

Methods for Improved Imaging and Analysis of Tissue-based Biomarkers

by

Mehrnoush Khojasteh-Lakelayeh

B.Sc., Sharif University of Technology, Tehran, Iran

M.A.Sc., The University of British Columbia, Vancouver, BC, Canada

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES
(Electrical and Computer Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

December 2012

© Mehrnoush Khojasteh-Lakelayeh, 2012

Abstract

Study of molecular biomarkers can provide insight into the molecular complexity of cancer, create new cancer screening tools, monitor treatment's effectiveness, and predict patient's response to treatment. This thesis proposes novel methods for the improved quantitative analysis of labeled molecular biomarkers in tissue sections. This is a necessary step towards the ultimate goal of personalized treatment of cancer.

As 85% of all cancers arise in epithelial tissue, we have developed means for objectively and quantitatively assessing the distribution of a molecular biomarker in epithelial tissue sections. We have applied this means to characterize the spatial distribution of proliferating cells in 613 normal and pre-neoplastic bronchial epithelial biopsies. We have demonstrated, for the first time ever, that the knowledge of the spatial distribution of proliferating cells enables prediction of the outcome of lung intraepithelial lesions.

We have developed methods for the automated and quantitative assessment of the expression of tissue-based molecular biomarkers on a cell-by-cell basis. This is achieved by multispectral imaging of labeled tissue sections. We have proposed methods for unsupervised linear spectral unmixing of multispectral images for the purpose of identifying individual labels in a multiple labeled tissue section. We have demonstrated that the use of multispectral imaging combined with our proposed analysis methods quantitatively improves the results of cell nuclei identification compared to three-color RGB imaging, in more than 22,000 cells in 58 tissue sections with nuclear, cytoplasmic, or membrane bound biomarkers.

Finally, we have developed an imaging method for capturing images representing biomarkers in a tissue. Compared to multispectral imaging, our proposed imaging method significantly reduces the number of captured images required for the identification of a biomarker in a tissue. This method uses images captured under a series of narrow-band illumination spectra $\varphi_i, i = 1, 2, \dots, N$ to find a weighted linear combination of the images that represents a certain component in a tissue. The weights in the weighted linear combination of images are then used to design one or two illumination spectra as weighted linear combinations of the narrow-band spectra φ_i . Images representing the component of interest are then captured under the designed illuminations.

Preface

This dissertation consists of three research chapters: chapters 2,3, and 4. Below I list my contributions towards the study reported in each chapter:

Chapter 2: I designed and developed the analysis methods. The patient samples had been collected in the British Columbia Cancer Agency as part of the Lung Health Study program led by Dr Stephen Lam. Tissue samples had been sectioned, stained, and imaged. I used the images of the immunostained tissue sections as input to the programs that I developed. I analyzed the data and drew conclusions based on the analysis results.

Chapter 3: I designed and developed the analysis methods. Four sets of immunostained tissue sections were available for this study. I set up the imaging system and built tools for the automated image capture of the tissue sections. I captured images of the tissue sections using the imaging system. I analyzed the images using the analysis programs that I had developed. I designed evaluation methods, and generated evaluation results and drew conclusions based on those results.

Chapter 4: I designed and developed the analysis methods, set up the imaging system and wrote software for acquiring images using the imaging system. I acquired images of several samples that were available to this study. I analyzed the images, discussed the results, and drew conclusions.

The work presented in this thesis has been partially published in different journals or conference proceedings. The list of these publications is provided below:

- **Khojasteh, M.**, Ward, R., and MacAulay, C. Quantification of membrane IHC stains through multi-spectral imaging. *9th IEEE International Symposium on Biomedical Imaging*, 752-755 (2012).
- **Khojasteh, M.**, Buys, P.H. T, le Riche J, Lam, S., Guillaud, M., and MacAulay, C. A framework for quantitative assessment of Ki67 distribution in preneoplastic bronchial epithelial lesions. *Analytical and Quantitative Cytology and Histology* **34** (2012).
- **Khojasteh, M.** and MacAulay, C. Selective Excitation Light Fluorescence (SELF) Imaging. *Biomedical Optics, OSA Technical Digest, Optical Society of America* (2010).

- **Khojasteh, M.** and MacAulay, C. Selective Excitation Light Fluorescence (SELF) Imaging. *Proc. SPIE 7568, Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues VIII*, 75680A (2010).

I have been the main author for all publications and have had the main role in generating the ideas, developing the methodologies, processing the data, and analyzing the results.

Parts of Chapter 2 have been published in: Khojasteh, M., Buys, P.H. T, le Riche J, Lam, S., Guillaud, M., and MacAulay, C. A framework for quantitative assessment of Ki67 distribution in preneoplastic bronchial epithelial lesions. *Analytical and Quantitative Cytology and Histology* 34, 120-138 (2012). Figures 2-1 to 2-17 and Tables 2-1 and 2-2 are reprinted with permission of the journal of Analytical and Quantitative Cytology and Histology.

Table of Contents

Abstract	ii
Preface	iii
Table of Contents	v
List of Tables	xii
List of Figures	xiii
Acknowledgements	xxi
Dedication	xxii
Chapter 1. Introduction	1
1.1 Personalized Medicine and Molecular Biomarkers	1
1.2 Need for Automated Tissue-based Biomarker Analysis	2
1.3 Need for Cellular-Level Analysis of Tissue-based Biomarkers	3
1.4 State of the Art in Quantitative Analysis of IHC	4
1.5 Aims of the Thesis	7
1.5.1 Quantification of the Spatial Distribution of Biomarkers	9
1.5.2 Quantification of the Expression of Biomarkers in Tissue Sections	9
1.5.3 Design of “Selective Spectral Illumination” for Imaging of Biomarkers	10
1.6 Dissertation Organization	11
Chapter 2. Quantitative Assessment of Spatial Distribution of Tissue-based Biomarkers	13
2.1 Introduction	13
2.2 Background	14

2.2.1	Lung Cancer	14
2.2.2	Squamous Cell Carcinoma Preneoplastic Lesions	15
2.2.3	Natural Course of Preneoplastic Lesions in Bronchial Epithelium.....	16
2.2.4	Proliferative Biomarker Ki67 in Bronchial Epithelial Lesions.....	18
2.2.5	Ki67 Scoring	18
2.3	Materials	20
2.3.1	Study Population	20
2.3.2	Bronchoscopic Examination	20
2.3.3	Histologic Examination.....	20
2.3.4	Ki67 Expression	21
2.3.5	Data	21
2.4	Methods.....	22
2.4.1	Interpretation of Immunohistochemistry.....	22
2.4.2	Spatial Characterization of the Epithelium	23
2.4.3	Voronoi Parameters.....	23
2.4.3.1.1	Voronoi Neighbors	24
2.4.3.1.2	Voronoi Layers.....	24
2.4.3.2	Distance Parameters	24
2.4.4	Quantitative Ki67 Features.....	25
2.5	Results.....	28
2.5.1	Reproducibility of Ki67 Features.....	28
2.5.2	Association of Ki67 Features with Histology	30
2.5.3	Ki67 Expression Pattern as a Function of Depth	40
2.5.4	Ki67-Positive Cells in the Basal Layer	43
2.5.5	Association of Ki67 Features with Progression	44

2.5.5.1	Definition of Progression/Regression.....	44
2.5.5.2	Outcome Analysis	46
2.6	Discussion.....	54
2.7	Summary.....	60
Chapter 3.	Quantitative Tissue-based Biomarker Analysis Using Multispectral Imaging	62
3.1	Introduction.....	62
3.2	Background.....	65
3.2.1	Immunohistochemistry.....	65
3.2.2	IHC Assessment.....	66
3.2.3	Quantitative IHC	68
3.2.4	Image Analysis Solutions for Quantitative IHC	68
3.2.5	Quantitative IHC Using Multispectral Imaging.....	71
3.2.5.1	Multispectral Imaging Hardware.....	72
3.2.5.2	Multispectral Image Analysis.....	73
3.2.5.3	Applications of Linear Spectral Unmixing in Quantitative IHC.....	73
3.3	Image Acquisition.....	74
3.3.1	Data	74
3.3.2	Imaging System.....	74
3.4	Linear Spectral Unmixing of Multispectral Images.....	77
3.4.1	Methods.....	77
3.4.1.1	Relationship between Transmission Images and Concentration of Dyes.....	77
3.4.1.2	Solving the Linear Mixing Model	79
3.4.1.3	MCR-ALS	80
3.4.1.4	Initial Estimates	81
3.4.2	Results.....	83

3.4.3	Ki67 Set.....	83
3.4.3.1.1	MCR-ALS with Manually Captured Spectra	84
3.4.3.1.2	Obtaining the Initial Estimates of Components Spectra.....	90
3.4.3.1.3	Obtaining the Initial Estimates of Components Concentration Maps ...	92
3.4.3.1.4	Linear Spectral Unmixing Using RGB Images	93
3.4.3.2	P16 Set.....	95
3.4.3.3	CD8 Set.....	107
3.4.3.4	Ki67 & P16 Set.....	114
3.4.4	Discussion	127
3.5	Linear Discriminant Analysis of Multispectral Images	130
3.5.1	Methods.....	130
3.5.2	Results	131
3.5.2.1	Ki67 Set	131
3.5.2.2	P16 Set.....	134
3.5.2.3	CD8 Set.....	136
3.5.2.4	Ki67 & P16 set	138
3.5.3	Discussion	145
3.6	Segmentation of Nuclei.....	146
3.6.1	Methods.....	146
3.6.1.1	Initial Binarization.....	146
3.6.1.2	Edge Relocation.....	150
3.6.1.3	Distance-based Watershed Segmentation.....	152
3.6.1.4	Merging	153
3.6.2	Results	153
3.6.2.1	Segmentation Results	154

3.6.2.2	Validation	165
3.6.3	Discussion	172
3.7	Quantification of Biomarker Expression	175
3.7.1	Methods.....	175
3.7.1.1	Voronoi Tessellation.....	176
3.7.1.2	Nearest Neighbour Tessellation.....	177
3.7.1.3	Classification of Cells.....	177
3.7.2	Results	179
3.7.3	Discussion	188
3.8	Summary	190
Chapter 4.	Tissue Biomarker Imaging through Selective Illumination	192
4.1	Introduction.....	192
4.1.1	Transmission Imaging	193
4.1.2	Fluorescence Imaging	194
4.2	Selective Illumination Transmission Imaging	195
4.2.1	Methods.....	196
4.2.2	Results	200
4.2.3	Discussion	221
4.3	Fluorescence Excitation Imaging.....	223
4.3.1	Background on Fluorescence Imaging.....	223
4.3.2	Multispectral Excitation Imaging.....	227
4.3.3	Selective Excitation Light Fluorescence (SELF) Imaging.....	229
4.3.4	Spectral Unmixing of Multispectral Excitation Images	233
4.3.4.1	Estimating the Number of Pure Components	236
4.3.5	Results	238

4.3.5.1	Fluorescence Unmixing through Multispectral Excitation Imaging	238
4.3.5.1.1	Immunofluorescence	238
4.3.5.1.2	Fluorescence-labeled cells	243
4.3.5.1.3	Autofluorescence Removal from FISH-labeled Tissue Section	247
4.3.5.2	SELF Imaging	252
4.3.5.2.1	PCA Imaging of Plastic Material	252
4.3.5.2.2	Spectrally Unmixed Imaging of Fluorescence-labeled Cells	257
4.3.5.3	Excitation Imaging of a Healing Wound	261
4.3.6	Discussion	263
4.4	Summary	266
Chapter 5.	Conclusion	268
5.1	Quantification of the Spatial Distribution of Biomarkers	268
5.1.1	Summary of Results	268
5.1.2	Suggested Future Directions	269
5.2	Cellular Level Quantification of the Expression of Biomarkers	269
5.2.1	Summary of Results	269
5.2.2	Suggested Future Directions	274
5.3	“Selective Spectral Illumination” for Imaging of Tissue-based Biomarkers	275
5.3.1	Summary of Results	275
5.3.2	Suggested Future Directions	277
References	279
Appendices	289
Appendix A.	Additional Spectral Unmixing Results	289
A.1.	Ki67 Set	289
A.2.	P16 Set	294

A.3.	CD8 Set	300
A.4.	Ki67 & P16 Set	302
Appendix B. Additional Nuclei Segmentation Results		305
B.1.	Ki67 Set.....	305
B.2.	P16 Set.....	307
B.3.	CD8 Set	310

List of Tables

Table 2.1 Inter-observer effects on the reproducibility of the features. P-values from 2-way ANOVA.	30
Table 2.2 P-values from the mean rank comparisons using a Kruskal-Wallis ANOVA test. Groups 1, 2, 3, 4 and 5 respectively correspond to normal, hyperplasia, metaplasia, mild dysplasia, and a combined group of moderate and severe dysplasia.	39
Table 2.3 Outcome of lesions with different baseline histology grades. Entries highlighted in pink represent numbers of the progressing lesions that were included in the analyses. Entries highlighted in blue represent the numbers of regressing lesions that were included in the analyses.	45
Table 2.4 Means and standard deviations of each Ki67 feature in the two groups of progressing and regressing lesions. 4 th and 5 th columns represent p-values of t-test with equal and unequal estimates of the variance of feature in each group, respectively. Last column represents the <i>p</i> -value of a test comparing the variance of the feature in each group. Where the p-value is less than 0.05 (highlighted in gray), the hypothesis of equal variances can be rejected; therefore, the <i>p</i> -value reported in the 5 th column (separate estimates of variance) should be used. <i>p</i> -values less than 0.05 are highlighted in blue.	47
Table 2.5 Features selected by the stepwise linear discriminant analysis that best discriminate between the progressing lesions and regressing lesions.	48
Table 2.6 Results of the comparison of Kaplan-Meier survival function of two groups of cases with high and low Ki67 feature values. The comparison was done using a log rank test. The log-rank test statistic and the corresponding p-value are presented. <i>p</i> -values smaller than 0.05 are highlighted. LDA represents the linear discriminant function.	51
Table 2.7 Sensitivity and specificity of each Ki67 feature in identifying progressing and regressing lesions in two groups of cases with high and low Ki67 feature values. LDA represents the linear discriminant function.	54
Table 2.8 Hazard ratio and 95% confidence interval of each Ki67 feature obtained using the Cox proportional hazards model. LDA represents the linear discriminant function.	55

List of Figures

Figure 2.1 Voronoi tessellation (blue polygons) superimposed on the image of a tissue section. Green polygons represent the cells in the first layer above the basement membrane (basal layer). Pink polygons represent cells in the 4th layer. Yellow polygons are Voronoi neighbors of the nucleus marked with a yellow star.....	23
Figure 2.2 (a) Image of the ROI in a tissue section, (b) color-coded representation of distances (in microns) of the pixels of the ROI in (a) to the basement membrane. Nuclei centers are marked with black dots.	26
Figure 2.3 Estimated variances of different components (subject, inter-observer, and intra-observer) in the random effects ANOVA model, for different Ki67 features.	29
Figure 2.4 Box plot of <i>Ki67-index</i> grouped by Histology.	31
Figure 2.5 Box plots of <i>Layer-average</i> , and <i>Layer-90</i> grouped by Histology, Y-axis represents layer number.	32
Figure 2.6 Box plots of <i>Distance-average</i> and <i>Distance-90</i> grouped by Histology, Y-axis represents distance in microns.	33
Figure 2.7 Box plots of <i>Relative-distance-average</i> , and <i>Relative-distance-90</i> , grouped by histology.....	33
Figure 2.8 Box plots of <i>Relative-layer-average</i> , and <i>Relative-layer-90</i> , grouped by histology...	34
Figure 2.9 Box plots of Basal-layer-fraction, Lower-half-layer-fraction, and Upper-half-layer-fraction grouped by Histology.	34
Figure 2.10 Box plots of <i>Deep-third-fraction</i> , <i>Middle-third-fraction</i> , and <i>Upper-third-fraction</i> grouped by Histology.....	36
Figure 2.11 Box plot of <i>Positive-neighbor-number</i> grouped by Histology.....	36
Figure 2.12 Box plots of <i>Positive-neighbor-distance-overall</i> , and <i>Positive-neighbor-distance-positive</i> grouped by Histology, Y-axis represents distances in microns on logarithmic scale.	37
Figure 2.13 Box plot of <i>Positive-neighbor-fraction</i> , grouped by histology.	37
Figure 2.14 Box plot of <i>Positive-per-basal-length</i> , grouped by histology.....	38
Figure 2.15 Box plot of <i>Basal-positive-per-basal-length</i> , grouped by histology.....	38
Figure 2.16 Distribution of Ki67 expressing cells in the epithelium, for different histology grades. Box plots of the Ki67 positivity indices in 5-micron intervals above the basement membrane. Each box represents the inter-quartile range of the Ki67 positivity indices. The circles correspond to the median of Ki67 positivity indices. Medians are connected by a black line to demonstrate the pattern.	41
Figure 2.17 Distribution of Ki67 expressing cells in the epithelium, for different histology grades. Box plots of the Ki67 positivity indices in Voronoi cell layers above the basement membrane. Each box represents the inter-quartile range of the Ki67 positivity indices. The circles correspond to the median of Ki67 positivity indices. Medians are connected by a black line to demonstrate the pattern.	42
Figure 2.18 Box plot of <i>Distance-90</i> for the two groups of progressing and regressing lesions..	48
Figure 2.19 Box plot of the linear discriminant function values for the two groups of progressing and regressing lesions.	49

Figure 2.20 Kaplan-Meier survival functions for two groups of lesions with low and high <i>Ki67-index</i> values.....	51
Figure 2.21 Kaplan-Meier survival functions for two groups of lesions with low and high <i>Distance-average</i> values.....	52
Figure 2.22 Kaplan-Meier survival functions for two groups of lesions with low and high LDA values.....	53
Figure 3.1 Illumination profiles for the capture of (a) multispectral, and (b) RGB images.....	76
Figure 3.2 RGB Image of a typical <i>Ki67</i> sample.....	84
Figure 3.3 (a) manually obtained spectra from blue and brown regions in Figure 3-2,	85
Figure 3.4 Concentration maps of the stains obtained by the spectral unmixing algorithm, the left column shows concentration maps of H, and the right column shows concentration maps of DAB, (b) and (c) generated using S_{blue} and S_{brown} as initial estimates, (d) and (e) generated using S_{blue} and S_{brown} after the appropriate proportion of S_{blue} was subtracted from S_{brown} , (f) and (g) generated using initial concentration maps of H and DAB.....	87
Figure 3.5 Spectra estimated by the MCR-ALS algorithm,.....	88
Figure 3.6 Resulting spectra after subtracting increasing proportions of S_{blue} from S_{brown} , arrows show the directions of change at both ends of the spectrum, with increasing k in equation (3.15).....	89
Figure 3.7 Absorption spectra of blue and brown pixels, captured manually, and through the SIMPLISMA algorithm.....	91
Figure 3.8 RGB unmixing results, concentration maps of the stains obtained by the spectral unmixing algorithm, the left column shows concentration maps of H, and the right column shows concentration maps of DAB, (b) and (c) generated using S_{blue} and S_{brown} as initial estimates, (d) and (e) generated using S_{blue} and S_{brown} after the appropriate proportion of S_{blue} was subtracted from S_{brown}	94
Figure 3.9 RGB unmixing, spectra estimated by the MCR-ALS algorithm,.....	94
Figure 3.10 RGB image of a typical P16 sample.....	95
Figure 3.11 (a) manually obtained spectra from blue nuclei, brown cytoplasm, and brown nuclei regions in Figure 3.10, (b) spectra in (a) normalized by Euclidean norm.....	96
Figure 3.12 Absorption spectra of blue and brown pixels, captured manually, and through the SIMPLISMA algorithm.....	97
Figure 3.13 Concentration maps of the stains obtained by the spectral unmixing algorithm, the left column shows concentration maps of DAB, and the right column shows concentration maps of H, (a) and (b) generated using S_{blue} and S_{brown} as initial estimates, (c) and (d) generated using initial concentration maps of H and DAB, (e) and (f) generated using RGB spectral unmixing.....	98
Figure 3.14 P16 spectral unmixing results, Initial spectra, and the spectra estimated by the MCR-ALS algorithm.....	99
Figure 3.15 P16 RGB unmixing, initial and estimated spectra.....	100
Figure 3.16 Multispectral unmixing, concentration maps of the stains.....	101
Figure 3.17 RGB unmixing concentration maps of the stains.....	102
Figure 3.18 Spectra estimated from a sample with dark DAB staining, (a) multispectral, and (b) RGB.....	103
Figure 3.19 (a) spectra from blue nuclei, brown cytoplasm, brown nuclei, and dark brown nuclei regions in Figure 3-16 (a), (b) spectra in (a) normalized by length, (c) average of the spectra in (b), (d) RGB spectra.....	104

Figure 3.20 Dark DAB stain unmixing spectra. Dashed lines represent the initial estimates of spectra.	106
Figure 3.21 RGB Image of a typical CD8 sample	107
Figure 3.22 Spectra from a typical CD8 sample, (a) manually obtained spectra from blue and brown regions, (b) spectra in (a) normalized by length, (c) manually and automatically obtained reference spectra.....	109
Figure 3.23 Spectral unmixing on a typical CD8 sample, concentration maps of the stains.....	110
Figure 3.24 Spectra estimated from a CD8 sample, optimization started with (a) initial estimates of spectra, and (b) initial estimates of concentration maps.....	111
Figure 3.25 RGB Spectra estimated from a CD8 sample	112
Figure 3.26 Example of CD8 sample with dark DAB, incomplete removal of DAB signal from the H channel, arrows point to traces of DAB absorption attributed to the H channel	113
Figure 3.27 RGB unmixing results: concentration maps of (b) H, (c) VR, and (d) DAB. Note the incomplete unmixing of DAB and VR.	115
Figure 3.28 RGB unmixed spectra.....	116
Figure 3.29 Double immunostained sample, unmixed spectra.	116
Figure 3.30 Concentration maps of (b) H, (c) DAB, and (d) VR. Note the holes in (c) and (d) at the locations of the red nuclei.	118
Figure 3.31 Double immunostained sample, (a) S_{red} with increasing proportions of the S_{blue} subtracted from it, (b) S_{red} with increasing proportions of the S_{brown} subtracted from it. ...	119
Figure 3.32 S_{red} after the appropriate proportions of S_{blue} and S_{brown} are subtracted from it.	121
Figure 3.33 The two surfaces $S1p1, p2$ (shown in green) and $S2p1, p2$ (shown in pink) and their intersection curve (shown in blue), the point $p1, p2$, where the blue curve is equal to zero, is marked with a yellow dot.....	122
Figure 3.34 Concentration maps of (b) H, (c) DAB, and (d) VR, obtained using the initial estimates of spectra. The initial estimate of VR spectrum was obtained using the “subtraction algorithm”.....	123
Figure 3.35 Double immunostained sample, unmixed spectra. Spectral unmixing with initial estimates of spectra. Initial estimate for VR spectrum obtained using the “subtraction algorithm”.	124
Figure 3.36 (a) RGB image of a Ki67 & P16 sample, (b), (c), and (d) concentration maps of H, DAB, and VR. Unmixing done using initial estimates of spectra. Initial estimate of VR spectrum obtained using the “subtraction algorithm”.....	125
Figure 3.37 Unmixing results obtained with thresholded optical densities as initial estimates of concentration maps. Concentration maps of (b) VR, (c) H, and (d) DAB.	126
Figure 3.38 Double immunostained sample, unmixed spectra. Spectral unmixing with Initial estimates of concentration maps.	127
Figure 3.39 Coefficients of the linear discriminant functions obtained from the multispectral image stacks of Ki67 samples.....	133
Figure 3.40 Coefficients of the linear discriminant functions obtained from the RGB images of Ki67 samples.....	133
Figure 3.41 (a) RGB image of a Ki67 sample, (b) and (d) grayscale images representing nuclei and DAB generated by the linear discriminant functions obtained from the multispectral image stacks, (c) and (e) grayscale images representing nuclei and DAB generated by the linear discriminant functions obtained from the RGB images.....	134

Figure 3.42 Coefficients of the linear discriminant functions obtained from the multispectral image stacks of P16 samples.....	135
Figure 3.43 Coefficients of the linear discriminant functions obtained from the RGB images of P16 samples.	135
Figure 3.44 (a) RGB image of a P16 sample, (b) and (d) grayscale images representing nuclei and DAB generated by the linear discriminant functions obtained from the multispectral image stacks, (c) and (e) grayscale images representing nuclei and DAB generated by the linear discriminant functions obtained from the RGB images.....	136
Figure 3.45 Coefficients of the linear discriminant functions obtained from the multispectral image stacks of CD8 samples.	137
Figure 3.46 Coefficients of the linear discriminant functions obtained from the RGB images of CD8 samples.	137
Figure 3.47 (a) RGB image of a CD8 sample, (b) and (d) grayscale images representing nuclei and DAB generated by the linear discriminant functions obtained from the multispectral image stacks, (c) and (e) grayscale images representing nuclei and DAB generated by the linear discriminant functions obtained from the RGB images.....	138
Figure 3.48 (a) Coefficients of the linear discriminant functions obtained from the multispectral image stacks of Ki67 & P16 samples , (b) coefficients of the linear discriminant functions obtained from the RGB images of Ki67 & P16 samples.	140
Figure 3.49 (a) RGB image of a Ki67 & P16 sample, (b), (c), and (d) grayscale images representing nuclei, VR, and DAB generated by the linear discriminant functions obtained from the RGB images.	141
Figure 3.50 (a) RGB image of a Ki67 & P16 sample, (b), (c), and (d) grayscale images representing nuclei, VR, and DAB generated by the linear discriminant functions obtained from the multispectral image stacks, (e) grayscale image representing DAB and DAB+VR.	142
Figure 3.51 Coefficients of the linear discriminant functions for Ki67 & P16 samples. Red bars represent coefficients for the identification of DAB and DAB+H. Blue bars represent coefficients for the identification of DAB, DAB+H, and DAB+VR.....	143
Figure 3.52 (a) RGB image of a Ki67 & P16 sample, (b), (c), and (d) grayscale images representing nuclei, VR, and DAB generated by the linear discriminant functions obtained from the multispectral image stacks, (e) grayscale image representing DAB and DAB+VR.	144
Figure 3.53 (a) RGB image of the sample with the ROI marked on it,	149
Figure 3.54 Steps of the edge relocation algorithm: (a) gray-scale image, (b) cluster of objects in the initial binarized image, (c) object mask, (d) initial border, (e) dilated border, initial donut, (f) corrected donut (high gradient and/or low intensity objects that are connected to the donut are added to it), (g) thinned donut, (b) final objects after merging, and (j) final borders.....	152
Figure 3.55 Two examples of segmentation of clustered nuclei. The images on the left show the objects identified by the binarization method and images on the right show the final segmented objects.	155
Figure 3.56 Segmentation of a CD8 sample: (a) RGB image, (b) unmixed “nuclei image”, (c) unmixed “DAB image”, and (d) binarized DAB image – continued in Figure 3.56.....	157
Figure 3.57 Segmentation of a CD8 sample, continued from Figure 3.56, (e) areas in the binarized DAB image that were enclosed by ring objects marked with red, (f) initial	

binarization result, (g) initial binarization result corrected by incorporating information from (e), and (h) final segmentation result.	158
Figure 3.58 Ki67 segmentation example one: (a) initial binarization result, objects with white boundaries are further separated, (b) final segmentation result.....	159
Figure 3.59 Ki67 segmentation example two: (a) initial binarization result, objects with white boundaries are further separated, (b) final segmentation result.....	160
Figure 3.60 P16 segmentation example one: (a) initial binarization result, objects with white boundaries are further separated, (b) final segmentation result.....	161
Figure 3.61 P16 segmentation example two: (a) RGB image, (b) initial binarization result, objects with white boundaries are further separated, (c) final segmentation result.....	162
Figure 3.62 P16 segmentation example with dark DAB staining: (a) unmixed nuclei image to be segmented (b) final segmentation result superimposed on the RGB image.	163
Figure 3.63 CD8 segmentation example: (a) RGB image, (b) unmixed nuclei image, (c) initial binarization result, and (d) final segmentation result.....	164
Figure 3.64 Example of evaluating the segmentation accuracy, green dots are manually marked centers of nuclei, blue circles are centers of automatically segmented nuclei, objects with white borders are under-segmented objects, objects with black borders are over-segmented or falsely detected objects, and red stars are missed objects.	167
Figure 3.65 Example of how nuclei marked by observer one were compared against the reference, i.e. nuclei marked by observer two. Green circles are nuclei centers marked by observer two (the reference). Red dots are nuclei centers marked by observer one. Polygons represent the Voronoi tessellation based on the nuclei centers identified by observer one. Nuclei inside blue polygons are correctly identified. White polygons contain missed nuclei (false negatives). Dark blue polygons contain falsely detected objects (false positives). ..	169
Figure 3.66 Ki67 segmentation: F-scores of segmentation results.	170
Figure 3.67 P16 segmentation: F-scores of segmentation results.....	171
Figure 3.68 CD8 segmentation: F-scores of segmentation results.	172
Figure 3.69 Example of the nuclei map used for the measurement of the immunostain: Voronoi tessellation superimposed on the binarized image, foreground objects bound by each Voronoi polygon are considered as nuclei.....	181
Figure 3.70 Ki67 Example, (a) cell classification, white dot indicates a positive nucleus, and black dot indicates a negative nucleus, (b) map of the measured staining inside the ROI of each cell. Red dots represent negative nuclei and yellow dots represent positive nuclei... ..	182
Figure 3.71 P16 Example, (a) cell classification, red dots represent positive cells and blue dots represent negative cells, (b) map of the measured staining inside the ROI of each cell.	183
Figure 3.72 CD8 Example, (a) cell classification, white dot indicates a positive nucleus and black dot indicates a negative nucleus, (b) map of the measured staining inside the ROI of each cell.....	184
Figure 3.73 P16 Example two, (a) cell classification, red dot indicates a positive nucleus and blue dot indicates a negative nucleus, (b) map of the measured staining inside the ROI of each cell, (c) refined cell classification, taking into account the standard deviation of staining intensity inside the ROI of each cell, (d) map of the standard deviation of staining intensity inside the ROI of each cell. Arrows point to two positive cells that were identified as negative in (a).	185

Figure 3.74 Comparison of Ki67 immunostain images: Kruskal-Wallis test statistics of immunostain images for separating the positive and negative pixels shown for each sample in Ki67 set.	187
Figure 3.75 Comparison of P16 immunostain images: Kruskal-Wallis test statistics of immunostain images for separating the positive and negative pixels shown for each sample in P16 set.	187
Figure 3.76 Comparison of CD8 immunostain images: Kruskal-Wallis test statistics of immunostain images for separating the positive and negative pixels shown for each sample in CD8 set. Image 13 did not contain any positive pixels, so no chi-square statistic is reported for it.	188
Figure 4.1 RGB image of a typical Ki67 immunostained sample, captured with three-channel illumination generated with OneLight light engine.	202
Figure 4.2 Linear discriminant function coefficients obtained from the multispectral image stack of the Ki67 sample, (a) for the identification of nuclei and (b) for the identification of DAB stain.	202
Figure 4.3 Linear discriminant function coefficients obtained from the RGB image of the Ki67 sample, (a) for the identification of nuclei and (b) for the identification of DAB stain.	203
Figure 4.4 Illumination profiles, (a) multispectral, and (b) three-channel RGB.	204
Figure 4.5 Illumination profiles for the capture of linear discriminant images, based on the multispectral image stack, (a) for the identification of nuclei, and (c) for the identification of DAB, (b) and (d) demonstrate how the illumination profiles in (a) and (b) were partitioned into positive and negative segments.	205
Figure 4.6 Illumination profiles for the capture of Linear Discriminant Images, based on the three-channel RGB image, (a) for the identification of nuclei, and (c) for the identification of DAB stain, (b), and (d) demonstrate how the illumination profiles in (a) and (b) were partitioned into two segments: positive, and negative, for the capture of two images under each.	206
Figure 4.7 (a) Illumination profiles generated with OneLight light engine for the identification of nuclei from multispectral image stack, (b) illumination profiles after properly scaled to resemble the designed illumination profiles, (c) and (d) generated vs. designed positive and negative illumination profiles.	209
Figure 4.8 Capture of the image of nuclei, with illumination generated from the multispectral image stack, (a) image captured under the positive segment of the designed illumination, (b) image captured under the negative segment of the designed illumination.	210
Figure 4.9 Image of nuclei, generated with illumination designed from the multispectral image stack.	211
Figure 4.10 (a) Illumination profiles generated by OneLight light engine for the identification of DAB from multispectral image stack, (b) illumination profiles after being corrected for OneLight scaling factor, (c) and (d) generated vs. designed positive and negative illumination profiles.	212
Figure 4.11 Capture of image of DAB, with illumination generated from the multispectral image stack, (a) image captured under the positive segment of the designed illumination, (b) image captured under the negative segment of the designed illumination.	213
Figure 4.12 Image of DAB, generated with illumination designed from the multispectral image stack.	214

Figure 4.13 (a) Illumination profiles generated by OneLight light engine for the identification of nuclei from RGB image stack, (b) illumination profiles after being corrected for OneLight scaling factor, (c) and (d) generated vs. designed positive and negative illumination profiles.	215
Figure 4.14 Image of nuclei, generated with illumination designed from the RGB image.	216
Figure 4.15 (a) Illumination profiles generated with OneLight light engine for the identification of DAB stain from RGB image stack, (b) illumination profiles after being corrected for OneLight scaling factor, (c) and (d) generated vs. designed positive and negative illumination profiles.	217
Figure 4.16 Image of DAB, generated with illumination designed from the RGB image.	218
Figure 4.17 RGB image of the second Ki67 immunostained sample.	218
Figure 4.18 Images of nuclei (a) and DAB (b), generated with illumination designed from the multispectral image stack.	219
Figure 4.19 Images of nuclei (a) and DAB (b), generated with illumination designed from the RGB image.	220
Figure 4.20 EEM of normal cervical tissue. Blue horizontal line represents an emission profile, and red vertical line represents an excitation profile. Photo obtained with permission from private communication with Dr Calum MacAulay and Dr Michelle Follen.	226
Figure 4.21 Schematic diagram of the system setup.	228
Figure 4.22 PCA composite image of the multi-color immunofluorescence labeled tissue section.	239
Figure 4.23 One hundred randomly selected regions in the image stack, for the extraction of purest spectra through SIMPISMA analysis.	240
Figure 4.24 Purest spectra obtained from one hundred regions, from left to right, top to bottom: 1 st , 2 nd , 3 rd , and 4 th purest spectra in each region. The relative total intensity of the standard deviation spectra for each one of the four purest spectra for each region was calculated, and then averaged across all images. These average values are presented on their corresponding plots.	241
Figure 4.25 (a), (b), and (c) unmixed concentrations of 1 st , 2 nd , and 3 rd components. The third component is tissue autofluorescence. (d) initial spectra for the three components (dotted line) and pure spectra calculated through unmixing.	242
Figure 4.26 (a) Composite RGB image with 1 st , 2 nd , and 3 rd components assigned to Red, Green, and Blue channels, (b) Composite RGB image of only the first two components (autofluorescence removed).	243
Figure 4.27(a) PCA composite image of AO-labelled cells, (b) fifty selected regions in the image on which the SIMPLISMA analysis was performed to extract purest spectra.	244
Figure 4.28 Purest spectra obtained from fifty regions. From left to right, top to bottom: 1 st , 2 nd , 3 rd , and 4 th purest spectra in each region. The relative total intensity of the standard deviation spectra for each one of the four purest spectra for each region was calculated, and then averaged across all images. These average values are presented on their corresponding plots. Note that the third and fourth values are significantly lower than the first two.	245
Figure 4.29 (a) and (b) unmixed images of AO fluorescent signal based on its different absorption spectra in the nucleus and cytoplasm, (c) initial spectra obtained from the image stack (dotted line) and pure spectra obtained through unmixing.	246
Figure 4.30 Excitation-Emission Images of the FISH-labeled tissue section.	248

Figure 4.31 Unmixing results from excitation image stack performed using the initial spectra directly obtained from the image. (a) and (b) unmixed images of tissue autofluorescence and FITC fluorophore, (c) initial spectra (dashed line) and estimated spectra (solid line) obtained through unmixing. Note the incomplete unmixing (holes in the autofluorescence image in the location of FITC-labeled nuclei).....	249
Figure 4.32 Unmixing results from excitation image stack, performed after the appropriate fraction of tissue autofluorescence spectrum is subtracted from the FITC spectrum, to remove the holes in the autofluorescence image of Figure 4-30(a). (a) and (b) unmixed images of tissue autofluorescence and FITC fluorophore. (c) initial spectra, and pure spectra obtained through unmixing. Note the homogeneous autofluorescence image.....	250
Figure 4.33 Unmixing results from emission image stack, performed after the appropriate fraction of tissue autofluorescence spectrum is subtracted from the FITC spectrum, to remove the holes in the autofluorescence image of Figure 4-30(a). (a) and (b) unmixed images of tissue autofluorescence and FITC fluorophore, (c) initial spectra, and pure spectra obtained through unmixing.....	251
Figure 4.34 (a) Narrow-band illumination spectra, (b) PCA composite image.....	252
Figure 4.35 (a), (b), and (d) Designed illumination profile for the capture of 1 st , 2 nd , and 3 rd principal component image, (c) illumination spectrum in (b), with positive and negative portions separated and the negative portion inverted, and (e) illumination spectrum in (d), with positive and negative portions separated and the negative portion inverted.....	253
Figure 4.36 Generated illumination profile for the capture of 1 st principal component image. .	254
Figure 4.37 (a) and (b) Generated illumination profiles for the capture of 2 nd and 3 rd principal components,	255
Figure 4.38 (a) and (b) 1 st principal component image, (c) and (d) 2 nd principal component image, (e) and (f) 3 rd principal component image, left column: calculated images, right column: images captured through SELF.....	256
Figure 4.39 Calculating the SELF illumination profiles: (a) unmixed spectra, (b) pseudo-inverse of the spectra matrix (to be multiplied by exposure times and narrow-band illumination channels to generate SELF illumination), (c) narrow-band illumination channels, and (d) calculated illumination profiles for the capture of the 1 st and 2 nd unmixed components. ..	258
Figure 4.40 (a) and (b) Designed illumination profiles for the capture of unmixed components, (c) and (d) generated illumination profiles, (e) and (f) generated illumination profiles adjusted for OneLight scaling factor, left column, (a), (c), and (e), corresponds to the first component, right column, (b), (d), and (f), corresponds to the second component.....	259
Figure 4.41(a) and (b) Unmixed images of AO fluorescent signal. (c) and (d) same images captured using SELF imaging.....	261
Figure 4.42 Composite RGB images of a healing finger wound, (a) and (b) show the same area with different illumination angles. Arrows point from palm to fingertip. Note that the colors of the same features stay the same regardless of the illumination angle.....	262
Figure 4.43 Composite RGB image of each sample obtained using coefficients from the other sample. (a) and (d) White light images of healing wounds, (b) and (e) PCA composite images generated from the corresponding excitation image stack, (c) composite PCA image of (b) generated using the PCA coefficients obtained from (e), and (f) composite PCA image of (e) generated using the PCA coefficients obtained from (b).....	263

Acknowledgements

I would like to express my deep appreciation and gratitude to my supervisor, Dr. Calum MacAulay, for the patience, guidance, valuable mentorship and continuous encouragement he provided to me. His dedication and enthusiasm about science and his frequent reassurances that I was not the only one going through this path kept me going. I always ended our meetings feeling as though I could tackle any problems.

I would like to thank my co-supervisor, Dr. Rabab Ward, for giving me the opportunity to do my graduate studies at the University of British Columbia, for all her assistance when I received a student visa to come to Canada, and for her continued support over the years.

I am also thankful to the members of my PhD departmental exam, Dr. Jose Marti, Dr. Jane Wang, and Dr. Shuo Tang for their useful feedback. I am thankful to Dr. Timon Buys for his time in reviewing this thesis. I would like to acknowledge the help of the scientists and staff at the Cancer Imaging department of the BC Cancer Research Center. Specially, I am thankful to Judit Banáth, Anita Carraro, Deanna Ceron, Dr. Martial Guillaud, Jagoda Korbelik, Sylvia Lam, Dr. Anthony Lee, and Dr. Michael Short.

The financial support for this thesis was partly provided by the Canadian Institute of Health Research, the Canadian Institute for Photonic Innovations, and the British Columbia Innovation Council, and I thank them for their investment.

Dedication

To my dear Derakhshandeh and Ebrahim,
To get here, I have paid the highest price of living so far away from you. Every single day, I
am missing you.

Chapter 1. Introduction

1.1 Personalized Medicine and Molecular Biomarkers

The aim of this thesis is to develop methods for the quantitative assessment of expression of labeled molecular biomarkers in tissue sections. This is a necessary step towards the ultimate goal of personalized treatment of cancer.

Cancer is the leading cause of premature death in Canada. 1,048,900 years of life were lost in 2006 as a result of cancer. In 2007, cancer surpassed cardiovascular disease as the leading cause of death in Canada. An estimated 186,400 new cases of cancer and 75,700 deaths will occur in Canada in 2012.¹

Almost every type of human cancer is comprised of biological subsets that differ in their clinical behavior and their response to treatment. There are many important examples of treatment regimens that produce better results in some tumor subtypes than others.² One example of tumor subtypes that must be recognized in order to optimize the treatment is breast cancer tumors positive for estrogen receptor (ER) or Human Epidermal Growth Factor Receptor 2 (HER2), i.e. tumors that consist of cells that express ER or HER2.³ Another example is non-small-cell lung cancer (NSCLC) tumors with mutations in the Epidermal Growth Factor Receptor protein (EGFR). Mutations in the EGFR protein can identify patients with NSCLC who are likely to have a good response to a particular treatment targeting the EGFR domain.⁴

The biological heterogeneity of human cancer and the inter-individual variations in cancerous tumors and in the human genome call for a personalized approach to the treatment of cancer. The expectation is that this approach will improve treatment efficacy, reduce toxicity and minimize cost.²

Personalized medicine for the management and treatment of cancer requires the discovery and use of prognostic, predictive and pharmacodynamic biomarkers to inform therapeutic decisions. Biomarkers are generally quantifiable molecules or processes indicative of a certain biological state or condition. In the context of cancer, various molecular analytes (such as DNA, mRNA, microRNA, and proteins) and physiological processes (e.g. angiogenesis and proliferation) have proven useful for cancer detection and management.

Ultimately, all these genetic and molecular alterations result in altered expression of protein products. One approach to the assessment of the expression of such proteins in a tissue is to employ antibody-based assays, such as immunohistochemistry (IHC).⁵

IHC is a method used to determine in which cells or parts of cells a particular protein is located. It uses antibodies as specific reagents to bind to antigens (proteins) through specific antigen-antibody interactions. In order to detect these bound antibodies, antibodies are “labeled”. In the most common instance, a very thin slice of tissue is fixed onto glass, incubated with antibody or a series of antibodies, the last of which is chemically linked to (labeled with) an enzyme, such as hydrogen peroxidase. The enzyme catalyzes a color producing reaction by adding a chemical substrate, such as Diamonobenzidine (DAB). After the color is developed, the distribution of the enzyme label can be examined by microscopy. The detection of the enzyme label in a cell or part of a cell is indicative of presence of the target protein in that cell. Alternatively, the antibody can be labeled with a fluorophore, radioactive element, or nano-particles.⁶

1.2 Need for Automated Tissue-based Biomarker Analysis

New molecular biomarkers for diagnosis, screening, or predicting response to treatment are being identified every day. Many of these molecular biomarkers need to be identified in intact cells and tissue. Tissue-based detection of molecular biomarkers is therefore important. In addition, the biological complexity of tumors requires more than one molecular biomarker to be studied at once, allowing the examination of the relationships between different biomarkers. Multiplexed studies that report on multiple biomarkers simultaneously are needed to untangle the biology underpinning diseases and to improve clinical performance. An example of such a multiplexed study is immunohistochemistry (IHC) experiments for simultaneous detection of two or more antigens, using a different label for each antigen. High throughput tissue microarrays and whole-slide digital scanners are increasingly applied in pre-clinical research to improve the reproducibility and utility of tissue pathology. All these factors place great demands on automated image analysis methods for the quantitative assessment of molecular biomarkers.

1.3 Need for Cellular-Level Analysis of Tissue-based Biomarkers

Cancer is a disease of the genome involving changes in genetic and epigenetic structures, which are transferable to subsequent generations of neoplastic progeny. Cancer initiation and progression is caused by the evolution of expanding genetically modified clones of cells. These cells undergo selective population pressure⁷ such that they gain a selective growth advantage over normal cells. This selective growth advantage happens as a result of cells accumulating specific alterations in at least two groups of genes, the proto-oncogenes and the tumor suppressor genes (i.e. genes that protect a cell from a step on the path to cancer and genes that have the potential to gain mutations that can cause cancer).⁸ Modern molecular technology has made it possible to identify many of these genetic alterations in the human tissue.

A single region within a tumor does not represent the genetic state and/or expression profile of the tumor as a whole. The variation in a tumor's genotype and/or phenotype is known as tumor heterogeneity.⁹ Intra-tumor heterogeneity has been reported in many types of cancers and has been shown to have important clinical implications.^{10,11}

There is a need for methodologies to identify the clonal populations of cells with genetic and molecular signatures/profiles associated with dangerous biological behaviour. These need to be readily measured across the entire lesions in a high throughput fashion. Immunohistochemistry is one of the existing methodologies by which these forms of measurements can be obtained at the individual cell level, in a high throughput fashion, across an entire lesion.

Ultimately the efficient detection of the clone or clones that are relevant to available treatment options will be required to enable genetically based personalized medicine to be of utility to the majority of cancer patients.

For a true measure of tumor heterogeneity, it is necessary to find the means for the detection of distinct clonal populations of cells with varied biological attributes. This should be attainable down to the individual cell level and across entire tumor/tissue sections. Recognition of such clonal population of cells requires the identification of the cells sharing specific attributes over a specific domain area. The spatial associations of cells are determined by geometric distance measures.

It has been suggested that cell-by-cell analysis increases the biological relevance of molecular biomarker measurement. As an example, Emily et al.¹² measured the protein expression of DARPP-32 using IHC in a series of 31 breast cancer patients to differentiate between patients who remained disease free after 5 years and those who experienced recurrence or death within 5 years. This study showed that while a mean measure of protein expression could detect poor prognosis for patients in 83% of cases, it did so with a poor specificity of 44%. This was in contrast to the use of a cell-by-cell based heterogeneity measure, which had the same detection rate of 83% while maintaining a specificity of 76%.

1.4 State of the Art in Quantitative Analysis of IHC

For immunohistochemistry (IHC) stained tissue sections, the simplest method of biomarker quantification is visual inspection by a pathologist. Quantification involves determining the expression levels of proteins in an area of interest within a tissue section. In general, expression of these biomarkers is usually scored on an ordinal 0 – 3 scale. This score, in some cases, is combined with a scored interpretation of the overall labeling intensity of the biomarker.¹³

At best, manual inspection is only semi-quantitative and reduces continuous biomarker expression to an ordinal scale. Visual inspection can also be confounded by the inherently subjective nature of human observation. It can, for example, be affected by factors such as the intensity of staining, the area of the tumor present, non-specific background staining, and stromal staining. These issues can lead to low inter- and intra-observer variability.¹⁴ In some cases, subtle sub-populations may not be identified using manual analysis. Visual scoring is also time and labour intensive, and since the introduction and the increased use of whole slide images and high throughput tissue microarray slides, it is becoming a limiting factor.

The combination of IHC and computer-assisted image analysis systems provides the possibility of objective and reproducible quantification of IHC staining and clonal distribution.

Several commercially available software and hardware systems for the quantitative assessment of immunohistochemistry (IHC) using computer-assisted image analysis exist. Some of them are offered as part of a whole imaging and analysis system (whole slide scanners or microscope-based imaging systems), while others are dedicated software

packages. Some of these have clearance by the USA Food and Drug Administration (FDA) for testing specific biomarkers. The most commonly seen application of image analysis-based quantification of IHC labels is for the quantification of human epidermal growth factor receptor 2 (HER2)/neu expression. Other current applications include assessment of the percentage of ER-, PR-, and Ki67-positive nuclei.¹⁵ One such system is the Automated Cellular Imaging System (ACIS) (Dako, Denmark A/S), which has FDA approval for detection of HER2/neu protein over-expression in breast cancer tissues. Virtuoso (Ventana Medical Systems inc., Tucson, AZ, US) has FDA approval for the progesterone receptor (PR) protein IHC. ScanScope XT System (Aperio Technologies, Inc. Vista, CA) has FDA approval for ER and PR IHC.

Recent studies have shown automated image analysis systems to be an effective substitute for manual scoring: Rizzardi et al.¹⁶ assessed 215 ovarian serous carcinoma specimens stained for S100A1 using Genie Histology Pattern Recognition software, (Aperio Technologies, Vista, CA, USA). They demonstrated that automated analysis can produce data that are highly similar to that obtained by visual evaluation of a pathologist. Mohammed et al.¹⁷ compared visual and computerized image analysis of HER2 IHC on tissue microarrays for 431 breast cancer patients and demonstrated that computerized image analysis of HER2 IHC gives results comparable to those obtained by visual assessment. The Slidepath Tissue image analysis system, version 3.0 (Slidepath, Dublin, Ireland) was used for the automated image analysis. In the study of Diaz et al.¹⁸, estrogen receptor was quantified on invasive breast cancer tumors. Visual scores were compared with those obtained using the QCA image analysis system (QCA, Lake Bluff, IL, USA), and it was shown that manual and computer-aided scoring techniques are comparable.

The degree of immunostaining, i.e. the expression of an IHC biomarker, can be measured using an area-based or cell-based methodology. With an area-based measure of immunostaining, for each IHC biomarker, the area within a tissue section that is labeled by that biomarker's corresponding label is reported at the pixel level, along with the total area of the region of interest. With the cell-based approach, the staining is reported at the individual cell level. i.e. for each IHC biomarker, the number of cells that are labeled with that biomarker's corresponding label is reported, along with the total number of cells within the region of interest in the tissue section. This approach is advantageous since cells are the

fundamental units in which the biological processes occur. Knowledge of the presence of molecular biomarkers in each cell or cell compartment allows for the identification of spatially associated/connected cells with the same molecular characteristics. This potentially useful information is discarded in the area-based approach. The biological significance of this and other patterns of biomarkers revealed quantitatively on a per cell basis is currently unknown for most applications. Development of tools to enable the exploration of the molecular inter-relationships in individual cells using multi-color IHC-based techniques allows the assessment of biological significance of such information. Most traditional image analysis applications, including those reported in the previous paragraph (studies of Rizzardi et al.¹⁶, Mohammed et al.¹⁷, and Diaz et al.¹⁸) still use the area-based approach. This may be due to the fact that the cell-based approach requires the delineation (segmentation) of cell nuclei or other cell compartments (cytoplasm and membrane), which despite all the progress in imaging and image analysis techniques has proven challenging and remains an active area of research.

Only recently have there been a few reports on cell-based analysis of IHC biomarkers using automated image analysis systems. Gokhale et al.¹⁹ assessed the accuracy of two automated systems, the ChromaVision Automated Cellular Imaging System and the Applied Imaging Ariol SL-50, and compared them to manual scoring by performing immunostaining for ER on 64 cases of breast cancer. The percentages of positive cells were scored by the two imaging systems. Their study indicates that ER scores calculated by the Ariol SL-50 and ACIS automated systems almost perfectly correlate with the manually determined semi-quantitative scores. In another study, the commercial image analysis system Ariol (Applied Imaging Inc., San-Jose, CA) is compared to visual scoring of ER in invasive breast cancer. The percentage of positive tumor cells was calculated using the automated program. It was reported that the visual and the automated scores were highly consistent.²⁰

More recently, the use of multispectral imaging in digital pathology has been explored. Multispectral imaging is the capture of image data at multiple specific wavelengths across the light spectrum. The wavelengths may be separated by optical filters or by the use of instruments that are sensitive to particular wavelengths, e.g. liquid crystal tunable filters. Multispectral imaging can allow extraction of additional information the human eye fails to capture with its receptors for red, green and blue. A few automated image analysis systems

are available that enable multispectral image acquisition and analysis of histopathological tissue sections. One example of such a multispectral imaging system is the Nuance system (Caliper Life Sciences, Hopkinton, MA, USA), which is a multispectral imaging system and analysis software. The system uses a liquid crystal tunable filter. Nuance software incorporates tools for building custom spectral libraries as well as unmixing and analysis algorithms allowing spectral characterization and quantification for each of the multi-labeled components in an image. This system was used in the study of Fiore et al.²¹ for the quantification of cytoplasmic biomarker Stathmin and the nuclear biomarker Ki67 on tissue microarrays of prostate adenocarcinoma. This study suggested some beneficial role for the use of a multispectral imaging system in the automated analysis of immunohistochemistry.

In summary, the quantification of biomarkers in IHC stained tissues is possible using automated analysis. For this analysis, either the RGB images or multispectral images of tissue sections can be used. Although some of the newer automated image analysis systems claim to perform cell-counting according to morphological and color criteria; these features, however, are not yet widely used. Only a few studies, such as those done by Gokhale et al.¹⁹ and Turbin et al.²⁰, validate the capability of some of these systems in performing cell-based analysis of IHC images. There is a need for automated analysis systems that are robust and easy to use. Such systems would enable the wider use of cell-level quantitative assessment of molecular biomarkers. The analysis methods for the extraction of the quantitative data need to be validated. The validation should not be done through the comparison of the quantitative data generated by the automated system with those obtained by the visual semi-quantitative methods, as has been done in a few recent studies. It should be done through the association of the obtained biomarker data with the biological and clinical information, such as correlation of the biomarker data with progression or regression of a tumor, or with the survival rates of the patients, etc.

1.5 Aims of the Thesis

For achieving personalized treatment and management of cancer, cell-level identification of clonal populations of cells with relevant molecular profiles need to be attainable across tumors/tissue sections. For the recognition of clonal population of cells, the cells which 1) share the same molecular profile (based on one or more molecular biomarkers), and 2) are in

close proximity to each other (based on some geometric distance measure) need to be identified. This involves the following two tasks: the first task is the quantitative assessment of labeled molecular biomarkers on a cell-by-cell basis within a tissue section.

The second task is the assessment of the spatial distribution of the cells that express molecular biomarker of interest within a tissue section.

This dissertation focuses on the above mentioned tasks. Specifically, the goals of this dissertation are as follows:

1) Our first goal is to develop methods for finding and quantifying the spatial associations of the cells, which express a molecular biomarker, within a tissue section. We hypothesize that quantification of such spatial associations increases the correlation (association) of biomarker expression data with biological behaviour of a lesion. Currently, research has mostly concentrated on the measurement of the expression of the molecular biomarkers on a “global” basis, i.e. assessing the total amount of the molecular biomarker within an entire area of interest in a tissue section. The spatial associations of cells that express a biomarker have not been commonly assessed.

2) Our second goal is to develop methods for the automated and quantitative assessment of the amount (expression) of tissue-based molecular biomarkers on a cell-by-cell basis. This involves the automated identification of the locations of cell nuclei (i.e. cell nuclei segmentation) followed by the assessment of the biomarker abundance within each cell’s nucleus, cytoplasm, or membrane. We hypothesize that the use of multispectral imaging combined with proper analysis methods improves the results of cell nuclei identification as well as biomarker measurement, compared to three-color RGB imaging.

3) Finally, our goal is to develop an imaging system that can significantly reduce the number of captured images for the identification of biomarkers in tissue, compared to the number of images captured to generate a multispectral image. This is based on the design of a specific light spectrum for illuminating the tissue sample. We refer to an illumination spectrum designed this way as “selective spectral illumination”. The hypothesis is that imaging a tissue using selective spectral illumination, which is specifically designed for the purpose of identifying a certain biomarker, would enable the extraction of relevant information about that biomarker with a smaller number of captured images, compared to multispectral imaging.

1.5.1 Quantification of the Spatial Distribution of Biomarkers

In Chapter 2, we develop methods of objectively and quantitatively assessing the distribution of a molecular biomarker within tissue. This is done through the characterization of spatial associations of cells using two geometric distance metrics; one that measures the distance between two cells in terms of the physical length and another that measures that distance in terms of the number of cells. Based on these two spatial characterizations, new features that explain the spatial distribution of a biomarker within the tissue will be presented.

We have applied the quantification procedure we developed to perform the first ever analysis of spatial distribution of Ki67-expressing cells in normal and pre-neoplastic bronchial epithelial regions.

The presence of pre-neoplastic lesions of the bronchus increases the likelihood of developing Lung cancer. However, there are no validated endpoint biomarkers that can reliably predict lung cancer incidence at present. The utility of Ki67 expression, a proliferation biomarker in conjunction with histological grade to predict the malignancy potential of a pre-neoplastic lesion has recently been studied by several research groups.²² We demonstrate that the spatial distribution of Ki67 expressing cells increases the relevance of Ki67 expression in predicting the outcome of bronchial epithelial lesions, over the use of the “global” ki67 expression (expression of Ki67 within the entire region of interest in a tissue section). Our newly proposed Ki67 features have higher correlation with the biological outcome, i.e. progression or regression, of the pre-neoplastic lesions.

1.5.2 Quantification of the Expression of Biomarkers in Tissue Sections

For IHC experiments to be able to make biomarker measurements at the individual cell level in a high throughput fashion, automated image analysis systems are needed to robustly and reliably quantify IHC biomarkers on a cell-by-cell basis.

Multispectral imaging has been recently employed to improve and increase the capabilities of computerized image analysis for the quantification of tissue-based biomarkers such as IHC biomarkers.¹⁴ The extra information provided by the multispectral images over the three-channel RGB images could benefit cell-by-cell measurement of biomarkers, in particular for multiple labeled specimens. The degree to which quantitative tissue-based

biomarker analysis may benefit from multispectral imaging has not yet been investigated. Similarly, the techniques through which multispectral images of histopathological samples can be analyzed are not fully explored, nor streamlined.

In chapter 3, we present methods for the analysis of multispectral images of IHC stained tissue sections. In a tissue section with N labels, this analysis involves generating N images from the multispectral stack of images, where each image represents one of the N labels. In each IHC tissue section, one of the N labels is a counterstain that is used to label all the nuclei. The image corresponding to that label (the nuclei counterstain) represents the cell nuclei. This image will be used for the automated identification and segmentation of nuclei. The remaining $N-1$ labels correspond to $N-1$ biomarkers. In the samples available in our study, either one or two labeled IHC biomarkers were present. The image representing the label of each IHC biomarker will be used to obtain the abundance of that IHC biomarker within each cell.

If the IHC biomarker is expressed in a cell compartment other than the nucleus (i.e. cytoplasm or membrane), then the location and boundaries of that cell compartment needs to be identified prior to the measurement of IHC biomarker within that cell compartment. For the localization of the cell compartments, we partition the area of interest within a tissue section into regions equivalent to cells. This partitioning is done using the locations of the cell nuclei and is used as an approximation of the locations of the cells, since in the absence of a label that marks the entire boundary of all cells, the cell boundaries are not visible. Therefore, the exact location of cells cannot be determined.

Multispectral imaging and RGB imaging are compared with respect to the accuracy of nuclei segmentation and IHC biomarker measurement.

1.5.3 Design of “Selective Spectral Illumination” for Imaging of Biomarkers

Multispectral imaging involves the sequential capture of several images of the sample. These images are typically acquired using several different narrow-band wavelength channels, i.e. the light transmitted through (in transmission imaging) or emitted from (in fluorescence imaging) the specimen is detected in several narrow-band wavelength channels. Alternatively, multispectral images can be acquired using several different narrow-band

illumination channels, i.e. sequential illumination of the sample with a number of spectrally narrow illumination spectra, followed by the image capture at each illumination channel.

In chapter 4, we employ the second multispectral imaging approach to acquire multispectral images of tissue samples. We demonstrate how a weighted sum of these multispectral images can generate a representation of a particular component of interest in a tissue sample. This approach, however, requires the acquisition and processing of a large number of images. We then explore the possibility of capturing a fewer number of images for the identification of the particular component of interest (e.g. labeled biomarker or labeled nuclei) in the tissue. This is done by designing “selective spectral illumination” spectra. Such a selective spectral illumination spectrum is calculated as a weighted sum of the spectrally narrow illumination spectra which were used for the capture of the multispectral images. This weighted sum spectrum of light could be generated by a spectrally programmable light engine which has the ability to generate a mixture of spectrally narrow illumination bands with controlled intensity at every spectral illumination band. We hypothesize that using this imaging approach we can capture the most relevant information from all spectral bands using a smaller number of images, compared to the number of images acquired using the conventional multispectral stack of images. In our approach, usually two captured images are required for the identification of a component of interest within a tissue sample.

1.6 Dissertation Organization

This dissertation is composed of five chapters. Chapter 1 (this chapter) is a general introduction and provides the motivation and significance of the research.

Chapter 2 focuses on the first aim of this research: to develop a framework for the quantification of the spatial distribution of biomarkers within the tissue. Since this framework is demonstrated on Ki67 expressing cells within the epithelium of lung precancerous lesions, some background on lung cancer and cell proliferation is provided. The framework is then presented, followed by its application to the case of Ki67 expression in lung preneoplastic lesions. The biological relevance of the new Ki67 features is studied and presented.

In Chapter 3, we present the quantitative analysis of tissue-based biomarkers at the cellular level through multispectral imaging. Methods for the analysis of multispectral

images are presented. A fully automated procedure for the segmentation of nuclei is then presented, followed by methods for the cell-based assessment of the abundance of biomarkers.

Chapter 4 presents our new imaging approach for the targeted imaging of biomarkers, in both transmission and fluorescence imaging modes.

Chapter 5 provides a summary of the results and conclusions for the research presented in this dissertation.

Finally, appendices A and B present additional information on the results reported in chapter 3.

Chapter 2. Quantitative Assessment of Spatial Distribution of Tissue-based Biomarkers

2.1 Introduction

The neoplastic process is a cumulative acquisition of genetic and epigenetic alterations which through selective local evolution gives rise to a population or populations of epithelial cells with improved cellular survival characteristics.²³ Evidence of this clonal selection and evolution process can be evident in the spatial distribution of the cells exhibiting molecular profiles associated with altered biological behaviors. The assessment of a molecular biomarker of interest is therefore not complete without considering the spatial distribution of that biomarker with regard to the histological structures within the tissue and the spatial inter-relations of the cells expressing the biomarker. However, most scoring systems for the assessment of tissue-based biomarkers provide global data only, such as percentage of cells, within an area of interest in a tissue section, expressing a particular biomarker; the spatial distribution of the biomarker is disregarded.

We developed a framework for the objective and quantitative assessment of the distribution of a molecular biomarker within tissue. In this framework, we use geometric measures of distance to define mathematical structures that characterize the architecture of tissue and arrangement of cells. Quantitative features that explain the location of the biomarkers with respect to such structures are proposed.

We used this framework to evaluate the spatial distribution of proliferating cells, as identified by immunohistochemical staining of Ki67, a proliferation biomarker, in normal and pre-neoplastic epithelial regions. To date, the spatial distribution of proliferating cells in a Ki67-stained biopsy cross-section had not been evaluated to determine whether this information had utility for disease grading and patient management in the context of early stage lung cancer.

Cell proliferation, as determined by immunohistochemical staining for Ki67, has been studied in a number of lung cancer prevention clinical trials as a potential “surrogate endpoint biomarker”. The clinical endpoint, in lung cancer chemoprevention studies, is the

occurrence of lung cancer. A “surrogate endpoint biomarker” is a biomarker that is intended to substitute the clinical endpoint. We hypothesized that the knowledge of spatial distribution of Ki67 expression would increase the biological relevance of Ki67 expression as a surrogate endpoint biomarker in bronchial epithelial lesions. We studied the relevance of the spatial distribution of proliferating cells to the histological grade, and more importantly to the malignant potential of the lesion.

This chapter is organized as follows: Section 2.2 provides some background on lung preneoplastic lesions, and the related research on the Ki67 expression in lung preneoplasia. Section 2.3 provides an explanation of the data used in this study. Section 2.4 presents the framework for the assessment of spatial distribution of cells, and the proposed features explaining this spatial distribution. Section 2.5 presents the results of the application of this framework to the case of Ki67 expression in lung precancerous lesions. The results are discussed in section 2.6. Finally section 2.7 provides a summary of the results.

2.2 Background

2.2.1 Lung Cancer

Lung cancer remains the leading cause of cancer death for both men and women. According to Canadian Cancer Society, in 2012, an estimated 25,600 Canadians will be diagnosed with lung cancer and 20,100 will die of it.²⁴ The American Cancer Society's estimates for lung cancer in the United States in 2012 indicate that about 226,160 new cases of lung cancer will be diagnosed, and there will be an estimated 160,340 deaths from lung cancer, accounting for about 28% of all cancer deaths.²⁵

Based on the Surveillance Epidemiology and End Results (SEER) Cancer Statistics Review of the National Cancer Institute (NCI), the overall 5-year relative survival for 2002-2008 from 18 SEER geographic areas was 15.9%. The 5-year relative survival by stage at diagnosis for 2002-2008 are as follows: localized (confined to primary site) is 52.2%, regional (spread to regional lymph nodes) is 25.1%, and distant (cancer has metastasized) is 3.7%.²⁶ The outcomes for this disease are significantly better for earlier stage tumors relative to advanced ones.

If lung cancer could be identified earlier, at a preneoplastic stage, before angiogenesis, invasion and micro-metastasis can occur, there would be a greater chance of improving

survival rates.²⁷ Experience with other epithelial cancers, such as uterine, cervical, esophageal, and colon carcinomas, suggests that the detection of the disease at premalignant stages will afford better survival rates.²⁸ While the majority of lung cancer cases are diagnosed at inoperable metastatic stages, much effort is being made to establish screening programs for disease detection at precancerous stages.²⁹

Lung cancers are classified according to their histological type. This classification has important implications for clinical management and prognosis. Most primary lung tumors are epithelial tumors. Epithelial malignant tumors are classified into non-small cell lung carcinomas (NSCLC) and small cell lung carcinomas (SCLC). Small cell carcinoma most often metastasizes more rapidly and widely than non-small cell lung carcinoma (and hence is staged differently). Small cell lung carcinoma has long been divided into two clinico-pathological stages, including limited stage and extensive stage. The most common types of NSCLC are squamous cell carcinoma, large cell carcinoma, and adenocarcinoma, but there are several other types that occur less frequently. All types can occur in unusual histologic variants and as mixed cell-type combinations.³⁰

2.2.2 Squamous Cell Carcinoma Preneoplastic Lesions

This study is limited to squamous cell carcinomas. A squamous cell carcinoma is often preceded for years by morphological changes in the bronchial epithelial region. The term preneoplasia is used to identify groups of phenotypically altered cells that may reflect sequential changes leading to neoplasia. The sequence of morphological changes that may lead to squamous cell carcinoma has long been thought to be basal cell hyperplasia, squamous metaplasia, mild, moderate and severe dysplasia, and carcinoma *in situ* (CIS). The World Health Organization (WHO) histological classification of tumors of the lung lists squamous dysplasia and CIS as main morphological forms of preinvasive lung lesions.³¹

Normal bronchial epithelium is represented by pseudo-stratified ciliated columnar epithelium. Basal Cell Hyperplasia (BCH) is present when the basal epithelial layer in the epithelium of bronchial lining is more than three cells thick. Squamous metaplasia usually starts in a zone of pre-existing BCH. In these zones, cytoplasmic changes consistent with squamous differentiation, and stratification of the epithelium occur.³² WHO classification of lung tumors defines mild, moderate, and severe dysplasia and CIS as follows: mild dysplasia

is represented by mildly increased thickness, mild pleomorphism, basilar zone expanded with cellular crowding in lower one-third of the epithelium, with nuclei vertically oriented, and absence of mitoses. Moderate dysplasia is presented by moderately increased thickness, moderate anisocytosis and pleomorphism; basilar zone expanded with cellular crowding in lower two-thirds of epithelium with nuclei vertically oriented, and mitotic figures in the lower third. Severe dysplasia is presented by markedly increased thickness, marked anisocytosis and pleomorphism, basilar zone expanded with cellular crowding into upper third, little progression of maturation, irregular nuclei vertically oriented in lower third, and mitosis in the lower two-thirds. CIS presents as markedly increased cell size and pleomorphism, no progression of maturation, cellular crowding throughout epithelium, no consistent orientation of nuclei, irregular nuclei with coarse, uneven chromatin, and mitotic figures through full thickness.³¹

These lesions are asymptomatic and become evident only through screening. One screening tool for detection and localization of preinvasive and early invasive lesions is bronchoscopy. These lesions are often not detected by conventional white light bronchoscopy. However, the use of fluorescence bronchoscopy, such as Lung-Imaging Fluorescence Endoscopy (LIFE)³³, greatly increases the sensitivity of detection. This technique uses the difference between the autofluorescence spectra of normal bronchial tissue and that of the pre-malignant/malignant tissue to discriminate between normal and pre-malignant/malignant tissue.

2.2.3 Natural Course of Preneoplastic Lesions in Bronchial Epithelium

The presence of preinvasive lesions of the bronchus increases the likelihood of developing lung cancer.³⁴ But there is a discrepancy between the prevalence of preinvasive lesions and the incidence of lung cancer, implying that not all lesions inevitably develop to lung cancer.

Even though lung cancers are believed to arise after a sequence of pathological stages (preneoplastic lesions), there is still little information about which lesions progress to neoplasia, how and why they progress, and whether or not there is any treatment that can reverse or prevent progression to neoplasia.

One critical need for a screening program is to be able to effectively grade premalignant lesions detected in bronchial epithelia and identify the ones that are likely to progress. That, along with developing appropriate therapies, may significantly improve survival rates.

Chemoprevention, the use of agents to treat the early phases of carcinogenesis and thereby prevent the development of invasive cancer, offers a promising means of controlling lung cancer by intervening during the earlier, potentially treatable phases of the neoplastic process.³⁵ Crucial to the success of cancer prevention clinical trials is the identification of intermediate endpoint biomarkers (surrogate endpoint biomarkers) that are sufficiently predictive for cancer incidence, to serve as study endpoints. These biomarkers should be involved in the process of carcinogenesis. Modulation of expression of these biomarkers should correlate with an altered risk of lung cancer. Their expression should differ in normal versus preneoplastic or at-risk epithelium, and they should be easily, quantitatively, and reproducibly measurable from biological specimens obtained during the trials.^{36,37}

There are no validated surrogate endpoint biomarkers that can reliably predict lung cancer incidence at present.³⁸ Bronchial preneoplastic lesions have been frequently used as a surrogate endpoint biomarker, but the utility of bronchial histopathology as the primary surrogate endpoint biomarker has recently been challenged. Although the risk of cancer progression increases with histological grade of preneoplastic lesion, these lesions may also regress. It is unknown whether a low-grade lesion like metaplasia or mild dysplasia progresses to lung cancer more slowly than high grade dysplasia. Thus, the histological grade of a preinvasive bronchial lesion might not be predictive of lung cancer. The potentially more malignant preneoplastic lesions among the many preneoplastic lesions present in the bronchial mucosa cannot be differentiated using the histological grade only.^{39,40}

Therefore, it is important to understand the carcinogenesis process and correlate the histopathology and the molecular profile of a lesion with its biological behavior to define what constitutes an important early cancer (preinvasive lesion that progresses to clinical cancer) to detect and treat.⁴¹

Carcinogenesis is characterized by genomic instability and accumulation of several key mutation events, giving advantages to malignant cells in cellular growth or survival. This process involves: (i) activation of growth-signaling pathways related to proto-oncogenes, (ii)

inactivation of tumor suppressor pathways, (iii) resistance to apoptosis, (iv) DNA repair alteration and immortalization, (v) angiogenesis and (vi) cellular migration.⁴² Cancer prevention trials generally assess multiple biomarkers as surrogate endpoint biomarkers including premalignant histologic lesions, pharmacodynamic endpoints, proliferation indices (i.e. fraction of proliferating cells within a tumor), and other attributes of the process of carcinogenesis.⁴³ One such biomarker that is frequently used for lung cancer risk assessment and chemopreventive response is the proliferative biomarker, Ki67.³⁵

2.2.4 Proliferative Biomarker Ki67 in Bronchial Epithelial Lesions

Ki67 is an immuno-marker that is used to identify proliferating cells in tissue sections. Ki67 is detectable in cells in nearly all stages of the cell-division cycle and is therefore a direct indicator of the growth fraction. Mounting of success evidence is driving calls for a wider use of proliferative indices (fraction of proliferating cells) based on Ki67 staining for clinical management strategies against a variety of cancer types.⁴⁴

Dysregulated proliferation is a well established hallmark of carcinogenesis.⁴⁵ The high Ki67 labeling index is associated with poor prognosis in lung cancer⁴⁶. Expression of Ki67 increases in bronchial pre-neoplasias of increasing grades of severity up to and including invasive lung cancer, even though it displays large inter-individual variation.²² These characteristics make Ki67 a potential endpoint biomarker, however there is currently no evidence that shows that modulation of this biomarker correlates with an altered risk for lung cancer.^{36,37}

2.2.5 Ki67 Scoring

Proliferative indices based on immunohistochemical detection and scoring of Ki67 have gained a lot of attraction in diagnostic pathology and have been studied for clinical management strategies against a variety of cancer types, including breast, cervical, and prostate cancer.^{47,48}

The Ki67 labeling index is typically calculated as the ratio of the number of Ki67-positive nuclei to the number of all nuclei in the entire epithelial area of a biopsy tissue section. Typical scoring systems for quantifying this biomarker provide only “global” data, i.e. one score for the entire epithelial area, disregarding the spatial inter-relations of the Ki67 expressing cells.

The spatial distribution of Ki67 staining has recently been used to evaluate cell proliferation and tumorigenesis in a few cancer types, with differing metrics applied towards this end.

A “topological index” of Ki67 staining in esophageal squamous dysplasias, based on a measure defined as the relative value of the vertical position of immuno-reactive cells within the epithelium, revealed that the distribution of Ki67 staining expanded in the lower layer of the epithelium with progression.⁴⁹

The combined analysis of the proliferating fraction and multiple quantitative features in Ki67 IHC-stained cervical intraepithelial neoplasias (CIN) found a correlation between these features and dysplastic grade.⁵⁰ Another study of CINs found that multivariate combination of features from the following list of Ki67-positive cell descriptors was associated with different CIN grades: the distance of the nucleus to the basal membrane, the thickness of the epithelium at the location of the nucleus, the distance of the nucleus to the lumen, the stratification index, the density of Ki67-positive nuclei per 100 μm basal membrane, and the percentage of the Ki67-positive nuclei in the deep third, middle third and upper third of the epithelium.⁵¹ In separate studies, this same group of researchers 1) used the presence or absence of MIB1-positive clusters (a cluster of at least two nearby MIB1-positive nuclei) in the upper two thirds of the epithelial layer to distinguish CIN and non-CIN⁵² and 2) showed that the Ki67 spatial distribution features have stronger predictive value for progression in the low grade CIN than histopathological grading and/or HPV presence.⁵³ For vulvar premalignant lesions, the basal cell layer expression of Ki67 has also been found to have diagnostic utility.⁵⁴

Regarding the bronchial epithelial lesions, the percentage of epithelial cells expressing Ki67 has been reported to be progressively elevated with increasing lesion severity (though large inter-individual variation has been observed).^{22,55-57} Meert et al.²² previously studied Ki67 expression to assess the proliferative activity of mild, moderate and severe dysplasias and CIS.²² They used parameters such as the presence of Ki67-positive nuclear clusters (a cluster was defined as a group of at least two contiguous cells showing strongly Ki67-positive nuclei located in the upper third of the abnormal epithelium) and a Ki67 score expressed as the sum of a score reflecting the proportion of Ki67-positive cells plus a score for staining intensity. This research group reported that when metrics for the distribution of

Ki67-positive cells and staining intensity were combined with the percentage of Ki67-positive cells, there are statistically significant differences between the category of mild and moderate dysplasia versus the one for severe dysplasia and CIS.

2.3 Materials

2.3.1 Study Population

This study used immunostained biopsy sections originally acquired as part of a chemoprevention clinical trial. The aim of the clinical trial was to determine the efficacy and safety of inhaled budesonide (Pulmicort Turbuhaler) as a chemopreventive agent in smokers with premalignant lesions in their bronchial epithelia.⁵⁸ Individuals at risk of developing lung cancer based on smoking history of higher than 30 pack-years were recruited for the chemoprevention study.

2.3.2 Bronchoscopic Examination

Autofluorescence bronchoscopy was performed in subjects with sputum atypia to localize areas of dysplasia using the LIFE-Lung device manufactured by Xillix Technologies Corp. (Richmond, British Columbia, Canada). Biopsy samples were then taken from areas with abnormal fluorescence that were at least 1.2 mm in size.

2.3.3 Histologic Examination

The biopsy samples were fixed in buffered formalin, embedded in paraffin, and serially sectioned. Hematoxylin and Eosin stained sections were systematically reviewed by two pathologists. All biopsy samples were classified into one of the following seven histology groups: normal, basal cell hyperplasia, squamous metaplasia, mild/moderate/severe dysplasia, or CIS). Normal tissue was represented by pseudo-stratified ciliated columnar epithelium. Basal cell hyperplasia was represented by an increase in the number and stratification of normal-appearing basal cells still covered with normal ciliated or mucin secreting cells. Metaplasia was represented by a stratified epithelium and cytoplasmic changes consistent with squamoid differentiation but lacking dysplastic changes. Mild, moderate, or severe dysplasia and CIS were classified according to criteria of the World

Health Organization.⁵⁹ Because individual biopsies frequently contained more than one histological condition, the diagnosis was based on the most advanced histology present.

2.3.4 Ki67 Expression

Five micron bronchial biopsy sections were mounted on silanized glass slides (HistoBond, Marienfeld, Germany). Tissue sections used for Ki67 immunohistochemistry were adjacent cuts to the ones from which histological diagnoses were made. Slides were baked overnight at 60°C and de-waxed in xylenes, followed by antigen retrieval in DakoCytomation Target Retrieval Solution. Antigen retrieval was performed in the Biocare Medical Decloaking Chamber pressure cooker at 120°C and 15 psi, 2.5 minutes at full pressure. Immuno-staining was performed on the DakoCytomation Autostainer using the EnVision Detection system (DakoCytomation, Carpinteria, CA). MIB1 antibody was provided by DakoCytomation and used at 1:100 dilution. Endogenous peroxidase was blocked with treatment in 3% aqueous hydrogen peroxide for 10 minutes. Tissue sections were incubated with the primary antibody for 30 minutes followed by a 30-minute incubation with EnVision+ (mouse). Color development was achieved by Dako Cytomation 3,3'-diaminobenzidine chromogen (10 minutes). Tissue sections were then counterstained with Hematoxylin. The slides were dehydrated, cleared with xylene, and cover-slipped.

2.3.5 Data

A total of 745 immunostained biopsy sections were available for this study. The set consisted of 27 normals, 158 hyperplasias, 194 metaplasias, 341 mild, 22 moderate, two severe dysplasias, and one CIS. Biopsies with unacceptable quality for the analysis were excluded; these were slides, in which the epithelium was torn, folded, or the basement membrane and the superficial membrane could not be demarcated, or the biopsy was tangentially cut. This left 613 biopsies (25 normals, 138 hyperplasias, 154 metaplasias, 273 mild, 21 moderate, and 2 severe dysplasias). Due to the small number of severe dysplasia specimens, moderate and severe dysplasias were combined into one category. These biopsies came from a total of 105 subjects.

Among the 613 immunostained biopsies, sequential histological changes in 323 lesions were available through biopsies performed during multiple follow-up bronchoscopies that were performed as part of the Lung Health Study led by Dr. Stephen Lam in the BC Cancer

Agency (Vancouver, BC, Canada). Different numbers of follow-up biopsies were available for different patients (ranging between 1 to 6, with median of 2). The length of follow-up period ranged from 5.5 months to 119.3 months (median of 15.7 months and inter-quartile range of 13.5 to 19.2 months). The median time between two consecutive follow-up visits was 7.9 months. The histology grades of follow-up biopsies will be used in the analysis of association of Ki67 quantitative features with progression in section 2.5.5.

2.4 Methods

2.4.1 Interpretation of Immunohistochemistry

Getafics system⁶⁰ (Integrative Oncology, BC Cancer Research Center), an in-house tissue section image capture and analysis system, was used to collect the images. Getafics system consists of a monochrome Charge Coupled Device (CCD) camera mounted on top of a light microscope controlled through an interactive image capture and analysis computer program.

In each case, the area within the tissue section with the highest number of Ki67-positive nuclei was carefully located by a trained technician. Then the image of the area was captured with a 40X objective, camera pixel size of $6.7\mu\text{m} \times 6.7\mu\text{m}$, and resolution of 1280×1024 , resulting in field of views of $0.21\text{mm} \times 0.17\text{mm}$. Using the image analysis utilities of Getafics, the epithelial region was interactively delineated on this image. This formed the region of interest (ROI) in the image. Both the basement membrane and the most superficial apical surface were separately marked. The centers of all visible nuclei inside the region were also individually marked by the technician. Ki67-positive nuclei were then recorded. The average number of cells in the marked epithelial regions across the 613 cases was 151, with a standard deviation of 97. The high standard deviation was due to the variability in the epithelial thicknesses of the samples, e.g. normal biopsies had pseudo-stratified epithelium with only one layer of cells. The coordinates of the nuclei centers, their labels (Ki67-positive or Ki67-negative) and the coordinates of the ROI were recorded in an output file. This information was subsequently used to calculate quantitative features describing the distribution of the Ki67-positive nuclei in the epithelium.

2.4.2 Spatial Characterization of the Epithelium

The spatial arrangement of cells within the epithelium was characterized in two ways: 1) a nucleus-centered Voronoi tessellation⁶¹ and 2) a distance map representing the distance of each pixel in the region of interest (ROI) from the basement membrane.⁶²

2.4.3 Voronoi Parameters

Given a set of points S in a plane, a Voronoi tessellation for S is the partition of the plane which associates a polygon $V(p)$ with each point p of S in such a way that all locations in $V(p)$ are closer to p than to any other point in S (see Figure 2.1).⁶³

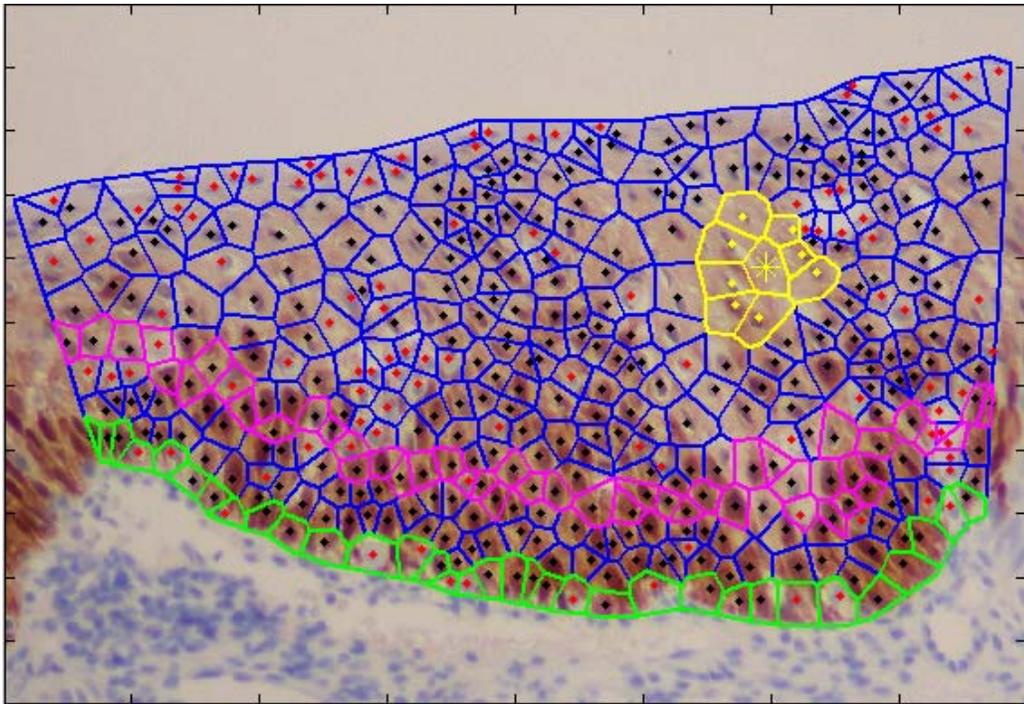


Figure 2.1 Voronoi tessellation (blue polygons) superimposed on the image of a tissue section. Green polygons represent the cells in the first layer above the basement membrane (basal layer). Pink polygons represent cells in the 4th layer. Yellow polygons are Voronoi neighbors of the nucleus marked with a yellow star.

Applied to centers of nuclei in the ROI, the Voronoi tessellation allows each nucleus to be associated with a Voronoi polygon that can be viewed as its area of influence. In this way, the Voronoi diagram gives a mathematical representation of the epithelium (see Figure

2.1). With this diagram, it is straightforward to analyze the distribution of the nuclei within a given region. Voronoi tessellation has been previously used to develop structural features to assess the possible diagnostic and prognostic value of the quantitative analysis of the architecture of a tissue.⁶³⁻⁶⁵

Here, we use structural features based on the Voronoi diagram and on Ki67 positivity to describe the spatial inter-relations of Ki67-positive cells in conjunction with the tissue architecture.

Given a finite set of points, the Voronoi polygons of the points on the outside edges of the set are unbounded, i.e. one vertex of those polygons will be located at infinity. The polygons corresponding to these points are called marginal polygons. For this reason, the intersection of a marginal polygon with the ROI (consisting of the basement membrane, the external membrane, and the lines joining the ends of these two membranes) was found. Each marginal polygon was replaced with the polygon resulting from the intersection of the marginal polygon with the ROI boundary.

2.4.3.1.1 Voronoi Neighbors

Based on the nucleus-centered Voronoi tessellation, the Voronoi neighbors were defined as those cells whose corresponding Voronoi polygons have a common side, i.e. touching Voronoi polygons. Voronoi neighbors were not defined for marginal polygons, so for those features that were calculated based on Voronoi neighbors, the nuclei with marginal polygons were not taken into account.

2.4.3.1.2 Voronoi Layers

Voronoi layers were defined as follows: all cells whose corresponding Voronoi polygons touched the basement membrane were assigned to layer 1, which was the basal layer. Layer 2 consisted of those cells which were not in layer 1 and had a neighbor in layer 1. Layer 3 consisted of those cells which were not in layer 2 or 1 and had a neighbor in layer 2. Higher number layers were defined in a similar fashion (as demonstrated in Figure 2.1).

2.4.3.2 Distance Parameters

We estimated the physical distance of each pixel P in the ROI to the basement membrane as follows: we formed a binary image BW of the basement membrane, in which pixels

corresponding to the basement membrane were presented by 1s, and all other pixels were presented by 0s. We then computed the Euclidean distance transform of the binary image BW. For each pixel in BW, the distance transform assigns a number that is the distance between that pixel and the nearest nonzero pixel of BW, which is the nearest pixel of the basement membrane. Matlab (The MathWorks Inc.) was used to calculate the distance transform. Figure 2.2 demonstrates an example of a typical distance map generated in this way.

Based on the distance maps of the ROI, several quantitative features were automatically calculated as listed in the following section.

2.4.4 Quantitative Ki67 Features

From the above mentioned spatial characterizations and by the assignment of the Ki67 positivity or negativity label to each cell nucleus, a number of quantitative features were derived that describe the spatial distribution of the positive and negative cells within the ROI. The metrics that we investigated were as follows:

1. *Ki67-Index*: the ratio of the number of the Ki67-positive nuclei to the number of all nuclei in the ROI within the epithelium.
2. *Layer-average*: the layer numbers of Ki67-positive nuclei, averaged over all Ki67-positive nuclei.
3. *Relative-layer-average*: *Layer-average* divided by the total number of layers in the ROI.
4. *Layer-90*: the 90th percentile of the layer numbers of Ki67-positive nuclei.
5. *Relative-layer-90*: *Layer-90* divided by the total number of epithelial cell layers in the ROI.
6. *Basal-layer-fraction*: the ratio of the number of Ki67-positive nuclei to the number of all nuclei within the basal layer (the first layer of Voronoi polygons above the basement membrane).
7. *Lower-half-layer-fraction*: the ratio of the number of Ki67-positive nuclei to the number of all nuclei in the lower half of the epithelial cell layers (excluding the basal layer).

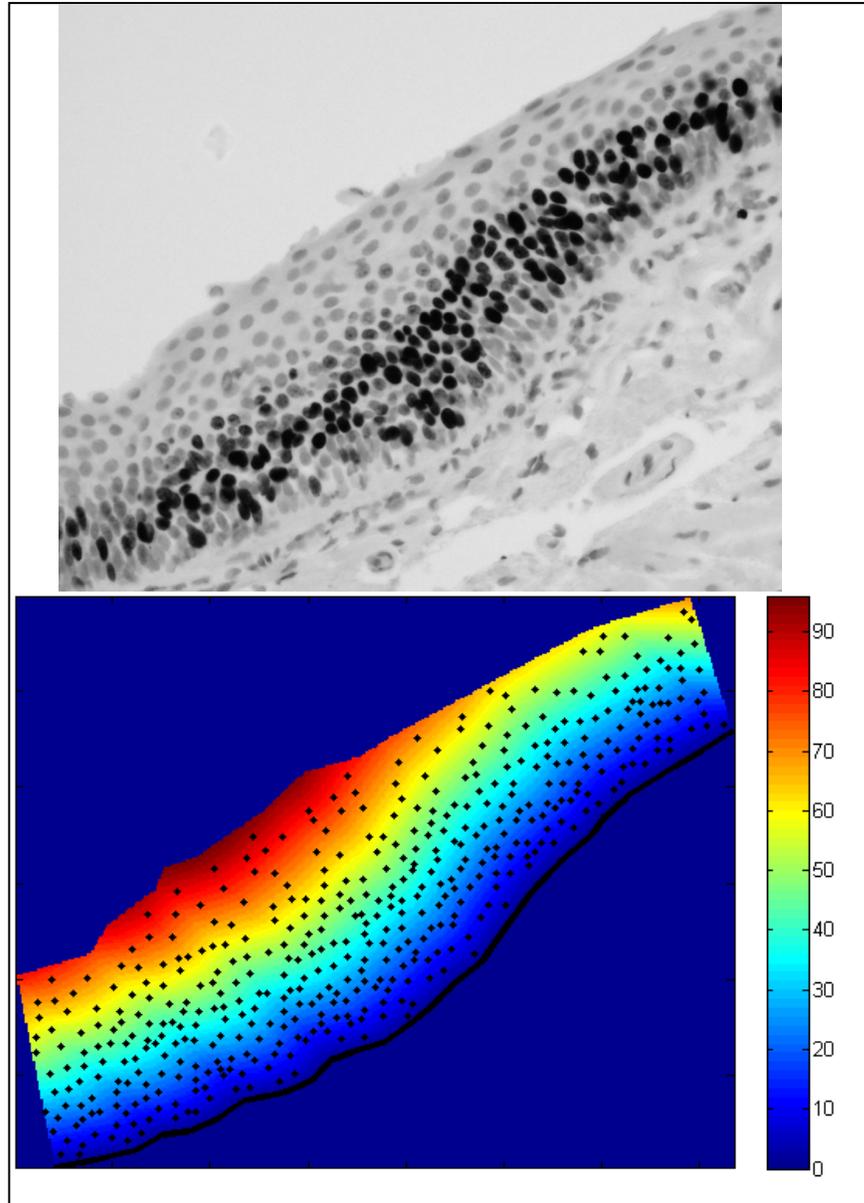


Figure 2.2 (a) Image of the ROI in a tissue section, (b) color-coded representation of distances (in microns) of the pixels of the ROI in (a) to the basement membrane. Nuclei centers are marked with black dots.

8. *Upper-half-layer-fraction*: the ratio of the number of Ki67-positive nuclei to the number of all nuclei in the upper half of the epithelial cell layers (excluding the basal layer).
9. *Distance-average*: the distances of Ki67-positive nuclei from the basement membrane, averaged over all Ki67-positive nuclei.

10. *Relative-distance-average*: the “relative distances” of Ki67-positive nuclei from the basement membrane, averaged over all Ki67-positive nuclei (“relative distance” of each nucleus from the basement membrane refers to the distance of that nucleus from the basement membrane divided by the thickness of the epithelium at the position of that nucleus).
11. *Distance-90*: the 90th percentile of the distances of Ki67-positive nuclei from the basement membrane.
12. *Relative-distance-90*: the 90th percentile of the relative distances of Ki67-positive nuclei from the basement membrane.
13. *Deep-third-fraction*: the ratio of the number of Ki67-positive nuclei to the number of all nuclei in the lower third of the epithelium.
14. *Middle-third-fraction*: the ratio of the number of Ki67-positive nuclei to the number of all nuclei in the middle third of the epithelium.
15. *Upper-third-fraction*: the ratio of the number of Ki67-positive nuclei to the number of all nuclei in the upper third of the epithelium.
16. *Positive-neighbor-number*: the average, over all Ki67-positive nuclei in the ROI, of the following: number of Ki67-positive neighbors that each Ki67-positive nucleus has. Expressed as number of neighbors.
17. *Positive-neighbor-fraction*: the average, over all Ki67-positive nuclei in the ROI, of the following: number of the Ki67-positive neighbors of each Ki67-positive nucleus, divided by the total number of neighbors of that nucleus. Expressed as a percentage.
18. *Positive-neighbor-distance-overall*: the average, over all nuclei in the ROI, of the distance of each nucleus from its nearest Ki67-positive neighbor.
19. *Positive-neighbor-distance-positive*: the average, over all Ki67-positive nuclei in the ROI, of the following: the distance of each Ki67-positive nucleus from its nearest Ki67-positive neighbor.
20. *Positive-per-basal-length*: the number of Ki67-positive nuclei in the ROI divided by the length of the basement membrane.
21. *Basal-positive-per-basal-length*: the number of Ki67-positive nuclei in the basal layer divided by the length of the basement membrane.

2.5 Results

2.5.1 Reproducibility of Ki67 Features

To investigate the reproducibility of the Ki67 features, 20 biopsies with Ki67 indices spanning 0 to 100% were selected. Two expert technicians analyzed these biopsies separately and this procedure was repeated three times, hence, each of these biopsies was analyzed six times, i.e. three times by each of the technicians. After the first round of analysis for each biopsy, a picture of the field of view was taken. This picture was later used by the two technicians to find the same field of view on the slide. The analysis consisted of marking the ROI, marking the center of all nuclei, and labeling the nuclei as negative and positive for Ki67. The Ki67 features were then calculated for each analyzed biopsy.

A two-way (subject \times observer) analysis of variance (ANOVA) was performed on each Ki67 feature to test for the effect of the inter-observer differences using Matlab (The Mathworks Inc.). The factors included in the model were the subject (biopsy) effect, the observer effect, and the interaction between the two. The subject main effect had 20 levels corresponding to 20 analyzed biopsies. The observer main effect had two levels corresponding to the two observers. There were three observations for each combination of factors, observer and subject, corresponding to the three replicate measurements by each observer. In an ordinary ANOVA model, each grouping variable represents a fixed factor, i.e. the levels of that factor are a fixed set of values. In a random effects model a factor's levels represent a random selection from a larger set of possible levels.⁶⁶ Since both “subject” and “observer” were random effects, the random effects model was used to estimate the variances caused by different factors: the inter-observer variance, the subject variance, and the intra-observer variance which was the remaining variance after the variances of observer, subject and the subject-observer interaction were removed.

The percentages of subject (biopsy) variance, inter-observer variance, and intra-observer variance for each feature are shown in Figure 2.3. Note that the three percentages for each feature do not add to 100% because there was also some variance due to the interaction between observer and subject (not shown).

The significance levels of the effects in the ANOVA analysis for every feature are shown in Table 2.1. Ideally, we would want the subject-effect to be statistically significant and the observer-effect to be non-significant. As can be seen in Table 2.1, the subject effect was statistically significant for all features. The observer effect, however, was only statistically significant for some Ki67 features (i.e. the observer had a significant effect on the value of several features). We note that for the *Ki67-index*, which is the “global” Ki67 feature that is typically used, the observer effect was significant, i.e. there was statistically significant difference between the values of Ki67 obtained by each observer. This indicated low inter-observer agreement for the *Ki67-index* feature. However, for some of our proposed features that take into account the distribution of Ki67-positive cells, the observer effect was non-significant. These features were: *Relative-layer-average*, *Upper-half-layer-fraction*, *Distance-average*, *Relative-distance-average*, *Distance-90*, *Relative-distance-90*, *Upper-third-fraction*, and *Positive-neighbor-distance-positive*. The non-significant observer effect for these features indicated low inter-observer variation.

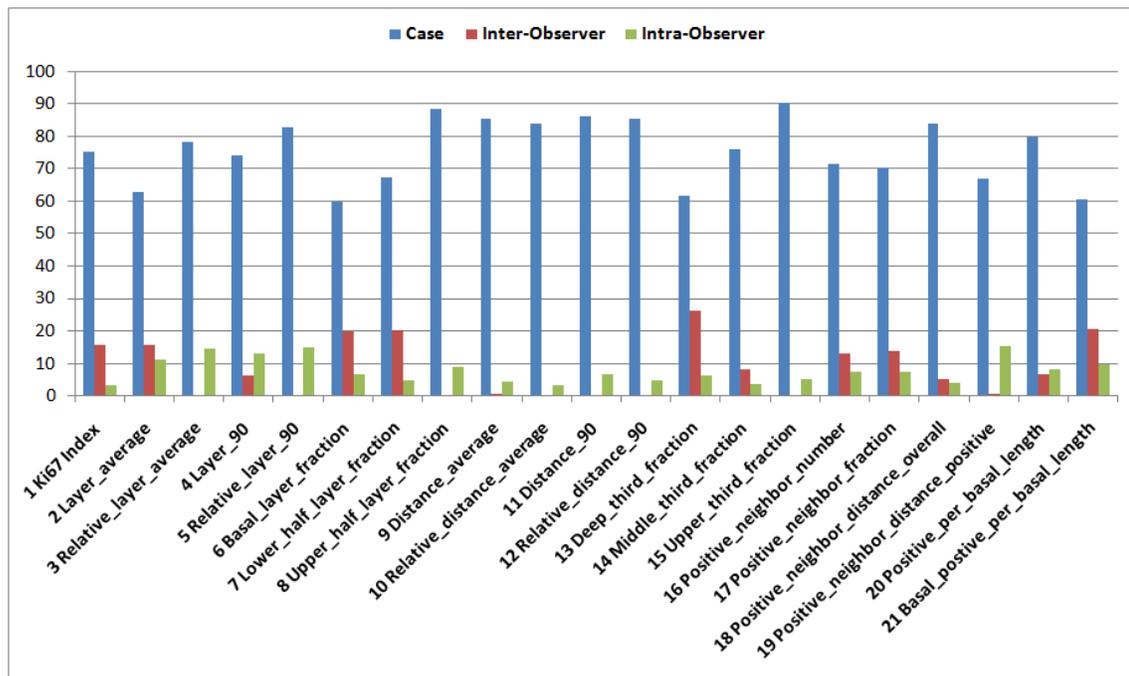


Figure 2.3 Estimated variances of different components (subject, inter-observer, and intra-observer) in the random effects ANOVA model, for different Ki67 features.

Table 2.1 Inter-observer effects on the reproducibility of the features. P-values from 2-way ANOVA.

No.	Feature Name	Observer	Subject
1	<i>Ki67 Index</i>	<0.001	<0.001
2	<i>Layer-average</i>	<0.001	<0.001
3	<i>Relative-layer-average</i>	0.464	<0.001
4	<i>Layer-90</i>	0.003	<0.001
5	<i>Relative-layer-90</i>	0.218	<0.001
6	<i>Basal-layer-fraction</i>	<0.001	<0.001
7	<i>Lower-half-layer-fraction</i>	<0.001	<0.001
8	<i>Upper-half-layer-fraction</i>	0.201	<0.001
9	<i>Distance-average</i>	0.177	<0.001
10	<i>Relative-distance-average</i>	0.298	<0.001
11	<i>Distance-90</i>	0.240	<0.001
12	<i>Relative-distance-90</i>	0.286	<0.001
13	<i>Deep-third-fraction</i>	<0.001	<0.001
14	<i>Middle-third-fraction</i>	0.002	<0.001
15	<i>Upper-third-fraction</i>	0.460	<0.001
16	<i>Positive-neighbor-number</i>	<0.001	<0.001
17	<i>Positive-neighbor-fraction</i>	<0.001	<0.001
18	<i>Positive-neighbor-distance-overall</i>	0.003	<0.001
19	<i>Positive-neighbor-distance-positive</i>	0.231	<0.001
20	<i>Positive-per-basal-length</i>	0.001	<0.001
21	<i>Basal-positive-per-basal-length</i>	<0.001	<0.001

2.5.2 Association of Ki67 Features with Histology

This section investigates the differences of each Ki67 feature across the five histology grades. These histology grades were numbered as 1, 2, 3, 4, and 5, which referred to normal, hyperplasia, metaplasia, mild dysplasia, and the combined group of moderate and severe dysplasia, respectively. To compare the value of each feature in five histology groups, a univariate Kruskal-Wallis test was first performed on each individual feature, followed by multiple comparisons of mean ranks using Statistica (StatSoft Inc., Tulsa, OK, US). The p-values of the Ki67 features for pair-wise comparisons between the histology groups are

shown in Table 2.2. To compensate for the multiple comparisons performed on each feature, p-values for each individual feature were Bonferroni-corrected⁶⁷ for the 10 comparisons between pair-wise combinations of the five groups. We did not, however, correct the p-values for the multiple features analyzed, because these features were highly correlated, mostly measured the same trend, and were not independently defined. (Most of the p-values are so small that even after Bonferroni correction they would still be significant [multiplication by a factor of 21, for the 21 features studied here]). Table 2.2 shows that the mean ranks of these features were nearly all significantly different between groups 1 and 3, 1 and 4, 1 and 5, 2 and 3, 2 and 4, and 2 and 5. None of the features were strongly different between groups 4 and 5 (mild dysplasia versus moderate and severe dysplasia). Only certain features were significantly different between groups 1 and 2 (normal versus hyperplasia) and between groups 3 and 4 (metaplasia versus mild dysplasia).

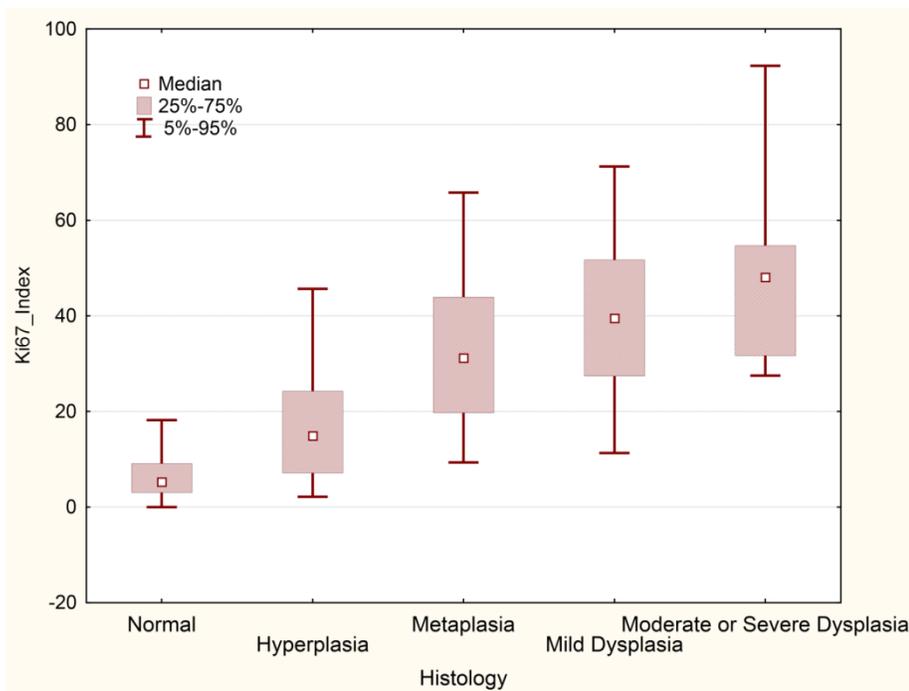


Figure 2.4 Box plot of *Ki67-index* grouped by Histology.

The box plots of the *Ki67* features grouped by histology are shown in Figures 2.4 to 2.15. The *Ki67-index* was seen to increase with increasing histology grades from normal to moderate and severe dysplasia (Figure 2.4). *Layer-average* and *Layer-90* results are shown

in Figure 2.5, while *Distance-average* and *Distance-90* results are shown in Figure 2.6. All these features followed the same trend: with increasing histology grades, the average or maximum distances of proliferating cells from the basement membrane increased (in terms of both measured distance (μm) and cell layers). Relative distances both in terms of distance (μm) and cell layers followed the same trend (Figures 2.7 and 2.8).

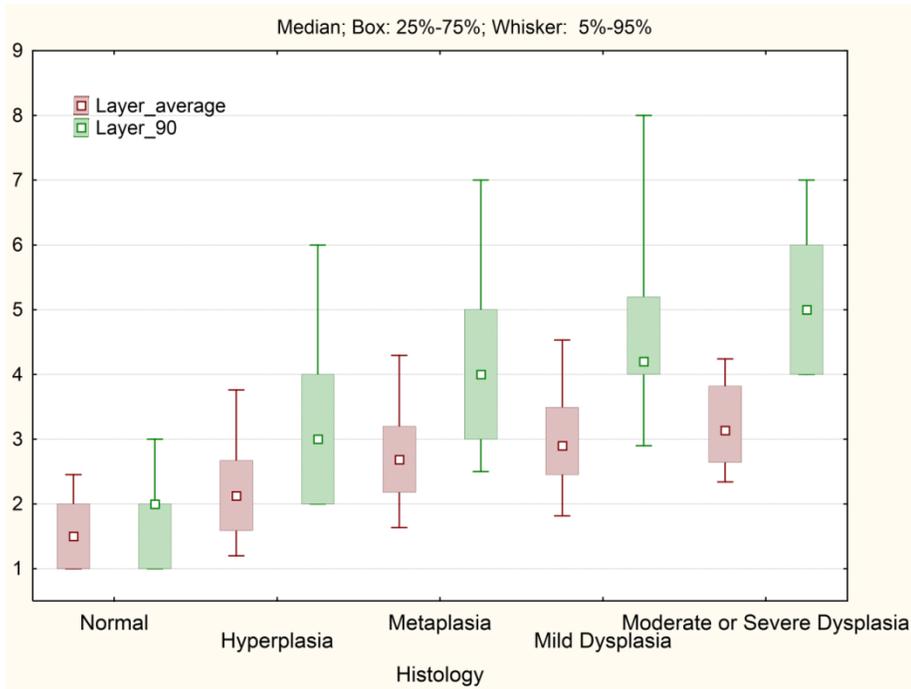


Figure 2.5 Box plots of *Layer-average*, and *Layer-90* grouped by Histology, Y-axis represents layer number.

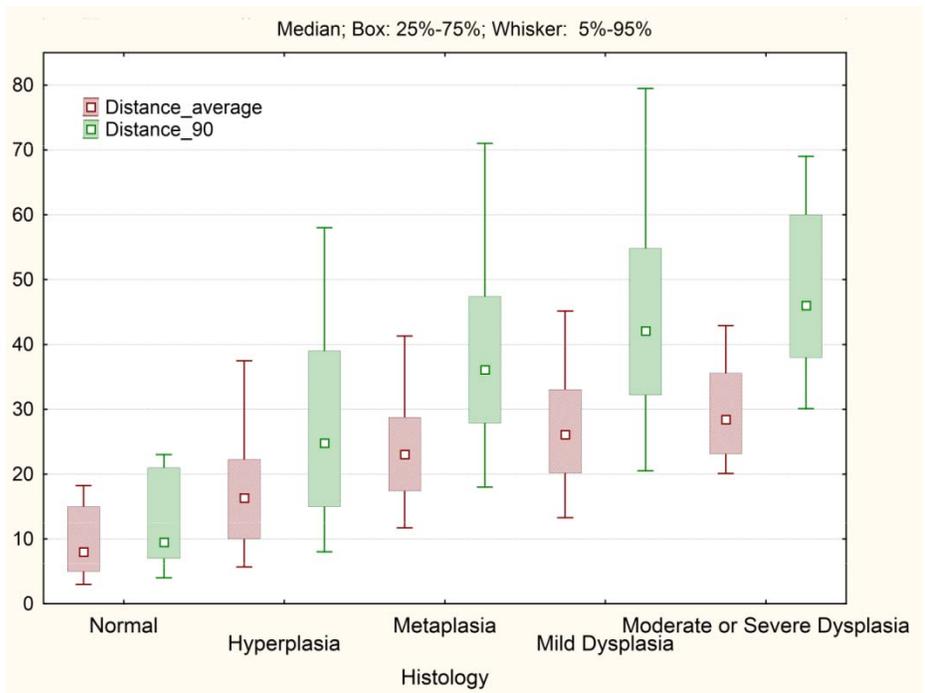


Figure 2.6 Box plots of *Distance-average* and *Distance-90* grouped by Histology, Y-axis represents distance in microns.

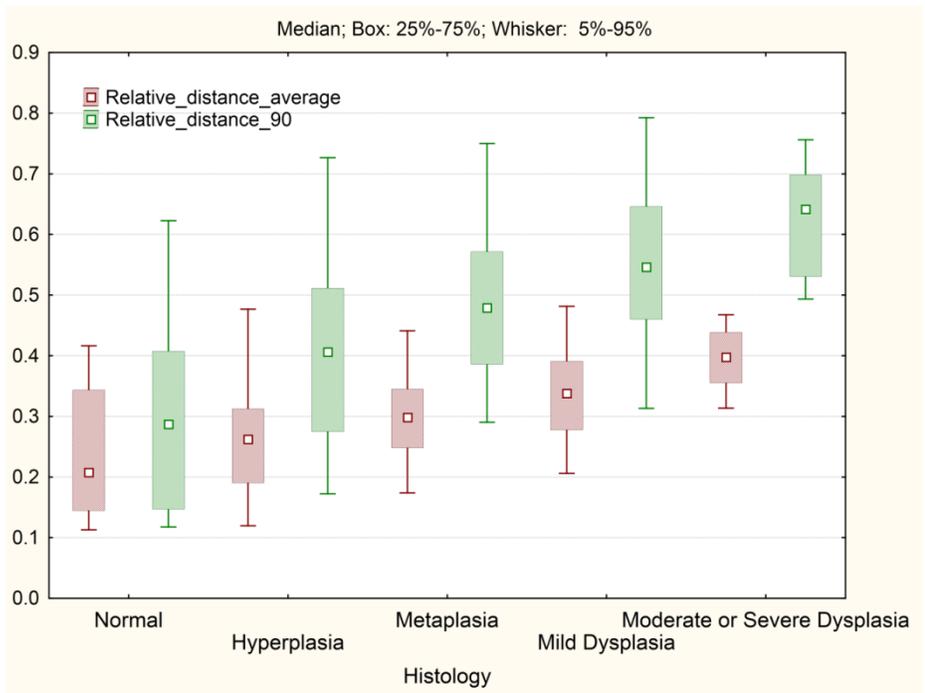


Figure 2.7 Box plots of *Relative-distance-average*, and *Relative-distance-90*, grouped by histology.

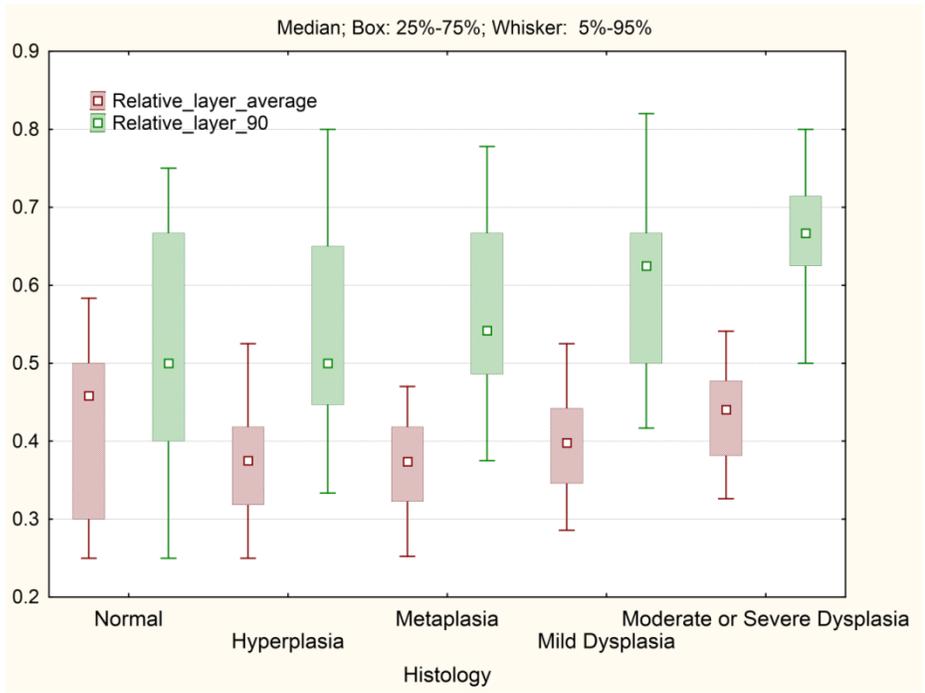


Figure 2.8 Box plots of *Relative-layer-average*, and *Relative-layer-90*, grouped by histology.

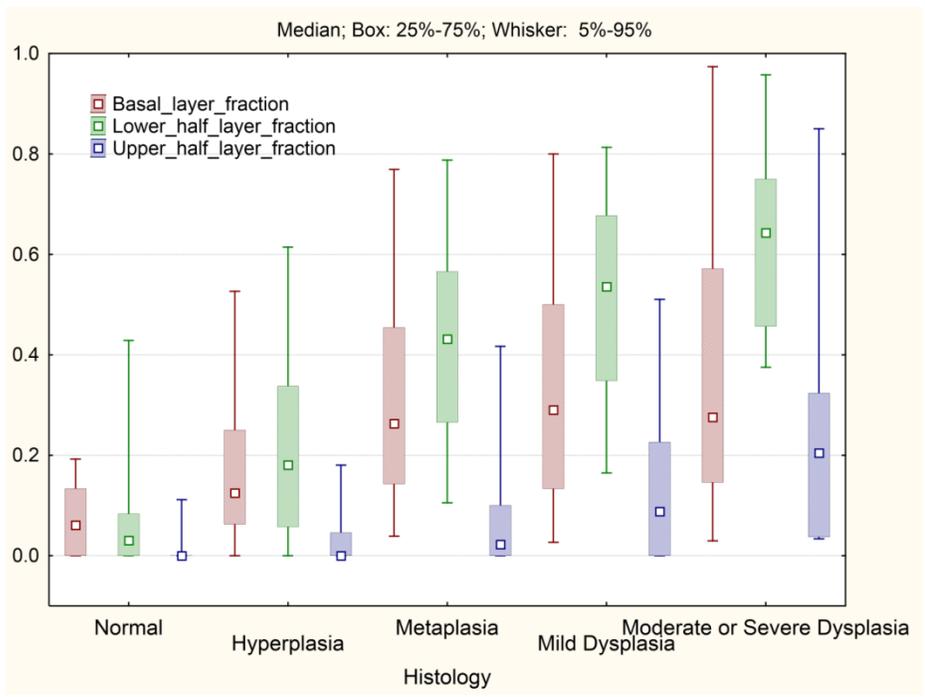


Figure 2.9 Box plots of *Basal-layer-fraction*, *Lower-half-layer-fraction*, and *Upper-half-layer-fraction* grouped by Histology.

Figure 2.9 shows the box plots for the *Basal-layer-fraction*, *Lower-half-layer-fraction*, and the *Upper-half-layer-fraction*. The median of the *Basal-layer-fraction* increased from normal to hyperplasia and to metaplasia, but remained the same for metaplasia and dysplasias.

Figure 2.10 shows the box plots for fractions of Ki67-positive cells in the deep, middle, and upper third of the epithelium. The fraction of positive cells in the lower third of the epithelium was highest, followed by the middle third, and the upper third (regardless of the histology grade). With increasing histology grade, both the number and the fraction of positive neighbors of positive cells increased.

Figure 2.11 shows the trend for the number of positive neighbors each cell had. Also, the distance to the closest positive neighbor was seen to decrease with increasing histology grade (Figure 2.12). Figure 2.13 shows that *Positive-neighbor-fraction* increased with increasing histology grade.

The number of Ki67 positive nuclei divided by the length of the basement membrane, both in the entire ROI and in the basal layer, increased with increasing histology grade (Figures 2.14 and 2.15).

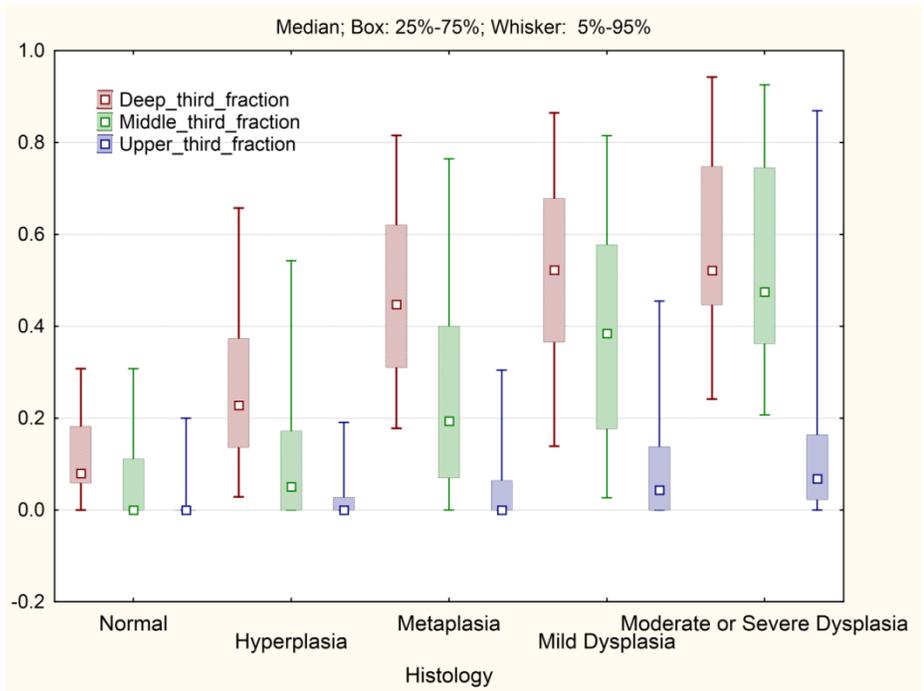


Figure 2.10 Box plots of *Deep-third-fraction*, *Middle-third-fraction*, and *Upper-third-fraction* grouped by Histology.

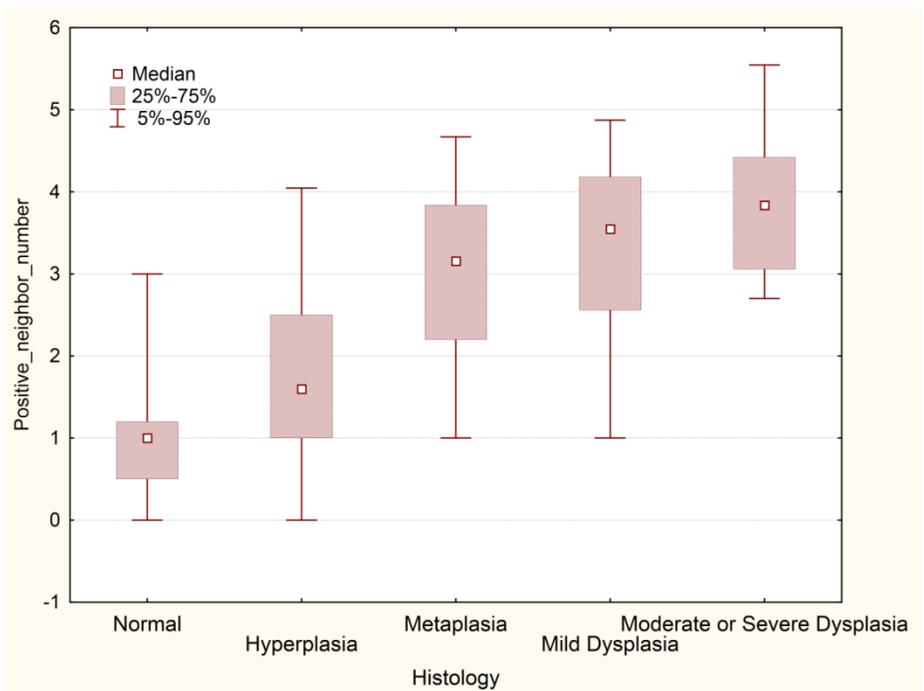


Figure 2.11 Box plot of *Positive-neighbor-number* grouped by Histology.

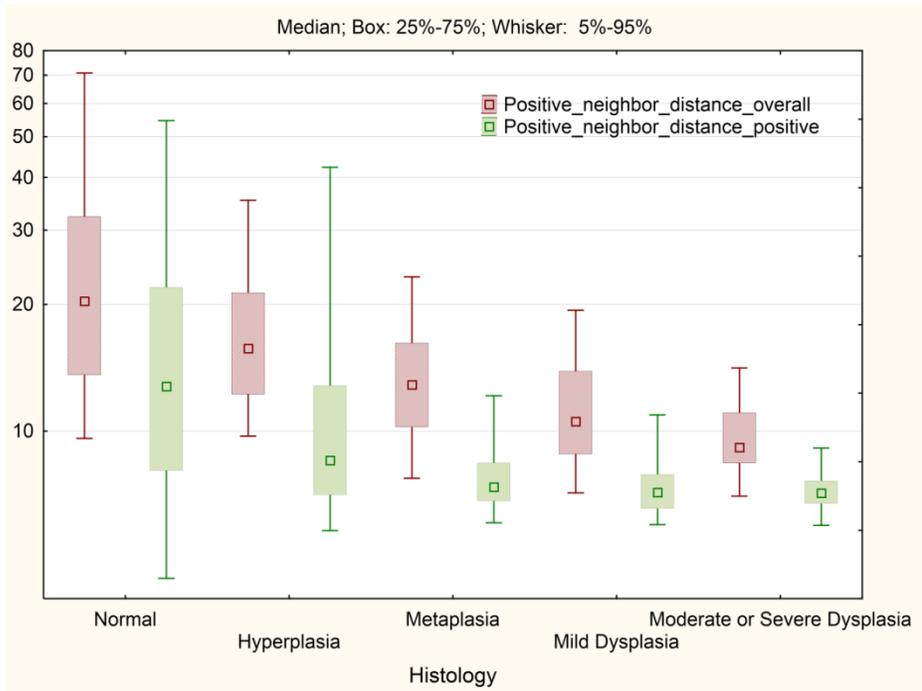


Figure 2.12 Box plots of *Positive-neighbor-distance-overall*, and *Positive-neighbor-distance-positive* grouped by Histology, Y-axis represents distances in microns on logarithmic scale.

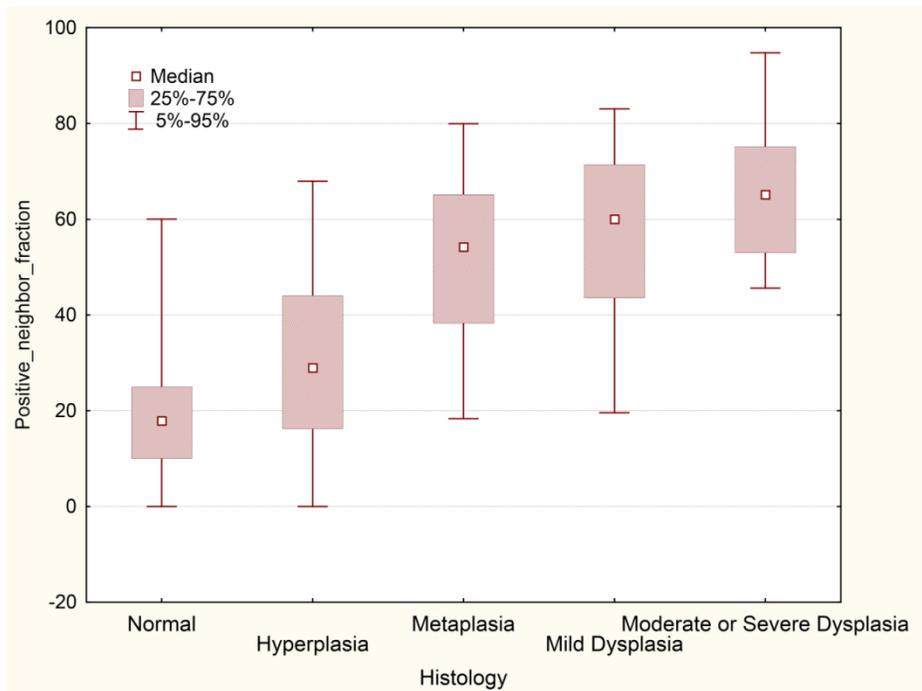


Figure 2.13 Box plot of *Positive-neighbor-fraction*, grouped by histology.

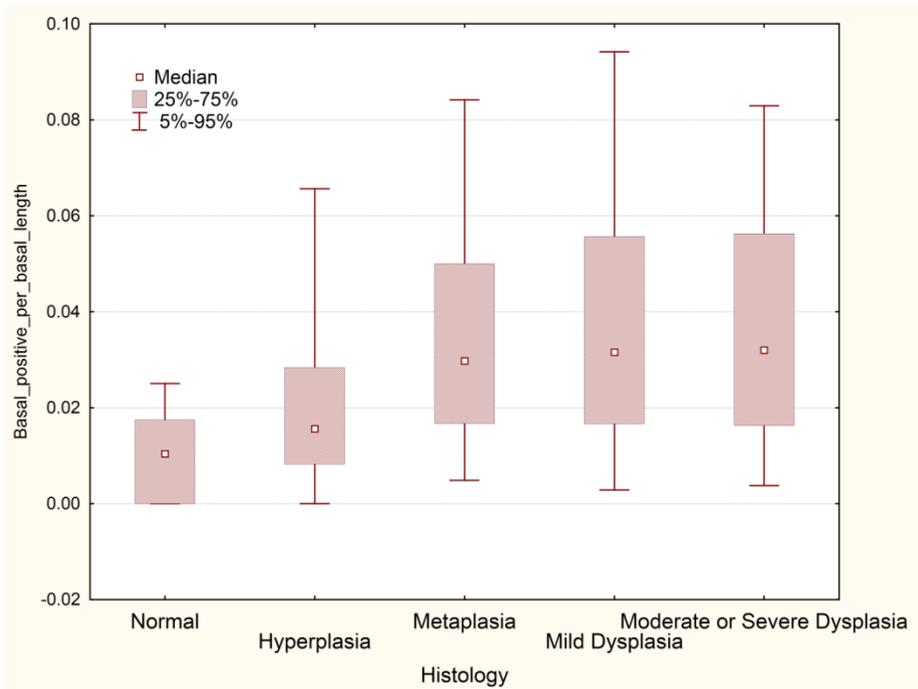


Figure 2.14 Box plot of *Positive-per-basal-length*, grouped by histology.

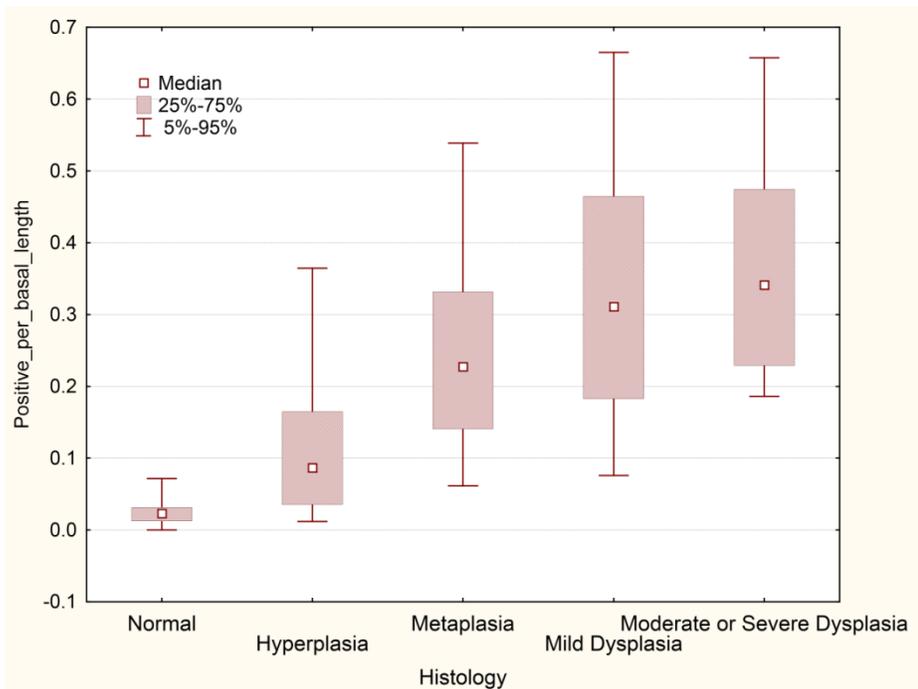


Figure 2.15 Box plot of *Basal-positive-per-basal-length*, grouped by histology.

Table 2.2 P-values from the mean rank comparisons using a Kruskal-Wallis ANOVA test. Groups 1, 2, 3, 4 and 5 respectively correspond to normal, hyperplasia, metaplasia, mild dysplasia, and a combined group of moderate and severe dysplasia.

Feature Name	1 vs. 2	1 vs. 3	1 vs. 4	1 vs. 5	2 vs. 3	2 vs. 4	2 vs. 5	3 vs. 4	3 vs. 5	4 vs. 5
<i>Ki67 Index</i>	0.082	0.000	0.000	0.000	0.000	0.000	0.000	0.007	0.017	0.972
<i>Layer-average</i>	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.032	0.110	1.000
<i>Relative-layer-average</i>	0.051	0.038	1.000	1.000	1.000	0.009	0.005	0.003	0.003	0.451
<i>Layer-90</i>	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.016	0.052	1.000
<i>Relative-layer-90</i>	1.000	1.000	0.097	0.004	1.000	0.000	0.000	0.001	0.001	0.254
<i>Basal-layer-fraction</i>	0.318	0.000	0.000	0.000	0.000	0.000	0.004	1.000	1.000	1.000
<i>Lower-half-layer-fraction</i>	0.396	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.008	0.677
<i>Upper-half-layer-fraction</i>	1.000	0.067	0.000	0.000	0.025	0.000	0.000	0.000	0.000	0.093
<i>Distance-average</i>	0.013	0.000	0.000	0.000	0.000	0.000	0.000	0.039	0.088	1.000
<i>Relative-distance-average</i>	1.000	0.173	0.000	0.000	0.006	0.000	0.000	0.000	0.000	0.018
<i>Distance-90</i>	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.025	0.109	1.000
<i>Relative-distance-90</i>	0.453	0.002	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.123
<i>Deep-third-fraction</i>	0.026	0.000	0.000	0.000	0.000	0.000	0.000	0.949	0.998	1.000
<i>Middle-third-fraction</i>	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.350
<i>Upper-third-fraction</i>	1.000	0.364	0.001	0.000	0.359	0.000	0.000	0.003	0.003	0.365
<i>Positive-neighbor-number</i>	0.862	0.000	0.000	0.000	0.000	0.000	0.000	0.066	0.050	1.000
<i>Positive-neighbor-fraction</i>	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.066	1.000
<i>Positive-neighbor-distance-overall</i>	1.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.364
<i>Positive-neighbor-distance-positive</i>	1.000	0.020	0.000	0.006	0.002	0.000	0.008	0.092	1.000	1.000
<i>Positive-per-basal-length</i>	0.017	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.104	1.000
<i>Basal-positive-per-basal-length</i>	0.339	0.000	0.000	0.001	0.000	0.000	0.025	1.000	1.000	1.000

While the significant p -values of the Kruskal-Wallis test for each feature (Table 2.2) indicated statistically significant differences in the mean ranks between the groups, they did not provide a measure of the discriminating power of each feature. The box plots provided a quick and simple visual means of demonstrating the discrimination power of evaluated features. As can be seen in the box-plots, there was considerable overlap between adjacent histology groups. The overall classification rates were generally low for either individual features or combinations of features selected by stepwise discriminant analysis (data not shown). It should be noted that the features were grouped by histology grade, even though it is well known that the reproducibility of histologic grading system is not high. Therefore the considerable overlap of the Ki67 features across histology groups may have been due to the uncertainties in the histologic grade assignments as well as measurement uncertainty and biological relevance.

2.5.3 Ki67 Expression Pattern as a Function of Depth

To assess the spatial distribution of the Ki67-positive cells within the bronchial epithelium, the epithelium was divided into bands of equal thickness (5 μm) starting from the basement membrane. For each tissue section, within each band, the Ki67 positivity index was calculated as the number of positive nuclei within that band divided by the number of all the nuclei (positive or negative) within the same band. These measures were then plotted for different histology grades. The box plots of the Ki67 positivity indices for the different regions of the epithelium are shown in Figure 2.16.

Similar plots were also generated using Voronoi cell layers as the distance measure, i.e. the different Ki67 positivity indices were calculated within the epithelial Voronoi layers as opposed to bands of equal thickness. These plots are shown in Figure 2.17.

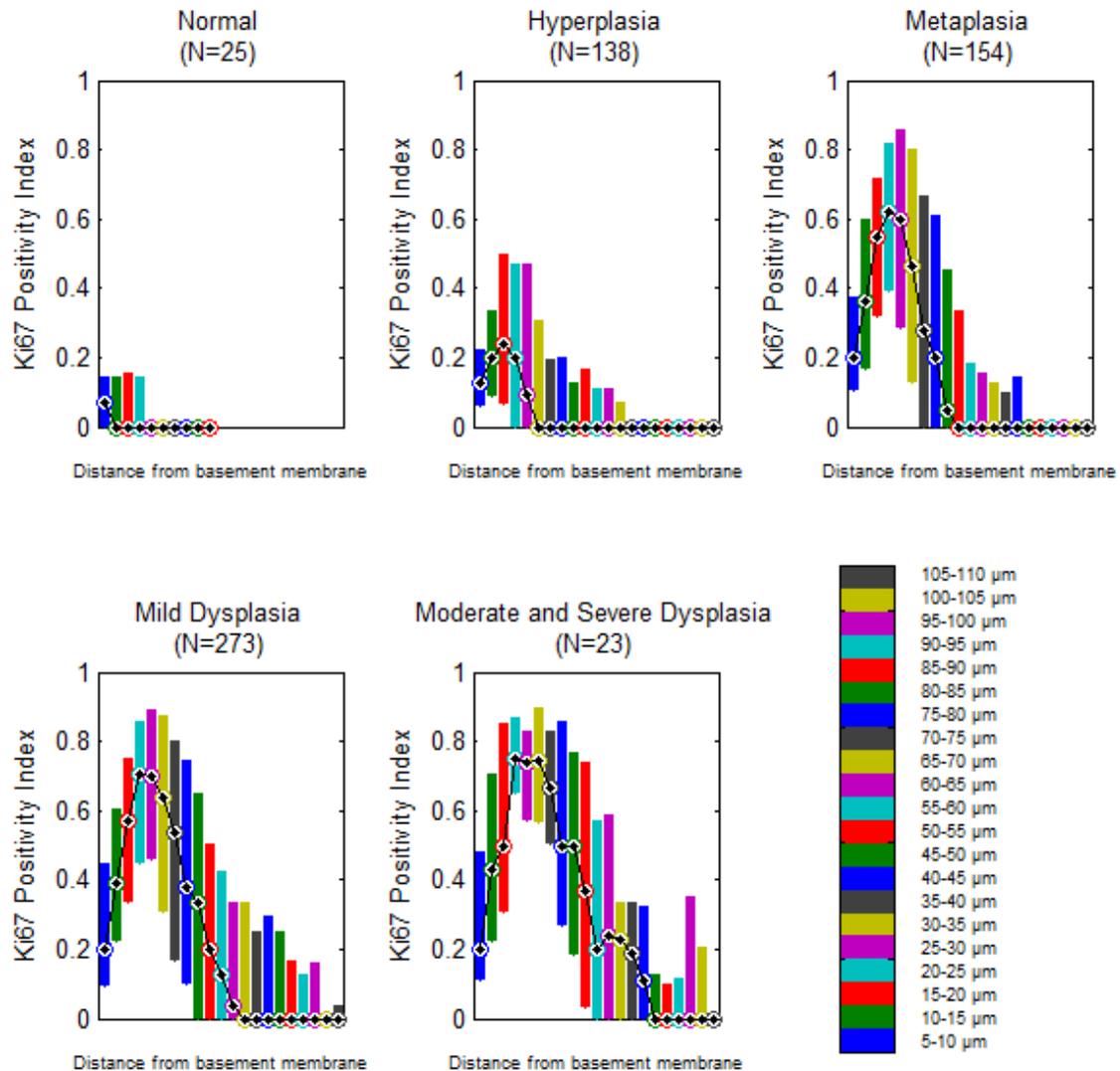


Figure 2.16 Distribution of Ki67 expressing cells in the epithelium, for different histology grades. Box plots of the Ki67 positivity indices in 5-micron intervals above the basement membrane. Each box represents the interquartile range of the Ki67 positivity indices. The circles correspond to the median of Ki67 positivity indices. Medians are connected by a black line to demonstrate the pattern.

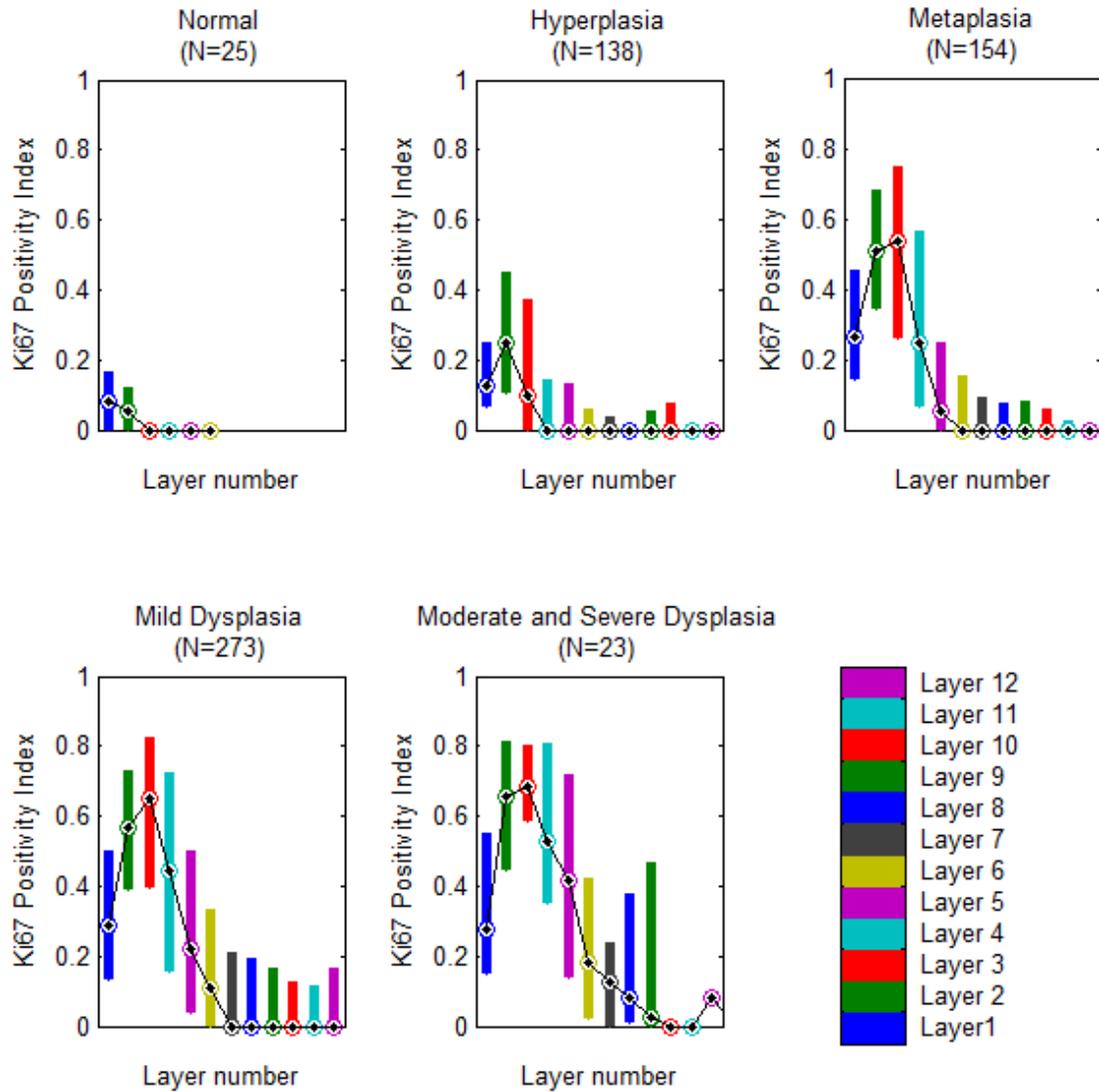


Figure 2.17 Distribution of Ki67 expressing cells in the epithelium, for different histology grades. Box plots of the Ki67 positivity indices in Voronoi cell layers above the basement membrane. Each box represents the inter-quartile range of the Ki67 positivity indices. The circles correspond to the median of Ki67 positivity indices. Medians are connected by a black line to demonstrate the pattern.

2.5.4 Ki67-Positive Cells in the Basal Layer

According to the results shown in Figure 2.17, the percentage of Ki67-positive cells in the basal layer was lower than that in the next two layers, in groups of metaplasia, mild, moderate, and severe dysplasia. This suggested that the distribution of the proliferating cells followed a certain pattern. We attempted to find out the likelihood of this pattern being generated by random chance, i.e. the probability of observing this pattern if the proliferating cells were randomly distributed in the epithelium. To find this probability, we performed the following randomization test: for each biopsy section, we generated 1000 simulated biopsy sections, with the same ROI, the same number of cells within the ROI, and the same location of the nuclei centers. In each simulated biopsy section, we then randomly re-assigned the Ki67-positive versus Ki67-negative labels to the cells, while keeping the number of the positive cells the same as the original biopsy section.

For the j -th simulated biopsy, a statistic S_{ij} was defined as:

$$S_{ij} = \text{percentage of positive cells in the } i\text{-th layer} - \text{percentage of positive cells in the basal layer} \quad (2.1)$$

where i was a layer number. Therefore, S_{2j} represented the difference between the percentage of positive cells in layer two and in the basal layer, in the j -th simulated biopsy section. Similarly, S_{3j} represented the difference between the percentage of positive cells in layer three and in the basal layer. $S_{i-original}$ ($i = 2,3$) was defined as the difference between the percentage of positive cells in layers 2 and 3, respectively, and in the basal layer in the original biopsy section.

For each original biopsy section, we then defined two p -values p_i ($i = 2,3$) which represented the fraction of the simulated biopsies in which S_{ij} was larger than $S_{i-original}$. p_i was therefore calculated as follows:

$$X = \{j \mid S_{ij} > S_{i-original}\} \\ p_i = \frac{n(X)}{1000}, \quad \text{for } i = 2,3 \quad (2.2)$$

where $n(X)$ represents the number of elements in the set X .

Since a significant number of positive nuclei would be needed for the pattern to be apparent, we limited our analysis to only those biopsy sections with more than 30% positive nuclei. In 70% of these cases (i.e. cases with more than 30% positive nuclei), either p_2 or p_3 were smaller than 0.05. Thus it appeared that in proliferating lung epithelium, for a majority of cases, with a significant number of positive nuclei, the second or third layer of cells proliferated significantly more than cells in the basal layer. This pattern was significantly different from that possible by random chance.

This strongly suggests that a biological rule must be in force (this pattern being observed by random chance is very unlikely). One possible biological interpretation of this observed pattern is that it demonstrates that asymmetric cell divisions must be occurring in the basal layer. The daughter cells that remains in the basal layer divide less frequently than the daughter cells in layer 2 or daughter cells in layer 3 (or possibly daughter cells of layer 2 daughter cells).⁶⁸

2.5.5 Association of Ki67 Features with Progression

2.5.5.1 Definition of Progression/Regression

The outcome for every biopsied site was defined based on the highest histology grade of the follow-up biopsies obtained from the same site. Table 2.3 shows the outcome for lesions with different baseline histology grades.

As Table 2.3 shows, for this dataset, the overall number of lesions that progressed to moderate or severe dysplasia was very limited (16 in total). Typically, in studies on the history of preneoplastic lung lesions, progression is defined as the change from a low-grade lesion (hyperplasia, metaplasia, mild dysplasia, or moderate dysplasia) to a high-grade lesion (severe dysplasia and CIS).^{40,69} We altered these criteria according to the availability of lesions in our cohort by applying the following logic: only lesions with initial (baseline) histology of metaplasia and mild dysplasia were included in the following analyses. The reason was the statistical requirement to have a cohort with enough cases for both progression and regression. Thus, lesions with initial histology of hyperplasia were excluded because they could only progress into higher histology grades. Similarly, lesions with initial histology of moderate and severe dysplasia were excluded because they only regressed. Including the mentioned biopsies in the data would falsely make the initial histology appear

as the dominant effect, i.e. it would make the data appear as if the majority of lesions with initial histology of moderate dysplasia regress to lower grades, while this only happened in our dataset because the numbers of moderate and severe dysplasias were very limited.

Table 2.3 Outcome of lesions with different baseline histology grades. Entries highlighted in pink represent numbers of the progressing lesions that were included in the analyses. Entries highlighted in blue represent the numbers of regressing lesions that were included in the analyses.

Baseline Histology	Total No.	Follow-up histology (outcome)					
		Normal	Hyperplasia	Metaplasia	Mild Dysplasia	Moderate Dysplasia	Severe Dysplasia
Hyperplasia	39	0	12	6	20	1	0
Metaplasia	92	1	43	17	29	1	1
Mild Dysplasia	178	6	86	43	31	10	2
Moderate Dysplasia	13	0	6	3	4	0	1
Severe Dysplasia	1	0	0	1	0	0	0

A case with initial histology of mild dysplasia would be considered a progression if it persisted as a mild dysplasia or changed to a higher grade of dysplasia at any point in the follow-up period. It would be considered a regression if it changed to normal or hyperplasia. Lesions that changed from mild dysplasia to metaplasia were not included in the analyses because the difference between the histology of metaplasia and mild dysplasia is not significant and the inter-observer agreement for grading metaplasia versus mild dysplasia is very low. Moreover, taking into account the “biopsy effect”, i.e. partial or complete removal and/or disruption of a lesion through multiple biopsies⁷⁰, the change from mild dysplasia to metaplasia could be due to the biopsy effect and therefore is not a reliable regression.

For lesions with initial histology of metaplasia, a change to dysplasia was considered progression, while a change to hyperplasia or normal was considered regression. Lesions that started and remained at metaplasia were excluded from analyses. A change from metaplasia to mild dysplasia was considered progression, contradictory to the exclusion of lesions that changed from mild dysplasia to metaplasia. The justification for this was that an increase in the grade from metaplasia to mild dysplasia happened in spite of the “biopsy effect”, which made it significant.

There were several lesions where the time sequence of histology grades contained both increase and decrease of the histology grade. This made the assignment of progression versus regression uncertain, e.g. mild dysplasia → hyperplasia → metaplasia. Such cases were removed from analyses.

2.5.5.2 Outcome Analysis

The association of each individual Ki67 feature with the outcome of the lesions was studied by performing univariate t-tests between groups of lesions that “progressed” and lesions that “regressed” based on our definition of progression versus regression explained in above section. Those features which were not normally distributed were transformed in a way that the new variable had a distribution as close to normality as possible, using the Box-Cox transformation.⁷¹ The analysis was done by STATISTICA (data analysis software system), V.10 (StatSoft, Inc.). The means and standard deviations of the Ki67 features, along with the p-values for the t-tests are listed in Table 2.4. Several features were found to be significantly different between the two groups (progressing vs. regressing lesions). The *p*-values were not Bonferroni adjusted. The reason for this was that these features were highly correlated and were not randomly selected. Yet, some of the features would still have low *p*-values after Bonferroni correction, which is equivalent to multiplying *p*-values by the total number of the features tested, in this case 21. According to the test results, several of our proposed Ki67 features had lower *p*-values than *Ki67-index*.

Even though the mean values of several features were significantly different between the two groups (progression vs. regression), there was considerable overlap between the two groups due to the high variability of each feature within each group, suggesting that the discriminating power of these features was not high. As an example, Figure 2.18 shows the box plot for one Ki67 feature, *distance-90*.

Table 2.4 Means and standard deviations of each Ki67 feature in the two groups of progressing and regressing lesions. 4th and 5th columns represent p-values of t-test with equal and unequal estimates of the variance of feature in each group, respectively. Last column represents the *p*-value of a test comparing the variance of the feature in each group. Where the *p*-value is less than 0.05 (highlighted in gray), the hypothesis of equal variances can be rejected; therefore, the *p*-value reported in the 5th column (separate estimates of variance) should be used. *p*-values less than 0.05 are highlighted in blue.

Feature	Follow-up status		<i>p</i> -value	<i>p</i> -value separate variance estimate	<i>p</i> -value Variances
	Regression mean (SD)	Progression mean (SD)			
<i>Ki67-Index</i>	0.80 (0.26)	0.86 (0.23)	0.064	0.053	0.147
<i>Layer-average</i>	1.50 (0.51)	1.70 (0.53)	0.011	0.012	0.780
<i>Relative-layer-average</i>	2.60 (1.03)	2.52 (0.82)	0.959	0.956	0.031
<i>Positive-neighbor-number</i>	1.64 (0.58)	1.88 (0.44)	0.004	0.002	0.007
<i>Positive-neighbor-fraction</i>	0.94 (0.24)	1.04 (0.15)	0.003	0.001	<0.001
<i>Positive-neighbor-distance-overall</i>	0.56 (0.27)	0.58 (0.23)	0.817	0.810	0.210
<i>Positive-neighbor-distance-positive</i>	0.47 (0.21)	0.43 (0.11)	0.120	0.064	<0.001
<i>Distance-average</i>	0.87 (0.21)	0.96 (0.20)	0.006	0.005	0.674
<i>Distance-90</i>	0.84 (0.21)	0.91 (0.21)	0.042	0.042	0.989
<i>Relative-distance-average</i>	3.90 (1.66)	3.90 (1.40)	0.958	0.956	0.110
<i>Relative-distance-90</i>	5.74 (2.71)	5.34 (2.25)	0.379	0.353	0.080
<i>Deep-third-fraction</i>	2.96 (1.56)	3.37 (1.54)	0.069	0.068	0.938
<i>Middle-third-fraction</i>	0.94 (0.71)	1.05 (0.66)	0.224	0.215	0.512
<i>Upper-third-fraction</i>	0.22 (0.26)	0.20 (0.24)	0.765	0.759	0.372
<i>Positive-per-basal-length</i>	1.04 (0.55)	1.19 (0.53)	0.061	0.059	0.748
<i>Basal-positive-per-basal-length</i>	0.87 (0.50)	0.81 (0.46)	0.421	0.409	0.404
<i>Basal-layer-fraction</i>	0.96 (0.59)	0.94 (0.58)	0.884	0.883	0.881
<i>Lower-half-layer-fraction</i>	2.79 (1.65)	3.30 (1.57)	0.026	0.024	0.645
<i>Upper-half-layer-fraction</i>	0.29 (0.31)	0.30 (0.29)	0.760	0.756	0.593
<i>Layer-90</i>	1.14 (0.30)	1.23 (0.31)	0.065	0.068	0.705
<i>Relative-layer-90</i>	5.11 (2.58)	4.76 (2.03)	0.492	0.462	0.026

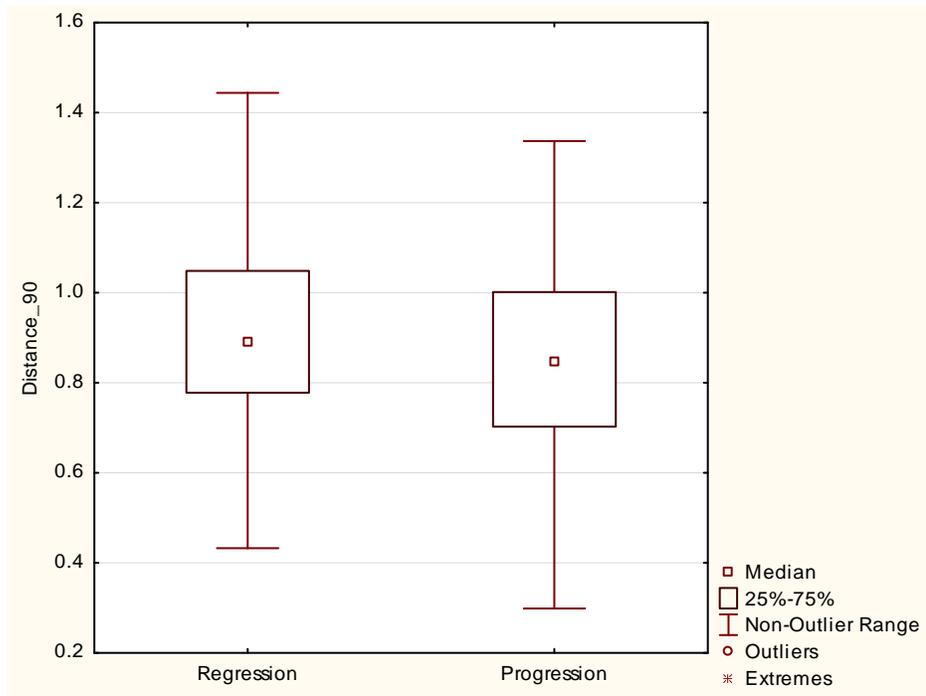


Figure 2.18 Box plot of *Distance-90* for the two groups of progressing and regressing lesions.

Among those features that were significantly different between the two groups at significance level of 0.05, a stepwise linear discriminant analysis selected *positive-neighbor-fraction*, *distance-average* and *positive-per-basal-length* as the best linear discriminant set of three features to distinguish between progressing lesions and regressing lesions (Table 2.5). We defined a new variable equal to the linear discriminant function obtained using the three selected features. This new variable will be referred to as the linear discriminant function in the following analyses. Figure 2.19 shows the box plot of the linear discriminant function consisting of these three features. There was still considerable overlap between the linear discriminant function values in the two groups (progressing vs. regressing).

Table 2.5 Features selected by the stepwise linear discriminant analysis that best discriminate between the progressing lesions and regressing lesions.

	Step	<i>p</i> -value
<i>Positive-neighbor-fraction</i>	1	0.003
<i>Distance-average</i>	2	0.004
<i>Positive-per-basal-length</i>	3	0.002

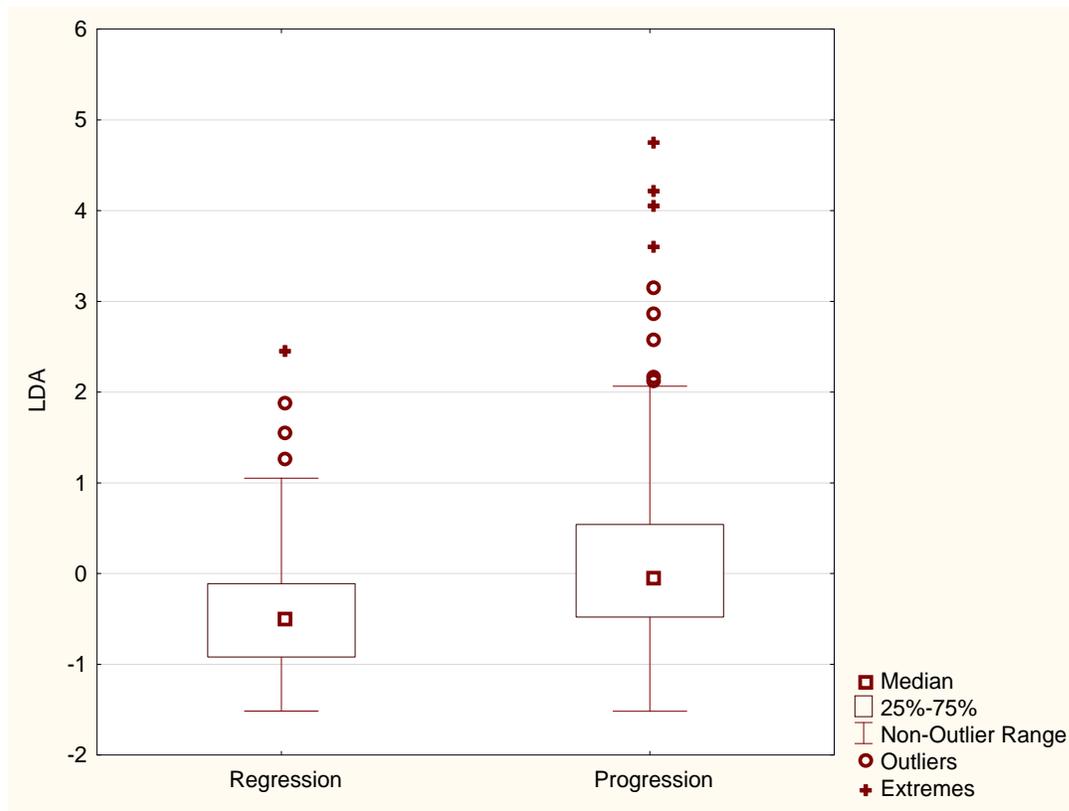


Figure 2.19 Box plot of the linear discriminant function values for the two groups of progressing and regressing lesions.

To evaluate the predictive value of Ki67 features, survival analysis was performed. The goal was to determine whether there was a statistically significant difference between the “survival times” of lesions with high and low Ki67 features. The endpoint in the survival analysis was progression to mild dysplasia or higher. The “survival time” for each “complete” (uncensored) case was therefore the duration of the period between the time of the baseline biopsy and the time of progression of the lesion to mild dysplasia or higher. The lesions that progressed to mild dysplasia or a higher histological grade were considered “complete” cases. The lesions that did not progress were considered “censored” cases.

For each Ki67 feature, two groups of biopsies were formed using the upper 25 and lower 25 percentiles of that Ki67 feature as high and low thresholds, respectively. Group one contained all lesions whose Ki67 feature values were higher than the high threshold, while group two contained lesions whose Ki67 feature values were lower than the low threshold.

The survival curves of the two groups of lesions were estimated using the Kaplan-Meier method and were compared using the log-rank test.⁷² The sensitivity and specificity of each Ki67 feature in identifying the progressing and regressing lesions in these two groups of cases were also determined. Sensitivity and specificity were calculated as follows:

	High Ki67 feature value (group 1)	Low Ki67 feature value (group 2)
No. of cases progressed	H1	L2
No. of cases regressed	H2	L1

(2.3)

$$\text{Sensitivity} = \frac{H1}{H1 + H2}$$

$$\text{Specificity} = \frac{L1}{L1 + L2}$$

Table 2.6 shows the results of the comparison of survival curves. In addition to individual Ki67 features, the survival curves were also estimated and compared for the linear discriminant function of Table 2.5. For *Ki67-index*, the Kaplan-Meier survival curves were not significantly different between groups one and two. However, for several newly proposed Ki67 features the survival curves were significantly different between groups one and two. These features were: *Positive-neighbor-number*, *Distance-average*, and *Positive-per-basal-length*. The survival curves for the *Ki67-index*, the *distance-average*, and the linear discriminant function are shown in Figures 2.20 to 2.22.

Table 2.6 Results of the comparison of Kaplan-Meier survival function of two groups of cases with high and low Ki67 feature values. The comparison was done using a log rank test. The log-rank test statistic and the corresponding p-value are presented. *p*-values smaller than 0.05 are highlighted. LDA represents the linear discriminant function.

	Log Rank	<i>p</i>-value
<i>Ki67-Index</i>	1.511	0.131
<i>Layer-average</i>	1.585	0.113
<i>Positive-neighbor-number</i>	2.112	0.035
<i>Positive-neighbor-fraction</i>	1.725	0.085
<i>Distance-average</i>	2.531	0.011
<i>Distance-90</i>	1.776	0.076
<i>Positive-per-basal-length</i>	2.042	0.041
<i>Lower-half-layer-fraction</i>	1.532	0.125
<i>Layer-90</i>	0.492	0.622
LDA	3.214	0.001

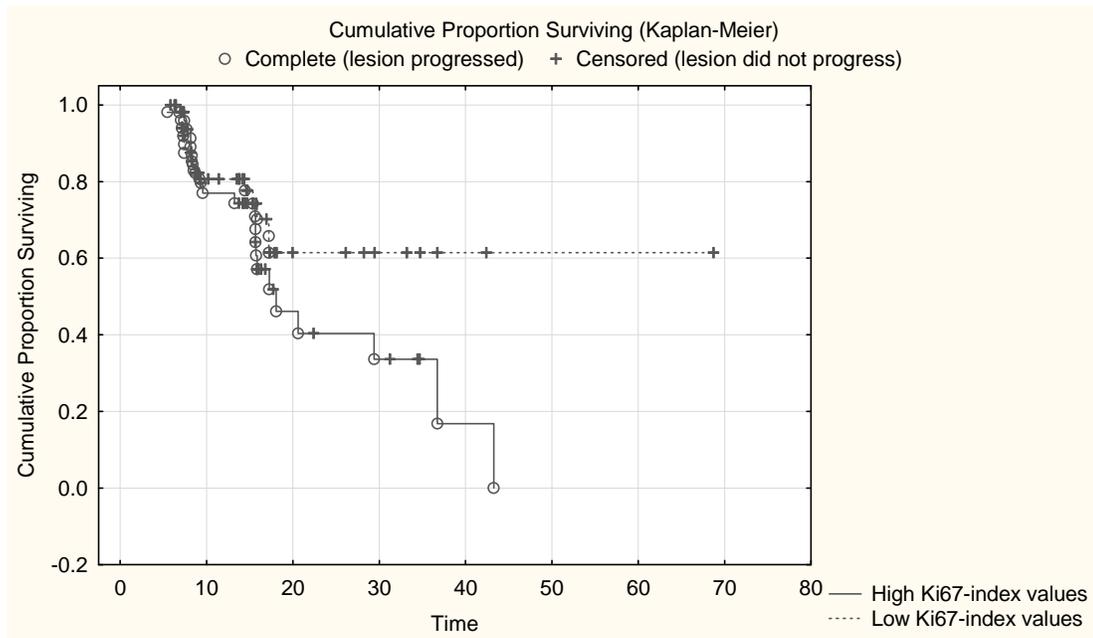


Figure 2.20 Kaplan-Meier survival functions for two groups of lesions with low and high *Ki67-index* values.

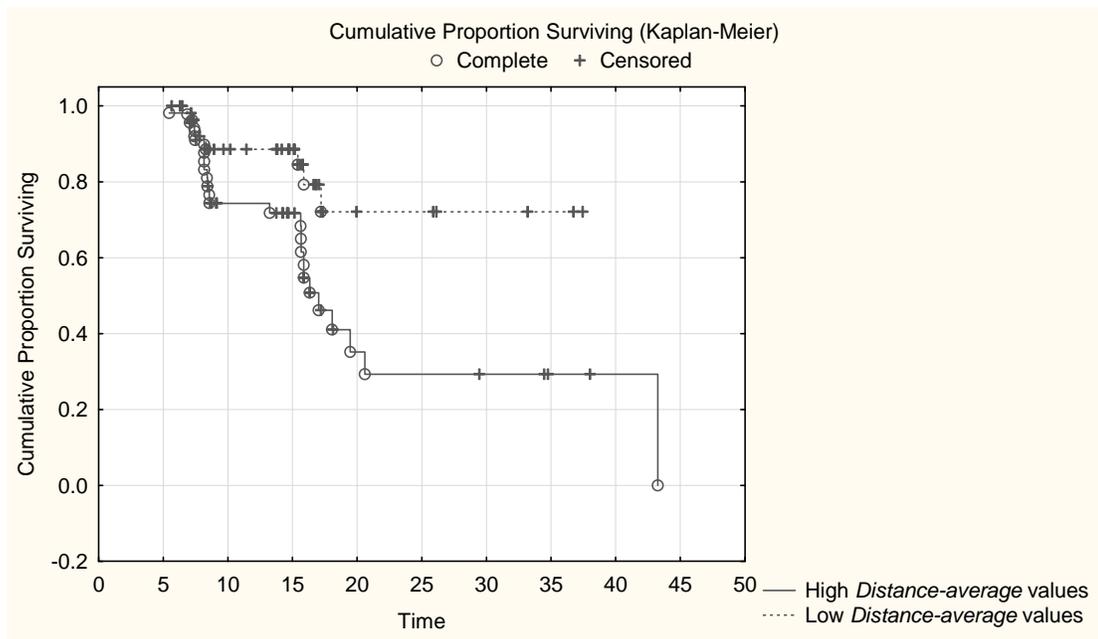


Figure 2.21 Kaplan-Meier survival functions for two groups of lesions with low and high *Distance-average* values.

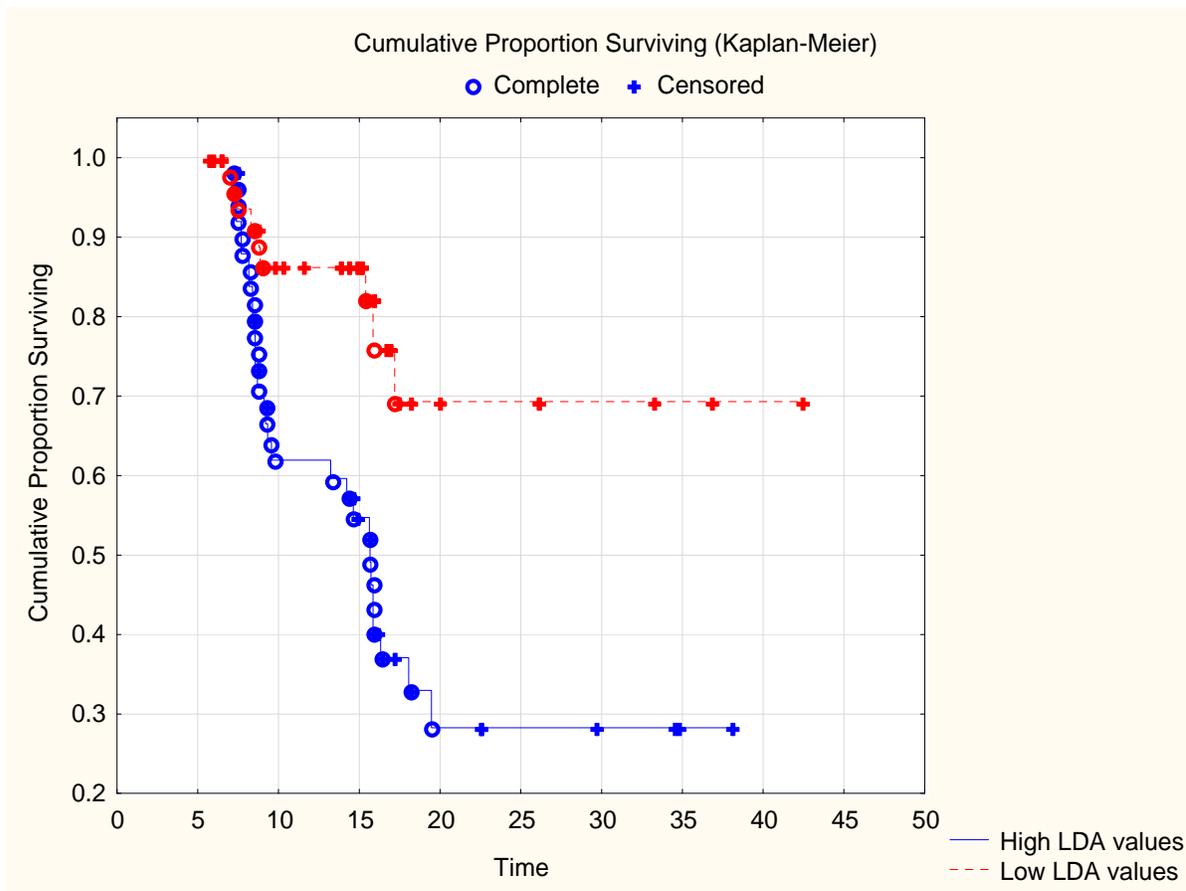


Figure 2.22 Kaplan-Meier survival functions for two groups of lesions with low and high LDA values.

Table 2.7 presents the sensitivity and specificity of each feature, as defined in equation (2.3). Even though only 50% of the cases with the highest and lowest Ki67 feature values were considered in calculating the sensitivity and specificity, the obtained sensitivity and specificity values were generally low (see Table 2.7). The linear discriminant function provided the highest sensitivity and specificity (56% and 83%, respectively). The linear discriminant function, however, was trained and tested on the same data; therefore, it might have been over-trained (with the limited number of cases available in our cohort, cross-validation was not possible). The sensitivity and specificity of the linear discriminant function are therefore only considered the upper bounds of the sensitivity and specificity achievable through a linear combination of these Ki67 features. Among the individual features, the *Distance-average* provided the highest sensitivity and specificity, respectively 44% and 84%.

Table 2.7 Sensitivity and specificity of each Ki67 feature in identifying progressing and regressing lesions in two groups of cases with high and low Ki67 feature values. LDA represents the linear discriminant function.

	Sensitivity	Specificity
<i>Ki67-Index</i>	0.42	0.74
<i>Layer-average</i>	0.39	0.77
<i>Positive-neighbor-number</i>	0.45	0.75
<i>Positive-neighbor-fraction</i>	0.41	0.75
<i>Distance-average</i>	0.44	0.84
<i>Distance-90</i>	0.43	0.78
<i>Positive-per-basal-length</i>	0.39	0.78
<i>Lower-half-layer-fraction</i>	0.40	0.76
<i>Layer-90</i>	0.36	0.73
<i>LDA</i>	0.56	0.83

Due to the low sensitivity and specificity of Ki67 features in identifying the progressing lesions, instead of attempting to discriminate progressing lesions based on these features, we attempted to determine if Ki67 features correlated with the survival times. Cox proportional hazards regression model⁷² was used to model the progression hazard rate as a function of each Ki67 feature.

Before this analysis was performed, all features were linearly normalized to have a mean of zero and standard deviation of one, so the hazard ratios were comparable from one feature to another. Table 2.8 shows the hazard ratio for a one unit change in the predictor (individual Ki67 feature) as well as its 95% confidence interval and *p*-value. For comparison, the same parameters were also estimated for the linear discriminant function of Table 2.5. Higher *Ki67-index* was not significantly associated with higher hazard ratio; however, several proposed Ki67 features were significantly correlated with the hazard ratio.

2.6 Discussion

Uncontrolled proliferation is a hallmark of cancer and is believed to be a biologically-plausible risk biomarker for epithelial malignant transformation. The Ki67 immuno-marker is expressed in all active parts of the cell cycle and can be used to interpret the growth process of tissue. Further, Ki67 immunostaining has been optimized for use in clinical histopathology labs, meaning that novel analyses of Ki67 expression with clinical utility will

be rapidly adopted for patient management, specially that they are relatively inexpensive to implement.⁴⁴

Table 2.8 Hazard ratio and 95% confidence interval of each Ki67 feature obtained using the Cox proportional hazards model. LDA represents the linear discriminant function.

	Hazard Ratio	Hazard-Ratio Confidence Interval		p-value
		Lower limit	Upper Limit	
<i>Ki67-Index</i>	1.190	0.952	1.487	0.127
<i>Layer-average</i>	1.251	1.013	1.546	0.038
<i>Relative-layer-average</i>	0.889	0.716	1.103	0.286
<i>Positive-neighbor-number</i>	1.382	1.077	1.774	0.011
<i>Positive-neighbor-fraction</i>	1.389	1.083	1.782	0.010
<i>Positive-neighbor-distance-overall</i>	1.013	0.794	1.293	0.917
<i>Positive-neighbor-distance-positive</i>	0.564	0.327	0.973	0.040
<i>Distance-average</i>	1.273	1.026	1.578	0.028
<i>Distance-90</i>	1.223	0.979	1.526	0.076
<i>Relative-distance-average</i>	0.923	0.735	1.159	0.491
<i>Relative-distance-90</i>	0.851	0.674	1.075	0.177
<i>Deep-third-fraction</i>	1.207	0.967	1.506	0.097
<i>Middle-third-fraction</i>	1.123	0.894	1.411	0.319
<i>Upper-third-fraction</i>	0.921	0.709	1.197	0.539
<i>Positive-per-basal-length</i>	1.196	0.970	1.475	0.094
<i>Basal-positive-per-basal-length</i>	0.962	0.759	1.219	0.749
<i>Basal-layer-fraction</i>	1.019	0.812	1.279	0.871
<i>Lower-half-layer-fraction</i>	1.257	1.002	1.576	0.048
<i>Upper-half-layer-fraction</i>	1.029	0.812	1.304	0.812
<i>Layer-90</i>	1.240	0.995	1.546	0.055
<i>Relative-layer-90</i>	0.882	0.700	1.110	0.283
<i>LDA</i>	1.869	1.280	2.729	0.001

We analyzed Ki67 expression in preneoplastic bronchial lesions to obtain an objective and quantitative representation of its distribution using two different mathematical methods of structure analysis in microscopic images of immuno-stained tissue cross-sections: the distance transform and Voronoi tessellation.

The Ki67 index is the percentage of epithelial cells expressing Ki67. It is the main Ki67 quantitative feature that has been previously investigated in preneoplastic and neoplastic

lesions in the bronchial epithelium.²² We observed that the Ki67 index increased from normal to metaplasia to hyperplasia to mild dysplasia and, finally, to moderate/severe dysplasias (analyzed in combination), with a wide range of variation in a given histological grade (Figure 2.4). These findings agree with previous reports on the proliferative activity in preneoplastic bronchial lesions.^{22,57}

We also studied and assessed additional features of the distribution of the Ki67 expressing cells. Overall, the highest distance from the basement membrane of the Ki67-positive nuclei, both in terms of distance (μm) and Voronoi cell layers, was found to increase with increasing histology grade (Figures 2.5 and 2.6). This was true for both the absolute distance and the distance normalized by the thickness of the epithelium. Also, on average, the number of the positive neighbors of the positive cells, a measure of how closely Ki67-positive nuclei were clustered, was higher in lesions with higher histological grades (Figure 2.9).

Our results suggest that, overall, individual Ki67 features are incapable of providing exact discrimination between histological grades. However, this does not necessarily mean that Ki67 expression is not able to predict malignant potential. Histopathological grading, though considered the “gold standard” for assessing the malignant potential of a given epithelial lesion, has significant difficulty in accurately predicting the progression likelihood for the individual pre-neoplastic lesions.⁷³ For example, in a study involving repeated longitudinal bronchoscopy examinations and histologic sampling of 52 at-risk individuals, the per-patient analysis showed that there was no significant difference between the rate of progression to CIS/invasive carcinoma for individuals with severe dysplasia and for the individuals with lower grade dysplasia.⁴⁰ Another factor is that the reproducibility of the histology grading system is not high, a reality that also affects the utility of this metric in a clinical context. A previous study on observer variability in histopathological reporting of preinvasive bronchial neoplasia reported Kappa values of 0.44, 0.31, 0.02, 0.46, and 0.57, respectively for the agreement on diagnosis of hyperplasia, mild, moderate, severe dysplasias, and CIS.⁷⁴ In another study, the Kappa values for inter-observer agreement on a 5-point grading system (metaplasia; mild, moderate, and severe dysplasia; and CIS) was reported to be 0.55.⁷⁵ Hence, the histological grading for preneoplastic lung lesions is neither accurate nor reproducible. In addition, the histological grading cannot reliably

differentiate the true malignant potential of preneoplastic squamous lesions.⁷⁶ Consequently, our ultimate goal was not the application of these Ki67 features to discriminate among histology grades. Instead, these features were used to evaluate the progression likelihood, using early lesions with longitudinal follow-up and known progression status. The fact that Ki67 predicts progression in early Cervical Intra-epithelial Neoplasia (CIN) supports this approach.⁵³

We also studied the inter-observer and intra-observer effects on the reproducibility of the Ki67 features. We found that some of the Ki67 features that describe the spatial distribution of Ki67-positive cells in the epithelium (e.g. those explaining the distance or the relative distance of positive nuclei from the basement membrane) were more reproducible than the Ki67 index (Table 2.1 and Figure 2.3). Specifically, these distance features were less susceptible to inter-/intra-observer bias than the Ki67 index.

We also calculated the fractions of Ki67-positive cells at different distances from the basement membrane and obtained plots of the distribution of Ki67 expressing cells in the epithelium for different histology grade categories (Figures 2.16 and 2.17). From these plots, we observed that most proliferation was generally limited to layers 2, 3, and 4 of the epithelium, suggesting that these were the transient amplifying (TA) cells produced by the stem cells in the basal layer. We also observed that the number of layers involved with proliferation increased with histology grade. These data demonstrated that the percentage of basal cells that proliferate was tightly controlled; the probability of this observed pattern being generated by chance on a sample-by-sample basis was very low ($p < 0.05$) for the majority (>70%) of samples. This statistically demonstrated that the basal cells operated under asymmetric division rules consistent with stem cell-like behavior.⁶⁸ In other words, our results suggested that, as a basal cell divides and produces daughter cells, one of these cells stays attached to the basement membrane and divides infrequently relative to the TA daughter cell that rises within the epithelium. This latter cell continues to divide more frequently.

The work we present here is consistent with some of the previous studies evaluating the expression of Ki67 in other epithelial pre-cancers. Ki67 has been extensively studied in CIN and is known to be a sensitive biomarker for grading these lesions into the escalating dysplastic groups of CIN1, CIN2, or CIN3.^{50,77} These studies have also shown that Ki67-

positive cells are found at progressively higher levels within the epithelium during progression through CIN stages. These results from CIN are consistent with the patterns of Ki67 expression we observed in bronchial epithelia. Our findings also align well with the results reported for esophageal squamous dysplasias, where the topological distribution of Ki67-expressing cells expanded in the lower layers of the epithelium during progression, a pattern we also observed in precancerous lung epithelia (Figure 2.11).

Interestingly, previously reported Ki67 expression patterns in vulvar intraepithelial neoplasia (VIN) did not align with our results. A group determining Ki67 positivity indices for these lesions (calculated in 5 μm strata from the basement membrane and using the area of positive nuclei pixels divided by the area of all nuclei pixels) reported that 1) Ki67 expression differed among the various vulvar premalignancies and that 2) normal vulvar epithelia had a Ki67-negative basal cell layer while other vulvar premalignancies had basal cell layers with significantly different Ki67 indices.⁵⁴ The results from bronchial precancerous lesions we reported here are dissimilar to those reported for VIN. One interpretation of this result is that the Ki67 expression patterns of epithelial premalignancies may differ based on tissue type. We also postulate that the use of the first Voronoi cell layer as the basal cell layer might be more accurate for identifying basal cells than dividing the stratum into 5 μm layers – and that this difference in the approach may account for the observed differences.

After validating the Ki67 features through evaluating their association with histology, we studied the predictive power of these features in terms of progression to a higher histological grade. Three different tests were used:

- 1) A T-test was performed on each individual feature to evaluate the differences in the means between a group of progressing lesions and another group of regressing lesions. Several Ki67 features showed small but statistically significant differences in their means.

- 2) For each feature, a survival analysis was performed on 25% of the highest feature values and 25% of the lowest feature values to see if there are any significant differences between the progression-free survival probability of lesions with low and high ki67 feature values.

3) The effect of each Ki67 feature on the hazard rate of progression was studied through Cox proportional hazards model. The results showed that some Ki67 features were strongly associated with progression, while the global Ki67 index was not.

Overall, the sensitivity and specificity of all Ki67 features were too low to allow the use of these features for prediction of progression versus regression, but the use of spatial characteristic increased the association of Ki67 features with progression.

There are limitations to the study in this chapter: 1) WHO regards severe dysplasia and CIS to be preinvasive lesions of lung cancer. Consequently intraepithelial lesions are typically divided to two main groups in chemoprevention studies or studies that examine the natural history of such lesions: low grade lesions consisting of metaplasia, mild and moderate dysplasia, and high grade lesions consisting of severe dysplasia and CIS. In the data available to this study, the number of cases progressing to high grade lesions was limited. Therefore the endpoint that was used in the outcome analysis was progression to or persisting as mild dysplasia or higher. The risk of progressing from metaplasia to dysplasia may not be the same as risk of developing severe dysplasia or CIS. Thus, even though some Ki67 features are associated with progression to dysplasia, their associations with progression to high grade lesions still need to be verified. For this verification, access to a wider range of intraepithelial lesions is needed. 2) The follow-up method used in this study (autofluorescence bronchoscopy followed by biopsy) is not optimal as we cannot exclude the possibility that some of the modifications observed in this study were in fact related to the bronchial sampling procedure, by disruption, partial or complete removal of the lesion. This limitation is not unique to our study. As long as histology remains the gold standard in identifying intraepithelial lesions, there is no other means of following up the lesions. Therefore there is a need for non-invasive *in vivo* imaging methods than can replace the histologic examination of tissue. The effect of biopsy will not be known until accurate and reproducible methods of lesion classification using noninvasive means are developed.

Ultimately, different analytical measures of Ki67 expression will need to be evaluated across epithelial premalignancies. This will give us a clearer picture of both the cell division behaviors underpinning disease progression and the most effective approach for using Ki67 expression as a biomarker in the management of early stage lesions. To interpret the growth processes, not only must we evaluate the proportion of proliferating cells, but we must also

assess their distribution throughout the tissue as well as the inter-relationships amongst these cells.

2.7 Summary

We reported means for objectively and quantitatively assessing the distribution of an immunohistochemical biomarker in the epithelial tissue. Several features recording the spatial behavior of proliferating cells were proposed.

We applied this means to evaluate and characterize the spatial distribution of proliferating cells in normal and pre-neoplastic epithelial regions, for the first time ever.

We validated these features using their correlation with “gold standard” histological evaluation. Given the specific challenges associated with reproducibility of histopathological evaluation of early lung lesions, and the Ki67 labeling index specifically, we also evaluated the reproducibility of these new features. Our results indicated that these features were indeed more reproducible than the Ki67 labeling index. These features are therefore more likely to have utility as biomarkers for disease management.

These features were then evaluated as a means of predicting the progression likelihood in early stage lesions with follow-up data. Ki67 features that explain the spatial distribution of proliferating cells in the bronchial epithelium have significant association with progression, while neither the histology nor the Ki67 index does. This proves our hypothesis that the knowledge of the spatial behavior of proliferating cells increases the biological relevance of proliferation. However, even though there is a statistically significant association between the proposed Ki67 features and the progression, the predictive power of Ki67 features alone is insufficient for discriminating between progressing and non-progressing preneoplastic lesions. To be used as a definitive surrogate endpoint biomarker, the proliferation information may need to be accompanied by the apoptosis and cell survival biomarkers and other risk-factors.

Given that our approach is based on an immuno-marker that is already widely used in clinical laboratories, rapid post-validation adoption of this biomarker for clinical management should be highly feasible.

Based on our results pertaining to the various measures of the distance of Ki67-positive cells from the basement membrane, we also conclude that the percentage of basal cells that

proliferate during progression is tightly controlled and that the basal cells operate under asymmetric division rules consistent with stem cell like behavior and the production of transient amplifying daughter cells.

In summary, assessment of the spatial distribution of proliferating cells provides a useful adjunct to the standard evaluation of proliferation in lung pre-cancers based on Ki67, an immunohistochemical biomarker that is already widely available, robust, and optimized for use in clinical histopathology laboratories.

Chapter 3. Quantitative Tissue-based Biomarker Analysis Using Multispectral Imaging

3.1 Introduction

The objective of this chapter is to develop an automated image analysis framework for the cellular-level quantification of molecular biomarkers expression in histopathological tissue sections in an objective and reproducible fashion. The analysis methods are ultimately aimed to be used in routine clinical settings. Thus, the robustness of these methods is an important consideration.

Factors that affect the performance of this image analysis system are considered. These factors include the choice of camera and illumination source, the optical performance of the labels used to visualize the biomarkers, and the presence and degree of multiplexing of biomarkers. Algorithms for the extraction of quantitative information about biomarker abundance and location from the imaging data are developed.

The biomarkers in the samples available in this study are labeled by the Immunohistochemistry (IHC) assay. In this thesis, we address the hypothesis that the additional information provided by multispectral imaging over RGB imaging can be used to improve the quantitative analysis of IHC biomarkers in tissue sections, on a cell by cell basis. The quantitative analysis of an IHC biomarker in a tissue section that is immunohistochemically labeled for that biomarker and is also counterstained with a nuclei label involves the following tasks:

- 1) To identify and separate the label representing the nuclei and the label representing the biomarker based on the absorption properties of the labels,
- 2) To robustly and reliably delineate (i.e. segment) the labeled nuclei in the histological tissue section, to enable the measurement of biomarkers on a cell-by-cell basis, and
- 3) To measure the abundance of the labeled biomarker in each cell within the cell compartment in which it is present (a biomarker of interest may be present in the cell nucleus, cytoplasm, or membrane).

We hypothesize that multispectral imaging improves the accuracy of biomarker measurement results. To prove this hypothesis we compare the effectiveness of the multispectral imaging approach against RGB imaging in terms of the accuracy of the ultimate results of the nuclei segmentation and measurement of the IHC biomarkers.

The development of algorithms for the automated segmentation of labeled cell nuclei in histological specimens has been the focus of a lot of research. In spite of all these efforts, the robust and fully automated segmentation of nuclei is still a challenging problem. This is because of the inherent variability in biological specimens, as well as variations in preparation and staining procedures. Additionally, nuclei are often found in dense clusters in which they are touching or appear to be overlapping.

For a true cellular-level quantitative analysis, once the nuclei are delineated (segmented), the locations of the cells are determined, and then the biomarker abundance is measured within each cell.

The rest of this chapter is organized as follows:

In section 3.2, some background and a summary of the related work on imaging and quantification of immunohistochemistry labeled tissue sections is presented.

Section 3.3 presents the samples available for this study, and the imaging system that was assembled and used for the capture of all the images.

Section 3.4 focuses on the task of identifying the distribution of the labels in a tissue sample through “linear spectral unmixing”. The utility of the spectral unmixing methods is demonstrated on four sets of IHC stained tissue samples. The first three sets contain two labels: one chromogen for the visualization of the IHC biomarker (Diaminobenzidine or DAB), and one nuclear counterstain (Hematoxylin) to label all the nuclei. These three sets are immunostained for different biomarkers. In the first set the biomarker (Ki67 antigen) is bound to the cell nucleus, in the second set the biomarker (P16 antigen) is in both cell nucleus and cytoplasm, and in the third set the biomarker (CD8 antigen) is bound to the cell membrane. The localization of the IHC biomarker in the cell changes the spatial co-localization of the labels (the counterstain and the chromogen): in the first set, the IHC label spatially overlaps the nuclei label in all the immuno-positive cells. In the second set, there are areas of spatial overlap between the labels, while in other areas; pure (unmixed) labels are present. In the third set, no spatial overlap between the labels is expected. The fourth set

contains three labels: one counterstain to label all nuclei (Hematoxylin), one chromogen (Vector Red) to label a nucleus bound IHC biomarker (Ki67 antigen), and another chromogen (DAB) to visualize the second IHC biomarker (P16 antigen), which is localized in nuclei and cytoplasm.

The immunohistochemical markers (Ki67, P16, and CD8) were chosen so that examples of antigens that are localized in the three main compartments of cells (nucleus, cytoplasm, and membrane) would be analyzed. Lung dysplastic tissue biopsies immunostained for Ki67 were used as examples of nuclear IHC biomarkers. Ki67 is one of the most commonly used markers of cell proliferation. It is routinely applied to biopsies of cancerous or pre-cancerous lung and cervix lesions.^{28,78} Cervical dysplastic tissue biopsies immunostained for P16 were used as examples of cytoplasmic IHC biomarkers. P16 is an important tumour suppressor protein that regulates the cell cycle. It is commonly used as a diagnostic marker of cervical neoplasia.⁷⁸ Among the proteins that are expressed in the cell membrane, the most widely used is Her2-neu that is the target of the drug Herceptin. Herceptin is prescribed for breast cancer tumours that over-express Her2-neu. Since samples immunostained for Her2-neu were not available for this study, as an alternate protein that is expressed in the cell membrane, we used lymphoma tumour biopsies that were immunostained for CD8. CD8 is expressed in the cell membranes of T cell lymphoblastic lymphoma.⁷⁹

Section 3.5 presents the application of the Linear Discriminant Analysis as an alternative way of discriminating amongst the labels.

In section 3.6, we present a nuclear segmentation algorithm and validate its performance on three sets of IHC images.

Section 3.7 focuses on the measurement of the IHC biomarker on a cell-by-cell basis in tissue samples. We present an approach to classify each cell as positive or negative, based on the presence or absence of the IHC biomarker of interest in that cell.

Finally, section 3.8 provides a summary of the results.

3.2 Background

3.2.1 Immunohistochemistry

Interpretation of tissue specimens by light microscopy is a common approach in clinical cancer pathology. Histopathological examination of human tissues with a potential malignancy is performed on millions of samples world-wide to provide a definitive diagnosis after preliminary symptoms or early cancer screening tests to indicate the possible existence of cancer or pre-cancer. Histopathologic evaluation involves examination of the tissue under the microscope by a pathologist. This tissue is obtained by biopsy, which is the surgical removal of a small piece of tissue for microscopic examination. Cancerous or pre-cancerous tissues have distinctive morphological features and appearance under the microscope. Among the traits the pathologist looks for are a large number of irregularly shaped dividing cells, variation in nuclear size and shape, variation in cell size and shape, loss of specialized cell features, loss of normal tissue organization, and a poorly defined tumor boundary.

However, there are many cases where this information may be supplemented by the evaluation of tissue-based molecular biomarkers. Molecular biomarkers are biological molecules that are indicators of the physiological state and of changes during a disease process. Molecular biomarkers may have uses beyond diagnosis only. They can provide insight into the molecular complexity of the disease process, create new cancer screening tools, inform the design of new treatments, monitor treatment effectiveness, and predict patient response to treatment. Patients may receive significant benefits from the routine application of the information contained in these biomarker measurements. One successful use of molecular biomarker information in personalizing treatments is the development of the drug Herceptin.⁸⁰ Herceptin targets tumor cells that overproduce a protein called Human Epidermal growth factor Receptor 2 (HER2) and interferes with their growth. It is approved for the treatment of early-stage breast cancer that is Human Epidermal growth factor Receptor 2 (HER2)-positive. Its success is leading scientists to search for other protein targets and quantitative protein assessment methods that could be used to personalize treatment and improve predictions of patient prognosis upon diagnosis.

Molecular biomarkers could potentially be represented by changes in gene copy number, mRNA expression or protein expression levels. Immunohistochemistry (IHC) is the most

practical and widely used method of assessing protein expression changes in histopathological specimens.

IHC is the process of localizing proteins/antigens in tissue sections. It uses labeled antibodies as specific reagents through antigen-antibody interactions that are visualized by an enzyme, fluorescent dye, radioactive element, or nano-particles.⁶

IHC involves a series of steps, typically beginning with antigen retrieval. Methods of antigen retrieval vary in terms of reagents and methods with the standard goal of unmasking antigens hidden by formalin cross-links or other availability limitations due to fixation methods. The first definitive step of IHC following antigen retrieval is the application of a specific primary antibody (typically produced by immunizing a mouse or rabbit with a peptide/antigen of interest), followed by extensive washing to remove excess amounts of primary antibody. A species-specific secondary antibody (a general secondary antibody that recognizes all antibodies originating from the same animal species as the primary), which binds to the primary antibody, is then applied.

Visualizing an antibody-antigen interaction can be accomplished in a number of ways. In the most common instance, an antibody is conjugated to an enzyme, such as peroxidase, that can catalyze a color-producing reaction (immunoperoxidase staining). A detection reagent is then applied that includes a chromogen to visualize the localization of the primary antibody. Alternatively, the antibody can also be tagged to a fluorophore (also referred to as immunofluorescence).⁶

3.2.2 IHC Assessment

Immunohistochemical reactivity can be within the cells' nuclei, membranes, or cytoplasm. Assessment of reactivity can range from the presence or absence of the IHC biomarker to the extent of IHC staining, the intensity of staining, or a combination of extent and intensity. If intensity is being assessed then there has to be a positive control having different levels of staining that is included with each staining batch so that the staining intensity in test cases is compared against it. Having this standard should overcome the problem of batch variation and allow comparisons.

Assessing the extent of reactivity can also range from the quick eye-balling to careful counting. It can involve counting up to 1000 cells for reactivity to give a percentage or

counting the number of positive cells in 10 high-power fields. Less precise is the visual assessment of the proportion of cells staining throughout the tissue section, with user subdivision of this assessment into percentage ranges. There are variations in the ease of this process between antigens, e.g. counting of nuclear reactivity is easier than counting cytoplasmic reactivity.

Typically a percentage of positive IHC staining is derived as a method of assessment of the IHC biomarker. These derived percentages are divided into ranges which are reported as grades, e.g. 0 to 25 percent as grade 1, 25 to 50 as grade 2, and so on. This method is referred to as the “semi-quantitative” scoring process.¹³

Visual scoring systems were introduced in order to overcome variations, particularly for biomarkers that are used for making therapeutic selections. For example, the Histochemical Score (HSCORE) was introduced by McCarty et al.⁸¹ HSCORE incorporates both the intensity and the distribution of specific staining. It was introduced for the assessment of Estrogen Receptor, but has been used for a range of other biomarkers. The evaluation is recorded as the percentage of positively stained target cells in each of the four intensity categories which are graded as 0 (no staining), 1 (weak but detectable above control), 2 (distinct), and 3 (intense). For each tissue section, HSCORE is derived by summing the percentages of cells stained at each intensity, weighted by the grade of intensity of staining. That is: $HSCORE = \sum P_i(i+1)$ where $i=0, 1, 2, 3$ is the grade of intensity of staining and P_i is the percentage of cells with intensity grade i and varies from 0 to 100%.⁸² The Immunoreactive score (IRS) suggested for Estrogen Receptors in breast cancer tissue subdivides staining intensity into four categories (0: negative, 1: weakly positive, 2: moderately positive, 3: strongly positive). The percentage of positive cells is graded as 0,1,2,3, and 4 corresponding to 0% (negative), 1-10%, 11-50%, 51-80%, and more than 80% positive cells present. The IRS score is calculated by multiplying the intensity grade by the percent of positive cells grade.⁸³ Reiner’s system adds these two values together.⁸⁴ A different system has been introduced for another biomarker used for therapeutic selection in breast cancer, HER2. This is the HercepTest, United States Food and Drug Administration-approved grading system of 0, 1+, 2+, and 3+. For 3+ there must be strong membrane staining of the complete cell in >10% of tumor cells; 2+ is moderate staining, whilst 1+ is

weaker staining and the complete cell is not stained. Cytoplasmic staining is not considered.⁸⁵

These subjective scoring systems suffer from inter-observer reproducibility problems.⁸⁶⁻⁸⁹ Also they do not provide a truly continuous measurement of protein expression.

3.2.3 Quantitative IHC

While IHC has become an established method in research and clinical pathology, a surprising lack of standardization and reproducibility exists among different laboratories.¹³

The need for reliability and reproducibility of IHC findings has expanded due to the new roles of IHC which include: 1) the identification of prognostic molecular biomarkers that can identify subsets of patients at high or low risk upon diagnosis, as well as new targets for potential future therapeutic development and 2) the accurate distinguishing of the expression levels of biomarkers to determine whether or not a patient is a candidate for a specific treatment.^{90,91}

Enormous variability exists in terms of the reagents available, the detection methods used, and the interpretation and reporting of IHC findings. In order to achieve a useful degree of reliability and reproducibility with IHC staining, the total process must be standardized. This includes sample preparation (transportation, fixation, sectioning) and antigen retrieval, protocols, reagents, and controls.¹⁴

The lack of reliability of IHC methods can be partly attributed to the nature of the “semi-quantitative” scoring process and the intrinsic deficiencies of an observer-based, subjective manual method. The reporting of the outcome will vary from pathologist to pathologist or even for the same pathologist from day to day. Quantification of IHC however, can help eliminate the inherent variability of subjective visual scoring systems and may also increase the sensitivity and dynamic range of in situ measurement of protein expression.^{13,14,92} Computer-assisted image analysis is the solution to the scoring of IHC stains, where a quantifiable result is the desired outcome.

3.2.4 Image Analysis Solutions for Quantitative IHC

The main assumption in all approaches to quantitative IHC is that the IHC experiment has generated a valid signal for the imaging system to capture and that the signal on the slide

is representative and in some way quantitatively related to the abundance of the antigen of interest in the tissue section. The abundance of the antigen in the tissue section is related (in an unknown way) to the absolute amount of the antigen present in the original tissue.

The combination of IHC and computer-assisted image analysis systems provides the possibility of quantifying the IHC staining. Automated or semi-automated image analysis methods have been developed for quantification of IHC over the years with the goals of increasing objectivity, reproducibility and accuracy of results.

Analysis techniques to extract the quantitative intensity and area measurements from the acquired images need to consider factors such as the choice of camera, the illumination source, the optical properties of the labels, and the presence and degree of multiplexing, i.e. the number of labels used on the same slide. In a typical single-antibody IHC experiment, two dyes are used: one labels the specific molecular biomarker, which we refer to as the immunostain (typically DAB), and one that labels all the nuclei, referred to as the counterstain (typically Hematoxylin). The minimum number of labels used is therefore two. More recently, practical methods have been developed for double or triple antibody staining in absorbance.⁹³

The quantification begins with the separation and/or classification of multiple labels. Once labeled areas are located on the images, the quantification then involves obtaining the degree of labeling (staining) in a region of interest in a tissue. The region of interest can be the area of tumor region versus benign cells, or epithelial region versus stroma.

The typical approaches to the separation and/or classification of IHC signals are now briefly discussed:

- 1) One approach to separating these labels is to use specific wavelength filters so that one filter only allows the transmission of the IHC label signal and the other only allows the transmission of the counterstain signal. This results in two images: image of the counterstain and image of the IHC label. The counterstain image is used to identify nuclei. The IHC label image is used to identify the immuno-positive areas. Typically the ratio of the areas of immuno-positive regions to the nuclear regions is reported as the measure of average IHC staining per cell. In the case of antigens that are localized within the cell nucleus, quantification may be done using a technique known as “nuclear masking”. Briefly, the counterstain image is used to create a mask of the nuclear regions. The IHC signal is

quantified within this mask on a cell by cell basis. Any influence from the counterstain is removed by calibration techniques (e.g. use of control samples).⁹⁴⁻⁹⁷

This approach only works when the individual labels have non-overlapping absorption spectra. It thus requires specific labels and specific narrow-band filters. Even with specifically matched filters it is difficult to design a system where each label appears completely separated because partial overlap in the absorption spectra of the labels may exist. In the case where there is an extra-nuclear staining component, the quantification of IHC cannot be done on a per cell basis due to the fact that the automated identification of cytoplasm boundaries have proven to be too difficult and manual delineation of boundaries is subjective and time consuming. Thus, in such cases, quantification of IHC results in an average section-wide score.

2) In another approach, the true-color RGB images of the tissue section are captured. In some studies, only one of the red, green, or blue filtered grayscale images - the one with highest contrast between the colors of interest - is used to separate these colors.⁹⁸ In the work of Kuyatt et al., the blue image was used to identify the IHC stained regions and the red image was used to identify all counterstained regions.⁹⁹ In the work of Elmoataz et al. the green image was used to perform segmentation on the nuclei using active contours and mathematical morphology and the blue image was used to discriminate positive nuclei from other counterstained nuclei.¹⁰⁰

In other studies, the color information from each of the red, green and blue grayscale images has been used along with a combination of separate thresholds on individual grayscale images for the separation of the labels.¹⁰¹ Conversion of the RGB image to a hue, saturation, and intensity (HSI) image has also been used and was reported to improve the separation.^{102,103} Elie et al.¹⁰⁴ used a combination of thresholds set on the components of the RGB image and the HSI representation of the RGB image.

Converting the color image to a grey-scale image that provides better separation of positive areas has also been used. In the study of Brey et al.¹⁰⁵, a normalized blue image was defined by dividing the blue image by the sum of the three red, green, and blue images and was shown to outperform methods that use thresholds set on the individual components of the color image. In the work of Ruifrok et al.¹⁰⁶, a linear combination of the three components of the RGB image was used to define a “brown image” that highlighted the

DAB-labeled brown areas. Ranefall et al.^{105,106} used a principal component transform of the color space. The first principal component (PC1) was then used to extract objects of interest. PC1 was then regarded as lightness (as in the HSI color space) and the angle between every color in the color space and the second principal component was regarded as hue. The hue component was then used to distinguish the positive pixels from negative pixels.^{107,108}

Supervised and unsupervised classifiers have also been used for separating the labels based on the color information. In the work of Arambula et al.¹⁰⁹, a K-Nearest Neighbor classifier was used for color separation. Watershed segmentation combined with edge detection was then used to isolate individual nuclei. Ranefal et al.¹¹⁰ applied a pixel-wise classifier to the color images to distinguish between positive nuclei pixels and negative nuclei pixels.

None of the methods discussed thus far result in the separation of the contributions of two or more labels to the resulting color at each pixel of the image. Areas that are labeled with two or more dyes are designated to one of the colors depending on the threshold settings. Color deconvolution was proposed by Ruifrok et al.^{111,112}, as a means of separating the contribution of the labels to the intensity of each pixel. Color deconvolution is a method to deconvolve the color information acquired with RGB cameras, to calculate the contribution of each of the applied stains, based on the stain-specific RGB absorption. It is a special case of “linear spectral unmixing” performed on the three channels of the RGB image.

3.2.5 Quantitative IHC Using Multispectral Imaging

To quantify IHC biomarkers in labeled tissue sections, either a monochrome camera with 2 or 3 filters in a filter wheel or a color RGB (Red-Green-Blue) camera has traditionally been employed. However, the three broad bands of RGB cameras or the narrow bands of select filters may not result in best separation of the labeled components from the non-labeled components and the background. RGB color information only partially represents the true spectral content of an image.¹¹³

The optical density of a labeled area of the tissue section is calculated as the logarithm of the inverse transmittance of that area (i.e. the ratio of the incident light intensity to the transmitted light intensity). According to the Beer-Lambert law, the optical density is correlated with the concentration of the label. The concentration of a label is correlated with

the amount of the labeled substance. Therefore, the optical density can potentially provide a means for measuring the labeled substance (biomarker). However, the same law states that besides concentration, the optical density is also a function of the thickness of the specimen and of the molar absorptivity of the label(s). In the case where a single dye is used, these parameters can be considered as unknown constants and optical density can be used as a quantitative index of the amount of the label. But nearly in all cases, modern histology employs multiple dyes, which are used for labeling different tissue substances and structures. In these cases, the assessment of the amount of each particular label at any spatial point of the examined field is not possible, since the absorption spectra of the labels are typically broad and overlapping, while components labeled with different dyes may also be spatially overlapping. As a result, the color or the gray-level of the pixels of the conventional color or grayscale microscopic images and consequently the calculated optical density is determined by the total absorption from the multiple dyes. Therefore, quantitative information cannot be directly derived for each labeled substance and structure.¹¹³

Multispectral imaging and analysis techniques can be employed to extract quantitative information about the amount of the labeled components, despite the spatial and spectral overlap of the labels. For a given number of mixture components, imaging at multiple wavelengths enables the development of an equal number of simultaneous equations, which can then be solved for component concentrations. Spectral imaging may also contribute to the solution of challenges in segmenting histological tissue sections, since it may enable the enhancement of contrast between features under discrimination. The enhancement is based on the selection of proper imaging spectral band(s), at which the maximum difference in the spectral characteristics of these features is recorded.¹¹³

3.2.5.1 Multispectral Imaging Hardware

Several approaches are used for spatially resolved multispectral imaging of clinical samples. The main approaches employ: 1) multiple band-pass and/or narrow-pass optical filters (filter-wheels). Such filters lack flexibility and may not have enough spectral resolution to be useful for the analysis of spectrally complex specimens typical of labeled histological samples. 2) Spatially-resolved Fourier transform spectroscopy. This approach may have the best combination of sensitivity and spectral resolution because it captures

photons from all wavelengths simultaneously, although it requires the capture of multiple full-frame images followed by subsequent Fourier transform computation. And 3) tunable (acousto-optical or liquid crystal) filters. These filters have typically been used on the detection side, i.e. the light from the object is filtered into several narrowband wavelength channels. Therefore only photons that can pass through the wavelength window are detected and all other photons are rejected.¹¹⁴

3.2.5.2 Multispectral Image Analysis

Regardless of the technology used, a spectral imager delivers three-dimensional image sets (x, y, and wavelength intensity) that contain spectral information at every pixel. Spectral image data are rich in information and require appropriate analysis methods. Analysis methods that are typically applied are 1) pixel classification and 2) linear spectral unmixing. Pixel classification methods assign every pixel to a particular class based on its spectral information. The classes may be previously known through a spectral library, or may be derived from the image data, either through interactive selection or by unsupervised clustering and/or classification techniques. Linear Spectral Unmixing is the most common analytical tool used with multispectral biomedical images and will be discussed separately in the following section.

3.2.5.3 Applications of Linear Spectral Unmixing in Quantitative IHC

Spectral unmixing is the process of resolving a signal that is a mixture of several components into its constituents. The first step in the linear unmixing algorithms is the estimation of the spectra of the individual components in the mixture (also referred to as pure components). In the context of the immunostained samples, these components are the labels used for the visualization of the nuclei or the IHC biomarkers. Different methods have been used to estimate the spectra of the labels. These include: 1) measuring the spectral properties of the labels prior to the experiment, either in solution¹¹⁴ or in singly labeled tissue sections¹¹⁵, 2) interactive selection of the areas thought to represent pure labeling of each of the dyes by the user and using the spectra from those areas as the pure component spectra (this method only works if pure labeled regions exist in the sample)¹¹⁶, or 3) blind source separation¹¹⁷ techniques.^{118,119} There are also proprietary algorithms incorporated into

commercial systems that estimate the spectra of pure components computationally under interactive control.¹¹⁶

3.3 Image Acquisition

3.3.1 Data

Our dataset contained four sets of IHC labeled samples. These four sets are listed below:

1) Ki67 set: 15 Lung preneoplastic epithelial tissue sections immunostained for Ki67 antigen. Ki67 is expressed in the nuclei.

2) P16 set: 23 cervical dysplastic epithelial tissue sections, immunostained for P16 Antigen. P16 is expressed in both nuclei and cytoplasm.

3) CD8 set: 20 Lymphoma tumor tissue microarray cores immunostained for CD8 antigen. CD8 is expressed in the membrane.

4) Ki67 & P16 set: 7 dysplastic cervical epithelial tissue sections, immunostained for P16 and Ki67 antigens. P16 antibody is visualized by DAB, and Ki67 is visualized by VR chromogen.

In the first three sets, the antigen of interest was labeled with Diaminobenzidine (DAB), and nuclei were counterstained with Hematoxylin (H). In the fourth set, tissue sections were immunostained for two antigens, one of which was labeled with DAB and the other one was labeled with Vector Red (VR) chromogen. Nuclei were counterstained with Hematoxylin. All samples were formalin fixed, paraffin embedded tissue sections of 5 micron thickness.

3.3.2 Imaging System

The tissue samples were imaged under a Zeiss AxioScope microscope (Carl Zeiss Canada Ltd., Canada) with motorized z-stage using 20X or 40X magnification. A monochrome QImaging Retiga CCD digital camera (QImaging, Vancouver BC) was used for image capture. Multispectral illumination was generated by a tunable light engine: OneLight Spectra (OneLight Corp., Vancouver, BC, Canada). OneLight Spectra is a spectrally and temporally programmable light engine that can provide any desired mix of wavelengths in the 400nm-680nm range, can control the intensity at each wavelength independently, and uses a spectrometer based feedback loop to determine the wavelength intensity distribution output by the system. It uses the Texas Instruments Digital Micro-mirror Device (DMD), a

type of spatial light modulator, in combination with a lamp and related optics to control the wavelength dependent energy distribution of light.¹²⁰

Custom software was developed to control and synchronize the light engine, the microscope, and the CCD camera. The wavelength range of the illumination light, number of illumination channels, and the bandwidth of each illumination channel were specified through the software user interface. Under the software control, the light engine sequentially creates a narrow-band illumination profile. Under each illumination, the software automatically focuses the light on the sample, and an image is captured. A stack of images are thus generated.

Multispectral illumination performed in this work consisted of 12 narrow-band illumination spectra with central wavelengths of 420 nm, 440 nm, ..., and 640 nm. The bandwidth at half maximum (BWHM) of each illumination channel was 20 nm. For Ki67, P16, and Ki67 & P16 sets, images from the same field of view were also taken at three wideband channels corresponding to red, green, and blue channels of three-color RGB imaging, to form equivalent of RGB images. For CD8, RGB images were synthesized by adding up the images in the first five channels as the blue channel, the 6th to 10th channels as the green channel, and the last two channels (11th and 12th) as the red channel. Figure 3.1 shows the illumination profiles of both multispectral images and RGB images.

Each image in the 12-channel multispectral image stack and the three-channel RGB image was a 1280×1024 grayscale image with bit-depth of 8.

Before acquiring the multispectral image stacks, an auto-exposure routine was performed while imaging a blank (transparent) area of a slide to determine the exposure time necessary to illuminate the blank area to about 90% of the saturation gray level, 255. A multispectral “flat-field” image stack was then acquired, followed by capture of a multispectral image stack of the sample, with both image stacks being acquired using the same exposure times at the same spectral bands. A “dark-field” image stack was also captured by turning off the light engine, and using the same exposure times as the “flat-field” image stack. Each image in the multispectral image stack was then corrected for non-uniform illumination and background light as follows:

$$\text{Corrected Image} = \frac{\text{Captured image} - \text{darkfield image}}{\text{flat field image} - \text{dark field image}} \quad (3.1)$$

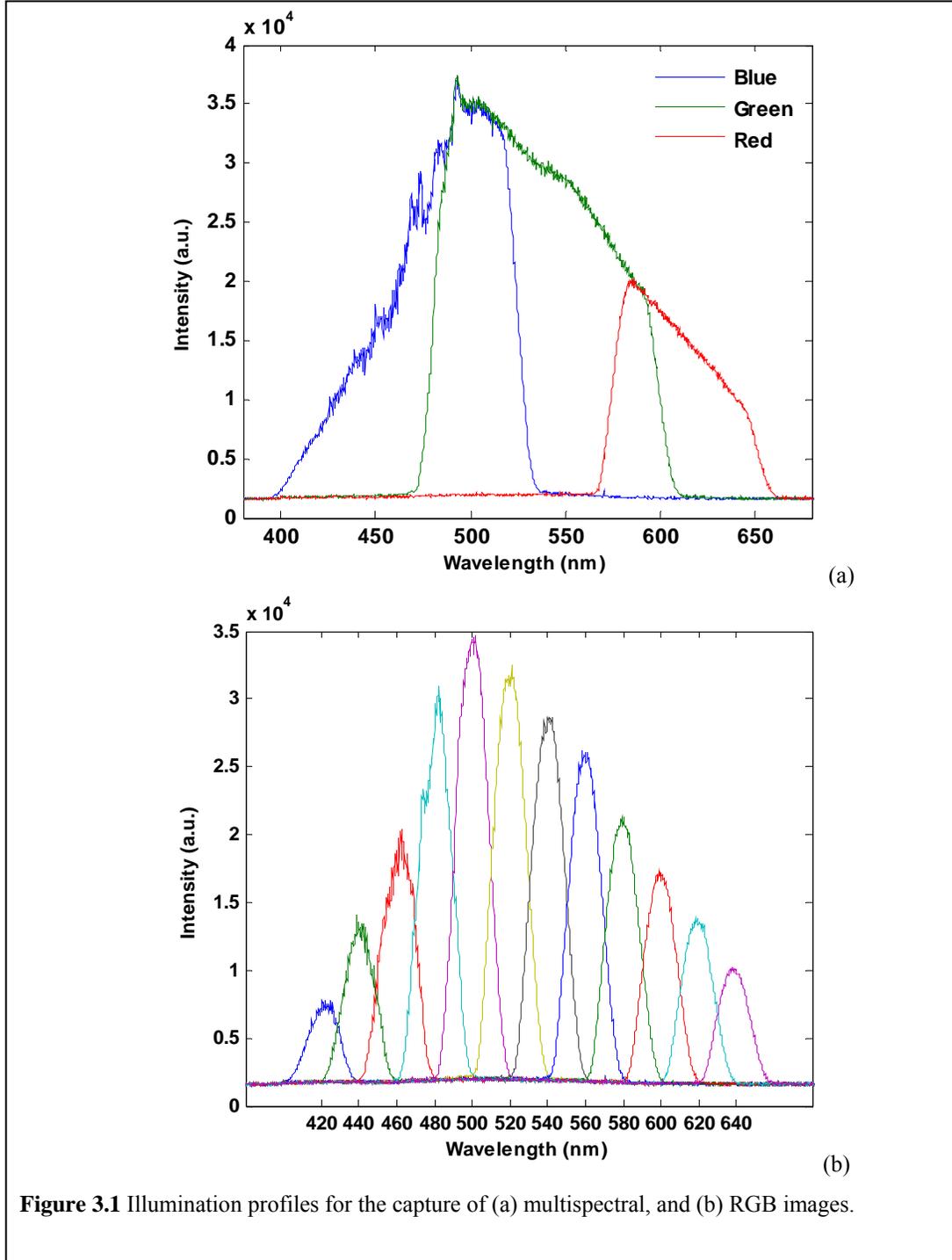


Figure 3.1 Illumination profiles for the capture of (a) multispectral, and (b) RGB images.

3.4 Linear Spectral Unmixing of Multispectral Images

Let's assume that a multispectral image stack of a sample that is a mixture of several components is available. The intensities of each pixel in the images within the multispectral image stack form the spectrum of that pixel. Linear spectral unmixing is the procedure by which the spectrum at each pixel is decomposed into the spectra of the components constituting the sample at that pixel. It also yields a set of corresponding "concentrations", that indicate the concentration of each component present in that pixel. Spectral unmixing has been investigated in hyperspectral remote sensing¹²¹, and in analytical chemistry^{122,123}.

Our approach is to apply linear spectral unmixing to separate the contributions of the labels (chromogens, counterstains, etc.) to the absorption at each pixel of the image. The distribution (concentration) map of each label can thus be generated. Then, the distribution map of the nuclear label can be used for the identification and segmentation of the nuclei. The distribution map of each IHC label can be used to measure the abundance (amount) of that IHC biomarker. Using the location of nuclei obtained by nuclei segmentation, the abundance of the IHC biomarker can be measured within each cell corresponding to each nucleus.

In the following section, methods for the application of linear spectral unmixing to the multispectral images of IHC stained tissue sections are presented.

3.4.1 Methods

3.4.1.1 Relationship between Transmission Images and Concentration of Dyes

In conventional optical microscopy, the sample is illuminated with visible light. The light transmitted from the sample travels through a system of lenses and creates an enlarged image of the sample. The transmission image may be viewed by the microscope eyepiece lens, or recorded by a light detector, e.g. charge coupled (CCD) camera. Such image is referred to as the transmission image.

In this section, we derive the relationship between transmission images of a sample and concentration of dyes (labels) within that sample.

We consider a sample (or specimen) that is a mixture of L components (dyes or labels). Let's consider the transmission images acquired at N illumination wavelengths. Let $\lambda_n, n =$

$1, 2, \dots, N$ be the wavelengths, and $\varphi_n(x, y), n = 1, 2, \dots, N$ be the captured images, where (x, y) refers to the coordinates of each pixel within the image.

According to the Beer-Lambert law¹²⁴, the optical density of a sample consisting of multiple absorbing components is linearly dependent on each component's molar concentration, and its absorption coefficient. Molar concentration is the amount of a component divided by the volume of the mixture. When light of wavelength λ_n passes through the sample consisting of L components, with component molar concentration $c_i(x, y)$ (mass per volume) and thickness $b_i(x, y)$, the absorbance (or optical density) $A_n(x, y)$ at wavelength λ_n is given by Beer-Lambert law:¹²⁴

$$A_n(x, y) = \log \left(\frac{\varphi_{0n}(x, y)}{\varphi_n(x, y)} \right) = \sum_{i=1}^L \varepsilon_{in} c_i(x, y) b_i(x, y), n = 1, 2, \dots, N \quad (3.2)$$

where $\varphi_n(x, y)$ and $\varphi_{0n}(x, y)$ are the captured images in the presence of the L components and in their absence, at wavelengths $\lambda_n, n = 1, 2, \dots, N$, respectively, and ε_{in} is the molar absorptivity coefficient of the i -th component at wavelength λ_n . If we define the mass density (mass per area, i.e. amount of a component divided by the surface area of the mixture) as $\delta_i(x, y) = c_i(x, y) b_i(x, y)$, equation (3.2) can then be written as:

$$A_n(x, y) = \sum_{i=1}^L \delta_i(x, y) \varepsilon_{in}, \quad n = 1, 2, \dots, N \quad (3.3)$$

From equation (3.3), it follows that if N sets of images $\varphi_n(x, y), n = 1, 2, \dots, N$, are acquired at N different wavelengths $\lambda_n, n = 1, 2, \dots, N$, we obtain a set of N linear equations. This set of linear equations can be written in the matrix form as:

$$A = \Delta E^T \quad (3.4)$$

where A is the $M \times N$ matrix of optical densities for M pixels within the image in N wavelengths, Δ is the $M \times L$ matrix of mass densities of the L constituents at each one of the M pixels ($\delta_i(x, y), i = 1, 2, \dots, L$), and E^T is the $L \times N$ matrix of mass absorption coefficients of the L constituents at N wavelengths ($\varepsilon_{in}, i = 1, 2, \dots, L$, and $n = 1, 2, \dots, N$).

According to equation (3.4), the relationship between the transmitted light from a sample consisting of several components and the components concentrations follows the “linear mixing model”. The linear mixing model states that the spectrum of a mixed pixel is a linear combination of the spectra of the components present in that pixel weighted by the concentration of those components in that pixel.

In the context of the immunostained samples studied in this chapter, the components are the labels used for the counterstaining of nuclei or for the visualization of IHC biomarkers. The spectrum of each pixel is the optical density of that pixel at multiple illumination wavelengths.

3.4.1.2 Solving the Linear Mixing Model

The linear mixing model can be formulated as:

$$D = CS^T + E \quad (3.5)$$

where D is the multispectral data matrix with m rows corresponding to m pixels in the image and n columns corresponding to n different wavelengths. C is the m by k matrix of concentrations of the k components where each row represents the composition of the k components at the corresponding pixel. S^T is the k by n pure component spectra matrix where each row is the pure spectrum of the corresponding component. E is the experimental error or noise.

Linear spectral unmixing is the procedure by which the measured spectrum at each pixel of a multispectral image is decomposed into the collection of spectra S^T of the components, and a set of corresponding fractions C that indicate the proportion or concentration of each component present in the pixel. In other words, linear spectral unmixing solves the linear mixing model of equation (3.5) for C and S^T .

The set of linear equations in equation (3.5) is solved for concentrations C of the components when the pure component spectra S^T are known. Alternatively, it is solved for both C and S^T , when no prior knowledge of the pure component spectra is available.

If spectra of the components comprising the sample are known, then the spectral unmixing problem is essentially finding the concentrations of those components, i.e. when S

is known in equation (3.5), the problem is finding the matrix of concentrations C from the observed spectra D .

The standard methodology that has been developed over the past several decades is to solve the above problem through finding \hat{C} that minimizes the second norm of the matrix of error e . e is calculated as:

$$e = \|D - CS^T\| \quad (3.6)$$

This is a classical least squares problem. The unconstrained least squares solution for C is:

$$\hat{C} = DS(S^T S)^{-1} \quad (3.7)$$

This estimate exists if there are more rows than columns in matrix D , and when S has full column rank. In practice, both spectra and concentrations are non-negative quantities, so a non-negativity constraint is applied to the above least squares problem. The constrained problem is typically solved through a non-negative least squares algorithm.^{125,126}

When neither the pure component spectra nor the concentrations are known, then spectral unmixing can be done through Multivariate Curve Resolution (MCR) techniques. MCR comprises a family of techniques that decompose signals from multi-component mixtures. Curve resolution methods extract simultaneously the pure component spectra and their concentrations. Among the different MCR techniques, iterative resolution methods are currently considered the most popular due to their flexibility in accommodating external information in the resolution process. These methods start from initial estimates of C or S^T that evolve to yield profiles with chemically meaningful shapes, tailored according to chemical or mathematical information included in the optimization process, under the form of constraints. The Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) is among these methods.¹²⁷

3.4.1.3 MCR-ALS

MCR-ALS finds the solution to equation (3.5) through an iterative Alternating Least Squares (ALS) algorithm. To initiate the iterative ALS procedure, an initial estimate is needed for the spectra or the concentration profiles of each component.

If the initial estimates are the spectral profiles S^T , the unconstrained least squares solution for the concentration profiles can be calculated from equation (3.7). Similarly, if the initial estimates are the concentration profiles C , the unconstrained least squares solution for the spectra can be calculated from the expression:

$$\hat{S}^T = (C^T C)^{-1} C^T D \quad (3.8)$$

when C is of full rank.

Both equations (3.7) and (3.8) are implemented in an alternating least squares cycle so that in each iteration, new matrices of C and S^T are obtained. However, after each iteration, a series of constraints with the purpose of resulting in solutions which are physically meaningful and limiting their possible complexity are applied, such as unimodality and equality. Iterations continue until an optimal solution is obtained that fulfils the constraints and the established convergence criteria. Convergence is usually based on the residual sum of squares RSS , which is calculated after each iteration as follows:

$$\hat{E} = D - \hat{C}\hat{S}^T, \text{RSS} = \sum_i \sum_j e_{ij}^2 \quad (3.9)$$

When RSS reaches a threshold fixed by the user according to the noise level, the system is considered to have converged.

3.4.1.4 Initial Estimates

The MCR-ALS algorithm requires initial estimates of either the pure concentration matrix C or the pure spectra matrix S as the first step of the optimization procedure. One technique for extracting estimates for the spectra of the pure components from the observed spectra D is the Simple-to-use Self modeling Mixture Analysis (SIMPLISMA), introduced by Windig et al.¹²⁸ SIMPLISMA is one of the first self-modeling curve resolution methods used in spectroscopy. The key point of SIMPLISMA is the selection of “pure” variables from the data matrix D . A “pure” variable is a variable which has contributions from only one component of the mixture. The multi-step SIMPLISMA algorithm works by calculating a purity value for each variable at each step. At the k -th step, the variable with the highest purity value is chosen as the k -th pure variable. The number of “pure” variables selected is equal to the number of components in the mixture. “Pure” variables can be “pure” pixels (rows of matrix D) or “pure” spectral bands (columns of matrix D). The spectra of “pure”

pixels or the images corresponding to “pure” spectral bands can then be used as initial estimates of spectra or concentrations, respectively, in the MCR-ALS procedure.

Here, we provide a summary of how the SIMPLISMA algorithm works: when searching for the purest variables in the rows of matrix D , the first purity value of a variable x_j , referred to as $p_j^{(1)}$, is defined by:

$$p_j^{(1)} = \frac{\sigma_j}{\mu_j + \alpha}, \text{ for } j = 1, \dots, m \quad (3.10)$$

where α is an offset parameter, m is the number of rows in the matrix D and:

$$\sigma_j = \sqrt{\frac{\sum_{i=1}^n (d_{ji} - \mu_j)^2}{n}} \text{ and } \mu_j = \frac{\sum_{i=1}^n d_{ji}}{n}, \text{ for } j = 1, \dots, m \quad (3.11)$$

where n is the number of columns in matrix D , and d_{ji} is the element in the j -th row and i -th column of matrix D . The offset parameter α is a user-defined value that is added to the denominator in equation (3.10) to avoid assigning a high purity value to a variable with a low mean value.

After calculating the first purity value for all variables, the spectrum of the variable with the highest purity value is chosen as the first pure spectrum. In the next steps, the contributions of the previously selected pure components are removed from all variables by using a weighting parameter $w_j^{(k)}$ that reduces the influence of variables that are correlated with the selected pure variables. $w_j^{(k)}$ is the weighting parameter for the j -th variable in the k -th step, i.e. after $k-1$ pure variables are selected. The k -th purity $p_j^{(k)}$ of a variable x_j is defined as:

$$p_j^{(k)} = w_j^{(k)} \frac{\sigma_j}{\mu_j + \alpha}, \text{ for } j = 1, 2, \dots, m \text{ and } k = 2, 3, \dots, n \quad (3.12)$$

The purest variable at the k -th iteration is the variable for which $p_j^{(k)}$ is the highest. Let's call the purest variable at the k -th iteration $P^{(k)}$. The weight vector is a vector with m elements containing the weights $w_j^{(k)}$ of each spectrum x_j (j -th row in the matrix D). Here is

how the weighting parameter $w_j^{(k)}$ is calculated: the matrix $Y_j^{(k)}$ is defined for each variable x_j , at the k -th iteration, as follows:

$$Y_j^{(k)} = [P^{(1)}, P^{(2)}, \dots, P^{(k-1)}, z_j] \quad (3.13)$$

where $z_j = x_j / \|x_j\|_2$. $\|x_j\|_2$ is the Euclidean norm of x_j : $\|x_j\|_2 = \sqrt{\sum_{i=1}^n x_{ji}^2}$, where n is the number of columns of D . In other words, $Y_j^{(k)}$ contains all the “pure spectra” selected in the previous $k-1$ iterations and the normalized spectrum of x_j . The covariance matrix of $Y_j^{(k)}$ is $Y_j^{(k)} Y_j^{(k)T}$. The weighting parameter $w_j^{(k)}$ is calculated as the determinant of the covariance matrix of $Y_j^{(k)}$.^{129,130}

3.4.2 Results

In this section, the results obtained by applying the spectral unmixing methods to our four sets of multispectral IHC images are demonstrated. The multispectral image stacks consist of 12 images captured at 12 wavelength channels spanning the range of 420 nm to 640 nm. For comparison, the spectral unmixing was also performed on three-channel RGB color images of the same field of view. Each image was converted to optical densities by taking the negative logarithm of the image intensities.

3.4.3 Ki67 Set

We start by illustrating the results of samples immunostained for Ki67. The Ki67 antigen is visualized using DAB and the sample is counterstained with Hematoxylin (H). Figure 3.2 shows the RGB image of a typical sample. The sample, like all the subsequent samples, was imaged under multispectral illumination generated by the OneLight Spectra light engine. Images were taken using 12 channels with the center wavelengths of 420 nm through 640 nm with a step size of 20 nm. The “full width at half maximum”, i.e. the width of the illumination spectrum between two wavelengths at which the intensity is half of the peak intensity, of each illumination channel was 20 nm. Besides the images captured at the 12 multispectral channels, images from the sample were taken at three wideband channels corresponding to red, green, and blue channels of three-color RGB imaging.

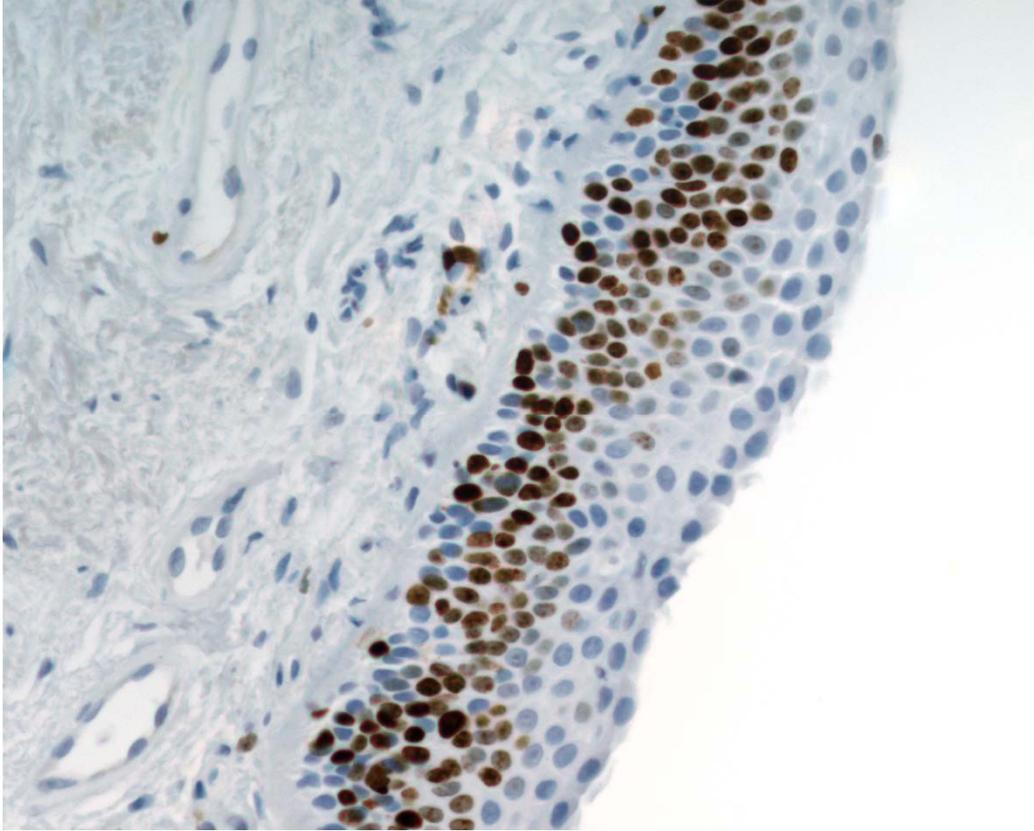
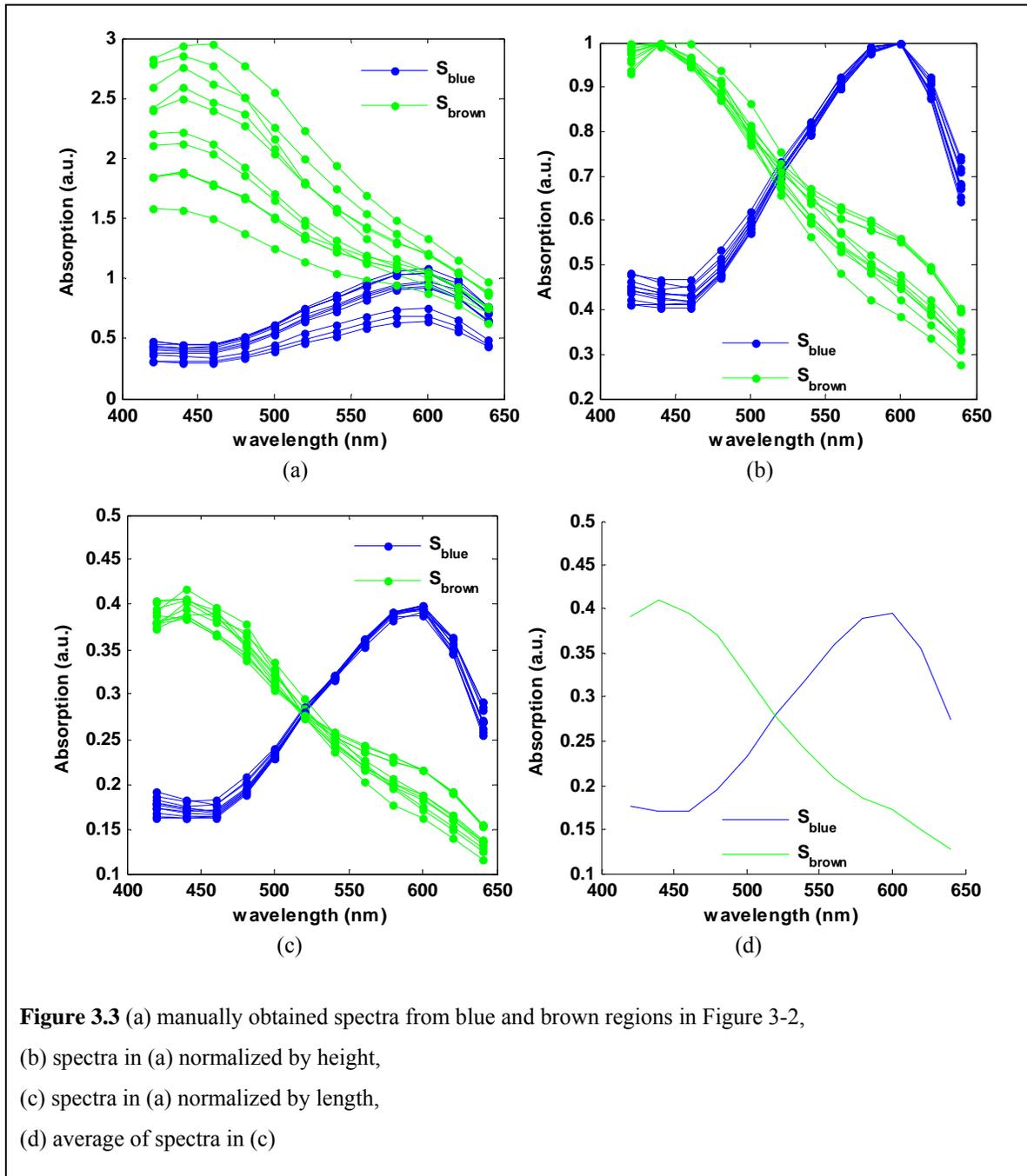


Figure 3.2 RGB Image of a typical Ki67 sample

3.4.3.1.1 MCR-ALS with Manually Captured Spectra

The absorption spectrum for each pixel consists of its optical density at each image in the multispectral image stack. To obtain the absorption spectra of the blue and brown pixels in Figure 3.2, we selected small regions with blue and brown staining. To reduce the effect of noise, the spectra of all pixels inside each region were averaged. Figure 3.3 (a) shows the spectra of ten blue and ten brown regions, after the spectra of all pixels within each region were averaged. The spectra appear to have the same shape but different overall intensity. To maintain the shape of the spectra, but discard the differences due to the brightness of each pixel, the spectra should be normalized. If S_i is the spectrum of the i -th pixel, then $\check{S}_i = S_i / \max(S_i)$, and $\hat{S}_i = S_i / \|S_i\|_2 = S_i / \sqrt{\sum S_i^2}$ are the spectrum of the i -th pixel normalized by the “magnitude” of the spectrum and by the “Euclidean norm” of the spectrum, respectively. We observed that normalizing the spectra by the Euclidean norm generated

spectra that had lower variation in their shape. Thus we chose to normalize the spectra by the Euclidean norm throughout this work. Figures 3.3 (b) and (c) show the normalized spectra.



The spectrum captured from a blue region provides a pure absorption spectrum of H, since we know that this spectrum is caused by H only. The absorption in the brown regions, however, is the result of absorption of both H and DAB. Let S_H and S_{DAB} be the spectra of H and DAB, respectively, and S_{blue} and S_{brown} be the spectra from the blue and brown regions in the image, then we have:

$$\begin{aligned} S_{blue} &= S_H \\ S_{brown} &= S_{DAB} + p \cdot S_H \end{aligned} \quad (3.14)$$

p is an unknown fraction of H absorption spectrum that is added to S_{DAB} to generate S_{brown} .

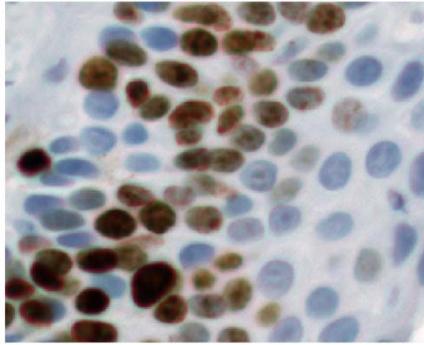
Next we apply the MCR-ALS algorithm to estimate the concentration maps of H and DAB, and their estimated spectra. Figures 3.4 (b) and (c) show the concentration maps, and Figure 3.5 (a) shows the estimated spectra. The terms \hat{S}_H and \hat{S}_{DAB} are used for the S_H and S_{DAB} estimated by the MCR-ALS algorithm. When S_{blue} and S_{brown} are used as initial estimates of spectra for the MCR-ALS algorithm, the concentration of H is incorrectly low in the areas where DAB staining is present (nuclei appear as black holes). This is because most of the absorbance at these areas is attributed to the brown label.

From a mathematical point of view, if \hat{S}_H and \hat{S}_{DAB} satisfy equation (3.5), any linear combinations of them would also satisfy this equation, but would result in incorrect concentrations. Thus, the ALS optimization procedure that attempts to minimize the error in equation (3.6) converges to a local minimum close to the initial spectra: S_{blue} and S_{brown} , instead of converging to the true S_H and S_{DAB} .

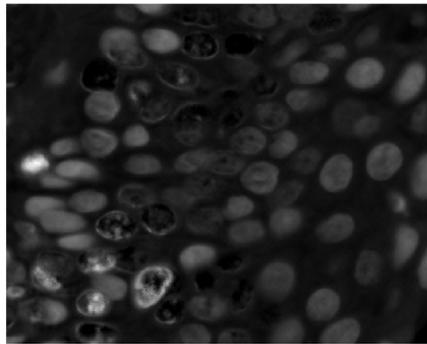
Our proposed solution for this problem is as follows: we subtract a fraction of S_{blue} from S_{brown} to find the initial estimate for S_{DAB} . This is equivalent to estimating p in equation (3.14). Figure 3.6 shows the normalized spectra resulting from subtracting increasing fractions of S_{blue} from S_{brown} . These normalized spectra are calculated as follows:

$$S_k = \frac{S_{brown} - k \cdot S_{blue}}{\sqrt{\sum (S_{brown} - k \cdot S_{blue})^2}}, \text{ for } k = 0:0.01:1 \quad (3.15)$$

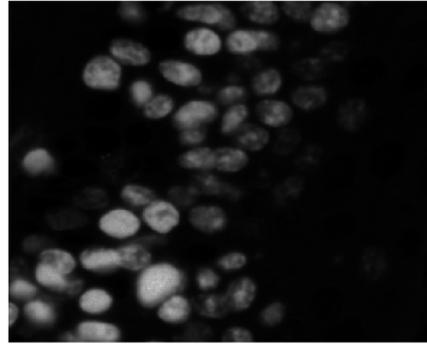
We note how the shape of the resulting spectrum changes as higher fractions of S_{blue} are subtracted from S_{brown} . For higher k , the right end of the spectrum starts to go up, generating a valley at ~ 610 nm. When k is further increased, the right end becomes negative.



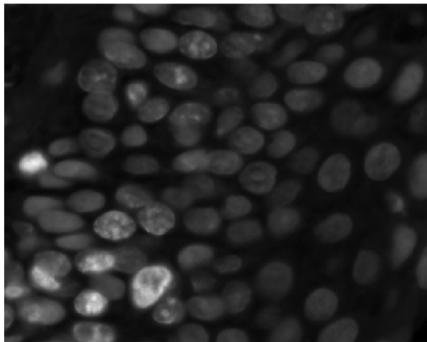
(a)



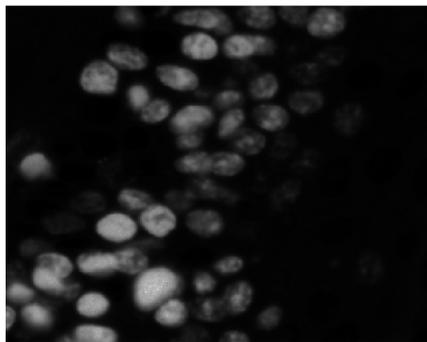
(b)



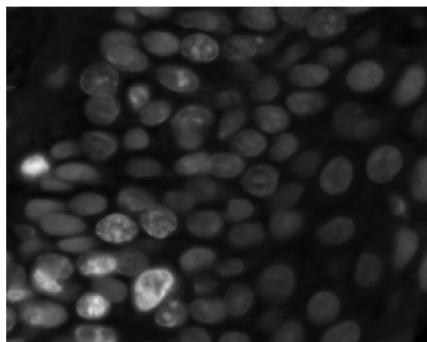
(c)



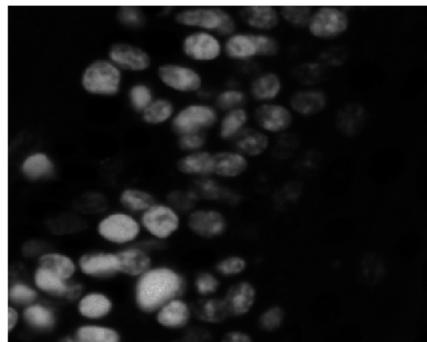
(d)



(e)

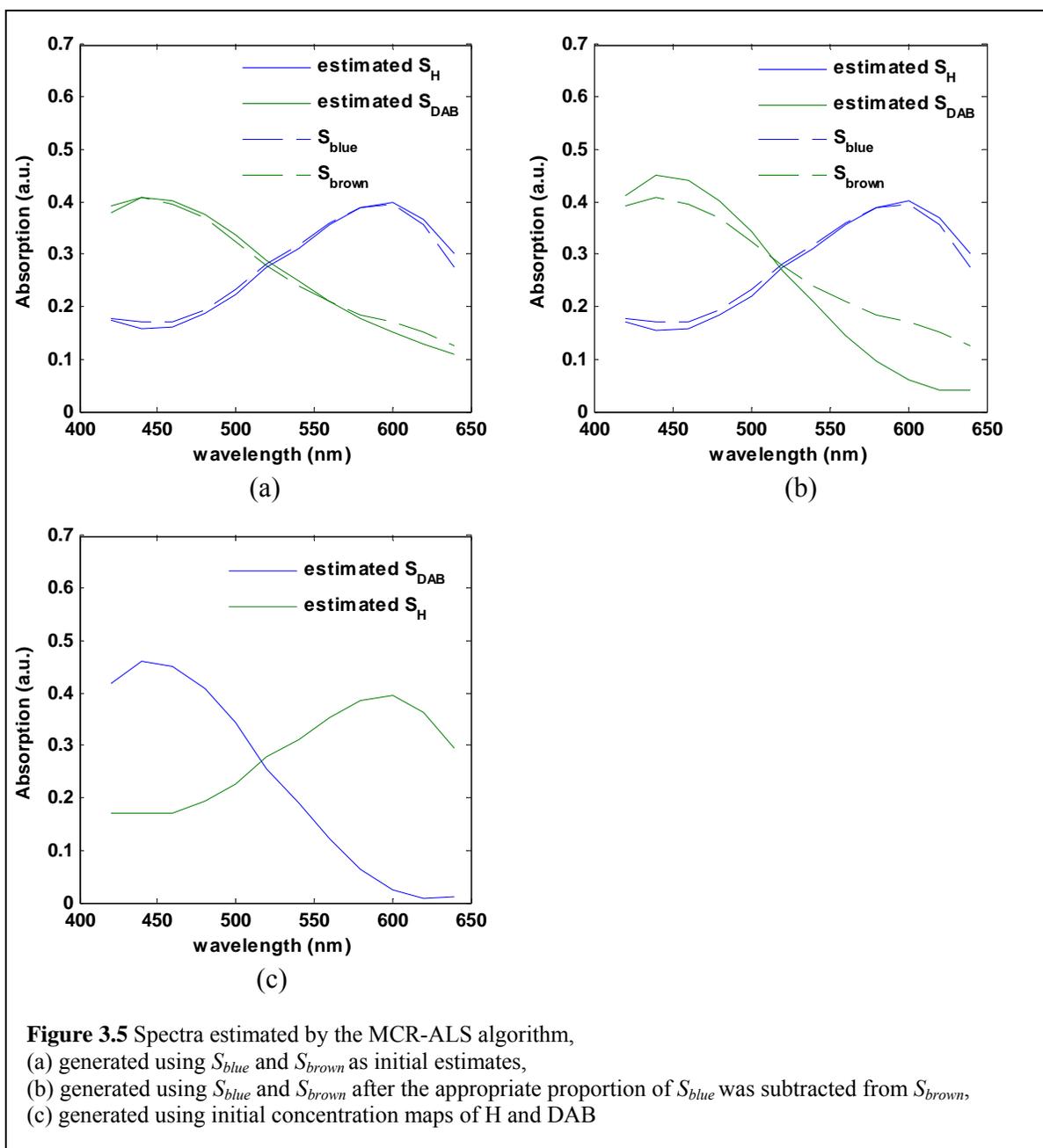


(f)



(g)

Figure 3.4 Concentration maps of the stains obtained by the spectral unmixing algorithm, the left column shows concentration maps of H, and the right column shows concentration maps of DAB, (b) and (c) generated using S_{blue} and S_{brown} as initial estimates, (d) and (e) generated using S_{blue} and S_{brown} after the appropriate proportion of S_{blue} was subtracted from S_{brown} , (f) and (g) generated using initial concentration maps of H and DAB



Without prior knowledge of the absorption spectrum of DAB in this sample, we could give the solution physical meaning by applying two constraints 1) non-negativity and 2) unimodality of the resulting absorption spectrum. We can calculate k_{max} , the highest value of k in equation (3.15) that results in a unimodal and positive spectrum. If k_{max} generates $S_{k_{max}}$ in equation (3.15), then $S_{k_{max}}$ can be used as the initial estimate for S_{DAB} . $S_{k_{max}}$ is

highlighted in Figure 3.6. We will refer to this procedure as the “subtraction algorithm” in the following sections.

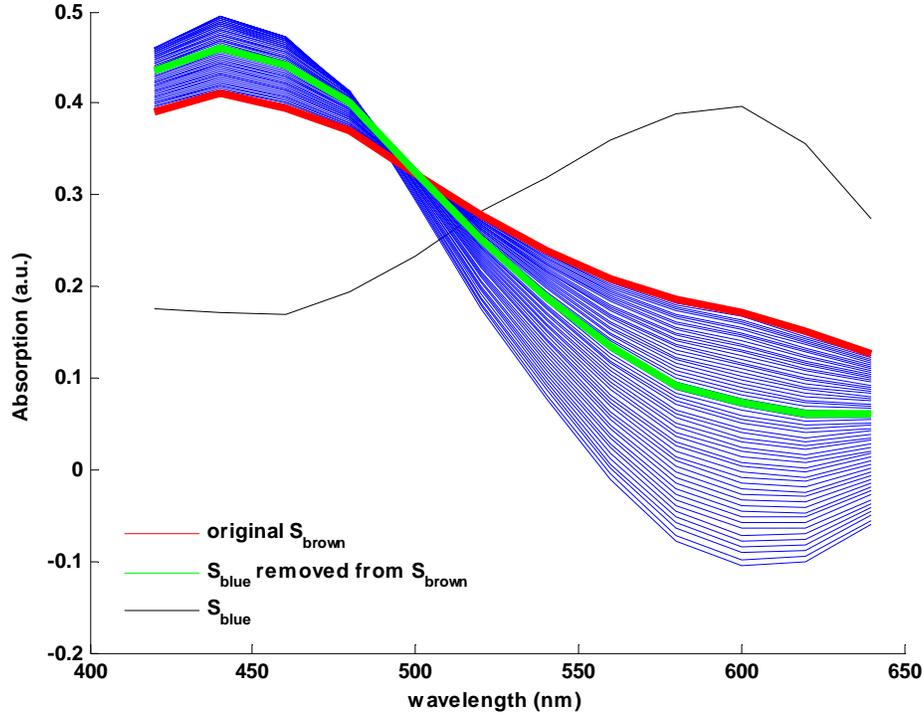


Figure 3.6 Resulting spectra after subtracting increasing proportions of S_{blue} from S_{brown} , arrows show the directions of change at both ends of the spectrum, with increasing k in equation (3.15)

This is the best initial estimate of S_{DAB} without knowing the spectrum of DAB in the same sample. The ALS optimization procedure will refine this estimate until positive concentrations for both H and DAB channels are obtained. Figures 3.4 (d) and (e) show the unmixed concentration maps. Figure 3.5 (b) shows the estimated spectra. Note that in the estimated concentration map of H, black holes in the position of nuclei disappear, and labels are accurately separated.

An additional constraint that may be applied is to force \hat{S}_H to be equal to the input S_{blue} . The equality constraint on \hat{S}_H does not greatly change the output of the unmixing spectra. It does however result in faster convergence of the iterative optimization procedure.

3.4.3.1.2 Obtaining the Initial Estimates of Components Spectra

So far, we have demonstrated how our proposed subtraction algorithm when performed on the spectra captured from the blue and brown regions provides initial estimates of the pure spectra. The initial estimates are then further refined by the MCR-ALS procedure. To minimize the user intervention in this procedure, it is desirable to automate the capture of spectra from the image stack.

We apply the SIMPLISMA algorithm to find initial estimates of the spectra. The algorithm works by finding the pixels that have the purest spectra, based on image statistics: mean and standard deviation. If the algorithm is applied to the whole image stack, the resulting spectra will be based on two pixels only (the two pixels with the purest brown and blue spectra), and will be easily impacted by the noise in the image. Instead, we apply the algorithm to a number of smaller regions chosen randomly across the image. Then we average the resulting spectra after normalizing them. This provides a more robust estimate of the spectra of pure pixels. This procedure will be referred to as the “multi-region SIMPLISMA” procedure.

Our automated algorithm randomly selects a total of 100 regions resulting in one hundred candidate spectra for the blue label and one hundred candidate spectra for the brown label. For each label, the trimmed average of the spectra is calculated by excluding the lowest and highest 10% of the candidate spectra values at each wavelength. Before applying this algorithm, we applied a Gaussian filter to the image stack to reduce the noise. In addition, we need to ensure that the small randomly selected regions contain both labels (blue and brown). This is done by only accepting those randomly selected regions within the image that 1) contain more than a pre-determined number of pixels (e.g. 50) whose absorption in the blue wavelength region is higher than a threshold, representing brown pixels and 2) contain more than a pre-determined number of pixels (e.g. 50) whose absorption in the blue wavelength region is lower than a second threshold, representing blue pixels. The choice of the low and high thresholds was done heuristically on one image. A low threshold value of 0.3 and high threshold value of 0.8 were used. The same thresholds were used for all images in the Ki67 set.

The SIMPLISMA algorithm finds the purest brown and the purest blue pixels in each randomly selected region. In implementing this procedure, it is important to take into account

that the first and second pure pixels identified by the SIMPLISMA algorithm in each region do not necessarily correspond to those identified in another region, i.e. in one region the first and second pure pixels may correspond to brown and blue, respectively, while in another the first and second pixels may correspond to blue and brown. Therefore, the two spectra obtained from each region need to be properly assigned to the right group of blue or brown spectra. This can be done through an ad-hoc algorithm such as this one: after the two purest spectra (referred to as S_1 and S_2) are obtained from the first randomly selected region, S_1 and S_2 are assumed to represent the first and second groups of spectra, respectively. Let the two spectra obtained from a new randomly selected region be S_i and S_j . Each one of S_i and S_j is assigned to the group that it has the highest correlation with its representing spectrum (S_1 or S_2). The representing spectrum for each group is then updated to be the average of the newly added spectrum and the old representing spectrum. Alternatively, once two spectra are obtained from each one of 100 randomly selected regions, a k-means clustering algorithm¹³¹ can be applied to the set of 200 obtained spectra to partition the spectra into two groups.

The spectra obtained automatically in this manner were very similar to those obtained manually from the image by selecting regions corresponding to different labels (Figure 3.7).

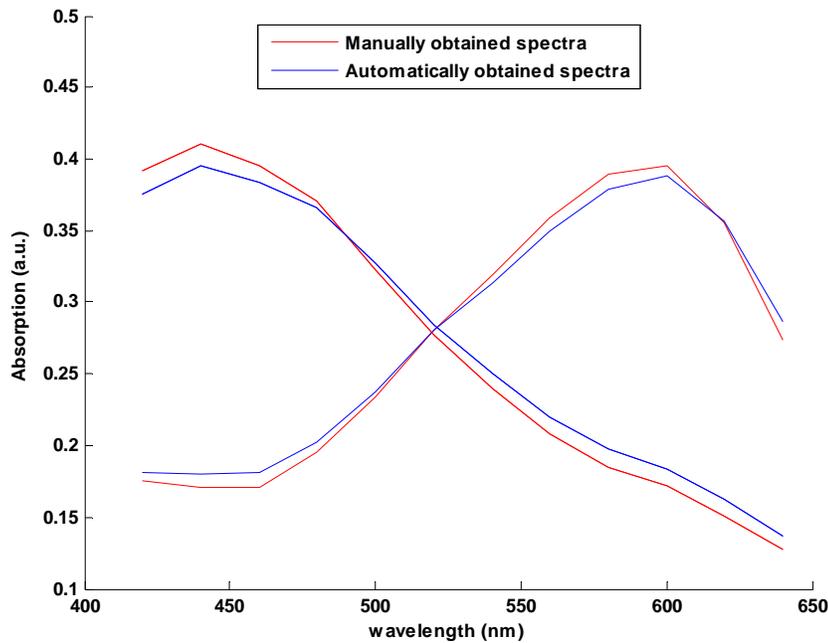


Figure 3.7 Absorption spectra of blue and brown pixels, captured manually, and through the SIMPLISMA algorithm.

3.4.3.1.3 Obtaining the Initial Estimates of Components Concentration Maps

The MCR-ALS optimization procedure can start using the initial estimates of either C or S^T . In the previous section (3.4.2.1.2), we demonstrated the results of using initial estimates of S^T . In this section we propose the use of initial estimates of the concentration maps C , which are denoted by \hat{C}^0 .

Looking at the absorption spectra of H and DAB, we know that the absorption of H is highest in the red wavelength range, and lowest in the blue wavelengths and the absorption of DAB is highest in the blue and lowest in the red wavelengths. Roughly speaking, the concentration map of H is most similar to the optical density of the image in the red wavelengths (where the absorption of H is highest). Similarly, the concentration map of DAB is most similar to the optical density of the image in the blue wavelengths. Thus the optical densities at these two wavelength regions (red and blue, e.g. 620 nm, and 440 nm) may be provided to the MCR-ALS algorithm as initial estimates of the concentration maps:

$$\hat{C}_H^0 = OD_{620nm} \text{ and } \hat{C}_{DAB}^0 = OD_{440nm} \quad (3.16)$$

Alternatively, the SIMPLISMA algorithm may be applied to the image stack to find the purest spectral bands instead of the purest pixels. In this case, the algorithm finds the first and second “purest” spectral bands, i.e. for every component, the algorithm finds the spectral band at which the absorptions of all other components are significantly lower than the absorption of that component. When SIMPLISMA is applied to find the purest spectral bands, the number of variables is significantly smaller than the case where the algorithm finds the purest pixels (12 wavelengths compared to 1024×1280 pixels), so the computational cost is lower. In addition, the multiple random region selection and the multiple applications of SIMPLISMA are not necessary. As expected, the SIMPLISMA analysis results in the selection of a blue and a red wavelength channel as the purest wavelength channels. Also, unlike using the initial estimates of spectra, the proposed subtraction algorithm is not necessary.

We observed, however, that using the optical density of the channels selected by SIMPLISMA did not result in accurate separation of the labels (the estimated distribution maps of the labels were not correct). Therefore, we tried a slightly different approach: we

used thresholded versions of the optical density in the blue (440 nm) and the red channel (620 nm), as follows:

$$\begin{aligned}\hat{C}_H^0 &= \begin{cases} OD_{620nm} & \text{if } OD_{620nm} > t_1 \\ 0 & \text{otherwise} \end{cases} \\ \hat{C}_{DAB}^0 &= \begin{cases} OD_{440nm} & \text{if } OD_{440nm} > t_2 \\ 0 & \text{otherwise} \end{cases}\end{aligned}\quad (3.17)$$

where OD_{620nm} and OD_{440nm} are the optical density representations (absorption maps) of the images captured at 620 nm and 440 nm spectral bands, and t_1 and t_2 are two threshold levels that can be determined interactively. This is equivalent to setting the background equal to zero. By setting the background equal to zero, the initial estimates of the concentration maps will be closer to the true concentration maps of DAB and H. Using the two thresholded absorption maps as the initial estimates results in proper separation of the labels. Figures 3.4 (f) and (g) show the unmixed concentration maps. Figure 3.5 (c) shows the estimated spectra. The choice of t_1 and t_2 in calculating \hat{C}_H^0 and \hat{C}_{DAB}^0 is not very critical, since these are only initial estimates which will be refined through the optimization process. We heuristically determined t_1 and t_2 so the thresholded absorption maps (\hat{C}_H^0 and \hat{C}_{DAB}^0) are non-zero in the locations of the corresponding labeled structures and zero elsewhere. The same values of t_1 and t_2 were used for the entire samples in the Ki67 set.

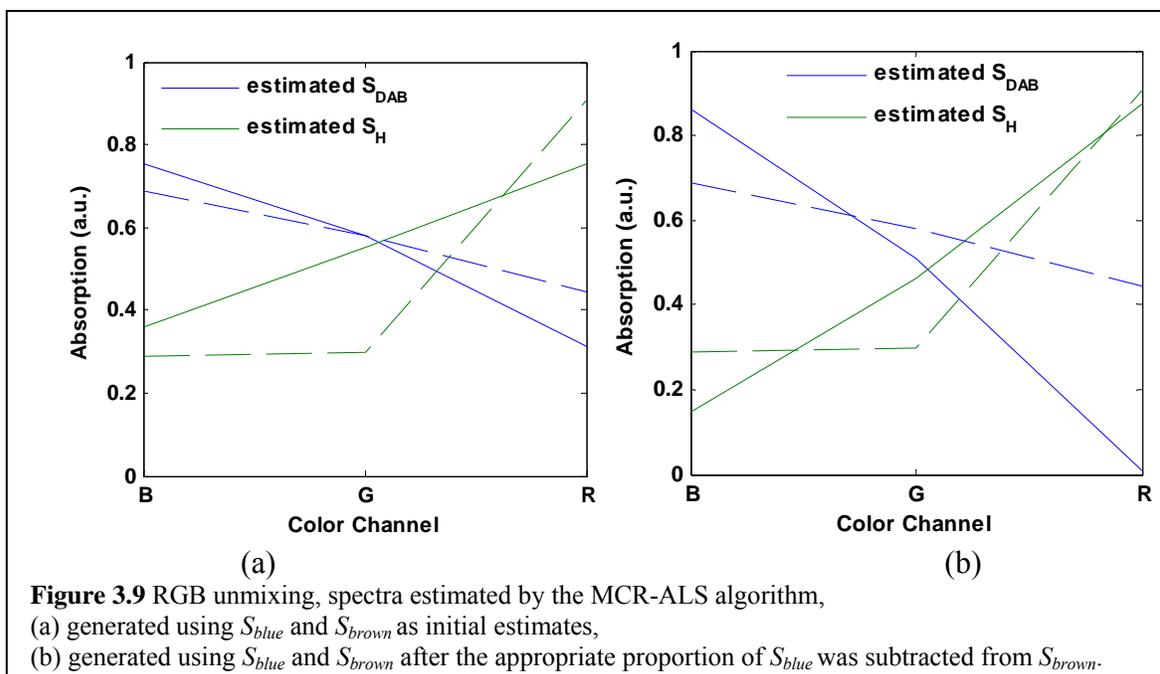
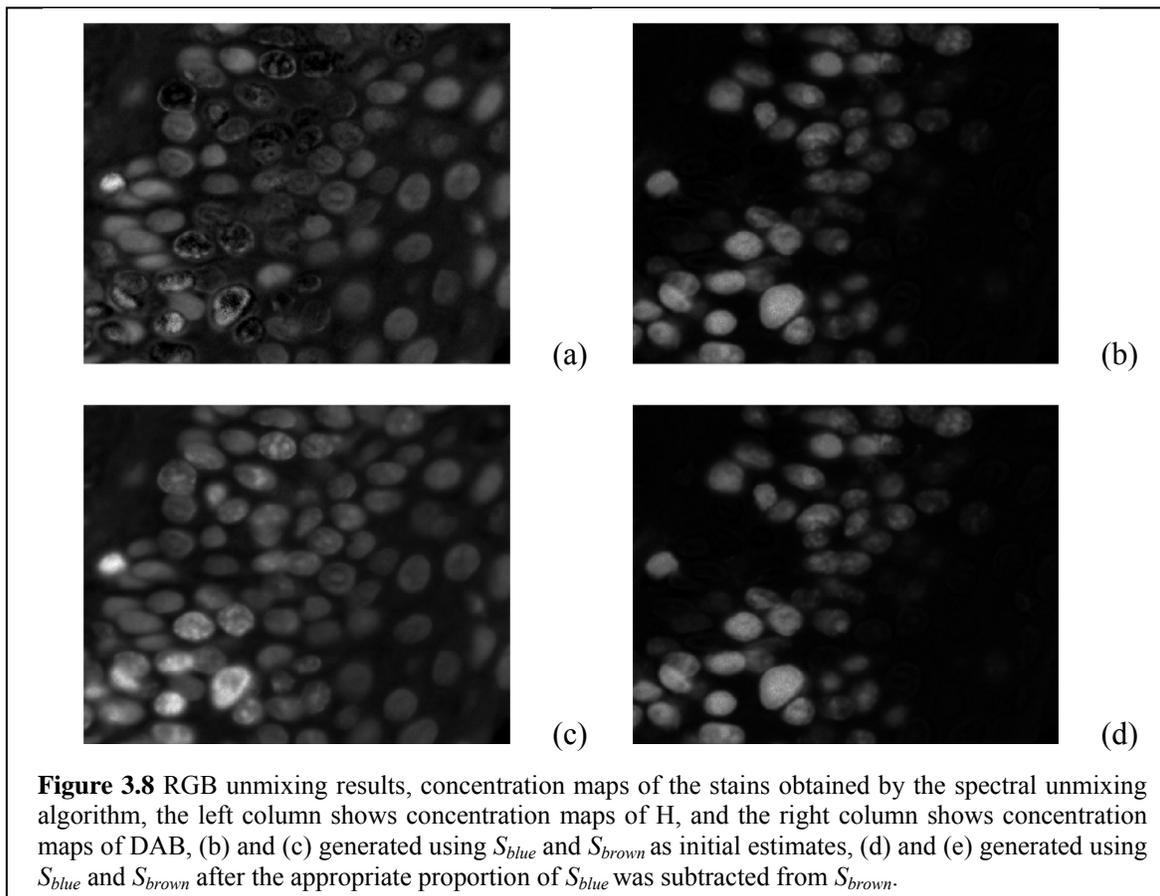
3.4.3.1.4 Linear Spectral Unmixing Using RGB Images

For each sample, in addition to performing the linear spectral unmixing on the multispectral stack of images, we also performed the linear spectral unmixing procedure on the three-channel RGB image of the same field of view. We will refer to the former as “multispectral unmixing” and to the latter as “RGB unmixing”. Similar to “multispectral unmixing”, in the case where the initial estimates of spectra were used for unmixing, the nuclei with both H and DAB labeling appeared darker than they should in the image representing the concentration map of nuclei label (H). So, the “subtraction algorithm” was used to subtract a proportion of the H spectrum from the DAB spectrum.

Figures 3.8 and 3.9 illustrate the results from the application of spectral unmixing to the Ki67 sample in Figure 3.4.

The visual evaluation of the two concentration maps obtained from the multispectral images and the concentration maps obtained from the RGB images did not show any

significant differences. Both multispectral and RGB unmixing approaches generated accurate separation of the labels, after the “subtraction algorithm” was applied to the initial spectra.



3.4.3.2 P16 Set

Images of the samples immunostained for P16 contain three types of pixels: blue nuclei pixels, brown nuclei pixels, and brown cytoplasm pixels. Unlike Ki67 samples, where no pixel with pure DAB staining was present in the images, in the P16 samples the brown cytoplasm pixels contain only DAB staining.

Figure 3.10 shows an RGB image of a typical P16 sample. Figure 3.11 shows the spectra of the blue and brown nuclei and the brown cytoplasm regions.

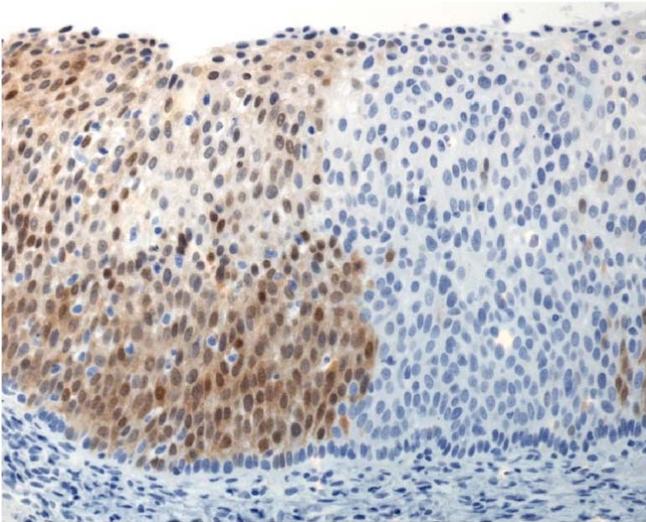


Figure 3.10 RGB image of a typical P16 sample

The automated pure pixel extraction algorithm of section 3.4.2.1.2, the “multi-region SIMPLISMA” procedure, was used to extract initial estimates of the spectra of the two components (H and DAB). Multi-region SIMPLISMA ensured that brown cytoplasm pixels were selected as the pixels with pure brown labeling, as opposed to brown nuclei pixels, which contained contributions from both DAB and H absorptions. Thus, the “subtraction algorithm” was not necessary. Figure 3.12 shows the spectra obtained manually and those obtained using the automated multi-region SIMPLISMA algorithm. Another way of performing the spectral unmixing was to use the initial estimates of concentration maps, obtained using the thresholded absorption maps corresponding to the two purest spectral bands (the same technique that was proposed in section 3.4.2.1.3). Figures 3.13 and 3.14 illustrate the concentration maps and unmixed spectra of the labels obtained through the above methods.

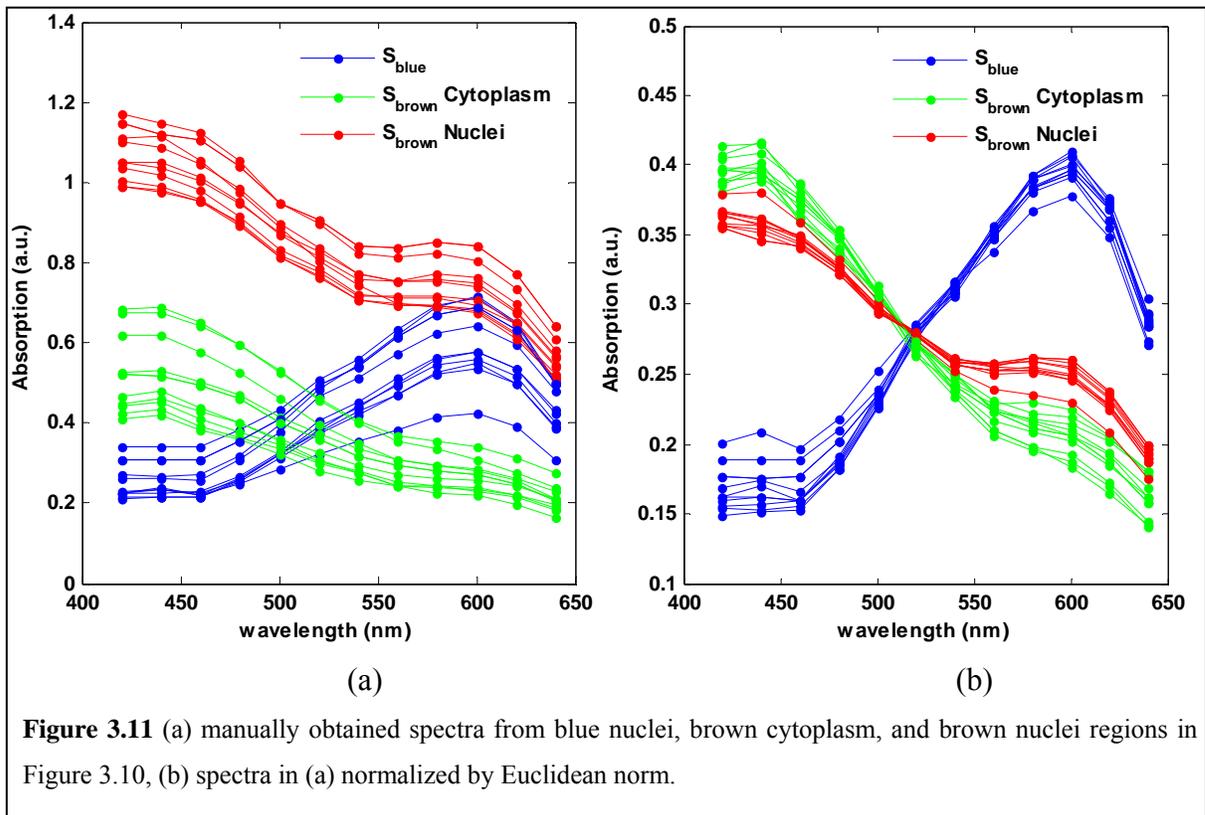


Figure 3.11 (a) manually obtained spectra from blue nuclei, brown cytoplasm, and brown nuclei regions in Figure 3.10, (b) spectra in (a) normalized by Euclidean norm.

The spectral unmixing procedure was also performed on the three channels of RGB images of the same field of view (“RGB unmixing”). Figure 3.15 shows the estimated RGB

spectra. Panels (e) and (f) of Figure 3.13 show the concentration maps of the two labels (H and DAB) obtained by “RGB unmixing”. No significant differences were observed between the unmixing results obtained through the “RGB unmixing” and “multispectral unmixing” (Figure 3.13).

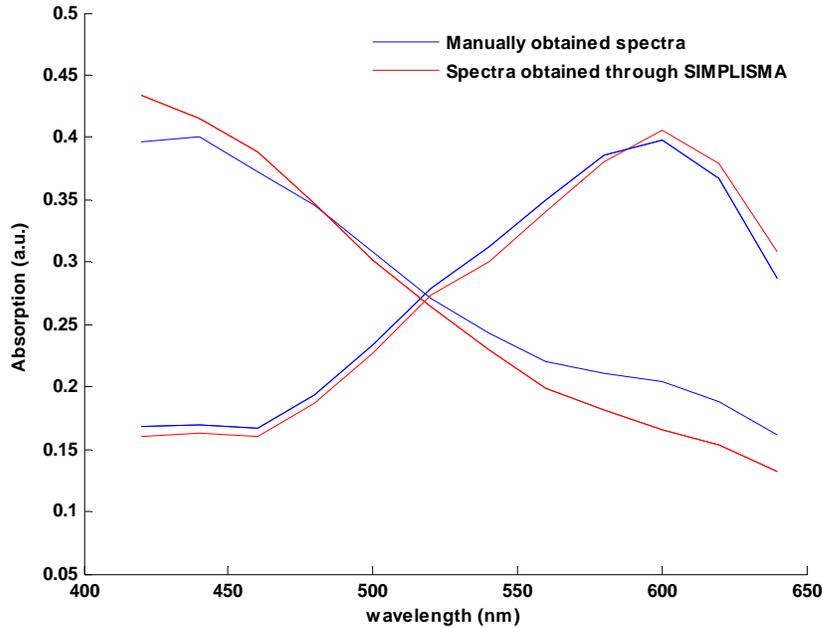


Figure 3.12 Absorption spectra of blue and brown pixels, captured manually, and through the SIMPLISMA algorithm.

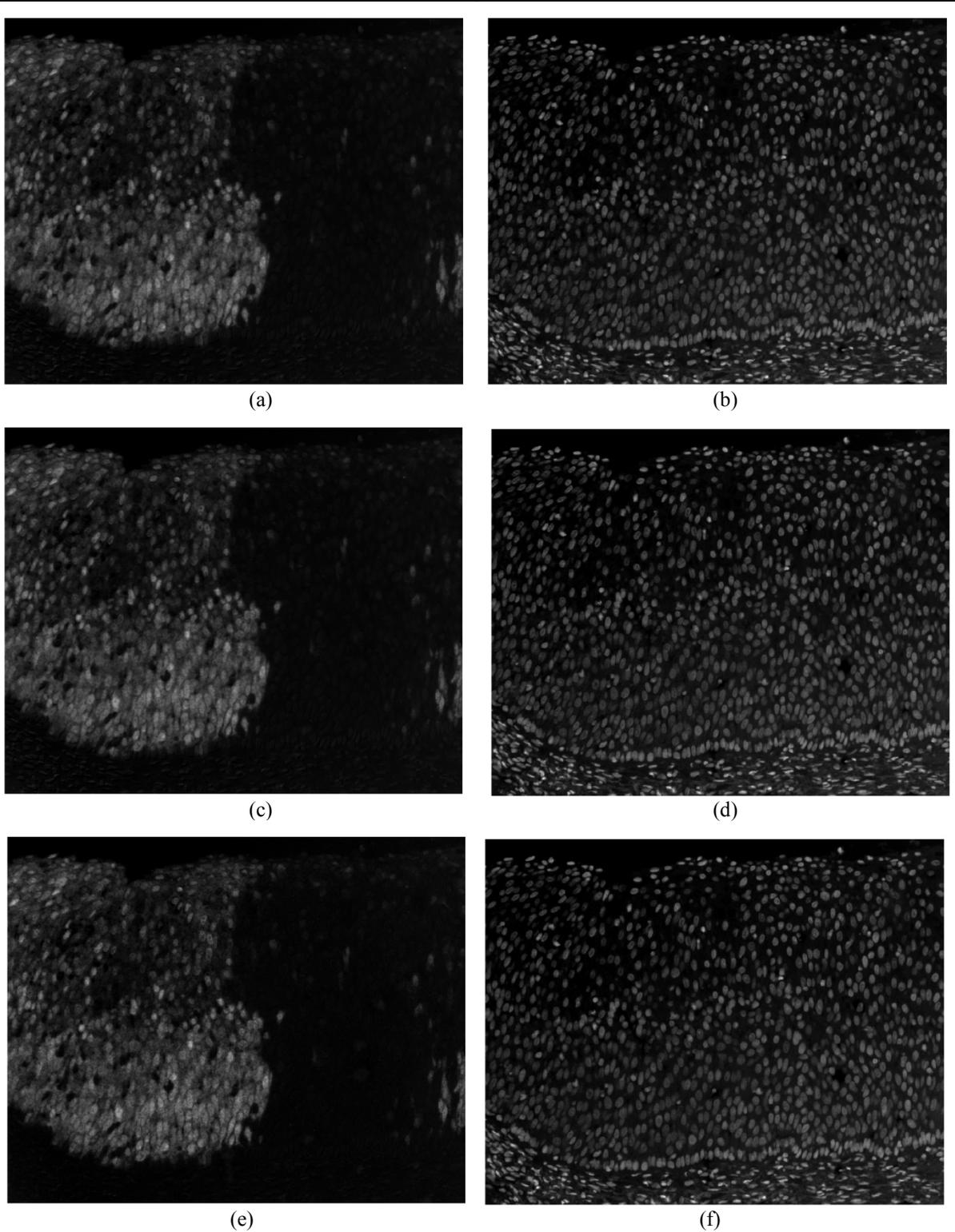


Figure 3.13 Concentration maps of the stains obtained by the spectral unmixing algorithm, the left column shows concentration maps of DAB, and the right column shows concentration maps of H, (a) and (b) generated using S_{blue} and S_{brown} as initial estimates, (c) and (d) generated using initial concentration maps of H and DAB, (e) and (f) generated using RGB spectral unmixing.

The methods explained above worked well on P16 samples with weak to moderate DAB staining. In samples with strong DAB staining, however, we noticed that these methods did not result in complete unmixing of the two labels. Figures 3.16 (b) and (c) show results of “multispectral unmixing” on a sample with dark DAB staining. Figures 3.17 (b) and (c) show concentration maps of the same sample obtained through “RGB unmixing”. Note that most of the cytoplasmic DAB staining is incorrectly assigned to the concentration map of H. Figure 3.18 shows the MCR-ALS estimated RGB and multispectral spectra for the same sample. The optimization took significantly more iterations to converge, compared to the samples with moderate DAB staining. As panel (a) in Figure 3.18 shows, the estimated spectra of H and DAB are different from the true H and DAB absorption spectra.

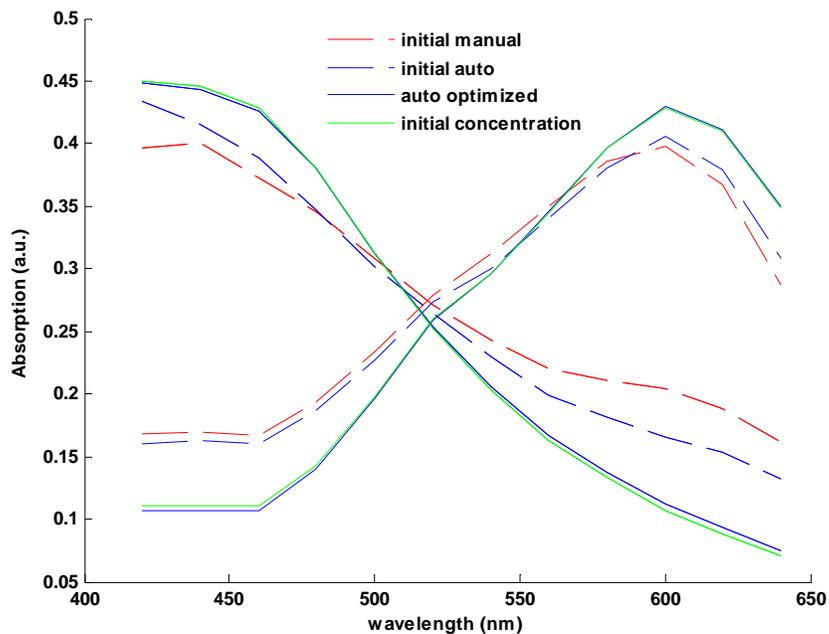


Figure 3.14 P16 spectral unmixing results, Initial spectra, and the spectra estimated by the MCR-ALS algorithm.

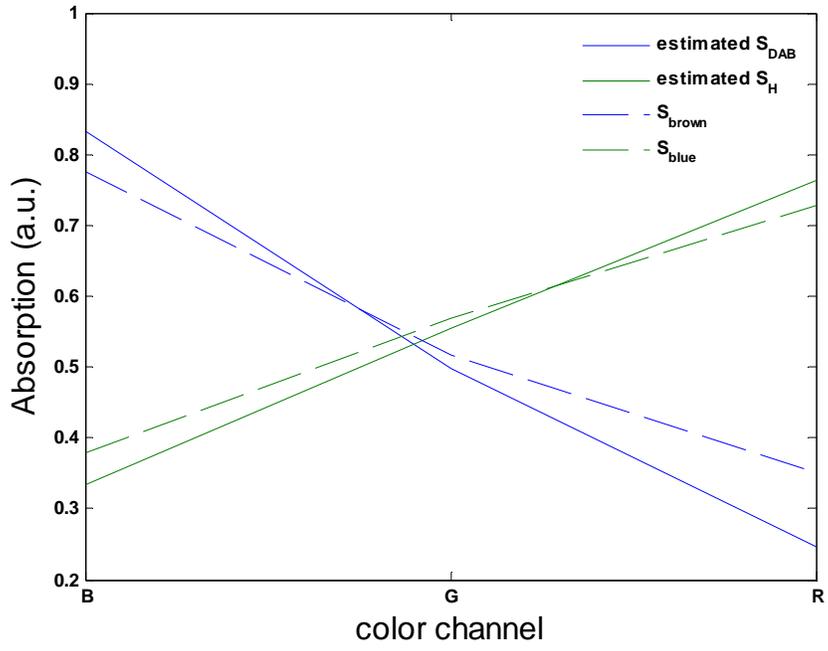


Figure 3.15 P16 RGB unmixing, initial and estimated spectra.

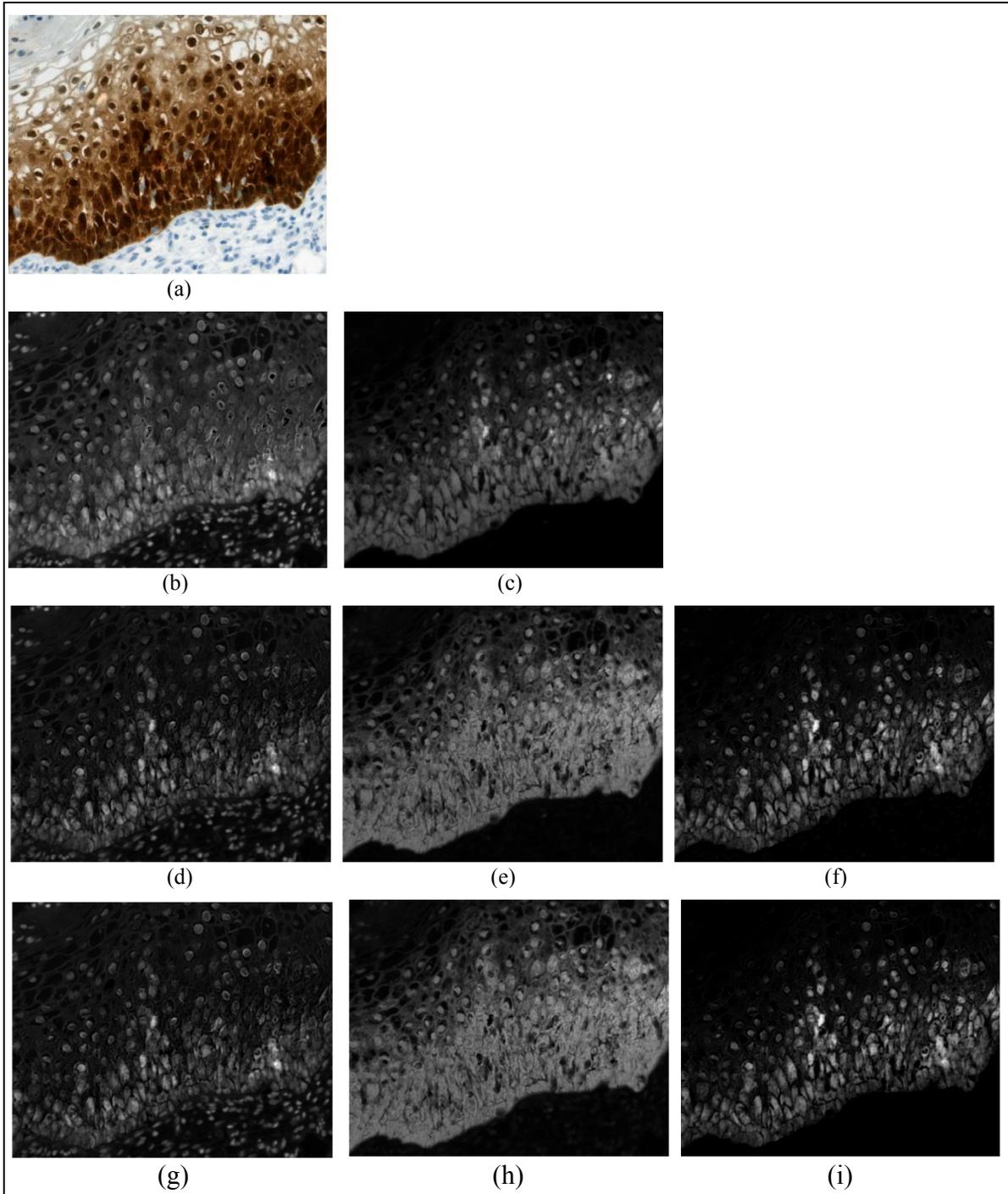
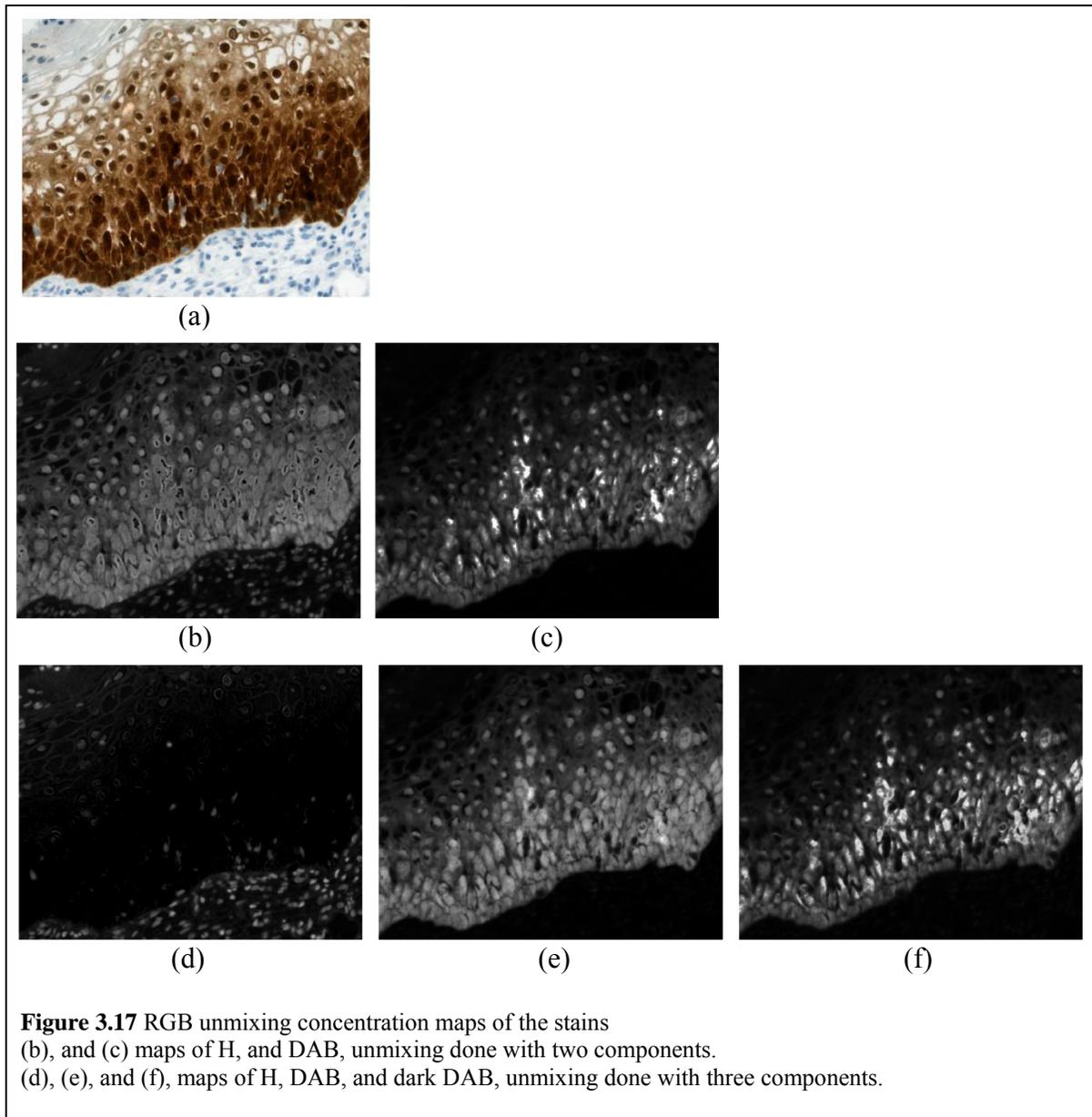


Figure 3.16 Multispectral unmixing, concentration maps of the stains (b), and (c) maps of H, and DAB, unmixing done with two components. (d), (e), and (f), maps of H, DAB, and dark DAB, unmixing done with three components, initial estimates of spectra. (g), (h), and (i), maps of H, DAB, and dark DAB, unmixing done with three components, initial estimates of concentration maps.



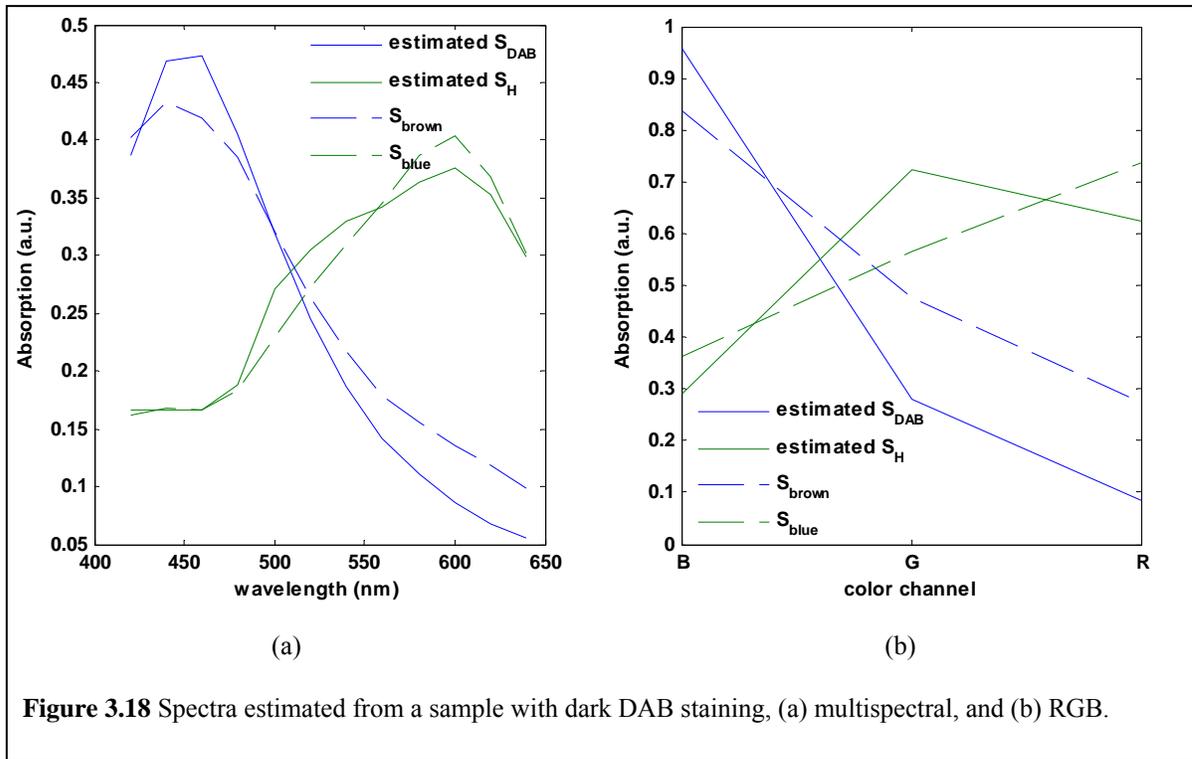


Figure 3.18 Spectra estimated from a sample with dark DAB staining, (a) multispectral, and (b) RGB.

We found that the reason was that the dark DAB label had an absorption spectrum that was different from that of weak or moderate DAB label. Figure 3.19 shows the spectra captured from four different regions: blue nuclei, brown cytoplasm, brown nuclei with moderate DAB labeling, and brown nuclei with dark DAB labeling. After normalizing the spectra, it was evident that the dark DAB spectra were different from moderate DAB spectra. As a result, the absorption in dark DAB pixels could not be unmixed into a linear combination of only two spectra. Thus, to minimize the least squares error, the optimization converged to an incorrect solution.

One solution is to consider dark DAB a third component and to perform the optimization based on three pure components: H, DAB, and dark DAB. We performed the multispectral unmixing procedure using the three spectra of Figure 3.19 (c) as initial estimates of spectra. In the resulting concentration maps (shown in Figure 3.16 panels (d), (e), and (f)) there was better contrast between the nuclei and cytoplasm in the resulting H image and the concentration map of the dark DAB label clearly shows the nuclei with strong DAB labeling.

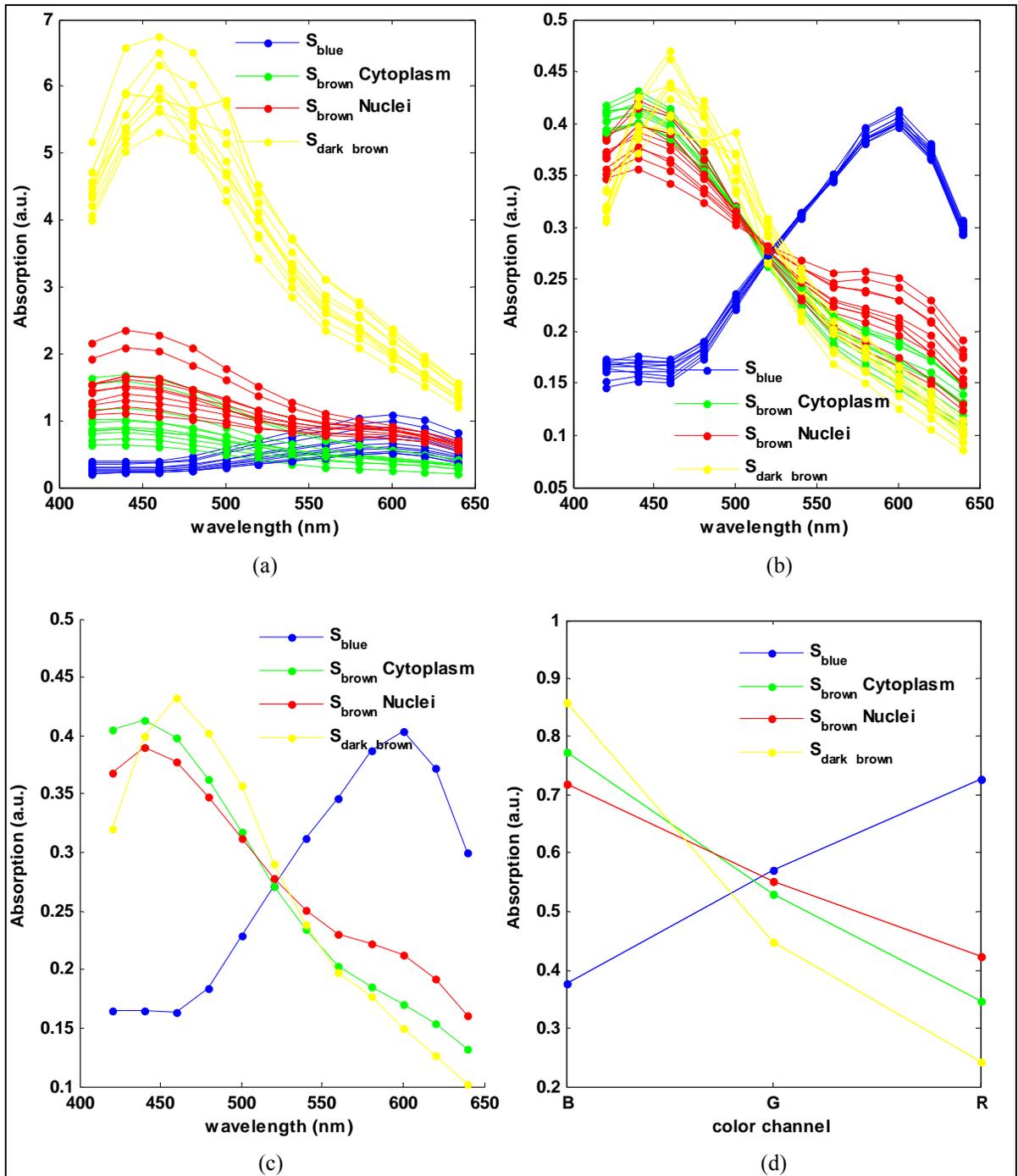


Figure 3.19 (a) spectra from blue nuclei, brown cytoplasm, brown nuclei, and dark brown nuclei regions in Figure 3-16 (a), (b) spectra in (a) normalized by length, (c) average of the spectra in (b), (d) RGB spectra.

The multispectral unmixing could also be initiated using three concentration maps, by setting thresholds on optical densities of blue and red channels:

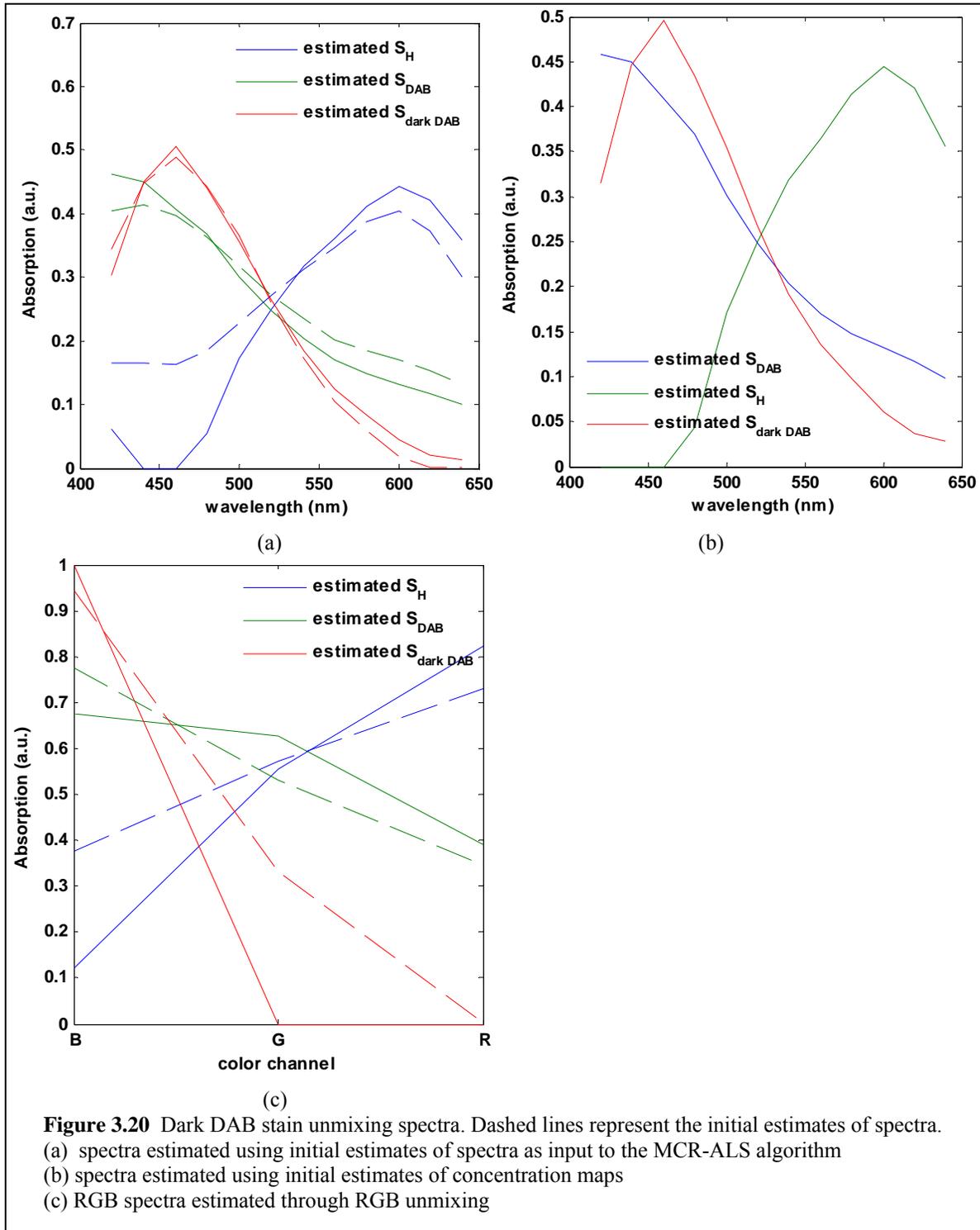
$$\begin{aligned}
\hat{C}_H^0 &= \begin{cases} OD_{red} & \text{if } OD_{red} > t_1 \\ 0 & \text{otherwise} \end{cases} \\
\hat{C}_{DAB}^0 &= \begin{cases} OD_{blue} & \text{if } OD_{blue} > t_2 \\ 0 & \text{otherwise} \end{cases} \\
\hat{C}_{Dark\ DAB}^0 &= \begin{cases} OD_{blue} & \text{if } OD_{blue} > t_3 \\ 0 & \text{otherwise} \end{cases}
\end{aligned} \tag{3.18}$$

where t_1 , t_2 , and t_3 were determined heuristically.

Figure 3.16 panels (g), (h), and (i) show the resulting concentration maps. These results were similar to those obtained using initial estimates of spectra, shown in panels (d), (e), and (f). Even though introducing the third component in the unmixing corrected the concentration maps, the estimated absorption spectrum of H remained different from its true spectrum. Panels (a) and (b) in Figure 3.20 show the estimated spectra. In Figure 3.20 (a), where the spectral unmixing was performed using the initial estimates of spectra, the estimated absorption of H is lower than its true absorption in the blue region of the spectrum, indicating that most of the absorption in this wavelength region was attributed to DAB. In Figure 3.20 (b), where the spectral unmixing was performed using the initial estimates of concentration maps, however, the estimated absorption of H is higher than its true absorption in the blue region of the spectrum.

For comparison, we also performed the ‘‘RGB unmixing’’ with three components: H, DAB, and dark DAB. Unlike multispectral unmixing results, introducing the third component did not result in separation of the DAB signal from the dark DAB signal. The high overlap between the RGB spectra of DAB and dark DAB labels did not allow the spectral unmixing algorithm to accurately unmix the signals arising from each one of them. Estimated concentration maps using RGB unmixing for a typical sample with dark DAB staining are shown in Figures 3.17 (d), (e), and (f). Note that almost all the absorption in regions with dark DAB staining was attributed to DAB, and dark DAB components. Since concentration of H in such areas was incorrectly estimated to be zero, the nuclei with dark DAB staining did not appear in the H image. Unmixed RGB spectra are shown in Figure

3.20 (c). Note the high overlap between the initial estimates of the RGB spectra of DAB and dark DAB.



3.4.3.3 CD8 Set

CD8 is expressed in the membrane of CD8 positive cells. So, no spatial overlap between the blue and brown label was expected. In practice, however, because of the finite (non-zero) thickness of a tissue section, absorption at each pixel is caused by all material in the vertical path of light. When light travels through a half cut cell, it may pass through both nucleus and membrane, so there is some spatial overlap between the two labels. There is also non-specific staining that can cause spatial overlap. Figure 3.21 shows an RGB image of a typical CD8 sample.

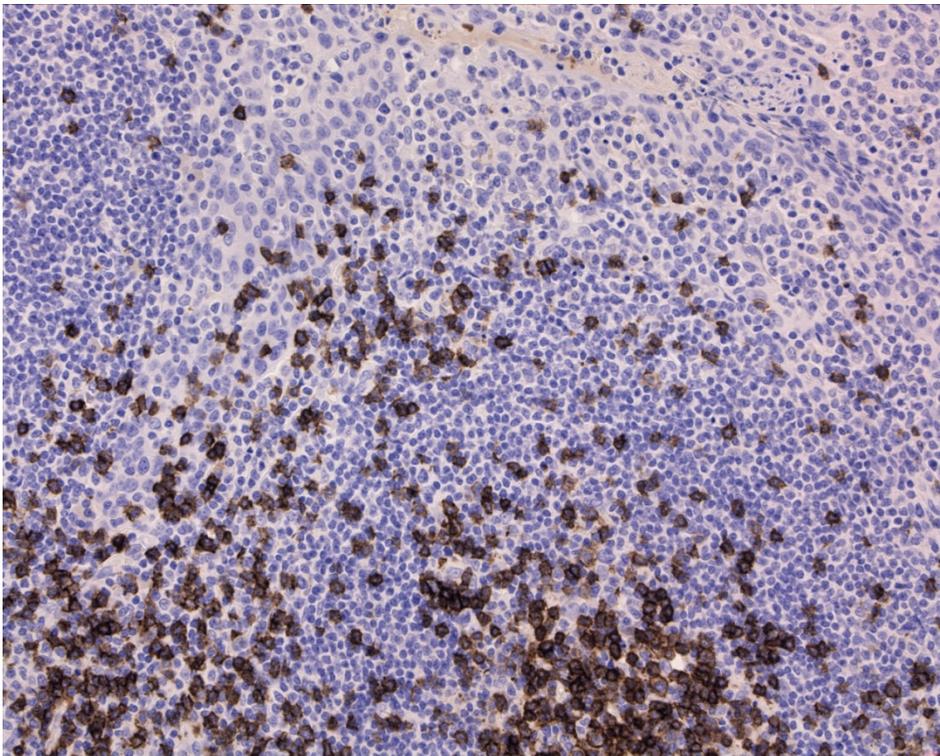


Figure 3.21 RGB Image of a typical CD8 sample

Figure 3.22 shows the manually captured spectra for blue nuclei and brown membrane pixels. The multi-region SIMPLISMA procedure (the automated pure pixel extraction algorithm of section 3.4.2.1.2) could also be used to extract pure spectra. These spectra are shown along with the manually captured spectra in Figure 3.22. In the case of CD8 samples, the manual capture of the spectra from regions with DAB staining was difficult. This was

due to the image being captured at 20X magnification. As a result, the membranes were only 2-3 pixels thick. Therefore it was difficult to select a region that contained only brown pixels and no blue pixels. As Figure 3.22 (c) shows, the automatically captured brown spectrum looks more similar to the true DAB spectrum compared to the brown spectrum captured manually (the true DAB spectrum can be seen in Figure 3.11 (b). It was obtained from a P16 sample where regions of brown cytoplasm with pure DAB staining were present). The brown spectrum that was obtained manually appears to be a mix of the H spectrum and the DAB spectrum, because from the results of the experiment shown in Figure 3.6, we know that adding increased fractions of the H spectrum to the DAB spectrum makes the right end of the spectrum go up and the left end of the spectrum go down, which is the case here. This demonstrates the advantage of the automated multi-region SIMPLISMA algorithm over the manual selection of pure pixels.

Similar to P16 and Ki67 samples, unmixing was also performed using the initial estimates of the concentration maps. The concentration maps resulting from both initializations are shown in Figure 3.23. Also shown in Figure 3.23 is the concentration maps obtained from the “RGB unmixing”. Figures 3.24 and 3.25 present the unmixed spectra. Again similar to both Ki67 and P16 cases, no significant differences were observed in the concentration profiles obtained by RGB unmixing versus multispectral unmixing.

All spectral unmixing methods discussed above (multispectral unmixing using initial estimates of spectra, multispectral unmixing using the initial estimates of the concentration maps, and RGB unmixing) were performed on all the samples in the CD8 set. All three methods resulted in accurate unmixing of the labels in samples with moderate DAB staining. Similar to the samples in the P16 set, in samples with dark DAB staining, complete unmixing of the labels was not possible with only two components (H and DAB), which caused incomplete removal of the DAB signal from the H channel. An example of this is demonstrated in Figure 3.26.

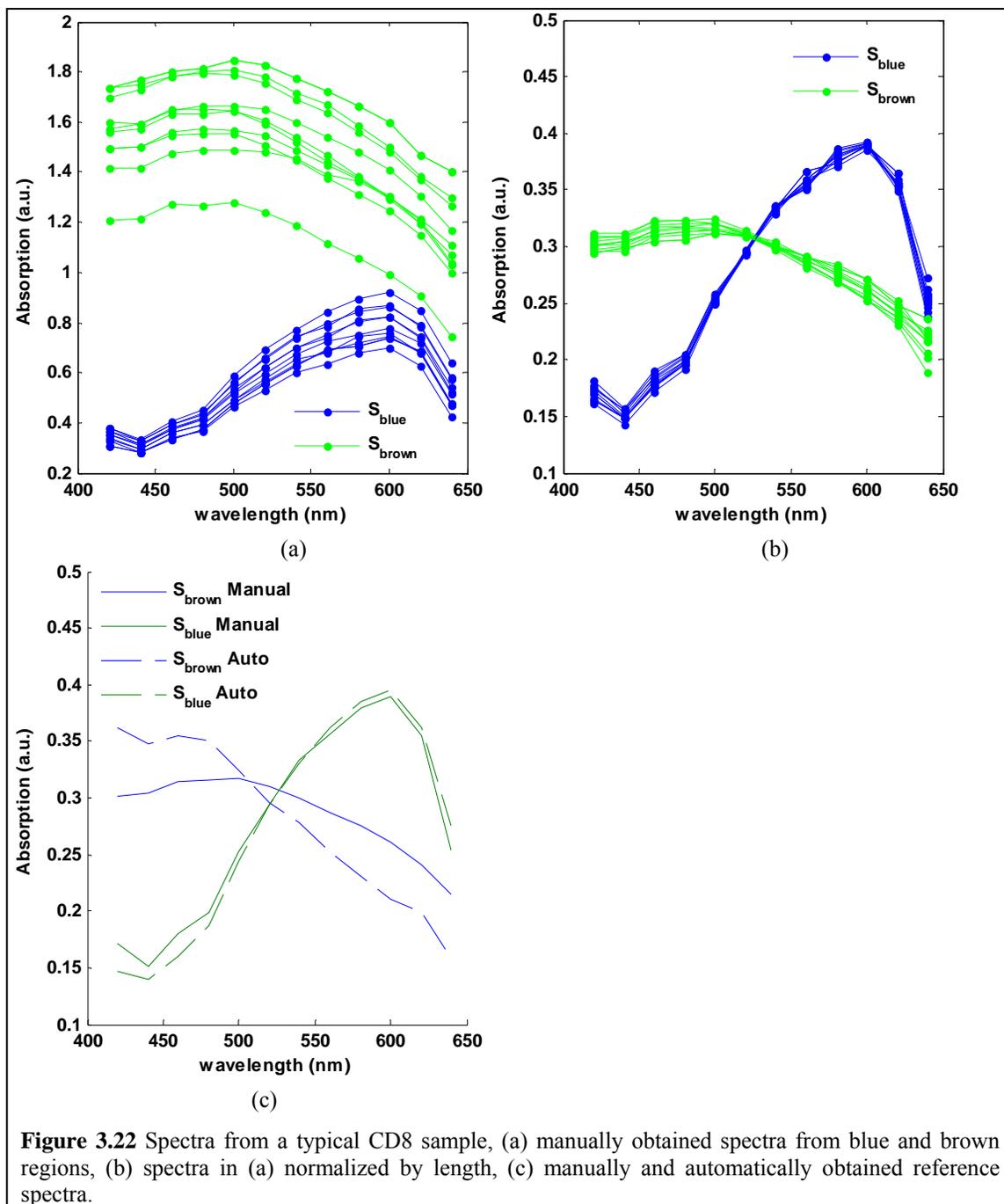
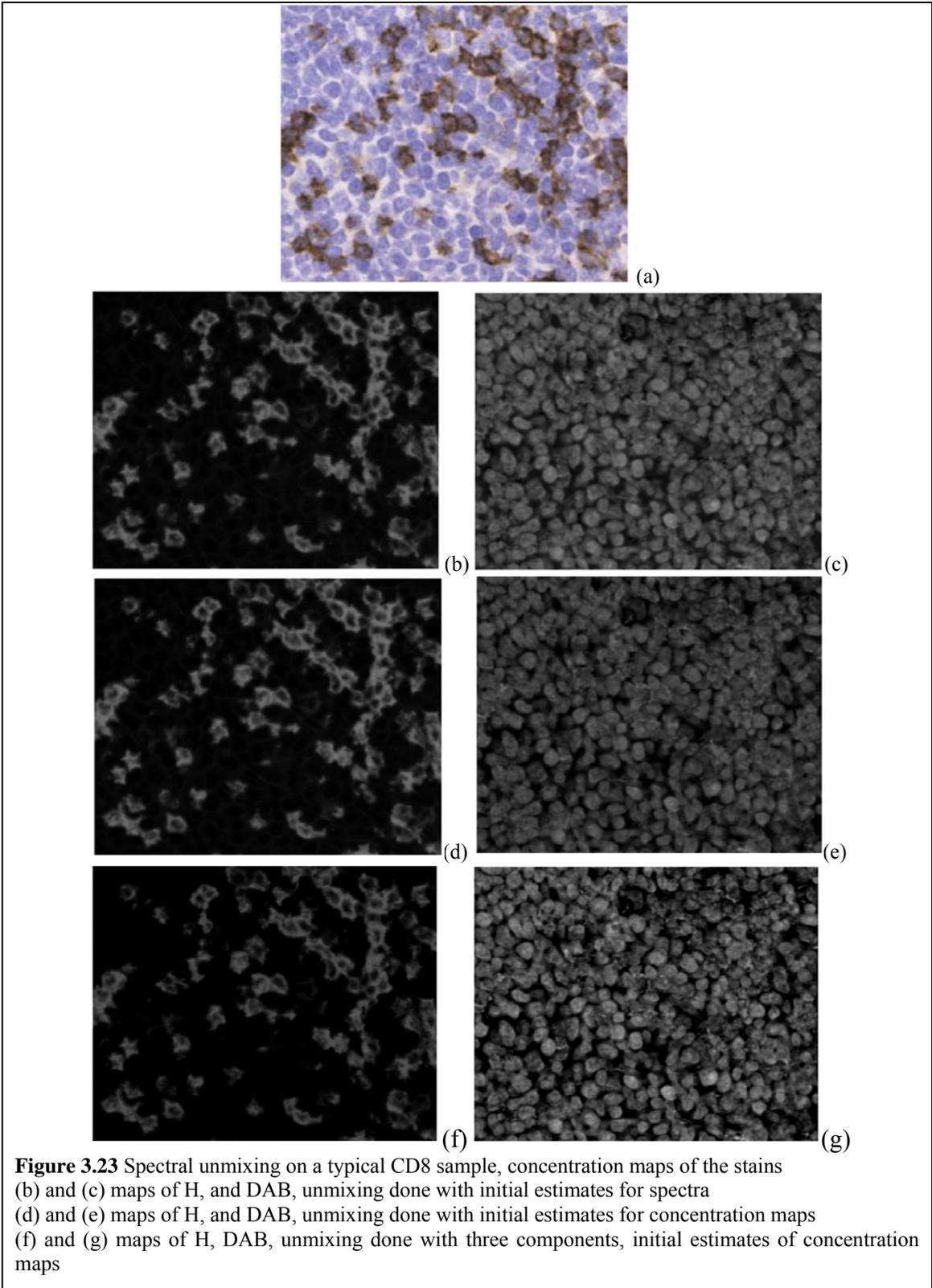
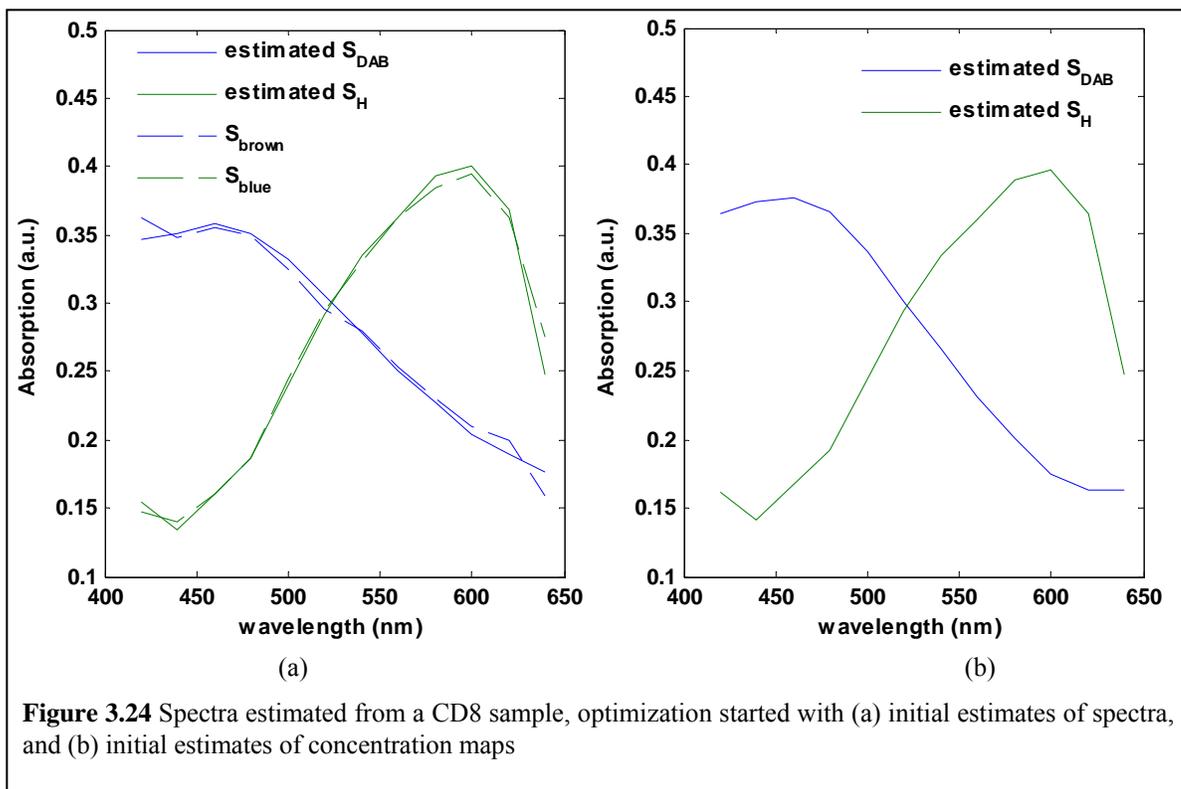


Figure 3.22 Spectra from a typical CD8 sample, (a) manually obtained spectra from blue and brown regions, (b) spectra in (a) normalized by length, (c) manually and automatically obtained reference spectra.





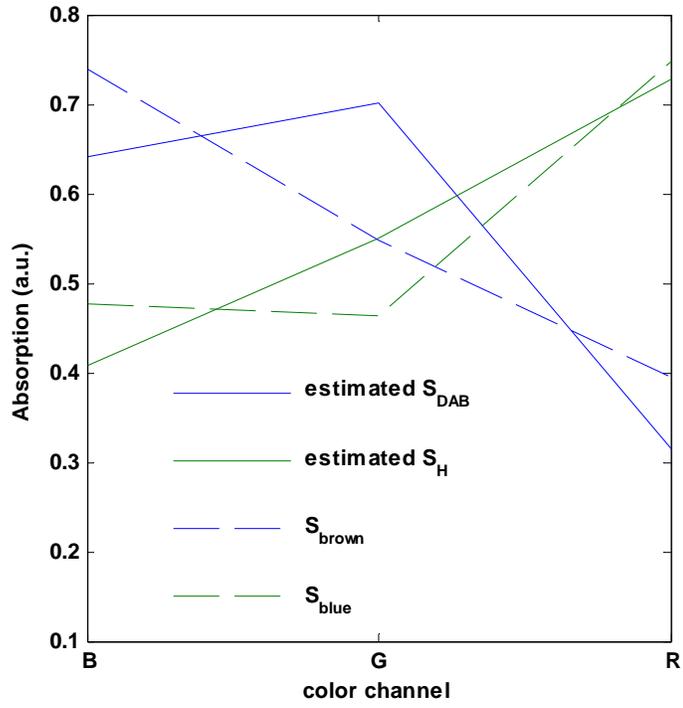


Figure 3.25 RGB Spectra estimated from a CD8 sample

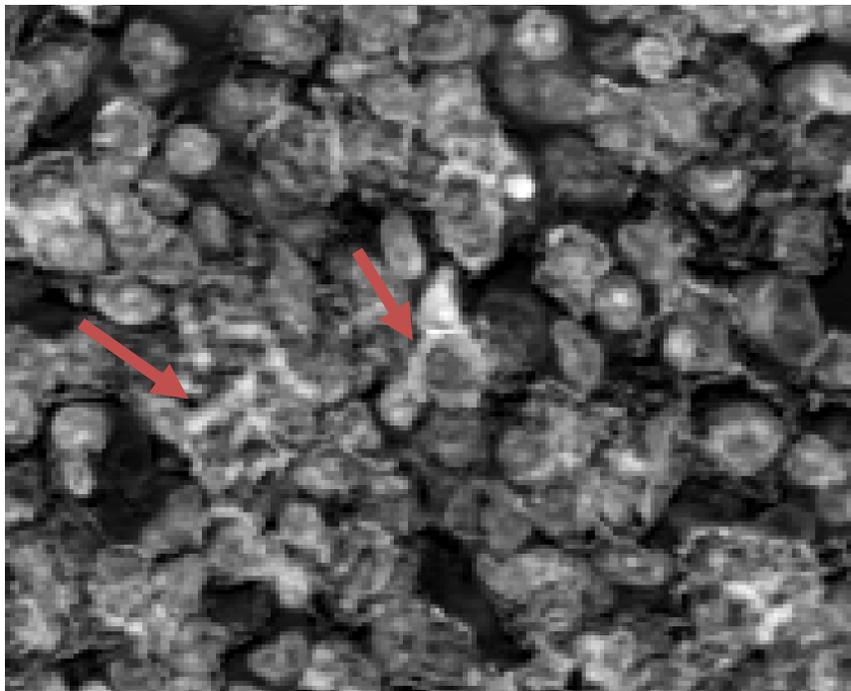
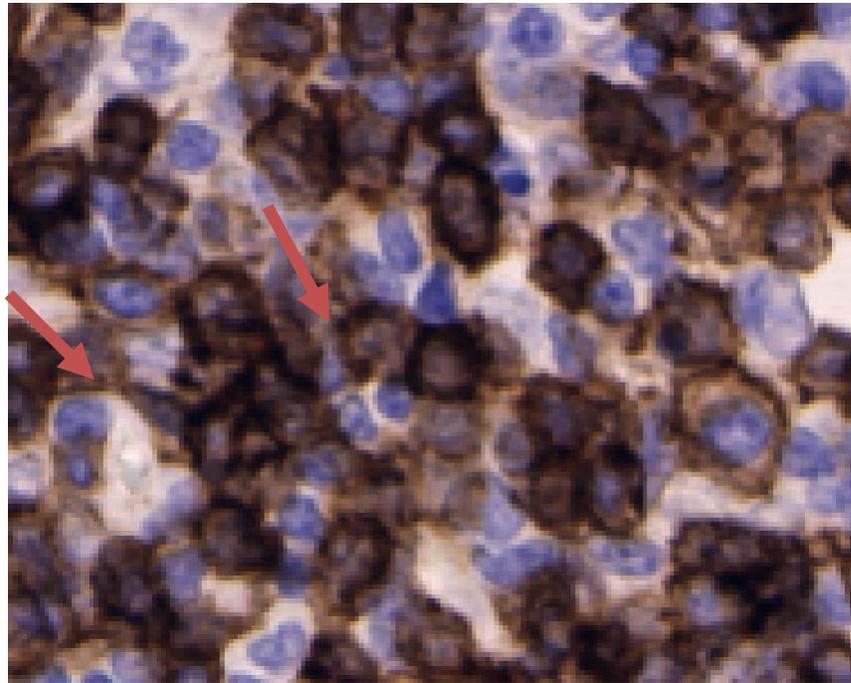


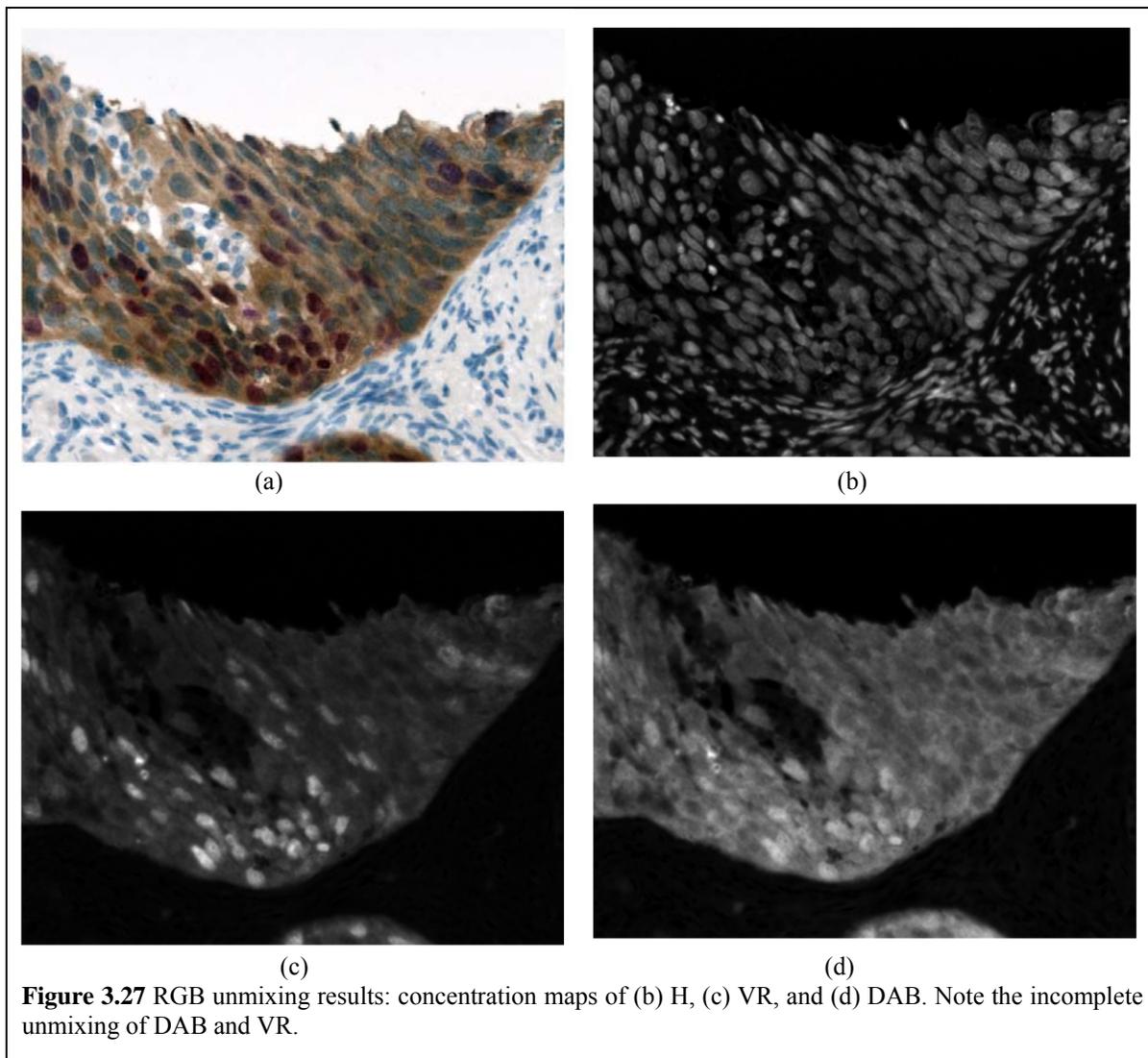
Figure 3.26 Example of CD8 sample with dark DAB, incomplete removal of DAB signal from the H channel, arrows point to traces of DAB absorption attributed to the H channel

3.4.3.4 Ki67 & P16 Set

The images analyzed thus far contained only two labels (H and DAB). The double immunostained samples contained three labels: H, DAB, and Vector Red (VR). Ki67 was labeled with H, P16 was labeled with DAB and nuclei were labeled with H. There was spatial overlap between H and DAB in the Ki67 expressing nuclei and there was spatial overlap between H, DAB and VR in the nuclei that expressed both Ki67 and P16.

For the double-immunostained samples in this set, linear spectral unmixing performed on the RGB images did not generate completely unmixed concentration maps of the labels. This was because of the high overlap between the RGB spectra of DAB, and VR.

Figure 3.27(a) shows the RGB image of a typical double-immunostained sample. Panels (b), (c), and (d) show the concentration maps of H, DAB, and VR obtained using RGB unmixing. Note that the DAB and VR labels are not separated properly. Figure 3.28 shows the initial RGB spectra obtained from the image, along with the spectra refined by the MCR-ALS algorithm.



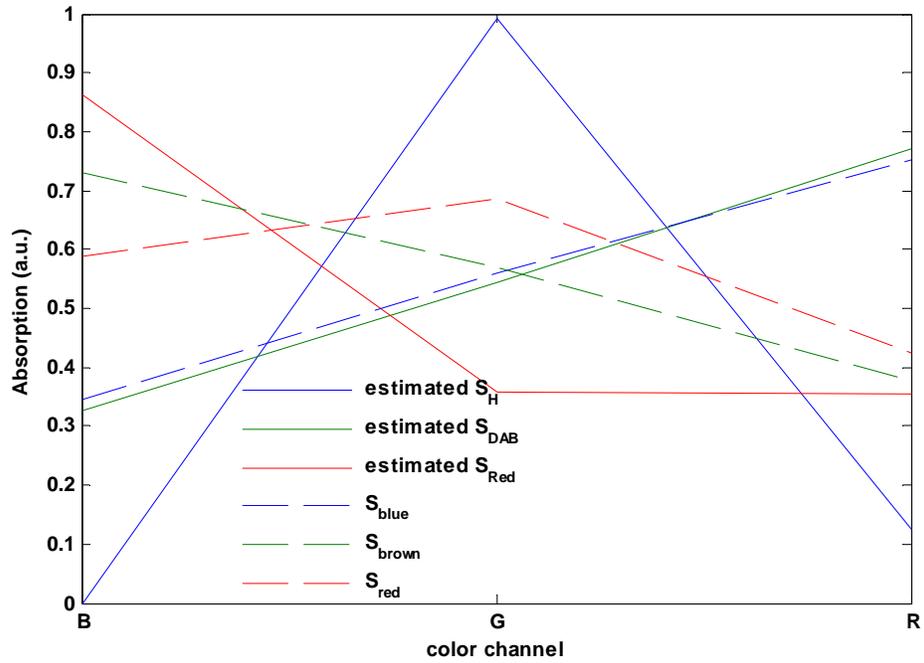


Figure 3.28 RGB unmixed spectra

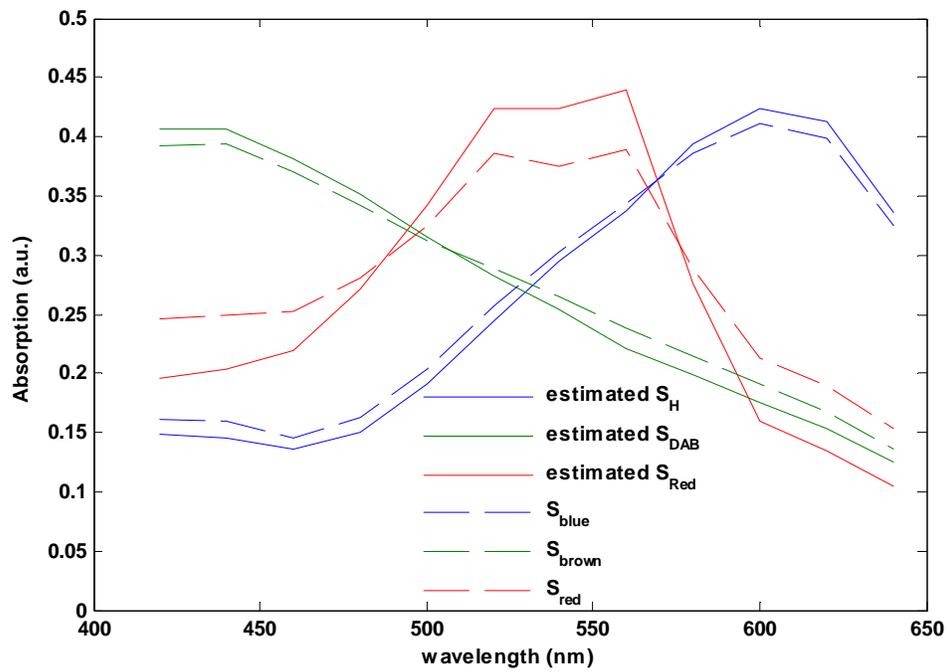
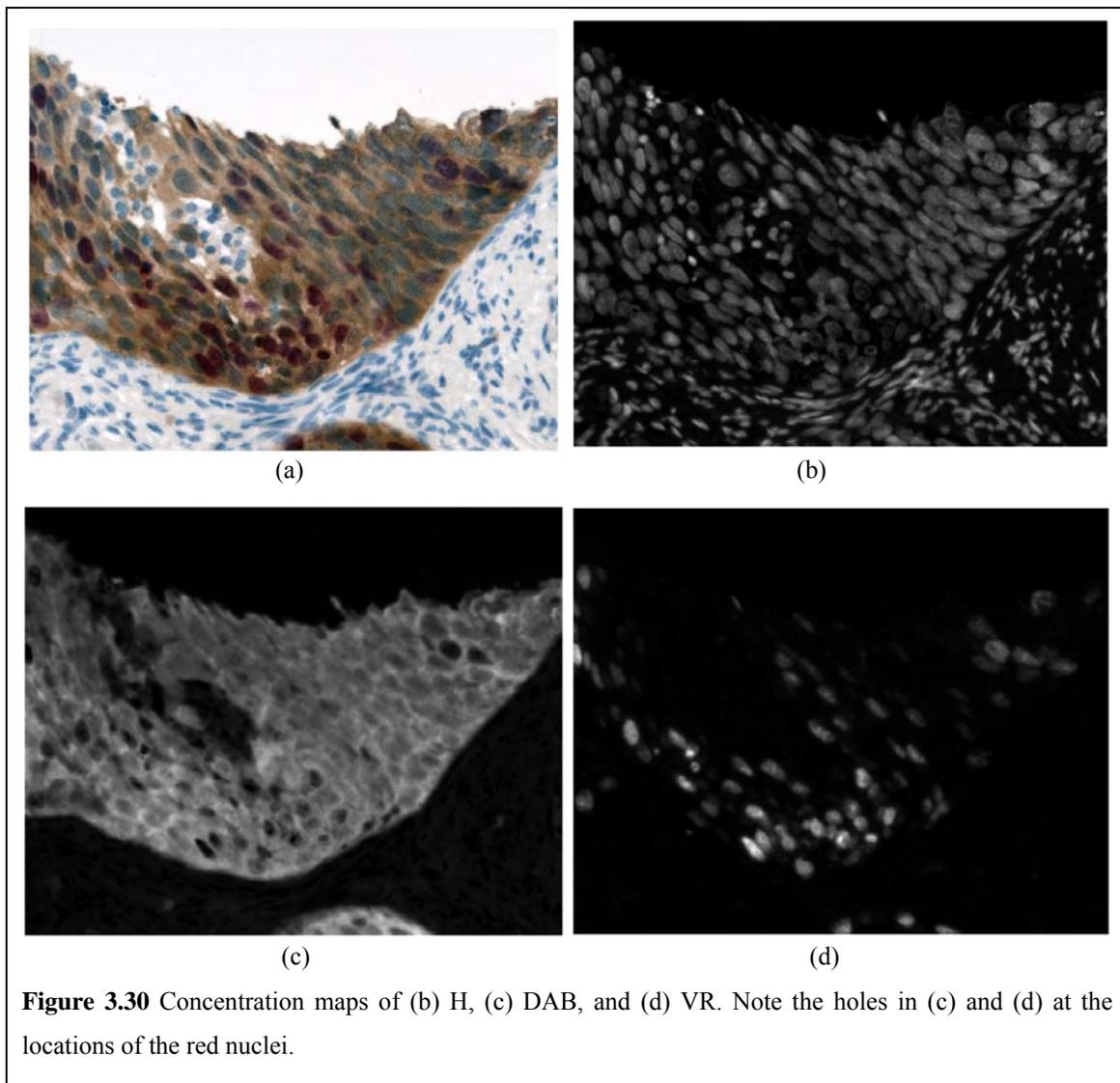
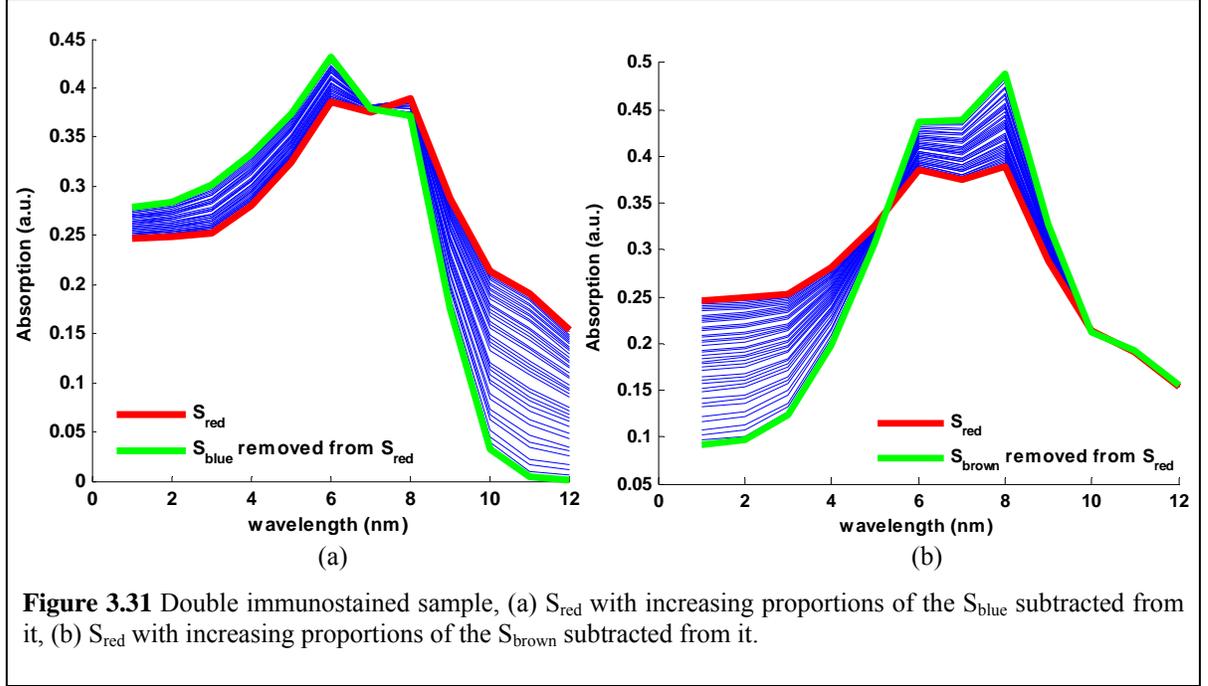


Figure 3.29 Double immunostained sample, unmixed spectra.

We now demonstrate the results of application of MCR-ALS to the multispectral images. The first set of unmixed maps was generated by using the initial estimates of spectra obtained directly from the image (shown in Figure 3.29). Let these spectra be S_{blue} , S_{brown} , and S_{red} . S_{blue} represents S_{H} . S_{brown} was obtained from brown cytoplasm regions where the absorption was only due to DAB, as opposed to brown nuclei, where absorption was from H and DAB. Thus S_{brown} represented S_{DAB} . Because of the spatial overlap of both H and DAB with VR, the spectrum of VR obtained from the sample was not a pure spectrum and contained contributions from both H and DAB. When these three spectra were used to unmix the labels, accurate unmixing was not achieved. Concentration maps are shown in Figure 3.30. Note the holes in the locations of ki67-expressing nuclei (represented with the reddish color in the RGB image) in the H and DAB channels.

Similar to the case of Ki67 samples, we needed to subtract appropriate fractions of S_{blue} , and S_{brown} from S_{red} , to obtain the initial estimate of the VR spectrum. Panels (a) and (b) in Figure 3.31 illustrate the resulting spectra after increasing proportions of S_{blue} and S_{brown} are subtracted from S_{red} . We note that subtracting too much of the S_{blue} from S_{red} resulted in the right end of the spectrum to become negative and subtracting too much of S_{brown} from S_{red} resulted in the left end of the spectrum to go down, and eventually become negative.





The initial estimate of the VR spectrum S_{VR}^0 can be obtained by subtracting the appropriate proportions p_1 and p_2 of the H and DAB spectra from the red spectrum:

$$S_{VR}^0(p_1, p_2) = \frac{S_{red} - p_1 S_{blue} - p_2 S_{brown}}{\|S_{red} - p_1 S_{blue} - p_2 S_{brown}\|} \quad (3.19)$$

To find p_1 and p_2 , in the absence of any prior information about the VR spectrum, the only assumption we can make about S_{VR}^0 is non-negativity, i.e. the proportions p_1 and p_2 should be chosen so that the resulting VR spectrum is not negative. However, the VR spectrum has been recorded from singly labeled tissue sections and is known from the literature.^{93,132} Based on the published spectrum the local minimum values of the absorption spectrum within the wavelength range of our experiment (420 nm to 640 nm) occur at 420 nm and 640nm, where absorption is approximately 0.1 and 0 respectively (when the spectrum is normalized by its Euclidean norm). We chose to use these two values at the two ends of the spectrum as *a priori* knowledge to estimate p_1 and p_2 . There may be slight differences between the published absorption spectrum of VR and the one that is used in our sample (caused by different illumination, or differences in the composition of the chromogen). However, the spectrum calculated in this manner is only provided as a starting

estimate in the MCR-ALS algorithm, and will be further refined to result in positive concentrations of the labels.

To find p_1 and p_2 , we use the values of S_{VR}^0 at 420 nm and 640nm: S_{VR420}^0 and S_{VR640}^0 . So the problem is to find maximum values for p_1 and p_2 in such a way that:

$$S_{VR420}^0 \geq 0.1 \text{ and } S_{VR640}^0 \geq 0 \quad (3.20)$$

This is equivalent to subtracting the highest fractions of H and DAB spectra from the VR spectrum, while the lowest points of the VR spectrum are greater than or equal the values in equation (3.20). We used a numerical approach to find p_1 and p_2 : let's define the following two surfaces S_1 and S_2 in the space of p_1 and p_2 as follows:

$$\begin{aligned} S_1(p_1, p_2) &= S_{VR420}^0(p_1, p_2) - 0.1 \\ S_2(p_1, p_2) &= S_{VR640}^0(p_1, p_2) \end{aligned} \quad (3.21)$$

p_1 and p_2 are the points at which $S_1(p_1, p_2) = S_2(p_1, p_2) = 0$.

$S_1(p_1, p_2) = S_2(p_1, p_2)$ is a one-dimensional curve that is the intersection of the two surfaces S_1 and S_2 . $S_1(p_1, p_2) = S_2(p_1, p_2) = 0$ is the point where the intersection curve is equal to zero. The coordinates of this point (\hat{p}_1, \hat{p}_2) are the right proportions to use in equation (3.19) to find S_{VR}^0 . We found (\hat{p}_1, \hat{p}_2) numerically, by choosing a two-dimensional uniform grid consisting of points (p_{1i}, p_{2i}) with spacing $h = 1/n$:

$$(p_{1i}, p_{2i}), 0 \leq i \leq n, \text{ such that } p_{1i} = ih \text{ and } p_{2i} = ih \quad (3.22)$$

$S_1(p_1, p_2)$ and $S_2(p_1, p_2)$ were calculated on this grid. (\hat{p}_1, \hat{p}_2) would be the point in the grid where: $S_1(p_1, p_2) \cong 0$ and $S_2(p_1, p_2) \cong 0$. The value of n determines the precision of the solution. We used $n=20$. The $S_{VR}^0(\hat{p}_1, \hat{p}_2)$ obtained this way is the initial estimate of the VR spectrum. This solution is the two-dimensional equivalent of the ‘‘subtraction algorithm’’ that was previously used for the samples in the Ki67 set.

This initial estimate of the VR spectrum in the sample of Figure 3.30 (a) is illustrated in Figure 3.32. Figure 3.33 demonstrates the two surfaces S_1 and S_2 and their intersection. Figures 3.34 and 3.35 demonstrate the unmixing results using the calculated S_{VR}^0 as the initial

estimate of the VR spectrum along with S_{blue} and S_{brown} as initial estimates of H and DAB. Note in the regions of spatial overlap between VR and H and DAB, the labels are correctly separated resulting in uniform concentration maps of H and DAB labels. Figure 3.36 demonstrates a second example of spectral unmixing using the proposed subtraction algorithm.

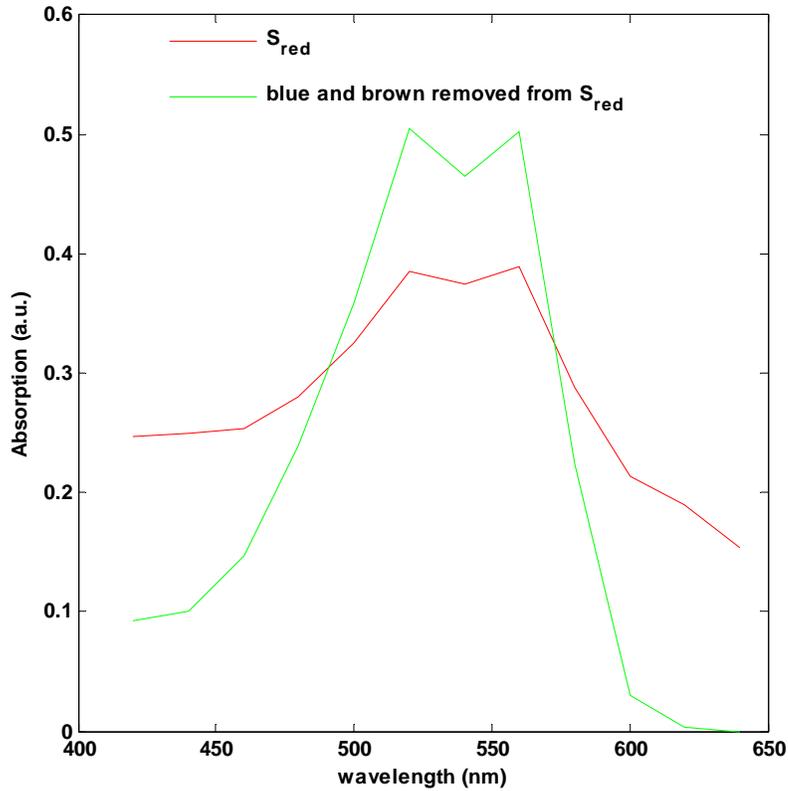


Figure 3.32 S_{red} after the appropriate proportions of S_{blue} and S_{brown} are subtracted from it.

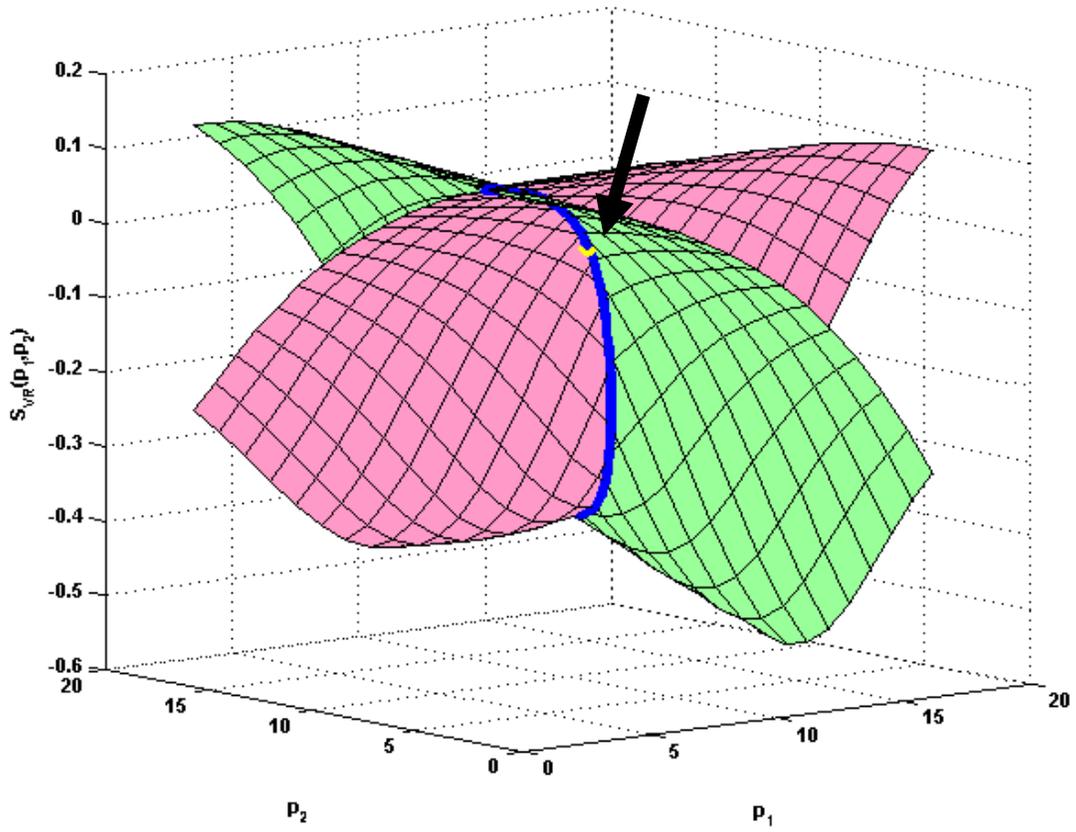
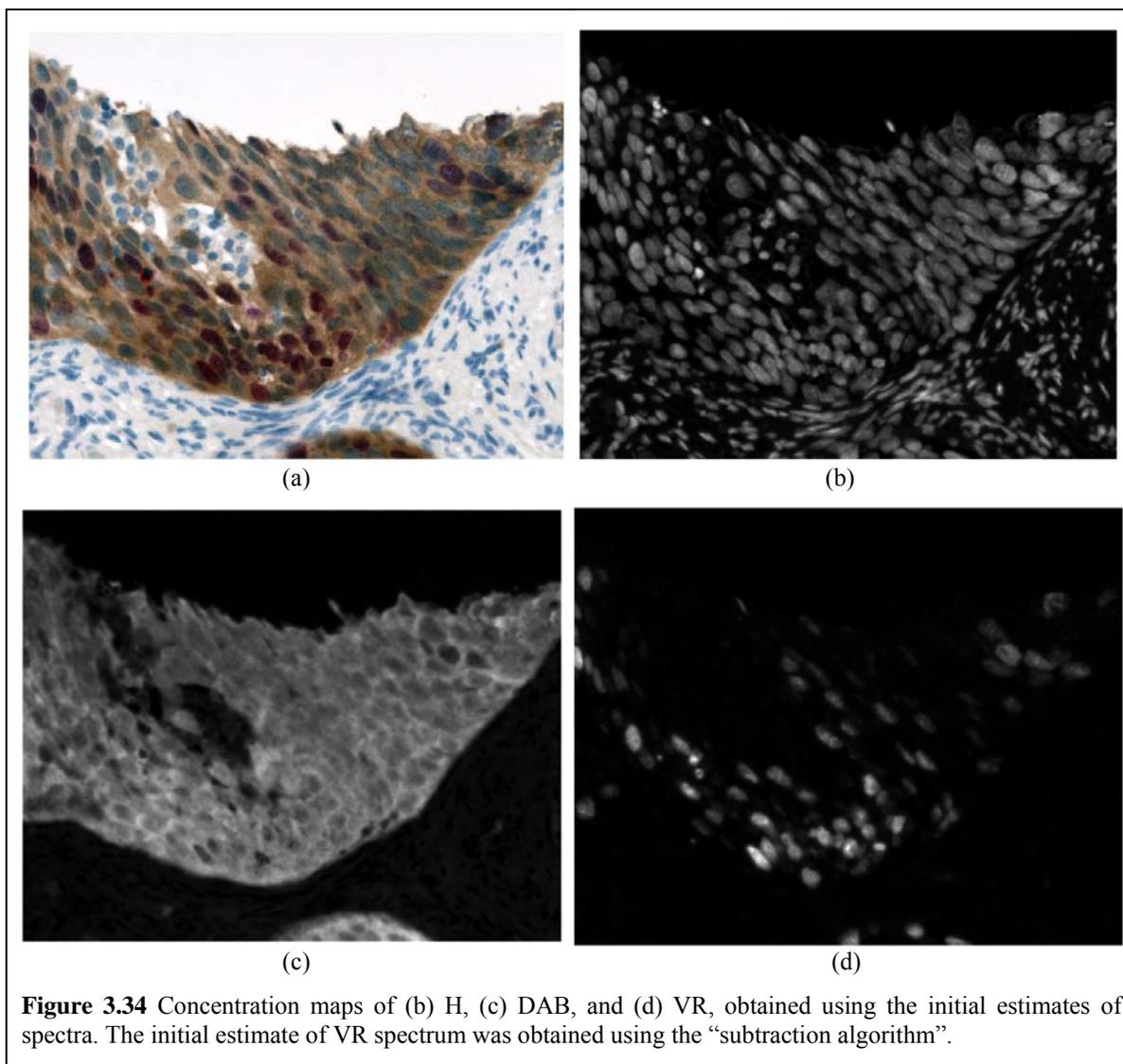


Figure 3.33 The two surfaces $S_1(p_1, p_2)$ (shown in green) and $S_2(p_1, p_2)$ (shown in pink) and their intersection curve (shown in blue), the point (\hat{p}_1, \hat{p}_2) , where the blue curve is equal to zero, is marked with a yellow dot.



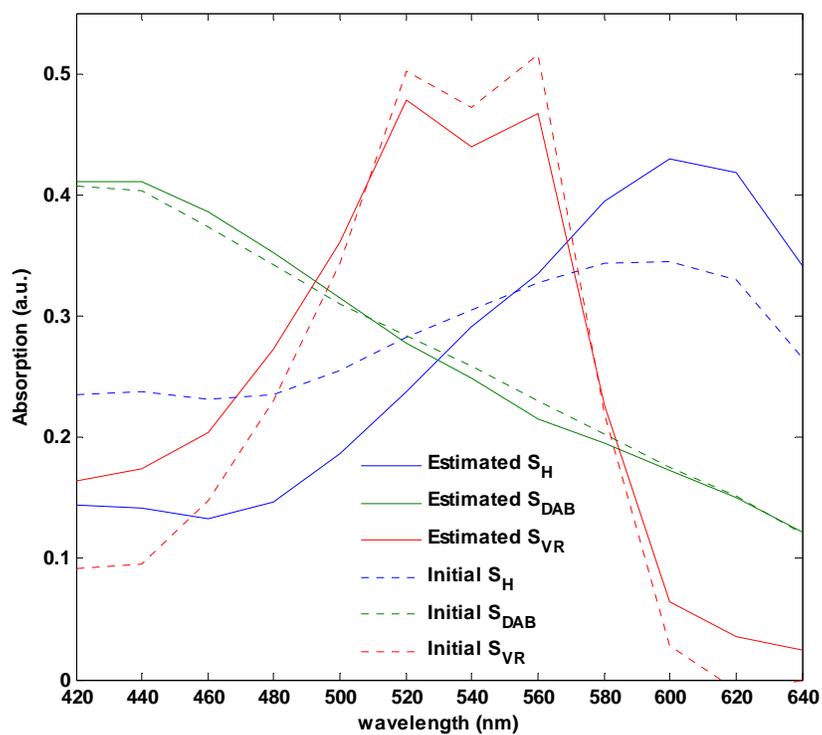
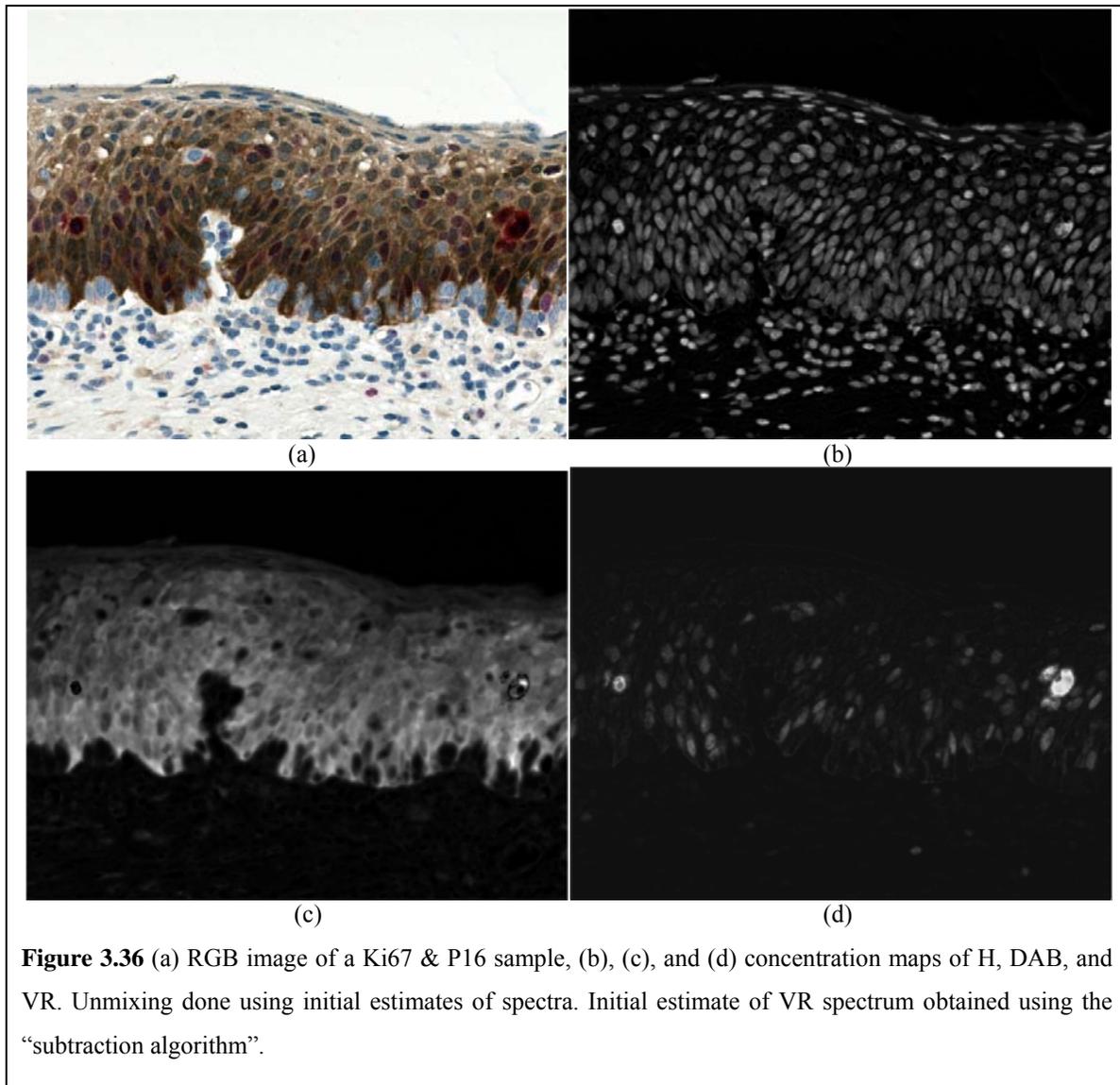
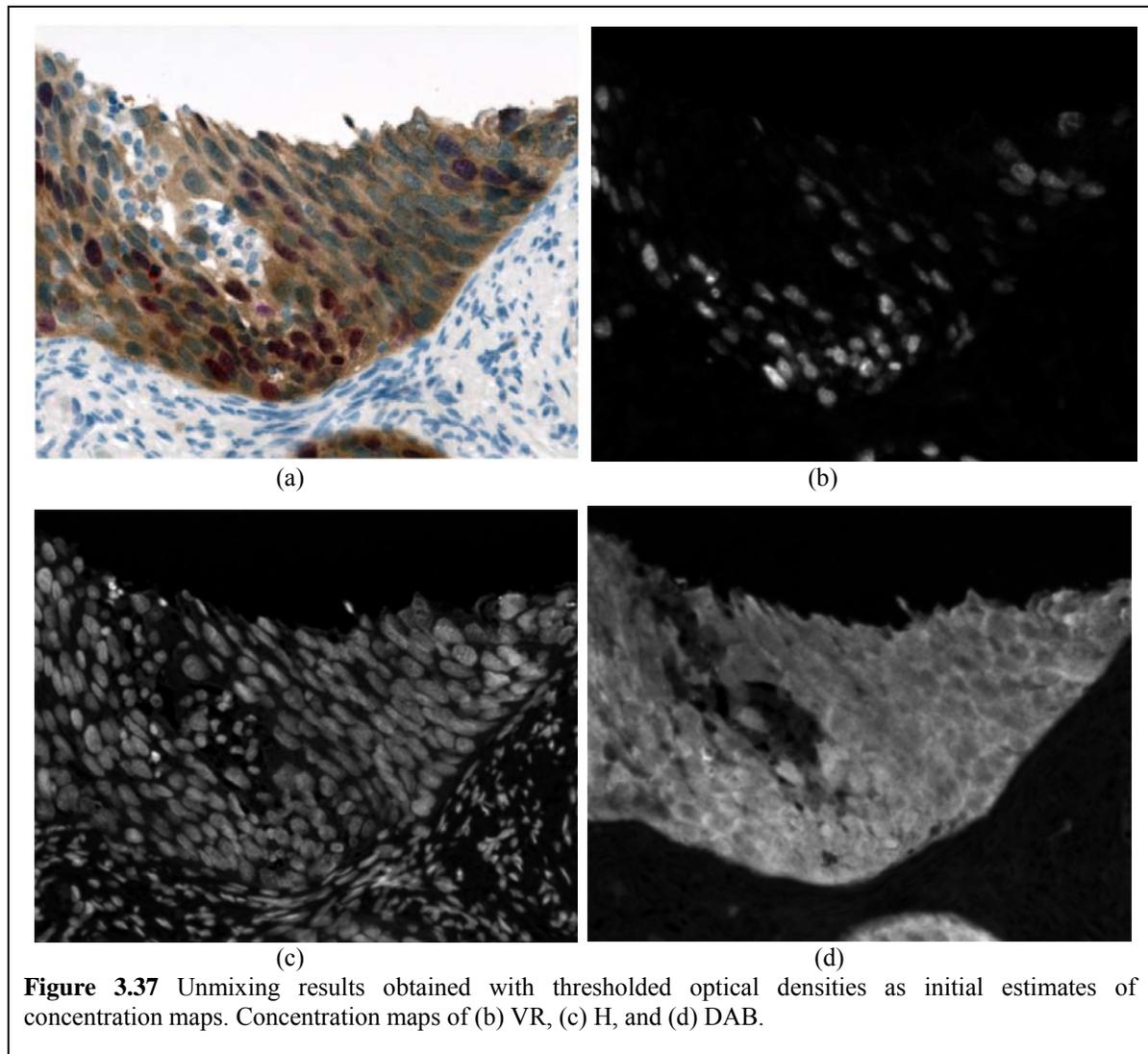


Figure 3.35 Double immunostained sample, unmixed spectra. Spectral unmixing with initial estimates of spectra. Initial estimate for VR spectrum obtained using the “subtraction algorithm”.



Alternatively, the spectral unmixing process could be initiated using the initial estimates of the concentration maps. To obtain the initial concentration maps, we applied the SIMPLISMA algorithm to the multispectral image stack to find the purest spectral bands. These spectral bands for the sample shown in Figure 3.30 (a) were 520 nm, 640 nm, and 420 nm. We then used the thresholded optical densities at these three spectral bands as the initial estimates of concentration maps. Figures 3.37 and 3.38 demonstrate the resulting concentration maps and unmixed spectra. The estimated VR absorption spectrum was slightly different from the one obtained using the initial estimates of spectra to initiate the optimization process. The left end of the spectrum was zero here, while in the previous

solution (using the initial estimates of spectra) it was constrained to be 0.1. In fact, without the prior knowledge of the absorption of VR at 420 nm, the only assumption we could make to find p_1 and p_2 in equation (3.19) would be non-negativity, which would result in constraining the left end of the spectrum (absorption of VR at 420 nm) to be equal to zero. Therefore the results obtained using the initial estimates of the spectra would be the same as the ones obtained here (using the initial estimates of the concentration maps). In the absence of any prior knowledge of the pure component spectra, thresholded optical densities at purest spectral bands provide a simple and effective way of initiating the spectral unmixing optimization procedure.



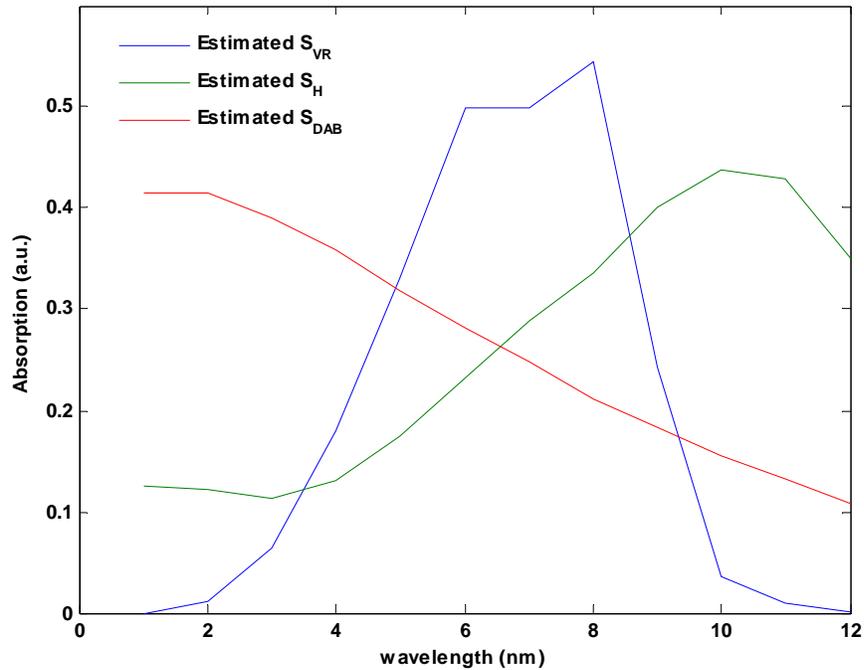


Figure 3.38 Double immunostained sample, unmixed spectra. Spectral unmixing with Initial estimates of concentration maps.

3.4.4 Discussion

We performed the linear spectral unmixing procedure on the multispectral images of our entire sample set that included 15 samples in the Ki67 set, 23 samples in the P16 set, 20 samples in the CD8 set, and 7 samples in the Ki67 & P16 set (total of 65 immunostained samples). The first three sets of samples differed with respect to the spatial co-localization of the labels: in the first set, the DAB label was only present in mixture with the H label. In the second set, DAB was present both as a pure label and as mixed with H. In the third set, no spatial overlap of the labels was expected. Obtaining the pure label spectra was straightforward in the third set, because of no spatial overlap. While in the first two sets, spatial overlap of the labels was taken into account. The fourth set contained three labels. H and DAB were present as pure labels, while VR was only present mixed with both H and DAB.

For each sample, the MCR-ALS procedure was done using both initial estimates of spectra and initial estimates of concentrations. The MCR-ALS Matlab toolbox developed by Jamout et al.¹³³ was used to perform the MCR-ALS procedure. Initial estimates of spectra were obtained using our proposed “multi-region SIMPLISMA” analysis. This analysis was followed by the “subtraction algorithm” for samples in the Ki67 and Ki67 & P16 sets, since the DAB label in Ki67 set and the VR label in the Ki67 & P16 set were not present in their pure forms. Therefore, the subtraction algorithm was used to estimate the pure spectra of those labels. The initial estimates of concentration maps were obtained by selecting the “purest” spectral bands using the SIMPLISMA algorithm followed by thresholding the absorption maps (optical densities) in those spectral bands.

Using the proposed methods for obtaining the initial estimates of either spectra or concentration maps from image data removed the need for obtaining the spectra of labels from singly labeled tissue sections.

Typical examples of the linear spectral unmixing results were demonstrated in the results section. Appendix A further presents the unmixed components spectra obtained for each sample using the proposed methods.

Visually, the concentration maps of the labels obtained using the initial estimates of the spectra were very similar to those obtained using the initial estimates of the concentration maps. When initial estimates of the concentration maps were used, however, the optimization algorithm typically converged in a smaller number of iterations, compared to when the initial estimates of the spectra were used. Also, obtaining the initial estimates of the concentration maps involved fewer computational steps compared to obtaining the initial estimates of spectra. This was because the SIMPLISMA analysis was done only once to find the “purest spectral bands” among the 12 spectral bands, while obtaining the initial estimates of the spectra involved performing the SIMPLISMA analysis on multiple regions within a tissue section. This was also followed in some instances by the “subtraction algorithm”.

In samples with dark DAB staining, correct unmixing of the labels was not achieved (Figures 3.16 and 3.26). This was shown to be caused by the difference between the absorption spectra of the weak DAB stain and strong DAB stain (as shown in Figure 3.19), which indicated that DAB does not follow the Beer-Lambert law, i.e. its absorption does not linearly increase with its increased concentration. This observation is in line with findings of

Van Der Loos et al.^{93,132} We proposed to perform unmixing on samples with dark DAB staining by considering the dark DAB stain a separate component. This significantly improved the unmixing results (as can be seen in Figure 3.16 second and third rows).

Linear spectral unmixing was performed on all 65 samples within this study in two ways: 1) spectral unmixing was performed on the multispectral image stack of each sample, which consisted of 12 images acquired at 12 narrow-band illumination channels and 2) spectral unmixing was performed on the three-channel RGB image of each sample that was acquired at three broad-band illumination channels within the red, green, and blue regions of the light spectrum. These two methods were referred to as “multispectral unmixing” and “RGB unmixing” respectively. For the samples labeled with DAB and H, no significant differences were observed in the unmixed images of the individual labels generated by each method. This is an expected result. H and DAB have been the stains of choice for pathologists for a long time because they are easily distinguishable with unaided eye or within RGB images. Even though the results are visually similar, we further compared the effectiveness of each imaging approach in generating accurate nuclei segmentation results. The quantitative results of this comparison will be presented in section 3.6. As will be shown, multispectral imaging results in more accurate segmentation of nuclei.

In samples labeled with H, DAB, and VR labels, however, RGB imaging was unable to unmix DAB and VR, because of the high overlap of their absorption spectra within the three channels of the RGB image (see Figures 3.27 and 3.28). Multispectral imaging resulted in successful unmixing of the three labels (Figure 3.29). Ultimately, the number of spectral bands necessary for the unmixing of the labels in a sample depends on the number and the spectral properties of the labels.

3.5 Linear Discriminant Analysis of Multispectral Images

As an alternative to linear spectral unmixing, in this section we examine the potential of Linear Discriminant Analysis (LDA) for obtaining the distribution maps of labels in multiple labeled tissue sections.

3.5.1 Methods

Linear Discriminant Analysis (LDA) is a statistical method of finding a linear combination of variables which characterizes or separates two or more classes of objects. It involves finding a weighted linear combination of variables that maximizes the ratio of between-class variance to within-class variance with the objective of maximizing class separability.¹³⁴ The weighted linear combination of variables is referred to as the linear discriminant function. The weights in the linear discriminant function are calculated based on training data.

In the case of multispectral image data, the pixel intensities at multiple spectral bands of the multispectral image form the variables. Two or more classes may be defined based on the absorption spectrum or color of the pixels, e.g. class of brown pixels vs. class of blue pixels.

Our approach is to use LDA to find a linear discriminant function that linearly transforms the images in a multispectral image stack into a grayscale image that emphasizes the spectral differences between a component of interest and the rest of the tissue section. The component of interest can be a certain label (such as H or DAB); the grayscale image would thus represent the distribution of that label in the tissue section.

For a tissue section that is labeled with N labels, N linear discriminant functions corresponding to the N labels are obtained. Each linear discriminant function generates an image that represents the distribution of the corresponding label in the tissue section.

If one label (e.g. H) labels nuclei and another one (e.g. DAB) labels an IHC biomarker, then one grayscale image represents the labeled nuclei, and the other one represents the labeled biomarker. The image representing the labeled nuclei can then be used for the automated segmentation of the nuclei. The image that represents the labeled biomarker can be used for the automated measurement of the biomarker expression in the tissue sample.

The procedure for generating the image that represents a label using LDA involves two steps: 1) a training set of pixels is formed that consists of two groups of pixels: the first group

contains pixels that are labeled with the label of interest. The second group contains pixels with labels other than the label of interest and background pixels and 2) a two-class LDA is performed on the training set and a linear discriminant function is obtained that transforms the images in the multispectral image stack into one grayscale image.

3.5.2 Results

In the following sub-sections the results of using LDA to obtain images representing different labeled structures in multiple labeled tissue sections are presented.

3.5.2.1 Ki67 Set

In the case of Ki67 samples, the pixels can be divided into three groups:

Group 1: Nuclei pixels labeled with H. These pixels look blue in the RGB image,

Group 2: Nuclei pixels labeled with both H, and DAB. These pixels look brown in the RGB image, and

Group 3: Cytoplasm and background pixels. The cytoplasm pixels have a faint blue color in the RGB image, which is the result of non-specific H staining. Background pixels look white.

A three-class LDA would determine two linear discriminant functions with some optimal combination of variables so that the first function provides the most overall discrimination between groups, and the second provides the second most. Alternatively, we could specify how to combine the groups to form the different discriminant functions, i.e. we could perform two separate two-class discriminant analyses: 1) one to separate the labeled nuclei pixels (including blue and brown) from all other pixels, which is equivalent to separating the pixels in groups 1 and 2 from the pixels in group 3 and 2) one to separate all DAB labeled pixels from all other pixels, which is the equivalent of separating the pixels in group 2 from the pixels in groups 1 and 3 combined. We chose the second approach. This was because with this approach, each discriminant function transforms the stack of multispectral images into one grayscale image that “highlights” a specific labeled structure or a certain label of interest. With the first approach, neither of the two discriminant functions would necessarily separate one group from the other two combined. Instead, the combination of the two

discriminant functions would maximize the overall discrimination between all groups, which is not what we want.

To determine the discriminant functions, a training set was required. For each image stack, areas representing each group of pixels were selected. The multispectral intensities for every pixel in these areas and the group that the pixel belongs to were recorded, and formed the training set. A total of approximately 1000 pixels per group (3000 in total) were recorded for each image. A training set for the entire dataset (15 image stacks of the 15 Ki67 samples) was built by combining the individual training sets.

The training set consisted of three groups. A linear discriminant function was designed to discriminate between the combination of groups 1 and 2 and group 3. This function would represent all pixels that are labeled with either H alone or a mixture of H and DAB. A second linear discriminant function was designed to separate group 2 from the combination of groups 1 and 3. This function would represent DAB-labeled pixels. The two discriminant functions were calculated as:

$$F_1 = \sum_{i=1}^{12} c_{1i}I_i, \text{ and } F_2 = \sum_{i=1}^{12} c_{2i}I_i \quad (3.23)$$

where F_1 and F_2 are the first and second discriminant functions, c_{1i} and c_{2i} , $i = 1, 2, \dots, 12$, are the corresponding coefficients, and I_i , $i = 1, 2, \dots, 12$ are the pixel intensities within the 12 spectral bands.

The two discriminant functions were then used to transform each image stack into two grayscale images. For comparison, we also performed the same LDA analysis on the three-channel RGB images. Figures 3.39, and 3.40 show the coefficients for the first and second discriminant functions and for the multispectral and the RGB images, respectively. Figure 3.41 shows an example of the generated grayscale images for a typical Ki67 sample.

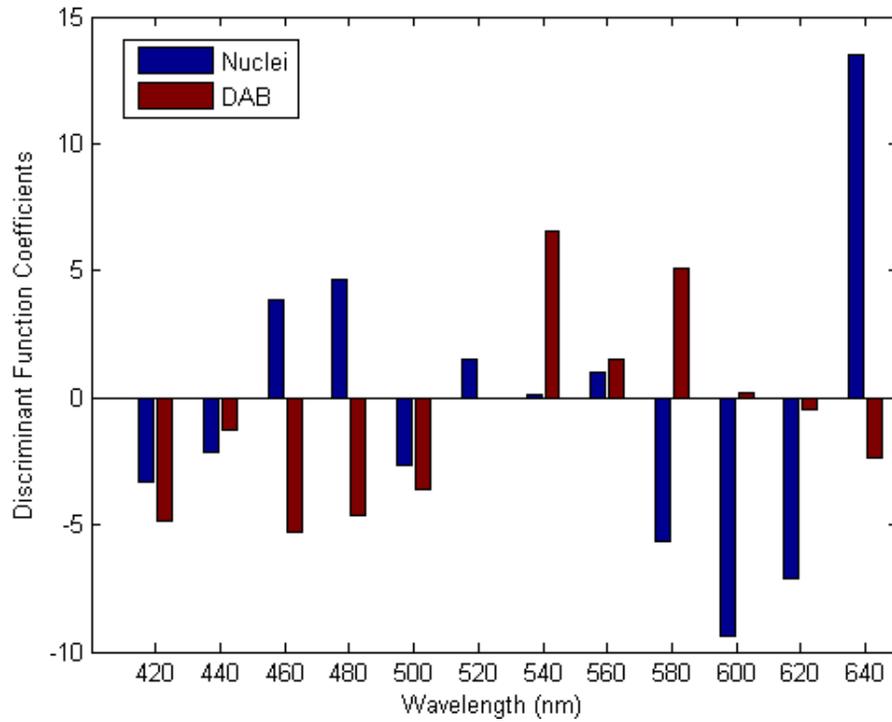


Figure 3.39 Coefficients of the linear discriminant functions obtained from the multispectral image stacks of Ki67 samples.

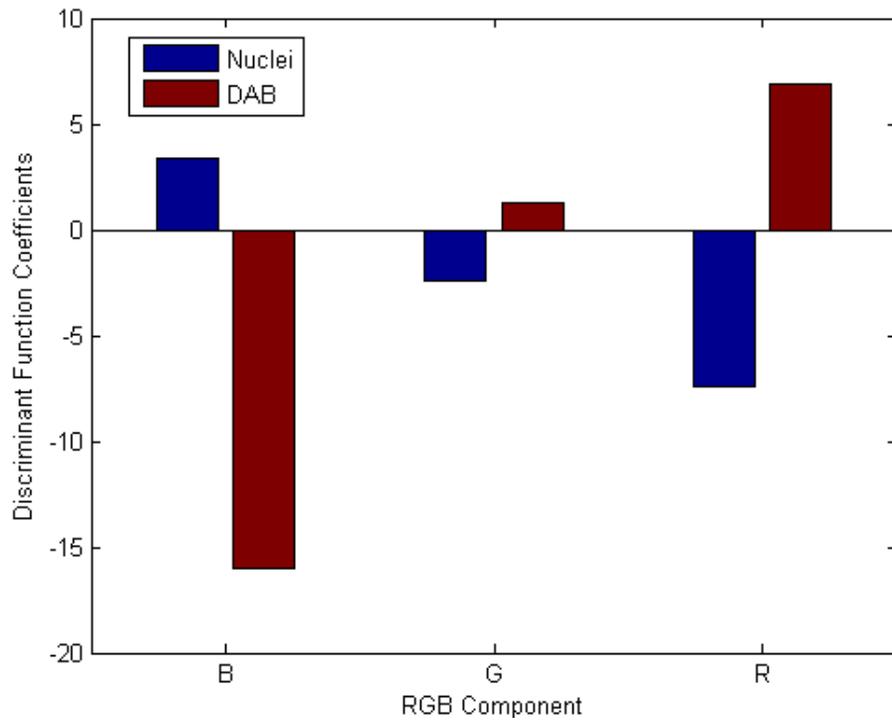


Figure 3.40 Coefficients of the linear discriminant functions obtained from the RGB images of Ki67 samples.

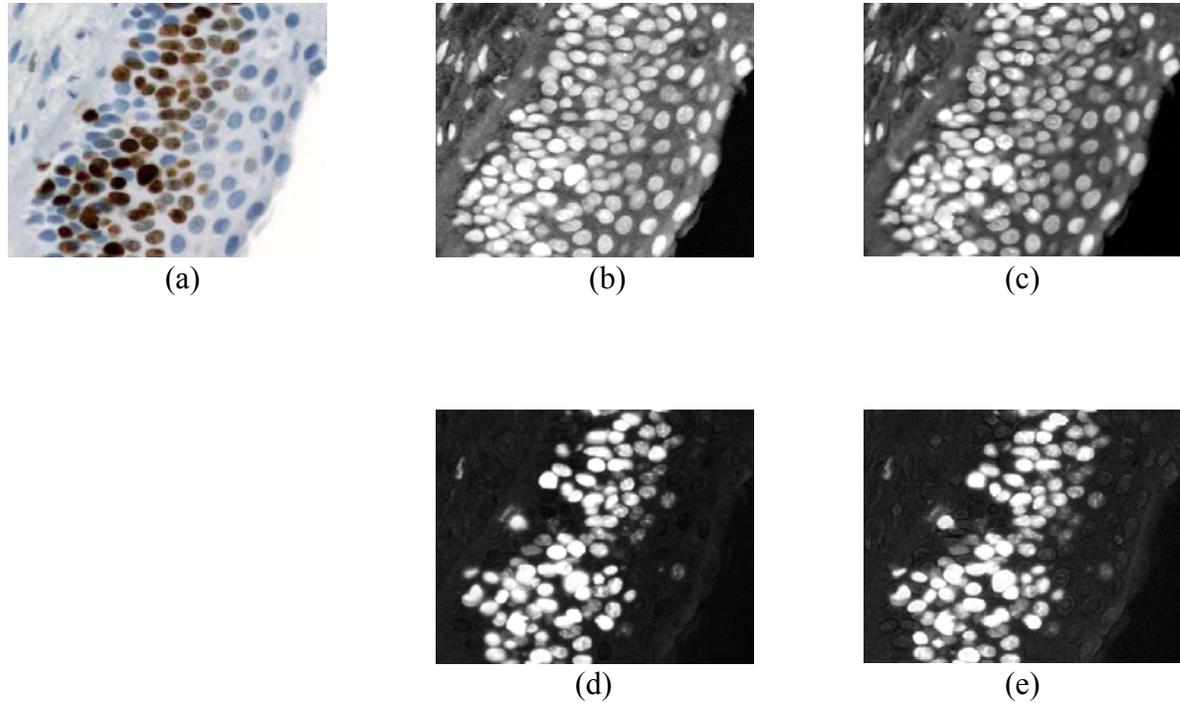


Figure 3.41 (a) RGB image of a Ki67 sample, (b) and (d) grayscale images representing nuclei and DAB generated by the linear discriminant functions obtained from the multispectral image stacks, (c) and (e) grayscale images representing nuclei and DAB generated by the linear discriminant functions obtained from the RGB images.

3.5.2.2 P16 Set

In the case of P16 samples, there were four groups of pixels: 1) blue nuclei pixels, 2) brown nuclei pixels, 3) brown cytoplasm pixels, and 4) blue cytoplasm pixels. In this case, we determined one discriminant function that separated groups 1 and 2 from groups 3 and 4 combined. This function transformed the image stack into a grayscale image that highlighted (represented) all nuclei (both blue and brown). We then determined a second discriminant function that separated groups 2 and 3 from groups 1 and 4. This resulted in a grayscale image that highlighted all DAB-labeled pixels (either nuclei or cytoplasm).

The discriminant functions for P16 samples were determined in a way similar to those for Ki67 samples. Figures 3.42 and 3.43 show the coefficients for the first and second discriminant functions. Figure 3.44 shows an example of the grayscale images for one typical image stack.

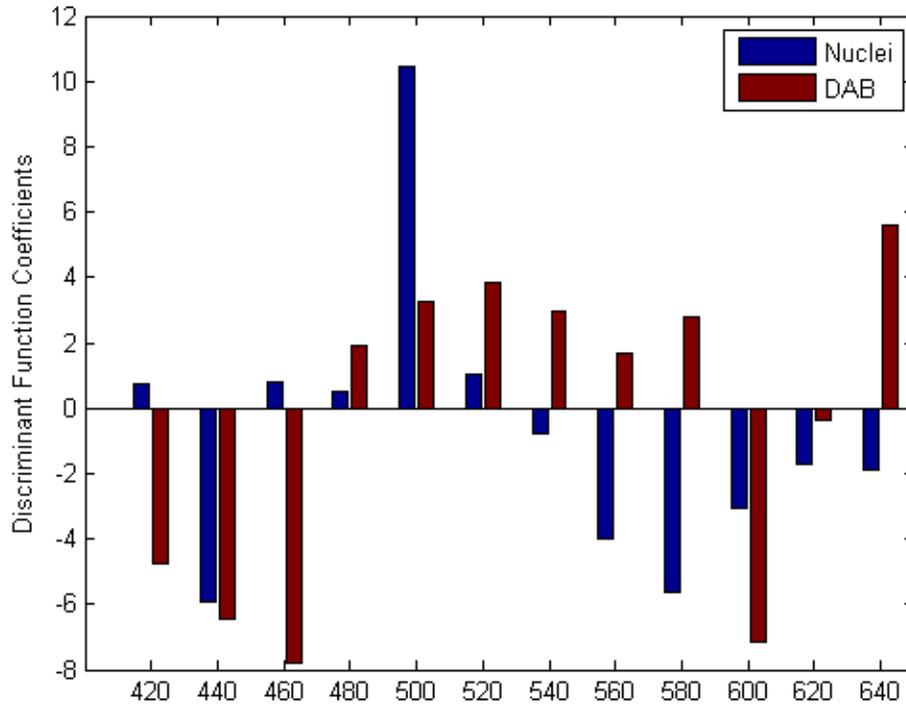


Figure 3.42 Coefficients of the linear discriminant functions obtained from the multispectral image stacks of P16 samples.

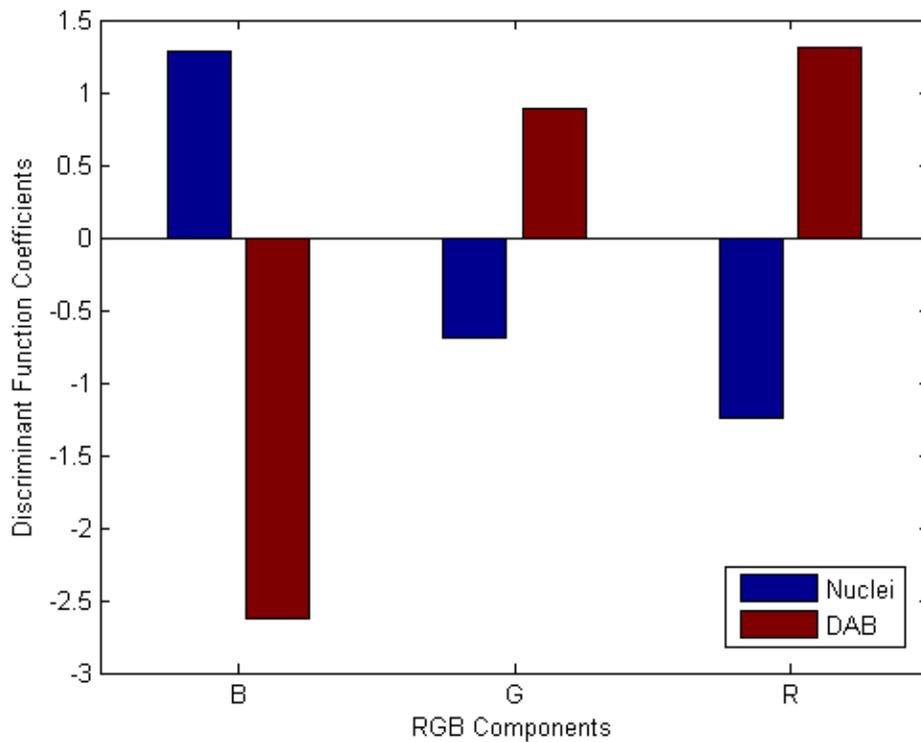


Figure 3.43 Coefficients of the linear discriminant functions obtained from the RGB images of P16 samples.

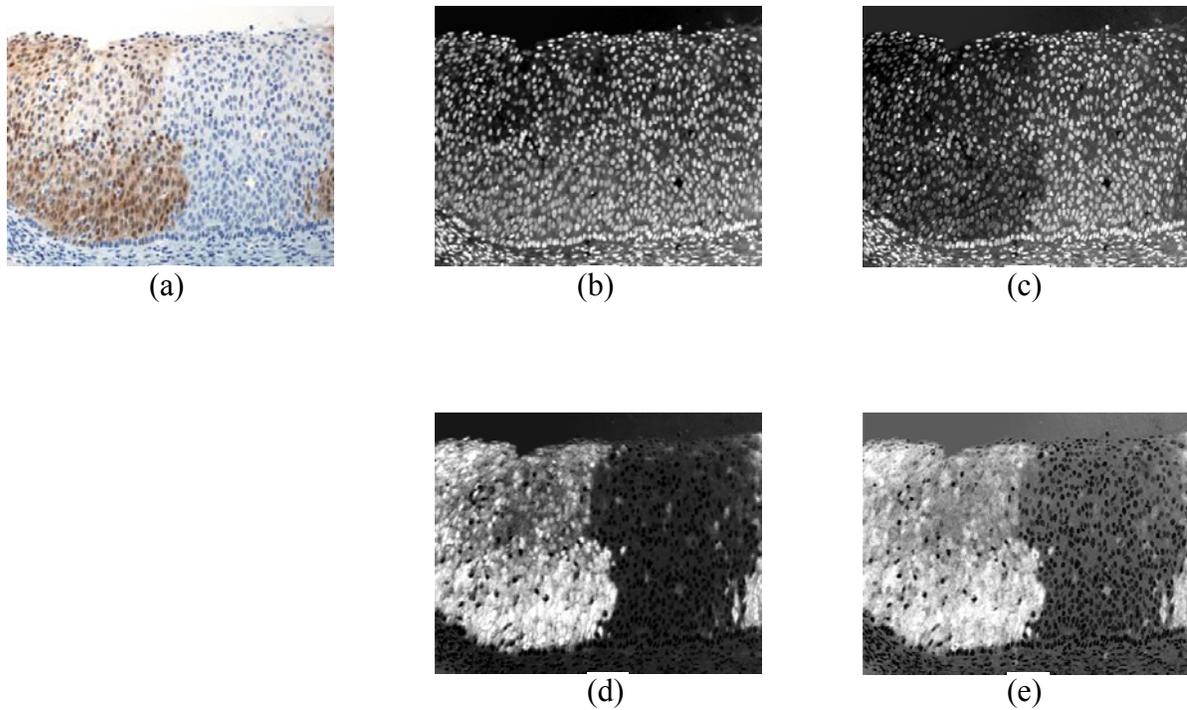


Figure 3.44 (a) RGB image of a P16 sample, (b) and (d) grayscale images representing nuclei and DAB generated by the linear discriminant functions obtained from the multispectral image stacks, (c) and (e) grayscale images representing nuclei and DAB generated by the linear discriminant functions obtained from the RGB images.

3.5.2.3 CD8 Set

In case of the membrane samples, there are three groups of pixels: 1) blue nuclei pixels, 2) brown membrane pixels, and 3) cytoplasm pixels. So the first discriminant function was designed to separate group 1 from the others, and the second discriminant function was designed to separate group 2 from the others.

The discriminant functions were determined in a similar way to those of the Ki67 samples. Figures 3.45 and 3.46 show the coefficients for the first and second discriminant functions. Figure 3.47 shows an example of the grayscale images for one of the image stacks.

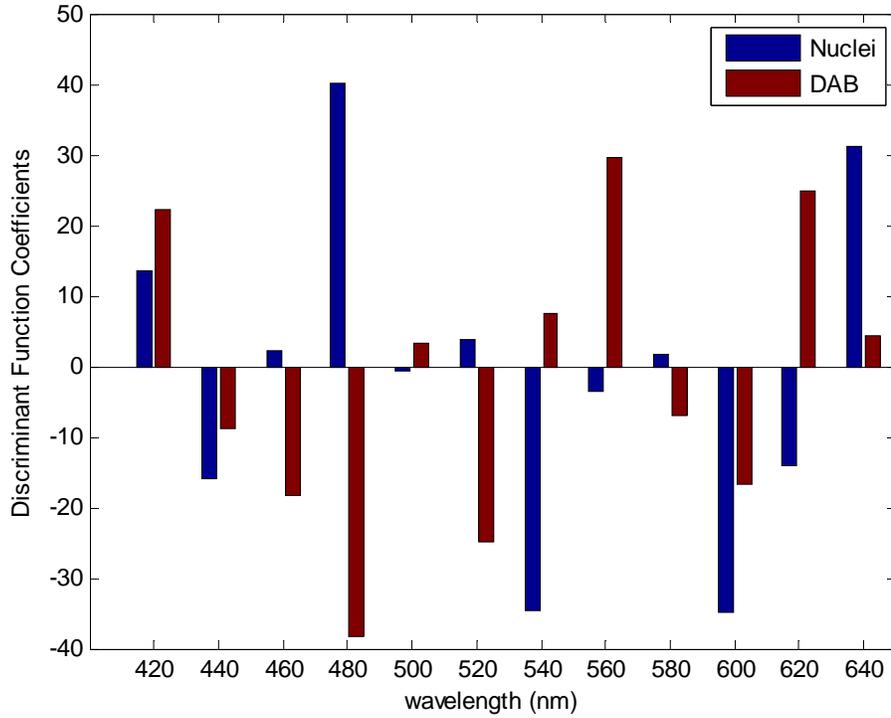


Figure 3.45 Coefficients of the linear discriminant functions obtained from the multispectral image stacks of CD8 samples.

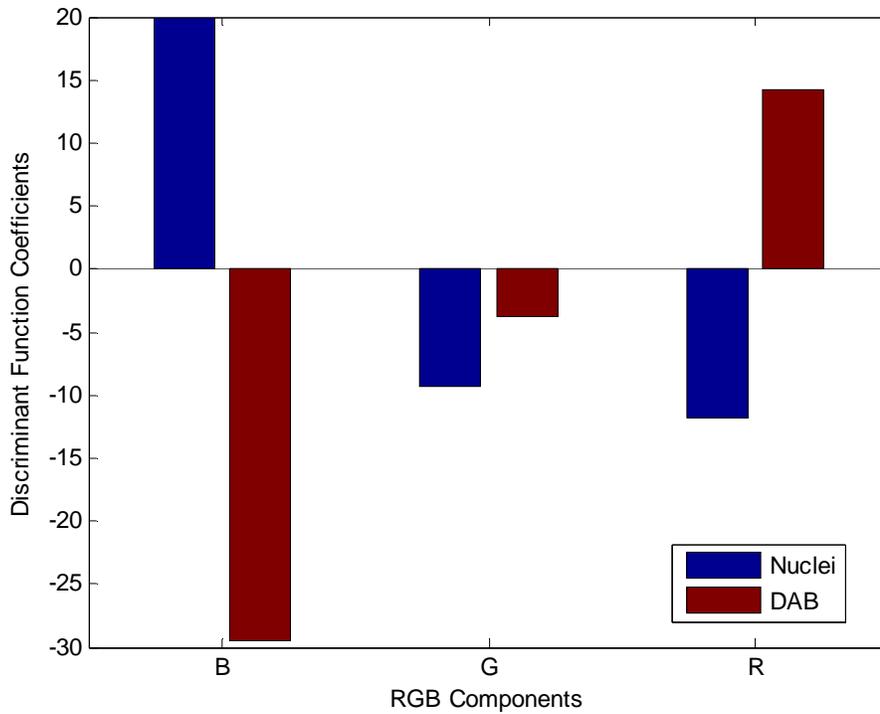


Figure 3.46 Coefficients of the linear discriminant functions obtained from the RGB images of CD8 samples.

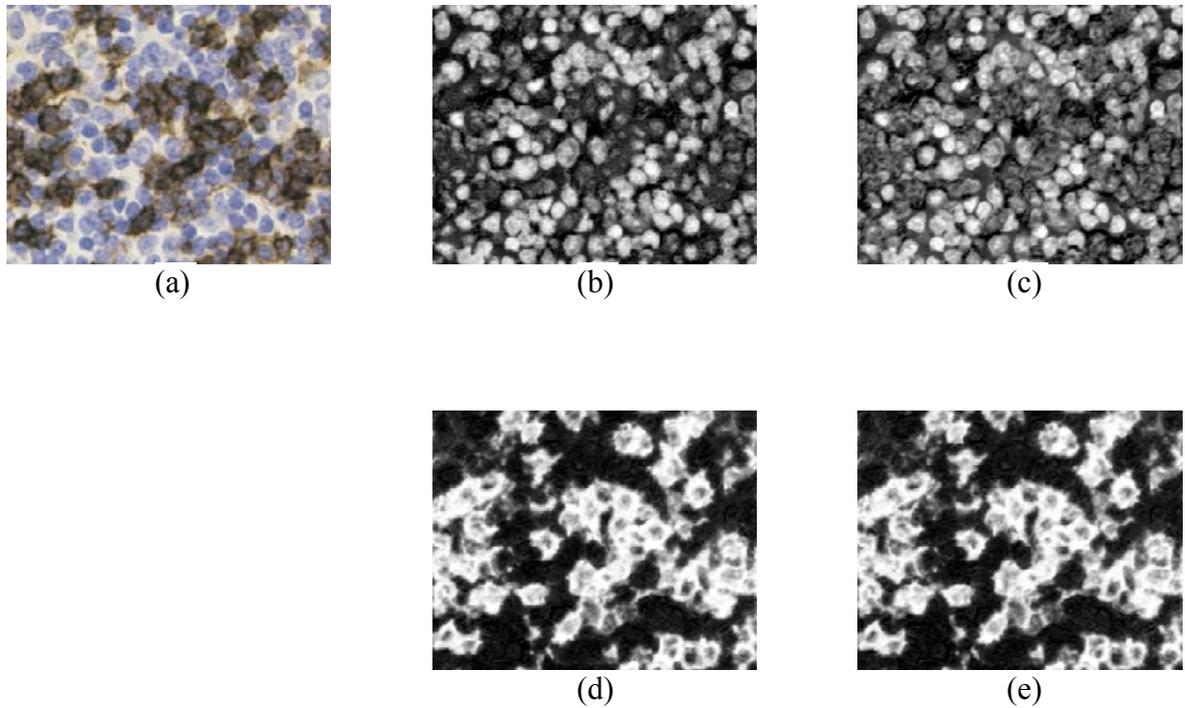


Figure 3.47 (a) RGB image of a CD8 sample, (b) and (d) grayscale images representing nuclei and DAB generated by the linear discriminant functions obtained from the multispectral image stacks, (c) and (e) grayscale images representing nuclei and DAB generated by the linear discriminant functions obtained from the RGB images.

3.5.2.4 Ki67 & P16 set

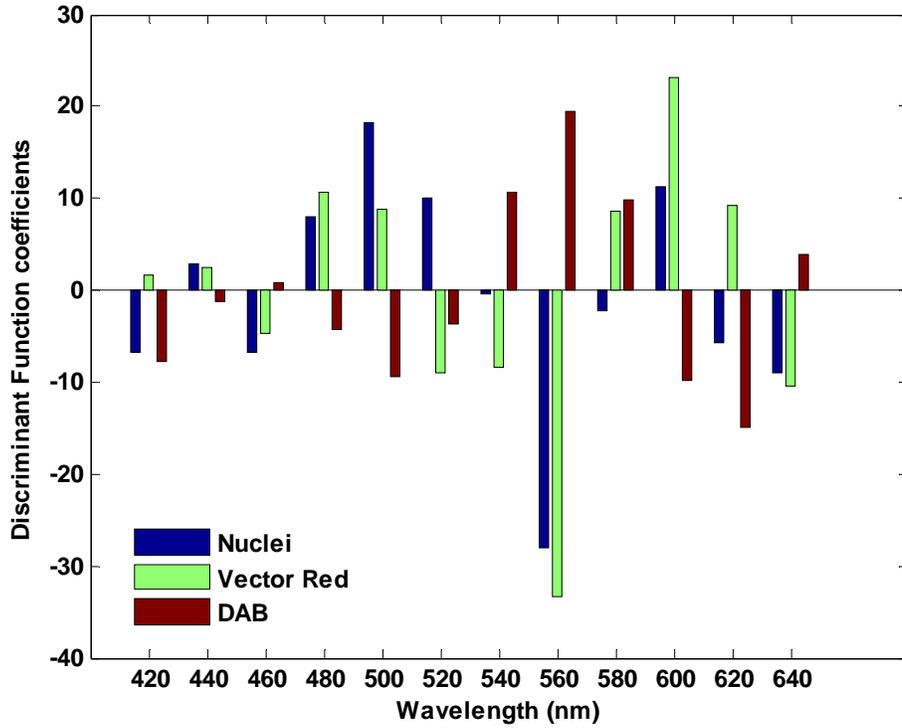
In the case of double-immunostained samples, there are five groups of pixels:

- 1) Blue nuclei pixels,
- 2) Brown nuclei pixels,
- 3) Red nuclei pixels,
- 4) Brown cytoplasm pixels, and
- 5) Other cytoplasm and background pixels.

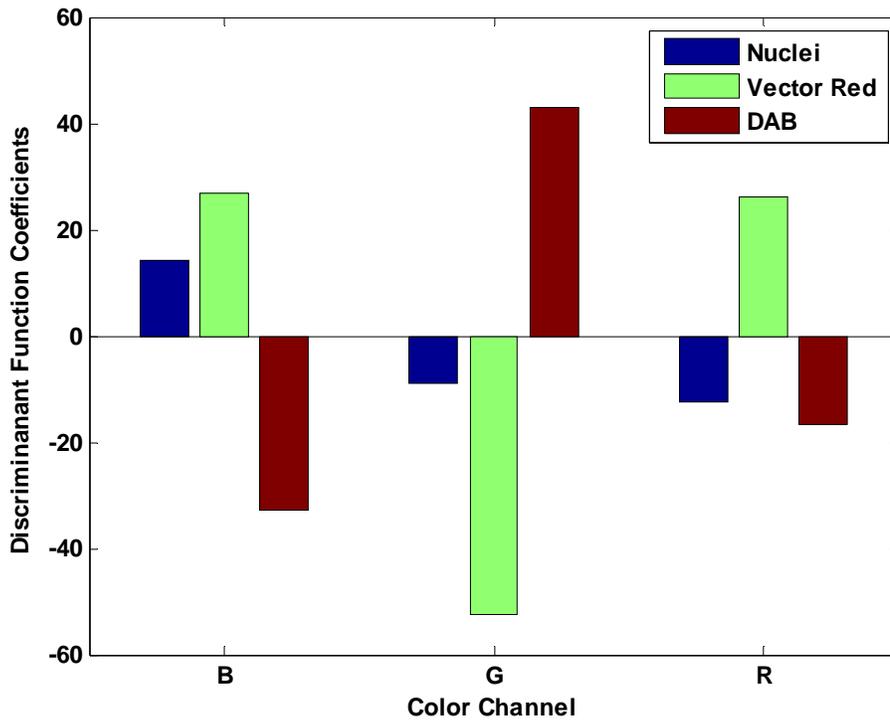
Three linear discriminant functions were designed: 1) to separate pixels in groups 1, 2, and 3 from all other pixels, 2) to separate red nuclei pixels, group 3, from all other pixels, and 3) to separate brown nuclei and cytoplasm pixels, groups 2 and 4, from all other pixels. Figure 3.48 shows the coefficients of these discriminant functions. Figures 3.49 and 3.50 demonstrate a typical example of grayscale images obtained from one sample's RGB image and multispectral image stack, respectively. Note in Figure 3.50 (d) that in the position of red

nuclei pixels no DAB signal is detected by the linear discriminant function. This is because the linear discriminant function was not designed to detect those nuclei with both DAB and VR staining. We know, however, that DAB and VR spatially overlap in those nuclei pixels. We therefore designed the third discriminant function such that it separated groups 2, 4, and 3 from all other pixels. In this manner, the coefficients would be chosen so that they emphasized the features similar among these three groups of pixels but different in all other pixels. Figure 3.50 (e) shows the result of the new discriminant function. Note that the holes in the location of red nuclei disappeared. Figure 3.51 shows the coefficients for the new linear discriminant function.

There were no pure VR-labeled pixels in our images. Hence, we could not verify if LDA was able to discriminate between the mixture of DAB + VR and the pure VR. In the example of Figure 3.52, there was one nucleus that was labeled with VR in the stroma which did not appear to have DAB labeling. However, in the generated DAB image, the nucleus appeared brighter than the background, indicating that the DAB signal was detected in that nucleus. Thus, it appears that the linear discriminant function was not able to discriminate between the Vector Red, and the combination of Vector Red and DAB signals. This may be a limitation of the approach due to the fact that spectral profiles of these two labels were too close. Alternatively, it may be because there were no pure VR pixels in the training set, so LDA was not trained to find the features that would discriminate pure VR pixels from the combination of the VR and DAB. This needs to be confirmed in samples where these labels are present in pure form as well as in mixtures with other labels.

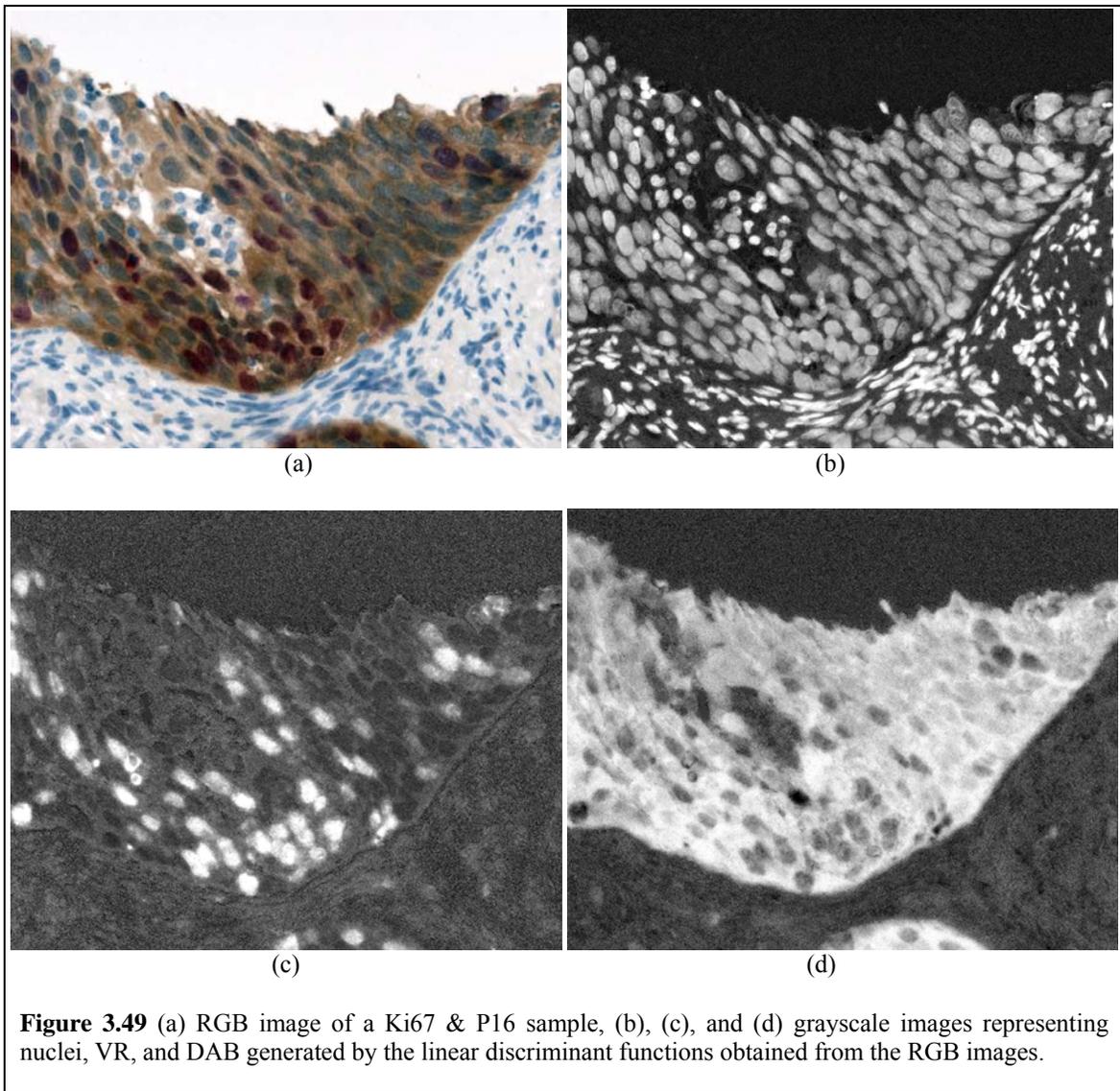


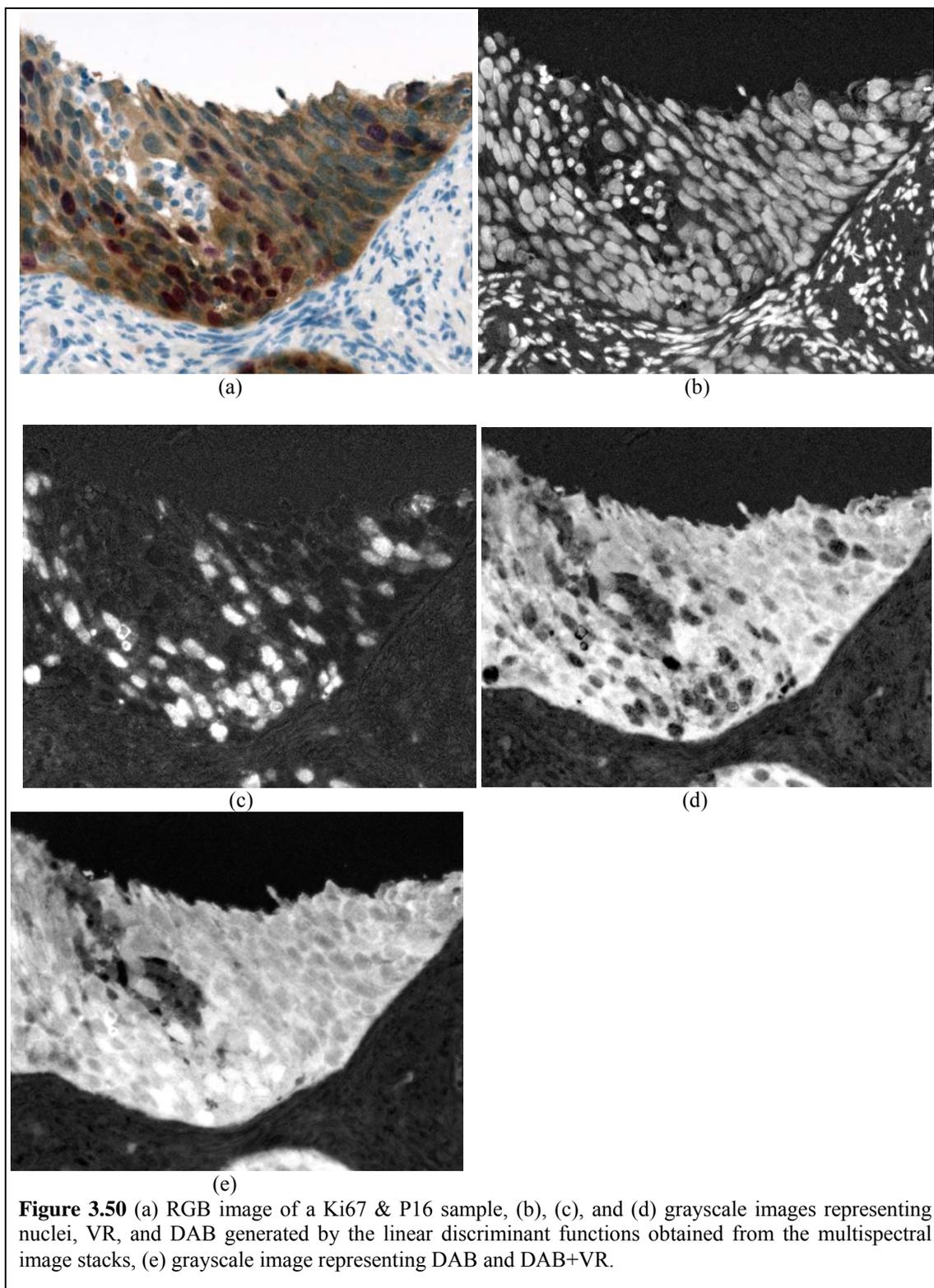
(a)



(b)

Figure 3.48 (a) Coefficients of the linear discriminant functions obtained from the multispectral image stacks of Ki67 & P16 samples , (b) coefficients of the linear discriminant functions obtained from the RGB images of Ki67 & P16 samples.





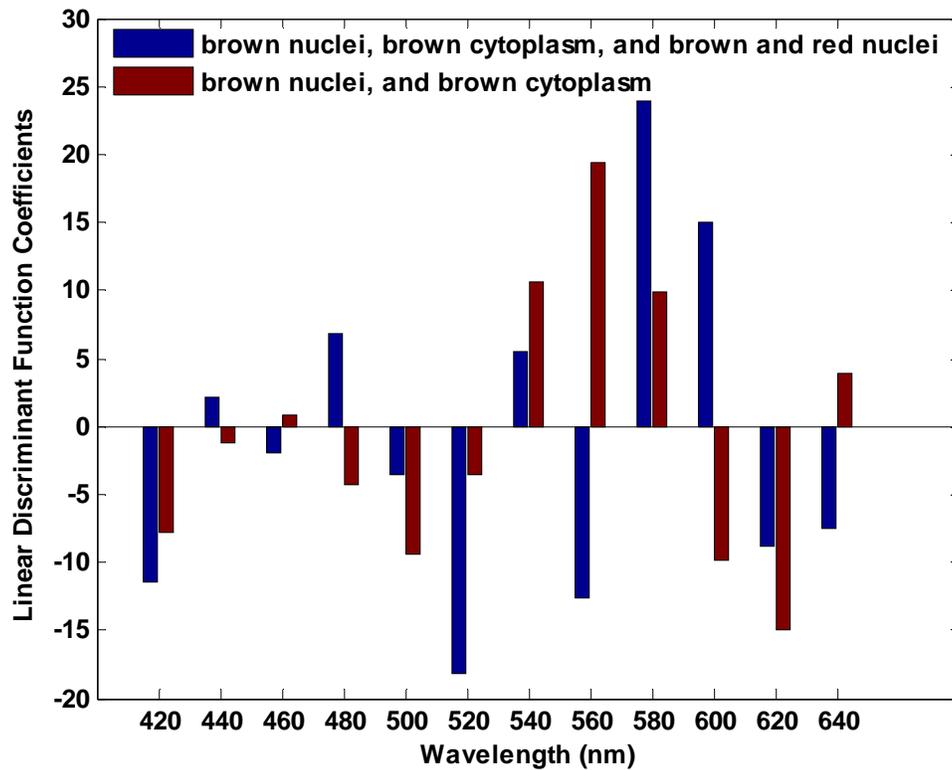


Figure 3.51 Coefficients of the linear discriminant functions for Ki67 & P16 samples. Red bars represent coefficients for the identification of DAB and DAB+H. Blue bars represent coefficients for the identification of DAB, DAB+H, and DAB+VR.

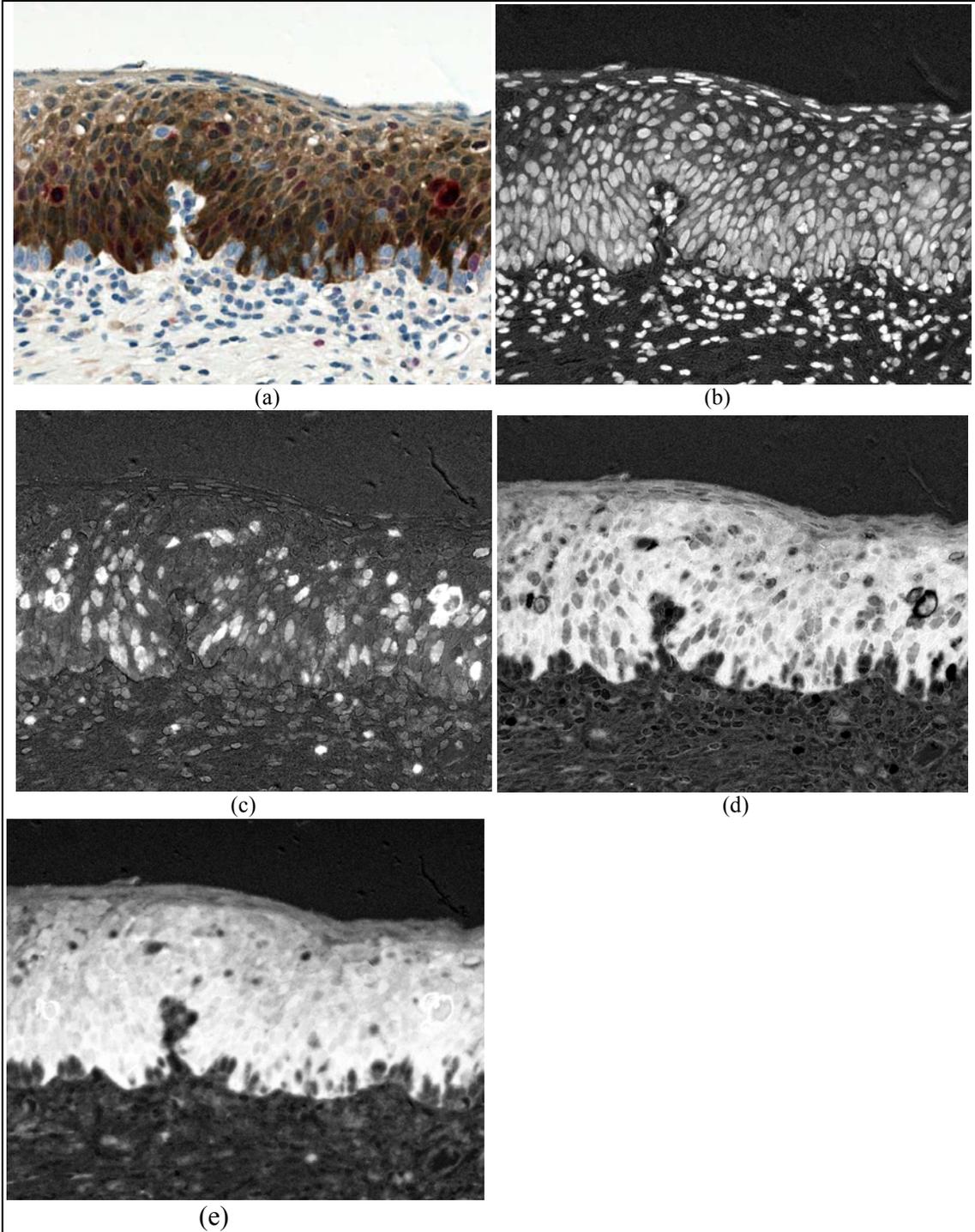


Figure 3.52 (a) RGB image of a Ki67 & P16 sample, (b), (c), and (d) grayscale images representing nuclei, VR, and DAB generated by the linear discriminant functions obtained from the multispectral image stacks, (e) grayscale image representing DAB and DAB+VR.

3.5.3 Discussion

LDA was used as an alternative to the linear spectral unmixing approach (discussed in section 3.4) to obtain images that represented the labels in a multiple labeled tissue section. The difference between the two approaches is that linear spectral unmixing divides the absorption at each pixel into absorptions of the components (labels) present in that pixel. A multi-class LDA, however, would assign each pixel, in which multiple labels are present, to only one of the multiple labels (one class of the pixels). As a result, co-localization of the labels would not be taken into account. Hence, we used LDA in a different way: we performed multiple linear discriminant analyses. Each linear discriminant function was designed to discriminate between those pixels in which a certain label was present (either alone or mixed with other labels) and all the other pixels. In this way, LDA generated an image that represented the distribution of that certain label within a tissue section. If, for example, two labels were present in one sample, each label was separately represented in the image generated by one of the two linear discriminant analyses. The co-localization information of the labels was thus preserved.

Linear discriminant functions were trained on training sets obtained from several tissue samples within each set. The reason for this was to take into account the possible variations among the spectral properties of labels from sample to sample.

Linear discriminant functions were calculated for all samples within this study, using two image sets: 1) multispectral images consisting of images acquired at 12 narrow-band illumination channels and 2) three-channel RGB images that were acquired with three broad-band illuminations within the red, green, and blue regions of the light spectrum.

Visual evaluation of the resulting grayscale images demonstrated that LDA was successful in generating image representations of different labels in each tissue section (typical examples shown in Figures 3.40, 3.43, 3.46, and 3.48). No significant differences between the results from the RGB images and the results from the multispectral images were observed.

We then compared the effectiveness of each imaging approach (multispectral versus RGB) in generating accurate nuclei segmentation results. The quantitative results of this comparison will be presented in section 3.6. These results will also be compared to those of the linear spectral unmixing from the previous section.

3.6 Segmentation of Nuclei

3.6.1 Methods

Once images representing the distribution maps of the labels are available, the distribution map of the nuclear label was used for the identification of nuclei. The segmentation algorithm consisted of four main steps: 1) initial binarization, 2) edge relocation, 3) distance-based watershed transform, and 4) merging.

It should be noted that in the segmentation algorithm, it was assumed that the objects to be segmented were brighter than the background.

3.6.1.1 Initial Binarization

Our fully automated procedure for the segmentation of nuclei starts with an initial binarization of the image. Due to the non-uniform background intensities and the significant changes in the brightness of the objects across the image (varying from very bright to very dim), global binarization methods, e.g. Otsu's method, will not perform well. One local binarization method is the local average method that uses the average intensity of pixels in the local neighborhood as a threshold for local binarization of the image. Pixels with intensities higher than the average are assigned to the foreground and pixels with intensities lower than average are assigned to the background.¹³⁵ A modification of this method is to replace the average by a gradient weighted average, so that pixels that have higher gradient values have a higher weight in calculating the weighted average. In this way, the threshold would be closer to the intensity of the pixels that have higher gradients and are thus likely to represent edges of objects. The threshold is therefore close to the intensity of the pixels on the border of the objects.

For every window of pixels, the threshold T is calculated as:

$$T = \frac{\sum I G}{\sum G} \quad (3.24)$$

where I is the pixel intensity, G is the pixel gradient, and the summation is over all the pixels inside the window. The threshold T computed in this way is assigned to the pixel in the center of the window. The window is then moved across the image and at each location the threshold for the center pixel in the window is calculated. To reduce the computational cost,

the windows do not overlap, so the threshold is only calculated for the pixels at the center of the non-overlapping windows. Finally, a bilinear interpolation is applied to these thresholds to find the threshold levels for all the pixels.

A Sobel gradient filter was applied to the image to find the gradient G . The Sobel filter is a 3×3 filter that emphasizes horizontal edges by approximating a vertical gradient:

$$h = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (3.25)$$

The transpose of the filter above, is a horizontal gradient which emphasizes the vertical edges. The gradient G was calculated as follows:

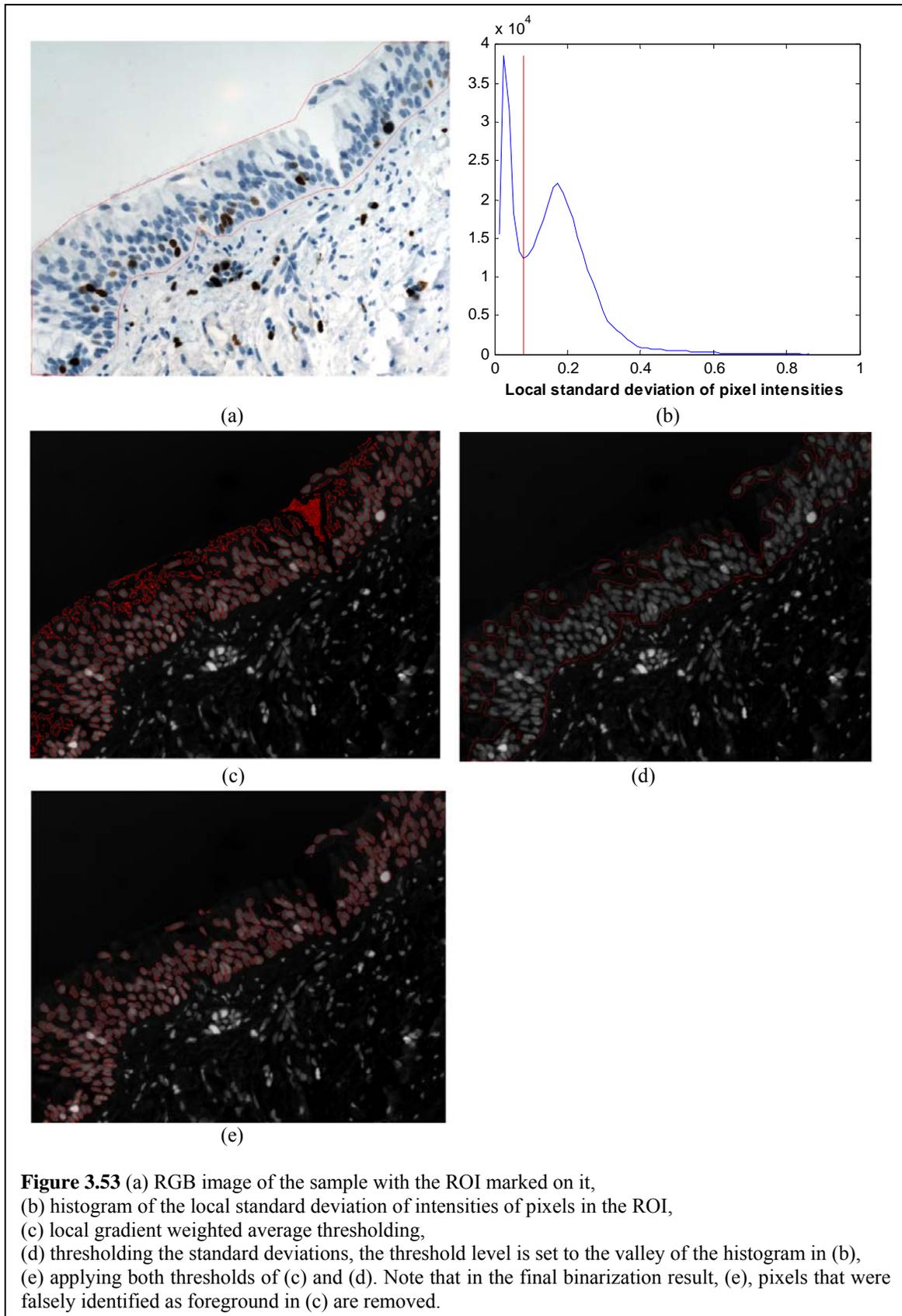
$$G_{vertical} = I * h, \quad G_{vertical} = I * h^T$$

$$G = \sqrt{G_{vertical}^2 + G_{horizontal}^2} \quad (3.26)$$

where the symbol $*$ refers to convolution. The size of the window affects the sensitivity of the thresholding procedure. Smaller windows result in more local thresholding and make the estimated threshold more sensitive to local variations in image intensity. The smaller window sizes are more effective in finding those objects that have less contrast with background, but they are also more sensitive to noise. We found that a window size that is approximately the size of an individual nucleus provides a good compromise.

Note that in regions of the image that are significantly bigger than the window size and do not contain any objects, the average intensity of the local pixels is very low. For these regions the above thresholding procedure will consider small background variations as foreground objects. If, in addition to the average intensity in each neighborhood, the standard deviation is also taken into account, this problem can be overcome. This is because areas that contain no objects will have lower local standard deviation. Consequently a pixel is considered a foreground pixel if 1) its intensity is higher than the average intensity of the local neighborhood and 2) the standard deviation of intensities of the pixels in the local neighborhood is higher than a second threshold. We calculated the local standard deviation for the image by moving a window across the image and by assigning the standard deviation of the pixels in each window to the center pixel of that window. As with the local gradient

weighted average, to reduce the computational cost, the windows did not overlap. Hence the standard deviation was only calculated for the pixels in the center of the non-overlapping windows. We then used a bilinear interpolation to find the local standard deviation for all pixels. The threshold for the local standard deviation was chosen based on the histogram of the local standard deviation values, as the first valley in the histogram. In case of unimodal histograms, the threshold was set to zero. Figure 3.53 shows different steps of the binarization procedure.



3.6.1.2 Edge Relocation

The above initial binarization results in many clustered objects and will not be precise. Next, we apply an implementation of the “edge relocation” algorithm originally developed by MacAulay et al.¹³⁶ The algorithm makes use of intensity information, gradient magnitude information and both object and edge connectivity information and generates a closed contour precisely along the edge of the nucleus. The algorithm is applied to each connected component in the binary image that was generated by the initial binarization. It consists of three main steps:

1) The first step generates a potential “edge mask” around the boundary of the initial connected component, C . This is done by morphologically dilating and eroding C a number of times resulting in $C_{Dilated}$ and C_{Eroded} . The “edge mask” is then calculated as the set of pixels that belong to $C_{Dilated}$ but do not belong to C_{Eroded} . The number of dilations and erosions should be matched to the magnification and size of the objects to be segmented.

2) Pixels that belong to C_{Eroded} belong to the object foreground based on the initial segmentation. Amongst these pixels, those which are connected to the “edge mask” and satisfy the criteria to be a potential edge pixel are added to the edge mask. These criteria are: 1) low intensity and 2) high gradient. This step results in the cluster of objects to be broken apart into smaller objects.

The thresholds for determining low intensity and high gradient pixels may be set in different ways. In this work, the threshold for finding the low intensity pixels was based on the Otsu threshold¹³⁷ for the pixels included in the “edge mask”. This threshold was calculated as follows:

$$T_{Low Intensity} = T_{Otsu} + (P - th \text{ percentile of edge mask intensities} - T_{Otsu})/2 \quad (3.27)$$

The higher the value of P is in equation (3.27), the higher the threshold $T_{Low Intensity}$. Higher threshold results in the addition of more pixels to the edge mask, which increases the possibility of the object being split into more objects. If the object foregrounds are textured and/or noisy, a smaller value of P is preferred because higher P would result in over-

segmentation. We found that 75th percentile was suitable for segmenting distribution maps of Hematoxylin that were obtained from the linear spectral unmixing procedure.

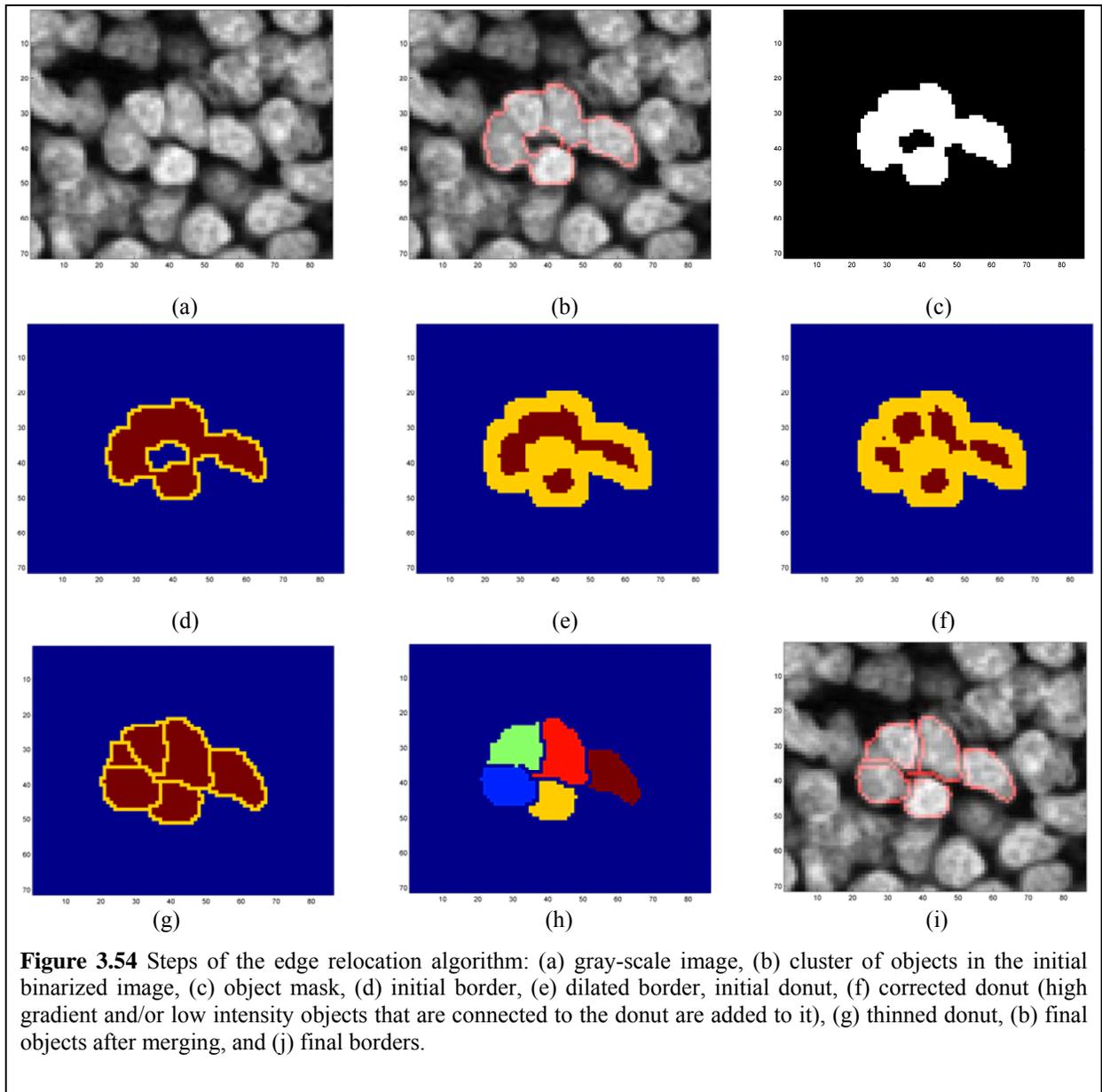
The threshold for determining high gradient pixels was chosen as a certain percentile of the foreground normalized gradient values. Normalized gradient is the gradient divided by the pixel intensity. A 3×3 Sobel filter was used to calculate the gradient. The 60th percentile was used throughout this work.

The set of low intensity pixels determined in this manner are combined with the set of high gradient pixels. Those pixels in this set that are connected to the “edge mask” are then added to the “edge mask”.

3) The algorithm then starts to conditionally “thin” the edge mask by sequentially removing the pixels with low gradient from the mask, provided that the connectivity of the object and the mask are maintained. The exact way this was implemented is as follows: at the beginning of the thinning process a low threshold was chosen for the normalized gradient and those pixels included in the edge mask that touched the outside or the inside pixels were compared against this threshold. Pixels with normalized gradient values below the threshold were considered for being removed from the edge mask. The removal of pixels started from the one with the lowest normalized gradient. The pixel would be removed if its removal preserved the connectivity of the edge mask. When no more pixels below the threshold could be removed, the threshold was raised slightly and the conditional erosion process was repeated. The amount by which the threshold was raised each time was set to one sixteenth of the range of the normalized gradient values in the edge mask. This process continued until the normalized gradient threshold reached the maximum of the normalized gradient values in the edge mask.

The result of this conditional thinning process was: 1) possible dividing of the initial object, if there were closed contours with low intensity or high gradient inside the object and 2) a closed contour around each identified object that followed the path with the highest gradient inside the search area, i.e. “edge mask”.

The above 3-step algorithm was run iteratively until no more objects could be divided further and every object’s boundary had been refined. Figure 3.54 demonstrates this algorithm.



3.6.1.3 Distance-based Watershed Segmentation

After edge relocation, there may still be objects that are likely to be a cluster of smaller objects. In this work, such objects were identified based on their area, i.e. if the object area was larger than a pre-determined value, then it was assumed that it was a cluster of touching objects. More robust measures may be defined and calculated for the identification of such objects. The future versions of the segmentation procedure will take this into account.

A distance-based watershed segmentation¹³⁸ was performed on such objects to possibly divide them further. The procedure was as follows: the binary mask of the object was inverted. A distance transform of the inverted binary mask was then calculated. For each pixel in the image, the distance transform assigned a number that was the distance between that pixel and the nearest non-background pixel. The resulting distance map was inverted, i.e. dark areas in the input image became light and light areas became dark. In the inverted distance image, the object appeared as a hole. A watershed transform then segmented the distance image into watershed regions and their boundaries. Considering the gray scale inverted distance map as a surface, each local minimum can be thought of as a point to which water falling on the surrounding region drains. The boundaries of the watersheds would lie on the tops of the ridges. To avoid over-segmentation, before applying the watershed transform on the distance image, a median filter was applied to the distance image to remove smaller local minima.

3.6.1.4 Merging

After both edge relocation and distance-based watershed transform steps, a merging procedure was performed on the objects that were thought to be over-segmented as follows: 1) if the object size was smaller than a user-determined threshold and it had common boundaries with a bigger object, then that object would be merged into the bigger object. 2) If the ratio of the length of the common boundary of an object A with another object B to the length of the boundary of object A was higher than a threshold, the object A would be merged into the object B.

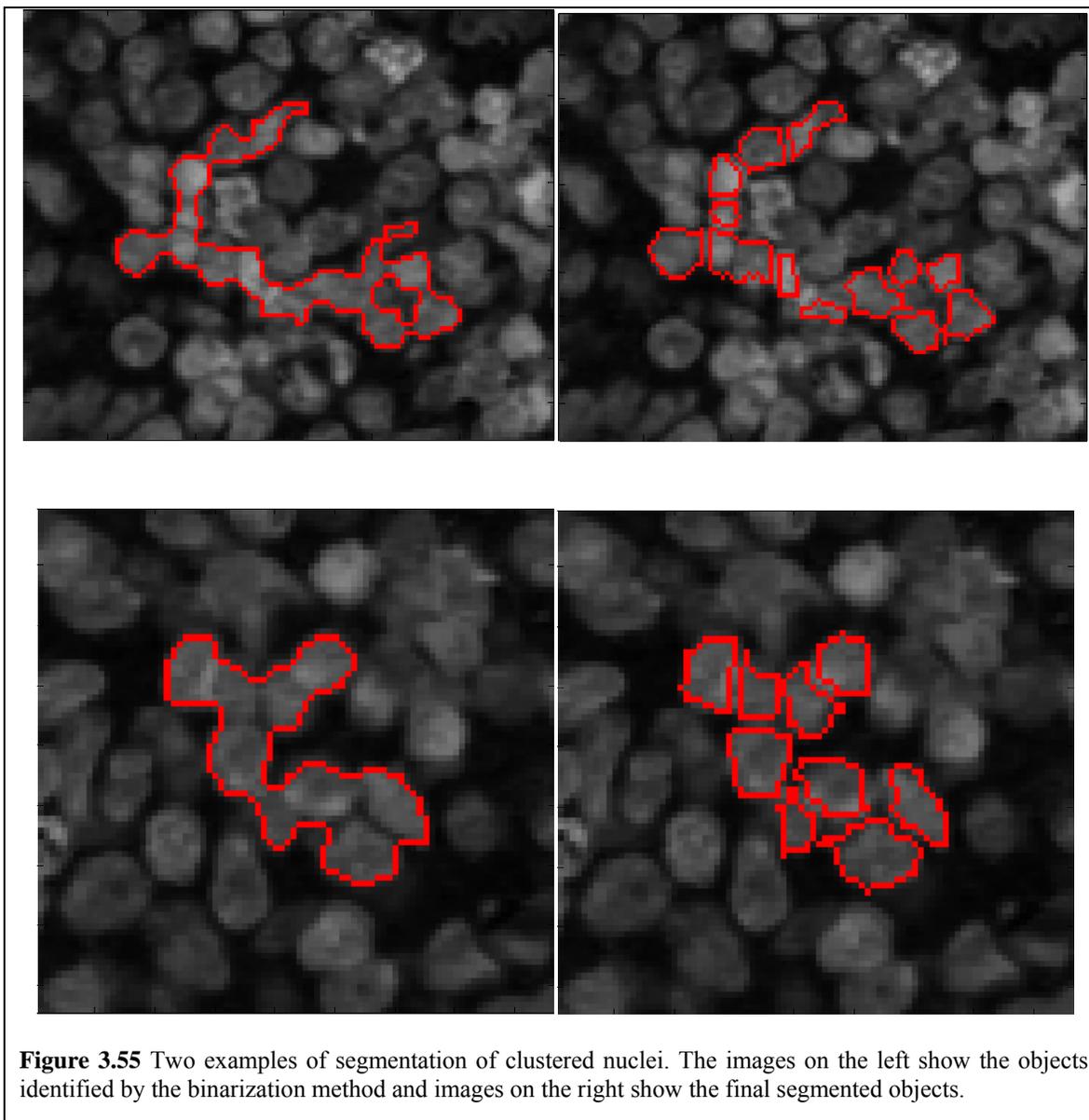
3.6.2 Results

The nuclear segmentation algorithm requires a single grayscale image of nuclei, where the nuclei are brighter than the background. In the previous sections, we explained how a grayscale map of nuclei can be constructed through linear spectral unmixing of 12 channel multispectral image cubes, spectral unmixing of three channel RGB color images, and linear discriminant transforms of either the 12-channel multispectral or three-channel RGB images. Any such grayscale images which highlight the nuclei may be used for segmentation of nuclei.

3.6.2.1 Segmentation Results

Figure 3.55 illustrates two examples of the effectiveness of the segmentation algorithm in separating touching nuclei.

Segmentation of nuclei in the CD8 samples proved to be more challenging. We consider these cases to be of high complexity due to the high number of closely packed nuclei, and occasional high density and unspecific localization of the antibody. This has made the identification and separation of nuclei so difficult that even manual identification of nuclei has proven to be quite difficult and subjective. Neither spectral unmixing nor the linear discriminant functions had resulted in the perfect separation of the labels. So traces of the membrane label remained in the nuclei map. The nuclei in these images were located very close to each other. In the nuclei image, in the places where only the labeled membranes separated several nuclei, incomplete separation of the labels could result in the nuclei to appear as one big cluster of objects. The edge gradients were not strong enough to allow for the separation of touching nuclei. Thus, the segmentation algorithm resulted in less than ideal identification of nuclei.



To improve the segmentation in these cases, we used the information from the immunostain concentration map. Membranes that were positive for CD8 appeared as rings in the immunostain map. Each ring encompassed a nucleus. Thus the rings could be used to separate clusters of nuclei where there was positive staining.

The procedure that we used to employ this information was as follows:

Step 1) nuclei were detected in the nuclei map using the binarization technique of section 3.6.1.1.

Step 2) positive membrane pixels were detected in the immunostain map using the same binarization technique of section 3.6.1.1. The choice of the windows size for the adaptive local thresholding was important. In the detection of nuclei pixels, a window that was about the size of the average nucleus was used. This window size, however, was too big for detecting rings of positive pixels in the immunostain map, as it resulted in detection of rings that were too thick or in circular masks containing the nuclei. The window dimensions should be equal to the approximate thickness of the membrane in pixels. Here, a window of 4×4 pixels was able to detect the rings of positive pixels. Every closed ring detected in this manner was assumed to enclose a nucleus. But the thresholding could also result in incomplete rings that were missing a few pixels. To avoid discarding such incomplete rings, several rounds of morphological image closing were performed and after each round, the binary map was searched for closed rings. After binarization of the immunostain image, the closed rings were detected and their encompassed round objects were considered to be masks of nuclei. Thus in the binary map of nuclei from step 1, the pixels corresponding to the closed ring were set to zero, and the pixels corresponding to the enclosed objects were set to one. This ensured that the enclosed nuclei were included in the binary map and were separated from neighboring nuclei.

Step 3) the refined binary map of nuclei was passed on to the edge relocation algorithm of section 3.6.1.2 for the refined separation and identification of nuclei. Figures 3.56 and 3.57 illustrate these steps.

Typical results of the application of the nuclear segmentation algorithm to nuclei images of Ki67, P16, and CD8 samples are demonstrated in Figures 3.58 to 3.63.

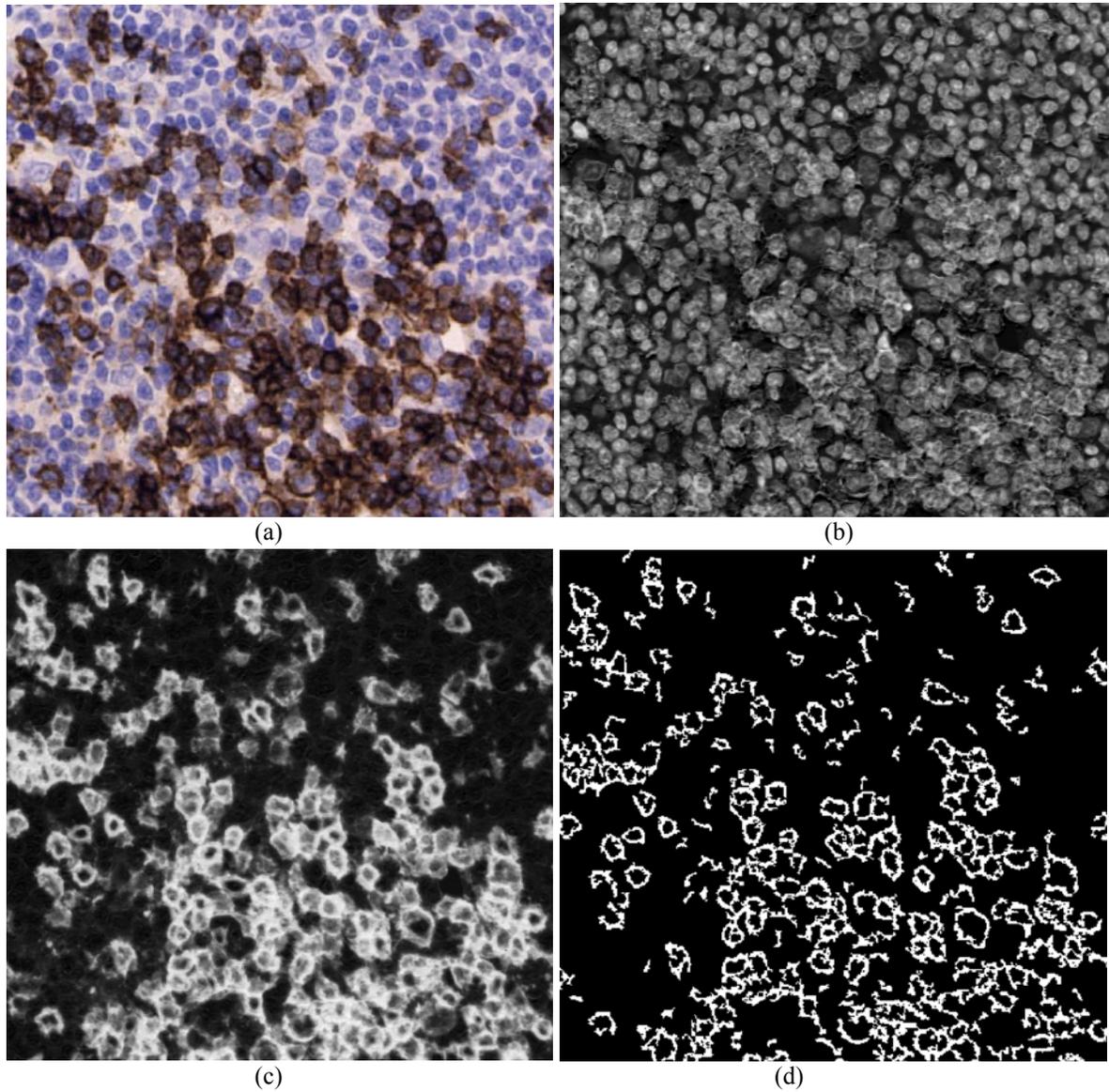


Figure 3.56 Segmentation of a CD8 sample: (a) RGB image, (b) unmixed “nuclei image”, (c) unmixed “DAB image”, and (d) binarized DAB image – continued in Figure 3.56.

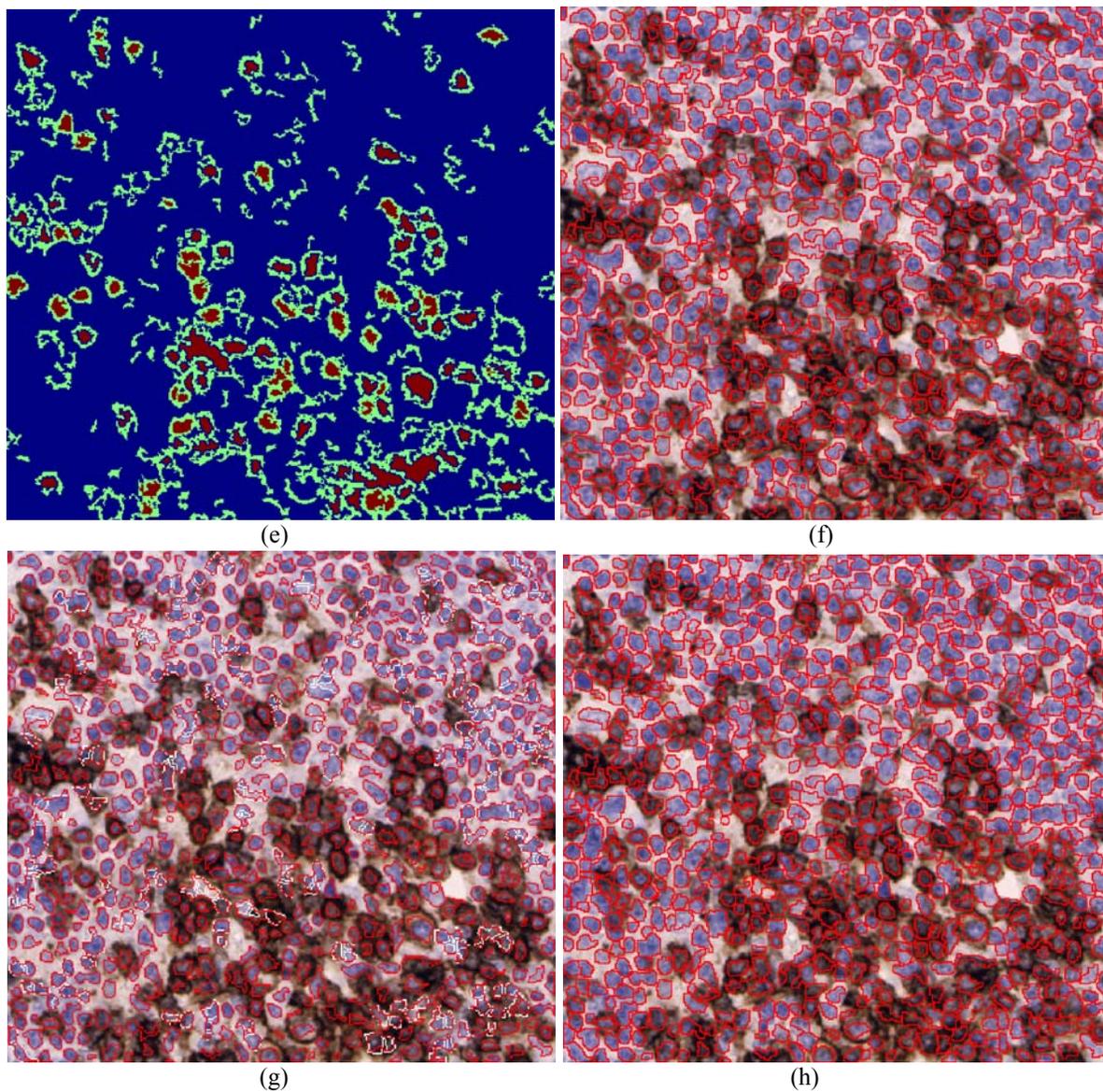
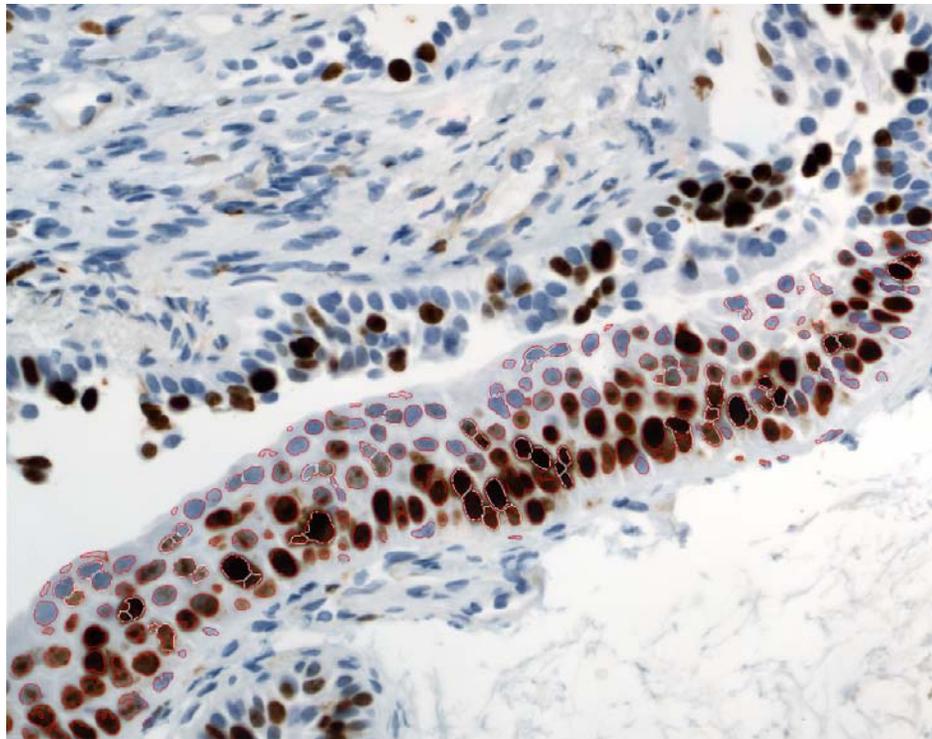
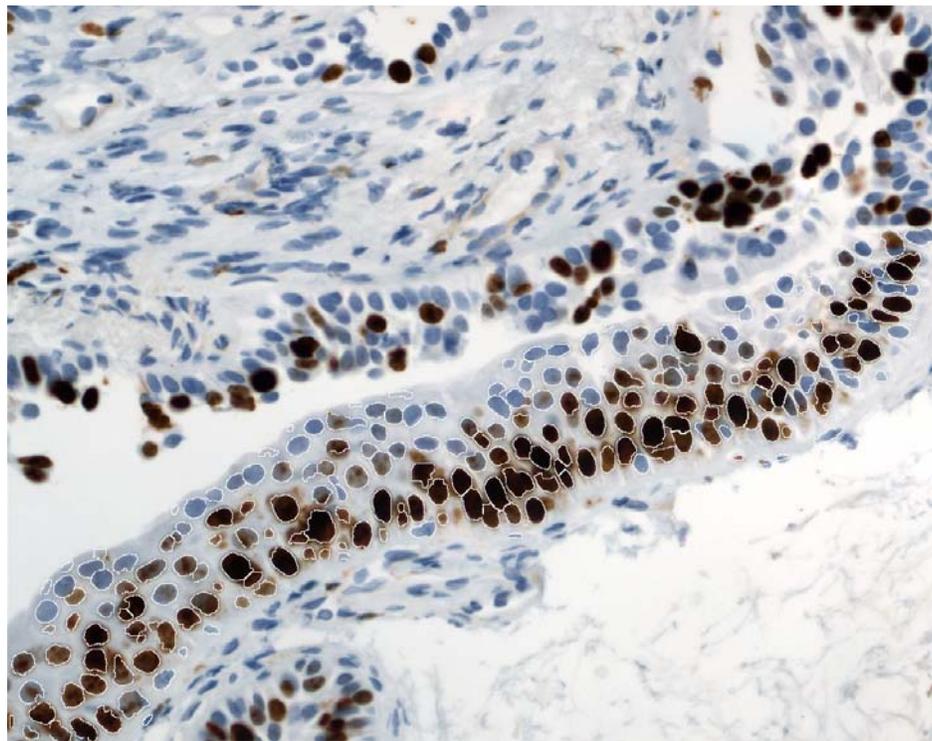


Figure 3.57 Segmentation of a CD8 sample, continued from Figure 3.56, (e) areas in the binarized DAB image that were enclosed by ring objects marked with red, (f) initial binarization result, (g) initial binarization result corrected by incorporating information from (e), and (h) final segmentation result.

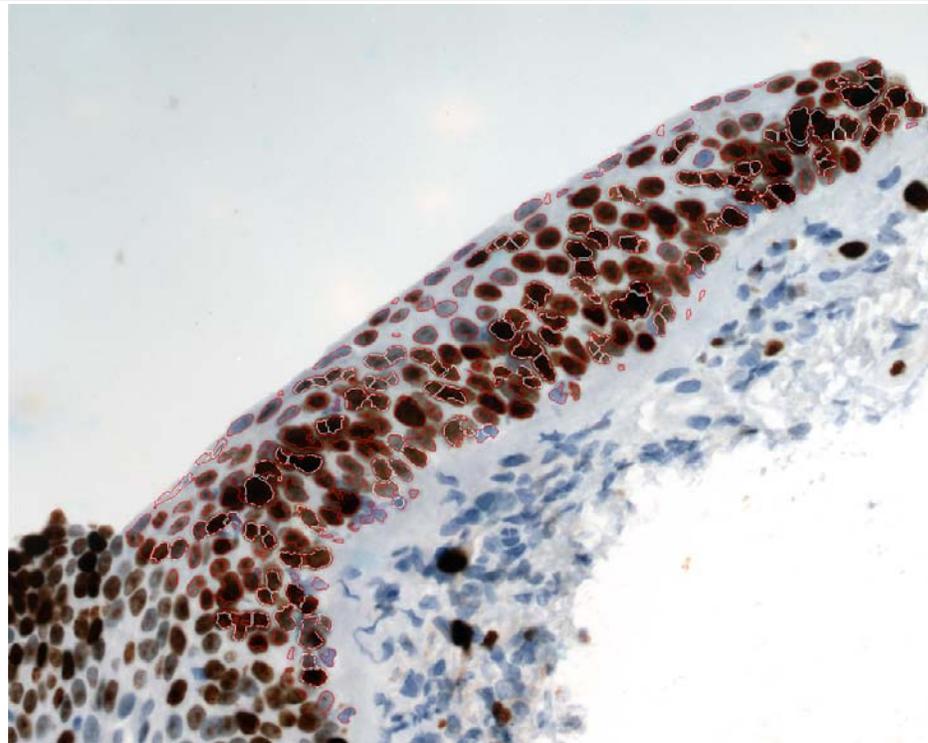


(a)

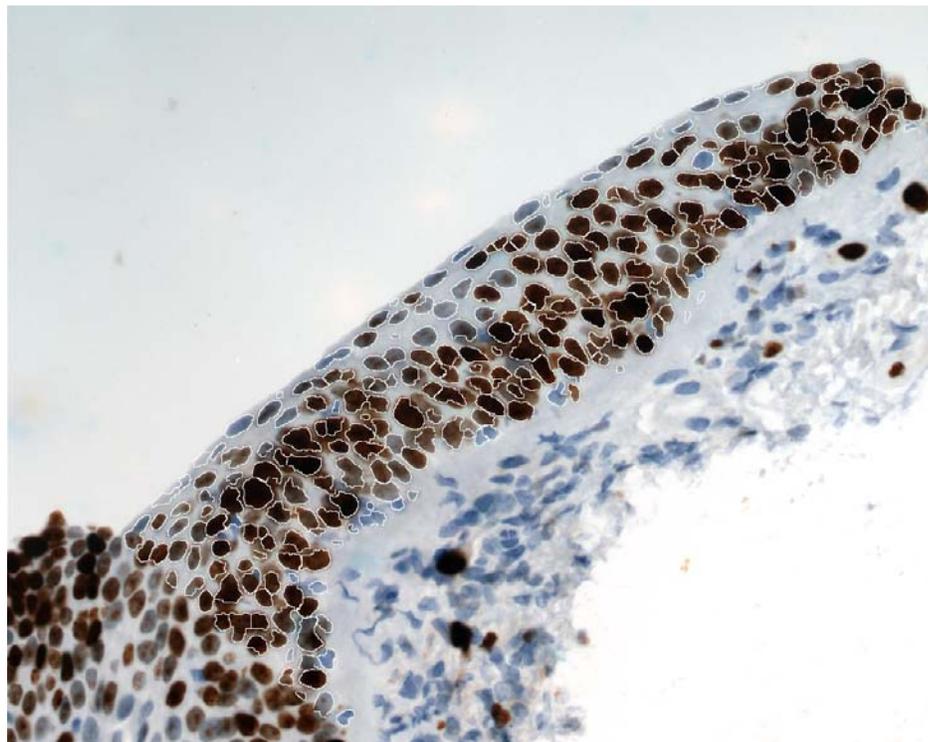


(b)

Figure 3.58 Ki67 segmentation example one: (a) initial binarization result, objects with white boundaries are further separated, (b) final segmentation result

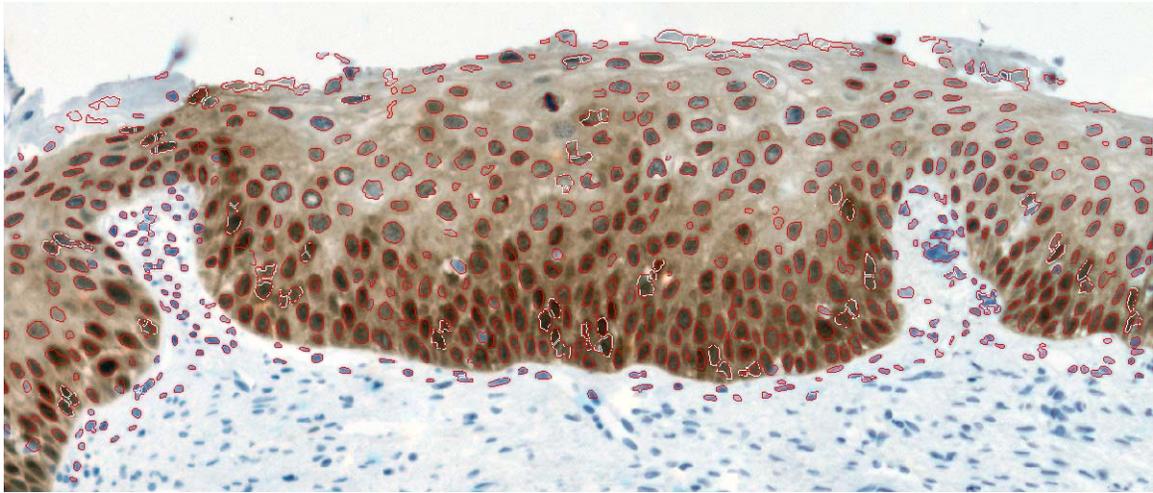


(a)

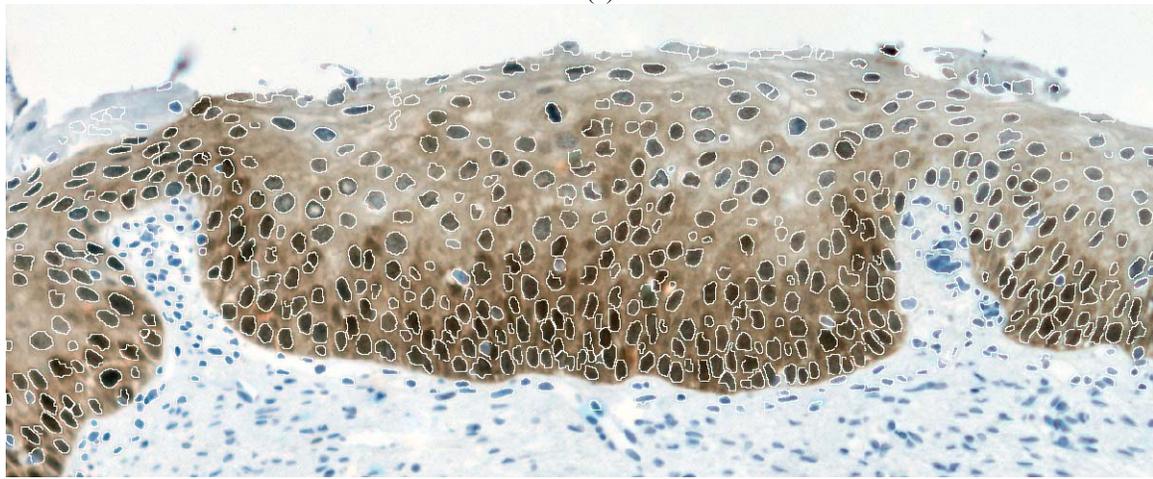


(b)

Figure 3.59 Ki67 segmentation example two: (a) initial binarization result, objects with white boundaries are further separated, (b) final segmentation result.

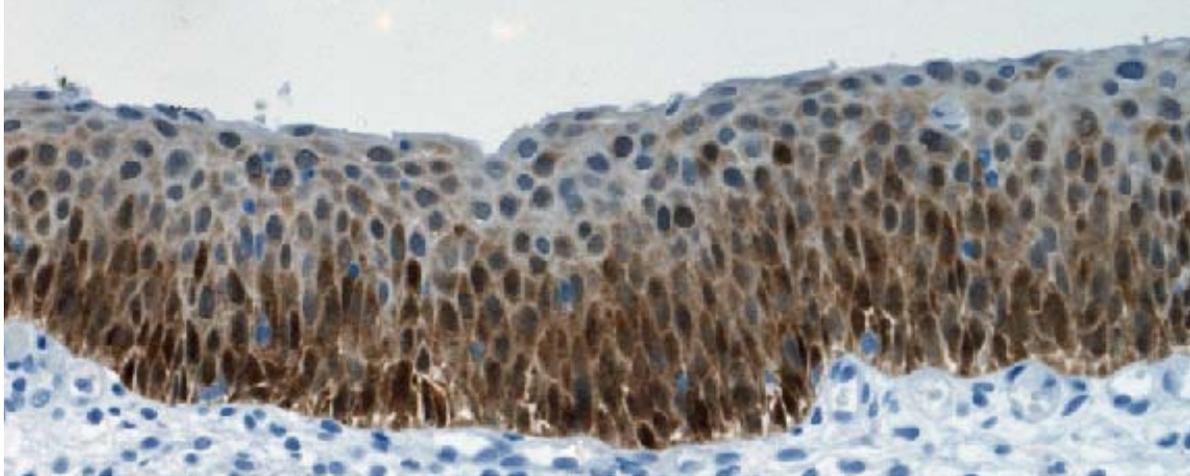


(a)

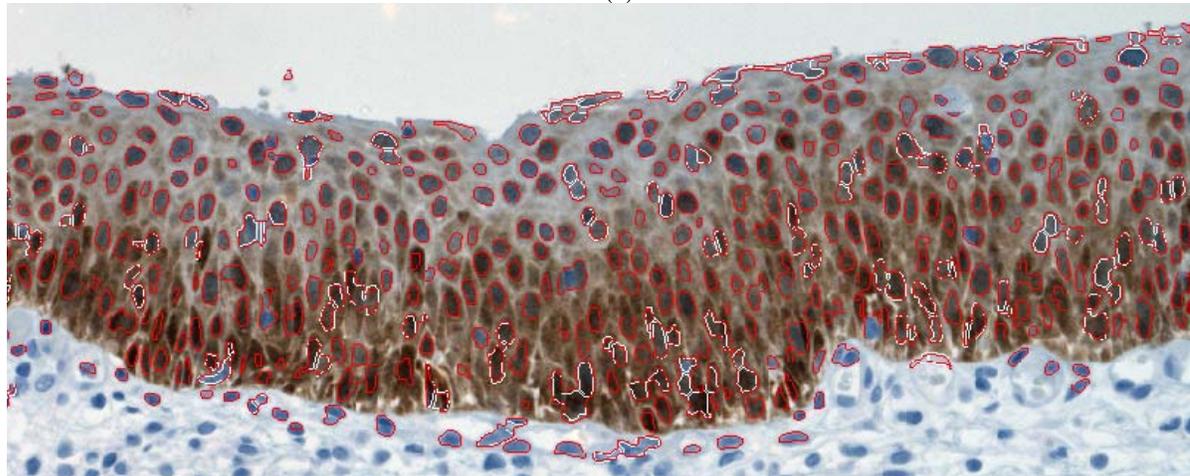


(b)

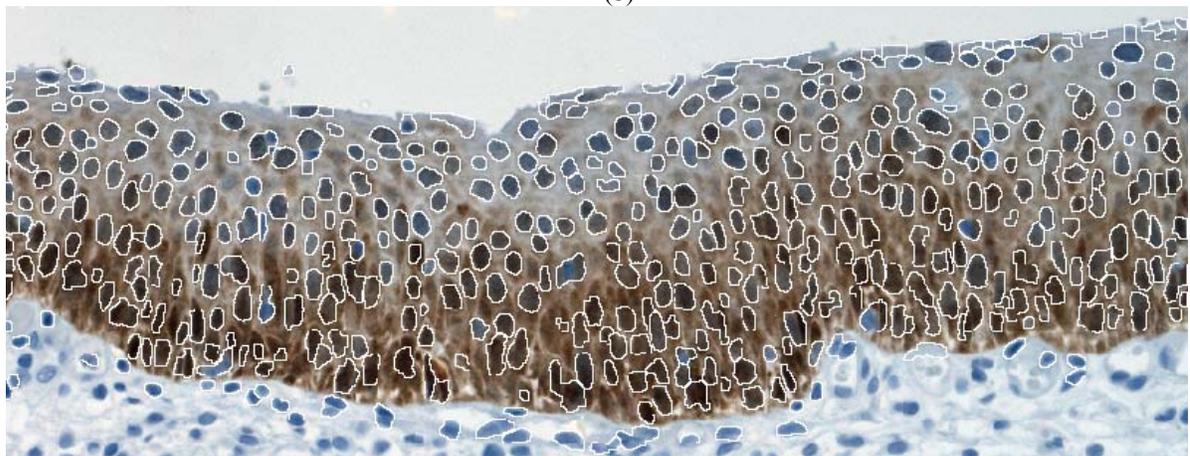
Figure 3.60 P16 segmentation example one: (a) initial binarization result, objects with white boundaries are further separated, (b) final segmentation result.



(a)



(b)



(c)

Figure 3.61 P16 segmentation example two: (a) RGB image, (b) initial binarization result, objects with white boundaries are further separated, (c) final segmentation result.

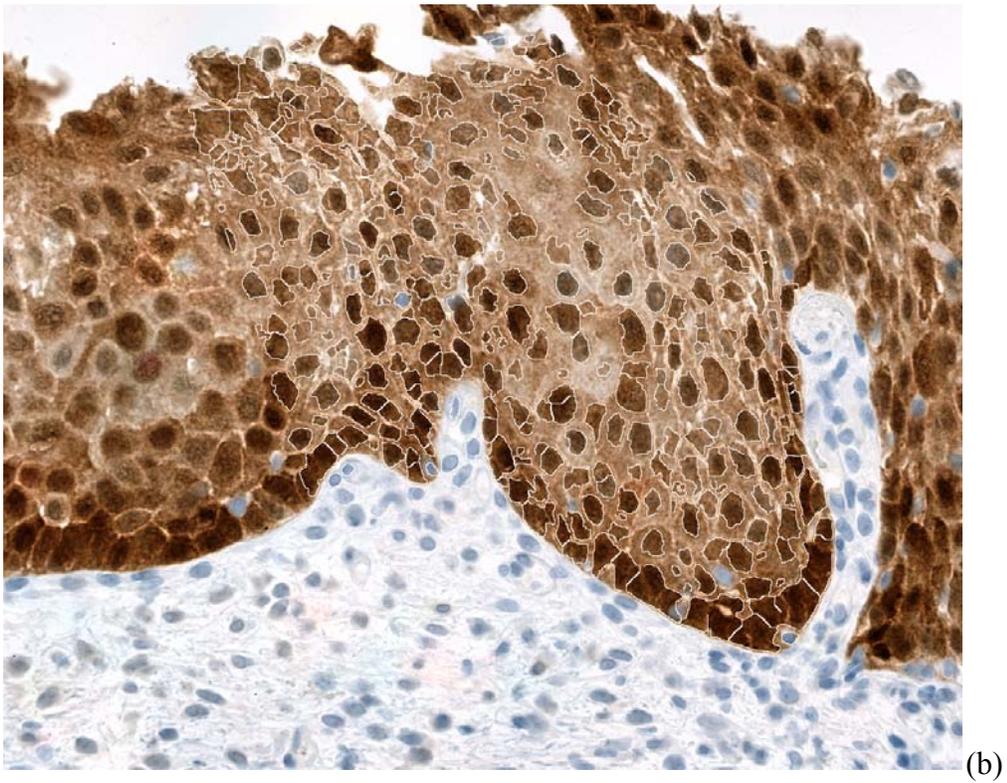
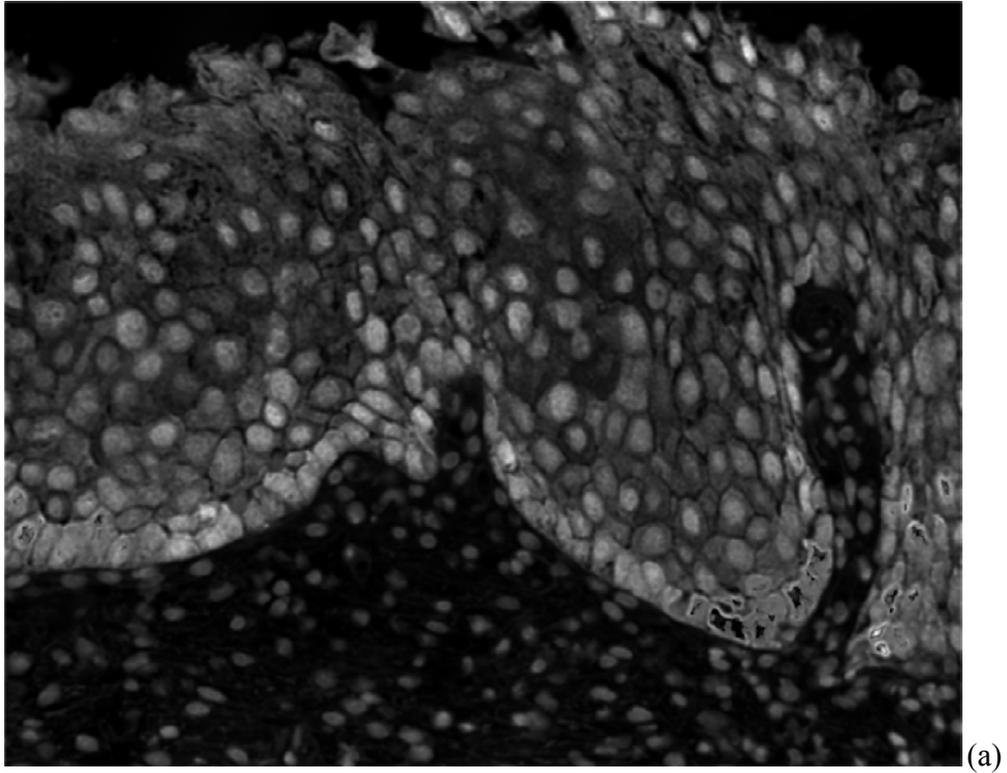


Figure 3.62 P16 segmentation example with dark DAB staining: (a) unmixed nuclei image to be segmented (b) final segmentation result superimposed on the RGB image.

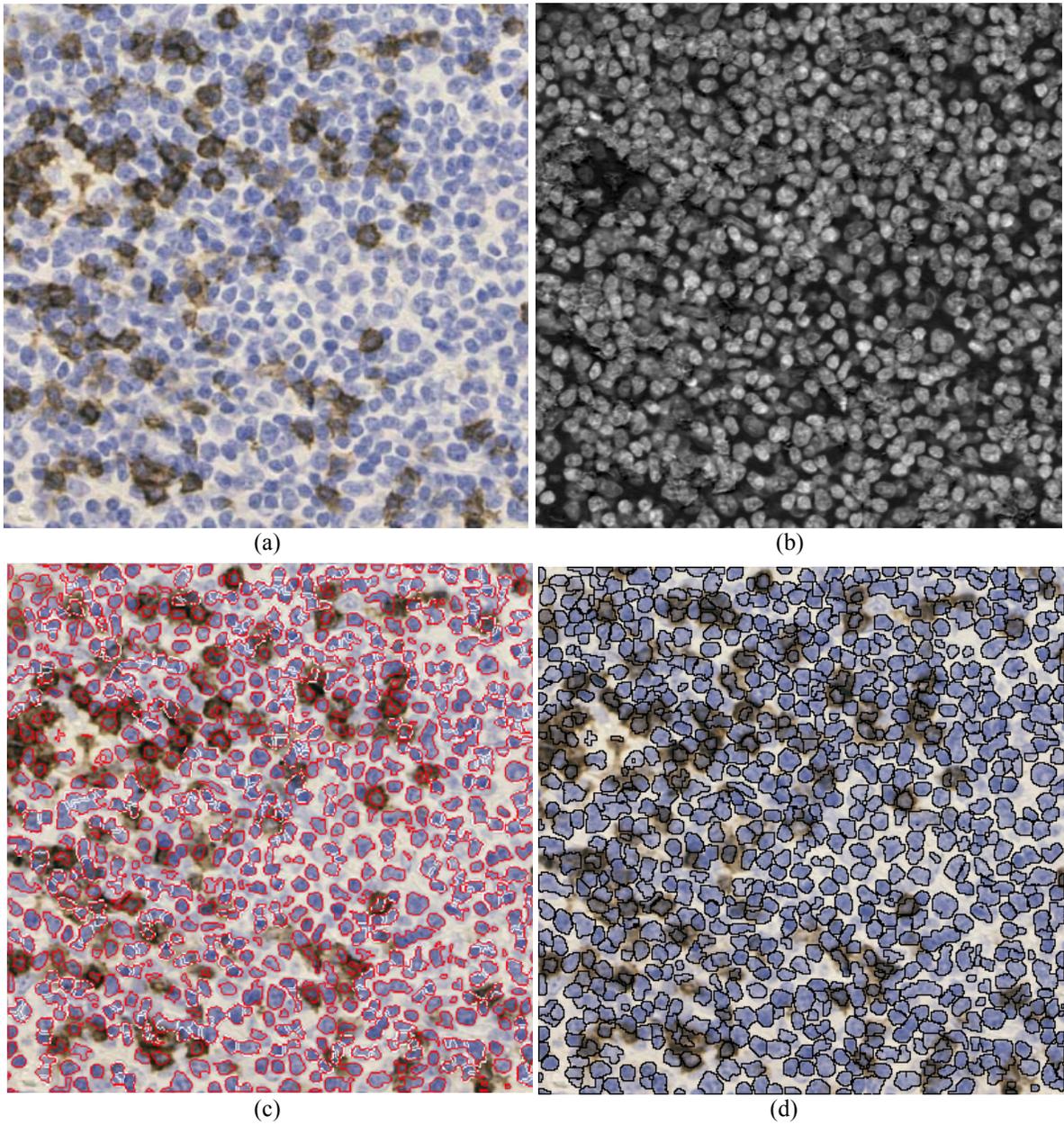


Figure 3.63 CD8 segmentation example: (a) RGB image, (b) unmixed nuclei image, (c) initial binarization result, and (d) final segmentation result.

3.6.2.2 Validation

Manually identified nuclei were the only available gold standard for validating and evaluating the automated analysis. Our validation approach is to have two domain experts manually mark the center of nuclei using computer-assisted annotation.

We chose to mark the center of nuclei instead of delineating the membrane of nuclei, for two reasons: 1) delineation of all nuclei was labour intensive and time consuming and 2) for the purpose of quantifying the amount of immunostain per cell, knowledge of the exact border of nuclei was not critical and minor differences between the actual border, and the automatically generated border were acceptable.

To obtain a measure of inter-observer variability in identifying the nuclei, the nuclei were marked by two expert observers. Firstly one of the observers (observer one) marked the nuclei in all images. By comparing manual segmentation results with automated segmentation results, we noticed that there were several instances where this observer had missed a nucleus. It was only after the automated procedure that we were able to identify that nucleus. Therefore, a second observer (observer two) marked the nuclei by comparing the manually marked nuclei and the automatically segmented nuclei and correcting any manual or automatic segmentation errors. The segmentation results of the second expert observer were used as the reference set. As will be explained later, the segmentation of observer one was later compared against this reference to provide a measure of observer error in manual identification of nuclei.

For comparison, we used several grayscale images as input to the nuclear segmentation algorithm, with the goal of finding the approach that generated the most accurate segmentation. These images were:

- 1) The nuclei map (i.e. the image representing the nuclear label H) generated by the application of the linear spectral unmixing algorithm to the 12-channel multispectral images (“multispectral unmixing”),
- 2) The nuclei map (i.e. the image representing the nuclear label H) generated by the application of the spectral unmixing algorithm to the three-channel RGB images (“RGB unmixing”),

3) The nuclei map (i.e. the image representing the nuclear label H) generated by the application of the LDA algorithm to the 12-channel multispectral images (“multispectral LDA”),

4) The nuclei map (i.e. the image representing the nuclear label H) generated by the application of the LDA algorithm to the three-channel multispectral images (“RGB LDA”),

5) The red component of each RGB image. The red component was chosen since, compared to the blue and green components, it had the highest absorption for H and the lowest for DAB.

For each sample, the accuracy of the segmentation performed on each of the five separately obtained nuclei images was assessed. In evaluating the performance of segmentation, four different segmentation scenarios may happen:

- 1) Correct segmentation: exactly one manually marked nucleus center is inside one automatically generated border,
- 2) Under-segmentation: more than one manually marked nuclei centers are inside one automatically generated border,
- 3) Over-segmentation: no manually marked nuclei centers are inside one automatically generated border,
- 4) Missed nucleus: a manually marked nucleus center that is not inside any of the automatically generated borders.

Figure 3.64 demonstrates an example of these segmentation errors. These four scenarios were automatically counted for each segmented image, resulting in four metrics for every automatically segmented image:

N_1 = number of correctly segmented nuclei,

N_2 = number of under-segmented nuclei,

N_3 = number of over-segmented nuclei, and

N_4 = number of missed nuclei.

To combine all above four metrics into one metric that facilitates comparison of different approaches, we used the F-score¹³⁹, which is the harmonic mean of precision and recall and is calculated as follows:

$$\begin{aligned}
 \textit{precision} &= \frac{\textit{true positive}}{\textit{true positive} + \textit{false positive}} = \frac{N_1}{N_1 + N_3} \\
 \textit{recall} &= \frac{\textit{true positive}}{\textit{true positive} + \textit{false negative}} = \frac{N_1}{N_1 + (N_2 + N_4)} \\
 F &= 2 \cdot \frac{\textit{precision} \cdot \textit{recall}}{\textit{precision} + \textit{recall}}
 \end{aligned}
 \tag{3.28}$$

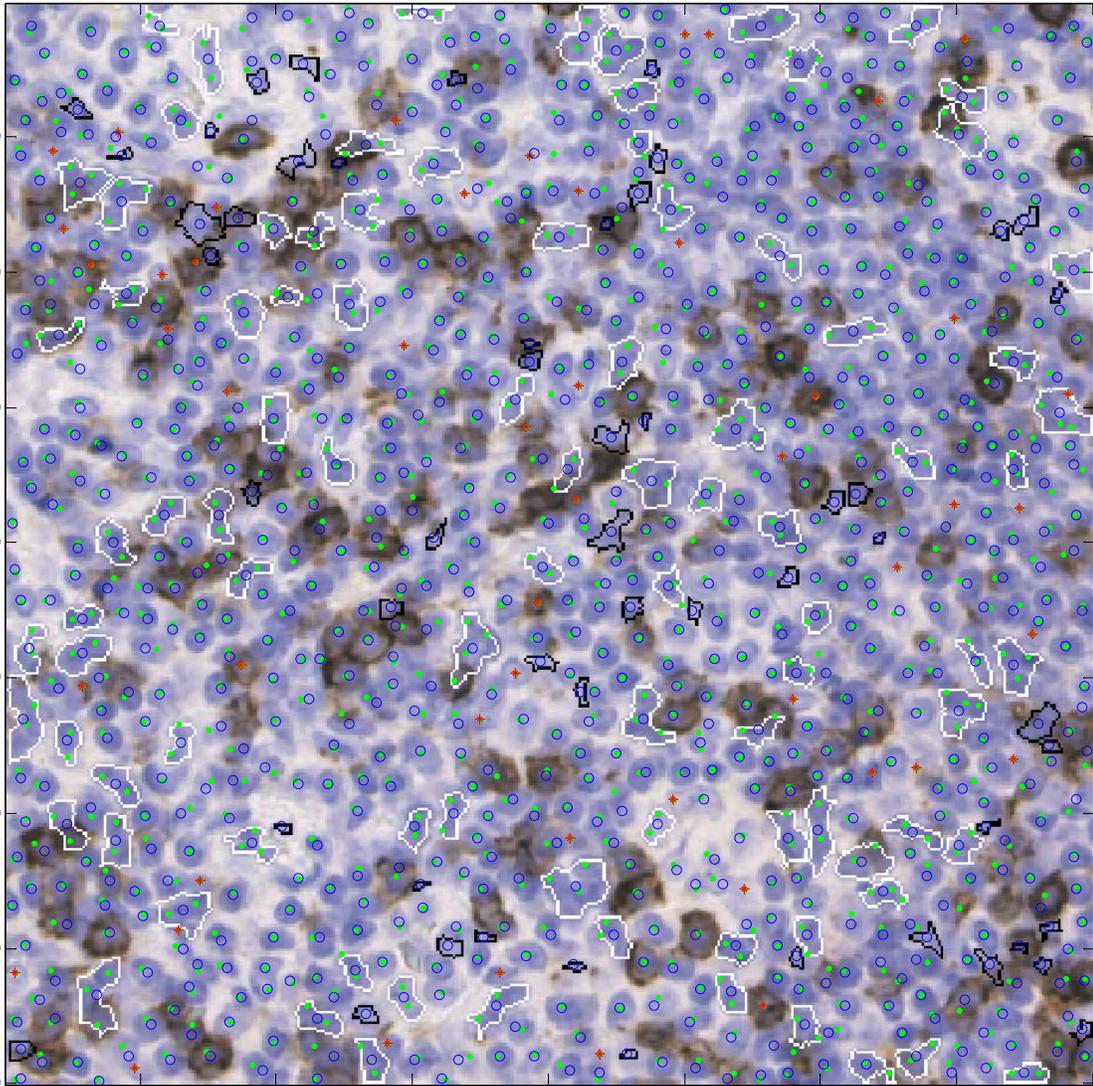


Figure 3.64 Example of evaluating the segmentation accuracy, green dots are manually marked centers of nuclei, blue circles are centers of automatically segmented nuclei, objects with white borders are under-segmented objects, objects with black borders are over-segmented or falsely detected objects, and red stars are missed objects.

To estimate the degree to which the manual cell identification was different from one observer to the other, we used the nuclei marked by observer two as the reference and compared the manual results of observer one against it. The observers had only marked the centers of nuclei, so the boundaries of nuclei were unknown in the manual segmentation results. Without the boundary of nuclei, the segmentation metric based on the four scenarios explained above could not be calculated. As a workaround, we generated a Voronoi tessellation map based on the centers of nuclei marked by observer one. Each Voronoi polygon was considered the nucleus boundary for the corresponding nucleus. We realize that the Voronoi polygon encompasses a larger area than the cell nucleus. However, for the purpose of evaluating the manual segmentation results, if a manually marked nucleus centroid was inside the Voronoi polygon of that nucleus, we considered that nucleus correctly marked. Using the Voronoi polygon as the nucleus boundary, three scenarios may happen:

1) Correctly marked nucleus: if there is only one reference nucleus within a polygon, that nucleus is identified correctly.

2) Missed nuclei: if there are more than one reference nuclei within a polygon, all but one of those reference nuclei are counted as missed nuclei.

3) Falsely detected nucleus: if there are not any reference nuclei within a polygon, that polygon's corresponding nucleus is a false positive.

These three scenarios resulted in the following three metrics for every manually marked image:

N_1 = number of correctly marked nuclei,

N_2 = number of missed nuclei (false negatives), and

N_3 = number of falsely marked nuclei (false positives).

Precision, recall, and F-score were calculated in the same way they were calculated for automatic segmentation results:

$$\begin{aligned}
 \textit{precision} &= \frac{\textit{true positive}}{\textit{true positive} + \textit{false positive}} = \frac{N_1}{N_1 + N_3} \\
 \textit{recall} &= \frac{\textit{true positive}}{\textit{true positive} + \textit{false negative}} = \frac{N_1}{N_1 + N_2} \\
 F &= 2 \cdot \frac{\textit{precision} \cdot \textit{recall}}{\textit{precision} + \textit{recall}}
 \end{aligned}
 \tag{3.29}$$

Figure 3.65 shows an example of how the manually marked nuclei by observer one were compared against the reference, i.e. manually marked nuclei by observer two, using this approach.

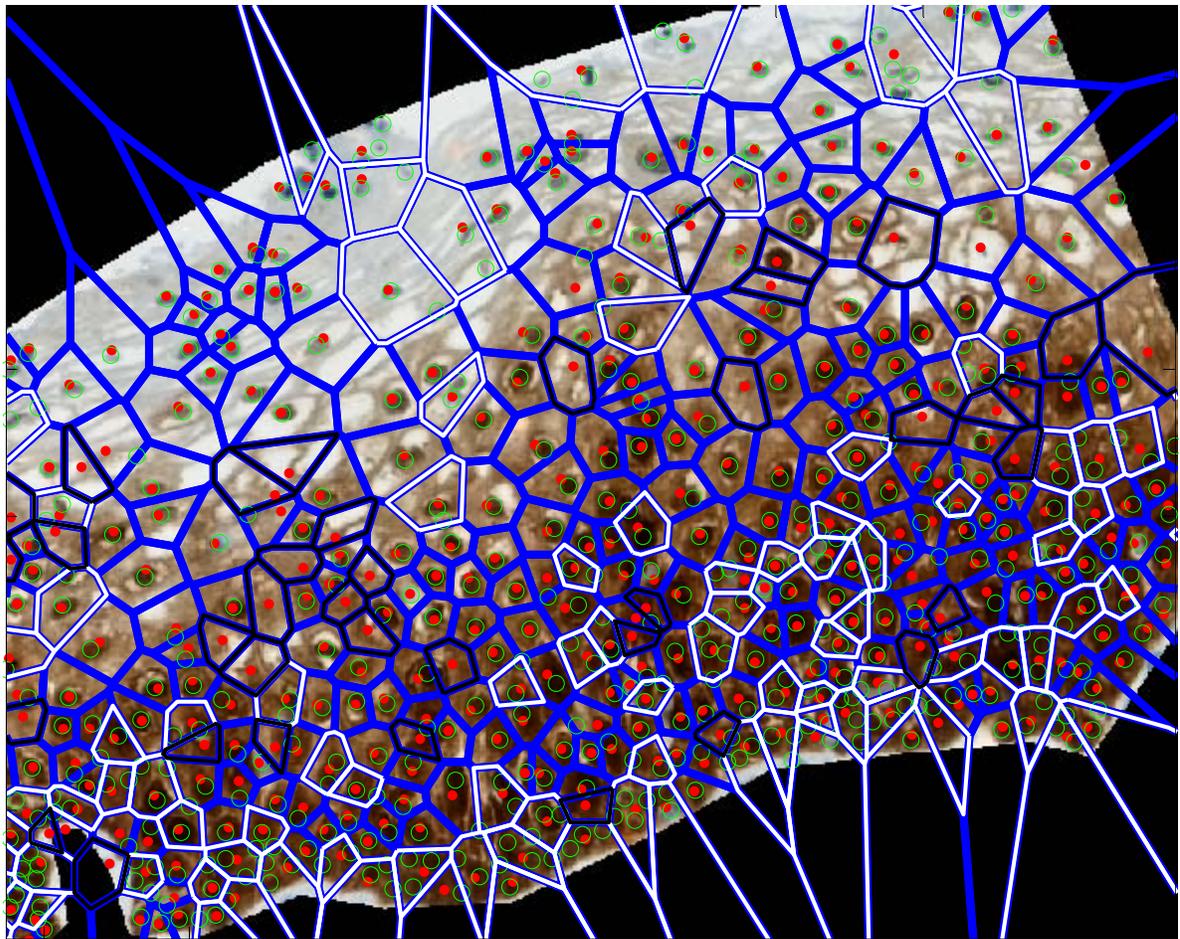


Figure 3.65 Example of how nuclei marked by observer one were compared against the reference, i.e. nuclei marked by observer two. Green circles are nuclei centers marked by observer two (the reference). Red dots are nuclei centers marked by observer one. Polygons represent the Voronoi tessellation based on the nuclei centers identified by observer one. Nuclei inside blue polygons are correctly identified. White polygons contain missed nuclei (false negatives). Dark blue polygons contain falsely detected objects (false positives).

Figures 3.66 to 3.68 present the F-scores for the five different images generated as explained above and for the manual identification of nuclei performed by observer one, for samples in Ki67, P16, and CD8 sets. The percentages of missed, under-segmented, over-segmented, and missed nuclei for five automated and one manual nuclei identification methods for all samples are presented in Appendix B.

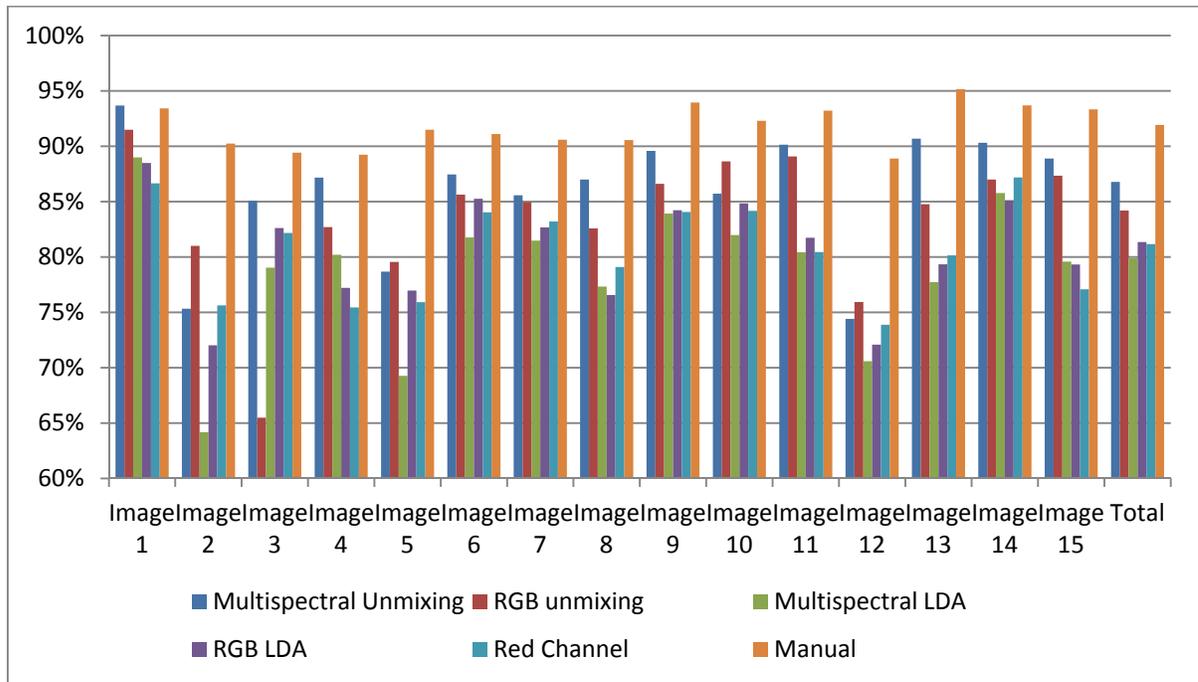


Figure 3.66 Ki67 segmentation: F-scores of segmentation results.

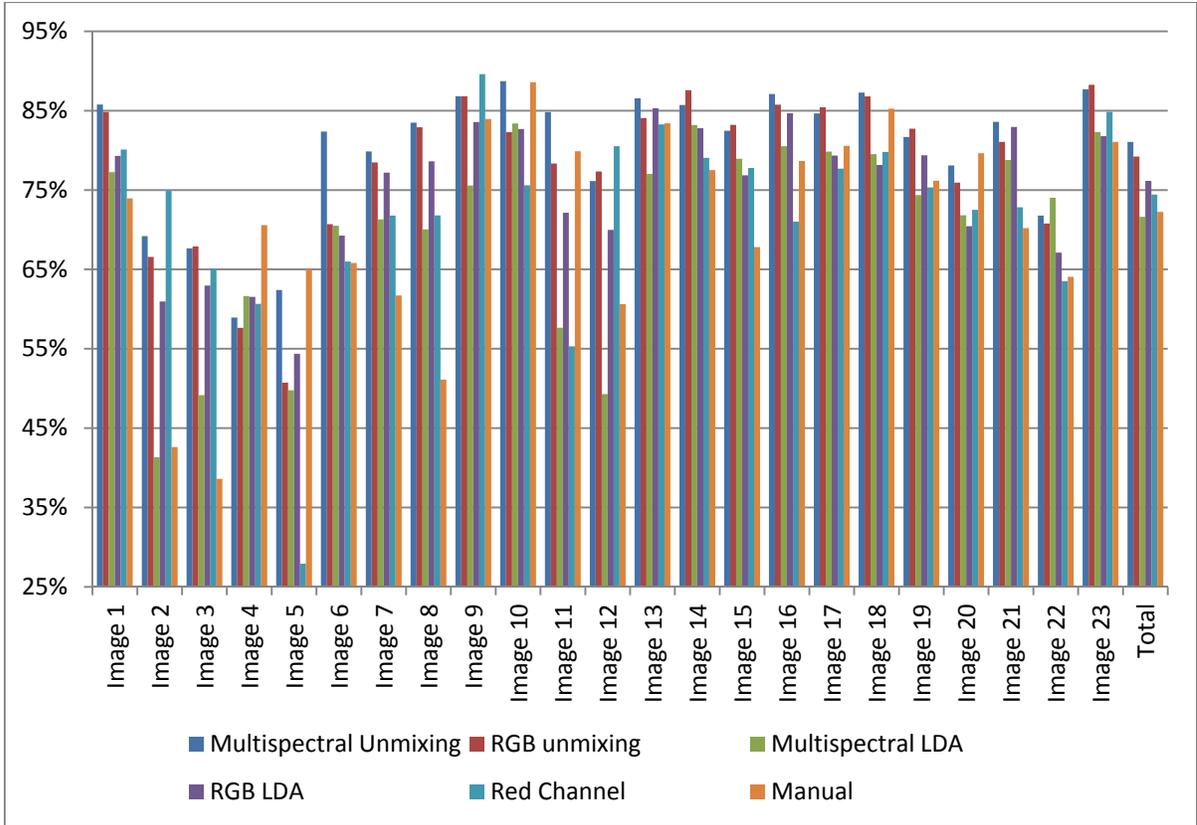


Figure 3.67 P16 segmentation: F-scores of segmentation results.

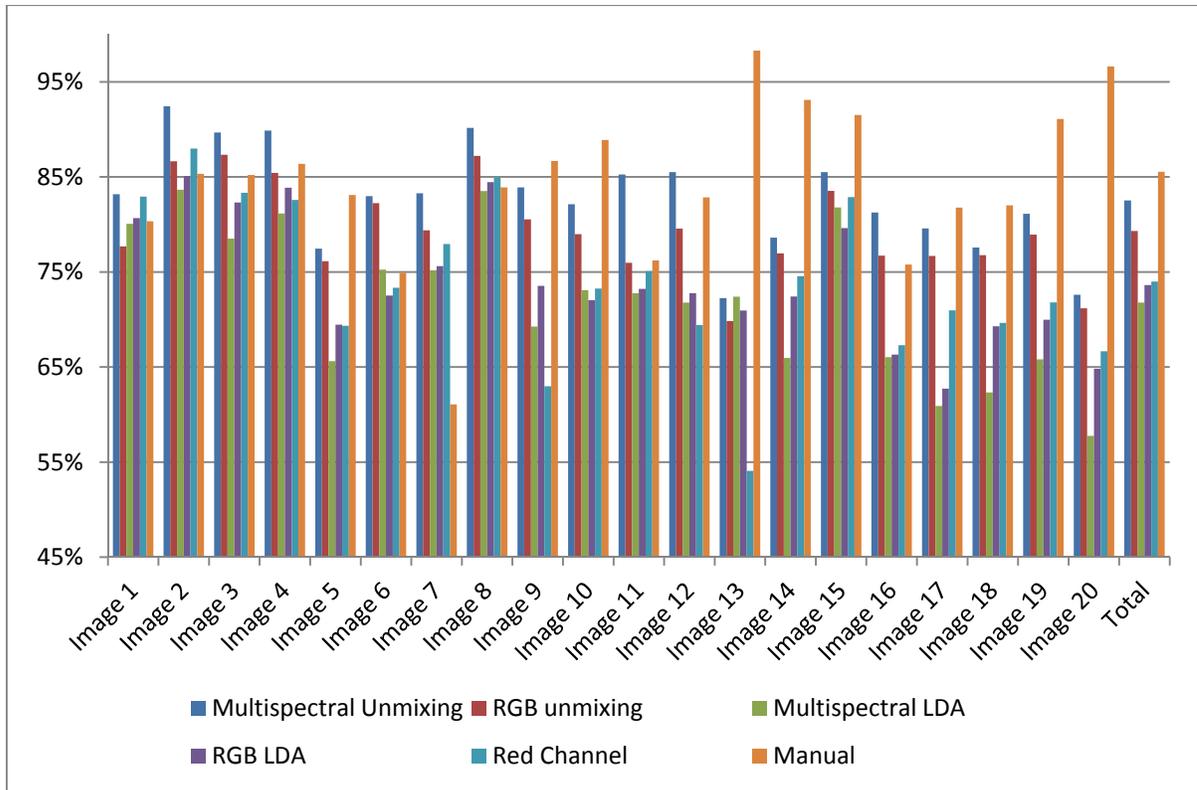


Figure 3.68 CD8 segmentation: F-scores of segmentation results.

3.6.3 Discussion

An automatic nuclear segmentation procedure for the identification and delineation of labeled nuclei in grayscale images of histopathological tissue section was presented. The procedure is based on the edge relocation algorithm of MacAulay et al.¹³⁶ It starts with an image binarization step to separate objects from the background. An edge relocation algorithm is then applied to refine the boundaries and separate touching or clustered nuclei. This may be followed by a distance-based watershed segmentation to further separate clustered objects. A merging procedure based on 1) the size of the objects and 2) the intersection of boundaries of neighboring objects is used to merge back those objects that are over-segmented.

In the previous two sections (3.4 and 3.5), we presented four approaches for generating a grayscale image representation of the labeled nuclei. There four approaches were: 1) linear

spectral unmixing performed on images in the multispectral image stack, 2) linear spectral unmixing performed on the RGB images, 3) LDA performed on multispectral images, and 4) LDA performed on RGB images. A fifth representation of labeled nuclei that was studied was the red channel of the RGB images. The red channel was chosen because it provided the highest contrast between the blue counterstain, Hematoxylin, and the brown chromogen, DAB. The segmentation procedure was performed on each of these five images representing the labeled nuclei, to find the one that resulted in the most accurate identification of nuclei.

All images of the samples within this study were manually segmented by two observers, to provide a reference set for the evaluation of automated segmentation results. After observer one had manually marked the nuclei within the image of each sample, observer two reviewed the results and compared them against the automatic segmentation results, to generate a final annotated set. The annotation by observer two was used as the reference, against which the annotation of observer one was later compared. This provided a measure of human error in the identification of nuclei.

The segmentation accuracy measures of the five different approaches plus the manually annotated results were all calculated by comparing the results against the reference set. We used the F-score for the comparison of different approaches. The results for each of the three sample sets: Ki67, P16, and CD8 are presented in Figures 3.66 to 3.68. Overall, in all the three sets, spectral unmixing (performed on the multispectral image stacks) resulted in the most accurate segmentation, in terms of the highest F-score. Spectral unmixing performed on the RGB images resulted in the second best segmentation. LDA performed on RGB images and the individual red channel resulted in roughly the same F-scores, on average. LDA performed on multispectral images resulted in the lowest F-scores. This could be due to two factors: 1) multispectral images generally have lower signal to noise ratios, due to the narrow illumination spectra, which results in longer exposure times. In addition, with multiple image captures, the detector readout noise is multiplied. 2) Because there were only two labels in the samples, the use of all 12 channels of multispectral images may have resulted in over-training of the linear discriminant functions. Therefore optimal contrast between the nuclei and background might not have been achieved, resulting in less accurate segmentation results.

The segmentation accuracy results may not appear very high initially. But these results should be evaluated keeping in mind the complexity and variable nature of histopathological tissue sections. This is reflected in the accuracy measures of the manual annotation. The manual segmentation results are less than optimal, especially in the more complex images of P16 and CD8. In P16 samples, the best of the five automatic segmentation results outperforms the manual segmentation by a margin of more than 8% in the F-score. The low F-scores of the manual nuclei identification results represent the low agreement between the two observers and the high human error in manual identification of nuclei in histopathological samples.

The differences between the F-scores of different approaches may not be very high. But it is important to remember that all different approaches perform equally well on the majority of the nuclei. Those complex cases that distinguish among the different approaches do not comprise a high percentage of the nuclei in the images.

The segmentation procedure was performed on all images within the same sample set, consisting of about 3,000 nuclei in the Ki67 samples, 10,000 nuclei in the P16 samples, and 9,000 nuclei in the CD8 samples, using the same set of parameters and with no user interaction. This demonstrates its robustness and suitability for use in high-throughput settings with minimal user intervention.

The presented segmentation procedure did not include any quality control steps. To make it more robust, the procedure would benefit from adding a quality control step in which the quality of the segmentation of each object is evaluated. A quality metric can be defined using a series of features that explain the shape of a segmented object. Such quality metric would identify well segmented objects and would also provide a measure of reliability for the segmentation of each object, so that the segmentation can be reversed or redone, if necessary.

3.7 Quantification of Biomarker Expression

3.7.1 Methods

In this section, we present methods for the automated measurement of the abundance of three IHC biomarkers. These three biomarkers are different with respect to the cell compartment in which they are expressed. One is bound to the nuclei, one to both the nuclei and cytoplasms, and the last one to the cell membranes. All three types are labeled with DAB for the visualization of the immunohistochemical biomarker and with H to counterstain the nuclei. Our approach is to obtain the distribution map of each label, referred to as I_{Nuclei} and $I_{Immunostain}$. In sections 3.4 and 3.5, we presented several methods for obtaining these distribution maps. In section 3.6, we presented a segmentation procedure for the identification of nuclei in the I_{Nuclei} image. After identification of all the nuclei, the IHC biomarker is measured for every cells associated with an identified nucleus. The cells can then be classified as positive or negative, based on the presence or absence of the IHC biomarker.

In the case of a nucleus-bound IHC biomarker, the task is straightforward: once the nuclei in the distribution map of the nuclei label I_{Nuclei} are segmented, a binary “nuclear map” is obtained. In the binary map, every nucleus is associated with a mask, where the nuclei pixels are represented by 1’s and the background pixels by 0’s. By superimposing the binary nuclear map on the immunostain distribution map $I_{Immunostain}$, the amount of the immunostain contained in each nucleus mask can be calculated based on the intensity of $I_{Immunostain}$ contained in the nucleus mask.

In the case of cytoplasmic or membrane biomarkers, a mask for the particular cell compartment (cytoplasm or membrane) is necessary to quantify the amount of the immunostain associated with each cell. Since the cytoplasms or the membranes are not specifically labeled, the exact boundaries are not identifiable. If a cell is positive for the biomarker then the cytoplasm or the membrane will be labeled by the immunostain, thus they can be delineated. However, not all cells are positive for the immunostain, so delineating the immunostained membrane or cytoplasm does not identify all cells.

In the absence of a specific label that marks the membrane or surface of every cell in the region of interest in the tissue, we resort to inferring the approximate boundary of the cells

from the position of the nuclei. Since nuclei are generally located in the middle of the cells, we can assume that the surface of each cell includes those pixels that are closest to the nucleus of that cell than to the nuclei of all other cells.

We used two tessellations approaches that partition the image into regions including points that are closest to the nuclei:

1) Voronoi tessellation⁶¹ based on the “centers” of nuclei, resulting in tessellation of image into regions corresponding to the nuclei. In the region corresponding to each nucleus, each point is closer to that nucleus center than to any other nuclei centers.

2) A tessellation based on the minimum distance from the boundary of the nuclei. In this approach the image is partitioned into regions in which every point is closest to the corresponding nucleus boundary (as opposed to the nucleus centroid, which was the case with Voronoi tessellation).

3.7.1.1 Voronoi Tessellation

Voronoi tessellation¹⁴⁰ based on the center of nuclei can be used as an approximation of the cell boundaries. This approach was applied by Guillaud et al.¹⁴¹ to human breast cancer sections immunohistochemically labeled for bcl-2 protein and counterstained with nuclear methyl green stain. In their study, they demonstrated that analysis of the mean staining densities correlated well with the visual scoring performed by two independent pathologists.

Let $P = (p_1, p_2, \dots, p_n)$ be a set of points in a two-dimensional Euclidean space, each point being referred to as a “site”. Each coordinate within the specified region of interest is then assigned to its nearest site. The set of coordinates associated with site p_i thus forms the Voronoi region $V(p_i)$. Each region $V(p_i)$ therefore consists of all coordinates at least as close to p_i as to any other site. $V(P)$ thus forms a tessellation of the space according to the positions of the P sites. Each site has an associated Voronoi polygon which represents the area of influence of that site.

The prerequisite information for constructing the Voronoi tessellation was the knowledge of a defined region of interest (ROI) and a series of sites within this region. Centroids of cell nuclei were used as the sites. Those sites that lie on the convex hull of P have unbounded Voronoi polygons, i.e. one of the vertices of their Voronoi polygon lies in infinity. These are called marginal polygons. In this work, each marginal polygon was replaced with the region

of overlap of that polygon with the ROI. The Voronoi tessellation was performed using the Convex Hull implementation¹⁴² in MATLAB R2012a (The Mathworks Inc.).

3.7.1.2 Nearest Neighbour Tessellation

In this approach, each coordinate within the specified region of interest is assigned to a nucleus whose boundary is closest to that coordinate.

This tessellation was performed using the implementation of the Nearest Neighbour Transform¹⁴³ in MATLAB R2012a (The Mathworks, Inc.). The required information for constructing the nearest neighbour map is the binary map of the segmented nuclei.

3.7.1.3 Classification of Cells

Using either of the tessellation methods explained earlier, the image is partitioned into several regions, where each region corresponds to a nucleus. We use these regions as approximations of cell's surfaces. The abundance of the IHC biomarker within each cell can then be estimated.

The integrated intensity of the immunostain within each cell normalized by the immunostained area within that cell can be used as the measure of the abundance of the IHC biomarker within that cell. The abundance of the IHC biomarker within each cell will be used to classify that cell as negative or positive with regard to the IHC biomarker of interest.

We define the region of interest (ROI) of each cell as the region or compartment in the cell in which the IHC biomarker of interest is expressed. The amount of immunostaining for each cell is measured inside the ROI of that cell.

Let $I_{Immunostain}$ be the immunostain image and ROI_i be the ROI of the i -th cell. For the nucleus bound Ki67 biomarker, the ROI of each cell is its nucleus. Let N_i be the nucleus of the i -th cell:

$$ROI_i = \{p: p \in N_i\} \tag{3.30}$$

where p represents a pixel the immunostain image.

P16 is expressed in both nuclei and cytoplasm; therefore the ROI for each cell is its entire Voronoi polygon. Let V_i be the Voronoi polygon of the i -th cell:

$$ROI_i = \{p: p \in V_i\} \quad (3.31)$$

In the CD8 images in our study, the Voronoi polygon corresponding to each positive cell contains a blue nucleus, and a brown region around the nucleus which is the labeled membrane. The Voronoi polygon corresponding to the negative cells consists of the blue nucleus, and a white region around it. Thus membrane pixels are those pixels inside the Voronoi polygon that are not in the nucleus mask of that polygon. Therefore the ROI is determined as:

$$ROI_i = \{p: p \in V_i, \text{ and } \sim(p \in N_i)\} \quad (3.32)$$

Once the ROI corresponding to each cell within the image is determined, the abundance of the IHC biomarker for the i -th cell S_i is calculated as follows:

$$S_i = \frac{\sum_{p \in ROI_i} I_{Immunostain}(p)}{Area(ROI_i)} \quad (3.33)$$

where $Area(ROI_i)$ is the number of pixels in ROI_i .

The measured immunostain intensity in the ROI of each cell S_i can be used to classify the cells into positive and negative for the IHC biomarker of interest. If the sample contained both positive and negative cells, with neither group dominating, then an automated thresholding technique could be used to find the optimal threshold to separate the two groups based on the measured immunostain. However, in many cases, most of the cells belonged to one group (either positive or negative). Therefore, an automated thresholding algorithm based on the histogram of cells measured immunostaining intensities (e.g. Otsu's method) could not be used for cell classification. Our approach was to find a threshold based on a region R with no immunostaining in the immunostain image of a sample. We calculated the average and the standard deviation of the immunostaining intensities in this region. The threshold was then calculated as follows:

$$T = M_{background} + k SD_{background} \quad (3.34)$$

where $M_{background}$ and $SD_{background}$ are the average and the standard deviation of the intensities of the pixels within the region R in the immunostain image and the factor k can be

heuristically set by the user based on visual evaluation of the results. We used $k = 6$ for the cases in our study.

3.7.2 Results

Both Voronoi tessellation and the nearest neighbor tessellation approaches were examined. The tessellations generated by these two approaches were similar. The only instances where the tessellations could be different would be where the sizes of the two neighboring nuclei are significantly different. In that case, it makes a difference whether the distance of the points to be assigned to those nuclei are calculated to the nucleus boundary (the nearest neighbor tessellation) or to the nucleus center (Voronoi tessellation). If the distance is calculated to the boundary, the region assigned to the larger nucleus would be bigger than the region assigned to the smaller nucleus. However, in the samples in this study, no significant differences existed in the sizes of nuclei within each image; therefore, we demonstrate the results of Voronoi tessellation only.

For each sample, we used five grayscale images as the immunostain maps $I_{Immunostain}$ (i.e. the image representing the IHC label). These were:

- 1) The immunostain map that was generated by the application of the spectral unmixing algorithm to the 12-channel multispectral images (“multispectral unmixing”),
- 2) The immunostain map that was generated by the application of the spectral unmixing algorithm to the three-channel RGB images (“RGB unmixing”),
- 3) The immunostain map that was generated by the application of the LDA algorithm to the 12-channel multispectral images (“multispectral LDA”),
- 4) The immunostain map that was generated by the application of the LDA algorithm to the three-channel multispectral images (“RGB LDA”),
- 5) The blue component of the RGB image of the sample. The blue component was used because it had the highest absorption of DAB and lowest of Hematoxylin amongst the three channels of the RGB images.

To measure the IHC staining, we needed the binary map of nuclei. This would allow the association of the immunostaining with individual cells. In order to be able to compare the quality of the five immunostain maps listed above, we chose to use the same nuclei segmentation for all of them. In this way, only the differences in the immunostain map would

be compared, regardless of the quality of the nuclear segmentation. We generated the binary map of the nuclei based on the manually identified nuclei centers. However, because only the centers of the nuclei were manually marked, we took the following approach to obtain the nuclei surfaces: the image representing the nuclei generated through spectral unmixing of the 12-channel multispectral images was used as the nuclei image, since it resulted in the most accurate segmentation in the previous section. We performed the binarization procedure of section 3.6.1.1 on this nuclei image. The result was used as the initial segmentation. We then performed a Voronoi tessellation based on the manually marked nuclei centers and superimposed this tessellation on the binarized nuclei image generated above. The tessellation was used to separate the touching nuclei in the binarized nuclear image, i.e. the foreground region of the part of the binary map inside each Voronoi polygon was assumed to be the mask of the nucleus corresponding to that Voronoi polygon. This procedure is illustrated in Figure 3.69.

Figures 3.70 to 3.72 demonstrate examples of immunostain measurement for each set of IHC-labeled samples in this study.

The Voronoi polygons are “approximate” boundaries. Consequently, in some instances, the Voronoi polygon corresponding to a cell contains some pixels from a neighboring cell. In the P16 samples, this can sometimes cause the immunostaining intensity within a negative cell’s ROI that neighbors a positive cell to be estimated higher than what it really is. This can cause the negative cell to be falsely classified as positive. This can be avoided by taking into account the standard deviation of the immunostaining intensities of pixels in each cell’s ROI. In a negative cell that neighbors a positive cell, the immunostaining intensities of the pixels in the cell’s ROI would have a higher standard deviation. This is because in addition to the non-labeled ROI pixels, pixels with high DAB immunostaining that belong to the neighboring positive cell are included in its ROI. The high standard deviation enables the recognition of such negative cells. Figure 3.73 demonstrate the map of standard deviations of staining intensity in each cell’s ROI, along with the map of average staining intensity. Negative cells that are surrounded by positive cells and would otherwise be classified as positive can be identified by their higher standard deviation.

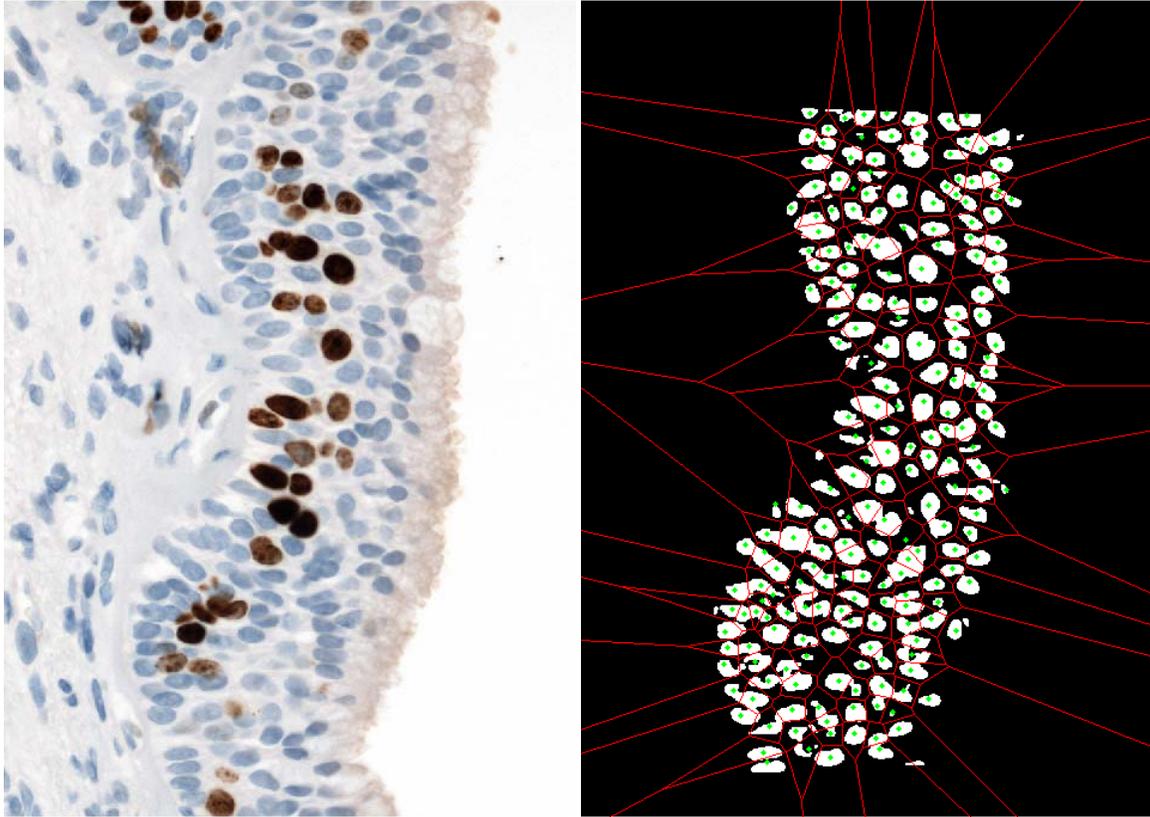


Figure 3.69 Example of the nuclei map used for the measurement of the immunostain: Voronoi tessellation superimposed on the binarized image, foreground objects bound by each Voronoi polygon are considered as nuclei.

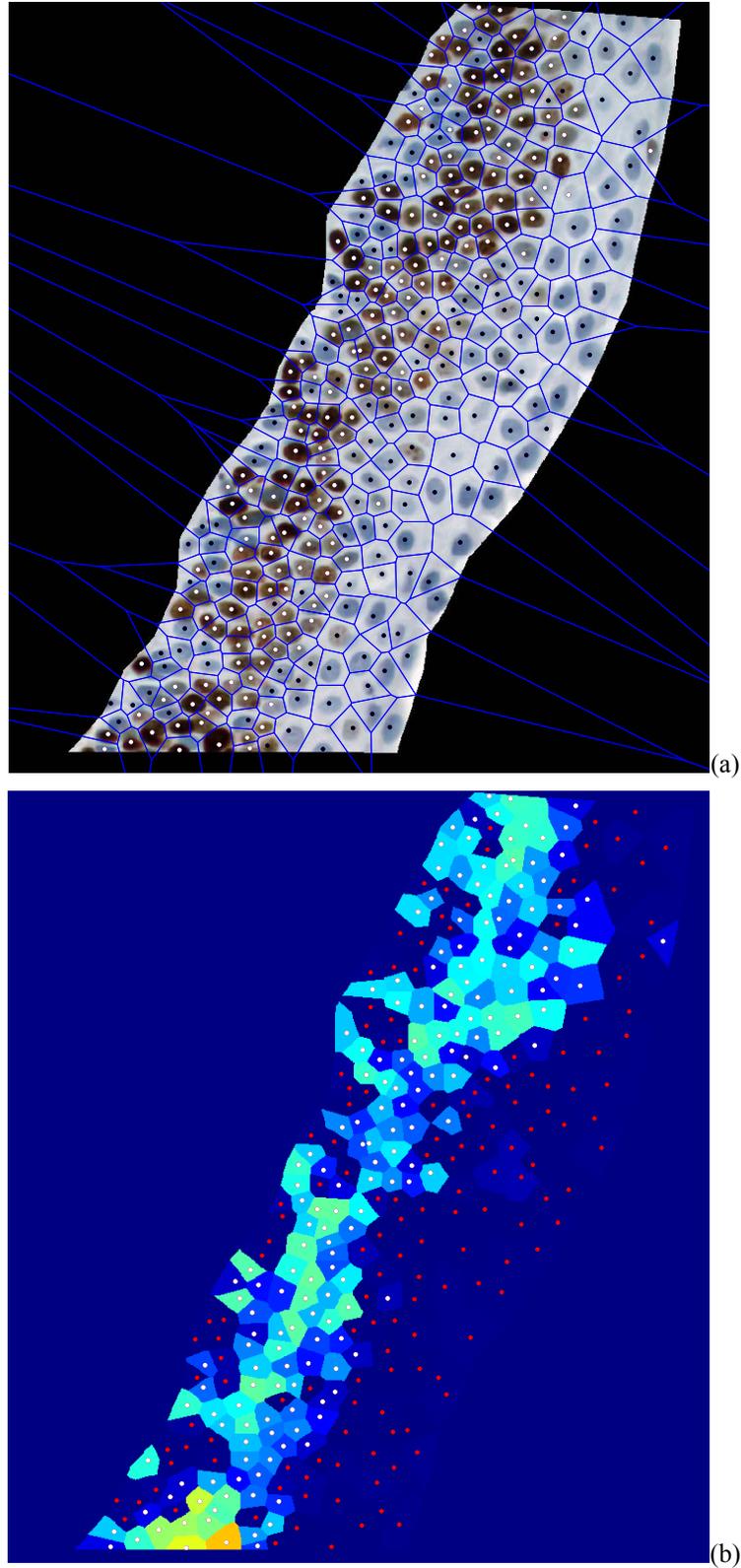
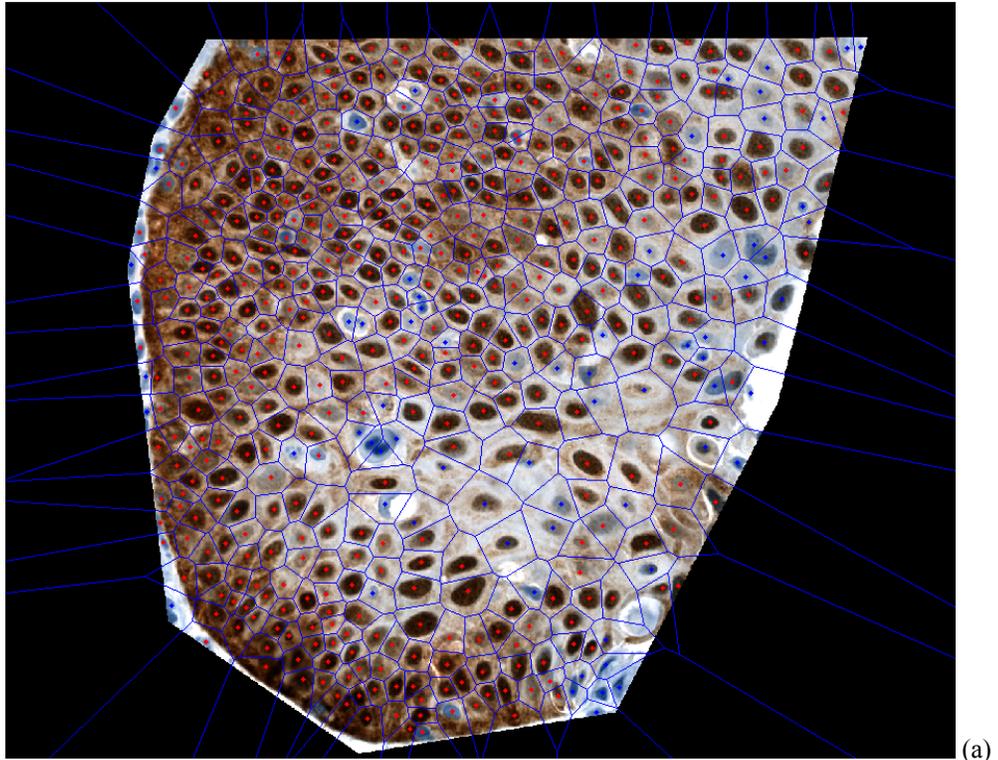
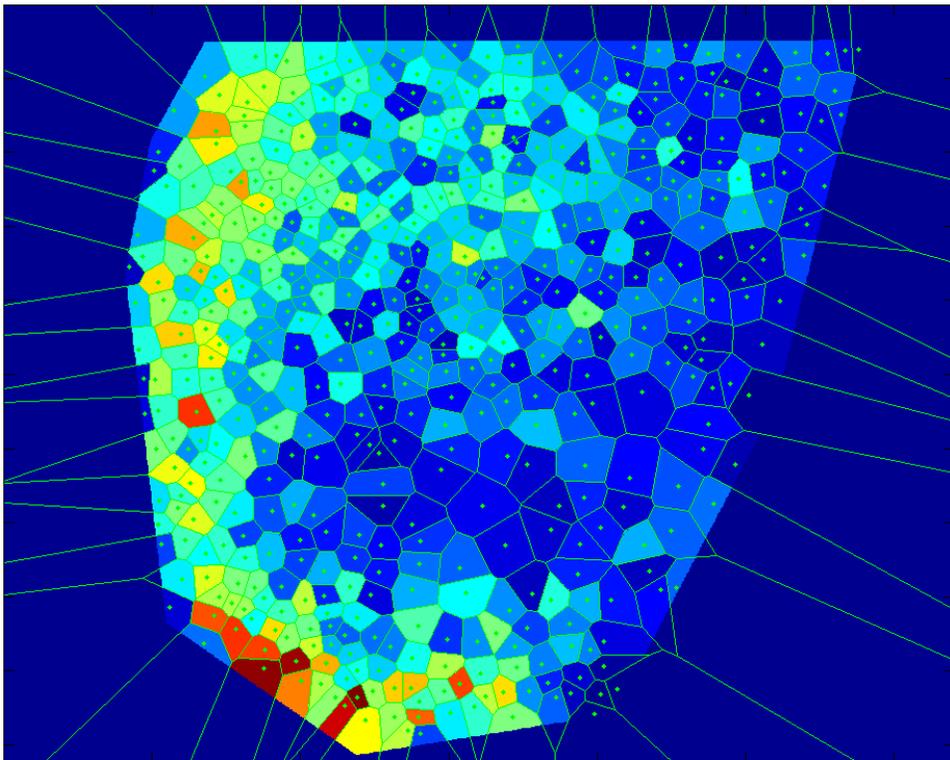


Figure 3.70 Ki67 Example, (a) cell classification, white dot indicates a positive nucleus, and black dot indicates a negative nucleus, (b) map of the measured staining inside the ROI of each cell. Red dots represent negative nuclei and yellow dots represent positive nuclei.



(a)



(b)

Figure 3.71 P16 Example, (a) cell classification, red dots represent positive cells and blue dots represent negative cells, (b) map of the measured staining inside the ROI of each cell.

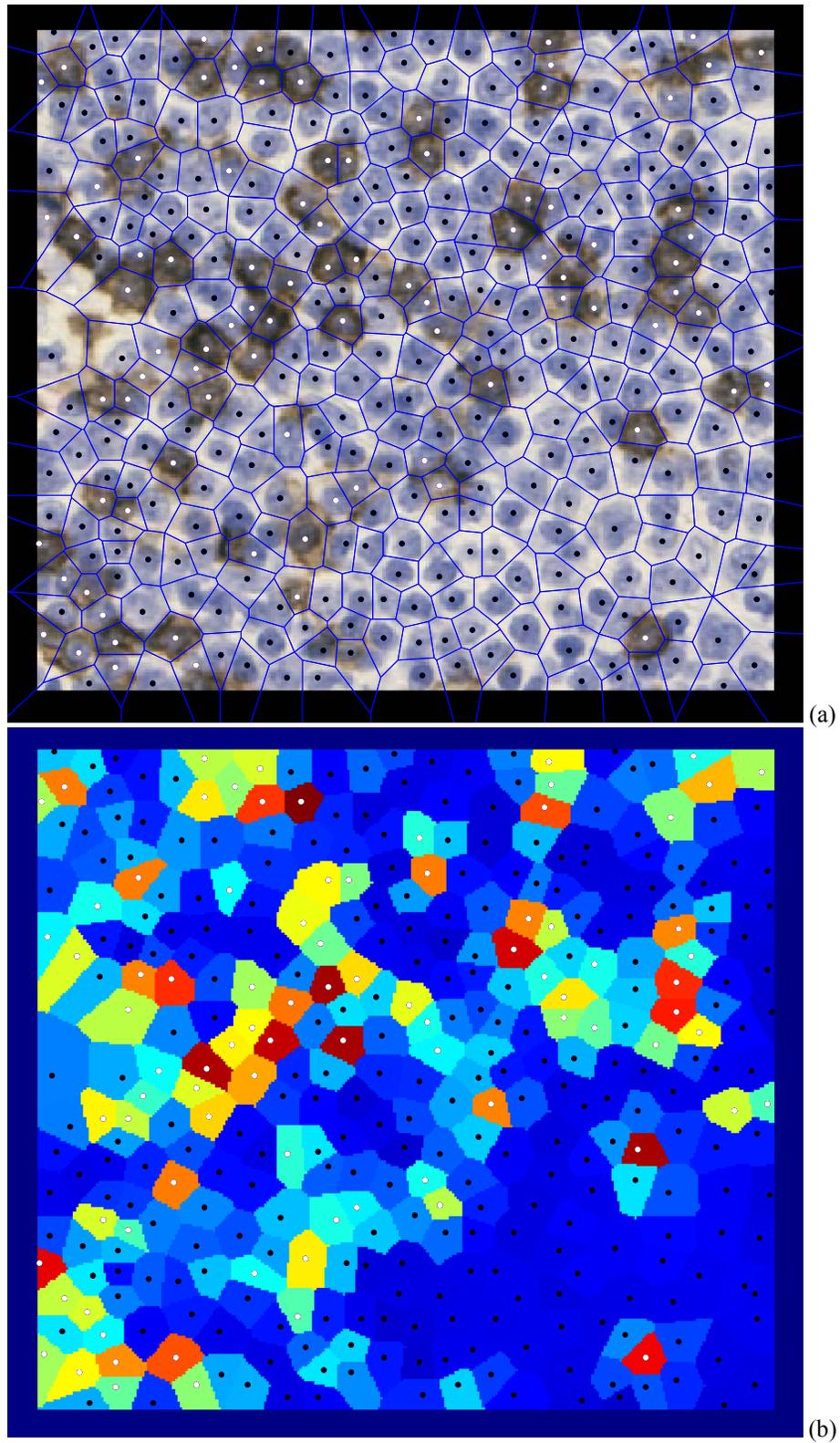


Figure 3.72 CD8 Example, (a) cell classification, white dot indicates a positive nucleus and black dot indicates a negative nucleus, (b) map of the measured staining inside the ROI of each cell.

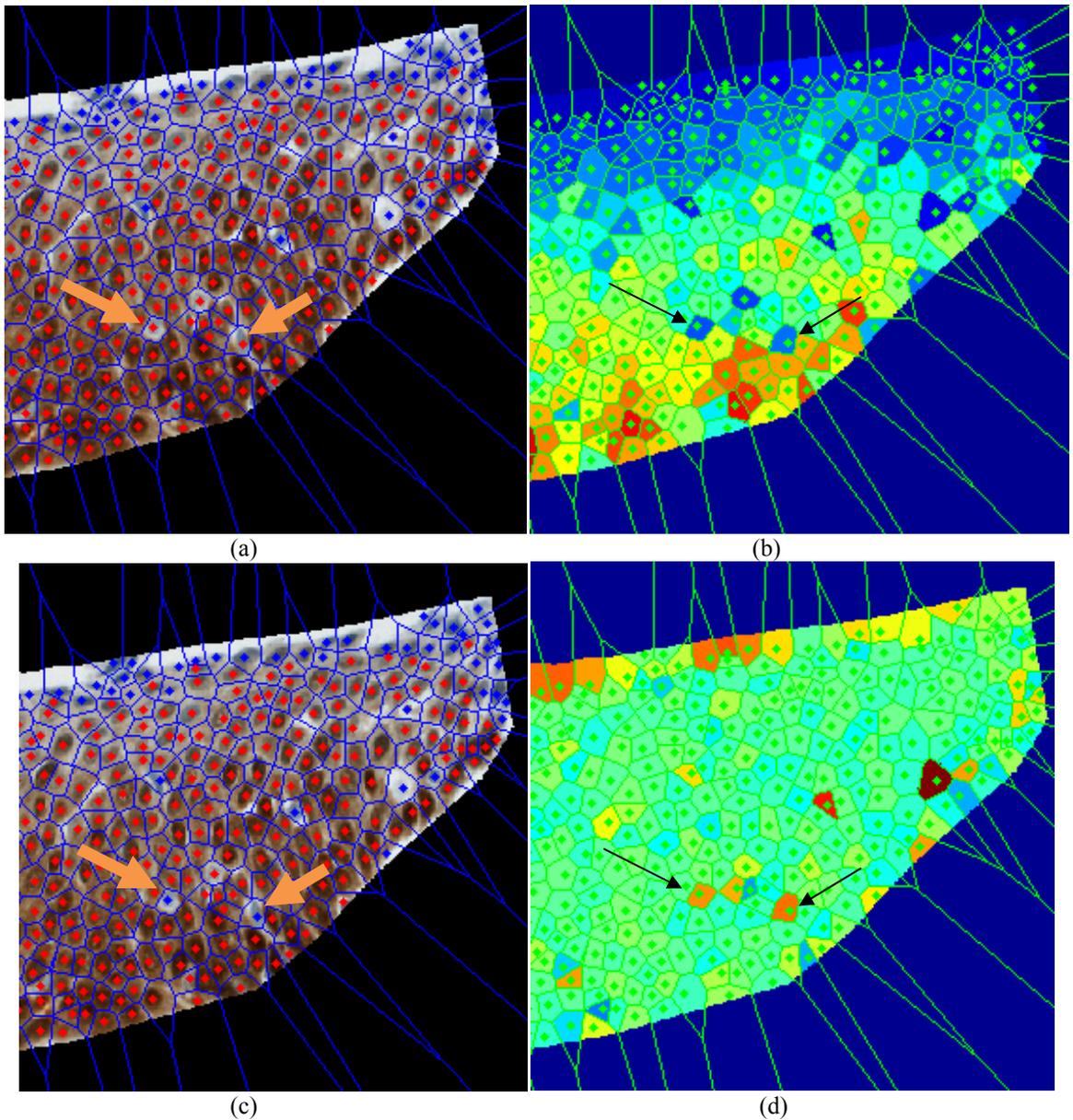


Figure 3.73 P16 Example two, (a) cell classification, red dot indicates a positive nucleus and blue dot indicates a negative nucleus, (b) map of the measured staining inside the ROI of each cell, (c) refined cell classification, taking into account the standard deviation of staining intensity inside the ROI of each cell, (d) map of the standard deviation of staining intensity inside the ROI of each cell. Arrows point to two positive cells that were identified as negative in (a).

All five immunostain images resulted in equally accurate classification results, when evaluated visually. To find out which one of the five immunostain images results in the best pixel-level contrast between positive and negative pixels in the immunostain image, we used the following methods:

1) For Ki67 and P16 biomarkers, we determined the positive and negative cells based on the immunostain image obtained using spectral unmixing of the 12-channel multispectral image. There was minimal difference between the classification results based on this immunostain image and that of the other four images. Therefore any of the immunostain images could be used. We assigned all the pixels from the ROI of positive cells to the positive pixels group, and all the pixels from the ROI of negative cells to the negative pixels group. We then calculated a Kruskal-Wallis test statistic to evaluate the differences in the mean ranks of the two groups of pixels. The Kruskal-Wallis test statistic is the non-parametric equivalent of the F-statistic in ANOVA and explains the between-group variability divided by the within-group variability.¹⁴⁴

2) For the CD8 biomarker, we took a slightly different approach. After identifying the positive and negative cells, we considered the positive cells only. Looking at positive cells, we assigned all the nuclei pixels to one group and all membrane pixels to another group. Those Voronoi polygon pixels that did not belong to the nuclei were considered membrane pixels. The larger the difference in the immunostain intensities of these two groups of pixels were, the more efficient the immunostain image was in identifying the positive pixels. The reason the positive and negative pixels were chosen this way was that this choice allowed for a more similar number of pixels in each group. The Kruskal-Wallis test statistic was then calculated to evaluate the differences in mean ranks of the two groups of pixels.

The Kruskal-Wallis test statistic values obtained from the five different immunostain images are presented in Figures 3.74 through 3.76.

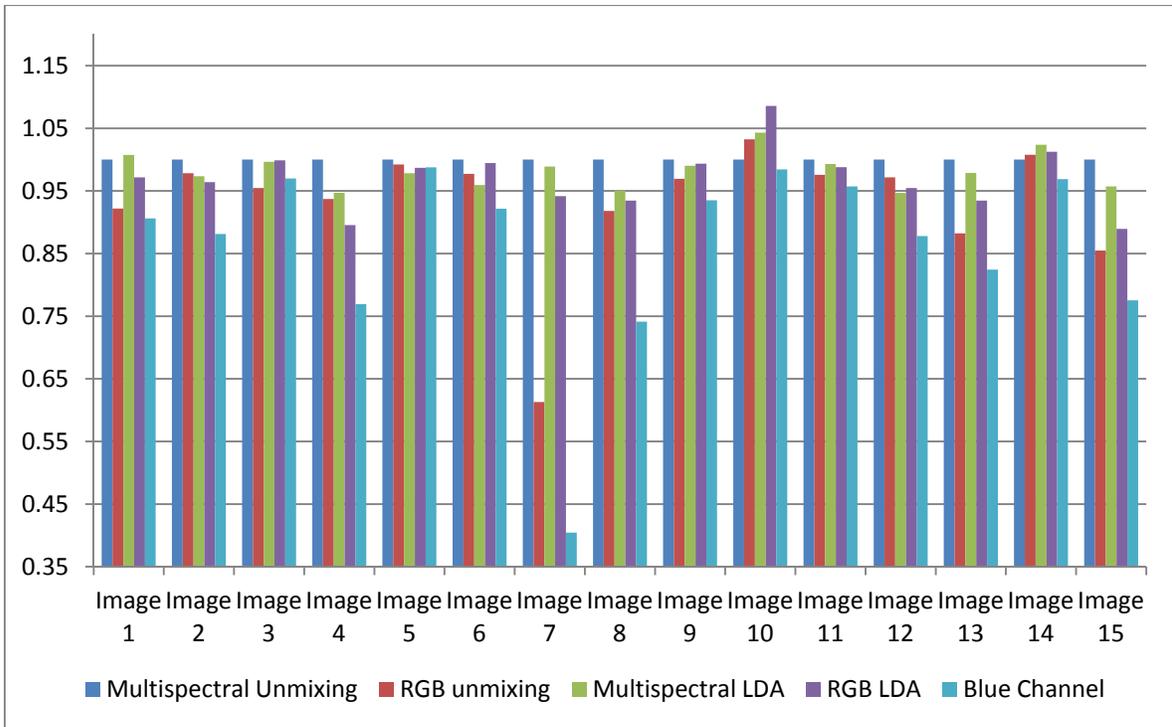


Figure 3.74 Comparison of Ki67 immunostain images: Kruskal-Wallis test statistics of immunostain images for separating the positive and negative pixels shown for each sample in Ki67 set.

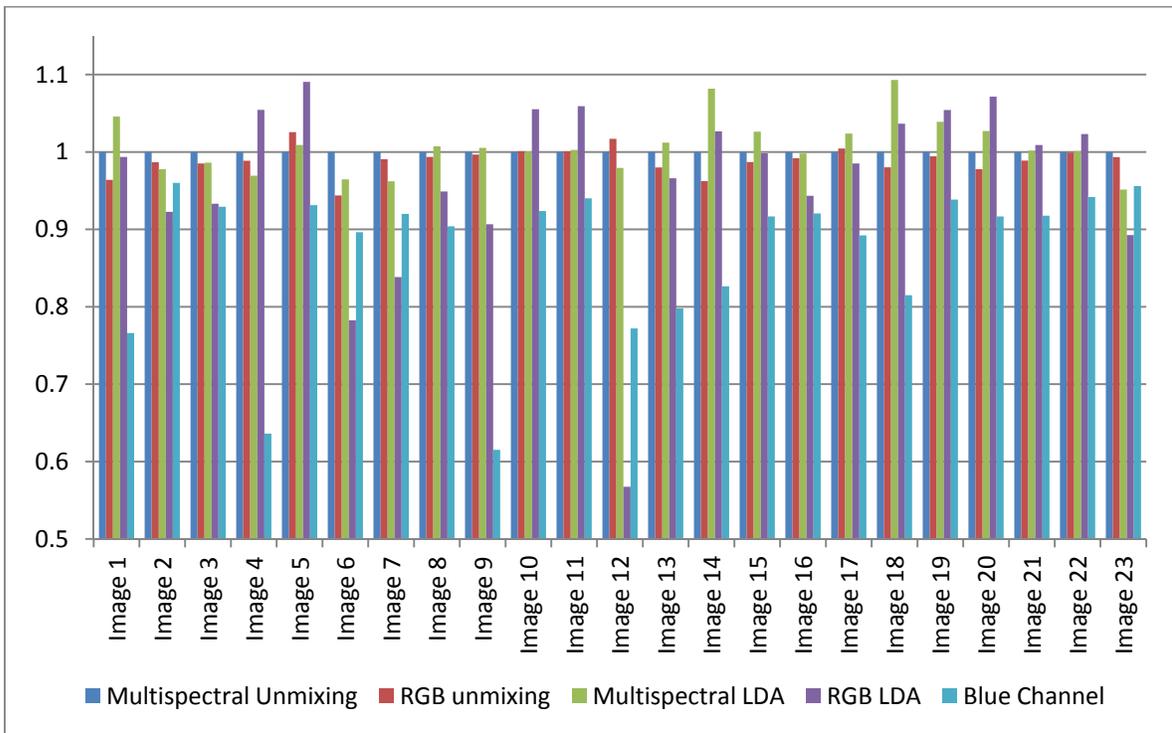


Figure 3.75 Comparison of P16 immunostain images: Kruskal-Wallis test statistics of immunostain images for separating the positive and negative pixels shown for each sample in P16 set.

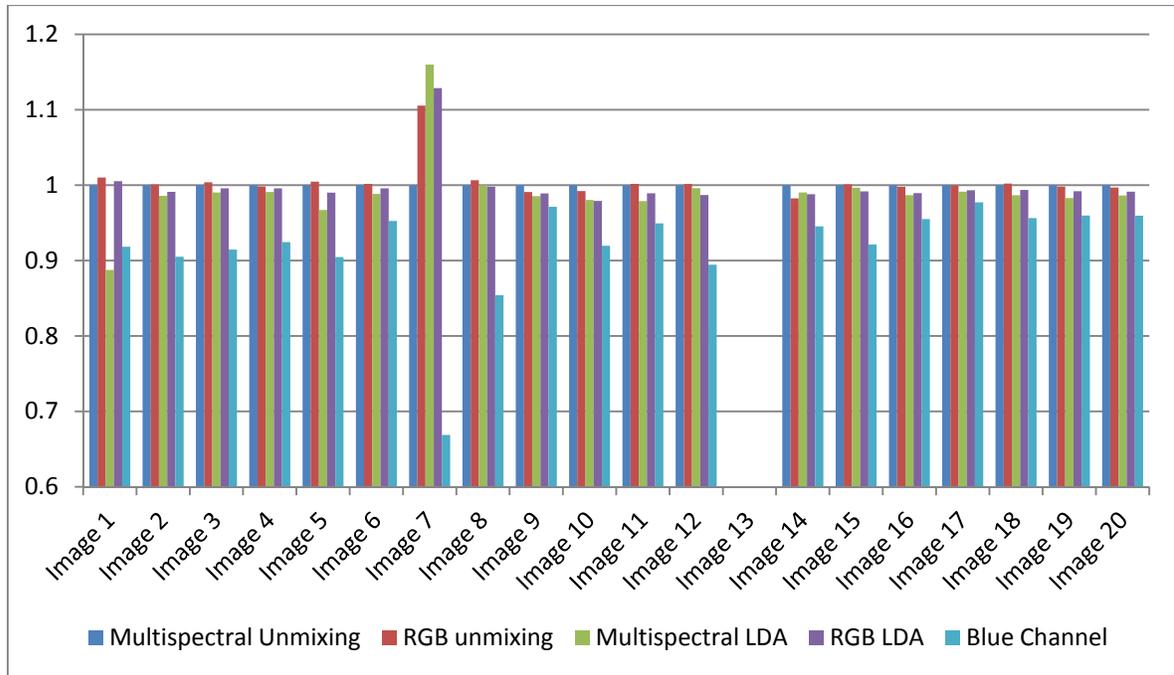


Figure 3.76 Comparison of CD8 immunostain images: Kruskal-Wallis test statistics of immunostain images for separating the positive and negative pixels shown for each sample in CD8 set. Image 13 did not contain any positive pixels, so no chi-square statistic is reported for it.

3.7.3 Discussion

The true gold standard for the identification of positive and negative cells would be the biological behavior of the lesion, i.e. progression or regression of a lesion, disease free survival of the cancer patient, effectiveness of a treatment, etc. It is only by correlation with biological behavior or outcome that one can compare several methods of IHC biomarker quantification. The best method would be the one that has the highest correlation with outcome. In the absence of this information, visual evaluation of the classification results is the only available reference. In our experience, the decision of positivity and negativity of the cells by a human observer is very subjective, especially in more complex samples such as P16 or CD8 immunostained samples. Because of the high uncertainty in the manual classification of the cells, we did not quantify the degree of similarity between the automated classification and the manual classification. All five immunostain images resulted in equally accurate classification results, when evaluated visually. We did, however, quantify the contrast between the intensity of the pixels that belonged to the ROI of positive and negative

cells. The method that resulted in the highest degree of separation between the positive and negative pixels would be most robust to the value of a threshold used for the classification of cells and would therefore be the preferable approach.

Based on our results (shown in Figures 3.74 to 3.76), the blue channel in the RGB images provided the least contrast between the positive and negative pixels. For Ki67 and CD8, the Kruskal-Wallis test statistic values for other four immunostain images were very close and did not follow the same order from one image to the other. For P16, in samples 4, 5, 10, 11, 14, 18, 19, and 20 LDA outperformed Spectral unmixing. In five out of these eight images, strong DAB immunostaining was present, which as discussed before, could cause inaccurate unmixing. This suggests that when dark DAB staining was present, LDA outperformed spectral unmixing. In some samples, “RGB LDA” outperformed “multispectral LDA”. This could be due to overtraining in the LDA, which may have been caused by the use of a small training set or inclusion of too many variables in the discriminant function.

3.8 Summary

In this chapter, we presented methods for the analysis of multispectral images of IHC stained tissue sections. The goal of the analysis was to assess the amount of a labeled IHC biomarker within the nucleus, cytoplasm, or membrane of each cell in an IHC-labeled tissue section. The analysis involved the following tasks:

1) The first task was to generate N grayscale images from the multispectral image stack of a tissue section that was labeled with N labels. Each grayscale image represented the distribution of one of the N labels within the tissue section. The image that represented the nuclear label (the “nuclei image”) could then be used for the identification of nuclei and the image that represented an IHC label (the “immunostain image”) could then be used for the assessment of that IHC biomarker in the vicinity of every nucleus (which was identified in the “nuclei image”).

Two approaches were investigated for performing the above task: a) linear spectral unmixing and b) linear discriminant analysis (LDA). The two approaches were applied to four sets of IHC labeled tissue sections. These four sets consisted of Ki67, P16, CD8, and Ki67 & P16 samples (65 IHC-labeled tissue sections in total). We were able to robustly generate the “nuclei” and “immunostain” images (as defined in the previous paragraph) of each IHC labeled tissue section using our developed methods.

2) The second task was the automated segmentation of nuclei in “nuclei images” obtained by one of the above mentioned approaches. For each tissue section, the location of all nuclei were needed to determine the location of cells, which were in turn needed for the measurement of the IHC biomarker in the cells. For this task, a fully automated nuclear segmentation algorithm was developed.

Using our developed segmentation procedure, nuclei in 58 IHC-labeled tissue sections (15 Ki67, 23 P16, and 20 CD8 samples) were segmented. For each tissue section, the nuclei image to be used by the segmentation procedure was generated in five ways: 1) by performing linear spectral unmixing on the sample’s 12-channel multispectral image stack, 2) by performing linear spectral unmixing on the sample’s three-channel RGB image, 3) by performing LDA on the sample’s 12-channel multispectral image stack, 4) by performing LDA on the sample’s three-channel RGB image, and 5) the red component of the RGB image of the tissue section was used as the fifth representation of the nuclei image. Each one

of the five nuclei images generated by the five methods listed above was segmented by our developed automated segmentation procedure. The accuracy of the automated segmentation of each nuclei image was measured by comparing the segmentation results with manually annotated nuclei. Our results showed that the nuclei image generated from the multispectral images by linear spectral unmixing resulted in the most accurate segmentation.

3) The third task was the measurement of the abundance of a labeled IHC biomarker in the cell compartment where the IHC biomarker was expressed (nucleus, cytoplasm, or membrane). This was done by partitioning the area of interest in a tissue section into several regions, where each region approximated the surface of a cell. The abundance of the biomarker present in each cell was then determined by measuring the intensity of the “immunostain image” within the region corresponding to that cell. According to this measured value, cells were classified into positive and negative for the biomarker of interest.

For each tissue section, the immunostain image to be used for the measurement of the IHC biomarker was generated in five ways: 1) by performing linear spectral unmixing on the 12-channel multispectral image stack of the sample, 2) by performing linear spectral unmixing on the sample’s three-channel RGB image, 3) by performing LDA on the sample’s 12-channel multispectral image stack, 4) by performing LDA on the sample’s three-channel RGB image, and 5) the blue component of the RGB image of the tissue section was used as the fifth representation of the immunostain image. All five immunostain images listed above resulted in good classification of the cells into positive and negative, as judged visually. The contrast between the pixel intensities with and without the IHC label, in each immunostain image, was also evaluated. There were no significant differences between the results of methods 1 to 4. Method 5, however, results in the lowest contrast. The true gold standard for the classification of cells would be correlation of cell classification results with the biological behavior or outcome of lesions, i.e. progression from dysplasia to cancer, disease free survival of the cancer patient, etc. This information was not available in this study. Evaluation of the ultimate results of the IHC quantification procedure was therefore not possible.

Chapter 4. Tissue Biomarker Imaging through Selective Illumination

4.1 Introduction

In chapter 3, we use multispectral imaging to quantify the expression of biomarkers in images of histopathological tissue sections. There are a number of approaches to acquire multispectral images. The main ones are:

1) Use of multiple band-pass optical filters (filter wheels). Filter-wheel based systems have been applied to immuno-phenotyping and to whole genome karyotyping. However, they lack flexibility in the choice of wavelength channels. Also because of the moving mechanical parts, the image acquisition can take a long time.

2) Interferometry. In this approach the light transmitted from or reflected by the sample enters a Sagnac interferometer where it is split into two beams. These beams are sent in opposite directions around a common light path and are allowed to interfere with one another after a small and variable optical path difference has been introduced. The interference pattern is then detected by a camera. An interferogram is generated at each pixel by optically changing the optical path difference and recording the signals of successive frames. The Fourier transform of the interferogram at each pixel creates a spectral image cube.

3) Use of tunable filters such as Acousto-Optical Tunable Filters (AOTFs) and Liquid Crystal Tunable Filters (LCTFs) to filter the light transmitted or reflected by the sample into a series of narrow-band wavelength channels. Instead of filtering the light transmitted or reflected by the sample, it is also possible to perform multispectral imaging by illuminating the sample with a series of narrow-wavelength band illumination profiles. LCTFs or AOTFs are however not suitable for this task, since they only pass a small portion of the illuminating light for each wavelength band, and the portion that does not pass through the filter generates heat that may damage the filter.

The approach presented here is different from above. We acquire multispectral images using multispectral illumination. We use a spectrally programmable light engine to generate spectrally narrow-band illumination light. OneLight Spectra (OneLight corp., Vancouver,

Canada) is a tunable light engine that can generate illumination of any desired mixture of wavelengths in the 400nm-680nm range, and can control the intensity at each wavelength independently. It uses the Texas Instruments Digital Micro-mirror Device (DMD), a type of spatial light modulator, in combination with a lamp and related dispersive and conventional optics to control the wavelength dependent energy distribution of light.¹²⁰

There are advantages to the use of multispectral illumination for spectral imaging: unlike using tunable filters on the detection side, the optical path from the sample to the camera is unaffected by additional optical elements. So no artifacts are introduced to the detected light. In addition, with no filters on the detection side, all the light transmitted from or reflected by the sample is recorded by the camera. This allows for more efficient use of the light, i.e. capture of maximum light without illuminating the sample with unnecessary power.

One particular advantage of multispectral illumination for spectral imaging that is the subject of this chapter is that it allows for imaging under tailored spectral illumination. Multispectral imaging generates a large amount of data through the sequential capture of multiple images at multiple illumination bands. Substantial computer processing such as pixel classification or linear spectral unmixing is required to extract the information relevant to the identification of features of interest in the samples under investigation. The objective of the study reported in this chapter is to reduce the number of captured images to reduce the analysis cost (i.e. time and memory) of the samples. We present an imaging system for the acquisition of images using selective spectral illumination for both transmission imaging and fluorescence imaging. In the following two subsections, the objectives of this study will be specified for each imaging mode.

4.1.1 Transmission Imaging

Transmission imaging refers to the detection of the light transmitted through a light-absorbing specimen. We acquire multispectral transmission images of tissue samples using multispectral illumination generated by a spectrally programmable light engine. We demonstrate how representations of particular objects of interest in a tissue sample can be generated through weighted sums of the multispectral images of that tissue sample. We then explore the possibility of capturing a fewer number of images by designing selective spectral illumination profiles for the identification of particular objects of interest (e.g. labeled

biomarker or labeled nuclei) in the tissue. These selective spectral illumination profiles are calculated as weighted sums of the spectrally narrow illumination profiles which were used for the capture of the multispectral images. To achieve this, we leverage the ability of a spectrally programmable light engine to generate a mixture of spectrally narrow illumination bands with controlled intensity at every spectral illumination band. We hypothesize that using this imaging approach; we can obtain the most relevant information from all spectral bands in a smaller number of images, over the number of images within a multispectral stack of images. Each one of these images is captured under a specific weighted sum of the spectral illumination bands. We hypothesize that the use of a tunable light engine for illumination allows the design of selective spectral illumination profiles that highlight particular objects of interest in the image of the sample. We demonstrate how under specific mathematically designed illumination profiles, specific biomarkers may be imaged.

4.1.2 Fluorescence Imaging

In transmission imaging, multispectral illumination generates multispectral images that are the same as those that would be generated by multispectral detection. However, this is not the case in fluorescence imaging.

Typically, fluorescence imaging systems use tunable filters on the detection side, i.e. fluorescence light emitted from the object is detected in several spectrally narrow wavelength channels. This results in the collection of a multispectral emission image stack. Illuminating an object with several spectrally narrow illumination profiles enable the collection of a multispectral excitation image stack. Such multispectral excitation image stack carries information about the excitation spectrum at each pixel of the image, while multispectral emission image stack carries information about the emission spectrum at each pixel. The excitation spectrum refers to the intensity of collected fluorescence emissions at different excitation wavelengths, while the emission spectrum refers to the intensity of the collected emissions at different emission wavelengths.

Multispectral image analysis techniques can be performed on the excitation image stack as they can be performed on an emission image stack; the difference being that with the emission image stack, differences in the emitted images are due to the different emission spectra of the underlying fluorescent material. On the other hand, with the excitation image

stack, the differences in the emitted images are caused by the different excitation profiles of the fluorescing material. Therefore, excitation imaging generates information that is independent from that of emission imaging.

In this chapter we hypothesize and show that 1) multispectral fluorescence imaging through multispectral illumination enables the detection of changes in the absorption/excitation profile of the fluorescent material and 2) spectral illumination profile may be selectively designed to capture images that improve the identification of particular objects of interest in the sample.

4.2 Selective Illumination Transmission Imaging

As discussed in Chapter 3, multispectral imaging generates a large amount of data. Substantial computer processing is required to extract the information relevant to the identification of features of interest in the samples under investigation. Multispectral Imaging involves sequential capture of several images resulting in increased data volume and decreased speed of operation. Subsequent computer analysis may consist of pixel classification or linear spectral unmixing. It is of interest to reduce the number of captured images to reduce the analysis cost (i.e. time and memory) of the samples.

Miller et al.¹⁴⁵ reported the use of an agile lamp (i.e. a spectral illuminator) for imaging which produces illumination with any desired spectral flux distribution. Using a method denoted “matched filtering”, they proposed choosing the spectral illuminator’s settings to produce illumination spectra that allow the extraction of full spectral information from many bands at once. The design of the illuminator, however, was not discussed in their report.

In this section, we explore a new approach to the imaging and identification of specific labeled objects in a sample. We illuminate the sample with a selective spectral illumination provided by a spectrally programmable light engine. The basic idea is as follows: we acquire a stack of multispectral images of a tissue sample using a series of spectrally narrow illumination bands. Assume that we calculate a weighted sum of the multispectral images which highlights specific objects (e.g. labeled nuclei or labeled biomarkers), i.e. maximizes the contrast between the specific objects and the rest of the image. We show that we can directly capture this weighted sum of images if we illuminate the tissue sample with a

specific spectral illumination profile. We show that this specific spectral illumination profile can be calculated as a weighted sum of the spectrally narrow illumination profiles.

In the following sections, we first derive the mathematical formula for calculating the coefficients for the weighted sum of illumination profiles. Then, as a proof of principle example, we demonstrate the use of Linear Discriminant Analysis (LDA) in finding a weighted linear combination of the multispectral images that maximizes the contrast between a labeled object of interest and the rest of the sample. We show that we can design a specific illumination profile for the capture of such image.

4.2.1 Methods

Consider a system consisting of a spectrally programmable light source, a camera, and a light absorbing specimen in front of the camera. The light engine sequentially illuminates the sample with L narrow-wavelength bands of light. At each illumination band, the camera captures an image of the sample. The output of the camera at the i -th illumination band I_i can be described as:

$$I_i = t_i (\int \varphi_i(\lambda) T(\lambda) S(\lambda) d\lambda) + D_i, \text{ for } i = 1, 2, \dots, L \quad (4.1)$$

where $\varphi_i(\lambda)$ is the incident energy density per unit time at wavelength λ of the i -th wavelength channel, $T(\lambda)$ is the transmittance of the sample, and $S(\lambda)$ is the spectral sensitivity of the camera. t_i is the exposure time at the i -th illumination channel. D_i is the dark-frame image, i.e. the image captured by the camera sensor for exposure time t_i when no light is present. If the dark-frame image intensities are significantly lower than the sample image (the offset of the camera is close to zero), we can disregard the dark-frame image, $D_i = 0$, in which case there is a linear relationship between the incident light, $\varphi_i(\lambda)$, and the captured image of the sample I_i .

Now, let the incident light be $\psi(\lambda)$, where:

$$\psi(\lambda) = \sum_{i=1}^L c_i \varphi_i(\lambda) \quad (4.2)$$

Let the image captured under this illumination be F . F can be expressed as:

$$F = t_F \left(\int \psi(\lambda) T(\lambda) S(\lambda) d\lambda \right) \quad (4.3)$$

where t_F is the exposure time for the capture of F . Replacing $\psi(\lambda)$ with its value in equation 4.2 gives:

$$\begin{aligned} F &= t_F \left(\int \left(\sum_{i=1}^L c_i \varphi_i(\lambda) \right) T(\lambda) S(\lambda) d\lambda \right) \\ &= t_F \sum_{i=1}^L \left(\int c_i \varphi_i(\lambda) T(\lambda) S(\lambda) d\lambda \right) \\ &= t_F \sum_{i=1}^L \left(c_i \int \varphi_i(\lambda) T(\lambda) S(\lambda) d\lambda \right) \\ &= t_F \sum_{i=1}^L \left(c_i \frac{I_i}{t_i} \right) = t_F \sum_{i=1}^L \frac{c_i}{t_i} I_i = t_F \sum_{i=1}^L v_i I_i \end{aligned} \quad (4.4)$$

where $v_i = \frac{c_i}{t_i}$, for $i = 1, 2, \dots, L$. t_F is a scaling factor in image F , and without loss of generality we can assume $t_F = 1$. Therefore we have:

$$\begin{aligned} \text{Illumination} &\rightarrow \text{Image} \\ \varphi_i(\lambda) &\rightarrow I_i, i = 1, 2, \dots, L \\ \psi(\lambda) = \sum_{i=1}^L c_i \varphi_i(\lambda) &\rightarrow F = \sum_{i=1}^L v_i I_i \end{aligned} \quad (4.5)$$

Let's assume that by using some mathematical technique, we found some linear combination of the multispectral images by which we can highlight a certain object of interest in the sample, or reduce the dimensionality of the multispectral image set, etc. If the coefficients in this weighted sum are $v_i, i = 1, 2, \dots, L$, then we can calculate the illumination profile, $\psi(\lambda)$, so that if the sample is illuminated with $\psi(\lambda)$, image F can be captured directly. $\psi(\lambda)$ is calculated using equation 4.2, where coefficients c_i are:

$$c_i = v_i t_i \quad (4.6)$$

The above derivation was based on “raw”, i.e. non-calibrated, captured images. Usually, transmission images are calibrated by dividing them by the flat-field image. A flat-field image is the image captured from a transparent area, where the sample is absent. The

exposure times, t_i are selected so that the flat-field image reaches just below the maximum pixel intensity (e.g. 90% of the maximum pixel intensity: 255 in an 8-bit image). The same exposure time t_i is used for the capture of both the flat-field and the sample image. Similar to equation 4.1 the flat-field image can be expressed as:

$$IF_i = t_i \left(\int \varphi_i(\lambda) S(\lambda) d\lambda \right) + D_i, \text{ for } i = 1, 2, \dots, L \quad (4.7)$$

The calibrated image $\bar{I}_i(\lambda)$ is obtained by dark-field D_i , and flat-field IF_i correction as follows:

$$\begin{aligned} \bar{I}_i(\lambda) &= \frac{I_i - D_i}{IF_i - D_i} = \frac{t_i \int \varphi_i(\lambda) T(\lambda) S(\lambda) d\lambda}{t_i \int \varphi_i(\lambda) S(\lambda) d\lambda} \\ &= \frac{\int \varphi_i(\lambda) T(\lambda) S(\lambda) d\lambda}{\int \varphi_i(\lambda) S(\lambda) d\lambda} \end{aligned} \quad (4.8)$$

We assume that each wavelength channel is narrow enough so that $T(\lambda)$ is constant within that channel: $T(\lambda) = T_i$. Another assumption that we make is that the camera response $S(\lambda)$ is perfectly linear and is independent of illumination wavelength: $S(\lambda) = S$, where S is a constant. So, the i -th illumination channel, $\varphi_i(\lambda)$, generates the following image and flat-field-corrected image:

$$I_i = T_i t_i S \int \varphi_i(\lambda) d\lambda \quad (4.9)$$

$$\bar{I}_i = \frac{T_i S \int \varphi_i(\lambda) d\lambda}{S \int \varphi_i(\lambda) d\lambda} = T_i \quad (4.10)$$

Now we derive the image F captured under illumination $\psi(\lambda)$ using the corrected images, \bar{I}_i :

$$\begin{aligned}
F &= t_F S \int T(\lambda) \psi(\lambda) d\lambda = t_F S \int T(\lambda) \left(\sum_{i=1}^L c_i \varphi_i(\lambda) \right) d\lambda \\
&= t_F S \sum_{i=1}^L \int T(\lambda) c_i \varphi_i(\lambda) d\lambda \\
&= t_F S \sum_{i=1}^L \left(c_i T_i \int \varphi_i(\lambda) d\lambda \right) \stackrel{\text{Substituting Eq (4.10)}}{=} t_F S \sum_{i=1}^L c_i \bar{I}_i \left(\int \varphi_i(\lambda) d\lambda \right) \\
&= t_F S \sum_{i=1}^L v_i \bar{I}_i
\end{aligned} \tag{4.11}$$

where $v_i = c_i \left(\int \varphi_i(\lambda) d\lambda \right)$, $i = 1, 2, \dots, L$. Therefore using the corrected images, \bar{I}_i , the coefficients c_i for the generation of illumination $\psi(\lambda)$ are calculated as follows:

$$c_i = \frac{v_i}{\int \varphi_i(\lambda) d\lambda} \quad i = 1, 2, \dots, L \tag{4.12}$$

Thus image F may be generated by illuminating the sample with the following weighted sum of narrow-band illumination channels $\varphi_i(\lambda)$:

$$\psi(\lambda) = \sum_{i=1}^L c_i \varphi_i(\lambda) \tag{4.13}$$

Therefore, once the coefficients v_i , $i = 1, 2, \dots, L$, are calculated for generating the image F of a labeled object of choice within a sample using $F = \sum_{i=1}^L v_i I_i$, instead of capturing L images I_i at L illumination channels, image F can be captured under illumination $\psi(\lambda)$. Assuming that coefficients v_i do not change from one sample to another sample with the same composition, the labeled object of interest can be imaged in subsequent samples with the same composition under the same illumination (λ) .

This approach, which we refer to as ‘‘Selective Illumination Transmission Imaging’’, leverages the ability of the spectrally programmable light engine to generate a mixture of wavelength bands with controlled intensity at every spectral band. It provides a significant advantage over band-sequential multispectral imaging, by enabling the collection of the most relevant information from all spectral bands in a smaller number of images, compared to the number of images in a multispectral image stack.

One technique for reducing the dimensionality of a spectral image set is Principal Component Analysis (PCA) which compresses the information in the multispectral channels into a few channels through a linear transformation. This technique is useful when we want to obtain the highest overall contrast in the image. An alternative approach that allows the identification of specific targets in the sample is Linear Discriminant Analysis (LDA). In chapter 3, we demonstrated the use of LDA in the identification and separation of the labels in multiply labeled tissue sections. Linear discriminant functions can be generated based on training sets consisting of groups of pixels that correspond to the objects which we wish to separate.

One problem with the “Selective Illumination Transmission Imaging” approach is that not all the PCA or LDA coefficients will be positive, so the illumination intensity versus wavelength profile that is calculated as a weighted sum of the excitation profiles may have negative intensities. Negative light is a nonsensical concept. To overcome this limitation, two illumination profiles will need to be generated: one using the positive part of the calculated illumination profile, and one using the inverse of the negative part:

$$\psi_{pos}(\lambda) = \begin{cases} \psi(\lambda), & \text{if } \psi(\lambda) > 0 \\ 0 & \text{otherwise} \end{cases}, \text{ and } \psi_{neg}(\lambda) = \begin{cases} -\psi(\lambda), & \text{if } \psi(\lambda) < 0 \\ 0 & \text{otherwise} \end{cases} \quad (4.14)$$

For each of these two illumination profiles, one image is captured. The image captured under $\psi_{neg}(\lambda)$ is then subtracted from the image captured under $\psi_{pos}(\lambda)$. Using this method, instead of capturing several images (one for each excitation wavelength) and then calculating the weighted sum of the images, the weighted sum images are captured directly. This reduces the number of images that need to be captured and results in a shorter image acquisition time.

In the following section we present the application of Selective Illumination Transmission Imaging to the identification of labeled objects of interest in the sample through LDA.

4.2.2 Results

In the previous section we derived the illumination profile under which the image of a weighted sum of multispectral images can be directly captured.

As stated before, the coefficients in the weighted sum may be obtained in different ways. One way of obtaining such coefficients is through LDA. Let us assume that multispectral

images of a sample are captured under a series of narrow-band illumination profiles spanning the range of visible wavelengths. We aim to discriminate two groups of objects based on their multispectral intensities. We can build a training set consisting of pixels from each group. Through LDA, we can calculate a linear discriminant function as a weighted sum of pixel intensities in the multispectral images. Using the “Selective Illumination Transmission Imaging” approach explained in the previous section, we can calculate what we refer to as the “linear discriminant illumination profile”, by calculating the appropriate weights in equation 4.2 from the LDA coefficients v_i using either $c_i = v_i t_i$ or $c_i = \frac{v_i}{\int \phi_i(\lambda) d\lambda}$, depending on whether the “raw”, i.e. non-calibrated, or “flat-field corrected” images were used in the LDA analysis.

We demonstrate this approach using images from Ki67 immunostained tissue sections, as an example. These tissue sections are labeled for nuclei and Ki67 protein with Hematoxylin and DAB chromogens respectively.

A multispectral image stack was captured under multispectral illumination that consisted of 12 narrow-band illumination spectra with central wavelengths of 420 nm, 440 nm, ..., and 640 nm. The bandwidth at half maximum (BWHM) of each illumination channel was 20 nm. Images from the same field of view were also taken at three wideband channels corresponding to red, green, and blue channels of three-color RGB imaging, to form equivalent of RGB images. Figure 4.1 shows the RGB image of a typical Ki67 sample.

In chapter 3, we presented how LDA can be used to build two discriminant functions: one to separate the labeled nuclei pixels from all other pixels and one to separate DAB labeled pixels from all other pixels. Such discriminant functions provide the coefficients of the linear transformation that transforms the multispectral images into one image that provides the maximum separation between the objects of interest and the rest of the sample. Therefore discrimination of the objects that would require the collection of a multispectral image stack is achieved by collecting a significantly smaller number of images captured under specifically designed illumination profiles.

The coefficients for the linear discriminant functions were obtained using selected pixels from this image for training using the methods described in section 3.5. The linear discriminant analysis was performed on the 12-dimensional multispectral image stack as well as on the three channels of the RGB image. LDA coefficients are shown in Figures 4.2 and

4.3. Figure 4.4 demonstrates the illumination profiles for the capture of multispectral images and RGB images.

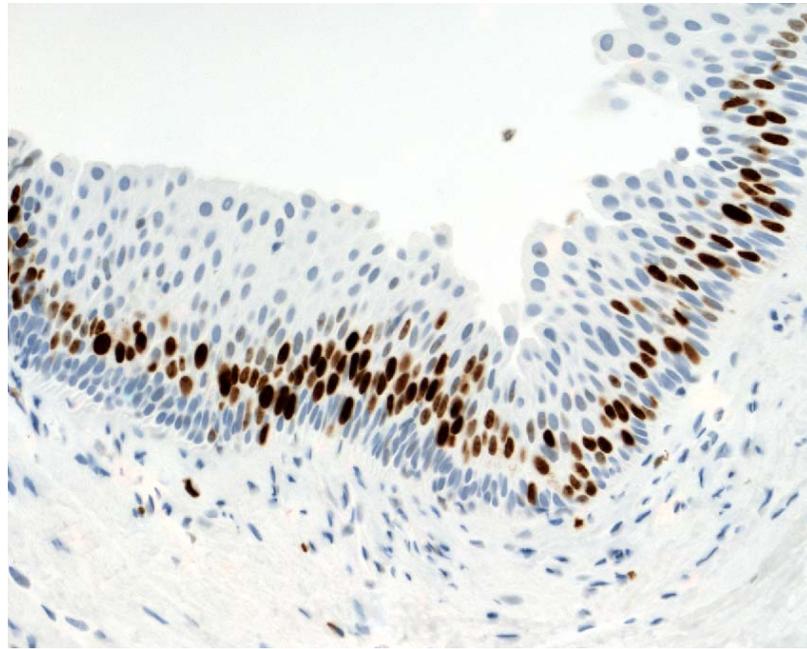


Figure 4.1 RGB image of a typical Ki67 immunostained sample, captured with three-channel illumination generated with OneLight light engine

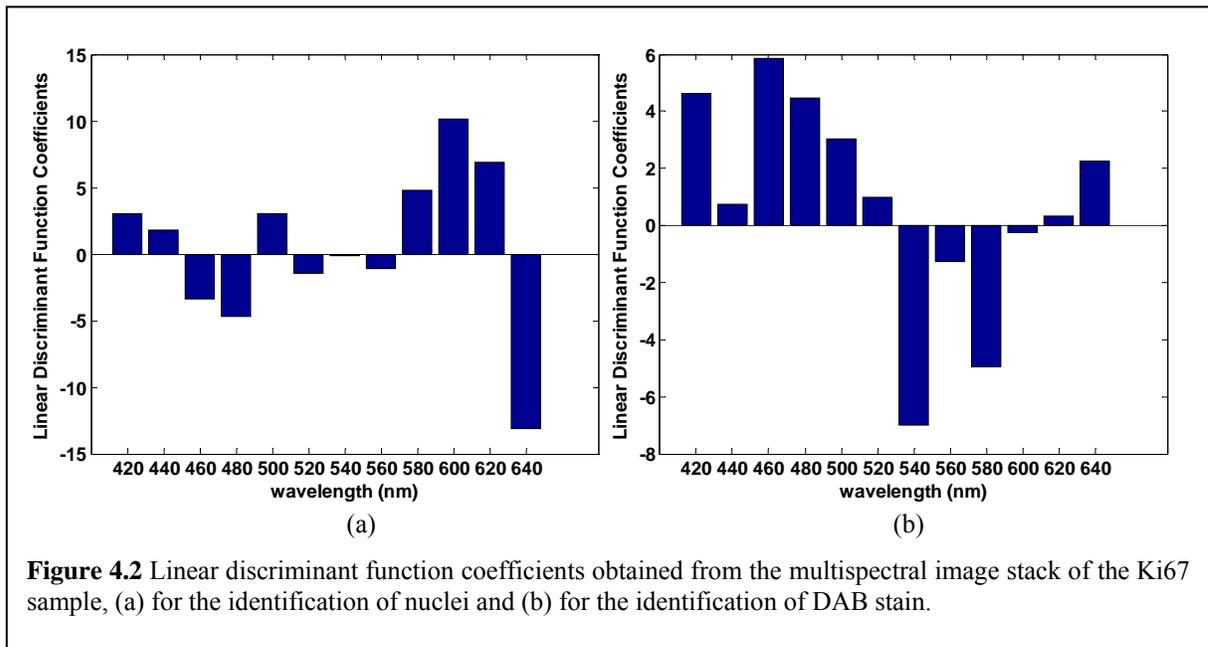
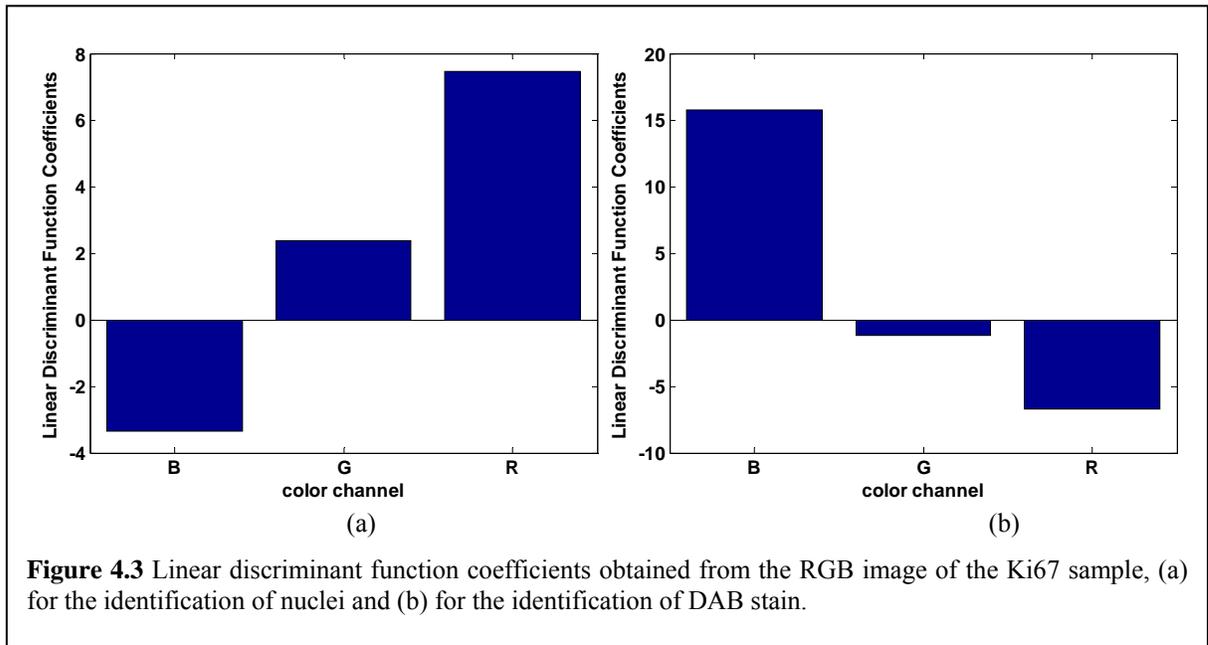
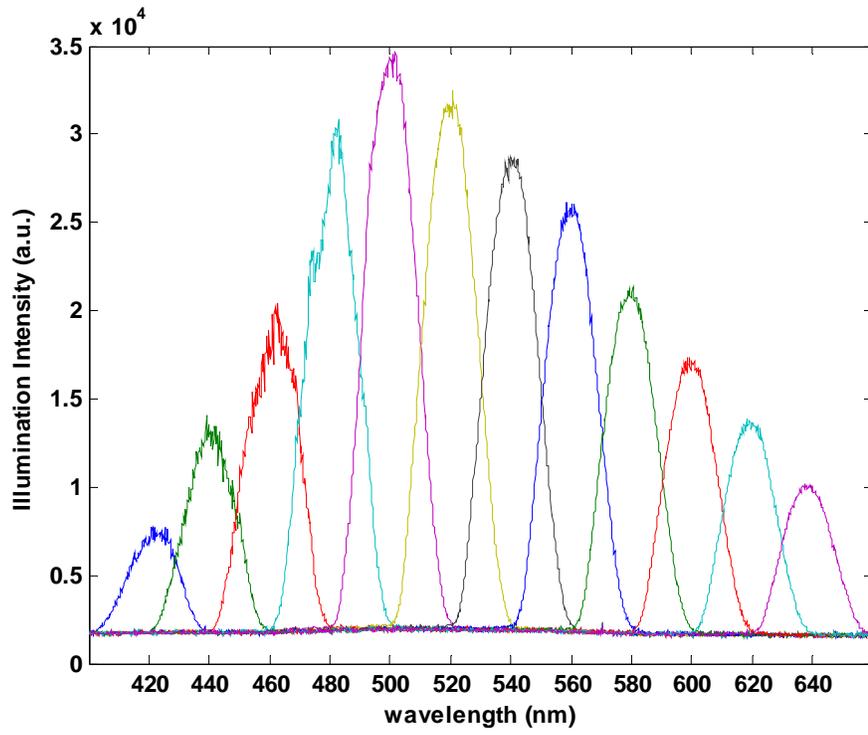
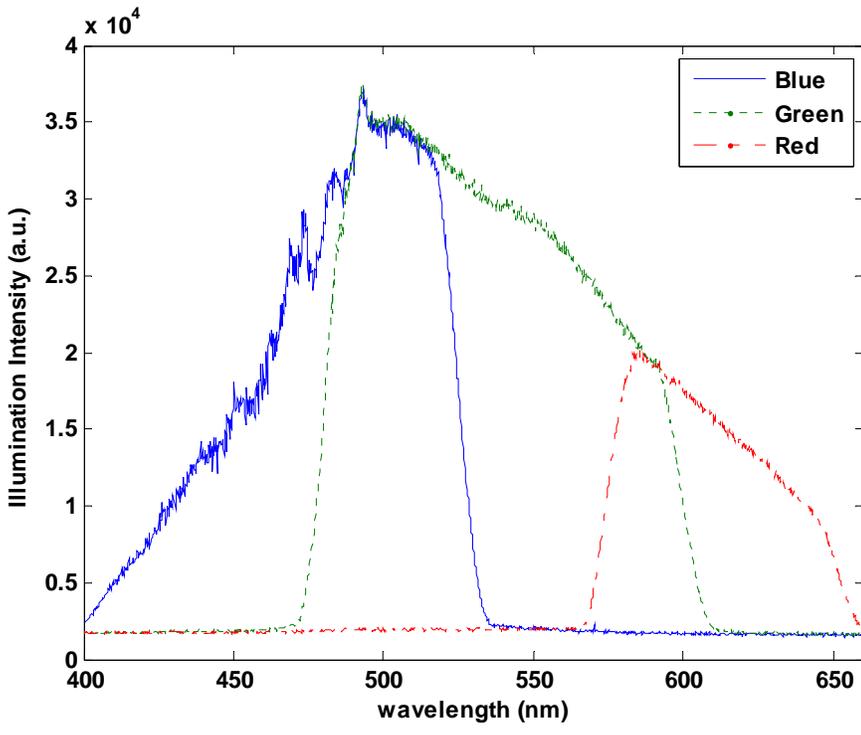


Figure 4.2 Linear discriminant function coefficients obtained from the multispectral image stack of the Ki67 sample, (a) for the identification of nuclei and (b) for the identification of DAB stain.





(a)



(b)

Figure 4.4 Illumination profiles, (a) multispectral, and (b) three-channel RGB.

Since the images that were used for the LDA were flat-field corrected, we used equation 4.12 to calculate the coefficients for the “linear discriminant illumination spectrum”. So, each LDA coefficient corresponding to each wavelength channel was normalized by the total intensity of the illumination profile for that channel. These coefficients were then used to calculate the weighted sum of the illumination profiles. The resulting illumination profiles for the capture of the nuclei image and the DAB image are shown in Figures 4.5 and 4.6, corresponding to the multispectral image stack and the RGB image, respectively.

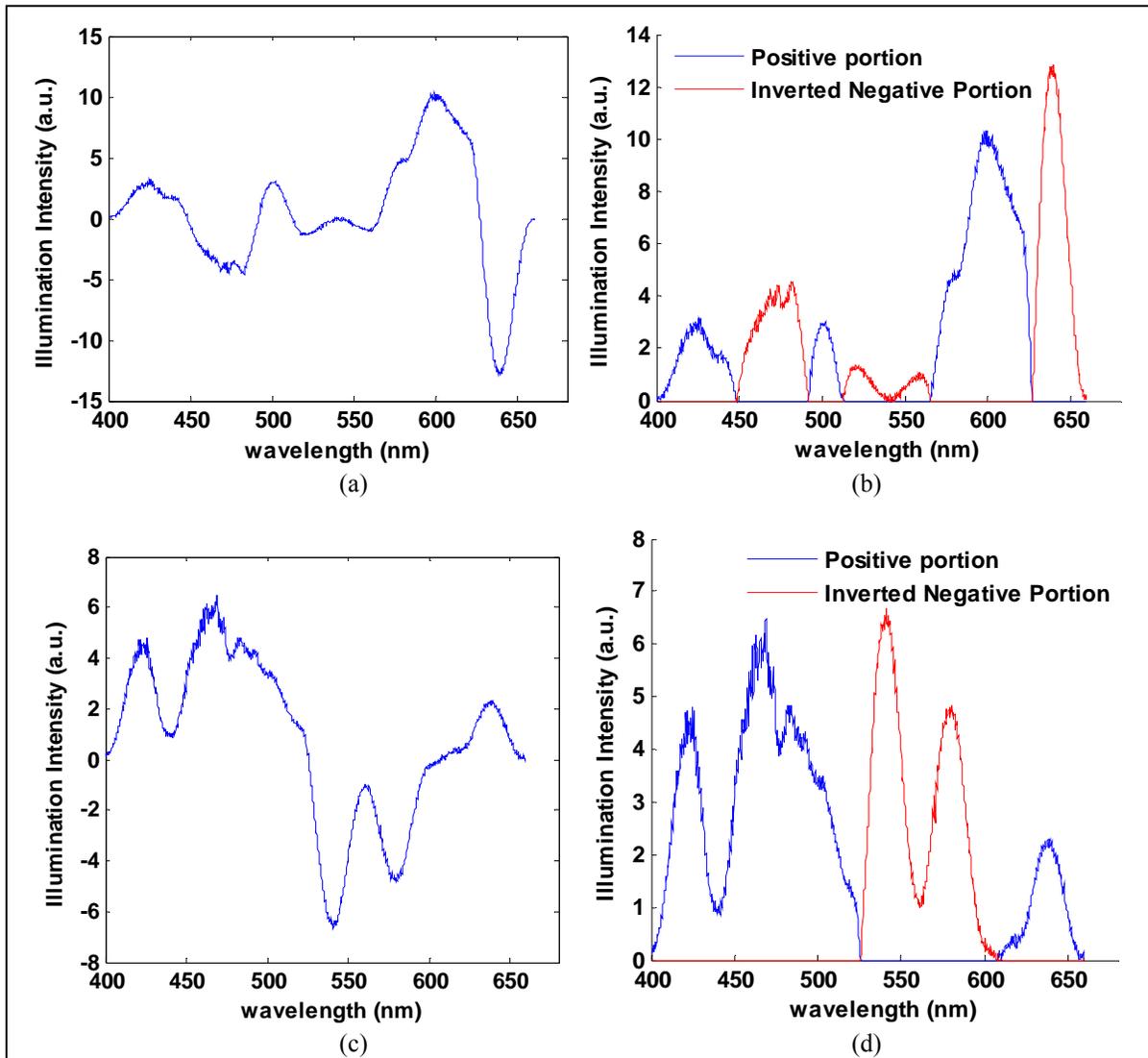
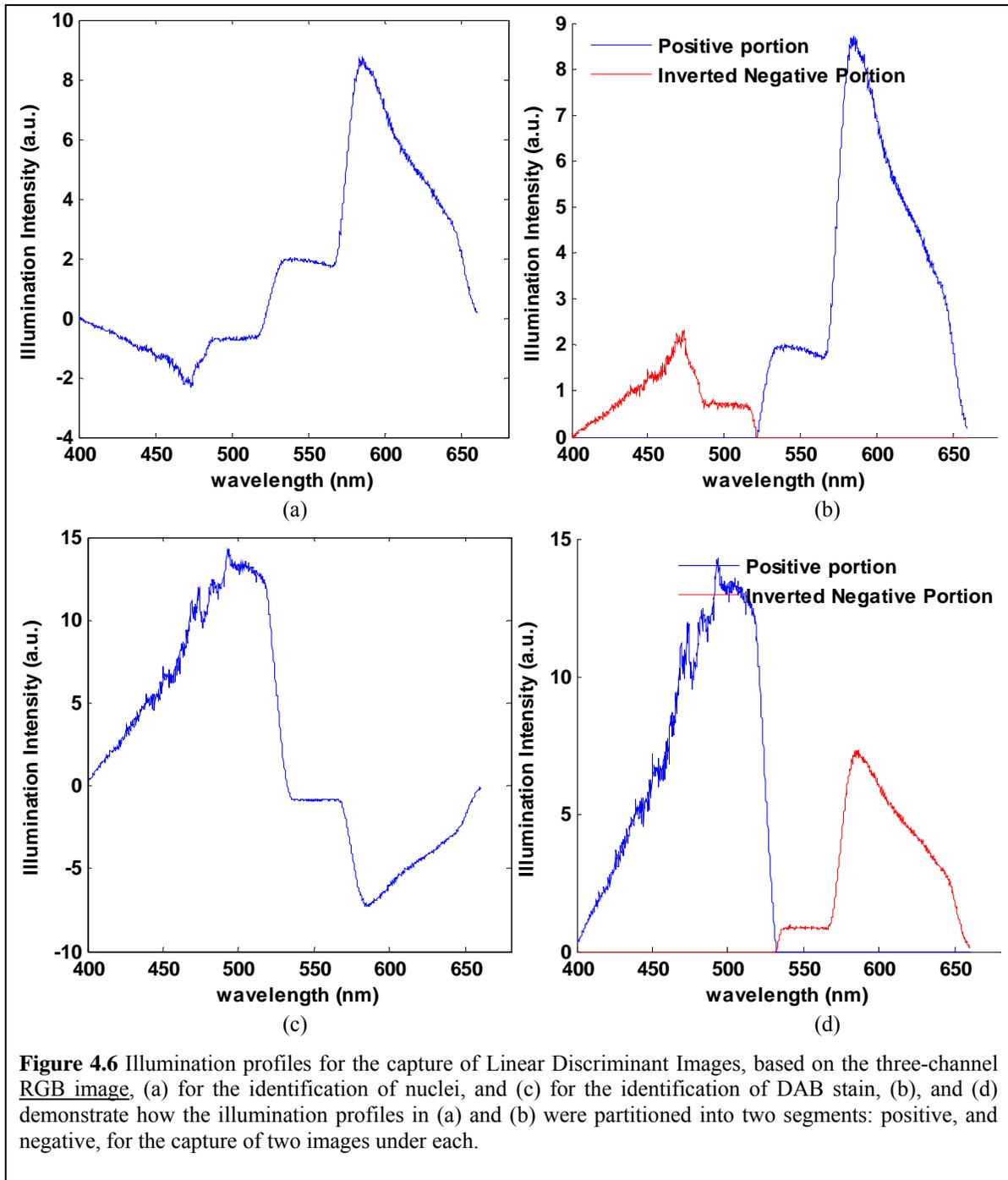


Figure 4.5 Illumination profiles for the capture of linear discriminant images, based on the multispectral image stack, (a) for the identification of nuclei, and (c) for the identification of DAB, (b) and (d) demonstrate how the illumination profiles in (a) and (b) were partitioned into positive and negative segments.



As explained before, two images were captured under the positive and the inverted negative segments of the designed illumination profile. We refer to these two illumination profiles as the “positive” and the “negative” illumination profiles and to the images captured under each as “positive” and “negative” images. The subtraction of the “negative” image from the “positive” image yields the image of the object of interest (nuclei or DAB). However, there are two scaling factors that must be taken into account when subtracting the two images:

1) Exposure-time: The exposure time for the capture of the positive and the negative images are selected to generate the maximum dynamic range, and are not the same. But for the difference image to correspond to the designed illumination profile, the exposure times need to be equivalent. So each image was normalized by its exposure time.

2) OneLight Scaling: The maximum intensity that the OneLight light engine can output at each wavelength is determined by the intensity of its internal Xenon lamp at that wavelength. All spectra generated by OneLight are bounded in intensity by the intensity profile of the lamp. Therefore, once the specification of the desired illumination profile is provided to the OneLight light engine, it automatically scales the illumination to fit it under the illumination profile of the Xenon lamp. This changes the illumination power ratio of the positive and the negative illumination profiles. But for the difference image to correspond to the designed illumination profile, the ratio of the positive and negative illumination profiles should be maintained. So, we determine the scaling factor that transforms the OneLight generated illumination profile to the desired illumination profile and apply that scaling factor to the image captured under the generated illumination. Let the illumination profiles generated by the OneLight light engine be: $\hat{\psi}_{pos}(\lambda)$ and $\hat{\psi}_{neg}(\lambda)$. If \hat{I}_{pos} is the image captured under $\hat{\psi}_{pos}(\lambda)$ and \hat{I}_{neg} is the image captured under $\hat{\psi}_{neg}(\lambda)$, then the “positive” and “negative” images are scaled as follows:

$$\begin{aligned}
 K_{pos} &= \frac{\hat{\psi}_{pos}(\lambda)}{\psi_{pos}(\lambda)} \rightarrow I_{pos} = \frac{\hat{I}_{pos}}{K_{pos}} \\
 K_{neg} &= \frac{\hat{\psi}_{neg}(\lambda)}{\psi_{neg}(\lambda)} \rightarrow I_{neg} = \frac{\hat{I}_{neg}}{K_{neg}}
 \end{aligned}
 \tag{4.15}$$

Once the “positive” and the “negative” images are scaled properly, their subtraction, $I_{pos} - I_{neg}$, corresponds to the image of the object of interest (H or DAB). The illumination profiles, which were calculated based on the multispectral illuminations, for the capture of the first discriminant function image, which identifies the nuclei, are shown in Figure 4.7. Also shown in Figure 4.7 are the illumination profiles generated by the light engine, along with the same generated profiles adjusted for OneLight scaling factor. The captured “positive” and “negative” images under the generated illumination profile are shown in Figure 4.8. The image generated by subtracting the “negative” image from the “positive” image which represents the image of the nuclei is shown in Figure 4.9. Note that without the proper scaling, the difference between the images in Figure 4.8 would not generate the nuclei image of Figure 4.9. Figures 4.10 to 4.12 illustrate the illumination profiles, and the captured images for the identification of DAB.

As stated before, the linear discriminant analysis was performed once on the multispectral image stack, and once on the three channels of the RGB image. Figures 4.13 to 4.16 show the illumination profiles for the identification of nuclei and DAB obtained using the three-channel RGB illuminations and the images captured under these illuminations.

To demonstrate that the LDA coefficients and therefore the illumination profiles for the capture of the LDA images do not significantly change from one sample to the other, we used the same illumination profiles to capture images from a different Ki67 sample, shown in Figure 4.17. The resulting images clearly correspond to the nuclei and the DAB stain. These images are demonstrated in Figures 4.18 and 4.19.

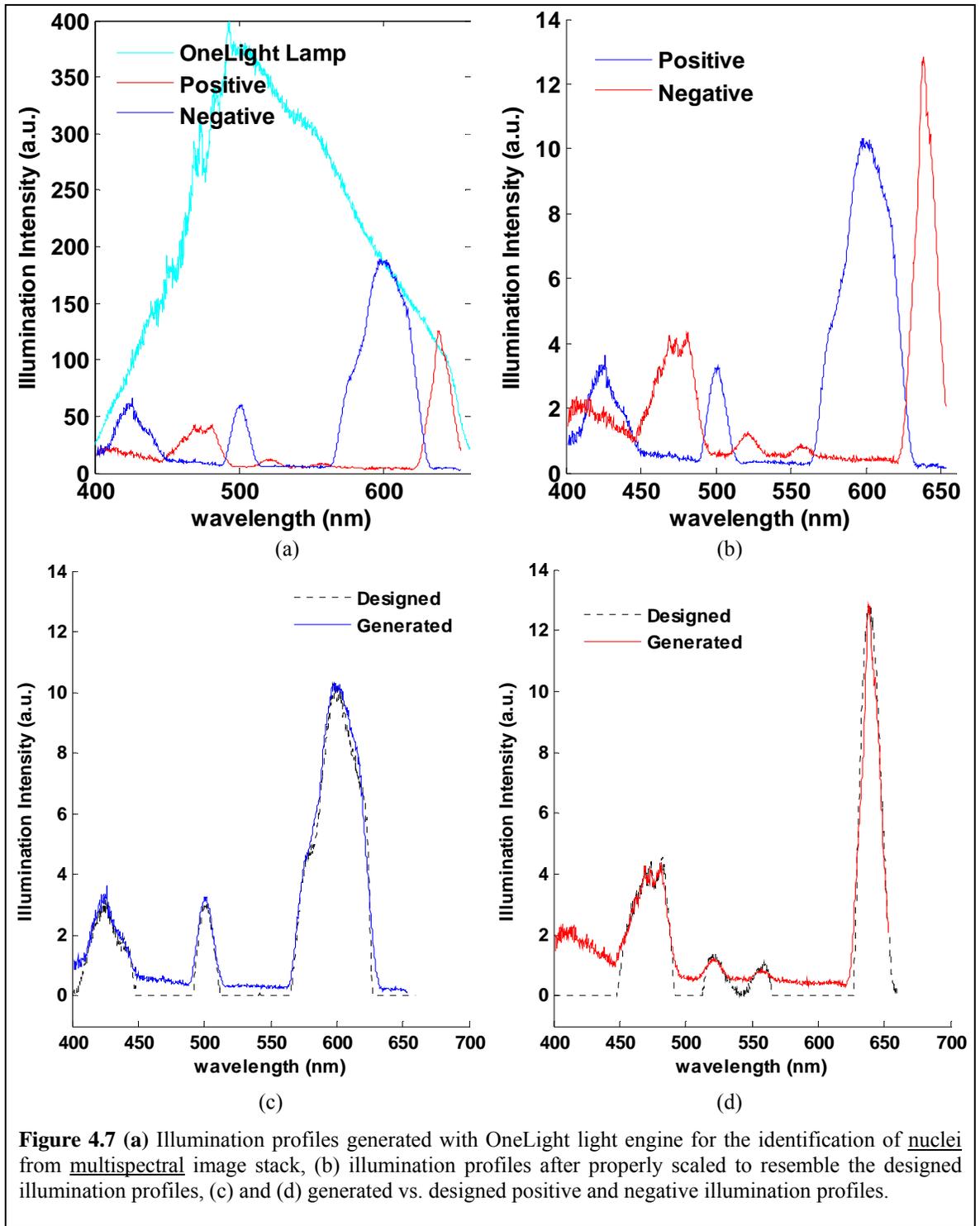
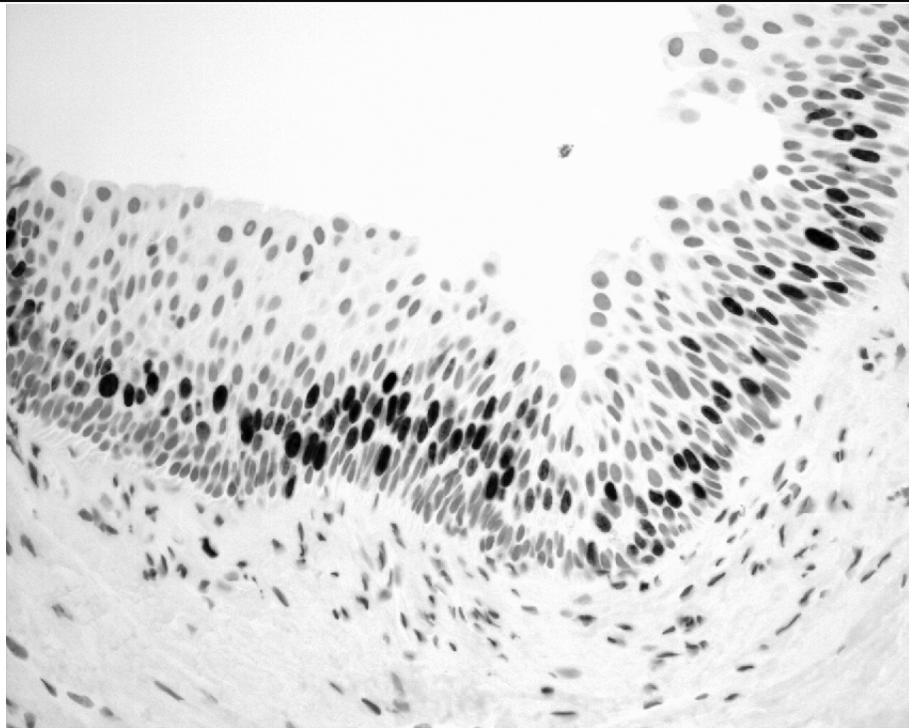
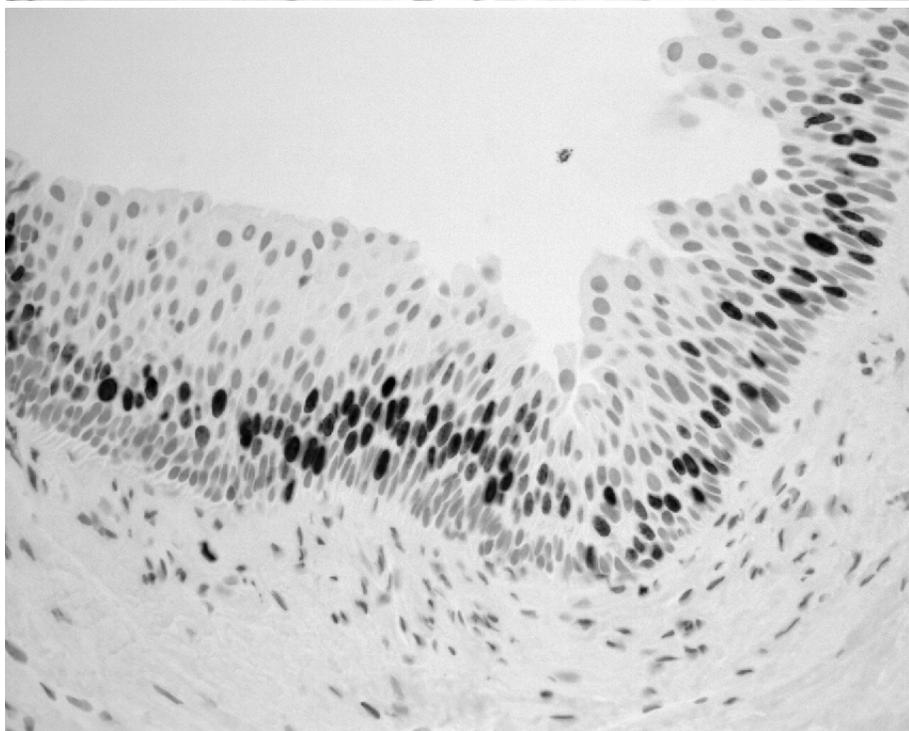


Figure 4.7 (a) Illumination profiles generated with OneLight light engine for the identification of nuclei from multispectral image stack, (b) illumination profiles after properly scaled to resemble the designed illumination profiles, (c) and (d) generated vs. designed positive and negative illumination profiles.



(a)



(b)

Figure 4.8 Capture of the image of nuclei, with illumination generated from the multispectral image stack, (a) image captured under the positive segment of the designed illumination, (b) image captured under the negative segment of the designed illumination.

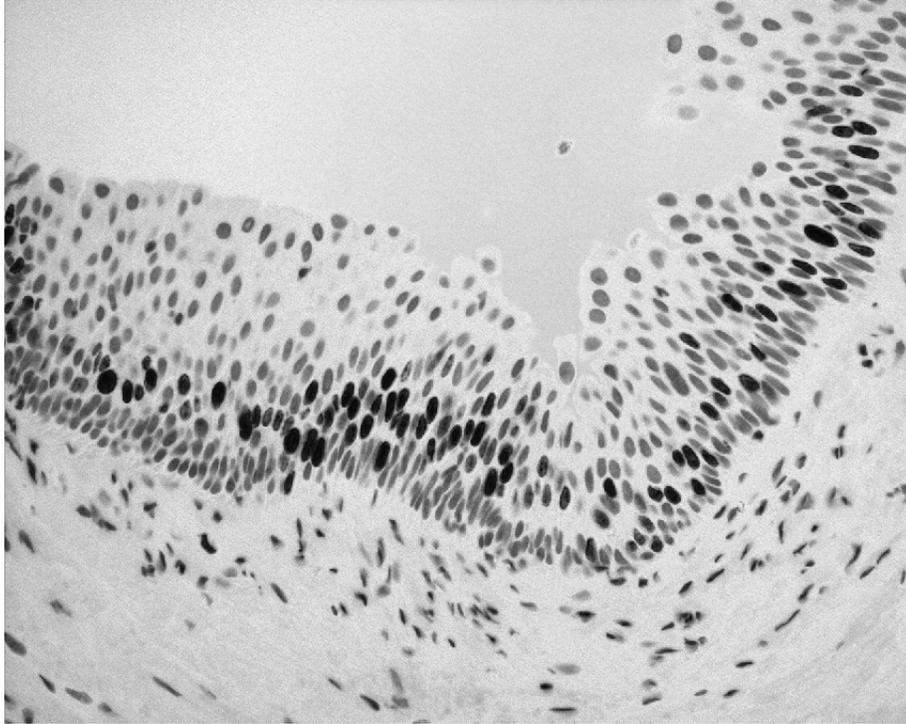


Figure 4.9 Image of nuclei, generated with illumination designed from the multispectral image stack.

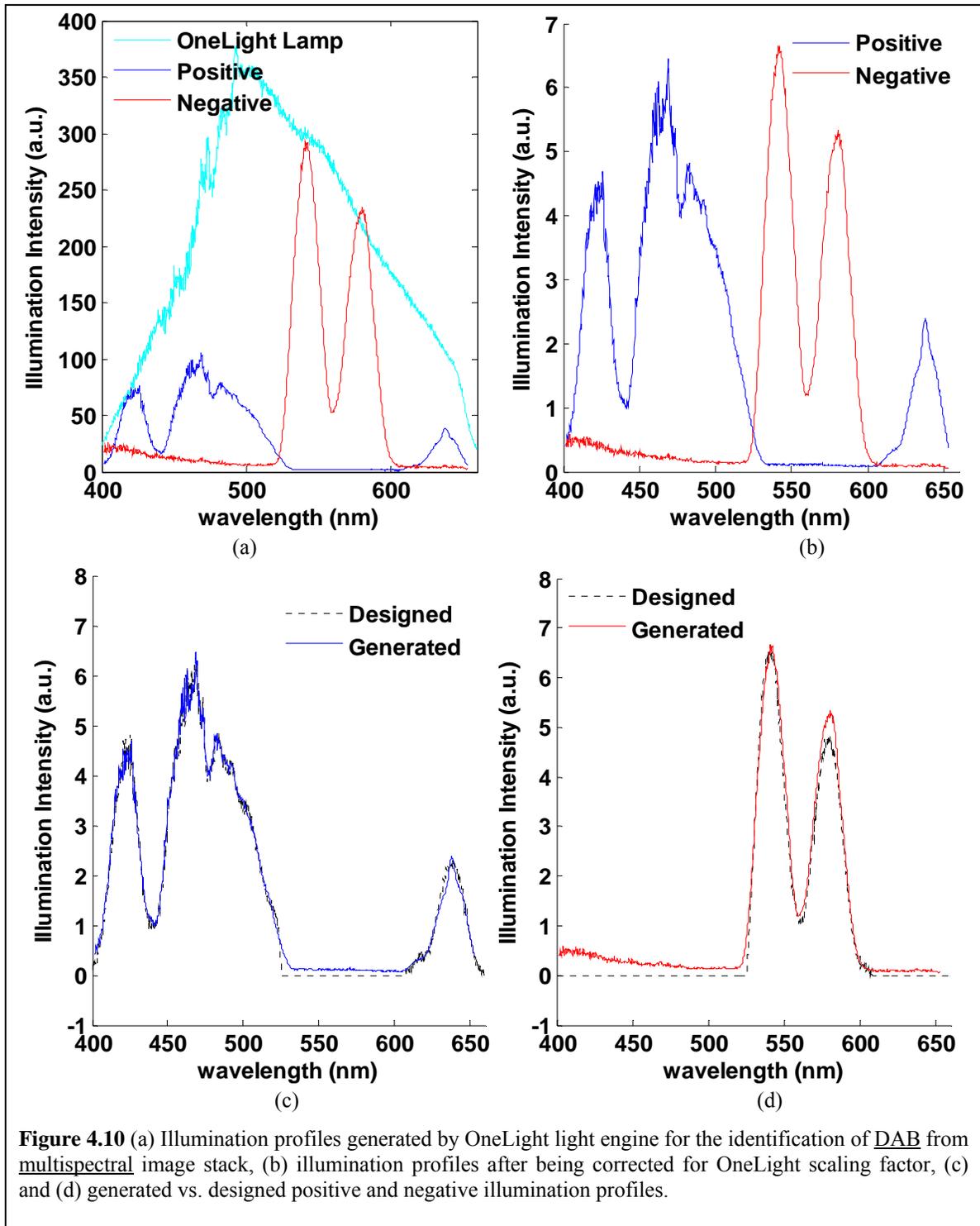
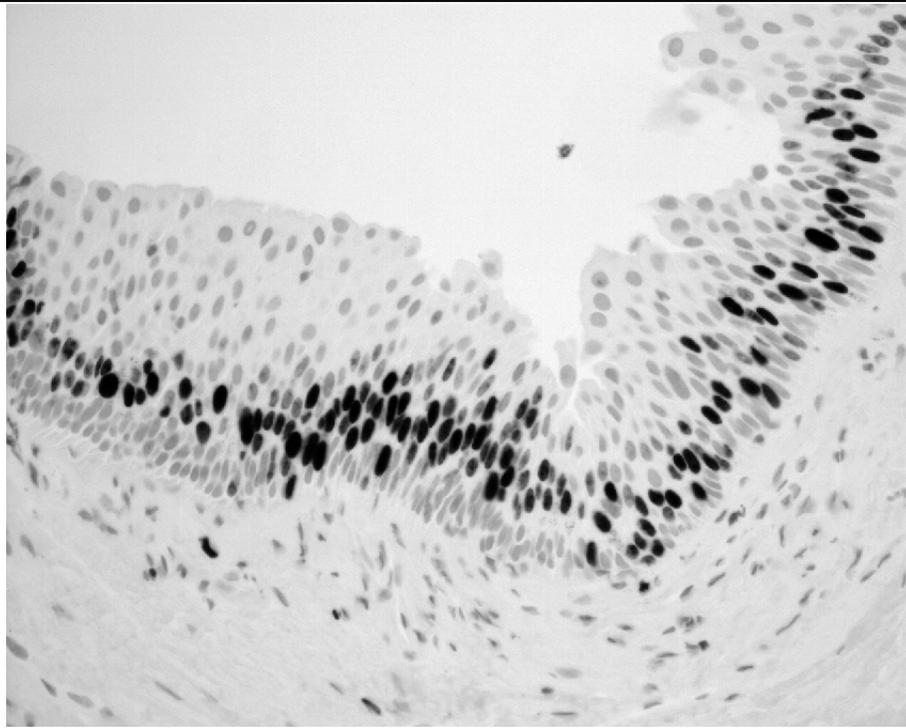
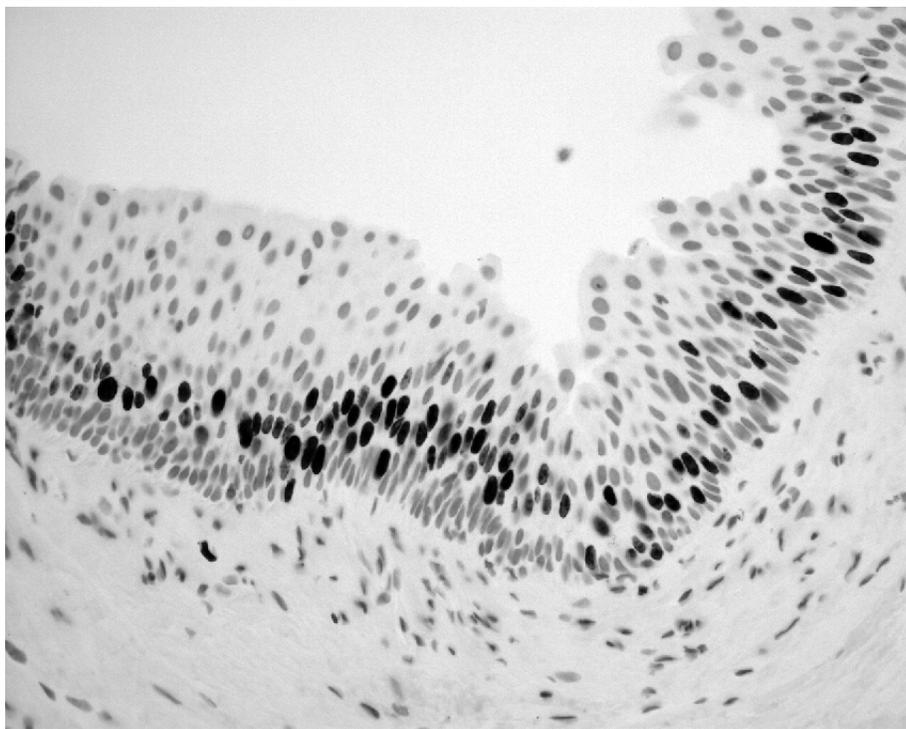


Figure 4.10 (a) Illumination profiles generated by OneLight light engine for the identification of DAB from multispectral image stack, (b) illumination profiles after being corrected for OneLight scaling factor, (c) and (d) generated vs. designed positive and negative illumination profiles.



(a)



(b)

Figure 4.11 Capture of image of DAB, with illumination generated from the multispectral image stack, (a) image captured under the positive segment of the designed illumination, (b) image captured under the negative segment of the designed illumination.

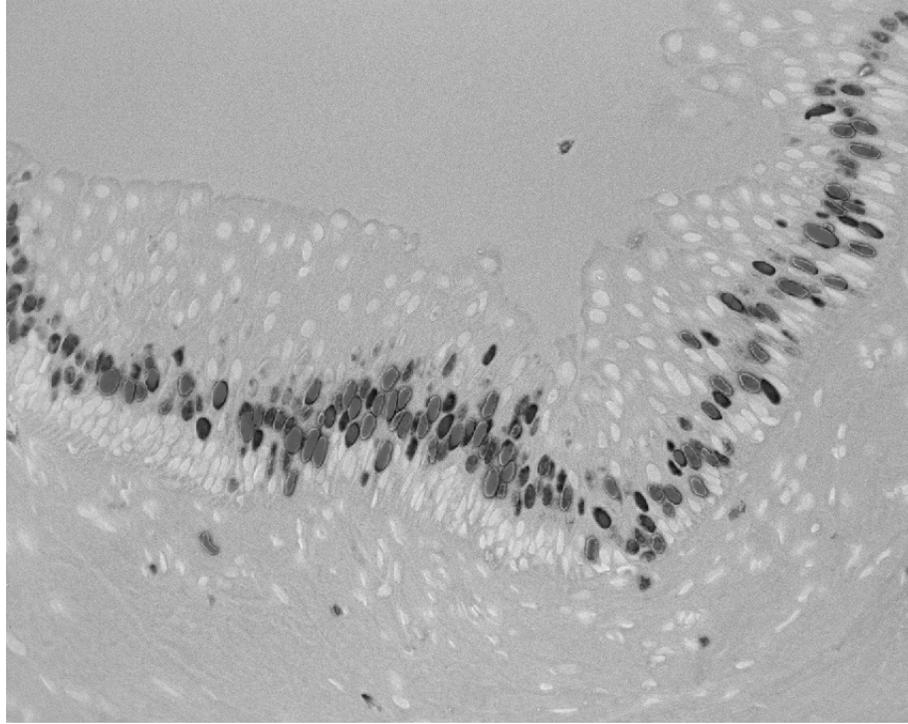


Figure 4.12 Image of DAB, generated with illumination designed from the multispectral image stack.

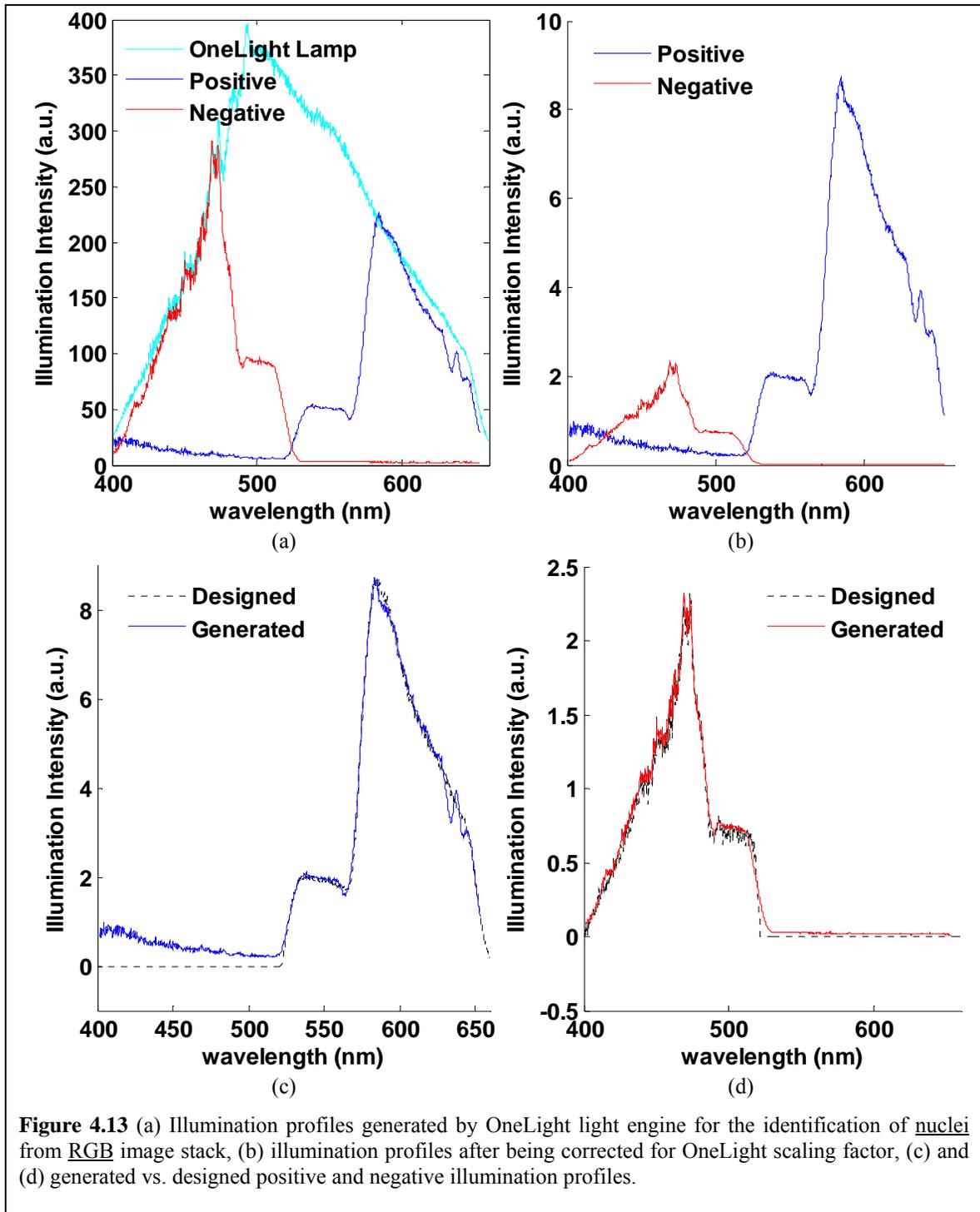


Figure 4.13 (a) Illumination profiles generated by OneLight light engine for the identification of nuclei from RGB image stack, (b) illumination profiles after being corrected for OneLight scaling factor, (c) and (d) generated vs. designed positive and negative illumination profiles.



Figure 4.14 Image of nuclei, generated with illumination designed from the RGB image.

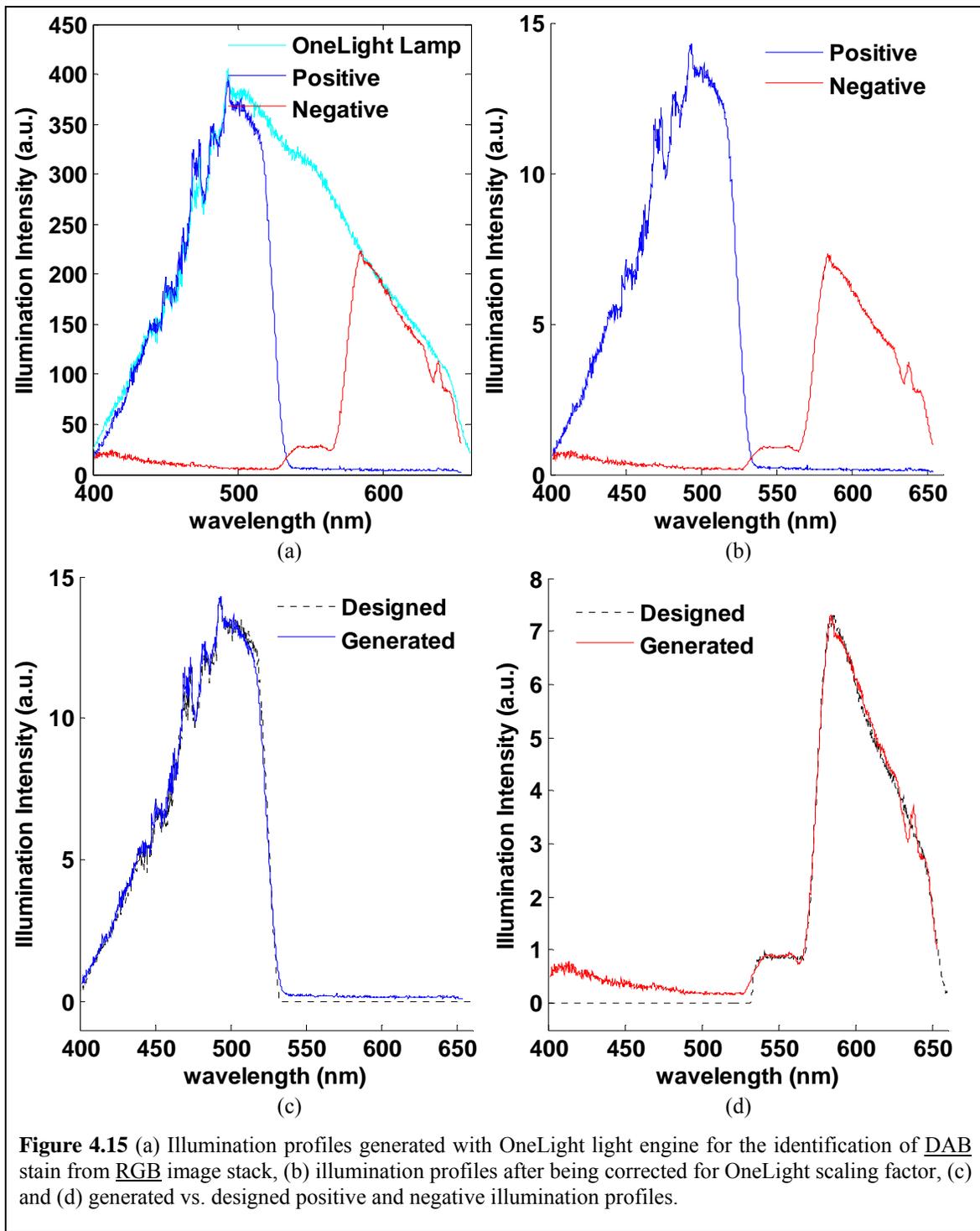


Figure 4.15 (a) Illumination profiles generated with OneLight light engine for the identification of DAB stain from RGB image stack, (b) illumination profiles after being corrected for OneLight scaling factor, (c) and (d) generated vs. designed positive and negative illumination profiles.

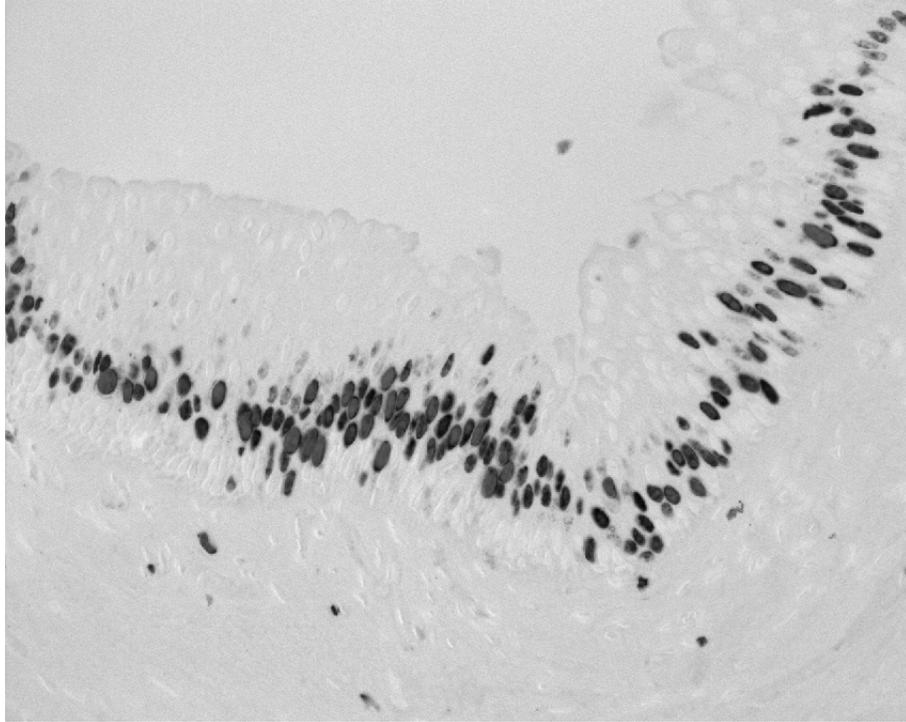


Figure 4.16 Image of DAB, generated with illumination designed from the RGB image.

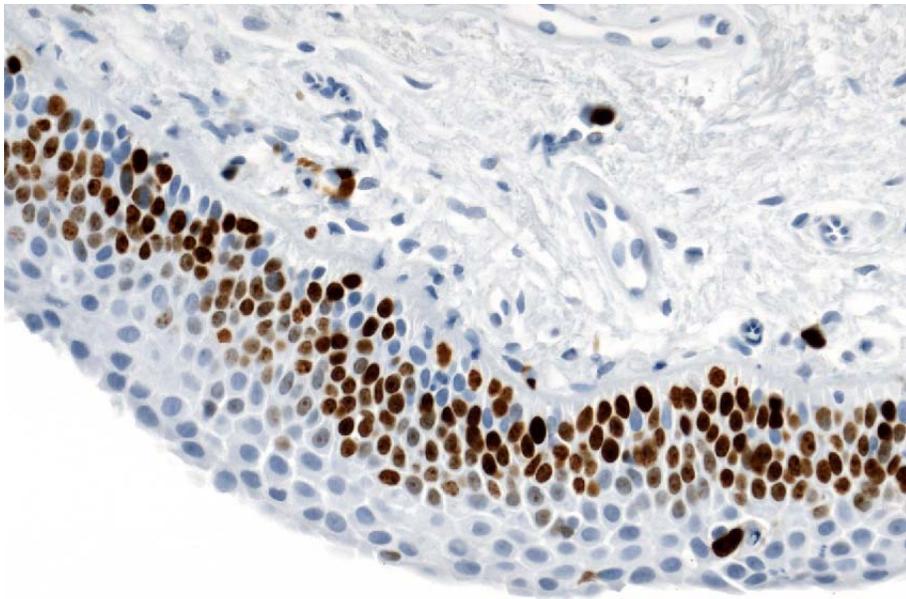


Figure 4.17 RGB image of the second Ki67 immunostained sample.

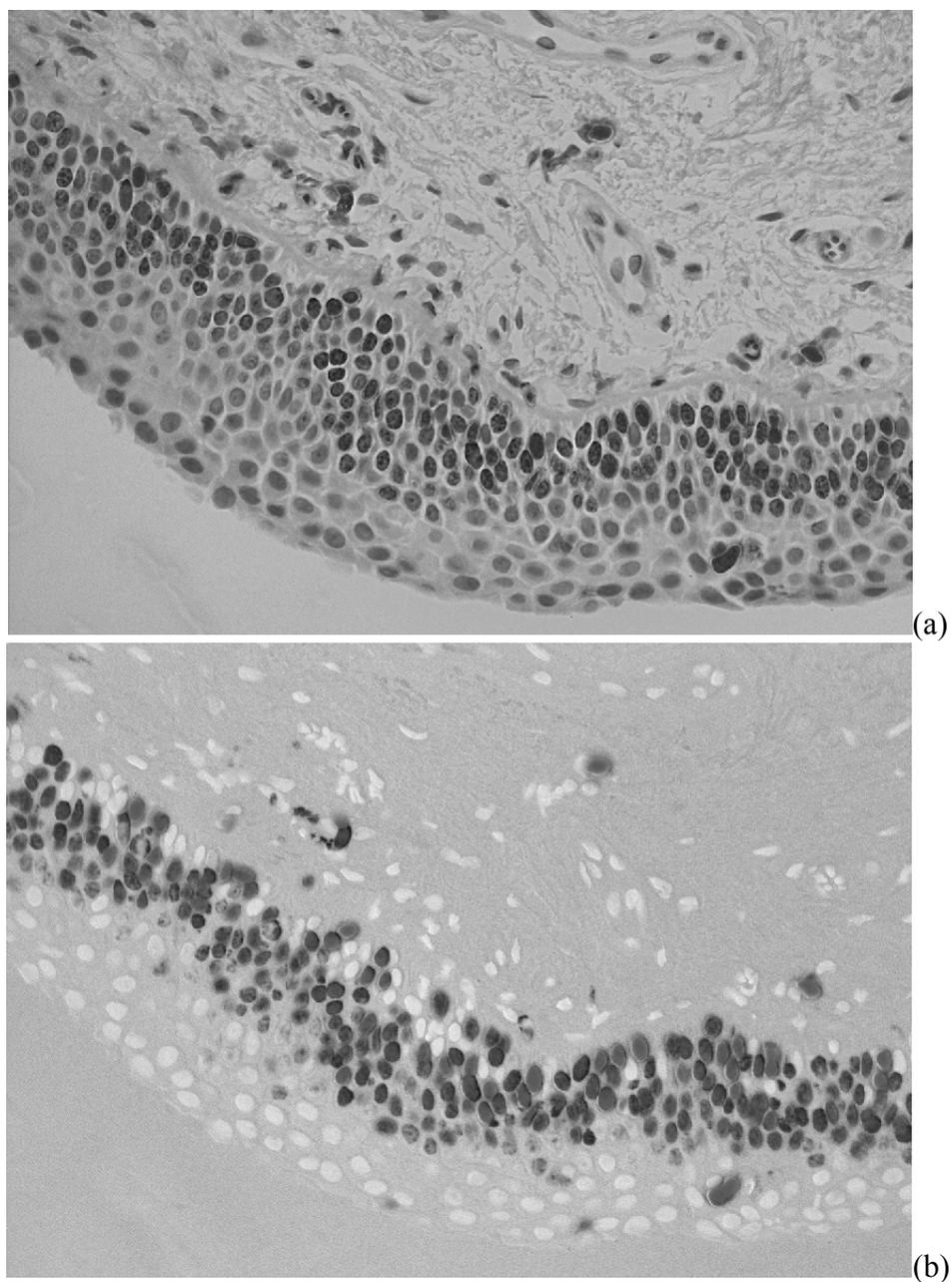


Figure 4.18 Images of nuclei (a) and DAB (b), generated with illumination designed from the multispectral image stack.

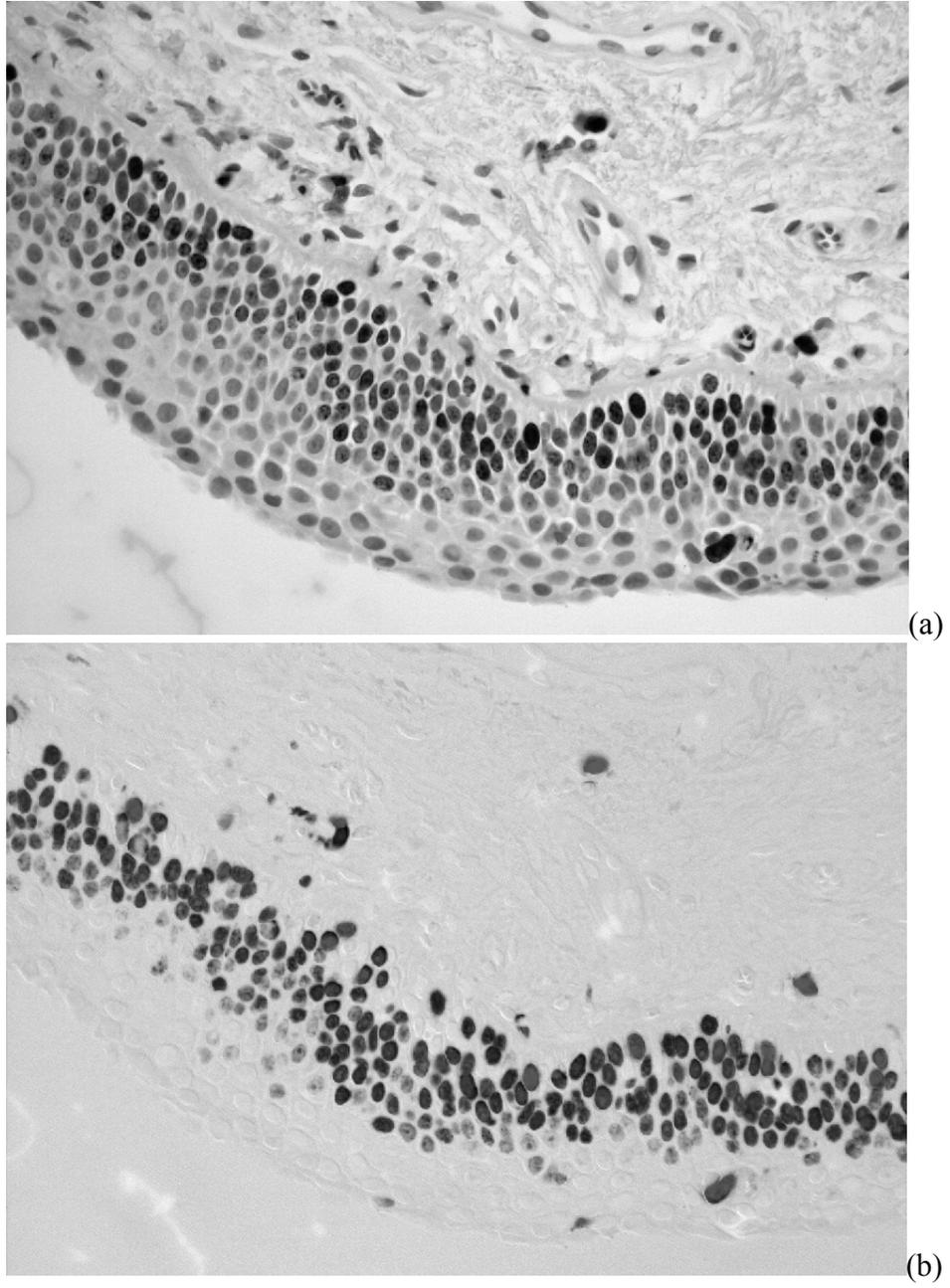


Figure 4.19 Images of nuclei (a) and DAB (b), generated with illumination designed from the RGB image.

4.2.3 Discussion

The Selective Illumination Transmission Imaging approach presented here allows us to directly capture images that are equivalent to a weighted sum of images in a multispectral image stack, using specifically designed and generated illumination. We demonstrated this approach on immunostained tissue sections stained with H and DAB chromogens. While the captured images do highlight the parts of the tissue stained with these chromogens, there are slight differences between the calculated and captured weighted sum images. The difference between the calculated and captured images is caused by the difference between the designed and the generated illumination profiles. The illumination light as a function of wavelength generated by OneLight light engine is never zero. There is always some background light, even when the designed illumination is zero. The difference between the designed and generated illumination light becomes more pronounced through scaling of the images to correct for OneLight's scaling of generated illumination light. Equation 4.15 formulates this scaling factor. When K_{pos} or K_{neg} is less than one, the scaled image has higher intensities than the captured image. This means that the contribution to image intensities from the unwanted "background" light is being amplified. The difference between the calculated and captured images becomes more pronounced as a result.

The images captured under illumination which was calculated using the 3-channel RGB illuminations (Figures 4.14, 4.16, and 4.19) were less noisy and more similar to the calculated LDA images, than those captured under illumination generated using the 12-channel multispectral illuminations (Figures 4.9, 4.12, and 4.18). We speculate that this is because the illumination spectra generated using the RGB illumination channels were broader, since the three illumination channels (red, green, and blue illumination spectra) used for the capture of RGB images are broader than the narrow-band illuminations used for the capture of 12-channel multispectral images. Broader illumination profile means higher light power, and shorter exposure times, which increases the signal to noise ratios. When correcting for the OneLight's scaling factor, the values of K_{pos} and K_{neg} were closer to one, therefore the captured images were not scaled significantly, thus the background light effect was not amplified as much as in the case of the illumination designed using the 12-channel narrow multispectral illumination bands. In addition, the illumination profiles designed based on the

RGB images (shown in Figures 4.13 and 4.15) contain less structure in the illumination profiles (peaks and valleys) compared to the ones based on the multispectral images (shown in Figures 4.7 and 4.10), since they are linear combinations of only three channels as opposed to 12 channels for the multispectral images. Less peaks and valleys result in better fit of OneLight's generated illumination to the designed illumination. This in turn minimizes the differences between the calculated and captured illuminations, and consequently the calculated and captured images.

Based on the above discussion, it is preferable to generate the selective illumination profiles based on as few and broad illumination bands as possible. Of course the choice of the number of the illumination channels and their bandwidth depends on the spectral properties of the labels in the sample. In our experiment, the labels are Hematoxylin and DAB, which have been traditionally used together because they are easily discriminated by the eye and in RGB images. In other words these two chromogens were chosen so that they are distinguishable in an RGB image. Therefore, three broad bands of an RGB image are sufficient for the discrimination of these labels. Thus, selective illumination profiles can be reliably generated using the three-channel RGB illuminations. There may be cases however where more than three channels is necessary, particularly if more than two labels are present in the sample.

We successfully applied the illumination profiles designed for one Ki67-immunostained sample to capture images from another Ki67 sample. This verifies that the LDA coefficients, and therefore the illumination profiles, are the same for the samples that are prepared and stained with the same protocol and labels. We have also demonstrated this fact in chapter 3, where we trained the linear discriminant functions on a training set comprised of data from several samples, and showed that the discriminant functions resulted in satisfactory contrasts between objects of interest. Therefore, once the LDA coefficients are learnt on a training set, the illumination profiles generated based on those coefficients can be used to capture "LDA images" of other samples with the same preparation and staining procedures.

4.3 Fluorescence Excitation Imaging

4.3.1 Background on Fluorescence Imaging

Fluorescence imaging is being applied for tissue diagnostics in a wide variety of clinical and preclinical situations.^{33,146-150} In the past decade, multispectral fluorescence imaging has been a rapidly expanding field.^{113,151,152} Multispectral imaging combines spectroscopy with single wavelength imaging and generates a data cube of spectrally resolved image information across wavelengths of interest. Multispectral imaging systems typically involve optical filter wheels or electronic filters such as acousto-optical tunable filters or liquid crystal tunable filters in the optical path.

Multispectral fluorescence imaging has been applied to multi-probe analysis in microscopy and *in vivo* imaging (including small animal imaging). In multi-probe analysis, multispectral imaging enables separation of multiple fluorophores from each other as well as from any autofluorescence that overlap both spectrally and spatially, thus removing crosstalk amongst the different sources of fluorescence.^{116,153}

Multispectral autofluorescence imaging is also advantageous for *in vivo* imaging. The degree to which the fluorescence image appearance changes *in vivo* due to metabolic, morphologic, and structural changes depends strongly on the excitation light used and emission wavelength(s) detected. Thus there maybe one or more excitation or emission wavelengths that highlight changes of interest in a specific tissue. Multispectral imaging enables the detection and application of such wavelengths.^{146,154,155}

There are many devices for the detection and characterization of early neoplastic lesions using *in vivo* fluorescence spectroscopy and imaging, some of which have been adopted clinically. These include: 1) the Xillix Onco-LIFE and Pentax SAFE-3000 devices for detecting dysplasia and carcinoma in situ (CIS) in the central airways of the lung (and also for staging prior to surgery looking for synchronous lesions), and 2) the VELScope (LED dental, Burnaby, BC, Canada) and Identafi 3000 (Trimira, Houston, Texas) for direct fluorescence visualization of the oral cavity in order to detect lesions at risk of transforming into cancer and determining the extent of the surgical field of oral cancer containing the at risk tissue needing to be treated. The cervix, bladder, GI tract, esophagus and skin are also

areas of active research and clinical development for the use of fluorescence imaging by many groups.¹⁵⁶⁻¹⁶⁰

For wider acceptance of fluorescence imaging of tissue *in vivo*, improvements are needed. Barriers to wider use include the high false positive rate when used in IEN (intra-epithelial neoplasia) screening settings (due to low prevalence of the disease in screening populations) and the confounding effects of inflammation (as in almost all *in vivo* tissue autofluorescence settings inflammation can appear as a false positive). Additionally, significant clinical skill can be required to recognize true positive areas from confounding factors such as inflammation, infection, ulceration, pigmentation, bruising, normal changes in tissue structure (e.g. the transformation zone of the cervix), and tissue transformations (e.g. due to Barrett's esophagus or lichen planus in the oral cavity).

Tissues are a mixture of several fluorophores with different concentrations and at different depths. Consequently, tissue autofluorescence changes with tissue morphology, vasculature, and biochemical changes caused by neoplastic processes. The degree of this change is strongly wavelength dependent, so there may be regions in the fluorescence excitation emission matrix (EEM) of the tissue where changes are more pronounced.¹⁴⁷

Almost all previously described *in vivo* fluorescence imaging techniques make use of a single excitation wavelength band and detection of one or more emitted wavelengths. Alternatively, they may rely on the sequential application of a few excitation wavelength bands through filter wheels and the synchronized detection of one or more emitted wavelengths. This allows for detection of the differences in the emission profile of the tissue fluorophores (rows in the EEM).

Central to the work presented here is the extraction of independent (and potentially useful) information in the differences in the excitation profiles of tissue (columns in the EEM). As an example, Figure 4.20 illustrates the EEM of the cervical tissue along with an excitation profile and an emission profile at two arbitrary emission and excitation wavelengths. The addition of multispectral excitation data alone or together with the multispectral emission data may facilitate detection of areas of pathological changes in the tissue or additional fluorophores in microscopy. Put another way, images of individual fluorophore distributions within a mixture of fluorophores may be detected based upon their absorption/excitation spectra by using a multitude of illumination wavelengths. This can be

done by sequential illumination followed by calculating a weighted sum of the detected images which highlights specific desired fluorophores.

Qu et al.¹⁶¹ proposed an optical processing method based on Principal Component Analysis for the classification of autofluorescence and the characterization of tissue. This was done by designing optical filters to process autofluorescence signals optically and generate principal component images of the examined tissue simultaneously. The major advantage of this method is that it completely eliminates the dispersion instrument and computer processing required to generate principal component scores.

In this work, we propose the use of specifically designed illumination spectra to capture an image of a sample that corresponds to a linear combination of the images in the multispectral excitation image stack. Similar to the concept of PCA imaging proposed by Qu et al. this method removes the need for computer processing. The difference is that instead of designing optical filters to filter the light emitted from a sample, we design illumination spectra to excite the fluorescent sample with. Our proposed selective spectral illumination imaging technique is not specific to PCA imaging. Illumination spectra can be designed to capture any linear combination of the images in the multispectral excitation image stack.

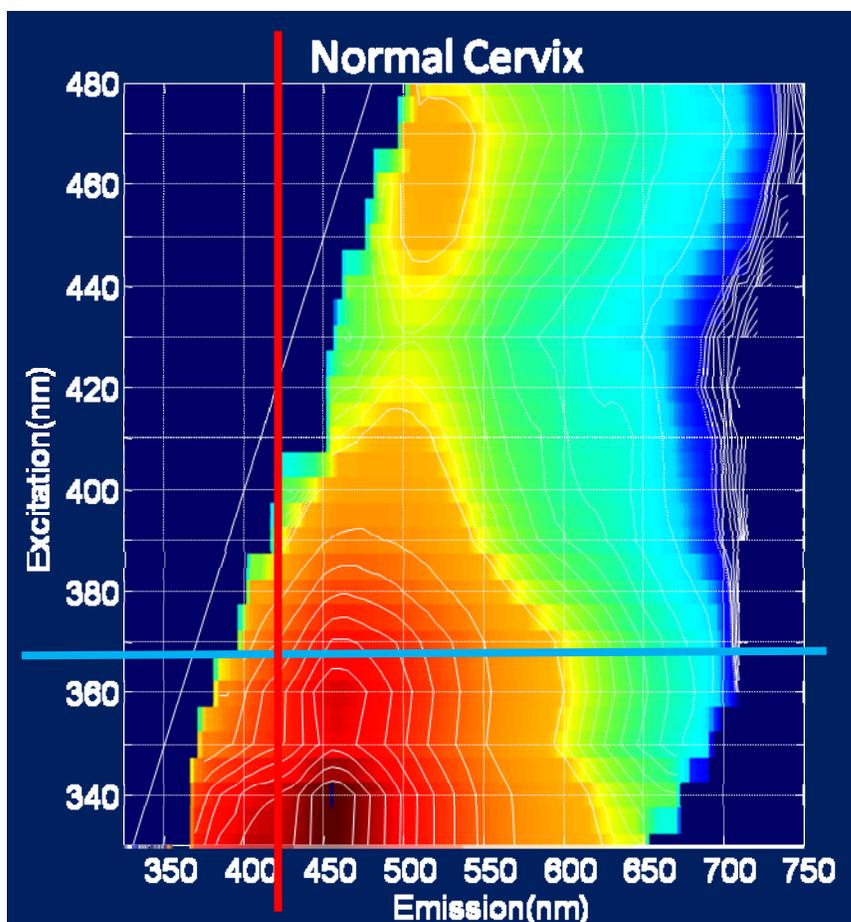


Figure 4.20 EEM of normal cervical tissue. Blue horizontal line represents an emission profile, and red vertical line represents an excitation profile. Photo obtained with permission from private communication with Dr Calum MacAulay and Dr Michelle Follen.

The hypotheses of this work are:

1) We can unmix the total fluorescence signal detected from the sample to the fluorescence signals from each of the distinct fluorescing components comprising the sample, using the excitation image stack, based on the difference in the excitation spectra of the underlying fluorophores, or the same fluorophore attached to different material.

2) If a fluorescence signal of interest can be generated by a weighted linear combination of images in the excitation image stack, then this specific fluorescence signal could be directly recorded by simultaneously illuminating the tissue sample with a specifically determined weighted sum of illumination wavelengths and then detecting the emitted light.

4.3.2 Multispectral Excitation Imaging

We present a multispectral excitation imaging method that is based on multi-channel illumination generated by a spectrally programmable light engine. Using this setup, excitation image stacks can be collected in a rapid and efficient manner, similar to the collection of emission image stacks using tunable filters on the detection side of the imaging system. Spectral unmixing techniques can be performed on the excitation image stack as they can be performed for an emission image stack; the difference being that unmixing is based on the differences in the emitted images that are caused by the variable excitation of fluorophores at different wavelengths. This enables detection of changes in the absorption profile of the fluorescent material.

The multispectral excitation imaging system consists of a spectrally programmable light source that projects excitation light with a predetermined intensity-wavelength profile onto the sample through a short-pass excitation filter (because the OneLight system by itself has insufficient out of band rejection for fluorescence imaging). A lens system collects the light emitted from the sample. This light passes through a long-pass detection filter and is then recorded by a monochrome CCD camera. A schematic diagram of the system is shown in Figure 4.21.

We developed custom software to control and synchronize the light engine with the CCD camera. The wavelength range of the excitation light, number of excitation channels, and the bandwidth of each excitation channel are specified through the software user interface. The light engine can sequentially create a narrow-band excitation illumination band and for each excitation band, a fluorescence image is captured. A stack of excitation images are thus generated.

In the multispectral excitation imaging system explained above, a tunable filter may be used after the long-pass detection filter. This allows collection of fluorescence images at different emission wavelengths for each one of the excitation wavelengths used. This corresponds to collecting an EEM at every pixel of the image, and results in a two-dimensional stack of excitation-emission images. In our prototype system, we used the CRI's VariSpec LCTF (CRI, MA, USA) as the tunable detection filter.

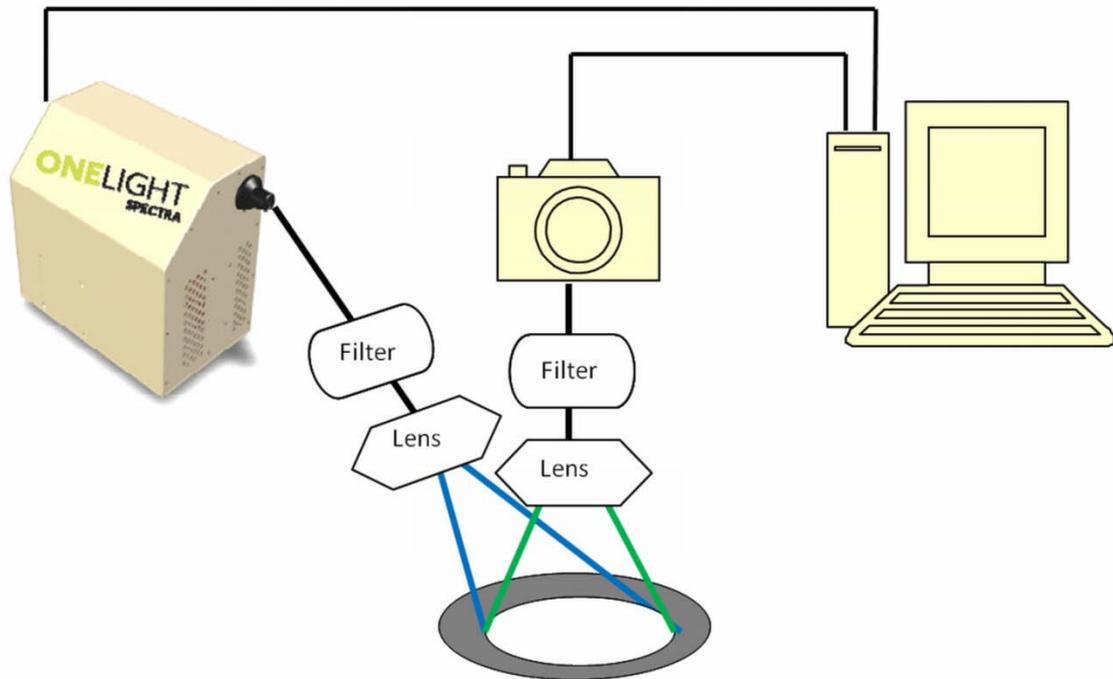


Figure 4.21 Schematic diagram of the system setup.

Throughout this chapter we use PCA as a way to reduce the dimensionality of the multispectral data. PCA is a widely used technique for finding patterns in high-dimensional data.¹⁶² Applied to the stack of excitation images captured across multiple wavelengths, PCA finds several weighted sums of these images, the so called principal components, with the following property: the first principal component accounts for as much variability across the images as possible (which translates to image contrast in this context), and each succeeding component accounts for as much of the remaining variability as possible. The number of principal components is equal to the number of the images in the multispectral image stack; however, most of the information is usually contained in the first few principal components.

Any three principal components may be used to generate a composite RGB image by assigning one principal component image to each color channel. Different colors in different regions of the RGB image generated this way correspond to the differences in the excitation spectra of the underlying objects. Most of the information is usually contained in the first few principal components, therefore one way to visualize most of the information in the image cube is to assign the first three principal components to the RGB image channels. We refer to

the RGB image generated using the first three principal components as the “PCA composite image”.

4.3.3 Selective Excitation Light Fluorescence (SELF) Imaging

According to the Beer-Lambert law, when light with intensity I_0 travels through a substance, the transmitted light I can be written as:

$$I = I_0 e^{-\epsilon l c} \quad (4.16)$$

where ϵ is the molar absorptivity of the material, l is the path length of light, and c is the molar concentration of the absorbing material. The optical density (absorption) A and absorbed light $I_{absorbed}$ can therefore be written as:

$$I_{absorbed} = I_0 - I = I_0(1 - e^{-\epsilon l c}) \quad (4.17)$$

$$A = -\log\left(\frac{I}{I_0}\right) = \epsilon l c \quad (4.18)$$

For fluorescent material, assuming that the optical density ($\epsilon l c$) is very low (less than 10), the absorbed light $I_{absorbed}$ can be approximated as:

$$I_{absorbed} = I_0(1 - (1 - \epsilon l c)) = I_0 \epsilon l c \quad (4.19)$$

The emitted fluorescence light intensity $I_{emitted}$ can be written as a product of the absorbed light intensity $I_{absorbed}$ and the quantum efficiency Φ :

$$I_{emitted} = \Phi I_{absorbed} \stackrel{\text{Eq. 4.19}}{=} \Phi \epsilon l c I_0 \quad (4.20)$$

Therefore, for low absorbing fluorescent material, the emitted light intensity is linearly related to the exciting light intensity.

Based on equation 4.20 and taking the excitation and emission wavelengths into account, the fluorescence emission intensity $E(\lambda_e, \lambda_m)$ of the emitted light from a low absorbing fluorescent object at any excitation wavelength λ_e and emission wavelength λ_m is given by:¹⁶³

$$E(\lambda_e, \lambda_m) = \Phi(\lambda_m) \varphi(\lambda_e) \epsilon(\lambda_e) l c \quad (4.21)$$

where $\varphi(\lambda_e)$ is the intensity of the excitation light, $\Phi(\lambda_m)$ is the quantum efficiency of the sample at the excitation wavelength of λ_e and emission wavelength λ_m , $\varepsilon(\lambda_e)$ is the molar absorptivity, l is the path length of light, and c is the molar concentration of the fluorophore.

Let's assume a fluorescent sample is sequentially excited with L narrow-wavelength bands of light $\varphi_i(\lambda_e)$, and emits $E_i(\lambda_e, \lambda_m)$, $i = 1, 2, \dots, L$, where λ_e is the wavelength of the excitation light, and λ_m is the wavelength of the emitted light. At each excitation band, an image I_i of the sample is captured. The set of images, I_i , $i = 1, 2, \dots, L$, form the multispectral excitation images of the sample. For every pixel with coordinates (x, y) , $\{I_i(x, y)\}$, $i = 1, 2, \dots, L$, is the excitation profile of pixel (x, y) . The output of the camera at the i -th illumination band is the i -th excitation image I_i , and can be described as:

$$I_i = t_i \int_{\lambda_e} \int_{\lambda_m} E_i(\lambda_e, \lambda_m) S(\lambda_m) d\lambda_m d\lambda_e + D_i \quad (4.22)$$

where $S(\lambda_m)$ is the spectral response of the detector, and t_i is the exposure time at the i -th band. D_i is the captured image when the excitation light is turned off. From substituting equation 4.21 into equation 4.22, it follows:

$$I_i = t_i \int_{\lambda_e} \int_{\lambda_m} (\Phi(\lambda_m) \varphi_i(\lambda_e) \varepsilon(\lambda_e) l c) S(\lambda_m) d\lambda_m d\lambda_e + D_i \quad (4.23)$$

If D_i is significantly lower than the captured image of the sample, we can assume $D_i = 0$. Then, the relationship between the incident light $\varphi_i(\lambda_e)$ and the detected fluorescence intensity I_i at each pixel becomes linear. If the incident light is $\psi(\lambda_e) = \sum_{i=1}^L c_i \varphi_i(\lambda_e)$, the image F captured under this illumination can be expressed as:

$$\begin{aligned}
F &= t \int_{\lambda_e} \int_{\lambda_m} \Phi(\lambda_m) \psi(\lambda_e) \varepsilon(\lambda_e) l c S(\lambda_m) d\lambda_m d\lambda_e \\
&= t \int_{\lambda_e} \int_{\lambda_m} \Phi(\lambda_m) \left(\sum_{i=1}^L c_i \varphi_i(\lambda_e) \right) \varepsilon(\lambda_e) l c S(\lambda_m) d\lambda_m d\lambda_e \\
&= t \int_{\lambda_e} \int_{\lambda_m} \left(\sum_{i=1}^L \Phi(\lambda_m) c_i \varphi_i(\lambda_e) \varepsilon(\lambda_e) l c S(\lambda_m) \right) d\lambda_m d\lambda_e \tag{4.24} \\
&= t \sum_{i=1}^L c_i \left(\int_{\lambda_e} \int_{\lambda_m} \Phi(\lambda_m) \varphi_i(\lambda_e) \varepsilon(\lambda_e) l c S(\lambda_m) d\lambda_m d\lambda_e \right) \\
&\quad \stackrel{\text{Eq (4.18)}}{=} t \sum_{i=1}^L c_i \frac{I_i}{t_i} = t \sum_{i=1}^L v_i I_i
\end{aligned}$$

where t is the exposure time for the capture of the image F , and $v_i = c_i/t_i$.

Let us assume that a weighted sum of the images in the multispectral image stack are obtained, using some mathematical or experimental technique, that transform the multispectral image stack into an image F . For example, Image F may be designed to represent the distribution of a certain component (e.g. a fluorophore) in the sample, to maximize the contrast between a certain component and the rest of the sample, or to contain the largest possible variance. Image F can be expressed as follows:

$$F = \sum_{i=1}^L v_i I_i \tag{4.25}$$

where v_i , $i = 1, 2, \dots, L$ are the coefficients to form the weighted linear combination of excitation images. Then according to equation (4.24), the image F can be captured directly using the following illumination:

$$\psi(\lambda) = \sum_{i=1}^L c_i \varphi_i(\lambda) \quad \text{where} \quad c_i = v_i t_i \tag{4.26}$$

We refer to this imaging method as Selective Excitation Light Fluorescence (SELF) imaging (patent pending). In this technique, the sample is illuminated simultaneously with a

specific weighted sum of multiple excitation wavelengths and the emitted image that represents the fluorophores (components) desired is detected.

Similar to the Selective Illumination Transmission Imaging example in the previous section, an image of the fluorescing sample, formed by the weighted linear combination of the excitation images, may be calculated in a way that it has the highest contrast between a certain object of interest and the rest of the sample. Such an image can be calculated by applying the LDA to the multispectral image stack. In a sample with unknown material, however, we would not know what object to look for, i.e. an object of interest is not known *a priori*. Therefore, LDA cannot be applied. In this case, PCA can be applied to the images in the multispectral image stack, to calculate the first few principal component images. The principal components images are linear combinations of multispectral images which best explain the data. Each principal component image is a weighted sum of the excitation images. Therefore, one can illuminate the sample with the illumination resulting from a weighted sum of some of the multispectral illumination bands and directly capture the principal component image. Therefore, the information from all spectral bands are obtained by only a few captured images (corresponding to a few principal components) as opposed to having to sequentially illuminate and capture each image at each spectral band.

Similar to LDA, not all the PCA coefficients are positive, so the illumination profile (representing the illumination intensity vs. wavelength) that is calculated as a weighted sum of the excitation profiles may have negative intensities. Negative light is a nonsensical concept. To overcome this limitation, two illumination profiles are generated: one using the positive part of the calculated illumination profile, and one using the inverse of the negative part:

$$\psi_{pos}(\lambda) = \begin{cases} \psi(\lambda), & \text{if } \psi(\lambda) > 0 \\ 0 & \text{otherwise} \end{cases}, \text{ and } \psi_{neg}(\lambda) = \begin{cases} -\psi(\lambda), & \text{if } \psi(\lambda) < 0 \\ 0 & \text{otherwise} \end{cases} \quad (4.27)$$

Under each of these two illumination profiles, one image is captured. The image captured under $\psi_{neg}(\lambda)$ is then subtracted from the image captured under $\psi_{pos}(\lambda)$. Using this method, instead of capturing several images (one for each excitation wavelength) and then calculating the principal components, the principal component images are captured directly,

reducing the number of images that need to be captured and resulting in a shorter image acquisition time.

4.3.4 Spectral Unmixing of Multispectral Excitation Images

In bright-field (transmission) images, the light transmitted by the object is captured by the detector. The transmitted light is not linearly related to the concentration of the absorbing material in the sample. According to equation (3.2), the absorption (optical density) of the object is linearly related to the concentration of the absorbing material. As a result, before performing linear spectral unmixing on the transmission images of a sample, the transmitted intensity at each pixel is converted to the optical density. However, in fluorescence images, the relationship between fluorescence light intensity emitted from the sample and the concentration of the fluorescent dyes in the sample is linear (equation 4.21). Linear spectral unmixing can therefore be directly performed on the detected fluorescence intensities.

Let us assume a fluorescent sample is sequentially excited with L narrow-wavelength bands of light $\varphi_i(\lambda_e)$, and emits $E_i(\lambda_e, \lambda_m)$, $i = 1, 2, \dots, L$, where λ_e is the wavelength of the excitation light, and λ_m is the wavelength of the emitted light. At each excitation band, an image D_i of the sample is captured. If there are K fluorophores in the sample, then according to equation (4.21) $E_i(\lambda_e, \lambda_m)$ can be expressed as:

$$E_i(\lambda_e, \lambda_m) = \left(\sum_{j=1}^K \Phi_j(\lambda_m) \varepsilon_j(\lambda_e) C_j \right) \varphi_i(\lambda_e) \quad (4.28)$$

where $\Phi_j(\lambda_e, \lambda_m)$, $\varepsilon_j(\lambda_e)$, and C_j are the quantum efficiency, molar absorptivity, and density of the j -th fluorophore. $C_j = c_j l_j$, where l_j is the path length of light, and c_j is the concentration of the j -th fluorophore.

The output of the camera at the i -th illumination band, D_i , i.e. the i -th excitation image, can be described as:

$$\begin{aligned}
D_i &= t_i \int_{\lambda_e} \int_{\lambda_m} E_i(\lambda_e, \lambda_m) S(\lambda_m) d\lambda_m d\lambda_e \\
&= t_i \int_{\lambda_e} \int_{\lambda_m} \left(\sum_{j=1}^K \Phi_j(\lambda_m) \varepsilon_j(\lambda_e) C_j \right) \varphi_i(\lambda_e) S(\lambda_m) d\lambda_m d\lambda_e \\
&= \sum_{j=1}^K C_j \int_{\lambda_e} \int_{\lambda_m} t_i \Phi_j(\lambda_m) \varepsilon_j(\lambda_e) \varphi_i(\lambda_e) S(\lambda_m) d\lambda_m d\lambda_e \\
&= \sum_{j=1}^K C_j S_{ij}
\end{aligned} \tag{4.29}$$

where $S_{ij} = \int_{\lambda_e} \int_{\lambda_m} t_i \Phi_j(\lambda_m) \varepsilon_j(\lambda_e) \varphi_i(\lambda_e) S(\lambda_m) d\lambda_m d\lambda_e$.

In equation (4.29), D_i and C_j refer to the image intensity at the i -th excitation band and the concentration of the j -th fluorophore at any of the M pixels of the image. If the image intensity at the i -th excitation band and the m -th pixel is D_{im} , and the j -th fluorophore concentration at m -th pixel is C_{mj} , equation (4.29) can be rewritten as:

$$D_{mi} = \sum_{j=1}^K C_{mj} S_{ij}, \quad \text{for } \begin{cases} m = 1, 2, \dots, M \\ i = 1, 2, \dots, L \\ j = 1, 2, \dots, K \end{cases} \tag{4.30}$$

Equation (4.30) can be expressed in the matrix form as:

$$D = CS^T \tag{4.31}$$

where D is the $M \times L$ matrix of multispectral intensities D_{mi} at M pixels and L wavelengths, C is the $M \times K$ matrix of K fluorophore concentrations C_{mj} (or more precisely molar densities) at M pixels and S^T is the $K \times L$ matrix of spectral coefficients S_{ij} for the K fluorophores.

So the signal from multiple fluorescent dyes follows the linear mixing model. Spectral unmixing methods can thus be employed to unmix the concentrations of fluorophores at every pixel of the multispectral image. Unlike the transmission images, conversion to optical density is not necessary.

When more than one fluorophore is used in an experiment, or when autofluorescence is present, multispectral excitation imaging can be used to find the relative contribution of each fluorophore to the fluorescence signal at each pixel in the image stack.

Linear spectral unmixing estimates the pure spectra and concentration maps of each of the components comprising the fluorescence signal from the sample. Let \hat{C} and \hat{S} be the estimated concentration, and spectra matrices. According to the bilinear model of equation (4.31), these matrices satisfy this equation:

$$D = \hat{C}\hat{S}^T \quad (4.32)$$

Consequently, multiplying the multispectral image matrix by the pseudo-inverse of \hat{S}^T , generates the concentration maps:

$$\hat{C} = D\hat{S}(\hat{S}^T\hat{S})^{-1} \quad (4.33)$$

Lets define the $L \times K$ matrix V as follows:

$$V = \hat{S}(\hat{S}^T\hat{S})^{-1} \quad (4.34)$$

It follows from equations (4.33) and (4.34) that the concentrations of fluorophores \hat{C}_k , i.e. the columns of matrix \hat{C} , can be calculated by a weighted sum of the excitation images D_i , where the weights are given by the corresponding columns in the matrix V , as follows:

$$\hat{C}_k = \sum_{i=1}^L V_{ik}D_i, \quad k = 1, 2, \dots, K \quad (4.35)$$

Since the concentrations of the k -th fluorophore \hat{C}_k is a weighted sum of the excitation images D_i , an image corresponding to \hat{C}_k can be directly captured using our Selective Excitation Light Fluorescence Imaging approach: the illumination spectrum (the intensity of the illumination as a function of the wavelength) for the detection of the k -th component (fluorophore) can thus be designed as follows:

$$\psi_k(\lambda) = \sum_{i=1}^L V_{ik}t_i\varphi_i(\lambda), \quad k = 1, 2, \dots, K \quad (4.36)$$

t_i is the exposure time for the capture of the i -th excitation image.

If the sample is illuminated by these illumination profiles, the image of the concentration map of each component can be captured directly. It is worth mentioning that even though

the MCR-ALS algorithm is applied to solve the linear mixing model and to find non-negative estimates for C and S^T , the weights in the weighted sum in equations (4.35) and (4.36), which are the columns in the pseudo-inverse of matrix \hat{S}^T , are not necessarily non-negative, so we may have to capture two images using the positive and negative parts of the illumination profiles.

4.3.4.1 Estimating the Number of Pure Components

In chapter 3, we performed spectral unmixing on transmission images of histopathological samples after they were converted to optical densities. In those cases, the number of pure absorbing components in the sample was known. We used an algorithm based on the SIMPLISMA analysis to obtain the purest variables in the image stacks, and used them as initial estimates to the MCR-ALS unmixing procedure. This involved performing SIMPLISMA on several randomly selected regions of the image, and obtaining the purest spectra in those regions, followed by averaging the obtained spectra for all regions.

In the fluorescence excitation imaging examples in this chapter, there are instances where we do not know *a priori* how many pure fluorescence sources are detected in the sample. SIMPLISMA analysis can be employed to determine the number of pure variables in the sample. In section 3.4.1.4, a description of the SIMPLISMA analysis¹²⁸ for finding the purest variables in the data was provided. Here, we use the same algorithm for estimating the number of components in a mixture. To preserve the continuity, we repeat the explanation of the algorithm here, before explaining how it can be used to find the number of components in a mixture.

The key point of SIMPLISMA is the selection of “pure” variables from the data matrix D , in equation (4.31). A “pure” variable is a variable to which only one component of the mixture contributes. It can be a pure pixel (a row of matrix D) or a pure spectral band (a column of matrix D). The multi-step algorithm works by calculating a purity value for each variable at each step. At the i -th step, the variable with the highest purity value is chosen as the i -th pure variable. When searching for the purest variable in the rows of matrix D , the first purity value of a variable j , $p_j^{(1)}$ is defined by:

$$p_j^{(1)} = \frac{\sigma_j}{\mu_j + \alpha}, \text{ for } j = 1, \dots, M \quad (4.37)$$

Where α is an offset parameter and:

$$\sigma_j = \sqrt{\frac{\sum_{i=1}^n (d_{ij} - \mu_j)^2}{n}} \text{ and } \mu_j = \frac{\sum_{i=1}^n d_{ij}}{n}, \text{ for } j = 1, \dots, M \quad (4.38)$$

The offset parameter α is a user-defined value that is added to the denominator in equation (4.37) to avoid assigning a high purity value to a variable with a low mean value. After calculating the first purity value for all variables, the spectrum of the variable with the highest purity value is chosen as the first pure component spectrum. In the next steps, the contribution of the previously selected pure components are removed from all variables by using a weighting parameter, $w_j^{(k)}$ that reduces the influence of variables that are correlated with the selected pure variables. The k -th purity $p_j^{(k)}$ of a variable j is defined as:

$$p_j^{(k)} = w_j^{(k)} \frac{\sigma_j}{\mu_j + \alpha}, \text{ for } j = 1, \dots, M \text{ and } k = 2, \dots, L \quad (4.39)$$

The purest variable at k -th iteration is the variable for which $p_j^{(k)}$ is the highest. The weight vector is a vector with M elements containing the weights of each spectrum x_i (i -th row in spectra matrix D). Each element of the weight vector is defined as the determinant of the dispersion matrix of Y_i ($Y_i Y_i^T$), which contains all the selected “pure spectra” and each normalized spectrum $z_i = x_i / \|x_i\|$.

Windig et al.¹²⁸ used the term “purity spectrum” for $p_j^{(k)}$. They defined the “standard deviation spectrum” $s_j^{(k)}$ as:

$$s_j^{(k)} = w_j^{(k)} \sigma_j \quad (4.40)$$

To determine the number of pure components in the mixture, Windig et al.¹²⁸ suggested the following procedure: if, for example, two pure variables are representative of the whole mixture system, all other variables in the data set will be linear combinations of the two pure variables, which will result in zero values for the determinant function for calculating the weight vector. Consequently, the resulting purity and standard deviation spectra, equations 4.39 and 4.40, should also result in zero values. In practice, the purity values and standard deviation spectra will have certain positive values due to noise. A possible candidate for

determining the rank of the system after determining k pure variables is based on the relative total intensity of the standard deviation spectra:

$$R_k = \frac{\sum_{j=1}^M S_j^{(k)}}{\sum_{j=1}^M S_j^{(1)}} \quad (4.41)$$

This function will have a maximum value of 1, which is the value before the contribution of the first variable is removed from the data. The value of R_k becomes close to zero after using the proper number of pure variables. Visual evaluation of extracted spectra will facilitate the determination of number of components.

4.3.5 Results

4.3.5.1 Fluorescence Unmixing through Multispectral Excitation Imaging

In the following sections, three examples are presented for the fluorescence unmixing using multispectral excitation images.

4.3.5.1.1 Immunofluorescence

In the first example, the goal is to remove the autofluorescence (fluorescence produced by the tissue itself) signal from the fluorescence excitation image detected from an immunofluorescence labeled tissue section.

The tissue section was labeled with three fluorophores: 1) Alexa 488, and 2) a mix of Alexa 546 and Alexa 594, directed respectively against GFP, CD45, and 3) Hoechst (which stains nuclei). The slide was imaged under the microscope with eight narrow illumination spectra with center wavelengths 415 nm, 425 nm, ..., and 485 nm. The full width at half maximum (FWHM) of each illumination band was 15 nm. A band-pass filter with center wavelength of 450 nm and band pass of 70 nm was used as the illumination filter. A long-pass 510 nm filter was used as the detection filter.

To perform spectral unmixing, we need to first estimate the number of pure fluorescent components that are present in the sample as well as their initial spectra. One way to do this is to use the composite PCA image of the sample. In the composite PCA image, Figure 4.22, three distinct colors are identified which correspond to nuclei (green), GFP antigen (red), and tissue autofluorescence (blue), indicating that only three different fluorescence sources

(components) can be differentiated in this image stack. We were not able to differentiate the mix of Alexa 546 and Alexa 594 from the autofluorescence because their excitation profiles in the range of 415-485 nm were very similar.

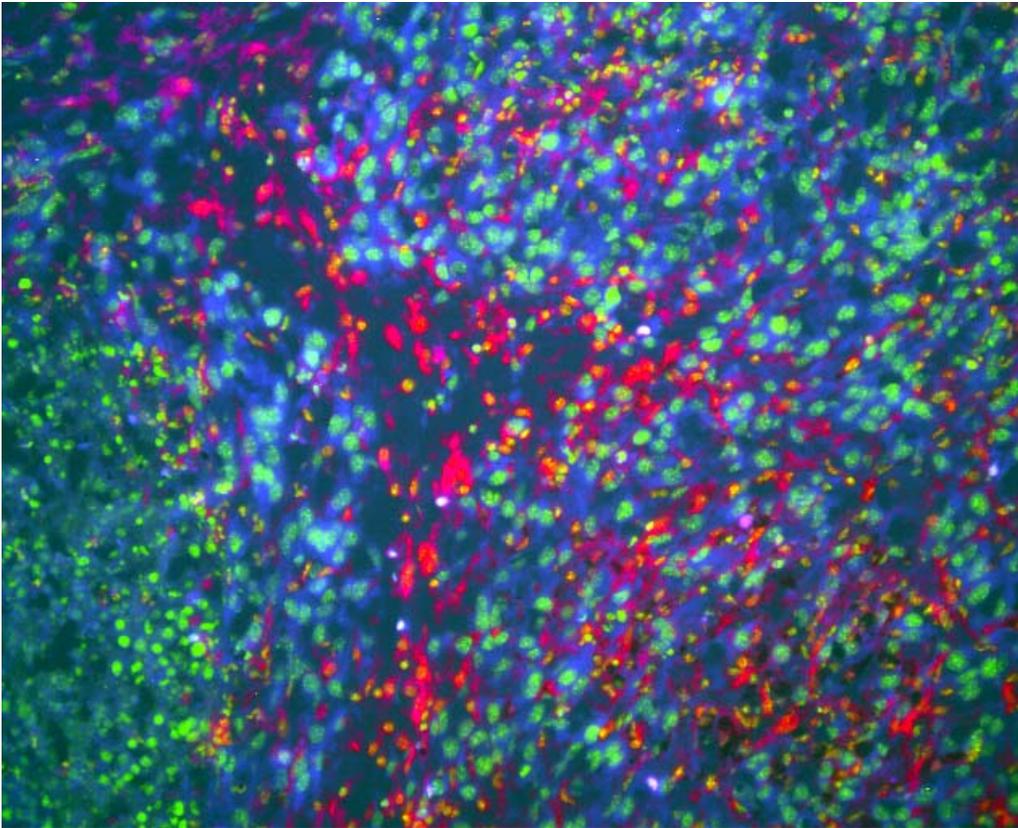


Figure 4.22 PCA composite image of the multi-color immunofluorescence labeled tissue section.

Another approach to determine the number of the pure fluorescent components present in the sample is the algorithm based on SIMPLISMA analysis, as explained in the methods section. In summary, we randomly selected one hundred regions in the image (Figure 4.23), and through SIMPLISMA analysis, we obtained the spectra of four “purest” pixels in each region. These spectra are shown in Figure 4.24. There are distinct differences among the first three spectra for each region. However, the fourth purest spectrum in each region, is closely correlated with the third, indicating that there are only three pure components present in each region. Alternatively, for each randomly selected region, the relative total intensity of the standard deviation spectra, as formulated in equation (4.41), for every pure variable (pixel) can be calculated. The averages of these values across all regions are noted in Figure 4.24

along with the corresponding extracted spectra. The average value for the fourth variable is very small and close to zero, another indicator that only three pure variables could be detected.

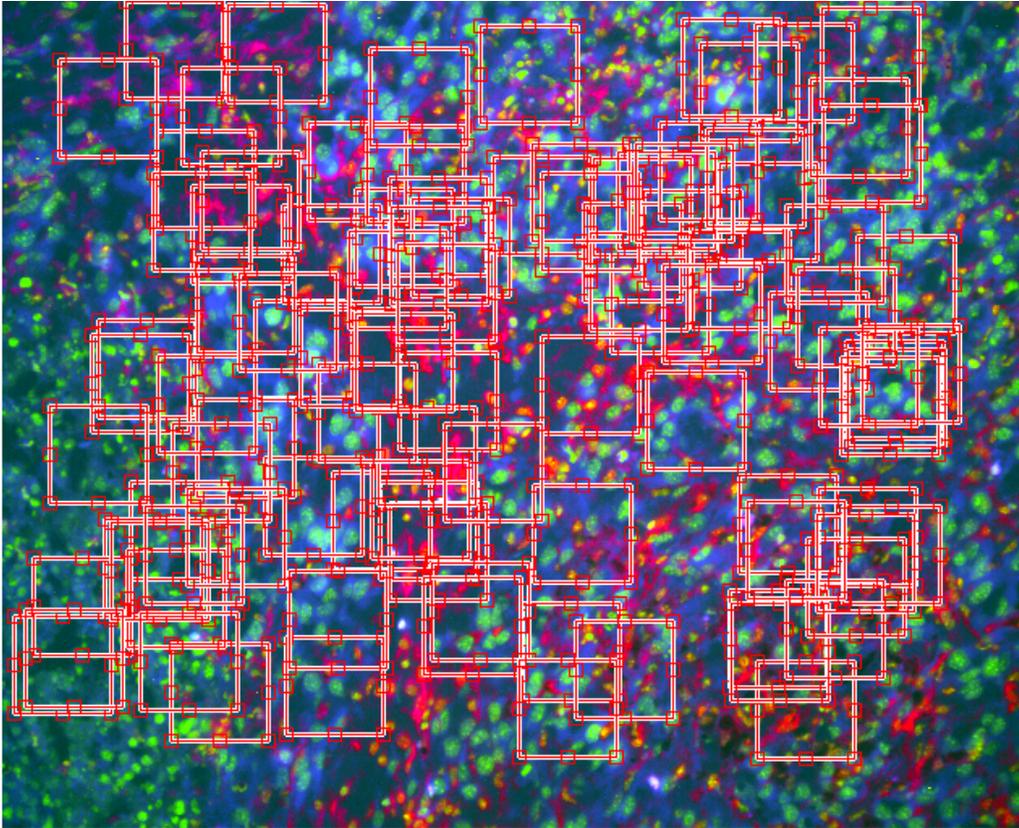


Figure 4.23 One hundred randomly selected regions in the image stack, for the extraction of purest spectra through SIMPISMA analysis

After determining the number of pure spectra to be equal to three, we calculated the average of the first purest spectrum for all the regions. This spectrum was used as the initial estimate for the first pure spectrum in the image. Similarly, the average of the second and third purest spectra for all regions represented the initial estimates for the second and third pure spectra for the entire image.

Using spectral unmixing, the contributions of each of these three fluorescence spectra to every pixel in the image stack were calculated. Figure 4.25 shows the concentration maps of the three fluorescence sources. The composite RGB image, with each of the concentration maps assigned to one channel is shown in Figure 4.26 (a). The composite image of the nuclei and the antibody without the tissue autofluorescence is shown in Figure 4.26 (b).

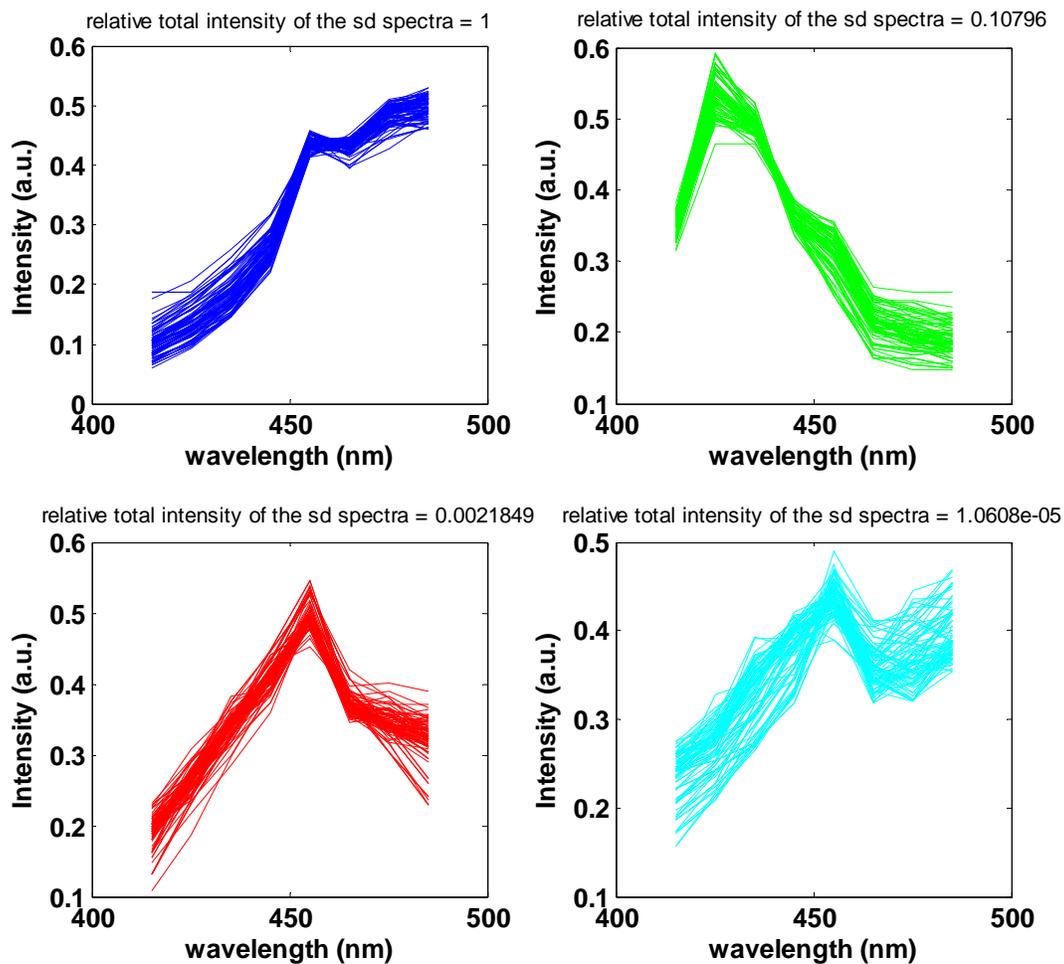


Figure 4.24 Purest spectra obtained from one hundred regions, from left to right, top to bottom: 1st, 2nd, 3rd, and 4th purest spectra in each region. The relative total intensity of the standard deviation spectra for each one of the four purest spectra for each region was calculated, and then averaged across all images. These average values are presented on their corresponding plots.

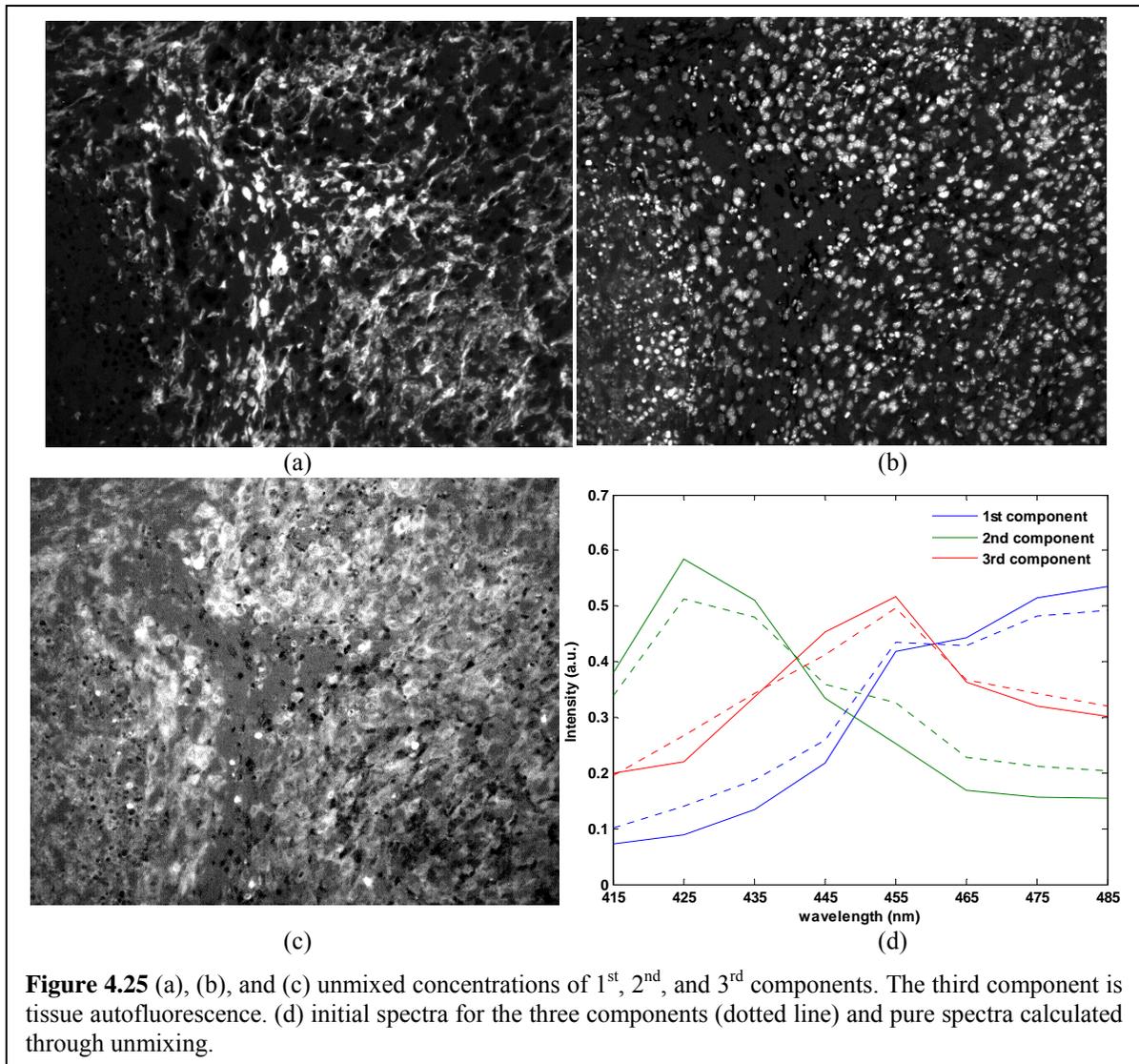
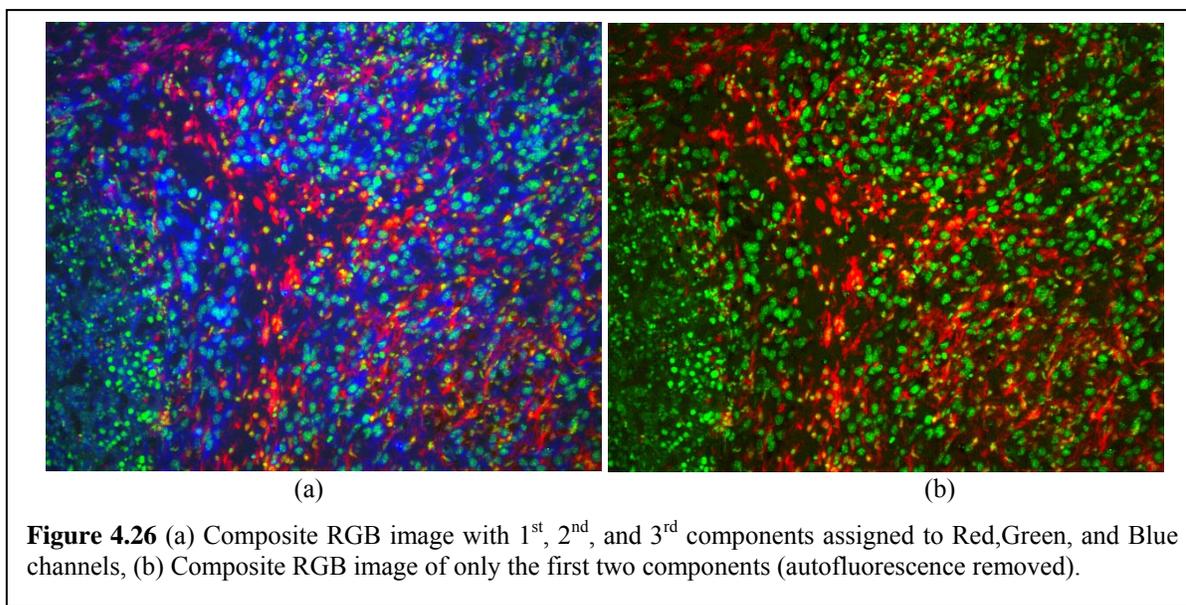


Figure 4.25 (a), (b), and (c) unmixed concentrations of 1st, 2nd, and 3rd components. The third component is tissue autofluorescence. (d) initial spectra for the three components (dotted line) and pure spectra calculated through unmixing.

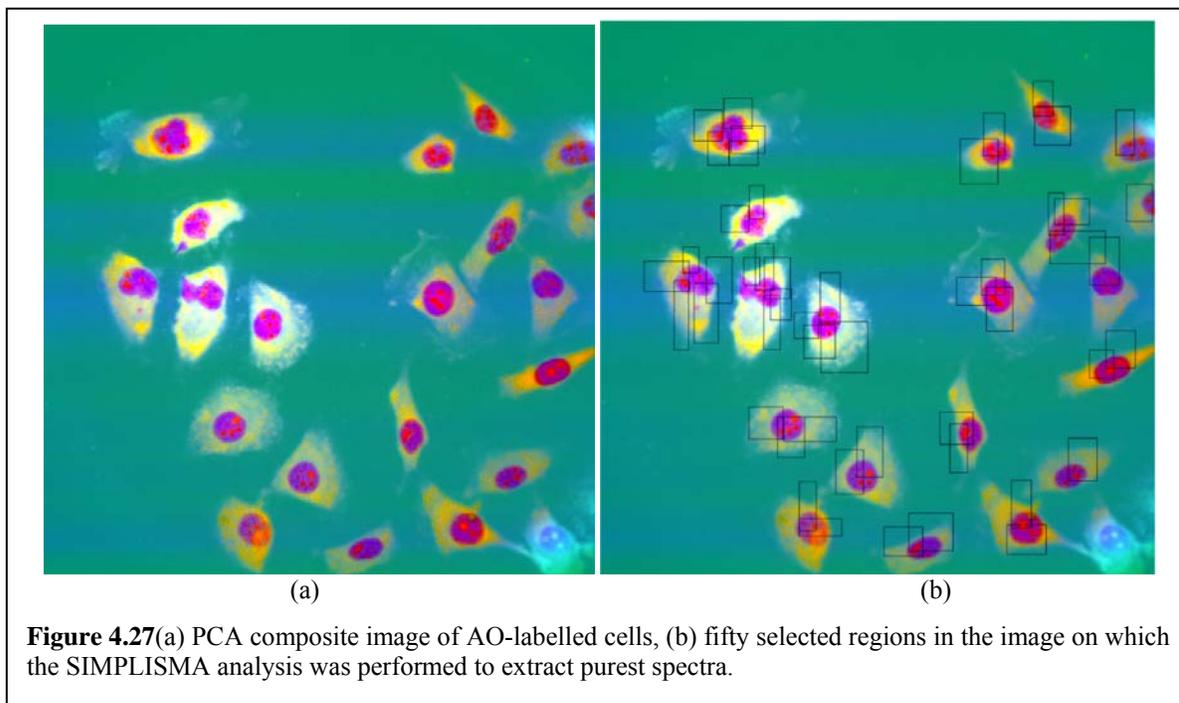


4.3.5.1.2 Fluorescence-labeled cells

The second example demonstrates how we separated the nucleus from the cytoplasm in the images of living cells stained with the fluorescent dye: Acridine Orange (AO) by making use of the differences in the absorption spectra of AO in nucleus vs. cytoplasm. AO is a fluorescent dye that interacts with both DNA and RNA. It stains the nuclei and cytoplasm differently. Human leukemia cells were stained with AO and imaged with our system under fifteen excitation wavelengths (390-530 nm with a step size of 10 nm).

The PCA composite image is shown in Figure 4.27 (a). The Nuclei and cytoplasm appear in two different colors. We applied the SIMPLISMA algorithm to fifty regions in the image (Figure 4.27 (b)), and extracted the four purest spectra in each region. These spectra are shown in Figure 4.28. The third and fourth spectra appear to be closely correlated with the second and the first spectra, respectively. The relative total intensity of the standard deviation spectra, as formulated in equation (4.41), is also very small for the third and fourth spectra, indicating that only two pure spectra exist in each selected region. Therefore, we used the averages of the first and the second spectra for all regions as the initial estimates to be used in the MCR-ALS unmixing algorithm.

The unmixed concentration maps are shown in Figure 4.29 (a) and (b). Figure 4.29 (c) shows the initial estimates for the excitation spectra, along with the spectra refined by the MCR-ALS algorithm.



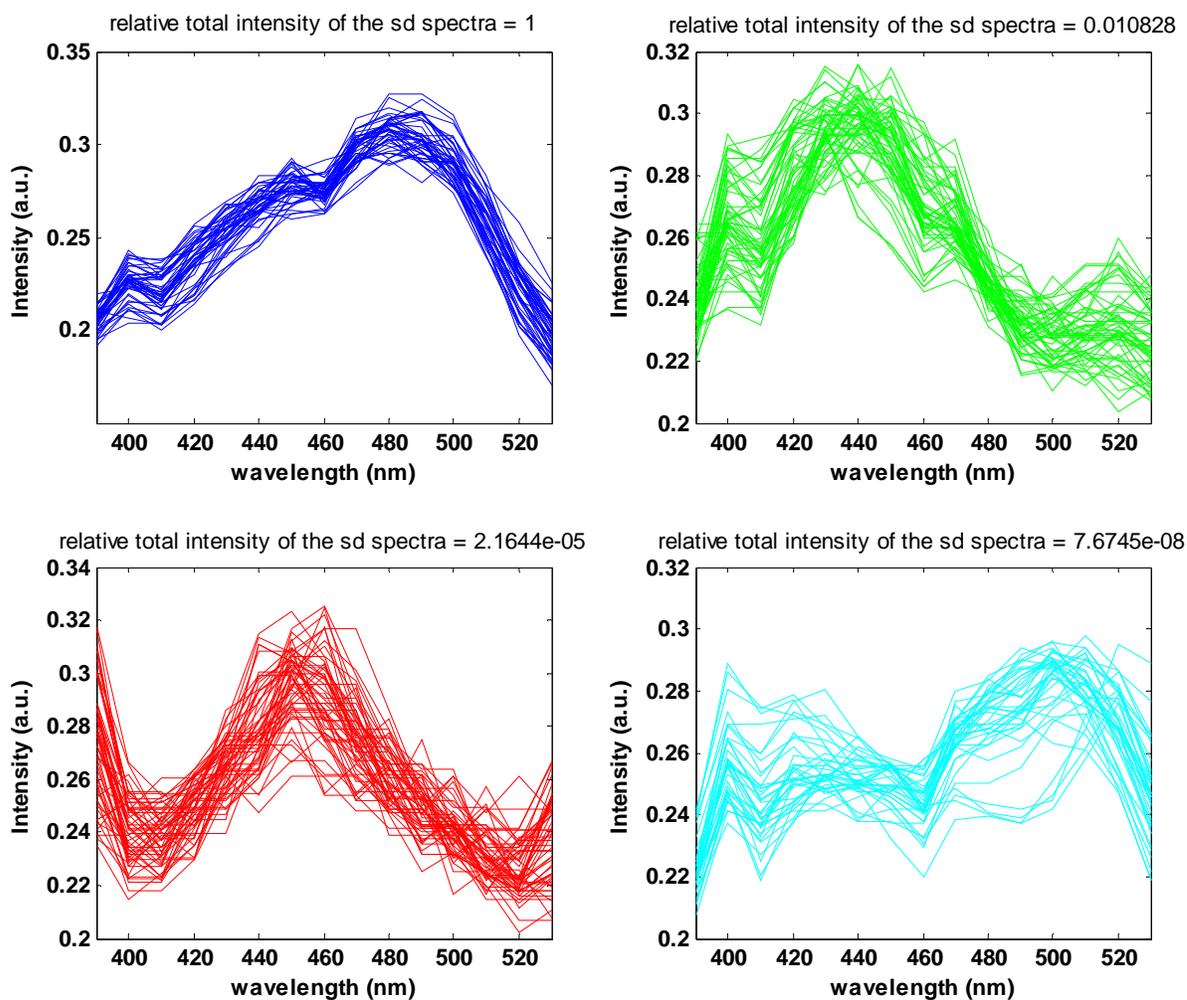
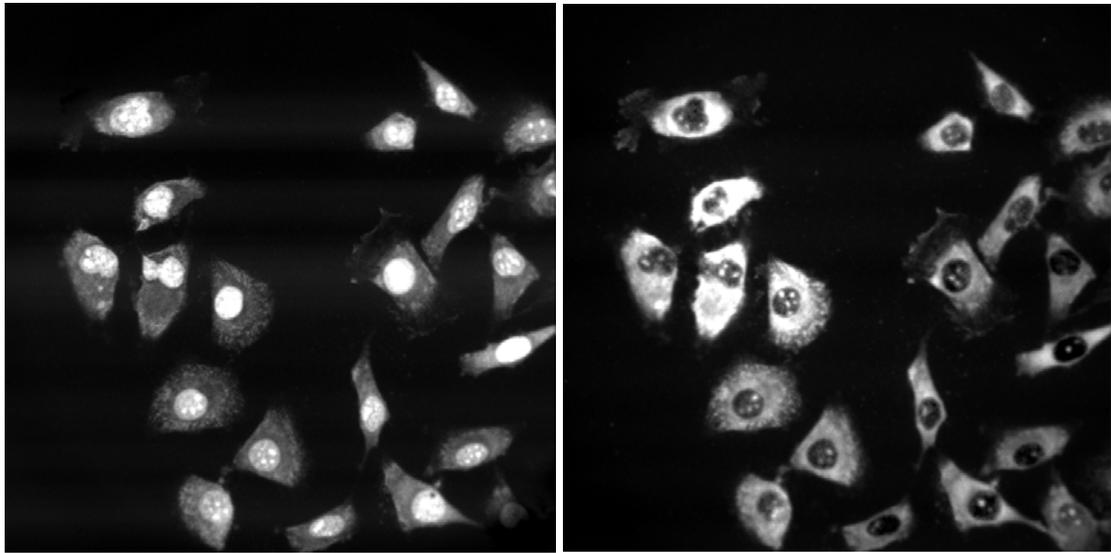
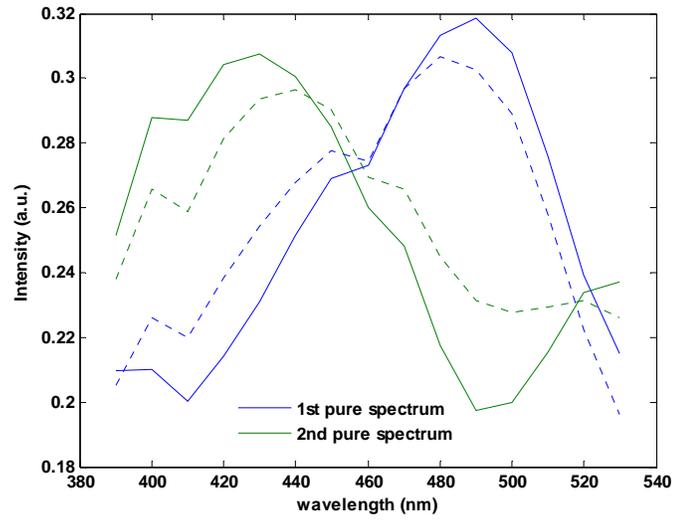


Figure 4.28 Purest spectra obtained from fifty regions. From left to right, top to bottom: 1st, 2nd, 3rd, and 4th purest spectra in each region. The relative total intensity of the standard deviation spectra for each one of the four purest spectra for each region was calculated, and then averaged across all images. These average values are presented on their corresponding plots. Note that the third and fourth values are significantly lower than the first two.



(a)

(b)



(c)

Figure 4.29 (a) and (b) unmixed images of AO fluorescent signal based on its different absorption spectra in the nucleus and cytoplasm, (c) initial spectra obtained from the image stack (dotted line) and pure spectra obtained through unmixing.

4.3.5.1.3 *Autofluorescence Removal from FISH-labeled Tissue Section*

This example demonstrates the results of spectral unmixing performed both on the excitation and the emission stack of images captured with our system from a tissue section labeled with FITC through telomere Fluorescence *in situ* Hybridization (FISH). Using a single FITC filter-cube, the bright tissue autofluorescence makes it difficult to differentiate FITC-labeled material. The aim of this experiment was to remove tissue autofluorescence. A conventional FITC filter cube with an excitation filter with center wavelength of 480 nm and band pass of 40 nm and a detection filter with center wavelength of 535 nm and band pass of 50 nm was used. We captured an excitation emission stack of images within the band pass allowed by the filter cube. With the use of a tunable filter (CRI's VariSpec) three emission wavelengths – 520 nm, 540 nm, and 560 nm – were used and the light engine generated three excitation wavelengths (450 nm, 470 nm, and 490 nm). A total of nine images were captured to form a 3×3 excitation emission image matrix (Figure 4.30). From the nine images, the image with the best contrast between nuclei and background appears to be the image at the excitation of 490 nm and emission of 520 nm. This image still has significant background caused by tissue autofluorescence.

We can perform spectral unmixing on the entire set of nine excitation-emission images. Alternatively, spectral unmixing can be performed on either the excitation image stack, or the emission image stack. We can synthesize an excitation image stack, by adding together the three images detected at each excitation wavelength. Similarly, an emission image stack can be generated by adding the three images detected at each emission wavelength. Linear spectral unmixing was performed on both excitation, and emission image stacks using the spectra obtained interactively from the areas thought to be representative of FITC and the tissue autofluorescence.

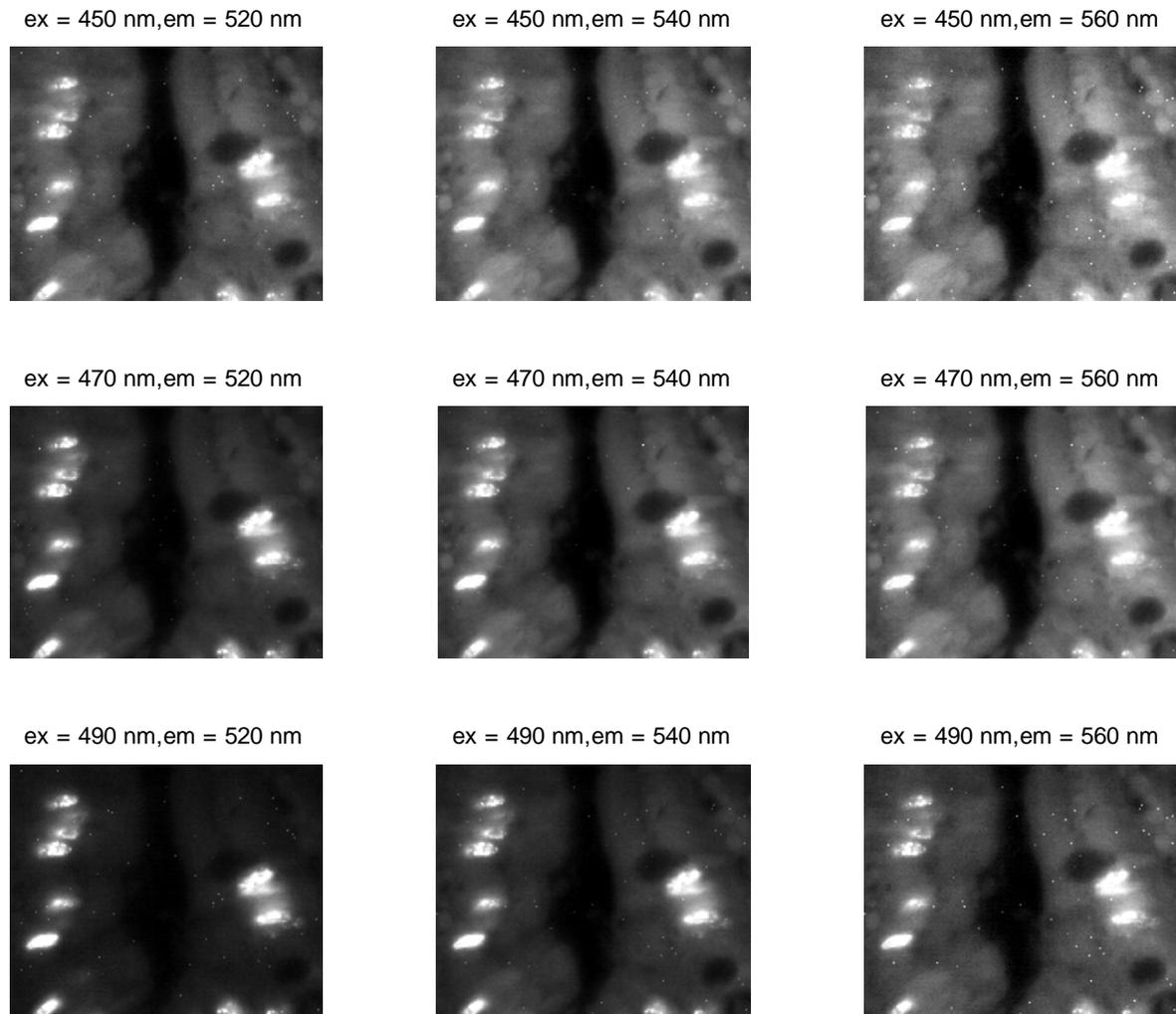


Figure 4.30 Excitation-Emission Images of the FISH-labeled tissue section

Figure 4.31 demonstrates the results of performing spectral unmixing on the excitation image stack, using spectra obtained directly from the image as the initial estimates for pure spectra. Since the spectra obtained from the FITC labeled tissue contain some contribution from the tissue autofluorescence as well as FITC, complete unmixing is not achieved. Note the holes in the position of FITC labeled nuclei in the autofluorescence channel. This is because the entire fluorescence signal in such areas is assigned to the FITC channel. If an appropriate fraction of the tissue-autofluorescence spectrum is subtracted from the FITC spectrum obtained from the image, the “holes” in the autofluorescence channel will disappear. We found this fraction through examining a series of fractions and choosing the

one which results in an almost homogeneous autofluorescence image. The results are shown in Figure 4.32. Figure 4.33 demonstrates the results of performing spectral unmixing on the emission image stack, after the appropriate fraction of tissue autofluorescence spectrum was subtracted from the FITC spectrum, to remove the holes in the autofluorescence image.

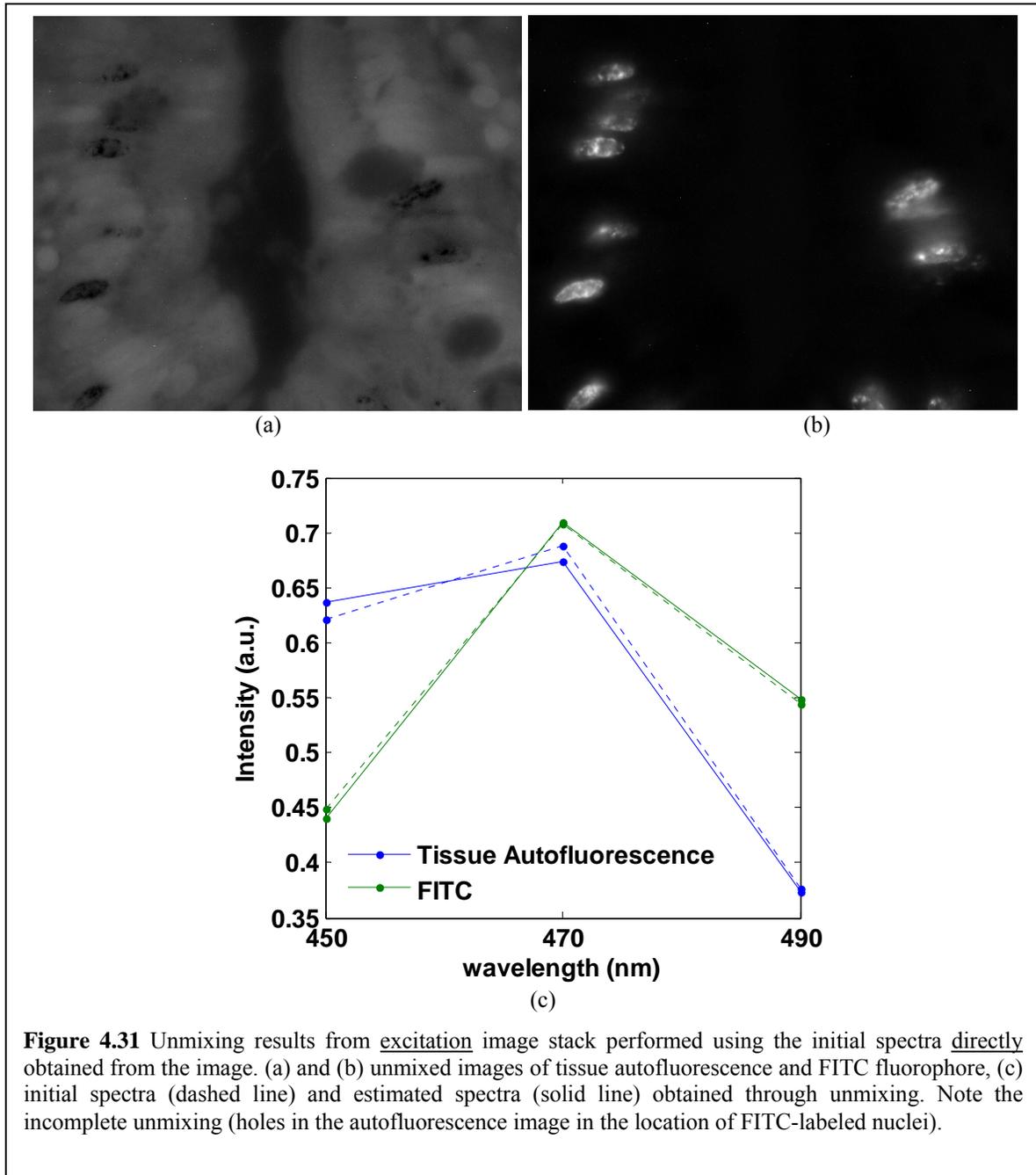
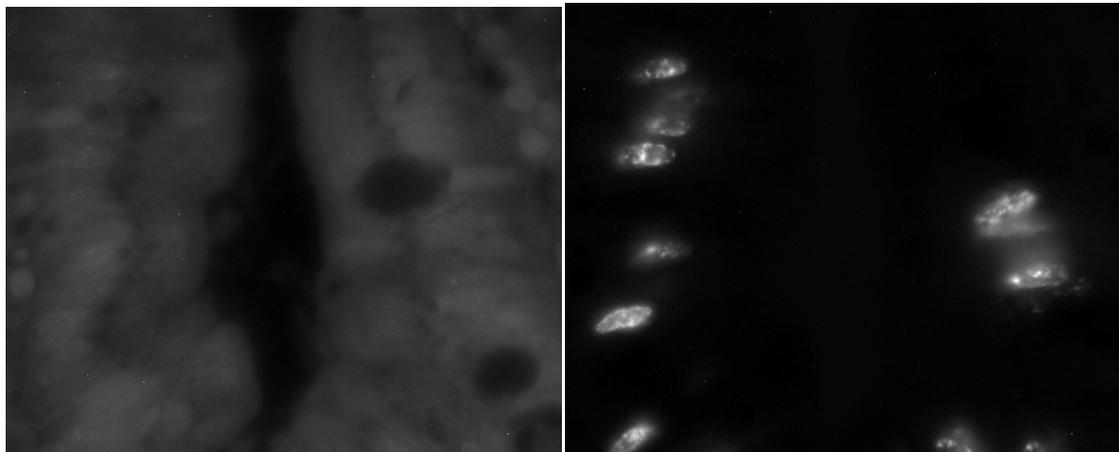
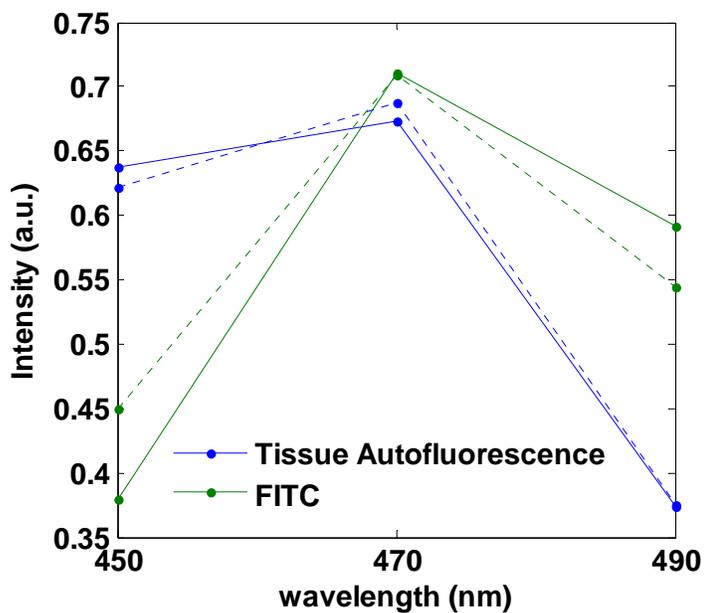


Figure 4.31 Unmixing results from excitation image stack performed using the initial spectra directly obtained from the image. (a) and (b) unmixed images of tissue autofluorescence and FITC fluorophore, (c) initial spectra (dashed line) and estimated spectra (solid line) obtained through unmixing. Note the incomplete unmixing (holes in the autofluorescence image in the location of FITC-labeled nuclei).



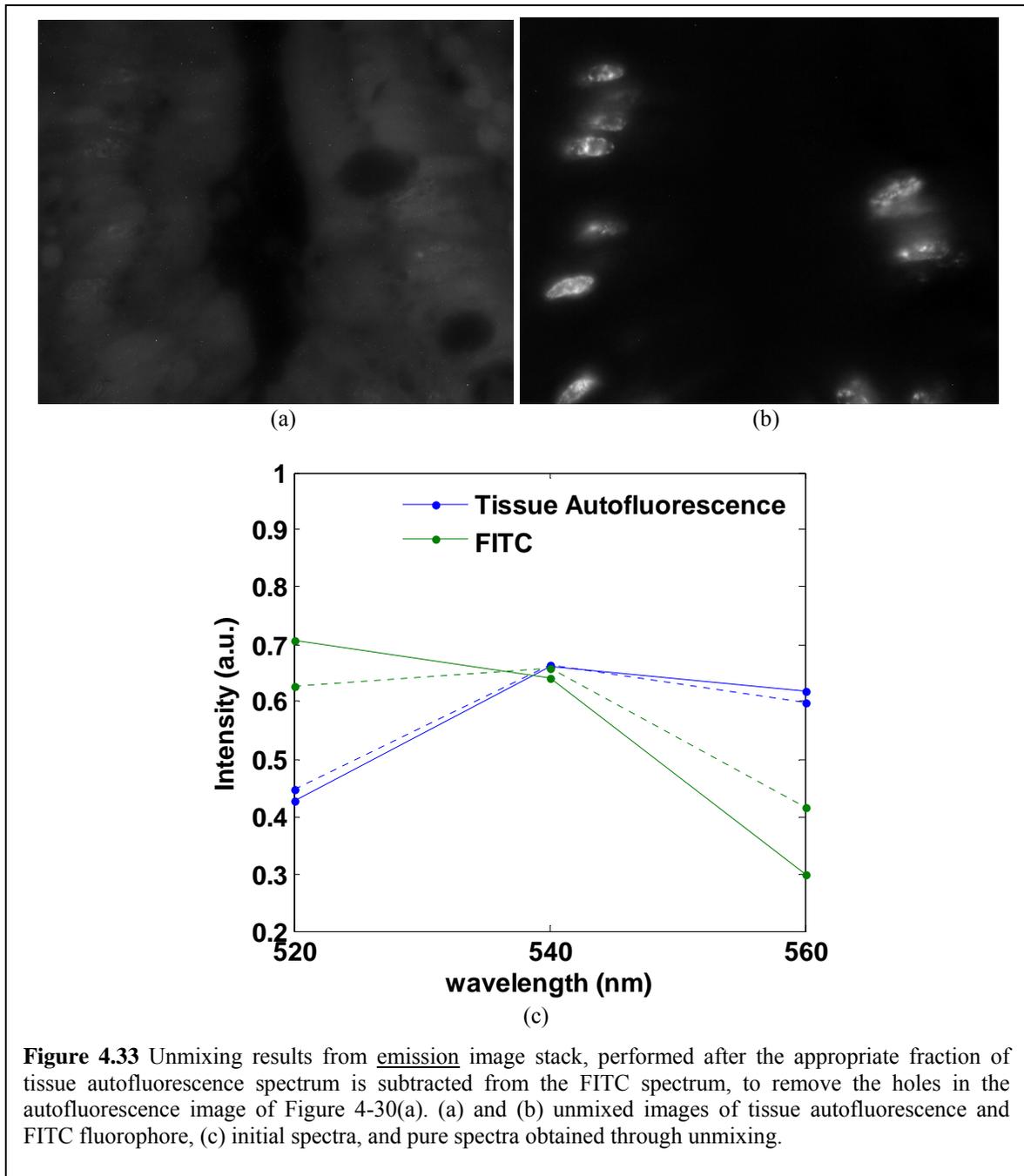
(a)

(b)



(c)

Figure 4.32 Unmixing results from excitation image stack, performed after the appropriate fraction of tissue autofluorescence spectrum is subtracted from the FITC spectrum, to remove the holes in the autofluorescence image of Figure 4-30(a). (a) and (b) unmixed images of tissue autofluorescence and FITC fluorophore. (c) initial spectra, and pure spectra obtained through unmixing. Note the homogeneous autofluorescence image.



Through this example, we demonstrated that unmixing can be performed on the excitation image stack in the same way that it can be performed on the emission image stack. The spectral properties of the two fluorescence sources in this experiment were such that both excitation and emission unmixing resulted in similar unmixing results. However, they

may be instances where the differences in the excitation spectra of two fluorophores are more significant than the differences in their emission spectra, and vice-versa.

4.3.5.2 SELF Imaging

4.3.5.2.1 PCA Imaging of Plastic Material

To demonstrate that we can resolve different fluorescence signals using excitation imaging, we captured the excitation image stack of several different plastic pens and plastic pieces. Excitation images were acquired under thirteen spectral illumination channels with center wavelengths of 415 nm, 425 nm, ..., and 535 nm. Figure 4.34 (a) shows spectra of the illumination channels that were used to acquire the multi-spectral data cube. An RGB image was constructed by assigning the first three principal component images to the three color channels. Figure 4.34 (b) shows the composite RGB image. Different colors result in the composite image because of differences in the absorption spectra of the different fluorophores in the different plastic materials.

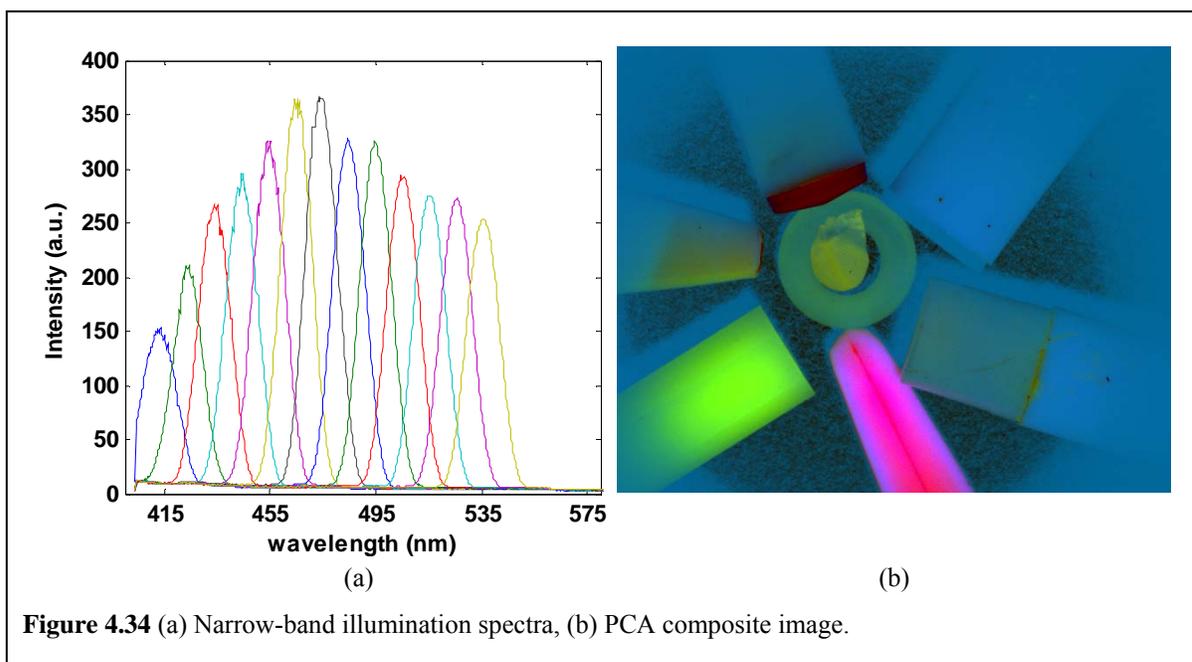


Figure 4.34 (a) Narrow-band illumination spectra, (b) PCA composite image.

Using SELF imaging, we calculated three illumination profiles for the direct capture of the first three PCA images. These illumination profiles are shown in Figure 4.35. Panels (c) and (e) show how the second and third profiles that were partly negative, were separated into the positive and negative portions.

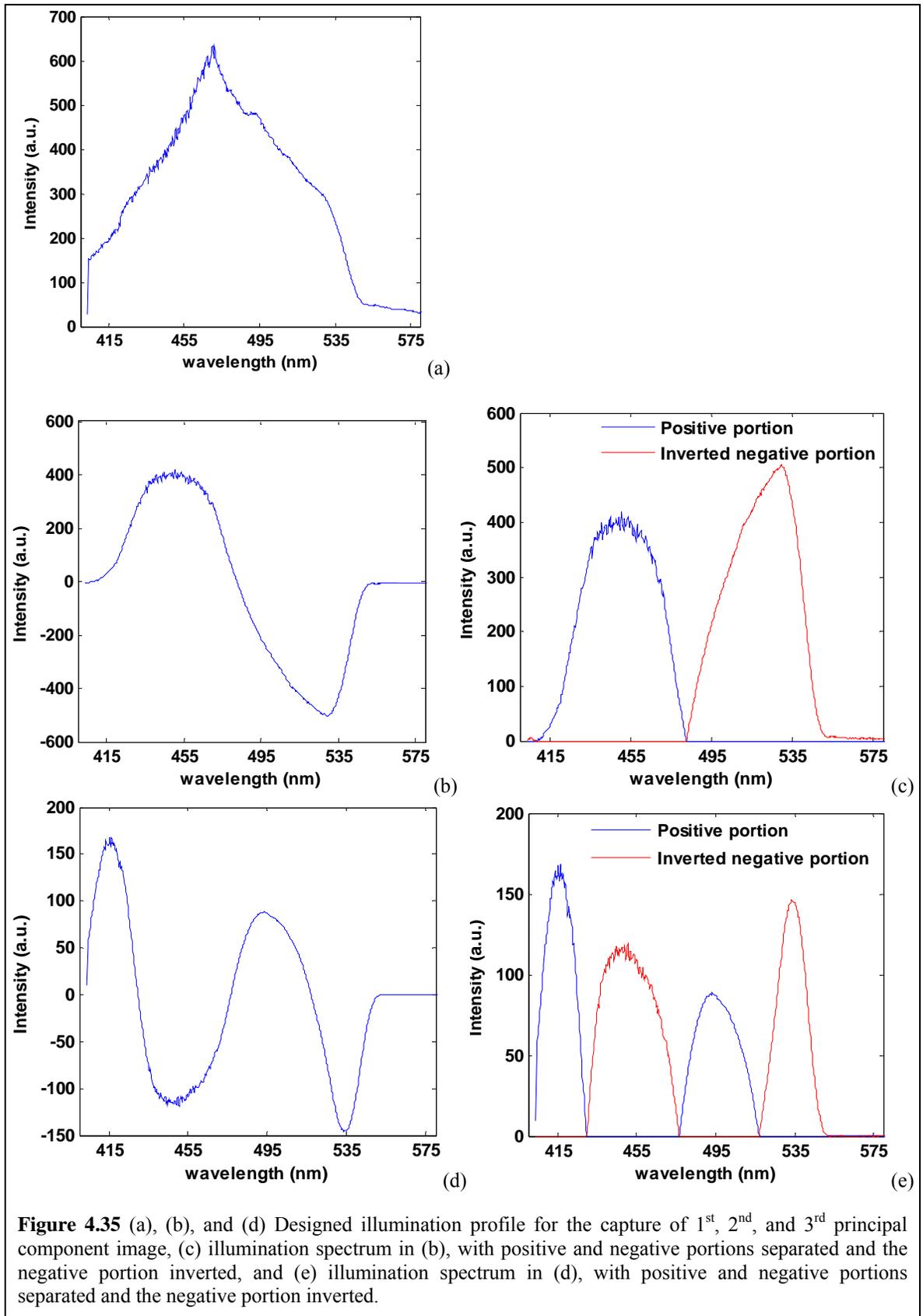


Figure 4.35 (a), (b), and (d) Designed illumination profile for the capture of 1st, 2nd, and 3rd principal component image, (c) illumination spectrum in (b), with positive and negative portions separated and the negative portion inverted, and (e) illumination spectrum in (d), with positive and negative portions separated and the negative portion inverted.

When the designed illumination profile contained negative intensities, the same procedure for the capture of Selective Illumination Transmission Images was performed for the capture of SELF images: one image was captured under the “positive” illumination, and another under the “negative” illumination. The difference of the two images was calculated after each was normalized by its exposure time, and scaled by the correction factor due to OneLight’s scaling of generated illuminations (as explained in section 4.2.2, and equation 4.15).

Figure 4.36 shows the generated illumination profile for the capture of the first principal component image. Since the designed illumination for the capture of the first principal component was non-negative, only one image had to be captured.

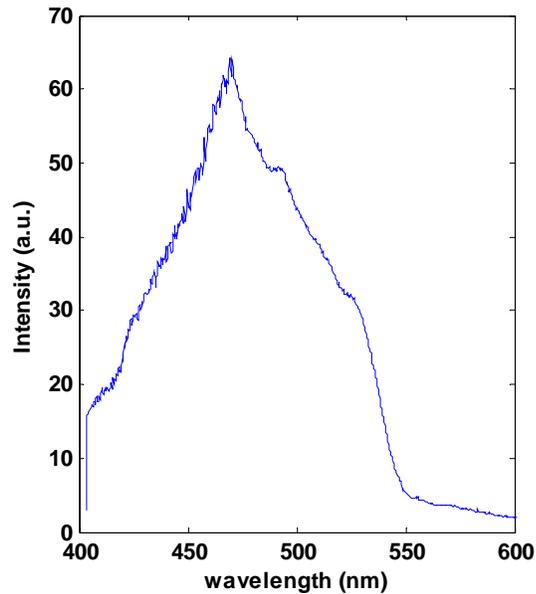


Figure 4.36 Generated illumination profile for the capture of 1st principal component image.

Figure 4.37 (a) and (b) show the generated illumination profiles for the capture of second and third principal components. Panels (c) and (d) show the same illumination profiles after they were corrected for OneLight’s scaling factor.

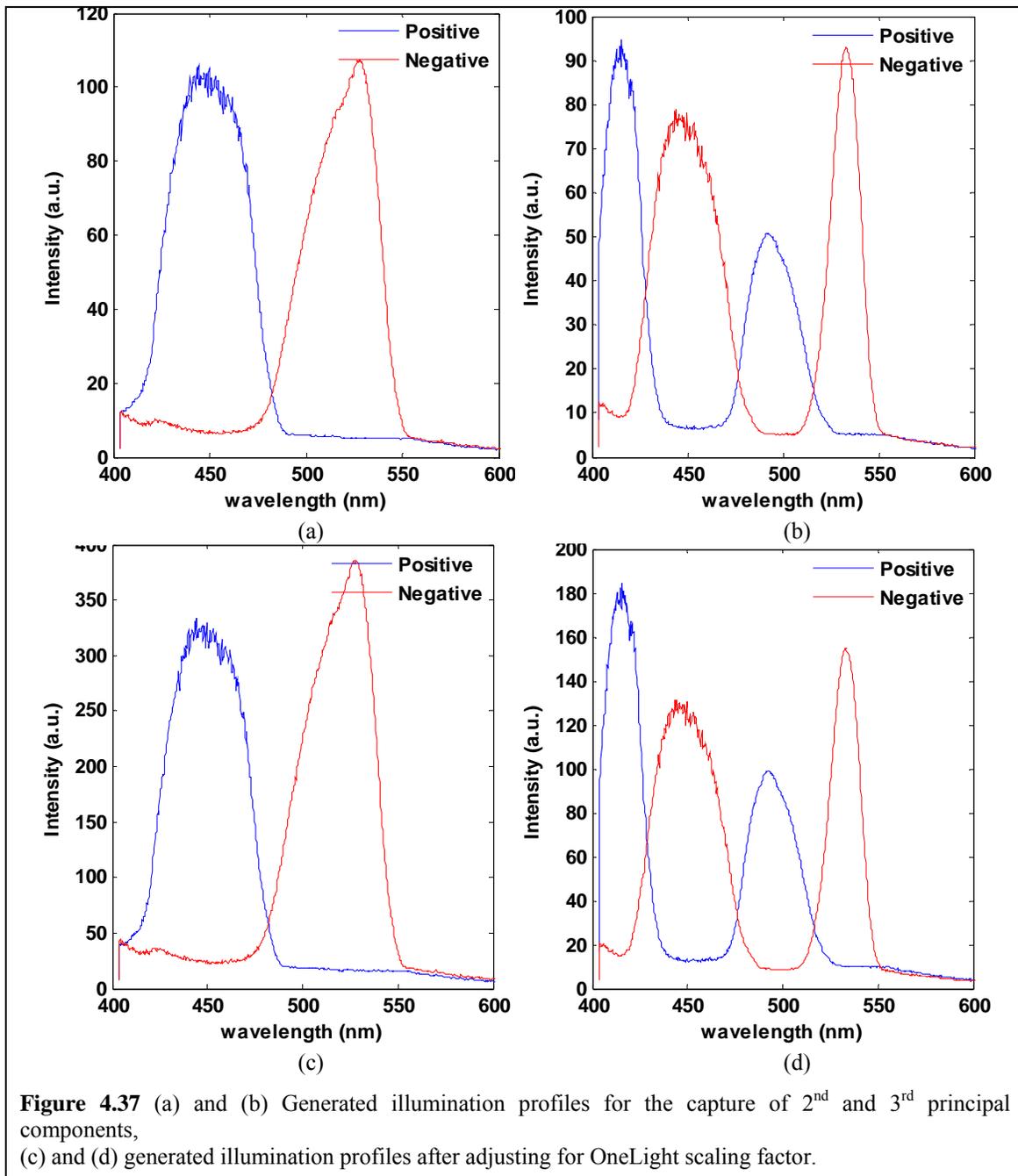
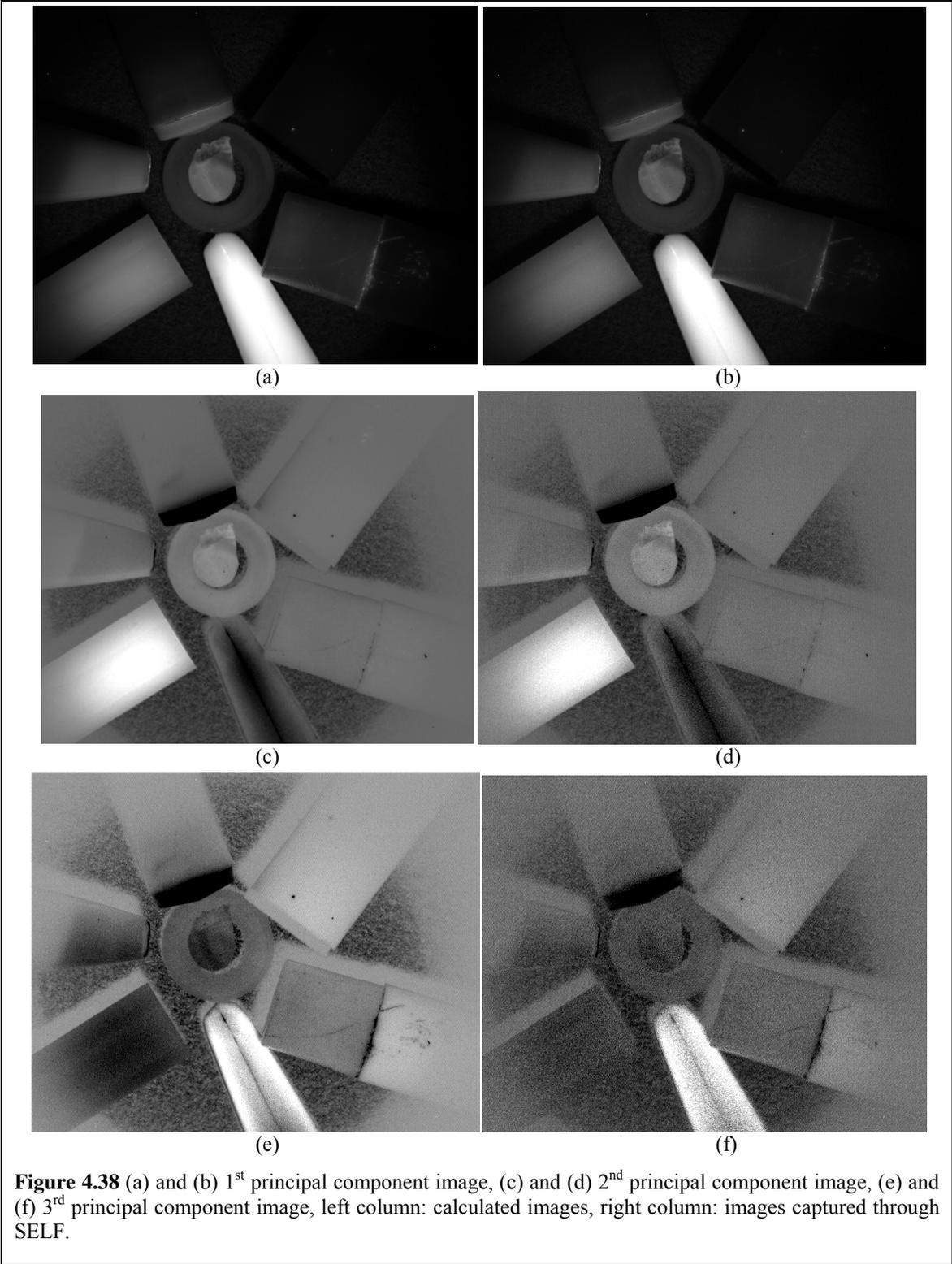


Figure 4.38 shows the first three principal component images that were calculated by applying PCA to the thirteen excitation images, along with the corresponding principal component images that were captured under the SELF illumination profiles.



The coefficients for the first principal component were all positive, while the second two principal components contained negative coefficients as well. Hence, a total of five images were captured to directly construct the first three principal components as opposed to the initial thirteen (one for the first principal component and two per each of the second and third principal components). As the results show, there is no significant difference between the captured and calculated principal component images.

4.3.5.2.2 Spectrally Unmixed Imaging of Fluorescence-labeled Cells

As discussed in section 4.3.4, spectral unmixing can be performed on the excitation image stack to unmix the concentrations of fluorophores that have different excitation spectra. Using equations (4.30) and (4.31) the illumination profiles can be designed for the direct capture of the unmixed fluorescence signals. In this section, we demonstrate this approach on cells that were fluorescently labeled with AO. In section 4.3.5.1.2 we demonstrated the application of spectral unmixing to unmix AO in the nuclei and the cytoplasms of the cells based on its different excitation spectra in the nucleus and in the cytoplasm.

Here, we captured an excitation image stack of AO-labeled cells, at thirteen excitation channels with center wavelengths of 420 nm, 430 nm, ..., and 540 nm. We performed linear spectral unmixing in the multispectral excitation image stack to estimate the pure spectra, and their concentration maps. We then calculated the illumination profiles for the direct capture of the unmixed fluorescence signals. These illumination profiles are demonstrated in Figure 4.39. For the illumination profile corresponding to each unmixed fluorescence signal, two images were captured under the positive and inverted negative portion of the calculated illumination profile.

As explained before, the two images were normalized by the exposure times, and corrected for OneLight's scaling of generated illuminations. Figure 4.40 shows the designed illumination profiles, and the illuminations generated by the light engine. As can be seen in this figure, in the illuminations generated by OneLight, the ratio between the positive and negative designed illuminations is not maintained; therefore the images need to be corrected for OneLight's scaling of illuminations. It is the captured positive and negative images that are scaled, but, the scaling factor is calculated using the generated and the designed

illumination profiles, as formulated in equation (4.15). The scaling factors are calculated in such a way that if they are applied to the generated positive and negative illuminations, the ratio between the generated positive and negative illuminations is the same as the ratio between the designed positive and negative illuminations. Figure 4.40 panels (e) and (f) demonstrate the scaling factors applied to the generated illumination profiles. The same scaling factor is applied to the positive and negative captured images. Note that after scaling, the designed positive and negative illuminations are proportional to the generated positive and negative illuminations.

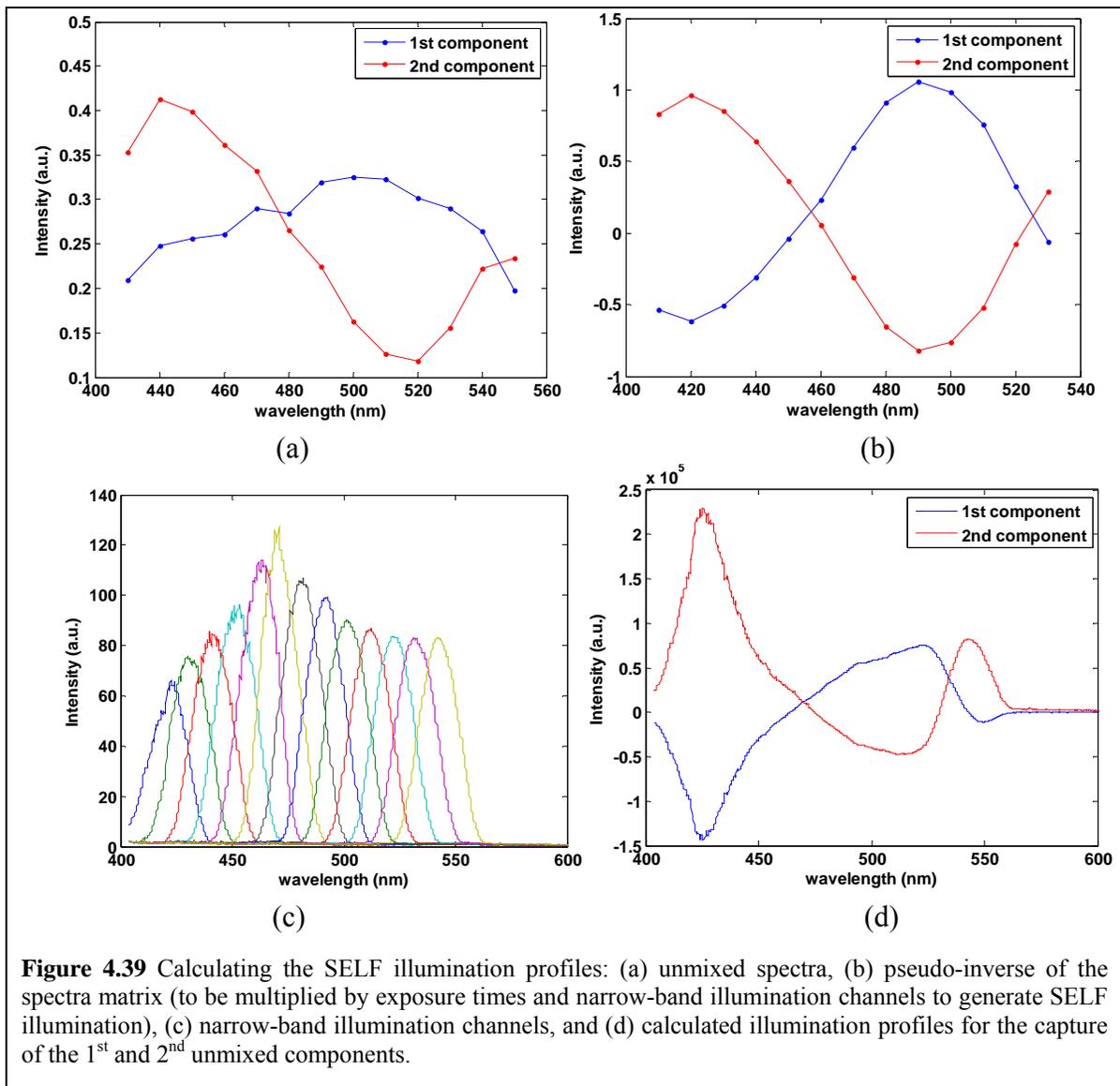


Figure 4.39 Calculating the SELF illumination profiles: (a) unmixed spectra, (b) pseudo-inverse of the spectra matrix (to be multiplied by exposure times and narrow-band illumination channels to generate SELF illumination), (c) narrow-band illumination channels, and (d) calculated illumination profiles for the capture of the 1st and 2nd unmixed components.

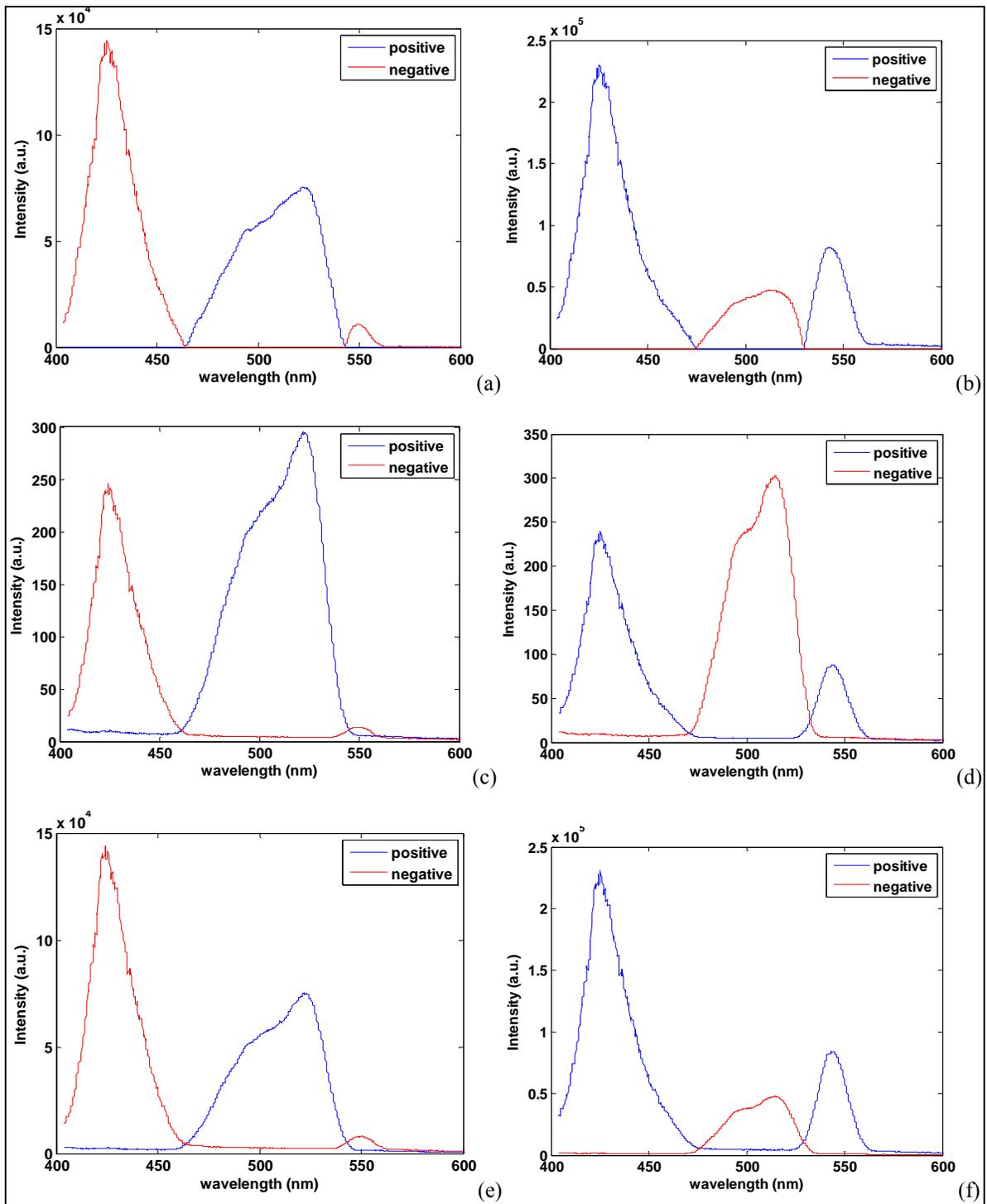
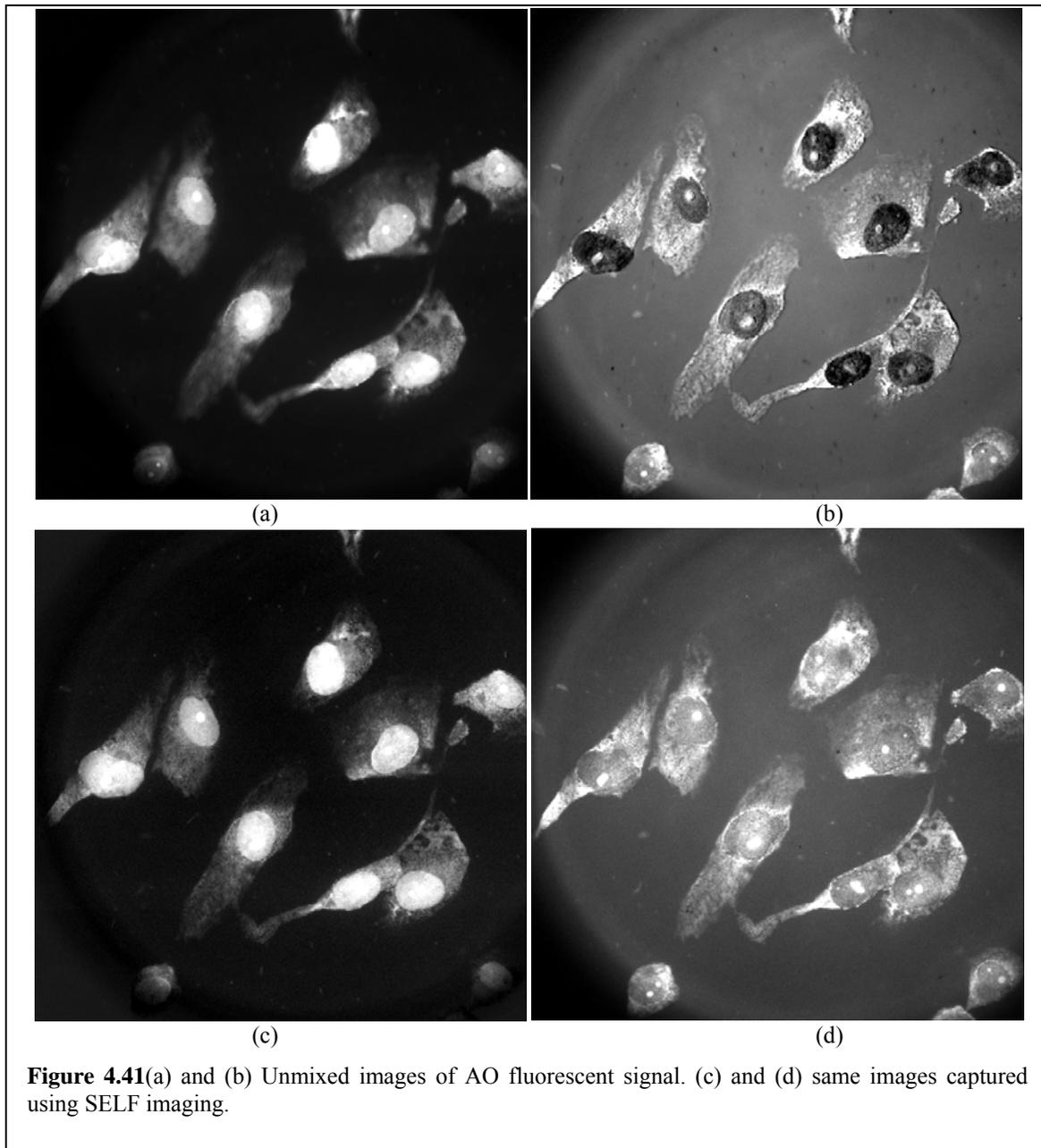


Figure 4.40 (a) and (b) Designed illumination profiles for the capture of unmixed components, (c) and (d) generated illumination profiles, (e) and (f) generated illumination profiles adjusted for OneLight scaling factor, left column, (a), (c), and (e), corresponds to the first component, right column, (b), (d), and (f), corresponds to the second component.

The difference between the corrected positive and negative images represents the image of the unmixed fluorescence signal captured under the designed illumination. The images of the two unmixed components captured this way are shown in Figure 4.41 along with the images of component concentrations calculated using spectral unmixing.

The captured image of the cytoplasm is not exactly the same as the calculated image of cytoplasm. This might be due to the slight difference in the designed and the generated illumination profiles for the capture of this image. As can be seen in Figure 4.40 panel (d), the generated illuminations have some “background” light where the illumination is expected to be zero. This background light can cause some fluorescence emission, which can in turn cause the difference between the captured and calculated images.

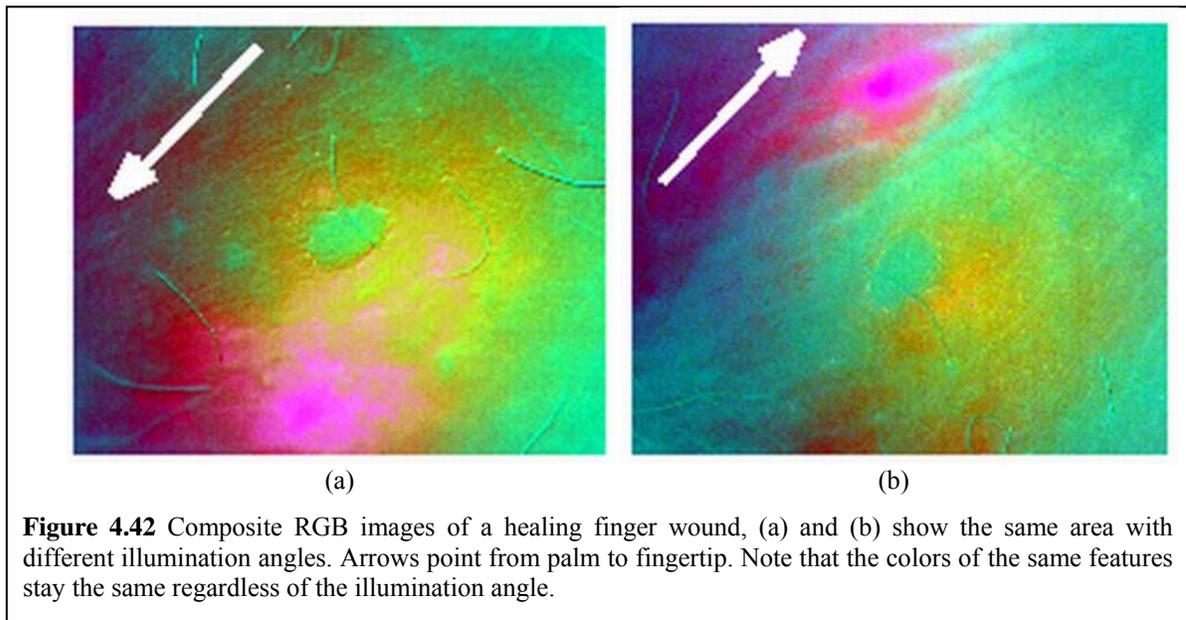


4.3.5.3 Excitation Imaging of a Healing Wound

In this last experiment, *in vivo* multispectral excitation autofluorescence images were captured from a healing wound on a subject's finger. The goal of this experiment was to demonstrate that excitation imaging is able to show differences in the autofluorescence signal between healthy and healing tissue. Thirteen illumination wavelengths starting from 415 nm to 535 nm with a step size of 10 nm were used. PCA was performed on the multispectral data

cube. Figure 4.42 shows the composite PCA images. Different colors exist in the RGB image. Although the exact cause of these differences is not known to us, we speculate they could be due to: a) different underlying fluorophores, b) the same fluorophores at different depths (due to the scattering and absorption properties of the overlying tissue these fluorophores may be exposed to differing excitation profiles), or c) the same fluorophores in biochemically different material which alter their absorption spectra in the tissue.

To verify that the colors were not dependent on the illumination angle and were truly indicative of different excitation/absorption spectra, we captured two images of the same tissue illuminated at different angles. The colors of the different locations in the tissue stayed the same (shown in Figure 4.42).



Also, to examine whether or not the PCA coefficients varied significantly from one tissue to another, we captured an image stack at two different positions and then used the coefficients from the first image stack to calculate a weighted sum of the images in the second image stack, and vice versa. No significant changes were observed in the composite PCA images (Figure 4.43).

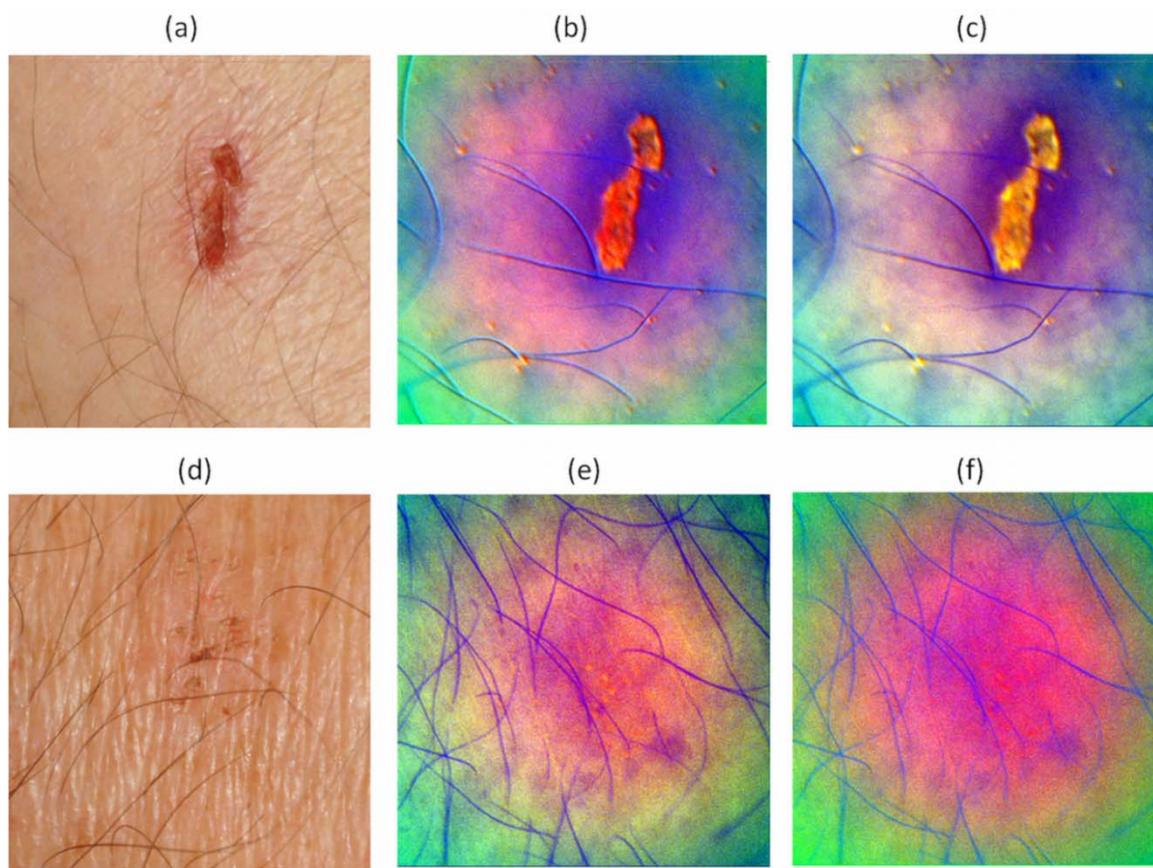


Figure 4.43 Composite RGB image of each sample obtained using coefficients from the other sample. (a) and (d) White light images of healing wounds, (b) and (e) PCA composite images generated from the corresponding excitation image stack, (c) composite PCA image of (b) generated using the PCA coefficients obtained from (e), and (f) composite PCA image of (e) generated using the PCA coefficients obtained from (b).

4.3.6 Discussion

Multispectral fluorescence imaging through multi-channel detection captures the fluorescence emission spectrum at each pixel. The emission spectrum corresponds to only one orientation in an EEM. The fluorescence excitation imaging approach described in this work enables us to capture the excitation spectrum at each pixel. The fluorescence excitation spectrum corresponds to the orthogonal orientation in the EEM. The information present in each orientation is mathematically independent and very likely contains independent biological information that cannot be obtained by a different view. This approach may prove advantageous in certain applications, e.g. where the fluorescence change in pathologically changed tissue is more pronounced in the excitation spectrum of the tissue rather than the

emission. Use of this orthogonal information can also help increase the number of simultaneous labels which can be distinctly detected on a single microscopy slide as labels can be separated by their absorption spectra and not just their emission spectra. This method of fluorescence imaging may also result in better signal to noise ratio as there are no optical or mechanical elements in the imaging light path, which results in minimal image degradation, whereas in emission imaging approaches there is a filter wheel or tunable filter in the detection light path.

Through the use of a programmable light source such as the OneLight Spectra, in which not only the wavelengths of the illumination light, but also their individual intensities (alone or in combination) can be selected under computer control, it is possible: 1) to rapidly illuminate a sample with multiple separate excitation wavelengths to acquire a multispectral excitation image stack and 2) to illuminate a sample with a specifically designed illumination spectrum that allows the direct capture of an image that highlights a specific feature in the sample. To design such an illumination spectrum, first a mathematical algorithm is used to find a weighted linear combination of the multispectral excitation images that represents a feature of interest in a sample. Using the weights in this weighted combination of images, an illumination spectrum is formed by calculating a weighted combination of the excitation channels. Under this illumination spectrum, an image of the sample is captured that represents the feature of interest. Examples of mathematical algorithms that can be used to calculate the weights in the weighted linear combination of excitation images include: 1) linear spectral unmixing to identify certain targets, 2) PCA to obtain the maximum-contrast image, and 3) LDA to enhance the contrast between a target and the rest of the sample. This approach, that was termed SELF imaging in our work, enables the real-time acquisition of fluorescence images of targeted structures. It also increases the signal to noise ratio by reducing the number of images captured. Both of these issues are important in particular for *in vivo* imaging, where fluorescence signals of interest are typically very dim, and subject motion can be a challenge.

Once a set of coefficients and a particular illumination spectrum has been generated to highlight a specific component, it is applicable to other samples comprised of the same components. This approach may ultimately be used to enable imaging of specific objects/areas/tissue types/tissue components of interest in a given object (e.g. a region of

neoplastic tissue, collagen type, NADH, FAD or blood absorption/vascular structures, etc.). This would be achieved through the use of different weighted excitation wavelength illuminations, which can be found mathematically through linear spectral unmixing techniques trained on representative samples.

For *in vivo* macroscopy experiments, we showed that our method can detect different structures in a tissue sample. The method is shown to be robust from one sample to another and under different illumination angles; hence, the detected structures are physical tissue areas which likely represent underlying biological differences.

4.4 Summary

In summary, we presented an imaging approach based on selective spectral illumination generated by a spectrally programmable light engine for the capture of specific targets in the sample, in two imaging modes: transmission and fluorescence.

For the transmission imaging mode, we derived the formula for calculating the illumination profile (the illumination intensity as a function of the wavelength) under which an image equivalent to the weighted sum of a series of multispectral images can be captured. We refer to this imaging approach as “Selective Illumination Transmission Imaging”. We used LDA for calculating the weighted sum of the images in the multispectral image stack that maximizes the contrast between a specific component in the sample and the rest of the sample. We successfully generated illumination profiles for the capture of images of specific objects of interest in the sample. Using this approach, for the capture of one linear discriminant function to highlight one specific object, two image captures were required: one image under the positive segment of the calculated illumination profile, and one under the inverted negative portion of the illumination profile. This enables the spectral discrimination of a specific chromogen to require as few as two image captures, resulting in reduced image acquisition time. This also reduces the image processing time, since the LDA image is directly captured, and will not be calculated. Minimal processing is thus required for the identification of the object of interest in the captured LDA image.

For the fluorescence imaging mode, we presented a multispectral excitation imaging system based on spectrally programmable illumination and its suitability for *in vitro* and *in vivo* fluorescence microscopy and macroscopy. This approach makes use of the differing absorption spectra of different fluorophores or similar fluorophores in different environments (or at different depths within tissue).

For fluorescence microscopy, we showed that it is feasible to successfully remove tissue autofluorescence in the multispectral stack of excitation images, separate the multiple fluorescence signals, and detect the signals from the same fluorophore in different environments using only one filter-cube. We also demonstrated how to perform spectral unmixing on a multispectral excitation image stack using the non-negative least squares techniques, with initial estimates found automatically through the SIMPLISMA algorithm. While spectral unmixing can also be performed on routinely acquired multispectral emission

image stacks, spectral unmixing performed on excitation images separates the fluorescence signals based on their absorption spectra. This can prove useful in cases where the emission spectra of the fluorophores are similar but their excitation spectra are different.

We proposed the SELF imaging approach for the capture of specific fluorescence targets. In a proof of principle experiment with plastic pens, we performed PCA on a multispectral stack of excitation images, to reduce the dimensionality of the multispectral data, and to present the information in as few images as possible. Once we determined the coefficients for the principal components, we used those coefficients to calculate illumination profiles (spectra) for the direct capture of the principal component images. These illumination profiles were made using the OneLight Spectra light engine. Under these specific illuminations, we were able to directly capture the principal component images. In another experiment, we generated illumination profiles for the direct capture of unmixed fluorescence signals that have different excitation spectra, where the linear weights were calculated through linear spectral unmixing. Again, we were able to directly capture the unmixed images under these designed illumination profiles.

This is a significant result that allows the capture of images of fluorescing samples representing particular components (fluorophores) in those samples in a fast and more photon and memory efficient manner, making this method very attractive for real-time *in vivo* imaging. It should be noted that we are assuming that the required coefficients (weights) are generated from one image stack, and are then used for direct capture of images of other similar objects.

SELF imaging has the potential to be further developed to be ultimately used to quickly and accurately extract information regarding the character and composition of tissues *in vivo*. This information may, in turn, have the potential to impact the management of a variety of diseases, contributing to more powerful diagnostic algorithms or more accurate disease boundary delineations. We see great potential clinical utility for this novel imaging approach.

Chapter 5. Conclusion

This research was motivated by the need for quantitative analysis of molecular biomarkers in histopathological tissue sections with the ultimate goal of achieving personalized treatment and management of cancer.

This dissertation was based on the following three hypotheses:

1) The quantification of the spatial associations of the cells that lie within a tissue section and share the same molecular profile, e.g. express the same protein, increases the biological relevance of available biomarker data.

2) The use of multispectral imaging combined with the appropriate analysis methods improves the results of cell nuclei identification and cellular level biomarker measurement.

3) Specifically designed spectral illumination enables the extraction of the relevant information about biomarkers using a smaller number of captured images compared with the number of images in a multispectral image stack.

These hypotheses were addressed in chapters 2 to 4, respectively. In the following three sections, we present a summary of the results of each chapter, and highlight the contributions of this thesis, and suggest some future directions for continuing and building upon this research.

5.1 Quantification of the Spatial Distribution of Biomarkers

5.1.1 Summary of Results

- In chapter 2, we developed a framework for the objective and quantitative assessment of the distribution of a biomarker in a tissue section. Several novel features describing the spatial behavior of a biomarker within epithelial tissue were proposed.
- Using these features, we assessed the spatial distribution of the Ki67 expressing cells in normal and pre-neoplastic bronchial epithelial regions, for the first time ever. We showed that the proposed features correlate with the histological evaluation, which is the “gold standard”.
- Given the specific challenges associated with the reproducibility of histopathological evaluation of early lung lesions, and the Ki67 labeling index specifically, we evaluated

the reproducibility of these new features. Our results indicated that these features were indeed more reproducible than the Ki67 labeling index. These features are therefore more likely to have utility as biomarkers for disease management, compared to Ki67 labeling index.

- The performance of the proposed features was then evaluated as a means of predicting the progression likelihood in early stage lesions using the follow-up data of the lesions. Several of these features were shown to have significant association with progression, while neither the histology nor the Ki67 labeling index does. This demonstrated our hypothesis that the knowledge of the spatial behavior of the Ki67 expressing cells increases the biological relevance of the expression of the proliferation biomarker, Ki67.

5.1.2 Suggested Future Directions

- In the data available for this study, the number of cases progressing to the high grade lesions was limited. Therefore the endpoint that was used in the outcome analysis was progression to or persisting as mild dysplasia or higher. The risk of progressing from metaplasia to dysplasia may not be the same as the risk of developing severe dysplasia or CIS. Thus the associations of the Ki67 proposed features with progression to high grade lesions still need to be verified. For this verification, access to a wider range of intraepithelial lesions is needed.
- The proposed Ki67 features may have higher prediction power when combined with risk factors such as age, smoking status, history of lung diseases such as Chronic Obstructive Pulmonary Lung Disease (COPD), etc. A multivariate analysis of Ki67 features with these factors, when available, may therefore be worthwhile.

5.2 Cellular Level Quantification of the Expression of Biomarkers

5.2.1 Summary of Results

In chapter 3, we developed methods for the quantitative assessment of immunohistochemical biomarkers abundance in IHC-labeled tissue sections. These methods address the following:

1) Identification and separation of the labeled nuclei and labeled biomarkers

- We captured a multispectral image stacks from each IHC-labeled tissue section available in our study. Each multispectral image stack was acquired by capturing 12 images from a tissue section under 12 narrow illumination spectra. These illumination spectra spanned the range of 420nm to 640 nm and were generated by a spectrally programmable light engine.
- We explored the application of two methods for the extraction of the information from the multispectral image stacks: linear spectral unmixing, and linear discriminant analysis. Given the multispectral image stack of a sample with N labels, each of these methods transforms the images in the multispectral image stack into N images, where each image represents the distribution of one label.
- We studied the application of the MCR-ALS algorithm for the linear spectral unmixing of multispectral images of immunostained tissue sections. For each sample with N labels, linear spectral unmixing generates N images, where each image represents the distribution (concentration) of one label, e.g. in the case of a sample with a nuclear label and two IHC labels, linear spectral unmixing generates one image representing the nuclear label and two images representing the two IHC labels. The first image therefore represents the nuclei (referred to as the “nuclei image”) while the second and third images represent the first and second IHC biomarkers (referred to as the “immunostain images”). Linear spectral unmixing works by unmixing the contribution of each label present in a pixel to the absorption (optical density) in that pixel.
- The MCR-ALS algorithm requires initial estimates of either the spectra of the components (labels) or the concentration maps of the components. We proposed three methods for calculating initial estimates. The first two methods calculate initial estimates of spectra of the components (labels) present in a sample. The third method calculates the concentration maps of the components. All three methods use only the image data. These methods are as follows:
 - 1) The first method is for calculating the initial estimates of the spectra. It involves performing the SIMPLISMA analysis of Windig et al.¹²⁸ in several randomly selected regions within a labeled tissue section. When for each label there are pixels in the image stack that have an absorption contribution from only that label (“pure” pixels), the

spectral unmixing procedure performed using the initial estimates calculated by this method was shown to result in correctly unmixed concentration maps.

2) The second method is for finding the initial estimate of the spectrum of a component C when that component is not present by any “pure” pixels in the image stack, i.e. there are no pixels in the image stack that have an absorption contribution from only component C. We proposed a subtraction algorithm that automatically subtracts estimated fractions of the spectra of other components present in the sample from the spectrum of component C. The fractions of the spectra are estimated such that the value of the resulting spectrum at certain wavelengths is equal to zero or to some *a priori* known value. The resulting spectrum is used as the initial estimate of the spectrum of this component. We showed the utility of this algorithm for estimating the spectrum of a component when mixed with one other component (DAB mixed with H in the Ki67 samples) and when mixed with two other components (VR mixed with H and DAB in the “Ki67 & P16” samples).

3) The third method provides the initial estimates of the concentration maps of the components. It starts by employing the SIMPLISMA analysis to find the spectral bands in which the absorption of one of the components is significantly higher than the absorption of the other components. It then uses the thresholded optical density maps in those spectral bands as the initial estimates of the concentration maps.

These three proposed methods remove the need for the use of singly labeled tissue sections for estimating the spectra of the pure components.

- We compared the label distribution maps obtained by performing linear spectral unmixing on multispectral images with those obtained from RGB images. We demonstrated that, for samples with two labels, multispectral images and RGB images result in visually similar distribution maps. However, for samples with three labels, only the spectral unmixing performed on multispectral images was able to correctly generate images representing the distribution of each label.
- We presented a new way of using the linear discriminant analysis (LDA) for the identification of the labeled objects in a tissue sample. For a sample with N labels, N linear discriminant functions were trained to calculate N greyscale images where each grayscale image represents one label. We showed that, for each sample with multiple labels, LDA was able to generate images representing each label, e.g. an image

representing the labeled nuclei, and another image representing a labeled biomarker. The images representing the labeled objects were visually similar to those obtained by spectral unmixing.

- The images representing the labeled nuclei and the labeled biomarker(s) that were generated using either method (spectral unmixing or LDA) were later quantitatively compared in terms of: 1) the accuracy of nuclei segmentation results and 2) the contrast in the images of the labeled biomarkers.

2) Automated segmentation of nuclei

- We presented a fully automated nuclei segmentation procedure. The segmentation procedure uses an adapted version of the edge relocation algorithm originally developed by MacAulay et al.¹³⁶ It is able to robustly delineate the nuclei and separate the touching nuclei from each other.
- Five different representations of the labeled nuclei were studied. The first four were generated using: 1) spectral unmixing on multispectral images, 2) spectral unmixing on RGB images, 3) LDA of multispectral images, 4) LDA of RGB images, and 5) the fifth representation of the labeled nuclei that was studied was the red component of the RGB image, which has the highest contrast between the nuclei and the rest of the image amongst the three RGB components. This is because in the red channel the absorption of the brown IHC biomarker label is lowest amongst the red, green, and blue channels and the absorption of the blue counterstain used for labeling the nuclei is highest amongst the three RGB channels. For each of the 58 immunostained samples, the segmentation procedure was performed on the five different representations of nuclei listed above.
- The segmentation procedure was validated using the set of fifty eight immunostained images. The nuclei in these images had been previously biomarker by an expert observer. This set comprised more than 22,000 cells. The most accurate segmentation results were obtained using the first one of the five representations of nuclei listed above, which was generated by spectral unmixing of the multispectral images.
- The accuracy of the segmentation, when applied to the nuclei image generated by spectral unmixing of the multispectral images, in the tissue sections immunostained for Ki67, P16, and CD8 proteins was 87%, 81%, and 82% respectively. The accuracy of the manual identification of the nuclei performed by a different expert observer was 92%,

72%, and 86% (data shown in Appendix B). The automated segmentation accuracy measures are slightly lower than the manual segmentation in the cases of Ki67 and CD8 proteins and higher than the manual segmentation in the case of P16 protein. The comparison of the accuracy of the automated segmentation with that of the manual identification indicates the efficiency of the automated procedure in the segmentation of the complex histological samples.

3) Measurement of the abundance of the labeled biomarkers on a cell by cell basis

- Methods for measuring the abundance (amount) of labeled biomarkers on a cell-by-cell basis were proposed. These methods use the previously calculated images which represent the distribution of each label in the sample. These are the “nuclei image” which represents the labeled nuclei and the “immunostain image” which represents the distribution of a labeled IHC biomarker.

For each tissue section, the region or compartment within each cell in which the IHC biomarker of interest is expressed was identified. The abundance of the biomarker within its corresponding cell compartment, in the “immunostain image”, was then measured for each cell. For the nucleus bound biomarkers, the cell compartment within which the biomarker is expressed is the nucleus. The regions in which the biomarker is present are therefore found by the segmentation of the nuclei in the “nuclei image”. If the biomarker is expressed in the cytoplasm or the cell membrane, the location of the boundary of each cell needs to be identified as well as the boundary of the cell nucleus. This is to enable the identification of the region in the “immunostain image” in which the biomarker corresponding to each cell needs to be measured. Because neither the cell surface nor the cell boundary was labeled, the exact location of the cell boundary could not be found. Therefore we approximated the location of the boundary of each cell by partitioning the area of interest within each tissue section into regions around each of the segmented nuclei using either the Voronoi tessellation or the nearest neighbour tessellation methods. The region around each nucleus was assumed to correspond to a cell.

- A thresholding scheme was suggested to classify each cell into positive or negative according to the abundance of the biomarker within that cell (A positive cell indicates the presence of the biomarker of interest in that cell).The threshold was calculated using the mean and standard deviation of the pixel intensities in an area in the immunostain

image where no immunostaining was present. The thresholding scheme was able to reliably identify the biomarker-expressing cells, as evaluated by a human observer.

- For each tissue section, five different images representing the distribution of the labeled biomarker (“immunostain image”) were studied. The first four representations were generated by 1) spectral unmixing of multispectral images, 2) spectral unmixing of RGB images, 3) LDA of multispectral images, 4) LDA of RGB images, and 5) the fifth “immunostain image” that was studied was the blue component of the RGB image, which has the highest contrast between the biomarker label and the rest of the image, amongst the three RGB components. This is because in the blue channel the absorption of the brown IHC biomarker label was the highest amongst the red, green, and blue channels and the absorption of the blue counterstain used for labeling the nuclei was the lowest amongst the three RGB channels.
- The first four methods performed equally well in identifying the positive cells when judged by visual examination of the results. The last representation, however, resulted in the lowest contrast between the intensities of the labeled and non-labeled pixels in the “immunostain image”.
- The above results are based on visual identification of positive cells based on their color in the RGB images. We note that in the absence of relevant biological information about the outcome of the lesions or tumors, no other standard for the definition and determination of the positive cells exists. Therefore without the outcome information, the optimal method for the generation of the “immunostain image” cannot be selected.

5.2.2 Suggested Future Directions

- The segmentation procedure presented in this study would benefit from a quality control step. We envision such a step to depend on a quality metric that measures the degree to which a certain object resembles a nucleus. Such a metric would determine if the segmentation of that object is satisfactory or if the object can be further segmented into smaller objects. This metric can be based on features that explain the shape of a segmented object. One way of defining such a metric would be to generate a training set of objects consisting of two groups. The first group consists of objects that are correctly segmented and the second group is formed by clusters of objects that need to be further

broken apart. For every object, several features explaining the shape of the objects are calculated. A feature selection algorithm can then be employed to find the subset of features, or a combination of them, that best identifies the objects that need to be further split. In the current version of the segmentation procedure, only the area (size) of the object is used to determine if the object needs further segmentation. The area of an object is not the best descriptor due to variability in the areas of the nuclei within a tissue section. Other shape features such as the existence of concavities in the boundary of the object may be more reliable. The addition of a quality metric in the segmentation procedure allows for a conditional iterative algorithm that repeats the segmentation of the object until based on the quality metric of the object, no more iterations are needed. This would improve the robustness of the segmentation procedure, since the object would not go through unnecessary rounds of splitting.

- The ultimate biological relevance of the quantitative biomarker expression data can only be determined by using the correlation of such data with the true biological outcome, e.g. progression or regression of the lesion into higher or lower grades of neoplasia or dysplasia, improved survival rate, improved effectiveness of the treatment, etc. To assess the effectiveness of the quantitative analysis of a certain biomarker in a tissue the association of the biomarker data with the biological outcome needs to be studied in large sample sets. It is only via such associations that the impact of different segmentation and biomarker measurement methods can be compared.

5.3 “Selective Spectral Illumination” for Imaging of Tissue-based Biomarkers

5.3.1 Summary of Results

- In chapter 4, we presented the concept of “selective spectral illumination” for two modes of imaging: transmission (bright-field) and fluorescence. This concept is based on the use of a spectrally programmable light engine that can control the wavelengths and intensities of the illumination it generates, and can generate any mix of wavelengths in the visible light wavelengths. We performed proof of principle experiments for both the transmission and the fluorescence imaging modes to show this concept.

- We studied the case when a sample is sequentially illuminated by a series of spectrally narrow illumination bands $\varphi_i, i = 1, 2, \dots, N$ and at each illumination band φ_i an image of the sample is captured to generate a stack of multispectral images $I_i, i = 1, 2, \dots, N$. We demonstrated that we can generate an image representing a particular component in the sample by calculating a specific weighted sum of the multispectral images ($\sum_{i=1}^N c_i I_i$). We showed that we can then design an illumination spectrum composed of a specifically determined weighted sum of the illumination bands ($\sum_{i=1}^N v_i \varphi_i$). This illumination spectrum is generated by a spectrally programmable light engine. Under this illumination, the image representing the component of interest would be directly captured. Different mathematical methods may be used to obtain the weighted linear combination of the multispectral images that highlights a particular component of interest. We presented examples of the use of the linear discriminant analysis (LDA) in transmission imaging, and the use of either principal component analysis (PCA) or linear spectral unmixing in fluorescence imaging for transforming the multispectral image stack into images representing the distribution of certain components in the sample. Regardless of the mathematical method of finding a weighted sum of the multispectral image, once these weights are determined, they are used to design a corresponding illumination spectrum for the direct capture of the specific image of interest. This approach is referred to as “selective spectral illumination imaging”.

Typically, several images of a tissue specimen are captured to generate a multispectral image stack. Not all the information within this multispectral image stack is relevant or necessary for the extraction of the distribution of a certain component within the specimen. Using our selective spectral illumination imaging approach, we can obtain only the most relevant information, for the identification of a certain component within the sample, using one or two captured images. Once a particular illumination spectrum is designed for imaging a specific component, the same spectrum can be used for imaging other samples comprised of the same components.

- Conventional multispectral fluorescence imaging involves the acquisition of images of fluorescence emissions in a series of narrow spectral bands. This is referred to as multispectral emission imaging. Linear spectral unmixing can be performed on the multispectral fluorescence emission images to unmix the total fluorescence signal

detected from a fluorescing sample and to obtain the fluorescence signals from each of the distinct fluorescing components comprising that sample. In this way, fluorescence signals are separated using the differences in their emission spectra. We have presented a multispectral excitation imaging system for the acquisition of images of fluorescence emissions when the sample is illuminated with a series of narrow-band illumination spectra. We demonstrated that, similar to the spectral unmixing using the multispectral emission image stack, we can perform linear spectral unmixing using the multispectral excitation images to obtain the fluorescence signals from each of the distinct fluorescing components comprising a sample. The unmixing, however, is based on the differences in the excitation spectra of the underlying fluorescing components.

This approach can potentially prove to be advantageous in certain applications, e.g. when the fluorescence change in pathologically changed tissue is more pronounced in the excitation spectrum of the tissue rather than the emission spectrum. Use of the differences in the excitation spectra of fluorescence components can also help increase the number of simultaneous labels that can be distinctly detected on a single microscopy slide (as labels can be separated by their excitation spectra and not just their emission spectra). This method of fluorescence imaging may also result in better signal to noise ratio as there are no optical or mechanical elements in the imaging light path. This results in minimal image degradation, whereas in emission imaging approaches there is a filter wheel or tunable filter in the detection light path.

5.3.2 Suggested Future Directions

By employing proof of principle experiments, we demonstrated the utility of selective spectral illumination. Experiments with a larger number of samples are needed to assess the accuracy of this imaging approach. Differences between the image calculated as a weighted sum of the multispectral images and that captured under the designed illumination spectrum need to be quantified. These differences are likely to be caused by the background illumination generated by the light engine, due to the fact that the OneLight system has insufficient out of band rejection. The impact of such differences on the extraction of the biomarker data needs to be determined. For example, in transmission imaging, the accuracy

of nuclei segmentation and biomarker measurement can be compared between the multispectral images, and the images captured under selective illumination.

In the fluorescence excitation imaging case, the advantage of using the excitation imaging needs to be studied by imaging specimens containing fluorophores with similar emission spectra but different excitation spectra. Direct comparison of the unmixing results using the multispectral emission images and the multispectral excitation images is necessary to compare the accuracy of unmixing, and signal to noise ratio of each imaging technique.

Selective Excitation Fluorescence Imaging may prove useful in detecting the changes in the tissue autofluorescence caused by early malignancies. To investigate these potential utilities, multispectral excitation images can be collected from non-cancerous and various grades of pre-cancerous and cancerous tissues. Regions of changes in the tissue autofluorescence detected in the multispectral excitation images can then be compared with the histology grades and disease boundaries as determined by the pathologist. If regions with certain autofluorescence changes that are specific to histology changes are found, the excitation spectrum of such regions can be used to design certain illumination spectra to directly highlight these changes.

References

- 1 Canadian-Cancer-Society. *General cancer statistics at a glance*, <http://www.cancer.ca/Canada-wide/About%20cancer/Cancer%20statistics/Stats%20at%20a%20glance/General%20cancer%20stats.aspx?sc_lang=en> (2012).
- 2 Schilsky, R. L. Personalized medicine in oncology: the future is now. *Nature Reviews Drug Discovery* **9**, 363-366 (2010).
- 3 Slamon, D. J. *et al.* Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *New England Journal of Medicine* **344**, 783-792 (2001).
- 4 Gately, K. *et al.* The role of the molecular footprint of EGFR in tailoring treatment decisions in NSCLC. *Journal of clinical pathology* **65**, 1-7 (2012).
- 5 Brennan, D. J., O'Connor, D. P., Rexhepaj, E., Ponten, F. & Gallagher, W. M. Antibody-based proteomics: fast-tracking molecular diagnostics in oncology. *Nature Reviews Cancer* **10**, 605-617 (2010).
- 6 Ramos-Vara, J. Technical aspects of immunohistochemistry. *Veterinary Pathology Online* **42**, 405-426 (2005).
- 7 Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194**, 23-28 (1976).
- 8 Kent, O. & Mendell, J. A small piece in the cancer puzzle: microRNAs as tumor suppressors and oncogenes. *Oncogene* **25**, 6188-6196 (2006).
- 9 Coons, S. W. & Johnson, P. C. Regional heterogeneity in the DNA content of human gliomas. *Cancer* **72**, 3052-3060 (1993).
- 10 Morikawa, K. & Fidler, I. Heterogeneous response of human colon cancer cells to the cytostatic and cytotoxic effects of recombinant human cytokines: interferon-alpha, interferon-gamma, tumor necrosis factor, and interleukin-1. *Journal of biological response modifiers* **8**, 206 (1989).
- 11 Zhou, G. H. K., Sechrist, G. L., Brattain, M. G. & Mulder, K. M. Clonal heterogeneity of the sensitivity of human colon carcinoma cell lines to TGF β isoforms. *Journal of cellular physiology* **165**, 512-520 (1995).
- 12 Emily, M., Morel, D., Marcelpoil, R. & François, O. Spatial correlation of gene expression measures in tissue microarray core analysis. *Journal of Theoretical Medicine* **6**, 33-39 (2005).
- 13 Walker, R. Quantification of immunohistochemistry—issues concerning methods, utility and semiquantitative assessment I. *Histopathology* **49**, 406-410 (2006).
- 14 Taylor, C. & Levenson, R. Quantification of immunohistochemistry—issues concerning methods, utility and semiquantitative assessment II. *Histopathology* **49**, 411-424 (2006).
- 15 Al-Janabi, S., Huisman, A. & Van Diest, P. J. Digital pathology: current status and future perspectives. *Histopathology* (2011).
- 16 Rizzardi, A. E. *et al.* Quantitative comparison of immunohistochemical staining measured by digital image analysis versus pathologist visual scoring. *Diagnostic Pathology* **7**, 42 (2012).

- 17 Mohammed, Z. *et al.* Comparison of visual and automated assessment of HER2 status and their impact on outcome in primary operable invasive ductal breast cancer. *Histopathology* (2012).
- 18 Diaz, L. K., Sahin, A. & Sneige, N. Interobserver agreement for estrogen receptor immunohistochemical analysis in breast cancer: a comparison of manual and computer-assisted scoring methods. *Annals of Diagnostic Pathology* **8**, 23-27 (2004).
- 19 Gokhale, S. *et al.* Assessment of two automated imaging systems in evaluating estrogen receptor status in breast carcinoma. *Applied Immunohistochemistry & Molecular Morphology* **15**, 451 (2007).
- 20 Turbin, D. A. *et al.* Automated quantitative analysis of estrogen receptor expression in breast carcinoma does not differ from expert pathologist scoring: a tissue microarray study of 3,484 cases. *Breast cancer research and treatment* **110**, 417-426 (2008).
- 21 Fiore, C. *et al.* Utility of multispectral imaging in automated quantitative scoring of immunohistochemistry. *Journal of clinical pathology* **65**, 496-502 (2012).
- 22 Meert, A. P. *et al.* Ki67 expression in bronchial preneoplastic lesions and carcinoma in situ defined according to the new 1999 WHO/IASLC criteria: a preliminary study. *Histopathology* **44**, 47-53 (2004).
- 23 Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481**, 306-313 (2012).
- 24 *Lung cancer statistics at a glance*, <http://www.cancer.ca/canada-wide/about%20cancer/cancer%20statistics/stats%20at%20a%20glance/lung%20cancer.aspx?sc_lang=en#> (2012).
- 25 *What are the key statistics about lung cancer?*, <<http://www.cancer.org/Cancer/LungCancer-Non-SmallCell/DetailedGuide/non-small-cell-lung-cancer-key-statistics>> (
- 26 Institute, N. C. *Cancer Stat Fact Sheets, Cancer of the Lung and Bronchus* <<http://seer.cancer.gov/statfacts/html/lungb.html#survival>> (2012).
- 27 Greenberg, A. K., Yee, H. & Rom, W. N. Preneoplastic lesions of the lung. *Respir Res* **3**, 20 (2002).
- 28 Wistuba, II & Gazdar, A. F. Lung cancer preneoplasia. *Annu Rev Pathol* **1**, 331-348, doi:10.1146/annurev.pathol.1.110304.100103 (2006).
- 29 Tomaszek, S. C. & Wigle, D. A. Surgical management of lung cancer. *Semin Respir Crit Care Med* **32**, 69-77, doi:10.1055/s-0031-1272871 (2011).
- 30 Collins, L. G., Haines, C., Perkel, R. & Enck, R. E. Lung cancer: diagnosis and management. *American family physician* **75**, 56 (2007).
- 31 Brambilla, E., Travis, W., Colby, T., Corrin, B. & Shimosato, Y. The new World Health Organization classification of lung tumours. *European Respiratory Journal* **18**, 1059-1068 (2001).
- 32 Politi, E. N. & Syrigos, K. N. Preneoplastic Lesions of the Lung. *Tumors of the Chest*, 81-91 (2006).
- 33 Lam, S. *et al.* Localization of bronchial intraepithelial neoplastic lesions by fluorescence bronchoscopy. *Chest* **113**, 696-702 (1998).
- 34 Banerjee, A. K. Preinvasive lesions of the bronchus. *Journal of Thoracic Oncology* **4**, 545 (2009).
- 35 Szabo, E. Lung epithelial proliferation: a biomarker for chemoprevention trials? *Journal of the National Cancer Institute* **93**, 1042-1043 (2001).

- 36 Lam, S., MacAulay, C., LeRiche, J. C. & Gazdar, A. F. Key issues in lung cancer chemoprevention trials of new agents. *Tumor Prevention and Genetics*, 182-195 (2003).
- 37 Szabo, E. Assessing efficacy in early-phase cancer prevention clinical trials: the case of Ki-67 in the lung. *Cancer Prevention Research* **3**, 128-131 (2010).
- 38 Mao, J. T. & Durvasula, R. Lung cancer chemoprevention: current status and future directions. *Current Respiratory Care Reports*, 1-12 (2012).
- 39 Jeanmart, M. *et al.* Value of immunohistochemical markers in preinvasive bronchial lesions in risk assessment of lung cancer. *Clinical cancer research* **9**, 2195-2203 (2003).
- 40 Breuer, R. H. *et al.* The natural course of preneoplastic lesions in bronchial epithelium. *Clin Cancer Res* **11**, 537-543 (2005).
- 41 McWilliams, A., MacAulay, C., Gazdar, A. F. & Lam, S. Innovative molecular and imaging approaches for the detection of lung cancer and its precursor lesions. *Oncogene* **21**, 6949-6959 (2002).
- 42 Lantuéjoul, S., Salameire, D., Salon, C. & Brambilla, E. Pulmonary preneoplasia—sequential molecular carcinogenetic events. *Histopathology* **54**, 43-54 (2009).
- 43 Szabo, E. Altered Histology Provides a Positive Clinical Signal in the Bronchial Epithelium. *Cancer Prevention Research* **4**, 775 (2011).
- 44 Jonat, W. & Arnold, N. Is the Ki-67 labelling index ready for clinical use? *Ann Oncol* **22**, 500-502, doi:10.1093/annonc/mdq732 (2011).
- 45 Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *cell* **100**, 57-70 (2000).
- 46 Martin, B. *et al.* Ki-67 expression and patients survival in lung cancer: systematic review of the literature with meta-analysis. *British journal of cancer* **91**, 2018-2025 (2004).
- 47 Madewell, B. R. Cellular proliferation in tumors: a review of methods, interpretation, and clinical applications. *Journal of Veterinary Internal Medicine* **15**, 334-340 (2001).
- 48 van Diest, P. J., Brugal, G. & Baak, J. Proliferation markers in tumours: interpretation and clinical value. *Journal of clinical pathology* **51**, 716-724 (1998).
- 49 Shirakawa, Y. *et al.* Topological analysis of p21WAF1/CIP1 expression in esophageal squamous dysplasia. *Clin Cancer Res* **6**, 541-550 (2000).
- 50 Bulten, J. *et al.* MIB1, a promising marker for the classification of cervical intraepithelial neoplasia. *J Pathol* **178**, 268-273, doi:10.1002/(sici)1096-9896(199603)178:3<268::aid-path482>3.0.co;2-4 (1996).
- 51 Kruse, A. J. *et al.* Ki-67 immunoquantitation in cervical intraepithelial neoplasia (CIN): a sensitive marker for grading. *J Pathol* **193**, 48-54, doi:10.1002/1096-9896(2000)9999:9999<::aid-path719>3.0.co;2-e (2001).
- 52 Kruse, A. J. *et al.* Evaluation of MIB-1-positive cell clusters as a diagnostic marker for cervical intraepithelial neoplasia. *Am J Surg Pathol* **26**, 1501-1507 (2002).
- 53 Kruse, A. J. *et al.* Ki67 predicts progression in early CIN: validation of a multivariate progression-risk model. *Cell Oncol* **26**, 13-20 (2004).
- 54 van der Avoort, I. A. *et al.* MIB1 expression in basal cell layer: a diagnostic tool to identify premalignancies of the vulva. *Mod Pathol* **20**, 770-778, doi:10.1038/modpathol.3800796 (2007).

- 55 Hittelman, W. N. *et al.* Proliferative changes in the bronchial epithelium of former smokers treated with retinoids. *J Natl Cancer Inst* **99**, 1603-1612, doi:10.1093/jnci/djm205 (2007).
- 56 Hoshino, H. *et al.* Biological features of bronchial squamous dysplasia followed up by autofluorescence bronchoscopy. *Lung Cancer* **46**, 187-196, doi:10.1016/j.lungcan.2004.04.028 (2004).
- 57 Miller, Y. E. *et al.* Bronchial epithelial Ki-67 index is related to histology, smoking, and gender, but not lung cancer or chronic obstructive pulmonary disease. *Cancer Epidemiol Biomarkers Prev* **16**, 2425-2431, doi:10.1158/1055-9965.epi-07-0220 (2007).
- 58 Lam, S. *et al.* A randomized phase IIb trial of pulmicort turbuhaler (budesonide) in people with dysplasia of the bronchial epithelium. *Clin Cancer Res* **10**, 6502-6511, doi:10.1158/1078-0432.ccr-04-0686 (2004).
- 59 Travis, W., Colby, T., Corrin, B., Shimosato, Y. & Brambilla, E. in *World Health Organization Pathology Panel: World Health Organization. International Classification of Tumors*. 5-30 (Springer Verlag, 1999).
- 60 Kamalov, R. *et al.* A Java application for tissue section image analysis. *Comput Methods Programs Biomed* **77**, 99-113, doi:S0169260704001026 [pii] 10.1016/j.cmpb.2004.04.003 [doi] (2005).
- 61 Aurenhammer, F. Voronoi Diagrams - A Survey of a Fundamental Geometric Data Structure. *ACM Computing Surveys* **23**, 345-405 (1991).
- 62 Gonzalez, R. C. & Woods, R. E. *Digital Image Processing*. 2nd edn, (Prentice Hall, 2002).
- 63 Marcelpoil, R. & Usson, Y. Methods for the study of cellular sociology: Voronoi diagrams and parametrization of the spatial relationships. *Journal of Theoretical Biology* **154**, 359-369 (1992).
- 64 Guillaud, M. *et al.* Quantitative histopathological analysis of cervical intra-epithelial neoplasia sections: methodological issues. *Cell Oncol* **26**, 31-43 (2004).
- 65 Guillaud, M. *et al.* Quantitative architectural analysis of bronchial intraepithelial neoplasia. *Proceedings of SPIE*, 74 (2000).
- 66 Hogg, R. V., and J. Ledolter. *Engineering Statistics*. (MacMillan, 1987).
- 67 Dunn, O. J. Multiple comparisons among means. *Journal of the American Statistical Association*, 52-64 (1961).
- 68 Neumüller, R. A. & Knoblich, J. A. Dividing cellular asymmetry: asymmetric cell division and its implications for stem cells and cancer. *Genes & development* **23**, 2675-2699 (2009).
- 69 Bota, S. *et al.* Follow-up of bronchial precancerous lesions and carcinoma in situ using fluorescence endoscopy. *American journal of respiratory and critical care medicine* **164**, 1688-1693 (2001).
- 70 Alaa RM, M. *et al.* Risk of lung cancer in patients with preinvasive bronchial lesions followed by autofluorescence bronchoscopy and chest computed tomography. *Lung Cancer* **72**, 303-308 (2011).
- 71 Box, G. E. P. & Cox, D. R. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 211-252 (1964).
- 72 StatSoft-Inc. *Electronic Statistics Textbook*, <<http://www.statsoft.com/textbook/>> (2012).

- 73 Banerjee, A. K. Preinvasive lesions of the bronchus. *J Thorac Oncol* **4**, 545-551, doi:10.1097/JTO.0b013e31819667bd (2009).
- 74 Venmans, B. J. *et al.* Observer variability in histopathologic reporting of bronchial biopsy specimens: influence on the results of autofluorescence bronchoscopy in detection of preinvasive bronchial neoplasia. *Journal of Bronchology & Interventional Pulmonology* **7**, 210-214 (2010).
- 75 Nicholson, A. G. *et al.* Reproducibility of the WHO/IASLC grading system for preinvasive squamous lesions of the bronchus: a study of inter-observer and intra-observer variation. *Histopathology* **38**, 202-208 (2001).
- 76 Wistuba, II. Histologic evaluation of bronchial squamous lesions: any role in lung cancer risk assessment? *Clin Cancer Res* **11**, 1358-1360, doi:10.1158/1078-0432.ccr-04-2156 (2005).
- 77 Baak, J. P. *et al.* Dynamic behavioural interpretation of cervical intraepithelial neoplasia with molecular biomarkers. *J Clin Pathol* **59**, 1017-1028, doi:10.1136/jcp.2005.027839 (2006).
- 78 Van Niekerk, D. *et al.* p16 and MIB1 improve the sensitivity and specificity of the diagnosis of high grade squamous intraepithelial lesions: methodological issues in a report of 447 biopsies with consensus diagnosis and HPV HCII testing. *Gynecologic oncology* **107**, 233 (2007).
- 79 Kumar, S. *et al.* Subcutaneous panniculitic T-cell lymphoma is a tumor of cytotoxic T lymphocytes. *Human pathology* **29**, 397-403 (1998).
- 80 Jackisch, C. HER-2-positive metastatic breast cancer: optimizing trastuzumab-based therapy. *The oncologist* **11**, 34-41 (2006).
- 81 McCarty, K. S. *et al.* Use of a monoclonal anti-estrogen receptor antibody in the immunohistochemical evaluation of human tumors. *Cancer research* **46**, 4244s (1986).
- 82 McCarty Jr, K., Miller, L., Cox, E., Konrath, J. & McCarty Sr, K. Estrogen receptor analyses. Correlation of biochemical and immunohistochemical methods using monoclonal antireceptor antibodies. *Archives of pathology & laboratory medicine* **109**, 716 (1985).
- 83 Remmele, W. & Stegner, H. Recommendation for uniform definition of an immunoreactive score (IRS) for immunohistochemical estrogen receptor detection (ER-ICA) in breast cancer tissue]. *Der Pathologe* **8**, 138 (1987).
- 84 Reiner, A. *et al.* Estrogen receptor analysis on biopsies and fine-needle aspirates from human breast carcinoma. Correlation of biochemical and immunohistochemical methods using monoclonal antireceptor antibodies. *The American journal of pathology* **125**, 443 (1986).
- 85 Paik, S. *et al.* Real-world performance of HER2 testing—national surgical adjuvant breast and bowel project experience. *Journal of the National Cancer Institute* **94**, 852-854 (2002).
- 86 Biesterfeld, S., Veuskens, U., Schmitz, F. J., Amo-Takyi, B. & Böcking, A. Interobserver reproducibility of immunocytochemical estrogen-and progesterone receptor status assessment in breast cancer. *Anticancer research* **16**, 2497-2500 (1996).
- 87 Sallinen, P. *et al.* Sources of variation in the assessment of cell proliferation using proliferating cell nuclear antigen immunohistochemistry. *Analytical and quantitative*

- cytology and histology/the International Academy of Cytology [and] American Society of Cytology* **16**, 261 (1994).
- 88 Ravn, V. *et al.* Reproducibility of subjective immunohistochemical estrogen-and progesterone receptor determination in human endometrium. *Pathology-Research and Practice* **189**, 1015-1022 (1993).
- 89 Kay, E., Walsh, C., Cassidy, M., Curran, B. & Leader, M. C-erbB-2 immunostaining: problems with interpretation. *Journal of clinical pathology* **47**, 816-822 (1994).
- 90 Seidal, T., Balaton, A. J. & Battifora, H. Interpretation and quantification of immunostains. *The American journal of surgical pathology* **25**, 1204 (2001).
- 91 O'Leary, T. J. Standardization in immunohistochemistry. *Applied Immunohistochemistry & Molecular Morphology* **9**, 3 (2001).
- 92 Cregger, M., Berger, A. J. & Rimm, D. L. Immunohistochemistry and quantitative analysis of protein expression. *Archives of pathology & laboratory medicine* **130**, 1026-1030 (2006).
- 93 Van Der Loos, C. M. Multiple immunoenzyme staining: methods and visualizations for the observation with spectral imaging. *Journal of Histochemistry & Cytochemistry* **56**, 313 (2008).
- 94 Bacus, S., Flowers, J., Press, M., Bacus, J. & McCarty Jr, K. The evaluation of estrogen receptor in primary breast carcinoma by computer-assisted image analysis. *American journal of clinical pathology* **90**, 233 (1988).
- 95 Bacus, S., Bacus, J., Slamon, D. & Press, M. HER-2/neu oncogene expression and DNA ploidy analysis in breast cancer. *Archives of pathology & laboratory medicine* **114**, 164 (1990).
- 96 Makkink-Nombrado, S., Baak, J., Schuurmans, L., Theeuwes, J. & Van der Aa, T. Quantitative immunohistochemistry using the CAS 200/486 image analysis system in invasive breast carcinoma: a reproducibility study. *Analytical cellular pathology: the journal of the European Society for Analytical Cellular Pathology* **8**, 227 (1995).
- 97 Esteban, J., Battifora, H., Warsi, Z., Bailey, A. & Bacus, S. Quantification of estrogen receptors on paraffin-embedded tumors by image analysis. *Modern pathology: an official journal of the United States and Canadian Academy of Pathology, Inc* **4**, 53 (1991).
- 98 Charpin, C. *et al.* Multiparametric study (SAMBA 200) of estrogen receptor immunocytochemical assay in 400 human breast carcinomas: analysis of estrogen receptor distribution heterogeneity in tissues and correlations with dextran coated charcoal assays and morphological data. *Cancer research* **48**, 1578-1586 (1988).
- 99 Kuyatt, B., Reidy, C., Hui, K. & Jordan, W. Quantitation of smooth muscle proliferation in cultured aorta. A color image analysis method for the Macintosh. *Analytical and quantitative cytology and histology/the International Academy of Cytology [and] American Society of Cytology* **15**, 83 (1993).
- 100 Elmoataz, A., Schüpp, S., Clouard, R., Herlin, P. & Bloyet, D. Using active contours and mathematical morphology tools for quantification of immunohistochemical images. *Signal Processing* **71**, 215-226 (1998).
- 101 Willemse, F., Nap, M., Henzen-Logmans, S. C. & Eggink, H. F. Quantification of area percentage of immunohistochemical staining by true color image analysis with application of fixed thresholds. *Analytical and quantitative cytology and histology* **16**, 357-364 (1994).

- 102 Goto, M. *et al.* Chromaticity analysis of immunostained tumor specimens. *Pathology-Research and Practice* **188**, 433-437 (1992).
- 103 Kohlberger, P. *et al.* Modified true-color computer-assisted image analysis versus subjective scoring of estrogen receptor expression in breast cancer: a comparison. *Anticancer research* **19**, 2189-2193 (1999).
- 104 Elie, N., Plancoulaine, B., Signolle, J. P. & Herlin, P. A simple way of quantifying immunostained cell nuclei on the whole histologic section. *Cytometry Part A* **56**, 37-45 (2003).
- 105 Brey, E. M. *et al.* Automated selection of DAB-labeled tissue for immunohistochemical quantification. *Journal of Histochemistry & Cytochemistry* **51**, 575 (2003).
- 106 Ruifrok, A. C. Quantification of immunohistochemical staining by color translation and automated thresholding. *Analytical and quantitative cytology and histology/the International Academy of Cytology [and] American Society of Cytology* **19**, 107 (1997).
- 107 Ranefall, P., Wester, K. & Bengtsson, E. Automatic quantification of immunohistochemically stained cell nuclei using unsupervised image analysis. *Analytical Cellular Pathology* **16**, 29-43 (1998).
- 108 Ranefall, P., Wester, K., Andersson, A., Busch, C. & Bengtsson, E. Automatic quantification of immunohistochemically stained cell nuclei based on standard reference cells. *Analytical Cellular Pathology* **17**, 111-124 (1998).
- 109 Arámbula Cosío, F., Marquez Flores, J., Padilla Castaneda, M., Solano, S. & Tato, P. Automatic analysis of immunocytochemically stained tissue samples. *Medical and Biological Engineering and Computing* **43**, 672-677 (2005).
- 110 Ranefall, P., Egevad, L., Nordin, B. & Bengtsson, E. A new method for segmentation of colour images applied to immunohistochemically stained cell nuclei. *Anal Cell Pathol* **15**, 145-156 (1997).
- 111 Ruifrok, A. C. & Johnston, D. A. Quantification of histochemical staining by color deconvolution. *Analytical and quantitative cytology and histology* **23**, 291-299 (2001).
- 112 Ruifrok, A. C., Katz, R. L. & Johnston, D. A. Comparison of quantification of histochemical staining by hue-saturation-intensity (HSI) transformation and color-deconvolution. *Applied Immunohistochemistry & Molecular Morphology* **11**, 85 (2003).
- 113 Levenson, R. M. Spectral imaging and pathology: seeing more. *Lab Medicine* **35**, 244-251 (2004).
- 114 Farkas, D. L. *et al.* Non-invasive image acquisition and advanced processing in optical bioimaging. *Computerized medical imaging and graphics* **22**, 89-102 (1998).
- 115 Rothmann, C. *et al.* Potential use of spectral image analysis for the quantitative evaluation of estrogen receptors in breast cancer. (2000).
- 116 Mansfield, J. R., Gossage, K. W., Hoyt, C. C. & Levenson, R. M. Autofluorescence removal, multiplexing, and automated analysis methods for in-vivo fluorescence imaging. *Journal of Biomedical Optics* **10**, 041207 (2005).
- 117 Cardoso, J. F. Blind signal separation: statistical principles. *Proceedings of the IEEE* **86**, 2009-2025 (1998).

- 118 Rabinovich, A., Agarwal, S., Laris, C., Price, J. H. & Belongie, S. Unsupervised color decomposition of histologically stained tissue samples. *Advances in Neural Information Processing Systems* (2003).
- 119 Rabinovich, A., Agarwal, S., Price, J. & Belongie, S. Quantitative spectral decomposition for stained tissue analysis. *Proceedings of the University of California* (2005).
- 120 MacKinnon, N., Stange, U., Lane, P., MacAulay, C. & Quatrevalet, M. Spectrally programmable light engine for in vitro or in vivo molecular imaging and spectroscopy. *Applied optics* **44**, 2033-2040 (2005).
- 121 Keshava, N. & Mustard, J. F. Spectral unmixing. *Signal Processing Magazine, IEEE* **19**, 44-57 (2002).
- 122 de Juan, A. & Tauler, R. A. Multivariate curve resolution (MCR) from 2000: progress in concepts and applications. *Critical Reviews in Analytical Chemistry* **36**, 163-176 (2006).
- 123 Jiang, J. H. & Ozaki, Y. Self-modeling curve resolution (SMCR): principles, techniques, and applications. *Applied Spectroscopy Reviews* **37**, 321-345 (2002).
- 124 Vo-Dinh, T. *The Biomedical Photonics Handbook, Second Edition Volume 1: Fundamentals, Devices, and Techniques*. (Taylor & Francis, 2010).
- 125 Charles L Lawson, R. J. H. *Solving Least Squares Problems*. (John Wiley and Sons, Inc., 1974).
- 126 Bro, R. & De Jong, S. A fast non-negativity-constrained least squares algorithm. *Journal of Chemometrics* **11**, 393-401 (1997).
- 127 de Juan, A. & Tauler, R. Multivariate curve resolution (MCR) from 2000: progress in concepts and applications. *Critical reviews in analytical chemistry* **36**, 163-176 (2006).
- 128 Windig, W. & Guilment, J. Interactive self-modeling mixture analysis. *Analytical chemistry* **63**, 1425-1432 (1991).
- 129 Sánchez, F. C. & Massart, D. Application of SIMPLISMA for the assessment of peak purity in liquid chromatography with diode array detection. *Analytica chimica acta* **298**, 331-339 (1994).
- 130 Duponchel, L., Elmi-Rayaleh, W., Ruckebusch, C. & Huvenne, J. Multivariate curve resolution methods in imaging spectroscopy: Influence of extraction methods and instrumental perturbations. *Journal of chemical information and computer sciences* **43**, 2057-2067 (2003).
- 131 MacQueen, J. B. *Some methods for classification and analysis of multivariate observations*. 281-297 (University of California Press, 1967).
- 132 van der Loos, C. M. Chromogens in Multiple Immunohistochemical Staining Used for Visual Assessment and Spectral Imaging: The Colorful Future. *Journal of Histotechnology* **33**, 31-40 (2010).
- 133 Jaumot, J., Gargallo, R., de Juan, A. & Tauler, R. A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemometrics and Intelligent Laboratory Systems* **76**, 101-110 (2005).
- 134 McLachlan, G. J. & Wiley, J. *Discriminant analysis and statistical pattern recognition*. Vol. 5 (Wiley Online Library, 1992).
- 135 Bradley, D. & Roth, G. Adaptive thresholding using the integral image. *Journal of Graphics, GPU, & Game Tools* **12**, 13-21 (2007).

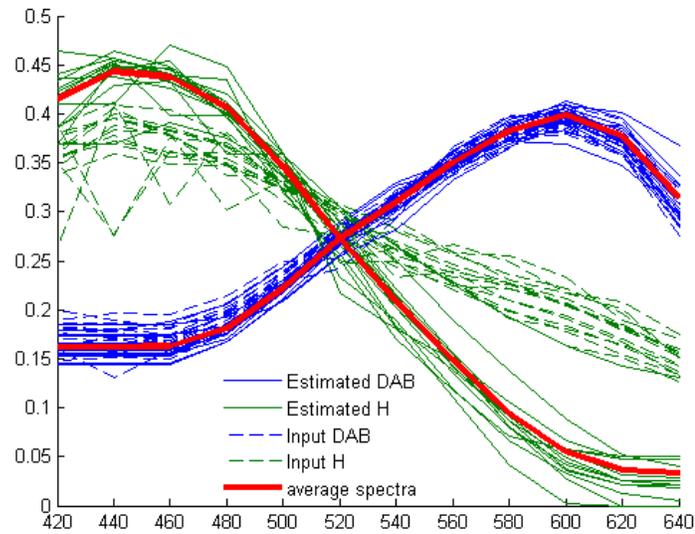
- 136 MacAulay, C. & Palcic, B. An edge relocation segmentation algorithm. *Analytical and quantitative cytology and histology/the International Academy of Cytology [and] American Society of Cytology* **12**, 165 (1990).
- 137 Otsu, N. A threshold selection method from gray-level histograms. *Automatica* **11**, 285-296 (1975).
- 138 Meyer, F. Topographic distance and watershed lines. *Signal Processing* **38**, 113-125 (1994).
- 139 Rijsbergen, C. J. v. *Information Retrieval* (Butterworth, 1979).
- 140 Marcelpoil, R. & Usson, Y. Methods for the study of cellular sociology: Voronoi diagrams and parametrization of the spatial relationships. *Journal of theoretical biology* **154**, 359-369 (1992).
- 141 Guillaud, M., Matthews, J., Harrison, A., MacAulay, C. & Skov, K. A novel image cytometric method for quantitation of immunohistochemical staining of cytoplasmic antigens. *Analytical Cellular Pathology* **14**, 87-100 (1997).
- 142 Barber, C. B., Dobkin, D. P. & Huhdanpaa, H. The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software (TOMS)* **22**, 469-483 (1996).
- 143 Maurer Jr, C. R., Qi, R. & Raghavan, V. A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **25**, 265-270 (2003).
- 144 Hollander, M., and D. A. Wolfe. *Nonparametric Statistical Methods*. (John Wiley & Sons, Inc., 1999).
- 145 Miller, P. J. & Levenson, R. M. in *BiOS 2001 The International Symposium on Biomedical Optics*. 1-7 (International Society for Optics and Photonics).
- 146 Benavides, J. *et al.* Multispectral digital colposcopy for in vivo detection of cervical cancer. *Optics Express* **11**, 1223-1236 (2003).
- 147 DaCosta, R. S., Wilson, B. C. & Marcon, N. E. Fluorescence and spectral imaging. *ScientificWorldJournal* **7**, 2046-2071 (2007).
- 148 Kato, M., Kaise, M., Yonezawa, J., Yoshida, Y. & Tajiri, H. Autofluorescence endoscopy versus conventional white light endoscopy for the detection of superficial gastric neoplasia: a prospective comparative study. *Endoscopy* **39**, 937-941 (2007).
- 149 Pavlova, I., Williams, M., El-Naggar, A., Richards-Kortum, R. & Gillenwater, A. Understanding the biological basis of autofluorescence imaging for oral cancer detection: high-resolution fluorescence microscopy in viable tissue. *Clinical Cancer Research* **14**, 2396-2404 (2008).
- 150 Wagnieres, G. A., Star, W. M. & Wilson, B. C. In vivo fluorescence spectroscopy and imaging for oncological applications. *Photochemistry and photobiology* **68**, 603-632 (1998).
- 151 R Mansfield, J. Distinguished photons: a review of in vivo spectral fluorescence imaging in small animals. *Current Pharmaceutical Biotechnology* **11**, 628-638 (2010).
- 152 Zhou, L. & El-Deiry, W. S. Multispectral fluorescence imaging. *Journal of Nuclear Medicine* **50**, 1563-1566 (2009).
- 153 Zimmermann, T. Spectral imaging and linear unmixing in light microscopy. *Microscopy techniques*, 1308-1310 (2005).
- 154 Roblyer, D. *et al.* Multispectral optical imaging device for in vivo detection of oral neoplasia. *Journal of Biomedical Optics* **13**, 024019 (2008).

- 155 Montcuquet, A. S., Herve, L., Navarro, F., Dinten, J. M. & Mars, J. I. Nonnegative matrix factorization: a blind spectra separation method for in vivo fluorescent optical imaging. *Journal of Biomedical Optics* **15**, 056009 (2010).
- 156 Thekkek, N. & Richards-Kortum, R. Optical imaging for cervical cancer detection: solutions for a continuing global problem. *Nature Reviews Cancer* **8**, 725-731 (2008).
- 157 Grimbergen, M., Aalders, M. & Leeuwen, T. Optical imaging and diagnosis in bladder cancer. *Imaging in Oncological Urology*, 407-419 (2009).
- 158 Goetz, M. & Wang, T. D. Molecular imaging in gastrointestinal endoscopy. *Gastroenterology* **138**, 828-833. e821 (2010).
- 159 Wang, K. K. *et al.* Endoscopic Evaluation and Advanced Imaging of Barrett's Esophagus. *Gastrointestinal Endoscopy Clinics of North America* **21**, 39-51 (2011).
- 160 Fawzy, Y. & Zeng, H. Intrinsic fluorescence spectroscopy for endoscopic detection and localization of the endobronchial cancerous lesions. *Journal of biomedical optics* **13**, 064022 (2008).
- 161 Qu, J. Y., Chang, H. & Xiong, S. Fluorescence spectral imaging for characterization of tissue based on multivariate statistical analysis. *JOSA A* **19**, 1823-1831 (2002).
- 162 Jolliffe, I. T. *Principal Component Analysis*. 2 edn, (Springer, 2002).
- 163 Guilbault, G. G. *Practical Fluorescence, Second Edition*. (Taylor & Francis, 1990).

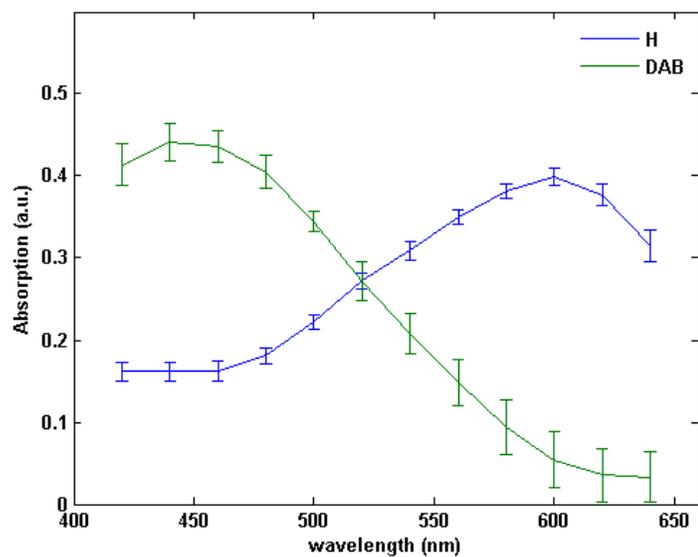
Appendices

Appendix A. Additional Spectral Unmixing Results

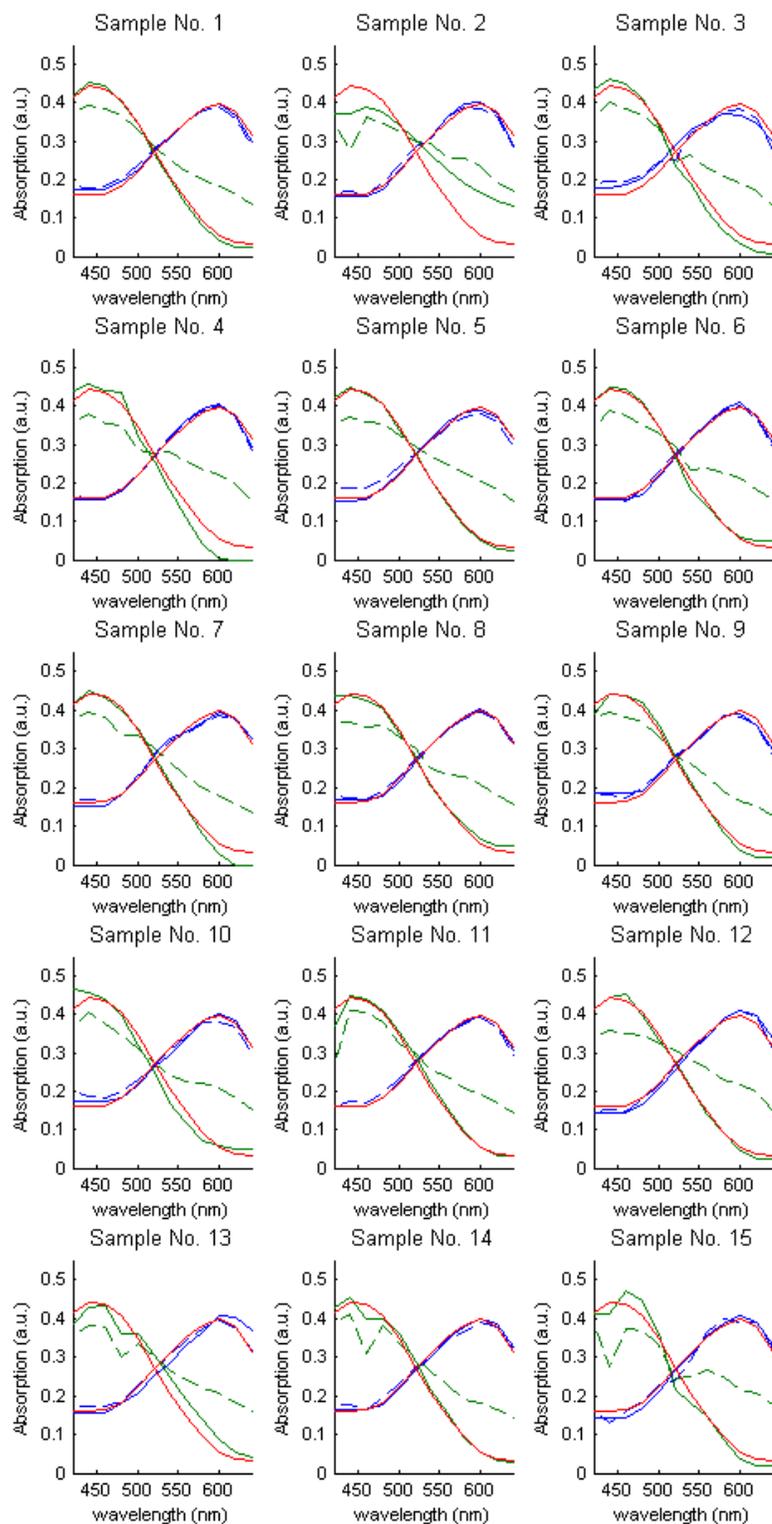
A.1. Ki67 Set



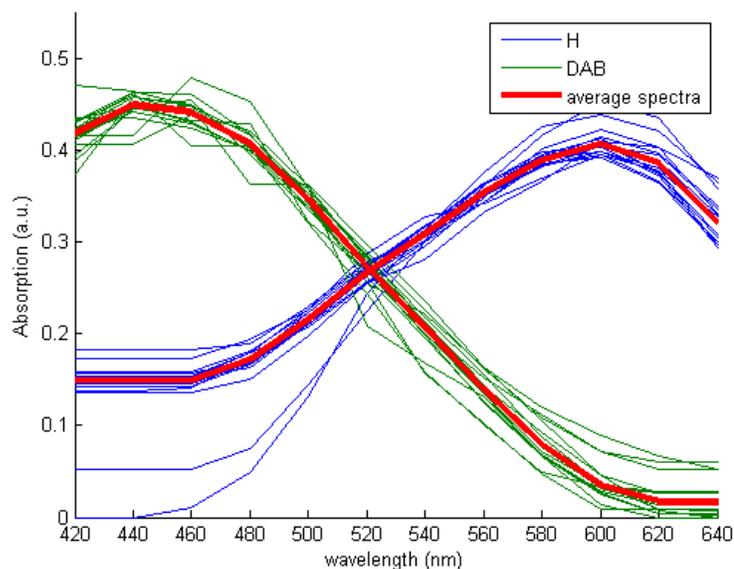
Appendix Figure 1 Unmixed spectra of H and DAB for all 15 Ki67 samples. The averages of H and DAB spectra obtained from 15 samples are also shown. Linear spectral unmixing was performed using initial estimates of spectra obtained using “multi-region SIMPLISMA” procedure followed by the “subtraction algorithm”.



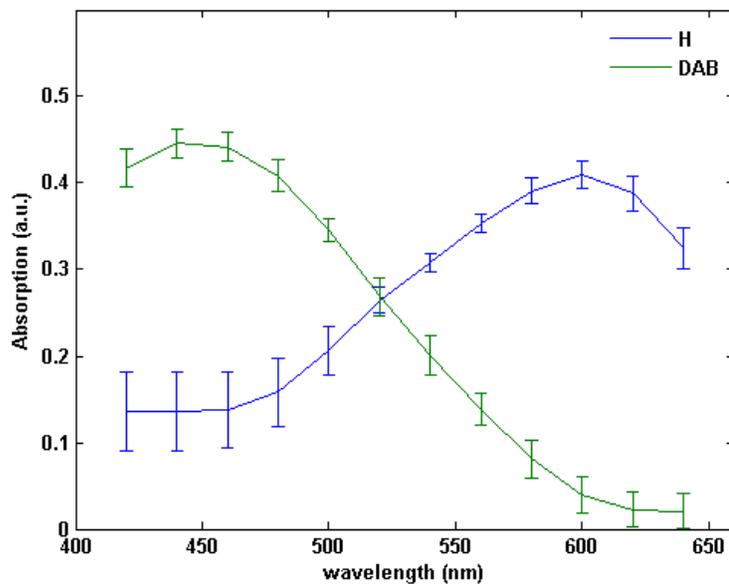
Appendix Figure 2 Average of H and DAB spectra obtained from 15 Ki67 samples. For each label, the error bar at each wavelength represents the standard deviation of that label’s absorption in that wavelength. Linear spectral unmixing was performed using initial estimates of spectra obtained using “multi-region SIMPLISMA” procedure followed by the “subtraction algorithm”.



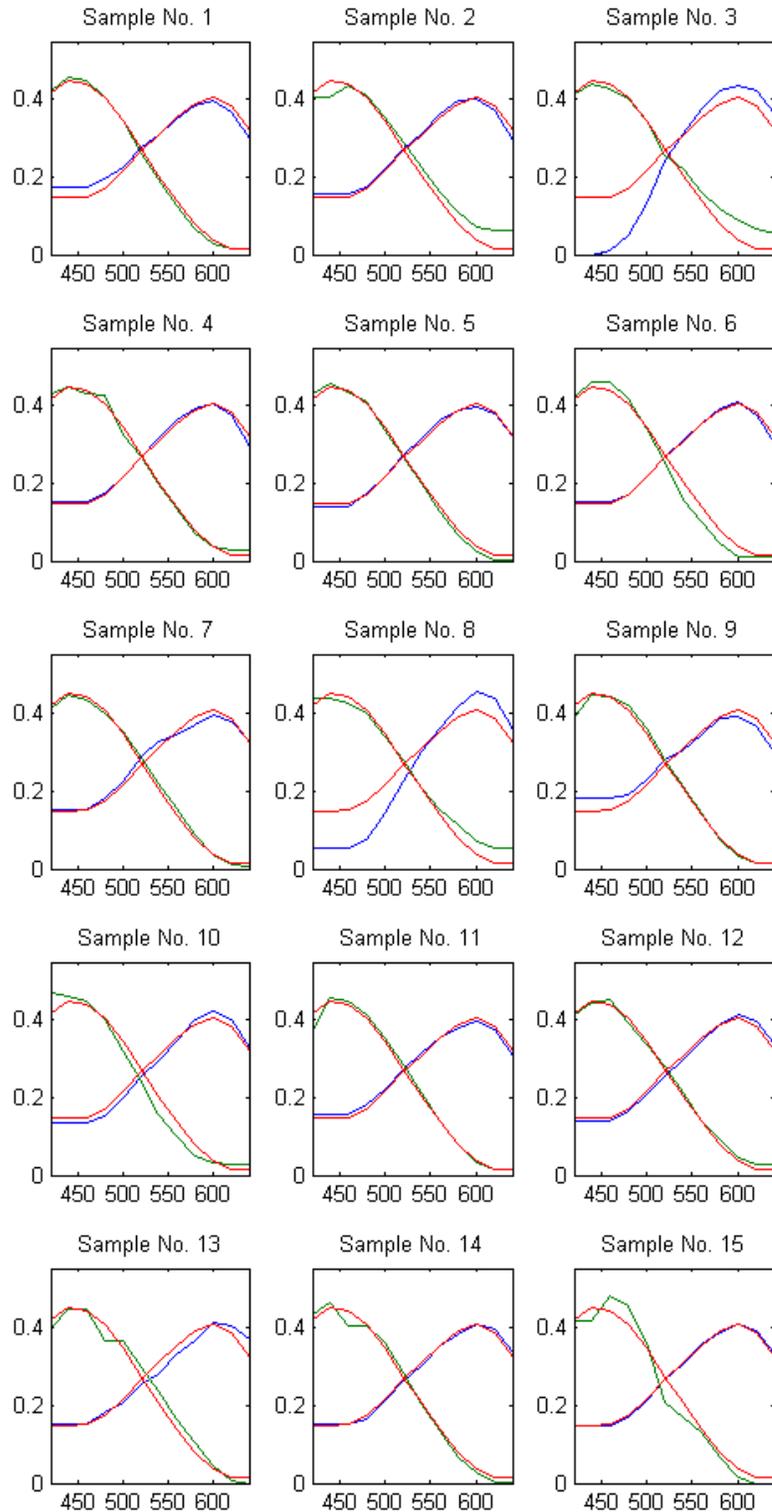
Appendix Figure 3 Spectra obtained from each Ki67 sample. Linear spectral unmixing was performed using initial estimates of spectra obtained using “multi-region SIMPLISMA” procedure followed by the “subtraction algorithm”. Dashed lines represent initial estimates of spectra and solid lines represent unmixed spectra. The spectra shown in red are the average across 15 Ki67 samples.



Appendix Figure 4 Unmixed spectra of H and DAB for all 15 Ki67 samples. The averages of H and DAB spectra obtained from 15 samples are also shown. Linear spectral unmixing was performed using initial estimates of concentration maps.

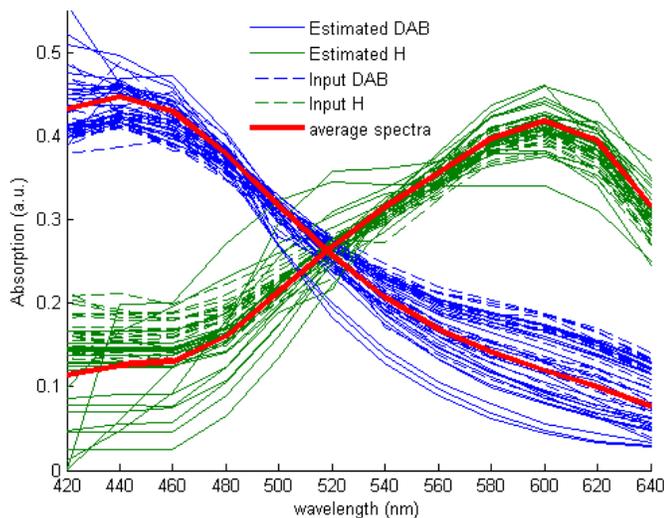


Appendix Figure 5 Average of H and DAB spectra obtained from 15 Ki67 samples. For each label, the error bar at each wavelength represents the standard deviation of that label's absorption in that wavelength. Linear spectral unmixing was performed using initial estimates of concentration maps.

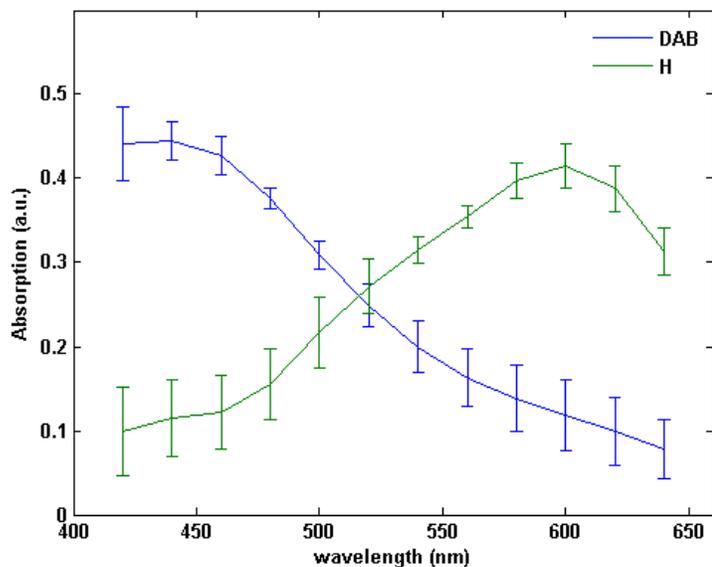


Appendix Figure 6 Spectra obtained from each Ki67 sample. Linear spectral unmixing was performed using initial estimates of concentration maps. Dashed lines represent initial estimates of spectra and solid lines represent unmixed spectra. The spectra shown in red are the average across 15 Ki67 samples.

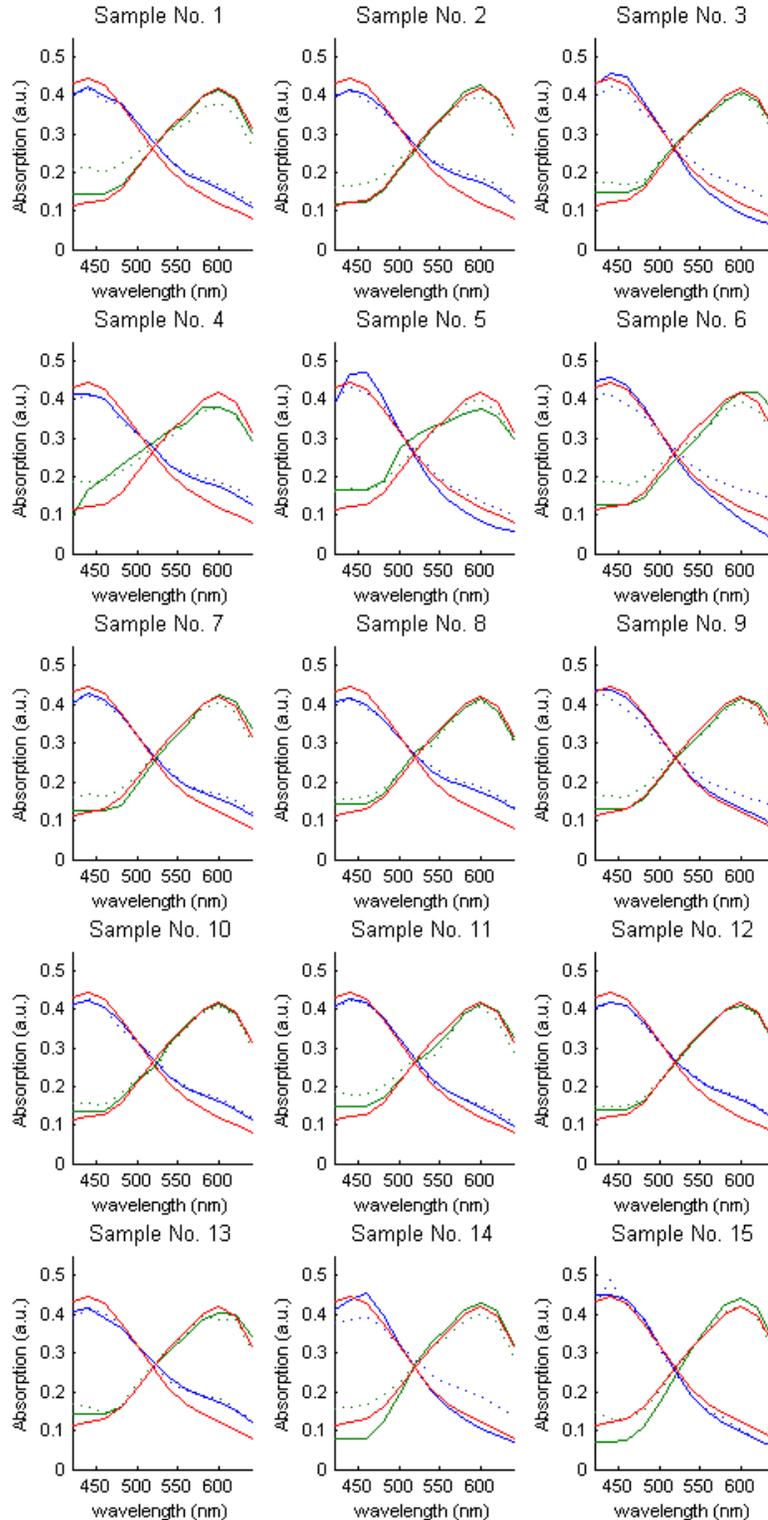
A.2. P16 Set



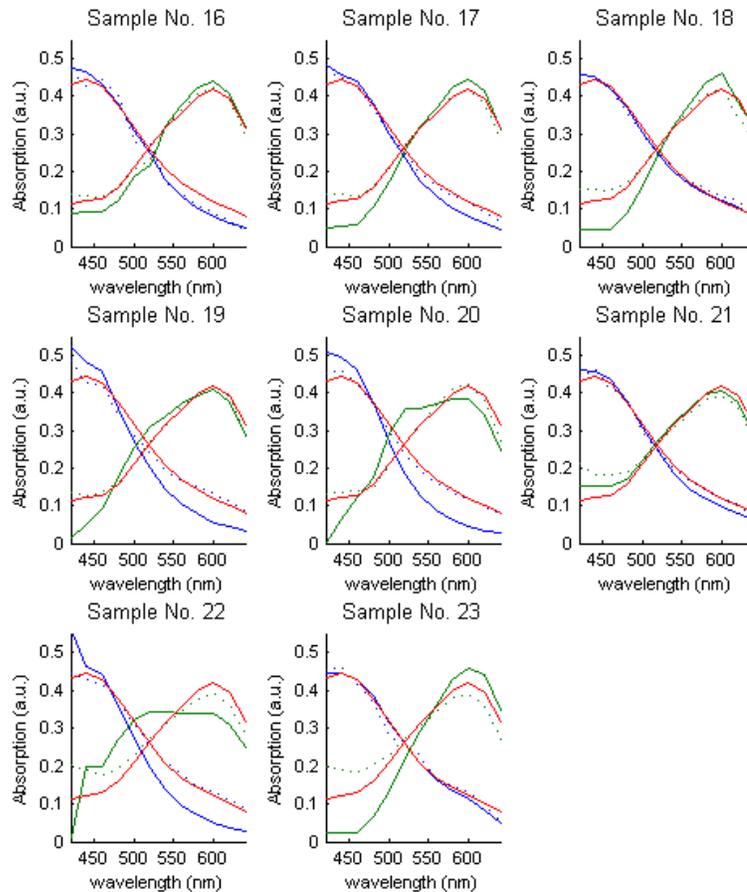
Appendix Figure 7 Unmixed spectra of H and DAB for all 23 P16 samples. The averages of H and DAB spectra obtained from 23 samples are also shown. Linear spectral unmixing was performed using initial estimates of spectra obtained using “multi-region SIMPLISMA” procedure.



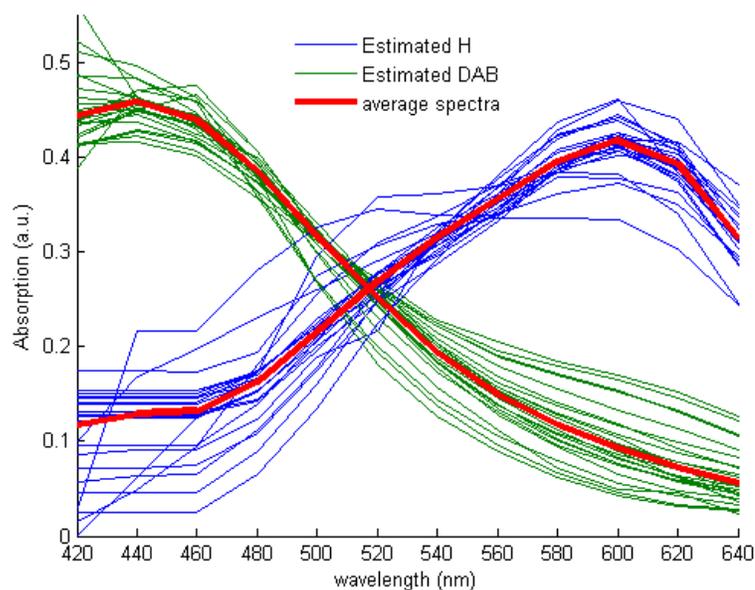
Appendix Figure 8 Average of H and DAB spectra obtained from 23 P16 samples. For each label, the error bar at each wavelength represents the standard deviation of that label’s absorption in that wavelength. Linear spectral unmixing was performed using initial estimates of spectra obtained using “multi-region SIMPLISMA” procedure.



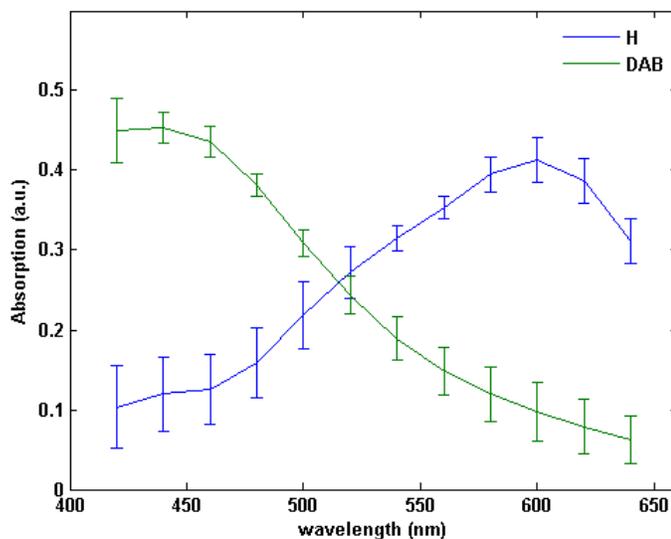
Appendix Figure 9 Spectra obtained from each P16 sample. Linear spectral unmixing was performed using initial estimates of spectra obtained by the “multi-region SIMPLISMA” procedure. Dashed lines represent initial estimates of spectra and solid lines represent unmixed spectra. The spectra shown in red are the average across 23 P16 samples. To be continued in Appendix Figure 10.



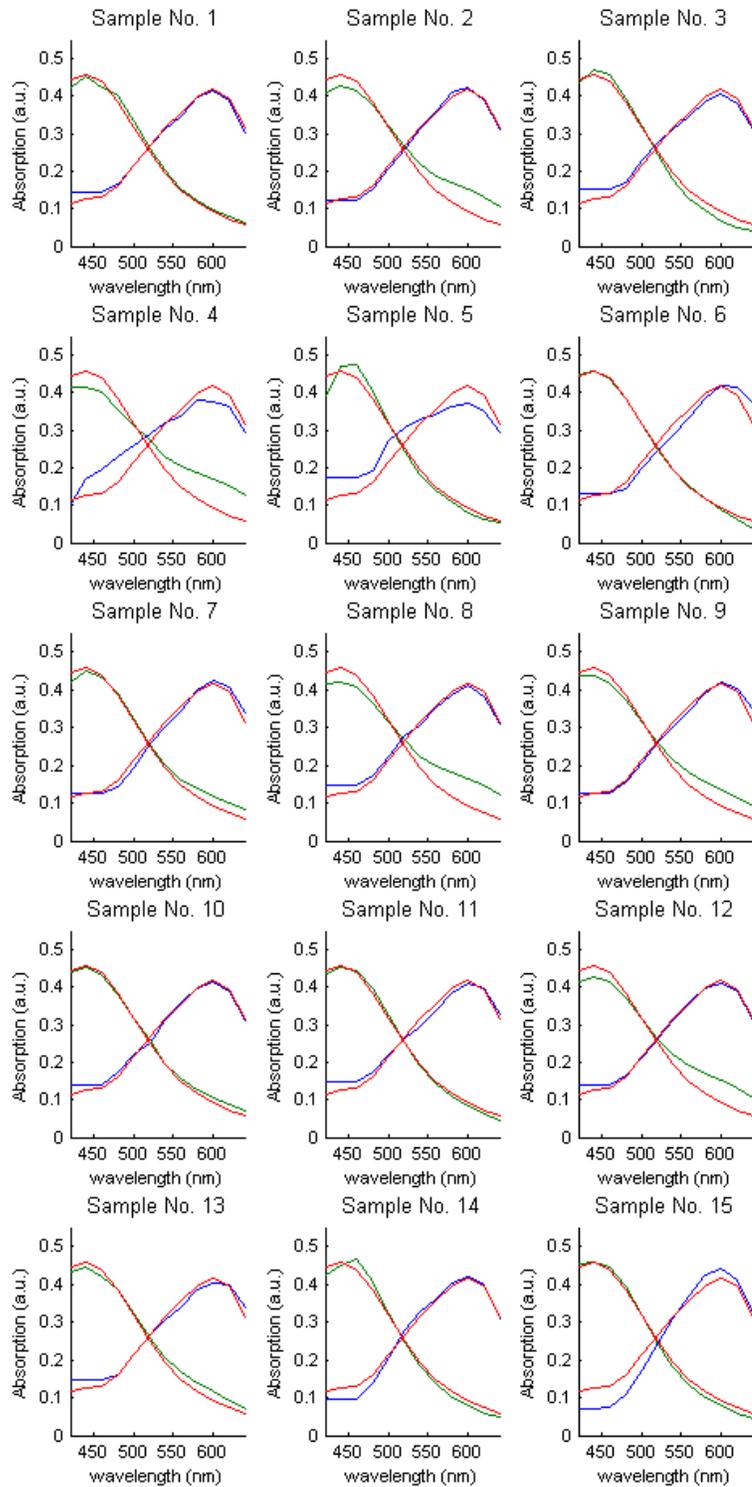
Appendix Figure 10 Continued from Appendix Figure 9.



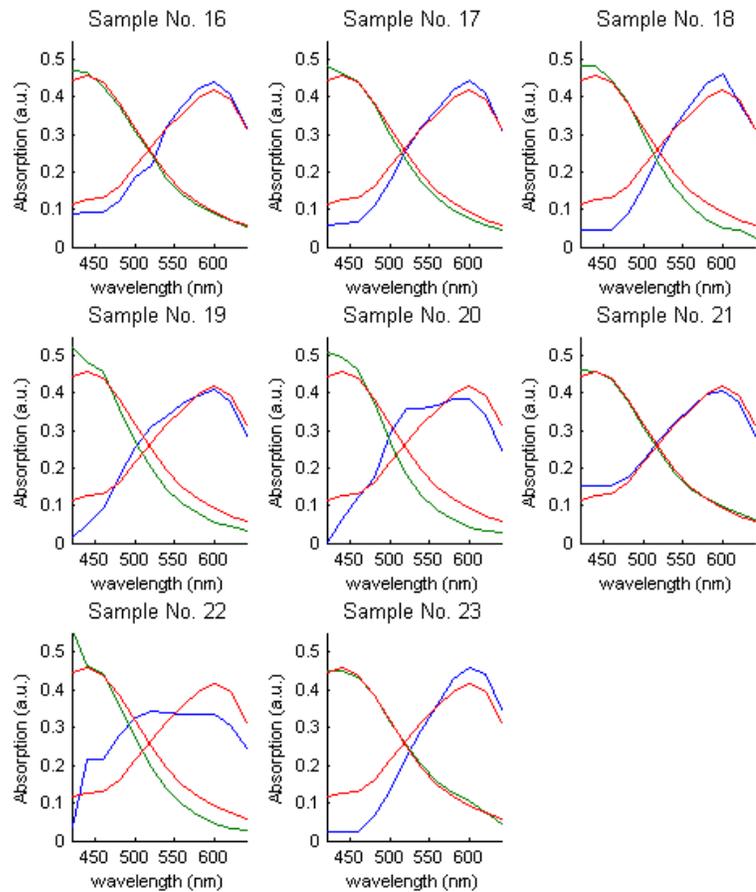
Appendix Figure 11 Unmixed spectra of H and DAB for all 23 P16 samples. The averages of H and DAB spectra obtained from 23 samples are also shown. Linear spectral unmixing was performed using initial estimates of concentration maps.



Appendix Figure 12 Average of H and DAB spectra obtained from 23 P16 samples. For each label, the error bar at each wavelength represents the standard deviation of that label's absorption in that wavelength. Linear spectral unmixing was performed using initial estimates of concentration maps.

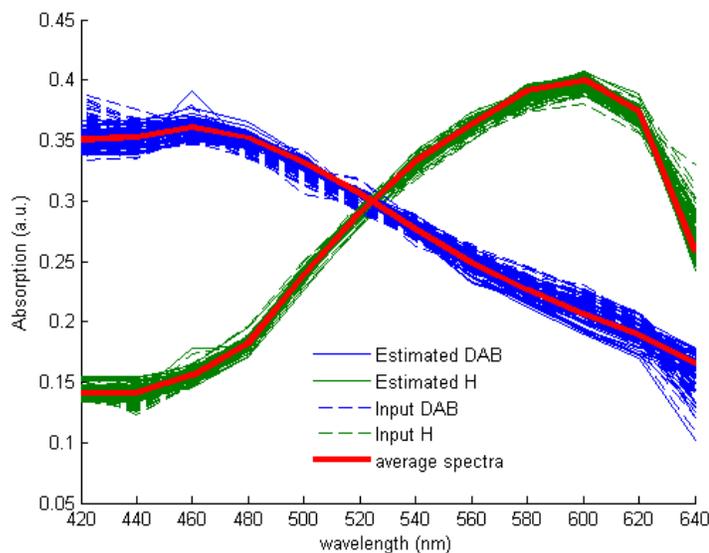


Appendix Figure 13 Spectra obtained from each P16 sample. Linear spectral unmixing was performed using initial estimates of concentration maps. Dashed lines represent initial estimates of spectra and solid lines represent unmixed spectra. The spectra shown in red are the average across 23 P16 samples. To be continued in Appendix Figure 14.

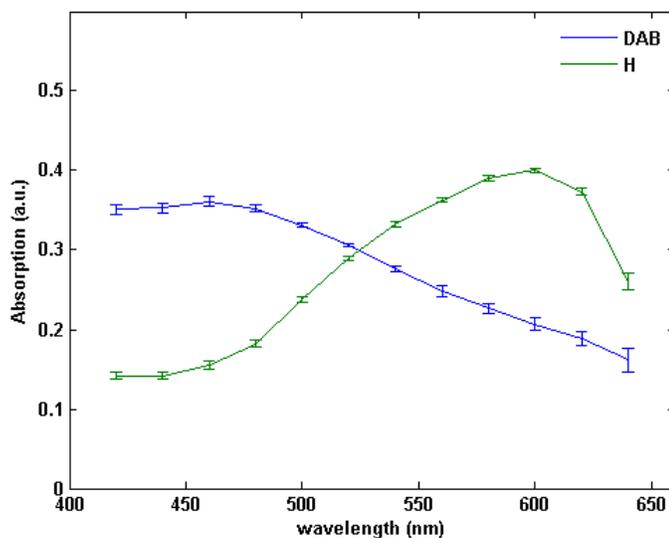


Appendix Figure 14 Continued from Appendix Figure 13.

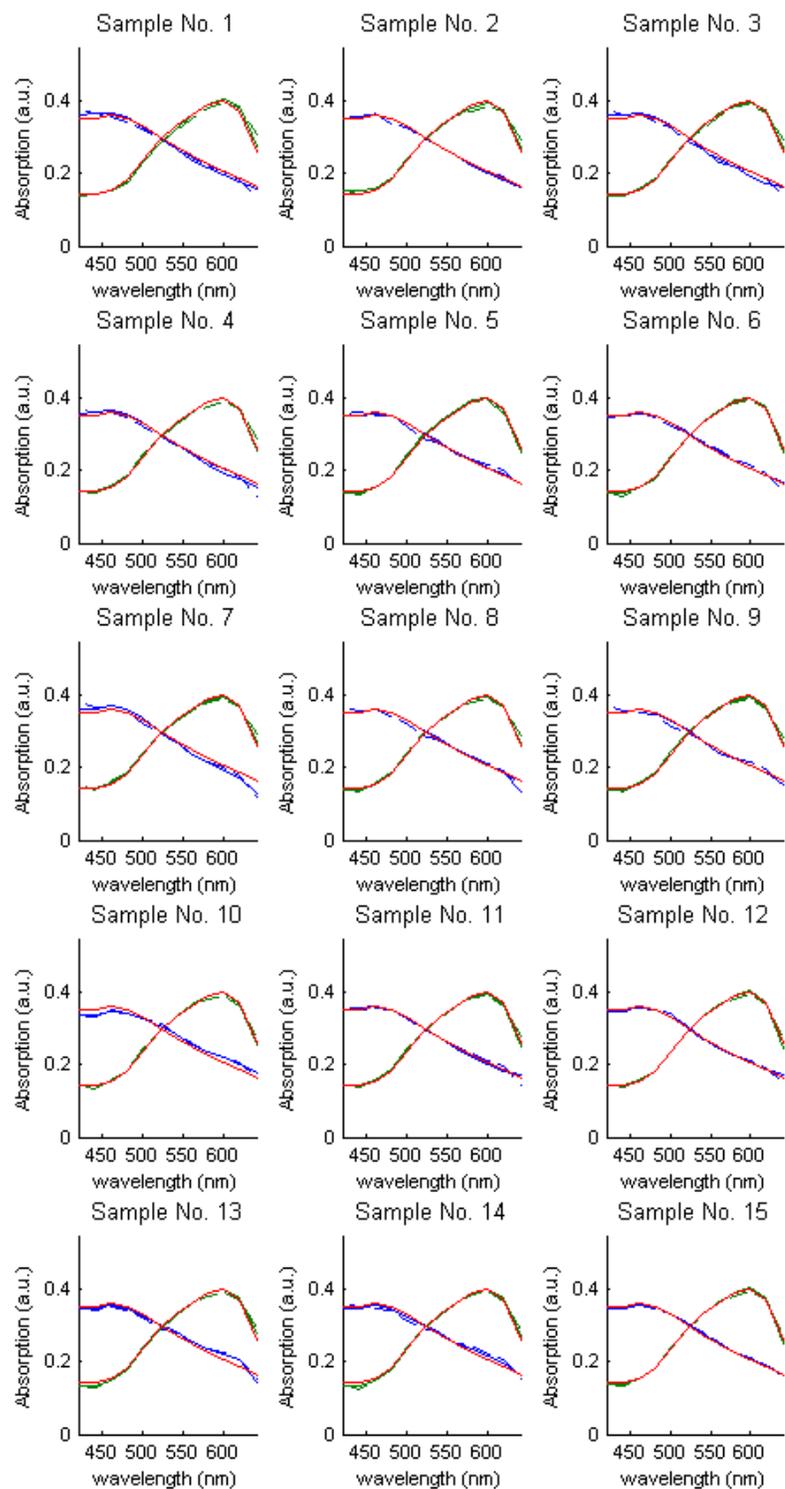
A.3. CD8 Set



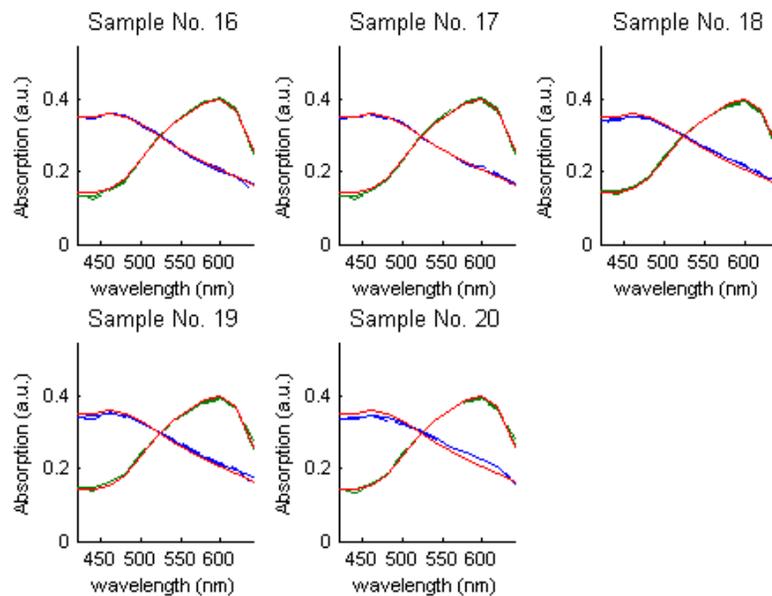
Appendix Figure 15 Unmixed spectra of H and DAB for all 20 CD8 samples. The averages of H and DAB spectra obtained from 20 samples are also shown. Linear spectral unmixing was performed using initial estimates of spectra obtained using “multi-region SIMPLISMA” procedure.



Appendix Figure 16 Average of H and DAB spectra obtained from 20 CD8 samples. For each label, the error bar at each wavelength represents the standard deviation of that label’s absorption in that wavelength. Linear spectral unmixing was performed using initial estimates of spectra obtained using “multi-region SIMPLISMA” procedure.

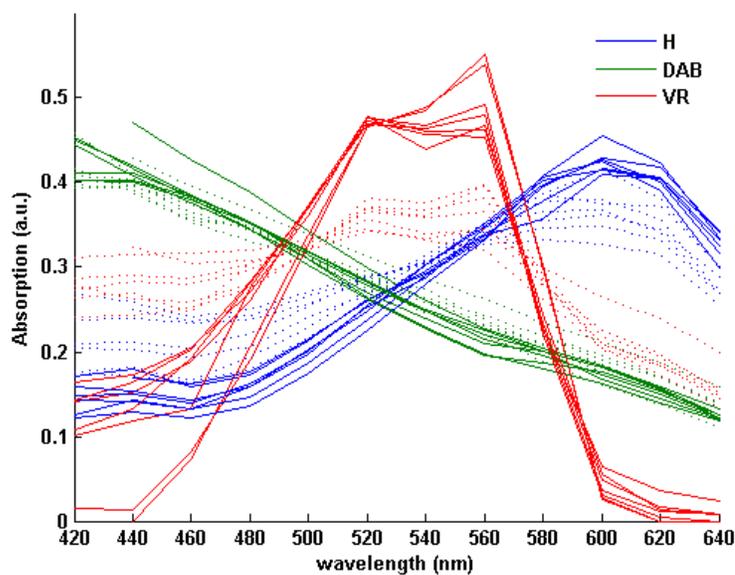


Appendix Figure 17 Spectra obtained from each CD8 sample. Linear spectral unmixing was performed using initial estimates of spectra obtained by the “multi-region SIMPLISMA” procedure. Dashed lines represent initial estimates of spectra and solid lines represent unmixed spectra. The spectra shown in red are the average across 20 CD8 samples. To be continued in Appendix Figure 18.

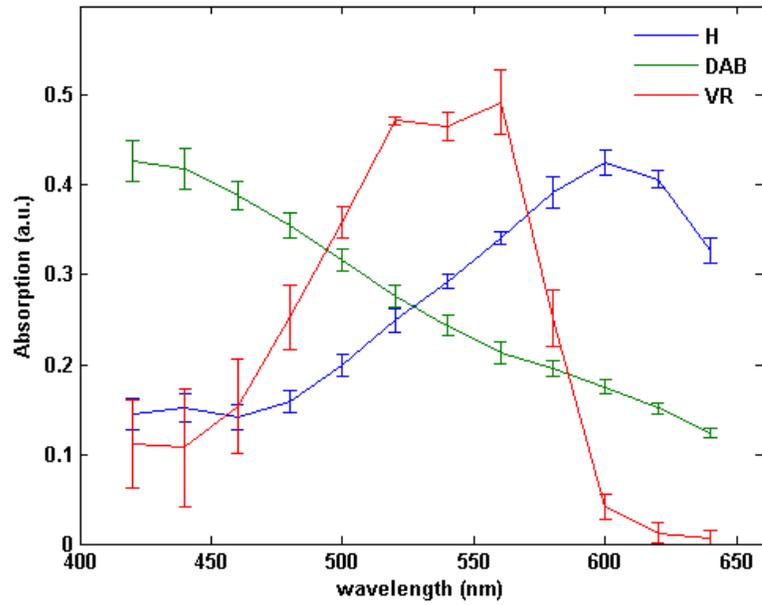


Appendix Figure 18 Continued from Appendix Figure 17.

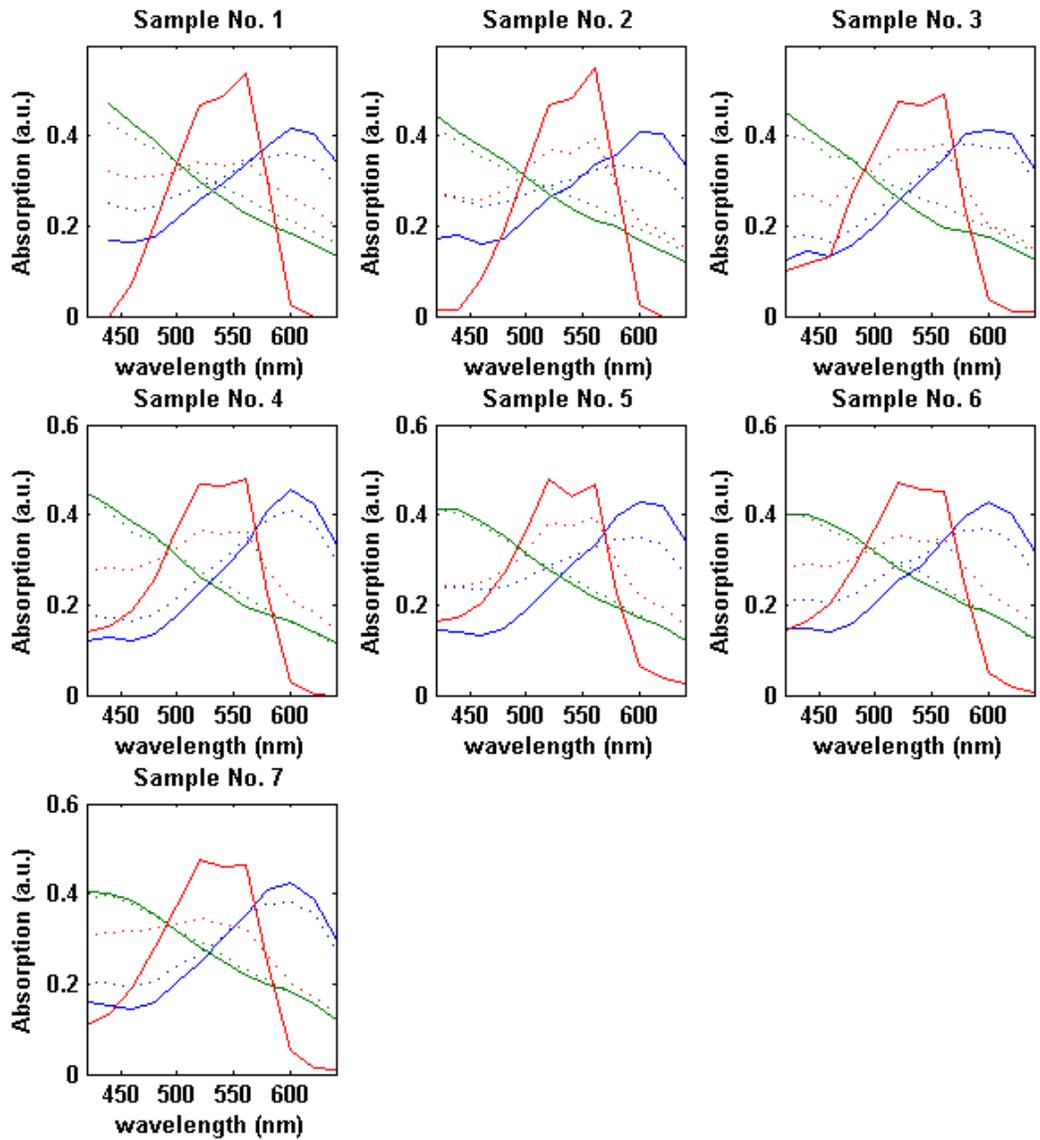
A.4. Ki67 & P16 Set



Appendix Figure 19 Unmixed spectra of H, DAB, and VR for all 7 “Ki67 & P16” samples. Linear spectral unmixing was performed using initial estimates of spectra. Initial estimate of VR spectrum was obtained using the two-dimensional version of the “subtraction algorithm”.



Appendix Figure 20 Average of H, DAB, and VR spectra obtained from 7 “Ki67 & P16” samples. For each label, the error bar at each wavelength represents the standard deviation of that label’s absorption in that wavelength. Linear spectral unmixing was performed using initial estimates of spectra.



Appendix Figure 21 Spectra obtained from each “Ki67 & P16” sample. Linear spectral unmixing was performed using initial estimates of spectra. Dashed lines represent initial estimates of spectra and solid lines represent unmixed spectra.

Appendix B. Additional Nuclei Segmentation Results

B.1. Ki67 Set

Appendix Table 1 Ki67 segmentation results

		Multi spectral Unmixing	RGB unmixing	Multi spectral LDA	RGB LDA	Red Channel	Manual	No. Of Nuclei
Image 1	Correct	89.68%	92.04%	85.84%	88.50%	85.25%	87.91%	339
	Missed	1.77%	9.14%	7.08%	11.50%	11.50%	0.29%	
	Under	3.83%	2.65%	0.59%	1.18%	1.18%	0.00%	
	Over	6.49%	5.31%	13.57%	10.32%	13.57%	12.09%	
	F-Score	93.68%	91.50%	88.99%	88.50%	86.66%	93.42%	
Image 2	Correct	62.78%	71.11%	52.78%	62.22%	65.56%	82.22%	180
	Missed	3.89%	4.44%	11.67%	10.56%	7.78%	0.00%	
	Under	8.89%	8.33%	1.11%	2.78%	1.67%	0.00%	
	Over	28.33%	20.56%	46.11%	35.00%	32.78%	17.78%	
	F-Score	75.33%	81.01%	64.19%	72.03%	75.64%	90.24%	
Image 3	Correct	83.82%	82.35%	84.56%	78.68%	77.94%	80.88%	136
	Missed	13.24%	69.12%	29.41%	11.76%	11.76%	0.00%	
	Under	7.35%	11.76%	2.21%	4.41%	4.41%	0.00%	
	Over	8.82%	5.88%	13.24%	16.91%	17.65%	19.12%	
	F-Score	85.07%	65.50%	79.04%	82.63%	82.17%	89.43%	
Image 4	Correct	79.13%	74.27%	74.76%	67.48%	70.87%	80.58%	206
	Missed	2.43%	5.34%	11.65%	7.28%	16.99%	0.00%	
	Under	2.43%	0.97%	1.46%	0.97%	0.49%	0.00%	
	Over	18.45%	24.76%	23.79%	31.55%	28.64%	19.42%	
	F-Score	87.17%	82.70%	80.21%	77.22%	75.45%	89.25%	
Image 5	Correct	70.94%	71.92%	56.65%	67.49%	66.01%	84.73%	203
	Missed	9.36%	8.87%	6.90%	7.88%	7.88%	0.49%	
	Under	10.34%	8.87%	2.96%	5.42%	6.90%	0.00%	
	Over	18.72%	19.21%	40.39%	27.09%	27.09%	15.27%	
	F-Score	78.69%	79.56%	69.28%	76.97%	75.92%	91.49%	
Image 6	Correct	81.19%	81.19%	77.72%	80.20%	78.22%	83.66%	202
	Missed	4.46%	8.42%	12.38%	7.92%	7.92%	0.00%	
	Under	0.99%	2.48%	0.00%	0.50%	0.50%	0.00%	
	Over	17.82%	16.34%	22.28%	19.31%	21.29%	16.34%	
	F-Score	87.47%	85.64%	81.77%	85.26%	84.04%	91.11%	
Image 7	Correct	82.79%	82.79%	76.74%	77.67%	77.21%	82.79%	215
	Missed	10.70%	12.09%	11.63%	10.23%	8.37%	0.00%	

		Multi spectral Unmixing	RGB unmixing	Multi spectral LDA	RGB LDA	Red Channel	Manual	No. Of Nuclei
Image 8	Under	7.91%	7.91%	5.12%	5.12%	4.65%	0.00%	202
	Over	9.30%	9.30%	18.14%	17.21%	18.14%	17.21%	
	F-Score	85.58%	84.96%	81.48%	82.67%	83.21%	90.59%	
	Correct	81.19%	72.77%	71.78%	66.34%	69.31%	83.17%	
	Missed	5.45%	3.47%	13.86%	6.93%	5.94%	0.50%	
Image 9	Under	1.98%	4.46%	2.97%	3.96%	2.97%	0.00%	237
	Over	16.83%	22.77%	25.25%	29.70%	27.72%	16.83%	
	F-Score	87.00%	82.58%	77.33%	76.57%	79.10%	90.57%	
	Correct	85.23%	81.86%	79.32%	78.90%	76.79%	88.61%	
	Missed	5.06%	7.17%	9.70%	8.44%	5.91%	0.00%	
Image 10	Under	5.06%	7.17%	2.53%	5.06%	1.69%	0.00%	140
	Over	9.70%	10.97%	18.14%	16.03%	21.52%	11.39%	
	F-Score	89.58%	86.61%	83.93%	84.23%	84.06%	93.96%	
	Correct	79.29%	83.57%	76.43%	80.00%	77.86%	85.71%	
	Missed	5.71%	5.00%	10.00%	8.57%	7.14%	0.00%	
Image 11	Under	5.00%	4.29%	4.29%	5.00%	3.57%	0.00%	189
	Over	15.71%	12.14%	19.29%	15.00%	18.57%	14.29%	
	F-Score	85.71%	88.64%	81.99%	84.85%	84.17%	92.31%	
	Correct	84.66%	82.01%	77.25%	74.60%	75.13%	87.30%	
	Missed	3.17%	2.12%	14.81%	7.94%	11.64%	0.00%	
Image 12	Under	4.76%	5.29%	3.70%	0.53%	1.06%	0.00%	160
	Over	10.58%	12.70%	19.05%	24.87%	23.81%	12.70%	
	F-Score	90.14%	89.08%	80.44%	81.74%	80.45%	93.22%	
	Correct	60.00%	63.13%	60.00%	58.13%	61.88%	80.00%	
	Missed	1.25%	3.13%	10.00%	3.13%	5.63%	0.00%	
Image 13	Under	6.88%	6.25%	2.50%	3.13%	1.88%	0.00%	130
	Over	33.13%	30.63%	37.50%	38.75%	36.25%	20.00%	
	F-Score	74.42%	75.94%	70.59%	72.09%	73.88%	88.89%	
	Correct	90.00%	79.23%	79.23%	73.85%	80.77%	90.77%	
	Missed	8.46%	7.69%	24.62%	12.31%	20.77%	0.00%	
Image 14	Under	3.85%	1.54%	1.54%	1.54%	0.77%	0.00%	236
	Over	6.15%	19.23%	19.23%	24.62%	18.46%	9.23%	
	F-Score	90.70%	84.77%	77.74%	79.34%	80.15%	95.16%	
	Correct	88.98%	86.44%	81.78%	80.08%	82.20%	88.14%	
	Missed	8.05%	12.29%	8.90%	8.05%	6.36%	0.00%	
	Under	5.93%	5.93%	2.97%	1.69%	2.12%	0.00%	
	Over	5.08%	7.63%	15.25%	18.22%	15.68%	11.86%	

		Multi spectral Unmixing	RGB unmixing	Multi spectral LDA	RGB LDA	Red Channel	Manual	No. Of Nuclei
	F-Score	90.32%	86.99%	85.78%	85.14%	87.19%	93.69%	
Image 15	Correct	81.16%	81.64%	73.43%	68.60%	69.08%	87.92%	207
	Missed	1.45%	5.31%	11.11%	4.35%	10.14%	0.48%	
	Under	2.42%	1.93%	1.93%	5.31%	0.97%	0.00%	
	Over	16.43%	16.43%	24.64%	26.09%	29.95%	12.08%	
	F-Score	88.89%	87.34%	79.58%	79.33%	77.09%	93.33%	
Total	Correct	80.75%	79.91%	74.51%	74.38%	74.82%	85.18%	2982
	Missed	5.33%	9.89%	12.01%	8.48%	9.52%	0.13%	
	Under	5.06%	5.16%	2.31%	3.02%	2.25%	0.00%	
	Over	14.19%	14.92%	23.17%	22.60%	22.94%	14.82%	
	F-Score	86.79%	84.20%	79.90%	81.35%	81.17%	91.93%	

B.2. P16 Set

Appendix Table 2 P16 segmentation results

		Multi spectral Unmixing	RGB unmixing	Multi spectral LDA	RGB LDA	Red Channel	Manual	No. Of Nuclei
Image 1	Correct	82.17%	82.73%	67.83%	74.25%	82.49%	59.27%	1262
	Missed	9.43%	12.28%	7.77%	13.00%	23.45%	1.03%	
	Under	11.65%	7.13%	6.26%	11.17%	4.28%	0.16%	
	Over	6.18%	10.14%	25.91%	14.58%	13.23%	40.57%	
	F-Score	85.77%	84.84%	77.26%	79.31%	80.11%	73.95%	
Image 2	Correct	53.15%	50.35%	26.11%	44.06%	61.54%	27.27%	429
	Missed	0.47%	0.93%	0.23%	0.47%	2.80%	0.70%	
	Under	17.25%	20.05%	10.02%	12.82%	6.06%	0.00%	
	Over	29.60%	29.60%	63.87%	43.12%	32.40%	72.73%	
	F-Score	69.20%	66.56%	41.33%	60.97%	74.89%	42.62%	
Image 3	Correct	51.88%	52.78%	32.78%	46.02%	50.23%	24.06%	665
	Missed	1.50%	2.71%	0.60%	0.15%	4.21%	0.60%	
	Under	14.74%	15.64%	12.78%	17.74%	11.13%	0.00%	
	Over	33.38%	31.58%	54.44%	36.24%	38.65%	75.94%	
	F-Score	67.65%	67.89%	49.15%	62.96%	65.04%	38.60%	
Image 4	Correct	69.72%	70.42%	69.01%	73.24%	78.17%	67.61%	142
	Missed	66.90%	73.94%	54.93%	64.79%	79.58%	23.94%	
	Under	15.49%	16.20%	2.82%	8.45%	2.82%	0.00%	
	Over	14.79%	13.38%	28.17%	18.31%	19.01%	32.39%	
	F-Score	58.93%	57.64%	61.64%	61.54%	60.66%	70.59%	
Image 5	Correct	67.39%	48.14%	41.46%	51.71%	22.83%	51.55%	644
	Missed	48.60%	41.61%	25.16%	38.51%	40.68%	6.99%	

		Multi spectral Unmixing	RGB unmixing	Multi spectral LDA	RGB LDA	Red Channel	Manual	No. Of Nuclei
	Under	14.44%	6.68%	6.06%	7.14%	70.96%	0.00%	
	Over	18.17%	45.19%	52.48%	41.15%	6.21%	48.45%	
	F-Score	62.40%	50.74%	49.77%	54.37%	27.92%	65.03%	
Image 6	Correct	72.50%	57.37%	56.58%	56.19%	56.78%	50.88%	509
	Missed	3.54%	4.91%	3.93%	6.09%	15.32%	3.73%	
	Under	7.47%	5.89%	4.52%	11.98%	4.91%	0.00%	
	Over	20.04%	36.74%	38.90%	31.83%	38.31%	49.12%	
	F-Score	82.37%	70.70%	70.50%	69.25%	65.98%	65.82%	
Image 7	Correct	72.20%	72.36%	60.81%	69.92%	69.27%	46.67%	615
	Missed	8.62%	12.03%	9.76%	11.22%	23.74%	4.55%	
	Under	16.10%	11.71%	5.85%	7.64%	3.58%	0.00%	
	Over	11.71%	15.93%	33.33%	22.44%	27.15%	53.33%	
	F-Score	79.86%	78.48%	71.31%	77.20%	71.78%	61.72%	
Image 8	Correct	75.99%	75.24%	56.14%	69.19%	66.92%	34.40%	529
	Missed	6.05%	6.24%	4.16%	6.81%	19.47%	0.19%	
	Under	17.01%	15.88%	7.37%	9.26%	6.05%	0.00%	
	Over	6.99%	8.88%	36.48%	21.55%	27.03%	65.60%	
	F-Score	83.49%	82.92%	70.05%	78.63%	71.81%	51.12%	
Image 9	Correct	77.65%	77.52%	61.21%	72.38%	84.41%	72.45%	1539
	Missed	1.23%	1.04%	0.78%	0.84%	3.96%	0.13%	
	Under	10.72%	11.44%	10.01%	11.96%	5.91%	0.00%	
	Over	11.63%	11.05%	28.78%	15.66%	9.68%	27.55%	
	F-Score	86.81%	86.83%	75.57%	83.57%	89.62%	83.96%	
Image 10	Correct	90.54%	86.44%	79.18%	81.39%	80.13%	79.50%	317
	Missed	13.56%	23.66%	10.73%	15.46%	31.86%	0.00%	
	Under	6.94%	7.26%	4.10%	4.42%	3.79%	0.00%	
	Over	2.52%	6.31%	16.72%	14.20%	16.09%	20.50%	
	F-Score	88.72%	82.28%	83.39%	82.69%	75.60%	88.58%	
Image 11	Correct	78.72%	75.00%	43.09%	60.64%	45.74%	67.55%	188
	Missed	6.91%	16.49%	6.38%	7.45%	19.68%	1.60%	
	Under	15.96%	18.62%	3.72%	9.04%	4.79%	0.00%	
	Over	5.32%	6.38%	53.19%	30.32%	49.47%	32.45%	
	F-Score	84.81%	78.33%	57.65%	72.15%	55.31%	79.87%	
Image 12	Correct	61.66%	63.26%	32.91%	53.99%	68.69%	43.77%	313
	Missed	0.32%	0.32%	0.64%	0.32%	1.92%	0.64%	
	Under	10.54%	10.54%	7.35%	8.31%	4.47%	0.00%	
	Over	27.80%	26.20%	59.74%	37.70%	26.84%	56.23%	
	F-Score	76.13%	77.34%	49.28%	69.98%	80.52%	60.62%	
Image 13	Correct	82.12%	77.50%	64.23%	77.05%	80.48%	72.73%	671
	Missed	7.60%	6.86%	2.53%	3.58%	12.82%	1.64%	
	Under	12.07%	15.20%	9.09%	8.35%	4.92%	0.00%	
	Over	5.81%	7.30%	26.68%	14.61%	14.61%	27.27%	
	F-Score	86.57%	84.07%	77.03%	85.31%	83.27%	83.42%	

	Multi spectral Unmixing	RGB unmixing	Multi spectral LDA	RGB LDA	Red Channel	Manual	No. Of Nuclei
Image 14 Correct	90.44%	90.81%	84.19%	89.34%	86.40%	96.69%	544
Missed	20.59%	16.54%	18.20%	26.47%	32.17%	52.76%	
Under	2.57%	2.21%	1.47%	1.10%	1.84%	0.00%	
Over	6.99%	6.99%	14.34%	9.56%	11.76%	3.31%	
F-Score	85.71%	87.59%	83.20%	82.79%	79.06%	77.52%	
Image 15 Correct	86.64%	86.23%	78.95%	83.40%	85.02%	95.14%	247
Missed	23.48%	21.05%	21.05%	33.60%	33.60%	85.43%	
Under	0.40%	0.00%	0.81%	0.81%	0.81%	0.00%	
Over	12.96%	13.77%	20.24%	15.79%	14.17%	4.86%	
F-Score	82.47%	83.20%	78.95%	76.87%	77.78%	67.82%	
Image 16 Correct	87.24%	84.87%	83.38%	83.68%	77.45%	94.66%	337
Missed	13.06%	13.06%	23.74%	13.95%	40.65%	45.99%	
Under	3.26%	4.75%	1.19%	4.15%	1.19%	0.00%	
Over	9.50%	10.39%	15.43%	12.17%	21.36%	5.34%	
F-Score	87.11%	85.76%	80.52%	84.68%	71.02%	78.67%	
Image 17 Correct	92.31%	86.29%	72.91%	88.63%	92.64%	86.62%	299
Missed	25.75%	15.72%	9.70%	34.78%	45.82%	28.43%	
Under	3.01%	7.02%	18.06%	5.69%	1.00%	0.00%	
Over	4.68%	6.69%	9.03%	5.69%	6.35%	13.38%	
F-Score	84.66%	85.43%	79.85%	79.34%	77.70%	80.56%	
Image 18 Correct	84.81%	84.81%	82.33%	84.10%	87.99%	92.93%	283
Missed	9.54%	10.60%	24.73%	31.10%	32.51%	25.09%	
Under	2.12%	2.47%	0.71%	3.89%	1.06%	0.00%	
Over	13.07%	12.72%	16.96%	12.01%	10.95%	7.07%	
F-Score	87.27%	86.80%	79.52%	78.16%	79.81%	85.25%	
Image 19 Correct	80.84%	83.23%	70.36%	77.25%	78.14%	97.60%	334
Missed	17.07%	17.96%	18.86%	17.37%	29.34%	58.68%	
Under	3.89%	2.99%	0.60%	2.40%	1.50%	0.00%	
Over	15.27%	13.77%	29.04%	20.36%	20.36%	2.40%	
F-Score	81.69%	82.74%	74.37%	79.38%	75.32%	76.17%	
Image 20 Correct	73.36%	71.50%	64.95%	66.82%	69.63%	82.24%	214
Missed	14.49%	16.82%	15.89%	22.90%	22.43%	24.30%	
Under	11.68%	10.75%	7.01%	7.94%	8.41%	0.00%	
Over	14.95%	17.76%	28.04%	25.23%	21.96%	17.76%	
F-Score	78.11%	75.93%	71.83%	70.44%	72.51%	79.64%	
Image 21 Correct	88.64%	83.15%	75.46%	84.62%	79.49%	82.42%	273
Missed	23.44%	21.98%	16.12%	19.41%	38.83%	52.38%	
Under	1.10%	5.13%	0.73%	2.20%	5.49%	0.00%	
Over	10.26%	11.72%	23.81%	13.19%	15.02%	17.58%	
F-Score	83.59%	81.07%	78.78%	82.94%	72.82%	70.20%	
Image 22 Correct	88.89%	95.37%	88.43%	93.98%	87.50%	94.91%	216
Missed	58.80%	74.07%	50.46%	86.11%	87.96%	101.39%	
Under	2.78%	0.93%	1.39%	2.31%	1.39%	0.00%	

		Multi spectral Unmixing	RGB unmixing	Multi spectral LDA	RGB LDA	Red Channel	Manual	No. Of Nuclei
	Over	8.33%	3.70%	10.19%	3.70%	11.11%	5.09%	
	F-Score	71.78%	70.79%	74.03%	67.11%	63.53%	64.06%	
Image 23	Correct	80.54%	80.54%	75.10%	75.10%	82.88%	89.11%	257
	Missed	3.11%	1.95%	7.39%	8.56%	12.45%	30.74%	
	Under	2.33%	3.11%	1.17%	3.11%	1.17%	0.00%	
	Over	17.12%	16.34%	23.74%	21.79%	15.95%	10.89%	
	F-Score	87.71%	88.27%	82.30%	81.78%	84.86%	81.06%	
Total	Correct	76.81%	74.30%	61.58%	70.45%	72.56%	65.24%	10827
	Missed	12.69%	13.25%	10.35%	14.57%	22.42%	15.36%	
	Under	10.03%	9.37%	6.47%	8.50%	8.49%	0.02%	
	Over	13.16%	16.34%	31.95%	21.05%	18.95%	34.75%	
	F-Score	81.06%	79.23%	71.63%	76.15%	74.43%	72.24%	

B.3. CD8 Set

Appendix Table 3 CD8 segmentation results

		Multi spectral Unmixing	RGB unmixing	Multi spectral LDA	RGB LDA	Red Channel	Manual	No. Of Nuclei
Image 1	Correct	86.96%	77.08%	84.19%	84.98%	82.61%	92.89%	253
	Missed	22.13%	21.34%	26.09%	25.69%	16.60%	38.34%	
	Under	5.14%	18.18%	5.53%	5.14%	13.44%	0.00%	
	Over	7.91%	4.74%	10.28%	9.88%	3.95%	7.11%	
	F-Score	83.18%	77.69%	80.08%	80.68%	82.94%	80.34%	
Image 2	Correct	90.76%	82.63%	79.55%	81.51%	84.03%	88.80%	357
	Missed	5.60%	8.12%	10.64%	10.08%	7.00%	19.33%	
	Under	3.08%	3.36%	2.24%	2.80%	7.84%	0.00%	
	Over	6.16%	14.01%	18.21%	15.69%	8.12%	11.20%	
	F-Score	92.44%	86.64%	83.65%	85.09%	87.98%	85.33%	
Image 3	Correct	85.15%	82.07%	68.07%	75.63%	75.63%	81.51%	357
	Missed	4.76%	5.88%	5.32%	8.12%	5.88%	9.80%	
	Under	4.76%	6.16%	2.80%	3.92%	9.80%	0.00%	
	Over	10.08%	11.76%	29.13%	20.45%	14.57%	18.49%	
	F-Score	89.68%	87.33%	78.51%	82.32%	83.33%	85.21%	
Image 4	Correct	85.34%	80.13%	72.96%	77.85%	74.92%	84.69%	307
	Missed	4.56%	7.49%	6.84%	7.82%	6.51%	11.40%	
	Under	3.58%	6.51%	2.93%	2.93%	7.82%	0.00%	
	Over	11.07%	13.36%	24.10%	19.22%	17.26%	15.31%	
	F-Score	89.88%	85.42%	81.16%	83.86%	82.59%	86.38%	

		Multi spectral Unmixing	RGB unmixing	Multi spectral LDA	RGB LDA	Red Channel	Manual	No. Of Nuclei
Image 5	Correct	64.40%	64.13%	51.36%	56.52%	54.08%	78.80%	368
	Missed	1.90%	4.35%	5.16%	6.25%	1.90%	10.87%	
	Under	6.52%	10.33%	4.62%	5.71%	10.87%	0.00%	
	Over	29.08%	25.54%	44.02%	37.77%	35.05%	21.20%	
	F-Score	77.45%	76.13%	65.63%	69.45%	69.34%	83.09%	
Image 6	Correct	73.56%	73.85%	64.66%	61.78%	60.06%	64.37%	348
	Missed	3.74%	5.75%	7.18%	8.62%	3.74%	7.47%	
	Under	9.20%	9.77%	6.03%	6.61%	15.23%	0.29%	
	Over	17.24%	16.38%	29.31%	31.61%	24.71%	35.34%	
	F-Score	82.98%	82.24%	75.25%	72.51%	73.33%	74.92%	
Image 7	Correct	74.55%	69.76%	65.27%	68.26%	70.36%	44.61%	334
	Missed	4.49%	5.99%	8.38%	12.28%	10.18%	1.50%	
	Under	6.89%	8.98%	2.69%	3.29%	4.79%	0.00%	
	Over	18.56%	21.26%	32.04%	28.44%	24.85%	55.39%	
	F-Score	83.28%	79.39%	75.17%	75.62%	77.94%	61.07%	
Image 8	Correct	86.63%	84.30%	77.33%	80.52%	81.98%	72.67%	344
	Missed	5.52%	9.01%	7.85%	10.17%	11.05%	0.58%	
	Under	3.20%	4.07%	2.03%	3.49%	3.20%	0.00%	
	Over	10.17%	11.63%	20.64%	15.99%	14.83%	27.33%	
	F-Score	90.17%	87.22%	83.52%	84.45%	84.94%	83.89%	
Image 9	Correct	74.22%	70.93%	56.98%	63.57%	50.78%	77.52%	516
	Missed	2.71%	5.23%	7.56%	9.30%	10.47%	1.36%	
	Under	5.43%	8.33%	6.98%	6.20%	43.99%	0.00%	
	Over	20.35%	20.74%	36.05%	30.23%	5.23%	22.48%	
	F-Score	83.90%	80.53%	69.26%	73.54%	62.98%	86.67%	
Image 10	Correct	73.47%	69.79%	63.05%	62.21%	61.68%	83.37%	950
	Missed	5.47%	6.95%	9.47%	10.53%	6.74%	4.21%	
	Under	5.26%	5.47%	5.89%	6.74%	10.32%	0.00%	
	Over	21.26%	24.74%	31.05%	31.05%	28.00%	16.63%	
	F-Score	82.12%	78.98%	73.09%	72.03%	73.25%	88.89%	
Image 11	Correct	79.93%	68.09%	62.83%	62.50%	65.46%	69.08%	304
	Missed	7.57%	11.18%	9.87%	8.22%	8.88%	12.17%	
	Under	3.29%	7.89%	7.24%	6.25%	18.09%	0.00%	
	Over	16.78%	24.01%	29.93%	31.25%	16.45%	30.92%	
	F-Score	85.26%	75.96%	72.76%	73.22%	75.09%	76.23%	
Image 12	Correct	76.58%	69.62%	59.18%	60.44%	55.70%	77.22%	316
	Missed	2.53%	5.38%	5.70%	5.70%	4.75%	9.18%	
	Under	6.01%	8.86%	6.01%	6.33%	12.97%	0.00%	
	Over	17.41%	21.52%	34.81%	33.23%	31.33%	22.78%	
	F-Score	85.51%	79.57%	71.79%	72.76%	69.43%	82.85%	

		Multi spectral Unmixing	RGB unmixing	Multi spectral LDA	RGB LDA	Red Channel	Manual	No. Of Nuclei
Image 13	Correct	61.77%	57.68%	63.14%	64.16%	43.00%	98.63%	293
	Missed	9.22%	7.51%	11.26%	16.72%	16.04%	2.05%	
	Under	4.44%	5.80%	9.56%	23.55%	53.58%	0.00%	
	Over	33.79%	36.52%	27.30%	12.29%	3.41%	1.37%	
	F-Score	72.26%	69.83%	72.41%	70.94%	54.08%	98.30%	
Image 14	Correct	67.57%	65.72%	54.33%	62.25%	63.49%	89.36%	808
	Missed	4.33%	5.07%	10.40%	9.65%	6.81%	2.60%	
	Under	4.83%	7.80%	4.58%	3.34%	5.45%	0.00%	
	Over	27.60%	26.49%	41.09%	34.41%	31.06%	10.64%	
	F-Score	78.62%	76.96%	65.97%	72.43%	74.56%	93.10%	
Image 15	Correct	83.97%	82.24%	78.62%	77.07%	82.24%	87.41%	580
	Missed	12.41%	14.66%	13.62%	16.55%	16.21%	3.62%	
	Under	3.45%	3.97%	4.31%	5.69%	6.72%	0.00%	
	Over	12.59%	13.79%	17.07%	17.24%	11.03%	12.59%	
	F-Score	85.51%	83.54%	81.79%	79.61%	82.88%	91.52%	
Image 16	Correct	73.24%	69.23%	54.63%	54.74%	55.07%	63.99%	897
	Missed	7.02%	11.26%	10.81%	10.37%	8.58%	4.91%	
	Under	6.69%	11.04%	9.36%	8.47%	21.07%	0.00%	
	Over	20.07%	19.73%	36.01%	36.79%	23.86%	36.01%	
	F-Score	81.26%	76.71%	66.04%	66.31%	67.30%	75.78%	
Image 17	Correct	69.30%	67.48%	47.11%	48.33%	60.18%	78.42%	329
	Missed	4.86%	8.51%	7.60%	5.78%	9.42%	13.37%	
	Under	5.78%	8.21%	5.78%	6.08%	9.42%	0.00%	
	Over	24.92%	24.32%	47.11%	45.59%	30.40%	21.58%	
	F-Score	79.58%	76.68%	60.90%	62.72%	70.97%	81.77%	
Image 18	Correct	67.77%	67.77%	50.92%	58.24%	57.14%	80.95%	273
	Missed	6.96%	8.79%	12.45%	9.89%	6.96%	16.48%	
	Under	6.96%	13.55%	7.69%	8.06%	17.95%	0.00%	
	Over	25.27%	18.68%	41.39%	33.70%	24.91%	19.05%	
	F-Score	77.57%	76.76%	62.33%	69.28%	69.64%	82.00%	
Image 19	Correct	72.76%	71.40%	54.28%	58.95%	59.92%	89.49%	514
	Missed	6.61%	9.53%	10.70%	9.53%	7.00%	7.00%	
	Under	7.78%	12.84%	11.67%	11.48%	18.68%	0.00%	
	Over	19.46%	15.76%	34.05%	29.57%	21.40%	10.51%	
	F-Score	81.13%	78.92%	65.80%	69.98%	71.79%	91.09%	
Image 20	Correct	58.41%	56.87%	42.17%	49.90%	51.45%	94.20%	517
	Missed	2.51%	2.90%	3.87%	4.06%	2.90%	0.77%	
	Under	8.32%	9.67%	6.38%	5.42%	6.77%	0.00%	
	Over	33.27%	33.46%	51.45%	44.68%	41.78%	5.80%	
	F-Score	72.60%	71.19%	57.75%	64.82%	66.67%	96.63%	

		Multi spectral Unmixing	RGB unmixing	Multi spectral LDA	RGB LDA	Red Channel	Manual	No. Of Nuclei
Total	Correct	74.47%	71.02%	61.28%	64.15%	63.52%	80.09%	8965
	Missed	5.99%	8.06%	9.45%	10.11%	8.19%	7.17%	
	Under	5.60%	8.31%	5.97%	6.49%	14.52%	0.01%	
	Over	19.93%	20.67%	32.75%	29.36%	21.95%	19.90%	
	F-Score	82.53%	79.31%	71.79%	73.63%	73.99%	85.54%	