University of Iowa

Iowa Research Online

Theses and Dissertations

2008

Lung cancer: an evaluation of volumetric histopathological architecture with correlation to computed tomography

Jessica Corinne De Ryk
University of Iowa

Recommended Citation

De Ryk, Jessica Corinne. "Lung cancer: an evaluation of volumetric histopathological architecture with correlation to computed tomography." PhD (Doctor of Philosophy) thesis, University of Iowa, 2008.
https://doi.org/10.17077/etd.ng77sz9f

LUNG CANCER: AN EVALUATION OF VOLUMETRIC HISTOPATHOLOGICAL

ARCHITECTURE WITH CORRELATION TO COMPUTED TOMOGRAPHY

by

Jessica Corinne de Ryk

An Abstract

Of a thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Biomedical Engineering
in the Graduate College
of The University of Iowa

December 2008

Thesis Supervisor: Professor Geoffrey McLennan

ABSTRACT

Over 190,000 Americans die every year from lung cancer, making it the number one cause of death from cancer in the United States of America. Lung cancer has maintained the same low five year survival rate, of 13-15%, over the last thirty years. There is therefore desperate need for improvement in diagnostic and therapeutic techniques for lung cancer. Multidetector computed tomography (MDCT) is being increasingly used for lung cancer detection and characterization. While national lung cancer screening trials have shown MDCT to be effective in detecting even very small lung nodules, the characterization achievable through this modality is poor. The majority of non-small cell lung cancer nodules are histologically heterogeneous and consist of malignant tumor cells, necrosis, stromal tissue, and inflammation; however, the extent of this heterogeneity is unknown. Geometric and tissue density heterogeneity are underutilized in MDCT representations of lung tumors for distinguishing between malignant and benign nodules because there has been no thorough investigation into the correlation between radiographic heterogeneity and corresponding histological content in 3D. To understand and to make more effective this lung cancer characterization by MDCT, two vital steps must be taken. Firstly, an understanding of the 3D structure and content of tissue types that constitute a lung nodule must be established. Secondly, this knowledge must then be used to assess how nodule tissue content corresponds to the heterogeneity apparent in MDCT data, impacting diagnosis, planning biopsy procedures and nodule change analysis.

In this study we have developed a process model for establishing a direct correlation between histopathology and non-destructive radiological imaging. We

provide the 3D structural and pathological detail of lung cancer nodules and surrounding tissues using a purpose built Large Image Microscope Array (LIMA). This information served as the basis for registration of MDCT images of the human nodule before and after resection, computed micro-tomography (micro-CT) detail and histopathology.

Abstract Approved: _____

Thesis Supervisor

_____

Title and Department

_____

Date

LUNG CANCER: AN EVALUATION OF VOLUMETRIC HISTOPATHOLOGICAL

ARCHITECTURE WITH CORRELATION TO COMPUTED TOMOGRAPHY

by

Jessica Corinne de Ryk

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Biomedical Engineering
in the Graduate College
of The University of Iowa

December 2008

Thesis Supervisor: Professor Geoffrey McLennan

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL
_____

PH.D. THESIS
_____

This is to certify that the Ph.D. thesis of

Jessica Corinne de Ryk

has been approved by the Examining Committee for the thesis requirement
for the Doctor of Philosophy degree in Biomedical Engineering at the
December 2008 graduation.

Thesis Committee: _____
                       Geoffrey McLennan, Thesis Supervisor

                       _____
                       Joseph M. Reinhardt

                       _____
                       Eric A. Hoffman

                       _____
                       Milan Sonka

                       _____
                       Barry De Young

                       _____
                       Michael B. Cohen

To my parents, for inspiring me to shoot for the stars and your unwavering
faith that I can reach them.

Nothing in life is to be feared, it is only to be understood. Now is the time
to understand more, so that we may fear less.

Marie Curie

I was never travelling alone. A special thank you to Melissa Suter, who was my first friend in America and has proven to be lucky find ever since. To Jacqueline Thiesse Namati, I do not know how I will survive my next job without our ritual 'lap around the hospital' – I will miss you.

To my parents, Lya and Frans de Ryk and my sister Jade, words just can not describe how thankful I am to have been blessed with you as my family.

Finally, thank you to my husband, Jered Sieren, for being my rock, for knowing just what to say, for making me laugh even when I felt like crying and for being the exceptional man you are.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

x

xi

CHAPTER 1

INTRODUCTION

The mortality from lung cancer is higher in both men and women than from any other form of cancer, despite the fact that the most commonly diagnosed cancers are prostate cancer in men and breast cancer in women. Multidetector computed tomography (MDCT) is widely used for the detection and evaluation of lung nodules and there are major efforts underway to assess this modality in screening and clinical approaches, including the National Lung Screening Trial, the Lung Image Database Consortium, and the Early Lung Cancer Action Program. However, most detected lung nodules are not malignant. The lung cancer paradox remains unsolved – early detection of suspect nodules is achievable, but early therapy is not. Currently, the recommended standard of care is radiological surveillance for evidence of 'growth'. Some computer aided detection and diagnosis algorithms based on MDCT data have been developed and achieved promising results. However, a considerable barrier to assisting with diagnosis is that the 'ground truth' of lung nodule structure is not known, significantly inhibiting the development of more sophisticated algorithms with further informative outputs.

Non-small cell lung cancer nodules are considered by most clinicians to be histologically homogeneous (consisting only of cancer). However, heterogeneity exists, a factor not appreciated by many investigators and one which is also poorly studied. The nodules typically consist of malignant tumor cells, necrotic tumor, fibroblastic stromal tissue, and inflammation with possibly only a small proportion of the nodule containing viable cancer cells. However, the precise registration between radiographic heterogeneity and corresponding histological content has been non-existent and is not available for the

two-dimensional (2D) or three-dimensional (3D) image space. It is vital, for the continued maturation of lung cancer diagnosis technologies and treatment outcome measures, that a 3D understanding of the structural content of lung nodules is obtained. This information is not just self evident for MDCT approaches but also for other imaging techniques such as PET, MRI, and optical scanning.

The aim of this investigation is to establish a process for examining the 3D structure of a human lung cancer nodule within a multimodal environment, by linking the ground truth from 3D histopathology to *in vivo* MDCT. Through the creation of a dataset which contains registered radiological density, structural, color and cellular information, correlations can be made between the representations of the tissue types in the different modalities.

The *first hypothesis* to be examined in this study is that lung nodules contain both cancerous and non-cancerous regions with a three dimensional complex architecture which is not evident through traditional histological processing. The *second hypothesis* is that the histopathologically determined tissue types within a lung nodule is related to the nodule's representation in computed tomography imaging, with a correlation between subtle density variations and cancer versus non-cancerous tissue types.

The specific aims of this research proposal that have been developed to address the hypotheses presented above are:

Specific Aim 1: *Establish a reliable tissue acquisition, processing and multimodal imaging method.* This aim involves the establishment of an interdisciplinary study team, gaining approval for the study from the University of Iowa Institutional Review Board (IRB) and identifying and consenting patients for study enrollment.

Tissue processing methods also need to be developed for tissue fixation and specimen stabilization between imaging modalities. The resulting datasets are to include *in vivo* MDCT, MDCT scans of the resected lobe both fresh and fixed, and the resected nodule imaged via MDCT, computed micro-tomography (micro-CT), a novel large-scale image microscope array (LIMA) system with associated histology. Through these datasets density, color and cellular information relating to the cancer nodule will be gained.

Specific Aim 2: *The development of an automated classification approach for histopathological data to improve speed and repeatability of nodule tissue type identification*. Manual tracing of histopathology data in order to generate a ground truth mapping of nodule tissue types is time consuming and subjective. Hence, an approach for the automated classification of the histological data will be established.

Specific Aim 3: *The construction of a registration processing pipeline to register the multimodal image datasets to a common coordinate system.* The novel LIMA system has been developed to serve as the ground truth reference for the registration of the datasets, establishing structural correspondence between the micro-CT and histology datasets. Rigid and elastic registration techniques will be employed to meet the registration goals.

Specific Aim 4: *The analysis of nodule histopathological tissue type properties including the correlation to MDCT graylevel heterogeneity.* The histopathological tissue type maps will be used to statistically evaluate if the MDCT graylevel heterogeneity is related to the cellular composition in lung nodules.

CHAPTER 2

BACKGROUND

2.1   The Respiratory System

The lungs are the major organ of the respiratory system within which gas transfer occurs. Oxygen is delivered to the vascular system such that it can be transported to the cells and carbon dioxide, a by-product of cell metabolism, is removed. At total lung capacity the average human lung volume is 4-5 liters, of this volume approximately 10 percent is tissue, 10 percent is blood and 80 percent is air [1]. The lungs are divided into sections, or lobes. The right human lung has three lobes, right upper lobe (RUL), right middle lobe (RML) and right lower lobe (RLL). The smaller left lung has two lobes, left upper lobe (LUL) and left lower lobe (LLL). The airways consist of a conducting portion, which transports gases in and out of the lung, and the respiratory region in which gas exchange occurs. The conducting zone incorporates the trachea, the bronchi, the bronchioles and the terminal bronchioles. The respiratory zone contains the acini, which includes the respiratory bronchioles, the alveolar ducts and the alveoli, as depicted in **Figure 2-1**. There are approximately 480 million alveoli in an adult human lung, which are arranged in clusters to form alveolar sacs [2, 3].

The tissue content of the lung includes the walls of the circulatory vessels, conducting airways and the acini. The characteristics and cellular composition of these lung components varies widely. The large conducting airways are thick, multi-layer walls containing a mucosa, smooth muscle and cartilage. As the airways decrease in size, less cartilage is present. The terminal acini units consist of the respiratory bronchioles and alveoli, as shown in **Figure 2-2**. The respiratory bronchiole walls still contain

smooth muscle and cuboidal epithelium, but unlike the conducting bronchioles no cartilage is present. The alveolar walls are composed of an endothelium lining of the capillaries, an epithelium lining the airspaces and an interstitial layer containing connective tissue fibers. The epithelium of the alveoli is made up of type I pneumocytes and type II pneumocytes (secretory cells). The type I pneumocytes provide a large surface area for gas exchange while the alveolar type II pneumocytes are responsible for the synthesis, storage and secretion of phosopholipid rich surfactant.

2.2   Lung Cancer

Cancer is a disease in which cells acquire genetic alterations and divide without control. Cell division of normal lung tissue is necessary to retain the structure and functionality of the organ. Normal cells undergo controlled transitions between resting and dividing states. Exposure to cigarette smoke, excess radiation and other environmental carcinogens along with genetic factors can cause malignant transformation (carcinogensis) of normal cells. Malignant or cancerous cells grow and divide independent of the needs and limitations of the body, avoiding the resting state typical of normal cells. These cancerous cells have the ability to travel via the blood stream to other parts of the body where they continue to grow as metastases.

2.2.1   Radiological Classification

A solitary pulmonary nodule is defined as a discrete area of pulmonary opacity appearing on chest x-ray or MDCT [4]. Opaque lesions less than three centimeters are defined as nodules while a larger lesion is referred to as a mass [4, 5]. Solitary pulmonary nodules may be cancerous or non-cancerous (usually a granuloma).

With increasing biopsies being performed due to the early detection of lung cancer through MDCT screening, the MDCT appearances have been described with nodule classifications of solid, non-solid (also sometimes referred to as pure ground glass opacities – GGO) and part solid (sometimes referred to as GGO with a solid central component) being associated somewhat with histological subtypes [6-9].

2.2.2   Cellular Classification

There are two main histological groups of lung cancer; small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). Small cell lung cancer is an aggressive cancer that accounts for around 15% of all lung cancers. SCLC's are derived from a common neuroendocrine precursor cell within the airways and is often fast growing and metastasize easily. Approximately 85% of all lung cancers are NSCLC, of which there are three sub-types. These subtypes are grouped together under NSCLC due to their similarity in treatment and prognosis. NSCLC's are derived from a common precursor epithelial cell that may be poorly differentiated or differentiated into; squamous cell carcinoma, large cell carcinoma or adenocarcinoma,. Squamous cell carcinomas typically arise in the bronchial epithelium and are typically centrally located and comprise about 30% of all lung cancers. Large cell carcinomas account for about 10% of all lung cancers and show no evidence of squamous or glandular maturation and typically present in the mid to peripheral regions of the lung. Adenocarcinomas are the most common type of NSCLC comprising of approximately 40% of all lung cancers and are the primary focus for this lung nodule study. Adenocarcinomas arise from the glandular cells located in the epithelium lining of the bronchi and are typically peripherally located, often near the pleural surface. Adenocarcinomas were sub typed by Noguchi into

pathological sub-types A-F [10]. These sub-types, A to F, are associated with progressively poorer prognoses in terms of long term survival:

A. Localized bronchioalveolar carcinoma (LBAC)

B. LBAC with foci of collapsed alveolar structures

C. LBAC with foci of active fibroblastic proliferation

D. Poorly differentiated adenocarcinoma

E. Tubular adenocarcinoma

F. Papillary adenocarcinoma with compressive and destructive growth

The World Health Organization (WHO) updated the list of pathological sub-types of adenocarcinoma in 2004 [11] as:

1. Adenocarcinoma with mixed subtypes

2. Acinar adenocarcinoma

3. Papillary adenocarcinoma

4. Bronchioalveolar carcinoma (BAC) – nonmucinous, mucinous or intermediate

5. Solid adenocarcinoma with mucin production

The difficulty with attempts to categorize the pathological sub-types of adenocarcinoma is that most lung adenocarcinomas are histopathplogically heterogeneous and contain multiple sub-types. As a further complication adenocarcinoma nodules sometimes present containing squamous cell or even NSCLC components. In addition, categorization is based on pathologist's classification of nodule content from histopathology data, which examines only small sub samples of tissue.

2.2.3 Lung Cancer Staging

Lung cancer staging is a method by which the extent of disease is classified. This process is important in identifying appropriate treatment approaches and determining prognoses. All available factors, including clinical factors (physical exam, imaging and laboratory findings) and pathological finding (from tissue specimens obtained via bronchoscopy, mediastinoscopy or surgery) are used to determine stage.

The methods for staging differ based on cellular classification. For non-small cell lung cancers (NSCLC) the American Joint Committee on Cancer has designated staging by tumor, node and metastases (TNM) classification. This staging system takes into account the extent of the tumor (T), the level of regional lymph node involvement (N) and the presence of metastases (M) [12]:

- Primary tumor (T)

    o TX: Positive malignant cytology finding with no observable lesion

    o Tis: Carcinoma in situ

    o T1: Diameter of 3 cm or smaller, is surrounded by lung or visceral pleura, and is without invasion more proximal than the lobar bronchus

    o T2: Diameter greater than 3 cm and/or has extension to the visceral pleura, atelectasis, obstructive pneumonitis that extends to the hilar region but does not involve the whole lung or tumor of a main bronchus more than 2 cm distal from the carina

    o T3: A tumor of any size that directly invades any of the following: chest wall (including superior sulcus tumors), diaphragm, mediastinal pleura, parietal pericardium; or, associated atelectasis or obstructive pneumonitis of the entire

lung or, tumor in the main bronchus less than 2 cm distal to the carina but without involvement of the carina

- o T4: A tumor of any size that invades any of the following: mediastinum, heart, great vessels, trachea, esophagus, vertebral body, carina; or, separate tumor nodules in the same lobe; or, tumor with a malignant pleural effusion.

- Regional lymph nodes (N)

  - o NX: Regional lymph nodes cannot be assessed

  - o N0: No regional lymph node metastasis

  - o N1: Metastasis to ipsilateral peribronchial and/or ipsilateral hilar lymph nodes, and intrapulmonary nodes including involvement by direct extension of the primary tumor

  - o N2: Metastasis to ipsilateral mediastinal and/or subcarinal lymph node(s)

  - o N3: Metastasis to contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene, or supraclavicular lymph node(s)

- Metastasis (M)

  - o MX: Distant metastasis cannot be assessed

  - o M0: No distant metastasis

  - o M1: Distant metastasis present.

Using the TNM classification to describe the level of tumor invasion, a staging system [13] and these stages can be linked to estimated survival rates [14];

- Stage 0: TisN0M0

- Stage IA: T1N0M0, 5 year survival rate of 60-80%

- Stage IB: T2N0M0, 5 year survival rate of 50-60%

- Stage IIA: T1N1M0, 5 year survival rate of 40-50%

- Stage IIB: T2N1M0 or T3N0M0, 5 year survival rate of 25-40%

- Stage IIIA: T3N1M0 or T(1-3)N2M0, 5 year survival rate of 10-35%

- Stage IIIB: T4N(0-3)M0 or T(1-4)N3M0, 5 year survival rate of 5%

- Stage IV: T(1-4)N(0-3)M1, 5 year survival rate less than 5%

Due to the aggressive nature of small cell lung cancers (SCLC), the majority of diagnosed patients also have metastases and hence a simple two stage classification (limited versus extensive) is typically favored over the detailed TNM staging used for NSCLC.

Limited stage SCLC is assigned for small tumors which are confined to the chest (including mediastinum and supraclavicular node) with no pleural effusion [15]. Limited SCLC stage is associated with a 2 year survival rate of 20% [16].

Extensive stage SCLC is assigned for an occurrence of distant metastases and/or for any tumor too extensive to be incorporated into the limited stage. The prognosis for extensive stage SCLC is a 2 year survival rate of 5% [16].

2.3 Radiological Lung Imaging and Evaluation

2.3.1 *In Vivo* Lung Imaging Modalities

Both x-ray and nuclear imaging are commonly used to acquire *in vivo* lung image data. These modalities provide fast and relatively non-invasive methods for the detection of lung nodules.

Positron emission tomography (PET) is a form of nuclear imaging which detects biochemical changes in body tissue. This non-invasive imaging method involves the

administering of radionuclides, which are taken up by healthy tissue at different rates to cancerous tissue. The radionuclides emit positrons which interact with electrons to generate two opposing photons, a process called *annihilation*. These photon pairs are detected by the tomography systems to create an image.   PET imaging is often used clinically to scan the whole body and evaluate for metastatic disease or recurrences. PET scanning is increasingly utilized clinically to inform whether a lung nodule is metabolically active (for glucose metabolism) which can be used to distinguish between malignant (cancerous) and benign (non-cancerous) cases. PET studies have achieved an overall sensitivity of 96%, specificity of 79% and accuracy of 91% in distinguishing benign and malignant nodules between 1 and 3cm in diameter [17-19] however, the resolution of the PET scanner means that only nodules greater than 1cm should be scanned in this manner. The limitations of PET systems include poor spatial resolution, a lack of structural detail, and surprisingly, no one knows whether the PET scan positivity relates primarily to the inflammatory component, or to the malignant one [19]. This is an area that could benefit in the future from the methodologies developed in this study.

Techniques such as X-ray and MDCT, involve the detection of x-rays transmitted through tissue. Areas of high density, such as bone, limit x-ray transmission while low density tissues, such as lung parenchyma, permit x-rays to pass through. The resulting image data reveals the structural composition of the subject based on densities of the different tissues. These techniques provide information regarding nodule size, margin characteristics, calcification and cavitation. Through multiple studies conducted over time, MDCT can be used to calculate nodule growth rate.

2.4 Computed Tomography as a Screening Method
for Lung Cancer

The successes in screening for cancers such as cancer of the cervix, breast and the prostate, along with the widespread introduction of MDCT scanning into large and small medical centers, has lead to the re-discussion of lung cancer screening. A number of projects investigating the feasibility and best methodology of lung cancer screening are being conducted both in the United States of America and in Europe. The Early Lung Cancer Action Project (ELCAP) began in 1992 to assess the usefulness of annual MDCT screening for lung cancer. In 2002 the National Lung Screening Trial (NLST) was launched to investigate if lung cancer screening by either x-ray or MDCT actually saves lives [20]. A similar study was initiated in the Netherlands in 2003; Nederlands- Leuvens Lonkanker Screenings ONdersoek (NELSON) [21]. These lung cancer screening studies have shown MDCT can detect early stage lung cancer [22, 23]. While the complete mortality data for these studies is not yet available, an evaluation in 2004 by the U.S. Preventive Services Task Force of all available lung cancer screening data found they 'could not determine the balance between the benefits and harms of screening for lung cancer' [24]. One of the primary concerns with lung cancer screening using MDCT is what to do with an identified new nodule and in many cases the recommendation is to watch the nodule for signs of growth over 3 to 24 months. This time period is required, as a large number of early detected nodules (97%) are false positives (non-cancerous) and will not require treatment. Tracking growth is one way of separating non-cancerous nodules, which are unlikely to change over this time period, from cancerous ones. Unfortunately, it is obvious that early detection (finding a suspect nodule) does not

translate into immediate therapy (resection) – the lung cancer paradox. Thus, it is important to develop the most effective possible ways of staging these early detected nodules as well as reliable methods for tracking nodule growth.

2.4.1   Current Computer Aided Detection and Diagnosis Methodologies

Computer aided detection (CAD) algorithms incorporate image analysis and pattern recognition techniques in order to identify and describe areas of interest in an image dataset. In many cases, the area of interest is a specific pathology such as a lung nodule in a chest MDCT scan. CAD systems are becoming increasingly important in the clinical setting, serving as a second reader in image interpretation, effectively improving detection accuracy and consistency [25, 26]. In some cases CAD systems also incorporate a diagnosis element which aims to classify an identified structure as normal or abnormal, or in some cases will stage pathology in degrees of severity. For example, a lung nodule may be classified as malignant or benign [27] while emphysema may be staged from none to extensive [28-30]. In order to develop a CAD system a complete dataset is required containing numerous cases for which the final diagnoses are known, this is known as a 'training dataset'. From the training set a group of features are established to describe each of the diagnoses, or 'classes'. The training dataset is used to guide the development of the CAD algorithm, by comparing the features of an unknown case to the features of the training set, and seeing which class it most closely matches. Therefore, if the final diagnosis in the training set has only two classes – malignant and benign – then they are the two possible outputs when the CAD algorithm is applied to a

new case. What is needed for the improvement of current CAD technology are training sets with more specific classes including; cancer, necrosis, inflammation and fibrosis.

2.4.2 High Resolution Imaging

During the last several years, higher resolution radiological techniques have came to market to image small anatomical structures and animal models, including micro-CT [31-33]. These systems have a restricted field of view and hence can image only small animal models *in vivo*. For human studies using micro-CT systems, tissue samples less than 5cm in diameter are examined *ex vivo*.

Micro-CT imaging has been employed in osteoporosis research, analysis and quantification of bone growth and bone repair models, dentistry research and the analysis of cardiac vascularization in mice and rats [34, 35] to mention only a few topics. Recently micro-CT is being increasingly employed to investigate pulmonary structure and function in normal and diseased mouse models [36-39]. Chang et al. recently used a microPET/micro-CT system to measure the growth and response of lung carcinoma to a variety of treatments in mice [40].

Electromagnetic radiation generated in a synchrotron system can be used to generate x-ray beams of specific energy, and hence much higher resolution tomography image dataset than achievable on stand alone MDCT or micro-CT systems. At the Advanced Photon Source (Chicago, IL, USA) a synchrotron micro-CT system has been developed, capable of acquiring datasets at 100 nm resolution [41]. At the Photon Factory (Tsukuba, Japan) contrast synchrotron CT has been used to examine human colon and liver carcinomas in great detail [42].

In summary, MDCT, micro-CT and synchrotron CT systems all function under the same principle of the detection of x-rays transmitted through the tissue. However, a compromise is made between imaging field of view and resolution. MDCT, micro-CT and synchrotron CT have respectively lower fields of view and higher resolutions.

2.5 Histopathology

2.5.1 Traditional Pathology Techniques

Biopsy pathology provides a means for a more detailed investigation of lung nodule structure and cellular composition. Typically, pulmonary pathology involves the collection of a small biopsy sample of tissue, which is then fixed, embedded, sectioned, stained, mounted on a slide and examined via microscopy.

Histological staining is used to provide contrast between the structures being examined. The most widely used histopathological staining technique used for medical diagnosis is hematoxylin and eosin (H&E). The H&E stain combines the basic stain, hematoxylin which binds to structures in the nuclei and an acidic dye, eosin which binds to the proteins in the intracellular or extracellular spaces. This staining results in nuclei stained blue, cytoplasm pink and red blood cells intense red. H&E is the most common staining approach used in clinical pathology for the diagnosis of lung cancer, however immunohistochemistry is also being increasingly employed in the clinical setting to aid with diagnosis [43]

2.5.2 Immunohistochemistry

Immunohistochemistry is a process used to identify proteins in a sample by using tagged antibodies which bind specifically to antigens in the tissue. There are two main

approaches to visualizing the antibody-antigen interaction, immunoperoxidase staining and immunofluorescence [44]. Immunoperoxidase staining involves an antibody conjugated to an enzyme which catalyses a color producing reaction. Immunofluorescence uses an antibody tagged to a fluorophore which emits fluorescence when stimulated with the appropriate wavelength of light [44].

The immunohistochemical tagging of an antigen can be accomplished using either a direct or indirect method. The direct method is a single step process in which the labeled antibody binds directly to the tissue antigen. The indirect method involves two stages; an unlabeled primary antibody tags the tissue antigen and then a second, labeled antibody tags the primary antibody. While the direct method is simpler, the indirect method is most commonly used as it is more sensitive. A higher sensitivity can be achieved with the indirect method, through signal amplification where the labeled secondary antibodies bind to several different sites on the primary antibody.

Immunohistochemistry was first utilized as a diagnostic application to classify renal disease in the 1940's [45]. Since then immunohistochemistry has been used in the clinical environment for the diagnosis of immunodeficiency disorders, renal diseases and skin diseases [45]. Immunohistochemistry also has a role in surgical pathology for the diagnosis of tumors, particularly those which are poorly differentiated. These techniques are also utilized to determine the primary site of metastatic epithelial tumors.

A number of studies have been conducted examining the use of immunohistochemical staining for the identification and staging of lung cancer nodules. Marson et al. reported there was a significant correlation between the immunohistological expression of cytokeratins (CK), of a range of molecular weights, and thyroid

transcription factor (TTF1) to the histological type and grade of lung tumors [46]. From this study the CK7 positivity was presented for various primary lung tumors as follows; adenocarcinoma (100%), squamous cell carcinoma (23%), bronchio-alveolar carcinoma (93%), small-cell carcinoma (37%), large-cell carcinoma (40%), large-cell neuroendocrine carcinoma (70%) and carcinoid tumor (4%). Chu et al. conducted a similar study in which the expression of CK7 and CK20 were evaluated in the carcinomas from various organ systems. This study also concluded 100% of lung adenocarcinomas were CK7 positive [47].

Scarpatetti et al. noted there was no precise relationship between the origin of the tumor and the immunohistochemical response to the panel of CK antibodies. However, with statistical analysis insightful relationships were uncovered showing lung adenocarcinomas could be separated from metastatic adenocarcinomas with a specificity of 95% and a sensitivity of 63% when using a sequential staining scheme, using CK18, CK5 and high molecular weight (HMW) CK [48].

It has been reported that CK8 is expressed in all NSCLC; Pendleton et al. reported 100% of NSCLC biopsy specimens were CK8 positive [49] while Blobel et al found all adenocarcinomas of the lung positive for CK7, CK8, CK18 and CK19. Also, positive CK8 and CK18 staining were reported for squamous cell carcinomas [50].

2.5.3 Limitations of Traditional Pathology Techniques

Issues involved with the traditional pathology techniques include the analysis of small individual sections of tissue, the absence of spatial correlation between samples, severe shrinkage of tissues due to the dehydration processes associated with fixation and the very high level of manual manipulation of the tissue required during processing. The

usual pathology sample from transbronchial forceps biopsy is the size of a match head and from needle aspiration cytology, a small sampling of cells. These sample sizes are too small to inform on the heterogeneity of the lung nodule. Even a pathology slide made from the lung nodule completely removed at surgery is typically 1 cm x 1cm x 4 microns. This is a trivial amount of tissue, less than 0.0004% of the nodule volume, if the nodule is 1cm^3.

2.5.4   Previous 3D Pathological Investigation of Tumors

With the recent movement towards the 3D radiological analysis of lung cancer nodules has come the realization that nodule growth is likely best assessed as a volume change rather than as a diameter change [51, 52]. The complex structural content of cancer nodules often hinders the ability of 2D planar data to be representative of the 3D volumetric case. Traditional pathology techniques fall directly into the category of taking a small 2D sample of tissue to represent the entire nodule content. It thus follows that steps must be taken to create 3D histopathological datasets. To date, 3D lung histopathology has not been investigated. The complexity of histopathology imaging studies of non-lung nodules in 3D range from the basic segmentation and rendering of gastric carcinoma volumes [53] to complete reconstructions of human uveal melanoma content [54]. Volume rendering can provide some valuable information regarding tumor size, shape and surface topography but is not useful beyond these parameters. Three-dimensional histological renderings of prostatic adenocarcinoma [55], hypopharyngeal cancer [56] and brain tumors [57] have been created. Boag et al. used digitized microscopic images from serial sections of paraffin embedded prostate adenocarcinomas to create 3D reconstructions [55]. In doing so they found that 'although apparently

separate in the 2D images of routine microscopy, 3D renderings demonstrated that the glands of Gleason grade 3 adenocarcinoma in fact interconnect to form a complex array of intertwined tubules.' Hence, while there are limited publications relating to cancer structure in 3D, it is clearly an important area for future investigation. This is particularly true for lung cancer, the structure of which has not yet been investigated in three-dimensions.

2.5.5 Lung Nodule Heterogeneity

The term 'lung nodule heterogeneity' is loosely defined and interpreted differently based on medical specialty. On a cellular level, the term heterogeneity has been linked to cancer nodules containing more than one histological sub-type [58]. In this case the heterogeneity is in reference to histological classification of mixed sub-types in the cancerous portion of the nodule. The purpose of describing this form of cancerous tissue heterogeneity was, and remains, important for determining prognostic and treatment guidelines [43]. More recently, lung nodule heterogeneity is being described and evaluated, extending beyond the cancerous component of the nodule to included vascularity, fibrosis and necrotic tissue [59, 60]. These initial studies have indicated there is a statistically significant correlation between the proportions of necrosis and fibrosis within a lung nodule and patient outcome. This may lead to future sub-type classifications of lung nodules extending beyond the cancerous portion to also include non-cancerous tissues of the nodule. However, large challenges are already faced in obtaining consistent classifications of histological sub-typing due to a large dependence on subjective assessment and the reliance on 2D assessment of a potentially complex 3D biomass. Automated systems which aid in the identification and quantification of

important features with in the histopathological data would not only be highly valuable for the future exploration of possible sub-type to outcome relationships but also necessary if the complexity of sub-type classifications incorporated into the clinical setting continues to escalate.

From a radiological perspective, lung nodule heterogeneity is a term used in reference to the grayscale Hounsfield Unit (HU) variance as viewed in MDCT data. It is widely accepted that the radiological representation of lung nodules may be linked to the histological sub-type, such as ground-glass opacities correlated to BAC [61-63]. What is poorly understood is if the correlation between HU variance and underlying tissue structure extends beyond the solid and non-solid nodule components, to the level of cellular tissue types.

From a pharmacological perspective, the lack of consideration given to lung cancer heterogeneity has lead to the acceptance of misleading standards for the evaluation of treatment approaches [64]. Currently, clinical research has been directed towards shortening the drug assessment cycle by replacing the survival statistic with radiologically obtained indications of response [65]. Many pharmacological studies evaluating new cancer treatment agents, use a reduction of nodule dimension in MDCT as indication of positive response, however this measure does not take into consideration that the nodule biomass may contain a large proportion of other non-cancerous cell types [66]. Hence a drug may be effective in destroying all cancer cells without significantly reducing the nodule mass, an occurrence not factored into the assessment methods, likely leading to effective drugs being falsely classified as un-effective.

2.6   Medical Image Registration

Registration is the process of aligning one image to another. More specifically, it involves finding the transformation which map points in the moving image to corresponding points in a fixed source image. The basic components of a medical image registration algorithm are the two image sets (moving and fixed), an interpolator, a metric, an optimizer and a transform. The transform determines how the points (or pixels) from the moving image are mapped to the fixed image. The interpolator extrapolates the image intensities of the moving image. The metric is used to evaluate how well the moving image corresponds to the fixed image and the optimizer modifies the parameters of the transform based on feedback from the metric [67].

The process of image registration is widely used for medical applications such as tracking disease progression over time, fusing multimodal images, inter-patient comparison and atlas based segmentation [68-70]. The selection of the registration components is greatly influenced by the type of images at the input and the level of accuracy required at the output.

2.6.1   Histopathology to Computed Tomography Registration Methods

For many years, it has been acceptable to correlate pathological findings with external imaging observations by general comparison. Kennel et al. presented histology sections of mouse lungs next to roughly corresponding micro-CT and used arrows to highlight suspected tumor foci in the micro-CT data and confirmed foci in the histology images [33]. For this case, registration of the datasets may have made a more compelling statement but it was not necessary to meet the objectives of the study. A prospective

pilot study was conducted by Chan et al. to correlate tumor size on CT with tumor sizes measured microscopically in non-small cell lung cancers [71]. In this basic study the resected tumor was manually aligned to the approximate orientation of the *in vivo* CT imaging plane and transverse sections for histological preparation were made. The boundary from the CT data was then compared to that of the histological sections with the conclusion that CT over-estimated the size of the tumors to a degree greater than what could be accounted for by the significant tissue shrinkage through histological preparation. This study effectively avoids the possibility of creating false matching by using only rigid registration, but in doing so the investigators are unable to make any quantitative conclusions regarding boundary representation. While the objectives of this study are highly important, it is clear that a methodology with greater precision, repeatability and accuracy is required.

For more specific comparative studies such as boundary representation or spatial distribution of tissues, the distortions caused by histological processing must be compensated for in some way. Some studies have investigated the non-rigid alignment of histology to non-destructive data, such as MRI and micro-CT. Jacobs et al. employed a modified head and hat surface-based registration to align rat brain MRI data to corresponding histological sections [72]. Clarke et al. composed a multimodal mapping of carotid atherosclerotic plaque components [73]. The approximate matching MRI and micro-CT slices were located by visual inspection and non-rigidly registered using the Delaunay triangle method. Registration approaches such as these, involve a high risk of introducing false information due to the non-rigid registration of the histology data to a

non-corresponding image slice. A spatial reference system is required such that the correct corresponding base image can be located for each warped image.

2.7 Significance and Innovation

It is the goal of this study to create 3D, cross-registered, multimodal datasets relating the complex histopathology of lung nodules from a cellular level to density based representations. To facilitate the creation of this dataset an automated tissue characterization process for the histopathology data is required to replace manual segmentation. The datasets will be used to investigate the ability to distinguish the tissue types in micro-CT and MDCT data.

The quantified distribution of tissues within complete, lung cancer nodules sections has not been previously investigated. Defining tumor tissue distributions in 2D and 3D will reveal insight into the projected growth patterns of these nodules. This information may impact treatment development in forms of delivery methods and how best to target the malignant components of a nodule. The spatial relationship between tissue types with the nodule is also extremely important for establishing guidelines for traditional biopsy sampling and new optical biopsy techniques. If large degrees of heterogeneity occurred in lung cancer tumors, it is likely that cancerous regions may be missed by biopsy. This is particularly pertinent for new optical biopsy techniques which possess a relatively small field of view. With knowledge of the average variance in cancerous tissue distribution versus other tissues in lung nodules, statistical measures may be used to guide the resolution and sampling field required for reliable biopsy results.

The creation of a process model for the acquisition and registration of histopathological data to corresponding *ex vivo* and *in vivo* MDCT data has high potential to impact clinical and research arenas. As there is no complete database containing registered lung nodule pathology to corresponding MDCT data, it is impossible to evaluate the extent of information which is captured in an MDCT scan. This correlation may have a clinical significance in bridging radiological and pathological classifications of lung nodules. This is of particular interest in the case of classifying BAC subtypes which are linked to differing prognostic statistics [11]. Further understanding of the relationship between lung nodule tissue type and MDCT representation will impact research in computer aided diagnosis algorithms while also providing a valuable means of validation.

This method of correlation between non-destructive imaging and histopathology would also be of value to industry. Research and technological development are largely driven by a 'needs' basis. It is required to evaluate the capabilities and limitations of available systems in order to determine new 'needs' for the non-invasive imaging and evaluation of lung nodules.

The generation of an automated segmentation algorithm for lung nodules in histological data may have a significant impact on patient care in the future. This may be an effective method of evaluating the ability of pathologists to correctly estimate the proportion of tissue types within a nodule. Research may also be conducted, as to the prognostic significance of varying proportions of tissue type within a nodule. Current research has already indicated that the proportion of fibrotic tissue within a nodule can be correlated to increased grades of invasion, while the proportion of cancerous tissue

remains relatively constant [74]. However, the manual identification of these tissue

proportions by expert pathologists is neither time nor cost efficient and the results are

also susceptible to human error.
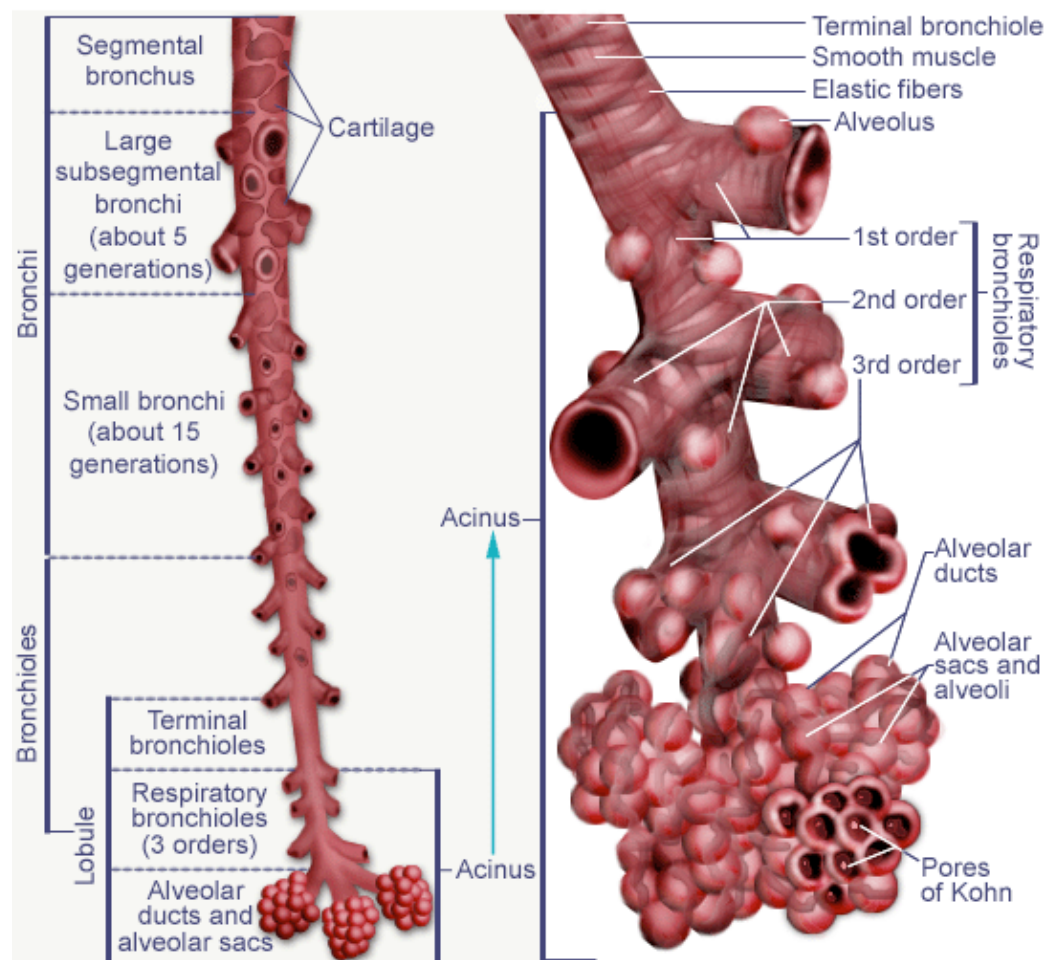
Figure 2-1: The pulmonary airway structure. Graphical depiction of the pulmonary airways showing the conducting and respiratory airway segments. The Acinus is the terminal, respiratory region of the airways and includes the respiratory bronchioles, alveolar ducts, alveolar sacs and alveoli.

Reproduced from Elderlman 2006 [75]
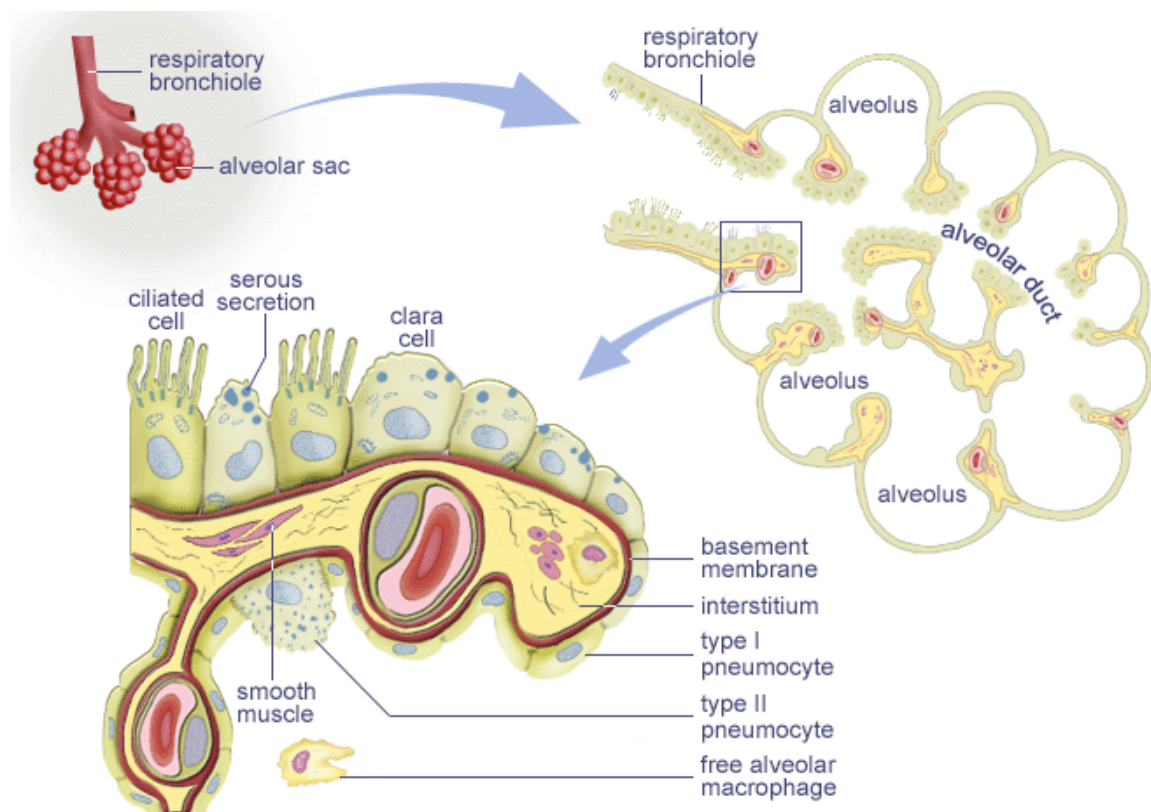
Figure 2-2: The terminal acinus. Highlighted is the cellular composition of the respiratory bronchiole wall which contains ciliated cells, serous secretion and clara cells while the alveolar wall consists of type I and type II pneymocytes.

Reproduced from Mckee 2007 [76]

CHAPTER 3

DATA ACQUISITION PIPELINE

3.1   Tissue Fixation

The structural integrity of the lung tissue must be maintained through the fixation process if direct comparisons are to be made between the *in vivo* image datasets and those following fixation. Formaldehyde fixation or air drying fixation affect the size and density of the lung, causing discrepancies between pre- and post-fixation measurements. A modified Heitzman technique has been employed to fix lung tissue, maintaining the structural integrity of the parenchyma [77]. Using a glycol rich fixation solution the water in the lung tissue is replaced by oil and fixed at a lower inflation pressure than traditionally used – a pressure we have established after similar work in mouse and sheep lung.

We have conducted a number of studies to investigate the optimal fixation pressure for radiological and pathological correlations. Imaging is performed with lungs *in situ* at an applied airway pressure. Then fixation is performed at that airway pressure, using the glycol fixation technique and image data is collected again. Volumetric comparisons are drawn between the lung volumes *in situ* and following fixation. Tracing of the lung boundaries for volumetric assessment in MDCT and where applicable, micro-CT datasets are conducted using our PASS software. PASS is a pulmonary analysis software developed to aid the objective of other studies within our laboratory, and which features a number of analysis tools, including boundary tracing. We have demonstrated that for sheep lungs the inflation pressure to fix the lungs, and keep them at the same lung volume as total lung capacity in the intact animal, with maintenance of the lung density

on MDCT scanning is of the order of 18 cm H_2O pressure [78]. Using mice models it has been demonstrated that a statistically insignificant difference of 0.6% between *in situ* and fixed lung volumes can be achieved with a fixation pressure between 10 and 15 cm H_2O [79]. In accordance with these results a lung fixation pressure of 15 cm H_2O will be used for the fixation of the resected human lung lobes used in this investigation.

3.2   Large-scale Image Microscope Array (LIMA)

A system was required to bridge the gap between non-destructive imaging techniques, such as MDCT, and destructive histopathological imaging. Current processing techniques do not retain any spatial correspondence between the generated histopathological slides and the original 3D structure of the tissue of interest. Hence, a purposefully built system was developed – the large image microscope array (LIMA). The LIMA system consists of both innovative hardware and software developments to allow tissue to be sectioned and imaged with the relationship between the generated slides and the original 3D structure maintained.

3.2.1   LIMA Hardware

Due to the lack of an appropriate sectioning instrument in the commercial market, a series of products were purchased, modified and assembled to constructed the LIMA system [80]. In summary, the adaptation of the Leica SM2500 microtome involved the incorporation of a vibrating blade, stepper motor control of the stage and a stage locking mechanism. The vibrating blade system incorporates a base bracket that allows adjustment of the height of the knife with respect to the stage. The knife itself is a 26cm surgical trimming blade held via a vice mechanism in the vibration frame. Hence the

LIMA system is capable of sectioning a very large field of 25 x 25 x 25cm. The stepper

motor control of the stage ensures precision by generating a uniform cutting speed and

also permits the interfacing to a computer. The stage locking mechanism is also

computer controlled and acts as a "photo stop" where by each time a section is made the

stage is brought back to the same location and locks into place.

The image acquisition component of the LIMA system hardware is composed of a

Leica MZ16a stereomicroscope set up with a 1x planar achromatic objective and a 10x

eyepiece. This specific microscope carrier system was selected due to its flexibility in

performing bright field and fluorescent microscopy of stained and non-stained specimens.

The microscope has a magnification range of 7.11x to 115x at a maximum resolution of

420 line pairs in this configuration. Digital image attainment is achieved through the use

of a JenOptik charge-coupled device (CCD) digital camera. The active image area of the

CCD is a 1300 by 1030 pixel array with a pixel pitch of 6.7µm. The camera has a multi-

shot feature, which can be used to increase the effective pixel count from 1.3 mega pixels

to 12 mega pixels. Image data acquired by the camera is transferred via a FireWire IEEE

1394 to the computer.

The microscope and CCD camera coupling is maneuvered over the sectioned

tissue surface via a Velmex Bislide parallel coupled gantry system. The gantry consists

of a 94cm parallel bislides in the y-axis, an 112 cm single bislide in the x-axis and a 25

cm single z-axis bislide. Control of the gantry motion is automated via three computer

interfaced stepper motors. At present a standard computer containing an AMD 1800+

chip and 60 GB hard drive is being utilized to control and capture the images. **Figure 3-1**

is a photograph of the hardware setup.

A high level of precision is needed in sectioning tissue samples and in positioning the microscope and camera couple. Add to this the highly repetitive and time-consuming nature of multiple, sequential image acquisition and the necessity of system automation is apparent. Microsoft's .NET framework, specifically Managed C++, was chosen as the programming language used to design the controller for the automation of the LIMA hardware and to provide an efficient graphical user interface.

3.2.2   LIMA Sectioning and Imaging Process

The LIMA system is advantageous due to maintenance of the spatial correspondence between subsequent tissue slices. This is achieved by the sequential imaging of the tissue surface en bloc, followed by the sectioning and removal of that tissue slice. The images are acquired via an automated raster scan, for which the microscope and CCD camera is moved over the tissue surface via the bi-slides. Once the images of the surface are acquired and stored, the vibrating microtome cuts the tissue and the user removes the cut tissue slice. This process is repeated to image and section the complete specimen. An automated image processing approach has been developed to generate large complete composite images from the multiple raster acquired sub images, as described below. The tissue slices from sectioning can be processed via standard histological processing to generate histological sections. This specimen sectioning and imaging procedure using the LIMA system is summarized in **Figure 3-2**.

3.2.3   LIMA Image Content

The LIMA system was designed to provide a means of gaining 'ground truth' data for comparison to MDCT images. The LIMA images can be representative of structural ground truth as image content is in no way dependent on reconstruction algorithms. The

images produced by the LIMA system also serve to bridge the gap between whole organ density data from MDCT and small scale cellular data from histopathological processing. A large benefit of the LIMA data is the color content, which aids in the distinction between structures within the image. The large depth of field in the images can also aid the understanding of spatial relationships between structures. **Figure 3-3** (A) illustrates the detail captured in a low magnification LIMA image of glycol fixed sheep lung, reflecting the branching of a bronchi followed by a pulmonary blood vessel. **Figure 3-3** (B) shows the detail achievable at the highest magnification, where the shape of individual alveoli can be seen.

3.2.4   LIMA Post Processing

Our aim to generate image datasets consisting of MDCT images along with high-resolution images generated from the LIMA system, required the construction of a system physically capable of creating large-scale sections and collecting high magnification images. To capture the microscopic structural detail of nodule configuration and surrounding alveolar sacs in the resected lung tissue, a 10x magnification is required and consequently the field of view is sacrificed. Many small high-resolution sub images are therefore captured per slice of lung. The system hardware alone could not be relied upon to generate the precisely contiguous digital color sub images. Image registration was employed to correct misalignment in adjacent sub images such that a high-resolution large field composite could be generated.

For the completion of a full LIMA system dataset a large number of images would be required and hence image processed. With this in mind, the technique chosen needed to be robust and require little to no human interaction. With regards to the

creation of three-dimensional renderings as well as the integration of segmentation and analysis tools, a flexible, open source C or C++ based toolkit was desired. The National Library of Medicine Insight Segmentation and Registration Toolkit (ITK) fulfilled these requirements.

An image mosaic algorithm has been developed for creating large field composite images for the LIMA, **Figure 3-4** [81]. The algorithm consists of a number of separate filters required to perform the various stages involved in creating the large field composite image. A filter exits to extract the overlapping regions within two image tiles and converts these extracted regions to grayscale such that further image processing computation is simplified. A registration algorithm consisting of a translation transform, a linear interpolator, a regular step gradient optimizer and a mutual information metric determines the alignment transform required to match the two overlap regions. The alignment transform is then applied to the images to correct the respective misalignment. These filters have been linked such that no human interaction is required after initiation, to create a complete large image composite corresponding to each slice of tissue.

To test the ability of the registration algorithm to reliably align the color images produced by the large image microscope array system, a simple test was performed using sheep lung tissue. For a low (7.11x) and a high (115x) magnification source images, the same process was applied to examine the registration algorithm accuracy. From the source image of dimensions 1300 x 1030 pixels, two cropped images of dimensions 500 x 500 pixels were obtained using the region crop filter. One image was obtained with a 40 pixel displacement in the x-axis and a 30 pixel displacement in the y-axis with respect to the other. The two cropped images were then input to the registration filter. The

registration algorithm was able to exactly recover the applied translation parameters for both low and high magnification images, requiring computation time of less than two seconds.

Figure 3-5 illustrates the cropping of the regions and the result of the difference filters before and after the registration process was applied. The differences in image content are created by subtracting the pixel graylevels in one image from the corresponding pixel graylevels in the other. If both pixel graylevels are equal then the result of the subtraction should be a black pixel. As can be seen below, the difference is great in the before registration image and removed after registration. The band of gray pixels at the border of the after registration image is due to the correction in misalignment between the images.

3.3   Lung Nodule Data Acquisition Process

3.3.1   Lung Nodule Selection

A multi-disciplinary team was established to facilitate the identification of suitable patients and the collection of resected tissue. The team consisted of cardiothoracic surgeons, nurses, surgical pathologists, pulmonologists, and biomedical engineers. The University of Iowa Institutional Review Board (IRB) approved the collection of resected lung tissue for the purpose of this study.

Patients identified for the study by the multi-disciplinary team were required to meet the following inclusion criteria:

- Requiring surgical resection
- Peripherally located lung nodule

- Estimated nodule cross section less than 2.5cm in at least one axis

Exclusion from the study resulted in the case of;

- Cellular classification of type other than adenocarcinoma

- Pregnancy

- Under the age of eighteen

Sputum, risk factors, pre-operative x-ray, MDCT and/or PET imaging and in some cases biopsy samples were used by the cardiothoracic team to evaluate a suspicious nodule and determine if surgery was required. Surgical resection may occur via wedge resection or lobectomy based on the location, size, stage of the nodule and preference of the cardiothoracic surgeon. A wedge resection involves the removal of a small portion of the lobe while in a lobectomy procedure the entire lobe is removed. However, wedge resected specimens were utilized for protocol experimentation and testing only as these specimens could not be accurately registered. The optimal cases for the study were lobectomy procedures such that inflation of the lobe for imaging and fixation purposes was possible. This inflation was necessary in the registration of the datasets to a common coordinate system.

A restriction to the size of the nodule was applied because a pathology slide has a maximum sample field dimension of 2.5 cm. Hence, as it was desired for a complete nodule cross-section to mount on a single slide, there had to be at least one orientation of the nodule less than 2.5 cm in diameter.

Efforts were made to target nodules of specific cellular classification, namely adenocarcinomas and benign nodules, based on the peripheral location restriction, pre-

surgical data and opinions of the multi-disciplinary team. If available via pre-surgical biopsy, cellular classification was used to identify patients for exclusion.

3.3.2   Patient Recruitment and Consent

Recruitment for study participants is conducted through the University of Iowa Hospitals and Clinics cardiothoracic team, the Holden Comprehensive Cancer Center and through the Bioengineering Research Partnership Grant study. During the patient's pre-surgical visits a brief history and physical exam was performed, the study was described in detail and consent was obtained. Women of child bearing age were requested to undergo a pregnancy test.

Between October 2005 and May 2008, 17 suitable patients were identified and consented for participation in the study, 8 males and 9 females. Six of these 17 potential datasests were not collected due to changes in surgical plan or surgical pathology processing requirements. Of the remaining 11 datasets, there were 7 adenocarcinomas, 3 squamous cell carcinomas and 1 neuroendocrine carcinoma.

The age of the patients ranged from 51 years to 79 years with an average body mass index of 26. The study pool included nine females and two males. The ethnicities of the subject pool included ten Caucasian patients and one African American patient. Ten out of the eleven patients had a significant smoking history with the average consumption being one package of cigarettes per week for the past thirty years. The only patient enrolled in the study without a smoking history, lived for over fifty years with a heavily smoking spouse.

From pre-surgical MDCT imaging, the nodules diameters ranged from10mm to 50mm. The staging of these nodules ranged from stage IA to stage IIIA, based on the

final surgical pathology reports. **Table 3-1** presents a summary of the characteristics, diagnoses and staging of study patients from whom data was collected. Of the patients listed in Table 3-1, one squamous cell carcinoma was used to experiment with fresh and fixed micro-CT imaging protocols, one adenocarcinoma case was used for a synchrotron imaging study and another adenocarcinoma case had a poor fixation result affecting the histopathological data collection. Hence, extensive data was collected for these datasets but as they were not complete they were excluded from the analysis portion of this study.

3.3.3 Image Acquisition and Tissue Processing Overview

An extensive image acquisition and tissue processing scheme was established in order to generate a comprehensive dataset for human lung nodules which is summarized in **Figure 3-6**. **Table 3-2** depicts a timeline chart of the process highlighting the work required to obtain one dataset. In all, it requires approximately 236 hours of processing time to acquire all the data for a single nodule. This time estimate does not include the lag time often encountered when waiting for specific equipment to become available for use.

3.3.3.1 *Pre-Surgical Scan Data Obtained*

As a component of a lung cancer patient's clinical care, a pre-surgical MDCT scan is required by the multi-disciplinary cancer service at the University of Iowa. This pre-surgical dataset may be collected prior to or following consent of the patient. In consenting participation in the study the patient permited access to previously collected imaging pertaining to the lung nodule. From the pre-surgical MDCT data, information relating to the density of the nodule *in vivo* was gained.

3.3.3.2 Surgery

On the patient's day of surgery, they were transported to the operating room for the resection procedure. The resection procedure was conducted by the cardiothoracic surgical and anesthesiology teams. Their involvement, including administering pre-procedure medications, conscious sedation, performing tissue biopsy, obtaining consent for surgery and post surgical care, is part of clinical care. The amount of tissue removed from the patient during surgery was independent of study participation and was determined by the cardiothoracic surgeon.

The surgical pathology team evaluated the specimen following resection. In the case of a wedge resection procedure, the pathologist immediately removed a segment of the nodule for diagnosis such that the surgical team may determine if a full lobectomy was required. The removal of this tissue prevented the subsequent inflation of the tissue and hence deemed it insufficient for inclusion in the complete study protocol. These tissue samples were however useful in the development of the study protocols. With a lobectomy, initial measurements and removal of the accessible lymph nodes were conducted immediately following the patients surgery but the nodule was not disturbed. As no immediate action was required for patient care, the final diagnosis was made using the histology sections generated by this study's processing procedures.

3.3.3.3 Tissue Imaging and Fixation

Following evaluation of the tissue in surgical pathology, the lung lobe specimen was cannulated through the major airway and inflated with air to 15 cm H_2O pressure. With the resected lobes, in which there was some connection to surrounding lobes, surgical staples were used to prevent air pressure escaping. With the lung lobe inflated, it

was MDCT scanned. The tissue was then fixed via the modified Heitzman fixation technique at the same airway pressure and air-dried in a drying oven. Following fixation the fixed lung lobe was MDCT imaged again.

Wedge resection specimens could not be inflated due to the absence of a major airway inlet and due to the removal of part of the nodule in pathology. Hence, these specimens were fixed and dried without inflation.

3.3.3.4   *Nodule Separation and Imaging*

Following fixation and using MDCT scan control, the nodule and the immediate surrounding tissue was dissected *en-bloc* from the lung lobe using a thin autopsy knife. By using MDCT scan control the nodule could be resected as a cube of tissue with precision and without damaging the structural or spatial relationships. The isolated nodule was then imaged again using the MDCT, and immediately afterwards using the micro-CT scanner. Removal of the nodule from the lung lobe was required due to the small imaging field of the micro-CT. Then the nodule was sectioned and imaged using the LIMA system. Of the sections created by the LIMA, every intact slice underwent histological processing.

3.3.4   Nodule Stabilization

It is desirable to minimize the complexity of registering the multimodal datasets. Hence some hardware was created to stabilize the isolated nodule between MDCT, micro-CT scanning and LIMA sectioning. The MDCT and micro-CT imaging systems, acquire data in the transverse plane, while the LIMA system cuts and images along the coronal plane. To ensure there was a consistency between the direction and angle of the imaging planes, the specimen was mounted to a polyethylene base plate. The

polyethylene material was selected due to its very low attenuation in radiological imaging systems. Parallel grooves were etched into the base plate to facilitate tissue attachment. Nylon screws were used to vertically attach the base plate to a polymethyl methacrylate mount arm. The specimen was imaged in this configuration on both the MDCT and micro-CT systems.

For sectioning on the LIMA system, the base plate was removed from the mount arm and attached horizontally, with metal screws, to the specimen stage. Interlocking metal stacking support brackets were attached to the base plate and the interior space was filled with agarose. When set, the agarose formed a solid support media to prevent tissue motion during sectioning. The metal stacking support brackets and metal base plate became cold as the agarose solidified in a refrigerated environment. During sectioning the metal insulated the agarose, keeping it cool and hence at an ideal solid consistency. The 4 mm metal stacking brackets were removed one by one as the specimen was sectioned. Images of the developed nodule stabilization system are shown in **Figure 3-7**.

3.3.5   Imaging Protocols

3.3.5.1   Computed Tomography

The scanning protocols for the clinical *in vivo* MDCT scan varied across institutions and also with in the University of Iowa Hospitals and Clinics. All clinical MDCT scans were conducted with a 120 kV. However, the current ranged from 60 to 127 mAs, the pixel spacing from 550 to 1000 microns, slice thickness from 2.5 to 5 mm. There also existed a great variation in the reconstruction kernel, including B20, B30, B31, B40 and B70. The MDCT scanner systems also ranged including models from Siemens, Phillips and Genesis. For an average sized patient a chest field of view

measured approximately 30 cm^2. The majority of patients were imaged with some form of contrast agent.

IRB approval was gained to conduct high resolution, perfusion imaging of the patient prior to surgery. This protocol began with a low dose volume MDCT on a Siemens Somatom Sensation 64 with 120 kV, 40 mAs with 750 micron slice thickness, to locate the nodule. Following, a 3-second bolus delivery of contrast agent would be delivered and 15 scans over 2.8cm would be acquired using cardiac gating, 80 kV and 150 mAs. Unfortunately no patients participating in the study consented to this pre-surgical imaging portion of the study.

For the MDCT scanning of the resected lobe containing the nodule the protocol and the effective field of view were different from that of the *in vivo* case. As the lobe was resected from the patient the field of view was reduced to approximately 15 cm by 15 cm, and hence the resolution increased. Also, radiation dose was no longer a factor in the protocol selection. A Siemens Somatom Sensation 64 MDCT scanner was used, scanning 120 kV, 140 mAs with 0.6 mm slice collimation for the fresh resected lobe, inflated to 15 cmH$_2$O, and fixed lung lobes.

For the imaging of the isolated nodule the protocol again uses 120 kV, 140 mAs, with a slice collimation of 0.6 mm on a Siemens Somatom Sensation 64 MDCT scanner. The field of view for the isolated nodule scans could be greatly reduced and hence the highest resolution from the MDCT was achieved.

From the MDCT scans conducted we gained density information, with increasing in plane resolution (pixel sizes of approximately 600 micron to 100 micron)

relating to the cancer nodule in vivo, in the resected lobe and in the nodule physically

isolated from the lobe.

3.3.5.2 Micro-CT Processing

We developed protocols for the SkyScan1076 micro-CT scanner system (SkyScan, Kontich, Belgium) and more recently the MicroCAT II (Siemens Pre-Clinical, Knoxville, TN) system. The micro-CT modality provided density based information from x-ray but with a much higher resolution than that achievable with conventional MDCT systems. As the lung nodules were resected and fixed prior to micro-CT imaging the exposure time and radiation dose were not influencing factors in determining protocol settings.

The SkyScan system featured an x-ray source capable of 20-100 kVp and a maximum power of 10W. Settings for voltage and current were restricted such that increasing one would cause a decrease in the other to balance the required power. The final protocol for the isolated fixed lung nodules using the Skyscan 1076 micro-CT scanner involved a 50 kVp, 200 microA, an exposure time of 1770 ms and a binning of 2 to produce isotropic voxels of 18 micron. Three of the eleven collected datasets were imaged using this system and protocol.

The MicroCAT II scanner had a greater flexibility as the current and voltage could be adjusted independently from 0-500 microA and 0-130 kVp, up to a maximum power output of 65 W. With a power level of 8 W or less, the focal spot size at the source is 10 microns. It was desired to identify a protocol which achieved a high spatial resolution while minimizing ring artifact and noise. **Figure 3-8** illustrates the resulting data from four imaging protocols on a wedge resected lung nodule specimen. The image

in **Figure 3-8A** shows the best compromise between maintaining a high spatial resolution, when compared to **Figure 3-8C** and **Figure 3-8D**, and minimizing ring artifact, as occurs in **Figure 3-8B**. Hence, it was experimentally determined for the MicroCAT II system 80 kVp, 100 microA, an exposure time of 4000ms and a binning of 2 to produce an isotropic voxel of 28 micron.

From the micro-CT scanning of the isolated lung nodule, high resolution (isotropic voxels of 18 or 28 micron) density information as well as surface border characteristics for the isolated lung nodule was generated.

3.3.5.3   LIMA Processing

Following previously defined imaging stages, the isolated nodule was sectioned at 250 micron using the LIMA system. The in-plane resolution achievable with the LIMA system was dependent on the degree of magnification chosen and ranged from 15 micron to 0.9 micron. A compromise had to be reached between gaining the most amount of valuable information and the time required to section and image the tissue. It was determined through experimentation with the LIMA system that a magnification of 10x, producing a pixel size of 8.5 micron was appropriate for this study. The images were obtained using white light and indicate the color of the specimen.

Sectioning and imaging of the isolated nodule using the LIMA system resulted in a low magnification, light microscopy dataset featuring color information of the nodule.

3.3.5.4   Histopathology Processing

The tissue sections created by the LIMA were retained for processing via traditional pathology techniques, with staining, to reveal cellular detail. The pathology

data was utilized as confirmation of nodule content and regions. For the non-destructive (micro-CT) and destructive (LIMA and pathology) imaging of the isolated lung nodule the slice thickness between images was very different. The micro-CT dataset was created using a protocol providing a slice thickness of 28 micron while the LIMA datasets had a slice thickness of 250 micron. Every slice of tissue from the LIMA system underwent histological processing, however due to the delicate nature of the tissue sections, not all submitted slices produced usable histology slides.

Each tissue section generated by the LIMA was submitted in plastic cassettes for histological processing. The sections were fixed for at least 60 minutes before processing. Processing of the tissue involved the use of an ethanol-based express tissue processing system. Paraffin wax was used as an embedding medium to form the tissue 'block'. The blocks were sectioned at 4 µm using a conventional microtome, with three sections per tissue block slide mounted. The orientation of the tissue blocks was important as it was desired to collect the sections for mounting from the top of the block to ensure the nodule content captured in the pathology slide would closely match the content in the corresponding LIMA image.

Once the sections were mounted on glass slides they underwent staining. The staining protocol involved the creation of one haematoxylin and eosin (H&E) stained slide and one immunohistochemical stain. H&E is a common and easy to perform staining technique which is used for cancer diagnosis in pathology, while the immunohistochemical stain was experimentally identified as described in CHAPTER 4 below. The unstained slides were collected and retained for the investigation of alternative staining techniques.

The slides produced were digitized using a ScanScope system (Aperio Technology). The light microscopy imaging was performed with a 20x objective which produced pixel sizing of 0.5 micron. As these resulting images were extremely large and contain a much higher resolution than required, bi-cubic resampling was used to reduce the resolution of the images to 2.5 micron.

The histopathological processing of each section produced by the LIMA system resulted in two digitized histopathology datasets (2.5 micron pixels); one traditional H&E stained and one immunohistochemical stain with contrast between tissue types.
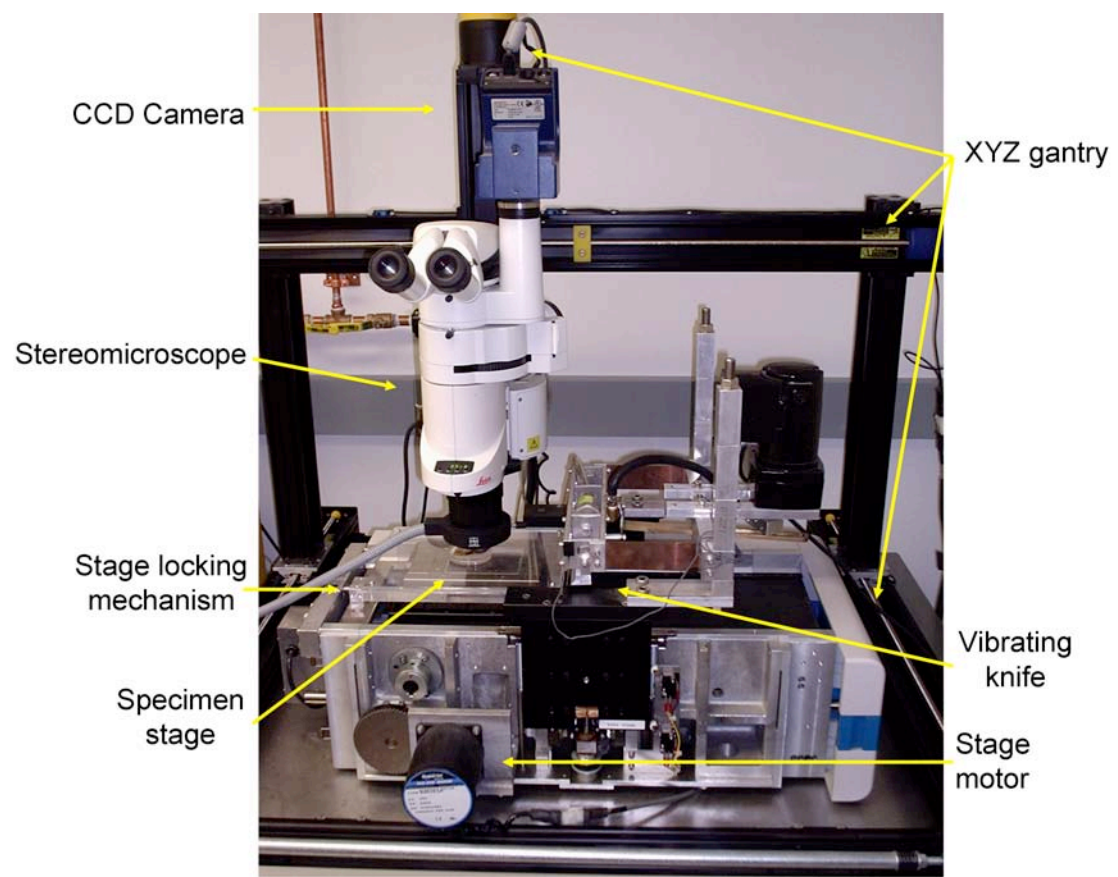
Figure 3-1: The hardware of the large image microscope array (LIMA) system, featuring a unique microtome and a stereoscope coupled with a digital camera.
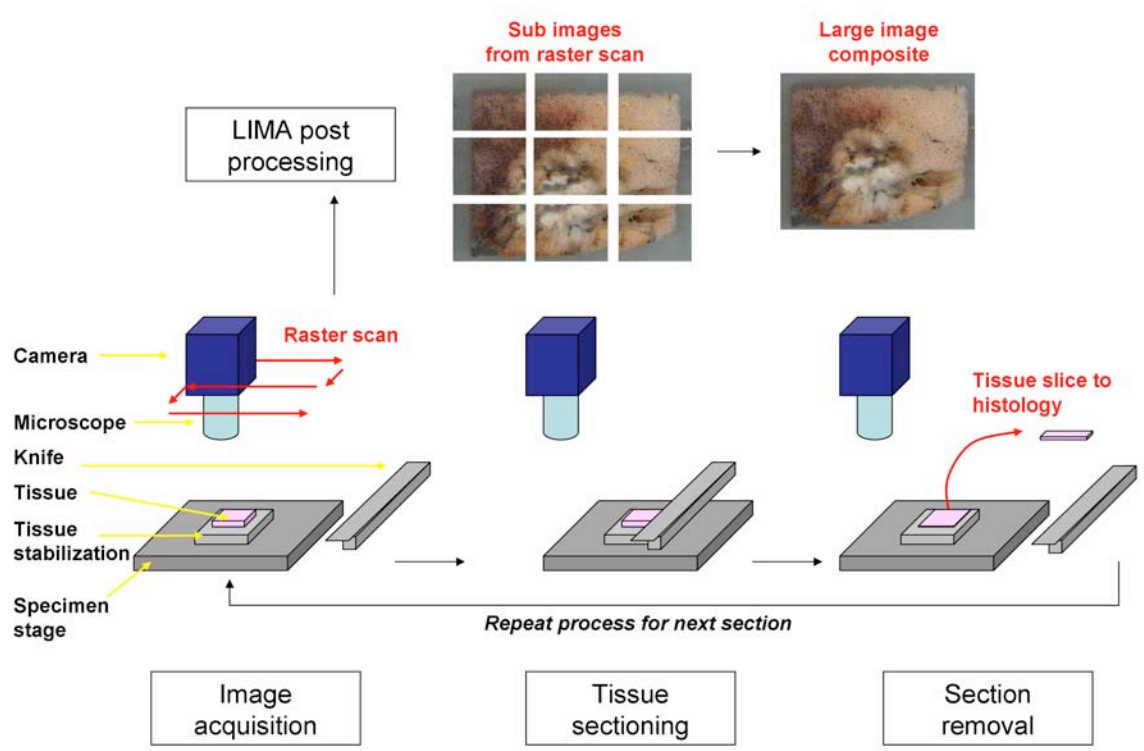
Figure 3-2: Summary of the large image microscope array (LIMA) system sectioning and imaging process. This process results in the generation of a color dataset which can be used as the ground truth for linking non-destructive image sets, such as CT to destructive histopathology image sets.
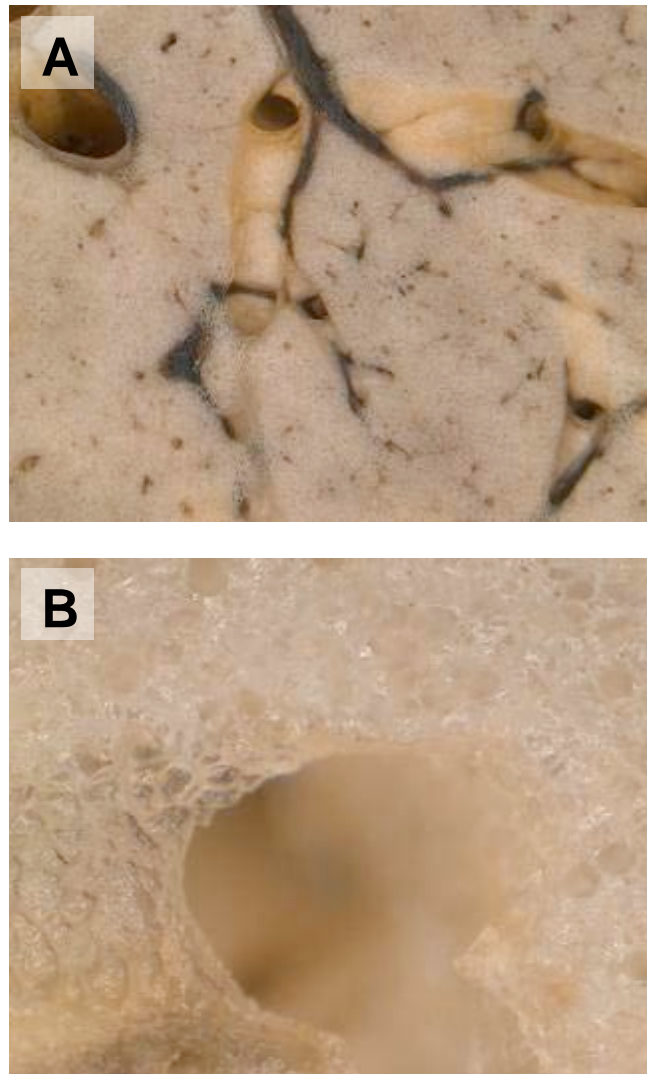
Figure 3-3: Sheep lung imaged using the minimum and maximum magnification of LIMA system. An example of images of fixed sheep lung from the large image microscope array (LIMA) system at original magnifications of 7.11x (A) and a 115x (B).
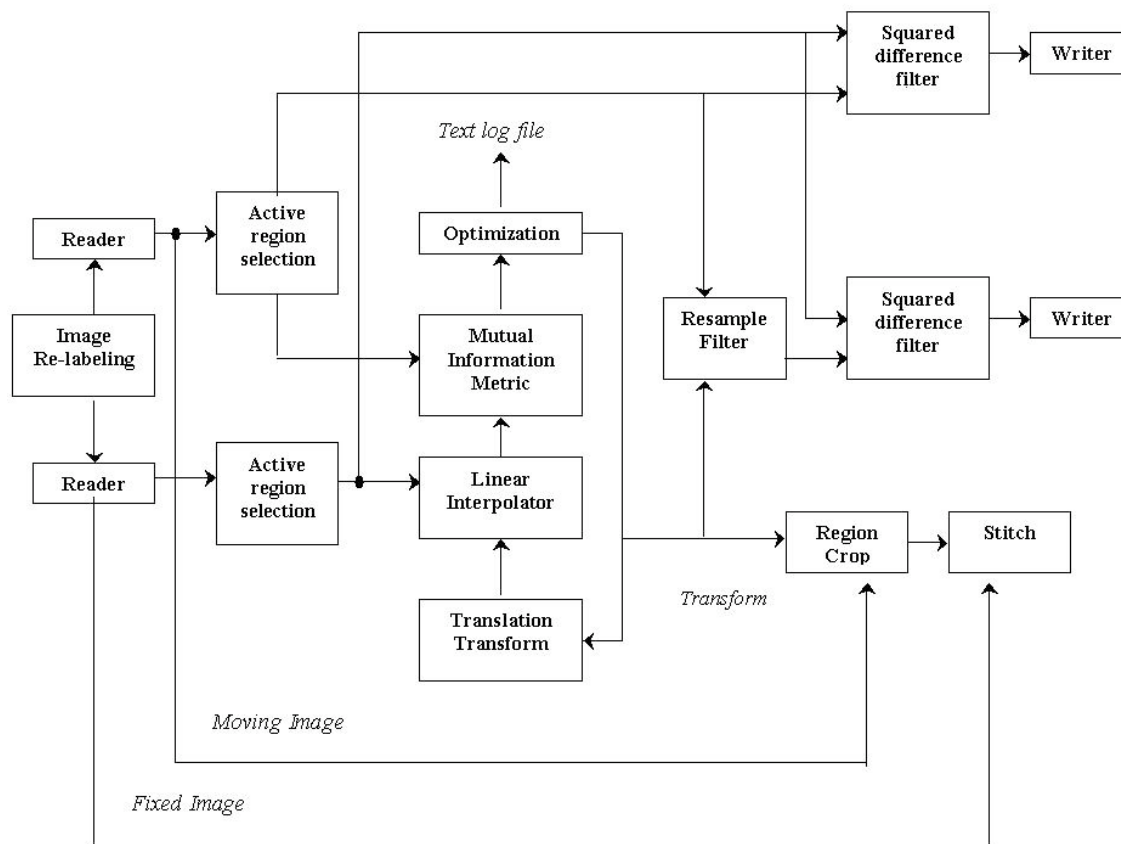
Figure 3-4: Flow chart illustrating the major components of the image mosaic algorithm. A fixed reference image and a moving image are loaded into memory and the overlapping regions common to both images are identified. The overlapping regions are fed into the registration algorithm that consists of a transform, interpolator, metric and optimizer. The resulting stitched image is delivered as an output from the system, along with an image of the squared difference between the overlapping regions before and after registration.
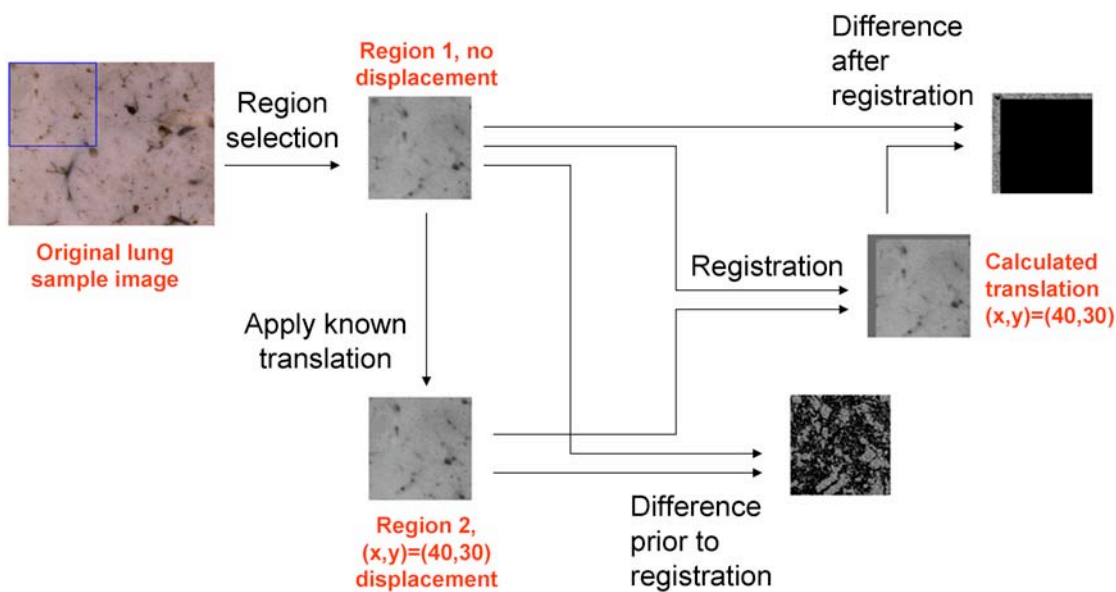
Figure 3-5: Assessment of the registration algorithm. The assessment of the registration algorithm in realigning a translated region from a 7.11x magnified image of lung tissue. The difference in image content after registration illustrates the perfect realignment of the two regions.

**Table 3-1: Summary of characteristics and diagnoses
for study patient group**

Characteristic		mean	range
Age (years)		66	51-79
Body Mass Index (BMI)		26	21-33
Nodule Size from Radiology Report			
	Max. diameter	27	14-50
	Min. diameter	19	10-15
		n	**%**
Gender			
	Male	2	18
	Female	9	82
Race			
	Caucasian	10	91
	African American	1	9
Site			
	Right Upper Lobe	1	9
	Right Middle Lobe	0	0
	Right Lower Lobe	6	55
	Left Upper Lobe	3	27
	Left Lower Lobe	1	9
Diagnosis			
	Adenocarcinoma	7	64
	Squamous Cell Carcinoma	3	27
	Neuroendocrine Carcinoma	1	9
Stage			
	Stage IA	5	46
	Stage IB	3	27
	Stage IIA	0	0
	Stage IIB	2	18
	Stage IIIA	1	9
	Stage IIIB	0	0
	Stage IV	0	0
Smoking History			
	Never	1	9
	<1 pack per week	2	18
	1 pack per week	6	55
	2 pack per week	1	9
	3 pack per week	1	9
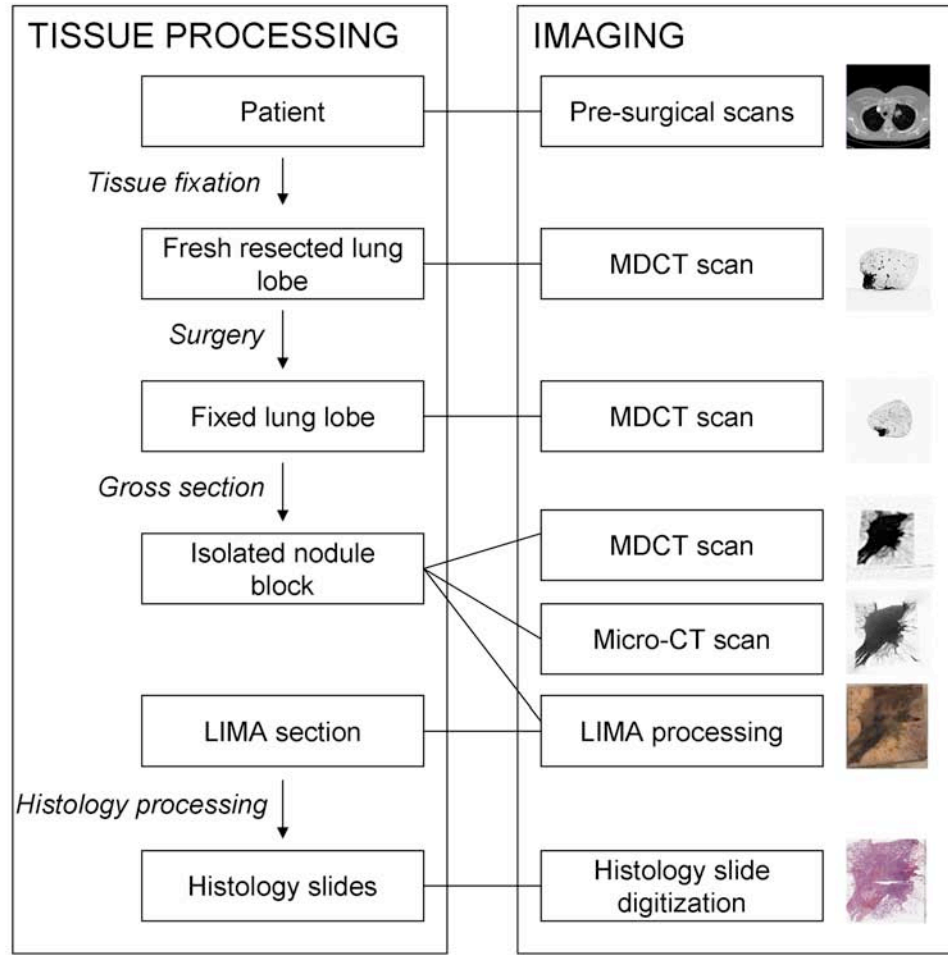
Figure 3-6: A summary diagram outlining the image acquisition and tissue processing procedure.

Table 3-2: Timeline of the processes required to acquire the multi-modal data for one complete lung nodule dataset.

Process	Time
Experimentation preparation	
Gather required materials	10 min
Measuring and mixing of fixative solution	30 min
Set up of fixation station	30 min
Patient	
Identification	10 min
Consent	1 hr
Download and save presurgical data	2 hr
Fresh Lobe	
Surgical pathology preliminary assessment	30 min
MDCT of fresh lobe at inflation	1 hr
Modified Heitzman fixation; in solution	52 hr
Modified Heitzman fixation; in dryer	90 hr
Clean up of fixation station	30 min
Fixed Lobe	
MDCT of fixed lobe	1 hr
MDCT guided lung nodule gross section	1 hr
Isolated Nodule	
Adhere nodule to bracket with glue	3 hr
MDCT of isolated nodule	1 hr
Micro-CT of isolated nodule (with recon)	5 hr
Embed nodule in agar and set	8 hr
Calibration and set up of LIMA system	1 hr
LIMA sectioning and imaging	6 - 8 hr
Post processing LIMA images	2 hr
LIMA sections	
Place sections in histology cassettes	30 min
Ethanol based express processing	1 hr
Paraffin embedding/sectioning/mounting	24 hr
Slide drying	12 hr
Histology Slides	
H&E staining	2 hr
Immunohistochemical staining	3 hr
Slide digitization	7 hr
All image files	
Transfer and backup of datasets	8 hr
TOTAL TIME FOR SINGLE DATASET	236 hr
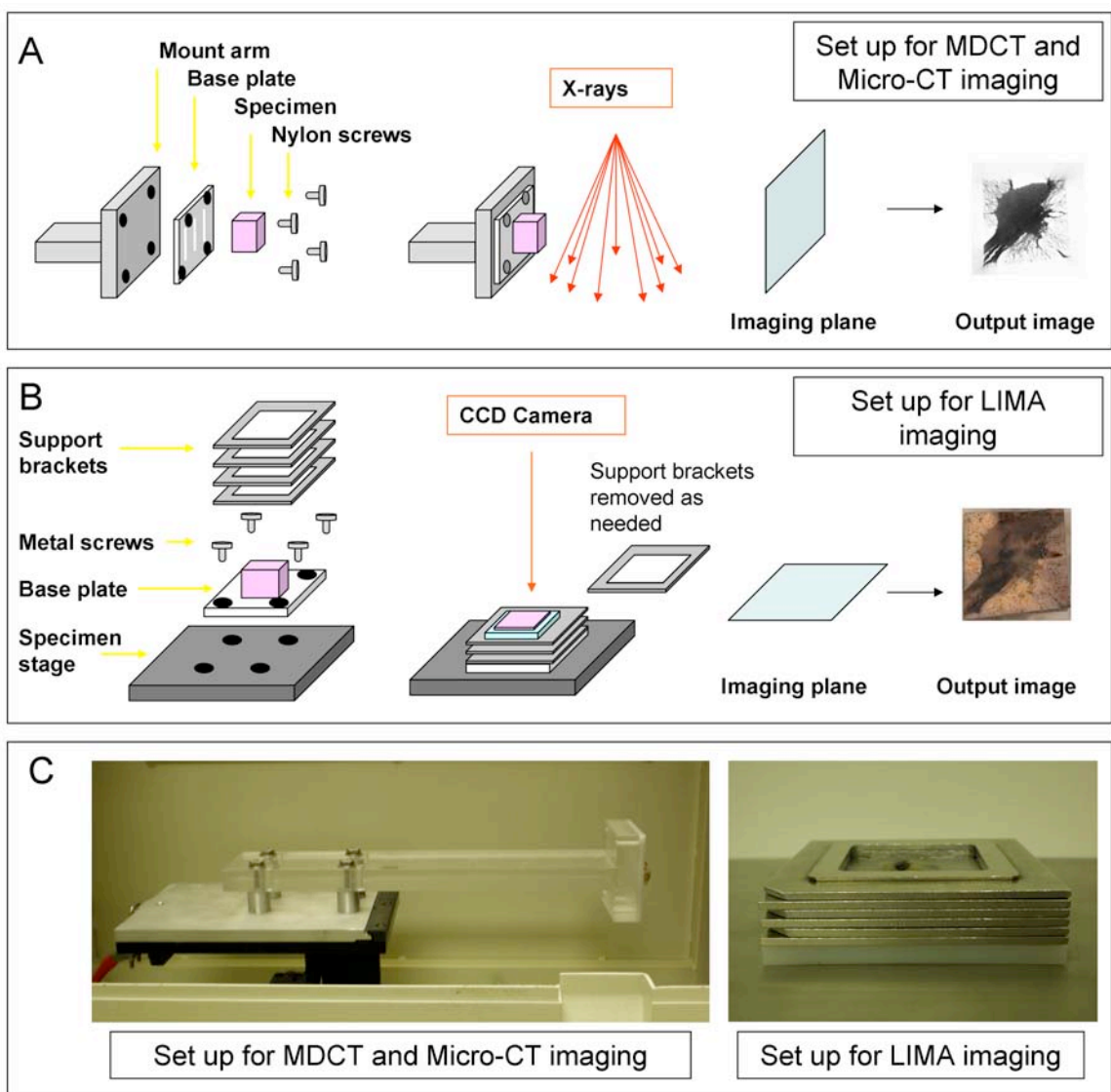
Figure 3-7: Schematic diagram of the nodule stabalization system developed for maintaining the alignment between the imaging planes from MDCT. Micro-CT and the large image microscope array (LIMA) system. Shown are the components and imaging plane of the MDCT and micro-CT modalities (A) and the LIMA system (B). Photographs of the actual hardware is shown in (C) which show the detail of the mount arm and the support brackets.

Voltage: 80 kVp, Current: 100 µA, Power: 8 W

Voltage: 40 kVp, Current: 200 µA, Power: 8 W

Voltage: 60 kVp, Current: 200 µA, Power: 12 W

Voltage: 80 kVp, Current: 500 µA, Power: 40 W
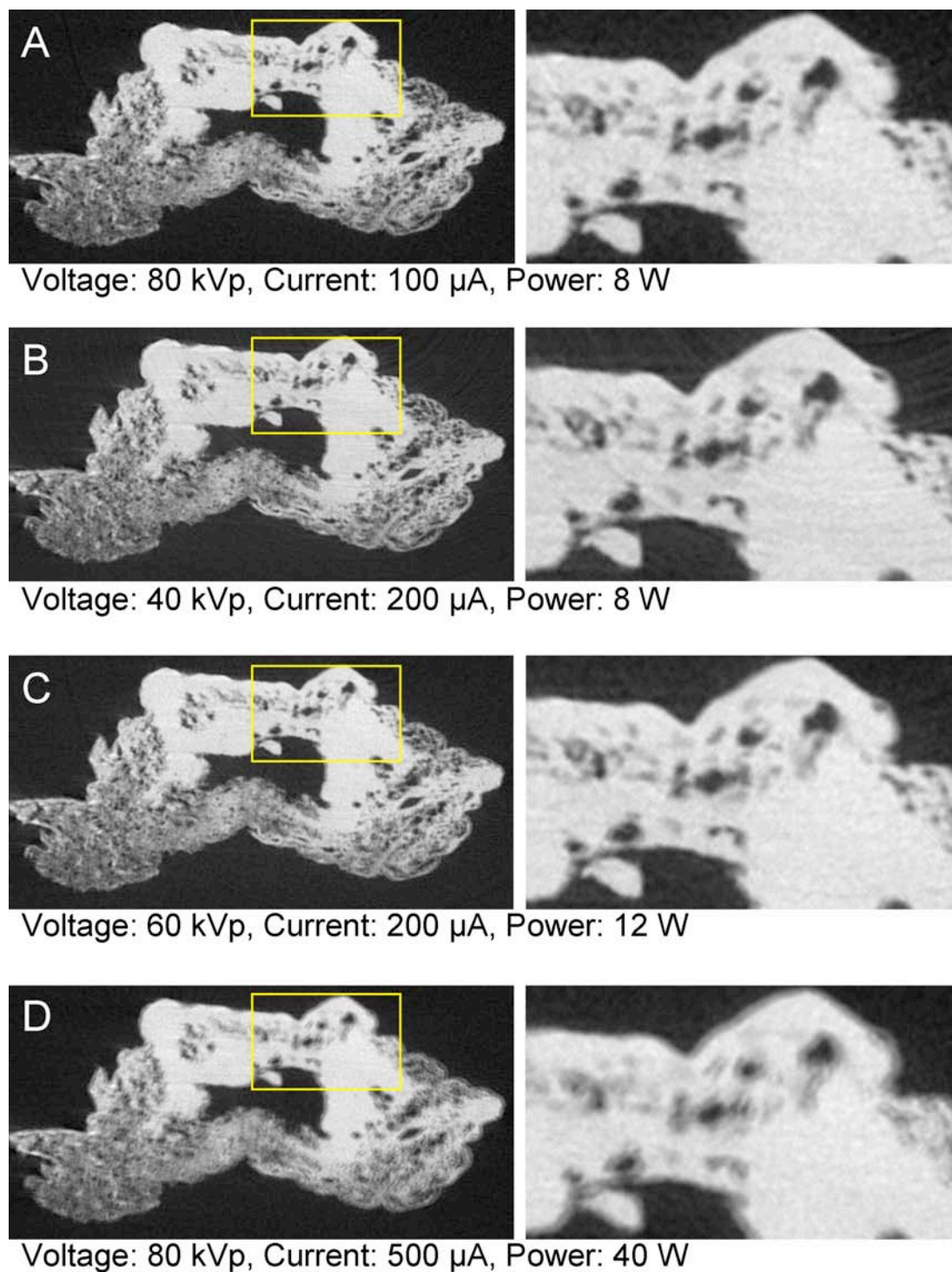
Figure 3-8: Comparison between micro-CT imaging protocols using a wedge resected lung nodule. The protocol deemed best suited to the study (A) had an optimal power of 8 W and resulted in less ring artifact and blurring than the other protocols (B), (C) and (D).

Table 3-3: Comparison between the specifications of the datasets, including resolution and slice separation.

	Modality	*Destructive imaging*	*Pixel size (micron)*	*Slice thickness (micron)*	*Slice separation (micron)*
	In vivo CT	No	550 to 1000	600	2500 to 5000
	Resected Lobe MDCT	No	300	600	300
	Isolated nodule MDCT	No	100	600	300
	Isolated nodule micro-CT	No	18 or 28	18 or 28	18 or 28
	LIMA	Yes	8.5	250	250
	Histology	Yes	2.5	4	250 to750

CHAPTER 4

HISTOPATHOLOGICAL CLASSIFICATION

Developing a method for the controlled, volumetric evaluation of lung nodule architecture is advantageous to the clinical and research environment. Recent studies have found a valuable link between the proportion of tissue types, such as fibrosis or necrosis with in a lung nodule and prognosis [59, 60, 82]. Maeshima et al. found the fibrotic proportion of nodule content was an independent prognostic factor, with direct correlation between increased fibrotic component and decreased survival. However, studies such as this have relied on the qualitative evaluation of histopathological sections to score or grade the appearance of a particular tissue type which provided a limited understanding of the nodules complete composition.

In order to gain a comprehensive understanding of the content and architecture of the tissue type distributions throughout a nodule biomass, the cellular content of the nodule needed to be identified and segmented. The histopathology datasets provided information content on the cellular content of the nodule. Based on these histophatological datasets, a manual and an automated segmentation approach were developed for the purpose of segmenting out the tissue types present.

4.1 Manual Segmentation

The most widely used staining approach for the clinical diagnosis of lung cancer is hematoxylin and eosin (H&E). H&E staining results in the blue coloration of nuclei, pink staining of cytoplasm structures and the intense red coloring of blood cells. This

broad stain is effective in highlighting the cellular structures such that a pathologist is able to distinguish cells of differing types.

A surgical pathologist with a subspecialty expertise in pulmonary pathology analyzed the digitized H&E histological images. An Intuos graphics tablet and pen (Wacom) were used to generate mask images corresponding to the different tissues identified. An example of a tissue type mask generated by the manual segmentation of a H&E section of lung adenocarcinoma tumor is shown in **Figure 4-1**. The cellular based tissue types identified and segmented included; solid regions of cancerous tumor cells, cancerous tumor cells in a borncioalveolar carcinoma (BAC) configuration, necrotic cells, active fibroblastic stromal tissue, inactive (hyaline) fibrosis , red blood cells and normal tissue.

For the manual tracing of these tissue classes, a set of definitions was established to promote consistency across the datasets.

- Cancer (solid): solid groupings of cancer cells
- Cancer (BAC): cancer cells following a bronchio-alveolar carcinoma invasion pattern, which is a non-solid, alveolar pattern.
- Necrotic tumor: dead cells of any origin, including cancerous
- Active fibrosis: majority of fibroblasts (greater than 50% of cells fibroblasts)
- Inactive fibrosis: fibroblasts intermixed with collagen (less than 50% of cells fibroblasts)
- Blood: Groupings of erythrocytes (red blood cells)

The task of manually tracing histology sections was extremely time consuming, with each section taking well over an hour to complete. This was a costly process as a pathologist with expertise in pulmonary pathology was required to conduct these tracings due to the complexity involved in defining the tissue boundaries. Due to the human observer component there may be some inconsistency in the definition of tissue type boundaries and there also exists a trade off between the level of detail and the time required to complete the traced maps. These human observer limitations were particularly true for the case of H&E stained histology as the boundary between one tissue type and another was not clearly highlighted, **Figure 4-2**.

4.2 Automated Segmentation

4.2.1 Immunohistochemical Staining for Increased Contrast

Immunohistochemistry has been used in research studies and in the clinical pathology environment to identify lung cancer. Cytokeratins are found in the cytoskeleton of epithelial cells and are hence effective in identifying epithelial derived carcinoma. Furthermore, there are twenty different forms of human cytokeratin which are found in normal and malignant epithelial cell lines, different combinations of which can be used to characterize poorly differentiated carcinoma [83]. Hence cytokeratin was chosen as the immunohistochemical target for highlighting cancerous tumor cells. It should be noted that cytokeratin immunohistochemistry does not selectively target neoplastic epithelial cells (versus normal epithelial cells), however, in this study cohort it is expected that normal, non-neoplastic cells make up an insignificant portion of the biomass. The primary purpose of this immunohistochemical staining approach was to

provide some contrast in the appearance of cancerous tumor cells and all other types present within the nodule. To ensure a positive immunohistochemical staining of all the cancerous tumor cells in any NSCLC nodule, a number of cytokeratin weights were targeted. A pan-cytokeratin cocktail was used consisting of Cytokeratin 7 (1:50, Dako Corporation, Carpinteria, CA), CK 8/18 (1:200, Abcam, Cambridge, MA) and AE 1/3 (1:200, Chemicon Int, Tenecula, CA). To produce the most consistent and dependable immunohistochemical staining possible an automated immunostaining system (Dako Corporation, Carpinteria, CA) was utilized.

A number of counterstains were explored to examine the maximal achievable contrast between the tissue types of interest. Hematoxylin and variations of the Masson's Trichrome were tested for suitability. Hematoxylin is a simple nuclear stain and hence highlighted the nuclei of all cell types present. Masson's Trichrome is a more complicated staining process involving multiple dyes for the differentiation of muscle, collagen, fibrin and erythrocytes. The reagents for this staining technique include; Bouin's fixative, Biebrich Scarlet, Weigert's Iron Hematoxylin, Phosphotungstic-Phosphomolybdic acid solution and Aniline Blue and when applied sequentially the resulting stain colors nuclei - black, cytoplasm, erythrocytes and muscle - red and collagen - blue. Two modifications of the traditional Masson's Trichrome counterstain were explored. The first (Mod 1) involved the exclusion of the Bouin's fixative step and the second (Mod 2) included the Bouin's fixative but excluded the Weigert's Iron Hematoxylin. Of these staining combinations, the pan-cytokeratin cocktail counterstained with the Masson's Trichrome (Mod 2) counterstain was found to produce the greatest contrast between the tissue types of interest, **Figure 4-3**. Of particular

advantage was the increased contrast between areas of inactive and active fibrosis with dense collagen fibers stained blue and predominantly elastic type matrix stained pink, **Figure 4-4**. To increase the consistency of the counterstaining process an automated slide stainer, DRS-601 (Sakura Finetek, CA, USA) was used.

4.2.2   Algorithm Design

The immunohistochemical staining approach was developed to increase the contrast between different tissue types with in the nodule. The chosen, pan-cytokeratin cocktail with Masson's Trichrome (Mod 2) counterstain produced the greatest color contrast. Automated staining equipment was utilized for the application of the pan-cytokeratin cocktail with modified Masson's Trichrome counterstain in an effort to minimize the variation in stain intensity between sections and also between nodule datasets. While these efforts proved successful in minimizing the stain variation between sections of a single dataset, the staining across datasets varied greatly. This proved to be a significant challenge in the development of an automated segmentation approach as the variation in the staining intensity across the datasets was larger than the separation of the color based feature set.

The developed algorithm used the *Lab* and *HSV* color spaces as features for the segmentation of the tissue types. Both these color systems differ from the *RGB* color space in that they were developed to more closely represent the human perception of color. The *a* and *b* axes of the *Lab* system mark the variation from red to green and from yellow to blue respectively, while the third channel (*L*) reflects the luminance. In the *HSV* color model the three channels represent hue, saturation and value (intensity). In the case of features for the classification of immunohistochemically stained samples the *Lab*

color space is advantageous in that the luminance can be excluded and color is explained by orthogonal axes. The *HSV* model is preferable when a single value is desired to represent the hue in the image. **Figure 4-5** shows an example of representation of the colors in the developed immunohistochemical stain approach, in the *RGB*, *Lab* and *HSV* color space.

The large variance in the staining of the different nodule cases meant supervised classification approaches were not appropriate for the segmentation of the complete, multi-nodule, immunohistochemical dataset. A k-means clustering approach was chosen as a suitable unsupervised algorithm capable of accommodating the variations in staining across the nodule datasets [22, 84, 85]. The k-means clustering algorithm attempts to optimally partitioned data into a set number of natural groups, *k* [84]. In general, this is achieved by initializing *k* centroids, *c*. Each point in the data is assigned to a group based on it's proximity to that groups centroid. Once all points are assigned the updated centroids for each group are calculated and the process is repeated until there is no change in the location of the centroids – indicating a stable partitioning of the data into *k* groups. The first step in the developed segmentation approach, which is summarized in Figure 4-6, was to use k-means clustering to segment the data into regions, using the *a* and *b* channels from the *Lab* color space. The Euclidean distance measure was used to determine the closest centroid for each point. To avoid the occurrence of partitions at local minima, the clustering was repeated five times. At each repetition, the total distance from all points to their centroids was calculated and the partition result with the lowest total distance was chosen.

The number of groups, k, was initialized by the user but was restricted to either three or four groups. All complete histopathological slides contained at least three tissue classes along with background data ($k = 4$). However, not all the small sample images used for validation contained background pixels ($k = 3$). All the histopathological slides in the dataset contained cancerous tumor, inactive fibrosis and active fibrosis. The color separation of these tissue types, based on the devised staining approach, was the greatest and hence directly correlated to the partitioning of the cluster algorithm. However, the labeling of the regions output from the k-means clustering algorithm was randomly assigned and bore no reference to the properties of the feature set. Hence, a hue based labeling step was created to assign labels based on the mean hue of the regions.

The mean hue for each region identified by the k-means clustering algorithm was calculated and these values sorted in descending order. The region with the highest mean hue was assigned a 'mixed' class label with a pixel value of 1. 'Active fibrosis' and 'inactive fibrosis' class labels were assigned to the following ranked hue values and identified by pixel values of 2 and 3 respectively. Finally, if a k of four was selected, the lowest ranked region based on mean hue, was assigned to 'background', with a pixel label of 4.

The accurate labeling of the cancerous tumor tissue was deemed the highest priority for the automated segmentation technique, however, based on the k-means clustering approach the blood and cancerous tumor pixels were grouped together. A second pass labeling approach was incorporated to further classify the mixed class into cancerous tumor and blood. A binary mask image was created containing only pixels with a 'mixed' tissue type label. This binary mask contained many sub-regions, most of

which corresponded to cancerous tumor and a few of which corresponded to red blood cells. Connected component labeling was used to assign new labels to each of these individual regions. For each new region the mode of the *a* channel (from *Lab* color space) and the area was calculated. Regions with a mode *a* value greater than 0.65 and an area greater than five pixels were assigned as 'red blood cells' and given a pixel value of 5 in the original label image. The threshold of 0.65 for the mode *a* value separating the cancerous tumor and blood classes was determined by locating the average minima of the mode *a* histogram from a number of sample images. Pseudocode summarizing this automated segmentation and labeling approach is in **Table 4-1**.

Mode filtering with a two by two neighborhood was applied to the final labeled image so that single pixel regions were removed. Some degree of smoothing of the image was desired as many single pixel regions would disrupt the further analysis of the dataset. For illustrative purposes, each label in the resulting segmented dataset was assigned the same color as used for the manually segmented result, making the distinction between classes clearer to view.

4.2.3 Results

The developed algorithm effectively overcame the challenge of accommodating different staining intensities across datasets and was able to successfully segment cancerous tumor, inactive fibrosis, active fibrosis and blood using the purpose developed immunohistochemical staining approach. The approach was time efficient, delivering a segmentation result for a complete histopathology section in less than one minute. **Figure 4-7** illustrates the automated segmentation result for two sample slices, from two different nodule cases.

The algorithm was limited to identifying a fixed number of tissue types. Normal and necrotic tissue could not be separated from the fibrosis group based on the color data in the developed histochemical staining approach.

For validation purposes, a testing set was created. Twenty five test sample images were randomly selected throughout the immunohistochemical dataset. Two surgical pathologists manually traced the tissue types present in the test samples using the Intuos graphics tablet and pen. The test sample images were limited to 200 by 200 pixels so that a high level of detail could be obtained through manual tracing, in a relatively short period of time. **Figure 4-8** illustrates some example cases from the testing set.

Confusion matrices were generated for the evaluation and comparison of the observer performances and the automated generated segmentation. Three matrices were generated corresponding to each image in the testing set, comparing the classification from observer 1 to observer 2, observer 1 to the automated result and observer 2 and the automated result. Summary confusion matrices, showing the percentage classifications across all the test sample images are show in **Table 4-2** and **Table 4-3**. The comparison of the two observer tracings to each other revealed an accuracy of 73 % between the pixel classifications across all tissues. An accuracy of 72 % was found between the pixel classifications of observer 1 and the automated result and 69 % between observer 2 and the automated result.

From the confusion matrices for each test image, the sensitivity and specificity for the classification of each tissue class was calculated. Comparable sensitivity and specificity results were obtained using the automated segmentation technique when

compared to the manual tracings. The average sensitivity and specificity, with standard error, of the segmentation result for each tissue type class are graphed in **Figure 4-9**.

The repeatability of the automated segmentation approach was also tested by repeatedly running the algorithm, six times over the testing set and comparing the confusion matricies. A standard error of 0 ± 0 pixels was found for all tissue classes, for all test images indicating 100% repeatability of the algorithm.
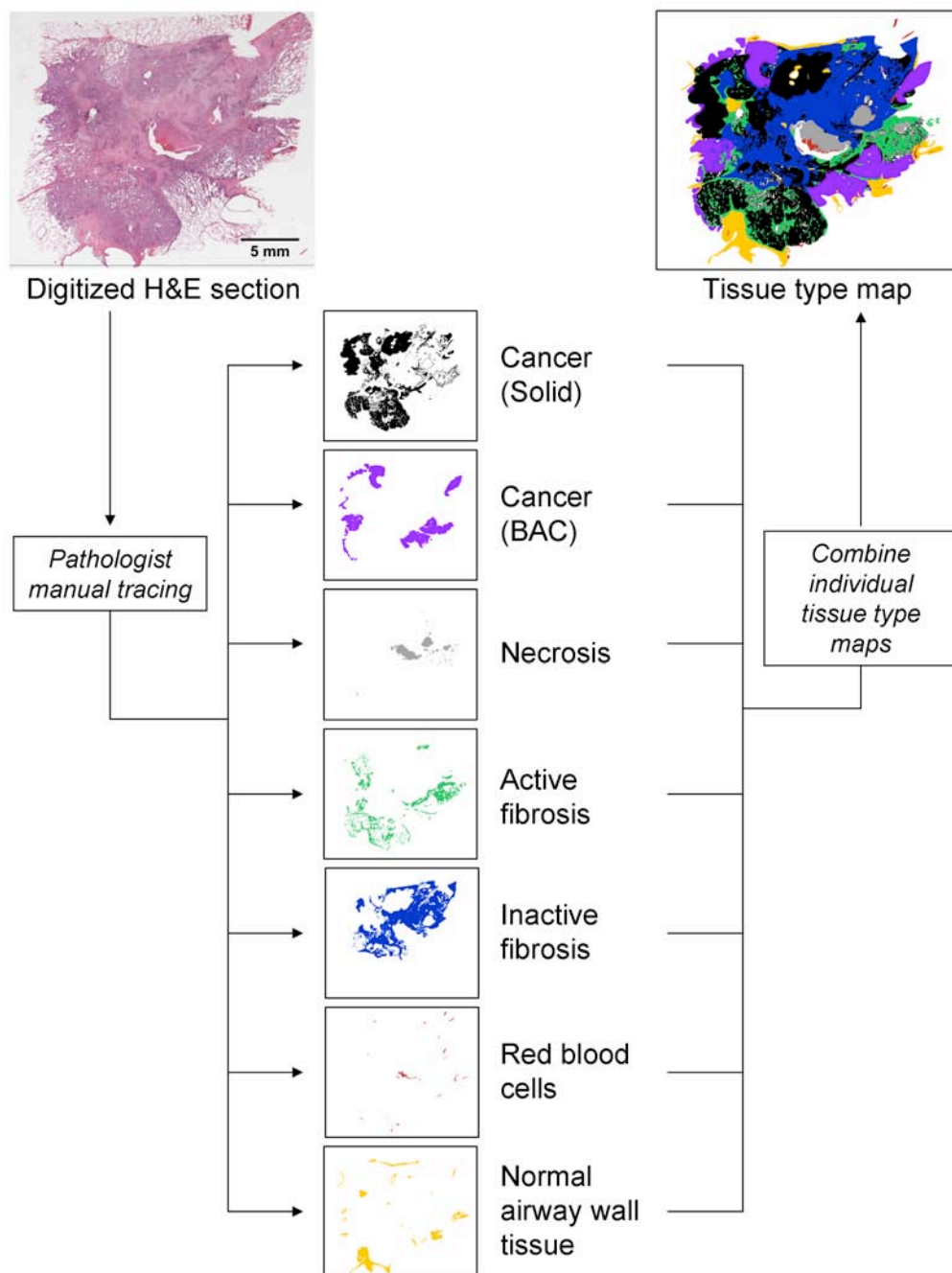
Figure 4-1: Digitized H&E sections are manually segmented by a pathologist with a pulmonary subspecialty. The result of the segmentation is a set of individual tissue type maps (binary images) which can be combined into a single representative image.
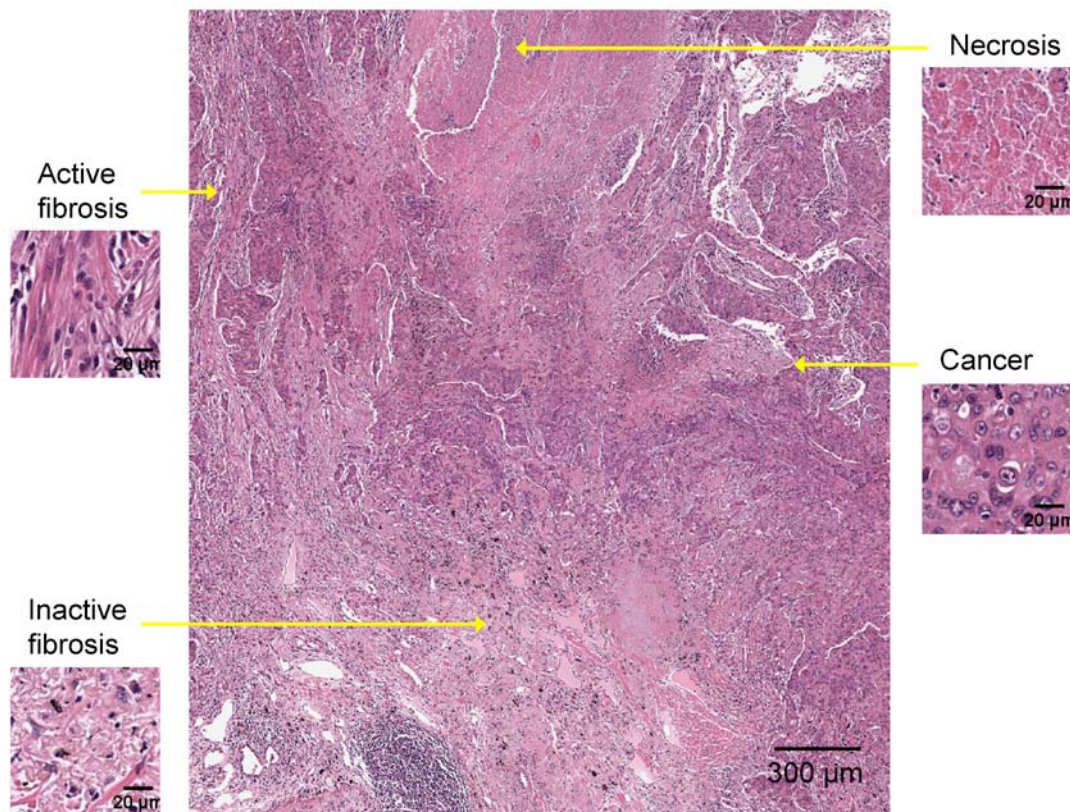
Figure 4-2: H&E histology image of a lung nodule showing the appearance of cancerous tumor cells, necrotic tumor, active fibrosis and inactive fibrosis in this dataset. (Original magnification 20x)
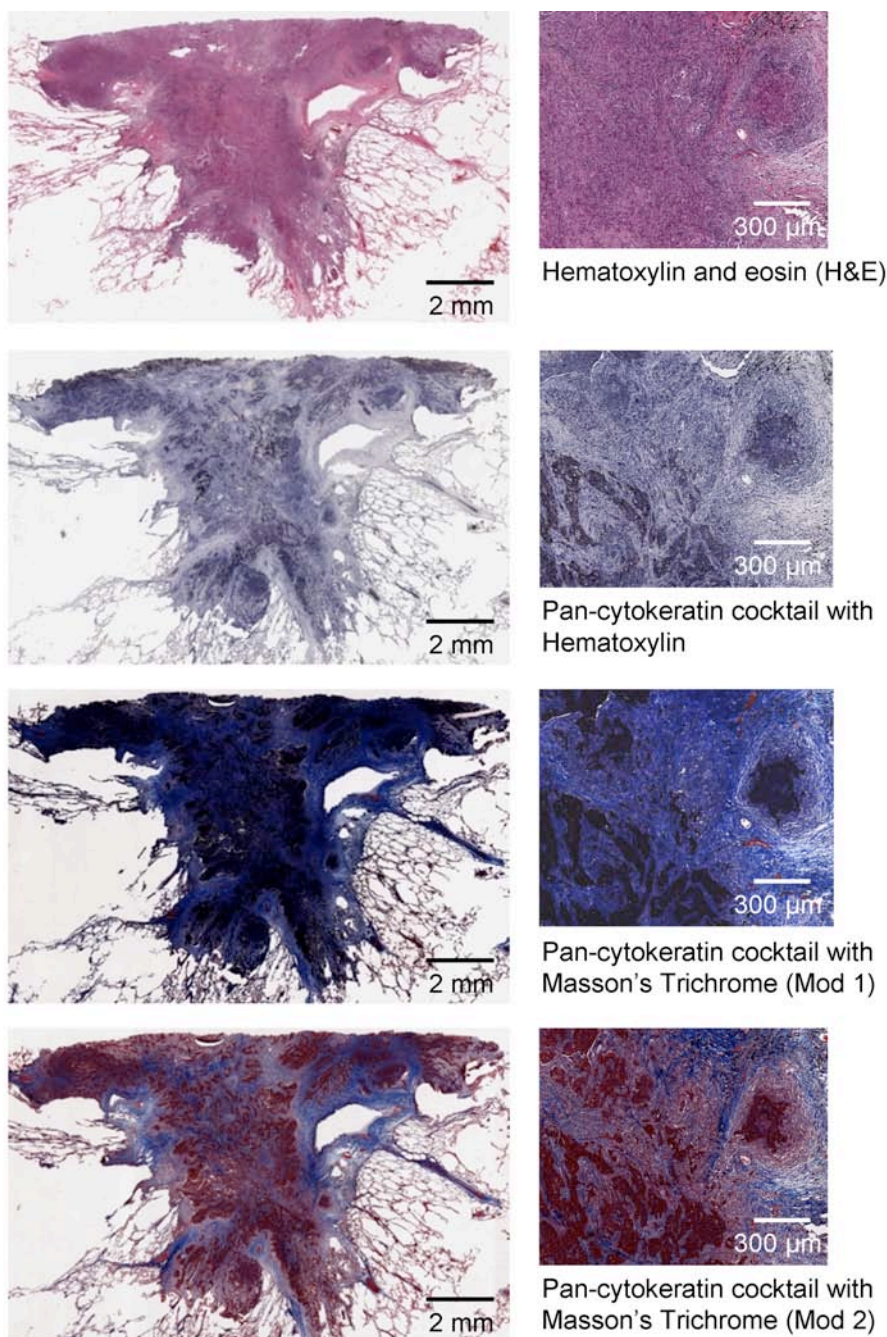
Figure 4-3: Comparison of different staining approaches for the identification of tissue types with in lung nodules. The immunohistochemical pan-cytokeratin cocktail highlights the cancerous tumor portion while the various counterstains add additional contrast. The first modification of the Masson's trichrome counterstain (Mod 1) involved the exclusion of the Bouin's fixative resulting in a predominantly blue counterstain. The optimal contrast between the tissue types was obtained using the Masson's trichrome counterstain without the Weigert's iron hematoxylin (Mod 2) (Original magnification, x20)
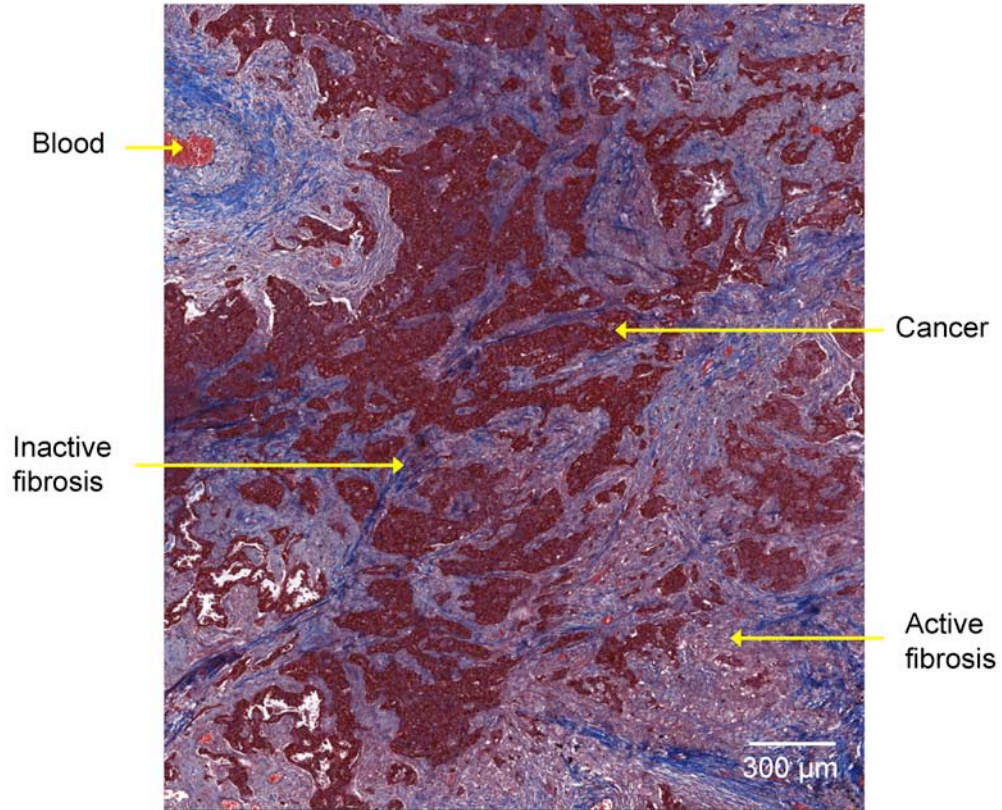
Figure 4-4: The pan-cytokeratin cocktail counterstained with Masson's Trichrome (Mod 2) was found to produce the optimal contrast between tissue type regions. The staining approach resulted in tumor cells stained maroon, inactive fibrotic regions stained blue, active fibrosis stained pink and red blood cells (erythrocytes) red. (Original magnification, x20)
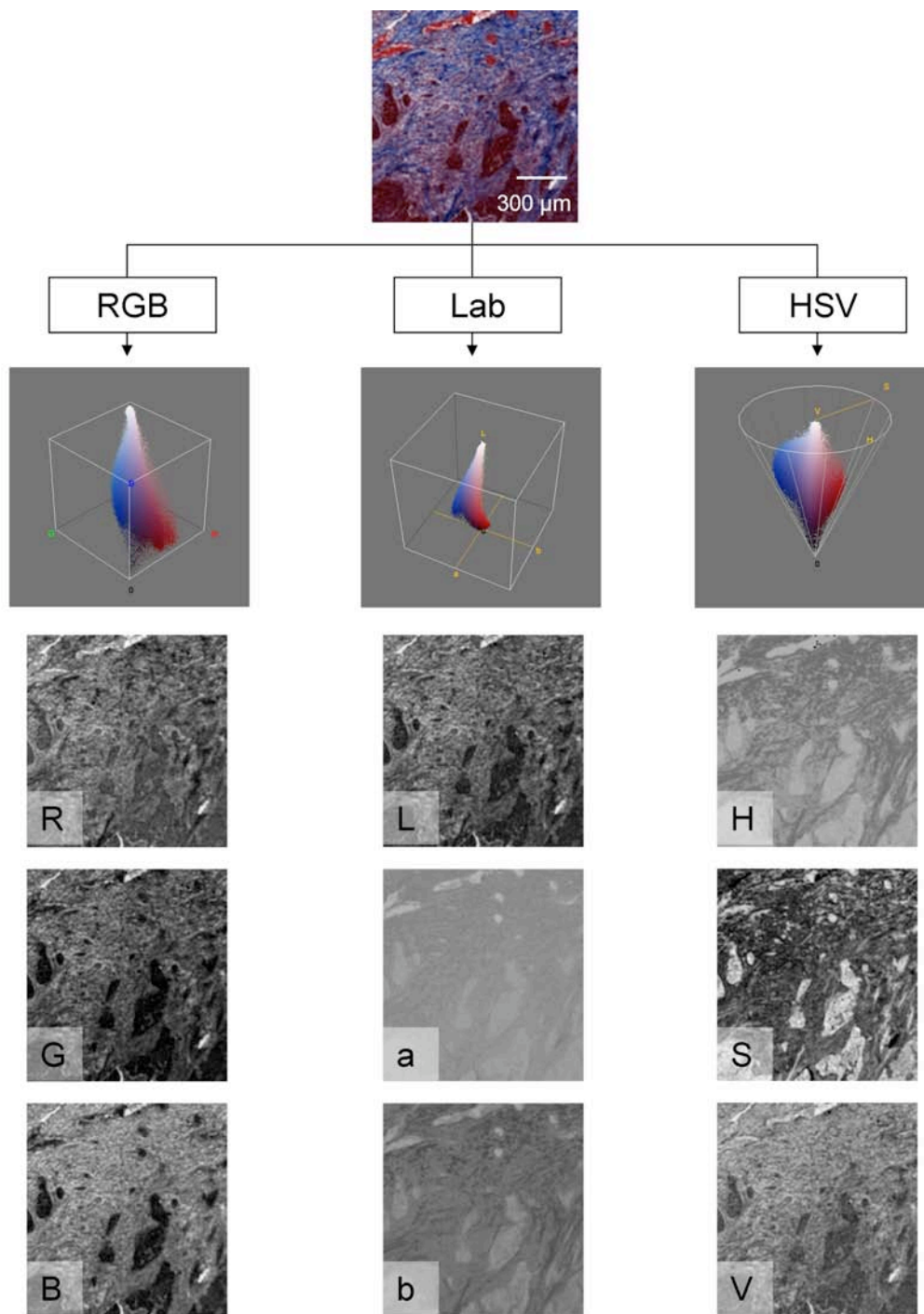
Figure 4-5: Comparison of three color spaces. Illustration of the pixel colors in a nodule sample, stained with the pan-cytokeratin with Masson's Trichrome (Mod 2) counterstain, represented in the RGB, Lab and HSV color spaces.
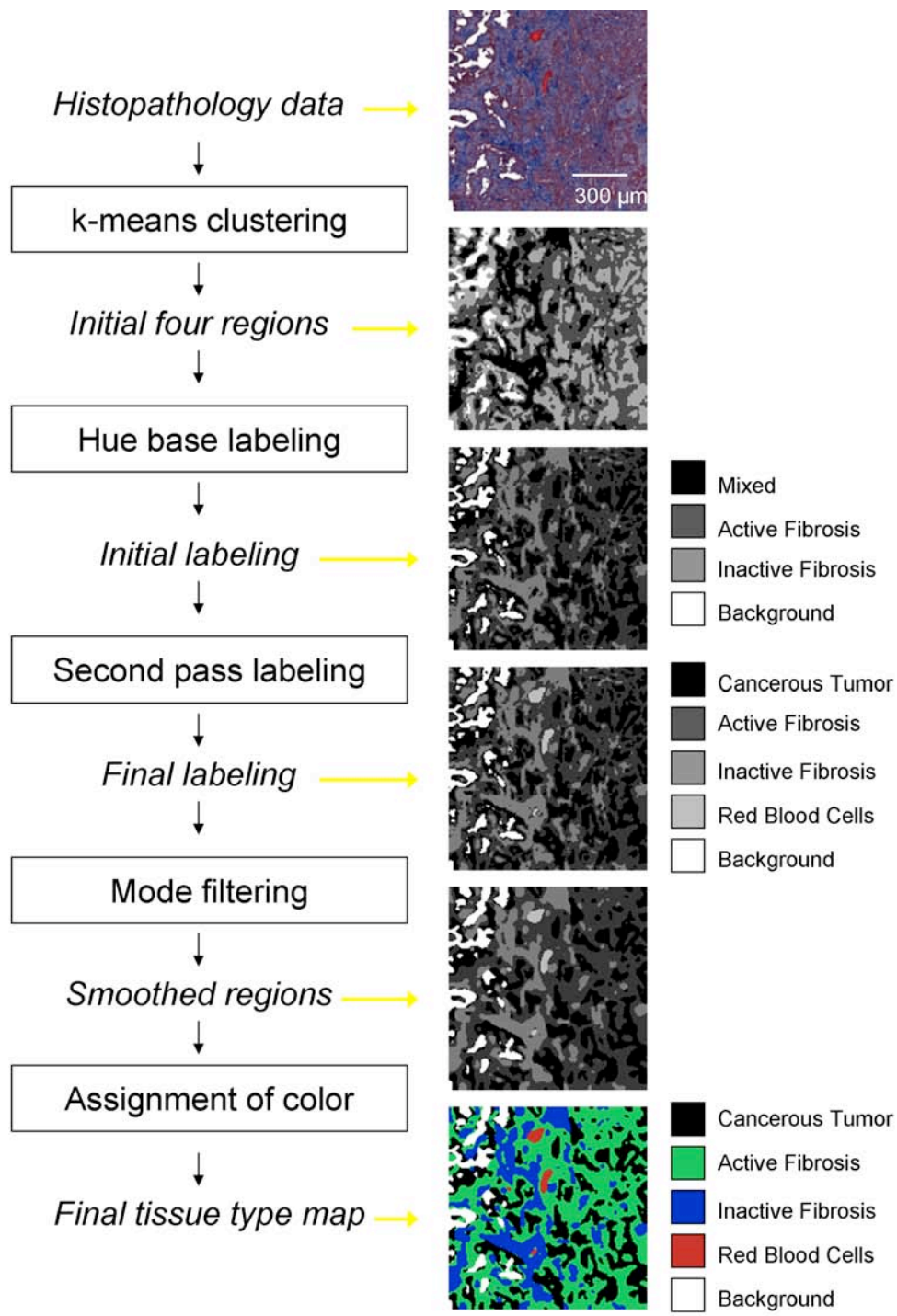
Figure 4-6: Summary diagram of the automated histopathology segmentation approach.

Table 4-1: Pseudocode for the automated segmentation approach.

Automated Segmentation Algorithm

INPUT: Image, *Im*, number of clusters, *k*, replication of k-mean calculation, *Rkm*

Convert *Im* to **Lab** color space
Create feature array using *a* and *b* values for each pixel in the image
FOR 1 to *Rkm*
    Randomly initialize *k* centroids, $c_1, c_2...c_k$
    UNTIL there are no changes in the centroids
            Group each element in the feature array to the closest centroid
            FOR 1 to *k*
                    Recalculate centroids based on all group members
    Calculated total distance between all elements and their group centroids
Find and use iteration which had minimum total distance

// Hue based labeling
Convert *Im* to **HSV** color space
FOR 1 to *k*
    Calculate mean *H* value for region
Sort the mean *H* values and label in order (1 to 3)
IF *k >3*
    Assign fourth ranked– 4

//Second pass labeling
Create binary image using pixels labeled 1
FOR each row and each column in the image
    IF pixel is not background
            Find the neighboring pixels
            IF no labeled neighbors
                    Assign pixel a new label
            ELSE Assign pixel lowest label of neighbors and store equivalence
    Resolve equivalence classes
FOR each row and column in the image
    IF pixel is not background
            Re-label pixel based on equivalence (1 to *r*)
FOR I to *r*
    Calculate the mode *a* value (from **Lab** color space) for region
    Calculate the area of region
    IF mode of *a* < 0.65 and region > 5 pixels
            Assign tissue type label 5
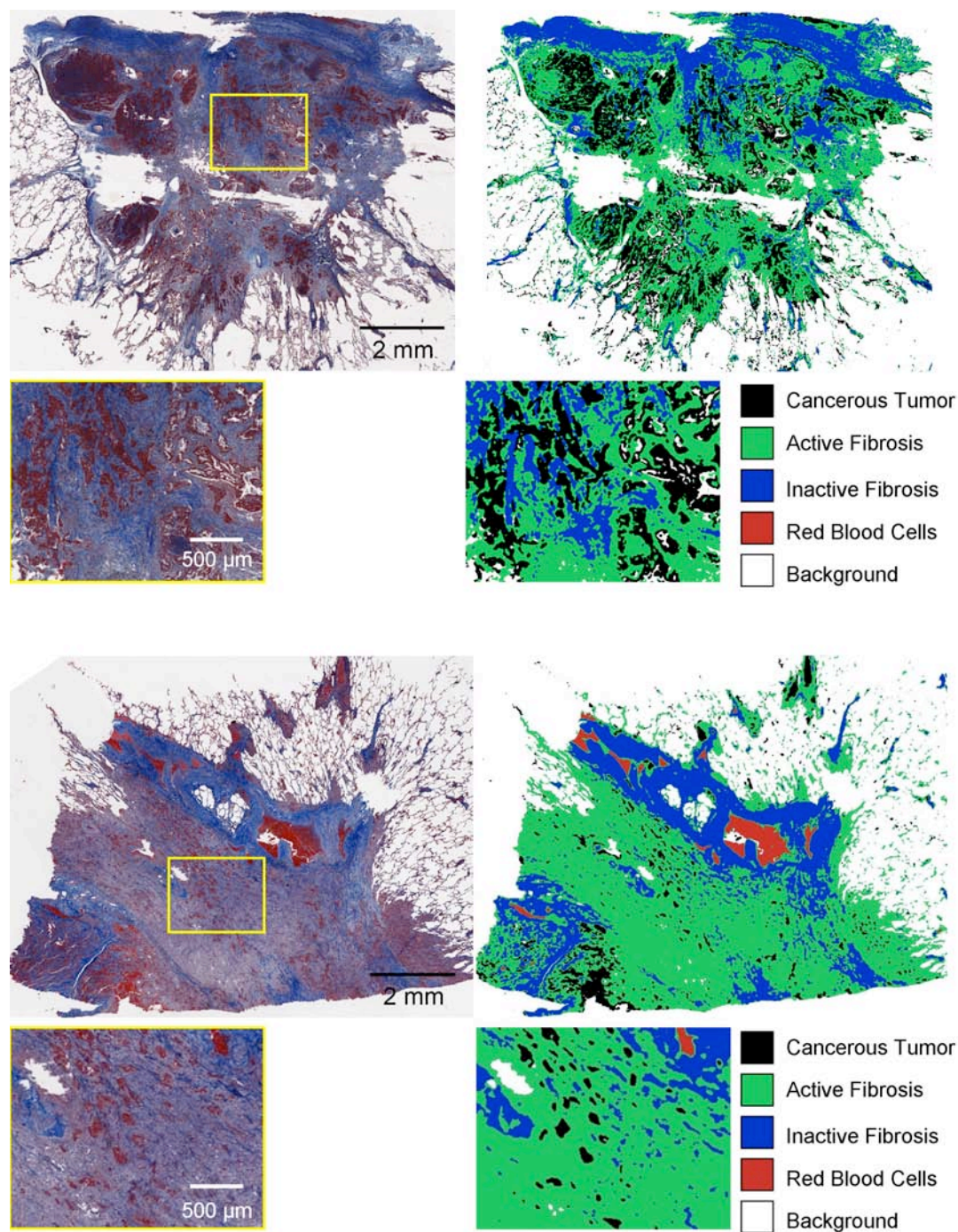
Figure 4-7: An example of the automated classification result for two different adenocarcinoma cases. Of particular interest is the ability of the classifier to achieve consistent classification result despite significant variations in the intensity of the immunohistochemical stain between datasets. A limitation of the automated approach is the exclusion of normal and necrotic tissue classes as there was insufficient distinction of these tissue types in the immunohistochemical staining.
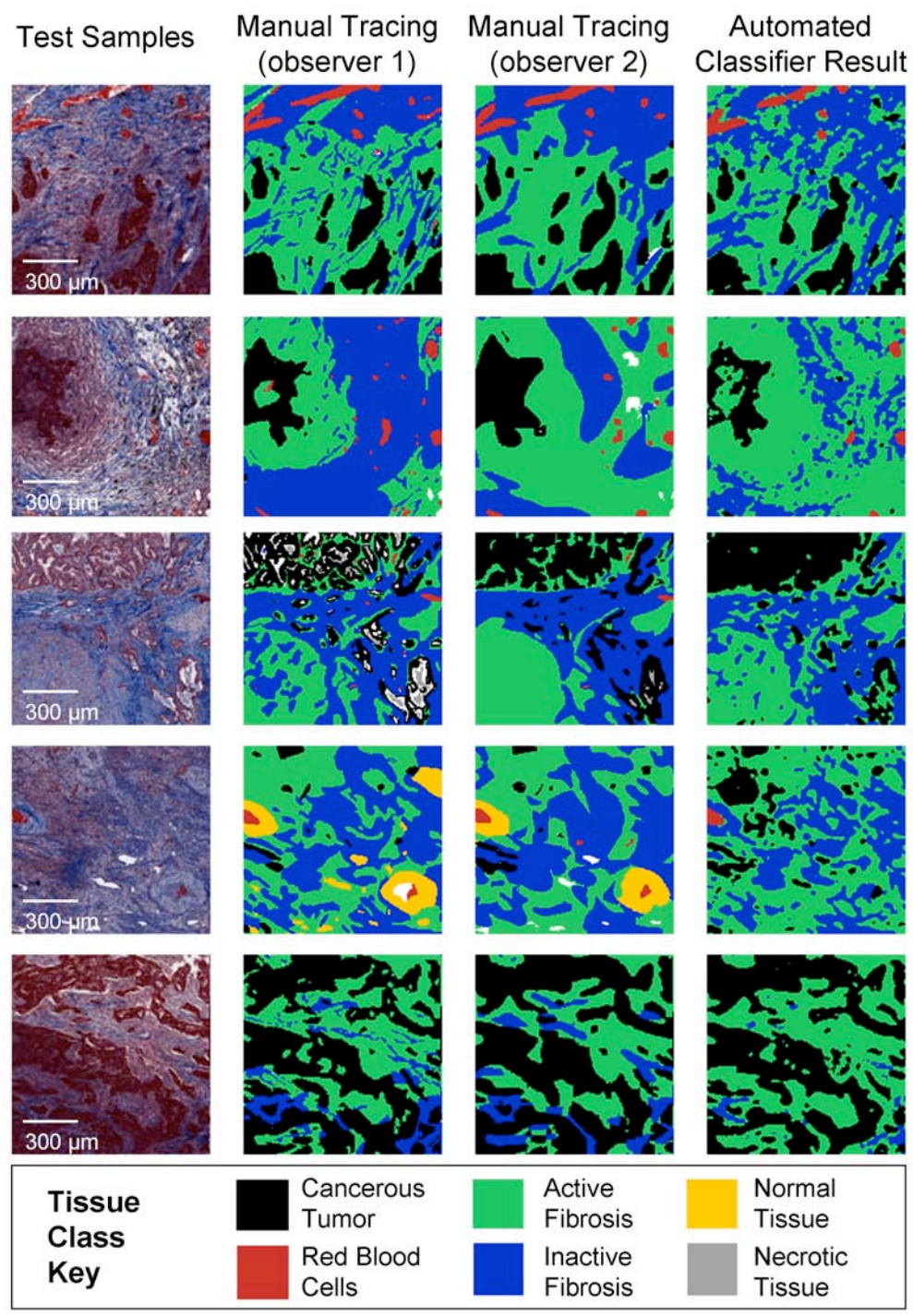
Figure 4-8: Some examples of the test sample images used for validation purposes.
Two manual observers manually traced regions of the different tissue types. These
results were compared to the output of the automated classification approach.

Table 4-2: A confusion matrix showing the percentage of pixels classified into each tissue type class by the two observers. The shaded cells show the agreement between the two observers. The accuracy was calculated by summing the percentage agreed upon, which was 73% between the two manual tracing results.

		Manual Tracing (observer 1)						
		Tumor (solid)	Active Fibrosis	Inactive Fibrosis	Blood	Back-ground	Necrosis	Normal
Manual Tracing (observer 2)	Tumor (solid)	17.7 %	1.6 %	0.8 %	0 %	0 %	0 %	0 %
	Active Fibrosis	2.8 %	28.4 %	7 %	0 %	0 %	0 %	0.1 %
	Inactive Fibrosis	0.5 %	10.7 %	24.5 %	0.2 %	0.2 %	0 %	0 %
	Blood	0.1 %	0.2 %	0.2 %	1 %	0 %	0 %	0 %
	Back-ground	0.9 %	0.1 %	0 %	0 %	0 %	0 %	0.1 %
	Necrosis	0.3 %	0 %	0 %	0 %	0 %	0.1 %	0 %
	Normal	0 %	0.7 %	0.2 %	0 %	0.2 %	0 %	1.3 %

Table 4-3: Confusion matrices comparing the percentage of pixels classified into each tissue type class by the automated segmentation approach and the respective observer. The shaded cells show the agreement between the automated segmentation and the observer. The accuracy for the automated segmentation compared to observer 1 was 72%, while the accuracy for the automated segmentation compared to observer 2 was 69%.

		Manual Tracing (observer 1)						
		Tumor (solid)	Active Fibrosis	Inactive Fibrosis	Blood	Back-ground	Necrosis	Normal
Automated Segmentation	Tumor (solid)	18.3 %	3.5 %	0.7 %	0.2 %	0.1 %	0 %	0 %
	Active Fibrosis	3.6 %	33 %	11.7 %	0.2 %	0.3 %	0.1 %	0.8 %
	Inactive Fibrosis	0.3 %	5.2 %	20.2 %	0.1 %	0 %	0 %	0.7 %
	Blood	0 %	0 %	0.1 %	0.7 %	0 %	0 %	0 %
	Back-ground	0 %	0 %	0 %	0 %	0 %	0 %	0 %
	Necrosis	0 %	0 %	0 %	0 %	0 %	0 %	0 %
	Normal	0 %	0 %	0 %	0 %	0 %	0 %	0 %

		Manual Tracing (observer 2)						
		Tumor (solid)	Active Fibrosis	Inactive Fibrosis	Blood	Back-ground	Necrosis	Normal
Automated Segmentation	Tumor (solid)	17.1 %	4.1 %	0.4 %	0.2 %	0.7 %	0.2 %	0.2 %
	Active Fibrosis	2.8 %	30 %	14.6 %	0.4 %	0.3 %	0.2 %	1.5 %
	Inactive Fibrosis	0.3 %	4.2 %	21 %	0.1 %	0.2 %	0 %	0.7 %
	Blood	0 %	0 %	0.1 %	0.7 %	0 %	0 %	0 %
	Back-ground	0 %	0 %	0 %	0 %	0 %	0 %	0 %
	Necrosis	0 %	0 %	0 %	0 %	0 %	0 %	0 %
	Normal	0 %	0 %	0 %	0 %	0 %	0 %	0 %
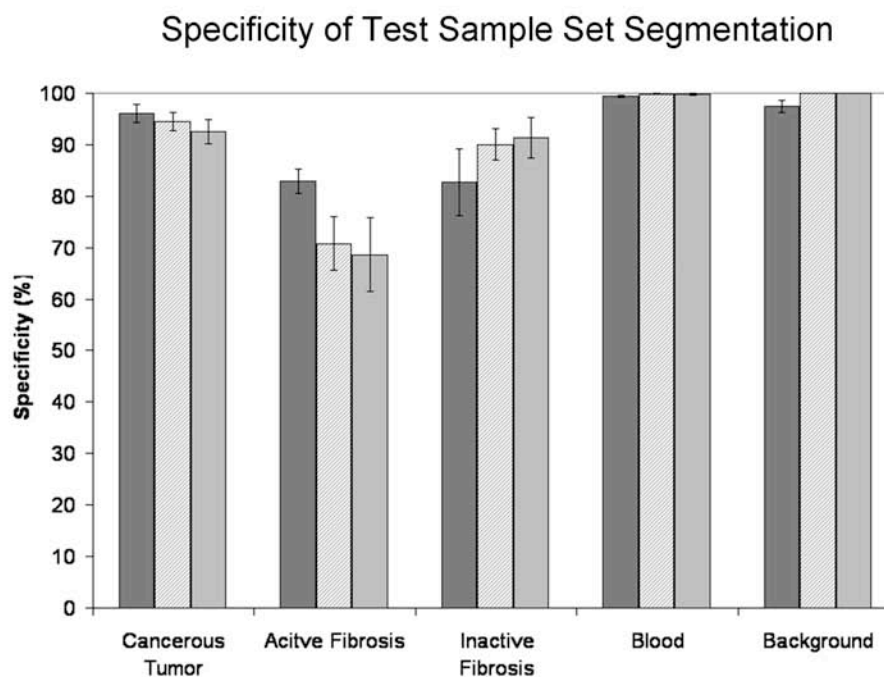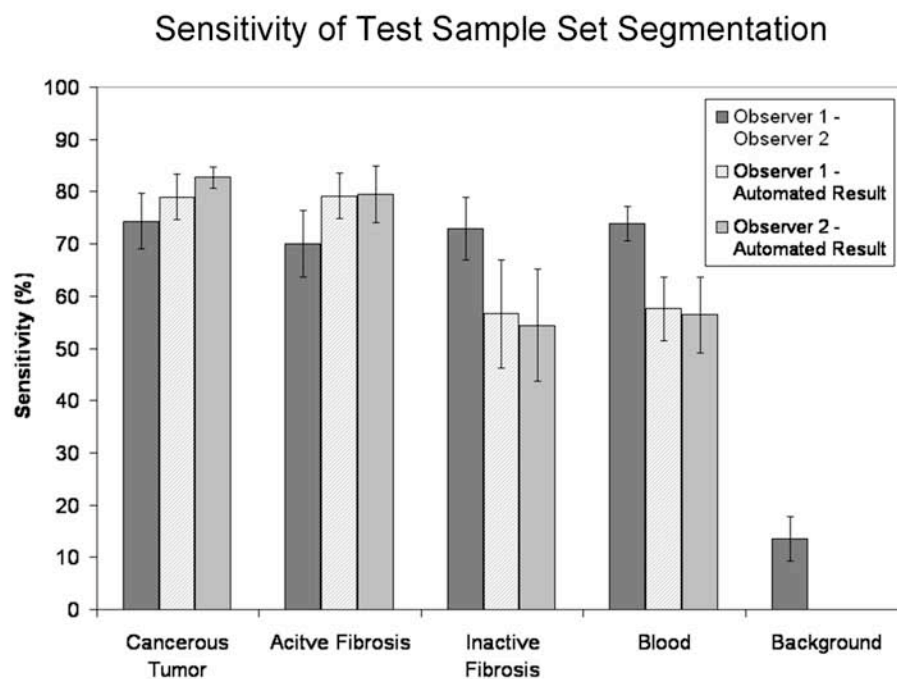
Figure 4-9: Sensitivity and specificity graphs. These graphs compare the segmentations of each tissue class between the two observers, observer 1 and the automated result and observer 2 and the automated result. The average and standard error of the performance across all test image sets is shown.

CHAPTER 5

REGISTRATION PIPELINE

In previous studies, a significant barrier to the registration of histological data to non-destructive data has been the lack of reliable spatial correspondence between the datasets. However, the development of the large-scale image microscope array (LIMA) system served to bridge this information gap for this study. **Figure 5-1** provides a summary of the registration pipeline developed to bring the multimodal lung nodule datasets to a common alignment. This pipeline incorporates three registration approaches; two-dimensional (2D) rigid registration, 2D non-rigid registration and three dimensional (3D) rigid registration.

5.1   Registration of the Micro-CT Data

The lung nodule tissue properties were not altered between the acquisition of the micro-CT data and the LIMA imaging and so a non-rigid registration was deemed unnecessary to correct the misalignment between the datasets. A rigid, similarity transform, incorporating translation, rotation and scaling was selected as appropriate for the LIMA and micro-CT registration [86].

Each slice of the LIMA dataset was acquired from the cut tissue block so that a large depth of field was incorporated into the image. Comparing an airway representation in both datasets, **Figure 5-2**, reveals the difference the incorporation of this depth of field makes on the resulting image set. This difference in image content can have a negative impact on intensity based registration metrics.

A three dimensional rigid registration approach, incorporating a normalized mutual information metric was explored for the registration of the micro-CT and LIMA datasets (see equation 5-10 below). Pre-processing was used to gain a rough alignment between the datasets and to transform the LIMA dataset to grayscale. While it was possible to achieve a visually acceptable registration result for some of the nodule datasets using this method, the registration of datasets which featured numerous or prominent airway structures failed. As it was desired to have a consistent registration approach that could be used to reliably align all nodule cases, a landmark based approach was chosen over the normalized mutual information approach.

Corresponding micro-CT and LIMA image slices were determined though the identification of the best matching micro-CT slice for the first LIMA image. This was guided by the amount of tissue grossed from the LIMA prior to beginning image acquisition. Subsequent micro-CT slices were determined based on z-axis depth, in microns, from this starting position. The nodule stabilization system, described as a portion of the data acquisition pipeline, prevented any motion in the z-plane, hence the rigid registration transform could be calculated in 2D and applied to all image slices to produce alignment between the three dimensional datasets.

The 2D rigid registration algorithm, summarized in **Figure 5-3**, consisted of a similarity transform in which the registration landmark points were used to solve for the transform parameters by minimizing the distance between corresponding moving and fixed landmarks. The similarity transform f_S, consists of a translation vector, t, rotation matrix, A, and scale matrix λ.

$$f_S(x) = \lambda A x + t \qquad\qquad \textbf{5-1}$$

The parameters of the similarity transform were found using the two corresponding registration landmark sets, Lr_m and Lr_f so that the distance measure $D(f_s)$ was minimized, where;

$$D(f_S) = \sum_{i=1}^{n} \left\| f_S(Lr_m^i) - Lr_f^i \right\|^2 = \sum_{i=1}^{n} \left\| \lambda A Lr_m^i + t - Lr_f^i \right\|^2 \qquad \textbf{5-2}$$

Landmark points were used to calculate the transform as the large difference in image content in the micro-CT and LIMA datasets deemed them unsuitable for other intensity based approaches. Four corresponding registration landmark points, *Lr,* and two corresponding evaluation landmark points, *Le,* were manually selected in the LIMA and micro-CT datasets for each nodule case. These points were saved in a registration landmark file and an evaluation landmark file and input to the registration algorithm along with the moving image set.

The output from the registration algorithm included the aligned moving dataset and the transform parameters applied. In addition, the fiducial registration error (FRE) and the target registration error (TRE) were output from the algorithm for validation purposes. The FRE directly equates to the distance measure $D(f_s)$ that was being minimized to solve for the transform, using the registration landmark set.

$$FRE = D(f_s) = \sum_{i=1}^{n} \left\| \lambda A Lr_m^i + t - Lr_f^i \right\|^2 \qquad \textbf{5-3}$$

The TRE was also calculated as a distance measure between fixed landmarks and transformed, moving landmarks. However, the TRE was based on the evaluation landmark set, Le_m and Le_f, which was not used to determine the registration parameters and hence reflects how well the determined transform parameters map other points in the image to their corresponding fixed target points.

$$TRE = \sum_{i=1}^{n} \left\| \lambda A L e_m^i + t - L e_f^i \right\|^2 \qquad \text{5-4}$$

In the ideal case where the determined transform exactly translated the provided moving image to the specified fixed image, the FRE and TRE values would be zero.

For the registration of the LIMA and micro-CT data the average and standard deviation of the FRE was 2.83 ± 0.84 pixels and the TRE was 3.63 ± 2.95 pixels for the SkyScan micro-CT datasets. In the global coordinate system this equated to a physical FRE distance of 51.01 ± 15.09 micron and a TRE of 65.4 ± 53.15 micron. For the micro-CT data generated from the MicroCAT II system the FRE was 1.59 ± 0.69 pixels and the TRE was 3.10 ± 1.40 pixels. In the global coordinate system this equated to a physical FRE distance of 44.57 ± 19.22 microns and a TRE of 86.83 ± 39.23 microns.

5.2   Registration of Histological Data

Due to the extensive processing involved with the creation of histology slides, a loss of structural integrity of the tissue sample occurs. Bending, shearing and tearing of the tissue is common as the tissue slice is transferred to the slide. This is especially so in the case of the preparation of lung tissue due to the very fragile nature of the alveolar walls. Thus, rigid registration alone would not result in the accurate alignment of the histopathology data to the LIMA images.

As with the micro-CT registration case, there exists a significant difference in the image content from the LIMA section and the histology data. Also, the process of histopathology slide creation can result in incomplete tissue sections with portions of the tissue missing from the slide. These factors together resulted in a landmark registration

approach being selected for the registration of the histopathology data to the global
coordinate system.

A landmark based thin-plate spline algorithm was developed to non-rigidly
register each histopathology, moving, image to its corresponding LIMA, fixed, image.
Figure 5-4 summarizes the 2D thin plate spline registration approach. The algorithm was
based on the mathematics originally presented by Bookstein [87], in which a spline, f_{TPS},
is found which interpolates the moving landmarks, Lr_m to the fixed landmarks, Lr_f, In
thin plate spline registration the smallest possible smooth deformation is found by
minimized bending energy, that permits the landmarks to be exactly mapped to each
other.

$$f_{TPS}(x) = Dx + \sum_{i-1}^{n} w_i U\left(\left|x - Lr_m^i\right|\right) \qquad \text{5-5}$$

where $U(r) = r^2 \log r^2$, D is a matrix representing the affine transformation, w_i is
the warping coefficient matrix which represents the non-affine deformation.

The bending energy function, E, is minimized to control the extent of warping to
the smallest deformation possible:

$$E[f(x)] = \iint_{\Re^2} \left(f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2\right) dxdy \qquad \text{5-6}$$

The correspondence between the LIMA and histology images was inherently
gained through the acquisition technique. For each LIMA and histology pair registration
landmarks, Lr, and evaluation landmarks, Le, were manually selected and saved to text
files. The number of landmarks chosen was dependent on the size of the histology slide
and the number of reliable landmark structures available, with the average and number of
landmarks being 36 ± 6. The landmark text files, the fixed base image and the moving

images were applied as inputs to the two dimensional thin plate spline registration algorithm. Multiple moving histological images could be input to the algorithm and have the same spline deformation applied. This was required in the case of registering the tissue type maps to the global coordinate system.

The output from the registration approach included the transformed histopathology image. The deformation field, which represented the applied spline, was also output, and example is shown in Figure 5-5. For evaluation purposes, the TRE was also calculated and output. The nature of the spline calculation is to force the moving registration landmarks Lr_m to the exact position of the fixed registration landmarks, Lr_f. Hence the FRE is always equal to zero for this approach.

$$FRE = \sum_{i=1}^{n} \left\| f_{TPS}(Lr_m^i) - Lr_f^i \right\|^2 = 0 \qquad \textbf{5-7}$$

The TRE utilizes evaluation landmarks which are not used to calculate the spline warping and hence reflects how closely the deformation field matches target points in the translated moving image to the corresponding positions in the fixed image space.

$$TRE = \sum_{i=1}^{n} \left\| f_{TPS}(Le_m^i) - Le_f^i \right\|^2 \qquad \textbf{5-8}$$

The average and standard deviation of the TRE for the landmark based thin plate spline warping of the histology image set to the LIMA base set was 5.04 ± 2.20 pixels which translated to 43.08 ± 18.83 microns in global space.

5.3   Registration of *Ex Vivo* MDCT Data

Registration of the *ex vivo* MDCT data to the global coordinate space was achieved using the volume dataset. A rigid registration algorithm was deemed

appropriate for the registration of the isolated lung nodule MDCT and the fixed lobe
MDCT data to the micro-CT dataset as the tissue properties were not altered between
imaging. The rigid registration transform, f_R, involved translation, t and rotation, A:

$$f_R(x) = Ax + t$$
                                        5-9

A Quasi-Newton optimizer was used to step the transformation towards a
minimum solution. As a significant difference in resolution was present between the
micro-CT dataset and the MDCT dataset, a multi-resolution optimization was included.
This involved re-sampling the data to four levels; (8,8,3), (4,4,2), (2,2,1) and (1,1,1). The
registration transform was found, first using the lower resolution data then that result was
used as the initial setting for the next resolution level, and so on. Incorporating the multi-
resolution approach into the optimization decreased the computation time while
increasing the robustness of the registration approach.

The normalized mutual information (NMI) metric, has been found to be more
robust than the standard mutual information metric [88]. This metric does not rely on the
direct correspondence between graylevel values in the dataset but rather compared the
overlap in information. The NMI of images A and B is calculated from the sum of the
marginal entropies of A and B, divided by the joint entropy:

$$NMI(A, B) = \frac{H(A) + H(B)}{H(A, B)}$$
                                        5-10

The entropy is based on the probability distribution of the gray values which are
found from the histogram of the image.

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log_b p(x_i)$$
                                        5-11

Validation of this registration approach was performed by repeating the registration process three times and calculating the standard error between the resulting transformation parameters. For the registration of the isolated lung nodule to the micro-CT dataset the standard errors for resulting translations in x, y and z were 0.9 microns, 4.5 microns and 5.4 microns respectively. The standard error for rotation was 6.3 microns. For the registration of the fixed lung lobe to the micro-CT data the standard errors in translation were 2.0 micron in x, 6.3 microns in y and 2.0 microns in z. For rotation the standard error was 2.2 microns which are all sub-pixel registration errors.

Qualitative validation was also performed using checkerboard and fused image sets of the fixed and moving datasets following registration. These datasets allowed the visualization of the matching between structural points of reference such as tumor boundaries and airways. The checkerboard image shows alternating squares from each dataset while the fused image shows a colormap of the overlap between the dataset. An example of the checkerboard and fused images for the registered isolated nodule MDCT with the micro-CT dataset is shown in **Figure 5-7**, and the fixed lobe MDCT with the micro-CT is shown in **Figure 5-8**.

5.4   Registration of *In Vivo* MDCT Data

The *in vivo* MDCT datasets were acquired pre-surgically for clinical purposes and added to the study once a patient was consented. Unfortunately, the quality and consistency of the scanning protocol across the nodule cases, was less than ideal. As a part of this study patients were asked to participate in a high resolution, pre-surgical perfusion MDCT – but all refused. As different voltage, current, reconstruction kernels

and slice thicknesses were used, this dataset was deemed inappropriate for the comparison of Hounsfield Unit variance to histopathological tissue type. Hence, the *in vivo* MDCT dataset was aligned to the global coordinate system using the 3D rigid registration approach. Although it could not be assumed that the tissue properties had not been altered between this data acquisition and the fixed tissue acquisition, the limited resolution of theses datasets deemed more sophisticated registration approaches unnecessary.

Manual alignment between the *in vivo* MDCT and the fixed lobe MDCT was first achieved using the large airways and vessels in the lobe as structural landmarks. Once a rough alignment was achieved the 3D rigid registration approach was applied, as described above.

Once again the error in the misalignment of the *in vivo* MDCT data to the global coordinate system was evaluated by repeating the registration three times and calculating the standard error in the resulting transformations. The calculated standard errors for resulting translations in x, y and z were 1.0 mm, 0.6 mm and 3 mm respectively. The standard error for rotation was 1.2 mm. As one would expect the standard errors reflect the difficulty in registering such sparse data. The overall tendency across the datasets, for the large slice separation to be in the global z-axis is reflected by the large standard error in the z translation.

5.5 Results

The developed registration pipeline was applied to register all the nodule datasets to a common coordinate system. The pipeline was developed to be flexible in

accommodating different types of lung cancer with varying structural features, such as the large airways found in many squamous cell carcinomas. Through this registration subsequent analysis could be conducted relating the tissue classes, as defined from the histopathological data, to the registered MDCT dataset on a pixel by pixel basis. Example 2D slices from the registered datasets are shown in **Figure 5-10** to **Figure 5-15**. The data can also be viewed as a set of registered volumes as shown in **Figure 5-16**.
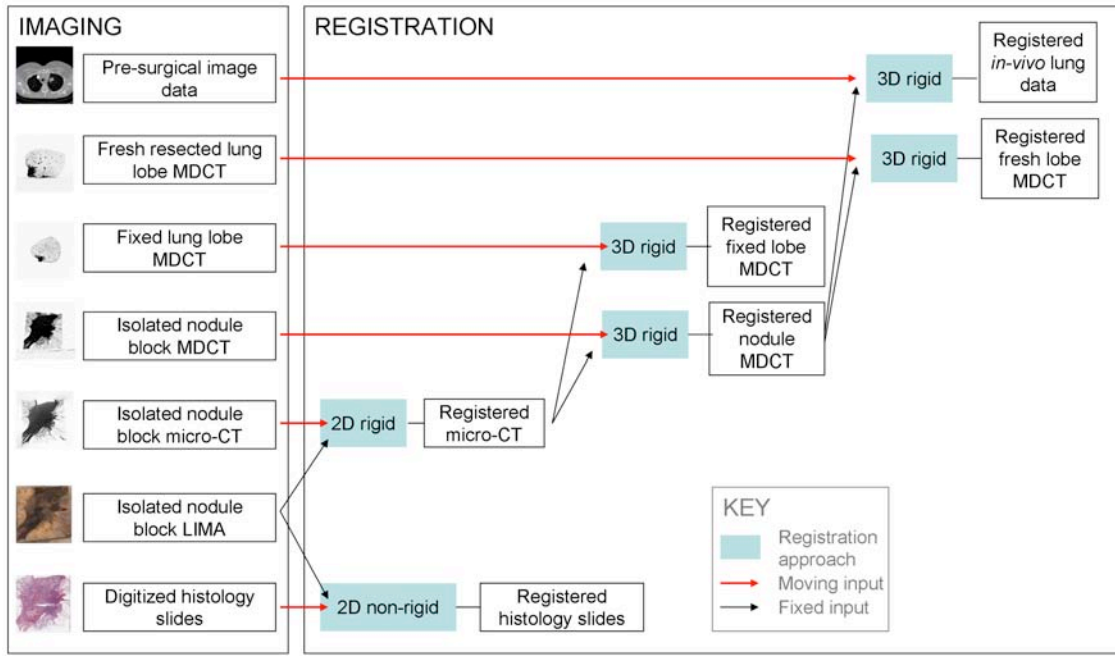
Figure 5-1: Summary diagram showing the registration approaches applied to map the multimodal image datasets to a common global image space. The red arrows indicate the moving image input while the black arrows show the fixed image input.
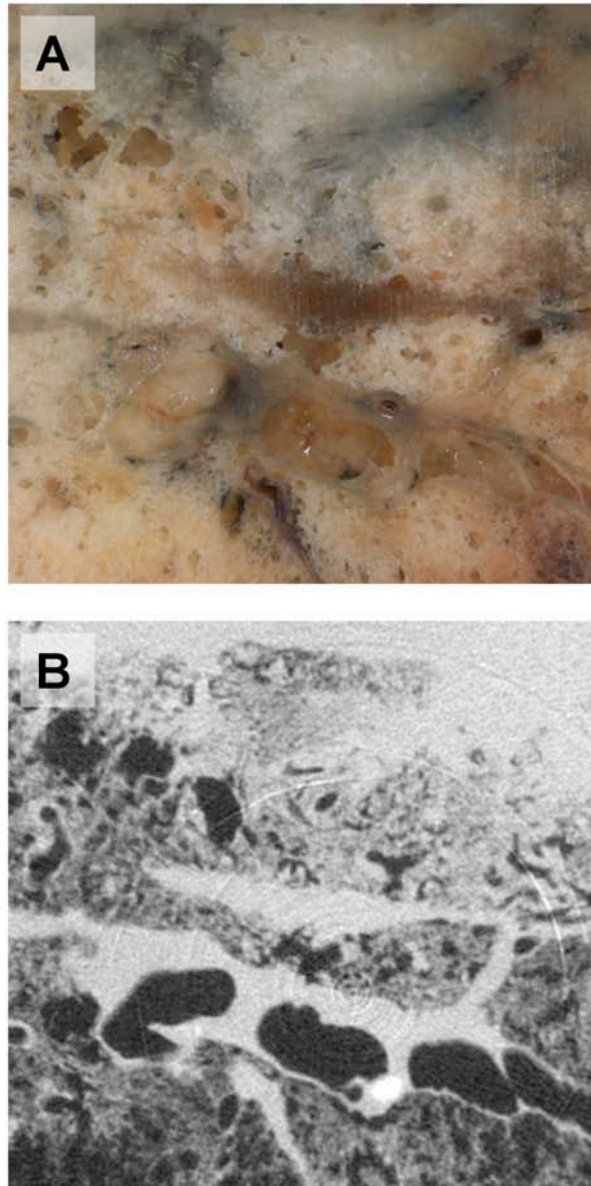
Figure 5-2: Comparing the image content of the LIMA and micro-CT datasets. A large difference in image content in the LIMA (A) and micro-CT (B) dataset, shown post registration such that the images are aligned. Due to the nature of image acquisition for the LIMA, capturing the tissue surface of the cut tissue block, a large depth of field is present which significantly alters the appearance of airway structures compared to the micro-CT representation.
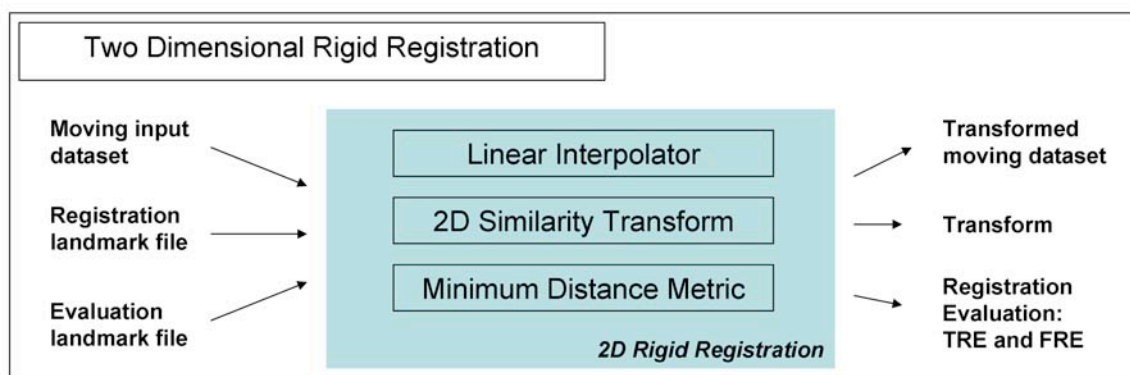
Figure 5-3: Diagram of the inputs and outputs of the two dimensional rigid registration approach used for registering the micro-CT dataset to the global coordinate system. The minimum distance metric is used to evaluate the parameters of the similarity transform based on the input registration landmarks. A linear interpolator maps the pixels from the input image to their transformed position and this transformed image in output from the algorithm. The target registration error (TRE) and fiducial registration errors (FRE) are also output for evaluation purposes.
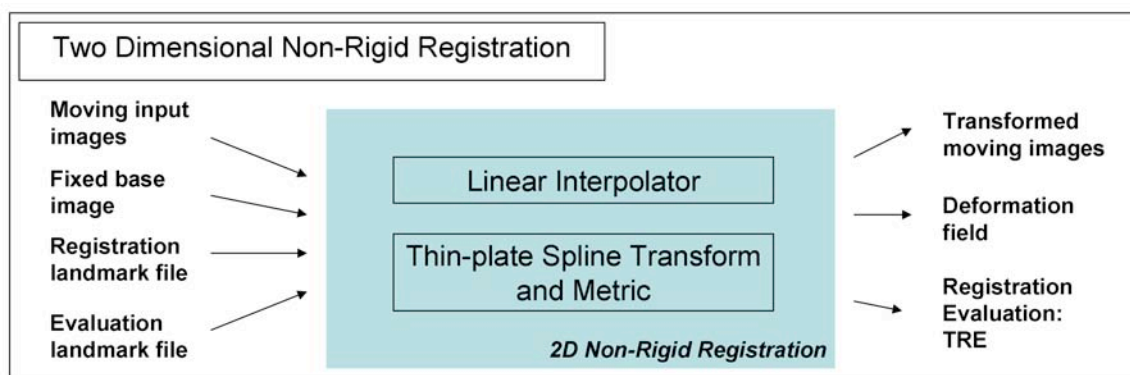
Figure 5-4: The two dimensional thin plate spline, non-rigid registration approach used to register the histopathology data to the global coordinate system. This approach took the moving and fixed images as well as the registration and evaluation landmark files as input. The registration landmarks were used to calculate the spline transform while minimizing the bending energy. A linear interpolator was used to map the moving image pixels to the transformed position. The resulting transformed image, the spline deformation field and the target registration error (TRE) were provided as outputs from the algorithm.
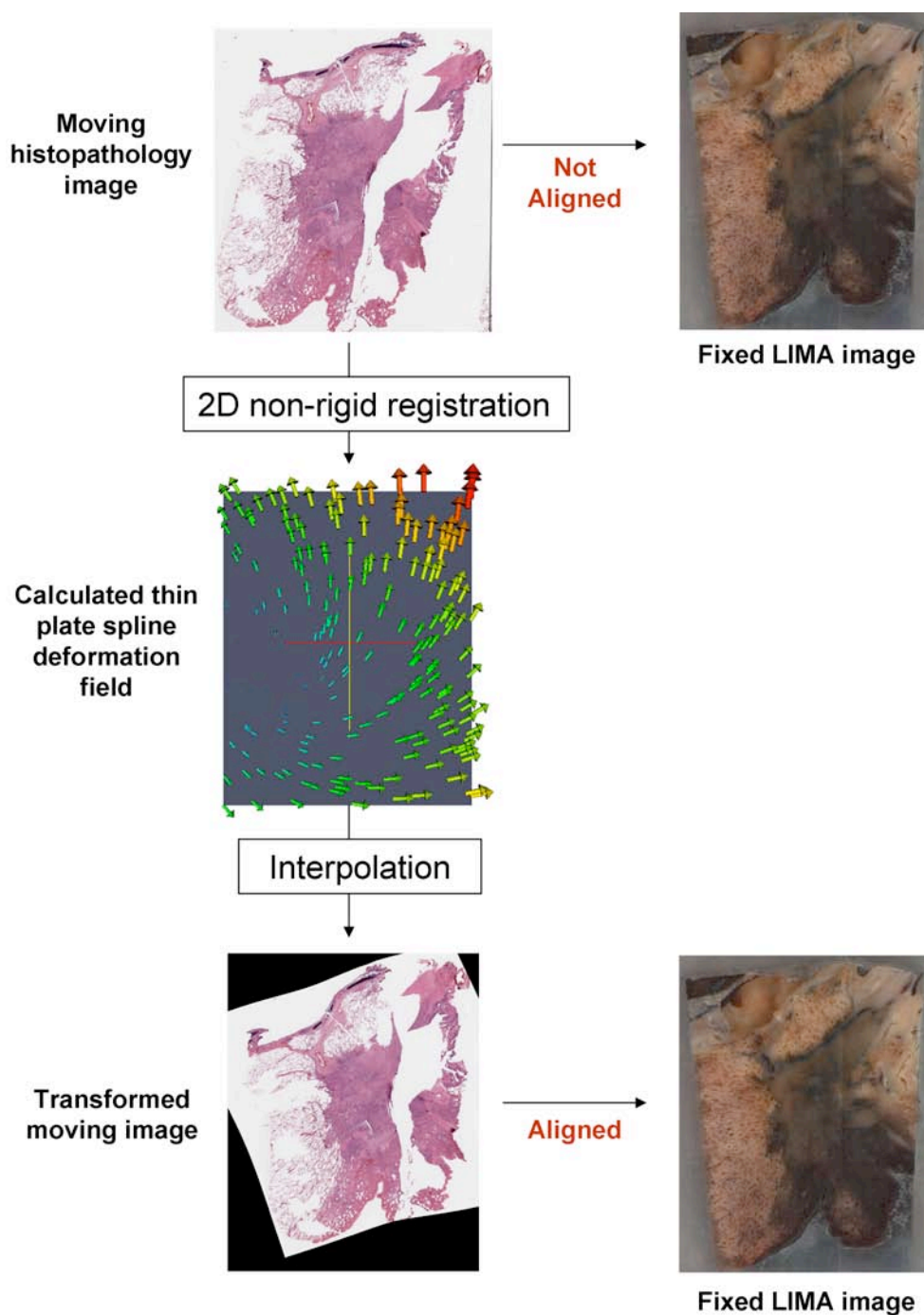
Figure 5-5: An example of the 2D thin plate spline, non-rigid registration approach.
This figure shows the original histopathology image, the calculated deformation field and
the resulting transformed histopathology image in alignment with the LIMA dataset.
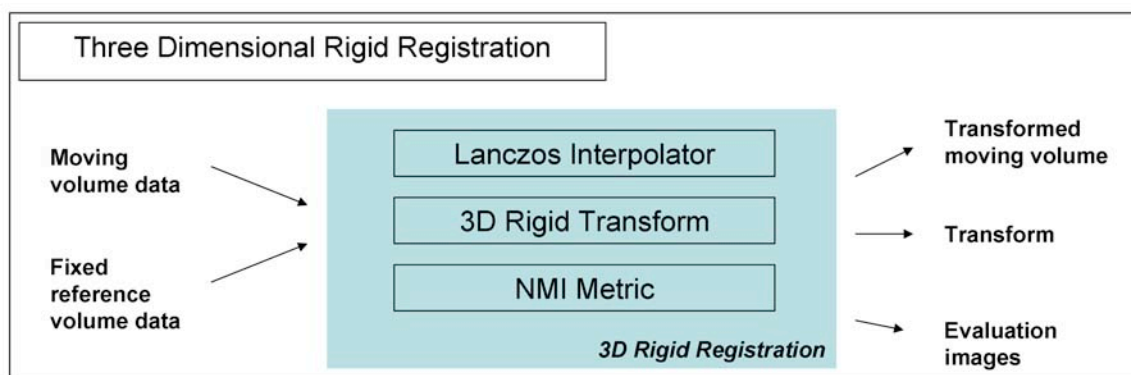
Figure 5-6: The three dimensional rigid registration approach used to register the MDCT data to the global coordinate system. This approach took the moving and fixed volumes as inputs. The normalized mutual information (NMI) metric was used to evaluate the registration. A lanczos interpolator was used to map the moving image pixels to the transformed position in 3D space. The resulting transformed volume, the parameters of the transform and the evaluation images were produced as output.
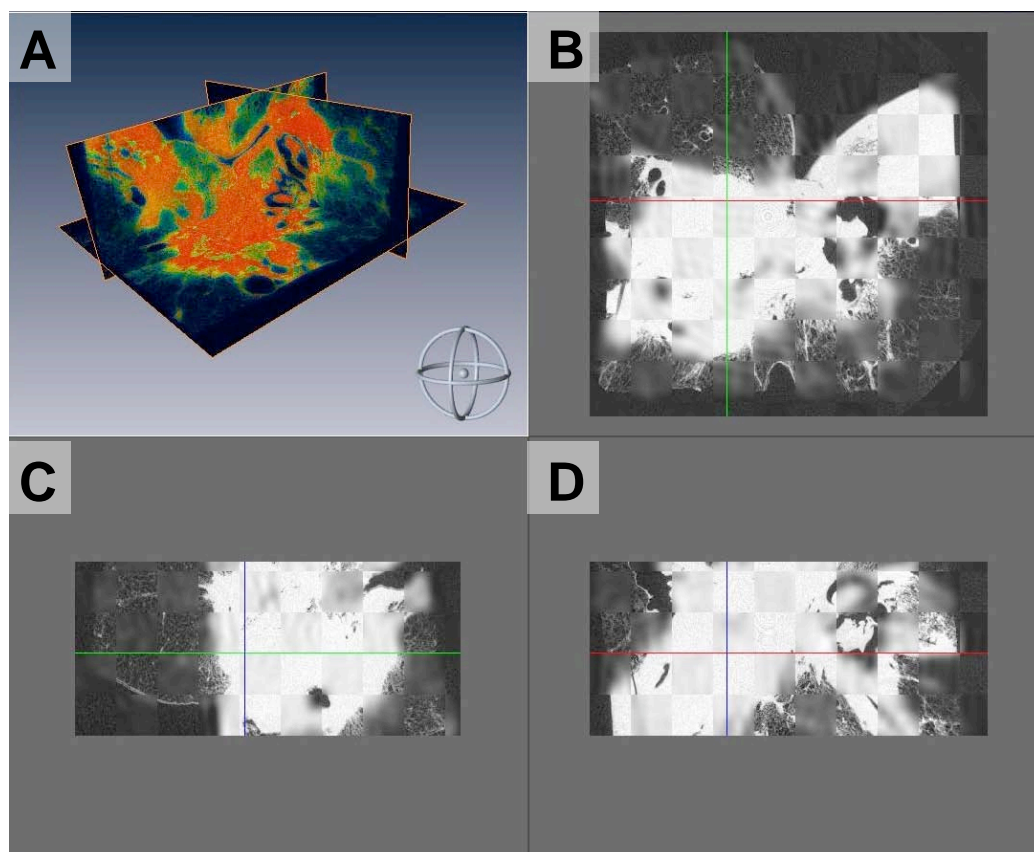
Figure 5-7: Qualitative validation of the 3D rigid registration approach was achieved through the use of fused and checkerboard image sets. The fused data (A) shows a colormap of the overlap between the registered isolated nodule MDCT and the micro-CT dataset. The checkerboard images show alternating squares from the two datasets in the axial (B), sagittal (C) and coronal (D) view.
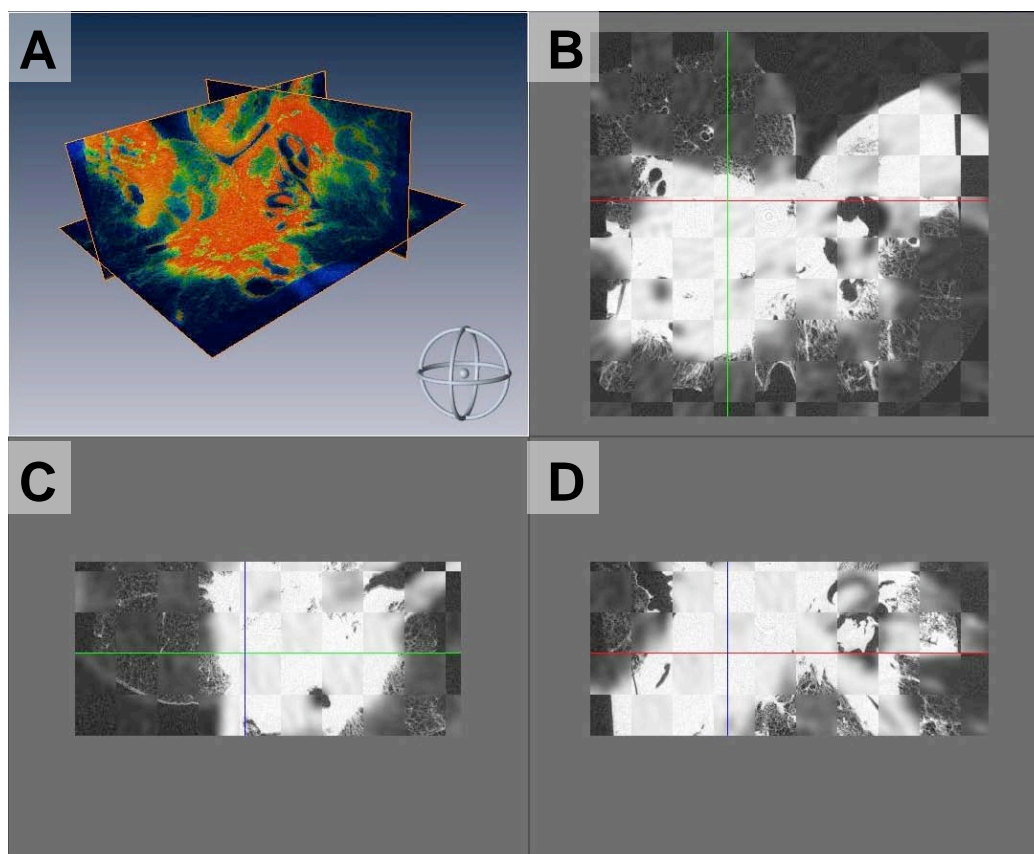
Figure 5-8: The evaluation of the fixed lobe MDCT registration to the micro-CT.
The alignment between the nodule boundaries and internal structures in the registered
fixed lobe MDCT and the micro-CT dataset can be viewed in the fused data (A) and the
axial (B), sagittal (C) and coronal (D) checkerboard images.
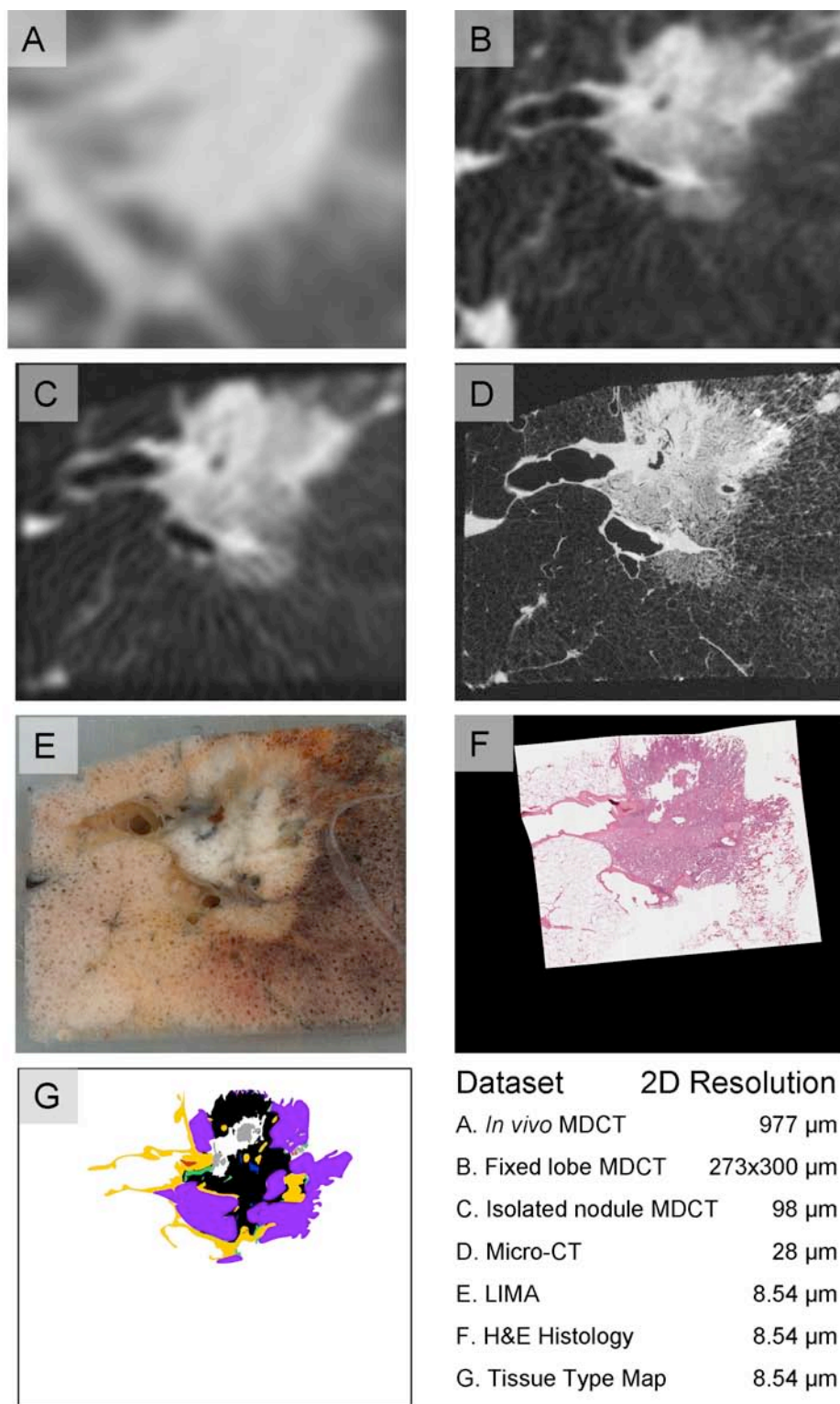
Figure 5-9: Registered adenocarcinoma case 1. A 2D slice from the volumetric, registered, multi-modal dataset with a tabulated account of the resolutions.
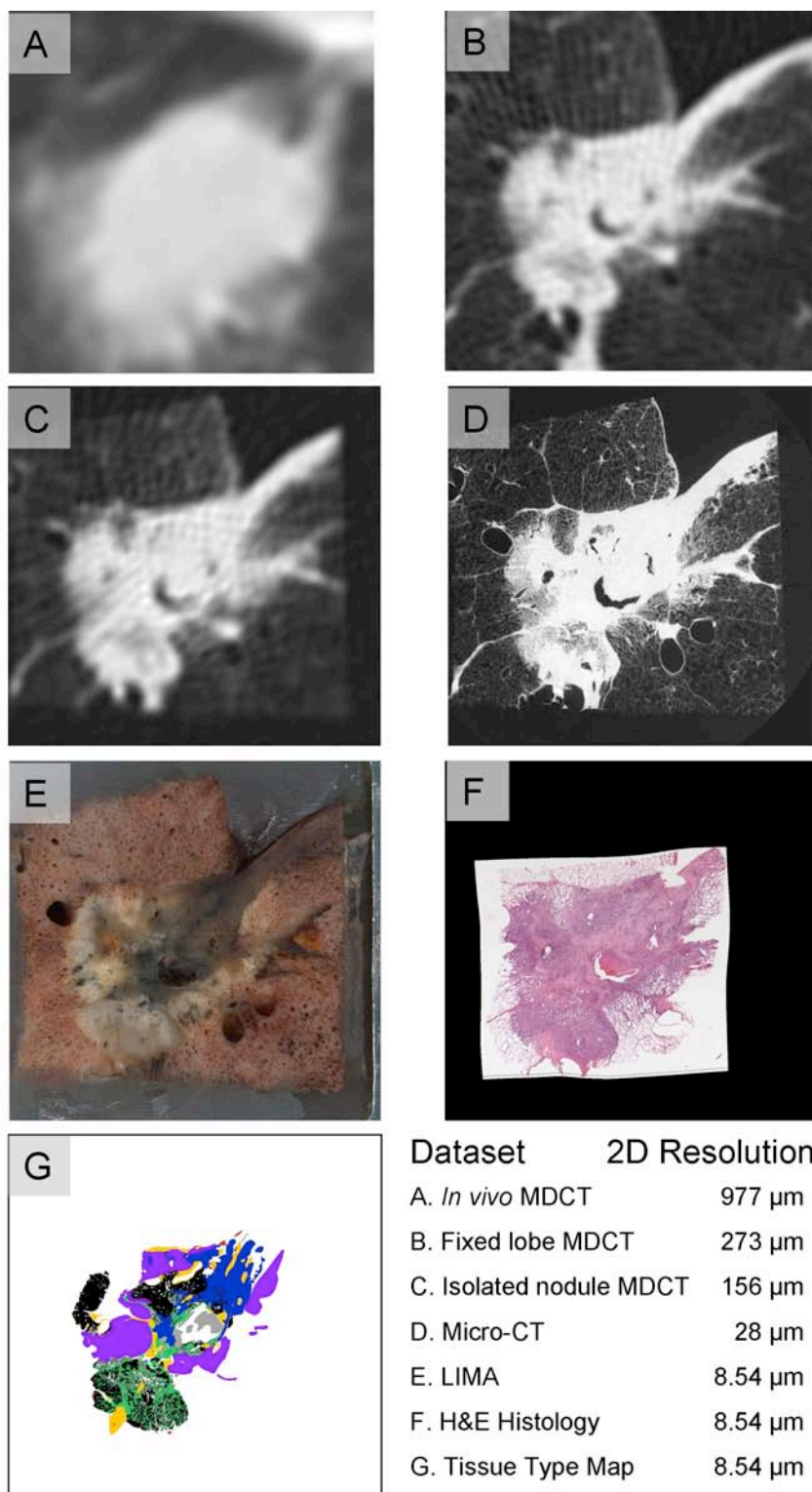
Figure 5-10: Registered adenocarcinoma case 2. A 2D slice from the volumetric, registered, multi-modal dataset with a tabulated account of the resolutions.
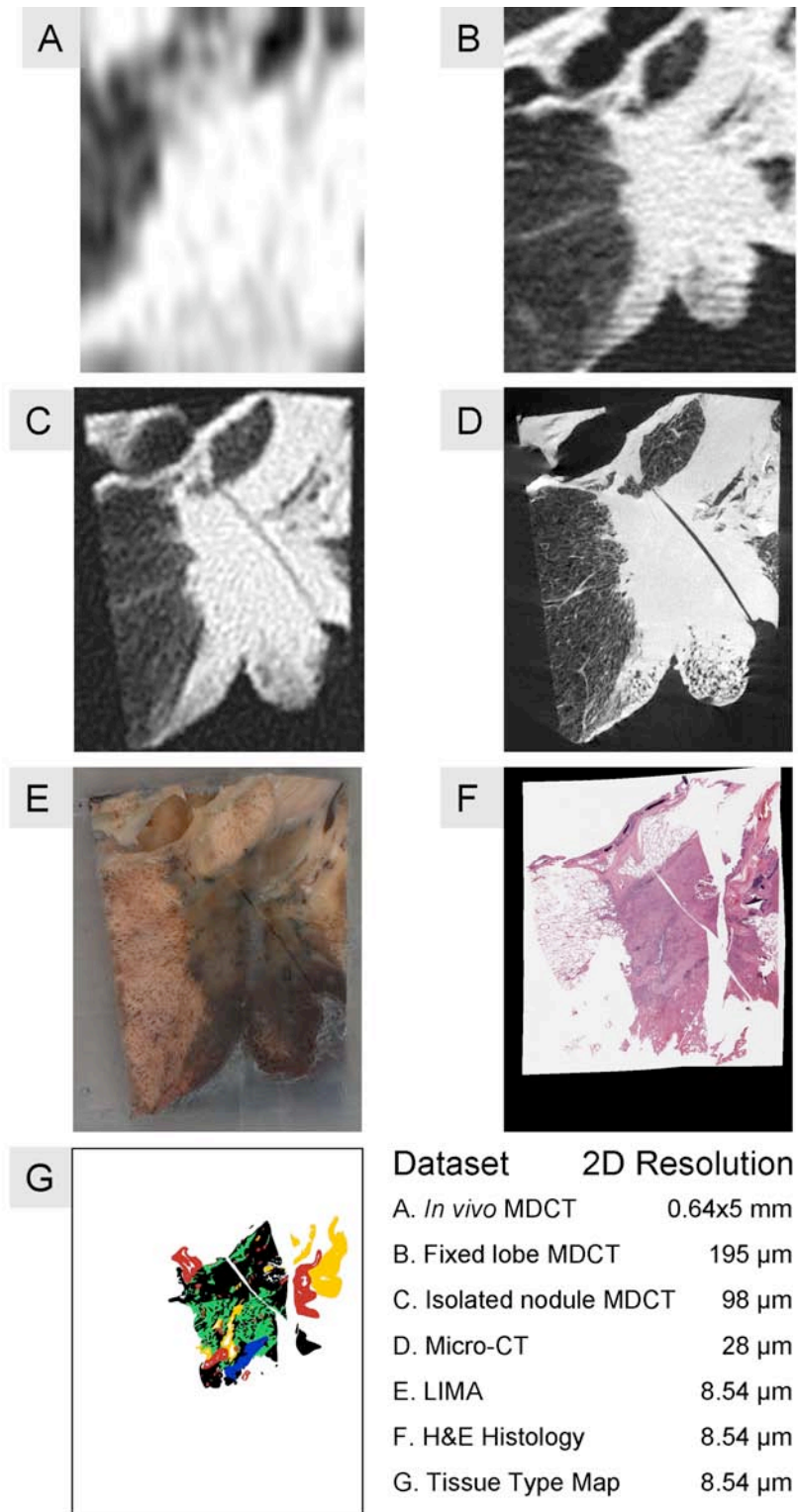
Figure 5-11: Registered adenocarcinoma case 3. A 2D slice from the volumetric, registered, multi-modal dataset with a tabulated account of the resolutions.
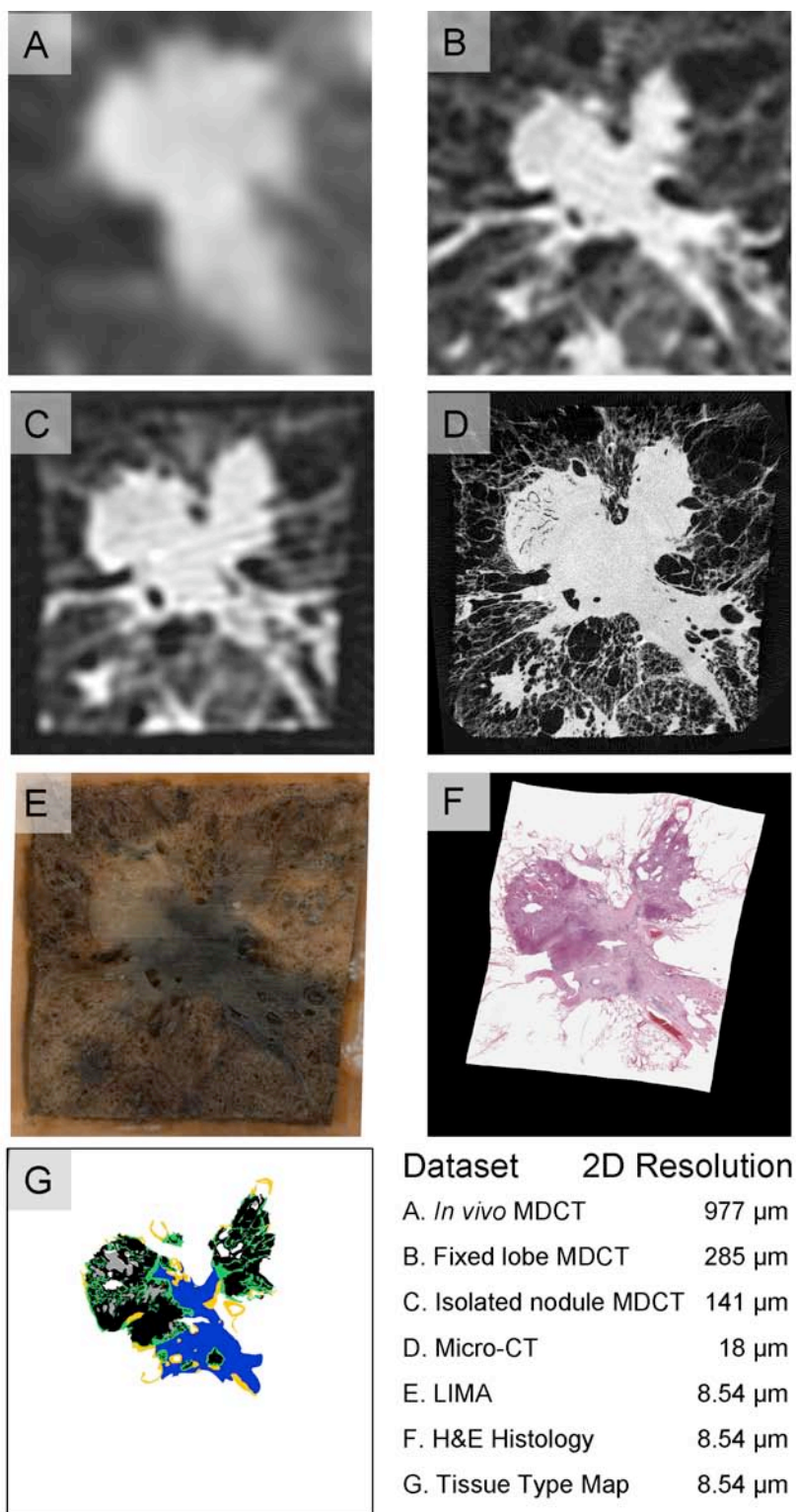
Figure 5-12: Registered adenocarcinoma case 4. A 2D slice from the volumetric, registered, multi-modal dataset with a tabulated account of the resolutions.
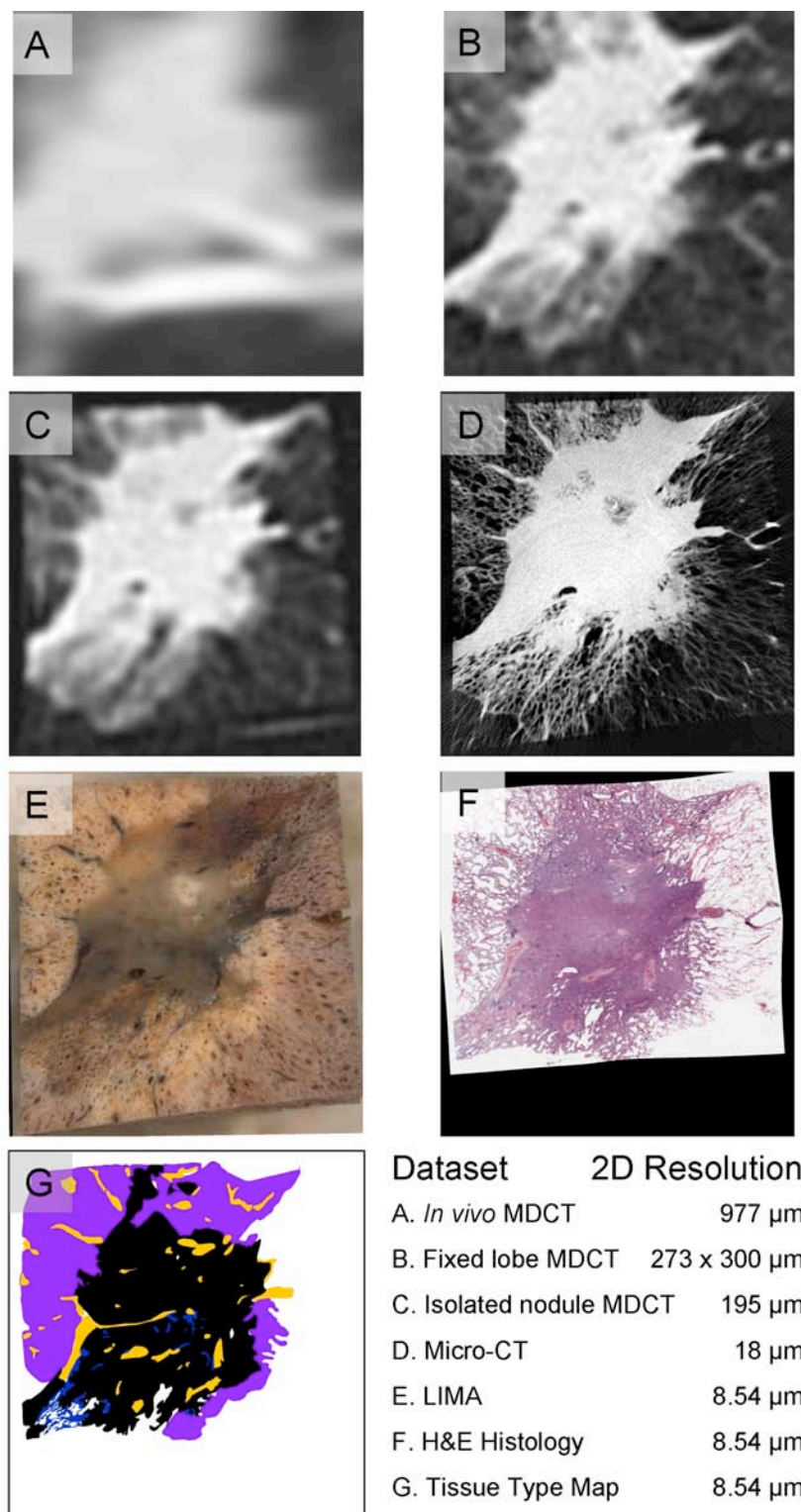
Figure 5-13: Registered adenocarcinoma case 5. A 2D slice from the volumetric, registered, multi-modal dataset with a tabulated account of the resolutions.
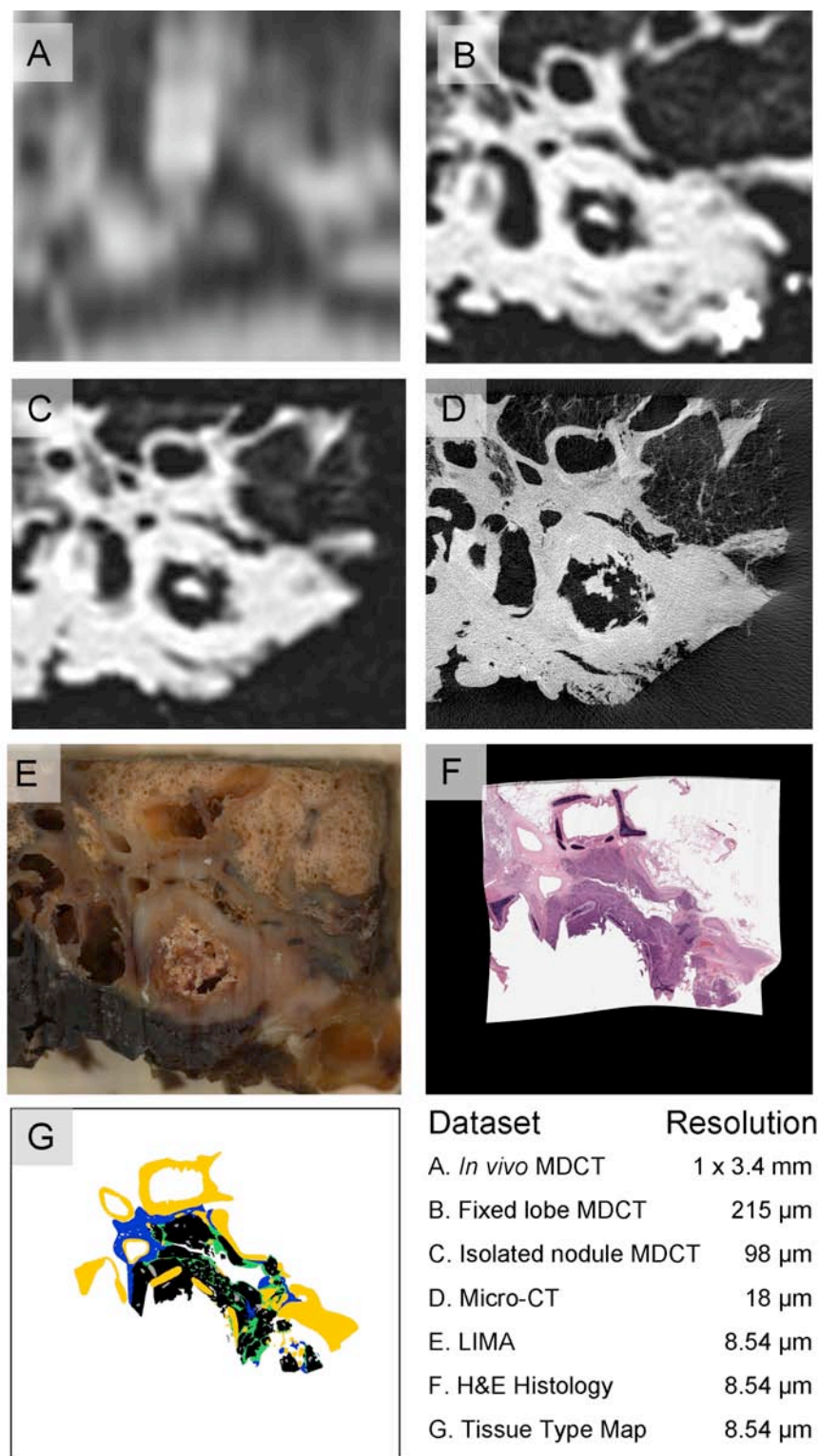
Figure 5-14: Registered squamous cell carcinoma. A 2D slice from the volumetric, registered, multi-modal dataset with a tabulated account of the resolutions.
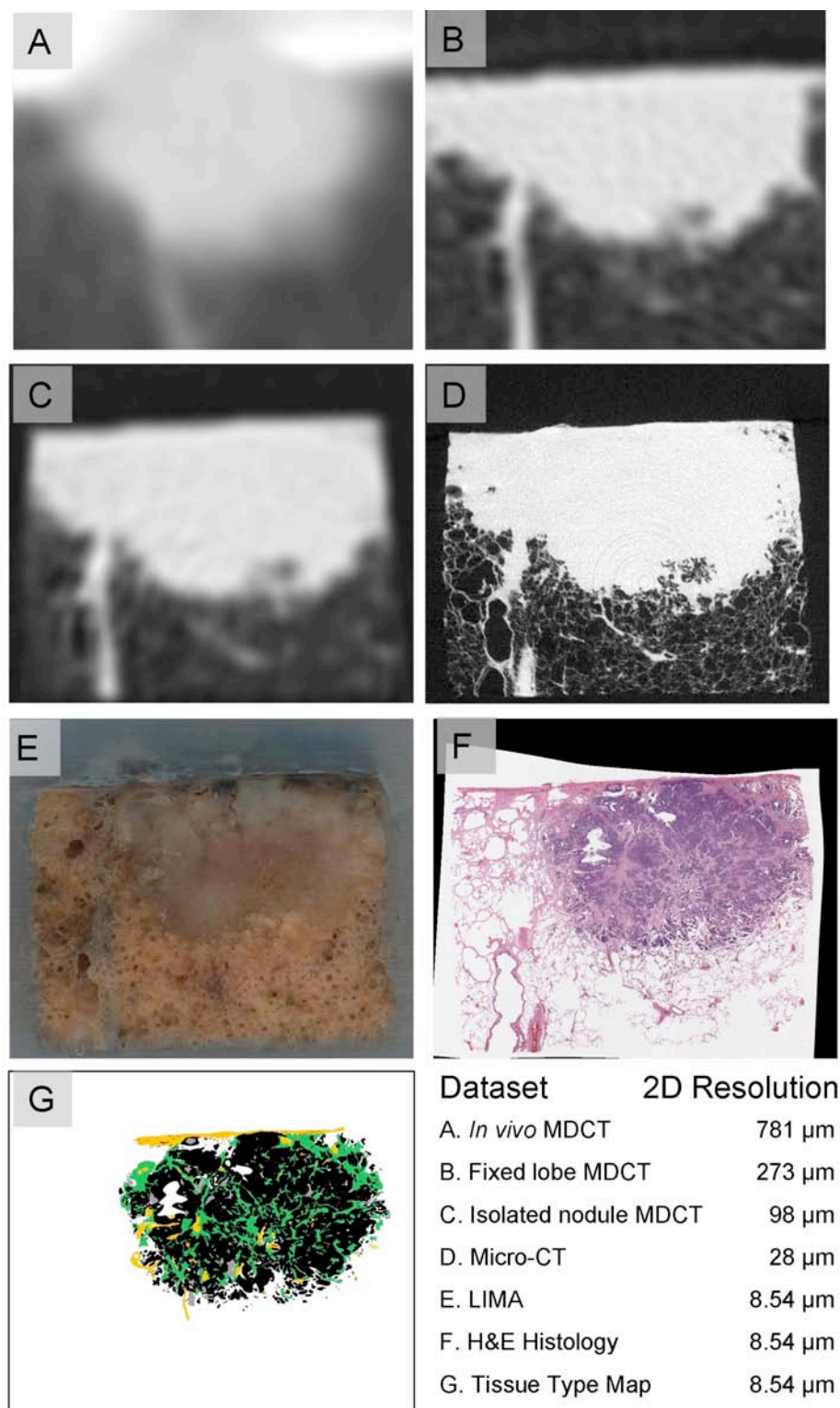
Figure 5-15: Registered neuroendocrine carcinoma. A 2D slice from the volumetric, registered, multi-modal dataset with a tabulated account of the resolutions.
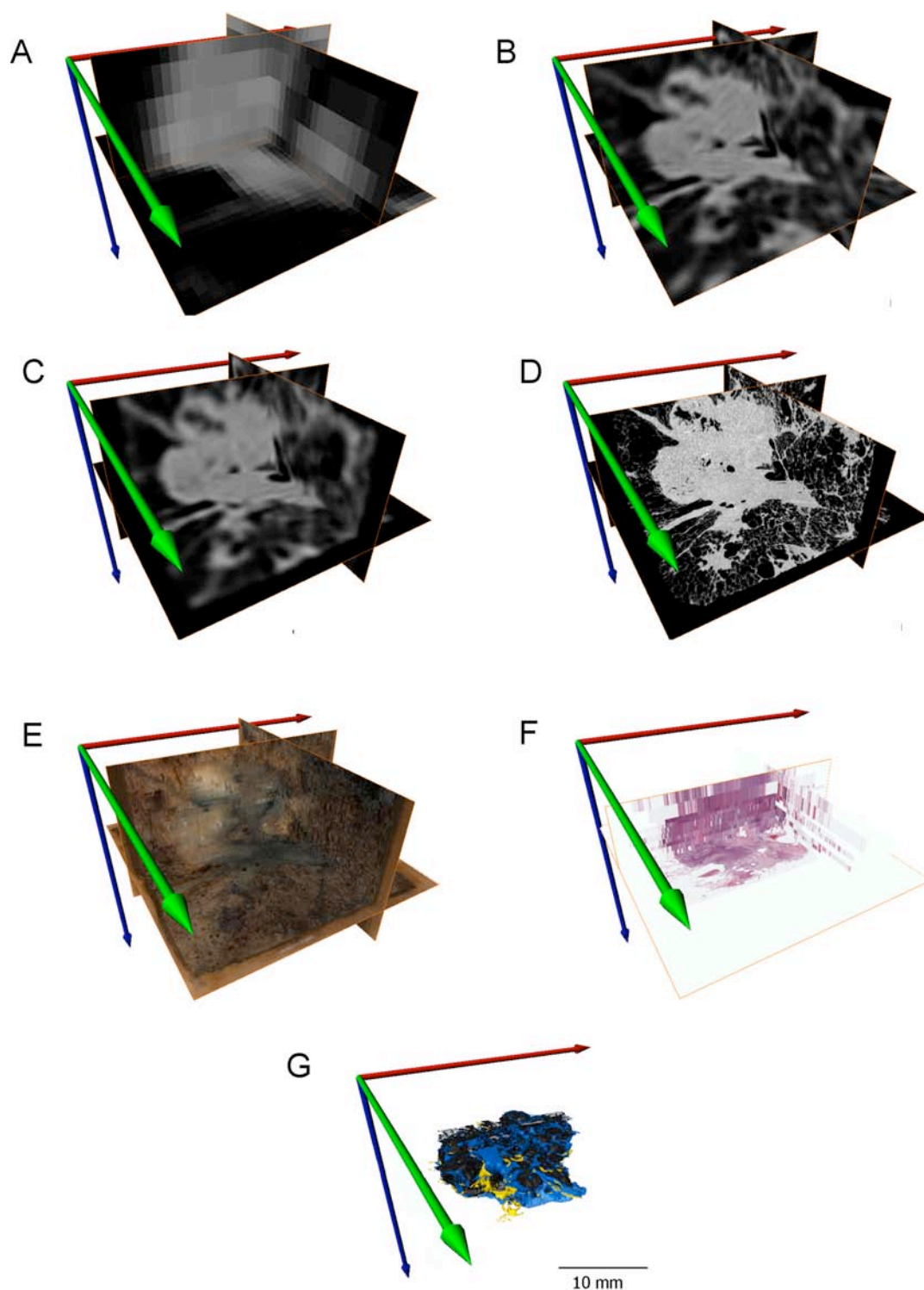
Figure 5-16: A volumetric depiction of the registered adenocarcinoma case 4. This figure illustrates the in vivo MDCT (A), fixed lobe MDCT (B), isolated nodule MDCT (C), micro-CT (D), LIMA (E), histology (F) and tissue type map (G).

CHAPTER 6

ANALYSIS OF THE MULTIMODAL DATASETS

6.1 Lung Nodule Architecture

The segmentation of the histopathological data, generating tissue type maps for each dataset, allowed the quantitative evaluation of the composition of the nodule.

6.1.1 Proportion of Each Tissue Type

The proportions of each tissue type were calculated by summing the number of pixels in a tissue type class and dividing this number by the sum of pixels in all tissue type classes. An average of eleven histopathological sections per nodule were analyzed. **Figure 6-1** features these results, tabulated and presented in a percentage bar graph. A wide range of distributions of tissue type percentages occurred across the dataset. The solid cancerous tumor portion for the adenocarcinoma cases ranged from 33% to 56%. The neuroendocrine carcinoma contained the highest proportion of solid cancerous tumor at 76%. The BAC component of the nodules ranged from 58% to 0% in the adenocarcinoma cases. The two adenocarcinoma cases with the highest total cancer proportion (solid plus BAC) also had the two smallest proportions of active fibrosis.

6.1.2 Three Dimensional Tumor Renderings

The histopathologically obtained tissue type maps were also used to generate 3D reconstructions of the tissue types with in a nodule, **Figure 6-2**. The creation of 3D reconstructions from histopathology data is challenging due to the distortion of the tissue which occurs during the creation of a histology slide. The developed process and registration models allow for the correction of this distortion and the alignment of

sequential histological sections in a global coordinate space. Hence volumetric histopathological datasets were created for each nodule.

Within the histopathological data, an anisotropy existed between the in-plane resolution (8.54 micron) and the separation between subsequent sections (approximately 504 microns). Lanczos resampling was used to decrease the in-plane resolution and cubic interpolation was used to bridge the information between subsequent sections. A 3D surface for each tissue type was then generated via smoothed triangular approximation. The 3D tissue type reconstructions are valuable in visualizing the distribution of the tissue types throughout the nodule volume. These reconstructions contribute towards gaining a comprehensive understanding of the nodules volumetric content and the relationships between tissue boundaries.

6.1.3   Regional Based Analysis

The construction of tissue type maps based on histopathology can not only be used to gain comprehensive estimates of tissue type proportions and 3D component based reconstructions, but may also be used for quantification of the regional properties of each tissue type.

The histological tissue type maps were split into a series of binary masks, one for each tissue type. Connected component analysis was conducted to label each region of each tissue type. For each region, a number of shape descriptors were calculated and these values were tabulated for analysis. Shape descriptors are widely used to simplify the description of complex objects. This is achieved by assigning a numerical value to represent a property of the object's shape. Area and perimeter are examples of simple shape descriptors.

Across the adenocarcinoma cases, the solid cancerous tumor and active fibrosis tissue types had the highest number of regions, 2299 and 2260 respectively. The average size of these regions however, was quite different with the solid cancerous tumor having a mean region size of 1.7 mm^2 while the active fibrosis had a mean region size of 0.17 mm^2. The number of regions for necrosis, inactive fibrosis, cancerous tumor (BAC) and red blood cells were 509, 448, 216 and 193 respectively. **Figure 6-3** shows the average areas of the tissue type regions.

The Euler number is a shape descriptor which is indicative of the number of holes with in a region. It is calculated by subtracting the number of holes in a region from the number of connected components [89]. For understanding the architecture of lung nodules, trends in the Euler number can be insightful for determining which tissue types are generally encompassing and which are encompassed. A structure that is encompassing of other regions would have a high number of holes and hence a very negative Euler number. Structures that exist primarily as 'islands' in an encompassing 'ocean' would have a Euler number of zero or above.

A plot of the average range of Euler numbers for each tissue type, across all datasets is shown in **Figure 6-4**. The plot clearly shows that the inactive fibrosis tissue type is the most encompassing and that necrotic and blood regions exist exclusively as islands (as seen by the global minimum Euler number). Solid cancerous tumor also presents as being more encompassing overall than active fibrosis and non-solid, BAC cancerous tumor.

Area to perimeter ratios have been commonly used to describe the boundary of an object. Compactness is a shape measure which represents the spread of a region and it is

calculated as the area times 4π divided by the squared perimeter [90]. A circular object has the maximum compactness value of one. Shapes with more complex boundaries have a compactness value less than one. An infinitely long and narrow shape would have a compactness value of zero. **Figure 6-5** shows the average compactness for each tissue type across all the nodule datasets. From this data, necrosis and blood regions have the simplest, most compact boundaries with average compactness values of 0.68 and 0.81 respectively. On average the inactive fibrosis regions have been found to contain the most complex boundaries, having the lowest overall compactness value at 0.45. Cancerous regions (both solid and BAC) and active fibrosis had comparable compactness values of 0.55, 0.49 and 0.53 respectively.

6.1.4   Regional Analysis Conclusions

As described, the lung cancer nodule represents a complex biomass. The static organization of histopathological tissue types with in the biomass is potentially informative for cancer biology.

Shape descriptor features were used to evaluate the regional structure of the nodules. Looking at the number and size of regions across the tissue types it was found that the active fibrosis regions were small but numerous. The inactive fibrosis regions were on average of a similar size to the solid cancerous regions, however, many more cancerous regions were present throughout the nodule volumes. BAC cancerous tumor had the largest regional areas of all tissue types.

Using the Euler number, inactive fibrosis was found to be the most encompassing tissue type. Solid cancerous tumor also had a large range of Euler numbers with the second lowest global minimum. One possible explanation for this is that the fibrosis is in

response to cancer regions, with the fibrosis replacing cancer as the tumor biomass develops. The small range and zero global minimum of the Euler numbers of necrotic regions indicated this tissue type does not encompass other tissue type regions. This may be indicative of the mechanism causing necrosis to occur. Were necrosis regions formed by the cell death of only cancerous tissue, we would expect the Euler numbers for necrosis to vary in the same extent as for solid cancerous regions. However, the near zero, very small range of Euler numbers for necrosis indicate that necrotic regions likely occur from a lack of sufficient blood supply to an area of the nodule, resulting in the cell death of all tissue types in that vicinity. As would be expected, groups of red blood cells were strongly presented as encompassed island regions. This is indicative of the blood being contained within vessel walls and separate from other tissue types.

Regions of red blood cells were also found to have the most compact boundaries out of the tissue types. Again, this is intuitive given the containment in relatively cylindrical blood vessels. The relatively high average compactness and small standard error found for necrotic regions suggests that necrosis is a singular event that occurs with a nutrient deficit from perhaps a single critical blood vessel.

The inactive fibrosis tissue type had the lowest average compactness value which may be related to the Euler number findings for this tissue type. The BAC cancerous tumor regions also had relatively low compactness values which are a reflection of the boundary of these regions following a complex alveolar structure. The solid cancerous regions had a compactness value close to the midpoint (0.5) which is likely to represent the random non-directed expansion of cancer cells, some being very complex and others being compact.

6.2 Investigation of Nodule Representation in MDCT

6.2.1 Hounsfield Unit Correlations to Tissue Types

Having described and evaluated the histopathological heterogeneity in tissue types within lung nodules, it was possible to relate this information to the MDCT representations. MDCT data also reveals heterogeneity within lung nodules, in the graylevel, or Hounsfield Unit (HU), **Figure 6-6**, however it is not known if a correlation exists between the two.

Following the registration of the histopathological tissue type maps and the radiological datasets to the common coordinate system, a direct correspondence was established between the voxels in the radiological data and the tissue type labels. This labeling was used to extract out the HU values corresponding to each tissue type and statistically evaluate if a significant difference could be found.

The tissue type map was split into a series of binary masks, one for each tissue type. Each binary mask was resampled, using Lanczos resampling, to match the resolution of the micro-CT, the isolated nodule MDCT dataset and the fixed lobe MDCT dataset. As the *in vivo* MDCT datasets were acquired using different scanning protocols, on different MDCT systems and reconstructed with low resolution using differing reconstruction kernels, this data could not be incorporated.

The binary masks were used to isolate pixels specific to a tissue type, from the radiological data. HU histograms were found for each tissue type and the histogram statistics were collected, including mean, median, standard deviation, skewness and

kurtosis. Ninety values were tabulated for each histological section (6 histopathological tissue types for 3 imaging modalities providing 5 histogram statistics), **Figure 6-7**.

Only the adenocarcinoma cases were evaluated for statistical separation in HU values based on histopathological tissue type. As only one complete dataset for squamous cell and one for neuroendocrine carcinoma were obtained, the datasets were not large enough to calculate the variance for these cancer types.

Linear mixed model analysis was used to compare the mean histogram parameters among the histological tissue types. This was followed by Tukey's test for pairwise comparison of means between tissue types. This analysis was performed separately for each of the imaging methods (micro-CT, isolated nodule MDCT, and fixed lobe MDCT). Figure 6-8 to **Figure 6-12** contain the tabulated and graphed statistical results showing the mean and standard error for each parameter analyzed and *p*-values from Tukey's test.

From this analysis, it was found that the HU heterogeneity in the MDCT data of lung nodules is informative for at least some tissue types. The mean HU measure provided the greatest ability to separate the histopathological tissue classes in the MDCT data of the isolated nodule, Figure 6-8. Statistically significant separability ($p < 0.0001$ to $p = 0.044$) was found between the BAC cancerous tumor (-376.2 HU), red blood cells (-74.7 HU), solid cancerous tissue (-10.9 HU) and inactive fibrosis (25.9 HU) classes. Active fibrosis and necrotic tissue regions could not be distinguished from the solid cancerous tumor using the mean HU from the isolated nodule dataset which were -17.5 HU, -3.4 HU and -10.9 HU respectively.

The BAC cancerous tumor was easily identified as different from the other solid nodule tissue types, due to the non-solid alveolar structural arrangement of the cancer

cells. A very high statistical difference in the mean, median, standard deviation, skewness and kurtosis of the HU values, in the micro-CT, isolated nodule MDCT and fixed lobe MDCT of the BAC tumor was found when compared to the other tissue types ($p < 0.0001$).

Blood could be separated from solid cancerous tissue using the mean HU in all image sets, median HU in the isolated nodule MDCT, HU standard deviation in the isolated nodule MDCT and fixed lobe MDCT, the HU skewness in the isolated nodule dataset and the kurtosis of the fixed lobe MDCT.

Inactive fibrosis could be separated from solid cancerous tissue using the HU standard deviation in all image sets and the HU kurtosis of the micro-CT. Active fibrosis and necrosis were the least separable from solid cancerous tumor. Active fibrosis was only significantly different form solid cancerous tissue in the HU kurtosis of the fixed lobe MDCT. Necrosis was not significantly different from solid cancerous tissue in any of the datasets.

The micro-CT dataset contained the greatest standard error in the HU due to a higher degree of noise within the dataset. This is due to the higher resolution of the micro-CT and also the absence of the same corrections built in with the Siemens Somatom Sensation 64 MDCT system, on which the isolated nodule and fixed lobe MDCT data was acquired. The raw micro-CT data was converted to HU using an air and water phantom to determine the linear correction of grayscale values. While phantoms are also used to calibrate the Siemens Somatom Sensation 64 MDCT system, we are not in control of how this is achieved. This resulted in different HU calibrations between the micro-CT and the MDCT image sets.

To our knowledge, this is the first work which presents a direct, statistically significant correlation between MDCT heterogeneity and histological tissue type.

Tissue Type Proportions
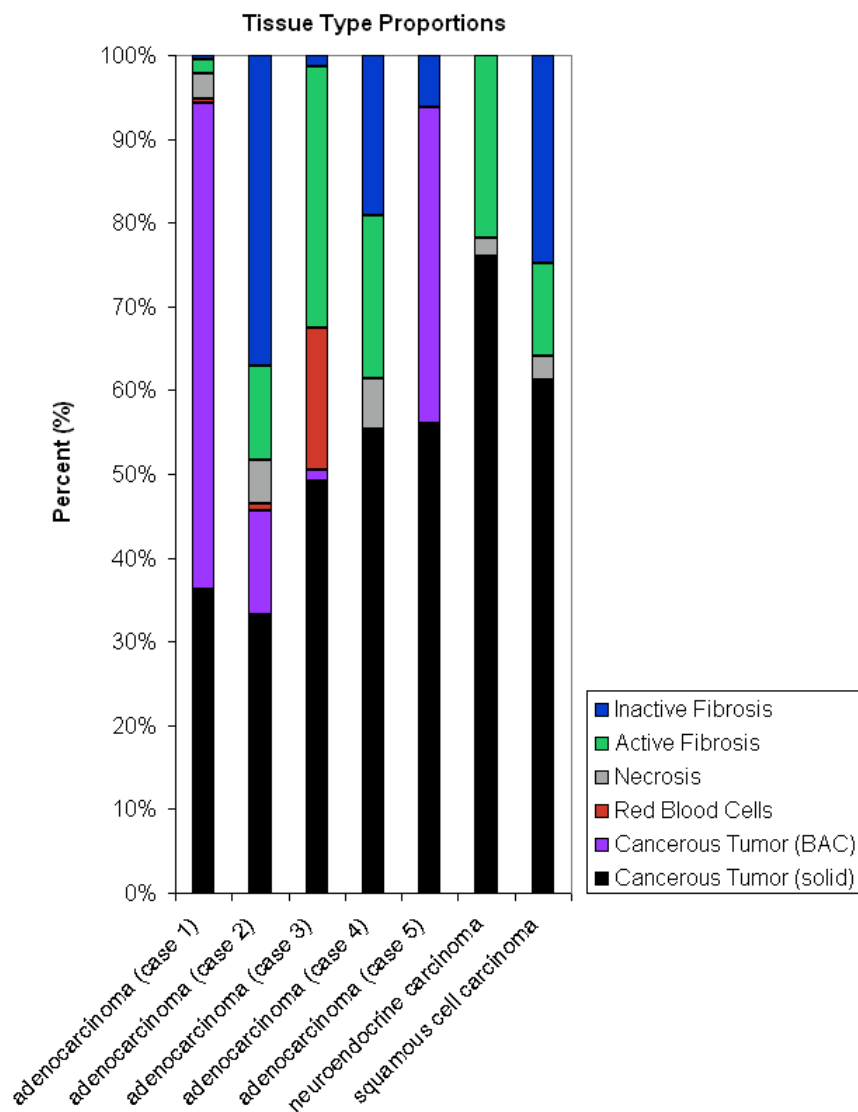
Tissue Type Proportions	Adenocarcinoma					Neuro-endocrine	Squamous Cell
	Case 1	Case 2	Case 3	Case 4	Case 5		
Cancer (solid)	36 %	33 %	49 %	55 %	56 %	76 %	61 %
Cancer (BAC)	58 %	12 %	1 %	0 %	38 %	0 %	0 %
Red Blood Cells	0 %	1 %	17 %	0 %	0 %	0 %	0 %
Necrosis	3 %	6 %	0 %	6 %	0 %	2 %	3 %
Active Fibrosis	2 %	11 %	31 %	20 %	0 %	22 %	11 %
Inactive Fibrosis	1 %	37 %	2 %	19 %	6 %	0 %	25 %

Figure 6-1: Comparison of tissue type proportions. The proportions of each tissue type for the five adenocarcinoma cases and example neuroendocrine and squamous cell carcinomas were determined using the 3D histopathology datasets.
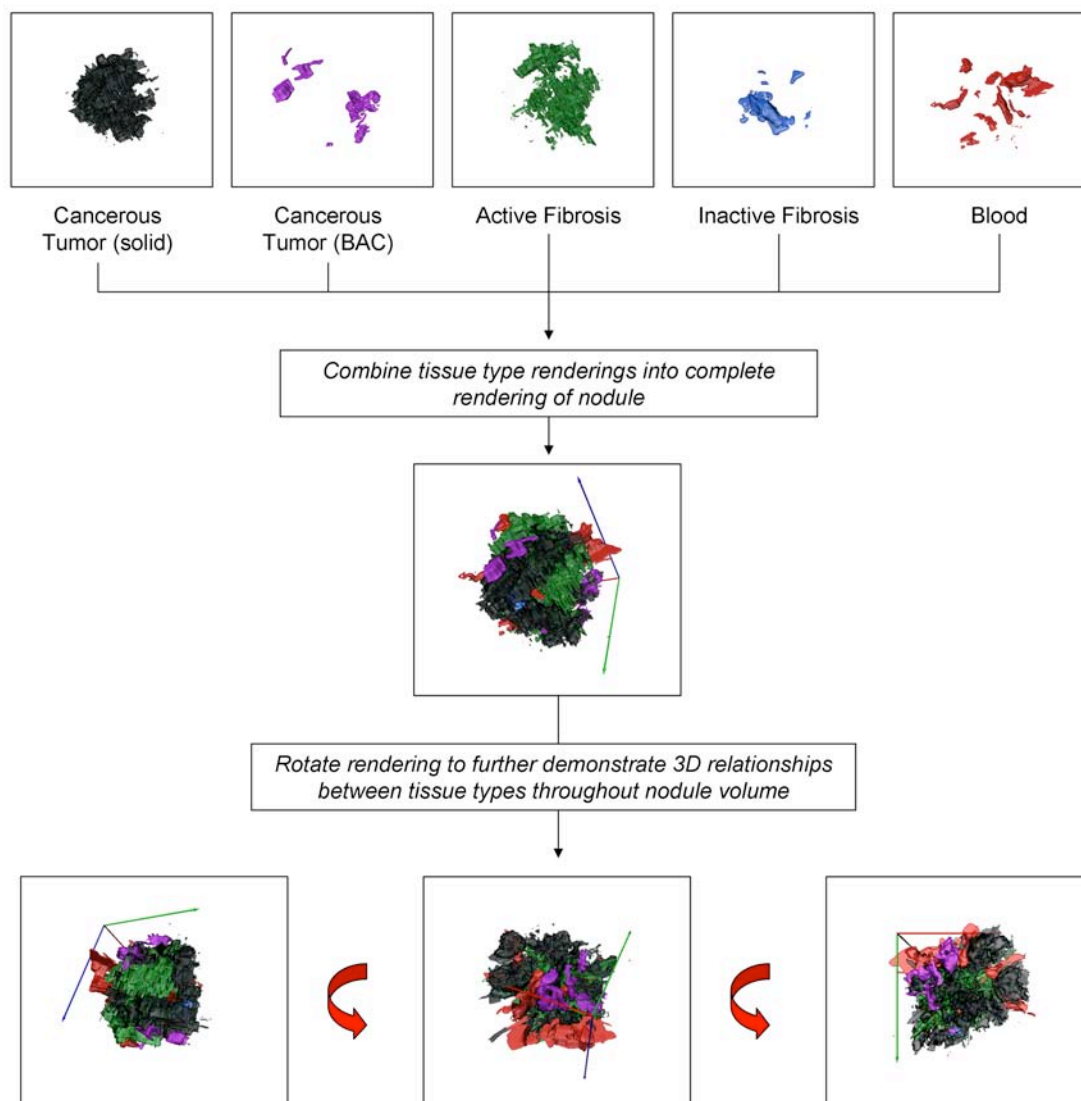
Figure 6-2: The 3D reconstruction of the tissue types with in an adenocarcinoma case. This reconstruction in 3D allows for a more comprehensive understanding of the relationships between the different tissue types and their distribution throughout the nodule volume.

Average Region Area
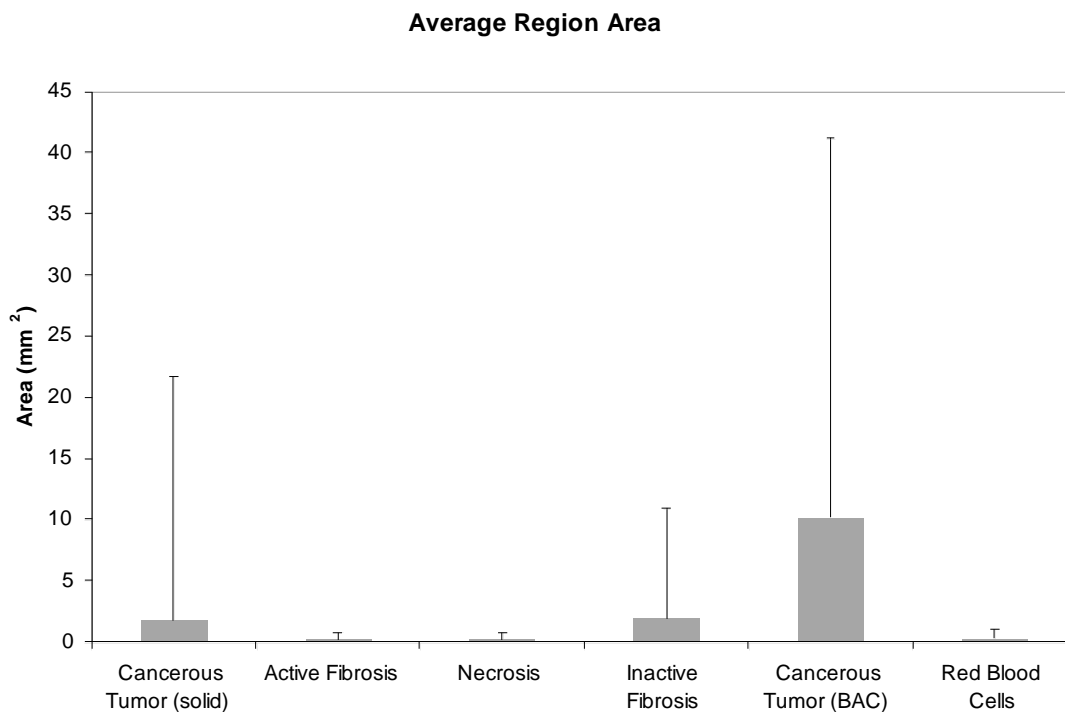
Figure 6-3: The average region areas across the tissue types. This plot shows the average area of the regions for each tissue type (with standard deviation). Regions of red blood cells, necrosis and active fibrosis are the smallest while BAC cancerous regions have the largest region areas.
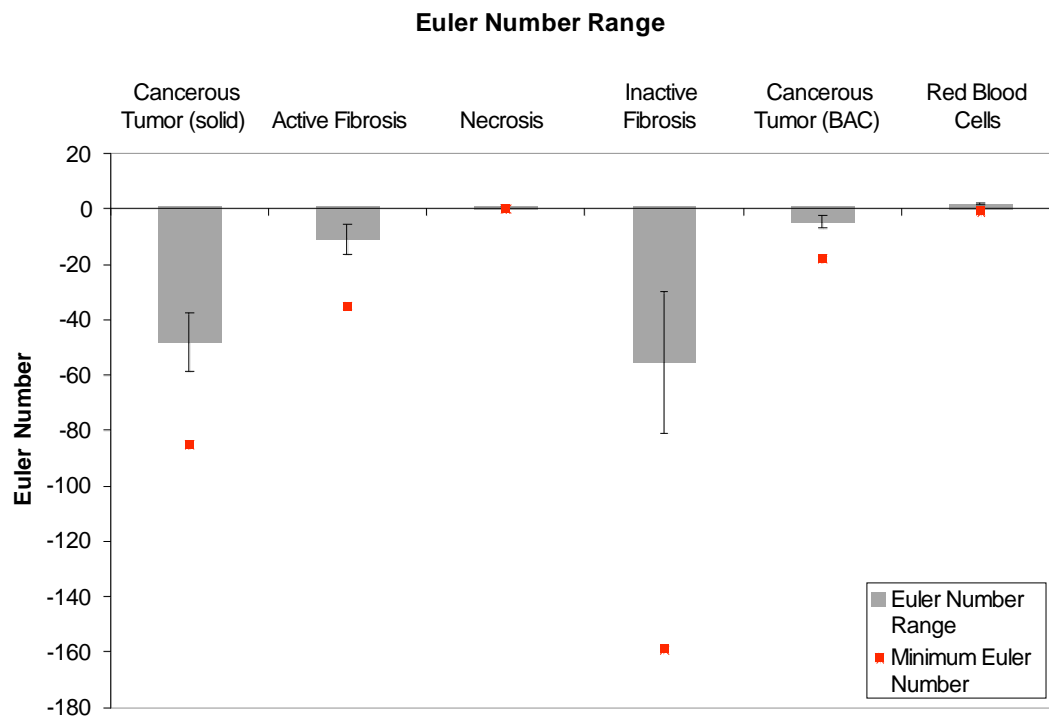
Euler Number Range

Figure 6-4: Euler number trends across the different tissue types. This plot shows the average range of Euler numbers in each tissue type (with standard error bars), across all nodule datasets. Also plotted is the overall global minimum Euler number. This plot reveals the trend that the inactive fibrosis and cancerous tumor tissue types tend to be encompassing while the other tissue types tend to be encompassed.

Average Compactness
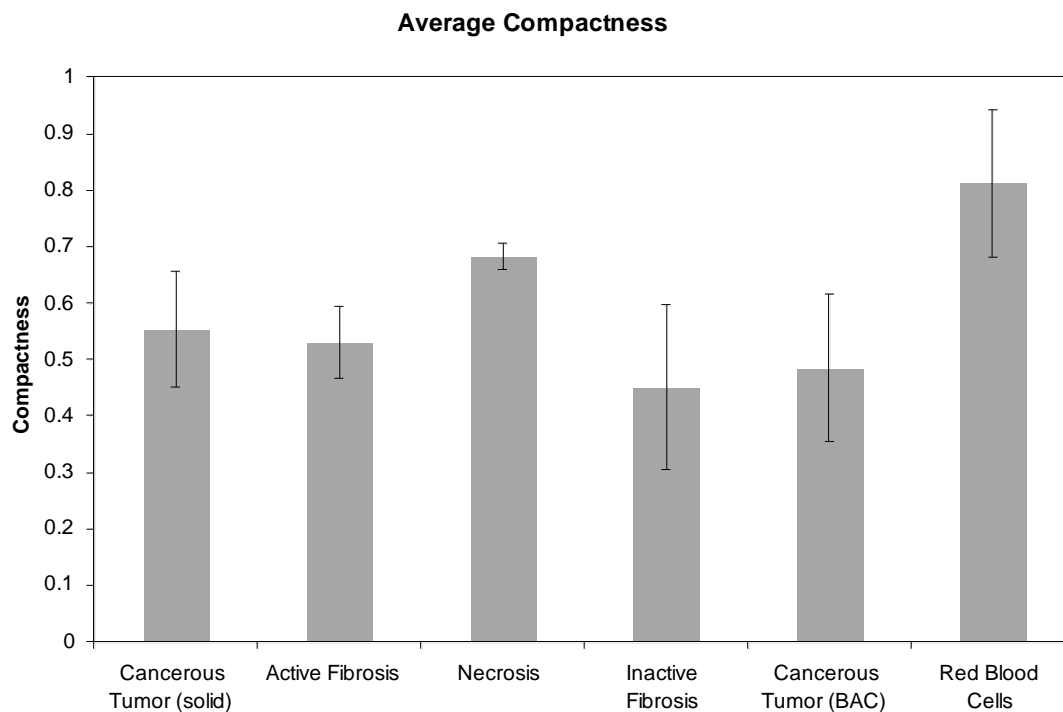
Figure 6-5: Compactness trends across different tissue types. This plot shows the average compactness values for each tissue type (with standard error bars), across all the nodule datasets. The red blood cells and necrosis regions have the most compact boundaries while the inactive fibrosis regions are the least complex.
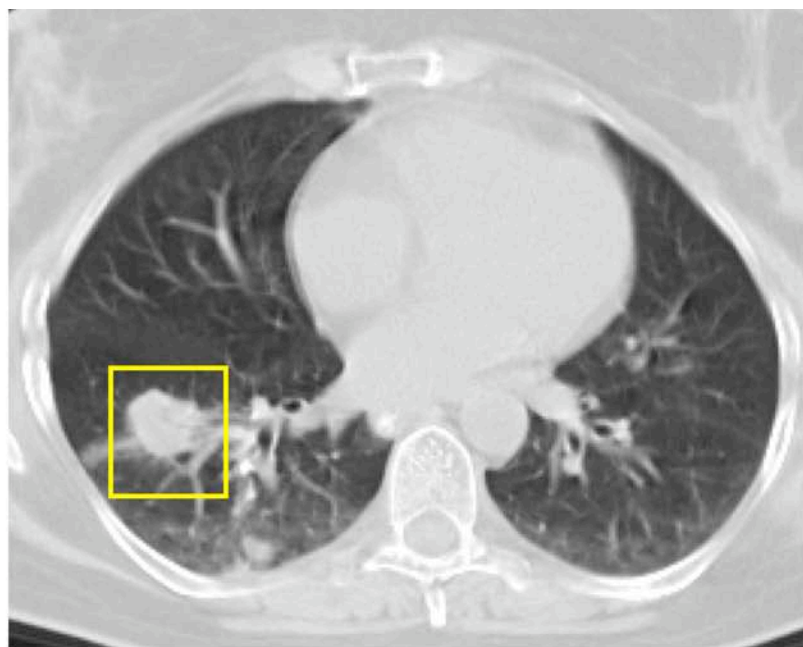
In vivo MDCT
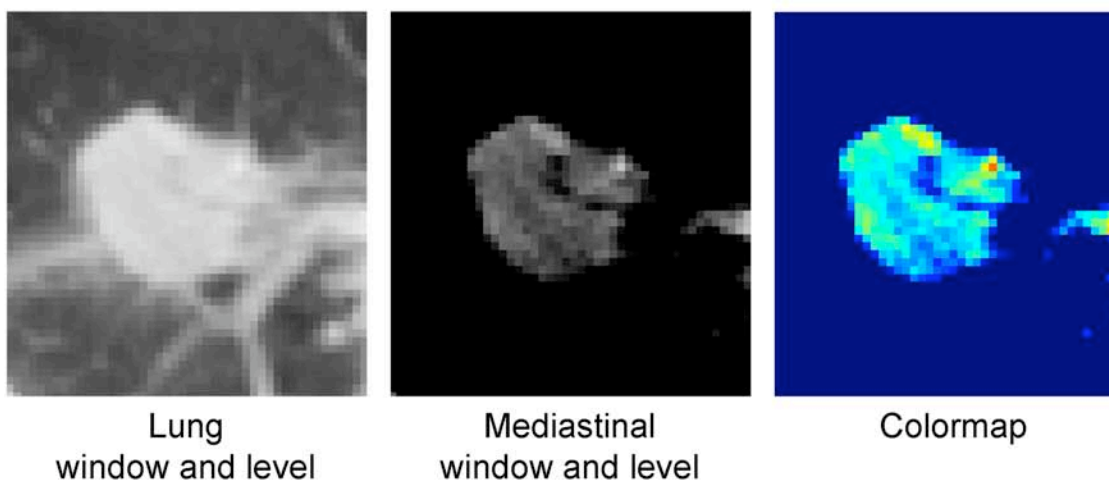
Lung
window and level

Mediastinal
window and level

Colormap

Figure 6-6: An example of the grayscale Hounsfield Unit (HU) heterogeneity in *in vivo* MDCT data of lung nodules. The grayscale heterogeneity is shown using different window and leveling; lung (window width, 1500 HU, window level, -500 HU) and mediastinal (window width, 300 HU, window level, 50 HU). The colormap highlights the difference between the graylevels (red, 350 HU to blue, -250 HU).
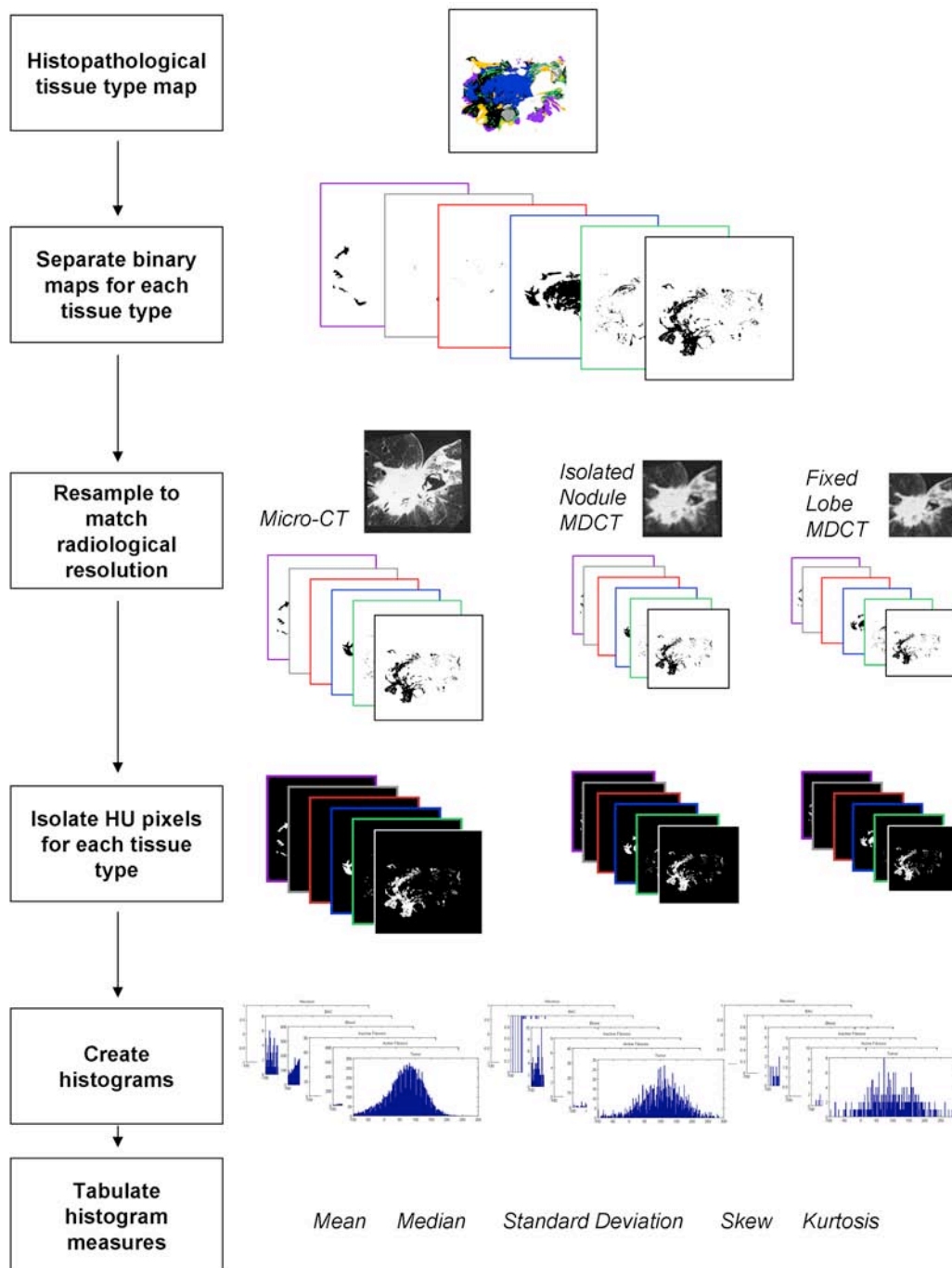
Figure 6-7: Summary diagram for the collection of the 90 histogram statistical measures for each histological section. This involves separating the histopathological tissue type map into 6 binary images, using them to isolate the tissue type specific Hounsfield Unit pixels in 3 CT datasets and collecting 5 histogram statistics from each histogram.

Mean Hounsfield

Mean Hounsfield								
Micro-CT			**MDCT (isolated nodule)**			**MDCT (fixed lobe)**		
Tissue	**Mean (SE)**	**Tukey***	**Tissue**	**Mean (SE)**	**Tukey***	**Tissue**	**Mean (SE)**	**Tukey***
Cancer (BAC)	-353.3 (45.0)	a	Cancer (BAC)	-376.2 (32.9)	a	Cancer (BAC)	-343.6 (35.0)	a
Blood	-60.1 (39.3)	b	Blood	-74.7 (29.1)	b	Necrosis	-59.2 (35.0)	bc
Necrosis	31.2 (36.8)	bc	Active Fibrosis	-17.5 (28.1)	c	Blood	-57.2 (28.4)	b
Active Fibrosis	-0.5 (35.1)	c	Cancer (solid)	-10.9 (27.8)	c	Active Fibrosis	-23.0 (27.7)	bd
Cancer (solid)	8.1 (35.1)	cd	Necrosis	-3.4 (31.7)	cd	Cancer (solid)	-5.3 (26.3)	cd
Inactive Fibrosis	34.7 (35.8)	d	Inactive Fibrosis	25.9 (27.9)	d	Inactive Fibrosis	27.8 (27.3)	d
Significant differences have p-value from p<0.0001 (BAC vs. all other tissue types) to p=0.044 for Blood vs. Cancer (solid)			*Significant differences have p-value from p<0.0001 (BAC vs. all other tissue types) to p=0.017 for Blood vs. Necrosis*			*Significant differences have p-value from p<0.0001 (BAC vs. all other tissue types; Blood vs. Inactive Fibrosis) to p=0.034 for Cancer (solid) vs. Inactive Fibrosis*		

* means with the same letter are not significantly different

Figure 6-8: Comparison of the statistically significant separations between the *mean* Hounsfield Unit for each tissue type in the micro-CT, isolated nodule MDCT and fixed lobe MDCT data.

Median Hounsfield

Median Hounsfield								
Micro-CT			**MDCT (isolated nodule)**			**MDCT (fixed lobe)**		
Tissue	**Mean (SE)**	**Tukey***	**Tissue**	**Mean (SE)**	**Tukey***	**Tissue**	**Mean (SE)**	**Tukey***
Cancer (BAC)	-355.8 (44.8)	a	Cancer (BAC)	-373.5 (34.1)	a	Cancer (BAC)	-361.5 (34.1)	a
Blood	-22.3 (38.6)	b	Blood	-29.6 (29.3)	b	Necrosis	-11.5 (34.4)	b
Active Fibrosis	33.3 (31.7)	c	Necrosis	0.2 (29.0)	bc	Blood	0.6 (26.9)	b
Necrosis	40.3 (31.3)	bc	Active Fibrosis	22.1 (26.4)	c	Active Fibrosis	13.2 (26.8)	b
Cancer (solid)	49.6 (31.3)	bc	Cancer (solid)	31.3 (25.5)	c	Cancer (solid)	32.0 (25.6)	b
Inactive Fibrosis	59.3 (32.5)	bc	Inactive Fibrosis	53.0 (25.7)	c	Inactive Fibrosis	51.6 (26.1)	b
Significant differences have p-value from p<0.0001 (BAC vs. all other tissue types) to p=0.041 for Blood vs. Active Fibrosis			*Significant differences have p-value from p<0.0001 (BAC vs. all other tissue types) to p=0.002 for Blood vs. Cancer (solid)*			*Significant differences have p-value of p<0.0001 (BAC vs. all other tissue types)*		

* means with the same letter are not significantly different

Figure 6-9: Comparison of the statistically significant separations between the *median* **Hounsfield Unit for each tissue type in the micro-CT, isolated nodule MDCT and fixed lobe MDCT data.**

Hounsfield Standard Deviation

Hounsfield Standard Deviation								
Micro-CT			**MDCT (isolated nodule)**			**MDCT (fixed lobe)**		
Tissue	Mean (SE)	Tukey*	Tissue	Mean (SE)	Tukey*	Tissue	Mean (SE)	Tukey*
Cancer (BAC)	260.2 (26.4)	a	Cancer (BAC)	201.1 (19.4)	a	Cancer (BAC)	197.1 (27.7)	ab
Necrosis	233.3 (30.0)	ab	Blood	193.0 (18.3)	a	Blood	195.6 (22.6)	a
Blood	201.5 (28.0)	b	Active Fibrosis	147.8 (16.6)	b	Active Fibrosis	155.2 (21.8)	b
Cancer (solid)	191.9 (25.3)	b	Cancer (solid)	147.3 (16.8)	b	Cancer (solid)	151.0 (21.2)	b
Active Fibrosis	187.0 (25.4)	b	Necrosis	132.9 (22.7)	bc	Necrosis	140.1 (29.2)	abc
Inactive Fibrosis	137.1 (25.7)	c	Inactive Fibrosis	114.5 (17.5)	c	Inactive Fibrosis	107.8 (22.2)	c
Significant differences have p-value from p<0.0001 (BAC vs. Cancer (solid), Active Fibrosis, and Inactive Fibrosis; Necrosis, Cancer (solid) vs. Inactive Fibrosis) to p=0.002 for Blood vs. Inactive Fibrosis			*Significant differences have p-value from p<0.0001 (BAC and Blood vs. Inactive Fibrosis) to p=0.018 for Blood vs. Necrosis*			*Significant differences have p-value from p<0.0001 (Blood vs. Inactive Fibrosis) to p=0.005 for Blood vs. Active Fibrosis*		

* means with the same letter are not significantly different

Figure 6-10: Comparison of the statistically significant separations between the *standard deviation* in the Hounsfield Unit histogram for each tissue type in the micro-CT, isolated nodule MDCT and fixed lobe MDCT data.

Hounsfield Histogram Skew

Hounsfield Histogram Skew								
Micro-CT			**MDCT (isolated nodule)**			**MDCT (fixed lobe)**		
Tissue	**Mean (SE)**	**Tukey***	**Tissue**	**Mean (SE)**	**Tukey***	**Tissue**	**Mean (SE)**	**Tukey***
Cancer (BAC)	-0.16 (0.24)	a	Cancer (BAC)	-0.04 (0.13)	a	Cancer (BAC)	0.03 (0.10)	a
Inactive Fibrosis	-1.50 (0.33)	b	Blood	-1.21 (0.16)	b	Necrosis	-0.94 (0.19)	b
Blood	-1.75 (0.36)	b	Active Fibrosis	-1.43 (0.16)	bc	Blood	-1.13 (0.09)	bc
Necrosis	-1.85 (0.27)	b	Necrosis	-1.43 (0.27)	bc	Active Fibrosis	-1.52 (0.11)	c
Cancer (solid)	-1.99 (0.38)	b	Inactive Fibrosis	-1.56 (0.25)	bc	Inactive Fibrosis	-1.55 (0.26)	bc
Active Fibrosis	-2.26 (0.31)	b	Cancer (solid)	-1.68 (0.14)	c	Cancer (solid)	-1.76 (0.11)	c
Significant differences have p-value from p<0.0001 (BAC vs. Active Fibrosis and Necrosis) to p=0.002 for BAC vs. Blood			*Significant differences have p-value from p<0.0001 (BAC vs. all other Tissue types) to p=0.034 for Blood vs. Cancer (solid)*			*Significant differences have p-value from p<0.0001 (Blood vs. Necrosis, Blood, Active Fibrosis, and Cancer (solid)) to p=0.046 for Necrosis vs. Active Fibrosis*		

* means with the same letter are not significantly different

Figure 6-11: Comparison of the statistically significant separations between the *skew* in the Hounsfield Unit for each tissue type in the micro-CT, isolated nodule MDCT and fixed lobe MDCT data.
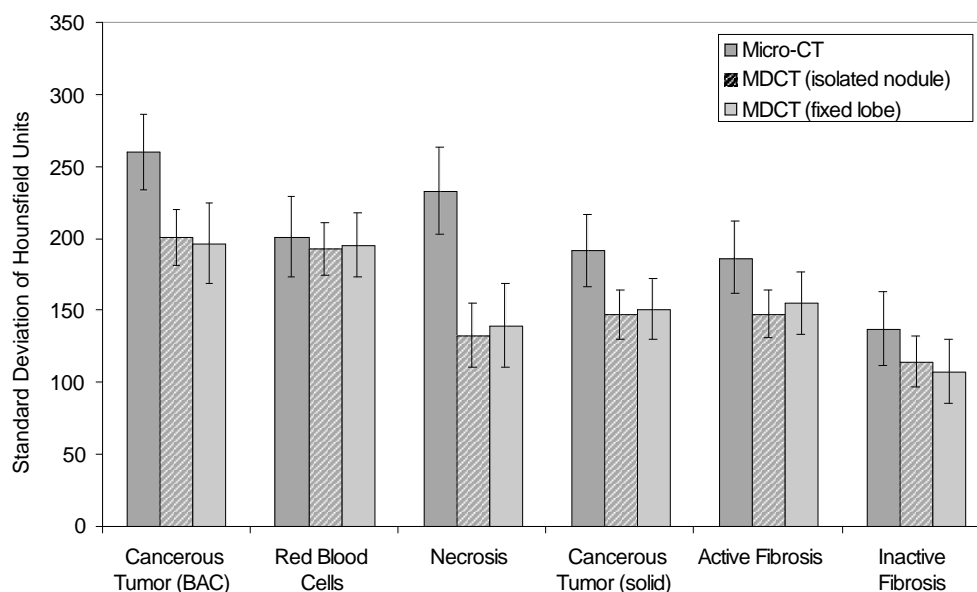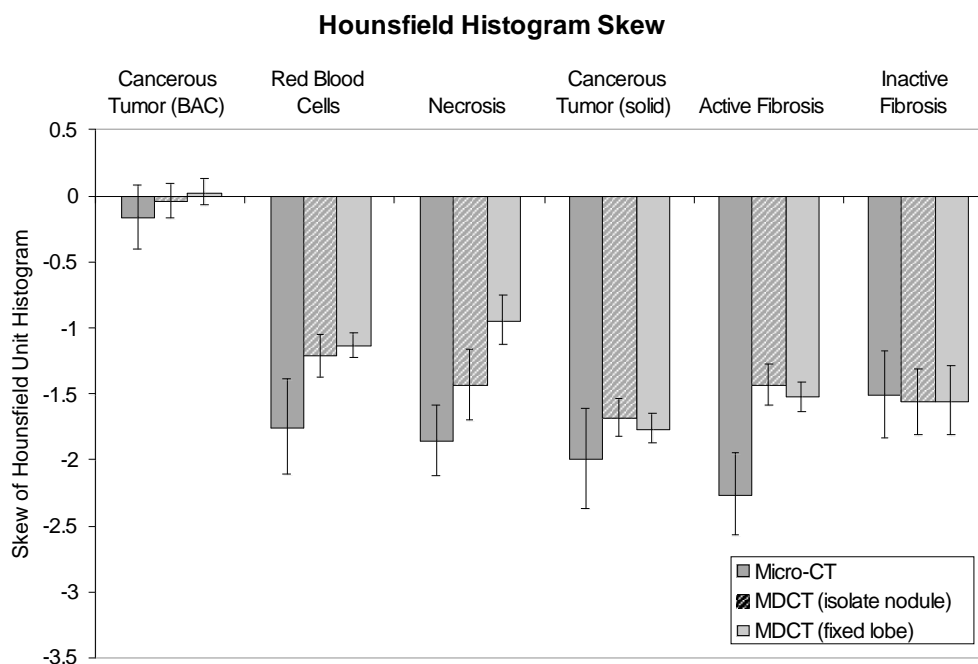
Hounsfield Histogram Kurtosis

Hounsfield Histogram Kurtosis								
Micro-CT			**MDCT (isolated nodule)**			**MDCT (fixed lobe)**		
Tissue	Mean (SE)	Tukey*	Tissue	Mean (SE)	Tukey*	Tissue	Mean (SE)	Tukey*
Cancer (BAC)	2.5 (0.2)	a	Cancer (BAC)	2.3 (0.2)	a	Cancer (BAC)	2.2 (0.1)	a
Necrosis	4.6 (0.4)	b	Blood	4.2 (0.5)	b	Blood	3.8 (0.3)	b
Inactive Fibrosis	5.6 (0.9)	b	Necrosis	5.4 (0.9)	bc	Necrosis	4.3 (0.7)	bcd
Blood	7.6 (1.3)	bc	Inactive Fibrosis	5.6 (1.0)	bc	Active Fibrosis	5.6 (0.5)	c
Cancer (solid)	12.7 (1.4)	cd	Active Fibrosis	5.9 (0.6)	bc	Inactive Fibrosis	5.8 (0.9)	bcd
Active Fibrosis	12.9 (1.5)	d	Cancer (solid)	6.7 (0.6)	c	Cancer (solid)	7.1 (0.5)	d
Significant differences have p-value from p<0.0001 (BAC vs. Blood, Cancer (solid), and Active Fibrosis; Necrosis vs. Cancer (solid) and Active Fibrosis) to p=0.014 for Blood vs. Active Fibrosis			*Significant differences have p-value from p<0.0001 (BAC vs. Blood, Active Fibrosis, Cancer (solid)) to p=0.002 for Blood vs. Cancer (solid)*			*Significant differences have p-value from p<0.0001 (BAC vs. Necrosis, Cancer (solid), and Active Fibrosis; Blood vs. Cancer (solid)) to p=0.031 for Active Fibrosis vs. Cancer (solid)*		

* means with the same letter are not significantly different

Figure 6-12: Comparison of the statistically significant separations between the *kurtosis* in the Hounsfield Unit histogram for each tissue type in the micro-CT, isolated nodule MDCT and fixed lobe MDCT data.

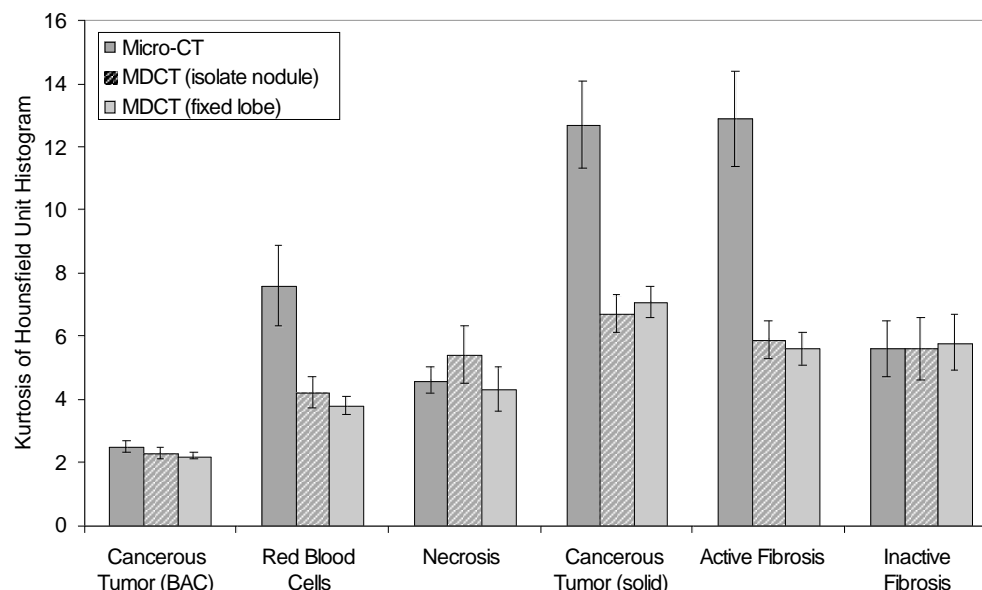# CHAPTER 7

CONCLUSION

7.1   Contributions

It was the purpose of this research to investigate two important hypotheses related to lung nodules:

The first hypothesis was that lung nodules contain both cancerous and non-cancerous regions with a three dimensional complex architecture which is not evident through traditional histological processing. The second hypothesis was that the histopathologically determined tissue types within a lung nodule is related to the nodule's representation in computed tomography imaging, with a correlation between subtle density variations and cancer versus non-cancerous tissue types.

To recall the specific aims of the research study:

- Specific Aim 1: Establish a reliable tissue acquisition, processing and multimodal imaging method.

- Specific Aim 2: The development of an automated classification approach for histopathological data.

- Specific Aim 3: The construction of a registration processing pipeline to register the multimodal image datasets to a common coordinate system.

- Specific Aim 4: The analysis of nodule histopathological tissue type properties including the correlation to MDCT graylevel heterogeneity.

A successful tissue acquisition, processing and multimodal imaging method for the acquisition of lung nodule data has been developed and described in Chapter 3. This process incorporates the use of a unique sectioning and imaging system, the LIMA,

which maintains the spatial correspondence between non-destructive radiological imaging and histopathological processing. This developed methodology was utilized to acquire complete and comprehensive datasets for five adenocarcinoma cases as well as an example neuroendocrine and squamous cell carcinoma case (Aim 1).

A multi-step immunohistochemical staining approach has been identified to provide contrast between the tissue types of interest in lung nodules. Using this immunohistochemical staining, an automated segmentation approach was developed (Chapter 4). The k-means clustering approach was capable of generating tissue type segmentations in under a minute with comparable accuracy, sensitivity and specificity to inter-observer values (Aim 2)

The issue of establishing a reliable basis for the non-rigid registration of histopathological data to non-destructive datasets has been overcome with the LIMA system, which served as the basis for the developed registration approach (Chapter 5). Hence for the first time, labeled histopathology data could be related, on a pixel to pixel basis, to radiological datasets (Aim 3).

Chapter 6 described the analysis of the histopathological data (Aim 4). This involved the comparison of tissue type proportions across the nodules datasets. Regional shape analysis was also conducted to provide quantitative data for the investigation of the lung nodule architecture. It was demonstrated that the registered histopathological tissue type maps could be used to generate volumetric reconstructions for the further understanding of the tissue type composition of the lung nodule biomass (first hypothesis).

Also presented in Chapter 6 was the statistical evaluation of the micro-CT, isolated nodule MDCT, and fixed lobe MDCT tissue specific histograms. Statistical evidence was found that a significant correlation does exist between MDCT graylevels and histopathologically defined tissue types (second hypothesis).

7.2   Limitations of the Research

There are a number of pertinent comments with regard to research in human subjects, including; the number of suitable cases, obtaining consent for pre-surgical MDCT scanning and the time required to acquire and process the data. This study required lobectomy specimens such that imaging and fixation could be achieved in an inflated state. At the projects conception lobectomy procedures were commonly performed at the University of Iowa for the removal of lung nodules, however a shift towards wedge resections has occurred over recent years. Currently at the University of Iowa a high number of nodules are resected via a wedge resection procedure and a full lobectomy is performed only if the surgical pathology diagnosis is positive for cancer. This approach is beneficial for patients as the greatest possible amount of lung tissue is retained, but it does have a negative affect on patient recruitment for this study.

The less than ideal quality of the clinically obtained *in vivo* MDCT data prevented the investigation of the statistical separation between tissue types in this dataset. It was known that the slice thickness for the clinical MDCT data would be approximately 3mm. What was not expected, however, was the large range in scanning protocols used clinically, even within the same institution.

IRB approval was sought and received to conduct, high resolution, perfusion MDCT scanning of the patient prior to surgery. Unfortunately, of the suitable patients identified for participation in this research study, all have been unwilling or unable to undergo an additional pre-surgical, perfusion MDCT scanning procedure. In a number of cases, patients expressed an unwillingness to undergo a PICC line placement as the reason for not consenting to pre-surgical imaging.

This multimodal study is a highly detailed, labor intensive process. The preparation of the tissue and acquisition of the image sets takes over 230 hours to complete. This time estimate is just to obtain the image data and does not include the time required for registration or analysis. When this is coupled with the issue of having no control over when cases become available, personal and equipment scheduling conflicts can arise and become a further restrictive issue.

7.3   Future Work

Future directions for the project would be to improve the pre-surgical clinical imaging by encouraging thinner slice reconstructions and the use of standardized protocols with in a single institution. Perfusion imaging should still be pursued but an alternative less invasive option could increase the number of subjects consenting to pre-surgical imaging. Dual energy MDCT imaging would be another potential approach for investigation.

We have established that statistically significant correlations exist between CT graylevel values and histopathologically determined tissue type. As a future direction of the research, it needs to be determined if this separability of the CT based data is

achievable using the histopathologically determined tissue type maps only for validation. That is, the development of a classifier which is trained using the segmented CT datasets and tested on un-segmented CT data. It is possible that incorporating texture based features as well as graylevel statistics could be highly valuable in this task.

The future expansion of the study to include other forms of pre-surgical imaging such as PET and also possibly MRI would be highly valuable. PET/CT systems are being increasingly incorporated into clinical practice for the evaluation of identified lung nodules [91]. Currently a single standard uptake value (SUV) with a cutoff of 2.5 or greater, is used to indicate malignancy [92]. However, a broad range of SUV values are found in the clinical arena. Higher SUV values have been associated with a decrease in survival for lung cancer patients [93]. Using the designed processed model presented, it would be possible to investigate if a correlation exists between SUV values and proportion of cancerous to non-cancerous nodule tissue.

Improvements may be made to the automated histopathology segmentation technique so that necrotic and normal airway and vessel wall tissue could be distinguished from fibrosis. This may require adjustments to the staining protocol. Alternatively a post-processing step may be added which further classed the image based on regional shape features and relationships.

The designed process model for the creation of multi-modal, volumetric datasets has extraordinary potential in exploring the structure of nodules in other organ systems.

The continued collection of lung nodule cases should also involve the tracking and incorporation of patient outcomes over time. With this, the link between patient

outcome and specific tissue type proportions with in the nodule could for the first time be quantitatively evaluated.

These comprehensive nodule datasets are able to provide a more specific ground truth to computer aided diagnosis developers, than the currently available 'malignant' versus 'benign' nodule diagnosis. It is the intention to make this dataset accessible to the public via future incorporation into the Lung Imaging Database Consortium (LIDC) dataset [94].

REFERENCES

[1]   E. Weibel and L. Cruz-Orive, "Morphometric methods," in *The Lung*, Crystal, West, Weibel, and Barnes, Eds., 2nd ed. Philadelphia: Lippincott-Raven, 1997.

[2]   R. R. Mercer, M. L. Russell, and J. D. Crapo, "Alveolar septal structure in different species," *J Appl Physiol*, vol. 77, pp. 1060-6, 1994.

[3]   M. Ochs, J. R. Nyengaard, A. Jung, L. Knudsen, M. Voigt, T. Wahlers, J. Richter, and H. J. Gundersen, "The number of alveoli in the human lung," *Am J Respir Crit Care Med*, vol. 169, pp. 120-4, 2004.

[4]   J. H. Austin, N. L. Muller, P. J. Friedman, D. M. Hansell, D. P. Naidich, M. Remy-Jardin, W. R. Webb, and E. A. Zerhouni, "Glossary of terms for CT of the lungs: recommendations of the Nomenclature Committee of the Fleischner Society," *Radiology*, vol. 200, pp. 327-31, 1996.

[5]   D. Ost, A. M. Fein, and S. H. Feinsilver, "Clinical practice. The solitary pulmonary nodule," *N Engl J Med*, vol. 348, pp. 2535-42, 2003.

[6]   R. Tsuchiya, "Implication of the CT characteristics of subcentimeter pulmonary nodules," *Semin Thorac Cardiovasc Surg*, vol. 17, pp. 107-9, 2005.

[7]   T. Ohtsuka, H. Nomori, H. Horio, T. Naruke, and K. Suemasu, "Radiological examination for peripheral lung cancers and benign nodules less than 10 mm," *Lung Cancer*, vol. 42, pp. 291-6, 2003.

[8]   K. Kishi, S. Homma, A. Kurosaki, N. Motoi, T. Kohno, K. Nakata, and K. Yoshimura, "Small lung tumors with the size of 1cm or less in diameter: clinical, radiological, and histopathological characteristics," *Lung Cancer*, vol. 44, pp. 43-51, 2004.

[9]   C. I. Henschke, D. F. Yankelevitz, R. Mirtcheva, G. McGuinness, D. McCauley, and O. S. Miettinen, "CT screening for lung cancer: frequency and significance of part-solid and nonsolid nodules," *AJR Am J Roentgenol*, vol. 178, pp. 1053-7, 2002.

[10]   M. Noguchi, A. Morikawa, M. Kawasaki, Y. Matsuno, T. Yamada, S. Hirohashi, H. Kondo, and Y. Shimosato, "Small adenocarcinoma of the lung. Histologic characteristics and prognosis," *Cancer*, vol. 75, pp. 2844-52, 1995.

[11]   W. D. Travis, K. Garg, W. A. Franklin, Wistuba, II, B. Sabloff, M. Noguchi, R. Kakinuma, M. Zakowski, M. Ginsberg, R. Padera, F. Jacobson, B. E. Johnson, F. Hirsch, E. Brambilla, D. B. Flieder, K. R. Geisinger, F. Thunnisen, K. Kerr, D. Yankelevitz, T. J. Franks, J. R. Galvin, D. W. Henderson, A. G. Nicholson, P. S. Hasleton, V. Roggli, M. S. Tsao, F. Cappuzzo, and M. Vazquez, "Evolving concepts in the pathology and computed tomography imaging of lung

adenocarcinoma and bronchioloalveolar carcinoma," *J Clin Oncol*, vol. 23, pp. 3279-87, 2005.

[12]   L. T. Tanoue, "Staging of non-small cell lung cancer," *Semin Respir Crit Care Med*, vol. 29, pp. 248-60, 2008.

[13]   O. Lababede, M. A. Meziane, and T. W. Rice, "TNM staging of lung cancer: a quick reference chart," *Chest*, vol. 115, pp. 233-5, 1999.

[14]   J. Minna, *Neoplasms of the lung*, 16th ed, 2005.

[15]   C. F. Mountain, "Revisions in the International System for Staging Lung Cancer," *Chest*, vol. 111, pp. 1710-7, 1997.

[16]   M. Dollinger, E. Rosenbaum, M. Tempro, and S. Mulvhill, *Everyone's Guide to Cancer Therapy*, 4th ed: Andrews McMeel, 2002.

[17]   B. M. Fischer, J. Mortensen, and L. Hojgaard, "Positron emission tomography in the diagnosis and staging of lung cancer: a systematic, quantitative review," *Lancet Oncol*, vol. 2, pp. 659-66, 2001.

[18]   J. F. Vansteenkiste and S. G. Stroobants, "The role of positron emission tomography with 18F-fluoro-2-deoxy-D-glucose in respiratory oncology," *Eur Respir J*, vol. 17, pp. 802-20, 2001.

[19]   M. Gould, C. Maclean, W. Kuschner, C. Rydzak, and D. Owens, "Accuracy of Positron Emission Tomography for Diagnosis of Pulmonary Nodules and Mass Lesions: A Meta-Analysis," *JAMA*, vol. 285, pp. 936-7, 2001.

[20]   T. R. Church, "Chest radiography as the comparison for spiral CT in the National Lung Screening Trial," *Acad Radiol*, vol. 10, pp. 713-5, 2003.

[21]   D. M. Xu, H. Gietema, H. de Koning, R. Vernhout, K. Nackaerts, M. Prokop, C. Weenink, J. W. Lammers, H. Groen, M. Oudkerk, and R. van Klaveren, "Nodule management protocol of the NELSON randomised lung cancer screening trial," *Lung Cancer*, vol. 54, pp. 177-84, 2006.

[22]   B. Karacali, A. P. Vamvakidou, and A. Tozeren, "Automated recognition of cell phenotypes in histology images based on membrane- and nuclei-targeting biomarkers," *BMC Med Imaging*, vol. 7, pp. 7, 2007.

[23]   S. J. Swensen, "Screening for cancer with computed tomography," *Bmj*, vol. 326, pp. 894-5, 2003.

[24]   L. L. Humphrey, S. Teutsch, and M. Johnson, "Lung cancer screening with sputum cytologic examination, chest radiography, and computed tomography: an update for the U.S. Preventive Services Task Force," *Ann Intern Med*, vol. 140, pp. 740-53, 2004.

[25]   D. Wormans, F. Beyer, S. Diederich, K. Ludwig, and W. Heindel, "Diagnositc performance of a commercially available computer-aided diagnosis system for automatic detection of pulmonary nodules: comparison with single and double reading," *Fortschr Rontgenstr*, vol. 176, pp. 953-958, 2004.

[26]   K. Awai, K. Murao, A. Ozawa, M. Komi, H. Hayakawa, S. Hori, and Y. Nishimura, "Pulmonary Nodules at Chest CT:Effect of computer-aided diagnosis on radiologists' detection performance," *Thoracic Imaging*, vol. 230, pp. 347-352, 2004.

[27]   S. K. Shah, M. F. McNitt-Gray, S. R. Rogers, J. G. Goldin, R. D. Suh, J. W. Sayre, I. Petkovska, H. J. Kim, and D. R. Aberle, "Computer aided characterization of the solitary pulmonary nodule using volumetric and contrast enhancement features," *Acad Radiol*, vol. 12, pp. 1310-9, 2005.

[28]   R. Uppaluri, T. Mitsa, M. Sonka, E. A. Hoffman, and G. McLennan, "Quantification of pulmonary emphysema from lung computed tomography images," *American Journal of Respiratory & Critical Care Medicine*, vol. 156, pp. 248-54, 1997.

[29]   R. Uppaluri, E. A. Hoffman, M. Sonka, G. W. Hunninghake, and G. McLennan, "Interstitial lung disease: A quantitative study using the adaptive multiple feature method," *American Journal of Respiratory & Critical Care Medicine*, vol. 159, pp. 519-25, 1999.

[30]   Y. Xu, M. Sonka, G. McLennan, J. Guo, and E. A. Hoffman, "MDCT-based 3-D Texture Classification of Emphysema and Early Smoking Related Lung Pathologies," *IEEE Transaction on medical imaging*, vol. 25, pp. 464-475, 2006.

[31]   A. Sasov and D. Van Dyck, "Desktop X-ray Microscopy and microtomography," *J Microsc*, vol. 191, pp. 151-158, 1998.

[32]   D. Sasov, "Full-body in-vivo Micro-CT system with 10 microns resolution.," presented at Proceedings of the High Resolution Meeting, Washington, DC, 2001.

[33]   S. Kennel, S. Davis, J. Branning, H. Pan, G. Kabalka, and M. Paulus, "High resolution computed tomography and MRI for monitoring lung tumor growth in mice undergoing radioimmunotherapy: Correlation with histology," *Med. Phys*, vol. 27, pp. 1101-1107, 2000.

[34]   E. L. Ritman, "Molecular imaging in small animals--roles for micro-CT," *J Cell Biochem Suppl*, vol. 39, pp. 116-24, 2002.

[35]   S. Jorgensen, O. Demirkaya, and E. L. Ritman, "Three-dimensional imaging of vasculature and parenchyma in intact rodent organs with X-ray micro-CT," *Am J Physiol*, vol. 275, pp. H1103-14, 1998.

[36]   E. Namati, D. Chon, J. Thiesse, E. A. Hoffman, J. de Ryk, A. Ross, and G. McLennan, "In vivo micro-CT lung imaging via a computer-controlled intermittent iso-pressure breath hold (IIBH) technique," *Phys Med Biol*, vol. 51, pp. 6061-75, 2006.

[37]   J. Thiesse, J. de Ryk, S. Bond, J. Vislisel, E. A. Hoffman, E. Nixon, J. Reinhardt, R. A., and G. McLennan, "Assessment of mouse lung fixation pressures using computed tomography," presented at American Thoracic Society, San Diego, CA, USA, 2005.

[38]   J. Thiesse, J. Reinhardt, J. de Ryk, E. Namati, J. Leinen, W. Recheis, E. A. Hoffman, and G. McLennan, "Three-dimensional visual truth of the normal airway tree for use as a quantitative comparison to micro-CT reconstructions," *Proceedings of SPIE: Medical Imaging*, vol. 5746, pp. 50-54, 2005.

[39]   S. Shofer, C. Badea, S. Auerbach, D. A. Schwartz, and G. A. Johnson, "A micro-computed tomography-based method for the measurement of pulmonary compliance in healthy and bleomycin-exposed mice," *Exp Lung Res*, vol. 33, pp. 169-83, 2007.

[40]   C. H. Chang, M. L. Jan, K. H. Fan, H. E. Wang, T. H. Tsai, C. F. Chen, Y. K. Fu, and T. W. Lee, "Longitudinal evaluation of tumor metastasis by an FDG-microPet/microCT dual-imaging modality in a lung carcinoma-bearing mouse model," *Anticancer Res*, vol. 26, pp. 159-66, 2006.

[41]   Y. Wang, S. Narayanan, J. Liu, D. Shu, A. Mashayekhi, J. Qian, and J. Wang, "A sagittally focusing double-multilayer monochromator for ultrafast X-ray imaging applications," *J Synchrotron Radiat*, vol. 14, pp. 138-43, 2007.

[42]   T. Takeda, A. Momose, K. Hirano, S. Haraoka, T. Watanabe, and Y. Itai, "Human carcinoma: early experience with phase-contrast X-ray CT with synchrotron radiation--comparative specimen study with optical microscopy," *Radiology*, vol. 214, pp. 298-301, 2000.

[43]   W. A. Franklin, "Diagnosis of lung cancer: pathology of invasive and preinvasive neoplasia," *Chest*, vol. 117, pp. 80S-89S, 2000.

[44]   *Immunohistochemistry II*, vol. 14. Chichester, New York: Wiley, 1993.

[45]   C. Taylor, *Immunomicroscopy: a diagnostic tool for the surgical pathologist*. Philadelphia: Saunders, 1994.

[46]   V. Marson, J. Mazieres, O. Groussard, O. Garcia, J. Berjaud, M. Dahan, P. Carles, and G. Daste, "Expression of TTF-1 and cytokeratins in primary and secondary epithelial lung tumors: correlation with histological type and grade," *Histopathology*, vol. 45, pp. 1250134, 2004.

[47]   P. Chu, E. Wu, and L. M. Weiss, "Cytokeratin 7 and cytokeratin 20 expression in epithelial neoplasms: a survey of 435 cases," *Mod Pathol*, vol. 13, pp. 962-72, 2000.

[48]   M. Scarpatetti, O. Tsybrovskyy, and H. H. Popper, "Cytokeratin typing as an aid in the differential diagnosis of primary versus metastatic lung carcinomas, and comparison with normal lung," *Virchows Arch*, vol. 440, pp. 70-76, 2002.

[49]   N. Pendleton, N. L. Occleston, M. J. Walshaw, J. A. Littler, C. I. Jack, M. W. Myskow, and J. A. Green, "Simple cytokeratins in the serum of patients with lung cancer: relationship to cell death," *Eur J Cancer*, vol. 30A, pp. 93-6, 1994.

[50]   G. A. Blobel, R. Moll, W. W. Franke, and I. Vogt-Moykopf, "Cytokeratins in normal lung and lung carcinomas. I. Adenocarcinomas, squamous cell carcinomas and cultured cell lines," *Virchows Arch B Cell Pathol Incl Mol Pathol*, vol. 45, pp. 407-29, 1984.

[51]   D. F. Yankelevitz, A. P. Reeves, W. J. Kostis, B. Zhao, and C. I. Henschke, "Small pulmonary nodules: volumetrically determined growth rates based on CT evaluation," *Radiology*, vol. 217, pp. 251-6, 2000.

[52]   D. Wormanns, G. Kohl, E. Klotz, A. Marheine, F. Beyer, W. Heindel, and S. Diederich, "Volumetric measurements of pulmonary nodules at multi-row detector CT: in vivo reproducibility," *Eur Radiol*, vol. 14, pp. 86-92, 2004.

[53]   S. Kikuchi, Y. Hiki, Y. Sakakibara, A. Kakita, and S. Kuwao, "Measuring the tumor volume of gastric carcinoma by computer image analysis: clinical significance," *World J Surg*, vol. 24, pp. 603-7, 2000.

[54]   X. Chen, Z. Ai, M. Rasmussen, P. Bajcsy, L. Auvil, M. Welge, L. Leach, S. Vangveeravong, A. J. Maniotis, and R. Folberg, "Three-dimensional reconstruction of extravascular matrix patterns and blood vessels in human uveal melanoma tissue: techniques and preliminary findings," *Invest Ophthalmol Vis Sci*, vol. 44, pp. 2834-40, 2003.

[55]   A. H. Boag, L. A. Kennedy, and M. J. Miller, "Three-dimensional microscopic image reconstruction of prostatic adenocarcinoma," *Arch Pathol Lab Med*, vol. 125, pp. 562-6, 2001.

[56]   M. Nakayama, H. Takahashi, K. Yao, K. Inagi, T. Makoshi, H. Nagai, and M. Okamoto, "Three-dimensional analysis of supracricoid hemilaryngopharyngectomized hypopharyngeal cancer," *Acta Otolaryngol Suppl*, pp. 35-40, 2002.

[57]   A. Mojsilovic, B. Rogowitz, J. Gomes, and T. S. Deisboeck, "Analysis and 3D reconstruction of heterogeneity in malignant brain tumors: an interdisciplinary case study using a novel computational visualization approach," *Anal Quant Cytol Histol*, vol. 24, pp. 125-33, 2002.

[58]   C. R. Olcott, "Cell types and histologic patterns in carcinoma of the lung; observations on the significance of tumors containing more than one type of cell," *Am J Pathol*, vol. 31, pp. 975-95, 1955.

[59]   A. M. Maeshima, T. Niki, A. Maeshima, T. Yamada, H. Kondo, and Y. Matsuno, "Modified scar grade: a prognostic indicator in small peripheral lung adenocarcinoma," *Cancer*, vol. 95, pp. 2546-54, 2002.

[60]   K. Okudera, Y. Kamata, S. Takanashi, Y. Hasegawa, T. Tsushima, Y. Ogura, K. Nakanishi, H. Sato, and K. Okumura, "Small adenocarcinoma of the lung: prognostic significance of central fibrosis chiefly because of its association with angiogenesis and lymphangiogenesis," *Pathol Int*, vol. 56, pp. 494-502, 2006.

[61]   H. Asamura, "Minimally invasive approach to early, peripheral adenocarcinoma with ground-glass opacity appearance," *Ann Thorac Surg*, vol. 85, pp. S701-4, 2008.

[62]   T. Aoki, Y. Tomoda, H. Watanabe, H. Nakata, T. Kasai, H. Hashimoto, M. Kodate, T. Osaki, and K. Yasumoto, "Peripheral lung adenocarcinoma: correlation of thin-section CT findings with histologic prognostic factors and survival," *Radiology*, vol. 220, pp. 803-9, 2001.

[63]   K. Ikeda, K. Awai, T. Mori, K. Kawanaka, Y. Yamashita, and H. Nomori, "Differential diagnosis of ground-glass opacity nodules: CT number analysis by three-dimensional computerized quantification," *Chest*, vol. 132, pp. 984-90, 2007.

[64]   J. de Ryk, E. Namati, J. Thiesse, and G. McLennan, "New imaging approaches for understanding lung cancer response to treatment," *Clin Pharmacol Ther*, vol. 84, pp. 517-22, 2008.

[65]   P. Therasse, S. G. Arbuck, E. A. Eisenhauer, J. Wanders, R. S. Kaplan, L. Rubinstein, J. Verweij, M. Van Glabbeke, A. T. van Oosterom, M. C. Christian, and S. G. Gwyther, "New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada," *J Natl Cancer Inst*, vol. 92, pp. 205-16, 2000.

[66]   R. S. Tuma, "Sometimes size doesn't matter: reevaluating RECIST and tumor response rate endpoints," *J Natl Cancer Inst*, vol. 98, pp. 1272-4, 2006.

[67]   L. Ibanez, W. Schroeder, L. Ng, and J. Cates, *The ITK Software Guide*. United States of America: Kitware Inc., 2005.

[68]   B. Godbout, J. A. de Guise, G. Soulez, and G. Cloutier, "3D elastic registration of vessel structures from IVUS data on biplane angiography," *Acad Radiol*, vol. 12, pp. 10-6, 2005.

[69]   F. Wang, B. C. Vemuri, A. Rangarajan, and S. J. Eisenschenk, "Simultaneous nonrigid registration of multiple point sets and atlas construction," *IEEE Trans Pattern Anal Mach Intell*, vol. 30, pp. 2011-22, 2008.

[70]   S. L. Aquino, J. C. Asmuth, R. H. Moore, S. B. Weise, and A. J. Fischman, "Improved image interpretation with registered thoracic CT and positron emission tomography data sets," *AJR Am J Roentgenol*, vol. 178, pp. 939-44, 2002.

[71]   R. Chan, Y. He, A. Haque, and J. Zwischenberger, "Computed tomographic-pathologic correlation of gross tumor volume and clinical target volume in non-small cell lung cancer: a pilot experience," *Arch Pathol Lab Med*, vol. 125, pp. 1469-72, 2001.

[72]   M. A. Jacobs, J. P. Windham, H. Soltanian-Zadeh, D. J. Peck, and R. A. Knight, "Registration and warping of magnetic resonance images to histological sections," *Med Phys*, vol. 26, pp. 1568-78, 1999.

[73]   S. E. Clarke, R. R. Hammond, J. R. Mitchell, and B. K. Rutt, "Quantitative assessment of carotid plaque composition using multicontrast MRI and registered histology," *Magn Reson Med*, vol. 50, pp. 1199-208, 2003.

[74]   H. Sakurai, A. Maeshima, S. Watanabe, K. Suzuki, R. Tsuchiya, A. M. Maeshima, Y. Matsuno, and H. Asamura, "Grade of stromal invasion in small adenocarcinoma of the lung: histopathological minimal invasion and prognosis," *Am J Surg Pathol*, vol. 28, pp. 198-206, 2004.

[75]   D. Eidelman and M. Molson Medical Informatics Project, "Pulmonary Airway Tree," MedEdPORTAL: http://www.aamc.org/meded/mededportal/downloads/MEP-ID-0008.htm   2006.

[76]   M. Mckee and M. Molson Medical Informatics Project, "Respiratory Bronchiole," MedEdPORTAL: http://www.aamc.org/meded/mededportal/downloads/MEP-ID-0009.htm   2007.

[77]   B. Markarian and E. Dailey, *Preparation of Inflated Lung Specimens. In The Lung. Radiologic Pathologic Correlations*. St Louis: Mosby Company, 1984.

[78]   s. Tomimitsu, S. Samrah, E. A. Hoffman, K. C. Beck, and G. McLennan, "Lung Preparation Technique for Computed Tomography - Pathologic Correlation," *American Journal of Respiratory & Critical Care Medicine*, vol. 167, pp. A875, 2003.

[79]   J. Thiesse, J. de Ryk, S. Bond, J. Vislisel, E. A. Hoffman, E. Nixon, J. M. Reinhardt, A. Ross, and G. McLennan, "Assessment of Mouse Lung Fixation Pressures Using Computed Tomography," presented at American Thoracic Society, San Diego, 2005.

[80]   E. Namati, J. De Ryk, J. Thiesse, Z. Towfic, E. Hoffman, and G. McLennan, "Large image microscope array for the compilation of multimodality whole organ image databases," *Anat Rec (Hoboken)*, vol. 290, pp. 1377-87, 2007.

[81]   J. de Ryk, E. Namati, J. Reinhardt, C. Piker, Y. Xu, L. Liu, E. A. Hoffman, and G. McLennan, "A Whole Organ Serial Sectioning and Imaging System for Correlation of Pathology to Computed Tomography," *Proceedings of SPIE: Three-Dimensional and Multidimensional Microscopy*, vol. 5324, pp. 224-234, 2004.

[82]   K. Suzuki, T. Yokose, J. Yoshida, M. Nishimura, K. Takahashi, K. Nagai, and Y. Nishiwaki, "Prognostic significance of the size of central fibrosis in peripheral adenocarcinoma of the lung," *Ann Thorac Surg*, vol. 69, pp. 893-7, 2000.

[83]   R. Moll, W. W. Franke, D. L. Schiller, B. Geiger, and R. Krepler, "The catalog of human cytokeratins: patterns of expression in normal epithelia, tumors and cultured cells," *Cell*, vol. 31, pp. 11-24, 1982.

[84]   A. Hartigan and M. A. Wong, "A K-means clustering algorithm," *Applied Statistics* vol. 28, pp. 100-108, 1979.

[85]   F. Guillemin, M. Devaux, and F. Guillon, "Evaluation of Plant Histology by Automatic Clustering Based on Individual Cell Morphological Features," *Image Analysis and Stereology*, vol. 23, pp. 13-22, 2004.

[86]   J. B. Maintz and M. A. Viergever, "A survey of medical image registration," *Medical Image Analysis*, vol. 2, pp. 1-36, 1998.

[87]   F. Bookstein, "Principal Warps: Thin-Plate Splines and the Decomposition of Deformations," *IEEE Transactions on pattern analysis and machine intellegence*, vol. 11, pp. 567-585, 1989.

[88]   C. Studholme, D. Hill, and D. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognition*, vol. 32, pp. 71 - 86, 1999.

[89]   R. Gonzalez and R. Woods, *Digital Image Processing*, Second ed. New Jersey: Prentice-Hall, Inc., 2002.

[90]   K. E. Selkirk, *Pattern and place an introduction to the mathematics of geography*. Cambridge: Cambridge University Press, 1982.

[91]   K. Manning, B. Tepfer, G. Goldklang, R. Loyd, P. Garimella, and R. Halkar, "Clinical practice guidelines for the utilization of positron emission tomography/computed tomography imaging in selected oncologic applications: suggestions from a provider group," *Mol Imaging Biol*, vol. 9, pp. 324-32; discussion 323, 2007.

[92]   M. Khalaf, H. Abdel-Nabi, J. Baker, Y. Shao, D. Lamonica, and J. Gona, "Relation between nodule size and 18F-FDG-PET SUV for malignant and benign pulmonary nodules," *J Hematol Oncol*, vol. 1, pp. 13, 2008.

[93]   A. Davies, C. Tan, C. Paschalides, S. F. Barrington, M. O'Doherty, M. Utley, and T. Treasure, "FDG-PET maximum standardised uptake value is associated with variation in survival: analysis of 498 lung cancer patients," *Lung Cancer*, vol. 55, pp. 75-8, 2007.

[94]   L. E. Dodd, R. F. Wagner, S. G. Armato, 3rd, M. F. McNitt-Gray, S. Beiden, H. P. Chan, D. Gur, G. McLennan, C. E. Metz, N. Petrick, B. Sahiner, and J. Sayre, "Assessment methodologies and statistical issues for computer-aided diagnosis of lung nodules in computed tomography: contemporary research topics relevant to the lung image database consortium," *Acad Radiol*, vol. 11, pp. 462-75, 2004.