Analyst

Identification of cancer associated molecular changes in histologically benign vulval disease found in association with vulval squamous cell carcinoma using Fourier transform infrared spectroscopy

Journal:	*Analyst*
Manuscript ID	Draft
Article Type:	Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Frost, Jonathan; Gloucestershire Hospitals NHS Foundation Trust, Biophotonics Research Unit Ludeman, Linmarie; Cheltenham General Hospital, Department of Histopathology Hillaby, Kathryn; Cheltenham General Hospital, Department of Gynaecology Gornall, Rob; Cheltenham General Hospital, Department of Gynaecology Lloyd, Gavin; Gloucestershire Hospitals NHS Foundation Trust, Biophotonics Research Unit Kendall, Catherine; Gloucestershire Hospitals NHS Foundation Trust, Biophotonics Research Unit Shore, Angela; NIHR Clinical Research Facility, University of Exeter Medical School Stone, Nick; University of Exeter, College of Engineering, Mathematics and Physical Sciences

Gloucestershire Hospitals NHS
NHS Foundation Trust

Dr Jonathan Frost
Biophotonics Research Unit
Gloucestershire Royal Hospital
Great Western Road
Gloucester
GL1 3NN
j.frost@nhs.net
03004225473

The Editor
Analyst

Thursday, 8 September 2016

Dear Referee

Thank you for considering this manuscript entitled 'Identification of cancer associated molecular changes in histologically benign vulval disease found in association with vulval squamous cell carcinoma using Fourier transform infrared spectroscopy' for publication in the optical diagnosis themed issue of Analyst (SPEC 2016).

In this paper we examine a novel method for assessing the biomolecular changes which occur in the vulva that lead to the development of invasive cancer. Detecting these changes has the potential to give valuable prognostic information for women with precancerous conditions such as vulval intraepithelial neoplasia and lichen sclerosus as for those who have already developed SCC.

This study explores the role of Fourier transform infrared spectroscopy (FTIR-S) and multivariate analysis in the evaluation of molecular changes in HPV dependent and HPV independent vulval carcinogenesis. The molecular information obtained through FTIR-S has the potential to work as an adjunct to histopathology to aid in the early diagnosis of cancer and offers additional information about the molecular risk profile of the tissue.

We establish there are molecular changes detectable by FTIR-S in vulval intraepithelial neoplasia and lichen sclerosus that are associated with the development of vulval SCC.

Abstract

This study evaluates the capability of Fourier transform infrared spectroscopy (FTIR-S) in the differentiation of molecular changes in vulval intraepithelial neoplasia (VIN) and lichen sclerosus (LS) found in association with vulval squamous cell carcinoma (SCC), compared with VIN and LS found in isolation.
48 sections of vulval epithelium with features of VIN (n=24) or LS (n=24) underwent FTIR-S micro-spectroscopic mapping. Spectra from each section were correlated with the pathological diagnoses and the presence of concurrent SCC. Spectral variance was explored using principal component analysis and a multivariate linear discriminant classification model was developed and validated with leave one sample out cross validation.

The discriminant model was able to correctly identify FTIR-S spectra taken from samples of VIN and LS found in association with SCC from those found in isolation with a sensitivity of 82% and specificity of 93% for LS and sensitivity of 75% and specificity of 94% for VIN. The discriminant model was adjusted to maximise sensitivity whilst conceding specificity on a per patient basis and could differentiate LS associated with SCC with a sensitivity of 100% and specificity of 84% and VIN associated with SCC sensitivity of 100% and specificity 58%.
In distinguishing VIN and LS found in association with SCC from that found in isolation FTIR-S offers a potential technique for the assessment of molecular changes in the vulva that predispose to the development of SCC. Further study is needed to assess the ability of FTIR-S to risk stratify patients with VIN or LS.

Contribution of authors

Dr J Frost – Main author responsible for all aspects of the study
Dr L Ludeman – Histopathological opinion and advice
Miss K Hillaby – Clinical guidance and advice
Mr R Gornall – Clinical guidance and advice
Dr G Lloyd – Cooperated with data analysis
Dr C Kendall – Spectroscopic guidance and advice
Prof. A C Shore – Scientific guidance and advice
Prof. N Stone – Spectroscopic guidance and advice

Ethical considerations
Research ethics committee approval was gained for the use of archived fixed vulval tissue (East of Scotland Research Ethics Service 14/ES/1066).

Yours Faithfully

Dr Jonathan Frost
Clinical Research Fellow

Title

Identification of cancer associated molecular changes in histologically benign vulval disease found in association with vulval squamous cell carcinoma using Fourier transform infrared spectroscopy

Authorship

Jonathan Frost [a,b], Linmaire Ludeman [c], Kathryn Hillaby [c], Robert Gornall [c], Gavin Lloyd BSc PhD [b], Catherine Kendall [b], Angela C Shore [d], Nick Stone [a]

a. Biomedical Physics, School of Physics and Astronomy, University of Exeter, Exeter, UK, EX4 4QL

b. Biophotonics Research Unit, Gloucestershire Hospitals NHS Foundation Trust, Great Western Road, Gloucestershire, UK, GL1 3NN

c. Cheltenham General Hospital, Gloucestershire Hospitals NHS Foundation Trust, Sandford Road, Gloucestershire, UK, GL53 7AN

d. Medical School and NIHR Exeter Clinical Research Facility, University of Exeter, Exeter, UK, EX4 4QL

Corresponding author

Dr Jonathan Frost

Biophotonics Research Unit, Gloucestershire Hospitals NHS Foundation Trust

Great Western Road, Gloucestershire, GL1 3NN, UK

Email: j.frost@nhs.net Telephone: +443004225473

Fax: +443004225485

Abstract

This study evaluates the capability of Fourier transform infrared spectroscopy (FTIR-S) in the differentiation of molecular changes in vulval intraepithelial neoplasia (VIN) and lichen sclerosus (LS) found in association with vulval squamous cell carcinoma (SCC), compared with VIN and LS found in isolation. 48 sections of vulval epithelium with features of VIN (n=24) or LS (n=24) underwent FTIR-S micro-spectroscopic mapping. Spectra from each section were correlated with the pathological diagnoses and the presence of concurrent SCC. Spectral variance was explored using principal component analysis and a multivariate linear discriminant classification model was developed and validated with leave one sample out cross validation.

The discriminant model was able to correctly identify FTIR-S spectra taken from samples of VIN and LS found in association with SCC from those found in isolation with a sensitivity of 82% and specificity of 93% for LS and sensitivity of 75% and specificity of 94% for VIN. The discriminant model was adjusted to maximise sensitivity whilst conceding specificity on a per patient basis and could differentiate LS associated with SCC with a sensitivity of 100% and specificity of 84% and VIN associated with SCC sensitivity of 100% and specificity 58%. In distinguishing VIN and LS found in association with SCC from that found in isolation FTIR-S offers a potential technique for the assessment of molecular changes in the vulva that predispose to the development of SCC. Further study is needed to assess the ability of FTIR-S to risk stratify patients with VIN or LS.

Full Paper

Introduction

Vulval cancer is a rare gynaecological malignancy that is increasing in incidence in the United Kingdom and worldwide. In England the incidence has risen from 2 per 100,000 women in 1990 to 2.5 per 100,000 women in 2009.[1] Squamous cell carcinoma (SCC) is the predominant malignancy found in the vulva and represents around 95% of all vulval cancers. There are thought to be two distinct types of vulval SCC each with unique identifiable precursor disorders.[2] The first is linked with oncogenic subtypes of the human papilloma virus (HPV) and HPV associated vulval intraepithelial neoplasia (VIN). This HPV associated vulval cancer includes basaloid and warty SCC; mainly affects younger women and accounts for 20-35% of all vulval cancer. Keratinising SCC accounts for the remaining 65-80% of vulval cancer; is associated with lichen sclerosus and non HPV associated (differentiated) VIN and typically occurs in older women.[3] Each type of vulval SCC has a unique set of molecular changes that occur as vulval tissue is driven towards a malignant phenotype.[4] The ability to readily detect these molecular changes would be invaluable in the risk stratification of preneoplastic vulval disease and in the early diagnosis of vulval SCC.

HPV dependent carcinogenesis in the vulva

In vulval intraepithelial neoplasia (VIN) precancerous changes occur in the vulva. These precancerous changes can affect women of all ages although the peak incidence occurs in women less than 50 years old.[5,6] The primary

concern with VIN is the potential to progress to vulval squamous cell carcinoma. The risk of women with VIN developing vulval cancer is difficult to determine and the reported incidence of malignant transformation varies from 9% to 16% for untreated disease and 2% to 6% for those receiving preventative destructive or excisional management.[5,7–9] The diagnosis of VIN is clinical, supplemented with histological assessment. The International Society for the Study of Vulvovaginal Disease (ISSVD) 2004 consensus on terminology of VIN lesions defined two high-grade subgroups of VIN, usual type and differentiated type VIN.[10] Usual type VIN (uVIN) is associated with oncogenic human papilloma virus (HPV) infection, typically HPV 16 and can be further subdivided into warty, basaloid or mixed types.[11,12] Newer terminology has now been recommended by the ISSVD in line with other HPV related anogenital lesions, however this terminology has not yet been widely adopted.[13] Within the new terminology usual type VIN is renamed vulval high-grade squamous intraepithelial lesion (HSIL) and the terminology for HPV independent (differentiated) VIN remains unchanged.

uVIN (HSIL) is more common in younger women and is a precursor for the majority of basaloid and warty type vulval SCC.[14–16] The rising incidence of vulval cancer is primarily due to an increase in vulval cancer in women aged 70 and below, suggestive of an increase in HPV related disease.[1] This increase in HPV related disease may also account for the recent increase in the incidence of VIN.[17,18]

Malignant transformation of vulval epithelial cells by high-risk HPV subtypes is mediated through integration of the HPV DNA with in the host genome leading to HPV induced expression of the E6 and E7 oncoproteins. These

oncoproteins have numerous oncogenic effects including interference with control of the cell cycle, producing numeric and structural oncogenic chromosomal abnormalities.[19,20] In addition, methylation or mutations of the hosts' genome can occur that distorts transcriptional control, cell differentiation and viral gene expression. This results in altered molecular expression in tissues undergoing malignant transformation.[21]

HPV independent carcinogenesis in the vulva

The typical precursor vulval disorders in cases of HPV independent vulval SCC are Lichen Sclerosus (LS) and differentiated VIN. Lichen sclerosus (LS) is an inflammatory condition that can affect any skin site but is most commonly found in the vulva.[22] The true prevalence of LS is unknown, however community studies in the elderly have demonstrated a prevalence as high as 3%.[23] The aetiology of LS is unknown although there is some evidence to suggest that the lichen sclerosus is an autoimmune condition.[24] Vulval LS may occur at any age but typically either appears in prepubertal girls or post-menopausal women.[25] Diagnosis can be clinical although biopsy is necessary if the diagnosis is uncertain or if there is a failure to respond to adequate treatment.[26] This condition is complicated by development of SCC in 4% to 5% of women.[22,27] This compares to a background risk of vulval cancer in the UK of 0.3%.[28]

LS and differentiated VIN are thought to develop into SCC through a HPV independent process which is not fully understood.[2] Numerous genetic and epigenetic changes have been noted to occur in areas of LS associated with SCC that are not present in LS found in isolation. These include mutation of

TP53 and hypermethylation of MGMT and RASSF2A, suggesting a potential role for these genes in non HPV associated carcinogenesis however the exact mechanism is not well understood.[21,29]

Overall the pathogenesis of vulval SCC is not as well described as that for more common cancers, however it is clear that in both the HPV dependent and HPV independent oncogenesis biomolecular changes occur in the vulva that lead to the development of invasive cancer. Detecting these changes may give valuable prognostic information for women with precursor conditions such as LS and VIN as well as for those who have already developed SCC.

This study explores the role of Fourier transform infrared spectroscopy (FTIR-S) and multivariate analysis in the evaluation of molecular changes in HPV dependent and HPV independent vulval carcinogenesis. The current gold standard for the diagnosis of vulval conditions is histopathological examination. Traditional histological examination gives little information on molecular profile of the tissue being examined. The molecular information obtained through FTIR-S has the potential to work as an adjunct to histopathology to aid in the early diagnosis of cancer and may offer additional information about the molecular risk profile of the tissue.

The aim of this paper is to establish if there are molecular changes detectable by FTIR-S in HPV related (uVIN) and HPV independent (LS) vulval disease that are associated with the development of vulval SCC.

Experimental

Fourier transform infrared spectroscopy (FTIR-S)

FTIR-S is a non-destructive analytical technique that can be used to probe the molecular composition of tissues. When infrared radiation enters tissue, if it has sufficient energy to excite a molecular vibration mode, then it can cause vibrations of the molecular bonds within the tissue. This results in the absorption of the infrared radiation with the specific energy or wavelength to excite that particular bond. Thus the absorption of light is dependent on the molecular composition of the tissue. Measuring the absorption of infrared radiation across a range of wavelengths results in an absorption spectrum that represents a molecular fingerprint for the tissue under examination. FTIR-S combined with chemometric modelling has already been used to identify multiple different cancers from the analysis of tissue biopsies.[30,31] In addition, FTIR-S has been shown to have promise as an adjunct to routine histopathology in identification of clinically aggressive cancer subtypes.[32]

Sample preparation and FTIR spectral acquisition

Research ethics committee approval was gained for the use of archived fixed vulval tissue (East of Scotland Research Ethics Service 14/ES/1066). The pathology database at Gloucestershire Hospitals NHS Foundation Trust was interrogated and 24 cases of uVIN and 24 cases of LS with and without concurrent vulval SCC were selected. A sample size calculation was performed and with a conjectured accuracy of 0.85, a desired power of 0.9 and an equal number of cases to controls this number of cases was sufficient

to ensure a type one error rate of ≤0.05 and type 2 error rate of ≤0.1.[33] uVIN was chosen as the commonest identifiable HPV dependent precancerous vulval disease and LS was chosen as the commonest HPV independent preneoplastic vulval disease. The haematoxylin and eosin (H&E) stained slides from the cases identified were scrutinised by a gynaecology histopathologist and areas typical of uVIN and LS selected. Contiguous 4 micron tissue sections were cut from the corresponding paraffin embedded blocks onto a glass slide, a calcium fluoride substrate and a further glass slide. The sections cut onto glass were H&E stained and examined by a histopathologist blinded to the previous histological diagnosis to confirm the pathology present. The sections cut onto to the calcium fluoride substrates underwent FTIR spectroscopic mapping using a Perkin Elmer Spectrum One Spotlight 400 imaging system in transmission mode. The tissue sections were mapped with a step size (image pixel) of 6.25 microns and a spectral resolution of 4 cm^{-1}. The spectrum collected from each point on the spectral map was a mean of eight scans to reduce the noise in the spectrum. The number of spectra collected from each sample was determined by the size of the lesion identified on the tissue sections. In total tissue sections from 48 different patients were analysed (Figure 1).

Data analysis

The data analysis was carried out using Matlab (2014b) Mathworks USA. Data collected for the LS group (1) and the uVIN group (2) were analysed independently (Figure 1). The measured spectra were converted into absorbance and data points outside of the 800 to 1800 cm^{-1} fingerprint range were excluded. The remaining spectra were then corrected for the paraffin content of the tissue using an extended multiplicative scatter correction

(EMSC) to remove the effect of the variation in the paraffin in the tissue samples.[34,35] The spectra were then vector normalized.

In order to select the areas of epidermis within the spectral maps (Figure 2) a two stage k-means cluster analysis was performed. In the first stage individual spectral points that contained a high density of paraffin were identified using k-means clustering seeded with spectra from pure paraffin. The spectra identified were then excluded from further analysis. A second round of supervised k-means cluster analysis seeded with manually selected epidermal and dermal spectra was used to identify the epidermis within the spectral maps. The areas of epidermis identified were then isolated for further analysis and the non-epidermal spectra excluded. These areas were selected as the bulk of molecular changes that result in malignant transformation occur within the cellular epidermis. Principal component analysis (PCA) was applied to reduce the dimensionality of the data into loadings and corresponding scores in preparation for linear discriminant analysis (LDA). An analysis of variance (ANOVA) was performed on the principal component (PC) scores to determine which of the PC scores were significantly different between those tissue samples from women with concurrent SCC (groups 1a and 2a) and those without (groups 1b and 2b). The confidence level for the ANOVA was set at 95%. LDA was applied to the significant principal component scores and their associated sample groups (Figure 1). The resultant LDA model was then used to predict whether each spectrum was taken from a tissue sample from a woman with concurrent vulval SCC.

The classification ability of FTIR-S to detect whether a tissue sample of either LS or uVIN was from a patient with concurrent SCC was evaluated using leave one sample out cross validation. In this validation cycle the data from a

whole tissue sample (test set) was excluded from the initial analysis (PCA, ANOVA and LDA) that was performed on the remaining data (training set). The test set were then classified according to the LDA model produced with the training dataset. This process was repeated excluding each sample in turn from the initial analysis. The predicted groups from each cycle of validation were then collated for analysis. The specificity and sensitivity of the technique for detecting tissue associated with a SCC was calculated across a range of diagnostic thresholds (i.e. the proportion of spectra classified as being associated with SCC for the whole sample to be classified as being associated with SCC). The receiver operator characteristic was used to evaluate the diagnostic performance of the technique. The probability of obtaining a false positive result was evaluated using the method described by Obuchowski et al.[33,36]

Results and discussion

The diagnostic models for both group 1 (uVIN) and group 2 (LS) were each derived from 24 tissue samples. The total number of FTIR spectra included in each group was 509,000 for group 1 and 65,000 for group 2. The two stage k-means clustering analysis was able to correctly isolate the epidermal spectra in all but one case where manual selection of the epidermal spectra was successfully employed. The result of the initial PCA ANOVA LDA analysis demonstrated that the technique was able to classify the majority of spectra correctly using the linear discriminant scores (Figure 3).

Defining the biochemical constituent peaks responsible for discrimination within the LDA model is difficult. FTIR-S does not allow simple biochemical

quantification of target tissues due to the compound effect on the absorption spectra from the numerous molecular constituents of the tissue. Examining the two most significant principal component loadings and the composite linear discriminant loading we can speculate on the molecular basis for classification within the LDA models produced (Table 1, Figure 4). The spectral differences between those with concurrent SCC and those without appear to occur primarily in the amide I, amide II and amide III regions of the absorbance spectra. In addition changes are seen in lipid, fatty acid, nucleic acid and carbohydrate regions (Figure 4). These findings are suggestive of a difference in protein expression as uVIN and LS progresses towards malignant transformation.

The ability of FTIR-S combined with multivariate analysis to identify tissue samples from women with SCC was evaluated using leave one sample out cross validation. After validation the discriminant model demonstrated FTIR-S was able to correctly differentiate spectra from uVIN associated with SCC with a sensitivity of 75% and specificity of 94%. Similarly FTIR-S was able to correctly differentiate spectra from LS associated with SCC with a sensitivity of 82% and specificity of 93%. Each tissue sample was then classified as being associated with SCC or found in isolation by the number of spectra the LDA model classified as being associated with SCC. The threshold number of spectra to classify the tissue sample was adjusted and the receiver operator characteristic was determined for both groups. This demonstrated the sensitivities and specificities that can be achieved using this technique per sample. LS associated with SCC can be differentiated from LS found in isolation with a sensitivity of 100% and a specificity of 84% with area under the curve (AUC) of 0.98 (false positive rate of <0.05). uVIN associated with

SCC can be differentiated from uVIN found in isolation with a sensitivity of 100% and a specificity of 58% with AUC of 0.87 (false positive rate of <0.05). The specificity of the technique can be improved by conceding sensitivity (Figure 5). In the uVIN group a higher threshold number of spectra to classify the samples can be used to give a sensitivity of 75% and a specificity of 83%.

This preliminary study has demonstrated FTIR-S is able distinguish between LS or uVIN found in isolation and LS or uVIN found with concurrent SCC at another anatomical location on the vulva. Discrimination between these groups is based on the molecular variations detected by FTIR-S and multivariate analysis. The ability to separate the groups according to the presence of concurrent SCC is suggestive of cancer associated molecular changes occurring across the vulva. The detection of these molecular changes supports the concept of preneoplastic field cancerisation in the vulva.[4,37,38] In the development of a preneoplastic field of cancerisation a stem cell acquires a genetic or epigenetic alteration that gives it a growth advantage over its neighbouring cells. Replication of the stem cell results in an expanding patch of altered daughter cells. With additional genetic or epigenetic alterations this patch replaces the normal epithelium and develops into a field of cancerisation. Within the abnormal preneoplastic field the epithelium may appear histologically benign however molecular alterations within cells result in an increased risk of malignant transformation. The development of preneoplastic field change can be used to explain the multiple primary tumours and distant tumour recurrences that are often observed in the vulva. The molecular basis for field cancerisation in the vulva has previously been explored by examining the distribution of X-chromosome inactivation, TP53 mutations and viral integration sites.[19,38,39] The data

published demonstrated molecular monoclonality of anatomically distant vulval lesions supporting the concept of preneoplastic field change in the vulva. These studies of specific molecular changes in the vulva only show us part of the picture as the transformation of normal cells to cancer is a complex multistep process involving more that just a few molecular variations. In this study we have used FTIR-S to demonstrate the molecular field changes associated with the development of vulval cancer. FTIR-S allows the analysis of a broad range of molecular changes within the vulva. This technique not only examines molecular changes resulting from known provocations (e.g. TP53 mutation and HPV integration) but also assesses other undescribed molecular changes.[40] This gives a broader picture of the molecular changes present, but at the expense of detailed biomolecular causation as FTIR-S does not allow us to determine the exact molecular differences underlying spectral differences identified.

Demonstrating the ability of FTIR-S to differentiate between vulval disease found with concurrent cancer and vulval disease found in isolation highlights the role FTIR-S could have as a tool for assessing preneoplastic molecular changes in the vulva.

Traditional histopathological analysis gives limited information about the biomolecular changes associated with an increased risk of malignant transformation. FTIR-S has the potential to be used to augment traditional histopathology giving pathologists additional molecular information from the tissue under examination. The FTIR spectroscopic analysis of preneoplastic fields and molecular risk stratification of patients with VIN or LS may facilitate the identification of those at high risk of developing vulval cancer. In addition,

FTIR-S has the potential to be used as a histopathological adjunct to improve the early diagnosis of SCC through the analysis of molecular changes that identify those in whom further investigation is likely to find an occult malignancy. This early diagnosis would be an advantage as when vulval SCC is diagnosed early, when the depth of invasion is less than 1.0mm (i.e. FIGO stage 1a) the risk of lymph node metastasis is very low.[41] The absence of lymph node metastasis is an important prognostic factor in vulval SCC, with reported five-year survival in those who are node negative ranges from 70-93% compared with 25-42% in node positive women.[42]

Conclusion

FTIR spectroscopy offers a technique for molecular assessment that is straightforward to apply to routinely prepared tissue; that has a relatively low cost; that only requires a small amount of tissue and that can be performed in a short amount of time.[30] These characteristics make FTIR-S suitable as an adjunct to routine histopathology where fast analysis of tissue samples is required. Previous research by Baker et al. demonstrated the potential of FTIR-S in evaluating the biopotiential of prostate cancer, however to our knowledge this is the first time FTIR-S has been used to assess the molecular changes associated with vulval cancer.[32]

FITR-S has the potential to provide a technique that detects the genesis of malignant disease rather than just the presence of cancer. Such a technique would be useful for early cancer detection and informing prognosis. This study highlights the potential of FTIR-S for the molecular risk stratification of women with LS or uVIN. Further work in the form of longitudinal studies is required to

determine if this preliminary work can be transformed into a technique that will improve outcomes for women.

Conflicts of interest / Sources of Support

All authors report a grant received from British Society for the Study of Vulval Disease which part funded the conduct of the study.

Acknowledgements

We acknowledge the British Society for the Study of Vulval Disease and the Gloucestershire Hospitals NHS Foundation Trust Research and Innovation Forum for financial support. We thank Jo Motte and the team at the Gloucestershire Hospitals NHS Foundation Trust histopathology department for the preparation of the tissue sections. The study was supported by the NIHR Exeter Clinical Research Facility. The interpretations of data in the paper are those of the authors and not of NIHR or the Department of Health.

References

1 J. Lai, R. Elleray, A. Nordin, L. Hirschowitz, B. Rous, C. Gildea and J. Poole, *BJOG An Int. J. Obstet. Gynaecol.*, 2014, **121**, 728–738.

2 Y. Ueda, T. Enomoto, T. Kimura, K. Yoshino, M. Fujita and T. Kimura, *J. Skin Cancer*, 2011, **2011**, 951250.

3 H. P. van de Nieuwenhof, I. a M. van der Avoort and J. a. de Hullu, *Crit. Rev. Oncol. Hematol.*, 2008, **68**, 131–156.

4 G. D. Dakubo, J. P. Jakupciak, M. a Birch-Machin and R. L. Parr, *Cancer Cell Int.*, 2007, **7**, 2.

5 R. W. Jones, D. M. Rowan and A. W. Stewart, *Obs. Gynecol*, 2005, **106**, 1319–1326.

6 M. Nygård, B. T. Hansen, J. Dillner, C. Munk, K. Oddsson, L. Tryggvadottir, M. Hortlund, K.-L. Liaw, E. J. Dasbach and S. K. Kjær, *PLoS One*, 2014, **9**, e88323.

7 M. Van Seters, M. Van Beurden and A. J. M. de Craen, *Gynecol. Oncol.*, 2005, **97**, 645–51.

8 J. J. Wallbillich, H. E. Rhodes, A. M. Milbourne, M. F. Munsell, M. Frumovitz, J. Brown, C. L. Trimble and K. M. Schmeler, *Gynecol. Oncol.*, 2012, **127**, 312–5.

9 M. K. Fehr, M. Baumann, M. Mueller, D. Fink, S. Heinzl, P. Imesch and K. Dedes, *J Gynecol Oncol*, 2013, **24**, 236–241.

10 M. Sideri, R. W. Jones, E. J. Wilkinson, M. Preti, D. S. Heller, J. Scurry, H. Haefner and S. Neill, *J. Reprod. Med.*, 2005, **50**, 807–10.

11 P. Lynch, M. Moyal-Barracco, F. Bogliatto, L. Micheletti and J. Scurry, 2014, **52**, 1–12.

12 S. de Sanjosé, L. Alemany, J. Ordi, S. Tous, M. Alejo, S. M. Bigby, E. A.

Joura, P. Maldonado, J. Laco, I. G. Bravo, A. Vidal, N. Guimerà, P. Cross, G. V. Wain, K. U. Petry, L. Mariani, C. Bergeron, V. Mandys, A. R. Sica, A. Félix, A. Usubutun, M. Seoud, G. Hernández-Suárez, A. M. Nowakowski, G. Wilson, V. Dalstein, M. Hampl, E. S. Kasamatsu, L. E. Lombardi, L. Tinoco, I. Alvarado-Cabrero, M. Perrotta, N. Bhatla, T. Agorastos, C. F. Lynch, M. T. Goodman, H.-R. Shin, H. Viarheichyk, R. Jach, M. O. L. E. Cruz, J. Velasco, C. Molina, J. Bornstein, A. Ferrera, E. J. Domingo, C.-Y. Chou, A. F. Banjo, X. Castellsagué, M. Pawlita, B. Lloveras, W. G. V. Quint, N. Muñoz and F. X. Bosch, *Eur. J. Cancer*, 2013, **49**, 3450–3461.

13 ISSVD, *The 2015 International Society for the Study of Vulvovaginal Disease Terminology of Vulvar Squamous Intraepithelial Lesions*, 2015.

14 I. a. M. Van der Avoort, H. Shirango, B. M. Hoevenaars, J. M. M. Grefte, J. a de Hullu, P. C. M. de Wilde, J. Bulten, W. J. G. Melchers and L. F. a. G. Massuger, *Int. J. Gynecol. Pathol.*, 2006, **25**, 22–29.

15 M. C. Reyes and K. Cooper, *J. Clin. Pathol.*, 2014, **67**, 290–294.

16 M. Hampl and H. Sarajuuri, 2006, **108**, 1361–1368.

17 T. Iversen and S. Tretli, *Obstet. Gynecol.*, 1998, **91**, 969–72.

18 E. A. Joura, A. Lösch, M. G. Haider-Angeler, G. Breitenecker and S. Leodolter, *J. Reprod. Med.*, 2000, **45**, 613–5.

19 S. Vinokurova, N. Wentzensen, J. Einenkel, R. Klaes, C. Ziegert, P. Melsheimer, H. Sartor, L. C. Horn, M. Höckel and M. von Knebel Doeberitz, *J. Natl. Cancer Inst.*, 2005, **97**, 1816–1821.

20 J. Zekan, M. Sirotkovic-Skerlev and M. Skerlev, in *DNA Replication-Current Advances*, InTech, 2011.

21 M. D. Trietsch, L. S. Nooij, K. N. Gaarenstroom and M. I. E. van Poelgeest, *Gynecol. Oncol.*, 2015, **136**, 143–157.

22 J. Powell and F. Wojnarowska, *Lancet*, 1999, **353**, 1777–1783.

23 A. Leibovitz, V. Kaplun, N. Saposhnicov and B. Habot, *Arch. Gerontol. Geriatr.*, 2000, **31**, 1–4.

24 R. H. Meyrick Thomas, C. M. Ridley, D. H. Mcgibbon and M. M. Black, *Br J Dermatol*, 1988, **118**, 41–46.

25 J. J. Meffert, B. M. Davis and R. E. Grimwood, *J. Am. Acad. Dermatol.*, 1995, **32**, 393-416–8.

26 S. Neill and F. Lewis, *Br J Dermatol*, 2010, 672–682.

27 J. a Carlson, R. Ambros, J. Malfetano, J. Ross, R. Grabowski, P. Lamb, H. Figge and M. C. Mihm, *Hum. Pathol.*, 1998, **29**, 932–48.

28 Cancer Research UK, 2016.

29 D. Guerrero, R. Guarch, A. Ojer, J. M. Casas, C. Méndez-Meca, M. Esteller, E. Barba-Ramos, F. Garcia-Bragado and A. Puras, *Int. J. Cancer*, 2011, **128**, 2853–64.

30 O. J. Old, L. M. Fullwood, R. Scott, G. R. Lloyd, L. M. Almond, N. A. Shepherd, N. Stone, H. Barr and C. Kendall, *Anal. Methods*, 2014, **6**, 3901.

31 C. Kendall, M. Isabelle, F. Bazant-Hegemark, J. Hutchings, L. Orr, J. Babrah, R. Baker and N. Stone, *Analyst*, 2009, **134**, 1029.

32 M. J. Baker, E. Gazi, M. D. Brown, J. H. Shanks, P. Gardner and N. W. Clarke, *Br. J. Cancer*, 2008, **99**, 1859–66.

33 N. A. Obuchowski, M. L. Lieber and F. H. Wians, *Clin. Chem.*, 2004, **50**, 1118–1125.

34 J. Nallala, G. R. Lloyd and N. Stone, *Analyst*, 2015, 2369–2375.

35 A. Tfayli, C. Gobinet, V. Vrabie, R. Huez, M. Manfait and O. Piot, *Appl. Spectrosc.*, 2009, **63**, 564–570.

36 N. A. Obuchowski and D. K. McClish, *Stat. Med.*, 1997, **16**, 1529–42.

37 B. J. M. Braakhuis, M. P. Tabor, J. A. Kummer and R. H. Brakenhoff, *Cancer Res*, 2003, **63**, 1727–1730.

38 A. N. Rosenthal, A. Ryan, D. Hopster and I. J. Jacobs, *Int. J. Cancer*, 2002, **99**, 549–554.

39 K. J. Rolfe, a B. MacLean, J. C. Crow, E. Benjamin, W. M. N. Reid and C. W. Perrett, *Br. J. Cancer*, 2003, **89**, 2249–53.

40 K. M. Ostrowska, A. Garcia, A. D. Meade, A. Malkin, I. Okewumi, J. J. O'Leary, C. Martin, H. J. Byrne and F. M. Lyng, *Analyst*, 2011, **136**, 1365–1373.

41 N. F. Hacker and J. Van der Velden, *Cancer*, 1993, **71**, 1673–1677.

42 J. M. Schilder and F. B. Stehman, in *Clinical Gynecologic Oncology: Eighth Edition*, 2012, pp. 219–244.

Tables and figures

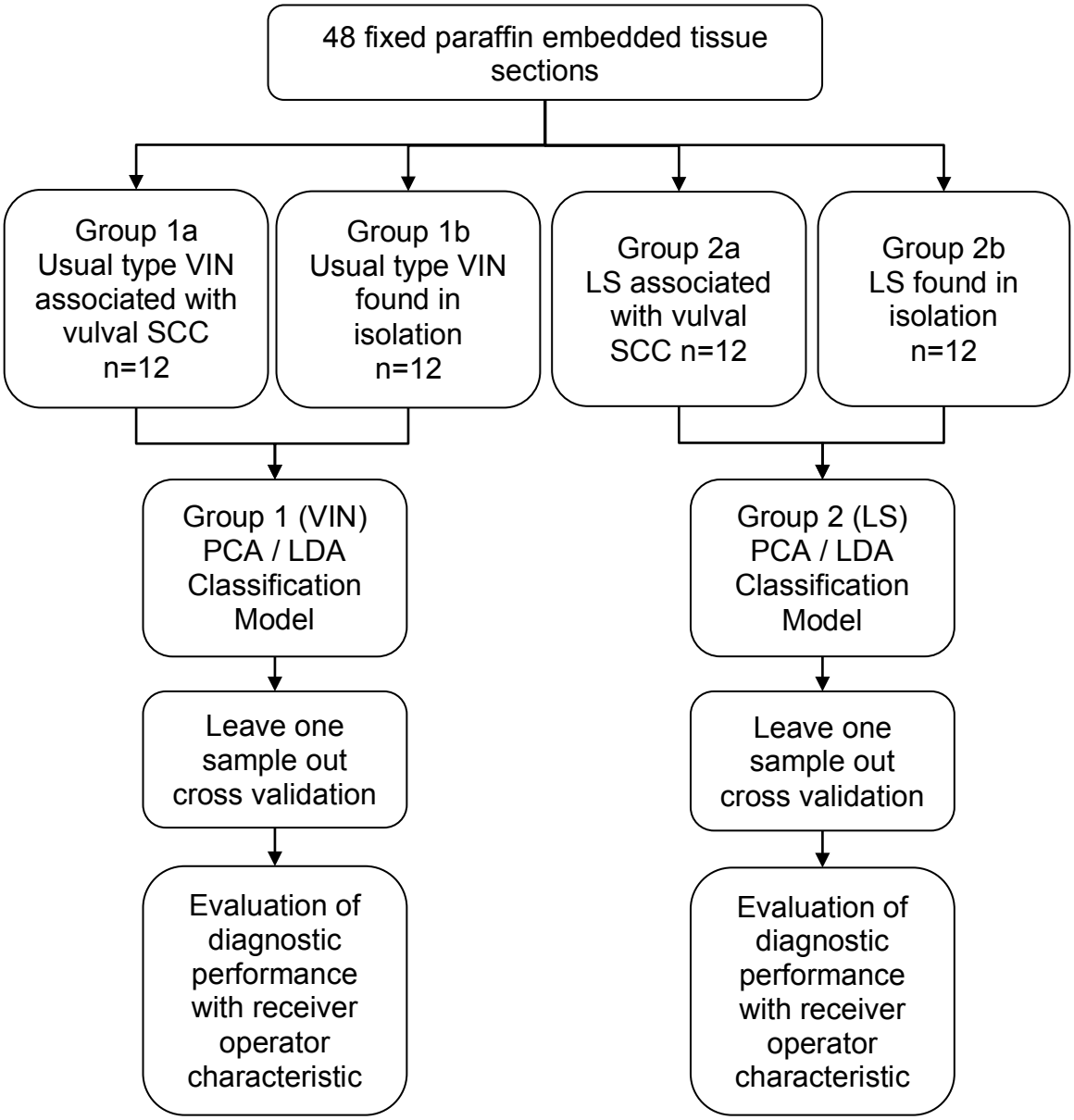

Figure 1, Study overview showing the four sample groups that underwent FTIR spectroscopic mapping and subsequent analysis (PCA - principal component analysis, LDA - Linear discriminant analysis)

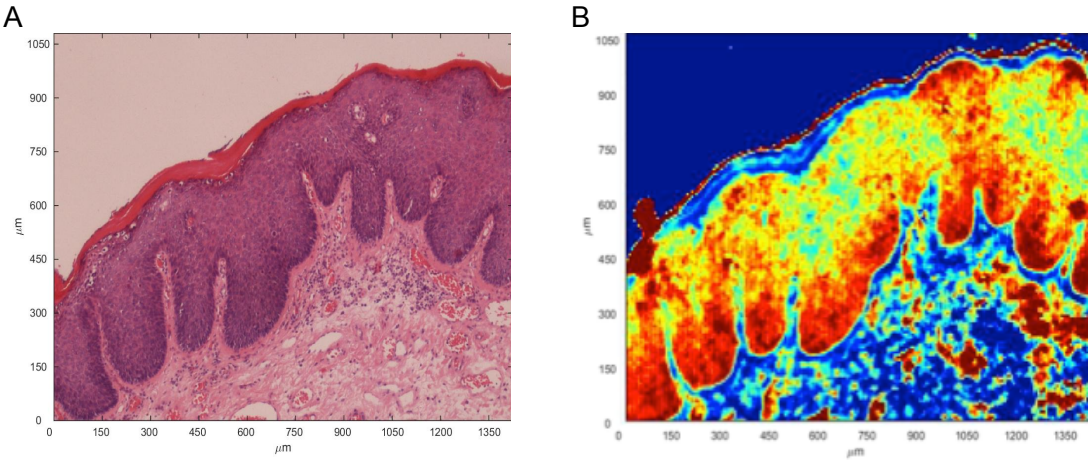

Figure 2, H&E stained tissue section (A) and corresponding spectral principal component scores plot (first principal component) (B)

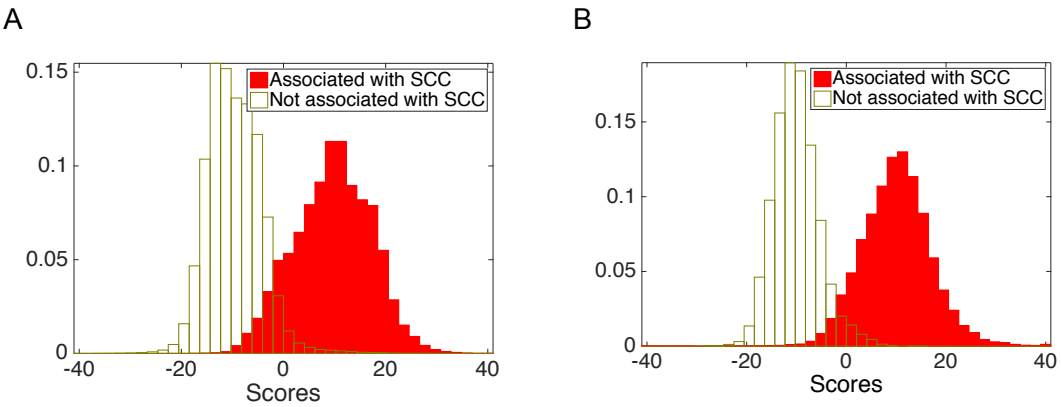

Figure 3, Histogram of score values from the linear discriminant function for individual spectra for the uVIN group (A) and the LS group (B).

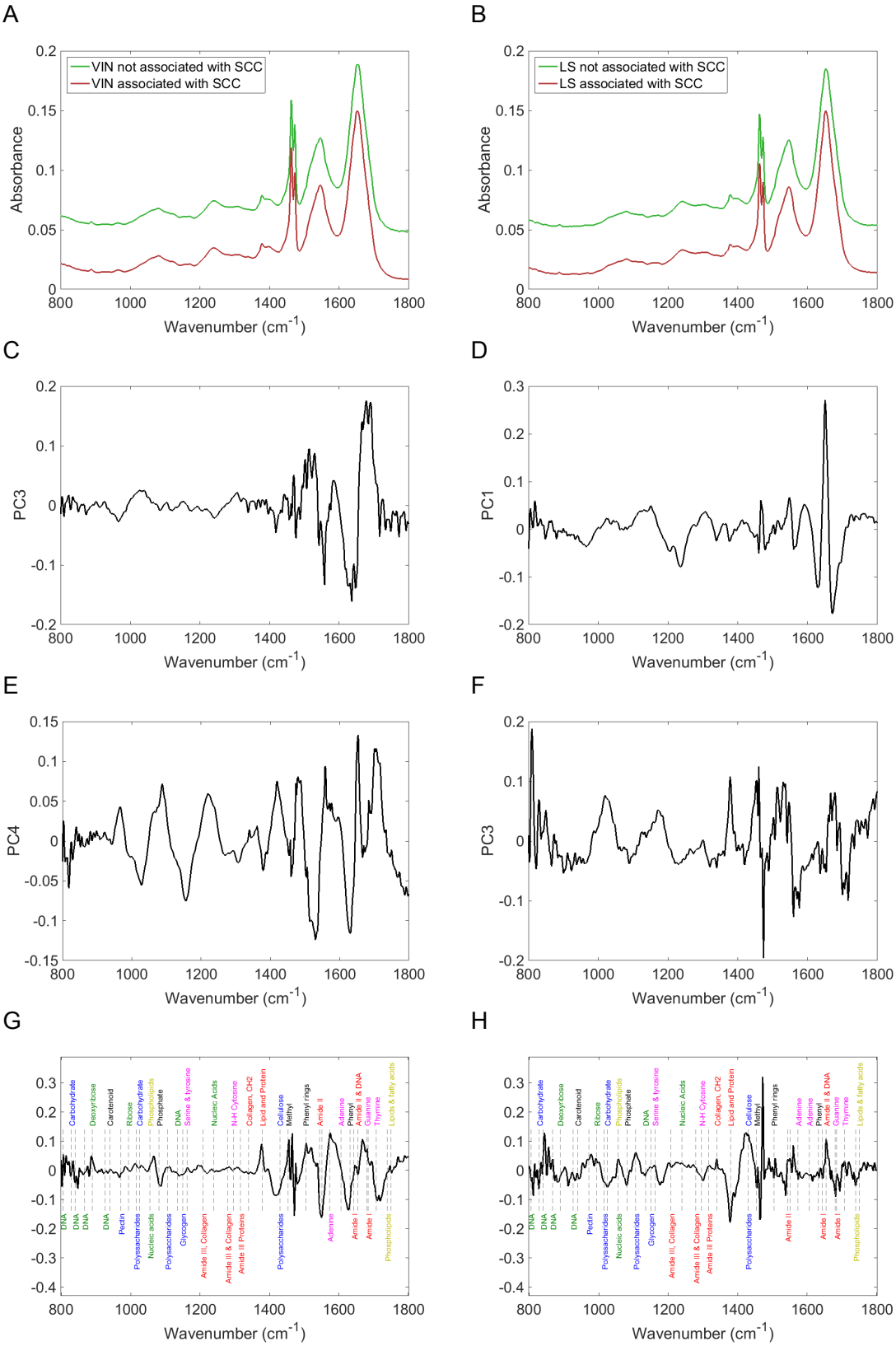

Figure 4, Mean spectra for the uVIN group (A) and the LS group (B) (offset to separate spectra), the two most significant principal components for the uVIN model (C and E) and for the LS model (D and F), and composite LDA loading for the uVIN group (G) and for the LS group (H)

Table 1, Proposed principal spectral changes responsible for discrimination between uVIN and LS found in association with SCC and that found in isolation with suggested biomolecular constituent assignments

Wavenumber (cm^{-1})	Suggested biomolecular constituent allocation
Group 1 (uVIN)	
966	C-O deoxyribose, C-C DNA
968	Left-handed helix DNA (Z form)
1000–1140	Protein amide I absorption
1022-4, 1028	Glycogen absorption
1040-100, 1025, 1107	Carbohydrates (including glucose, fructose, glycogen, etc.)
1056/7	Stretching C-O deoxyribose
1080	Collagen & phosphodiester groups of nucleic acids
1117	C-O stretching vibration of C-OH group of ribose (RNA)
1172	Serine, threonine, and tyrosine residues of cellular proteins
1204, 1339	Collagen, proteins-amide III and polysaccharides
1235	Composed of amide III as well as nucleic acids
1236	Amide III and nucleic acids
1236-42	Collagen and nucleic acids
1307-17	Amide III band components of proteins
1390	Carbon particle
1451, 1455-6	Methyl groups of proteins
1480-600, 1540, 1549	Amide II regions
1545	Protein band
1600-720, 1630–700	Amide I regions
1730, 1743	Fatty acid esters and lipids
Group 2 (LS)	
968	DNA
985	polysaccharides-cellulose
1018	polysaccharides, pectin
1000–140, 1600-720, 1630–700, 1656, 1670, 1717	Protein Amide I absorption
1030	Glycogen vibration, Collagen & phosphodiester groups of nucleic acids
1040-100	Carbohydrates (including glucose, fructose, glycogen, etc.)
1066, 1220, 1235	Nucleic acids
1084–6	Phosphate of nucleic acids
1164	Serine, threosine, & tyrosine of proteins
1200, 1284, 1339	Collagen
1235, 1284, 1307-17	Amide III
1396, 1455-6	Methyl groups of proteins
1419	Polysaccharides, pectin
1470	Methylene chains in lipids
1504	Phenyl rings
1545	Protein band
1480-600, 1517, 1540, 1549	Amide II regions
1670, 1750	lipids, fatty acids

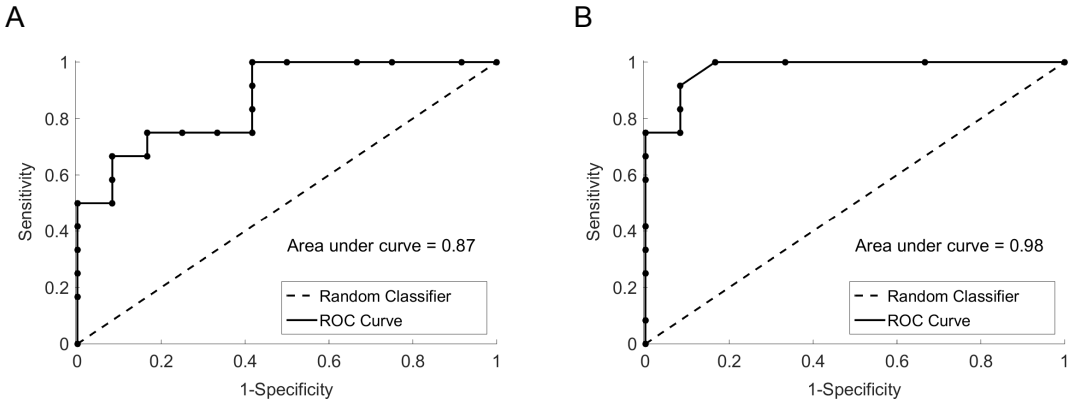

Figure 5, Receiver operating characteristic for the identification of tissue associated with SCC in the uVIN group (A) and the LS group (B)