Electronic Supplementary Material (ESI) for Molecular Omics.

Electronic Supplementary Material for

Cancer evaluation in dogs using cerumen as a source for volatile biomarkers prospection

João Marcos G. Barbosa*, Engy Shokry, Lurian C. David, Naiara Z. Pereira, Adriana R. da Silva, Vilma F.de Oliveira, Maria Clorinda S. Fioravanti, Paulo H. Jorge da Cunha, Anselmo E. de Oliveira, Nelson R. Antoniosi Filho*

*Corresponding author: joaomarcosquim.ufg@outlook.com (J.M.G.B); nelsonroberto@ufg.br (N.R.A.F.)

This file includes:

Figures S1-S17

Tables S1-S4

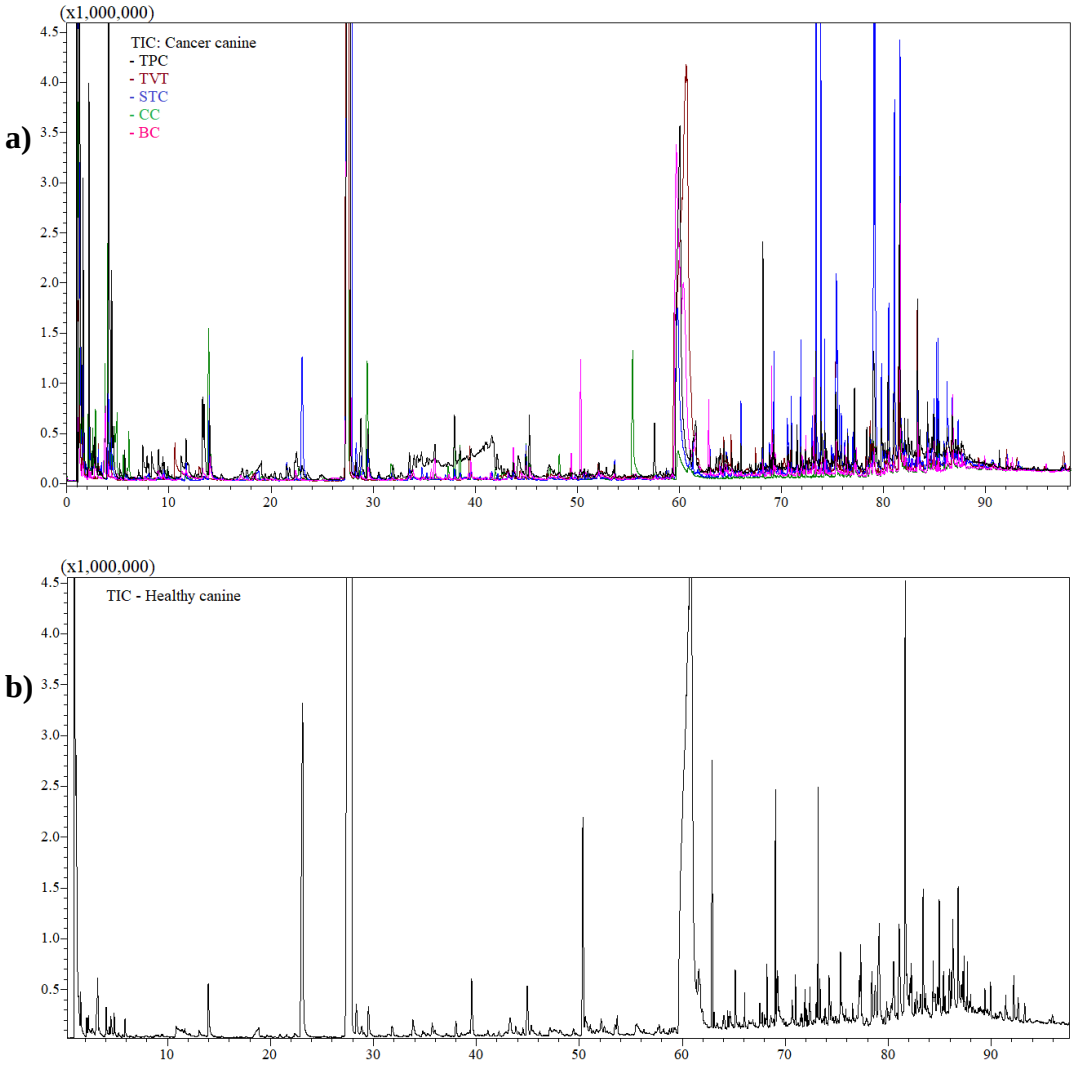

Figure S1. Total ion chromatograms (TICs) fingerprint of cerumen samples analysis from some examples of cancer dogs (a) and one example of healthy canines (b). TPC= Testicular & Prostate Cancer, TVT=Transmissible Venereal Tumor, STC=Soft Tissue Cancer, CC= Colorectal Cancer, BC=breast cancer.

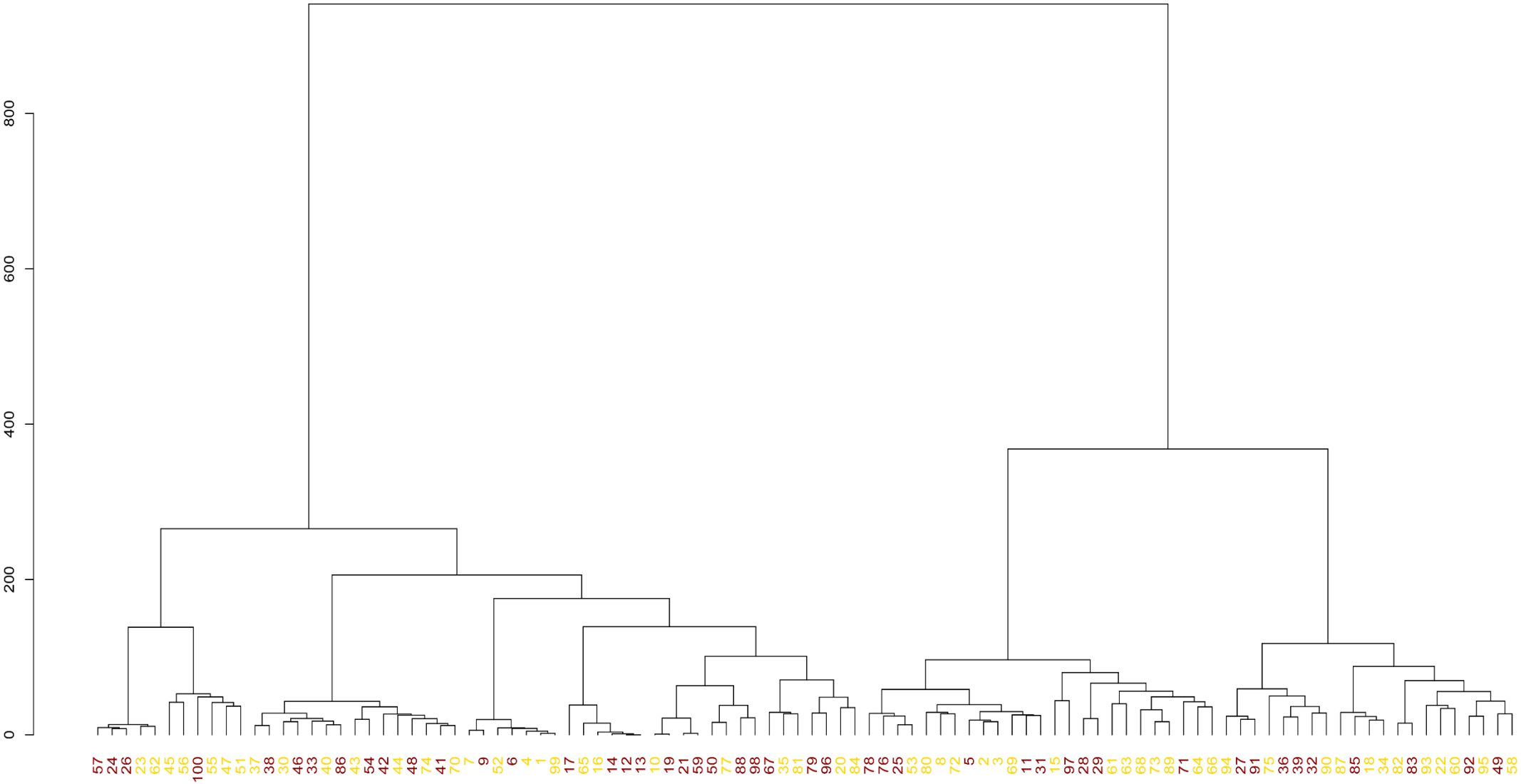

Figure S2. Hierarchical Cluster Analysis (HCA) using all the 128 Volatile Organic Metabolites (VOMs) annotated to evaluate the sex influence on sample discrimination. Cerumen samples: Gold = Female; Dark red = Male. Samples are enumerated by HS/GC-MS order of analysis.

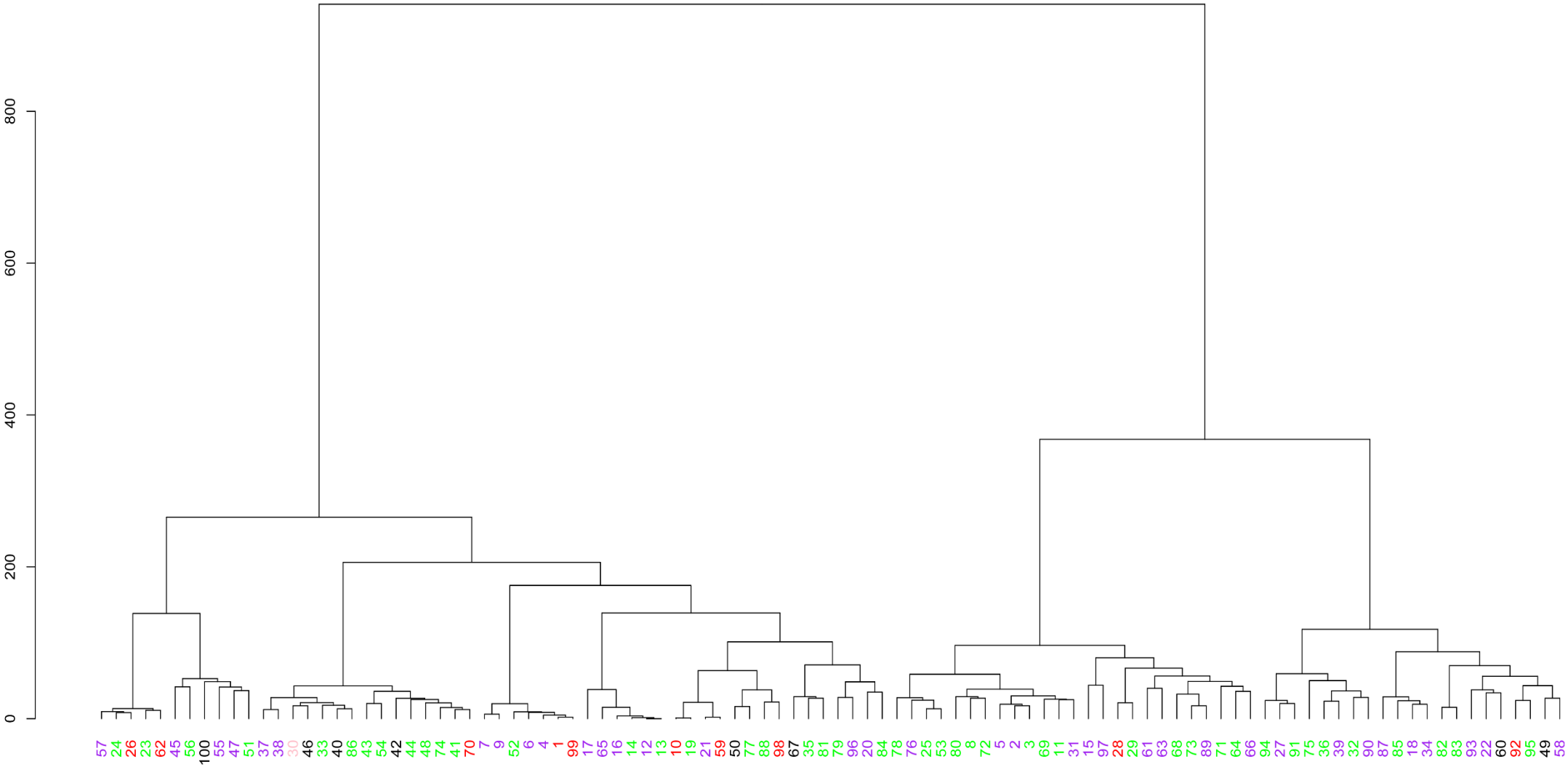

Figure S3. Hierarchical Cluster Analysis (HCA) using all the 128 Volatile Organic Metabolites (VOMs) annotated to evaluate the age bias on sample discrimination. Cerumen samples: Red = Adolescent (6–18 months) ; Purple = Adults (2–7 years); Green = Seniors (8–14 years); Pink = Geriatrics (>14 years); Black = Age Not available. Samples are enumerated by HS/GC-MS order of analysis.

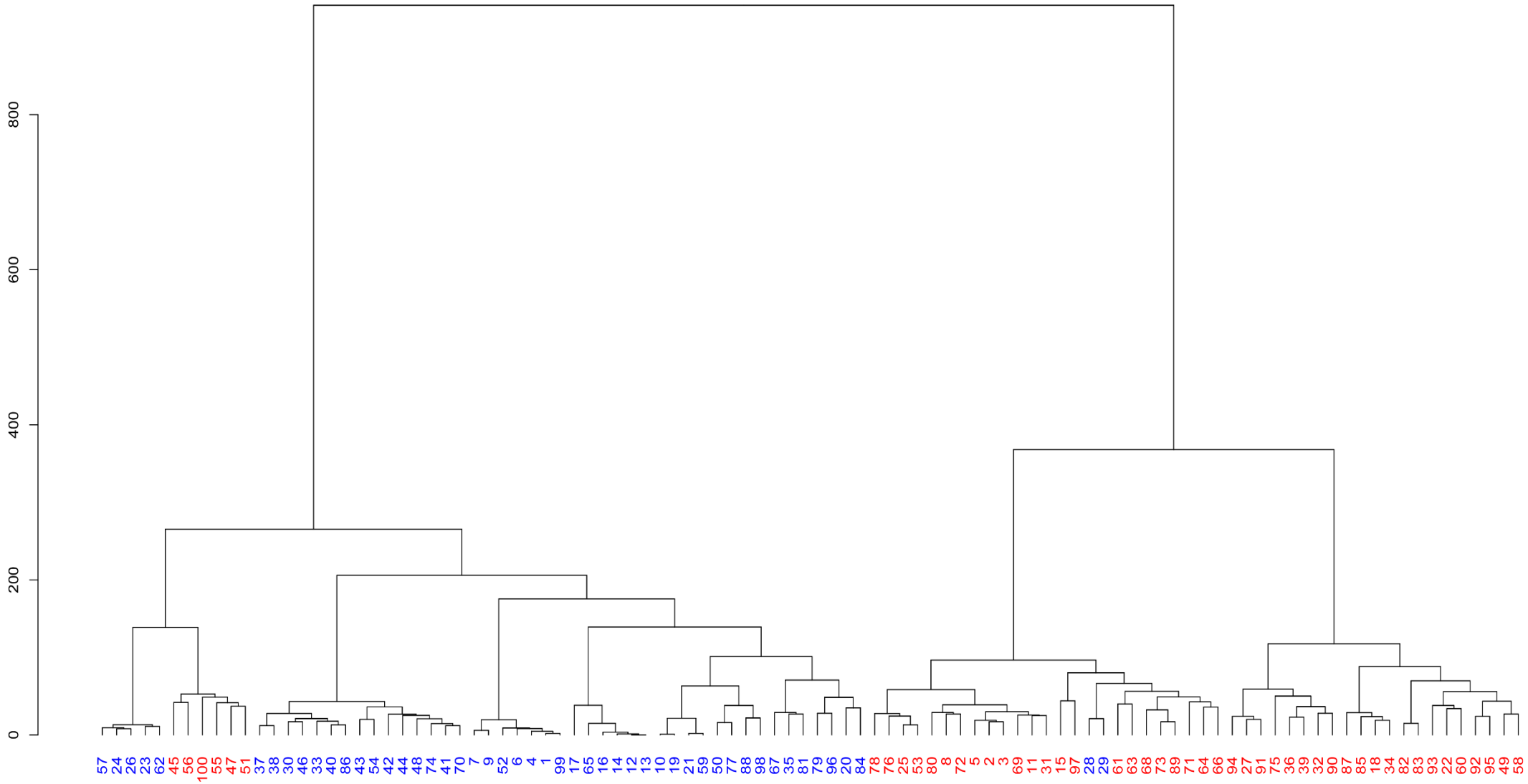

Figure S4. Hierarchical Cluster Analysis (HCA) using all the 128 Volatile Organic Metabolites (VOMs) annotated to evaluate the cancer influence on sample discrimination. Cerumen samples: Red = Cancer dogs, Blue = Healthy dogs. Samples are enumerated by HS/GC-MS order of analysis.

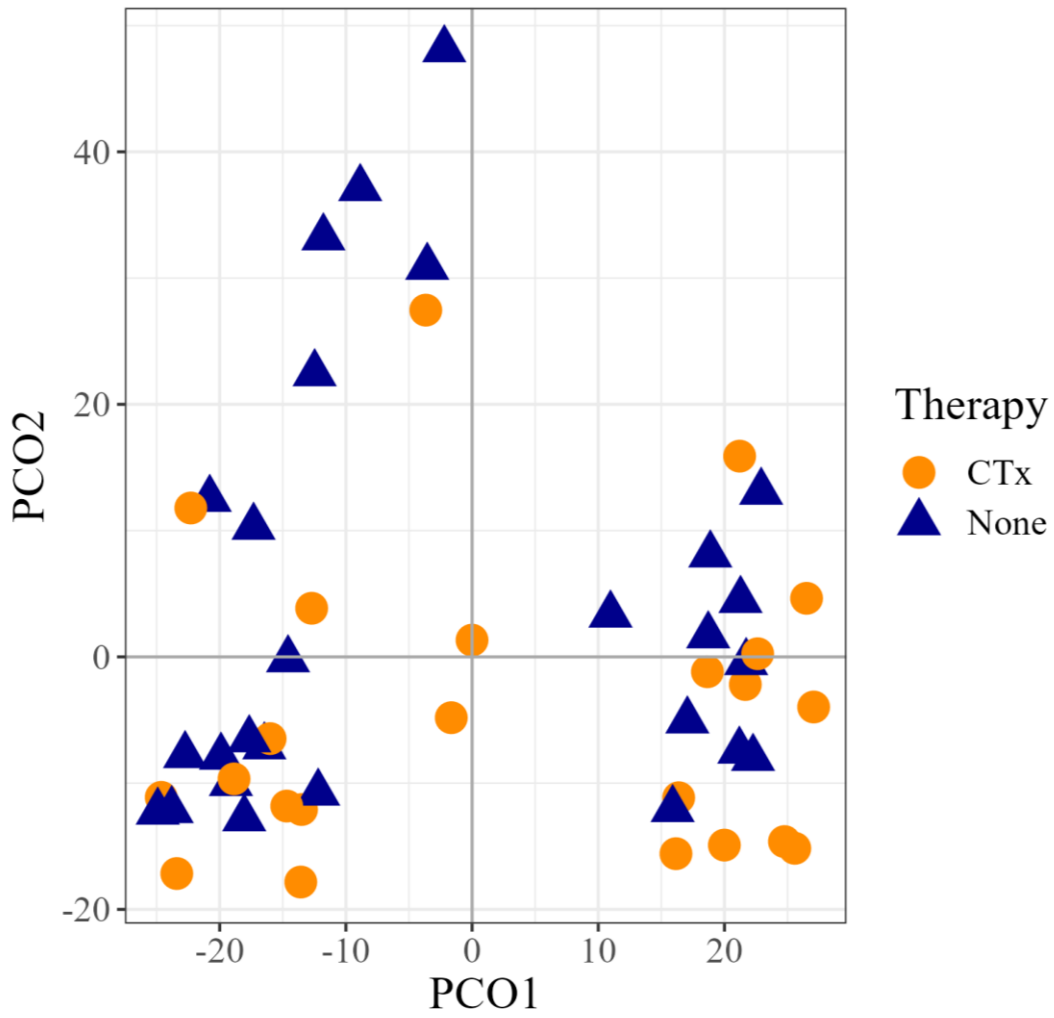

Figure S5. Principal coordinate analysis (PCoA) visualization of all the 128 Volatile Organic Metabolites (VOMs) annotated to evaluate the chemotherapy treatment influence in cancer sample discrimination. Cerumen samples: Orange dots = Cancer dogs under chemotherapy therapy (CTx), Dark blue triangles = Cancer dogs without chemotherapy treatment (None).

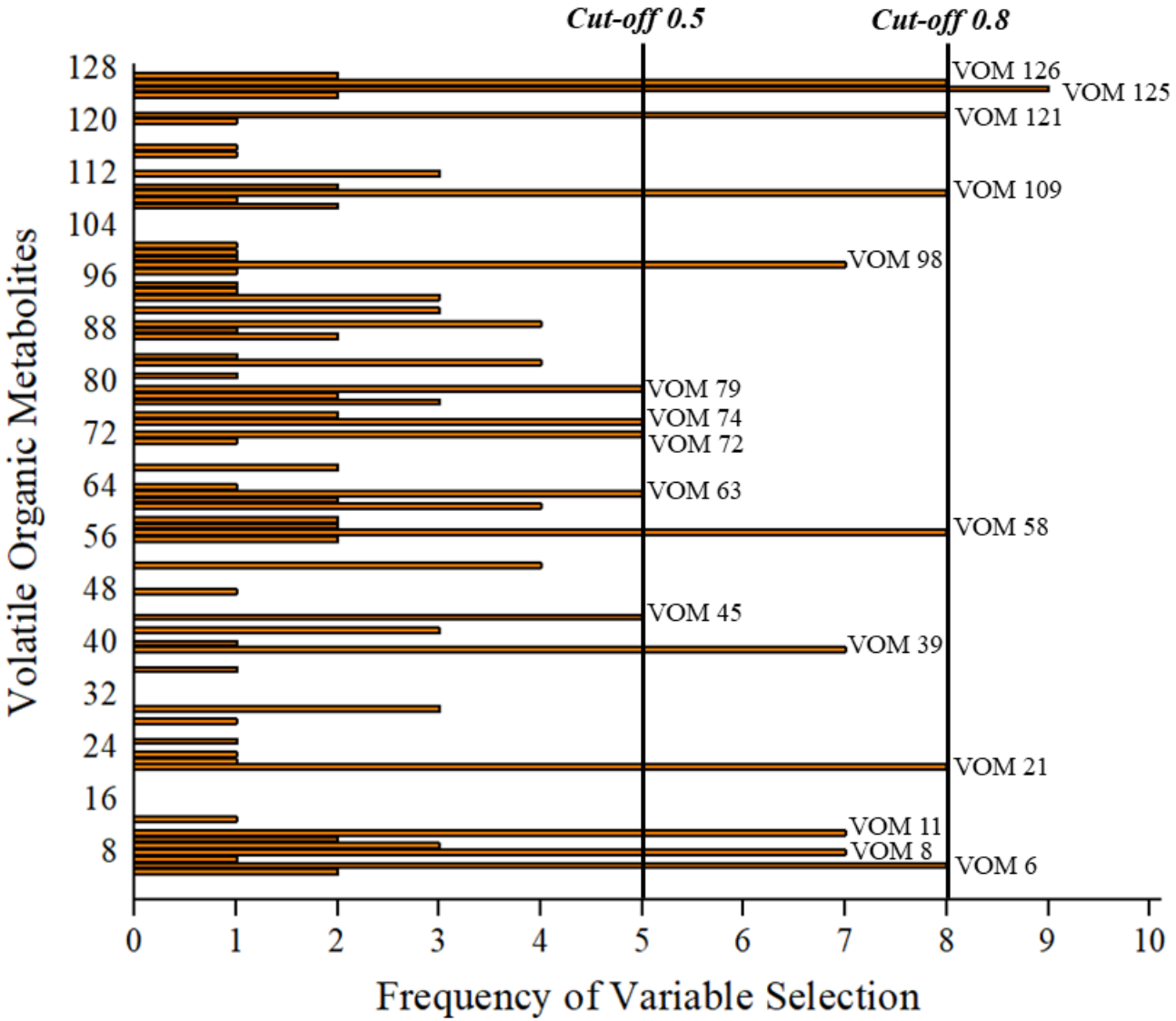

Figure S6. Variables/VOMs selected by GA-PLS to build the candidate panels of biomarkers. Chromosomes XI and XII hold the variables chosen at least 5 out of 10 (*cut-off* 0.5) and 8 out of 10 (*cut-off* 0.8) times during the ten random GA models.

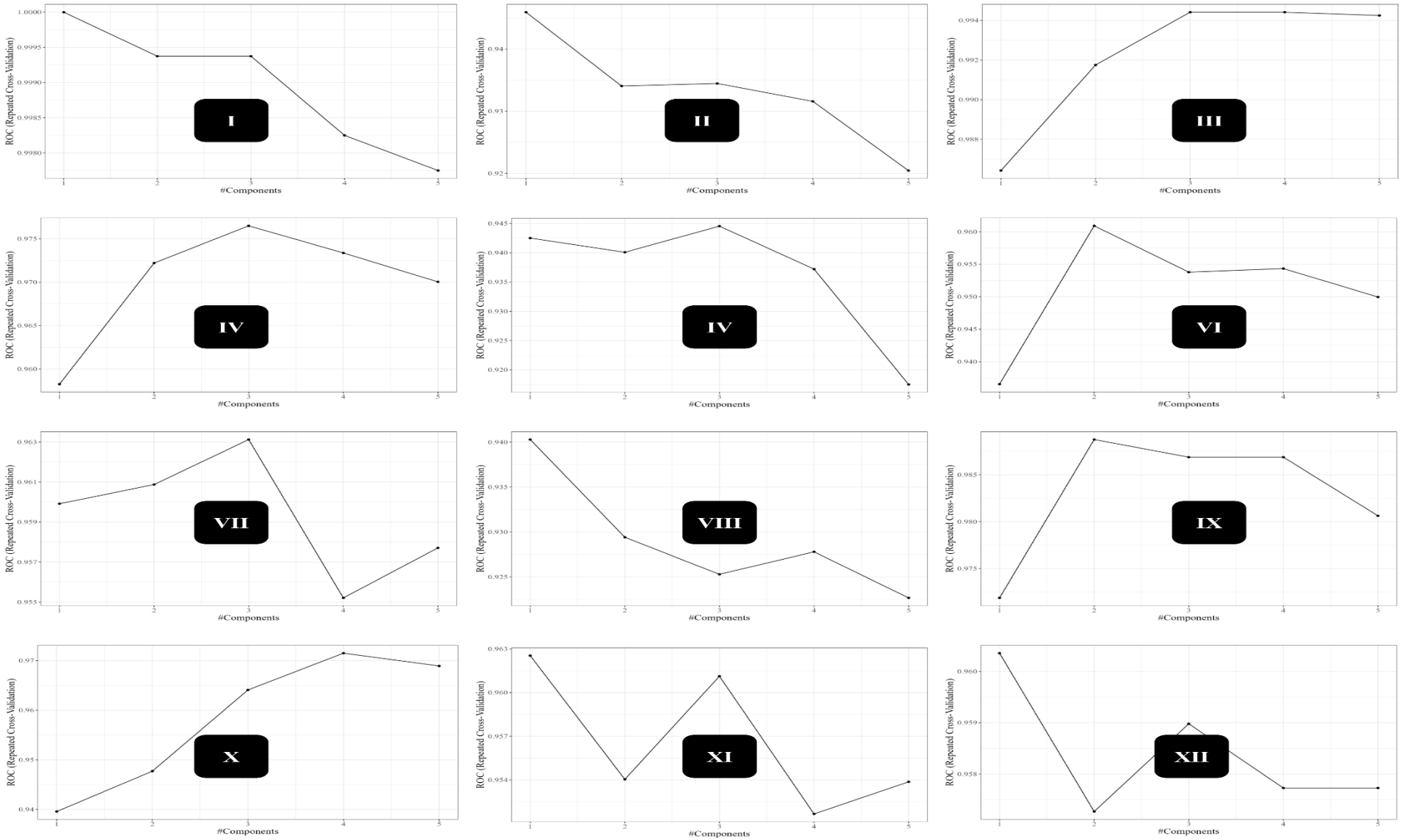

Figure S7. ROC results for the repeated k-fold cross-validation (k=10, repeats=10) for the cerumenogram model in dogs. The numbers I to XII represent the GA-PLS chromosomes containing the candidate panel biomarkers, as described in Table S3. Chromosome 1 represents the best ROC value (1.000) using 18 VOMs and 1 PLS component.

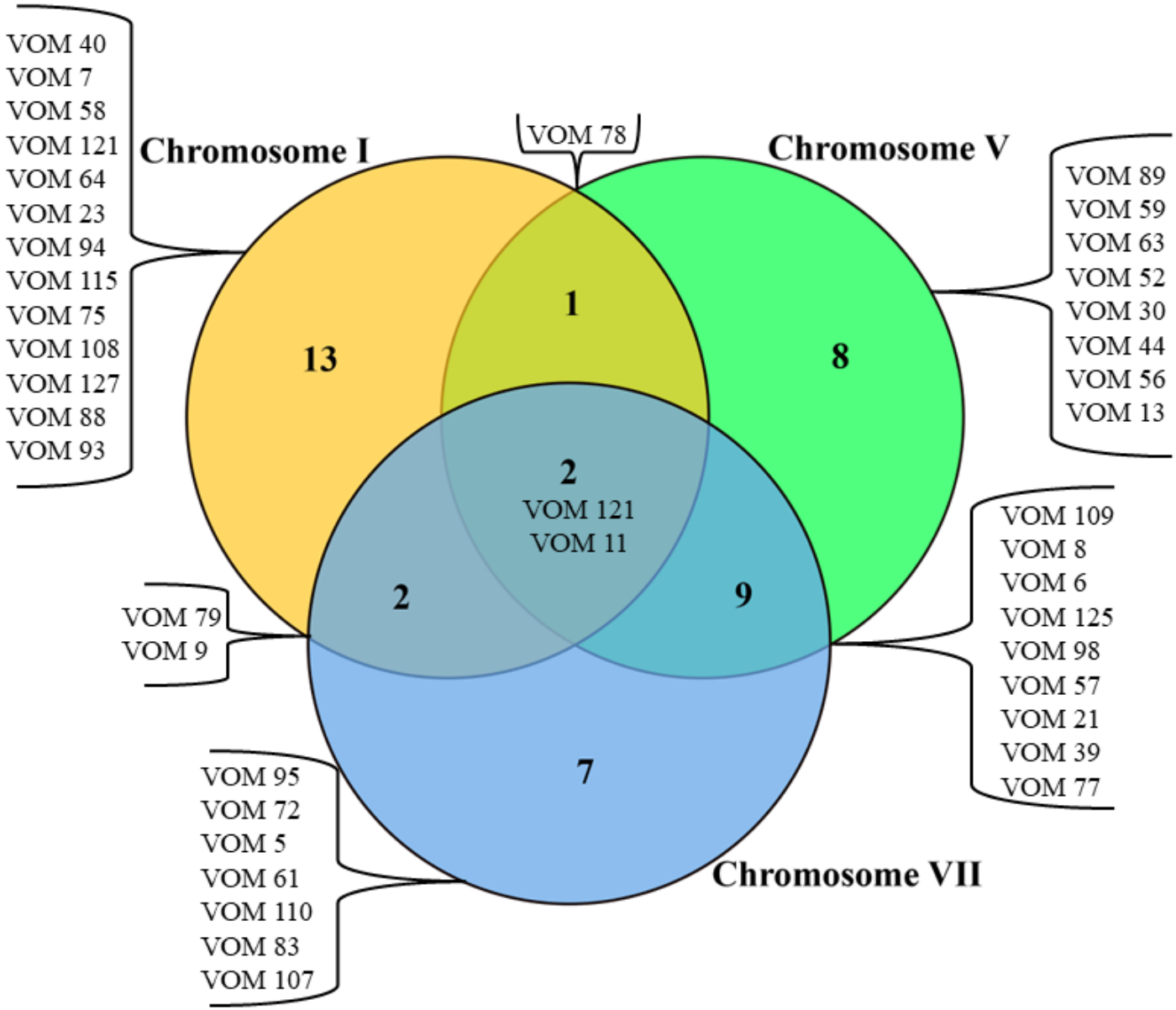

Figure S8. Venn diagram of the three best GA-PLS chromosomes (I, V, and VII Table S3) showing the shared VOMs enumerated by their retention time (Table S2).

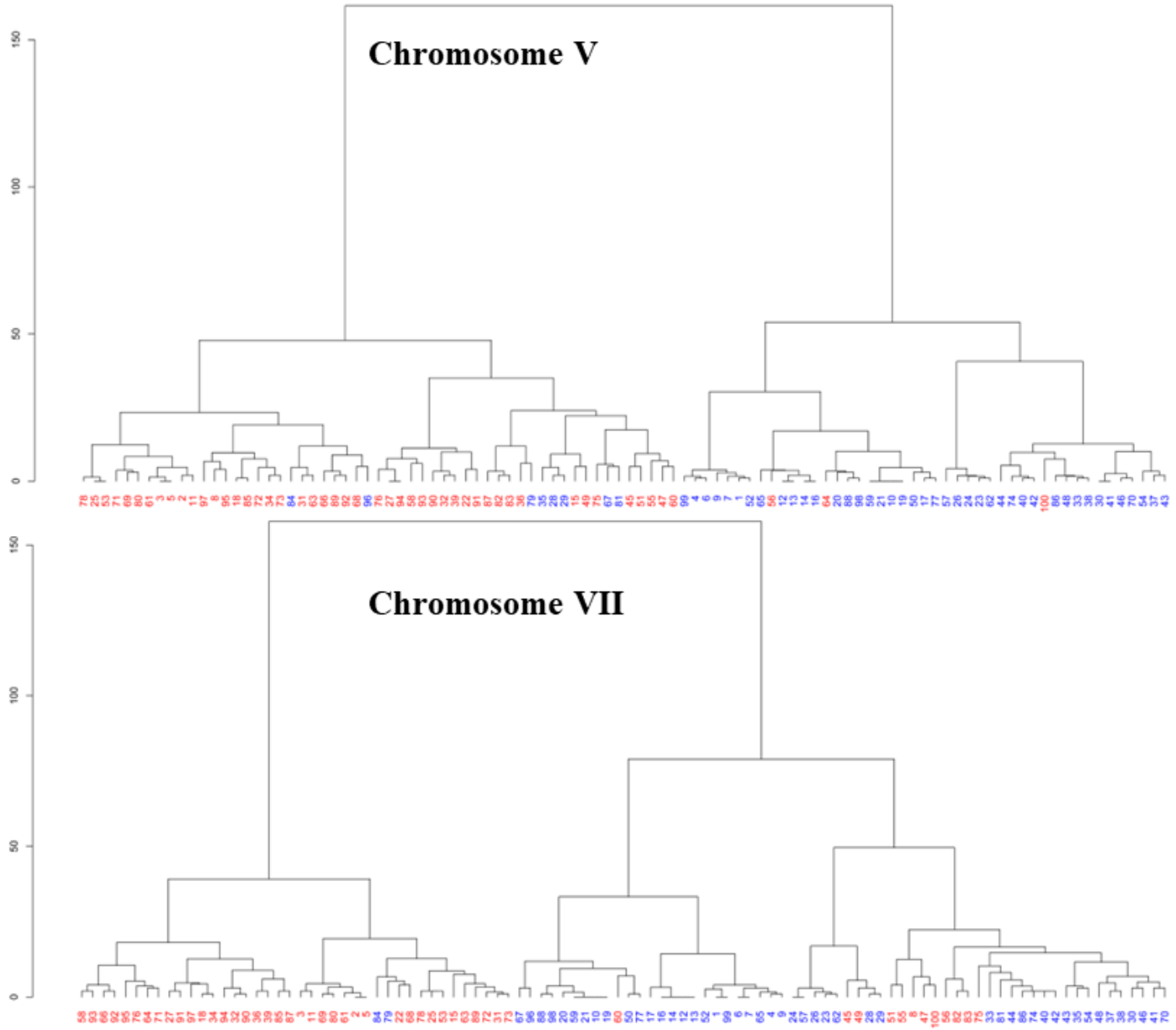

Figure S9. Dendrogram of the 100 cerumen samples using the VOMs in chromosomes V and VII pointed out as a panel of potential cancer biomarkers in canines. The candidate biomarkers did not successfully discriminate between the control (Blue) and cancer (red) samples.

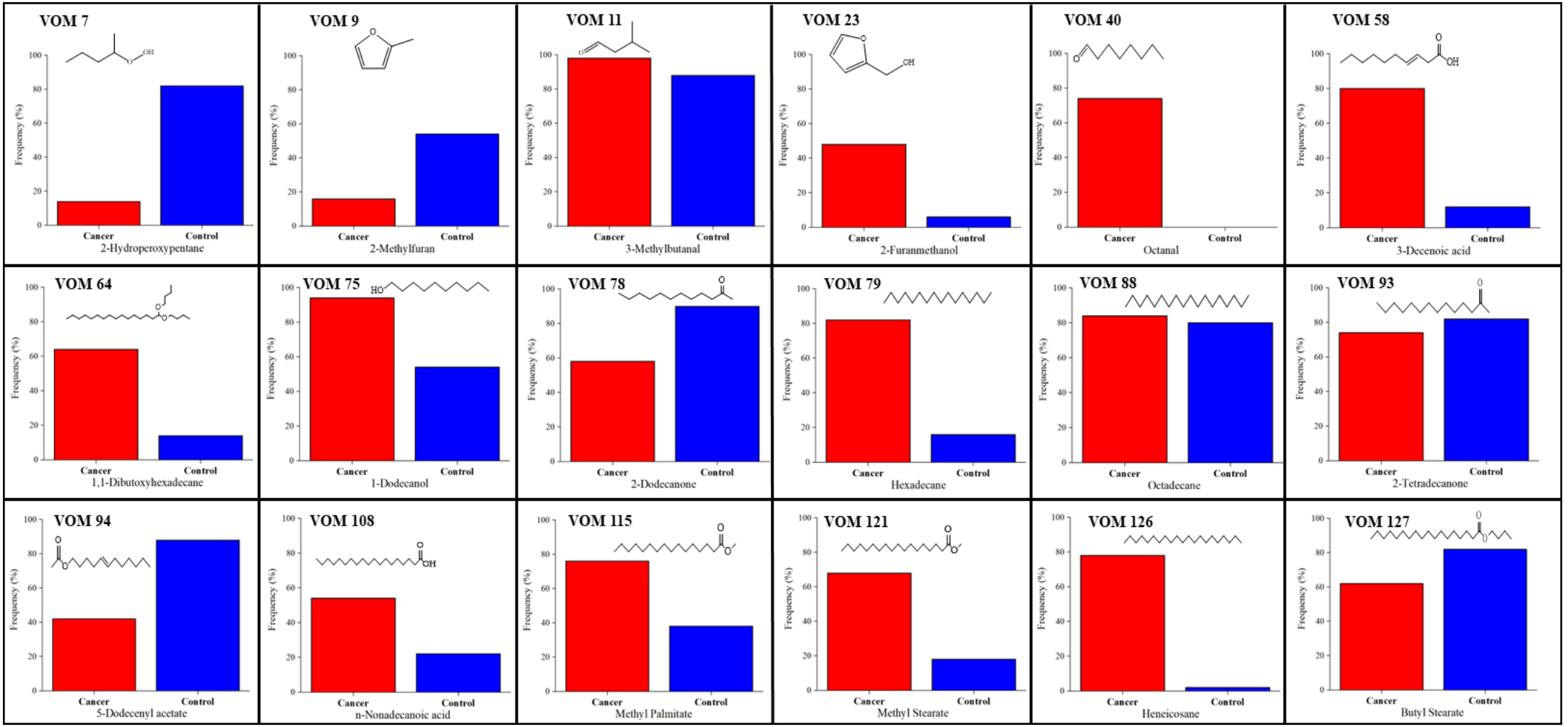

Figure S10. The panel of the potential cancer biomarkers detected in earwax samples of dogs. The metabolite occurrence (frequency, %) in a group (cancer and control) represents the sum of the presence of each volatile across the samples analyzed by HS/GC-MS. The VOMs are enumerated under their relative retention time (Table S2).

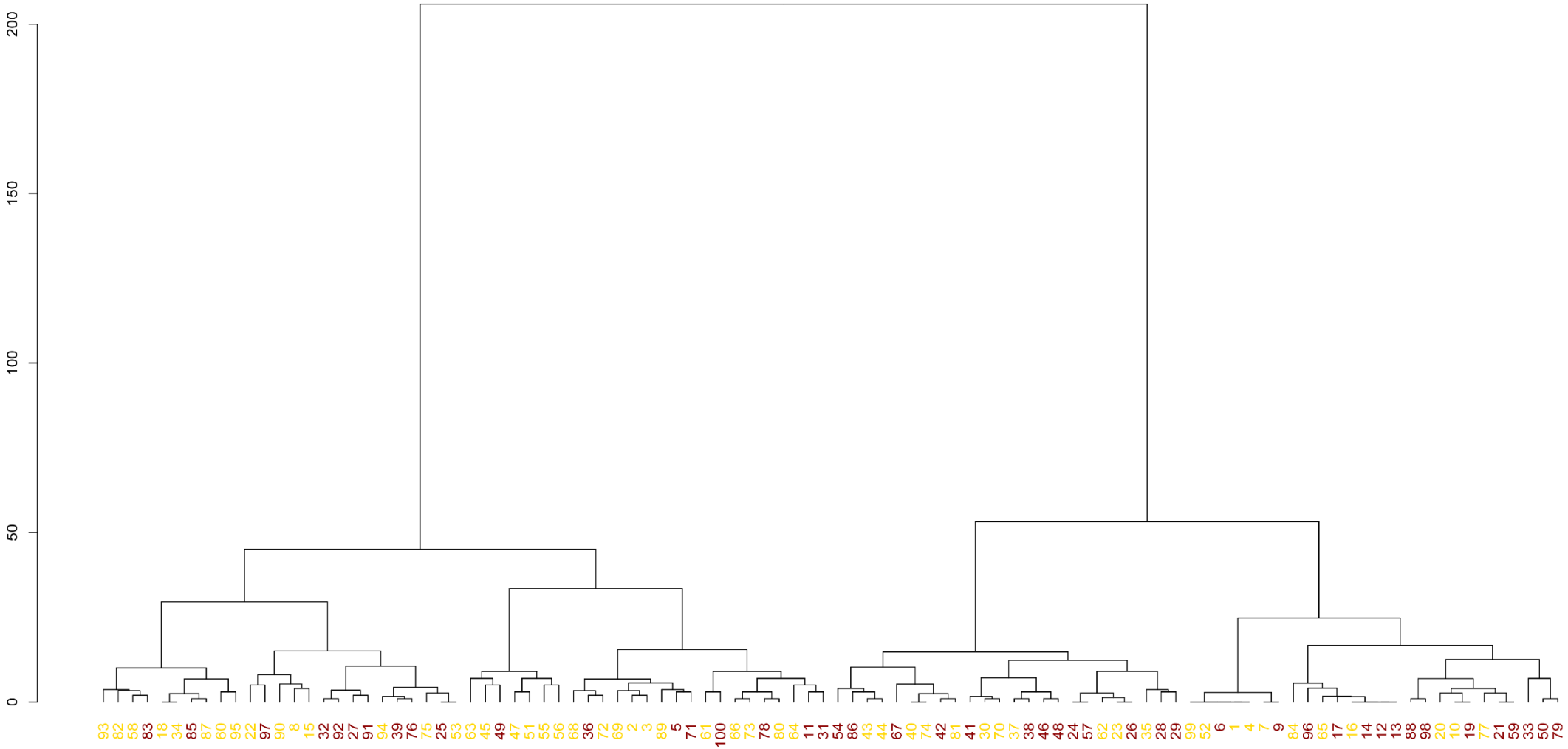

Figure S11. Hierarchical Cluster Analysis (HCA) using the 18 Volatile Organic Metabolites (VOMs) selected as potential cancer biomarkers in canines to evaluate the sex influence in the multivariate model. Cerumen samples: Gold = Female; Dark red = Male. Samples are enumerated by HS/GC-MS order of analysis.

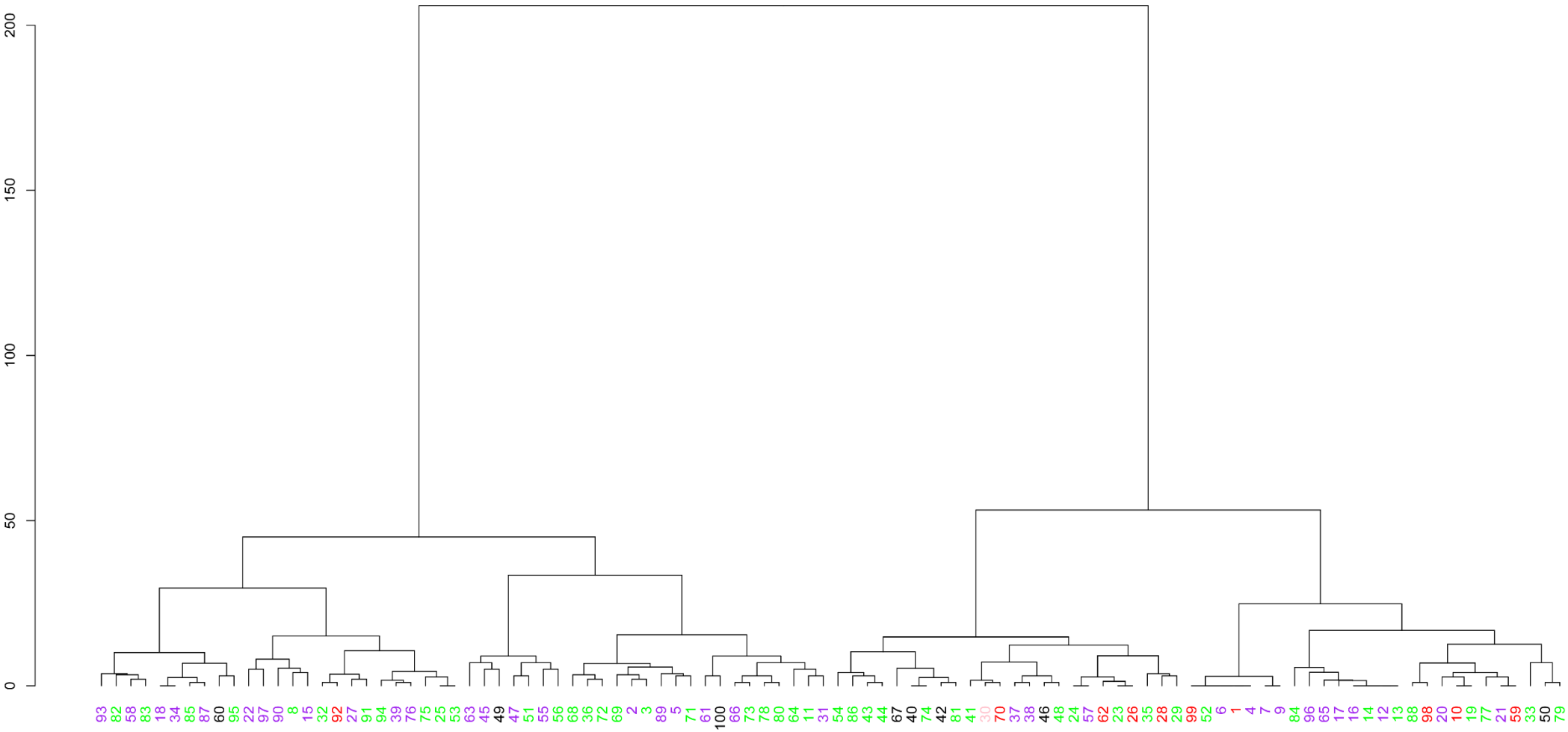

Figure S12. Hierarchical Cluster Analysis (HCA) using the 18 Volatile Organic Metabolites (VOMs) selected as potential cancer biomarkers in canines to evaluate the age bias in the multivariate model. Cerumen samples: Red = Adolescent (6–18 months) ; Purple = Adults (2–7 years); Green = Seniors (8–14 years); Pink = Geriatrics (>14 years); Black = Age not available. Samples are enumerated by HS/GC-MS order of analysis.

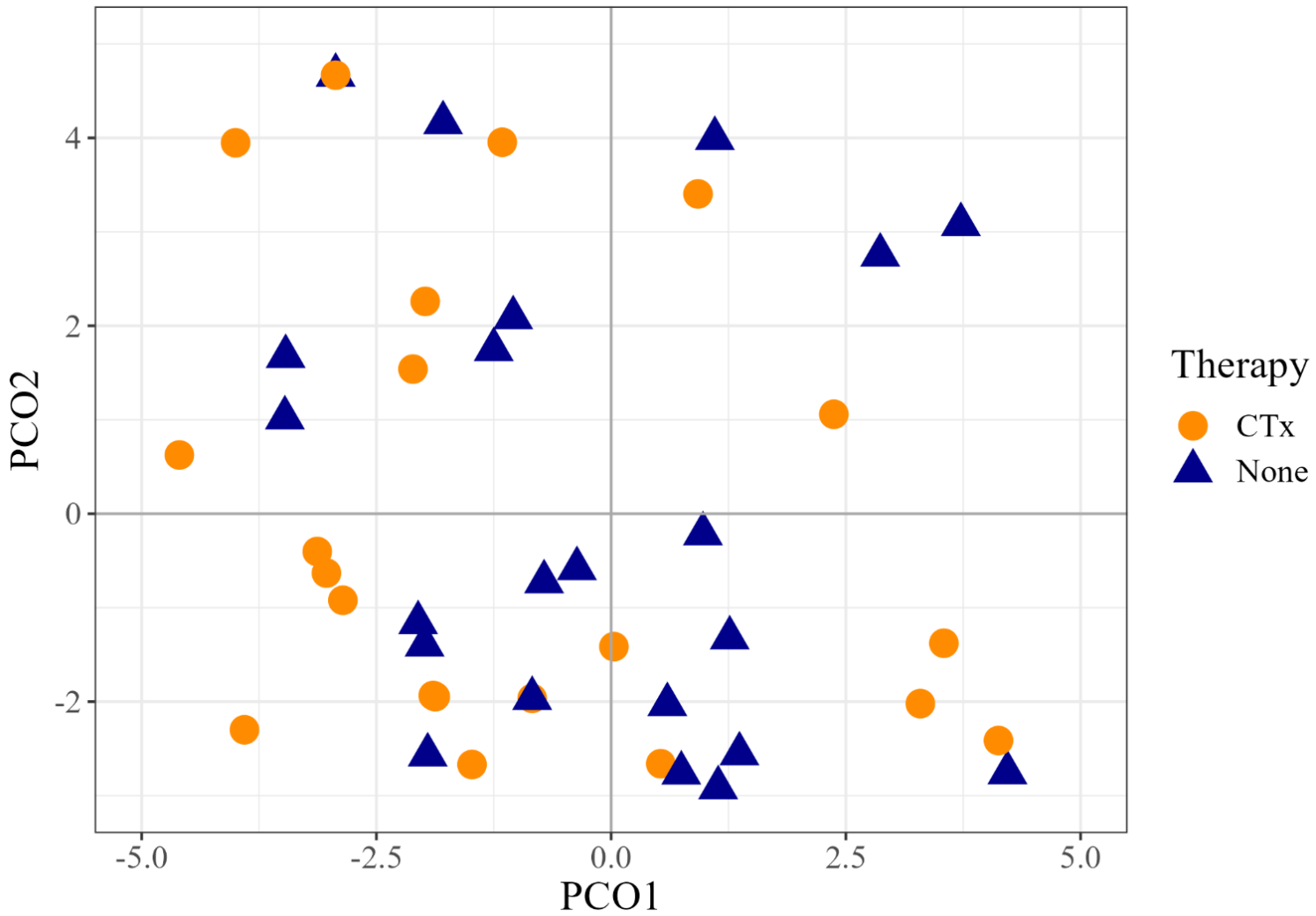

Figure S13. Principal coordinate analysis (PCoA) visualization to evaluate the chemotherapy treatment influence in cancer sample discrimination using the 18 Volatile Organic Metabolites (VOMs) selected as potential cancer biomarkers. Cerumen samples: Orange dots = Cancer dogs under chemotherapy therapy (CTx), Dark blue triangles = Cancer dogs with no chemotherapy treatment (None).

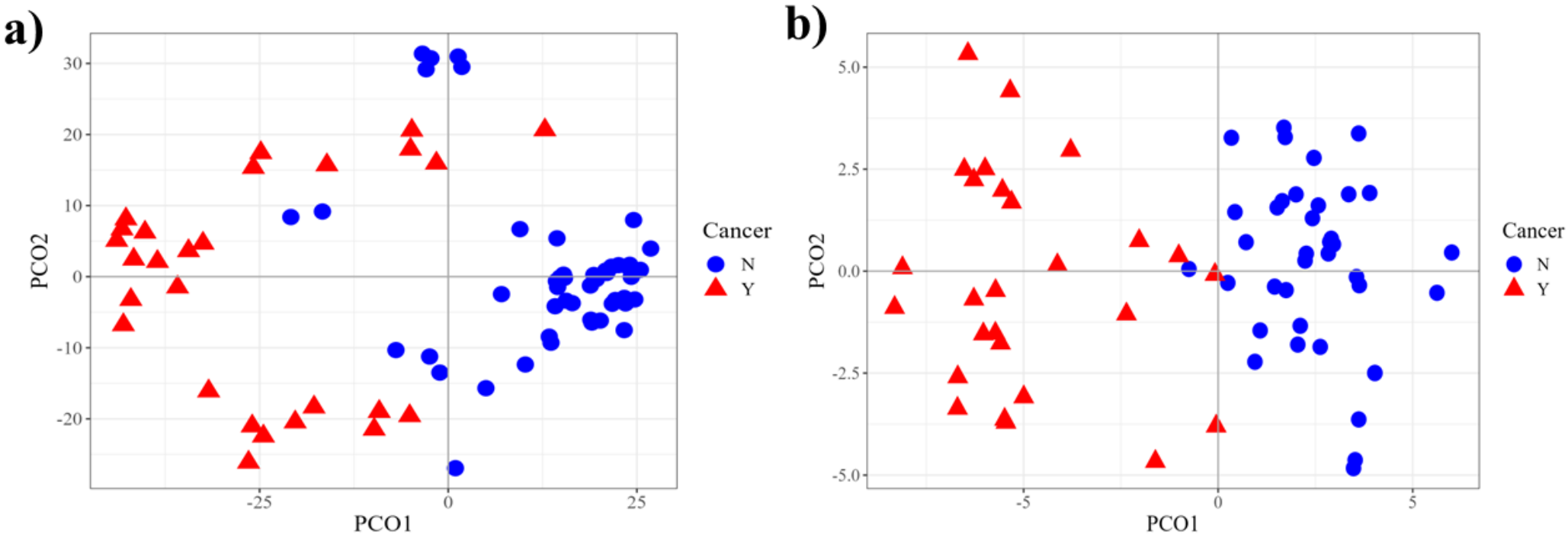

Figure S14. A discriminant trend showed in Principal coordinate analysis (PCoA) between the cerumen samples of the **control** and **cancer** group (without chemotherapy) using all the **a)** 128 Volatile Organic Metabolites (VOMs) annotated by HS/GC-MS. The trend is improved using the **b)** 18 VOMs selected as potential cancer biomarkers.

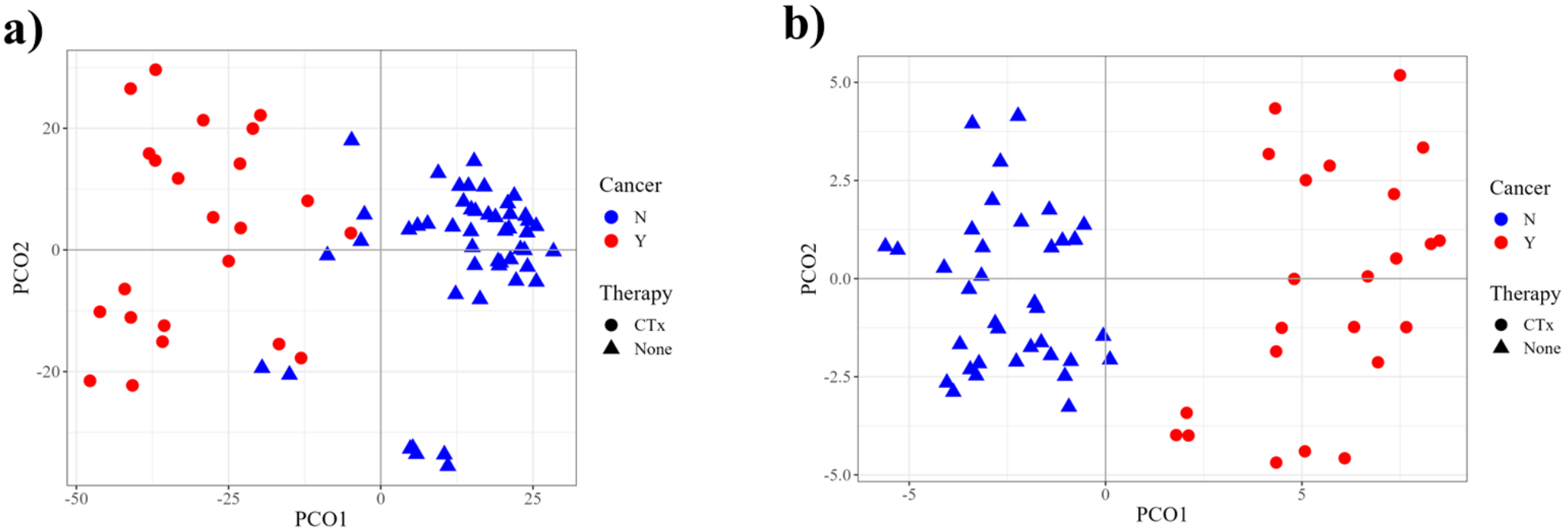

Figure S15. A discriminant trend showed in Principal coordinate analysis (PCoA) between the cerumen samples of the **control** and **cancer** group (with only cancer samples under chemotherapy - CTx) using all the **a)** 128 Volatile Organic Metabolites (VOMs) annotated by HS/GC-MS. The trend is improved using the **b)** 18 VOMs selected as potential cancer biomarkers.

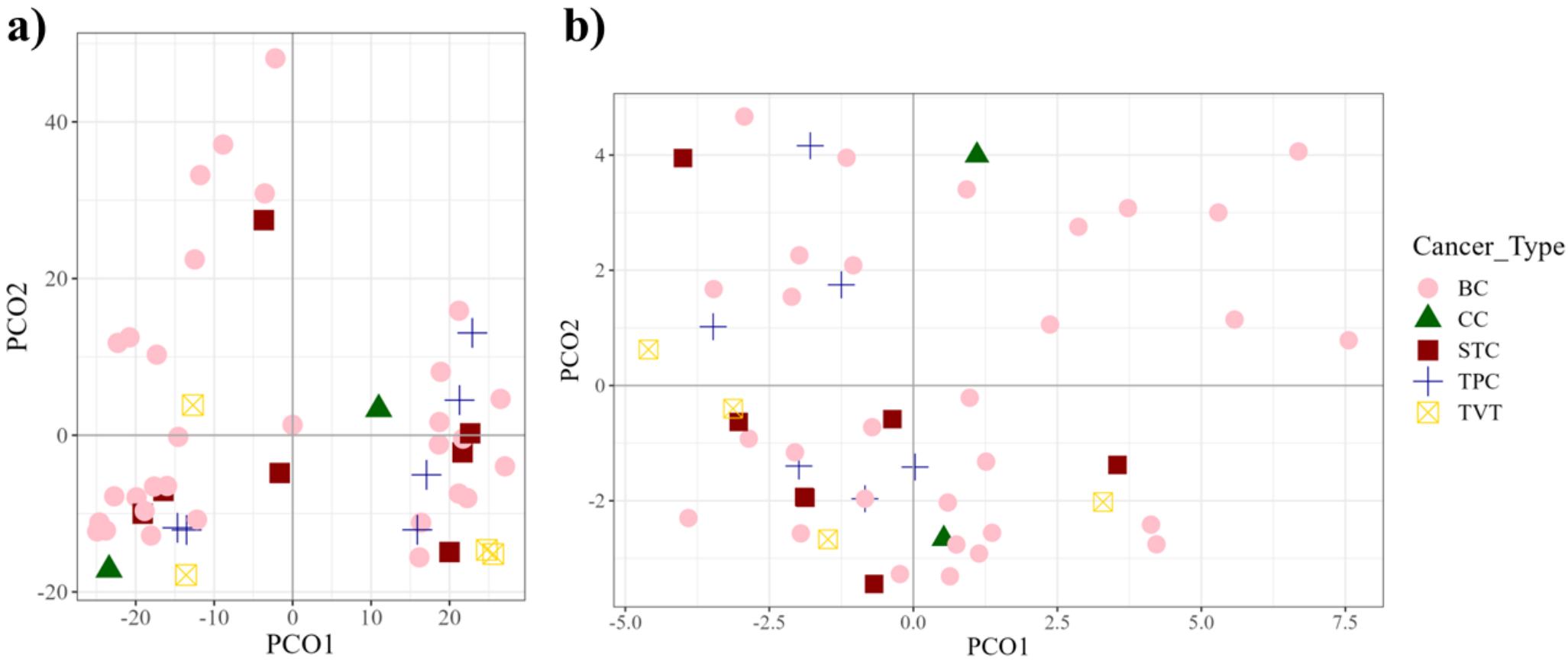

Figure S16. Principal coordinate analysis (PCoA) visualization aiming to evaluate the cancer type influence in cancer sample discrimination using the **a)** 128 Volatile Organic Metabolites (VOMs) annotated by HS/GC-MS and using the **b)** 18 VOMs selected as potential cancer biomarkers. BC = Breast Cancer; CC = Colorectal Cancer; STC = Soft tissue cancer; TPC = Testicular & prostate cancer; TVT = Transmissible venereal tumor.

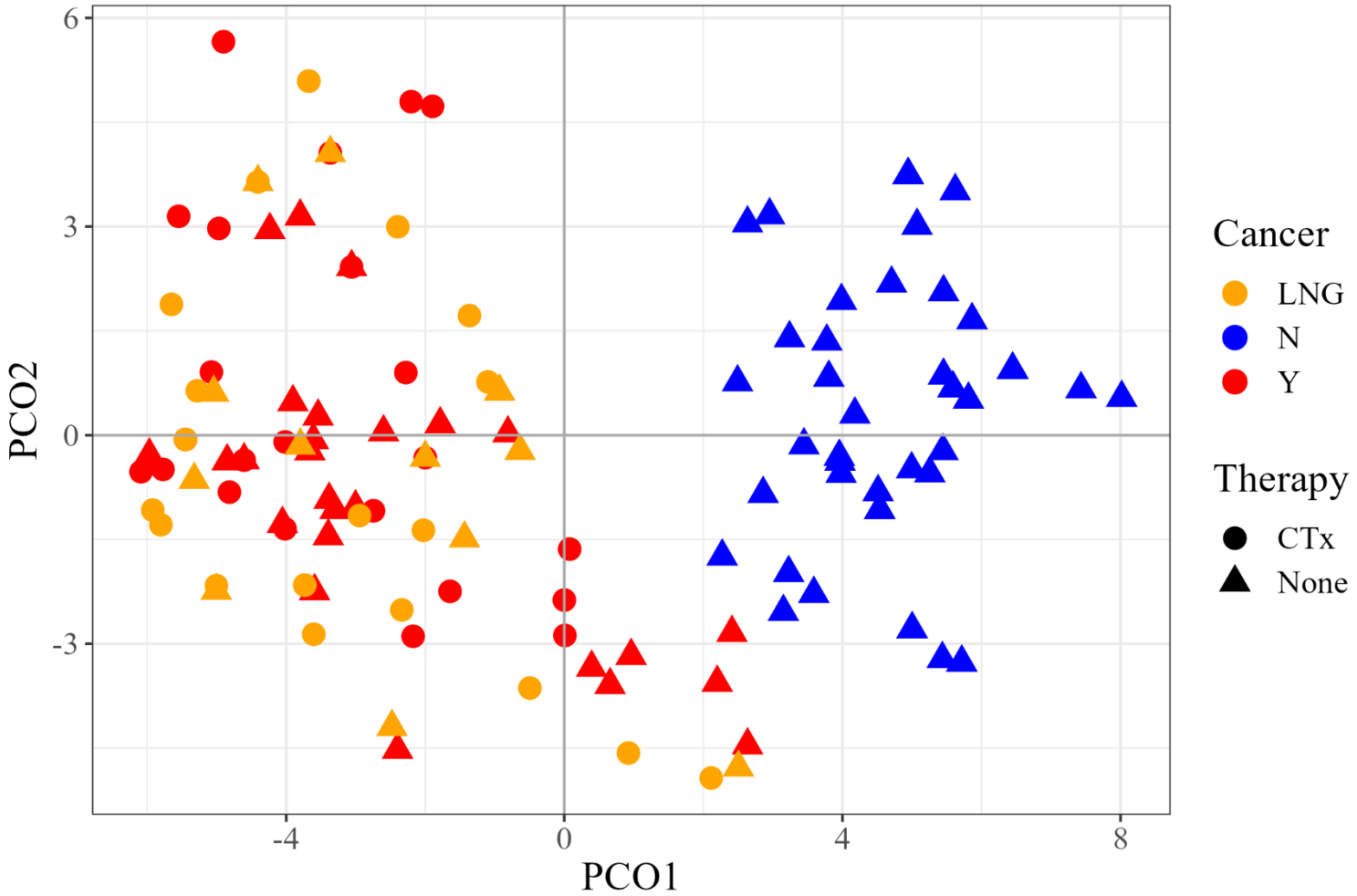

Figure S17. Principal coordinate analysis (PCoA) visualization of the 100 samples in the cerumenogram model (18 discriminant healthy/cancer VOMs) with the addition of the 33 cancer samples collected longitudinally (LNG – orange).

Table S1. Detailed characteristics of the dog cohort include breed, age, sex, and tumor type.

Characteristics	Healthy (Non-cancer)	Breast cancer (BC)	Soft tissue cancer (STC)	Transmissible venereal tumor (TVT)	Testicular & prostate cancer (TPC)	Colorectal cancer (CC)
Number of subjects, n (%)	50 (100.0)	31 (100.0)	7 (100.0)	4 (100.0)	6 (100.0)	2 (100.0)
Age range, n (%)						
Adolescents (6–18 months)	9 (18.0)	0 (0)	0 (0)	0 (0)	1 (16.7)	0 (0)
Adults (2–7 years)	14 (28.0)	16 (51.6)	3 (42.9)	2 (50.0)	0 (0)	1 (50.0)
Seniors (8–14 years)	21 (42.0)	15 (48.4)	2 (28.6)	2 (50.0)	5 (83.3)	0 (0)
Geriatrics (>14 years)	1 (2.0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Not available (NA)	5 (10.0)	0 (0)	2 (28.6)	0 (0)	0 (0)	1 (50.0)
Sex, n (%)						
Males	28 (56.0)	0 (0)	3 (42.9)	4 (100.0)	6 (100.0)	2 (100.0)
Females	22 (44.0)	31 (100.0)	4 (57.1)	0 (0)	0 (0)	0 (0)
Breed, n (%)						
1. American Hunter	0 (0)	0 (0)	0 (0)	0 (0)	1 (16.7)	0 (0)
2. Basset	0 (0)	1 (3.2)	0 (0)	0 (0)	0 (0)	0 (0)
3. Belgian shepherd	0 (0)	1 (3.2)	0 (0)	1 (25.0)	0 (0)	0 (0)
4. Bulldog	0 (0)	1 (3.2)	0 (0)	0 (0)	0 (0)	0 (0)
5. Dalmatian	0 (0)	4 (12.9)	0 (0)	0 (0)	0 (0)	0 (0)
6. Dachshund	0 (0)	4 (12.9)	0 (0)	0 (0)	0 (0)	0 (0)
7. Brazilian Mastiff	0 (0)	1 (3.2)	0 (0)	0 (0)	0 (0)	0 (0)
8. German spitz	0 (0)	1 (3.2)	0 (0)	0 (0)	0 (0)	0 (0)
9. Ilhaso Apso	0 (0)	1 (3.2)	0 (0)	0 (0)	0 (0)	0 (0)
10. Labrador Retreiver	0 (0)	1 (3.2)	0 (0)	0 (0)	1 (16.7)	1 (50.0)
11. Pitbull	1 (2.0)	0 (0)	0 (0)	0 (0)	1 (16.7)	0 (0)
12. Pinscher	2 (4.0)	2 (6.5)	2 (28.6)	0 (0)	2 (33.3)	0 (0)

Characteristics	Healthy (Non-cancer)	Breast cancer (BC)	Soft tissue cancer (STC)	Transmissible venereal tumor (TVT)	Testicular & prostate cancer (TPC)	Colorectal cancer (CC)
13. Poodle	2 (4.0)	3 (9.7)	1 (14.3)	1 (25.0)	0 (0)	0 (0)
14. Shar-pei	0 (0)	0 (0)	1 (14.3)	0 (0)	0 (0)	0 (0)
15. Chow chow	2 (4.0)	0 (0)	0 (0)	0 (0)	1 (16.7)	0 (0)
16. Shih Tzu	9 (18.0)	3 (9.7)	1 (14.3)	0 (0)	0 (0)	0 (0)
17. Yorkshire	5 (10.0)	1 (3.2)	0 (0)	0 (0)	0 (0)	0 (0)
18. Mixed	29 (58.0)	7 (22.6)	2 (28.6)	2 (50.0)	0 (0)	1 (50.0)
Castrated, n (%)	13 (26.0)	17 (54.8)	2 (28.6)	0 (0)	0 (0)	0 (0)
Histologic picture, n (%)						
Carcinoma	0 (0)	14 (45.2)	0 (0)	0 (0)	5 (83.3)	0 (0)
Adenocarcinoma	0 (0)	1 (3.2)	0 (0)	0 (0)	0 (0)	0 (0)
Mastocitoma	0 (0)	0 (0)	5 (71.4)	0 (0)	0 (0)	0 (0)
Liposarcoma	0 (0)	0 (0)	1 (14.3)	0 (0)	0 (0)	0 (0)
Lipoma	0 (0)	0 (0)	1 (14.3)	0 (0)	0 (0)	0 (0)
Carcinosarcoma	0 (0)	3 (9.7)	0 (0)	0 (0)	0 (0)	0 (0)
Lymphosarcoma	0 (0)	0 (0)	0 (0)	4 (100.0)	0 (0)	0 (0)
Lymphoma	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (100.0)
Neoplasia	0 (0)	13 (41.9)	0 (0)	0 (0)	1 (16.7)	0 (0)
Metastasis, n (%)						
Metastatic	0 (0)	3 (9.7)	1 (14.3)	0 (0)	1 (16.7)	1 (50.0)
Restricted/Nodular	0 (0)	28 (90.3)	6 (85.7)	4 (100.0)	5 (83.3)	1 (50.0)
Cancer Therapy, n (%)						
Chemotherapy	0 (0)	11 (35.5)	4 (57.1)	4 (100.0)	2 (33.3)	2 (100.0)
None	50 (100.0)	20 (64.5)	3 (42.9)	0 (0)	4 (66.7)	0 (0)

Table S2. List of VOMs encountered in canine earwax samples via HS/GC-MS analysis, their target MS peak, level of identification (ID), retention times, canonical SMILES, and the occurrence (%) in the groups: H=healthy (cancer-free), CCG = collective cancer (all cancer samples), BC=breast cancer, STC=Soft Tissue Cancer, TVT=Transmissible Venereal Tumor, TPC= Testicular & Prostate Cancer, CC= Colorectal Cancer.

Nº [a]	Volatile organic metabolites (VOMS)	Target peak (*m/z*)	Level of ID [b]	Absolute retention time (min)	Relative retention time	Canonical SMILES	CAS Number	Frequency of occurrence (%)						
								H	CCG	BC	STC	TVT	TPC	CC
1	Hydroxyurea	44	3	1.140	0.041	C(=O)(N)NO	127-07-1	100.0	100.0	100.0	100.0	100.0	100.0	100.0
2	Dimethylamine	44	3	1.225	0.045	CNC	124-40-3	96.0	96.0	93.5	100.0	100.0	100.0	100.0
3	Acetone	43	1	1.548	0.056	CC(=O)C	67-64-1	100.0	100.0	100.0	100.0	100.0	100.0	100.0
4	2-Methylpropanal	43	3	2.126	0.077	CC(C)C=O	78-84-2	90.0	82.0	90.3	71.4	50.0	66.7	100.0
5	Methylhydrazine	46	3	2.207	0.080	CNN	60-34-4	22.0	0.0	0.0	0.0	0.0	0.0	0.0
6	2,3-Dihydroxypropanoic acid	45	3	2.335	0.085	C(C(C(=O)O)O)O	473-81-4	30.0	76.0	74.2	85.7	100.0	66.7	50.0
7	**2-Hydroperoxypentane** [c]	43	3	2.570	0.093	CCCC(C)OO	14018-58-7	82.0	14.0	9.7	28.6	25.0	16.7	0.0
8	2-Butanone	43	3	2.684	0.098	CCC(=O)C	78-93-3	26.0	70.0	74.2	57.1	75.0	50.0	100.0
9	**2-Methylfuran**	82	3	2.840	0.103	CCC(=O)C	534-22-5	54.0	16.0	12.9	28.6	50.0	0.0	0.0
10	Acetic Acid	43	1	3.013	0.110	CC(=O)O	64-19-7	52.0	92.0	90.3	85.7	100.0	100.0	100.0
11	**3-Methylbutanal**	44	3	4.028	0.146	CC(C)CC=O	590-86-3	88.0	98.0	96.8	100.0	100.0	100.0	100.0
12	2-Methylbutanal	41	3	4.339	0.158	CCC(C)C=O	96-17-3	14.0	92.0	93.5	85.7	100.0	83.3	100.0
13	Pentanal	44	3	5.626	0.204	CCCCC=O	110-62-3	56.0	70.0	71.0	71.4	75.0	50.0	100.0
14	Heptane	43	1	5.889	0.214	CCCCCCC	142-82-5	18.0	10.0	6.5	14.3	25.0	0.0	50.0
15	Propylene Glycol	45	3	8.373	0.304	CC(CO)O	57-55-6	12.0	46.0	51.6	14.3	75.0	33.3	50.0
16	1H-Pyrrole	67	3	9.000	0.327	C1=CNC=C1	109-97-7	14.0	64.0	54.8	71.4	50.0	100.0	100.0
17	3-Methylheptane	43	3	9.540	0.347	CCCCC(C)CC	589-81-1	2.0	72.0	74.2	57.1	75.0	66.7	100.0

Nº [a]	Volatile organic metabolites (VOMS)	Target peak (*m/z*)	Level of ID [b]	Absolute retention time (min)	Relative retention time	Canonical SMILES	CAS Number	Frequency of occurrence (%)						
								H	CCG	BC	STC	TVT	TPC	CC
18	Acetamide	59	3	10.586	0.385	CC(=O)N	60-35-5	10.0	34.0	25.8	42.9	50.0	50.0	50.0
19	Hexanal	44	3	11.627	0.423	CCCCCC=O	66-25-1	16.0	82.0	83.9	71.4	75.0	83.3	100.0
20	N-Ethylacetamide	43	3	12.971	0.471	CCNC(=O)C	625-50-3	22.0	48.0	45.2	28.6	100.0	50.0	50.0
21	2-Methylpyrazine	94	3	13.430	0.488	CC1=NC=CN=C1	109-08-0	18.0	50.0	45.2	42.9	100.0	50.0	50.0
22	Furfural	96	3	13.980	0.508	C1=COC(=C1)C=O	98-01-1	6.0	86.0	83.9	100.0	75.0	100.0	50.0
23	**2-Furanmethanol**	98	3	16.836	0.612	C1=COC(=C1)CO	98-00-0	6.0	48.0	41.9	57.1	75.0	50.0	50.0
24	5-Methylhexanal	43	3	17.640	0.641	CC(C)CCCC=O	1860-39-5	4.0	42.0	38.7	28.6	75.0	66.7	0.0
25	3-Methylbutanoic acid	60	3	18.047	0.656	CC(C)CC(=O)O	503-74-2	6.0	36.0	38.7	14.3	50.0	33.3	50.0
26	2-Methylbutanoic acid	74	3	18.958	0.689	CCC(C)C(=O)O	116-53-0	4.0	4.0	0.0	14.3	25.0	0.0	0.0
27	1,2,3-Butanetriol	44	3	19.339	0.703	CC(C(CO)O)O	4435-50-1	4.0	30.0	25.8	42.9	50.0	33.3	0.0
28	Heptanal	70	3	21.555	0.783	CCCCCCC=O	111-71-7	12.0	52.0	48.4	71.4	50.0	50.0	50.0
29	3,4-Dimethyl-1,5-hexadiene	55	3	22.275	0.810	CC(C=C)C(C)C=C	4894-63-7	14.0	6.0	3.2	14.3	25.0	0.0	0.0
30	1,4-Benzoquinone	108	3	23.100	0.840	C1=CC(=O)C=CC1=O	106-51-4	12.0	34.0	41.9	42.9	25.0	0.0	0.0
31	Vinyl propionate	57	3	24.919	0.906	CCC(=O)OC=C	105-38-4	0.0	6.0	6.5	14.3	0.0	0.0	0.0
32	1-Heptanol	70	3	26.068	0.947	CCCCCCCO	111-70-6	16.0	20.0	16.1	28.6	50.0	16.7	0.0
33	5-Methyl-2-furaldehyde	110	3	29.515	1.073	CC1=CC=C(O1)C=O	620-02-0	14.0	44.0	48.4	42.9	25.0	33.3	50.0
34	Benzaldehyde	77	3	30.150	1.096	C1=CC=C(C=C1)C=O	100-52-7	10.0	46.0	45.2	42.9	0.0	66.7	100.0
35	2-Cyclohexen-1-one	68	3	31.751	1.154	C1CC=CC(=O)C1	930-68-7	10.0	46.0	38.7	57.1	75.0	50.0	50.0
36	5,6-Dihydro-2H-pyran-2-one	68	3	31.990	1.163	C1COC(=O)C=C1	3393-45-1	18.0	66.0	61.3	42.9	100.0	83.3	100.0
37	2-Ethyl-5-methylpyrazine	121	3	33.206	1.207	CCC1=NC=C(N=C1)C	13360-64-0	6.0	6.0	6.5	14.3	0.0	0.0	0.0
38	2,2-Dimethyldecane	57	3	33.465	1.216	CCCCCCCCC(C)(C)C	17302-37-3	26.0	60.0	51.6	85.7	75.0	50.0	100.0

Nº [a]	Volatile organic metabolites (VOMS)	Target peak (*m/z*)	Level of ID [b]	Absolute retention time (min)	Relative retention time	Canonical SMILES	CAS Number	Frequency of occurrence (%)						
								H	CCG	BC	STC	TVT	TPC	CC
39	2-Pentylfuran	81	3	34.584	1.257	CCCCCC1=CC=CO1	3777-69-3	60.0	46.0	48.4	71.4	0.0	33.3	50.0
40	**Octanal**	43	3	34.795	1.265	CCCCCCCC=O	124-13-0	0.0	74.0	74.2	71.4	50.0	100.0	50.0
41	Phenol	94	3	34.995	1.272	C1=CC=C(C=C1)O	108-95-2	6.0	42.0	48.4	28.6	25.0	33.3	50.0
42	2-Ethylhexanol	57	3	38.015	1.382	CCCCC(CC)CO	104-76-7	4.0	46.0	45.2	57.1	25.0	50.0	50.0
43	Phenylacetaldehyde	91	3	38.560	1.401	C1=CC=C(C=C1)CC=O	122-78-1	4.0	66.0	61.3	71.4	100.0	50.0	100.0
44	2-Methyl-2-heptenal	41	3	39.227	1.426	CCCCC=C(C)C=O	30567-26-1	8.0	32.0	35.5	0.0	25.0	50.0	50.0
45	3-Methyl-2-cyclohexen-1-one	82	3	39.439	1.433	CC1=CC(=O)CCC1	1193-18-6	8.0	66.0	67.7	71.4	75.0	50.0	50.0
46	Butoxybenzene	94	3	41.530	1.509	CCCCOC1=CC=CC=C1	1126-79-0	4.0	42.0	45.2	14.3	25.0	50.0	100.0
47	2-Nonen-1-ol	57	3	41.665	1.514	CCCCCCC=CCO	22104-79-6	2.0	40.0	35.5	42.9	75.0	33.3	50.0
48	2-Pyrrolidone	85	3	41.910	1.523	C1CC(=O)NC1	616-45-5	6.0	26.0	22.6	14.3	75.0	16.7	50.0
49	1-Octanol	56	3	42.210	1.534	CCCCCCCCO	111-87-5	6.0	64.0	64.5	85.7	50.0	33.3	100.0
50	Tetrahydro-6-methyl-2H-pyran-2-one	42	3	42.803	1.556	CC1CCCC(=O)O1	823-22-3	12.0	66.0	54.8	71.4	100.0	83.3	100.0
51	Tetrahydro-4-methyl-2H-pyran-2-one	42	3	43.137	1.568	CC1CCOC(=O)C1	1121-84-2	4.0	64.0	58.1	57.1	100.0	66.7	100.0
52	2-Nonanone	43	1	43.672	1.587	CCCCCCCC(=O)C	821-55-6	48.0	84.0	80.6	100.0	100.0	83.3	50.0
53	2-Oxepanone	42	3	44.398	1.614	C1CCC(=O)OCC1	502-44-3	4.0	58.0	51.6	57.1	100.0	50.0	100.0
54	Nonanal	57	3	44.891	1.632	CCCCCCCCC=O	124-19-6	14.0	50.0	58.1	42.9	25.0	33.3	50.0
55	4H-Pyran-4-one	96	3	45.342	1.648	C1=COC=CC1=O	108-97-4	2.0	10.0	6.5	14.3	0.0	33.3	0.0
56	Succinimide	99	3	46.891	1.704	C1CC(=O)NC1=O	123-56-8	10.0	30.0	25.8	28.6	75.0	16.7	50.0
57	Isoborneol	95	3	49.321	1.793	CC1(C2CCC1(C(C2)O)C)C	124-76-5	58.0	88.0	83.9	100.0	75.0	100.0	100.0
58	**3-Decenoic acid**	43	3	51.935	1.888	CCCCCCC=CCC(=O)O	15469-77-9	12.0	80.0	77.4	57.1	100.0	100.0	100.0

Nº [a]	Volatile organic metabolites (VOMS)	Target peak (*m/z*)	Level of ID [b]	Absolute retention time (min)	Relative retention time	Canonical SMILES	CAS Number	Frequency of occurrence (%)						
								H	CCG	BC	STC	TVT	TPC	CC
59	1-Decanol	70	3	53.490	1.944	CCCCCCCCCCO	112-30-1	0.0	36.0	32.3	28.6	75.0	33.3	50.0
60	Dodecane	57	3	54.502	1.981	CCCCCCCCCCCC	112-40-3	18.0	38.0	35.5	28.6	100.0	16.7	50.0
61	4-Hydroxyquinazoline	146	3	57.575	2.092	C1=CC=C2C(=C1)C(=O)NC=N2	491-36-1	12.0	28.0	19.4	42.9	50.0	50.0	0.0
62	Undecanal	41	3	58.765	2.136	CCCCCCCCCCC=O	112-44-7	30.0	76.0	71.0	71.4	75.0	100.0	100.0
63	Hydroquinone	110	3	59.874	2.176	C1=CC(=CC=C1O)O	123-31-9	40.0	74.0	71.0	85.7	100.0	66.7	50.0
64	**1,1-Dibutoxyhexadecane**	57	3	61.350	2.230	CCCCCCCCCCCCCCCCC(OCCCC)OCCCC	18302-68-6	14.0	64.0	64.5	57.1	100.0	50.0	50.0
65	2,4-Dodecadienal	81	3	61.724	2.243	CCCCCCCC=CC=CC=O	13162-47-5	28.0	80.0	80.6	57.1	75.0	100.0	100.0
66	8-Methyl-1-undecene	43	3	62.965	2.288	CCCC(C)CCCCCC=C	74630-40-3	50.0	64.0	58.1	71.4	100.0	50.0	100.0
67	3-Methylenetridecane	70	3	63.656	2.314	CCCCCCCCCCC(=C)CC	19780-34-8	52.0	68.0	67.7	57.1	75.0	83.3	50.0
68	5-Pentyldihydro-2(3H)-furanone	85	3	64.304	2.337	CCCCCC1CCC(=O)O1	104-61-0	30.0	64.0	58.1	85.7	75.0	66.7	50.0
69	2,6-Dimethyl-4-propyl-4-heptanol	129	3	64.573	2.347	CCCC(CCC)(CC(C)(C)C)O	54774-83-3	18.0	62.0	64.5	57.1	100.0	33.3	50.0
70	Resorcinol monoacetate	110	3	65.050	2.364	CC(=O)OC1=CC=CC(=C1)O	102-29-4	88.0	82.0	83.9	100.0	100.0	66.7	0.0
71	Tetradecane	57	1	65.818	2.392	CCCCCCCCCCCCCC	629-59-4	70.0	34.0	29.0	57.1	50.0	33.3	0.0
72	Dodecanal	41	3	66.050	2.401	CCCCCCCCCCCC=O	112-54-9	10.0	36.0	45.2	28.6	0.0	16.7	50.0
73	Pentadecane	57	1	66.240	2.407	CCCCCCCCCCCCCCC	629-62-9	24.0	32.0	32.3	57.1	25.0	0.0	50.0
74	Dimethyl phthalate	163	3	67.435	2.451	COC(=O)C1=CC=CC=C1C(=O)OC	131-11-3	40.0	68.0	67.7	85.7	50.0	50.0	100.0
75	**1-Dodecanol**	55	3	68.140	2.476	CCCCCCCCCCCCO	112-53-8	54.0	94.0	90.3	100.0	100.0	100.0	100.0

Nº [a]	Volatile organic metabolites (VOMS)	Target peak (*m/z*)	Level of ID [b]	Absolute retention time (min)	Relative retention time	Canonical SMILES	CAS Number	Frequency of occurrence (%)						
								H	CCG	BC	STC	TVT	TPC	CC
76	9-Decen-1-yl acetate	43	3	68.516	2.490	CC(=O)OCCCCCCCCC=C	50816-18-7	60.0	32.0	35.5	14.3	50.0	16.7	50.0
77	1-Tridecanol	55	3	68.625	2.494	CCCCCCCCCCCCCO	112-70-9	50.0	54.0	58.1	57.1	25.0	33.3	100.0
78	**2-Dodecanone**	58	3	68.776	2.500	CCCCCCCCCCC(=O)C	6175-49-1	90.0	58.0	51.6	71.4	100.0	50.0	50.0
79	**Hexadecane**	57	1	68.955	2.506	CCCCCCCCCCCCCCCC	544-76-3	16.0	82.0	74.2	100.0	100.0	100.0	50.0
80	Tridecanal	57	3	69.285	2.518	CCCCCCCCCCCCC=O	10486-19-8	12.0	78.0	80.6	85.7	75.0	66.7	50.0
81	Tetradecanal	57	3	69.669	2.532	CCCCCCCCCCCCCC=O	124-25-4	40.0	48.0	54.8	42.9	50.0	16.7	50.0
82	Uric acid	43	3	69.856	2.539	C12=C(NC(=O)N1)NC(=O)NC2=O	69-93-2	38.0	20.0	16.1	42.9	50.0	0.0	0.0
83	Dodecanoic acid	73	3	70.594	2.566	CCCCCCCCCCCC(=O)O	143-07-7	62.0	72.0	71.0	71.4	100.0	66.7	50.0
84	1-Tetradecanol	55	3	70.908	2.577	CCCCCCCCCCCCCCO	112-72-1	84.0	80.0	77.4	85.7	75.0	100.0	50.0
85	1-Cyclododecylethanone	43	3	71.180	2.587	CC(=O)C1CCCCCCCCCCC1	28925-00-0	64.0	46.0	51.6	57.1	25.0	33.3	0.0
86	2-Hexyl-1-decanol	57	3	71.379	2.594	CCCCCCCCC(CCCCCC)CO	2425-77-6	72.0	48.0	54.8	28.6	50.0	33.3	50.0
87	2-Tridecanone	58	3	71.450	2.597	CCCCCCCCCCCC(=O)C	593-08-8	24.0	62.0	58.1	71.4	25.0	83.3	100.0
88	**Octadecane**	57	3	71.546	2.600	CCCCCCCCCCCCCCCCCC	593-45-3	80.0	84.0	77.4	85.7	100.0	100.0	100.0
89	Pentadecanal	82	3	71.910	2.613	CCCCCCCCCCCCCCC=O	2765-11-9	26.0	82.0	80.6	71.4	100.0	83.3	100.0
90	Hexadecanal	82	3	72.270	2.627	CCCCCCCCCCCCCCCC=O	629-80-1	60.0	60.0	61.3	57.1	50.0	66.7	50.0

Nº [a]	Volatile organic metabolites (VOMS)	Target peak (*m/z*)	Level of ID [b]	Absolute retention time (min)	Relative retention time	Canonical SMILES	CAS Number	Frequency of occurrence (%)						
								H	CCG	BC	STC	TVT	TPC	CC
91	n-Tetradecanoic acid	73	3	73.020	2.654	CCCCCCCCCCCCCC(=O)O	544-63-8	70.0	80.0	77.4	71.4	100.0	100.0	50.0
92	1-Pentadecanol	43	3	73.286	2.663	CCCCCCCCCCCCCCCO	629-76-5	76.0	64.0	58.1	57.1	75.0	100.0	50.0
93	**2-Tetradecanone**	58	3	73.385	2.667	CCCCCCCCCCCCC(=O)C	2345-27-9	82.0	74.0	67.7	100.0	75.0	83.3	50.0
94	**5-Dodecenyl acetate**	43	3	73.643	2.676	CCCCCCC=CCCCCOC(=O)C	16676-96-3	88.0	42.0	45.2	28.6	25.0	33.3	100.0
95	2-Pentadecanone	58	3	73.880	2.685	CCCCCCCCCCCCCC(=O)C	2345-28-0	32.0	76.0	74.2	71.4	100.0	66.7	100.0
96	11-Dodecen-1-yl acetate	43	3	73.935	2.687	CC(=O)OCCCCCCCCCC=C	35153-10-7	14.0	52.0	51.6	42.9	50.0	50.0	100.0
97	3-Pyrrolidin-2-yl-propanoic acid	70	3	74.083	2.692	C1CC(NC1)CCC(=O)O	18325-18-3	24.0	36.0	32.3	57.1	0.0	66.7	0.0
98	2-(dodecyloxy)ethanol	57	3	74.218	2.697	CCCCCCCCCCCCOCCO	4536-30-5	26.0	56.0	58.1	42.9	25.0	66.7	100.0
99	2-Octyl-1-decanol	57	3	74.390	2.704	CCCCCCCCC(CCCCCCCC)CO	45235-48-1	52.0	76.0	71.0	57.1	100.0	100.0	100.0
100	n-Pentadecanoic acid	73	3	75.257	2.735	CCCCCCCCCCCCCCC(=O)O	1002-84-2	50.0	76.0	77.4	57.1	75.0	83.3	100.0
101	2-Hexadecanone	58	3	75.360	2.739	CCCCCCCCCCCCCCC(=O)C	18787-63-8	20.0	56.0	51.6	71.4	50.0	50.0	100.0
102	2-Heptadecanone	58	3	75.587	2.747	CCCCCCCCCCCCCCCCC(=O)C	2922-51-2	70.0	66.0	58.1	71.4	100.0	66.7	100.0
103	1-Octadecanol	43	3	75.875	2.758	CCCCCCCCCCCCCCCCCCO	112-92-5	66.0	68.0	64.5	71.4	50.0	83.3	100.0
104	Isopropyl myristate	43	3	76.177	2.769	CCCCCCCCCCCCCC(=O)OC(C)C	110-27-0	72.0	42.0	32.3	57.1	50.0	66.7	50.0
105	N-butylbenzenesulfonamide	77	3	76.801	2.791	CCCCNS(=O)(=O)C1	3622-84-2	28.0	64.0	64.5	42.9	75.0	66.7	100.0

Nº [a]	Volatile organic metabolites (VOMS)	Target peak (*m/z*)	Level of ID [b]	Absolute retention time (min)	Relative retention time	Canonical SMILES	CAS Number	Frequency of occurrence (%)						
								H	CCG	BC	STC	TVT	TPC	CC
						=CC=CC=C1								
106	2-Nonadecanone	58	3	77.037	2.800	CCCCCCCCCCCCCCCCCC(=O)C	629-66-3	56.0	76.0	80.6	57.1	75.0	66.7	100.0
107	n-Octadecanoic acid	43	3	78.244	2.844	CCCCCCCCCCCCCCCCCC(=O)O	57-11-4	90.0	96.0	93.5	100.0	100.0	100.0	100.0
108	**n-Nonadecanoic acid**	43	3	78.405	2.850	CCCCCCCCCCCCCCCCCCC(=O)O	646-30-0	22.0	54.0	58.1	57.1	25.0	33.3	100.0
109	Butyl isodecyl phthalate	149	3	78.636	2.858	CCCCOC(=O)C1=CC=CC=C1C(=O)OCCCCCCCC(C)C	89-18-9	26.0	80.0	77.4	71.4	100.0	83.3	100.0
110	Cyclopentadecanone	55	3	78.920	2.868	C1CCCCCCCC(=O)CCCCCC1	502-72-7	84.0	78.0	74.2	71.4	100.0	83.3	100.0
111	2-Hydroxycyclopentadecanone	55	3	79.047	2.873	C1CCCCCCC(C(=O)CCCCCC1)O	4727-18-8	16.0	82.0	77.4	85.7	100.0	100.0	50.0
112	6,10,14-Trimethyl-2-pentadecanone	43	3	79.708	2.897	CC(C)CCCC(C)CCCC(C)CCCC(=O)C	502-69-2	64.0	74.0	67.7	85.7	100.0	66.7	100.0
113	Methyl Palmitoleate	55	3	79.880	2.903	CCCCCCC=CCCCCCCC(=O)OC	1120-25-8	26.0	76.0	71.0	85.7	100.0	83.3	50.0
114	Eicosanoic acid	43	3	80.342	2.920	CCCCCCCCCCCCCCCCCCC(=O)O	506-30-9	46.0	60.0	58.1	71.4	50.0	50.0	100.0
115	**Methyl Palmitate**	74	3	80.510	2.926	CCCCCCCCCCCCCCCCCC(=O)OC	112-39-0	38.0	76.0	64.5	100.0	100.0	83.3	100.0
116	N-methyldodecanamide	73	3	81.830	2.974	CCCCCCCCCCCC(=O)NC	27563-67-3	66.0	42.0	32.3	57.1	50.0	66.7	50.0
117	Methyl Heptadecanoate	74	3	82.458	2.997	CCCCCCCCCCCCCCCCC(=O)OC	1731-92-6	32.0	50.0	41.9	71.4	50.0	66.7	50.0
118	Tetradecanamide	59	3	82.867	3.012	CCCCCCCCCCCCCC(=O)N	638-58-4	38.0	18.0	19.4	14.3	25.0	16.7	0.0

Nº [a]	Volatile organic metabolites (VOMS)	Target peak (*m/z*)	Level of ID [b]	Absolute retention time (min)	Relative retention time	Canonical SMILES	CAS Number	Frequency of occurrence (%)						
								H	CCG	BC	STC	TVT	TPC	CC
119	Eicosanol	43	3	84.955	3.088	CCCCCCCCCCCCCCCCCCCCO	629-96-9	38.0	62.0	54.8	85.7	75.0	66.7	50.0
120	Methyl Linoleate	67	3	85.449	3.106	CCCCCC=CCC=CCCCCCCCC(=O)OC	112-63-0	52.0	74.0	80.6	71.4	25.0	83.3	50.0
121	**Methyl stearate**	74	3	85.930	3.123	CCCCCCCCCCCCCCCCCCCC(=O)OC	112-61-8	18.0	68.0	58.1	85.7	75.0	100.0	50.0
122	Heneicosanoic acid	326	3	86.653	3.149	CCCCCCCCCCCCCCCCCCCCC(=O)O	2363-71-5	100.0	88.0	83.9	85.7	100.0	100.0	100.0
123	Octadecanamide	59	3	87.065	3.164	CCCCCCCCCCCCCCCCCCC(=O)N	124-26-5	48.0	78.0	67.7	85.7	100.0	100.0	100.0
124	Butyl palmitate	56	3	87.222	3.170	CCCCCCCCCCCCCCCCCC(=O)OCCCC	111-06-8	14.0	86.0	83.9	85.7	75.0	100.0	100.0
125	1-Heneicosanol	55	3	89.074	3.237	CCCCCCCCCCCCCCCCCCCCO	15594-90-8	24.0	58.0	58.1	42.9	50.0	83.3	50.0
126	**Heneicosane**	57	3	89.920	3.268	CCCCCCCCCCCCCCCCCCCCC	629-94-7	2.0	78.0	71.0	100.0	75.0	100.0	50.0
127	**Butyl stearate**	56	3	92.011	3.344	CCCCCCCCCCCCCCCCCC(=O)OCCCC	123-95-5	82.0	62.0	61.3	71.4	50.0	66.7	50.0
128	Bis(2-ethylhexyl) phthalate	149	3	97.594	3.547	CCCCC(CC)COC(=O)C1=CC=CC=C1C(=O)OCC(CC)CCCC	117-81-7	54.0	74.0	71.0	71.4	50.0	100.0	100.0

[a] VOMs are presented in order of increasing relative retention time to the internal standard (3-methylcyclohexanone); [b] Level of ID: 1 – Metabolites confirmed with authentic standards. 3 – Metabolites with at least 80% matching with the NIST17s libraries and confirmed by their retention times with the in-house cerumen database. [c] VOMs selected as potential cancer biomarkers using variable selection combining genetic algorithm (GA) and partial least squares (PLS) were highlighted in bold.

Table S3. Summary performance of the chromosomes (I–XII) in the GA-PLS for the cerumenogram model after repeated k-fold cross-validation (k=10, repeats=10) using 80% of the samples as the training set (80 samples) and 20% as the test set (20 samples).

GA-PLS models					Training set (80samples)						Test set (20 samples)					
[a]Chrom Nº	[b]Nº of comp	[c]ROC	[d]Nº of var	[e]Variable Importance, N (%)	Accuracy (95% CI)	[f]Sens	[g]Spec	Kappa value	[h]H. Pred Value	[i]CA. Pred Value	Accuracy (95% CI)	Sens	Spec	Kappa value	H. Pred Value	CA. Pred Value
I	1	1.000	18	VOMs: 126(100.0), 79(90.4), 40(89.4), 7(87.2), 58(83.4), 121(66.1), 64(62.5), 23(54.8), 94(51.7), 9(50.7), 115(50.2), 78(44.7), 75(41.3), 108(38.7), 127(20.8), 11(8.2), 88(4.1), 93(0)	1.000 (0.955-1.000)	1.000	1.000	1.000	1.000	1.000	0.950 (0.751-0.999)	1.000	0.929	0.900	0.909	1.000
II	1	0.946	27	VOMs: 79(100.0), 109(82.9), 121(75.5), 36(67.4), 6(66.5), 8(62.7), 42(60.7), 87(59.3), 9(49.1), 21(47.5), 125(46.7), 63(45.4), 74(41.4), 52(40.1), 57(39.9), 5(34.6), 30(31.7), 72(27.4), 116(25.8), 67(24.1)	0.900 (0.812-0.956)	0.923	0.878	0.800	0.878	0.913	0.900 (0.683-0.988)	1.000	0.778	0.794	0.846	1.000
III	3	0.994	20	VOMs: 126(100.0), 124(95.5), 109(75.4), 89(69.4), 121(67.0), 6(58.0), 62(57.3), 8(49.5), 125(48.6), 74(44.2), 57(34.4), 39(30.6), 21(26.9), 91(24.8), 83(24.7), 44(24.6), 112(16.7), 107(5.2), 11(3.5), 93(0)	0.975 (0.913-0.997)	1.000	0.944	0.949	0.956	1.000	0.750 (0.509-0.913)	1.000	0.643	0.519	0.545	1.000
IV	3	0.976	22	VOMs: 126(100.0), 89(76.9), 121(56.9), 10(49.6), 52(49.3), 71(46.1), 8(45.0), 57(42.3), 21(40.4), 125(39.6), 74(33.5), 98(25.6), 39(18.5), 81(17.8), 91(17.5), 93(17.0), 61(14.9), 11(13.6), 97(13.4), 112(12.4)	0.975 (0.913-0.997)	0.976	0.974	0.950	0.976	0.974	0.800 (0.563-0.943)	1.000	0.636	0.612	0.692	1.000
V	3	0.940	24	VOMs: 126(100.0), 109(61.6), 89(53.5), 8(52.6), 6(51.9), 59(43.5), 63(42.8), 52(42.5), 78(42.3), 30(34.3), 125(33.5), 98(32.8), 57(31.8), 44(28.3), 56(25.3), 21(23.5), 39(17.9), 77(12.9), 11(10.4), 13(10.3)	0.975 (0.913-0.997)	1.000	0.947	0.950	0.954	1.000	0.950 (0.751-0.999)	1.000	0.917	0.900	0.889	1.000
VI	2	0.961	24	VOMs: 126(100.0), 79(90.4), 109(64.6), 6(59.5), 42(56.2), 121(55.0), 8(54.8), 63(45.0), 101(43.0), 52(36.9), 125(36.8), 98(34.0), 44(31.8), 56(28.7), 127(26.6), 99(22.6), 39(18.8), 61(11.2), 84(7.4), 83(6.8)	0.962 (0.894-0.992)	0.976	0.949	0.925	0.952	0.974	0.900 (0.683-0.988)	1.000	0.818	0.802	0.818	1.000

GA-PLS models					Training set (80samples)						Test set (20 samples)					
[a]Chrom Nº	[b]Nº of comp	[c]ROC	[d]Nº of var	[e]Variable Importance, N (%)	Accuracy (95% CI)	[f]Sens	[g]Spec	Kappa value	[h]H. Pred Value	[i]CA. Pred Value	Accuracy (95% CI)	Sens	Spec	Kappa value	H. Pred Value	CA. Pred Value
VII	3	0.963	24	VOMs: 126(100.0), 79(80.6), 109(68.4), 8(54.6), 95(54.3), 6(51.1), 9(47.8), 57(45.3), 21(43.2), 125(41.0), 72(37.8), 98(32.8), 5(29.7), 61(23.2), 39(22.5), 11(18.3), 110(17.6), 83(16.5), 107(15.2), 77(14.1)	0.975 (0.913-0.997)	1.000	0.952	0.950	0.950	1.000	0.950 (0.751-0.999)	0.917	1.000	0.900	1.000	0.889
VIII	1	0.940	24	VOMs: 126(100.0), 79(81.0), 109(77.9), 6(74.7), 121(57.8), 8(54.9), 28(53.9), 42(53.3), 125(47.7), 59(46.2), 63(45.7), 98(37.8), 21(37.7), 57(36.6), 72(33.3), 74(28.9), 67(19.4), 83(16.7), 77(12.3), 11(12.1)	0.962 (0.894-0.992)	1.000	0.925	0.925	0.930	1.000	0.900 (0.683-0.988)	1.000	0.800	0.800	0.833	1.000
IX	2	0.989	25	VOMs: 58(100.0), 22(96.4), 126(95.1), 109(72.5), 121(57.3), 6(46.1), 125(45.9), 21(44.8), 98(39.1), 25(36.1), 57(35.8), 44(33.9), 30(33.8), 100(32.6), 72(26.0), 74(25.6), 110(23.1), 112(17.5), 61(16.5), 39(12.6)	0.966 (0.907-0.999)	0.975	0.955	0.921	0.929	0.985	0.926 (0.757-0.991)	0.900	0.941	0.841	0.900	0.941
X	4	0.971	20	VOMs: 124(100.0), 89(70.2), 109(62.4), 6(60.3), 121(57.5), 10(55.5), 87(55.4), 62(50.0), 21(45.0), 63(41.6), 125(39.4), 98(36.6), 57(36.3), 44(27.6), 120(24.7), 48(21.8), 72(19.8), 39(13.0), 91(11.8), 11(0)	0.962 (0.894-0.992)	0.976	0.949	0.925	0.952	0.974	0.900 (0.683-0.988)	0.889	0.909	0.798	0.889	0.909
XI	1	0.962	16	VOMs: 126(100.0), 79(76.3), 58(76.1), 109(59.3), 121(58.8), 45(54.6), 6(42.0), 8(38.3), 63(35.3), 125(34.4), 98(31.0), 74(25.3), 21(23.5), 72(15.9), 39(14.6), 11(0).	0.912 (0.828-0.964)	0.927	0.897	0.825	0.905	0.921	0.900 (0.683-0.988)	1.000	0.818	0.802	0.818	1.000
XII	1	0.960	7	VOMs: 126(100), 58(83.6), 109(56.3), 121(51.9), 6(24.4), 125(15.0), 21(0).	0.912 (0.828-0.964)	0.974	0.854	0.825	0.864	0.972	0.900 (0.683-0.988)	0.909	0.889	0.798	0.909	0.889

[a] The chromosome number of the GA algorithm; [b] The optimal number of PLS components; [c] Receiver operating characteristic; [d] The number of variables in the chromosome; [e] Top 20 variable Importance of the chromosome; [f] Sensitivity; [g] Specificity; [h] Healthy predictive value; [i] Cancer predictive value.

Table S4. MultiROC curve-based model evaluation (tester) results of the 33 earwax samples collected longitudinally from 18 tumor-bearing dogs. The potential cancer biomarkers discrimination set accurately predicts the samples as cancer (Y: Cancer group).

Sample number of the animal in the dendrogram*	Sample identification⁑	Probability	Predict Class	True Class	Chemotherapy	Cancer Type
39	**C67**	0.9865	Y	Y	None	**STC**
	C68	0.9988	Y	Y	CTx	
	C69	0.9549	Y	Y	CTx	
	C70	0.9558	Y	Y	CTx	
32	**C57**	0.9770	Y	Y	CTx	**STC**
66	**C82**	0.9816	Y	Y	CTx	**BC**
	C83	0.9816	Y	Y	CTx	
	C84	0.9682	Y	Y	CTx	
18	**C36**	0.9775	Y	Y	CTx	**BC**
	C52	0.9677	Y	Y	CTx	
	C53	0.8981	Y	Y	CTx	
	C55	0.9677	Y	Y	CTx	
92	**C63**	0.9928	Y	Y	None	**TPC**
53	**C73**	0.9992	Y	Y	None	**BC**
	C74	0.8514	Y	Y	None	
	C75	0.9354	Y	Y	None	
	C76	0.8255	Y	Y	CTx	
5	**C77**	0.9964	Y	Y	CTx	**CC**

Sample number of the animal in the dendrogram*	Sample identification‡	Probability	Predict Class	True Class	Chemotherapy	Cancer Type
80	**C28**	0.9929	Y	Y	None	**BC**
95	**C79**	0.9957	Y	Y	None	**BC**
39	**C56**	0.9549	Y	Y	CTx	**STC**
11	**C26**	0.9937	Y	Y	CTx	**STC**
87	**C51**	0.9890	Y	Y	CTx	**BC**
71	**C72**	0.6680	Y	Y	CTx	**TVT**
58	**C78**	0.9732	Y	Y	CTx	**BC**
91	**C62**	0.9871	Y	Y	CTx	**TVT**
27	**C59**	0.9946	Y	Y	None	**TVT**
	C60	0.9770	Y	Y	None	
	C61	0.9924	Y	Y	CTx	
45	**C64**	0.9571	Y	Y	None	**BC**
	C65	0.9928	Y	Y	CTx	
	C66	0.9164	Y	Y	CTx	
63	**C81**	1.0000	Y	Y	CTx	**BC**

*Sample number regarding the original number of the animal in the circular dendrogram (Figure 2 of the manuscript).

‡Codes for each sample collected during longitudinal evaluation.