

Approches virtuelles dédiées à la technologie des puces à tissus "Tissue MicroArrays " TMA : Application à l'étude de la transformation tumorale du tissu colorectal

Redha Heus

▶ **To cite this version:**

Redha Heus. Approches virtuelles dédiées à la technologie des puces à tissus "Tissue MicroArrays " TMA : Application à l'étude de la transformation tumorale du tissu colorectal. Sciences du Vivant [q-bio]. Université Joseph-Fourier - Grenoble I, 2009. Français. NNT : . tel-00429056

HAL Id: tel-00429056
https://theses.hal.science/tel-00429056

Submitted on 30 Oct 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT
DE
L'UNIVERSITÉ GRENOBLE 1 - JOSEPH FOURIER

ÉCOLE DOCTORALE INGÉNIERIE POUR LA SANTÉ, LA COGNITION ET L'ENVIRONNEMENT

Spécialité : " BIOTECHNOLOGIE, INSTRUMENTATION, SIGNAL ET IMAGERIE POUR LA BIOLOGIE, LA MEDECINE ET L'ENVIRONNEMENT"

Présentée par
Redha HEUS

Pour obtenir le titre de

DOCTEUR DE L'UNIVERSITÉ JOSEPH FOURIER

Approches virtuelles dédiées à la technologie des puces à tissus «Tissue MicroArrays » TMA : Application à l'étude de la transformation tumorale du tissu colorectal

Soutenue le 28 Septembre 2009
Devant le jury composé de :

François ESTÈVE	**Président**
Paulette HERLIN	**Rapporteur**
Philippe VAGO	**Rapporteur**
Jean-Christophe SABOURIN	**Examinateur**
Jacques KLOSSA	**Examinateur**
Ahlame DOUZAL	**Examinateur**
Françoise GIROUD	**Directrice de thèse**

Thèse préparée au Laboratoire : Techniques de l'Ingénierie Médicale et de la Complexité -Informatique, Mathématiques et Applications de Grenoble (TIMC-IMAG)

Remerciements

Je tiens à remercier, Madame Françoise Giroud, ma directrice de thèse, pour la confiance qu'elle m'a accordée durant les années de Master2 et de doctorat. Merci Françoise pour tes conseils pertinents et tes encouragements durant les moments difficiles de la thèse.

Je remercie également Monsieur Philippe Vago, Professeur de l'Université de Clermont-Ferrand, et Madame Paulette Herlin, du Centre Régional de Lutte Contre le Cancer (CRLCC) François Baclesse de Caen, d'avoir accepté avec beaucoup de spontanéité d'être les rapporteurs de cette thèse. Merci à Monsieur François Estève, Professeur de l'université Grenoble 1, d'avoir accepté la charge de présider cette thèse. Je leur adresse tous mes remerciements.

J'exprime mes remerciements aux autres membres du jury, pour avoir aimablement accepté de juger mon travail. Merci à Monsieur Jean-Christophe Sabourin, Professeur au CHU de Rouen, à Monsieur Jacques Klossa de la société TRIBVN et à Madame Ahlame Douzal de l'université de Grenoble 1, d'avoir accepté de participer à mon jury de thèse.

Merci au Docteur Joëlle Simony-Lafontaine, pour sa collaboration et son implication pour l'aboutissement de mes travaux de recherche. Merci pour ton aide Joëlle.

Un merci particulier à Mesdames Marie-Paule MontMasson et Malika Hamel pour leur contribution avec beaucoup de compétences dans la partie expérimentale de ma thèse. Merci Malika et Marie-Paule pour votre gentillesse et pour la bonne ambiance que vous faite régner dans l'équipe.

Merci à Olivier et à Noureddine pour leurs conseils, leur disponibilité et leur soutien durant la thèse. Merci pour vos remarques avisées concernant la rédaction de ce manuscrit. Merci pour votre bonne humeur et votre amitié.

Enfin, j'adresse mes remerciements à ma famille qui m'a soutenu et m'a supporté pendant ces années de thèse. Je dédie cette thèse à mon père.

Table des matières

INTRODUCTION GENERALE **21**

CHAPITRE 1 LE CANCER COLORECTAL 27
1.1 Introduction 28
1.1.1 Cancer et prolifération cellulaire 28
1.1.2 Progression tumorale 30
1.1.3 Les différentes étapes du développement d'un cancer 31
1.2 Anatomie générale du colon 32
1.3 Histologie du colon 33
1.3.1 La muqueuse 33
1.3.2 La musculaire muqueuse 34
1.3.3 La sous-muqueuse 34
1.3.4 La tunique musculaire propre 34
1.3.5 La séreuse 34
1.4 Organisation fonctionnelle de l'épithélium colique 34
1.5 Principales étapes de la transformation tumorale du colon 37
1.5.1 Etats précurseurs du cancer colorectal 37
1.5.1.1 Hyperplasie, dysplasie et néoplasie intra-épithélial 37
1.5.1.2 Les Cryptes Aberrantes 38
1.5.1.3 Les Adénomes 38
1.5.2 De l'adénome vers le cancer colorectal 39
1.5.3 Les cancers invasifs ou adénocarcinomes 41
1.6 Le système de classification TNM 41
1.7 Cancérogenèse colorectale 43
1.7.1 Réseau moléculaire de la prolifération cellulaire 44
1.7.2 Les altérations moléculaires associées au cancer colorectal 45
1.8 Conclusion 47

CHAPITRE 2 LA TECHNOLOGIE DES TMA : TISSUE MICROARRAYS 49
2.1 Introduction 50
2.2 Principe des TMA 52
2.3 Plan de construction des TMA 54
2.3.1 Elaboration du plan de construction des TMA 54
2.3.2 Confection du bloc TMA 56
2.3.3 Réalisation des coupes TMA 58
2.4 Domaines d'application des TMA 59
2.5 Problématiques liées à la technologie TMA 60
2.5.1 Echantillonnage des spots sur la biopsie 61
2.5.2 Numérisation des coupes TMA 62
2.5.3 Analyse des lames TMA numériques 63
2.5.4 Gestion et exploitation des données générées par la technologie des TMA 65
2.6 Conclusion 66

CHAPITRE 3 CONCEVOIR DES PUCES A TISSUS TMA : METHODOLOGIE D'ECHANTILLONNAGE BASEE SUR LE CONCEPT DE « CAROTTES VIRTUELLES » 67
3.1 Introduction 68
3.2 Matériel et méthodes 72
3.2.1 Matériel biologique 72
3.2.2 Coloration HES 72
3.2.3 Les marqueurs tumoraux 73
3.2.4 Acquisition des images histologiques 73

3.2.5 Analyse et traitement d'images ... 74
3.2.6 Délimitation des zones d'intérêts ... 75
3.2.6.1 Annotation sur lames HES ... 75
3.2.6.2 Les cartes de prolifération cellulaire ... 76
3.2.7 Echantillonnage des zones d'intérêts ... 77
3.2.7.1 Génération des carottes virtuelles ... 77
3.2.7.2 La procédure d'échantillonnage des carottes virtuelles ... 78
3.3 Résultats ... 81
3.3.1 Mise en œuvre de la stratégie d'échantillonnage virtuel : ... 81
3.3.2 Description des zones d'intérêts à échantillonner : ... 82
3.3.3 Résultats de l'échantillonnage virtuel des zones d'intérêts : ... 84
3.4 Discussion ... 88
3.4.1 Choix des marqueurs tumoraux ... 88
3.4.2 L'approche virtuelle par simulation de prélèvement de carottes TMA : Un outil efficace pour l'évaluation de la technologie TMA : ... 90
3.4.3 Le concept de « TMA virtuel » : un outil prometteur pour la construction des TMA et la validation de nouvelles méthodologies d'exploitation des TMA : ... 91
3.5 Conclusion ... 94

CHAPITRE 4 LA COULEUR ... 95
4.1 Introduction ... 96
4.2 La physiologie de la vision humaine de la couleur ... 98
4.3 Les modèles de la perception humaine de la couleur ... 100
4.3.1 Le modèle du tri-stimulus ... 100
4.3.2 Le modèle des couleurs opposées ... 101
4.3.3 Le modèle des compartiments ... 101
4.3.4 Le modèle perceptuel ... 102
4.4 La colorimétrie « science de mesure de la couleur » ... 103
4.4.1 Introduction ... 103
4.4.2 La trivariance visuelle ... 104
4.4.2.1 Principe ... 104
4.4.2.2 Expériences d'égalisation des couleurs ... 104
4.4.2.3 Lois colorimétriques de la trivariance visuelle ... 105
4.5 Les travaux de normalisation de la CIE ... 106
4.5.1.1 Introduction ... 106
4.5.1.2 Les fonctions colorimétriques r(λ), g(λ) et b(λ) ... 107
4.5.1.3 Les fonctions colorimétriques x(λ), y(λ) et z(λ) ... 111
4.5.1.4 Vers une représentation des couleurs perpétuellement uniforme ... 117

CHAPITRE 5 SYSTEMES DE REPRESENTATION DE LA COULEUR ... 123
5.1 Introduction ... 124
5.2 Les systèmes des primaires ... 124
5.2.1 Le système RGB ... 124
5.2.2 Le système XYZ ... 125
5.2.3 Le système LMS ... 125
5.3 Les systèmes luminance-chrominance ... 126
5.3.1 Les systèmes Y Cb Cr ... 126
5.3.2 Les systèmes antagonistes ... 127
5.3.3 Autres systèmes luminance-chrominance ... 129
5.4 Les systèmes perceptuels ... 129
5.4.1 Les systèmes uniformes ... 129
5.4.1.1 Le système Lab ... 130
5.4.1.2 Le système Luv : ... 130
5.4.2 Les représentations de type Luminance-Teinte-Saturation : ... 131
5.4.2.1 Les systèmes en coordonnées polaires ou cylindriques ... 132
5.4.2.2 Les systèmes triangulaires ... 134
5.4.2.3 Les systèmes en cône hexagonal ... 135
5.4.2.4 Les systèmes d'atlas des couleurs ... 137
5.5 Les systèmes d'axes indépendants ... 138
5.5.1 Le système en composantes principales ... 139
5.5.2 Le système I_1 I_2 I_3 ... 140

5.5.3 Le système en composantes indépendantes ... 141
5.6 Les systèmes hybrides ... 141
5.7 Choix du système couleur pour le traitement des images couleur ... 141
5.8 Conclusion ... 144

CHAPITRE 6 LA SEGMENTATION DES IMAGES COULEUR ... 147
6.1 Introduction ... 148
6.2 Approche région ... 151
6.2.1 Croissance de régions ... 151
6.2.2 Division-fusion des régions ... 155
6.2.3 Conclusion ... 158
6.3 Approche contours ... 158
6.3.1 Introduction ... 158
6.3.2 Fusion de résultats marginaux ... 161
6.3.3 Fusion des vecteurs gradients marginaux ... 162
6.3.4 Gradient couleur LTS ... 163
6.3.5 Méthodes basées sur les dérivées secondes ... 165
6.3.6 La méthode des contours actifs ... 166
6.3.7 Conclusion ... 167
6.4 Approche pixellaire ... 167
6.4.1 Introduction ... 167
6.4.2 Classification basée sur l'histogramme ... 168
6.4.3 Classification basée sur les méthodes de « clustering » ... 171
6.4.4 Conclusion ... 174
6.5 Approche hybrides ... 175
6.6 Conclusion ... 178

CHAPITRE 7 CLASSIFICATION COULEUR DES IMAGES HISTOLOGIQUES ... 183
7.1 Introduction ... 184
7.2 Méthodologies de travail ... 188
7.2.1 Matériel biologique ... 188
7.2.2 Immuno-marquage Ki67 ... 188
7.2.3 Préparation des lames de références ... 189
7.2.4 Acquisition des images histologiques ... 189
7.2.5 Génération des images de carottes virtuelles ... 190
7.2.6 Langage et environnement de développement informatique ... 190
7.3 Modélisation du système colorimétrique DAB-Hématoxyline : Démonstration ... 190
7.3.1 Description des noyaux cellulaires à classifier ... 190
7.3.2 Protocole de classification couleur ... 192
7.3.2.1 Détection des noyaux cellulaires des images : étape de segmentation ... 193
7.3.2.2 Séparation couleur entre le DAB et l'Hématoxyline : étape de ségrégation ... 195
7.3.2.2.1 Analyse colorimétrique du DAB et de l'Hématoxyline ... 197
7.3.2.2.1.1 Choix de l'espace couleur ... 197
7.3.2.2.1.2 Le profil couleur du DAB et de l'Hématoxyline ... 199
7.3.2.2.2 Modèles couleur associées au DAB et à l'Hématoxyline ... 203
7.3.2.2.3 Construction de la table de classification couleur associée au DAB et à l'Hématoxyline ... 207
7.3.2.2.4 Mise en application de la table de classification couleur ... 211
7.3.3 Protocole global de la classification couleur des images histologiques ... 213
7.3.3.1 Phase de segmentation ... 214
7.3.3.2 Phase d'apprentissage ... 214
7.3.3.3 Phase de classification couleur ... 215
7.4 Modélisation du système colorimétrique DAB-Hématoxyline : Tests de validation ... 216
7.4.1 Paramètre de comparaison ... 216
7.4.2 Méthodes de classification automatique ... 217
7.4.2.1 Méthode automatique basée sur la teinte ... 217
7.4.2.2 Méthode automatique basée sur la composante bleue normalisée ... 218
7.4.2.3 Méthode automatique basée sur la déconvolution couleur ... 219
7.4.3 Protocole de classification manuelle ... 221
7.4.4 Construction de la classification de référence ... 223
7.4.5 Critères de comparaison entre les méthodes de classification ... 224
7.4.5.1 L'indice de marquage ... 224

7.4.5.2 La différence absolue des indices de marquage ... 225
7.4.5.3 Le pourcentage des pixels mal classifiés ... 226
7.4.6 Les analyses statistiques ... 227
7.4.6.1 Analyse des indices de marquage (LI) ... 227
7.4.6.2 Analyse des différences absolues et des pourcentages des pixels mal classifiés : ... 228
7.5 Modélisation du système colorimétrique DAB-Hématoxyline : Tests de robustesse ... 228
7.5.1 Variation réelle des conditions d'éclairage ... 228
7.5.2 Variation simulée (virtuelle) des conditions d'éclairage ... 229
7.6 Modélisation du système colorimétrique DAB-Hématoxyline : Résultats des expérimentations 231
7.6.1 Introduction ... 231
7.6.2 Résultats des tests de validation ... 231
7.6.2.1 Les classifications manuelles ... 231
7.6.2.1.1 Indices de marquage ... 231
7.6.2.1.1.1 Description générale ... 231
7.6.2.1.1.2 Résultats statistiques ... 235
7.6.2.1.2 Différences absolues des indices de marquages par rapport aux indices de marquages de référence ... 236
7.6.2.1.3 Pourcentages des pixels mal classés par rapport à la classification de référence ... 238
7.6.2.2 Les classifications automatiques ... 241
7.6.2.2.1 Indices de marquage ... 241
7.6.2.2.1.1 Description générale ... 241
7.6.2.2.1.2 Résultats statistiques ... 242
7.6.2.2.2 Différences absolues des indices de marquages par rapport aux indices de marquages de référence ... 243
7.6.2.2.3 Pourcentages des pixels mal classés par rapport à la classification de référence ... 244
7.6.3 Résultats des tests de robustesse ... 247
7.6.3.1 Variation réelle des conditions d'éclairage du microscope ... 247
7.6.3.2 Variation simulée (virtuelle) de luminosité des images histologiques ... 251
7.7 Modélisation du système colorimétrique DAB-Hématoxyline : Discussion ... 255
7.8 Conclusion ... 262

CHAPITRE 8 SOCIOLOGIE CELLULAIRE : OUTIL D'INVESTIGATION SUR LA PROGRESSION TUMORALE DU CANCER COLORECTAL ... 265
8.1 Introduction ... 266
8.2 Matériel et méthodes ... 266
8.2.1 Analyse de l'architecture tissulaire : le concept de « sociologie cellulaire » ... 266
8.2.1.1 Diagramme de Voronoï ... 267
8.2.1.2 Application à l'étude de la transformation tumorale ... 268
8.2.2 Matériel biologique ... 269
8.2.3 Procédures de traitement d'images ... 269
8.2.4 Positionnement des germes pour construire le diagramme de Voronoï ... 270
8.2.4.1 Génération des germes en mode automatique ... 270
8.2.4.2 Génération des germes en mode semi-automatique ... 273
8.2.5 Les paramètres de sociologie cellulaire ... 275
8.2.5.1 Paramètres locaux liés aux polygones de Voronoï ... 275
8.2.5.2 Paramètres globaux du diagramme de Voronoï ... 276
8.3 Résultats ... 277
8.3.1 Disposition spatiale des cellules prolifératives selon le grade des lésions tumorales ... 278
8.3.1.1 Densité cellulaire en rapport avec le grade tumoral ... 278
8.3.1.2 Diminution du facteur de forme dans les hauts grades ... 280
8.3.1.3 Stabilité relative du facteur d'hétérogénéité d'occupation de l'espace par les cellules : ... 280
8.3.1.4 Diminution du facteur d'homogénéité géométrique des polygones dans les hauts grades : ... 280
8.3.1.5 Augmentation de la distance médiane aux cellules positives voisines avec le grade ... 281
8.3.2 Disposition spatiale des cellules hors cycle selon le grade des lésions tumorales ... 282
8.3.2.1 Topologie des cellules hors cycle ... 282
8.3.2.2 Rapport des cellules hors cycle avec leur entourage ... 282
8.4 Discussion ... 283
8.5 Conclusion ... 285

CHAPITRE 9 CONCLUSIONS ET PERSPECTIVES ... **287**
9.1 Conception d'un bloc TMA ... 288

9.2 Classification couleur des lames histologiques TMA 289
9.3 Exploitation des données générées par la technologie TMA 291
9.4 Le concept « TMA virtuel » : Alternatif et complément à la technologie TMA 292
9.4.1 TMA virtuel : outil d'investigation et d'évaluation de la technologie TMA 293
9.4.2 TMA virtuel : Support de validation de nouvelles approches de traitement d'images 294
9.4.3 TMA virtuel : Outil d'exploitation et de fouille de données 294

BIBLIOGRAPHIE 297

PUBLICATIONS 323

ANNEXE A IMMUNOHISTOCHIMIE : IHC 325

ANNEXE B TMA VIRTUEL : OUTIL COMPLEMENTAIRE A LA TECHNOLOGIE TMA 337

ANNEXE C ANALYSE STATISTIQUE DES METHODES DE CLASSIFICATIONS COULEUR 345

Table des figures

FIGURE 1.1: LES DIFFERENTES ETAPES DE LA PROGRESSION TUMORALE. 31
FIGURE 1.2: ANATOMIE GENERALE DU COLON. 32
FIGURE 1.3 : HISTOLOGIE DU COLON. 33
FIGURE 1.4 : SCHEMA DU RENOUVELLEMENT CELLULAIRE DE L'EPITHELIUM COLIQUE. 35
FIGURE 1.5 : LES DIFFERENTS STADES DE LA TRANSITION ADENOME-CANCER 39
FIGURE 1.6 : CORRESPONDANCE ENTRE LA CLASSIFICATION TNM ET LE STADE DES CANCERS COLORECTAUX 43
FIGURE 1.7: SCHEMA DE REGULATION DE LA B-CATENINE PAR LA VOIE WNT [L'ALLEMAIN 2006] 44
FIGURE 1.8 : SCHEMA DE CANCEROGENESE COLORECTALE DE VOGELSTEIN. 47
FIGURE 2.1 : PRINCIPE DE LA TECHNOLOGIE TMA 52
FIGURE 2.2 : COUPE HISTOLOGIQUE D'UN BLOC TMA 53
FIGURE 2.3 : LE TMA MICRO ARRAYER (BEECHERISTRUMENTSTM) 54
FIGURE 2.4 : ANNOTATION ANATOMOPATHOLOGIQUE D'UNE COUPE HISTOLOGIQUE HES. 55
FIGURE 2.5 : CONSTRUCTION D'UN BLOC TMA. 57
FIGURE 2.6 : PRINCIPE DE LA TECHNIQUE PSA®. 58
FIGURE 2.7 : AVANTAGE DE LA TECHNIQUE PSA®. 59
FIGURE 3.1 : STATION D'ACQUISITION D'IMAGES SAMBA2050TM. 73
FIGURE 3.2 : SYSTEME DE VISUALISATION DES LAMES HISTOLOGIQUES NUMERIQUES NAVIQAP™. 74
FIGURE 3.3 : ANNOTATION DES ZONES D'INTERETS SUR LES LAMES HES. 75
FIGURE 3.4: CARTE DE PROLIFERATION CELLULAIRE. 76
FIGURE 3.5 : PROCEDURE DE DELIMITATION DES ZONES D'INTERETS.. 77
FIGURE 3.6 : PROCEDURE DE GENERATION DES CAROTTES VIRTUELLES. 78
FIGURE 3.7 : PRINCIPE DU PRELEVEMENT ALEATOIRE DES CAROTTES VIRTUELLES. 79
FIGURE 3.8 : EXEMPLE DE DISTRIBUTION D'UNE SERIE D'ECHANTILLONNAGE VIRTUEL. 79
FIGURE 3.9 : ECHANTILLONNAGE VIRTUEL D'UNE ZONE D'INTERET. 81
FIGURE 3.10: EXEMPLE D'ECHANTILLONNAGE VIRTUEL DE DEUX ZONES D'INTERETS DE LAMES KI67 82
FIGURE 3.11: LIEN ENTRE L'HETEROGENEITE DES MARQUEURS ET LA SURFACE DES ZI. 82
FIGURE 3.12: LIEN ENTRE L'HETEROGENEITE DES MARQUEURS ET LE NIVEAU DE MARQUAGE DES ZI. 84
FIGURE 3.13: LIEN ENTRE LE NOMBRE OPTIMAL ET LA SURFACE DES ZI.. 85
FIGURE 3.14: LIEN ENTRE LE NOMBRE OPTIMAL ET LE LI DES ZI.. 85
FIGURE 3.15: LIEN ENTRE LE NOMBRE OPTIMAL ET L'HETEROGENEITE DES MARQUEURS. 86
FIGURE 3.16: PRESENTATION DU CONCEPT DU TMA VIRTUEL. 93
FIGURE 4.1: EXPERIENCE DE NEWTON. DECOMPOSITION DE LA COULEUR BLANCHE PAR UN PRISME 96
FIGURE 4.2: LE SPECTRE VISIBLE DES COULEURS. ETENDUE SPECTRALE DES COULEURS VISIBLES. 97
FIGURE 4.3: ANATOMIE GENERALE DE L'ŒIL. 98
FIGURE 4.4: LES SPECTRES DE SENSIBILITE DES CELLULES DE LA RETINE : CONES (S, M ET L) ET BATONNETS. 99
FIGURE 4.5: EXPERIENCES D'EGALISATION VISUELLE DES COULEURS. 105
FIGURE 4.6: FONCTIONS COLORIMETRIQUES R(Λ), G(Λ) ET B(Λ) CIE 1931 108
FIGURE 4.7: CUBE DES COULEURS RGB ET DIAGRAMME DE CHROMATICITE (R, G). 109
FIGURE 4.8: DIAGRAMME DE WRIGHT (R, G) 111
FIGURE 4.9: DIAGRAMME DE WRIGHT (R,G) ET LE TRIANGLE XYZ. 112
FIGURE 4.10: FONCTIONS COLORIMETRIQUES X(Λ), Y(Λ) ET Z(Λ) DE LA CIE 1931 113
FIGURE 4.11: DIAGRAMME DE CHROMATICITE XY CIE 1931. 114

FIGURE 4.12: ELLIPSES DE MACADAM. 116
FIGURE 4.13: LE DIAGRAMME UV CIE 1976. 119
FIGURE 4.14: LES ELLIPSES DE MACADAM UV. 119
FIGURE 4.15: DIAGRAMME DE CHROMATICITE *AB*. 120
FIGURE 5.1: REPRESENTATION 3D DE L'ESPACE *LAB*. 130
FIGURE 5.2: REPRESENTATION CYLINDRIQUE DE L'ESPACE TLS. 132
FIGURE 5.3: REPRESENTATION LCH DU SYSTEME *LAB*. 133
FIGURE 5.4: SYSTEME HSV EN CONE HEXAGONAL. 135
FIGURE 5.5: SYSTEME HSV EN CONE DOUBLE-HEXAGONAL. 135
FIGURE 5.6: SYSTEME OU ATLAS DES COULEURS DE MUNSELL. 138
FIGURE 7.1: LAME HISTOLOGIQUE IMMUNO-MARQUEE POUR LA DETECTION DE LA PROTEINE KI67. 185
FIGURE 7.2: EXEMPLES DE NOYAUX AYANT DIFFERENTES NUANCES DE COULEUR MARRON ET BLEUE. 191
FIGURE 7.3: LES PROFILS D'ABSORPTION ASSOCIES AU DAB ET A L'HEMATOXYLINE. 191
FIGURE 7.4: PROTOCOLE GENERAL DE CLASSIFICATION COULEUR. 192
FIGURE 7.5: SEGMENTATION BASEE SUR L'AXE DE LUMINOSITE. 193
FIGURE 7.6: SCHEMA DE SEGMENTATION POUR ISOLER LES NOYAUX CELLULAIRES DU FOND DE L'IMAGE. 194
FIGURE 7.7: FUSION DE LA SEGMENTATION DES PLANS MARGINAUX R, G ET B. 194
FIGURE 7.8: PROFIL DES NUAGES DES POINTS COULEUR ASSOCIES A UNE IMAGE HISTOLOGIQUE. 195
FIGURE 7.9: LE DIAGRAMME DE CHROMATICITE XY. 198
FIGURE 7.10: FORMULATION DES COULEURS AU NIVEAU DU DIAGRAMME DE CHROMATICITE XY. 199
FIGURE 7.11: NUAGE COULEUR D'UNE IMAGE HISTOLOGIQUE COMPORTANT LE DAB ET L'HEMATOXYLINE. 200
FIGURE 7.12: NUAGE COULEUR ASSOCIE A UNE IMAGE DE REFERENCE POSITIVE DAB. 201
FIGURE 7.13: NUAGE COULEUR ASSOCIE A UNE IMAGE DE REFERENCE NEGATIVE HEMATOXYLINE. 201
FIGURE 7.14: MISE EN RELATION DES DEUX NUAGES COULEUR DAB ET HEMATOXYLINE. 202
FIGURE 7.15: ANALYSE COLORIMETRIQUE LIEE AU DAB ET A L'HEMATOXYLINE. 204
FIGURE 7.16: NUANCES COULEUR ASSOCIEES AU MODELE COULEUR DAB. 205
FIGURE 7.17: NUANCES COULEUR ASSOCIEES AU MODELE COULEUR HEMATOXYLINE. 205
FIGURE 7.18: PROJECTION DES MODELES COULEUR SUR LES PLANS R, G ET B. 206
FIGURE 7.19: NUANCES COULEUR DU MODELE DU MELANGE COULEUR ENTRE LE DAB ET L'HEMATOXYLINE. 207
FIGURE 7.20: GENERATION DES POINTS COULEUR ASSOCIES A LA MIXTURE DAB-HEMATOXYLINE. 208
FIGURE 7.21: CONSTRUCTION DE LA TABLE DE CLASSIFICATION COULEUR. 209
FIGURE 7.22: TABLE DE CLASSIFICATION COULEUR ASSOCIEE AU DAB ET A L'HEMATOXYLINE. 211
FIGURE 7.23: PROTOCOLE D'UTILISATION DE LA TABLE DE CLASSIFICATION COULEUR DAB-HEMATOXYLINE. 212
FIGURE 7.24: PLAN GLOBAL DE LA CLASSIFICATION COULEUR DES IMAGES HISTOLOGIQUES. 213
FIGURE 7.25: EXEMPLE D'IMAGE HISTOLOGIQUE CLASSIFIEE. 216
FIGURE 7.26: PROTOCOLE DE CLASSIFICATION AUTOMATIQUE BASE SUR LA TEINTE. 218
FIGURE 7.27: PROTOCOLE DE CLASSIFICATION AUTOMATIQUE BASE SUR LA COMPOSANTE BLEUE NORMALISEE. 219
FIGURE 7.28: PRINCIPE DE LA DECONVOLUTION COULEUR. 220
FIGURE 7.29: PROTOCOLE DE CLASSIFICATION AUTOMATIQUE BASE SUR LA DECONVOLUTION COULEUR. 220
FIGURE 7.30: INTERFACE GRAPHIQUE DE LA CLASSIFICATION MANUELLE. 221
FIGURE 7.31: EXEMPLES DE RESULTATS DE CLASSIFICATION MANUELLE. 222
FIGURE 7.32: CRITERES DE COMPARAISON ENTRE LES CLASSIFICATIONS. 227
FIGURE 7.33: VARIATIONS VIRTUELLES DE LA LUMINOSITE D'UNE IMAGE HISTOLOGIQUE. 230
FIGURE 7.34: INDICES DE MARQUAGE DES CLASSIFICATIONS MANUELLES REGROUPES PAR OBSERVATEUR. 232
FIGURE 7.35: INDICES DE MARQUAGE DES 9 CLASSIFICATIONS MANUELLES. 233
FIGURE 7.36: EXEMPLES DE CLASSIFICATIONS APPARIEES. 234
FIGURE 7.37: DIFFERENCES ABSOLUES DES LI ASSOCIEES AUX CLASSIFICATIONS MANUELLES APPARIEES. 234

FIGURE 7.38: LES DIFFERENCES ABSOLUES (LI-LI(REF)) DES 9 CLASSIFICATIONS MANUELLES.....236
FIGURE 7.39: LES DIFFERENCES ABSOLUES (LI-LI(REF)) REGROUPEES PAR OBSERVATEUR.....237
FIGURE 7.40: LES DIFFERENCES ABSOLUES (LI-LI(REF)) DES CLASSIFICATIONS DE L'OBSERVATEUR FG.....237
FIGURE 7.41: LES DIFFERENCES ABSOLUES (LI-LI(REF)) DES CLASSIFICATIONS DE L'OBSERVATEUR MH. 237
FIGURE 7.42: LES DIFFERENCES ABSOLUES (LI-LI(REF)) DES CLASSIFICATIONS DE L'OBSERVATEUR MPM. 238
FIGURE 7.43: POURCENTAGES DES PIXELS MAL CLASSES ASSOCIES AUX CLASSIFICATIONS MANUELLES. 238
FIGURE 7.44: POURCENTAGES DES PIXELS MAL CLASSES REGROUPES PAR OBSERVATEUR.....239
FIGURE 7.45: POURCENTAGES DES PIXELS MAL CLASSES DES CLASSIFICATIONS DE L'OBSERVATEUR FG.....240
FIGURE 7.46: POURCENTAGES DES PIXELS MAL CLASSES DES CLASSIFICATIONS DE L'OBSERVATEUR MH. 240
FIGURE 7.47: POURCENTAGES DES PIXELS MAL CLASSES DES CLASSIFICATIONS DE L'OBSERVATEUR MPM.....240
FIGURE 7.48: DESCRIPTION GENERALE DES LI ASSOCIES AUX CLASSIFICATIONS AUTOMATIQUES.. 241
FIGURE 7.49: DISTRIBUTION DES LI ASSOCIES AUX CLASSIFICATIONS AUTOMATIQUES. 242
FIGURE 7.50: DIFFERENCES ABSOLUES (LI-LI(REF)) DES CLASSIFICATIONS AUTOMATIQUES. 243
FIGURE 7.51: DIFFERENCES ABSOLUES (LI-LI (REF)) ASSOCIEES AUX CLASSIFICATIONS AUTOMATIQUES.....244
FIGURE 7.52: POURCENTAGES DES PIXELS MAL CLASSES ASSOCIES AUX CLASSIFICATIONS AUTOMATIQUES. 244
FIGURE 7.53: POURCENTAGES DE PIXELS MAL CLASSES LIES AUX CLASSIFICATIONS AUTOMATIQUES. 245
FIGURE 7.54: POURCENTAGES DE PIXELS MAL CLASSES LIES AUX CLASSIFICATIONS AUTOMATIQUES. 246
FIGURE 7.55: CLASSIFICATION BN DES IMAGES HISTOLOGIQUES SOUS DIFFERENTS NIVEAUX D'ECLAIRAGE. 247
FIGURE 7.56: CLASSIFICATION CD DES IMAGES HISTOLOGIQUES SOUS DIFFERENTS NIVEAUX D'ECLAIRAGE 248
FIGURE 7.57: CLASSIFICATION CHR DES IMAGES HISTOLOGIQUES SOUS DIFFERENTS NIVEAUX D'ECLAIRAGE.....248
FIGURE 7.58: CLASSIFICATION HUE DES IMAGES HISTOLOGIQUES SOUS DIFFERENTS NIVEAUX D'ECLAIRAGE.....249
FIGURE 7.59: ROBUSTESSE DES CLASSIFICATIONS AUTOMATIQUES SELON LES CONDITIONS D'ECLAIRAGE. 250
FIGURE 7.60: CLASSIFICATION CHR DES IMAGES HISTOLOGIQUES SOUS DES VARIATIONS DE LUMINOSITE. 251
FIGURE 7.61: CLASSIFICATION CD DES IMAGES HISTOLOGIQUES SOUS DES VARIATIONS DE LUMINOSITE. 252
FIGURE 7.62: CLASSIFICATION BN DES IMAGES HISTOLOGIQUES SOUS DES VARIATIONS DE LUMINOSITE.....252
FIGURE 7.63: CLASSIFICATION HUE DES IMAGES HISTOLOGIQUES SOUS DES VARIATIONS DE LUMINOSITE.. 253
FIGURE 7.64: ROBUSTESSE DES CLASSIFICATIONS AUTOMATIQUES FACE AUX VARIATIONS DE LUMINOSITE. 254
FIGURE 8.1: PRINCIPE DU DIAGRAMME DE VORONOÏ.....267
FIGURE 8.2: EXEMPLE D'APPLICATION DU DIAGRAMME DE VORONOÏ EN HISTOLOGIE.....268
FIGURE 8.3: MISE EN APPLCATION DE LA TABLE DE CLASSIFICATION COULEUR.....270
FIGURE 8.4: GENERATION DES IMAGES (KI67+) ET (KI67-).. 271
FIGURE 8.5: DELIMITATION DES GERMES SUR LES NOYAUX CELLULAIRES.. 271
FIGURE 8.6: POSE EN MODE AUTOMATIQUE DES GERMES SUR LES NOYAUX CELLULAIRES.. 272
FIGURE 8.7: POSE EN MODE MANUEL DES GERMES SUR LES NOYAUX CELLULAIRES.....273
FIGURE 8.8: DIAGRAMME DE VORONOÏ ASSOCIE AUX GERMES POSES SUR LES NOYAUX EPITHELIAUX. 274
FIGURE 8.9: POLYGONES DE VORONOÏ ASSOCIES AUX CELLULES EPITHELIALES.....274
FIGURE 8.10: DIAGRAMME DE VORONOÏ DE TISSUS COLIQUES A DIFFERENTS STADES DE DIFFERENCIATION. 279

Table des tableaux

TABLEAU 2.1 : SYSTEMES D'ANALYSE D'IMAGES COMMERCIAUX DEDIES AUX TMA. 64
TABLEAU 3.1: CORRELATION DE L'HETEROGENEITE DES MARQUEURS AVEC LA SURFACE DES ZI. 83
TABLEAU 3.2 : CORRELATION DE L'HETEROGENEITE DES MARQUEURS AVEC LE NIVEAU DE MARQUAGE DES ZI. ..83
TABLEAU 3.3: DESCRIPTEURS GENERAUX DES NOMBRES DE CAROTTES A PRELEVER AU NIVEAU DES ZI. 84
TABLEAU 3.4: RELATION ENTRE LE NOMBRE OPTIMAL ET LA NATURE DU MARQUAGE DES ZI. 86
TABLEAU 3.5: RELATION ENTRE LE NOMBRE OPTIMAL ET LE TYPE DES ZI. 87
TABLEAU 3.6: RELATION DU NOMBRE OPTIMAL AVEC LA NATURE DU MARQUEUR TUMORAL. 88
TABLEAU 7.1: CORRELATION ENTRE LES CLASSIFICATIONS MANUELLES ET LA CLASSIFICATION DE REFERENCE..232
TABLEAU 7.2: DISTRIBUTIONS STATISTIQUES DES LI DES CLASSIFICATIONS MANUELLES. 235
TABLEAU 7.3: DISTRIBUTIONS DES LI DES CLASSIFICATIONS MANUELLES APPARIEES. 236
TABLEAU 7.4: CORRELATION DES CLASSIFICATIONS AUTOMATIQUES AVEC LA REFERENCE. 242
TABLEAU 7.5: DISTRIBUTIONS STATISTIQUES DES LI DES CLASSIFICATIONS AUTOMATIQUES. 243
TABLEAU 7.6: ROBUSTESSE DES CLASSIFICATIONS AUTOMATIQUES SOUS DIFFERENTS NIVEAUX D'ECLAIRAGE...250
TABLEAU 7.7: ROBUSTESSE DES CLASSIFICATIONS AUTOMATIQUES FACE AUX VARIATIONS DE LUMINOSITE. 255
TABLEAU 8.1: FACTEUR DE FORME MOYEN EN RELATION AVEC LE GRADE DU CANCER. 280
TABLEAU 8.2: HETEROGENEITE D'OCCUPATION DE L'ESPACE EN RELATION AVEC LE GRADE DU CANCER. 280
TABLEAU 8.3: HOMOGENEITE GEOMETRIQUE DES CELLULES DES DIFFERENTS GRADES DE CANCER. 281
TABLEAU 8.4: RELATION DE VOISINAGE IMMEDIAT DES CELLULES PROLIFERATIVES. 281
TABLEAU 8.5: DONNEES DE TOPOLOGIE CELLULAIRE ASSOCIEES AU CELLULES HORS CYCLE (KI67-). 282
TABLEAU 8.6: RELATION DE VOISINAGE IMMEDIAT DES CELLULES HORS CYCLE. 283

Abréviations

Ac : Anticorps.
ACI : Analyse en Composantes Indépendantes.
ACP : Analyse en Composantes Principales.
AD: Area Disorder.
AEC: 3-Amino-9-EthylCarbazole.
AFNOR : Association Française de Normalisation.
Ag : Antigène.
ANOVA : ANalysis Of Variance.
APC : Adenomatous Polyposis Coli.
API : Application Programming Interface.
Bcat : ß-caténine.
BN : Bleu Normalisé.
CCD: Charged Coupled Device.
CD: Color Deconvolution.
CD1 : Cycline D1.
CGL : Corps Genouillé Latéral.
CIE : Commission Internationale de l'Eclairage.
CIS : Carcinome *in Situ*.
CRLCC : Centre de Recherche de Lutte Contre le Cancer.
DAB: Di-Amino-Benzidine.
Dsh: Dishevelled.
EQM : Erreur Quadratique Moyenne.
ER : Estrogen Receptor (récepteur d'œstrogène).
PAF : Polyposes Adénomateuses Familiales.
Gb : Giga Byte.
GSK3β : Glycogène synthase kinase 3β.
HES : Hématoxyline Eosine Safran.
HLS : Hue Luminance Saturation.
HSB: Hue Saturation Brightness.
HSL: Hue Saturation Luminance.

HSV: Hue Saturation Value.

IHC : Immuno Histo Chimie.

INC : Institut National du Cancer.

ISH: Intensity Saturation Hue.

ISO: International Standard Organisation.

ITS : Intensité Teinte Saturation.

IVS : Institut de Veille Sanitaire.

JPEG: Joint Photographic Experts Group.

LCH: Luminance Chroma Hue.

LEF: Lymphoid Enhancer Factor.

LI: Labelling Index.

LOG : Laplacian Of Gaussian.

LPE : Ligne de Partage des Eaux.

LTS : Luminance Teinte Saturation.

MO : Méga Octet.

NA: Numerical aperture.

NTSC: National Television Standards Committee.

OMS : Organisation Mondiale de la Santé.

PAL: Phase Alternation by Line.

PPI : Pixels Per Inch.

PR : Progesteron Receptor (récepteur de progestérone).

RAM: Random Access Module.

RF: Roudness Factor (Facteur de Forme).

RFH: Roudness Factor Homogeneity.

RF_m : Mean Roudness Factor (Facteur de Forme Moyen).

RGB : Red Green Blue.

RN : Réseaux de Neurones.

SD : Standard Deviation.

TCF : T-Cell Factor.

TMA : Tissue Micro Arrays.

ZI : Zone d'Intérêt.

Introduction générale

Le cancer est un problème majeur de santé publique vu le nombre important de personnes atteintes. Il apparaît comme un désordre de la prolifération de certaines cellules qui se reproduisent de façon anarchique et excessive, conduisant à l'apparition d'une tumeur. Il est dû au dérèglement de la division d'une partie infime des 60.000 milliards de cellules qui constituent l'organisme humain. La complexité du réseau moléculaire impliqué dans le processus de cancérisation constitue un grand obstacle à la compréhension du mécanisme de la transformation tumorale.

L'étude de ce réseau moléculaire nécessite le recours à des technologies de haut-débit pour mener des études immunohistochimiques « IHC » sur un grand nombre de tumeurs. A cet égard, la technologie des puces à tissus « Tissue Micro Arrays » dite TMA apparaît comme le candidat idéal pour réaliser ce type d'études. Cette technique s'inscrit dans le cadre des micro-technologies qui se développent actuellement en biologie cellulaire, notamment pour valider les cartes d'identité des tumeurs. Elle permet de prendre en compte des informations relatives aux localisations tissulaires et cellulaires des marqueurs tumoraux et de leurs expressions en relation avec le développement tumoral.

La technologie TMA est un outil de recherche qui rend possible des études à grande échelle sur du matériel biologique inclus en paraffine. Cette technique permet ainsi de collecter sur un même bloc de paraffine, dit bloc TMA, plusieurs dizaines voire des centaines d'échantillons, qui sont prélevés sous forme de petites carottes cylindriques à partir de différents blocs de biopsies tumorales. L'utilisation de TMA rend alors envisageable des études simultanées sur de très grands nombres de tissus avec une large variété de marqueurs biologiques.

Dans le cadre des travaux de cette thèse, le concept de TMA virtuel par simulation d'un bloc TMA réel, est proposé pour l'étude des contraintes scientifiques et techniques soulevées par l'utilisation de la technologie TMA. Ce concept est mis en œuvre pour l'étude des marqueurs du cycle cellulaire du cancer colorectal. Le recours à cette approche virtuelle nous a permis de se focaliser sur les différents aspects de traitements d'images et de contrôle qualité liés à la technologie des TMA.

Le premier chapitre situe le contexte médical lié à notre sujet de thèse, il est consacré au cancer colorectal. Ce chapitre décrit les différentes altérations histologiques, à l'échelon tissulaire, allant du colon normal au colon cancéreux. Il retrace également la séquence d'événements moléculaires, à l'échelon cellulaire, responsable de ces altérations. Le chapitre se focalise sur les principaux marqueurs du cycle cellulaire et leur implication dans la progression tumorale du cancer colorectal.

Le deuxième chapitre introduit la technique TMA comme le choix biotechnologique à adopter pour l'étude à grande échelle des marqueurs tumoraux du cancer colorectal. Ce chapitre présente la technologie des TMA, notamment son principe et ses principaux domaines d'application. Il aborde également les problèmes techniques et scientifiques que soulève l'utilisation de cette biotechnologie à haut-débit.

La technologie TMA permet le traitement en masse des échantillons tissulaires. Elle permet l'accélération des procédures classiques : biologiques et histologiques, menant à la découverte de nouveaux marqueurs tumoraux. En revanche, le recours à une telle micro-technologie à haut-débit a fait émerger de nouvelles problématiques concernant :

- la représentativité des microéchantillons tissulaires, inclus dans le bloc TMA, pour répondre à une question biologique donnée,
- le traitement et l'exploitation du nombre considérable de coupes tissulaires (carottes) générées par la technologie TMA.

Notre sujet de recherche s'articule autour de ces problématiques. Au cours de ce travail nous avons adopté le concept de « TMA virtuel » pour répondre aux différents problèmes liés à la

conception des blocs de TMA et à l'analyse automatique des lames histologiques issues de cette technologie.

Le troisième chapitre se focalise particulièrement sur la problématique d'échantillonnage des blocs de tumeurs lors de la construction d'un bloc TMA. Ce chapitre introduit la notion de « carotte virtuelle » pour simuler le prélèvement réel de carottes. Un protocole de carottage aléatoire basé sur ce principe est développé pour désigner le nombre de carottes à prélever en fonction de l'hétérogénéité des blocs de tumeurs à échantillonner en rapport avec les différents marqueurs tumoraux envisagés. Ce protocole d'échantillonnage est mis en application au niveau de ce chapitre pour simuler un échantillonnage optimal d'un groupe de tumeurs colorectales en respectant la nature de la distribution spatiale de l'expression de certains marqueurs du cycle cellulaire.

L'étude des multiples marqueurs tumoraux impliqués dans le développement et la progression du cancer colorectale nécessite l'analyse des profils de leurs expressions sur un grand nombre de coupes histologiques. Le recours à des technologies haut-débit, telles que la technologie TMA et/ou les stations d'acquisition automatisées d'images histologiques, a raccourci considérablement le temps pour préparer et numériser une telle masse de matériel biologique. Cependant, l'un des facteurs limitant pour l'adoption de ces technologies dans les laboratoires reste le volume horaire important alloué à l'analyse manuelle du grand nombre d'images histologiques générées par ces technologies. C'est dans ce contexte que nous avons développé une nouvelle approche de classification couleur, adaptée à l'analyse automatique des images histologiques : une approche qui exploite les caractéristiques colorimétriques de l'immunomarquage et de la contre-coloration pour la quantification automatique de l'expression des marqueurs tumoraux.

Les chapitres 4, 5 et 6 comportent des rappels nécessaires sur la couleur et l'état de l'art sur son exploitation pour la représentation et le traitement automatique des images couleurs. Les chapitres 7 et 8 présentent notre approche de classification couleur et son application pour l'étude de la transformation tumorale du cancer colorectal.

Le quatrième chapitre aborde la notion de la couleur, le système visuel humain et ses propres mécanismes pour appréhender l'information couleur. Ce chapitre expose les différents travaux menés dans le domaine de la colorimétrie. Il s'attache particulièrement à présenter les lois colorimétriques associées à la synthèse additive des couleurs et les techniques fondamentales de représentation et de reproduction de la couleur.

Le cinquième chapitre se rapporte à la présentation des différents systèmes de représentation de la couleur, les atouts et les inconvénients de chacun d'entre eux. Ce chapitre décrit le lien qui existe entre tous ces systèmes couleur, un lien formalisé par différentes équations permettant le passage d'un système couleur vers un autre. Une partie de ce chapitre est consacrée à la présentation des systèmes couleur les plus utilisés dans le domaine du traitement des images couleurs.

Le sixième chapitre présente l'état de l'art sur les différentes méthodes de segmentation des images couleurs citées dans la littérature. Ce chapitre donne un aperçu global sur les principales approches et techniques développées en vue de la segmentation et de la classification des images couleurs. Le chapitre discute également les considérations à prendre en compte lors de la conception du protocole d'analyse automatique des images couleurs.

Le septième chapitre est réservé à la présentation de notre méthodologie pour l'analyse automatique des images histologiques. Il s'agit d'une nouvelle approche de classification couleur pour la quantification de l'expression des marqueurs tumoraux. Ce chapitre présente les modèles colorimétriques développés pour décrire le comportement chromatique du marquage et de la contre-coloration des images histologiques. Sur la base de ces modèles, le chapitre expose ensuite notre stratégie pour construire une table de classification 2D, permettant la séparation couleur entre le marquage et la contre-coloration. Une partie de ce chapitre est réservée à la validation de notre méthode de classification par sa comparaison à différentes approches de classifications couleur automatiques et manuelles.

Le huitième chapitre couronne notre travail de thèse. Il illustre la mise en œuvre de la table de classification 2D développée dans le **septième chapitre**, pour la détection automatique de l'expression du marqueur du cycle cellulaire Ki67 au niveau des carottes virtuelles

introduites dans **le troisième chapitre**. Ce chapitre aborde la notion de sociologie cellulaire et l'adapte pour l'étude des relations de voisinage entre les cellules cancéreuses au cours de la cancérogénèse colorectale (**premier chapitre**). Il met l'accent sur le lien entre les interactions élémentaires des cellules cancéreuses et les bouleversements architecturaux des tumeurs colorectales. Par cette approche, le chapitre illustre l'utilisation des coupes TMA (**deuxième chapitre**) par simulation (carottes virtuelles) pour l'étude de la transformation tumorale dans le cancer colorectal.

Chapitre 1

Le cancer colorectal

Le cancer colorectal est la forme la plus fréquente des tumeurs malignes en France. Avec 37 413 nouveaux cas enregistrés au cours de l'année 2005, en augmentation régulière au cours du temps, le cancer colorectal est un problème de santé public majeur en France.

Ce chapitre situe le contexte médical lié à notre sujet de thèse, il est consacré au cancer colorectal. Une première partie est réservée à la présentation générale du colon, notamment son histologie et le renouvellement physiologique de sa muqueuse.

Dans la deuxième partie, sont décrites les différentes manifestations d'ordre histologique accompagnant la formation, puis la propagation du cancer colorectale. A la fin est exposé l'état de l'art sur la séquence d'altérations moléculaires majeures mises en cause dans la transformation tumorale de la muqueuse colique normale.

La conclusion met l'accent sur la complexité du réseau moléculaire impliqué dans le cancer colorectal. L'étude de ce réseau moléculaire nécessite le recours à des technologies haut-débit pour mener des études immunohistochimiques sur un grand nombre de tumeurs. A cet égard la technologie des puces à tissus « TMA » apparaît comme le candidat idéal pour mener ce type d'études.

1.1 Introduction

A l'échelle mondiale, Plus de 940 000 nouveaux cas de cancers colorectaux sont signalés chaque année. On utilise le terme cancer colorectal, puisqu'on associe souvent le cancer du rectum à celui du colon dans les études épidémiologiques. La France se situe parmi les régions à risque élevé à l'image de l'Europe de l'Ouest, des USA et de l'Australie. Ce cancer est beaucoup plus rare en Amérique du Sud, en Asie et surtout en Afrique.

Le cancer colorectal est la forme la plus fréquente des tumeurs malignes rencontrées dans la population française. Selon les données fournies par l'Institut National du Cancer [INC 2009] et l'Institut de Veille Sanitaire [IVS 2009], 37 413 nouveaux cas ont été diagnostiqués au cours de l'année 2005 en France. A ce titre, le cancer colorectal est le troisième cancer le plus fréquent après le cancer de la prostate et le cancer du sein. Il reste cependant le deuxième cancer en termes de mortalité derrière le cancer du poumon. Il devance le cancer du sein et le cancer de la prostate avec 16 865 décès recensés en 2005.

En France le cancer colorectal est un problème de santé publique majeur. Ce qui explique l'intérêt que nous portons à ce type de cancer, au niveau de cette thèse. En effet, notre travail s'articule autour des différentes technologies associées à la mise en évidence du réseau moléculaire impliqué dans le développement de ce cancer.

Au niveau de ce chapitre, nous allons aborder en premier lieu des notions générales sur la prolifération cellulaire et son rôle dans le développement des cancers. Ensuite, nous allons nous focaliser sur le colon en abordant son anatomie et son histologie. Puis, nous allons présenter les différentes étapes menant à la formation et au développement du cancer colorectal. Enfin, nous nous attacherons à exposer les mécanismes moléculaires à l'échelle cellulaire mis en cause dans le développement tumoral de ce cancer.

1.1.1 Cancer et prolifération cellulaire

L'organisme humain est composé d'environ 60 000 milliards de cellules. Les cellules sont soumises à un renouvellement constant : régulièrement, une partie d'entre elles meurt

remplacée par de nouvelles cellules. Le cancer provient d'un déséquilibre de ce renouvellement. Il est dû à la prolifération anarchique d'une cellule « transformée », « anormale ». Cette prolifération échappe à l'homéostasie.

Les cancers apparaissent comme un désordre de la prolifération des cellules qui se reproduisent de façon excessive et provoquent l'apparition d'une tumeur. Leur étude contribue à élucider la prolifération cellulaire normale et à comprendre comment elle peut être perturbée.

Les organismes vivants sont constitués de très nombreuses cellules qui diffèrent entre elles : elles sont spécialisées pour la fonction qu'elles assurent et sont groupées en tissus. Dans un même tissu, il existe un ou plusieurs types cellulaires. Durant le développement embryonnaire, les différents types cellulaires sont déterminés chacun à un moment et à un endroit définis.
La plupart des cellules ont des durées de vie courtes et elles doivent se diviser pour être renouvelées. Les cellules de la peau, des muqueuses, des poumons ou du sang sont rapidement détruites et continuellement remplacées. Il existe une « mémoire cellulaire » suivant laquelle une cellule d'un certain type, garde tout au long de sa vie des caractéristiques qu'elle transmet à la cellule qui prend sa place lorsqu'elle meurt. Tous les types cellulaires de l'organisme ne sont pas aptes à se renouveler et à être remplacés.

La prolifération est donc nécessaire à la survie. Elle se fait à des vitesses et selon des mécanismes propres à chaque type cellulaire. Si la croissance cellulaire n'est plus contrôlée, les cellules alors capables de se diviser le font à un rythme anormalement élevé. Des cellules différenciées, normalement incapables de se multiplier, peuvent se dédifférencier et se multiplier activement. Il se développe alors une tumeur. On parle de cellules transformées. Leurs principales caractéristiques sont la poursuite de la prolifération dans des conditions où la croissance des cellules normales cesse et une diminution des besoins en facteurs de croissance (qui sont des substances stimulant la croissance et la prolifération). Elles ne peuvent pas arrêter leur croissance, même lorsque le milieu ne contient plus les éléments nécessaires (elles meurent alors en tentant de se diviser). Enfin elles n'ont plus besoin d'un

ancrage pour proliférer. Ces désordres ont une origine cellulaire, liée à une mutation : les cellules cessent d'obéir aux messages de freinage et se mettent à se reproduire en excès.

1.1.2 Progression tumorale

La progression tumorale désigne un phénomène anatomique ou biologique. La progression anatomique d'un cancer correspond à l'envahissement de l'organe, lieu de naissance du cancer, puis de tout l'organisme. Cette propagation se fait de proche en proche, en bordure de la tumeur vers les tissus voisins. Elle se produit aussi par détachement des cellules néoplasiques qui vont, à partir du foyer primitif, migrer, principalement par l'intermédiaire des vaisseaux et de la circulation sanguine, puis fonder des foyers secondaires ou métastases. D'autres tumeurs envoient, simultanément ou préférentiellement, des cellules cancéreuses dans les vaisseaux et ganglions lymphatiques. La progression biologique est un phénomène plus récemment reconnu. Elle se réalise, avant même l'apparition du caractère envahissant, proprement cancéreux, par des changements préalables à la modification décisive. Cette mutation déterminante à l'origine de la première cellule cancéreuse ne fige pas l'évolution tumorale, ce qui explique que, malgré son caractère monoclonal, une population cancéreuse qui descend d'une seule cellule présente une certaine hétérogénéité. D'autres modifications apparaissent, écartant de plus en plus les cellules de la normale, leur confèrent des avantages évolutifs allant dans le sens d'une plus grande malignité. De nouvelles propriétés permettent le développement de métastases, une tendance plus marquée à envahir les tissus normaux, la sécrétion de diverses substances (marqueurs), l'aptitude à résister à une influence thérapeutique (chimiorésistance). Inversement, les propriétés normales disparaissent comme la présence de récepteurs hormonaux, la sécrétion de substances caractéristiques des cellules d'origine, ou la capacité de réparer les lésions cellulaires induites par le traitement. Toutes ces altérations convergent pour constituer une prolifération de plus en plus maligne, se généralisant, évoluant rapidement et résistant aux traitements.

La transformation maligne de la cellule souche responsable de la tumeur survient en raison de l'accumulation de mutations dans certains de ses gènes. Ces gènes sont les clés des mécanismes responsables des cancers humains. Au fur et à mesure du développement du

cancer, les cellules tumorales acquièrent une instabilité génétique de plus en plus marquée, aboutissant à des anomalies génétiques de plus en plus nombreuses. Les tumeurs apparaissent quand plusieurs gènes participant au contrôle du cycle cellulaire sont mutés. Des anomalies d'autres classes de gènes participeraient également à l'apparition de la malignité en conférant un caractère invasif aux cellules tumorales. Capables de migrer, elles créent des métastases à distance de la tumeur initiale.

1.1.3 Les différentes étapes du développement d'un cancer

Beaucoup de cellules tumorales ont une malignité croissante avec le temps. Pour certaines tumeurs, on peut établir une progression ordonnée, depuis le tissu normal jusqu'au cancer invasif, en passant par des stades précancéreux comme la dysplasie, puis le carcinome *in situ*.

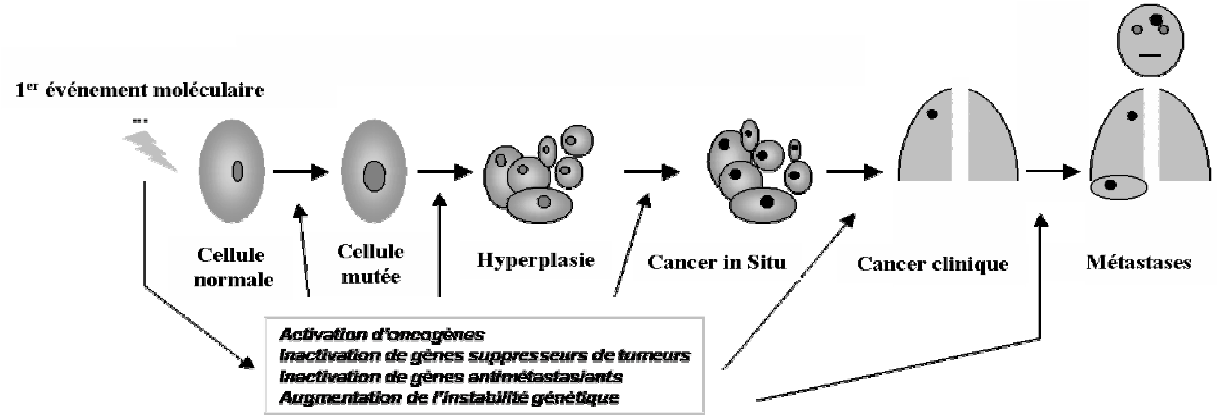


Figure 1.1: Les différentes étapes de la progression tumorale.

L'hyperplasie est la caractéristique d'un lot de cellules d'apparence normale qui présentent le défaut de se multiplier excessivement par rapport au tissu normal. Elle résulte de la mutation d'une seule cellule, suite à la surexpression d'un gène impliqué dans la prolifération cellulaire, ou à l'inactivation d'un gène suppresseur de tumeurs. L'anomalie peut en rester là ou être suivie d'une dysplasie : les cellules de la zone dysplasique ont des anomalies de forme et d'orientation. L'hyperplasie puis la dysplasie sont souvent qualifiées d'état précancéreux parce qu'un stade de plus doit être franchi, à l'occasion d'une mutation génétique supplémentaire, pour que les cellules deviennent franchement anormales. Dès lors, ces cellules forment une tumeur qui se développe localement, sans franchir les limites du tissu auquel elles appartiennent : on parle à ce stade d'un cancer *in situ*. Cette tumeur peut rester

confinée sur place ou peut s'étendre au-delà : elle devient alors un cancer invasif. Ce passage survient lorsque certaines cellules tumorales acquièrent la faculté de migration suite à une ultime mutation. Ces cellules migrent par voie sanguine ou lymphatique, vont plus ou moins rapidement former des métastases à distance **(**Figure 1.1).

La complexité des processus moléculaires impliqués dans le comportement cellulaire est un grand obstacle à la compréhension du processus de transformation tumorale. Dans les sections suivantes, nous allons présenter le cancer du colon et l'état actuel des recherches concernant le réseau moléculaire responsable de son déclanchement puis de sa propagation.

1.2 Anatomie générale du colon

Le colon est la partie de l'intestin qui fait suite à la portion jéjunale de l'intestin grêle et trouve sa terminaison au niveau de la naissance du rectum (Figure 1.2). Le colon mesure entre 1 m et 1,5 m et présente un trajet anatomique sous forme de plusieurs segments :

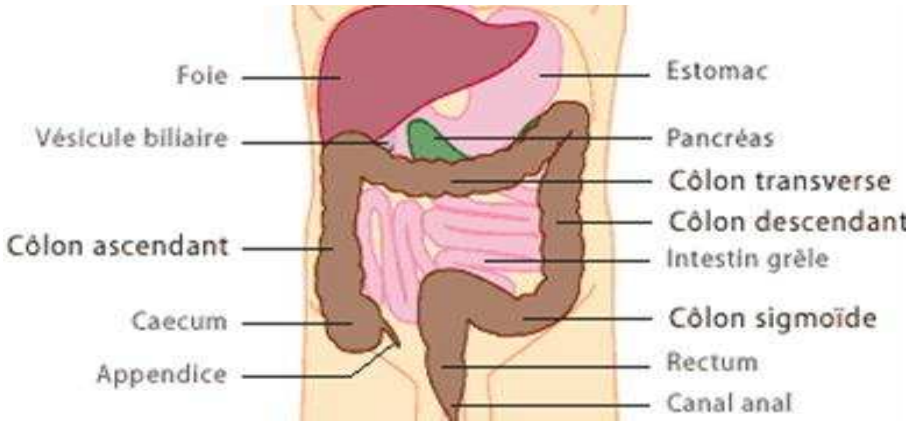


Figure 1.2: Anatomie générale du colon.

- le caecum, lieu d'implantation de l'appendice. C'est ce segment qui assure la jonction entre l'intestin grêle et le colon,
- le colon ascendant, ou colon droit, qui se termine par un angle sous forme de coude, au niveau du foie,
- le colon transverse, reliant l'angle droit hépatique à l'angle gauche splénique,
- le colon sigmoïde, qui se courbe de gauche vers la partie centrale basse de l'abdomen,

- le rectum, faisant 8 à 15 cm. Bien qu'il soit en réalité un organe à part, il peut aussi être considéré comme la partie terminale du colon.

Les cancers colorectaux se développent, par fréquence décroissante, dans le recto-sigmoïde (66%), puis dans le colon descendant (10%) et le caecum (10%). Le reste se répartit sur les autres segments du colon.

1.3 Histologie du colon

Le colon est constitué des structures histologiques suivantes (Figure 1.3) :

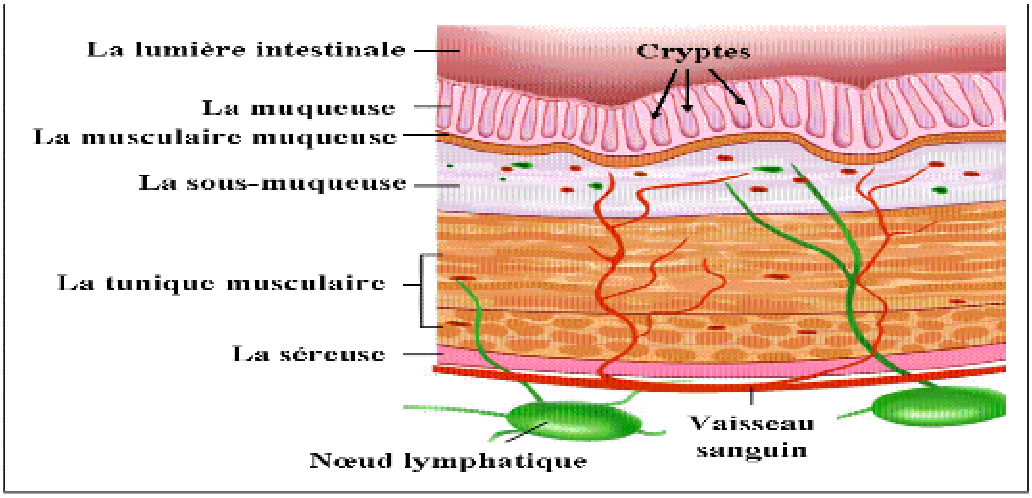


Figure 1.3 : Histologie du colon.

1.3.1 La muqueuse

La muqueuse colique est composée d'une couche épithéliale unique de cellules cylindriques. Elle borde la lumière colique et s'invagine pour former les cryptes (Figure 1.3). L'ensemble se pose sur une fine couche musculaire lisse, la musculaire muqueuse. Les cryptes s'ouvrent dans la lumière épithéliale le plus souvent par un abouchement unique. L'organisation de l'épithélium colique en cryptes se retrouve tout au long du colon. Cette disposition joue un rôle important dans le renouvellement de l'épithélium colique. Ce point sera abordé en détail par la suite.

1.3.2 La musculaire muqueuse

Elle sépare la muqueuse de la sous-muqueuse. Elle est constituée d'une mince couche de tissu musculaire lisse. Elle est traversée par les réseaux vasculaires, lymphatiques et nerveux.

1.3.3 La sous-muqueuse

Elle comporte globalement les mêmes constituants que la musculaire muqueuse. Mais, elle comporte quelques structures en plus, comme les cellules adipeuses. Deux plexus nerveux occupent la sous-muqueuse : le plexus sous-muqueux de Meissner, constitué de neurones et de cellules gliales dispersés dans les éléments de soutien, et le plexus profond sous-muqueux d'Henle.

1.3.4 La tunique musculaire propre

Elle est constituée par des muscles lisses organisés en une couche interne circulaire et une couche externe longitudinale, entre lesquelles se positionne le plexus nerveux d'Auerbach ayant une composition analogue à celle du plexus de Meissner. La tunique musculaire est irriguée par les vaisseaux sanguins et lymphatiques.

1.3.5 La séreuse

Elle est constituée par une fine couche de cellules mésothéliales, enveloppant le tube colique.

1.4 Organisation fonctionnelle de l'épithélium colique

Nous avons vu ci-dessus, que la muqueuse colique est organisée sous forme de cryptes, qui se présentent comme des structures ordonnées comportant un compartiment de cellules prolifératives, qui vont évoluer au cours du temps et de l'espace, en quittant le fond des cryptes pour migrer vers la surface tout en se différenciant. Les cellules de surface, complètement différenciées, sont confrontées au processus d'apoptose. Ce sont les cryptes qui constituent le compartiment de prolifération de l'intestin, grâce à leur population spécifique de cellules souches [Bach et al 2000] [Brittan et Wright 2002]. En effet, parmi les tissus

épithéliaux, la muqueuse du tube digestif est particulière par son renouvellement constant [Crosnier et al 2006].

La figure 1.4 représente le schéma d'organisation du renouvellement cellulaire des cryptes. Dans ce schéma, le fond de la crypte est peuplé par les cellules souches dévolues à la prolifération. La zone intermédiaire est occupée par des cellules toujours prolifératives, qui entament une phase de différenciation au fur et à mesure de leur migration vers la partie supérieure de la crypte. Les cellules de l'épithélium de surface, au niveau de la partie supérieure, sont définitivement différenciées. Elles ne se multiplient plus et ne se divisent plus. Par ce mécanisme, l'épithélium colique est renouvelé entièrement tous les 3 à 5 jours. C'est le temps nécessaire aux cellules prolifératives du fond des cryptes pour regagner la lumière intestinale [Crosnier et al 2006].

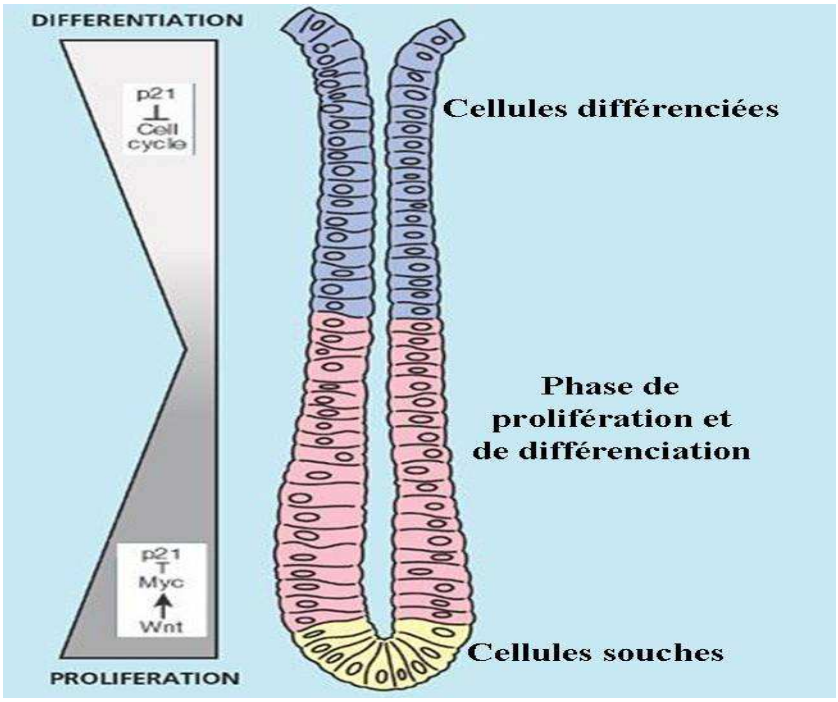


Figure 1.4 : Schéma du renouvellement cellulaire de l'épithélium colique.

A l'échelle moléculaire, les cellules souches reçoivent le signal mitotique via la voie de signalisation cellulaire Wnt. La voie Wnt, par une cascade de transduction du signal à l'intérieur des cellules, libère la ß-caténine qui va être ensuite véhiculée vers le noyau

cellulaire. La ß-caténine est une molécule clef de la prolifération cellulaire. Une fois présente dans le noyau, elle va s'associer aux facteurs de transcription TCF/LEF, modulant ainsi leur activité et activant (ou inactivant) l'expression d'un grand nombre de gènes cibles [Behrens et al 1996] [Korinek et al 1997]. c-Myc et les cyclines D sont parmi les premiers gènes qui répondent à une telle régulation [He et al 1998]. En conséquence, lorsque la voie de signalisation Wnt est activée, c-Myc est exprimé inactivant à son tour l'expression du gène p21 dont l'expression provoque l'arrêt du cycle cellulaire [Tetsu et McCormick 1999]. Le gène p21 n'exerçant plus aucun effet sur le cycle cellulaire, la division cellulaire peut s'amorcer dès lors, sous l'effet de la ß-caténine nucléaire.

Au contraire, le compartiment de différenciation des cryptes, se caractérise par l'absence des signaux Wnt. A ce niveau, le gène p21 est particulièrement actif. En absence du gène c-Myc, il donne un coup d'arrêt au cycle cellulaire, précipitant ainsi la différenciation cellulaire (Figure 1.4).

Il semble que ce mécanisme de régulation moléculaire, associé à l'expression de la voie de signalisation Wnt, joue un rôle majeur dans l'organisation spatiale de l'épithélium colique au niveau des cryptes. En effet, cette voie de signalisation est exprimée exclusivement au niveau de la partie basse des cryptes. Certaines études ont démontré que le blocage de la voie Wnt provoque l'arrêt de la prolifération cellulaire de l'épithélium intestinal. Dans ce cas, le fond des cryptes change de statut, devenant ainsi un compartiment de cellules épithéliales différenciées [Fre et al 2009].

La dérégulation de ce mécanisme peut conduire à des bouleversements tissulaires au niveau du colon. Le cancer colorectal trouve sans doute son origine dans la dérégulation de l'expression liée à une ou plusieurs molécules intervenant dans l'organisation fonctionnelle et spatiale de l'épithélium colique. A travers le cancer colorectal, nous aborderons la manifestation de cette dérégulation moléculaire à l'échelle tissulaire et macroscopique. Ensuite, nous présenterons l'état actuel de la recherche, concernant les principales altérations moléculaires, responsables du développement tumoral du cancer colorectal.

1.5 Principales étapes de la transformation tumorale du colon

Le renouvellement cellulaire de l'épithélium colique est nécessaire, en réponse à la perte cellulaire permanente au niveau de l'épithélium bordant la lumière intestinale. Ce perpétuel renouvellement cellulaire accroît les risques de mutations accidentelles à l'échelle cellulaire, avec ou sans manifestation tissulaire. Les plus mineures d'entre-elles donneront lieu à des lésions de dysplasie tandis que leur aggravation permettra la transformation maligne puis la progression tumorale. En parallèle, le colon subit des bouleversements tissulaires, allant des lésions les plus minimes (les cryptes aberrantes) au cancer invasif [Zukerberg 2002]. Nous allons tenter de décrire brièvement les différentes étapes de ces bouleversements à l'échelon histologique.

1.5.1 Etats précurseurs du cancer colorectal

1.5.1.1 Hyperplasie, dysplasie et néoplasie intra-épithélial

L'**hyperplasie** est une simple augmentation du volume du tissu par multiplication cellulaire. Elle ne s'accompagne pas d'altérations cellulaires.

La **dysplasie** associe un trouble de la prolifération cellulaire à des anomalies cellulaires. L'ensemble de ces lésions est commun à certaines proliférations bénignes et aux cancers. Contrairement à l'hyperplasie vraie, la dysplasie est un état précurseur du cancer. L'appellation « **néoplasie intra-épithéliale** » est introduite par l'organisation mondiale de la santé (OMS) [TNM 2003], pour mettre en accord les classifications japonaise et occidentale concernant le cancer colorectal [Schlemper 1998].

On distingue généralement trois degrés de dysplasie :

- la **dysplasie légère** ou **néoplasie intra-épithéliale de bas grade,** où la prolifération cellulaire reste modérée, les cellules restent différenciées,
- la **dysplasie moyenne** où la prolifération est active, les cellules indifférenciées occupent de larges zones,
- La dysplasie sévère ou **néoplasie intra-épithéliale de haut grade**, où la prolifération est très active avec un grand nombre de mitoses dans toutes les couches et sur toute la

hauteur des cryptes. La totalité de la lésion est indifférenciée. Il existe de nombreuses anomalies nucléo-cytoplasmiques.

Le stade ultime de la dysplasie sévère est le **carcinome *in situ*** (CIS), appelé encore **adénocarcinome intra-muqueux.** A ce stade, on assiste à une invasion de la membrane basale, soutenant l'épithélium colique, avec ou sans invasion de la musculaire muqueuse mais sans atteinte de la sous-muqueuse.

1.5.1.2 Les Cryptes Aberrantes

Alors que les dysplasies sont définies essentiellement à l'échelon cellulaire, les cryptes aberrantes sont les lésions définies par leurs bouleversements tissulaires.

La crypte aberrante est difficile à identifier, du fait de sa très petite taille. Elle n'a été individualisée que très récemment. Elle reste encore sous-diagnostiquée et méconnue par de nombreux pathologistes [Shpitz et al 1998]. C'est la première lésion adénomateuse identifiable. Elle est qualifiée parfois de micro-adénome [Roncucci et al 1991]. La crypte aberrante apparaît comme une anomalie de la surface colique, visible dès l'examen à la loupe. Elle est considérée, par plusieurs auteurs, comme un état précurseur initial de la transformation maligne [Nucci et al 1997] [Pretlow et al 1991] [Takayama et al 1998].

1.5.1.3 Les Adénomes

Nous avons souligné que la dysplasie et les cryptes aberrantes sont les deux facettes : cellulaire et tissulaire d'un stade précancéreux. Leur développement débouche sur la formation d'un adénome. L'adénome, est une tumeur néoplasique constituée par une prolifération épithéliale bénigne comportant des degrés variés de dysplasie. Les adénomes correspondent à des polypes, qui sont des excroissances faisant saillie dans la lumière colique.

Sur le plan étiologique, deux schémas différents de dérégulation du cycle cellulaire sont à l'origine de deux types de polypes :

- les polypes adénomateux de nature sporadique,

- les polypes adénomateux entrant dans le cadre des syndromes de polyposes héréditaires et familiales (gène APC).

Ces deux formes seront abordées plus loin à l'échelon moléculaire.

Les adénomes sont des lésions néoplasiques, c'est-à-dire qu'elles réalisent une croissance avec une perte du contrôle de la prolifération cellulaire. Les théories pathogénétiques actuelles suggèrent que les adénomes sont à l'origine de la majorité des cancers colorectaux [Kinzler et Vogelstein 1996]. On estime que 25 à 30% des adénomes se transforment en cancers.

1.5.2 De l'adénome vers le cancer colorectal

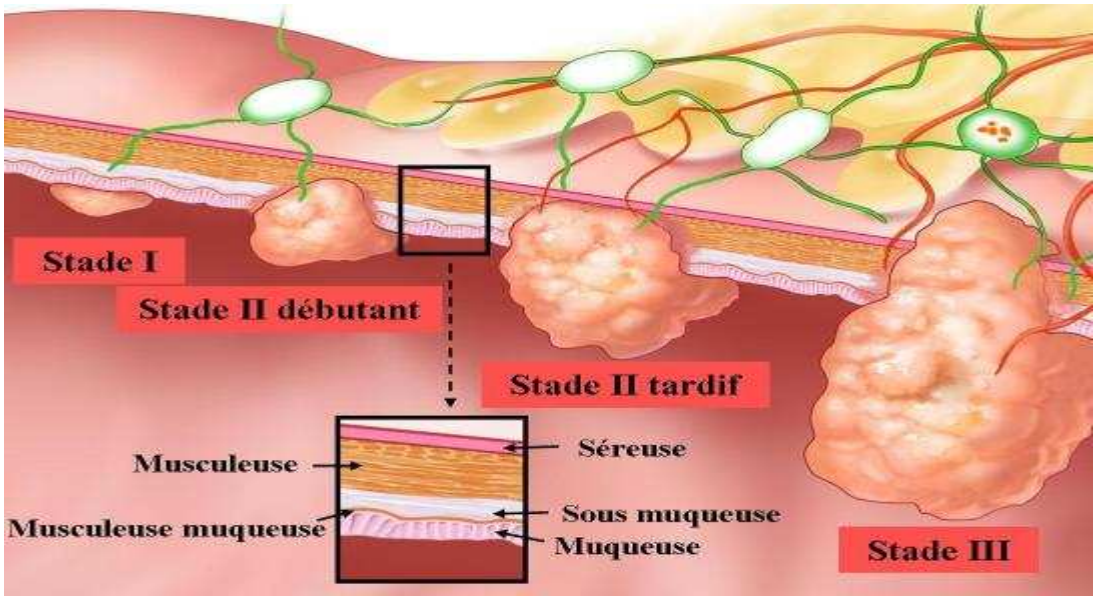


Figure 1.5 : Les différents stades de la transition adénome-cancer.

La séquence de transformation adénome-cancer a été décrite la première fois en 1971 [Potet et Souillard 1971] sur la base d'observations histo-pathologiques. Cette séquence a été reprise ensuite à l'échelle moléculaire par Vogelstein [Simony 2007]. La séquence schématisant le passage de l'adénome au cancer, sur le plan morphologique, est illustrée dans la Figure 1.5. Cette séquence de transition comporte plusieurs stades :

- **Stade I :** le cancer est encore limité par la musculaire muqueuse.

- **Stade II débutant :** À ce stade, le cancer effondre la musculaire muqueuse. Dans un premier temps le cancer atteint l'axe du polype ; il envahit ensuite sa partie distale implantée dans les structures histologiques du colon.
- **Stade II tardif :** Le cancer envahit totalement le polype puis infiltre les structures sous-jacentes de la muqueuse (musculeuse muqueuse et sous muqueuse). La musculeuse est attirée par la prolifération ; la surface peut s'ulcérer. A ce stade, le caractère malin de la lésion est évident dès l'examen macroscopique.
- **Stade III :** Le cancer envahit la musculeuse ; il peut rester végétant ou s'ulcérer plus ou moins en surface. On est alors en présence d'un cancer invasif classique qu'il est difficile de rattacher morphologiquement à l'adénome qui lui a donné naissance.

Le potentiel évolutif précancéreux des adénomes peut être estimé selon trois paramètres :

- **Degré de dysplasie** : Le taux de transformation maligne des adénomes s'accroît bien évidemment avec le degré de dysplasie épithéliale et son étendue dans l'adénome. Un adénome présentant des lésions de dysplasie légère peut ne pas évoluer pendant des années. En revanche, il en est tout autrement dès que la dysplasie s'accentue. On comprend donc qu'il est capital pour le pathologiste de s'attacher à rechercher un ou plusieurs foyers de cancérisation, et d'effectuer une analyse très fine de l'axe du polype, et de son pied d'implantation en muqueuse présumée saine, dont l'envahissement signe le passage à un degré de malignité supplémentaire.
- **Type histologique** : Il existe deux formes histologiques d'adénomes :
    - l'**adénome tubuleux,** où la prolifération épithéliale se fait de la surface vers la profondeur. Ce type d'adénome évolue rarement [Potet et Souillard 1971],
    - l'**adénome villeux,** où la prolifération se fait vers la lumière intestinale. Cette forme évolue vers un cancer dans 30% des cas.
- **Taille** : C'est le principal paramètre qui détermine le risque de la transformation maligne de l'adénome. Le risque de transformation cancéreuse d'un adénome augmente avec sa taille : de 0.5% pour une taille inférieure à 1cm, à 5% si la taille se situe entre 1 et 2 cm. Le risque passe à 30% pour les adénomes mesurant plus de 2cm.

1.5.3 Les cancers invasifs ou adénocarcinomes

Le stade ultime de la transformation maligne d'un adénome est un cancer invasif, appelé aussi « adénocarcinome ». Lorsque le cancer colorectal déborde les structures histologiques formant la paroi colique, il est qualifié d'invasif. Les cancers invasifs, dans leur progression, franchissent et détruisent les limites de l'organe qui leur a donné naissance.

Dans leur première phase d'extension, par voie lymphatique ou sanguine, les cancers se propagent en premier lieu vers les ganglions régionaux. Les métastases se forment principalement dans les ganglions (50%), ensuite dans le foie (25%), puis dans les autres organes, mais à moindre degré. En complément aux stades de progression tumorale, évoqués ci-dessus, la présence des métastases correspond au **stade IV**.

Le pronostic est directement dépendant de l'extension de la tumeur, locale, régionale ou générale. La nécessité d'adopter des classifications ayant pour but de faire entrer les tumeurs, donc les patients, dans des groupes homogènes, s'impose pour une prise en charge optimale des patients. De multiples systèmes de classification du cancer colorectal existent, pour estimer ce pronostic. Nous allons présenter dans les paragraphes suivants le système TNM établi par l'Union Internationale Contre le Cancer [TNM 2003], qui est le système de référence pour mettre en œuvre les protocoles thérapeutiques spécifiques.

1.6 Le système de classification TNM

La classification TNM s'appuie sur trois paramètres :

- T : degré d'extension de la tumeur primitive,
- N : degré d'envahissement des ganglions lymphatiques régionaux,
- M : ce paramètre nous renseigne, sur l'absence ou la présence de métastases à distance.

Chacun des trois paramètres est assorti d'un indice. Les différents cas de figure associés à ce système de codification sont les suivants :

- **T : tumeur primitive** :
  - Tx : renseignements insuffisants pour classer la tumeur primitive,

- T0 : pas de signe de tumeur primitive,
- Tis : carcinome *in situ* : intraépithélial, ne dépassant pas la membrane basale,
- T1 : envahissement de la sous-muqueuse,
- T2 : envahissement de la musculeuse,
- T3 : envahissement de la sous-séreuse et des tissus péricoliques, périrectaux et extrapéritonéaux,
- T4 : envahissement des organes ou structures de voisinage.

- **N : adénopathies régionales :**
  - Nx : renseignements insuffisants pour classer les adénopathies régionales,
  - N0 : pas d'adénopathie régionale métastatique,
  - N1 : 1 à 3 adénopathies régionales métastatiques,
  - N2 : 4 ou plus adénopathies régionales métastatiques.

- **M : métastases à distance :**
  - Mx : renseignements insuffisants pour affirmer la présence d'une métastase,
  - M0 : absence de métastase,
  - M1 : présence d'au moins une métastase.

La classification TNM est une évaluation clinique lors du diagnostic du cancer colorectal. Elle est réalisée pour proposer un protocole thérapeutique. Une seconde classification en post-opératoire existe. Elle se base sur des observations anatomopathologiques. Celle-ci permet d'apporter la preuve sur la nature exacte de l'état de propagation de la tumeur. Elle permet en outre l'affinage du protocole thérapeutique. On note par pTNM la classification post-opératoire. Tous les paramètres cités ci-dessus sont précédés de la lettre p lorsqu'il s'agit de ce type de classification.

En complément, la classification pTNM prend également en compte l'évaluation de la tumeur résiduelle (R) sur la pièce opératoire après un traitement de première intention :

- R0 : berges de résection microscopiquement indemnes de la tumeur résiduelle,
- R1 : berges de résection microscopiquement envahies,

- R2 : tumeur résiduelle macroscopique (constatation chirurgicale).

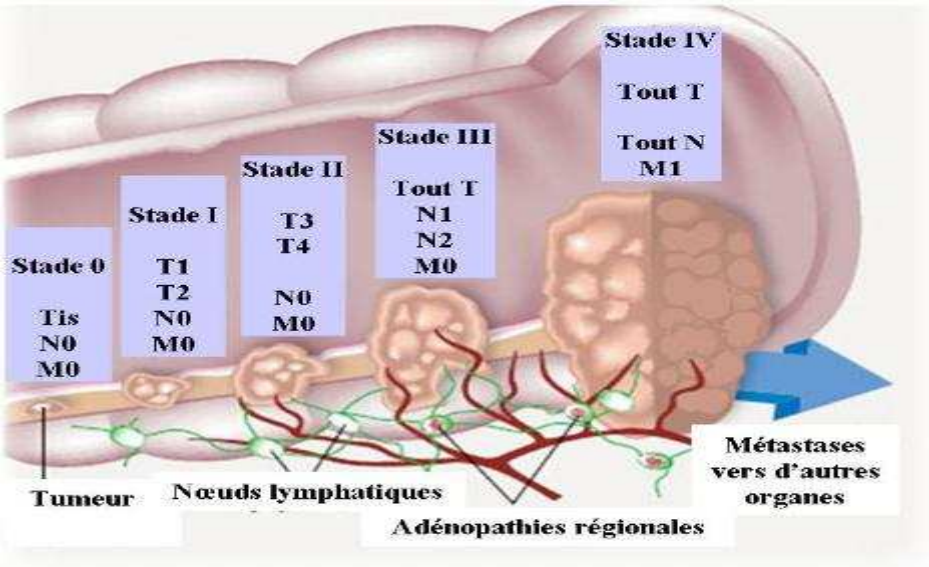


Figure 1.6 : Correspondance entre la classification TNM et le stade des cancers colorectaux.

La correspondance entre la classification TNM et les différents stades du cancer colorectal, est donnée comme suit (Figure 1.6) :

- Stade 0 : Tis N0 M0,
- Stade I : T1N0 M0, T2 N0 M0,
- Stade IIA : T3 N0 M0,
- Stade IIB : T4 N0 M0,
- Stade IIIA :T1 N1 M0, T2 N1 M0,
- Stade IIIC : T3 N1 M0, T4 N1 et tous T N2 M0,
- Stade IV : Tous T tous N M1.

1.7 Cancérogenèse colorectale

Le cancer colorectal résulte d'un déséquilibre au niveau du réseau moléculaire contrôlant la prolifération cellulaire. Il est provoqué par la mutation d'un ou plusieurs gènes appartenant à ce réseau. Les mécanismes qui aboutissent à la formation, puis au développement du cancer colorectal, sont aujourd'hui relativement bien étudiés. Afin d'illustrer les événements moléculaires accompagnant la transformation de l'épithélium colique normal vers un cancer,

nous allons en premier lieu, présenter les mécanismes moléculaires régissant la prolifération cellulaire normale. Ensuite nous allons tenter d'expliquer, les principales altérations moléculaires, perturbant cette prolifération cellulaire et conduisant vers le développement du cancer colorectal.

1.7.1 Réseau moléculaire de la prolifération cellulaire

Les cryptes coliques sont le siège d'une prolifération cellulaire active, rendue nécessaire par la perte fréquente des cellules épithéliales au contact de la lumière intestinale. Cette prolifération est contrôlée par un réseau complexe de molécules. Sous l'influence de la voie de signalisation cellulaire Wnt, ce réseau moléculaire subit des modifications majeures, précipitant la cellule dans la voie de mitose [L'Allemain 2006]. La Figure 1.7 illustre la cascade des évènements moléculaires, induite par les deux états de la voie de signalisation Wnt : ON et OFF.

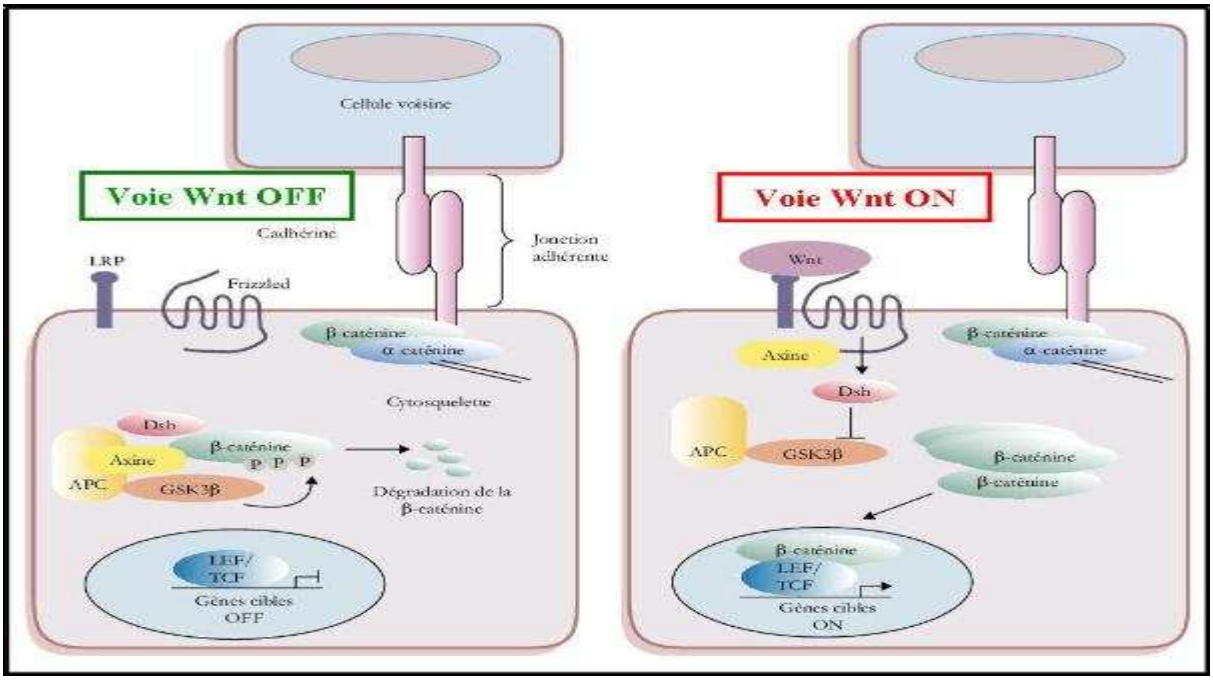


Figure 1.7: Schéma de régulation de la β-caténine par la voie Wnt [L'Allemain 2006].

La protéine β-caténine est la clef de voûte de la voie de signalisation Wnt. En absence du signal mitotique Wnt : « voie Wnt OFF », c'est-à-dire en phase de repos cellulaire, elle participe dans la régulation de l'adhérence cellulaire par son interaction avec la protéine

transmembranaire : E-cadhérine. Une autre fraction de la β-caténine, peut se trouver dans le cytoplasme. A ce niveau, par l'intermédiaire de sa partie N-terminale, la β-caténine est prise en charge par un complexe multi protéique, formé par l'Axine, le suppresseur de tumeurs APC (Adenomatose Plyposis Coli) et la GSK3β (Glycogène synthase kinase 3β). Au sein de ce complexe, l'interaction entre l'Axine et la GSK3β favorise la phosphorylation de la β-caténine, action qui aboutit à la dégradation de la β-caténine par protéolyse [Aberle et al 1997]. Le complexe APC-axine-GSK3β régule donc l'activité de la β-caténine. Il empêche son accumulation, sous sa forme active : non phosphorylée, dans le cytoplasme.

Dans le processus normal de prolifération cellulaire, le signal mitotique est véhiculé par la voie de signalisation Wnt : « voie Wnt ON » (Figure 1.7). Les Wnt sont des glycoprotéines sécrétées, qui se fixent sur et activent les récepteurs Frizzled, qui à leur tour participent à la phosphorylation de la protéine Dsh (Dishevelled). Cette dernière est un régulateur de la voie Wnt. Une fois phosphorylée, elle s'associe à l'axine, empêchant de ce fait la capture de la β-caténine cytoplasmique par le complexe APC-axine-GSK3β. La β-caténine ainsi libérée dans le cytoplasme peut exercer ses fonctions [Bellaïche et Perrimon 1997]. Elle migre vers le noyau où elle s'associe aux facteurs de transcription TCF-LEF (T-cell factors) en exerçant le rôle d'un co-activateur de la transcription génétique [Behrens et al 1996]. Par ce mécanisme, la β-caténine active un grand nombre de gènes du cycle cellulaire [Korinek et al 1997], parmi eux on peut citer c-Myc ou les Cycline D, particulièrement actifs pendant les phases de prolifération cellulaire [He et al 1998].

1.7.2 Les altérations moléculaires associées au cancer colorectal

La découverte de la relation fonctionnelle entre le gène APC avec la ß-caténine a constitué un pas décisif dans l'identification de la voie moléculaire associée à la cancérogénèse colique [Miyashiro et al 1995]. Ainsi, toute altération génétique libérant la β-caténine cytoplasmique, peut conduire à une activation effrénée de la prolifération cellulaire. Les mutations touchant les molécules antagonistes de la β-caténine peuvent conduire à la formation de tumeur. La perte de l'expression d'APC est relevée dans 70 à 80% des cancers colorectaux chez l'homme [Bienz et Clevers 2000]. L'APC est le premier gène isolé responsable de la cancérogénèse colorectale. Il a été mis en cause à l'occasion de l'étude des polyposes adénomateuses

familiales (PAF) [Simony 2007]. Les patients atteints de cette forme de cancer présentent une mutation du gène APC sur un seul allèle. La PAF est normalement un type de cancer bénin, mais elle peut évoluer vers un cancer invasif si le deuxième allèle du gène APC est muté également. La mutation des deux allèles de l'APC peut s'observer aussi dans les formes sporadiques du cancer colorectal. On comprend dès lors pourquoi les personnes ayant hérité un gène APC muté sont particulièrement suivies. En effet, ces personnes développent plus facilement un cancer colorectal par rapport aux autres personnes. Ils ont moins d'étapes (de dérégulation moléculaire) à franchir pour aboutir à une transformation maligne.

Les mutations affectant la partie N-terminale de la β-caténine provoquent les mêmes effets de la mutation du gène APC [Harada et al 1999]. Ces altérations sont relevées dans 10 à 15% des cas de cancers colorectaux [L'Allemain 2006]. Ceci prouve que la cancérogénèse est associée en premier lieu à la régulation de la β-caténine cytoplasmique. L'APC est un antagoniste de la β-caténine [Rubinfeld et al 1993]. C'est un gène suppresseur de tumeur. C'est sa mutation, donc son inactivation, qui est un évènement précoce de la cancérogénèse colorectale.

Si des doutes demeurent quant à l'ordre chronologique de survenue des mutations, l'altération du gène APC est souvent décrite comme initiatrice de la cancérogenèse : une fois muté, il déclencherait la prolifération excessive des cellules, ce qui marquerait le début du développement tumoral [Fearon et Vogelstein 1990]. Cependant, la promotion du statut tumoral du cancer colorectal, d'un simple adénome vers une forme tumorale plus sévère, nécessite l'accumulation d'autres altérations génétiques. Ainsi, la mutation de l'oncogène *k-ras,* acteur de la prolifération cellulaire, est observée dans 50 % des cancers coliques [Pretlow et al 1993]. Le gène suppresseur de tumeurs p53, acteur clé de l'apoptose, est muté dans 70 % des cancers du colon [Brunz et al 1999]. Vogelstein associe les mutations connues dans la cancérogénèse colorectale avec la progression tumorale du cancer colorectal (Figure 1.8) [Simony 2007].

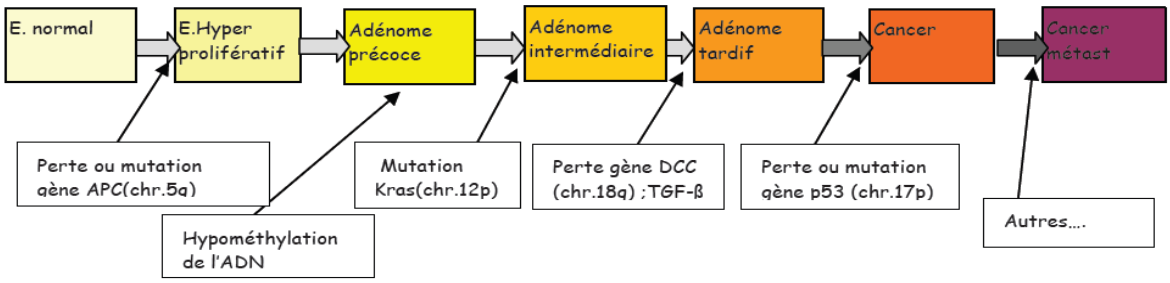


Figure 1.8 : Schéma de cancérogénèse colorectale de Vogelstein.

Même si cette séquence d'évènements génétiques est vérifiée dans la plupart des cancers colorectaux, des exceptions à cette hypothèse peuvent exister. Ainsi, des mutations de K-ras et de P53 sont relevées, dans des cas de cancers colorectaux sans qu'aucune mutation d'APC ne soit détectée [Smith et al 2002]. La séquence d'évènements peut également débuter par une altération de la β-caténine à la place de la mutation APC [Harada et al 1999].

1.8 Conclusion

Au cours de ce chapitre, nous avons abordé brièvement les principales étapes histologiques de la transformation du tissu colique normal vers un cancer colorectal. Nous avons décrit également les altérations génétiques majeures associées à ce type de cancer. Nous avons surtout remarqué qu'il existe plusieurs scénarios de mutations génétiques pouvant expliquer le développement d'une tumeur colorectale.

Si la plupart des gènes impliqués dans le cancer colorectal sont bien connus, la distribution spatiale de leurs profils d'expression à l'échelle tissulaire reste encore peu prise en compte [Gerike et al 1998] [Park et al 1995] [Smallwood et al 2004]. De telles informations peuvent expliquer pourquoi certains adénomes se transforment en cancers invasifs et pas d'autres bien qu'ils aient le même statut génétique.

L'étude des profils d'expression tissulaire de toutes les molécules impliquées dans le cancer colorectal nécessite la réalisation d'études IHC sur un nombre considérable de coupes histologiques. De telles études sont fastidieuses, consommatrices de ressources importantes de temps et d'argent. La technologie des puces à tissus ou des TMA « Tissue Microarrays »,

se montre une alternative séduisante, aux approches traditionnelles d'IHC. Elle permet l'analyse des profils d'expression de dizaines voire de centaines de tissus à la fois. Le chapitre suivant sera consacré à la description de cette technologie et sa pertinence pour étudier le profil tissulaire du cancer colorectal.

Chapitre 2

La technologie des TMA : Tissue MicroArrays

La progression tumorale, associée au cancer colorectal, est un phénomène complexe faisant intervenir plusieurs altérations génétiques et moléculaires. La mise en évidence du réseau moléculaire, responsable de la transformation tumorale d'un tissu sain, nécessite des études immunohistochimiques à grande échelle, sur un grand nombre de tissus.

La dernière décennie a connu l'émergence d'un nouveau genre de technologies dans le domaine de la biologie moléculaire in situ, telle que la technologie des puces à tissus ou TMA, permettant de regrouper plusieurs centaines de petits échantillons dans un seul bloc de paraffine. Cette technologie se situe à l'interface entre pathologistes et biologistes moléculaires, notamment pour la validation de nouveaux marqueurs tumoraux comme indicateurs de pronostic ou de diagnostic.

Le recours à une technologie à haut-débit, telle que la technologie TMA, a créé un besoin, concernant le choix des échantillons à inclure dans le bloc TMA, l'analyse et la gestion de la masse de données générée par cette technologie.

Ce chapitre sera consacré à la description de la technologie des TMA, notamment son principe et ses principaux domaines d'application. Nous allons évoquer également les problèmes techniques et scientifiques que soulève l'utilisation d'une micro-technologie à haut-débit, telle que la technologie TMA.

2.1 Introduction

A travers le chapitre précédent, nous avons vu que la transformation tumorale du tissu colique sain est causée à la base par une altération génétique, même si plusieurs molécules ont été identifiées comme responsables de la formation ou la progression du cancer colorectal. Néanmoins, la séquence exacte des évènements moléculaires menant d'une tumeur bénigne vers un cancer invasif est toujours matière à plusieurs hypothèses. La connaissance du pool de gènes, altérés à chaque étape de la progression tumorale, est un préalable indispensable pour développer des stratégies thérapeutiques qui soient adaptées au statut génétique des tumeurs.

La classification actuelle « TNM » [TNM 2006], basée sur des critères histologiques pour définir le grade des tumeurs, pose parfois un réel problème pour établir le protocole thérapeutique. Par exemple, le choix thérapeutique des patients atteints du cancer colorectal de stade II, est actuellement très controversé. Dans ce contexte, la classification génétique, donc moléculaire des tumeurs colorectales, apparaît comme un moyen prometteur pour l'organisation des patients en groupes homogènes d'un point de vue du suivi thérapeutique. Une telle classification nécessite la cartographie complète de l'expression génétique du cancer colorectal au cours des différentes phases de sa progression.

La mise en œuvre de la carte génétique du cancer colorectal nécessite l'analyse moléculaire de l'expression d'un nombre considérable de gènes, sur des centaines voire des milliers de cas de cancers. Les méthodes classiques d'analyse moléculaire, telle que le « *northern-blot* », sont inadaptées pour un tel projet. Elles permettent seulement l'analyse de quelques gènes à la fois. Cependant, les deux dernières décennies ont vu l'émergence d'un nouveau genre de technologies, qui ont révolutionné le domaine de l'analyse moléculaire. Parmi ces technologies, on peut citer les puces à ADN « cDNA Arrays », qui permettent l'analyse simultanée de plusieurs milliers de gènes. Dans ce contexte, ces technologies haut-débit peuvent prendre le relais pour l'analyse de l'expression de ces gènes. Ainsi, ces technologies permettent d'appréhender à partir d'une seule analyse le statut génomique et moléculaire d'une tumeur. Elles permettent en effet de mettre en lumière les mécanismes moléculaires

régulant la biologie tumorale spécifique à un cancer. Cependant, elles ne prennent pas en compte la distribution spatiale de l'expression des produits de ces gènes, sur le plan cellulaire et tissulaire. Les profils d'expression sont en effet extraits de façon globale sur le matériel biologique. Or, en plus de la région tumorale, le matériel prélevé pour l'analyse peut contenir du tissu normal ou des foyers de nécrose, ce qui peut représenter des artéfacts menant à des interprétations erronées. Toutes ces raisons ont conduit à développer, au service des pathologistes, un système haut-débit, appelé les puces à tissus ou TMA « Tissue MicroArrays », pour l'analyse des profils d'expression protéique des gènes à l'échelon tissulaire. Le TMA apparaît au bout de cette chaîne haut-débit, pour la validation rapide de nouvelles molécules, ayant un impact pronostic dans la progression tumorale, facilitant ainsi les choix thérapeutiques.

Ce chapitre sera consacré à la technologie des TMA. Nous allons aborder en premier lieu son principe et sa méthode de construction. Ensuite, nous exposerons les principaux domaines d'applications liés à cette technologie. Puis, nous présenterons quelques exemples de systèmes dédiés à la gestion des données générées par les TMA. Enfin nous discuterons la valeur des résultats obtenus par cette technique.

2.2 Principe des TMA

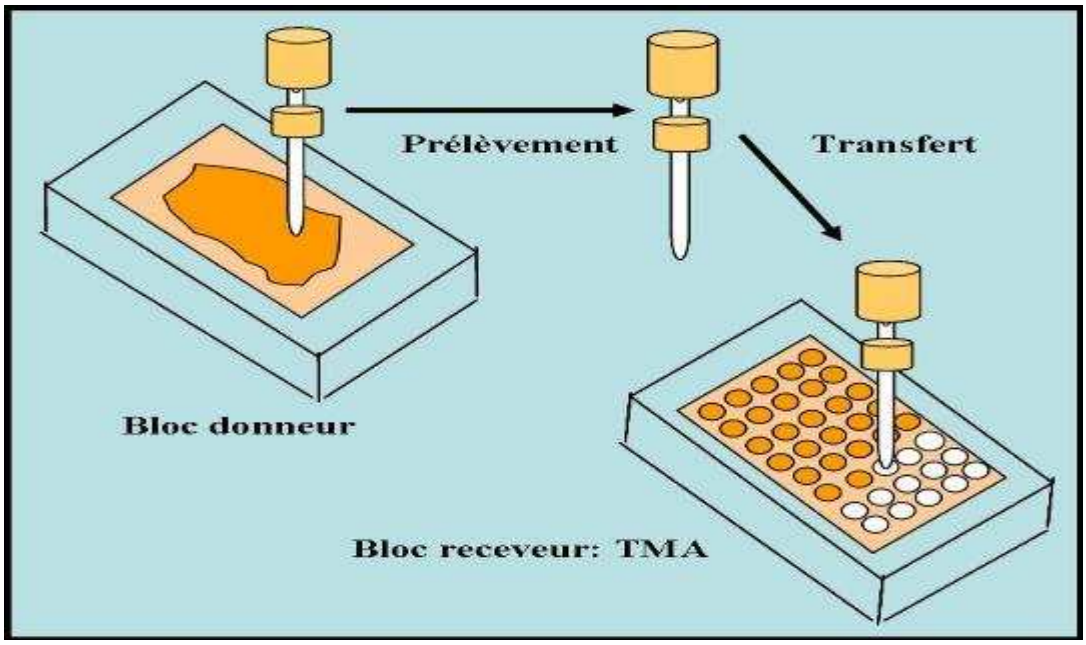


Figure 2.1 : Principe de la technologie TMA.

La technique des TMA est un outil indispensable pour la validation à grande échelle, des nouvelles découvertes concernant les altérations des gènes. Cette technique permet de prendre en compte aussi les informations relatives aux taux d'expression et de localisation tissulaire des produits d'expression de ces gènes.

La technologie des TMA s'inscrit dans le cadre des biotechnologies qui se développent actuellement dans le domaine de la biologie moléculaire *in situ*. Elle consiste à collecter en un seul bloc d'inclusion en paraffine des centaines d'échantillons tissulaires, sous forme de carottes cylindriques prélevées à partir de plusieurs blocs de paraffine classiques (Figure 2.1). Elle permet de réaliser ainsi des profils d'expression moléculaire à grande échelle. Le recours à des petits échantillons, qui varient entre 600µm et 2mm, permet la préservation des blocs donneurs, qui sont souvent des ressources précieuses. De ce fait, les blocs donneurs peuvent être ultérieurement intégrés dans de nouvelles études.

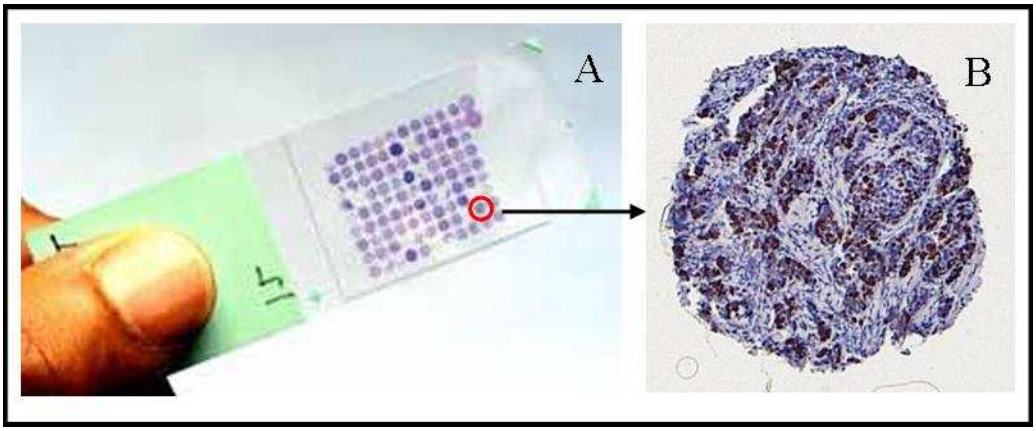


Figure 2.2 : Coupe histologique d'un bloc TMA. A : lame histologique d'un bloc TMA**. B :** image d'un seul spot (coupe d'une seule carotte).

Le bloc TMA, peut être utilisé comme n'importe quel bloc d'archive classique. Des coupes histologiques peuvent être pratiquées pour monter des lames de microscope comprenant plusieurs dizaines à plusieurs centaines de coupes de tissus (Figure 2.2). Ainsi, l'utilisation des méthodes d'IHC sur les coupes de TMA rend envisageable les études sur de très grands nombres de tissus en même temps. Cette approche permet de s'affranchir des facteurs de variabilité inter-lames. Tous les échantillons de la même lame TMA, subissent les mêmes conditions pendant les différentes étapes de coloration, telles que les concentrations des réactifs, les températures et les durées d'incubations dans les différents bains. On peut même introduire des échantillons de références sur la lame, pour le contrôle de qualité des traitements IHC (contrôles positifs et négatifs).

Les protocoles classiques d'IHC s'appliquent sans modification sur les lames de TMA. Il ne faut changer ni la concentration ni le volume des produits déposés par lame. Dès lors, les économies de réactifs sont considérables et directement proportionnelles au nombre de tissus par lame. Des études à grande échelle deviennent réalisables avec des budgets plus restreints et un gain important de temps.

La technique des TMA s'adresse particulièrement aux pathologistes qui disposent de banques tissulaires archivées. L'analyse par TMA de plusieurs centaines de tumeurs représentatives des différents stades tumoraux et des différents grades histologiques permet un gain de temps

et une économie de ressources financières et tissulaires, importants par rapport à l'analyse des blocs de tumeurs classiques.



Figure 2.3 : Le TMA Micro Arrayer (BeecherIstrumentsTM) : instrument pour réaliser des blocs TMA.

La méthodologie de rassembler plusieurs échantillons dans un seul bloc de paraffine remonte à 1986 [Bhattufira 1986]. Mais c'est en 1998 que Kononen et al ont publié le premier article décrivant la technique des TMA, avec un instrument dédié à la construction des blocs TMA [Kononen et al 1998] (Figure 2.3). Depuis, la technologie des TMA a connu un essor considérable. Actuellement les articles décrivant l'utilisation de cette technique se comptent par plusieurs centaines [Debault et Tubbs 2007]. La technologie TMA a été utilisée en particulier dans les schémas de validation des marqueurs tumoraux [Srivastava et al 2001] [Skacel et al 2002] [Mousses et al 2002] [Sugita et al 2002] [Bertucci et al 2004].

2.3 Plan de construction des TMA

2.3.1 Elaboration du plan de construction des TMA

La construction des TMA est simple et rapide. Cependant, la phase précédant cette construction est longue et fastidieuse. Elle nécessite une réelle réflexion, soigneusement motivée (expression des besoins) et guidée pour la spécification du protocole de construction du bloc TMA. Cette phase préliminaire comporte donc plusieurs étapes :

- formulation d'un objectif : tel que l'étude de l'expression de la protéine ß-caténine au cours de la progression tumorale,

- choix des blocs d’archives donneurs à inclure pour la confection du bloc TMA. Ces archives doivent être suffisamment documentées (ex. données clinique des patients et le grade de la tumeur),
- élaboration d’un plan détaillé, concernant le nombre et l’emplacement des carottes à prélever au niveau de chaque bloc donneur, puis leurs emplacements futurs sur le bloc TMA.

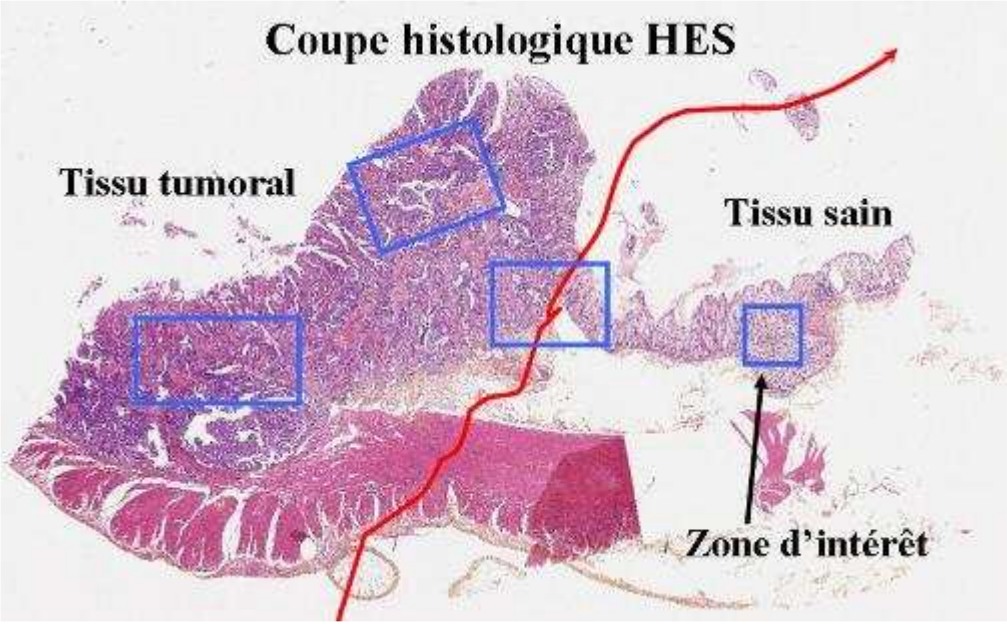


Figure 2.4 : Annotation anatomopathologique d'une coupe histologique HES.

Une fois le plan établi et avant de procéder à la construction du bloc TMA, une coupe de 4µm est pratiquée au niveau de chaque bloc d’archive donneur. A l’aide d’une coloration standard, HES (Hématoxyline-Eosine-Safran), le pathologiste observe les structures tissulaires des blocs donneurs. C’est une étape nécessaire qui permet d’une part le contrôle de la qualité des blocs recrutés, et d’autre part la délimitation des zones d’intérêts, où le technicien peut après réaliser le carottage.

La Figure 2.4 illustre un exemple d’annotation d’une coupe HES par le pathologiste. On peut observer 4 types de zones d’intérêts délimitées, deux zones tumorales, une zone normale et une zone à cheval entre le tissu sain et le tissu tumoral. Une telle annotation est adoptée dans le cas d’études s’intéressant à la variation de l’expression de certains marqueurs tumoraux

entre les zones de prolifération tumorales et les zones tissulaires saines en passant par toutes les zones intermédiaires.

Dès lors que les zones de carottage sont déterminées ainsi que le nombre de carottes à prélever au niveau de chaque zone, la construction du bloc TMA peut débuter. La construction du TMA est facilitée par l'élaboration d'un tableau présentant les entrées suivantes : identification du bloc donneur, identification de la zone d'intérêt dans ce bloc, position dans la matrice de carottes (colonne-ligne : A1, A2, C3…) et coordonnées géométriques (X, Y) des carottes exprimées en microns. Ce tableau facilite beaucoup le travail incombé aux techniciens chargés de la fabrication du bloc TMA.

2.3.2 Confection du bloc TMA

L'appareil permettant de construire des blocs TMA, le « MicroArrayer » est fabriqué par BeecherIstruments (www.beecherinstruments.com) et il est commercialisé par Alphelys (www.alphelys.com). Plusieurs modèles de cet appareil existent, nous disposons au niveau de notre laboratoire d'un modèle manuel (Figure 2.5). Celui-ci est constitué d'une tourelle comportant 2 aiguilles, ayant un diamètre variant entre 0,6 et 2 mm de diamètre. Le Micro Arrayer Manuel est mobile selon deux axes équipés de micromètres à affichage digital pour définir les points de carottage avec une précision de 1 µm. Il se compose d'un carrousel à 4 positions qui permet la construction en parallèle et simultanément de 4 blocs TMA. Ce carrousel facilite les opérations et réduit les manipulations de blocs. Il permet aussi de maintenir la précision de positionnement sur l'ensemble des blocs même avec la rotation.

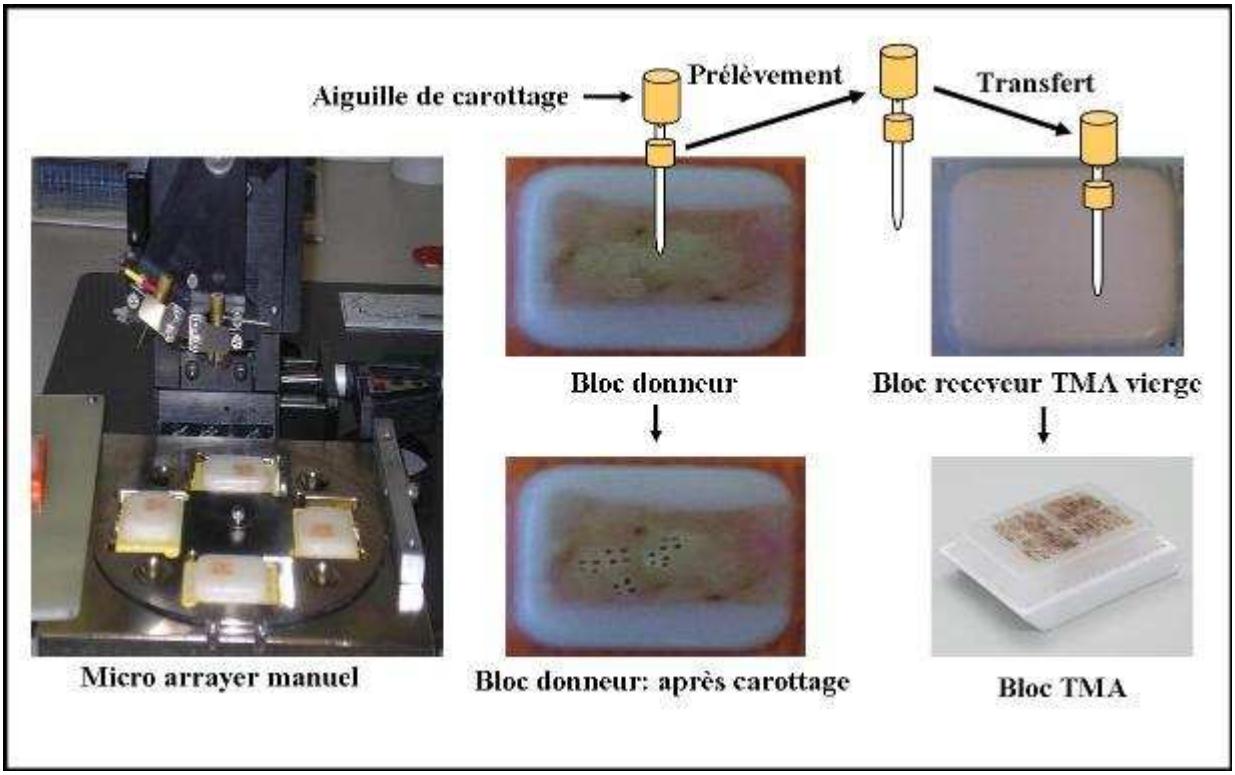


Figure 2.5 : Construction d'un bloc TMA. Les carottes sont prélevées à l'aide de l'aiguille du microArrayer, à partir des blocs de biopsies donneurs. Ensuite ces carottes sont transférées vers le bloc de TMA receveur.

De manière générale, la confection du bloc TMA est réalisée suivant 3 étapes :

- **Préparation du bloc receveur :** des trous sont pratiqués dans le bloc à l'aide de la première aiguille du MicroArrayer. Les trous sont pratiqués généralement au fur et à mesure de l'avancement du plan expérimental et selon les coordonnées calculées et définies par celui-ci.
- **Prélèvement des carottes des blocs de biopsies donneurs (**Figure 2.5**)** : la deuxième aiguille est enfoncée dans les blocs donneurs, au niveau des zones de carottages définies auparavant par le pathologiste. Puis l'aiguille est retirée prélevant ainsi une carotte ayant le même diamètre que l'aiguille. Soulignons que 600 µm est le diamètre le plus utilisé pour la réalisation des carottes. Des carottes de diamètres plus importants peuvent être prélevées, mais celui-ci réduit le nombre de prélèvements par bloc de biopsie donneur et surtout endommage plus rapidement ces blocs donneurs.
- **Transfert des carottes dans le bloc TMA receveur (**Figure 2.5**) :** l'aiguille comportant la carotte prélevée est amenée à proximité de la surface du bloc TMA

receveur, directement au-dessus du trou préalablement percé. Puis, la carotte est poussée hors de l'aiguille et transférée dans le trou du bloc receveur. L'opération est répétée jusqu'au transfert total des carottes.

Une fois fabriqué, le bloc TMA est incubé à 37°C pendant 10 à 15 minutes. Cela permet de réchauffer la paraffine afin d'établir une bonne cohésion entre les carottes et la paraffine du bloc receveur. Cette température ne devra pas dépasser une plage de valeurs prédéfinies afin de ne pas détériorer les protéines des tissus à étudier.

2.3.3 Réalisation des coupes TMA

Le bloc TMA est considéré comme un bloc d'archive classique, où des coupes fines de quelques microns d'épaisseur peuvent être pratiquées pour être ensuite montées sur des lames procurant une bonne adhésion de type « Super Frost ». Néanmoins, il faut prendre quelques précautions en manipulant les TMA. Le TMA est un bloc hétérogène formé de tissus et de paraffine non intimement cohésifs. La coupe de tels blocs entraîne fréquemment des plis voire des pertes de carottes rendant l'exploitation des lames difficiles. Des pertes de l'ordre de 50% peuvent être observées.

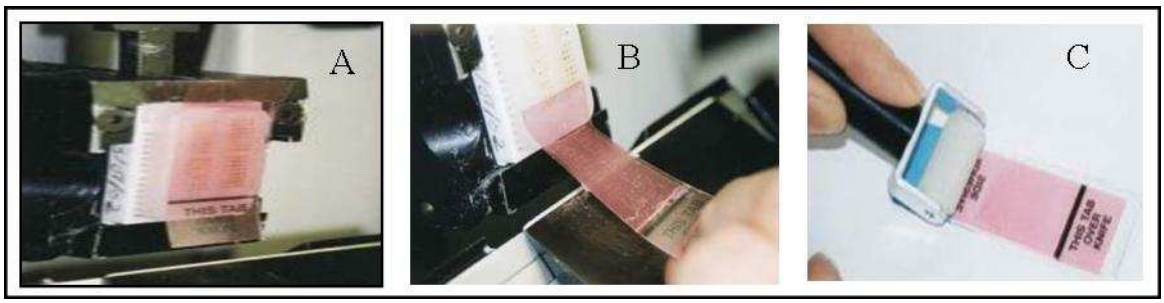


Figure 2.6 : Principe de la technique PSA®. A : Pose du scotch PSA® sur le bloc TMA avant coupe**. B :** Réalisation de la coupe et transfert de la coupe sur le scotch. C : Transfert de la coupe sur la lame PSA®.

L'utilisation de la technique du Scotch de transfert PSA® brevetée par Alphelys (www.alphelys.com), réduit fortement les pertes de carottes, élimine les plis et permet de surcroît de conserver la géométrie du TMA. La conservation de la géométrie du TMA facilite grandement son analyse au microscope. La Figure 2.6 illustre l'utilisation de la technique

PSA® pendant la réalisation des coupes TMA. La Figure 2.7 démontre l'amélioration que peut apporter la technique PSA® par rapport à l'approche classique de réalisation des coupes histologiques.

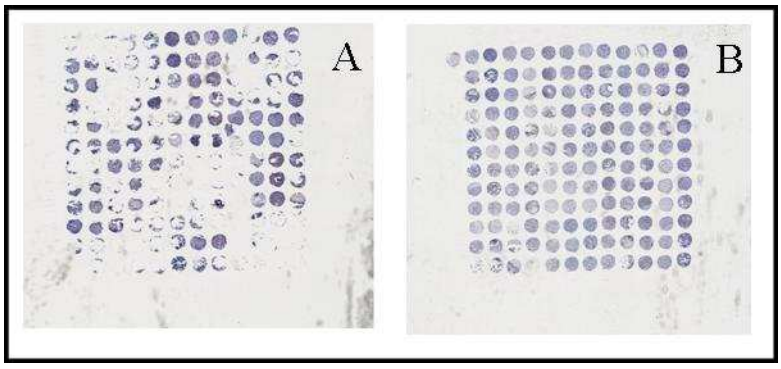


Figure 2.7 : Avantage de la technique PSA®. A : Coupe de TMA en technique classique**. B :** Coupe de TMA avec la technologie du scotch PSA®.

Malgré toutes ces dispositions, un certain nombre de spots (coupes de carottes) reste ininterprétable : il s'agit des spots, perdus pendant la coupe ou contenant du tissu nécrotique ou du stroma. Même au sein des équipes les plus chevronnées, la proportion de ces spots perdus pour l'analyse varie entre 10 et 30% [Jacquemier et al 2003] [Terris et Bralet 2002] [Engellau 2004] [Matysiak et al 2003].

Notons à la fin qu'on peut obtenir, en moyenne, 200 coupes histologiques de 4 à 5 µm à partir d'un bloc TMA [Watanabe 2005].

2.4 Domaines d'application des TMA

La technologie TMA est un outil de recherche qui rend possible des études à grande échelle sur le matériel archivé en paraffine. Les travaux utilisant cette technique se comptent par centaines [Al kuraya et al 2004] [Watanabe 2005]. Ils concernent surtout les études portant sur le cancer du sein et de la prostate, mais il existe aussi des travaux sur d'autres types de cancers [Sauter et al 2003].

La technique des TMA est un outil complémentaire des puces à ADN, pour la validation de la découverte des altérations génétiques. Le TMA peut valider à l'échelon protéique, une phase de criblage en masse des profils d'expression génétique par les puces à ADN [Bertucci et al 2004]. Elle permet d'évaluer l'impact clinique de ces altérations sur des centaines d'échantillons en même temps [Terris et Bralet 2002].

La technologie des TMA s'est principalement développée dans le domaine de la cancérologie. L'avantage majeur de cette technologie est qu'elle apporte de nouvelles informations concernant, le potentiel pronostic de certains marqueurs tumoraux (ou le confirme). C'est avec cette approche que l'expression de Cox-2 ou de la ß-caténine a été mise en évidence comme indicateur de pronostic dans le cancer colorectal [wu et al 2003] [Chung et al 2001]. La technique des TMA a été utilisée également pour la classification des tumeurs. A titre d'exemple, les profils d'expression de plusieurs marqueurs : tels que la CK7, la CK20 et la L-plastin, ont été établies comme paramètres, pour désigner le grade histologique du cancer colorectal [Park et al 2002] [Hoos et al 2002] [Otsuka et al 2001].

Les TMA permettent aussi l'étude de l'expression d'une molécule particulière, à travers plusieurs types de tumeurs. Par exemple, Anderson et al ont étudié les niveaux d'expression de la Cycline E à travers plus de 130 catégories de tumeurs [Simon et Sauter 2002].

Enfin les TMA sont également, utilisés comme un outil de contrôle de qualité, notamment pour la validation des nouveaux anticorps en IHC, sur des centaines d'échantillons tissulaires simultanément. On peut ainsi déterminer rapidement et à moindre coût les conditions idéales, telles que la concentration des anticorps, pour réaliser un immunomarquage optimal sur différents types tissulaires [Packeisen et al 2002] [De Marzo et al 2002] [Chin et al 2003]. Les TMA rendent plus facile la comparaison entre les laboratoires des pratiques usuelles telles que les techniques de colorations IHC ou l'évaluation des niveaux d'expression des marqueurs immunohistochimiques [Diaz et al 2004] [Parker et al 2002].

2.5 Problématiques liées à la technologie TMA

Nous avons souligné que le recours à la technologie TMA permet l'accélération des procédures classiques : biologiques et histologiques, notamment pour la découverte des

nouveaux marqueurs tumoraux à visée pronostic. En revanche, la technologie TMA pose de nouvelles problématiques concernant :

- la représentativité des échantillons tissulaires de petites tailles, inclus dans le bloc TMA,
- les nouveaux besoins pour l'analyse, l'exploitation et la gestion de la masse de données, grandissante générée par la technologie des TMA.

2.5.1 Echantillonnage des spots sur la biopsie

Une des difficultés de la technique TMA réside bien évidemment dans le choix optimal de l'échantillonnage à réaliser, échantillonnage en ce qui concerne d'une part les tumeurs à impliquer dans la fabrication des blocs TMA et d'autre part les divers prélèvements (ou carottes) à réaliser sur chacun des blocs donneurs.

En effet, un problème essentiel posé par la réalisation d'un bloc de paraffine unique à partir d'une ou plusieurs centaines de blocs donneurs est la taille de l'échantillon (carotte) qui peut varier de 2 mm à moins d'un ½ mm et qui limite de façon drastique ses possibilités d'être réellement représentative du potentiel de diversité morphologique d'une tumeur pouvant mesurer plusieurs cm de diamètre.

L'hétérogénéité tumorale est une donnée très fréquente en pathologie oncologique, donnée, à laquelle est particulièrement confronté le pathologiste qui veut évaluer, à partir d'un échantillon de tissu censé être le plus représentatif possible, des paramètres biologiques tels que le niveau d'expression de tel ou tel marqueur tumoral. L'accumulation des mutations au cours de l'évolution d'une tumeur à partir d'un clone originel donne naissance à une diversification du phénotype et du génotype. L'hétérogénéité qui en résulte peut alors être interprétée comme due, à la fois à l'émergence d'un clone dominant et à la persistance de clones antérieurs. Si on considère généralement que le clone dominant est celui qui doit être pris en compte pour cibler les thérapeutiques, l'histoire clinique de certains cancers montre qu'il n'en est pas toujours ainsi et que certains clones très peu représentés, « évènements rares », peuvent à eux seuls influencer de façon péjorative le cours de la maladie et donner naissance aux métastases.

Ce problème a des implications pratiques très importantes quant à la nécessité de s'assurer de la représentativité de l'échantillon que l'on doit analyser. Généralement et comme recommandé par Kononen et al [1998], trois est le nombre de carottes usuellement retenu dans les laboratoires pour échantillonner une tumeur ou une zone d'intérêt.

Plusieurs études ont comparé l'utilisation des carottes de 600 µm avec les coupes histologiques classiques des tumeurs, sur plusieurs types de cancer : le cancer de la prostate [Mucci et al 2000] [Rubin et al 2002], le cancer du sein [Camp et al 2000] [Torhorst et al 2001] et le cancer colorectal [Fernebro et al 2002] [Hendriks et al 2003]. Ces études concluent que 2 à 4 carottes sont suffisantes pour représenter une tumeur. Elles soulignent que le recours à un nombre plus important d'échantillons par tumeur, n'améliore pas ou très peu la représentativité de ces échantillons.

En revanche, certains auteurs recommandent le prélèvement d'un nombre plus conséquent de carottes pour mieux exprimer la distribution hétérogène ou focale de certains marqueurs, tels que le ki67, le Bcl2 et le p53 [Chiesa-Vottero et al 2003] [Charpin et al 2004] [Merseburger et al 2003] [García et al 2003]. L'étude de l'expression d'un marqueur à différents endroits, tels que le centre de la tumeur et sa périphérie, pour analyser le profil moléculaire d'une tumeur en progression, requiert aussi un nombre important d'échantillons [Giltnane et Rimm 2004] [Engellau 2004]. Ces résultats controversés, nous ont motivés pour réaliser un travail d'investigation sur le problème d'échantillonnage des marqueurs tumoraux du cancer colorectal. Le chapitre suivant sera entièrement consacré à ce sujet.

2.5.2 Numérisation des coupes TMA

Une fois les coupes de TMA obtenues, leur durée de conservation est limitée après la réalisation du marquage. L'intensité des colorations associées aux marqueurs baisse avec le temps. La solution est sans doute la numérisation des coupes de TMA, mais il faut que l'image obtenue contienne la totalité de l'information résidant dans la lame microscopique. Pour qu'une image numérique contienne toute l'information de la lame microscopique, il faut que le nombre de pixels numérisés *via* le capteur CCD (Charged-Coupled-Device) soit le

double du nombre de points capables d'être séparément résolus par l'optique du microscope [Leong et McGee 2001]. Le recours à l'imagerie permet une préservation illimitée des lames histologiques et favorise les échanges des ressources histologiques entre les laboratoires via le net. Les lames histologiques ainsi numérisées, peuvent être analysées sous un écran d'ordinateur, de la même manière qu'une lame histologique classique sous un microscope. Une étude de classification du cancer de la prostate a démontré une concordance de 99% entre l'utilisation des images numériques et des lames microscopiques traditionnelles [Bova et al 2001].

La numérisation d'une lame microscopique entière à haute résolution est un véritable défi technique. Une image numérique d'une résolution de 50,000 pixels par pouce (ppi) contient les mêmes détails qu'un homme peut voir en utilisant la lentille d'un microscope au grossissement x10 avec une ouverture numérique (NA) de 0.4. La taille d'une lame virtuelle créée à une résolution de balayage de 50,000 ppi (0,5 µm par pixel) est approximativement 11 Mo par mm^2. Une lame de 15mm x 15mm correspond à 2.5 Gb de mémoire [Soenksen 2003]. Le volume des données générées par la numérisation des blocs TMA, devient dès lors très important. Des solutions logicielles existent pour compresser cette masse de données [Watanabe et al 2005]. Par exemple, on peut utiliser la compression JPEG 2000 avec un rapport de 20:1 ce qui réduit à 125 Mo la taille mémoire d'une lame microscopique numérique.

2.5.3 Analyse des lames TMA numériques

La méthode la plus simple pour évaluer l'expression d'un marqueur sur les coupes TMA est l'inspection visuelle de chaque spot. Cependant, un bloc de TMA peut générer jusqu'à 200 coupes (100,000 spots), plusieurs centaines d'heures sont alors nécessaires pour analyser visuellement toutes les lames histologiques issues d'un bloc TMA. Cette durée présente 10 à 50 fois le temps alloué à la construction du bloc TMA [Watanabe et al 2005] [Rabinovich et al 2006]. L'apport des techniques d'imagerie, apparaît comme une solution adéquate, pour faciliter le travail des pathologistes afin de tirer profil rapidement de la technologie TMA.

L'expansion rapide de la technologie TMA a été accompagnée par la mise sur le marché, par plusieurs entreprises, de divers systèmes dédiés à l'acquisition et à l'analyse automatique des lames numériques TMA (Tableau 2.1). Le problème majeur de ces applications est leur coût qui n'est pas à la portée de toutes les institutions de recherche.

Tableau 2.1 : Systèmes d'analyse d'images commerciaux dédiés aux TMA.

Système d'analyse TMA	Compagnie & site web
Système SAMBA2050 ™, TMACrop™, SAMBA IPS™	TRIBVN, 39, rue Louveau - 92320 Châtillon http://www.tribvn.com
BLISS slide scanning workstation	Bacus Laboratories, Inc., Lombard, IL http://www.bacuslabs.com
Automated Cellular Imaging System (ACIS®)	ChromaVision, Inc, San Juan Capistrano, CA http://www.chromavision.com
ScanScope® Virtual microscopy solutions	Aperio Technologies, Inc., Vista, CA http://www.aperio.com/
TissueAnalytics™	Icoria™ Research Triangle Park, NC http://www.icoria.com/
GenoMX™ Vision	Biogenex San Ramone, CA http://www.biogenex.com/
Discovery TM™	Molecular Devices, Corp., Sunnyvale, CA http://www.moleculardevices.com/
Automated Quantitative Analysis (AQUA)	HistoRx, New Haven, CT http://www.historx.com/
Pathological Image Analysis and Management (PATHIAM)	Bioimagene, Inc., San Jose, CA http://www.bioimagene.com/
Ariol® Pathology Workstation	Applied Imaging Corp., San Jose, CA http://www.aicorp.com

Certaines applications existent dans la littérature, mais elles aussi sont dépendantes de plates-formes de traitement d'images commerciales [Vrolijk et al 2003] [Matkowskyj et al 2000]. De nombreuses applications de traitement d'images automatiques ont été destinées à l'analyse des marqueurs fluorescents [Camp et al 2002] [Jubb et al 2003]. Plusieurs articles démontrent la robustesse de ces applications dans l'analyse automatique des TMA dans de nombreux types de cancer : prostate, sein, colon etc. [Camp et al 2003] [Jubb et al 2003]. [DiVitoet al 2004] [Rubin et al 2004]. Néanmoins, la plupart des marqueurs sont révélés sur les lames TMA, à l'aide des techniques d'IHC, moins coûteuses, faisant intervenir des chromogènes tels que le Di-Amino-Benzidine (DAB) (Annexe A). Nous allons aborder l'analyse automatique de ce type de lames, au cours de la deuxième partie de ce manuscrit, où nous présenterons notre approche de quantification automatique de l'expression des marqueurs tumoraux.

2.5.4 Gestion et exploitation des données générées par la technologie des TMA

La création de TMA permet l'analyse immuno-histochimique rapide de plusieurs centaines de spots, avec de nombreux anticorps différents. Ce développement technique a créé un besoin d'outils pour faciliter la gestion et l'exploitation des grandes quantités de données produites par l'analyse des images numérisées des coupes de TMA.

Plusieurs travaux ont été menés pour proposer des systèmes de gestion des données concernant les TMA. Ainsi, Liu et al proposent une base de données pour stocker les images des coupes des TMA [Liu et al 2002]. Bova et al ont développé une application client-serveur qui permet à des utilisateurs distants (pathologistes) de télécharger des images numériques et de poser le diagnostic par l'intermédiaire d'une interface graphique à partir de la base de données gérée par le serveur [Bova et al 2001]. Manley et al proposent une base de données relationnelle, qui gère les images des spots, leurs emplacements dans le bloc TMA et les dossiers cliniques qui leurs sont associés [Manley et al 2001]. Le point commun de ces bases est leur simplicité et leur accessibilité via le net pour la visualisation des spots par les pathologistes. Néanmoins, elles ne prennent pas suffisamment en compte, l'exploitation ou la fouille des données générées par la technologie TMA.

Pour répondre à ce besoin, des systèmes complets ont été développés pour accompagner les équipes de recherche à partir de l'élaboration du plan de construction du bloc TMA jusqu'à l'exploitation statistique des niveaux d'expression des marqueurs tumoraux quantifiés sur les lames TMA. Dans cette catégorie, on peut citer le système Profiler développé par Kim et al [Kim et al 2005], qui peut associer des données cliniques à l'expression d'un marqueur donné sous forme de courbes ou de graphiques. Demichelis propose le système TMABoost, qui donne la possibilité aux utilisateurs de réaliser leurs propres expériences, selon les ressources stockées, telles que l'affichage d'une courbe de survie en relation avec l'évolution de l'expression d'un marqueur tumoral [Demichelis 2005]. On peut citer également Virtual Tissue Matrix, qui regroupe des spots provenant même de différents blocs de TMA pour réaliser divers types d'analyses [Conway et al 2006].

Dans ce contexte, nous mentionnons le système **TMA Explorer** développé au sein de notre équipe de recherche (RFMQ), qui permet de piloter les différentes phases en amont et en aval de la construction d'un bloc TMA [Giroud et al 2004]. **En amont,** en assistant l'utilisateur dans sa conception du bloc TMA, ce système donne diverses solutions pour construire un bloc TMA. A ce niveau, la méthodologie d'échantillonnage développée au cours de ce travail (§chapitre 3), apporte une assistance à l'utilisateur pour un échantillonnage optimal des blocs d'archives donneurs. **En aval**, en assistant l'utilisateur dans son exploitation des données générées par l'analyse automatique des lames TMA. Grâce à cette approche, M^{elle} Julie Bourbeillon qui a développé le module de fouille de données associé à TMA Explorer, a pu mettre en évidence la relation entre l'expression nucléaire de la ß-caténine et la progression tumorale du cancer colorectal [Bourbeillon 2007].

2.6 Conclusion

La technologie des TMA se situe à l'interface entre les pathologistes et les biologistes moléculaires, notamment pour l'investigation sur le rôle des altérations moléculaires, à l'échelon tissulaire, dans la progression tumorale. Elle consiste à regrouper, des centaines d'échantillons biopsiques (600 µm), dans un seul bloc de paraffine, à partir de plusieurs blocs d'archives donneurs. C'est une technique haut-débit, qui permet la réalisation simultanée des profils d'expression moléculaires, sur des centaines d'échantillons. L'analyse par TMA, à grande échelle, des marqueurs tumoraux sur des tumeurs ayant différents stades de cancérisation et grades histologiques, permet la mise en évidence rapide du réseau moléculaire impliqué dans la transformation tumorale d'un tissu sain.

En dépit des avantages multiples des TMA, la critique porte essentiellement sur la qualité des études réalisées, sur des échantillons tissulaires mesurant 600 µm de diamètre. Nous avons vu à travers ce chapitre, que les avis sont controversés sur ce sujet. Ce point sera discuté en détails au niveau du chapitre suivant. En effet, nous aborderons le problème d'échantillonnage des blocs de biopsies donneurs. Cet échantillonnage concerne, les zones tumorales à prendre en compte au sein de chaque bloc donneur et le nombre d'échantillons à prélever au niveau de chacune de ces zones.

Chapitre 3

Concevoir des puces à tissus TMA : Méthodologie d'échantillonnage basée sur le concept de « carottes virtuelles »

La technique des puces à tissus « TMA : Tissue Micro Arrays » permet de collecter, en un seul bloc d'inclusion en paraffine, des centaines d'échantillons sous forme de petites carottes cylindriques, prélevées à partir de différents blocs de tumeurs. La critique porte essentiellement sur la représentativité des résultats obtenus à partir d'aussi petits fragment tissulaires, (0.6 mm de diamètre environ), compte tenu de l'hétérogénéité de certaines tumeurs pouvant mesurer plusieurs cm de diamètre. Le but de ce chapitre est alors d'étudier l'impact de l'échantillonnage sur l'utilisation de la technologie TMA, notamment pour l'étude de la progression tumorale du cancer colorectal.

Le concept de « carottes virtuelles » est introduit dans ce chapitre, il s'agit de simuler le prélèvement réel de carotte. Un algorithme associé à ce concept est pratiqué au niveau des blocs de tumeur pour réaliser des prélèvements aléatoires de carottes virtuelles. L'objectif est de déterminer le nombre optimal de carottes à prélever en respectant l'hétérogénéité des zones tumorales à échantillonner.

En résultats, le nombre de carottes à prélever doit être adapté à l'hétérogénéité du matériel tumoral à échantillonner et aux marqueurs tumoraux considérés. Si le prélèvement d'une à deux carottes apparaît suffisant pour certaines zones (homogènes) et certains marqueurs à distribution homogène ou focale (Bcl2), en revanche, les marqueurs à distribution hétérogène, tels que le Ki67 ou la ß-caténine requièrent le prélèvement d'un nombre plus important d'échantillons.

Le concept de « carottes virtuelles » permet la conception des blocs de « TMA virtuel » pour proposer différents plans expérimentaux de construction de TMA réel. Nous avons démontré l'utilité de ce concept pour résoudre le problème d'échantillonnage des blocs de tumeurs.

3.1 Introduction

La technique des puces à tissus « TMA » s'inscrit dans le cadre des micro-technologies qui se développent actuellement en biologie moléculaire *in situ*. Elle permet aux chercheurs de visualiser et d'étudier des centaines d'échantillons tissulaires en même temps. Elle se situe à l'interface entre biologistes et pathologistes, notamment pour la validation, à l'échelon tissulaire, des nouveaux marqueurs biologiques découverts par les biotechnologies à haut-débit, telles que les DNA arrays ou les cDNA arrays [Skacel et al 2007] [Ginestier et al 2002]. La technologie TMA est un moyen indispensable d'investigation sur le rôle éventuel des différentes molécules qui sont impliquées dans le processus de développement tumoral. Elle permet ainsi d'utiliser ces molécules comme des indicateurs de diagnostic ou de pronostic, associés aux différents types de cancer [Alonso et al 2004] [Andersen et al 2002] [Camp et al 2000] [Chung et al 2001] [Nielsen et al 2003].

Comme nous l'avons décrit dans le chapitre précédent (§ 2.2), la technique des TMA permet de collecter en un seul bloc d'inclusion en paraffine des centaines d'échantillons tissulaires, sous forme de petites carottes cylindriques prélevées à partir de différents blocs de biopsies donneurs. Les carottes prélevées peuvent être de diamètres différents (de 0.6 mm à 2mm), mais néanmoins les carottes de 0.6 mm sont les plus souvent utilisées [Kononen al 1998].

En dépit des nombreux avantages de cette nouvelle technologie, la critique porte essentiellement sur la représentativité des résultats obtenus à partir de petits échantillons tissulaires (0,6-2 mm de diamètre), compte tenu de l'hétérogénéité de certaines tumeurs pouvant mesurer plusieurs cm de diamètre. Trois questions principales se posent concernant l'échantillonnage des blocs de biopsies donneurs :

- Comment définir les zones de carottage aux niveaux des blocs de biopsies donneurs ?
- Quel doit être la taille des échantillons à prélever ?
- Quel est le nombre d'échantillons nécessaires pour représenter un bloc de biopsie donneur ?

Le choix des zones de carottage sur les blocs donneurs se fait par un pathologiste sur la base des informations recueillies après la visualisation d'une coupe HES [Sauter et al 2003]. La

coloration HES fournie des informations histologiques d'ordre général, telles que la configuration spatiale des zones tumorales ou les désordres architecturaux en surface des blocs donneurs. Mais quelle est la valeur réelle de cette coloration classique (HES) pour délimiter, par exemple, les foyers de prolifération cellulaire ? Nous présenterons plus loin dans ce chapitre, un protocole pour assister le pathologiste, dans la délimitation des zones de carottages, en y apportant des informations concernant l'activité proliférative des blocs de biopsies donneurs.

Concernant la taille des échantillons à inclure dans le bloc de TMA, il existe un consensus sur l'utilisation des carottes de 600 µm de diamètre [Kallionemi et al 2001]. Le recours à un tel diamètre permet d'inclure jusqu'à mille échantillons dans un bloc TMA tout en préservant les blocs de biopsies donneurs. En effet, les carottes de 2 mm causent un dommage plus important aux blocs donneurs sans toutefois améliorer de façon significative le pouvoir de représenter l'hétérogénéité tissulaire des blocs donneurs, par rapport aux carottes de 600 µm [Kajdacsy-Balla et al 2007]. Plusieurs auteurs recommandent le prélèvement de plusieurs carottes de 600 µm, au lieu d'une carotte de plus grande taille [Camp et al 2000] [Giltnane et Rimm 2004] [Simon et al 2003]. Cependant, les avis divergent quant au nombre de carottes à prélever au niveau de chaque bloc donneur. En effet, la taille réduite des carottes (600µm) limite leurs possibilités d'être réellement représentatives du potentiel de la diversité morphologique des blocs donneurs. Plusieurs travaux ont été menés pour délimiter le nombre de carottes, de 600 µm, nécessaires, pour représenter une tumeur.

Lorsqu'il s'agit d'établir un lien entre l'expression des différents marqueurs tumoraux et le pronostic, ou le grade de certains types de cancers, la plupart des études suggèrent que 1 à 4 carottes sont suffisantes pour représenter correctement une tumeur. Ainsi, Torhorst et al trouvent qu'une seule carotte est suffisante pour établir les marqueurs ER (récepteur d'œstrogène), PR (récepteur de progestérone) et p53 comme des indices de pronostic, dans une étude de 553 cancers de sein [Torhorst et al 2001]. Nocito et al dans une étude portant sur 2317 cas de cancers de la vessie, démontrent que 4 carottes sont suffisantes pour désigner le grade tumoral à l'aide du marqueur Ki67 [Nocito et al 2001]. Camp et al comparaient les coupes TMA avec les coupes histologiques classiques, sur 38 cas de cancers du sein. Ils ont prélevé pour cela, 10 carottes au niveau de chaque bloc donneur. 95% de concordance a été

observée entre seulement deux carottes et les coupes entières des tumeurs. L'addition de carottes supplémentaires, augmente le taux de concordance jusqu'à 99,5% avec 5 carottes.

Le nombre de 3 carottes reste le nombre de prélèvements le plus recommandé par les chercheurs [Ginestier et al 2002] [Hoos et al 2001] [Engellau et al 2001] [Giltnane et Rimm 2004] [Simon et al 2003] [Fernebro et al 2002] [Hendriks et al 2003]. Par exemple, Ginestier et al ont souligné que 3 échantillons sont parfaitement adaptés pour l'analyse de l'expression des marqueurs, tels que le RP, le Bcl2 et le p53 dans le cas du cancer du sein [Ginestier et al 2002]. Hoos et al ou Engellau et al démontrent que trois carottes sont largement suffisantes pour l'étude de l'expression des marqueurs tumoraux dans le cas des tumeurs fibroblastiques, y compris les marqueurs ayant une distribution hétérogène, tels que le Ki67 [Hoos et al 2001] [Engellau et al 2001]. Concernant le cancer colorectal, Fernebro et al ou Hendriks et al démontrent que 3 échantillons sont capables de représenter les tumeurs, notamment pour le Ki67. Ils soulignent que le prélèvement de 3 carottes est nécessaire pour compenser les pertes d'échantillons, lors des différentes étapes de manipulation des TMA et pour tenir compte en même temps de l'hétérogénéité des tissus [Fernebro et al 2002] [Hendriks et al 2003].

En contraste avec ces travaux, certains auteurs considèrent qu'un nombre limité d'échantillons, n'est pas adapté pour prendre en considération la nature hétérogène ou focale de la distribution de certains marqueurs, tels que le ki67, le Bcl2 et le p53 [Chiesa-Vottero et al 2003] [Charpin et al 2004] [Merseburger et al 2003] [García et al 2003] [Singh et al 2004]. Ainsi, Singh et al recommandent la modulation du nombre d'échantillons en fonction de l'hétérogénéité des blocs de tumeurs donneurs. Ils notent qu'il faut prélever de 5 à 12 carottes pour tenir compte de l'hétérogénéité des marqueurs tumoraux Ki67 et AR [Singh et al 2004]. Merseburger et al soulignent que les blocs TMA, intégrant 3 carottes de chaque bloc donneur, ne sont pas adaptés pour l'analyse des marqueurs p53 et Bcl2, qui sont des marqueurs à distribution focale et localisée [Merseburger et al 2003].

Camp et al ou Engellau résument ces opinions qui paraissent controversées sur la technologie TMA [Camp et al 2000] [Engellau 2004]. Ils notent que dans la plupart des cas, 3 échantillons de 600 µm sont suffisants pour représenter une tumeur, notamment pour l'analyse de l'expression des marqueurs tumoraux à distribution homogène. Mais ils rajoutent que l'étude

de l'expression de certains marqueurs à distribution hétérogène, nécessite le prélèvement d'un nombre plus important d'échantillons. Par exemple, l'étude de l'hypoxie au niveau des tumeurs, requiert le prélèvement de plusieurs carottes, au niveau du centre de la tumeur et sur sa périphérie. Kajdacsy-Balla et al, et Waterworth et al notent qu'il faut prélever jusqu'à 14 carottes, pour représenter la distribution hétérogène de l'hypoxie cellulaire, ou l'étude du profil moléculaire des fronts d'invasion tumorale [Waterworth et al 2004] [Kajdacsy-Balla et al 2007]. Engellau souligne que le TMA devient même inadapté pour l'étude des marqueurs très hétérogènes, comme par exemple les marqueurs d'angiogénèse [Engellau 2004].

Nous avons vu à travers ces travaux, l'importance de l'échantillonnage dans le processus de fabrication d'un nouveau bloc de TMA. Les avis divergent parfois en ce qui concerne les zones d'échantillonnage et le nombre d'échantillons à prélever pour assurer une bonne représentativité des blocs de tumeurs donneurs. Nous développerons dans la suite de ce chapitre, une stratégie d'échantillonnage virtuelle, basée à la fois sur les annotations pathologiques et l'expression d'un marqueur à distribution hétérogène : le Ki67 [Gerdes et al 1984]. Le but de cette stratégie est de définir le nombre optimal d'échantillons à prélever, pour une représentation optimale de l'hétérogénéité tumorale. Pour cette raison, des outils méthodologiques dédiés à l'échantillonnage ont été développés, tels que « les cartes de prolifération cellulaire » et « Le prélèvement virtuel des échantillons ».

Ces outils d'échantillonnage virtuels vont être appliqués, pour l'étude de l'expression de quatre marqueurs ayant un intérêt particulier dans le cancer colorectal : le Ki67, la β-caténine, la Cycline D1 et le Bcl2. En faisant le choix de ces marqueurs tumoraux, nous avons voulu étudier l'impact de l'échantillonnage sur l'utilisation de la technologie TMA pour l'étude de la cancérogénèse colorectale.

Le Ki67, marqueur des cellules en phase active du cycle et susceptibles de proliférer (cellules prolifératives). Le Ki67 définit la cartographie du taux de prolifération cellulaire sur les coupes histologiques. C'est à ce titre qu'il a été retenu dans ce travail, car il nous renseigne sur le statut prolifératif des tumeurs. La β-caténine a été retenue du fait de son rôle clef dans la voie moléculaire de la cancérogénèse colique. En effet, la majorité des altérations moléculaires, initialisant le processus de cancérogénèse, affecte l'activité de la β-caténine

(§1.7.2). La protéine Cycline D1 a été retenue comme cible privilégiée de la β-caténine dans l'activation de la prolifération cellulaire. Le Bcl2 est une protéine anti-apoptotique, dont la présence dans le cytoplasme empêche la mort cellulaire. Il a été retenu comme témoin indirect de l'absence de l'expression des gènes suppresseurs de tumeurs, tels que le p53 muté dans 70 % des cancers du colon (§1.7.1) [Brunz et al 1999].

A la fin de ce chapitre, nous discuterons le concept de « TMA virtuel », notamment pour établir les plans de construction des TMA ou l'exploitation des données générées par la technologie TMA.

3.2 Matériel et méthodes

3.2.1 Matériel biologique

Les blocs d'archives de tumeurs sont recrutés au Centre de Recherche de Lutte Contre le Cancer de Montpellier (CRLCC Val d'Aurelle) sous la responsabilité du Dr. Joëlle Simony du service d'anatomopathologie. Les blocs sont préparés en suivant les grandes étapes classiquement utilisées en histologie : fixation, déshydratation et inclusion en paraffine. En totalité, nous avons intégré 61 blocs de cancer du colon dans cette étude. 5 coupes sériées ont été pratiquées au niveau de chaque bloc de tumeur pour la réalisation de la coloration HES et les différents marquages immunohistochimiques.

3.2.2 Coloration HES

Cette coloration comporte 3 constituants :

- **Hématoxyline** : contient un colorant appelé l'hématéine ou hemalun qui est une laque nucléaire qui vire au bleu noirâtre quand elle est bien différenciée.
- **L'éosine** : c'est un colorant acide faible qui colore en rose les cytoplasmes.
- **Le safran :** est un colorant naturel qui colore le collagène de type III en jaune.

La coloration HES permet la visualisation de l'étendue de la zone tumorale dans les structures histologiques du colon. C'est sur la lame HES que se base le pathologiste pour réaliser les

annotations anatomopathologiques, et pour définir les zones d'intérêts au niveau desquelles seront réalisés les prélèvements des carottes TMA.

3.2.3 Les marqueurs tumoraux

4 coupes histologiques sériées de 4-5 µm ont été pratiquées dans les blocs de tumeurs pour la révélation des marqueurs tumoraux. Les procédures de révélation de ces marqueurs, ainsi que les concentrations des différents anticorps dirigés contre eux sont détaillées dans l'annexe A. En finalité, 5 lames histologiques ont été obtenues à partir de chaque bloc de tumeur. Une lame HES et quatre lames pour les marqueurs tumoraux.

3.2.4 Acquisition des images histologiques

La numérisation des lames histologiques à haut-débit a été réalisée avec le système d'acquisition d'images Samba 2050^{TM} (www.tribvm.com) dont nous disposons au laboratoire. Ce système est équipé d'un microscope Olympus Bx-40F à lumière transmise avec les objectifs ×4, ×10, ×20 et ×40. Sur ce microscope a été montée une camera CCD Spot Insight 1600x1200 avec un focus permettant une vitesse d'acquisition de 2000 µm/sec, et une résolution de 0,5 µm. Le microscope reçoit les lames d'un chargeur d'une capacité de 50 lames relié à une platine automatisée, le tout permettant l'encodage des positions des lames. La précision du pas de navigation est supérieure à 5 µm. Toute la station d'acquisition est pilotée par ordinateur Pentium IV, 2.2 GHz, 512 Ko de cache et 512Mo de RAM (Figure 3.1).



Figure 3.1 : Station d'acquisition d'images Samba2050^{TM}. Le chargeur de lames est à gauche.

Le grossissement x10 a été retenu pour la numérisation de la surface totale des lames histologiques. Vu les dimensions des lames histologiques, qui peuvent atteindre 4 cm^2, la surface de ces lames est découpée en une grille de petits champs jointifs. Le système Samba 2050™ parcourt tous les champs, se focalise, à chaque étape, sur un champ pour le numériser, puis se déplace vers le champ voisin.

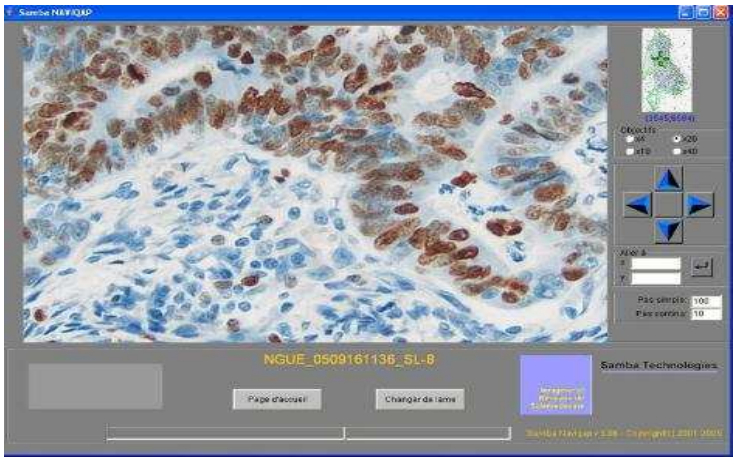


Figure 3.2 : Système de visualisation des lames histologiques numériques Naviqap™.

La station d'acquisition d'images est équipée d'un outil d'exploration des lames numérisées, appelé Naviqap™ (TRIBVN™). Ce système est une sorte de microscope virtuel, il permet la visualisation des lames histologiques numérisées sur l'écran d'un ordinateur. Ainsi, le pathologiste peut visualiser les lames HES sur Naviqap™ pour délimiter les zones d'intérêts (Figure 3.2).

Les zones d'intérêts délimitées par le pathologiste sont alors acquises, une deuxième fois, avec le grossissement x20.

3.2.5 Analyse et traitement d'images

Toutes les procédures de traitement d'images ont été développées avec le langage C, sous l'API Windows 32 à l'aide de l'environnement de programmation Visual C++ et en intégrant la librairie de traitement d'images Samba32.lib (TRIBVM™). Ces programmes ont été développés pour tracer les cartes de prolifération cellulaire des lames ki67. Ils étaient

également développés pour automatiser l'échantillonnage et la génération des données pour tracer les courbes de tirages aléatoires associées à l'ensemble des zones d'intérêts.

3.2.6 Délimitation des zones d'intérêts

3.2.6.1 Annotation sur lames HES

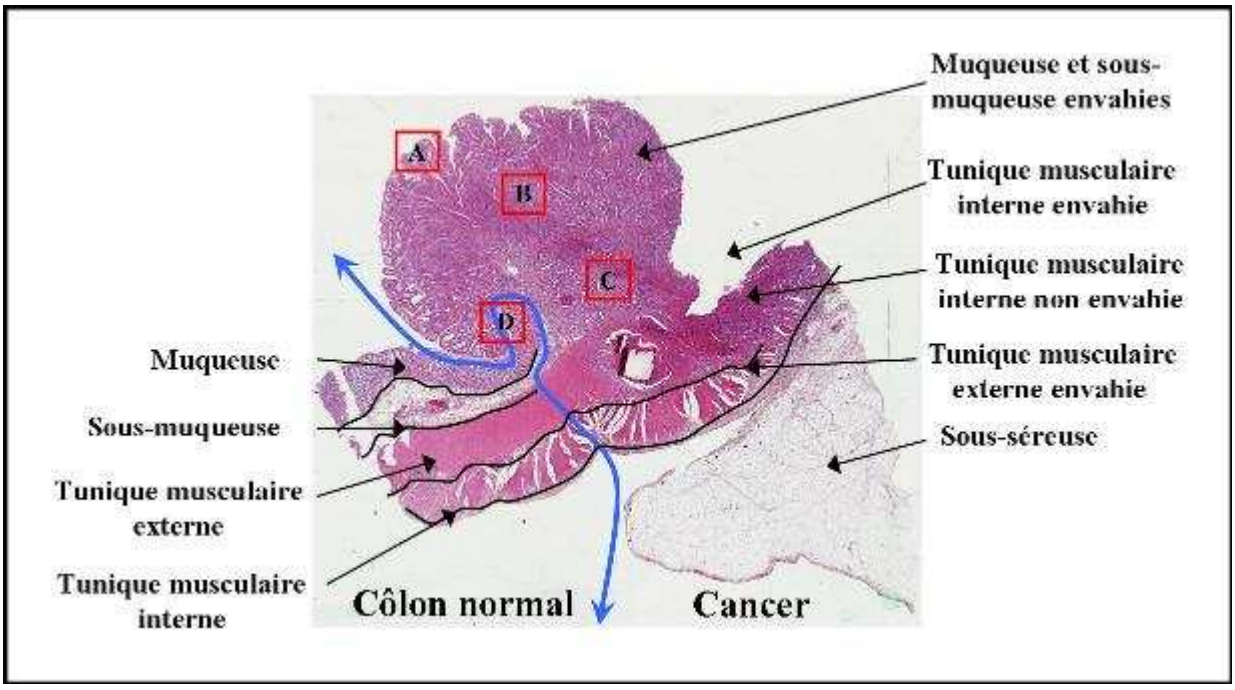


Figure 3.3 : Annotation des zones d'intérêts sur les lames HES. Zone A : en surface. Zone B : en pleine tumeur. Zone c : en profondeur de la tumeur. Zone D : à cheval entre tissu sain et tissu tumoral.

Les zones de carottage : zones d'intérêts (ZI), délimitées par le pathologiste sur la lame HES, ont été classées en 4 catégories, selon leur localisation par rapport à la surface tumorale. Ainsi, les zones en surface de la tumeur, vers la lumière colique sont annotées zones A. Les zones en plein développement tumoral, zones B ; les zones en profondeur vers les structures histologiques internes du colon, zones C et les zones non tumorales, à cheval entre tissu sain et tissu tumoral, sont notées zones D (Figure 3.3). En totalité 153 zones d'intérêts ont été délimitées sur les lames HES des 61 blocs de tumeurs. Ces zones comprennent 47 zones A, 57 zones B et 49 zones C. Leurs surfaces varient entre 1,44 mm^2 et 12,96 mm^2.

3.2.6.2 Les cartes de prolifération cellulaire

La délimitation des zones d'intérêts est parfois une tâche non triviale. S'il parait facile de désigner des zones de type B en contact de la lumière intestinale, la délimitation des zones en plein développement tumoral sur les fronts de progression de la tumeur est difficile. La délimitation de ces zones, oblige le pathologiste à parcourir la surface de toute la lame HES, sous différents grossissements. Cette opération nécessite une attention toute particulière du pathologiste, elle est fastidieuse et consomme beaucoup de temps.

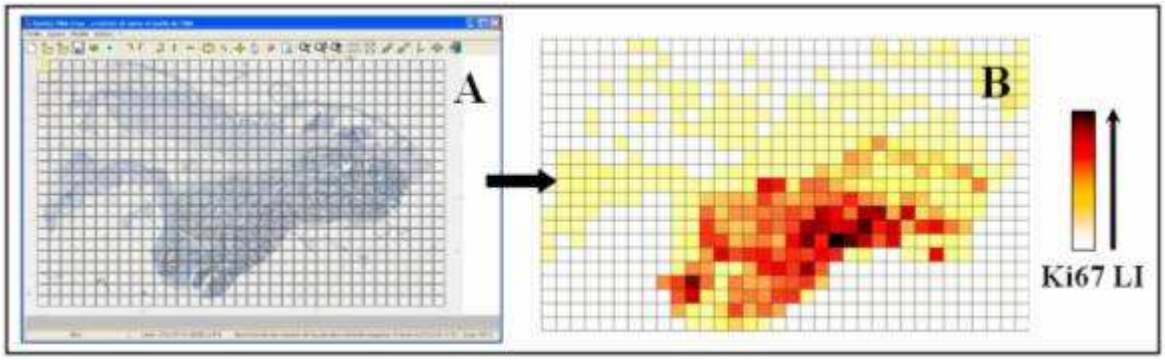


Figure 3.4: Carte de prolifération cellulaire. A : Découpage de la lame Ki67 en une grille de champs (600 µm). **B :** Carte de prolifération cellulaire affichée sur la base de la distribution spatiale de l'indice de marquage Ki67.

Dans le but de faciliter la tâche du pathologiste, nous avons introduit le concept de cartes de prolifération cellulaire pour assister ce dernier dans la délimitation des zones d'intérêts. Ces cartes de prolifération sont construites à partir des lames Ki67 acquises au grossissement x10. Après acquisition, ces lames numériques sont subdivisées en une grille régulière de champs carrés de 600 µm de coté, en mimant des carottes TMA, à l'aide du logiciel TMACrop™ (Samba Technologies^{TM}). L'indice de marquage (labelling index : LI), associé à l'expression de la protéine Ki67, est calculé au niveau de chaque champ. On obtient de cette façon, la topographie de la prolifération cellulaire sur toute la lame Ki67. Un exemple, de carte de prolifération, est illustré sur la Figure 3.4 où les champs sont affichés selon une échelle, de fausses couleurs, graduée selon le niveau d'indice de marquages Ki67. Ces cartes nous permettent de visualiser les foyers de prolifération cellulaire qui peuvent être les centres actifs de la prolifération tumorale.

Les cartes de prolifération cellulaire peuvent orienter le pathologiste pour le ciblage des zones à forts taux de prolifération, pour délimiter les zones A en plein développement. Avec ce type de cartes, la détection des fronts de l'invasion tumorale devient également plus aisée. Ainsi, le pathologiste peut délimiter également de façon facile et rapide les zones de type C et D. La Figure 3.5 résume notre méthodologie pour une délimitation, des zones d'intérêts, plus rapide et plus fine. Ainsi, le pathologiste peut visualiser, à la fois, les cartes de prolifération et les lames HES pour définir les zones de prélèvement des carottes (zones d'intérêts).

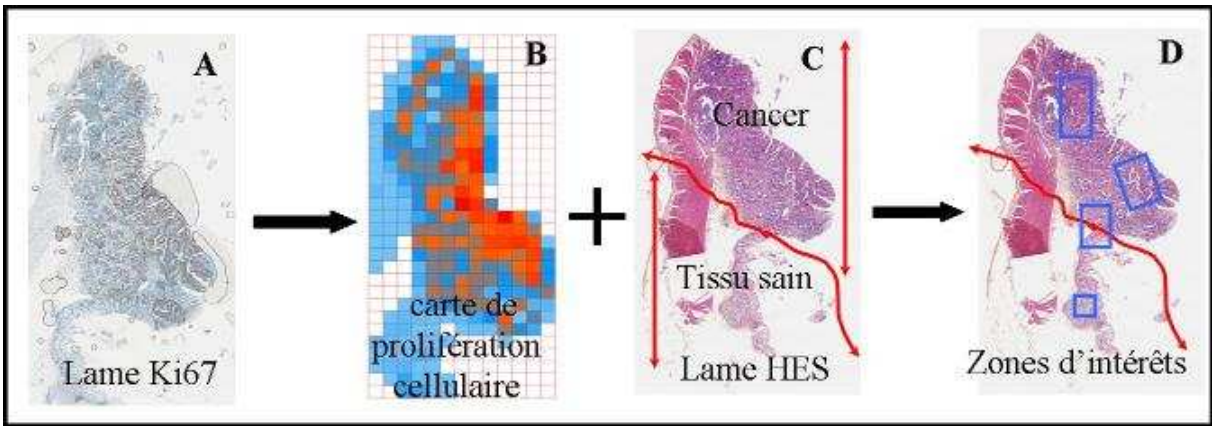


Figure 3.5 : Procédure de délimitation des zones d'intérêts. A : lame Ki67**. B :** génération de la carte de prolifération cellulaire**. C :** lame HES**. D :** délimitation des zones d'intérêts par le pathologiste après visualisation de la lame HES et la carte de prolifération cellulaire**.**

3.2.7 Echantillonnage des zones d'intérêts

Dans le but de déterminer le nombre de carottes à prélever au niveau de chaque zone d'intérêt, en tenant compte de son potentiel d'hétérogénéité, nous avons développé un protocole de carottage virtuel. Il s'agit de simuler le prélèvement de carottes réelles, au niveau des lames histologiques numériques ß-caténine, Bcl2, Cycline D1 et Ki67. C'est un outil d'investigation sur le nombre de carottes nécessaires pour représenter la distribution de ces marqueurs tumoraux au niveau des zones d'intérêts. Nous présenterons dans ce qui suit, les deux étapes du protocole de carottage virtuel.

3.2.7.1 Génération des carottes virtuelles

Le protocole consiste à reporter les zones d'intérêts, délimitées par le pathologiste au niveau de la lame HES, sur les lames des quatre marqueurs tumoraux étudiés. Ensuite, on découpe

virtuellement les zones d'intérêts, sur ces lames, en une grille de champs, où chaque champ correspond à un endroit éventuel de prélèvement et sa taille est égale au diamètre des carottes à prélever. Chaque champ représente alors une « carotte virtuelle » appelée également « spot virtuel » (Figure 3.6).

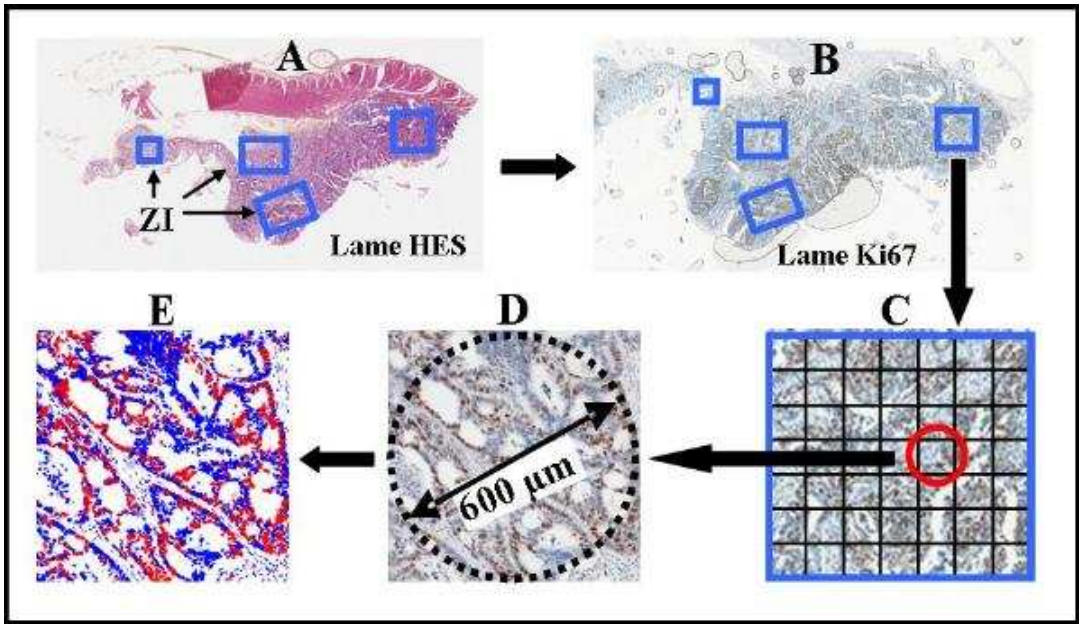


Figure 3.6 : Procédure de génération des carottes virtuelles. A : Zones d'intérêts (ZI) délimitées par le pathologiste sur la lame HES**. B :** Reproduction des ZI sur la lame Ki67**. C :** Découpage de la ZI en une grille de champs appelés « carottes virtuelles » avec TMACrop™. **D :** Chaque carotte virtuelle possède la taille d'une carotte réelle (600 µm de diamètre). **E :** Calcul de l'indice de marquage de la carotte virtuelle à l'aide d'IMMUNO™.

Le découpage des zones d'intérêts est réalisé par l'intermédiaire du logiciel TMACrop™ (TRIBVN™). L'indice de marquage, associé à l'expression du marqueur tumoral au niveau de chaque champ (carotte virtuelle), est calculé avec le logiciel IMMUNO™ (TRIBVN™). On obtient ainsi la topographie de la distribution spatiale du marqueur tumoral, au sein des zones d'intérêts. La Figure 3.6 illustre un exemple de découpage des zones d'intérêts sur une lame Ki67, puis la génération des carottes virtuelles.

3.2.7.2 La procédure d'échantillonnage des carottes virtuelles

Un algorithme de prélèvement aléatoire et progressif est appliqué au niveau de chaque zone d'intérêt. Le prélèvement virtuel débute par le tirage aléatoire d'un seul champ, puis un jeu de

deux champs et ainsi de suite jusqu'au tirage de la totalité des champs (carottes virtuelles) de la zone d'intérêt à échantillonner (Figure 3.7).



Figure 3.7 : Principe du prélèvement aléatoire des carottes virtuelles.

Le nombre de carottes, virtuellement prélevées, est ainsi incrémenté à chaque étape du déroulement de l'algorithme. La moyenne des indices de marquage, de chaque jeu de carottes virtuelles prélevées, est calculée. Le jeu de carottes le plus représentatif est celui qui produit la moyenne la plus proche de l'indice de marquage de la zone d'intérêt échantillonnée, qui est la moyenne des indices de marquage de toutes les carottes virtuelles appartenant à cette zone. Signalons que pour chaque nombre de carottes, le prélèvement virtuel est répété un millier de fois, avec replacement des échantillons pour le tirage suivant, pour avoir des résultats stables et reproductibles. Il s'agit de la technique du « bootstrap », adaptée à l'étude de l'échantillonnage au niveau des petites populations [Wonnacott et Wonnacott 1995]. En adoptant cette méthodologie, nous avons alors obtenu une distribution (1000 fois), de la moyenne des indices de marquage, pour chaque nombre de carottes virtuelles prélevées. Ces distributions reflètent la fluctuation de l'échantillonnage au niveau des zones d'intérêts. La Figure 3.8 montre un exemple d'une série de 1000 prélèvements aléatoires, de 3 carottes virtuelles, pratiqués au niveau d'une zone d'intérêt d'une lame Cycline D1. Du fait du tirage aléatoire, on remarque une fluctuation de la moyenne des indices de marquage des carottes prélevées autour de l'indice de marquage de la zone d'intérêt échantillonnée.

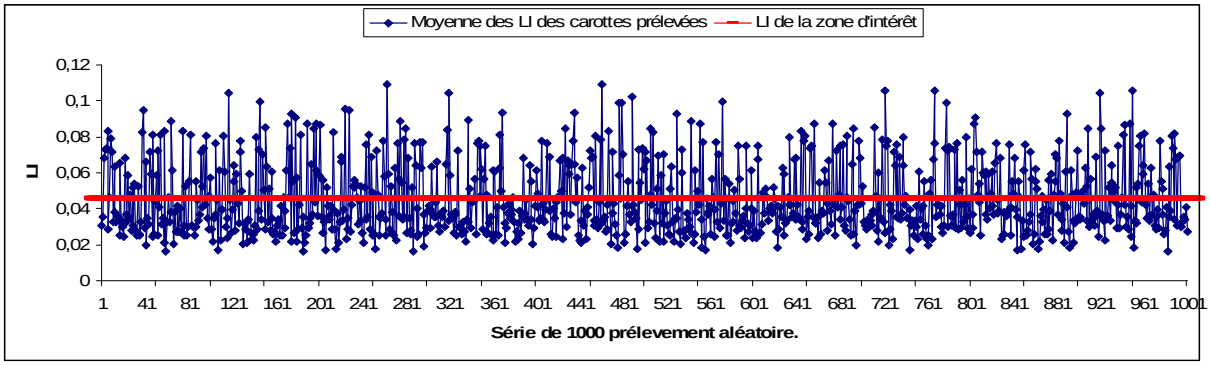


Figure 3.8 : Exemple de distribution d'une série d'échantillonnage virtuel. Distribution des indices de marquage moyens générés par le prélèvement de 3 carottes virtuelles (1000 fois). La ligne rouge présente l'indice de marquage de la zone d'intérêt à échantillonner. Cette zone comporte une grille de 25 champs (carottes virtuelles).

L'idée est d'évaluer, par l'intermédiaire de la distribution des prélèvements aléatoires, la précision d'un nombre donné de carottes à estimer la vraie valeur de l'indice du marquage de la zone à échantillonner. Dans ce cadre, nous avons utilisé l'erreur quadratique moyenne (EQM) associée aux distributions des prélèvements aléatoires. L'EQM est une mesure utilisée dans les processus de contrôle qualité, notamment dans l'industrie, pour vérifier la conformité des échantillons produits avec les objectifs voulus [Battaglia et Maynard 1992] [Wonnacott et Wonnacott 1995].

L'EQM est calculée comme suit : $EQM = \sigma^2 + biais^2$.

- $\boldsymbol{\sigma^2}$ est la variance de la distribution des moyennes des indices de marquage des carottes virtuelles prélevées. La variance (σ^2) est introduite pour la prise en compte de l'aspect aléatoire des prélèvements d'un coté, et de la distribution spatiale de l'expression du marquage dans les zones d'intérêts de l'autre coté. Une expression homogène d'un marqueur tumoral donné, à travers la surface de la zone d'intérêt, se traduit par une variance (σ^2) faible.
- **biais** = μ - T. μ est la moyenne de la distribution d'échantillonnage, c'est-à-dire la moyenne des moyennes des indices de marquage des échantillons prélevés. μ est un estimateur de l'indice du marquage de la zone d'intérêt à échantillonner : T. T est la cible de cet estimateur (μ), c'est la moyenne des indices de marquages associés aux champs (carottes virtuelles) appartenant à la zone d'intérêt. L'objectif est de trouver l'estimateur μ le plus proche de T, c'est-à-dire celui générant le plus faible biais.

La valeur d'EQM décroît avec la progression du nombre de carottes prélevées, jusqu'à ce qu'elle devienne nulle avec le prélèvement de toutes les carottes virtuelles de la zone d'intérêt. Notre objectif n'est pas de prélever tous les champs de la zone à échantillonner, pour avoir une erreur d'estimation de l'indice de marquage de cette zone, quasiment nulle (EQM=0), mais, le but est de déterminer le nombre de carottes virtuelles à prélever générant une erreur d'estimation acceptable. Un seuil EQM $(5\%)^2$ est déterminé avec un biais acceptable égal à 5 % (en intégrant la variabilité due à l'échantillonnage aléatoire), pour estimer la vraie valeur de l'indice du marquage de la région à échantillonner.

Une courbe est tracée représentant l'évolution de l'EQM, en relation avec le nombre de carottes virtuelles prélevées aléatoirement, pendant la progression de l'algorithme d'échantillonnage virtuel. En suivant la progression de la courbe d'EQM, le premier nombre, de carottes virtuelles prélevées, situé sous la valeur seuil d'EQM autorisée (avec un biais = 5%), est choisi comme le nombre optimal de carottes à prélever. En conséquence, le prélèvement d'un nombre plus significatif de carottes n'améliore pas la capacité des carottes à être plus représentatives de la zone d'intérêt, par rapport à un biais de 5%. La **Figure 3.9** résume le processus d'échantillonnage virtuel et la méthodologie retenue pour le choix du nombre optimal de carottes à prélever de la zone d'intérêts à échantillonner.

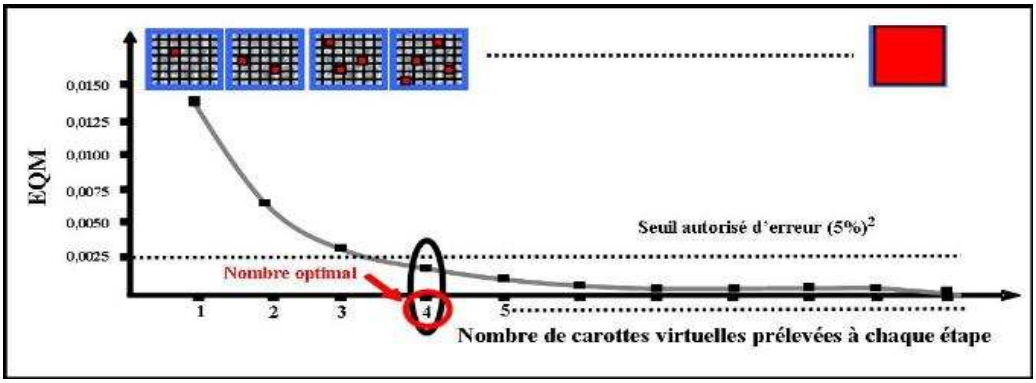


Figure 3.9 : Echantillonnage virtuel d'une zone d'intérêt. Le prélèvement aléatoire, est pratiqué 1000 fois pour chaque nombre de carottes virtuelles. Ensuite, la courbe traduisant la progression de l'erreur quadratique moyenne (EQM), en relation avec la distribution, d'échantillonnage, associée à chaque nombre de carottes, est tracée. Le premier nombre de carottes virtuelles générant un seuil d'erreur inférieur à 5%, pour estimer la vraie valeur d'indice de marquage de la zone à échantillonner, est retenu comme le nombre optimal de carotte à prélever.

Cette stratégie d'échantillonnage a été appliquée sur toutes les zones d'intérêts de l'ensemble des lames histologiques : Ki67, ß-caténine (Bcat), Bcl2 et Cycline D1 (CD1). Nous aborderons dans la partie suivante les résultats d'échantillonnage virtuel au niveau de ces zones.

3.3 Résultats

3.3.1 Mise en œuvre de la stratégie d'échantillonnage virtuel :

La Figure 3.10 illustre le fonctionnement de notre stratégie d'échantillonnage virtuel et aléatoire au niveau de deux zones d'intérêts Ki67, ayant des surfaces similaires. Les cartes de prolifération rattachées à ces zones reflètent la nature de la distribution spatiale du marqueur

Ki67. La zone d'intérêt 1 hétérogène nécessite le prélèvement de 6 carottes pour capter cette hétérogénéité. En revanche la zone d'intérêt 2, homogène selon sa carte de prolifération, requiert seulement le prélèvement de deux carottes.

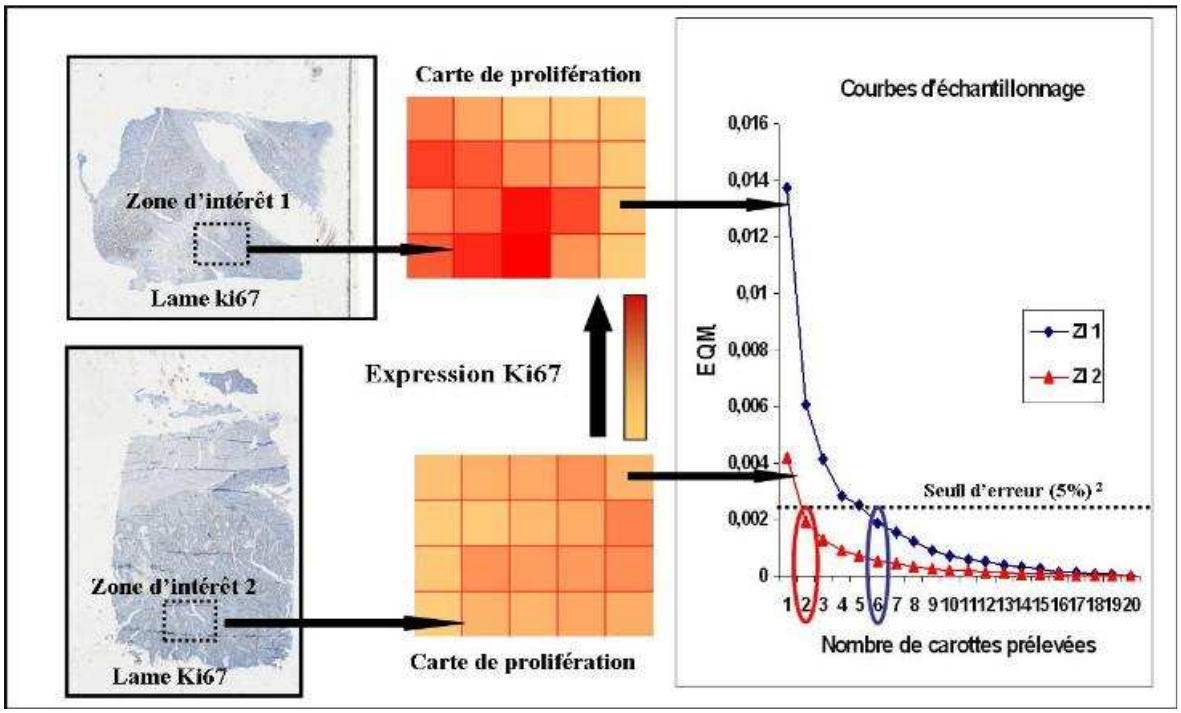


Figure 3.10: Exemple d'échantillonnage virtuel de deux zones d'intérêts de lames Ki67. La zone d'intérêt 1 hétérogène nécessite le prélèvement de 6 carottes. 2 carottes suffisent pour représenter la zone d'intérêt 2 homogène.

3.3.2 Description des zones d'intérêts à échantillonner :

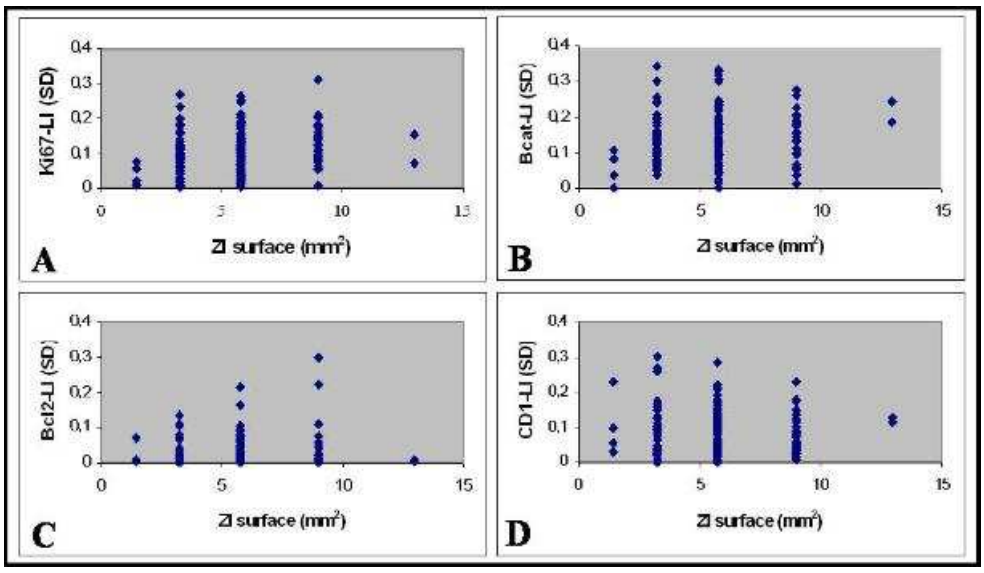


Figure 3.11: Lien entre l'hétérogénéité des marqueurs et la surface des ZI. Relation de l'hétérogénéité (écart type LI(SD)) des marqueurs tumoraux avec la surface des zones d'intérêts (ZI surface). A : Ki67. **B :** ß-caténine (Bcat). **C :** Bcl2. **D :** Cycline D1 (CD1).

La Figure 3.11 illustre la relation entre l'hétérogénéité de l'expression des marqueurs tumoraux et la surface des zones d'intérêts. L'hétérogénéité a été estimée par l'intermédiaire de l'écart type des indices de marquage (LI (SD)) mesurés individuellement sur chaque carotte virtuelle (champ). Ce paramètre traduit la distribution spatiale du marquage à travers la surface des zones d'intérêts. La Figure 3.11 montre que l'hétérogénéité des zones d'intérêts n'est pas reliée à leur surface. Des zones d'intérêts ayant des surfaces similaires expriment la plupart du temps des hétérogénéités de marquage très variées. Les faibles corrélations calculées entre la distribution des marqueurs tumoraux et les surfaces des zones d'intérêts (Tableau 3.1) confirment ces observations.

Tableau 3.1: Corrélation de l'hétérogénéité des marqueurs avec la surface des ZI. Coefficient de corrélation entre l'écart type des LI des champs des zones d'intérêts (LI (SD)) et leur surface (ZI surface).

	Ki67	ß-caténine	Bcl2	Cycline D1
Coefficient de Corrélation LI(SD) vs ZI surface	0,20788	0,10207	0,04459	-0,06213

Il semble qu'en général la distribution des marqueurs tumoraux ne soit pas liée directement avec les niveaux de marquage (Tableau 3.2). Cependant et en se référant à la Figure 3.12, nous remarquons que les zones d'intérêts à trop faible ou trop fort niveau de marquage sont relativement homogènes. Les zones les plus hétérogènes sont celles dont l'indice de marquage se situe entre 40 % et 60%, particulièrement pour les marqueurs ß-caténine (Figure 3.12.B) et Ki67 (Figure 3.12.A). L'hétérogénéité peut s'expliquer par la coexistence de champs, au sein de ces zones, ayant différents taux de marquage. Ce fait se traduit d'une part, par un indice de marquage moyen (LI : 40-60 %), des zones d'intérêts et d'autre part, une distribution spatiale hétérogène du marquage.

Tableau 3.2 : Corrélation de l'hétérogénéité des marqueurs avec le niveau de marquage des ZI. Coefficient de corrélation entre l'écart type des indices de marquage des champs des zones d'intérêts (LI (SD)) et les indices de marquage globaux de ces zones (LImoy).

	Ki67	ß-caténine	Bcl2	Cycline D1
Coefficient de Corrélation LI(SD) vs LImoy	0,03249	0,00853	-0,03545	-0,21890

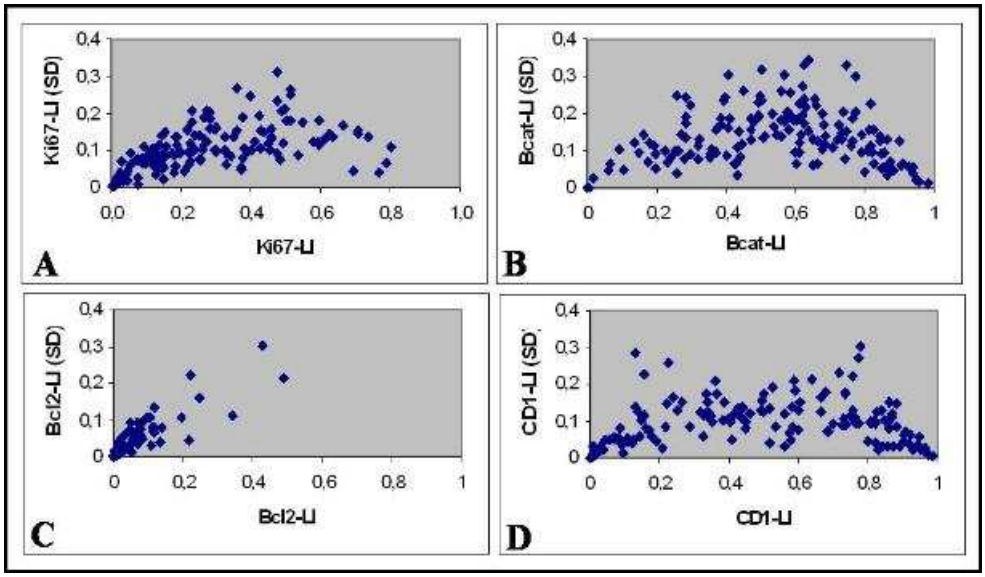


Figure 3.12: Lien entre l'hétérogénéité des marqueurs et le niveau de marquage des ZI. Relation de l'hétérogénéité (écart type LI (SD)) des marqueurs tumoraux avec l'indice de marquage des zones d'intérêts (LI). A : Ki67**. B :** ß-caténine (Bcat). **C :** Bcl2**. D :** Cycline D1 (CD1).

3.3.3 Résultats de l'échantillonnage virtuel des zones d'intérêts :

Le Tableau 3.3 montre les résultats généraux concernant le nombre de carottes à prélever au niveau des zones d'intérêts pour une représentation optimale des marqueurs tumoraux. Le nombre de carottes à prélever varie entre 1 et 16. En moyenne quatre carottes suffisent pour représenter une zone d'intérêt Ki67 ou Cycline D1. Les zones d'intérêts Bcl2 s'avèrent les plus homogènes puisque deux carottes (1,56) sont capables de représenter l'expression de ce marqueur. En revanche, une représentation adéquate de l'expression spatiale de la ß-caténine nécessite en moyenne le prélèvement de 5 carottes.

Tableau 3.3: Descripteurs généraux des Nombres de carottes à prélever au niveau des ZI.

Nombre de carottes à prélever	Ki67	ß-caténine	Bcl2	Cycline D1
nombre optimal moyen	4,01	5,42	1,56	3,40
nombre optimal min	1,00	1,00	1,00	1,00
nombre optimal max	16,00	15,00	15,00	12,00

La Figure 3.13 montre que le nombre optimal de carottes à prélever, n'est pas affecté par la surface de la zone d'intérêt à échantillonner. Des zones ayant des surfaces identiques peuvent nécessiter le prélèvement d'un nombre variable de carottes.

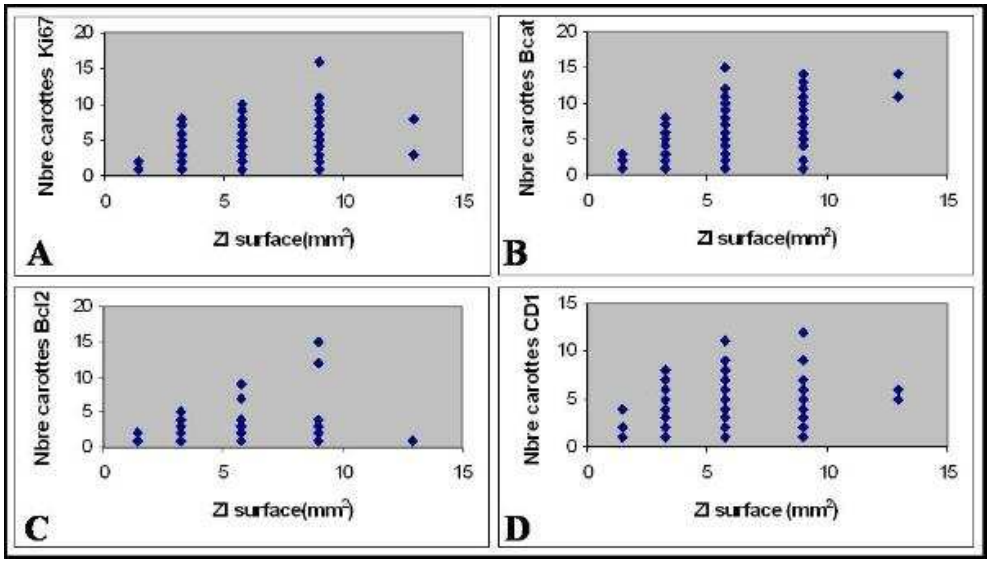


Figure 3.13: Lien entre le nombre optimal et la surface des ZI. Relation entre le nombre optimal de carottes à prélever (Nbre carottes) et la surface des zones d'intérêts (ZI surface) pour chacun des marqueurs tumoraux. A : Ki67**. B :** ß-caténine/Bcat. **C :** Bcl2**. D :** Cycline D1/CD1**.**

De manière globale, le nombre de carottes à prélever est indépendant du niveau de marquage des zones d'intérêts (Figure 3.14). Néanmoins, les niveaux de marquage, trop bas ou trop élevés, impliquent le prélèvement d'un nombre réduit de carottes. Ce constat est particulièrement vérifié pour les marqueurs Ki67 (Figure 3.14.A) et Bcl2 (Figure 3.14.C).

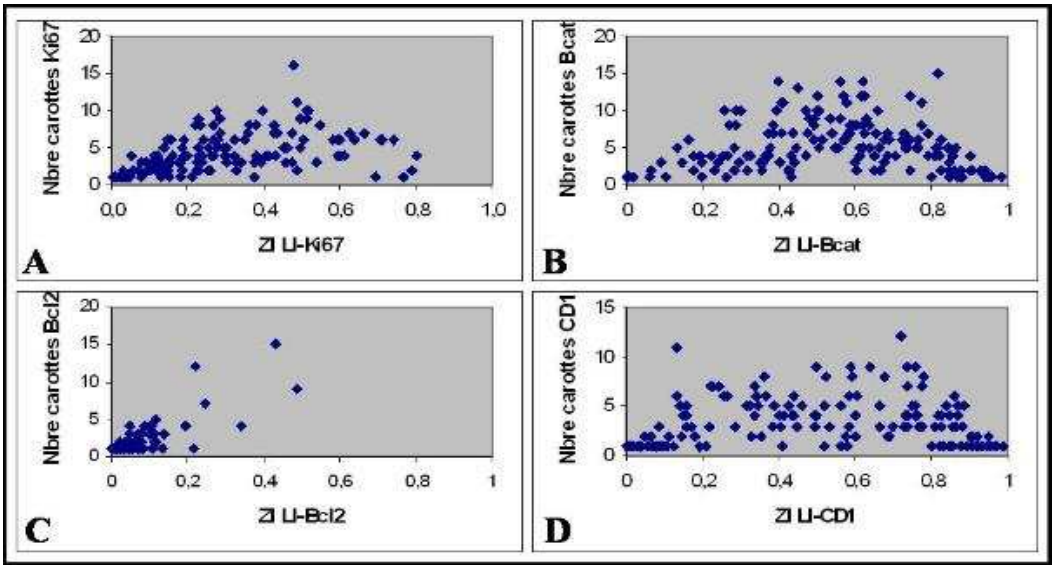


Figure 3.14: Lien entre le nombre optimal et le LI des ZI. Relation entre le nombre optimal de carottes à prélever (Nbre carottes) et l'indice de marquage (LI) des zones d'intérêts pour chacun des marqueurs tumoraux. A : Ki67**. B :** ß-caténine/Bcat. **C :** Bcl2**. D :** Cycline D1/CD1**.**

La Figure 3.15 démontre que le nombre optimal de carottes à prélever au niveau des zones d'intérêts est déterminé en fonction de la distribution spatiale du marquage. Une carotte seule suffit pour représenter les zones homogènes, ayant une dispersion minimale des indices de marquage au niveau des champs des ZI (LI (SD) faible). Le nombre de carottes à prélever augmente avec le degré d'hétérogénéité des zones d'intérêts (LI (SD) augmentant).

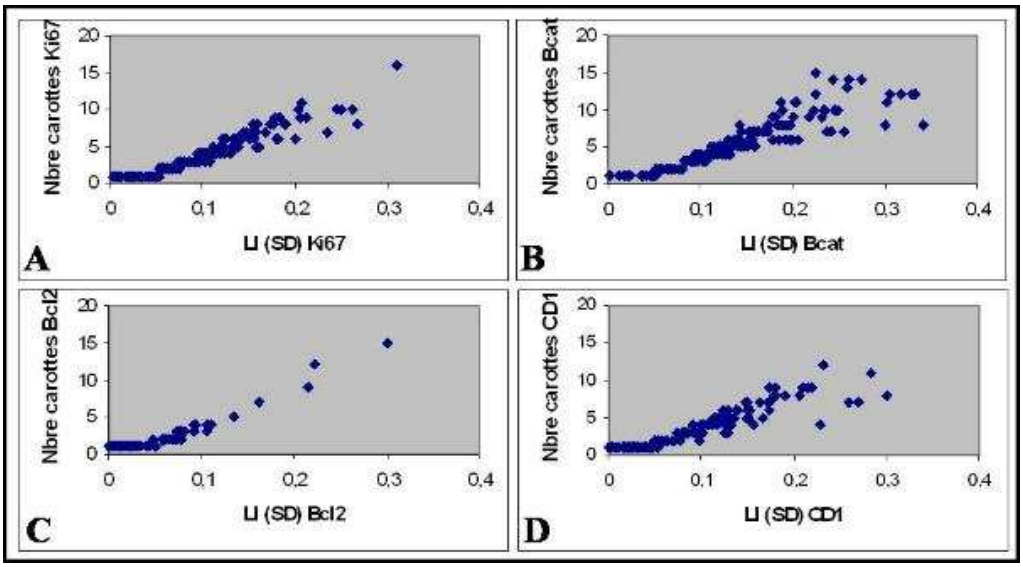


Figure 3.15: Lien entre le nombre optimal et l'hétérogénéité des marqueurs. Relation entre le nombre optimal de carottes à prélever (Nbre carottes) et l'hétérogénéité du marquage (LI(SD)) au niveau des zones d'intérêts pour chacun des marqueurs tumoraux. A : Ki67. **B :** ß-caténine/Bcat. **C :** Bcl2**. D :** Cycline D1/CD1**.**

Tableau 3.4: Relation entre le nombre optimal et la nature du marquage des ZI. Coefficient de corrélation entre le nombre optimal de carottes à prélever (Nbre Optimal) et l'indice de marquage (LI), la surface (ZI surface) et l'hétérogénéité (LI(SD)) des zones d'intérêts à échantillonner pour chacun des marqueurs tumoraux : Ki67, ß-caténine/Bcat, Bcl2 et Cycline D1/CD1.

Coefficient de Corrélation Nbre Optimal avec :	Ki67	ß-caténine	Bcl2	Cycline D1
Indice de marquage (LI)	0,49460	-0,00019	0,82105	0,14591
Surface (ZI surface)	0,35827	0,34879	0,11469	0,12364
Ecart type (LI (SD))	0,94943	0,91415	0,92948	0,92525

Le Tableau 3.4 confirme nos observations sur les figures 3.13, 3.14 et 3.15. On remarque bien que c'est le degré d'hétérogénéité, lié à la distribution spatiale du marquage, qui détermine le nombre de carottes à prélever. La surface, ou le niveau de marquage, des zones à échantillonner semble avoir un effet négligeable dans la désignation du nombre optimal de carottes à prélever.

Les exemples donnés dans la Figure 3.10 démontrent bien que c'est l'hétérogénéité de la distribution du marquage qui détermine le nombre de carottes à prélever. Ce nombre est déterminé de façon indépendante de la surface et du niveau de marquage des zones à échantillonner.

Le type des zones d'intérêts semble influencer peu le nombre de carottes à prélever (Tableau 3.5). Sauf pour les marqueurs Ki67 et Cycline D1, où les zones A en surface des tumeurs impliquent le prélèvement d'un nombre plus important de carottes par rapport aux zones B et C. Il parait que les zones A, siège de l'épithélium colique, sont plus hétérogènes, en comparaison avec les zones B et C respectivement en plein développement tumoral et à la périphérie de la tumeur. Les marqueurs Ki67 et Cycline D1 sont particulièrement exprimés au niveau des zones A, ce qui explique le nombre élevé de carottes à prélever, par rapport aux zones B et C.

Tableau 3.5: Relation entre le nombre optimal et le type des ZI. La moyenne du nombre optimal de carottes à prélever (Nbre Optimal) au niveau des zones d'intérêts de type A (surface de la tumeur), B (pleine tumeur) et C (en profondeur) pour chacun des marqueurs tumoraux : Ki67, ß-caténine, Bcl2 et Cycline D1.

	Ki67	ß-caténine	Bcl2	Cycline D1
Zones A	4,32	5,60	1,57	3,87
Zones B	3,77	5,37	1,65	2,98
Zones C	4,00	5,33	1,43	2,40

Le Tableau 3.6 résume les résultats d'échantillonnage. Il démontre que le nombre optimal de carottes à prélever est lié principalement à la nature du marqueur tumoral. On remarque que dans la plupart des zones d'intérêts (81.05%) une carotte suffit pour représenter un marqueur ayant une expression focale (expression en foyers localisés) comme le Bcl2. En revanche, deux tiers des zones d'intérêts (67,97%) requièrent le prélèvement d'un nombre plus conséquent de carottes (>3) pour représenter de manière adéquate l'expression de la ß-caténine. Les résultats concernant le Ki67 et la Cycline D1 sont plus mitigés. Ainsi, on note que dans la moitie des cas, 3 carottes sont capables de capter la distribution spatiale des marqueurs Ki67 (49,67%) et Cycline D1 (59,48%). Mais de l'autre coté la deuxième moitié des zones d'intérêts nécessite le prélèvement d'un nombre plus important de carottes.

Tableau 3.6: Relation du nombre optimal avec la nature du marqueur tumoral. Répartition des zones d'intérêts selon le nombre optimal de carottes à prélever pour chacun des marqueurs tumoraux : Ki67, ß-caténine, Bcl2 et Cycline D1.

	Ki67	ß-caténine	Bcl2	Cycline D1
Nbre optimal = 1	18,95%	11,11%	81,05%	32,68%
Nbre optimal <= 3	49,67%	32,03%	92,81%	59,48%
Nbre optimal > 3	50,33%	67,97%	7,19%	40,52%

En récapitulatif et selon les résultats obtenus, nous pensons que le nombre de 3 carottes, retenu la plupart du temps pour construire les blocs de TMA, n'est pas adéquat tout le temps pour représenter le potentiel d'hétérogénéité des zones d'intérêts. Bien que 3 carottes soient largement suffisantes pour représenter l'expression des marqueurs homogènes (Bcl2), néanmoins le prélèvement de ce nombre de carottes sous-estime parfois la distribution spatiale des marqueurs du cycle cellulaire, tels que le Ki67 et la Cycline D1. Ce nombre sous-estime aussi l'hétérogénéité de l'expression de la ß-caténine qui est le marqueur clé pour étudier le cancer colorectal.

3.4 Discussion

L'objectif principal de ce chapitre met en exergue la technologie des TMA pour mener des études IHC, notamment dans le cas des marqueurs tumoraux du cancer colorectal tout en se focalisant plus particulièrement sur la simulation du prélèvement des carottes pour assurer une stratégie d'échantillonnage optimale des blocs de tumeurs donneurs.

La discussion se portera sur :

- le choix des marqueurs tumoraux,
- la méthodologie de prélèvement des carottes par simulation comme outil d'évaluation de la technique des TMA, et enfin
- la présentation de l'approche de TMA virtuel, qui découle du concept de carottes virtuelles, comme outil d'assistance à la conception ou l'exploitation des TMA réels.

3.4.1 Choix des marqueurs tumoraux

La prolifération cellulaire accrue est l'un des signes de la transformation maligne des tumeurs. Elle est classiquement employée pour établir le pronostic lié à différents types de cancers

[Salminen et al 2005]. Le Ki67 est le marqueur tumoral le plus utilisé en IHC pour analyser la prolifération cellulaire. Une expression élevée de ce marqueur est souvent associée à un pronostic réservé [Pollack et al 2004] [Bui et al 2004] [Tisell at al 2003]. A ce titre, le marqueur Ki67 est un excellent candidat pour l'évaluation et la validation des nouvelles technologies, telles que la technologie des TMA, comme outil d'investigation sur l'expression IHC des marqueurs tumoraux. Le Ki67 a été le marqueur tumoral de choix pour des études comparatives entre l'utilisation des coupes histologiques TMA et les coupes histologiques classiques [Chiesa-Vottero et al 2003]. Il a été utilisé pour valider le recours à la technologie des TMA pour l'étude de différents types de tumeurs, tels que le cancer de la vessie [Nocito et al 2001], le cancer de la prostate [Singh et al 2004], les tumeurs fibroblastiques [Engellau et al 2001] ou le cancer colorectal [Hendriks et al 2003]. En accord avec cette logique, le Ki67 a été retenu comme le premier marqueur à intégrer au niveau de notre stratégie d'échantillonnage.

Malgré ce consensus établi pour l'usage du marqueur Ki67 dans de tels protocoles d'évaluation, on peut se poser la question sur la généralisation de ces résultats pour adopter la technologie des TMA à l'étude des autres marqueurs tumoraux. Le Ki67 est une protéine à expression spatiale hétérogène, il est parfois couplé à un autre marqueur à expression focale et localisée. Ainsi Fernebro et al joignent le marqueur p53 (suppresseur de tumeurs) au Ki67 pour valider la technique des TMA [Frenebro et al 2002]. Merseburger et al utilisent de leur coté, à la fois le p53 et la Bcl2 pour évaluer les TMA [Merseburger et al 2003]. Suivant cette approche, nous avons intégré le marqueur anti-apoptotique Bcl2 dans notre stratégie d'échantillonnage. Le choix a été porté sur le Bcl2 pour deux raisons : sa localisation focale et son rôle comme témoin indirect de l'absence de l'expression des gènes suppresseurs de tumeurs, tels que le p53 muté dans 70 % des cancers du colon [Brunz et al 1999].

Certains chercheurs ont intégré les marqueurs tumoraux ayant un rôle particulier dans certains types de cancers pour valider la technologie TMA. Par exemple, Torhorst et al ou Ginestier et al trouvent que 3 carottes sont suffisantes pour l'analyse de l'expression des récepteurs à l'œstrogène (ER) et à la progestérone (PR) dans le cas du cancer du sein [Ginestier et al 2002] [Torhorst et al 2001].

En ce qui nous concerne, nous avons joint la ß-caténine et la Cycline D1, aux marqueurs ki67 et Bcl2, notamment pour leur rôle clé dans le développement tumoral du cancer colorectal.

3.4.2 L'approche virtuelle par simulation de prélèvement de carottes TMA : Un outil efficace pour l'évaluation de la technologie TMA :

Les résultats obtenus prouvent que notre approche de prélèvement virtuel de carottes tient réellement compte de l'hétérogénéité du matériel biologique à échantillonner. Ils démontrent que le nombre optimal de carottes à prélever est avant tout lié à la nature de la distribution spatiale des marqueurs tumoraux, bien que trois carottes suffisent pour représenter l'expression des marqueurs tumoraux dans la moitié des zones d'intérêts. Néanmoins, un nombre plus important de carottes est nécessaire pour représenter des zones d'intérêts hétérogènes.

Le nombre de carottes à prélever est aussi fonction du marqueur tumoral étudié. Par exemple, une à deux carottes sont largement suffisantes dans le cas du Bcl2. Ces résultats confirment ceux de Torhorst et al, concernant l'étude des marqueurs tumoraux à expression focale, tels que le p53 [Torhorst et al 2001]. Les marqueurs tumoraux Ki67 et Cycline D1 nécessitent le prélèvement de 3 à 4 carottes en moyenne, ce qui parait en accord avec plusieurs travaux concernant les marqueurs de prolifération cellulaire [Hoos et Cordon-Cardo 2001] [Engellau et al 2001] [Fernebro et al 2002] [Hendriks et al 2003]. Le prélèvement en moyenne de 5 carottes est nécessaire pour une représentation adéquate de l'expression de la ß-caténine.

A première vue, ces résultats paraissent conformes aux travaux qui stipulent que 3 à 4 voir 5 carottes, au maximum, sont requises pour la prise en compte de l'expression des différents marqueurs tumoraux [Nocito et al 2002] [Ginestier et al 2002] [Hoos et Cordon-Cardo 2001] [Engellau et al 2001] [Giltnane et Rimm 2004] [Simon et al 2003] [Fernebro et al 2002] [Hendriks et al 2003]. Cependant, nos résultats démontrent clairement que le nombre de carottes est très variable, il dépend en premier lieu de la nature hétérogène des zones à échantillonner. Ce nombre varie entre un et seize, il est faible pour les zones homogènes, par contre les zones très hétérogènes requièrent le prélèvement de plusieurs carottes (jusqu'à 16 carottes).

Nous pensons que le recours à un nombre fixe de carottes pour représenter toutes les zones d'intérêts n'est pas la stratégie la plus judicieuse à adopter. Par exemple, le prélèvement systématique de trois carottes au niveau de chaque zone d'intérêt, comme préconisé par les concepteurs de la technique des TMA [Kononen et al 1998], revient en effet à sur représenter certaines zones (homogènes) et certains marqueurs à distribution focale (Bcl2). En revanche, ce nombre (3) de carottes s'avère insuffisant pour capter le potentiel d'hétérogénéité de certaines zones d'intérêts.

Nos résultats sont en accord avec les travaux plus récents sur la technologie TMA, considérant qu'un nombre limité d'échantillons, n'est pas adapté pour prendre en considération la nature hétérogène de la distribution de certains marqueurs, tels que le ki67 et les Cyclines D, le Bcl2 et le p53 [Chiesa-Vottero et al 2003] [Charpin et al 2004] [Singh et al 2004]. Nos résultats sont particulièrement conformes à ceux obtenus par Singh et al qui recommandent la modulation du nombre d'échantillons en fonction de l'hétérogénéité des blocs de tumeurs donneurs. Ils notent qu'il faut prélever de 5 à 12 carottes pour tenir compte de l'hétérogénéité des marqueurs tumoraux Ki67 et AR [Singh et al 2004].

Certains auteurs résument bien l'adaptation du nombre de carottes à prélever avec la nature de la zone à échantillonner. Ils notent que le prélèvement de 3 carottes est adapté pour la plupart des zones d'intérêts (58,49% dans notre cas § 3.2.2). Néanmoins, il faut prélever jusqu'à 14 carottes, pour représenter la distribution hétérogène de certains marqueurs tumoraux (ß-caténine), dans l'étude du profil moléculaire des fronts d'invasion tumorale [Engellau 2004] [Waterworth et al 2004] [Kajdacsy-Balla et al 2007].

3.4.3 Le concept de « TMA virtuel » : un outil prometteur pour la construction des TMA et la validation de nouvelles méthodologies d'exploitation des TMA :

La stratégie de prélèvement de carottes virtuelles, que nous avons présentée dans ce chapitre, tire profit des avantages des développements actuels des outils de microscopie virtuelle et des stations d'acquisitions d'images haut-débit. Cette stratégie permet la simulation, par des essais

multiples et non limités, de prélèvements réels des carottes. Elle a l'avantage d'éliminer les nombreuses étapes, longues et coûteuses, nécessaires pour construire un bloc de TMA réel. Notre méthodologie de simulation nous permet de tester toutes les combinaisons possibles de prélèvement de carottes réelles. Elle n'est limitée que par la taille du matériel biologique à échantillonner. A cet égard, elle apparaît plus exhaustive par rapport à d'autres études utilisant l'approche de combinaison entre un jeu limité de carottes réelles [Camp et al 2000] [Rubin et al 2002]. Ces auteurs se sont adonnés à des jeux de permutations possibles, entre 5 carottes [Camp et al 2000] ou 10 carottes [Rubin et al 2002], pour trouver le nombre de carottes nécessaire pour représenter une zone d'intérêt.

Le concept de TMA virtuel introduit au niveau de ce chapitre nous permet de proposer différents plans expérimentaux de TMA, pour la construction d'un bloc de TMA réel. Ainsi, et à l'aide d'outils statistiques ou de traitement d'images rapides, nous pouvons valider le bloc de TMA avant même sa construction effective.

Dans ce contexte, nous avons présenté une méthodologie de simulation d'échantillonnage pour construire des blocs TMA spécifiques pour l'étude du cancer colorectal. Cette méthodologie a été focalisée en premier lieu sur la délimitation des zones d'intérêts en se basant à la fois sur les annotations anatomopathologiques et sur les cartes de densité de prolifération cellulaire. Ensuite, le nombre de carottes à prélever, au niveau de chaque zone d'intérêt, a été déterminé selon la distribution spatiale de l'expression de certains marqueurs tumoraux du cancer colorectal. Nous avons ainsi validé statistiquement le recours à la technologie des TMA pour l'étude du cancer colorectal.

Kajdacsy et al soulignent dans une revue récente, que le plan de construction de TMA doit être adapté aux objectifs du projet de recherche [Kajdacsy et al 2007]. Ils notent qu'il existe différents types de TMA (plans TMA), tels que :

- le TMA pour établir les indices de pronostic en relation avec le suivi clinique des patients,
- le TMA pour étudier la progression tumorale d'un type donné de cancer, ce genre de TMA doit contenir tout le spectre des lésions tumorales associées au cancer étudié,

- le TMA pour étudier le grade tumoral du cancer incluant tous les stades de progression histologique de la tumeur,
- le TMA pour mener des investigations sur la distribution spatiale de certains marqueurs tumoraux. Ce type de TMA est construit pour étudier l'hétérogénéité tumorale à l'égard des marqueurs de l'hypoxie cellulaire par exemple. Il requiert le prélèvement d'un nombre important de carottes par bloc de biopsie donneur.

Bien que la technologie TMA a été conçue spécialement pour préserver le matériel biologique, en ne prélevant que de petits échantillons dans chaque bloc donneur, la réalisation d'un grand nombre de TMA pour répondre à différentes problématiques de recherches comme suggéré ci-dessus par Kajdacsy et al [Kajdacsy et al 2007], conduirait à l'épuisement rapide du précieux stock de blocs de tumeur d'archives.

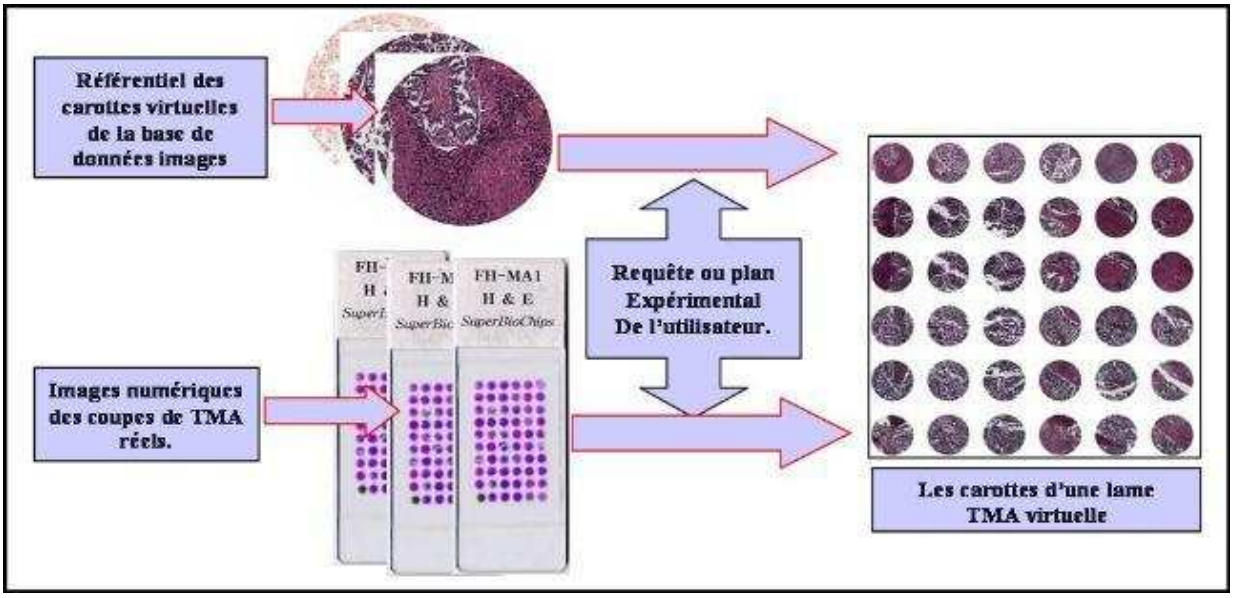


Figure 3.16: Présentation du concept du TMA virtuel. Le TMA virtuel consiste à construire de nouvelles images de lames TMA à partir du référentiel de carottes réelles (coupes TMA) ou de carottes virtuelles (à partir de coupes histologiques classiques).

Nous proposons de réadapter notre concept de TMA virtuel pour générer des blocs de TMA virtuels à partir de lames histologiques classiques ou de lames de TMA réels. Nous pouvons ainsi regrouper virtuellement des coupes de carottes réelles (coupes TMA) ou virtuelles (à partir de coupes histologiques) pour répondre à une problématique de recherche donnée. La Figure 3.16 illustre la construction d'un bloc de TMA virtuel à partir des images de coupes

TMA, ou de coupes histologiques classiques, disponibles dans une base de données images. Nous avons développé un outil dédié à la fabrication de TMA virtuels. Cet outil et son guide d'utilisation sont abordés de façon plus détaillée dans l'annexe B. Nous exploiterons ce concept de TMA virtuel pour proposer une nouvelle méthodologie d'analyse automatique de l'expression des marqueurs tumoraux au niveau des carottes virtuelles.

3.5 Conclusion

Nous avons proposé au niveau de ce chapitre un protocole d'échantillonnage concernant le nombre optimal de prélèvements de carottes à réaliser pour obtenir un échantillon réellement représentatif en tenant compte de l'hétérogénéité des blocs de biopsies donneurs. Le protocole se base sur une approche de simulation pour prélever de façon virtuelle les carottes. Un échantillonnage aléatoire de ces carottes virtuelles nous a permis la détermination du nombre de carottes à prélever dans chaque zone d'intérêt en relation avec son hétérogénéité.

Cette approche, basée sur le concept de carottes virtuelles, a été appliquée pour l'étude de l'échantillonnage de certains marqueurs tumoraux du cancer colorectal. Les résultats obtenus démontrent que le nombre de carottes à prélever de chaque zone d'intérêt est avant tout fonction de la zone à échantillonner. Les marqueurs à distributions hétérogènes, tels que le Ki67 ou la ß-caténine, requièrent le prélèvement d'un nombre plus important d'échantillons par rapport aux marqueurs à expression focale tels que le Bcl2.

En conclusion, le recours à trois carottes par zone d'intérêt, comme utilisé par plusieurs auteurs, n'est pas adapté pour la représentation de toutes les zones à échantillonner. S'il est suffisant pour représenter les zones homogènes, il se montre inadapté dans le cas des zones hétérogènes.

Les concepts de « TMA virtuel » et de « carottes virtuelles » ont été introduits pour valider les différents plans de construction des TMA. L'approche des carottes virtuelles sera utilisée dans la deuxième partie de ce manuscrit, pour illustrer et valider des méthodologies dédiées à l'analyse automatique et l'exploitation des coupes TMA. Une nouvelle technique de classification couleur sera exposée dans cette deuxième partie à travers la quantification des marqueurs tumoraux au niveau des images de carottes virtuelles.

Chapitre 4

La couleur

La complexité des mécanismes physiologiques qui régissent la vision humaine a donné lieu à la formulation de plusieurs théories ou modèles pour expliquer la perception humaine des couleurs. Il existe principalement quatre modèles : le modèle du tri-stimulus, le modèle des couleurs opposées, le modèle des compartiments et le modèle perceptuel.

En parallèle à ces théories, toute une série d'expérimentations a été menée pour comprendre les règles physiques de la couleur. Ces travaux ont permis l'émergence d'une science nouvelle qui s'attache à la formulation des lois fondamentales, des techniques de représentation et de reproduction des couleurs : il s'agit de la colorimétrie.

La colorimétrie a connu un essor considérable, notamment grâce aux expériences d'égalisation de la couleur liées au principe de la synthèse additive des couleurs, menées par la commission internationale de l'éclairage « CIE ». Cette organisation est à l'origine des fonctions colorimétriques CIE et des systèmes couleur CIE de référence qui en découlent.

Nous allons présenter au niveau de ce chapitre la physiologie de la vision humaine des couleurs et les principaux modèles expliquant celle-ci. Puis, nous exposerons le principe de la synthèse additive des couleurs. Enfin, nous aborderons les travaux de la CIE dans le domaine de la colorimétrie, notamment pour développer et normaliser les systèmes couleur dits de référence.

4.1 Introduction

Depuis l'antiquité, plusieurs savants et artistes ont tenté de donner une explication rationnelle au phénomène de la couleur. Les premières tentatives d'explication étaient signées par les philosophes grecs. Pythagore, au cinquième siècle avant notre ère, émettait la théorie du « feu visuel » selon laquelle l'œil émet un rayon « le quid » qui permet la vision de la couleur des objets. Aristote, un siècle plus tard, considérait la couleur comme le produit d'un « feu externe ». A l'opposé de Pythagore, il s'agissait, selon lui, de particules infimes de l'objet observé, qui en se frottant, émettaient un rayon lumineux coloré.

Néanmoins, il fallait attendre la fin du dix-septième siècle pour qu'une première explication scientifique du concept de la couleur, soit formulée. Le mérite revient sans doute au physicien britannique Isaac Newton (1642-1727), qui met fin aux interprétations approximatives de la couleur. C'est en 1672 que Newton a réussit à poser les fondements scientifiques sur la nature électromagnétique de la couleur en décomposant la lumière blanche à travers un prisme. Il a prouvé par cette expérience que la lumière blanche visible est la synthèse additive d'une multitude d'ondes monochromatiques représentant le spectre des couleurs pures (Figure 4.1). D'ailleurs, si l'on pose un deuxième prisme sur le trajet optique de ce spectre de couleur on reconstitue à nouveau la lumière blanche.

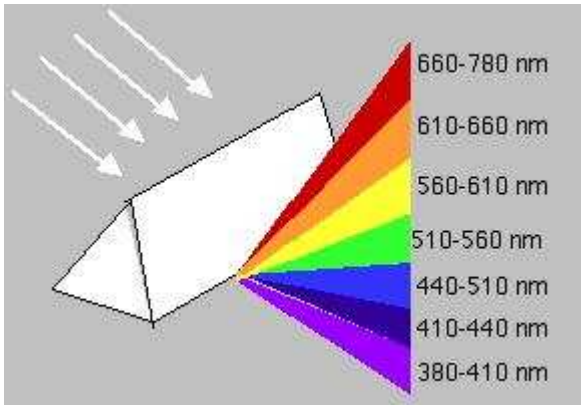


Figure 4.1: Expérience de Newton. Décomposition de la couleur blanche par un prisme.

En réalité ce que nous qualifions comme couleur n'est autre que le stimulus lumineux correspondant à la partie visible des ondes électromagnétiques, l'œil humain est sensible et ne

peut percevoir que les longueurs d'onde comprises entre 380 nm et 780 nm (Figure 4.2). Cet intervalle représente toute la gamme du spectre visible des couleurs, il s'étend du violet au rouge en passant par toutes les nuances des couleurs intermédiaires.

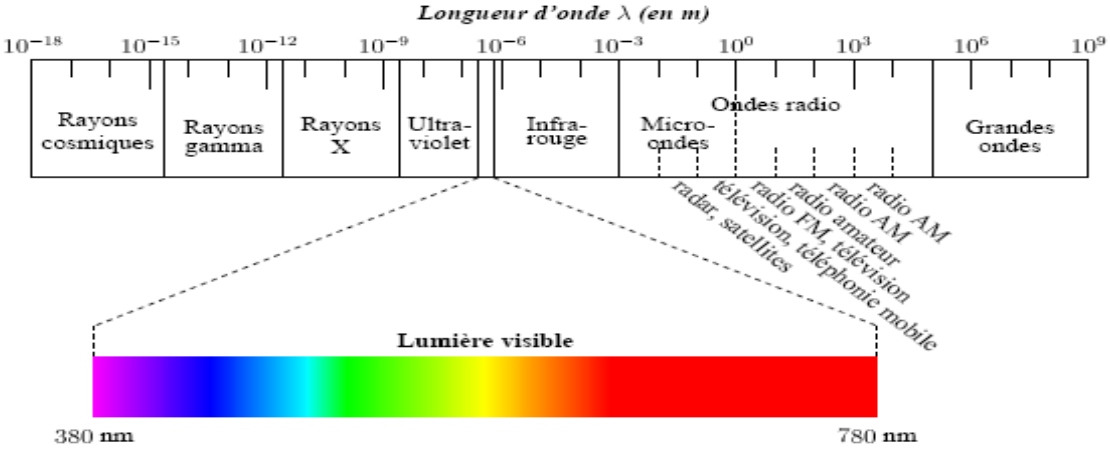


Figure 4.2: Le spectre visible des couleurs. Etendue spectrale des couleurs visibles.

Au début du dix-neuvième siècle, Tomas Young (1793-1829) formulait la première hypothèse sur la représentation de la couleur suite à ses travaux sur les interférences des ondes lumineuses. Il a démontré qu'il suffisait de trois couleurs primaires pour égaliser ou reconstituer pratiquement toute couleur. Les études de Young ont été poursuivies plus tard par James Clerk Maxwell (1831-1879) et Hermann Van Helmholtz. Maxwell (1856). Ce dernier a démontré que les couleurs rouge, vert et bleu, sont le jeu des primaires qui est le plus approprié pour l'égalisation couleur. Il a pu réaliser la première photographie en couleur grâce à ce principe. Helmholtz a repris et approfondi les travaux de Young. Sur la base des expériences de Maxwell, il a établi les bases de la théorie trichromatique qui stipule que tout stimulus couleur peut être reproduit par le mélange des trois stimuli de base : le rouge, le vert et le bleu. Cette théorie porte le nom de théorie de Young- Helmholtz ou théorie de synthèse additive des couleurs, elle est la base de tous les systèmes de représentation des couleurs utilisés en colorimétrie.

La colorimétrie a connu un essor important au cours de la première moitié du vingtième siècle, notamment grâce aux travaux de la commission internationale de l'éclairage « CIE ». La colorimétrie est devenue une science à part entière, elle est consacrée aux différentes

mesures concernant la couleur et s'intéresse tout particulièrement au développement de systèmes de représentation des couleurs.

Après ce bref préambule historique sur la couleur, on enchaîne ce chapitre par la présentation de la vision humaine des couleurs et les principaux modèles proposés dans la littérature concernant notre perception de la couleur.

4.2 La physiologie de la vision humaine de la couleur

La capacité de l'œil à distinguer les couleurs est assurée par la rétine, qui est une structure neuronale photosensible située au fond de l'œil. Néanmoins, plusieurs composants de l'œil interviennent avant que la lumière incidente ne parvienne à la rétine (Figure 4.3). Ces parties permettent une vision efficace des couleurs, essentiellement grâce à :

- L'iris, ou pupille, qui agit comme un diaphragme régulant la quantité de lumière qui atteint la rétine.
- Le cristallin en forme de lentille biconvexe, qui en modifiant sa forme grâce aux muscles ciliaires permet la mise au point précise des images des objets observés sur la rétine.

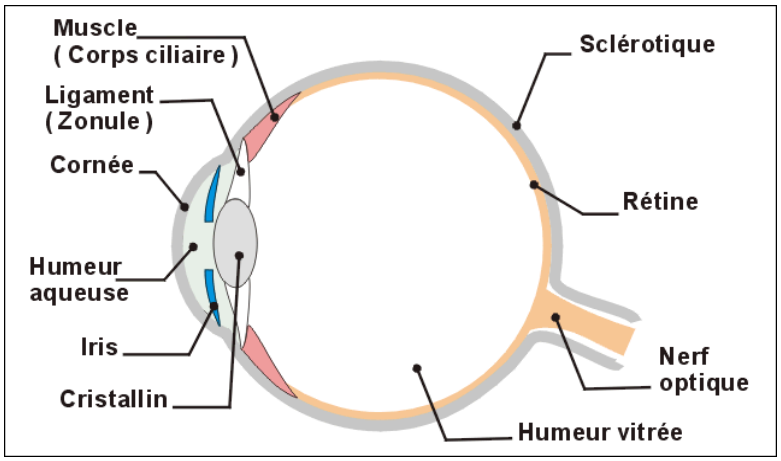


Figure 4.3: Anatomie générale de l'œil.

Après avoir parcouru le trajet optique à travers les différentes structures de l'œil, le stimulus lumineux vient se projeter sur la rétine, c'est à ce niveau que l'énergie lumineuse est transformée en signal nerveux transmis au cerveau.

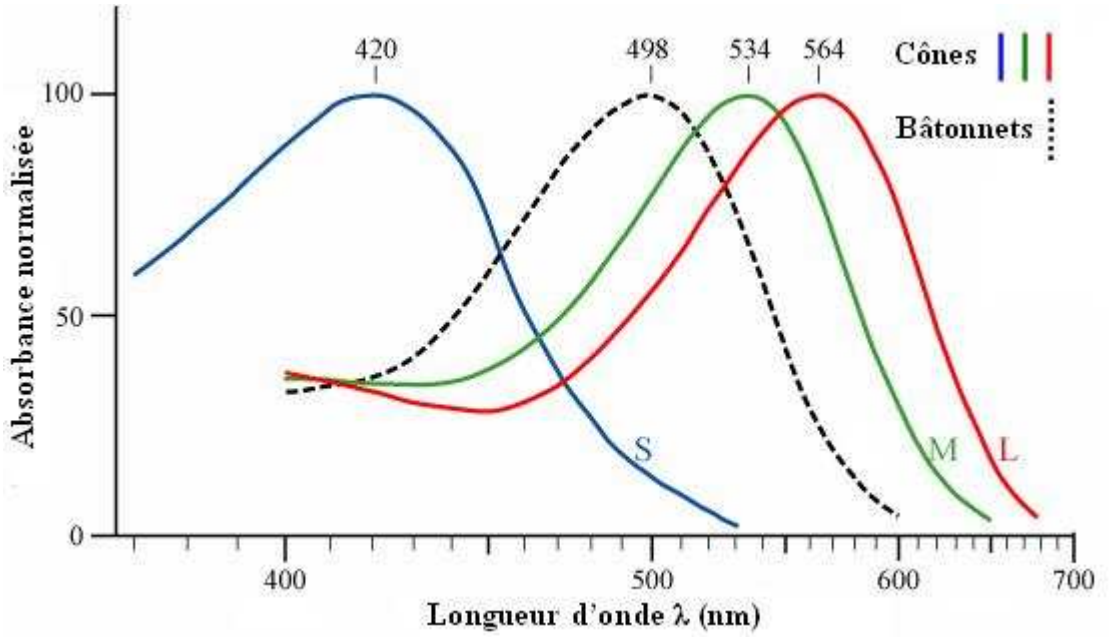


Figure 4.4: Les spectres de sensibilité des cellules de la rétine : cônes (S, M et L) et bâtonnets.

La rétine possède deux types de cellules photosensibles (Figure 4.4) [Trouvé 1989] :

- Les bâtonnets qui sont associés à la perception de l'intensité de la lumière, sont très sensibles et sont très vite saturés en lumière du jour. Ce qui explique leur responsabilité de la vision scotopique ou nocturne. Le nombre des bâtonnets varie entre 110 et 125 millions [Kowaliski 1990], ils se concentrent essentiellement au niveau de la périphérie rétinienne. La photosensibilité des bâtonnets est due à la présence d'un pigment particulier, la rhodopsine qui présente un pic de sensibilité spectrale vers 500 nm.
- Les cônes sont responsables de la perception des couleurs. Ils sont peu sensibles aux bas niveaux de lumière et sont réservés à la vision photopique ou diurne. Ils sont moins nombreux que les bâtonnets, leur nombre oscille entre 4 et 7 millions [Kowaliski 1990]. Ils se situent en grande majorité au centre de la rétine (*fovea*) avec une densité avoisinant 150000 cônes par mm^2 [Trémeau et al 2004]. Il existe trois types de cônes selon leurs sensibilités spectrales [Bowmarker et Dartnall 1980] :

- Les cônes de type S « Short-wavelength » possèdent un pigment photosensible « cyanolabe » qui présente un pic de sensibilité à 420 nm, ce qui correspond au ton de teinte violet-bleu.
- Les cônes de type M « Middle-wavelength » ont une sensibilité spectrale vers les longueurs d'ondes moyennes grâce au pigment « chlorolabe » qui manifeste un maximum d'efficacité spectrale à 534 nm correspondant à la couleur verte.
- Les cônes de type L « Long-wavelength » par l'intermédiaire de leur pigment « erythrolabe » présentent une réponse maximale dans les teintes rouges à 564 nm.

Les différentes cellules photosensibles rétiniennes (cônes et bâtonnets) réagissent à la lumière incidente. Elles engendrent en association avec d'autres types de cellules au niveau des couches les plus profondes de la rétine, des impulsions électriques véhiculées jusqu'au cerveau par l'intermédiaire du nerf optique, qui englobe environ un million de fibres optiques. Toutes ces fibres convergent vers une structure neuronale appelée corps genouillé latéral « CGL » qui se charge d'organiser les connexions avec les fibres nerveuses allant vers le cerveau. Il semble que le CGL soit le siège d'une première partie de l'analyse de l'information couleur [Vandenbroucke 2000]. Le signal couleur parvient enfin au cortex visuel situé dans la partie postérieure du cerveau. C'est à ce niveau que nous interprétons le signal couleur et donnons un sens aux scènes observées par nos yeux. Le processus de cette interprétation est un phénomène complexe et encore peu connu. *Néanmoins les connaissances acquises sur la vision humaine des couleurs ont donné naissance à plusieurs modèles, qui tentent d'expliquer les mécanismes qui régissent notre vision des couleurs. Nous exposons dans la suite les principaux modèles ou théories concernant la vision humaine des couleurs.*

4.3 Les modèles de la perception humaine de la couleur

4.3.1 Le modèle du tri-stimulus

Ce modèle s'appuie sur le fait que tout stimulus couleur qui excite la rétine est réduit à trois courbes de réponses spectrales délivrées par les trois types de cônes L, M et S. Ces réponses correspondent aux trois canaux rouge, vert et bleu. La vision des couleurs est par nature

trichromatique. Ce modèle repose aussi sur la théorie trichromatique (tri-stimulus) qui stipule que la plupart des couleurs peuvent être égalisées par synthèse additive de trois primaires fixes. Cette théorie s'appuie sur les travaux de Young, Helmotz et Maxwell [Trémeau 1998] [Lozano 1997]. Elle se base sur un principe expérimental d'égalisation de la lumière colorée par mélange additif ou superposition de trois lumières primaires : rouge, vert et bleu. En théorie, cette synthèse permet de reproduire l'ensemble des couleurs visibles par le système visuel humain. La représentation numérique des images couleur se base principalement sur ce modèle. Le système RGB qui découle de ce principe est sans doute le plus répandu.

4.3.2 Le modèle des couleurs opposées

Le nombre des fibres optiques est largement inférieur au nombre des récepteurs rétiniens (cônes LMS), il y a donc forcément une association d'informations de plusieurs cônes parvenant à chaque fibre optique. Le modèle des couleurs opposées suppose que l'information couleur est transmise selon une impulsion de luminosité et deux autres types d'impulsions opposées : rouge-vert et bleu-jaune. Ce modèle correspond à la théorie du physiologiste allemand Hering (1878), qui se base sur une représentation antagoniste des couleurs. Selon ce modèle les couleurs sont organisées selon trois axes :

- le premier axe achromatique qui décrit l'opposition noir-blanc,
- le deuxième axe correspond à l'axe d'opposition rouge-vert,
- le troisième axe correspond à l'axe d'opposition bleu-jaune.

Ce modèle est fondé sur des constatations neurobiologiques, résultat de nombreuses études psychophysiques [Boynton et Gordon 1965] [Jameson et Hurvich 1955] [Jameson et Hurvich 1961] et qui expliquent notre perception des couleurs. En effet, il existe des couleurs qui peuvent être des nuances des couleurs rouge et bleue, telles que le violet ou des nuances des couleurs jaune et rouge comme le cas de l'orange. En revanche, les couleurs rouge et vert sont des sensations exclusives, on ne peut trouver une couleur qui serait une nuance du rouge et du vert en même temps. Le même constat s'applique au couple de couleur jaune et bleue, d'où les deux axes chromatiques : rouge-vert et bleu-jaune.

4.3.3 Le modèle des compartiments

Ce modèle est un compromis entre le modèle du tri-stimulus et le modèle des couleurs opposées. Cette théorie divise le système visuel en deux compartiments : l'œil et le cerveau,

chacun possède son propre modèle couleur. Le modèle du tri-stimulus est le modèle couleur de la rétine. Celle-ci répond à la lumière incidente par trois signaux couleur par le biais des cônes L, M et S. Tandis que le cerveau code l'information couleur selon le modèle des couleurs opposées, la question qui se pose alors est : comment la transition se fait-elle entre ces deux modèles de couleur au niveau du système visuel humain ? L'hypothèse la plus simple propose des transformations linéaires des composantes du modèle du tri-stimulus pour donner lieu au modèle des couleurs opposées [Angulo-Lopez 2003]. Ainsi, l'axe des couleurs rouge-vert résulte de la différence des réponses des cônes L et M. L'axe bleu-jaune peut être obtenu par la différence de la réponse des cônes S et la somme des réponses des cônes L et M. L'axe achromatique est une combinaison des réponses des cônes L et M. Faugeras propose des transformations basées sur des opérateurs logarithmiques, car des études physiologiques ont démontré que la réponse spectrale des cônes est proportionnelle au logarithme du stimulus [Angulo-Lopez 2003]. L'espace couleur AC1C2 introduit par Faugeras [Faugeras 1979] puis développé par Bumbaca et Smith [Bumbaca et Smith 1987] découle de ce principe.

4.3.4 Le modèle perceptuel

Bien que les modèles du tri-stimulus et des couleurs opposées reposent sur une réalité physiologique, il est difficile de faire le lien entre les grandeurs physiques d'une couleur dans ces modèles et la sensation visuelle de cette même couleur [Pratt 1978]. Ce modèle repose sur les notions naturelles et intuitives liées à la couleur, la perception humaine de la couleur étant une réaction subjective au stimulus lumineux. Plutôt que de décrire une couleur selon des attributs difficiles à interpréter (tri-stimulus et couleurs opposées), il paraît plus intuitif de désigner celle-ci par les attributs : teinte, luminosité et saturation.

- La teinte correspond à la longueur d'onde prédominante de la couleur : rouge, jaune, bleue etc.,
- la luminosité traduit l'aspect clair, foncé ou terne d'une couleur,
- la saturation désigne si la couleur est pâle, vive, délavée, éclatante, etc.

Selon les définitions de l'AFNOR (Association Française de Normalisation) [Afnor 1989] :

- la teinte est l'attribut de la sensation visuelle qui a suscité des dénominations de couleur telles que : bleu, vert, jaune, rouge, pourpre, etc.
- la luminosité est l'attribut de la sensation visuelle selon lequel une surface éclairée par une source lumineuse déterminée paraît émettre plus au moins de lumière. C'est aussi le correspondant psychosensoriel (approximatif) de la grandeur photométrique « luminance lumineuse »,
- la saturation est l'attribut de la sensation visuelle permettant d'estimer la proportion de couleur chromatiquement pure contenue dans la sensation totale.

Beaucoup d'espaces couleur sont basés sur ces attributs, l'espace le plus connu est l'espace de Munsell « Atlas de Munsell » édité en 1909 par le peintre Albert Henry Munsell (1858-1918).

En parallèle aux différentes expérimentations, associées aux modèles couleur, menées pour décortiquer les mécanismes de la vision humaine des couleurs. Toute une série de travaux a été réalisée pendant les deux derniers siècles pour comprendre les règles qui régissent l'aspect physique de la couleur. Ces travaux ont permis l'émergence d'une science nouvelle qui s'attache à la formulation des lois fondamentales, des techniques de représentation et de reproduction des couleurs : il s'agit de la **colorimétrie**.

La section suivante est consacrée aux différents aspects de la colorimétrie, en particulier aux fondements expérimentaux des lois physiques régissant les systèmes de représentation des couleurs.

4.4 La colorimétrie « science de mesure de la couleur »

4.4.1 Introduction

La colorimétrie est la science de mesure de la couleur, elle s'appuie sur la théorie trichromatique. Elle a pour but d'associer à chaque couleur non pas un spectre de réflectance ou de transmittance mais trois valeurs. Cette trivariance est liée à notre mécanisme de perception des couleurs (§ 4.3.1). De ce fait, la colorimétrie qui est consacrée à la recherche des systèmes de représentation des couleurs, repose à la fois sur les propriétés physiques des

couleurs (ondes électromagnétiques) : la radiométrie, et sur les propriétés physiologiques de la vision humaine des couleurs : la photométrie.

La plupart des travaux scientifiques de normalisation liés au domaine de la colorimétrie ont été menés par la Commission Internationale de l'Eclairage « CIE ». C'est une organisation indépendante basée à Vienne, dédiée à l'étude de la lumière et de l'éclairage, *de facto* à l'étude de la couleur. Les travaux de la CIE sont publiés sous forme de recommandations qui font autorité et sont approuvées par l'ISO (International Standard Organisation).

Au cours de cette section, nous allons exposer en premier lieu le principe des expériences d'égalisation de la couleur liées à la trivariance visuelle. Ces expériences ont donné lieu à la formulation des lois de colorimétrie. Ensuite, nous enchaînerons sur la présentation des fonctions colorimétriques de la CIE et les systèmes couleur CIE de référence qui en découlent.

4.4.2 La trivariance visuelle

4.4.2.1 Principe

Les peintres et les teinturiers ont remarqué depuis l'antiquité qu'il suffisait de trois couleurs pour reconstituer l'ensemble des teintes. L'analyse de l'apparence visuelle de la couleur confirme ce caractère tridimensionnel de la couleur. Ce fait est lié à la physiologie de notre système de vision, qui distingue les couleurs grâce aux trois types de cellules photosensibles (cônes LMS). Les travaux de Young, Helmotz et Maxwell démontrent que la plupart des lumières colorées peuvent être reproduites par le mélange, en quantités appropriées de trois lumières ou primaires. Cependant, le mérite revient à Maxwell (1855) qui a définit le premier les couleurs : rouge, vert et bleu comme étant les primaires les mieux appropriées pour la synthèse ou la reproduction de l'ensemble des stimuli couleurs.

4.4.2.2 Expériences d'égalisation des couleurs

L'expérience d'égalisation ou d'appariement des couleurs consiste à présenter à un observateur donné, un champ visuel plat et blanc divisé en deux parties à l'aide d'un écran noir. L'observateur doit égaliser visuellement un stimulus couleur projeté sur la première

moitié du champ visuel, par le réglage des quantités de trois lumières dites primaires projetées sur l'autre moitié du champ visuel. Les trois couleurs primaires utilisées sont le rouge, le vert et le bleu, en accord avec la sensibilité spectrale des trois types de cônes de la rétine. L'observateur utilise sa vision comme appareil de zéro, il règle l'intensité ou la quantité des primaires afin de ne percevoir aucune différence visuelle entre les deux plages du champ visuel à égaliser (Figure 4.5). Signalant que les expériences sont menées pour égaliser les couleurs ou stimuli monochromatiques tout au long du spectre du visible : de 380 nm à 700 nm, l'expérience d'égalisation peut être modélisée ainsi par la relation :

$$S(\lambda) \equiv r(\lambda).R + g(\lambda).G + b(\lambda).B$$

S(λ) représente la répartition spectrale du stimulus couleur à égaliser, λ est la longueur d'onde et varie entre 380 et 700 nm, R, G et B sont les trois primaires et r, g et b correspondent à leurs quantités respectives nécessaires pour l'égalisation du stimulus S (λ). Le signe ≡ désigne l'équivalence visuelle entre deux stimuli couleurs.

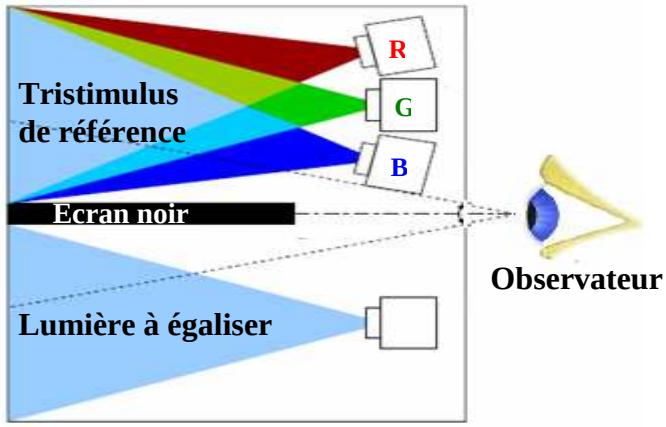


Figure 4.5: Expériences d'égalisation visuelle des couleurs.

4.4.2.3 Lois colorimétriques de la trivariance visuelle

Selon les expériences d'égalisation présentées dans le paragraphe précédent, il est possible de déterminer les quantités des trois primaires RGB nécessaires à égaliser toutes les longueurs d'ondes monochromatiques de la lumière visible. Sachant que toute couleur n'est autre qu'une somme de divers rayonnements monochromatiques, il est possible dès lors de calculer les quantités des primaires pour égaliser cette couleur. C'est-à-dire, qu'il est possible de

déterminer les composantes trichromatiques de n'importe quelle couleur pour le jeu des trois couleurs primaires. Donc, une couleur quelconque peut être représentée tout naturellement dans un espace couleur à trois dimensions correspondant aux vecteurs des trois couleurs primaires. Cependant, les lois colorimétriques régissant les espaces couleur à trois dimensions ont été formulées par Grassmann en 1853 [Grassman 1853] et complétées par Abney en 1913 [Abney 1913]. En effet, ces lois ont permis d'appliquer aux mélanges des couleurs les propriétés des égalités algébriques telles que : la réflexivité, la symétrie, l'additivité, l'associativité et la transitivité. Ainsi, l'espace colorimétrique possède les mêmes propriétés que les équations en arithmétique concernant les mélanges additifs des couleurs.

Une couleur notée C peut être définie au niveau d'un espace couleur à trois dimensions par la relation suivante :

$$C = r.R + g.G + b.B$$

Les paramètres r, g et b représentent les quantités relatives des primaires R, G et B. Ils déterminent les coordonnées trichromatiques de la couleur C dans l'espace couleur engendré par les couleurs primaires RGB. Notons que le mélange unitaire des trois primaires R, G et B produit la couleur blanche. Ce principe est à l'origine de tous les systèmes de représentation des couleurs.

Après ce bref exposé sur le principe de la trivariance visuelle, qui est à l'origine du développement des différents espaces colorimétriques. Nous nous attarderons dans la section suivante, sur la contribution de la CIE dans le domaine de la colorimétrie où, nous mettrons l'accent en particulier sur les fonctions colorimétriques et les différents espaces couleur normalisés par cette organisation.

4.5 Les travaux de normalisation de la CIE

4.5.1.1 Introduction

Pour rendre possible la comparaison des résultats d'expériences colorimétriques réalisées, par divers scientifiques, à des endroits et à des époques différents, il s'est avéré nécessaire de définir : des conditions, des paramètres communs et des références pour mener les

expériences, notamment, pour le choix des caractéristiques des trois primaires et des fonctions colorimétriques qui en découlent. Ce travail de normalisation a été effectué essentiellement, par la CIE et publié sous forme de recommandations. Parmi les nombreuses références, cette organisation a défini un observateur colorimétrique de référence en 1931. Le but était de définir un observateur ayant un comportement visuel moyen au sens statistique du terme. Cet observateur de référence est nécessaire afin d'établir des méthodes objectives et statistiques pour l'évaluation des couleurs, également lors des expériences d'égalisation visuelle des lumières colorées. Ainsi, dans le cas de l'égalisation trichromatique de la couleur, les conditions expérimentales sont strictes mais simplifiées : elles se basent sur un champ de vision de 2° ou de 10°, entouré d'un fond neutre (achromatique). Bien sûr, cela ne correspond pas le plus souvent aux situations naturelles de la vision humaine de la couleur, mais c'est avec ce formalisme très simplifié que la plupart des lois de la colorimétrie sont établies.

Nous allons aborder dans ce qui suit les différents travaux de normalisation de la CIE pour définir les fonctions colorimétriques régissant les lois physiques associées à la présentation des couleurs.

4.5.1.2 Les fonctions colorimétriques r(λ), g(λ) et b(λ)

La CIE a effectué une synthèse des expérimentations réalisées sur l'égalisation des couleurs, notamment les travaux de Guild et wright [Sharma et Trusell 1997] [Moroney et Fairchaild 1995]. Ainsi, la CIE a défini en 1931 les fonctions colorimétriques (Figure 4.6) relatives aux trois stimuli monochromatiques suivants [CIE 1986] :

- une couleur primaire rouge avec une longueur d'onde de 700 nm,
- une couleur primaire verte avec une longueur d'onde de 546,1 nm,
- une couleur primaire bleue avec une longueur d'onde de 435,8 nm.

Ces fonctions colorimétriques appelées encore composantes trichromatiques spectrales sont notées : r(λ), g(λ) et b(λ) et traduisent le comportement d'un observateur de référence pour égaliser l'ensemble des stimuli monochromatiques compris entre 400 nm et 700 nm. Notons que l'observateur de référence a été défini pour un champ de vision sous un angle d'observation de 2°.

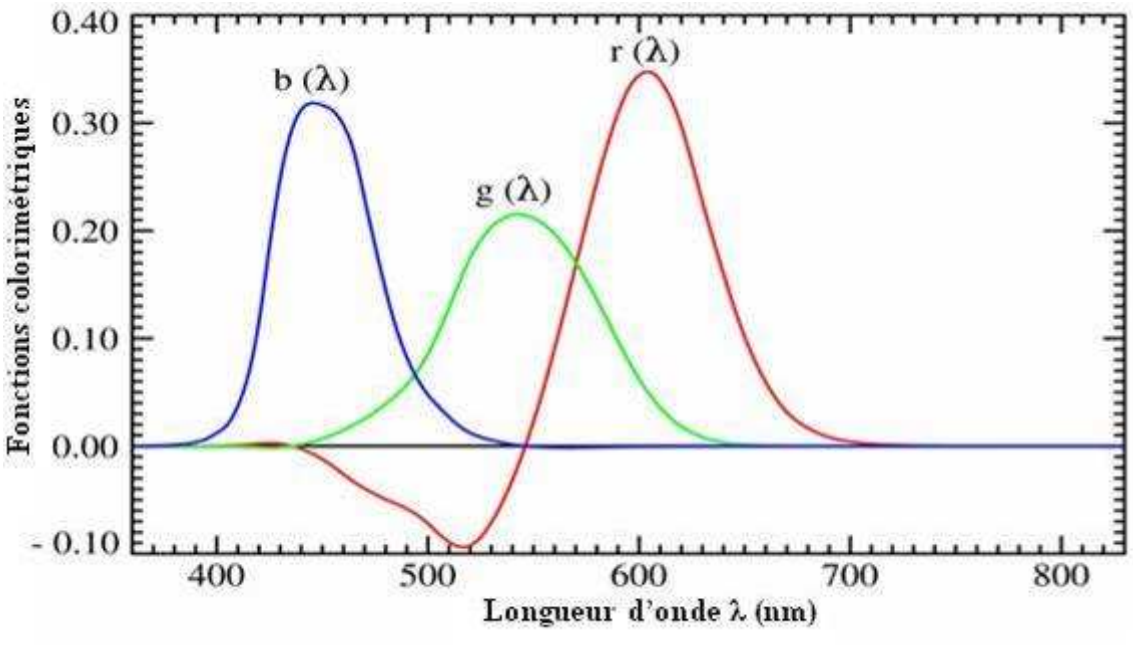


Figure 4.6: Fonctions colorimétriques r(λ), g(λ) et b(λ) CIE 1931.

Signalant que la fonction colorimétrique r(λ) présente une partie négative, ceci constitue une entorse au principe du mélange additif des couleurs spécifiées pour l'égalisation des couleurs. Prenons par exemple la couleur monochromatique bleu-vert ayant une longueur d'onde de 500 nm, on remarque très vite qu'il est impossible de reproduire cette couleur avec des composantes positives pour l'ensemble des trois primaires RGB. La solution ou l'artifice, consiste à rajouter à cette couleur à égaliser une certaine quantité de couleur rouge, de ce fait au lieu de noter :

$$S(\lambda) \equiv r(\lambda).R + g(\lambda).G + b(\lambda).B$$

On note :

$$S(\lambda) + r(\lambda).R \equiv g(\lambda).G + b(\lambda).B$$

Puisque l'égalisation des couleurs respecte le principe d'additivité on obtient :

$$S(\lambda) \equiv - r(\lambda).R + g(\lambda).G + b(\lambda).B$$

On pourrait en principe définir tout autre jeu de couleurs primaires afin d'être conforme au principe de la synthèse additive des couleurs. Mais il est pratiquement impossible de trouver les trois primaires qui satisfassent ce principe, pour certains jeux de primaires on est ramené à introduire deux composantes négatives pour égaliser certains stimuli couleur. Cependant, l'expérience prouve que la plage des égalisations purement positives est d'autant plus grande que les couleurs primaires sont espacées entre elles au niveau du spectre des couleurs. Pour

cette raison on choisit une primaire au milieu du spectre : la composante verte, et une primaire sur chaque extrémité du spectre : les composantes bleue et rouge.

Selon les lois établies par Grassmann et Abney [Trouvé 1989], les fonctions colorimétriques conduisent à la conception d'un espace tridimensionnel permettant la représentation géométrique des couleurs. L'espace couleur RGB ainsi engendré est un espace vectoriel à trois dimensions, où les trois primaires R, G et B sont représentées par trois vecteurs unitaires et orthonormés $\vec{R}$ R$^{\rightarrow}$ G$\vec{G}^{\rightarrow}$ et B$\vec{B}^{\rightarrow}$. Si l'on choisit O comme origine, toute couleur C produite par synthèse additive peut être formulée dans cet espace par le vecteur $\overrightarrow{OC}$:

$$\overrightarrow{OC} = R\vec{R} + G\vec{G} + B\vec{B}$$

Ou R, G et B sont appelées les coordonnées trichromatiques. D'une façon plus simple la couleur C est symbolisée le plus souvent dans l'espace RGB par le triplet (R, G, B). Rappelons que le mélange unitaire des trois primaires R, G et B produit la couleur blanche. Cet espace est appelé le système de primaires RGB CIE 1931, c'est l'espace RGB de référence. Ce système a été étendu par la CIE en 1964 en complétant les fonctions colorimétriques CIE 1931 pour un angle de vision plus large (10°).

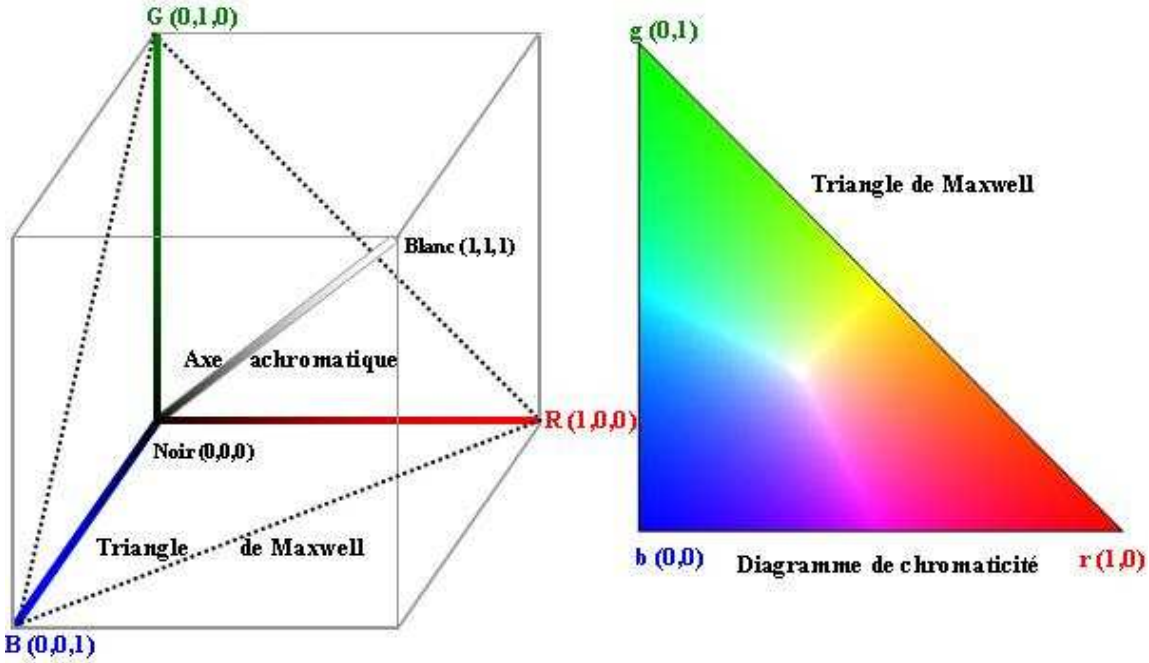


Figure 4.7: Cube des couleurs RGB et diagramme de chromaticité (r, g).

L’espace RGB délimité par les trois primaires R, G et B englobe l’ensemble des couleurs reproduites par synthèse additive et présente la forme d’un cube, d’ailleurs connu sous le nom de cube des couleurs (Figure 4.7). L’origine (0, 0, 0) symbolise la couleur noire, tandis que la couleur blanche correspond au triplet (1, 1, 1). Les couleurs rouge, vert et bleu, donc les primaires, ont respectivement comme coordonnées : (1, 0, 0), (0, 1, 0) et (0, 0, 1). La diagonale qui relie la couleur noire et la couleur blanche est appelée axe achromatique ou axe des gris ou encore axe de luminance. Cet axe présente toutes les nuances des niveaux de gris, et il peut être formalisé par l’équation : R = G = B.

Les coordonnées d’une couleur donnée au niveau de l’espace RGB dépendent à la fois de sa chrominance et sa luminance. Deux couleurs ayant une même chrominance peuvent avoir des coordonnées trichromatiques différentes si elles possèdent des niveaux de luminance différents. Dans le but de spécifier des composantes qui ne tiennent compte que de la chrominance, il convient d’utiliser les coordonnées normalisées :

$$r = \frac{R}{R+G+B} \qquad g = \frac{G}{R+G+B}\ b = \frac{B}{R+G+B}$$

Puisque r+g+b=1, il suffit alors de deux variables pour décrire une couleur, car la troisième variable peut se déduire des deux autres. On spécifie le plus souvent la chromaticité dans le plan (r, g), b peut se calculer par l’équation : b=1-r-g.

Maxwell a formalisé à partir du cube de couleurs RGB un plan représentant les différentes couleurs. Ce plan est défini par l’équation R+G+B=1 dans l’espace RGB. Son intersection avec le cube de couleurs forme un triangle appelé Triangle de Maxwell. Ce triangle est perpendiculaire à l’axe achromatique, ce qui signifie que toutes les couleurs contenues dans ce triangle, possèdent la même luminance (Figure 4.7). La position d’une couleur donnée ne dépend donc que de sa chrominance. L’intérêt du triangle de Maxwell est qu’il réduit l’espace couleur tridimensionnel à seulement un plan à deux dimensions. On dissocie ainsi la luminosité de la sensation couleur (Figure 4.7).

Wright a proposé un plan chromatique en projetant le triangle de Maxwell sur le plan (OR, OG) de l’espace RGB. Ce plan correspond au plan (r, g) défini à partir des coordonnées normalisées r et g. La Figure 4.8 montre l’étendue du triangle de Maxwell dans ce plan de chromaticité (r, g) appelée également le diagramme de Wright. La Figure 4.8 montre toute l’étendue du diagramme de Wright incluant même les valeurs négatives.

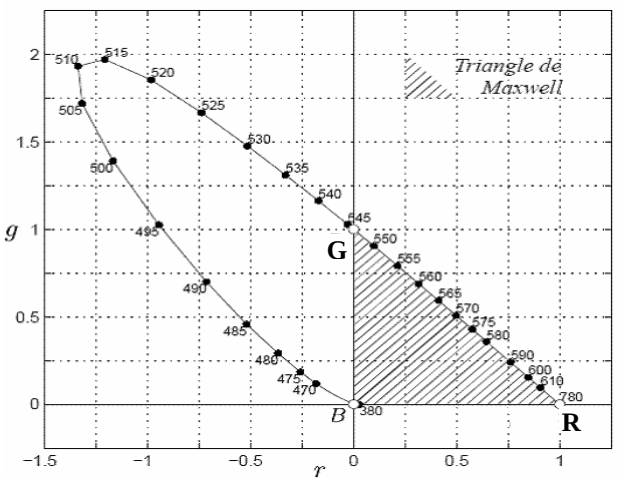


Figure 4.8: Diagramme de Wright (r, g).

Les couleurs monochromatiques du spectre visible (380nm-780nm) dessinent une courbe sous la forme d'un fer à cheval appelé le spectrum locus ou lieu du spectre, la ligne reliant les couleurs pures bleue et rouge est appelée la droite des pourpres. Le spectrum locus et la droite des pourpres, délimitent toutes les couleurs naturelles. On remarque très vite que la synthèse additive ne permet de reconstituer qu'une petite partie seulement des couleurs naturelles représentées par le triangle de Maxwell (partie hachurée). Les couleurs, en dehors de ce triangle, nécessitent l'introduction d'une composante négative, ce qui na pas de réalité physique. Et qui signifie que les technologies liées au système RGB dont nous disposons actuellement, telles que les moniteurs des ordinateurs, ne peuvent reproduire que les couleurs contenues dans le triangle de Maxwell.

4.5.1.3 Les fonctions colorimétriques x(λ), y(λ) et z(λ)

Les fonctions colorimétriques x(λ), y(λ) et z(λ) ont été établies par la CIE suite aux travaux de Judd [Judd 1930]. Elles ont été introduites afin de pallier aux inconvénients du système RGB puisque pour ce dernier :

- les coordonnées et les composantes trichromatiques peuvent prendre des valeurs négatives, ce qui constitue une exception au principe de la théorie trichromatique basée sur la synthèse additive,

- les composantes trichromatiques RGB sont liées à la luminance qui est une combinaison linéaire des composantes trichromatiques et non une composante en elle-même.

En observant la Figure 4.9 on remarque que l'aire couverte par le triangle de maxwell n'occupe même pas la moitié de la surface couverte par les couleurs naturelles. Il apparaît clair que la reproduction d'une grande gamme de couleurs n'est pas réalisable avec les primaires RGB, puisque on ne peut générer des lumières négatives.

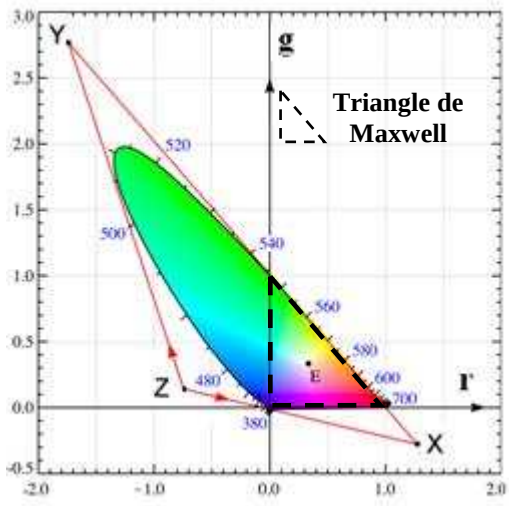


Figure 4.9: Diagramme de Wright (r, g) et le triangle XYZ.

En partant de ce constat, Judd a entrepris toute une série de transformations géométriques de l'espace RGB, dans le but de définir un nouveau système couleurs ne prenant en compte que des composantes positives pour la restitution de l'ensemble des couleurs naturelles. Il a commencé par remplacer le triangle de maxwell au niveau du plan de chromaticité (r, g) par un autre triangle plus grand qui englobe entièrement la totalité des couleurs naturelles (Figure 4.9). De ce fait, il a tracé trois droites qui sont les tangentes des deux parties latérales du spectrum locus. Par cette démarche il est parvenu à dessiner le plus petit triangle, qui puisse contenir l'ensemble des couleurs naturelles. Les sommets de ce triangle sont désignés par les lettres X, Y et Z et représentent les trois primaires du nouveau système couleur. Ces primaires sont virtuelles, elles n'existent pas physiquement et ne constituent pas des rayonnements réels, puisqu'elles se situent en dehors de la courbe des couleurs naturelles (spectrum locus) (Figure 4.9). Ces primaires ont pour coordonnées :

$$X = \begin{cases} r = +1.27 \\ g = -0.28 \end{cases} \qquad Y = \begin{cases} r = -1.74 \\ g = +2.77 \end{cases} \qquad Z = \begin{cases} r = -0.74 \\ g = +0.14 \end{cases}$$

A partir des coordonnées chromatiques r et g, sur le diagramme de Wright (r, g) (Figure 4.9), on peut facilement retrouver les composantes RGB des ces nouvelles primaires ainsi on a :

$$X = \begin{cases} R = 2.77 \\ G = 1.75 \\ B = 1.13 \end{cases} \qquad Y = \begin{cases} R = 1.00 \\ G = 4.59 \\ B = 0.06 \end{cases} \qquad Z = \begin{cases} R = 0.00 \\ G = 0.06 \\ B = 5.59 \end{cases}$$

La connaissance des coordonnées RGB des primaires XYZ permet la transformation de toutes les couleurs du système RGB vers le nouveau système trichromatique XYZ via une simple matrice de passage :

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{matrix} 2.77 & 1.75 & 1.33 \\ 1.00 & 4.59 & 0.06 \\ 0.00 & 0.06 & 5.59 \end{matrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Ce type de transformations est linéaire, ce qui signifie que la transformation est réversible. On peut donc revenir au système initial RGB par le biais de l'inverse de la matrice ci-dessus. De ce fait il y a conversion simple des lois du mélange des couleurs entre les deux systèmes.

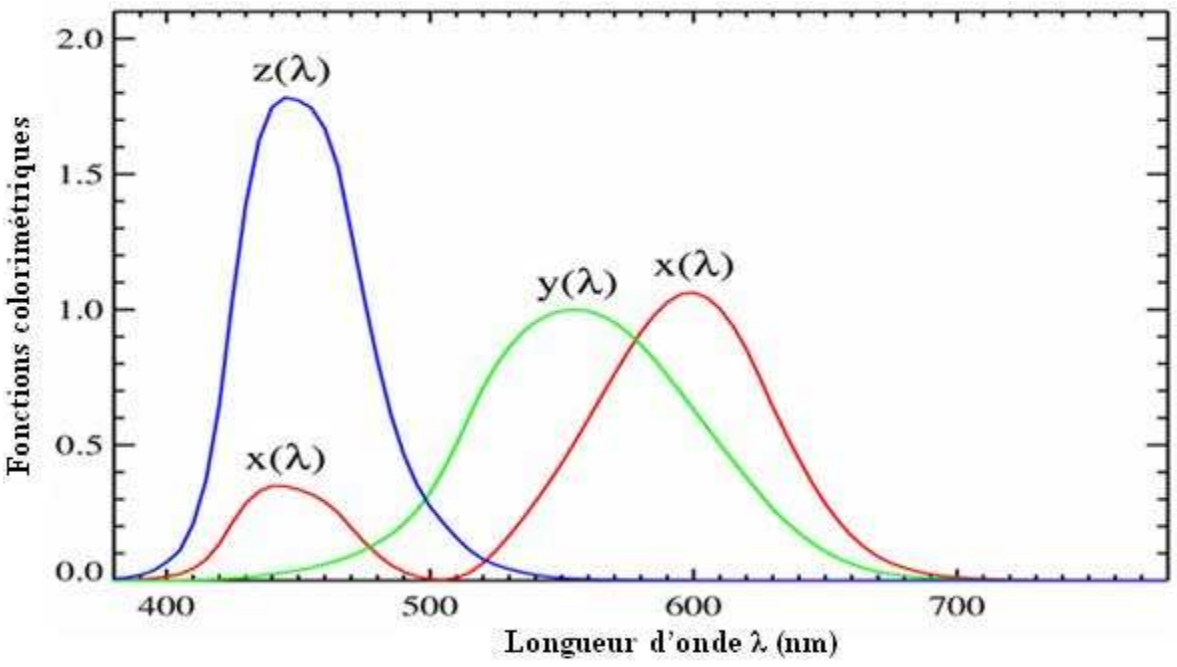


Figure 4.10: Fonctions colorimétriques x(λ), y(λ) et z(λ) de la CIE 1931.

On peut appliquer toutes les lois de Grassman au nouveau système XYZ. En 1931 la CIE a calculé les composantes XYZ des valeurs RGB de toutes les couleurs monochromatiques (spectrum locus), de ce fait la CIE a défini les courbes x(λ), y(λ) et z(λ) qui décrivent la variation des valeurs X, Y et Z en fonction de la longueur d'onde λ. Ces courbes sont connues sous le nom de fonctions colorimétriques du système de primaires XYZ CIE 1931 (Figure 4.10). Ces fonctions colorimétriques x(λ), y(λ) et z(λ) ne sont pas le résultat d'expériences d'égalisation visuelle des couleurs comme c'était le cas des fonctions associées au système RGB, elles sont déduites uniquement par simple calcul à partir du système RGB. Elles ont la particularité d'être toujours positives quelque soit la longueur d'onde monochromatique à égaliser (Figure 4.10).

Le système XYZ CIE 1931 constitue la base de la colorimétrie, tous les autres systèmes de primaires sont reliés à ce système [Trémeau et al 2004]. A ce titre il est le système de référence pour les mesures colorimétriques. Toutes les lois de calculs, établies par Grassman et Abney, qui régissent les relations entre les couleurs s'appliquent à ce système [Vandenbroucke 2000].

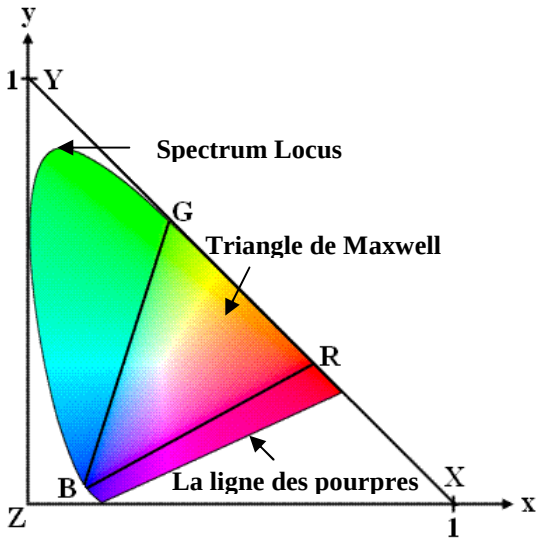


Figure 4.11: Diagramme de chromaticité xy CIE 1931. Le triangle de Maxwell est délimité par les couleurs R, G et B. Le spectrum locus est la courbes qui défini les couleurs pures. La ligne des pourpres relie les nuances des couleurs rouges et bleues.

De la même façon que le système RGB, un plan chromatique peut être obtenu à partir des primaires XYZ grâce aux primaires normalisées :

$$x = \frac{X}{X+Y+Z}, \qquad y = \frac{Y}{X+Y+Z} \text{ et} \qquad z = \frac{Z}{X+Y+Z}$$

Puisque x+y+z=1, deux composantes suffisent pour caractériser les couleurs. Le plan de chromaticité (x, y) est défini en analogie au plan chromatique (r, g) déduit de l'espace RGB. Le plan (x, y) appelé communément le diagramme de chromaticité xy est la représentation 2D la plus utilisée en colorimétrie (Figure 4.11).

Le diagramme de chromaticité xy présente l'avantage de respecter la synthèse additive en comparaison avec le plan chromatique rg (diagramme de Wright (r, g)). La Figure 4.11 montre clairement que toutes les couleurs naturelles peuvent être reconstruites avec des composantes positives, le triangle XYZ couvre une surface bien plus importante que le triangle RGB (triangle de Maxwell). En plus des couleurs naturelles, le triangle XYZ englobe d'autres couleurs imaginaires qui n'ont pas de réalité physique, cela correspond aux plages des aires blanches dans ce triangle. Les couleurs naturelles sont délimitées par la courbe des couleurs pures (spectrum locus) et la ligne des pourpres. La couleur blanche occupe le centre et a les coordonnées chromatiques (x, y) : (0.33, 0.33). Les couleurs les plus pures, autrement dit les plus saturées, se situent à la périphérie de la courbe des couleurs. Les couleurs deviennent de moins en moins saturées en se déplaçant vers la partie centrale où se trouve la couleur blanche. On peut dire que les couleurs se délavent en se rapprochant du blanc. Une caractéristique très importante du diagramme xy est de permettre l'application des lois qui décrivent le mélange des couleurs. En effet, toute couleur peut être considérée comme un mélange d'une ou plusieurs couleurs pures et la couleur blanche. Le mélange des couleurs est possible grâce aux jeux de constructions géométriques. Il suffit juste de tracer une ligne entre deux couleurs pour modéliser toute la gamme des couleurs qui résultent du mélange additif de ces deux couleurs. Le diagramme xy permet également de mesurer la distance entre les couleurs, et la comparaison des couleurs entre elles. On peut juger de ce fait si une « couleur a » est plus proche d'une « couleur b » par rapport à une « couleur c ». Ces caractéristiques seront exploitées afin de réaliser la classification des cellules tumorales au cours de ce travail.

Le diagramme xy procure une description chromatique cohérente des couleurs, à ce titre il est considéré comme la représentation 2D des couleurs la plus utilisée pour effectuer les différentes mesures colorimétriques [Trémeau et al 2004]. Cependant il manque d'uniformité perceptuelle. En effet une distance constante entre différents points colorimétriques ne se traduit pas toujours par le même écart perceptuel des couleurs. Les couleurs ne sont pas

réparties selon une logique perceptuelle. La sensibilité de l'œil à différencier les couleurs évolue suivant leurs emplacements au niveau du diagramme chromatique xy [Trouvé 1989] [Lezoray 2000]. Pour résoudre ce problème, différents travaux ont été menés en premier lieu pour mieux caractériser cette non-uniformité, puis pour établir une nouvelle représentation couleur perpétuellement uniforme pour laquelle tout écart couleur serait ramené à une distance métrique. Dans ce cadre on peut citer les travaux de Wright [Wright 1941] et MacAdam [Brown et MacAdam 1949] qui ont mené des études parallèles sur la non uniformité du diagramme xy. Malgré leurs approches différentes, tous les deux ont abouti à des résultats très similaires.

Nous allons nous contenter de présenter seulement les travaux de MacAdam [MacAdam 1943]. En effet, ce dernier a étudié 24 positions spécifiques réparties sur la surface du diagramme xy. Le principe était de se déplacer selon différentes directions à partir de ces 24 positions tant que l'observateur ne perçoit pas de différences couleurs. Le résultat est illustré dans la Figure 4.12 :

- les couleurs jugées non différenciables sont organisées selon des régions sous forme d'ellipses appelées les ellipses de MacAdam,
- les ellipses ont des tailles différentes et le rapport entre le grand axe et le petit axe, l'excentricité, vaut en moyenne 3.12 [Lozano 1997].

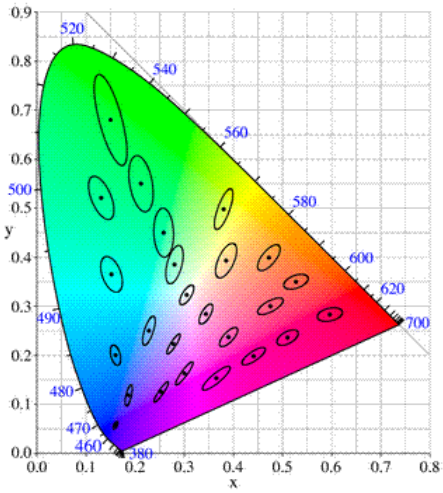


Figure 4.12: Ellipses de Macadam. Les ellipses sont agrandies 10 fois sur le diagramme de chromaticité xy de la CIE 1931.

Ces ellipses sont très petites dans le bleu, petites dans le rouge et grandes dans le vert [Trémeau et al 2004], ce qui signifie que l'œil est plus sensible aux écarts couleurs aux extrémités (bleu et rouge) qu'au centre du spectre (vert).

4.5.1.4 Vers une représentation des couleurs perpétuellement uniforme

Dans le but d'améliorer l'uniformité perceptuelle du diagramme de chromaticité xy, diverses transformations géométriques ont été proposées. Ces travaux ont donné lieu en finalité, à deux nouveaux espaces couleur, l'espace *Luv* et l'espace *Lab* normalisés par la suite par la CIE [Kowaliski 1990] qui les nomme respectivement l'espace *CIELUV* 1976 et l'espace *CIELAB* 1976.

Bien que les deux espaces *Luv* et *Lab* découlent d'études parallèles, ils présentent néanmoins beaucoup de similitudes, notamment :

- *L* représente l'axe de la luminance,
- le plan de chrominance est défini par les composantes *uv* et *ab*,
- les deux systèmes sont construits sur le principe d'opposition des couleurs d'Herring. L'opposition rouge-vert est représentée par les axes *a* et *u*, l'opposition bleu-jaune est reportée sur les axes *b* et *v*,
- les deux espaces offrent une représentation des couleurs selon un illuminant de référence. Ce qui signifie que les coordonnées d'une couleur donnée peuvent varier au niveau de ces deux espaces selon l'illuminant utilisé pour générer la couleur blanche,
- ces deux espaces ne sont qu'approximativement uniformes, bien que le but de leur développement ait été de proposer une représentation uniforme de la perception des écarts couleur. Ce point sera abordé plus en détail au cours de cette section.

Cependant ces deux espaces présentent quelques différences. Suivant l'application considérée ou la gamme de couleurs étudiées tel espace est privilégié plutôt que tel autre [Pointer 1981] [Robertson 1990] [Fairchild 1998] [Plataniotis et Venetsanoupoulos 2000].

L'espace *Luv* découle des travaux de Judd et MacAdam [MacAdam 1937] [MacAdam 1942], en effet ce dernier a développé un diagramme de chromaticité uv étendu par la suite à un espace 3D par Wyszecki noté l'espace (u,v,w) [Trouvé 1989]. Cet espace intermédiaire a été transformé et amélioré par Eastwood [Eastwood 1973] pour donner lieu à l'espace *Luv* ou CIELUV 1976. La transformation permettant le passage du système XYZ vers le système *Luv* est donné par les équations [CIE 1986] :

$$L = 116\left(\frac{Y}{Yo}\right)^{\frac{1}{3}} - 16 \qquad \text{pour} \quad \frac{Y}{Yo} > 0{,}008856$$

$$L = 903{,}3\left(\frac{Y}{Yo}\right) \qquad \text{pour} \quad \frac{Y}{Yo} \leq 0{,}008856$$

$$u = 13L\,[U(X, Y, Z) - U(Xo, Yo, Zo)]$$

$$v = 13L\,[V(X, Y, Z) - V(Xo, Yo, Zo)]$$

Avec :

$$U(X, Y, Z) = \frac{4X}{X + 15Y - 3Z} \qquad \text{et} \qquad V(X, Y, Z) = \frac{9X}{X + 15Y - 3Z}$$

Les équations sont définies en fonction des coordonnées (*Xo, Yo, Zo*), du blanc de référence, qui peuvent varier suivant l'illuminant de référence utilisé pour générer la couleur blanche. La Figure 4.13 montre bien l'opposition des couleurs sur les deux axes chromatiques *u* (rouge-vert) et *v* (bleu-jaune).

Si nous observons la disposition des ellipses de Macadam sur le diagramme *uv* pour juger de l'uniformité de cet espace (Figure 4.14), nous constatons que l'excentricité moyenne de ces ellipses vaut 2.48 [Pointer 1981]. Rappelons que le but de la création de l'espace *Luv* était de proposer un espace couleur, où ces ellipses auraient des formes circulaires avec une excentricité qui tend vers 1, traduisant ainsi l'uniformité des écarts perceptuels de la couleur autour des centres des ellipses de MacAdam. En effet, l'excentricité de ces ellipses est améliorée par rapport au diagramme xy, puisque celle-ci est ramenée de 3.12 à 2.48. Mais elle reste cependant loin de la valeur espérée de 1. La Figure 4.14 montre bien que les ellipses de MacAdam sur le diagramme uv restent de taille et d'excentricité variables suivant leurs emplacements. D'ailleurs, leurs formes et dispositions ressemblent à celles du diagramme xy.

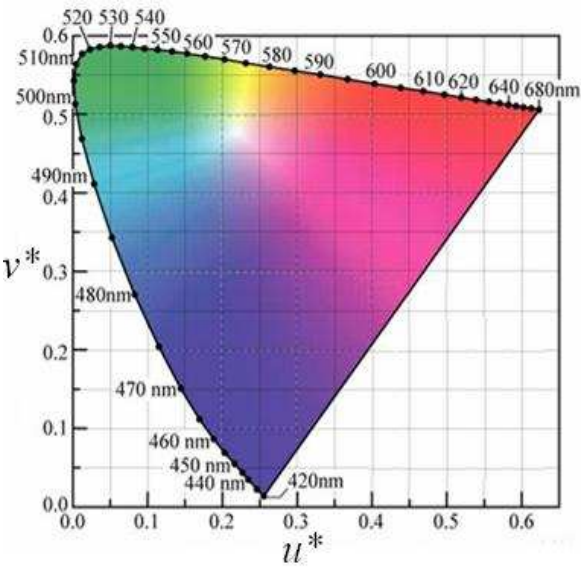


Figure 4.13: Le diagramme uv CIE 1976.

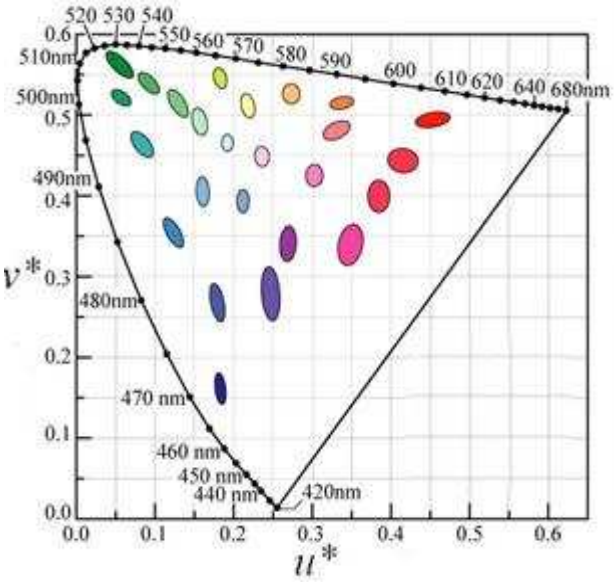


Figure 4.14: Les ellipses de Macadam uv.

Le système *Lab* a été adopté par la CIE en 1976. C'est une approximation d'un espace développé auparavant par Adams et Nickerson [Vandenbroucke 2000]. Il est obtenu à partir de l'espace XYZ selon les équations suivantes [CIE 1986] :

$$L = 116\left(\frac{Y}{Yo}\right)^{\frac{1}{3}} - 16 \qquad \text{pour} \quad \frac{Y}{Yo} > 0{,}008856$$

$$L = 903{,}3\left(\frac{Y}{Yo}\right) \qquad \text{pour} \quad \frac{Y}{Yo} \leq 0{,}008856$$

$$a = 500\left[f\left(\frac{X}{Xo}\right) - f\left(\frac{Y}{Yo}\right)\right]$$

$$v = 200\left[f\left(\frac{Y}{Yo}\right) - f\left(\frac{Z}{Zo}\right)\right]$$

Avec :

$$f(X) = X^{\frac{1}{3}} \qquad \text{pour } X > 0{,}008856$$

$$f(X) = 7{,}87\,X + \frac{16}{116} \qquad \text{pour } X \leq 0{,}008856$$

De même que l'espace *Luv* la transformation est définie selon les coordonnées trichromatiques (*Xo, Yo, Zo*) du blanc de référence, signalant que toute coordonnée (*X, Y, Z*)

est normalisée par les coordonnées de cette couleur blanche. Le système *Lab* offre aussi une représentation antagoniste des couleurs : l'opposition vert-rouge étant reportée sur l'axe a, le deuxième axe (b) défini l'opposition bleu-jaune (Figure 4.15).

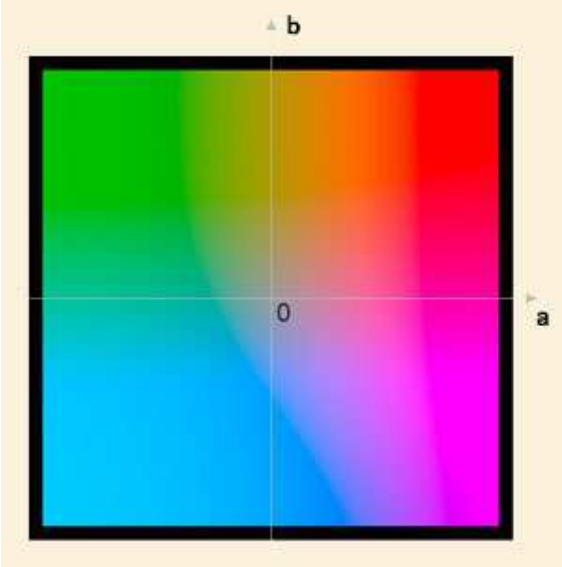


Figure 4.15: Diagramme de chromaticité *ab*.

Si on analyse l'uniformité perceptuelle de cet espace avec les ellipses de Macadam, on constate que l'espace *Lab* n'est pas proprement dit uniforme puisque la circularité absolue (excentricité = 1) des ellipses de Macadam n'est pas atteinte. L'excentricité moyenne des ellipses est de 3.50 [Pointer 1981]. De même que le diagramme xy les écarts couleur sont mieux détectés dans le bleu et le rouge par rapport aux couleurs proches du vert.

Le but initial de la proposition des deux systèmes *Lab* et *Luv* était de définir une représentation uniforme selon la perception humaine des différentes couleurs. Or on remarque que selon la forme et la disposition des ellipses de Macadam, ces deux espaces couleur offrent bel et bien une amélioration de l'uniformité par rapport au diagramme de chromaticité xy. En revanche l'uniformité absolue n'est pas atteinte. Ces deux systèmes ne décrivent plus une couleur de façon absolue. Ils la définissent désormais en fonction d'un illuminant de référence. Cet aspect est très important, on ne peut plus par exemple changer l'espace couleur d'une image numérique vers les espaces *Lab* et *Luv* sans connaître à *priori* l'illuminant qui a servi à éclairer l'image numérique pendant son acquisition. De plus, ces deux espaces

nécessitent toute une chaîne de transformations non linéaires à partir de l'espace XYZ pour décrire une couleur donnée. Ces transformations sont très coûteuses en temps de calcul, pour cette raison ces deux espaces couleur sont peu utilisés en traitement d'images couleur [Vandenbroucke 2000]. Cependant, les espaces *Lab* et *Luv* constituent les deux systèmes de référence pour effectuer les mesures colorimétriques puisque ce sont les deux systèmes qui se rapprochent actuellement le plus d'une représentation uniforme des couleurs. Selon l'application considérée, l'un ou l'autre des espaces *Lab* et *Luv* est préféré. À titre d'exemples, le système *Luv* est privilégié pour le calibrage des moniteurs, l'industrie des pigments utilise quand à elle le système *Lab*.

Jusqu'à présent nous nous sommes concentrés sur les principaux travaux menés dans le domaine de la colorimétrie et les systèmes couleur dits de référence recommandés par la CIE. En réalité, il existe un éventail très large de systèmes couleur. Ces systèmes sont développés par différents organismes et selon telles ou telles contraintes technologiques. Dans le chapitre suivant, nous allons présenter de façon globale les systèmes couleur. Nous allons nous pencher en particulier sur les principales caractéristiques qui ont conduit à leur organisation suivant plusieurs classes de systèmes couleur.

Chapitre 5

Systèmes de représentation de la couleur

L'existence de plusieurs modèles, expliquant les mécanismes qui régissent la vision humaine des couleurs, conjuguée aux contraintes technologiques, liées à certaines industries, a conduit à la coexistence d'un grand nombre de systèmes couleur dans la littérature. D'une façon générale ces systèmes sont organisés en cinq familles :

- ***les systèmes des primaires*** *qui sont définis selon le principe de la synthèse additive des couleurs à partir de trois couleurs de base appelées également primaires : le système le plus connu est le système RGB,*
- ***les systèmes luminance chrominance****, famille de systèmes couleur qui dissocient l'information chromatique des couleurs de leur intensité lumineuse,*
- ***les systèmes perceptuels*** *qui décrivent la couleur selon la perception humaine de la couleur : ils se basent sur des notions subjectives telles que la teinte, la luminance et la saturation,*
- ***les systèmes d'axes indépendants*** *qui utilisent des composantes couleurs décorrélées,*
- ***Les systèmes hybrides*** *qui consistent à retenir trois composantes, parmi les composantes couleurs de plusieurs systèmes, pour former un nouveau système couleur.*

Tous ces systèmes sont liés entre eux par divers types de transformations (linéaires ou non, réversibles ou non). Ce chapitre est dédié à la présentation de ces différents systèmes couleur.

Les différents travaux menés pour le choix d'un système couleur universel, qui soit le mieux approprié pour le traitement de toutes les images couleur, ne convergent pas vers un seul système couleur. Il est difficile à priori de déterminer un espace couleur qui soit le mieux adapté pour le traitement des images couleur en microscopie optique : qui est notre domaine d'application. Le choix d'un système couleur est, avant tout, fonction de l'application abordée et le type des images à traiter. Néanmoins, plusieurs études convergent vers le choix des systèmes couleur décorrélées, dissociant l'information chromatique de l'information achromatique pour le traitement automatique des images couleur.

5.1 Introduction

Le but de ce chapitre n'est pas de recenser tous les systèmes couleur existants. Ce travail dépasse le cadre de notre sujet. En effet, il existe une gamme très large de systèmes couleur. Certains systèmes possèdent même plusieurs définitions. Nous allons uniquement présenter les systèmes couleur les plus répandus dans le domaine de l'imagerie.

D'une façon générale, les systèmes de représentation de la couleur sont organisés en cinq familles selon leurs caractéristiques :

- les systèmes des primaires,
- les systèmes luminance chrominance,
- les systèmes perceptuels,
- les systèmes d'axes indépendants,
- les systèmes hybrides.

Notons que certains systèmes couleur peuvent appartenir à plusieurs familles à la fois.

5.2 Les systèmes des primaires

Les systèmes dits des primaires sont définis selon le principe de la trivariance visuelle qui stipule que chaque couleur peut être matérialisée par le mélange additif de trois couleurs de base appelées également primaires (§4.4.2.1).

5.2.1 Le système RGB

Parmi les différents systèmes de primaires RGB, seul le système RGB de la CIE doit être considéré comme le système de référence pour la colorimétrie [Trémeau et al 2004] (§4.3.2). Les autres systèmes RGB sont définis par différents organismes dans plusieurs domaines d'applications, tels que la technologie des téléviseurs couleur. C'est le système le plus commun pour l'acquisition, le stockage et les opérations d'échanges des images couleur numérisées.

Le système RGB se présente sous la forme d'un espace tridimensionnel porté par trois axes : rouge, vert et bleu. En informatique chaque axe est codé par un octet, lui donnant ainsi une dynamique de 0 à 256. Suivant ce codage, l'espace RGB peut définir plus de 16 millions de

couleurs. Remarquons que cet espace ne présente que les coordonnées positives, et que pour cette raison, une partie seulement des couleurs naturelles peut être définie, l'autre partie correspondant à des stimuli de couleurs non réalisables par synthèse additive.

Signalons enfin que d'autres espaces RGB sont définis par rapport à des primaires différentes de celles de la CIE. A titre d'exemple, les tubes cathodiques des téléviseurs et les moniteurs d'ordinateurs émettent des rayonnements (R, G, B) bien différents des primaires CIE. De même pour la technologie des téléviseurs, plusieurs normes de primaires coexistent. Ainsi, les téléviseurs européens utilisent la norme PAL (Phase Alternation by Line) et les téléviseurs américains fonctionnent selon la norme NTSC (National Television Standards Committee) [Vandenbroucke 2000]. Les moniteurs et les caméras utilisent aussi leurs propres primaires. D'ailleurs la CIE a défini un système de primaires spécifiques aux moniteurs couleurs [Lozano 1997].

5.2.2 Le système XYZ

Ce système a été établi pour pallier aux défauts du système RGB, notamment pour résoudre le problème des coordonnées négatives inhérentes au système RGB. Les primaires du système XYZ ont été crées de telle sorte que toutes les couleurs soient exprimées par des composantes trichromatiques positives, et de telle sorte que l'une de ces primaires représente l'information de luminance. Le système XYZ permet de représenter toutes les couleurs du visible reproduites par synthèse additive. La composante Y correspond à l'axe de luminance. Ce système a été amplement décrit dans la section 4.5.1.3. Signalons juste ici, que c'est le système de référence colorimétrique défini par la CIE [Trémeau et al 2004], et à ce titre, il est toujours utilisé dans le cadre de la colorimétrie.

5.2.3 Le système *LMS*

Bien que le système XYZ constitue le système de référence pour effectuer les mesures colorimétriques, le système *LMS* est la référence concernant la physiologie visuelle. Les primaires de ce dernier sont reliées directement avec les courbes de sensibilité des trois types de cônes de la rétine : *L, M* et *S*. C'est un système qui reflète le plus la perception visuelle humaine des couleurs. D'ailleurs, il est bien adapté pour analyser les petits écarts couleur. Les

primaires *LMS* se différencient des primaires RGB par la valeur de leurs longueurs d'ondes dominantes ainsi :

- 570 nm pour la primaire *L*,
- 540 nm pour la primaire *M*,
- 445 nm pour la primaire *S*.

Plusieurs fonctions colorimétriques dites fondamentales ($L(\lambda)$, $M(\lambda)$ et $S(\lambda)$) ont été proposées, notamment par Stiles et Burch [Trémeau et al 2004] ou par Stockman et Sharp [Stockman et Sharpe 2000]. Cependant, ce système ne doit pas être étendu aux applications de traitement d'images car il n'est pas encore normalisé. Sa normalisation est en cours d'étude par la CIE [Viénot 2001]. Notons enfin que les primaires *LMS* sont reliées aux primaires XYZ par la transformation suivante :

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{matrix} 0.15514 & 0.54312 & -0.03286 \\ -0.15514 & 0.45684 & 0.03286 \\ 0.00 & 0.00 & 0.00801 \end{matrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

5.3 Les systèmes luminance-chrominance

Cette famille de systèmes couleur dissocie l'information chromatique des couleurs de leur intensité lumineuse. En effet ce type de systèmes décrit un stimulus couleur selon une composante de luminance, et deux autres composantes allouées à la quantification de l'aspect chromatique du stimulus. Les composantes de ces systèmes sont calculées à partir des composantes trichromatiques RGB. On distingue différents types de systèmes luminance-chrominance :

5.3.1 Les systèmes Y Cb Cr

Les téléviseurs noir et blanc utilisaient seulement la luminosité. La diffusion des émissions en couleurs a poussé les chaînes de télévision à développer un système permettant de coder l'information couleur selon un système qui prend en charge les longueurs d'onde d'émission des tubes cathodiques. En effet, ces tubes possèdent des luminophores qui synthétisent la couleur selon des primaires couleurs différentes des primaires RGB. Ces primaires sont variées selon les normes de diffusion de chaque pays (standards télé : PAL, SECAM…..). Le

système devait assurer une compatibilité entre les téléviseurs couleur et les téléviseurs noir et blanc, puisque ces derniers devaient continuer à diffuser les émissions en noir et blanc même si le signal reçu était en couleur. Le passage entre les systèmes RGB et YC_bC_r nécessite une simple transformation linéaire. Notons qu'Y représente la luminance, et les deux composantes C_b et C_r représentent la chrominance. De façon générale Y représente la composante de luminosité du système XYZ, les composantes C_b et C_r se calculent selon la formule :

$$C_b = a_1 (R-Y) + b_1 (B-Y)$$

$$C_r = a_2 (R-Y) + b_2 (B-Y)$$

Les coefficients a_1 a_2 b_1 b_2 sont liés aux standards de télévision considérés.

Il existe plusieurs systèmes de type YC_bC_r :

- le système YIQ qui correspond à la norme NTSC,
- le système *YUV* lié à la norme PAL,
- le système YD_bD_r de la norme PAL.

Les téléviseurs émettent le signal reçu selon une loi exponentielle. Afin de restituer et d'afficher correctement ce signal il faut au préalable introduire une correction gamma. Par exemple pour le système *YIQ* de la norme NTSC on obtient :

$$\begin{bmatrix} R' \\ V' \\ B' \end{bmatrix} = \begin{bmatrix} R^{\frac{1}{2,2}} \\ V^{\frac{1}{2,2}} \\ B^{\frac{1}{2,2}} \end{bmatrix}$$

La transformation vers le système YIQ s'opère à partir de ces primaires corrigées R'V'B' [Trémeau et al 2004] :

$$Y = 0,299R' + 0,587V' + 0,114B'$$

$$I = 0,74(R'-Y) - 0,27(B'-Y)$$

$$Q = 0,48(R'-Y) + 0,41(B'-Y)$$

5.3.2 Les systèmes antagonistes

Ces systèmes couleur reposent sur la théorie des couleurs opposées d'Herring (§4.3.2). Ils divisent l'information couleur selon trois canaux : un achromatique pour l'opposition couleur

noir-blanc, et deux chromatiques pour les oppositions vert-rouge et bleu-jaune [Judd 1930] [Kowaliski 1990]. Plusieurs systèmes appartiennent à cette famille [Bedat 1998]. Le système AC_1C_2 introduit par Faugeras [Faugeras 1979] et développé après par Bumbaca et Smith est le système antagoniste le plus utilisé en traitement d'images couleur. Les composantes A, C_1 et C_2 sont évaluées à partir du système des primaires *L, M et S* qui traduisent les réponses spectrales des trois types de cônes de la rétine. La composante A englobe l'ensemble des trois primaires *LMS* :

- $A = a\,(\alpha \log(L) + \beta \log(M) + \gamma \log(S))$

Les deux composantes chromatiques C_1 et C_2 sont obtenues par le calcul de différences entre les primaires *L, M et S* [Faugeras 1979] [Bumbaca et Smith 1987]. Ainsi C_1 et C_2 sont données par les équations :

- $C_1 = \mu_1 (\log(L) - \log(M))$
- $C_2 = \mu_2 (\log(L) - \log(S))$

Les valeurs des coefficients a, α, β, γ, μ_1 et μ_2 varient selon l'application considérée, ainsi : Faugeras propose les valeurs : a = 22.6, α = 0.602, β = 0.369, γ = 0.019, $\mu_1 = 64$ et $\mu_2 = 10$ dans le domaine du traitement d'images couleur [Faugeras 1979].

Dans le domaine de la vision artificielle, d'autres systèmes antagonistes ne tenant pas compte de la vision humaine sont proposés. Ces systèmes présentent souvent une simplification du système AC_1C_2. De ce fait Garbay propose un système simplifié qui se calcule directement des primaires RGB [Garbay 1979] :

- $A = \frac{1}{3} \times (\log(R) + \log(G) + \log(B))$
- $C_1 = \frac{\sqrt{3}}{2} \times (\log(R) - \log(G))$
- $C_2 = \log(B) - \frac{\log(R) + \log(G)}{2}$

Signalons que les systèmes *Luv* et *Lab* uniformes peuvent aussi être considérés comme des systèmes antagonistes, puisqu'ils décrivent un axe achromatique et deux autres axes chromatiques basés sur les oppositions vert-rouge et bleu-jaune.

5.3.3 Autres systèmes luminance-chrominance

Cette catégorie de systèmes ne fait pas partie des deux types de systèmes cités plus haut. En effet il existe une multitude de systèmes couleur proposés dans le cadre de différentes études et appartenant à cette catégorie. Ainsi Carron propose un système noté YCh_1Ch_2 défini à partir du système RGB [Carron 1995] :

- A = $\frac{R+G+B}{3}$
- Ch_1 = R - $\frac{R+B}{2}$
- Ch_2= $\frac{\sqrt{3}}{2}$ x (B – G)

La CIE recommande l'utilisation de l'espace (Y, x, y), où Y est la composante de luminosité de l'espace XYZ, et les deux composantes x et y définissent le diagramme de chromaticité CIE. Littmann et Ritter proposent un système formé par le triangle de Maxwell pour représenter le plan chromatique, et l'axe achromatique représenté par l'axe de luminance qui relie les couleurs noir et blanc de l'espace RGB [Littmann et Ritter 1997].

5.4 Les systèmes perceptuels

Cette famille de systèmes couleur décrit la couleur selon la perception humaine de la couleur. Elle regroupe deux catégories de systèmes :

- les systèmes uniformes au sens de la perception visuelle,
- les systèmes de type Luminance, Teinte et Saturation.

5.4.1 Les systèmes uniformes

Ces systèmes peuvent décrire fidèlement, au sens de la perception visuelle, les écarts couleur proches. Ils possèdent en effet une métrique qui reflète au mieux les distances couleur selon la vision humaine des couleurs. Parmi ces systèmes, on peut citer :

5.4.1.1 Le système *Lab*

C'est le système de référence de la CIE pour calculer les écarts couleur. Ce système appartient aussi aux systèmes luminance-chrominance. Ce système a été amplement détaillé dans les sections précédentes, notamment les caractéristiques de son diagramme de chromaticité *(ab).* On peut juste rajouter que la représentation géographique de cet espace prend la forme d'une sphère (Figure 5.1).

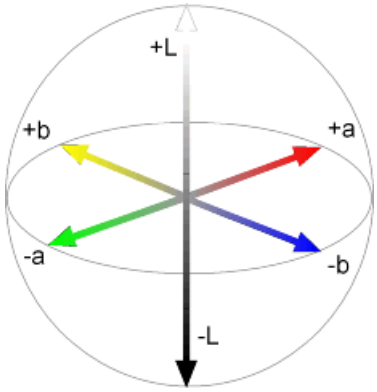


Figure 5.1: Représentation 3D de l'espace *Lab*.

Rappelons que le système *Lab* est obtenu par une transformation non linéaire à partir du système XYZ. En conséquence la dynamique varie selon les trois axes de cet espace. Ainsi la dynamique de la luminance *L* s'étend de 0 à100, tandis que la dynamique de la composante (*a)* s'étend de -130 à 200, et que la dynamique de la composante (*b)* s'étend de -160 à 160 [Raffy 1999].

Notons enfin que du fait de la dynamique variable selon l'axe considéré de cet espace, une transformation d'un espace RGB codé sur 8 bits par composante induit un espace *Lab* de très mauvaise sensibilité vis-à-vis des écarts couleur. Il faut *à priori* un espace RGB codé au moins sur 12 bits par composante pour obtenir un espace *Lab* avec un niveau acceptable de sensibilité. On peut conclure que l'utilisation d'un espace *Lab* n'est pas préconisée s'il dérive d'un système RGB en dessous de 12 bits par composante [Meyer et al 1986].

5.4.1.2 Le système *Luv* :

Ce système présente beaucoup de similitudes avec le système *Lab*, il est lui aussi recommandé par la CIE en colorimétrie. En effet, la plupart des caractéristiques du système *Lab* s'appliquent aussi au système *Luv*, notamment en ce qui concerne la configuration

géométrique des deux systèmes. Cependant, il existe quelques différences entre les deux systèmes selon certaines régions du spectre de la couleur [Tominaga 1996].

5.4.2 Les représentations de type Luminance-Teinte-Saturation :

Il paraît difficile selon la perception humaine des couleurs d'évaluer une couleur selon des coordonnées RGB abstraites. Il convient d'utiliser à la place des notions subjectives telles que la teinte, la luminance et la saturation. En effet, ces notions sont faciles à interpréter et sont liées à nos habitudes quotidiennes pour différencier les couleurs entre elles. Plusieurs systèmes quantifiant ces notions sont proposés dans la littérature. Ces systèmes fournissent une description des images numériques couleur qui est facile à assimiler. D'ailleurs les ordinateurs adoptent ce type de système pour communiquer avec l'homme lorsqu'il s'agit de décrire une couleur.

Les systèmes de type Luminance-Teinte-Saturation « LTS » sont basés sur un partitionnement uniforme de l'espace couleur selon un pas de discrétisation arbitraire, et qui n'est d'ailleurs pas relié à des grandeurs physiologiques de la vision humaine. Cependant, l'objectif de ces systèmes est de classer les couleurs selon un principe basé sur la perception psycho-visuelle. Ils sont surtout considérés comme des représentations descriptives de la couleur. La famille des systèmes LTS est mentionnée en littérature sous différentes nominations, citons à titre d'exemple :

- HSV pour Hue, Saturation and Value,
- ISH pour Intensity, Saturation and Hue,
- ITS pour Intensité, Teinte et Saturation,
- LCH pour Luminance, Chroma and Hue,
- HSL pour Hue, Saturation and Luminance,
- HSB pour Hue, Saturation and Brightness.

Tous ces systèmes se différencient par les formules qui permettent le calcul des grandeurs de leurs trois composantes [Rogers 1985] [Carron 1995]. Comme nous allons le constater dans ce qui suit, tous ces systèmes s'appuient la plus part du temps sur des transformations non-

inversibles. En effet, les opérations liées à ces transformations exigent des calculs de minima et de maxima.

Il existe différents modes de représentations géométriques liés à ce type de systèmes, de ce fait on peut les subdiviser en 4 familles :

- les systèmes cylindriques,
- les systèmes triangulaires,
- les systèmes en cône hexagonal et double-hexagonal,
- les systèmes d'atlas des couleurs.

5.4.2.1 Les systèmes en coordonnées polaires ou cylindriques

Comme leur nom l'indique, ces systèmes représentent l'ensemble des couleurs selon un cylindre centré sur l'axe de luminance (Figure 5.2). La saturation peut être formulée par la distance qui sépare le point couleur donné et l'axe de luminance. Les couleurs deviennent de plus en plus saturées en s'éloignant de l'axe de luminance. La teinte est formulée selon un angle qui varie entre 0 et 2π autour de l'axe de luminance.

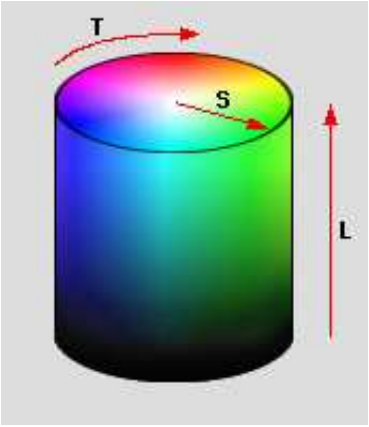


Figure 5.2: Représentation cylindrique de l'espace TLS.

Ces systèmes représentent en général les coordonnées cylindriques des espaces de type luminance-chrominance, comme le système *YIQ* lié à la norme de télévision NTSC. Sachant qu'*Y* correspond à la luminance, la teinte et la saturation sont données par les équations [Foo 1997] :

- $H = \arctan\left(\frac{\left(B^{\frac{1}{2,2}} - Y\right)}{\left(R^{\frac{1}{2,2}} - Y\right)}\right)$ et $C = ((R^{\frac{1}{2,2}} - Y) + (B^{\frac{1}{2,2}} - Y))^{\frac{1}{2}}$
- la saturation peut être obtenue par S = C/L.

Les systèmes *Lab* et *Luv* peuvent eux aussi être décrits selon des coordonnées polaires grâce au système cylindrique LCH (Figure 5.3). Puisque la composante *L* correspond à la luminance, les composantes H et C sont calculées par les formules [Vrhel et al 1994] :

- $C_{ab} = \left(a^2 + b^2\right)^{\frac{1}{2}}$ et $C_{ab} = \left(u^2 + v^2\right)^{\frac{1}{2}}$
- $H_{ab} = \arctan\left(\frac{b}{a}\right)$ et $H_{uv} = \arctan\left(\frac{v}{u}\right)$

Du fait de la fonction arctan, la teinte (H : Hue) est définie dans l'intervalle [-90°, +90°]. Pour résoudre ce problème et permettre ainsi à la teinte de s'étendre sur l'intervalle [0°, +360°] il faut passer par les transformations suivantes :

- H_{ab} si a > 0 et b > 0 et H_{uv} si u > 0 et v > 0
- $H_{ab} + 180°$ si a < 0 et b> 0 et $H_{uv} + 180°$ si u < 0 et v > 0
- $H_{ab} + 180°$ si a < 0 et b < 0 et $H_{uv} + 180°$ si u < 0 et v < 0
- $H_{ab} + 360°$ si a > 0 et b < 0 et $H_{uv} + 360°$ si u > 0 et v > 0

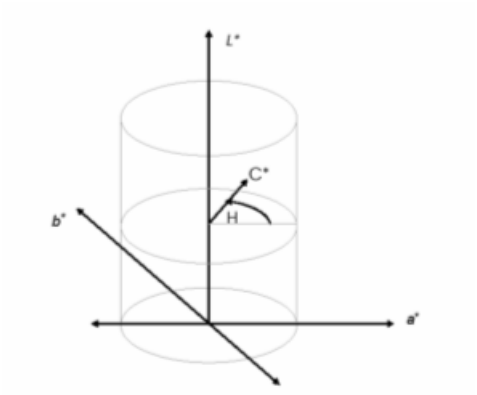


Figure 5.3: Représentation LCH du système *Lab.*

5.4.2.2 Les systèmes triangulaires

Ces systèmes notés ISH ou ITS sont calculés directement à partir du système RGB. IL suffit juste de modifier la disposition spatiale du cube de couleur RGB pour obtenir ce type de système. Si on pose le cube couleur sur le coin représentant la couleur noire, on peut définir ainsi un système ITS où l'intensité est représentée par l'axe achromatique de l'espace RGB, et les composantes S et T seront définies sur un plan perpendiculaire à l'axe d'intensité prenant la forme d'un triangle homothétique au triangle de Maxwell. Autrement dit, il faut juste rajouter au triangle de Maxwell une troisième dimension (Intensité : l'axe achromatique passant par le centre du triangle) afin de former cet espace. Ce système a été beaucoup utilisé en traitement d'images couleur [Carron 1995]. Différents modes de calcul existent pour exprimer les composantes IST à partir des primaires RGB :

- I = R + G + B [Ahmad et Reid 1996] [Kunt et al 1993] [Garbay 1979]

  ou I = $\frac{R+G+B}{3}$ [Lasserre 1996] [Adel 1994]

  ou encore I = $\frac{R+G+B}{\sqrt{3}}$ pour les formes pondérées.

- S = $1-\frac{3\times\min(R,G,B)}{R+G+B}$ [Claramont 1984].

  Certains auteurs remplacent cette formule par [Vandenbroucke 2000] :

  S = $1-3\times\min\left(\frac{R}{R+G+B},\frac{G}{R+G+B},\frac{B}{R+G+B}\right)$

  d'autres formules existent comme :

  S = $\sqrt{2}\times\frac{\sqrt{R^2+G^2+B^2-RG-GB-RB}}{R+G+B}$ ou S = $\frac{\sqrt{(R-G)^2+(G-B)^2+(B-R)^2}}{R+G+B}$

  Notons que la saturation n'est pas définie si R=G=B=0, pour cela on met S=0.

- T = arctan $\left(\frac{\sqrt{3}(G-B)}{2R-G-B}\right)$ [Ballard et Brown 1982]. Cette formule donne un angle compris entre 0° et 2π, l'angle 0° correspond à la couleur rouge. D'autres équations expriment la teinte selon le cosinus [Garbay 1979] :

$$T = \begin{cases} \arccos\left(\dfrac{0{,}5\times((R-G)+(R-B))}{\sqrt{(R-G)^2+(R-B)(G-B)}}\right) si B \le G \\ 2\Pi - \arccos\left(\dfrac{0{,}5\times((R-G)+(R-B))}{\sqrt{(R-G)^2+(R-B)(G-B)}}\right) si B > G \end{cases}$$

Une autre formule existe sans tenir compte des relations trigonométriques [Liang 1987] :

$$T = \begin{cases} \dfrac{B-R}{3(G+B-2R)} + \dfrac{1}{3} si \min(R,G,B) = R \\ \dfrac{R-G}{3(G+B-2R)} + \dfrac{2}{3} si \min(R,G,B) = G \\ \dfrac{G-B}{3(G+B-2R)} si \min(R,G,B) = B \end{cases}$$

5.4.2.3 Les systèmes en cône hexagonal

Cette famille de systèmes est connue sous le nom de systèmes HSV. Cette représentation a été définie par Travis. Il existe deux modèles de représentation : hexagonal (Figure 5.4) et double-hexagonal (Figure 5.5) :

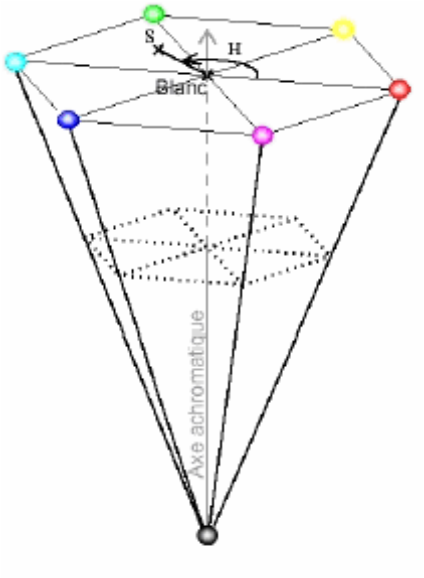


Figure 5.4: Système HSV en cône hexagonal.

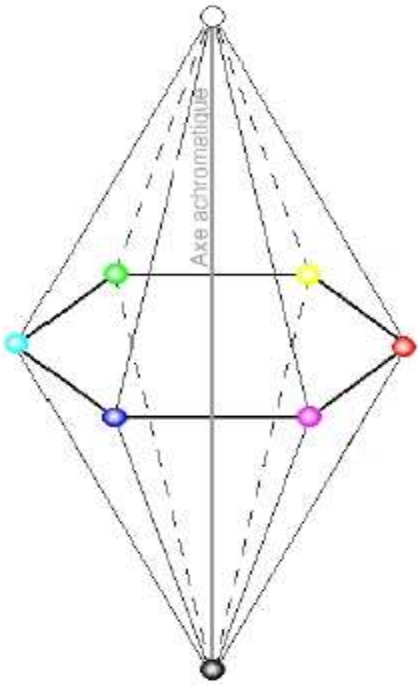


Figure 5.5: Système HSV en cône double-hexagonal.

❖ Le modèle hexagonal est obtenu en projetant le cube des couleurs RGB selon l'axe achromatique sur un plan perpendiculaire à cet axe et, contenant le coin du cube couleur représenté par la couleur blanche. On obtient ainsi une surface hexagonale dont les sommets sont occupés par les primaires RGB et leurs couleurs complémentaires et le centre est occupé par le blanc. La projection cumulée des couleurs sur tous les plans perpendiculaires à l'axe achromatique en partant du blanc vers le noir forme un cône dont le sommet est représenté par la couleur noire (figure 5.4). Dans ce modèle les composantes HSV sont données par les équations suivantes [Pujas 1996] :

- $V = \max(R, G, B)$
- $S = \dfrac{V - \min(R, G, B)}{V}$
- $$H = \begin{cases} \dfrac{G - B}{V - \min(R, G, B)} \, si V = R \\ 2 + \dfrac{B - R}{V - \min(R, G, B)} \, si V = G \\ 4 + \dfrac{R - G}{V - \min(R, G, B)} \, si V = B \end{cases}$$

On remarque que si V = R et min(R, G, B), H devient négative. La solution consiste à décaler les valeurs de la teinte pour qu'elle soit toujours positive, les équations suivantes matérialisent cette solution :

- $$H = \begin{cases} \dfrac{1}{6} \times \left[1 - \dfrac{R - G}{R - B} \right] si R \geq G \geq B \\ \dfrac{1}{6} \times \left[1 + \dfrac{G - R}{G - B} \right] si G \geq R \geq B \\ \dfrac{1}{6} \times \left[3 - \dfrac{G - B}{G - R} \right] si G \geq B \geq R \\ \dfrac{1}{6} \times \left[3 + \dfrac{B - G}{B - R} \right] si B \geq G \geq R \\ \dfrac{1}{6} \times \left[5 - \dfrac{B - R}{B - G} \right] si B \geq R \geq G \\ \dfrac{1}{6} \times \left[5 + \dfrac{R - B}{R - G} \right] si R \geq B \geq G \end{cases}$$

❖ Le modèle de cône en double-hexagonal (figure 5.5) est obtenu de la même manière que le modèle hexagonal. Mais au lieu de commencer à projeter les couleurs du cube RGB sur le plan contenant la couleur blanche, on commence par la projection sur le plan perpendiculaire à l'axe achromatique en passant par le milieu de cet axe. Puis on continue les projections vers les extrémités de cet axe. Les composantes HSV sont calculées de la façon suivante [Foley et al 1990] [Guzman de Leon 1997] :

- $V = \frac{\max(R,G,B) + \min(R,G,B)}{2}$

- $S = \begin{cases} \frac{\max(R,G,B) - \min(R,G,B)}{2V} si V \le \frac{V\max}{2} \\ \frac{\max(R,G,B) - \min(R,G,B)}{2(V\max - V)} si V > \frac{V\max}{2} \end{cases}$

  Vmax est la valeur maximale de la luminance.

  On remarque que si V=0 alors S=0.

- $H = \begin{cases} \frac{G-B}{\max(R,G,B) - \min(R,G,B)} siR = \max(R,G,B) \\ 2 + \frac{B-R}{\max(R,G,B) - \min(R,G,B)} siG = \max(R,G,B) \\ 4 + \frac{R-G}{\max(R,G,B) - \min(R,G,B)} siB = \max(R,G,B) \end{cases}$

La teinte n'est pas définie si S=0.

5.4.2.4 Les systèmes d'atlas de couleurs

Différents atlas de couleurs ont été construits afin de répertorier les couleurs : notamment, les atlas de Chevreuil et de Munsell. Ces atlas ont été principalement utilisés dans certains domaines industriels, comme le textile ou l'imprimerie. Par exemple, l'atlas de Munsell est un système couleur qui a été défini par le peintre Albert Henry Munsell en1909 [Trémeau et al 2004] puis repris et développé par la société américaine d'optique « Optical Society of American» en 1943. Ce système partitionne l'espace couleur selon un pas de discrétisation qui s'apparente au seuil différentiel du système visuel humain [MacAdam 1943]. En effet les couleurs sont décrites selon des quadrants couleur codés sur trois composantes : Luminance

(Value), Teinte (Hue) et Saturation (Chroma) représentant une forme qui ressemble à un cylindre (Figure 5.6). Bien que l'atlas de Munsell soit qualifié comme uniforme vis-à-vis de la perception visuelle, son inconvénient principal est le faible nombre de couleurs décrites. Puisque la luminance et la saturation représentent seulement une échelle à 10 niveaux. Seule la teinte peut varier dans une gamme à 100 unités. De ce fait, ce système peut décrire au maximum 10000 couleurs. Ce type de système ne peut être utilisé pour définir une couleur donnée de façon précise. Il ne décrit en fait que de manière approximative les couleurs.

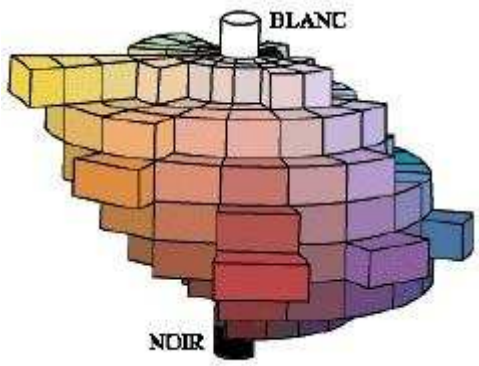


Figure 5.6: Système ou Atlas des couleurs de Munsell.

5.5 Les systèmes d'axes indépendants

Selon l'image considérée et le système couleur utilisé, les trois composantes liées à ce système peuvent présenter un degré de corrélation plus au mois important. A titre d'exemple, les composantes du système RGB sont fortement corrélées, du fait que la luminance est répartie sur chacune d'entre elles [Ohta et al 1980] [Lee et al 1994] [Kouassi et al 1997]. En effet, ces composantes sont corrélées car elles portent en elle une information commune. Dès lors, il devient difficile de traiter chaque composante indépendamment des deux autres. Plusieurs auteurs ont essayé de réaliser des systèmes de représentation de la couleur dont les composantes portent des informations différentes. Autrement dit, des systèmes couleur avec des composantes décorrélées. L'analyse en composantes principales peut apporter une bonne solution dans ce sens, notamment pour transformer le système couleur d'une image donnée vers un nouveau système décorrélé. Nous aborderons dans ce qui suit cette méthode, et le système qui en découle.

5.5.1 Le système en composantes principales

L'analyse en composantes principales (ACP) est une méthode d'analyse de données [Diday et al 1982] [Celeux et al 1989]. Elle consiste à analyser un ensemble de données représentées dans un espace multidimensionnel (ex. un espace couleur), pour obtenir un espace de représentation de ces données de dimension éventuellement réduite et dont les composantes sont décorrélées. Ces composantes sont appelées composantes principales. On peut ainsi transformer le système de représentation des couleurs d'une image donnée vers un système à composantes principales appelé encore système à axes indépendants. Tominaga applique l'analyse en composantes principales sur des images couleur codées par le système *Lab* afin de faciliter leur segmentation [Tominaga 1992]. Les différentes opérations nécessaires au calcul des nouveaux axes sont connues sous le nom de transformation de Karhunen-Loeve. Notons que l'information est reportée sur les 3 axes de façon déséquilibrée. Ainsi, le premier axe comporte le maximum d'information, puis suivi par le deuxième axe tandis que le troisième axe comporte peu d'information. De ce fait, la transformation de Karhunen-Loeve est utilisée parfois pour réduire l'espace couleur d'une image à traiter à un espace plus réduit ne comportant par exemple que deux axes [Savoji et Burge 1985].

L'inconvénient majeur de l'analyse en composantes principales est qu'elle dépend de la distribution des couleurs au niveau des images. En effet, les coefficients liés à la transformation de Karhunen-Loeve sont différents d'une image à l'autre car il faut *à priori* appliquer l'ACP à chaque image. De ce fait, le temps de calcul devient très important si pour une étude donnée, on doit appliquer la transformation de Karhunen-Loeve sur toutes les images à traiter. Certains auteurs ont tenté de contourner ce problème en appliquant une transformation proche de celle de Karhunen-Loeve et qui soit commune à l'ensemble des images à traiter [Guzman de Leon 1997] [Galichet 1989] [Orteu 1991]. De ce fait, et en fonction de l'application donnée, on obtient un seul système couleur en composantes principales pour toutes les images ; ce qui réduit de façon très significative le temps de traitement de ces images. Ainsi, Guzman de Leon propose un système d'axes indépendants spécifique à un jeu d'images coloscopiques [Guzman de Leon 1997]. À ce titre, on va aborder dans la section suivante le système proposé par Ohta et al [Ohta et al 1981] qui constitue le système d'axes indépendants le plus répandu.

5.5.2 Le système $I_1\ I_2\ I_3$

Ohta propose sur la base de l'analyse de plusieurs images couleur le système $I_1\ I_2\ I_3$ calculé par l'analyse en composantes principales. Il a pu démontrer que les équations générant ce système modélisent d'une manière efficace les trois premiers axes de l'ACP. Les composantes de ce système sont obtenues par les transformations suivantes :

- $I_1 = \frac{R+G+B}{3}$
- $I_2 = \frac{R-B}{2}$
- $I_3 = \frac{2G-R-B}{4}$

Ce système peut être classé avec la famille des systèmes de Luminance-Chrominance. Car si on inspecte de près l'équation générant la première composante, on peut constater que cette composante n'est autre que la luminance, tandis que les deux autres composantes représentent la chrominance. Ce système peut être également considéré comme un système de couleurs opposées, puisque les deux composantes, décrivant l'information chromatique, représentent un axe d'opposition des couleurs bleu-rouge et un autre axe pour les oppositions des couleurs magenta-vert. Signalons que la première composante, c'est-à-dire la luminance est la composante qui détient le plus d'information, suivie respectivement de la deuxième et la troisième composante. En conséquence, il convient de prendre quelques précautions pour la discrétisation de ce type de système. En effet, il faut *à priori* appliquer une dynamique qui serait en fonction du poids de l'information portée par chaque composante.

Pour résoudre le problème de la répartition non équilibrée de l'information sur les composantes du système, de même que pour les systèmes en composantes principales, un autre type de transformation modélisant l'ACP est proposé. Le but est justement une répartition plus équilibrée de l'information couleur sur l'ensemble des composantes du système couleur ainsi proposé. Nous aborderons dans la section suivante ce type de système.

5.5.3 Le système en composantes indépendantes

Au lieu de procéder à la transformation des systèmes couleur par l'ACP, une autre approche se base sur l'analyse en composantes indépendantes (ACI) [Liu 1991]. Le principal avantage de cette approche est le fait qu'elle permet la génération d'un système couleur, dont les composantes sont toujours indépendantes, mais sur lesquelles l'information est répartie de façon plus équilibrée.

5.6 Les systèmes hybrides

Enfin, au lieu d'utiliser un des espaces couleur précédemment présentés, certains auteurs retiennent trois composantes couleur parmi les composantes couleur de plusieurs systèmes (R, V, B, X, Y, Z, T, L, S …..). Ces trois composantes forment un système de couleur hybride [Vandenbroucke et al 1998] [Vandenbroucke 2000]. En effet, comme on a pu le constater tout au long de ce chapitre, tous les systèmes couleur sont reliés entre eux par divers types d'équations et tous ces systèmes sont définis dans un espace à trois dimensions. Dès lors, le choix d'un nouveau système couleur hybride paraît très intuitif. Néanmoins, la seule critique liée à ce type de systèmes est le fait d'être choisi de manière empirique, dans le cadre de certaines applications [Vandenbroucke 2000] [Lezoray 2000]. On ne peut les étendre à d'autres domaines du traitement d'images, car la sélection des composantes repose sur des critères propres au cadre de chaque étude.

5.7 Choix du système couleur pour le traitement des images couleur

Nous avons constaté tout au long de ce chapitre qu'il existe un grand éventail de systèmes couleur. D'ailleurs, on s'est limité à présenter les systèmes les plus communs. On a mis également en relief les points forts et les inconvénients de chacun de ces systèmes. En réalité, il n'existe pas un consensus autour d'un système de couleur qui soit recommandé pour effectuer le traitement des images couleur. De ce fait, le choix d'un système couleur pour traiter les images couleur liées à notre sujet devient une tâche non triviale. De nombreuses études ont été menées dans ce sens pour désigner le système de représentation des couleurs qui soit le plus satisfaisant et performant dans le traitement et la segmentation des images

couleur. Ces études touchaient des domaines d'applications très diversifiés. Leurs résultats divergeaient souvent sur le système couleur à recommander. En effet, ces travaux ont tenté de faire la comparaison entre les différents systèmes couleur afin de dégager le système le mieux adapté pour le traitement des images couleur.

Le système RGB a été recommandé par Lezoray [Lezoray 2000], Daily [Daily 1989] et Rodrigues [Rodrigues 1985]. Daily cherchait le système couleur qui évalue le mieux les différences de couleur d'images de scènes naturelles. Il conclue que le système RGB est celui qui donne le résultat le plus satisfaisant. Rodrigues a appliqué l'algorithme des nuées dynamiques pour la segmentation d'images couleur. Il démontre que les systèmes RGB et *YIQ* sont ceux qui permettent la meilleure segmentation. Lezoray a évalué la pertinence d'une segmentation manuelle en fonction du système utilisé pour représenter les images couleur. Il a mené son étude sur deux applications différentes. Il a trouvé que le système RGB est le mieux adapté pour extraire les différents compartiments cellulaires (noyau, cytoplasme) d'images cytologiques. En revanche, le système LTS est celui qui donnait le meilleur résultat pour extraire les nœuds au niveau d'images de planches à bois. On peut noter que si un système couleur est recommandé pour une application donnée, il peut s'avérer bien au contraire inadapté pour un autre type d'applications.

Le système Luminance-Teinte-Saturation (LTS) est considéré comme le plus performant d'après Pujas [Pujas 1996] et Ahmad et Reid [Ahmad et Reid 1996]. Il en est de même pour Lezoray [Lezoray 2000], mais dans le cadre de certaines applications seulement comme présenté dans le paragraphe précédent. En étudiant les fonctions de distribution des différents systèmes perceptuels, Pujas souligne que le système LTS est celui qui possède les meilleures propriétés statistiques. Ahmad et Reid constatent eux aussi que le système LTS est le plus efficace pour détecter les variations des couleurs des plantes de maïs lors de leur croissance. Néanmoins, il faut prendre quelques précautions en manipulant le système LTS, notamment pour le comportement instable de la teinte vis-à-vis des faibles niveaux de saturation. D'ailleurs, Pujas nous met en garde contre l'utilisation de ce type de système du fait de la périodicité de la teinte et son instabilité dans les bas niveaux de luminosité et de saturation.

Un autre groupe d'auteurs désignent les systèmes uniformes (*Lab*, *Luv*) comme les systèmes les mieux adaptés pour accomplir les tâches de segmentation d'images couleur [Rakotomalala 1999] [Takahashi et al 1995] [Ohta et al 1980] [Guzman de Leon 1997]. En effet, pour la détection des bordures des vaisseaux rétiniens, Rakotamalala note que, parmi plusieurs systèmes couleur, les systèmes *Lab* et *Luv* sont ceux qui offrent la meilleure détection [Rakotomalala 1999]. Le même résultat a été signalé par Takahashi et al pour la désignation automatique du nombre de classes d'objets au niveau des images couleur [Takahashi et al 1995]. Pour Ohta et al [Ohta et al 1980], le système *Lab* se dégage des autres systèmes couleur pour assurer une meilleure segmentation des images couleur. Guzman de Leon [Guzman de Leon 1997] rapporte aussi que le système *Lab* est le mieux adapté pour détecter les lésions au niveau d'un set expérimental d'images couleur coloscopiques. Cependant, ces deux derniers auteurs ont signalé que la transformation du système RGB vers le système *Lab* est très coûteuse en temps. Tous les deux préconisent le recours à d'autres systèmes couleur si la réduction du temps de traitement est nécessaire pour une application donnée. Dans ce cas de figure, Guzman de Leon préfère le système des primaires normalisées (r, g, b). En revanche Ohta recommande le système I_1 I_2 I_3. Ce système a été préconisé aussi par Lee et al [Lee et al 1994] à la place d'un système en composantes principales afin d'écourter les temps de traitement. En plus de ces contraintes liées au temps, le choix du système *Lab* est contrebalancé par les observations de Trémeau qui souligne que bien que ce système soit uniforme pour la perception visuelle, il n'apporte pas forcément les résultats les plus performants en traitement d'images couleur [Trémeau et al 2004]. En effet, il note que les transformations non linéaires qui permettent le passage du système RGB vers le système *Lab* engendrent un système couleur (*Lab*) avec des pas de discrétisation non uniformes. Si on dispose d'un système *Lab* calculé à partir d'un système RGB codé sur 8 bits, certaines couleurs se trouvent alors non représentées. Trémeau précise qu'il faut *à priori* disposer d'un système RGB codé sur 12 bits pour obtenir un espace *Lab* couvrant toute la gamme des couleurs perçues par l'œil humain [Trémeau et al 2004]. Il souligne à la fin que ce système n'est pas toujours approprié pour les traitements d'images couleur, mais de préférence qu'il est à recommander pour effectuer les mesures colorimétriques.

Notons enfin que certains auteurs ont entrepris une démarche marginale consistant à traiter les images couleur non plus selon un système couleur donné, mais selon la prise en compte

individuelle de chaque composante de plusieurs espaces couleur (RGB, LTS, *Lab*, *YIQ*.. ..). Avec cette démarche Ohlander a noté que la composante teinte présente souvent l'histogramme possédant le pic le plus significatif pour segmenter une image couleur [Ohlander 1975] [Ohlander et al 1978]. Dans un autre cadre d'étude, Robinson a constaté que la luminance est la composante la mieux adaptée pour détecter les contours au niveau des images couleur [Robinson 1977]. Ce résultat est conforté par le fait que selon les analyses en composantes principales, la première composante qui contient le maximum d'information se superpose le plus souvent avec l'axe de luminance. Enfin, il faut noter que la démarche marginale ne prend pas en considération l'aspect multidimensionnel des images couleur.

5.8 Conclusion

Nous avons vu dans le chapitre consacré à la couleur (chapitre 4), qu'il existe plusieurs modèles pour expliquer les mécanismes qui régissent la vision humaine des couleurs. Ceci conjugué aux contraintes technologiques de certaines industries (ex. Télévision) a conduit à la coexistence d'un grand nombre de systèmes couleur dans la littérature. Nous avons consacré ce chapitre pour présenter ces systèmes et les différentes classes auxquelles ils appartiennent. Nous avons constaté également qu'en finalité, tous ces systèmes sont liés entre eux par divers types de transformations (linéaires ou non, réversibles ou non).

Les différentes études abordées dans la section précédente (§5.7) ne convergeaient pas vers un seul système couleur universel approprié pour le traitement de toutes les images couleur. La principale conclusion que nous pouvons tirer de ces travaux, est le fait que le choix d'un système couleur est avant tout fonction de l'application abordée. Aussi, nous pouvons conclure que les critères retenus, pour juger si tel ou tel système couleur est le plus performant, influencent beaucoup ce choix.

Toutes ces études affirment qu'aucun espace couleur n'est le mieux placé qu'un autre pour une opération de traitement d'images données. Si un espace est retenu dans le cadre d'une étude donnée, cela ne veut pas dire pour autant que cet espace produira toujours les meilleurs résultats ; cela dépend de l'étude, du jeu d'images considérées et des critères retenus pour la comparaison des résultats.

Il est difficile *à priori* de déterminer un espace couleur qui soit le mieux adapté pour la segmentation des images couleur en microscopie optique, qui est le domaine d'application lié à notre sujet. Toutefois, plusieurs études convergent vers la conclusion suivante : l'axe de luminance est celui qui porte le maximum d'information, une partie du traitement d'image peut s'appuyer sur cet axe. Il reste à trouver le moyen de combiner les informations chromatiques de l'image couleur à l'information achromatique pour un traitement plus performant.

Chapitre 6

La segmentation des images couleur

La segmentation est l'étape clé de tout processus de traitement automatique d'images. La segmentation transforme l'image couleur en une image binaire objets versus non objets (fond de l'image). Les trois approches ***région, contour et pixellaire*** *constituent la pierre angulaire de toutes les procédures développées dans le domaine de la segmentation des images couleur. Les approches* ***région*** *et* ***contour*** *exploitent les relations de voisinage entre les pixels de l'image, c'est-à-dire qu'ils tiennent compte de la disposition spatiale des pixels. En revanche l'approche* ***pixellaire*** *s'appuie sur la disposition des pixels au niveau de l'espace couleur 3D. Ce chapitre donne un aperçu global de ces diverses méthodes de segmentation.*

A travers les travaux exposés au niveau de ce chapitre, nous pouvons dire qu'il n'existe pas de méthode de segmentation qui soit universelle et adaptée à tous les domaines de traitement d'images. Si un algorithme de segmentation peut donner des résultats satisfaisants pour une application donnée, il peut s'avérer défaillant et complètement inadapté pour un autre type d'application. Les procédures de segmentation les plus performantes sont celles conçues pour résoudre un problème de segmentation bien précis et pour un type d'images connu d'avance. En conséquence, il est primordial de procéder à une analyse minutieuse des caractéristiques des images à traiter avant toute conception de schémas de segmentation.

Dans le but de proposer une méthode de classification, qui soit adaptée à la quantification automatique de l'expression des marqueurs tumoraux, nous devons analyser au préalable les caractéristiques colorimétriques du marquage et de la contre-coloration. La prise en compte de ces caractéristiques est une condition indispensable pour concevoir un schéma de classification couleur performant pour l'analyse automatique des images histologiques.

6.1 Introduction

La segmentation est l'étape préliminaire de tout processus de traitement automatique d'images. Elle consiste à extraire des informations de bas niveau ou des primitives d'une image qui seront utilisées par la suite pour réaliser des schémas d'analyse d'images plus évolués, tels que la quantification du marquage immunohistologique ou la distribution tissulaire des cellules tumorales. C'est sur cette étape clé que repose la qualité des étapes ultérieures de la chaîne de traitement d'images. Par exemple, la finalité d'une analyse d'images histologiques d'un patient atteint d'un cancer donné, est d'établir un diagnostic concernant le grade de la tumeur en vue de mettre en place un protocole thérapeutique approprié. On comprend bien les conséquences néfastes d'une segmentation d'images inadéquate qui peut donner lieu à un diagnostic erroné. Un tel diagnostic peut conduire à une thérapie mal adaptée pouvant détériorer l'état du patient.

Pendant des décennies, les méthodes de segmentation traitaient exclusivement les images monochromatiques. La littérature foisonne de ce type de méthodes [Cheng et al 1999]. Cependant, la dernière décennie a connu l'émergence de la segmentation concernant les images couleur. De plus en plus d'articles relatifs à cette segmentation sont publiés chaque année. Cet essor peut s'expliquer par la baisse continuelle du coût des dispositifs d'acquisition des images couleur et la large diffusion des ordinateurs personnels équipés de puissantes capacités de calcul au niveau des laboratoires et centres de recherche. Ces machines sont nécessaires au développement des algorithmes de segmentation. De plus, les images couleur présentent des sources d'informations beaucoup plus importantes que les images en niveaux de gris.

Au début, les méthodes de segmentation couleur étaient de simples extensions des méthodes relatives à la segmentation en niveaux de gris, puisque l'image couleur RGB est composée de trois plans monochromes : R, G et B. Le principe consistait à segmenter chaque composante couleur à part, puis les résultats des trois segmentations dites marginales étaient fusionnés pour donner lieu à la segmentation couleur.

Une autre alternative consistait à réduire l'image couleur à segmenter en une image monochromatique. Dès lors, tout l'arsenal des méthodes de segmentation niveau de gris pouvait être appliqué sans adaptation préalable. Les transformations les plus mentionnées en littérature sont celles obtenues par l'analyse en composante principale et par la transformée de Karhunen-Loève [Trémeau et al 2004], ces dernières consistent à projeter les pixels couleur 3D sur l'axe portant le plus d'information par rapport à tous les axes de projection possibles. Ruifrok et Johnson [Ruifrok et Johnson 2001] ont opté pour une projection selon un axe représentant la couleur des structures à détecter en vue de segmenter les images de coupes histologiques. La réduction de l'image peut être obtenue également par une combinaison linéaire des trois composantes RGB de l'image couleur. Ainsi Bey et al [Brey et al 2003] utilisent une composante bleue normalisée (B normalisée= B/R+G+B) pour construire une image en niveau de gris. Montinomi et al [Montinomi et al 1996] préconisent l'utilisation de la composante verte normalisée par la composante bleue (G/B). Ruifrok [Ruifrok 1997] construit la composante Brown (Brown=B-0.3(R+G)) dans le but de quantifier le marquage immunohistochimique. L'approche la plus simple est sans doute l'utilisation directe d'une seule composante parmi les trois plans couleur. Cette méthode fait en effet abstraction des deux autres plans chromatiques mais en contraste, réduit le temps alloué au traitement des images. Dans cette catégorie, on peut citer le plan bleu utilisé par King et al [King et al 2002] ou le plan vert exploité par Smejkal et Shainoff [Smejkal et Shainoff 1997]. Ma et Loamoff ont porté leur choix sur la composante teinte de l'espace HSI pour réaliser la segmentation [Ma et Lozanoff 1997].

La décomposition de l'image couleur en trois plans séparés conduit à une perte d'information considérable. En effet, on perd toute relation entre les plans d'une même image et de ce fait, l'image couleur est réduite à une simple image multispectrale. Comparée aux niveaux de gris, la couleur est une source d'information beaucoup plus importante. En effet, le système visuel humain peut différencier plusieurs milliers de couleurs. En revanche, il ne peut guère distinguer plus de 30 niveaux de gris [Cheng et al 1999]. Actuellement, on trouve un grand nombre de méthodes dédiées à la segmentation des images couleur [Skarbek et Koschan 1994]. Ces méthodes exploitent justement le fort potentiel d'information inhérent à la couleur des objets à segmenter, prennent mieux en compte l'aspect tridimensionnel de l'information

couleur au niveau des images, et ne considèrent plus une image couleur comme un simple assemblage de trois images monochromes (R, G et B).

La segmentation transforme l'image couleur en une image binaire objets *versus* non objets (fond de l'image), ou vers une image de régions, où chaque région comporte des pixels ayant la même étiquette. Le processus de segmentation d'images consiste à partitionner l'image en régions homogènes et disjointes. Cette définition introduit en effet deux notions essentielles :

- homogénéité intra-région : qui signifie que les régions doivent respecter un prédicat d'uniformité couleur à l'intérieur de chaque **région**,
- hétérogénéité inter-région : deux régions adjacentes doivent être suffisamment différentes pour être séparées. Ce qui revient à établir un **contour** propre à chaque région.

En effet, c'est sur cette dualité région/contour que se base un grand nombre de méthodes de segmentation [Lucchese et Mitra 2001]. Les approches régions et contours exploitent les relations de voisinage entre les pixels de l'image, c'est-à-dire qu'ils tiennent compte de la disposition spatiale des pixels. En contraste à ces approches se basant sur l'espace 2D de l'image, une autre famille d'approches s'appuie sur la disposition des pixels au niveau de l'espace couleur 3D. Cette approche **pixellaire** considère chaque pixel comme un vecteur au niveau de l'espace couleur et fait abstraction de sa localisation au sein de l'image.

Les trois approches **région, contour et pixellaire** constituent la pierre angulaire de toutes les procédures développées dans le domaine de la segmentation des images couleur. Néanmoins, l'utilisation d'une seule approche conduit généralement à une segmentation peu satisfaisante. Les méthodes de segmentations les plus performantes combinent souvent les deux voire les trois approches.

Nous allons aborder tout au long de ce chapitre les approches région, contour et pixellaire, en citant pour chaque approche les procédures de segmentation les plus répandues dans la littérature. Ensuite, nous enchaînerons par la présentation des méthodes de segmentation mixtes combinant les trois approches. Le but de ce chapitre n'est pas une exposition

exhaustive de toutes les méthodes de segmentation couleur, mais de donner seulement un aperçu global sur les méthodes les plus connues. Les lecteurs les plus intéressés pour la segmentation des images couleur peuvent consulter les ouvrages de Trémeau et al [Trémeau et al 2004], Cheng et al [Cheng 1999] et de Skarbek et Koschan [Skarbek et Koschan 1994].

6.2 Approche région

Cette approche tente de partitionner l'image en zones homogènes. Elle a pour finalité de regrouper les pixels connexes ayant des caractéristiques couleurs proches en régions. En plus de cette cohérence colorimétrique et spatiale (connectivité des pixels), les régions doivent avoir un sens matériel dans l'image. Elles doivent correspondre à des objets réels, tels que des cellules ou des noyaux cellulaires au niveau d'une image de coupe histologique par exemple. L'approche région comprend essentiellement deux familles de méthodes :

6.2.1 Croissance de régions

C'est une approche ascendante qui consiste à agréger aux régions en formation, les pixels voisins qui respectent un prédicat de similarité couleur. La segmentation commence toujours par la pose des germes de régions au niveau de l'image. Ces germes peuvent être des points ou des groupes connexes de points. Ensuite, ces germes se propagent par des pas successifs de croissance dans l'espace 2D de l'image en agglomérant les pixels adjacents et proches au sens colorimétrique. La croissance s'arrête lorsque tous les pixels de l'image sont segmentés. Donc, la segmentation comprend 2 phases :

- phase d'initialisation pour la pose des germes,
- phase de la croissance des régions.

La phase de la sélection des germes est primordiale car elle conditionne la qualité de la segmentation qui en résulte. Les germes doivent être positionnés correctement sur les régions à extraire. Plusieurs stratégies existent pour générer les germes. Ils peuvent être définis par le moyen d'opérations morphologiques d'ouvertures et de fermetures suite à la binarisation de l'image à segmenter [Kanai 1998]. Gramariuc et al retiennent les minima locaux les plus significatifs de l'image gradient comme germes [Cramariuc et al 1997]. Cette approche

conduit parfois à définir plusieurs germes par région. En revanche, les petites régions ne sont représentées par aucun germe. Dans le but de segmenter des images de coupes histologiques par la ligne de partage des eaux (LPE), Kim et al [Kim et al 2006] s'appuient sur les germes correspondant aux maxima locaux de l'image des distances calculées après un seuillage de l'image originale. D'autres travaux adoptent une stratégie complètement différente. Certains auteurs distribuent les germes de façon régulière sur toute la surface de l'image [Amadasun et King 1988] [Ismaili et Gillier 1994]. Cette approche conduit souvent à une sur-segmentation et nécessite un post-traitement afin de fusionner les germes voisins et correspondant à une même région.

Une fois les germes définis au niveau de l'image couleur, la phase de croissance de régions proprement dite peut débuter. A ce stade, il subsiste également une multitude d'approches. Plusieurs stratégies de croissance de régions sont citées en littérature [Luccheese et Mitra 2001]. Hormis la proximité spatiale du pixel à agréger avec le germe en propagation, qui est une condition commune à toutes ces approches, la mesure de similarité couleur entre le pixel candidat et la région diffère d'une étude à l'autre. A titre d'exemple, Chassery et Garbay [Chassery et Garbay 1984] agrègent un pixel à une région sur la base d'une distance couleur calculée au niveau d'un système couleur de type LTS. Cette distance est formalisée par l'équation suivante :

$$Dist(P,R) = \sqrt{(L_P - L_R)^2 + (S_P - S_R)^2 + \left(\frac{S_P + S_R}{2}\right)^2 (T_P - T_R)^2}$$

R est la région en cours de construction et P représente le pixel candidat à l'agrégation. Cette équation calcule l'écart couleur entre le pixel et la région. On signale également que l'écart des teintes est pondéré par la saturation de P et R pour éviter les effets de fluctuation de la teinte pour les bas niveaux de saturation. Le pixel est ajouté à la région si cet écart est inférieur à un seuil fixé *ad hoc*. La couleur de la région est recalculée à chaque itération.

Trémeau et Borel [Trémeau et Borel 1997] utilisent la distance couleur euclidienne du système RGB comme critère de similarité colorimétrique. Cette distance couleur entre deux pixels P_1 et P_2 est formalisée par la relation :

$$Dist(\mathrm{P1},\mathrm{P2}) = \sqrt{(R_1 - R_2)^2 + (G_1 - G_2)^2 + (B_1 - B_2)^2}$$

Pour qu'un pixel soit agrégé à une région, il doit satisfaire trois conditions :

- similarité couleur : le pixel doit être proche au sens de la couleur de la région,
- homogénéité locale : le pixel doit aussi avoir une couleur proche de ces pixels voisins appartenant déjà à la région,
- densité locale : il faut qu'un certain nombre de voisins du pixel candidat appartienne déjà à la région pour que celui-ci soit rajouté à la région.

Ces trois conditions sont matérialisées par trois seuils. Les deux premiers définissent une distance couleur minimale tandis que le troisième définit le nombre minimal de pixels voisins. Tous ces seuils sont fixés par l'utilisateur, ce qui représente un facteur limitant pour une application automatisée de cette méthode.

Fesharaki et Hellestrand [Fesharaki et Hellestrand 1993] ont une approche de croissance de région totalement différente. Ils adoptent le test statistique de Kolmogorov et Smirnov. Chaque pixel à agréger est vu comme un individu d'un échantillon de la région. L'échantillon est formé par une fenêtre de 5x5 pixels autour du pixel considéré. Cette fenêtre de pixels est divisée en deux groupes selon différents schémas de partage. Puis, une hypothèse statistique est émise :

- H0 : Les deux sous-populations de pixels appartiennent à la même population de pixels. Ce qui signifie que cette fenêtre de pixels est homogène et fait partie de la même région,
- H1 : Les deux sous-populations sont hétérogènes et appartiennent à deux populations différentes de pixels. Dans ce cas, le pixel se trouve sur la limite de deux régions adjacentes, et servira à départager ces deux régions.

Au cours du test de Kolmogorov et Smirnov, les pixels sont représentés par leurs niveaux de gris sur les plans R, G et B. Cette méthode est marginale. La segmentation couleur est obtenue suite à la fusion des résultats de la segmentation réalisée sur chaque plan.

Bien que la ligne de partage des eaux (LPE) soit un outil de morphologie mathématique, elle s'apparente aussi à une méthode de croissance de régions, puisque la construction des régions

débute à partir des germes qui se propagent dans l'espace pour couvrir toute la surface des régions. La LPE a été largement utilisée pour segmenter les images en niveaux de gris [Serra 1982], puis elle a été étendue aux images couleur par Mayer [Mayer 1992]. Si on prend pour exemple une image en niveau de gris pour expliquer la LPE, l'image est vue comme une surface topographique avec les intensités des pixels comme altitudes. L'idée de la LPE se base essentiellement sur un processus d'immersion graduelle par l'eau de cette surface. Tout naturellement l'eau commence par occuper les parties les plus basses correspondant aux vallées. Puis, l'eau remplit progressivement ces vallées. Chaque vallée est considérée comme un bassin versant, le niveau d'eau augmentant continuellement au sein de chaque bassin avant de communiquer avec les eaux des autres bassins. Les points de communication de l'eau provenant de deux bassins voisins décrivent la ligne de partage des eaux : c'est la frontière entre deux vallées ou régions adjacentes. En vue d'appliquer l'algorithme de la LPE sur les images couleur, Saarinen [Saarinen 1995] adopte une approche marginale en fusionnant la LPE des 3 composantes R, G et B. Cependant, le mérite revient à Meyer [Meyer 1992] pour l'application de la LPE directement sur l'image couleur. Hormis les germes, il place tous les pixels de l'image couleur dans une file. Les pixels sont ordonnés dans la file selon la norme de leurs couleurs qui est formulée par l'équation :

$$Norme(\mathrm{P}) = \sqrt{(R)^2 + (G)^2 + (B)^2}$$

Chaque pixel de la file est attribué au germe ayant la différence de norme couleur la plus petite parmi les régions les plus proches. Belhomme et al [Belhomme et al 1997] remplacent la comparaison des normes couleur par la distance couleur euclidienne. Il revient en effet à un schéma de croissance de régions plus classique en gardant la structure de la file pour une construction plus rapide des régions. La LPE conduit souvent à une sur-segmentation conduisant à des objets morcelés en plusieurs régions. Saarinen [Saarinen 1995] utilise un graphe d'adjacence de régions pour fusionner les régions voisines et homogènes selon un critère de similarité couleur. Lezoray et al [Lezoray et al 1999] rajoutent, à la distance couleur classique (distance euclidienne), l'information gradient du pixel dans le but de construire un critère de similarité plus robuste. Ce critère est formalisé par l'équation suivante :

$$Dist(P,R) = (1-\alpha)\left\|\sqrt{(R_P - R_R)^2 + (G_P - G_R)^2 + (B_P - B_R)^2}\right\| + \alpha\left\|\nabla(R_P, G_P, B_P)\right\|$$

Le terme $\|\nabla(R_P, G_P, B_P)\|$ représente le gradient couleur au niveau du pixel P, et α est un coefficient de pondération. Cette distance prend en compte à la fois une information globale (la distance couleur entre le pixel candidat et sa région adjacente), et une information locale représentée par la variation de couleur dans l'entourage immédiat du pixel. Si ce pixel appartient au contour d'un objet, il va se situer dans une zone de variation de couleur abrupte. Son gradient couleur prendra alors une valeur très significative rendant ainsi la distance entre le pixel et la région très importante, ce qui empêche le rajout de ce pixel à la région en cours de construction. α permet de modifier et d'ajuster l'influence des informations globale et locale. S'il vaut 1, on retrouve un algorithme LPE classique opérant au niveau de l'image gradient couleur. En revanche, s'il prend la valeur 0, on retrouve une approche typique de croissance régions basée sur le critère de distance couleur.

L'approche de croissance de régions conduit en général à détecter les zones homogènes de l'image. Les régions texturées et les contours sont souvent mal définis. La construction précise des régions est fonction de plusieurs facteurs :

- la disposition initiale des germes,
- le choix de critère de similarité couleur,
- l'ordre de croissance des régions,
- les valeurs des seuils de similarité couleur pour agréger un pixel à une région.

Le réglage des différents seuils au cours de la croissance des régions est le principal facteur limitant pour une généralisation de cette approche dans le but de segmenter différents types d'images. Le réglage nécessite souvent des utilisateurs expérimentés. Des seuils inadaptés peuvent conduire à des résultats de segmentation erronés.

6.2.2 Division-fusion des régions

A l'inverse de la croissance des régions où l'on entame la construction d'une région à partir d'un jeu de pixels homogènes (germe), l'approche division-fusion débute la segmentation à partir de régions initialement hétérogènes. Au lieu de poser des germes sur toute la surface de l'image, on dispose d'un seul germe représentant l'image toute entière. La division-fusion

consiste dans un premier temps, à diviser le germe (région) s'il ne satisfait pas à un prédicat d'homogénéité colorimétrique. Le processus de division se poursuit au niveau des nouvelles régions générées, et s'achève une fois que toutes les régions sont homogènes. La deuxième phase de fusion vient affiner la segmentation en fusionnant les régions voisines selon un critère de similarité couleur.

L'approche division-fusion repose sur une structuration préalable de l'image couleur à segmenter. La structure la plus simple et la plus répandue en littérature est le tétra arbre [Jain et al 1995] [Horowitz et Pavlidis 1974]. C'est une structure arborescente avec une racine correspondant à l'image toute entière. Elle doit être de forme carrée avec des dimensions de type N^2xN^2 avec N un nombre entier positif. Chaque nœud correspond à un bloc carré de pixels, et possède 4 nœuds fils à l'exception des nœuds terminaux (régions homogènes). La phase de division débute au niveau du noeud racine (image entière), et par une opération de division récursive, chaque nœud non homogène de l'arborescence est scindé en 4 sous blocs carrés qui deviennent les 4 nœuds fils. La division se solde par un arbre ayant tous les nœuds terminaux appelés également les feuilles homogènes. Au terme de la phase de division, des blocs ou nœuds voisins ayant des couleurs proches peuvent exister. Le rôle de la phase de fusion est d'explorer tous les nœuds terminaux afin de fusionner ce type de blocs adjacents.

Les travaux optant pour la structure du tétra-arbre divergent sur le choix des procédés adoptés pour réaliser les opérations de division et de fusion. Ainsi, Coutance [Coutance 1991] se sert des variances R, G et B des pixels du nœud comme critère d'homogénéité. Le nœud est divisé si ces variances dépassent un seuil fixé auparavant. En revanche, il compare simplement les couleurs moyennes des régions pendant la phase de fusion.

Panjwani et healy [Panjwani et healy 1995] utilisent une modélisation en champs de Markov dans le but de segmenter des objets couleur texturés. Chaque bloc est divisé tant qu'il ne contient pas un modèle de texture prédéfini. La fusion est réalisée par une approche de relaxation considérant les blocs ou les régions comme des champs de Markov. Cette approche est très intéressante car elle ne se limite pas à prendre en compte seulement les valeurs individuelles des pixels, mais prend en considération aussi les relations d'entourage de chaque

pixel d'une région. Le critère de fusion repose sur la comparaison des vecteurs couleur, et des matrices de corrélations des deux régions à fusionner.

Round et al [Round et al 1997] adoptent une approche classique de division pour la segmentation des images de cancers de peau. En revanche, la phase de fusion repose sur l'organisation des régions selon un graphe d'adjacence de régions. Cette structuration accélère le traitement des images en donnant un accès plus rapide aux voisins de chaque région. Celenk [Celenk 1997] utilise la méthode de classification K-means pour réaliser les deux phases de segmentation division-fusion. Le critère se base sur le calcul de la variance des pixels pour les décisions de division et de fusion.

En dépit de la facilité de mise en œuvre des algorithmes se basant sur une structure en tétra-arbre, l'inconvénient majeur réside dans la rigidité d'une telle structure. Le découpage carré lié à cette approche déforme les objets présents dans l'image. En effet, la forme des régions générées tend à mimer la structure du tétra-arbre, et les contours des objets sont souvent mal définis. Dans le but de proposer des structures qui respectent au mieux la forme des objets, Gevers et Smeulders préfèrent une structuration de l'image selon une triangulation de Delaunay [Gevers et Smeulders 1997]. D'autres auteurs adoptent la partition de la surface de l'image selon le formalisme lié au diagramme de Voronoi [Schettini et Suardi 1994] [Itoh et Matsuda 1996]. Ce modèle offre une segmentation plus précise surtout dans les cas des images de coupes histologiques en délimitant bien les contours des structures cellulaires. Schettini emploie une procédure de division basée sur l'analyse d'histogrammes d'un certain nombre de caractéristiques chromatiques [Guo et al 1998]. L'algorithme repose sur la recherche récursive du mode ou pic le plus important au niveau de chaque histogramme. Les pixels de ce mode formant un ensemble de points connectés au niveau de l'image, sont retenus pour former les régions. L'histogramme contenant les pixels restant est analysé à nouveau et de façon récursive jusqu'au traitement complet de tous les pixels. La phase de fusion utilise la distance de Fisher comme critère de similarité couleur afin de regrouper les régions voisines. Les régions de petites tailles sont rattachées à la fin aux régions adjacentes ayant le vecteur couleur le plus proche.

6.2.3 Conclusion

Nous avons vu dans cette partie que l'approche de segmentation en régions tente de regrouper les pixels connexes ayant des couleurs similaires en régions homogènes. Elle se base essentiellement sur ce critère d'homogénéité couleur.

Les méthodes présentées requièrent le réglage de plusieurs seuils et paramètres régulant le critère de similarité couleur. Ce réglage n'est pas trivial et nécessite parfois l'intervention d'un utilisateur expérimenté. En effet, le réglage des seuils est délicat. Un mauvais ajustement de ces seuils conduit souvent à un mauvais résultat de segmentation [Di Zenzo 1986]. S'agissant de fusion ou de croissance de régions, l'ordre de construction des régions conditionne le résultat de la segmentation. Pour s'affranchir de cet inconvénient, il faut *à priori*, structurer les régions selon un graphe d'adjacence en cas de fusion, ou agréger les pixels selon un ordre de similarité décroissant dans le cas de croissance des régions.

Dans la plupart des cas, les régions obtenues ont des contours mal définis et ne délimitent pas la forme réelle des objets présents dans l'image. Cela est dû à la rigidité des structures utilisées pour accomplir la segmentation (tétra-arbre).

Les méthodes de croissance et de fusion des régions sont souvent utilisées en complément d'autres approches de segmentation, notamment des approches conduisant à une sur-segmentation. Elles interviennent comme méthode de post-traitement pour affiner le résultat de segmentation.

6.3 Approche contours

6.3.1 Introduction

L'approche contour est largement utilisée pour les images en niveau de gris [Fu et Mui 1981] [Rosenfeld et Kak 1982] [Pal et Pal 1993]. Elle repose essentiellement sur la détection des discontinuités des niveaux de gris à travers la surface de l'image. Elle met à nu les changements rapides des intensités qui caractérisent les bords ou limites des objets réels présents dans l'image. En effet, ce changement est la signature de déplacement d'un objet vers

un autre ou vers le fond de l'image. Les techniques de détections de contours suivent en général deux stratégies :

- **Approche globale :** cette catégorie fait partie des techniques de traitement de signal. Elle considère l'image comme un échantillonnage ponctuel et bidirectionnel d'un ensemble de signaux analogiques continus. Un signal peut être défini essentiellement par sa fréquence. Cette approche prend en compte tous les points de l'image simultanément, elle reçoit en entrée une matrice contenant tous les pixels de l'image, et délivre en sortie un signal de fréquences 2D. Le passage de l'image initiale vers le domaine fréquentiel est réalisé par la transformée de Fourier [Cheng et al 1999]. L'opération inverse est également possible grâce à la transformée de Fourier inverse. Au niveau fréquentiel, les régions homogènes de l'image sont converties en basses fréquences. En revanche, les hautes fréquences sont associées aux transitions rapides des niveaux de gris qui décrivent les bordures d'objets. Par conséquent, la détection des contours au niveau de l'image initiale se fait par le repérage des hautes fréquences au niveau du signal de fréquences 2D. Cette opération s'effectue simplement par l'utilisation d'un filtre passe-haut sur la transformée de Fourier de l'image initiale. De façon pratique la détection des contours se fait en deux étapes :
    - le filtrage de l'image des fréquences par un filtre passe haut,
    - l'application de la transformée de Fourier inverse au résultat du filtrage.

On obtient ainsi une image ne contenant que les pixels appartenant aux contours en sortie de cette chaîne de traitement. Les techniques liées à la transformée de Fourier trouvent leur principale application en imagerie médicale, notamment dans le domaine de la radiologie. L'inconvénient majeur de ce type de méthodes est qu'elles sont très restrictives. Elles ne détectent, en effet, que les changements abrupts des niveaux de gris. Elles sont incapables de déceler les contours d'objets matérialisés par un changement progressif d'intensité.

Comme alternative à la transformée de Fourier, on peut citer l'utilisation des champs de Markov comme moyen de modélisation de la matrice des pixels représentant l'image. Cependant, l'implémentation des algorithmes liés à cette approche est compliquée et nécessite

beaucoup d'étapes d'optimisation avant la détection effective des contours [Skarbek et Koschan 1994].

- **Approche locale :** ce type d'approche a besoin uniquement d'accéder à l'entourage immédiat d'un pixel pour affirmer si celui-ci fait partie ou non d'un contour. La détection des contours se fait par l'application directe d'un filtre 2D au niveau de l'image. Le filtre est représenté par une matrice caractérisée par sa largeur de fenêtre et le poids attribué à ses différents éléments. La convolution de l'image par le filtre se fait d'une manière séquentielle. Pour accomplir cela, on parcourt toute la surface de l'image. Au cours de chaque étape, le filtre est centré sur un nouveau pixel afin de calculer le rapport de celui-ci avec ses voisins. Il existe un grand nombre de filtres. A titre d'exemple, on peut citer les filtres de Sobel et Prewitt. Leur application sur une image revient à calculer l'opérateur gradient au niveau de chaque pixel. Le Laplacien correspond à un autre type de filtre, son principe repose sur la génération de la dérivée seconde de tous les pixels de l'image. Les contours coïncident avec les passages à zéro de cette dérivée seconde [Cheng et al 1999]. Ces filtres ne prennent en considération qu'un nombre restreint de voisins autour de chaque pixel. On peut élargir la fenêtre du filtre pour intégrer un entourage plus étendu. En procédant ainsi, on analyse le pixel dans un contexte régional, mais en revanche les temps de traitement se trouvent allongés. L'application d'un filtre 2D sur une image n'est pas un but en soi. Ce n'est qu'une étape dans un schéma de traitement global. Selon le filtre utilisé, on obtient comme résultat, soit une image de gradient soit une image de dérivée seconde. Il faut une étape supplémentaire exploitant ces données pour isoler les contours à terme. Les contours sont obtenus en deux étapes classiques de niveaux de gris [Canny 1986] :
    - extraction des maxima locaux de la norme du gradient,
    - seuillage par hystérésis afin de détecter les chaînes de pixels correspondant aux contours et élimination de celles faisant partie du bruit de l'image.

Au niveau des images en niveau de gris, les contours se manifestent par un changement rapide des niveaux de gris, ce qui traduit la différence d'intensité lumineuse entre deux régions

adjacentes. Cependant, en imagerie couleur l'information liée à la définition d'un contour est beaucoup plus riche. La signature d'un contour ne concerne plus un seul vecteur (niveau de gris) mais trois vecteurs à la fois. Par exemple, un contour séparant deux régions adjacentes ayant des luminosités proches et des teintes suffisamment distinctes est difficilement détectable sur une image en niveau de gris. En revanche, il est facilement décelé sur une image couleur [Macaire et al 1996]. Ainsi, un contour couleur peut être défini comme la discontinuité ou les changements abrupts de la couleur au niveau de l'image. La discontinuité n'est plus mono-composante (niveau de gris) mais multi-composantes. Elle doit être considérée selon les 3 axes de l'image couleur. En partant du principe qu'une image couleur est composée de trois images monochromatiques et, que les filtres ne peuvent être appliqués directement sur l'image couleur, l'approche marginale à partir des composantes monochromatiques semble la voie la plus adéquate pour générer les contours au niveau de l'image couleur. L'extraction des contours peut se faire par

- fusion des résultats des segmentations marginales.

En contraste à cette approche marginale, une autre stratégie consiste à construire une seule mesure traduisant les variations au niveau des trois composantes couleur à la fois [Cheng et al 1999] [Lucchese et Mitra 2001]. Ce type d'approche se base essentiellement sur les stratégies suivantes :

- fusion des vecteurs gradients marginaux pour calculer un vecteur gradient couleur,
- utilisation d'un vecteur gradient couleur basé essentiellement sur la teinte,
- les Méthodes basées sur le calcul des dérivées secondes,
- la méthode des contours actifs.

Dans ce qui suit, on détaillera toutes ces approches.

6.3.2 Fusion de résultats marginaux

Il existe en général deux schémas de fusion des résultats issus des trois composantes de l'image couleur :

- Fusion des trois images binaires de segmentation : Cette stratégie repose simplement sur des opérations binaires logiques, telles que l'union « or » ou l'intersection « and ». En effet, elle regroupe les résultats de segmentation des plans R, G et B sur une seule

image binaire. Cette technique est facile à mettre en œuvre et permet l'utilisation de n'importe quelle approche contour sur les composantes marginales de l'image couleur. Néanmoins, l'aspect tridimensionnel n'est pris en compte qu'au moment de la fusion.

- Fusion des images des normes des gradients marginaux : La norme du gradient traduit le degré de variabilité des niveaux de gris autour d'un pixel donné. Si la valeur de cette norme est importante, alors on est dans une zone de changement rapide d'intensité. Au contraire, les petites valeurs de norme définissent les zones homogènes de l'image. Un gradient est caractérisé par sa norme et par sa direction. La norme est donnée par la formule suivante :

$$N = \sqrt{(G_x)^2 + (G_y)^2} \qquad \text{et la direction par l'équation :} \qquad Direc = artg\left(\frac{G_x}{G_y}\right)$$

G_x et G_y sont respectivement les coordonnées verticale et horizontale du gradient marginal G. Souvent, la norme est normalisée entre 0 et 1.

La fusion des gradients peut s'effectuer en ne prenant en compte que les normes. Cette fusion peut se faire suivant les opérateurs min, max ou la moyenne. L'opérateur minimum est très restrictif et ne donne que peu de contours, mais il donne aussi peu de faux contours. Le maximum, en revanche, détecte beaucoup de pixels contour mais en même temps génère autant de faux contours introduits par le bruit de l'image. La moyenne peut être le moyen de compromis entre le min et le max. Il donne des résultats intermédiaires entre les deux derniers. Dans certains schémas de fusion, on introduit également la direction des gradients pour faire la différence entre les gradients ayant la même direction, des gradients orthogonaux ou avec des directions opposées.

6.3.3 Fusion des vecteurs gradients marginaux

La phase de fusion intervient très tôt dans le processus de segmentation. Un gradient vectoriel appelé également le gradient de Di Zenzo [Di Zenzo 1986] est construit à partir des gradients marginaux. Le gradient multi-composantes décrit plus fidèlement les transitions de couleurs au niveau d'une image multi-composantes. De ce fait, on prend mieux en compte l'aspect

tridimensionnel des images couleur. Cette approche a été développée par Di Zenzo [Di Zenzo 1986] et Lee et Cok [Lee et Cok 1991], elle est largement utilisée pour définir les contours au niveau des images couleur [Trémeau et al 2004]. En résumé, et sans entrer dans les détails des équations mathématiques, le gradient de Di Zenzo est généré à partir des trois gradients R, G et B en recherchant la direction des variations les plus prononcées. Ce gradient est applicable selon deux conditions [Lezoray 2001] :

- la couleur doit être répartie de façon corrélée sur les trois axes couleur de l'image. On peut appliquer ce gradient au niveau d'un espace de type RGB puisque les axes R, G et B sont, en général, fortement corrélés. Cependant, son application se trouve limitée au niveau des espaces de type HSI ou LTS, parce que cette famille d'espaces scinde la couleur en deux axes d'information complètement décorrélés : l'axe chromatique et l'axe achromatique,
- le niveau de bruit doit être distribué de manière équivalente sur les trois composantes de l'image couleur. Ce problème peut être surmonté par exemple, en lissant au préalable l'image couleur par un filtre Gaussien avant de procéder aux différents calculs, et pour obtenir des contours bien définis et moins sensibles aux bruits inhérents au processus de formation des images.

A titre d'exemple, on peut citer l'application de ce type de gradient (Di Zenzo) par Chapron [Chapron 1992] pour repérer les contours couleur au niveau des images couleur RGB.

6.3.4 Gradient couleur LTS

Le gradient couleur est calculé à partir des gradients luminance, teinte et saturation. Une attention particulière est donnée au gradient teinte, puisque la fiabilité des valeurs de teinte est fonction du niveau de saturation. Si une couleur est suffisamment saturée, la teinte est alors très pertinente. Au contraire, elle est très instable lorsque le niveau de saturation est faible, et devient très sensible au bruit. Carron et Lambert [Carron et Lambert 1994] proposent une formule pour calculer le gradient couleur en tenant compte des ces observations. Ils pondèrent le gradient teinte par un paramètre qui varie en fonction de la saturation. La formule est donnée par l'équation suivante :

$$\vec{G}_{LTS} = \vec{G}_L + \alpha\vec{G}_T + \vec{G}_S$$

$\vec{G}_L$, $\vec{G}_T$ et $\vec{G}_S$ représentent les gradients marginaux liés à la luminance, la teinte et la saturation. La valeur du paramètre α est proche de zéro pour les niveaux bas de saturation. Elle est ramenée à 1 quand la saturation est importante. On peut dire que le gradient couleur néglige le gradient teinte lorsque la saturation est faible. En dehors de cette situation, tous les gradients marginaux sont pris en compte.

Dans certains cas de figure, il est plus intéressant de ne travailler qu'avec la teinte pour définir les contours. Par exemple, les contours d'images couleur contenant des effets d'ombre sont mieux détectés avec le gradient teinte seul qu'avec l'association de celui-ci avec les gradients de luminance et de saturation. Ces deux derniers introduisent une variabilité liée aux ombres des objets et non aux contours réels de ces objets [Cheng et al 1999]. En accord avec ces considérations, Carron propose [Carron 1995] une deuxième formule avec un nouvel équilibre entre les gradients marginaux :

$$\vec{G}_{LTS} = (1-\alpha)\vec{G}_L + \alpha\vec{G}_T + (1-\alpha)\vec{G}_S$$

Dans cette deuxième équation, le gradient couleur se base essentiellement sur la teinte sous condition d'un bon niveau de saturation. Puisque le terme (1-α) tend vers zéro si la saturation est élevée, alors les gradients de saturation et de luminosité tendront à s'annuler, laissant le gradient couleur se construire à partir du gradient teinte. A l'inverse, sous un bas niveau de saturation, α prend la valeur de zéro. Le gradient de teinte est négligé au profit des deux autres gradients pour calculer le gradient couleur. Cette approche se montre très efficace, car en se basant sur la teinte, on privilégie la prise en compte de l'information chromatique qui, est stable et invariante à travers la surface de l'image, à l'opposé de la luminosité, qui peut varier si l'image a été acquise par exemple sous des conditions d'illumination non uniformes. Dans une image d'objets en relief, le gradient basé sur la teinte détecte correctement les contours. Par contre, un gradient basé sur la luminosité détecte en plus, des faux contours correspondant aux ombres, des objets. Pour cette raison, la deuxième version de Carron pour calculer le gradient couleur trouve tout son intérêt.

6.3.5 Méthodes basées sur les dérivées secondes

Au lieu de définir les contours par extraction des maxima locaux du gradient obtenu par la dérivée première, cette approche propose la délimitation des passages par zéro de la dérivée seconde directionnelle qui correspondent aux pixels contour. En effet, les passages à zéro de la dérivée seconde mettent en évidences les extrema de la dérivée première. De plus, cette approche présente l'avantage de générer des contours fermés et non pas des lignes ouvertes nécessitant un post-traitement de fermeture dans le cas des gradients couleur. La dérivée seconde peut être calculée de deux manières :

- calcul de la dérivée seconde directionnelle selon la direction d'un gradient de Di Zenzo, cette méthode a été adoptée par Cumani [Cumani 1991] et Alshati [Alshatti et Lambert 2000] pour la détection des contours,
- la dérivée seconde directionnelle est générée à partir des dérivées secondes marginales calculées par un filtre de Laplacien par exemple. Cette méthode est moins précise que la première, mais elle est très simple à mettre en œuvre et permet des temps de segmentation très rapides.

L'inconvénient majeur de la dérivée seconde est l'amplification du bruit de l'image lors des opérations de dérivation de premier et second ordre qui peut générer des faux contours. Si on admet que le bruit est un phénomène inhérent à tout processus de formation d'image, le moyen le plus simple pour éviter ce problème est l'application de filtres de lissage, pour atténuer les effets de ce bruit. Loukas et al [Loukas et al 2003] ont utilisé un filtre LOG (*Laplacian Of Gaussian*) dans le but de détecter les contours des noyaux cellulaires au niveau d'images histologiques. En effet, ce filtre (LOG) lisse l'image (filtre gaussien) avant la détection effective des contours par un filtre Laplacien. Byun et al [Byun et al 2006] adaptent les dimensions (largeur de fenêtre) du filtre LOG au diamètre moyen des objets à segmenter (noyaux cellulaires) afin de calculer la densité des cellules de la rétine au niveau des images microscopiques.

6.3.6 La méthode des contours actifs

Cette approche originale a été proposée par Kass et Witkin en 1988 [Kass et Witkin 1988]. Elle consiste à déformer un contour initial, jusqu'au moment où il coïncide parfaitement avec les bordures d'un objet réel. La déformation du contour est guidée par 3 forces :

- l'énergie propre du contour actif, c'est la force de tension et de torsion de la ligne formant ce contour,
- l'énergie potentielle de l'image qui vise à coller le contour actif sur les contours d'un objet,
- une énergie externe comportant les contraintes d'appariement entre le contour actif et le contour de l'objet. Les contraintes peuvent être spécifiques à la forme ou la taille du contour recherché par exemple.

Le contour actif évolue dans l'image dans le but de rechercher un compromis entre ces trois formes d'énergie. La détection d'un contour réel correspond à une énergie commune minimale combinant les trois énergies.

Sapiro a adapté l'approche des contours actifs aux équations dérivatives des images couleur [Sapiro 1996] [Sapiro 1997]. Il a introduit la métrique Riemannienne pour capturer l'information qui provient des trois composantes couleur et pour piloter la phase de déformation du contour actif au niveau d'une image couleur.

Malgré l'originalité de cette approche, elle reste néanmoins une méthode semi-automatique puisque les contours actifs sont posés sur l'image par un utilisateur. De plus, certaines contraintes sont rajoutées aux contours actifs qui sont fonction des objets recherchés. Dans le but d'améliorer cette approche, Yang et Jiang [Yang et Jiang 2001] utilisent l'information gradient pour détecter les contours des cellules biologiques en premier lieu. Ensuite, ces contours sont réajustés aux bordures réelles des cellules en introduisant un modèle ellipsoïdal qui décrit la forme des cellules.

6.3.7 Conclusion

Les résultats de l'approche contours sont généralement liés à la nature des objets présents dans l'image. Souvent, les contours sont bien délimités au niveau de certaines applications telles que les images de robotique ou d'environnement artificiel. Cependant, ces contours sont moins précis sur les images médicales et les images de scènes naturelles. Avec la présence d'images bruitées telles que les images d'ultrasons ou de radiologie, les contours s'avèrent complètement erronés. En général, les méthodes de détection de contours nécessitent deux types de traitement :

- un pré-traitement : on a vu que le bruit peut gêner une délimitation plus précise des contours. Pour certaines approches, ce bruit se trouve amplifié générant ainsi des faux contours. Il existe beaucoup de filtres permettant la réduction de l'ampleur du bruit au niveau des images avant l'application des propres méthodes d'extraction de contours. Les filtres les plus communs sont le filtre gaussien et le filtre médian,
- un post-traitement : pour la plupart des méthodes de cette approche, les contours sont imprécis. Ces contours sont souvent constitués d'une multitude de lignes ouvertes définissant mal les bordures d'un objet. Le traitement le plus répandu est un seuillage par hystérésis pour ne garder que les points les plus importants des contours et pour éliminer en même temps les chaînes de points correspondant au bruit. Ensuite, les points d'un même contour sont reliés entre eux pour former des contours fermés délimitant un objet réel au niveau de l'image.

La délimitation des contours au niveau d'une image n'est pas une segmentation en soi. Elle génère simplement une image binaire : contour/non contour. Il est nécessaire de procéder à un traitement supplémentaire pour obtenir une image partitionnée en régions homogènes.

6.4 Approche pixellaire

6.4.1 Introduction

Cette approche prend essentiellement en compte la couleur du pixel en vu de sa classification. Elle fait abstraction de la disposition spatiale des pixels au niveau de l'image. En effet, elle se base sur le principe que la couleur est une propriété constante de la surface des objets. En

conséquence, l'analyse des coordonnées couleur rattachées aux pixels d'une image donnée est le moyen indispensable pour effectuer la segmentation. L'approche pixellaire comprend principalement deux stratégies :

- une première stratégie exploite la distribution statistique des couleurs présentes au sein de l'image couleur. Les diverses classes de pixels se manifestent à travers les **histogrammes** de l'image par des modes plus au moins importants selon le nombre de pixels qui leurs sont rattachés. Le but de la segmentation est d'établir les différents seuils permettant la séparation de ces modes. Différentes publications ont été consacrées aux algorithmes qui décrivent le seuillage de l'histogramme [Skarbek et Koschan 1994].
- Selon une seconde stratégie, il est considéré qu'un objet ou une région de l'image est globalement composée par des pixels ayant des couleurs similaires, cet objet ayant tendance à former un nuage compact de points au niveau de l'espace couleur RGB. Dès lors, la segmentation peut être réalisée directement dans l'espace couleur 3D. Une étiquette est alors affectée à chaque nuage de points. La segmentation de l'image couleur est alors obtenue par une simple correspondance entre les pixels de l'image et l'espace couleur étiqueté. La segmentation de l'espace couleur fait appel aux techniques de classification des données multidimensionnelles appelées aussi méthodes de **clustering**.

Les deux stratégies de segmentation seront détaillées dans ce qui suit :

6.4.2 Classification basée sur l'histogramme

Le seuillage de l'histogramme est parmi les techniques les plus utilisées en traitement d'images en niveaux de gris [Lucchese et Mitra 2001] [Cheng et al 1999]. L'histogramme peut être divisé en plusieurs parties selon le nombre de modes qu'il comporte. Chaque mode correspond à une classe de pixels au niveau de l'image définissant ainsi une ou plusieurs régions. Le mode ressemble plus au moins à une forme de cloche, il comporte un pic ou un sommet. Les modes sont séparés entre eux par des vallées. La segmentation consiste à établir des seuils au niveau de ces vallées, pour réaliser la séparation des différents modes.

La méthode la plus simple pour segmenter les images couleur est de réaliser un seuillage sur l'histogramme 1D de l'une de ses composantes couleur RGB [King et al 2002]. Une autre alternative consiste à construire une nouvelle composante couleur à partir des composantes RGB. Dès lors, on revient à un schéma de seuillage d'histogramme classique en niveau de gris.

Les approches les plus performantes analysent les histogrammes des trois composantes couleur simultanément pour une meilleure prise en compte de l'aspect multidimensionnel des images couleur. Parmi ces approches, on peut citer les techniques qui se basent sur l'analyse récursive des histogrammes 1D en vue de définir les différents modes existants. Ces techniques partagent les étapes suivantes [Trémeau et al 2004] [Skarbek et Koschan 1994] :

1. poser un masque sur toute la surface de l'image couleur,
2. réalisation des histogrammes 1D correspondant au masque,
3. recherche du mode le plus important parmi les modes de tous les histogrammes,
4. si le mode comporte un nombre important de pixels :
    5. création d'une nouvelle classe avec les pixels de ce mode,
    6. retirer les pixels de ce mode du masque de l'image,
    7. revenir à l'étape 2.

Par contre, le choix des composantes couleur, la manière de déterminer le mode le plus important et le critère d'arrêt de l'analyse récursive, diffèrent d'une technique à l'autre. Ainsi, Ohlander [Ohlander et al 1978] recherche les modes au niveau de 9 histogrammes issus des 3 espaces couleur : RGB, YIQ et LTS. Le mode le plus important est détecté selon des mesures liées principalement à son amplitude et sa largeur. L'analyse d'un tel nombre d'histogrammes (9), au lieu des trois classiques (RGB), augmente la probabilité de capturer des modes importants et correspondants à des régions connexes au niveau de l'image. On note que cette méthode commence à détecter les grandes régions, puis progresse selon un ordre décroissant vers les petites régions. L'analyse récursive s'arrête lorsque la population du mode est inférieure à un seuil.

Ohta et al [Ohta et al 1980] adoptent la même approche d'Ohlander, mais ils l'appliquent seulement sur les trois histogrammes de l'espace couleur $I_1I_2I_3$. Les 3 composantes s'obtiennent par de simples transformations linéaires des composantes RGB :

$$I_1 = (R + G + B) / 3, \qquad I_2 = R - B, \qquad I_3 = (2G - R - B) / 2$$

En effet, ce système est une approximation des 3 axes décorrélés obtenus par la transformation de Karhunen-Loeve (KL). Son utilisation, comparée à l'utilisation d'un espace issu directement de la transformation KL, permet un gain de temps important. Les auteurs notent que les résultats de segmentation sont meilleurs avec ce type de système décorrélé.

Tominaga [Tominaga 1986] propose d'effectuer l'analyse récursive sur les histogrammes de l'espace de Munsell. Le mode le plus important est celui qui maximise la fonction suivante :

$$F = \frac{S_M}{S_H} * \frac{100}{L_M}$$

Où S_M est la surface du mode, L_M est la largeur à mi-hauteur du mode et S_H est la surface totale de l'histogramme. Le processus itératif s'arrête lorsque la surface du mode le plus important est suffisamment petite par rapport à la surface totale de l'image. Les pixels restants sont annexés à l'une des régions voisines selon un critère de similarité couleur. L'inconvénient de cette approche est la manière dont sont calculées les composantes HCV de l'espace de Munsell. En effet, il n'existe pas de formule pour générer ces composantes. Elles sont obtenues grâce à une table de conversion entre les composantes RGB et HCV [Wyszecki et Stiles 1982]. Tominaga prend en compte ces considérations et a modifié sa technique sur une autre publication [Tominaga 1990]. Il a remplacé en premier lieu le système de Munsell par l'espace couleur *Lab*. Ensuite, il a réalisé une ACP lors de chaque itération de l'algorithme. Les histogrammes générés correspondent aux trois premiers axes issus de l'ACP. Cette approche rallonge les temps de traitement, mais en contrepartie, offre une analyse plus fine des pixels.

Au lieu d'adopter une approche marginale basée sur l'étude des histogrammes 1D, la méthode qui semble la plus intuitive est l'analyse directe de l'histogramme 3D. Des classes de pixels séparables dans un espace couleur ne se manifestent pas toujours en modes distincts [Trémeau et al 2004]. En dépit des avantages de l'histogramme 3D qui prend parfaitement compte de

l'aspect tricolore des pixels, l'inconvénient majeur d'une telle structure est l'espace mémoire qu'elle peut occuper. Sachant que chaque axe de l'histogramme 3D comprend 256 valeurs possibles, l'espace total de cet histogramme comprend 256^3 cellules. Le découpage de l'histogramme 3D n'est pas une tâche triviale et la délimitation des classes de couleur nécessite un temps de calcul non négligeable [Haralick et Shapiro 1985]. Dans le but de faciliter l'analyse de l'histogramme 3D, Schacter et al structurent celui-ci en arbre binaire [Schacter et al 1975]. Chaque nœud stocke un couple : une valeur RGB et son occurrence au sein de l'image. Ce type de structures s'adapte bien aux algorithmes récursifs, offrant ainsi une accélération de la segmentation.

Un autre groupe de chercheurs suggère de porter l'analyse sur un histogramme 2D comme alternative pour éviter la complexité liée au traitement de l'histogramme 3D [Holla 1982] [Von Stein et Reimers 1983] [Shafarenko et al 1998]. Toutes ces méthodes opèrent dans des espaces couleur de type luminosité-chromaticité. La segmentation est effectuée toujours au niveau de l'histogramme 2D du plan chromatique. La séparation des classes couleur est réalisée essentiellement grâce à la LPE.

6.4.3 Classification basée sur les méthodes de « *clustering* »

Cette approche dérive des méthodes de classification multidimensionnelles. En général, ces méthodes se servent de différentes primitives afin de partager les données à classifier en classes d'individus similaires. Concernant l'espace de représentation des images couleur, les coordonnées trichromatiques des pixels sont tout naturellement les primitives utilisées pour réaliser ce clustering. En effet, l'espace de couleur 3D est abordé comme un espace de variables multiples. Les techniques de clustering tentent d'organiser les nuages de points couleur associés aux pixels de l'image afin qu'ils forment un nombre de classes données. Généralement, les méthodes de clustering sont des techniques de classification non supervisées. On ne possède aucune connaissance *à priori* sur les données à classifier. Les procédures de segmentation basées sur le clustering s'adaptent avec ce type de classifications non supervisées en injectant le nombre de classes de couleur à retrouver. Dans la plupart des cas, la connaissance de ce nombre nécessite l'intervention des approches supervisées telles que la visualisation d'un set d'images d'apprentissage par un expert, par exemple pour définir

le nombre de classes. En résumé, les techniques de clustering partitionnent l'espace couleur 3D en sous-régions correspondantes au nombre de classes d'appartenance des pixels de l'image à segmenter.

Parmi les méthodes de clustering les plus utilisées en classification des images couleur, on peut citer la méthode des k-means [McQueen 1967]. Cette approche consiste à construire une collection de k classes, où k est le nombre de classes possibles associées aux types d'objets présents au niveau de l'image. Les classes de couleur doivent respecter deux conditions :

- les couleurs doivent être similaires au sein d'une même classe,
- les couleurs sont dissimilaires si elles appartiennent à deux classes différentes.

En conséquence, le k-means repose essentiellement sur la minimisation de la variabilité couleur à l'intérieur des classes et sur la maximisation de cette variabilité couleur entre les différentes classes. Lors de la phase initiale, pour chaque classe, une couleur est choisie aléatoirement parmi les couleurs de l'image. Cet ensemble de couleurs sélectionnées forme les centres de gravité initiaux des classes. Ensuite, et pour chaque pixel de l'image une distance couleur est calculée vers les centres de gravité des classes ; le pixel est attribué à la classe ayant le centre de gravité le plus proche. Le centre de gravité d'une classe est recalculé et mis à jour à la suite de chaque rajout de pixel. Le k-means converge vers un minimum local de la somme des distances couleur à l'intérieur de chaque classe. Ce minimum ne coïncide pas toujours avec le minimum absolu, cela dépend des positions initiales des centres de gravité. Leydier et al [Leydier et al 2004] ont utilisé l'algorithme du k-means dans le but de segmenter les images couleur de manuscrits médiévaux.

Il existe dans la littérature [Wu et al 1994] [Kamei-Harada et al 1994] une variante du k-means, le c-means ou encore le *fuzzy* c-means, se basant sur une approche de classification floue. Au lieu de rattacher un pixel vers une seule classe, le c-means peut l'attribuer à plusieurs classes à la fois selon des degrés d'appartenance qui varient entre 0 et 1. Il est basé sur la minimisation de l'erreur quadratique à l'intérieur de chaque classe. Comme le k-means, le c-means requiert aussi la connaissance préalable du nombre de classes. La désignation finale de la classe d'appartenance d'un pixel donné nécessite la fixation d'un seuil d'appartenance. Ce seuil permet de trouver la classe ayant le maximum de probabilité ou de

chance à contenir ce pixel. On note que Huntsberger et al [Huntsberger et al 1985] ont adopté l'approche c-means dans le but de la segmentation couleur au niveau des espaces couleur RGB et $I_1I_2I_3$.

La segmentation par les réseaux de neurones (RN) constitue un autre groupe qui dérive des techniques de classification multidimensionnelle et d'intelligence artificielle. Le concept des RN se base sur le mécanisme qui régit la transmission de l'information entre les cellules nerveuses. Ainsi, les RN sont constitués d'un grand nombre de cellules interconnectées entre elles par de simples fonctions. Les RN interviennent d'habitude après la phase de segmentation. Ils sont utilisés en imagerie comme outil de classification se servant des primitives générées par la segmentation. A titre d'exemple, les RN ont été adoptés pour réaliser la classification des cellules tumorales dans le but de définir leurs grades de cancérisation [Kok et al 2001] [Romeo et al 1998]. Les RN sont constitués de plusieurs couches de cellules, mais le nombre de trois couches est le plus commun, car il est démontré que ce nombre suffit pour prendre des décisions complexes de classification [Lucchese et Mitra 2001]. La première couche sert de point d'entrée des primitives. Dans notre cas, ce sont les coordonnées couleur des pixels. La couche externe délivre les résultats de la classification. Les RN nécessitent une phase d'apprentissage. Le plus souvent, des utilisateurs ou des experts d'une application d'imagerie donnée spécifient manuellement les différentes classes d'appartenance des pixels. Pendant cette phase, les RN fonctionnent par rétro-propagation, c'est-à-dire de la couche de sortie vers la couche d'entrée. En effet, les RN ajustent les fonctions qui serviront par la suite à assurer la classification. Ensuite, et durant la phase de décision, les RN attribuent les pixels vers leurs classes respectives en minimisant une fonction d'énergie. A titre d'exemple, Funakubo [Funakubo 1994] utilise des RN pour séparer les cellules du fond d'images médicales. On peut citer dans ce cadre l'article de Zheng et al [Zheng et al 2004] qui présentent un RN capable de définir différents types cellulaires au niveau des images histologiques. Une fenêtre de taille fixe suffisamment large pour englober une cellule, glisse à travers toute la surface de l'image, les RN décident du type de la cellule cadrée par la fenêtre en se basant seulement sur les couleurs des pixels qui la composent.

La partition de l'espace couleur 3D peut être effectuée également par la technique de classification issue de l'analyse en composante principale (ACP). Ainsi, Umbaugh et al

[Umbaugh et al 1993] utilisent le premier axe principal obtenu par l'ACP pour une division récursive de l'espace couleur. Cette approche est adoptée pour identifier le type tumoral du cancer de la peau. La division débute par l'espace couleur entier, puis opère au niveau du sous-espace le plus volumineux. Pour chaque itération, une ACP est réalisée sur le sous-espace à diviser. Le découpage est effectué selon le premier axe principal qui est le plus discriminant. Le processus de division s'arrête lorsque le nombre de sous-espaces ou boxes correspond au nombre de classes ou au nombre de types tumoraux qui est connu d'avance.

6.4.4 Conclusion

Nous avons vu dans cette partie, que la classification pixellaire est basée essentiellement sur les deux approches : seuillage de l'histogramme et clustering de l'espace couleur. Hormis les classes de couleur bien localisées en régions connexes, les techniques de seuillage automatique génèrent souvent des petits groupes de pixels mal classifiés. En effet, cette approche néglige la disposition spatiale des pixels au niveau de l'image, et ne considère que la couleur des pixels. De ce fait, un post-traitement est nécessaire pour rattacher les petites régions ou groupes de pixels à la région la plus similaire en couleur parmi l'ensemble des régions voisines.

Nous avons signalé également que l'inconvénient majeur des techniques de clustering est leur incapacité à définir le nombre de classes d'appartenance des couleurs pour une application donnée. De ce fait, elles nécessitent une intervention de l'utilisateur pour spécifier le nombre de classes avant de réaliser la segmentation. Une solution est apportée par Lim et al [Lim et al 1990], qui associent une technique de seuillage sur histogramme au clustering pour réaliser une méthode non supervisée. Dans un premier temps, ils analysent les histogrammes de plusieurs composantes couleur (RGB, YIQ et $I_1I_2I_3$) dans le but de définir les modes les plus importants. Les modes dont leurs populations de pixels dépassent un certain seuil sont retenus comme classes valides. Les pics de ces modes constituent les centres de gravité de ces classes. Ainsi, on obtient le nombre de classes représentées par leurs centres de gravité. Dès lors, la deuxième phase de classification couleur peut débuter. Au cours de cette phase, les auteurs [Lim et al 1990] ont adopté une approche de classification floue (fuzzy c-means), pour rattacher les pixels couleur au centre de gravité le plus probable.

6.5 Approche hybrides

A travers les sections précédentes, nous avons passé en revue les trois principales approches adoptées dans le domaine de la segmentation des images couleur. Nous avons noté également que ces trois approches : pixellaire, régions et contours correspondent à trois conceptions différentes de l'expression colorimétrique et de la disposition spatiale des objets au niveau des images. En dépit des spécificités et avantages de ces approches, chacune d'elle présente pourtant des inconvénients. En effet, l'approche pixels souffre de son manque de prise en charge de l'interaction du pixel avec son voisinage. Les performances liées à l'approche régions sont tributaires des positions initiales des germes. Celles-ci représentent les points de départ pour la construction des régions. Un dépôt de germes non étudié, et qui ne correspond pas aux centres réels des objets, conduit à une segmentation erronée. Nous avons signalé aussi que les techniques qui s'appuient sur l'approche contours sont très sensibles au bruit inhérent au processus d'acquisition des images.

Plusieurs travaux ont été menés pour développer de nouvelles techniques dans le but de résoudre les problèmes cités ci-dessus. En général, les techniques proposées répondent principalement à deux exigences :

- elles doivent tirer profit des avantages des approches classiques,
- elles doivent être suffisamment robustes pour éviter les inconvénients liés aux trois approches traditionnelles.

La voie qui semble la plus commune entre ces techniques est celle qui associe deux approches voire trois simultanément. De manière globale, les travaux proposés dans la littérature se subdivisent en deux groupes essentiels :

- **Approche mixte régions/contours :**

Lancet [Skarbek et Koschan 1994] applique une détection de contours suivie d'une croissance de régions pour effectuer la segmentation d'images couleur codées dans l'espace *Lab*. La stratégie de segmentation fonctionne en deux temps : en premier lieu, il adopte une approche

marginale en calculant la dérivée première sur les plans couleur L, a et b grâce au filtre de Deriche. Un seuillage est appliqué pour extraire les contours qui correspondent aux maxima locaux. Il réitère toute l'opération depuis le début, en calculant cette fois-ci, la dérivée seconde des plans marginaux pour extraire d'avantage de points contours qui coïncident avec le passage par zéro de la dérivée seconde. Les contours sont ensuite obtenus par union des pixels contours extraits par les dérivées première et seconde au niveau des trois plans L, a et b. A l'issue de la première partie, certains contours restent formés de segments de lignes non fermées. Dès lors, la deuxième phase prend le relais. Lancet adopte une stratégie de croissance de régions pour fermer en premier temps, ensuite pour remplir la surface des objets délimités par les contours ainsi fermés.

Zugaj et Lattuati [Zugaj et Lattuati 1998] adoptent une approche similaire pour segmenter les images couleur RGB. La différence réside dans la manière d'extraction des contours. En effet, ils remplacent la méthode de Lancet par l'application directe du gradient couleur de Di Zenzo au niveau de l'image couleur.

Dans ce cadre, on peut évoquer également l'approche de Zhu et Yuille [Zhu et Yuille 1996] qui combinent la technique des contours actifs ou déformables avec une croissance de régions. La stratégie adoptée est basée sur la déformation des régions en minimisant un critère qui décrit la longueur des contours.

- **Approche mixte pixels/régions :**

Dans cette catégorie d'approche, on peut citer le travail de Tseng et Chang [Tseng et Chang 1992] qui associent un seuillage d'histogramme avec une croissance de régions. En effet, ils se servent des modes les plus importants des histogrammes de teinte et de luminosité pour un premier repérage des régions chromatiques et non chromatiques. Une analyse des régions est nécessaire en deuxième temps pour une délimitation plus fine des régions. Enfin, ils appliquent une procédure de fusion de régions pour surmonter la sursegmentation et pour rattacher les petites surfaces de pixels connexes aux régions adjacentes les plus similaires.

Holla [Holla 1982] applique la ligne de partage des eaux qui est assimilée à une croissance de régions pour segmenter l'histogramme 2D d'un espace couleur basé sur le principe des couleurs opposées. La LPE sert à définir les différentes classes couleur présentes au niveau de l'histogramme du diagramme de chromaticité. Holla suggère de faire un post-traitement qui prend en charge les relations de voisinage pour intégrer les groupes de pixels isolés dans les régions proches.

Malpica et al [Malpica et al 1997] appliquent un seuillage pour isoler les noyaux cellulaires du fond de l'image. Ensuite, ils calculent une carte de distances à partir du masque binaire des noyaux. Ainsi, les pixels des noyaux prennent comme valeur leur distance respective du fond de l'image. Les maxima locaux, c'est-à-dire les pixels les plus éloignés du fond, sont retenus comme germes pour la construction des régions par la LPE. Nilsson et Heyden [Nilsson et Heyden 2005] améliorent l'algorithme de Malpica et al en intégrant des conditions supplémentaires telles que le diamètre des objets à reconstruire. Ils ont adopté cette stratégie pour détecter les noyaux cellulaires de la moelle épinière regroupés en amas. Kim et al [Kim et al 2006] se réfèrent aussi à l'algorithme de Malpica et al, dans le but de calculer le pourcentage de cellules en cycle au niveau des images histologiques. De plus, ils affinent la phase de seuillage de l'histogramme grâce à une analyse préalable des couleurs des cellules en cycle et en repos.

Pour terminer cette section, nous citons deux exemples illustrant des cas d'études qui intègrent les trois types d'approches : pixellaire, régions et contours à la fois. Ainsi, Loukas et al [Loukas et al 2003] décrivent une stratégie coopérative pour calculer le nombre des cellules prolifératives au sein des coupes biologiques. Ils débutent la segmentation par un repérage des contours suite à l'application du filtre *LOG* (Laplacian of Gaussian). Il s'agit d'un opérateur de dérivée seconde (filtre Laplacien) réalisé après le lissage de l'image par un filtre gaussien. Dans un deuxième temps, une approche régions est utilisée pour remplir la surface des cellules ou amas de cellules entourées par les contours délimités auparavant. Les auteurs se servent de la carte des distances pour détecter les maxima locaux qui correspondent aux centres de gravité des cellules. En effet, la carte des distances favorise le repérage les formes circulaires tels que des noyaux cellulaires. Elle permet aussi une séparation optimale des cellules organisées en amas. La dernière phase consiste à effectuer une ACP sur les pixels

appartenant aux noyaux cellulaires, suivie d'un seuillage réalisé sur l'histogramme du premier axe issu de l'ACP. Cette classification pixellaire conduit à séparer correctement les cellules prolifératives des cellules en repos.

Huntsberger et Descalzi [Huntsberger et Descalzi 1985] adoptent une technique originale qui applique une classification floue pour définir en même temps les régions et les contours. En effet, Ils utilisent le *fuzzy* c-means qui est à l'origine une méthode de classification pixellaire pour réaliser la segmentation au niveau des espaces couleur RGB et XYZ. Le point fort de leur approche est l'adaptation de ce type de classification floue aux deux approches : régions et contours. En premier lieu, le degré d'appartenance de chaque pixel à ses différentes régions voisines est calculé. Ensuite, les pixels sont rattachés soit aux régions, soit aux contours. Les règles de décisions sont très simples, un pixel donné est rajouté à la région pour laquelle il possède le degré d'appartenance le plus élevé. S'il appartient à deux régions avec le même degré, il est rattaché aux contours. Cela signifie qu'il fait partie de la frontière qui sépare deux régions voisines. Les contours sont définis également dans les endroits où deux pixels voisins appartiennent à deux régions différentes.

6.6 Conclusion

Tout au long de ce chapitre, nous avons tenté de présenter les méthodes de segmentation couleur les plus connues en littérature. Nous avons vu également que ces méthodes sont organisées selon trois principales approches : pixellaire, régions et contours. Chaque approche a été mise en relief. Nous avons exposé notamment sa conception théorique, son principe d'action et quelques exemples de travaux illustrant son application dans différents domaines d'imagerie. Nous nous sommes aussi attachés à souligner ses avantages et points forts, mais aussi ses principaux inconvénients et limitations. Des tentatives de faire coopérer ces approches ont été également signalées à travers quelques exemples d'algorithmes de segmentation couleur. À la lumière des différents travaux et éléments présentés au cours de ce chapitre, nous pouvons tirer les conclusions suivantes :

- Il n'existe pas de méthode de segmentation qui soit universelle et adaptée à tous les domaines de traitement d'images. Si un algorithme de segmentation peut donner des

résultats satisfaisants pour une application donnée, il peut s'avérer défaillant et complètement inadapté pour un autre type d'application.

- La plupart des techniques exposées nécessitent la fixation de certains paramètres et seuils. Le réglage de ces seuils est fonction du type d'images considérées, il n'est pas toujours trivial et requiert l'intervention d'un utilisateur expérimenté. De ce fait, la généralisation de ces techniques pour segmenter d'autres types d'images n'est pas toujours possible.
- Les procédures de segmentation les plus performantes sont celles conçues pour résoudre un problème de segmentation bien précis et pour un type d'images connu d'avance. En conséquence, il est primordial de procéder à une analyse minutieuse des caractéristiques des images à traiter avant toute conception de schémas de segmentation. *A priori*, il faut savoir la finalité de cette segmentation et connaître parfaitement les primitives des objets et des structures à extraire de l'image.
- D'une manière générale, les auteurs des travaux cités dans ce chapitre se contentent de présenter leurs approches. Ils les appliquent sur un jeu d'images données en ignorant le plus souvent de les comparer à d'autres méthodes ou algorithmes existants dans la littérature.
- Certains auteurs ont tenté de proposer différents critères pour comparer les résultats de segmentation [Borsotti et al 1998] [Foliguet et Guigues 2006]. Ces critères sont basés sur des mesures quantitatives telles que : l'homogénéité et la surface des objets détectés. Nous pensons que ces critères ne prennent pas suffisamment en compte l'aspect psycho-visuel de la perception humaine des couleurs. Un critère intégrant de telles mesures qualitatives est difficile à réaliser. D'autre part, il paraît insensé de comparer avec ce type de critères des segmentations qui sont développées à la base pour des domaines d'application complètement différents. Nous pensons que la comparaison doit être limitée aux méthodes de segmentation qui sont spécifiques à une application donnée. Dans certains domaines comme par exemple l'imagerie médicale, la comparaison doit parfois faire appel à des évaluations ou des expertises visuelles.
- Il semble que la séparation des processus de segmentation et de reconnaissance des objets soit la raison des mauvaises performances de certaines méthodes de segmentation. À notre sens, la segmentation et la reconnaissance des objets doivent

être intégrées dans un seul schéma de classification globale. C'est la manière la plus naturelle pour prendre en charge des contraintes liées à une application particulière.

- Les performances des procédures de segmentation sont en fonction de l'approche adoptée, mais également de l'espace couleur choisi. Nous avons abordé plusieurs exemples de segmentation qui opèrent sur un espace couleur autre que l'espace RGB. Nous avons cité même des techniques qui utilisent plusieurs espaces couleur à la fois. Comme il n'existe pas de méthode de segmentation universelle, le même constat est valable pour les espaces couleur. A ce niveau, nous réitérons la conclusion tirée dans le chapitre consacré aux espaces couleur : il n'existe pas un espace couleur qui soit adapté à tout type d'images et à tout schéma de segmentation couleur. Cependant, la tendance générale des exemples de segmentation traités au cours de ce chapitre est l'utilisation des espaces couleur décorrélés. En effet, ces espaces présentent l'avantage de séparer la couleur selon deux informations : chromatique et non chromatique qui correspond à l'axe de luminance.

A la lumière de toutes ces considérations et dans le but de proposer un schéma de segmentation qui serait adapté à notre sujet de thèse, soit l'analyse automatique des images histologiques, il paraît important de prendre en compte tous les points relevés au niveau de cette section.

En premier lieu, nous devons bien analyser nos images à traiter. Ensuite, nous devons identifier exactement les objets et structures à extraire des images, notamment leurs caractéristiques colorimétriques, leurs formes et leurs organisations sur la surface des images.

Dans le cadre de notre sujet de thèse, nous travaillons sur les approches automatiques pour la quantification de l'expression des marqueurs tumoraux au niveau des coupes histologiques. Suivant les différents enseignements tirés au niveau de ce chapitre, nous devons analyser au préalable les caractéristiques colorimétriques du marquage et de la contre-coloration, avant la conception du schéma du traitement d'images automatique.

Une fois que toutes les contraintes liées au sujet sont connues, nous pouvons entamer la phase de segmentation proprement dite. On peut parler à ce stade de classification, puisque, comme mentionné ci-dessus, la segmentation peut être fusionnée avec la phase de classification. A ce niveau, nous proposons de réaliser la classification au niveau d'un système de couleur décorrélé en accord avec une grande partie des exemples de segmentation abordés dans ce chapitre.

Finalement, nous devons valider le schéma de segmentation ou de classification retenu. La méthode proposée doit être comparée à d'autres techniques de segmentation développées pour le même type d'images et pour le même type de problématiques.

Toutes ces étapes seront reprises en détails dans le chapitre suivant qui sera consacré à la classification des images de coupes histologiques. En effet, nous allons tenter au niveau de ce chapitre de proposer une nouvelle méthode de classification capable de faire la ségrégation automatique entre les cellules positives et les cellules négatives pour les marqueurs tumoraux en IHC. Notre approche sera confrontée à d'autres types de segmentation cités dans la littérature et développés pour le même but.

Chapitre 7

Classification couleur des images histologiques

Le facteur limitant pour l'adoption de la technologie TMA dans les laboratoires reste le volume horaire important nécessaire à l'analyse manuelle du grand nombre d'images histologiques produites par ce type de biotechnologies. La solution est sans doute l'analyse automatique des images histologiques. Dans ce contexte, nous avons développé une approche automatique pour l'analyse des images histologiques. Cette approche est illustrée à travers la quantification de l'expression du marqueur nucléaire Ki67.

Le diagramme de chromaticité xy est adopté pour l'analyse colorimétrique du marquage et de la contre-coloration. En se basant sur ce modèle couleur, nous avons construit une table de classification 2D permettant la séparation couleur entre le marquage et la contre-coloration.

Notre méthode de classification automatique a été comparée avec une approche de classification manuelle de référence. Elle a été comparée également à d'autres types de classifications automatiques de la couleur citées dans la littérature.

Les résultats obtenus témoignent de la conformité de notre méthode de classification avec la méthode de classification de référence. Notre approche s'est avérée, la classification automatique, générant le moins d'erreurs de classification par rapport à la classification de référence. La table de classification 2D révèle notamment son efficacité et sa stabilité, pour classer correctement les images histologiques, même dans les conditions les plus extrêmes de luminosité ou de saturation.

7.1 Introduction

L'IHC est une technique d'immunomarquage largement utilisée dans les laboratoires de recherche. Elle permet la révélation au niveau des coupes histologiques des profils d'expression des marqueurs du cycle cellulaire. L'amélioration constante de cette technique (IHC) a permis la réalisation d'importantes avancées pour l'investigation des mécanismes de fonctionnement des différents marqueurs du cycle cellulaire [Seidal et al 2001]. La dérégulation de l'expression de ces marqueurs est à l'origine du processus de cancérogénèse (§ 1.7.2). Dès lors, ces marqueurs deviennent des indicateurs de diagnostic, de pronostic et de suivi thérapeutique, pour les différents types de cancers [Fitzgibbons et al 2000] [Hensel et al 2002] [Korkolopoulou et al 2002]. L'interprétation visuelle des coupes histologiques sous un microscope par un pathologiste, pour l'évaluation de l'expression et de la distribution de ces marqueurs, reste la procédure classique au niveau des laboratoires d'anatomopathologie clinique.

Toutefois, le recours à la numérisation des lames histologiques est devenu une pratique courante au sein des centres médicaux et des laboratoires de recherche. Les lames numériques présentent l'avantage d'être conservées indéfiniment, faciles à stocker et disposées aux techniques d'analyse d'images automatiques par rapport aux lames histologiques classiques. De plus en plus de microscopes sont reliés à des dispositifs d'acquisition d'images couleur. Ce processus de numérisation trouve toute sa place dans les techniques de biotechnologies haut-débit telles que la technologie des TMA. Actuellement, une station d'imagerie moderne peut numériser les coupes histologiques d'un bloc TMA dans un temps record, permettant ainsi l'étude accélérée d'un grand éventail de molécules à la fois. Le temps pour la mise en évidence des marqueurs tumoraux caractéristiques d'un type de cancer donné s'en trouve alors réduit. C'est grâce à ce type d'approche, que des marqueurs tumoraux tels que le Ki67, la ß-Cathénine, la Cycline D1, le Bcl2, le RP, le ER, le P53 et le P16 ont été établis comme des indices de pronostic dans différents types de cancers [Angellau 2000] [Straume et al 2000] [Maelandsmo et al 1996] [Schneider-Stock et al 2005] [Chung et al 2001] [Ginestier et al 2002] [Torhost et al 2001].

Cependant, le facteur limitant de cette chaîne de technologies dites haut-débit reste le volume horaire important nécessaire à l'analyse manuelle du grand nombre d'images histologiques produites [Vrojlik et al 2003]. Ainsi, Watanabe et al rapportent que l'analyse visuelle de toutes les coupes d'un bloc de TMA peut nécessiter plusieurs centaines d'heures à un pathologiste [Watanabe et al 2005]. Rabinovich et al rapportent que le temps nécessaire pour l'évaluation manuelle des scores de marquage d'un bloc TMA, occupe un volume horaire 10 fois plus important que le temps alloué à la construction de ce bloc TMA [Rabinovich et al 2006].

En IHC, le Di-Amino-Benzidine (DAB) est le chromogène de choix pour révéler l'expression des marqueurs tumoraux au niveau des coupes histologiques. Le DAB se manifeste par une couleur marron proportionnelle à la quantité du marqueur étudié. Ce type de coloration est hautement spécifique, stable et offre des résultats reproductibles [Aziz et Barathur 1994]. L'Hématoxyline est un colorant bleu non spécifique. Il est souvent utilisé comme contre-colorant en IHC pour teinter les noyaux négatifs au marquage. La Figure 7.1 illustre l'utilisation du couple DAB-Hématoxyline pour la révélation de l'expression de la protéine Ki67, qui est un marqueur nucléaire de la prolifération cellulaire.

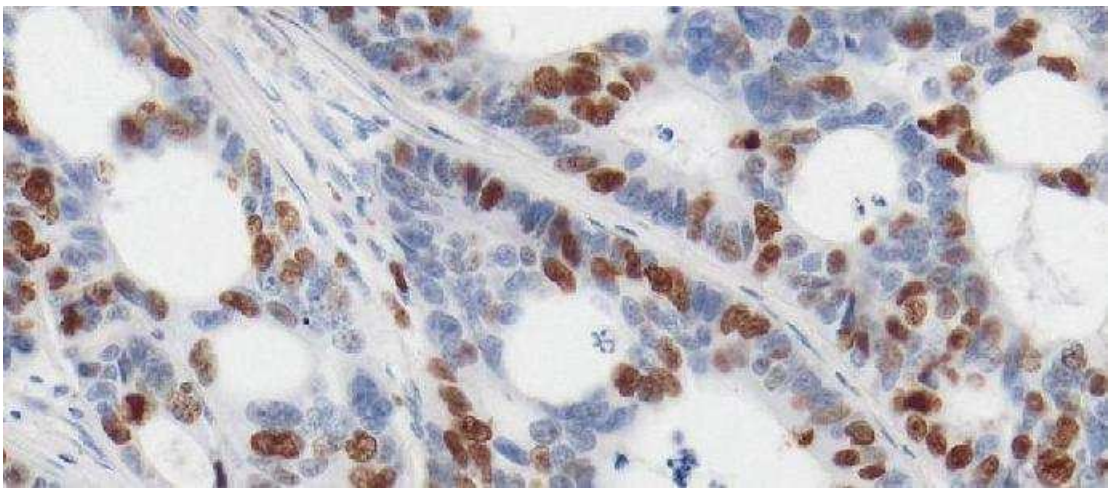


Figure 7.1: Lame histologique immuno-marquée pour la détection de la protéine ki67. Le marquage ki67 permet l'étude de la prolifération cellulaire : les noyaux en cycle apparaissent marqués par la couleur marron caractéristique de l'immuno-marquage DAB, les noyaux hors cycle prennent la couleur bleue de la contre-coloration Hématoxyline.

Après la préparation et la coloration des coupes histologiques, l'analyse de l'expression d'un marqueur donné à travers le couple DAB-Hématoxyline peut être effectuée par un pathologiste ou un utilisateur expérimenté en histologie (histo-technicien). L'analyse se base sur l'inspection visuelle de la lame histologique sous un microscope optique ou encore sur un écran d'ordinateur si le microscope est équipé d'un dispositif de numérisation des lames histologiques. L'évaluation de l'expression du marquage se résume habituellement à faire la distinction entre les cellules positives et négatives au marquage sur la base de la différence de couleur entre le DAB et l'Hématoxyline [Seigneurin et al 1989].

Un comptage manuel des cellules positives et négatives fournit une quantification chiffrée de l'expression du marqueur étudié. Cependant, il existe plusieurs méthodes de comptage. Certaines consistent à calculer l'indice de marquage qui correspond au pourcentage de cellules positives au marquage par rapport à la totalité des cellules analysées [Nocito et al 2001]. D'autres se limitent à utiliser un score binaire : le champ microscopique ou l'image numérique inspectée est considéré comme positif au marquage si le pourcentage de cellules positives, évalué globalement sans dénombrement, dépasse un seuil donné, celui-ci pouvant varier d'une méthode à une autre [Camp et al 2000] [Torhost et al 2001] [Hoos et al 2001] [Fernebro et al 2002] [Abd-El-Hameed 2005]. Un autre groupe de méthodes divise l'expression du marquage selon plusieurs niveaux établis selon différents seuils d'indice de marquage. Pour cela, certains auteurs utilisent des adjectifs comme : absent, faible, modéré ou fort [Engellau et al 2001], d'autres utilisent un score ordinal : 0, +1, +2 ou +3 [Simony 2007] [Hilbe et al 2003]. D'autres encore se servent d'un score mixte : négatif, rare, +1, +2, +3 ou +4 [Merseberger et al 2003]. Au vu de toutes ces méthodes, on constate qu'il n'existe pas de méthode manuelle standard dans la littérature pour évaluer l'expression d'un marqueur donné. Cette variabilité de méthodologies conduit à une diversité de comptage du marquage entre les laboratoires [Seidal et al 2001] [Wick et Mills 2001].

L'inconvénient que l'on peut rattacher à l'évaluation manuelle de l'expression d'un marqueur donné, est la subjectivité inhérente à la vision humaine pour classifier les couleurs. En effet, il est plus facile à l'œil humain de détecter la couleur DAB si l'expression du marquage est forte. Par contre, une faible quantité de marquage se traduit par une coloration marron (DAB) très atténuée qui se confond avec la couleur bleue de l'Hématoxyline, ce qui rend plus

subjectif la séparation entre le marquage et la contre-coloration, et conduit souvent à des variabilités de classification couleur entre les observateurs. De plus, la vision du pathologiste est influencée par le contexte dans lequel se présentent les cellules à classifier. L'œil est influencé par la couleur du fond du champ microscopique ou de l'image à inspecter, par la lumière qui éclaire le champ microscopique [Divito et Camp 2005]. En effet, la perception psycho-visuelle d'une couleur varie selon les conditions d'éclairage de la scène à visualiser. Ce phénomène est connu sous le nom de l'effet de Bezold-Brücke, qui démontre que la teinte d'une couleur varie en relation avec le niveau de luminosité [Trémeau et al 2004].

Tous les facteurs cités ci-dessus conjugués à la fatigue visuelle et mentale de l'observateur, après des heures d'inspection et de classification répétitives de nombreuses coupes histologiques, sont une source de variabilités intra et inter observateurs importante. En plus, l'approche manuelle limite considérablement la quantité de matériel biologique analysé. L'automatisation de la phase d'analyse des images histologiques parait nécessaire et évidente pour franchir ces obstacles et affranchir les pathologistes de cette tâche contraignante. Le rôle du pathologiste reste néanmoins central dans l'interprétation et l'expertise des résultats issus de l'analyse automatique. L'analyse automatique des images, permet le traitement d'un grand nombre d'images en un temps réduit, comparée à l'analyse visuelle. Elle discrimine le marquage de la contre-coloration d'une manière objective. En conséquence, l'approche automatique d'analyse d'images permet le développement de méthodes standardisées et reproductibles. C'est alors un moyen indispensable pour réduire la variabilité de quantification du marquage en IHC liée aux évaluations visuelles.

Nous allons présenter au niveau de ce chapitre une nouvelle approche de classification couleur des images histologiques. Une approche qui consiste à séparer le marquage de la contre-coloration en se basant sur les caractéristiques colorimétriques du DAB et de l'Hématoxyline. Cette approche sera illustrée à travers la quantification automatique de l'expression d'un marqueur immunohistochimique : le marqueur nucléaire Ki67, indicateur de la prolifération cellulaire au niveau des images de carottes virtuelles.
Notre méthode de classification couleur sera comparée à d'autres méthodes automatiques citées dans la littérature, ainsi qu'aux procédures de classification manuelles classiques réalisées par trois observateurs.

7.2 Méthodologies de travail

Les méthodes de préparation du matériel biologique : réalisation des coupes histologiques, immuno-marquage et acquisition des images, ont été détaillées dans le chapitre 3 (§ 3.2). Néanmoins, nous allons reprendre ces étapes de manière succincte juste pour rappeler les principes généraux surtout de la coloration immuno-histologique. Ces rappels nous serviront aussi pour introduire les éléments spécifiques à l'étude biologique illustrée dans ce chapitre.

7.2.1 Matériel biologique

42 blocs d'archives de tumeurs du côlon ont été recrutés pour cette étude. Les blocs sont préparés en suivant les grandes étapes classiques utilisées en histologie : fixation, déshydratation et inclusion. Une coupe histologique de 4-5 µm d'épaisseur est réalisée pour chaque bloc de tumeur pour pratiquer un immuno-marquage Ki67.

7.2.2 Immuno-marquage Ki67

La protéine Ki67 est une protéine nucléaire présente au niveau des cellules en cycle lors des phases G1, S et G2, absente au niveau des cellules en repos (cellules en G0) [Salminen et al 2005]. L'expression du marqueur Ki67 est associée à la prolifération tumorale et l'indice de marquage lié à cette protéine présente un indice de pronostic important pour différents types de tumeurs. Un indice de marquage très élevé est souvent associé à un pronostic réservé [Pollack et al 2004] [Bui et al 2004][Tisell at al 2003] [Burger et al 1986] [Cobb et al 1996].

L'expression de la protéine Ki67 est révélée grâce à un anticorps primaire anti-Ki67 lui-même révélé à l'aide d'un anticorps secondaire couplé à une enzyme : la peroxydase. Le Di-Amino-Benzidine (DAB) permet la mise en évidence des sites de la réaction antigénique. En effet, le DAB est un substrat de la peroxydase, l'interaction entre celle-ci et le DAB délivre un signal de couleur brune. L'Hématoxyline est utilisée en un deuxième temps comme contre-colorant pour teinter en bleu les noyaux des cellules qui ne sont pas en cycle. L'évaluation visuelle de

l'expression d'un marqueur donné tel que le Ki67 se base ainsi sur la différence de couleur entre le DAB marron et la contre-coloration bleue.

7.2.3 Préparation des lames de références

Nous avons eu recours à deux lames de référence pour les besoins liés à ce travail : une lame de référence positive et une autre lame de référence négative.

La préparation de la lame de référence positive comporte toutes les étapes usuelles pour la révélation du marquage Ki67, excepté la phase de contre-coloration. En conséquence, une lame de référence positive est obtenue n'exprimant que la couleur marron liée au DAB.

La lame de référence négative a été préparée en suivant toutes les étapes de colorations liées au marquage Ki67 et à la contre-coloration, sauf l'étape d'application de l'anticorps primaire dirigé contre la protéine Ki67. Cette action a pour effet de couper la chaîne de révélation de l'expression du ki67. Ainsi, une lame de référence négative est obtenue exprimant seulement la couleur bleue associée à l'Hématoxyline appliqué lors de la phase de contre-coloration.

7.2.4 Acquisition des images histologiques

La numérisation des 42 coupes histologiques a été réalisée avec le système de microscopie virtuelle SAMBA2050^{TM} (TRIBVIN^{TM}) dont nous disposons au laboratoire. Les images numériques ont été acquises au grossissement x 20.
Différentes zones d'intérêts ont été délimitées par le Dr. Simony au niveau des images des 42 coupes histologiques. Ces zones d'intérêts sont des zones tumorales correspondant à trois types de régions :

- zone A : zone tumorale à la surface de la tumeur dirigée vers la lumière intestinale,
- zone B : zone tumorale en plein développement au milieu de la tumeur,
- zone C : zone tumorale située dans la partie distale de la tumeur au contact des couches internes de l'épithélium du côlon.

Au final, 32 zones A, 24 zones B et 23 zones C ont été ainsi annotées, soit un total de 79 zones.

7.2.5 Génération des images de carottes virtuelles

Des images carrées de la taille d'une carotte TMA (600 µm) sont prélevées au niveau des zones tumorales délimitées par le pathologiste. Ces images baptisées « carottes virtuelles » sont générées à partir du logiciel TMACrop™ (TRBVIN™) (§3.2.7.1). Un ensemble de 88 images de carottes virtuelles de 1600x1600 pixels a été ainsi généré.

7.2.6 Langage et environnement de développement informatique

Nous soulignons que toute la partie développement informatique liée à ce chapitre a été réalisée en langage java sous l'environnement de programmation fourni par le logiciel ImageJ. Ce logiciel gratuit offre une interface de programmation très riche. Il englobe une bibliothèque de classes java couvrant la majorité des procédures usuelles de traitement d'images. Le recours à cet environnement nous a permis de nous concentrer sur les méthodes de classification couleur présentées dans ce chapitre, et de bénéficier des fonctions élémentaires liées au traitement d'images. Ces méthodes de classification ont ainsi été programmées sous forme de plugins exécutables sous ImageJ.

7.3 Modélisation du système colorimétrique DAB-Hématoxyline : Démonstration

7.3.1 Description des noyaux cellulaires à classifier

Avant de présenter le schéma de classification adopté, nous proposons de commencer par décrire les caractéristiques couleurs du DAB et de l'Hématoxyline puisque c'est sur la couleur que repose la discrimination entre les noyaux positifs et négatifs au marqueur Ki67. En se référant à la Figure 7.2 qui contient un échantillon de noyaux cellulaires marqués pour la détection de l'expression de la protéine Ki67, on remarque que les couleurs DAB et Hématoxyline ne sont pas uniformes. Les deux présentent un large spectre de nuances.

La couleur DAB varie du marron très sombre, pour les noyaux très marqués, au marron clair tirant vers le beige, pour les noyaux peu marqués. Les noyaux positifs présentent également des nuances de bordeaux qui contient du bleu, des nuances de violet et des nuances de jaune.

L'Hématoxyline liée aux noyaux négatifs présente toutes les nuances de la couleur bleue, en partant du bleu très pâle qui tire vers le gris, en passant par le bleu turquoise jusqu'au bleu franc et le bleu qui tire vers le violet. Ces observations sont en accord avec les constatations rapportées par Weaver et Au [Weaver et Au 1997] et Schüpp [Schüpp 2000]. La séparation des couleurs DAB et Hématoxyline devient difficile avec la présence de toutes ces nuances. Ces variations de couleurs sont exacerbées par la variation de la densité de coloration DAB et Hématoxyline à travers les lames histologiques. Cette dernière est essentiellement due aux variations biologiques et aux différentes caractéristiques des procédures de fixation et de coloration. A cet effet, s'ajoute le mélange de couleur qui peut se produire entre le DAB et l'Hématoxyline lors de leur application sur une même lame histologique. Si on observe de plus près la Figure 7.2, on constate sur la ligne médiane de noyaux un panel de couleurs qui se situe entre le DAB et l'Hématoxyline. La classification de ces noyaux est difficile même pour l'œil entraîné du pathologiste.



Figure 7.2: Exemples de noyaux ayant différentes nuances de couleur marron et bleue.

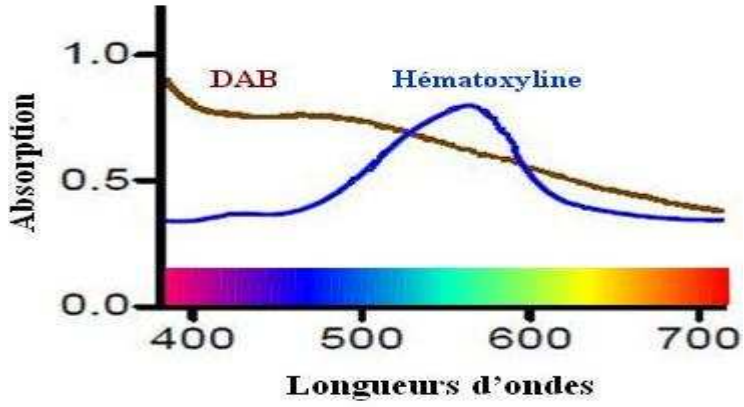


Figure 7.3: Les profils d'absorption associés au DAB et à l'Hématoxyline.

Les couleurs DAB et Hématoxyline se chevauchent beaucoup dans le système couleur RGB. Les spectres d'absorption liés à ces deux couleurs réaffirment nos observations visuelles sur les noyaux (Figure 7.3). En effet, le DAB contient un peu de couleur bleue puisque son spectre d'absorption recouvre partiellement le spectre de l'Hématoxyline au niveau des plages couleur correspondant au bleu.

7.3.2 Protocole de classification couleur

Nous proposons un protocole de classification couleur qui se divise en deux parties (Figure 7.4). En premier lieu, on segmente le matériel biologique, constitué des noyaux cellulaires du fond de l'image. Ensuite, on applique une procédure de classification couleur qui s'appuie sur les attributs couleur du DAB et de la contre-coloration (Hématoxyline), pour la ségrégation des pixels nucléaires positifs au marquage Ki67 des pixels nucléaires négatifs.

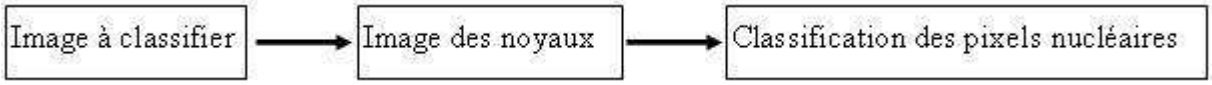


Figure 7.4: Protocole général de classification couleur.

A travers les études présentées dans le chapitre 5 consacré aux systèmes couleur, nous avons vu, qu'aucun espace couleur n'est mieux placé qu'un autre pour réaliser les opérations de traitement d'images (§5.7). Le choix de l'espace couleur dépend du domaine de l'application et des caractéristiques des images à traiter. Il paraît difficile de sélectionner un espace couleur qui soit mieux adapté pour la classification des images couleur en microscopie. Nous avons constaté également que plusieurs auteurs convergent vers la même conclusion : l'axe de luminance est celui qui porte le maximum d'information. Ils notent que pour la plupart des images couleur, l'axe principal issu de l'analyse en composantes principales tend à se superposer avec l'axe de luminance [Trémeau et al 2004] [Ohta et al 1980] [Ranefall et al 1998]. Ils soulignent qu'une grande partie du traitement d'images peut s'appuyer sur cet axe. Au cours des sections suivantes et conformément à ces études, nous allons tenter de combiner l'information achromatique fournie par la luminosité, à l'information chromatique liée aux caractéristiques couleurs du DAB et de l'Hématoxyline, pour réaliser la classification couleur des images histologiques recrutées pour cette étude.

7.3.2.1 Détection des noyaux cellulaires des images : étape de segmentation

En accord avec les études mentionnées ci-dessus, nous avons utilisé l'axe de luminance pour séparer l'ensemble du matériel nucléaire du fond de l'image. Nous avons changé, en premier lieu, le système de représentation des images à classifier de l'espace RGB vers l'espace HLS. Ensuite nous avons utilisé la composante luminance de l'espace HLS pour effectuer une segmentation automatique. En observant de plus près la Figure 7.5, on remarque globalement que les noyaux positifs (DAB) sont détectés alors que les noyaux négatifs (Hématoxyline) possédant une faible coloration bleue ne sont pas correctement segmentés.

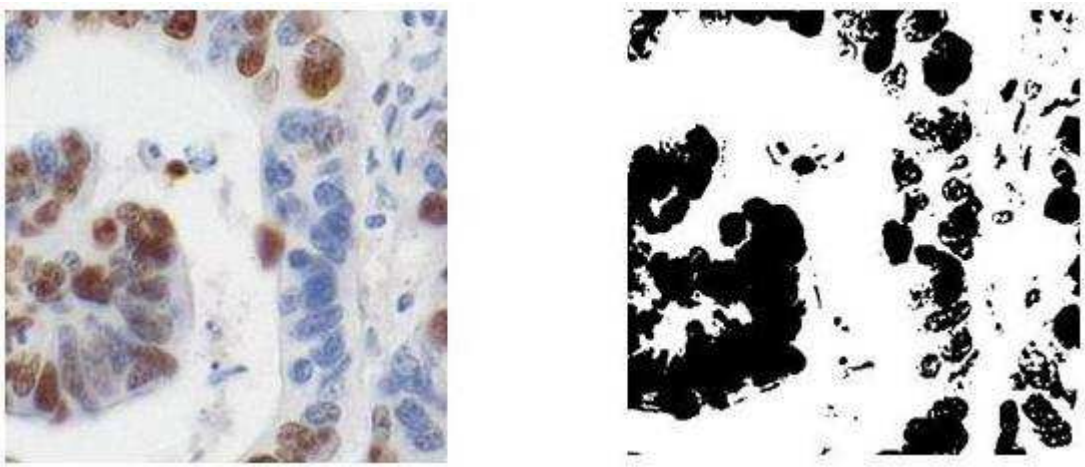


Figure 7.5: Segmentation basée sur l'axe de luminosité.

La segmentation basée sur l'axe de luminance ne permet pas une séparation correcte par rapport au fond de l'image de la totalité de la surface occupée par les noyaux : les pixels nucléaires possédant une faible quantité de couleur se retrouvent rattachés à tort au fond de l'image. Paraddo et al [Paraddo et al 1996] trouvent aussi des résultats similaires. Ils notent que la segmentation basée sur l'axe de luminance permet seulement la segmentation correcte des noyaux fortement marqués. De plus, lors de la phase d'acquisition des images, la lampe du microscope, éclairant les lames histologiques, n'est pas une source de lumière totalement uniforme. Elle génère ainsi une fluctuation du niveau de lumière à travers la surface des images acquises. Cette fluctuation introduit une variation et un bruit supplémentaire au niveau de l'axe de luminance.

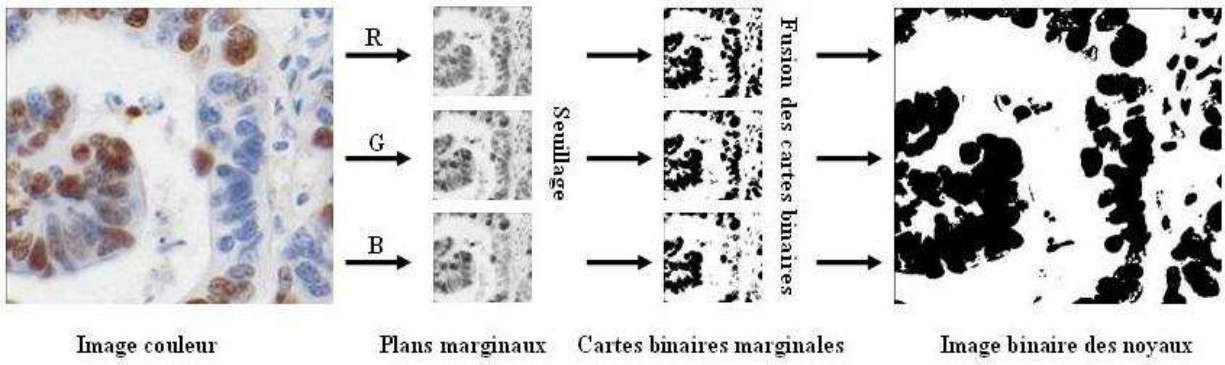


Figure 7.6: Schéma de segmentation pour isoler les noyaux cellulaires du fond de l'image.

Plutôt que d'utiliser la composante luminance, nous adopterons une approche marginale de segmentation en réalisant un seuillage automatique des trois plans couleur R, G et B puis en fusionnant les cartes binaires issues du seuillage des plans marginaux : R, G et B (Figure 7.6). La fusion est basée sur un simple opérateur de fusion logique « opérateur min ». En effet, on compare les trois cartes de segmentations marginales pixel à pixel, l'image binaire résultante prend la valeur du pixel le plus faible. Avec cette approche, on détecte le maximum de matériel nucléaire.

Nous remarquons qu'en comparant la Figure 7.5 et la Figure 7.7, que l'approche de segmentation marginale détecte mieux les noyaux que la simple segmentation sur la base de la luminance. Nous observons, que les noyaux négatifs et peu colorés ont été mal segmentés par la méthode classique. Par contre, ils sont bien pris en compte par notre méthode de segmentation.

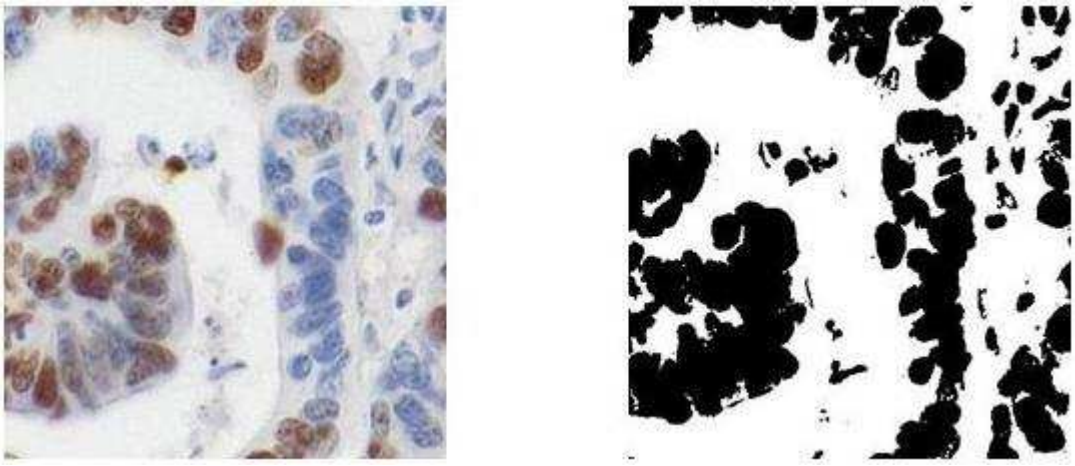


Figure 7.7: Fusion de la segmentation des plans marginaux R, G et B.

7.3.2.2 Séparation couleur entre le DAB et l'Hématoxyline : étape de ségrégation

La couleur est une caractéristique essentielle en microscopie optique pour mettre en évidence les structures biologiques, telles que les noyaux cellulaires. C'est sur la couleur que se basent les pathologistes pour séparer les cellules positives des cellules négatives pour un marqueur immunohistochimique donné [Seigneurin et al 1989] [Tucker et al 1987] [Salminen et al 2005] [Frenebro et al 2002] [Nocito et al 2001]. Notre méthodologie de classification va donc se focaliser sur la couleur pour réaliser la distinction entre les cellules positives et les cellules négatives au marquage ki67.

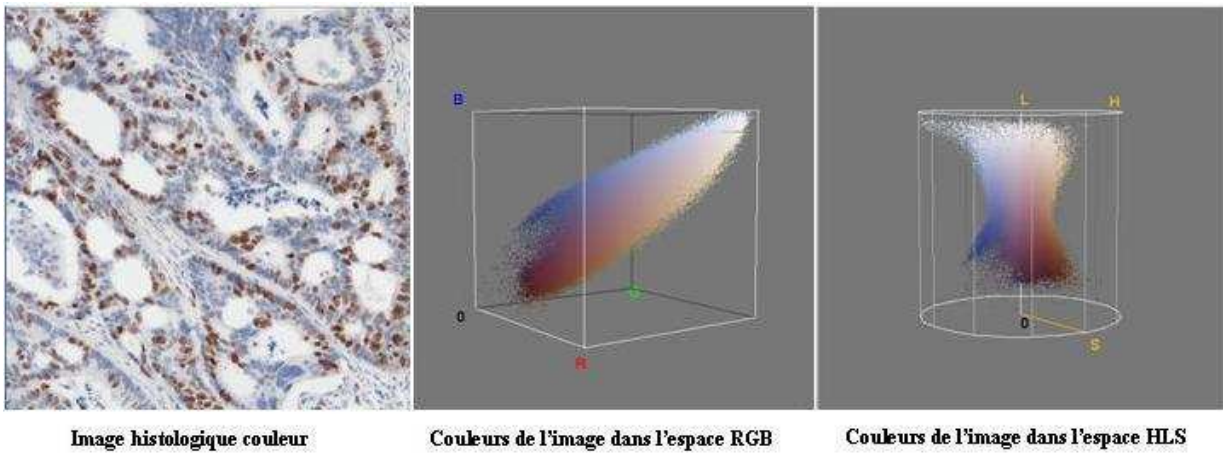


Figure 7.8: Profil des nuages des points couleur associés à une image histologique.

La couleur des images histologiques est codée dans l'espace RGB puisque les stations d'acquisition des images en microscopie optique stockent les lames numériques dans cet espace couleur. L'exploitation en premier lieu de l'espace RGB pour l'analyse, des couleurs rattachées au marquage et à la contre-coloration, semble très logique. La Figure 7.8 illustre la distribution des points couleur associés à une image de marquage Ki67 au niveau du cube couleur RGB. On remarque surtout que le DAB et l'Hématoxyline ne se manifestent pas par deux nuages couleur bien distincts. Les nuages marron du DAB et bleu de l'Hématoxyline se chevauchent et se recouvrent beaucoup dans l'espace RGB. On peut même juger qu'ils ne forment qu'un seul nuage couleur représentant deux domaines diffus de marron et de bleu avec toute une étendue de couleurs intermédiaires entre le marron et le bleu. Ces observations sont en accord avec nos observations visuelles relevées sur les noyaux cellulaires illustrés sur

la Figure 7.2. En s'appuyant sur ces constatations, la séparation couleur du DAB et de l'Hématoxyline est une tâche difficile. De plus, la réalisation d'algorithmes de classification couleur dans un espace 3D requiert des méthodes de calcul très complexes et accroît considérablement le temps alloué au traitement des images [Sarabi et Aggarwal 1989].

Le recours à un autre espace couleur semble nécessaire pour contourner la difficulté de classification lié au phénomène de recouvrement couleur entre le DAB et l'Hématoxyline dans l'espace RGB. Le système HLS « Hue Luminance Saturation » est parmi les systèmes couleur les plus utilisés dans le domaine du traitement des images couleur. La littérature abonde de publications adoptant ce type de système couleur pour accomplir la classification couleur des images biologiques [Garbay et Brugal 1981] [Fermin et Degraw 1995] [Poston 1996] [Weaver et Au 1997] [Johansson et al 2000] [Van der Laak et al 2000] [Elie et al 2003] [Ruifrok et al 2003] [Lehr et al 1997] [Lehr et al 1999].

On constate sur la base de la disposition des couleurs DAB et Hématoxyline (Figure 7.8), que l'espace HLS ne permet pas lui non plus de dégager deux nuages bien différenciés de couleur marron et bleue. Dans la plupart du temps, les travaux de classification opérant dans un espace de type HLS, s'appuient essentiellement sur la composante teinte « Hue ». Mais, même sur cette composante les couleurs marron et bleue se chevauchent. Les profils d'absorption liés au DAB et à l'Hématoxyline montrent bien le recouvrement spectral entre les deux couleurs (Figure 7.3). Ce fait est aussi partiellement dû, à la nature de la couleur du DAB qui contient en partie une fraction de couleur bleue.

Nous avons relevé la difficulté de la tâche de séparation des couleurs marron et bleue liées au DAB et à l'Hématoxyline au niveau des espaces couleur RGB et HLS. Cette difficulté peut être attribuée à plusieurs facteurs :

- la nature physique du DAB et de l'Hématoxyline : ces deux colorants ne délivrent pas deux couleurs complémentaires qui pourraient faciliter leur séparation. A l'inverse, ils se manifestent par des spectres d'absorption qui se recoupent beaucoup,
- le DAB et l'Hématoxyline ne se traduisent pas au niveau des images histologiques par deux points couleur uniques. Ils se manifestent par un large spectre de couleurs marron

et bleue lié aux différentes concentrations et quantités de DAB et d'Hématoxyline fixées sur les noyaux des coupes histologiques,

- l'application consécutive du DAB et de l'Hématoxyline sur la surface d'une même lame histologique, rend possible la formation de mixture entre les deux colorants. Ce phénomène peut expliquer l'existence du large panel de nuances couleur entre le marron du DAB et le bleu de l'Hématoxyline.

Tous ces facteurs doivent être pris en considération lors de la conception de tout protocole de classification capable de séparer correctement les couleurs DAB et Hématoxyline. La compréhension et la maîtrise des mécanismes qui régissent ces facteurs sont sans doute la clé pour proposer et réaliser une stratégie de classification couleur performante. La section suivante sera consacrée à développer des modèles colorimétriques qui tentent d'expliquer les facteurs de variation des couleurs rattachées au DAB et à l'Hématoxyline.

7.3.2.2.1 Analyse colorimétrique du DAB et de l'Hématoxyline

7.3.2.2.1.1 Choix de l'espace couleur

Nous avons porté notre choix sur le diagramme de chromaticité xy lié à l'espace couleur XYZ pour analyser les caractéristiques colorimétriques du DAB et de l'Hématoxyline. Comme nous l'avons déjà signalé au niveau du chapitre consacré à la couleur (§4.5.1.3), l'espace couleur XYZ reste toujours le système couleur de référence pour effectuer les mesures colorimétriques [Vandenbrouk 2000]. Les colorimètres et les spectromètres utilisent généralement ce système pour mesurer une couleur.

Le diagramme de chromaticité xy permet la localisation des couleurs au niveau d'un repère géométrique à deux dimensions. Il est à ce titre la représentation 2D la plus utilisée en colorimétrie [Trémeau et al 2004]. Le diagramme xy présente l'avantage de dissocier l'information chromatique de l'information achromatique. Il se focalise essentiellement, sur la chromaticité pour représenter une couleur donnée, abstraction faite de sa luminosité. Le diagramme de chromaticité xy permet la visualisation de toutes les couleurs du visible (Figure 7.9.A). Ces couleurs sont délimitées par une courbe appelée le lieu spectral ou « Spectrum

locus », et la ligne des pourpres. Cette courbe définit toutes les couleurs monochromatiques dites couleurs pures. Toute couleur à l'intérieur de la surface délimitée par le lieu spectral peut être représentée comme le résultat d'un mélange de couleurs, entre la couleur blanche et une couleur pure appartenant au lieu spectral. Les couleurs qui peuvent être générées par la synthèse additive liée au système RGB, qui est le système utilisé pour l'acquisition et le stockage des images histologiques traitées au niveau de se travail, se situent dans un triangle délimité par les points couleurs R, G et B (Figure 7.9.B). Ce triangle de couleurs est appelé le triangle de Maxwell.

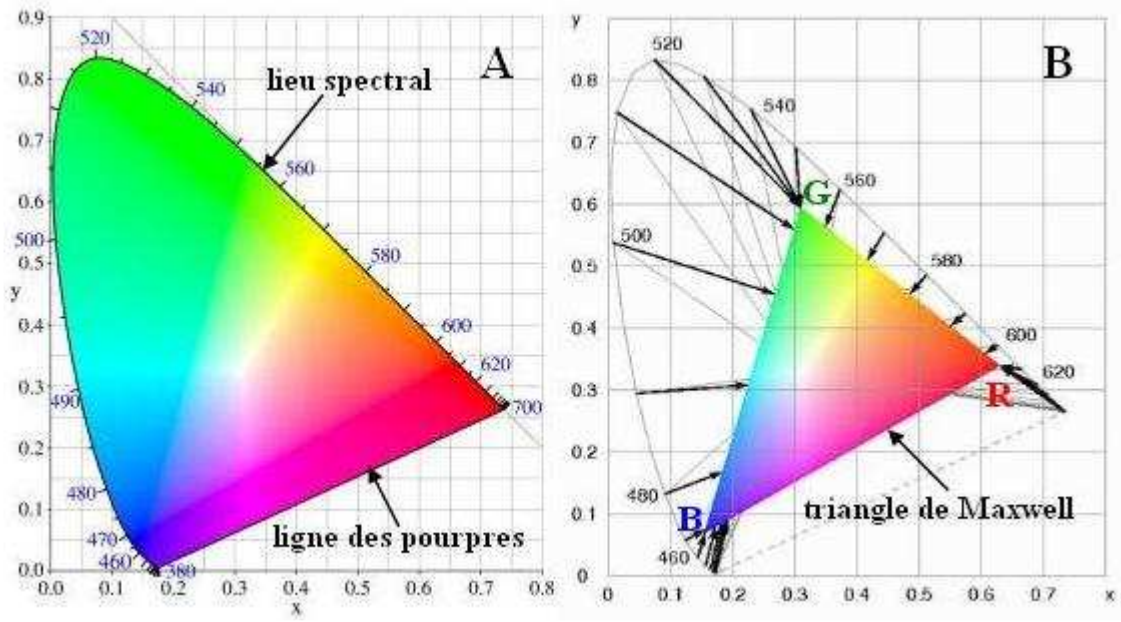


Figure 7.9: Le diagramme de chromaticité xy. A : les couleurs du visible sont délimitées par le lieu spectral et la ligne des pourpres**. B :** le triangle de Maxwell délimite les couleurs reproduites par synthèse additive**.**

Le diagramme de chromaticité est très utile, car il permet de décrire aisément les relations entre les différentes couleurs. Il permet surtout la modélisation des opérations de mélange des couleurs par de simples constructions géométriques. Ce formalisme peu compliqué, décrit par de simples relations géométriques, satisfait les lois de Grassman et Abney qui régissent les règles élémentaires de la synthèse additive des couleurs [Grassman 1853] [Abney 1913].

Par exemple, sur la Figure 7.10, on peut obtenir la couleur pure dominante de la couleur A notée D, en prolongeant la ligne reliant la couleur blanche B et la couleur A vers le lieu

spectral. La couleur Ā complémentaire de la couleur A, est obtenue par simple symétrie par rapport à la couleur blanche. La couleur M résulte du mélange additif entre les couleurs A et C. Toutes les nuances de ce mélange se trouvent sur la ligne reliant les deux couleurs A et C.

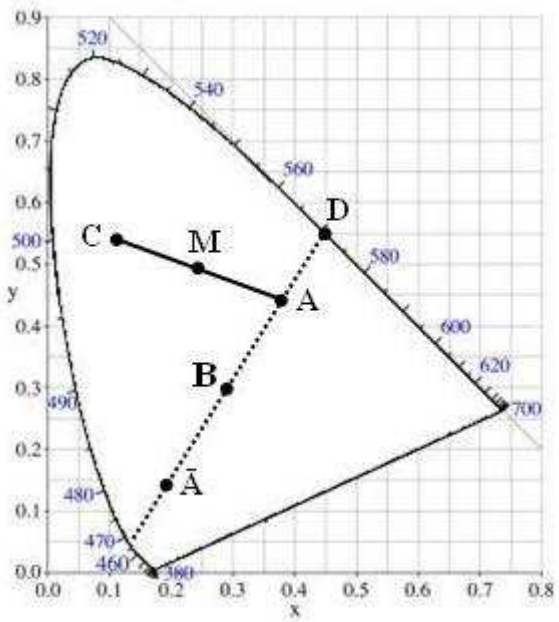


Figure 7.10: Formulation des couleurs au niveau du diagramme de chromaticité xy. La couleur D est la couleur monochromatique dominante de la couleur A. La couleur Ā est la couleur complémentaire de la couleur A. La couleur M résulte du mélange additif entre les couleurs A et C.

7.3.2.2.1.2 Le profil couleur du DAB et de l'Hématoxyline

Notre idée de base est d'utiliser les caractéristiques chromatiques de la couleur marron du DAB associée au marquage ki67, et de la couleur bleue de l'Hématoxyline, associée à la contre-coloration, pour effectuer la discrimination entre ces deux couleurs.

Nous allons en premier lieu, étudier le profil des deux couleurs DAB et Hématoxyline au niveau du diagramme de chromaticité xy. La projection du matériel nucléaire d'une image histologique comportant à la fois, le DAB et l'Hématoxyline sur le diagramme xy (figure 7.11) montre un large nuage de couleurs qui s'allonge du marron vers le bleu, en passant par toute une gamme de couleurs intermédiaires entre ces deux couleurs.

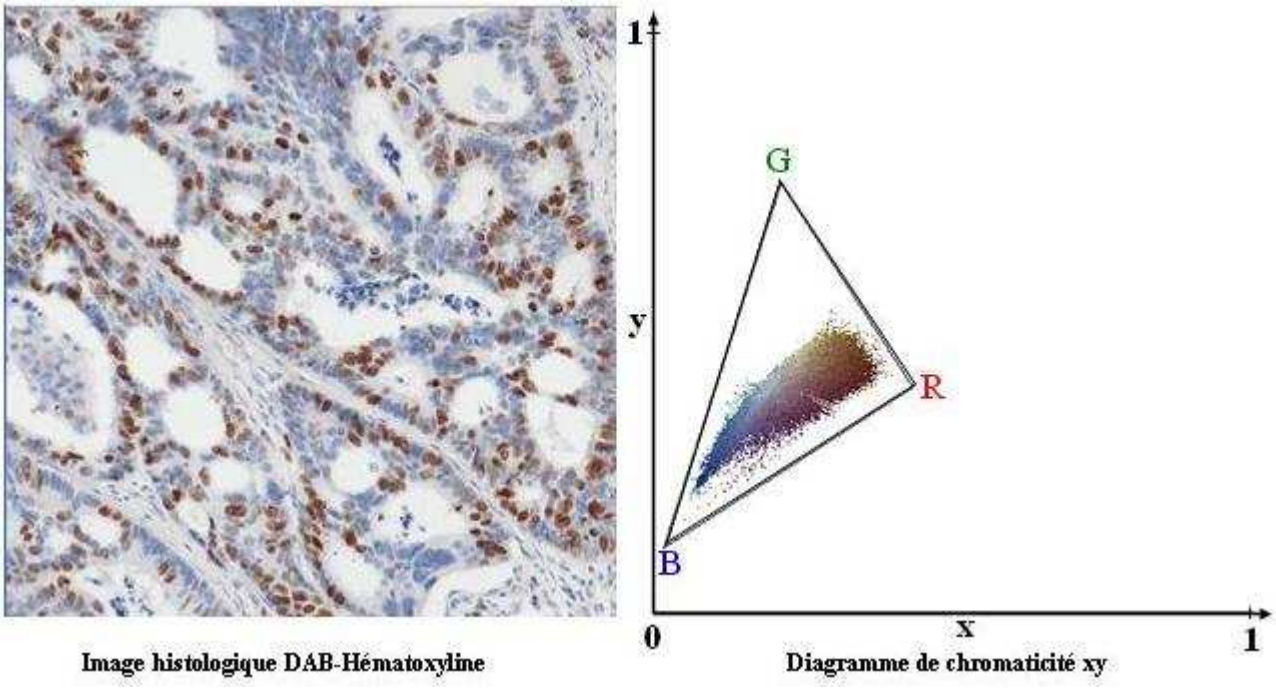


Figure 7.11: Nuage couleur d'une image histologique comportant le DAB et l'Hématoxyline.

La Figure 7.11 illustre bien la difficulté de séparer le DAB de l'Hématoxyline sur la base du diagramme xy. A l'instar des espaces couleur RGB et HLS, le diagramme de chromaticité xy ne permet pas de dégager deux domaines distincts liés au DAB et à l'Hématoxyline. Nous assistons à un recouvrement entre les domaines couleur marron et bleu associées au DAB et à l'Hématoxyline. Ce recouvrement est peut être dû aux nuances couleurs résultantes du phénomène de mélange couleurs qui peut se produire lors des phases d'application du DAB et de l'Hématoxyline sur une même surface biologique.

Afin de restreindre l'analyse colorimétrique au DAB et à l'Hématoxyline, tout en évitant l'analyse des nuances couleurs issues du mélange couleur entre le DAB et l'Hématoxyline, nous avons fait appel aux images de références positives et négatives. Une image référence comporte un seul colorant. Ainsi, une image de référence positive est une image histologique comportant exclusivement la coloration DAB. La phase de contre-coloration a été supprimée lors de la préparation de la coupe. A l'inverse, une image référence négative comporte uniquement l'Hématoxyline appliquée pendant la phase de contre-coloration. La chaîne de révélation du marquage Ki67 est altérée, par la suppression de l'anticorps primaire anti-Ki67.

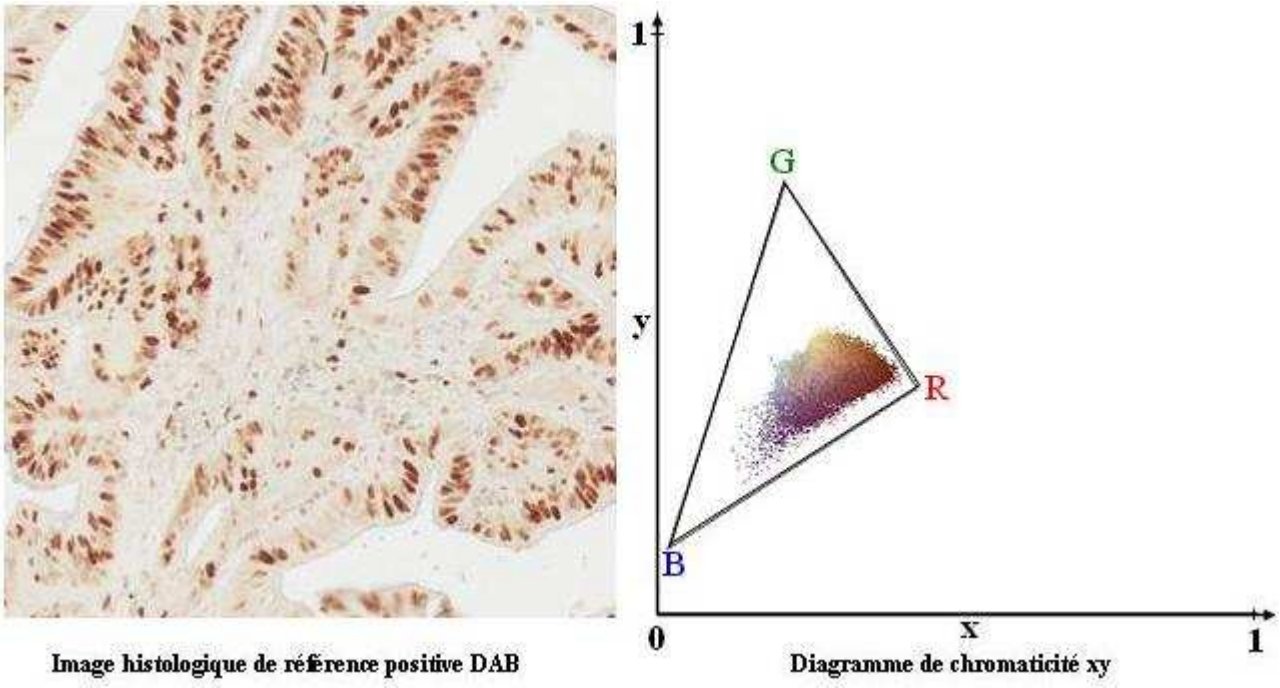


Figure 7.12: Nuage couleur associé à une image de référence positive DAB.

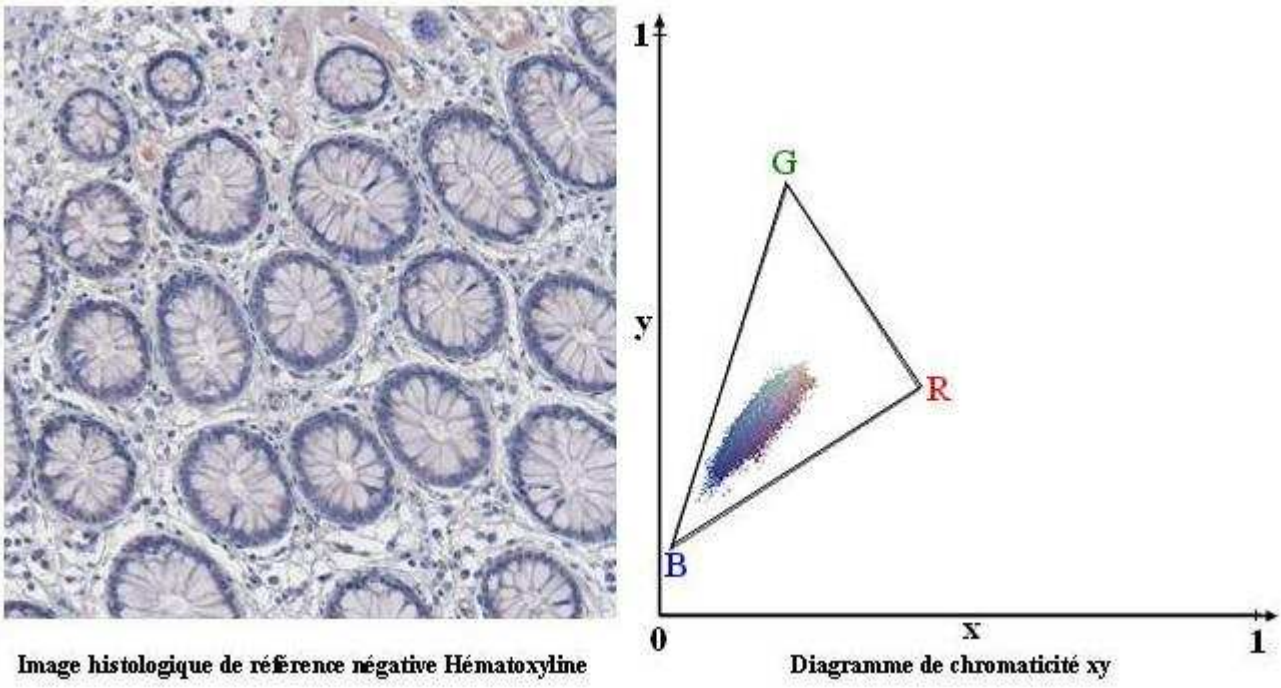


Figure 7.13: Nuage couleur associé à une image de référence négative Hématoxyline.

La Figure 7.12 montre le profil couleur de l'image référence positive. On note que le DAB se manifeste par un large nuage de couleurs. Bien que le marron soit la couleur principale du DAB, le nuage couleur montre l'existence d'autres nuances couleurs telles que le brun orangé, le mauve et le violet qui tire vers le gris.

Le profil du nuage couleur Hématoxyline, issu de l'image référence négative, ne se limite pas à présenter uniquement une seule nuance de bleu (Figure 7.13:). En plus du bleu, le nuage couleur associé à l'Hématoxyline comporte du violet et du mauve.

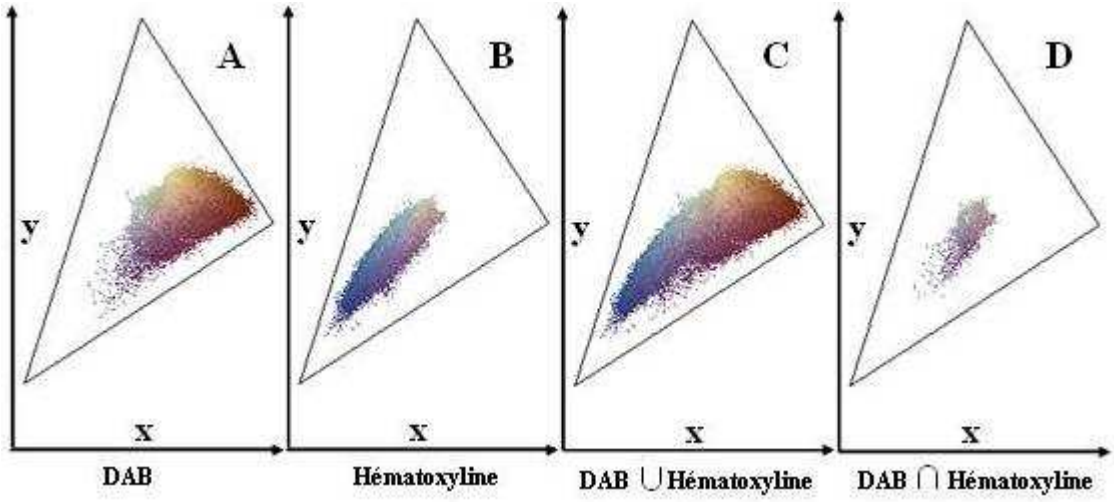


Figure 7.14: Mise en relation des deux nuages couleur DAB et Hématoxyline. A : Nuage couleur associé au DAB**. B :** Nuage couleur associé à l'Hématoxyline**. C :** Union des deux nuages couleur DAB et Hématoxyline**. D :** Intersection des deux nuages couleur DAB et Hématoxyline**.**

Nous remarquons, sur la base des observations visuelles, que les nuages DAB et Hématoxyline partagent certaines nuances couleur comme le mauve et le violet. L'affichage des deux nuages couleur DAB et Hématoxyline sur un même diagramme de chromaticité xy (Figure 7.14.C), illustre parfaitement le recouvrement produit entre les deux nuages couleur au niveau du mauve et du violet (Figure 7.14.D). Le recouvrement chromatique entre le DAB et l'Hématoxyline (Figure 7.14.D), se traduit par une surface commune mesurant 15,967% de la surface globale générée par les nuages couleur (nuage couleur DAB additionné au nuage couleur Hématoxyline) (Figure 7.14.C). La surface commune occupe 21,815% de la surface du nuage couleur DAB, et 37,330% de la surface occupée par le nuage couleur associé à l'Hématoxyline. Ces chiffres confirment le profil d'absorption spectrale associé au DAB et à

l’Hématoxyline, illustré dans les sections précédentes (Figure 7.3).

L’usage des nuages couleur associés aux images références montre clairement que le recouvrement couleur qui se produit entre le DAB et l’Hématoxyline est dû principalement à la nature même de ces deux colorants. La quantité des colorants fixés n’est pas constante à travers la surface des coupes histologiques. En microscopie, et sous les conditions d’éclairage en lumière transmise, deux quantités différentes d’un même colorant se traduisent généralement par deux nuances de couleurs différentes. Le recouvrement entre le DAB et l’Hématoxyline peut être dû, secondairement à l’action des différents agents chimiques utilisés lors des phases de préparation des coupes histologiques. Par l’intermédiaire de leurs caractéristiques telles que leur PH, ces agents peuvent modifier le comportement colorimétrique du DAB et de l’Hématoxyline. Ces effets éventuels doivent être étudiés de plus près.

En résumé, le DAB et l’Hématoxyline ne se matérialisent pas par deux couleurs bien localisées au niveau du diagramme de chromaticité xy. Il coexiste tout un panel de nuances couleurs rattachées à chacun deux. La variation colorimétrique inhérente à l’application du DAB et de l’Hématoxyline au niveau des coupes histologiques, doit être prise en compte dans les procédures de classification couleur. Dans la section suivante, nous proposerons des modèles couleur qui expliquent à la fois, la variation couleur du DAB et de l’Hématoxyline, et le mélange couleur qui peut se produire entre ces deux colorants (DAB et Hématoxyline).

7.3.2.2.2 Modèles couleur associées au DAB et à l’Hématoxyline

Les modèles couleur que nous proposons pour expliquer les variations colorimétriques associées au DAB et à l’Hématoxyline, se basent sur le concept des constructions géométriques lié au diagramme de chromaticité xy. Le formalisme des constructions géométriques est très utile pour situer les couleurs les unes par rapport aux autres. Il est très simple à réaliser, intuitif et puissant à la fois. Il permet notamment la modélisation des quantités de DAB et d’Hématoxyline déposées et fixées sur les coupes histologiques. Les relations géométriques permettent également la modélisation des nuances couleur générées par le mélange couleur entre le DAB et l’Hématoxyline.

Nous nous sommes servis des images références positives et négatives pour générer les couleurs moyennes DAB et Hématoxyline. En effet, pour chaque image de référence, nous avons calculé les coordonnées RGB de la couleur moyenne du matériel nucléaire. Les couleurs moyennes DAB et Hématoxyline sont ensuite projetées sur le diagramme de chromaticité xy (Figure 7.15.A). Les coordonnées chromatiques associées au DAB et à l'Hématoxyline, constituent les centres de gravité liés aux deux nuages couleur de références (DAB et Hématoxyline) illustrés dans la Figure 7.12 et la Figure 7.13.

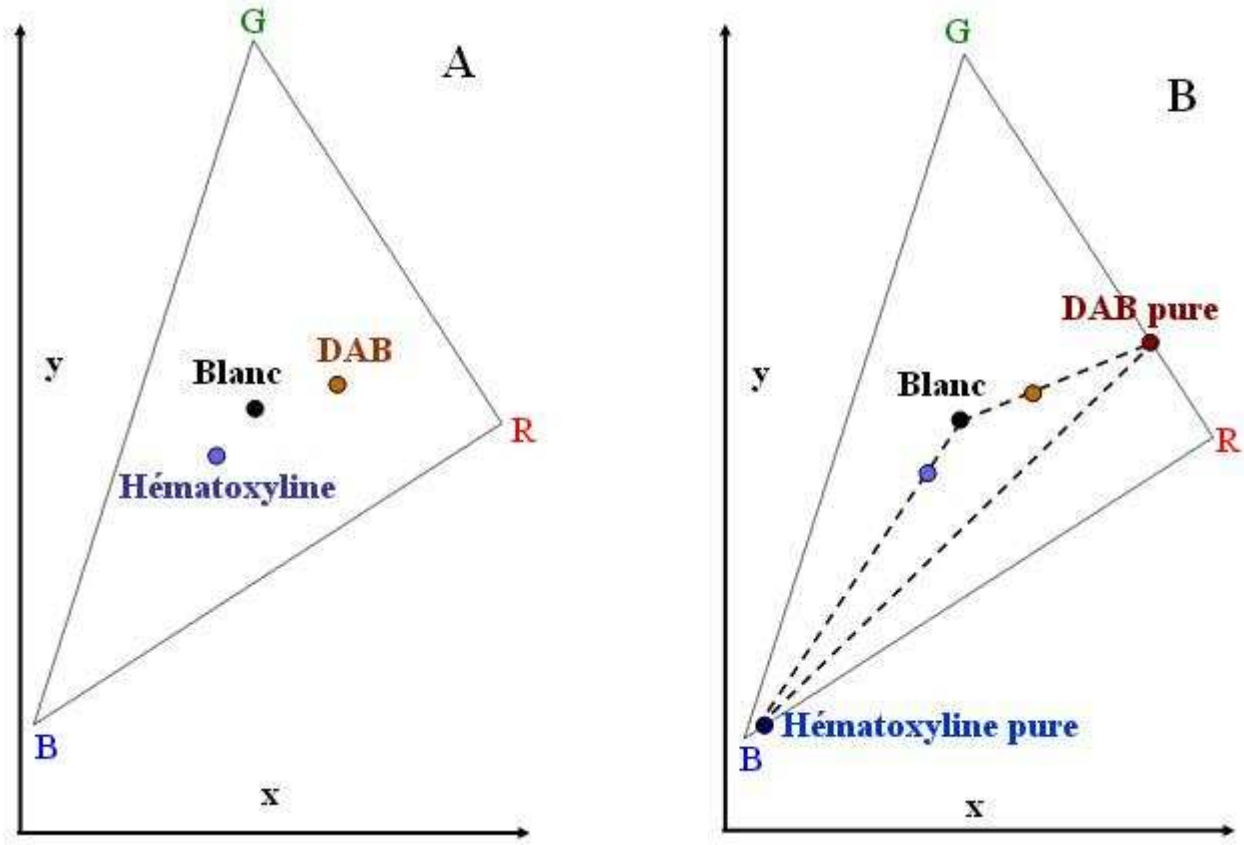


Figure 7.15: Analyse colorimétrique liée au DAB et à l'Hématoxyline. A : Les deux points chromatiques DAB et Hématoxyline**. B :** Détection des deux points couleur DAB pure et Hématoxyline pure**.**

La puissance du diagramme de chromaticité xy réside dans le fait que les modèles couleur associés au DAB et à l'Hématoxyline sont construits uniquement à partir de trois points couleur : le DAB, l'Hématoxyline et la couleur blanche. En accord avec les principes de colorimétrie, toute couleur est un mélange additif d'une couleur monochromatique et de la couleur blanche. Toutes les nuances de ce mélange couleur peuvent être modélisées par une simple ligne reliant les deux couleurs : blanche et monochromatique. Selon ce principe, les couleurs DAB et Hématoxyline résultent de deux mélanges couleur entre le blanc et deux

couleurs monochromatiques pures DAB et Hématoxyline. Ces deux couleurs pures s'obtiennent facilement en traçant deux lignes qui partent du blanc, et passent par le point DAB ou Hématoxyline, puis se prolongent jusqu'aux limites du triangle des couleurs de Maxwell. De cette façon, on arrive à situer les couleurs DAB et Hématoxyline pures sur les limites du triangle de Maxwell (Figure 7.15.B). Ce sont les couleurs les plus pures qui puissent être générées par synthèse additive liée au système couleur RGB. Les couleurs baptisées DAB pure et Hématoxyline pure (Figure 7.15.B) sont les deux couleurs dominantes associées au DAB et à l'Hématoxyline.



Figure 7.16: Nuances couleur associées au modèle couleur DAB.

La ligne joignant la couleur DAB pure et la couleur blanche est reportée sur la Figure 7.16. Elle modélise les nuances de la couleur marron liée au DAB, elle illustre également les quantités de DAB pure associées aux nuances de la couleur marron. La quantité de DAB décroit en se déplaçant vers la couleur blanche.



Figure 7.17: Nuances couleur associées au modèle couleur Hématoxyline.

De même que pour le DAB, la Figure 7.17 illustre la ligne reliant la couleur blanche à la couleur Hématoxyline pure. Cette ligne expose toutes les nuances de la couleur bleue associées à l'Hématoxyline. La quantité de couleur bleue est maximale au niveau de l'Hématoxyline pure, puis diminue au fur et à mesure qu'on se rapproche de la couleur blanche.

En microscopie optique, l'image de la coupe histologique captée par le dispositif de numérisation (caméra) est formée à partir de la lumière émise par la source d'éclairage du microscope puis transmise à travers la lame histologique. L'intensité de la lumière qui arrive jusqu'aux capteurs de la caméra d'acquisition, est fonction de la quantité du matériel biologique déposé sur lame. Ainsi, l'intensité est maximale et la lumière est proche du blanc lorsque celle-ci passe à travers peu de matériel biologique ou à travers un champ vide de la

lame histologique. A l'inverse, le signal lumineux est modifié s'il passe à travers une surface contenant une grande quantité de matériel biologique. Par exemple, les noyaux absorbent une partie du signal lumineux, réduisant ainsi l'intensité de la lumière transmise à travers la coupe histologique. En conséquence, l'intensité de la lumière transmise, est une mesure pertinente de la quantité de matériel biologique.

Les deux lignes associées aux modèles couleur DAB et Hématoxyline (Figure 7.16 et Figure 7.17) sur le diagramme de chromaticité xy, reflètent les différentes quantités de DAB et d'Hématoxyline, soient, les quantités du matériel biologique fixé par ces colorants. La quantité de couleur ou de matériel biologique est maximale au niveau des couleurs DAB pure et Hématoxyline pure, laissant ainsi passer peu de lumière à travers la lame histologique. Ensuite, la quantité de DAB et d'Hématoxyline diminue progressivement en s'éloignant des couleurs pures, en transmettant de plus en plus de lumière à travers la lame histologique. La quantité de DAB et d'Hématoxyline est nulle au niveau du point couleur blanc, là où la transmission de lumière est maximale.

La mise en relation des niveaux de luminosité (transmittance) avec les coordonnées xy des deux lignes couleurs DAB et Hématoxyline est très intéressante, car elle nous permet l'accès aux valeurs RGB liées aux modèles couleur DAB et Hématoxyline. Nous pouvons étudier la possibilité des plans R, G et B à séparer correctement les couleurs DAB et Hématoxyline. La figure 7.18 montre la mise en relation des valeurs RGB des modèles DAB et Hématoxyline, sur les plans R, G et B avec les niveaux de transmittance de lumière des deux lignes couleurs DAB et Hématoxyline. La Figure 7.18 montre clairement, qu'aucun plan couleur n'est approprié pour séparer les deux couleurs DAB et Hématoxyline. Aucun seuil basé sur les niveaux de gris R, G et B ne permet une discrimination entre le DAB et l'Hématoxyline.

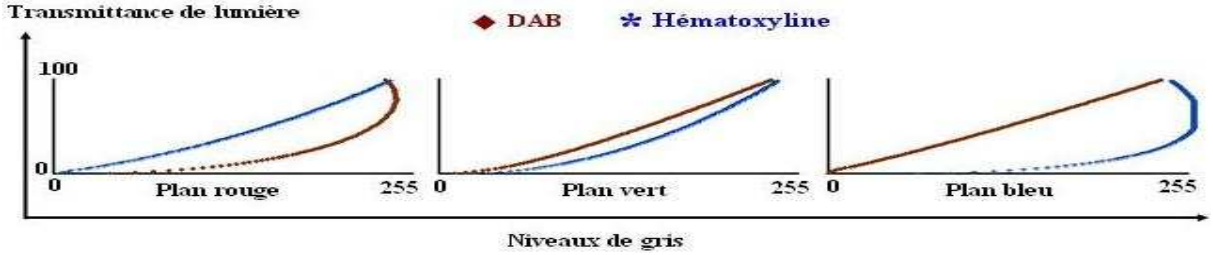


Figure 7.18: Projection des modèles couleur sur les plans R, G et B. Mise en relation des niveaux de gris R, G et B et des niveaux de transmittance de lumière associés aux modèles couleur DAB et Hématoxyline.

La Figure 7.19 représente la ligne reliant les deux couleurs DAB pure et Hématoxyline pure. Cette ligne illustre parfaitement les nuances couleurs intermédiaires entre la couleur marron et la couleur bleue associées au DAB et à l'Hématoxyline. La ligne DAB pure-Hématoxyline pure (Figure 7.15.B et Figure 7.19) est un modèle du mélange additif entre le DAB et l'Hématoxyline. C'est aussi un modèle des nuances couleur communes entre le DAB et l'Hématoxyline. Nous retrouvons également au niveau de cette ligne, ou modèle couleur, les nuances couleur mauve et violet présents sur la surface de recouvrement entre les nuages couleur DAB et Hématoxyline exposés auparavant (Figure 7.14.D).



Figure 7.19: Nuances couleur du modèle du mélange couleur entre le DAB et l'Hématoxyline.

7.3.2.2.3 Construction de la table de classification couleur associée au DAB et à l'Hématoxyline

Les modèles couleur associés au DAB et à l'Hématoxyline nous ont permis l'analyse et la prise en compte de la variation colorimétrique liée à l'application des colorants au niveau des coupes histologiques. Nous avons pu démontrer, selon ces deux modèles couleur, l'insuffisance du seuillage des composantes monochromatiques RGB à séparer de manière satisfaisante le DAB et l'Hématoxyline.

Nous allons développer au niveau de cette section, une approche de classification couleur qui réalise la séparation chromatique entre deux couleurs différentes. Cette approche sera illustrée à travers l'exemple des deux couleurs marron et bleue associées au DAB et à l'Hématoxyline. Notre méthode exploite essentiellement les attributs des modèles couleur DAB et Hématoxyline développés au cours de la section précédente, pour construire une table couleur 2D capable de classifier le matériel nucléaire des lames histologiques, comme appartenant au DAB ou à l'Hématoxyline.

Grâce au formalisme des constructions géométriques lié au diagramme de chromaticité xy, nous avons pu définir les deux couleurs pures et dominantes associées au DAB et à l'Hématoxyline. L'idée de base est de se servir de ces deux points couleur, comme repères ou

points d'ancrage initiaux pour construire deux polygones qui divisent le triangle des couleurs (triangle de Maxwell) en deux surfaces, une comportant les couleurs proches du DAB, et l'autre pour les couleurs proches de l'Hématoxyline.

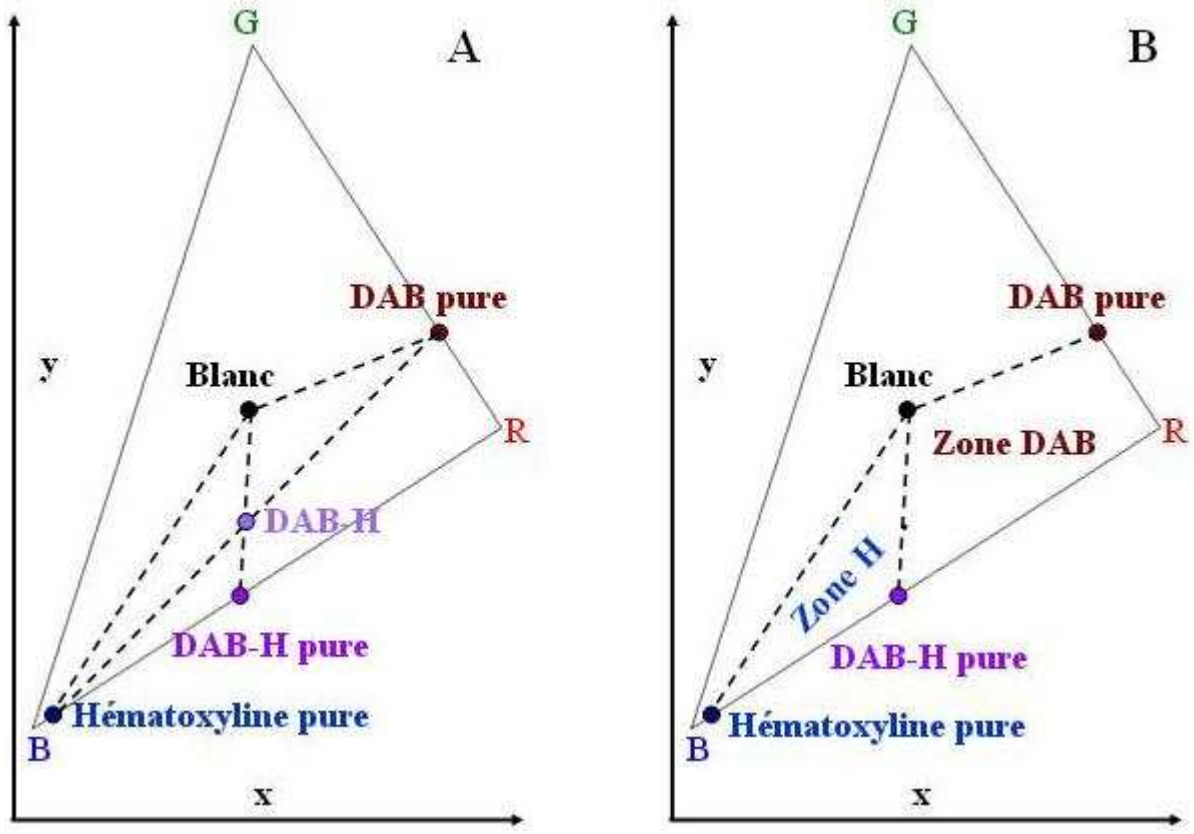


Figure 7.20: Génération des points couleur associés au mélange DAB-Hématoxyline. A : Calcul des points couleur DAB-H et DAB-H pure. **B :** Ligne couleur modélisant la frontière entre le DAB et l'Hématoxyline**.**

Le schéma expliquant les étapes de construction des polygones couleur est illustré sur la Figure 7.20 et la Figure 7.21. A partir des couleurs DAB pure et Hématoxyline pure, nous pouvons tracer une ligne modélisant les mélanges couleur issue de ce couple de couleurs. La couleur DAB-H représente le mélange paritaire entre le DAB pure et l'Hématoxyline pure. Elle se situe à mi-distance entre ces deux couleurs pures (Figure 7.20.A). De part et d'autre de la couleur DAB-H, le mélange couleur contient une partie plus importante de DAB pure ou d'Hématoxyline pure. La couleur notée DAB-H pure est la couleur dominante de la couleur DAB-H, elle s'obtient en prolongeant la ligne qui relie le blanc et la couleur DAB-H, jusqu'aux limites du triangle des couleurs (Figure 7.20.A). La ligne entre les deux points couleur blanche et DAB-H pure démarque la frontière entre deux zones couleurs :

- une zone DAB délimitée par les couleurs : blanche, DAB pure, rouge et DAB-H pure. Cette zone comporte les couleurs plus proches du DAB par rapport à l'Hématoxyline,
- une zone Hématoxyline notée Zone H définie par le triangle formé par les couleurs : blanche, DAB-H pure et Hématoxyline pure. Ce triangle englobe les couleurs proches de l'Hématoxyline par rapport au DAB.

Nous remarquons sur la Figure 7.20.B que les zones DAB et Hématoxyline ne couvrent pas la totalité du triangle des couleurs. Dans le but d'étendre ces deux zones couleur, pour qu'elles occupent toute la surface du triangle des couleurs, nous avons fait appel aux couleurs complémentaires associées au DAB pure et à l'Hématoxyline pure (Figure 7.21.A).

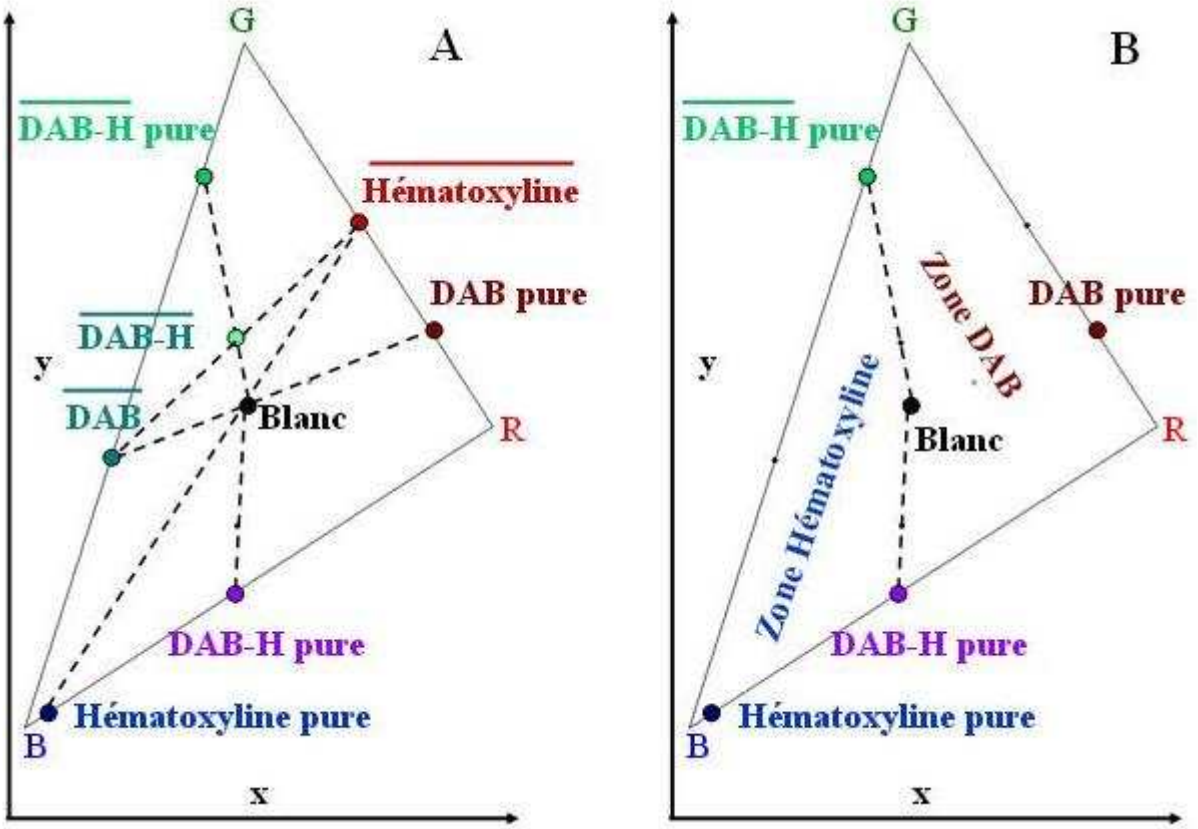


Figure 7.21: Construction de la table de classification couleur. A : Génération des points repères couleur**. B :** Dessin de la ligne séparant les zones DAB et Hématoxyline**.**

La couleur complémentaire du DAB pure, notée $\overline{DAB}$, est chromatiquement plus proche de la couleur Hématoxyline pure par rapport à la couleur DAB pure. Egalement, la couleur

complémentaire rattachée à l'Hématoxyline pure notée $\overline{Hématoxyline}$ est plus proche du DAB pure, par rapport à la couleur Hématoxyline (Figure 7.21.A). Les deux couleurs complémentaires $\overline{DAB}$ et $\overline{Hématoxyline}$ définissent elles aussi, une ligne modélisant leur mélange couleur. De la même façon que le DAB pure et l'Hématoxyline pure, la couleur notée $\overline{DAB-H}$ représente le mélange égalitaire entre les couleurs $\overline{DAB}$ et $\overline{Hématoxyline}$. La couleur dominante pure liée à la couleur $\overline{DAB-H}$ est également détectée sur les frontières du triangle des couleurs. Elle est notée $\overline{DAB-H}$ pure (Figure 7.21.A). La ligne reliant la couleur blanche et la couleur $\overline{DAB-H}$ pure, délimite la frontière entre les couleurs proches du $\overline{DAB}$, en conséquence les couleurs plus proche de l'Hématoxyline pure, de celles proches de la couleur $\overline{Hématoxyline}$, c'est-à-dire proche de la couleur DAB (Figure 7.21.B).

En finalité, les trois points couleur : blanc, DAB-H pure et $\overline{DAB-H}$ pure, divisent le triangle des couleurs en deux régions (Figure 7.21.B) :

- une zone DAB délimitée par le polygone formé à partir des points couleur : DAB pure, R, DAB-H pure, blanc, $\overline{DAB-H}$ pure et G, cette zone comporte toutes les couleurs proches du DAB par rapport à l'Hématoxyline,
- une zone Hématoxyline délimitée par le polygone construit à partir des points couleur : Hématoxyline pure, DAB-H pure, blanc, $\overline{DAB-H}$ pure et B. La zone Hématoxyline englobe toutes les couleurs proches de l'Hématoxyline par rapport au DAB.

Le triangle des couleurs de Maxwell divisé en deux zones d'influence associées aux couleurs DAB et Hématoxyline constitue en réalité une table de classification qui permet d'affilier les différentes couleurs au DAB ou à l'Hématoxyline. Nous allons illustrer au cours de la section suivante, l'utilisation de cette table de classification dans un schéma global de classification couleur des lames histologiques.

7.3.2.2.4 Mise en application de la table de classification couleur

La Figure 7.22 illustre la table de classification qui différencie entre les couleurs marron et bleue associées au DAB et à l'Hématoxyline. Les deux surfaces DAB et Hématoxyline sont codées selon deux couleurs :

- la couleur rouge, est affectée à tous les points couleur appartenant à la zone DAB,
- la couleur bleue est attribuée aux points couleur de la zone Hématoxyline.

En réalité, la table de classification partitionne l'espace couleur en deux sous espaces, l'un rattaché au DAB et l'autre à l'Hématoxyline. Cette table est essentiellement utilisée pour déterminer si une couleur donnée est affiliée au DAB ou à l'Hématoxyline.

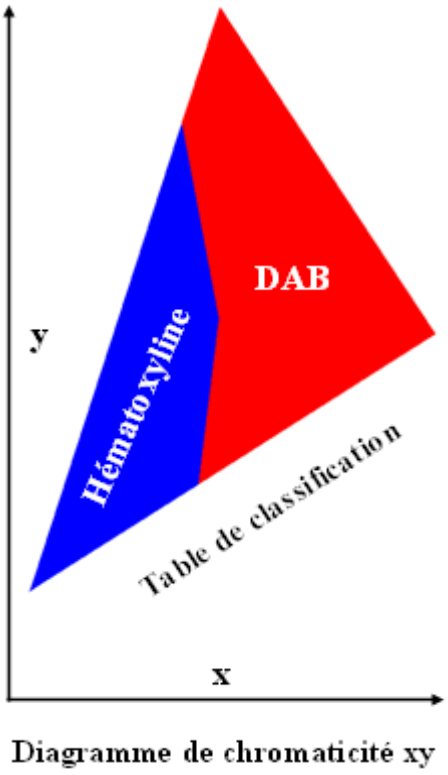


Figure 7.22: Table de classification couleur associée au DAB et à l'Hématoxyline.

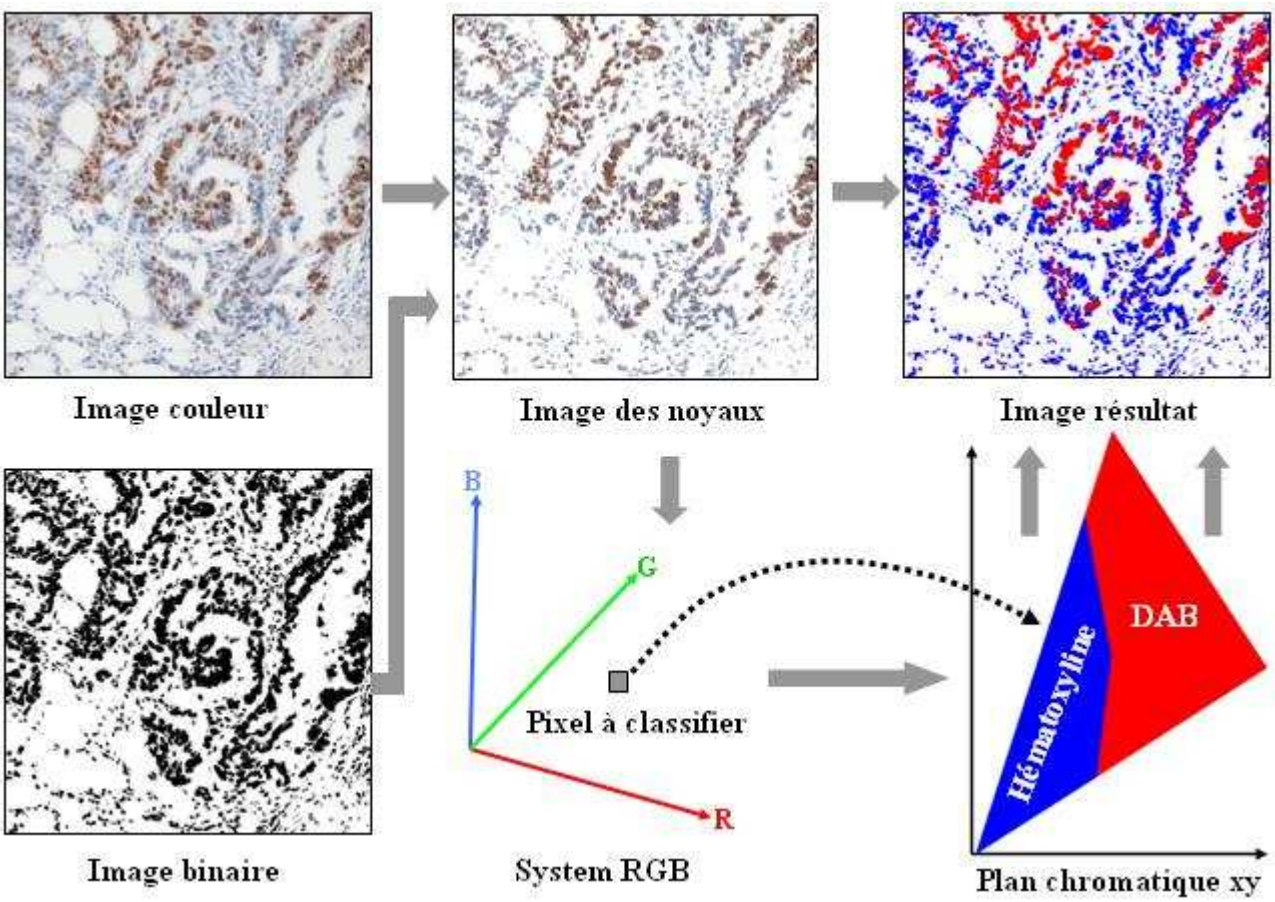


Figure 7.23: Protocole d'utilisation de la table de classification couleur DAB-Hématoxyline.

La Figure 7.23 démontre l'utilisation de la table de classification dans le traitement d'une image histologique. La première étape consiste à séparer le matériel nucléaire du fond de l'image histologique à classifier conformément au protocole de segmentation présenté auparavant (Figure 7.6). Les coordonnées chromatiques associées au diagramme de chromaticité xy des pixels appartenant aux noyaux sont calculées. Projetées dans la table de classification des couleurs DAB-Hématoxyline (Figure 7.22), ces coordonnées permettent d'affecter chaque pixel à l'une des deux composantes : DAB ou Hématoxyline. Enfin, une image résultante fausses couleurs sera construite avec des pixels rouges lorsqu'ils sont affiliés à la zone DAB et des pixels bleus lorsqu'ils sont affiliés à la zone Hématoxyline (Figure 7.23).

7.3.3 Protocole global de la classification couleur des images histologiques

La Figure 7.24 illustre les différentes phases nécessaires pour réaliser la classification couleur des images histologiques. La classification comprend globalement trois phases : segmentation, apprentissage et classification. Nous détaillerons dans ce qui suit les trois étapes et nous présenterons leurs algorithmes associés.

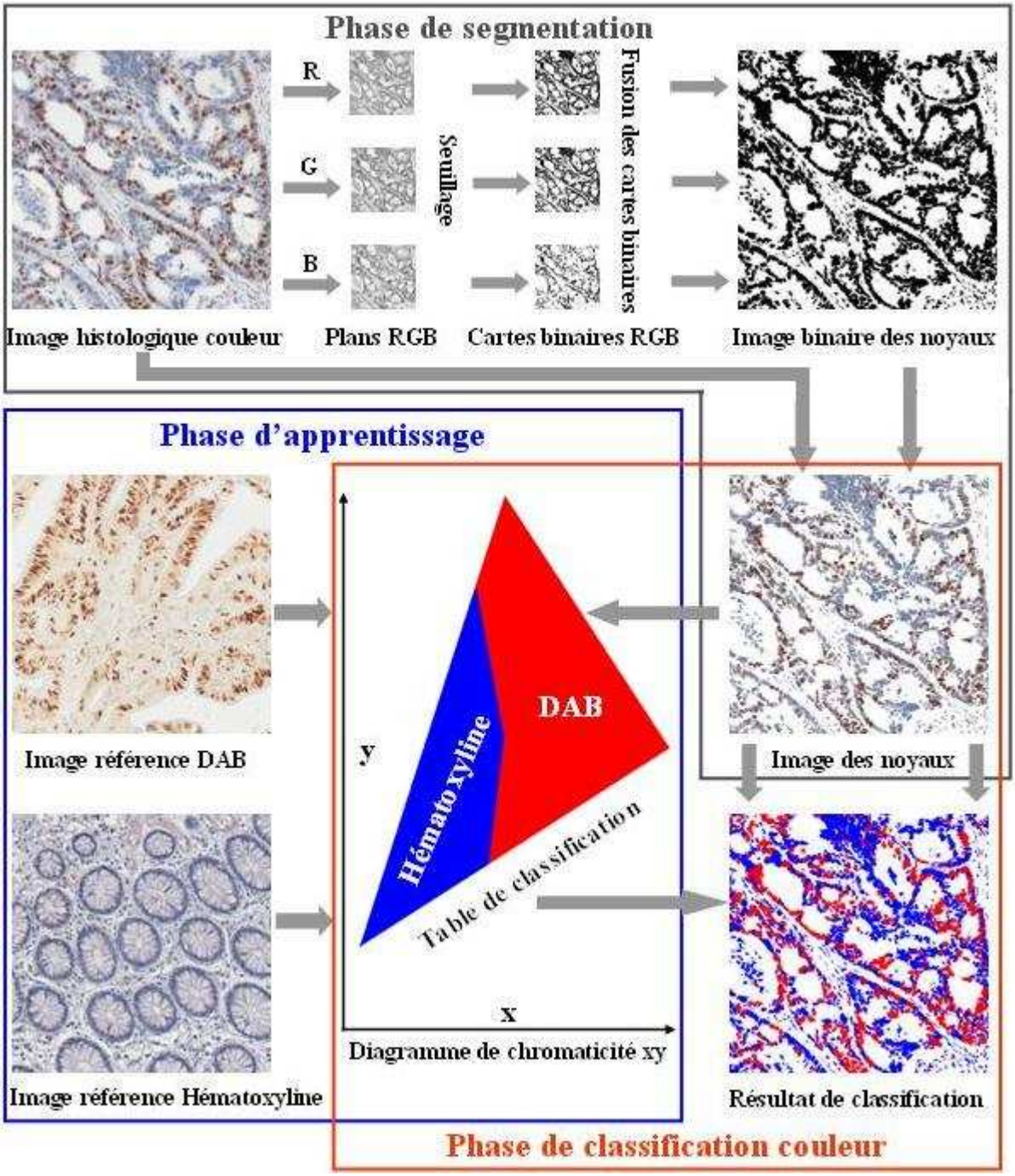


Figure 7.24: Plan global de la classification couleur des images histologiques.

7.3.3.1 Phase de segmentation

Cette phase consiste essentiellement à isoler le matériel nucléaire du fond des images histologiques à traiter. Nous avons adopté une approche marginale de segmentation, se basant sur la segmentation automatique des plans R, G et B. Voici l'algorithme concernant cette phase :

- génération des plans R, G et B associés à l'image histologique à traiter,
- segmentation automatique des trois plans marginaux R, G et B,
- union logique par l'opérateur logique « MIN » des cartes binaires issues de la segmentation des plans marginaux. L'union produit une image binaire comportant le masque binaire des noyaux cellulaires,
- génération de l'image couleur comportant uniquement les noyaux cellulaires, à l'aide de l'union logique entre l'image histologique d'origine et l'image des masques binaires des noyaux. L'union est effectuée selon l'opérateur logique « MAX » « ET ». De ce fait, l'image résultat comporte uniquement le matériel nucléaire et le fond de l'image est remplacé par un fond blanc uniforme (niveau de gris : 255).

7.3.3.2 Phase d'apprentissage

Cette phase se base essentiellement sur le recueil des attributs couleur liés aux DAB et à l'Hématoxyline dans le but de construire la table de classification couleur. La phase d'apprentissage fait appel à deux images de référence : une pour le marquage DAB, et l'autre pour la contre coloration Hématoxyline. L'algorithme associé à cette phase se déroule comme suit :

- génération des images comportant seulement les noyaux des images de référence DAB et Hématoxyline, conformément à la phase de segmentation. Cette opération nous permet d'écarter le fond de l'image de l'analyse des couleurs de référence marron et bleue liées au DAB et à l'Hématoxyline,
- calcul de la couleur moyenne des pixels associés à chaque image de référence,
- conversion des deux couleurs moyennes : DAB et Hématoxyline, de l'espace RGB vers l'espace XYZ, puis vers les coordonnées chromatiques xy,
- localisation des couleurs DAB et Hématoxyline sur le diagramme de chromaticité xy,

- génération des points repères couleur : DAB pure, Hématoxyline pure, DAB-H pure, $\overline{DAB}$, $\overline{Hématoxyline}$ et $\overline{DAB-H}$ pure (§7.3.2.2.3).
- utilisation des points couleur rouge, vert, bleu et blanc, en plus des points repères couleur cités ci-dessus, pour la délimitation des deux zones couleur DAB et Hématoxyline. En effet, ces deux zones, sous forme de polygones, divisent le triangle de Maxwell en deux régions : une région DAB et une région Hématoxyline,
- génération de la table de classification couleur. Cette étape consiste à attribuer la couleur bleue à la région associée à l'Hématoxyline, et la couleur rouge à la région DAB. Ensuite, la table de classification est sauvegardée sous forme d'image pour réaliser la phase de la classification couleur.

7.3.3.3 Phase de classification couleur

Cette phase représente la phase proprement dite de classification. Elle consiste à attribuer le matériel nucléaire associé aux images histologiques au DAB ou à l'Hématoxyline. L'algorithme de cette phase comporte essentiellement les étapes suivantes :

- conversion des coordonnées RGB de chaque pixel de l'image des noyaux (n'appartenant pas au fond de l'image), vers l'espace XYZ, puis vers les coordonnées chromatiques xy,
- projection des coordonnées xy du pixel courant sur la table de classification,
- le pixel prend la couleur bleue, si ses coordonnées chromatiques xy sont localisées dans la zone Hématoxyline. Au contraire, le pixel est coloré en rouge, s'il possède des coordonnées xy appartenant à la zone DAB,
- au terme de cette phase, tous les pixels appartenant aux noyaux cellulaires sont, soit rattachés au DAB, soit rattachés à l'Hématoxyline,
- le résultat de classification couleur est stocké dans une image couleur (image résultat sur la Figure 7.24) et sauvegardé pour des exploitations ultérieures, comme par exemple pour le calcul de l'indice de marquage.

Nous avons présenté au cours des sections précédentes, notre conception colorimétrique pour la réalisation d'une table couleur capable de différencier entre les couleurs associées au DAB

et à l'Hématoxyline. Nous avons également exposé le plan de classification couleur intégrant cette table de classification. Nous entamons au niveau des sections suivantes la phase de validation de notre approche en la comparant à d'autres approches de classification couleur.

7.4 Modélisation du système colorimétrique DAB-Hématoxyline : Tests de validation

Au cours de cette section, nous allons présenter la stratégie retenue pour la validation de notre approche de classification couleur. La stratégie se base sur la comparaison de notre approche de classification avec trois familles de classification couleur automatique citées dans la littérature. Notre méthode sera confrontée également à la classification manuelle des images histologiques. Cette section est organisée en 6 parties :

- définition du paramètre de comparaison entre les différentes méthodes,
- présentation des méthodes de classification automatique,
- protocole pour la méthode de classification manuelle,
- construction de la classification de référence à partir des classifications manuelles,
- critères de comparaison entre les différentes méthodes,
- tests statistiques retenus pour réaliser les comparaisons entre les différentes méthodes.

7.4.1 Paramètre de comparaison

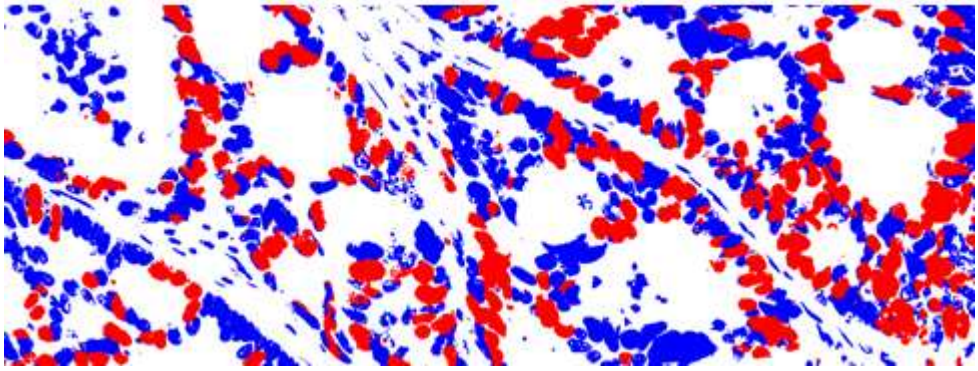


Figure 7.25: Exemple d'image histologique classifiée.

Le paramètre de comparaison adopté est l'indice de marquage « Labelling Index » noté « LI ». L'indice de marquage peut être défini comme le pourcentage des noyaux positifs

(DAB), en rapport avec le nombre total des noyaux (DAB et Hématoxyline). En mode surface, l'indice de marquage se traduit par le rapport entre la somme des surfaces nucléaires positives au marquage DAB et la somme de la totalité des surfaces nucléaires (surfaces DAB + surfaces Hématoxyline).

Le mode de calcul de l'indice de marquage retenu au cours de ce travail est le mode par surface, car notre approche de classification est une méthode de classification des pixels. A ce stade de travail, notre méthode permet seulement la classification des pixels comme DAB ou Hématoxyline. Elle ne permet pas d'isoler chaque noyau, pour adopter le mode nucléaire dans le calcul de l'indice de marquage. En se référant à la Figure 7.25, l'indice de marquage est le pourcentage de la somme des pixels rouges (DAB) en rapport avec la somme des pixels rouges et bleus (DAB + Hématoxyline). Il se traduit par la formule :

$$LI = \frac{\sum PixelsDAB}{\sum PixelsDAB + \sum PixelsHématoxyline}$$

7.4.2 Méthodes de classification automatique

Nous avons porté notre choix sur trois types de classifications automatiques développées pour des problématiques similaires à notre sujet à savoir : la classification et la quantification automatiques de l'expression du marquage immunohistochimique. Nous allons présenter brièvement les trois méthodes automatiques retenues.

7.4.2.1 Méthode automatique basée sur la teinte

Les approches s'appuyant sur la teinte pour accomplir la classification des images biologiques sont très répandues dans la littérature [Garbay et Brugal 1981] [Fermin et Degraw 1995] [Poston 1996] [Weaver et Au 1997] [Johansson et al 2000] [Van der Laak et al 2000] [Elie et al 2003] [Ruifrok et al 2003] [Lehr et al 1997] [Lehr et al 1999].Toutes ces approches se basent sur la composante teinte pour réaliser la séparation entre les plages couleurs affiliées au marquage ou à la contre-coloration. Conformément à ces approches, nous avons développé une méthode de classification automatique qui sépare la couleur bleue de l'Hématoxyline de la couleur marron du DAB (Figure 7.26).

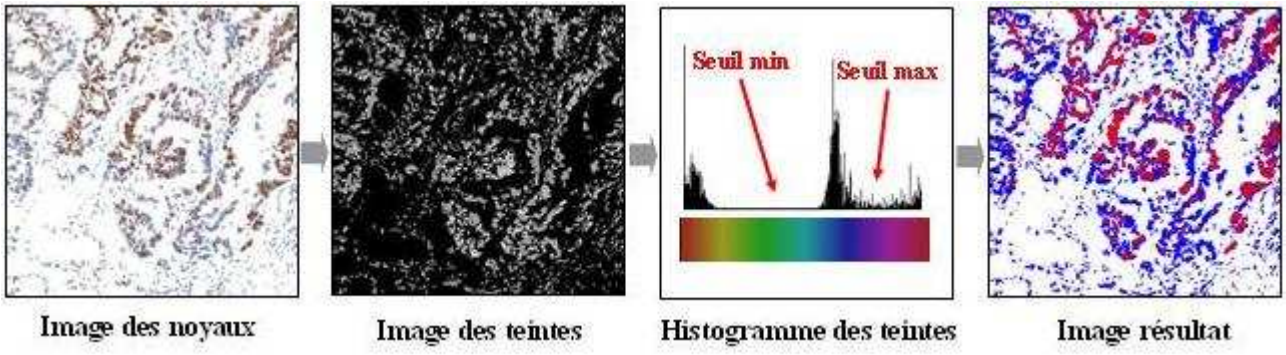


Figure 7.26: Protocole de classification automatique basé sur la teinte.

L'algorithme de classification se déroule comme suit :

- conversion de l'image des noyaux, de l'espace RGB vers l'espace HLS,
- réalisation de l'histogramme de la composante des teintes « H »,
- calcul des deux seuils : min et max au niveau de l'histogramme. Si on observe l'histogramme des teintes sur la Figure 7.26, on remarque que les seuils min et max, correspondent aux deux vallées qui délimitent le pic de l'histogramme associé à la couleur bleue,
- attribution à l'Hématoxyline des pixels qui se situent entre les seuils min et max. Ces pixels sont colorés en bleu sur l'image résultat,
- attribution au DAB des pixels qui se trouvent à l'extérieur de l'intervalle délimité par les seuils min et max. Ces pixels sont colorés en rouge sur l'image résultat.

7.4.2.2 Méthode automatique basée sur la composante bleue normalisée

Cette méthode est utilisée par Brey et al [Brey et al 2003] pour la classification des images histologiques. Selon les spécifications fournies par les auteurs, la composante monochromatique bleue normalisée notée « BN » est construite selon la formule suivante :

$$BN = \frac{B * 255}{R + G + B}$$

Cette transformation génère une image en niveaux de gris, où les nuances de marron associées au DAB sont représentées par des nuances de gris sombre, alors que les nuances de la couleur bleue de l'Hématoxyline sont représentées par des nuances de gris clair (Figure 7.27).

Conformément aux notes des auteurs de cet article [Brey et al 2003], nous avons développé l'algorithme suivant :

- génération des plans couleur R, G et B à partir de l'image des noyaux,
- construction de l'image bleue normalisée à partir des images R, G et B (Figure 7.27),
- réalisation de l'histogramme de l'image bleue normalisée,
- calcul du seuil automatique associé à l'histogramme de l'image bleue normalisée,
- attribution des pixels des noyaux dont le niveau de gris est supérieur au seuil de classification à l'Hématoxyline (pixels clairs),
- attribution des pixels des noyaux dont le niveau de gris est inférieur au seuil de classification au DAB (pixels sombres),
- attribution de la couleur bleue aux pixels Hématoxyline, et la couleur rouge aux pixels DAB (image résultat de la Figure 7.27).

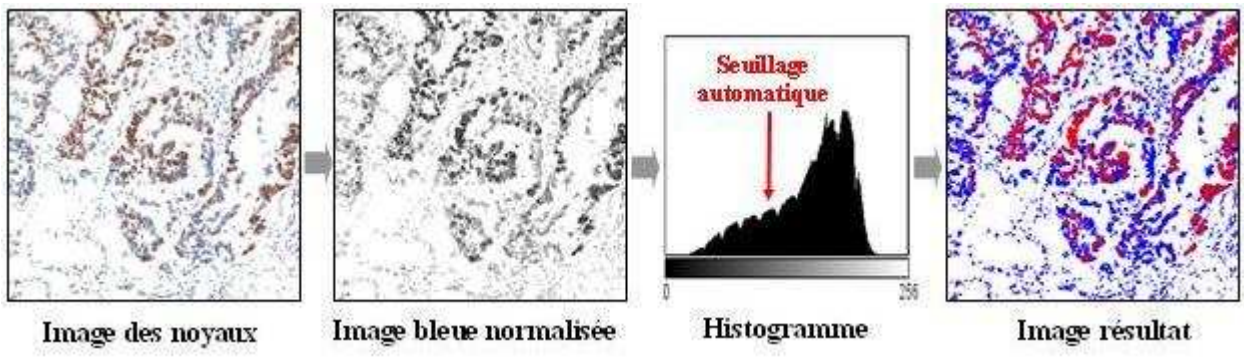


Figure 7.27: Protocole de classification automatique basé sur la composante bleue normalisée.

7.4.2.3 Méthode automatique basée sur la déconvolution couleur

Le principe de la déconvolution couleur a été développé par Ruifrok et Johnston [Ruifrok et Johnston 2001]. Il consiste à calculer la contribution d'une couleur donnée au niveau d'une image couleur. Ruifrok et Johnston ont utilisé ce principe pour l'étude de la contribution du DAB et de l'Hématoxyline au niveau des images histologiques. Cette technique a été reprise après par plusieurs auteurs pour réaliser la classification couleur [Ruifrok et al 2003] [Kim et al 2006] [Landini et Othman 2003] [Landini 2006] [Berger et al 2006].

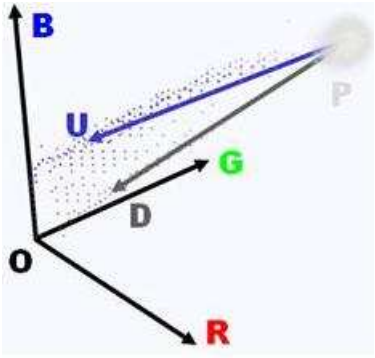


Figure 7.28: Principe de la déconvolution couleur.

La déconvolution couleur consiste à projeter les coordonnées RGB des points couleur associés à une image sur un axe représentant une couleur donnée, dans le but de mesurer la contribution de cette couleur au niveau de tous les points de l'image. La Figure 7.28 montre deux axes représentant deux couleurs notées D et U. La projection des pixels d'une image sur ces axes fournit la contribution des couleurs D et U sur chaque pixel de l'image. Notons que la construction des deux axes D et U nécessite deux images références D et U.

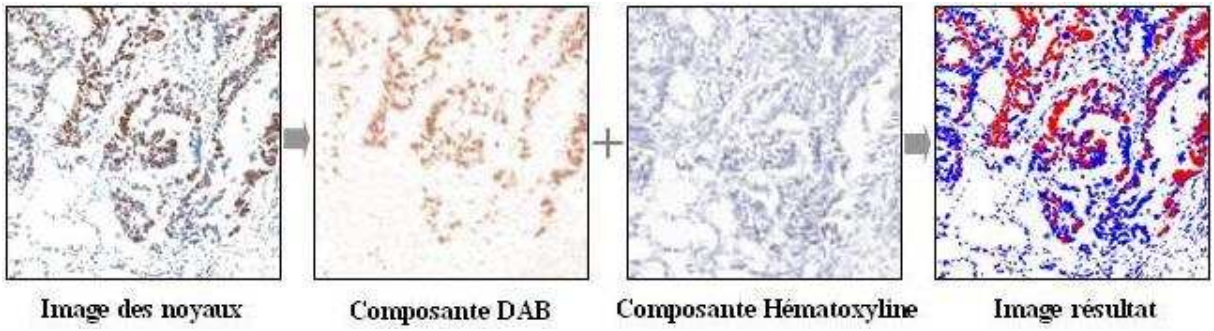


Figure 7.29: Protocole de classification automatique basé sur la déconvolution couleur.

Une implémentation en java, réalisant la déconvolution couleur est disponible sous forme de plugin imageJ [Landini 2004]. Nous avons fourni en entrée à ce programme, les deux images de références DAB et Hématoxyline utilisées pour notre méthode, pour construire les axes couleur qui leurs sont associés. Puis, nous avons intégré le plugin de Landini [Landini 2004], dans un programme que nous avons développé pour effectuer la classification couleur. L'algorithme associé à ce programme est le suivant :

- génération des deux images : DAB et Hématoxyline à partir de l'image des noyaux à l'aide de la déconvolution couleur (Figure 7.29),

- génération des images inverses associées aux composantes DAB et Hématoxyline. Tout comme les images en niveau de gris traduisent la transmission de la lumière, les images inverses traduisent l'absorption de la lumière par le DAB ou l'Hématoxyline,
- union logique entre les deux images inverses DAB et Hématoxyline par l'opérateur « Subtract ». Cette opération nous permet de réaliser la différence pixel à pixel entre les images inverses DAB et Hématoxyline, c'est-à-dire : image DAB inverse – image Hématoxyline inverse. Si l'absorption du DAB est plus importante que l'absorption liée à l'Hématoxyline, en conséquence, la quantité de couleur DAB est plus importante que la quantité de couleur Hématoxyline, alors les pixels seront positifs. Dans le cas inverse, c'est-à-dire que la quantité d'Hématoxyline est plus importante que la quantité de DAB, les pixels prendront des valeurs négatives,
- attribution des pixels positifs au DAB : ces pixels sont colorés en rouge,
- attribution des pixels négatifs à l'Hématoxyline : ces pixels sont colorés en bleu.

7.4.3 Protocole de classification manuelle

L'évaluation visuelle des lames histologiques reste la manière la plus classique et courante pour réaliser la quantification des marqueurs tumoraux. A ce titre, elle est considérée comme la méthode de référence « Golden Standard » pour classifier les images histologiques. Toute méthode de classification nouvelle, pour être validée, doit être comparée à cette méthode manuelle. Dans le but de mesurer les performances de notre méthode de classification basée sur une table de classification couleur, nous avons développé une interface graphique qui offre aux utilisateurs (pathologistes et techniciens) des outils simples pour réaliser la classification couleur sur la base de l'observation visuelle des images histologiques (Figure 7.30).

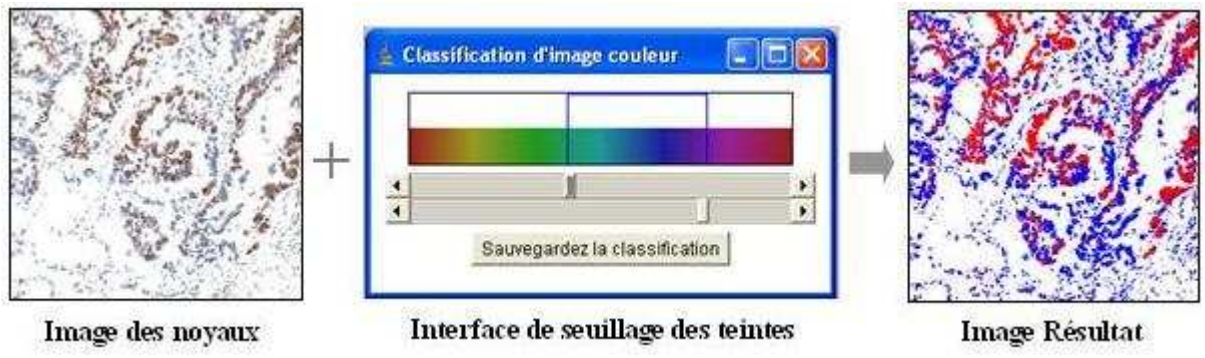


Figure 7.30: Interface graphique de la classification manuelle.

Le développement de cet outil informatique est nécessaire pour deux raisons :

- pour imposer aux utilisateurs de classifier manuellement les mêmes surfaces d'objets nucléaires classifiées (traitées) par les méthodes automatiques,
- pour récupérer le résultat de classification manuelle sous forme d'image couleur résultat afin de comparer ultérieurement les différentes méthodes manuelles et automatiques.

L'outil graphique de classification manuelle est une boîte de seuillage comportant deux curseurs glissants. L'utilisateur peut bouger et ajuster ces curseurs pour définir la largeur d'une fenêtre bleue (interface de seuillage sur la Figure 7.30). Le but est de faire varier la largeur de cette fenêtre, pour qu'elle englobe les nuances de couleur bleue associées à l'Hématoxyline. La partie du spectre couleur qui reste en dehors de la fenêtre bleue est rattachée au DAB qui est représenté par la couleur rouge.

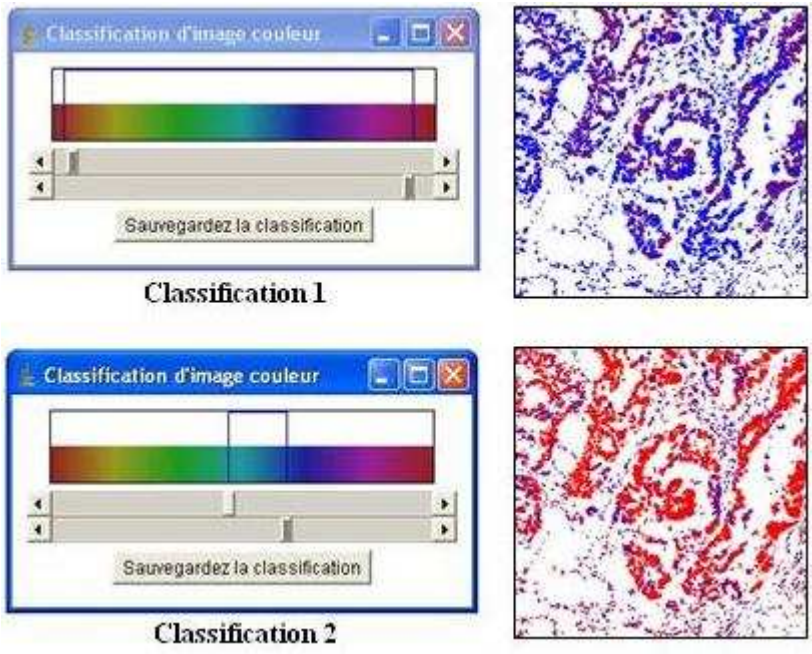


Figure 7.31: Exemples de résultats de classification manuelle.

Nous signalons que la classification manuelle est interactive. Les plages couleur (rouge et bleue) associées au DAB et à l'Hématoxyline au niveau de l'image à classifier se mettent à jour en temps réel en interaction avec les positions des curseurs de seuillage et à la largeur de la fenêtre bleue. En effet, le programme récupère les seuils de teintes définis par l'utilisateur

et modifie l'apparence de l'image suivant la valeur de teinte associée à chaque pixel de l'image. Par exemple, la Figure 7.31 montre les résultats associés à deux états de seuillages manuels différents. L'utilisateur modifie la position des curseurs de seuillage jusqu'à ce que les surfaces rouges et bleues couvrent respectivement les surfaces DAB et Hématoxyline. Nous notons aussi, que l'image originale est affichée pendant la phase de classification. Cela permet à l'utilisateur de comparer l'image qui comporte le résultat de seuillage avec limage d'origine avant de valider sa classification. Voici les étapes de l'algorithme lié à la classification manuelle :

- conversion de l'image des noyaux de l'espace RGB vers l'espace HLS,
- affichage de l'image des teintes « H » sur laquelle l'utilisateur va réaliser sa classification,
- affichage de l'outil graphique de classification (Figure 7.30 et Figure 7.31),
- mise à jour dynamique et interactive de l'image des teintes grâce aux fausses couleurs rouges et bleues,
- sauvegarde de l'image résultat de classification manuelle si l'utilisateur presse sur le bouton « Sauvegardez la classification » de l'outil graphique de classification (Figure 7.30).

La classification manuelle de l'ensemble des 88 images a été confiée à trois observateurs afin de mesurer la variation de classification inter-observateurs que nous avons abordée au cours de l'introduction de ce chapitre. Nous notons que chaque observateur a réalisé trois séances de classification sur chaque image pour analyser la variabilité de classification intra-observateur. Un intervalle de temps minimum (un jour) a été retenu entre deux séances de classification sur une même image. Cet intervalle est nécessaire pour que l'utilisateur ne soit pas influencé par les seuils de la classification précédente.

7.4.4 Construction de la classification de référence

Nous avons adopté les classifications manuelles comme classifications de référence, pour mesurer les performances des méthodes de classifications automatiques. Le recours à 9 classifications manuelles fournit 9 indices de marquage différents pour chaque image. Or, nous avons besoin d'un seul indice de marquage retenu comme référence pour le comparer

aux indices de marquage issus des classifications automatiques. Pour résoudre ce problème, nous avons proposé de calculer un seul indice de marquage de référence à partir des 9 classifications manuelles. En premier lieu, nous avons comparé les images résultats correspondant aux 9 classifications. La mise en relation des 9 images de résultats associées à chaque image d'origine, nous permet la comparaison des classifications manuelles pixel par pixel. Ensuite et pour chaque pixel, nous avons adopté la règle suivante : **si le pixel est classé comme DAB par au moins 5 classifications manuelles, alors ce pixel est considéré comme un pixel DAB. Sinon, il est considéré comme un pixel Hématoxyline**. En adoptant le nombre de 5 classifications comme seuil, nous avons en réalité retenu la classification manuelle majoritaire pour chaque pixel. Enfin, les classifications de références sont stockées dans de nouvelles images résultats. Au cours des sections suivantes, nous allons aborder les critères retenus pour comparer l'indice de marquage de la classification de référence avec l'ensemble des classifications automatiques et manuelles.

7.4.5 Critères de comparaison entre les méthodes de classification

On signale que la classification de toutes les méthodes (manuelles et automatiques) est réalisée sur les mêmes surfaces nucléaires pour chaque image. En effet, nous avons appliqué la phase de segmentation, sur l'ensemble des 88 images retenues pour le plan de comparaison des classifications. Cette opération est nécessaire pour appliquer les classifications couleurs uniquement sur les noyaux. Elle nous permet de comparer les performances des méthodes de classification sur les mêmes surfaces, ce qui réduit les paramètres de comparaison à l'indice de marquage seulement.

7.4.5.1 L'indice de marquage

Nous rappelons que le paramètre retenu pour juger les performances de chaque méthode, qu'elle soit manuelle ou automatique, est **l'indice de marquage « LI »**. En effet, l'indice de marquage de chaque méthode est comparé à celui de la méthode de référence, qui est obtenu à partir de la classification de référence, générée à partir des classifications manuelles. La comparaison des méthodes de classifications automatiques avec la classification de référence, sur la base de l'indice de marquage, devrait dégager la méthode de classification automatique

la plus performante, c'est-à-dire celle qui fournit l'indice de marquage le plus proche de celui de la classification de référence. De même, les classifications manuelles sont comparées à la classification de référence pour mesurer l'ampleur de la variabilité de classification entre les différents observateurs et la variation inhérente aux classifications réalisées par un même observateur. Pour des raisons de lisibilité et de facilité de lecture des résultats de comparaison, nous avons adopté les abréviations suivantes pour noter les classifications. Ainsi :

- La classification de référence est notée : **REF**.
- Pour les classifications automatiques :
    - la classification basée sur la composante bleue normalisée est notée **BN**,
    - la classification basée sur la composante teinte est notée **Hue**,
    - la classification associée à la déconvolution couleur est notée **CD**,
    - notre méthode de classification est notée **CHR** en rapport avec de plan de chromaticité xy utilisé pour générer la table de classification couleur.

- Les classifications manuelles :
    - les classifications manuelles réalisées par l'observateur 1 sont notées : **FG1, FG2 et FG3**,
    - Les classifications manuelles réalisées par l'observateur 2 sont notées : **MH1, MH2 et MH3**.
    - Les classifications manuelles réalisées par l'observateur 3 sont notées : **MPM1, MPM2 et MPM3**.

En plus de la comparaison directe à l'aide de l'indice de marquage, nous avons développé d'autres critères de comparaison pour déterminer la méthode de classification la plus proche de la classification de référence. Ces critères et leurs méthodes de calcul seront détaillés dans les paragraphes suivants.

7.4.5.2 La différence absolue des indices de marquage

Le principe est de calculer pour chaque image histologique, la différence absolue entre l'indice de marquage d'une méthode de classification donnée et l'indice de marquage de

référence. Si on prend une méthode de classification x par exemple, la différence absolue est donnée par la formule suivante :

$$DifAbs = \left| LI(x) - LI(REF) \right|$$

Le recours à la différence absolue est nécessaire, car l'utilisation de la simple différence peut nous conduire à des résultats de comparaisons erronés. En effet, soient les valeurs 40%, 50%, et 60%, les indices de marquage des images im1, im2 et im3 fournis par la classification de référence. Et 20%, 50% et 80% les indices respectifs de ces mêmes images mesurés par une méthode de classification x. Les différences respectives entre ces deux méthodes donnent 20%, 0%, et -20%. La moyenne de ces différences est nulle (0%). Elle peut nous conduire à considérer à tort les deux méthodes parfaitement concordantes. Tandis que, les différences absolues associées aux deux méthodes sont dans l'ordre : 20%, 0% et 20%. La moyenne liée à ces différences est de 13,33%, ce qui reflète la différence réelle qui existe entre les deux méthodes : référence et x.

7.4.5.3 Le pourcentage des pixels mal classifiés

Un pixel mal classifié au niveau des images histologiques, est un pixel appartenant à la surface objet (surfaces nucléaires) qui est :

- soit un pixel DAB classifié à tort comme un pixel Hématoxyline,
- soit le cas inverse, un pixel Hématoxyline rattaché à tort au DAB.

Le calcul des pixels mal classifiés nécessite la mise en relation de l'image résultat référence avec celle de la classification à évaluer. Le pourcentage des pixels mal classifiés peut être calculé par : le rapport entre la somme des pixels mal classés par rapport à l'image de référence, et la somme totale des pixels de la surface objet (noyaux). Ce paramètre varie entre 0% et 100% : 0% si la classification correspond parfaitement à la classification de référence et 100%, si les deux divergent complètement. En conséquence, la classification présentant le pourcentage de pixels mal classés le plus faible, se rapprochant de 0, est la meilleure classification.

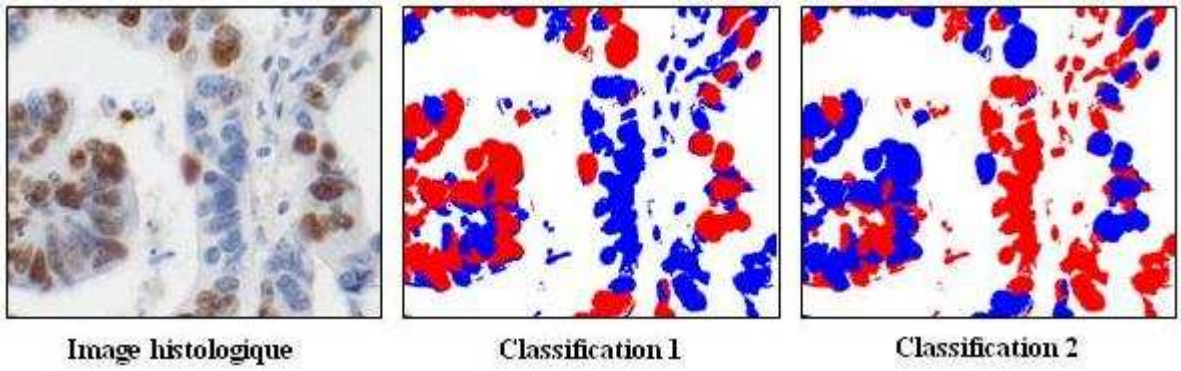


Figure 7.32: Critères de comparaison entre les classifications.

Le critère reposant sur le pourcentage des pixels mal classifiés est complémentaire de celui basé sur le calcul des différences absolues. Ce dernier utilisé seul, peut s'avérer parfois insuffisant pour évaluer la qualité d'une classification couleur. Si nous prenons comme exemple la Figure 7.32, les classifications 1 et 2 présentent des indices de marquages très proches (1 :50,25%, 2 :49,75%). La différence absolue va nous donner la valeur de 0,5%, qui est le signe d'une très bonne concordance entre les deux classifications. Or, nous avons juste inversé les couleurs bleue et rouge entre les deux classifications. Le pourcentage des pixels mal classifiés qui est de 100%, nous permet de nous rendre compte que les deux classifications sont en réalité complètement différentes. Si nous nous limitons à utiliser uniquement le critère de différence absolue, nous pourrions évaluer des classifications entièrement erronées comme performantes. Pour cette raison, le recours aux deux critères est nécessaire pour réaliser de bons tests de comparaison entre les différentes classifications.

7.4.6 Les analyses statistiques

7.4.6.1 Analyse des indices de marquage (LI)

Dans le but de comparer les différentes classifications (manuelles et automatiques à la classification de référence), nous avons utilisé le test statistique de Kruskal-Wallis (seuil de significativité : $p < 0.05$) adapté à l'analyse des échantillons appariés < LI classification x, LI classification référence>. Le test de Kruskal-Wallis est un test non paramétrique, il a été retenu puisque le test W de Shapiro-Wilk (seuil de significativité : $p < 0.05$) a révélé que la majorité de nos échantillons appariés ne suivent pas une loi de distribution normale (Annexe C).

7.4.6.2 Analyse des différences absolues et des pourcentages des pixels mal classifiés :

Chaque classification, manuelle ou automatique, est représentée par un échantillon de 88 images résultats. La mise en relation de cet échantillon avec celui des images résultats de référence, nous a permis de générer deux nouveaux échantillons :

- le premier représentant les différences absolues de l'indice de marquage,
- le second comportant les pourcentages des pixels mal classifiés.

Ces échantillons sont considérés comme appariés et sont analysés directement grâce aux descripteurs statistiques généraux tels que : les moyennes et les variances.

7.5 Modélisation du système colorimétrique DAB-Hématoxyline : Tests de robustesse

L'objectif de cette étude, est l'analyse du comportement des méthodes de classifications couleur automatiques sous des conditions de luminosité variées. Nous voulons répondre à l'aide de cette étude à la question suivante : « est-ce que nous aurons le même indice de marquage pour une méthode de classification donnée, si une image est traitée sous des niveaux d'éclairage différents ? ». De même, nous voulons également savoir qu'elle est la méthode de classification automatique la plus robuste pour traiter des images histologiques acquises sous des conditions d'éclairage détériorées. Nous qualifions de robuste une méthode de classification qui fournit le même indice de marquage pour la même image histologique malgré les variations d'intensité lumineuse qui pourraient affecter cette image.

7.5.1 Variation réelle des conditions d'éclairage

L'étude porte sur 10 champs microscopiques sélectionnés de façon à illustrer toute la dynamique observée pour le marquage Ki67 au sein de notre jeu de données initial. Nous avons réalisé 7 protocoles d'acquisitions d'images pour chaque champ microscopique, toujours au grossissement x20. Les 7 acquisitions associées à chaque champ, correspondent à :

- une acquisition selon les conditions optimales d'éclairage,
- 3 acquisitions associées à trois niveaux d'éclairages bas. Nous avons en effet diminué, après chaque acquisition, l'intensité de la lampe qui éclaire le champ microscopique,

- 3 acquisitions avec trois niveaux d'éclairages hauts, en augmentant à chaque acquisition, l'intensité de la lampe d'éclairage.

L'ensemble des 7 séries d'acquisitions a été effectué selon des niveaux de luminosité croissants. Nous avons reproduit les mêmes niveaux de luminosité pour la globalité des 10 champs microscopiques étudiés. En finalité, nous avons acquis 70 images histologiques.

Les niveaux de luminosité, autorisés par la station d'imagerie microscopique Samba 2050™ (TRIBVN^{TM}), ne nous ont pas permis des variations importantes des conditions d'acquisitions d'images. Le logiciel d'acquisition d'image rattaché à cette station est équipé d'un module « assurance qualité » qui nous empêche de travailler dans des conditions d'éclairage trop éloignées de l'éclairage optimal. De plus, la station d'imagerie effectue au moment de chaque acquisition une opération de « balance des blancs » qui consiste à normaliser l'image du champ microscopique par l'image d'un champ vide. Cette opération tend en réalité à compenser le manque ou l'excès de luminosité, elle diminue en partie l'effet de notre action pour modifier la luminosité des images.

7.5.2 Variation simulée (virtuelle) des conditions d'éclairage

En complément aux séries d'acquisitions d'images associées aux 10 champs microscopiques, nous avons adopté une approche virtuelle, en simulant la modification des conditions d'éclairage des images. En effet, nous avons récupéré les images des 10 champs microscopiques acquises sous des conditions d'éclairage optimales. Ensuite, nous avons fait varier de façon virtuelle la luminosité de chaque image : nous avons utilisé 15 niveaux de luminosité différents. Si on estime à 100% le niveau initial de luminosité associé à l'image optimale, les 15 niveaux de luminosité générés sont gradués de 10% à 150%. 15 images ont été ainsi générées suivant ces niveaux de luminosité. En totalité, nous avons généré 150 images pour les 10 champs microscopiques.

Le recours à cette méthodologie, nous a permis de simuler des conditions d'éclairage plus étendues que celles obtenues par le la station d'imagerie microscopique. Si on observe la

Figure 7.33, on remarque que l'approche virtuelle permet ainsi de simuler des conditions d'éclairage très variées pour une image histologique donnée.

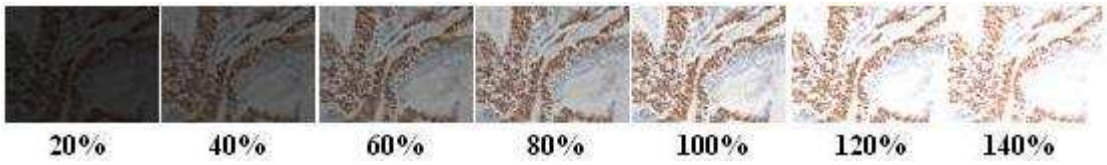


Figure 7.33: Variations virtuelles de la luminosité d'une image histologique.

En récapitulatif, nous avons utilisé en totalité :

- 70 images pour les variations réelles des niveaux de luminosité,
- 150 images pour les variations virtuelles des niveaux de luminosité.

L'ensemble de ces images a été classifié selon les quatre méthodes de classification couleur automatiques (§7.4.2) afin de désigner la méthode automatique la plus stable face aux variations des conditions de luminosité. Nous signalons, qu'un masque binaire est calculé sur la surface objets (noyaux) au niveau des images acquises dans les conditions optimales d'éclairage (images de référence). Ensuite, ce masque est superposé sur les autres images ayant des niveaux de luminosité modifiés. Ces opérations nous permettent d'appliquer les classifications couleur automatiques au niveau de la même surface objet, à travers toutes les images d'un champ microscopique donné.

Au niveau de la section suivante, nous allons analyser les résultats de classifications couleur générés par l'ensemble des méthodes de classification abordées au cours de ce chapitre. Notre méthode sera essentiellement confrontée à la méthode de classification de référence. Elle sera également comparée aux autres méthodes de classification : manuelles et automatiques.

7.6 Modélisation du système colorimétrique DAB-Hématoxyline : Résultats des expérimentations

7.6.1 Introduction

Au niveau de cette partie, nous allons aborder les résultats de classification générés par les différentes méthodes de classification couleur présentées au cours des sections précédentes. Les résultats concernent les deux études présentées dans les sections précédentes :

- tests de validation (§7.4),
- tests de robustesse (§7.5).

Les tests de validation concernent l'ensemble des méthodes de classification couleur : manuelles et automatiques, abordées dans ce chapitre (§7.4.2 et 7.4.3). Toutes ces méthodes, seront analysées selon les trois paramètres (§ 7.4.5) :

- l'indice de marquage « LI » (§ 7.4.5.1),
- la différence absolue de l'indice de marquage par rapport à l'indice de marquage de la classification de référence (§ 7.4.5.2),
- le pourcentage des pixels mal classifiés associé à chaque classification par rapport à la classification de référence (§ 7.4.5.3).

Les tests de robustesse, concernent seulement les méthodes de classification automatiques. Les résultats de classification, devront nous indiquer les méthodes de classification les plus stables. Les résultats de cette étude prennent en compte uniquement les indices de marquage.

7.6.2 Résultats des tests de validation

7.6.2.1 Les classifications manuelles

7.6.2.1.1 Indices de marquage

7.6.2.1.1.1 Description générale

Globalement, les séries de classifications manuelles sont fortement corrélées avec la classification de référence (Tableau 7.1). Ce qui semble naturel, puisque cette dernière découle des classifications manuelles.

Tableau 7.1: Corrélation entre les classifications manuelles et la classification de référence.

Classif. Manuelles	MPM 1	MPM 2	MPM 3	MH 1	MH 2	MH 3	FG 1	FG 3	FG 3
Fact. Corrélation	0,983	0,991	0,989	0,99	0,993	0,995	0,979	0,973	0,986

La Figure 7.34 regroupe les classifications manuelles relatives à chaque observateur. Elle nous indique que la moyenne des indices de marquages propres à FG, est la plus proche de la moyenne associée à la classification de référence (ligne horizontale) : MH surévalue cette moyenne, à l'inverse MPM la sous-évalue.

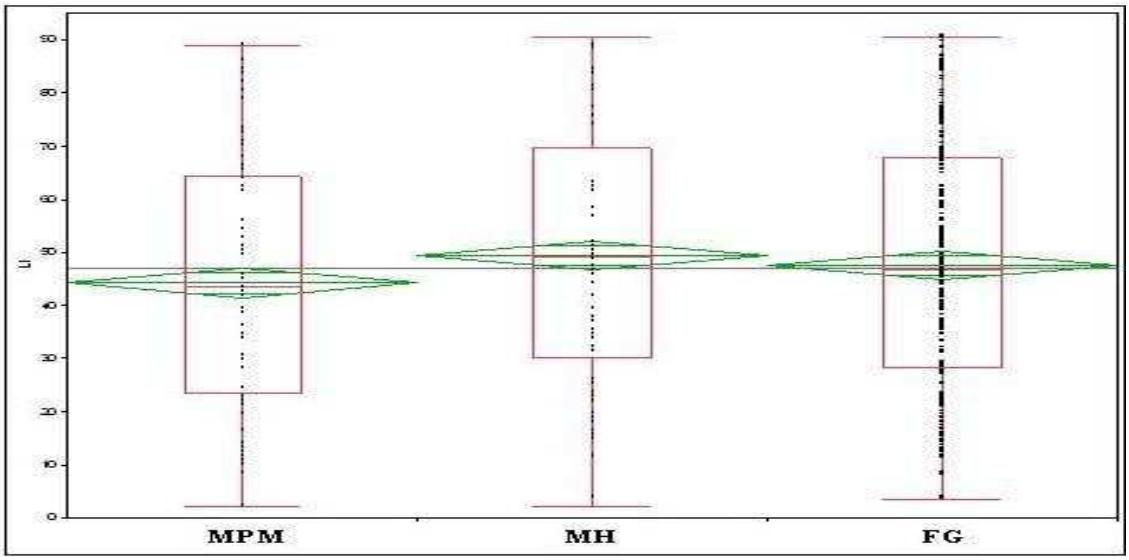


Figure 7.34: Indices de marquage des classifications manuelles regroupés par observateur. Représentation « Box-Plot » de la distribution des séries de classifications regroupées pour chacun des 3 observateurs MPM, MH et FG. Le diamant vert représente la position de la moyenne et de son intervalle de confiance à 95%.

La Figure 7.35 illustre la distribution globale des résultats de classifications manuelles réalisées par les trois observateurs. En se focalisant sur la moyenne associée à chaque série de classification manuelle, nous pouvons constater que :

- les classifications de FG (FG1, FG2 et FG2) sont les plus concordantes à la classification de référence,
- il existe plus de régularité dans les séries de classifications de FG et MPM par rapport aux classifications de MH,
- les séries de classifications de MPM sous-estiment les indices de marquage par rapport aux classifications de FG. Ce constat confirme les observations rapportées par [Furness 1997] [Van-der-Laak 2001], concernant la variation d'évaluation du

marquage due à la différence de perception psycho-visuelle des couleurs entre les différents observateurs,

- les scores de classifications réalisés par MH, représentent des séries d'indices de marquage variables. Cette fluctuation démontre l'existence d'une variabilité intra-observateur en addition à la variabilité inter- observateurs.

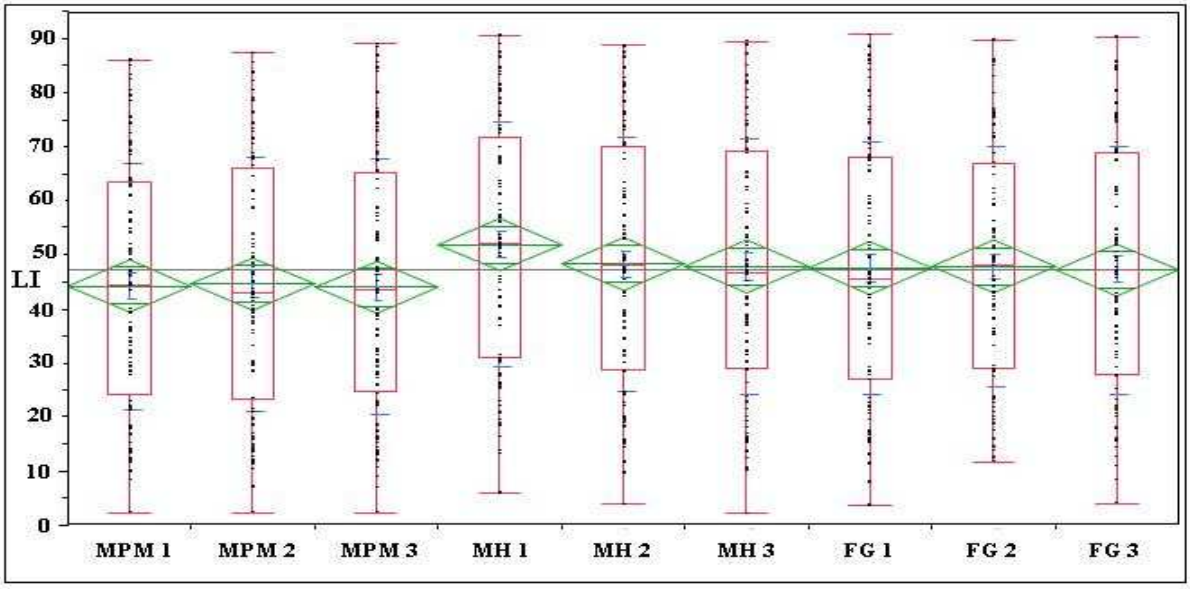


Figure 7.35: Indices de marquage des 9 classifications manuelles. Représentation « Box-Plot » de la distribution des 88 indices de marquage associés aux 9 classifications manuelles effectuées par les 3 observateurs : MPM, MH et FG (3 répétitions par observateur).

En dépit de leurs fortes corrélations avec la classification de référence (tableau 7.1), les classifications manuelles manifestent cependant des variations à la fois entre les différents observateurs (inter-observateurs) et entre les différentes séries propres aux observateurs (intra-observateurs) (Figure 7.34 et Figure 7.35). En effet, cette variabilité est inhérente à la subjectivité associée à la perception des sensations couleur, qui varie selon les sujets et selon le contexte dans lequel les couleurs sont présentées.

La Figure 7.36 montre deux exemples de séries de classifications appariées selon la méthode de représentation de « Bland et Altman » [Bland et Altman 1986 1996]. La figure 7.36 illustre deux cas de comparaison : un cas incluant deux séries de classifications (MH2-MH3) réalisées par le même observateur (Figure 7.36.A) et un autre présentant la comparaison de deux séries de classifications (MH1-MPM3) propres à deux observateurs différents (Figure 7.36.B). Par exemple, les classifications MH2 et MH3 (Figure 7.36.A) sont associées à une

différence moyenne de LI égale à : 3,557 ± 0,363. Les classifications appariées MH1 et MPM3 (Figure 7.36.B) fournissent une différence égale à : 7,902 ± 0,618.

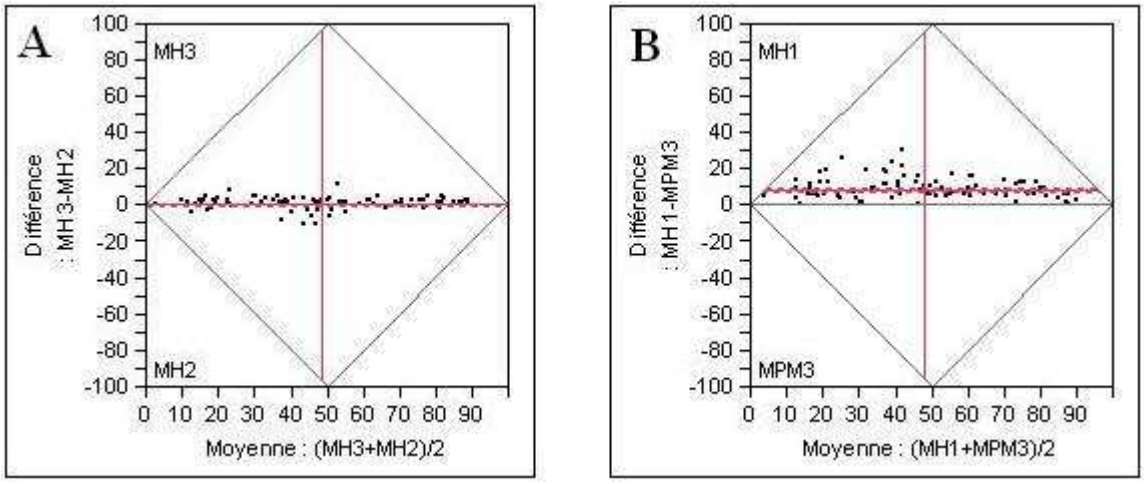


Figure 7.36: Exemples de classifications appariées. A : Comparaison de deux classifications relatives à un seul observateur (MH2,MH3). **B :** Comparaison de deux classifications relatives à deux observateurs différents (MH1, MPM3).

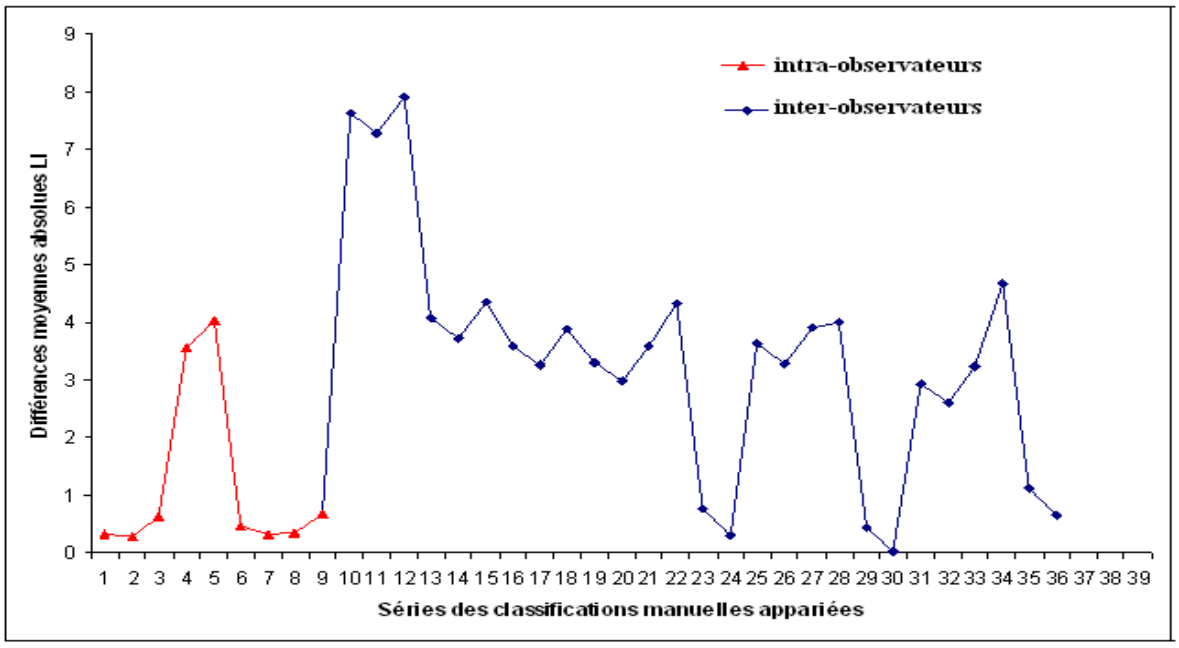


Figure 7.37: Différences absolues des LI associées aux classifications manuelles appariées. En rouge (triangle) sont rapportées les classifications appariées d'un même observateur. En bleu (losange) sont rapportées les classifications appariées relatives à des observateurs différents.

La mise en correspondance de l'ensemble des classifications manuelles nous a permis le calcul de la différence moyenne absolue des indices de marquage appariés relative à chaque

couple de classifications (Figure 7.37). La variabilité de classification entre les observateurs (différence (moyenne absolue) : 3,384 ± 0,574) est plus importante que la variabilité issue des classifications d'un même observateur (différence (moyenne absolue) : 1,182 ± 0,477). L'ampleur de la variabilité inter-observateurs (max : 7,902 ± 0,618) peut atteindre presque le double de l'ampleur de la variabilité intra-observateur (max : 4,026 ± 0,411) (Figure 7.37).

7.6.2.1.1.2 Résultats statistiques

L'application du test de Kruskal-Wallis sur les classifications manuelles appariées avec la classification de référence, revient en réalité à tester la validité statistique de notre approche pour construire la classification de référence. Les « valeurs p » rapportées au niveau du Tableau 7.2 démontrent clairement que toutes les classifications manuelles possèdent une distribution identique à celle associée à la classification de référence.

Tableau 7.2: Distributions statistiques des LI des classifications manuelles. Les valeurs p du test Kruskal-Wallis associées aux classifications manuelles appariées avec la classification de référence.

	MPM1/REF	MPM2/REF	MPM3/REF	MH1/REF	MH2/REF	MH3/REF	FG1/REF	FG2/REF	FG3/REF
val. p	0,4356	0,455	0,3475	0,1246	0,6275	0,7583	0,8626	0,7628	0,8673

Ces résultats valident en effet, notre classification de référence qui représente toutes les classifications manuelles à la fois. Au lieu de comparer une classification automatique donnée à toutes les classifications manuelles, il suffit de la comparer seulement à la classification de référence pour qu'elle soit validée.

En dépit de la variabilité visuelle observée au niveau de la distribution des indices de marquage liée aux classifications manuelles (Figure 7.37), le test statistique de Kruskal-Wallis révèle que globalement toutes ces classifications possèdent la même distribution des indices de marquage, excepté les couples formés par la classification MH1 d'une part, et les séries de classifications de MPM d'autre part (Tableau 7.3). Ces résultats démontrent l'absence de variabilité significative entre les classifications d'un même observateur.

Tableau 7.3: Distributions des LI des classifications manuelles appariées. Valeurs P (test Kruskal-Wallis) associées aux couples de classifications manuelles.

	MPM1	MPM2	MPM3	MH1	MH2	MH3	FG1	FG2	FG3
MPM1									
MPM2	0,9269								
MPM3	0,8988								
MH1	0,0281	0,0312	0,0261						
MH2	0,2343	0,2384	0,1865	0,2775					
MH3	1,0847	0,9419	0,2609	0,2245	0,8707				
FG1	0,3676	0,3763	0,3201	0,2012	0,7969	0,9387			
FG2	0,2841	0,2895	0,2184	0,2001	0,8637	0,9740	0,9210		
FG3	0,4074	0,4057	0,3289	0,1513	0,7314	0,8637	0,9340	0,8037	

7.6.2.1.2 Différences absolues des indices de marquages par rapport aux indices de marquages de référence

La Figure 7.38 illustre la distribution des différences absolues (entre les LI et les LI de référence) associées à l'ensemble des séries de classifications manuelles. Nous pouvons remarquer à travers ces distributions une variabilité apparente entre les trois observateurs. Cette variabilité se manifeste également à travers les résultats de classifications réalisées par un même observateur (ex. MH).

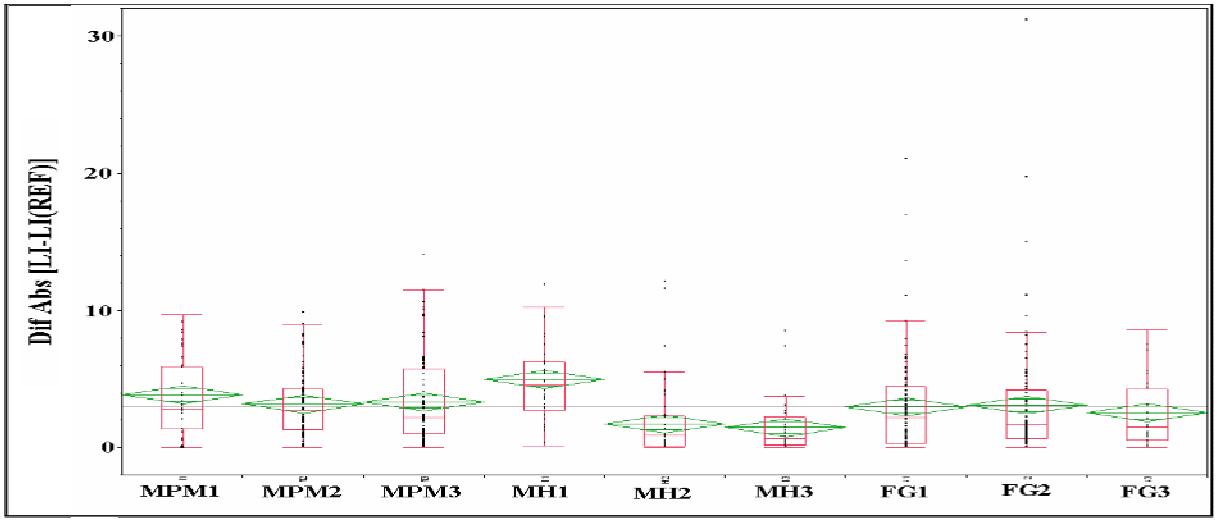


Figure 7.38: Les différences absolues (LI-LI(REF)) des 9 classifications manuelles. Représentation « Box-Plot » de la distribution des 88 différences absolues (LI-LI(REF)) associés aux 9 classifications manuelles effectuées par les 3 observateurs : MPM, MH et FG (3 répétitions par observateur).

En regroupant les séries de classifications propres à chaque observateur (moyenne), on confirme la variabilité entre les trois observateurs à classifier les images, mais on observe également la fluctuation du niveau d'erreur (LI-LI (REF)), commise par chaque observateur, à travers la classification des différentes images histologiques (Figure 7.39).

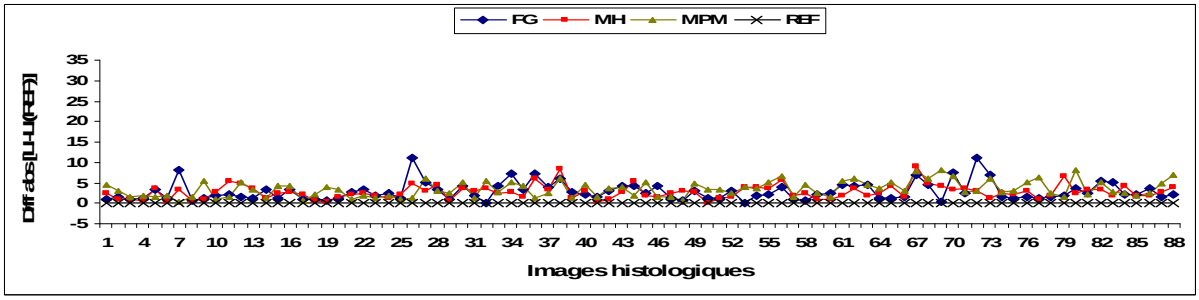


Figure 7.39: Les différences absolues (LI-LI(REF)) regroupées par observateur. Distribution des différences absolues (LI-LI(REF)) moyennes associées aux classifications manuelles des trois observateurs : MPM, MH et FG pour la classification des 88 images histologiques.

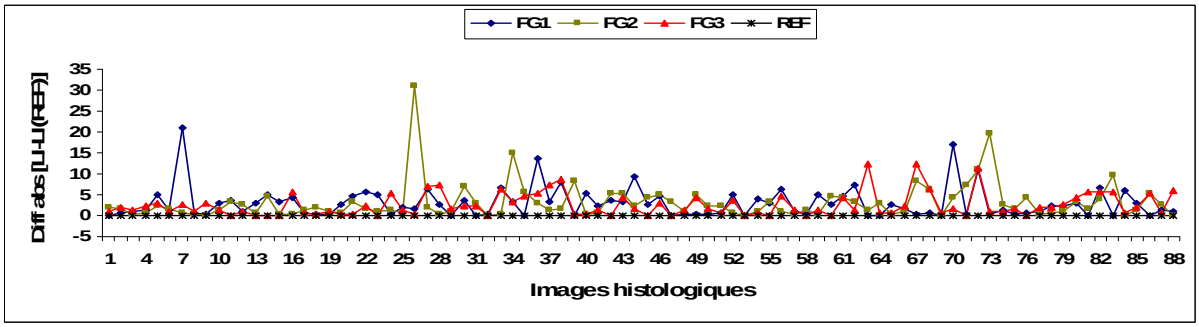


Figure 7.40: Les différences absolues (LI-LI(REF)) des classifications de l'observateur FG. Distribution des différences absolues (LI-LI(REF)) associées aux trois classifications de FG des 88 images histologiques.

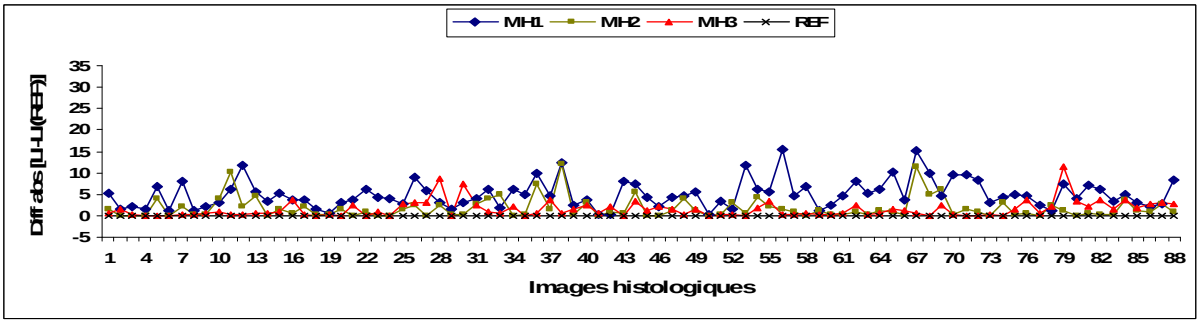


Figure 7.41: Les différences absolues (LI-LI(REF)) des classifications de l'observateur MH. Distribution des différences absolues (LI-LI(REF)) associées aux trois classifications de MH des 88 images histologiques.

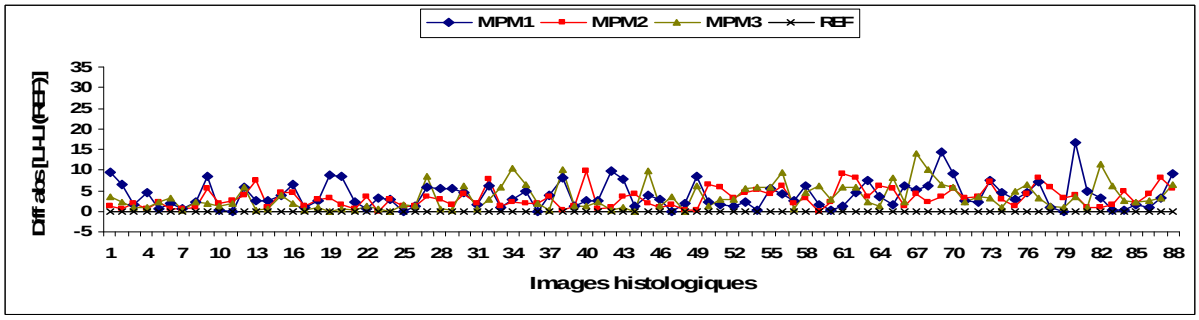


Figure 7.42: Les différences absolues (LI-LI(REF)) des classifications de l'observateur MPM. Distribution des différences absolues (LI-LI(REF)) associées aux trois classifications de MPM des 88 images histologiques.

Les Figures Figure 7.40, Figure 7.41 et Figure 7.42 illustrent la variabilité de classification associée à un même observateur. Cette variabilité est une donnée constante rattachée aux classifications manuelles. En effet, il semble difficile à un observateur de déterminer le même indice de marquage s'il classe plusieurs fois la même image histologique.

7.6.2.1.3 Pourcentages des pixels mal classés par rapport à la classification de référence

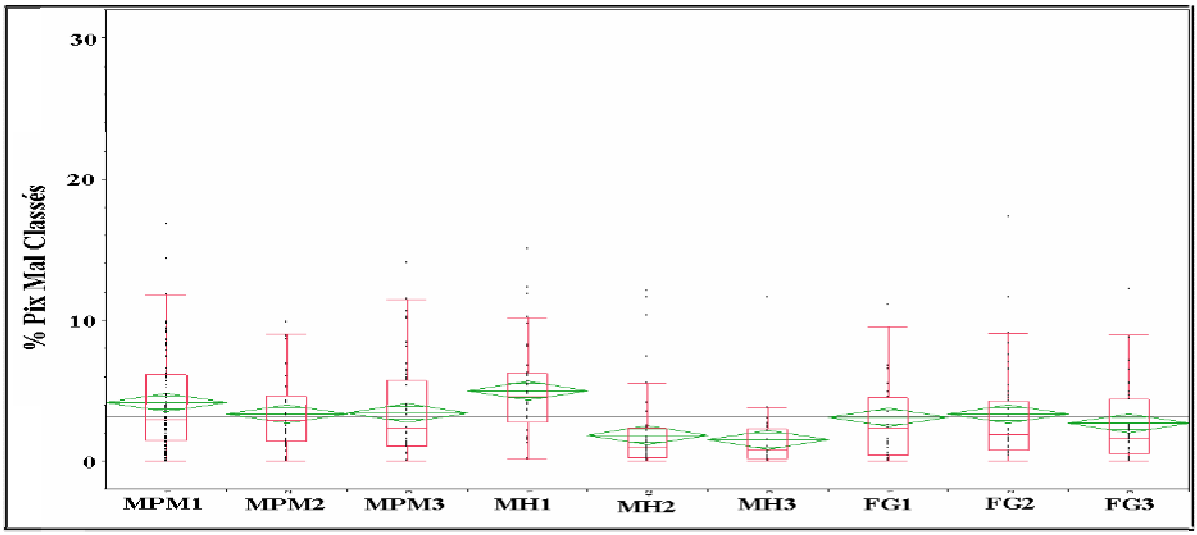


Figure 7.43: Pourcentages des pixels mal classés associés aux classifications manuelles. Représentation « Box-Plot » de la distribution des 88 % des pixels mal classés associés aux 9 classifications manuelles effectuées par les 3 observateurs : MPM, MH et FG (3 répétitions par observateur).

La Figure 7.43 révèle l'existence d'une variabilité du pourcentage de pixels mal classés entre les différentes classifications manuelles. Néanmoins, les classifications associées à FG présentent plus de stabilité par rapport aux deux autres observateurs. Nous observons que les deux séries de classification MH2 (1,818 ± 2,410) et MH3 (1,498 ± 1,887), sont les séries qui génèrent le moins de pixels mal classés. Cependant, les classifications de MH sont les plus variables, si nous prenons en compte seulement les classifications propres à chaque observateur.

La mise en relation des scores du pourcentage moyen de pixels mal classés associés aux trois observateurs, propres à chaque image histologique classifiée, confirme l'existence d'une variabilité de classification entre les observateurs (Figure 7.44).

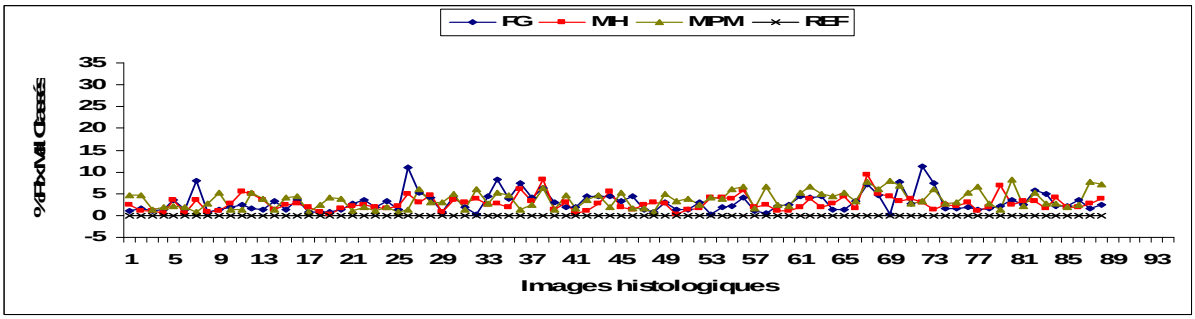


Figure 7.44: Pourcentages des pixels mal classés regroupés par observateur. Distribution des Pourcentages des pixels mal classés moyens associés aux trois observateurs : MPM, MH et FG pour la classification des 88 images histologiques.

L'affichage conjoint des scores liés aux trois séances de classifications, propres à chaque observateur, met en relief une variabilité nette entre les classifications réalisées par un même observateur (Figure 7.45, Figure 7.46 et Figure 7.47). A cette variabilité, s'additionne un autre type de variation à travers le lot d'images histologiques classifiées. En effet, le pourcentage de pixels mal classés, varie d'une image à une autre. Cette fluctuation est due probablement à l'appréciation visuelle des nuances couleur associées au DAB et à l'Hématoxyline. Il semble parfois difficile à un observateur de répéter la séparation du DAB et de l'Hématoxyline avec la même précision, notamment s'ils manifestent des nuances couleur proches.

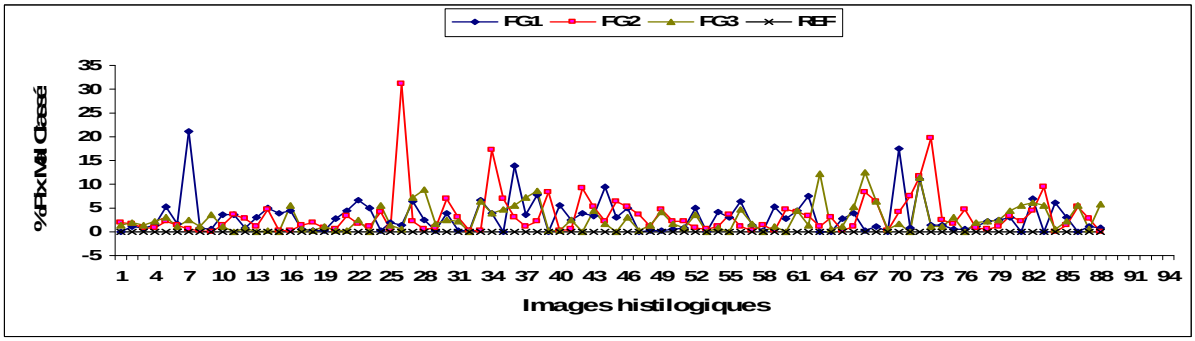


Figure 7.45: Pourcentages des pixels mal classés des classifications de l'observateur FG. Distribution des Pourcentages des pixels mal classés associés aux trois classifications de FG des 88 images histologiques.

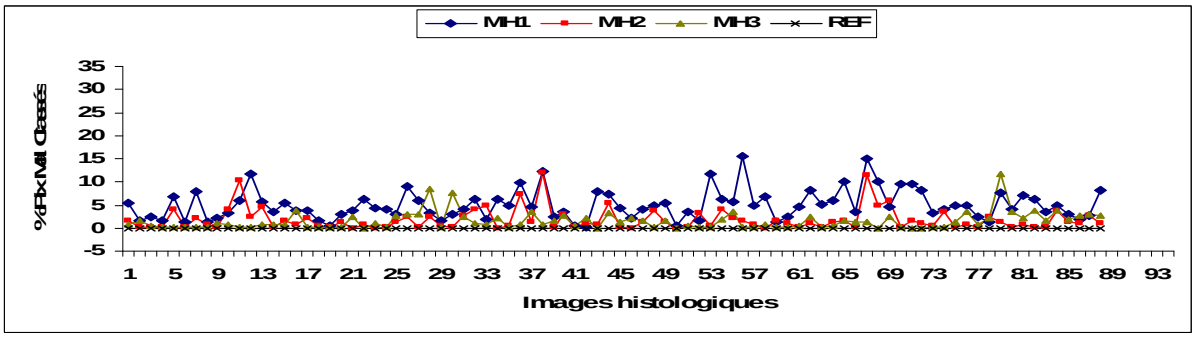


Figure 7.46: Pourcentages des pixels mal classés des classifications de l'observateur MH. Distribution des Pourcentages des pixels mal classés associés aux trois classifications de MH des 88 images histologiques.

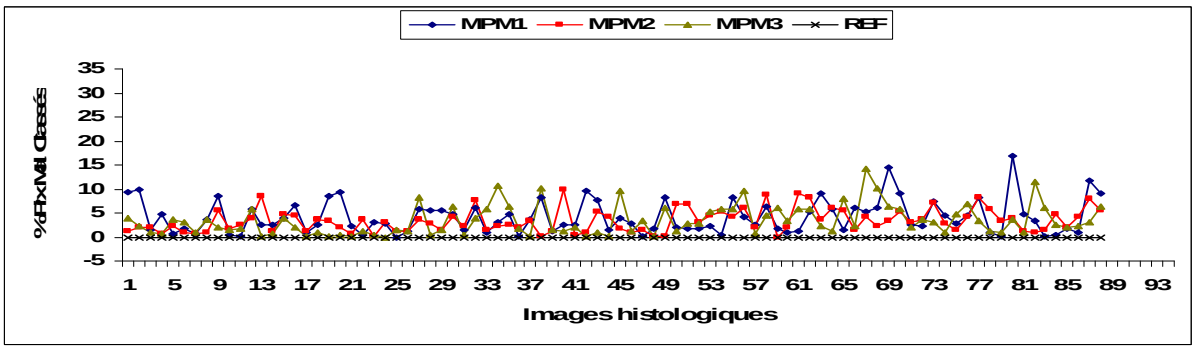


Figure 7.47: Pourcentages des pixels mal classés des classifications de l'observateur MPM. Distribution des Pourcentages des pixels mal classés associés aux trois classifications de MPM des 88 images histologiques.

7.6.2.2 Les classifications automatiques

7.6.2.2.1 Indices de marquage

7.6.2.2.1.1 Description générale

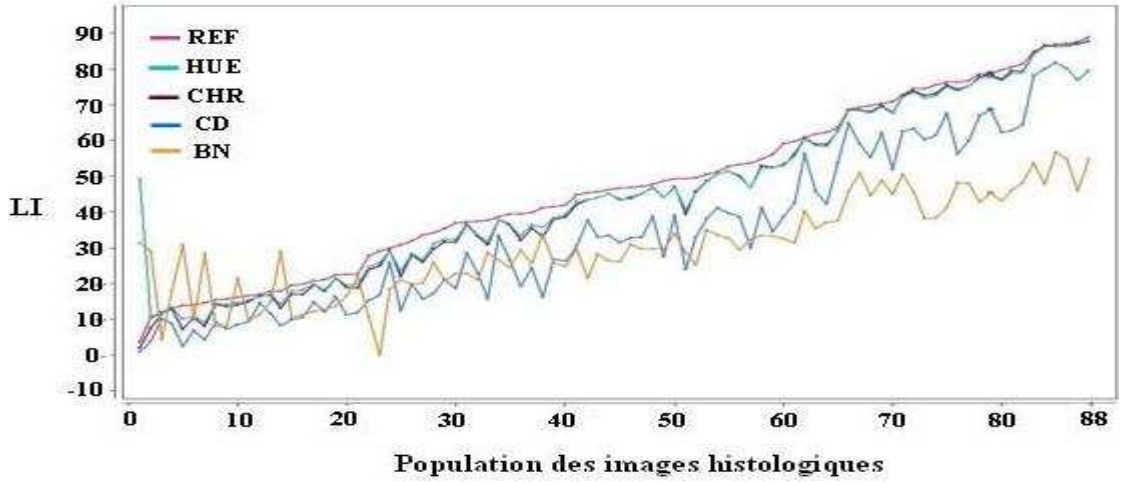


Figure 7.48: Déscription générale des LI associés aux classifications automatiques. Par soucis de clarification de la représentation, les 88 échantillons ont été ordonnés selon les valeurs croissantes de LI (indice de marquage) pour la classification de référence (REF).

La Figure 7.48 illustre l'estimation des indices de marquage associés aux classifications automatiques de l'ensemble des 88 images histologiques. Elle démontre globalement que toutes les classifications automatiques sous-estiment les niveaux d'indice de marquage Ki67 par rapport à la classification de référence. Les deux méthodes automatiques HUE et CHR se détachent des autres méthodes automatiques CD et BN, et se rapprochent le plus des scores de classifications associés à la classification de référence. De plus, ces deux classifications (HUE et CHR) se superposent le plus souvent : elles sont plus proches l'une de l'autre que l'une et l'autre de la classification de référence. Nous notons également, que la classification BN est la classification la moins stable. Elle est la méthode de classification qui oscille le plus.

Si on se réfère à la Figure 7.49, on note là aussi que les classifications HUE et CHR distancent les classifications CD et BN. Ce sont les deux méthodes qui présentent les distributions les plus similaires à la distribution des indices de marquage associées à la classification de référence. La classification CD se place en troisième position après les classifications HUE et CHR, devançant ainsi la méthode de classification BN.

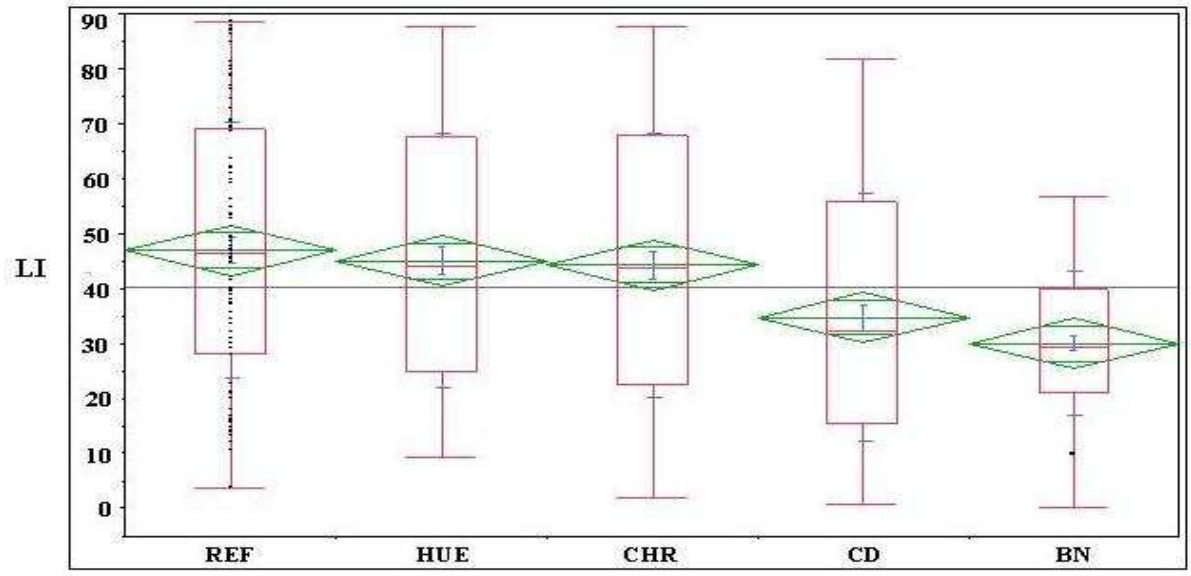


Figure 7.49: Distribution des LI associés aux classifications automatiques. Représentation « Box-Plot » de la distribution des indices de marquage (LI) associés aux classifications automatiques des 88 images histologiques.

Les facteurs de corrélation des classifications automatiques avec la classification de référence placent la méthode CHR en premier suivie de la méthode CD (Tableau 7.4). La méthode HUE, quand à elle, est décalée à la troisième position.

Tableau 7.4: Corrélation des classifications automatiques avec la référence. Les facteurs de corrélation des classifications automatiques par rapport à la classification de référence.

Classif. Auto	CHR	CD	HUE	BN
Fact. Corrélation	0,984	0,972	0,953	0,892

7.6.2.2.1.2 Résultats statistiques

Le test de Kruskal-Wallis démontre que seules les classifications CHR et HUE, sont conformes à la classification de référence. Les distributions des indices de marquage générées par les classifications CD et BN ne sont pas identiques à celle des indices de marquage de la classification de référence (Tableau 7.5). La distribution de la classification HUE est plus proche de la distribution de référence (p : 0.5018) que ne l'est la distribution CHR (p : 0.3795). L'analyse statistique des classifications CHR/HUE appariés, démontre que ces deux méthodes sont encore plus semblables (p=0,8209) que ne l'est HUE à la référence (p=0,5018) et à moindre degré CHR à la référence (p=0,3795).

Tableau 7.5: Distributions statistiques des LI des classifications automatiques. Les valeurs p du test Kruskal-Wallis associées aux classifications automatiques appariées avec la classification de référence.

Classifications appariées	CHR/REF	HUE/REF	CD/REF	BN/REF
valeurs P	0,3795	0,5018	0,0005	<,0001

7.6.2.2.2 Différences absolues des indices de marquages par rapport aux indices de marquages de référence

La Figure 7.50 démontre que les méthodes CHR et HUE sont les classifications les plus proches de la référence. Excepté une image histologique (image n^{o} 39), les courbes associées à CHR et HUE se superposent le plus souvent. La classification BN semble celle qui fournit les différences absolues (LI-LI (REF)) les plus importantes.

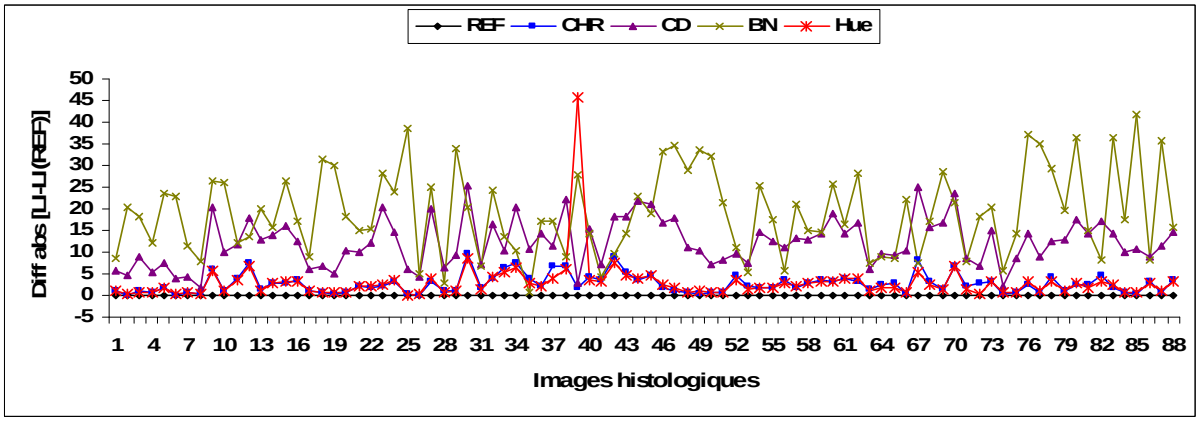


Figure 7.50: Différences absolues (LI-LI(REF)) des classifications automatiques. Variabilité des différences absolues (LI-LI(REF)) associées aux classifications automatiques des 88 images histologiques.

La Figure 7.51 démontre aussi que les deux classifications CHR et HUE se détachent des autres classifications, ce sont celles qui génèrent les différences de LI les moins élevées.

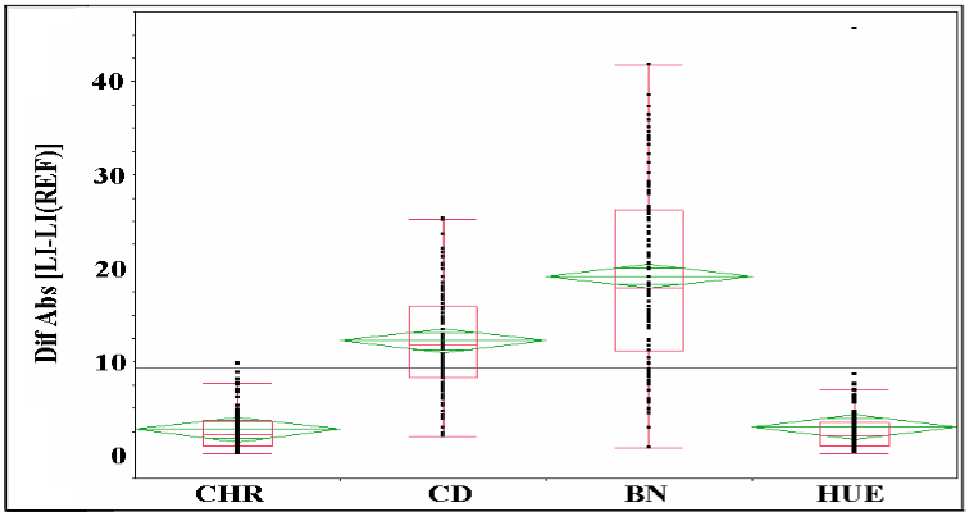


Figure 7.51: Différences absolues (LI-LI (REF)) associées aux classifications automatiques. Représentation « Box-Plot » de la distribution des différences absolues (LI-LI (REF)) associées aux classifications automatiques des 88 images histologiques.

7.6.2.2.3 Pourcentages des pixels mal classés par rapport à la classification de référence

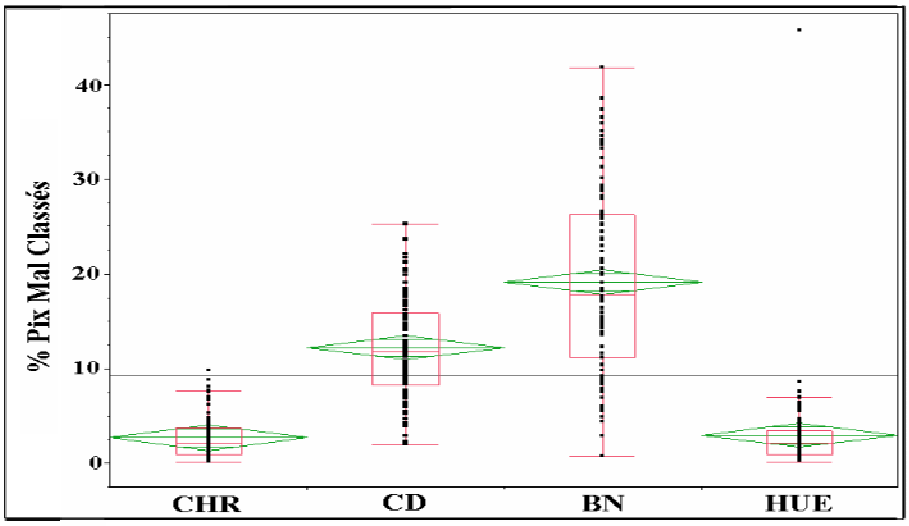


Figure 7.52: Pourcentages des pixels mal classés associés aux classifications automatiques. Représentation « Box-Plot » des pourcentages des pixels mal classés associés aux classifications automatiques des 88 images histologiques.

La Figure 7.52 montre une grande variabilité de la méthode BN et à moindre mesure la méthode CD. Les classifications CHR et HUE sont celles engendrant le moins de pixels mal classés. Nous observons aussi, une variabilité réduite rattachée à ces deux types de classifications.

La classification CHR génère le moins de pixels mal classés (3,341 ± 2,24), elle devance légèrement la classification HUE. La classification BN produit la plus mauvaise performance (21,883 ± 8,331).

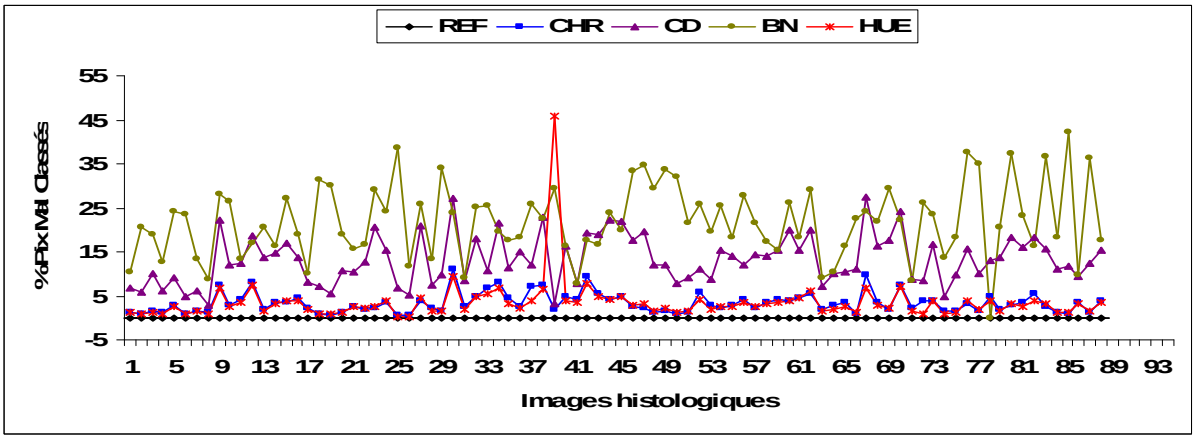


Figure 7.53: Pourcentages de pixels mal classés liés aux classifications automatiques. Variabilité des pourcentages de pixels mal classés associés aux classifications automatiques des 88 images histologiques.

L'observation directe, des pourcentages de pixels mal classés des images histologiques, révèle une grande similitude entre les deux classifications CHR et HUE (Figure 7.53). Les classifications CD et BN comportent beaucoup de variabilité, elles génèrent la plupart du temps un pourcentage de pixels mal classés dépassant les 5%.

Les classifications CHR et HUE manifestent des pourcentages de pixels mal classés très proches pour la majorité des images, en revanche, elles divergent pour classer une image en particulier (image n° 39 sur la Figure 7.53). La Figure 7.54 montre les résultats de classification de cette image par les méthodes CHR et HUE. Cette figure illustre la défaillance de la classification HUE à classer correctement les surfaces de contre-coloration ; elle les attribue à tort au marquage DAB (Figure 7.54.D). Visuellement, la méthode CHR semble classer correctement l'image histologique, elle produit une classification très proche de la référence (pourcentage de pixel mal classés : 1,994%) (Figure 7.54.C). Par contre, HUE donne un résultat de classification très éloigné de la référence (pourcentage de pixel mal classés : 45,667%). Nous observons, qu'elle classifie les nuances bleues de contre-coloration qui tirent vers le gris comme appartenant au DAB (Figure 7.54). En effet, ces nuances couleurs se caractérisent en particulier par leurs faibles niveaux de saturation, puisqu'elles sont proches de l'axe achromatique (nuances de gris). En outre, la teinte se singularise par son

instabilité particulière face aux basses saturations [Angulo 2003] [Pujas 1996]. Une variation minime au niveau de ces plages de saturation, induit des variations significatives de la teinte. A notre avis, se sont ces spécificités liées au comportement singulier de la teinte qui semblent expliquer un indice de marquage anormalement élevé (49,265%) par rapport à la référence (3,598%).

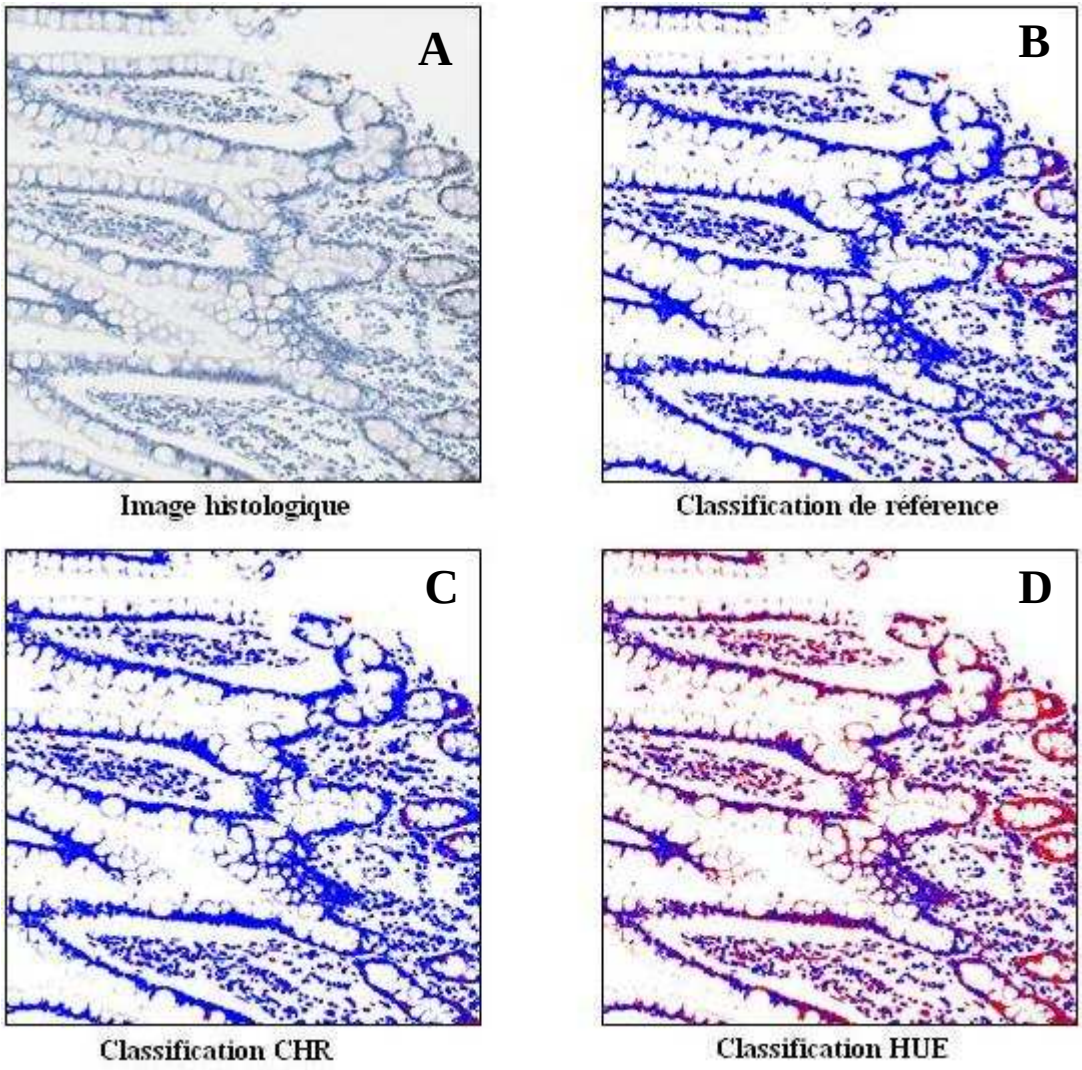


Figure 7.54: Pourcentages de pixels mal classés liés aux classifications automatiques. Variabilité des pourcentages de pixels mal classés associés aux classifications automatiques des 88 images histologiques. A : image histologique avant classification. B : résultat de classification de référence manuelle. C : résultat de classification automatique CHR. D : résultat de classification automatique HUE.

7.6.3 Résultats des tests de robustesse

7.6.3.1 Variation réelle des conditions d'éclairage du microscope

Les courbes retraçant les fluctuations des indices de marquage reflètent une divergence dans le comportement des classifications automatiques vis-à-vis de la variation du niveau de luminosité (Figure 7.55, Figure 7.56, Figure 7.57 et Figure 7.58). Ainsi, les classifications BN et CD, en dépit de leur non-conformité statistique avec la classification de référence, démontrent une relative robustesse (Figure 7.55 et Figure 7.56). Les courbes associées à ces classifications indiquent une légère déviation vers les niveaux extrêmes de luminosité. Cependant, les courbes BN paraissent plus stables (Figure 7.55).

Visuellement, la classification CHR semble celle générant les courbes les plus stables (Figure 7.57). Toutefois, les indices de marquage rattachés à ces courbes ont tendance à baisser légèrement au niveau des luminosités les plus fortes. La classification HUE présente globalement une stabilité à proximité des conditions optimales de luminosité (niveau 0). A l'inverse, on observe une diminution du marquage sur les niveaux bas et hauts de luminosité (Figure 7.58).

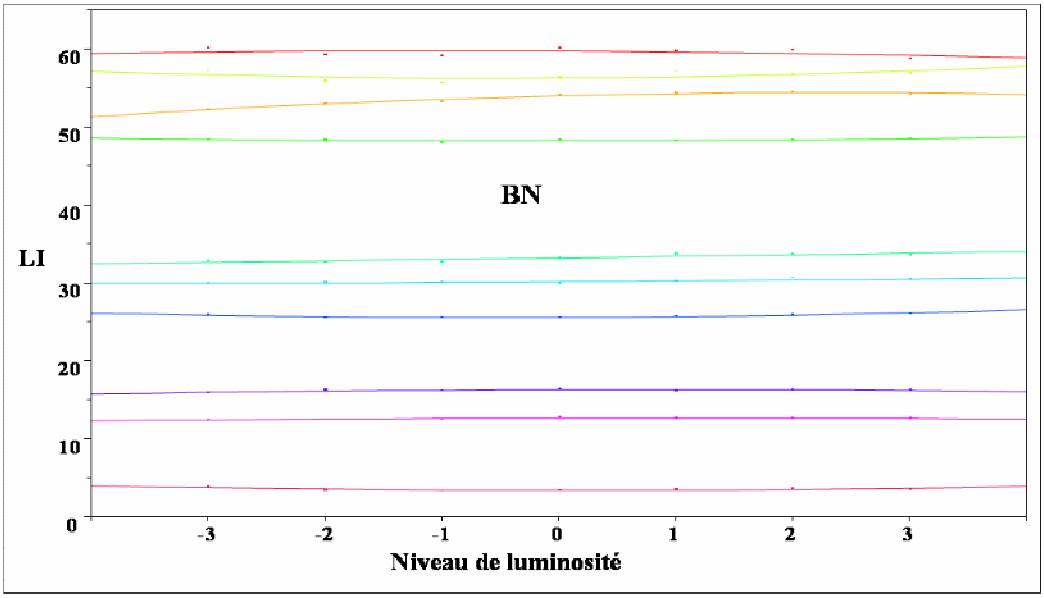


Figure 7.55: Classification BN des images histologiques sous différents niveaux d'éclairage. Indices de marquage générés par la classification BN de 10 champs microscopiques numérisés sous différents niveaux de luminosité. Le niveau de luminosité 0 est le niveau optimal d'éclairage.

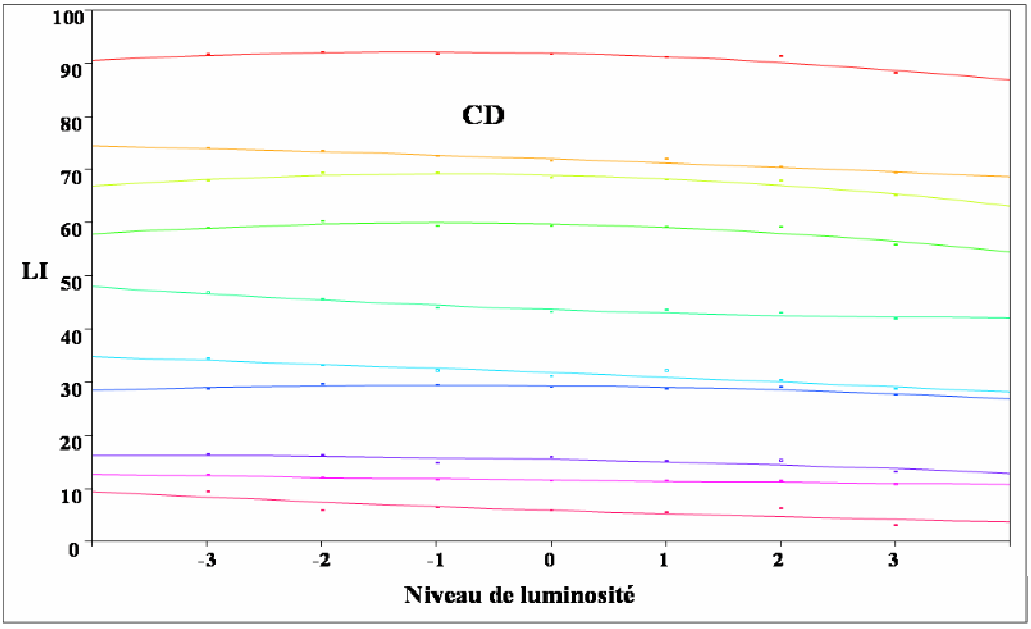


Figure 7.56: Classification CD des images histologiques sous différents niveaux d'éclairage. Indices de marquage générés par la classification CD de 10 champs microscopiques numérisés sous différents niveaux de luminosité. Le niveau de luminosité 0 est le niveau optimal d'éclairage.

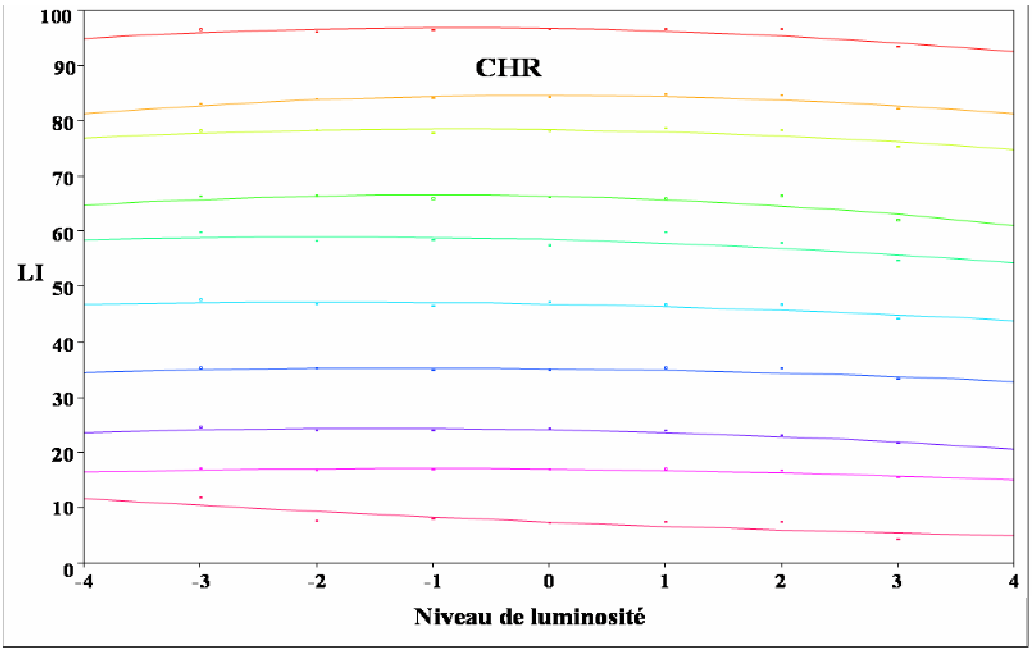


Figure 7.57: Classification CHR des images histologiques sous différents niveaux d'éclairage. Indices de marquage générés par la classification CHR de 10 champs microscopiques numérisés sous différents niveaux de luminosité. Le niveau de luminosité 0 est le niveau optimal d'éclairage.

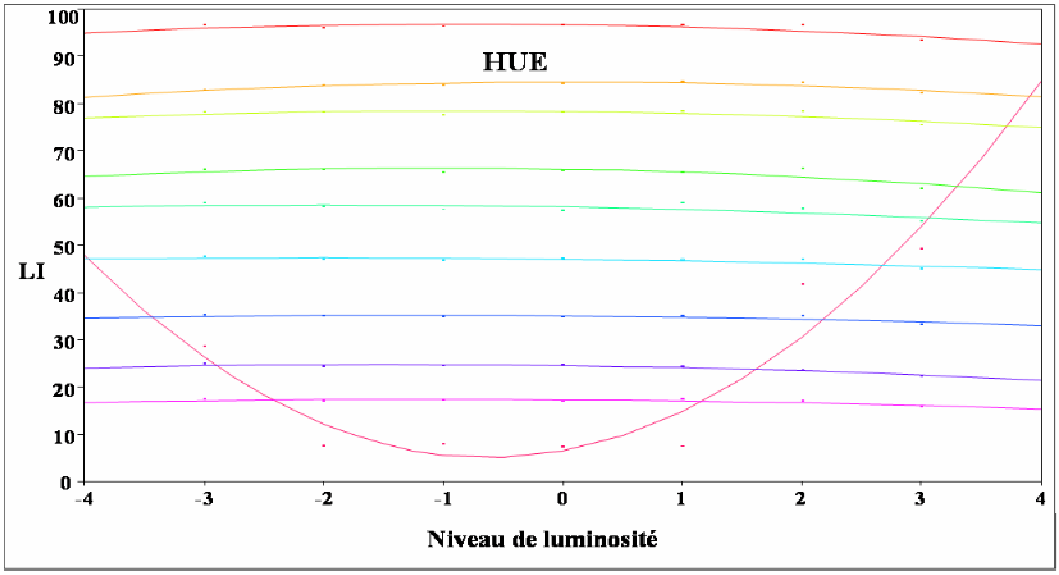


Figure 7.58: Classification HUE des images histologiques sous différents niveaux d'éclairage. Indices de marquage générés par la classification HUE de 10 champs microscopiques numérisés sous différents niveaux de luminosité. Le niveau de luminosité 0 est le niveau optimal d'éclairage.

Nous avons remarqué une flexion globale des courbes associée à toutes les classifications, lorsque le niveau de lumière s'écarte des conditions optimales d'éclairage (Figure 7.55, Figure 7.56, Figure 7.57 et Figure 7.58). Afin d'analyser en détail ce phénomène, nous avons calculé les écarts de classification, engendrés sous les différents niveaux de luminosité, par rapport au niveau optimal.

La comparaison directe, des classifications automatiques sur le même graphique, révèle leurs comportements totalement différents, face aux variations de luminosité (Figure 7.59). En effet, la classification CHR semble la plus stable, face à des variations modérées de la luminosité. Néanmoins, elle a tendance à s'écarter du niveau de marquage initial dans les hautes luminosités et à un degré moindre dans les basses luminosités. Les classifications BN et CD commencent à produire des écarts de classification avec les petites modifications de luminosité. Ces écarts s'amplifient si les variations de luminosité, deviennent importantes. La classification CD surévalue l'indice de marquage quand on diminue la luminosité. Au contraire, elle le sous-évalue si on augmente la luminosité. La classification BN adopte un comportement diamétralement opposé : l'augmentation de luminosité provoque la baisse d'indice de marquage et sa baisse provoque l'inverse.

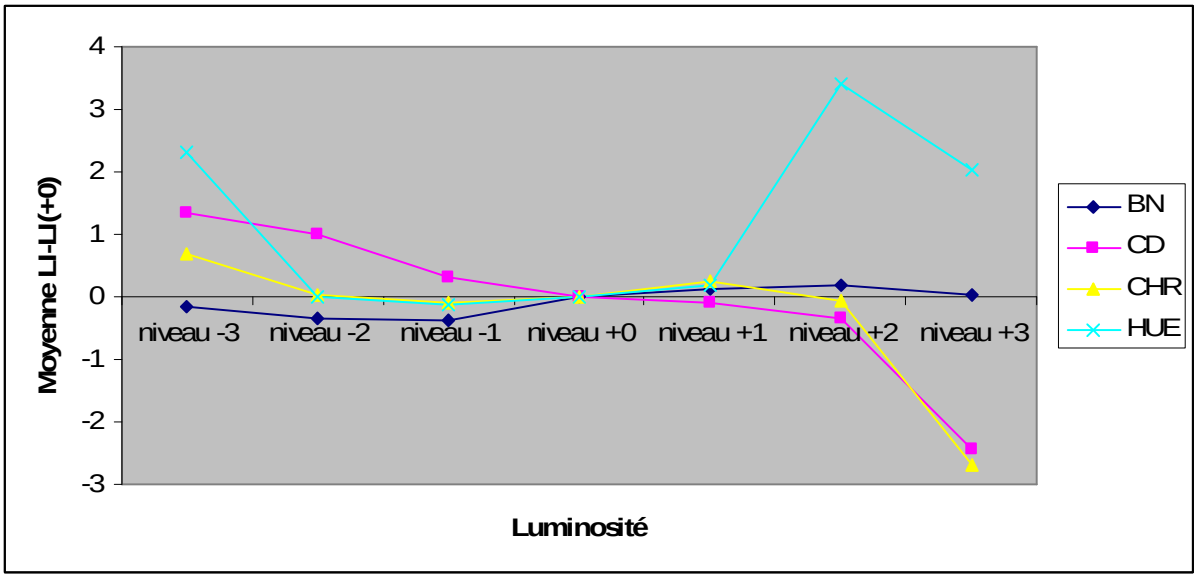


Figure 7.59: Robustesse des classifications automatiques selon les conditions d'éclairage. Ecarts moyens des indices de marquage des classifications automatiques, par rapport à la classification sous un niveau optimal des conditions d'éclairage du microscope (niveau +0).

La classification HUE se montre la méthode la plus sensible aux variations de luminosité. Les indices de marquage commencent à s'écarter du niveau initial, avec les premières variations de lumière. Ensuite, les erreurs de classification deviennent très importantes, dans les niveaux extrêmes de luminosité.

Tableau 7.6: Robustesse des classifications automatiques sous différents niveaux d'éclairage. Erreurs moyennes (± écart type) des indices de marquage générés par les classifications automatiques par rapport à la classification sous un niveau optimal de luminosité (niveau +0).

	BN	CD	CHR	HUE
Erreur moyenne	0,6667	0,7938	0,5387	1,1547
Ecart type moyen	0,4017	0,7378	0,6493	4,7159

Le Tableau 7.6 confirme nos analyses visuelles de la figure 7.59. Ce tableau révèle que la classification CHR est la méthode la plus robuste. C'est celle qui engendre l'erreur de classification la plus faible, si on varie la luminosité d'une image couleur (0,5387 ± 0,6493). Les classifications BN et CD se montrent aussi robustes. Par contre, la classification HUE se révèle la méthode de classification la plus sensible aux variations de luminosité. Elle produit en moyenne, une erreur de classification deux fois plus importante (1,1547 ± 4,7159), que celle associée à la classification CHR.

7.6.3.2 Variation simulée (virtuelle) de luminosité des images histologiques

La classification des images histologiques sous des variations de luminosité contrôlées et graduées, nous révèle une stabilité particulière de notre méthode de classification CHR (Figure 7.60). Nous observons que cette méthode, maintien à un niveau constant les indices de marquage. Nous notons seulement un léger infléchissement de ce niveau, vers les luminosités les plus élevées (150%).

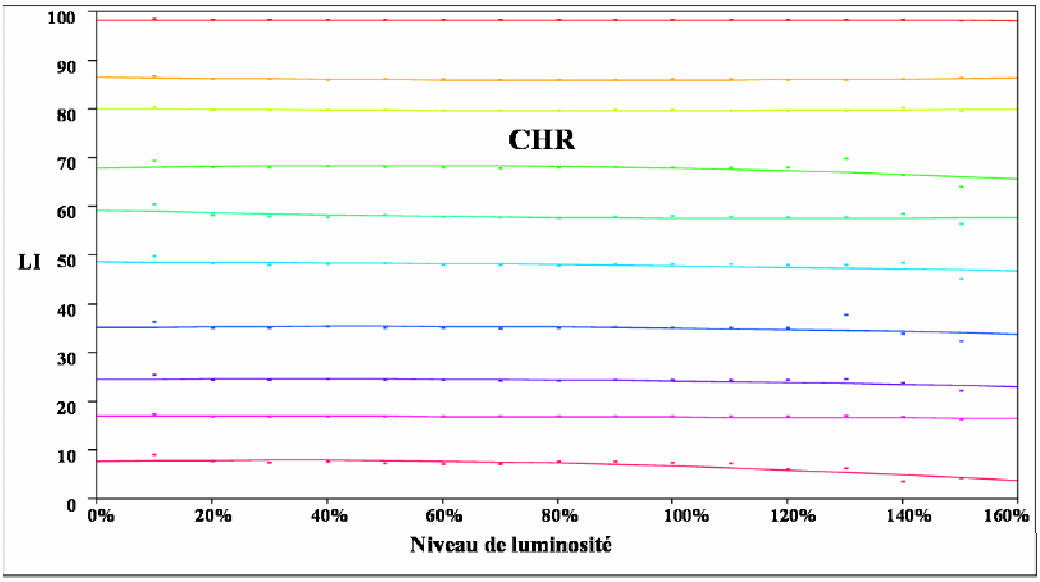


Figure 7.60: Classification CHR des images histologiques sous des variations de luminosité. Indices de marquage générés par la classification CHR de 10 images histologiques sous différents niveaux simulés de luminosité. Le niveau de luminosité 100% est le niveau initial des images histologiques.

La classification CD reflète également une stabilité apparente, même si les indices de marquage ont tendance à baisser vers les luminosités supérieures à la normale (Figure 7.61). Nous pouvons remarquer que la classification CD reste stable avec les baisses de luminosité. Par contre, elle devient sensible avec la progression de la luminosité. Mais, globalement cette sensibilité n'est pas très importante. Nous remarquons une baisse légère mais soutenue des indices de marquage avec l'augmentation de la luminosité.

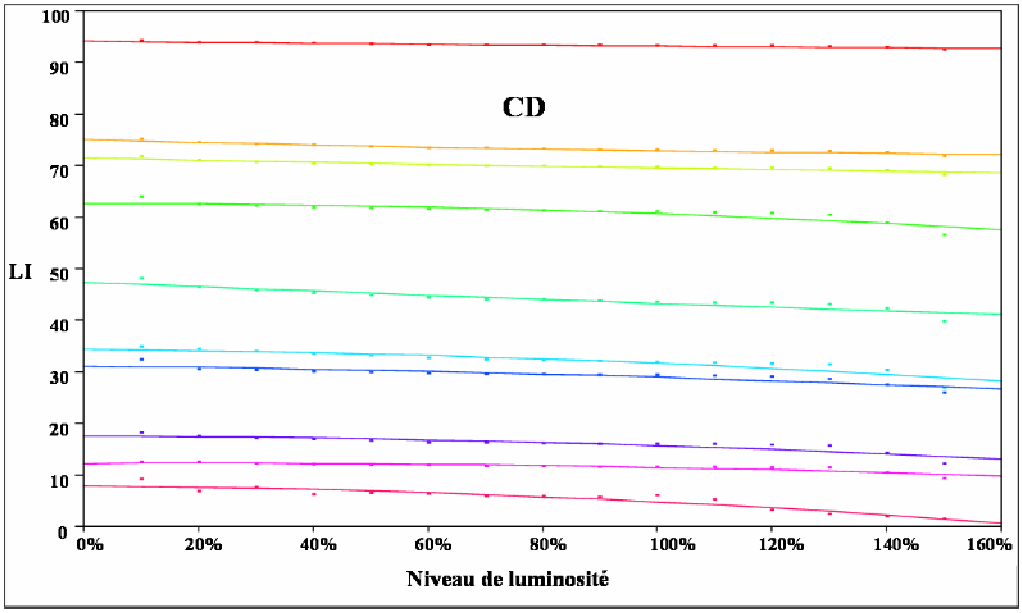


Figure 7.61: Classification CD des images histologiques sous des variations de luminosité. Indices de marquage générés par la classification CD de 10 images histologiques sous différents niveaux simulés de luminosité. Le niveau de luminosité 100% est le niveau initial des images histologiques.

La classification BN tend à sous-évaluer les indices de marquage, dans les basses et hautes luminosités à la fois (Figure 7.62). Les indices de marquage générés par cette méthode ont tendance à baisser dès les premières variations du niveau de luminosité.

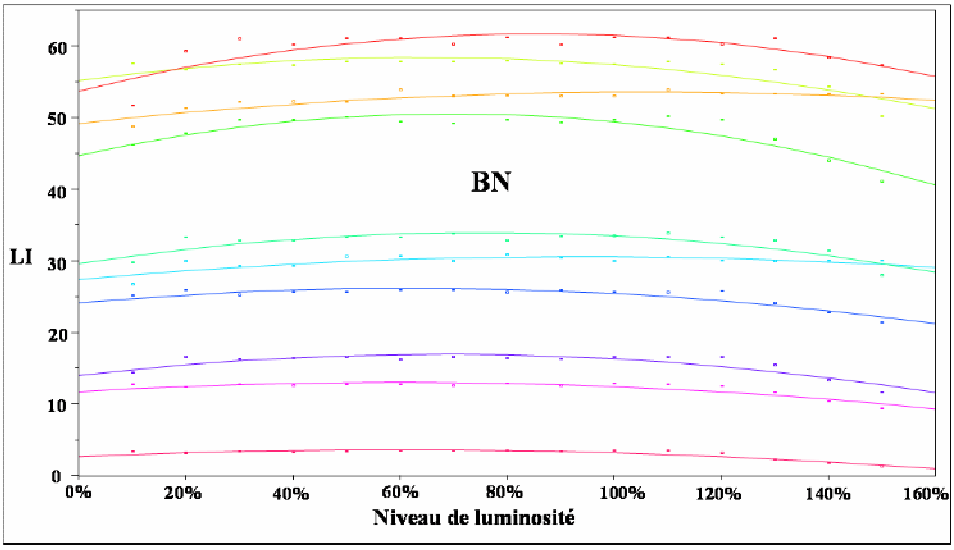


Figure 7.62: Classification BN des images histologiques sous des variations de luminosité. Indices de marquage générés par la classification BN de 10 images histologiques sous différents niveaux simulés de luminosité. Le niveau de luminosité 100% est le niveau initial des images histologiques.

La classification HUE se montre particulièrement sensible aux variations de luminosité (Figure 7.63). Le niveau de marquage se maintient stable dans un intervalle restreint de part et d'autre du niveau normale de lumière, puis il augmente très vite, et devient très élevé pour les niveaux extrêmes de luminosité (bas et hauts). Ce constat est particulièrement valable, pour les images ayant initialement de faibles indices de marquage. Ces observations mettent le doigt en effet, sur la sensibilité singulière de la teinte aux niveaux très bas et très hauts de luminosité.

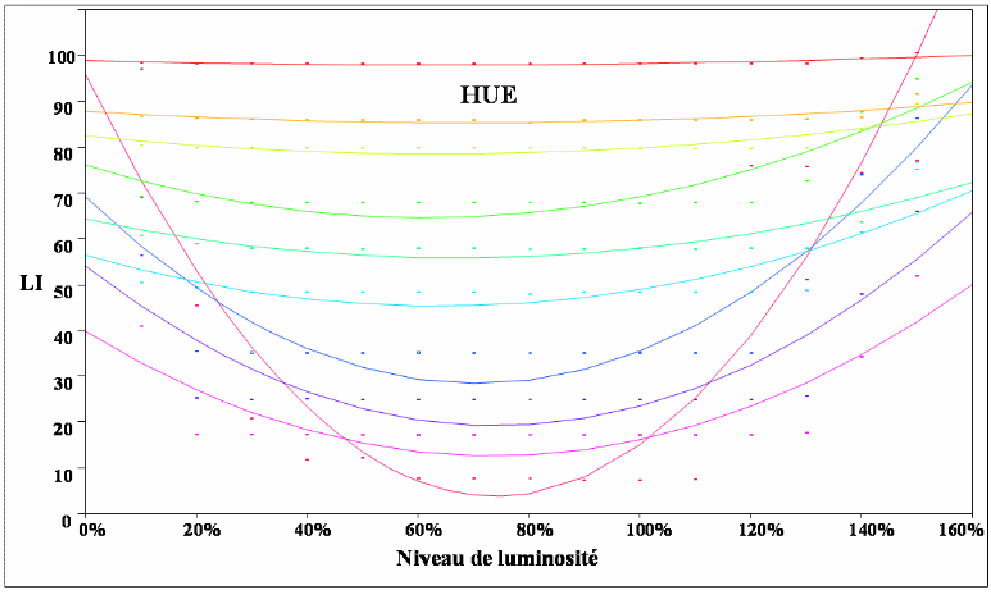


Figure 7.63: Classification HUE des images histologiques sous des variations de luminosité. Indices de marquage générés par la classification HUE de 10 images histologiques sous différents niveaux simulés de luminosité. Le niveau de luminosité 100% est le niveau initial des images histologiques.

La Figure 7.64 regroupe des erreurs moyennes de classification, spécifiques à chaque niveau de luminance pour les quatre types de classification. La classification CHR s'avère la méthode la plus robuste : elle garde un niveau stable de classification (Figure 7.64). Elle ne génère pratiquement pas d'écart de classification, excepté, pour les niveaux trop élevés de luminosité. Même à ce niveau, les écarts de classification demeurent faibles. Les classifications BN et CD démontrent aussi des bonnes performances dans une large gamme de niveaux de lumière. Toutefois, ces deux classifications génèrent des erreurs de classification en réponse à des fortes variations de luminosité (vers le bas et vers le haut).

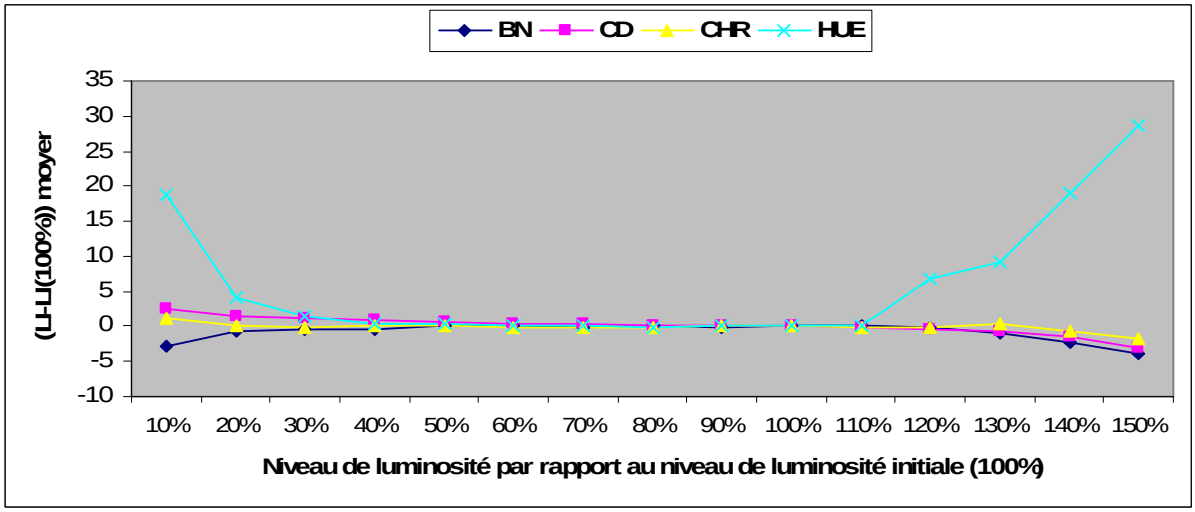


Figure 7.64: Robustesse des classifications automatiques face aux variations de luminosité. Ecarts moyens des indices de marquage générés par les classifications automatiques, par rapport à la classification automatique sous le niveau initial de luminosité (100%).

La classification HUE se montre la méthode générant le plus d'erreurs de classification (Figure 7.64). Elle est à la fois sensible aux faibles et forts niveaux de lumière. Dés que le niveau de luminosité s'écarte des conditions normales d'éclairage, la classification HUE tend à surestimer l'indice de marquage. Les erreurs d'estimation du niveau de marquage deviennent très importantes vers les niveaux extrêmes de luminosité. Toutefois, la classification HUE semble plus sensible aux fortes luminosités par rapport aux faibles luminosités. Les explications d'une telle réaction de la classification HUE peuvent trouver leur origine dans la nature même de la teinte. En effet, la teinte a la spécificité d'être particulièrement instable aux bas niveaux de luminosité et de saturation. Il parait logique de voir une performance aussi erronée de la classification HUE lorsque la luminosité est très basse. D'un autre côté, et du fait des spécificités des couleurs, l'augmentation du niveau de luminosité entraîne la diminution de la saturation des couleurs. Les fortes luminosités ne peuvent se produire que pour les couleurs faiblement saturées. Par ailleurs, les couleurs les plus saturées ne s'expriment qu'aux bas niveaux de luminosité. C'est ce phénomène qui semble le plus plausible pour expliquer les écarts de classification aussi importants au niveau des luminosités les plus importantes. Ces dernières s'accompagnant par voie de conséquence des saturations les plus faibles. Ces explications corroborent les remarques rapportées par

plusieurs auteurs sur les spécificités de la teinte face aux bas niveaux de luminosité et de saturation [Trémeau et al 2004] [Angulo 2003] [Ruifrok et al 2003] [Pujas 1996].

Tableau 7.7: Robustesse des classifications automatiques face aux variations de luminosité. Erreurs moyennes (± écart type) des indices de marquage générés par les classifications automatiques par rapport à la classification sous des conditions optimales de luminosité (100%).

	BN	CD	CHR	HUE
Erreur moyenne	0,8958	0,9096	0,3548	5,9666
Ecart type moyen	0,7696	0,5513	0,3429	8,8436

En récapitulatif, la classification CHR commet une erreur moyenne de 0,3548 ± 0,3429 par rapport à la classification des images histologiques sous des conditions normales de luminosité (100%) (Tableau 7.7). Cet écart de marquage semble très insignifiant, il nous révèle en effet, la grande stabilité de notre approche de classification CHR face aux variations de luminosité, même extrêmes. Notre méthode devance les autres méthodes. Elle génère une erreur de classification pratiquement 3 fois moins importante que celles produites par les classifications BN (0,8958 ± 0,7696) et CD (0,9096 ± 0,3429). Enfin, la classification HUE démontre son instabilité face aux variations de luminosité, elle s'écarte en moyenne de presque 6% (5,9666 ± 8,8436) par rapport au niveau de marquage initial. Ce taux d'erreur est extrêmement important comparé aux erreurs engendrées par les autres classifications. Il dépasse par exemple, 15 fois le taux d'erreur généré par la classification CHR. Ce chiffre traduit la sensibilité singulière de la teinte face aux variations de luminosité (Tableau 7.7).

7.7 Modélisation du système colorimétrique DAB-Hématoxyline : Discussion

En analyse automatique des images histologiques, la reconnaissance des molécules d'intérêts occupe une place capitale [Lucchese et al Mitra 2001]. La couleur est le critère essentiel sur lequel repose l'attribution ou non des pixels de l'image aux molécules recherchées [Poston et Gall 1990]. Si un seul colorant est appliqué au niveau des coupes histologiques, la classification des pixels devient relativement aisée. Elle se base sur un simple seuillage monochromatique, en particulier avec la présence d'un fond d'image clair. Cependant, les techniques d'IHC utilisent généralement deux colorants différents. Le premier pour mettre en

évidence la molécule recherchée, le deuxième comme contre-colorant pour teinter le reste. Dans ce cas de figure, le recours au seuillage monochromatique pour réaliser la classification devient insuffisant. La solution peut être apportée par :

- soit le choix de colorants ayant des spectres d'absorption qui ne se recouvrent pas dans le spectre du visible, ce qui faciliterait leur séparation. Néanmoins, l'utilisation de tels colorants est limitée en IHC [Lehr etal 1999] [Kai et al 1999]. De plus, des mixtures couleurs peuvent se produire lors de l'application de deux colorants même spectralement distants. Ce phénomène peut gêner la reconnaissance des molécules cibles [Van-Der-Laak et al 2000],
- soit l'utilisation des filtres avec des bandes passantes spécifiques aux longueurs d'ondes des colorants. Le défaut de cette technique, est qu'elle ne prend en considération qu'une partie du spectre d'absorption des colorants à mesurer. Et même au niveau de la bande passante utilisée, des recouvrements spectraux peuvent avoir lieu entre les colorants,
- soit le recours à l'imagerie spectrale des coupes histologiques. Cette technique possède l'avantage de fournir le spectre d'absorption lié à chaque pixel. De ce fait, elle rend plus aisée la distinction, sur des critères spectraux, entre deux colorants. Cependant, les techniques d'imagerie spectrale ne sont pas à la portée de tous les laboratoires. Elles nécessitent un appareillage extrêmement onéreux. En effet, la plupart des stations d'acquisition d'images histologiques sont équipées de caméras tri-CCD RGB. Le coût abordable de ce type de caméras a contribué à leur large diffusion.

L'espace couleur de stockage des images histologiques est l'espace RGB. En toute logique, c'est le système RGB qui est le plus exploité pour réaliser la classification des images couleur. La manière la plus précise pour séparer deux colorants est la prise en compte de leur répartition 3D au niveau de l'espace RGB : chaque colorant peut occuper un sous-espace particulier du système RGB. La classification passe par la délimitation de l'enveloppe 3D associée à chaque colorant. Mais, cette approche fait appel à des opérations de calcul complexes, ce qui accroît considérablement le temps alloué à la classification [Sarabi et Aggarwal 1981] [Fermin et Degraw 1995].

Afin de réduire la complexité liée à l'utilisation d'un espace 3D, plusieurs auteurs ont eu recours aux plans marginaux pour réaliser la classification des images histologiques tels que le plan bleu [King et al 2002] ou le plan vert [Smejkal et Shainoff 1997]. Or, au cours de l'analyse de la distribution du DAB et de l'Hématoxyline au niveau des trois plans couleur : R, G et B (Figure 7.18, § 7.3.2.2.2), nous avons démontré la carence de ces plans à réaliser une séparation correcte entre le marquage et la contre-coloration. Aucun seuillage ne semble assurer cette séparation du fait du recouvrement du DAB et de l'Hématoxyline au niveau des trois plans R, G et B.

D'autres techniques combinent les plans R, G et B pour former de nouvelles composantes qui seraient plus appropriées à séparer le DAB de l'Hématoxyline. Nous pouvons citer à titres d'exemples : la composante bleue normalisée (BN : B/R+G+B) [Brey et al 2003], la composante (G/B) [Montironi et al 1996], la composante Brown (Brown=B-0.3(R+G)) [Ruifrok 1997] ou la composante marron (brune) (marron= magenta+jaune+rouge) [Raneffall 1998]. A travers la composante BN, préconisée par Brey et al [Brey et al 2003], nous avons démontré les limites de ce type d'approche à accomplir une classification performante. L'application de la classification BN sur des images histologiques a dévoilé l'insuffisance d'une telle approche à produire des classifications conformes aux classifications manuelles (intervalle de confiance 95%). Elle produit en effet un score de marquage significativement plus faible que les classifications manuelles (§7.6.2.2.1). Elle a tendance à attribuer des pixels DAB à l'Hématoxyline. D'ailleurs, ce point est souligné par le concepteur de cette approche [Brey et al 2003] : il note que les objets sombres sont attribués le plus souvent à l'Hématoxyline même s'ils sont des objets DAB. Nous avons pu confirmer ce constat en remarquant que les indices de marquage ont tendance à baisser quand on diminue la luminosité d'une image (§7.6.3.2). Ceci prouve que cette classification attribue effectivement des pixels DAB à l'Hématoxyline dans les basses luminosités et provoque ainsi la baisse de l'indice de marquage.

Au lieu d'avoir recours à un seul plan, une autre approche utilise à la place deux plans couleurs, un plan pour chaque révélateur [Sont et al 2003] [Elie et al 2003]. Dans cette catégorie, nous pouvons citer l'approche originale de Ruifrok et Johnson [Ruifrok et Johnson

2001] qui consiste à projeter l'information couleur sur deux axes : un, représentant le marquage et l'autre, la contre-coloration. Le but de leur approche était de mesurer la contribution à la fois du marquage et de la contre-coloration au niveau de chaque pixel couleur. Nous avons repris leur travail pour réaliser la méthode de classification CD qui affilie le pixel au colorant qui contribue le plus à la constitution de la couleur de ce pixel (§7.4.2.3). Les résultats obtenus, sur la classification des images histologiques, ont démontré une stabilité générale de cette méthode face aux variations modérées des conditions de luminosité (§7.6.2.2.1). Cependant, et d'un point de vue statistique (intervalle de confiance de 95%), cette méthode n'est pas conforme aux classifications manuelles (§7.6.2.2.1.2).

Dans le modèle RGB, le mélange de l'information chromatique et lumineuse gène la reconnaissance de la coloration liée au marquage. En plus, dans le domaine de l'imagerie microscopique, et du fait de la nature des sources de lumière et du système du trajet optique, associées au dispositif d'acquisition d'images, la lumière n'est pas répartie de façon homogène à travers la surface des lames histologiques à numériser. Des solutions logicielles existent pour réduire cette fluctuation, telles que « La balance des blancs » qui normalise l'image des lames par l'image d'un champ vide. Cependant, le succès d'une telle opération est partiel : une partie de cette variation subsiste, elle est conjuguée avec le bruit du signal inhérent aux dispositifs électroniques d'acquisition d'images. A notre avis, la solution réside dans le découplage de la lumière et de l'information chromatique, surtout dans un protocole de classification couleur qui se limite à attribuer les objets au marquage ou à la contre-coloration. L'intensité lumineuse doit être intégrée dans le plan de classification, si en complément, on cherche à mesurer la quantité de colorant disposée au niveau de chaque objet.

Il existe toute une famille de systèmes couleur dissociant l'intensité lumineuse de l'information chromatique : il s'agit des systèmes couleur « luminance chrominance » (§5.3.3). Le système HLS est le système couleur le plus utilisé de cette famille pour la classification des images histologiques [Garbay et al 1981] [Fermin et Degraw 1995] [Poston 1996] [Weaver et Au 1997] [Johansson et al 2001] [Van der Laak et al 2000] [Elie et al 2003] [Ruifrok et al 2003] [Lehr et al 1997] [Lehr et al 1999]. De manière générale, les méthodes de classification couleur opérant dans ce système utilisent essentiellement la composante teinte « Hue ». En effet, c'est sur la notion de teinte que s'appuie l'œil humain pour différencier les

couleurs. Les procédures de traitement d'images manipulant la teinte essaient d'imiter le mécanisme de la vision humaine à reconnaître les couleurs. Au cours de ce chapitre, nous avons développé une procédure de classification couleur automatique, baptisée « HUE », exploitant la teinte pour séparer le marquage de la contre-coloration. La teinte a démontré des résultats satisfaisants pour la distinction entre le DAB et l'Hématoxyline. Elle exprime des scores de classification couleur très conformes aux classifications manuelles des images histologiques. Cependant, nous avons pu démontrer que les performances de la teinte se dégradaient très vite avec de moindres variations de luminosité. La classification basée sur la teinte a montré ces limites, particulièrement dans les bas niveaux de luminosité et de saturation. Ces résultats sont conformes aux constatations rapportées par plusieurs auteurs [Trémeau et al 2004] [Angulo 2003] [Ruifrok et al 2003] [Pujas 1996]. En effet, la teinte exprime un comportement singulier face aux bas niveaux de lumière et de saturation : à ces niveaux, la teinte devient très sensible aux moindres variations de saturation et de luminosité. La valeur de teinte devient même indéfinie lorsque le niveau de luminosité ou de saturation se rapproche de 0. Nous avons pu vérifier toutes ses propriétés en modifiant les niveaux de luminosité des images histologiques. Plus particulièrement, la sensibilité de la teinte dans les très basses et très hautes luminosités, en sachant que ces dernières s'expriment toujours avec les basses saturations. Dès lors, nous pouvons confirmer la fragilité de la teinte au niveau des basses saturations.

Le système HLS n'est pas le seul espace couleur de la famille « luminance chrominance » qui peut être utilisé dans le domaine du traitement des images. Afin d'éviter les problèmes soulevés par l'utilisation de la teinte, nous avons fait le choix de travailler avec le système XYZ qui est à la fois un système de primaires et un système de type « luminance chrominance », puisque Y est la composante luminance. Nous avons travaillé plus particulièrement avec le plan de chromaticité xy associé à ce système couleur. Le choix de l'espace couleur XYZ et du diagramme xy qui lui est associé est motivé par le fait qu'ils sont les représentations 3D (XYZ) et 2D (xy) de référence pour effectuer les mesures colorimétriques [Trémeau et al 2004] [Vandenbrouk 2000]. Le diagramme de chromaticité xy permet, à l'aide de simples constructions géométriques, de localiser les couleurs les unes par rapport aux autres.

Nous avons passé en revue, tout au long du chapitre 6, les principales méthodes de classification couleur. Nous avons surtout retenu qu'il n'existe pas de classification qui soit universelle et adaptée à tous les domaines de traitement d'images. Les méthodes de classification couleur les plus performantes sont celles spécifiques à un problème de traitement d'image bien précis. Conformément à cette logique, nous avons entrepris une stratégie d'analyse des images histologiques à traiter. L'analyse colorimétrique préalable des objets à classifier est la clé pour définir une classification qui soit performante en exploitant les caractéristiques couleurs de ces objets.

L'analyse du DAB et de l'Hématoxyline, au niveau du diagramme de chromaticité xy, nous a permis la modélisation du comportement chromatique associé à ces deux colorants. La phase d'analyse a débouché surtout sur la proposition de modèles colorimétriques caractérisant la disposition chromatiques du DAB et de l'Hématoxyline par rapport aux différents facteurs de variabilité tels que les niveaux de luminosité et de saturation. Les interactions colorimétriques du DAB avec l'Hématoxyline ont été prises en compte dans ces modèles, pour expliquer le phénomène de mélange couleur entre ces deux colorants.

Une stratégie de classification couleur a été développée à partir de ces modèles couleur. Le but était de définir une procédure de classification intégrant à la fois :

- les caractéristiques couleurs du DAB et de l'Hématoxyline,
- les conditions de préparation et d'acquisition d'images des lames histologiques qui peuvent faire varier les paramètres couleur liés aux DAB et à l'Hématoxyline, tels que la luminosité et la saturation,
- le mélange couleur qui peut se produire entre le DAB et l'Hématoxyline.

En suivant cette approche, nous avons construit une table de classification 2D permettant la séparation couleur entre le DAB et l'Hématoxyline. L'utilisation des plans chromatiques pour réaliser la classification des images couleur n'est pas nouvelle [Holla 1982] [Von Stein et Reimers 1983] [Shafarenko et al 1998]. La plupart de ces approches exploitent la distribution statistique des couleurs au niveau de l'image à traiter. Elles utilisent en réalité des histogrammes 2D au lieu d'exploiter réellement le plan chromatique. Notre approche, au

contraire, utilise de manière directe le diagramme de chromaticité xy. Son originalité réside dans l'intégration des spécificités couleurs des objets à classer pour formuler le protocole de classification. Elle intègre également les facteurs de variations de luminosité et de saturation qui détériorent souvent les performances des méthodes de classification couleur.

Les résultats obtenus témoignent de l'efficacité de notre approche dans le calcul des niveaux de marquage Ki67 au niveau des images histologiques. Les tests statistiques attestent de la conformité de notre méthode de classification avec les méthodes de classification manuelles (§7.6.2.2.1.2). La table de classification 2D « DAB/Hématoxyline » révèle notamment son efficacité et sa stabilité pour classer correctement les images histologiques, même dans les conditions les plus extrêmes de luminosité et/ou de saturation (§7.6.2.2). A notre avis, l'intégration du comportement chromatique du DAB et de l'Hématoxyline, en tenant compte des différents types de variations dans le plan de classification, est à l'origine de ces bonnes performances.

L'étude des profils d'expression des multiples marqueurs tumoraux, ayant un rôle potentiel dans le développement et la progression du cancer du côlon, nécessite le marquage d'un grand nombre de coupes histologiques. Le recours à des technologies haut-débits, tels que les TMA ou les stations d'acquisition de lames microscopiques virtuelles, a raccourci considérablement le temps pour préparer et numériser un tel volume de matériel biologique. C'est dans ce contexte que nous avons proposé une méthode de classification automatique, comme complément indispensable à ces technologies. L'utilisation d'une classification automatique est la solution pour traiter les masses d'images histologiques générées par de telles technologies. Comparée aux classifications manuelles, la classification automatique est rapide et objective. En effet, elle délivre des scores de marquage reproductibles, en contraste avec une classification manuelle dont le score peut changer au cours du temps (variabilité des classifications d'un observateur) (§7.6.2.1.1.1). A travers la table de classification 2D, nous avons démontré l'efficacité d'une approche automatique à classifier correctement les images histologiques. Une approche qui nous permet de travailler sous des conditions extrêmes d'éclairage, là où une classification manuelle devient impossible.

La stratégie de classification, que nous avons développée au cours de ce chapitre, se focalise sur l'information couleur des pixels en vue de leur classification. C'est une approche pixellaire, elle fait abstraction de la disposition spatiale des pixels au niveau de l'image. De ce fait, les indices de marquage issus de la classification ont été calculés en mode surface. Le pourcentage de surface marquée est corrélé avec le pourcentage des noyaux marqués [Ranefall 1998]. Ce mode de calcul a été adopté par plusieurs auteurs pour analyser l'expression des marqueurs tumoraux [Sont et al 2003] [Elie et al 2003] [Underwood et al 2001] [Rubin et al 2002]. Cependant, le comptage du pourcentage de cellules positives, exprimant un marqueur tumoral, nécessite l'addition d'une étape supplémentaire de traitement d'image à notre approche de classification. Une étape qui doit détecter chaque cellule au niveau de l'image histologique, puis déterminer si elle est positive ou négative au marquage. Cette phase peut être effectuée avec des opérateurs de morphologie mathématique ou des routines de modélisation géométrique.

7.8 Conclusion

Nous avons présenté au niveau de ce chapitre, une approche automatique de classification couleur. Une approche de classification comprenant 3 phases de traitement d'images :

- une phase d'apprentissage : pour le recueillement des caractéristiques chromatiques du marquage DAB et de la contre-coloration Hématoxyline. Cette phase a débouché sur la construction d'une table de classification couleur 2D,
- une phase de segmentation : pour isoler les noyaux cellulaires du fond de l'image,
- une phase de classification couleur : qui consiste à exploiter la table de classification 2D pour la séparation couleur du marquage et de la contre-coloration.

Notre approche de classification couleur a été illustrée par la quantification de l'expression du marqueur de prolifération cellulaire : la protéine Ki67. A travers cet exemple de marquage, nous avons confronté notre approche avec une approche de classification manuelle de référence. Notre méthode de classification a été comparée également à d'autres types de classifications automatiques de la couleur citées dans la littérature.

L’application de différents tests statistiques intégrant plusieurs paramètres de comparaison, tels que l’indice de marquage ou le pourcentage de pixels mal classés, a révélé la conformité de notre approche avec la classification de référence. Notre approche s’est montrée la classification automatique générant le moins d’erreur de classification par rapport à la classification de référence.

Nous avons pu mesurer ensuite la stabilité de notre méthode de classification par rapport aux variations des niveaux de luminosité. Nous avons démontré que c’est la méthode de classification automatique la plus robuste. Elle présente un degré de stabilité plus important que les autres classifications. Elle demeure stable même dans les conditions les plus extrêmes de luminosités, là ou la réalisation d’une classification manuelle est non pratiquable.

Ces résultats plaident pour l’adoption de notre approche de classification dans la quantification des niveaux d’expression des marqueurs tumoraux. Elle est adaptée pour analyser un grand nombre d’images histologiques dans un temps record. C’est le type d’approche nécessaire pour analyser le profil d’expression du réseau moléculaire impliqué dans la progression tumorale d’un cancer, tel que le cancer colorectal.

Chapitre 8

Sociologie cellulaire : Outil d'investigation sur la progression tumorale du cancer colorectal

Une tumeur peut être considérée comme une société de cellules où l'hétérogénéité serait une conséquence des interactions entre les cellules qui composent la population tumorale. La sociologie cellulaire est un outil de modélisation tissulaire qui permet de formaliser les relations de voisinage des cellules entre elles.

Le concept de sociologie cellulaire, par l'intermédiaire du diagramme de Voronoï, a été adopté au niveau de ce chapitre pour analyser les interactions élémentaires entre les cellules cancéreuses et les bouleversements architecturaux des tumeurs colorectales. Il s'agit de la première étude de sociologie cellulaire appliquée au cancer colorectal.

Les résultats obtenus démontrent que les remaniements tissulaires, observés au cours de la progression tumorale du cancer colorectal, sont principalement induits par la dérégulation de la disposition spatiale des cellules cancéreuses prolifératives. Les cellules hors cycle semblent participer peu à ces bouleversements tissulaires.

Aussi, les paramètres de forme et de voisinage des cellules cancéreuses ont pu être utilisés comme outil de classification entre les différents grades du cancer colorectal.

8.1 Introduction

Nous avons vu au cours du premier chapitre, que le cancer colorectal trouvait son origine dans la dérégulation de l'expression liée à une ou plusieurs molécules intervenant dans le cycle cellulaire de l'épithélium colique. Cette dérégulation, à l'échelle cellulaire, se répercute sur l'organisation fonctionnelle et spatiale de l'épithélium colique à l'échelle tissulaire. En conséquence, elle conduit à des bouleversements tissulaires de la tumeur colorectale en développement.

Nous pensons que la prise en compte des interactions élémentaires entre toutes les cellules d'un tissu, est l'élément clé pour comprendre le phénomène de désordre architectural accompagnant le tissu colique dans sa transformation tumorale. En effet, et selon une approche ascendante du bas (cellule) vers le haut (tissu), l'hétérogénéité cellulaire souvent rencontrée dans les cancers est la cause principale de ces bouleversements tissulaires.

Nous présenterons, au niveau de ce chapitre, le concept de sociologie cellulaire comme outil d'investigation sur les interactions spatiales entre les cellules cancéreuses. Ensuite, nous appliquerons ce concept sur des images de « carottes virtuelles » introduites au niveau du chapitre 3. Enfin, nous mettrons en relation les informations de sociologie cellulaire avec les remaniements de l'architecture tissulaire au cours de la progression tumorale du cancer colorectal.
Nous signalons, qu'au niveau de ce chapitre, la partie expérimentale et les interprétations des résultats ont été réalisées grâce à la collaboration du Dr. Simony [Simony 2007].

8.2 Matériel et méthodes

8.2.1 Analyse de l'architecture tissulaire : le concept de « sociologie cellulaire »

Les outils mathématiques représentés par la sociologie cellulaire nous permettent l'exploration des relations qui existent entre les fonctions des cellules et la structure de leur environnement, c'est-à-dire du tissu. Il convient donc de s'intéresser à la cellule non seulement comme entité

primordiale de la vie, mais aussi comme entité sociale incluse au sein d'un groupe (le tissu) [Chandebois 1977].

La sociologie cellulaire est un outil de modélisation tissulaire qui permet de formaliser les relations de voisinage des cellules entre elles. La sociologie cellulaire repose sur la théorie des graphes dont la caractéristique principale est la détermination des jeux de points de voisinage qui permettent de représenter l'espace occupé par les cellules autour des centres de gravité des cellules, matérialisés par les centres de leurs noyaux [Sudbo et al 2000a et 2000b] [Marcelpoil et Usson 1992]. Le modèle le plus utilisé dans le domaine de la sociologie cellulaire est le diagramme de Voronoï.

8.2.1.1 Diagramme de Voronoï

Le diagramme de Voronoï repose sur la définition la plus informative du voisinage [Toussaint 1980]. C'est une définition du type « plus proche voisin » : chaque point (ou germe) est associé non pas à son plus proche voisin, mais à la région de l'espace qui lui est la plus proche (Figure 8.1). Dans notre approche, chaque point, matérialisé par le centre de gravité d'un noyau cellulaire, est associé à la région de l'espace qui lui est la plus proche : appelée polygone de Voronoï.

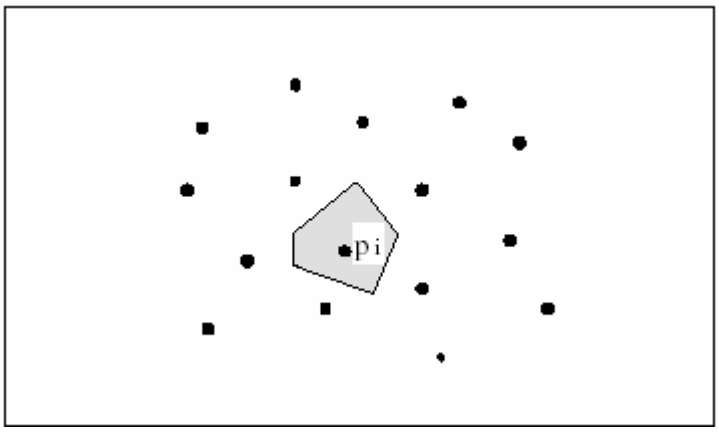


Figure 8.1: Principe du Diagramme de Voronoï. Le polygone de Voronoï du point pi est matérialisé par la région grisée polygonale et convexe, il représente la zone d'influence du point pi.

En termes mathématiques : soit *S* un ensemble de *n* points (germes) du plan. Pour chaque point *p* de *S*, la cellule ou le polygone de Voronoï *V(p)* de *p* est l'ensemble des points de l'espace qui sont plus proches de *p* que de tous les autres points de *S*. Le diagramme de Voronoï de *V(S)* est une partition du plan en un ensemble de polygones de Voronoï, où chaque polygone délimite la zone d'influence d'un point (germe).

8.2.1.2 Application à l'étude de la transformation tumorale

Une tumeur peut être considérée comme une société de cellules où l'hétérogénéité serait une conséquence des interactions entre les cellules qui la composent. L'environnement d'une cellule peut alors être assimilé à un lieu organisé par les échanges élémentaires entre membres de la population tumorale, en termes de communications mais aussi de régulation des comportements et des échanges.

La première étape de l'analyse de la sociologie cellulaire consiste à obtenir la position des centres de gravité des noyaux cellulaires (les germes) à partir d'une lame histologique. En deuxième temps, et suite à la mise en place des germes, l'application de l'algorithme du diagramme de Voronoï permet de simuler l'espace occupé par chaque cellule (Figure 8.2).

Notons que la délimitation des germes doit parfois se focaliser sur un type cellulaire en particulier. Par exemple, dans le cas du cancer colorectal, les germes ne doivent être posés que sur des cellules épithéliales, car c'est l'altération de l'épithélium colique qui est à l'origine de la cancérogénèse.

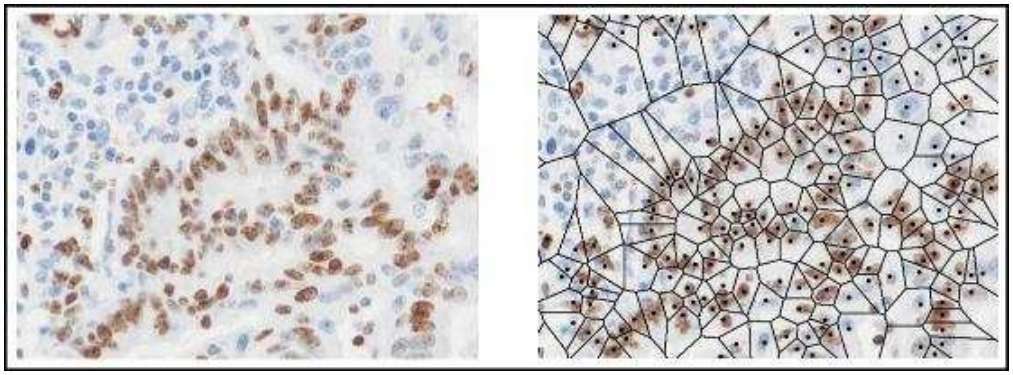


Figure 8.2: Exemple d'application du diagramme de Voronoï en histologie. Diagramme de Voronoï (à droite) associé aux noyaux cellulaires épithéliaux, d'une coupe histologique Ki67 (à gauche).

La construction des graphes permet d'extraire des paramètres décrivant l'ordre et le désordre dans les structures biologiques [Hopkins et Skellam 1954] [Dussert et al 1986] [Tezuka et al 1990] [Carvajal-Gonzales et al 1992] [Kayser et al 1993] [Raymond et al 1993]. Ainsi, le degré de désordre de l'architecture tissulaire peut être évalué lors de la transformation tumorale, en se basant sur les renseignements fournis par la sociologie cellulaire.

Le concept de sociologie cellulaire, par l'intermédiaire du diagramme de Voronoï, a été appliqué sur un ensemble d'images de carottes virtuelles révélant le marqueur de la prolifération cellulaire Ki67. L'approche a été d'analyser la disposition spatiale des cellules, en cycle cellulaire (Ki67+) et hors cycle cellulaire (Ki67-), et son influence sur les paramètres d'architecture tissulaire dans différents stades de cancérisation colorectale.

8.2.2 Matériel biologique

L'étude de sociologie cellulaire a porté sur un groupe de 84 images de « carottes virtuelles ». Ces carottes ou spots virtuels correspondent à des images focales prises de façon virtuelle (§ 3.2.7.1). Ainsi, ont été générées :

- 22 carottes virtuelles au niveau de muqueuses non cancéreuses adjacentes à la tumeur,
- 41 carottes virtuelles dans des cancers de grade 1 et 2 comportant des cellules cancéreuses malignes moyennement et bien différenciées, et
- 21 carottes virtuelles au niveau de zones de cancers de grade 3 exprimant des cellules malignes peu différenciées.

8.2.3 Procédures de traitement d'images

Toutes les procédures de traitement d'images, présentées au niveau de ce chapitre, ont été programmées sous le langage JAVA, sous forme de plugins sous le logiciel de traitement d'images imageJ. La pose de germes, pour construire le diagramme de Voronoï, a fait l'objet de la programmation d'une interface graphique pour une délimitation semi-automatique des noyaux cellulaires. Une procédure de traitement d'images a été développée en vue d'automatiser cette phase de pose de germes basée sur la détection des noyaux cellulaires.

Le diagramme de Voronoï est construit grâce à l'adaptation d'une applet java proposée par L. Paul Chew (chew@cs.cornell.edu version 2005). Il s'agit d'une applet, qui prend en entrée les coordonnées des germes, et fournit en sortie la partition de l'image en polygones de Voronoï. Cette applet a été modifiée et réadaptée pour fonctionner en mode plugins sous imageJ. Des modules supplémentaires lui ont été rajoutés sous forme de classes java pour l'extraction de différents paramètres à partir du diagramme de Voronoï.

8.2.4 Positionnement des germes pour construire le diagramme de Voronoï

L'étape primordiale pour construire le diagramme de Voronoï est le positionnement des germes. Dans le cas du cancer colorectal, ces germes doivent être posés exclusivement sur les cellules épithéliales : les seules concernées par le phénomène de cancérisation. La phase de pose doit ignorer les autres types de cellules, telles que les cellules du tissu conjonctif de soutien ou les lymphocytes, deux autres types cellulaires souvent rencontrés sur les coupes histologiques du côlon. Nous présenterons dans ce qui suit, une approche automatique pour extraire les germes des images de « carottes virtuelles ».

8.2.4.1 Génération des germes en mode automatique

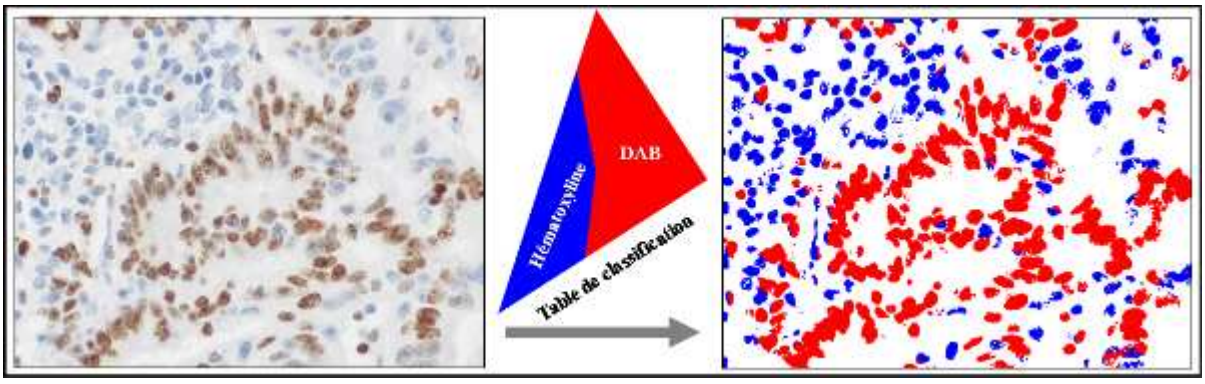


Figure 8.3: Mise en application de la table de classification couleur. Classification couleur de l'image histologique à l'aide de la table de classification couleur : les pixels DAB (Ki67) sont représentés en rouge et les pixels Hématoxyline de la contre-coloration sont représentés en bleu.

La détection automatique des germes comporte plusieurs étapes. En premier lieu, nous avons mis en oeuvre la table de classification couleur « DAB-Hématoxyline » développée dans le

chapitre précédent (§7.3.2.2.3), pour effectuer la classification couleur des images des carottes virtuelles. Ainsi, les surfaces positives au marqueur Ki67 (DAB) sont colorées en rouge et les structures négatives au marquage (cellules hors cycle) sont contre-colorées en bleu (Figure 8.3).

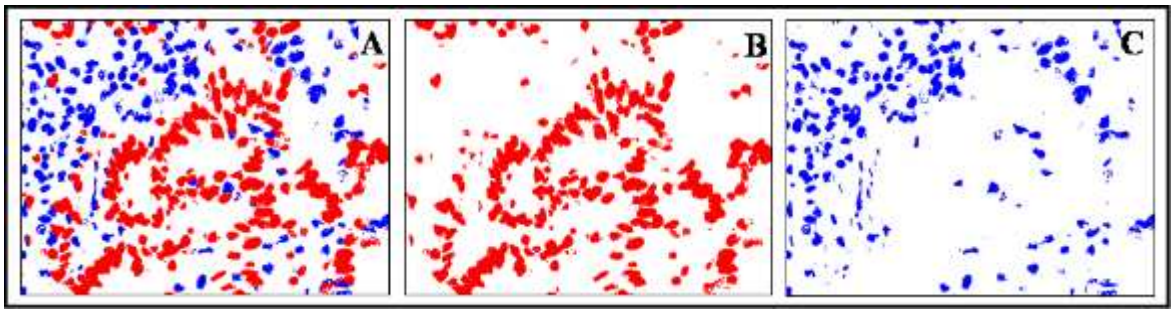


Figure 8.4: Génération des images (Ki67+) et (Ki67-). A : image classifiée**. B :** image des cellules positives Ki67+ (rouge). **C :** image des cellules négatives Ki67- (bleu).

Ensuite, nous avons découplé les cellules positives des cellules négatives : chaque groupe cellulaire a été reporté sur une nouvelle image (Figure 8.4). Le but de cette phase est de dissocier la détection des germes positifs et des germes négatifs.

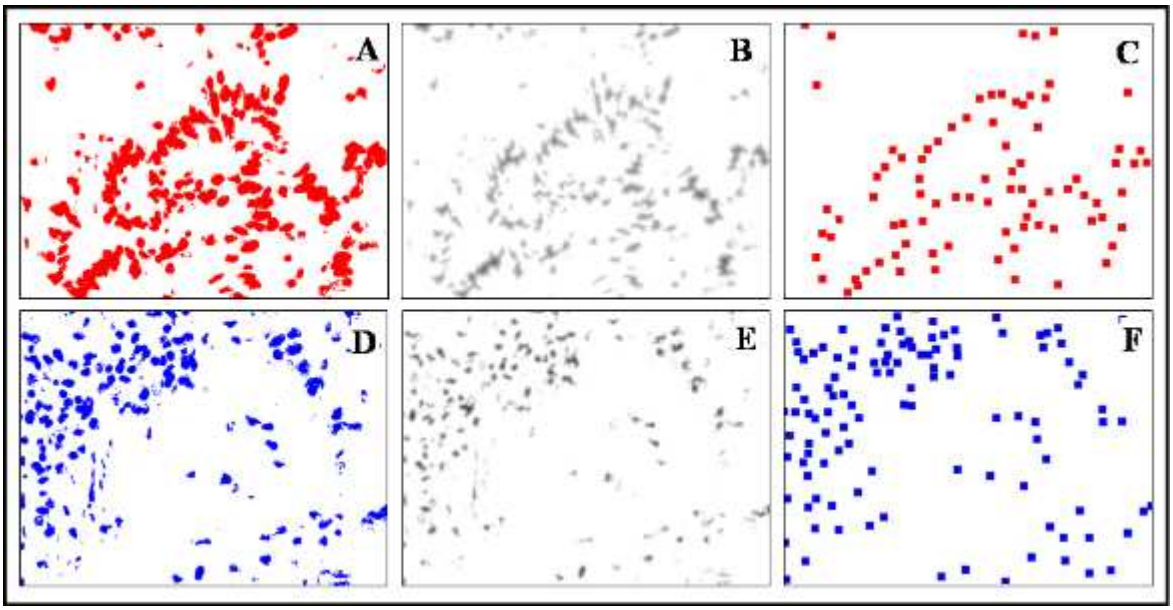


Figure 8.5: Délimitation des germes sur les noyaux cellulaires. A : image binaire des noyaux positifs (Ki67+). **B :** la carte de distance des noyaux positifs**. C :** germes positionnés sur les noyaux positifs**. D :** image binaire des noyaux négatifs (Ki67-). **E :** la carte de distance des noyaux négatifs**. C :** germes positionnés sur les noyaux négatifs.

Nous avons ensuite calculé les cartes de distances (Figure 8.5.B et Figure 8.5.E), associées aux images des cellules positives d'une part (figure 8.5.A) et des cellules négatives d'autre part (Figure 8.5.D). Une carte de distance consiste simplement à calculer la distance de chaque pixel, appartenant à la surface objet (noyaux), par rapport au fond de l'image (surface blanche sur les Figures Figure 8.5.A et Figure 8.5.D). Cette technique est très adaptée pour détecter les centres de gravité des objets circulaires, tels que les noyaux cellulaires. Les pixels qui se localisent au centre des noyaux, c'est-à-dire les plus éloignés du fond de l'image, sont ceux qui possèdent les valeurs de distance les plus élevées. Ils coïncident, en effet, avec les maxima locaux de la carte de distance (Figure 8.5.C et Figure 8.5.F). Ainsi, nous avons posé nos germes sur les emplacements de ces maxima locaux.

Signalons, que le recours aux maxima locaux comme germes conduit le plus souvent à positionner plusieurs germes sur un même noyau. En effet, les noyaux qui sont étirés ne présentant pas la forme d'un cercle ou ayant des petits trous, contiennent parfois plusieurs maxima locaux. Pour cette raison, nous avons pratiqué une dilatation de ces maxima locaux avec cinq itérations, afin que les germes proches, appartenant à un même noyau, fusionnent.

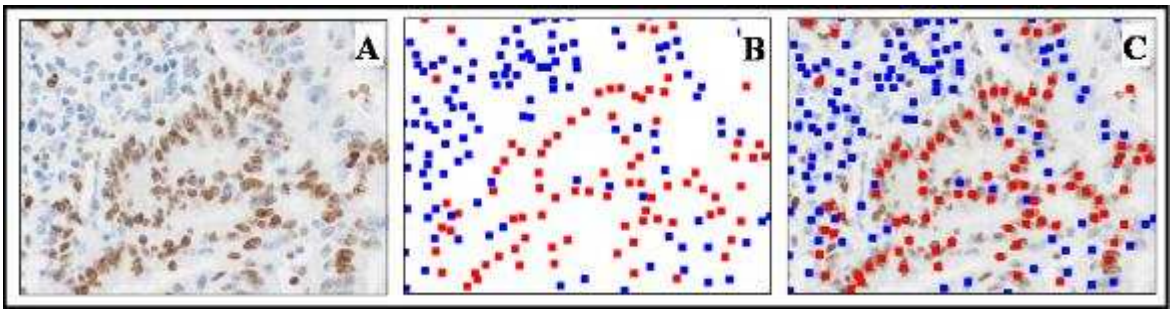


Figure 8.6: Pose en mode automatique des germes sur les noyaux cellulaires. A : image histologique**. B :** images des germes positifs et négatifs**. C :** superposition à l'image histologique des germes représentant les centres de gravité des noyaux**.**

La Figure 8.6 illustre un exemple de détection automatique des germes. Nous remarquons que les cellules sont correctement détectées à part quelques noyaux cellulaires très clairs (sur le bord supérieur droit de la Figure 8.6.C). Cependant, notre stratégie ne fait pas de différence entre les noyaux des cellules épithéliales et ceux du stroma cellulaire (tissu conjonctif de soutien bordant l'épithélium). On remarque sur le coté gauche de la Figure 8.6.C que les

germes sont également posés sur des cellules du tissu conjonctif. Nous savons que le cancer colorectal concerne essentiellement les cellules épithéliales, elles seules doivent donc être prises en considération pour la construction du diagramme de Voronoï.

8.2.4.2 Génération des germes en mode semi-automatique

A l'état actuel, nous continuons nos travaux pour écarter les cellules non épithéliales lors de la phase de détection des germes. Nous travaillons actuellement sur des procédures de traitements d'images, basées sur des critères de forme, de taille et de position pour éliminer les germes non épithéliaux. Comme solution temporaire, dans l'attente d'une solution en mode complètement automatique, nous avons conçu une interface graphique qui permet à l'utilisateur de reprendre l'image de germes superposés sur l'image histologique. A l'aide d'outils simples, tels qu'un cercle ou un polygone, l'utilisateur peut alors entourer un groupe de germes pour l'éliminer.

Pour le travail exposé au niveau de ce chapitre, antérieur aux améliorations présentées dans la section précédente (§ 8.2.4.1), nous avons eu recours à une alternative semi-automatique pour poser les germes. En effet, nous avons développé une interface graphique sous imageJ pour que le pathologiste (Dr. Simony) puisse poser les germes sur les cellules épithéliales de façon manuelle. Cette interface permet aussi au pathologiste de revisiter les images déjà traitées pour rajouter de nouveaux germes ou en supprimer d'autres mal posés.

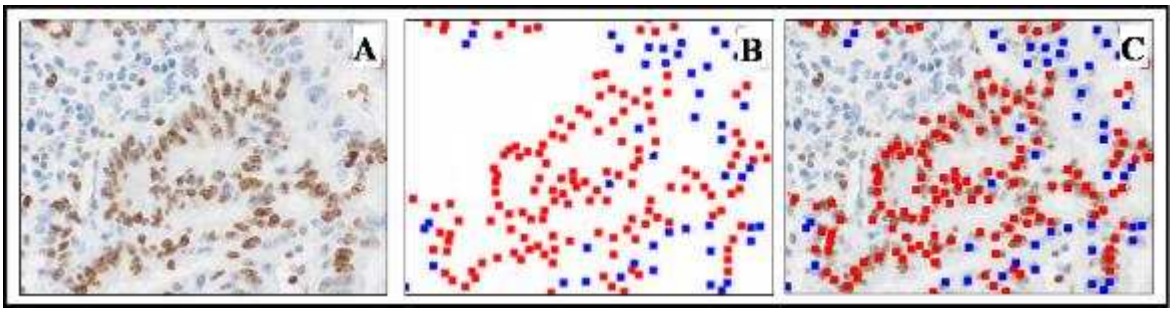


Figure 8.7: Pose en mode manuel des germes sur les noyaux cellulaires. A : image histologique**. B :** image des germes positifs et négatifs**. C :** superposition à l'image histologique des germes représentant les centres de gravité des noyaux**.**

La Figure 8.7 illustre le résultat de cette détection supervisée (pathologiste) des germes. En opposition à la Figure 8.6 (mode automatique), nous remarquons que les noyaux du tissu conjonctif, sur le tiers supérieur et gauche de l'image 8.7.C, sont ignorés par le pathologiste. Parallèlement, les germes sont posés de façon précise sur les noyaux épithéliaux, au niveau du coin supérieur droit de la Figure 8.7.C, noyaux non détectés en mode automatique.

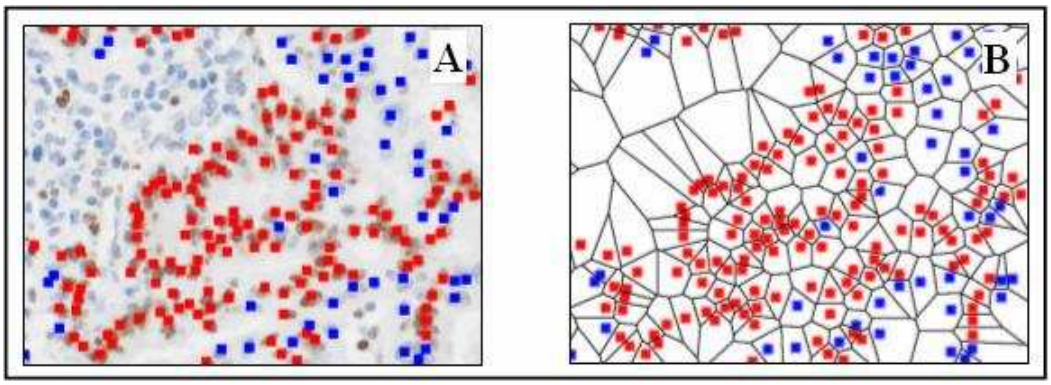


Figure 8.8: Diagramme de Voronoï associé aux germes posés sur les noyaux épithéliaux. A : image histologique d'un tissu colique (zone d'une carotte virtuelle) sur laquelle sont posés les germes par le pathologiste**. B :** partition de l'image histologique en polygones convexes par le diagramme de Voronoï.

En dépit de ces avantages, la pose manuelle des germes reste longue et fastidieuse, elle dure une heure à deux heures par « carotte virtuelle ». Cette durée est liée à la densité cellulaire des images à traiter, qui varie entre 972 et 1952 cellules par image (carotte virtuelle). Ainsi, les germes ont été positionnés sur l'ensemble des 84 images de carottes virtuelles de l'étude menée au niveau de ce chapitre. Le diagramme de Voronoï a été ensuite généré sur ces images à partir des coordonnées des germes (Figure 8.8).

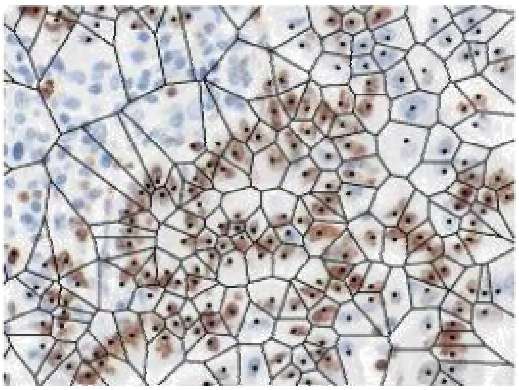


Figure 8.9: Polygones de Voronoï associés aux cellules épithéliales.

La Figure 8.9 démontre que les polygones de Voronoï sont générés exclusivement pour les noyaux épithéliaux, ce qui est une amélioration par rapport à la pose automatique de germes qui ne différencie pas entre les noyaux épithéliaux et ceux du tissu conjonctif. Cependant, ces polygones qui modélisent l'espace occupé par une cellule épithéliale, débordent parfois sur le tissu conjonctif. La figure 8.9 montre que les polygones, associés aux noyaux épithéliaux (germes) à proximité du stroma, comportent plusieurs noyaux de tissu conjonctif. Nous pensons qu'une segmentation préalable du tissu épithélial est la solution pour résoudre ce problème. Des travaux sont en cours pour générer des polygones qui soient restreints à la surface du tissu épithélial.

8.2.5 Les paramètres de sociologie cellulaire

Le diagramme de Voronoï modélise chaque noyau cellulaire à l'aide d'un polygone. Il décompose l'image histologique en une mosaïque de cellules convexes (polygones). Ainsi, par l'intermédiaire des caractéristiques locales de ces éléments, nous pouvons décrire la topographie globale des cellules peuplant une coupe histologique.

Notre stratégie est basée sur la partition de la surface tissulaire selon les polygones de Voronoï. Cette partition nous donne accès à l'analyse de la topographie tissulaire à deux niveaux : à l'échelle locale d'une cellule et son voisinage immédiat, et à l'échelle globale tissulaire décrivant la disposition générale des cellules pour occuper l'espace.

8.2.5.1 Paramètres locaux liés aux polygones de Voronoï

Chaque polygone peut être décrit par sa surface A, son périmètre L et son facteur de forme RF qui découle des deux premiers paramètres (A et L). RF peut capter la complexité de la forme des polygones, il est donné par la formule :

$$RF = \frac{4\pi * A}{L^2}$$

Le facteur de forme est maximal pour un cercle (RF=1), il décroît avec le niveau de complexité du polygone. Ainsi, un polygone étiré est associé à un facteur de forme beaucoup moins important que celui d'un polygone arrondi.

En plus de ces trois paramètres, nous avons également intégré pour chaque polygone les paramètres suivants :

- les coordonnées du germe qui lui est associé,
- le statut positif ou non du germe au marqueur Ki67,
- le nombre de polygones (cellules) voisins positifs,
- le nombre de polygones voisins négatifs,
- la distance aux polygones (distance germe-germe) voisins positifs,
- la distance aux polygones voisins négatifs.

8.2.5.2 Paramètres globaux du diagramme de Voronoï

Plusieurs paramètres peuvent s'extraire de la disposition topographique des polygones de Voronoï. Ainsi, La densité cellulaire D, d'une coupe histologique, peut être estimée à partir de la surface des polygones à l'aide de la formule :

$$D = \frac{1}{\frac{1}{N} * \sum_{i}^{N} A(i)}$$, ou N est le nombre total des polygones (cellules).

Le facteur de forme moyen (RF_m) nous renseigne sur la forme moyenne d'occupation de l'espace par les cellules. Il est donné par la formule :

$$RF_m = \frac{1}{N} * \sum_{i}^{N} RF(i)$$

D'un point de vue histologique, la topographie d'un tissu peut être analysée à l'aide de deux paramètres : la forme moyenne d'occupation de la surface tissulaire (RF_m) et le désordre intrinsèque des cellules engendré par la disposition spatiale des polygones de Voronoï. Ce désordre est décrit selon deux paramètres :

- l'homogénéité des facteurs de forme des polygones à travers le diagramme de Voronoï : RFH (Roudness Factor Homogeneity), et

- l'hétérogénéité des surfaces associées aux polygones de Voronoï : AD (Area Disorder).

Les deux paramètres sont donnés par les formules :

$$RFH = \left(1 + \frac{\sigma_{RF}}{RF_m}\right)^{-1} \quad \text{et} \quad AD = 1 - \left(1 + \frac{\sigma_A}{A_m}\right)^{-1}$$

Où σ_A est l'écart type des surfaces des polygones, σ_{RF} est l'écart type sur le facteur de forme, A_m est la surface moyenne des polygones et RF_m leur facteur de forme moyen. Les paramètres AD et RFH sont bornés sur l'intervalle [0,1]. Ces deux paramètres nous permettent de mesurer la topographie d'une population donnée, notamment la description de l'hétérogénéité d'occupation de l'espace (AD) et l'hétérogénéité des distributions locales au sein de cette population (RFH).

Notre approche est d'analyser, par l'intermédiaire des paramètres de sociologie cellulaire, la répercussion des bouleversements de l'architecture tissulaire, observés lors de la progression tumorale, sur les relations de voisinage entre les cellules cancéreuses. Nous voulons démontrer par cette analyse, à l'échelon cellulaire, le rôle des interactions locales entre cellules voisines dans la disposition générale de la topologie d'une tumeur en développement. Nous exposerons dans la partie suivante, les résultats d'application des paramètres locaux et globaux, de la sociologie cellulaire sur des images de tumeurs colorectales à différents niveaux de cancérisation.

8.3 Résultats

Avant de présenter les résultats, nous signalons que les polygones marginaux des diagrammes de Voronoï ayant un voisinage incomplet (bords des images), ont été ignorés dans cette étude pour éliminer les effets de bord.

Nous rappelons que les images des carottes virtuelles concernées par cette étude ont été regroupées selon le degré d'altération des cellules cancéreuses, en trois classes de cancer :

- tumeurs bénignes : regroupant les muqueuses normales et dysplasiques,
- tumeurs maligne grade 1-2 : ce sont des tumeurs bien et moyennement différenciées,

- tumeurs malignes grade 3 : qui correspondent aux tumeurs peu différenciées.

Les informations fournies par les paramètres de sociologie cellulaire ont été analysées à l'aide du test statistique de Kruskal-Wallis (seuil de test d'hypothèse p=0,05). Les résultats de sociologie cellulaire se focalisent sur la disposition spatiale des deux contingents de cellules prolifératives (Ki67+) ou non (Ki67-). Le but est d'analyser lequel des deux groupes cellulaires contribue le plus aux remaniements architecturaux observés dans le processus de cancérisation colorectal.

8.3.1 Disposition spatiale des cellules prolifératives selon le grade des lésions tumorales

8.3.1.1 Densité cellulaire en rapport avec le grade tumoral

La figure 8.10 illustre le diagramme de Voronoï sur des exemples de cellules épithéliales coliques à différents stades de différenciation tumorale. Nous observons que la surface des polygones augmente avec le degré de différenciation de la tumeur. Ainsi, les tailles de ces polygones ne sont pas homogènes pour l'épithélium colique normal (Figure 8.10.A). Cette taille diminue et devient assez régulière dans les lésions dysplasiques (Figure 8.10.B). Les polygones gardent leurs tailles relativement homogènes dans les cancers coliques bien différenciés (Figure 8.10.C). La Figure 8.10.D démontre que la taille des polygones augmente de manière significative dans le cas des tumeurs peu différenciées.

Cette variation dans la taille des polygones peut être interprétée par l'évolution de la densité cellulaire au cours de la transformation tumorale. De manière générale, la densité cellulaire (1/A) diminue avec le degré de différenciation tumorale. Elle augmente légèrement pendant les premières phases de cancérisation, puisqu'elle passe de 0,057 en muqueuse saine (figure 3.10.A) à 0,058 pour les muqueuses dysplasiques (Figure 8.10.B) et pour les cancers bien différenciés (Figure 8.10.C). La densité cellulaire baisse ensuite, pour passer de 0,052 pour les cancers moyennement différenciés, à 0,033 dans le cas des cancers peu différenciés (Figure 8.10.D). Ces mesures ne représentent qu'une tendance, ils n'ont cependant pas de signification statistique.

Notons que la taille disproportionnée, liée à certains polygones, affecte un peu ces résultats. Ce constat est particulièrement vérifié pour les polygones des cellules épithéliales bordant le tissu conjonctif (Figure 8.10.A et Figure 8.10.D). Ces effets ont été déjà discutés dans la section consacrée à la pose des germes (§8.2.4.2).

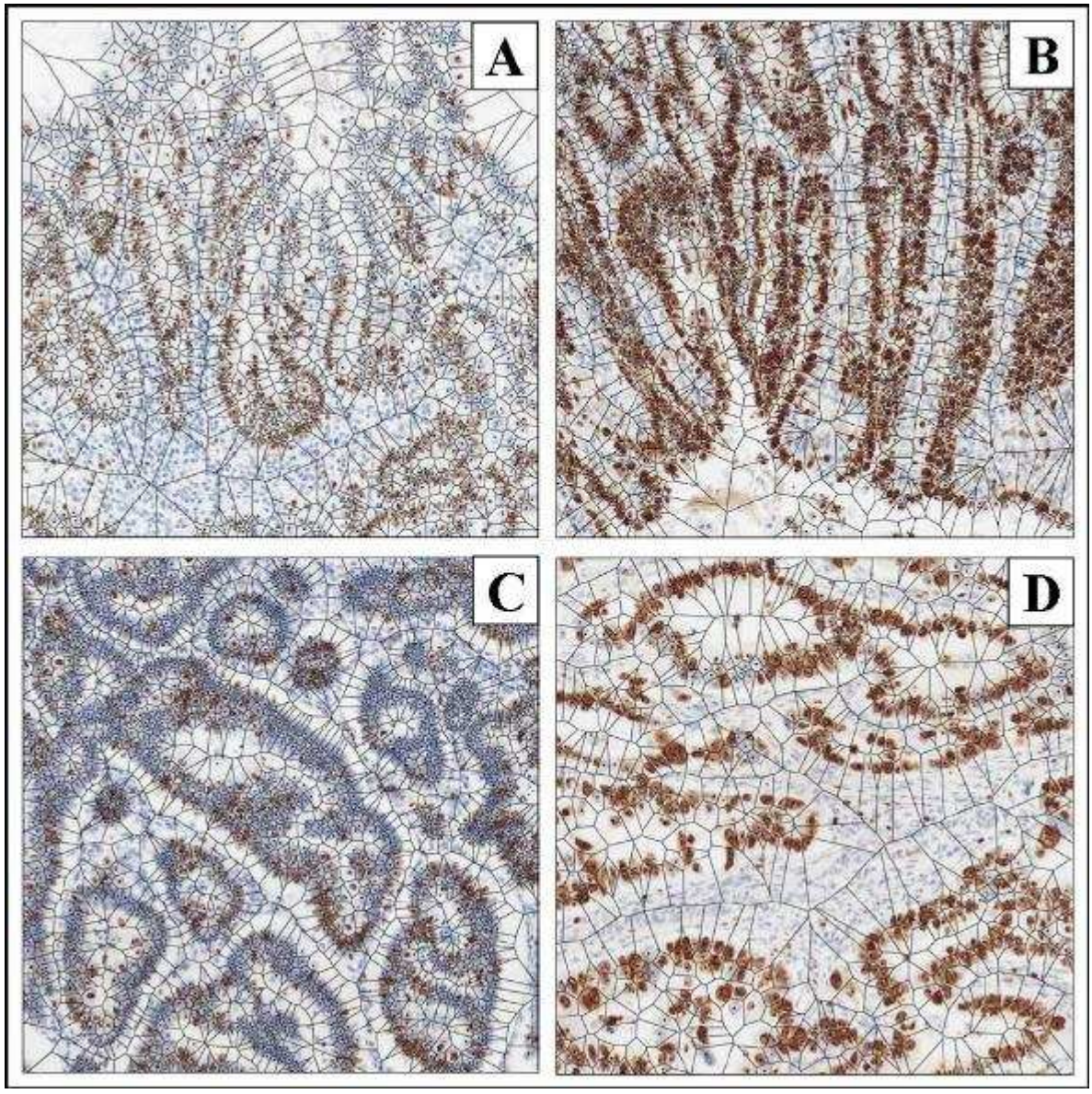


Figure 8.10: Diagramme de Voronoï de tissus coliques à différents stades de différenciation. A : Muqueuse normale**. B :** Lésion dysplasique**. C :** Tumeur bien différenciée**. D :** Cancer peu différencié.

8.3.1.2 Diminution du facteur de forme dans les hauts grades

Le facteur de forme moyen (RF_m) est relativement stable dans les lésions bénignes et les tumeurs de grade 1 et 2. Cependant, ce paramètre diminue dans les cancers de grade 3 (Tableau 8.1). Le facteur de forme moyen discrimine les bas grades de cancers (bénins et grade 1-2), des cancers de haut grade de manière significative (p=0.009). Cette diminution significative du facteur de forme moyen démontre que les cellules cancéreuses tendent à se déformer et à s'étirer lors de la progression tumorale, vers les grades les plus élevés de cancérisation.

Tableau 8.1: Facteur de forme moyen en relation avec le grade du cancer.

	grade bénin	cancer grade 1-2	cancer grade 3
RF_m (moyenne)	0,63	0,65	0,62
RF_m (médiane)	0,65	0,66	0,6

8.3.1.3 Stabilité relative du facteur d'hétérogénéité d'occupation de l'espace par les cellules :

L'occupation de la surface tissulaire par les cellules prolifératives, estimée par le facteur d'hétérogénéité des surfaces : AD, ne montre pas de différences significatives entre les trois groupes de lésions tumorales (Tableau 8.2).

Tableau 8.2: Hétérogénéité d'occupation de l'espace en relation avec le grade du cancer.

	grade bénin	cancer grade 1-2	cancer grade 3
AD (moyenne)	0,57	0,55	0,54
AD (médiane)	0,57	0,55	0,54

8.3.1.4 Diminution du facteur d'homogénéité géométrique des polygones dans les hauts grades :

Nous constatons une homogénéité relativement stable des formes géométriques des cellules dans les bas grades de cancers (bénins et grade 1-2). Le facteur RFH diminue sensiblement dans les hauts grades de cancer (Tableau 8.3).

Tableau 8.3: Homogénéité géométrique des cellules des différents grades de cancer.

	grade bénin	cancer grade 1-2	cancer grade 3
RFH (moyenne)	0,84	0,85	0,84
RFH (médiane)	0,84	0,86	0,80

Cette diminution est statistiquement significative (p=0.019) et montre que les hauts grades de cancer sont les plus désordonnés.

8.3.1.5 Augmentation de la distance médiane aux cellules positives voisines avec le grade

Le nombre des cellules voisines positives ou négatives reste stable dans les différents grades de cancer (Tableau 8.4). Ce paramètre ne permet pas une distinction statistique entre les trois groupes de cancer.

Les résultats montrent que la distance médiane aux cellules prolifératives voisines augmente de manière significative (p=0.008) avec la progression du grade tumoral. La distance médiane aux cellules voisines hors cycle cellulaire augmente aussi. Cependant, elle n'est pas statistiquement significative pour différencier les trois grades de cancer. Ces résultats démontrent bien que c'est la perturbation des relations de voisinage entre les cellules cancéreuses (prolifératives) qui est à l'origine des désordres architecturaux observés à l'échelle du tissu tumoral.

Tableau 8.4: Relation de voisinage immédiat des cellules prolifératives. Ce tableau illustre pour chaque cellule positive (+) le nombre moyen et médian de ces cellules « voisines (+) » et de ces cellules « voisines (-) ». Il illustre aussi la distance moyenne et médiane de chaque cellule (+) vers ces « voisines (+) » et vers ces « voisines (-) ».

	Nombre « voisines (+) »	Nombre « voisines (-) »	Distance « voisines (+) »	Distance « voisines (-) »
grade bénin (moyenne)	3,45	2,32	35,66	30,03
grade bénin (médiane)	4	2	36,36	29,8
cancer grade 1-2 (moyenne)	3,9	1,83	37,25	28,02
cancer grade 1-2 (médiane)	4	2	37,28	32,93
cancer grade 3 (moyenne)	3,436	2,14	42,35	32,92
cancer grade 3 (médiane)	4	2	42,8	37,4

8.3.2 Disposition spatiale des cellules hors cycle selon le grade des lésions tumorales

8.3.2.1 Topologie des cellules hors cycle

Tous les descripteurs de la topologie cellulaire liés aux cellules hors cycle sont statistiquement non significatifs pour discriminer les trois groupes de cancers colorectaux (Tableau 8.5). Les cellules hors cycle préservent les niveaux de leurs surfaces et surtout de leurs facteurs de forme (RF et RFH) stables au cours de la progression tumorale. Ils indiquent une forme de similarité des cellules hors cycles même dans les grades de cancer les plus élevés. Nous pouvons dire que la topologie des cellules hors cycle varie peu et de façon non significative au cours de la progression tumorale. Ces cellules semblent peu participer dans le bouleversement architectural d'une tumeur en développement.

Tableau 8.5: Données de topologie cellulaire associées aux cellules hors cycle (Ki67-).

	grade bénin	cancer grade 1-2	cancer grade 3	Kruskal-Wallis
RF_m (moyenne)	0,61	0,63	0,62	>0,05
RF_m (médiane)	0,63	0,65	0,6	>0,05
AD (moyenne)	0,58	0,55	0,54	>0,05
AD (médiane)	0,57	0,55	0,5	>0,05
RFH (moyenne)	0,82	0,84	0,84	>0,05
RFH (médiane)	0,83	0,84	0,8	>0,05

8.3.2.2 Rapport des cellules hors cycle avec leur entourage

Le statut des cellules voisines des cellules hors cycle, ne permet pas une distinction statistique entre les trois types de lésions tumorales (Tableau 8.6). La distance médiane des cellules hors cycle avec les cellules voisines prolifératives se montre statiquement significative entre les trois groupes de lésions (p=0.001). Cette distance médiane augmente avec le grade de la tumeur (Tableau 8.6). La distance vers les cellules hors cycle se révèle statiquement non significative pour dissocier les trois groupes de grade tumoral.

Tableau 8.6: Relation de voisinage immédiat des cellules hors cycle. Ce tableau illustre pour chaque cellule hors cycle (-) les nombres moyen et médian de ces cellules « voisines (+) » et de ces cellules « voisines (-) ». Il illustre aussi la distance moyenne et médiane de chaque cellule (-) vers ces « voisines (+) » et vers ces « voisines (-) ».

	Nombre « voisines (+) »	Nombre « voisines (-) »	Distance « voisines (+) »	Distance « voisines (-) »
grade bénin (moyenne)	1,91	3,73	28,18	31,34
grade bénin (médiane)	2	4	29,84	29,08
cancer grade 1-2 (moyenne)	2,36	3,2	34,76	32,7
cancer grade 1-2(médiane)	2	3	35,5	32,78
cancer grade 3 (moyenne)	2,25	3,52	39,53	37,26
cancer grade 3 (médiane)	2	4	40,3	37,9

8.4 Discussion

Les résultats obtenus démontrent que les remaniements de l'architecture tissulaire des cancers colorectaux s'accompagnent, à l'échelle locale, par des réorganisations topologiques des cellules cancéreuses. Les paramètres de sociologie cellulaire décrivant cette réorganisation peuvent discriminer entre les différents grades du cancer colorectal.

La progression tumorale s'accompagne dans un premier stade par une augmentation de la densité cellulaire. Puis, cette densité chute au niveau des cancers les plus différenciés. A ce stade avancé de cancérisation, les cellules malignes sont dispersées. Elles délimitent des polygones de taille plus importante. Cette dynamique topologique est la traduction du gradient de prolifération cellulaire observé dans le développement tumoral du cancer du côlon. En effet, la prolifération connaît un pic dans les grades T1 et T2 pour diminuer ensuite dans les grades plus avancés du cancer colorectal : T3-T4 [Simony 2007].

Les bouleversements de l'architecture tissulaire des tumeurs colorectales sont induits principalement par les cellules cancéreuses prolifératives, malgré l'homogénéité dans les formes géométriques des cellules (AD) prolifératives d'une même lésion lors de la progression tumorale. Nous avons observé cependant, une augmentation de la variabilité (diminution du RFH) des formes des cellules prolifératives. Cette augmentation ne s'accompagne pas d'une variation de l'occupation de l'espace. Nous pensons que ce phénomène est dû à une altération dans le mécanisme régulant l'agencement spatial des

cellules prolifératives. Ces cellules, bien engagées dans le processus de cancérisation (grade 3), perdent leur régularité de forme. Ce désordre n'est pas répercuté de façon homogène sur toute la surface tumorale. Bien qu'il soit observé à un niveau local, à l'échelon cellulaire, il participe cependant à la mise en place d'une forme d'hétérogénéité à un niveau plus global : à l'échelon tissulaire.

L'hétérogénéité tumorale est une donnée très fréquente en pathologie oncologique. Nous avons pu analyser son ampleur dans le cancer colorectal, à travers la distribution d'expression de certains marqueurs tumoraux (§ 3.3.2). Cette hétérogénéité peut être expliquée par l'accumulation des mutations au cours de l'évolution d'une tumeur conduisant à l'émergence de plusieurs clones de cellules cancéreuses. Le désordre architectural, incriminé par les paramètres de sociologie cellulaire, trouve peut être son origine dans le développement simultané de plusieurs clones de cellules cancéreuses au niveau de la tumeur. Chaque clone cellulaire adopte ses propres paramètres pour occuper l'espace tumoral, conduisant *in fini* à l'hétérogénéité de la topologie cellulaire d'une tumeur peuplée par plusieurs clones à la fois.

Les résultats sur le voisinage immédiat des cellules hors cycle confirment ceux du voisinage des cellules prolifératives. Ils démontrent que ce sont principalement les cellules prolifératives (cancéreuses) qui induisent les remaniements de l'architecture tissulaire des tumeurs. Les cellules hors cycle gardent leurs caractéristiques topologiques et leurs relations spatiales avec le voisinage (sauf avec les cellules cancéreuses) tout au long de la transformation tumorale de la muqueuse colique normale.

Ces observations sont en accord avec les résultats rapportés dans la littérature concernant l'application des paramètres de sociologie cellulaire sur différents types de cancer. Ainsi, le diagramme de Voronoï a été utilisé pour classifier les différents types d'une même lésion tumorale, dans le cas du cancer prostatique [Bartels et al 1998], pulmonaire [Bibras et al 1996] ou dans le cancer de la langue [Sudbo et al 2000c].

Des travaux sont en cours pour améliorer la mise en application de la sociologie cellulaire sur le cancer colorectal. Nos efforts se concentrent actuellement sur la segmentation de

l’épithélium colique afin de générer des polygones de Voronoï traduisant le plus fidèlement possible l’espace occupé par une cellule épithéliale.

8.5 Conclusion

Nous avons présenté au niveau de ce chapitre, une approche originale basée sur l’application des paramètres de sociologie cellulaire pour analyser la dynamique de la progression tumorale du cancer colorectal. Ce travail a été réalisé conjointement avec le Dr. Simony et il s’agit de la première étude de sociologie cellulaire menée sur le cancer colorectal.

Les paramètres de sociologie cellulaire, calculés à partir du diagramme de Voronoï, ont permis la classification des tumeurs colorectales selon le grade de différenciation des cellules cancéreuses. Ainsi, nous avons pu démontrer que les remaniements architecturaux tissulaires, observés lors de la progression tumorale, sont principalement induits par l’hétérogénéité de la distribution spatiale des cellules cancéreuses en cycle.

Pour finir, la topologie associée à la disposition spatiale des cellules cancéreuses prolifératives a permis de distinguer les tumeurs bénignes des tumeurs malignes.

Chapitre 9

Conclusions et perspectives

La cancérogenèse colorectale est un phénomène complexe faisant intervenir plusieurs altérations génétiques et moléculaires. L'appréhension du réseau moléculaire, responsable de ce type de cancer nécessite des études IHC à grandes échelles, sur un grand nombre de tumeurs colorectales. Dans ce contexte, la technologie TMA se montre une alternative séduisante, aux approches traditionnelles d'IHC. Elle permet l'analyse simultanée des profils d'expression des marqueurs tumoraux sur des centaines d'échantillons tissulaires.

L'objet de cette thèse est porté sur les divers aspects de traitement d'images liés à la technologie TMA et à son application pour l'étude de la progression tumorale dans le cancer colorectal. Le sujet de thèse s'intéresse particulièrement aux différentes problématiques soulevées par le recours à une micro-technologie à haut-débit, telle que la technologie TMA.

Dans le cadre des travaux liés à notre sujet de thèse, nous avons adopté une stratégie virtuelle basée sur le concept de « carottes virtuelles » pour l'étude des différentes contraintes scientifiques et techniques associées à l'utilisation de la technologie TMA. Ce concept a été mis en application pour le développement et l'illustration de nouvelles approches de traitement d'images dédiées à la conception, l'analyse automatique et l'exploitation des coupes TMA.

Le travail réalisé au cours de la thèse peut être organisé en trois grandes parties selon les différentes phases : la conception, l'analyse automatique et l'exploitation d'un bloc TMA.

9.1 Conception d'un bloc TMA

L'obstacle majeur pour l'adoption de la technologie TMA dans la pratique courante des laboratoires est bien la question sur la capacité des petits échantillons tissulaires prélevés à capter le potentiel d'hétérogénéité des tumeurs. Pour répondre à cette problématique, nous avons développé une approche originale basée sur le concept de « carottes virtuelles » afin de simuler le prélèvement des échantillons tissulaires (carottes) réels. Ce concept permet de tester toutes les combinaisons possibles de prélèvement des carottes réelles.

Dans ce contexte, nous avons développé un protocole d'échantillonnage aléatoire basé sur le concept de « carottes virtuelles » pour étudier l'impact de l'hétérogénéité tumorale sur l'utilisation de la technologie TMA, notamment pour l'étude du cancer colorectal. Ce protocole a été appliqué pour l'étude de l'échantillonnage d'un groupe de tumeurs colorectales en respectant la nature de la distribution spatiale de l'expression des marqueurs du cycle cellulaire.

Grâce à notre outil de simulation, nous avons démontré que le nombre de carottes à prélever doit être adapté à l'hétérogénéité du matériel tumoral à échantillonner et aux marqueurs tumoraux considérés. Si le prélèvement d'une à deux carottes apparaît suffisant pour les zones de marquage homogènes et pour certains marqueurs à distribution focale (Bcl2), en revanche les marqueurs à distributions hétérogènes nécessitent le prélèvement d'un nombre plus important d'échantillons. Le recours à trois carottes, comme utilisé par plusieurs auteurs, n'est pas adapté pour la représentation de toutes les zones à échantillonner. S'il est suffisant pour représenter les zones homogènes, il se montre inadapté dans le cas des zones hétérogènes.

Notre stratégie de simulation nous a permis de prendre en considération l'hétérogénéité tumorale liée au cancer colorectal. Si cette hétérogénéité ne remet pas en cause le recours à la technologie TMA, elle démontre cependant les limites de cette technologie pour la représentation des marqueurs tumoraux les plus hétérogènes.

9.2 Classification couleur des lames histologiques TMA

Les différents travaux abordés dans l'état de l'art sur les espaces couleur et les techniques de la segmentation couleur, ne convergent pas vers un seul système couleur et vers une seule méthode de segmentation, que l'on pourrait préconiser de manière générale pour le traitement de tout type d'images histologiques. Les techniques de classification couleur les plus performantes sont celles adaptées à un problème de traitement d'image bien spécifique.

A la lumière de ce constat, nous avons développé une nouvelle approche de classification couleur spécifique à la quantification de l'expression des marqueurs tumoraux au niveau des coupes histologiques. Cette approche a été proposée après une phase d'analyse colorimétrique, du marquage IHC et de la contre-coloration associée, au niveau du diagramme de chromaticité xy. Cette analyse colorimétrique a débouché sur la proposition des modèles couleur, caractérisant la disposition chromatique du marquage et de la contre-coloration, par rapport aux différents facteurs de variabilité, tels que les niveaux de luminosité et de saturation. Le phénomène de mélange couleur entre le marquage et la contre-coloration a été également pris en compte dans ces modèles.

Sur la base de ces modèles couleur, nous avons construit une « table de classification 2D » capable de séparer le marquage de la contre-coloration. Notre stratégie de classification utilise de manière directe le diagramme de chromaticité xy comme table de classification couleur. Son originalité réside, dans l'intégration des spécificités couleur du marquage et de la contre-coloration.

Notre approche de classification couleur a été illustrée à travers la quantification du marqueur Ki67 révélé par le système DAB (immunomarquage) / Hématoxyline (contre-coloration). La table de classification 2D a démontré son efficacité pour classifier les images histologiques. Elle s'est avérée la méthode la plus conforme aux classifications manuelles par rapport à trois méthodes de classification couleur automatique citées dans la littérature. Des méthodes qui réalisent la classification couleur selon la teinte, la composante bleue normalisée [Brey et al 2003] de l'image couleur ou selon le principe de la déconvolution couleur [Ruifrok et Johnston 2001]. Notre approche de classification s'est révélée la méthode de classification

couleur automatique la plus stable et la plus robuste, dans les conditions de luminosité les plus extrêmes, là où la classification manuelle est non praticable.

Notre méthode de classification couleur peut être utilisée dans des domaines d'application autres que la classification des images histologiques. Elle peut parfaitement être appliquée dans le domaine de la vision artificielle dans un schéma de reconnaissance de couleur, notamment dans un environnement de luminosité détériorée. Notre approche peut être intégrée dans le processus de fabrication industrielle pour la reconnaissance automatique des défauts de fabrication, lorsque les outils de contrôle qualité de la production font intervenir des protocoles basés sur l'analyse colorimétrique.

Notre stratégie de classification couleur a été illustrée à travers le couple DAB : Hématoxyline. Techniquement, la table de classification couleur 2D peut être générée à partir de n'importe quel couple de marqueur et contre-colorant. Il faut juste fournir en entrée de notre procédure de traitement d'images, les images de références positive et négative associées au couple marqueur/contre-colorant voulu. Les démonstrations développées pour la construction de la table de classification 2D à partir des modèles couleur DAB et Hématoxyline, sont valables pour tout couple marqueur/contre-colorant (§7.3.2.2.3). Les tests statistiques réalisés, pour le couple DAB : Hématoxyline valident le recours à la table de classification 2D pour l'analyse automatique de l'expression des marqueurs tumoraux mise en évidence par tout système de révélation : immunomarqueur/contre-colorant.

Nous signalons que notre stratégie d'analyse se focalise sur l'information couleur des pixels en vue de leur classification. C'est donc une approche pixellaire, elle fait abstraction de la disposition spatiale des pixels au niveau de l'image. Elle attribue les pixels au marquage ou à la contre-coloration. A l'état actuel, nous segmentons les surfaces nucléaires du fond de l'image. L'indice de marquage est calculé en mode surface, c'est-à-dire le pourcentage de pixels nucléaires exprimant le marquage. Cependant, le comptage du pourcentage de cellules positives, exprimant un marqueur tumoral dans une image histologique, nécessite l'amélioration de notre étape de segmentation pour une détection individualisée des noyaux cellulaires. La difficulté réside dans la séparation correcte des amas de noyaux rencontrés souvent dans les images histologiques. Des opérateurs de morphologie mathématique, tels que

la ligne de partage des eaux (LPE) ou les routines de modélisations géométriques, peuvent être appliqués pour séparer les noyaux accolés. Une fois les noyaux détectés, on peut les attribués au marquage ou à la contre coloration selon le pourcentage de leurs pixels exprimant le marquage (>50%).

L'approche de classification couleur proposée est basée essentiellement sur l'information chromatique pour différencier les pixels DAB des pixels Hématoxyline. C'est une approche de classification qualitative. A l'état actuel, notre approche ne permet pas d'effectuer des mesures densitométriques pour déterminer la quantité du marquage et de la contre-coloration au niveau de chaque pixel. Des perspectives sont envisagées pour améliorer notre table de classification 2D, pour qu'elle soit capable de déterminer le pourcentage de contribution du marquage et de la contre-coloration dans la constitution de la couleur associée à chaque pixel du matériel histologique classifié. Ce pourcentage peut être mis en relation avec les intensités lumineuses des pixels (composante luminance) pour mesurer la quantité de marquage et de contre-coloration au niveau de chaque pixel.

9.3 Exploitation des données générées par la technologie TMA

Nous avons présenté le concept de sociologie cellulaire comme outil d'investigation sur les interactions spatiales entre les cellules cancéreuses au cours de la progression tumorale du cancer colorectal. Ce concept a été mis en application sur des images de carottes virtuelles présentant différents stades de cancer colorectal. Notre objectif était d'étudier la répercussion des bouleversements de l'architecture tissulaire, observés lors de la progression tumorale, sur les relations de voisinage entre les cellules cancéreuses.

Les paramètres de sociologie cellulaire, calculés à partir du diagramme de Voronoï, démontrent que les remaniements de l'architecture tissulaire des cancers colorectaux s'accompagnent, par des réorganisations topologiques des cellules cancéreuses. Ainsi, cette topologie associée au désordre spatial des cellules cancéreuses a permis de distinguer entre les tumeurs bénignes et malignes.

Il s'agit de la première étude de sociologie cellulaire menée sur le cancer colorectal. Ces résultats sont préliminaires, ils doivent être confirmés par une étude à plus grande échelle. Ils devraient être corrélés avec les études de survie, afin de pouvoir introduire les paramètres de sociologie cellulaire comme des indices de pronostic.

Nous soulignons que la phase de pose de germes sur les centres de gravité des cellules, pour générer le diagramme de Voronoï, est un obstacle à l'application de notre approche en routine. Cette opération était initialement réalisée en mode semi-automatique et nécessite l'intervention d'un pathologiste pour positionner ces germes.

Nous avons démontré au cours de cette thèse, une approche pour automatiser la phase de pose de germes qui exploite et met en œuvre la table de classification 2D. L'obstacle majeur rencontré est la difficulté de restreindre la pose des germes en mode automatique sur les cellules épithéliales. Le cancer colorectal concerne essentiellement les cellules épithéliales, elles seules doivent être prises en considération pour la construction du diagramme de Voronoï. Nous pensons qu'une segmentation préalable du tissu épithélial est la solution pour générer des polygones qui soient restreints aux surfaces des cellules épithéliales. A l'état actuel, nous continuons nos travaux pour écarter les cellules non épithéliales (stroma) lors de la phase de détection des germes. Nous travaillons actuellement sur des procédures de traitement d'images, basées sur des critères de forme, de taille et de position des noyaux cellulaires pour éliminer les germes non épithéliaux.

9.4 Le concept « TMA virtuel » : Alternatif et complément à la technologie TMA

Le recours à la technologie TMA permet l'accélération des procédures classiques : biologiques et histologiques, notamment pour la découverte des nouveaux marqueurs tumoraux à visée pronostic. En revanche, l'élaboration d'un TMA est longue et fastidieuse, elle doit donc être soigneusement motivée (expression d'un besoin), guidée (spécification d'un protocole) et validée. Plusieurs étapes interviennent dont : 1) la formulation d'un objectif (ex : examiner la relation entre grade et facteur de prolifération), 2) la spécification d'un plan d'expérience, 3) la réalisation du TMA et 4) l'exploitation du TMA.

Dans le cadre des travaux de thèse, nous avons introduit le concept de « TMA virtuel » pour répondre aux différentes problématiques posées par le recours à la technologie TMA. Un outil permettant la construction de TMA virtuel a été développé pour illustrer l'exploitation des données générées par les TMA (§ Annexe B). Un TMA virtuel peut regrouper plusieurs carottes virtuelles ou réelles pour répondre à une question scientifique donnée. Au cours de cette thèse, nous avons illustré l'utilisation de ce concept :

- comme moyen d'évaluation de la technologie TMA pour l'étude du cancer colorectal,
- comme support pour la validation de nouvelles approches de traitement d'images dédiées à la technologie TMA,
- et comme un outil de fouille et d'exploitation des données générées par les TMA.

9.4.1 TMA virtuel : outil d'investigation et d'évaluation de la technologie TMA

La stratégie de prélèvement des carottes virtuelles, présentée dans le troisième chapitre, tire profit des avantages des développements actuels des outils de microscopie virtuelle et des stations d'acquisitions d'images haut-débit. Cette stratégie permet la simulation, par des essais multiples et non limités, des prélèvements réels des carottes. Cette stratégie a été mise en application pour l'étude de l'échantillonnage des tumeurs colorectales lors de la conception d'un bloc TMA. Bien que notre approche par simulation ne remettent pas en cause la technologie TMA, mais elle nous met en garde contre l'utilisation de cette technologie pour l'étude des marqueurs tumoraux hétérogènes.

Nous proposons d'adopter le concept de TMA virtuel comme complément à la technologie TMA. A titre d'exemple, un bloc TMA peut être construit pour l'étude des marqueurs tumoraux à distribution homogène, avec la procédure classique de trois carottes par bloc de biopsie. Des coupes histologiques peuvent être pratiquées au niveau des blocs de biopsie donneur même après carottage (pour la fabrication du TMA réel), pour la révélation des marqueurs tumoraux les plus hétérogènes. En complément, un TMA virtuel peut être généré par le prélèvement de carottes virtuelles à partir des images de ces coupes histologiques. La mise en correspondance finale des carottes réelles (TMA réel) avec des carottes virtuelles

(TMA virtuel), peut être réalisée pour l'étude par exemple de l'ensemble des marqueurs tumoraux (homogènes et hétérogènes) impliqués dans la progression d'un type de cancer donné.

9.4.2 TMA virtuel : Support de validation de nouvelles approches de traitement d'images

Au lieu de procéder à la confection d'un bloc TMA réel pour tester ou développer de nouvelles procédures de traitement d'images dédiées à la technologie TMA. Nous pouvons faire appel au TMA virtuel pour appliquer ces procédures sur des images de carottes virtuelles. Cette stratégie a l'avantage d'éliminer les nombreuses étapes, longues et coûteuses, nécessaires pour construire un bloc de TMA réel. Nous avons eu recours à cette approche pour présenter et valider notre approche de classification couleur (chapitre 7). Le concept de sociologie cellulaire pour l'étude de la progression tumorale du cancer colorectale a été également mis en application sur des images de carottes virtuelles (chapitre 8).

9.4.3 TMA virtuel : Outil d'exploitation et de fouille de données

Le TMA virtuel peut être utilisé comme un outil d'exploitation et de fouille de données. Nous pouvons ainsi regrouper des coupes de carottes réelles (coupes TMA) ou virtuelles (à partir des coupes histologiques), pour répondre à une problématique de recherche donnée. Cette stratégie a été illustrée à travers l'étude de la nature tissulaire en profondeur des carottes TMA (Annexe B).

Bibliographie

A. Abd-El-Hameed. Survivin Expression in Colorectal Adenocarcinoma Using Tissue Microarray. *Journal of the Egyptian Nat. Cancer Inst.*, vol. 17(1): 42-50, 2005.

H. Aberle, A. Bauer, J. Stappert, A. Kispert et R. Kemler. b-catenin is a target for the ubiquitin-proteasome pathway. *EMBO J*, vol. 16: 3797-804, 1997.

W.W. Abney. Researches in colour vision. *Longmans*, Green, London, 1913.

M. Adel. Détection de défauts de placage d'Okoumé par traitement d'images couleur et analyse de texture. Thèse de doctorat, Institut National Polytechnique de Lorraine, janvier 1994.

AFNOR. Couleur et colorimétrie. *Afnor*, 1989.

I.S. Ahmad et J.F. Reid. Evaluation of color representations for maize images. *Journal of Agricultural Engineering Research*, vol. 63(3): 185–195, 1996.

K. Al Kuraya, R. Simon et G. Sauter. Tissue microarrays for high-throughput molecular pathology. *Ann Saudi Med*, vol. 24(3), May-June 2004.

S.R. Alonso, P. Ortiz, M. Pollan, B. Perez-Gomez, L. Sanchez, M.J. Acuna, R. Pajares, F.J. Martınez-Tello, C.M. Hortelano, M.A. Piris et J.L. Rodrıguez-Peralto. Progression in Cutaneous Malignant Melanoma Is Associated with Distinct Expression Profiles *A Tissue Microarray-Based Study. American Journal of Pathology,* vol. 164(1), January 2004.

W. Alshatti et P. Lambert**.** Using Eigenvectors of a Vector Field for Deriving a Second Directional Derivative Operator for Color Images. *Journal of Intelligent and Robotic Systems*, vol. 28(1-2): 5-20, June 2000.

M. Amadasun et R.A. King. Low level segmentation of multispectral images via agglomerative clustering of uniform neighbourhoods. *Pattern Recognition*, vol. 21(3): 261-268, 1988.

C.L. Andersen, O. Monni, U. Wagner , J. Kononen , M. Bärlund, C. Bucher, P. Haas, A. Nocito, H. Bissig, G. Sauter et A. Kallioniemi. High-throughput copy number analysis of 17q23 in 3520 tissue specimens by fluorescence in situ hybridization to tissue microarrays. *Am J Pathol*, vol. 161 : 73-9, 2002.

J. Angulo Lopez. Morphologie mathématique et indexation d'images couleur. Application à la microscopie en médecine. Thèse de doctorat, Ecole des mines de Paris, France, décembre 2003.

D.C Aziz et R.B. Barathur. Quantification and Morphometric analysis of tumors by image analysis. *J Anal Biochem*, vol. 19: 120-125, 1994.

S.P. Bach, A.G. Renehan et C.S. Potten. Stem cells: The intestinal stem cell as a paradigm. *Carcinogenesis*, vol. 21: 469-476, 2000.

D.H. Ballard et C.M. Brown. Computer vision, *Prentice-Hall*, New Jersey, 1982.

P.H. Bartels, R. Montironi, P.W. Hamilton, D. Thompson, L. Vaught et H.G Bartels. Nuclear chromatin texture in prostatic lesions. II. PIN and maligancy associated changes. *Anal Quant Cytol Histol*, vol. 20: 389-396, 1998.

G.J. Battaglia et J.M. Maynard. Means Square Error: A useful tool for statistical process management. *AMP Journal of Technology*, vol. 2, November 1992.

L. Bedat. Aspects psychovisuels de la perception des couleurs. Application au codage d'images couleur fixes avec compression de l'information. Thèse de doctorat, Université de Nantes, octobre 1998.

J. Behrens, J.P. von Kries, M. Kuhl, L. Bruhn, D. Wedlich, R. Grosschedl et W. Birchmeier. Functional interaction of β-catenin with the transcription factor LEF-1. *Nature*, vol. 382: 638-42, 1996.

P. Belhomme, A. Elmoataz, P. Herlin et D. Bloyet. Generalized region growing operator with optimal scanning: application to segmentation of breast cancer images. *Journal of Microscopy*, vol. 186: 41-50, 1997.

Y. Bellaïche et N. Perrimon. La voie de signalisation Wingless chez la drosophile. *Med Sci*, vol. 13: 166-74, 1997.

C.E.H. Berger, J.A. de Koeijer, W. Glas et H.T. Madhuizen. Color Separation in Forensic Processing. *J Forensic Sci*, vol. 51(1), January 2006.

F. Bertucci, S. Salas, S. Eysteries, V. Nasser, P. Finetti, C. Ginestier, E. Charafe-Jauffret, B. Loriod, L. Bachelart, J. Montfort, G. Victorero, F. Viret, V. Ollendorff, V. Fert, M. Giovaninni, J.R. Delpero, C. Nguyen, P. Viens, G. Monges, D. Birnbaum et R. Houlgatte. Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters. *Oncogene*, vol. 23(7): 1377-91, 2004.

H. Bhattufira. The multitumor (sausage) tissue block: novel method for immunohistochemical antibody esting. *Lab Invest*, vol. 55: 244–248, 1986.

M. Bienz et H. Clevers. Linking colorectal cancer to Wnt signaling. *Cell*, vol. 103: 311-20, 2000.

G. Bigras, R. Marcepoil, E. Brambilla et G. Brugal. Cellular sociology applied to neuroendocrine tumors of the lung: Quantitative model of neoplastic architecture. *Cytometry*, vol. 24: 74-82, 1996.

J.M. Bland et D.G. Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, vol. 1: 307-310, 1986.

J.M. Bland et D.G. Altman. Measurement error and correlation coefficients. *BMJ*, vol. 313: 41–42, 1996.

M. Borsotti, P. Campadelli et R. Schettini. Quantitative evaluation of color image segmentation results. *Pattern Recognition Letters*, vol. 19 (8): 741 – 747, June 1998.

G.S. Bova, G. Parmigianni, J.I. Epstein, T. Wheeler, N.R. Mucci et M.A. Rubin. Web-Based Tissue Microarray Image DataAnalysis: Initial Validation Testing Through Prostate Cancer Gleason Grading. *Hum Pathol*, vol. 32: 417-427, 2001.

J.K. Bowmarker et H.J.A Dartnall. Visual Pigments of rods and cones in a human retina. *J Physiol*, vol. 298: 501-511, 1980.

R.M. Boynton et J. Gordon. Bezold-brucke hue shift measured by color-naming technique. *Journal of Optical Society of America*, vol. 55: 78-86, 1965.

E.M. Brey, Z. Lalani, C. Johnston, M. Wong, L.V. McIntire, P.J. Duke et C.W. Patrick. Automated Selection of DAB-labeled Tissue for Immunohistochemical Quantification. *The Journal of Histochemistry & Cytochemistry*, vol. 51(5): 575–584, 2003.

M. Brittan et N.A. Wright. Gastrointestinal stem cells. *J Pathol*, vo.197: 492-509, 2002.

W.R.J. Brown et D.L. MacAdam. Visual sensitivities to combined chromaticity and luminance differences. *Journal of Optical Society of America*, vol. 39(10): 808-834, 1949.

F. Brunz, P.M. Hwang, C. Torrance, T. Waldman, Y. Zhang, L. Dillehay, J. Williams, C. Lengauer, K.W. Kinzler et B. Vogelstein. Disruption of p53 in human cancer cells alters the response to therapeutic agents. *J Clin Inves*, vol. 104(3): 263-269, 1999.

M.H. Bui, H. Visapaa, D. Seligson, H. Kim, K.R. Han, Y. Huang, S. Horvath, E.J. Stenbridge, A. Palotie, R.A. Figlin et A.S. Belldegrun. Prognostic value of kroniek anhydrase IX and Ki67 as predictor of survival for renal clear cell cancer. *J Urol*, vol. 171: 2461-2466, 2004.

F. Bumbaca et K.C. Smith. Design and implementation of a colour vision model for computer vision applications. *Computer Vision, Graphics and Image processing*, vol. 39(2): 226-245, august 1987.

P.C. Burger, T. Shibata et P. Kleihues. The use of monoclonal antibody Ki-67 in the identification of proliferating cells: application to surgical neuropathology. *Am J Surg Pathol*, vol. 10: 611-617, 1986.

J. Byun, M.R. Verardo, B. Sumengen, G.P. Lewis, B.S. Manjunath et S.K. Fisher. Automated tool for the detection of cell nuclei in digital microscopic images: Application to retinal images. *Molecular Vision*, vol. 12 : 949-60, 2006.

R.L. Camp, L.A. Charette et D.L. Rimm. Validation of tissue microarray technology in breast carcinoma. *Lab Invest*, vol. 80: 1943-9, 2000.

R.L. Camp, G.G. Chung et D.L. Rimm. Automated subcellular localization and quantification of protein expression in tissue microarrays. *Nat Med*; vol. 8: 1323–1327, 2002.

R.L. Camp, M. Dolled-Filhart, B.L. King et D.L. Rimm.Quantitative analysis of breast cancer tissue microarrays shows that both high and normal levels of HER2 expression are associated with poor outcome. *Cancer Res*, vol. 63: 1445–1448, 2003.

J. Canny. A computational approach to edge detection. *IEEE, Pattern analysis and Machine Intelligence*, vol. 8(6): 679-698, 1986.

T. Carron et P. Lambert. Color edge detector using jointly hue, saturation and Intensity. *International Conference on Image Processing*, vol. 3: 977-981, 1994.

T. Carron. Segmentation d'images couleur dans la base Teinte Luminance Saturation : approche numérique et symbolique, Thèse de doctorat, Université de Savoie, Chambéry, France, décembre 1995.

S. Carvajal-Gonzales, N. Delisle, D. König et J.P. Rigaut. Direct 3-D processing of histological images obtained by confocal scanning laser microscopy. *Acta Stereol*, 1992.

M. Celenk. Hierarchical Color Clustering for Segmentation of Textured Images. *29th Southeastern Symposium on System Theory*, pp. 483-487, Cookeville, TN, 9-11 Mar 1997.

G. Celeux, E. Diday, G. Govaert, Y. Lechevallier, et H. Ralambondrainy. Classification .automatique des données. *Dunod informatique*, 1989.

R. Chandebois. Cell sociology and the problem of position effect : pattern formation, origin and role of gradients. *Acta Biotheoretica*, vol. 26: 203-238, 1977.

M. Chapron. A new chromatic edge detector used for color image segmentation.*11th International Conference on Pattern Recognition*, vol. 3: 311-314, Den Hague, Netherlands, 30 Aug-3 Sept. 1992.

C. Charpin, J.P. Dales, S. Garcia1, S. Carpentier, A. Djemli, L. Andrac, M.N. Lavaut, C. Allasia et P. Bonnier. Tumor Neoangiogenesis by CD31 and CD105 Expression Evaluation in Breast Carcinoma Tissue Microarrays. *Clinical Cancer Research*, vol. 10: 5815-5819, September 2004.

J.M. Chassery et C. Garbay. An interative segmentation method based on a contextual color and shape criterion. *IEEE, Pattern Analysis &Machine Intelligence*, vol. 6(6): 794-800, 1984.

H.D. Cheng, X.H. Jiang, Y. Sun et J. Wang. Color image segmentation: advances and prospects. *Pattern recognition*, vol. 18: 1857-1872, 1999.

A.G. Chiesa-Vottero, L.A. Rybicki, et R.A. Prayson. Comparison of Proliferation Indices in Glioblastoma Multiforme by Whole Tissue Section vs Tissue Microarray. *American journal of clinical pathology*, vol. 120(6): 902-908, 2003.

S.F Chin, Y. Daigo, H.E. Huang, N.G Iyer, G. Callagy, T. Kranjac, M. Gonzalez, T. Sangan, H. Earl et C Caldas. A simple and reliable pretreatment protocol facilitates fluorescent in situ hybridisation on tissue microarrays of paraffin wax embedded tumour samples. *Molecular Pathology*, vol. 56: 275-279, 2003.

G.G. Chung, E. Provost, E.P. Kielhorn, L.A. Charette, B.L. Smith et D.L. Rimm. Tissue Microarray Analysis of βcatenin in Colorectal Cancer Shows Nuclear Phospho-βcatenin Is Associated with a Better Prognosis. *Clinical Cancer Research*, vol. 7: 4013–4020, December 2001.

G. Claramont. Développement de techniques de segmentation chromatique et implantation dans un module de vision. Thèse de doctorat, Institut National Polytechnique de Toulouse, décembre 1984.

M.A. Cobb, M. Husain, B.J. Andersen et O. Al-Mefty. Significance of proliferating cell nuclear antigenin predicting recurrence of intracranial meningioma. *J Neurosurg*, vol. 84: 85-90, 1996.

Commission Internationale de l'Eclairage. « Colorimetry ». *Tech.Rep. CIE 15.2*, 1986.

C.M. Conway, D. O'Shea, S. O'Brien, D. K. Lawler, G. D. Dodrill, , A. O'Grady, H. Barrett, C. Gulmann, L. O'Driscoll, W.M. Gallagher, E.W. Kay et D.G. O'Shea. The development and validation of the Virtual Tissue Matrix, a software application that facilitates the review of tissue microarrays on line. *BMC Bioinformatics*, vol. 7 : 256-267, 2006.

V. Coutance. La couleur en vision par ordinateur: Application à la robotique. Thèse de doctorat, Université Paul Sabatier de Toulouse, janvier 1991.

B. Cramariuc, M. Gabbouj, et J. Astola. Clustering Based Region Growing Algorithm for Color Image Segmentation. *The 13th International Conference on Digital Signal Processing*, Santorini, Greece, vol. 2: 857-860, 2-4 July 1997.

C. Crosnier, D. Stamataki et J. Lewis. Organizing cell renewal in the intestine: stem cells, signals and combinatorial control. *Nat Rev Genet*, vol.7: 349-359, 2006.

A. Cumani. Edge Detection in Multispectral Images. *Graphical Models and Image Processing*, vol. 53(1): 40–51, 1991.

M.J. Daily. Color image segmentation using markov random fields. *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 304–312, 1989.

L.E. DeBault et R.R. Tubbs. Tissue Microarrays: Robust Molecular Morphology Tools. *J Mol Hist*, vol. 38: 111–112, 2007.

A.M. De Marzo, H.H. Fedor, W.R. Gage et M.A. Rubin. Inadequate formalin fixation decreases reliability of $p27^{Kip1}$ immunohistochemical staining: Probing optimal fixation time using high-density tissue microarrays. *Human pathology*, vol. 33(7): 756-760, 2002.

F. Demichelis. On information organization and information extraction for the study of gene expressions by tissue microarray technique. PhD thesis, DIT - University of Torento, 2005.

L.K. Diaz, R. Gupta, N. Kidwai, N. Sneige et E.L. Wiley. The use of TMA for interlaboratory validation of FISH testing for detection of HER2 gene amplification in breast cancer. *Journal of Histochemistry and Cytochemistry*, vol. 52(4): 501-507, 2004.

E. Diday, J. Lemaire, J. Pouget et F. Testu. Eléments d'analyse de données. *Dunod*, 1982.

K.A. DiVito, A.J. Berger, L.R. Camp, M. Dolled-Filhart, D.L. Rimm et H.M. Kluger. Automated Quantitative Analysis of Tissue Microarrays Reveals an Association between High Bcl-2 Expression and Improved Outcome in Melanoma. *Cancer Research*, vol. 64: 8773-8777, December 2004.

K.A. DiVito et R.L. Camp. Tissue microarrays – automated analysis and future directions. *Cambridge University Press, ISSN 1470-903. Breast Cancer Online*, vol. 8(7), 2005.

S. Di Zenzo. A note on the gradient of a multi-image. *Computer Vision, Graphics, and Image Processing*, vol. 33: 116-125, 1986.

C. Dussert, G. Rasigni, M. Rasigni et J. Palmari. Minimal spanning tree: A new approach for studying order and disorder. *Phys. Rev. B*, vol. 34: 3528-3531, 1986.

D. Eastwood. A simple modification to improve the visual uniformity of the CIE 1964 *U* V* W** colour space. In *Second congress of the Association Internationale de la Couleur, Colour'73*, pp. 293–296, York, 1973.

N. Elie, B. Plancoulaine, J.P. Signolle et P. Herlin. A Simple Way of Quantifying Immunostained Cell Nuclei on the Whole Histologic Section. *Cytometry*, vol. 56(A): 37–45, 2003.

J. Engellau, M. Akerman, H. Anderson, H.A. Domanski, E. Rambech, Alvegard TA et M. Nilbert. Tissue microarray technique in soft tissue sarcoma: immunohistochemical Ki-67 expression in malignant fibrous histiocytoma. *Appl Immunohistochem Mol Morphol*, vol. 9(4): 358–363, 2001.

J. Engellau. Prognostic factors in soft tissue sarcoma: Tissue microarray for immunostaining, the importance of whole-tumor sections and time-dependence. Ph.D. Thesis, Lund University, Sweden. *Acta Orthopaedica Scandinavica Supplementum*, vol. 75(314), 2004.

M.D. Fairchild. Color appearance models. *Addison-Wesley,* 1998.

O.D. Faugeras. Digital color image processing within the framework of a human visual model. *IEEE Transactions on Acoustics, Speech, and Signal processing*, vol. 27(4): 380–393, 1979.

E.R. Fearon et B. Vogelstein. A genetic model for colorectal tumorigenesis. *Cell*, vol. 61, 1990.

C. D. Fermin et S. Degraw. Colour tresholding in video imaging. *J Anat*, vol. 186: 469-481, 1995.

E. Fernebro, M. Dictor, P.O. Bendahl, M. Ferno et M. Nilbert. Evaluation of the Tissue Microarray Technique for Immunohistochemical Analysis in Rectal Cancer. *Arch Pathol Lab Med*, vol. 126, June 2002.

M.N. Fesharaki et G.R. Hellestrand. Real-time colour image segmentation. *School of Computer Science and Engineering, University of New South Wales, Technical Report: 9316,* Australia, December 1993.

P.L. Fitzgibbons, D.L. Page, D. Weaver, A.D. Thor, D.C. Allred, G.M. Clark, S.G. Ruby, F. O'Malley, J.F. Simpson, J.L. Connolly, D.F. Hayes, S.B. Edge, A. Lichter et S.J. Schnitt. Prognostic factors in breast cancer. *College of American Pathologists Consensus Statement 1999. 1. Arch Pathol Lab Med*, vol.124: 966–978, 2000.

J.D. Foley, A. Van Dam et S.K. Feiner. Computer graphics, principles and practice. *Addison-Wesley, second edition,* 1990.

S.P. Foliguet et L. Guigues. Evaluation de la Segmentation d'Images : Etat de l'Art, Nouveaux Indices et Comparaison**.** *Traitement du Signal*, vol. 23(2): 109-124, 2006.

S.C. Foo. A gonioreflectometer for measuring the bidirectional reflectance of materiel for use in illumination computation. Ph.D. Thesis, Cornell University, august 1997.

S. Fre, S.K. Pallavi, M. Huyghe, M. Lae, K.P. Janssen, S. Robine, S. Artavanis-Tsakonas et D. Louvard. Notch and Wnt signals cooperatively control cell proliferation and tumorigenesis in the intestine. *PNAS : Proc Natl Acad Sci U SA* , vol. 106(15): 6309-6314, April 2009.

K.S. Fu et J.K. Mui. A Survey on Image Segmentation. *Pattern Recognition*, vol. 13: 3-16, 1981.

N. Funakubo. Feature Extraction of Color Texture Using Neural Networks for Region Segmentation. *The 20th Annual Conference of IEEE Industrial Electronics*, vol. 2: 852-856, Bologna, Italy, 5-9 Sept. 1994.

P.N. Furness. The use of digital images in pathology. *J Pathol*, vol.183: 253–263, 1997.

S. Galichet. Etude d'un système d'incrustation d'images vidéo couleur. Thèse de doctorat, Université de Technologies de Compiègne, décembre 1989.

C. Garbay. Modélisation de la couleur dans le cadre de l'analyse d'images et de son application à la cytologie automatique. Thèse de doctorat, Institut National Polytechnique de Grenoble, décembre 1979.

C. Garbay, G. Brugal et C. Choquet. Application of colored image analysis to bone marrow cell recognition. *Anal Quant Cutol*, vol. 3: 272-280, 1981.

J.F. García, F.I. Camacho, M. Morente, M. Fraga, C. Montalbán, T. Álvaro, C. Bellas, Á. Castaño, A. Díez, T. Flores, C. Martín, M.A. Martínez, F. Mazorra, J. Menárguez, M.J. Mestre, M. Mollejo, A.I. Sáez, L. Sánchez et A. Miguel. Hodgkin and Reed-Sternberg cells harbor alterations in the major tumor suppressor pathways and cell-cycle checkpoints: analyses using tissue microarrays. *Blood*, vol. 101(2): 681-689, January 2003.

J. Gerdes, H. Lemke, H. Baish, H.H. Wacker, U. Schwab et H.Stein. Cell cycle analysis of a cell proliferation-associated human antigen defined by the monoclonal antibody ki-67. *J Immunology*, vol. 133: 1710-1715, 1984.

T.G. Gerike, U. Paulus, C.S. Potten et M. Loeffler. A dynamic model of proliferation and differentiation in the intestinal crypt based on a hypothetical intraepithelial growth factor. *Cell Prolif*, vol. 31: 93-110, 1998.

T. Gevers et A.W.M. Smeulders. Combining Region Splitting and Edge Detection Through Guided Delaunay Image Subdivision. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1021-1026, San Juan, Puerto Rico, 17-19 June 1997.

J.M. Giltnane et D.L. Rimm. Technology insight: Identification of biomarkers with tissue microarray technology. *Nat Clin Pract Oncol*, vol. 1(2): 104-11, 2004.

C. Ginestier, E. Charafe-Jauffret, F. Bertucci, F. Eisinger, J. Geneix, D. Bechlian, N. Conte, J. Adélaïde, Y. Toiron, C. Nguyen, P. Viens, M.J. Mozziconacci, R. Houlgatte, D. Birnbaum et J. Jacquemier. Distinct and complementary information provided by use of tissue and DNA microarrays in the study of breast tumor markers. *Am J Pathol*, vol. 161: 1223-33, 2002.

F. Giroud, J. Simony-Lafontaine, M.P. Montmasson et C. Garbay. The TMA-Explorer plateform: an integrated tool integrating a virtual TMA concept. Poster Abstract in *Cytometry, 22th ISAC Congress*, vol. 59(1): 153. Montpellier (France), 22-27 Mai, 2004.

H. Grassman. On the theory of compound colours. *Philosophical Magazine Series*, vol. 4(7): 254–264, 1853.

G.D. Guo, S. Yu et S.D. Ma. Unsupervised Segmentation of Color Images. *IEEE International Conference on Image Processing*, pp. 299-302, 1998.

A. Guzman de Leon. Etude d'espaces couleur pour la segmentation et l'aide à l'analyse d'images colposcopiques. Thèse de doctorat, Université de Technologies de Compiègne, juillet 1997.

N. Harada, Y. Tamai, T. Ishikawa, B. Sauer, K. Takaku, M. Oshima, et M.M. Taketo. Intestinal polyposis in mice with a dominant stable mutation of the b-catenin gene. *EMBO J*, vol. 18: 5931-42, 1999.

R.M. Haralick et L.G. Shapiro. Image Segmentation Techniques. *Computer Vision, Graphics and Image Processing*, vol. 29: 100-132, 1985.

T.C. He, A.B. Sparks, C. Rago, H. Hermeking, L. Zawel, L.T. da Costa, P.J. Morin, B. Vogelstein et K.W. Kinzler. Identification of c-MYC as a target of the APC pathway. *Science*, vol. 281: 1509-12, 1998.

Y. Hendriks, P. Franken, J.W. Dierssen, W. de Leeuw, J. Wijnen, E. Dreef, C. Tops, M. Breuning, A. Bröcker-Vriends, H. Vasen, R. Fodde et H. Morreau. Conventional and Tissue Microarray Immunohistochemical Expression Analysis of Mismatch Repair in Hereditary Colorectal Tumors. *American Journal of Pathology*, vol. 162: 469-477, 2003.

M. Hensel, A. Schneeweiss, H.P. Sinn, G. Egerer, E. Solomayer, R. Haas, G. Bastert et A.D. Ho. P53 is the strongest predictor of survival in highrisk primary breast cancer patients undergoing high-dose chemotherapy with autologous blood stem cell support. *Int J Cancer*, vol.100: 290–296, 2002.

W. Hilbe, A. Gächter, H.C. Duba, S. Dirnhofer, W. Eistrer, T. Schmid, A. Mildner, J. Bodner et E. Wöll. Comparison of automated cellular imaging system and manual microscopy for immunohistochemically stained cryostat sections of lung cancer specimens applying p53, ki-67 and p120. *Oncology Reports*, vol. 10: 15-20, 2003.

K. Holla. Opponent colors as a 2-dimensional feature within a model of the first stages of the human visual system. *The 6th International Conference on Pattern Recognition*, pp. 561-563, Munich, Germany, 19-22 Oct. 1982.

A. Hoos, M.J. Urist, A. Stojadinovic, S. Mastorides, M.E. Dudas, D.H.Y. Leung, D. Kuo, M.F. Brennan, J.J. Lewis et C. Cordon-Cardo. Validation of Tissue Microarrays for Immunohistochemical Profiling of Cancer Specimens Using the Example of Human Fibroblastic Tumors. *American Journal of Pathology*, vol. 158(4), April 2001.

A. Hoos et C. Cordon-Cardo. Tissue microarray profiling of cancer specimens and cell lines: opportunities and limitations. *Lab Invest*, vol. 81: 1331-38, 2001.

A. Hoos, A. Nissan, A. Stojadinovic, J. Shia, C.V. Hedvat, D.H.Y. Leung, P.B. Paty, D. Klimstra, C. Cordon-Cardo et W.D. Wong. Tissue Microarray Molecular Profiling of Early, Node-negative Adenocarcinoma of the Rectum: A Comprehensive Analysis. *Clinical Cancer Research*, vol. 8: 3841-3849, December 2002.

B. Hopkins et J.G. Skellam. A New Method for determining the Type of Distribution of Plant Individuals. *Ann. Bot*, vol.18: 213-227, 1954.

S.L. Horowitz et S. Pavlidis. Picture segmentation by a directed split and merge procedure. *2th International Conference on Pattern Recognition*, pp. 424-433, 1974.

T.L. Huntsberger, C.L. Jacobs et R.L Cannon. Iterative fuzzy image Segmentation. *Pattern Recognition*, vol. 18(2): 131-138, 1985.

T.L. Huntsberger et M.F. Descalzi. Colour edge detection", *Pattern Recognition*, vol. 3: 205-209, 1985.

INC, Institut National du Cancer. http://www.e-cancer.fr, 2009.

IVS, Institut de Veille Sanitaire. http://www.invs.sante.fr, 2009.

I.A. Ismaili et D.F. Gillies.Colour image segmentation using regression analysis in RGB space. *Machine Graphics & Vision*, vol. 3: 373-384, 1994.

S. Itoh et I. Matsuda. Segmentation of Colour Still Images Using Voronoi Diagrams. *8th Conference European Signal Processing*, vol. 3: 1869-1872, Trieste, Italy, 10-13 September 1996.

J. Jacquemier, C. Ginestier, E. Charafe-Jauffret, F. Bertucci, T. Bege, J. Geneix et D. Birnbaum. Small but high throughput: how «tissue microarrays " became a favorite tool for pathologists and scientists. *Ann Pathol*, vol. 23(6): 623-32, 2003.

R. Jain, R. Kasturi, et B.G. Schunck, Machine Vision, McGraw-Hill, New York, 1995.

D. Jameson et L. M. Hurvich. Some quantitative aspects of an opponent colors theory. I chromatic response and spectral saturation. II brightness, saturation and hue in normal and dichromatic vision. *Journal of Optical Society of America*, vol. 45(7-8), August-July 1955.

D. Jameson et L. M. Hurvich. Opponent chromatic induction: Experimental evaluation and theorical account. Journal of Optical Society of America, vol. 51(1): 46-53, January 1961.

A.C. Johansson, E. Visse, B. Widegren, H.O. Sjogren et P. Siesjo. Computerized image analysis as a tool to quantify infiltrating leukocytes: a comparison between high- and low-magnification images. *J Histochem Cytochem*, vol. 49: 1073–1079, 2001.

A.M. Jubb, T.H. Landon, J. Burwick, T.Q. Pham, G.D. Frantz, B. Cairns, P. Quirke, F.V. Peale et K.J. Hillan. Quantitative analysis of colorectal tissue microarrays by immunofluorescence and *in situ* hybridization. *J Pathol*; vol. 200: 577–588, 2003.

D.B. Judd. Reduction of data on mixture of color stimuli. *Bureau of Standards Journal Research*, vol. 4(163): 515–548, 1930.

M. Kai, O. Nunobiki, E. Tanuguchi, Y. Sakamoto et K. Kakudo. Quantitative and qualitative analysis of stain color using RGB computer color specification. *Anal Quant Cytol Histol*, vol. 21: 477-480, 1999.

A. Kajdacsy-Balla, J.M. Geynisman, V. Macias, S. Setty, N.M. Nanaji, J. J. Berman, K. Dobbin, J. Melamed, X. Kong, M. Bosland, J. Orenstein, J. Bayerl, M.J. Becich, R. Dhir et M.W. Datta. The Cooperative Prostate Cancer Tissue Resource. Practical aspects of planning, building, and interpreting tissue microarrays: The Cooperative Prostate Cancer Tissue Resource experience. *J Mol Hist*, vol. 38: 113–121, 2007.

O.P. Kallioniemi, U. Wagner, J. Kononen et G. Sauter. Tissue microarray technology for high-throughput molecular profiling of cancer. Human Molecular Genetics, vol. 10(7): 657-662, 2001.

Y. Kamei-Harada, M. Miyakoshi, et M. Shimbo. A Fuzzy Clustering Method for Automatic Segmentation of Color Scenes. *Japanese Journal of Fuzzy Theory and Systems*, vol. 6(5): 651-672, 1994.

Y. Kanai. Image Segmentation Using Intensity and Color Information. *The International Society for Optical Engineering. Visual Communications and Image Processing*, San Jose, California, pp. 709-720, 28-30 Jan 1998.

M. Kass, A. Witkin, et D. Terzopoulos. Snakes: Active Contour Models. *International Journal of Computer Vision*, vol. 1: 321-331, 1988.

K. Kayser, H.J. Gabius et O. Hagemeyer. Malignant teratoma of the lung with lymph node metastasis of the ectodermal compartment : a case report. *Anal. Cell. Pathol*, vol. 5: 31-37, 1993.

Y.J. Kim, B.F.M. Romeike, J. Uszkoreit et W. Feiden. Automated nuclear segmentation in the determination of the Ki-67 labeling index in meningiomas.*Clinical Neuropathology*, vol. 25(2): 67-73, 2006.

T.W. King, E.M. Brey, A. Youssef., C. Johnston et C.W. Patrick. Quantification of Vascular Density Using a Semiautomated Technique for Immunostained Specimens. *Analytical and quantitative cytology and histology*, vol. 24(1): 39-48, 2002.

K.W. Kinzler et B. Vogelstein. Lessons from Hereditary Colorectal Cancer. *Cell*, vol. 87: 159-170, 1996.

M.R. Kok, M.E. Boon, P.G. Schreiner-Kok, J. Hermans, D.E. Grobbee et L.P. Kok. Less medical intervention after sharp demarcation of grade 1-2 cervical intraepithelial neoplasia smears by neural network screening. *Cancer Cytopathology*, vol. 93: 173–178, 2001.

J. Kononen, L. Bubenfort, A. Kallioniemi, M. Barlund, P. Schraml, S. Leighton, J. Torhorst, M.J. Mihatsch, G. Sauter et O.P. Kallioniemi. Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med*, vol. 4(7): 844–877, 1998.

V. Korinek, N. Barker, P.J. Morin, D. Van Wichen, R. De Weger, K.W. Kinzler, B. Vogelstein et H. Clevers. Constitutive transcriptional activation by a β-catenin-Tcf complex in APC-/- colon carcinoma. *Science*, vol. 275: 1784-7, 1997.

P. Korkolopoulou, I. Vassilopoulos, A.E. Konstantinidou, H. Zorzos, E. Patsouris, E. Agapitos et P. Davaris. The combined evaluation of p27Kip1 and Ki-67 expression provides independent information on overall survival of ovarian carcinoma patients. *Gynecol Oncol*, vol. 85: 404–414, 2002.

R.K. Kouassi, J.C. Devaux, P. Gouton et M. Paindavoine. Application of the Karhunen-Loeve transform for natural colour images analysis. *Irish Machine Vision and Image Processing Conference*, vol. 1: 20–27, Londonderry, 1997.

P. Kowaliski. Vision et mesure de la couleur. Physique fondamentale et appliquée. *Masson, seconde edition*, 1990.

M. Kunt, G. Granlund, M. Kocher et C. Horne. Traitement de l'information : Traitement numérique des images. Presses polytechniques et universitaires romandes, vol. 2, 1993.

G. L'Allemain. Rôle des voies Wnt dans l'oncogenèse. *Bull Cancer hors série*, pp. 88-9, Avril 2006.

G. Landini et I.E. Othman. Estimation of tissue layer level by sequential morphological reconstruction. *Journal of Microscopy*, vol. 209: 118–125, 2 February 2003.

G. Landini. http://www.den tistry.bham.ac.uk/landinig/soft ware/software.html. 2004.

G. Landini. Quantitative analysis of the epithelial lining architecture in radicular cysts and odontogenic keratocysts. *Head & Face Medicine*, vol. 2(4), 2006.

P. Lasserre. Vision pour la robotique en environnement naturel. Thèse de doctorat, Université Paul Sabatier de Toulouse, septembre 1996.

H. C. Lee et D. Cok. Detection boundaries in a vector field. *IEEE Trans. Signal Processing*, vol. 39(5), 1991.

J.H. Lee, B.H. Chang et S.D. Kim. Comparison of colour transformations for image segmentation. *Electronics Letters*, vol. 30(20): 1660–1661, 1994.

H.A. Lehr, D.A. Mankoff, D. Corwin, G. Santeusanio et A.M. Gown. Application of Photoshop-based image analysis to quantification of hormone receptor expression in breast cancer. *J Histochem Cytochem*, vol. 45: 1559–1565, 1997.

H.A. Lehr, C.M. Van Der Loos, P. Teeling et A.M. Gown. Complete chromogen separation and analysis in double immunohistochemical stains using Photoshop-based image analysis. *J Histochem Cytochem*, vol. 47: 119–126, 1999.

F.J.W-M. Leong et J.O'D. McGee: Automated complete slide digitization: a medium for simultaneous viewing by multiple pathologists. *J Pathol*, 195: 508–514, 2001.

Y. Leydier , F. Lebourgeois et H. Emptoz. Serialized k-means for adaptative color image segmentation : application to document images and others. *Int. Workshop on Documents Analysis Systems (DAS2004)*, Florence, Italy. vol. 3163: 252-263, September 2004.

O. Lezoray, A. Elmoataz, H. Cardot et M. Revenu. A.R.C.T.I.C : Un système automatique de Tri Cellulaire par Analyse d'Images. *Conference of Vision Interface 99*, Trois-Rivières, Canada, 1999.

O. Lezoray. Segmentation d'images couleur par morphologie mathématique et classification de données par réseaux de neurones : application à la classification de cellules en cytologie des séreuses. Thèse de doctorat, Université de Caen, janvier 2000.

Y.W. Lim et S.U. Lee. On the Color Image Segmentation Algorithm Based on the Thresholding and the Fuzzy c-means Techniques. *Pattern Recognition*, vol. 23(9): 935-952, 1990.

L. Liang. Contribution à la segmentation chromatique des images. Application à l'identification d'objets colorés. Thèse de doctorat, Institut National Polytechnique de Toulouse, juillet 1987.

E. Littmann et H. Ritter. Adaptive color segmentation. A comparison of neural and statistical methods. *IEEE Transactions on Neural Networks*, vol. 8(1): 175–185, 1997.

X. Liu. Analyse d'images couleur en composantes indépendantes par réseau de neurones. Thèse de doctorat, Institut National Polytechnique de Grenoble, décembre 1991.

C.L. Liu, W. Prapong, Y. Natkunam, A. Alizadeh, K. Montgomery, C.B. Gilks et M. van de Rijn. Software tools for highthroughput analysis and archiving of immunohistochemistry staining data obtained with tissue microarrays. American Journal of Pathology, vol. 161(5): 1557-1565, 2002.

C.G. Loukas, G.D. Wilson, B. Vojnovic et A. Linney. An image analysis-based approach for automated counting of cancer cell nuclei in tissue sections. *Cytometry*, vol. 55(A): 30-42, 2003.

V. Lozano. Contribution de l'analyse d'image couleur au traitement des images textiles. Thèse de doctorat, Université Jean Monnet, Saint-Etienne, France, novembre 1997.

L. Lucchese et S.K. Mitra. Color Image Segmentation: A State-of-the-Art Survey. T*he Indian National Science Academy), Image Processing, Vision, and Pattern Recognition*, New Delhi, India, vol. 67(2): 207-221, March 2001.

W. Ma et S. Lozanoff A full color for quantitative assessment of histochemical and immunohistochemical staining patterns. *Biotech Histochem*, vol. 74: 1–9, 1998.

D.L. MacAdam. Projective transformations of I.C.I. color specifications. *Journal of the Optical Society of America*, vol. 27: 294–299, 1937.

D.L. MacAdam. Projective transformations of color-mixture diagram. *Journal of the Optical Society of America*, vol. 32:2–6, 1942.

D.L. MacAdam. Specification of small chromaticity differences. *Journal of the Optical Society of America*, vol. 33: 18–26, 1943.

L. Macaire, V. Ultre et J.G. Postaire. Determination of Compatibility Coefficients for Color Edge Detection by Relaxation. *International Conference on Image Processing*, pp.1045-1048, 1996.

G.M. Maelandsmo, V.A. Florenes, E. Hovig, T. Oyjord, O. Engebraaten, R. Holm, A.L. Borresen et O. Fodstad. Involvement of the pRb/p16/cdk4/ cyclin D1 pathway in the tumorigenesis of sporadic malignant melanomas. *Br J Cancer*, vol. 73: 909–916, 1996.

N. Malpica, C. Ortiz de Solorzano, J.J. Vaquero, A. Santos, I. Vallcorba, J.M. Garcıa-Sagredo et F. del Pozol. Applying Watershed Algorithms to the Segmentation of Clustered Nuclei. *Cytometry*, vol. 28: 289–297, 1997.

S. Manley, N.R. Mucci, A.M. De Marzo et M.A. Rubin. Database Structure to Manage High-Density Tissue Microarray Data and Images for Pathology Studies Focusing on Clinical Outcome. *Am J Pathol*, vol. 159(3), 2001.

R. Marcelpoil et Y. Usson. Methods for the study of cellular sociology: voronoi diagrams and parametrization of the spatial relationships. *J Theor Biol*, vol. 154: 359-369, 1992.

K.A. Matkowskyj, D. Schonfeld et R.V. Benya. Quantitative immunohistochemistry by measuring cumulative signal strength using commercially available software photoshop and matlab. *J Histochem Cytochem*; vol. 48: 303–312, 2000.

B.E. Matysiak, T. Brodzeller, S. Buck, A. French, C. Counts, B. Boorsma, M.W. Datta et A.A. Kajdacsy-Balla. Simple, inexpensive method for automating tissue microarray production provides enhanced microarray reproducibility. *Appl Immunohistochem Mol Morphol*, vol. 11(3): 269-73, Sep 2003.

J. McQueen. Some Methods for Classification and Analysis of Multivariate Observations. *The 5th Symposium on Mathematics, Statistics and Probabilities*, vol. 1: 281-296, Berkeley, 1967.

A.S. Merseburger, M.A. Kuzyk, J. Serth, C.Y Young, L. Sun, R.R. Connely, D.G. Mcleod, F.K. Mostofi, S.K. Srivastava, A. Stenzi, J.W. Moul et I.A. Sesterhenn. Limitations of tissue microarrays in the evaluation of focal alterations of bcl-2 and p53 in whole mount derived prostate tissues. *Oncology Reports*, vol. 10: 223-228, 2003.

G.W. Meyer, H.E. Rushmaeir, M.F. Cohen, D.P. Greenberg et K.T. Torrance. An experimental evaluation of computer graphics imagery. *ACM Transactions on Graphics*, vol. 5(1): 30-50, January 1986.

F. Meyer. Color Image Segmentation. *Proceedings of the 4th Conference Image Processing and its Applications*, pp. 303-306, 1992.

I. Miyashiro, T. Senda, A. Matsumine, G.H. Baeg, T. Shimano, S. Muira, T. Noda, S. Kobayashi et M. Monden. Subcellular localization of the APC protein: immunoelectron microscopic study of the association of the APC protein with catenin. *Oncogene*, vol. 11: 89-96, 1995.

R. Montironi, L. Diamanti, D. Thompson, H.G. Bartels et P.H. Bartels. Analysis of the capillary architecture in the precursors of prostate cancer: recent findings and new concepts. *European journal of Urology*, vol. 30:191–200, 1996.

N.M. Moroney et M.D. Fairchaild. Color space selection for jpeg compression. *Journal of electronic imaging*, vol. 4(4): 373-381, 1995.

S. Mousses, L. Bubendorf, U. Wagner, G. Hostetter, J. Kononen, R. Cornelison, N. Goldberger, A.G. Elkahloun, N. Willi, P. Koivisto, W. Ferhle, M. Raffeld, G. Sauter et O.P Kallioniemi. Clinical Validation of Candidate Genes Associated with Prostate Cancer Progression in the CWR22 Model System using Tissue Microarrays. *Cancer Research*, vol. 62: 1256-1260, March 2002.

N. Mucci, G.Akdas, S.Manely et M.Rubin. Neuroendocrine expression in metastaticprostate cancer: Evaluation of high throughput tissue microarrays to detect heterogeneous protein expression. *Human Pathology*, vol. 31(4): 406-414, 2000.

T.O. Nielsen, F.D Hsu, J.X O'Connell, C.B. Gilks, P.H.B. Sorensen, L. Sabine, R.B. West, C.L. Liu, D. Botstein, P.O. Brown et M. DeRijn. Tissue microarray validation of epidermal growth factor receptor and SALL2 in synovial sarcoma with comparison to tumors of similar histology. *Am J Pathol*, vol.163: 1449-56, 2003.

B. Nilsson et A. Heyden. Segmentation of Complex Cell Clusters in Microscopic Images: Applicationto Bone Marrow Samples. *Cytometry*, vol. 66(A): 24–31, 2005.

A. Nocito, L. Bubendorf, E.M. Tinner, K. Suess, U. Wagner, T. Forster, J. Kononen, A. Fijan, J. Bruderer, U. Schmid, D. Ackerman, R. Maurer, G. Alund, H. Knonagel, M. Rist, M. Anabitarte, F. Hering, T. Hardmeier, A.J. Schoenenberger, R. Flury, P. Jager, J.L. Fehr, P. Schraml, H. Moch, M.J. Mihatsch, T. Gasser et G. Sauter. Microarrays of bladder cancer tissue are highly representative of proliferation index and histological grade. *J Pathol*, vol. 194: 349-57, 2001.

M.R. Nucci, C.R. Robinson, P. Longo, P. Campbell et S. Mamilton. Phenotypic and genotypic characteristics of aberrant crypt foci in human colorectal mucosa. *Hum Pathol*, vol. 28: 1396-1407, 1997

R. Ohlander. Analysis of natural scenes. Ph.D. Thesis, Carnegie-Mellon University, 1975.

R. Ohlander, K. Price et D.R. Reddy. Picture segmentation using a recursive region splitting method. *Computer Graphics and Image Processing*, vol. 8: 313–333, 1978.

Y.I. Ohta, T. Kanade, et T. Sakai. Color information for region segmentation. *Computer Graphics and Image Processing*, vol. 13: 222–241, 1980.

J.J. Orteu. Application de la vision par ordinateur à l'automatisation de l'abattage dans les mines. Thèse de doctorat, Université Paul Sabatier de Toulouse, novembre 1991.

M. Otsuka, M. Kato, T. Yoshikawa, H. Chen, E.J. Brown, Y. Masuho, M. Omata et N. Seki. Differential Expression of the L-Plastin Gene in Human Colorectal Cancer Progression and Metastasis. *Biochemical and Biophysical Research Communications*, vol. 289(4): 876-881, December 2001.

J. Packeisen, H. Buerger, R. Krech et W. Boecker. Tissue microarrays: a new approach for quality control in immunohistochemistry. *Journal of Clinical Pathology*, vol. 55(8): 613-615, 2002.

N.R. Pal et S.K. Pal. Review on Image Segmentation Techniques. *Pattern Recognition*, vol. 26(9): 1277-1294, 1993.

D.K. Panjwani et G. Healey. Markov Random Field Models for Unsupervised Segmentation of Textured Color Images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 17(10): 939-954, October 1995.

H.S. Park, R.A. Goodlad et N.A. Wright. Crypt fission in the small intestine and colon: A mechanism for the emergence of G6PD locus-mutated crypts after treatment with mutagens. *Am J Pathol*, vol. 147(5): 1416–1427, 1995.

S.Y. Park, H.S. Kim, E.K. Hong et W.H. Kim. Expression of cytokeratins 7 and 20 in primary carcinomas of the stomach and colorectum and their value in the differential diagnosis of metastatic carcinomas to the ovary. *Hum Pathol*, vol. 33: 1078-1085, 2002.

R.L. Parker, D.G. Huntsman, MD, D.W. Lesack, J.B. Cupples, D.R. Grant, M. Akbari et C.B. Gilks. Assessment of Interlaboratory Variation in the Immunohistochemical Determination of Estrogen Receptor Status Using a Breast Cancer Tissue Microarray. *American journal of clinical pathology*, vol. 117(5): 723-728, 2002.

C. Parrado, U.G. Falkmer, A. Höög, S. Falkmer, O. Ahrens, F. Dius et L. Grimelius, A technique for automatic/interactive assessment of proliferation fraction of neoplastic cells in solid tumors. *Gen Diagn Pathol*, vol. 141: 215-227, 1996.

K. N. Plataniotis et A. N. Venetsanoupoulos. Color image and applications. *Digital Signal Processing*. *Springer-Verlag*, 2000.

M. R. Pointer. A comparison of the cie 1976 colour space. *Color research and Application*, vol. 6(2): 108-118, 1981.

A. Pollack, M. DeSilvio, L.Y. Khor, R. Li, T.I. Al-Saleem, M.E. Hammond, V. Venkatesan, C.A. Lawron, M. Roach, W.U. Shipley, G.E. Hanks et H.M. Sandler. Ki-67 staining is a strong predictor of distant metastasis and mortality for men with prostate cancer treated with radiotherapy and androgen deprivation: Radiation Therapy Oncology Group Trial 92-02. *J Clin Oncol*, vol. 22: 2133-2140, 2004.

R.N. Poston et N.P. Gall. Hue-saturation-intensity colour image analysis for the quantitation of immunoperoxidase staining. *In: Elder HY, editor. Micro 90. Bristol: Adam Hilger*, pp. 525–528, 1990.

R.N. Poston. Colour image analysis for histology: the advantages of the hue-saturation-intensity mode. *Image processing*, pp. 4-8, 1996.

F. Potet et J. Souillard. Polyps of the rectum and colon. *Gut*, vol.12: 468-482, 1971.

W.K Pratt. « Digital Image Processing ». *John Wiley & Sons*, New-York, 1978.

T.P. Pretlow, B.J. Barrow, W.S. Ashton, M.A. O'riordan, T.G. Pretlow, J.A. Jurcisek et T.A. Stellato. Aberrant crypts. Putative preneoplastic foci in human colonic mucosa. *Cancer Res*, vol. 51(5): 1564-1567, 1991.

T.P. Pretlow, T.A. Brasitus, N.C. Fulton, C. Cheyer et E.L. Kaplan. K-ras mutations in putative preneoplastic lesions in human colon. *J Natl Cancer Inst*, vol. 85: 2004-7, 1993.

P. Pujas. Analyse d'images couleur et fusion d'images 3D et couleur. Thèse de doctorat, Université de Montpellier II Sciences et Techniques du Languedoc, février 1996.

A. Rabinovich, S. Krajewski, M. Krajewska, A. Shabaik, S.M. Hewitt, S. Belongie, J.C. Reed et J.H. Price. Framework for Parsing, Visualizing and Scoring Tissue Microarray Images. *IEEE Transactions On Information Technology In Biomedecine*, vol. 10(2), April 2006.

G. Raffy. Vision numérique couleur appliquée à l'évaluation de la qualité de carcasses de dindes. Thèse de doctorat, Université Blaise Pascal de Clermont-Ferrand, mars 1999.

V. Rakotomalala. Reconstruction bidimensionnelle de vaisseaux rétiniens par analyse d'images couleur de fond d'oeil. Thèse de doctorat, Université des Sciences et Technologies de Lille, décembre 1999.

P. Ranefall, K. Wester et E. Bengtsson. Automatic quantification of immunohistochemically stained cell nuclei using unsupervised images analysis. *Anal Cell Pathol*, vol.16 (1): 29-43, 1998.

P. Ranefall. Towards automatic quantification of immunohistochemistry using color image analysis. Ph.D. Thesis, Uppsala University 1998.

E. Raymond, M. Raphaël, M. Grimaud, L. Vincent, J.L. Binet et F. Meyer. Germinal center analysis with the tools of mathematical morphology on graphs. *Cytometry,* 1993.

A.R. Robertson. Historical development of cie recommended color difference equations. *Color research and Application,* vol. 15: 170-197, 1990.

G.S. Robinson. Color edge detection. *Optical Engineering*, vol. 16(5): 479–484, 1977.

V. Rodrigues. L'analyse du mouvement dans les images couleurs. Thèse de doctorat d'état, Université Paul Sabatier de Toulouse, novembre 1985.

D.F. Rogers. Procedural elements for computer graphics. *Mc Graw Hill*, 1985.

M. Romeo, F. Burden, M. Quinn, B. Wood et D. McNaughton. Infrared microspectroscopy and artificial neural networks in the diagnosis of cervical cancer. *Cellular Molecular Biology*, vol. 44: 179–187, 1998.

L. Roncucci, D. Stamp, A. Medline, J.B. Cullen et W.R. Bruce. Identification and quantification of aberrant crypt foci and microadenomas in the human colon. *Hum Pathol*, vol. 2: 387-394, 1991.

A. Rosenfeld et A. Kak, Digital Picture Processing. *Academic Press*, vol. 2, New York, NY, 1982.

A.J. Round, A.W.G. Duller et P.J. Fish. Colour Segmentation for Lesion Classification. *19th Annual Conference of the IEEE Engineering in Medicine and Biology Society,* vol. 2: 582-585, Chicago, Illinois, 30 Oct.-2 Nov. 1997.

M.A. Rubin, M.P. Zerkowski, R.L. Camp, R. Kuefer, M.D. Hofer, A.M. Chinnaiyan et D.L. B. Rubinfeld, B. Souza, I. Albert, O. Muller, S.H. Chamberlain, F.R. Masiarz, S. Munemitsu et P. Polakis. Association of the APC gene product with b-catenin. *Science*, vol. 262: 1731-4, 1993.

M.A. Rubin, R. Dunn, M. Strawderman et K.J. Pienta**.** Tissue microarray sampling strategy for prostate cancer biomarker analysis. *Am J Surg Pathol*, vol. 26(3): 312-9, Mar 2002.

M.A. Rubin, M.P. Zerkowski, R.L. Camp, R. Kuefer, M.D. Hofer, A.M. Chinnaiyan et D.L. Rimm. Quantitativedetermination of expression of the prostate cancer protein alpha-methylacyl-CoA racemase using automated quantitative analysis (AQUA): a novel paradigm for automated and continuous biomarker measurements. *Am J Pathol*, vol. 164: 831–840, 2004.

A.C. Ruifrok. Quantification of immunohistochemical staining by color translation and automated thresholding. *Anal Quant Cytol Histol*, vol. 19: 107–113, 1997.

A.C. Ruifrok et D.A. Johnston. Quantification of histochemical staining by color deconvolution. *Anal Quant Cytol Histol*, vol. 23: 291–299, 2001.

A.C. Ruifrok, R.L. Katz et D.A. Johnston. Comparison of quantification of histochemical staining by Hue-Saturation-Intensity (HSI) transformation and Color-Deconvolution. *Applied Immunohistochemistry and Molecular Morphology*, vol. 11(1): 85-91, Marsh 2003.

K. Saarinen. Color Image Segmentation. *Internal Report*, 1995.

E. Salminen, S. Palmu, T. Vahlberg, P.J. Roberts et K.O. Söderström. Increased proliferation activity measured by immunoreactive Ki67 is associated with survival improvement in rectal/recto sigmoid cancer. *World J Gastroenterol*, vol. 11(21): 3245-3249, 2005.

G. Sapiro. Vector (Self) Snakes: A Geometric Framework for Color, Texture, and Multiscale Image Segmentation. The International Conference on Image Processing, vol. 1: 817-820, Lausanne, Switzerland, 16-19 Sept. 1996.

G. Sapiro. Color Snakes. *Computer Vision and Image Understanding*, vol. 68(2): 247-253, Nov. 1997.

A. Sarabi et J.K. Aggarwal. Segmentation of chromatic images. *Pattern Recognition*, vol. 13: 417–427, 1981.

G. Sauter, R. Simon et K. Hillan. Tissue microarrays in drug discovery. *Nat Rev Drug Discov*, vol. 2(12): 962-72, Dec 2003.

M.H. Savoji et R.E. Burge. Note on differents methods based on the Karhunen-Loeve expansion and used in image analysis. *Computer Vision, Graphics, and Image Processing*, vol. 29: 259–269, 1985.

B. Schacter, L. Davis et A. Rosenfeld. Scene segmentation by cluster detection in color space. Department of Computer Science, University of Maryland, College Park, MD, Nov 1975.

R. Schettini et M. Suardi. A Low-level Segmentation Procedure for Color Images. *7th Conference of European Signal Processing*, vol. 1: 26-29, Lausanne, Switzerland, 13-16 September 1994.

R.J. Schlemper, M. Itabashi, Y. Kato, K.J. Lewin, R.H. Riddell, T. Shimoda, P. Sipponen, M. Stolte et H. Watanabe. Differences in the diagnostic criteria used by Japanese and Western pathologists to diagnose colorectal carcinoma. *Cancer*, vol. 82: 60-69, 1998.

R. Schneider-Stock, C. Boltze, J. Lasota, B. Peters, C.L. Corless, P. Ruemmele, L. Terracciano, M. Pross, L. Insabato, D. Di Vizio, I. Iesalnieks, S. Dirnhofer, A. Hartmann, M. Heinrich, M. Miettinen, A. Roessner et L. Tornillo. Loss of p16 Protein Defines High-Risk Patients with Gastrointestinal Stromal Tumors: A Tissue Microarray Study. *Clinical Cancer Research*, vol. 11: 638–645, January 15, 2005.

S. Schüpp. Application en imagerie microscopique biomédicale. Thèse de doctorat, Université de Caen, décembre 2000.

T. Seidal, A.J. Balaton et H. Battifora. Interpretation and quantification of immunostains. *Am J Surg Pathol*, vol.25: 1204–1207, 2001.

D. Seigneurin, O. Cohen et J. Louis. Imminocytochemistry and image cytometry of progesterone receptor in breast carcinomas imprints. *Anal Cell Pathol*, vol. 1: 97-104, 1989.

J. Serra. Image Analysis and Mathematical Morphology. *Academic Press, Londres*, 1982.

L. Shafarenko, M. Petrou et J.V. Kittler. Histogram based segmentation in a perceptually uniform color space. *IEEE Trans. Image Processing*, vol. 7 (9): 1354-1358, Sept 1998.

G. Sharma et H.J. Trusell. Digital color imaging. *IEEE Transactions on Image Processing*, vol. 6(7): 901-932, 1997.

B. Shpitz, Y. Bornstein, Y. Mekori, R. Cohen, Z. Kaufman, D. Neufeld, M. Galkin, et J. Bernheim. Aberrant crypt foci in human colons: distribution and histomorphologic characteristics. *Hum Pathol*, vol. 29: 469-475, 1998.

R. Simon et G. Sauter. Tissue microarrays for miniaturized high-throughput molecular profiling of tumors. *Exp Hematol*, vol. 30: 1365-1372, 2002.

R. Simon, M. Mirlander et G. Sauter. Tissue microarrays in cancer diagnosis. *Expert Rev. Mol. Diagn.*, vol. 3(4): 421-430, 2003.

J. Simony. Bouleversement architecturaux induits dans la muqueuse colique normale et tumorale par la transformation maligne et la progression tumorale : Approche morphologique. Thèse de doctorat, Université Joseh Fourier, Grenoble, France, 24 Octobre 2007.

S.S. Singh, B. Qaqish , J.L. Johnson , O.H. Ford, J.F. Foley , S.J. Maygarden et J.L. Mohler. Sampling strategy for prostate tissue microarrays for Ki-67 and androgen receptor biomarkers. *Anal Quant Cytol Histol.*, vol. 26(4): 194-200, Aug 2004.

M. Skacel, B. Skilton, J.D. Pettay et R.R. Tubbs. Tissue microarrays: a powerful tool for high-throughput analysis of clinical specimens: a review of the method with validation data. *Appl Immunohistochem Mol Morphol*, vol. 10(1): 1-6, Mar 2002.

M. Skacel, A. Siva, B. Xu et R.R. Tubbs. From array to array: Confirmation of genomic gains and losses discovered by array-based comparative genomic hybridization utilizing fluorescence in situ hybridization on tissue microarrays. *J Mol Hist*, vol. 38: 135–140, 2007.

W. Skarbek et A. Koschan. Colour Image Segmentation: A Survey. *Technical University of Berlin, Germany, Technical Report*: 94-32, 1994.

R.H. Smallwood, W.M.L. Holcombe et D.C. walker. Development and validation of computational models of cellular interaction. *J Mol Histol*, vol. 35: 659-665, 2004.

G.B. Smejkal et J.R. Shainoff. Enhanced digital imaging of diaminobenzidene-stained immunoblots. *Biotechniques*, vol. 22(462), 1997.

G. Smith , F.A. Carey , J. Beattie , M.J. Wilkie , T.J. Lightfoot , J. Coxhead , R.C. Garner , R.J. Steele et C.R. Wolf . Mutations in APC, Kirsten-ras and p53-alternative genetic pathways to colorectal cancer. *PNAS : Proc Natl Acad Sci U S A*, vol. 99(14): 9433-8., 2002.

D.G. Soenksen. Automated Microscopic Inspection of Tissue Microarrays Using Virtual Microscopy. *Feature article*, 2003.

J.K. Sont, W.I. De Boer, W.A.A.M. Van Schadewijk, K. Grunberg, J.H.J.M Van Krieken, P.S. Hiemstra et P.J. Sterk. Fully automated assessmant of inflammatory cell counts and cytokine expression in bronchial tissue. *AM J Respir Crit Care Med*, vol. 67: 1496-1503, 2003.

M. Srivastava, L. Bubendorf, V. Srikantan, L. Fossom, L.Nolan, M. Glasman, X. Leighton, W. Fehrlei, S. Pittalugai, M. Raffeldi, P. Koivisto, N. Willi, T.C. Gasser, J. Kononen, G. Sauter, O.P. Kallioniemi, S. Srivastava et H.B. Pollard. ANX7, a candidate tumor suppressor gene for prostate cancer. *Proc Natl Acad Sci U S A: PNAS*, vol. 98(8): 4575-80, Apr 2001.

A. Stockman et L.T. Sharpe. The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision research*, vol. 4: 1711-1737, 2000.

O. Straume, L. Sviland et L.A. Akslen. Loss of nuclear p16 protein expression correlates with increased tumor cell proliferation (Ki-67) and poor prognosis in patients with vertical growth phase melanoma. *Clin Cancer Res*, vol. 6: 1845–1853, 2000.

J. Sudbo, R. Marcelpoil et A. Reith. Caveats : Numerical requirements in graph theory based quantitation of tissue architecture. *Analytical Cellular Pathology*, vol. 21: 59-69, 2000a.

J. Sudbo, R. Marcelpoil et A. Reith. New algorithms based on the Voronoi diagram applied in a pilot study on normal mucosa and carcinomas. *Analytical Cellular Pathology*, vol. 21: 71-86, 2000b.

J. Sudbo, A. Bankfalvi, M. Bryne, R. Marcelpoil, M. Boysen, J. Piffko, J. Hemmer, K. Kraft et A. Reith. Prognostic Value of Graph Theory-Based Tissue Architecture Analysis in Carcinomas of the Tongue. *Lab Invest*, vol. 80: 1881-1889, 2000c.

M. Sugita, M. Geraci, B. Gao, R.L. Powell, F.R. Hirsch, G. Johnson, R. Lapadat, E. Gabrielson, R. Bremnes, P.A. Bunn et W.A. Franklin. Combined Use of Oligonucleotide and Tissue Microarrays Identifies Cancer/Testis Antigens as Biomarkers in Lung Carcinoma. *Cancer Research*, vol. 62: 3971-3979, July 2002.

K. Takahashi, H. Nakatani et K. Abe. Color image segmentation using ISODATA clustering method. *Second Asian Conference on Computer Vision*, vol. 1: 523–527, Singapore, 1995.

T. Takayama, S. Katsuki, Y. Takahashi, M. Ohi, S. Nojiri, S. Sakamaki, J. Kato, K. Kogawa, H. Miyake et Y. Niitsu. Aberrant crypt foci of the colon as precursors of adenoma and cancer. *N Engl J Med*, vol. 339: 1227-1284, 1998.

B. Terris et M.P. BRALET. Tissue microarrays : Ou l'arrivée des puces en anatomie pathologique. *Annales de pathologie*, vol. 22(1): 69-72, 2002.

O. Tetsu et F. McCormick. Betacatenin regulates expression of cyclin D1 in colon carcinoma cells. *Nature*, vol. 398: 422-6, 1999.

F. Tezuka, I. Sato, H. Higashiiwai, N. Endo, K. Ito et M. Kasai. Method for the Quantitative Evaluation of the Distribution Pattern of Nuclei in Normal and Malignant Endometrial Epithelia. *Analyt. Quant. Cytol. Histo*, vol. 12: 237-241, 1990.

L.E. Tisell, A. Oden, A. Muth, G. Altiparmak, J. Molne, H. Ahlman et O. Nilsson. The Ki67 index a prognostic marker in medullary thyroid carcinoma. *Br J Cancer*, vol. 89: 2093-2097, 2003.

TNM. Classification des tumeurs malignes. *6° édition Paris : Cassini*, p. 63-106, 2003.

S. Tominaga. Color image segmentation using three perceptual attributes.*Conference of Computer Vision and Pattern Recognition*, pp. 628-630. Miami Beach, Florida, USA, 22-26 June 1986.

S. Tominaga. A colour classification method for color images using a uniform color space. *10th International Conference on Pattern Recognition*, vol. 1: 803-807, Atlantic City, New Jersey, 16-21 June 1990.

S. Tominaga. Color classification of natural color images. *Color Research and Application*, 17(4): 230–239, 1992.

S. Tominaga. Dichromatic reflection models for rendering object surfaces. *Journal of Imaging Science and Technology*, vol. 40(6): 549-555, December 1996.

J. Torhorst, C. Bucher, J. Kononen, P. Haas, M. Zuber, O.R. Kochli, F. Mross, H. Dietrich, H. Moch, M. Mihatsch, O.P. Kallioniemi et G. Sauter. Tissue microarrays for rapid linking of molecular changes to clinical endpoints. *J Pathol*, vol. 159: 2249–2256, 2001.

G.T Toussaint. Pattern recognition and geometrical complexity. In *Proceedings of the 5th International Conference on Pattern Recognition*, IEEE Catalog. 80, pp.1324-1347, 1980.

A. Trémeau et N. Borel. A Region Growing and Merging Algorithm to Color Segmentation. *Pattern Recognition*, vol. 30(7): 1191-1204, July 1997.

A. Trémeau. Analyse d'images couleurs : du pixel à la scène. Thèse de doctorat, Université Jean Monnet, Saint-Etienne, France, 1998.

A. Trémeau, C. Fernandez-Maloigne et P. Bonton. Image numérique couleur : De l'acquisition au traitement : Cours et applications. *Editions Dunod*, Février 2004.

A. Trouvé. La mesure de la couleur : principes, technique et produits du marché. *Afnor-Cetim*, 1989.

D.C. Tseng et C.H. Chang. Color segmentation using perceptual attributes. *11th Inernational Conerence. on Pattern Recognition*, vol. 3: 228-231, Den Hague, Netherlands, 30 Aug.-3 Sept. 1992.

J. Tucker, U. Schenck, W. Eiermann et G. Burger. Quantitative image analysis on immunocytochemically labeled cells. *Clinical cytometry and histometry. San Diego: Academic Press*, pp. 402–406, 1987.

S.E. Umbaugh, R.H. Moss, W.V. Stoecker et G.A. Hance. Automatic colour segmentation algorithms with application to skin tumor feature identification. *IEEE Engineering in Medicine and Biology*, vol. 12(3): 75-82, 1993.

R.A. Underwood, N.S. Gibran, L.A. Muffley, M.L. Usui et J.L. Olerud. Color subtractive-computer-assisted image analysis for quantification of cutaneous nerves in a diabetic mouse model. *J Histochem Cytochem*, vol. 49: 1285–1291, 2001.

N. Vandenbroucke, L. Macaire et J.-G. Postaire. Color pixels classification in an hybrid color space. In *IEEE International Conference on Image Processing*, vol. 1: 176–180, Chicago, 1998.

N. Vandenbroucke, Segmentation d'images couleur par classification de pixels dans des espaces d'attributs colorimétriques adaptés. Application à l'analyse d'images de football. Thèse de doctorat, Université de Lille 1, France, décembre 2000.

J. Van der Laak, M Pahlplatz, A.. Hanselaar et P. de Wilde. Hue-Saturation-Density (HSD) Model for Stain Recognition in Digital Images From Transmitted Light Microscopy. *Cytometry*, vol. 39: 275–284, 2000.

J. Van Der Laak. Automated Identification of Cell and Tissue Components in Pathology. Ph.D Thesis. University of Nijmegen, The Netherlands, 2001.

F. Viénot. Colorimétrie et image couleur. *Ecole de printemps-images couleurs*, Pau, GDR-ISIS, pp. 2-3, Mars 2001.

H.D. Von Stein et W. Reimers. Segmentation of color pictures with the aid of color information and spatial neighbourhoods. *Signal processing II: theories and applications, Elsevier Science Publ*., pp. 271-273, North-Holland, Amsterdam, Netherlands, 1983.

M.J. Vrhel, R. Gershon et L.S. Iwan. Measurement and analysis of object reflectance spectra. *Color Research and Application*, vol. 19(1): 4-9, 1994.

H. Vrolijk, W. Sloos, W. Mesker,P. Franken, R. Fodde, H. Morreau et H. Tanke . Automated acquisition of stained tissue microarrays for high-throughput evaluation of molecular targets. *J Mol Diagn*, vol. 5: 160–167, 2003.

A. Watanabe, R. Cornelison et G. Hostetter. Tissue microarrays: applications in genomic research. *Expert Rev. Mol. Diagn,* vol. 5(2): 171-181; 2005.

A. Waterworth, K. Horgan, V. Speirs et A.M. Hanby. Tissue microarrays: big potentialfrom small samples. *International Journal Of Oncology*, vol. 25: 167-171, 2004.

J.R. Weaver et J.L.S. Au. Application of automatic thresholding in image analysis scoring of cells in human solid tumors labeled for proliferation markers. *Cytometry*; vol.29: 128–135, 1997.

M.R. Wick et S.E. Mills. Consensual interpretive guidelines for diagnostic immunohistochemistry. *Am J Surg Pathol*, vol. 25: 1208–1210, 2001.

T.H. Wonnacott et R.J. Wonnacott. Statistique : économie-gestion-sciences-médecine. *Economica Paris,* 4ème édition 1995.

W.D. Wright. The sensitivity of the eye to small color differences. *Proceedings of Physical Society*, vol. 53(2): 93-112, 1941.

J. Wu, H. Yan, et A.N. Chalmers. Color Image Segmentation Using Fuzzy Clustering and Supervised Learning. *Journal of Electronic Imaging*, pp. 397-403, Oct. 1994.

A.W. Wu, J. Gu, J.F. Ji, Z.F.Li et G.W. Xu. Role of COX-2 in carcinogenesis of colorectal cancer and its relationship with tumor biological characteristics and patients' prognosis. *World Journal of Gastroenterology*, vol. 9(9): 1990-1994, 2003.

G. Wyszecki et W. S. Stiles. Color science, *Wiley*, New York, 1982.

F. Yang et T. Jiang. Cell Image Segmentation with Kernel-Based Dynamic Clustering and an Ellipsoidal Cell Shape Model. *Journal of Biomedical Informatics*, vol. 34: 67–73, 2001.

Q. Zheng, B.K. Milthorpe et A.S. Jones. Direct Neural Network Application for Automated Cell Recognition. *Cytometry*, vol. 57(A): 1–9, 2004.

S.C. Zhu et A.L. Yuille. Region competition. *IEEE Trans. Pattern Analysis and Machine Intelligenc*, vol. 18(9): 884-900, 1996.

D. Zugaj et V. Lattuati. A new approach of color images segmentation based on fusing region and edge segmentation outputs. *Pattern Recognition*, vol. 31(2): 105-113, 1998.

L. Zukerberg. The Molecular Basis of Dysplasia. *Sem Diagn Pathol*, vol. 19: 48-55, 2002.

Publications

- **R. Heus**, M.P. Montmasson, J. Simony-Lafontaine et F. Giroud. Concevoir des puces à tissus TMA : Méthodologie d'échantillonnage. *10ième congrès annuel de l'Association Française de Cytométrie, In : Actes de la conférence AFC 05 : pp.22, Grenoble, France*, 4-7 octobre 2005.

- J. Simony-Lafontaine, M.P. Montmasson, **R. Heus**, J. Bourbeillon, L. Khadari et F. Giroud. La plateforme « tmas-explorer » : Un outil d'aide à la validation et à l'analyse des tissue-arrays pour le praticien. *Annales de Pathologie*, 24(1) :164, 2004.

- F. Giroud, J. Simony-Lafontaine, M.-P. Montmasson, **R. Heus**, J. Bourbeillon et C. Garbay. The TMAs-Explorer Platform: an integrated tool including a virtual TMA concept. *In: Abstract Book of the 7th European Congress on Virtual Microscopy and 1st International Congress on Virtual Microscopy. Poznan, Poland*, July 8-11, 2004.

- **R. Heus**, M-P. Montmasson, J. Simony-Lafontaine, F. Giroud. To design microarrays: Methodology of sampling. Article soumis une première fois pour publication au journal *Cellular oncology* (version courte à ICISP 2008). Cet article est en cours de modification selon les suggestions proposées par les reviewers pour une nouvelle soumission.

- Une procédure est en cours auprés de l'Agence pour la Protection des Programmes (APP) pour un transfert de technologie autour du logiciel développé pour réaliser notre approche, d'analyse automatique d'images histologiques, présentée dans le septième chapitre.

Annexe A Immunohistochimie : IHC

A.1 Introduction :

L'immunohistochimie (IHC) est un mot composé : du grec histos, tissu et khêmê, qui fait référence à un produit chimique. L'IHC consiste à révéler sur une coupe histologique divers sites antigéniques (Ag) mis en évidence par des anticorps (Ac) spécifiques dirigés contre eux. Les Ag recherchés sont des constituants antigéniques de molécules nucléaires, cytoplasmiques ou membranaires. La réaction ou liaison Antigène-Anticorps (Ag-Ac) est mise en évidence grâce à un procédé colorimétrique : un traceur émettant un signal couleur est fixé directement ou indirectement sur l'anticorps ; cette révélation couleur fait généralement appel à une réaction enzymatique (ex. peroxydase). La figure A.1 résume le principe de l'IHC.

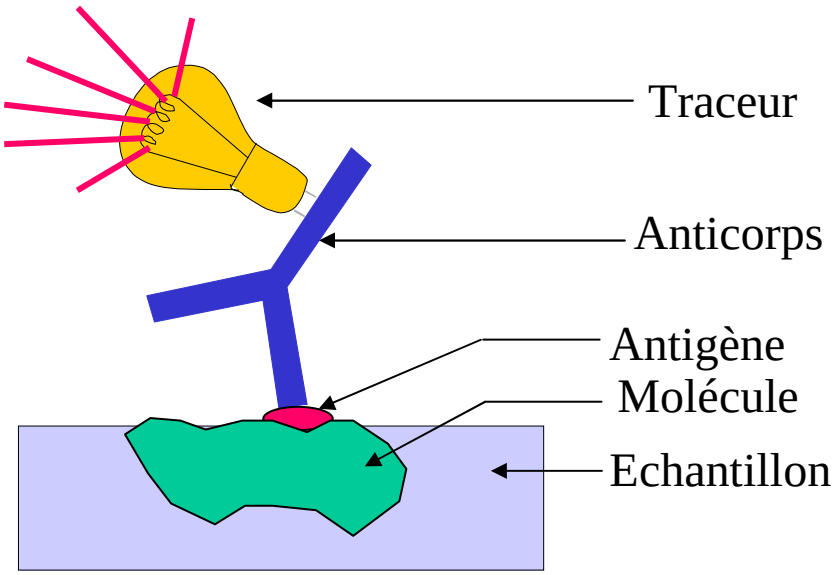


Figure A.1 : Principe d'une réaction immunohistochimique.

La technique d'IHC trouve ses origines dans les débuts des années 1940, pour la première fois des antigènes cellulaires étaient révélés par des anticorps marqués par un traceur fluorescent [Coons et al 1941]. Aujourd'hui, l'IHC est largement diffusée et est devenue une technique de routine dans les laboratoires d'anatomie pathologique et de recherche à des fins multiples parmi lesquelles :

- la classification des tumeurs pour établir le diagnostic par la mise en évidence des antigènes de différenciation cellulaire,
- la quantification de certains marqueurs de prolifération pour dresser le pronostic de la tumeur étudiée.

La réaction immunohistochimique comporte trois étapes fondamentales :

1- la préparation du tissu contenant l'antigène recherché,
2- le choix de l'anticorps dirigé contre l'antigène recherché,
3- le système de révélation du complexe antigène-anticorps.

Ces trois éléments sont détaillés dans ce qui suit :

A.2 La préparation des tissus à étudier :

Il est peu commun de marquer les spécimens tissulaires avec des anticorps directement sans préparation préalable. Les échantillons tissulaires doivent subir le plus souvent une longue série de traitements avant l'immunomarquage proprement dit. Il existe un grand panel de techniques de préparation des tissus, néanmoins la technique de base comporte plusieurs étapes : la fixation du tissu, l'inclusion dans la paraffine, la réalisation des coupes et enfin l'application de l'immunomarquage.

A.3 La fixation :

Elle est primordiale pour conserver la morphologie cellulaire et figer les structures des tissus étudiés. La fixation permet de garder les antigènes à marquer dans leur environnement naturel. Elle doit être effectuée juste après le prélèvement des tissus afin que les structures ne soient pas endommagées. En dépit des avantages de la fixation, elle peut conduire néanmoins à une modification de l'affinité des antigènes pour leurs anticorps spécifiques et peut rendre le tissu fixé moins perméable ce qui peut compromettre la diffusion des anticorps vers leurs cibles (antigènes). Les produits chimiques utilisés pour la fixation peuvent inhiber la liaison anticorps-antigènes [Brandtzag 1984].

Il n'existe pas d'agent chimique de fixation idéal ayant la capacité de conserver tous les constituants tissulaires (protéines, lipides …..). Le plus souvent, on a recours à un mélange de différents fixateurs, notons que la fixation est pratiquée à froid pour éviter l'autolyse des tissus.

Parmi les fixateurs utilisés en IHC, on retrouve les fixateurs classiques utilisés en histologie et les fixateurs développés spécialement pour l'IHC ; le traitement par la chaleur est considéré actuellement comme un fixateur idéal en IHC.

A.3.1 Les Fixateurs classiques :

Les déshydratation : si l'on soustrait l'eau des cellules, les constituants cellulaires (le plus souvent les protéines) adhèreront entre eux figeant ainsi la structure des cellules. Il existe plusieurs façons d'effectuer la déshydratation, on peut citer :

- **la chaleur** : les blocs de tissus sont déshydratés dans un four micro-onde ou dans un bain-marie,
- **les alcools et les cétones** : c'est des solvants miscibles à l'eau ; les blocs de tissus sont plongés dans ces solvants avec un gradient de concentrations croissantes de ces solvants jusqu'à élimination totale de l'eau. L'éthanol est un agent de fixation couramment employé. L'acétone et le méthanol qui sont moins utilisés.

A.3.2 Les Fixateurs développés pour l'IHC :

L'utilisation des fixateurs classiques rend les tissus peu perméables à la liaison antigène-anticorps ; ce qui provoque la diminution fréquente de la formation de complexes antigène-anticorps. Plusieurs équipes de recherche [Pollak et al 1972] [Pearse et al 1974] ont mené des recherches pour développer des fixateurs permettant une meilleure conservation de l'immunoréaction des tissus fixés. Les principaux fixateurs utilisés sont :

- **Le carbodiimide** : développé par [Pollak et al 1972], il préserve les tissus de façon excellente,

- **Le périodate-lysine-formaldéhyde** : qui fixe les sucres et les protéines en donnant des résultats remarquables pour les observations en microscopie photonique,
- **Le zinc** : meilleur que les fixateurs histologiques classiques pour la préservation des antigènes,
- **la chaleur** : pendant longtemps, elle a été considérée comme nocive pour la préservation des blocs de tissus. Des travaux menés depuis quelques années [Garaud et Roussel 2007] ont démontré que la chaleur est un excellent fixateur pour les réactions d'immunomarquage. La chaleur peut être fournie par différentes méthodes : four micro-onde, autoclave ou bain-marie. Les micro-ondes pénètrent rapidement en profondeur les blocs de tissus stoppant ainsi l'autolyse des tissus ; ce qui permet de réduire le temps de fixation ainsi que la concentration des agents fixants chimiques. Le recours à l'autoclave donne des résultats équivalents ; cela permet même une meilleure conservation de la morphologie des noyaux cellulaires.

De manière générale, pour garder la perméabilité des tissus et pour favoriser la diffusion des anticorps vers leurs antigènes, nous pouvons s'orienter vers des fixations douces c'est à dire en diminuant les concentrations des fixateurs, leur temps de fixation ou en combinant différents fixateurs comme, par exemple, la chaleur et un agent chimique en concentration réduite. Nous pouvons se permettre une fixation incomplète dans le but de conserver la perméabilité des tissus. Des traitements perméabilisant à base de solvants ou de détergents peuvent être utilisés après la fixation.

A.4 L'inclusion en paraffine :

Les blocs de tissus ainsi fixés sont dans la plupart des cas des tissus mous, rendant difficile l'obtention de coupes fines de l'ordre de 4-5 μm pour les observations microscopiques. Il faut au préalable rendre la consistance des blocs de tissus plus solide pour réussir à obtenir des coupes suffisamment fines. Généralement, les tissus sont imprègnés dans un milieu d'inclusion liquide. Ce milieu s'infiltre dans les tissus et se solidifie en se refroidissant rendant les tissus plus solides. Une fois les tissus solidifiés, ils peuvent être débités en coupes fines à l'aide d'un microtome.

La paraffine est un milieu d'inclusion hydrophobe. Une étape de déshydratation est obligatoire avant de procéder à l'inclusion. La technique la plus courante consiste à déshydrater les tissus par passage dans des concentrations croissantes d'agents de déshydratation miscibles à l'eau comme l'éthanol. Notons que la déshydratation est réalisée à température ambiante. Un bain supplémentaire dans un solvant comme le toluène ou le xylène, est nécessaire pour éliminer toute trace d'eau avant l'imprégnation dans le bain de paraffine. Les tissus sont alors plongés dans un bain de paraffine liquide ramené à une température qui varie entre 56°c et 60°c. L'inclusion est réalisée dans des moules ou cassettes en plastique, le bloc de tissu imprégné de paraffine est solidifié à température ambiante.

A.5 La réalisation des coupes :

L'observation des réactions immunohistochimiques sous microscope en lumière transmise nécessite la réalisation de coupes relativement transparentes. Les coupes doivent être suffisamment fines afin de permettre à la pénétration de la lumière transmise par le microscope. Notons que la consistance de la paraffine et sa nature cristalline rend difficile la réalisation de coupes d'épaisseur inférieure à 4 µm ; on utilise le plus souvent un microtome pour réaliser des coupes d'environ 4 à 5 µm d'épaisseur. Il existe plusieurs types de microtome : microtomes à chariot coulissant, à bascule et rotatifs. Les microtomes rotatifs sont les plus courants. Depuis peu, un microtome à laser est commercialisé (www.rowiak.com) permettant la réalisation de coupes d'épaisseur régulière.

Il arrive que des coupes soient plissées, comprimées ou fragmentées, malgré l'utilisation d'un microtome de dernière génération. Des films de transfert adhésif sont disponibles pour obtenir de belles coupes [Kawamoto 1986, 1990]. La technique consiste à déposer le film adhésif sur la surface du bloc de tissu à débiter. Ce film, sous forme de scotch, consolide la coupe lors de sa section par le microtome et permet son transfert sur la lame sans dégâts. Notons que les blocs de tissus présentent, dans la plupart des cas, une surface rugueuse et en reliefs ; il est alors nécessaire de débiter plusieurs fois le bloc de tissus pour obtenir une surface réellement plane, avant de procéder à la première coupe tissulaire proprement dite.

A.6 Le traitement des coupes :

La paraffine est un milieu d'inclusion hydrophobe. Sans traitement préalable des coupes, les anticorps ont beaucoup de mal à se lier aux antigènes. Il faut impérativement dissoudre la paraffine pour que la surface des coupes retrouve sa nature hydrophile. La paraffine se dissout parfaitement dans de nombreux solvants aromatiques, tels que le toluène ou l'éther de pétrole. La dissolution du milieu d'inclusion est suivie généralement par un traitement perméabilisant pour favoriser l'accessibilité des antigènes. Ce traitement appelé communément « Démasquage des sites antigéniques » se fait de plusieurs façons par :

- **digestion enzymatique** : ce traitement peut dégager un espace autour des antigènes pour mieux exposer leurs épitopes aux anticorps. Parmi les enzymes utilisées, on peut citer : la pronase [Finly et al 1978], la trypsine [Curaan et gregory 1978] et la pepsine [Reading 1977],
- **traitements dénaturants** : des traitements à base d'agents dénaturants tels que le SDS (Sodium Dodecyl Sulfate) [Robinson et Vandre 2001] ou le thiocyanate de guanidine peuvent améliorer le marquage,
- **traitements par la chaleur** : La technique de démasquage antigénique par la chaleur [Shi et al 1991] renforce et remplace le traitement enzymatique. L'utilisation de la chaleur est devenue une pratique systématique pour le prétraitement des coupes avant l'immunoréaction ; la chaleur permettant la rupture des liaisons établies. Différentes sources de chaleur sont utilisées : micro-onde, bain-marie, autoclave ou cocotte-minute, aboutissant au même résultat.

A.7 L'immunomarquage :

Nous avons survolé les différentes étapes de traitement des tissus nécessaires avant d'effectuer l'immunomarquage proprement dit. La mise en évidence de la molécule ou l'antigène recherché sur une coupe histologique consiste à utiliser un anticorps spécifique capable de se lier et d'interagir avec cet antigène. La réaction ou la liaison antigène-anticorps est détectable grâce à un traceur fixé directement ou indirectement à l'anticorps en émettant un signal couleur. L'immunomarquage nécessite donc trois éléments : l'antigène, l'anticorps et le traceur ou le système de révélation :

A.7.1 Les antigènes :

C'est la molécule cible du marquage. En effet, le but essentiel des réactions immunohistochimiques est l'étude de la distribution de ces antigènes à la surface des coupes. L'antigène est une macromolécule capable d'engendrer la formation d'anticorps qui réagissent spécifiquement avec lui. Les antigènes peuvent être des protéines, des polysaccharides ou des lipides. Ceux utilisés en immunohistochimie sont généralement des protéines. On peut citer comme antigène la protéine Ki67, étudiée au cours de cette thèse. La protéine Ki67 est nucléaire et s'exprime pendant toute la durée des phases de prolifération cellulaire. Elle est donc absente au cours de la phase du repos cellulaire : G0. L'immunomarquage de la protéine Ki67 permet le calcul du % des cellules en cycle au niveau des coupes histologiques ; sa quantification nous renseigne notamment sur l'activité de prolifération cellulaire d'un tissu donné. Les antigènes doivent conserver leur antigénicité et résister aux différents traitements que subissent les tissus pendant les multiples étapes de l'immunomarquage. Notons que, dans le cas des antigènes protéiques, l'anticorps reconnaît une partie seulement de l'antigène nommée épitope ou déterminant antigénique, l'antigène pouvant comporter plusieurs épitopes.

A.7.2 Les anticorps :

Les anticorps sont des glycoprotéines de la famille des immunoglobulines utilisées par le système immunitaire pour se lier aux antigènes de façon spécifique. Les anticorps sont secrétés par des cellules issues des lymphocytes B. Ils ont une structure sous forme d'un Y, la partie terminale (pied du Y) a une structure constante. Par contre, les deux extrémités (les deux bras du Y) possèdent des domaines très variables. Ce sont ces parties extrêmement variables qui constituent le site de reconnaissance (paratope) des antigènes. Cette reconnaissance est mise à profit pour détecter divers antigènes (ex. marqueurs tumoraux) en immunohistochimie. Un anticorps est caractérisé par 4 paramètres principaux : son titre (capacité), sa spécificité, son affinité et son avidité :

- **titre** : ou capacité, c'est la concentration de l'anticorps dans une solution (en mg/ml),
- **spécificité** : l'anticorps doit être spécifique à l'épitope qu'il reconnait au niveau de l'antigène,

- **affinité** : reflète la force de liaison entre l'anticorps avec son épitope spécifique, elle traduit la stabilité du complexe Ag-Ac,
- **avidité** : représente le nombre de déterminants antigéniques ou épitopes fixés par un anticorps sur le même antigène.

Un antigène possède le plus souvent plusieurs épitopes différents. On peut classer les anticorps selon leur capacité à se lier à un seul ou plusieurs épitopes. On parle alors respectivement des anticorps polyclonaux et monoclonaux :

- **Anticorps polyclonaux** : correspondent à un mélange d'anticorps reconnaissant différents épitopes sur le même antigène. Chaque type d'anticorps est fabriqué par un clone cellulaire (plasmocyte). Les anticorps polyclonaux sont obtenus en injectant un antigène à un animal. Au cours de la réponse immunitaire, l'organisme de l'animal produit des anticorps dirigés contre différents épitopes de l'antigène : on parle de réponse polyclonale. Ces anticorps sont recueillés à partir du sang de l'animal immunisé.
- **Anticorps monoclonaux** : ce sont des anticorps ne reconnaissant qu'un seul type d'épitope sur un antigène donné. Ils sont tous identiques et produits par un seul clone de plasmocyte. L'immunisation de l'animal est identique à celle utilisée pour la production des anticorps polyclonaux. Cependant, au lieu d'isoler les anticorps du sang de l'animal, on récupère des lymphocytes B de la rate ou des ganglions sécrétant des anticorps dirigés spécifiquement contre l'antigène injecté. Les lymphocytes B ne se multiplient pas naturellement. Elles sont cultivées sur un milieu approprié en les hybridant à des cellules myélomateuses, c'est-à-dire des cellules tumorales immortalisées. Ces hybrides lymphocytes B - cellules myélomateuses vont alors se multiplier indéfiniment, constituant ainsi un clone cellulaire qui produira des anticorps monoclonaux. On peut également produire des anticorps monoclonaux *in vitro* pour éviter l'utilisation d'animaux. On immunise alors des cellules en culture (issues de la rate) en suivant la technique citée ci-dessus [Mishell et Shiigi 1980][Ling 2000].

A.7.3 Le mécanisme de révélation :

Le système de révélation permet de rendre visible les complexes Ag-Ac ainsi formés. Ce système est constitué de deux éléments : un anticorps et son traceur, les deux éléments constituant le marqueur. Le traceur est fixé généralement sur la partie constante de l'anticorps (pied du Y), trois types de traceurs sont utilisés en immunohistochimie : les fluorophores, les enzymes et l'or colloïdal. Nous nous intéresserons uniquement aux enzymes car ce sont les seuls traceurs pouvant être exploités en microscopie photonique à transmission. Et nous ne traitons dans le cadre de ce travail que les images des immunoréactions numérisées par ce type de microscopie. Les enzymes procurent une longue conservation stable du du marquage et sont peu onéreuses comparées au fluorophores ou à l'or colloïdal. Parmi ces enzymes, on peut citer la peroxydase de Raifort et la phosphatase alcaline. La peroxydase de Raifort est de loin l'enzyme la plus utilisée dans les laboratoires d'anatomie pathologique. Elle est préférée à la phosphatase alcaline pour le marquage des tissus animaux car ils contiennent une quantité non négligeable de la phosphatase endogène faussant le résultat final du marquage. La peroxydase est révélée en présence d'eau oxygénée et d'un chromogène incolore (substrat de la peroxydase) qui, par oxydation, délivre un signal coloré. Il existe deux principaux chromogènes pour la peroxydase de Raifort : le 3-amino-9-éthylcarbazole (AEC) et la 3,3'-dialinobenzidine (DAB). L'AEC donne une couleur rouge et la DAB émet une couleur marron. La DAB est le chromogène le plus utilisé. Elle autorise un large spectre de contre-colorants et l'utilisation d'un milieu de montage permanent. La DAB s'oxydant facilement à l'air libre, Il faut la conserver à l'abri de l'air dans des flacons bouchés hermétiquement.

A.7.4 Les différentes méthodes d'immunomarquage :

Il existe deux types d'immunomarquage :

1- **Immunomarquage direct** : le traceur est fixé directement à l'anticorps primaire spécifique à l'antigène. C'est une immunoréaction qui se déroule en une seule étape. C'est donc la réaction la plus simple et la plus rapide. La distance est très étroite entre le traceur et l'antigène. Cette réaction donne une très bonne résolution spatiale du marquage. Néanmoins on note que peu d'anticorps directement marqués sont disponibles dans le commerce (figure A.2).

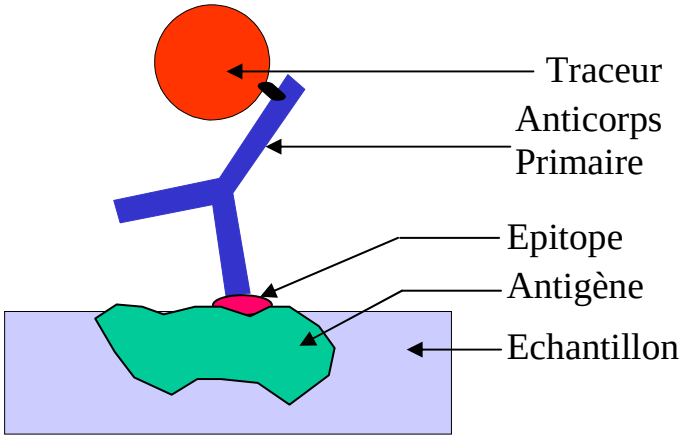


Figure A.2 : Immunomarquage direct.

2- Immunomarquage indirect :

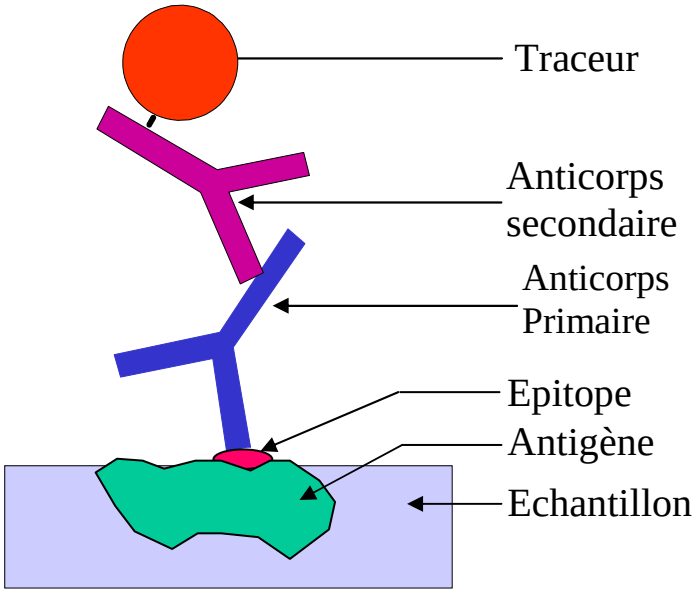


Figure A.3 : Immunomarquage indirect.

C'est la méthode la plus utilisée avec un grand choix de marqueurs (anticorps et traceur) disponibles chez de nombreux fournisseurs. La technique la plus simple consiste à révéler l'anticorps primaire par un anticorps secondaire qui porte une enzyme, par exemple la peroxydase de Raifort (figure A.3). Cette méthode est simple, peu onéreuse, et plus sensible comparée à l'immunoréaction directe, car plusieurs anticorps secondaires peuvent venir se fixer à l'anticorps primaire. Il existe d'autres types d'immunoréaction indirecte mais qui nécessitent d'autres d'anticorps intermédiaires ou d'autres systèmes de révélation.

3- Le système de révélation *biotine-avidine* :

Au cours de ma thèse, nous avons utilisé un système indirect à base d'avidine et de biotine pour la révélation de l'expression des marqueurs tumoraux. Cette technique se base sur la fixation de la biotine sur l'anticorps secondaire. L'avidine conjuguée au traceur possède une grande affinité à la biotine. C'est ce principe de liaison avidine-biotine qui est exploité pour la révélation des marqueurs tumoraux (figure A.4).

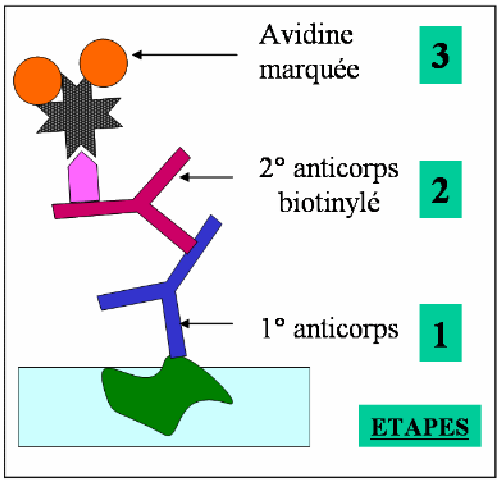


Figure A.4 : Système de révélation biotine-avidine.

A.8 Les anticorps utilisés au cours de la thèse :

Les marqueurs du cycle cellulaire suivants : ß-caténine, Bcl2, Cycline D1 et Ki67 ont été révélés par les techniques classiques d' immunohistochimie, présentées au niveau de cette annexe. Le tableau A.1 présente les différents anticorps utilisés pour la révélation des marqueurs tumoraux, du cancer colorectal, étudiés au cours de mon travail de thèse.

Tableau A.1 : Anticorps des marqueurs tumoraux du cancer colorectal. Dako[1] : Dakocytomation, Inc. USA. Novocastra[2] : Novocastra Laboratories Ltd. UK.

Anticorps	Laboratoire	Clone	Type	Dilution	Dénaturation par la chaleur
Ki67	Dako[1]	MIB-1	Monoclonal souris	1/50	Oui
ß-caténine	Novocastra[2]	17C2	Monoclonal souris	1/100	Oui
Cycline D1	Dako	DCS-6	Monoclonal souris	1/30	Oui
Bcl2	Dako	124	Monoclonal souris	1/50	Oui

Pour réaliser le marquage IHC, nous avons utilisé et suivi leprotocole du kit Streptavidine Biotine LSAB2 (fourni par la société Dako), comprenant les anticorps secondaires et le système de révélation..

A.9 Bibliographie :

P. Brandtzaeg et T.O. Rognum. Evaluation of nine different fixatives. 1. Preservation of immunoglobulin isotypes, J chain, and secretory component in human tissues. *Pathol Res Pract*, 179, 250-266, 1984.

A.H. Coons, H.J. Creech et R.N. Jones. Immunological properties of an antibody containing a fluorescent group. *Proc Soc Exp Biol*, vol.47:200-202, 1941.

R.C. Curran et J. Gregory .Demonstration of immunoglobulin in cryostat and paraffin sections of human tonsil by immunofluorescence and immunoperoxidase techniques. Effects of processing on immunohistochemical performance of tissues and on the use of proteolytic enzymes to unmask antigens in sections. *J Clin Pathol*, vol. 31:974-983, 1978.

J.C. Finley, G.H. Grossman, P. Dimeo et P. Petrusz .Somatostatin-containing neurons in the rat brain: widespread distribution revealed by immunocytochemistry after pretreatment with pronase. *Am J Anat*, vol. 153: 483-488, 1978.

J.C. Garaud et G. Roussel. Livre IHC: Immunohistochimie en microscopie photonique et électronique.Théorie et pratique. Version 2007.1.

T. Kawamoto et M. Shimizu.A method for preparing whole-body sections suitable for autoradiographic, histological and histochemical studies. *Stain Technol*, vol.61: 169-183, 1986.

T. Kawamoto. Light microscopic autoradiography for study of early changes in the distribution of water-soluble materials. *J Histochem Cytochem*, vol. 38: 1805-1814, 1990.

N.R. Ling. Immunoglobulin production by cultured human lymphocytes.J.Immunol. Methods, vol. 238: 3-15, 2000.

B.B. Mishell et S.M. Shiigi. Selected methods in immunology.Ed. *WHH Freeman and company, San Francisco*, 1980.

A.G. Pearse, J.M. Polak, C. Adams et P.A. Kendall. Diethylpyrocarbonate, a vapour-phase fixative for immunofluorescence studies on polypeptide hormones. *Histochem J*, vol. 6: 347-352, 1974.

J.M. Polak, P.A. Kendall, C.M. Heath et A.G.E Pearse. Carbodiimide fixation for electron microscopy and immunoelectron cytochemistry. *Experientia*, vol. 28, 368, 1972.

M. Reading .A digestion technique for the reduction of background staining in the immunoperoxidase method. *J Clin Pathol*, vol. 30: 88-90, 1977.

J.M. Robinson et D.D. Vandre. Antigen retrieval in cells and tissues: enhancement with sodium dodecyl sulfate.Histochem *Cell Biol*, vol. 116: 119-30, 2001.

S.R. Shi, M.E. Key et K.L. Kalra. Antigen retrieval in formalin-fixed, paraffin-embedded tissues: an enhancement method for immunohistochemical staining based on microwave oven heating of tissue sections. *J Histochem Cytochem*, vol. 39: 741-748, 1991.

Annexe B TMA virtuel : Outil complémentaire à la technologie TMA

B.1 Le concept de TMA virtuel :

L'élaboration d'un TMA est longue et fastidieuse ; elle doit donc être soigneusement motivée (expression d'un besoin), guidée (spécification d'un protocole) et validée. Plusieurs étapes interviennent dont : 1) la formulation d'un objectif (ex : examiner la relation entre grade et facteur de prolifération), 2) la spécification d'un plan d'expérience, 3) la réalisation du TMA et 4) l'exploitation du TMA.

Dans le cadre des travaux de thèse, nous avons introduit le concept de « TMA virtuel » pour répondre aux différentes problématiques posées par la technologie TMA. Un TMA virtuel peut regrouper plusieurs carottes virtuelles ou réelles pour répondre à une question scientifique donnée. Nous avons proposé l'utilisation du TMA virtuel comme outil complémentaire de la technologie TMA. Au cours de cette thèse, nous avons illustré l'utilisation de ce concept :

- comme moyen d'évaluation de la technologie TMA pour l'étude du cancer colorectal,
- comme support pour la validation de nouvelles approches de traitement d'images dédiées à la technologie TMA,
- et comme outil d'exploitation de données générées à partir des coupes TMA.

Nous présentons au niveau de cette annexe un outil dédié à la construction des TMA virtuels. Un outil qui permet le prélèvement de crottes virtuelles à partir de lames histologiques classiques avec différents marqueurs tumoraux ou à partir de lames TMA réelles. Ainsi, les carottes virtuelles peuvent être regroupées pour mener des analyses automatiques en réponse à une problématique scientifique ou médicale donnée.

B.2 L'outil de construction des TMA virtuels :

La conception d'un TMA virtuel suit trois étapes essentielles numérotées de 1 à 3 dans la figure B.1 :



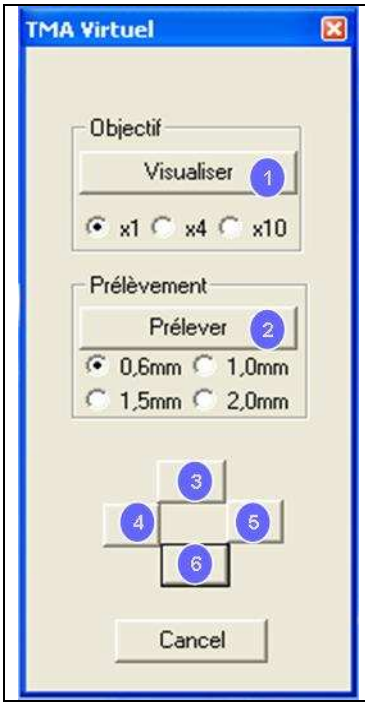
Figure B.1 : Outil TMA virtuel. Les trois boutons permettant la conception et l'analyse des TMA virtuels.

B.2.1 Prélèvement virtuel des carottes :

Pour construire un TMA virtuel, il faudra disposer d'un échantillon numérique complet des biopsies (blocs donneurs). Ceci requiert la création d'une mosaïque d'images à différents niveaux de résolution (objectifs x4, x10) reconstituant la totalité de l'échantillon. Lorsque les lames histologiques sont numérisées par le système d'acquisition d'image Samba2050^{TM}, elles sont découpées en morceaux d'images appelées vignettes. Ces vignettes ont des coordonnées contenues dans leur nom de fichier en hexadécimal. Lorsque le logiciel charge une image réduite de la lame histologique, il récupère l'arborescence des répertoires de la base des données images qui contiennent les vignettes constituants la lame numérique.

Le logiciel selon le grossissement courant (x1, x4 ou x10) utilise les vignettes numériques pour l'affichage de la vue courante, et permet grâce à quatre boutons de navigation de parcourir la biopsie de façon transparente et continue pour l'utilisateur (figure B.2).

Le prélèvement d'une carotte se fait par simple clic sur la partie voulue de la lame histologique. Les carottes ainsi prélevées sont stockées dans des répertoires prédéfinis dans la base de données images. Les carottes virtuelles peuvent être également prélevées à partir des coupes de blocs TMA.



Le bouton **1** permet de changer l'objectif de visualisation de la lame numérique et zoomer sur les parties de l'image qu'on veut visualiser.

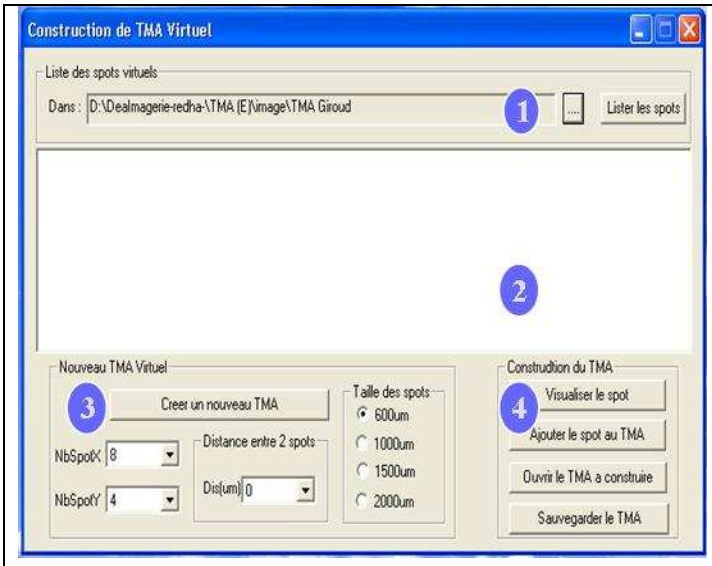
Le bouton **2** permet le prélèvement virtuel des carottes, l'utilisateur peut choisir différents diamètres pour la carotte à prélever en cochant des boutons radio.

Les boutons **3**, **4**, **5** et **6** permettent le déplacement dans les quatre directions au sein de la lame numérique.
3 : vers le haut.
4 : vers la gauche.
5 : vers la droite.
6 : vers le bas.

Figure B.2 : Boîte de dialogue permettant le prélèvement virtuel des carottes.

B.2.2 Construction des TMA virtuels :

Une boîte de dialogue a été conçue pour assurer cette étape (figure B.3). Le logiciel propose différentes options pour la conception du TMA virtuel à construire, l'utilisateur peut choisir le nombre de carottes sur la largeur et la hauteur ainsi que l'espace inter-carottes de la trame du TMA. L'utilisateur peut aussi sélectionner le diamètre des carottes à inclure dans le nouveau TMA virtuel. Le TMA ainsi créé est vierge, il ne contient aucune carotte pour le moment.



La partie **1** permet la sélection des lames pour lesquelles on veut lister la totalité des carottes virtuelles prélevées.

La zone de texte **2** affiche les carottes virtuelles déjà prélevées.

La zone **3** permet la sélection les options de création d'un nouveau TMA virtuel.

La zone **4** permet l'ouverture et la sauvegarde des TMA. Elle permet également la visualisation des carottes affichées dans la zone **2** et l'ajout des carottes sélectionnées.

Figure B.3 : Boîte de dialogue qui assure la construction des TMA virtuels.

La boîte de dialogue permet de consulter la totalité des carottes virtuelles répertoriées dans la base de données images ou seulement les carottes prélevées sur une seule lame ou sur un groupe de lames spécifiques (figure B.3). L'utilisateur peut également sélectionner les carottes de la base de données images et les visualiser pour guider son choix sur des carottes pertinentes pour son plan expérimental afin de les inclure dans le TMA en construction. En fin de construction du TMA, un fichier image représentant la lame TMA virtuelle est stocké dans la base de données images. Le TMA virtuel pourra alors être soumis à la procédure d'analyse d'images.

B.2.3 Analyse quantitative des carottes du TMA virtuel :

Cette partie a été développée dans le but de donner à l'expérimentateur un outil pour la validation du TMA virtuel précédemment construit.

Une boîte de dialogue (figure B.4) permet la visualisation de la totalité des carottes de l'image du TMA virtuel. Un simple clic sur la surface d'une carotte de l'image réduite du TMA permet l'affichage de la carotte avec un plus fort grossissement (x10).



1 Ce bouton permet le choix d'une carotte pour l'afficher en x10. **2** Option de seuillage de la carotte. **3** Zone d'affichage du résultat du marquage de la carotte. **4** Zone déroulante pour la sélection du marquage à détecter. **5** Retour à la lame de TMA totale.

Figure B.4 : Boîte de dialogue permettant l'analyse de TMA virtuel.

Sur tous types de carottes, une quantification du marquage peut êtr réalisée à l'aide d'une procédure de seuillage interactif. Les régions positives pour le marquage (ou la coloration) sont affichées selon un dégradé de couleurs qui va du bleu au rouge pour les régions les plus marquées (figure B.5).

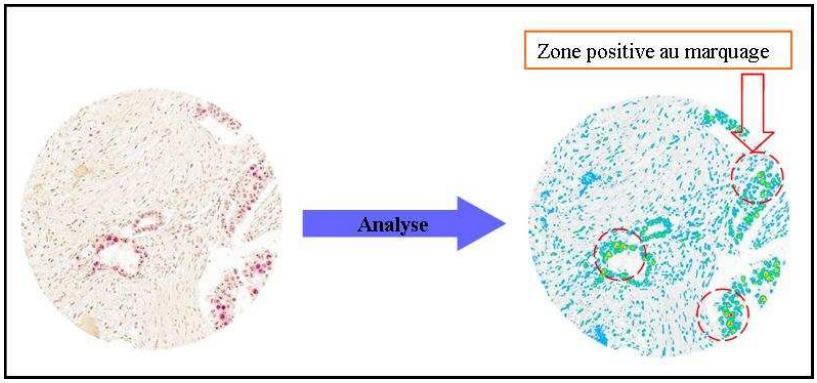


Figure B.5 : Analyse d'images d'une carotte de TMA virtuel.

B.3 La validation de l'outil de TMA virtuel par la création d'une lame TMA virtuelle :

B.3.1 Le concept des lames virtuelles :

Une lame virtuelle consiste à regrouper des carottes ayant un point commun afin de répondre à une question donnée. Les points communs sont formulés sous forme de requêtes par l'utilisateur. Les requêtes peuvent être de type « regrouper toutes les carottes d'un patient « donné » ou « les tumeurs d'un grade donné ». Les carottes d'une lame virtuelle peuvent être constituées d'une juxtaposition de prélèvements tumoraux issus d'un même bloc donneur mais de localisations différentes, ou de blocs donneurs différents, ou de toute composition au choix de l'expérimentateur selon la question biologique et/ou médicale posée. De plus, une lame virtuelle peut regrouper des carottes de TMA réels et des carottes virtuelles issues de lames histologiques numériques et répertoriées dans la base de données images (figure B.6).

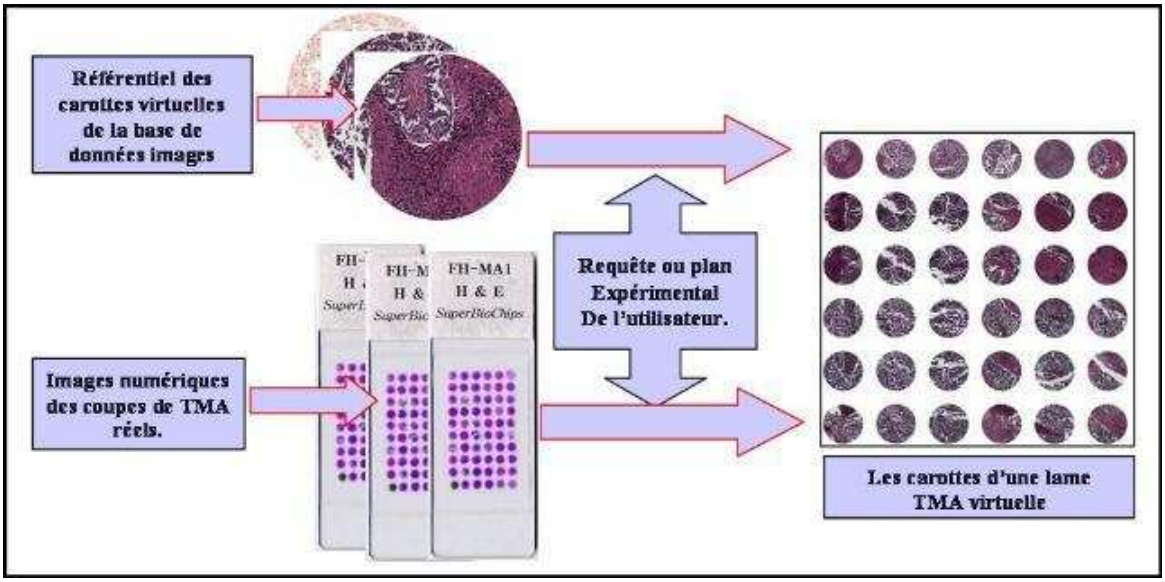


Figure B.6 : Processus de construction d'une lame virtuelle.

B.3.2 Exemple de construction d'une lame TMA virtuelle :

Pour illustrer et valider l'utilisation de notre outil de construction de TMA virtuel, nous avons généré une lame TMA virtuelle. Voici les étapes de sa construction :

B.3.2.1 Création d'un bloc TMA : plan d'expérience :

Huit blocs de biopsies donneurs, indicés de A à H sont utilisés dans notre expérience. Sur chacun de ces blocs, 3 à 4 zones d'intérêts sont sélectionnées (régions 1 à 4). Pour chacune de ces zones, 3 carottes sont prélevées et mises en place dans le bloc receveur selon le plan d'expérience présenté dans la figure B.7.

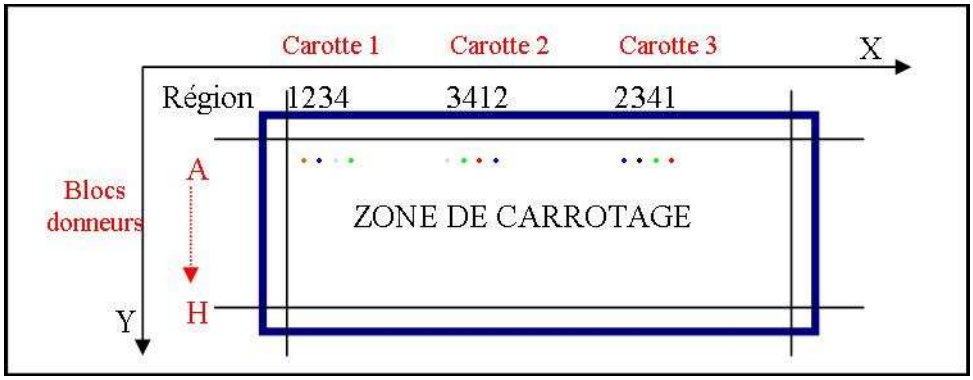


Figure B.7 : Plan d'expérience, représentation du bloc TMA. 8 blocs donneurs (A à H), 3 à 4 zones de carottage (carottes de couleurs différentes), 3 carottes par zone.

B.3.2.2 Acquisition de coupes TMA sériées (coloration HES) :

A partir de 20 coupes successives réalisées dans le bloc de TMA et colorées en HES, on a réalisé l'acquisition de 3 couples de lames sériées : lames 4-5, 14-15 et 22-23.

B.3.2.3 Création d'une lame TMA virtuelle :

En utilisant les lames TMA précédemment sélectionnées, une lame TMA virtuelle de 63 carottes est générée pour l'étude de l'homogénéité du tissu constituant les carottes dans la profondeur des carottes du bloc de TMA réel. La matrice (9x7 carottes) de la lame ainsi réalisée (figure B.8) présente sur l'axe des abscisses les carottes déposées par groupe de trois carottes issues de la même zone d'intérêt délimitée précedemment par l'anatomopathologiste. Sur l'axe des ordonnées, les carottes sont déposées selon leur ordre croissant du niveau de

coupe (0, 4, 5, 14, 15, 22, 23). La position 0 correspond à des carottes virtuelles prélevées sur les lames histologiques des blocs de biopsies donneurs.

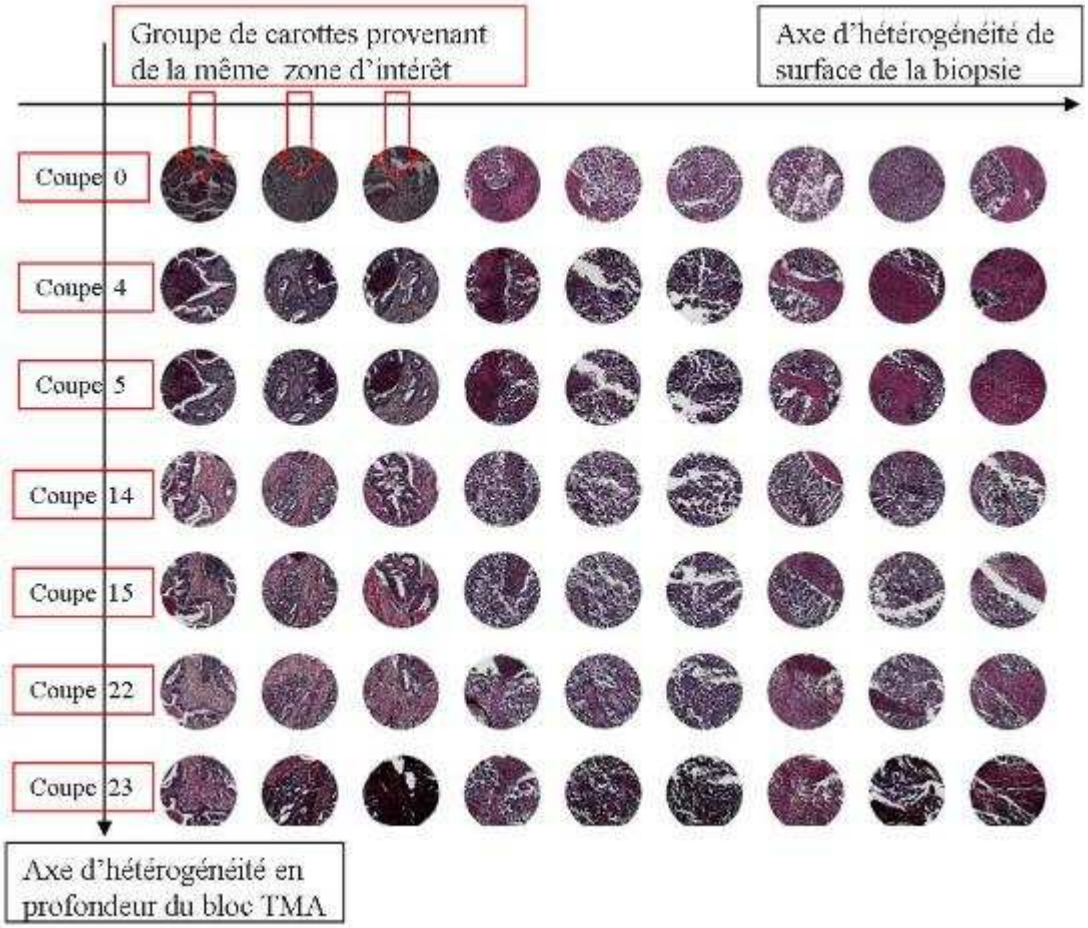


Figure B.8 : Lame virtuelle composée de carottes virtuelles (ligne de la coupe 0) et de carottes réelles correspondant aux coupes 4, 5, 14, 15, 22 et 23. Sur chaque ligne, les carottes sont alignées par bloc de 3 correspondant aux 3 carottes prélevées dans chacune des 3 zones.

La figure B.8 illustre bien la variation de la nature tissulaire des carottes en profondeur. Cet exemple de lame TMA virtuelle démontre l'efficacité de notre outil de construction de TMA virtuel pour répondre à une question d'ordre histologique bien précise. Ce principe peut être adapté pour mener d'autres types de questions biologiques ou médicales.

Annexe C Analyse statistique des méthodes de classifications couleur

L'objectif de cette annexe est de présenter la méthodologie statistique requise pour le choix des tests statistiques à appliquer et pour la comparaison des différentes méthodes de classification couleur présentées dans le septième chapitre (§7.4.6.1).

Le but est de comparer les différentes classifications : manuelles et automatiques à la classification de référence. Les analyses statistiques réalisées sont basées sur des comparaisons effectuées par paires < LI classification x, LI classification référence> pour chaque image histologique des 88 images que comporte l'étude. En conséquence, les analyses statistiques concernent uniquement des **échantillons appariés**.

L'idée de base est d'utiliser les tests classiques sur les échantillons appariés, tels que les tests de Fischer, t-Student ou d'ANOVA II. Ces tests sont des tests paramétriques. Il faut veiller à vérifier la nature de la distribution des échantillons, car ces tests ne peuvent s'appliquer que sur des échantillons présentant des distributions normales.

C.1 Tests de normalité liés à la distribution des indices de marquage :

En vue de comparer deux échantillons de même taille, plusieurs études mettent en application directement le test de t Student, adapté en réalité qu'aux échantillons ayant une distribution normale [Altman et Bland 1996]. L'utilisation d'un test statistique inadéquat peut conduire à des résultats erronés. Il faut *à priori* toujours contrôler la distribution des échantillons dont on dispose, pour choisir, en conséquence, les tests statistiques les plus adaptés à ces échantillons. La figure C.1 illustre la distribution de la classification de référence, ainsi que deux exemples de classification automatique et manuelle. On observe déjà que les indices de marquage, associés aux trois types de classification, ne suivent pas une loi gaussienne (normale), représentée par la courbe rouge sur la figure C.1. On peut aussi constater que les histogrammes associés aux classifications ne présentent pas d'axe de symétrie (sont dissymétriques).

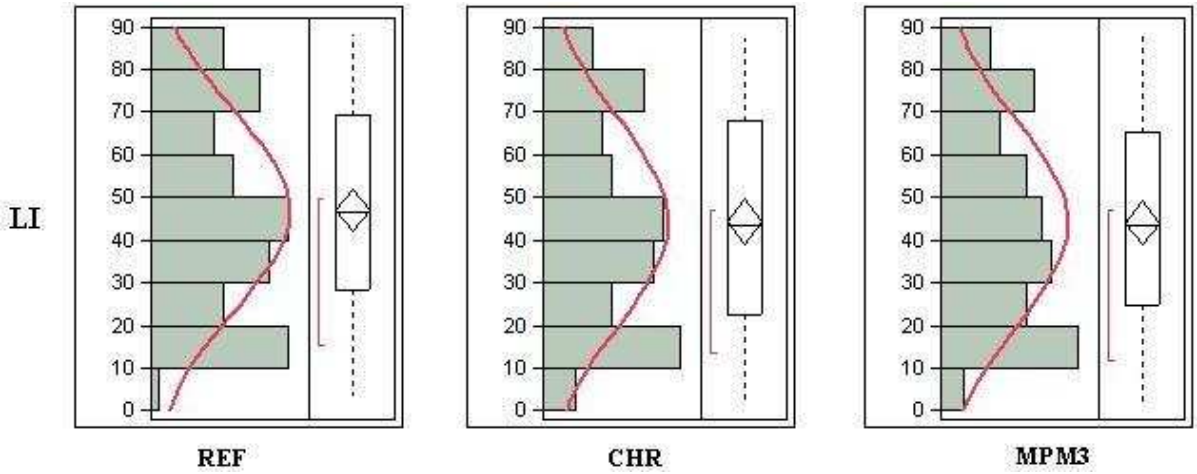


Figure C.1 : Histogrammes de distribution des indices de marquage liés à quelques classifications.

Afin d'analyser de manière exacte et chiffrée les distributions des séries de données associées à l'ensemble des classifications, nous avons choisi le test de Shapiro-Wilk. Ce test est un moyen puissant pour étudier les caractéristiques d'une distribution donnée, en particulier dans le cas des échantillons ayant des distributions asymétriques.

Le test de Shapiro-Wilk a été appliqué avec un intervalle de confiance de 95%. L'hypothèse nulle Ho sous-jacente à cet intervalle de confiance est la suivante : **la distribution de l'échantillon suit une loi normale avec un risque d'erreur de 5% :** les valeurs p supérieures au seuil 5% (p>0.05) rejettent Ho, l'hypothèse H1 alternative est alors acceptée.

Tableau C.1 : W valeurs et p valeurs du test Shapiro-Wilk associées aux résultats des classifications.

	MPM1	MPM2	MPM3	MH1	MH2	MH3	FG1	FG2	FG3	REF	CHR	CD	BN	Hue
W	0,9622	0,9557	0,957	0,9624	0,957	0,9595	0,9623	0,9577	0,9626	0,9564	0,9525	0,9348	0,9821	0,9478
Pval.	0,0115	0,0044	0,0053	0,0119	0,0053	0,0077	0,0117	0,0059	0,0123	0,0048	0,0028	0,0003	0,2671	0,0014

Excepté la classification automatique BN qui (satisfait Ho : p>0,05), les échantillons d'indices de marquage associés aux classifications rejettent l'hypothèse Ho (p<0,05), donc ils ne suivent pas une distribution normale (tableau C.1). Ces résultats imposent le choix d'un test non paramétrique pour comparer ces échantillons, c'est-à-dire un test ne posant aucune condition préalable sur la distribution de ces échantillons. Dans la section suivante, nous aborderons le test statistique adopté pour comparer les résultats générés par les différentes méthodes de classification.

C.2 Comparaison des indices de marquages associés aux classifications :

L'objectif est de tester si les différentes classifications manuelles et automatiques fournissent des indices de marquage suivant les mêmes distributions que celles rattachées à la classification de référence ; c'est-à-dire tester si ces méthodes de classification sont conformes statistiquement à la référence.

Notre choix s'est porté sur le test de Wilcoxon qui est un test non paramétrique d'identité adapté aux échantillons appariés :

- non paramétrique : puisqu'il ne pose aucune hypothèse sur la distribution des échantillons à comparer,
- d'identité : il teste en effet si deux échantillons sont issus de la même distribution.

Le test de Wilcoxon est adapté pour comparer deux échantillons. Pour appliquer ce test à plusieurs échantillons à la fois, nous avons eu recours au test de Kruskal-Wallis, appelé également le test de Wilcoxon/Kruskal-Wallis. En effet, Le test de Kruskal-Wallis est une généralisation du test de Wilcoxon simple à plus de deux échantillons. Il est considéré comme l'équivalent du test ANOVA univarié classique adapté aux distributions normales.

Le test de Wilcoxon/Kruskal-Wallis est appliqué sur nos séries de classifications avec un intervalle de confiance de 95%. Ce test est associé à l'hypothèse Ho selon laquelle les échantillons analysés proviennent de la même distribution. L'hypothèse alternative H1, rejetant Ho, signifie que les échantillons ne suivent pas la même distribution : elle prouve avec un risque de 5% que ces échantillons proviennent de distributions différentes.

- Si $p > 0,05$ alors on accepte l'hypothèse Ho au risque de commettre une erreur d'estimation de 5%. Les échantillons analysés (2 ou plusieurs) présentent des distributions identiques d'indices de marquage LI.
- Si $p < 0,05$ alors on rejette l'hypothèse Ho. Les échantillons analysés (2 ou plusieurs) présentent des distributions d'indices de marquage LI non identiques.

C.3 Bibliographie :

D.G. Altman et J.M Bland. Statistics Notes: Comparing several groups using analysis of variance. *BJM*, vol. 312 :1472-1473, 8 June 1996.

Résumé : La technique récente des puces à tissus « Tissue Micro Arrays » TMA apparaît comme un moyen indispensable d'investigation pour la validation des profils d'expression des marqueurs tumoraux en relation avec la dynamique de l'architecture tissulaire lors de la transformation tumorale. Brièvement, cette technologie consiste à regrouper, dans un seul bloc de paraffine, plusieurs centaines de petits échantillons tissulaires sous forme de carottes cylindriques prélevées à partir de différents blocs de biopsies classiques. Le sujet de thèse s'intéresse aux différents aspects de traitement d'images et de contrôle qualité liés à la technologie TMA. Les travaux de thèse sont réalisés selon deux grands axes qui retracent l'enchaînement chronologique des opérations : conception des blocs TMA, puis analyse automatique des coupes TMA. Dans un premier temps, le concept de TMA virtuel est introduit pour simuler les protocoles de prélèvement de carottes afin d'évaluer la technologie TMA. En deuxième temps, une modélisation originale de la couleur associée au diagramme de chromaticité xy, est proposée pour la détection automatique des marqueurs tumoraux au niveau des coupes TMA. Le concept de sociologie cellulaire, modélisé par le diagramme de Voronoï, est finalement adopté pour illustrer la dynamique de la distribution spatiale des cellules cancéreuses au cours de la transformation tumorale. Les méthodes développées sont appliquées à l'étude des marqueurs tumoraux du cancer colorectal.
Mots clés : Cancer colorectal, puce à tissus « TMA », TMA virtuel, modèles couleur, sociologie cellulaire, diagramme de Voronoï.

Abstract: The recent technique of Tissue Micro Arrays "TMA" seams an essential way to investigate about the precise knowledge of tumour markers expression and their relation with the tissue architecture dynamic in the tumour development process. Briefly, this technology allows gathering, in a single block of paraffin, several hundreds of small cylindrical-shaped tissue-samples from different blocks of conventional tumour biopsies. This thesis focuses on different aspects of image processing and quality control related to TMA technology. The work is realized according to two main sections in agreement with chronological sequence of operations: design of TMA and automated analysis of TMA sections. As a first step, the concept of virtual TMA is introduced to simulate the protocols of core sampling to evaluate the TMA technology. In the second step, an original colour model associated with the chromaticity diagram xy is proposed for automatic detection of tumour markers in numerical TMA slides. The concept of sociology cell, modelled by Voronoï diagram, is finally adopted to illustrate the dynamics of the spatial distribution of cancer cells in the process of tumour development. The different methods developed in this work are applied to study the tumour markers of colorectal cancer.
Keywords: Colorectal cancer, Tissue Micro Arrays "TMA", virtual TMA, colour models, sociology cell, Voronoï diagram.