparison of the canine and human olfactory receptor gene repertoires. Genome Biol. 2003; 4(12): R80.

20 Tacher S, Quignon P, Rimbault M, Dreano S, Andre C, Galibert F. Olfactory receptor sequence polymorphism within and between breeds of dogs. J Hered. 2005;96(7):812-6. Epub 2005 Oct 26.

21 Robin S, Tacher S , Rimbault M, Vaysse A, Dréano S, André A, Hitte C, Galibert F. *BMC Genomics* 2009, **10**:21 doi:10.1186/1471-2164-10-21Genetic diversity of canine olfactory receptors

22 Quignon P and Galibert F. in Canine Olfaction Science and Law Ed by Jezierski et al. CRC ptess, chapter 4 Genetics of canine olfaction chap pp 39-48

23 Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol 27, 182-189 (2009).

24 CanFam3.1/Ensembl/release-101/hisat2/Canis_lupus_familiaris.CanFam3.1.6

25 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 2013 29, 1, 15–21, https://doi.org/10.1093/bioinformatics/bts635

26 Li H, Handsaker B,Wysoker A, Fennell T, Ruan J, Homer N, Marth, G. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics, 2009 25 (16), 2078-9.

27 Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 2012 28 (12), 1647– 1649, DOI: 10.1093/bioinformatics/bts199.

28 Strauss T, Michael Johan von Maltitz MJ. Generalising Ward's Method for Use with Manhattan Distances. Plos One2017 2017 Jan 13;12(1):e0168288.doi: 10.1371/journal.pone.0168288. eCollection 2017.

29 Farbman AL,Brunjes PC, RentfroL, Michas J,Ritz S 1988.The effect of unilateral l naris occlusion on cell dynamics in the developing rat olfactory epithelium. the journal of neuroscience 8(9),3290-3295)

30 Kavoi B, Makanya A, Hassanali J, Carlsson HE, Kiama S. Comparative functional structure of the olfactory mucosa in the domestic dog and sheep Ann Anat. 2010;192(5):329-37. doi: 10.1016/j.aanat.2010.07.004. Epub 2010 Aug 5.

31 Coppola DM, Waggener CT. The effects of unilateral naris occlusion on gene expression profiles in mouse olfactory mucosa. J Mol Neurosci. 2012 Jul;47(3):604-618

32 Hoppman-Chaney N, Peterson LM, Klee EW, Middha S, Courteau LK, Ferber MJ. Evaluation of Oligonucleotide Sequence Capture Arrays and Comparison of Next-Generation Sequencing Platforms for Use in Molecular Diagnostics. Clin Chem 2010 56 (8)1297–1306.

Tables

Table 1 Summary of RNA-Seq data

	Read number	Read length	Uniquely mapped %	Mismatch %
Golden Retriever	299,595,502	263	88.79	0.27
Bichon	291,999,387	245	89.99	0.31
Murine OSN [10]	58,234,129	Not documented	59.28	Not documented

The percentage of mismatch noted for the Bichon and Golden Retriever samples were obtained by comparison of their sequences to that of the reference dog genome (CanFam 3.1 Ensembl.org 2011). The percentage of uniquely mapped genes were calculated with STAR.v2(24).

Table 2 Percentage of known and unknown genes

The known genes are genes identified and named in the ENSEMBL.org database. The unknown genes correspond to transcripts covering sequences with the characteristics of coding genes but not identified as such yet and quoted “novel gene” in the database. FPKM (Fragments Per Kilobase of Exon per Million of Fragments Mapped) were calculated with GENEIOUS (26).

	Known genes	% of genes with FPKM >0.1	Unknown genes	% of unknown genes with FPKM >0.1	Highest FPKM
Bichon	14.540	88	13.159	62	62.980
Golden Retriever	14.554	90	16.230	56	52.340

Table 3 The 10 most highly transcribed genes

Genes ranking at the 10 first positions but present in one sample only are underlined.

BICHON			GOLDEN RETRIEVER		
Gene Name	FPKM	Description	Gene Names	FPKM	Description
SCGB1A1	62,98	Secretoglobin family 1A member 1	TFF1	52,34	trefoil factor 1
VMO1	56,515	vitelline membrane outer layer 1 homolog	VMO1	47,833	vitelline membrane outer layer 1 homolog
TAGLN2	41,031	Transgelin 2	ENSCAFG00000030140	46,288	Novel gene
ENSCAFG00000009876	39,431	Novel gene	ENSCAFG00000009876	42,706	Novel gene
BPIFA1	39,131	BPI fold containing family A member	B2M	31,104	Beta-2-microglobulin
GSTM4	23,355	GLUTATHIONE S TRANSFERASE MU 4	TAGLN2	26,382	Transgelin 2
GSTM3	22,523	GLUTATHIONE S TRANSFERASE MU 3	CSTB	24205	Cystatin B
CYP2A13	18820,5	Cytochrome P450 family 2 subfamily A polypeptide 13	GAPDH	23,532	Glyceraldehyde-3-phosphate dehydrogenase
UBB	15,585	Ubiquitin B	FTH1	22,835	Ferritin heavy chain Ferritin heavy chain, N-terminally processed
GAPDH	15,574	Glyceraldehyde-3-phosphate dehydrogenase	GPX2	21,69	Glutathione peroxidase 2

Table 4 Level of expression of a couple of key OSN transcripts.

This table shows that the ratio of the level of expression between the most expressed OR and the a subunit of the Golf (GNAL) differs according to the species. In the case

	OMP/ Rank	OMP/FPKM	Highest OR/Rank	Highest OR/FPKM	GNAL/Rank	GNAL/FPKM
Bichon	1,796	173	8,767	10	4,415	50
Golden Retriever	2,502	152	10,680	10	4,167	68
Murine OSN	3	1,185	9	1,072	380	89
Murine OE	16	1,370	55	399	1,427	35

Figures

Figure 1

Schematic view of the localizations of the Cane Corso and Golden Retriever euthanized dog

Samples made by the brushing approach correspond to positions 3 In this figure

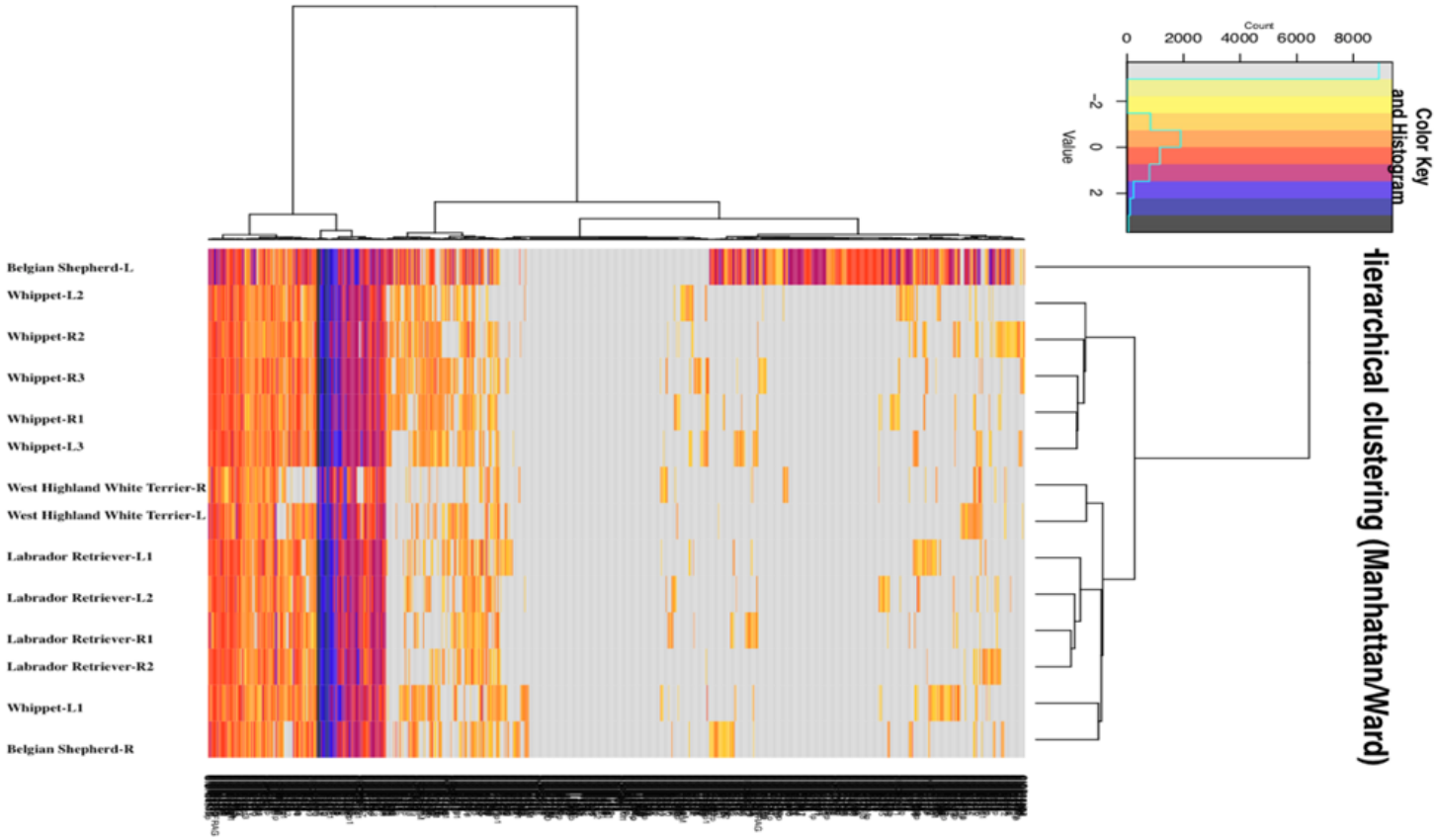

Figure 2

Multiple samples clustering

Hierarchical clustering constructed with all OR FPKM log values, that correspond to the number of times each transcript was sequenced, which itself depend on their concentration in the libraries (Additional data file 3). Most of the samples from the same animal are in the same cluster, the main exception are the two Belgian Shepherd samples, one of them being clustered with the Labrador Retriever samples, the second being alone. This overall good clustering indicates the good reproducibility of the sampling and of the analysis. The distribution of the two Belgian Shepperd samples in two different clusters suggests a pathological problem (30)

	Columns			
	1	2	3	4
Breeds	FPKM max	# >1%	# >1‰	>0.1
West Highland White Terrier-R	4876,85	29	94	242
West Highland White Terrier-L	4559,58	13	41	301
Whippet-R1	19821,35	22	78	366
Whippet-R2	26811,21	18	63	383
Whippet-R3	18838,6	24	82	369
Whippet-L1	21545,87	19	70	373
Whippet-L2	17589,41	24	78	353
Whippet-L3	21427,53	21	81	366
Belgian Shepherd-R	17892,72	20	71	347
Belgian Shepherd-L	32152,87	19	211	702
Labrador Retriever-R1	12945,53	23	94	310
Labrador Retriever-R2	15040,52	29	90	312
Labrador Retriever-L1	18336,93	22	73	315
Labrador Retriever-L2	13659,01	28	95	344

2a

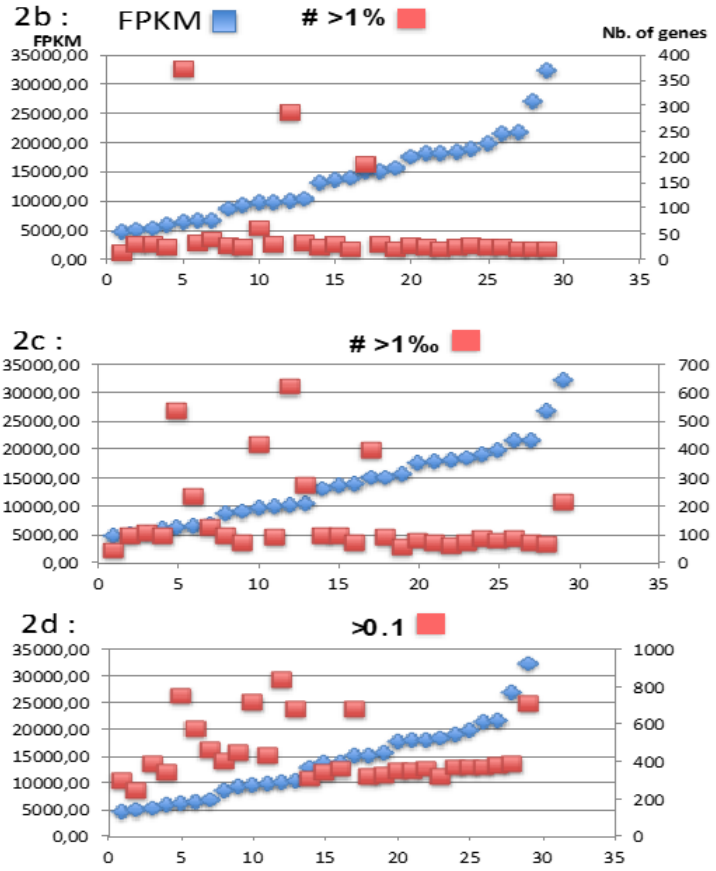

Figure 3

Absence of correlation between the FPKM values and the number of expressed OR

This figure is made of four Tables (2a to 2D). Table 2a for each sample, the FPKM values of the most highly expressed OR (column 1). In column 2 are the number of OR expressed at a FPKM value ³ to 1%. Column 3 indicates the number of OR expressed with an FPKM value ³ to $1^0/_{00}$ and column 4 gives the total number of detected OR (FPKM ³ 0.1). To check whether the FPKM values to which the highest expressed OR are detected impact the number of detected OR, we compared the number of detected OR (column 2, 3 or 4) to that of the highest FPKM values (1). As shown in plots 2b, c and d there is no correlation between the highest FPKM values and the number of expressed OR genes indicating that the low number of detected OR is a reality and might be a characteristic of dogs.

Image not available with this version

Figure 4

This image is not available with this version.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionaldatafile1a.xlsx
- Additionaldatafile1b.xlsx
- Additionaldatafile2.xlsx
- Additionaldatafile3.xlsx
- Additionaldatafile4.xlsx
- Additionaldatafile5.xlsx
- Additionaldatafile6a.xlsx
- Additionaldatafile6b.xlsx
- Additionaldatafile7.xlsx