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

Running head: Canine olfactory detection of melanoma in-vivo

CM Willis[1], LE Britton[1], MA Swindells[2], EM Jones[3], AE Kemp[1], NL Muirhead[1], A Gul[1], RN Matin[4], L Knutsson[5], M Ali[1]

[1]Department of Dermatology, Amersham Hospital, Buckinghamshire Healthcare NHS Trust, Amersham, HP7 0JD, UK

[2]Search Dogs UK, 9 Church Road, Thornton-Cleveleys, Lancashire, FY5 2TX, UK

[3]Department of Statistical Science, University College London, Gower Street, London, WC1E 6BT, UK

[4]Department of Dermatology, Churchill Hospital, Oxford University Hospitals NHS Foundation Trust, OX3 7LE, UK

[5]Faculty of Medicine and Health Sciences, Linköping University, 581 83 Linköping, Sweden

Corresponding author: Dr Carolyn Willis, Department of Dermatology, Amersham Hospital, Amersham, HP7 0JD, UK; Tel. No.: +44 (0)1494 734618; Fax. No.: +44 (0)1494 734635; Email: carolyn.willis@buckshealthcare.nhs.uk

Funding: The study was funded by the Amerderm Research Trust (Registered UK charity no. 1114684). The charity played no part in the design or conduct of the study, nor in the manuscript preparation and/or publication decisions.

Conflict of interest disclosure: MAS owns Search Dogs UK, a commercial company specialising in the training of dogs for olfactory detection purposes; however, final accuracy testing was managed by researchers based at Amersham Hospital and conducted in a double-blind fashion. The remaining authors declare no competing interests.

What's already known about this topic?

- Some tumours have been shown to emit characteristic patterns of volatile organic compounds (VOCs).
- There is anecdotal evidence of canine olfactory detection of melanoma, and some supporting analytical findings, but robust data for a distinct profile of melanoma-associated VOCs are lacking.

What does this study add?

- We demonstrated by canine olfaction that invasive melanoma in-vivo emits a profile of odorous VOCs which differs from those of basal cell carcinoma, benign naevi and healthy skin.
- The study provides further evidence of the potential for VOCs to be utilised as biomarkers for the objective, non-invasive diagnosis of melanoma.

Abstract

Background Volatile organic compounds (VOCs) are continuously released by the body during normal metabolic processes, but their profiles change in the presence of cancer, offering an objective, non-invasive approach to diagnosis. Robust evidence that invasive melanoma in-vivo emits a characteristic VOC signature is currently lacking.

Objective To conduct a canine olfactory, proof of principle study, to investigate whether VOCs captured from the surface of invasive melanoma are distinguishable from those of basal cell carcinoma (BCC), benign naevi and healthy skin in-vivo.

Methods After a 13 month training period, the dog's ability to discriminate melanoma was evaluated in a series of 20 double-blind tests, each requiring the selection of one melanoma sample from amongst nine controls (three each of BCC, naevi and healthy skin; all samples being new to the dog), assuring a power of 88% with respect to the analysis.

Results The dog correctly selected the melanoma sample on nine of the 20 occasions, a success rate of 45% (exact 95% CI [0.23, 0.68]), compared with the 10% expected by chance alone. A one-sided exact binomial test gave a p-value of <0.001, supporting the hypothesis that samples were not chosen at random, but that some degree of VOC signal from the melanoma samples significantly increased the probability of their detection. Use of a discrete choice model confirmed melanoma as the most influential of the recorded medical/personal covariates in determining the dog's choice of sample. Notably, accuracy rates based on familiar samples during training were not a reliable indicator of the dog's ability to distinguish melanoma, when confronted with new, unknown samples.

Conclusions Our study demonstrates that invasive melanoma in-vivo releases odorous VOCs which are distinct from those of BCC, benign naevi and healthy skin, adding to the body of evidence that the volatile metabolome of melanoma contains diagnostically useful biomarkers.

Keywords: malignant melanoma; volatile organic compound; canine olfactory detection; cancer biomarker; diagnosis

Background

Accurate and early detection of malignant melanoma is essential, both to maximise survival rates and to reduce the number of pigmented lesions unnecessarily excised.[1] However, with diagnostic accuracy rates based on visual evaluation running from as low as 22% amongst general practitioners, to between 40% and 95% for dermatologists,[2-6] there is a need for more reliable, objective methods to assist preliminary diagnosis.[7,8]

A non-invasive diagnostic approach which is currently showing promise for a number of disorders, including cancer, is the profiling of volatile organic compounds (VOCs). These high vapour pressure compounds are continuously released from the body during normal metabolic processes,[9,10] but show qualitative and quantitative changes in the presence of disease.[11-13] Analytical techniques, such as gas chromatography mass spectrometry (GC-MS), are commonly used in their investigation,[9-14] but, since many VOCs are odorous, olfactory detection is also possible.[15-22]

The concept that malignant melanoma may emit a different pattern of VOCs from surrounding healthy skin and benign lesions, in fact, originated from canine behaviour linked to olfaction. In their anecdotal report published in The Lancet in 1989, Williams and Pembroke described the persistent interest shown by a pet dog toward a lesion on its owner's leg, which, on excision, proved to be a melanoma.[23]

Building on this and other preliminary findings,[24] D'Amico et al applied a gas sensor array – otherwise known as an electronic nose – to compare the spectrum of VOCs emitted from the surface of in-vivo melanomas with those of benign naevi in the same individuals, achieving accuracy rates for their identification of 70% and 90%, respectively, albeit with small sample sizes.[25] Further supporting evidence for a characteristic profile of VOCs for melanoma has also come from more recent GC-MS and electronic nose studies of biopsy material and cultured cell lines.[26-29]

Advancing the development of VOC-based diagnostic aids for melanoma, however, calls for a greater understanding of the volatile metabolite signatures released directly from the surface of a variety of different skin lesions. To that end, we decided to make use of dogs' exquisite olfactory sensitivity, innate pattern recognition skills and proven cancer detection ability,[16-22] to determine, with greater rigour than previous studies, whether the odour of melanoma differs from those of healthy skin and benign naevi, and to examine, for the first

time, whether the odour of melanoma differs from that of another skin cancer, namely basal cell carcinoma (BCC). It also gave us the opportunity to assess the suitability of a method for skin VOC collection, pragmatically chosen as the most clinically practical, in preparation for future GC-MS studies.

An experimental design similar to that of our earlier proof of principle for canine olfactory detection of bladder cancer[16] was employed, the dog, in this case, being required to distinguish one melanoma sample from within a line-up of nine controls.

Materials and Methods

The study was approved by the National Research Ethics Committee London – Brent (REC Ref. 13/LO/1491), and all participants gave written, informed consent. Procedures were carried out in accordance with the Declaration of Helsinki of 1975, as revised in 2013.

Participants

In total, 741 study participants were recruited between October 2013 and June 2015. Eligible patients were identified from dermatology outpatient clinics, whilst healthy control subjects were recruited from amongst staff at Amersham Hospital and their families and friends. Overall, 593 patients were recruited from within the Buckinghamshire Healthcare NHS Trust, with a further nine patients coming from Oxfordshire University Hospitals NHS Foundation Trust. Of the healthy controls, 33 individuals had benign melanocytic naevi, diagnosed visually by an experienced dermatologist, but not surgically removed, with the remaining 106 subjects having normal, healthy skin.

In order to provide a logical progression of samples during training, patients and those control subjects with naevi were assigned to one of eight subgroups (Groups 1-8), depending on the diagnosis of their skin lesion, whilst volunteers with normal skin were categorised into one of four subgroups (Groups 9-12), depending on their age and Fitzpatrick skin type[30] (Table 1). Details of the characteristics of the subjects within the 12 groups are given in Table 2.

Demographic details and medical histories (including oral and topical medication history, and personal history of diabetes) were obtained from all participants. Lesion details, including site, maximum diameter and histological diagnosis, where appropriate, were also recorded.

Comprehensive dietary and lifestyle-related metadata were collected to allow for evaluation of potential confounding factors. An emphasis was placed on those factors known to affect the profile of VOCs released from the skin, including smoking, alcohol consumption, active dieting/fasting, anxiety level, strenuous exercise, and ingestion of certain foodstuffs, such as garlic and onions.[31] Given the ubiquity of perfumed personal and household products, subjects were only asked to refrain from the use of perfume itself on the day of sampling.

Exclusion criteria included lesion ulceration, a current or past history (< 5 years) of any non-dermatological malignancy, mental incapacitation and age < 18 years. Unless specifically recruited for the condition, individuals were excluded if they had a current or past history (< 5 years) of malignant melanoma, melanoma in-situ, lentigo maligna, squamous cell carcinoma or Bowen's disease, or a current history of BCC, actinic keratosis or dysplastic naevi. Concurrent benign lesions, such as seborrhoeic keratoses, solar lentigines and benign melanocytic naevi, as well as inflammatory conditions, such as hand eczema, were permitted, providing they were located a minimum of 20 cm from the lesion of interest.

Skin VOC sampling

Volatile organic compounds were collected from patients, immediately prior to excision or biopsy of their lesion, using 100% cotton gauze pads (plain sterile gauze swabs BP [12 ply], Synergy Health, Chorley, Lancashire, UK),[32,33] cut to size and applied directly to the surface of the lesion. Overlaid on top of the gauze to prevent VOC loss into the atmosphere were slightly larger pieces of low odour Nalophan® plastic,[14] cut from Bacofoil roasting bags (Wrap Film Systems Ltd., Telford, Shropshire, UK), which were secured to the skin on all sides using 2.5 cm wide Micropore tape (3M Healthcare, Bracknell, Berkshire, UK).

Healthy skin of normal control subjects (matched to patients' melanoma sites) was sampled in an identical manner, using a standard 2 cm^2 sized gauze pad. Procedural blanks, designed to collect background VOCs, were prepared in parallel with the subjects' skin samples, using gauze pads exposed to the atmosphere for 10 min, and then sandwiched between two layers of Nalophan® and sealed with Micropore tape for a further 10 min.

Gauze pads were left in contact with the skin for 15 min, after which time they (and the procedural blanks) were placed immediately in 7 mL, screw neck, amber glass vials (SCHOTT; Adelphi Healthcare Packaging, Haywards Heath, West Sussex, UK), closed with 18 mm black urea caps with foil liners (Wheaton UK Ltd., Rochdale, UK), frozen and stored at -40 ^{o}C until required.

Over 92% of subjects (and 100% of those used in the final testing phase) were sampled in a dedicated clinical research room (Amersham Hospital), at a temperature maintained at 22 ^{o}C +/- 1 ^{o}C. At other hospital sites, the temperature ranged from 20 ^{o}C to 24 ^{o}C. Powder-free vinyl gloves were worn by researchers and dog trainers throughout.

Training of the dog

Two dogs began training in January 2014, but, due to the time commitment required, the owner/trainer of one of them (LK) withdrew from the study later that year. Only one dog, a working Labrador (Ronnie, male, 2 years old at the start of the study; Fig. 1), which was obedience trained, but had no prior training in scent discrimination, therefore completed the study. His owner/trainer (MAS) has over 40 years' experience of selecting and training dogs for scent discrimination tasks on behalf of such organisations as the police and customs and excise, and has also been involved in a number of university-led research projects in the fields of forensics and wildlife conservation.

Training was conducted over a 13 month period, the ultimate goal being that the dog should be able to distinguish one melanoma odour sample from within a line-up containing nine non-melanoma controls (all samples being entirely new to the dog) with a success rate, after 20 test sets, greater than would be expected by chance alone.

Operant conditioning to attain the desired alert of 'freezing' in front of the positive melanoma sample was achieved by positive reinforcement, using ball play as the reward, in conjunction with the clicker. Early recognition of the melanoma scent was facilitated by search and find games using samples from several different patients. These were gradually replaced over a period of a few weeks by a more structured discriminatory search for the melanoma from amongst controls, along a line of, first two, and then up to ten, samples. The gauze-containing vials, which remained stored at -40 ^{o}C at the training facility when not in use, were placed in 2.5 cm diameter holes drilled 40 cm apart along wooden planks to a depth sufficient to prevent the dog dislodging them. The vial necks were too narrow to allow the dog to make direct contact with the samples. Prior to training (and testing), the samples were allowed to warm up to ambient temperature. To preserve the odour as much as possible (as samples were used on multiple occasions), the vial tops were removed immediately before the dog entered, and replaced as soon as the search was complete.

Control samples, against which successive melanoma samples were run, were introduced in a step-wise fashion of increasing pathological significance (i.e. descending order of the sample groups given in Table 1), beginning with procedural blanks, and gradually progressing to normal skin controls and then to the various non-melanoma lesion types. Initially, we did not restrict the range of non-melanoma control lesions used in training; however, as explained below, this later became necessary.

At no time was there more than one melanoma sample in a run; however, in order to minimize expectation bias, the trainer's assistant included random, blinded, no-melanoma runs (i.e. all negative controls) during each training session. Training took place four times a week, each session lasting about 30 minutes, with the dog being worked, off-lead, on three or four runs of different combinations of samples.

Preliminary assessment of the dog

Six months into training, the dog had been exposed to the odours from 17 invasive melanomas and almost 200 control odours, taken from over 50 different types of non-melanoma skin lesion. At this stage, MAS reported back a near 100% success rate in identifying the melanomas, and so it was decided to begin a series of double-blind tests (methodology as described below, all samples being new to the dog, with the nine controls being selected from Groups 3 to 12). However, after 13 runs, the dog had successfully identified only one of the melanoma samples, with no particular pattern to the chosen control samples, suggesting that the dog had been merely scent-matching known samples, and had failed to learn a generic odour for melanoma.

We speculated that the use of so many diverse control lesions might have been hampering the dog's ability to find the 'common denominator' between melanoma samples, and so took the decision to simplify the task, restricting the types of control samples to just three categories: BCC (Group 4), benign naevi (Group 6) and normal skin (Groups 9 to 12). Training with all other samples (i.e. those in Groups 2,3,5,7 and 8) was stopped, and fresh melanoma and appropriate control samples were collected and passed to MAS to further train the dog.

Final testing of the trained dog

In early 2015, the dog had been trained on a total of 41 melanoma samples, 114 BCCs, 51 benign naevi and 46 normal skin samples, and the results of mock tests were more encouraging. Testing was recommenced between February and June 2015, and 20

consecutive double-blind tests were carried out using new samples throughout. No duplicate sample was used within or between test sets, and all samples came from different individuals. All lesions had histological diagnoses.

Each test set consisted of one melanoma, three BCCs, three benign naevi and three normal skin samples. Careful consideration was given to the combination of controls to ensure that key characteristics of the melanoma samples were adequately matched (Table 3; Supplementary File 1). Where gender-specific panels were not possible (due to the on-going nature of sample collection), at least two controls with the same gender as the melanoma patient were used. Two or more controls were of equivalent age or older than the melanoma subject, and at least three controls were matched with respect to the site of the lesion and the size of area sampled. Matching with two or more controls was also carried out for Fitzpatrick skin type, diabetes status, smoking, alcohol and garlic consumption.

Sample vials were numerically coded from 0 to 9 at Amersham Hospital, using the random number generator in Excel, and were then sent by express delivery to the training centre, where they were immediately refrozen. For each test, the assistant at the training centre (who was not part of the hospital team) laid out the thawed samples in numerical order (although the positions could be changed, if requested by MAS, to confirm the dog's alert to a particular vial, or to check for possible positional bias). As soon as MAS was satisfied with the dog's alert, he named the putative positive sample, and CW, in telephone contact (from Amersham Hospital) with the assistant during the test, broke the sample codes. This allowed the dog to be immediately rewarded if correct, and for MAS to further train the dog on the line if incorrect. All samples were then added to the training pool. On the two occasions when the dog failed to make a clear alert, the codes remained unbroken, and the tests were re-run on another day. One new test was conducted per week, with training sessions in between.

Statistical analysis

Power and sample size determination

Following consultation with the trainer, it was decided that ten samples per test set (one melanoma plus nine controls) was the maximum that a dog could realistically be expected to manage in a single line-up. The primary outcome measure was thus the proportion of successes for the dog, compared with an expected value of 1 in 10 (0.1).

The package, nQuery Advisor 6.02 (Statistical Solutions Ltd., Cork, Ireland), was used to assess the powers for various sample size combinations, assuming a significance level of 5% for all calculations. Since in our original bladder cancer study, the dogs were correct 41% (0.41) of the time,[15] we calculated the powers associated with predicted rates of success of 0.35, 0.4 and 0.5, which, for 20 separate tests, were 88%, 94% and >99%, respectively. Therefore, 20 test sets, each consisting of one melanoma sample, and nine non-melanoma samples, were deemed to provide sufficient power.

Exact binomial test

Following the completion of the double-blind tests, the exact binomial test was used to examine the hypothesis that the samples were chosen at random in each test set (with the hypothesised probability of choosing the melanoma sample being 0.1). The one-sided test was deemed most suitable here as interest is in whether the melanoma samples were chosen *more often* than would be expected under the assumption of choosing samples at random. However, for completeness, a two-sided test was also included.

Model of sample choice

A discrete choice model (conditional logit) was used to explore the covariates most influential in determining the dog's choice of sample from each test set. The outcome was the sample chosen, with the characteristics of the sample (either melanoma, BCC, benign naevus or normal skin, with no further subtype classification), together with medical/personal information on the participants, used as potential covariates. The latter were first screened to assess their suitability for inclusion; binary covariates for which there was little variability between participants, or which were only weakly associated with the chosen samples, were excluded.

Seventeen covariates, in addition to lesion diagnosis, remained after this screening (Supplementary File 2), so a backward stepwise approach was used to choose an initial model.[34] Covariates with a p-value greater than 0.2 were sequentially deleted, followed by further elimination of the remaining covariates based on likelihood ratio tests.[35]

As the effective sample size was small, we applied Estrella's goodness-of-fit measure,[36] which penalises for both small sample size and the number of estimated parameters, to assess the fit of the final model to the data. There is evidence, however, that a sample size of 20 may be sufficient under the assumption that each test panel of ten skin samples is equivalent.[37]

Results

Table 4 gives a summary of the samples chosen by the dog during the 20 double-blind test runs (see also Supplementary File 1 for metadata on the 200 subjects), with Table 5 summarizing the tumour characteristics of the melanoma samples.

With 20 test sets, each containing ten samples, and only one classified as melanoma in each set, the dog would be expected to correctly identify the melanoma samples by chance alone on four occasions *at most* (one-sided binomial test for four successes in 20 tests gives a p-value of 0.13). However, the dog correctly selected the melanoma sample in nine of the 20 tests (proportion correctly identified was 0.45, with exact 95% CI [0.23, 0.68]). The one-sided exact binomial test gave a p-value of less than 0.001, supporting the hypothesis that samples were not chosen at random from each test set, but that some degree of VOC signal from the melanoma samples significantly increased the probability of their detection. The equivalent two-sided test resulted in a p-value which was also less than 0.001.

In the 11 tests where an alternative sample category to melanoma was chosen, six were BCC samples, and five were benign naevi. Since no normal healthy skin samples were chosen from any test set, the exact binomial test was further used to investigate whether samples were chosen at random from within the seven non-normal skin samples. Again, both one- and two-sided tests yielded p-values of less than 0.001, and even after applying the very conservative Bonferroni correction for multiple testing,[38] all p-values were still significant at the 0.01 level.

The discrete choice model pointed to a number of covariates which were seemingly associated with a sample being chosen (Table 6); nonetheless, melanoma remained the most significant predictor in sample choice, based on the p-value obtained for each variable in the model. Hence, the above result using an exact binomial test, where covariates were not considered, was not contradicted. Applying Estrella's goodness-of-fit measure gave a value of 0.73 (where a value of 0 indicates no fit, and 1 indicates perfect fit), providing evidence in favour of our model.

Although of clinical interest, there were insufficient data to examine the relationship between dog's sample choice and tumour characteristics of the melanoma (Breslow thickness etc.).

Nor was it considered appropriate to estimate the sensitivity and specificity of the dog, since this experimental design (where the dog alerts to one sample out of ten) precludes the independent assessment (i.e. a 50-50 chance of being correct) of each sample.

Discussion

In this study, we have shown that a dog can be trained to distinguish patients with invasive melanoma from those with BCC, benign naevi and healthy skin, on the basis of skin surface odour, with significantly greater accuracy than would be expected by chance alone. From this, we can infer that melanoma releases a profile of VOCs from its surface which is different from those of the three control groups included in the study – in the case of BCC, the first time, to our knowledge, that this has been reported for lesions in-vivo. The findings therefore add weight to a novel VOC-based diagnostic approach to melanoma, whilst at the same time validating our method for skin VOC collection, but, importantly, do not at this stage demonstrate clinical usefulness.

Our experimental design, in which the dog was remote from the subjects at all times, necessitated the capture and subsequent storage of the VOCs emitted from the skin. Although initially we considered using an absorptive fibre-filled funnel upturned over the area of interest, in a similar set-up to that designed by Gallagher et al,[9] it became apparent that too many of the lesions were located on uneven, awkward body sites for this to be successful. For this reason, we focused on direct contact sampling patches. These have the advantage of conforming to the anatomical contours of the body, can be cut to the appropriate size of the lesion, are low-cost and easy to handle.

With regard to the choice of material comprising the patches, this has been the subject of a number of studies pertaining to canine scenting in forensic science,[32,33] and the optimum fibre, in terms of the greatest variety and quantity of skin surface VOCs collected and readily released, appeared at the outset of this study to be 100% cotton; we therefore employed a widely available, sterile, pure cotton gauze throughout. As for the chosen sampling time of 15 mins, we were again guided by the forensic science literature.[32] Multiple use of the same sample 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.

Although the original aim was to train and test the dog against the widest possible range of control lesions, the dog failed to learn a generic odour for melanoma after six months, prompting the decision to limit the breadth of controls. Importantly, this lack of success had not been the impression of the trainer on the strength of his results using a large battery of samples on a repeated basis; only through double-blind testing using new samples did it become apparent. Our experience closely mirrors that of Elliker et al when attempting to train dogs to detect prostate cancer,[39] and serves to highlight the danger of citing canine accuracy rates based on training samples.

The need to reduce the diversity of control odours presented to the dog is perhaps not surprising, since, for over half of the lesion types, only one or two representatives were available during training. The challenge was compounded by the wide variation in lesion site, which is known to give rise to significant qualitative differences in VOC emission.[9] Furthermore, there is a high likelihood of overlapping VOC profiles amongst melanocytic lesions displaying the full spectrum of cellular abnormality, from mild dysplasia through in-situ disease to invasive malignancy. A more generous and continuous supply of known positives and negatives during training may have overcome this, since the most successful period during the double-blind tests coincided with the simultaneous introduction of a large number of supplementary training samples, but, unfortunately, resources and the availability of suitable patients were limited.

Resources permitting, we would also have employed additional dogs to possibly increase the chances of a higher level of discrimination. We acknowledge that using only one dog was a study limitation; however, we had confidence in the ability of our trainer to select a dog with the necessary attributes.[40] Failing to achieve the high accuracies reported by others for different types of cancer[18,21,22] may therefore be a reflection of the difficulty of this particular task, rather than deficiencies on the part of the dog.

Restricting the scope of the controls had the indirect advantage, however, of providing a greater robustness to the statistical analysis, particularly in view of the many potential confounding factors, in addition to site, within the melanoma and control groups. When recruiting the subjects, we collected detailed and comprehensive medical and personal information, with around 50 variables recorded. Common to such high-dimensional datasets,[41] several covariates of statistical significance emerged in the discrete choice model, two of which, room temperature and skin type, are of dubious relevance to the data. The former fluctuated to such a minor degree (+/- 1 ^{o}C) that it is unlikely to have influenced the temperature of the occluded skin during sampling, whilst the latter would have become a

concern only if the dog had selected patients at either end of the Fitzpatrick scale; a bias towards subjects with skin types IV-VI, for example, may have been an indication that the dog was generalising on the odour of melanin itself, but this was clearly not the case. Alcohol consumption within the previous 24 hours does warrant consideration, however, but is unlikely to be a factor in the dog's choice of sample. None of the participants admitted to consuming alcohol on the actual day of sampling/surgery, and so the probability of significant amounts of alcohol remaining in their bloodstreams is low.

Although our findings lend support to the concept of a distinctive VOC signature for melanoma, our aim in conducting this study was not to potentially add weight to the notion that dogs can be used in actual clinical diagnosis. High sensitivities and specificities for some cancers have been reported in several research studies,[18,21,22] but the suitability of dogs for high-volume screening of unknown samples has yet to be determined.[42] Introducing canine diagnosis of cancer in the absence of adequate validation, and without external quality control assurance mechanisms 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, more generally.[43] However, within the context of research, we believe that utilising the exquisite olfactory sensitivity and innate pattern recognition skills of dogs represents a logical and potentially valuable step in the search for volatile biomarkers of cancer.

Acknowledgements

We greatly appreciate the assistance of Linda Swindells in training the dog, and the clinical support given by Cherry Grout, Kay Mumby, Dr Sophie Grabczynska, Dr Cristina Rodriguez, Dr Amal Eissa and Dr Vincy Abraham. We would also like to thank all the patients and volunteers who kindly took part in the study.

References

1. Argenziano G, Cerroni L, Zalaudek I *et al*. Accuracy in melanoma detection: a 10-year multicenter survey. *J Am Acad Dermatol* 2012; **67**:54-59.

2. Koelink CJ, Vermeulen KM, Kollen BJ *et al*. Diagnostic accuracy and cost-effectiveness of dermoscopy in primary care: a cluster randomized clinical trial. *J Eur Acad Dermatol Venereo*l 2014; **28**:1442-1449.

3. Kittler H, Pehamberger H, Wolff K, Binder M. Diagnostic accuracy of dermoscopy. *Lancet Oncol* 2002; **3**:159-165.

4. Rajpara SM, Botello AP, Townend J, Ormerod AD. Systematic review of dermoscopy and digital dermoscopy/ artificial intelligence for the diagnosis of melanoma. *Br J Dermatol* 2009; **161**:591-604.

5. Vestergaard ME, Macaskill P, Holt PE, Menzies SW. Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting. *Br J Dermatol* 2008; **159**:669-76.

6. Rolfe HM**.** Accuracy in skin cancer diagnosis: a retrospective study of an Australian public hospital dermatology department. *Australas J Dermatol* 2012; **53**:112-117.

7. Langley R, Rajadhyaksha M, Dwyer P *et al*. Confocal scanning laser microscopy of benign and malignant melanocytic skin lesions *in vivo*. *J Am Acad Dermatol* 2001; **45**:365-376.

8. Mayer JE, Swetter SM, Fu T, Geller AC**.** Screening, early detection, education, and trends for melanoma: current status (2007-2013) and future directions: Part I. Epidemiology, high-risk groups, clinical strategies, and diagnostic technology. *J Am Acad Dermatol* 2014; **71**:599-608.

9. Gallagher M, Wysocki CJ, Leyden JJ *et al*. Analyses of volatile organic compounds from human skin. *Br J Dermatol* 2008; **159**:780-791.

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 Mar; **8(1)**:014001. doi: 10.1088/1752-7155/8/1/014001.

11. Kataoka H, Saito K, Kato H, Masuda K. Noninvasive analysis of volatile biomarkers in human emanations for health and early disease diagnosis. *Bioanalysis* 2013; **5**:1443-1459.

12. Amann A, De Lacy Costello B, Miekisch W, Schubert J, Buszewski B, Pleil J, Ratcliffe N, Risby T**.** The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. *J Breath Res* 2014 Sep; **8(3)**:034001. doi: 10.1088/1752-7155/8/3/034001.

13. 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.

14. Turner C, Parekh B, Walton C *et al*. An exploratory comparative study of volatile compounds in exhaled breath and emitted by skin using selected ion flow tube mass spectrometry. *Rapid Commun Mass Spectrom* 2008; **22**:526-532.

15. Bijland LR, Bomers MK, Smulders YM. Smelling the diagnosis: a review on the use of scent in diagnosing disease. *Neth J Med* 2013 Jul-Aug; **71(6)**:300-7.

16. Willis CM, Church SM, Guest CM *et al*. Olfactory detection of human bladder cancer by dogs: proof of principle study. *BMJ [Br Med J]* 2004; **329**:712-715.

17. McCulloch M, Jezierski T, Broffman M *et al*. Diagnostic accuracy of canine scent detection in early- and late-stage lung and breast cancers. *Integr Cancer Ther* 2006; **5**:30-39.

18. Horvath G, Järverud GA, Järverud S, Horváth I. Human ovarian carcinomas detected by specific odor. *Integr Cancer Ther*. 2008 Jun; **7(2)**:76-80. doi: 10.1177/1534735408319058.

19. Willis CM, Britton LE, Harris R *et al*. Volatile organic compounds as markers of bladder cancer: sensitivity and specificity using trained sniffer dogs. *Cancer Biomark* 2010/2011; **8**:145-153.

20. Cornu J-N, 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.

21. Sonoda H, Kohnoe S, Yamazato T, et al. Colorectal cancer screening with odour material by canine scent detection *Gut* 2011; **60**:814-819. doi 10.1136/gut.2010.218305

22. 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.

23. Williams H, Pembroke A. Sniffer dogs in the melanoma clinic? *Lancet* 1989; **1**:734.

24. Pickel D, Manucy GP, Walker DB *et al*. Evidence for canine olfactory detection of melanoma. *Appl Anim Behav Sci* 2004; **89**:107-116.

25. D'Amico A, Bono R, Pennazza G *et al*. Identification of melanoma with a gas sensor array. *Skin Res Technol* 2008; **14**:226-236.

26. Abaffy T, Duncan R, Riemer DD *et al*. Differential volatile signatures from skin, naevi and melanoma: a novel approach to detect a pathological process. *PLoS One* 2010; doi: 10.1371/journal.pone.0013813.

27. Pennazza G, Santonico M, Martinelli E *et al*. Monitoring of melanoma released volatile compounds by a gas sensors array: from in vitro to in vivo experiments. *Sensors Actuators B: Chem* 2011; **154**:288-294.

28. Kwak J, Gallagher M, Ozdener MH *et a*l. Volatile biomarkers from human melanoma cells. *J Chromatogr B* 2013; **931**:90-96.

29. Abaffy T, Möller MG, Riemer DD *et al*. Comparative analysis of volatile metabolomics signals from melanoma and benign skin: a pilot study. *Metabolomics* 2013; **9**:998-1008.

30. Fitzpatrick TB. The validity and practicality of sun-reactive skin types I through VI. *Arch Dermatol* 1988; **124**:869-871.

31. Jiang R, Cudjoe E, Bojko B *et al*. A non-invasive method for in vivo skin volatile compounds sampling. *Anal Chim Acta* 2013; **804**:111-119.

32. Prada PA, Curran AM, Furton KG. The evaluation of human hand odor volatiles on various textiles: a comparison between contact and noncontact sampling methods. *J Forensic Sci* 2011; **56**:866-881.

33. DeGreeff LE, Curran AM, Furton FG. Evaluation of selected sorbent materials for the collection of volatile organic compounds related to human scent using non-contact sampling mode. *Forensic Sci Int* 2011; **209**:133-142.

34. Hosmer DW, Lemeshow S, Sturdivant RX. *Applied Logistic Regression*. Wiley, 2013.

35. Afifi A, May S, Clark V. *Practical Multivariate Analysis*. Chapman and Hall/CRC, 2012.
.
36. Estrella A. A new measure of fit for equations with dichotomous dependent variables. *J Bus Econ Stat* 1998; **16**:198-205.

37. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. *PharmacoEconomics* 2008; **26**:661–77.

38. McDonald JH*. Handbook of Biological Statistics (3rd ed.)*. Sparky House Publishing, 2014.

39. Elliker KR, Sommerville BA, Broom DM *et al*. 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. doi: 10.1186/1471-2490-14-22.

40. Rooney NJ, Bradshaw JWS, Almey H. Attributes of specialist search dogs – A questionnaire survey of UK dog handlers and trainers. *J Forensic Sci* 2004; **49**:300-306.

41. Fan J, Han F, Liu H. Challenges of big data analysis. *Natl Sci Rev* 2014; **1(2)**:293-314.

42. Jezierski T[1], 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 May 6; **9(2)**:027001. doi: 10.1088/1752-7155/9/2/027001.

43. Doctors warn patients against unregulated screening tests. http://web.bma.org.uk/pressrel.nsf/wall/585A1F1D25B59C468025706500559F03?OpenDocument [Accessed 15 June 2016].

Table 1. Categorisation of the different skin lesion types and normal skin samples to provide a logical progression of samples for dog training (N=741).

Diagnosis (no. of subjects)	Group	No. of subjects
Melanoma (T1-T4, Clark level > 2) *(superficial spreading [52], lentigo maligna melanoma [5], nodular [3], Spitzoid [1])*	1	61
Melanoma in-situ (Tis, Clark level 1)	2	27
Squamous cell carcinoma	3	22
Basal cell carcinoma *(subtypes: nodular [65], nodular and superficial [27], nodular and infiltrative [23], superficial [24], infiltrative [10], nodular, superficial and infiltrative [5], nodulocystic [5], pigmented superficial [5], pigmented nodular [4], nodulocystic and superficial [2], pigmented, nodular and superficial [2], nodulocystic and infiltrative [1], pigmented, nodular, superficial and infiltrative [1])*	4	174
Dysplastic lesions *(dysplastic naevus [63], pigmented actinic keratosis [3], actinic keratosis [38], Bowen's disease [6])*	5	110
Benign naevus *(benign melanocytic naevus [61], intradermal naevus [16], compound naevus [11], blue naevus [10], Spitz naevus [8], acral naevus [2], lentiginous naevus [1], benign junctional naevus [1], Meyerson's naevus [1])*	6	111
Benign non-*naevus (solar lentigo [12], pigmented seborrhoeic keratosis [4], late stage lichenoid keratosis [3], large cell acanthoma [2] seborrhoeic keratosis [40], cellular dermatofibroma [9], haemangioma [8],keratoacanthoma [7], early stage lichenoid keratosis [5], eccrine poroma [2], hidrocystoma [2], pyogenic granuloma [2], trichoepithelioma [1], clear cell acanthoma [1], fibroepithelial polyp [1], papilloma [1], hydradenoma [1])*	7	101
Inflammatory and miscellaneous *(inflammatory changes [8], post-inflammatory pigmentation [3], scarring [3], sebaceous hyperplasia [3], lichen planus [3], trauma [1], , mucous cyst [1], fibrous papule [1], varicose eczema [1], venostasis [1], arteriovenous malformation [1], follicular keratosis [1], folliculitis [1], foreign body reaction [1])*	8	29
Healthy skin from subjects ≥ 31 y, skin types IV – VI	9	5
Healthy skin from subjects ≥ 31 y, skin types I – III	10	73
Healthy skin from subjects ≤ 30 y, skin types IV – VI	11	11
Healthy skin from subjects ≤ 30 y, skin types I – III	12	17

Table 2. Baseline characteristics of the subjects within each group (training and test samples combined; N = 741)

Group	No. of subjects	No. of males	No. of females	Age range (y)	Median age (y)
1	61	34	27	24 - 87	65
2	27	15	12	38 - 82	72
3	22	17	5	65 - 90	76
4	174	100	74	36 - 87	70
5	110	58	52	20 - 88	67
6	111	35	76	18 - 72	35
7	101	50	51	21 - 88	65
8	29	10	19	30 - 86	66
9	5	1	4	41 - 57	52
10	73	14	59	31 - 76	57
11	11	4	7	18 - 28	24
12	17	3	14	20 - 30	27

(Group 1 = invasive melanoma; 2 = melanoma in-situ; 3 = squamous cell carcinoma; 4 = basal cell carcinoma; 5 = dysplastic lesions; 6 = benign naevi; 7 = benign non-naevus lesions; 8 = inflammatory and miscellaneous lesions; 9-12 = normal healthy controls of varying skin types and ages)

Table 3. Baseline characteristics of the participants/samples selected for the 20 double-blind tests, showing age, gender, body site *(Abd = abdomen)* and max. diameter (mm). Full medical and lifestyle details are given in Supplementary File 1.

Sample type	**Test Sets**									
	1	2	3	4	5	6	7	8	9	10
Melanoma	74F, Leg, 12	42F, Arm, 12	46F, Arm, 9	50M, Leg, 20	61M, Back, 15	75M, Back, 18	77F, Arm, 10	41F, Leg, 8	66M, Abd, 20	65M, Back, 10
BCC	79F, Leg, 12	79F, Arm, 10	69F, Arm, 9	77M, Face, 20	73M, Chest, 7	84M, Face, 8	78F, Face, 11	73M, Scalp, 6	75M, Face, 9	83M, Face, 7
	73F, Face, 5	5 F, Arm, 9	67F, Face, 4	69M, Leg, 6	59M, Back, 6	76M, Face, 6	65F, Arm, 10,	63M Face, 6	60M, Face, 20	70M, Leg, 7
	65F, Leg, 12	47F, Leg, 7	67F, Chest, 7	66M, Face, 3	44M, Abd, 11	74M, Back, 18	62M, Arm, 8,	50M, Leg, 3	60M, Face, 6	69M, Face, 12
Benign Naevus	74F, Face, 5,	35F, Back, 15	48F, Abd, 6	55M, Abd, 15	63M, Arm, 8	49F, Abd, 6	78F, Arm, 4,	43F, Leg, 6	45M, Back, 6	46M, Back, 5
	56F, Arm, 4,	31F, Arm, 6	40F, Arm, 3	25M, Back, 5	54M, Back, 6	44M, Back, 6	49F, Back, 6	31F, Abd, 4	40F, Arm, 5	36F, Back, 6
	18F, Foot, 12	21F, Leg, 4	22F, Back, 4	21M, Back, 3	37M Back, 6	33F, Neck, 7	35F, Back, 7	18F, Arm, 9	32F, Abd, 7	30M, Arm, 6
Normal Skin	72F, Leg, 20	41F, Arm, 20	59F, Arm, 20	38M, Leg, 20	60F, Back, 20	70F, Back, 20	76F, Arm, 20	41F, Leg, 20	67F, Abd, 20	71F, Back, 20
	30F, Leg, 20	40F, Arm, 20	57F, Arm, 20	34M, Leg, 20	48M, Back, 20	68F, Back, 20	57F, Arm, 20	32F, Leg, 20	60F, Abd, 20	51F, Back, 20
	20F, Leg, 20	27F, Arm, 20	42F, Arm, 20	30M, Leg, 20	45F, Back, 20	20M, Back, 20	50M, Arm, 20	28F, Leg, 20	50F, Abd, 20	39F, Back, 20
Sample type	**Test Sets**									
	11	12	13	14	15	16	17	18	19	20
Melanoma	75F, Leg, 7	71M, Chest, 12	79M, Leg, 4	70F, Leg, 15	75M, Back, 10	80M, Chest, 10	83M, Abd, 18	45M, Arm, 10	51M, Back, 15	76F, Leg, 12
BCC	77F, Face, 8	73M, Arm, 7	80M, Chest, 12	85F, Face, 7	78M, Face, 10	82F, Face, 11	84F, Hand, 6	77M, Back, 10	71M, Leg, 15	78M, Face, 15
	68F, Neck, 6	63M, Abd, 12	82F, Leg, 12	77M, Back, 9	76F, Neck, 8	81M, Face, 5	85F, Face, 20	66M, Face, 10	67M, Chest, 13	77M, Chest, 10
	48F, Leg, 4	52F, Chest, 10	51M, Face, 5	68M, Scalp, 18	67F, Back, 8	79F, Arm, 20	80F, Face, 6	60M, Back, 20	55M, Face, 10	72F, Face, 6
Benign Naevus	36F, Arm, 5	74M, Back, 5	49F, Face, 5	49F, Face, 5	35F, Neck, 9	49F, Back, 4	62F, Face, 5	44M, Back, 5	56M, Back, 6	49F, Face, 5
	32F, Neck, 6	32F, Abd, 6	33F, Abd, 5	29F, Abd, 9	34F, Back, 7	34F, Face, 6	34F, Ear, 6	40M, Back, 4	32F, Back, 7	41F, Leg, 8
	20F, Leg, 4	28M, Arm, 6	29M, Arm, 6	26F, Leg, 9	27F, Chest, 7	23F, Neck, 4	22F, Neck, 7	26M, Back, 6	22F, Leg, 4	34F, Leg, 6
Normal Skin	78F, Leg, 20	56F, Chest, 20	51F, Leg, 20	71F, Leg, 20	65M, Back, 20	66M, Chest, 20	71M, Abd, 20	38M, Arm, 20	40F, Back, 20	73F, Leg, 20
	56F, Leg, 20	51F, Chest, 20	24F, Leg, 20	48F, Leg, 20	50F, Back, 20	26F, Chest, 20	68M, Abd, 20	28M, Arm, 20	25F, Back, 20	49F, Leg, 20
	55F, Leg, 20	44F, Chest, 20	24F, Leg, 20	32F, Leg, 20	24F, Back, 20	24F, Chest, 20	28F, Abd, 20	27M, Arm, 20	23M, Back, 20	30F, Leg, 20

Table 4. Summary of the baseline characteristics of the samples selected by the dog handler on the basis of the dog's alert in the 20 double-blind tests (further metadata given in Supplementary File 1).

Test Set No.	**Lesion *(sample no.)***	**Age (y)**	**Sex**	**Skin type (I-VI)**	**Site**	**Max. diam. (mm)**
1	Intradermal melanocytic naevus *(1/4)*	74	F	I	Face	5
2	Malignant melanoma *(2/0)*	42	F	II	Arm	12
3	Malignant melanoma *(3/0)*	46	F	II	Arm	9
4	Basal cell carcinoma *(4/2)*	69	M	III	Leg	6
5	Common blue naevus *(5/4)*	63	M	III	Arm	8
6	Malignant melanoma *(6/0)*	75	M	III	Back	18
7	Common blue naevus *(7/4)*	78	F	I	Arm	4
8	Basal cell carcinoma *(8/2)*	63	M	I	Face	6
9	Basal cell carcinoma *(9/1)*	75	M	III	Face	9
10	Malignant melanoma *(10/0)*	65	M	III	Back	10
11	Malignant melanoma *(11/0)*	75	F	III	Leg	7
12	Malignant melanoma *(12/0)*	71	M	II	Chest	12
13	Malignant melanoma *(13/0)*	79	M	III	Leg	4
14	Malignant melanoma *(14/0)*	70	F	III	Leg	15
15	Basal cell carcinoma *(15/1)*	78	M	III	Face	10
16	Basal cell carcinoma *(16/3)*	79	F	II	Arm	20
17	Compound melanocytic naevus *(17/6)*	22	F	III	Neck	7
18	Basal cell carcinoma *(18/2)*	66	M	II	Face	10
19	Malignant melanoma *(19/0)*	51	M	II	Back	15
20	Benign intradermal naevus *(20/6)*	34	F	II	Leg	6

Table 5. Characteristics of the malignant melanomas (all superficial spreading) used in the double-blind tests (further metadata given in Supplementary File 1).

Test Set No.	Correctly chosen by dog handler (X)	Tumour site	Breslow thickness (mm)	Clark level	Micro-stage	Max. diam. (mm)
1		Leg – left, anterior lower	0.8	4	pT1a	12
2	X	Arm – left posterior upper	1.5	4	pT2a	12
3	X	Arm – right posterior upper	0.8	4	pT1b	9
4		Leg – left anterior upper	1.8	4	pT2a	20
5		Back – right scapula	3.7	4	pT3a	15
6	X	Back – right scapula	1.4	3	pT2a	18
7		Arm – left posterior forearm	6.0	4	pT4a	10
8		Leg – right anterior upper	0.4	2	pT1a	8
9		Abdomen – left lower	0.4	3	pT1a	20
10	X	Back – left lower	1.2	4	pT2a	10
11	X	Leg – right anterior lower	0.5	2	pT1a	7
12	X	Chest – right upper	0.3	3	pT1a	12
13	X	Leg – left posterior upper	0.2	2	pT1a	4
14	X	Leg – left posterior lower	5.0	4	pT4b	15
15		Back – right upper	3.9	4	pT3b	10
16		Chest – right anterior	0.5	3	pT1b	10
17		Abdomen – left flank	1.1	3	pT2a	18
18		Arm – right anterior forearm	0.9	4	pT1b	10
19	X	Back – left upper	1.2	3	pT2a	15
20		Leg – left anterior upper	0.3	2	pT1a	12

Table 6. Covariates which remain in the discrete choice model. The outcome of interest is the sample chosen, and the covariates were all suitable characteristics of each of the samples and their donors.

Variable	Regression coefficient	p-value	Comments
Melanoma (yes/no)	**1.71**	**0.002**	**Increased probability of choosing melanoma over other sample types**
Skin type III (yes/no)	1.44	0.024	Increased probability of choosing those who had skin type III over other skin types.
Alcohol	1.40	0.025	Increased probability of choosing those who had consumed alcohol in the previous 24h
Sampling temperature	-0.73	0.057	Probability of being chosen decreases as sampling temperature increases.
Age	0.03	0.102	Probability of being chosen increases with increased age (not statistically significant)