ARTICLE   OPEN

Check for updates

Molecular Diagnostics

Investigation of urinary volatile organic compounds as novel diagnostic and surveillance biomarkers of bladder cancer

Lauren Lett [1,6], Michael George [1,2,6], Rachael Slater[1], Ben De Lacy Costello[3], Norman Ratcliffe[3], Marta García-Fiñana[4], Henry Lazarowicz[5] and Chris Probert [1 ✉]

BACKGROUND: The diagnosis and surveillance of urothelial bladder cancer (UBC) require cystoscopy. There is a need for biomarkers to reduce the frequency of cystoscopy in surveillance; urinary volatile organic compound (VOC) analysis could fulfil this role. This cross-sectional study compared the VOC profiles of patients with and without UBC, to investigate metabolomic signatures as biomarkers.
METHODS: Urine samples were collected from haematuria clinic patients undergoing diagnostic cystoscopy and UBC patients undergoing surveillance. Urinary headspace sampling utilised solid-phase microextraction and VOC analysis applied gas chromatography-mass spectrometry; the output underwent metabolomic analysis.
RESULTS: The median participant age was 70 years, 66.2% were male. Of the haematuria patients, 21 had a new UBC diagnosis, 125 had no cancer. In the surveillance group, 75 had recurrent UBC, 84 were recurrence-free. A distinctive VOC profile was observed in UBC patients compared with controls. Ten VOCs had statistically significant abundances useful to classify patients (false discovery rate range 1.9×10^{-7}–2.8×10^{-2}). Two prediction models were evaluated using internal validation. An eight-VOC diagnostic biomarker panel achieved AUROC 0.77 (sensitivity 0.71, specificity 0.72). A six-VOC surveillance biomarker panel obtained AUROC 0.80 (sensitivity 0.71 and specificity 0.80).
CONCLUSIONS: Urinary VOC analysis could aid the diagnosis and surveillance of UBC.

BACKGROUND

Urothelial bladder cancer
Patients with suspected bladder cancer require urological evaluation including diagnostic cystoscopy [1, 2]. In Western societies, three-quarters of primary bladder tumours are non-muscle-invasive urothelial bladder cancers (UBCs) [3]; these patients are managed endoscopically. A proportion of cases will be recurrent in nature, and some may progress to muscle invasive or metastatic disease. Therefore, lifelong endoscopic surveillance is mandated in those at risk of progression. Cystoscopy is invasive, labour-intensive, expensive and risk associated. The National Institute for Health and Care Excellence outlined the need for a urinary bladder cancer biomarker for use in surveillance, to reduce the frequency of cystoscopy [4].

Urinary volatile organic compounds
Several studies report the diagnostic potential of urinary volatile organic compounds (VOCs) in malignancies such as bladder [5–14], prostate [15, 16] and renal cancer [17, 18]. The study of VOCs in bladder cancer has more commonly utilised pattern recognition techniques to date [5–11]; a chemical analytical approach, such as

solid-phase microextraction (SPME) and gas chromatography-mass spectrometry (GC-MS), is required for the identification of individual VOC trends. SPME facilitates the isolation of VOCs, whilst GC-MS enables VOC analysis [19]. SPME-GC-MS is rapid, cost-effective and delivers reliable semi-quantitative data [20].

Aim
This cross-sectional study compares the urinary VOC profiles of patients with and without UBC, to establish metabolomic signatures for use as diagnostic and surveillance biomarkers.

METHODS
The study flow and data analysis pipeline have been summarised in Supplementary Fig. 2.

Participant recruitment
Patients were recruited from urology clinics at the Royal Liverpool and Broadgreen University Hospitals NHS Trust between August 1, 2016 and July 31, 2018. These patients represented two clinical opportunities: (i)

[1]Department of Molecular and Clinical Cancer Medicine, Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool L69 3GE, UK. [2]School of Medicine, Cedar House, University of Liverpool, Liverpool L69 3GE, UK. [3]Centre of Research in Biosciences, University of the West of England, Frenchay Campus, Bristol BS16 1QY, UK. [4]Department of Health Data Science, University of Liverpool, Liverpool L69 3GE, UK. [5]Department of Urology, Royal Liverpool University Hospital, Liverpool University Hospitals NHS Trust, Liverpool L7 8XP, UK. [6]These authors contributed equally: Lauren Lett, Michael George. ✉email: Chris.Probert@liverpool.ac.uk

diagnostic patients undergoing cystoscopy to investigate visible and non-visible haematuria; (ii) surveillance patients with previously diagnosed and treated UBC attending for follow-up cystoscopy as recommended by the current national guideline. A sample size calculation [21], assuming the recruitment ratio 1:2 (UBC cases: controls) and anticipated sensitivity and specificity of 0.90, determined 330 participants (110 cases and 220 controls) were required to achieve a 95% confidence interval width of 0.05 for the area under the receiver operating characteristic (AUROC). The study endpoint was the presence or absence of UBC. Participants were not followed-up beyond their initial cystoscopy appointment. Participants were over 18 years old. Exclusion criteria included: urinary tract infection, as cystoscopy was postponed; concurrent non-UBC malignancy, as these cases might introduce confounding VOCs; HIV infection, due to local health and safety guidance. Participants were assigned a unique identifier to maintain anonymity and allow blinding during data analysis.

Ethical approval was granted by the National Research Ethics Committee North West—Haydock (REC 14/NW/1212). Participants provided informed, written consent prior to enrolment.

Sample collection and storage
Urine samples were collected prior to cystoscopy in the urology clinic. Samples were stored in a refrigerator until the end of the clinic and then transported to the Liverpool Tissue Bank for long-term storage at −80 °C; freezer storage of urine within 12 h of sample collection is recommended to minimise VOC alteration secondary to metabolic and degradative processes [22]. Long-term storage of urine samples at −80 °C preserves VOC signatures and promotes matrix stability [23]. Urine samples were transferred to the analytical laboratory for storage at −20 °C in the days preceding analysis.

Chemicals, materials and instrumentation
In all, 0.3 mL 3 M sulphuric acid solution (Fisher Scientific, Loughborough, UK) was added to 0.9 mL urine in 10 mL vials (Sigma Aldrich, Dorset, UK); the acid increases the number and diversity of VOCs present in the headspace [24, 25].

A PerkinElmer Clarus 500 GC-MS single quadruple system (PerkinElmer, Beaconsfield, UK) and PAL COMBI-xt autosampler (CTC Analytics, Switzerland) were used to analyse samples. A Zebron ZB-624 GC column (0.25-mm inner diameter, 60 m length and 1.4-μm film thickness) (Phenomenex, Macclesfield, United Kingdom) was utilised. The carrier gas was 99.996% pure helium (BOC, Sheffield, UK). A 50-μm-divinylbenzene-30-μm-carboxen-polydimethylsiloxane SPME fibre (Sigma Aldrich, Dorset, UK) was used for headspace sampling; this SPME fibre maximises the number and diversity of VOCs extracted from the headspace of urinary samples, whilst minimising the prevalence of contaminant degradation products [26]. The SPME fibre underwent pre-conditioning before use; the effectiveness of this process was assessed using blank runs.

Volatile organic compound analysis
Samples were thawed at room temperature, then heated to 60 °C for 30 min. The SPME fibre was exposed to the headspace above the urine for 20 min and subsequently introduced to the GC injection port at 220 °C for 5 min, for VOC desorption. The GC oven was set to 40 °C with a 2-min hold time; the temperature was increased by 5 °C/min to 220 °C and held for 4 min. A solvent delay of 4 min was used. The MS system employed positive electron impact ionisation mode: scanning mass fragments 10–300 m/z, interscan delay 0.1 s, resolution 1000 full-width at half maximum. The helium carrier gas flowed at 1 mL/min. Samples were randomly ordered for analysis. GC-MS data were converted to NetCDF format.

Data processing
A project-specific VOC library was built using the Automated Mass Spectral Deconvolution and Identification System (AMDIS, version 2.73, 2017) and the National Institute of Standards and Technology Mass Spectral Library (NIST-MSL, version 2.0, 2011). VOCs were added to the library when the forward and reverse match values exceeded 800, or 700 with probability >80%; this was accompanied by inspection of the fragmentation pattern in AMDIS. A unique identifier ("*tentative name, probability = x%, RT = y minutes*") was recorded for VOCs that did not meet these criteria, to enable their review in subsequent samples. VOCs were recorded using their International Union of Pure and Applied Chemistry (IUPAC) synonym.

A batch report was generated which was re-processed with the NIST-MSL and NetCDF files using R studio (version 1.2.1335, 2019) with the Metab extension (version 1.24.0, 2011). Metab is a deconvolution software that aligns output data by assessing fragments and their retention times and then allowing a small time window to be applied to remove any compounds that fall outside of this, common in GC work [27].

Statistical analysis
VOCs with a prevalence of <50% in all groups were excluded, as they lack discriminatory power. Missing data values were replaced with one-fifth of the VOC's minimum level of detection. VOC data were normalised by median, log-transformed and auto-scaled for all analyses. The hypothesis of normality was rejected for all VOCs (Shapiro–Wilk test), whilst equality of variance between groups was confirmed (Levene's test). Thereafter, the non-parametric Wilcoxon rank-sum test was applied for two-group comparisons, using Metaboanalyst (version 5.0, https://www.metaboanalyst.ca). Statistical analysis was adjusted for multiple comparison using the false discovery rate (FDR) threshold of 0.05.

Biomarker analysis was performed using Metaboanalyst. A diagnostic model was constructed using the cancer versus control dataset, whilst a surveillance model utilised the recurrence versus no recurrence dataset. Demographic and clinical participant data were not included given the exploratory nature of this research. VOC selection for inclusion in the models utilised random forest ranking, Lasso analysis and K-means clustering functions in Metaboanalyst. The models were built using 70% of samples and 30% were held back for internal validation. For both models, a Random Forest classifier with 100 cross-validations was applied. We report the average AUROC, accuracy (probability of correct classification), sensitivity and specificity for the training and hold-out datasets.

RESULTS
Participant characteristics
A total of 305 samples were included (Supplementary Fig. 1), 96 UBC cases and 209 controls: non-UBC haematuria ($n = 125$), new UBC diagnosis ($n = 21$), no recurrence of UBC in surveillance ($n = 84$) and recurrence of UBC in surveillance ($n = 75$). Table 1 shows the participant characteristics. The median age was 70 years (interquartile range [IQR] 63–78). Kruskal–Wallis testing identified variation in age between groups ($P = 0.021$), although post hoc Bonferroni-adjusted pairwise comparisons did not reveal statistically significant results. There were slightly more male (66.2%) than female participants in the study: a male predominance was observed across all four groups, with no statistically significant difference in the male to female ratio between them ($P = 0.059$). Participant tobacco-smoking habits are summarised in Table 1; statistical analysis was limited for this variable due to the high non-disclosure rate (as discussed within the study limitations).

Cancer versus control
Urinary VOC profiles of all patients with UBC ($n = 96$) and those without ($n = 209$) were compared. Wilcoxon rank-sum testing found ten VOCs that differed significantly between the two groups (Table 2 and Fig. 1a). PLS-DA showed relatively poor group separation (Fig. 1b).

No recurrence versus recurrent UBC
The National Institute for Health and Care Excellence recommended the development of biomarkers for surveillance. The VOC profiles of patients undergoing surveillance were compared: recurrence ($n = 75$) and no recurrence ($n = 84$) of UBC. Wilcoxon rank-sum testing demonstrated six statistically significant VOCs (Table 3 and Fig. 2a). All of these VOCs were also statistically significant in the cancer versus control comparison. Limited group separation was visualised by PLS-DA (Fig. 2b).

Non-UBC haematuria versus new UBC diagnosis
Patients presenting with haematuria will undergo a full complement of investigations, so a diagnostic biomarker for UBC is less imperative, although it may be a useful adjunct to flexible cystoscopy. Thus, VOC profiles were compared for newly

Table 1. Participant characteristics.

	Cancer		Control	
	New UBC diagnosis	Recurrence of UBC	Non-UBC haematuria	No recurrence of UBC
Number, frequency (%)				
General	21	75	125	84
Male	14 (66.7%)	54 (72.0%)	72 (57.6%)	62 (73.8%)
Female	7 (33.3%)	21 (28.0%)	53 (42.4%)	22 (26.2%)
Age median (interquartile range), years				
General	67 (60–77.5)	73 (65–80)	69 (62–76)	71.5 (65.25–81)
Male	68.5 (61–79)	73 (65.75–79)	69 (62–75.75)	73 (64.75–81)
Female	65 (58–73)	71 (62.5–84.5)	67 (60.5–77)	70.5 (65.75–83.25)
Tobacco smoking, frequency (%)				
Current smoker	4 (19.0%)	6 (8.0%)	12 (9.6%)	11 (13.1%)
Ex-smoker	8 (38.1%)	35 (46.7%)	53 (42.4%)	42 (50.0%)
Never smoker	6 (28.6%)	15 (20.0%)	52 (41.6%)	29 (34.5%)
Unknown	3 (14.3%)	19 (25.3%)	8 (6.4%)	2 (2.4%)
Indication for investigation				
Haematuria clinic diagnostic cystoscopy (visible)	17	1	61	0
Haematuria clinic diagnostic cystoscopy (non-visible)	0	0	33	0
Haematuria clinic diagnostic cystoscopy (other or unknown)	4	0	12	0
Non-haematuria clinic diagnostic cystoscopy	0	1	17	0
Cancer surveillance programme	0	73	0	84
Unknown	0	0	2	0
Procedure performed, frequency (%)				
Flexible cystoscopy (+/− biopsy)	12 (57.2%)	18 (24.0%)	114 (91.2%)	76 (90.5%)
Cystoscopy (+/− general anaesthesia +/− biopsy)	0	1 (1.3%)	2 (1.6%)	0
TURBT	9 (42.9%)	54 (72.0%)	4 (3.2%)	6 (7.1%)
Other	0	2 (2.7%)	5 (4.0%)	2 (2.4%)
Tumour stage as per TNM, frequency (%)				
Ta	9 (42.9%)	46 (61.3%)	N/A	N/A
T1	7 (33.3%)	12 (16.0%)	N/A	N/A
T2 (non-specified)	1 (4.8%)	1 (1.3%)	N/A	N/A
T2a	3 (14.3%)	5 (6.7%)	N/A	N/A
Unknown	1 (4.8%)	11 (14.7%)	N/A	N/A
Tumour grade, frequency (%)				
G1	1 (4.8%)	3 (4.0%)	N/A	N/A
G2 not specified	0	1 (1.3%)	N/A	N/A
G2 low-grade	4 (19.0%)	25 (33.3%)	N/A	N/A
G2 high-grade	2 (9.5%)	9 (12.0%)	N/A	N/A
G3	12 (57.1%)	28 (37.3%)	N/A	N/A
G3 carcinoma in situ	1 (4.8%)	1 (1.3%)		
Unknown	1 (4.8%)	8 (10.7%)	N/A	N/A

presenting patients with haematuria, prior to diagnostic cystoscopy: UBC diagnosis ($n = 21$) and non-UBC haematuria ($n = 125$). The large discrepancy in group size, with the small number of new UBC diagnoses, was a potential source of inaccuracy and the comparison lacked statistical power. Wilcoxon rank-sum testing did not identify VOCs with statistically significant differences between the groups. Despite this, several of the tabulated VOC trends were similar to those observed in the cancer versus control comparison (Supplementary Table 1). In this study, visible and non-visible haematuria cases were not compared.

Biomarker analysis
Statistically significant VOC trends in the recurrence versus no recurrence, and cancer versus no cancer, comparisons supported the hypothesis that VOCs may be of value as surveillance and diagnostic biomarkers respectively.

Table 2. Statistically significant VOCs on Wilcoxon rank-sum testing of the cancer versus control comparison.

VOC	CAS number	Relative change in UBC (fold change)	Wilcoxon rank-sum P value	Wilcoxon rank-sum false discovery rate
Nonanal	124-19-6	Decreased (0.306)	3.78×10^{-9}	1.85×10^{-7}
2-ethylhexan-1-ol	104-76-7	Decreased (0.413)	2.42×10^{-5}	3.95×10^{-4}
1,1,4a-trimethyl-4,5,6,7-tetrahydro-3H-naphthalen-2-one	4668-61-5	Decreased (0.388)	7.81×10^{-5}	9.57×10^{-4}
5-ethyl-3-methyloxolan-2-one	2610-98-2	Increased (1.808)	6.36×10^{-8}	1.56×10^{-6}
Phenol	108-95-2	Increased (1.991)	1.94×10^{-4}	1.66×10^{-3}
4-methylpent-3-enoic acid	504-85-8	Increased (1.457)	2.03×10^{-4}	1.66×10^{-3}
2-methoxyphenol	8021-39-4	Increased (1.287)	2.60×10^{-4}	1.82×10^{-3}
3-methylheptan-2-one	2371-19-9	Increased (1.285)	4.68×10^{-3}	2.67×10^{-2}
1,2,4,5-tetramethylbenzene	95-93-2	Increased (1.119)	4.90×10^{-3}	2.67×10^{-2}
Heptan-2-one	110-43-0	Increased (1.177)	5.70×10^{-3}	2.79×10^{-2}
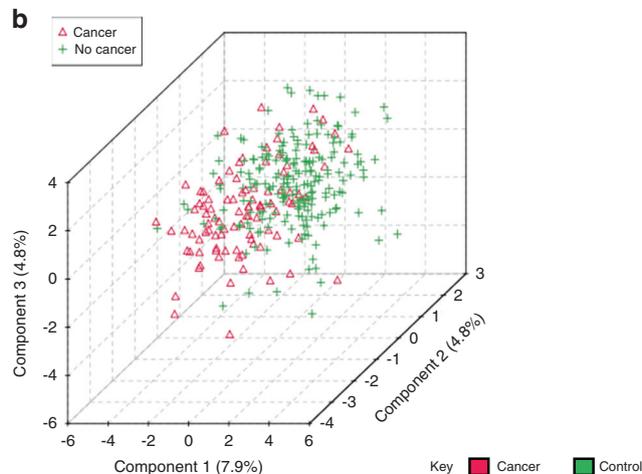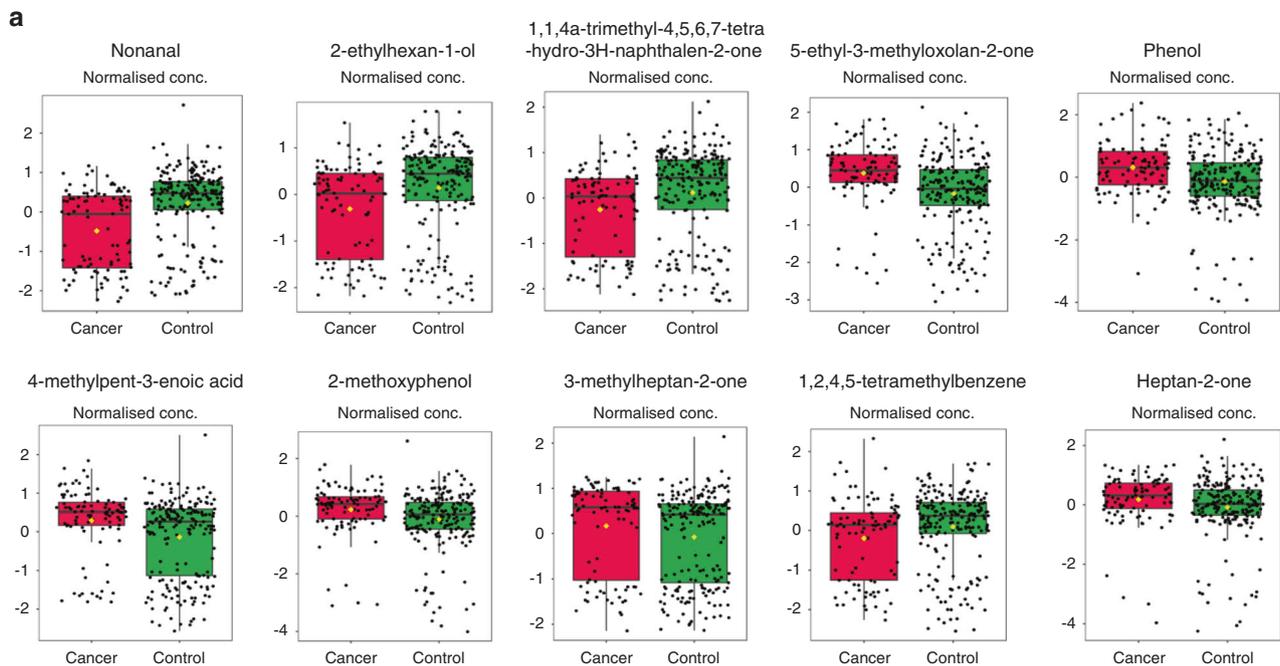
Fig. 1 Comparison of the urinary VOC profiles of all patients with and without UBC. a Box and whisker plots illustrating statistically significant VOCs in UBC on Wilcoxon rank-sum testing. **b** 3D clustering of cancer versus control groups ($n = 96$ and $n = 209$, respectively) on PLS-DA.

Table 3. Statistically significant VOCs in the recurrence of UBC in surveillance on Wilcoxon rank-sum testing.

VOC	CAS number	Relative change in recurrence of UBC (fold change)	Wilcoxon rank-sum P value	Wilcoxon rank-sum false discovery rate
Nonanal	124-19-6	Decreased (0.314)	$P = 2.18 \times 10^{-5}$	1.07×10^{-3}
2-ethylhexan-1-ol	104-76-7	Decreased (0.424)	$P = 6.50 \times 10^{-5}$	1.24×10^{-3}
1,1,4a-trimethyl-4,5,6,7-tetrahydro-3H-naphthalen-2-one	4668-61-5	Decreased (0.332)	$P = 1.88 \times 10^{-4}$	1.84×10^{-3}
5-ethyl-3-methyloxolan-2-one	2610-98-2	Increased (1.426)	$P = 7.57 \times 10^{-5}$	1.24×10^{-3}
4-methylpent-3-enoic acid	504-85-8	Increased (1.111)	$P = 1.01 \times 10^{-3}$	8.29×10^{-3}
Heptan-2-one	110-43-0	Increased (1.560)	$P = 1.44 \times 10^{-4}$	1.77×10^{-3}

All VOCs were significant in the cancer versus control comparison (bold).
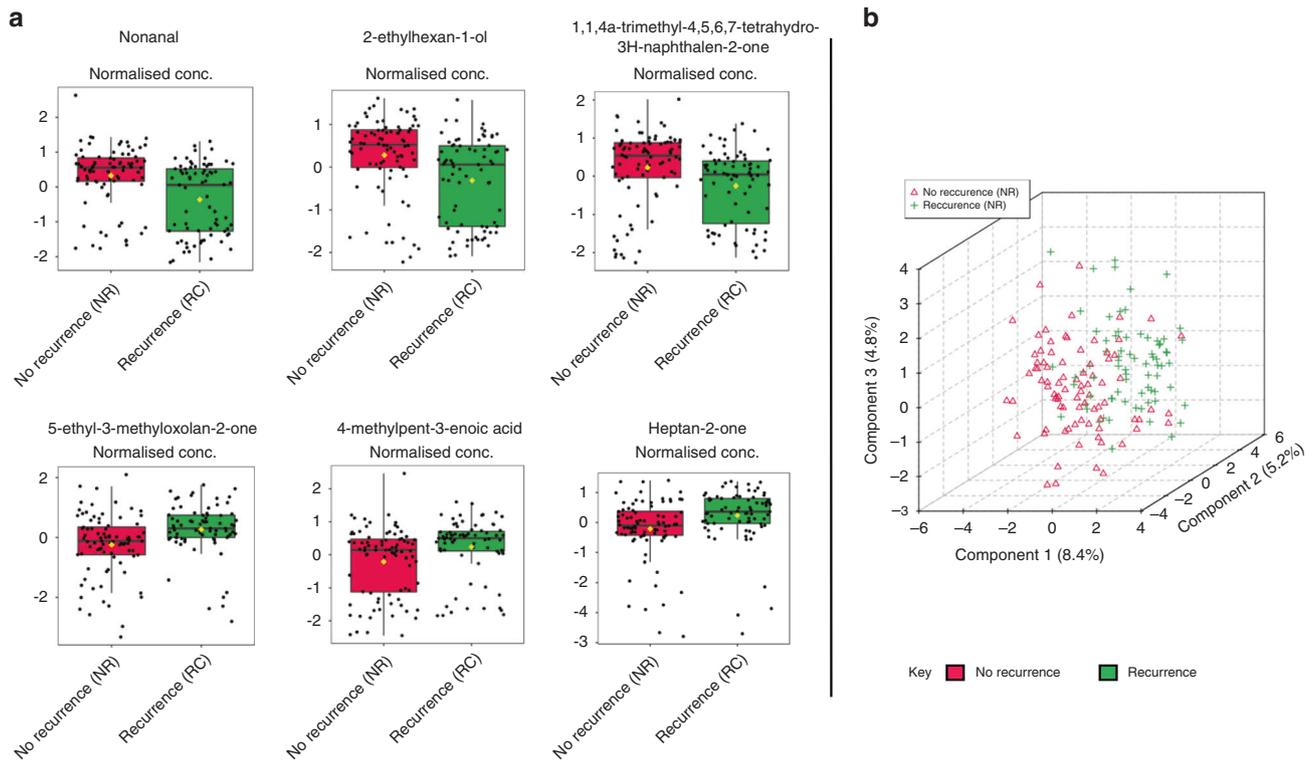
Fig. 2 Comparison of the urinary VOC profiles of surveillance patients with recurrent and non-recurrent UBC. a Box and whisker plots illustrating statistically significant VOCs in the recurrence of UBC during surveillance on Wilcoxon rank-sum testing; **b** 3D clustering of no recurrence versus recurrence of UBC in surveillance ($n = 84$ and $n = 75$, respectively) on PLS-DA.

A six-VOC surveillance biomarker model was developed using nonanal, 2-ethylhexan-1-ol, 5-ethyl-3-methyloxolan-2-one, heptan-2-one, 1,1,4a-trimethyl-4,5,6,7-tetrahydro-3H-naphthalen-2-one and propan-2-one; it had an overall AUROC of 0.80 (95% confidence interval 0.70–0.88) (Fig. 3a). Five of these VOCs were statistically significant in the recurrence versus no recurrence comparison (Table 3). The rationale for this model has been tabulated in Supplementary Table 2. The surveillance biomarker model had an average accuracy of 75% (sensitivity 0.71, specificity 0.80) when applied to the recurrence versus no recurrence dataset after 100 cross-validations. This model then underwent internal validation, achieving a hold-out AUROC of 0.86 with accuracy 78%, sensitivity 0.76 and specificity 0.80 (Fig. 3b).

An eight-VOC diagnostic model combined nonanal, phenol, 5-ethyl-3-methyloxolan-2-one, 2-ethylhexan-1-ol, 1,1,4a-trimethyl-4,5,6,7-tetrahydro-3H-naphthalen-2-one, 1-methyl-4-propan-2-ylcyclohexan-1-ol, benzaldehyde and 2,6-dimethyloct-7-en-2-ol; the overall AUROC was 0.77 (95% confidence interval 0.71–0.81)).

The rationale for VOC selection in this model is described in Supplementary Table 3. Five of these VOCs were statistically significant in the cancer versus control (Table 2). The diagnostic biomarker had an average accuracy of 72% (sensitivity 0.71, specificity 0.72) when applied to the cancer versus control dataset after 100 cross-validations. This model then underwent internal validation, obtaining a hold-out AUROC of 0.80 with accuracy 72%, sensitivity 0.56 and specificity 0.87.

DISCUSSION
Overview
This exploratory study investigated the urinary VOC profile of UBC using SPME-GC-MS. Statistically significant VOC trends provided some level of discrimination between patients with and without UBC. No significant differences were found between new and recurrent UBC cases (Supplementary Table 4); similar VOC patterns were observed between these groups, that differed from the
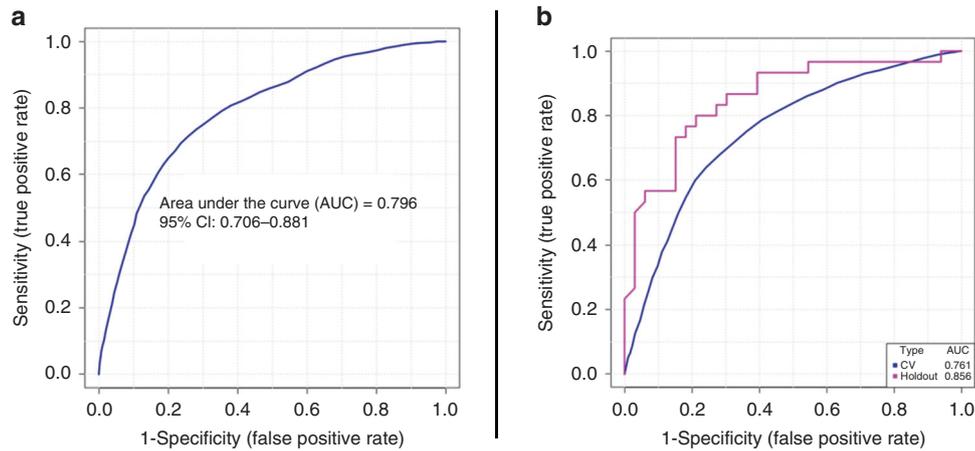
Fig. 3 Surveillance biomarker model performance. a Training evaluation using the recurrence versus no recurrence dataset; **b** internal validation using the recurrence versus no recurrence comparison with a hold-out dataset for internal validation. CI confidence interval, CV cross validation.

controls, suggesting a distinct VOC signature for UBC. VOC-based diagnostic and surveillance biomarker models have been reported and evaluated.

Urinary VOCs in UBC
Rodrigues et al. investigated the metabolomic signature of bladder cancer using the extracellular media of three distinct in-vitro cancer cell lines (Scaber, urinary bladder squamous cell carcinoma; J82 and 5637, UBC) [28]. The authors used principal component analysis to illustrate the separation of cancer and normal cell line groups [28]. Phenol and nonanal levels were different and the patterns were cell line-specific [28]; these VOCs also demonstrated the potential to differentiate UBC from urinary bladder squamous cell carcinoma [28]. Thus, the study of VOCs in the media of cell lines can be used to identify VOCs worthy of investigation in urine of patients. Furthermore, Rodrigues et al. found the concentration of nonanal was lower in high-grade versus low-grade disease [28], suggesting a grade-dependent relationship between the urinary volatilome and UBC; this is consistent with the significant decrease in nonanal in UBC we report in this study.

Several studies have evaluated urinary VOCs in patients with UBC [5–14]; the majority of published research reports the use of pattern recognition techniques, such as electronic nose technology, in urinary VOC analysis. A chemical analytical approach, such as GC-MS, allows a comprehensive description of the urinary volatilome. The sample sizes of previously published studies range from 10 to 277 participants [5–14]; we report data from 305 patients making this the largest study of its kind. Furthermore, this study demonstrates novelty in the metabolomic evaluation of both newly diagnosed and follow-up UBC patients. The biomarker models reported in this study achieved comparable performance to those in existing literature, further supporting the role of urinary VOC analysis.

The comparison of our findings with the existing literature is challenging due to differences in the treatment of urine samples and the analytical conditions implemented. An optimised and universal approach to urinary VOC analysis is desirable to further the evidence base and strengthen future conclusions. Tyagi et al. applied gas chromatography time-of-flight mass spectrometry to evaluate the VOC profile of untreated urinary samples [12]. The authors identified the statistical significance of nonanal in bladder cancer, although the directional trend in its abundance is not reported [12]. Pinto et al. investigated the metabolomic signature of untreated urinary samples provided by patients with bladder cancer ($n = 53$) and controls ($n = 56$) using SPME-GC-MS [13] and

reported increased levels of alkanes and aromatic compounds, as well as the decreased abundance of aldehydes, ketones and monoterpenes [13]; there were no common VOCs between that research and this study. Cauchi et al. applied GC-MS to establish the UBC-VOC profile of urine samples following treatment with hydrochloric acid [14]; the significant VOCs reported in that study lacked likeness with our findings and demonstrated little similarity with the previously discussed research.

The treatment of urine samples using sulphuric acid has been shown to optimise VOC analysis by enhancing the number and diversity of potential biomarkers detectable in the headspace [24, 25]. This technique remained constant across all analyses performed within this study and facilitated the identification of distinctive VOC trends in UBC. However, the acidification process makes it difficult to establish the biological origins of urinary VOCs with confidence and some compounds might be generated in-vitro. This could be addressed by the study of VOCs in untreated urine. A comprehensive review published by Janfaza et al. provides an insight into the chemical classes and metabolic origins of cancer-associated VOCs across several biological matrices [29].

Clinical applications of urinary VOCs in UBC
The National Institute for Health and Care Excellence recommended the development of a urinary biomarker to reduce the frequency of endoscopic UBC surveillance from annually to biennially [4]. Our surveillance biomarker model had an AUROC 0.80 (sensitivity 0.71, specificity 0.80). Therefore, 20% of negative test results would be a false negative, with a similar proportion of false positives. This model would reduce the demand for cystoscopy, at the expense of delaying the diagnosis of recurrent UBC by up to 12 months in the false negatives and unnecessarily exposing patients to the risks of cystoscopy in the false positives. There is an inherent trade-off between sensitivity and specificity: a longitudinal study of urinary VOCs and cystoscopy is required to determine the best cut-off point for clinical use. Although the sensitivity and specificity of this model are insufficient for use as a standalone investigation, these results show promise. Urinary VOC analysis may find utility in combination with more established surveillance tests.

In a research laboratory, GC-MS is labour-intensive, however, we used a Combipal system to control temperature, fibre exposure, sample extraction and injection; this reduces manual aspects of the analysis and may increase its reproducibility and precision. Alternatively, full-automated GC-sensor systems (based on the device reported by Aggio et al. [10]) should enable the workflow

to be streamlined. However, such development requires significant investment and proof of concept studies, such as that reported here, are necessary to secure funding.

The diagnosis of UBC is currently reliant on the symptomatic presentation of patients. Visible haematuria is the greatest symptomatic predictor of UBC: positive predictive value 3.9% and 6.1% on first and second presentation, respectively, in patients aged over 65 years [30]. However, this is not pathognomonic of bladder cancer and there are other clinically significant causes of haematuria, so the investigation is mandated.

A meta-analysis by Chou et al. summarised the performance of established urinary biomarkers in the diagnosis and surveillance of UBC [31]. The overall sensitivities and specificities varied between assays [31]. Improved performance is observed on combining urinary biomarkers [32, 33].

Urinary VOCs in other urological malignancies
The investigation of urinary metabolites as potential biomarkers of urinary tract disease is logical. The urinary VOC signatures of renal cancer and prostate cancer have undergone evaluation, demonstrating distinctive trends that may offer clinical utility.

Urinary VOCs have been reported in renal cell carcinoma patients [18]. This cancer shares presenting features with UBC, particularly haematuria. Wang et al. reported the significantly increased abundance of nonanal in renal cell cancer [18]; we found it in significantly decreased abundance in UBC. These contrasting results suggest nonanal may be capable of differentiating these two urological malignancies.

The urinary VOC profile of prostate cancer has also been explored. Tyagi et al. reported the statistical significance of 2-ethylhexan-1-ol and phenol amongst cases [12]. Lima et al. identified several significant prostate cancer-associated VOCs and proposed a panel of six compounds that achieved discrimination of cases from controls with an accuracy of 86% [34]. Khalid et al. reported a different set of discriminating VOCs, with a predictive model achieving an accuracy of 63–65% [9]. The prostate cancer-associated VOCs discussed in the latter two studies demonstrated no cross-over with the UBC-associated VOCs we report in this study.

Limitations
This study aimed to identify the urinary VOC profile of UBC as diagnostic and surveillance biomarkers. The findings reported meet these predetermined aims but require external validation with larger and independent datasets.

Having benchmarked the metabolomic signature of UBC, future research should investigate its association with disease stage, grade and prognosis. Statistical analysis demonstrated no significant differences between the VOCs of patients with localised-indolent and advanced-aggressive disease in this study. These analyses were limited by the small advanced disease sample size.

We optimised our protocol to maximise the number and range of VOCs from control samples. It is possible that an alternative SPME may be better suited to specific VOCs relevant to the diagnosis of UBC.

Construction of the diagnostic model using the new UBC versus non-UBC haematuria comparison was not feasible given the small number of cancer cases; consequently, this model utilised the cancers versus control dataset in order to appraise a diagnostic biomarker capable of detecting UBC regardless of its clinical presentation.

Following exclusion of samples provided by patients with unknown smoking status, statistical analysis exhibited one significant VOC in the smoker ($n = 33$) versus non-smoker ($n = 240$) comparison: 1,1,5-trimethyl-2H-naphthalene was present in significantly decreased abundance in smokers (fold change 0.434, $P = 8.51 \times 10^{-4}$, false discovery rate 0.043). This compound conferred no significance in any UBC group comparisons. This comparison may have been limited by the proportionately small number of cigarette smokers; further research into the VOC profile of tobacco smoking is desirable.

Statistical analysis demonstrated that the male predominance observed within the study cohort did not differ significantly between the participant groups ($P = 0.059$). As a result, the proposed biomarker models were not adjusted for patient gender; future research may wish to evaluate the relationship between gender and VOC detection, with the view to further optimising predictive models.

Data on participant medication use and past medical history were incomplete. However, this pragmatic study recruited a large patient cohort that would accurately represent the real-world demographic variability in patients that would engage with a non-invasive biomarker test such as urinary VOC analysis. Participant exclusion based upon such factors was therefore avoided. The future interpretation of urinary VOCs in patients with co-morbidities and concurrent medication use would benefit from further research: an investigation of the relationship between medications and urinary VOCs, and a prospective study of urinary VOCs in dynamic disease states such as progressing or resolving pathophysiology.

Urine dilution provides an area of uncertainty in this research. On completion of this work, very dilute samples were excluded, and the statistical analysis repeated; their removal did not change the interpretation of the results. In future studies, it may be beneficial to place patients on a restricted diet to reduce 'noise' from food, and to obtain a first-pass urine collection in order to obtain more concentrated samples: however, the study was pragmatic and reflects the real world.

CONCLUSION
The metabolomic signature of UBC is distinctive. Significant trends in urinary VOCs enable the identification of some but not all UBC samples from clinically relevant controls. The biomarker models proposed within this study show promise. Urinary VOC analysis may have the potential to influence future clinical practice in assisting the diagnosis and surveillance of UBC.

REFERENCES
1. Gorin MA, Ayyathurai R, Soloway MS. Diagnosis and treatment of bladder cancer: how can we improve? Postgrad Med. 2012;124:28–36.
2. Ouzaid I, Panthier F, Hermieu JF, Xylinas E. Contemporary surgical and technical aspects of transurethral resection of bladder tumor. Transl Androl Urol. 2019;8:21–4.
3. Boustead GB, Fowler S, Swamy R, Kockelbergh R, Hounsome L. Stage, grade and pathological characteristics of bladder cancer in the UK: British Association of Urological Surgeons (BAUS) urological tumour registry. BJU Int. 2014;113:924–30.
4. NICE. Bladder cancer: research recommendations. National Institute for Health and Care Excellence; 2015 [Available from: https://www.nice.org.uk/guidance/ng2/chapter/2-research-recommendations]. Accessed 09/03/2021.
5. Heers H, Gut JM, Hegele A, Hofmann R, Boeselt T, Hattesohl A, et al. Non-invasive detection of bladder tumors through volatile organic compounds: a pilot study with an electronic nose. Anticancer Res. 2018;38:833–7.
6. Zhu S, Corsetti S, Wang Q, Li C, Huang Z, Nabi G. Optical sensory arrays for the detection of urinary bladder cancer-related volatile organic compounds. J Biophotonics. 2019;12:e201800165.
7. 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. 2010;8:145–53.
8. Weber CM, Cauchi M, Patel M, Bessant C, Turner C, Britton LE, et al. Evaluation of a gas sensor array and pattern recognition for the identification of bladder cancer from urine headspace. Analyst. 2011;136:359–64.
9. Khalid T, White P, De Lacy Costello B, Persad R, Ewen R, Johnson E, et al. A pilot study combining a GC-sensor device with a statistical model for the identification of bladder cancer from urine headspace. PLoS ONE. 2013;8:e69602.

10. Aggio RB, de Lacy Costello B, White P, Khalid T, Ratcliffe NM, Persad R, et al. The use of a gas chromatography-sensor system combined with advanced statistical methods, towards the diagnosis of urological malignancies. J Breath Res. 2016;10:017106.

11. Bannaga AS, Kvasnik F, Persaud K, Arasaradnam RP. Differentiating cancer types using a urine test for volatile organic compounds. J Breath Res. 2020;15:017102.

12. Tyagi H, Daulton E, Bannaga AS, Arasaradnam RP, Covington JA. Urinary volatiles and chemical characterisation for the non-invasive detection of prostate and bladder cancers. Biosensors. 2021;11:1.

13. Pinto J, Carapito Â, Amaro F, Lima AR, Carvalho-Maia C, Martins MC, et al. Discovery of volatile biomarkers for bladder cancer detection and staging through urine metabolomics. Metabolites. 2021;11:4.

14. Cauchi M, Weber CM, Bolt BJ, Spratt PB, Bessant C, Turner DC, et al. Evaluation of gas chromatography mass spectrometry and pattern recognition for the identification of bladder cancer from urine headspace. Anal Methods. 2016;8:4037–46.

15. Gao Q, Su X, Annabi MH, Schreiter BR, Prince T, Ackerman A, et al. Application of urinary volatile organic compounds (VOCs) for the diagnosis of prostate cancer. Clin Genitourin Cancer. 2019;17:183–90.

16. Khalid T, Aggio R, White P, De Lacy Costello B, Persad R, Al-Kateb H, et al. Urinary volatile organic compounds for the detection of prostate cancer. PLoS ONE. 2015;10:e0143283.

17. Monteiro M, Moreira N, Pinto J, Pires-Luís AS, Henrique R, Jerónimo C, et al. GC-MS metabolomics-based approach for the identification of a potential VOC-biomarker panel in the urine of renal cell carcinoma patients. J Cell Mol Med. 2017;21:2092–105.

18. Wang D, Wang C, Pi X, Guo L, Wang Y, Li M, et al. Urinary volatile organic compounds as potential biomarkers for renal cell carcinoma. Biomed Rep. 2016;5:68–72.

19. Smith S, Burden H, Persad R, Whittington K, de Lacy Costello B, Ratcliffe NM, et al. A comparative study of the analysis of human urine headspace using gas chromatography-mass spectrometry. J Breath Res. 2008;2:037022.

20. Zhang S, Raftery D. Headspace SPME-GC-MS metabolomics analysis of urinary volatile organic compounds (VOCs). Methods Mol Biol. 2014;1198:265–72.

21. Machin D, Capmbell MJ, Tan SB, Tan SH. Reference intervals and receiver operating curves. In: Machin D, Capmbell MJ, Tan SB, Tan SH. (eds). Sample size tables for clinical studies. Third Edition edn. (Wiley-Blackwell: Chichester, West Sussex, United Kingdom, 2009) pp 158–178.

22. McFarlanE M, MozdiaK E, Daulton E, Arasaradnam R, Covington J, Nwokolo C. Pre-analytical and analytical variables that influence urinary volatile organic compound measurements. PLoS ONE. 2020;15:e0236591-e.

23. Esfahani S, Sagar NM, Kyrou I, Mozdiak E, O'Connell N, Nwokolo C, et al. Variation in gas and volatile compound emissions from human urine as it ages, measured by an electronic nose. Biosens. 2016;6:1.

24. Aggarwal P, Baker J, Boyd MT, Coyle S, Probert C, Chapman EA. Optimisation of urine sample preparation for headspace-solid phase microextraction gas chromatography-mass spectrometry: altering sample pH, sulphuric acid concentration and phase ratio. Metabolites 2020;10:12.

25. Drabińska N, Młynarz P, de Lacy Costello B, Jones P, Mielko K, Mielnik J, et al. An optimization of liquid-liquid extraction of urinary volatile and semi-volatile compounds and its application for gas chromatography-mass spectrometry and proton nuclear magnetic resonance spectroscopy. Molecules. 2020;25:3651.

26. Natalia D, Starowicz M, Krupa-Kozak U. Headspace solid-phase microextraction coupled with gas chromatography–mass spectrometry for the determination of volatile organic compounds in urine. J Anal Chem. 2020;75:792–801.

27. Aggio R, Villas-Bôas SG, Ruggiero K. Metab: an R package for high-throughput analysis of metabolomics data generated by GC-MS. Bioinformatics. 2011;27:2316–8.

28. Rodrigues D, Pinto J, Araújo AM, Monteiro-Reis S, Jerónimo C, Henrique R, et al. Volatile metabolomic signature of bladder cancer cell lines based on gas chromatography–mass spectrometry. Metabolomics 2018;14:62.

29. Janfaza S, Khorsand B, Nikkhah M, Zahiri J. Digging deeper into volatile organic compounds associated with cancer. Biol Methods Protoc. 2019;4:bpz014.

30. Shephard EA, Stapley S, Neal RD, Rose P, Walter FM, Hamilton WT. Clinical features of bladder cancer in primary care. Br J Gen Pr. 2012;62:e598–604.

31. Chou R, Gore JL, Buckley D, Fu R, Gustafson K, Griffin JC, et al. Urinary biomarkers for diagnosis of bladder cancer: a systematic review and meta-analysis. Ann Intern Med. 2015;163:922–31.

32. Todenhöfer T, Hennenlotter J, Esser M, Mohrhardt S, Aufderklamm S, Böttge J, et al. Stepwise application of urine markers to detect tumor recurrence in patients undergoing surveillance for non-muscle-invasive bladder cancer. Dis Markers. 2014;2014:973406.

33. Horstmann M, Patschan O, Hennenlotter J, Senger E, Feil G, Stenzl A. Combinations of urine-based tumour markers in bladder cancer surveillance. Scand J Urol Nephrol. 2009;43:461–6.

34. Lima AR, Pinto J, Azevedo AI, Barros-Silva D, Jerónimo C, Henrique R, et al. Identification of a biomarker panel for improvement of prostate cancer diagnosis by volatile metabolic profiling of urine. Br J Cancer. 2019;121:857–68.

COMPETING INTERESTS
The authors declare no competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE
Ethical approval was granted by the National Research Ethics Committee North West–Haydock (REC 14/NW/1212). Participants provided informed, written consent prior to enrolment. This study was performed in accordance with the Declaration of Helsinki.

CONSENT TO PUBLISH
Not applicable.

ADDITIONAL INFORMATION
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41416-022-01785-8.

Correspondence and requests for materials should be addressed to Chris Probert.

Reprints and permission information is available at http://www.nature.com/reprints

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