Classification of ultraviolet irradiated mouse skin histological stages by bimodal spectroscopy: multiple excitation autofluorescence and diffuse reflectance

Marine Amouroux
Gilberto Díaz-Ayil
Walter C. P. M. Blondel
Nancy-University
Centre de Recherche en Automatique de Nancy
CRAN UMR 7039 CNRS
2 avenue de la Forêt de Haye
54516 Vandoeuvre-les-Nancy cedex
France

Geneviève Bourg-Heckly
Université Pierre et Marie Curie Paris 6
Laboratoire de Biophysique Moléculaire, Cellulaire et
    Tissulaire
BioMoCeTi UMR 7033 CNRS
Genopole Campus 1, 5 rue Henri Desbruères
91030 Evry cedex
France

Agnès Leroux
Alexis Vautrin Cancer Center
Avenue de Bourgogne
54511 Vandoeuvre-les-Nancy, cedex
France

François Guillemin
Nancy-University
Centre de Recherche en Automatique de Nancy
    and
Alexis Vautrin Cancer Center
Avenue de Bourgogne
54511 Vandoeuvre-les-Nancy, cedex
France

Abstract. Histopathological analysis and *in vivo* optical spectroscopy were used to discriminate several histological stages of UV-irradiated mouse skin. At different times throughout the 30-week irradiation, autofluorescence (AF) and diffuse reflectance (DR) spectra were acquired in a bimodal approach. Then skin was sampled and processed to be classified, according to morphological criteria, into four histological categories: normal, and three types of hyperplasia (compensatory, atypical, and dysplastic). After extracting spectral characteristics, principal component analysis (data reduction) and the *k*-nearest neighbor classifying method were applied to compare diagnostic performances of monoexcitation AF (based on each of the seven excitation wavelengths: 360, 368, 390, 400, 410, 420, and 430 nm), multiexcitation AF (combining the seven excitation wavelengths), DR, and bimodal spectroscopies. Visible wavelengths are the most sensitive ones to discriminate compensatory from precancerous (atypical and dysplastic) states. Multiexcitation AF provides an average 6-percentage-point increased sensitivity compared to the best scores obtained with monoexcitation AF for all pairs of tissue categories. Bimodality results in a 4-percentage-point increase of specificity when discriminating the three types of hyperplasia. Thus, bimodal spectroscopy appears to be a promising tool to discriminate benign from precancerous stages; clinical investigations should be carried out to confirm these results. © *2009 Society of Photo-Optical Instrumentation Engineers.* [DOI: 10.1117/1.3077194]

1 Introduction

Skin cancer is the most prevalent cancer among fair-skinned populations.[1] Two main types of skin cancerous lesions originate from cells found in the epidermis (the skin epithelium): keratinocytic tumors and melanocytic tumors.[2] The former type of lesion is the most common one, and among this type, squamous cell carcinoma (SCC) is the most dangerous cancer because it metastasizes in up to 7.4% of cases.[3] The best chance to fully cure SCC is to excise the tumor during its first stages of development before dermis invasion, where the vascularization network starts. The so-called "first stages" include precancerous (e.g., actinic keratosis) and *in situ* (e.g., Bowen's disease) stages. Thus, detecting these early stages is of paramount interest for clinical diagnosis. Ideally, this detection would be as specific and sensitive as possible. A survey of general practioners found that 30% had no confidence in correctly triaging skin lesions; 22% of cancers would not have been biopsied or referred, while benign lesions would have been wrongly biopsied or referred 34% of the time.[4] False-negative diagnoses may result in further growth of a lesion, making it too extensive for surgical resection. A delayed diagnosis may even allow for metastasis to occur, possibly with fatal consequences. False-positive diagnoses result in unnecessary procedures, patient inconvenience and stress, increased morbidity, and substantial additional costs.[4]

Several diagnostic techniques are currently under study (biopsy techniques, molecular markers, ultrasonography, confocal microscopy, spectroscopy, fluorescence imaging, electrical impedance, etc.), but the noninvasive techniques are particularly well adapted to everyday clinical practice since they give the clinician a chance to follow the evolution of lesions and give immediate information. Among them, dermoscopy is

Address all correspondence to: Walter C. P. M. Blondel, PhD, Centre de Recherche en Automatique de Nancy, CRAN UMR 7039 CNRS-Nancy University, 2 avenue de la Forêt de Haye, 54516 Vandoeuvre-les-Nancy, France. Tel: +33383595638; Fax: +33383595644; E-mail: walter.blondel@ensem.inpl-nancy.fr.

the most widespread technique in everyday practice. In dermoscopy, the lesion is examined with a $10\times$ to $100\times$ magnification lens placed directly against skin to which immersion oil has been applied for index matching. However, dermoscopy of keratinocytic lesions is still in its infancy, because no consensus regarding diagnostic criteria has been established.[1] Furthermore, such a technique is still based on visual appreciation, and therefore depends on the physician's personnal experience. That is why *in vivo* optical spectroscopy (also called "optical biopsy") is a potentially helpful technique because it may provide physicians with quantitative, i.e., more objective, criteria. To assess spectroscopy's ability to detect *in vivo* precancerous skin stages, the entire carcinogenesis in an animal model must be followed because it is the only way to obtain tissue sampling from each carcinogenesis stage. Two ways of inducing carcinogenesis have been widely described: chemically induced (using dimethyl benz(a)anthracene–tetradecanoylphorbol-13-acetate DMBA-TPA application)[5] and UV-induced[6] carcinogenesis of mouse skin. However, Pentland et al. showed that models to test chemoprevention strategies must specifically consider cancer etiology in their design.[7] The main etiological factor for SCC is UV solar (often referred to as "actinic") radiation.[8] Therefore, we chose to study skin neoplastic changes of UV-irradiated SKH1 (hairless and albinos) mice throughout the entire carcinogenesis.

UV radiation is a complete carcinogen, i.e., it is responsible for cancer induction, promotion, and progression. The typical UV-induced DNA damage is the generation of dimeric photoproducts between adjacent pyrimidine bases. Tumor suppressor gene p53 is a common target of UV-induced mutations. A proliferative advantage of p53 mutant keratinocytes over normal keratinocytes results in neoplastic transformation.[9] DNA damage also arises spontaneously from endogenous chemicals produced during metabolism, oxidative stress, and chronic inflammation.[10] In preclinical studies, exposure of murine skin to UV radiation for 30 weeks aimed to first induce an acute inflammation, followed by an unresolved or "chronic" inflammation that resulted in continuous exposure of tissue to activated immune cells. Such cells secrete soluble mediators that result in excessive tissue remodeling, loss of tissue architecture, and enhanced risk for development of neoplastic tissue states.[11] Most importantly, the recognition that a given tissue abnormality is associated with cancer development does not necessarily imply that all cancers at that site will develop from such lesions. Indeed, validation of cancer precursors as intermediate endpoints for cancer development is challenging, since tissue sampling necessarily interrupts or alters the natural course of the neoplastic process. In such cases, spectroscopy also can be a useful tool of investigation.

In the current study, autofluorescence (AF) spectroscopy and diffuse reflectance (DR) spectroscopy, alone and in combination, were investigated as noninvasive techniques to discriminate different stages of UV-irradiated mouse skin. The diagnostic efficiency of each technique has been studied separately on a wide variety of organs, but Chang et al.,[12] Breslin et al.,[13] and Georgakoudi et al.[14] showed that combining the two techniques in a "bimodal" approach improved the diagnostic accuracy of cervical, breast, and esophagus cancer, respectively. To the best of our knowledge, only one study has been published on both spectroscopic and histological evaluations of mouse skin carcinogenesis; the main result of this study was that AF spectroscopy using several excitation wavelengths discriminated early stages of chemically induced SCC with improved classification accuracy compared to single-wavelength excitation.

The aim of our study was twofold: (1) establish histopathological criteria to classify different stages of UV-irradiated murine skin: compensatory, atypical, and dysplastic hyperplasia; and (2) compare the accuracy (sensitivity and specificity) of monoexcitation AF, multiexcitation AF, DR, and bimodal spectroscopies in discriminating pairs of histological classes thanks to a reduced set of selected spectral characteristics.

2 Materials and Methods

2.1 Tumor Model

This experiment was approved by the French Ethical Committee on Animal Experimentation. Twelve-week-old SKH1 female mice (Charles River Laboratories, France) were divided into two groups. The first group included 8 sham-irradiated mice (control group), and the second group included 20 UV-irradiated mice. UV irradiation was provided by a bank of eight fluorescent tubes (TL/12, Philips, The Netherlands) set into a UV research unit (Daavlin, Belgium) emitting mainly in the UVB range (50% of the emission in the 290 to 320-nm spectral range). Mice were irradiated with a fluence of 3 mW/cm^2, 23 cm away from the lamps, for 20 seconds each time. The dose was controlled using an integrated dosimeter. The mice were kept in a small cage placed each time at the same place underneath the UV research unit in order to deliver as homogenous a dose as possible to all the mice.

Irradiation was performed once a week during the first 15 weeks, then twice a week during the last 15 weeks. Irradiation was always performed at least two days before spectroscopic measurements to avoid an acute reaction bias of spectroscopic measurements. Ten mice were euthanized after one month of irradiation and ten others after six months of irradiation to sample acute and chronic UV-induced skin modifications, respectively.

2.2 Experimental Protocol

The mice were anesthesized with a mix of ketamine (110 mg/kg, Panpharma, France) and xylazine (11 mg/kg, Rompun, Bayer, France) for a 75-minute immobilization. They were placed on a 35 °C heated blanket underneath the optical fiber probe that was fixed on a motorized stage to control the probe's angle (90 deg) and pressure repeatability on the skin [Fig. 1(a)]. The probe was put in gentle contact with the skin. An experimental protocol was developed to ensure that spectroscopic measurements and histological evaluations were obtained from the same tissue site. Since tattoo ink is a light absorber and the needles used to tattoo the skin could modify the skin's microscopic morphology, a plastic pattern was used to localize the measurement sites. As shown in Figs. 1(a) and 1(b), two rows of six holes each were perforated into the pattern; then the pattern was adhered to each mouse's back to ensure that spectroscopic measurements were performed every 5 mm in a row. Once spectroscopic measurements were done, the beginning of each six-point row was tattooed to mark the beginning of length measurements on histological slides [Figs. 1(b) and 1(c)]. Within six hours of
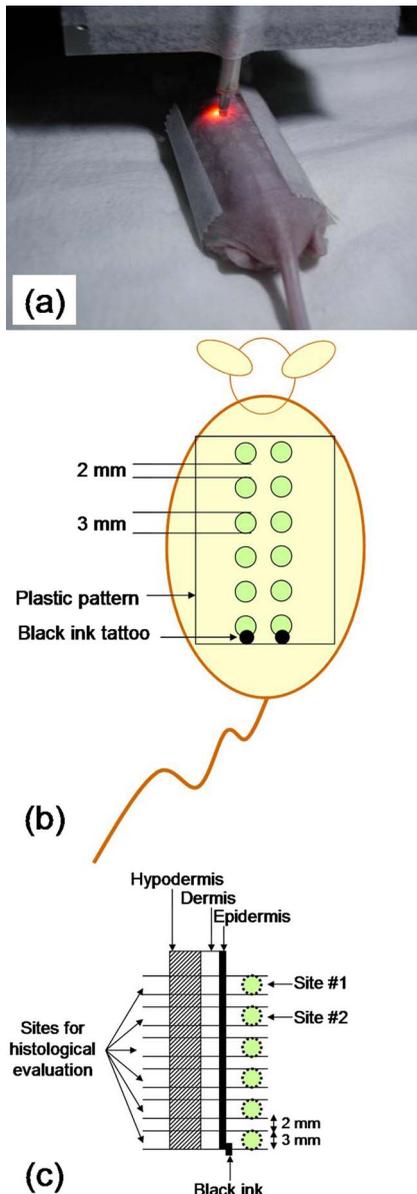
Fig. 1 (a) An anesthesized mouse on the heated blanket during spectroscopic measurement. The optical fiber probe was put in gentle contact with the skin through a plastic pattern stuck on the mouse's back to localize spectroscopic evaluation spots. (b) Schematic of the plastic pattern dimensions and black ink tattoo used to localize each 6-spot row along the backbone. (c) Schematic showing the method used to localize histological evaluation sites.

spectroscopic measurements, the mice were euthanized by performing cervical dislocation. Within 10 minutes of death, the back skin was excised. To prevent the skin from shrinking, each piece of skin was held on a polystyrene lump with needles. The skin was further fixed in a formaldehyde solution before embedding in paraffin. Then 5-μm slices were stained with standard hematoxylin and eosin (H&E) for histological classification based on epidermis morphological features. Dermal elastic fibers were also stained using Weigert specific staining.

2.3 Histological Classification

Hyperplasia is defined as an increase in the size of a tissue or organ due to an increase in the number of cells. It may be physiologic, compensatory, or pathologic.[10] We performed a two-step classification. The first step to discriminate a hyperplastic from a nonhyperplastic skin epidermis. To do so, we needed to first determine a normal (N) epidermis thickness range. Since skin does not display the same thickness on the whole body surface, we determined a normal thickness range for each spot number, i.e., for each anatomical location from a mouse's neck to tail along the backbone. The epidermis thickness of spot #y ($y=1$, 2, 3, 4, 5, or 6) on UV-irradiated mice was then compared to the epidermis thickness of the corresponding spot #y on sham-irradiated mice, thus taking into account physiological variations. Thickness measurements were performed on $10\times$- to $100\times$-magnified images acquired via a microscope (Axioskop, Zeiss, Germany) equipped with a CCD camera (DXC-390P, Sony, Japan). Image acquisition software (Tribvn ICS, Tribvn, France) and data processing software (Osiris, Digital Imaging Unit of the Geneva Hospital, Switzerland) were used. A Gaussian distribution of the data set was checked, then the normal range was defined according to what is commonly recommended for data having a Gaussian distribution: mean $\pm 2 \times$ standard deviation (SD).[15] The epidermis of a specific spot was classified as hyperplastic if its thickness was above the "mean$+2 \times$SD" cutoff value.

Since atypical and dysplastic hyperplasias (AH and D) are true cancer precursors, whereas compensatory hyperplasia (CH) is not,[10] discriminating CH from AH and D is of utmost medical value. That is why the second step of our classification method aimed at classifying hyperplastic skin samples into the CH, AH, and D histological classes. Atypia is defined as an abnormal cellular proliferation in which there is a loss of normal architecture and orientation.[10] Histological features of skin dysplasia are given by Kumar et al.:[16] because skin dysplasia is usually the result of chronic exposure to sunlight and is associated with the buildup of excess keratin, these lesions are called actinic keratoses. The dermis contains thickened blue-gray elastic fibers (elastosis), a probable result of abnormal dermal elastic fiber synthesis by sun-damaged fibroblasts within the superficial dermis. The stratum corneum is thickened and, unlike normal skin, nuclei in the cells are often retained (a pattern called "parakeratosis"). Table 1 displays the morphological features on which we based our classification of hyperplastic samples in one of the three classes. All the criteria were evaluated by a pathologist. Three fields of view (FOV) were examined from each sample. A sample was assigned to a class only if classifications of the three FOV were consistent with one another. If they were not consistent, then the sample was discarded from histological classification.

2.4 Spectroscopy and Statistical Analysis Methods

Figure 2 shows the spectroscopy instrumentation developed to acquire spatially colocalized AF and DR spectra. A fiber probe made of 200-μm core diameter optical fibers was used. One of these fibers was chosen as the excitation fiber and another was chosen to collect AF and backscattered light at a 271-μm collecting to excitation fiber separation (CEFS). The spectrometer was an iHR320 spectrograph (Horiba Jobin Yvon,

Table 1 Morphological features used to classify skin samples into the three different classes of hyperplasia (CH, AH, and D). ⊕: feature that must be found to classify the sample into the corresponding histological class, ⊗: additional features that may be found as well. Sham-irradiated mice were considered normal (N) and used as negative cortrols. Histological features displayed by SCC were used as positive controls.

	Normal (N)	Compensatory hyperplasia (CH)	Atypical hyperplasia (AH)	Dysplasia (D)	Bowen's disease/SCC
Thickened epidermis		⊕	⊕	⊕	⊕
Mitosis in upper layers		⊗	⊗	⊗	⊗
Hyperkeratosis		⊗			
Nucleoted upper layers			⊕	⊕	⊕
Heterogenous chromatin			⊕	⊕	⊕
Cell layer disorganization			⊕	⊕	⊕
Parakeratosis				⊗	⊕
Dyskeratosis				⊗	⊕
Elastosis				⊗	⊕

France) including a back-illuminated CCD detector that could attain a spectral resolution better than 5 nm in the 350- to 800-nm spectral range.

Our goal was to construct a tunable multiexcitation source based on a short-arc 300 W xenon lamp, remotely software-controlled. The source also needed to be capable of generating adjustable center wavelengths and bandwidths and of automatically switching from one excitation wavelength to the other [for AF spectroscopy: seven excitation wavelengths at 360, 368, 390, 400, 410, 420, and 430 nm, with a full width at half maximum (FWHM) of 15 nm] and from one illumination band to the other (for DR spectroscopy: 365 to 545, 450 to 640, and 550 to 740 nm to cover the UV-visible range) in a series of sequential steps. For this purpose, we used bandpass linear filters (LVF series, Ocean Optics) composed of a long-pass filter and a short-pass filter (respectively referred to as FL1 and FL2 in Fig. 2). Both filters were fixed on motorized translation stages, and wavelength tuning was achieved by sliding the filters with respect to each other using a programmable motorized micropositioning system. The motorized stages slid on a 45-mm-long bench that limited their excursion, resulting in a maximal transmitted bandwidth of 190 nm. The three illumination bands resulted from this mechanical limitation: 3×45 mm were needed to cover the UV-
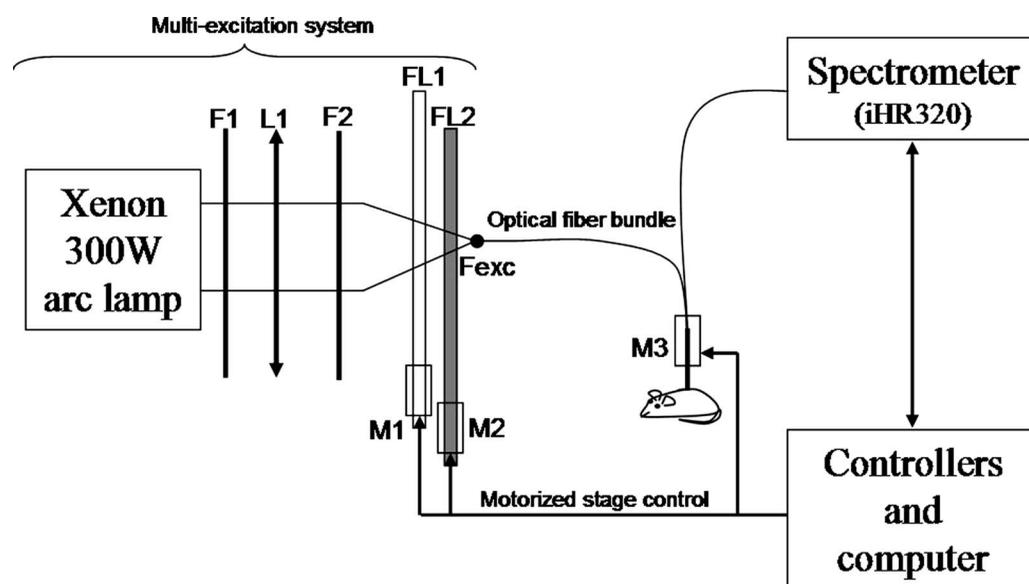

Fig. 2 Schematic diagram of the bimodal spectroscopy system used to acquire mouse skin spectra *in vivo* using a short arc lamp (300-W xenon). F1: anticaloric filter; L1: convex lens; FL1 and FL2: high-pass and low-pass linearly variable filters; Fexc: excitation optical fiber; M1, M2, and M3: motorized micrometric translation stages.

Table 2 Spectral features (19 for AF and 4 for DR) used for spectra classification. Total energy is the sum of all spectra intensities in the $(\lambda_i$-$\lambda_j)$ bandwidth. Area ratio is the ratio between the sum of intensities in the $(\lambda_i$-$\lambda_j)$ bandwidth and the sum of intensities in the $(\lambda_j$-$\lambda_k)$ bandwidth. Peak ratio is the ratio of intensities at λ_m and λ_n. Slope is the slope between the intensities at λ_p and λ_q.

Autofluorescence (AF)

λ_{exc} (nm)	Total energy (a.u.)	Area ratio	Peak ratio
360	$A_{410-700}$	$A_{410-459}/A_{459-700}$	
368	$A_{410-700}$	$A_{410-470}/A_{470-700}$	
390	$A_{460-700}$	$A_{460-545}/A_{545-700}$	$R_{740/634}$
400	$A_{461-700}$	$A_{461-550}/A_{550-700}$	$R_{485/633.5}$
410	$A_{460-700}$	$A_{460-542.5}/A_{542.5-700}$	$R_{492/633}$
420	$A_{462.5-700}$	$A_{462.5-550}/A_{550-700}$	$R_{505/633}$
430	$A_{476-700}$	$A_{476-585}/A_{585-700}$	$R_{507/633}$

Diffuse reflectance (DR)	Total energy	Slope	Peak ratio
	$A_{390-720}$	$S_{580-600}$	$R_{423/577}$
		$S_{615-720}$	

visible spectral range. The reflectance signal was extracted from the spectral ranges for which filter transmission was at maximum, resulting in the following three partly overlapping selected bands: 390 to 490 nm (illumination band #1), 490 to 590 nm (illumination band #2), and 590 to 720 nm (illumination band #3). Then these three bands were simply concatenated. The DR signal in each band was obtained by dividing the raw signal acquired *in vivo* on mouse skin by the signal acquired on a Lambertian surface (WS-1 diffuse reflection standard, Ocean Optics). Such reference spectra were collected every day before starting the experiment. This normalization accounted for the nonuniform spectral response of the acquisition system and for variations of the light-source intensity over the 30-week period.

For AF spectroscopy, spectra were recorded as a whole over the entire acquisition spectral range. In order to reject skin backscattered excitation light, three different long-pass filters were used with the following cutoff wavelengths: 400 nm for 360- and 368-nm excitation wavelengths; 435 nm for 390- and 400-nm excitation wavelengths; and 455 nm for 410-; 420-, and 430-nm excitation wavelengths. AF spectra were corrected for the nonuniform spectral response of the acquisition system using a radiometric calibration source (HL-2000, Ocean Optics). A correction factor was calculated by dividing the HL-2000 experimentally acquired spectrum by the calibration spectrum, then used to correct spectra measured on mouse skin. Variations of the light source intensity were taken into account by measuring excitation intensity at the probe tip using a calibrated photodiode (818-UV, Newport Research Corp.) prior to each experiment. Raw acquired AF spectra were divided by the excitation light power. Fluorescence data from a single measurement spot are represented as an excitation-emission matrix (EEM). Columns of the matrix spectra correspond to emission spectra at each excitation wavelength.

Our spectroscopic statistical analysis was a three-step process: (1) spectral features extraction, (2) data reduction, and (3) classification. As shown in Table 2, 19 spectral features and 4 spectral features, respectively, were extracted from AF and DR spectra that were acquired on each skin sample (resulting in 23 spectral features when the two modalities were combined for the bimodal approach). The spectral parameters were selected to be sensitive to the intensity and spectral shape modifications induced by neoplastic transformation. The total area under the spectrum between λ_i and λ_j $(A_{\lambda i-\lambda j})$ was related to the global backscattered intensity (DR) and emission intensity (AF). Area ratios were $A_{\lambda i-\lambda j}/A_{\lambda j-\lambda k}$. Slopes were between λ_p and λ_q $(S_{\lambda p-\lambda q})$, and peak ratios λ_m and λ_n $(R_{\lambda m/\lambda n})$ were related to the spectral shape. For the case of area ratios $(A_{\lambda i-\lambda j}/A_{\lambda j-\lambda k})$ extracted from AF emission intensity spectra: (1) λ_i was chosen as the shortest wavelength from which the emission signal recorded was considered to be no longer affected by any residual backscattered excitation light; (2) λ_j was the wavelength at which spectra from two different histological classes "naturally" crossed when normalized to unitary surface under the curve; and (3) λ_k was chosen as the longest wavelength that still provided significant AF signal from the fluorescence spectrum tail.

The large number of features was reduced by applying principal component analysis (PCA). This linear transformation technique employed for the dimensionality reduction of the parameter space allowed us to represent most of the information from the original data set in the first few uncorrelated principal components (PCs).[17] When data reduction was done, a *k*-nearest neighbor (*k*-NN) method was used to build the

Table 3 Number of PCs and nearest neighbors (k) tested for each of the four spectroscopic modalities: monoexcitation AF, multiexcitation AF, DR, and bimodality (DR+multiexcitation AF).

Monoexcitation AF		Multiexcitation AF	DR	DR+multiexcitation AF
360, 368 nm	390, 400, 410, 420, 430 nm			
PCs=1,2.	PCs=1,2,3.	PCs=1,2...,7.	PCs=1,2,3,4.	PCs=1,2...,10.
k=1,2,…15.	k=1,2,…,15.	k=1,2,…,15.	k=1,2,…,15.	k=1,2,…,15.

algorithm for classification of spectra in one of the following classes: N, CH, AH, and D. The k-NN method classifies an unknown sample into the class that has the most similar or "nearest" sample points in the training set of data. Euclidean distance was used to measure the distance between points. In this case, we randomly chose 66% of spectra in each histological class as the training set of data. Then any new sample from the validation data set was classified as N, CH, AH, or D by comparing it with the standard models fore each of the four classes.

The algorithm accuracy to classify samples was expressed through two parameters: sensitivity (Se) and specificity (Sp). Sensitivity was defined as the number of samples classified as pathological by statistical analysis divided by the number of pathological samples determined as such by histology. Specificity was defined as the number of samples classified as normal by statistical analysis divided by the number of normal samples determined as such by histology. When the two histological classes to be discriminated were two pathological classes, the less advanced state in the carcinogenic process was used as the "normal" sample. For instance, when discriminating between AH and D, AH was chosen as "normal" and D as "pathological" in the definitions above. To obtain an optimal value of the (Se, Sp) couple, we studied the influence of the number of neighbors ($1 < k < 15$) and of PCs on Se and Sp for each modality (Table 3) when discriminating the six pairs of histological classes. Since the first few PCs usually accounted for most of the total variance of the original data set, and the goal was to obtain optimum classification results with as few PCs as possible, the classification algorithm was performed using only 7 and 10 PCs for multiexcitation AF and bimodality, respectively, accounting for 99% of the variance of the original data set. In the case of monoexcitation AF and DR used alone, since at most four or fewer spectral features were extracted, 100% of the PCs were used.

3 Results and Discussion

3.1 Histology

Because of inter-individual variable response to UV radiation, we decided to classify skin samples exclusively according to histological evaluation and not according to irradiation dose or exposure duration. Figure 3 and Table 4 show the cutoff values (maximum normal epidermis thickness) used to define hyperplasia. Any UV-irradiated epidermis displaying a thickness value above such "normal" thresholds was considered hyperplastic. A Student's t-test was performed to compare mean thickness values between anatomical spots #1 [mouse head, Fig. 4(a)] and #6 [mouse tail, Fig. 4(b)] and showed that

the difference in mean epidermal thickness was significantly different with a level of significance $p=0.0006 < 0.05$. This emphasizes how important it is to base histological comparison of normal and irradiated skin on corresponding anatomical locations, especially in our case, where epidermis thickness was one of the criteria used to define hyperplasia.

Each class of hyperplasia was characterized by qualitative features relative to the epidermis and dermis observed in stained histological sections. At the beginning of UV exposure (the first three to six weeks of exposure), skin displayed cell proliferation that attempted to regenerate the tissue, leading to a compensatory type of hyperplasia (CH). A thickened epidermis as well as a thickened keratin layer (hyperkeratosis) could be observed [Fig. 4(c)], but neither cytological (cell nuclei and layer organization still look normal) nor structural (normal dermal elastic fibers) carcinogenic features were displayed [Fig. 5(a)]. After six months of UV exposure, skin not only became hyperplastic, but it also displayed cancerous characteristics. We distinguished two types of cancerous characteristics relative to nuclei and keratin maturation, respectively. The first type (heterogenous chromatin, mitosis, cell layer disorganization) were related to the AH class. The second type [parakeratosis, Fig. 5(b); dyskeratosis, Fig. 5(c), and elastosis, Fig. 5(d)], appearing in addition to nuclei modifications, were related to the D class. Using the classification criteria defined in Table 1, the 224 mice skin samples were classified as follows: 84 N, 47 CH, 37 AH, and 56 D.

We used SCC that developed on a few mice to check that the sampling protocol allowed a precise match of spectroscopic measurements and histological evaluations. As shown in Fig. 6, SCC appears as a macroscopic abnormal growth on
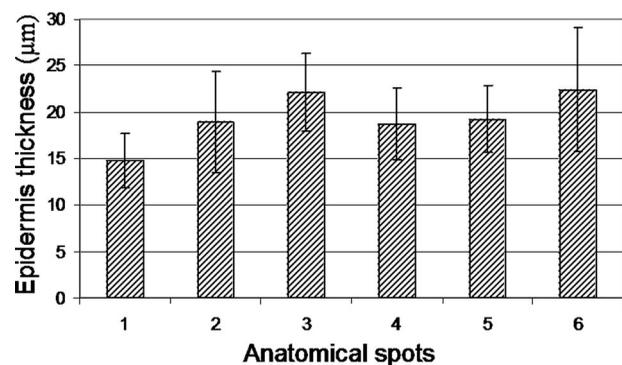

Fig. 3 Histogram showing mean ($n=160$) normal epidermis thicknesses of sham-irradiated mice for the six anatomical spots. Error bars represent ±1SD.

Table 4 Normal epidermis cutoff values [defined as mean ($n=160$)+2 SD] used to define hyperplasia for each anatomical spot (spot #1 is toward the mouse's neck and spot #6 is towards the mouse's tail) to account for physiological variations of epidermis thickness along the backbone.

Anatomical spot number	1	2	3	4	5	6
Mean epidermis thickness+2SD (μm)	21	29	30	27	27	40

mouse's skin [Fig. 6(a)] on spot #5 of spectroscopic measurement. The abnormal growth shown in Fig. 6(b) and further identified as a SCC by histology can also be located on spot #5 of the histological slide. Thus histological classification was provided for each of the corresponding spot of spectroscopic measurements.

3.2 Spectroscopy and Statistical Analysis Results

Figures 7(a)–7(d) show the EEMs (mean values) obtained for each histological class respectively: N, CH, AH, and D. Figure 8 displays 410-nm monoexcitation AF (a) and DR (b) mean spectra for each histological class. As expected, AF emission spectra exhibit characteristic flavins and keratin (410-nm excitation, 510-nm emission) as well as porphyrins (410-nm excitation, double-peak emission at 633 and 672 nm), while DR spectra exhibit characteristic hemoglobin absorption peaks (420, 542, and 577 nm). It should be noted that such spectra are obtained for a 271-μm CEFS.

The aim of our spectroscopic classification was to determine Se and Sp for discrimination of the four histological classes using (1) monoexcitation AF spectroscopy alone based on each of the seven excitation wavelengths, (2) multiexcita-

tion AF spectroscopy based on a combination of the seven excitation wavelengths, (3) DR spectroscopy alone, and (4) a combination of DR and multiexcitation AF spectroscopies (bimodal spectroscopy). We performed a pair-wise comparison: normal versus each type of hyperplasia (N versus CH, N versus AH, and N versus D) and each type of hyperplasia versus
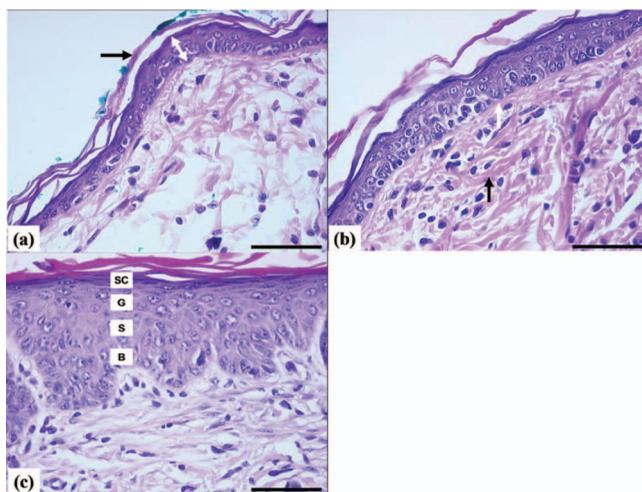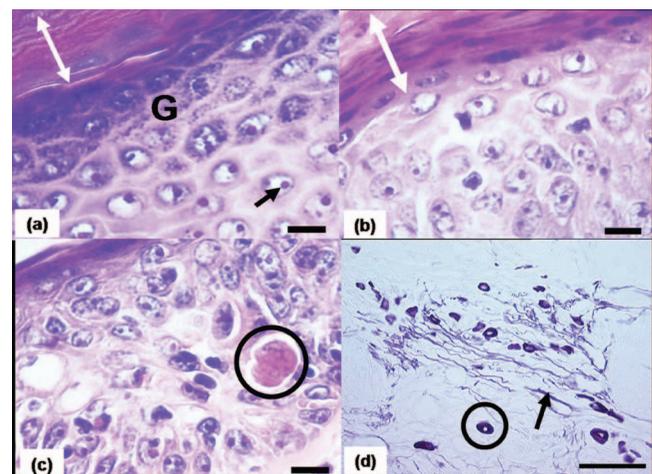

Fig. 5 Representative microscopic images of H&E (a, b, c) and Weigert (d) stained histological sections of mouse skin after (a) 1 month of UV irradiation; and (b, c, d) after 6 months' UV irradiation. Images display the morphological epidermis and dermis features used to classify skin samples in each of the three histological classes (CH, AH, D). (a) Compensatory hyperplastic epidermis featuring thickened epidermis (above normal cutoff values), hyperkeratosis (white arrow), normal keratin maturation (G, granulous layer), organized cell layers, and chromatin confined within nuclei (black arrow). (b) Dysplastic epidermis featuring keratin dysmaturation (parakeratosis: white arrow, and no visible granulous layer), cell layer disorganization, and heterogenous nuclei chromatin. (c) Dysplastic epidermis featuring dyskeratosis (black circle). (d) Dermis displaying elastosis (black arrow: short and thick elastic fibers) and chronic inflammation (circle: mast cell). Scale bars: 10 μm (a, b, c) and 50 μm (d).

Fig. 4 Representative microscopic images of H&E stained histological sections of mouse skin. (a) Sham-irradiated spot #1 displays the epidermis (delimited by the white arrow) and keratin on top of the epidermis (black arrow). (b) Sham-irradiated spot #6 displays basement membrane (white arrow) and dermal collagen fibers (black arrow). The epidermis is significantly thickened compared to spot #1. (c) Spot #6 after 1 month of UV irradiation: SC: stratum corneum; G: granulous layer; S: spinous layer; B: basal cell layer. The epidermis is significantly thickened compared to sham-irradiated spot #6. Scale bar: 50 μm.
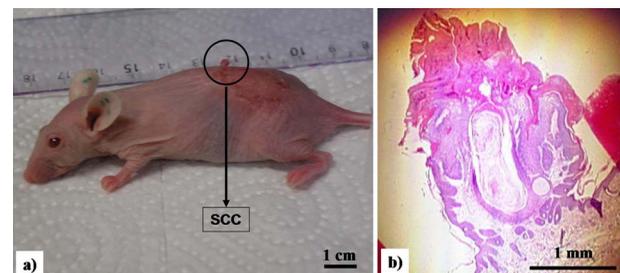

Fig. 6 Matching of (a) anatomical spectroscopic measurement spots, and (b) histological evaluation sites.
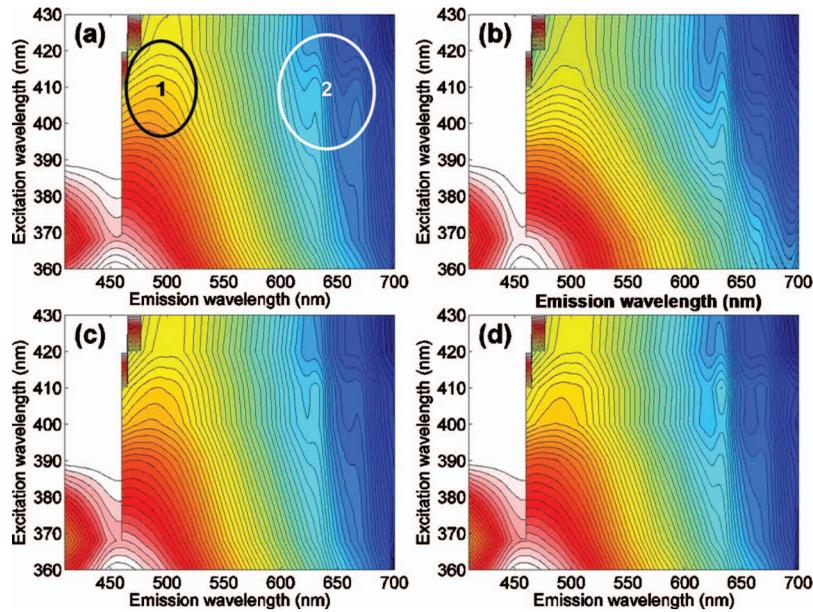
Fig. 7 Mean EEMs from multiexcitation AF spectroscopy of (a) normal, (b) compensatory hyperplastic, (c) atypical hyperplastic, and (d) dysplastic mouse skin. Areas attributable to flavins and keratin (1) as well as porphyrins (2) are circled.

one another (CH versus AH, CH versus D, and AH versus D).

3.2.1 Monoexcitation AF spectroscopy

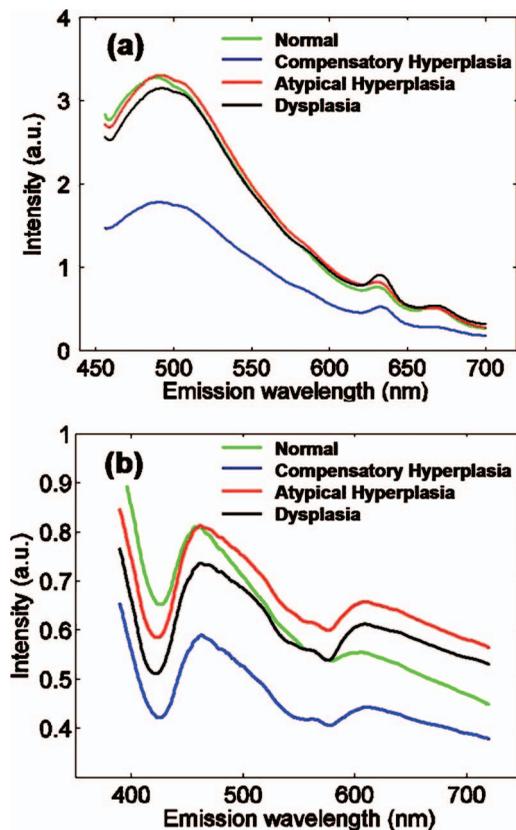Focusing of AF spectroscopy, results concerning Se and Sp are summarized in Figs. 9(a) and 9(b), respectively. When

Fig. 8 Mean (a) AF spectra at 410-nm excitation, and (b) DR spectra obtained for each histological class (N, CH, AH, and D).

discriminating CH from each of the other three histological classes (N, AH, and D), the longer wavelengths (visible wavelengths, i.e., all wavelengths except for 360 and 368 nm) reach higher sensitivity values than the shorter ones (360 and 368 nm). Compensatory states are at low risk to evolve into SCC; therefore, the use of visible wavelengths should help to reduce the number of unnecessary procedures. Such visible wavelengths achieve their lowest sensitivity level (around 50%) when discriminating between N and AH. It can be hypothesized that the keratin layer plays an important role in AF spectral discrimination. Indeed, N and AH are the only two classes that usually display no abnormally thick keratin layer, whereas CH may display hyperkeratosis and D often displays parakeratosis. The keratinized layer absorbs and scatters the excitation light, thus reducing its penetration into deeper layers. This effect can be seen in Fig. 8, where CH, characterized by the thickest keratin layer (hyperkeratosis), exhibits a much lower average fluorescence intensity than all other classes.

Based on information from several studies that investigated fluorescence of keratin and keratinizing layers from various tissues,[18–21] it is most likely that keratin fluorescence is involved in our results concerning AF spectral discrimination. However, its precise role is difficult to determine, because the AF signal results from the overlap of various fluorophores located at different depths. A tentative explanation that accounts for the better sensitivity achieved using visible rather than UV wavelengths (360 and 368 nm) when discriminating CH from each of the other three histological classes (N, AH, and D) would be as follows: the keratin layer plays a shielding role, most pronounced in UV, reducing the excitation of other fluorophores in the epidermis (NADH) and dermis (collagen, elastin), which are spectral markers of neoplastic transformation. Under UV excitation, the observed AF signal is a mixture of NADH, collagen, and keratin emissions with a high contribution of keratin possibly blurring the differences in spectral signatures due to other fluorophores.
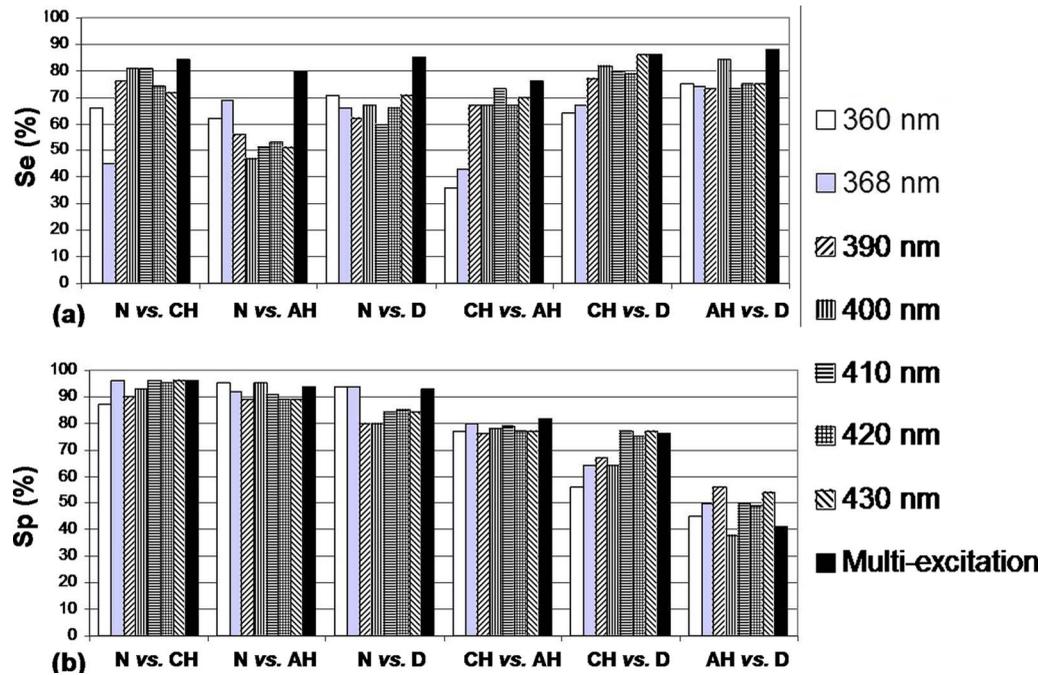
Fig. 9 Average (a) Se and (b) Sp achieved by monoexcitation (for each of the seven excitation wavelengths) and multiexcitation AF spectroscopy for the discrimination of the six pairs of histological classes.

Using either monoexcitation or multiexcitation AF, a drastic decrease in Sp can be observed in Fig. 9(b) when discriminating between AH and D compared to all other pair-wise comparisons. Such a low Sp might be explained by the close histological characteristics of these two classes (see Table 1).

3.2.2 Multiexcitation AF spectroscopy

As shown in Fig. 9(a), multiexcitation AF always provides higher Se compared to any single excitation wavelength, resulting in an average 6 percentage points increased Se over all pairs of histological classes. Conversely, Sp [Fig. 9(b)] is not improved by multiexcitation (except for CH versus AH: +2 percentage points): Sp remains higher than 92% when discriminating between normal skin and the three types of hyperplasia (CH, AH, and D) but drastically decreases when

discriminating the three types of hyperplasia from one another (82%, 76%, and 41%, respectively, for each pair of histological classes).

3.2.3 DR spectroscopy

As shown in Fig. 10, DR spectroscopy is somewhat more sensitive (+5.6 percentage points) when discriminating the three types of hyperplastic tissues from one another (75.3% average sensitivity over the three pairs of histological classes: CH versus AH, CH versus D, and AH versus D) than when discriminating normal skin from any type of hyperplasia (69.7% over the three pairs of histological classes: N versus CH, AH, and D, respectively). Due mainly to the low Sp achieved when discriminating AH from D (46%), a drastic fall of Sp (−24 percentage points) is observed when discriminat-
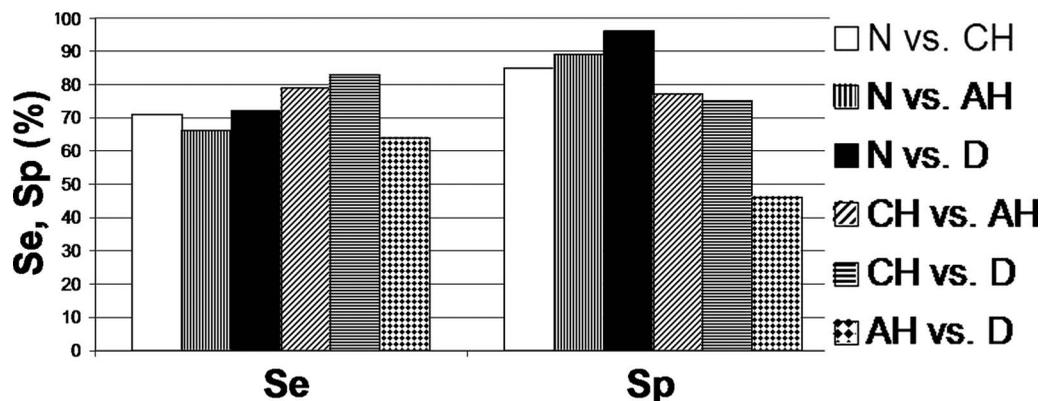

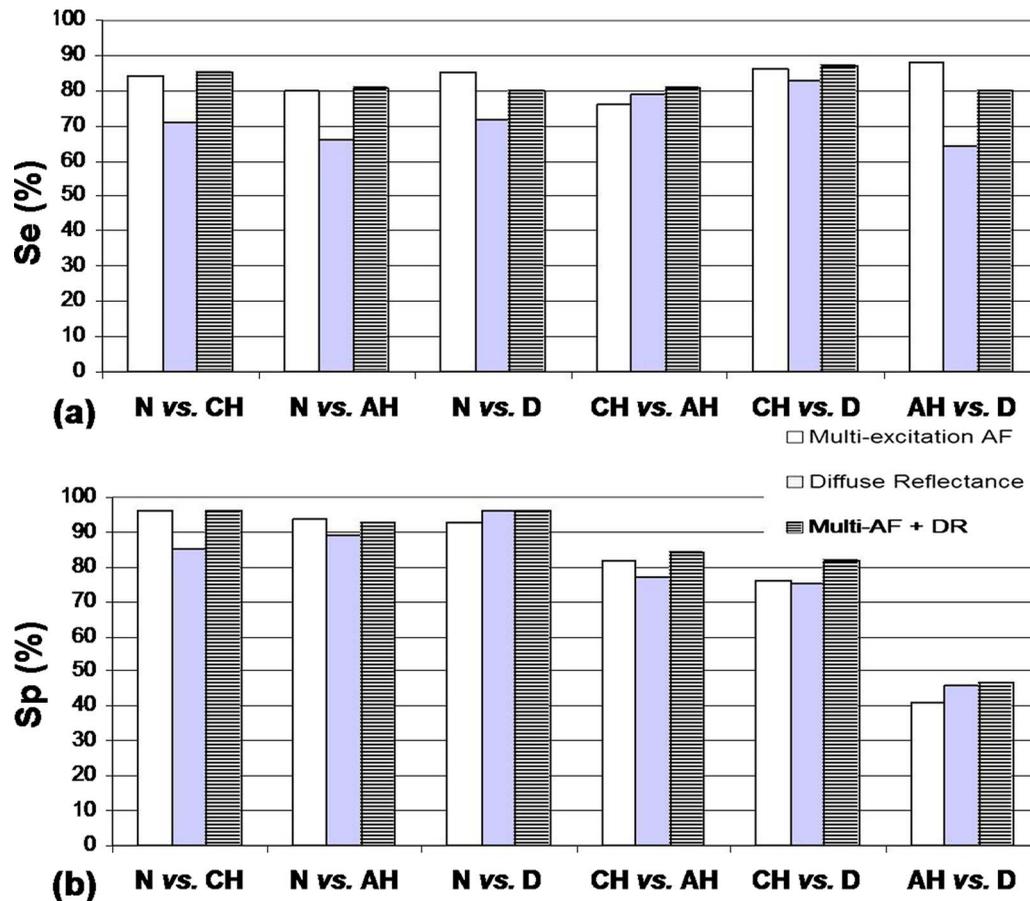**Fig. 10** Average Se and Sp achieved by DR for each pair of histological classes.

Fig. 11 Average (a) Se and (b) Sp of multiexcitation AF (white), DR (grey), and bimodal spectroscopy (multiexcitation AF+DR: dash) when discriminating each pair of histological classes.

ing the three types of hyperplasia from one another (average Sp of 66% over the three pairs of histological classes). Indeed, the Sp achieved when discriminating normal tissue from each type of hyperplasia is as high as 90% (average Sp over the three histological classes).

3.2.4 Multiexcitation AF, DR, and bimodality

A comparative analysis between the diagnostic performances of multiexcitation AF alone, DR alone, and bimodal spectroscopies can be seen in Fig. 11. If we were to select only one modality on the basis of sensitivity results, multiexcitation AF would be selected. Indeed, on average, over the six pairs of histological classes to discriminate, multiexcitation AF was the most sensitive modality with an average sensitivity of 82%, which is +1 and +10 percentage points compared to bimodality and DR used alone, respectively. Bimodality is most interesting when considering Sp. It is either as specific or more specific than multiexcitation AF and DR used alone over all pairs of histological classes (except for N versus AH: 94% and 93% for multiexcitation AF and bimodality, respectively). Even though Sp remains low, bimodality slightly improves Sp (+4 percentage points) when discriminating the three types of hyperplasia from one another, and achieves a 71% Sp compared to 66% for both multiexcitation AF and DR used alone (average over the three pairs of histological classes).

To evaluate the global diagnostic performance and taking into account both Se and Sp as a function of the numbers of k and PCs, we used a receiver operating characteristic (ROC)
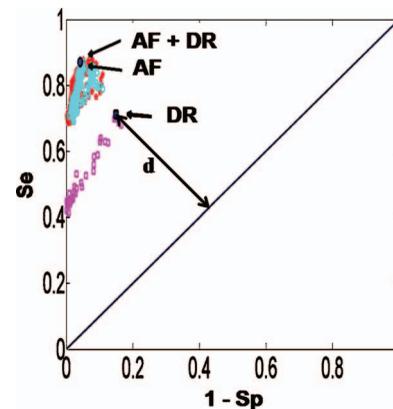

Fig. 12 ROC plot showing how the optimum classification accuracy couple (Se and Sp) is determined for each spectroscopic modality: blue data cluster, AF; pink, DR; and red, bimodality (DR+AF). The current plot illustrates the determination of the optimum (Se, Sp) couple for the discrimination of N and AH using DR; the data point that is the farthest away at distance d is kept as the optimum classification result.
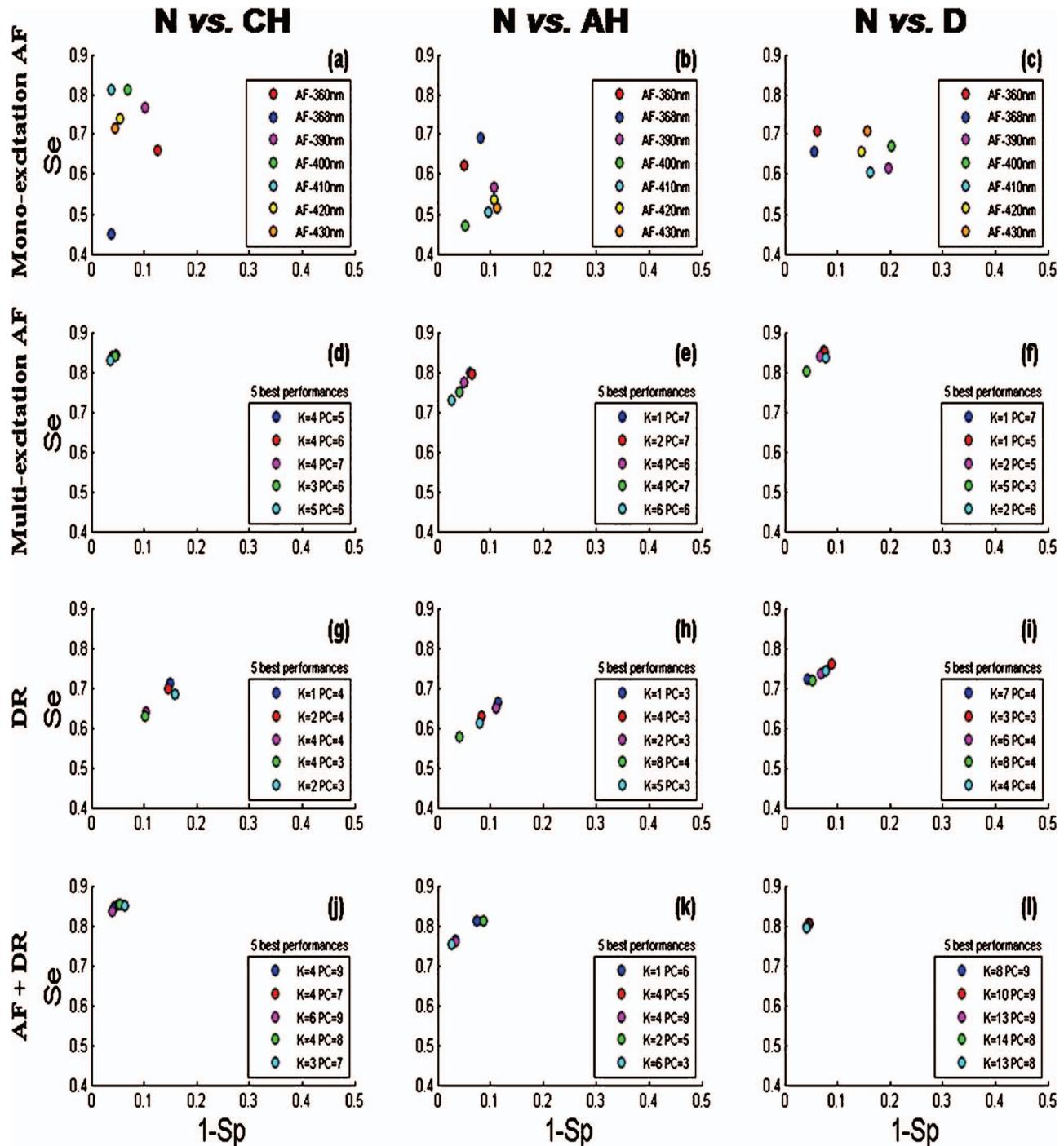
Fig. 13 ROC curves for the discrimination between N mouse skin and the three types of hyperplasia (CH, AH, D). First row (a, b, and c) shows monoexcitation AF spectroscopy with the optimum (Se, Sp) couple displayed for each excitation wavelength (red, 360 nm; dark blue, 368 nm; pink, 390 nm; green, 400 nm; light blue, 410 nm; yellow, 420 nm; and orange, 430 nm). Second row (d, e, and f) shows multiexcitation AF with the best five classification performances displayed with corresponding numbers of neighbors (k) and PCs (dark blue, red, pink, green, and light blue). Third row (g, h, and i) shows DR with the same color coding as above. Fourth row (j, k, and l) shows bimodality, i.e., multiexcitation AF +DR, with the same color coding as above. In (1), the points are very close and nearly superposed.

representation.[22] We evaluated the best combination of Se and Sp corresponding to the data point that was the furthest away from the median line, as shown with distance *d* in Fig. 12. Figures 13 and 14 display ROC representations for each of the four spectroscopic modalities (corresponding to each row of Figs. 13 and 14) and for each pair of histological classes

(corresponding to each column of Figs. 13 and 14). For every case, all the combinations of numbers of PCs and nearest neighbors were tested to determine which ones achieved the best diagnostic accuracy, i.e., the optimum [Se, Sp] couple. The best five performances are indicated for each pair-wise analysis in Figs. 13 and 14.
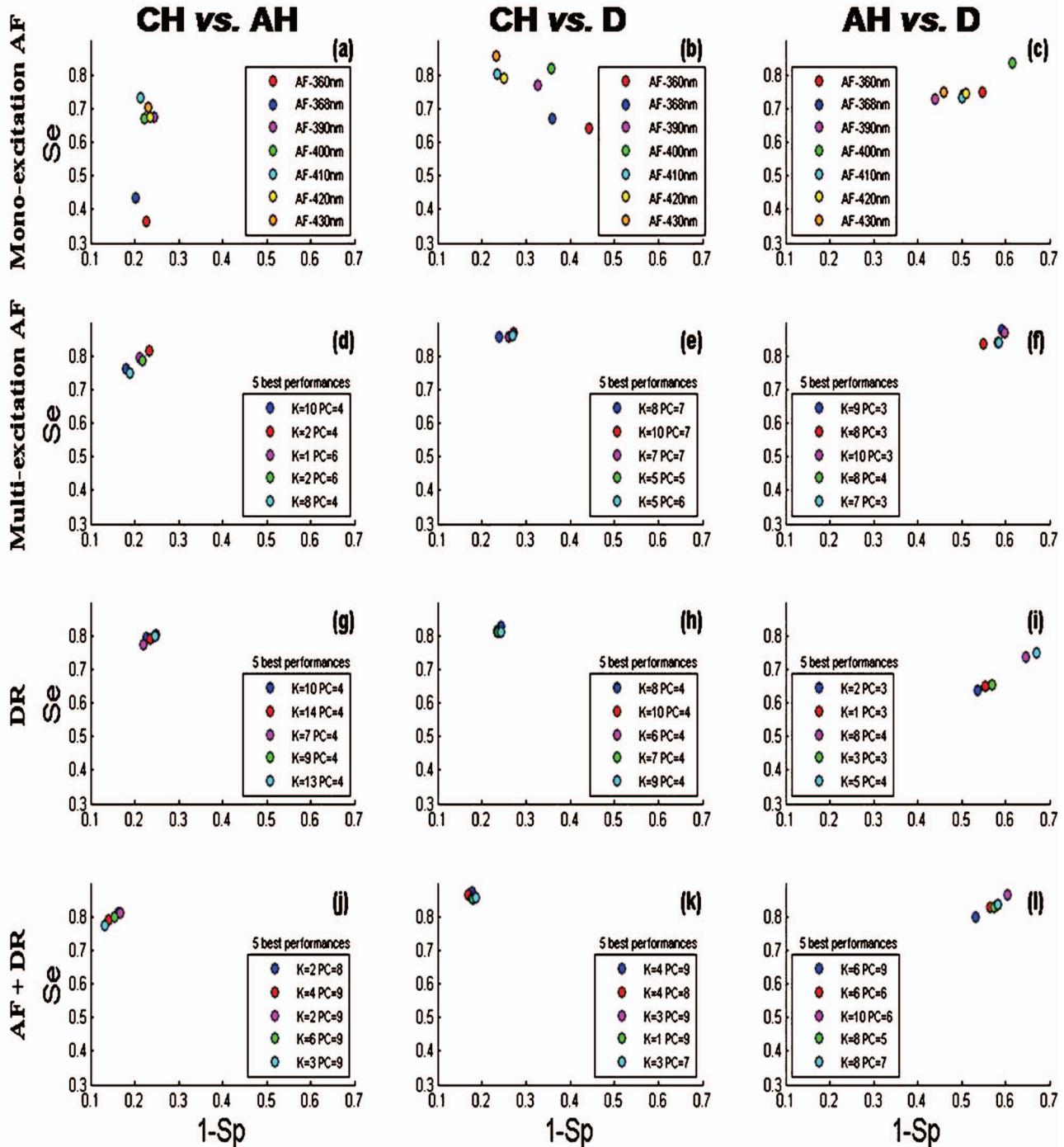
Fig. 14 ROC curves for the discrimination between the three types of hyperplasia from one another (CH, AH, D). First row (a, b, and c) shows monoexcitation AF spectroscopy with the optimum (Se, Sp) couple displayed for each excitation wavelength (red, 360 nm; dark blue, 368 nm; pink, 390 nm; green, 400 nm; light blue, 410 nm; yellow, 420 nm; and orange, 430 nm). Second row (d, e, and f) shows multiexcitation AF with the best five classification performances displayed with corresponding numbers of neighbors (k) and PCs (dark blue, red, pink, green, and light blue). Third row (g, h, and i) shows DR with the same color coding as above. Fourth row (j, k, and l) shows bimodality, i.e., multiexcitation AF +DR with the same color coding as above.

Our results are consistent with those reported by Diagaradjane et al. in a study that characterized early neoplastic changes in a DMBA/TPA-induced mouse skin tumor model.[5] They used multiexcitation AF in conjunction with a multivariate statistical method. Out of the 19 excitation wavelengths (ranging from 280 to 460 nm) used by the authors, the best classification accuracy was obtained for 280 and 410 nm (74.3% and 73.1%, respectively). We did not test wavelengths as short as 280 nm because their use

might be questionable in a clinical trial due to the risk of genetic mutations. But in the spectral range (360 to 430 nm) used by both our teams, Diagaradjane et al. also reported that visible wavelengths (410 nm) achieved better classification accuracy than shorter ones (e.g., 360 nm: 63.9% accuracy).[5] When comparing AF spectroscopy's classification ability using either one excitation wavelength or a combination of the five most discriminating excitation wavelengths (out of the 19 they tested), they showed that statistical analysis of the combination wavelengths resulted in an 11.6% increase in the overall classification accuracy when compared to the highest classification accuracy obtained with single-wavelength analysis. On UV-irradiated mouse skin, our results showed an average 9.3 percentage points increased sensitivity when considering the discrimination of normal skin from each of the three classes of UV-irradiated skin.

Chang et al. evaluated the diagnostic potential of multiexcitation AF alone, DR alone, and bimodality to discriminate human (columnar and squamous) normal cervical tissue from different types of precancerous tissues.[12] Their algorithms were tested to discriminate pairs of histological classes as well: normal (columnar or squamous) versus low-or high-grade intraepithelial lesions (columnar or squamous). They showed that fluorescence alone gave a superior performance compared to reflectance alone, and that the combination of reflectance spectra and fluorescence spectra provided a modest improvement in diagnostic performance. Similarly, in our case we showed that multiexcitation AF yielded higher Se than DR and that Sp was only slightly improved by bimodality compared to multiexcitation AF or DR used alone when discriminating the different types of hyperplasia from one another. This is precisely the use for which bimodality is of particular value: increasing the diagnostic Sp when discriminating benign (i.e., CH) from precancerous stages (i.e., AH and D) would be of great medical interest to reduce unnecessary biopsies.

4 Conclusion

To our knowledge, the current study is the first one to show bimodal spectroscopic evaluation of different stages of UV-irradiated mouse skin. We used an animal model to study changes in tissue throughout the carcinogenesis sequence because such information is not available with human patients. Based on epidermis and dermis morphological characteristics, a classification of UV-irradiated mouse skin stages was performed. With this classification, compensatory hyperplastic (CH) states were distinguished from atypical hyperplastic and dysplastic ones (AH and D). The former types are considered to be benign and the second types to be precancerous. Our results showed that visible wavelengths tended to achieve better classification performance than the shorter ones (360 and 368 nm) in discriminating benign from precancerous skin samples. These are encouraging results for clinical application since such visible wavelengths are harmless. Combining several excitation wavelengths improved diagnostic sensitivity and combining modalities (DR and multiexcitation AF) improved diagnostic specificity, especially when discriminating the three types of hyperplasia from one another. These results suggest that bimodal spectroscopy could be an accurate technique to discriminate normal and benign states (H and CH)

from precancerous ones (AH and D). We believe the sensitivity and specificity of the method could be improved using several collection optical fibers instead of a single one. Indeed, combining information from a number of points located at variable distances from the excitation optical fiber will probe the tissue at several depths,[23] resulting in a potentially better discrimination between the different histological classes. This work is still in progress: data from collection fibers set at 536 and 834 μm from the excitation fiber are being combined with those collected at 271 μm (presented in the current paper) to investigate whether this combination improves the classification accuracy. To confirm the diagnostic potential of the method, a further step will be needed to assess it in a clinical trial.

References

1. M. Mogensen and B. E. Gregor, "Diagnosis of nonmelanoma skin cancer/keratinocyte carcinoma: a review of diagnostic accuracy of nonmelanoma skin cancer diagnostic tests and technologies," *Dermatol. Surg.* **33**(10), 1158–1174 (2007).
2. E. LeBoit, G. Burg, D. Weedon, and A. Sarasin, *Pathology and Genetics of Skin Tumours*, IARC Press, Lyon, France (2006).
3. R. Lewis Kelso, M. I. Colome-Grimmer, T. Uchida, H. Q. Wang, and R. F. Wagner Jr., "p75(NGFR) immunostaining for the detection of perineural invasion by cutaneous squamous cell carcinoma," *Dermatol. Surg.* **32**(2), 177–183 (2006).
4. H. Tehrani, J. Walls, S. Cotton, E. Sassoon, and P. Hall, "Spectrophotometric intracutaneous analysis in the diagnosis of basal cell carcinoma: a pilot study," *Int. J. Dermatol.* **46**(4), 371–375 (2007).
5. P. Diagaradjane, M. A. Yaseen, J. Yu, M. S. Wong, and B. Anvari, "Autofluorescence characterization for the early diagnosis of neoplastic changes in DMBA/TPA-induced mouse skin carcinogenesis," *Lasers Surg. Med.* **37**, 382–395 (2005).
6. V. E. Reeve, S. Widyarini, D. Domanski, E. Chew, and K. Barnes, "Protection against photoaging in the hairless mouse by the isoflavone equol," *Photochem. Photobiol.* **81**(6), 1548–1553 (2005).
7. A. P. Pentland, G. Scott, J. A. Vanbuskirk, C. Tanck, G. LaRossa, and S. Brouxhon, "Cyclooxygenase-1 deletion enhances apoptosis but does not protect against ultraviolet light-induced tumors," *Cancer Res.* **64**, 5587–5591 (2004).
8. F. R. de Gruijl and J. C. van der Leun, "Physical variables in experimental photocarcinogenesis and quantitative relationships between stages of tumor development," *Front. Biosci.* **7**, 1525–1530 (2002).
9. V. Molho-Pessach and M. Lotem, "Ultraviolet radiation and cutaneous carcinogenesis," *Curr. Probl Dermatol.* **35**, 14–27 (2007).
10. E. L. Franco and T. E. Rohan, *Cancer Precursors, Epidemiology, Detection, and Prevention*, Springer-Verlag, New York (2002).
11. K. E. de Visser, L. M. Coussens, K. Dittmar, K. S. Zaenker, and A. Schmidt, *The Inflammatory Tumor Microenvironment and Its Impact on Cancer Development. Infection and Inflammation: Impacts on Oncogenesis*, Karger, Basel, Switzerland (2006).
12. S. K. Chang, Y. N. Mirabal, E. N. Atkinson, D. Cox, A. Malpica, M. Follen, and R. Richards-Kortum, "Combined reflectance and fluorescence spectroscopy for in vivo detection of cervical pre-cancer," *J. Biomed. Opt.* **10**(2), 024031 (2005).
13. T. M. Breslin, F. Xu, G. M. Palmer, C. Zhu, K. W. Gilchrist, and N. Ramanujam, "Autofluorescence and diffuse reflectance properties of malignant and benign breast tissue," *Ann. Surg. Oncol.* **11**, 65–70 (2004).
14. I. Georgakoudi, B. C. Jacobson, J. Van Dam, V. Backman, M. B. Wallace, M. G. Müller, Q. Zhang, K. Badizadegan, D. Sun, G. A.

Thomas, L. T. Perelman, and M. S. Feld, "Fluorescence, reflectance, and light-scattering spectroscopy for evaluating dysplasia in patients with Barrett's esophagus," *Gastroenterology* **120**, 1620–1629 (2001).

15. W. J. Marshall and S. K. Bangert, *Clinical Biochemistry: Metabolic and Clinical Aspects*, Churchill Livingston, Edinburgh (2008).
16. V. Kumar, R. S. Cotran, and S. L. Robbins, *Robbins Basic Pathology*, Saunders, Philadelphia (2007).
17. S. D. Kamath and K. K. Mahato, "Optical pathology using oral tissue fluorescence spectra: classification by principal component analysis and k-means nearest neighbor analysis," *J. Biomed. Opt.* **12**(1), 014028 (2007).
18. Y. Wu, P. Xi, J. Y. Qu, T.-H. Cheung, and M.-Y. Yu, "Depth-resolved fluorescence spectroscopy reveals layered structure of tissue," *Opt. Express* **12**(14), 3218–3223 (2004).
19. Y. Wu, P. Xi, J. Y. Qu, T.-H. Cheung, and M.-Y. Yu, "Depth-resolved fluorescence spectroscopy of normal and dysplastic cervical tissue," *Opt. Express* **13**(2), 382–388 (2005).
20. A.-M. Pena, M. Strupler, T. Boulesteix, G. Godeau, and M.-C. Schanne-Klein, "Spectroscopic analysis of keratin endogenous signal for skin multiphoton microscopy," *Opt. Express* **13**(16), 6268–6274 (2005).
21. M. C. Skala, J. M. Squirrell, K. M. Vrotsos, J. C. Eichkhoff, A. Gendron-Fitzparrick, K. W. Eliceiri, and N. Ramanujam, "Multiphoton microscopy of endogenous fluorescence differenciates normal, precancerous, and cancerous squamous epithelial tissues," *Cancer Res.* **65**(4), 1180–1186 (2005).
22. D. C. de Veld, M. Skurichina, M. J. Witjes, R. P. Duin, H. J. Sterenborg, and J. L. Roodenburg, "Clinical study for classification of benign, dysplastic, and malignant oral lesions using autofluorescence spectroscopy," *J. Biomed. Opt.* **9**(5), 940–950 (2004).
23. Q. Liu and N. Ramanujam, "Sequential estimation of optical properties of a two-layered epithelial tissue model from depth-resolved ultraviolet-visible diffuse reflectance spectra," *Appl. Opt.* **45**, 4776–4790 (2006).