Defining Novel Molecular Events in Leiomyosarcoma

by

Amir Hossein Safavi Ardebili

A thesis submitted in conformity with the requirements for the degree of Master of Science
Institute of Medical Science
University of Toronto

© Copyright by Amir Hossein Safavi Ardebili 2016
Defining Novel Molecular Events in Leiomyosarcoma

Amir Hossein Safavi Ardebili

Master of Science

Institute of Medical Science
University of Toronto

2016

Abstract

Leiomyosarcoma (LMS) is a smooth muscle neoplasm and a common histological subtype of soft tissue sarcoma (STS). LMS primarily arises in the retroperitoneum and abdominal cavity, the extremities, and the uterine wall. Current treatment outcomes are poor; consequently, personalized medicine informed by the molecular basis of the disease is needed. Here, YAF2-PRICKLE1 and ESR1-CCDC92, two novel fusion transcripts, were discovered in an LMS cell line and were detected as rare events in the tumor from which that cell line was derived. Furthermore, unbiased cluster analysis indicated that abdominal and extremity LMS transcriptomes group distinctly from uterine LMS transcriptomes, suggesting potential molecular subtypes and supporting the notion that uterine LMS is biologically distinct from abdominal and extremity LMS. Lastly, decreased FOXO4 expression correlated with increased smooth muscle differentiation in LMS. Collectively, these preliminary findings present avenues for future investigations which may facilitate diagnostic and therapeutic innovations for this disease.
Acknowledgments

I am indebted to my supervisor, Dr. Rebecca Gladdy, who gave me this opportunity to challenge myself and develop my critical thinking and professional skills. Thank you for being my strongest advocate and toughest critic. I have been so fortunate to have you as a mentor and to learn from your resilience, patience, passion, and work ethic. I hope to continue building on the lessons you have taught me.

I am grateful to Dr. David Malkin, Dr. Steven Gallinger, and Dr. Adam Shlien, who were members of my Program Advisory Committee. Your scholarly analysis and enthusiastic engagement with my studies were invaluable and reassuring. Many thanks to Dr. Jeff Wrana and Jess Shen for their dedicated collaboration; much of the work in this thesis would not be possible without your efforts and insights.

My day-to-day progress in the lab was supported by the guidance and technical expertise of Rosemarie Venier, Dr. Yael Babichev and Dr. Timothy McKinnon. Warmest thanks to my dear friends Andrew Kim and Anthony Choi for their daily dose of humorous banter and camaraderie. Here’s to another decade of friendship.

My parents provided me the educational opportunities that have led me to this point; for this, I am very grateful. In particular, I am immeasurably thankful to my mother, Homeyra; your love, unwavering dedication and integrity give me strength when all hope seems lost. The completion of this thesis is a testament to your faith in me and your encouragement. I truly cannot thank you enough.

I would like to acknowledge the following scholarship and award programs which supported my graduate studies: Ontario Student Opportunity Trust Fund from the Lunenfeld-Tanenbaum Research Institute, Dalton Whitebread Scholarship Fund from the University of Toronto’s Faculty of Medicine, and the Institute of Medical Science Entrance Scholarship. Additionally, this work was supported by an OICR operating grant.

Lastly, I am deeply appreciative of the brave and generous leiomyosarcoma patients who donated their tissue samples to Mount Sinai Hospital’s Sarcoma Tissue Bank. Your contributions grant projects such as this one the opportunity to solve the leiomyosarcoma puzzle and discover therapeutic targets for this terrible disease. May these studies lead to benefits for these patients in the near future.
Contributions

Amir Safavi (author) prepared this thesis. The planning, execution, and analysis of all studies were performed in whole or in part by the author. The following contributions from mentors, collaborators and colleagues are gratefully acknowledged:

Dr. Rebecca Gladdy (Supervisor and Thesis Committee Member) – mentorship; laboratory resources; guidance and assistance in planning and analysis of experiments and thesis preparation

Dr. David Malkin (Thesis Committee Member) – mentorship; guidance in interpretation of results and thesis preparation

Dr. Steven Gallinger (Thesis Committee Member) – mentorship; guidance in interpretation of results and thesis preparation

Dr. Adam Shlienz (Thesis Committee Member) – mentorship; guidance in interpretation of results and thesis preparation

Dr. Jeff Wrana – laboratory resources; guidance and assistance in planning experiments

Jess Shen – analysis of RNA-sequencing data for gene fusion detection and transcriptome profiling, unsupervised hierarchical clustering of LMS transcriptomes

Kin Chan – cDNA library preparation and RNA-sequencing of LMS patient samples and cell lines

Dr. Yael Babichev – immunocytochemistry staining of LMS cell lines for SMA, desmin and h-caldesmon, guidance and assistance in the execution and analysis of experiments

Rosemarie Venier – guidance and assistance in the planning, execution, and analysis of experiments

Leah Kabaroff – guidance and assistance in the execution of experiments

Dr. Timothy McKinnon – guidance and assistance in the execution of experiments

Dr. Brendan Dickson – pathological assessment of LMS patient samples
# Table of Contents

Acknowledgments ........................................................................................................ iii  
Contributions ................................................................................................................ iv  
Table of Contents ......................................................................................................... v  
List of Tables ................................................................................................................ x  
List of Figures ................................................................................................................. xi  
List of Abbreviations ................................................................................................... xiii

Chapter 1  General Introduction ......................................................................................1

  1.1 Preamble ..................................................................................................................2  
  1.2 Thesis Organization .................................................................................................3

Chapter 2  Literature Review ............................................................................................4

  2.1 Abstract ....................................................................................................................5  
  2.2 Soft Tissue Sarcoma .................................................................................................5
    2.2.1 Overview .............................................................................................................5
    2.2.2 Etiology ...............................................................................................................7
    2.2.3 Diagnosis: Histopathology and Molecular Genetics .........................................8
    2.2.4 Therapeutic Management ..................................................................................9
    2.2.5 Molecular Categories .....................................................................................11
  2.3 Clinical Overview of Leiomyosarcoma .................................................................13
    2.3.1 Epidemiology ....................................................................................................13
    2.3.2 Diagnosis: Histopathology ..............................................................................13
    2.3.3 Prognostic Factors, Survival, and Recurrence ...............................................16
    2.3.4 Therapeutic Management ...............................................................................19
2.4 Molecular Characteristics of Leiomyosarcoma ................................................................. 21
  2.4.1 Molecular Alterations ...................................................................................................... 21
  2.4.2 Molecular Subtypes ......................................................................................................... 24
2.5 Gene Fusions .......................................................................................................................... 28
  2.5.1 Mechanisms of Formation .............................................................................................. 28
  2.5.2 Functional Consequences of Gene Fusions ........................................................................ 32
  2.5.3 Gene Fusions as Diagnostic Biomarkers and Therapeutics .............................................. 33
2.6 Overview of PI3K/AKT/mTOR Pathway .................................................................................. 34
  2.6.1 PI3K/AKT/mTOR Pathway ............................................................................................... 34
  2.6.2 PI3K .................................................................................................................................. 36
  2.6.3 AKT .................................................................................................................................. 38
  2.6.4 mTOR ............................................................................................................................... 40
2.7 Cell of Origin of Leiomyosarcoma ......................................................................................... 40
2.8 Overview of Smooth Muscle Differentiation ............................................................................. 42
  2.8.1 Smooth Muscle Cells ....................................................................................................... 42
  2.8.2 Regulation of SMC Differentiation ................................................................................... 43
  2.8.3 MYOCD ............................................................................................................................ 45
  2.8.4 FOXO4 ............................................................................................................................. 47

Chapter 3  Thesis Aims and Hypotheses ..................................................................................... 50
  3.1 Thesis Aims ............................................................................................................................ 51
  3.2 Hypotheses ........................................................................................................................... 51
    3.2.1 YAF2-PRICKLE1 and ESR1-CCDC92 Gene Fusions are Recurrent in LMS ............... 51
    3.2.2 LMS Transcriptome Profiles Cluster by Site ................................................................. 52
    3.2.3 FOXO4 Expression and Phosphorylation Regulates Smooth Muscle Differentiation in LMS ............... 53

Chapter 4  Investigating Novel Gene Fusions and Molecular Subtypes in LMS ....................... 54
4.1 Abstract..................................................................................................................55

4.2 Introduction.............................................................................................................57

4.3 Materials and Methods..........................................................................................61
  4.3.1 Tumor Samples.....................................................................................................61
  4.3.2 Cell Line Establishment and Culture Conditions..................................................61
  4.3.3 RNA- Sequencing (RNA-Seq)................................................................................62
  4.3.4 Gene Fusion Detection..........................................................................................62
  4.3.5 RT-PCR..................................................................................................................63
  4.3.6 Unsupervised Hierarchical Clustering ..................................................................64
  4.3.7 Analysis of Differential Gene Expression ..............................................................64
  4.3.8 Gene Set Enrichment Analysis and Functional Annotation Clustering....................65

4.4 Results.......................................................................................................................66
  4.4.1 YAF2-PRICKLE1 and ESR1-CCDC92 are Present in the Index Cell Line ............66
  4.4.2 YAF2-PRICKLE1 and ESR1-CCDC92 are not Recurrent in Additional LMS Patient Samples ..........................................................68
  4.4.3 Transcriptomes of Abdominal and Extremity LMS Form a Cluster .....................70
  4.4.4 Gene Set Enrichment and Functional Annotation Clustering of Differentially Expressed Genes of Differentially Expressed Genes................................................73
  4.4.5 PRKAA2 and CDKN2A are Differentially Expressed Between the Abdominal and Extremity Cluster and the Uterine Samples .........................................................74

4.5 Discussion................................................................................................................76

Chapter 5 Investigating the Role of FOXO4 in LMS Smooth Muscle Differentiation .....83

5.1 Abstract....................................................................................................................84

5.2 Introduction..............................................................................................................86

5.3 Materials and Methods............................................................................................90
  5.3.1 Tumor Samples.....................................................................................................90
  5.3.2 Cell Line Establishment and Culture Conditions..................................................90
5.3.3 RNA-Seq (RNA-Seq) .......................................................................................... 91
5.3.4 Analysis of Gene Expression .............................................................................. 91
5.3.5 Protein Extraction from Cell Lines .................................................................... 92
5.3.6 siRNA Transfection ......................................................................................... 92
5.3.7 Immunoblotting ............................................................................................... 93
5.3.8 Immunocytochemistry (ICC) ........................................................................... 93

5.4 Results .................................................................................................................. 95
5.4.1 Characterizing Smooth Muscle Differentiation in LMS ................................. 95
5.4.2 The PI3K/AKT Pathway is Activated in LMS .................................................. 100
5.4.3 MYOCDS Expression Correlates with Smooth Muscle Differentiation in LMS Cell Lines ................................................................. 101
5.4.3 FOXO4 Expression Correlates with Smooth Muscle Differentiation in LMS Cell Lines ................................................................. 104

5.5 Discussion ............................................................................................................. 107

Chapter 6 Concluding Summary, General Discussion, Future Directions, and Translational Relevance ................................................................. 113

6.1 Concluding Summary .......................................................................................... 114

6.2 General Discussion .............................................................................................. 118
6.2.1 Defining Molecular Events in LMS ................................................................. 118
6.2.2 Limitations ..................................................................................................... 123

6.3 Future Directions .................................................................................................. 130
6.3.1 Identifying Recurrent Gene Fusions in LMS ................................................... 130
6.3.2 Characterizing the Oncogenic Potential of Recurrent Gene Fusions in LMS ...... 131
6.3.3 Validating LMS Transcriptome Clustering ..................................................... 132
6.3.4 Investigate Alternative Splicing in LMS ....................................................... 132
6.3.5 Whole Genome Profiling of LMS ................................................................. 133
6.3.6 Epigenomic Profiling of LMS ...................................................................... 134
6.3.7 Investigating the Regulation of Smooth Muscle Gene Expression in LMS ........135

6.4 Translational Relevance ........................................................................137

References ..................................................................................................139
List of Tables

Table 2-1. Genetic Predisposition to Soft Tissue Sarcoma .......................................................8

Table 2-2. Molecular Categories of Soft Tissue Sarcomas ..........................................................12

Table 2-3. FNCLCC Grading System for Soft Tissue Sarcomas ..................................................19

Table 2-4. Targets Unique to LMS Subtypes ..............................................................................28

Table 2-5. Targeted Agents that Produce Responses in Sarcoma ..............................................34

Table 4-1. Clinical and Histopathological Features of LMS Patient Samples ...........................71
List of Figures

Figure 2-1. Taxonomy of Soft Tissue Sarcoma ................................................................. 6

Figure 2-2. Histology of Well-Differentiated to Poorly Differentiated LMS ......................... 15

Figure 2-3. Genomic Profiles of “Rearranged” and “Arm” LMS ....................................... 25

Figure 2-4. Unsupervised Hierarchical Clustering of LMS Samples Reveals Three Molecular Subtypes ........................................................................................................... 27

Figure 2-5. Triggers for Genomic Rearrangements ............................................................ 31

Figure 2-6. Chromosomal Basis of Gene Fusions .............................................................. 32

Figure 2-7. PI3K/AKT/mTOR Signalling Pathway .............................................................. 32

Figure 2-8. Regulation of Smooth Muscle Differentiation .................................................. 44

Figure 4-1. Two Novel Fusion Transcripts are Detected in a Primary LMS Cell Line using RNA-Seq ................................................................................................................. 59

Figure 4-2. YAF2-PRICKLE1 and ESR1-CCDC92 Transcripts are Confirmed in the Index Cell Line using RT-PCR .......................................................... 67

Figure 4-3. YAF2-PRICKLE1 and ESR1-CCDC92 Spanning Reads are Detected in the Index Case by RNA-Seq .................................................. 69

Figure 4-4. Unsupervised Hierarchical Clustering of 15 LMS Patient Samples Reveals Two Molecular Subtypes ................................................................. 72

Figure 4-5. PRKAA2 and CDKN2A are Differentially Expressed between the Uterine and Non-Uterine Samples ................................................................. 75

Figure 5-1. Smooth Muscle Cell Phenotypes ..................................................................... 94

Figure 5-2. STS39 is the Most Differentiated LMS Cell Lines ................................... 96-97
Figure 5-3. MYLK, MYH11, and CALD1 are Differentially Expressed between Abdominal and Extremity LMS .................................................................98

Figure 5-4. AKT is Phosphorylated at Thr308 in LMS .................................................................99

Figure 5-5. Myocardin and SRF Expression in LMS Cell Lines and Patient Samples .. 102-103

Figure 5-6. FOXO4 Expression in LMS .....................................................................................105
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4E-BP1</td>
<td>Eukaryotic translation initiation factor 4E-binding protein 1</td>
</tr>
<tr>
<td>β-actin</td>
<td>Beta-actin</td>
</tr>
<tr>
<td>ACTA2</td>
<td>Alpha smooth muscle actin</td>
</tr>
<tr>
<td>AR</td>
<td>Androgen receptor</td>
</tr>
<tr>
<td>AMPK 5′</td>
<td>AMP-activated protein kinase</td>
</tr>
<tr>
<td>BH p-value</td>
<td>Benjamini-Hochberg p-value</td>
</tr>
<tr>
<td>CALD1</td>
<td>Caldesmon 1</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary DNA</td>
</tr>
<tr>
<td>CCDC92</td>
<td>Coiled-coil domain containing 92</td>
</tr>
<tr>
<td>CDKN2A</td>
<td>Cyclin-dependent kinase inhibitor 2A</td>
</tr>
<tr>
<td>CDK4</td>
<td>Cyclin-dependent kinase 4</td>
</tr>
<tr>
<td>CGH</td>
<td>Comparative genomic hybridization</td>
</tr>
<tr>
<td>Chr6</td>
<td>Chromosome 6</td>
</tr>
<tr>
<td>Chr12</td>
<td>Chromosome 12</td>
</tr>
<tr>
<td>CK1</td>
<td>Casein kinase 1</td>
</tr>
<tr>
<td>CNN1</td>
<td>Calponin 1</td>
</tr>
<tr>
<td>COBRA-FISH</td>
<td>Combined binary ratio labeling fluorescence in situ hybridization</td>
</tr>
<tr>
<td>COL1A1</td>
<td>Collagen, Type I, Alpha 1</td>
</tr>
<tr>
<td>DAPI</td>
<td>4′,6-diamidino-2-phenylindole</td>
</tr>
<tr>
<td>DAVID</td>
<td>Database for Annotation, Visualization and Integrated Discovery</td>
</tr>
<tr>
<td>DES</td>
<td>Desmin</td>
</tr>
<tr>
<td>DFSP</td>
<td>Dermatofibrosarcoma protuberans</td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco’s modified eagle medium</td>
</tr>
<tr>
<td>DSS</td>
<td>Disease-specific survival</td>
</tr>
<tr>
<td>DYRK1</td>
<td>Dual specificity tyrosine-phosphorylation-regulated kinase 1</td>
</tr>
<tr>
<td>ER</td>
<td>Estrogen receptor</td>
</tr>
<tr>
<td>ERK</td>
<td>Extracellular signal-regulated kinase</td>
</tr>
<tr>
<td>ESR1</td>
<td>Estrogen receptor 1</td>
</tr>
<tr>
<td>FBS</td>
<td>Fetal bovine serum</td>
</tr>
<tr>
<td>FFPE</td>
<td>Formalin-fixed, paraffin-embedded</td>
</tr>
<tr>
<td>FNCLCC</td>
<td>La Fédération Nationale des Centres de Lutte Contre le Cancer</td>
</tr>
<tr>
<td>FOXO1</td>
<td>Forkhead box protein 01</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>FOXO4</td>
<td>Forkhead box protein 04</td>
</tr>
<tr>
<td>GAPDH</td>
<td>Glyceraldehyde-3-phosphate dehydrogenase</td>
</tr>
<tr>
<td>GIST</td>
<td>Gastrointestinal stromal tumor</td>
</tr>
<tr>
<td>HPRT</td>
<td>Hypoxanthine-guanine phosphoribosyltransferase</td>
</tr>
<tr>
<td>ICC</td>
<td>Immunocytochemistry</td>
</tr>
<tr>
<td>IGF-1</td>
<td>Insulin-like growth factor 1</td>
</tr>
<tr>
<td>IGF-2</td>
<td>Insulin-like growth factor 2</td>
</tr>
<tr>
<td>IGF1R</td>
<td>Insulin-like growth factor 1 receptor</td>
</tr>
<tr>
<td>IGF2R</td>
<td>Insulin-like growth factor 2 receptor</td>
</tr>
<tr>
<td>IHC</td>
<td>Immunohistochemistry</td>
</tr>
<tr>
<td>IRS-2</td>
<td>Insulin receptor substrate 2</td>
</tr>
<tr>
<td>LMS</td>
<td>Leiomyosarcoma</td>
</tr>
<tr>
<td>LPP</td>
<td>Lim domain containing preferred translocation partner in lipoma</td>
</tr>
<tr>
<td>MADS</td>
<td>MCM1, Agamous, Deficiens, SRF</td>
</tr>
<tr>
<td>MAPK</td>
<td>p38 mitogen-activated protein kinase</td>
</tr>
<tr>
<td>mlgG</td>
<td>Mouse immunoglobulin G</td>
</tr>
<tr>
<td>MSC</td>
<td>Mesenchymal stem cell</td>
</tr>
<tr>
<td>mTOR</td>
<td>Mechanistic target of rapamycin</td>
</tr>
<tr>
<td>mTORC1</td>
<td>Mammalian target of rapamycin complex 1</td>
</tr>
<tr>
<td>MYH11</td>
<td>Myosin, heavy chain 11, smooth muscle</td>
</tr>
<tr>
<td>MYLK</td>
<td>Myosin light-chain kinase, smooth muscle</td>
</tr>
<tr>
<td>MYOCD</td>
<td>Myocardin</td>
</tr>
<tr>
<td>NF1</td>
<td>Neurofibromatosis 1</td>
</tr>
<tr>
<td>NHEJ</td>
<td>Non-homologous end joining</td>
</tr>
<tr>
<td>OS</td>
<td>Overall survival</td>
</tr>
<tr>
<td>p16</td>
<td>Cyclin-dependent kinase inhibitor 2A</td>
</tr>
<tr>
<td>P70S6K</td>
<td>P70 ribosomal protein S6 kinase</td>
</tr>
<tr>
<td>p-AKT(Thr308)</td>
<td>AKT phosphorylated at Threonine 308</td>
</tr>
<tr>
<td>PAX3</td>
<td>Paired Box 3</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PDGFβ</td>
<td>Platelet-derived growth factor beta</td>
</tr>
<tr>
<td>PDGFRA</td>
<td>Platelet-derived growth factor receptor alpha</td>
</tr>
<tr>
<td>PDK1</td>
<td>Phosphoinositide-dependent kinase 1</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>PFS</td>
<td>Progression-free survival</td>
</tr>
<tr>
<td>PI3K</td>
<td>Phosphoinositide 3-kinase</td>
</tr>
<tr>
<td>PIP₂</td>
<td>Phosphatidylinositol (4,5)-biphosphate</td>
</tr>
<tr>
<td>PIP₃</td>
<td>Phosphatidylinositol (3,4,5)-triphosphate</td>
</tr>
<tr>
<td>PR</td>
<td>Progesterone receptor</td>
</tr>
<tr>
<td>PRICKLE1</td>
<td>Prickle-like protein 1</td>
</tr>
<tr>
<td>PRKAA1</td>
<td>5'-AMP-activated protein kinase catalytic subunit alpha-1</td>
</tr>
<tr>
<td>PRKAA2</td>
<td>5'-AMP-activated protein kinase catalytic subunit alpha-2</td>
</tr>
<tr>
<td>PTEN</td>
<td>Phosphatase and tensing homolog</td>
</tr>
<tr>
<td>PVDF</td>
<td>Polyvinylidene fluoride</td>
</tr>
<tr>
<td>Rb</td>
<td>Retinoblastoma</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNA-Seq</td>
<td>Ribonucleic acid sequencing</td>
</tr>
<tr>
<td>RPTOR</td>
<td>Regulatory associated protein of mTOR</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Reverse transcription polymerase chain reaction</td>
</tr>
<tr>
<td>RTK</td>
<td>Receptor tyrosine kinase</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulfate</td>
</tr>
<tr>
<td>SGK1</td>
<td>Serum/glucocorticoid regulated kinase 1</td>
</tr>
<tr>
<td>SH2</td>
<td>SRC homology 2 domain</td>
</tr>
<tr>
<td>siFOXO4</td>
<td>Forkhead box protein 04 short interfering ribonucleic acid</td>
</tr>
<tr>
<td>siRNA</td>
<td>Short interfering ribonucleic acid</td>
</tr>
<tr>
<td>siScr</td>
<td>Non-targeting short interfering ribonucleic acid</td>
</tr>
<tr>
<td>SMA</td>
<td>Smooth muscle actin</td>
</tr>
<tr>
<td>SMC</td>
<td>Smooth muscle cell</td>
</tr>
<tr>
<td>SRF</td>
<td>Serum response factor</td>
</tr>
<tr>
<td>STS</td>
<td>Soft Tissue Sarcoma</td>
</tr>
<tr>
<td>TAGLN</td>
<td>Transgelin</td>
</tr>
<tr>
<td>TP53</td>
<td>Tumor protein p53</td>
</tr>
<tr>
<td>TSC2</td>
<td>Tuberous sclerosis 2</td>
</tr>
<tr>
<td>YAF2</td>
<td>YY1-associated factor 2</td>
</tr>
</tbody>
</table>
Chapter 1  General Introduction
1.1 Preamble

Leiomyosarcoma (LMS) is a tumor comprised of cells displaying smooth-muscle differentiation and accounts for 11% of soft tissue sarcomas (Ducimetière et al., 2011). LMS can have an aggressive clinical course and outcomes for patients with LMS are poor: 39% of patients have local and/or distant recurrence, 6-9% have late (>5 year) recurrence and the 5-year overall survival rate is 63% (Gladdy et al., 2013). The mainstay of curative treatment of LMS is surgical resection; treatment for advanced and metastatic LMS using chemotherapy and radiation therapy is largely ineffective, resulting in poor outcomes (Bathan et al., 2013). Defining molecular events in LMS may lead to identifying molecular signatures which may serve as diagnostic markers and/or facilitate the development of targeted therapeutics, thereby improving outcomes for patients. Currently, there is a limited understanding of the genetic drivers and molecular subtypes in LMS, as well as if developmental pathways are dysregulated in this disease. Traditionally, LMS has been characterized as a genomically unstable neoplasm. Recent genomic studies have identified two potential subtypes of LMS: a subtype predominantly featuring extremity tumors and characterized by complex, unstable genomes, and a subtype primarily consisting of retroperitoneal tumors and featuring simple, stable genomes (Italiano et al., 2013). An independent study identified three potential subtypes of LMS based on expression profiling of LMS samples (Beck et al., 2009). These findings may challenge the notion that LMS is uniformly characterized by genomic instability and a dearth of simple genetic alterations and suggest the possibility of redefining the genetic basis of LMS.
1.2 Thesis Organization

The overarching purpose of this thesis is to define molecular events in LMS. This purpose manifests itself in two distinct, yet synergistic aims: 1) to investigate novel molecular characteristics of LMS and 2) to elucidate the mechanism regulating smooth muscle differentiation in LMS. To best reflect the rationale for these inquiries, this thesis follows the “multiple paper format” in place of the continuous “traditional format”. Chapter 2 serves to contextualize the work in this thesis by introducing soft tissue sarcoma (STS), presenting a clinical overview of LMS, and reviewing the known molecular characteristics of LMS, gene fusions, the PI3K/AKT/mTOR pathway, and smooth muscle differentiation. Chapter 3 outlines the research aims and hypotheses which direct this thesis. Chapter 4 presents findings from studies that investigated novel recurrent gene fusions and defined transcriptomic subtypes in LMS. Chapter 5 introduces aims, preliminary data, and future directions for elucidating how smooth muscle differentiation is regulated in LMS. In addition to complementing the discussion sections of the previous two chapters, Chapter 6 summarizes the key findings of the thesis, posits conclusions, presents a general discussion and postulates future directions and translational relevance.
Chapter 2    Literature Review
2.1 Abstract

This review chapter consists of seven sections. The first three sections (2.2, 2.3 and 2.4) present overviews of soft tissue sarcoma (STS) and leiomyosarcoma (LMS) from both clinical and molecular perspectives. The following section (2.5) introduces gene fusions, which are clinically relevant molecular signatures of numerous subtypes of STS. This section focuses on mechanisms of formation, functional consequences and diagnostic and therapeutic utility of gene fusions. Section 2.6 summarizes the PI3K/AKT/mTOR pathway, which is known to play a critical role in LMS. Finally, sections 2.7 and 2.8 discuss the cell of origin of LMS and present an overview of smooth muscle differentiation.

2.2 Soft Tissue Sarcoma

2.2.1 Overview

The term 'sarcoma' is derived from the Greek words meaning fleshy (sarcos) tumor (oma). STS are a diverse group of neoplasms with over 50 recognized subtypes (Figure 2-1) (reviewed in Fletcher et al., 2013). They are derived from mesenchymal tissues such as smooth muscle, skeletal muscle and adipocytes. STS constitute 1% of all human cancers but are more prevalent in children than in adults as they represent 15% of all pediatric cancers (reviewed in Brennan, 2005). The most common sites of STS
include extremity (50%), retroperitoneum (15%) trunk (10%), and the head and neck region (10%) (DeVita et al., 2011).

Figure 2-1. Taxonomy of Soft Tissue Sarcoma. This phylogeny consists of ~60 subtypes of soft tissue sarcoma as defined by the World Health Organization (WHO), with branching by lineage, similarity in prognosis, if they feature translocations, and if so, the genes shared among distinct fusions (in this order). MFH refers to malignant fibrous histocytoma, which is now a historical term and has been reclassified to
undifferentiated pleomorphic sarcoma (UPS) in the most recent WHO Soft Tissue Sarcoma Manual (Fletcher et al., 2013). PNET refers to primitive neuroectodermal tumor. (Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Cancer, 11, Taylor et al., Advances in sarcoma genomics and new therapeutic targets, pages 541-557, copyright (2011): License #3863161033508)

2.2.2 Etiology

STS are believed to arise sporadically from somatically acquired molecular alterations, though there are several genetic-predisposition syndromes and environmental exposures that are associated with sarcomas (reviewed in Helman and Meltzer, 2003). However, the genetic drivers of many subtypes of STS have not yet been elucidated. STS may arise from genetic pre-disposition syndromes including Li-Fraumeni Syndrome, caused by heterozygous germline TP53 mutations (Malkin et al., 1990), hereditary retinoblastoma, caused by RB1 mutations, and neurofibromatosis type 1, caused by mutations in NF1 (Table 2-1) (reviewed in Brennan, 2005). Environmental factors, such as exposure to vinyl chloride (reviewed in Helman and Meltzer, 2003), phenoxy herbicides (Hoppin et al., 1998) and dioxins may lead to STS formation. Ionizing radiation from radiation therapy may cause radiation associated sarcomas (RAS), with a median interval of 10 years between radiation therapy and RAS formation (Gladdy et al., 2010). Radiation-associated STS have significantly worse disease-specific survival (Gladdy et al., 2010) and are thought to arise from different molecular events (Guo et al., 2011). STS may also arise from infection, such as leiomyosarcomas (LMS) caused by Epstein-Barr virus in immunocompromised patients (McClain et al.,
1995; Lee et al., 1995) and Kaposi’s sarcomas caused by human herpesvirus 8 in AIDS patients (Chang et al., 1994).

Table 2-1. Genetic Predisposition to Soft Tissue Sarcoma.

<table>
<thead>
<tr>
<th>Diagnosed Disease (von Recklinghausen disease)</th>
<th>Gene</th>
<th>Chromosome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neurofibromatosis type I</td>
<td>MPNT</td>
<td>11q12</td>
</tr>
<tr>
<td>Retinoblastoma</td>
<td>Rb-1</td>
<td>13q14</td>
</tr>
<tr>
<td>Li-Fraumeni Syndrome</td>
<td>PS3</td>
<td>17p13</td>
</tr>
<tr>
<td>Gardner Syndrome</td>
<td>APC</td>
<td>5q21</td>
</tr>
<tr>
<td>Werner Syndrome (adult proportion)</td>
<td>WRN</td>
<td>8p12</td>
</tr>
<tr>
<td>Gorlin Syndrome (nevoid basal cell carcinoma syndrome)</td>
<td>PTC</td>
<td>9q22.3</td>
</tr>
<tr>
<td>Caneys Triad</td>
<td>GIST</td>
<td>Unknown</td>
</tr>
<tr>
<td>Tuberous Sclerosis (Bouwenli’s disease)</td>
<td>TSC1</td>
<td>9q34 16p13.3</td>
</tr>
<tr>
<td>MPNT: Malignant peripheral nerve sheath tumour</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIST: Gastrointestinal stromal tumour</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Adapted from The Surgeon: Journal of the Royal Colleges of Surgeons of Edinburgh and Ireland, 3, Brennan et al., Soft tissue sarcoma: Advances in understanding and management, pages 216-223, copyright (2005), with permission from Elsevier: License #3863170126917)

2.2.3 Diagnosis: Histopathology and Molecular Genetics

Histological evaluation of tumor biopsies using immunohistochemical (IHC) staining is essential to diagnosing soft tissue sarcomas. Positive staining for vimentin and negative staining for cytokeratin aid in diagnosing sarcomas but are not specific to STS. Specific IHC markers are used to identify some subtypes of STS, such as SMA, desmin and h-caldesmon in LMS and myogenin and MyoD in rhabdomyosarcoma (reviewed in Fletcher et al., 2013); however, disease-specific IHC markers have not been identified for many STS subtypes. A thorough understanding of the molecular and genetic
features of STS may be of more value than current classifications which rely on histologic descriptions (reviewed in Helman et al., 2003). For example, analysis of the gene expression patterns of translocation negative alveolar rhabdomyosarcoma (ARMS) identified a molecular signature that is more consistent with embryonal rhabdomyosarcoma (ERMS), which are PAX3-FOXO1 or PAX7-FOXO1 negative (Williamson et al., 2010). Furthermore, clinical outcomes of translocation negative ARMS patients were consistent with ERMS (Williamson et al., 2010). Thus, the authors of this study suggest that these patients should be reclassified based on their genetic profiles rather than with their histopathology alone (Williamson et al., 2010).

Determining the genetic alterations driving sarcomagenesis may improve STS detection and reduce overtreatment of patients with favourable biology, while providing less toxic and more effective treatment options for patients with aggressive disease where the genetic pathways driving the disease are known and therefore targeted.

2.2.4 Therapeutic Management

Surgery is the primary modality for curative treatment of most subtypes of STS. Some subtypes, such as Ewing’s sarcoma, are sensitive to chemotherapy and radiation therapy and thus may be effectively treated with these modalities; surgery may be reserved for the management of residual disease in this clinical context. One of the greatest challenges in the surgical management of STS patients is that most tumors are greater than 5 cm at presentation (DeVita, Lawrence et al. 2011), and therefore may require the resection of adjacent organs in the abdomen, and sacrifice of key
neurovascular structures and muscle groups in the extremity causing functional impairment. Neoadjuvant radiation therapy may be used to decrease the chance of local recurrence and to facilitate less extensive surgical resections (reviewed in Bathan et al., 2013). Chemotherapy use in STS varies according to the chemosensitivity of the subtype being treated. Neoadjuvant chemotherapy may be beneficial for chemosensitive histologies such as LMS and can be used to treat locally advanced disease prior to surgical resection. Adjuvant chemotherapy however is not routinely recommended for high-grade STS as it does not improve overall survival (OS) (Woll et al., 2012). Chemotherapy is also an important cornerstone of palliative treatment. Conventional chemotherapy for STS consists of cytotoxic drugs such as doxorubicin, a DNA intercalating agent, and ifosfamide, an alkylating agent (reviewed in Bathan et al., 2013). Limitations of current chemotherapeutic agents include limited efficacy and significant systemic toxicity, in particular doxorubicin induced cardiotoxicity and encephalopathy with ifosfamide use. Combination therapy with doxorubicin and ifosfamide has higher response rates than doxorubicin alone (26% versus 14%, p<0.0006) (Judson et al., 2014). Targeted molecular therapy for gastrointestinal stromal tumor (GIST), a common sarcoma of the gastrointestinal tract, has greatly improved outcome with the identification of the c-kit mutation and the application of imatinib, a selective tyrosine kinase inhibitor that inhibits KIT activation (reviewed in Corless et al., 2011). Over time, resistance to imatinib emerges in GIST patients as a result of prolonged use; less selective second line tyrosine kinase inhibitors, such as sunitinib, have emerged as effective treatments for these patients, though they have more side effects (Demetri et al., 2007). However, targeted molecular therapies have yet to be established for most STS (reviewed in Linch et al., 2014).
2.2.5 Molecular Categories

STS are classified into two broad genomic categories: a genomically stable category characterized by a near-diploid karyotype with few chromosomal arrangements and a genomically unstable category featuring complex karyotypes with arm or whole-chromosome gains or losses (Table 2-2) (reviewed in Taylor et al., 2011; reviewed in Helman et al., 2003; reviewed in Borden et al., 2003). Most genomically stable STS are defined by tumor-specific, translocations, many of which encode aberrant transcription factors that dysregulate the transcription of target genes, chimeric protein tyrosine kinases, or activating mutations in autocrine growth factors.

The protein products of these translocations may be central to tumor pathogenesis (reviewed in Taylor et al., 2011; reviewed in Helman et al., 2003). Translocation-associated STS include Ewing’s sarcoma (*EWSR1-FLI1, EWSR1-ERG, EWSR1-ETV1, EWSR1-ETV4, EWSR1-FEV*), solitary fibrous tumor (*NAB2-STAT6*), dermatofibrosarcoma protuberans (*COL1A1-PDGFβ*), ARMS (*PAX3-FOXO1, PAX7-FOXO1*), myxoid liposarcoma (*FUS-DDIT3*), synovial sarcoma(*SS18-SSX1, SS18-SSX2*), clear cell sarcoma (*EWSR1-CREB1, EWSR1-ATF1*) and alveolar soft parts sarcoma (*ASPSCR1-TFE3*) (reviewed in Mertens et al., 2016). Some genomically stable STS, such as GIST, may be defined by specific activating mutations such as *KIT* and, less commonly, *PDGFRα* mutations (reviewed in Taylor et al., 2011). Patients with genomically stable STS tend to be younger (27 year old) (reviewed in Borden et al., 2003). Genomically unstable STS may arise from a less aggressive form and progressively become more genomically complex, though high-grade lesions may
present without prior progression (reviewed in Taylor et al., 2011). These tumors typically arise in older patients (57 years old) and may be characterized by their prevalence of loss of function mutations in the p53 and Rb signaling pathways (reviewed in Borden et al., 2003). These aneuploid tumors often have hundreds of genetic mutations (Barretina et al., 2010) and lack reciprocal, tumor-specific translocations (reviewed in Helman et al., 2003). Thus, the molecular mechanisms responsible for sarcomagenesis are unknown for most of these tumors. Genomically unstable STS include liposarcoma, ERMS, and leiomyosarcoma, which is the focus of this thesis.

Table 2-2. Molecular Categories of Soft Tissue Sarcomas.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Genomically Stable STS</th>
<th>Genomically Unstable STS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Karyotype</td>
<td>Simple</td>
<td>Complex</td>
</tr>
<tr>
<td>Translocations</td>
<td>Reciprocal and tumor-specific</td>
<td>Non-reciprocal and non-specific</td>
</tr>
<tr>
<td>Average age at diagnosis</td>
<td>27</td>
<td>57</td>
</tr>
<tr>
<td>Prevalence of p53 pathway alterations</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Prevalence of Rb pathway alterations</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Representative STS subtypes</td>
<td>Dermatomiosarcoma Protuberans, Ewing's Sarcoma, Alveolar Rhabdomyosarcoma</td>
<td>Leiomyosarcoma, Liposarcoma, Embryonal Rhabdomyosarcoma</td>
</tr>
</tbody>
</table>

(Adapted from Clinical Cancer Research, 2003, 9, pages 1941-1956, Borden et al., Soft Tissue Sarcomas of Adults: State of the Translational Science, with permission from AACR: License #3863200813821)
2.3 Clinical Overview of Leiomyosarcoma

2.3.1 Epidemiology

LMS is a neoplasm composed of cells displaying some degree of smooth muscle differentiation (reviewed in Fletcher et al., 2013). It is one of the most common histological subtypes of soft tissue sarcoma (STS) (reviewed in Singer et al., 2000) and accounts for 11% of STS (Ducimetière et al., 2011). LMS primarily occurs in middle-aged and older adults (reviewed in Weiss et al., 2008) with a median age of 57 years (range, 18-88 years) (Gladdy et al., 2013). Though immunocompromised patients infected by Epstein-Barr virus, particularly pediatric patients, may be predisposed to LMS (McClain et al., 1995; Lee et al., 1995), LMS is not considered a hereditary disease and is thought to arise from somatically acquired genomic alterations. The disease is more common in women than in men: approximately 66% of retroperitoneal LMS (Hashimoto et al., 1985; Shmookler et al., 1983) and more than 75% of leiomyosarcomas of the vena cava (Kevorkian et al., 1973) occur in women.

2.3.2 Diagnosis: Histopathology

Histopathological analysis of a biopsy is essential to diagnosing LMS (reviewed in Bathan et al., 2013). Macroscopically, LMS forms a fleshy, grey to white to tan mass and larger tumors may display hemorrhage and necrosis (reviewed in Fletcher et al., 2013). Microscopically, LMS is comprised of varying percentages of spindle and
pleomorphic cells, as determined by the degree of differentiation (reviewed in Guillou and Aurias, 2009). Well-differentiated tumors consist of spindle cells grouped in longitudinally-oriented fascicles, with centrally located, elongated nuclei and eosinophilic or fibrillar cytoplasm (reviewed in Guillou and Aurias, 2009; reviewed in Batan et al., 2013). High-grade, poorly differentiated tumors feature cytologic atypia such as pleomorphic nuclei and abnormal mitoses (Figure 2-2) (reviewed in Guillou and Aurias, 2009; reviewed in Fletcher et al., 2013). Smooth muscle actin (SMA), desmin and h-caldesmon are immunohistochemical markers of smooth muscle differentiation and are used to diagnose LMS as they are positive in more than 70% of LMS. (reviewed in Fletcher et al., 2013). SMA is detected in most LMS (Swanson et al., 1991; Oda et al., 2001), desmin is found to be expressed in >50% of LMS (Oda et al., 2001; Hisaoka et al., 2001) and h-caldesmon is present in approximately 40% of LMS; its expression varies with tumor site and degree of differentiation (Hisaoka et al., 2001). De-differentiated areas are usually negative for SMA and desmin (Chen et al., 2001).

Estrogen receptor and progesterone receptor have been noted to be co-expressed in approximately 80% of uterine LMS (Kelley et al., 2004). Historically, gastrointestinal LMS had been confused with GIST due to similar gross and microscopic appearance (Katz et al., 2008; Gladdy et al., 2013); GISTs are distinguished from LMS by the presence of activating mutations in KIT and PDGFRα and positive staining for CD117 (KIT), DOG1 and/or CD34 (Gladdy et al., 2013). Additional diagnostic factors include an assessment of cellular atypia, necrosis and the number of mitoses per high-powered field (reviewed in Batan et al., 2013).
Figure 2-2. Histology of Well-Differentiated to Poorly Differentiated LMS.
Histopathological assessment by Demicco et al.: A. Well-differentiated tumor with intersecting fascicles of spindle cells with eosinophilic cytoplasm. B. Moderately differentiated tumor with variable nuclei and disorganized fascicles. C. Moderately differentiated myxoid tumor with spindle cells in a storiform or fascicular arrangement. D. Moderately differentiated tumor with pleomorphic, multinucleated giant cells in a scattered arrangement. E. Poorly differentiated tumor with rounded to epithelioid cells and loss of fascicular arrangement. F. Poorly differentiated tumor with epithelioid features. (Modified and reprinted from Histopathology, 66, Demicco et al., Progressive loss of myogenic differentiation in leiomyosarcoma has prognostic value, pages 627-638, copyright (2015), with permission from John Wiley & Sons Ltd.: License #3863210139874)
2.3.3 Prognostic Factors, Survival, and Recurrence

2.3.3.1 Site of Disease

Site of disease is one of the most important prognostic factors in LMS, as there are significant clinical and biological differences based on site of disease (reviewed in Weiss et al., 2008). The major site-related subgroups are retroperitoneal/abdominal, extremity, uterine, vascular, and cutaneous LMS. Retroperitoneal/abdominal disease is the most common subgroup, as 50-75% of all LMS arise at these sites (reviewed in Weiss et al., 2008). Retroperitoneal LMS is an aggressive lesion which is characterized by local recurrence and distant metastasis (reviewed in Weiss et al., 2008). The 5-year disease-specific survival (DSS) for retroperitoneal disease is 67%, yet patients continue to succumb to disease over long term (Gladdy et al., 2013). 51% of abdominal/retroperitoneal patients will have at least one local or distant recurrence of LMS over their lifetime (Gladdy et al., 2013). The rate of local recurrence is 21% at 5 years and increases to 27% at 10 years (Gladdy et al., 2013). The rate of distant metastasis in abdominal/retroperitoneal patients is 43% at 5 years and drops to 9% after 5 years, with the most common site of metastasis being the lungs, followed by the liver (Gladdy et al., 2013).

The extremities are another common LMS disease site. In one study of 170 extremity LMS patients, the 5-year DSS for extremity disease was 75%, while disease-related death was not observed after 8 years. 33% of extremity patients will have at least one recurrence of LMS (Gladdy et al., 2013). The rate of local recurrence is 10% at 5 and 10
years for extremity patients (Gladdy et al., 2013). The rate of distant metastasis in extremity patients is 32% at 5 years and drops to 6% after 5 years (Gladdy et al., 2013). Uterine LMS is a rare uterine malignancy and arises from the smooth muscle of the uterine wall. The median OS for patients with uterine LMS is 45 months (Tirumani et al., 2014). Local recurrence occurs in 50% of uterine patients, while distant metastasis occurs in 81.4% of patients after a median interval of 7 months (Tirumani et al., 2014). The most common site of uterine LMS metastasis is the lungs (74%) and a common site of local recurrence is the peritoneum (41%) (Tirumani et al., 2014). Vascular LMS is a rare group of tumors that arise from major vessels such as the inferior vena cava and the large veins of the lower extremity, including the saphenous, iliac and femoral veins (reviewed in Weiss et al., 2008; reviewed in Fletcher et al., 2013). Though they are not classified as vascular LMS, many retroperitoneal and extremity LMS are thought to arise from small to medium size blood vessels such as the renal or gondal vessels (reviewed in Weiss et al., 2008). Cutaneous LMS have a good prognosis because of their superficial location (reviewed in Weiss et al., 2008). Case reports indicate rare sites including the thyroid gland (Mouaqit et al., 2013), gallbladder (Park et al., 2012), base of tongue (Croce et al., 2012), liver (Takehara et al., 2012), kidney (Yodonawa et al., 2012), and the pancreas (Moletta et al., 2012)

2.3.3.2 Histological Grade and Tumor Size

Histological grade is another important prognostic factor in LMS. The most common grading system for STS was created by La Fédération Nationale des Centres de Lutte Contre le Cancer (FNCLCC) (Coindre, 2006). Tumors are graded on a scale from
Grade 1 to Grade 3. The grade is determined based on tumor differentiation, mitotic count and the percentage of necrosis (Table 2-3). Each criterion is assigned a score from one to three for differentiation and mitotic count and from zero to two for necrosis and the cumulative score is used to determine the grade of the tumor. Low-grade tumors are usually differentiated like normal cells and consequently are slower growing. High-grade tumors are aggressive, as they are often poorly differentiated, fast growing and likely to metastasize. Uterine LMS are not graded; uterine smooth muscle tumors with necrosis are classified as leiomyosarcoma. This poses a diagnostic challenge to clinicians, as it is difficult to discriminate between benign leiomyomas and malignant leiomyosarcomas, particularly in the absence of LMS-specific diagnostic markers.

Grade is an independent predictor of DSS and a prognostic factor of distant recurrence (Gladdy et al., 2013; Svarvar et al., 2006). The 5-year DSS for patients with low-grade LMS was 98.5 % (Gladdy et al., 2013). Tumor size is also an important prognostic factor. The median size of LMS is 6.0 cm (range, 0.3-45 cm) (Gladdy et al., 2013). Large tumors (>10 cm) are difficult to excise with clear margins. Size >10 cm is an independent predictor of DSS and a prognostic factor for local and distant recurrence (Gladdy et al., 2013; Svarvar et al., 2006).
2.3.4 Therapeutic Management

2.3.4.1 Surgery, Chemotherapy, and Radiation Therapy

Surgery with wide (>1.0 cm) margins is the principal curative form of treatment for LMS (reviewed in Bathan et al., 2013). Achieving clear margins is difficult when tumors are large and surrounded by vital organs (reviewed in Bathan et al., 2013). In the management of primary LMS, neoadjuvant radiation therapy is used only to treat abdominal disease. Palliative radiation therapy may be used to alleviate symptoms for patients with metastatic disease or unresectable local recurrence (reviewed in Bathan et al., 2013). Systemic chemotherapy, including doxorubicin and ifosfamide in
combination, has marginal efficacy with respect to local recurrence, distant recurrence, overall recurrence and OS (Pervaiz et al., 2008). Currently, a randomized trial is underway to determine if adjuvant chemotherapy can improve OS in patients with high-grade uterine LMS (Clinicaltrials.gov Identifier: NCT01533207). Doxorubicin and ifosfamide are the first line treatment for advanced or metastatic disease (Antman et al., 1993). The less toxic combination of gemcitabine, a nucleoside analog, and docetaxel, an anti-mitotic compound, is an alternate systemic therapy and can be used to treat uterine LMS (reviewed in Bathan et al., 2013). Despite improvements in surgery and systemic therapy for LMS, patients with advanced disease continue to have poor outcomes.

2.3.4.2 Hormonal Therapy

Studies have also demonstrated that LMS can be characterized by hormone-receptor positivity. Specifically, 42% of uterine LMS have been reported to be estrogen receptor (ER) positive, while 41% of uterine LMS was reported to be progesterone receptor (PR) positive (Leitao et al., 2012). In addition, ER and PR were found to be co-expressed in 12 of 15 uterine LMS using IHC (Kelley et al., 2004). Hormone-receptor positivity may be associated with improved outcomes. After stratifying for stage, PR expression in uterine LMS was significantly associated with progression-free survival (PFS; p = 0.002) and overall survival (OS), while ER expression in uterine LMS was significantly associated with PFS (p=0.01) (Leitao et al., 2012). In one study, androgen receptor (AR) was positive in 40% of LMS and found to be predictive of a lower risk of recurrence (p = 0.035) (Leitao et al., 2004). Additionally, ER and/or PR expression was detected in 4 of 16 extrauterine LMS (Kelley et al., 2004). Hormonal therapy may be used for
hormone-receptor positive LMS. One case report noted a response to medroxyprogesterone, a progestational agent, in a patient with metastatic LMS (Uchida et al., 1996). Another study found that aromatase inhibitors achieved a significant clinical benefit rate as a first line and second line regimen in 62.5% and 50% of patients, respectively (Thanopoulou et al., 2014).

2.3.4.3 Targeted Agents

Several targeted agents have been investigated in LMS. Tyrosine-kinase inhibitors have shown promising results. Pazopanib, a selective and multi-targeted tyrosine-kinase inhibitor, has shown to improve PFS in patients with metastatic or unresectable STS, including advanced LMS (van der Graaf et al., 2012).

2.4 Molecular Characteristics of Leiomyosarcoma

2.4.1 Molecular Alterations

LMS has been characterized as a genomically unstable STS. The disease features complex karyotypic alterations, which may differ from one tumor to another (reviewed in Guillou et al., 2009). These alterations include chromosomal gains (chromosomes 1, 5, 6, 8, 15, 16, 17, 19, 20, 22, X), losses (chromosomes 1p, 2, 3, 4, 6q, 8, 9, 10p, 11p, 11q, 12q, 13, 16, 17p, 18, 19, 22q), and amplifications (chromosomes 1, 5, 8, 12, 13,
The most frequent cytogenetic changes are gains of 17p, 8q, and 5p14-pter and losses of 1p12-pter, loss of 2p, loss of 13q14-q21, loss of 10q, and loss of 16q.

There is also a prevalence of alterations in tumor suppressor pathways in LMS. Specifically, loss of 13q (location of RB) was observed in 17 of 29 samples using SKY and CGH (Wang et al., 2001). In an independent study, loss of pRB has been reported in 70% of LMS using IHC (Dei Tos et al., 1996). Decreased p16 expression has been noted in 32% of LMS using IHC (Kawaguchi et al., 2004). p16 is an inhibitor of cyclin-dependent kinase 4 (CDK4), which phosphorylates pRB and results in cell cycle progression. CDK4 has kinase activity when it is bound by Cyclin D. Nuclear Cyclin D3 was positive in 22% of LMS assayed by IHC (Dei Tos et al., 1996). Loss of TP53 has been observed in 43% of LMS samples (Agaram et al., 2015). TP53 missense or frameshift mutations are prevalent in LMS, with reports of mutations in 24-37% of LMS samples (Agaram et al., 2015; Kawaguchi et al., 2004). Gains in 6q and 8q were detected by CGH in 8 of 29 LMS samples; these included gains in 6q24 and 8q24, where the oncogenes MYB and MYC are located (El-Rifai et al., 1998).

The PI3K/AKT/mTOR pathway is known to be active in LMS and is believed to play a role in leiomyosarcomagenesis (Hernando et al., 2007; reviewed in Guillou et al., 2009; reviewed in Yang et al., 2008). Loss of 10q is found in 60-70% of LMS (Agaram et al., 2015; Wang et al., 2001); PTEN, whose protein product negatively regulates the PI3K/AKT pathway by dephosphorylating phosphatidylinositol 4,5-biphosphate (PIP3), is located at 10q23. Partial or complete loss of PTEN protein expression was found in
68% of LMS (Gibault et al., 2012), indicative of PI3K/AKT pathway dysregulation. Insulin signaling can also activate the PI3K/AKT/mTOR pathways. Insulin receptor substrate-2 (IRS2) and phospho-AKT (p-AKT) were present in 78% of LMS samples, suggesting pathway activation (Hernando et al., 2007). Conditional Pten knockout mice developed LMS; these tumors had overexpression of P70S6K and 4E-BP1, demonstrating that the mTOR pathway is also activated in LMS. Recent studies have confirmed PI3K/AKT/mTOR pathway activation in LMS and found that the use of a dual PI3K/mTOR inhibitor can reduce LMS xenograft tumor volume by 68% compared to vehicle controls (Babichev et al., 2016).

Focal amplification of 17p11.2 has been reported in 41-73% of LMS samples using CGH (Italiano et al., 2013; Pérot et al., 2009). Myocardin (MYOCD) is located at 17p11.2 and is overexpressed in LMS with genomic amplification at that locus (Pérot et al., 2009). MYOCD is a transcriptional coactivator which preferentially activates smooth or cardiac muscle reporter genes through its interaction with serum response factor (SRF) (Wang et al., 2003). Specifically, knockdown of MYOCD in an LMS cell line resulted in a decrease of smooth muscle differentiation (Pérot et al., 2009). One study investigating LMS independent of grade found that MYOCD was the most differentially expressed gene in retroperitoneal LMS compared to extremity LMS (Italiano et al., 2013). Thus, MYOCD may play an important role in regulating smooth muscle differentiation in LMS.
2.4.2 Molecular Subtypes

Traditionally, LMS has been classified as a genomically unstable disease with complex karyotypic alterations. A recent genomic study using CGH identified two recurrent genomic profiles: one group of tumors, the “arm” profile group, had few (<30) chromosomal alterations involving gains or losses of full arms or entire chromosomes, while the other group, the “rearranged” profile group, featured >30 chromosomal alterations and thus was more genomically complex (Figure 2-3) (Italiano et al., 2013). A significant correlation was identified between site of disease and genomic profile, as 69% of “arm” profile group tumors were retroperitoneal, while 76% of “rearranged” profile group tumors were located in the extremities (p = 0.02) (Italiano et al., 2013). The majority of genes overexpressed in retroperitoneal LMS encode proteins involved in muscle differentiation and MYOCD is the most overexpressed gene in retroperitoneal LMS as compared extremity LMS.
Figure 2-3. Genomic Profiles of “Rearranged” and “Arm” LMS. A. Genomic profile of an extremity LMS with a “rearranged” profile. B. Genomic profile of a retroperitoneal LMS with an “arm” profile. (Reprinted from Clinical Cancer Research, 2013, 19, pages 1190-1196, Italiano et al., Genetic Profiling Identifies Two Classes of Soft-Tissue Leiomyosarcomas with Distinct Clinical Characteristics, with permission from AACR: License #3863240715801)

In a separate study, unsupervised hierarchical clustering of expression profiles from 51 LMS samples indicated three clusters of samples (Figure 2-4) (Beck et al., 2009). The expression profile of group I tumors was characterized by enrichment of muscle differentiation and function genes (Beck et al., 2009). The gene set most highly and differentially expressed in group II tumors was enriched for genes relating to protein metabolism, regulation of cell proliferation, and organ development (Beck et al., 2009). The gene set most highly and differentially expressed in group III tumors was enriched
for genes relating to organ and system development, proteins involved in wound response and ribosomal proteins involved in protein synthesis (Beck et al., 2009). A later study by the same group validated these subtypes in additional samples and reported that group I is associated with favourable outcomes in extrauterine LMS while group II is associated with poor prognosis in both uterine and extrauterine LMS (Guo et al., 2015). Leiomodin 1 (LMOD1) and ADP-ribosylation factor-like 4C (ARL4C) were identified as diagnostic immunohistochemical markers for group I and group II tumors, respectively (Guo et al., 2015). In order to explore the potential clinical implications of identifying these subtypes, the authors also analyzed differential expression of genes for which targeted therapies are available or under development. A large number of these genes were found to be differentially expressed among the three subtypes (Table 2-4), suggesting that LMS subtypes may respond differently to targeted therapies and may play a role in determining the success rate of therapies in clinical trials (Guo et al., 2015)
Figure 2-4. Unsupervised Hierarchical Clustering of LMS Samples Reveals Three Molecular Subtypes. Clustering was performed on 51 LMS samples with 3038 genes that showed one standard deviation across the samples. Group I tumors are shown in red, group II in blue and group III in green. Five paired primary-metastasis samples are indicated by paired symbols (#, $, &, !, ^). Yellow indicates increased expression, black indicates median expression, blue indicates decreased expression. (Adapted and reprinted by permission from Macmillan Publishers Ltd: Oncogene, 29, Beck et al., Discovery of molecular subtypes in leiomyosarcoma through integrative molecular profiling, pages 1-10, copyright (2009): License # 3863240906676)
Table 2-4. Targets Unique to LMS Subtypes

(Reprinted from Clinical Cancer Research, 2015, 21, pages 3501-3511, Guo et al., Clinically Relevant Molecular Subtypes in Leiomyosarcoma, with permission from AACR: License #3863241069217)

2.5 Gene Fusions

2.5.1 Mechanisms of Formation

Gene fusions are produced by genomic rearrangements that result in the joining of distant DNA sequences (Mani et al., 2010). Genomic rearrangements can occur as a
result of many factors including spatial proximity, cellular stress, DNA sequence features and inappropriate DNA repair or recombination (Figure 2-5). Spatial proximity has been found to correlate with translocation frequency. Sequences undergoing rearrangement need to be in close spatial proximity in the nucleus to become joined (reviewed in Mani et al., 2010). The proximity between loci can be cell type-specific or induced by particular signaling pathways (reviewed in Mani et al., 2010). A DNA break or a switch in the template strand used during DNA replication may also trigger genomic rearrangements. Cellular stresses, encompassing genotoxic, oxidative, and replication stresses, may cause double-stranded DNA breaks and template switching (reviewed in Mani et al., 2010).

Genotoxic stress can result from ionizing radiation or chemical compounds and can lead to random or site-specific double-stranded DNA breaks throughout the genome. Fragile DNA sites include regions that are AT-rich, harbor CGG repeats, contain CpG sites, and are comprised of repetitive elements such as Alu and LINE-1 elements (reviewed in Mani et al., 2010). Oxidative stress is caused by an imbalance in the production and degradation of reactive oxidative species (ROS) which results in increased levels of cellular ROS. Increased levels of superoxide radicals have been directly correlated to genomic rearrangements (reviewed in Mani et al., 2010). Replication stress occurs when inefficient DNA replication results in stalling or collapse of replication forks. Replication stress can be caused by insufficient levels of nucleotide precursors or DNA replication enzymes.
Aberrations in DNA repair pathways are well known molecular mechanisms that generate chromosomal rearrangements. DNA repair pathways are generally homology independent, such as non-homologous end joining (NHEJ), or homology directed, as in homologous recombination. NHEJ is a DNA double-strand break repair mechanism that rejoins two broken DNA ends in a template independent manner. It is an error-prone pathway occurs predominantly at the G1 phase of the cell cycle, making it the most commonly used DNA double stranded break repair pathway in multicellular eukaryotes (reviewed in Mani et al., 2010).

The C-NHEJ pathway repairs intrachromosomal double-stranded breaks, while the error-prone microhomology-dependent A-NHEJ pathway generates chromosomal translocations (reviewed in Mani et al., 2010). Defects in C-NHEJ will promote use of alternative error-prone repair pathways, thereby facilitating chromosomal translocations (reviewed in Mani et al., 2010). Homologous recombination is a template-dependent DNA double-strand break repair mechanism that occurs predominantly in S phase. Non-allelic homologous recombination (NAHR) results from defects in the homologous recombination pathway and may also result in chromosomal rearrangement (reviewed in Mani et al., 2010).
Figure 2-5. Triggers for Genomic Rearrangements. Genomic rearrangements are triggered by combinations of spatial proximity between translocation partners, cellular stress, inappropriate DNA repair or recombination and DNA sequence and chromatin features. The outer rings depict components of each category. (Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Genetics, 11, Mani et al., Triggers for genomic rearrangements: insights into genomic, cellular, and environmental influences, pages 819-829, copyright (2010): License #3863250006401)

Gene fusions can be hundreds to several megabases long and can be the result of insertion (a chromosome segment in a new interstitial position in the same or another chromosome), deletion, duplication, inversion (a rotation of a chromosome segment by 180 degrees) or translocation (the transfer of chromosome segments between chromosomes); combinations of these events result in complex rearrangements (Figure 2-6A).
2.5.2 Functional Consequences of Gene Fusions

Gene fusions may lead to the deregulation of the rearranged genes by the juxtaposition of the coding sequences of one gene with the regulatory sequences of the other gene in the other breakpoint, or by the creation of a chimeric gene through the fusion of parts of the two genes, one in each breakpoint (Figure 2-6B) (reviewed in Mertens et al., 2015).

Gene fusions may also result in truncation of one of the genes, resulting in
haploinsufficiency or dominant-negative isoforms. Functional consequences of genomic rearrangements include gene activation, repression or the formation of fusion proteins that have novel functions (reviewed in Mani et al., 2010).

2.5.3 Gene Fusions as Diagnostic Biomarkers and Therapeutics

Gene fusions can provide growth and survival advantages to cancer cells (reviewed in Mani et al., 2010). The correlation between recurrent gene fusions and tumor subtype makes gene fusions ideal for diagnostic purposes (reviewed in Mertens et al., 2015). Gene fusions are also important for treatment stratification, as morphologically homogeneous disease may respond differently to treatment based on positivity for gene fusions, as demonstrated by PAX3-FOXO1 positive versus negative ARMS (Williamson et al., 2010). Some chimeric proteins, directly or indirectly, are excellent treatment targets (Table 2-5). In dermatofibrosarcoma protuberans (DFSP), a dermal sarcoma, the COL1A1-PDGFB translocation is identified in more than 90% of tumors and deregulates expression of PDGFB (McArthur et al., 2005). DFSP has a high response rate to imatinib, a tyrosine kinase inhibitor that can be used to neoadjuvantly downstage tumors by targeting the activated PDGFR or to systematically treat metastatic disease (McArthur et al., 2005; Labropoulos and Razis, 2007). Thus, identifying disease-specific gene fusions can be critical for improving patient outcomes. Gene fusions have been detected in one-third of STS subtypes and half of these events are recurrent (reviewed in Mertens et al., 2016). Traditionally, it has been thought that gene fusions are a feature of genomically stable STS, such as dermatofibrosarcoma protuberans (DFSP)
and absent in genomically unstable STS, including liposarcoma and leiomyosarcoma. However, next-generation sequencing based studies of cancers with complex genomic profiles has shown that pathologically important gene fusions may be found in genomically unstable STS (reviewed in Mertens et al., 2015).

### Table 2-5. Targeted Agents that Produce Responses in Sarcoma.

<table>
<thead>
<tr>
<th>Agent</th>
<th>Tumour types</th>
<th>Known alteration or recognized or proposed target</th>
<th>Approximate response rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imatinib</td>
<td>Dermatofibrosarcoma protuberans</td>
<td>COL1A1–PDGFβ fusions</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Tenosynovial giant cell tumour and pigmented villonodular synovitis</td>
<td>COL6A3–CSF1 fusions</td>
<td>High</td>
</tr>
<tr>
<td>Sunitinib</td>
<td>Alveolar soft-part sarcoma</td>
<td>ASPL–TFE3 fusion</td>
<td>Unknown</td>
</tr>
<tr>
<td>Crizotinib</td>
<td>Inflammatory myofibroblastic tumour</td>
<td>ALK fusions</td>
<td>Unknown</td>
</tr>
<tr>
<td>MET inhibitor</td>
<td>Alveolar soft-part sarcoma</td>
<td>ASPL–TFE3 fusion</td>
<td>Unknown</td>
</tr>
<tr>
<td></td>
<td>Clear-cell sarcoma</td>
<td>EWS–ATF1 fusions</td>
<td>Unknown</td>
</tr>
<tr>
<td>IGF1R antibody</td>
<td>Ewing’s sarcoma</td>
<td>EWS–FLI1 or EWS–ERG</td>
<td>10–15%</td>
</tr>
<tr>
<td>Bevacizumab</td>
<td>Angiosarcoma</td>
<td>VEGFR2 mutations; VEGF or its receptors</td>
<td>~15%</td>
</tr>
<tr>
<td>Cediranib</td>
<td>Alveolar soft-part sarcoma</td>
<td>ASPL–TFE3 fusion</td>
<td>Unknown</td>
</tr>
</tbody>
</table>


### 2.6 Overview of PI3K/AKT/mTOR Pathway

### 2.6.1 PI3K/AKT/mTOR Pathway

The PI3K/AKT/mTOR pathway (Figure 2-7) is a critical signaling pathway in eukaryotic cells which regulates cell proliferation and survival (reviewed in Vivanco et al., 2002).
This pathway is dysregulated in several solid cancers and has become a focus of the development of targeted therapies (reviewed in Nitulescu et al., 2015). The PI3K/AKT/mTOR pathway can be activated by signals, including hormones, growth factors and extracellular matrix proteins (reviewed in Nicholson et al., 2002). Receptor tyrosine kinases (RTKs), such as type-1 and type-2 insulin-like growth factor receptors (IGFR), activate class I phosphatidylinositol 3-kinase (PI3K) through binding or tyrosine phosphorylation of adaptor proteins, such as IRS1 or IRS2, which then bind and activate PI3K. Phosphatidylinositol-4,5-bisphosphate (PIP₂) is phosphorylated by PI3K to generate phosphatidylinositol-3,4,5-trisphosphate (PIP₃). PIP₃ can be dephosphorylated to PIP₂ by the phosphatase PTEN, which is a negative regulator of the pathway. AKT binds to PIP₃ at the plasma membrane and is phosphorylated at Thr308 by PDK1. mTOR complex 2 (mTORC2), which is also activated by RTK signaling, phosphorylates AKT at Ser473, which can be dephosphorylated by PHLPP. mTOR complex 1 (mTORC1) is activated by AKT through phosphorylation of TSC2 at S939 and T1462. This blocks TSC2 from acting as a GTPase and allows Rheb-GTP to accumulate. Rheb-GTP activates mTORC1, which phosphorylates downstream targets such as 4E-BP1 and P70S6K. mTORC1 is also activated by AKT phosphorylation of PRAS40, an inhibitor of mTORC1. Activated mTORC1 and P70S6K can phosphorylate serine residues on IRS1, causing IRS1 degradation. This acts as a negative feedback mechanism to decrease PI3K/AKT signaling (reviewed in Manning and Cantley, 2007).
2.6.2 PI3K

2.6.2.1 Structure and Function of PI3K

There are three classes of PI3K. Class I PI3K consist of two subgroups, IA and IB; class IA PI3K transmit signals from tyrosine kinases while class 1B PI3K transmit signals from G-protein coupled receptors. Here, the focus will be Class IA PI3K, as it is involved in oncogenesis (reviewed in Vivanco et al., 2002). Class IA PI3Ks catalyze the
phosphorylation of PIP2 at the 3-position. PI3K is a heterodimer consisting of a regulatory and catalytic subunit. p85, the regulatory subunit, can directly associate with RTKs through the physical interaction of its SRC homology 2 domain (SH2) with phosphotyrosine residues on the RTK or through intermediate phosphoproteins, such as IRS1 and IRS2 (reviewed in Vivanco et al., 2002). p110, the catalytic subunit, is found in a complex with p85. p110 is activated when p85 interacts with the RTK, as it is in close proximity to its lipid substrates in the cell membrane (reviewed in Vivanco et al., 2002). Mammals contain three different genes for the p85 regulatory subunit (p85α, p85β, and p55γ) and three different genes for the p110 catalytic subunit (p110α, p110β, and p110δ).

2.6.2.2 PI3K Alterations in Cancer

PI3K alterations have been noted in numerous cancers. Alterations detected in cancer include p110α amplification and p85α mutation in ovarian cancer, PI3K pathway activation in breast cancer, and p85α mutation in colon cancer (reviewed in Vivanco et al., 2002). PIK3CA gene amplifications, deletions, and missense mutations have been reported in many cancers of the colon, breast, brain, liver, stomach and lung (reviewed in Karakas et al., 2006). PIK3CA missense mutations are believed to increase the kinase activity of PIK3CA contributing to cellular transformation. PIK3CA mutations were found to occur in 18% of myxoid/round-cell liposarcomas and were associated with AKT activation and shorter duration of disease-specific survival than those with wild-type PIK3CA (Barretina et al., 2010).
2.6.3 AKT

2.6.3.1 Structure and Activation of AKT

AKT, also known as protein kinase B, is a serine/threonine kinase with three isoforms: AKT1, AKT2, and AKT3 (reviewed in Nitulescu et al., 2015). AKT consists of a pleckstrin homology (PH) domain that interacts with PIP$_3$, a central kinase domain containing a Thr308 residue, and a carboxy-terminal regulatory domain containing a Ser473 domain (reviewed in Nitulescu et al., 2015). AKT plays a critical role in regulating cell growth, proliferation, survival, transcription and protein synthesis (reviewed in Nitulescu et al., 2015). AKT can be phosphorylated at two regulatory sites: Thr308 and Ser473. AKT and PDK1 bind to PIP$_3$ at the plasma membrane, and PDK1 phosphorylates the activation loop of AKT at Thr308 (reviewed in Manning and Cantley, 2007). AKT can also be phosphorylated at Ser473 by mTORC2. The kinase activity of AKT is regulated by the phosphorylation at Thr308 and Ser473.

2.6.3.2 Nuclear Localization of Activated AKT

Generally, AKT and its activated, phosphorylated form p-AKT (Thr308 and Ser473) are located in the cytoplasm. However, p-AKT has been reported to translocate to the endoplasmic reticulum, the mitochondria, the Golgi apparatus, and the nucleus (reviewed in Martelli et al., 2012). Some p-AKT substrates, including the FOXO family of transcription factors, are located in the nucleus and are phosphorylated by p-AKT. AKT
has been reported to translocate to the nucleus in response to a stimuli such as IGF-1, hypoglycemia, insulin, and nerve growth factor (NGF) (reviewed in Martelli et al., 2012).

2.6.3.3 Downstream Effects of AKT Activation

AKT plays diverse roles in cellular function through the substrates it phosphorylates. The physiological responses downstream of AKT are mediated by multiple phosphorylation targets. For example, AKT promotes cellular survival by activating MDM2 and inhibiting GSK3, FOXO, BAD and CASP9 (reviewed in Manning and Cantley et al., 2007). Furthermore, some AKT substrates can activate multiple cellular functions. AKT-phosphorylated TSC2 and PRAS40 promote cell growth, angiogenesis and metabolism (reviewed in Manning and Cantley et al., 2007). AKT activation has biological consequences relevant to cancer cell growth, as it promotes survival, proliferation and growth (reviewed in Vivanco et al., 2002). In addition, AKT affects tumor angiogenesis through hypoxia-inducible factor-1 (HIF-1) and vascular endothelial growth factor (VEGF) (reviewed in Vivanco et al., 2002). Elevated AKT1 kinase activity has been reported in ovarian, breast, and thyroid cancer (Sun et al., 2001; Ringel et al., 2001).
2.6.4 mTOR

Mammalian target of rapamycin (mTOR) is a serine/threonine kinase that regulates cell growth and protein synthesis on the basis of the availability of nutrients (reviewed in Vivanco et al., 2002). mTOR activity is modulated by AKT through inhibition of TSC1/TSC2 and PRAS40. mTOR enhances mRNA translation by activating P70S6K and inhibiting 4E-BP1, which is a translational repressor (reviewed in Vivanco et al., 2002). mTOR is noted to be active in cancer cells. mTOR signaling was found to be required for sarcomagenesis in conditional PTEN knockout mice that developed LMS (Hernando et al., 2007). Recent studies have identified a PI3K/mTOR inhibitor which reduces LMS xenograft tumor volume by 68% (Babichev et al., 2016). CCI-779, an analog of the mTOR inhibitor rapamycin, has been used in phase III clinical trials to determine its effect on time to progression and overall survival in patients with advanced renal-cell carcinoma (reviewed in Bjornsti et al., 2004).

2.7 Cell of Origin of Leiomyosarcoma

Since LMS is composed of cells featuring smooth muscle differentiation, it is believed that LMS arises from smooth muscle cells (de Graaf et al., 2015; Guo et al., 2015; reviewed in Bathan et al., 2013; Hernando et al., 2007). Traditionally, mature somatic cells have been considered the cell-of-origin of their corresponding tumors. More recently, progenitor or stem cells have been considered as the target of neoplastic
transformation, based on the observation that tumor cells need to accumulate genetic and phenotypic alterations over extended time periods (reviewed in Xiao et al., 2013; reviewed in Pardal et al., 2003). One study reported that p53-deficient mesenchymal stem cells (MSC) gave rise to LMS-like tumors after injection into immunodeficient mice (Rubio et al., 2010). It may be possible that aberrations in the normal differentiation process of these progenitors occur at early stages of tumorigenesis. A group of 20 miRNAs were found to be required for smooth muscle cell (SMC) differentiation of human MSCs and these miRNAs were similar to those found in uterine LMS (Danielson et al., 2010). In contrast, uterine leiomyomas had similar miRNA profiles to those of mature SMCs and myometrium, suggesting that uterine LMS and uterine leiomyomas may have different origins and that uterine LMS may originate from mesenchymal stem cells (Danielson et al., 2010). However, overexpression of MYOCD in three undifferentiated pleomorphic sarcoma (UPS) cell lines and a liposarcoma (LPS) cell line results in smooth muscle differentiation and increased cell migration, suggesting that LMS may not necessarily derive from smooth muscle cells, but could stem from undifferentiated mesenchymal cells driven to smooth muscle differentiation.
2.8 Overview of Smooth Muscle Differentiation

2.8.1 Smooth Muscle Cells

SMCs are elongated, spindle-shaped cells that form involuntary, non-striated muscle. They constitute the contractile component of hollow organs, including blood vessels, esophagus, colon, and bladder, and the myometrium of the uterine wall (reviewed in Owens et al., 2004). While cardiac and skeletal muscles terminally differentiate, SMCs maintain their ability to modulate their phenotype from contractile to synthetic in response to growth factors and environmental cues (Wang et al., 2011).

2.8.1.1 Molecular Characteristics

SMCs are characterized by the expression of genes including alpha smooth muscle actin (ACTA2), desmin (DES), transgelin (TAGLN), myosin light-chain kinase (MYLK), myosin heavy chain 11 (MYH11), caldesmon 1 (CALD1), and calponin 1 (CNN1) (Robin et al., 2013; Liu et al., 2005; Wang et al., 2003). Smooth muscle genes are regulated by serum response factor (SRF), a MADS box transcription factor that binds to the DNA consensus sequence CC(A/T)\_6GG (CArG) (Wang et al., 2003). Smooth muscle genes contain two CArG boxes, which are both bound by homodimers of SRF and act cooperatively to regulate smooth muscle gene expression (Wang et al., 2003). MYOCD is the transcriptional coactivator of SRF and thereby is the master regulator of smooth muscle gene expression (Wang et al., 2003).
2.8.1.2 Contractile and Synthetic Phenotypes

SMCs transition between a quiescent, contractile phenotype and a proliferative, synthetic phenotype in response to physiological stimuli (Liu et al., 2005). The modulatory nature of the SMC phenotype is crucial to their function in vascular remodeling and development. SMCs express contractile proteins required to maintain cardiovascular homeostasis and can de-differentiate and proliferate in response to vascular injury (reviewed in Owens et al., 2004). Abnormal SMC proliferation and migration contributes to the pathogenesis of atherosclerotic lesions, hypertension and potentially the development of LMS (reviewed in Owens et al., 2004).

2.8.2 Regulation of SMC Differentiation

The PI3K/AKT, extracellular signal-regulated kinase (ERK), and p38 mitogen-activated protein kinase (MAPK) signaling pathways are known to be involved in controlling SMC phenotypes (Wang et al., 2003). Specifically, IGF-1 stimulates differentiation of SMCs by activating PI3K/AKT signaling. In turn, AKT phosphorylates forkhead box protein 04 (FOXO4) at Thr32, Ser197 and Ser262, which results in translocation of FOXO4 from the nucleus to the cytoplasm. FOXO4 represses SMC differentiation by forming a ternary complex with MYOCD and SRF and inhibiting the activity of MYOCD (Liu et al., 2005). Nuclear exclusion of FOXO4 results in MYOCD activity and SMC differentiation (Liu et al., 2005). Thus, FOXO4 is a signal-dependent regulator of SMC differentiation (Figure 2-8).
**Figure 2-8. Regulation of Smooth Muscle Differentiation.** IGF signaling activates the PI3K/AKT pathway. Once AKT is phosphorylated at threonine 308, it enters the nucleus, where it phosphorylates FOXO4 at threonine 32, serine 197 and serine 262. Phosphorylated FOXO4 is transported out of the nucleus and cleaved or targeted for ubiquitination and proteasome-mediated degradation. Once FOXO4 is excluded from the nucleus, MYOCDE transcriptionally coactivates SRF and smooth muscle genes are expressed (Liu et al., 2005).
2.8.3 MYOCD

2.8.3.1 Structure and Function of MYOCD

MYOCD belongs to the SAP domain family of transcription factors (reviewed in Miano et al., 2015). The amino-terminus binds to SRF and the carboxy-terminus mediates strong transcriptional activity (Wang et al., 2001). Dimerization of MYOCD through a leucine zipper motif exposes the carboxy-terminus transcription activating domains (TAD), resulting in transcriptional activation (Wang et al., 2003). MYOCD is both necessary and sufficient for the development and differentiation of most SMCs (Wang et al., 2001). MYOCD is the transcriptional coactivator of SRF and primarily functions as the master regulator of cardiac and smooth muscle differentiation (reviewed in Miano et al., 2015). It preferentially activates smooth and cardiac muscle reporter genes controlled by pairs of CArG boxes through its interaction with SRF (Wang et al., 2003). In addition, MYOCD has been reported to repress skeletal muscle differentiation (Miano et al., 2015).

2.8.3.2 Roles of MYOCD in Cancer

The role of MYOCD in cancer may be different depending on the disease context. Inactivation of MYOCD has been associated with malignant tumor growth (Milyavsky et al., 2007). For example, MYOCD expression levels are decreased in several nasopharyngeal carcinoma cell lines and this decrease was correlated with hypermethylation of its promoter (Chen et al., 2011). Treatment of these cells with 5-
azacytidine, a demethylating agent, increased MYOCD expression. Thus, MYOCD may act as a tumor suppressor in nasopharyngeal carcinoma (reviewed in Miano et al., 2015). Additionally, MYOCD has been reported to act as a growth inhibitor in uterine LMS (reviewed in Miano et al., 2015). MYOCD was found to be highly downregulated in uterine LMS cell lines, which have a de-differentiated smooth muscle phenotype (Kimura et al., 2010). Exogenous MYOCD expression resulted in expression of smooth muscle genes and upregulation of p21, a cell cycle regulator (Kimura et al., 2010). Mutation of the CArG box sequence of the p21 promoter eliminated the upregulation of p21 in response to exogenous MYOCD expression, suggesting that MYOCD transactivates the p21 promoter and thereby may have a tumor suppressive role in uterine LMS (Kimura et al., 2010). In contrast, MYOCD was reported to be highly amplified and overexpressed in 53% of well-differentiated LMS (Pérot et al., 2009). An independent study reported MYOCD amplification in 13% of LMS samples, which included both retroperitoneal and extremity samples (Agaram et al., 2015). Inactivation of MYOCD in an LMS cell line led to a decrease in smooth muscle differentiation and reduced cell migration (Pérot et al., 2009). Thus, the function of MYOCD in LMS may be different depending on the site of disease and suggests that uterine LMS is biologically distinct from LMS of other sites.
2.8.4 FOXO4

2.8.4.1 Function of FOXO4

Class O of forkhead box transcription factors (FOXO) consists of FOXO1, FOXO3, FOXO4 and FOXO6 (reviewed in van der Horst and Burgering, 2007). FOXO transcription factors bind to the conserved consensus sequence 5’TTGTTTAC3’ at helix three of the DNA-binding domain (reviewed in van der Horst and Burgering, 2007). The expression of FOXO-regulated genes can be modulated by any member of the FOXO transcription factor family; specificity is obtained either by the specific expression pattern of each isoform or by isoform-specific regulation (reviewed in van der Horst and Burgering, 2007). FOXO proteins interact with DNA and other proteins to regulate transcription of specific genes through numerous mechanisms, such as recruiting transcriptional co-activators or cooperating DNA-binding transcription factors and repressing transcription by competing with transcription factors for a binding site on a gene’s promoter (reviewed in Glauser and Schlegel, 2007). FOXO proteins contribute to regulation of cell cycle progression, proliferation, cell size determination, apoptosis, cell differentiation, reactive oxidative species detoxification, and metabolism (reviewed in Tzivion et al., 2011; reviewed in Calnan et al., 2008).
2.8.4.2 Subcellular Localization and Regulation of FOXO4

FOXO protein activity is regulated by phosphorylation, which results in shuttling of FOXO proteins between the nucleus and the cytoplasm (reviewed in Glauser and Schlegel, 2007). FOXO proteins are phosphorylated in response to growth factors including IGF-1, erythropoietin, EGF and nerve growth factor. Phosphorylation of FOXO4 by AKT at Thr32, Ser197, and Ser262 results in exclusion from the nucleus via binding to 14-3-3 proteins (Tzivion et al., 2011). The exact subcellular location of phosphorylation by AKT has not been clearly elucidated; there is evidence for both nuclear and cytoplasmic phosphorylation, though activated AKT is found primarily in the cytoplasm (Manning and Cantley et al., 2007). Phosphorylation of FOXO proteins by AKT leads to reduced FOXO transcriptional activity and binding with other target proteins. Phosphorylation at Thr32, Ser197, and Ser262 by serum/glucocorticoid regulated kinase 1 (SGK1) also leads to the exclusion of FOXO4 from the nucleus (Glauser and Schlegel, 2007; van der Horst and Burgering, 2007). Phosphorylation of FOXO4 by c-Jun N-terminal kinases (JNK) at T447 and T451 results in translocation to the nucleus and increased transcriptional activity (Tzivion et al., 2011). Furthermore, MDM2 binds phosphorylated FOXO proteins in the cytoplasm and induces polyubiquitination and ultimately protein degradation, thereby regulating levels of FOXO (Calnan and Brunet, 2008). Subcellular localization of FOXO4 has been reported to regulate the expression of smooth muscle marker genes; nuclear FOXO4 forms a complex with MYOCD and SRF which results in the repression of smooth muscle gene expression, while translocation of FOXO4 to the cytoplasm results in the transcriptional activity of MYOCD and SRF and the expression of smooth muscle genes (Liu et al.,
2005). Additionally, siRNA-mediated knockdown of FOXO4 in a rat aortic smooth muscle cell line has been reported to downregulate smooth muscle marker expression (Liu et al., 2005), further suggesting a regulatory role for FOXO4 in smooth muscle differentiation.
Chapter 3  Thesis Aims and Hypotheses
3.1 Thesis Aims

Leiomyosarcoma (LMS) has been described as a genomically unstable cancer. It has been characterized by multiple genomic aberrations within a single tumor and heterogeneity of these aberrations between tumors. As a result, there is a limited understanding of the genetic drivers, molecular subtypes and dysregulated developmental pathways in LMS. Defining molecular events in LMS may have implications for diagnosis and disease management. The primary aims of this thesis are to uncover the molecular characteristics of LMS and to elucidate the mechanism(s) regulating smooth muscle differentiation in LMS. Three discrete studies were performed to investigate the following hypotheses.

3.2 Hypotheses

3.2.1 YAF2-PRICKLE1 and ESR1-CCDC92 Gene Fusions are Recurrent in LMS

Gene fusions are found in one-third of soft tissue sarcomas (STS) (reviewed in Mertens et al., 2016). Gene fusions can be valuable diagnostic tools for refining STS subtype stratification and may facilitate targeted therapeutics to improve patient outcomes. Next-generation sequencing of genomically complex cancers has identified that gene fusions may be also found in genomically unstable STS (reviewed in Mertens et al., 2015). Preliminary transcriptome analysis performed by the Gladdy and Wrana Laboratories at
the Lunenfeld-Tanenbaum Research Institute identified two novel fusion transcripts, *YAF2-PRICKLE1* and *ESR1-CCDC92*, in a primary LMS cell line derived from a metastatic buttock lesion. *YAF2*, *PRICKLE1*, and *ESR1* have noted roles in cancer biology. Since these fusion transcripts are present in an LMS cell line, I hypothesized that they are recurrent in LMS. In order to test this hypothesis and confirm the presence of the fusion transcripts in the index case, we analyzed the transcriptomes of additional LMS patient samples for the expression of these gene fusions.

### 3.2.2 LMS Transcriptome Profiles Cluster by Site

Recent genomic studies have identified two subtypes of LMS: a genomically stable subtype comprised primarily of retroperitoneal tumors and a genomically unstable subtype comprised primarily of extremity tumors (Italiano *et al.*, 2013). An independent study of expression profiles from 51 LMS samples identified three subtypes of LMS which were not site-specific (Guo *et al.*, 2015). Site is one of the most important prognostic factors in assessing outcome for patients with LMS. Patients with extremity tumors have better outcomes than those with retroperitoneal tumors; retroperitoneal tumors have higher rates of recurrence than extremity tumors (51% and 33%) and have worse long-term disease-specific survival (DSS) compared to extremity or trunk (Gladdy *et al.*, 2013). Since there may be site-associated genomic subtypes of LMS, I hypothesized that there may be site-associated transcriptomic subtypes of LMS. Transcriptome profiles from a discovery set of LMS patient samples underwent unsupervised hierarchical clustering to determine if they cluster by site. We also
investigated differential gene expression and gene set enrichment of these clusters to identify the molecular characteristics of these clusters.

3.2.3 FOXO4 Expression and Phosphorylation Regulates Smooth Muscle Differentiation in LMS

Traditionally, LMS has been thought to arise from smooth muscle cells. However, recent evidence suggests that LMS may arise from other cell types, including mesenchymal stem cells (Danielson et al., 2010; Pérot et al., 2009), and that dysregulation of smooth muscle differentiation may be associated with leiomyosarcomagenesis. Myocardin (MYOCD), the master regulator of smooth muscle gene expression in smooth muscle and cardiac muscle, is amplified and overexpressed in well-differentiated LMS (Italiano et al., 2013) and downregulated in de-differentiated LMS (Kimura et al., 2010). A separate study found that knockdown of MYOCD in an LMS cell line resulted in a decrease of smooth muscle differentiation (Pérot et al., 2009). Studies using rat aortic smooth muscle cells elucidated that IGF-mediated PI3K/AKT signaling results in the phosphorylation and nuclear exclusion of FOXO4, thereby negating the inhibition of MYOCD transcriptional activity by FOXO4 and resulting in smooth muscle differentiation (Liu et al., 2005). Furthermore, knockdown of FOXO4 has been reported to result in increased expression of smooth muscle markers (Liu et al., 2005). I hypothesized that FOXO4 expression and phosphorylation may regulate smooth muscle differentiation in LMS. To investigate this, we characterized smooth muscle differentiation, PI3K/AKT pathway activation and MYOCD and FOXO4 expression in LMS cell lines and patient samples.
Chapter 4  Investigating Novel Gene Fusions and Molecular Subtypes in LMS
4.1 Abstract

Leiomyosarcoma (LMS) is a smooth muscle tumor and one of the most common subtypes of soft tissue sarcoma (STS) (reviewed in Weiss et al., 2008). Currently, there is a limited understanding of the genetic drivers of LMS. Additionally, it is not clear if clinical features of LMS, such as site of disease, may be associated with particular molecular characteristics. Several studies have challenged the traditional characterization of LMS as a genetically heterogeneous disease through the identification of potential molecular subtypes (Guo et al., 2015; Italiano et al., 2013). Furthermore, a novel recurrent translocation has recently been reported in LMS (de Graaff et al., 2015). A preliminary examination by our group identified YAF2-PRICKLE1 and ESR1-CCDC92 fusion transcripts in a primary LMS cell line. Here, we investigated if 1) YAF2-PRICKLE1 and ESR1-CCDC92 fusion transcripts are recurrent in LMS and 2) LMS transcriptomes cluster by site of disease. RNA-Seq and RT-PCR were performed on RNA extracted from 15 LMS patient samples and the transcriptomes were analyzed for fusion transcripts. Unsupervised hierarchical clustering analysis was performed on the transcriptomes to determine if they cluster by site. YAF2-PRICKLE1 and ESR1-CCDC92 were confirmed in the primary LMS cell line they were discovered in using RT-PCR. Spanning reads with the YAF2-PRICKLE1 and ESR1-CCDC92 sequences were detected in the tumor from which the cell line was derived, but neither fusion transcript was classified as a fusion event based on the parameters used for fusion detection with RNA-Seq. YAF2-PRICKLE1 and ESR1-CCDC92 were not recurrent in the additional LMS patient samples tested. Cluster analysis of the transcriptomes from the LMS samples revealed that abdominal and extremity samples
form a cluster distinct from uterine samples. Differentially expressed genes between the
clusters were analyzed for gene set and functional annotation clustering enrichment and
several statistically significant clusters were identified. In summary, YAF2-PRICKLE1
and ESR1-CCDC92 are not recurrent in the LMS patient samples assayed in this study,
though this does not preclude them from being rare events which may be recurrent in a
larger sample set. Furthermore, these findings suggest that abdominal and extremity
LMS may be molecularly distinct from uterine disease.
4.2 Introduction

LMS is characterized as a genomically unstable cancer with complex chromosomal alterations (reviewed in Yang et al., 2009). To date, the known molecular alterations in LMS include alterations in the p53 and RB tumor suppressor pathways and activation of the PI3K/AKT/mTOR pathway (reviewed in Yang et al., 2009; Hernando et al., 2007); however, there is a limited understanding of the molecular subtypes in this disease. In a recent study of 68 LMS patient samples, two LMS subtypes were identified based on comparative genomic hybridization (CGH) and gene expression profiles (Italiano et al., 2013). One subtype, consisting of extremity tumors, is characterized by chromosomal complexity and features many (>30) chromosomal alterations. The other subtype, consisting of retroperitoneal tumors, has a simpler genomic profile, featuring fewer (<30) chromosomal alterations that involve full arm or entire chromosomal gain or loss (Italiano et al., 2013). Sarcomas with simple genomic alterations may include disease-specific gene fusions. Furthermore, the retroperitoneal subgroup was characterized by overexpression of muscle differentiation genes, whereas the extremity subgroup had overexpression of genes involved in extracellular matrix, wounding healing and adhesion pathways (Italiano et al., 2013). These findings may suggest that (a) a subgroup of LMS is characterized by recurrent gene fusions and (b) LMS molecular subtypes may exist which correlate with site of disease.

Recently, a recurrent t(6;14) translocation was identified in two LMS using combined binary ratio labeling fluorescence in situ hybridization (COBRA-FISH), though this event was not identified in 37 additional LMS samples using a tissue microarray (de Graaff et
In a preliminary investigation, the Gladdy Lab discovered two novel fusion transcripts (YAF2-PRICKLE1 and ESR1-CDC92) in a primary LMS cell line using RNA Sequencing (RNA-Seq) conducted and analyzed in collaboration with the Wrana Lab at the Lunenfeld-Tanenbaum Research Institute (Figure 4-1). YY1-associated factor 2 (YAF2) (Madge et al., 2003), prickle-like protein 1 (PRICKLE1) (Kaucka et al., 2013), and estrogen receptor 1 (ESR1) (Leitao et al., 2012; Leitao et al., 2004; Kelley et al., 2004) have known biological functions that contribute to cancer development and progression. YAF2 is a negative regulator of muscle-restricted genes, PRICKLE1 is involved in WNT planar cell polarity signaling and ESR1 is overexpressed in LMS, while coiled-coil domain containing 92 (CCDC92) is a protein that is localized to the cytoskeleton with unknown function.
Figure 4-1. Two Novel Fusion Transcripts are Detected in a Primary LMS Cell Line using RNA-Seq. Analysis of RNA-Seq data identified a fusion event between exon 4 of YAF2 with exon 1 of PRICKLE1. Both exons are located on 12q12 approximately 300,000 bp apart. The second fusion transcript linked exon 2 and a downstream intronic region of ESR1 with exon 1 of CCDC92, which are located on 6q25.1 and 12q24.31, respectively. SK-LMS-1, STS210, STS39 and STS54 are LMS cell lines. STS256, STS123, STS107, STS269 are dermatofibrosarcoma protuberans (DFSP) cell lines. DFSP cell lines were used as a positive control for fusion detection, as >90% of DFSP samples are positive for the COL1A1-PDGFβ gene fusion.

Three LMS molecular subtypes were identified in a study which performed unsupervised hierarchical clustering analysis of expression profiles from 51 LMS samples (Beck et al., 2009). A later study by the same group utilized 3SEQ, a modified RNA-Seq protocol in which only the 3'-ends of genes are sequenced, to confirm these three LMS subtypes and externally validated these findings using TCGA expression...
data (Guo et al., 2015). The limitations of this study included the absence of clustering by clinical parameters such as site of disease and the lack of differential expression of biologic pathways to distinguish the subgroups. These limitations may have been due to limited clinical annotation, as site was unknown for 20% of samples, and the nature of the transcriptomic analysis performed.

Clinical observations suggest that site of disease is one of the most important prognostic factors in LMS. Patients with extremity tumors have better outcomes than those with retroperitoneal tumors; retroperitoneal tumors have higher rates of recurrence than extremity tumors (51% and 33%) and have worse long-term disease-specific survival (DSS) compared to extremity or trunk (Gladdy et al., 2013). However, whether there is a molecular basis for the site-associated differences in outcome is not well established.

Since YAF2-PRICKLE1 and ESR1-CDC92 fusion transcripts are present in a cell line derived from an LMS tumor and genomic and transcriptomic subtypes of LMS have been reported, I hypothesized that (1) YAF2-PRICKLE1 and ESR1-CDC92 are recurrent in LMS and (2) LMS transcriptomes may cluster by site of disease. We performed RNA-Seq on 15 LMS patient samples, including the tumor from which the fusion-positive cell line was derived, and analyzed the transcriptomes for expression of the gene fusions. We confirmed the detection of gene fusions using RT-PCR. In addition, we performed unsupervised hierarchical clustering analysis of the transcriptomes from these patient samples to see if they group by site.
4.3 Materials and Methods

4.3.1 Tumor Samples

Formalin-fixed, paraffin-embedded (FFPE) tumors were obtained from the Biospecimen Repository and Processing Lab (Mount Sinai Hospital, Toronto, CA). Histologic slides from all FFPE samples were evaluated and scored by a dedicated sarcoma pathologist at Mount Sinai Hospital. 7 abdominal, 4 extremity and 4 uterine tumors were classified as LMS. Samples were scored for degree of differentiation, amount of necrosis, and mitotic rate in accordance with the FNCLCC grading system (Coindre, 2006). Total RNA was isolated from the FFPE samples using an RNeasy Mini Kit (Qiagen, Hilden, DEU). RNA yield and quality were assessed using Nanodrop (Thermo Fisher Scientific, Waltham, USA) and BioAnalyzer (Agilent Technologies, Santa Clara, USA).

4.3.2 Cell Line Establishment and Culture Conditions

Research Ethics Board consent was obtained to derive cell lines from resected LMS tumors STS210 (metastatic buttock) and STS39 (paravaginal) and DFSP tumors STS256, STS123, STS107 and STS269. Primary cell lines were derived from these tumors by manual dissociation of patient samples which were incubated overnight at 37°C in supplemented DMEM/F-12 10% FBS (Life Technologies, Carlsbad, USA) with 2 μg/ml collagenase (Sigma-Aldrich, St. Louis, USA). Cell suspensions were centrifuged at 1000 rpm for 5 min and plated on T75 flasks. SKLMS1 cells were obtained from
ATCC and cultured in DMEM (Life Technologies), 10% FBS media. RNA was extracted using the RNeasy Mini Kit and RNA yield and quality were assessed by Nanodrop.

4.3.3 RNA-Sequencing (RNA-Seq)

200 ng of total RNA was used per sample for RNA-Seq. rRNAs were removed from the total RNA of FFPE samples and cell lines using Ribo-Zero rRNA Removal Kit (Illumina, San Diego, USA). rRNA removal was indicated by a GAPDH:S18 ratio >1. cDNA libraries were prepared using TruSeq RNA Library Prep Kit v2 (Illumina). Libraries displaying successful removal of rRNA were then loaded as 2 indexed samples per lane on a HiSeq 2000 (Illumina) for RNA-Seq (Liu et al., 2014). Raw sequence reads were obtained in FASTQ format using standard CASAVA software (Illumina). Standard algorithms for RNA-Seq data production and quality control were used with an average read depth of approximately 100 million reads per sample.

4.3.4 Gene Fusion Detection

FASTQ format raw reads were mapped onto human genome build 19 (UCSC) using TopHat 2.0.7 (JHU) fusion junction mapper with Bowtie 1.01. The following parameters were used for fusion detection: a minimum of 1 spanning read, 2 paired-end reads, and 5 paired-end reads with one end spanning the exon junction of two genes that are $\geq$100,000 base pairs apart or on separate chromosomes, with each read mapping to at least 13 bases on both sides of the breakpoint. Preliminary fusion events were then
further examined manually to filter intron retention events and cross-alignments to yield high confidence gene fusion events.

4.3.5 RT-PCR

100 ng of RNA extracted from FFPE samples and cell lines was reverse transcribed using the qScript cDNA Synthesis kit (Quanta Biosciences, Gaithersberg, USA) and SuperScript III First Strand Synthesis Kit (Thermo Fisher Scientific) with oligoDT$_{20}$, Random Hexamer, and Gene Specific primers. The following primers were designed for PCR amplification of the fusion breakpoints using Primer3web version 4.0.0 (University of Tartu, Tartu, Estonia) and synthesized by Eurofins Genomics (Huntsville, USA): ESR1: CCDC92 primers 5’TCTGGATGCCTCCTATGAC3’ (forward) and 5’ACTTCCAAACCCμCTTTCAGC3’ (reverse), YAF2: PRICKLE1 primers 5’GACTTTCTGGGCTCTGGATG3’ (forward) and 5’TGCTGAGTAACCTGCTGTGC3’ (reverse). GAPDH primers 5’GGTGCTGAGTATGTCGTGGA3’ (forward) and 5’ACAGTCTTCTGGGTGGCAGT3’ (reverse) were used as an internal control. 25 μL PCR mixes were comprised of 1 μL 1x ThermoPol Reaction Buffer (Thermo Fisher Scientific), 2 mM additional MgCl$_2$, 200 nM dNTP and 200 nM of each of the primers. PCR was performed using a C1000 Thermal Cycler (BioRad, Hercules, USA) with the following conditions: 1 cycle at 95°C for 5 minutes, followed by 30 cycles of denaturation at 95 °C for 30 seconds, annealing at 61 °C for 30 seconds, and elongation at 68 °C for 30 seconds, and 1 cycle of elongation at 68 °C for 5 minutes. 15 μL of PCR products underwent electrophoresis using 2% agarose gels and visualized using SYBR Safe DNA Gel Stain (Thermo Fisher Scientific). PCR products were purified using the
QIAQuick PCR Purification Kit (Qiagen) and underwent Sanger sequencing at The Centre for Applied Genomics (Hospital for Sick Children, Toronto, CAN). Fusion sequences were then aligned against human genome and transcriptome using the BLAST alignment tool (NCBI, Bethesda, USA).

4.3.6 Unsupervised Hierarchical Clustering

Pairwise comparison of gene expression profiles of each LMS sample was performed using Pearson correlation coefficients (Liu et al., 2014). Unsupervised hierarchical clustering was performed with Euclidean distances and complete linkage and the results were plotted as a heat map matrix. Site, and differentiation, necrosis and mitotic rate scores were assigned to each sample to correlate clusters with clinical and histopathological features.

4.3.7 Analysis of Differential Gene Expression

The expression level of each gene was estimated by counting the number of reads mapped onto the gene regardless of transcription isoforms and normalizing to total mapped reads to obtain transcript union Read Per Million total reads (truRPMs) (Liu et al., 2014). Reads mapping onto both exons and introns were all counted for truRPM calculations using a custom R script. Differential gene expression was analyzed by calculating fold change of abdominal vs. uterine, extremity vs. uterine, abdominal + extremity vs. uterine, and abdominal vs. extremity samples. p-values were calculated for
the larger fold change in each comparison to identify genes with significant differential gene expression. The false discovery rate associated with multiple hypothesis testing was controlled for using the Benjamini-Hochberg (BH) procedure. Differential gene expression with a BH p-value < 0.05 was considered significant.

4.3.8 Gene Set Enrichment Analysis and Functional Annotation Clustering

Gene set enrichment analysis was performed using the enrichment analysis tool on the Gene Ontology Consortium homepage (http://geneontology.org/). Enrichment tests were performed using the top 100, 500, 1000, and 1471 differentially expressed genes to determine if gene sets from particular cellular components, molecular functions, or biological processes were enriched by site. Functional annotation clustering analysis was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID). The top 100, 500, 1000 and 1471 differentially expressed genes were analyzed by annotation groups including disease, functional categories, gene ontology, pathways, protein domains, and protein-protein interactions. p-values were calculated to examine the significance of gene-term enrichment and BH corrections were made to account for multiple hypothesis testing. Enrichment scores were calculated to rank the enrichment of the different annotation clusters. Each enrichment scores is the geometric mean of all the enrichment p-values of each annotation term in the group. Enrichment scores above 1.3 (equivalent to non-log scale 0.05) were considered to be statistically significant.
4.4 Results

4.4.1 YAF2-PRICKLE1 and ESR1-CCDC92 are Present in the Index Cell Line

*YAF2-PRICKLE1* and *ESR1-CCDC92* are two novel fusion transcripts detected in the primary LMS cell line STS210 by the Gladdy and Wrana Labs using RNA-Seq (Figure 4-1). In order to confirm *YAF2-PRICKLE1* and *ESR1-CCDC92* in the primary cell line, RT-PCR was performed using primers designed to amplify the breakpoints of these fusion transcripts in STS210. Both *YAF2-PRICKLE1* and *ESR1-CCDC92* were amplified at the expected sizes (220 bp and 175 bp, respectively) in STS210 and were not amplified in SK-LMS-1 or STS39, confirming the results of the RNA-Seq (Figure 4-2A and Figure 4-2B). The PCR products from both reactions using cDNA from STS210 underwent Sanger sequencing at the Toronto Centre for Applied Genomics (TCAG, Hospital for Sick Children) and the resulting sequences were aligned against human genome and transcriptome using the BLAST alignment tool. Sanger sequencing using the forward primer confirmed that the breakpoint of the *YAF2-PRICKLE1* fusion transcript occurred between exon 4 of *YAF2* and exon 1 of *PRICKLE1*. It also confirmed that the breakpoint of the *ESR1-CCDC92* fusion transcript occurred between the intronic region downstream of exon 2 of *ESR1* and the inverted exon 1 of *CCDC92*. Sanger sequencing with the reverse primers only detected one gene from each fusion (*PRICKLE1* and *ESR1*). The primer design necessitated that the reverse primers for both fusion PCRs be chosen close to the breakpoint, as the sequence downstream of the breakpoint in *CCDC92* was GC rich and the sequence downstream of the breakpoint in *YAF2* contained many repeats. To determine if *YAF2-PRICKLE1* and
ESR1-CCDC92 encode in-frame protein products, both sequences were translated in all reading frames and compared against a protein sequence database using the blastx program (NCBI). The YAF2-PRICKLE1 transcript codes for YAF2 protein while the ESR1-CCDC92 transcript does not code for either ESR1 or CCDC92 protein. Thus, the formation of these fusions may account for the upregulation of YAF2 and downregulation of ESR1, CCDC92 and PRICKLE1 (Figure 4-1).

Figure 4-2. YAF2-PRICKLE1 and ESR1-CCDC92 Transcripts are Confirmed in the Index Cell Line using RT-PCR. Both fusion genes were amplified in STS 210 (retroperitoneal site) but not in SK-LMS-1 (uterine site) or in STS39 (pelvic site) (A and B), confirming the transcriptome analysis. Sequencing confirmed the identity of the amplified products. GAPDH PCR was performed on all samples as an internal control (C).
4.4.2 YAF2-PRICKLE1 and ESR1-CCDC92 are not Recurrent in Additional LMS Patient Samples

YAF2-PRICKLE1 and ESR1-CCDC92 were confirmed in STS210, an LMS cell line derived from a metastatic buttock lesion, henceforth referred to as the index case, by the Gladdy Lab. Since these transcripts were identified in the cell line derived from the index case, it is anticipated that they are also present in the index case. Furthermore, since these transcripts were found in an LMS cell line, it may be possible that they are a recurrent feature of LMS. In order to determine if these fusion transcripts are present in the index case and recurrent in LMS, RT-PCR and RNA-Seq were performed in parallel on RNA extracted from an FFPE blocks from 15 LMS patient samples, including the index case. Analysis of the RNA-Seq data indicated that 7 spanning reads contained the YAF2-PRICKLE1 sequence and 18 spanning reads contained the ESR1-CCDC92 sequence (Figure 4-3). However, neither YAF2-PRICKLE1 nor ESR1-CCDC92 was detected as a fusion transcript in the index case using the parameters defined for gene fusion detection (Section 4.3.4). The minimum number of paired-end reads required to detect a high-confidence gene fusion event is two; neither fusion transcript was detected in the index case by paired-end reads, and thus neither was detected as a gene fusion event in the index case (Figure 4-3). In addition, neither YAF2-PRICKLE1 nor ESR1-CCDC92 was detected in the additional LMS patient samples using RNA-Seq (data not shown). Putative, recurrent fusion transcripts consisting of genes from the same chromosome were detected in several LMS patient samples. However, upon manual examination of the mapping of these fusions using ENSEMBL, it was determined that these genes were <1000 base pairs apart. The parameters for gene fusion detection required that these genes be ≥100,000 base pairs apart; these invalid
fusion transcripts may have met the parameters due to mapping errors by the TopHat 2.0.7 fusion junction mapper. RT-PCR confirmed these findings in the index case and the additional LMS patient samples.

**Figure 4-3. YAF2-PRICKLE1 and ESR1-CCDC92 Spanning Reads are Detected in the Index Case by RNA-Seq.** RNA-Seq data from STS210 and the index case was analyzed for spanning reads, paired reads, and paired reads with one end spanning each fusion. 7 YAF2-PRICKLE1 and 18 ESR1-CCDC92 spanning reads were detected in the Index Case. However, no paired reads of either fusion transcript were detected in the Index Case. Only 3 YAF2-PRICKLE paired reads with one end spanning the fusion were detected in the index case. A minimum of 1 spanning read, 2 paired reads, and 5 paired reads with one end spanning the fusion were required for fusion transcripts to be identified through the analytical pipeline.
4.4.3 Transcriptomes of Abdominal and Extremity LMS Form a Cluster

Several recent studies have identified a number of novel molecular subtypes of LMS. In order to determine if there are transcriptomic subtypes of LMS, unsupervised hierarchical clustering analysis of the transcriptome profiles generated from RNA-Sequencing of 15 LMS patient samples was performed. Pairwise correlation values were calculated for pairs of samples and Euclidean distance values were calculated to identify clusters of samples. Samples were classified by site of disease, grade, differentiation score, mitotic score, necrosis score, and metastasis to determine if LMS transcriptome clusters correlate with these disease characteristics (Table 4-1). These characteristics were selected because site and grade (which is determined by the differentiation, mitotic rate and necrosis of a tumor) are important prognostic factors in LMS. These samples were untreated and included 7 abdominal, 4 extremity and 4 uterine samples. Notably, uterine samples are not graded. A cluster of 11 samples was identified: these samples were the abdominal and extremity samples (Figure 4-4). The other histopathological and clinical features did not correlate with the clustering of the samples. The uterine samples were a distinct group with more heterogeneous transcriptomes. Two of the uterine samples (Uterine 2A and Uterine 2B) are from different blocks of the same primary tumor. Interestingly, these two samples did not have identical transcriptome profiles; this may be due to intratumor heterogeneity.
Table 4-1. Clinical and Histopathological Features of LMS Patient Samples.

<table>
<thead>
<tr>
<th>Site</th>
<th>Differentiation Score</th>
<th>Mitotic Score</th>
<th>Necrosis Score</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uterine 1</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>N/A</td>
</tr>
<tr>
<td>Uterine 2</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>Uterine 3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>Uterine 4</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>N/A</td>
</tr>
<tr>
<td>Extremity 1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Extremity 2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Extremity 3</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Abdomen 1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Abdomen 2</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Extremity 4</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Abdomen 3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Abdomen 4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Abdomen 5</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Abdomen 6</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Abdomen 7</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

(FNCLCC grading system reproduced with permission from Fletcher CDM, Bridge JA, Hogendoorn PCW, Mertens F. World Health Organization Classification of Tumours, Volume 5. Tumours of Soft Tissue and Bone. Fourth Edition. Lyon, IARC; 2013)
Figure 4-4. Unsupervised Hierarchical Clustering of 15 LMS Patient Samples Reveals Two Molecular Subtypes. Transcriptomes from abdominal/retroperitoneal (n=7) and extremity (n=4) samples cluster together whereas uterine site (n=4) has a more molecularly distinct profile.
4.4.4 Gene Set Enrichment and Functional Annotation Clustering of Differentially Expressed Genes of Differentially Expressed Genes

Given the tight clustering of the abdominal and extremity samples compared to uterine samples, it may be possible that particular gene sets, signaling pathways, or protein interaction networks may be differentially expressed by site. Differential gene expression was calculated between the cluster of abdominal and extremity samples and the uterine samples. 1471 genes were identified with significant differential expression (BH p-value < 0.05). The top 100, 500, 1000 and 1471 genes underwent gene set enrichment analysis and functional annotation clustering. 13 annotation clusters were identified with enrichment scores greater than 1.3, though these scores were not high (the largest enrichment score was 3.24). The annotation terms with significant differential expression did not seem to be biologically relevant to leiomyosarcomagenesis or progression.
4.4.5 PRKAA2 and CDKN2A are Differentially Expressed Between the Abdominal and Extremity Cluster and the Uterine Samples

Several recurrent molecular alterations in LMS have been reported in the literature. The PI3K/AKT/mTOR pathway is known to be activated in LMS (Hernando et al., 2007) and is believed to play an important role in leiomyosarcomagenesis. Other molecular alterations in LMS include alterations in tumor suppressor pathways, with loss of RB (70%) (Dei Tos et al., 1996), p16 (32%) (Kawaguchi et al., 2004), and p53 (43%) (Agaram et al., 2015). To determine if these alterations were more prevalent in the transcriptome cluster, the genes with significant differential expression (BH p-value < 0.05) were manually investigated to determine if PI3K/AKT/mTOR pathway and tumor suppressor genes are differentially expressed in LMS. PRKAA2, which codes for the catalytic subunit of AMPK, a negative regulator of mTORC1, was found to have 11.3 fold greater expression in abdominal and extremity samples than in uterine samples (Figure 4-5). CDKN2A, which codes for p16, was found to have 5.16 fold greater expression in abdominal and extremity samples than in uterine samples (Figure 4-5).
Figure 4-5. *PRKAA2* and *CDKN2A* are Differentially Expressed between the Uterine and Non-Uterine Samples. *PRKAA2* and *CDKN2A* have significant differential expression in the LMS patient samples (*p<0.01; ** p<0.005). *PRKAA2* codes for the catalytic subunit of AMPK, which negatively regulates mTORC1 by phosphorylating RPTOR and TSC2. *CDKN2A* codes for p16, a tumor suppressor that is dysregulated in LMS. Relative expression was calculated by dividing the number of smooth muscle gene reads by the number of *ACTB* reads. Errors bars indicate the mean ± standard deviation.
4.5 Discussion

Currently, it is unclear if there are recurrent gene fusions in LMS. In this study, I hypothesized that YAF2-PRICKLE1 and ESR1-CCDC92, two novel fusion transcripts detected in a primary LMS cell line by our group using RNA-Seq, may be recurrent in LMS. While YAF2-PRICKLE1 and ESR1-CCDC92 were confirmed in cell line STS210 using RT-PCR and were detected with spanning reads in the index case, these fusion transcripts ultimately were not recurrent in a discovery set of 15 LMS patient samples. Additionally, potential molecular subtypes have been identified in LMS. I hypothesized that LMS transcriptomes cluster by site of disease. Unsupervised hierarchical clustering of the transcriptomes from the discovery set of LMS samples suggested that abdominal and extremity LMS form a cluster, while uterine LMS is a distinct group.

YAF2-PRICKLE1 and ESR1-CCDC92 were amplified in STS210 using RT-PCR primers designed to amplify the breakpoints of the fusion transcripts. Each PCR reaction resulted in the amplification of a single PCR product and the sequencing of each product resulted in the sequence expected from the initial RNA-Seq data. This finding both confirmed the presence of YAF2-PRICKLE1 and ESR1-CCDC92 in STS210 and validated the use of RT-PCR to confirm the detection of novel fusion transcripts.

Determining if YAF2-PRICKLE1 and ESR1-CCDC92 are present in the index case is crucial to characterizing these fusion transcripts as a feature of the patient’s disease, as opposed to an artifact that may have arisen from creating a cell line from a patient’s resected metastatic tumor and performing serial passaging of the cell lines. RNA-Seq of
the index case resulted in 7 YAF2-PRICKLE1 and 18 ESR1-CCDC92 spanning reads and no paired-end reads of either transcript. In comparison, RNA-Seq of STS210 resulted in 35 spanning reads and 7 paired-end reads of YAF2-PRICKLE1 and 101 spanning reads and 5 paired-end reads of ESR1-CCDC92. Not only are there fewer spanning reads of each fusion transcript in the index case, but also there are no paired-end reads of the fusion transcripts. The criteria for fusion detection using RNA-Seq requires a minimum of 2 paired-end reads per fusion. As a result, neither fusion was detected in this analysis. The low number of spanning reads of the fusion transcripts in the index case suggests that these fusion transcripts are present at low levels in the index case and cells expressing these transcripts may have been clonally selected in the derivation of the STS210 cell line from the index case.

YAF2-PRICKLE1 and ESR1-CCDC92 were not recurrent in the discovery set of 15 additional LMS samples. Neither spanning reads nor paired-end reads with the sequences for either fusion transcript were detected in the additional samples. Although putative recurrent fusion transcripts were detected in these LMS patient samples, manual assessment of the mapping of the genes of these fusion events using ENSEMBL revealed that these genes were <1000 base pairs apart, suggesting that these events may have been misidentified as fusion genes due to mapping errors in the TopHat 2.0.7 fusion junction mapper. Moreover, this indicates that mapping errors may result in false-positive detection of fusion transcripts involving genes that are close when using this RNA-Seq based approach.
Neither YAF2-PRICKLE1 nor ESR1-CCDC92 was detected in the additional LMS patient samples using RT-PCR. PCR primers were designed for exon 2 of ESR1, exon 1 of CCDC92, exon 4 of YAF2 and exon 1 of PRICKLE1. It is possible that there may be fusion transcript isoforms arising from alternate exons of these genes which we did not assay for. Additional PCR primers would have to be designed to detect these fusion isoforms in the patient samples. One of the challenges of using RT-PCR to determine if YAF2-PRICKLE1 and ESR1-CCDC92 are present in the index case and recurrent in additional LMS samples was difficulty of amplifying a housekeeping gene using the RNA from the majority of the LMS patient samples. GAPDH was not amplified in 11/15 LMS patient samples, including the index case, using RT-PCR. cDNA synthesis was performed using different cDNA strategies, including qScript cDNA synthesis kit, 1:1 mix of oligotdT and random hexamers, oligodT alone and random hexamers alone. The RNA was extracted from FFPE samples, so it is possible that the RNA may have been fragmented in the process of formalin fixation. RNA yield and quality were assessed by Mount Sinai Pathology Services, re-assessed using Nanodrop and Bioanalyzer prior to preparation of cDNA libraries for RNAseq of these samples and re-analyzed using Nanodrop in the Gladdy Lab. The RNA yield and quality were adequate at each evaluation, as demonstrated by 260/280 ratios of ~2.0. Hypoxanthine-guanine phosphoribosyltransferase (HPRT) was used as an alternate housekeeping gene. Amplification of HPRT in the 4 GAPDH+ samples was weak compared to GAPDH and no HPRT was amplified in the other 11 samples. Since YAF2-PRICKLE1 and ESR1-CCDC92 fusion transcripts may be rare events, it is crucial to utilize RNA of adequate quality so that they may be detected by RT-PCR and RNA-Seq. Fresh frozen samples
may yield higher quality RNA upon extraction and thus should be considered for future studies.

A discovery set of 15 patient samples was used to investigate whether gene fusions are recurrent in LMS. Although recurrent gene fusions were not detected in this study, several strategies may be used in future investigations to examine this further. Firstly, using a larger sample set may increase the likelihood of finding recurrences of \( YAF2-PRICKLE1 \) and \( ESR1-CCDC92 \). Recently, COBRA-FISH karyotyping identified a translocation t(6;14) in two cases of LMS, though structural alterations were not identified using a tissue microarray of 37 additional LMS samples (de Graaff et al., 2015). Interestingly, gene fusions involving genes from chromosomes 6 and 14 were not identified in the 15 patient samples that underwent RNA-Seq in this study. This further supports the notion that gene fusions may be rare events (<5%) and suggests that a large number of samples must be assayed to determine whether recurrent fusion transcripts are present. Secondly, future studies investigating the recurrence of \( YAF2-PRICKLE1 \) and \( ESR1-CCDC92 \) fusion transcripts should use not only a larger number of LMS patient samples, but also a cohort of metastatic samples. It is important to note that these fusion transcripts were identified in a cell line derived from a metastatic buttock lesion. Aside from the index case, the additional LMS samples used in this study did not include metastatic samples. \( YAF2-PRICKLE1 \) and \( ESR1-CCDC92 \) may be a feature of metastatic LMS, and thus metastatic samples should be included in future studies investigating the recurrence of these fusion transcripts.
Unsupervised hierarchical clustering of the transcriptomes from the 15 LMS patient samples resulted in the grouping of abdominal and extremity samples. The uterine samples did not cluster with the abdominal and extremity samples and their transcriptomes were heterogeneous amongst each other. This finding may support the notion of uterine LMS being biologically different from LMS originating at other sites. It may also suggest that abdominal and extremity samples form a distinct molecular subtype of LMS, though they have different rates of recurrence and long-term DSS. It could be possible that abdominal and extremity have similar transcriptome profiles but respond to treatment differently due to differential expression of a few genes or differences in translational regulation, post-translational modification, and inter and intracellular signaling.

The discovery set of LMS samples used to test this hypothesis consisted of only 15 samples and did not include an equivalent number of samples from each disease site. Additional samples should undergo RNAseq to determine if the abdominal and extremity samples continue to cluster as one group when there are equivalent numbers of samples from each site and many more samples per site. Unsupervised hierarchical clustering of transcriptomes from a larger number of samples may confirm the heterogeneity of uterine LMS transcriptomes and support the finding that they do not cluster with abdominal and extremity samples. Furthermore, increasing the sample size would facilitate the investigation of clustering by grade and by the pathological factors that are used to determine grade, including degree of differentiation, mitotic rate, and degree of necrosis.
Gene set enrichment analysis and functional annotation clustering of genes with significant differential expression by cluster identified statistically significant enriched clusters of genes. The 13 annotation clusters that were identified had low enrichment scores (1.33-3.24). The annotation terms with significant differential expression were primarily mitochondrial and organelle membrane proteins. Annotation terms with biological relevance to leiomyosarcomagenesis were not identified in this list of differentially expressed genes. To further investigate significant differential gene expression in this discovery set of transcriptomes, I analyzed the expression of PI3K/AKT/mTOR pathway genes. The PIK/AKT/mTOR pathway has been shown to play a critical role in LMS. 70% of LMS has been reported to feature deletions in 10q, the chromosomal arm where PTEN is located (Hernando et al., 2007). Conditional PTEN knockouts result in LMS formation (Hernando et al., 2007). pS6K (Thr389), and p4EBP1 (Thr37/46), indicative of mTOR activity, have also been correlated with LMS formation (Hernando et al., 2007). The Gladdy Lab has shown that mTOR pathway is activated in two LMS cell lines and that a dual PI3K/mTOR inhibitor significantly inhibited LMS growth in vivo (Babichev et al., 2016). PRKAA2 was found to be differentially expressed in this discovery set of LMS samples, with 11-fold greater expression in the abdominal and extremity cluster compared to uterine samples. PRKAA2 codes for the catalytic subunit of AMPK, which negatively regulates mTORC1 by phosphorylating RPTOR and TSC2. Lower expression of PRKAA2 in uterine LMS may suggest that there is less mTORC1 inhibition and, consequently, increased mTOR activity in uterine LMS.
Lastly, the discovery set of LMS transcriptomes was analyzed for differential expression of known dysregulated genes in LMS. *CDKN2A* was found to be differentially expressed, with 5-fold greater expression in the abdominal and extremity cluster than in the uterine samples. p16, which is coded by *CDKN2A*, is a tumor suppressor which has been reported to have decreased expression in 32% of LMS. The differential expression of *CDKN2A* may suggest that the decrease in p16 may be more prominent in uterine LMS than in abdominal and extremity LMS, and may contribute to uterine disease being biologically different from abdominal and extremity LMS.

The purpose of this chapter is to determine if 1) *YAF2-PRICKLE1* and *ESR1-CDC92* are recurrent in LMS and 2) LMS transcriptomes cluster by site of disease. Both fusion transcripts were confirmed in the primary LMS cell line they were discovered in but were not detected in the additional patient samples used in this study and thus were not recurrent. Considering the low number of *YAF2-PRICKLE1* and *ESR1-CCDC92* spanning reads in the index case, the quality of the RNA extracted from FFPE samples, the limited number of samples studied, and the lack of a metastatic cohort of samples, it is possible that these fusion transcripts may be found to be recurrent upon further investigation. Abdominal and extremity LMS transcriptomes clustered while uterine LMS transcriptomes formed a separate, heterogeneous group. This may suggest that uterine LMS is biologically different from LMS originating at other sites. The validation of these clusters with a larger sample set composed of equal numbers of samples from different sites may suggest that LMS has site-associated molecular subtypes. Ultimately, these findings may aid in better defining the molecular events in LMS.
Chapter 5  Investigating the Role of FOXO4 in LMS Smooth Muscle Differentiation
5.1 Abstract

Leiomyosarcoma (LMS) is a tumor characterized by distinct smooth muscle features (reviewed in Fletcher et al., 2013); as such, it has been considered to arise from smooth muscle cells undergoing neoplastic transformation. Evidence from several independent studies suggests that LMS may arise from more stem-like cells, such as mesenchymal stem cells or undifferentiated cells committed to other lineages (Danielson et al., 2010; Pérot et al., 2009). Furthermore, miRNA signatures associated with smooth muscle differentiation have been identified that distinguish uterine LMS from uterine leiomyoma (Danielson et al., 2010). These findings suggest that smooth muscle differentiation may be dysregulated and therefore associated with leiomyosarcomagenesis (Danielson et al., 2010). Gene amplification of myocardin (MYOCD), the master regulator of smooth muscle gene expression, has been reported in several LMS patient studies (Agaram et al., 2015; Pérot et al., 2009) and dysregulated MYOCD expression has been found in a number of additional studies (Kimura et al., 2010; Italiano et al., 2013). Additionally, knockdown of MYOCD in an LMS cell line led to a loss of smooth muscle differentiation (Pérot et al., 2009). PI3K/AKT pathway activation via IGF signaling regulates smooth muscle differentiation by phosphorylating and translocating nuclear FOXO4 to the cytoplasm, thereby negating the inhibitory effect of FOXO4 on MYOCD transcriptional activity (Liu et al., 2005). Knockdown of FOXO4 has also been reported to result in increased expression of smooth muscle genes in a rat aortic smooth muscle cell line (Liu et al., 2005). Here, we investigated if FOXO4 regulates smooth muscle differentiation in LMS by characterizing PI3K/AKT pathway activation, insulin signaling, MYOCD and smooth muscle marker expression, and FOXO4 expression in LMS cell
lines and patient samples. Specifically, we characterized smooth muscle differentiation in 3 LMS cell lines and confirmed differential expression of smooth muscle markers between abdominal and extremity LMS. We report that AKT is phosphorylated at Thr308 in 3 LMS cell lines, indicating PI3K/AKT pathway activation in LMS. Additionally, we found that MYOCDC expression is correlated with smooth muscle differentiation in the LMS cell lines. Finally, we report that FOXO4 expression negatively correlates with smooth muscle differentiation in the LMS cell lines. These findings may suggest that dysregulated FOXO4 expression regulates smooth muscle differentiation in LMS and support further investigation of the role of FOXO4 in LMS smooth muscle differentiation.
5.2 Introduction

LMS is characterized as a smooth muscle tumor. However, there is a spectrum of LMS histology as well-differentiated tumors primarily consist of spindle cells grouped in longitudinally-oriented fascicles, with centrally located, elongated nuclei and eosinophilic or fibrillary cytoplasm (reviewed in Guillou and Aurias, 2009; reviewed in Batan et al., 2013), while poorly differentiated tumors have pleomorphic cells with abnormal mitoses (reviewed in Guillou and Aurias, 2009; reviewed in Fletcher et al., 2013). The diagnosis of LMS requires immunohistochemical staining for smooth muscle actin (SMA), desmin or h-caldesmon. SMA, desmin and h-caldesmon are markers of smooth muscle differentiation and are expressed in more than 70% of LMS (reviewed in Fletcher et al., 2013). Smooth muscle actin is detected in the majority of LMS samples (Swanson et al., 1991; Oda et al., 2001) while desmin is expressed in >50% of LMS (Oda et al., 2001; Hisaoka et al., 2001). H-caldesmon is expressed in 40% of LMS and this expression varies with tumor site and degree of differentiation (Hisaoka et al., 2001). De-differentiated areas typically lack SMA and desmin expression (Chen et al., 2001).

LMS is thought to arise from smooth muscle cells, since LMS is composed of cells featuring smooth muscle differentiation (de Graaf et al., 2015; Guo et al., 2015; reviewed in Batan et al., 2013; Hernando et al., 2007). The cell-of-origin of tumors has traditionally been considered to be the corresponding mature somatic cells. Increasingly, progenitor or stem cells have been considered as the target of neoplastic transformation (reviewed in Xiao et al., 2013; reviewed in Pardal et al., 2003). LMS-like tumors have been reported to arise in p53-deficient mesenchymal stem cells (MSC).
injected into immunodeficient mice (Rubio et al., 2010). It may be possible that tumorigenesis in these progenitors results from the dysregulation of developmental pathways. A group of 30 miRNAs were found to be required for smooth muscle cell (SMC) differentiation of human MSCs and 20/76 miRNAs identified in uterine LMS were among those identified with an earlier state of smooth muscle differentiation in the former group (Danielson et al., 2010). In contrast, the miRNA profiles of uterine leiomyomas featured miRNAs associated with mature SMCs and myometrium, suggesting that uterine LMS and uterine leiomyomas may have different origins and that uterine LMS may originate from mesenchymal stem cells (Danielson et al., 2010).

MYOCD has been reported to be focally amplified and overexpressed in 53% of well-differentiated (WD) LMS (Pérot et al., 2009). A separate study found MYOCD amplification in 13% of LMS but did not report MYOCD amplification in high-grade LMS from different disease sites (Agaram et al., 2015). Another group investigating LMS independent of grade found that MYOCD was the most differentially expressed gene in retroperitoneal LMS compared to extremity LMS (Italiano et al., 2013). MYOCD protein expression has been reported to be highly downregulated in uterine LMS cell lines, which generally have a de-differentiated smooth muscle phenotype (Kimura et al., 2010). Furthermore, exogenous MYOCD expression resulted in expression of smooth muscle genes and upregulation of p21, a cell cycle regulator (Kimura et al., 2010). MYOCD is a known transcriptional cofactor of serum response factor (SRF). The MYOCD-SRF complex has been reported to regulate smooth muscle differentiation by transcriptionally activating smooth muscle differentiation genes including ACTA2, DES, TAGLN, MYLK, MYH11, CALD1, and CNN1 (Wang et al., 2003). Interestingly,
overexpression of *MYOCD* in three undifferentiated pleomorphic sarcoma (UPS) cell lines and a liposarcoma (LPS) cell line resulted in smooth muscle differentiation and increased cell migration, suggesting that LMS can arise from mesenchymal lineages driven to smooth muscle differentiation and does not have to arise exclusively from smooth muscle cells.

Smooth muscle cells modulate between a differentiated, contractile phenotype and a de-differentiated, synthetic phenotype (reviewed in Owens *et al.*, 2004). The phenotypic modulation of smooth muscle cells is characterized by changes in smooth muscle gene expression. In rat smooth muscle cells (SMCs), the MYOCD-SRF complex becomes transcriptionally active due to the translocation of FOXO4 from the nucleus to the cytoplasm, resulting in smooth muscle marker expression (Liu *et al*., 2005). FOXO proteins are involved in cell proliferation, apoptosis and cell cycle regulation (reviewed in Calnan *et al*., 2008). FOXO4 is known to repress rat SMC differentiation by binding to the MYOCD-SRF complex and inhibiting the transcriptional activity of MYOCD (Liu *et al*., 2005). Insulin signaling results in the phosphorylation of AKT at Thr308; in turn, FOXO4 is phosphorylated at 3 sites (Thr32, Ser197 and Ser262) by activated AKT. This results in translocation of FOXO4 to the cytoplasm and the expression of smooth muscle markers (Liu *et al*, 2005). This mechanism regulates smooth muscle differentiation in rat smooth muscle and may play a role in the pathogenesis of atherosclerotic lesions and leiomyosarcoma (Liu *et al*., 2005). Additionally, knockdown of FOXO4 in a rat aortic smooth muscle cell line resulted in upregulation of smooth muscle genes (Liu *et al*., 2005), indicating that dysregulation of FOXO4 expression may regulate smooth muscle differentiation in smooth muscle cells.
The PI3K/AKT pathway plays an important role in regulating cell growth, metabolism, and proliferation and is activated in LMS (Hernando et al., 2007). Specifically, Insulin Receptor Substrate 2 (IRS2) can act as an upstream modulator of the pathway in LMS (Gibault et al., 2012). Additionally, PTEN, a phosphatase that negatively regulates the pathway by dephosphorylating PIP3 to PIP2, is deleted as part of chromosomal arm 10q in 70% of LMS [6]. Furthermore, 80% of conditional PTEN knockout mice develop LMS with a latency of 2 months [6]. Thus, the PI3K/AKT pathway, MYOCD and FOXO4 may be involved in regulating smooth muscle differentiation in human LMS as postulated in rat SMCs. I hypothesize that (1) dysregulated FOXO4 expression and (2) phosphorylation of FOXO4 may regulate SMC differentiation in LMS.

Elucidating the mechanism driving smooth muscle differentiation in LMS may improve our understanding of the pathogenesis of leiomyosarcoma and reveal potential therapeutic targets.
5.3 Materials and Methods

5.3.1 Tumor Samples

Formalin-fixed, paraffin-embedded (FFPE) tumors were obtained from the Biospecimen Repository and Processing Lab (Mount Sinai Hospital, Toronto, CA). Histologic slides from all FFPE samples were evaluated and scored by a dedicated sarcoma pathologist at Mount Sinai Hospital (Dr. Brendan Dickson) and 7 abdominal, 4 extremity and 4 uterine tumors were classified as LMS. Samples were scored for degree of differentiation, amount of necrosis, and mitotic rate in accordance with the FNCLCC grading system (Coindre, 2006). Total RNA was isolated from the FFPE samples using an RNeasy Mini Kit (Qiagen, Hilden, DEU). RNA yield and quality were assessed using Nanodrop (Thermo Fisher Scientific, Waltham, USA) and BioAnalyzer (Agilent Technologies, Santa Clara, USA).

5.3.2 Cell Line Establishment and Culture Conditions

Research Ethics Board consent was obtained to derive cell lines from resected LMS tumors STS210 (metastatic buttock) and STS39 (paravaginal). STS210 and STS39 cell lines were derived by manual dissociation of patient samples which were incubated overnight at 37°C in supplemented DMEM/F-12 10% FBS (Life Technologies, Carlsbad, USA) with 2 μg/ml collagenase (Sigma-Aldrich, St. Louis, USA). Cell suspensions were centrifuged at 1000 rpm for 5 min and plated on T75 flasks. SKLMS1, SW480, and C2C12 cells were obtained from ATCC and cultured in DMEM (Life Technologies), 10%
FBS media. Myoblastic differentiation was induced in C2C12 cells by incubating in DMEM, 5% horse serum for 5 days.

5.3.3 RNA-Sequencing (RNA-Seq)

200 ng of total RNA was used per sample for RNA-Seq. rRNAs were removed from the total RNA of FFPE samples and cell lines using Ribo-Zero rRNA Removal Kit (Illumina, San Diego, USA). rRNA removal was indicated by a GAPDH:S18 ratio >1. cDNA libraries were prepared using TruSeq RNA Library Prep Kit v2 (Illumina). Libraries displaying successful removal of rRNA were then loaded as 2 indexed samples per lane on a HiSeq 2000 (Illumina) for RNA-seq (Liu et al., 2014). Raw sequence reads were obtained in FASTQ format using standard CASAVA software (Illumina). Standard algorithms for RNA-seq data production and quality control were used with an average read depth of approximately 100 million reads per sample.

5.3.4 Analysis of Gene Expression

The expression level of each gene was estimated by counting the number of reads mapped onto the gene regardless of transcription isoforms and normalizing to total mapped reads to obtain transcript union Read Per Million total reads (truRPMs) (Liu et al., 2014). Reads mapping onto both exons and introns were all counted for truRPM calculations using a custom R script. Relative expression of each gene was calculated by dividing the number of reads of the gene by the number of reads of β-actin (ACTB). Differential gene expression was analyzed by calculating fold change of abdominal vs
uterine, extremity vs uterine, abdominal + extremity vs uterine, and abdominal vs extremity samples. p-values were calculated for the larger fold change in each comparison to identify genes with significant differential gene expression. The false discovery rate associated with multiple hypothesis testing was controlled for using the Benjamini-Hochberg (BH) procedure. Differential gene expression with a BH p-value < 0.05 was considered significant.

5.3.5 Protein Extraction from Cell Lines

SW480, STS210, STS39, SKLMS1 and C2C12 cells were grown in p100 dishes until almost confluent. After a cold PBS (Sigma-Aldrich) wash, cells were lysed for 30 minutes on ice with RIPA buffer (50mM pH 8 Tris-HCl, 150mM NaCl, 1% Triton X-100, 0.1% SDS, 0.5% Deoxycholate, 1mM EDTA) supplemented with phosphatase (Sigma-Aldrich) and protease inhibitors (Roche, Basel, CHE). Protein concentrations were measured using the DC Protein Concentration Assay (BioRad, Hercules, USA).

5.3.6 siRNA Transfection

ON-TARGETplus SMARTpool siRNA targeting FOXO4 and siGENOME Control Non-Targeting siRNA pool #3 were purchased from Dharmaco (Chicago, USA) and used to perform FOXO4 expression knockdown experiments in SW480 cells. SW480 cells were seeded at a density of $2 \times 10^5$ per well onto 6-well plates 24 hours before transfection. Cells were transfected for 24 hours at 37°C with siRNA duplex at a concentration of 40
nM using Lipofectamine 2000 (Invitrogen, Carlsbad, USA). Cells were harvested 72 hours after transfection for immunoblotting using a FOXO4 antibody.

5.3.7 Immunoblotting

Protein extracted from cell lines and frozen samples was used for immunoblotting. Electrophoresis was performed using 8-10% polyacrylamide gels and transferred to PVDF by wet transfer. Immunoblots were performed with the following antibodies: p-AKT(Thr308) (1:500; #9275), total AKT (1:500; #9272), total FOXO4 (1:1000; #9472), all from Cell Signaling Technology (Denver, USA), p-FOXO4(S197) (1:500; ab192183) from Abcam (Cambridge USA), MYOCD (1:200; A-13) and β-actin (1:5000; AC-15) from Santa Cruz Biotechnology (Dallas, USA). Secondary antibodies were horseradish peroxidase conjugated anti-rabbit IgG (1:5000; NA934), anti-mouse (1:5000, NA931) from GE Healthcare Life Sciences (Little Chalfont, UK) and anti-goat IgG (1:10000; sc-2020) from Santa Cruz Biotechnology.

5.3.8 Immunocytochemistry (ICC)

Protein expression of smooth muscle actin (SMA), desmin, and h-caldesmon in the cell lines was analyzed using ICC. Cells grown on 8-well chamber slides were fixed with 4% paraformaldehyde in PBS for 10 min at room temperature and then permeabilized in Permeabilization Solution for 20 min at room temperature. Immunocytochemistry was performed with the following primary antibodies: SMA [clone 1A4] (1:200; M0851) and h-caldesmon (1:200; M3557) from Dako (Burlington, CAN) and desmin [clone DE-U-10]
(1:200; ab6322) from Abcam. Mouse IgG (1:400; sc-2025) from Santa Cruz Biotechnology was used to control background. Nuclei were counterstained with DAPI (1:2000; D9542) from Sigma-Aldrich.
5.4 Results

5.4.1 Characterizing Smooth Muscle Differentiation in LMS

Smooth muscle cells (SMC) do not terminally differentiate; rather, they transition between a quiescent, differentiated, contractile phenotype and a proliferative, de-differentiated, synthetic phenotype (reviewed in Milewicz et al., 2010; reviewed in Owens et al., 2004). This phenotypic modulation is characterized by changes in smooth muscle marker expression (Figure 5-1). Fully differentiated SMCs express the highest levels of markers such as smooth muscle actin ($\text{ACTA2}$), desmin ($\text{DES}$), transgelin ($\text{TAGLN}$), myosin light-chain kinase ($\text{MYLK}$), myosin heavy chain 11 ($\text{MYH11}$), h-caldesmon ($\text{CALD1}$), and calponin ($\text{CNN1}$), while proliferating smooth muscle cells express reduced levels of these genes (reviewed in Owens et al., 2004).

![Figure 5-1. Smooth Muscle Cell Phenotypes.](image)

**Figure 5-1. Smooth Muscle Cell Phenotypes.** Smooth muscle cells modulate between a proliferative, de-differentiated, synthetic phenotype and a quiescent, differentiated, contractile phenotype. Expression of SMC markers ($\text{ACTA2}$, $\text{DES}$, $\text{TAGLN}$, $\text{MYLK}$, $\text{MYH11}$, $\text{CALD1}$, and $\text{CNN1}$) is higher in contractile SMC and lower in synthetic SMC. (Adapted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: Genetics in Medicine, 12, Milewicz et al., Genetic variants promoting smooth muscle cell proliferation can result in diffuse and diverse vascular diseases: Evidence for a hyperplastic vasculomyopathy, pages 196-203, copyright (2010): License # 3937650653529)
To characterize the degree of smooth muscle differentiation in LMS, transcriptome profiles of three LMS cell lines and a discovery set of 15 LMS patient samples generated previously by RNA-Seq (Chapter 4) were analyzed for the expression of ACTA2, DES, TAGLN, MYLK, MYH11, CALD1, and CNN1 (Figure 5-2 and Figure 5-3). Publically available transcriptome data from human coronary artery smooth muscle cells (HCASMC) was also analyzed to determine smooth muscle marker expression differences between LMS cell lines/samples and normal smooth muscle. The expression of smooth muscle markers was analyzed relative to that of beta-actin (ACTB), a housekeeping gene, so that expression differences between samples would be due to biological differences and not due to sequencing differences. Among the cell lines, STS39 has the highest expression of ACTA2 (Figure 5-2A), MYLK (Figure 5-2D), TAGLN (Figure 5-2C), and CNN1 (Figure 5-2G), while STS210 has the highest expression of DES (Figure 5-2B) and CALD1 (Figure 5-2F) and SK-LMS-1 has the highest expression of MYH11 (Figure 5-2E). Since STS39 also has the second highest expression of DES (Figure 5-2B) and CALD1 (Figure 5-2F), these results suggest that STS39 is the most differentiated of the three LMS cell lines, while SK-LMS-1 is the least differentiated. Furthermore, protein expression of smooth muscle actin (SMA), desmin and h-caldesmon was characterized in the cell lines using immunocytochemical (ICC) staining. STS39 expresses SMA, desmin and h-caldesmon, while STS210 and SK-LMS-1 only express desmin (Figure 5-2H). The protein expression of these markers also supports the finding that STS39 is the most differentiated of the three LMS cell lines.
Figure 5-2. STS39 is the Most Differentiated LMS Cell Line. ACTA2 is expressed in all three LMS cell lines, with highest expression in STS39 (A). DES is expressed in STS210 and STS39, with higher expression in STS210, and not expressed in SK-LMS-1 or HCASMC (B). TAGLN is expressed in all three LMS cell lines, with the highest expression in STS39 among the LMS cell lines (C). MYLK is expressed in all three LMS cell lines, with the highest expression in STS39 (D). MYH11 is expressed in all three LMS cell lines, with the highest expression in SK-LMS-1 (E). CALD1 is expressed in all three LMS cell lines, with the highest expression in STS210 (F). CNN1 is expressed in all three LMS cell lines, with the highest expression in STS39 (G). Relative expression was calculated by dividing the number of smooth muscle gene reads by the number of ACTB reads. Error bars indicate the mean ± standard deviation. ICC of LMS cell lines demonstrates that STS39 expresses SMA, desmin and h-caldesmon, while STS210 and SK-LMS-1 only express desmin (H). DAPI staining indicates cell nuclei. Scale bars for all images, 50 μm.
One study has reported that muscle differentiation genes are differentially expressed between retroperitoneal LMS and extremity LMS, with overexpression observed in the former (Italiano et al., 2013). To characterize the differential expression of smooth muscle genes by site, transcriptome data from 15 LMS patient samples was analyzed for the expression of ACTA2, DES, TAGLN, MYLK, MYH11, CALD1, and CNN1. Of these markers, MYLK, MYH11, and CALD1 have significant differential expression between abdominal and extremity LMS, with statistically significant overexpression in abdominal samples (Figure 5-3). Smooth muscle genes were not differentially expressed between abdominal and uterine samples, extremity and uterine samples, or uterine and non-uterine samples (data not shown). Collectively, these results indicate that STS39 is the most differentiated of the LMS cell lines and confirm the overexpression of smooth muscle differentiation genes in abdominal LMS.

Figure 5-3. MYLK, MYH11, and CALD1 are Differentially Expressed between Abdominal and Extremity LMS. MYLK, MYH11, and CALD1 expression is significantly higher in abdominal LMS than in extremity LMS, while ACTA2, DES, TAGLN, and CNN1 expression is not differentially expressed between abdominal and extremity LMS. Relative expression was calculated by dividing the number of smooth muscle gene
reads by the number of *ACTB* reads. Error bars indicate the mean ± standard deviation. *p<0.05; **p<0.005.

### 5.4.2 The PI3K/AKT Pathway is Activated in LMS

Smooth muscle differentiation in LMS may be regulated by PI3K/AKT pathway activation resulting from phosphorylation of AKT at Thr308 (Liu *et al.*, 2005). To characterize the activation of the PI3K/AKT pathway in LMS, protein lysates from three LMS cell lines (STS210, STS39, and SK-LMS-1) were assayed by western blotting with total and p-AKT(Thr308) antibodies. Pathway activation is indicated by AKT phosphorylation. p-AKT(Thr308) is present in all 3 LMS cell lines and thus the PI3K/AKT pathway is activated in these cell lines (Figure 5-4). Pathway activation does not correlate with smooth muscle differentiation in the LMS cell lines, since STS39, the most differentiated of the LMS cell lines, was characterized by the lowest level of p-AKT(Thr308) in the LMS cell lines (Figure 5-4).

**Figure 5-4. AKT is Phosphorylated at Thr308 in LMS.** Western blot shows phosphorylation of AKT at Thr308. SW480 was used as a positive control for p-AKT(Thr308) and AKT expression. β-actin was used as a loading control.
5.4.3. MYOCD Expression Correlates with Smooth Muscle Differentiation in LMS Cell Lines

MYOCD, a SRF transcriptional cofactor, is the master regulator of smooth muscle gene expression (Wang et al., 2003). The MYOCD gene has been reported to be overexpressed in well-differentiated retroperitoneal LMS (Pérot et al., 2009) and downregulated in dedifferentiated uterine LMS cell lines (Kimura et al., 2010). Moreover, one study found that MYOCD is the most differentially expressed gene in retroperitoneal LMS compared to extremity LMS (Italiano et al., 2013). Here, we sought to determine if 1) MYOCD expression correlates with smooth muscle differentiation and 2) MYOCD is differentially expressed by site.

To characterize SRF and MYOCD expression in LMS, transcriptome profiles from the 3 LMS cell lines and 15 patient samples were analyzed and MYOCD western blots were performed on protein lysates from cell lines. MYOCD and SRF are expressed in the three LMS cell lines (Figure 5-5A and Figure 5-5B). The mean expression level of MYOCD and SRF across the 15 LMS samples is higher than the corresponding levels in HCASMC (Figure 5-5C and Figure 5-5D). The significance of the differential expression of these genes between LMS and normal smooth muscle could not be assessed, as only one control smooth muscle cell line was used in this study.

STS39 has the highest expression of MYOCD among the cell lines, followed by STS210 and SK-LMS-1. Furthermore, MYOCD is overexpressed in STS39 compared to expression in HCASMC (Figure 5-5A). STS39 is also the most differentiated of the LMS cell lines, while SK-LMS-1 is the least differentiated (Section 5.4.2). Thus, MYOCD expression correlates with smooth muscle differentiation in the LMS cell lines tested.
Interestingly, STS39, the cell line which had the highest mRNA expression of MYOCDD, has the lowest expression of level of MYOCDD protein, as detected by western blotting (Figure 5-5G). Analysis of the transcriptome data from the LMS patient samples suggests that mean expression level of MYOCDD is higher in well-differentiated LMS samples with a differentiation score of 1 than in moderately differentiated and de-differentiated samples with differentiation scores of 2 and 3, respectively (Figure 5-5E), though this differential expression was not statistically significant. Notably, MYOCDD expression did not correlate with ACTA2, DES, TAGLN, MYLK, MYH11, CALD1, and CNN1 expression (data not shown) in the patient samples. To determine if MYOCDD is differentially expressed by site, we analyzed expression differences between abdominal and extremity, abdominal and uterine, abdominal and non-abdominal samples, and uterine and non-uterine samples. Though the mean expression of MYOCDD is higher in abdominal samples than in extremity or uterine samples, statistically significant differential expression by site was not detected (Figure 5-5F). These findings indicate that MYOCDD expression correlates with smooth muscle differentiation in our LMS cell lines and may suggest that MYOCDD expression is greater in well-differentiated and abdominal LMS samples. Investigating additional patient samples would increase the statistical power and validate these findings.
Figure 5-5. Myocardin and SRF Expression in LMS Cell Lines and Patient Samples. *MYOCD* is expressed in all three LMS cell lines, with highest expression in
STS39 (A). SRF is expressed in all three LMS cell lines, with the highest expression level in SK-LMS-1 (B). Mean MYOCD and SRF expression levels were higher in the LMS samples (n=15) than in HCASMC (C and D). The statistical significance of the differential expression between LMS samples and normal smooth muscle could not be determined, as data from only a single smooth muscle cell line (HCASMC) was used to examine this difference. Mean MYOCD expression was higher in samples with differentiation score of 1 than in samples with differentiation scores of 2 and 3, though this difference was not statistically significant (E). Mean MYOCD expression was higher in abdominal samples than in extremity or uterine samples, though this difference was not statistically significant (F). Mean MYOCD expression was higher in samples with differentiation score of 1 than in samples with differentiation scores of 2 and 3, though this difference was not statistically significant (E). Relative expression was calculated by dividing the number of MYOCD and SRF reads by the number of ACTB reads. Error bars indicate the mean ± standard deviation. MYOCD is expressed in all three LMS cell lines (G). SW480 and terminally-differentiated C2C12 mouse myoblasts were used as positive and negative controls, respectively, for MYOCD expression. β-actin was used as a loading control.

5.4.3 FOXO4 Expression Correlates with Smooth Muscle Differentiation in LMS Cell Lines

The phosphorylation of FOXO4 at Thr32, Ser197 and Ser262 has been reported to regulate smooth muscle gene expression and differentiation in rat aortic smooth muscle by resulting in the cytoplasmic localization of FOXO4 and transcriptional activation of MYOCD (Liu et al., 2005). Additionally, transfection of a rat aortic smooth muscle cell line with FOXO4 siRNA has been found to increase expression of smooth muscle genes (Liu et al., 2005). FOXO4 is an ubiquitously expressed transcription factor that contributes to the regulation of cellular processes such as cell cycle progression, proliferation, and apoptosis. To determine if FOXO4 expression negatively correlates with smooth muscle differentiation in LMS, transcriptome profiles from 3 LMS cell lines and 15 patient samples were analyzed and FOXO4 western blots were performed on protein lysates from the cell lines. FOXO4 is expressed in the three LMS cell lines
STS39, the most differentiated of the LMS cell lines, has the lowest expression of FOXO4, while SK-LMS-1, the least differentiated of the LMS cell lines, has the highest expression of FOXO4 (Figure 5-6A). Thus, FOXO4 expression negatively correlates with smooth muscle differentiation in the LMS cell lines. Notably, FOXO4 protein expression did not correlate with smooth muscle differentiation in the LMS cell lines (Figure 5-6D). Furthermore, FOXO4 expression did not correlate with smooth muscle differentiation in the LMS patient samples (Figure 5-6B). Smooth muscle differentiation genes have been reported to be overexpressed in retroperitoneal compared to in extremity disease (Italiano et al, 2013) and we found that MYLK, MYH11, and CALD1 are differentially expressed between abdominal and extremity LMS (Figure 5-3). To determine if FOXO4 is differentially expressed by site, we analyzed expression differences between abdominal and extremity, abdominal and uterine, abdominal and non-abdominal samples, and uterine and non-uterine samples. Though the mean expression of FOXO4 is higher in abdominal samples than in extremity or uterine samples, statistically significant differential expression by site was not detected (Figure 5-6C). These findings indicate that FOXO4 expression negatively correlates with smooth muscle differentiation in the LMS cell lines. p-FOXO4(S197) western blots are underway to determine if the phosphorylation of FOXO4 at Thr32, Ser197 and Ser262 correlates with smooth muscle gene expression and differentiation in LMS.
Figure 5-6. FOXO4 Expression in LMS. FOXO4 is expressed in all three LMS cell lines (A). The highest FOXO4 expression level is detected in SK-LMS-1 and lowest in STS39 (A). Mean FOXO4 expression was higher in samples with differentiation score of 2 than in samples with differentiation scores of 1 and 3, though this difference was not statistically significant (B). Mean FOXO4 expression was higher in abdominal samples than in extremity or uterine samples, though this difference was not statistically significant (C). Relative expression was calculated by dividing the number of FOXO4 reads by the number of ACTB reads. Error bars indicate the mean ± standard deviation. Western blots indicate FOXO4 expression in LMS (D). SW480 was used as a positive control for FOXO4 expression. FOXO4 siRNA knockdown was performed in SW480 to generate a negative control for FOXO4 expression. β-actin was used as a loading control.
5.5 Discussion

Several findings suggest that smooth muscle differentiation may be associated with leiomyosarcomagenesis (Danielson et al., 2010; Pérot et al., 2009). Currently, the mechanism regulating smooth muscle differentiation in LMS has not been elucidated. Smooth muscle differentiation in rat aortic smooth muscle cells is thought to be regulated by the phosphorylation-induced translocation of FOXO4 to the cytoplasm (Liu et al., 2005). Dissociation of FOXO4 from the MYOCD-SRF complex results in transcriptional activation and the expression of smooth muscle markers (Liu et al., 2005). siRNA-mediated knockdown of FOXO4 in a rat aortic smooth muscle cell lines has also been reported to downregulate smooth muscle marker expression (Liu et al., 2005). Here, I hypothesized that (1) dysregulated FOXO4 expression and (2) phosphorylation of FOXO4 may regulate SMC differentiation in LMS. p-AKT(Thr308) was detected in 3 LMS cell lines, indicating that the PI3K/AKT pathway is activated in LMS. MYOCD expression correlated with smooth muscle differentiation in the LMS cell lines. Finally, FOXO4 expression negatively correlated with smooth muscle differentiation in the LMS cell lines. Collectively, these findings support further study of the hypotheses.

Contractile and synthetic smooth muscle cells are characterized by changes in smooth muscle gene expression; differentiation is indicated by upregulation of these genes. STS39 was found to be the most differentiated of the LMS cell lines, as it had the highest expression of ACTA2, MYLK, TAGLN, and CNN1, the second highest expression of DES and CALD1, and protein expression of all three SMC markers (SMA,
desmin, and h-caldesmon) that were assayed. STS210 was the second most
differentiated cell line, as it had the highest expression of DES and CALD1, the second
highest expression of ACTA2, TAGLN, MYLK, MYH11, and CNN1, and protein
expression of one of three smooth muscle markers (desmin). SK-LMS-1 was the least
differentiated of the LMS cell lines. It has been reported that uterine LMS cell lines have
a less-differentiated smooth muscle cell phenotype (Kimura et al., 2010); the
downregulation of smooth muscle genes in SK-LMS-1 supports this notion. MYLK,
MYH11, and CALD1 were differentially expressed between abdominal and extremity
samples, with overexpression in abdominal samples. This finding supports the
differential expression of smooth muscle genes between abdominal and extremity
samples that is noted in the literature (Italiano et al., 2013). Smooth muscle genes were
not differentially expressed between abdominal and uterine samples, extremity and
uterine samples, or uterine and non-uterine samples. Only 15 samples were used in this
study and equal numbers of samples from each site or differentiation score were not
included (Table 4-2). To further investigate if the smooth muscle genes are differentially
expressed by site in the patient samples, a larger sample set should be used with equal
numbers of samples from different sites and differentiation scores. Additionally, the
mean expression of DES in extremity patient samples was less than the standard
deviation. The mean expression level calculated from a small number of samples is
sensitive to outlier expression values; as a result, the standard deviation is greater than
the mean. Expression profiling of additional samples may result in a larger mean than
the standard deviation.
Since FOXO4 is phosphorylated at Thr32, Ser197, and Ser262 by p-AKT(Thr308), it was necessary to determine if AKT is phosphorylated at Thr308. The phosphorylation of AKT at Thr308 in LMS cell lines STS210, STS39 and SK-LMS-1 indicates that the PI3K/AKT pathway is activated in these cell lines. These findings complement the detection of p-AKT(Ser473) in STS39 and SK-LMS-1 (Babichev et al., 2016), confirming that this pathway is activated in LMS. In the regulatory mechanism elucidated in rat aortic smooth muscle, the PI3K/AKT pathway is activated by insulin signaling (Liu et al., 2005). IGF-1 expression correlated with smooth muscle differentiation in the LMS cell lines tested and IGF1R was differentially expressed between LMS samples with differentiation scores of 1 and 3 (data not shown). Functional studies must be conducted to confirm whether insulin signaling is a characteristic of LMS and activates the PI3K/AKT pathway.

The MYOCD gene has been reported to be overexpressed in well-differentiated LMS (Pérot et al., 2009) and downregulated in dedifferentiated uterine LMS cell lines (Kimura et al., 2010). An independent study determined that MYOCD is the most differentially expressed gene in retroperitoneal LMS compared to extremity LMS (Italiano et al., 2013). We investigated whether MYOCD expression correlates with smooth muscle differentiation and if MYOCD is differentially expressed by site. MYOCD was detected by RNA-Seq and western blotting in cell lines STS210, derived from a metastatic buttock lesion, STS39, derived from a paravaginal tumor, and SK-LMS-1, derived from a uterine LMS. MYOCD protein has been reported to not be detected in uterine cell lines including SK-LMS-1 by western blotting (Kimura et al., 2010). Our finding challenges this and suggests that MYOCD may be expressed in LMS cell lines,
regardless of the disease site. MYOCD expression was correlated with smooth muscle differentiation in the LMS cell lines. STS39 had the highest expression levels of MYOCD, ACTA2, TAGLN, MYLK, and CNN1 among the LMS cell lines, while SK-LMS-1 had the lowest expression levels of MYOCD and all smooth muscle genes analyzed aside from MYH11. Furthermore, SMA, desmin and h-caldesmon were detected in STS39 by ICC while desmin was weakly detected in SK-LMS-1. This correlation supports the findings that myocardin is markedly downregulated in de-differentiated LMS cell lines and overexpressed in well-differentiated LMS (Kimura et al., 2010; Pérot et al., 2009).

Analysis of MYOCD expression in the LMS patient samples indicated that mean MYOCD expression is greater in abdominal samples than in extremity or uterine samples, though this differential expression was not statistically significant. The study that reported MYOCD overexpression in well-differentiated retroperitoneal LMS did not examine moderately differentiated or de-differentiated LMS (Pérot et al., 2009). Mean MYOCD was higher in well-differentiated LMS than in de-differentiated LMS, though statistically significant differential expression was not detected. The standard deviation of MYOCD expression in extremity, uterine, and differentiation score 3 patient samples was greater than the mean expression; since the sample size is limited, the mean is sensitive to outlier expression values, and thus the standard deviation is greater than the mean. Analyzing a larger sample set may result in mean expression levels larger than the standard deviation. Additionally, the sample set should be expanded to include equal numbers of samples from different sites and differentiation score in order to
confirm the absence of differential expression of *MYOCD* by site and degree of differentiation in LMS.

Loss of *FOXO4* expression has been reported to result in decreased smooth muscle marker expression (Liu *et al.*, 2005). *FOXO4* expression was lowest in STS39, the most differentiated LMS cell line, and highest in SK-LMS-1, the least differentiated LMS cell line. Thus, *FOXO4* negatively correlated with smooth muscle differentiation in LMS cell lines. In future studies, the sample set of LMS cell lines should be expanded and protein lysates from LMS samples should be included in order to confirm this correlation. Statistically significant differential expression of *FOXO4* by degree of differentiation or site was not detected in the LMS patient samples. The analysis of differential *FOXO4* expression was constrained by the limited number of samples examined and the unequal number of samples representative of each differentiation score or disease site. Sequencing a larger cohort of LMS samples with equal representation from each differentiation score may aid in identifying differential expression of *FOXO4*.

In rat aortic smooth muscle cells, phosphorylation-induced translocation of *FOXO4* to the nucleus results in dissociation of *FOXO4* from the *MYOCD*-SRF complex and the expression of smooth muscle genes (Liu *et al.*, 2005). Currently, p-FOXO4(Ser197) western blots are being performed on protein lysates from the three LMS cell lines to characterize *FOXO4* phosphorylation in the LMS cell lines and determine if it correlates with smooth muscle differentiation and *MYOCD* expression in the cell lines. One of the challenges of performing these studies is identifying an adequate antibody for western
blotting that specifically detects p-FOXO4(Ser197). Two p-FOXO4(Ser197) antibodies have been tested thus far. Further optimization of this assay is required.

The aim of this chapter was to determine if dysregulation of FOXO4 expression and nuclear exclusion of FOXO4 induced by PI3K/AKT pathway activation enhance MYOCD transcriptional activity resulting in SMC differentiation in LMS. Smooth muscle differentiation was characterized in the LMS cell lines and patient samples. p-AKT(Thr308) was detected in the LMS cell lines, indicating activation of the PI3K/AKT pathway in LMS. MYOCD expression correlated with smooth muscle differentiation in the LMS cell lines. FOXO4 expression negatively correlated with smooth muscle differentiation in the LMS cell lines. Confirming these correlations and characterizing FOXO4 phosphorylation with a sample set comprised of additional LMS cell lines and patient samples would support further investigation of the hypothesis. Ultimately, elucidating how smooth muscle differentiation is mechanistically regulated in LMS may advance our understanding of the molecular events in LMS and present the possibility of developing differentiation therapies for this disease.
Chapter 6  Concluding Summary, General Discussion, Future Directions, and Translational Relevance
6.1 Concluding Summary

Leiomyosarcoma (LMS) is primarily treated with surgery, while conventional chemotherapy is used to treat advanced and metastatic disease and radiation therapy is used in the neoadjuvant/adjuvant setting and for palliation. Unfortunately, outcomes for LMS patients are limited. The 5-year disease-specific survival for retroperitoneal LMS is 67% yet patients continue to succumb to disease in the long term (Gladdy et al., 2013). Recurrence occurs in 33% of patients with extremity LMS (Gladdy et al., 2013) and the median overall survival for uterine LMS patients is 45 months (Tirumani et al., 2014). Personalized medicine informed by the molecular events that drive LMS is needed to improve outcomes. However, the genetic drivers, molecular subtypes and dysregulated developmental pathways in LMS are largely undefined. The purpose of this thesis is to better define the molecular events in LMS by 1) investigating if gene fusions and molecular subtypes exist and 2) elucidating the mechanism regulating smooth muscle differentiation in LMS. In my thesis, three distinct but interrelated studies were performed to examine these two aims.

A preliminary investigation by our group identified two novel fusion transcripts (YAF2-PRICKLE1 and ESR1-CDC92) in a primary LMS cell line using RNA Sequencing (RNA-Seq). YY1-associated factor 2 (YAF2) (Madge et al., 2003), prickle-like protein 1 (PRICKLE1) (Kaucka et al., 2013), and estrogen receptor 1 (ER) (Leitao et al., 2012; Leitao et al., 2004; Kelley et al., 2004) are known to biologically function in cancer development and progression. Since YAF2-PRICKLE1 and ESR1-CDC92 fusion transcripts were discovered in a primary LMS cell line, I hypothesized that they are
recurring in LMS. Firstly, YAF2-PRICKLE1 and ESR1-CCDC92 were confirmed in the primary LMS cell line they were discovered in using RT-PCR. Secondly, spanning reads with the YAF2-PRICKLE1 and ESR1-CCDC92 sequences were detected in the index case, but neither fusion transcript was detected as a fusion event based on the parameters used for fusion detection with RNA-Seq. Lastly, YAF2-PRICKLE1 and ESR1-CCDC92 were not recurrent in the additional LMS patient samples tested. These findings may suggest that these fusion transcripts are rare events and thus may be recurrent in a larger cohort of samples.

Several potential molecular LMS subtypes have been identified by several research groups (Guo et al., 2015; Italiano et al., 2013; Beck et al., 2010). One study has reported that unsupervised hierarchical clustering of expression profiles of 51 LMS patient samples indicated three molecular subtypes (Beck et al., 2010). Though these subtypes were confirmed in a later study (Guo et al., 2015), they were not found to associate with clinical parameters such as site of disease, possibly due to limited clinical annotation of the patient samples used. A recent study has reported that comparative genomic hybridization (CGH) analysis of 68 LMS patient samples indicated two genomic subtypes which may be associated with site of disease: a genomically stable subtype primarily consisting of retroperitoneal samples (69%) and a genomically unstable subtypes primarily consisting of extremity samples (76%) (Italiano et al., 2013). The association of these genomic subtypes with site, an important prognostic factor in LMS, may suggest that there is a molecular basis for the site-associate differences in outcomes. I hypothesized that LMS transcriptomes may cluster by site of disease. Firstly, a cluster of abdominal and extremity LMS formed as a result of unsupervised
hierarchical clustering of the transcriptomes of from the discovery set of patient samples. Secondly, transcriptomes from uterine LMS formed a separate, heterogeneous group. Lastly, differentially expressed genes between the clusters were analyzed for gene set and functional annotation clustering enrichment and several statistically significant clusters were identified. These findings may suggest that LMS features transcriptomic subtypes and that abdominal and extremity LMS may be molecularly distinct from uterine disease.

Evidence from numerous studies suggests that smooth muscle differentiation may be dysregulated in LMS and thus associated with leiomyosarcomagenesis (Danielson et al., 2010; Pérot et al., 2009). Myocardin (MYOCD), the master regulator of smooth muscle gene expression, has been reported to be amplified and overexpressed in 53% of well-differentiated LMS (Italiano et al., 2013) and downregulated in de-differentiated uterine LMS cell lines (Kimura et al., 2010). Additionally, knockdown of MYOCD has been reported to result in a loss of smooth muscle differentiation in an LMS cell line (Pérot et al., 2009). PI3K/AKT pathway activation due to IGF signaling regulates smooth muscle differentiation by phosphorylating nuclear FOXO4. Phosphorylated FOXO4 is translocated to the cytoplasm, thereby negating the inhibitory effect of FOXO4 on MYOCD transcriptional activity (Liu et al., 2005). Furthermore, knockdown of FOXO4 in rat aortic smooth muscle cell lines has been reported to result in upregulation of smooth muscle markers (Liu et al., 2005). I hypothesized that (1) dysregulated FOXO4 expression and (2) phosphorylation of FOXO4 may regulate SMC differentiation in LMS. Firstly, AKT was found to be phosphorylated at Thr308 in 3 LMS cell lines, indicating that the PI3K/AKT pathway is activated in LMS. Additionally, MYOCD expression
correlates with smooth muscle differentiation in the LMS cell lines. Finally, we report that FOXO4 expression negatively correlates with smooth muscle differentiation in the LMS cell lines. These findings may suggest that dysregulated FOXO4 expression regulates smooth muscle differentiation in LMS and support further investigation of the role of FOXO4 in LMS smooth muscle differentiation.
6.2 General Discussion

Per the “multiple paper format”, the findings in this thesis have been examined in separate discussion sections in each data chapter (sections 4.5 and 5.5). Here, this work will be discussed broadly in terms of its contributions to the literature and to defining molecular events in LMS. In addition, the limitations of the approaches undertaken will be addressed.

6.2.1 Defining Molecular Events in LMS

The discovery of a potential LMS subtype characterized by genomic stability (Italiano et al., 2013) suggested the possibility of identifying translocations in LMS, as many soft tissue sarcomas with stable genomes have been characterized by recurrent gene fusions (Mertens et al., 2016). In this study, two novel fusion transcripts, YAF2-PRICKLE1 and ESR1-CCDC92, were identified in a primary LMS cell line and detected spanning reads of these transcripts in the tumor from which the cell line was derived. Though these transcripts were not recurrent in the small sample set examined in this study, they may be rare events that could be found to be recurrent upon investigation of a larger group of samples. One study has identified a translocation t(6;14) in 5% of the LMS samples that were investigated (de Graaff et al., 2015), though these translocations were not identified in our sample set. Our findings support the notion that a subset of LMS may be characterized by translocations and complement the recently reported translocation in the literature. It would be informative to determine if the YAF2-PRICKLE1, ESR1-CCDC92, and t(6;14) events occur in genomically stable LMS.
samples, so that future studies of LMS-specific translocations can focus their investigations on this subset of LMS.

Several LMS subtypes have been reported in the literature. In particular, two subtypes were discovered based on the genomic stability of LMS (Italiano et al., 2013). The genomically stable subtype was enriched for retroperitoneal tumors (69%), while the genomically unstable subtype primarily consisted of extremity tumors (76%). Uterine LMS was not investigated in the study which identified two genomic subtypes. Three subtypes were identified based on clustering of expression profiles in two studies by another research group (Guo et al., 2015; Beck et al., 2009). These studies were limited by the absence of clustering by prognostic factors such as site of disease and grade and the lack of analysis of the differential expression of biologic pathways. In this study, transcriptomes from abdominal and extremity LMS clustered tightly and distinctly, while those from uterine LMS formed a separate group. These clusters did not associate with grade, differentiation, mitotic rate, or necrosis, possibly due to the limited number of samples representing the degrees of these factors. These findings may suggest that abdominal and extremity samples together form a distinct molecular subtype of LMS. Interestingly, the similarity of abdominal and extremity LMS transcriptomes contrasts with the site-specific genomic subtypes reported in the literature. Retroperitoneal and extremity LMS have different rates of recurrence and long term disease specific-survival; the site-specific genomic subtypes may underlie these differences, while the differential expression of a subset of genes may account for the differences in outcome despite the transcriptomic similarity of LMS from these sites. These findings may also support the notion of uterine LMS being biologically different from LMS originating at
other sites (Skubitz and Skubitz, 2003). In this study, the clusters did not associate with grade, differentiation, mitotic rate, or necrosis; this may be due to the limited sample set and unequal distribution of the samples from different the differentiation, mitotic and necrosis scores. In contrast to studies that identified three LMS subtypes based on expression profiles, this study was able to associate the transcriptomic clustering with site of disease, since the samples used here were well-annotated while 20% of the samples used in the former studies were of unknown site. Furthermore, the subtypes that those studies reported were not confirmed by the cluster analysis of the samples used in this study. Notably, those studies performed expression profiling using 3SEQ, a modified RNA-Seq protocol in which only the 3'-ends of genes are sequenced, while this study performed transcriptome profiling using total RNASeq. As a result, the subtypes identified in those studies might not emerge from the transcriptome profiles compiled in this study, as the data acquired from 3SEQ is dissimilar from the data acquired from total RNA-Seq. Thus, the site-specific clustering of LMS transcriptomes must be validated with a larger sample set in order to determine if LMS is characterized by these transcriptomic subtypes.

Recent evidence suggests that smooth muscle differentiation and oncogenesis may be associated in LMS (Danielson et al., 2010; Rubio et al., 2010). However, the regulation of smooth muscle gene expression and differentiation in LMS has not been investigated. In rat aortic smooth muscle cells, phosphorylation-induced translocation of FOXO4 to the cytoplasm negates the inhibitory effect of FOXO4 on MYOCD transcriptional activity and results in the expression of smooth muscle genes. The study in this thesis examined whether FOXO4 expression and phosphorylation may regulate
smooth muscle differentiation in LMS. Expression levels of smooth muscle genes in the three LMS cell lines (STS210, STS39, and SK-LMS-1) indicated that STS39, derived from a paravaginal LMS, was the most differentiated of the cell lines, while SK-LMS-1, derived from a uterine LMS, was the least differentiated of the cell lines. This confirms the finding that SK-LMS-1 has a dedifferentiated phenotype due to the weak expression of smooth muscle genes (Kimura et al., 2010). Smooth muscle genes have been reported to be differentially expressed between retroperitoneal and extremity LMS (Italiano et al., 2013). However, smooth muscle markers were not differentially expressed by site; a small sample set was used with an uneven distribution of samples across uterine, extremity and retroperitoneal sites, so it is possible that analyzing smooth muscle marker expression in additional samples may indicate statistically significant differential expression by site. FOXO4 is known to be phosphorylated by p-AKT(Thr308) at Thr32, Ser197, and Ser262. p-AKT (Thr308) was present in the three LMS cell lines examined. p-AKT(Ser197) has been reported in ~80% of LMS (Hernando et al., 2007) and was detected in two of the three LMS cell line in a previous study (Babichev et al., 2016). These findings suggest that p-AKT(Thr308) is also prevalent in LMS. p-AKT(Thr308) levels did not correlate with smooth muscle differentiation in the LMS cell lines. This may suggest that the phosphorylation and nuclear exclusion of FOXO4 does not correlate with smooth muscle differentiation in these cell lines. p-FOXO4(Ser197) western blots must be performed on the LMS cell lines to ascertain whether FOXO4 phosphorylation correlates with smooth muscle differentiation in these cell lines. MYOCD has been reported to be overexpressed in 53% of well-differentiated LMS (Pérot et al., 2009) and expressed at low levels in uterine LMS cell lines (Kimura et al., 2010). MYOCD expression correlated with smooth muscle differentiation in the LMS cell lines:
it was highly expressed in STS39 and weakly expressed in SK-LMS-1. This confirms
the finding the MYOCD is weakly expressed in SK-LMS-1 (Kimura et al., 2010).
Interestingly, MYOCD protein was detected in SK-LMS-1 by western blot; this
challenges the finding that MYOCD protein is not detected in SK-LMS-1 by western blot
(Kimura et al., 2010). This discrepancy may be due to several factors. Firstly, Kimura et al
generated their own anti-MYOCD antibody using recombinant human MYOCD
proteins as an antigen, while we used the commercial antibody which has been used to
assay MYOCD protein levels in LMS patient samples (Pérot et al., 2009). It is possible
that the antibody used in that study may not have been as sensitive as the one used in
this study. Secondly, the SK-LMS-1 protein lysates used for western blotting may have
been extracted from cells at different passage numbers. MYOCD protein expression
may have decreased in the cells which Kimura et al extracted protein from; conversely,
MYOCD protein expression may have increased in the cells which we extracted protein
from. This possibility highlights a limitation in using cell lines as a biological reagent.
MYOCD has been reported to be overexpressed in 53% of well-
differentiated LMS
(Pérot et al., 2009). Pérot et al did not examine LMS samples moderately differentiated
or de-differentiated LMS samples. MYOCD has also been found to be differentially
expressed between retroperitoneal and extremity samples, with greater expression in
retroperitoneal samples (Italiano et al., 2013). Italiano et al did not include samples from
other sites, such as the uterus. In this study, we examined the differential expression of
MYOCD by degree of differentiation and site. Though the mean expression level of
MYOCD was higher in both well-differentiated and abdominal samples, MYOCD was
not differentially expressed by degree of differentiation or by site. A small number of
samples across different sites and differentiation scores was used in this analysis;
analyzing MYOC expression in a larger number of samples may indicate increase the statistical power and indicate differential expression by degree of differentiation or site. Finally, we report that FOXO4 expression negatively correlates with smooth muscle differentiation in the LMS cell lines. This correlation may suggest FOXO4 expression regulates smooth muscle differentiation in LMS. FOXO4 was not differentially expressed by degree of differentiation or site. It should be noted that a limited number of samples were used with an uneven distribution across different sites and differentiation scores, so it is possible that analyzing smooth muscle marker expression in additional samples may indicate statistically significant differential expression by degree of differentiation or site.

6.2.2 Limitations

The approaches used to investigate the aims of this thesis have several limitations. Much of the work in this thesis was based upon the transcriptome data generated from the RNA-Seq of 15 formalin-fixed, paraffin-embedded (FFPE) LMS patient samples. RNA extracted from FFPE samples may not be ideal for RNA-Seq. The chemical reaction between formaldehyde and RNA results in the crosslinking of RNA with protein and the covalent modification of RNA through the addition of monomethyl groups to nucleotide bases. These modifications make RNA rigid and prone to shearing, and thus may compromise the ability of RNA-Seq to accurately profile the transcriptomes from FFPE samples. In this study, RNA quality was determined to be adequate using Nanodrop and BioAnalyzer. However, GAPDH and HPRT, housekeeping genes used as internal controls for the RT-PCR experiments, were only amplified in 4/15 patient
samples, suggesting that the RNA from the FFPE samples may have degraded after extraction and quality assessment but before cDNA synthesis. Additionally, the fragmented RNA extracted from FFPE samples may inhibit the PCR reaction. RNA from FFPE samples has been used in a study profiling transcriptomes from bladder cancer patient samples (Liu et al., 2014). The RNA extracted from fresh frozen patient samples is often of a higher quality from that extracted from FFPE samples due to the lack of chemical modification. However, fresh frozen samples are limited for research studies; FFPE samples were chosen instead of fresh frozen samples because the nature of the exploratory nature of the work in this thesis. Findings from this work could then be validated in larger studies utilizing fresh frozen samples.

The preliminary investigation of the fusion transcripts and the regulation of smooth muscle differentiation involved the use of LMS cell lines. Three of these (STS210, STS39, and STS54) were primary cell lines derived in the lab from resected tumors, while the fourth cell lines (SK-LMS-1) is commercially available. LMS cell lines are valuable research tools because they are sustainable, unlike banked patient samples which are finite, and model the disease. For these reasons, cell lines are ideal reagents for biomarker discovery and characterization studies. However, these cell lines differ from patient samples in several respects. Firstly, cells in culture may acquire genomic alterations through repeated passaging, resulting in cells that no longer resemble the tumors from which they were derived. Secondly, these cultured cells are grown in monolayers, while tumors cells grow in a three-dimensional microenvironment. As a result, the cultured cells cannot recapitulate the cell-cell and cell-matrix interactions that are characteristic of the tumor cells from which they were derived. Thirdly, cells cultured
in vitro are not grown with associated stromal cells, further contributing to a different microenvironment for the cultured cells. For these reasons, it is possible that these LMS cell lines may not fully recapitulate the disease. As a result, it is crucial that LMS patient samples be used in future studies to validate the findings from the cell lines.

RNA-Seq allows for unbiased gene fusion detection by providing a nucleotide-level and transcriptome-wide profile of a cell line or a patient sample. Unlike cytogenetic techniques relying on chromosomal banding analysis, RNA-Seq enables the detection of unbalanced chromosomes and intrachromosomal rearrangements involving genes in close proximity (Mertens et al., 2015). However, fusion detection using RNA-Seq is constrained by a number of limitations. Firstly, errors may be introduced during cDNA preparation, resulting in inaccurate sequences. Secondly, errors may occur when analyzing sequences as a result of using reference genomes with errors. Errors in sequence analysis may also occur due to sequence similarities between genes. Thirdly, false-negative and false-positives may arise during fusion detection due to suboptimal RNA quality decreasing the number and quality of reads. Lastly, rare events, those that result in transcriptional silencing, or subclonal fusions are difficult to detect using RNA-Seq (Mertens et al., 2015). Collectively, these factors impose limitations on fusion detection using RNA-Seq.

PCR-based fusion detection approaches allow for the detection of fusion variants. More than 90% of Ewing’s sarcomas contain EWSR1-FLI1 gene fusions, of which there are many variants arising from different combinations of exons from EWSR1 and FLI1 (de Alava et al., 1998). In this investigation, PCR primers were designed to amplify only one
breakpoint of both $YAF2$-$PRICKLE1$ and $ESR1$-$CCDC92$ fusion transcripts. Thus, the PCR strategies may not have detected variants of these fusion transcripts in the LMS patient samples. Neither fusion transcript was detected in the patient samples using RNA-Seq; however, fusion transcript detection with RNA-Seq may not be sensitive enough to detect rare events. Future studies investigating the recurrence of these fusion transcripts should incorporate PCR strategies that will assay for variants of the $YAF2$-$PRICKLE1$ and $ESR1$-$CCDC92$ fusion transcripts.

Another limitation of these studies was the size and composition of the set of LMS patient samples used. Banked patient samples are limited, finite resources and consequently must be used judiciously, often to validate hypotheses generated from preliminary investigations using cell lines. Here, a discovery set of 15 LMS patient samples was used to determine if the fusion transcripts detected in a primary LMS cell line were present in the tumor from which the cell line was derived. The transcriptomes generated from these samples also underwent unbiased clustering analysis to determine if they group by site. Additionally, this transcriptomic data was used towards investigating whether FOXO4 phosphorylation regulates smooth muscle differentiation in LMS. The sample set was small due to the cost of performing RNA-Seq on each sample. The limited sample size may have contributed to not finding recurrences of $YAF2$-$PRICKLE1$ and $ESR1$-$CCDC92$ fusion transcripts in LMS patient samples. A translocation t(6;14) was recently identified in 2/39 cases of LMS (de Graaff et al., 2015). Since $YAF2$-$PRICKLE1$ and $ESR1$-$CCDC92$ may be rare events, more than 15 samples may be required to determine if they are recurrent. Furthermore, the discovery set did not include a cohort of metastatic samples. $YAF2$-$PRICKLE1$ and $ESR1$-
CCDC92 fusion transcripts were discovered in a cell line derived from a metastatic buttock lesion and spanning reads of these transcripts were detected in the index case. Notably, the primary tumor of the inferior vena cava from which the metastatic index case arose was not accessible as surgery was performed at another institution. Since YAF2-PRICKLE1 and ESR1-CCDC92 were detected in metastatic sample, it is possible that they are recurrent in metastatic LMS. Thus, future studies examining the recurrence of these fusion transcripts should include a cohort of metastatic samples.

The limited number of LMS patient samples and normal smooth muscle cell lines also constrained the statistical power when analyzing the differential expression of MYOCD and FOXO4 between samples from different disease sites or different differentiation scores. The mean expression levels of DES in extremity LMS and MYOCD in differentiation score 3, extremity and uterine LMS were less than the standard deviations of the expression levels. This may have been due to the small sample size in conjunction with outlier expression levels which were much higher or lower than the mean. Since the mean is sensitive to these outliers and the sample size is not large enough to reduce their impact on the mean, the standard deviation is larger than the mean. Additionally, the analysis of the differential expression of these genes was constrained by the limited number of normal smooth muscle cell lines. Publicly available expression data from one normal smooth muscle cell line, HCASMC, was used to analyze differential expression between LMS samples and smooth muscle. However, the statistical significance of the expression differences between the LMS samples and the normal smooth muscle cannot be calculated when only one smooth muscle sample is used. Future investigation of the differential expression of these
genes should be preceded by a power calculation in order to determine an adequate number of samples so that statistically significant conclusions can be drawn.

The cost of performing RNA-Seq poses a constraint on the analysis of the expected variance in the primary tumor and cell line data. Technical replicates were not performed for any of the samples due to the high cost of sequencing. Without technical replicates, it is difficult to distinguish the variance inherent to the sensitivity of the sequencing chemistry from the biological variance associated with the RNA stability from the FFPE samples.

RNA and protein extracted from LMS cell lines STS210, STS39 and SK-LMS-1 were used to investigate the mechanism regulating smooth muscle gene expression in LMS. The RNA and protein were extracted from cells with different passage numbers. As cultured cells are passaged, they may undergo genotypic and phenotypic alterations. Thus, it is possible that results generated from RNA-Seq and western blotting may be representative of incongruous transcriptomic and proteomic states. In order to characterize complementary RNA and protein expression from cell lines, RNA and protein should be extracted from the same pool of cells.

Lastly, the investigation of the mechanism regulating smooth muscle marker expression in LMS was limited by the use of only three LMS cell lines. Three LMS cell lines have been derived from resected tumors in the Gladdy Lab; two of these were used for this study, as the third is slow-growing and thus was not ideal to use given the time frame of these studies. The third LMS cell line used was the commercially available SK-LMS-1.
Future studies can expand this sample set of cell lines by deriving new cell lines from tumors as they are resected at Mount Sinai Hospital and by collaborating with other research groups to include their cell lines. Patient samples could not be used to study this aim for two reasons. Firstly, adequate MYOCD, FOXO4 and p-FOXO4(Ser197) IHC antibodies are not available to characterize MYOCD expression and FOXO4 phosphorylation in slides cut from FFPE blocks of the patient samples. Secondly, fresh frozen samples were not used for western blotting due to the discovery nature of this project, as these samples are rare and non-renewable. Should the hypothesized regulatory mechanism be confirmed in the LMS cell lines, future studies may utilize fresh frozen samples to validate these findings.
6.3 Future Directions

6.3.1 Identifying Recurrent Gene Fusions in LMS

To determine if there are recurrent fusion transcripts in LMS, RNAseq and RT-PCR will be performed in parallel on 120 fresh frozen, untreated LMS patient samples. These samples will include 30 abdominal, 30 extremity, and 30 uterine samples and 30 LMS metastases. This investigation would include many more samples than used in the study conducted in the thesis and would have an equal number of samples from the disease sites of interest. Furthermore, it would use fresh frozen samples instead of FFPE samples, with the aim of harvesting higher quality RNA. The inclusion of a cohort of metastatic samples may increase the likelihood of identifying recurrence of YAF2-PRICKLE1 and ESR1-CCDC92, since these fusion transcripts were identified in a cell line derived from a metastatic buttock lesion.

Should recurrent fusion transcripts be detected in LMS, it would be necessary to determine if they are LMS-specific and result from genomic rearrangements. To this end, RT-PCR will be performed on 80 patient samples from four other sarcoma subtypes: 20 undifferentiated pleomorphic sarcomas (UPS), 20 myxofibrosarcoma, 20 pleomorphic rhabdomyosarcomas and 20 fibrosarcomas. These are other common sarcoma subtypes that are not associated with known gene fusions. To determine if these fusion transcripts are somatically acquired, RT-PCR will be performed on peripheral blood matched to the LMS patient samples. To determine if identified fusion transcripts result from genomic rearrangements, FISH experiments will be designed to
show that the fusion transcripts arise from genomic rearrangements. Long-range PCR will be performed to characterize the genomic breakpoints of each gene fusion.

6.3.2 Characterizing the Oncogenic Potential of Recurrent Gene Fusions in LMS

To determine the oncogenic potential of recurrent gene fusions, these gene fusions will be knocked down in fusion-positive LMS cell lines using siRNA or CRISPR interference (CRISPRi). These cell lines will undergo MTT assay to measure the cell proliferation rate, Boyden chamber assay to assess changes in migration and invasion capabilities, and soft agar assay to characterize the anchorage independent growth of these cells. Colony formation assay and cell cycle analysis will also be performed to characterize changes in the oncogenic potential of fusion positive LMS cell lines upon knockdown of the gene fusions. Furthermore, these gene fusions will be ectopically expressed in fusion-negative LMS cell lines by lentiviral transduction and the above assays will also be performed on these cells. Finally, fusion positive cells will be transplanted into immunodeficient mice and to determine if the fusion is associated with tumor growth. Collectively, these experiments will characterize the oncogenic potential of the gene fusions in LMS. To determine which oncogenic pathways are activated in fusion-positive tumors, differential gene expression between fusion positive and fusion negative tumors will be analyzed. Gene set enrichment analysis will be performed on differentially expressed genes and enriched pathways will be identified and further analyzed to identify how these gene fusions mechanistically contribute to leiomyosarcomagenesis.
6.3.3 Validating LMS Transcriptome Clustering

Unsupervised hierarchical clustering analysis of transcriptomes from 30 abdominal, 30 extremity and 30 uterine primary LMS samples will be performed to validate the clustering of abdominal and extremity samples and confirm that uterine transcriptomes form a separate group. In the preliminary study reported in this thesis, the unsupervised hierarchical clustering analysis was performed on 7 abdominal, 4 extremity and 4 uterine samples; the number of samples per site was small and unequal across the sites. A limited sample set may not be representative of the transcriptomic heterogeneity in LMS. Furthermore, the uneven distribution of samples from different disease sites may limit conclusions about the association of these clusters with site, as the composition of the sample set impacts how the samples cluster. Increasing the sample size and using equal numbers of samples from the different disease sites would facilitate the confirmation of the clustering observed in the preliminary study and determine if LMS features site-specific transcriptomic subtypes. Additionally, a separate clustering analysis will be performed which will include 30 metastatic samples to see if metastatic LMS clusters as a separate molecular subtype from primary samples.

6.3.4 Investigate Alternative Splicing in LMS

Alternative splicing of mRNA transcripts has not been investigated in LMS. Splicing can regulate gene expression by producing cell-specific functional transcripts. Cancer-associated splicing events were identified in a study of hundreds of matched tumor and normal RNA-Seq samples from eight solid tumor subtypes (Danan-Gotthold et al.,
2015). Subtype-specific differentially spliced genes and splice isoforms of genes such as CDK4 have been discovered in breast cancer (Eswaran et al., 2013). Differentially spliced genes and splice isoforms may also underlie biological differences and clinical parameters in LMS. To determine if there are LMS subtypes based on differential splicing of genes, RNA-Seq data from fresh frozen LMS patient samples will be analyzed for splice isoforms and these will undergo unsupervised clustering analysis to uncover clustering of samples based on differential splicing. These clusters will be examined to determine if they correlate with clinical parameters of LMS, such as site, grade, recurrence, and metastasis. Additionally, cluster specific splice isoforms will be identified to serve as makers for these subtypes of LMS.

6.3.5 Whole Genome Profiling of LMS

There is a limited understanding of the genomic landscape of LMS. Thus, there is a clear need to perform whole genome profiling of an adequate number of LMS patient samples. The Gladdy and Shlien Labs have collaboratively planned to characterize the genomic complexity of LMS by performing whole genome sequencing. These experiments aim to identify driver mutations in LMS and molecular subtypes by mutation signature. 13 retroperitoneal and 13 extremity primary LMS patient samples will undergo whole genome sequencing. Untreated, fresh frozen samples will be used with matched blood as a germline control. The sequencing of these samples will result in a survey of the mutational landscape in LMS, from single nucleotide substitutions to genomic rearrangements. In order to identify potential driver mutations, three analytical approaches will be used. Firstly, the genomic profiles from these samples will be
analyzed for recurrently mutated genes, such as genes with truncating mutations or those with missense changes at the same amino acid. Secondly, the genomic profiles will be analyzed for driver mutations identified in other cancers, as documented in databases such as COSMIC. Thirdly, the genomic profiles will be analyzed for recurrent non-coding mutations with gene regulatory potential, such as those occurring at transcription factor binding sites, DNase I hypersensitive sites, or histone H3 acetylated at lysine 27 (H3K27ac). In order to identify mutational subtypes of LMS, unsupervised hierarchical clustering will be performed on the mutation profiles from the LMS patient samples. Additionally, clusters will be analyzed for correlation with clinical parameters such as disease site. The genomic studies planned by the Gladdy and Shlien Labs would complement the transcriptomic studies to create a more complete “omics” landscape of LMS.

6.3.6 Epigenomic Profiling of LMS

Little is known about the epigenomics of LMS. One study performed genome-wide DNA methylation analysis of 3 uterine LMS, 3 uterine leiomyoma, and 3 myometrial samples and reported that uterine LMS exhibits epigenomic features known to be common among malignant solid tumors, including hypermethylation of PcG target genes and hypomethylation of large genomic blocks (Miyata et al., 2015). This study was limited by the number of LMS samples used and the use of only uterine LMS. Therefore, epigenomic profiling of an adequate number of LMS samples is required to better understand the epigenomics of LMS. Genome-wide DNA methylation analysis will be performed on 30 abdominal, 30 extremity, and 30 uterine samples to characterize the
methylation profiles of the LMS samples. Unsupervised hierarchical clustering of the methylation profiles will be performed to identify potential epigenomic subtypes in LMS. Differentially methylated regions will be identified between the subtypes and analyzed to identify recurrent methylation signatures for each cluster of samples. The epigenomic subtypes will be analyzed for correlation with clinical parameters such as site of disease and grade. These experiments will further our understanding of how gene expression is regulated in LMS and present an additional approach to identifying molecular subtypes of LMS.

6.3.7 Investigating the Regulation of Smooth Muscle Gene Expression in LMS

To further examine the mechanism regulating smooth muscle marker expression in LMS, several experiments can be performed. Firstly, immunohistochemical staining for AKT and smooth muscle markers including SMA, desmin, transgelin, h-caldesmon, smooth muscle myosin heavy chain, and smooth muscle myosin light-chain kinase can be performed on FFPE slides from patient samples. MYOCD and p-FOXO4 (Ser197) western blots will be performed using protein extracted from fresh frozen patient samples matched to the FFPE blocks. These experiments will indicate whether FOXO4 phosphorylation and MYOCD overexpression correlate with smooth muscle marker expression. To determine if FOXO4 phosphorylation regulates smooth muscle marker expression, FOXO4 with mutated phosphorylation sites (Thr32, Ser197, Ser262) will be expressed in LMS cell lines grown in serum-free conditions and smooth muscle marker expression will be analyzed. Additionally, FOXO4 will be knocked down and smooth
muscle marker expression will be analyzed. To determine if the inhibitory effect of FOXO4 on smooth muscle gene expression is mediated through an interaction with MYOCD and SRF, coimmunoprecipitation experiments will be performed.

A recent study has reported that nuclear PTEN may regulate smooth muscle differentiation by interacting with the amino terminal domain of SRF and promoting SRF binding to the promoters of smooth muscle genes (Horita et al., 2016). Translocation of nuclear PTEN from to the cytoplasm resulted in decreased expression of smooth muscle markers. Moreover, overall decreased expression of PTEN was found in proliferating smooth muscle cells with lower smooth muscle marker expression (Horita et al., 2016). Notably, 70% of LMS is characterized by loss of PTEN; given these findings, it may be possible that the loss of PTEN results in loss of smooth muscle marker expression and de-differentiation in LMS. PTEN western blots will be performed using fractionated nuclear and cytoplasmic protein lysates from fresh frozen patient samples. These experiments will indicate whether nuclear PTEN correlates with smooth muscle marker expression in LMS. To determine if nuclear PTEN regulates smooth muscle marker expression, PTEN will be knocked down and smooth muscle marker expression will be analyzed. To determine if the nuclear PTEN enhances smooth muscle gene expression through an interaction with SRF, coimmunoprecipitation experiments will be performed.
6.4 Translational Relevance

Defining molecular events in LMS may have profound implications on diagnosis and treatment of this disease. Surgery is the primary treatment for LMS, while chemotherapy and radiation are used to treat advanced disease. There is a demand for improved diagnostic tools and targeted therapies. Currently, the outcomes for LMS patients are poor. While the 5-year disease-specific survival for retroperitoneal LMS is 67%, patients continue to succumb to disease in the long term (Gladdy et al., 2013). 33% of patients with extremity LMS will have at least one recurrence of LMS (Gladdy et al., 2013) and the median overall survival for patients with uterine LMS is 45 months (Tirumani et al., 2014). This work identified two novel fusion transcripts which may be recurrent in LMS, potential transcriptomic subtypes of LMS which suggest that uterine LMS is biologically distinct from abdominal and extremity LMS, and a correlation between decreased FOXO4 expression and increased smooth muscle differentiation in LMS which supports further study of the regulatory role of FOXO4 in leiomyosarcomagenesis. Recurrent gene fusions and transcriptomic subtypes would categorize LMS into subgroups which may be clinically relevant. Fluorescence in situ hybridization (FISH) and RT-PCR assays can be developed to classify fusion-positive tumors and thereby diagnose LMS patients. Pharmacological compounds may be repurposed or developed to target the protein product of the gene fusion or a downstream effector in order to treat the disease. Site-associated molecular subtypes may aid in stratifying in LMS and identifying biological differences which may be used as diagnostic markers or therapeutic targets. Elucidating the mechanism regulating smooth muscle marker expression in LMS may challenge the concept of LMS arising from mature smooth muscle cells and pinpoint
targets which could become the focus of differentiation therapies for high grade disease. While these are preliminary findings, they present promising avenues for future investigations of the molecular events in LMS and may be stepping stones towards facilitating effective diagnostic and therapeutic innovations for this disease.
References


Epigenomic, and Transcriptomic Profiling towards Identifying Omics Features and Specific Biomarkers That Distinguish Uterine Leiomyosarcoma and Leiomyoma at Molecular Levels." **Sarcoma 2015**: 412068.


