Remote Medical Scent Detection of Cancer and Infectious Diseases With Dogs and Rats: A Systematic Review

Pierre Bauër, PhD[1,2]†, **Michelle Leemans, PhD**[2]† , **Etienne Audureau, MD, PhD**[2],
Caroline Gilbert, DVM, PhD[3,4], **Carole Armal**[1], and **Isabelle Fromantin,
Nurse PhD**[1,2]

Abstract

Background: Remote medical scent detection of cancer and infectious diseases with dogs and rats has been an increasing field of research these last 20 years. If validated, the possibility of implementing such a technique in the clinic raises many hopes. This systematic review was performed to determine the evidence and performance of such methods and assess their potential relevance in the clinic. **Methods:** Pubmed and Web of Science databases were independently searched based on PRISMA standards between 01/01/2000 and 01/05/2021. We included studies aiming at detecting cancers and infectious diseases affecting humans with dogs or rats. We excluded studies using other animals, studies aiming to detect agricultural diseases, diseases affecting animals, and others such as diabetes and neurodegenerative diseases. Only original articles were included. Data about patients' selection, samples, animal characteristics, animal training, testing configurations, and performances were recorded. **Results**: A total of 62 studies were included. Sensitivity and specificity varied a lot among studies: While some publications report low sensitivities of 0.17 and specificities around 0.29, others achieve rates of 1 sensitivity and specificity. Only 6 studies were evaluated in a double-blind screening-like situation. In general, the risk of performance bias was high in most evaluated studies, and the quality of the evidence found was low. **Conclusions:** Medical detection using animals' sense of smell lacks evidence and performances so far to be applied in the clinic. What odors the animals detect is not well understood. Further research should be conducted, focusing on patient selection, samples (choice of materials, standardization), and testing conditions. Interpolations of such results to free running detection (direct contact with humans) should be taken with extreme caution. Considering this synthesis, we discuss the challenges and highlight the excellent odor detection threshold exhibited by animals which represents a potential opportunity to develop an accessible and non-invasive method for disease detection.

Keywords

medical scent detection, VOCs (=volatile organic compounds), diagnostic, screening, cancer, infectious disease, smell, odor

1. Background

1.1.The Burden of Cancer and Infectious Diseases Worldwide

Cancer and infectious diseases are considered major health issues among men and women and are among the most common cause of morbidity and mortality worldwide. On the one hand, there were an estimated 18.1 million (95% UI: 17.5-18.7 million) new cases of cancer (17 million excluding non-melanoma skin cancer) and 9.6 million (95% UI: 9.3-9.8 million) deaths from cancer (9.5 million excluding non-melanoma skin cancer) worldwide in 2018.[1] On the

[1]Institut Curie, Paris, France
[2]Univ Paris Est Creteil, INSERM, IMRB, Team CEpiA
[3]Muséum National d'Histoire Naturelle, Brunoy, France
[4]Ecole nationale vétérinaire d'Alfort, Maisons-Alfort cedex, France

†These authors contributed equally to this work and share the first authorship

Corresponding Authors:
Michelle Leemans, Univ Paris Est Creteil, INSERM, IMRB, Team CEpiA, 61 Av. du Général de Gaulle, 94000 Créteil, F-94010 Créteil, France.
Email: michelle.leemans91@gmail.com

Pierre Bauër, Institut Curie, Ensemble hospitalier, Unité Recherche Plaies et Cicatrisation, 26, rue d'Ulm, 75005 Paris, France.
Email: pierre.bauer.1@gmail.com

other hand, infectious diseases such as urinary tract infections (UTI), tuberculosis, Clostridium difficile infections, Methicillin-resistant Staphylococcus aureus (MRSA) infection, pandemic outbreaks like Ebola, and lately severe acute respiratory syndrome coronavirus 2 (SARS-COV-2), are also claiming the lives of many people.

1.2. Early and Rapid Detection of Diseases

Disease detection is the first step before diagnosis and care. However, detection is not easily accessible everywhere. Efforts to control infectious diseases or detect cancers early would benefit from new screening technologies.[2,3] For instance, early diagnosis could reduce mortality for many cancer types[4] As well, quick, reliable, and widespread testing is vital to control a pandemic.

1.3. Need for a Noninvasive Low Cost and Reliable Detection Method

Diagnostics rely on direct imaging or on collecting samples from individuals or contaminated environments, transportation of samples to a laboratory, and subsequent laboratory testing to demonstrate the presence or the absence of the pathogen of interest. This results in a significant delay in response times and containment efforts. These procedures can be invasive and require skilled human resources and costly equipment and consumables depending on the diseases. A desirable screening method should be noninvasive, painless, inexpensive, and easily accessible to many patients. In addition, it should allow diagnosis at an early stage.[5]

1.4. Diseases Emit VOCs

The human body emits Volatile Organic Compounds (VOCs) originating as a result of normal biochemical or physiological processes (endogenous processes), after absorption of external contaminants (eg, food), and by bacterial metabolism (eg, armpit odor).[6-9] VOCs are organic chemicals with a high vapor pressure at typical room temperature, resulting in evaporation or sublimation of the molecules into the air surrounding the source. It has been shown previously that some diseases emit specific VOCs.[9] Disease-related VOCs may be found in the blood, breath, feces, skin, sputum, sweat, urine, and vaginal secretions of affected individuals. Such a signal could pave the way for a new detection technique: using VOCs as biomarkers for disease detection. Research investigating the VOCs profiles associated with various human diseases is underway, primarily driven by the goal of developing instrumentation for use in clinical diagnostics.[10,11]

Currently, intensive studies are being carried out to identify compounds that could be markers of cancer[12-17] and

could eventually support or even replace traditional screening methods. To do so, techniques such as gas chromatography-mass spectrometry (GC-MS) have already been developed, and several research teams and companies aim at developing electronic noses (e-noses).[18-20] Currently, the development of this technology is limited by the high cost of the necessary laboratory instrumentation, difficulties in standardizing sample collection, and preparation procedures in clinical settings.[21] These limitations can, for instance, be due to threshold, non-optimized odor-capturing materials, low signal-to-noise ratio, costs, and the complexity of both the chemical signature and the subsequent data analysis. It is worth noticing that the origin and the nature of VOCs emitted by cancers are not well understood. Whether the chemical signature originates from the tumor, from the tumor environment, or both is still under investigation[15,22]

1.5. The Consistent Use of Animals to Detect Diseases

Several studies evaluating and reporting the potential ability of trained animals to detect certain diseases thanks to their sense of smell have raised many hopes.[23,24] Sense of smell has been extensively studied and is reported to be highly developed in certain species.[25] Primarily, the canine sense of smell has been deeply investigated.[26-28] Dogs have been trained to locate explosives, illicit drugs, banknotes, missing persons, and disaster victims.[29,30] Rats and several other animals have also been successfully trained to identify targeted substances.[31]

Animal olfactory detection of human diseases has attracted increasing interest from researchers in recent years.[28] In 1989, the first case was reported where a dog seemed to have detected his master's melanoma.[32] Similar cases have been reported in the following years.[33,34]

Anecdotal findings allow emitting the hypothesis that some animals could potentially be used to detect diseases. However, these findings alone do not ensure that animals can be employed as reliable tools to detect diseases. Furthermore, it should be noted that training dogs incurs costs and require time and an appropriate facility. Therefore, this potential new tool must be further explored and developed following a scientific method.

Several structured research programs have reported the abilities of dogs, rats, ants, and other animals to detect diseases such as cancers, diabetes, epilepsy, tuberculosis, malaria, urinary tract infections (UTI), and SARS-COV-2 among others.[35-40] These studies focus on animal capabilities and research and optimize sampling protocols and materials, storage and use of odors, scent lineup parameters, animal welfare, and testing conditions. To do so, research programs usually gather several professionals such as medical staff, chemists, biologists, physicists, statisticians, data scientists, veterinarians, ethologists, and dog handlers.

1.6. Still Many Unanswered Questions

Because of the inconsistent findings reported in this body of research and the complexity of scent detection research, it seems complicated to ascertain the potential value of animal detectors in diagnostic.[22,41,42] To our knowledge, no disease-specific VOC has been identified so far despite the number of studies reporting the ability of trained animals to detect diseases. With a rising number of publications tackling this issue, carrying out a structured and objective state-of-the-art seemed necessary.

1.7. Objectives

In this systematic review, we aim at outlining the performances (sensitivity, specificity) of trained dogs and rats in distinguishing cases of cancers or infectious diseases cases from controls in humans, thanks to their sense of smell, published in peer-reviewed research. Additionally, methodological issues leading to inconsistencies in research are reviewed, and further recommendations to improve performances are given. We excluded studies using other animals (nematodes,[43] insects[44,45]), studies aiming at detecting agricultural diseases, animal infections (dogs, cows, ducks[46-49]), and other diseases (hypo/hyperglycemia, neurodegenerative diseases, where the dogs are mostly serving for both assistance and detection). Only original articles published between 01/01/2000 and 01/05/2021 were included, and reviews were excluded.

2. Methods

2.1. Literature Search

Pubmed and Web of Science were independently searched based on the standards of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).[50] Studies about detecting cancer and infectious diseases (hypo/hyperglycemia, neurodegenerative diseases were excluded) in humans (agricultural diseases, animal infections, and animal cancers were excluded) by dogs', rats' sense of smell, in the databases from 01/01/2000 to 01/05/2021 were retrieved. The search strategy was adjusted for each database. For PubMed the following string was employed: (("Dogs"[Mesh]) OR (canine*) OR ("Rats"[Mesh]) OR ("Mice"[Mesh])) AND (("Volatile Organic Compounds"[Mesh]) OR (Volatile AND Organic AND Compound*) OR ("Odorants"[Mesh]) OR (odor*) OR (odour*) OR ("Smell"[Mesh]) OR (nose) OR (scent*) OR (sniff*) OR (olfact*)) AND (("Disease"[Mesh]) OR ("Neoplasms"[Mesh]) OR (cancer)). For Web of Science the subsequent sequence was researched: (TI= ((dog* OR canine* OR rat* OR mouse OR mice) AND (cancer* OR neoplasm* OR disease*) AND (smell* OR scent* OR sniff* OR olfact* OR odo$r* OR volatomic* OR (volatile organic compound*) OR (volatile* AND organic* AND

compound*)))) OR (AB= ((dog* OR canine* OR rat* OR mouse OR mice) AND (cancer* OR neoplasm* OR disease*) AND (smell* OR scent* OR sniff* OR olfact* OR odo$r* OR volatomic* OR (volatile organic compound*) OR (volatile* AND organic* AND compound*)))). The retrieved articles were further reviewed for original articles. Non-human disease detection and articles written in a language other than English were excluded.

2.2. Study Selection and Eligibility Criteria

In total, 5665 records were identified, 2057 from PubMed, and 3608 from Web of Science, of which 210 were duplicates (Figure 1). We reviewed the remaining 5455 titles and abstracts to identify relevant studies. Three authors (P.B., M.L., and I.F.) reviewed the abstracts and/or full-text manuscripts independently and selected those that were regarded as relevant. No disagreement on the selection of articles was seen between the 3 reviewers. The inclusion criteria were for studies on disease detection with dogs and rats in original articles. Articles that described original research involving animal olfactory detection of human disease using samples collected from human participants were selected for inclusion. Review articles and articles not directly relevant to the topic were excluded. Sixty-two full-text papers were reviewed for inclusion; none were excluded after full-text review. A total of 62 papers were included in the systematic review.

2.3. Data Extraction

The relevant data were extracted from the 62 selected peer-reviewed journal articles. A standardized table was designed to abstract the studies of interest. Information abstracted from each study included: details on the articles (Table 1), patients (Supplementary Table 1), samples (Supplementary Table 2), animal details (Supplementary Table 3), test setting (Supplementary Table 4), performances (Supplementary Table 5), and the number of samples (Supplementary Table 6) used. All graphics and tables were made in Office Excel.

3. Results

3.1. Systematic Presentation and Synopsis of the Characteristics and Findings of the Included Studies

The last 20 years have seen an increase in the number of publications dealing with disease detection in dogs and rats (Figure 2). Starting with case reports in 1989, several proofs of principle in the early 2000s were reported, followed by more complex studies in the last decade. A summary of key information from each of the studies is provided in Table 1, including the disease targeted for detection, the type of
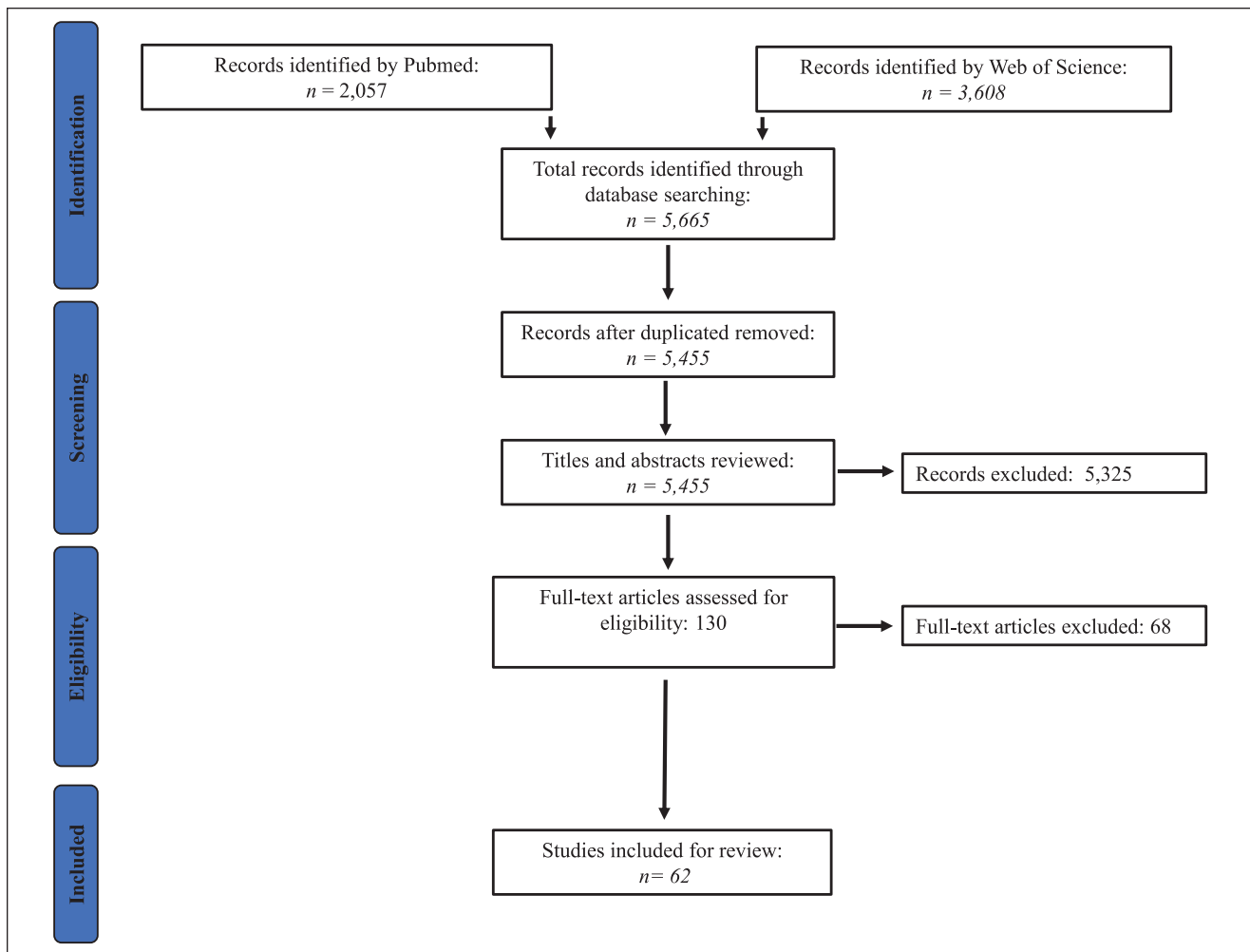
Figure 1. PRISMA flowchart.

body matrices used, the animal detector, and the sensitivity and specificity reported.

Cancer detection has received the most attention, with 2/3 of the studies targeting one or more cancers.[5,52-55,57-62,66-69,71-74,77,78,80-82,85-88,91-97,99-101,105] The remaining studies targeted tuberculosis, MRSA,[84] Malaria,[36] UTI,[79] *Clostridium difficile*,[63,70,83,89,98] and Covid 19[102-104,106-110] (Figure 3). However, since the Covid-19 outbreak in 2020, already 8 original articles reporting the ability of dogs to detect Covid 19 have been published, and more will possibly follow.

3.2. Selection of Subjects

A diagnostic test is designed to accurately discriminate patients from controls. Therefore, the choice of patients and controls is critical. Patient and control description among reviewed studies is reported in Supplementary Table 1.

Positive patients selected are subjects diagnosed with the disease of interest before any treatment. Diagnosis is mainly done with a reference test corresponding to the gold standard (histology, imaging, PCR & immunoassays). Histopathologic diagnosis is usually the reference test for cancer. The accuracy of the reference tests is, however, not systematically reported among reviewed studies.

Controls are of several types: (i) Healthy volunteers (healthy = absence of the disease of interest) who do not have and never had the disease; (ii) Healthy volunteers who do not have the disease anymore; (iii) Volunteers diagnosed with other diseases than the one of interest.

The absence of the targeted disease is not the only criterion for the selection of controls. Some teams often report to match age, gender, skin color, smoker status, diet, symptoms, and other comorbidities to limit confounders. Several studies, however, included controls with unmatched criteria compared to patients. A major drawback is the absence of

Table 1. Studies Included Within the Systematic Review.

Author	Animal - class	Disease	Sample	Sensitivity	Specificity
Willis et al[51]	Dog	Cancer - Bladder	Urine	0.41	UC
Pickel et al[52]	Dog	Cancer - Melanoma	Tissue	1.0	1.0
			Human	0.8	UC
McCulloch et al[53]	Dog	Cancer - Breast	Breath	0.97	1.0
		Cancer - Lung	Breath	0.99	0.98
Gordon et al[54]	Dog	Cancer - Breast	Urine	0.22	FC
		Cancer - Prostate	Urine	0.18	FC
Horvath et al[55]	Dog	Cancer - Ovarian	Tissue	1	0.98
Weetjens et al[56]	Rat	Tuberculosis	Sputum	0.73	0.93
Horvath et al[57]	Dog	Cancer - Ovarian	Tissue	1	0.96
			Plasma	1	0.94
Cornu et al[58]	Dog	Cancer - Prostate	Urine	0.91	0.91
Willis et al[59]	Dog	Cancer - Bladder	Urine	0.64	FC
Sonoda et al[60]	Dog	Cancer - Colorectal	Stool	0.97	0.99
			Breath	0.91	0.99
Buszewski et al[5]	Dog	Cancer - Lung	Breath	0.82	0.82
Ehmann et al[61]	Dog	Cancer - Lung	Breath	0.9	0.72
Walczak et al[62]	Dog	Cancer (several)	Breath	0.37	FC
Bomers et al[63]	Dog	Clostridium difficile	Stool	1	1
			Patients	0.83	0.98
Mgode et al[64]	Rat	Tuberculosis	Sputum	0.804	0.724
Mahoney et al[65]	Rat	Tuberculosis	Sputum	0.68	0.87
Horvath et al[66]	Dog	Cancer - Ovarian	Plasma (serie 1)	0.97	0.99
			Plasma (serie 2)	0.7	0.95
Amundsen et al[67]	Dog	Cancer - Lung	Breath	0.7	0.083
			Urine	0.657	0.25
Rudnicka et al[68]	Dog	Cancer - Lung	Breath	0.86	0.72
Elliker et al[69]	Dog	Cancer - Prostate	Urine	0.17	0.72
Bomers et al[70]	Dog	Clostridium difficile	Patients	0.86	0.97
Rudnicka et al[71]	Dog	Cancer - Lung	Breath	0.8554	0.72
Taverna et al[72]	Dog	Cancer - Prostate	Urine	0.986	0.964
Taverna et al[73]	Dog	Cancer - Prostate	Urine and serum	1 (pre-operatively)	UC
Urbanová et al[74]	Dog	Cancer - Prostate	Urine	0.935	0.916
Yoel et al[75]	Dog	Cancer (several)	Cell culture	1	1
Reither et al[76]	Rat	Tuberculosis	Sputum	0.569	0.805
Hackner et al[77]	Dog	Cancer - Lung	Breath	0.786	0.344
Willis et al[78]	Dog	Cancer - Melanoma	Skin	0.45	UC
Maurer et al[79]	Dog	UTI	Urine	0.996	0.915
Guerrero-Flores et al[80]	Dog	Cancer - Cervical	Tissue: biopsies Smear samples Surgical bandages	Scrapes: 0.928 Surgical bandages: 0.9636	Scrapes: 0.99 Surgical bandages: 0.99
Kitiyakara et al[81]	Dog	Cancer - Liver	Breath	0.78	UC
Guirao Montes et al[82]	Dog	Cancer - Lung	Breath	0.95	0.98
Bryce et al[83]	Dog	Clostridium difficile	Cultures & Feces	1	0.97
			Cultures & Feces	0.8	0.929
Koivusalo et al[84]	Dog	Methicillin-resistant Staphylococcus aureus	Cultures	0.75 - 0.97	0.83 - 0.96
Seo et al[85]	Dog	Cancer - Breast	Cell culture	0.94	0.986
		Cancer - Colorectal	Cell culture	0.93	0.981
Fischer-Tenhagen et al[86]	Dog	Cancer - Lung	Breath	0.94	0.6

(continued)

Table 1. (continued)

Author	Animal - class	Disease	Sample	Sensitivity	Specificity
Pacik et al[87]	Dog	Cancer - Prostate	Artificially created urine samples	0.7	0.85
Thuleau et al[88]	Dog	Cancer - Breast	Sweat	0.90	FC
Taylor et al[89]	Dog	Clostridium difficile	Stool	0.85	0.85
Edwards et al[90]	Rat	Tuberculosis	Sputum	UC	UC
Schoon et al[91]	Dog	Cancer - Colorectal	Stool	0.795	UC
Biehl et al[92]	Dog	Cancer - Lung	Breath	0.56	0.83
Feil et al[93]	Dog	Cancer - Lung	Urine	0.88	0.98
			Breath	0.78	0.96
Guirao et al[94]	Dog	Cancer - Lung	Breath	0.97	0.99
Junqueira et al[95]	Dog	Cancer - Lung	Blood Serum	0.97	0.98
Murarka et al[96]	Dog	Cancer - Ovarian	Cell lines	UC	UC
Protoshchak et al[97]	Dog	Cancer - Prostate	Urine	0.9785	0.95
Li et al[98]	Dog	Clostridium difficile	Free running	UC	UC
Guest et al[36]	Dog	Malaria	Foot odor	0.769-0.903	~0.9
Kure et al[99]	Dog	Cancer - Breast	Urine	1	1
Yamamoto et al[100]	Dog	Cancer - Cervical	Urine	1	1
Mazzola et al[101]	Dog	Cancer - Lung	Urine	0.45-0.73	0.89-0.91
Grandjean et al[102]	Dog	COVID-19	Sweat	0.76 - 1	UC
Jendrny et al[103]	Dog	COVID-19	Saliva or tracheobronchial secretions	0.83	0.96
Vesga et al[104]	Dog	COVID-19	Respiratory secretions	0.96	0.99
Guest et al[105]	Dog	Cancer - Prostate	Urine	0.71	0.70-0.76
Eskandary et al[106]	Dog	COVID-19	Throat and pharyngeal secretions	0.65	0.89
			Breath & sweat	0.86	0.929
Essler et al[107]	Dog	COVID-19	Urine and saliva	0.18	0.41
Grandjean[108]	Dog	COVID-19	Sweat	0.71-1	UC
Guest et al[109]	Dog	COVID-19	Breath & sweat	0.82-0.94	0.76-0.92
Jendrny et al[110]	Dogs	COVID-19	Saliva, sweat, urine	Saliva = 0.82 Sweat = 0.91 Urine = 0.95	Saliva = 0.96 Sweat = 0.94 Urine = 0.98

Abbreviations: UC, unclear; FC, forced choice.

control screening in most reviewed studies. This can lead to false negative samples.

3.3. Type of Body Matrix and Logistics

3.3.1. Diversity of sampled body matrices. When detection was designed to be done without contact between patients and animals, several types of body matrices have been collected to present odors to the animal detector. Urine is the main body fluid used (n=20), followed by breath (n=18), saliva (n=10), skin secretions (n=8), cell cultures (n=7), feces (n=6), blood/serum (n=5), tissue (n=4), and smear (n=1). Direct contact with patients or infected areas was conducted in 4 publications. The total is superior to the

number of publications, as some studies report to have used several types of samples. These data are reported in Supplementary Table 2, as well as in Figure 4. Three studies performed detection in direct contact between animals and humans (patients and controls).

3.3.2. Sampling materials and protocols. Sampling materials and devices to capture VOCs are reported in Supplementary Table 2. For urine, blood, and feces, no material was specially designed to capture VOCs efficiently. Only a receptacle (recipient, jar, cup, vial) was used. For sweat sampling, cotton pads are usually used. The composition of these pads is not always well-described. For breath, 3 types of materials/recipients were used: (i) Only a container (eg,
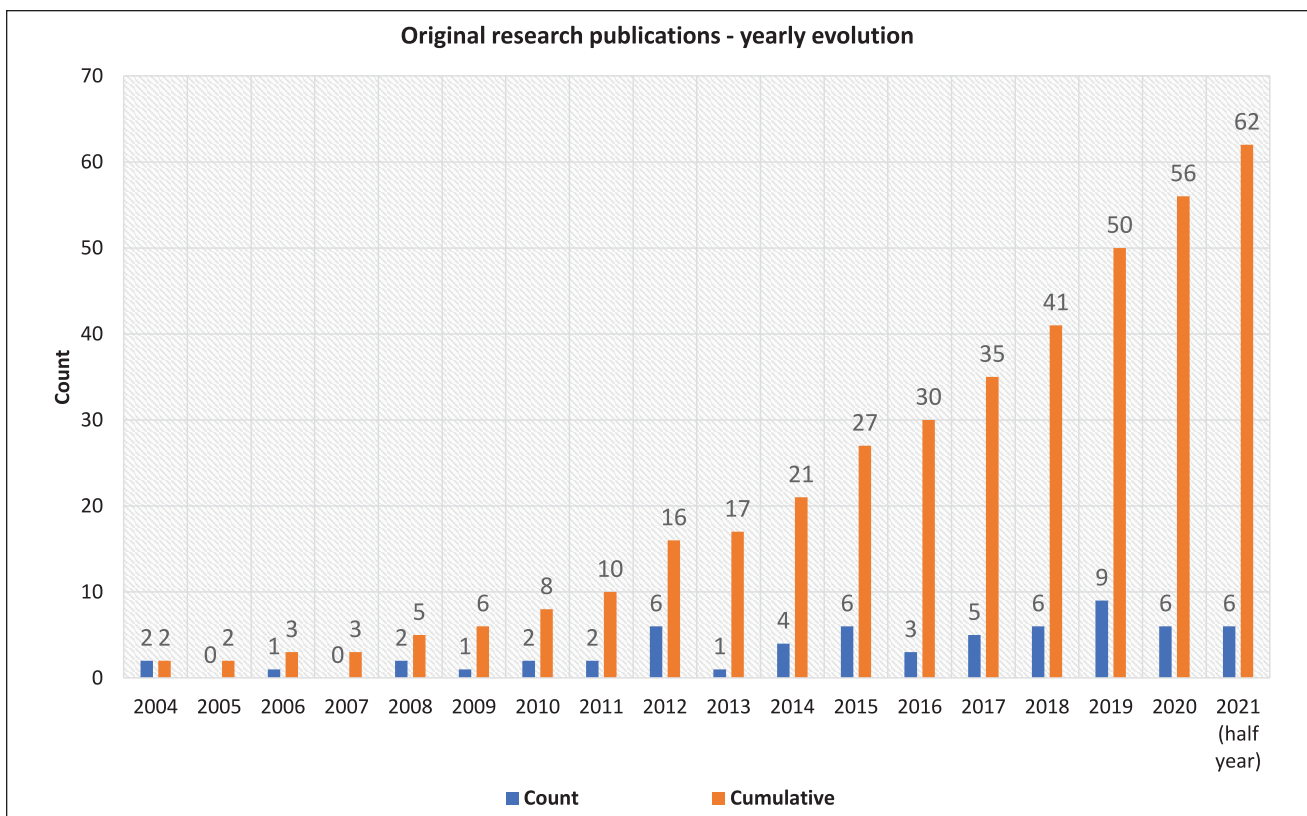
Figure 2. Evolution of the number of peer-reviewed publications per year among the selection.

breath sampling bag[60]); (ii) A tube filled with an absorber (eg, cylindrical polypropylene organic vapor sampling tube (Defencetek, Pretoria, South Africa)[53]; (iii) Other materials (ex: Face mask taken off and placed into a Ziploc® bag[81]). Tissues do not require specific sampling materials. In general, the choices of sampling materials are poorly justified, and material characterization and properties that capture VOCs are inaccurately described.

Sampling protocols are essential for reproducibility and limiting bias. They are reported for each study in Supplementary Table 2. They are well described primarily for 2 types of samples: exhaled air and skin secretions. For instance, Thuleau et al report that all patients and controls must shower with identical odorless soap before skin secretions sampling.[88] As well, some studies add information about fasting requirements before sampling, to limit biases (see Supplementary Table 2, column Sampling protocol).[82,92-94]

3.3.3. Sample storage conditions and conservation duration.
Storage temperatures are reported in Supplementary Table 2. Storage temperatures (T) are described for 73% (45 out of 62 articles) of the reviewed articles. We chose to classify them into 3 categories: (i) room temperature; (ii) cold: 0°C < T< 8°C; (iii) frozen: T <0°C (Figure 5). However, the choices of storage temperatures are not explained except in

Willis et al[59]: The team primarily stored samples at −80◦C, which has been the most desirable for retaining volatile chemicals.[111]

When stored at room temperature, all studies specify that samples were stored in the absence of light. Even if not specified for cold and frozen conditions, we assumed it was stored in the absence of light in a fridge or a freezer. Such a parameter is essential, as light is known to alter VOCs.[112] Air humidity data and atmospheric pressure were not described.

Also, sample conservation duration is poorly described among the reviewed articles and varies from a few days to several months. No data has been found about the quantification and the variation of VOCs captured in samples.

3.4. Animal Types

The data concerning animal details can be found in Supplementary Table 3.

3.4.1. Dogs.
Dogs are used in 92% (n=57, total=62) of the reviewed studies. In total, 226 dogs have participated in the studies. Among these 226 dogs, 186 (82%) have completed the whole study process (ie, have undergone full training and have participated in final testing). Most of the studies reported that dogs have been trained by professional dog handlers (data not shown).
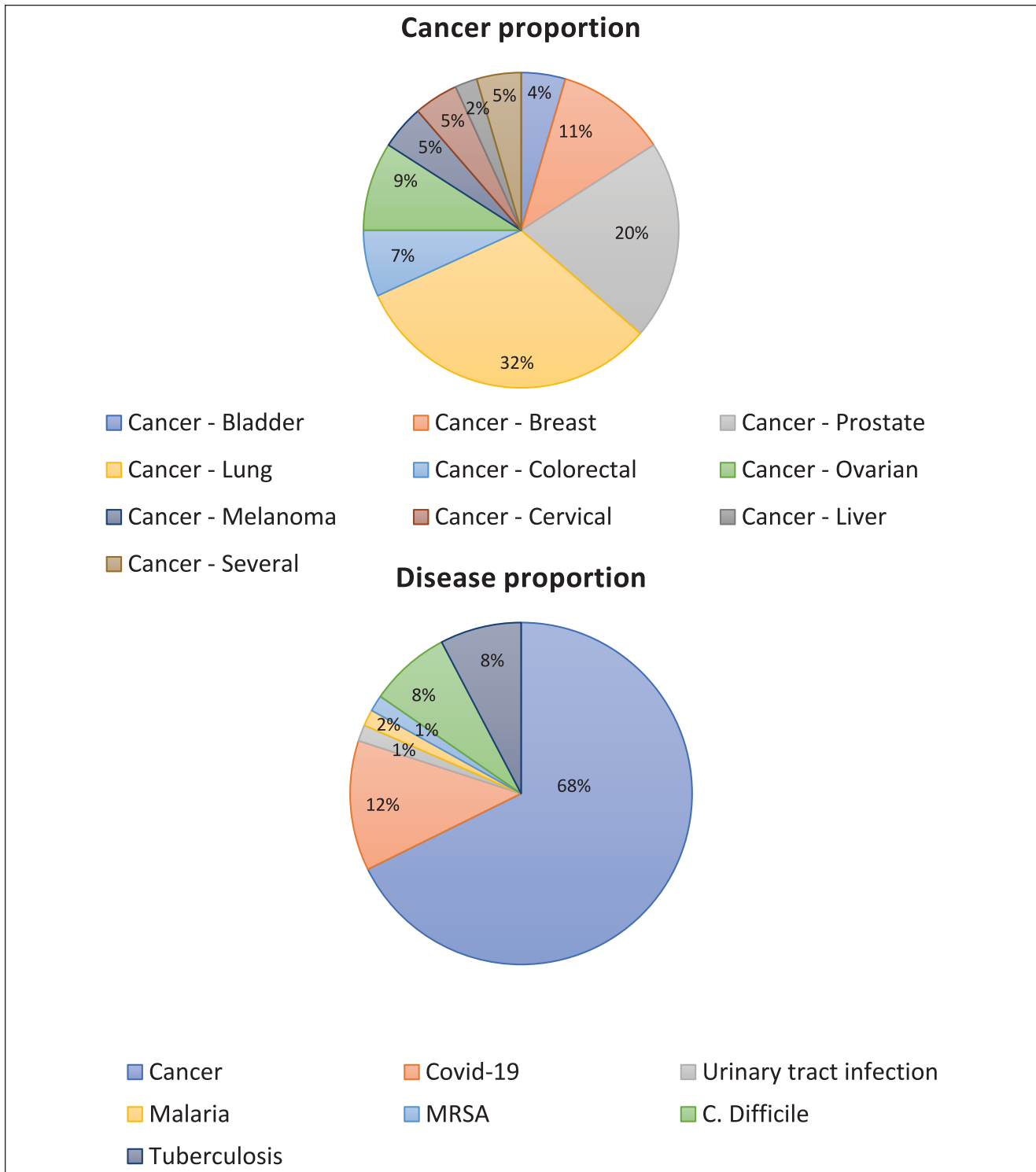
7

Cancer proportion

Legend (Cancer proportion):
- Cancer - Bladder
- Cancer - Breast
- Cancer - Prostate
- Cancer - Lung
- Cancer - Colorectal
- Cancer - Ovarian
- Cancer - Melanoma
- Cancer - Cervical
- Cancer - Liver
- Cancer - Several

Disease proportion

Legend (Disease proportion):
- Cancer
- Covid-19
- Urinary tract infection
- Malaria
- MRSA
- C. Difficile
- Tuberculosis

Figure 3. Proportion of diseases (upper) and different cancers (lower) over the included studies.

Dogs seem to be the first choice when it comes to using animals to detect diseases. This choice might come from the extensive use of dogs in drugs and explosives detection, the availability of dog trainers in many countries worldwide, and therefore the accumulated knowledge concerning their education. However, little information is provided about dog selection, except those choices are based on motivation (willing to search and play) and sense of smell. Standard selection tests to evaluate animal capacities are not described.
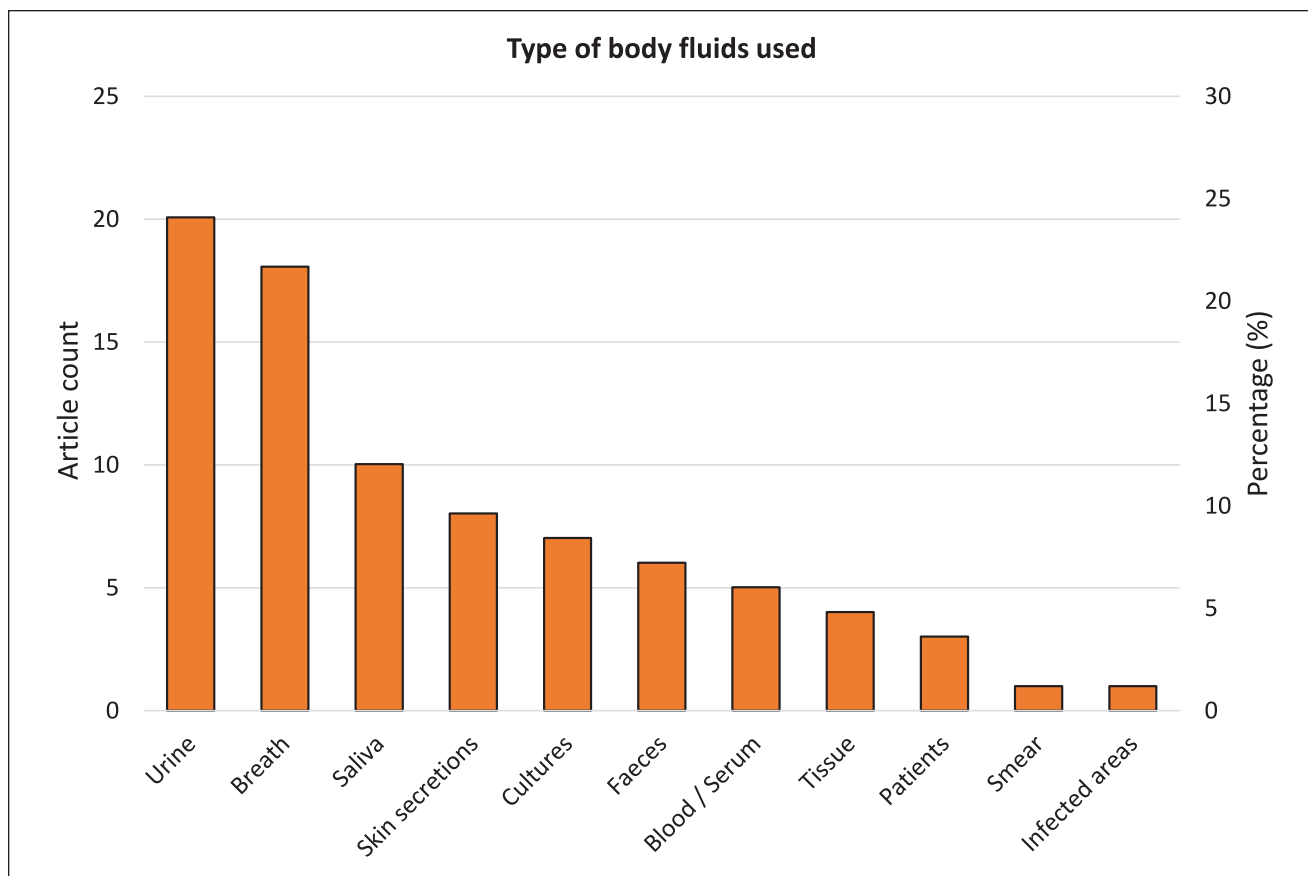
Figure 4. Bar chart of the different body matrices employed within the included studies.

Some breeds such as German and Belgian shepherds, Labradors, and Springers seem to be extensively used (see Supplementary Table 3). The distribution between males and females is the following: 52% males, 48% females. The number of dogs per study varies between 1 and 10, with an average of 3.96 (SD=2.84) dogs per study. These numbers are reported in Figure 5 and Supplementary Table 3.

3.4.2. Rats. African Giant Pouched Rats (*Cricetomys gambianus*) have been extensively used for tuberculosis detection in Tanzania (5 studies conducted with the same organization: APOPO). Rats were chosen for their sense of smell, easiness of operant conditioning, and availability in Tanzania. It is reported that such animals can live approximately 8 years and they can be trained within a few weeks. Mice studies were not included in the inclusion criteria, it is worth noting that only one mouse study was discarded.[113]

3.5. Animal Training and Testing

All reviewed studies report positive operant conditioning methods for training, the reward being food or a toy. A clicker training method is reported in 45% (total=62) of the studies. Animal living conditions were, however, not or poorly described. A few teams mentioned dogs' housing conditions (for instance, dogs being hosted in families). Education durations vary from a few weeks[53] to 5 years[60] depending on the teams and the difficulty of the exercise. The frequency of training ranged from once a week to 2 sessions per day, each day. These results are reported in Supplementary Table 3.

3.5.1. Sample presentation/stations. How individual samples were presented to the dogs is reported in Supplementary Table 4. A trade-off between odor intensity and contact avoidance between samples and the dogs' noses to limit sample pollution or dog contamination is usually reported to have led to station design. Stations' cleaning is not always described. When reported, no rationale is given. For instance, Horvath et al[55] chose to clean both the boxes and the containers with hot water after each exercise. Two years later, the same team[57] switched and cleaned with 95% alcohol. No standardized protocol has been identified.

3.5.2. Scent line-up characteristics. Scent line-up characteristics are reported in Supplementary Table 4. Scent line-ups are usually composed of 2 to 10 stations, disposed in a line or a circle. Most of the studies report a forced-choice design, that is, a fixed number of positive samples (> 0) per scent
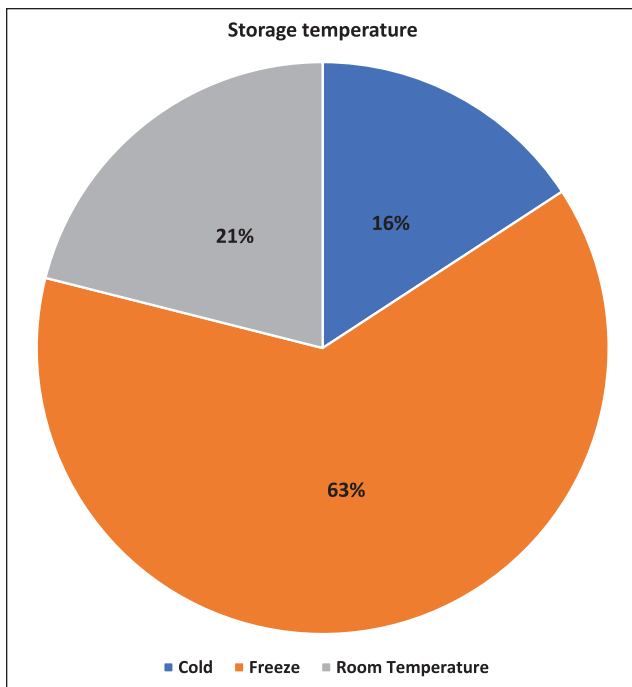
Figure 5. Storage temperatures of the employed samples from the included studies.

line-up. This concept of forced-choice design has been described by Edwards et al.[112] In a forced-choice design, the handler knows that the animal must find a fixed number of samples, which can induce bias. Some studies report the possibility of having zero positive samples per line (called "*blank runs*"), which corresponds to an unforced choice. Another type of unforced choice is to be able to vary the number of positive samples per line. Unforced choices are less common and often lead to worse performances (see Supplementary Table 5).

Within scent line-ups, several types of samples can be found: (i) positive samples; (ii) controls (healthy, other diseases); (iii) distractors. Distractors are samples different from positive samples and control samples. For instance, Murarka et al[96] used paper clips, paper towels, cotton balls, and screws as distractors. They are used to stimulate the animals to search.

3.5.3. Number of samples used for training and testing. The average number of samples used for training and testing is reported in Supplementary Table 6 as well as the standard deviation, minimum, and maximum numbers. These numbers are not systematically reported among studies, especially for training. Indeed, only 37% (15 out of 41) of cancer studies and 9.5% (2 out of 21) of infectious diseases studies gave information about the number of samples used for training. Information was more exhaustive for testing: 78% (32 out of 41) of cancer studies and 72% (15 out of 21) of infectious diseases studies gave the exact number of

samples used. Considering only studies which provided information, the mean number of samples used for training per study was 258 (SD = 560; min = 20; max = 2600), and the mean number for testing was 184 (SD = 186; min = 14; max = 902).

The number of times samples are used is not well reported. For training, some studies report training with only new samples to avoid 2 biases: (1) the memory effect (ie, animals do not learn to generalize, but remember each sample), and (2) the "novel object preference" effect (ie, animals select every sample they never encountered before). For instance, Ehmann et al[61] report that during the training and later in the testing, every test tube containing a human breath sample was used only once to preclude simple memory recognition of participants' unique odor signatures.[61] Even if not always described, when looking at the numbers, it is evident that most of the studies reuse some samples for training.

For testing, most teams used samples (positive and controls) only once per animal. However, a few studies report sample re-uses (eg, Cornu et al[58] (p. 201), Supplementary Table 5). The potential issues brought by reusing samples are discussed below. The mean number of rats used per publication was 11 for training and testing. In the case of dogs, on average 4 dogs were employed for training and testing (see Figure 6). The number of employed animals plays a crucial role in the feasibility of the study.

3.5.4. Blinding conditions. Results on blinding conditions can be consulted in Supplementary Tables 4 and 5. Several studies report having worked in blinded or double-blinded conditions. However, these terms do not seem to be used the same way among studies. In this review, we chose to classify the blinded conditions with the following terms:

- Unblinded conditions (UB): the dog handler knows the nature or the position of the sample to evaluate.
- Single blinded conditions (SB): an operator in the room (visible by the dog) knows the nature and the position of the samples to analyze, but the dog handler does not.
- Double-blinded conditions (DB): nobody in the room knows the nature or the position of the samples to analyze. This can be subdivided as follows:
  ○ Someone outside the room (or at least completely hidden) knows the nature of the sample to analyze (DB1) and can give feedback
    ■ In this configuration, the animals' indications can be evaluated each time, and therefore the handler can:
      • Reward his animal (= positive reinforcement)
      • Decide whether to continue or not the evaluations (because he knows if his animal is doing well or not)
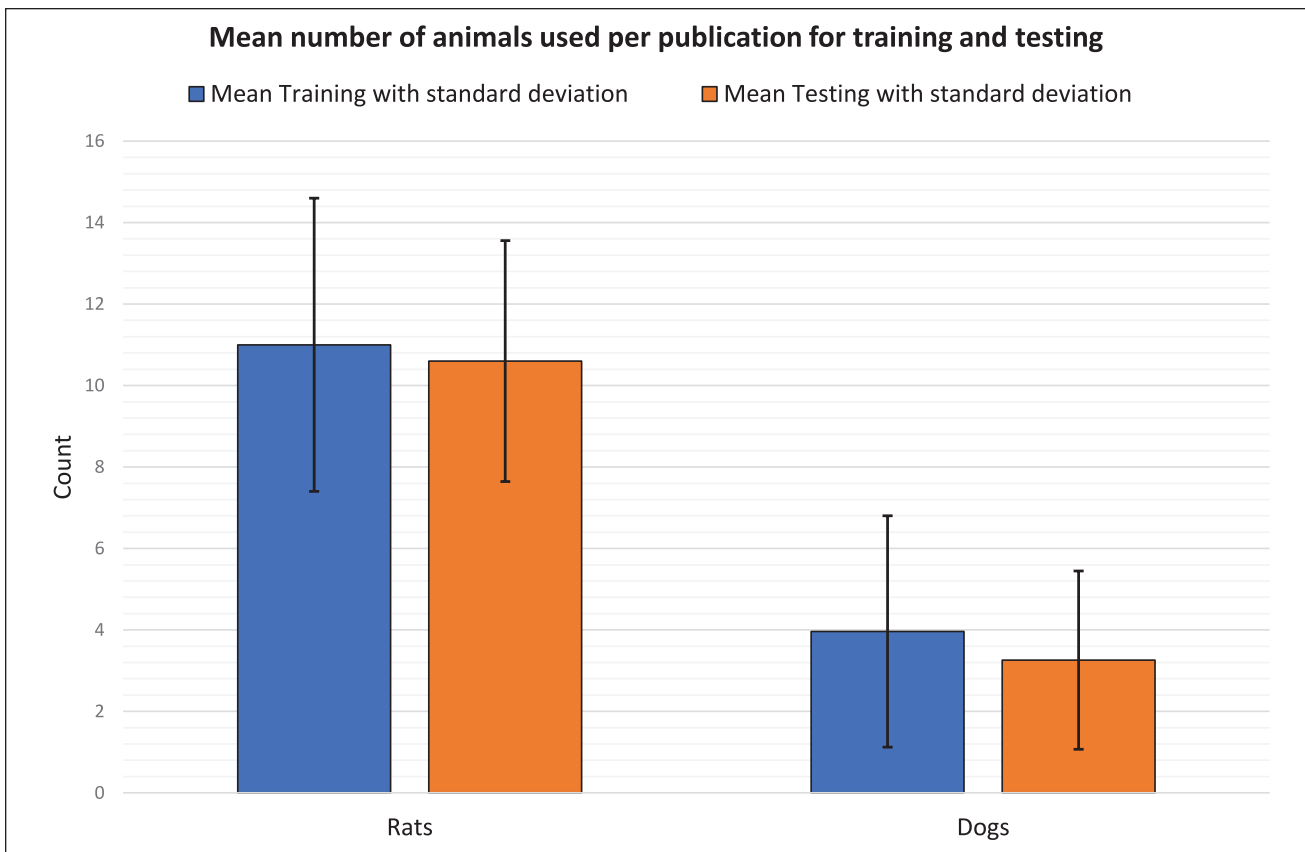
Mean number of animals used per publication for training and testing

Figure 6. The mean number of animals used per publication for training and testing.

○ Nobody knows the nature of the sample to analyze, or at least, cannot communicate it to the field (DB2). This condition is similar to a diagnostic condition.
  ■ In this configuration, the animals' indications cannot be evaluated each time, and therefore the handler:
    • Does not know whether to reward his animal or not
    • Does not know when to continue or to stop testing

In 78% of reviewed studies (48 out of 62), evaluations were done in double-blinded conditions to avoid the "Clever Hans" bias.[114,115] When well described, DB1 is the major double-blinded subtype reported (42%). Such conditions have limitations (see discussion). In 58% of studies (38 out of 62), the scent line-up had a forced-choice configuration.

3.6. Performances

Sensitivity and specificity varied widely, ranging from perfect to chance performance, with considerable variation among studies examining the same disease, employed
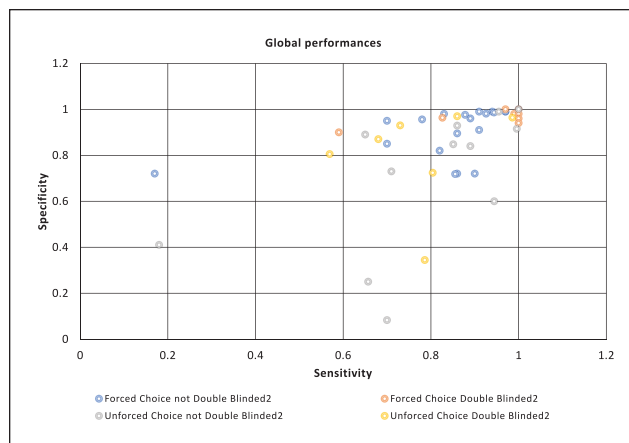

Figure 7. Global performances in different blinding conditions. Data are shown only for studies providing both sensitivity and specificity results.

body matrices, and detectors. Results are reported in Supplementary Table 5 and represented in Figure 7 when both sensitivity and specificity were available.

For "Forced Choice" situations, we report the specificity numbers from the original articles. Configurations using an unforced choice line-up in double-blind type DB2

correspond to a true "screening-like situation." Only 6 studies were found with the latter design: 3e with rats detecting tuberculosis, 2 with dogs detecting cancer, and 1 with dogs detecting *C. difficile* infections.

4. Discussion

4.1. General Comments

Since 2004, many proofs of concept have been published about the ability of dogs or rats to detect diseases. Furthermore, Buszewski et al[5] employed a chromatographic method for the identification of VOCs, and the results were compared with canine smell recognition. There are often great discrepancies among results. While some publications report low sensitivities of 0.17 (Gordon et al[54]) and specificities around 0.29 (Amundsen et al[67]), others achieve rates of 1 in sensitivity (Horvath et al[55]; Cornu et al[58]; Sonoda et al[60]) and specificity (Sonoda et al[60]; Yamamoto et al[100]).

Only a few studies reported testing performed in blinding conditions (DB2, unforced choice), and those usually enrolled small numbers of animals. This could be explained by the fact that screening conditions in double-blind testing combined with unforced choices are more challenging for the animals, the handlers, and the operators. Unfortunately, the little data available regarding this screening condition limits the possibility to validate such a method. These results are discussed in the following paragraphs.

4.2. Considerations About Patients' Selection and Samples

4.2.1. Patient and control selection: Reference test and populations matching.
First, careful diagnosis of patients and controls is critical to avoid bias. Making sure that patients have the disease of interest is usually confirmed with the gold standard. The accuracy of such a test must be high to avoid false-positive inclusions. Also, confirmed negative samples are critical, and all controls should be tested for the disease of interest. However, very few studies reported having rigorously tested controls. This can be explained by the fact that asking volunteers to perform non-required detection tests is costly, time-consuming, tedious, and invasive, poses ethical issues, and could lead to volunteer disengagement. However, from a scientific point of view, non-tested volunteers could be a source of false-negative samples. Such samples would be detrimental to animal training and testing. Indeed, the animal must be educated with samples with known status. An inaccurate reference test might lead to sample status errors and mislead the detector. For instance, Thuleau et al[88] reported they educated dogs to detect breast cancer from patients with cancer confirmed by histology and from volunteers with a recent (<12 months) negative

mammography. Even if mammography is reliable, false negatives can occur, or cancer can appear within a few months following the screening. Within the training phase, dogs are rewarded for not identifying the samples identified as negative from the mammography, therefore mammographic false negatives induce mistakes in training phases and consequently for further cancer identification.

If the reference test has poor accuracy, then animal training can be impacted. For instance, the results reported by Cornu et al[58] show that training a dog with potential "rogue" controls affected final performances. Selected controls were patients aged > 50 with elevated Prostate-Specific Antigen (PSA, comparable with cancer patients regarding these characteristics) levels. Control patients had a mean PSA value of 8.3 ± 4.1 [range: 2-16.8]. Given these values, it can be considered that 20% to 30% of these control patients with negative prostate biopsies had prostate cancer.[58]

Similarly, Willis et al[51] reported they were concerned that "rogue" control specimens from people with undiagnosed cancer elsewhere in the body might be inadvertently added to pooled samples. They did have an occasion during training in which all dogs unequivocally indicated as positive a sample from a participant recruited as a control based on negative cystoscopy and ultrasonography. After further tests, a transitional cell carcinoma was discovered. As such detection method with animals is not yet validated, not all false positives indicated by animals can be double-checked. More recently, Grandjean et al[102] had a similar issue, with 2 of their supposed SARS-CoV-2 negative controls turning out to be positive.

Second, the importance of matching the characteristics of patients and control groups to make sure that animals detect the disease itself and not a confounding factor.[27,112] Matching has been reported with age, sex, skin color, other diseases, comorbidities, symptoms, smoker status, and diet.

For instance, Bomers et al[63] worked on *C. difficile* detection with dogs at a hospital. They reported that on the day of the detection round all cases had diarrhea compared with 6% of the controls. In such a situation, we can wonder if the dog successfully indicated the targeted disease (C. difficile), or just the presence of diarrhea.

To prevent such bias, Willis et al[51] exposed the dogs to urine from patients with a broad range of transitional cell carcinomas, in terms of grade and stage, to increase their likelihood of recognizing the common factor or factors. They took particular care to train the dogs with control samples containing elements likely to be present in urine from patients with bladder cancer and commonly occurring in other non-malignant pathologies. This way, they could teach the dogs to ignore non-cancer-specific odors. This led to the inclusion of urine samples from a variety of patients, such as people with diabetes to control for glucose, those with chronic cystitis to deal with the influence of leukocytes

Table 2. Disease Odors Localization and Specificity Hypothesis.

	Disease odor localized	Disease odor widespread
Disease odor: specific	Sample choice critical	Sample choice is less critical
	High test specificity	High test specificity
Disease odor: common in several diseases	Sample choice critical	Sample choice is less critical
	Localization can give an alert to a shortlist of diseases	Low test specificity

and protein, and healthy menstruating women to control for blood.

Several years later, the same team (Willis et al[59]) assumed that body matrices, tissues, and emissions from young, healthy individuals differ in composition from those of older cancer patients to a greater extent than do samples from age-matched individuals with non-cancerous disease of the same organ. They performed an e-nose study in which the classification accuracy dropped when more diseased individuals were added to the healthy control group.[116] This shows that the choice of controls can markedly affect the level of specificity achieved.

4.2.2. Disease-specific odor and types of body matrices chosen.
To our knowledge, it is not known whether specific cancer has a characteristic chemical signature or not, and, if so, what would be the source of such signature. The odor of cancer could come from the tumor itself, the modified environment surrounding the tumor, or both. Moreover, it is still not known yet whether all cancers have shared odors or not. For instance, McCulloch et al's group reported good dogs' performances after being trained to alert to 2 cancers rather than for single cancer discrimination.[53] This could mean that there is a general biochemical marker common to all cancers, with individual-specific cancers having additional markers.[54]

There are different interpretations considering the localization of disease odor within the body: is it localized, organ-specific or spread? For instance, Horvath et al[55] report that one important observation during the training period was that use of fat from the same individuals from whom the carcinomas were removed did not increase the number of failures. The absence of reaction by the dog suggests that a general body odor including all organs did not exist. However, 2 years later, the same team[57] reported that for the same cancer (ovarian), dogs trained with tumors could discriminate blood samples and vice versa. Their study strongly suggests that the characteristic odor emitted by ovarian cancer samples is also present in the blood (plasma). Similarly, after observing that canine scent judgment can be used on both breath samples and watery stool samples, Sonoda et al[60] concluded that chemical compounds may be circulating throughout the body for colorectal cancer.

Murarka et al[96] comment that Yoel et al[75] found that after being trained on the breast cancer cell line, the dogs were able to detect both skin cancer and lung cancer cell lines, suggesting the possible presence of a general cancer olfactory cue within cancer cell lines. However, this study did not explore whether these dogs could also then detect cancer in patient-derived samples. In this case, there is also the possibility that the dog learned to disregard control samples (which were probably similar) instead of recognizing malignant cell cultures. This seems possible in the observations from Murarka et al,[96] whose research suggests that after training on cell lines to prepare the dogs, there were no spontaneous recognitions of cancer in blood plasma samples.

From these observations, 4 situations can be considered depending on odor specificity and localization, which are presented in Table 2.

Table 2 shows that the body matrix choice is critical. This also affects the choice of control cohorts. From this table, we see that an odor widespread throughout the body and non-specific to disease will lead to low specificity tests. In such situations, indications of a sample by a trained animal will not give much information on what disease to look for, and therefore will have low added value.

Body matrices used in the reviewed articles are dominated by breath and urine (Figure 4). These have the advantage of being easy to sample (liquid, air, noninvasive), easy to split into several samples, and therefore allow several trainings and tests per sample without encountering pollution or odor decrease. Liquids like urine are also easy to dilute, for instance, to increase detection difficulty by reducing the number of VOCs per sample. These dilutions also allowed to study animal detection thresholds,[113] and comparisons with GC-MS and e-noses. Urine samples have the additional advantage that they can be aliquoted and frozen for later usage. Despite this, we regret that the reasons that lead to the choices of body matrices were not or poorly documented within the selected articles.

4.2.3. Sampling protocols.
After body matrices and sampling localization choice, sampling protocols and materials are key to have high-quality samples. Most of the studies report the importance of applying the same sampling procedures both for patients and controls to eliminate potential bias and

confounders. For instance, Ehmann et al[61] showed that, at first, trained dogs were not discriminating against disease state, but sampling location which was different for patients (at the hospital) versus healthy volunteers (at home).

If the sampling protocol is made at home with no supervision, the risk of error can occur, leading to poor sample quality. Thuleau et al[88] report that to sample sweat they asked patients and volunteers to shower with an odorless soap, before sleeping with a cotton pad on the breast overnight. In this case, researchers cannot be sure that the person has followed each step correctly or that no incident occurred. In this example, the pad could have fallen during the night, resulting in pollution and a limited contact time of the pad with the skin, and therefore a limited number of VOCs. As well, other odors could have been impregnated on the sample, such as bedsheets' odors, partners' odors, and pets' odors without any possibility of quality control. Such unsupervised sampling protocols add difficulties and should be controlled as much as possible.

A non-exhaustive list of parameters that can induce bias are smoking status, sex, age, ethnicity, diet, different sampling locations, different sampling protocols for patients and controls, and treatments. For instance, to limit diet bias, Hackner et al[77] report that for homogeneous sampling, the tested persons were constrained not to drink, eat and smoke within 90 minutes before breath sample collection. It is important to note that receiving this data among patients requires regulatory approvals.

4.2.4. Odor sampling materials. All types of body matrices do not necessarily require odor-sampling materials. For instance, urine, feces, and blood can be sampled and presented untransformed to animal detectors. However, breath and skin secretions need optimized materials to capture VOCs without releasing other odors that could disturb detection. Some sampling materials have been presented in chapter 3.5.2 and Supplementary Table 2. For instance, Willis et al[78] report that their choice of material comprising their patches came from studies on canine scenting in forensic science. In terms of the greatest variety and quantity of skin surface VOCs collected and readily released, the optimum fiber appeared at the outset of their study to be 100% cotton, so they employed a widely available, sterile, pure cotton gauze throughout. For the chosen sampling time of 15 minutes, they were again guided by the forensic science literature.

However, such a description is an exception, and as for the choice of body fluids, we regret that the choice of materials is poorly documented. The vast discrepancies among material types strongly suggest this part of the research is still empirical and needs better understanding, characterization, and standardization. In the future, this field of research would benefit from a better description of material parameters, as it is often done in publications reporting VOC detection by GC-MS.[117,118]

4.2.5. Sample conservation. In chemistry, it is known that temperature variations, light, and air humidity can modify VOC profiles.[119-123] Such parameters are crucial but neither well described nor consensual.

Most of the reviewed studies stored samples at low temperatures ($<0°C$), and only a few stored them at room temperature (see Supplementary Table 2 and Figure 5). This choice is usually not justified, except in a few studies. Willis et al[59] report that urine samples were stored primarily at $-80°C$, which has been the most desirable temperature for retaining chemical species.[111] Mahoney et al[65] report that their samples were frozen at $-20°C$ until the evaluation day (up to 7 days). Though there is some controversy surrounding the cellular impact of freezing and thawing sputum, past research suggests that samples may be kept frozen without significant alteration of cell viability or cell counts.[124] Not much information is given about light. However, most studies report storing samples in a refrigerator or in a freezer, where an absence of light is evident. No information has been found about air humidity or atmospheric pressure. Conservation time and the number of sample openings lack description. The heterogeneity of VOC conservation procedures shows this part is still empirical and needs better understanding and evaluation. Guidelines on minimum, maximum, and optimum conservation conditions would undoubtedly be helpful for standardization.

4.2.6. Considerations about odor threshold. Selected animals have a sense of smell superior to that of humans.[25] For instance, Horvath et al[57] observed that trained dogs could detect a quantity of 20 ovarian carcinoma cells on the abdominal fat. However, the sense of smell is not unlimited, and it loses efficiency below a certain VOC threshold. This threshold effect has been studied in Sato et al[113] (article excluded from this systematic review). Willis et al[51] also report that they had to consider the physical state of the urine when presented to the dog. They opted to train one cohort of dogs on wet samples and another on dry samples. When tested, the dogs trained on liquid urine performed significantly better, suggesting that the more volatile molecules are important in the cancer odor signature.

Odor threshold also plays a role in dog training progression. Some teams chose to directly use the same types of samples at the training start and for testing.[54,58,77] On the contrary, others started detection work with samples with a higher number of VOCs and decreased the intensity step by step.[55,78,80,87] The latter strategy is supposed to be easier for the animals before lowering the threshold. These samples with more VOCs can be (i) bigger (bigger in volume, surface, quantity); (ii) more concentrated (exhaled air, sweat, etc); (iii) other types of samples, such as tumors or materials directly in contact with the tumor. However, the diversity of samples used before the final configuration is not systematically reported within studies. In addition, there may have been differences in odor intensity between diseases,

especially infectious and viral diseases with strong diffusion (to be related to contagion) versus hidden tumors.

4.2.7. Frequency of sample re-use: pollution and memory effect.
The number of times samples are used is not always well reported. It is evident, however, that some studies reused samples at least for some training. Here, different types of "reuses" are to consider:

- Case 1: The same sample is presented several times to the same animal detector[58]
- Case 2: The same sample is presented to several dogs (several times per dog or not)[53,86]
- Case 3: Sample replicates of the same patient are presented to an animal detector[60]

In cases 1 and 2, there is a risk of pollution (by direct contact with the animal or by its breath, by the atmosphere), which lead to sample alteration each time the sample is used. Therefore, once smelled, samples are not identical to "new" samples. Moreover, opening a sample several times may lead to a decrease in VOCs quantity. In cases 1 and 3, samples from the same person are presented several times to an animal. By doing so, there is a risk of training animals' memory instead of discrimination. This latter issue has been reported by several teams who saw their results plummet in double-blind situations with only new samples.[78,107]

On the contrary, however, Willis et al[78] report that multiple uses of the same sample (melanoma samples) during training did not appear to lead to a significant loss of volatile signature since the dog continued to successfully select known melanoma samples used up to 15 times over a period of 18 months post-collection. With such observation, one can assume that the dog did not learn to discriminate samples but instead memorized one specific sample.

Ideally, an animal should smell only new (uncontaminated) samples, only once per patient (to avoid memory effect). The advantage of urine, feces, blood, and breath is that these body matrices are easy to sample or to aliquotye, allowing to have several samples very quickly. This way, several dogs can be trained with samples from the same person, while preserving their quality.

In some studies (eg, Cornu et al[58]), some control samples were reused during testing. This does not seem to be a problem in an unforced choice configuration (cf scent line-up characteristics, part 3.7). However, in a forced-choice configuration, reusing some control samples might reduce the number of new possibilities for the dogs, leading to an easier design and higher success rate just by chance.

4.3. Animals

4.3.1. Animals.
Giant pouched rats have been extensively used by one team working on tuberculosis detection in Tanzania.[64,65,90] Little reasoning has been given in literature regarding the choice of animal except for their high sense of smell. Dogs are the most used animals worldwide. This choice can be justified by the availability and experience of dog trainers in many countries, for instance, for drugs and explosives detection. Dogs have the advantage of being adaptable to different fields (battle, airports, rescue, remote scent tracing, and contact with humans). However, for remote disease detection only (detection done in a controlled configuration, at a distance from patients), there is no need for such adaptation. To our knowledge, no validation study has been conducted comparing rats versus dogs. Authors generally report looking for motivated dogs with high olfaction capabilities. However, there seems to be no standard validated tests for dog selection, which so far remains empirical in the absence of clear guidelines.

Gordon et al[54] mention that it has been an ongoing theory that certain breeds are better at scent detection than others.[125] However, studies have shown a greater difference in scenting ability between dogs within a breed than between breeds. We observe variation in performance in selected studies between breeds[5,101] and within the same breeds[54,86,92] This has been described in Jamieson et al,[126] who concluded that a dog should not be solely chosen based on its breed due to individual variation. In addition, if we consider that evaluated dogs were for the majority selected among the best, under the watchful eyes of an experienced professional, we can assume that even more discrepancies would exist without such selection. There are an estimated 500 million dogs worldwide and, so far, less than 200 dogs have been considered potentially adapted to conduct disease screening tasks in controlled studies and achieved varying results. Such a method seems to have huge potential; however, these low numbers preclude extrapolation.

4.3.2. Selection success.
In Elliker et al,[69] only 3 out of 10 dogs initially recruited for the study passed the first stage of training. According to this research team, high failure rates are common when training dogs for specialist roles because of the specific behavior/temperament attributes required.[69,127,128]

Despite this low selection rate, 82% of the dogs mentioned in the studies completed all the exercises requested. This number may seem high but several factors might not be included within this percentage. For example, it is likely that some studies only mention the dogs who performed well and do not mention all the dogs they evaluated before selecting their champions. The loss rate is greater when the difficulty of the exercise increases (blank runs, double-blind). As most of the studies report forced choice scent line-ups, more dogs succeed. Interestingly, Murarka et al[96] report that all dogs leaving the disease detection program and switched to other odors (eg, narcotics, bed bugs, accelerants, blood plasma) have been
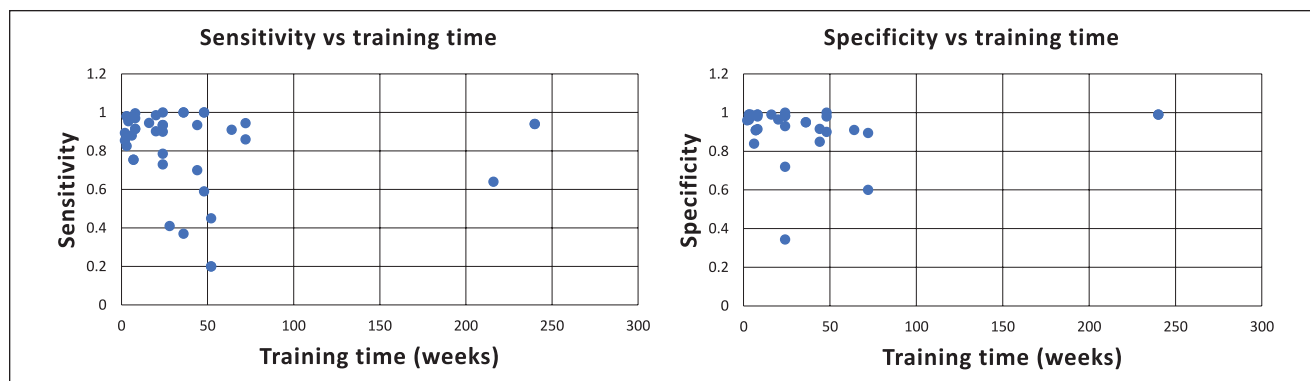
Figure 8. Sensitivity (left) and specificity (right) are set out as a function of the training time.

rapidly and successfully trained. This strongly illustrates the difficulty of disease detection with dogs compared to other odors.

Elliker et al[69] report that it has been suggested that it may be useful to breed dogs specifically for cancer odor detection,[129] which may help to increase the proportion of suitable dogs available for future studies of this type.

4.3.3. Training duration. Considerable differences in training durations are observed within studies, going from a few weeks to several years. Such differences can be explained by the type of disease to detect (Supplementary Table 1), the difference between patients and controls (Supplementary Table 1), the choice of body matrices (Supplementary Table 2), the quality of samples (Supplementary Table 2), training differences (Supplementary Table 3), animal abilities (Supplementary Table 5). No correlation was observed between training duration (Supplementary Table 3) and specificity and sensibility (Table 1) among studies (n=34) (see Figure 8). However, Ehmann et al[61] identified an improvement in lung cancer identification capabilities along with the test series and conclude that an ongoing training effect must be assumed, calling for even more extended dog training in future studies.

4.4. Scent line-up

4.4.1. Scent line-up: Number of samples and line versus circle, the distance between samples. The number of samples presented to animals ranges from 2 (Bomer et al[63]) to 12 (Essler et al[107]). No justification was provided considering these numbers. No study was performed with only one sample. It has been shown in the literature that dogs were able to perform tests with one sample only.[130] In such a test, dogs have to make an absolute choice. They are asked to "evaluate." On the contrary, when several samples are presented, the dog can perform a discrimination task and is probably more stimulated. In this situation, they are asked to "search." All studies reviewed used the latter configuration.

Samples were presented in a line, circle, or randomly (Supplementary Table 4). The choice of a line can be explained by the easiness of designing "blank" runs, such that, at the end of the line, the dog can indicate that no positive sample was found. Blank runs can also be done in a circle configuration. The advantage of the latter is that there is no start or end, so all samples are equivalent.

Space between samples is fundamental for several reasons. The most obvious reason is to preclude cross-contamination between samples. Another less apparent reason is that it gives enough time for latency and persistent olfaction times. Latency is defined as the necessary time to get an olfactive stimulus, estimated at 0.5 seconds for dogs. Persistence time is the duration the olfactive sensation stays. If samples are too close together, these durations cannot be respected, and the dogs risk either missing a sample or mixing signals.

4.4.2. Scent line-up configurations: Forced versus unforced choice. Using forced versus unforced choices scent line-ups have a strong influence on performances.[112] Unforced choice exercises are more complex. In a forced-choice exercise, the animal learns only one configuration. They know they must "find" the odor of the disease. As a result, they chose the sample which resembles the most to the target or the one which is the odd one out. Moreover, Bomers et al[63] report that anticipation of a single positive result could have influenced the trainer's behavior, thereby unintentionally influencing the dog's response.[116] Such configuration is therefore not only easier for the animal but also for the handler. On the contrary, animals must evaluate each sample in an unforced choice configuration and cannot choose only by simple comparison. This is a difficulty that not all animals can overcome. An unforced choice situation is, however, the only one that could be applied for screening.

With the particular configuration reported by Murarka et al[96] (see results), the dog has only one sample to evaluate, while the distractor is there for stimulation.[96] Such configuration is an interesting tradeoff between one versus several

samples scent line-ups described above and can easily be applied for screening (see Supplementary Table 4).

4.4.3. Atmospheric conditions.

Atmospheric conditions during training and testing are known to affect dogs' sense of smell. These conditions are poorly documented within the reviewed studies. Those who did, however, reported working with controlled temperatures between 12°C and 20°C (see Supplementary Table 4). It can also be seen that, when not under control, this can negatively impact scent detection work, for instance, reported by Sonoda et al[60] where tests were conducted from 13 November 2008 to 15 June 2009 because the dog's concentration tended to decrease during the hot summer season. As well, Hackner et al[77] observed some limiting influences including high humidity and elevated ambient temperature, which were found to be detrimental to the dogs' performance. They suggest that testing should not be performed during unfavorable weather conditions.

4.4.4. Blinding conditions

4.4.4.1. Proofs of principle versus double-blind clinical trials in a screening-like situation.

For a potential deployment of disease detection with animals, only double-blind clinical trials in a screening-like situation (ie, unforced choice) might be useful (see chapter 3.9 for blinded conditions). To date, only 6 studies meet these expectations (Supplementary Table 5). Focusing on these studies, the results usually decreased at first when shifting to double-blind. This drop between training and double-blind testing has often been explained by the Clever Hans effect.[115] To avoid failure, teams must train as much as they can in blind situations, as suggested by Gordon et al[54] who report that the use of blinding during the training should be initiated early to preclude unintended clues by the trainers that may contaminate the process. Willis et al[78] reported that after training the dog in a non-blinded situation, their trainer reported back a near 100% success rate in identifying the melanomas. It was decided to begin a series of double-blind tests. However, after 13 runs, the dog had successfully identified only one of the melanoma samples.[78] Implementing blinded conditions is not easy during training because dog handlers need to know when to reinforce positive behavior. To do so, a non-blinded assistant hidden from the dog and who can quickly tell the handler when to reinforce is needed.

4.4.4.2. Rewarding or not the dogs in a screening-like situation: a puzzling question.

In a screening-like situation, nobody knows whether the animal's indication is correct or not, which can be an issue for the reward. Indeed, if the trainer decides not to reward the animal, the latter can little by little lose interest. On the contrary, if the animal is rewarded every time, this might reinforce biases in case of incorrect indica-

tions. Therefore, several strategies are adopted among teams.

For instance, McCulloch et al[53] report that, since the experimenters no longer knew the status of the target breath sample, they did not activate the clicker device after a sitting indication by the dog, and therefore the handler did not reward the dog with any food. Bomers et al,[63] in the case of *C. difficile* infections, search in hospital wards, confirm that surveillance is principally different from the type of case-directed diagnosis in their study design because the dog cannot immediately receive a reward after a positive identification, potentially extinguishing the trained alert. The same solution was adopted by Willis et al[59]: "Both the trainers and researchers remained blinded throughout the trial, only breaking the sample and positional codes at the very end, meaning that the dogs could not be rewarded for a correct indication immediately after each test run. The trainers reported that, over time, this led to a loss of confidence in the dogs, with a deterioration in their performance." On the contrary, Elliker et al[69] performed 2 types of tests. On the first one, they were in a DB2 situation and decided to reward the dog for each indication. However, during 3 rigorously controlled double-blind tests involving urine samples from new donors, the dogs did not indicate cancer samples more frequently than expected by chance. The team finally switched to a DB1 situation, to be able to reward the dogs only for positive responses. These are exceptions because most of the studies were conducted in DB1 configuration, which allowed trainers to know whether to reward the dog or not after each line.

According to Biehl et al,[92] rewarding dogs' work has to be independent of the results achieved and should refer only to the work done. If dogs are only rewarded for positive indications, they will quickly learn to achieve more rewards through positive indications, which could easily lead to higher false-positive results. Hackner et al[77] attributed the inferior results to the true double-blind and screening-like conditions. They report that this factor posed immense stress on the dogs and their handlers, and therefore suggest positive feedback mechanisms for future study designs. According to them, it seems to be favorable to confront dogs relatively often with the pattern odors. Their results suggest that a test situation where dogs will always find an unblinded positive and ignore an unblinded negative sample in the line-up would probably be better. The positive sample would create the opportunity to earn a reward and would reinforce the dogs' motivation. The negative sample assures the handler that the dog is still performing well. The other samples in the line-up should be the blinded test samples.

Another similar solution would be to alternate training lines and test lines. It could be decided that one test line has to be performed only after an amount (to determine) of successful training lines. Another training line could be

performed right after the test to ensure the dog is still doing well. Such a pattern is feasible for implementation; however, it would slow the testing throughput.

This subject is crucial for implementing such a method, and no consensus nor solution has been achieved so far.

4.5. Applications/Implementation

Pickel et al[52] published a proof of concept with dogs sniffing directly human melanoma. Even if scientifically feasible, such a technique seems hardly applicable in the field. Since then, several studies using remote disease detection have step by step built a new scientific discipline. This review shows that no scientific study has validated that animals can be used as a first-line remote detection tool prior to existing technologies. Only APOPO, the organization supervising Giant Pouched Rats detecting tuberculosis in Tanzania, has found its place as a second-line screener, which makes sense for tuberculosis detection.[56,131]

4.5.1. How many dog decisions are needed to identify the target condition? Most studies focus on the performances of each animal separately. However, as animals are living organisms, their performances can be subject to variations. Biehl et al[92] reported that literature data show that some dog trainers included only one dog in scent detection, whereas others had 5 to 6 dogs and collected the individual dogs' data. McCulloch et al[53] stated that the sniffing quality of all dogs was comparable, and therefore the results obtained were similar. However, Ehmann et al[61] found differences in hit rates between individual dogs and consequently defined a "corporate dog decision" that required at least 3 out of 5 dogs with an identical decision. Amundsen et al,[67] as well as Hackner et al,[77] also showed considerable variations in single dogs' results. These variations might be due to the dogs' different sniffing capabilities and the dogs' different daily conditions and training.

Biehl et al[92] report that in their study, single dogs' results showed great differences concerning sensitivity in the range of 0.22 to 0.67 and concerning the specificity of 0.71 to 0.89. They conclude that it is advisable not to rely on a single dog's decision but to define a corporate decision to minimize variations arising from the single dogs. This choice is not straightforward. Indeed, Mahoney et al[65] report that sensitivity declines and specificity increases when 2 individual animals are employed because a positive sample can be indicated twice. On the contrary, if only the indication of 1 of the 2 animals is needed, the sensitivity will increase, but specificity will drop. The argument can be declined for more animals and indications. For instance, Gordon et al[54] report that, at the time, their study was the only one to incorporate replicates for assessing specificity. There were 3 and 2 replicates (33 and 18 runs) for the prostate- and breast cancer patients, respectively. The team adds that any study,

ultimately attempting to prove canine superiority over conventional cancer screening, must include replicates and, in the future, go head-to-head with standard screening methods. Another example is Mgode et al,[64] where for tuberculosis detection, a sample is considered positive if selected by 2 rats. Such a corporate decision is a tradeoff that has not found a consensus yet.

4.5.2. Number of samples to train a dog and maintain performances. The number of samples available for training is crucial. Indeed, many samples are needed so that animals learn to generalize and do not memorize each sample. Quantity is essential to work as often as possible with new (non-polluted) samples and limit the "novel object preference." Willis et al[59] report their protocol also avoids the phenomenon of novel-object preference, whereby dogs preferentially chose unfamiliar items over familiar ones.[132]

This is not straightforward, as organizing efficient logistics to gather samples continuously can be challenging to implement. For instance, Gordon et al[54] report that it took longer than anticipated to obtain enough samples to prepare for the final testing. This resulted in the training being spread over an extended period, 12 to 14 months. Possibly, the animals were periodically memorizing individual patients rather than recognizing an "odor signature" for cancer despite utilizing a large number of training samples. An ongoing system of recruitment of patients with cancer and control patients needs to be established, so the dogs have adequate numbers of new samples to maintain their proficiency even after the conclusion of the study. This has also been reported by Ehmann et al,[61] who wrote that during the training and also later in the testing, every test tube containing a human breath sample was used only once to preclude simple memory recognition of participants' unique odor signatures.

This need for the continuous arrival of new samples is a huge limitation. Indeed, if intended to be implemented in countries with low access to diagnostics, this arrival of new samples from screened patients and controls will be limited. This implies continuous logistics and partnerships with hospitals that might not be cost-effective.

4.5.3. Field implementation. If scientifically validated, remote scent medical detection implementation will have to overcome several issues. First, if implemented in populations with low health access, such detection will have sense only if care can follow. We saw that many known samples, both from patients and controls, are required to train animals. If implemented in an area with low access to gold standard detection, sample recruitment might be compromised.

Routine adoption of such detection raises the question of the number of samples which can be screened every day and its cost. From the studies reviewed, it seems that one dog, if efficient, could screen roughly a dozen of new

samples per day. Willis et al[78] report that only one new test was conducted per week, with training sessions in between, which is not very efficient for mass screening. Rats, however, seem to be able to screen more samples, as reported by Weetjens et al[56] "The use of trained rats to detect tuberculosis is reliable, potentially cheaper, and faster than sputum smear microscopy. One evaluation cage can contain more than 12 rats per day, and one rat can screen 140 samples in 40 minutes. The evaluation set-up can therefore process up to 1680 samples per day, while a microscopist can process up to only a maximum of 40 samples per day (WHO recommends an average of 20 samples per day).[133,134]

Another important consideration is the prevalence of the disease to be detected. Indeed, if very few positive samples are present, this could lower animals' motivation and accuracy. Hence the importance of training sessions with regular new known positive samples.

As discussed in chapter 4.2.2, such detection will be helpful if the odor and/or the sampling localization is specific to a shortlist of diseases. If not, then in the case of an alert, medical staff will not know what to look for.

Free-running rapid detection might be useful for infectious diseases. Free-running proofs of concept have been published for *C. difficile* infection detection with encouraging results[70] However, such detection has not been proven yet to work in the field for other diseases. So far, published articles report successful proofs of concept in remote conditions (like for cancer). Free running detection has recently been presented as an objective by several teams working on SARS-COV-2 detection. For instance, Guest et al[109] report that their preparatory work indicates that 2 dogs could screen 300 people in 30 minutes, for example, the time it takes to disembark from a plane, and PCR would only need to be used to test those individuals identified as positive by the dogs. However, no study has demonstrated such application in real screening conditions in contact with people in public places so far. On a different disease, Taylor et al[89] report that despite being highly trained, dogs are vulnerable to distractions and other foreign stimuli in a unique social environment.[135] Concerning their study, Essler et al[107] report that though dogs have previously been shown to be able to discriminate between saliva samples of SARS-CoV-2 positive and negative patients, these studies are also using repeated presentations of the same samples. Thus, it is possible dogs can discriminate between their training set of positive and negative patient samples but are unable to generalize this odor to new samples. These considerations are major limitations that preclude short-term implementation. However, this relatively new field of research is progressing quickly, and future studies may address these issues.

Finally, disease diagnostics can be expensive and complex to implement because of costly infrastructure and instruments, the need for consumables, and high-skilled professionals (MSc, PhD, MD). In this context, several teams claim that medical remote scent detection with animals might be cheap, however, this has yet to be proven. Cornu et al[58] report that in their proof-of-principle study, they tested a limited number of subjects in a costly, long study that makes it difficult to conceive of extended use for this test in clinical practice. Similarly, Sonoda et al[60] declare "it may be difficult to introduce canine scent judgment into clinical practice owing to the expense and time required for the dog trainer and dog education." No socio-economic study on the subject was found.

5. Conclusions

According to Hackner et al,[77] a suitable screening method should provide a true negative rate of near to 100% to be sufficient for safe use. Despite the number of studies reporting the potential capacity of trained animals to be used as disease detectors in a clinical setting, no validation has been issued so far. Willis et al[78] alert that introducing canine diagnosis of cancer in the absence of adequate validation, and without external quality assurance measures in place, may raise some of the same patient safety issues as those highlighted by the British Medical Association in their 2005 report on unregulated screening tests.[136]

Interestingly, several teams do not recommend the use of such a technique routinely. For instance, Horvath et al[55] wrote that they do not believe that dogs may be used in clinical practice. Dogs as "living instruments" may be influenced by several factors before and during their work, leading to changes in the accuracy rates. However, under controlled circumstances, they may be used in experiments to further explore the odor of malignancies. In Willis et al,[59] researchers do not advocate the use of dogs in a clinical setting. The authors hope that a greater understanding of the VOC biomarkers associated with bladder cancer, and urological disease more widely, will help optimize the design of an electronic nose. It has been suggested by Taverna et al[72,73] that dogs could be used to explore the response to cancer treatment or relapses in conjunction with VOCs identification. Although no direct comparison studies have been performed, for now, dogs appear to outperform e-noses.[35,137-139]

The implementation of dogs to detect infections in a free-running setting (in contact with humans) has still to prove efficiency. For instance, Bomers et al[63] report that a limitation of using an animal as a diagnostic tool is that behavior is not fully predictable. The dog's reaction to other stimuli (eg, children's play, being beckoned, being offered a treat) illustrates that dogs are still prone to distraction despite a high level of training.

However, this research field has made considerable progress since 2004, research teams, programs, and networks are constituted, and the main scientific obstacles seem to have been identified. By carrying out studies on

materials, VOCs conservation conditions, and by better mastering the selection and variability of dogs, a rigorous process will undoubtedly lead to possible implementation. Medico-economic studies still need to be conducted.

Finally, the work done in chemistry on the olfactory signature of diseases is complementary and will probably help to understand better and standardize research conducted with animals. Subsequent progress on this subject should determine more clearly what will be possible to implement in the future.

Authors' Information

P.B. and I.F. are members of the KDOG program, a research project led by Institut Curie aiming at evaluating the capacity of trained dogs to detect breast cancer from skin secretion samples. This program receives funding and support from Royal Canin company and Seris Security company, and private donors.

Authors' Contributions

P.B.: Protocol writing, title, and abstract screening, data extraction, manuscript writing, discussion, and submission. M.L.: title and abstracts screening, data extraction, discussion, and full-text reading. C.A.: data extraction, discussion, and full-text reading. C.G., E.A.: Discussion and full-text reading. I.F.: Research idea, data extraction, discussion, and full-text reading. All authors read and approved the final manuscript.

Availability of Data and Materials

All data generated or analyzed during this study are included in this published article.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethics Approval and Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

ORCID iD

Michelle Leemans [ID] https://orcid.org/0000-0002-2028-2610

Supplemental Material

Supplemental material for this article is available online.

References

1. Ferlay J, Colombet M, Soerjomataram I, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer*. 2019;144:1941-1953.
2. Angle C, Waggoner LP, Ferrando A, Haney P, Passler T. Canine detection of the volatilome: a review of implications for pathogen and disease detection. *Front Vet Sci*. 2016;3:47.
3. Angle TC, Passler T, Waggoner PL, et al. Real-time detection of a virus using detection dogs. *Front Vet Sci*. 2016;2:79.
4. Yu C, Fan S, Sun Y, Pickwell-Macpherson E. The potential of terahertz imaging for cancer diagnosis: A review of investigations to date. *Quant Imaging Med Surg*. 2012;2: 33-45.
5. Buszewski B, Ligor T, Jezierski T, Wenda-Piesik A, Walczak M, Rudnicka J. Identification of volatile lung cancer markers by gas chromatography-mass spectrometry: comparison with discrimination by canines. *Anal Bioanal Chem*. 2012;404: 141-146.
6. Abd El, Qader A, Lieberman D, Shemer Avni Y, et al. Volatile organic compounds generated by cultures of bacteria and viruses associated with respiratory infections. *Biomed Chromatogr*. 2015;29:1783-1790.
7. Blanchet L, Smolinska A, Baranska A, et al. Factors that influence the volatile organic compound content in human breath. *J Breath Res*. 2017;11:016013.
8. Jain RB. Levels of selected urinary metabolites of volatile organic compounds among children aged 6-11 years. *Environ Res*. 2015;142:461-470.
9. Shirasu M, Touhara K. The scent of disease: volatile organic compounds of the human body related to disease and disorder. *J Biochem*. 2011;150:257-266.
10. de Lacy Costello B, Amann A, Al-Kateb H, et al. A review of the volatiles from the healthy human body. *J Breath Res*. 2014;8:014001.
11. Peel AM, Wilkinson M, Sinha A, Loke YK, Fowler SJ, Wilson AM. Volatile organic compounds associated with diagnosis and disease characteristics in asthma - A systematic review. *Respir Med*. 2020;169:105984.
12. Cainap C, Pop LA, Balacescu O, Cainap SS. Early diagnosis and screening in lung cancer. *Am J Cancer Res*. 2020;10: 1993-2009.
13. Campanella A, De Summa S, Tommasi S. Exhaled breath condensate biomarkers for lung cancer. *J Breath Res*. 2019;13: 044002.
14. Gouzerh F, Bessière JM, Ujvari B, Thomas F, Dujon AM, Dormont L. Odors and cancer: current status and future directions. *Biochim Biophys Acta - Rev Cancer*. 2022; 1877:188644.
15. Hanna GB, Boshier PR, Markar SR, Romano A. Accuracy and methodologic challenges of volatile organic compound-based exhaled breath tests for cancer diagnosis: a systematic review and meta-analysis. *JAMA Oncol*. 2019;5:e182815.

16. Marzorati D, Mainardi L, Sedda G, Gasparri R, Spaggiari L, Cerveri P. A review of exhaled breath: a key role in lung cancer diagnosis. *J Breath Res*. 2019;13:034001.

17. Ratiu IA, Ligor T, Bocos-Bintintan V, Mayhew CA, Buszewski B. Volatile organic compounds in exhaled breath as fingerprints of lung cancer, asthma and COPD. *J Clin Med*. 2021;10:1-41.

18. Behera B, Joshi R, Anil Vishnu GK, Bhalerao S, Pandya HJ. Electronic nose: a non-invasive technology for breath analysis of diabetes and lung cancer patients. *J Breath Res*. 2019;13:024001.

19. Leemans M, Bauër P, Cuzuel V, Audureau E, Fromantin I. Volatile organic compounds analysis as a potential novel screening tool for breast cancer: a systematic review. *Biomark Insights*. 2022;17:11772719221100709.

20. Nakhleh MK, Amal H, Jeries R, et al. Diagnosis and classification of 17 diseases from 1404 subjects via pattern analysis of exhaled molecules. *ACS Nano*. 2017;11:112-125.

21. Sethi S, Nanda R, Chakraborty T. Clinical application of volatile organic compound analysis for detecting infectious diseases. *Clin Microbiol Rev*. 2013;26:462-475.

22. Schmidt K, Podmore I. Current challenges in volatile organic compounds analysis as potential biomarkers of cancer. *J Biomark*. 2015;2015:981458.

23. Cambau E, Poljak M. Sniffing animals as a diagnostic tool in infectious diseases. *Clin Microbiol Infect*. 2020;26: 431-435.

24. Wells DL. Domestic dogs and human health: an overview. *Br J Health Psychol*. 2007;12:145-156.

25. 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. *Genome Res*. 2014;24:1485-1496.

26. Lesniak A, Walczak M, Jezierski T, Sacharczuk M, Gawkowski M, Jaszczak K. Canine olfactory receptor gene polymorphism and its relation to odor detection performance by sniffer dogs. *J Hered*. 2008;99:518-527.

27. Pirrone F, Albertini M. Olfactory detection of cancer by trained sniffer dogs: a systematic review of the literature. *J Vet Behav*. 2017;19:105-117.

28. Pomerantz A, Blachman-Braun R, Galnares-Olalde JA, Berebichez-Fridman R, Capurso-García M. The possibility of inventing new technologies in the detection of cancer by applying elements of the canine olfactory apparatus. *Med Hypotheses*. 2015;85:160-172.

29. Browne C, Stafford K, Fordham R. The use of scent-detection dogs. *Ir Vet J*. 2006;59:97-104.

30. Williams M, Johnston JM. Training and maintaining the performance of dogs (Canis familiaris) on an increasing number of odor discriminations in a controlled setting. *Appl Anim Behav Sci*. 2002;78:55-65.

31. Poling A, Weetjens BJ, Cox C, Beyene NW, Sully A. Using giant African pouched rats (Cricetomys Gambianus) to detect landmines. *Psychol Rec*. 2010;60:715-728.

32. Williams H, Pembroke A. Sniffer dogs in the melanoma clinic? *Lancet*. 1989;333:734.

33. Campbell LF, Farmery L, George SM, Farrant PB. Canine olfactory detection of malignant melanoma. *BMJ Case Rep*. 2013;2013:bcr2013008566.

34. Church J, Williams H. Another sniffer dog for the clinic? *Lancet*. 2001;358:930.

35. Catala A, Grandgeorge M, Schaff JL, Cousillas H, Hausberger M, Cattet J. Dogs demonstrate the existence of an epileptic seizure odour in humans. *Sci Rep*. 2019; 9:4103.

36. Guest C, Pinder M, Doggett M, et al. Trained dogs identify people with malaria parasites by their odour. *Lancet Infect Dis*. 2019;19:578-580.

37. Kanaan R, Farkas N, Hegyi P, et al. Rats sniff out pulmonary tuberculosis from sputum: a diagnostic accuracy meta-analysis. *Sci Rep*. 2021;11:1877.

38. Kantele A, Paajanen J, Turunen S, et al. Scent dogs in detection of COVID-19: triple-blinded randomised trial and operational real-life screening in airport setting. *BMJ Glob Health*. 2022;7:e008024.

39. Lippi G, Plebani M. Diabetes alert dogs: a narrative critical overview. *Clin Chem Lab Med*. 2019;57:452-458.

40. Piqueret B, Bourachot B, Leroy C, et al. Ants detect cancer cells through volatile organic compounds. *iScience*. 2022; 25:103959.

41. Jendrny P, Twele F, Meller S, Osterhaus ADME, Schalke E, Volk HA. Canine olfactory detection and its relevance to medical detection. *BMC Infect Dis*. 2021;21:838.

42. Jezierski T, Walczak M, Ligor T, Rudnicka J, Buszewski B. Study of the art: canine olfaction used for cancer detection on the basis of breath odour. Perspectives and limitations. *J Breath Res*. 2015;9:027001.

43. Hirotsu T, Sonoda H, Uozumi T, et al. A highly accurate inclusive cancer screening test using Caenorhabditis elegans scent detection. *PLoS One*. 2015;10:e0118699.

44. Desikan P. Rapid diagnosis of infectious diseases: the role of giant African pouched rats, dogs and honeybees. *Indian J Med Microbiol*. 2013;31:114-116.

45. Suckling DM, Sagar RL. Honeybees Apis mellifera can detect the scent of Mycobacterium tuberculosis. *Tuberculosis*. 2011;91:327-328.

46. Dorman D, Foster M, Fernhoff K, Hess P. Canine scent detection of canine cancer: a feasibility study. *Vet Med*. 2017;8: 69-76.

47. Fischer-Tenhagen C, Wetterholm L, Tenhagen BA, Heuwieser W. Training dogs on a scent platform for oestrus detection in cows. *Appl Anim Behav Sci*. 2011;131:63-70.

48. Fischer-Tenhagen C, Tenhagen BA, Heuwieser W. Short communication: ability of dogs to detect cows in estrus from sniffing saliva samples. *J Dairy Sci*. 2013;96:1081-1084.

49. Golden GJ, Grady MJ, McLean HE, et al. Biodetection of a specific odor signature in mallard feces associated with infection by low pathogenic avian influenza A virus. *PLoS One*. 2021;16:e0251841.

50. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for systematic reviews and meta-analyses: the PRISMA Statement. *PLoS Med*. 2009;6:e1000097.

51. Willis CM, Church SM, Guest CM, et al. Olfactory detection of human bladder cancer by dogs: proof of principle study. *BMJ*. 2004;329:712-714.

52. Pickel D, Manucy GP, Walker DB, Hall SB, Walker JC. Evidence for canine olfactory detection of melanoma. *Appl Anim Behav Sci*. 2004;89:107-116.

53. McCulloch M, Jezierski T, Broffman M, Hubbard A, Turner K, Janecki T. Diagnostic accuracy of canine scent detection in early- and late-stage lung and breast cancers. *Integr Cancer Ther*. 2006;5:30-39.

54. Gordon RT, Schatz CB, Myers LJ, et al. The use of canines in the detection of human cancers. *J Altern Complement Med*. 2008;14:61-67.

55. Horvath G, Järverud GA, Järverud S, Horváth I. Human ovarian carcinomas detected by specific odor. *Integr Cancer Ther*. 2008;7:76-80.

56. Weetjens BJ, Mgode GF, Machang'u RS, et al. African pouched rats for the detection of pulmonary tuberculosis in sputum samples. *Int J Tuberc Lung Dis*. 2009;13:737-743.

57. Horvath G, Andersson H, Paulsson G. Characteristic odour in the blood reveals ovarian carcinoma. *BMC Cancer*. 2010; 10:643.

58. Cornu JN, Cancel-Tassin G, Ondet V, Girardet C, Cussenot O. Olfactory detection of prostate cancer by dogs sniffing urine: A step forward in early diagnosis. *Eur Urol*. 2011; 59:197-201.

59. Willis CM, Britton LE, Harris R, Wallace J, Guest CM. Volatile organic compounds as biomarkers of bladder cancer: sensitivity and specificity using trained sniffer dogs. *Cancer Biomark*. 2011;8:145-153.

60. Sonoda H, Kohnoe S, Yamazato T, et al. Colorectal cancer screening with odour material by canine scent detection. *Gut*. 2011;60:814-819.

61. Ehmann R, Boedeker E, Friedrich U, et al. Canine scent detection in the diagnosis of lung cancer: revisiting a puzzling phenomenon. *Eur Respir J*. 2012;39:669-676.

62. Walczak M, Jezierski T, Górecka-Bruzda A, Sobczyńska M, Ensminger J. Impact of individual training parameters and manner of taking breath odor samples on the reliability of canines as cancer screeners. *J Vet Behav*. 2012;7:283-294.

63. Bomers MK, van Agtmael MA, Luik H, van Veen MC, Vandenbroucke-Grauls CM, Smulders YM. Using a dog's superior olfactory sensitivity to identify Clostridium difficile in stools and patients: proof of principle study. *BMJ*. 2012;345:e7396-e7396.

64. Mgode GF, Weetjens BJ, Nawrath T, et al. Diagnosis of tuberculosis by trained African giant pouched rats and confounding impact of pathogens and microflora of the respiratory tract. *J Clin Microbiol*. 2012;50:274-280.

65. Mahoney A, Weetjens BJ, Cox C, et al. Pouched rats' detection of tuberculosis in human sputum: comparison to culturing and polymerase chain reaction. *Tuberc Res Treat*. 2012;2012:716989.

66. Horvath G, Andersson H, Nemes S. Cancer odor in the blood of ovarian cancer patients: a retrospective study of detection by dogs during treatment, 3 and 6 months afterward. *BMC Cancer*. 2013;13:396.

67. Amundsen T, Sundstrøm S, Buvik T, Gederaas OA, Haaverstad R. Can dogs smell lung cancer? First study using exhaled breath and urine screening in unselected patients with suspected lung cancer. *Acta Oncol*. 2014;53: 307-315.

68. Rudnicka J, Walczak M, Kowalkowski T, Jezierski T, Buszewski B. Determination of volatile organic compounds as potential markers of lung cancer by gas chromatography–mass spectrometry versus trained dogs. *Sens Actuators B Chem*. 2014;202:615-621.

69. Elliker KR, Sommerville BA, Broom DM, Neal DE, Armstrong S, Williams HC. Key considerations for the experimental training and evaluation of cancer odour detection dogs: lessons learnt from a double-blind, controlled trial of prostate cancer detection. *BMC Urol*. 2014;14:22.

70. Bomers MK, van Agtmael MA, Luik H, Vandenbroucke-Grauls CM, Smulders YM. A detection dog to identify patients with Clostridium difficile infection during a hospital outbreak. *J Infect*. 2014;69:456-461.

71. Rudnicka J, Walczak M, Jezierski T, Buszewski B. Is it possible to detect lung cancer by trained dogs? *Health Probl Civiliz*. 2015;2:19-26.

72. Taverna G, Tidu L, Grizzi F, et al. Olfactory system of highly trained dogs detects prostate cancer in urine samples. *J Urol*. 2015;193:1382-1387.

73. Taverna G, Tidu L, Grizzi F, et al. The ability of dogs to detect human prostate cancer before and after radical prostatectomy. *EC Vet Sci*. 2015;1:47-51.

74. Urbanová L, Vyhnánková V, Krisová Pacík D, Nečas A. Intensive training technique utilizing the dog's olfactory abilities to diagnose prostate cancer in men. *Acta Vet Brno*. 2015;84:77-82.

75. Yoel U, Gopas J, Ozer J, Peleg R, Shvartzman P. Canine scent detection of volatile elements, characteristic of malignant cells, in cell cultures. *Isr Med Assoc J*. 2015;17: 567-570.

76. Reither K, Jugheli L, Glass TR, et al. Evaluation of giant African pouched rats for detection of pulmonary tuberculosis in patients from a high-endemic setting. *PLoS One*. 2015;10:e0135877.

77. Hackner K, Errhalt P, Mueller MR, et al. Canine scent detection for the diagnosis of lung cancer in a screening-like situation. *J Breath Res*. 2016;10:046003.

78. Willis CM, Britton LE, Swindells MA, 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;175:1020-1029.

79. Maurer M, McCulloch M, Willey AM, Hirsch W, Dewey D. Detection of Bacteriuria by canine olfaction. *Open Forum Infect Dis*. 2016;3:ofw051.

80. Guerrero-Flores H, Apresa-García T, Garay-Villar, et al. A non-invasive tool for detecting cervical cancer odor by trained scent dogs. *BMC Cancer*. 2017;17:79.

81. Kitiyakara T, Redmond S, Unwanatham N, et al. The detection of hepatocellular carcinoma (HCC) from patients' breath using canine scent detection: a proof-of-concept study. *J Breath Res*. 2017;11:046002.

82. Guirao Montes Á, Molins López-Rodó L, Ramón Rodríguez I, et al. Lung cancer diagnosis by trained dogs. *Eur J Cardiothorac Surg*. 2017;52:1206-1210.

83. Bryce E, Zurberg T, Zurberg M, Shajari S, Roscoe D. Identifying environmental reservoirs of Clostridium difficile with a scent detection dog: preliminary evaluation. *J Hosp Infect*. 2017;97:140-145.

84. Koivusalo M, Vermeiren C, Yuen J, Reeve C, Gadbois S, Katz K. Canine scent detection as a tool to distinguish meticillin-resistant Staphylococcus aureus. *J Hosp Infect*. 2017;96:93-95.

85. Seo IS, Lee HG, Koo B, et al. Cross detection for odor of metabolic waste between breast and colorectal cancer using canine olfaction. *PLoS One*. 2018;13:e0192629.

86. Fischer-Tenhagen C, Johnen D, Nehls I, Becker R. A proof of concept: are detection dogs a useful tool to verify potential biomarkers for lung cancer? *Front Vet Sci*. 2018; 5:52.

87. Pacik D, Plevova M, Urbanova L, et al. Identification of sarcosine as a target molecule for the canine olfactory detection of prostate carcinoma. *Sci Rep*. 2018;8:4958.

88. Thuleau A, Gilbert C, Bauër P, et al. A new transcutaneous method for breast cancer detection with dogs. *Oncology*. 2019;96:110-113.

89. Taylor MT, McCready J, Broukhanski G, Kirpalaney S, Lutz H, Powis J. Using Dog Scent detection as a point-of-care tool to identify toxigenic Clostridium difficile in stool. *Open Forum Infect Dis*. 2018;5:ofy179.

90. Edwards TL, Ellis H, Watkins EE, et al. Tuberculosis detection by pouched rats: Opportunities for reinforcement under low-prevalence conditions. *Behav Processes*. 2018;155: 2-7.

91. Schoon GAA, De Jonge D, Hilverink P. How dogs learn to detect colon cancer—Optimizing the use of training aids. *J Vet Behav*. 2020;35:38-44.

92. Biehl W, Hattesohl A, Jörres RA, et al. VOC pattern recognition of lung cancer: a comparative evaluation of different dog- and eNose-based strategies using different sampling materials. *Acta Oncol*. 2019;58:1216-1224.

93. Feil C, Stein T, Forster A, et al. Diagnosis of lung cancer by canine olfactory detection in urine and breath samples. *J Clin Oncol*. 2019;37:e13067-e13067.

94. Guirao A, Molins L, Ramón I, et al. Trained dogs can identify malignant solitary pulmonary nodules in exhaled gas. *Lung Cancer*. 2019;135:230-233.

95. Junqueira H, Quinn TA, Biringer R, et al. Accuracy of canine scent detection of Non-Small cell lung cancer in blood serum. *J Am Osteopath Assoc*. 2019;119:413.

96. Murarka M, Vesley-Gross ZI, Essler JL, et al. Testing ovarian cancer cell lines to train dogs to detect ovarian cancer from blood plasma: A pilot study. *J Vet Behav*. 2019;32: 42-48.

97. Protoshchak VVP, Andreev EAA, Karpushhenko EGK, et al. Prostate cancer and dogs' sense of smell: opportunities of noninvasive diagnostics. *Urologiia*. 2019;5:22-26.

98. Li C, Zurberg T, Kinna J, et al. Using scent detection dogs to identify environmental reservoirs of Clostridium difficile: Lessons from the field. *Can J Infect Control*. 2019;34: 93-95.

99. Kure S, Iida S, Yamada M, et al. Breast cancer detection from a urine sample by dog sniffing. *Biology*. 2021; 10:517.

100. Yamamoto A, Kamoi S, Kurose K, et al. The trained sniffer dog could accurately detect the urine samples from the patients with cervical cancer, and even cervical intraepithelial neoplasia grade 3: a pilot study. *Cancers*. 2020; 12:3291.

101. Mazzola SM, Pirrone F, Sedda G, et al. Two-step investigation of lung cancer detection by sniffer dogs. *J Breath Res*. 2020;14:026011.

102. Grandjean D, Sarkis R, Lecoq-Julien C, et al. Can the detection dog alert on COVID-19 positive persons by sniffing axillary sweat samples? A proof-of-concept study. *PLoS One*. 2020;15:e0243122-NaN1.

103. Jendrny P, Schulz C, Twele F, et al. Scent dog identification of samples from COVID-19 patients - a pilot study. *BMC Infect Dis*. 2020;20:536.

104. Vesga O, Valencia AF, Mira A, et al. Dog savior: immediate scent-detection of SARS-COV-2 by trained dogs. *bioRxiv*. 2020;158105. doi:10.1101/2020.06.17.158105

105. Guest C, Harris R, Sfanos KS, et al. Feasibility of integrating canine olfaction with chemical and microbial profiling of urine to detect lethal prostate cancer. *PLoS One*. 2021;16:e0245530.

106. Eskandari E, Ahmadi Marzaleh M, Roudgari H, et al. Sniffer dogs as a screening/diagnostic tool for COVID-19: a proof of concept study. *BMC Infect Dis*. 2021;21:243.

107. Essler JL, Kane SA, Nolan P, et al. Discrimination of SARS-CoV-2 infected patient samples by detection dogs: A proof of concept study. *PLoS One*. 2021;16:e0250158.

108. Grandjean D. Use of canine olfactory detection for covid-19 testing study on U.A.E. Trained detection dog sensitivity. *Open Access J Vet Sci Res*. 2021;6:1-14.

109. Guest C, Dewhirst SY, Lindsay SW, et al. Using trained dogs and organic semi-conducting sensors to identify asymptomatic and mild SARS-CoV-2 infections: an observational study. *J Travel Med*. 2022;29:taac043.

110. Jendrny P, Twele F, Meller S, et al. Scent dog identification of SARS-CoV-2 infections in different body fluids. *BMC Infect Dis*. 2021;21:707.

111. Petri AL, Høgdall C, Christensen IJ, et al. Sample handling for mass spectrometric proteomic investigations of human urine. *Proteomics Clin Appl*. 2008;2:1184-1193.

112. Edwards TL, Browne CM, Schoon A, Cox C, Poling A. Animal olfactory detection of human diseases: guidelines and systematic review. *J Vet Behav*. 2017;20:59-73.

113. Sato T, Katsuoka Y, Yoneda K, et al. Sniffer mice discriminate urine odours of patients with bladder cancer: a proof-of-principle study for non-invasive diagnosis of cancer-induced odours. *Sci Rep*. 2017;7:14628.

114. Lazarowski L, Krichbaum S, DeGreeff LE, et al. Methodological considerations in canine olfactory detection research. *Front Vet Sci*. 2020;7:1-17.

115. Lit L, Schweitzer JB, Oberbauer AM. Handler beliefs affect scent detection dog outcomes. *Anim Cogn*. 2011;14: 387-394.

116. Weber CM, Cauchi M, Patel M, et al. Evaluation of a gas sensor array and pattern recognition for the identification of bladder cancer from urine headspace. *Analyst*. 2011;136: 359-364.

117. Xiang L, Wu S, Hua Q, Bao C, Liu H. Volatile organic compounds in human exhaled breath to diagnose gastrointestinal cancer: a meta-analysis. *Front Oncol*. 2021;11:606915.

118. Zhou W, Tao J, Li J, Tao S. Volatile organic compounds analysis as a potential novel screening tool for colorectal cancer: a systematic review and meta-analysis. *Medicine*. 2020;99:e20937.

119. Colquhoun TA, Schwieterman ML, Gilbert JL, et al. Light modulation of volatile organic compounds from petunia flowers and select fruits. *Postharvest Biol Technol*. 2013;86:37-44.

120. Kozicki M, Guzik K. Comparison of VOC emissions produced by different types of adhesives based on test chambers. *Materials*. 2021;14:1924.

121. Lan H, Li S, Yang J, Li J, Yuan C, Guo A. Effects of light exposure on chemical and sensory properties of storing meili rosé wine in colored bottles. *Food Chem*. 2021; 345:128854.

122. Markowicz P, Larsson L. Influence of relative humidity on VOC concentrations in indoor air. *Environ Sci Pollut Res Int*. 2015;22:5772-5779.

123. Niu Z, Kong S, Zheng H, et al. Temperature dependence of source profiles for volatile organic compounds from typical volatile emission sources. *Sci Total Environ*. 2021;751:141741.

124. Holz O, Mücke M, Zarza P, Loppow D, Jörres RA, Magnussen H. Freezing of homogenized sputum samples for intermittent storage. *Clin Exp Allergy*. 2001;31: 1328-1331.

125. Wilson AD, Baietto M. Advances in electronic-nose technologies developed for biomedical applications. *Sensors*. 2011;11:1105-1176.

126. Jamieson LTJ, Baxter GS, Murray PJ. Identifying suitable detection dogs. *Appl Anim Behav Sci*. 2017;195:1-7.

127. Slabbert JM, Odendaal JS. Early prediction of adult police dog efficiency—a longitudinal study. *Appl Anim Behav Sci*. 1999;64:269-288.

128. Weiss E. Selecting Shelter Dogs for service dog training. *J Appl Anim Welf Sci*. 2002;5:43-62.

129. Turcsán B, Kubinyi E, Miklósi. Trainability and boldness traits differ between dog breed clusters based on conventional breed categories and genetic relatedness. *Appl Anim Behav Sci*. 2011;132:61-70.

130. Fischer-Tenhagen C, Johnen D, Heuwieser W, Becker R, Schallschmidt K, Nehls I. Odor perception by dogs: evaluating two training approaches for odor learning of Sniffer Dogs. *Chem Senses*. 2017;42:435-441.

131. Mahoney AM, Weetjens BJ, Cox C, et al. Using giant African pouched rats to detect tuberculosis in human sputum samples: 2010 findings. *Pan Afr Med J*. 2011;9:28.

132. Kaulfuss P, Mills DS. Neophilia in domestic dogs (Canis familiaris) and its implication for studies of dog cognition. *Anim Cogn*. 2008;11:553-556.

133. Fend R, Kolk AH, Bessant C, Buijtels P, Klatser PR, Woodman AC. Prospects for clinical application of electronic-nose technology to early detection of Mycobacterium tuberculosis in culture and sputum. *J Clin Microbiol*. 2006;44:2039-2045.

134. Hawken MP, Muhindi DW, Chakaya JM, Bhatt SM, Ng'ang'a LW, Porter JDH. Under-diagnosis of smear-positive pulmonary tuberculosis in Nairobi, Kenya. *Int J Tuberc Lung Dis*. 2001;5:360-363.

135. Hackner K, Pleil J. Canine olfaction as an alternative to analytical instruments for disease diagnosis: understanding "dog personality" to achieve reproducible results. *J Breath Res*. 2017;11:012001.

136. Abergavenny RD, London LE. BMA warns against unnecessary screening tests in private sector plan for genetic testing of German civil servants stirs controversy Global Fund pulls grants to Myanmar. *BMJ*. 2005;331:475.

137. Bijland LR, Bomers MK, Smulders YM. Smelling the diagnosis a review on the use of scent in diagnosing disease. *Neth J Med*. 2013;71: 300-307.

138. Szulejko JE, McCulloch M, Jackson J, McKee DL, Walker JC, Solouki T. Evidence for cancer biomarkers in exhaled breath. *IEEE Sens J*. 2010;10:185-210.

139. Walker DB, Walker JC, Cavnar PJ, et al. Naturalistic quantification of canine olfactory sensitivity. *Appl Anim Behav Sci*. 2006;97:241-254.