Title

Characterizing the transcriptomic and genomic signatures of immune cell infiltration in undifferentiated sarcoma

Author

Graham, Danielle Schwarzbein

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Characterizing the transcriptomic and genomic signatures of immune cell infiltration in undifferentiated sarcoma

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Molecular, Cellular & Integrative Physiology

by

Danielle Graham

2022

ABSTRACT OF THE DISSERTATION

Characterizing the transcriptomic and genomic signatures of immune cell infiltration in undifferentiated sarcoma

by

Danielle Graham

Doctor of Philosophy in

Molecular, Cellular & Integrative Physiology

University of California, Los Angeles, 2022

Professor Thomas G. Graeber, Chair

There is a critical need for more effective systemic therapies for the treatment of soft tissue sarcoma and, specifically, undifferentiated sarcoma. Immunotherapy has shown signs of efficacy for the treatment of soft tissue sarcoma, particularly in undifferentiated sarcoma, though selecting the patients who will benefit from immunotherapy remains difficult and unclear. Further studies are needed to characterize the immune landscape of soft tissue sarcoma and to develop strategies to identify patients that are likely to benefit from immunotherapy.

In this study, I investigated the immunologic heterogeneity and identify transcriptomic and genomic correlates of immune cell infiltration in undifferentiated sarcoma. In doing so, I determined the immune cell landscape, the optimal high-throughput tools, and the transcriptomic

and genomic changes associated with high and low immune cell infiltration in soft tissue sarcoma.

This study synthesized many datasets and data types for a comprehensive analysis of the immune landscape in soft tissue sarcoma. I first characterized the immune cell landscape in soft tissue sarcoma using flow cytometry data from fresh operative samples (n=105) of multiple soft tissue sarcoma subtypes. I then generated a tissue microarray with matched RNA sequencing data from 60 samples of untreated undifferentiated sarcoma to determine the optimal method of in-silico immune deconvolution, which allows for the expansion of this analysis to other next generation sequencing data. Finally, I synthesized multiple publicly available datasets containing next generation sequencing data (RNA sequencing and whole exome sequencing) from undifferentiated sarcoma samples. This data is combined with data from the aforementioned samples for a total of 193 samples and is used to determine the transcriptomic and genomic correlates of immune cell infiltration in undifferentiated sarcoma.

In the analysis of the flow cytometry data, I found that undifferentiated sarcoma tumors are characterized by a myeloid predominance and a relative abundance of suppressor cells, such as Treg cells and CD11b cells. I additionally found that the immune composition of peripheral blood was associated with intratumoral leukocyte infiltration, and specifically that myeloid-predominant tumor and lymphocyte-predominant blood are mutually exclusive. I then determined the optimal in-silico immune deconvolution tool in undifferentiated sarcoma by determining the correlation between mIF and in-silico immune deconvolution scores. Based on these findings, I suggest the following practices when applying in-silico immune deconvolution tools to undifferentiated sarcoma: (1) Use TIMER to define overall immune cell infiltration. (2) Use MCP counter to define monocyte infiltration or use CIBERSORTx, EPIC, quanTIseq,

TIMER, or xCell to define macrophage infiltration. (3) Use caution when using in-silico immune deconvolution tools to define CD8+ T cell infiltration. CIBERSORTx most accurately defines CD8+ T cell immune infiltration, however, there are still many instances when tumors with high CD8+ T cell infiltration will be missed using this technique. (4) Avoid applying in-silico immune deconvolution results to define B cell or CD4+ T cell immune infiltration. Finally, I found that increased copy number changes were associated with low immune cell infiltration in undifferentiated sarcoma. These findings were suggested in both transcriptomic and genomic analyses. Interestingly, this association between CNA and immune invasion were unique to the UPS and DDLPS subtypes of STS, but it was not seen in other subtypes of STS. The mechanisms underlying this association are not clear and warrant further study.

These insights provide necessary information to understand which patients may benefit from immunotherapy and guide future studies to further the treatment of soft tissue sarcoma. These studies provide the groundwork for further investigation in this study of immune cell infiltration in soft tissue sarcoma and provide insights into how we may be able to improve outcomes in this rare and devastating disease. The mechanisms underlying these findings remain unclear and warrant further investigation. A deeper understanding of the drivers of immune cell infiltration, the unique tumor microenvironment in soft tissue sarcoma, and role that chromosomal instability plays in soft tissue sarcoma will hopefully ultimately lead to insights to new, and much-needed, treatments for this disease.

The dissertation of Danielle Graham is approved.

James Economou

Anusha Kalbasi

April Pyle

Thomas G. Graeber, Chair

University of California, Los Angeles

2022

TABLE OF CONTENTS

List of Figures

List of Tables

List of Symbols and Acronyms

BAM - binary tab delimited file containing sequence alignment data

BO112 - double stranded synthetic RNA used in immunotherapy

CIBERSORTx - in-silico immune deconvolution tool

Class. Mono. classical monocyte

CN - copy number

CNA - copy number alteration

CONSORT - Consolidated Standards of Reporting Trials

CTLA4 - cytotoxic T-lymphocyte associated protein 4

DC - dendritic cell

DDLPS - dedifferentiated liposarcoma

DESeq - differential expression analysis

DNA - deoxyribonucleic acid

EGA - European Genome-Phenome Archive

EPIC - in-silico immune deconvolution tool

Fastq - text-based file that contains biological sequence and quality score data

FDCS - follicular dendritic cell sarcoma

FFPE - formalin-fixed paraffin-embedded

GIST - gastrointestinal stromal tumor

gMDSC granulocytic myeloid-derived suppressor cell

GO - gene ontology

GSEA - gene set enrichment analysis

H&E - Hematoxylin and Eosin stain

HLA DR - an MHC class II cell surface receptor

IHC - immunohistochemistry

Intermed. Mono. intermediate monocyte

LPS - liposarcoma

MCP Counter - in-silico immune deconvolution tool

mDC - myeloid dendritic cell

MDSC - myeloid-derived suppressor cell

MFS - myxofibrosarcoma

mIF - multiplex immunofluorescence

mIHC - multiplex immunohistochemistry

MLS - myxoid liposarcoma

moMDSC monocytic myeloid-derived suppressor cell

MPNST - malignant peripheral nerve sheath tumor

Nonclass. Mono. - nonclassical monocyte

OS - osteosarcoma

PC - principal component

PCA - principal component analysis

PD1 - programmed cell death protein 1

PDL1 - programmed cell death ligand 1

QC - quality control

quanTIseq - in-silico immune deconvolution tool

RMS - rhabdomyosarcoma

RNA - ribonucleic acid

RNA-Seq - ribonucleic acid sequencing

RRHO - rank-rank hypergeometric overlap

SARC028 - phase 2 clinical trial investigating pembrolizumab in sarcoma

scRNASeq - single-cell ribonucleic acid sequencing

SRA - Sequence Read Archive

SS - synovial sarcoma

STLMS - non-gynecologic leiomyosarcoma

STS - soft tissue sarcoma

TARGET - Therapeutically Applicable Research to Generate Effective Treatments

TCGA - The Cancer Genome Atlas

Tcm - central memory T cell

Tem - effector memory T cell

Temra - effector memory re-expressing T cell

Th - helper T cell

TIMER - in-silico immune deconvolution tool

TMA - tissue microarray

TME - tumor microenvironment

Treg - regulatory T cell

ULMS - uterine leiomyosarcoma

UPS - undifferentiated pleomorphic sarcoma

USARC - undifferentiated sarcoma

WES - whole exome sequencing

xCell - in-silico immune deconvolution tool

Acknowledgements

I want to sincerely my PhD committee and mentors for their support. Specifically, I want to thank Drs. Jim Economou, Tom Graber, Anusha Kalbasi, and April Pyle for their guidance. Thank you to Dr. Kalbasi for the countless hours he spent teaching and coaching me. His dedication to science and teaching is unparalleled and my training would not be the same without him. Thank you to Dr. Graeber for his guidance and support; his ability to distill matters, both in life and in science, to the most important aspects is incredible. Thank you to Drs. Economou and Pyle for their guidance and wisdom, throughout the PhD process and continuing to this day. I particularly want to thank Dr. Fritz Eilber for his mentorship. None of this would have been possible without his support.

I want to acknowledge and thank the many people for their hard work that made this work possible. Thank you to all the members of the Kalbasi lab, particularly Mito Tariveranmoshabad, Sarah Kremer, Jack Tang and Ben DiPardo, to all the members of the Graeber lab, particularly Katherine Sheu, Nick Bayley, and Henan Zhu, to all the members of the Ribas lab, particularly Katie Campbell and Egmidio Medina, to all the member of the Schaue lab, to Dr. McBride, to the clinical sarcoma research team, particularly Dr. Eilber and Dr. Dry for her help with the construction of the TMA, and finally to the UCLA core facilities, particularly Xinmin Li from the TCGB (Technology Center for Genomics & Bioinformatics) and to Yunfeng Li and Joanne Olivas from the TPCL (Translational Pathology Core Laboratory).

Thank you to Dr. Joe Hines and to Dr. Timothy Donahue for supporting me in all my career goals and for allowing the extra time during my surgical residency to allow me to pursue my research goals. I particularly want to thank Chi Quach and my co-residents for making it all possible. Thank you to the UCLA STAR program and MCIP Department for their support,

particularly Drs. Linda Demer, Mitchell Wong, Tamer Sallam, and Mark Frye. I also want to thank Yesinia Rayos and Nadia Barreda for their constant guidance – I would not have been able to navigate the system without them.

I would like to thank my family and friends for their ongoing support throughout research and residents. Thank you to Arely Landaverde, Silvia Mayorga, and Ligia Hernandez for always being available and willing to take care of our children, for loving them, and for giving me peace of mind always knowing they are happy and safe. You all are the reasons I have been able to pursue my dreams.

Preface

Chapter 2: A version of this manuscript is in preparation for publication: Graham DS, DiPardo B, Ganapathy E, Li M-H, Tariveranmoshabad M, Tang P, Deng J, Kremer SV, Chin S, Crompton JG, Singh A, Li Y, Dry S, Bernthal N, Soragni A, Graeber T, Eilber F, Schaue D, Kalbasi A. Modest and Time-dependent Effects of Radiation on Tumor Immune Composition of Human Sarcoma.

Chapter 3: A version of this manuscript is in preparation for publication: Graham DS, DiPardo B, McCaw T, Tariveranmoshabad M, Tang P, Deng J, Campbell K, Medina E, Kremer SV, Chin S, Crompton JG, Singh A, Li Y, Dry S, Bernthal N, Soragni A, Graeber T, Eilber F, Schaue D, Kalbasi A. Validation of in-silico immune deconvolution methods in undifferentiated sarcoma

Chapter 4: A version of this manuscript is in preparation for publication. Graham DS, DiPardo B, McCaw T, Tariveranmoshabad M, Tang P, Deng J, Campbell K, Medina E, Kremer SV, Chin S, Crompton JG, Singh A, Li Y, Dry S, Bernthal N, Soragni A, Graeber T, Eilber F, Schaue D, Kalbasi A. The transcriptomic and genomic signatures of immune cell infiltration in undifferentiated sarcoma

This work was supported by the T32 Tumor Immunology Training Grant, University of California Los Angeles, T32 CA009120 (PI: Anusha Kalbasi, MD & Thomas Graeber, PhD), the Gerald S. Levey Surgical Research Award, Department of Surgery, University of California Los

Angeles (PI: Fritz Eilber, MD), and the Alan B. Slifka Foundation (ABSF) Grant, Alan B. Slifka Foundation (PI: Owen Witte, MD)

Biographical Sketch

EDUCATION AND TRAINING

University of Pennsylvania, Philadelphia, Pennsylvania
Bachelor of Arts in Biological Basis of Behavior and in Economics, December 2010

University of Pennsylvania, Philadelphia, Pennsylvania
Masters of Business Administration, May 2016

University of Pennsylvania, Philadelphia, Pennsylvania
Doctor of Medicine, May 2016

SELECT GRADUATE AWARDS AND HONORS

Bendheim Family Fellowship Award "for outstanding academic accomplishments, integrity, and leadership," The Wharton School, University of Pennsylvania, Philadelphia, PA (2014-2016)

Medical Student Teaching Award. David Geffen School of Medicine. University of California Los Angeles, Los Angeles, CA (2019, 2020, 2021)

Jonathan R. Hiatt Excellence in Medical Student Education Award. University of California Los Angeles, Los Angeles, CA (2020)

RESEARCH GRANTS

Gerald S. Levey Surgical Research Award, Department of Surgery, University of California Los Angeles, Los Angeles, CA. Project: "Personalized sarcoma therapy: Developing patient-derived orthotopic xenografts (PDOX) from core needle biopsies." Amount: $110,000. Role: Investigator. PI: Fritz Eilber, MD. 2018-2020.

T32 Tumor Immunology Training Grant, University of California Los Angeles, Los Angeles, CA. "Characterizing the Molecular and Genetic Signatures of Immune Evasion in Undifferentiated Pleomorphic Sarcoma." Amount: $57,456. Role: Investigator. PI: Thomas Graeber, PhD & Anusha Kalbasi, MD. 2020-2021.

Alan B. Slifka Foundation (ABSF) Grant, Alan B. Slifka Foundation, New York, NY. "Delineating the origin and genomics of sarcomas to reveal therapeutic targets." Amount: $10,000. PI: Owen Witte, MD. 2021.

SELECT PUBLICATIONS

Russell TA, Eckardt MA, Murakami T, Elliott IA, Kawaguchi K, Kiyuna T, Igarashi K, Li Y, Crompton JG, **Graham DS**, Dry SM, Bernthal N, Yanagawa J, Kalbasi A, Federman N, Chmielowski B, Singh AS, Hoffman RM, Eilber FC. "Clinical Factors That Affect the Establishment of Soft Tissue Sarcoma Patient-Derived Orthotopic Xenografts: A University of

California, Los Angeles, Sarcoma Program Prospective Clinical Trial." *JCO Precision Oncology.* June 2017.

Herbert G, Karakousis G, Bartlett E, Zaheer S, **Graham D**, Czerniecki B, Fraker D, Ariyan C, Aoit D, Brady S. Transected thin melanoma: implications for sentinel lymph node staging. *Journal of Surgical Oncology.* March 2018.

Shurrell-Linehan E, DiPardo BJ, Elliott IA, **Graham DS**, Eckardt MA, Dry SM, Nelson SD, Singh AS, Kalbasi A, Federman N, Bernthal NM, Eilber FC. "Pathologic Response to Neoadjuvant Therapy is Associated with Improved Long-term Survival in High-risk Primary Localized Malignant Peripheral Nerve Sheath Tumors." *American Journal of Clinical Oncology.* May 2019.

Graham DS, Kadera BE, Eilber FC. "Irregular Lipomatous Extremity Tumor." *JAMA.* May 2019.

Graham DS, Russell TA, Eckardt MA, Motamedi K, Seeger LL, Singh AS, Bernthal NM, Kalbasi A, Dry SM, Nelson SD, Elashoff D, Levine BD, Eilber FC. "Oncologic Accuracy of Image-Guided Percutaneous Core-Needle Biopsy of Peripheral Nerve Sheath Tumors at a High-Volume Sarcoma Center." *American Journal of Clinical Oncology*. October 2019.

Graham DS, DiNome ML, Ganz PA. "Breast Cancer Risk-Reducing Medications." *JAMA.* July 2020.

Graham DS, Onyshchenko M, Eckardt MA, DiPardo BJ, Venigalla S, Nelson SD, Chmielowski B, Singh AS, Shabason JE, Eilber FC, Kalbasi A. "Low rates of chemotherapy utilization for primary high-grade soft tissue sarcoma: A National Cancer Database Analysis." *JNCCN.* August 2020.

Graham DS, van Dams R, Jackson NJ, Onyschenko M, Eckardt MA, DiPardo BJ, Nelson SD, Chmielowski B, Shabason JE, Singh AS, Eilber FC, Kalbasi A. "Chemotherapy and Survival in Patients with Primary High-Grade Extremity and Trunk Soft Tissue Sarcoma." *Cancers.* August 2020.

Graham DS, Qorbani A, Eckardt MA, Klingbeil KD, Chen LY, Chopra S, Eilber FC, Dry SM. Does "Low-Grade" Dedifferentiated Liposarcoma Exist? A Single Institution 25-Year Retroperitoneal Liposarcoma Experience. *In submission.*

Eckardt M, **Graham D**, Klingbeil K, Lofftus S, McCaw T, Bailey M, Goldring C, Kendal J, Kadera B, Nelson S, Dry S, Kalbasi A, Singh A, Chmielowski B, Eilber F, Eilber F. "Lifelong imaging surveillance is indicated for patients with primary retroperitoneal liposarcoma." *In submission.*

Crane JN*, **Graham DS***, Mona CE, Nelson SD, Samiei A, Dawson DW, Dry SM, Masri MG, Crompton JG, Benz MR, Czernin J, Eilber FC, Graeber TG, Calais J, Federman NC. "Fibroblast activation protein expression in sarcomas." *In submission.*

CHAPTER 1

Introduction

1.1 Overview of sarcoma

Sarcoma is a rare, diverse, and aggressive group of malignancies, with over 60 subtypes, presumed to be of mesenchymal origin[1–4]. The two primary categorizations of sarcoma are osteosarcoma and soft tissue sarcoma (STS) and, together, they comprise only one percent of adult cancers and 15-20 percent of pediatric cancers[5]. Adult STS, and more specifically undifferentiated sarcoma (USARC), will be the primary focus of this study.

Undifferentiated sarcoma is one of the most common subtypes of STS. This rare tumor is difficult to characterize and is often a diagnosis of exclusion, specifically, it is a group of sarcomas that can't otherwise be classified. There is no classical immunohistochemistry (IHC) staining pattern. Rather, they are defined as "undifferentiated" because they do not fit another sarcoma profile[6,7]. Further, there are multiple different subtypes of undifferentiated sarcoma, including pleomorphic, epithelioid, and spindled, or any combination of these[7]. The nuances in diagnosis coupled with the rarity of this tumor pose significant challenges when studying USARC. As such, the underlying tumor biology of USARC remains poorly understood. Yet, there appears to be tumor behavior that links these tumors together.

1.2 Clinical considerations in the treatment of soft tissue sarcoma

Most patients with STS are treated with surgery with or without radiation and/or chemotherapy. Despite aggressive local therapy, median survival is approximately five years for patients with localized disease and approximately two to three years for patients with metastatic disease[8–15]. The efficacy of chemotherapy and radiation in the treatment of STS remains unclear. As such, there is wide variation in treatment patterns throughout the country. Overall in the United States, approximately 20% of patients with STS are treated with chemotherapy and/or radiation, and there is no clear association between the use of systemic therapy and survival in STS overall[16,17]. Undifferentiated sarcoma, in particular, has no widely successful systemic therapies. The prognosis for patients with advanced or metastatic disease is notably poor, with a median survival of less than one year[3]. There is a critical need for more effective systemic therapies for the treatment of STS.

1.3 The role of immunotherapy in the treatment of soft tissue sarcoma

Immune-based therapies have dramatically changed the treatment of many cancers in recent decades. These include blockade of immune checkpoint molecules such as PD-1/PD-L1 and CTLA-4, cellular therapies using chimeric antigen receptor or T-cell receptor modified T-cells, and emerging therapies modulating tumor-associated macrophages. While early studies of immunotherapy in STS have not had broad success, there are signals of efficacy that warrant further exploration[18–21]. The SARC028 study evaluated the efficacy of the anti-PD1 antibody pembrolizumab in patients with advanced sarcoma, and identified the undifferentiated pleomorphic sarcoma (UPS) subtype as the most responsive to anti-PD1 immunotherapy, with 40% of patients demonstrating complete or partial response[20]. This was notable given that UPS,

unlike most tumors that are most responsive to immune checkpoint blockade, have a low mutational burden.

Why a fraction of patients with UPS and other tumors with low mutational burden instigate an anti-tumor immune response is unclear. At the same time, 60% of patients with UPS in this study did not respond to anti-PD1 therapy, suggesting that within UPS there are drivers of immune evasion. For instance, UPS can arise from mesenchymal stem cells, which are characterized by immunosuppressive features that may persist in the malignant state. Furthermore, chromosomal instability is a hallmark of UPS and other STS, which has been associated with immune evasion in other tumor types.

In order to stratify the patients with STS that may benefit from immunotherapy, we need alternate strategies to study sarcoma in a high throughput fashion. For STS, alternatives to categorizing patients based on "response" versus "non-response" to immunotherapy are needed, as this disease is extremely rare and large studies of immune checkpoint blockade are not readily available as in other disease[18–27]. In order to guide further studies and to develop optimal immunotherapy strategies for patients with STS, we must first understand the landscape of immune cell infiltration and the associated genomic changes associated with immune evasion in this disease.

1.4 The association between chromosomal instability and immune cell infiltration

A comprehensive study of sarcoma using the TCGA demonstrated the landscape of chromosomal instability, specifically referring to copy number changes, in soft tissue sarcoma. STS is, overall, characterized by a paucity of mutations. Some STS subtypes, such as UPS, MFS, DDLPS, and MPNST are characterized by particularly high numbers of copy number mutations. In fact, even compared to many other types of cancer, these subtypes of sarcoma are uniquely characterized by low mutations and high copy number changes[6].

Prior studies have demonstrated an association between copy number changes and immune evasion, specifically demonstrating that high copy number change is associated with low immune cell infiltration in multiple tumor types. There are multiple hypotheses explaining this association, including immunoediting, loss of antigen presenting machinery, silencing of neoantigen promotors, and the activation of cGAS-STING pathway[28–35]. Yet, despite the examination of this association in other tumor types, the association between copy number changes and immune cell infiltration has not been described in STS and warrants further study.

1.5 Dissertation overview

In this study, I aim to investigate the immunologic heterogeneity and identify transcriptomic and genomic correlates of immune evasion in undifferentiated sarcoma. In doing so, I will determine the immune cell landscape, the optimal high-throughput tools, and the transcriptomic and genomic changes associated with high and low immune cell infiltration in STS. These insights will provide needed information to understand which patients may benefit from immunotherapy and guide future studies to further the treatment of this disease.

Specific Aim 1: Characterize the landscape of immune cell infiltration in soft tissue sarcoma. My goal is to determine the overall immune cell landscape and the association between clinical and treatment factors and immune cell infiltration in STS. (Project 1, Chapter 2)

Specific Aim 2: Develop and apply an optimal in-silico immune deconvolution technique for sarcoma. My goal is to assess the concordance between various immune deconvolution techniques and compare their ability to recapitulate the results of immunohistochemistry on paired USARC specimens. (Project 2, Chapter 3)

Specific Aim 3: Define transcriptomic signatures and genomic changes associated with immune cell infiltration in undifferentiated sarcoma. I hypothesize that tumors with low levels of immune cell infiltration are characterized by high copy number changes. (Project 3, Chapter 4)

A summary of the experimental design addressing these specific aims is shown in Figure 1-1. The details of the experimental methods used in each project are described in detail in the corresponding chapters.

1.6 Figures

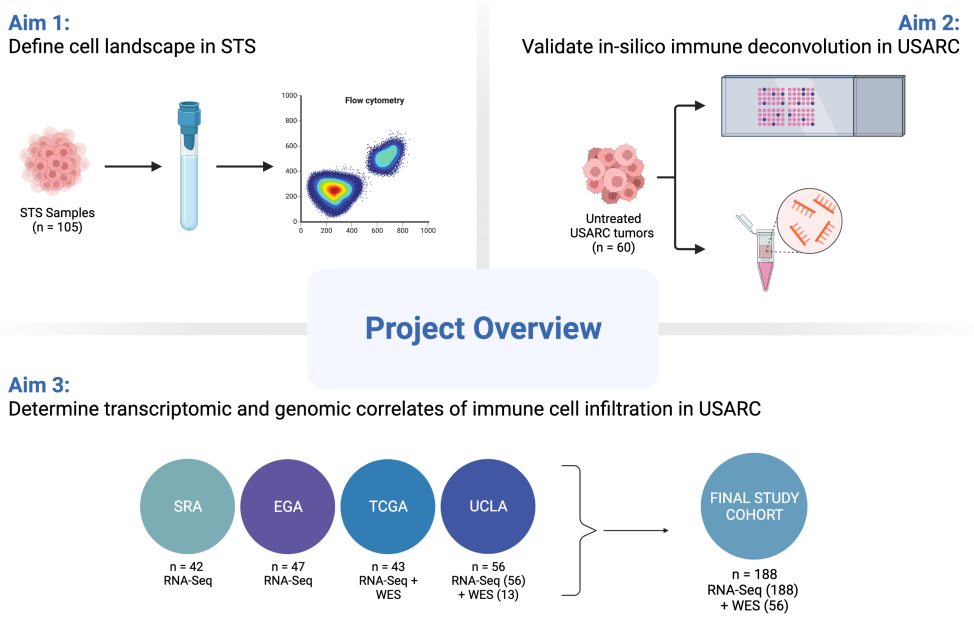


Figure 1-1: Overview of experimental design. Aim 1 characterizes the landscape of immune cell infiltration in soft tissue sarcoma using flow cytometry data generated from 105 fresh tumor samples. Aim 2 determines the optimal in-silico immune deconvolution technique for sarcoma by comparing in-silico and mIF results from 60 untreated undifferentiated sarcoma samples. Aim 3 defines transcriptomic signatures and genomic changes associated with immune cell infiltration in undifferentiated sarcoma using RNA-Seq and WES or WGS data from four datasets, ultimately including 188 unique undifferentiated sarcoma samples.

1.7 Bibliography

1. Cancer Facts & Figures 2017. Accessed May 5, 2021. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2017.html
2. Sarcomas, Soft Tissue - Statistics. Cancer.Net. Published June 25, 2012. Accessed May 5, 2021. https://www.cancer.net/cancer-types/sarcomas-soft-tissue/statistics
3. Ferguson PC, Deheshi BM, Chung P, et al. Soft tissue sarcoma presenting with metastatic disease: outcome with primary surgical resection. *Cancer*. 2011;117(2):372-379. doi:10.1002/cncr.25418
4. Taylor BS, Barretina J, Maki RG, Antonescu CR, Singer S, Ladanyi M. Advances in sarcoma genomics and new therapeutic targets. *Nat Rev Cancer*. 2011;11(8):541-557. doi:10.1038/nrc3087
5. Soft Tissue Sarcoma. Accessed September 27, 2022. https://www.cancer.org/cancer/soft-tissue-sarcoma.html
6. Abeshouse A, Adebamowo C, Adebamowo SN, et al. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. *Cell*. 2017;171(4):950-965.e28. doi:10.1016/j.cell.2017.10.014
7. Steele CD, Tarabichi M, Oukrif D, et al. Undifferentiated Sarcomas Develop through Distinct Evolutionary Pathways. *Cancer Cell*. 2019;35(3):441-456.e8. doi:10.1016/j.ccell.2019.02.002

8. Brennan MF, Antonescu CR, Moraco N, Singer S. Lessons learned from the study of 10,000 patients with soft tissue sarcoma. *Ann Surg*. 2014;260(3):416-421; discussion 421-422. doi:10.1097/SLA.0000000000000869
9. Trovik CS, Bauer HC, Alvegård TA, et al. Surgical margins, local recurrence and metastasis in soft tissue sarcomas: 559 surgically-treated patients from the Scandinavian Sarcoma Group Register. *Eur J Cancer Oxf Engl 1990*. 2000;36(6):710-716. doi:10.1016/s0959-8049(99)00287-7
10. Rööser B, Gustafson P, Rydholm A. Is there no influence of local control on the rate of metastases in high-grade soft tissue sarcoma? *Cancer*. 1990;65(8):1727-1729. doi:10.1002/1097-0142(19900415)65:8<1727::aid-cncr2820650811>3.0.co;2-x
11. Gustafson P. Soft tissue sarcoma. Epidemiology and prognosis in 508 patients. *Acta Orthop Scand Suppl*. 1994;259:1-31.
12. Gustafson P, Rööser B, Rydholm A. Is local recurrence of minor importance for metastases in soft tissue sarcoma? *Cancer*. 1991;67(8):2083-2086. doi:10.1002/1097-0142(19910415)67:8<2083::aid-cncr2820670813>3.0.co;2-5
13. Alamanda VK, Crosby SN, Archer KR, Song Y, Schwartz HS, Holt GE. Predictors and clinical significance of local recurrence in extremity soft tissue sarcoma. *Acta Oncol Stockh Swed*. 2013;52(4):793-802. doi:10.3109/0284186X.2012.711953
14. Weitz J, Antonescu CR, Brennan MF. Localized extremity soft tissue sarcoma: improved knowledge with unchanged survival over time. *J Clin Oncol Off J Am Soc Clin Oncol*. 2003;21(14):2719-2725. doi:10.1200/JCO.2003.02.026
15. Zaidi MY, Ethun CG, Tran TB, et al. Assessing the Role of Neoadjuvant Chemotherapy in Primary High-Risk Truncal/Extremity Soft Tissue Sarcomas: An Analysis of the Multi-

institutional U.S. Sarcoma Collaborative. *Ann Surg Oncol*. 2019;26(11):3542-3549. doi:10.1245/s10434-019-07639-7

16. Graham DS, Onyshchenko M, Eckardt MA, et al. Low Rates of Chemotherapy Use for Primary, High-Grade Soft Tissue Sarcoma: A National Cancer Database Analysis. *J Natl Compr Cancer Netw JNCCN*. 2020;18(8):1055-1065. doi:10.6004/jnccn.2020.7553
17. Graham DS, van Dams R, Jackson NJ, et al. Chemotherapy and Survival in Patients with Primary High-Grade Extremity and Trunk Soft Tissue Sarcoma. *Cancers*. 2020;12(9):E2389. doi:10.3390/cancers12092389
18. Pollack SM, Ingham M, Spraker MB, Schwartz GK. Emerging Targeted and Immune-Based Therapies in Sarcoma. *J Clin Oncol Off J Am Soc Clin Oncol*. 2018;36(2):125-135. doi:10.1200/JCO.2017.75.1610
19. Patnaik A, Kang SP, Rasco D, et al. Phase I Study of Pembrolizumab (MK-3475; Anti-PD-1 Monoclonal Antibody) in Patients with Advanced Solid Tumors. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2015;21(19):4286-4293. doi:10.1158/1078-0432.CCR-14-2607
20. Tawbi HA, Burgess M, Bolejack V, et al. Pembrolizumab in advanced soft-tissue sarcoma and bone sarcoma (SARC028): a multicentre, two-cohort, single-arm, open-label, phase 2 trial. *Lancet Oncol*. 2017;18(11):1493-1501. doi:10.1016/S1470-2045(17)30624-1
21. Ben-Ami E, Barysauskas CM, Solomon S, et al. Immunotherapy with single agent nivolumab for advanced leiomyosarcoma of the uterus: Results of a phase 2 study. *Cancer*. 2017;123(17):3285-3290. doi:10.1002/cncr.30738
22. Gandhi L, Rodríguez-Abreu D, Gadgeel S, et al. Pembrolizumab plus Chemotherapy in Metastatic Non–Small-Cell Lung Cancer. *N Engl J Med*. 2018;378(22):2078-2092. doi:10.1056/NEJMoa1801005

23. Mok TSK, Wu YL, Kudaba I, et al. Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial. *The Lancet*. 2019;393(10183):1819-1830. doi:10.1016/S0140-6736(18)32409-7
24. Cramer-van der Welle CM, Verschueren MV, Tonn M, et al. Real-world outcomes versus clinical trial results of immunotherapy in stage IV non-small cell lung cancer (NSCLC) in the Netherlands. *Sci Rep*. 2021;11(1):6306. doi:10.1038/s41598-021-85696-3
25. Middleton G, Brock K, Savage J, et al. Pembrolizumab in patients with non-small-cell lung cancer of performance status 2 (PePS2): a single arm, phase 2 trial. *Lancet Respir Med*. 2020;8(9):895-904. doi:10.1016/S2213-2600(20)30033-3
26. Robert C, Ribas A, Schachter J, et al. Pembrolizumab versus ipilimumab in advanced melanoma (KEYNOTE-006): post-hoc 5-year results from an open-label, multicentre, randomised, controlled, phase 3 study. *Lancet Oncol*. 2019;20(9):1239-1251. doi:10.1016/S1470-2045(19)30388-2
27. Pembrolizumab versus Ipilimumab in Advanced Melanoma | NEJM. Accessed May 23, 2021. https://www.nejm.org/doi/full/10.1056/nejmoa1503093
28. Rosenthal R, Cadieux EL, Salgado R, et al. Neoantigen-directed immune escape in lung cancer evolution. *Nature*. 2019;567(7749):479-485. doi:10.1038/s41586-019-1032-7
29. Anagnostou V, Smith KN, Forde PM, et al. Evolution of Neoantigen Landscape during Immune Checkpoint Blockade in Non-Small Cell Lung Cancer. *Cancer Discov*. 2017;7(3):264-276. doi:10.1158/2159-8290.CD-16-0828

30. Davoli T, Uno H, Wooten EC, Elledge SJ. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science*. 2017;355(6322). doi:10.1126/science.aaf8399

31. Hatch EM, Fischer AH, Deerinck TJ, Hetzer MW. Catastrophic nuclear envelope collapse in cancer cell micronuclei. *Cell*. 2013;154(1):47-60. doi:10.1016/j.cell.2013.06.007

32. Woo SR, Fuertes MB, Corrales L, et al. STING-dependent cytosolic DNA sensing mediates innate immune recognition of immunogenic tumors. *Immunity*. 2014;41(5):830-842. doi:10.1016/j.immuni.2014.10.017

33. Wu T, Ji Y, Moseman EA, et al. The TCF1-Bcl6 axis counteracts type I interferon to repress exhaustion and maintain T cell stemness. *Sci Immunol*. 2016;1(6):eaai8593. doi:10.1126/sciimmunol.aai8593

34. Santaguida S, Richardson A, Iyer DR, et al. Chromosome Mis-segregation Generates Cell-Cycle-Arrested Cells with Complex Karyotypes that Are Eliminated by the Immune System. *Dev Cell*. 2017;41(6):638-651.e5. doi:10.1016/j.devcel.2017.05.022

35. Senovilla L, Vitale I, Martins I, et al. An immunosurveillance mechanism controls cancer cell ploidy. *Science*. 2012;337(6102):1678-1684. doi:10.1126/science.1224922

CHAPTER 2

Local and systemic immune survey of surgically resected soft tissue sarcoma

2.1 Abstract

Immune-based therapies have revolutionized cancer therapy. However, early studies immunotherapy in soft tissue sarcoma (STS) have not had broad success. Yet, there are signals of efficacy of immunotherapy in STS that warrant further exploration[1–3]. A comprehensive understanding of immune cell infiltration in STS will likely provide insights into the diverse responses to immune checkpoint blockade. Yet, the immune cell landscape in STS remains poorly understood. In this study, we aim to characterize the landscape of immune cell infiltration in soft tissue sarcoma and to determine the association between clinical and treatment factors and immune cell infiltration in soft tissue sarcoma. We conducted multiparametric flow cytometry of tumor and peripheral blood specimens from patients undergoing surgical resection for STS at our institution and compared these variables across various patient, tumor, and treatment factors. We found that USARC tumors are characterized by a myeloid predominance and a relative abundance of suppressor cells, such as Treg cells and CD11b cells. We found similar trends when comparing the subtypes of STS that characteristically have high numbers of copy number alterations (CNA) (DDLPS, USARC, and MFS), compared to those that typically have low numbers of CNA (MLS, LMS, SS, and GIST)[4]. We additionally found that the immune composition of peripheral blood was association with intratumoral leukocyte infiltration, and specifically that myeloid-predominant tumor and lymphocyte-predominant blood are mutually exclusive. Finally, we observed that the effects of radiotherapy on the tumor microenvironment

evolve over time and ultimately yield a CD8 predominant phenotype. These findings warrant further study, though they may provide the basis for new treatment paradigms in STS.

2.2 Introduction

Soft tissue sarcoma (STS) is a rare and diverse group of malignancies with over 60 subtypes, presumed to be of mesenchymal origin. Patients with localized STS are managed with surgery with or without radiation[5–7]. Despite aggressive local therapy, up to 50% of patients with high-risk (≥ 5cm, high-grade) primary STS develop metastases, highlighting the critical need for more effective systemic therapies[8–15].

Immune-based therapies have dramatically changed the treatment of many cancers in recent decades. These include blockade of immune checkpoint molecules such as PD-1/PD-L1 and CTLA-4, cellular therapies using chimeric antigen receptor or T-cell receptor modified T-cells, and emerging therapies modulating tumor-associated macrophages. While early studies of immunotherapy in STS have not had broad success, there are signals of efficacy that warrant further exploration[1–3,16].

Immune cell infiltration has been shown to be associated with response to immunotherapy in various cancers. Infiltration by T cells and a cytolytic T cell signature have been shown to be biomarkers for response to immune checkpoint blockade in cancers such as melanoma[17,18]. Petitprez et al. demonstrated that B cell infiltration is associated with improved survival and immunotherapy response in sarcoma[19]. A comprehensive understanding of immune cell

infiltration in soft tissue sarcoma (STS) will likely provide insights into the diverse responses to immune checkpoint blockade. Yet, the immune cell landscape in STS remains poorly understood.

In this study, we aim to characterize the landscape of immune cell infiltration in soft tissue sarcoma and to determine the association between clinical and treatment factors and immune cell infiltration in soft tissue sarcoma.

2.3 Materials and Methods

2.3.1 Human Tissue Sample Collection

Patients with planned surgical resection or core needle biopsy for soft tissue sarcoma at our institution between April 10, 2017 through November 25, 2020 were identified and consented to an IRB approved blood and tissue collection protocol. A subset of patients was consented to a separate IRB approved prospective clinical trial, as a part of which blood and tissue were collected. At the time of surgery or core needle biopsy, tumor and blood specimens were obtained from enrolled patients. Fresh tumor samples were placed into RPMI on ice and blood was collected in tubes containing EDTA and transported for processing and analysis.

2.3.2 Clinicopathologic Data and Study Population

Patient and tumor clinical-pathologic data, including age, sex, tumor pathology, tumor size, and treatment duration and type, was collected for each enrolled patient. Tumors with final pathology demonstrating no residual tumor, final pathology demonstrating low-grade disease, and histologies with $n \leq 2$ were excluded from the final analysis.

2.3.3 Tumor and blood processing and antibody staining

Tumor samples were minced and up to 1 gram of tumor was placed in a GentleMACS C tube with RPMI, Collagenase D (1 mg/mL), and DNase I (0.3mg/mL). Tumors were dissociated using a GentleMACS Octo Dissociator for 1h at 37C. Cells were washed and filtered through a 100 micron strainer to obtain a single cell suspension, which were then treated with red blood cell lysis buffer (BD RBC Lysing Buffer). Cells were stained with viability stain, and then divided into two samples to be stained with each antibody panel (Table 2-1, myeloid panel; Table 2-2; lymphocyte panel). Cells were stained for 30 mins on ice, washed, and analyzed on a BD LSR Fortessa flow cytometer. Fresh blood samples were directly incubated with staining antibodies for 30 mins on ice (no viability stain), followed by incubation with red blood cell lysis buffer for 10 mins at room temperature. Cells were washed and analyzed on a BD LSR Fortessa flow cytometer. The gating strategy for the lymphoid and myeloid cells is shown in Figure 2-7. The following lists contain the following information: antigen, fluorochrome (company, catalog#) or reagent (company, catalog#).

<u>Lymphoid Panel</u>

Live/dead (tumor only), FVS 510 (BD, 564406)

CD3, BV711 (BD, 563725)

CD8, Alex700 (BD, 557945)

CD4, FITC (BD, 555346)

PD-1, BV421 (Biolegend, 329920)

CD45RA, APC (Biolegend, 304112)

CCR7, BV605 (Biolegend, 353224)

CD25, PE (BD, 555432)

CD127, PE-Cy7 (BD, 560822)

CXCR3, PE-CF594 (BD, 562451)

CCR6, BV650 (BD, 563922)

PD-L1, PerCP-Cy5.5 (Biolegend, 329738)

CD45, APC-H7 (BD, 650178)

Myeloid Panel

Live/dead, FVS 510 (BD, 564406)

CD3, BV711 (BD, 563725)

HLA-DR, FITC (BD, 555811)

CD11b, PerCP-Cy5.5 (Biolegend, 301328)

CD14, PE (BD, 562691)

CD16, BV650 (BD, 563692)

CD56, PE-CF594 (BD, 562289)

CD19, APC-H7 (BD, 641395)

CD15, APC (BD, 551376)

CD11c, Alex-700 (BD, 561352)

PD-L1, BV421 (BD, 563738)

CD45, BV605 (BD, 564047)

CD68 (tumor only), PE-Cy7 (BD, 565595)

CD163 (tumor only), BV786 (Biolegend, 333632)

CD66b (blood only), PE-Cy7 (Biolegend, 305116)

CD33 (blood only), BV786 (Biolegend, 303428)

Additional Reagents:

Collagenase (Sigma, 11088866001)

DNaseI (Roche, 10104159001)

2.3.4 Multiplex Immunofluorescence

Formalin-fixed paraffin embedded tissue sections were generated from tumors from a subset of patients. The TSA-based Opal method was used for immunofluorescence (IF) staining (Opal Polaris 7-Color Automation IHC Kit; Akoya Biosciences, Marlborough, MA, USA; Catalogue No. NEL871001KT). Because TSA and DAB oxidation are both peroxidase-mediated reactions, the primary antibody conditions and order of staining determined using DAB detection were directly applied to the fluorescent assays. Unlike conventional IHC in which a chromogenic peroxidase substrate is used for antigen detection, each antibody is paired with an individual Opal fluorophore for visualization. The Opal fluorophores were used at a 1 in 150 dilution, as recommended by Akoya when using the Leica BOND RX. As such, a fluorescent singleplex was

performed for each biomarker and compared to the appropriate chromogenic singleplex to assess staining performance.

Once each target was optimized using uniplex slides, the Opal 6 multiplexed assay was used to generate multiple staining slides. We applied primary antibodies to normal human tonsil specimens as controls at optimized concentrations previously determined on the uniplex control tissues. TMA slides Staining was performed consecutively Leica BOND RX by using the same steps as those used in uniplex IF, and the detection for each marker was completed before application of the next antibody.

All fluorescently labelled slides were scanned on the Vectra Polaris (Akoya Biosciences) at 40× magnification using appropriate exposure times. The data from the multispectral camera were analyzed by the imaging InForm software (Akoya Biosciences).

Regions of tumor tissue confirmed by corresponding H&E slide were selected using Phenochart 1.0.12 (Akoya Bioscience), exported as multi-layer TIFF images using inForm 2.4.10 (Akoya Bioscience), and stitched in HALO v3 (Indica Labs) for quantitative image analysis. Cell inclusion and segmentation criteria were optimized using the "real-time tuning" feature to define the nuclear contrast threshold, minimum nuclear intensity, nuclear segmentation aggressiveness, and maximum cytoplasmic radius. For each marker, specific positivity thresholds were defined based on a human tonsil positive control. The entire image was analyzed using these parameters and the generated data included the percentage of cells positive for each marker, expressed as a

percentage of all nucleated cells. The list below contains the following information regarding the materials used: antibody (company, catalog number), clone/lot (when available), opal dye.

Antibodies:

CD4 (Dako, M7310), Opal 570

CD8 (Dako, M7103), C8/144B, Opal 620

CD163 (Cell Marque, 163m-16), Opal 480

CD68 (Dako, m0876), PG-M1, Opal 520

CD45 (Dako, M0701), 2B11+PD7/26, Opal 690

2.3.5 Statistical Analysis

Clinical and flow data comparisons were performed using Wilcoxon rank sum test or chi squared test, as appropriate. Correlations between continuous variables were determined using a linear fit model, with Pearson correlation coefficients shown. The correlation plots shown depict Spearman correlation coefficients using complete observations. Heatmaps were scaled across the rows. Clustering results, when shown, were calculated using Euclidean distances. In the flow data heatmaps, missing values were set to the row median such that they appear white on the graph. Missing data per sample and per variable was analyzed and summarized in Figure 2-8. Linear regression models were used to determine the association between histology group (copy number high versus low) and select intratumoral leukocyte populations. The leukocyte populations that were investigated using multivariate linear regression models were those that were statistically significant in the univariate models. The covariates selected in the models were

the clinical variable that were statistically significant in the univariate analysis and those that are standard in the literature. Throughout the study, multiple hypothesis testing was performed using Bonferroni correction and is noted where applicable.

2.3.6 Software

The majority of statistical analyses and data visualizations were performed and generated in R (Version 4.0.5).

2.3.7 Approval

This study was approved by the UCLA Institutional Review Board (IRB #10-001857).

2.4 Results

2.4.1 Characteristics of study cohort

Of 125 patients from which specimens and data were collected, we excluded patients from whom no residual tumor was observed on pathology (n=2), patients with low-grade tumors (n=9) and patients with poorly represented (n≤2) sarcoma subtypes (n=9; Figure 2-1A). The characteristics of the final study cohort (n=105) are summarized in Table 2-1. The majority of patients underwent resection of primary tumors (n=62), though locally recurrent (n=17) and metastatic (n=26) tumors were also represented. Except for one lung metastasis, tumors were located in

either the extremity/trunk (n=72), or retroperitoneum (including abdomen/pelvis, n=32). Undifferentiated sarcoma (USARC, n=45), was the predominant histologic subtype, followed by myxoid liposarcoma (MLS, n=16), dedifferentiated liposarcoma (DDLPS, n=14), leiomyosarcoma (LMS, n=13), myxofibrosarcoma (MFS, n=8), synovial sarcoma (SS, n=5) and GIST (n=4) (Figure 2-1B).

Thirty-one patients were treatment-naïve, and the remainder received preoperative radiation (n=35), systemic therapy (n=21), or a combination of systemic therapy and radiation (n=18). Systemic therapy included cytotoxic chemotherapy, tyrosine kinase inhibitors, and immunotherapy. Radiation therapy included conventionally fractionated radiation (2 Gy per fraction, 50 Gy total) or hypofractionated radiation (6 Gy per fraction, 30 Gy total) (Figure 2-1B).

Aside from histologic grouping, sarcoma subtypes were grouped according to putative biological subtypes. Although most STS subtypes have low mutational burden, they can be categorized according to high and low prevalence of copy number alterations (CNA), with DDLPS, USARC, and MFS categorized as CNA-high (n=67) and MLS, LMS, SS, and GIST as CNA-low (n=38)[4]. Alternatively, soft tissue sarcomas can be categorized by the presence or absence of simple genomic driver events (e.g. fusion event or mutation), as observed in synovial sarcoma (SYT-SSX fusion), myxoid liposarcoma (FUS-DDIT3 fusion) and GIST (KIT or PDGFRA mutants) (Figure 2-1B)[20].

2.4.2 Validation of multiparametric flow cytometry data from tumor samples

To assess the internal validity of the flow cytometry data from digested tumor samples, we examined the association between pathologic treatment effect and the proportion of live cells as assessed by flow cytometry. There was a significant negative correlation between treatment effect scores and live cell content of single cells from tumor digests, whether assessed as a continuous or categorical variable. The effect was most prominent at the extremes, in which single cells from samples with pathologic treatment effect ≥90% had substantially lower viability by flow cytometry than those from samples with pathologic treatment effect ≤10% (Figure 2-1C). Incidentally, we also observed a significant correlation between tumor size and treatment effect, which we hypothesize may be related to the increased necrosis in large tumors (Figure 2-1D).

To evaluate the validity of the immunologic content as assessed by flow cytometry, we compared immune content assessed by multiplex immunohistochemistry (mIHC) in a subset of 24 samples. We did not observe a significant correlation between CD45+ cells as a percentage of total cells between mIHC and flow cytometry ($R = 0.2$, $p = 0.38$). This as expected, given that identification of live CD45- non-immune cells by flow cytometry can be challenging. This is also consistent with the presence of samples with high CD45 content by flow cytometry that was not captured by mIHC, where there were no samples with high CD45 content by mIHC and low CD45 content by flow cytometry. We did observe a statistically significant correlation between CD8 T cell content as a percentage of CD45+ cells between flow cytometry and mIHC results ($R = 0.56$, $p = 0.008$). Similarly, we observed a correlation between myeloid cell content as assessed by flow cytometry and CD163+ cells as a percentage of CD45+ cells by multiplex IHC, though

this did not reach statistical significance (R = 0.48, p = 0.061). And while we did not find a statistically significant correlation between CD4 T cell content across mIHC and flow cytometry datasets (R = 0.27, p = 0.26), in both datasets the proportion of CD4 T cells among CD45+ cells was <20% (Figure 2-1E).

2.4.3 Overview of intratumoral leukocyte populations in soft tissue sarcoma

Myeloid cells are the predominant immune cell in soft tissue sarcoma, comprising an average of 52.9% of the leukocyte population in the samples in our study. CD8 T cells and CD4 T cells comprise 19.4% and 16.2% of the immune cell infiltrates, respectively. There is a paucity of B cells, 1.9% (Figure 2-2A). Of the myeloid cells, the plurality of the cells are macrophages (21.7%), granulocytic myeloid derived suppressor cells (gMDSCs; 21.4%), intermediate monocytes (19.9%), and classical monocytes (15.3%) (Figure 2-2B). Of the CD4 and CD8 T cell populations, the majority of the cells are CD4 T effector memory (61.0%) and CD8 T effector memory (54.9%) cells, respectively. The minority of the CD4 and CD8 T cell populations are CD4 naïve (5.0%) and CD8 naïve (6.3%) cells, respectively (Figure 2-2C & 2D).

Figure 2-2E summarizes the correlation between the lymphocyte and myeloid populations of the tumors. In general, the myeloid and lymphoid populations, measured as a percentage of total CD45 positive cells, are inversely correlated, as expected. We further examined the correlation between lymphocytes (as a percent of total CD45) and the subpopulations of myeloid cells (as a percent of myeloid cells) to further understand the associations between various cell types. The results are summarized the correlation plot in Figure 2-2F. This correlation plot highlights that

lymphocytes (as a percent of CD45) are in fact positively correlated with myeloid dendritic cells (as a percent of myeloid cells) ($R = 0.38$, $p < 0.001$). When examining the correlation between myeloid cells and lymphoid cell subpopulations in a similar manner, we found that NK cells (as a percent of lymphocytes) and CD8 Temra cells (as a percent of CD8 T cells) were positively correlated with myeloid cells (as a percent of CD45), $R = 0.27$ ($p = 0.012$) and $R = 0.23$ ($p = 0.024$), respectively (Figure 2-2G). Tables 2-4 & 2-5 summarize the correlation values depicted in Figure 2-2F & 2-2G, respectively. Figure 2-10 depicts an expanded correlation analysis, including all tumor and blood samples.

2.4.4 CNA-high tumors have a more prominent immune infiltrate characterized by higher myeloid cells

Although most STS subtypes have low mutational burden, they can be categorized according to high and low prevalence of copy number alterations (CNA). DDLPS, USARC, and MFS categorized as CNA-high (n=67) and MLS, LMS, SS, and GIST as CNA-low (n=38)[4]. The CNA-high tumors have more intratumoral leukocytes overall, with higher CD45 cells as a percent of live cells (Figure 2-3A). On further examination, this difference appears to be driven by a higher proportion of myeloid and suppressor cells (Figure 2-3B-F). Specifically, there is a higher myeloid to lymphoid ratio, more CD11b+ cells, and more T regulatory cells in the CNA-high tumor group when compared to the CNA-low tumor group (Figure 2-3B & 2-3D). However, there are fewer CD4+ T cells and NK cells in the CNA-high tumor group when compared to the CNA-low tumor group (Figure 2-1C). Further analysis of the association between each subtype of sarcoma is depicted in Figure 2-11.

Multivariate linear regression analyses were performed to determine association between various intratumoral leukocyte population and CNA histology groups (as defined above). Cell types (dependent variables) included in this analysis included CD11b (% of CD45), myeloid cells (% of CD45), CD8 T cells (% of CD45), CD8PD1 cells (% of CD8), CD4 (% of CD45), CD4PD1 (% of CD4), and Treg (% of CD4). Independent variable included in the model were CNA histology group, XRT, chemotherapy, and lesion type (primary, recurrence, metastasis). The results of these models are summarized in Table 2-2 and show that CNA-high tumors are associated with increased CD11b (% of CD45), myeloid cells (% of CD45), CD4 (% of CD45), and Treg (% of CD4) populations when compared with the CAN-low histology group tumors. Sensitivity and subgroup analyses of these multivariate models are shown in Tables 2-6 through 2-8.

2.4.5 The lymphocyte content of tumors is positively correlated with the proportion of lymphocyte among peripheral blood leukocytes, and myeloid-rich tumor and lymphocyte-rich blood are mutually exclusive

Figure 2-4 depicts the associations between intratumoral and peripheral immune states in patients with soft tissue sarcoma. The proportion of lymphocytes among peripheral blood leukocytes is positively correlated with the lymphocyte content of tumors ($R = 0.25$, $p = 0.037$; Figure 2-4A) and negatively correlated with the myeloid content of tumors ($R = -0.25$, $p = 0.035$; Figure 2-4B). The associations between tumor lymphocytes and tumor myeloid cells and peripheral granulocytes and monocytes are not statistically significant.

Heatmaps of flow cytometry results from blood samples (n= 72; Figure 2-4C) and tumor samples (n = 100; Figure 2-4D) from patients with STS highlight three patterns of immune composition: lymphocyte-predominant, myeloid-predominant, and other. The alluvial plot in Figure 2-4E depicts the associations between these groups in the paired blood and tumor samples in this study. Lymphocyte-predominant peripheral samples were associated with either lymphocyte predominant or unclassified paired tumor samples. Lymphocyte-predominant peripheral samples were never paired with a myeloid predominant tumor sample. Myeloid-predominant and unclassified peripheral blood samples were paired with lymphocyte-predominant, myeloid-predominant, and unclassified tumor samples. In summary, myeloid-predominant tumor and lymphocyte-predominant blood were mutually exclusive.

2.4.6 Neoadjuvant therapy results in subtle effects on tumor-infiltrating leukocytes

Figure 2-5 summarizes the association between neoadjuvant therapy and intratumoral leukocytes in soft tissue sarcoma. In each comparison, no neoadjuvant therapy is used as the reference group. Overall, neoadjuvant therapy results in subtle effects on tumor-infiltrating leukocytes. Notably, there was no significant association between CD45+ content, CD11b+ myeloid content, or myeloid/lymphoid ratio and neoadjuvant therapy (Figure 2-5A & 2-5B). There was a significant decrease in B cells and an increase in CD8 T cell content associated with the combination of systemic therapy and radiation when compared with no neoadjuvant therapy (Figure 2-5C). However, the composition of the myeloid cells, CD8 T cells and CD4 T cells remained largely the same, except for a decrease in the proportion of naive CD4 T cells that was

associated with the combination of systemic therapy and radiation (Figure 2-5D-F). Patient who underwent immunotherapy treatment were excluded from the analysis in Figure 2-5D-F. Additional analyses exploring the association between treatment status and intratumoral leukocyte populations are depicted in Figure 2-18 through 2-21.

2.4.7 Effects of radiotherapy on the tumor microenvironment evolve over time, and ultimately yield a CD8 predominant phenotype

Overall, there were only subtle differences in the intratumoral leukocyte populations associated with treatment. However, on further examination of the association between radiotherapy and intratumoral leukocytes, we found that the tumor microenvironment evolved over time. We examined the association between XRT-surgery interval (days between radiotherapy and surgery or tumor sample acquisition) and the intratumoral leukocyte populations in a subset of 51 patients. There were 53 patients who underwent neoadjuvant radiotherapy, and of these patients, 2 outliers were excluded as they underwent XRT more than 19 weeks prior to surgery (Figure 2-9). The results of this analysis are shown in Figure 2-6A-D. Notably, there was a positive association between CD8 T cells and XRT-surgery interval ($R = 0.48$, $p < 0.001$) and between CD8/CD4 T cell ratio and XRT-surgery interval ($R = 0.48$, $p < 0.001$). We then validated these findings from our flow cytometry data in a subset of tumors using IHC. These results are summarized in Figure 2-6E & 2-6F and show that there is an increase in CD8 T cell infiltration associated with increased XRT-surgery interval.

2.5 Discussion

Although soft tissue sarcoma (STS) encompasses a broad set of clinically and biologically diverse malignancies, most subtypes lack the typical features associated with response to immune checkpoint blockade therapy. That is, most STS subtypes have a low mutational burden, are mismatch repair proficient, lack UV- or carcinogen-induced mutational signatures, and are not associated with viral infections[4,21,22]. However, the absence of these features does not preclude response to immune checkpoint blockade (ICB), especially among specific histologic subtypes of sarcoma such as undifferentiated pleomorphic sarcoma (UPS)[3]. Early evidence indicates that STS subtypes responding to ICB have higher levels of tumor-infiltrating immune cells at baseline, consistent with observations in patients with other malignancies[4,19,23–26]. However, the clinical and biological determinants of tumor immune infiltration in STS are not well understood.

In this study, we characterize the landscape of immune cell infiltration in soft tissue sarcoma and to determine the association between clinical and treatment factors and immune cell infiltration in soft tissue sarcoma. We conducted multiparametric flow cytometry of tumor and peripheral blood specimens from patients undergoing surgical resection for STS at our institution. Most existing studies of immune composition of STS, a rare tumor, have utilized large publicly available datasets like the TCGA, from which tumor immune profiles are estimated using immune deconvolution of bulk transcriptomic data. While informative, immune deconvolution is limited in its ability to define precise cellular contributions within the tumor microenvironment, especially for low frequency cell types. Furthermore, immune deconvolution methods have not been trained or validated in the context of sarcoma, in which tumor cells are known to share gene

expression profiles with myeloid cells. Lastly, analyses restricted to the tumor do not capture the effects of a tumor on the systemic immune state of a patient.

In this study, we summarize the intratumoral and peripheral immune profiles of tumors from our institutional cohort patient with STS. Our findings indicate that USARC tumors are characterized by a myeloid predominance and a relative abundance of suppressor cells, such as Treg cells and CD11b cells. A similar trend was seen when comparing all tumor subtypes that are classified as having high levels of copy number alterations overall (DDLPS, USARC, and MFS; n=67) versus those that are characterized as having low levels of copy number alterations (MLS, LMS, SS, and GIST as CNA-low (n=38)[4]. These findings may provide the basis for insights into improving systemic treatment strategies in sarcoma. Macrophage depletion and macrophage reprogramming are two proposed mechanisms for overcoming immunotherapy resistance[27–31]. It is likely that myeloid-targeted strategies such as these would be more relevant for tumors that are more myeloid-rich. It is possible that these strategies should be considered in STS tumors and this topic warrants further study.

We further explored the association between intratumoral and peripheral immune profiles of tumors with clinical features and preoperative exposures, including RT. We found that standard preoperative cytotoxic therapy is only modestly associated with the STS immune composition, but this association appears to be time-dependent. Specifically, we found that an increase in the time from XRT to surgery was associated with an increase in the CD8 composition of intratumoral leukocytes. These findings suggest a dynamic association between XRT and the tumor immune microenvironment, and perhaps suggest that waiting longer between XRT and

surgery should be considered to derive the optimal immunomodulatory effects. Though, further studies are warranted.

Limitations of this study include the small sample size overall and the imbalance in various histologies. Further, there were very small sample sizes of each of the comparison groups of the various treatment factors. This study was an exploratory and descriptive analysis, and further studies validating and confirming the findings in this study are needed.

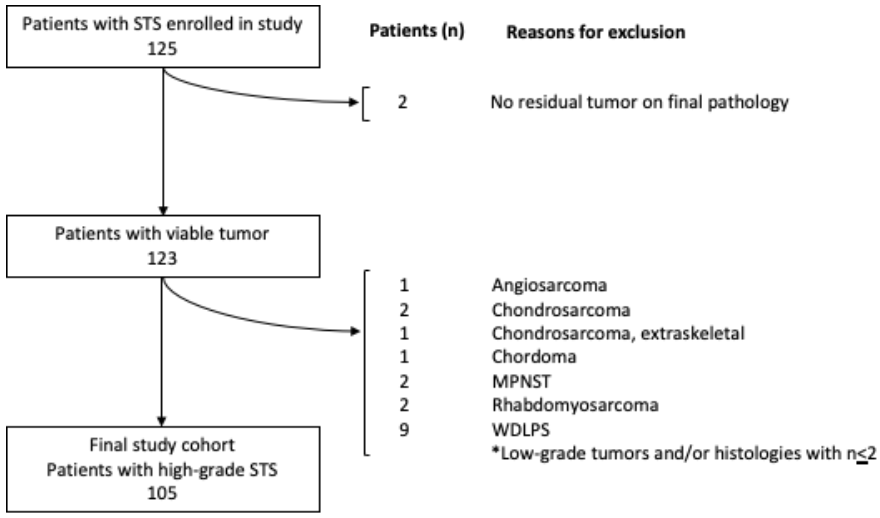
2.6 Conclusion

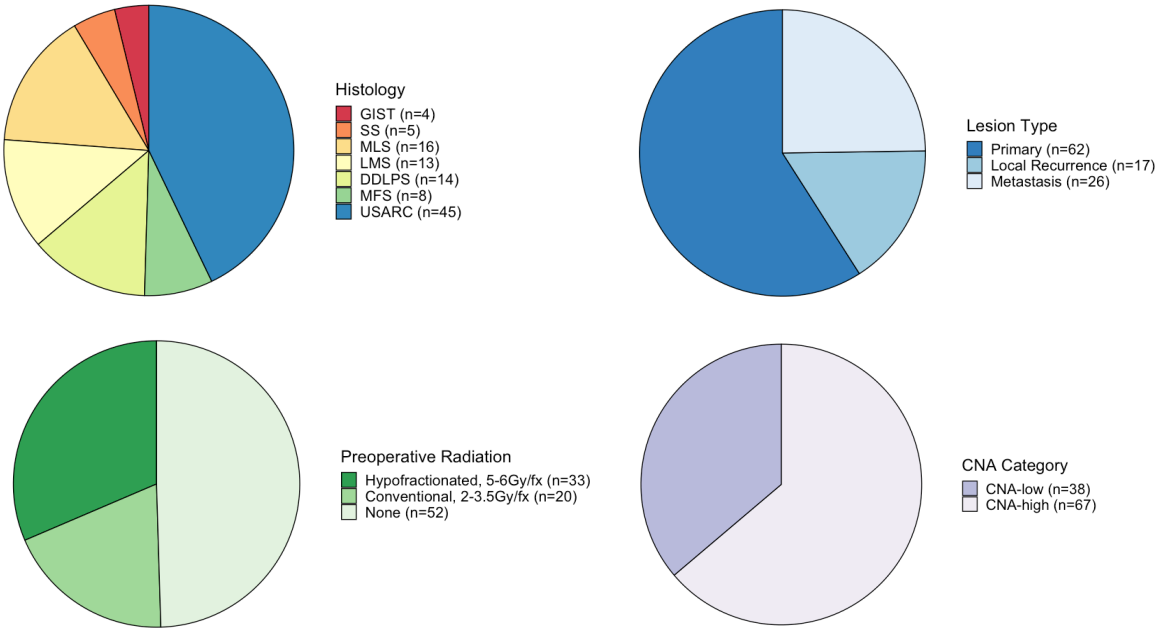
In this study, we characterized the landscape of immune cell infiltration in soft tissue sarcoma and determined the association between clinical and treatment factors and immune cell infiltration in soft tissue sarcoma. We found that USARC tumors are characterized by a myeloid predominance and a relative abundance of suppressor cells, such as Treg cells and CD11b cells. We found similar trends when comparing STS subtypes with high levels of copy number alterations overall (DDLPS, USARC, and MFS; n=67) versus those that are characterized as having low levels of copy number alterations (MLS, LMS, SS, and GIST) as CNA-low (n=38)[4]. We additionally found that the immune composition of peripheral blood was associated with intratumoral leukocyte infiltration, and specifically that myeloid-predominant tumor and lymphocyte-predominant blood are mutually exclusive. Finally, we observed that the effects of radiotherapy on the tumor microenvironment evolve over time and ultimately yield a CD8 predominant phenotype. These findings warrant further study, though they may provide insights into new treatment paradigms for STS. First, it is possible that macrophage depletion and

macrophage reprogramming may be promising in the treatment of these myeloid-rich tumors. It may also be possible to identify which tumors are particularly myeloid-rich by examining peripheral blood. Finally, waiting longer between XRT and surgery may derive the optimal immunomodulatory effects.

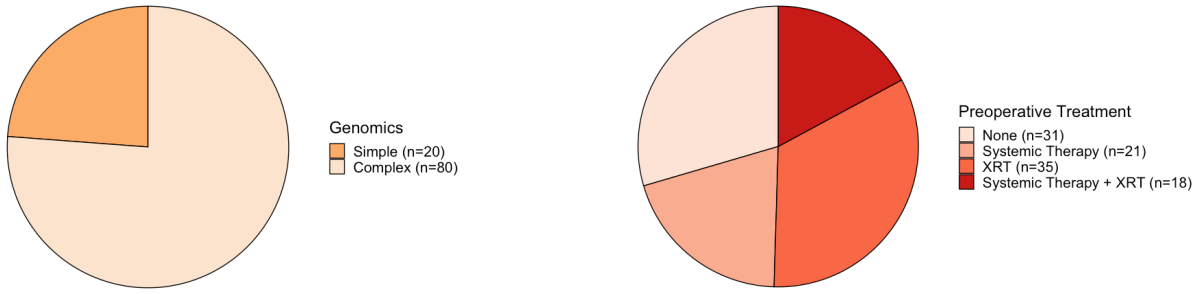
2.7 Figures

A

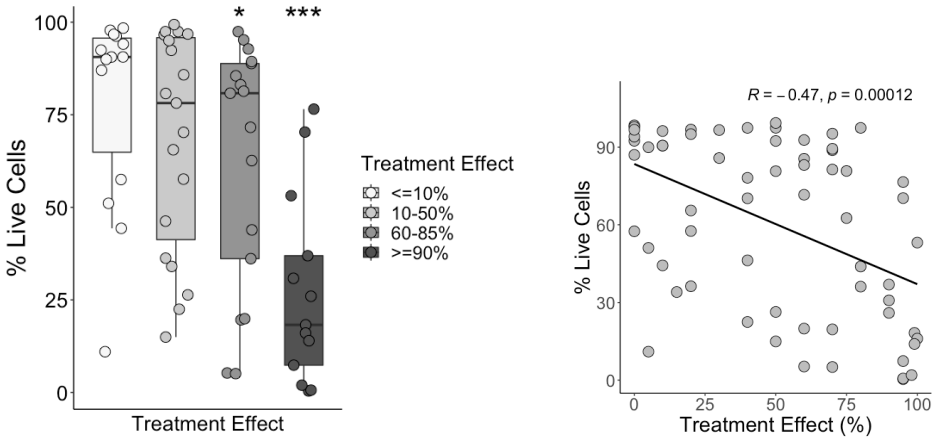




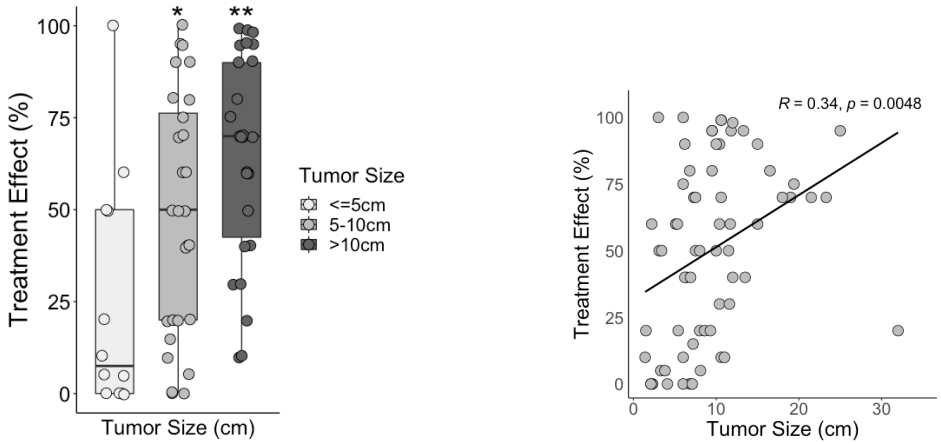
B

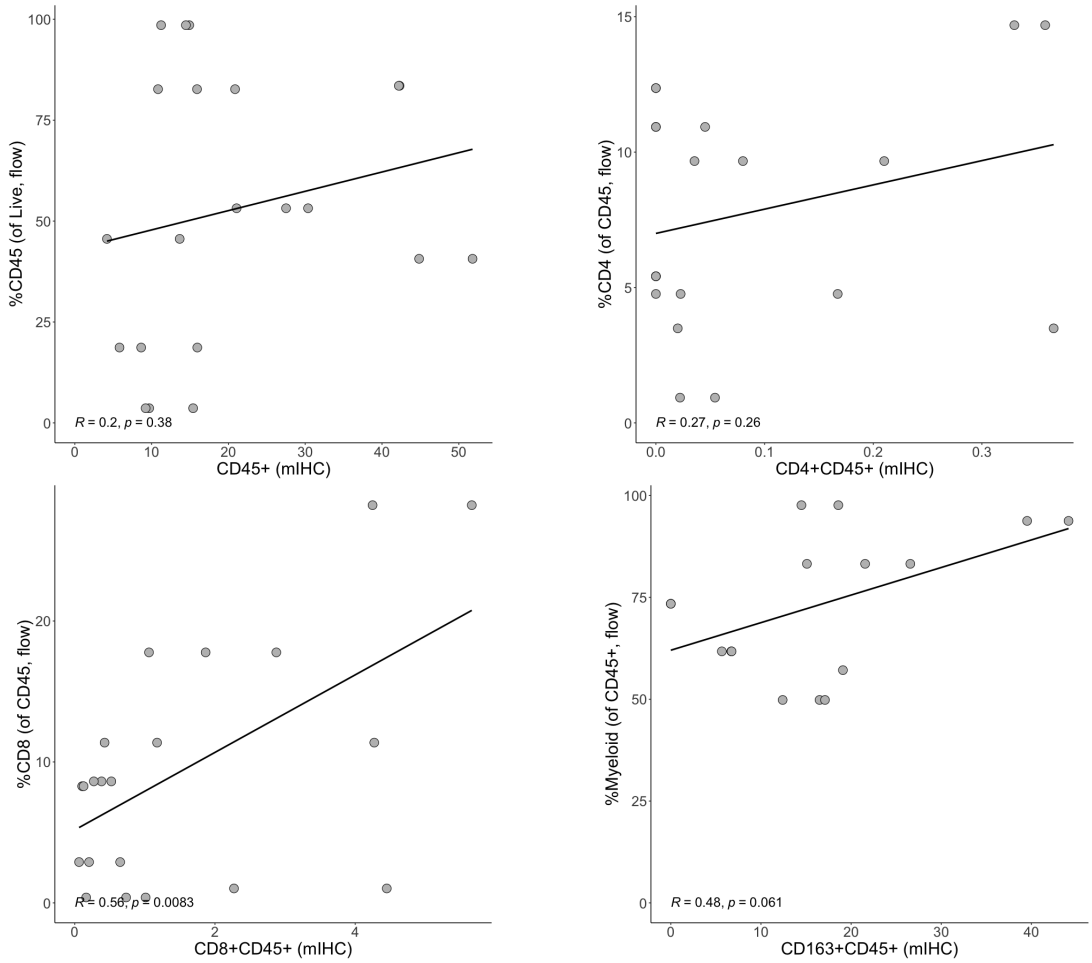

Genomics
Simple (n=20)
Complex (n=80)
Preoperative Treatment
None (n=31)
Systemic Therapy (n=21)
XRT (n=35)
Systemic Therapy + XRT (n=18)

C


% Live Cells
Treatment Effect
*
***
Treatment Effect
<=10%
10-50%
60-85%
>=90%
R = −0.47, p = 0.00012
% Live Cells
Treatment Effect (%)

D

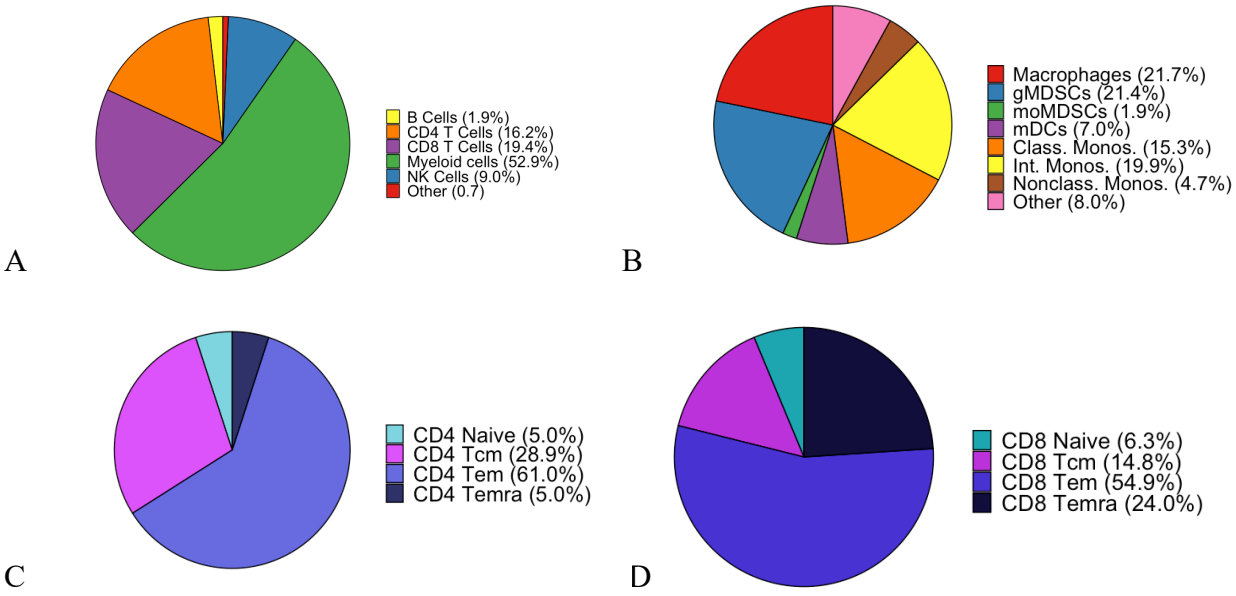

Treatment Effect (%)
Tumor Size (cm)
*
**
Tumor Size
<=5cm
5-10cm
>10cm
R = 0.34, p = 0.0048
Treatment Effect (%)
Tumor Size (cm)

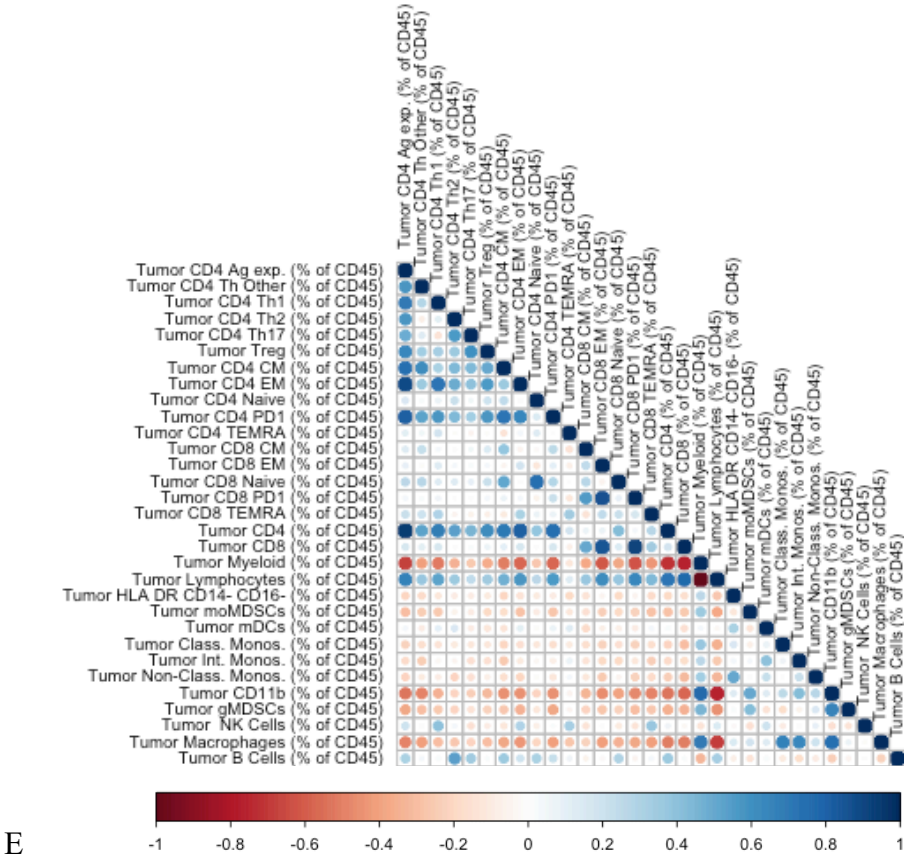


E

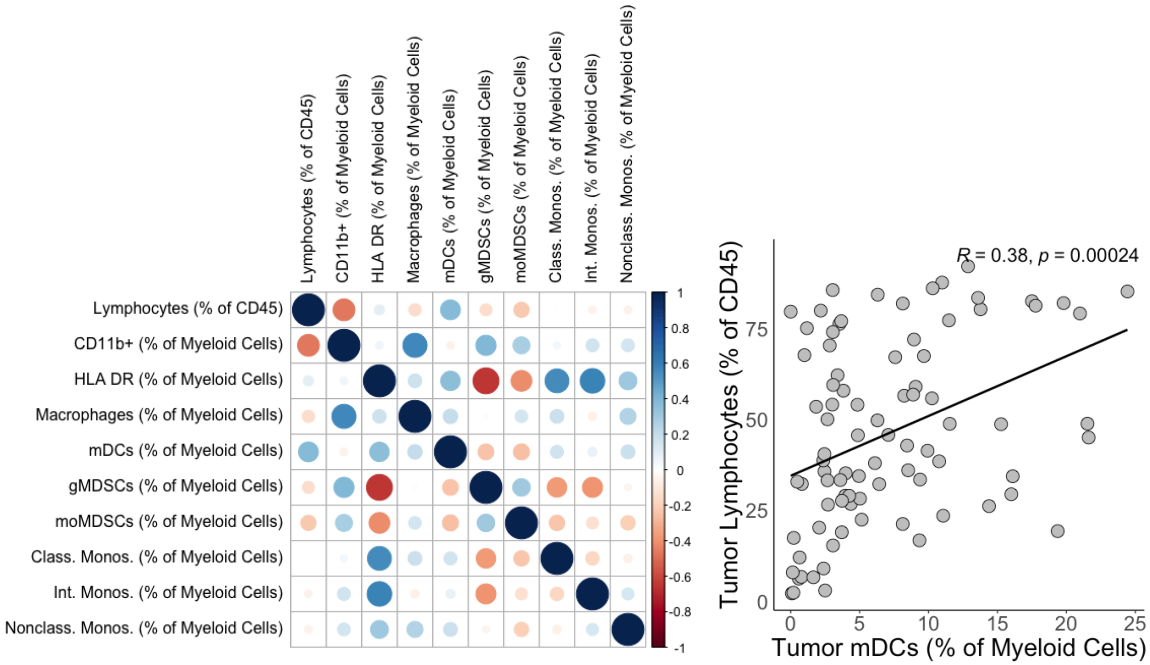
Figure 2-1: Clinical characteristics of study cohort of patients with soft tissue sarcoma. (A) Consolidated Standards of Reporting Trials (CONSORT) diagram showing definition of study cohort from a single institution. (B) The distribution of select characteristics, including tumor histologies, lesion type (primary versus recurrence versus metastasis), preoperative radiation fractionation, CNA histology group, genomic histology group, and neoadjuvant therapy type. (C) The association between live cells (% of total) and treatment effect (%). (D) The association between treatment effect (%) and tumor size (cm). (E) The association between flow cytometry and mIHC results for cell types as labeled. Significance testing for categorical data was performed using Wilcoxon test with treatment effect ≤10% or tumor size ≤5cm as the reference

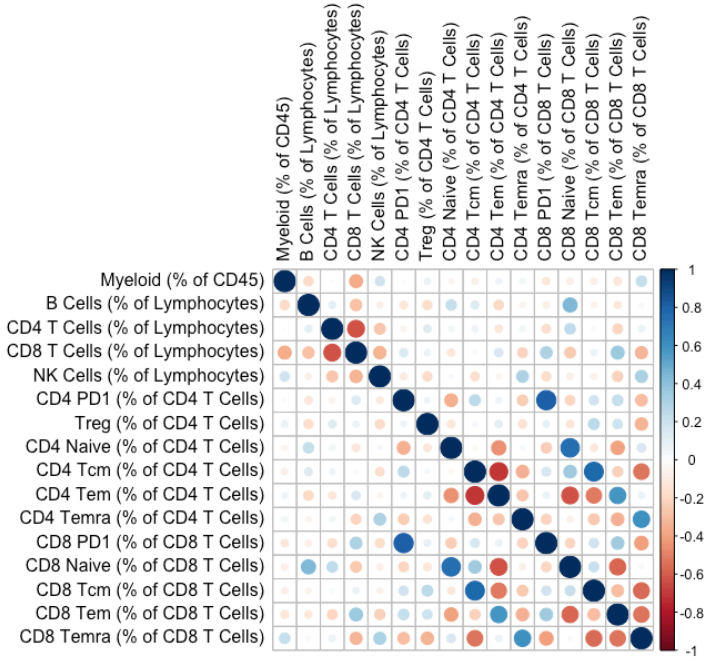
group. Regression lines for continuous data demonstrate linear models fit to data. *, P < 0.05; **, P < 0.01; ***, P < 0.001 ; ****, P < 0.0001.





F


Lymphocytes (% of CD45)
CD11b+ (% of Myeloid Cells)
HLA DR (% of Myeloid Cells)
Macrophages (% of Myeloid Cells)
mDCs (% of Myeloid Cells)
gMDSCs (% of Myeloid Cells)
moMDSCs (% of Myeloid Cells)
Class. Monos. (% of Myeloid Cells)
Int. Monos. (% of Myeloid Cells)
Nonclass. Monos. (% of Myeloid Cells)
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
R = 0.38, p = 0.00024
Tumor Lymphocytes (% of CD45)
Tumor mDCs (% of Myeloid Cells)
0
25
50
75
0
5
10
15
20
25


Myeloid (% of CD45)
B Cells (% of Lymphocytes)
CD4 T Cells (% of Lymphocytes)
CD8 T Cells (% of Lymphocytes)
NK Cells (% of Lymphocytes)
CD4 PD1 (% of CD4 T Cells)
Treg (% of CD4 T Cells)
CD4 Naive (% of CD4 T Cells)
CD4 Tcm (% of CD4 T Cells)
CD4 Tem (% of CD4 T Cells)
CD4 Temra (% of CD4 T Cells)
CD8 PD1 (% of CD8 T Cells)
CD8 Naive (% of CD8 T Cells)
CD8 Tcm (% of CD8 T Cells)
CD8 Tem (% of CD8 T Cells)
CD8 Temra (% of CD8 T Cells)
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

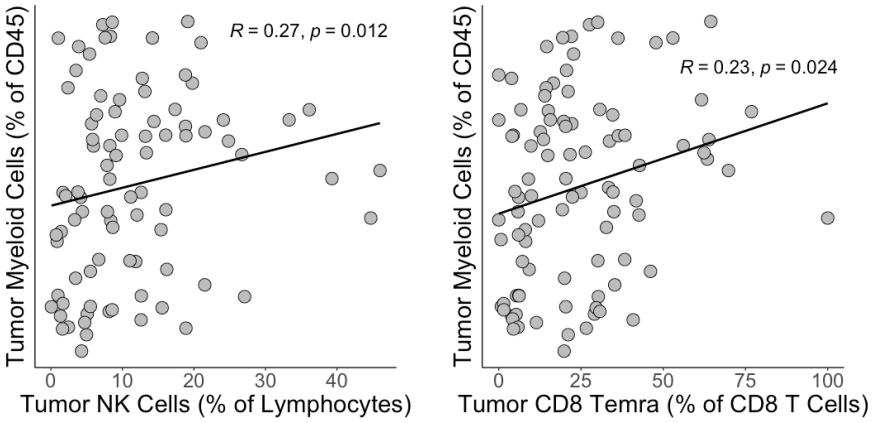
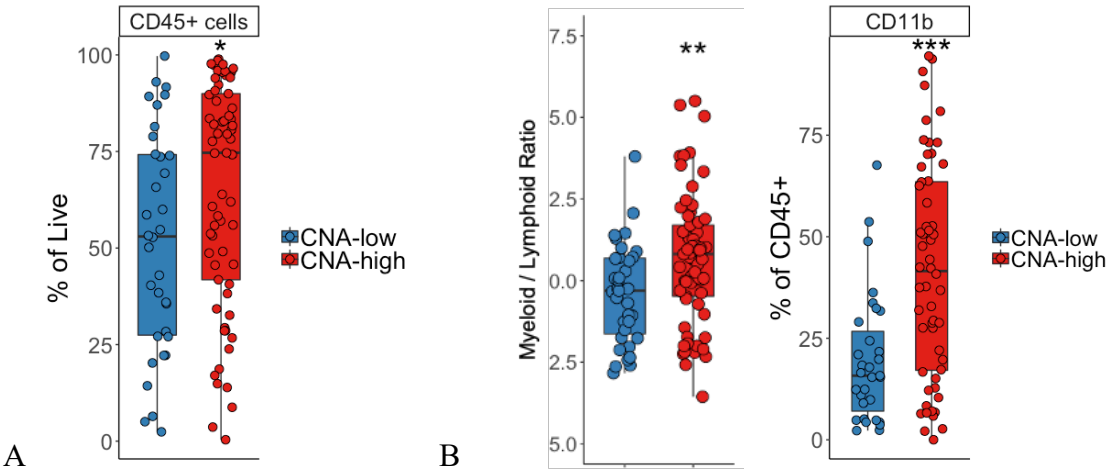
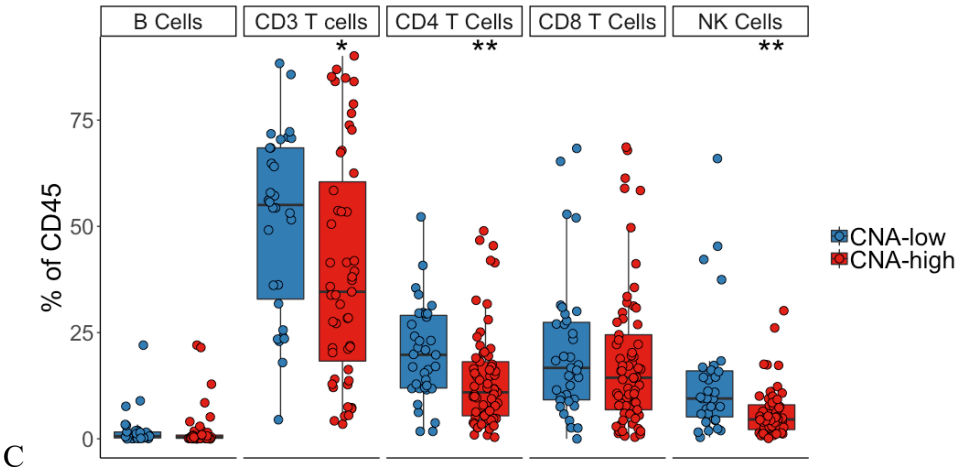
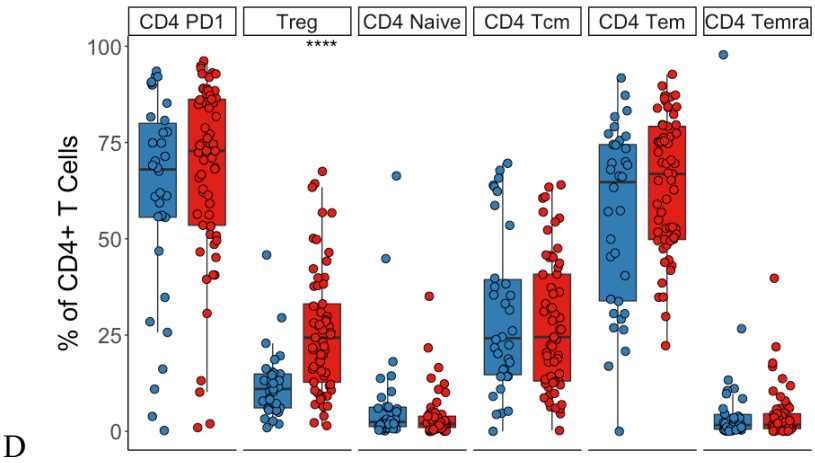
G 

Figure 2-2: Overview of the immune composition of soft tissue sarcoma. Soft tissue sarcomas demonstrate a myeloid predominance. (A-D) The relative amounts of select subsets of immune cells are summarized – lymphoid cells (A), myeloid cells (B), CD4+ T cells (C), and CD8+ T cells (D) in STS tumor samples. (E) Correlation plot depicting the correlation between intratumoral leukocyte populations in STS tumor samples. (F) Correlation plot depicting the association between intratumoral lymphocytes (% of CD45) and subpopulations of myeloid cells. These results highlight the positive association between lymphocytes (% of CD45) and mDCs (% of myeloid cells). (G) Correlation plot depicting the association between intratumoral myeloid cells (% of CD45) and subpopulations of lymphocytes. These results highlight the positive association between myeloid cells (% of CD45) and NK cells (% of lymphocytes) and between myeloid cells (% of CD45) and CD8 Temra (% of CD8 T cells). Spearman correlation coefficients are depicted in the correlation plots. Regression lines for continuous data demonstrate linear models fit to data. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.


A
CD45+ cells
% of Live
CNA-low
CNA-high
B
Myeloid / Lymphoid Ratio
CD11b
% of CD45+
CNA-low
CNA-high


C
B Cells
CD3 T cells
CD4 T Cells
CD8 T Cells
NK Cells
% of CD45
CNA-low
CNA-high


D
CD4 PD1
Treg
CD4 Naive
CD4 Tcm
CD4 Tem
CD4 Temra
% of CD4+ T Cells

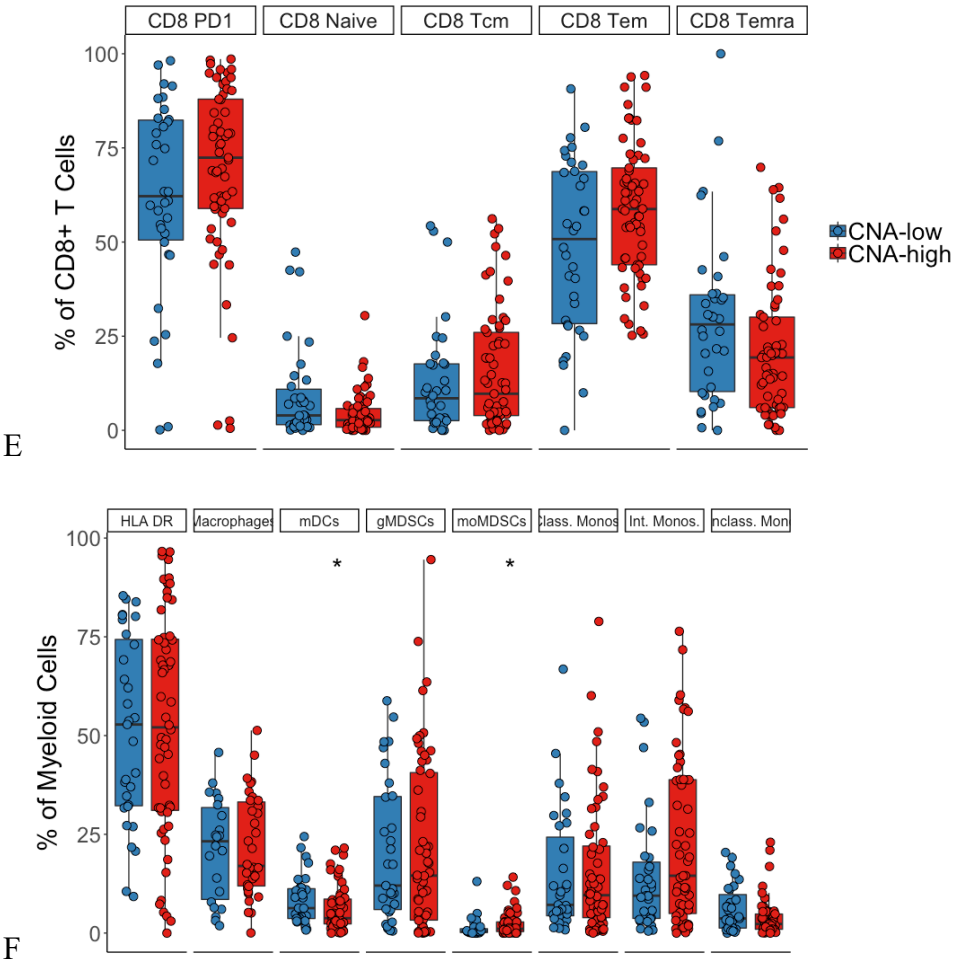
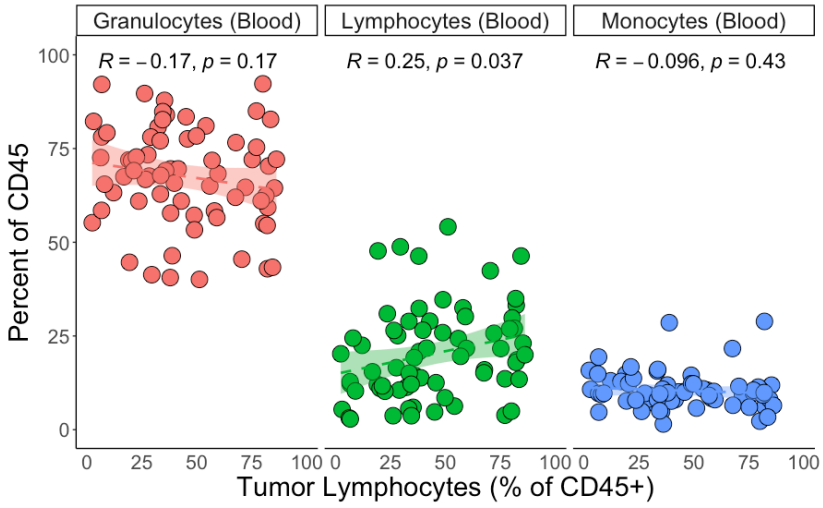


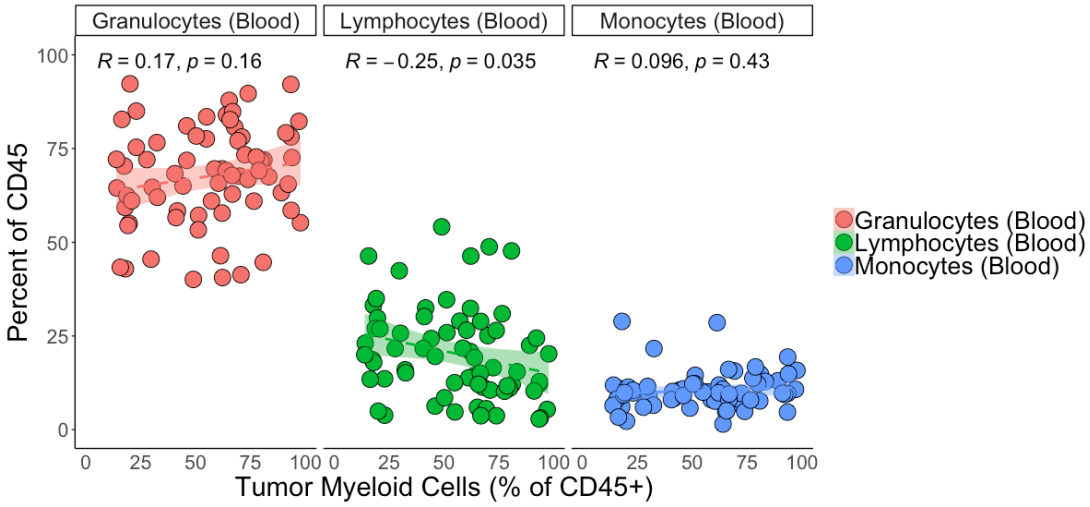
Figure 2-3: Summary of immune cell infiltrates across CNA histology groups of STS tumor samples. CNA-high histology group tumors have a more prominent immune infiltrate, which is characterized by higher myeloid cells and fewer CD4 T and NK cells when compared to CNA-low histology group tumors. Among tumor-infiltrating CD4 T cells, Tregs are more prominent in CNA-high tumors. (A) The proportion of CD45 (% of live) cells and (B) the myeloid/lymphoid ratio and CD11b quantities across CNA histology groups in STS tumor samples. Values shown are log2(myeloid/lymphoid ratio). The relative abundances of various cell populations as a % of lymphocytes (C), % of CD4+ T cells (D), % of CD8+ T cells (E), % of myeloid cells (F) across CNA histology groups. Significance testing was performed using Wilcoxon test with CNA-low

histology group as the reference group. Using Bonferroni correction, statistical significance is defined as <0.002. *, P < 0.05; **, P < 0.01; ***, P < 0.001 ; ****, P < 0.0001.

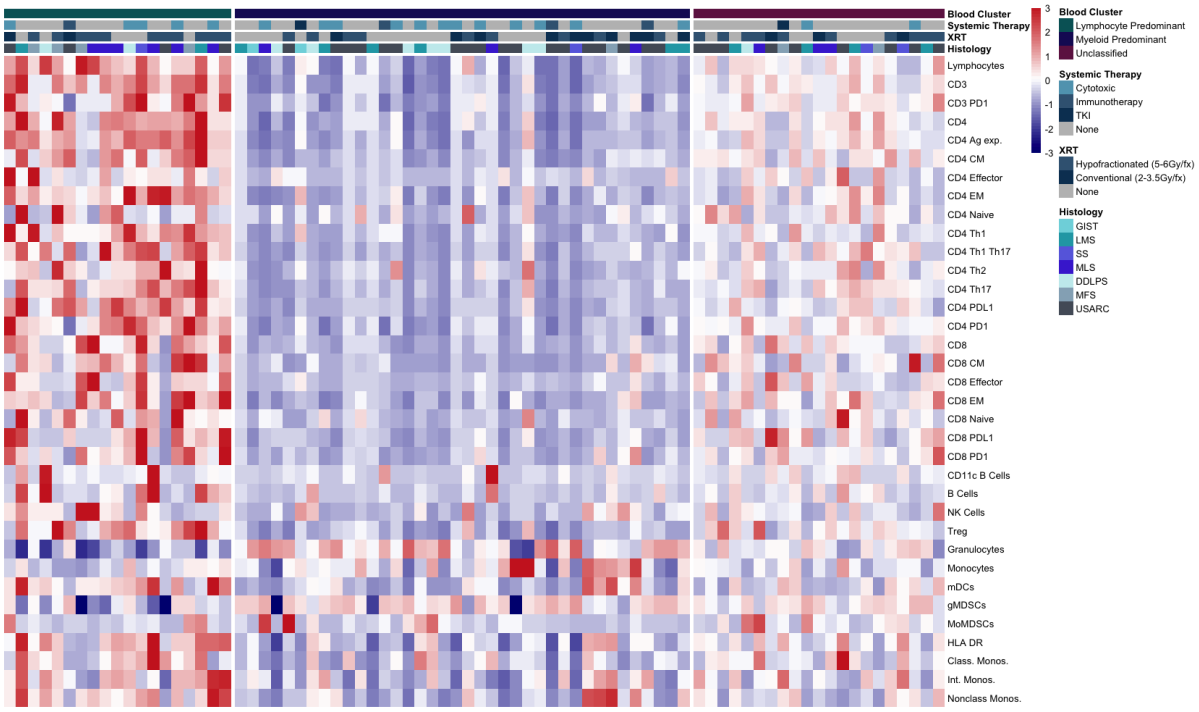
A



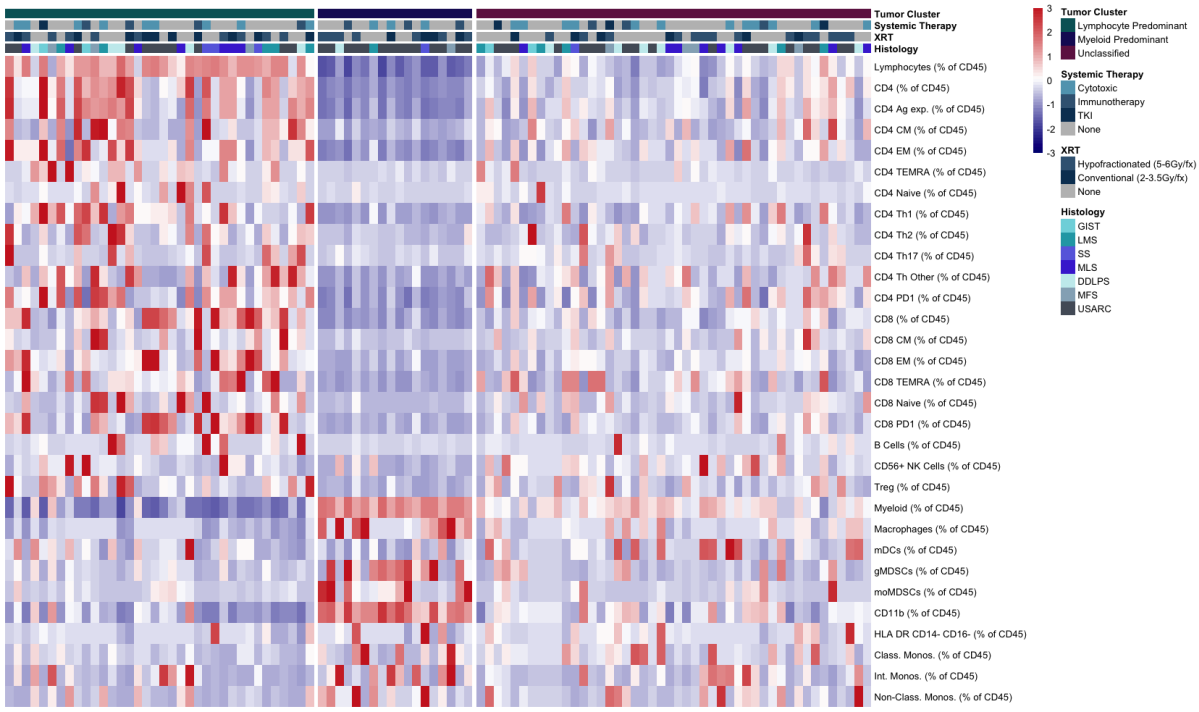
B

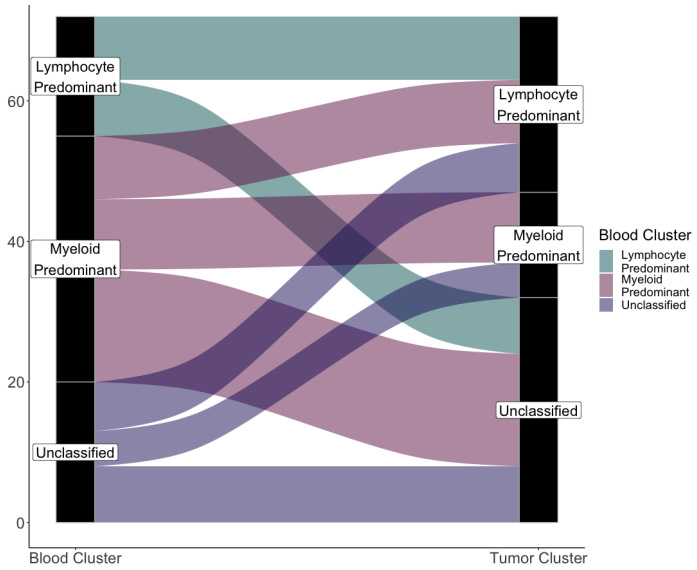


C


Blood Cluster
Systemic Therapy
XRT
Histology
Lymphocytes
CD3
CD3 PD1
CD4
CD4 Ag exp.
CD4 CM
CD4 Effector
CD4 EM
CD4 Naive
CD4 Th1
CD4 Th1 Th17
CD4 Th2
CD4 Th17
CD4 PDL1
CD4 PD1
CD8
CD8 CM
CD8 Effector
CD8 EM
CD8 Naive
CD8 PDL1
CD8 PD1
CD11c B Cells
B Cells
NK Cells
Treg
Granulocytes
Monocytes
mDCs
gMDSCs
MoMDSCs
HLA DR
Class. Monos.
Int. Monos.
Nonclass Monos.
3
2
1
0
-1
-2
-3
Blood Cluster
Lymphocyte Predominant
Myeloid Predominant
Unclassified
Systemic Therapy
Cytotoxic
Immunotherapy
TKI
None
XRT
Hypofractionated (5-6Gy/fx)
Conventional (2-3.5Gy/fx)
None
Histology
GIST
LMS
SS
MLS
DDLPS
MFS
USARC

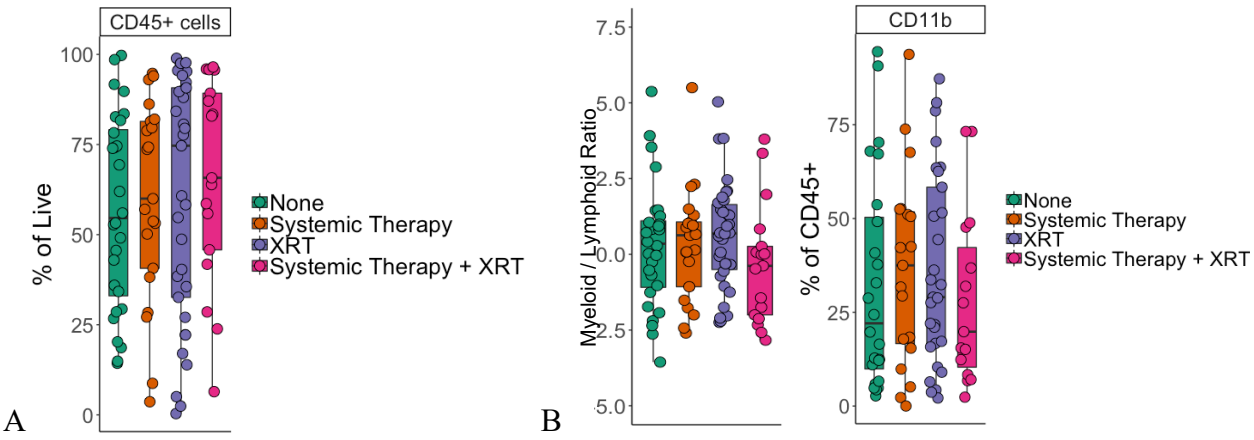
D

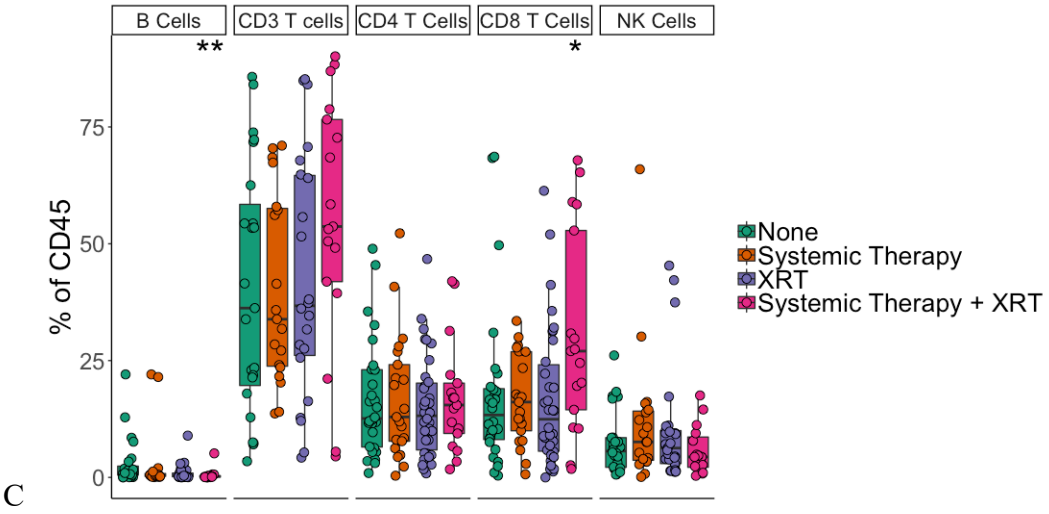

Tumor Cluster
Systemic Therapy
XRT
Histology
Lymphocytes (% of CD45)
CD4 (% of CD45)
CD4 Ag exp. (% of CD45)
CD4 CM (% of CD45)
CD4 EM (% of CD45)
CD4 TEMRA (% of CD45)
CD4 Naive (% of CD45)
CD4 Th1 (% of CD45)
CD4 Th2 (% of CD45)
CD4 Th17 (% of CD45)
CD4 Th Other (% of CD45)
CD4 PD1 (% of CD45)
CD8 (% of CD45)
CD8 CM (% of CD45)
CD8 EM (% of CD45)
CD8 TEMRA (% of CD45)
CD8 Naive (% of CD45)
CD8 PD1 (% of CD45)
B Cells (% of CD45)
CD56+ NK Cells (% of CD45)
Treg (% of CD45)
Myeloid (% of CD45)
Macrophages (% of CD45)
mDCs (% of CD45)
gMDSCs (% of CD45)
moMDSCs (% of CD45)
CD11b (% of CD45)
HLA DR CD14- CD16- (% of CD45)
Class. Monos. (% of CD45)
Int. Monos. (% of CD45)
Non-Class. Monos. (% of CD45)
3
2
1
0
-1
-2
-3
Tumor Cluster
Lymphocyte Predominant
Myeloid Predominant
Unclassified
Systemic Therapy
Cytotoxic
Immunotherapy
TKI
None
XRT
Hypofractionated (5-6Gy/fx)
Conventional (2-3.5Gy/fx)
None
Histology
GIST
LMS
SS
MLS
DDLPS
MFS
USARC

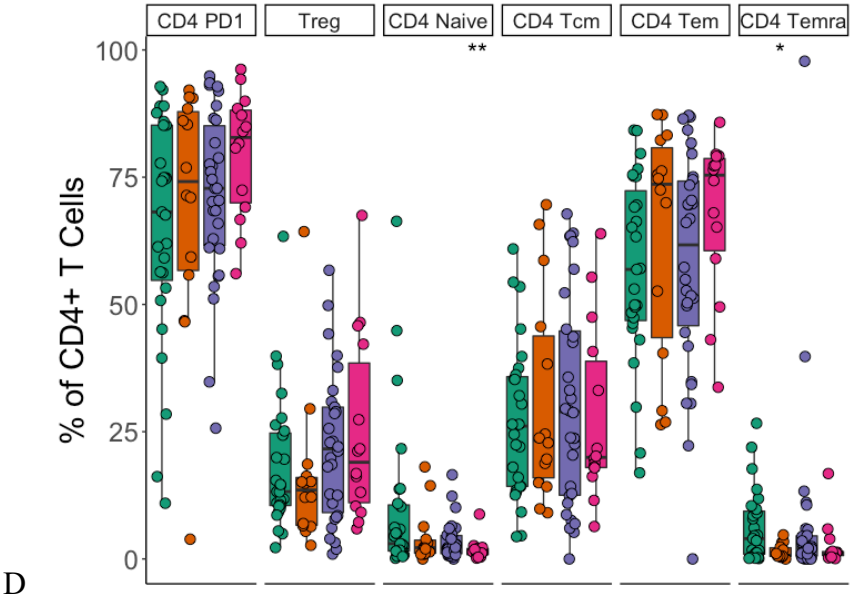


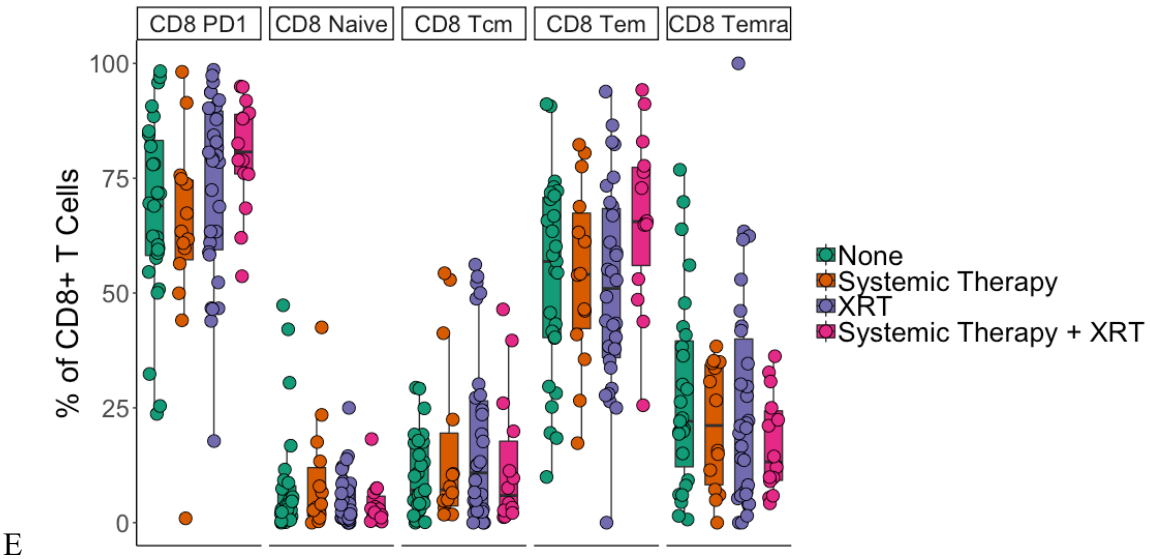
E

Figure 2-4: Associations between intratumoral and peripheral immune states in patients with STS. The lymphocyte content of tumors is associated with the proportion of lymphocyte among peripheral blood leukocytes. (A & B) The association between cell types in paired tumor and blood samples from patients with STS (n = 72). Pearson correlation coefficients are shown. (C & D) Heatmaps of flow cytometry results from blood samples (n= 72) (C) and tumor samples (n = 100) (D) from patients with STS highlight three patterns of immune composition: lymphocyte-predominant, myeloid-predominant, and unclassified. Heatmap values are scaled across rows. Missing data is set to the median value of the row and is white on the heatmap. (E) The alluvial plot shows the association between these defined blood and lymphocyte clusters in paired STS samples (n = 72). These findings show that myeloid-predominant tumor and lymphocyte-predominant blood are mutually exclusive.

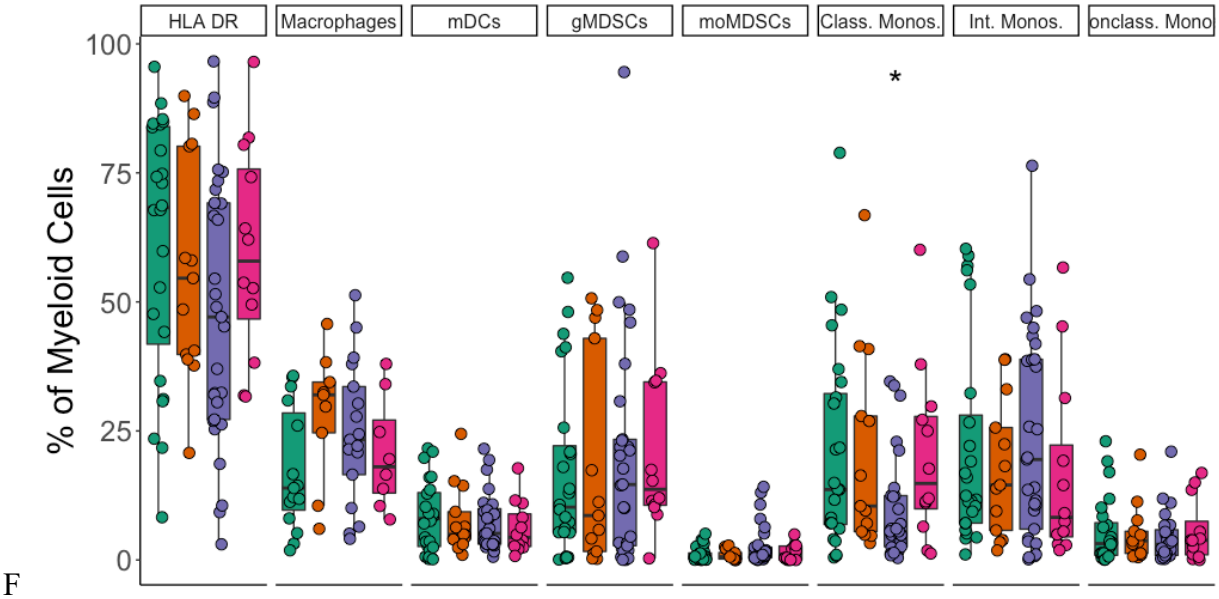

A
CD45+ cells
% of Live
100
75
50
25
0
None
Systemic Therapy
XRT
Systemic Therapy + XRT
B
Myeloid / Lymphoid Ratio
7.5
5.0
2.5
0.0
2.5
5.0
CD11b
% of CD45+
75
50
25
0
None
Systemic Therapy
XRT
Systemic Therapy + XRT


C
B Cells
CD3 T cells
CD4 T Cells
CD8 T Cells
NK Cells
**
*
% of CD45
75
50
25
0
None
Systemic Therapy
XRT
Systemic Therapy + XRT


D
CD4 PD1
Treg
CD4 Naive
CD4 Tcm
CD4 Tem
CD4 Temra
**
*
% of CD4+ T Cells
100
75
50
25
0



E

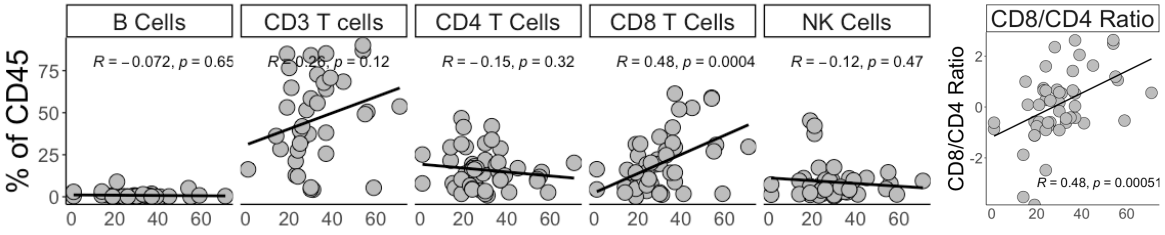


F

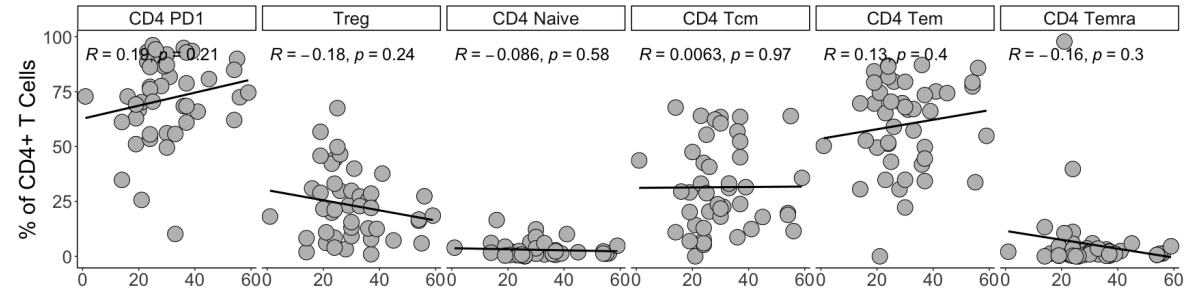
Figure 2-5: Summary of immune cell infiltrates across neoadjuvant therapy types STS tumor samples. There was no significant change in CD45+ content, CD11b+ myeloid content, or myeloid/lymphoid ratio across the groups. There was a decrease in B cells and an increase in CD8 T cell content after combination of systemic therapy and radiation. However, the composition of the myeloid cells, CD8 T cells and CD4 T cells remained largely the same, except for a decrease in the proportion of naïve CD4 T cells. (A) The proportion of CD45 (% of live) cells and (B) the myeloid/lymphoid ratio and CD11b quantities across neoadjuvant

treatment groups in STS tumor samples. Values shown are log2(myeloid/lymphoid ratio). The relative abundances of various cell populations as a % of lymphocytes (C), % of CD4+ T cells (D), % of CD8+ T cells (E), % of myeloid cells (F) across neoadjuvant groups. Significance testing was performed using Wilcoxon test with no neoadjuvant treatment group as the reference. Using Bonferroni correction, statistical significance is defined as <0.002. *, P < 0.05; **, P < 0.01; ***, P < 0.001 ; ****, P < 0.0001.

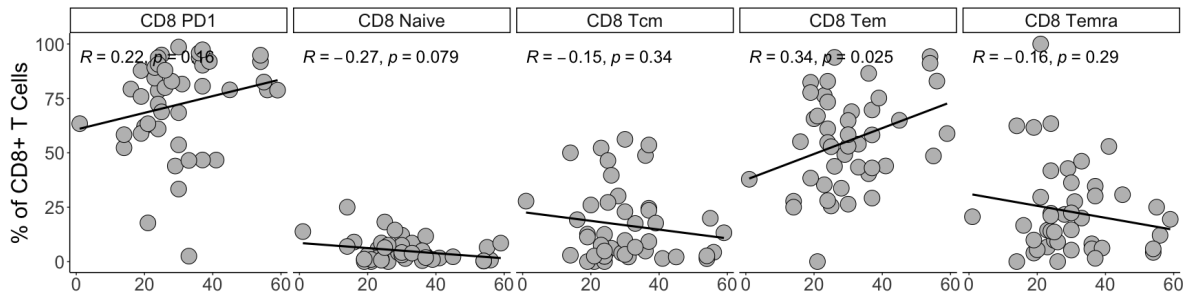
A



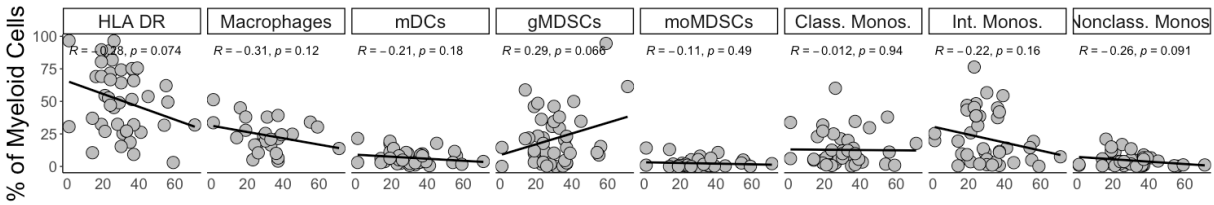
B

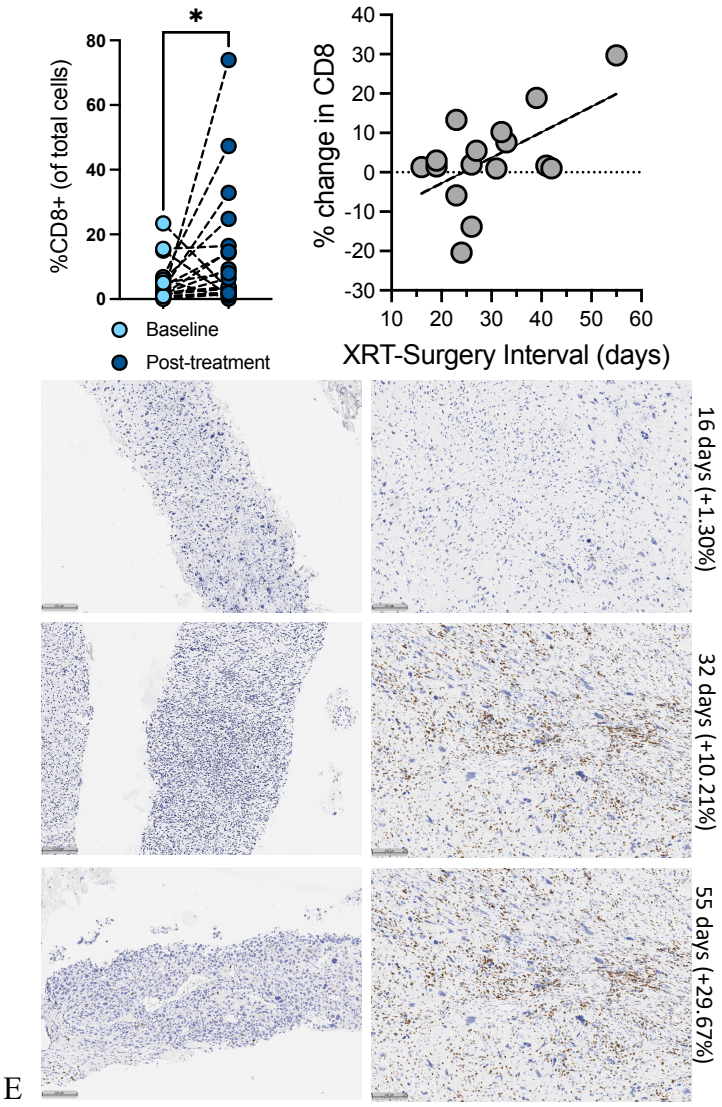


C



D





E

Figure 2-6: Effects of radiotherapy on the tumor microenvironment evolve over time, and ultimately yield a CD8 predominant phenotype. (A-D) The association between XRT-surgery interval (in days) and the relative abundances of various cell populations as a % of lymphocytes (A), % of CD4+ T cells (B), % of CD8+ T cells (C), % of myeloid cells (D). Outliers with XRT-surgery interval 19 weeks were excluded from the analysis. (E) There is an increase in CD8 population of immune cells after XRT, and this increase in CD8 T cells is positively correlated with XRT-surgery interval. These associations are shown in the graphs and IHC figures.

Regression lines for continuous data demonstrate linear models fit to data with Pearson correlation coefficients shown.

2.8 Tables

	n or median	% or range
Total	105	
Sex		
F	42	40.0
M	63	60.0
Age (years)	63	18-91
Tumor size (max dimension, cm)		
≤5cm	28	26.7
5-10cm	42	40.0
>10cm	35	33.3
Grade		
Intermediate	22	20.9
High	60	57.1
Missing	23	21.9
Tumor Site		
Lung	1	1.0
Extremity/Trunk	72	68.6
RP/Abdomen/Pelvis	32	30.5
Histology		
GIST	4	3.8
LMS	13	12.4
SS	5	4.8
MLS	16	15.2
DDLPS	14	13.3
MFS	8	7.6
USARC	45	42.9
Lesion Type		
Primary	62	59.0
Locally Recurrent	17	16.2
Metastatic	26	24.8
CNA Group		
CNA-High	67	63.8
CNA-Low	38	36.2
Genomic Type		
Simple	25	23.8
Complex	80	76.2
Preoperative Radiation		
Hypofractionated	33	31.4
Conventional	20	19.0
None	52	49.5
Radiation-Surgery Interval (Days)		
≤1 day	2	1.9
2-4 weeks	25	23.8

5-8 weeks	22	21.0
9-11 weeks	2	1.9
>19 weeks	2	1.9
NA	52	49.5
Preoperative Systemic Therapy		
Cytotoxic	25	23.8
Immunotherapy	10	9.5
TKI	4	3.8
None	66	62.9
Treatment Effect		
≤10%	14	13.3
10-50%	20	19.0
51-89%	18	17.1
≥90%	14	13.3
NA	39	37.1
TIL Growth		
0	38	36.1
1	21	20.0
2	26	24.8
NA	20	19.0
Treatment Status		
Untreated	31	29.5
Treated	74	70.5
Treatment Type		
None	31	29.5
Systemic Therapy	21	20.0
XRT	35	33.3
Systemic Therapy + XRT	18	17.1

Table 2-1: Clinical characteristics of study cohort of patients with soft tissue sarcoma.

Cell Type	Estimate (95% CI)	SE	p value
CD11b (% of CD45)	-21.45 (-32.34, -10.57)	5.47	<0.001
Myeloid (% of CD45)	-11.53 (-21.75, -1.31)	5.15	0.02
CD8 (% of CD45)	1.23 (-5.81, 8.28)	3.55	0.73
CD8PD1 (% of CD8)	-8.10 (-18.17, 1.97)	5.07	0.11
CD4 (% of CD45)	5.30 (0.31, 10.29)	2.57	0.04
CD4PD1 (% of CD4)	-7.48 (-17.78, 2.82)	5.18	0.15
Treg (% of CD4)	-13.75 (-19.89, -7.61)	3.09	<0.001

Table 2-2: Summary of results of multivariate linear regression analyses examining the association between various intratumoral leukocyte population and CNA histology groups. Cell types (dependent variables) included are listed on the left. Independent variables included in the model were CNA histology group, XRT, chemotherapy, and lesion type (primary, recurrence,

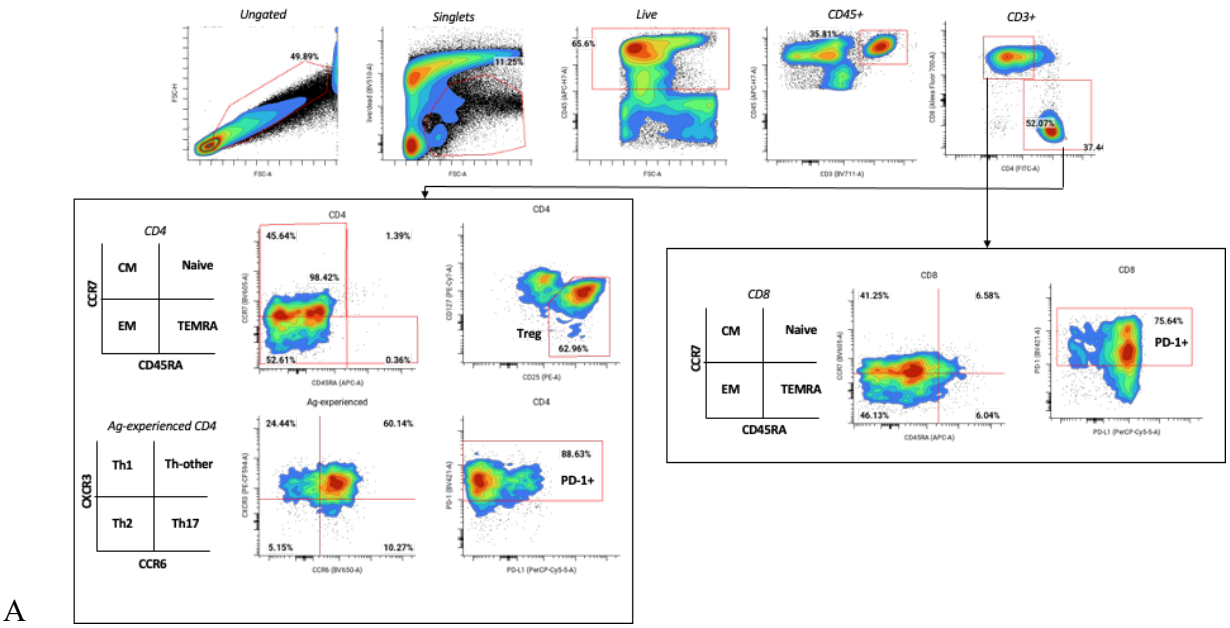
metastasis). The results of the CNA histology groups resulting from these models are listed. CNA-high tumors are associated with increased CD11b (% of CD45), myeloid cells (% of CD45), CD4 (% of CD45), and Treg (% of CD4) populations when compared with the CNA-low histology group tumors. The complete models and additional sensitivity results and subgroup analyses are shown in Tables 2-6 through 2-8.

2.9 Appendix to Chapter 2

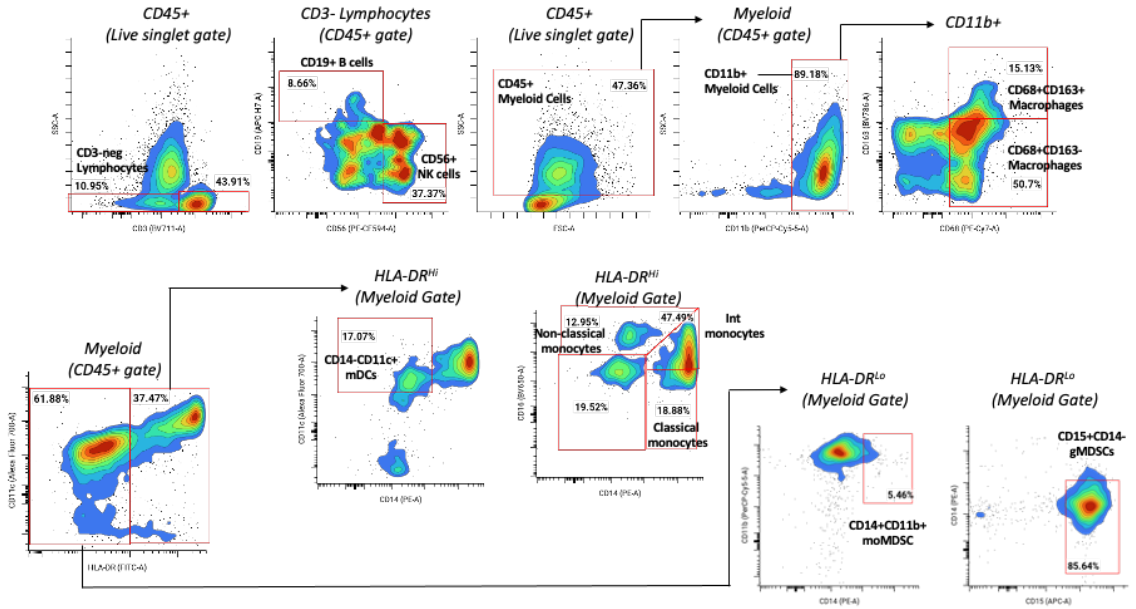
2.9.1 Myeloid and lymphoid gating strategies for tumor and blood samples

The gating strategy for this study is summarized in Figure 2-7. Figure 2-7A depicts the lymphoid gating strategy for both tumor and blood samples. We identified immune cells using CD45 and T cells using CD3, CD4, and CD8. Within the T cell subtypes, we defined regulatory T cells with CD25 + CD127 expression. Notably, we did not use FoxP3 to define Tregs, as this requires intracellular staining and is challenging when performing high throughput studies. We defined naïve and memory subsets, including central memory, effector memory, and TEMRA, using CD45RA and CCR7. We also defined PD1 expression on CD4 and CD8. Figure 2-7B describes the myeloid gating strategy. Immune cells were again identified using CD45. The absence of CD3 was used to identify myeloid cells. CD19 and CD56 were used to identify B and NK cells, respectively. Macrophages were identified using CD11b and CD68, with subtypes of macrophages identified using CD163. HLA DR was used to further subset the myeloid population. Of the HLA DR high cells, CD14 and CD11c were used to identify myeloid dendritic cells and CD16 and CD14 were used to identify subsets of monocytes. Finally, of the

HLA DR low cells, CD14 and CD11b and CD15 without CD14 defined monocytic and granulocytic myeloid derived suppressor cells, respectively.



A



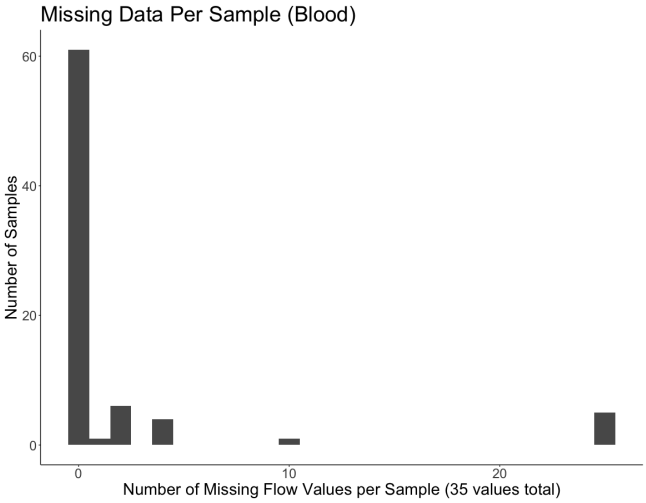
B

Figure 2-7: The gating strategy used in this study. (A) The gating strategy for lymphoid populations for tumor and blood samples. (B) The gating strategy for myeloid populations for tumor and blood samples.

2.9.2 Summary of missing flow data from blood and tumor samples

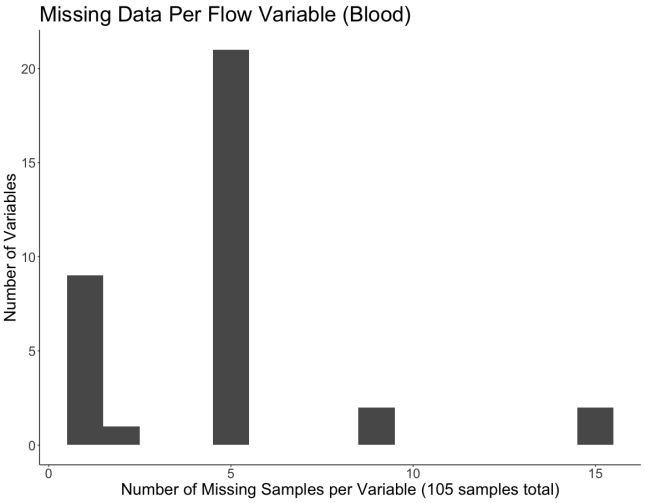
Below is a summary of the missing data in this study. The data is depicted both as missing data per blood sample and missing data per flow variable in each blood sample (Figure 2-8A & 2-8B) as well as missing data per tumor sample and missing data per flow variable in each tumor sample (Figure 2-8C & 2-8D).

A



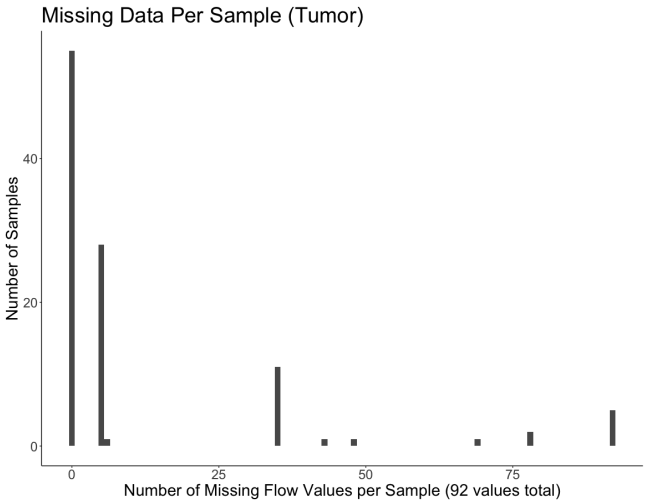
Number of missing flow variables	Frequency
0	61
1	1
2	6
4	4
10	1
25	5

B



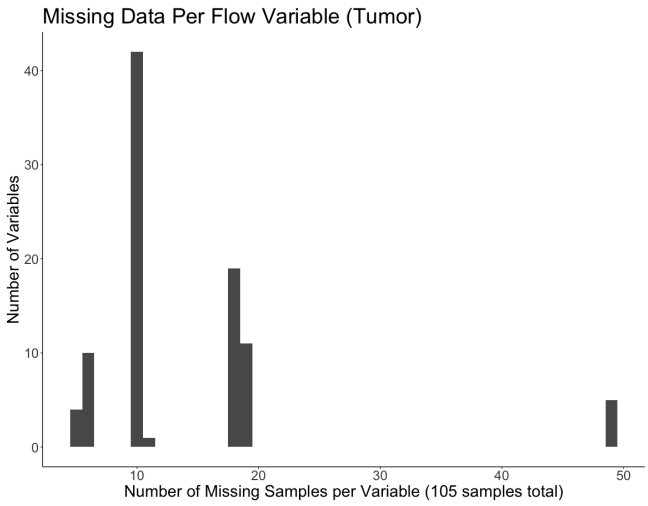
Number of missing samples per variable	Frequency
1	9
2	1
5	21
9	2
15	2

C



Number of missing flow variables	Frequency
0	55
5	28
6	1
35	11
43	1
48	1
69	1
78	2
92	5

D



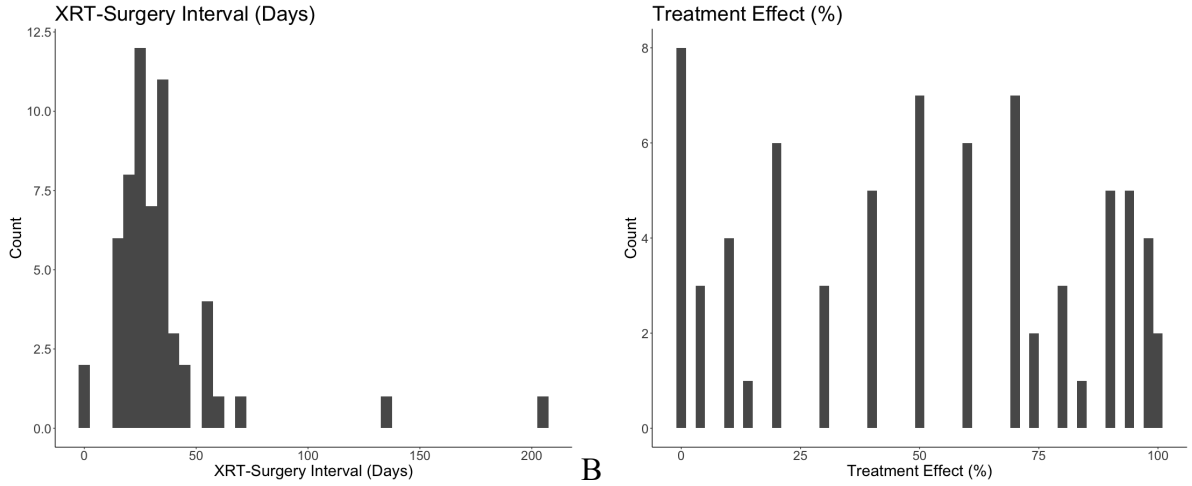
Number of missing samples per variable	Frequency
5	4
6	10
10	42
11	1
18	19
19	11
49	5

Figure 2-8: Summary of missing data. Bar graphs and tables represent (A) missing data per blood sample, (B) missing data per flow cytometry variable from blood samples, (C) missing data per tumor sample, and (D) missing data per flow cytometry variable from tumor samples.

2.9.3 Additional analysis of clinical characteristics of study cohort

We examined select clinical variables, specifically XRT-surgery interval and treatment effect, in additional detail. Both of these variables were originally continuous variables from which categorical variables were created. Figure 2-9A is a of the XRT-surgery interval (in days).

Patients with XRT-surgery interval (defined as days from end of XRT to surgery date) >133 days (19 weeks) were determined to be outliers and excluded from analyses of XRT-surgery interval when noted and as applicable. Categories were ≤1 day, 2-4 weeks, 5-8 weeks, 9-11 weeks, and >19 weeks, and were made based on clinical significance. Figure 2-9B depicts the histogram for treatment effect (%). Categories were ≤10%, 10-50%, 51-89%, and ≥90%, as these were approximately quartiles of the data. Finally, the table in Table 2-3 displays how various anatomic sites were recoded for this analysis.



A B

Figure 2-9: Additional analysis of clinical characteristics of study cohort. (A) Histogram summarizing XRT-Surgery interval in days. (B) Histogram summarizing treatment effect (%).

Anatomic Site (original)	n	Anatomic Site (new)
Abdomen	7	RP/Abdomen/Pelvis
Abdomen/Pelvis	4	RP/Abdomen/Pelvis
Abdomen/RP	4	RP/Abdomen/Pelvis
Extremity	58	Extremity/Trunk
Extremity/Groin	2	Extremity/Trunk
Flank / Gluteus	1	Extremity/Trunk
Gluteal	4	Extremity/Trunk
Gluteus	2	Extremity/Trunk
Groin	3	Extremity/Trunk
Lung	1	Chest
Pelvis	4	RP/Abdomen/Pelvis

Pelvis/Groin	1	RP/Abdomen/Pelvis
Pelvis	1	RP/Abdomen/Pelvis
RP	18	RP/Abdomen/Pelvis
RP/Abdomen	3	RP/Abdomen/Pelvis
Trunk	10	Extremity/Trunk

Table 2-3: Table summarizing the recoding of anatomic site variables.

2.9.4 Correlation analysis of various intratumoral leukocyte populations

Below is additional correlation analysis of myeloid cells with various lymphocyte subpopulations (Table 2-4) and of lymphoid cells with various myeloid subpopulations (Table 2-5). Values shown are Spearman correlation coefficients. Figure 2-10 shows the correlation between all measured cell types from blood and tumor samples included in this survey.

% of Myeloid Cells	**Correlation (R) with lymphocytes (% CD45)**
mDCs	0.38
HLA DR	0.10
Classical monocytes	-0.01
Intermediate monocytes	-0.05
Nonclassical monocytes	-0.06
gMDSCs	-0.14
Macrophages	-0.15
moMDSCs	-0.23
CD11b+	-0.47

Table 2-4: Correlation between lymphocytes (% of CD45+ cells) and the subsets of myeloid cells (% of myeloid cells). Myeloid cell types are ordered by descending Spearman correlation coefficient (R) with lymphocytes. This table corresponds with Figure 2-2F.

Lymphocyte populations	**Correlation (R) with myeloid cells (% CD45)**
CD8 Temra (% of CD8)	0.24

NK Cells (% of lymphocytes)	0.19
Treg (% of CD4)	0.08
CD4 Tem (% of CD4)	0.07
CD4 Temra (% of CD4)	0.05
CD4 PD1 (% of CD4)	0.01
CD4 T Cells (% of lymphocytes)	-0.01
CD4 Naïve (% of CD4)	-0.04
CD4 Tcm (% of CD4)	-0.08
CD8 Naïve (% of CD8)	-0.08
CD8 Tcm (% of CD8)	-0.08
CD8 PD1 (% of CD8)	-0.12
CD8 Tem (% of CD8)	-0.13
B Cells (% of lymphocytes)	-0.19
CD8 T Cells (% of lymphocytes)	-0.37

Table 2-5: Correlation between myeloid cells (% of CD45+ cells) and the subpopulations of lymphocytes. Lymphocyte cell types are ordered by descending Spearman correlation coefficient (R) with myeloid cells. This table corresponds with Figure 2-2G.

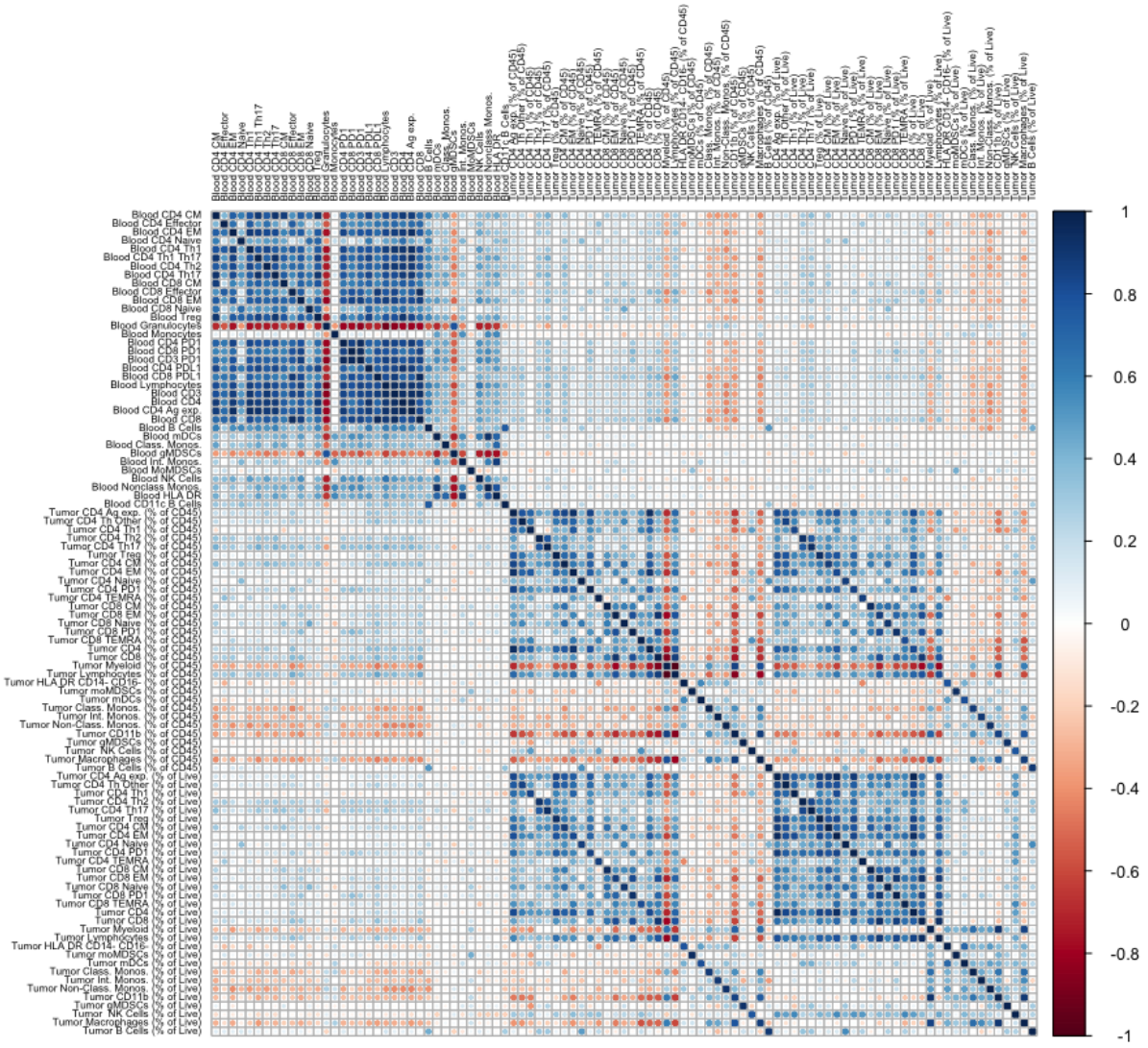


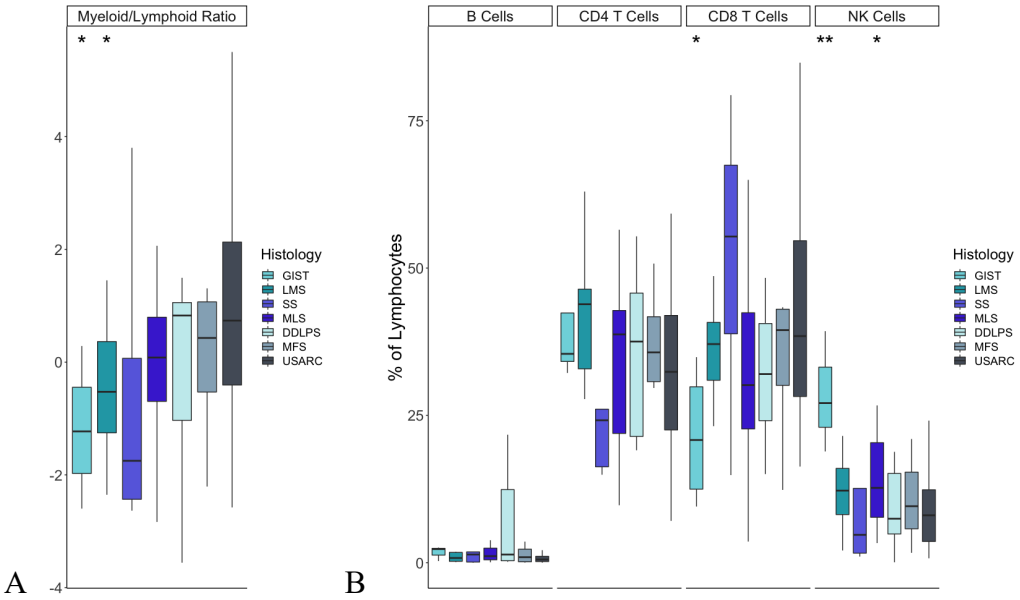
Figure 2-10: The correlation between immune cell types in blood and tumor samples of patients with STS. The Spearman correlation coefficient between each pair of cell types is depicted in the correlation plot.

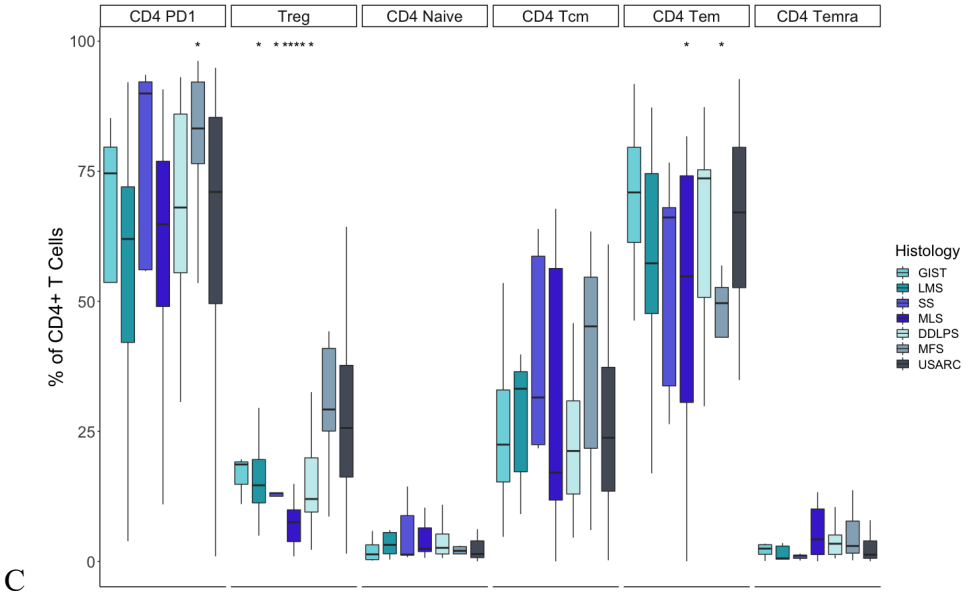
2.9.5 Correlation between histology and genomics and intratumoral leukocyte composition

Undifferentiated sarcomas are characterized by myeloid predominance and a relative abundance of suppressor cells. We examined the intratumoral leukocyte populations in each of the subtypes

of soft tissue sarcoma included in our study. The plurality of tumor samples in our study were undifferentiated sarcoma. In all comparisons, undifferentiated sarcoma is the reference group. These findings highlight the relative myeloid predominance of undifferentiated sarcoma. Specifically, Figure 2-5A demonstrates that undifferentiated sarcoma samples showed a trend toward having the highest myeloid/lymphoid ratio of all the subtypes included in this study (values depicted are log2 of the myeloid to lymphoid ratio). Other notable findings from Figure 2-11B-E are the relative abundance of suppressor cells, such as T regulatory cells, CD11b+ cells, and monocytic myeloid derived suppressor cells (moMDSCs). Figure 2-12 demonstrates additional analyses, including subgroup analyses, exploring the myeloid predominance of undifferentiated sarcoma cells as well as the association between myeloid/lymphoid ratio and tumor size by histology.

We similarly examined the intratumoral leukocyte populations between histology groups classified by simple versus complex genomics. Sarcomas with simple genomic driver events (e.g. fusion event or mutation) include synovial sarcoma (SYT-SSX fusion), myxoid liposarcoma (FUS-DDIT3 fusion) and GIST (KIT or PDGFRA mutants). Samples in this study were categorized as simple genomics (n = 25, 23.8%) or complex genomics (n = 80, 76.2%). The intratumoral leukocyte populations are summarized in Figure 2-13 below.


A
Myeloid/Lymphoid Ratio
* *
4
2
0
-2
-4
Histology
GIST
LMS
SS
MLS
DDLPS
MFS
USARC
B
B Cells
CD4 T Cells
CD8 T Cells
NK Cells
*
**
*
% of Lymphocytes
75
50
25
0
Histology
GIST
LMS
SS
MLS
DDLPS
MFS
USARC


C
CD4 PD1
Treg
CD4 Naive
CD4 Tcm
CD4 Tem
CD4 Temra
* * * * * *
* *
% of CD4+ T Cells
100
75
50
25
0
Histology
GIST
LMS
SS
MLS
DDLPS
MFS
USARC

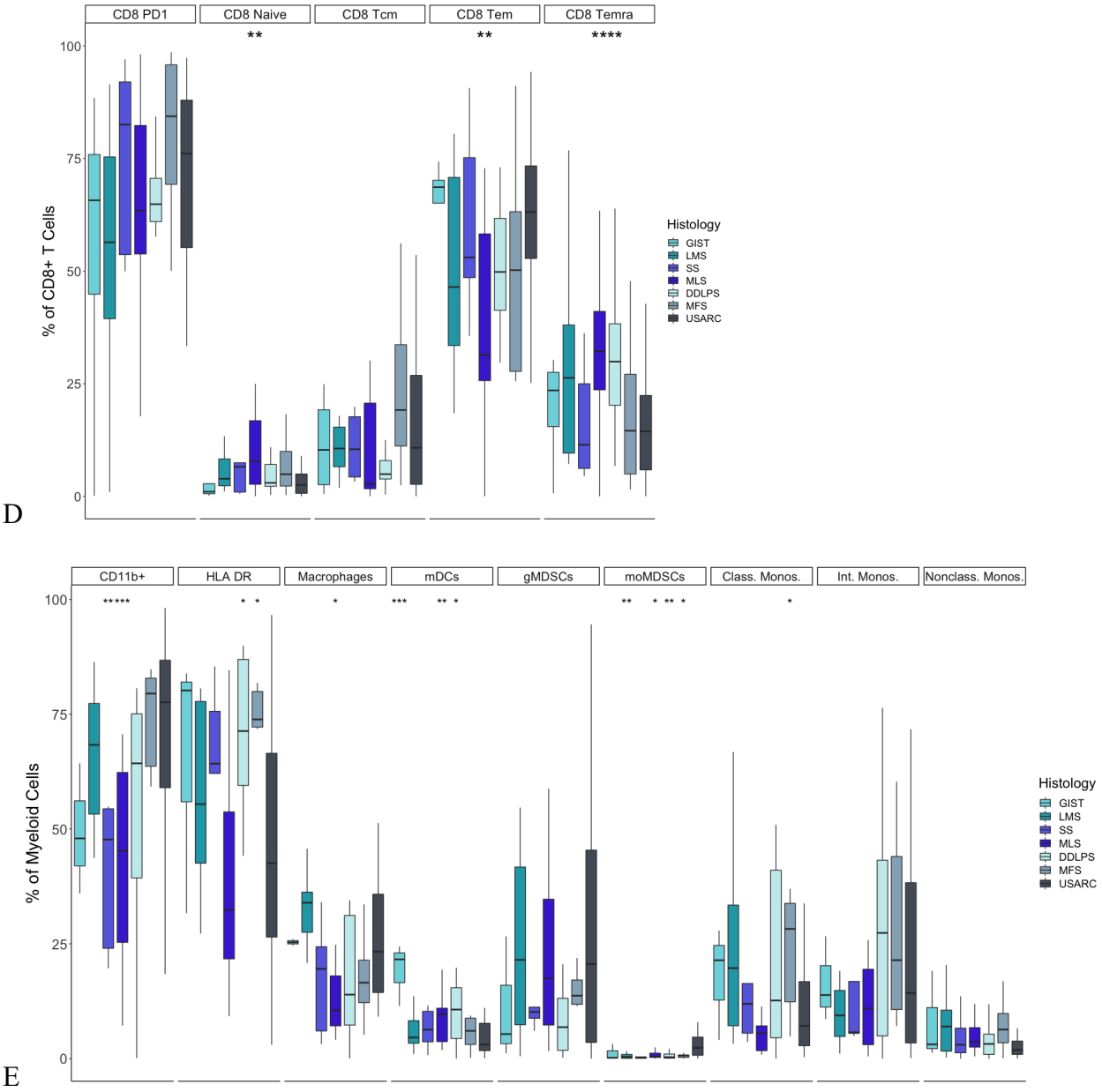
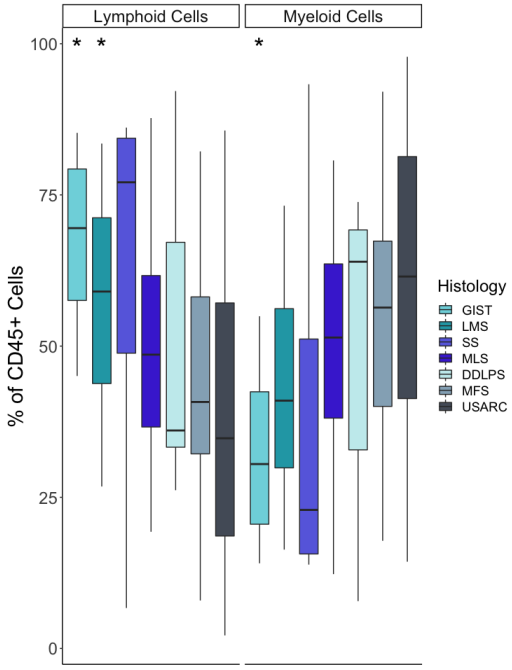
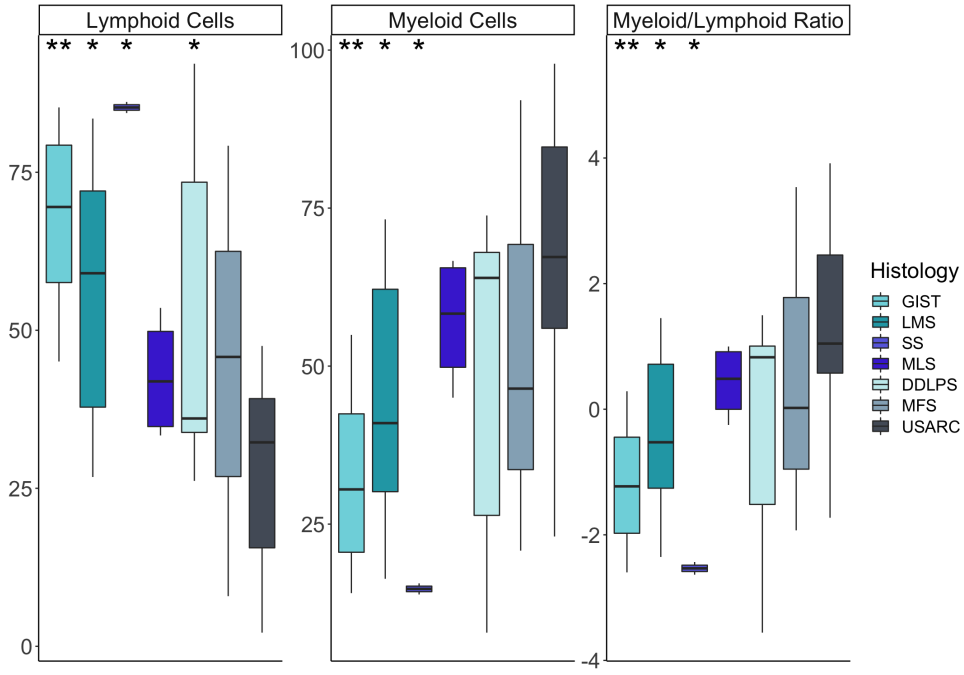


Figure 2-11: Undifferentiated sarcomas are characterized by myeloid predominance and a relative abundance of suppressor cells. (A) The myeloid/lymphoid ratio across STS histologies. Values shown are log2(myeloid/lymphoid ratio). The relative abundances of various cell populations as a % of lymphocytes (B), % of CD4+ T cells (C), % of CD8+ T cells (D), % of myeloid cells (E) across STS histologies. Significance testing was performed using Wilcoxon

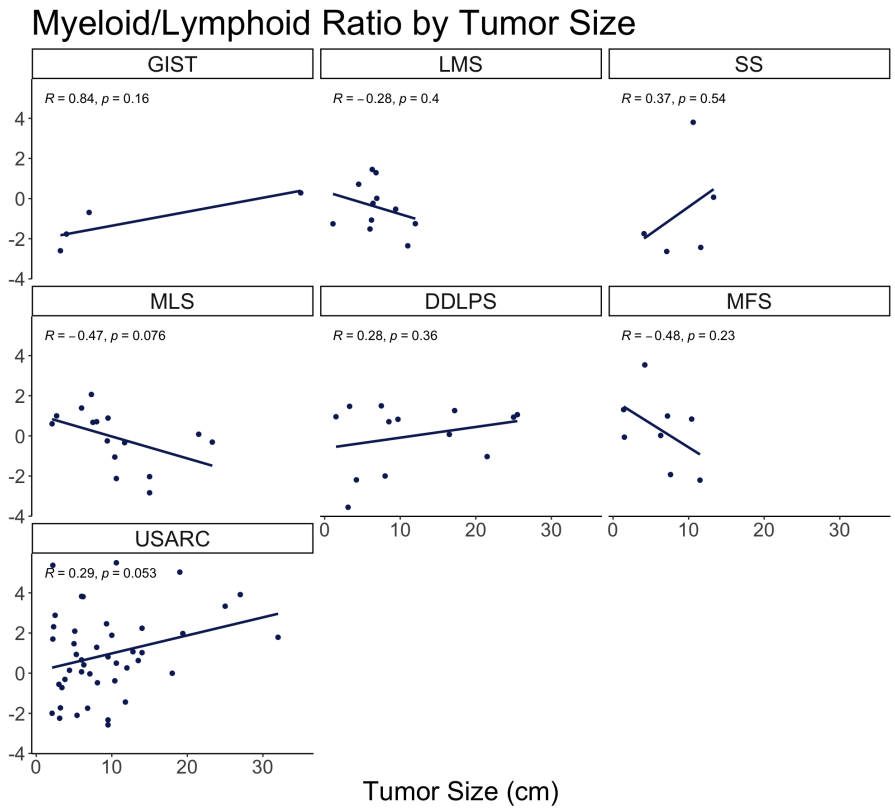
test with undifferentiated sarcoma as the reference group. Using Bonferroni correction, statistical significance is defined as <0.002. *, P < 0.05; **, P < 0.01; ***, P < 0.001 ; ****, P < 0.0001.



A

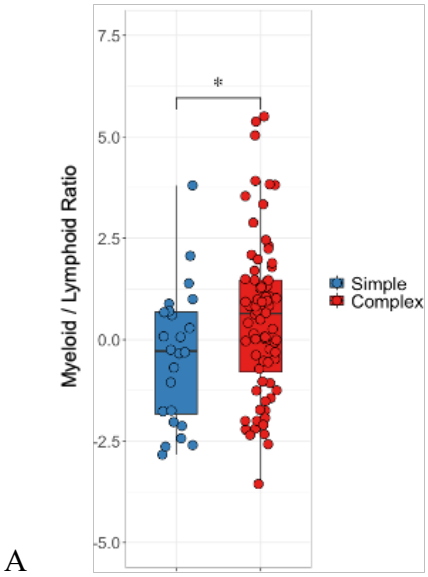


B

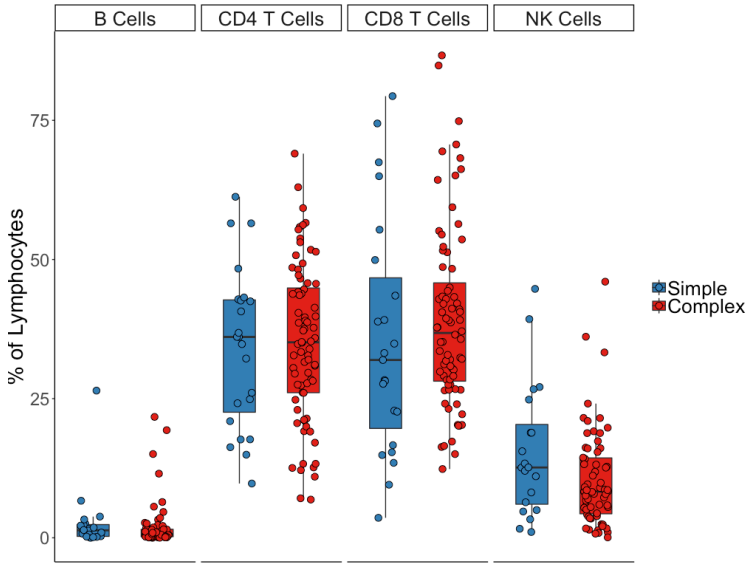


C

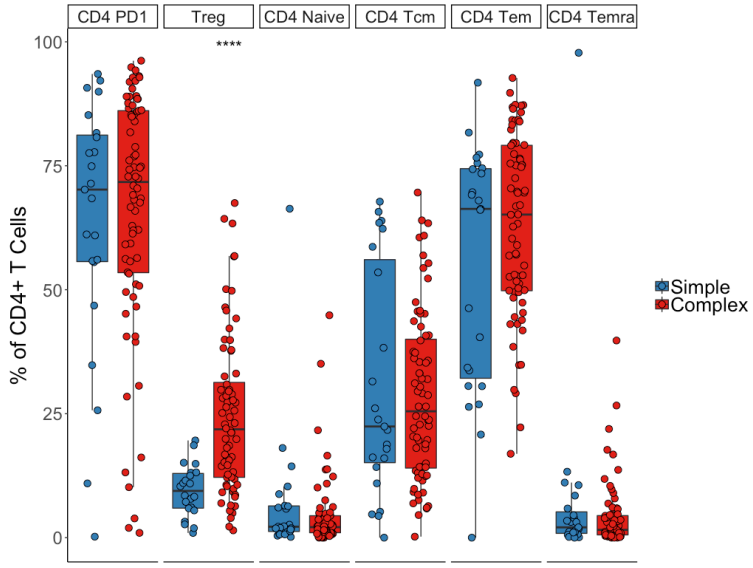
Figure 2-12: Additional overview of the immune cell infiltration landscape in tumor samples from patients with STS. (A) The relative abundances of myeloid and lymphoid cells across histologies. (B) The relative abundances of myeloid and lymphoid cells as well as the myeloid/lymphoid ratio across STS histologies in patients not treated with XRT. Myeloid and lymphoid values shown are % of CD45+ T cells. Myeloid/lymphoid ratio values shown are log2(myeloid/lymphoid ratio). (C) The association between myeloid/lymphoid ratio and tumor size (cm). Percentages shown are mean values across all STS samples. Significance testing was performed using Wilcoxon test with undifferentiated sarcoma as the reference group. Regression lines for continuous data demonstrate linear models fit to data. *, P < 0.05; **, P < 0.01; ***, P < 0.001; ****, P < 0.0001.


Myeloid / Lymphoid Ratio
7.5
5.0
2.5
0.0
-2.5
-5.0
*
Simple
Complex

A

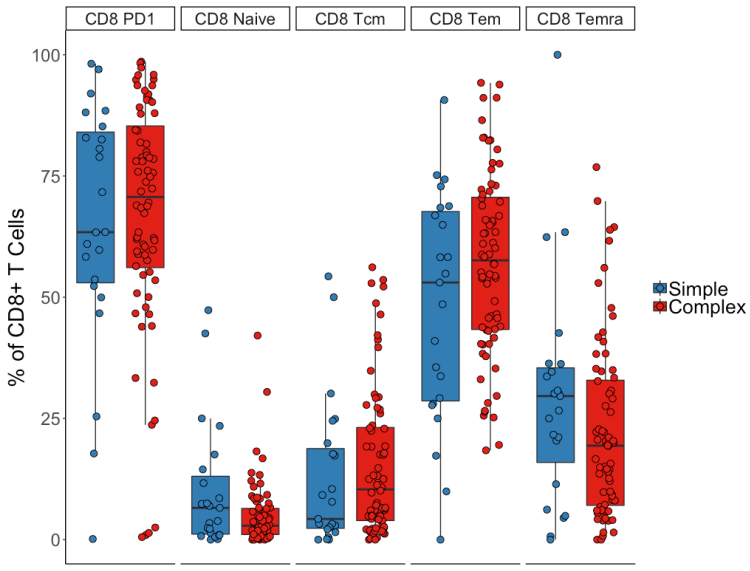

B Cells
CD4 T Cells
CD8 T Cells
NK Cells
% of Lymphocytes
75
50
25
0
Simple
Complex

B


CD4 PD1
Treg
CD4 Naive
CD4 Tcm
CD4 Tem
CD4 Temra
% of CD4+ T Cells
100
75
50
25
0
****
Simple
Complex

C

D



E

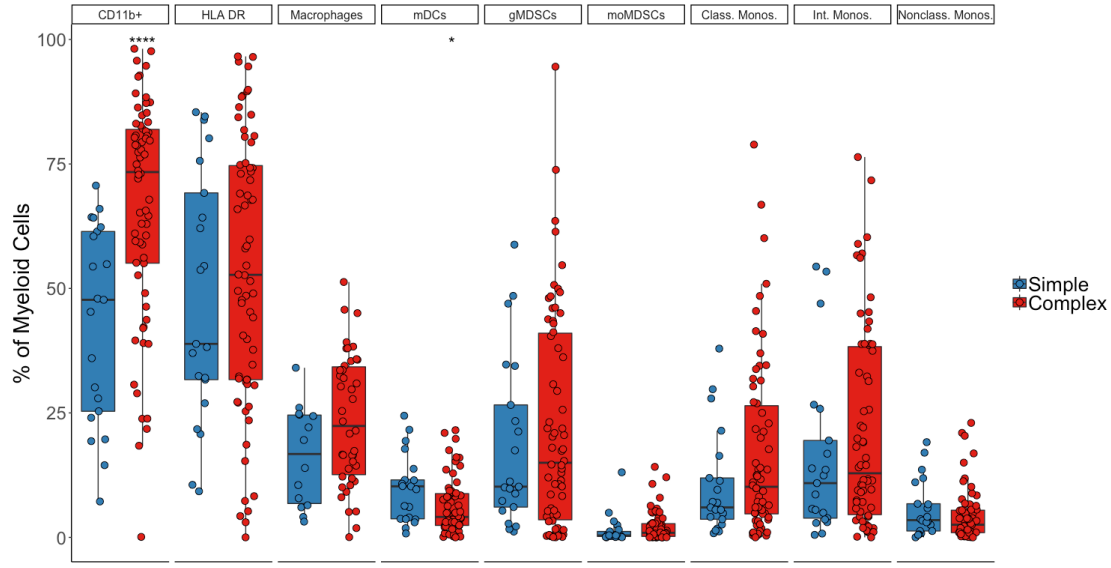


Figure 2-13: An overview of the immune cell infiltration landscape in tumor samples from patients with STS categorized by genomic subtype. (A) The myeloid/lymphoid ratio by genomic subtype; values shown are $\log_2$(myeloid/lymphoid ratio). (B-D) relative abundances of lymphocyte populations (B), CD4+ T cell subsets (C), and CD8+ T cell subsets (D) as a percentage of the parent population, grouped by genomic subtype. (E) Relative abundance of myeloid subsets grouped by genomic subtype. Significance testing was performed using Wilcoxon test with "Simple" genomic subtype as the reference group. Using Bonferroni

correction, statistical significance is defined as <0.002. *, P < 0.05; **, P < 0.01; ***, P < 0.001 ; ****, P < 0.0001.

2.9.6 Multivariate linear regression examining the correlation between CNA and intratumoral leukocyte populations

Multivariate linear regression analyses were performed to determine association between various intratumoral leukocyte population and CNA histology groups (as defined above). The CNA low group was associated with lower CD11b (% of CD45), myeloid cells (% of CD45), and Treg (% of CD4) and there was a trend toward association with higher CD4 (% of CD45) in all tumor samples and was associated with lower CD11b (% of CD45) and Treg (% of CD4) in samples from patient's who underwent pre-operative XRT.

The cell types (dependent variables) included in this analysis included CD11b (% of CD45), myeloid cells (% of CD45), CD8 T cells (% of CD45), CD8PD1 cells (% of CD8), CD4 (% of CD45), CD4PD1 (% of CD4), and Treg (% of CD4). Independent variable included in the model were CNA histology group, XRT, chemotherapy, and lesion type (primary, recurrence, metastasis). A subgroup analysis of patients who underwent XRT was performed. In this analysis, XRT-Surgery interval was included as an additional co-variate in the model. All models are described in the tables below.

	%CD11b of CD45		%Myeloid of CD45		%CD8 of CD45	
Covariate	*Estimate (95% CI)*	*p*	*Estimate (95% CI)*	*p*	*Estimate (95% CI)*	*p*
CNA low (vs CNA high)	-21.45 (-32.34, -10.57)	1.84E-03	-11.53 (-21.75, -1.31)	0.02	1.23 (-5.81, 8.28)	0.73
Systemic tx (vs no systemic tx)	0.85 (-10.61, 12.31)	0.88	-5.5 (-15.95, 4.95)	0.30	5.82 (-1.39, 13.02)	0.11
XRT (vs no XRT)	-1.95 (-14.45, 10.55)	0.76	1.05 (-10.76, 12.85)	0.86	4.08 (-4.06, 12.21)	0.32
Primary (of PRM)	6.57 (-7.62, 20.76)	0.36	5.86 (-7.08, 18.80)	0.37	-4.58 (-13.50, 4.34)	0.31
Recurrence (of PRM)	7.83 (-9.24, 24.90)	0.36	15.76 (-0.08, 31.61)	0.05	-9.55 (-20.48, 1.37)	0.09
	%CD8PD1 of CD8		%CD4 of CD45		%CD4PD1 of CD4	
Covariate	*Estimate (95% CI)*	*p*	*Estimate (95% CI)*	*p*	*Estimate (95% CI)*	*p*
CNA low (vs CNA high)	-8.1 (-18.17, 1.97)	0.11	5.3 (0.31, 10.29)	0.04	-7.48 (-17.78, 2.82)	0.15
Systemic tx (vs no systemic tx)	-10.25 (-20.62, 0.12)	0.05	1.02 (-4.09, 6.12)	0.69	-7.44 (-18.04, 3.16)	0.17
XRT (vs no XRT)	13.43 (1.73, 25.13)	0.03	-1.41 (-7.17, 4.36)	0.63	10.355 (-1.61, 22.32)	0.09
Primary (of PRM)	-11.24 (-24.21, 1.73)	0.09	-1.71 (-8.03, 4.60)	0.59	-8.12 (-21.38, 5.14)	0.23
Recurrence (of PRM)	-14.04 (-29.71, 1.63)	0.08	-5.35 (-13.09, 2.39)	0.17	-15.25 (-31.28, 0.77)	0.06
	%Treg of CD4					
Covariate	*Estimate (95% CI)*	*p*				
CNA low (vs CNA high)	-13.75 (-19.89, -7.61)	2.47E-05				
Systemic tx (vs no systemic tx)	0.61 (-5.70, 6.93)	0.85				
XRT (vs no XRT)	5.8 (-1.31, 12.90)	0.11				
Primary (of PRM)	-6.72 (-14.60, 1.16)	0.09				
Recurrence (of PRM)	-2.14 (-11.69, 7.42)	0.66				

Table 2-6: Summary of multivariate linear regression analyses investigating the association between various intratumoral leukocyte population and CNA histology groups. Dependent variables and covariates were selected based on the variables that were statistically significant on univariate analyses, as described above. The dependent variable for each model is noted at the top of each model. All models containing the XRT-surgery interval are the subgroup analyses that included only patients who underwent XRT.

	%CD11b of CD45		%Myeloid of CD45		%CD8 of CD45	
Covariate	*Estimate (95% CI)*	*p*	*Estimate (95% CI)*	*p*	*Estimate (95% CI)*	*p*
CNA low (vs CNA high)	-20.25 (-34.95, -5.55)	8.19E-03	-9.08 (-23.51, 5.35)	0.21	1.31 (-7.88, 10.49)	0.78
Systemic tx (vs no systemic tx)	-1.91 (-18.29, 14.46)	0.81	-6.61 (-22.24, 9.03)	0.40	4.87 (-5.08, 14.82)	0.33
RT-Surgery Interval (days)	-0.18 (-0.43, 0.07)	0.16	-0.21 (-0.46, 0.04)	0.10	0.33 (0.16, 0.49)	1.85E-04
Primary (vs metastasis)	3.4 (-18.84, 25.63)	0.76	6.29 (-14.26, 26.83)	0.54	-1.91 (-14.99, 11.17)	0.77
Covariate	**%CD8PD1 of CD8**		**%CD4 of CD45**		**%CD4PD1 of CD4**	
CNA low (vs CNA high)	*Estimate (95% CI)*	*p*	*Estimate (95% CI)*	*p*	*Estimate (95% CI)*	*p*
Systemic tx (vs no systemic tx)	-11.49 (-24.73, 1.74)	0.09	3.09 (-3.68, 9.85)	0.36	-8.84 (-21.02, 3.34)	0.15
RT-Surgery Interval (days)	-4.75 (-18.85, 9.34)	0.50	4.01 (-3.31, 11.34)	0.28	0.53 (-12.44, 13.50)	0.94
Primary (vs metastasis)	0.17 (-0.06, 0.41)	0.14	-0.1 (-0.22, 0.02)	0.09	-0.0557 (-0.27, 0.16)	0.60
	-2.53 (-22.59, 17.53)	0.80	-3.83 (-13.46, 5.80)	0.43	-2.206 (-20.67, 16.26)	0.81
	%Treg of CD4					
Covariate	*Estimate (95% CI)*	*p*				
CNA low (vs CNA high)	-17.24 (-25.50, -8.99)	1.29E-04				
Systemic tx (vs no systemic tx)	3.21 (-5.59, 12.00)	0.47				
RT-Surgery Interval (days)	-0.11 (-0.25, 0.04)	0.14				
Primary (vs metastasis)	-7.62 (-20.13, 4.89)	0.23				

Table 2-7: Summary of multivariate linear regression analyses investigating the association between various intratumoral leukocyte population and CNA histology groups within the subgroup of patients who received XRT. As in Table 2-6, dependent variables and covariates were selected based on the variables that were statistically significant on univariate analyses, as described above. The dependent variable for each model is noted at the top of each model. XRT-surgery interval was added as a covariate in this analysis.

Cell type	Model	p	R2 (adjusted)
%CD11b of CD45	All samples	0.0049	0.134
%CD11b of CD45	Subgroup (XRT only)	0.031	0.154
%Myeloid of CD45	All samples	0.0329	0.073
%Myeloid of CD45	Subgroup (XRT only)	0.0599	0.104
%CD8 of CD45	All samples	0.134	0.036
%CD8 of CD45	Subgroup (XRT only)	9.07E-05	0.344
%CD8PD1 of CD8	All samples	0.0075	0.113
%CD8PD1 of CD8	Subgroup (XRT only)	0.201	0.0471
%CD4 of CD45	All samples	0.151	0.032
%CD4 of CD45	Subgroup (XRT only)	0.403	0.0023
%CD4PD1 of CD4	All samples	0.031	0.078
%CD4PD1 of CD4	Subgroup (XRT only)	0.6549	-0.0347
%Treg of CD4	All samples	0.00038	0.181
%Treg of CD4	Subgroup (XRT only)	0.00145	0.2736

Table 2-8: Summary of model statistics from models described in Tables 2-6 & 2-7.

2.9.7 Additional data from the analysis of associations between intratumoral and peripheral immune states in patients with STS

In Figure 2-4 we examined the associations between intratumoral and peripheral immune states in patients with STS. Intratumoral leukocyte populations (as a percent of CD45) were compared to peripheral immune cells. The intratumoral leukocyte populations (as a percent of live cells) were also examined, and the heatmap summarizing this data is shown below. Similar distinct groups were seen in this data – leukocyte-predominant, myeloid-predominant, and unclassified.

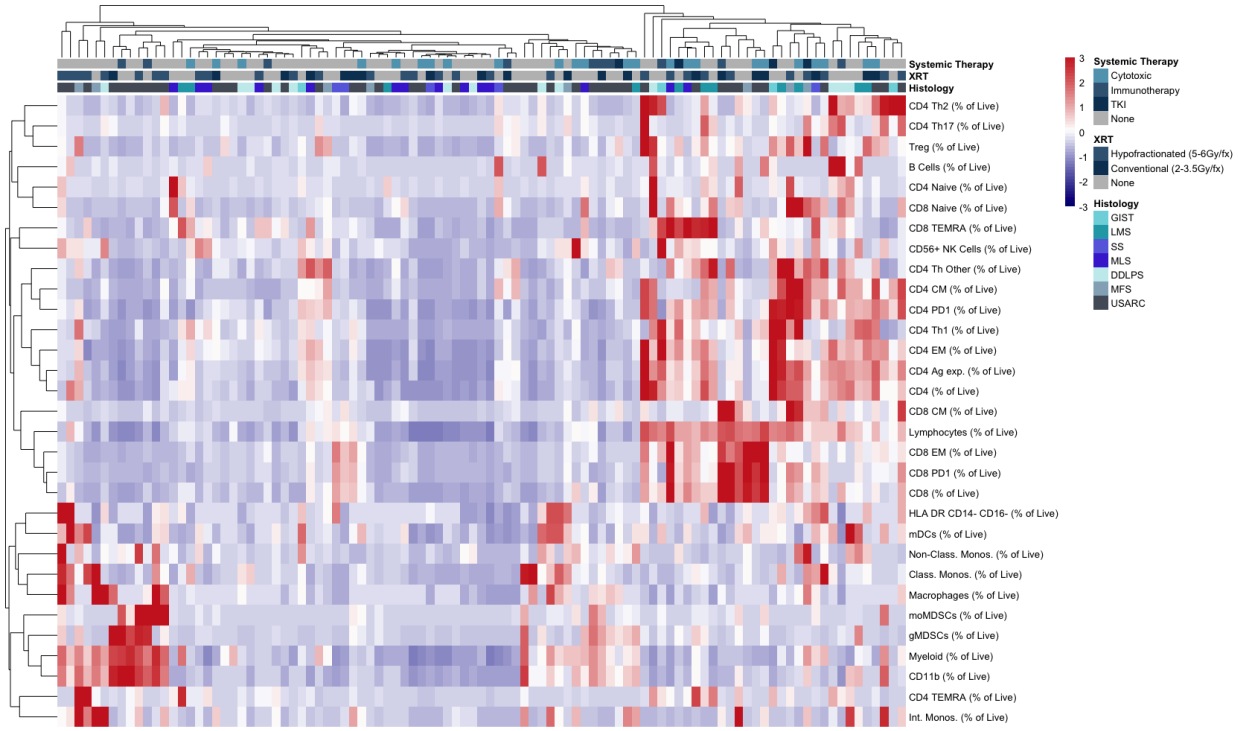
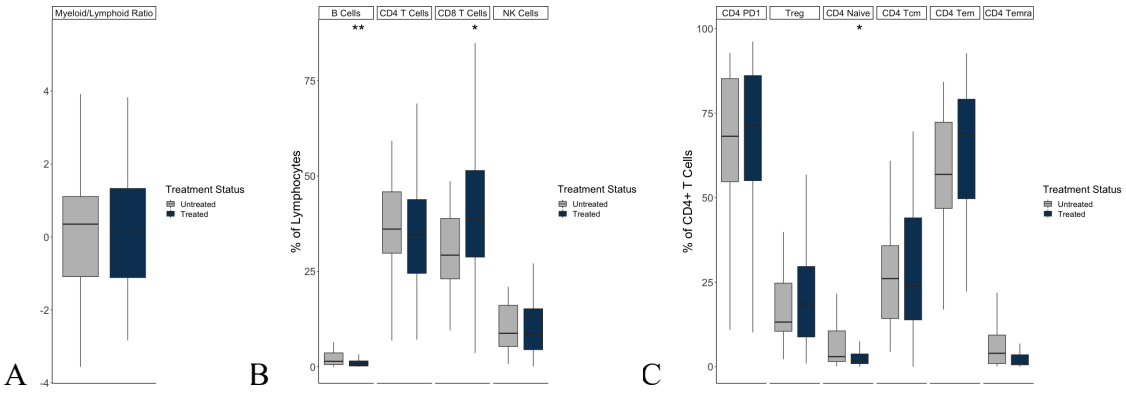


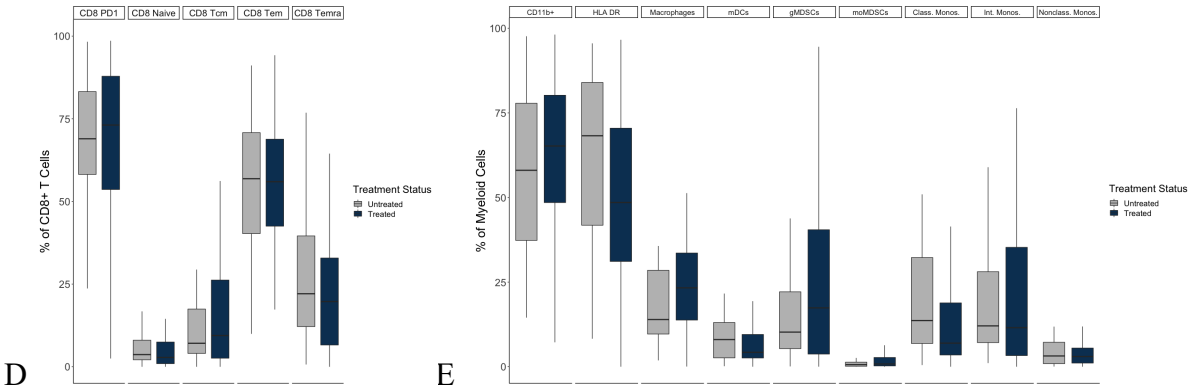
Figure 2-14: Heatmap of flow cytometry results as a percent of live cells from tumor samples (n=99) from patients with STS. Values are scaled across rows. Missing data is set to the median value of the row and is white on the heatmap. Rows are ordered based on unsupervised hierarchical clustering using Euclidean distances.

2.9.8 Additional analysis of the correlation between treatment and intratumoral leukocyte populations

In addition to exploring the association between variables such as histology, CNA histology group, and genomic histology group, as described above, we examined the associations between various treatment factors and immune cell infiltration. Overall, there were no consistent significant changes in the intratumoral leukocytes associated with various treatments, including treatment status (treated, untreated), XRT type (none, hypofractionated, conventional), systemic therapy type (none, TKI, cytotoxic, immunotherapy), and XRT (yes, no).

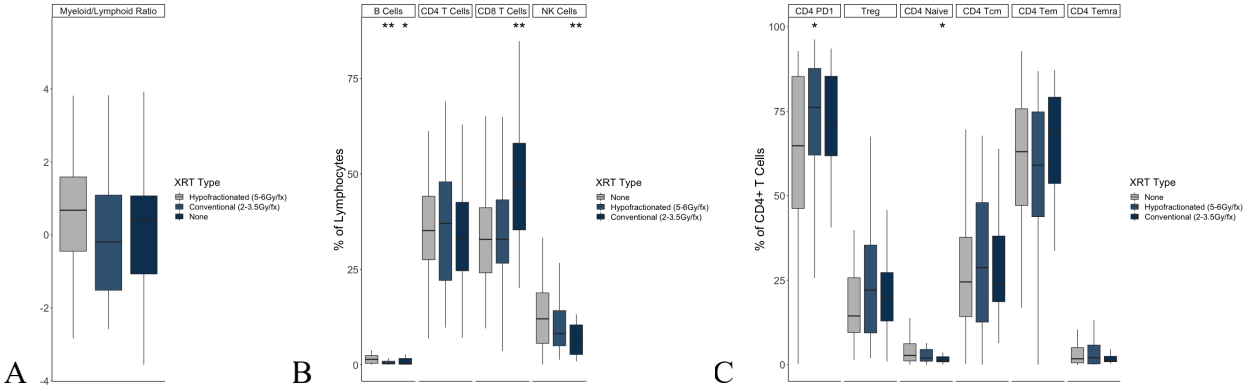
The use of XRT was associated with reduced B cells (% of CD45); increased CD4PD1 (% of CD4) and CD8PD1 (% of CD8) cells. XRT had no effect on overall immune infiltration, myeloid/lymphoid balance, or composition of myeloid cells. The association between XRT and increased CD8 PD1 was consistent in both standard and hypofractionated XRT groups.



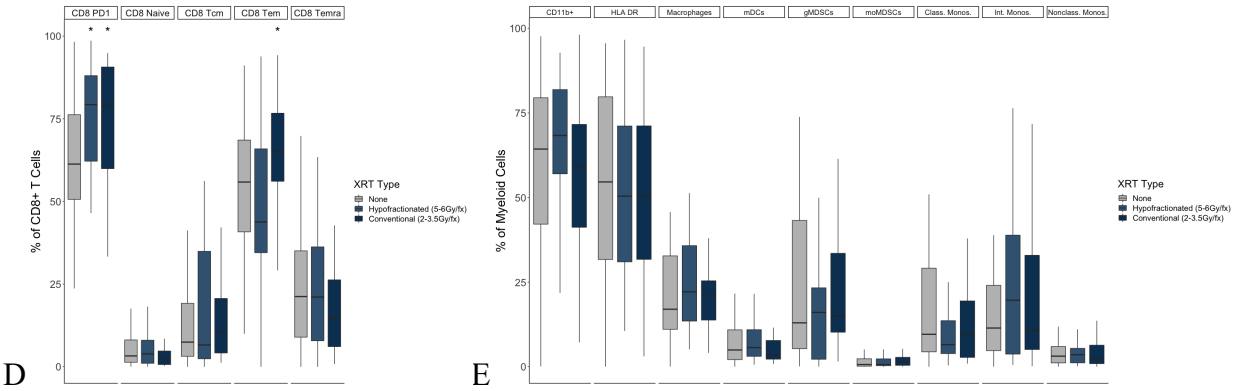


D E

Figure 2-15: The association between immune cell infiltration and treatment status in patients with STS. (A) The myeloid/lymphoid ratio by treatment status; values shown are $\log_2$(myeloid/lymphoid ratio). (B-D) relative abundances of lymphocyte populations (B), CD4+ T cell subsets (C), and CD8+ T cell subsets (D) as a percentage of the parent population, grouped by treatment status. (E) Relative abundance of myeloid subsets grouped by treatment status. Significance testing was performed using Wilcoxon test with "untreated" as the reference group. Using Bonferroni correction, statistical significance is defined as <0.002. *, P < 0.05; **, P < 0.01; ***, P < 0.001 ; ****, P < 0.0001.

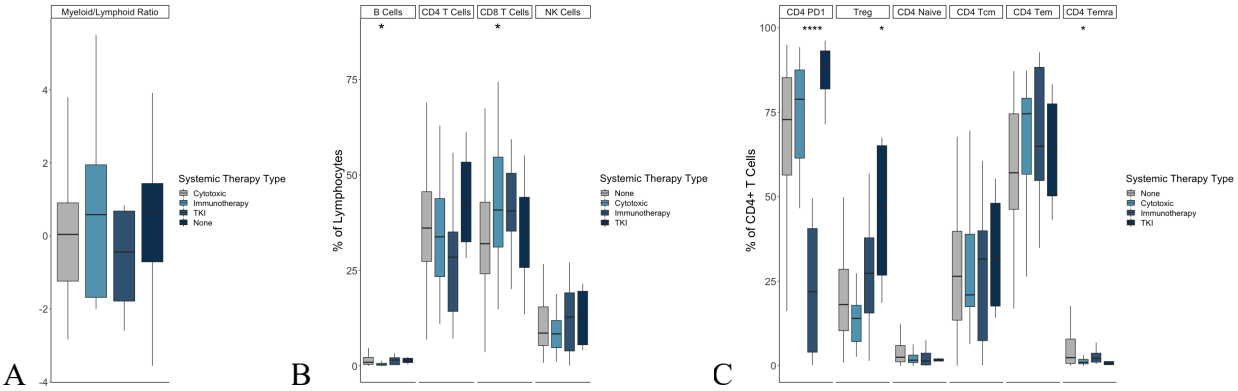


A B C



D E

Figure 2-16: The association between immune cell infiltration and XRT type in patients with STS. (A) The myeloid/lymphoid ratio by XRT type; values shown are $\log_2$(myeloid/lymphoid ratio). (B-D) relative abundances of lymphocyte populations (B), CD4+ T cell subsets (C), and CD8+ T cell subsets (D) as a percentage of the parent population, grouped by XRT type. (E) Relative abundance of myeloid subsets grouped by XRT type. Significance testing was performed using Wilcoxon test with "none" as the reference group. Using Bonferroni correction, statistical significance is defined as <0.002. *, P < 0.05; **, P < 0.01; ***, P < 0.001 ; ****, P < 0.0001.



A B C

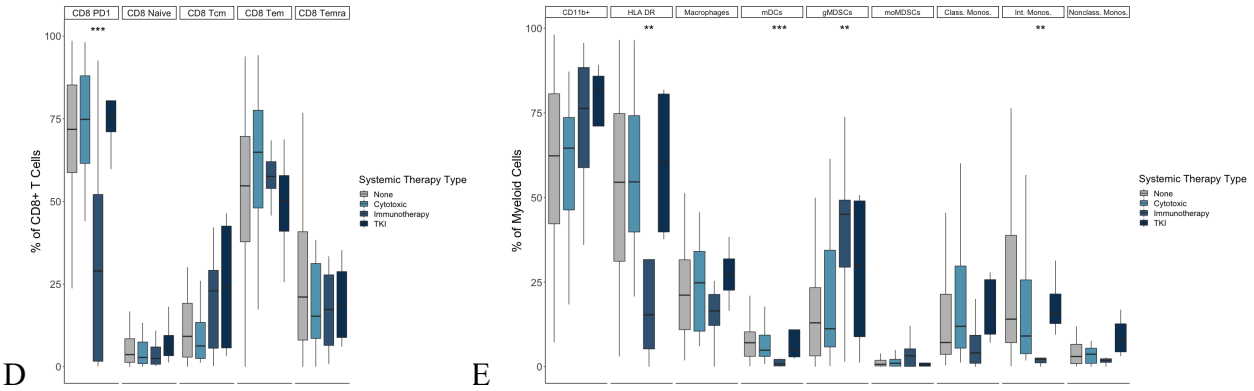
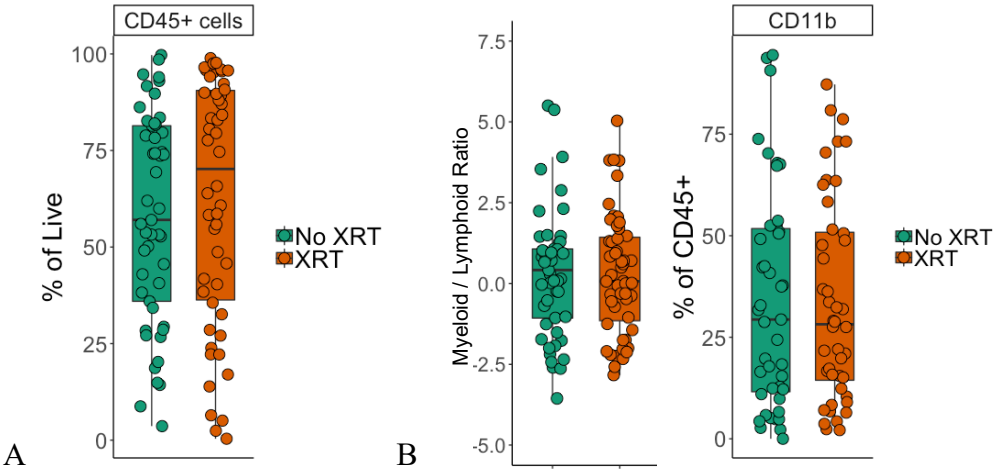


Figure 2-17: The association between immune cell infiltration and systemic therapy type in patients with STS. (A) The myeloid/lymphoid ratio by systemic therapy type; values shown are $\log_2$(myeloid/lymphoid ratio). (B-D) relative abundances of lymphocyte populations (B), CD4+ T cell subsets (C), and CD8+ T cell subsets (D) as a percentage of the parent population, grouped by systemic therapy type. (E) Relative abundance of myeloid subsets grouped by systemic therapy type. Significance testing was performed using Wilcoxon test with "none" as the reference group. Using Bonferroni correction, statistical significance is defined as <0.002. The association between immune cell infiltration and systemic therapy type. Treatment with anti-PD1 immunotherapy blocks the ability to detect PD1 on the surfaces of cells. Thus, the differences in PD1 demonstrated are likely due to technical artifact. *, P < 0.05; **, P < 0.01; ***, P < 0.001 ; ****, P < 0.0001.



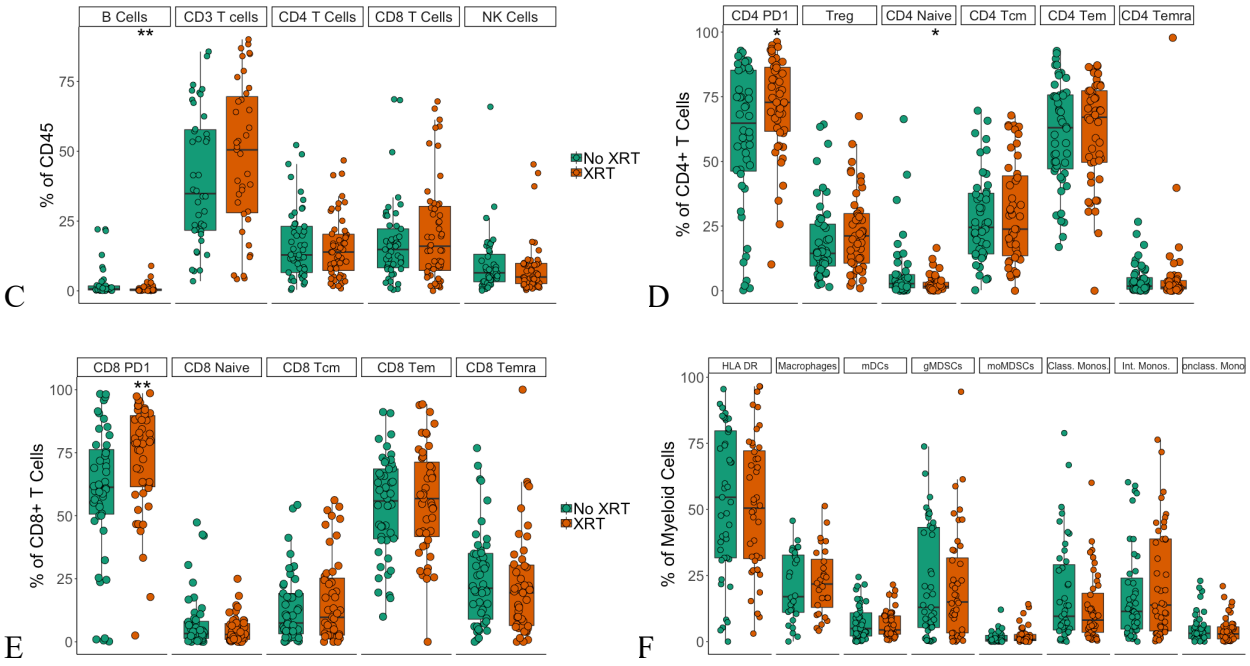


Figure 2-18: The association between immune cell infiltration and XRT status in patients with STS. (A) The proportion of CD45 (% of live) cells and (B) the myeloid/lymphoid ratio and CD11b quantities across XRT status in STS tumor samples. Values shown are log2(myeloid/lymphoid ratio). The relative abundances of various cell populations as a % of lymphocytes (C), % of CD4+ T cells (D), % of CD8+ T cells (E), % of myeloid cells (F) across XRT status. Significance testing was performed using Wilcoxon test with "no XRT" as the reference group. Using Bonferroni correction, statistical significance is defined as <0.002. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.

2.9.9 Additional analysis of the evolution of TME with time after XRT

Figure 2-6 demonstrates the analysis of the change in the intratumoral leukocyte population over time after the treatment with XRT. We found that the effects of radiotherapy on the tumor microenvironment evolve over time, and ultimately yield a CD8 predominant phenotype. Figure

2-19 below summarizes additional analyses performed. Specifically, we examined the association between the XRT-surgery interval (in days) and the proportion of CD45 (% of live) and CD11b (% of CD45) cells as well as the myeloid/lymphoid ratio (log2(value)). We also examined the association between treatment effect (%) and XRT-surgery interval. There was no association between the intratumoral leukocyte variables listed and XRT-Surgery interval. There was a trend toward a positive association (R = 0.27) between treatment effect with increased XRT-Surgery, however, this did not reach statistical significance (p = 0.062).

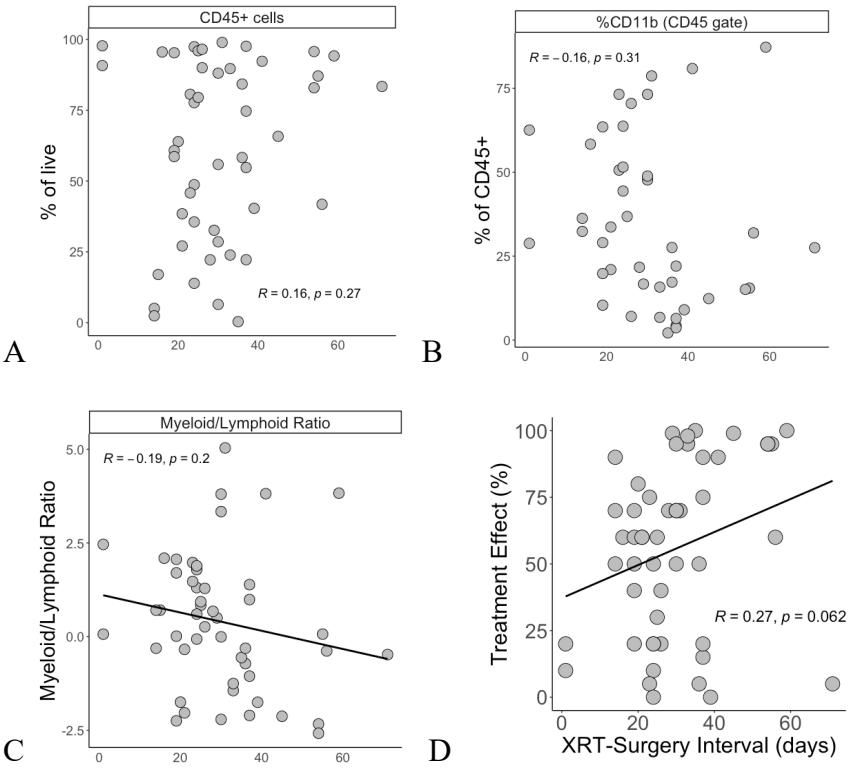


Figure 2-19: Summary of additional analysis examining the time-dependent effects of XRT. (A-C) Graphs depict the association between XRT-Surgery interval (days) and various intratumoral leukocyte populations: (A) CD45 (% of live), (B) CD11b (% of CD45), and (C) myeloid/lymphoid ratio (values depicted are log2 of the ratio). (D) The association between

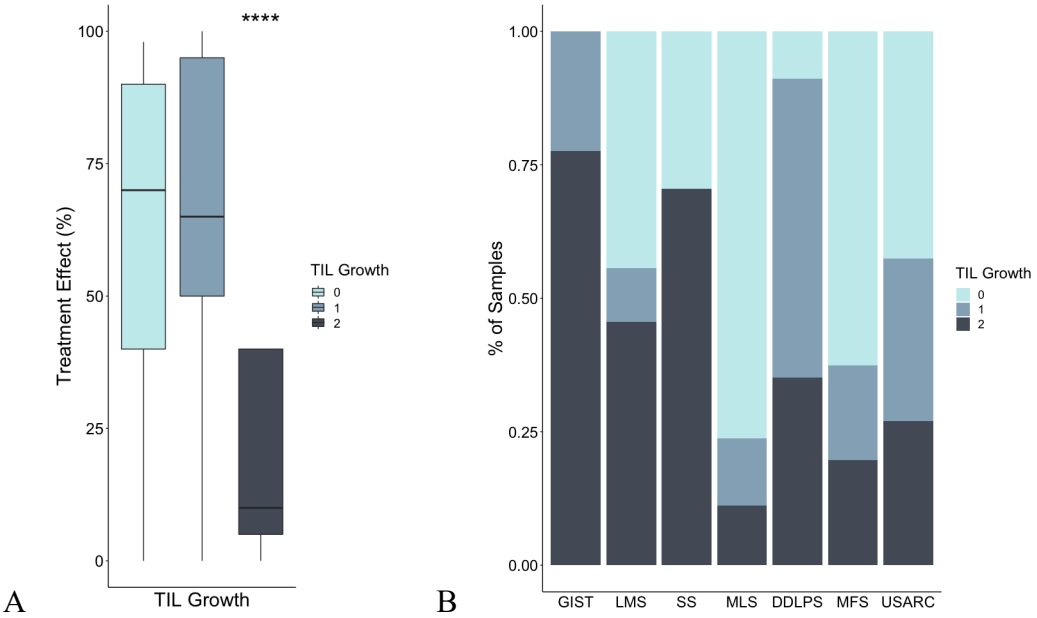
treatment effect (%) and XRT-surgery interval (days). Regression lines for continuous data demonstrate linear models fit to data.

2.9.10 Factors associated with TIL growth in patients with STS

Ex vivo tumor infiltrating lymphocyte cultures (TILs) were generated from patient tumor samples. Small tumor fragments (2-3mm^3) were collected from a larger tumor sample and placed individually in wells of a 24-well non-tissue culture treated plate with 2 mL of TIL media. TIL media consists of XVIVO with human IL2 at a concentration of 6000 U/mL TIL fragments were monitored three times a week (Monday, Wednesday, and Friday). On day five after plating, TIL media was changed for the first time. 1mL of media is removed carefully and 1mL of fresh media is added back. After day five, media was changed three times a week (Monday, Wednesday, and Friday). Once TILs were identified and began growing, they were transferred to additional wells in order to expand. TILs were collected and frozen in FBS with 5% DMSO once sufficiently expanded.

We explored the association between various clinical, tumor, and intratumoral leukocyte variables as demonstrated below. TIL growth was defined as 0, 1, and 2 (no growth, moderate growth, and strong growth, respectively). TIL growth was inversely associated with treatment effect. Specifically, samples with high TIL growth (Figure 2-20A). TIL growth was more effective from samples from patients that did not undergo neoadjuvant XRT. Figure 2-20B shows TIL growth from all samples, and Figures 2-21A & 2-21B demonstrate TIL growth from tumors from the subgroups of patients that did not and did undergo neoadjuvant XRT,

respectively. Finally, Figures 2-20C & 2-21C demonstrate the associations between various intratumoral leukocyte populations and TIL growth. Increased CD4 (% of CD45) and CD4PD1 (% of CD45) were associated with increased TIL growth.



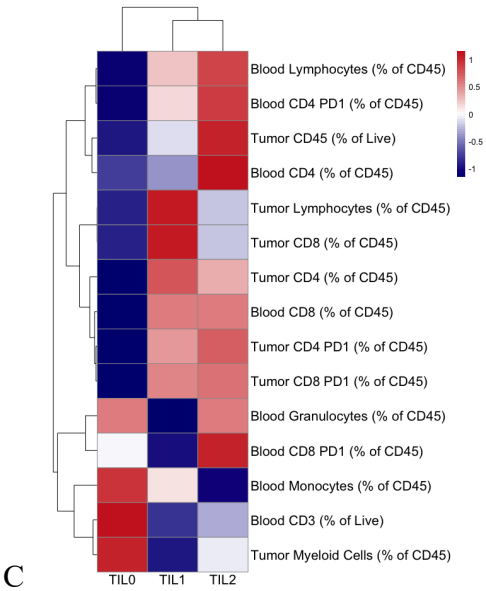
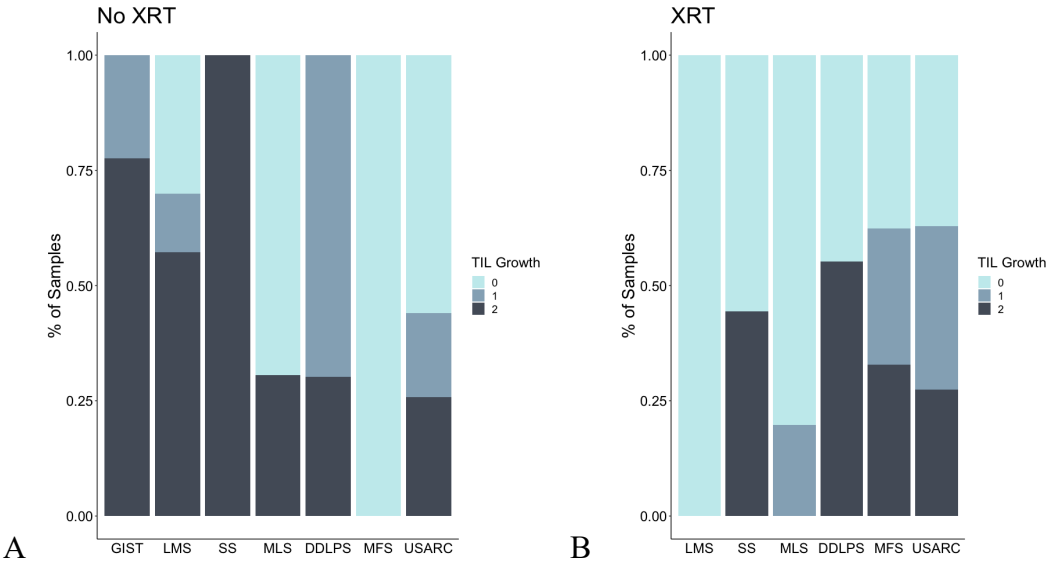


Figure 2-20: Factors associated with TIL growth in patients with STS. TIL growth is inversely associated with treatment effect. The highest TIL growth is seen in GIST tumors and the lowest in MLS. (A) The association between treatment effect (%) and TIL growth. (B) The association

between tumor histology and TIL growth. (C) Heatmap depicting the association between cell types in blood or tumor samples and TIL growth. Values are scaled across rows. Rows and columns are ordered based on unsupervised hierarchical clustering using Euclidean distances. Significance testing for categorical data was performed using Wilcoxon test with TIL growth 0 as the reference group. $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$. TIL Growth is defined as follows: 0, no growth; 1, moderate growth; 2, strong growth.



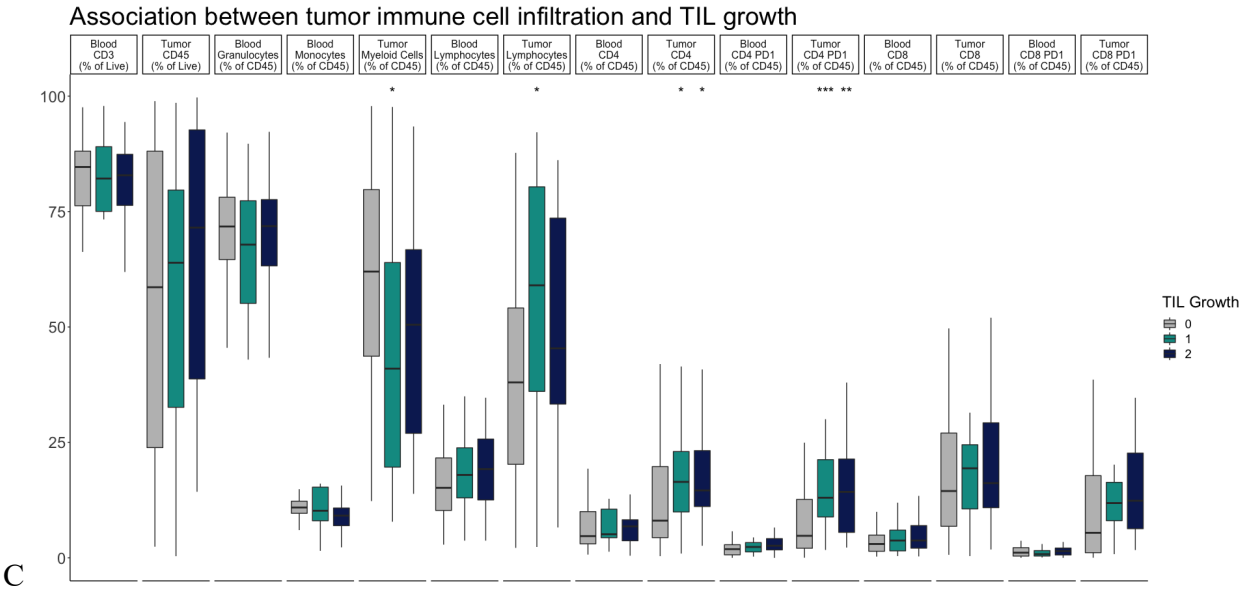


Figure 2-21: Factors associated with TIL growth in patients with STS. The association between tumor histology and TIL growth in patients who did not (A) and who did (B) undergo prior XRT. (C) The association between cell types in blood or tumor samples and TIL growth. Significance testing was performed using Wilcoxon test with TIL Growth 0 as the reference group. *, P < 0.05; **, P < 0.01; ***, P < 0.001 ; ****, P < 0.0001. TIL Growth is defined as follows: 0, no growth; 1, moderate growth; 2, strong growth.

2.9.11 Tables summarizing the data in the figures included in Chapter 2

		Treatment Effect (%)					
Tumor Size	*n*	*mean*	*median*	*min*	*max*	*sd*	*p value*
Total	105						0.003
<=5cm	28	25.00	7.50	0.00	100.00	32.54	Ref
5-10cm	42	48.04	50.00	0.00	100.00	32.95	0.042
>10cm	35	64.46	70.00	10.00	99.00	28.19	0.001

Table 2-9: The association between treatment effect and tumor size (cm). Kruskal-Wallis test and Wilcoxon tests were used to determine the overall and pairwise differences between the groups, respectively.

		Statistics for Cell Types Listed					
	n	*mean*	*median*	*min*	*max*	*sd*	*p*
All Patients							
MYELOID CELLS							
Total	105						0.119
GIST	4	32.50	30.49	14.08	54.94	18.00	0.039
LMS	13	43.43	40.98	16.35	73.21	18.94	0.051
SS	5	39.38	22.91	13.87	93.30	33.66	0.110
MLS	16	49.07	51.41	12.29	80.70	20.79	0.130
DDLPS	14	51.96	63.95	7.83	73.83	23.70	0.320
MFS	8	53.28	56.37	17.81	92.07	25.21	0.480
USARC	45	60.18	61.50	14.35	97.84	25.76	Ref
LYMPHOID CELLS							
Total	105						0.083
GIST	4	67.33	69.51	45.06	85.26	17.78	0.035
LMS	13	56.75	59.02	26.79	83.49	19.08	0.041
SS	5	60.62	77.09	6.69	86.13	33.67	0.096

MLS	16	50.72	48.59	19.30	87.71	20.72	0.100
DDLPS	14	48.32	36.05	26.17	92.17	23.43	0.223
MFS	8	45.46	40.76	7.93	82.19	25.25	0.462
USARC	45	39.03	34.76	2.16	85.65	25.67	Ref
MYELOID/LYMPHOID RATIO (log2 value)							
Total	105						0.099
GIST	4	-1.19	-1.23	-2.60	0.29	1.26	0.035
LMS	13	-0.43	-0.53	-2.35	1.45	1.21	0.043
SS	5	-0.59	-1.75	-2.63	3.80	2.68	0.103
MLS	16	-0.10	0.08	-2.83	2.06	1.39	0.112
DDLPS	14	0.00	0.83	-3.56	1.50	1.65	0.295
MFS	8	0.31	0.43	-2.21	3.54	1.84	0.493
USARC	45	0.93	0.74	-2.58	5.50	2.10	Ref
Subgroup (no neoadjuvant XRT)							
MYELOID CELLS							
Total	52						0.024
GIST	4	32.50	30.49	14.08	54.94	18.00	0.007
LMS	10	44.21	40.98	16.35	73.21	20.45	0.023
SS	2	14.75	14.75	13.87	15.63	1.24	0.013
MLS	4	57.06	58.30	45.00	66.64	10.57	0.249
DDLPS	12	49.07	63.95	7.83	73.83	24.65	0.089
MFS	3	53.10	46.44	20.80	92.07	36.10	0.359
USARC	17	68.55	67.26	23.04	97.84	21.51	Ref
LYMPHOID CELLS							
Total	52						0.009
GIST	4	67.33	69.51	45.06	85.26	17.78	0.003
LMS	10	56.01	59.02	26.79	83.49	20.62	0.012
SS	2	85.25	85.25	84.37	86.13	1.24	0.013
MLS	4	42.68	41.91	33.36	53.54	9.92	0.080
DDLPS	12	51.24	36.05	26.17	92.17	24.32	0.030
MFS	3	44.30	45.78	7.93	79.20	35.65	0.359
USARC	17	28.88	32.26	2.16	76.33	19.33	Ref
MYELOID/LYMPHOID RATIO (log2 value)							
Total	52						0.017
GIST	4	-1.19	-1.23	-2.60	0.29	1.26	0.005
LMS	10	-0.39	-0.53	-2.35	1.45	1.31	0.014
SS	2	-2.53	-2.53	-2.63	-2.43	0.14	0.013
MLS	4	0.43	0.48	-0.25	1.00	0.61	0.178
DDLPS	12	-0.20	0.83	-3.56	1.50	1.72	0.071
MFS	3	0.54	0.02	-1.93	3.54	2.77	0.359
USARC	17	1.67	1.05	-1.73	5.50	1.96	Ref

Table 2-10: Table summarizing the myeloid lineage, lymphoid lineage, and myeloid/lymphoid ratio (log2 values shown) in all STS samples in this study as well as in the subset of patients who

did not undergo neoadjuvant XRT. Kruskal-Wallis test and Wilcoxon tests were used to determine the overall and pairwise differences between the groups, respectively.

		Statistics for Cell Types Listed					
Tumor Size	*n*	*mean*	*median*	*min*	*max*	*sd*	*p value*
MYLEOID CELLS							
All STS Samples							
Total	105						0.904
<=5cm	28	50.50	50.68	7.83	97.65	27.28	Ref
5-10cm	42	53.46	61.25	13.87	93.42	22.86	0.730
>10cm	35	54.07	51.41	12.29	97.84	25.63	0.660
USARC Samples							
Total	45						0.257
<=5cm	13	54.74	49.19	17.41	97.65	28.32	Ref
5-10cm	17	56.06	61.25	14.35	93.42	26.66	0.980
>10cm	15	69.20	67.71	27.73	97.84	21.52	0.150
Non-USARC STS Samples							
Total	60						0.399
<=5cm	15	46.86	54.59	7.83	92.07	26.85	Ref
5-10cm	25	51.45	55.86	13.87	80.70	19.87	0.670
>10cm	20	42.72	43.92	12.29	93.30	22.76	0.620
LYMPHOID CELLS							
All STS Samples							
Total	105						0.840
<=5cm	28	49.67	49.32	2.35	92.17	27.27	Ref
5-10cm	42	45.73	38.75	6.58	86.13	22.77	0.650
>10cm	35	45.49	45.68	2.16	87.71	25.76	0.570
USARC Samples							
Total	45						0.222
<=5cm	13	45.52	52.19	2.35	82.59	28.24	Ref
5-10cm	17	42.61	38.75	6.58	85.65	26.46	0.880
>10cm	15	29.79	32.24	2.16	75.15	21.32	0.130
Non-USARC STS Samples							
Total	60						0.380
<=5cm	15	53.22	45.41	7.93	92.17	26.94	Ref
5-10cm	25	48.14	42.14	19.30	86.13	19.78	0.640
>10cm	20	57.26	54.98	6.69	87.71	22.63	0.640

MYELOID/LYMPHOID RATIO							
All STS Samples							
Total	105						0.877
<=5cm	28	0.10	0.04	-3.56	5.37	2.11	Ref
5-10cm	42	0.27	0.66	-2.63	3.83	1.61	0.680
>10cm	35	0.45	0.26	-2.83	5.50	2.09	0.640
USARC Samples							
Total	45						0.234
<=5cm	13	0.53	-0.08	-2.25	5.37	2.28	Ref
5-10cm	17	0.53	0.66	-2.58	3.83	1.97	0.910
>10cm	15	1.70	1.07	-1.44	5.50	2.00	0.140
Non-USARC STS Samples							
Total	60						0.405
<=5cm	15	-0.26	0.27	-3.56	3.54	1.97	Ref
5-10cm	25	0.08	0.35	-2.63	2.06	1.29	0.640
>10cm	20	-0.48	-0.32	-2.83	3.80	1.64	0.620

Table 2-11: Table summarizing the association between myeloid lineage, lymphoid lineage, and myeloid/lymphoid ratio (log2 values shown) and tumor size (cm). The associations were determined in the overall study cohort of all STS samples as well as in the subgroup of USARC versus non-USARC samples. Kruskal-Wallis test and Wilcoxon tests were used to determine the overall and pairwise differences between the groups, respectively.

		Statistics for Cell Types Listed					
Histology	*n*	*mean*	*median*	*min*	*max*	*sd*	*p value*
MYELOID AND LYMPHOID LINEAGES							
Lymphoid Cells (% of CD45)							
Total	105						0.083
GIST	4	67.33	69.51	45.06	85.26	17.78	0.035
LMS	13	56.75	59.02	26.79	83.49	19.08	0.041
SS	5	60.62	77.09	6.69	86.13	33.67	0.096
MLS	16	50.72	48.59	19.30	87.71	20.72	0.100
DDLPS	14	48.32	36.05	26.17	92.17	23.43	0.223
MFS	8	45.46	40.76	7.93	82.19	25.25	0.462
USARC	45	39.03	34.76	2.16	85.65	25.67	Ref

Myeloid Cells (% of CD45)							
Total	105						0.119
GIST	4	32.50	30.49	14.08	54.94	18.00	0.039
LMS	13	43.43	40.98	16.35	73.21	18.94	0.051
SS	5	39.38	22.91	13.87	93.30	33.66	0.111
MLS	16	49.07	51.41	12.29	80.70	20.79	0.133
DDLPS	14	51.96	63.95	7.83	73.83	23.70	0.322
MFS	8	53.28	56.37	17.81	92.07	25.21	0.477
USARC	45	60.18	61.50	14.35	97.84	25.76	Ref
Myeloid/Lymphoid Ratio (log2 value)							
Total	105						0.099
GIST	4	-1.19	-1.23	-2.60	0.29	1.26	0.035
LMS	13	-0.43	-0.53	-2.35	1.45	1.21	0.043
SS	5	-0.59	-1.75	-2.63	3.80	2.68	0.103
MLS	16	-0.10	0.08	-2.83	2.06	1.39	0.112
DDLPS	14	0.00	0.83	-3.56	1.50	1.65	0.295
MFS	8	0.31	0.43	-2.21	3.54	1.84	0.493
USARC	45	0.93	0.74	-2.58	5.50	2.10	Ref
LYMPHOCYTE LINEAGE							
B Cells (% of lymphocytes)							
Total	105						0.453
GIST	4	1.71	2.31	0.27	2.56	1.25	0.232
LMS	13	1.77	0.80	0.11	6.42	2.32	0.445
SS	5	5.95	1.39	0.00	26.44	11.48	0.677
MLS	16	1.80	1.11	0.07	6.65	1.95	0.121
DDLPS	14	6.21	1.38	0.15	21.72	8.27	0.057
MFS	8	1.35	0.93	0.01	3.58	1.47	0.777
USARC	45	0.88	0.54	0.00	4.64	0.99	Ref
CD4 T Cells (% of lymphocytes)							
Total	105						0.342
GIST	4	41.09	35.44	32.20	61.26	13.55	0.283
LMS	13	41.27	43.86	27.76	62.98	10.58	0.065
SS	5	25.96	24.17	14.93	48.36	13.42	0.293
MLS	16	34.90	38.75	9.75	56.50	14.52	0.547
DDLPS	14	36.96	37.53	19.08	55.38	13.13	0.351
MFS	8	37.19	35.69	6.86	69.02	17.87	0.558
USARC	45	32.76	32.39	7.10	59.24	14.14	Ref
CD8 T Cells(% of lymphocytes)							
Total	105						0.122

GIST	4	21.51	20.82	9.53	34.88	11.99	0.031
LMS	13	36.01	37.10	23.17	48.64	7.61	0.540
SS	5	51.17	55.35	14.86	79.35	25.23	0.375
MLS	16	33.87	30.14	3.59	74.44	19.35	0.150
DDLPS	14	31.73	32.01	15.04	48.34	9.97	0.093
MFS	8	40.53	39.47	12.35	86.66	21.26	0.852
USARC	45	41.92	38.45	16.32	84.85	17.91	Ref
NK Cells (% of lymphocytes)							
Total	105						0.043
GIST	4	28.42	27.09	18.88	39.29	10.27	0.006
LMS	13	13.53	12.23	2.07	33.30	8.92	0.097
SS	5	6.53	4.72	1.04	12.64	5.74	0.519
MLS	16	15.83	12.69	3.33	44.72	11.71	0.048
DDLPS	14	10.88	7.46	0.07	36.12	9.77	0.711
MFS	8	10.60	9.59	1.68	20.99	7.35	0.559
USARC	45	9.42	8.05	0.75	46.01	8.70	Ref
CD4 T CELLS							
CD4PD1 (% of CD4)							
Total	105						0.210
GIST	4	58.65	74.58	0.20	85.22	39.37	0.863
LMS	13	56.34	61.97	3.88	92.11	28.75	0.375
SS	5	77.49	89.94	55.79	93.52	19.74	0.211
MLS	16	60.01	64.79	10.95	90.72	23.15	0.333
DDLPS	14	68.04	68.02	30.63	93.08	19.05	0.925
MFS	8	81.54	83.22	53.50	96.19	13.90	0.042
USARC	45	65.05	71.01	0.97	94.87	25.34	Ref
Treg (% of CD4)							
Total							<0.001
GIST	4	16.41	18.62	11.01	19.61	4.70	0.239
LMS	13	17.33	14.62	4.95	45.81	11.79	0.036
SS	5	11.96	13.10	5.92	15.12	3.52	0.013
MLS	16	7.13	7.53	0.95	14.89	4.04	<0.001
DDLPS	14	15.64	11.98	2.23	38.23	10.56	0.015
MFS	8	33.26	29.20	8.61	67.49	17.61	0.344
USARC	45	27.53	25.63	1.47	64.30	16.21	Ref
CD4 Naïve (% of CD4)							
Total							0.372
GIST	4	2.17	1.34	0.15	5.84	2.63	0.735
LMS	13	7.53	3.17	0.33	44.85	12.93	0.179
SS	5	5.29	1.31	0.78	14.39	6.08	0.459

MLS	16	9.08	2.41	0.66	66.34	17.15	0.060
DDLPS	14	6.07	2.59	0.63	35.07	9.63	0.164
MFS	8	3.29	2.02	1.35	12.33	3.71	0.351
USARC	45	3.25	1.40	0.00	21.67	4.60	Ref
CD4 Tcm (% of CD4)							
Total							0.650
GIST	4	25.77	22.44	4.71	53.49	20.50	0.803
LMS	13	30.27	33.18	9.08	69.61	16.88	0.565
SS	5	39.64	31.50	21.72	63.91	20.21	0.186
MLS	16	29.02	17.05	0.00	67.77	25.23	0.767
DDLPS	14	24.62	21.21	4.56	64.00	16.86	0.622
MFS	8	37.94	45.17	6.02	63.42	22.05	0.175
USARC	45	27.09	23.76	0.20	60.92	15.87	Ref
CD4 Tem (% of CD4)							
Total							0.136
GIST	4	69.97	70.92	46.28	91.76	18.98	0.832
LMS	13	58.05	57.29	16.91	87.24	22.04	0.318
SS	5	54.17	66.10	26.39	76.65	22.52	0.224
MLS	16	50.25	54.78	0.00	81.70	26.64	0.032
DDLPS	14	64.63	73.62	29.82	87.32	19.28	0.793
MFS	8	50.33	49.63	22.24	86.86	18.02	0.022
USARC	45	66.48	67.07	34.86	92.71	15.92	Ref
CD4 Temra (% of CD4)							
Total							0.149
GIST	4	2.10	2.45	0.08	3.41	1.52	0.921
LMS	13	4.14	0.56	0.34	26.65	7.98	0.591
SS	5	0.90	1.09	0.15	1.49	0.53	0.323
MLS	16	11.64	4.23	0.00	97.79	25.17	0.061
DDLPS	14	4.68	3.38	0.56	17.71	4.98	0.083
MFS	8	8.45	2.95	0.22	39.76	13.38	0.250
USARC	45	3.18	1.27	0.00	21.94	4.62	Ref
CD8 T CELLS							
CD8PD1 (% of CD8)							
Total							0.073
GIST	4	55.01	65.72	0.14	88.46	38.43	0.452
LMS	13	54.45	56.40	0.95	91.43	27.20	0.080
SS	5	75.03	82.52	49.96	97.00	21.87	0.560
MLS	16	64.44	63.40	17.77	98.15	23.46	0.599
DDLPS	14	65.15	64.86	24.58	90.66	16.25	0.338
MFS	8	80.35	84.38	50.04	98.59	18.42	0.175

USARC	45	68.19	76.09	0.54	97.35	25.55	Ref
CD8 Naïve (% of CD8)							
Total							0.072
GIST	4	2.39	1.03	0.11	7.38	3.37	0.563
LMS	13	8.18	3.89	1.11	42.11	11.86	0.080
SS	5	11.61	6.55	0.57	42.52	17.56	0.306
MLS	16	11.99	7.75	0.00	47.33	13.06	0.008
DDLPS	14	6.78	2.99	0.30	30.46	8.81	0.198
MFS	8	6.68	4.91	0.28	18.21	6.15	0.120
USARC	45	3.35	2.51	0.00	13.80	3.42	Ref
CD8 Tcm (% of CD8)							
Total							0.461
GIST	4	11.50	10.31	0.50	24.90	11.58	0.536
LMS	13	13.71	10.62	1.92	52.88	14.03	0.982
SS	5	11.12	10.46	3.27	19.90	7.57	0.832
MLS	16	13.42	2.77	0.00	54.31	18.84	0.301
DDLPS	14	10.34	4.93	0.41	52.19	14.47	0.389
MFS	8	24.05	19.17	2.46	56.18	19.03	0.172
USARC	45	15.87	10.78	0.00	53.58	15.08	Ref
CD8 Tem (% of CD8)							
Total							0.021
GIST	4	66.61	68.64	54.86	74.31	8.28	0.475
LMS	13	49.88	46.47	18.42	80.49	22.81	0.129
SS	5	60.60	53.03	35.58	90.67	22.06	0.706
MLS	16	38.08	31.45	0.00	72.84	22.75	0.001
DDLPS	14	51.35	49.82	29.64	73.04	14.22	0.054
MFS	8	50.30	50.21	25.56	91.12	23.52	0.167
USARC	45	62.18	63.13	25.21	94.22	17.05	Ref
CD8 Temra (% of CD8)							
Total							0.032
GIST	4	19.50	23.52	0.69	30.28	13.19	0.631
LMS	13	28.22	26.32	7.19	76.84	21.55	0.089
SS	5	16.67	11.44	4.48	36.24	13.58	0.918
MLS	16	36.52	32.20	0.00	100.00	25.40	0.006
DDLPS	14	31.52	29.91	6.73	63.89	16.23	0.006
MFS	8	18.96	14.56	1.49	47.80	17.58	0.947
USARC	45	18.61	14.42	0.00	69.83	17.70	Ref
MYELOID LINEAGE							
CD11B (% of myeloid)							
Total							0.003

GIST	4	49.42	47.95	35.98	64.31	14.22	0.108
LMS	13	65.58	68.33	43.71	86.34	14.76	0.274
SS	5	40.14	47.72	19.69	54.87	17.00	0.005
MLS	16	42.67	45.29	7.21	70.66	22.59	0.001
DDLPS	14	55.31	64.29	0.09	80.65	25.68	0.064
MFS	8	68.32	79.49	23.82	84.76	23.72	0.855
USARC	45	69.85	77.60	18.40	98.14	22.12	Ref
HLA DR (% of myeloid)							
Total							0.012
GIST	4	65.23	80.16	31.68	83.85	29.12	0.274
LMS	13	57.53	55.42	27.21	80.63	20.02	0.120
SS	5	65.24	64.23	38.86	85.40	17.47	0.101
MLS	16	37.76	32.41	9.26	84.56	22.34	0.557
DDLPS	14	67.61	71.32	0.00	89.90	25.84	0.015
MFS	8	71.27	73.85	30.75	95.55	21.71	0.036
USARC	45	44.99	42.53	3.03	96.60	27.21	Ref
Macrophages (% of myeloid)							
Total							0.115
GIST	4	25.34	25.34	24.63	26.04	1.00	0.870
LMS	13	29.96	33.94	1.86	45.73	13.38	0.349
SS	5	17.43	19.55	3.16	34.07	12.88	0.308
MLS	16	12.82	10.51	4.10	24.79	7.93	0.017
DDLPS	14	17.31	13.94	0.05	34.47	13.44	0.168
MFS	8	17.80	16.54	5.20	33.61	10.66	0.374
USARC	45	24.92	23.30	9.14	51.29	12.78	Ref
mDCs (% of myeloid)							
Total							0.006
GIST	4	19.18	21.62	11.48	24.43	6.81	0.001
LMS	13	5.82	4.60	0.99	13.59	3.75	0.154
SS	5	6.52	6.31	0.77	11.54	4.49	0.241
MLS	16	8.83	9.66	1.84	19.37	5.73	0.006
DDLPS	14	10.07	10.69	0.00	19.80	6.50	0.012
MFS	8	7.51	6.05	0.15	20.99	7.37	0.374
USARC	45	4.69	3.07	0.08	21.53	4.72	Ref
gMDSCs (% of myeloid)							
Total							0.563
GIST	4	11.04	5.36	1.16	26.60	13.64	0.457
LMS	13	24.61	21.51	0.52	54.67	20.29	0.910
SS	5	14.14	10.19	6.08	34.43	11.50	0.449
MLS	16	21.91	17.47	1.65	58.79	19.47	0.856
DDLPS	14	12.65	6.86	0.21	45.04	15.88	0.126

MFS	8	13.21	13.72	0.49	21.90	7.28	0.393
USARC	45	26.10	20.59	0.01	94.54	24.27	Ref
moMDSCs (% of myeloid)							
Total							0.008
GIST	4	1.18	0.19	0.14	3.22	1.76	0.329
LMS	13	0.81	0.40	0.00	3.82	1.18	0.009
SS	5	1.13	0.22	0.03	4.95	2.14	0.061
MLS	16	1.62	0.41	0.00	13.07	3.51	0.016
DDLPS	14	0.62	0.32	0.00	2.15	0.70	0.003
MFS	8	0.61	0.68	0.02	1.28	0.46	0.035
USARC	45	3.12	2.35	0.00	14.16	3.41	Ref
Classical Monocytes (% of myeloid)							
Total							0.085
GIST	4	17.80	21.42	4.09	27.89	12.30	0.345
LMS	13	24.10	19.71	3.27	66.81	20.81	0.069
SS	5	15.08	11.94	3.64	37.91	13.73	0.519
MLS	16	6.84	5.53	0.82	29.75	7.65	0.225
DDLPS	14	19.88	12.66	0.00	50.94	19.59	0.328
MFS	8	23.52	28.25	4.84	36.98	13.81	0.047
USARC	45	13.07	7.12	0.34	78.89	16.31	Ref
Intermediate Monocytes (% of myeloid)							
Total							0.540
GIST	4	16.37	13.85	8.61	26.65	9.28	0.906
LMS	13	11.39	9.42	1.08	33.08	9.57	0.309
SS	5	17.48	5.73	4.95	54.39	21.22	0.840
MLS	16	15.17	10.89	0.50	53.39	17.35	0.474
DDLPS	14	29.07	27.37	0.00	76.39	25.36	0.339
MFS	8	28.16	21.46	7.11	60.30	22.27	0.305
USARC	45	20.20	14.28	0.12	71.72	19.31	Ref
Nonclassical Monocytes (% of myeloid)							
Total							
GIST	4	7.84	3.12	1.28	19.12	9.81	0.345
LMS	13	7.44	7.00	0.24	20.40	6.66	0.058
SS	5	4.91	3.03	0.00	13.59	5.46	0.495
MLS	16	5.51	3.66	0.50	17.01	5.00	0.061
DDLPS	14	5.08	3.21	0.00	22.99	6.56	0.425
MFS	8	7.19	6.34	0.09	16.86	6.03	0.072
USARC	45	2.88	1.91	0.04	20.98	3.56	Ref

Table 2-12: Table summarizing the association between immune cell infiltration and histology of STS tumor samples. Kruskal-Wallis test and Wilcoxon tests were used to determine the

overall and pairwise differences between the groups, respectively. USARC was the reference group for all pairwise comparisons.

		Statistics for Cell Types Listed					
Treatment status	*n*	*mean*	*median*	*min*	*max*	*sd*	*p value*
MYELOID AND LYMPHOID LINEAGES							
Lymphoid Cells (% of CD45)							
Untreated	31	46.73	41.68	2.35	92.17	25.78	0.979
Treated	74	46.64	46.60	2.16	87.71	24.68	
Myeloid Cells (% of CD45)							
Untreated	31	52.29	56.04	7.83	97.65	25.94	0.851
Treated	74	53.14	51.95	12.29	97.84	24.56	
Myeloid/Lymphoid Ratio (log2 value)							
Untreated	31	0.28	0.35	-3.56	5.37	2.05	0.905
Treated	74	0.30	0.20	-2.83	5.50	1.86	
LYMPHOCYTE LINEAGE							
B Cells (% of lymphocytes)							
Untreated	31	3.74	1.43	0.01	21.72	5.79	0.009
Treated	74	1.62	0.52	0.00	26.44	3.86	
CD4 T Cells (% of lymphocytes)							
Untreated	31	36.20	36.09	6.86	59.24	13.55	0.463
Treated	74	34.40	34.80	7.10	69.02	14.30	
CD8 T Cells(% of lymphocytes)							
Untreated	31	34.02	29.26	9.53	86.66	18.12	0.038
Treated	74	40.01	38.85	3.59	84.85	17.10	
NK Cells (% of lymphocytes)							
Untreated	31	12.38	8.77	0.75	46.01	11.14	0.769
Treated	74	11.26	8.46	0.07	44.72	9.26	
CD4 T CELLS							
CD4PD1 (% of CD4)							
Untreated	31	65.13	68.17	10.95	92.81	22.55	0.707
Treated	74	65.58	71.21	0.20	96.19	25.52	
Treg (% of CD4)							
Untreated	31	18.31	13.22	2.23	63.37	13.29	0.328
Treated	74	22.20	18.55	0.95	67.49	16.20	
CD4 Naïve (% of CD4)							
Untreated	31	9.83	3.00	0.15	66.34	15.57	0.017
Treated	74	3.12	1.92	0.00	18.09	3.78	
CD4 Tcm (% of CD4)							
Untreated	31	26.98	26.09	4.38	60.92	15.14	0.751

Treated	74	29.73	23.79	0.00	69.61	19.76	
CD4 Tem (% of CD4)							
Untreated	31	56.91	56.88	16.91	84.25	18.25	0.124
Treated	74	62.65	68.79	0.00	92.71	20.79	
CD4 Temra (% of CD4)							
Untreated	31	6.28	3.99	0.08	26.65	7.09	0.053
Treated	74	4.50	1.40	0.00	97.79	12.74	
CD8 T CELLS							
CD8PD1 (% of CD8)							
Untreated	31	67.31	68.95	23.68	98.30	20.33	0.833
Treated	74	66.17	73.10	0.14	98.59	26.13	
CD8 Naïve (% of CD8)							
Untreated	31	8.25	3.68	0.00	47.33	12.27	0.366
Treated	74	5.51	2.81	0.00	42.52	7.20	
CD8 Tcm (% of CD8)							
Untreated	31	10.60	7.11	0.00	29.40	8.75	0.438
Treated	74	16.49	9.46	0.00	56.18	17.11	
CD8 Tem (% of CD8)							
Untreated	31	53.24	56.87	9.95	91.12	21.79	0.726
Treated	74	55.61	55.97	0.00	94.22	20.00	
CD8 Temra (% of CD8)							
Untreated	31	27.92	22.09	0.69	76.84	20.97	0.220
Treated	74	22.39	19.72	0.00	100.00	19.19	
MYELOID LINEAGE							
CD11B (% of myeloid)							
Untreated	31	56.14	58.04	14.50	97.65	24.03	0.256
Treated	74	62.54	65.22	0.09	98.14	23.64	
HLA DR (% of myeloid)							
Untreated	31	61.14	68.24	8.27	95.55	25.10	0.066
Treated	74	48.72	48.53	0.00	96.60	26.58	
Macrophages (% of myeloid)							
Untreated	31	17.35	13.93	1.86	35.68	11.88	0.145
Treated	74	23.36	23.30	0.05	51.29	12.82	
mDCs (% of myeloid)							
Untreated	31	8.37	8.01	0.15	21.62	6.80	0.345
Treated	74	6.45	4.24	0.00	24.43	5.60	
gMDSCs (% of myeloid)							
Untreated	31	16.68	10.23	0.09	54.67	16.85	0.256
Treated	74	23.13	17.38	0.01	94.54	21.84	
moMDSCs (% of myeloid)							
Untreated	31	1.16	0.63	0.00	5.08	1.46	0.231
Treated	74	2.23	0.80	0.00	14.16	3.25	
Classical Monocytes (% of myeloid)							

Untreated	31	20.82	13.65	0.49	78.89	20.05	0.068
Treated	74	13.26	6.97	0.00	66.81	14.48	
Intermediate Monocytes (% of myeloid)							
Untreated	31	21.68	12.06	1.07	60.30	19.98	0.312
Treated	74	19.25	11.53	0.00	76.39	19.25	
Nonclassical Monocytes (% of myeloid)							
Untreated	31	5.44	3.18	0.04	22.99	6.44	0.823
Treated	74	4.39	3.01	0.00	20.98	4.82	

Table 2-13: Table summarizing the association between immune cell infiltration and treatment status in STS tumor samples. Treatment was defined by the receipt of neoadjuvant chemotherapy or radiation. Wilcoxon test was used to determine the differences between the groups.

		Statistics for Cell Types Listed					
XRT Fractionation	*n*	*mean*	*median*	*min*	*max*	*sd*	*p value*
MYELOID AND LYMPHOID LINEAGES							
Lymphoid Cells (% of CD45)							
Total	105						0.626
Hypofractionated (5-6Gy/fx)	33	44.43	38.50	2.96	87.71	24.51	0.776
Conventional (2-3.5Gy/fx)	20	50.82	53.33	6.58	85.65	26.26	0.419
None	52	46.39	40.50	2.16	92.17	24.81	Ref
Myeloid Cells (% of CD45)							
Total	105						0.652
Hypofractionated (5-6Gy/fx)	33	55.57	61.50	12.29	96.96	24.40	0.644
Conventional (2-3.5Gy/fx)	20	49.30	46.67	14.35	93.42	26.13	0.550
None	52	52.68	54.94	7.83	97.84	24.87	Ref
Myeloid/Lymphoid Ratio (log2 value)							
Total	105						0.644
Hypofractionated (5-6Gy/fx)	33	0.43	0.68	-2.83	5.03	1.83	0.709
Conventional (2-3.5Gy/fx)	20	0.07	-0.19	-2.58	3.83	1.97	0.491
None	52	0.30	0.42	-3.56	5.50	1.96	Ref
LYMPHOCYTE LINEAGE							
B Cells (% of lymphocytes)							
Total	105						0.008
Hypofractionated (5-6Gy/fx)	33	0.91	0.50	0.00	6.65	1.40	0.005
Conventional (2-3.5Gy/fx)	20	0.99	0.28	0.00	4.64	1.29	0.024
None	52	3.48	1.42	0.01	26.44	6.07	Ref
CD4 T Cells (% of lymphocytes)							
Total	105						0.683
Hypofractionated (5-6Gy/fx)	33	36.41	37.03	9.75	69.02	15.04	0.677

Conventional (2-3.5Gy/fx)	20	32.89	33.08	7.10	62.98	15.10	0.491
None	52	34.81	35.17	6.86	61.26	13.14	Ref
CD8 T Cells(% of lymphocytes)							
Total	105						0.015
Hypofractionated (5-6Gy/fx)	33	36.39	32.85	3.59	74.85	17.43	0.072
Conventional (2-3.5Gy/fx)	20	48.25	47.40	20.13	84.85	17.11	0.752
None	52	35.44	32.85	9.53	86.66	16.57	Ref
NK Cells (% of lymphocytes)							
Total	105						0.036
Hypofractionated (5-6Gy/fx)	33	11.77	8.15	1.34	44.72	10.98	0.293
Conventional (2-3.5Gy/fx)	20	6.73	6.98	0.87	13.22	4.23	0.009
None	52	13.48	12.03	0.07	46.01	10.16	Ref
CD4 T CELLS							
CD4PD1 (% of CD4)							
Total	105						0.120
Hypofractionated (5-6Gy/fx)	33	72.82	76.09	25.69	96.19	18.38	0.047
Conventional (2-3.5Gy/fx)	20	69.62	70.79	10.18	93.52	20.00	0.258
None	52	59.57	64.77	0.20	92.81	28.06	Ref
Treg (% of CD4)							
Total	105						0.296
Hypofractionated (5-6Gy/fx)	33	24.89	22.05	1.92	67.49	17.70	0.190
Conventional (2-3.5Gy/fx)	20	20.34	19.87	0.95	45.81	11.27	0.267
None	52	19.21	14.44	1.47	64.30	15.53	Ref
CD4 Naïve (% of CD4)							
Total	105						0.066
Hypofractionated (5-6Gy/fx)	33	3.50	1.96	0.00	16.54	3.98	0.287
Conventional (2-3.5Gy/fx)	20	1.98	1.28	0.10	8.79	2.00	0.024
None	52	7.16	2.74	0.00	66.34	12.35	Ref
CD4 Tcm (% of CD4)							
Total	105						0.928
Hypofractionated (5-6Gy/fx)	33	30.47	28.74	0.00	67.77	21.39	0.780
Conventional (2-3.5Gy/fx)	20	29.84	23.77	6.37	63.91	17.88	0.764
None	52	27.72	24.50	0.20	69.61	17.35	Ref
CD4 Tem (% of CD4)							
Total	105						0.520
Hypofractionated (5-6Gy/fx)	33	58.02	59.00	0.00	86.86	22.01	0.705
Conventional (2-3.5Gy/fx)	20	65.76	68.89	33.73	87.15	17.28	0.352
None	52	60.73	63.03	16.91	92.71	20.26	Ref
CD4 Temra (% of CD4)							
Total	105						0.701
Hypofractionated (5-6Gy/fx)	33	8.01	2.09	0.00	97.79	19.61	0.691
Conventional (2-3.5Gy/fx)	20	2.42	1.23	0.32	16.77	3.59	0.403
None	52	4.39	1.77	0.00	26.65	5.82	Ref

CD8 T CELLS							
CD8PD1 (% of CD8)							
Total	105						0.021
Hypofractionated (5-6Gy/fx)	33	74.44	79.28	17.77	98.59	19.32	0.013
Conventional (2-3.5Gy/fx)	20	71.39	78.83	2.50	94.85	24.44	0.050
None	52	59.98	61.30	0.14	98.30	25.78	Ref
CD8 Naïve (% of CD8)							
Total	105						0.365
Hypofractionated (5-6Gy/fx)	33	5.78	3.89	0.00	25.00	6.27	0.812
Conventional (2-3.5Gy/fx)	20	3.42	2.56	0.24	11.67	3.38	0.180
None	52	7.78	3.28	0.00	47.33	11.31	Ref
CD8 Tcm (% of CD8)							
Total	105						0.835
Hypofractionated (5-6Gy/fx)	33	18.98	6.60	0.00	56.18	20.46	0.643
Conventional (2-3.5Gy/fx)	20	13.34	10.49	1.21	42.17	10.78	0.617
None	52	13.08	7.46	0.00	54.31	13.46	Ref
CD8 Tem (% of CD8)							
Total	105						0.020
Hypofractionated (5-6Gy/fx)	33	48.97	43.79	0.00	93.86	21.85	0.315
Conventional (2-3.5Gy/fx)	20	65.79	65.13	29.18	94.22	16.21	0.033
None	52	53.77	55.87	9.95	91.12	19.82	Ref
CD8 Temra (% of CD8)							
Total	105						0.371
Hypofractionated (5-6Gy/fx)	33	26.27	21.09	0.00	100.00	24.48	0.774
Conventional (2-3.5Gy/fx)	20	17.45	14.50	0.79	42.78	12.57	0.138
None	52	25.37	21.25	0.00	76.84	19.03	Ref
MYELOID LINEAGE							
CD11B (% of myeloid)							
Total	105						0.456
Hypofractionated (5-6Gy/fx)	33	64.68	68.32	18.40	92.78	21.78	0.427
Conventional (2-3.5Gy/fx)	20	56.86	59.04	7.21	98.14	24.54	0.546
None	52	60.05	64.31	0.09	97.65	24.81	Ref
HLA DR (% of myeloid)							
Total	105						0.816
Hypofractionated (5-6Gy/fx)	33	52.96	50.40	10.55	96.60	25.84	0.878
Conventional (2-3.5Gy/fx)	20	48.75	50.50	3.03	94.58	26.14	0.536
None	52	53.07	54.60	0.00	95.55	27.78	Ref
Macrophages (% of myeloid)							
Total	105						0.720
Hypofractionated (5-6Gy/fx)	33	24.61	22.11	5.10	51.29	14.23	0.488
Conventional (2-3.5Gy/fx)	20	20.49	21.45	4.10	37.98	10.37	0.945
None	52	20.80	16.99	0.05	45.73	13.12	Ref
mDCs (% of myeloid)							
Total	105						0.213

Hypofractionated (5-6Gy/fx)	33	7.97	5.63	0.61	21.53	5.91	0.413
Conventional (2-3.5Gy/fx)	20	4.75	3.53	0.77	11.54	3.45	0.367
None	52	7.32	4.95	0.00	24.43	6.69	Ref
gMDSCs (% of myeloid)							
Total	105						0.636
Hypofractionated (5-6Gy/fx)	33	18.25	16.04	0.01	58.79	17.90	0.601
Conventional (2-3.5Gy/fx)	20	24.52	14.91	1.50	94.54	23.93	0.515
None	52	21.90	12.96	0.09	73.82	21.07	Ref
moMDSCs (% of myeloid)							
Total	105						0.537
Hypofractionated (5-6Gy/fx)	33	2.72	0.64	0.00	14.16	4.20	0.376
Conventional (2-3.5Gy/fx)	20	1.71	1.36	0.02	5.28	1.62	0.380
None	52	1.55	0.65	0.00	12.07	2.27	Ref
Classical Monocytes (% of myeloid)							
Total	105						0.571
Hypofractionated (5-6Gy/fx)	33	12.66	6.50	0.38	60.10	13.83	0.315
Conventional (2-3.5Gy/fx)	20	12.55	9.59	0.91	37.91	11.55	0.536
None	52	18.14	9.57	0.00	78.89	19.23	Ref
Intermediate Monocytes (% of myeloid)							
Total	105						0.667
Hypofractionated (5-6Gy/fx)	33	23.21	19.66	0.50	76.39	20.16	0.378
Conventional (2-3.5Gy/fx)	20	20.21	10.77	0.12	71.72	22.43	0.931
None	52	17.80	11.43	0.00	60.30	17.64	Ref
Nonclassical Monocytes (% of myeloid)							
Total	105						0.842
Hypofractionated (5-6Gy/fx)	33	4.73	3.53	0.19	20.98	5.16	0.849
Conventional (2-3.5Gy/fx)	20	4.10	2.42	0.00	15.01	4.51	0.624
None	52	4.90	3.12	0.00	22.99	5.76	Ref

Table 2-14: Table summarizing the association between immune cell infiltration and XRT fractionation in STS tumor samples. Kruskal-Wallis test and Wilcoxon tests were used to determine the overall and pairwise differences between the groups, respectively. No XRT was the reference group for all pairwise comparisons.

		Statistics for Cell Types Listed					
Systemic Therapy	*n*	*mean*	*median*	*min*	*max*	*sd*	*p value*
MYELOID AND LYMPHOID LINEAGES							
Lymphoid Cells (% of CD45)							

Total	105						0.507
Cytotoxic	25	51.45	49.30	6.69	87.71	23.38	0.192
Immunotherapy	10	45.34	40.21	2.16	79.98	30.88	0.916
TKI	4	56.83	54.91	32.24	85.26	26.81	0.382
None	66	44.37	39.11	2.35	92.17	24.48	Ref
Myeloid Cells (% of CD45)							
Total	105						0.398
Cytotoxic	25	48.62	50.70	12.29	93.30	23.44	0.258
Immunotherapy	10	54.85	59.64	19.93	97.84	30.59	0.916
TKI	4	38.42	37.87	14.08	63.86	22.55	0.192
None	66	55.18	60.89	7.83	97.65	24.55	Ref
Myeloid/Lymphoid Ratio (log2 value)							
Total	105						0.452
Cytotoxic	25	-0.10	0.04	-2.83	3.80	1.68	0.231
Immunotherapy	10	0.61	0.58	-2.00	5.50	2.54	0.903
TKI	4	-0.66	-0.44	-2.60	0.84	1.67	0.253
None	66	0.45	0.64	-3.56	5.37	1.89	Ref
LYMPHOCYTE LINEAGE							
B Cells (% of lymphocytes)							
Total	105						0.034
Cytotoxic	25	1.77	0.23	0.00	26.44	5.70	0.010
Immunotherapy	10	2.81	1.51	0.19	15.05	4.70	0.618
TKI	4	1.38	1.42	0.38	2.31	0.86	0.600
None	66	2.35	0.91	0.00	21.72	4.25	Ref
CD4 T Cells (% of lymphocytes)							
Total	105						0.174
Cytotoxic	25	33.42	33.79	10.98	62.98	13.54	0.410
Immunotherapy	10	27.13	28.47	7.10	55.85	14.76	0.058
TKI	4	43.53	42.33	28.21	61.26	15.21	0.390
None	66	36.21	36.10	6.86	69.02	13.78	Ref
CD8 T Cells(% of lymphocytes)							
Total	105						0.099
Cytotoxic	25	43.59	40.81	14.86	74.44	17.30	0.035
Immunotherapy	10	44.45	40.58	20.13	84.85	17.94	0.087
TKI	4	34.74	35.18	13.47	55.13	17.57	0.946
None	66	35.47	32.01	3.59	86.66	17.20	Ref
NK Cells (% of lymphocytes)							
Total	105						0.907
Cytotoxic	25	10.08	8.37	1.04	33.30	7.98	0.665
Immunotherapy	10	12.17	12.77	0.07	27.09	9.42	0.847
TKI	4	12.62	12.41	4.15	21.51	8.84	0.691
None	66	11.99	8.57	0.75	46.01	10.69	Ref
CD4 T CELLS							
CD4PD1 (% of CD4)							
Total	105						<0.001
Cytotoxic	25	72.23	78.81	3.88	94.25	20.56	0.388
Immunotherapy	10	23.64	21.89	0.20	49.54	20.38	<0.001

TKI	4	86.26	88.72	71.41	96.19	10.86	0.060
None	66	68.47	72.81	10.95	94.87	19.91	Ref
Treg (% of CD4)							
Total	105						0.035
Cytotoxic	25	16.55	14.00	2.67	46.46	12.40	0.262
Immunotherapy	10	28.09	27.34	1.47	56.81	18.25	0.217
TKI	4	44.98	46.90	18.62	67.49	24.59	0.032
None	66	20.21	18.12	0.95	63.37	13.97	Ref
CD4 Naïve (% of CD4)							
Total	105						0.195
Cytotoxic	25	3.26	1.64	0.00	18.09	4.50	0.118
Immunotherapy	10	2.32	1.41	0.00	7.55	2.52	0.098
TKI	4	1.75	1.68	1.35	2.28	0.46	0.374
None	66	6.48	2.52	0.00	66.34	11.41	Ref
CD4 Tcm (% of CD4)							
Total	105						0.953
Cytotoxic	25	28.86	20.96	6.37	69.61	19.06	0.905
Immunotherapy	10	27.04	31.55	0.20	60.55	20.28	0.785
TKI	4	33.50	32.21	14.21	55.36	20.12	0.610
None	66	29.00	26.45	0.00	67.77	18.36	Ref
CD4 Tem (% of CD4)							
Total	105						0.205
Cytotoxic	25	65.73	74.53	26.39	87.32	20.05	0.065
Immunotherapy	10	67.99	64.90	34.86	92.71	20.80	0.173
TKI	4	63.62	64.07	43.08	83.25	18.89	0.570
None	66	57.63	57.07	0.00	87.15	19.98	Ref
CD4 Temra (% of CD4)							
Total	105						0.095
Cytotoxic	25	2.14	1.01	0.00	16.77	3.46	0.049
Immunotherapy	10	2.65	2.20	0.47	6.88	2.00	0.654
TKI	4	1.14	0.46	0.22	3.41	1.52	0.149
None	66	6.89	2.41	0.00	97.79	14.27	Ref
CD8 T CELLS							
CD8PD1 (% of CD8)							
Total	105						0.004
Cytotoxic	25	70.69	74.83	0.95	98.15	21.18	0.880
Immunotherapy	10	31.17	28.96	0.14	92.62	31.26	<0.001
TKI	4	76.30	75.24	59.75	94.96	14.43	0.690
None	66	70.24	71.81	17.77	98.59	20.13	Ref
CD8 Naïve (% of CD8)							
Total	105						0.672
Cytotoxic	25	6.33	2.81	0.00	42.52	9.70	0.580
Immunotherapy	10	3.74	2.51	0.24	10.91	3.75	0.413
TKI	4	7.53	5.30	1.32	18.21	7.43	0.493
None	66	6.63	3.68	0.00	47.33	9.45	Ref

CD8 Tcm (% of CD8)							
Total	105						0.484
Cytotoxic	25	12.72	6.28	1.21	54.31	15.72	0.505
Immunotherapy	10	19.96	22.95	0.21	42.17	14.87	0.279
TKI	4	24.37	23.89	3.27	46.44	22.63	0.374
None	66	14.12	9.21	0.00	56.18	14.82	Ref
CD8 Tem (% of CD8)							
Total	105						0.256
Cytotoxic	25	61.96	64.88	17.31	94.22	20.03	0.065
Immunotherapy	10	55.82	57.57	33.06	68.48	10.33	0.705
TKI	4	48.67	50.16	25.56	68.79	18.04	0.673
None	66	52.27	54.72	0.00	93.86	21.69	Ref
CD8 Temra (% of CD8)							
Total	105						0.661
Cytotoxic	25	18.99	15.30	0.00	38.37	12.36	0.311
Immunotherapy	10	20.48	17.29	0.79	64.46	18.90	0.374
TKI	4	19.43	18.21	6.04	35.25	13.84	0.738
None	66	26.98	21.09	0.00	100.00	22.39	Ref
MYELOID LINEAGE							
CD11B (% of myeloid)							
Total	105						0.351
Cytotoxic	25	58.52	64.57	18.40	87.23	21.76	0.746
Immunotherapy	10	66.29	76.31	0.09	95.75	31.18	0.348
TKI	4	75.18	81.79	47.95	89.18	18.64	0.184
None	66	59.64	62.31	7.21	98.14	23.65	Ref
HLA DR (% of myeloid)							
Total	105						0.015
Cytotoxic	25	57.67	54.60	20.74	96.50	20.87	0.606
Immunotherapy	10	24.13	15.33	0.00	94.58	29.11	0.005
TKI	4	60.05	60.35	37.69	81.82	24.21	0.585
None	66	54.12	54.51	3.03	96.60	26.15	Ref
Macrophages (% of myeloid)							
Total	105						0.561
Cytotoxic	25	23.82	24.79	6.03	45.73	13.19	0.544
Immunotherapy	10	17.06	16.47	0.05	35.80	11.38	0.578
TKI	4	27.30	27.18	16.54	38.31	9.13	0.315
None	66	21.24	21.15	1.86	51.29	13.31	Ref
mDCs (% of myeloid)							
Total	105						0.006
Cytotoxic	25	6.61	4.87	0.77	17.77	4.87	0.649
Immunotherapy	10	2.09	0.81	0.00	11.48	3.63	<0.001
TKI	4	9.09	4.74	2.45	24.43	10.38	0.864
None	66	7.80	7.08	0.15	21.62	6.05	Ref
gMDSCs (% of myeloid)							
Total	105						0.029
Cytotoxic	25	18.14	11.22	0.21	61.40	17.53	0.981

Immunotherapy	10	42.09	45.04	1.50	73.82	21.20	0.004
TKI	4	27.93	29.93	1.16	50.70	25.34	0.408
None	66	18.61	12.96	0.01	94.54	19.86	Ref
moMDSCs (% of myeloid)							
Total	105						0.531
Cytotoxic	25	1.31	1.02	0.00	4.95	1.37	0.649
Immunotherapy	10	3.62	3.22	0.00	12.07	3.95	0.390
TKI	4	0.84	0.40	0.02	2.54	1.16	0.373
None	66	1.97	0.72	0.00	14.16	3.16	Ref
Classical Monocytes (% of myeloid)							
Total	105						0.061
Cytotoxic	25	20.80	11.94	1.25	66.81	19.19	0.157
Immunotherapy	10	5.98	4.09	0.00	20.00	6.42	0.089
TKI	4	17.61	17.72	7.10	27.89	10.36	0.310
None	66	14.60	7.14	0.38	78.89	16.27	Ref
Intermediate Monocytes (% of myeloid)							
Total	105						0.018
Cytotoxic	25	17.08	9.07	1.85	56.66	16.54	0.231
Immunotherapy	10	10.12	2.17	0.00	71.72	23.23	0.005
TKI	4	18.22	16.02	9.46	31.38	9.47	0.988
None	66	22.84	14.05	0.12	76.39	19.96	Ref
Nonclassical Monocytes (% of myeloid)							
Total	105						0.210
Cytotoxic	25	4.73	3.70	0.00	20.40	5.40	0.981
Immunotherapy	10	2.06	1.87	0.00	5.58	1.53	0.255
TKI	4	9.02	8.04	3.12	16.86	6.30	0.108
None	66	4.78	3.03	0.04	22.99	5.47	Ref

Table 2-15: Table summarizing the association between immune cell infiltration and systemic therapy type in STS tumor samples. Kruskal-Wallis test and Wilcoxon tests were used to determine the overall and pairwise differences between the groups, respectively. No systemic therapy was the reference group for all pairwise comparisons.

		Statistics for Cell Types Listed					
Treatment Type	*n*	*mean*	*median*	*min*	*max*	*sd*	*p value*
MYELOID AND LYMPHOID LINEAGES							
Lymphoid Cells (% of CD45)							
Total	105						0.285
None	31	46.73	41.68	2.35	92.17	25.78	Ref
Systemic Therapy	21	45.94	38.75	2.16	85.26	24.08	0.849
XRT	35	42.43	38.26	2.96	82.59	23.57	0.487
Systemic Therapy + XRT	18	55.93	56.57	6.69	87.71	26.49	0.196
Myeloid Cells (% of CD45)							
Total	105						0.307

None	31	52.29	56.04	7.83	97.65	25.94	Ref
Systemic Therapy	21	53.21	52.48	14.08	97.84	23.99	0.834
XRT	35	57.56	61.74	17.41	96.96	23.47	0.395
Systemic Therapy + XRT	18	44.21	43.43	12.29	93.30	26.36	0.280
Myeloid/Lymphoid Ratio (log2 value)							
Total	105						0.298
None	31	0.28	0.35	-3.56	5.37	2.05	Ref
Systemic Therapy	21	0.32	0.63	-2.60	5.50	1.88	0.820
XRT	35	0.60	0.69	-2.25	5.03	1.77	0.450
Systemic Therapy + XRT	18	-0.33	-0.38	-2.83	3.80	1.98	0.250
LYMPHOCYTE LINEAGE							
B Cells (% of lymphocytes)							
Total	105						0.004
None	31	3.74	1.43	0.01	21.72	5.79	Ref
Systemic Therapy	21	3.15	1.23	0.11	26.44	6.55	0.380
XRT	35	1.16	0.66	0.00	6.65	1.53	0.021
Systemic Therapy + XRT	18	0.54	0.20	0.00	2.69	0.77	0.001
CD4 T Cells (% of lymphocytes)							
Total	105						0.683
None	31	36.20	36.09	6.86	59.24	13.55	Ref
Systemic Therapy	21	32.96	33.91	13.29	61.26	12.67	0.330
XRT	35	36.21	36.82	9.75	69.02	14.19	0.890
Systemic Therapy + XRT	18	32.66	32.77	7.10	62.98	16.69	0.410
CD8 T Cells(% of lymphocytes)							
Total	105						0.021
None	31	34.02	29.26	9.53	86.66	18.12	Ref
Systemic Therapy	21	37.34	39.14	13.47	69.42	14.47	0.180
XRT	35	36.70	33.55	3.59	74.85	16.57	0.280
Systemic Therapy + XRT	18	49.73	51.30	20.13	84.85	18.32	0.003
NK Cells (% of lymphocytes)							
Total	105						0.022
None	31	12.38	8.77	0.75	46.01	11.14	Ref
Systemic Therapy	21	14.88	16.05	0.07	33.30	8.86	0.180
XRT	35	11.66	8.20	0.87	44.72	10.49	0.760
Systemic Therapy + XRT	18	5.93	5.52	1.04	12.61	3.49	0.047
CD4 T CELLS							
CD4PD1 (% of CD4)							
Total	105						0.164
None	31	65.13	68.17	10.95	92.81	22.55	Ref
Systemic Therapy	21	52.43	55.79	0.20	92.11	33.08	0.247
XRT	35	71.48	72.81	25.69	94.87	17.04	0.313
Systemic Therapy + XRT	18	71.41	80.75	10.18	96.19	22.47	0.364
Treg (% of CD4)							
Total	105						0.477
None	31	18.31	13.22	2.23	63.37	13.29	Ref
Systemic Therapy	21	20.44	15.01	1.47	64.30	18.41	1.000
XRT	35	21.93	21.62	0.95	56.72	14.56	0.360
Systemic Therapy + XRT	18	24.77	21.20	5.92	67.49	16.88	0.160
CD4 Naïve (% of CD4)							
Total	105						0.029
None	31	9.83	3.00	0.15	66.34	15.57	Ref

Systemic Therapy	21	3.72	2.14	0.00	18.09	4.65	0.152
XRT	35	3.46	2.09	0.00	16.54	3.79	0.105
Systemic Therapy + XRT	18	1.78	1.26	0.10	8.79	2.04	0.004
CD4 Tcm (% of CD4)							
Total	105						0.980
None	31	26.98	26.09	4.38	60.92	15.14	Ref
Systemic Therapy	21	28.68	23.76	0.20	69.61	20.18	0.890
XRT	35	30.82	28.93	0.00	67.77	20.93	0.710
Systemic Therapy + XRT	18	29.11	20.19	6.37	63.91	18.10	0.870
CD4 Tem (% of CD4)							
Total	105						0.194
None	31	56.91	56.88	16.91	84.25	18.25	Ref
Systemic Therapy	21	65.65	72.51	26.39	92.71	22.06	0.115
XRT	35	58.27	61.70	0.00	87.15	21.71	0.573
Systemic Therapy + XRT	18	66.67	74.33	33.73	85.78	16.83	0.074
CD4 Temra (% of CD4)							
Total	105						0.157
None	31	6.28	3.99	0.08	26.65	7.09	Ref
Systemic Therapy	21	1.95	1.52	0.00	6.88	1.81	0.061
XRT	35	7.45	2.20	0.00	97.79	18.64	0.303
Systemic Therapy + XRT	18	2.44	0.97	0.08	16.77	4.01	0.056
CD8 T CELLS							
CD8PD1 (% of CD8)							
Total	105						0.009
None	31	67.31	68.95	23.68	98.30	20.33	Ref
Systemic Therapy	21	50.55	56.40	0.14	98.15	29.30	0.041
XRT	35	72.87	79.04	17.77	98.59	19.92	0.290
Systemic Therapy + XRT	18	73.64	78.92	2.50	94.96	24.55	0.171
CD8 Naïve (% of CD8)							
Total	105						0.739
None	31	8.25	3.68	0.00	47.33	12.27	Ref
Systemic Therapy	21	7.17	2.71	0.00	42.52	10.22	0.625
XRT	35	5.18	3.29	0.00	25.00	5.74	0.492
Systemic Therapy + XRT	18	4.06	2.92	0.24	18.21	4.58	0.315
CD8 Tcm (% of CD8)							
Total	105						0.883
None	31	10.60	7.11	0.00	29.40	8.75	Ref
Systemic Therapy	21	16.28	7.80	0.21	54.31	17.53	0.580
XRT	35	17.29	10.85	0.00	56.18	18.26	0.480
Systemic Therapy + XRT	18	15.33	9.71	1.21	46.44	15.30	0.600
CD8 Tem (% of CD8)							
Total	105						0.221
None	31	53.24	56.87	9.95	91.12	21.79	Ref
Systemic Therapy	21	54.46	54.19	17.31	82.28	17.46	984.000
XRT	35	51.40	50.99	0.00	93.86	21.93	0.709
Systemic Therapy + XRT	18	64.48	64.92	25.56	94.22	17.43	0.118
CD8 Temra (% of CD8)							
Total	105						0.417
None	31	27.92	22.09	0.69	76.84	20.97	Ref
Systemic Therapy	21	22.09	20.38	0.00	64.46	16.11	0.421
XRT	35	26.13	20.85	0.00	100.00	23.92	0.562
Systemic Therapy + XRT	18	16.14	14.42	0.79	36.24	10.60	0.083
MYELOID LINEAGE							

CD11B (% of myeloid)							
Total	105						0.584
None	31	56.14	58.04	14.50	97.65	24.03	Ref
Systemic Therapy	21	64.98	73.64	0.09	95.75	25.55	0.220
XRT	35	62.54	63.01	7.21	98.14	23.34	0.320
Systemic Therapy + XRT	18	59.44	65.22	18.40	88.42	22.98	0.720
HLA DR (% of myeloid)							
Total	105						0.160
None	31	61.14	68.24	8.27	95.55	25.10	Ref
Systemic Therapy	21	42.88	39.81	0.00	89.90	28.26	0.054
XRT	35	48.30	47.07	3.03	96.60	25.99	0.091
Systemic Therapy + XRT	18	56.93	53.72	15.33	96.50	25.15	0.578
Macrophages (% of myeloid)							
Total	105						0.363
None	31	17.35	13.93	1.86	35.68	11.88	Ref
Systemic Therapy	21	24.79	29.73	0.05	45.73	13.80	0.160
XRT	35	24.67	23.30	4.10	51.29	13.88	0.180
Systemic Therapy + XRT	18	19.64	16.54	7.86	37.98	9.96	0.570
mDCs (% of myeloid)							
Total	105						0.327
None	31	8.37	8.01	0.15	21.62	6.80	Ref
Systemic Therapy	21	5.98	4.24	0.00	24.43	6.48	0.219
XRT	35	7.33	5.15	0.61	21.53	5.42	0.908
Systemic Therapy + XRT	18	5.35	3.04	0.77	17.77	4.82	0.234
gMDSCs (% of myeloid)							
Total	105						0.461
None	31	16.68	10.23	0.09	54.67	16.85	Ref
Systemic Therapy	21	28.49	26.60	0.21	73.82	24.31	0.161
XRT	35	20.20	14.62	0.01	94.54	22.21	0.703
Systemic Therapy + XRT	18	22.01	15.34	0.31	61.40	17.56	258.000
moMDSCs (% of myeloid)							
Total	105						0.508
None	31	1.16	0.63	0.00	5.08	1.46	Ref
Systemic Therapy	21	2.04	1.02	0.00	12.07	2.97	478.000
XRT	35	2.65	0.80	0.00	14.16	3.97	0.108
Systemic Therapy + XRT	18	1.64	1.18	0.00	5.28	1.79	0.862
Classical Monocytes (% of myeloid)							
Total	105						0.038
None	31	20.82	13.65	0.49	78.89	20.05	Ref
Systemic Therapy	21	14.76	6.61	0.00	66.81	18.09	0.161
XRT	35	9.46	5.55	0.38	34.58	10.04	0.014
Systemic Therapy + XRT	18	18.71	11.94	1.25	60.10	15.54	0.966
Intermediate Monocytes (% of myeloid)							
Total	105						0.198
None	31	21.68	12.06	1.07	60.30	19.98	Ref
Systemic Therapy	21	12.91	8.61	0.00	38.89	13.08	0.089
XRT	35	23.80	19.47	0.12	76.39	20.24	0.922
Systemic Therapy + XRT	18	18.48	7.44	1.36	71.72	22.46	0.230
Nonclassical Monocytes (% of myeloid)							
Total	105						0.995
None	31	5.44	3.18	0.04	22.99	6.44	Ref
Systemic Therapy	21	4.21	3.12	0.00	20.40	4.85	0.894
XRT	35	4.24	3.01	0.19	20.98	4.56	0.797
Systemic Therapy + XRT	18	4.94	2.75	0.00	16.86	5.54	0.943

Table 2-16: Table summarizing the association between immune cell infiltration and neoadjuvant treatment type in STS tumor samples. Kruskal-Wallis test and Wilcoxon tests were used to determine the overall and pairwise differences between the groups, respectively. No neoadjuvant therapy was the reference group for all pairwise comparisons.

		Statistics for Cell Types Listed					
XRT	*n*	*mean*	*median*	*min*	*max*	*sd*	*p value*
MYELOID AND LYMPHOID LINEAGES							
Lymphoid Cells (% of CD45)							
No XRT	52	46.39	40.50	2.16	92.17	24.81	0.830
XRT	53	46.93	48.83	2.96	87.71	25.15	
Myeloid Cells (% of CD45)							
No XRT	52	52.68	54.94	7.83	97.84	24.87	0.990
XRT	53	53.11	51.17	12.29	96.96	25.03	
Myeloid/Lymphoid Ratio (log2 value)							
No XRT	52	0.30	0.42	-3.56	5.50	1.96	0.920
XRT	53	0.29	0.07	-2.83	5.03	1.87	
LYMPHOCYTE LINEAGE							
B Cells (% of lymphocytes)							
No XRT	52	3.48	1.42	0.01	26.44	6.07	0.002
XRT	53	0.94	0.38	0.00	6.65	1.34	
CD4 T Cells (% of lymphocytes)							
No XRT	52	34.81	35.17	6.86	61.26	13.14	0.940
XRT	53	35.00	35.35	7.10	69.02	15.01	
CD8 T Cells(% of lymphocytes)							
No XRT	52	35.44	32.85	9.53	86.66	16.57	0.086
XRT	53	41.13	38.33	3.59	84.85	18.10	
NK Cells (% of lymphocytes)							
No XRT	52	13.48	12.03	0.07	46.01	10.16	0.032
XRT	53	9.66	7.58	0.87	44.72	9.08	
CD4 T CELLS							
CD4PD1 (% of CD4)							
No XRT	52	59.57	64.77	0.20	92.81	28.06	0.049
XRT	53	71.46	72.82	10.18	96.19	18.94	
Treg (% of CD4)							
No XRT	52	19.21	14.44	1.47	64.30	15.53	0.130
XRT	53	22.95	21.20	0.95	67.49	15.32	
CD4 Naïve (% of CD4)							
No XRT	52	7.16	2.74	0.00	66.34	12.35	0.049
XRT	53	2.85	1.65	0.00	16.54	3.34	

CD4 Tcm (% of CD4)							
No XRT	52	27.72	24.50	0.20	69.61	17.35	0.720
XRT	53	30.20	23.82	0.00	67.77	19.77	
CD4 Tem (% of CD4)							
No XRT	52	60.73	63.03	16.91	92.71	20.26	0.800
XRT	53	61.31	67.07	0.00	87.15	20.30	
CD4 Temra (% of CD4)							
No XRT	52	4.39	1.77	0.00	26.65	5.82	0.460
XRT	53	5.63	1.31	0.00	97.79	15.18	
CD8 T CELLS							
CD8PD1 (% of CD8)							
No XRT	52	59.98	61.30	0.14	98.30	25.78	0.006
XRT	53	73.14	78.92	2.50	98.59	21.44	
CD8 Naïve (% of CD8)							
No XRT	52	7.78	3.28	0.00	47.33	11.31	0.370
XRT	53	4.77	2.92	0.00	25.00	5.32	
CD8 Tcm (% of CD8)							
No XRT	52	13.08	7.46	0.00	54.31	13.46	0.550
XRT	53	16.58	9.71	0.00	56.18	17.11	
CD8 Tem (% of CD8)							
No XRT	52	53.77	55.87	9.95	91.12	19.82	0.630
XRT	53	56.13	56.81	0.00	94.22	21.19	
CD8 Temra (% of CD8)							
No XRT	52	25.37	21.25	0.00	76.84	19.03	0.310
XRT	53	22.52	19.43	0.00	100.00	20.58	
MYELOID LINEAGE							
CD11B (% of myeloid)							
No XRT	52	60.05	64.31	0.09	97.65	24.81	0.830
XRT	53	61.48	63.60	7.21	98.14	23.00	
HLA DR (% of myeloid)							
No XRT	52	53.07	54.60	0.00	95.55	27.78	0.660
XRT	53	51.24	50.50	3.03	96.60	25.74	
Macrophages (% of myeloid)							
No XRT	52	20.80	16.99	0.05	45.73	13.12	0.690
XRT	53	22.70	21.78	4.10	51.29	12.54	
mDCs (% of myeloid)							
No XRT	52	7.32	4.95	0.00	24.43	6.69	0.940
XRT	53	6.65	4.41	0.61	21.53	5.25	
gMDSCs (% of myeloid)							
No XRT	52	21.90	12.96	0.09	73.82	21.07	1.000
XRT	53	20.82	14.98	0.01	94.54	20.55	
moMDSCs (% of myeloid)							
No XRT	52	1.55	0.65	0.00	12.07	2.27	0.280
XRT	53	2.31	0.80	0.00	14.16	3.40	

Classical Monocytes (% of myeloid)							
No XRT	52	18.14	9.57	0.00	78.89	19.23	0.300
XRT	53	12.61	8.13	0.38	60.10	12.81	
Intermediate Monocytes (% of myeloid)							
No XRT	52	17.80	11.43	0.00	60.30	17.64	0.580
XRT	53	21.99	13.82	0.12	76.39	20.92	
Nonclassical Monocytes (% of myeloid)							
No XRT	52	4.90	3.12	0.00	22.99	5.76	0.899
XRT	53	4.48	2.88	0.00	20.98	4.86	

Table 2-17: Table summarizing the association between immune cell infiltration and neoadjuvant XRT in STS tumor samples. Wilcoxon test was used to determine the differences between the groups.

		Statistics for Cell Types Listed					
Genomics	*n*	*mean*	*median*	*min*	*max*	*sd*	*p value*
MYELOID AND LYMPHOID LINEAGES							
Lymphoid Cells (% of CD45)							
Simple	25	55.55	53.83	6.69	87.71	23.34	0.030
Complex	80	43.86	38.37	2.16	92.17	24.81	
Myeloid Cells (% of CD45)							
Simple	25	44.29	44.58	12.29	93.30	23.38	0.038
Complex	80	55.62	59.90	7.83	97.84	24.79	
Myeloid/Lymphoid Ratio (log2 value)							
Simple	25	-0.39	-0.28	-2.83	3.80	1.68	0.033
Complex	80	0.51	0.64	-3.56	5.50	1.94	
LYMPHOCYTE LINEAGE							
B Cells (% of lymphocytes)							
Simple	25	2.82	1.34	0.00	26.44	5.79	0.280
Complex	80	2.02	0.65	0.00	21.72	4.13	
CD4 T Cells (% of lymphocytes)							
Simple	25	34.03	36.08	9.75	61.26	14.38	0.710
Complex	80	35.17	35.12	6.86	69.02	14.03	
CD8 T Cells(% of lymphocytes)							
Simple	25	35.48	31.94	3.59	79.35	21.21	0.230
Complex	80	39.18	36.80	12.35	86.66	16.29	
NK Cells (% of lymphocytes)							
Simple	25	15.39	12.63	1.04	44.72	12.04	0.100
Complex	80	10.42	8.28	0.07	46.01	8.75	
CD4 T CELLS							

CD4PD1 (% of CD4)							
Simple	25	63.57	70.20	0.20	93.52	25.60	0.690
Complex	80	66.05	71.73	0.97	96.19	24.42	
Treg (% of CD4)							
Simple	25	9.50	9.43	0.95	19.61	5.16	5.60E-06
Complex	80	24.62	21.85	1.47	67.49	15.83	
CD4 Naïve (% of CD4)							
Simple	25	7.05	2.21	0.15	66.34	13.75	0.380
Complex	80	4.38	2.12	0.00	44.85	7.35	
CD4 Tcm (% of CD4)							
Simple	25	30.77	22.43	0.00	67.77	23.07	0.980
Complex	80	28.37	25.49	0.20	69.61	16.98	
CD4 Tem (% of CD4)							
Simple	25	54.53	66.30	0.00	91.76	24.81	0.130
Complex	80	63.09	65.17	16.91	92.71	18.17	
CD4 Temra (% of CD4)							
Simple	25	7.65	2.08	0.00	97.79	20.02	0.570
Complex	80	4.16	1.57	0.00	39.76	6.72	
CD8 T CELLS							
CD8PD1 (% of CD8)							
Simple	25	65.10	63.41	0.14	98.15	25.59	0.820
Complex	80	66.94	70.68	0.54	98.59	24.33	
CD8 Naïve (% of CD8)							
Simple	25	10.24	6.55	0.00	47.33	13.11	1.60E-01
Complex	80	5.03	2.87	0.00	42.11	6.79	
CD8 Tcm (% of CD8)							
Simple	25	12.58	4.28	0.00	54.31	15.48	0.220
Complex	80	15.52	10.37	0.00	56.18	15.41	
CD8 Tem (% of CD8)							
Simple	25	47.94	53.03	0.00	90.67	23.78	0.130
Complex	80	57.18	57.60	18.42	94.22	18.89	
CD8 Temra (% of CD8)							
Simple	25	29.24	29.62	0.00	100.00	22.92	0.150
Complex	80	22.27	19.38	0.00	76.84	18.50	
MYELOID LINEAGE							
CD11B (% of myeloid)							
Simple	25	43.03	47.72	7.21	70.66	19.81	6.00E-05
Complex	80	66.42	73.37	0.09	98.14	22.22	
HLA DR (% of myeloid)							
Simple	25	48.23	38.86	9.26	85.40	25.15	0.466
Complex	80	53.39	52.73	0.00	96.60	27.15	
Macrophages (% of myeloid)							
Simple	25	16.26	16.74	3.16	34.07	9.99	0.065
Complex	80	23.58	22.37	0.05	51.29	13.14	

mDCs (% of myeloid)							
Simple	25	9.76	10.25	0.77	24.43	6.70	0.010
Complex	80	6.10	4.10	0.00	21.53	5.50	
gMDSCs (% of myeloid)							
Simple	25	18.51	10.19	1.16	58.79	17.12	0.850
Complex	80	22.26	14.98	0.01	94.54	21.75	
moMDSCs (% of myeloid)							
Simple	25	1.44	0.38	0.00	13.07	2.94	0.110
Complex	80	2.09	0.91	0.00	14.16	2.90	
Classical Monocytes (% of myeloid)							
Simple	25	10.37	6.02	0.82	37.91	10.48	0.180
Complex	80	16.93	10.16	0.00	78.89	17.70	
Intermediate Monocytes (% of myeloid)							
Simple	25	15.89	10.89	0.50	54.39	16.74	0.370
Complex	80	21.20	12.87	0.00	76.39	20.08	
Nonclassical Monocytes (% of myeloid)							
Simple	25	5.70	3.46	0.00	19.12	5.61	0.270
Complex	80	4.36	2.54	0.00	22.99	5.20	

Table 2-18: Table summarizing the association between immune cell infiltration and genomics histology group (as defined in methods) in STS tumor samples. Wilcoxon test was used to determine the differences between the groups.

		Statistics for Cell Types Listed					
CNA	*n*	*mean*	*median*	*min*	*max*	*sd*	*p value*
MYELOID AND LYMPHOID LINEAGES							
Lymphoid Cells (% of CD45)							
Low	38	55.93	54.11	6.69	87.71	21.81	0.004
High	67	41.68	35.74	2.16	92.17	25.12	
Myeloid Cells (% of CD45)							
Low	38	44.02	44.16	12.29	93.30	21.81	0.006
High	67	57.69	62.00	7.83	97.84	25.18	
Myeloid/Lymphoid Ratio (log2 value)							
Low	38	-0.40	-0.31	-2.83	3.80	1.53	0.005
High	67	0.67	0.81	-3.56	5.50	2.00	
LYMPHOCYTE LINEAGE							
B Cells (% of lymphocytes)							
Low	38	2.47	1.26	0.00	26.44	4.89	0.300
High	67	2.07	0.65	0.00	21.72	4.39	
CD4 T Cells (% of lymphocytes)							
Low	38	36.38	36.46	9.75	62.98	13.55	0.440
High	67	34.14	33.91	6.86	69.02	14.34	

CD8 T Cells(% of lymphocytes)							
Low	38	35.65	34.04	3.59	79.35	17.82	0.290
High	67	39.71	36.49	12.35	86.66	17.32	
NK Cells (% of lymphocytes)							
Low	38	14.77	12.61	1.04	44.72	10.98	0.029
High	67	9.86	8.05	0.07	46.01	8.68	
CD4 T CELLS							
CD4PD1 (% of CD4)							
Low	38	61.23	68.00	0.20	93.52	26.44	0.260
High	67	67.80	72.82	0.97	96.19	23.40	
Treg (% of CD4)							
Low	38	12.11	11.01	0.95	45.81	8.65	7.90E-06
High	67	25.94	24.32	1.47	67.49	16.18	
CD4 Naïve (% of CD4)							
Low	38	7.21	2.44	0.15	66.34	13.30	0.150
High	67	3.81	1.93	0.00	35.07	5.83	
CD4 Tcm (% of CD4)							
Low	38	30.61	24.17	0.00	69.61	21.00	0.710
High	67	28.03	24.48	0.20	64.00	17.11	
CD4 Tem (% of CD4)							
Low	38	55.67	64.71	0.00	91.76	23.67	0.100
High	67	64.00	66.84	22.24	92.71	17.44	
CD4 Temra (% of CD4)							
Low	38	6.51	1.63	0.00	97.79	17.01	0.830
High	67	4.16	1.70	0.00	39.76	6.54	
CD8 T CELLS							
CD8PD1 (% of CD8)							
Low	38	61.66	62.18	0.14	98.15	26.20	0.160
High	67	69.19	72.41	0.54	98.59	23.32	
CD8 Naïve (% of CD8)							
Low	38	9.57	3.96	0.00	47.33	12.58	0.052
High	67	4.46	2.71	0.00	30.46	5.39	
CD8 Tcm (% of CD8)							
Low	38	12.95	8.51	0.00	54.31	14.82	0.380
High	67	15.85	9.71	0.00	56.18	15.73	
CD8 Tem (% of CD8)							
Low	38	48.57	50.79	0.00	90.67	23.14	0.066
High	67	58.49	58.76	25.21	94.22	18.00	
CD8 Temra (% of CD8)							
Low	38	28.91	28.12	0.00	100.00	22.16	0.050
High	67	21.20	19.33	0.00	69.83	17.88	
MYELOID LINEAGE							
CD11B (% of myeloid)							
Low	38	50.31	54.40	7.21	86.34	21.02	0.001

High	67	66.57	75.62	0.09	98.14	23.40	
HLA DR (% of myeloid)							
Low	38	51.23	52.80	9.26	85.40	23.69	0.840
High	67	52.65	52.08	0.00	96.60	28.31	
Macrophages (% of myeloid)							
Low	38	21.24	23.24	1.86	45.73	12.92	0.810
High	67	22.08	16.99	0.05	51.29	12.82	
mDCs (% of myeloid)							
Low	38	8.49	6.31	0.77	24.43	6.13	0.023
High	67	6.15	3.76	0.00	21.53	5.78	
gMDSCs (% of myeloid)							
Low	38	20.47	12.03	0.52	58.79	18.09	0.790
High	67	21.84	14.55	0.01	94.54	22.15	
moMDSCs (% of myeloid)							
Low	38	1.24	0.38	0.00	13.07	2.51	0.015
High	67	2.32	1.08	0.00	14.16	3.06	
Classical Monocytes (% of myeloid)							
Low	38	14.80	7.14	0.82	66.81	15.68	0.920
High	67	15.65	9.59	0.00	78.89	16.98	
Intermediate Monocytes (% of myeloid)							
Low	38	14.44	9.46	0.50	54.39	14.80	0.120
High	67	22.95	14.51	0.00	76.39	21.00	
Nonclassical Monocytes (% of myeloid)							
Low	38	6.26	3.70	0.00	20.40	5.91	0.056
High	67	3.81	2.30	0.00	22.99	4.76	

Table 2-19: Table summarizing the association between immune cell infiltration and CNA histology group (as defined in methods) in STS tumor samples. Wilcoxon test was used to determine the differences between the groups.

		Statistics for Cell Types Listed					
TIL Growth	*n*	*mean*	*median*	*min*	*max*	*sd*	*p value*
Blood CD3 (% of Live)							
Total	85						0.945
0	38	80.85	84.65	44.73	97.58	11.78	Ref
1	21	76.23	82.15	23.40	97.86	22.10	0.844
2	26	80.84	82.86	61.91	94.39	9.25	0.742
Blood CD4 (% of CD45)							
Total	85						0.729
0	38	7.14	4.70	0.70	20.73	6.09	Ref
1	21	7.04	5.11	1.33	12.79	3.78	0.472
2	26	6.30	6.82	0.46	13.73	3.47	0.694
Blood CD4 (% of Lymphocytes)							
Total	85						0.248
0	38	61.48	64.26	26.94	89.08	15.36	Ref
1	21	61.85	63.56	27.35	85.97	14.90	1.000

2	26	55.85	58.33	24.36	86.72	15.15	0.170
Blood CD4 PD1 (% of CD4)							
Total	85						0.068
0	38	35.45	35.18	0.11	73.50	19.58	Ref
1	21	36.17	29.93	20.05	64.78	14.26	0.825
2	26	44.79	50.23	0.04	80.41	19.12	0.064
Blood CD4 PD1 (% of CD45)							
Total	85						0.297
0	38	2.22	1.88	0.00	10.29	2.23	Ref
1	21	2.65	2.36	0.27	8.29	1.99	0.305
2	26	3.11	2.63	0.00	9.33	2.33	0.153
Blood CD8 (% of CD45)							
Total	85						0.491
0	38	3.53	3.00	0.27	9.96	2.72	Ref
1	21	3.97	3.75	0.39	11.94	3.31	0.847
2	26	5.08	3.75	0.31	16.45	4.38	0.245
Blood CD8 (% of Lymphocytes)							
Total	85						0.125
0	38	33.82	32.21	8.84	72.10	15.14	Ref
1	21	30.37	27.94	2.90	72.43	15.84	0.415
2	26	40.07	38.76	9.73	72.53	15.17	0.201
Blood CD8 PD1 (% of CD45)							
Total	85						0.538
0	38	1.34	1.12	0.00	5.35	1.29	Ref
1	21	1.26	0.81	0.07	4.86	1.32	0.825
2	26	2.06	1.47	0.00	9.03	2.32	0.418
Blood CD8 PD1 (% of CD8)							
Total	85						0.226
0	38	38.57	37.70	0.08	71.89	20.37	Ref
1	21	31.06	30.25	12.73	67.51	14.13	0.125
2	26	36.67	39.04	0.00	74.56	17.97	0.678
Blood Granulocytes (% of CD45)							
Total	85						0.492
0	38	70.22	71.76	40.59	92.10	11.65	Ref
1	21	65.79	67.83	42.94	89.68	14.72	0.280
2	26	70.44	71.84	43.33	92.26	11.80	0.965
Blood Lymphocytes (% of CD45)							
Total	85						0.605
0	38	17.15	15.17	2.86	46.32	11.49	Ref
1	21	18.51	17.96	3.73	34.97	9.17	0.391
2	26	18.93	19.22	3.72	46.33	11.01	0.431
Blood Monocytes (% of CD45)							
Total	85						0.281
0	38	10.97	10.91	4.69	21.65	3.73	Ref
1	21	12.44	10.20	1.52	28.89	7.82	0.868
2	26	9.43	9.14	2.27	19.38	3.83	0.118
Tumor CD4 (% of CD45)							
Total	85						0.048
0	38	13.53	8.06	0.37	46.73	12.33	Ref
1	21	19.77	16.45	0.93	48.93	13.18	0.035
2	26	17.95	14.60	2.62	52.23	12.00	0.049
Tumor CD4 (% of Lymphocytes)							
Total	85						0.449
0	38	32.32	30.52	6.86	62.98	14.83	Ref
1	21	36.31	39.01	10.98	55.38	13.69	0.267
2	26	35.86	35.99	7.10	61.26	13.47	0.341

Tumor CD4 PD1 (% of CD4)							
Total	85						0.006
0	38	56.62	61.05	0.97	92.81	28.42	Ref
1	21	72.94	71.91	50.78	92.83	12.45	0.044
2	26	76.13	85.22	28.45	96.19	19.40	0.003
Tumor CD4 PD1 (% of CD45)							
Total	85						<0.001
0	38	8.15	4.77	0.03	33.39	8.30	Ref
1	21	15.33	13.02	1.71	30.02	8.11	<0.001
2	26	15.40	14.28	2.23	37.96	9.99	0.002
Tumor CD45 (% of Live)							
Total	85						0.666
0	38	57.65	58.61	2.42	98.91	32.00	Ref
1	21	60.40	63.91	0.36	98.54	29.61	0.761
2	26	63.89	71.52	14.31	99.71	28.85	0.371
Tumor CD8 (% of CD45)							
Total	85						0.386
0	38	18.04	14.47	0.66	68.64	17.02	Ref
1	21	22.54	19.39	0.38	61.33	17.26	0.253
2	26	21.21	16.18	1.81	68.34	17.66	0.288
Tumor CD8 (% of Lymphocytes)							
Total	85						0.722
0	38	40.07	37.82	3.59	86.66	17.44	Ref
1	21	38.21	31.79	15.04	74.85	18.38	0.470
2	26	39.75	38.92	9.53	84.85	18.36	0.895
Tumor CD8 PD1 (% of CD45)							
Total	85						0.085
0	38	13.82	5.42	0.03	75.82	18.41	Ref
1	21	18.14	11.87	0.80	62.73	18.78	0.080
2	26	17.93	12.37	1.69	67.53	18.29	0.057
Tumor CD8 PD1 (% of CD8)							
Total	85						0.249
0	38	61.19	60.87	0.54	98.59	30.41	Ref
1	21	71.33	68.63	43.92	97.35	13.91	0.355
2	26	74.88	76.09	32.38	97.00	17.25	0.119
Tumor Lymphocytes (% of CD45)							
Total	85						0.080
0	38	40.28	38.00	2.16	87.71	25.69	Ref
1	21	55.26	59.02	2.35	92.17	25.10	0.049
2	26	50.49	45.37	6.58	86.13	23.99	0.131
Tumor Myeloid Cells (% of CD45)							
Total	85						0.070
0	38	59.67	62.00	12.29	97.84	25.66	Ref
1	21	44.79	40.98	7.83	97.65	25.05	0.039
2	26	48.39	50.48	13.87	93.42	23.82	0.087

Table 2-20: Table summarizing the association between immune cell infiltration and TIL growth in STS tumor samples. TILs were successfully generated from 85 samples in this study. TIL growth was defined as 0, 1, and 2 for no growth, moderate growth, and strong growth, respectively. Kruskal-Wallis test and Wilcoxon tests were used to determine the overall and

pairwise differences between the groups, respectively. No growth (0) was the reference group for all pairwise comparisons. The above analysis was also conducted comparing no TIL growth (0) to any TIL growth (1 & 2). There were no notable differences in the results (data not shown).

	TIL Growth 0		TIL Growth 1		TIL Growth 2		
	n or median	*% or range*	*n or median*	*% or range*	*n or median*	*% or range*	*p value*
Total	38	44.71	21	24.71	26	30.59	
Sex							0.968
F	14	45.16	8	25.81	9	29.03	
M	24	44.44	13	24.07	17	31.48	
Age (years)	63	27-83	66	33-81	60	18-91	0.987
Tumor size (max dimension, cm)							0.229
≤5cm	8	38.10	4	19.05	9	42.86	
5-10cm	13	39.39	12	36.36	8	24.24	
>10cm	17	54.84	5	16.13	9	29.03	
Grade							0.820
Low	1	25.00	1	25.00	2	50.00	
Intermediate	7	46.67	5	33.33	3	20.00	
High	20	42.55	12	25.53	15	31.91	
Missing	10	52.63	3	15.79	6	31.58	
Tumor Site							0.129
Lung	0	0.00	0	0.00	1	100.00	
Extremity/Trunk	30	52.63	12	21.05	15	26.32	
RP/Abdomen/Pelvis	8	29.63	19	33.33	10	37.04	
Histology							0.036
GIST	0	0.00	1	33.33	2	66.67	
LMS	5	45.45	1	9.09	5	45.45	
SS	2	40.00	0	0.00	3	60.00	
MLS	8	80.00	1	10.00	1	10.00	
DDLPS	1	9.09	6	54.55	4	36.36	
MFS	3	60.00	1	20.00	1	20.00	
USARC	19	47.50	11	27.50	10	25.00	
Lesion Type							0.602
Primary	23	46.94	14	28.57	12	24.49	
Locally Recurrent	6	46.15	3	23.08	4	30.77	
Metastatic	9	39.13	3	17.39	10	43.48	
Preoperative Radiation							0.538
Hypofractionated	12	54.55	5	22.73	5	22.73	
Conventional	9	47.37	6	31.58	4	21.05	
None	17	38.64	10	22.73	17	38.64	
Radiation-Surgery Interval (Days)							0.446

≤1	0	0.00	0	0.00	1	1.00	
14-30	12	57.14	6	28.57	3	14.29	
31-49	7	58.33	3	25.00	2	16.67	
50-71	1	20.00	2	40.00	2	40.00	
>133	1	50.00	0	0.00	1	50.00	
NA							
Preoperative Systemic Therapy							0.058
Cytotoxic	10	45.45	7	31.82	5	22.73	
Immunotherapy	6	75.00	0	0.00	2	25.00	
TKI	0	0.00	0	0.00	4	1.00	
None	22	43.14	14	27.45	15	29.41	
Treatment Effect							0.005
≤10%	3	25.00	1	8.33	8	66.67	
10-50%	8	50.00	3	18.75	5	31.25	
51-89%	8	57.14	6	42.86	0	0.00	
≥90%	7	58.33	4	33.33	1	8.33	
NA	12	38.71	7	22.58	12	38.71	
Treatment Status							0.292
Untreated	8	32.00	7	28.00	10	40.00	
Treated	30	50.00	14	23.33	16	26.67	
Treatment Type							0.617
None	8	32.00	7	28.00	10	40.00	
Systemic Therapy	9	47.37	3	15.79	7	36.84	
XRT	14	53.85	7	26.92	5	19.23	
Systemic Therapy + XRT	7	46.67	4	26.67	4	26.67	
Systemic Therapy							0.772
No	22	43.14	14	27.45	15	29.41	
Yes	16	47.06	7	20.59	11	32.35	
XRT							0.243
No	17	38.64	10	22.73	17	38.64	
Yes	21	51.22	11	26.83	9	21.95	

Table 2-21: Table summarizing the association between various clinical, tumor, and treatment factors and TIL growth in STS tumor samples. TILs were successfully generated from 85 samples in this study. TIL growth was defined as 0, 1, and 2 for no growth, moderate growth, and strong growth, respectively. Chi-squared, Fisher Exact, or Wilcoxon tests were used to compare the groups, as indicated. The above analysis was also conducted comparing no TIL growth (0) to any TIL growth (1 & 2). There were no notable differences in the results (data not shown).

2.10 Bibliography

1. Pollack SM, Ingham M, Spraker MB, Schwartz GK. Emerging Targeted and Immune-Based Therapies in Sarcoma. *J Clin Oncol Off J Am Soc Clin Oncol*. 2018;36(2):125-135. doi:10.1200/JCO.2017.75.1610
2. Patnaik A, Kang SP, Rasco D, et al. Phase I Study of Pembrolizumab (MK-3475; Anti-PD-1 Monoclonal Antibody) in Patients with Advanced Solid Tumors. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2015;21(19):4286-4293. doi:10.1158/1078-0432.CCR-14-2607
3. Tawbi HA, Burgess M, Bolejack V, et al. Pembrolizumab in advanced soft-tissue sarcoma and bone sarcoma (SARC028): a multicentre, two-cohort, single-arm, open-label, phase 2 trial. *Lancet Oncol*. 2017;18(11):1493-1501. doi:10.1016/S1470-2045(17)30624-1
4. Abeshouse A, Adebamowo C, Adebamowo SN, et al. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. *Cell*. 2017;171(4):950-965.e28. doi:10.1016/j.cell.2017.10.014
5. Cancer Facts & Figures 2017. Accessed May 5, 2021. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2017.html
6. Sarcomas, Soft Tissue - Statistics. Cancer.Net. Published June 25, 2012. Accessed May 5, 2021. https://www.cancer.net/cancer-types/sarcomas-soft-tissue/statistics
7. Ferguson PC, Deheshi BM, Chung P, et al. Soft tissue sarcoma presenting with metastatic disease: outcome with primary surgical resection. *Cancer*. 2011;117(2):372-379. doi:10.1002/cncr.25418

8. Brennan MF, Antonescu CR, Moraco N, Singer S. Lessons learned from the study of 10,000 patients with soft tissue sarcoma. *Ann Surg*. 2014;260(3):416-421; discussion 421-422. doi:10.1097/SLA.0000000000000869
9. Trovik CS, Bauer HC, Alvegård TA, et al. Surgical margins, local recurrence and metastasis in soft tissue sarcomas: 559 surgically-treated patients from the Scandinavian Sarcoma Group Register. *Eur J Cancer Oxf Engl 1990*. 2000;36(6):710-716. doi:10.1016/s0959-8049(99)00287-7
10. Rööser B, Gustafson P, Rydholm A. Is there no influence of local control on the rate of metastases in high-grade soft tissue sarcoma? *Cancer*. 1990;65(8):1727-1729. doi:10.1002/1097-0142(19900415)65:8<1727::aid-cncr2820650811>3.0.co;2-x
11. Gustafson P. Soft tissue sarcoma. Epidemiology and prognosis in 508 patients. *Acta Orthop Scand Suppl*. 1994;259:1-31.
12. Gustafson P, Rööser B, Rydholm A. Is local recurrence of minor importance for metastases in soft tissue sarcoma? *Cancer*. 1991;67(8):2083-2086. doi:10.1002/1097-0142(19910415)67:8<2083::aid-cncr2820670813>3.0.co;2-5
13. Alamanda VK, Crosby SN, Archer KR, Song Y, Schwartz HS, Holt GE. Predictors and clinical significance of local recurrence in extremity soft tissue sarcoma. *Acta Oncol Stockh Swed*. 2013;52(4):793-802. doi:10.3109/0284186X.2012.711953
14. Weitz J, Antonescu CR, Brennan MF. Localized extremity soft tissue sarcoma: improved knowledge with unchanged survival over time. *J Clin Oncol Off J Am Soc Clin Oncol*. 2003;21(14):2719-2725. doi:10.1200/JCO.2003.02.026
15. Zaidi MY, Ethun CG, Tran TB, et al. Assessing the Role of Neoadjuvant Chemotherapy in Primary High-Risk Truncal/Extremity Soft Tissue Sarcomas: An Analysis of the Multi-

institutional U.S. Sarcoma Collaborative. *Ann Surg Oncol*. 2019;26(11):3542-3549. doi:10.1245/s10434-019-07639-7

16. Ben-Ami E, Barysauskas CM, Solomon S, et al. Immunotherapy with single agent nivolumab for advanced leiomyosarcoma of the uterus: Results of a phase 2 study. *Cancer*. 2017;123(17):3285-3290. doi:10.1002/cncr.30738
17. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell*. 2015;160(1-2):48-61. doi:10.1016/j.cell.2014.12.033
18. Van Allen EM, Miao D, Schilling B, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science*. 2015;350(6257):207-211. doi:10.1126/science.aad0095
19. Petitprez F, de Reyniès A, Keung EZ, et al. B cells are associated with survival and immunotherapy response in sarcoma. *Nature*. 2020;577(7791):556-560. doi:10.1038/s41586-019-1906-8
20. Taylor BS, Barretina J, Maki RG, Antonescu CR, Singer S, Ladanyi M. Advances in sarcoma genomics and new therapeutic targets. *Nat Rev Cancer*. 2011;11(8):541-557. doi:10.1038/nrc3087
21. Doyle LA, Nowak JA, Nathenson MJ, et al. Characteristics of mismatch repair deficiency in sarcomas. *Mod Pathol*. 2019;32(7):977-987. doi:10.1038/s41379-019-0202-3
22. Lenze U, Pohlig F, Mühlhofer H, et al. Do human tumor-associated viruses play a role in the development of synovial sarcoma? *Clin Sarcoma Res*. 2015;5(1):11. doi:10.1186/s13569-015-0027-x

23. Tijhuis AE, Johnson SC, McClelland SE. The emerging links between chromosomal instability (CIN), metastasis, inflammation and tumour immunity. *Mol Cytogenet*. 2019;12(1):17. doi:10.1186/s13039-019-0429-1

24. Li J, Duran MA, Dhanota N, et al. Metastasis and Immune Evasion from Extracellular cGAMP Hydrolysis. *Cancer Discov*. 2021;11(5):1212-1227. doi:10.1158/2159-8290.CD-20-0387

25. Bakhoum SF, Ngo B, Laughney AM, et al. Chromosomal instability drives metastasis through a cytosolic DNA response. *Nature*. 2018;553(7689):467-472. doi:10.1038/nature25432

26. Davoli T, Uno H, Wooten EC, Elledge SJ. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science*. 2017;355(6322). doi:10.1126/science.aaf8399

27. Sharma P, Hu-Lieskovan S, Wargo JA, Ribas A. Primary, Adaptive and Acquired Resistance to Cancer Immunotherapy. *Cell*. 2017;168(4):707-723. doi:10.1016/j.cell.2017.01.017

28. Poh AR, Ernst M. Targeting Macrophages in Cancer: From Bench to Bedside. *Front Oncol*. 2018;8. doi:10.3389/fonc.2018.00049

29. Anfray C, Ummarino A, Torres Andón F, Allavena P. Current Strategies to Target Tumor-Associated-Macrophages to Improve Anti-Tumor Immune Responses. *Cells*. 2019;9(1). doi:10.3390/cells9010046

30. Zhang F, Parayath NN, Ene CI, et al. Genetic programming of macrophages to perform anti-tumor functions using targeted mRNA nanocarriers. *Nat Commun*. 2019;10. doi:10.1038/s41467-019-11911-5

31. Sarode P, Zheng X, Giotopoulou GA, et al. Reprogramming of tumor-associated macrophages by targeting β-catenin/FOSL2/ARID5A signaling: A potential treatment of lung cancer. *Sci Adv*. 2020;6(23):eaaz6105. doi:10.1126/sciadv.aaz6105

CHAPTER 3

Validation of in-silico immune deconvolution methods in undifferentiated sarcoma

3.1 Abstract

The composition of immune cells in the tumor microenvironment (TME) has been shown to be associated with prognosis and immunotherapy response in multiple cancers, including sarcoma[1–3]. Studies often use in-silico immune deconvolution to define the immune cell composition; however, the optimal in-silico tool for soft tissue sarcoma is not known. In this study, we assess the concordance between various in-silico immune deconvolution techniques and immunohistochemistry to determine the optimal in-silico immune deconvolution strategy in undifferentiated sarcoma. We created a tissue microarray (TMA) of 60 untreated sarcoma samples and performed RNA sequencing on these samples. Multiplex immunofluorescence was performed on the TMA, staining for CD20, CD68, CD4, CD8, and CD45. In-silico immune deconvolution was performed using six tools – CIBERSORTx, EPIC, MCP Counter, quanTIseq, TIMER, xCell[4–10]. The correlation between mIF and in-silico scores was determined for each cell type using a linear fit model and Pearson correlation coefficients. We found that, overall, the in-silico tools more accurately quantified overall immune cell infiltration than individual immune cell subtypes. We found that TIMER was the best tool for defining overall immune cell infiltration. Many tools performed well when defining the monocyte / macrophage cell population. Caution should be used when applying in-silico immune deconvolution tools to define CD8+ T cells, and in-silico immune deconvolution tools should be avoided when defining CD4+ T cells and B cells in undifferentiated sarcoma.

3.2 Introduction

The composition of immune cells in the tumor microenvironment (TME) has been shown to be associated with prognosis and immunotherapy response in multiple cancers, including sarcoma[1–3]. Flow cytometry and immunohistochemistry (IHC) are frequently used to characterize the immune composition of the TME; however, their limitations include the requirement for large amounts of fresh tumor (flow cytometry), low-throughput (flow cytometry and IHC), and the lack of publicly available data (flow cytometry and IHC)[10–12]. The ability to use tumor-derived RNA for immunologic analyses would circumvent the issues of tissue amount, throughput and data availability.

Multiple in-silico immune deconvolution methods, using bulk or single-cell RNA-seq (scRNA-Seq) data, have been developed. These methods allow for high-throughput quantification of immune cell populations and tools developed for bulk RNA-Seq facilitate the inclusion of preserved tumors and/or publicly available RNA-Seq data[4–12]. The ability to apply these methods to publicly available data and/or preserved tumors is particularly important for increasing sample sizes when studying rare diseases, such as soft tissue sarcoma (STS). However, the optimal in-silico immune deconvolution tool for studying sarcoma is not known. Prior studies have compared various in-silico immune deconvolution tools and suggest that histology-specific immune markers may improve estimates. Unfortunately, sarcoma was often omitted from test and/or validation cohorts when creating the in-silico tools[4–12].

The immune microenvironment in sarcoma, and specifically in undifferentiated sarcoma, remains poorly understood. Immune-based therapies have dramatically changed the treatment of many cancers in recent decades, and immunotherapy response has been shown to be associated with the composition of cells in the TME[1–3]. Early studies of immunotherapy in STS have not had broad success, however, there are findings that warrant further exploration[13–16]. The SARC028 study evaluated the anti-PD1 antibody pembrolizumab in patients with advanced sarcoma, and identified the undifferentiated sarcoma (specifically undifferentiated pleomorphic sarcoma, UPS) subtype as the most responsive to anti-PD1 immunotherapy. In this study, 40% of patients with UPS demonstrated complete or partial response[15]. To understand why the undifferentiated sarcoma subtype demonstrated the highest response rate and to elucidate the underlying mechanisms of response, we must first understand the immune microenvironment in undifferentiated sarcoma. In-silico immune deconvolution would allow the study of this rare tumor in an effective and efficient manner.

In this study, we aim to assess the concordance between various in-silico immune deconvolution techniques and immunofluorescence to determine the optimal in-silico immune deconvolution strategy in undifferentiated sarcoma.

3.3 Materials and Methods

3.3.1 Patient and sample identification

All patients treated at our institution for sarcoma between January 1, 2010 and December 30, 2020 were identified using a prospectively maintained database. Patient, tumor, and treatment data were extracted from the electronic medical record. Sixty patients with untreated undifferentiated sarcoma (USARC) with available formalin-fixed paraffin-embedded (FFPE) tissue blocks were identified for inclusion in this study (Figure 3-1A). Hematoxylin and eosin (H&E) stained sections of each tissue block were re-reviewed by an expert sarcoma pathologist to confirm USARC pathology.

3.3.2 Tissue microarray

The H&E sections of each tumor (selected as mentioned above) were again examined by an expert sarcoma pathologist to identify the region(s) of each slide that contained tumor. Three 1mm cores from each specimen were taken from the identified tumor region(s). A tissue microarray (TMA) consisting of triplicates of each sample was created.

3.3.3 Multiplex Immunofluorescence

Multiplex immunofluorescence was performed in the same manner as described in Chapter 2. The TSA-based Opal method was used for immunofluorescence (IF) staining (Opal Polaris 7-Color Automation IHC Kit; Akoya Biosciences, Marlborough, MA, USA; Catalogue No. NEL871001KT). Because TSA and DAB oxidation are both peroxidase-mediated reactions, the primary antibody conditions and order of staining determined using DAB detection were directly applied to the fluorescent assays. Unlike conventional IHC in which a chromogenic peroxidase

substrate is used for antigen detection, each antibody is paired with an individual Opal fluorophore for visualization. The Opal fluorophores were used at a 1 in 150 dilution, as recommended by Akoya when using the Leica BOND RX. As such, a fluorescent singleplex was performed for each biomarker and compared to the appropriate chromogenic singleplex to assess staining performance. Additional information regarding antibodies used are included in the following list: antibody, clone/lot (if applicable), dilution (company, catalog#).

Antibodies:

CD4, 1:80 (Dako, M7310)

CD8, C8/144B, 1:100 (Dako, M7103)

CD20, 26, 1:500 (Dako, M0755)

CD68, PG-M1, 1:200 (Dako, m0876)

CD45, 2B11+PD7/26, 1:200 (Dako, M0701)

Once each target was optimized using uniplex slides, the Opal 6 multiplexed assay was used to generate multiple staining slides. We applied primary antibodies to normal human tonsil specimens as controls at optimized concentrations previously determined on the uniplex control tissues. TMA slides Staining was performed consecutively Leica BOND RX by using the same steps as those used in uniplex IF, and the detection for each marker was completed before application of the next antibody.

All fluorescently labelled slides were scanned on the Vectra Polaris (Akoya Biosciences) at 40× magnification using appropriate exposure times. The data from the multispectral camera were analyzed by the imaging InForm software (Akoya Biosciences).

Regions of tumor tissue confirmed by corresponding H&E slide were selected using Phenochart 1.0.12 (Akoya Bioscience), exported as multi-layer TIFF images using inForm 2.4.10 (Akoya Bioscience), and stitched in HALO v3 (Indica Labs) for quantitative image analysis. Cell inclusion and segmentation criteria were optimized using the "real-time tuning" feature to define the nuclear contrast threshold, minimum nuclear intensity, nuclear segmentation aggressiveness, and maximum cytoplasmic radius. For each marker, specific positivity thresholds were defined based on a human tonsil positive control. The entire image was analyzed using these parameters and the generated data included the percentage of cells positive for each marker, expressed as a percentage of all nucleated cells.

3.3.4 Nucleic acid extraction and sequencing

RNA was extracted from FFPE-preserved tissues using the Covaris tNA Plus Kit. RNA QC was performed using High Sensitivity RNA ScreenTape (Agilent TapeStation Software v3.2). Library preparation for RNA was performed using hybrid capture technique using the Illumina TruSeq Exome kit. Paired-end RNA sequencing was performed using Illumina Hiseq 3000 platform with read length of 2x150 and a total of 40 million reads. All RNA-sequencing (RNA-seq) data was processed using the Toil pipeline, as described in previous publications[17]. RNA

outliers were identified using a combination of principal component analysis (PCA) and Pearson correlation.

3.3.5 Acquisition and processing of publicly available sequencing data

Raw RNA-Seq files were downloaded from Therapeutically Applicable Research to Generate Effective Treatments (TARGET), European Genome Archive (EGA), and Sequence Read Archive (SRA)[18–25]. BAM files from the EGA were converted to fastq files using Picard[26]. All fastq files were processed using the Toil pipeline, as noted above. Toil processed RNA sequencing data files from The Cancer Genome Atlas (TCGA) were downloaded from the Xena Browser from the University of California Santa Cruz Computational Genomics Lab. The number of samples and histologies included from each data source are summarized in Table 3-1.

3.3.6 In-silico immune deconvolution

Six in-silico immune deconvolution tools (MCP Counter, EPIC, CIBERSORTx, TIMER, quanTIseq, xCell) from R-package software were applied to the RNA-Seq data, using tpm or log2(tpm+1) data as appropriate[4–10]. MCP Counter was run using HUGO gene names and default probe sets and genes. CIBERSORTx deconvolution was run against the LM22 signature matrix files. Settings included no batch correction, disabled quantile normalization, 500 permutations, and absolute mode. xCell was run using default settings. EPIC, quanTIseq, and TIMER were run using the immunedeconv package. TIMER was run using the “sarc” cancer type setting.

Samples were defined as high, medium and low immune cell groups based on the tercile of the total immune score for each in-silico immune deconvolution tool, as defined in Table 3-2. A "Consensus Group" was defined using the following method: (a) a sample was labeled "high" if it was in the high group in four of six of the tools listed above and never low, (b) a sample was labeled "low" if it was in the low group in four of six of the tools listed above and never high, and (c) all remaining samples were labeled "medium."

3.3.7 Statistical analysis

The multiplex immunofluorescence (mIF) scores from the multiple cores from each sample were averaged to assign one score per sample. The cell type scores from each of the in-silico immune deconvolution methods were compared to the mIF results as noted in Table 3-2. In-silico scores that required summation were either scaled and then summed or the raw score was summed, as indicated based on the tool output. The correlation between mIF and in-silico scores was determined for each listed comparison (Table 3-2) using a linear fit model. Pearson correlation coefficients are noted. Wilcoxon rank-sum test was used to compare continuous variables across groups. Gene set enrichment analyses (GSEA) were performed using the Broad Institute GSEA Software (Version 4.2.3). Inputs included 1000 permutations, no collapse, and HALLMARK gene sets. Principal component analyses (PCAs) were performed on log2(tpm+1) data, using protein coding genes only. P values of <0.05 were deemed statistically significant. A FWER p value was used for GSEA. Otherwise, multiple hypothesis testing was performed using Bonferroni correction where applicable.

3.3.8 Sensitivity analysis

The correlation analyses mentioned above were performed using the average mIF score per sample as well as treating each core individually (i.e. up to three cores or mIF scores per sample). The results were similar using both methods (Table 3-3, Table 3-12). All the analyses were also performed using log of the mIF scores. Again, the results were essentially unchanged when using log scores when compared to raw scores (data not shown).

3.3.9 Software

The majority of statistical analyses and data visualizations were performed and generated in R (Version 4.0.5). Gene set enrichment analyses (GSEA) were performed using Broad Institute GSEA software (Version 4.2.3).

3.3.10 Approval

This study was approved by the UCLA Institutional Review Board (IRB #10-001857).

3.4 Results

3.4.1 Study cohort and sample characteristics

Sixty untreated undifferentiated sarcoma (USARC) samples were identified for inclusion in this study as demonstrated in Figure 3-1A and 3-1B. All 60 samples were included on the TMA, but mIF staining failed on 2 samples. Four samples were determined to be outliers on RNA-Seq and were excluded from the analysis. There was no overlap between the mIF failures and the RNA-Seq outliers. Thus, there were 54 samples included in the correlation analysis.

All subtypes of undifferentiated sarcoma, including ovoid, pleomorphic, spindled, and epithelioid, were included in the analysis. Representative H&E images of each subtype are shown in Figure 3-1C and the tissue microarray (TMA) is shown in Figure 3-2A. A principal component analysis of RNA-Seq data from multiple sarcoma subtypes and datasets is shown is Figure 3-1D. The comparison of PC2 and PC3 is shown and demonstrates that UCLA USARC samples cluster with USARC samples from other publicly available datasets.

The graph showing PC1 versus PC2 is shown in Figure 3-5. PC1 appears to capture the UCLA tumors. The statistically significant and largest magnitude enrichment scores (ES) results of the gene set enrichment analysis (GSEA) performed on PC1 are shown. This summarizes the difference between UCLA and the other study cohorts. Datasets included in these PCAs are summarized in Table 3-1.

3.4.2 Summary of multiplex immunofluorescence results

The multiplex immunofluorescence stains that performed the most reliably were CD45, CD68 and CD8, while CD20 and CD4 performed poorly. Specifically, there was relatively high

concordance between CD45, CD68 and CD8 between the mIF and the in-silico results. While the CD20 and CD4 mIF resulted in a paucity of stains (Figure 3-4). Triplicates of cores from each sample were included in on the TMA. The scores from each triplicate were similar overall, as shown in Figure 3-2C-G.

3.4.3 There is substantial discordance between various in-silico immune deconvolution tools in undifferentiated sarcoma

There is significant variability in the results of the various in-silico immune deconvolution tools when applied to USARC. This was true for both the categorical (tercile) results of total immune cell infiltration as well as the continuous data for each subtype.

The alluvial plot in Figure 3-3A demonstrates the correlation between the high, medium, and low total immune score terciles from each in-silico immune deconvolution tool and the Consensus Group, as defined above. Overall, there is significant change in the categorization of each sample across each of these different tools. Less than 20% of samples demonstrated concordance in the categorical (high versus medium versus low immune cell infiltration) group that they were assigned to across the six in-silico tools and the consensus group.

Figure 3-3B, 3-3C, and 3-3D show the summary of results of CIBERSORTx, EPIC, and quanTIseq, respectively. Cell types are summed as noted in Table 3-2. The results of the overall immune cell landscape as defined with each of these tools again showed significant variability across each of the tools. For example, the results of CIBERSORTx and EPIC suggest that CD4+

T Cells are the most abundant immune cell type in USARC, while quanTIseq does not. The results from quanTIseq suggest that B cells or macrophages are the most abundant immune cell type, while CIBERSORTx suggests that these are of moderate abundance, and the results of EPIC suggest they are relatively scarce. The results MCP Counter, TIMER, and xCell were not included in the comparison of the different cell lineages, as the score outputs from these tools are scaled, thus different cell types cannot be compared[10].

3.4.4 TIMER is the optimal tool for determining overall immune cell infiltration when using in-silico immune deconvolution tools in undifferentiated sarcoma

Figure 3-4 and Table 3-3 summarize the results of the correlation between mIF scores to the in-silico immune deconvolution scores for each tool and each cell subtype, as listed in Table 3-2. For overall immune cell infiltration, TIMER showed the highest correlation with the CD45+ mIF results (correlation coefficient 0.44). For the CD68+CD45+ results, MCP Counter monocyte score and the macrophage scores from the remainder of the tools performed well (correlation coefficient range 0.31-0.48). For CD8+CD45+ mIF results, CIBERSORTx and xCell CD8 T Cells scores show the highest correlation (correlation coefficients 0.34 and 0.27, respectively). However, the sequencing data identifies some 'high' T cell samples where the TMA does not. Comparisons of CD20+CD45+ and CD4+CD45+ mIF scores to B Cell and CD4+ T Cell in-silico scores showed overall poor concordance (correlation coefficient ranges -0.20-0.03 and -0.16-0.22, respectively). There were again many samples were the sequencing data identified 'high' B Cell or CD4 T Cell scores where the TMA mIF data did not.

3.5 Discussion

Immune cell infiltration has been found to be an important factor associated with survival or treatment response in various cancers, including sarcoma[1–3]. Notably, Petitprez et al found that B cells are associated with survival and immunotherapy response in sarcoma. Chen et al aimed to define factors in the undifferentiated sarcoma tumor microenvironment that might influence immunotherapy response . They described an immunosuppressive TME driven by TAMs. However, there is no consensus about the optimal in-silico tool that should be applied to sarcoma. For example, Petitprez et al used MCP Counter while Chen et al used CIBERSORTx. Further, many studies that validate the in-silico immune deconvolution tools do not include sarcoma[4–10].

In this study, we aimed to define the optimal in-silico immune deconvolution tool for undifferentiated sarcoma. We found that CD45+ multiplex immunofluorescence (mIF) results showed the highest correlation with total immune cell infiltration scores from TIMER in-silico immune deconvolution. There was relatively high correlation between CD68+CD45+ mIF results and monocyte (MCP Counter) or macrophage (CIBERSORTx, EPIC, quanTIseq, TIMER, xCell) immune deconvolution scores as well as between CD8+CD45+ mIF results and CD8+ T cell results from CIBERSORTx. There was very poor correlation between mIF and in-silico immune cell scores for B cells and CD4 + T cells. Further, the immune cell landscape of undifferentiated sarcoma defined by various in-silico immune deconvolution tools varies significantly across the different in-silico immune deconvolution tools used.

These results suggest that in-silico immune deconvolution can be used to define the immune cell populations in undifferentiated sarcoma (USARC), however, these tools should be applied to this tumor type in a judicious manner. In-silico immune deconvolution tools can define overall immune cell infiltration in undifferentiated sarcoma with relative success. However, the immune cell subtypes are more difficult to define accurately. Of the subtypes, the in-silico tools show the highest correlation with monocyte or macrophage scores, as determined by mIF. This may be because undifferentiated sarcoma is a relatively macrophage/monocyte-rich tumor and/or it may be because there is a high degree of similarity between macrophages/monocytes and sarcoma cells themselves[27-28]. Further, our flow cytometry results suggest there is a relative paucity of B cells, which may contribute to the overall poor performance of in-silico immune deconvolution tools for defining these subtypes.

Leveraging sequencing data and in-silico immune deconvolution tools is particularly important when studying rare diseases, such as sarcoma. Flow cytometry and immunohistochemistry (IHC) are frequently used to characterize the immune composition of tumors, however, they require large amounts of fresh tumor (flow cytometry), are low-throughput (flow cytometry and IHC), and they lack of publicly available data (flow cytometry and IHC)[10–12]. Using tumor-derived RNA for immunologic analyses would circumvent the issues and allow researchers to leverage publicly available data. However, these tools must be applied in a thoughtful and accurate manner.

Based on the findings in this paper, we suggest the following practices when applying in-silico immune deconvolution tools to undifferentiated sarcoma: (1) Use TIMER to define overall

immune cell infiltration. (2) Use MCP counter to define monocyte infiltration or use CIBERSORTx, EPIC, quanTIseq, TIMER, or xCell to define macrophage infiltration. (3) Use caution when using in-silico immune deconvolution tools to define CD8+ T cell infiltration. CIBERSORTx most accurately defines CD8+ T cell immune infiltration, however, there are still many instances when tumors with high CD8+ T cell infiltration will be missed using this technique. (4) Avoid applying in-silico immune deconvolution results to define B cell or CD4+ T cell immune infiltration.

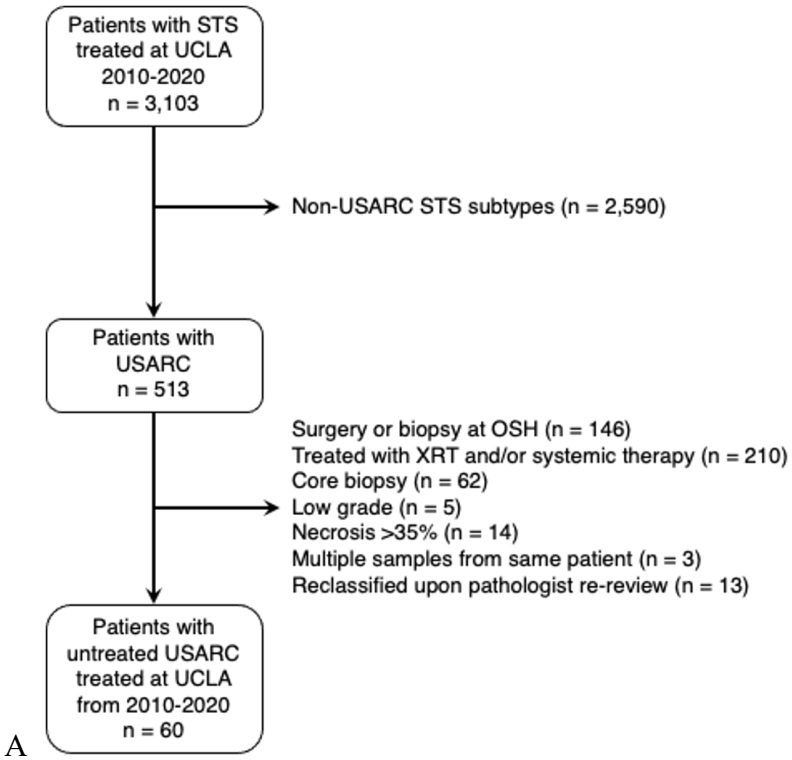
Limitations of this study include that tissue was obtained in a retrospective manner and that it was performed at a single institution. The samples were all formalin-fixed and paraffin embedded (FFPE) and stored for up to 10 years prior to nucleotide extraction. It is possible that this storage method had an impact on the results. Principal component analysis (PCA) was used to compare the in-house sequencing data to publicly available sequencing data from other sarcoma subtypes. While principal component (PC) 2 and PC3 showed that our in-house USARC data clustered with other USARC subtypes (Figure 3-1), PC1 versus PC2 showed that our in-house samples clustered together (Figure 3-5A), suggesting there are differences between the in-house RNA-Seq data and publicly available subtypes. These differences could include extraction techniques, sequencing techniques, duration of sample storage, and method of sample storage. Another possibility is that our samples selected areas of tumor and immune cell infiltration and a more directed way than the publicly available datasets. Our samples were highly curated and the portions of the tumor that were used to produce scrolls for nucleotide extraction and for inclusion on the tissue microarray (TMA) were selected deliberately by an expert sarcoma pathologist. All of these factors could limit the ability to extrapolate these findings to other undifferentiated

sarcoma samples. Finally, this project included only undifferentiated sarcoma samples. Thus, these findings may not be able to be extrapolated to other soft tissue sarcoma (STS) subtypes and further work is needed to define the optimal in-silico immune deconvolution tools for the application to STS more broadly. Further work should be done to refine the in-silico immune deconvolution tools to optimize them for application to STS.

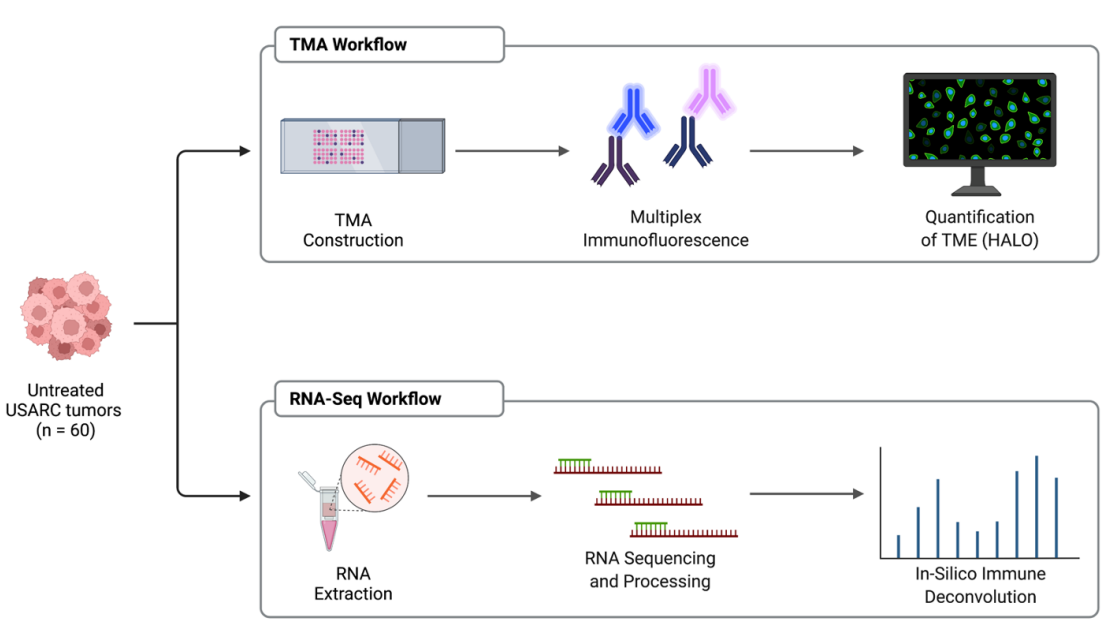
3.6 Conclusion

In this study, we aimed to determine the optimal in-silico immune deconvolution tool in undifferentiated sarcoma by determining the correlation between mIF and in-silico immune deconvolution scores. Based on our findings, we suggest the following practices when applying in-silico immune deconvolution tools to undifferentiated sarcoma: (1) Use TIMER to define overall immune cell infiltration. (2) Use MCP counter to define monocyte infiltration or use CIBERSORTx, EPIC, quanTIseq, TIMER, or xCell to define macrophage infiltration. (3) Use caution when using in-silico immune deconvolution tools to define CD8+ T cell infiltration. CIBERSORTx most accurately defines CD8+ T cell immune infiltration, however, there are still many instances when tumors with high CD8+ T cell infiltration will be missed using this technique. (4) Avoid applying in-silico immune deconvolution results to define B cell or CD4+ T cell immune infiltration.

3.7 Figures

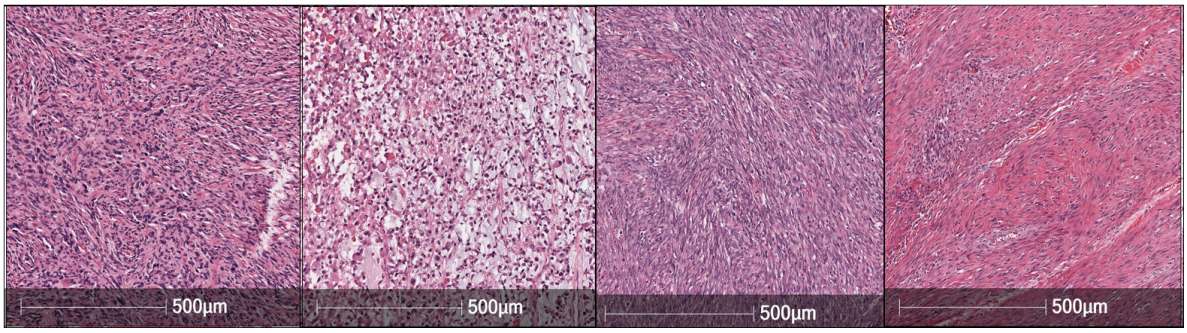

Patients with STS treated at UCLA 2010-2020 n = 3,103
Non-USARC STS subtypes (n = 2,590)
Patients with USARC n = 513
Surgery or biopsy at OSH (n = 146)
Treated with XRT and/or systemic therapy (n = 210)
Core biopsy (n = 62)
Low grade (n = 5)
Necrosis >35% (n = 14)
Multiple samples from same patient (n = 3)
Reclassified upon pathologist re-review (n = 13)
Patients with untreated USARC treated at UCLA from 2010-2020 n = 60

A


TMA Workflow
TMA Construction
Multiplex Immunofluorescence
Quantification of TME (HALO)
Untreated USARC tumors (n = 60)
RNA-Seq Workflow
RNA Extraction
RNA Sequencing and Processing
In-Silico Immune Deconvolution

B

C



D

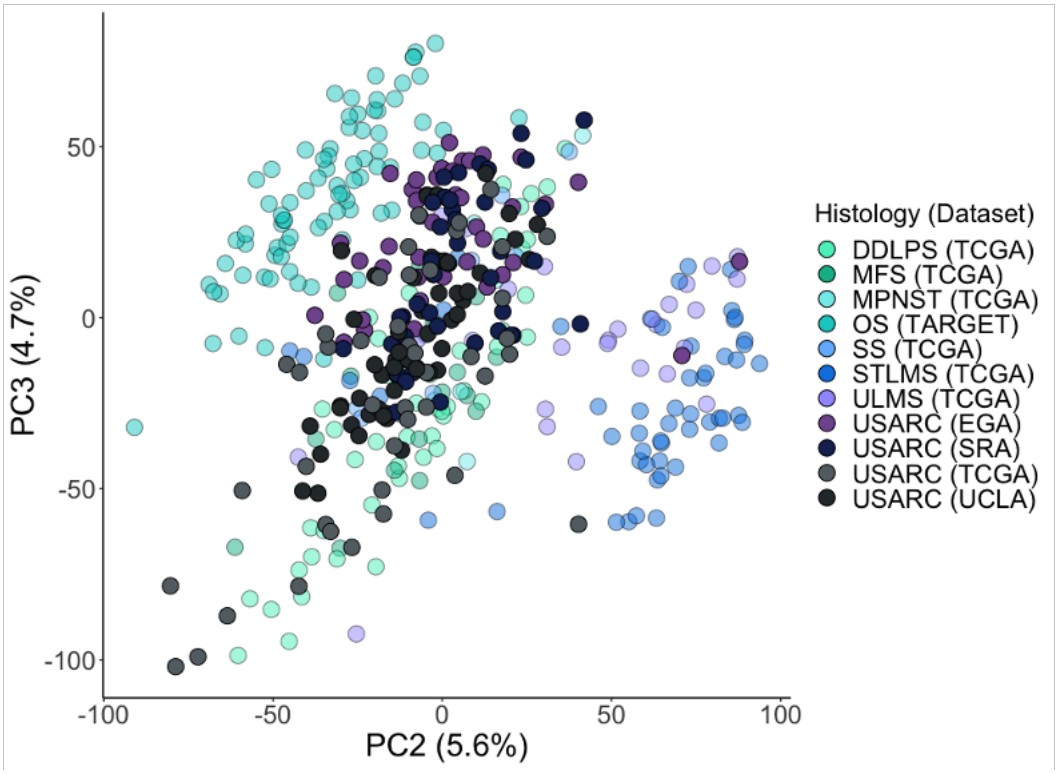
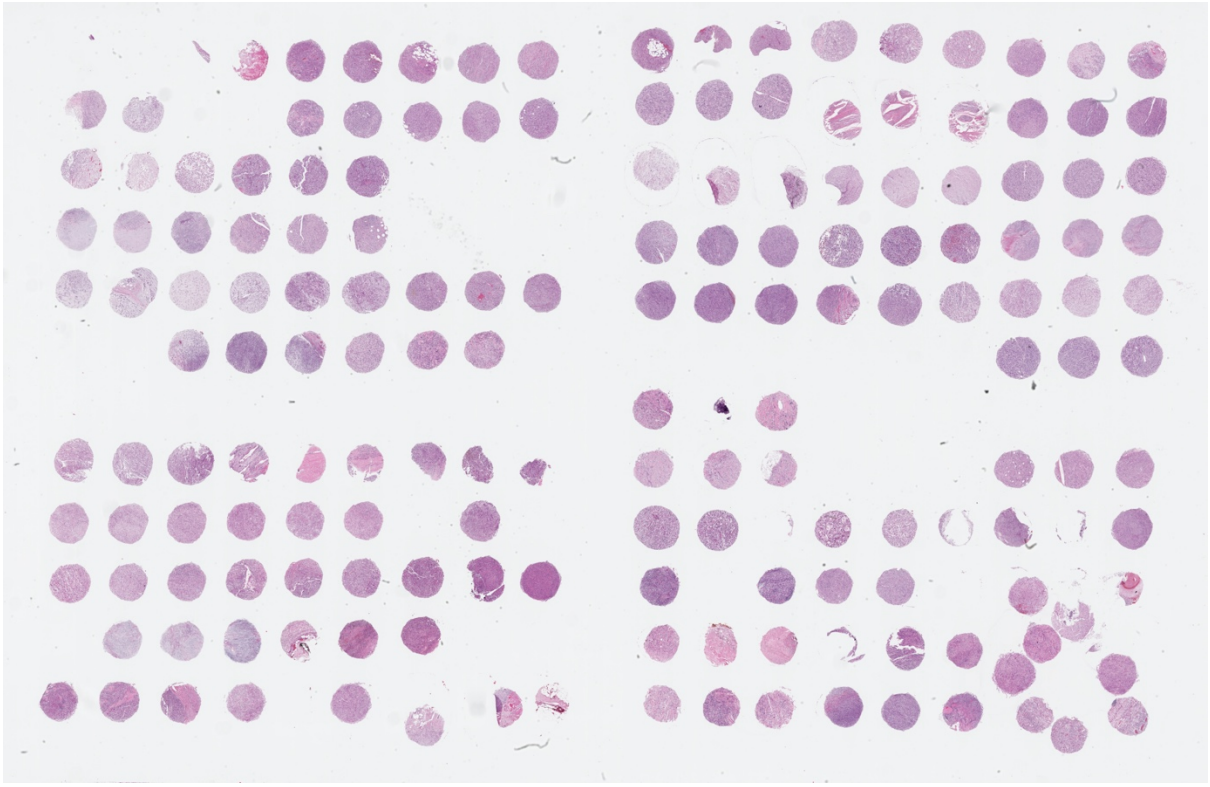
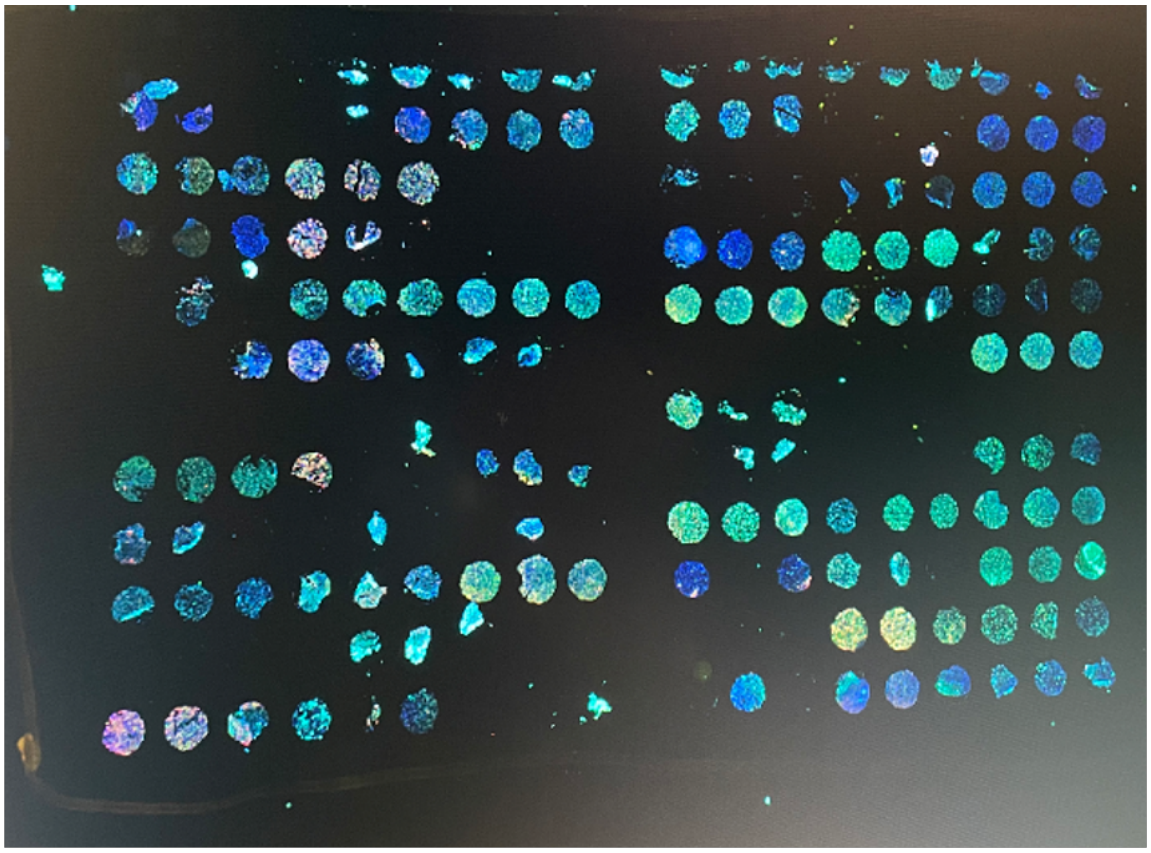


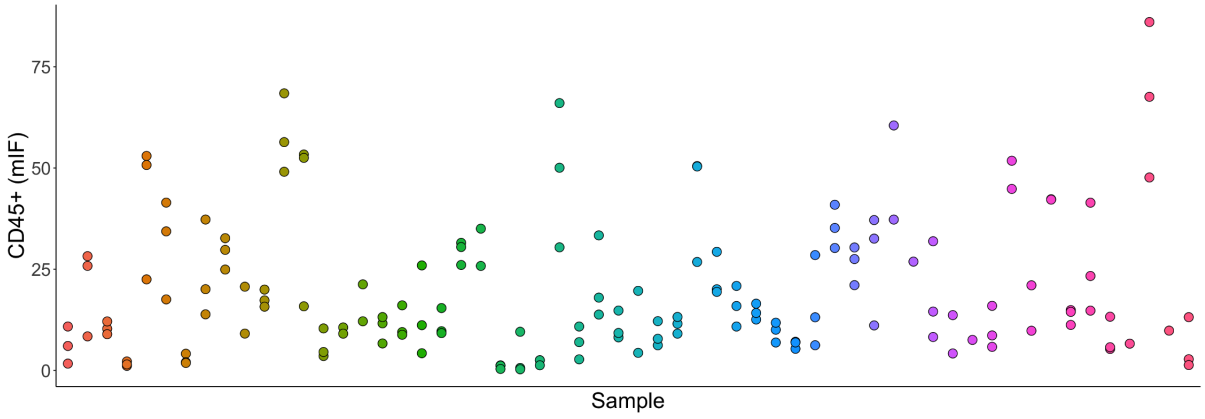
Figure 3-1: Study population and experimental design. (A) CONSORT diagram showing study population of 60 untreated USARC tumors. (B) Schematic of experimental design and analysis. All tumors were re-reviewed by an expert sarcoma pathologist. Three cores from selected sites from each tumor were used to create the TMA. RNA was extracted from FFPE-preserved samples of the same 60 tumors, the RNA was sequenced, and Toil processing was performed. Immune cell quantification was performed using mIF (CD45, CD68, CD4, CD8, CD20) on the

TMA and multiple in-silico immune deconvolution tools (CIBERSORTx, EPIC, MCP Counter, TIMER, quanTIseq, xCell) on the RNA-Seq data. (C) Representative images of the USARC tumors included in the TMA. From left to right, tumors are classified as spindled/pleomorphic, ovoid/epithelioid/focal pleomorphic, spindled, and ovoid, respectively. (D) PCA of RNA-Seq data from sarcoma tumors from multiple publicly available datasets (EGA, SRA, TARGET, TCGA) and UCLA USARC tumor samples. USARC tumors from multiple datasets cluster together. PCA run using protein coding genes.

A 

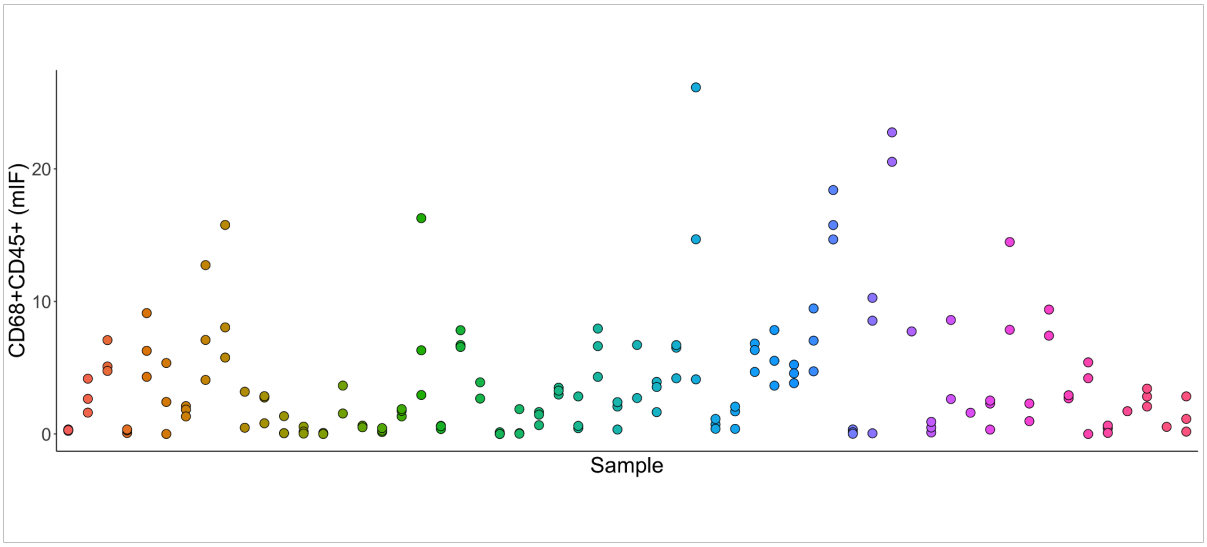
B



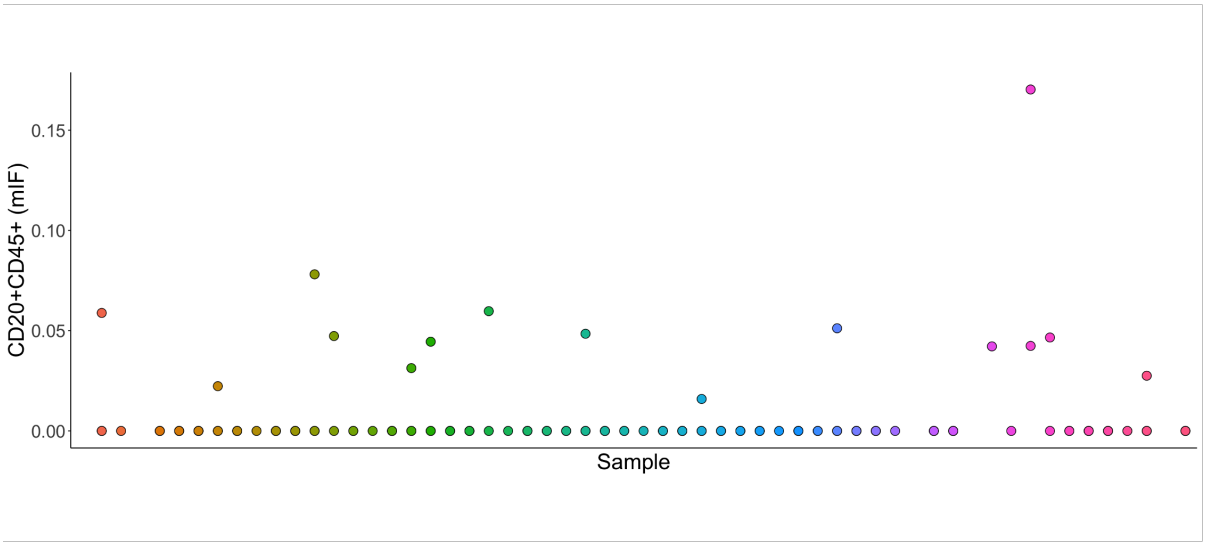


C

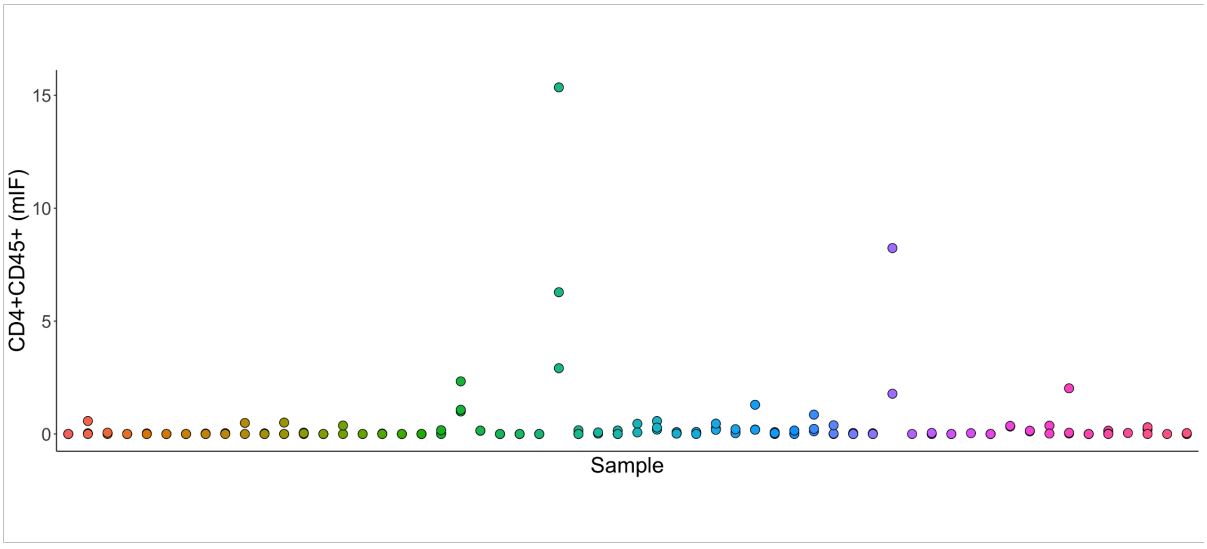
D

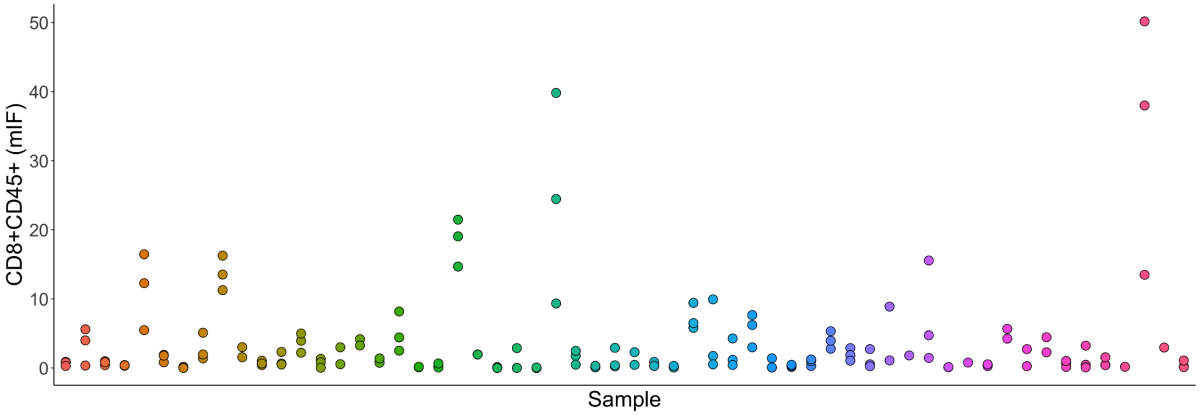

CD68+CD45+ (mIF)
20
10
0
Sample

E


CD20+CD45+ (mIF)
0.15
0.10
0.05
0.00
Sample

F

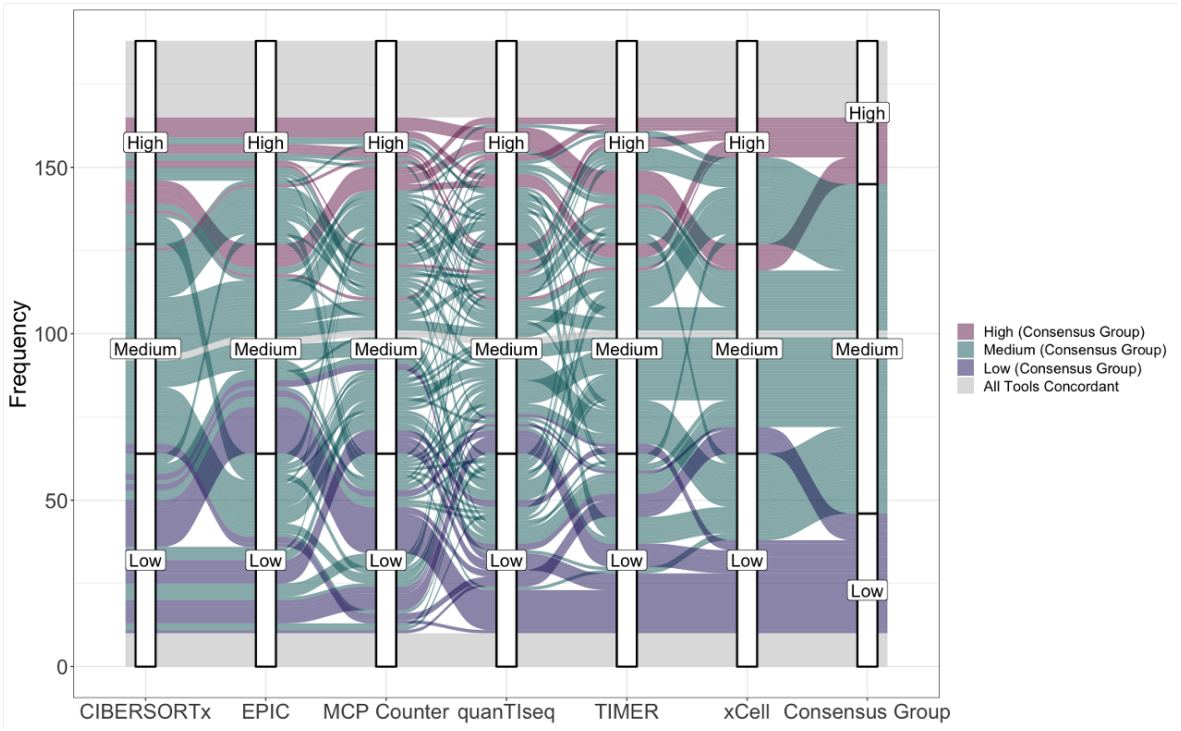

CD4+CD45+ (mIF)
15
10
5
0
Sample



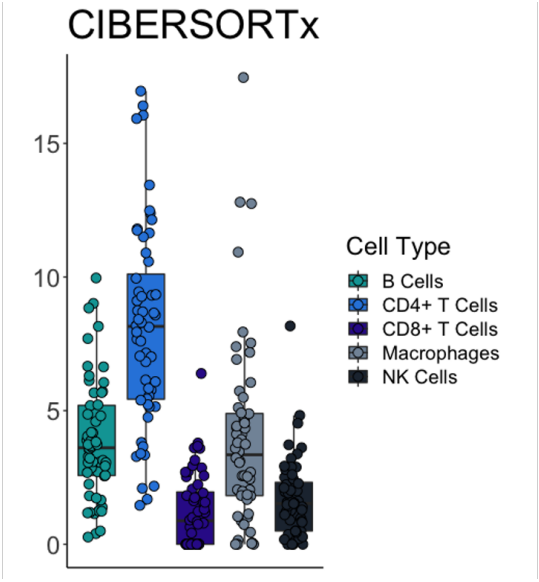
G

Figure 3-2: USARC TMA H&E and mIF analysis. CD8, CD68, and CD45 mIF performed reliably and consistently between the triplicates from each tumor. CD20 and CD4 mIF performed poorly. (A) H&E stain of the TMA with triplicate cores of 60 untreated USARC tumors from patients treated at UCLA. (B) A representative image of the multiplex mIF of the USARC TMA. (C-G) mIF results from the triplicate cores from each tumor. The y-axis represents CD45+ (C), CD68+CD45+ (D), CD20+CD45+ (E), CD4+CD45+ (F), and CD8+CD45+ (G) scores, respectively.

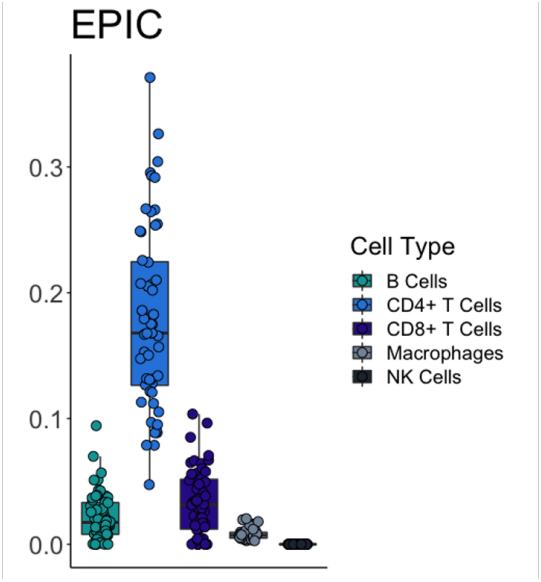
A


Frequency
150
100
50
0
High
Medium
Low
CIBERSORTx
EPIC
MCP Counter
quanTIseq
TIMER
xCell
Consensus Group
High (Consensus Group)
Medium (Consensus Group)
Low (Consensus Group)
All Tools Concordant

B


CIBERSORTx
15
10
5
0
Cell Type
B Cells
CD4+ T Cells
CD8+ T Cells
Macrophages
NK Cells

C


EPIC
0.3
0.2
0.1
0.0
Cell Type
B Cells
CD4+ T Cells
CD8+ T Cells
Macrophages
NK Cells

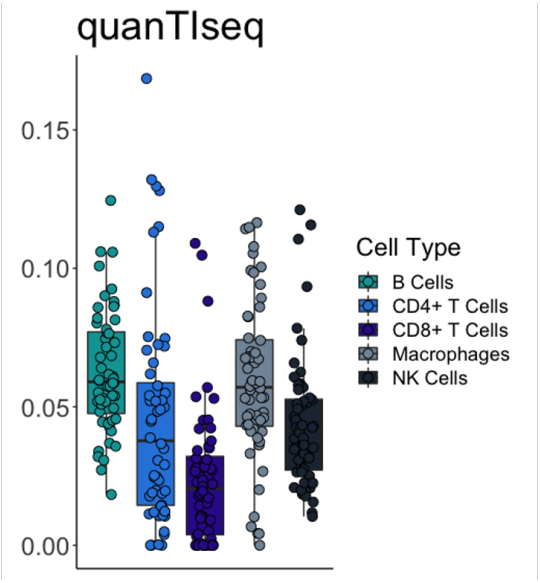
D 

Figure 3-3: There is discordance between the results of multiple in-silico immune deconvolution methods. (A) Alluvial plot demonstrating low, medium, and high samples, defined by terciles, of each of the in-silico immune deconvolution tools (CIBERSORTx, EPIC, MCP Counter, quanTIseq, TIMER, xCell). Consensus Group is a composite score developed based on the results of the six listed in-silico deconvolution methods, as described in the methods. Samples are colored by the Consensus Group tercile, and samples that were in the same group using all six tools are grey. 18.6% of the samples were concordant across all tools. Of the 188 total samples, 23, 2, and 10, were all high, medium, and low, respectively, across all in-silico immune deconvolution tools. (B-D) The immune cell landscape predicted using CIBERSORTx (B), EPIC (C) and quanTIseq (D), respectively. Cell types were summed as listed in Table 3-2. MCP Counter, TIMER, and xCell are not shown as the scores derived are scaled scores thus different cell types cannot be compared. See Tables 3-8, 3-9, & 3-11 for tables summarizing B-D, respectively.

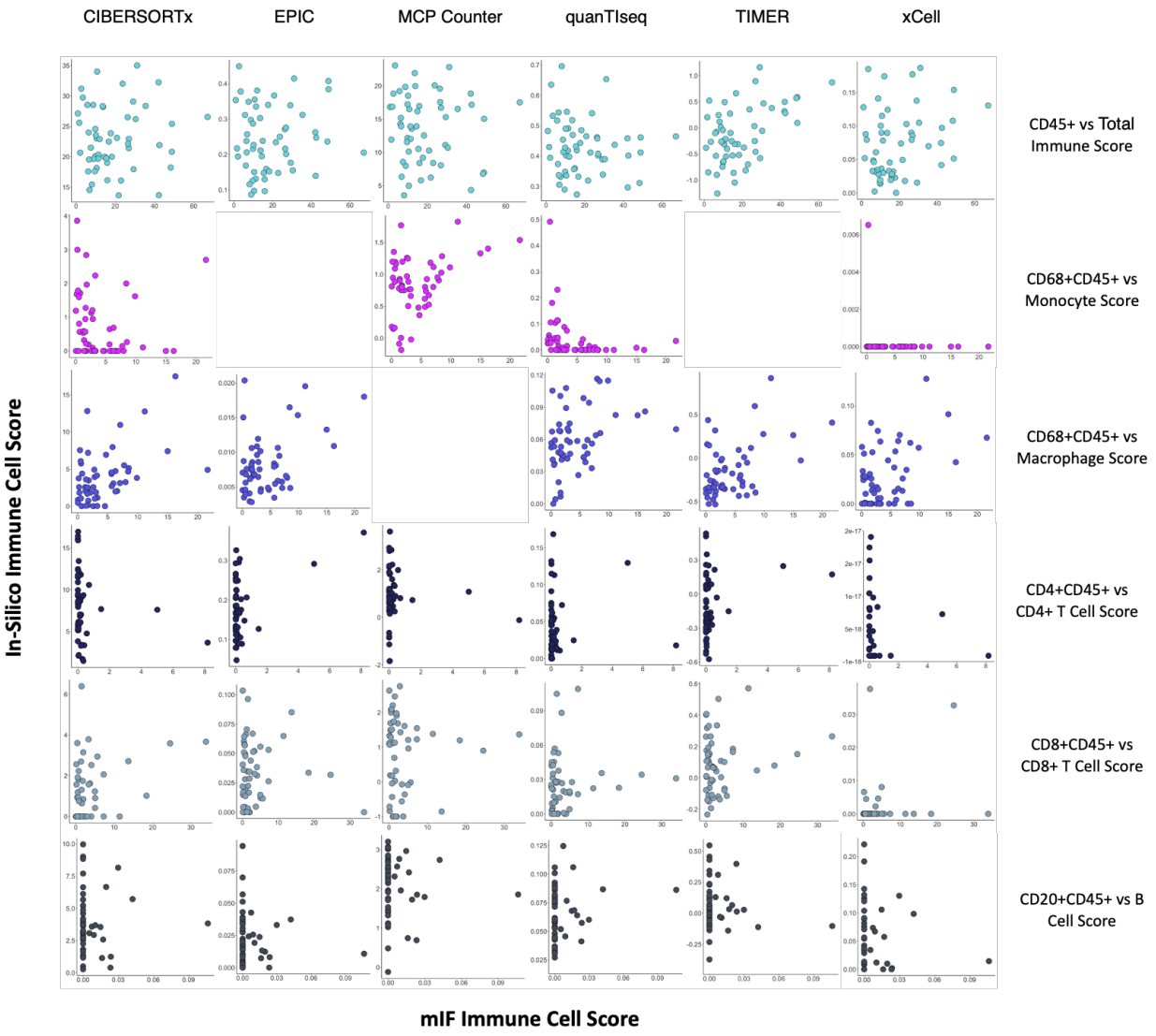


Figure 3-4: Different in-silico immune deconvolution tools demonstrate higher correlation with mIF results depending on the immune cell type. (A) Correlation between CD45 mIF score and in-silico Sum Immune Score. (B) Correlation between CD68+CD45+ mIF score and in-silico monocyte scores (when available). (C) Correlation between CD68+CD45+ mIF score and in-silico macrophage scores (when available). (D) Correlation between CD4+CD45+ mIF score and in-silico CD4 T Cell scores. (E) Correlation between CD8+CD45+ mIF score and in-silico CD8 T Cell scores. (F) Correlation between CD20+CD45+ mIF score and in-silico B Cell scores.

Tumors included in this analysis were 60 untreated USARCs (see Figure 3-1). The in-silico immune deconvolution tools used were CIBERSORTx, EPIC, MCP Counter, quanTIseq, TIMER, and xCell. X axis values are TMA mIF scores and y axis values are in-silico immune deconvolution tool results. Comparisons were made as listed in Table 3-2. Spearman correlation coefficients are listed in Table 3-3.

3.8 Tables

Dataset	**Histology**	**n**
EGA (EGAD00001004439)	USARC	48
SRA (GSE71121)	USARC	42
TARGET (dbGAP)	OS	81
TCGA (Xena Browser)	Multiple SARC subtypes	204
	DDLPS	*50*
	MFS	*17*
	MPNST	*5*
	SS	*9*
	STLMS	*53*
	ULMS	*27*
	UPS	*43*
UCLA	USARC	60

Table 3-1: Summary of datasets included in this analysis.

mIF Stain	**CIBERSORTx**	**EPIC**	**MCP Counter**	**quanTIseq**	**TIMER**	**xCell**
CD45	Sum Immune Score	Sum Immune Score*	Sum Immune Score*	Sum Immune Score*	Sum Immune Score*	Sum Immune Score
CD68+ CD45	Monocytes	-	Monocytic Lineage	Monocyte	-	Monocytes

CD68+ CD45	Macrophages M0 + Macrophages M1 + Macrophages M2	Macrophage	-	Macrophage M1 + Macrophage M2	Macrophage	Macrophages
CD4+ CD45	T Cells CD4 Naive + T Cells CD4 Memory Resting + T Cells CD4 Memory Activated + T Cells Follicular Helper + T Cells Regulatory Tregs	T Cell CD4	T Cells - CD8 T Cells	T Cell CD4 Non Regulatory + T Cell Regulatory Tregs	T Cells CD4	CD4 T Cells
CD8+ CD45	T Cells CD8	T Cell CD8	CD8 T Cells	T Cell CD8	T Cell CD8	CD8 T Cells
CD20+ CD45	B Cells Naive + B Cells Memory + Plasma Cells	B Cell	B Lineage	B Cell	B Cell	B Cells

Table 3-2: Pairwise comparisons between mIF stains and cell type scores from in-silico immune deconvolution. mIF, multiplex immunofluorescence. *Sum Immune Scores noted were calculated as the sum of all immune cell types scored by each tool.

mIF Stain	**CIBER-SORTx**	**EPIC**	**MCP Counter**	**quanTIseq**	**TIMER**	**xCell**
CD45 (vs Sum Immune Score)	-0.013 (0.92)	0.076 (0.57)	-0.12 (0.38)	-0.14 (0.29)	0.44 (<0.001)	0.14 (0.31)

CD68+CD45 (vs Monocytes)	-0.18 (0.17)	-	0.39 (0.0024)	-0.28 (0.031)	-	-0.11 (0.39)
CD68+CD45 (vs Macrophages)	0.48 (<0.001)	0.31 (0.019)	-	0.35 (0.0063)	0.41 (0.0013)	0.44 (<0.001)
CD4+CD45 (vs CD4 T Cell)	-0.16 (0.23)	0.39 (0.0023)	-0.072 (0.59)	0.070 (0.60)	0.22 (0.10)	-0.018 (0.89)
CD8+CD45 (vs CD8 T Cell)	0.34 (0.0091)	-0.020 (0.88)	0.024 (0.86)	0.11 (0.41)	0.16 (0.24)	0.27 (0.041)
CD20+CD45 (vs B Cell)	0.03 (0.83)	-0.087 (0.54)	-0.069 (0.62)	0.13 (0.34)	-0.20 (0.15)	-0.041 (0.77)

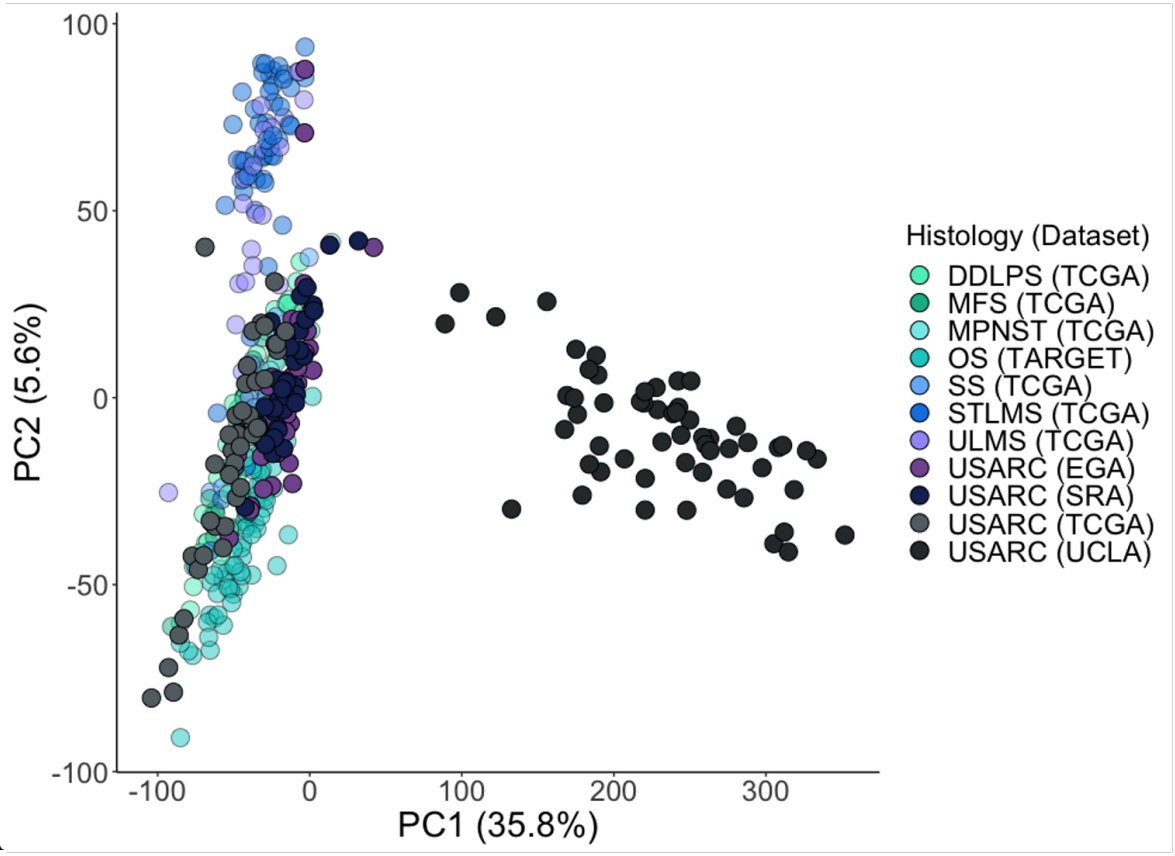
Table 3-3: Summary of Spearman correlation coefficients (R) and p values derived from the comparisons of mIF immune cell quantification and in-silico immune deconvolution scores of USARC tumors. Values shown are R (p value).

3.9 Appendix to Chapter 3

3.9.1 PC1 captures differences between RNA-Seq data from UCLA tumors compared to publicly available tumors

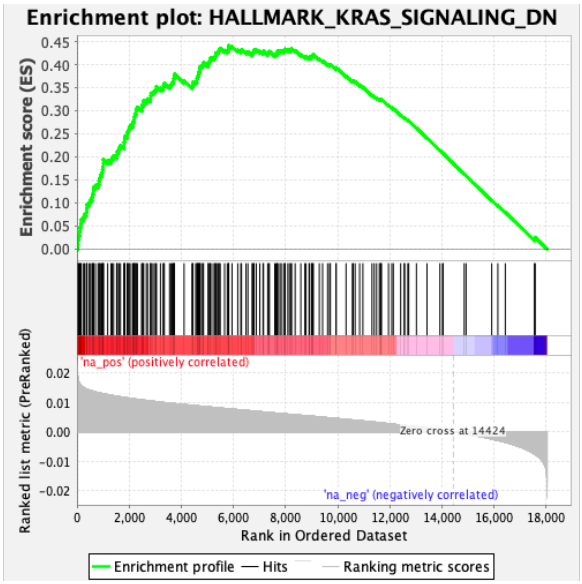
One of the potential limitations of this study is the differences seen between the UCLA RNA-Seq data and the publicly available RNA-Seq data. The PCA below demonstrates that the samples separate on PC1 primarily based on the dataset. This suggests that there are differences between the RNA-Seq data from the UCLA tumors when compared to the other tumors in the analysis. These differences could include extraction techniques, sequencing techniques, duration of sample storage, and method of sample storage. Another possibility is that our samples selected areas of tumor in a more directed way than the publicly available datasets, as described above.

The analyses below describe some of the differences between the UCLA samples when compared to the other publicly available samples. First, a PCA was performed, which demonstrated the clustering as described above. A GSEA was then performed on the genes driving PC1. Positive enrichment scores (ES) are associated with the UCLA tumors while the negative ES is associated with the publicly available tumors. The genes sets with the highest positive and negative ES are shown below. Finally, the top 100 genes driving positive PC1 scores and the top 100 genes driving negative PC1 scores (i.e. correlated the UCLA and publicly available samples, respectively) are listed below.



A

B



C

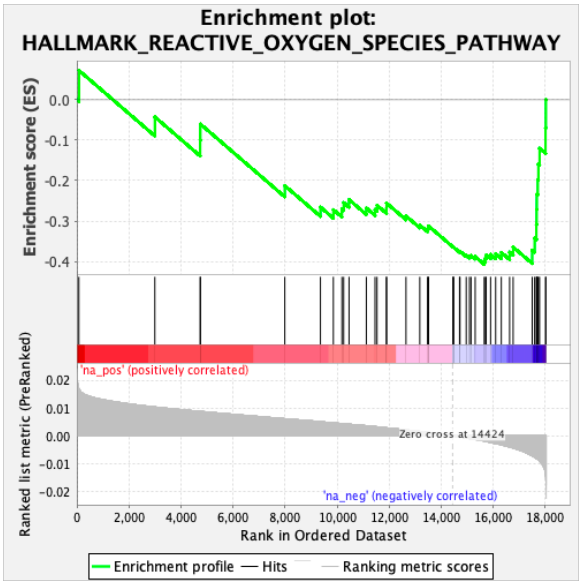


Figure 3-5: PC1 highlights differences between RNASeq from UCLA tumors versus RNASeq from publicly available sources (EGA, SRA, TARGET, and TCGA). (A) PCA of RNA-Seq data from sarcoma tumors from multiple publicly available datasets (EGA, SRA, TARGET, TCGA) and UCLA USARC tumor samples. UCLA USARC tumors cluster separately. PCA run using protein coding genes. (B-C) Select statistically significant (FEWR < 0.05) results of GSEA of ranked list of PC1 genes using HALLMARK gene sets. Positively scored genes and gene sets (B) are associated with UCLA tumors which negatively scored genes and gene sets (C) are associated with the other datasets (EGA, SRA, TARGET, TCGA).

TP53TG3	GOLGA6L4	CCDC144A	CEP295NL	TRIM49D1
OR4F17	CTAGE6	CP	MUC20	TP53TG3D
GRAPL	GOLGA8K	SPATA31A7	RIMBP3	ADGRF3
RGPD8	SPDYE2	KRT5	MBD3L2	PRSS48
STRC	KRT6A	ABCC8	FFAR3	TSSK4
TTN	TBC1D3B	GOLGA6L10	EGF	HYDIN
OR2A1	CYP21A2	PPIAL4F	MUC6	AMHR2
RGPD1	PPIAL4C	RGPD6	AGAP5	ZNF750

OR2A7	CDRT1	OR4F4	PCDH15	GPX2
CHRNA4	NPIPB8	FLG	FOXD4L6	AMT
RGPD2	CTAGE9	PPIP5K1	AGAP9	GRIN2A
HRNR	OR11H12	CLRN1	ANKRD36	NRXN1
CTAGE4	OR2T5	SPDYE2B	ANKRD36C	C9orf131
OR4F3	OR11H1	ALB	HSFX2	SEZ6
NPIPB4	SPATA31A3	TRPM1	GOLGA8N	CRB1
OR2T2	LYPD3	SLC35G6	CTAGE15	ALPK2
SPATA31A1	FOXD4L4	RGPD3	FAM86B1	GIPR
CNTNAP3	RGPD5	PPIAL4D	ANKRD20A1	CYP2F1
MBD3L3	GOLGA8R	PGA4	MPO	ARL17A
PGA3	ERBB3	PRAMEF15	NPIPB5	GP1BA

Table 3-4: Top 100 genes that are driving PC1 (i.e. associated with UCLA tumors).

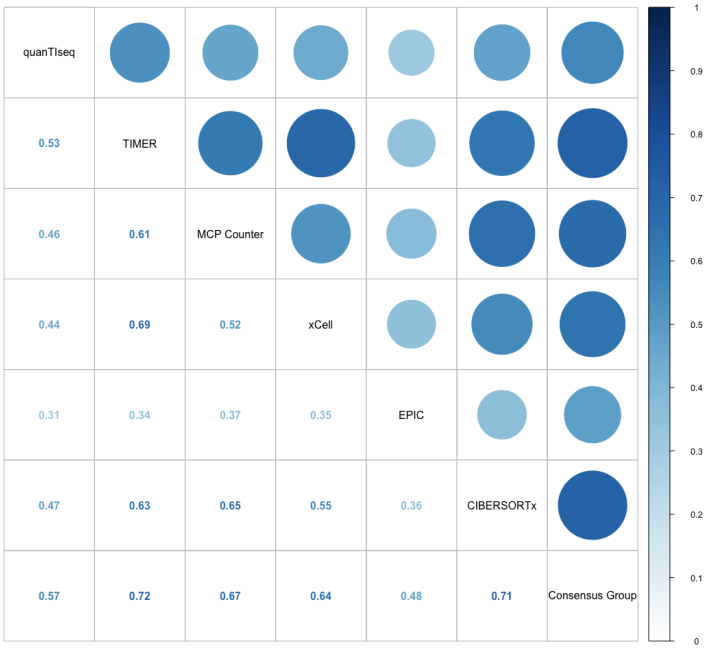
NME2	RPL38	SEC61G	FKBP2	FCER1G
IFI6	COX8A	PIGBOS1	ERH	PPIA
RPS21	UQCR11	RPL29	RPS9	MARCKSL1
LGALS1	PRDX1	PECAM1	BRK1	RPL30
HIGD1A	TMSB10	RPS27	S100A6	TSPO
RPL39	IGFBP4	POLR2L	UBL5	RPS10
COX6C	RPS14	DYNLRB1	GNG12	LAGE3
TOMM6	CREG1	C19orf53	BSG	CKS2
PTMS	SNRPB	S100A4	CAPNS1	COX4I1
TMSB4X	HLA-A	COX7C	NDUFA7	TAX1BP3
TXN	DNLZ	S100A11	GLO1	NDUFB1
MRPL12	RNF187	OST4	RPS3	MYL12B
GNG10	NDUFB7	HCST	C18orf32	VAMP8
ATP6V0E1	MANBAL	MT-ATP8	MT-CO2	MT-CO3
RPS19	LY6E	INAFM2	PSMB7	TMEM14A
PYURF	CRIP1	MIF	CMC4	ECHS1
TIMP2	CALM2	S100A13	TIMP3	EIF4EBP1
B2M	PSMA2	COX14	HMGN3	SMIM7
MYL9	RPS12	POMP	TUBB	MYL6
TMEM14C	CIB1	CDC42EP5	CST3	SNX3

Table 3-5: Top 100 genes that are driving negative PC1 (i.e. associated with publicly available USARC tumors from EGA, SRA, TCGA, and TARGET).

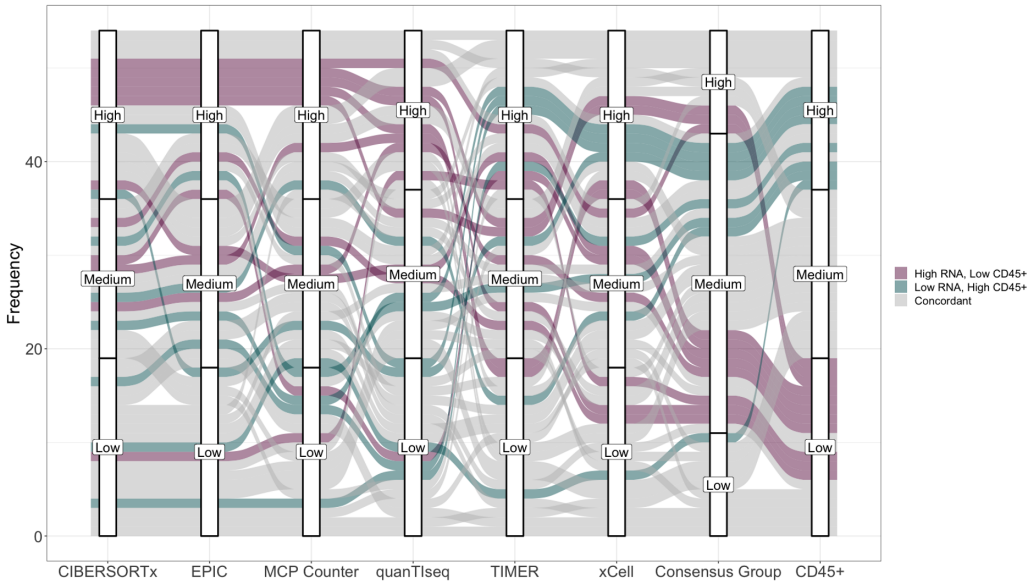
3.9.2 There is overall low correlation between various in-silico immune deconvolution methods when characterizing the tumor immune microenvironment in undifferentiated sarcoma

Figure 3-3 demonstrated that there is discordance between the results of multiple in-silico immune deconvolution methods. The findings here support the conclusions made in Figure 3-3 and provide additional more granular details. The alluvial plot below additionally includes the high, medium, and low terciles based on the mIF CD45+ score. The additional graphs below demonstrate the immune cell landscape as predicted by CIBERSORTx, EPIC, and quanTIseq immune deconvolution tools. All of the immune cell types described by these tools are included in these graphs, rather than just the sums of various cell types as shown in Figure 3-3.

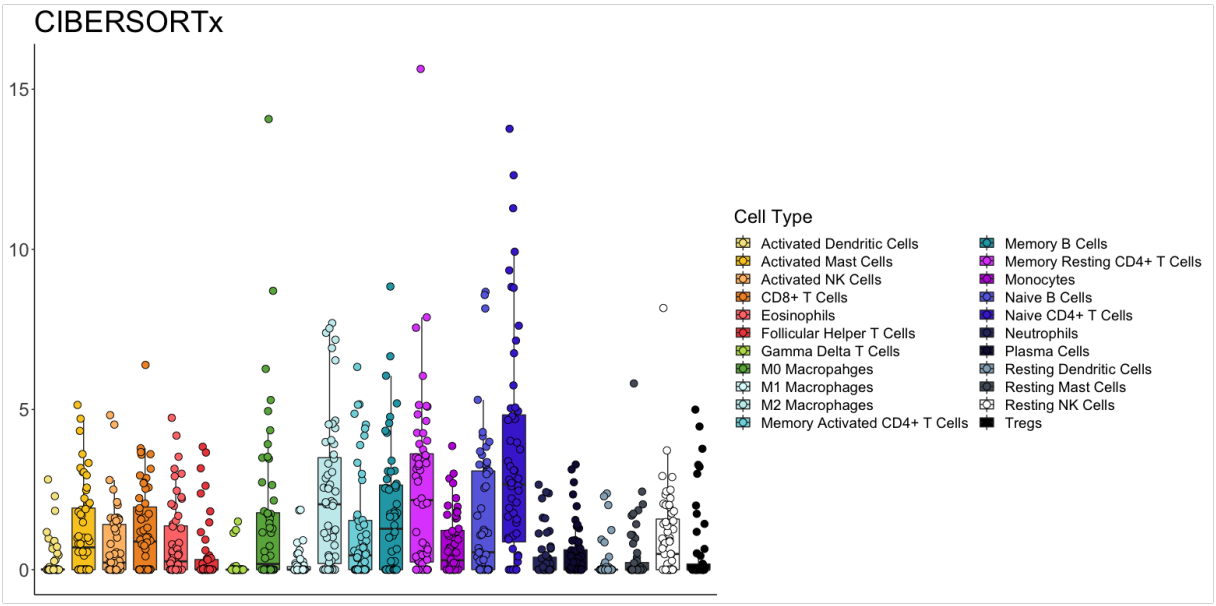
A



B


Frequency
0
20
40
High
Medium
Low
CIBERSORTx
EPIC
MCP Counter
quanTIseq
TIMER
xCell
Consensus Group
CD45+
High RNA, Low CD45+
Low RNA, High CD45+
Concordant

C


CIBERSORTx
0
5
10
15
Cell Type
Activated Dendritic Cells
Activated Mast Cells
Activated NK Cells
CD8+ T Cells
Eosinophils
Follicular Helper T Cells
Gamma Delta T Cells
M0 Macropahges
M1 Macrophages
M2 Macrophages
Memory Activated CD4+ T Cells
Memory B Cells
Memory Resting CD4+ T Cells
Monocytes
Naive B Cells
Naive CD4+ T Cells
Neutrophils
Plasma Cells
Resting Dendritic Cells
Resting Mast Cells
Resting NK Cells
Tregs

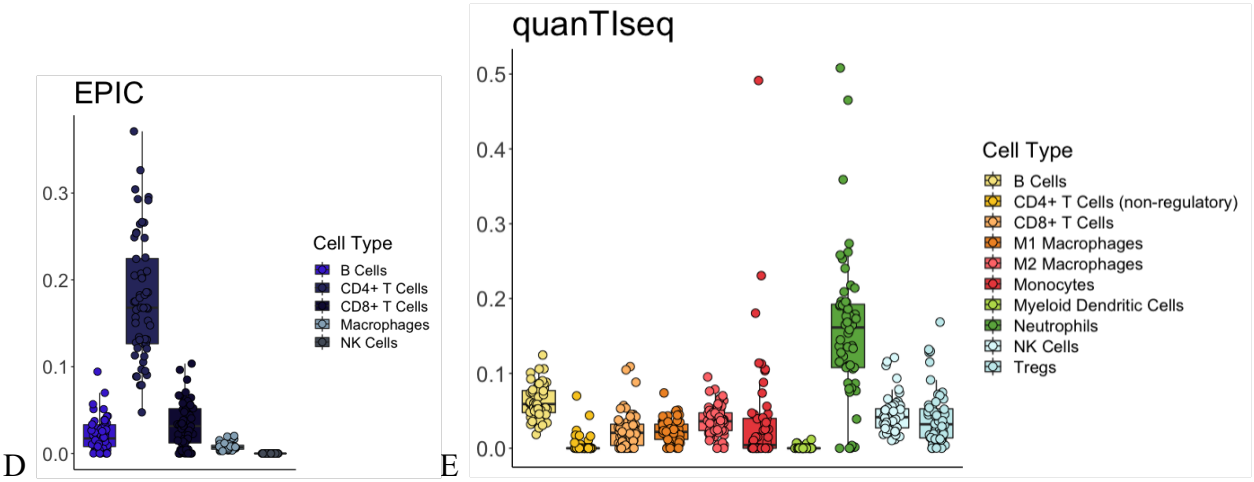


Figure 3-6: There is discordance between the results of multiple in-silico immune deconvolution methods. (B) A correlation plot demonstrating the pairwise correlation between the results of each of the six in-silico immune deconvolution tools as well as the Consensus Group. Values shown are the Cramer's V correlation coefficients between the results from each pair of tools. Larger and darker circles demonstrate higher correlation between samples. To calculate the immune groups shown, z-scores of each cell type were calculated and summed for each sample. The samples with in the highest, middle, and lowest tercile were labelled high, medium, and low, respectively. The Consensus Group score was high if four of the six tools were high and none were low. The Consensus Group score was low if four of the six tools were low and none were high. The remaining samples are medium. (A) Alluvial plot demonstrating low, medium, and high samples, defined by terciles, of each of the in-silico immune deconvolution tools (CIBERSORTx, EPIC, MCP Counter, quanTIseq, TIMER, xCell). Consensus Group is a composite score developed based on the results of the six listed in-silico deconvolution methods, as described in the methods. CD45+ results are from mIF data. Samples are colored by consensus between in-silico and mIF high versus low group definition. Samples are grey if they

were “high” based on CD45+ score and never “low” based on in-silico score or if they were “low” based on CD45+ score and never “high” based on in-silico score. (B-D) The immune cell landscape predicted using CIBERSORTx (B), EPIC (C) and quanTIseq (D), respectively. All cell types calculated by each of the respective tools are shown. MCP Counter, TIMER, and xCell are not shown as the scores derived are scaled scores thus different cell types cannot be compared.

In-silico Tool	High (in-silico tool) to low (CD45+)	Low (in-silico tool) to high (CD45+)
CIBERSORTx	6	3
EPIC	7	3
MCP Counter	6	4
quanTIseq	7	5
TIMER	4	2
xCell	4	1

Table 3-6: Data summarizing the alluvial plot in Figure 3-6B. The data above describes the number of samples that switched from high immune cell infiltration group defined by in-silico immune deconvolution tool terciles to low immune cell infiltration group defined by mIF CD45+ staining or any samples that switched from low to high, respectively.

Cell Type	Mean	Median	Range	Std
Activated Dendritic Cells	0.25	0.00	(0-2.81)	0.58
Activated Mast Cells	1.16	0.70	(0-5.14)	1.39
Activated NK Cells	0.73	0.22	(0-4.82)	1.08
CD8+ T Cells	1.20	0.88	(0-6.39)	1.40
Eosinophils	0.84	0.26	(0-4.74)	1.20
Follicular Helper T Cells	0.42	0.00	(0-3.84)	0.95
Gamma Delta T Cells	0.07	0.00	(0-1.50)	0.30
M0 Macrophages	1.51	0.18	(0-14.07)	2.54

M1 Macrophages	0.15	0.00	(0-1.87)	0.39
M2 Macrophages	2.27	2.04	(0-7.70)	2.27
Memory Activated CD4+ T Cells	1.19	0.45	(0-6.33)	1.70
Memory B Cells	1.72	1.28	(0-8.84)	2.00
Memory Resting CD4+ T Cells	2.48	2.17	(0-15.63)	2.76
Monocytes	0.76	0.29	(0-3.86)	0.95
Naive B Cells	1.68	0.54	(0-8.67)	2.24
Naive CD4+ T Cells	3.44	2.66	(0-13.76)	3.42
Neutrophils	0.38	0.00	(0-2.65)	0.70
Plasma Cells	0.48	0.00	(0-3.28)	0.82
Resting Dendritic Cells	0.19	0.00	(0-2.38)	0.54
Resting Mast Cells	0.40	0.00	(0-5.81)	0.96
Resting NK Cells	0.98	0.49	(0-8.17)	1.39
Tregs	0.55	0.00	0-5.00)	1.23

Table 3-7: Table summarizing values of CIBERSORTx results for UCLA USARC tumors. All immune cell types calculated by CIBERSORTx are shown.

Cell Type	**Mean**	**Median**	**Range**	**Std**
B Cells	3.88	3.61	(0.27-9.96)	2.24
CD4+ T Cells	8.09	8.15	(1.46-16.96)	3.76
CD8+ T Cells	1.20	0.88	(0-6.39)	1.40
Macrophages	3.94	3.35	(0-17.47)	3.45
NK Cells	1.72	1.53	(0-8.17)	1.47

Table 3-8: Table summarizing values of CIBERSORTx results for UCLA USARC tumors. Select and summed cell types (as shown in Figure 3-3) are shown.

Cell Type	**Mean**	**Median**	**Max**	**Std**
B Cells	0.02	0.02	(0-0.09)	0.02
CD4+ T Cells	0.18	0.17	(0.05-0.37)	0.07
CD8+ T Cells	0.03	0.03	(0-0.10)	0.03
Macrophages	0.01	0.01	(0-0.02)	0.00
NK Cells	0.00	0.00	(0-0)	0.00

Table 3-9: Table summarizing values of EPIC results for UCLA USARC tumors. All immune cell types calculated by EPIC are shown.

Cell Type	Mean	Median	Range	Std
B Cells	0.06	0.06	(0.02-0.12)	0.02
CD8+ T Cells	0.02	0.02	(0-0.11)	0.02
M1 Macrophages	0.02	0.02	(0-0.07)	0.02
M2 Macrophages	0.04	0.04	(0-0.10)	0.02
Monocytes	0.04	0.00	(0-0.49)	0.08
Myeloid dendritic cell	0.00	0.00	(0-0.01)	0.00
Neutrophil	0.16	0.16	(0-0.51)	0.10
NK Cells	0.04	0.04	(0.01-0.12)	0.02
T cell CD4+ (non-regulatory)	0.00	0.00	(0-0.07)	0.01
T cell regulatory (Tregs)	0.04	0.03	(0-0.17)	0.04

Table 3-10: Table summarizing values of quanTIseq results for UCLA USARC tumors. All immune cell types calculated by quanTIseq are shown.

Cell Type	Mean	Median	Range	Std
B Cells	0.06	0.06	(0.02-0.12)	0.02
CD4+ T Cells	0.04	0.04	(0-0.17)	0.04
CD8+ T Cells	0.02	0.02	(0-0.11)	0.02
Macrophages	0.06	0.06	(0-0.12)	0.03
NK Cells	0.04	0.04	(0.01-0.12)	0.02

Table 3-11: Table summarizing values of quanTIseq results for UCLA USARC tumors. Select and summed cell types (as shown in Figure 3-3) are shown.

3.9.3 Sensitivity analysis of USARC TMA data

In the primary analysis, the mIF scores were averaged across the triplicates of cores from each sample. The average of mIF scores was assigned to each sample and this value was compared to the in-silico immune deconvolution score for that sample. A sensitivity analysis was performed using the mIF score from each of the cores. The mIF score from each core was compared to the in-silico score from that core. The table below summarizes these results. Overall, the results were similar to those shown in Figure 3-4 and Table 3-3. The conclusions of the analysis would not change.

mIF Stain	**CIBER-SORTx**	**EPIC**	**MCP Counter**	**quanTIseq**	**TIMER**	**xCell**
CD45 (vs Sum Immune Score)	0.068 (0.42)	0.08 (0.34)	-0.02 (0.81)	-0.11 (0.20)	0.44 (3.8e-8)	0.19 (0.02)
CD68+CD45 (vs Monocytes)	-0.15 (0.07)	-	0.37 (4e-6)	-0.27 (0.001)	-	-0.12 (0.15)
CD68+CD45 (vs Macrophages)	0.48 (1.5e-9)	0.34 (3.7e-5)	-	0.33 (6.5e-5)	0.40 (7e-7)	0.39 (1.7e-6)
CD4+CD45 (vs CD4 T Cell)	-0.18 (0.03)	0.35 (1.9e-5)	-0.074 (0.38)	0.011 (0.90)	0.20 (0.014)	-0.48 (0.57)
CD8+CD45 (vs CD8 T Cell)	0.31 (<0.001)	-0.039 (0.64)	-0.024 (0.78)	0.081 (0.34)	0.16 (0.051)	0.32 (8.8e-5)
CD20+CD45 (vs B Cell)	0.0078 (0.93)	-0.091 (0.32)	-0.067 (0.47)	0.13 (0.15)	-0.11 (0.24)	-0.038 (0.68)

Table 3-12: Summary of Spearman correlation coefficients (R) and p values derived from the comparisons of mIF immune cell quantification and in-silico immune deconvolution scores of USARC tumors. Each sample was treated separately in this sensitivity analysis. Specifically, rather than averaging the scores across the triplicate cores for each sample (as was done above), each of the cores was treated separately. Values shown are R (p value).

3.9.4 Exploring the differences in TME between USARC cell lines, PDX models, and human tumors to better understand the limitation of in-silico immune deconvolution

RNA-Seq was performed on USARC cell lines and PDX models. Datasets and n are summarized in Table 3-7. All samples are included in the PCA below and only USARC samples are included in the in-silico immune deconvolution analysis. I hypothesized that the cell line RNA-Seq data would show the lowest immune scores using the in-silico immune deconvolution methods. I further expected that the PDX models will show a lower immune cell scores than the human tumors and especially low scores of T, B, and NK cells (as these tumors are generated in mouse models that lack mature T, B, and NK cells). However, I did expect there to be some immune cells identified using these tools. I predict this spillover effect, which is the detection of cell types due to similarities in marker genes between the tumor and immune cell types as described by Strum et al, will provide insights into the unique challenges when applying these in-silico tools to sarcoma tumors[10,11]. My predictions were generally seen in the data. Notably, EPIC demonstrated higher total immune cell infiltration seen in the cell lines and PDX data than in some of the tumor datasets, such as EGA and SRA. This is unsurprising, as EPIC showed the lowest correlation with the other in-silico tools as well as with the mIF results. Overall, EPIC performed poorly in sarcoma, so these results are not surprising. Rather, they support the findings above. The results below also provide further support for using the in-silico immune deconvolution tools to describe overall immune cell infiltration rather than to accurately quantify immune cell subtypes in sarcoma. Table 3 in the manuscript titled, "Comprehensive evaluation of transcriptome-based cell-type quantification methods for immune-oncology" by Sturm et al summarized important considerations and caveats when applying these tools[10]. Notably, all tools

used in this analysis can be used to compare across samples. CIBERSORTx (absolute mode), quanTIseq, and EPIC tools can be used to compare across cell types.

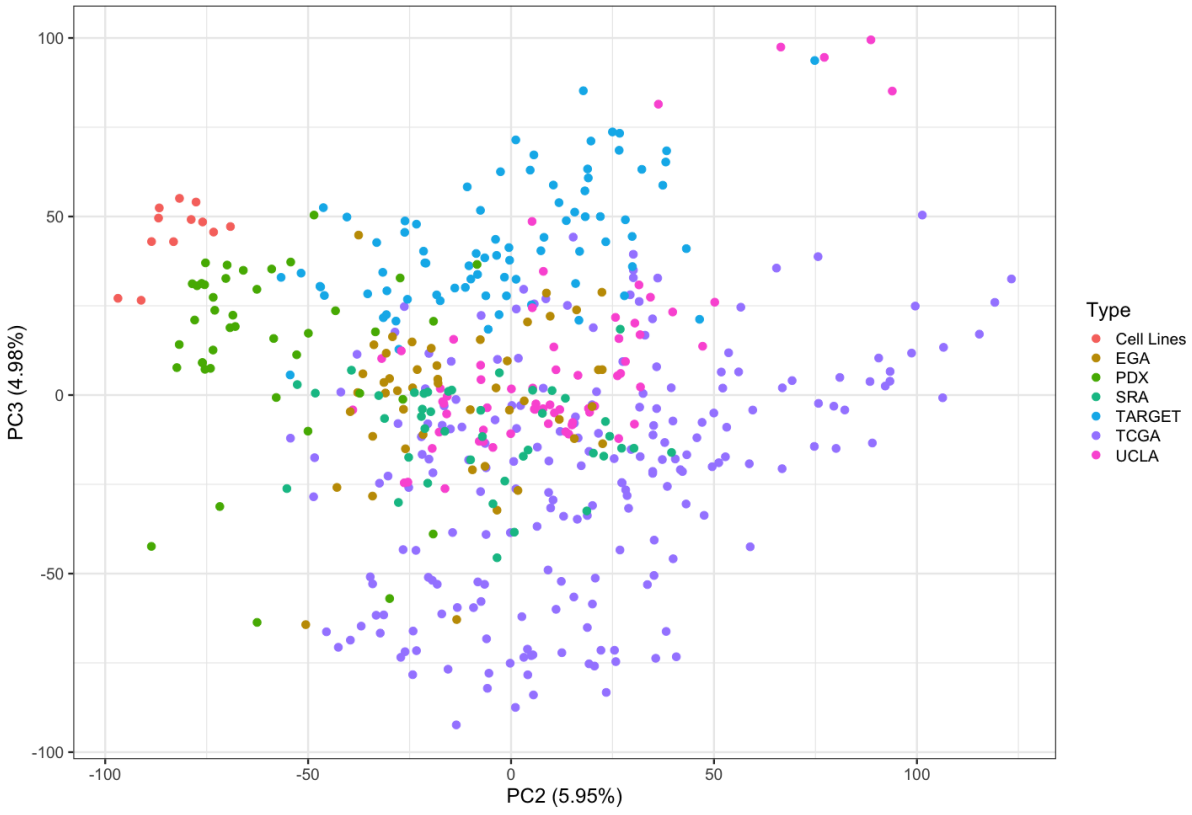
Dataset	Histology	n
Cell Lines	**Multiple sarcoma subtypes (pre/post BO112)**	**12**
	Fibrosarcoma	*1*
	GSC	*1*
	USARC	*4*
EGA (EGAD00001004439)	**USARC**	**48**
SRA (GSE71121)	**USARC**	**42**
TARGET (dbGAP)	**OS**	**81**
TCGA (Xena Browser)	**Multiple sarcoma subtypes**	**205**
	DDLPS	*50*
	MFS	*17*
	MPNST	*5*
	SS	*10*
	STLMS	*53*
	ULMS	*27*
	UPS	*43*
PDX	**Multiple sarcoma subtypes**	**42**
	DDLPS	*7*
	MFS	*5*
	OS	*2*
	SS	*2*
	STLMS	*5*
	ULMS	*1*
	USARC	*12*
	Other	*8*
UCLA	**USARC**	**64**
	Untreated	*60*
	Treated (pre/post immunotherapy pairs)	*4*

Table 3-13: A summary of the data sources, histologies, and sample sizes of the RNA-Seq data included in this analysis. All samples included in the PCA below. Only USARC samples (untreated cell lines and untreated UCLA human tumors only) are included in the in-silico

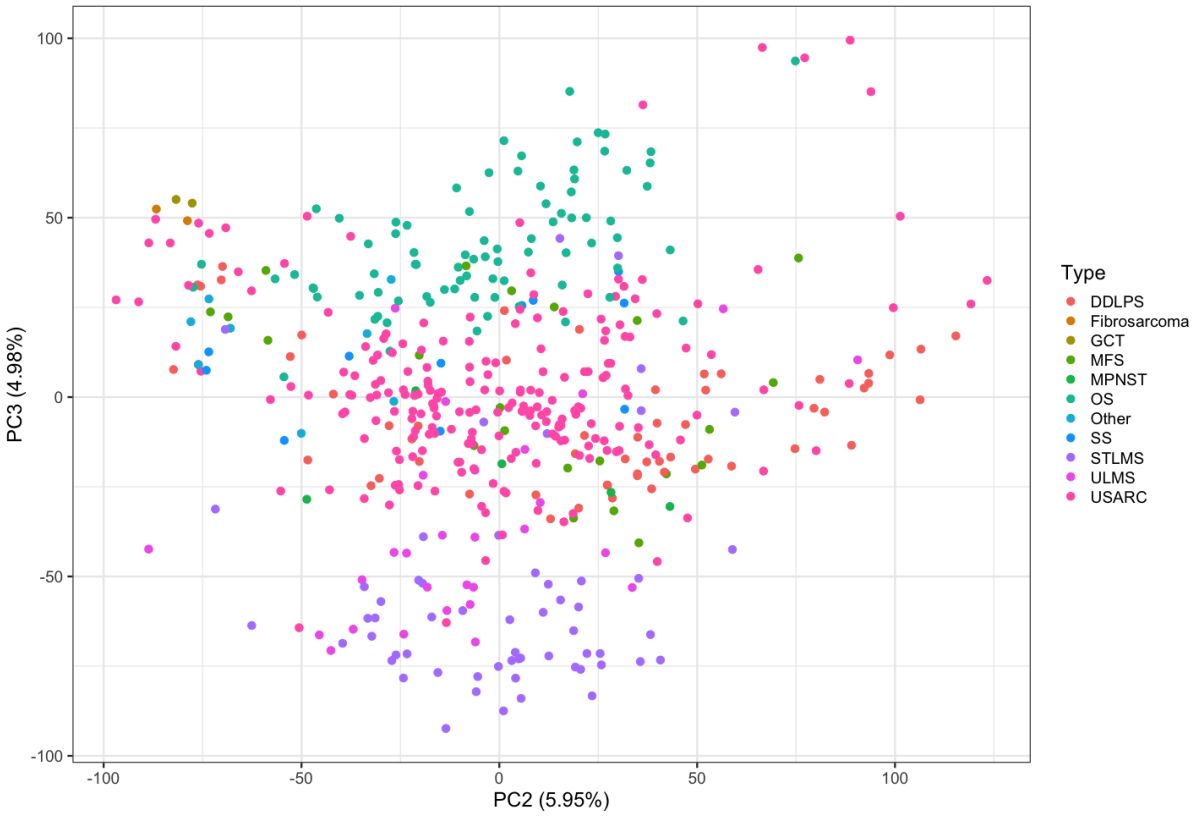
immune deconvolution analysis below. Other includes epithelioid sarcoma, Ewing's sarcoma, fibrosarcoma, myxoid LPS, FDCS, RMS.

Dataset	n
Cell Lines	4
EGA (EGAD00001004439)	47
SRA (GSE71121)	42
TCGA (Xena Browser)	43
PDX	12
UCLA	56

Table 3-14: A summary of the data sources and sample sizes of the RNA-Seq data included in the undifferentiated sarcoma in-silico immune deconvolution analysis. Outlier samples are removed from this table.



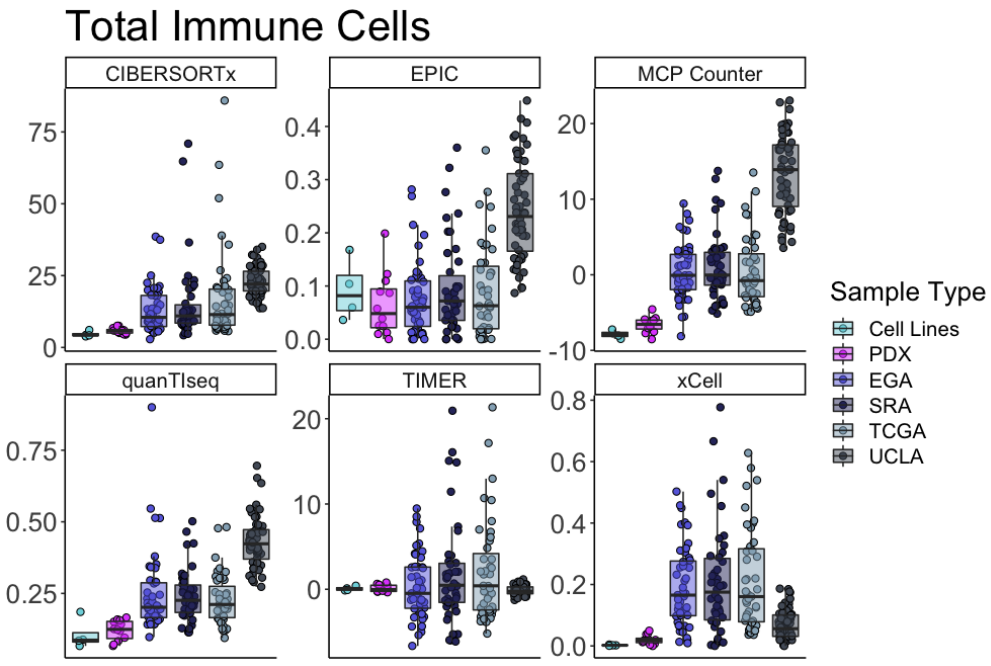
A



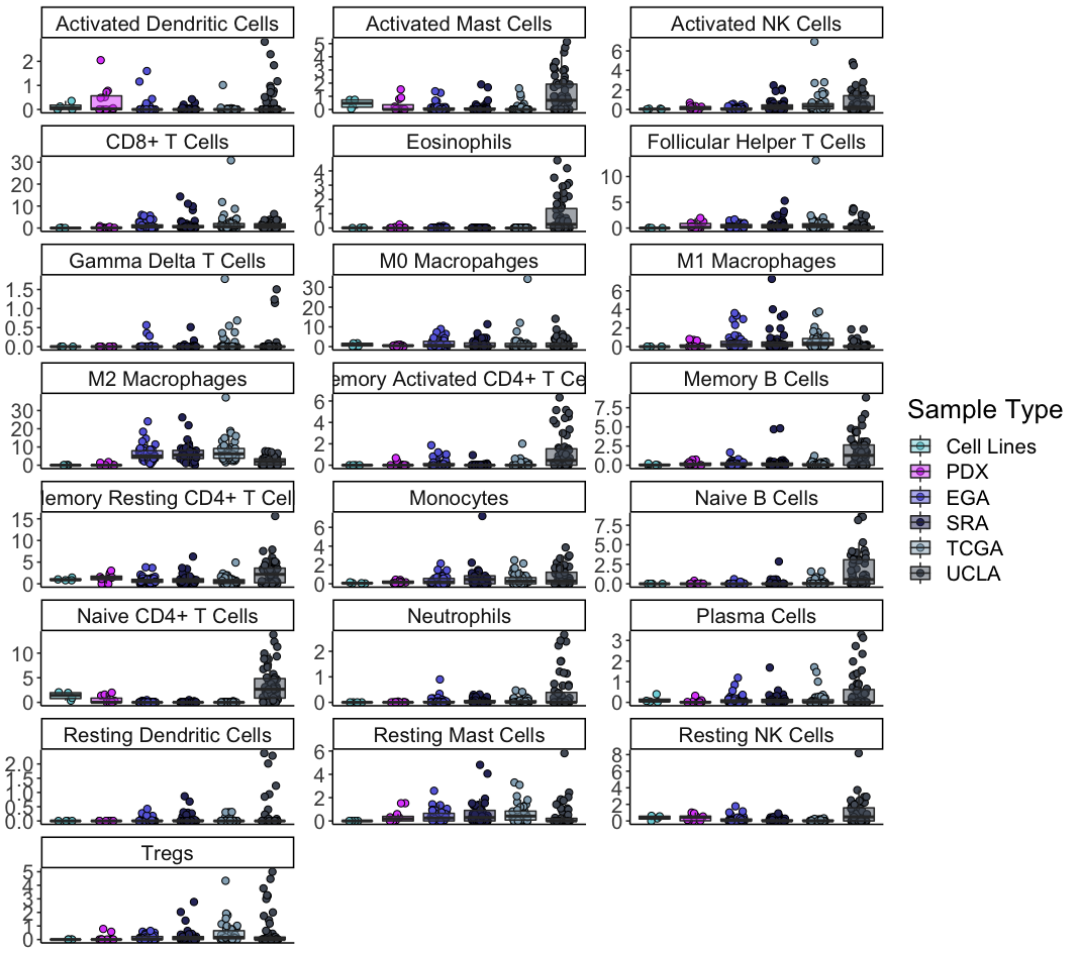
B

Figure 3-7: PCAs demonstrate clustering of cell lines and PDX models versus human tumors on PC2. (A & B) PCAs of RNA-Seq data of all samples listed in Table 3-7. PC2 appears to distinguish the human tumors from the other samples. From left to right (negative to positive direction) on PC2 the samples cluster from cell lines to PDX models to human tumors. PCAs are colored by dataset (A) and histology (B). The patterns of PC1 capturing UCLA tumors, as discussed extensively above, were also seen on these PCAs (data not shown).

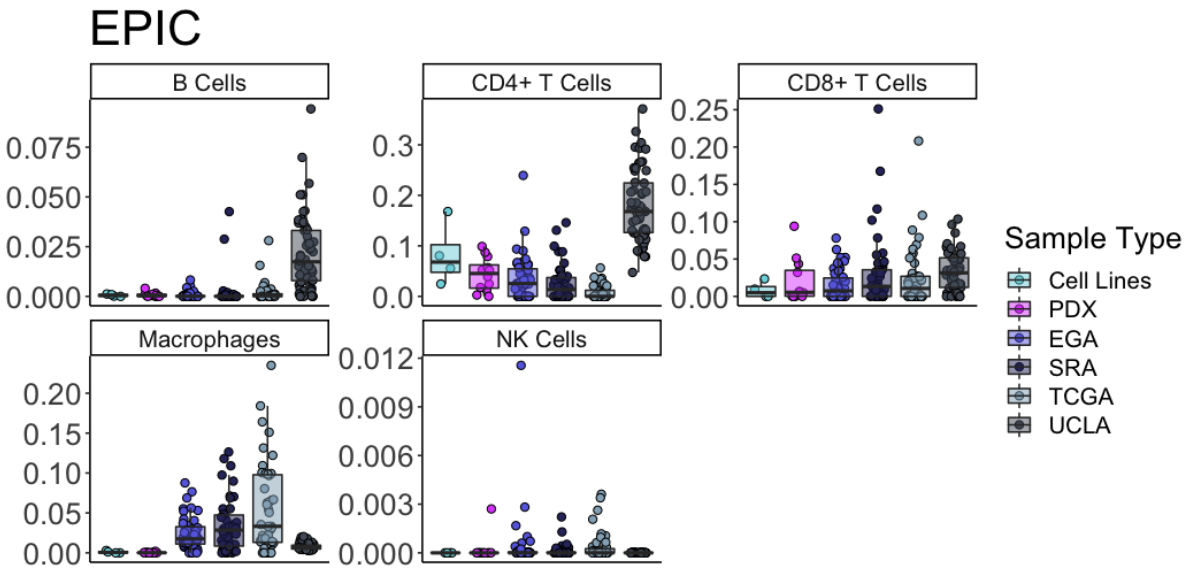
A


Total Immune Cells
CIBERSORTx
EPIC
MCP Counter
quanTIseq
TIMER
xCell
Sample Type
Cell Lines
PDX
EGA
SRA
TCGA
UCLA

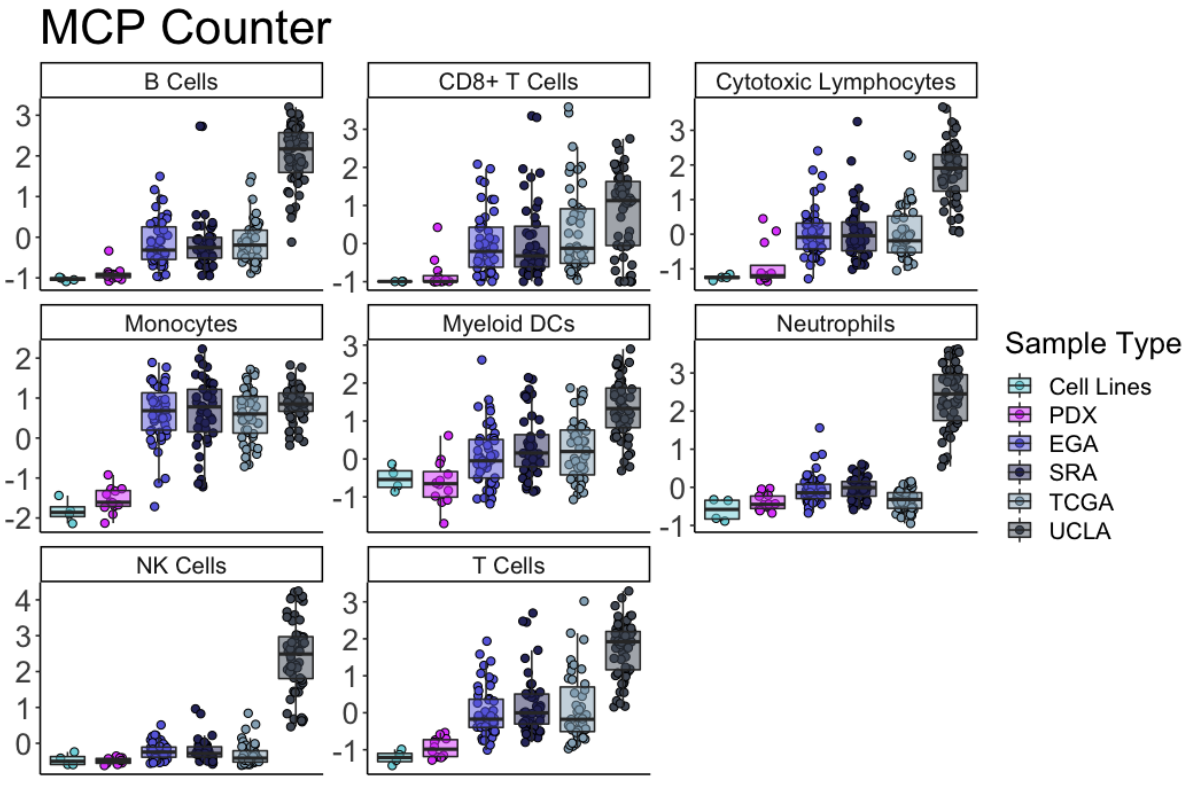
B


CIBERSORTx
Activated Dendritic Cells
Activated Mast Cells
Activated NK Cells
CD8+ T Cells
Eosinophils
Follicular Helper T Cells
Gamma Delta T Cells
M0 Macropahges
M1 Macrophages
M2 Macrophages
emory Activated CD4+ T Ce
Memory B Cells
emory Resting CD4+ T Cel
Monocytes
Naive B Cells
Naive CD4+ T Cells
Neutrophils
Plasma Cells
Resting Dendritic Cells
Resting Mast Cells
Resting NK Cells
Tregs
Sample Type
Cell Lines
PDX
EGA
SRA
TCGA
UCLA

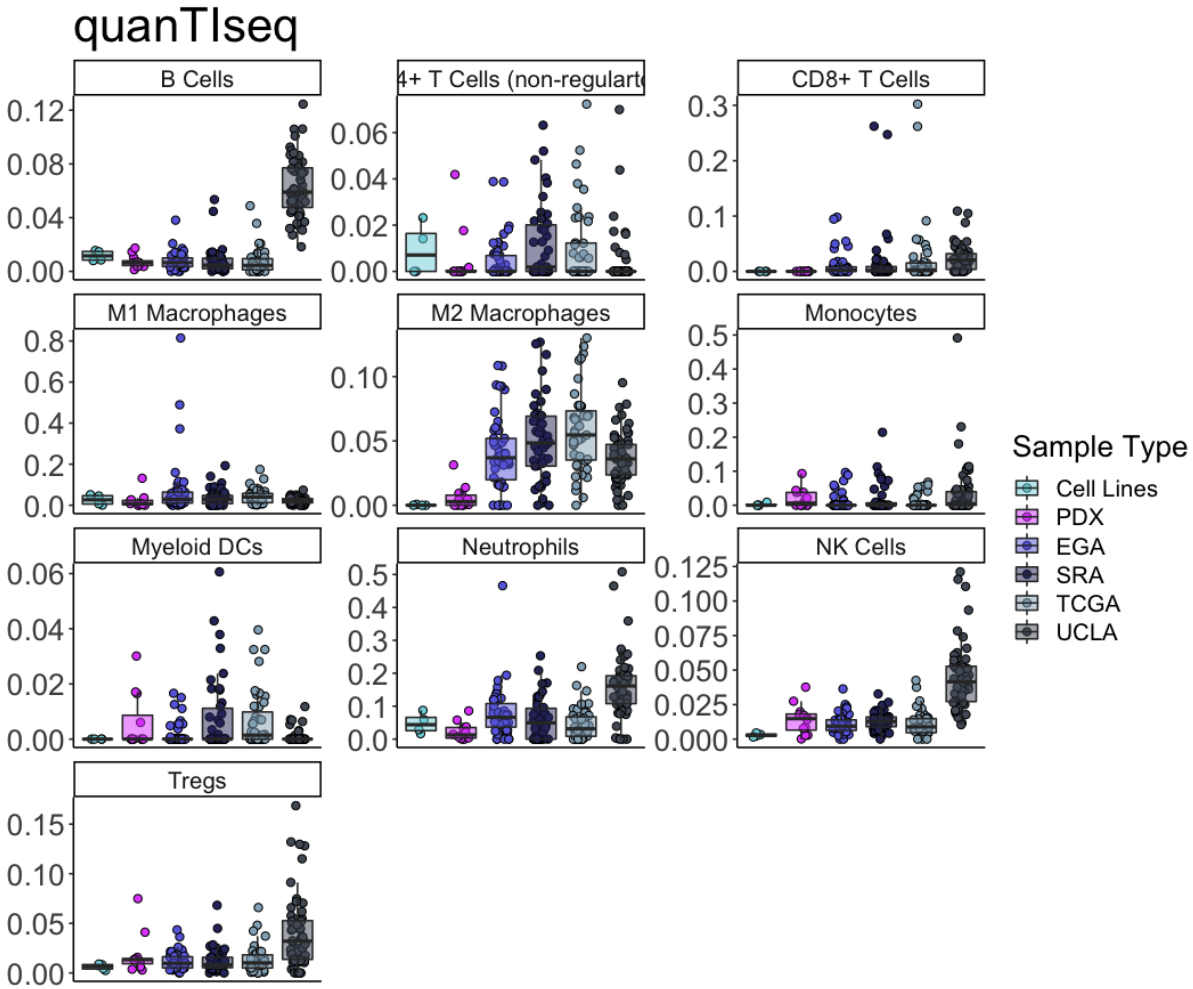
C


EPIC
B Cells
CD4+ T Cells
CD8+ T Cells
Macrophages
NK Cells
Sample Type
Cell Lines
PDX
EGA
SRA
TCGA
UCLA

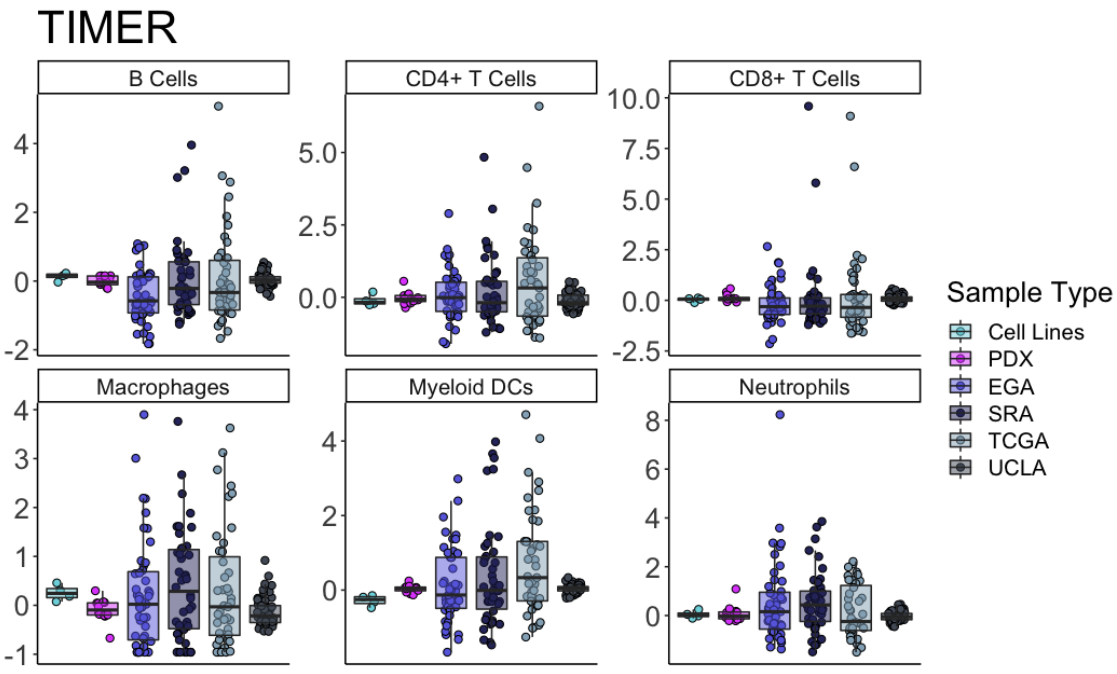
D

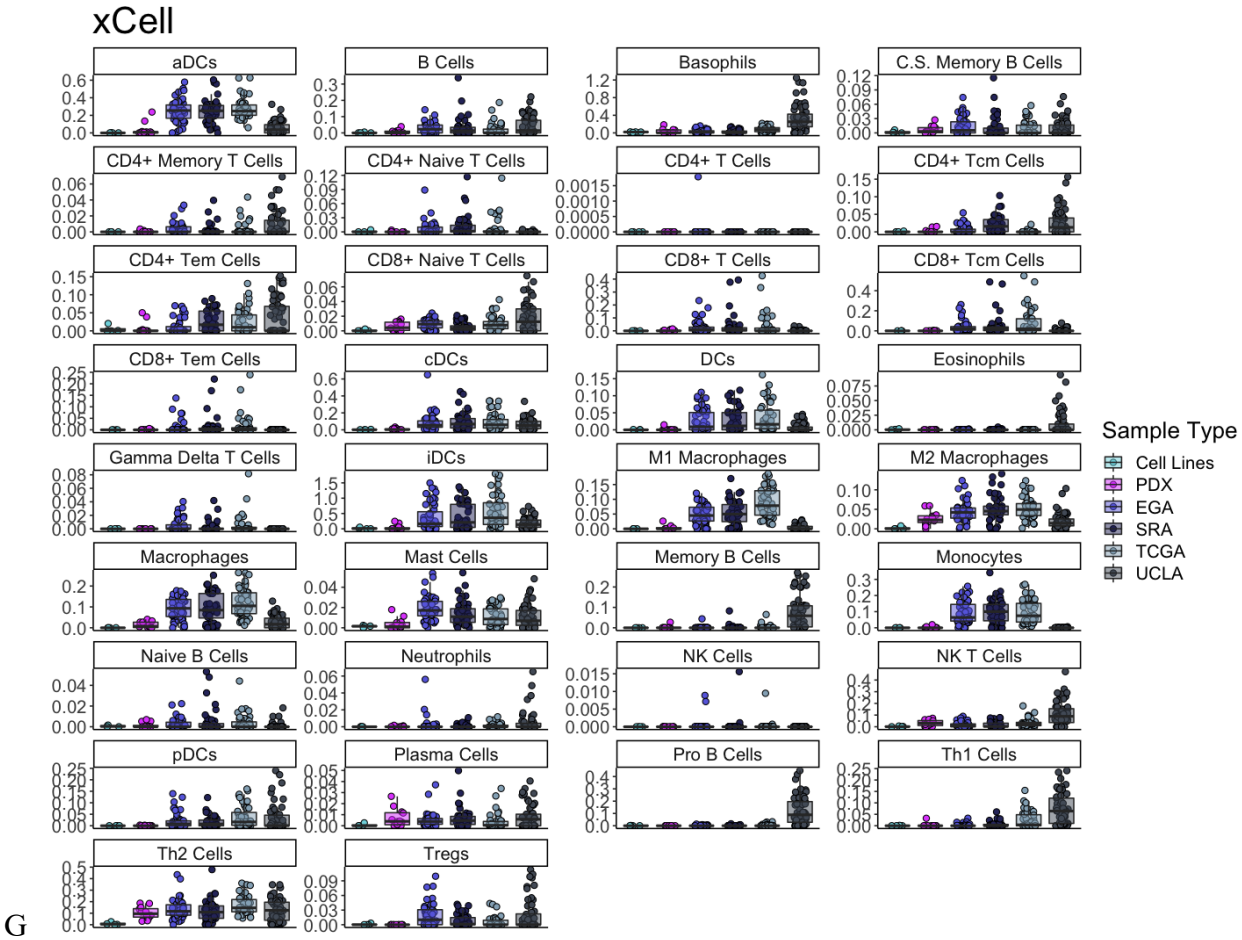

MCP Counter
B Cells
CD8+ T Cells
Cytotoxic Lymphocytes
Monocytes
Myeloid DCs
Neutrophils
NK Cells
T Cells
Sample Type
Cell Lines
PDX
EGA
SRA
TCGA
UCLA

E


quanTIseq
B Cells
4+ T Cells (non-regulart
CD8+ T Cells
M1 Macrophages
M2 Macrophages
Monocytes
Myeloid DCs
Neutrophils
NK Cells
Tregs
Sample Type
Cell Lines
PDX
EGA
SRA
TCGA
UCLA

F


TIMER
B Cells
CD4+ T Cells
CD8+ T Cells
Macrophages
Myeloid DCs
Neutrophils
Sample Type
Cell Lines
PDX
EGA
SRA
TCGA
UCLA



G

Figure 3-8: Bar graphs summarizing the immune cell landscape defined by the various in-silico immune deconvolution tools across all USARC samples included in this study – cell lines, PDX, EGA, SRA, TCGA, and UCLA. (A) Bar graph depicting the total immune cell infiltration defined by each tool. (B-G) Bar graphs depicting the immune cell populations predicted by the CIBERSORTx, EPIC, MCP Counter, quanTIseq, TIMER, and xCell tools, respectively.

Immune deconvolution tool	**Median**	**Min**	**Max**	**SD**	**p value**
CIBERSORTx					
UCLA	22.04	13.65	35.02	5.14	Ref
Cell Lines	4.28	3.85	6.07	0.99	0.03
EGA	5.83	4.49	7.55	1.03	<0.001
PDX	10.47	2.87	38.57	7.67	<0.001
SRA	10.97	4.23	70.92	13.53	<0.001

TCGA	11.45	5.65	85.92	16.23	<0.001
EPIC					
UCLA	0.23	0.09	0.45	0.09	Ref
Cell Lines	0.08	0.04	0.17	0.06	0.18
EGA	0.05	0.00	0.20	0.06	<0.001
PDX	0.06	0.00	0.28	0.07	<0.001
SRA	0.07	0.00	0.36	0.09	<0.001
TCGA	0.06	0.00	0.36	0.09	<0.001
MCP Counter					
UCLA	13.90	3.55	23.05	5.16	Ref
Cell Lines	-7.90	-8.48	-7.24	0.52	0.03
EGA	-6.56	-8.52	-4.57	1.06	<0.001
PDX	-0.07	-8.13	9.42	3.67	<0.001
SRA	-0.04	-5.14	13.75	4.53	<0.001
TCGA	-0.77	-4.89	13.53	4.51	<0.001
quanTIseq					
UCLA	0.42	0.27	0.70	0.09	Ref
Cell Lines	0.09	0.07	0.19	0.05	0.03
EGA	0.12	0.07	0.17	0.04	<0.001
PDX	0.20	0.10	0.90	0.14	<0.001
SRA	0.23	0.11	0.50	0.09	<0.001
TCGA	0.21	0.09	0.48	0.09	<0.001
TIMER					
UCLA	-0.25	-1.26	1.16	0.53	Ref
Cell Lines	0.01	-0.13	0.41	0.25	1.00
EGA	-0.04	-0.35	0.79	0.40	1.00
PDX	-0.48	-6.64	9.47	3.76	1.00
SRA	0.47	-6.12	20.94	6.14	1.00
TCGA	0.41	-5.19	21.34	5.84	1.00
xCell					
UCLA	0.06	0.00	0.19	0.05	Ref
Cell Lines	0.00	0.00	0.00	0.00	0.06
EGA	0.02	0.00	0.05	0.02	<0.001
PDX	0.17	0.01	0.50	0.13	0.01
SRA	0.18	0.00	0.78	0.18	<0.001
TCGA	0.16	0.04	0.63	0.17	<0.001

Table 3-15: Summary of total immune cell infiltration defined by each in-silico immune deconvolution tool (as listed above) across all USARC samples included in this study. This table corresponds to the graph in Figure 3-8A. Comparisons between groups were performed using the

Wilcoxon test. Bonferroni correction was used for multiple hypothesis testing correction.

Adjusted p values are shown.

Cell Type	Median	Min	Max	SD	p value
Activated Dendritic Cells					
Cell Lines	0.06	0.00	0.36	0.17	1.00E+00
PDX	0.02	0.00	2.05	0.61	1.00E+00
EGA	0.00	0.00	1.60	0.29	1.00E+00
SRA	0.00	0.00	0.43	0.08	1.00E+00
TCGA	0.00	0.00	1.02	0.15	1.00E+00
UCLA	0.00	0.00	2.81	0.58	Ref
Activated Mast Cells					
Cell Lines	0.47	0.06	0.74	0.35	1.00E+00
PDX	0.00	0.00	1.52	0.51	1.19E-02
EGA	0.00	0.00	1.38	0.29	1.00E+00
SRA	0.00	0.00	1.90	0.40	1.50E-02
TCGA	0.00	0.00	1.61	0.33	4.91E-04
UCLA	0.70	0.00	5.14	1.39	Ref
Activated NK Cells					
Cell Lines	0.04	0.00	0.08	0.04	1.00E+00
PDX	0.17	0.00	0.68	0.21	1.00E+00
EGA	0.01	0.00	0.57	0.14	1.00E+00
SRA	0.24	0.00	2.48	0.61	1.00E+00
TCGA	0.33	0.00	6.94	1.20	1.00E+00
UCLA	0.22	0.00	4.82	1.08	Ref
CD8+ T Cells					
Cell Lines	0.00	0.00	0.04	0.02	1.00E+00
PDX	0.10	0.00	0.87	0.27	1.00E+00
EGA	0.59	0.00	6.00	1.54	1.00E+00
SRA	0.51	0.00	14.33	3.18	1.00E+00
TCGA	0.73	0.00	30.81	5.05	1.00E+00
UCLA	0.88	0.00	6.39	1.40	Ref
Eosinophils					
Cell Lines	0.02	0.00	0.02	0.01	1.00E+00
PDX	0.00	0.00	0.26	0.07	2.20E-05
EGA	0.00	0.00	0.15	0.03	1.00E+00
SRA	0.00	0.00	0.03	0.00	2.04E-06
TCGA	0.00	0.00	0.00	0.00	5.67E-07

UCLA	0.26	0.00	4.74	1.20	Ref
Follicular Helper T Cells					
Cell Lines	0.00	0.00	0.00	0.00	1.00E+00
PDX	0.15	0.00	1.95	0.64	5.91E-03
EGA	0.22	0.00	1.67	0.42	1.00E+00
SRA	0.23	0.00	5.30	1.08	1.90E-02
TCGA	0.41	0.00	13.14	2.01	4.95E-04
UCLA	0.00	0.00	3.84	0.95	Ref
Gamma Delta T Cells					
Cell Lines	0.00	0.00	0.00	0.00	1.00E+00
PDX	0.00	0.00	0.00	0.00	1.00E+00
EGA	0.00	0.00	0.56	0.10	1.00E+00
SRA	0.00	0.00	0.51	0.08	1.00E+00
TCGA	0.00	0.00	1.78	0.30	1.00E+00
UCLA	0.00	0.00	1.50	0.30	Ref
M0 Macrophages					
Cell Lines	1.01	0.23	1.70	0.79	1.00E+00
PDX	0.49	0.00	1.00	0.31	1.00E+00
EGA	0.68	0.00	8.89	2.20	1.00E+00
SRA	0.06	0.00	11.34	2.32	1.00E+00
TCGA	0.12	0.00	34.24	5.52	1.00E+00
UCLA	0.18	0.00	14.07	2.54	Ref
M1 Macrophages					
Cell Lines	0.00	0.00	0.00	0.00	1.00E+00
PDX	0.00	0.00	0.81	0.33	8.48E-04
EGA	0.24	0.00	3.61	0.93	1.00E+00
SRA	0.23	0.00	7.28	1.42	1.69E-04
TCGA	0.37	0.00	3.77	0.84	2.50E-05
UCLA	0.00	0.00	1.87	0.39	Ref
M2 Macrophages					
Cell Lines	0.00	0.00	0.13	0.07	1.00E+00
PDX	0.00	0.00	1.74	0.61	8.47E-07
EGA	5.05	0.80	24.03	4.37	2.63E-02
SRA	5.61	0.13	26.22	4.98	1.75E-06
TCGA	6.26	1.63	37.16	6.27	5.42E-08
UCLA	2.04	0.00	7.70	2.27	Ref
Memory Activated CD4+ T Cells					
Cell Lines	0.00	0.00	0.00	0.00	1.00E+00
PDX	0.00	0.00	0.67	0.22	7.78E-02
EGA	0.00	0.00	1.87	0.40	5.50E-01

SRA	0.00	0.00	0.94	0.15	3.91E-05
TCGA	0.00	0.00	2.01	0.33	1.75E-02
UCLA	0.45	0.00	6.33	1.70	Ref
Memory B Cells					
Cell Lines	0.03	0.00	0.21	0.10	1.00E+00
PDX	0.08	0.00	0.73	0.27	1.00E+00
EGA	0.13	0.00	1.66	0.29	1.00E+00
SRA	0.07	0.00	4.81	1.01	8.80E-01
TCGA	0.01	0.00	1.21	0.24	3.71E-02
UCLA	1.28	0.00	8.84	2.00	Ref
Memory Resting CD4+ T Cells					
Cell Lines	0.92	0.82	1.43	0.29	1.00E+00
PDX	1.40	0.00	3.03	0.92	1.00E+00
EGA	0.63	0.00	3.82	0.79	1.00E+00
SRA	0.79	0.00	6.27	1.16	1.00E+00
TCGA	0.57	0.00	4.91	0.88	4.40E-01
UCLA	2.17	0.00	15.63	2.76	Ref
Monocytes					
Cell Lines	0.06	0.03	0.16	0.05	1.00E+00
PDX	0.15	0.06	0.43	0.13	1.00E+00
EGA	0.20	0.00	2.16	0.47	1.00E+00
SRA	0.46	0.00	7.20	1.15	1.00E+00
TCGA	0.29	0.00	2.50	0.60	1.00E+00
UCLA	0.29	0.00	3.86	0.95	Ref
Naive B Cells					
Cell Lines	0.00	0.00	0.00	0.00	1.00E+00
PDX	0.00	0.00	0.39	0.11	6.56E-05
EGA	0.00	0.00	0.60	0.10	2.20E-01
SRA	0.00	0.00	2.85	0.46	2.02E-02
TCGA	0.02	0.00	1.59	0.37	4.40E-01
UCLA	0.54	0.00	8.67	2.24	Ref
Naive CD4+ T Cells					
Cell Lines	1.42	0.37	2.05	0.81	1.00E+00
PDX	0.01	0.00	1.97	0.73	2.57E-11
EGA	0.00	0.00	0.47	0.08	7.62E-02
SRA	0.00	0.00	0.47	0.08	6.89E-11
TCGA	0.00	0.00	0.26	0.04	7.01E-12
UCLA	2.66	0.00	13.76	3.42	Ref
Neutrophils					
Cell Lines	0.00	0.00	0.00	0.00	1.00E+00

PDX	0.00	0.00	0.03	0.01	1.00E+00
EGA	0.00	0.00	0.90	0.14	1.00E+00
SRA	0.00	0.00	0.31	0.09	1.00E+00
TCGA	0.00	0.00	0.47	0.10	1.00E+00
UCLA	0.00	0.00	2.65	0.70	Ref
Plasma Cells					
Cell Lines	0.06	0.02	0.40	0.18	1.00E+00
PDX	0.00	0.00	0.32	0.09	1.00E+00
EGA	0.05	0.00	1.19	0.23	1.00E+00
SRA	0.04	0.00	1.68	0.28	1.00E+00
TCGA	0.00	0.00	1.70	0.36	1.00E+00
UCLA	0.00	0.00	3.28	0.82	Ref
Resting Dendritic Cells					
Cell Lines	0.00	0.00	0.00	0.00	1.00E+00
PDX	0.00	0.00	0.00	0.00	1.00E+00
EGA	0.00	0.00	0.42	0.08	1.00E+00
SRA	0.00	0.00	0.87	0.18	1.00E+00
TCGA	0.00	0.00	0.31	0.09	1.00E+00
UCLA	0.00	0.00	2.38	0.54	Ref
Resting Mast Cells					
Cell Lines	0.00	0.00	0.00	0.00	1.00E+00
PDX	0.18	0.00	1.52	0.56	1.10E-01
EGA	0.26	0.00	2.59	0.51	1.00E+00
SRA	0.29	0.00	4.81	1.01	1.10E-01
TCGA	0.42	0.00	3.30	0.78	1.01E-02
UCLA	0.00	0.00	5.81	0.96	Ref
Resting NK Cells					
Cell Lines	0.43	0.00	0.62	0.28	1.00E+00
PDX	0.44	0.00	1.01	0.36	6.60E-01
EGA	0.05	0.00	1.77	0.36	1.00E+00
SRA	0.00	0.00	0.91	0.20	2.57E-02
TCGA	0.00	0.00	0.31	0.08	3.87E-03
UCLA	0.49	0.00	8.17	1.39	Ref
Tregs					
Cell Lines	0.00	0.00	0.00	0.00	1.00E+00
PDX	0.00	0.00	0.78	0.27	1.00E+00
EGA	0.03	0.00	0.64	0.18	1.00E+00
SRA	0.08	0.00	2.77	0.55	1.00E+00
TCGA	0.16	0.00	4.33	0.78	2.74E-03
UCLA	0.00	0.00	5.00	1.23	Ref

Table 3-16: Summary of immune cell populations across all USARC samples included in this study predicted by CIBERSORTx in-silico immune deconvolution tools. This table corresponds to the graph in Figure 3-8B. Comparisons between groups were performed using the Wilcoxon test. Bonferroni correction was used for multiple hypothesis testing correction. Adjusted p values are shown.

Cell Type	Median	Min	Max	SD	p value
B Cells					
Cell Lines	5.05E-04	8.44E-07	1.21E-03	5.86E-04	1.50E-01
PDX	3.71E-04	4.01E-07	4.06E-03	1.16E-03	2.70E-04
EGA	3.53E-05	8.37E-09	8.25E-03	1.56E-03	2.08E-12
SRA	1.69E-05	3.26E-10	4.26E-02	7.81E-03	1.04E-10
TCGA	3.83E-04	2.01E-08	2.81E-02	4.88E-03	2.13E-09
UCLA	1.74E-02	3.63E-09	9.43E-02	1.91E-02	Ref
CD4+ T Cells					
Cell Lines	6.79E-02	2.46E-02	1.68E-01	6.17E-02	4.50E-01
PDX	4.53E-02	9.75E-05	9.88E-02	3.32E-02	7.48E-06
EGA	2.54E-02	2.86E-09	2.39E-01	4.35E-02	7.08E-15
SRA	1.36E-02	1.62E-07	1.46E-01	3.66E-02	2.31E-14
TCGA	8.66E-06	2.40E-09	5.64E-02	1.35E-02	5.23E-16
UCLA	1.68E-01	4.74E-02	3.71E-01	7.16E-02	Ref
CD8+ T Cells					
Cell Lines	4.89E-03	1.02E-07	2.35E-02	1.11E-02	9.75E-01
PDX	5.41E-03	3.64E-08	9.40E-02	2.97E-02	1.00E+00
EGA	7.37E-03	1.60E-08	7.81E-02	1.99E-02	8.63E-03
SRA	1.33E-02	1.01E-09	2.51E-01	4.98E-02	1.00E+00
TCGA	1.07E-02	1.35E-09	2.08E-01	3.97E-02	1.25E-01
UCLA	3.15E-02	2.54E-07	1.04E-01	2.63E-02	Ref
Macrophages					
Cell Lines	5.52E-04	9.89E-09	2.96E-03	1.39E-03	2.50E-02
PDX	1.58E-04	2.07E-07	1.92E-03	5.28E-04	1.69E-06
EGA	1.75E-02	3.54E-08	8.76E-02	2.07E-02	1.08E-05
SRA	2.85E-02	9.88E-08	1.26E-01	3.42E-02	1.83E-03
TCGA	3.33E-02	9.10E-09	2.35E-01	5.66E-02	1.29E-06
UCLA	7.20E-03	2.80E-03	2.04E-02	4.08E-03	Ref
NK Cells					

Cell Lines	3.58E-09	2.14E-10	8.35E-08	4.05E-08	1.00E+00
PDX	6.06E-08	4.66E-10	2.71E-03	7.81E-04	2.75E-01
EGA	1.34E-07	1.18E-09	1.16E-02	1.73E-03	1.94E-06
SRA	1.43E-07	3.33E-09	2.22E-03	4.00E-04	1.02E-05
TCGA	2.26E-06	1.96E-09	3.61E-03	8.88E-04	6.18E-09
UCLA	4.72E-09	8.99E-11	6.48E-06	1.07E-06	Ref

Table 3-17: Summary of immune cell populations across all USARC samples included in this study predicted by EPIC in-silico immune deconvolution tools. This table corresponds to the graph in Figure 3-8C. Comparisons between groups were performed using the Wilcoxon test. Bonferroni correction was used for multiple hypothesis testing correction. Adjusted p values are shown.

Cell Type	**Median**	**Min**	**Max**	**SD**	**p value**
B Cells					
Cell Lines	-1.03	-1.09	-0.98	0.05	3.81E-02
PDX	-0.92	-1.08	-0.34	0.19	2.71E-06
EGA	-0.32	-0.97	1.50	0.58	2.67E-15
SRA	-0.25	-0.95	2.73	0.75	6.48E-13
TCGA	-0.19	-0.89	1.49	0.54	1.15E-14
UCLA	2.17	-0.12	3.20	0.74	Ref
CD8+ T Cells					
Cell Lines	-0.99	-0.99	-0.99	0.00	1.20E-01
PDX	-0.99	-0.99	0.43	0.42	8.40E-04
EGA	-0.20	-0.99	2.08	0.81	7.92E-03
SRA	-0.32	-0.99	3.35	1.08	4.00E-02
TCGA	-0.12	-0.96	3.58	1.17	9.60E-01
UCLA	1.13	-0.99	2.75	1.09	Ref
Cytotoxic Lymphocytes					
Cell Lines	-1.25	-1.34	-1.16	0.07	3.81E-02
PDX	-1.20	-1.36	0.45	0.62	4.60E-06
EGA	-0.09	-1.29	2.41	0.73	1.81E-12
SRA	-0.04	-1.01	3.25	0.84	2.60E-11
TCGA	-0.19	-1.05	2.28	0.79	1.80E-11
UCLA	1.91	0.06	3.67	0.87	Ref

Monocytes					
Cell Lines	-1.86	-2.14	-1.44	0.29	3.81E-02
PDX	-1.61	-2.13	-0.92	0.33	2.71E-06
EGA	0.68	-1.71	1.89	0.74	1.00E+00
SRA	0.78	-1.22	2.23	0.88	1.00E+00
TCGA	0.60	-0.71	1.72	0.64	1.00E+00
UCLA	0.85	-0.19	1.82	0.43	Ref
Myeloid DCs					
Cell Lines	-0.54	-0.86	-0.13	0.33	4.00E-02
PDX	-0.65	-1.70	0.62	0.61	1.20E-05
EGA	-0.05	-1.19	2.61	0.78	1.82E-09
SRA	0.16	-0.86	2.14	0.82	5.80E-06
TCGA	0.20	-1.07	1.87	0.80	6.92E-07
UCLA	1.33	-0.31	2.90	0.77	Ref
Neutrophils					
Cell Lines	-0.58	-0.89	-0.32	0.30	3.81E-02
PDX	-0.45	-0.68	-0.03	0.22	2.71E-06
EGA	-0.14	-0.68	1.56	0.40	4.12E-16
SRA	-0.01	-0.59	0.61	0.29	1.45E-15
TCGA	-0.32	-0.95	0.16	0.28	7.84E-16
UCLA	2.45	0.54	3.63	0.83	Ref
NK Cells					
Cell Lines	-0.50	-0.59	-0.24	0.17	3.81E-02
PDX	-0.50	-0.62	-0.36	0.08	2.71E-06
EGA	-0.25	-0.56	0.51	0.23	1.29E-16
SRA	-0.28	-0.58	0.96	0.31	2.84E-15
TCGA	-0.41	-0.61	0.84	0.32	1.28E-15
UCLA	2.49	0.46	4.24	1.03	Ref
T Cells					
Cell Lines	-1.21	-1.43	-0.99	0.19	3.81E-02
PDX	-0.99	-1.28	-0.54	0.27	2.71E-06
EGA	-0.17	-1.01	1.94	0.68	4.76E-13
SRA	0.00	-0.80	2.69	0.85	1.16E-09
TCGA	-0.18	-0.97	3.02	0.93	6.72E-10
UCLA	1.92	0.16	3.29	0.75	Ref

Table 3-18: Summary of immune cell populations across all USARC samples included in this study predicted by MCP Counter in-silico immune deconvolution tools. This table corresponds

to the graph in Figure 3-8D. Comparisons between groups were performed using the Wilcoxon test. Bonferroni correction was used for multiple hypothesis testing correction. Adjusted p values are shown.

Cell Type	Median	Min	Max	SD	p value
B Cells					
Cell Lines	0.01	0.01	0.02	0.00	4.76E-02
PDX	0.01	0.00	0.02	0.00	3.39E-06
EGA	0.01	0.00	0.04	0.01	2.42E-16
SRA	0.00	0.00	0.05	0.01	1.19E-14
TCGA	0.00	0.00	0.05	0.01	3.67E-15
UCLA	0.06	0.02	0.12	0.02	Ref
CD4+ T Cells (non-regulatory)					
Cell Lines	0.01	0.00	0.02	0.01	1.00E+00
PDX	0.00	0.00	0.04	0.01	1.00E+00
EGA	0.00	0.00	0.04	0.01	1.00E+00
SRA	0.00	0.00	0.06	0.02	5.15E-03
TCGA	0.00	0.00	0.07	0.02	4.00E-01
UCLA	0.00	0.00	0.07	0.01	Ref
CD8+ T Cells					
Cell Lines	0.00	0.00	0.00	0.00	3.00E-01
PDX	0.00	0.00	0.00	0.00	3.26E-04
EGA	0.00	0.00	0.10	0.02	8.80E-03
SRA	0.00	0.00	0.26	0.06	2.74E-02
TCGA	0.00	0.00	0.30	0.06	4.50E-01
UCLA	0.02	0.00	0.11	0.02	Ref
M1 Macrophages					
Cell Lines	0.03	0.00	0.05	0.03	1.00E+00
PDX	0.01	0.00	0.13	0.04	1.00E+00
EGA	0.03	0.00	0.81	0.14	1.00E+00
SRA	0.03	0.00	0.19	0.04	1.00E+00
TCGA	0.04	0.00	0.17	0.04	2.00E-01
UCLA	0.02	0.00	0.07	0.02	Ref
M2 Macrophages					
Cell Lines	0.00	0.00	0.00	0.00	1.00E-01
PDX	0.00	0.00	0.03	0.01	2.17E-04
EGA	0.04	0.00	0.11	0.03	1.00E+00

SRA	0.05	0.00	0.13	0.03	1.00E+00
TCGA	0.05	0.00	0.13	0.03	1.70E-02
UCLA	0.04	0.00	0.10	0.02	Ref
Monocytes					
Cell Lines	0.00	0.00	0.01	0.00	1.00E+00
PDX	0.01	0.00	0.09	0.03	1.00E+00
EGA	0.00	0.00	0.10	0.02	2.00E-01
SRA	0.00	0.00	0.21	0.04	1.00E+00
TCGA	0.00	0.00	0.07	0.02	4.50E-01
UCLA	0.00	0.00	0.49	0.08	Ref
Myeloid DCs					
Cell Lines	0.00	0.00	0.00	0.00	1.00E+00
PDX	0.00	0.00	0.03	0.01	1.00E+00
EGA	0.00	0.00	0.02	0.00	1.00E+00
SRA	0.00	0.00	0.06	0.01	2.27E-02
TCGA	0.00	0.00	0.04	0.01	7.40E-04
UCLA	0.00	0.00	0.01	0.00	Ref
Neutrophils					
Cell Lines	0.04	0.02	0.09	0.03	6.50E-01
PDX	0.01	0.00	0.09	0.03	3.71E-04
EGA	0.07	0.00	0.47	0.08	1.04E-05
SRA	0.05	0.00	0.25	0.06	2.88E-06
TCGA	0.03	0.00	0.22	0.05	2.21E-08
UCLA	0.16	0.00	0.51	0.10	Ref
NK Cells					
Cell Lines	0.00	0.00	0.00	0.00	4.76E-02
PDX	0.01	0.00	0.04	0.01	3.21E-04
EGA	0.01	0.00	0.04	0.01	3.99E-14
SRA	0.01	0.00	0.03	0.01	1.87E-12
TCGA	0.01	0.00	0.04	0.01	9.10E-13
UCLA	0.04	0.01	0.12	0.02	Ref
Tregs					
Cell Lines	0.01	0.00	0.01	0.00	3.50E-01
PDX	0.01	0.00	0.07	0.02	7.50E-01
EGA	0.01	0.00	0.04	0.01	5.70E-06
SRA	0.01	0.00	0.07	0.01	5.75E-05
TCGA	0.01	0.00	0.07	0.01	2.05E-04
UCLA	0.03	0.00	0.17	0.04	Ref

Table 3-19: Summary of immune cell populations across all USARC samples included in this study predicted by quanTIseq in-silico immune deconvolution tools. This table corresponds to the graph in Figure 3-8E. Comparisons between groups were performed using the Wilcoxon test. Bonferroni correction was used for multiple hypothesis testing correction. Adjusted p values are shown.

Cell Type	Median	Min	Max	SD	p value
B Cells					
Cell Lines	0.17	-0.03	0.23	0.11	1.00
PDX	-0.03	-0.21	0.15	0.14	1.00
EGA	-0.58	-1.83	1.08	0.76	0.01
SRA	-0.21	-1.25	3.96	1.12	1.00
TCGA	-0.33	-1.67	5.09	1.38	0.60
UCLA	0.03	-0.44	0.55	0.19	Ref
CD4+ T Cells					
Cell Lines	-0.17	-0.26	0.19	0.20	1.00
PDX	-0.07	-0.36	0.56	0.23	1.00
EGA	-0.01	-1.61	2.89	0.84	1.00
SRA	-0.19	-1.21	4.84	1.17	1.00
TCGA	0.32	-1.41	6.60	1.62	1.00
UCLA	-0.19	-0.57	0.53	0.25	Ref
CD8+ T Cells					
Cell Lines	0.06	-0.11	0.12	0.10	1.00
PDX	0.07	-0.09	0.58	0.20	1.00
EGA	-0.31	-2.15	2.66	0.90	0.01
SRA	-0.29	-1.20	9.58	1.87	0.00
TCGA	-0.35	-1.63	9.10	2.00	1.00
UCLA	0.08	-0.23	0.57	0.17	Ref
Macrophages					
Cell Lines	0.24	0.07	0.46	0.16	0.33
PDX	-0.10	-0.67	0.30	0.24	1.00
EGA	0.02	-0.96	3.90	1.12	1.00
SRA	0.29	-0.96	3.76	1.12	1.00
TCGA	-0.03	-0.96	3.63	1.22	1.00
UCLA	-0.22	-0.53	0.92	0.30	Ref
Myeloid DCs					

Cell Lines	-0.26	-0.47	-0.15	0.15	0.03
PDX	0.04	-0.13	0.24	0.10	1.00
EGA	-0.13	-1.67	2.98	1.01	1.00
SRA	-0.01	-1.47	3.98	1.40	1.00
TCGA	0.33	-1.26	4.71	1.41	1.00
UCLA	0.05	-0.21	0.34	0.11	Ref
Neutrophils					
Cell Lines	0.02	-0.11	0.25	0.15	1.00
PDX	-0.02	-0.22	1.09	0.35	1.00
EGA	0.16	-1.37	8.23	1.67	1.00
SRA	0.43	-1.49	3.85	1.25	0.03
TCGA	-0.24	-1.50	2.21	1.06	1.00
UCLA	-0.05	-0.45	0.46	0.20	Ref

Table 3-20: Summary of immune cell populations across all USARC samples included in this study predicted by TIMER in-silico immune deconvolution tools. This table corresponds to the graph in Figure 3-8F. Comparisons between groups were performed using the Wilcoxon test. Bonferroni correction was used for multiple hypothesis testing correction. Adjusted p values are shown.

Cell Type	**Median**	**Min**	**Max**	**SD**	**p value**
aDCs					
Cell Lines	0.00	0.00	0.00	0.00	1.00
PDX	0.01	0.00	0.24	0.07	1.00
EGA	0.25	0.00	0.58	0.13	<0.001
SRA	0.25	0.00	0.60	0.14	<0.001
TCGA	0.25	0.06	0.63	0.12	<0.001
UCLA	0.04	0.00	0.33	0.07	Ref
B Cells					
Cell Lines	0.00	0.00	0.01	0.00	1.00
PDX	0.00	0.00	0.04	0.01	1.00
EGA	0.02	0.00	0.14	0.03	1.00
SRA	0.01	0.00	0.34	0.06	1.00
TCGA	0.00	0.00	0.19	0.04	1.00
UCLA	0.01	0.00	0.22	0.05	Ref

Basophils					
Cell Lines	0.02	0.01	0.02	0.01	0.51
PDX	0.00	0.00	0.18	0.06	<0.001
EGA	0.02	0.00	0.16	0.04	<0.001
SRA	0.01	0.00	0.14	0.04	<0.001
TCGA	0.07	0.00	0.21	0.06	<0.001
UCLA	0.25	0.00	1.25	0.30	Ref
C.S. Memory B Cells					
Cell Lines	0.00	0.00	0.01	0.00	1.00
PDX	0.00	0.00	0.03	0.01	1.00
EGA	0.01	0.00	0.07	0.02	1.00
SRA	0.00	0.00	0.12	0.02	1.00
TCGA	0.00	0.00	0.06	0.02	1.00
UCLA	0.00	0.00	0.08	0.02	Ref
CD4+ Memory T Cells					
Cell Lines	0.00	0.00	0.00	0.00	1.00
PDX	0.00	0.00	0.00	0.00	1.00
EGA	0.00	0.00	0.03	0.01	0.17
SRA	0.00	0.00	0.04	0.01	0.04
TCGA	0.00	0.00	0.04	0.01	1.00
UCLA	0.00	0.00	0.07	0.02	Ref
CD4+ Naive T Cells					
Cell Lines	0.00	0.00	0.00	0.00	1.00
PDX	0.00	0.00	0.00	0.00	1.00
EGA	0.00	0.00	0.09	0.01	1.00
SRA	0.00	0.00	0.12	0.02	1.00
TCGA	0.00	0.00	0.11	0.02	1.00
UCLA	0.00	0.00	0.01	0.00	Ref
CD4+ T Cells					
Cell Lines	0.00	0.00	0.00	0.00	1.00
PDX	0.00	0.00	0.00	0.00	0.85
EGA	0.00	0.00	0.00	0.00	1.00
SRA	0.00	0.00	0.00	0.00	1.00
TCGA	0.00	0.00	0.00	0.00	<0.001
UCLA	0.00	0.00	0.00	0.00	Ref
CD4+ Tcm Cells					
Cell Lines	0.00	0.00	0.00	0.00	1.00
PDX	0.00	0.00	0.02	0.01	1.00
EGA	0.00	0.00	0.05	0.01	1.00
SRA	0.02	0.00	0.10	0.02	1.00

TCGA	0.00	0.00	0.02	0.00	1.00
UCLA	0.01	0.00	0.16	0.04	Ref
CD4+ Tem Cells					
Cell Lines	0.00	0.00	0.02	0.01	1.00
PDX	0.00	0.00	0.05	0.02	1.00
EGA	0.00	0.00	0.07	0.02	1.00
SRA	0.02	0.00	0.09	0.03	0.51
TCGA	0.01	0.00	0.13	0.03	1.00
UCLA	0.00	0.00	0.15	0.05	Ref
CD8+ Naive T Cells					
Cell Lines	0.00	0.00	0.00	0.00	1.00
PDX	0.00	0.00	0.02	0.01	1.00
EGA	0.01	0.00	0.02	0.01	<0.001
SRA	0.00	0.00	0.02	0.01	<0.001
TCGA	0.01	0.00	0.04	0.01	1.00
UCLA	0.01	0.00	0.08	0.02	Ref
CD8+ T Cells					
Cell Lines	0.00	0.00	0.00	0.00	1.00
PDX	0.00	0.00	0.02	0.01	1.00
EGA	0.01	0.00	0.23	0.05	<0.001
SRA	0.01	0.00	0.39	0.09	<0.001
TCGA	0.00	0.00	0.43	0.09	<0.001
UCLA	0.00	0.00	0.04	0.01	Ref
CD8+ Tcm Cells					
Cell Lines	0.00	0.00	0.00	0.00	1.00
PDX	0.00	0.00	0.00	0.00	1.00
EGA	0.02	0.00	0.26	0.06	1.00
SRA	0.02	0.00	0.49	0.11	0.05
TCGA	0.02	0.00	0.55	0.13	0.12
UCLA	0.00	0.00	0.08	0.01	Ref
CD8+ Tem Cells					
Cell Lines	0.00	0.00	0.00	0.00	1.00
PDX	0.00	0.00	0.01	0.00	0.03
EGA	0.00	0.00	0.14	0.03	1.00
SRA	0.00	0.00	0.22	0.04	1.00
TCGA	0.00	0.00	0.24	0.05	1.00
UCLA	0.00	0.00	0.00	0.00	Ref
cDCs					
Cell Lines	0.00	0.00	0.00	0.00	1.00
PDX	0.00	0.00	0.03	0.01	1.00

EGA	0.05	0.00	0.65	0.10	1.00
SRA	0.07	0.00	0.45	0.11	1.00
TCGA	0.06	0.00	0.34	0.10	1.00
UCLA	0.05	0.00	0.34	0.06	Ref
DCs					
Cell Lines	0.00	0.00	0.00	0.00	1.00
PDX	0.00	0.00	0.01	0.00	1.00
EGA	0.01	0.00	0.11	0.03	0.34
SRA	0.01	0.00	0.12	0.03	0.17
TCGA	0.02	0.00	0.16	0.04	<0.001
UCLA	0.00	0.00	0.05	0.01	Ref
Eosinophils					
Cell Lines	0.00	0.00	0.00	0.00	1.00
PDX	0.00	0.00	0.00	0.00	1.00
EGA	0.00	0.00	0.00	0.00	1.00
SRA	0.00	0.00	0.00	0.00	1.00
TCGA	0.00	0.00	0.00	0.00	1.00
UCLA	0.00	0.00	0.09	0.02	Ref
Gamma Delta T Cells					
Cell Lines	0.00	0.00	0.00	0.00	1.00
PDX	0.00	0.00	0.00	0.00	1.00
EGA	0.00	0.00	0.04	0.01	1.00
SRA	0.00	0.00	0.04	0.01	1.00
TCGA	0.00	0.00	0.08	0.01	1.00
UCLA	0.00	0.00	0.00	0.00	Ref
iDCs					
Cell Lines	0.01	0.00	0.05	0.02	1.00
PDX	0.00	0.00	0.24	0.08	0.15
EGA	0.16	0.00	1.50	0.43	1.00
SRA	0.21	0.00	1.37	0.43	1.00
TCGA	0.36	0.00	1.80	0.52	0.02
UCLA	0.14	0.00	0.73	0.17	Ref
M1 Macrophages					
Cell Lines	0.00	0.00	0.00	0.00	1.00
PDX	0.00	0.00	0.03	0.01	1.00
EGA	0.05	0.00	0.12	0.03	<0.001
SRA	0.05	0.00	0.17	0.04	<0.001
TCGA	0.08	0.02	0.19	0.05	<0.001
UCLA	0.00	0.00	0.03	0.01	Ref
M2 Macrophages					

Cell Lines	0.00	0.00	0.01	0.00	1.00
PDX	0.02	0.00	0.06	0.02	1.00
EGA	0.04	0.00	0.12	0.03	<0.001
SRA	0.05	0.00	0.14	0.03	<0.001
TCGA	0.05	0.00	0.12	0.03	<0.001
UCLA	0.01	0.00	0.10	0.02	Ref
Macrophages					
Cell Lines	0.00	0.00	0.00	0.00	1.00
PDX	0.01	0.00	0.04	0.01	1.00
EGA	0.09	0.00	0.18	0.05	<0.001
SRA	0.08	0.00	0.25	0.07	<0.001
TCGA	0.10	0.01	0.26	0.07	<0.001
UCLA	0.02	0.00	0.13	0.03	Ref
Mast Cells					
Cell Lines	0.00	0.00	0.00	0.00	1.00
PDX	0.00	0.00	0.02	0.01	1.00
EGA	0.02	0.00	0.05	0.01	0.01
SRA	0.01	0.00	0.05	0.01	1.00
TCGA	0.01	0.00	0.03	0.01	1.00
UCLA	0.01	0.00	0.05	0.01	Ref
Memory B Cells					
Cell Lines	0.00	0.00	0.00	0.00	1.00
PDX	0.00	0.00	0.03	0.01	0.02
EGA	0.00	0.00	0.04	0.01	<0.001
SRA	0.00	0.00	0.08	0.01	<0.001
TCGA	0.00	0.00	0.06	0.01	<0.001
UCLA	0.06	0.00	0.27	0.08	Ref
Monocytes					
Cell Lines	0.00	0.00	0.00	0.00	1.00
PDX	0.00	0.00	0.02	0.01	1.00
EGA	0.06	0.00	0.26	0.07	<0.001
SRA	0.10	0.00	0.34	0.08	<0.001
TCGA	0.07	0.01	0.27	0.07	<0.001
UCLA	0.00	0.00	0.01	0.00	Ref
Naive B Cells					
Cell Lines	0.00	0.00	0.00	0.00	1.00
PDX	0.00	0.00	0.01	0.00	1.00
EGA	0.00	0.00	0.02	0.01	1.00
SRA	0.00	0.00	0.05	0.01	1.00
TCGA	0.00	0.00	0.04	0.01	1.00

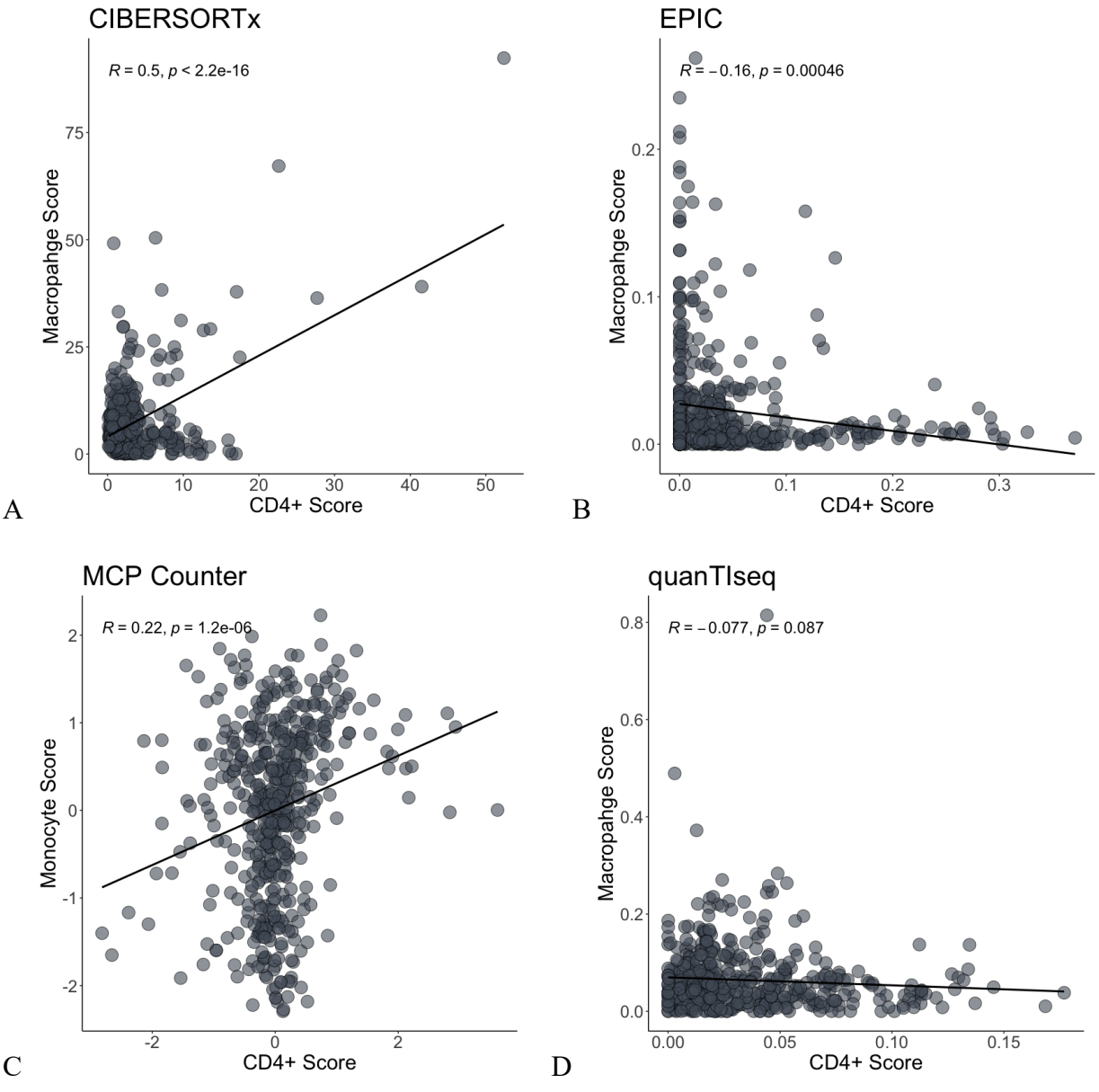
UCLA	0.00	0.00	0.02	0.00	Ref
Neutrophils					
Cell Lines	0.00	0.00	0.00	0.00	1.00
PDX	0.00	0.00	0.00	0.00	1.00
EGA	0.00	0.00	0.06	0.01	1.00
SRA	0.00	0.00	0.01	0.00	1.00
TCGA	0.00	0.00	0.01	0.00	1.00
UCLA	0.00	0.00	0.07	0.01	Ref
NK Cells					
Cell Lines	0.00	0.00	0.00	0.00	1.00
PDX	0.00	0.00	0.00	0.00	1.00
EGA	0.00	0.00	0.01	0.00	0.10
SRA	0.00	0.00	0.02	0.00	0.68
TCGA	0.00	0.00	0.01	0.00	1.00
UCLA	0.00	0.00	0.00	0.00	Ref
NK T Cells					
Cell Lines	0.00	0.00	0.01	0.00	1.00
PDX	0.03	0.00	0.07	0.03	1.00
EGA	0.01	0.00	0.09	0.02	<0.001
SRA	0.00	0.00	0.08	0.02	<0.001
TCGA	0.02	0.00	0.18	0.04	<0.001
UCLA	0.09	0.00	0.47	0.10	Ref
pDCs					
Cell Lines	0.00	0.00	0.00	0.00	1.00
PDX	0.00	0.00	0.00	0.00	1.00
EGA	0.01	0.00	0.14	0.03	1.00
SRA	0.01	0.00	0.12	0.02	1.00
TCGA	0.02	0.00	0.16	0.05	1.00
UCLA	0.00	0.00	0.24	0.06	Ref
Plasma Cells					
Cell Lines	0.00	0.00	0.00	0.00	1.00
PDX	0.00	0.00	0.03	0.01	1.00
EGA	0.00	0.00	0.04	0.01	1.00
SRA	0.00	0.00	0.05	0.01	1.00
TCGA	0.00	0.00	0.03	0.01	1.00
UCLA	0.01	0.00	0.04	0.01	Ref
Pro B Cells					
Cell Lines	0.00	0.00	0.00	0.00	0.34
PDX	0.00	0.00	0.00	0.00	<0.001
EGA	0.00	0.00	0.01	0.00	<0.001

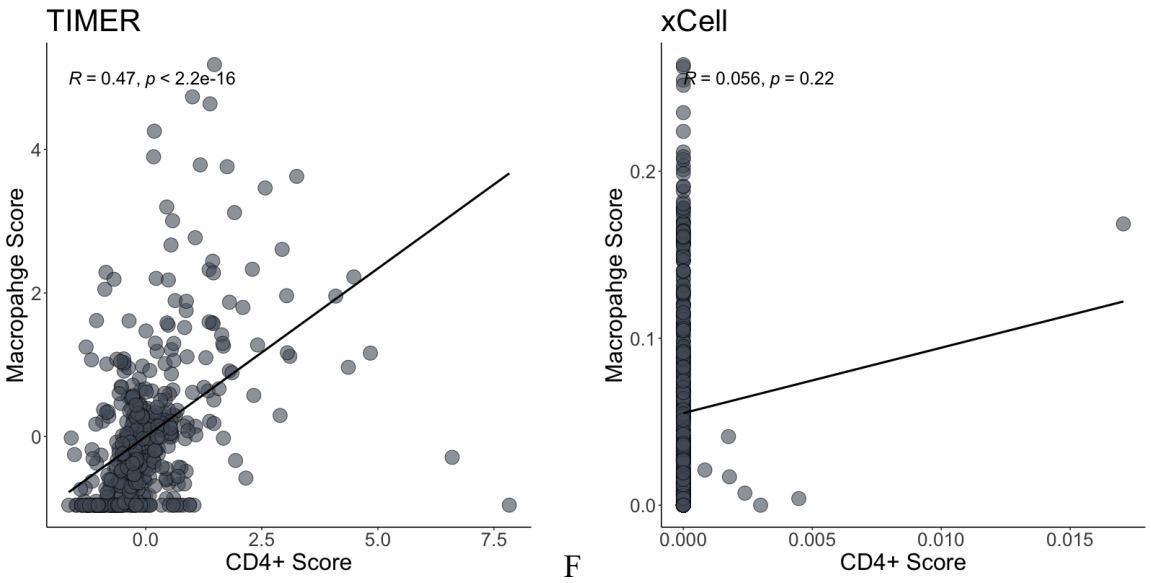
SRA	0.00	0.00	0.02	0.00	<0.001
TCGA	0.00	0.00	0.03	0.01	<0.001
UCLA	0.09	0.00	0.45	0.11	Ref
Th1 Cells					
Cell Lines	0.00	0.00	0.00	0.00	1.00
PDX	0.00	0.00	0.03	0.01	<0.001
EGA	0.00	0.00	0.03	0.01	<0.001
SRA	0.00	0.00	0.06	0.01	<0.001
TCGA	0.00	0.00	0.15	0.04	0.04
UCLA	0.06	0.00	0.24	0.07	Ref
Th2 Cells					
Cell Lines	0.00	0.00	0.03	0.01	1.00
PDX	0.10	0.03	0.19	0.05	1.00
EGA	0.12	0.00	0.43	0.09	1.00
SRA	0.11	0.00	0.48	0.09	1.00
TCGA	0.15	0.06	0.36	0.08	1.00
UCLA	0.12	0.00	0.35	0.10	Ref
Tregs					
Cell Lines	0.00	0.00	0.00	0.00	1.00
PDX	0.00	0.00	0.00	0.00	1.00
EGA	0.01	0.00	0.10	0.02	1.00
SRA	0.00	0.00	0.04	0.01	1.00
TCGA	0.00	0.00	0.04	0.01	1.00
UCLA	0.00	0.00	0.11	0.03	Ref

Table 3-21: Summary of immune cell populations across all USARC samples included in this study predicted by xCell in-silico immune deconvolution tools. This table corresponds to the graph in Figure 3-8G. Comparisons between groups were performed using the Wilcoxon test. Bonferroni correction was used for multiple hypothesis testing correction. Adjusted p values are shown.

3.9.5 Examination of CD4+ and macrophage in-silico correlation

We hypothesized that there would be a high correlation between macrophage and CD4+ T cell scores predicted by the in-silico immune deconvolution tools. We suspected this may be one reason for the discordance between CD4+ mIF and in-silico scores. We examined the correlation between CD4+ and macrophage scores in each tool using all USARC samples. The correlation between mIF and in-silico scores was determined using a linear fit model and Pearson correlation coefficients.





E F

Figure 3-9: Bar plots demonstrating the correlation between CD4+ and macrophage scores predicted by each of the in-silico immune deconvolution tools using all undifferentiated sarcoma samples in this study.

In-silico immune deconvolution tool	**R (Pearson correlation coefficient)**	**p value**
CIBERSORTx	0.50	<2.2e-16
EPIC	-0.16	0.00046
MCP Counter	0.22	1.2e-6
quanTIseq	-0.08	0.087
TIMER	0.47	<2.2e-16
xCell	0.06	0.22

Table 3-22: Correlation between CD4+ and macrophage scores in each of the in-silico immune deconvolution tools in this study. All undifferentiated sarcoma samples from all datasets were inclued in this analysis. This table summarizes the data depicted in Figure 3-9.

3.10 Bibliography

1. Petitprez F, de Reyniès A, Keung EZ, et al. B cells are associated with survival and immunotherapy response in sarcoma. *Nature*. 2020;577(7791):556-560. doi:10.1038/s41586-019-1906-8
2. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell*. 2015;160(1-2):48-61. doi:10.1016/j.cell.2014.12.033
3. Van Allen EM, Miao D, Schilling B, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science*. 2015;350(6257):207-211. doi:10.1126/science.aad0095
4. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453-457. doi:10.1038/nmeth.3337
5. Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *eLife*. 2017;6. doi:10.7554/eLife.26476
6. Becht E, Giraldo NA, Lacroix L, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol*. 2016;17(1):218. doi:10.1186/s13059-016-1070-5
7. Finotello F, Mayer C, Plattner C, et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med*. 2019;11(1):34. doi:10.1186/s13073-019-0638-6

8. Li B, Severson E, Pignon JC, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol*. 2016;17(1):174. doi:10.1186/s13059-016-1028-7
9. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol*. 2017;18(1):220. doi:10.1186/s13059-017-1349-1
10. Sturm G, Finotello F, Petitprez F, et al. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics*. 2019;35(14):i436-i445. doi:10.1093/bioinformatics/btz363
11. Avila Cobos F, Alquicira-Hernandez J, Powell JE, Mestdagh P, De Preter K. Benchmarking of cell type deconvolution pipelines for transcriptomics data. *Nat Commun*. 2020;11(1):5650. doi:10.1038/s41467-020-19015-1
12. Petitprez F, Sun CM, Lacroix L, Sautès-Fridman C, de Reyniès A, Fridman WH. Quantitative Analyses of the Tumor Microenvironment Composition and Orientation in the Era of Precision Medicine. *Front Oncol*. 2018;8:390. doi:10.3389/fonc.2018.00390
13. Pollack SM, Ingham M, Spraker MB, Schwartz GK. Emerging Targeted and Immune-Based Therapies in Sarcoma. *J Clin Oncol Off J Am Soc Clin Oncol*. 2018;36(2):125-135. doi:10.1200/JCO.2017.75.1610
14. Patnaik A, Kang SP, Rasco D, et al. Phase I Study of Pembrolizumab (MK-3475; Anti-PD-1 Monoclonal Antibody) in Patients with Advanced Solid Tumors. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2015;21(19):4286-4293. doi:10.1158/1078-0432.CCR-14-2607
15. Tawbi HA, Burgess M, Bolejack V, et al. Pembrolizumab in advanced soft-tissue sarcoma and bone sarcoma (SARC028): a multicentre, two-cohort, single-arm, open-label, phase 2 trial. *Lancet Oncol*. 2017;18(11):1493-1501. doi:10.1016/S1470-2045(17)30624-1

16. Ben-Ami E, Barysauskas CM, Solomon S, et al. Immunotherapy with single agent nivolumab for advanced leiomyosarcoma of the uterus: Results of a phase 2 study. *Cancer*. 2017;123(17):3285-3290. doi:10.1002/cncr.30738
17. Vivian J, Rao AA, Nothaft FA, et al. Toil enables reproducible, open source, big biomedical data analyses. *Nat Biotechnol*. 2017;35(4):314-316. doi:10.1038/nbt.3772
18. GenomeOC. Therapeutically Applicable Research to Generate Effective Treatments. Office of Cancer Genomics. Published January 18, 2013. Accessed May 5, 2021. https://ocg.cancer.gov/programs/target
19. Sayles LC, Breese MR, Koehne AL, et al. Genome-Informed Targeted Therapy for Osteosarcoma. *Cancer Discov*. 2019;9(1):46-63. doi:10.1158/2159-8290.CD-17-1152
20. Lesluyes T, Pérot G, Largeau MR, et al. RNA sequencing validation of the Complexity INdex in SARComas prognostic signature. *Eur J Cancer Oxf Engl 1990*. 2016;57:104-111. doi:10.1016/j.ejca.2015.12.027
21. EGA European Genome-Phenome Archive. Accessed July 29, 2022. https://ega-archive.org/
22. The Sequence Read Archive (SRA): Getting Started. Accessed July 29, 2022. https://www.ncbi.nlm.nih.gov/sra/docs/
23. The Cancer Genome Atlas Program - National Cancer Institute. Published June 13, 2018. Accessed May 5, 2021. https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga
24. Steele CD, Tarabichi M, Oukrif D, et al. Undifferentiated Sarcomas Develop through Distinct Evolutionary Pathways. *Cancer Cell*. 2019;35(3):441-456.e8. doi:10.1016/j.ccell.2019.02.002
25. UCSC Xena. Accessed July 29, 2022. https://xenabrowser.net/

26. SamToFastq (Picard). GATK. Accessed July 29, 2022. https://gatk.broadinstitute.org/hc/en-us/articles/360036485372-SamToFastq-Picard-
27. Chen L, Oke T, Siegel N, et al. The Immunosuppressive Niche of Soft-Tissue Sarcomas is Sustained by Tumor-Associated Macrophages and Characterized by Intratumoral Tertiary Lymphoid Structures. *Clin Cancer Res*. 2020;26(15):4018-4030. doi:10.1158/1078-0432.CCR-19-3416
28. Seomangal K, Mahmoud N, McGrath JP. Malignant fibrous histiocytoma, now referred to as Undifferentiated Pleomorphic Sarcoma: A Case Report of an unexpected histology of a subcutaneous lesion. *Int J Surg Case Rep*. 2019;60:299-302. doi:10.1016/j.ijscr.2019.06.035

CHAPTER 4

The transcriptomic and genomic signatures of immune cell infiltration in undifferentiated sarcoma

4.1 Abstract

Soft tissue sarcoma demonstrates poor response to immunotherapy overall. However, undifferentiated sarcoma is subtype as the most responsive to anti-PD1 immunotherapy, with 40% of patients demonstrating complete or partial response[1]. The drivers of this response are unclear and warrant further study. This study aimed to define the transcriptomic and genomic signatures associated with immune cell infiltration in undifferentiated sarcoma. We generated and aggregated next generation sequencing data from 192 unique undifferentiated sarcoma samples. Ultimately, 188 samples from four different datasets (EGA, SRA, TCGA, and UCLA) were included in this study. Immune cell infiltration was defined according to the total immune cell scores from the TIMER in-silico immune deconvolution tool, which was selected based on its superior performance described in Chapter 3. The transcriptomic and genomic analyses demonstrated that low immune cell infiltration is associated with high copy number changes in undifferentiated sarcoma.

4.2 Introduction

Immune-based therapies have dramatically changed the treatment of many cancers in recent decades. These include blockade of immune checkpoint molecules such as PD-1/PD-L1 and

CTLA-4, cellular therapies using chimeric antigen receptor or T-cell receptor modified T-cells, and emerging therapies modulating tumor-associated macrophages. While early studies of immunotherapy in STS have not had broad success, there are signals of efficacy that warrant further exploration[1–4]. The SARC028 study evaluated the efficacy of the anti-PD1 antibody pembrolizumab in patients with advanced sarcoma, and identified the undifferentiated pleomorphic sarcoma (UPS) subtype as the most responsive to anti-PD1 immunotherapy, with 40% of patients demonstrating complete or partial response[1]. This was notable given that UPS, unlike most tumors that are most responsive to immune checkpoint blockade, have a low mutational burden.

Soft tissue sarcoma (STS) is predominantly characterized by an abundance of genomic DNA copy number alterations. A few genes, specifically TP53, ATRX, and RB1, have been shown to be recurrently mutated across multiple sarcoma subtypes. However, overall mutational burden is low, particularly compared to other epithelial tumors and melanoma that respond to immune checkpoint blockade[5]. As such, the lower response rate of STS to immunotherapy in the SARC028 is unsurprising[1,6]. Yet, why there were higher response rates in UPS (40%) warrants further exploration[1].

Understanding the transcriptomic and genomic signatures associated with immune cell infiltration in undifferentiated sarcoma is critical to guiding future studies to develop optimal immunotherapy regimens for patients with this disease. Chromosomal instability, and in particular aneuploidy, has been shown to play a critical role in tumorigenesis, metastases, chronic inflammation, and immune cell infiltration in multiple cancer types[7–10]. For example,

Davoli et al. demonstrated that increased somatic copy number alterations (SCNA) was associated with decreased immune cell infiltration signatures across multiple cancer types[10]. Sarcoma, however, was notably missing from this analysis.

This study aimed to define the transcriptomic and genomic signatures associated with immune cell infiltration in undifferentiated sarcoma.

4.3 Materials and Methods

4.3.1 Patient and sample identification

Patients and samples were identified as described in the methods section of Chapter 3. All patients treated at our institution for sarcoma between January 1, 2010 and December 30, 2020 were identified using a prospectively maintained database. Patient, tumor, and treatment data were extracted from the electronic medical record. Sixty patients with untreated undifferentiated sarcoma (USARC) with available formalin-fixed paraffin-embedded (FFPE) tissue blocks were identified for inclusion in this study (Figure 4-1A). Hematoxylin and eosin (H&E) stained sections of each tissue block were re-reviewed by an expert sarcoma pathologist to confirm USARC pathology.

An additional four samples from before and after treatment with immunotherapy were included in portions of this analysis. These were paired samples obtained from two patients with undifferentiated sarcoma who were treated with immunotherapy.

4.3.2 Nucleic acid extraction, sequencing, and processing

RNA and DNA were simultaneously extracted from FFPE-preserved tissues using the Covaris tNA Plus Kit. DNA was extracted from blood or PBMCs, when available. RNA and DNA QC was performed using High Sensitivity RNA ScreenTape and Genomic DNA ScreenTape (Agilent TapeStation Software v3.2), respectively. Library preparation for RNA and whole exome sequencing was performed using hybrid capture technique using the Illumina TruSeq Exome kit. Paired-end RNA sequencing was performed using Illumina Hiseq 3000 platform with read length of 2x150 and a total of 40 million reads. All RNA-sequencing (RNA-seq) data was processed using the Toil pipeline, as described in previous publications[11]. Whole exome sequencing (WES) was performed on those samples for which matched-normal tissue (i.e. blood or PBMCs) were available. Paired-end whole exome sequencing was performed using the Illumina Hiseq 3000 platform with read lengths 2x150 with 100x coverage for blood and 200x coverage for tumors.

4.3.3 Processing of publicly available sequencing data

Transcriptomic data were obtained from the TCGA (SARC cohort), Lesluyes et al (GSE71121), and Steele et al (EGAD00001004439)[12–14]. TCGA SARC was restricted to the validated USARC samples defined by the TCGA group[12]. Myxofibrosarcoma (MFS) samples were excluded from the GSE71121 dataset[14]. Normal samples were excluded from the EGAD00001004439 dataset[13]. TOIL processed TCGA SARC data was downloaded from the UCSC Xena Browser portal[15]. All remaining RNA-sequencing data was Toil processed, as described in previous publications[11].

4.3.4 In-silico immune deconvolution

TIMER was used for in-silico immune deconvolution of the RNA sequencing data. This tool was selected based on its superior performance in undifferentiated sarcoma samples, as described in Chapter 3 (Figure 3-4). TIMER was run using the immunedeconv package in R using the "sarc" cancer type setting[16,17]. Samples were defined as high or low immune cell groups based on the median of the total immune score.

4.3.5 Transcriptomic and genomic analyses

Principal component analyses (PCAs) were performed on log2(tpm+1) data, using protein coding genes only. RNASeq data outliers were identified using PCA. Differential expression analysis (DESeq) was performed comparing samples with high versus low immune cell infiltration. Genes with sum of zero across all samples were dropped prior to the analysis. Dispersion estimates and MA plots were examined for each dataset. Statistical significance was defined as a q value less than 0.01. Rank-rank hypergeometric overlap was performed using the output of various DESeq results comparing samples with high versus low immune cell infiltration. The website RRHO version was used to perform the analyses. Step size was 100, as recommended[18]. GSEA and GO term analyses were performed using the Broad Institute Software[19]. GSEA analysis was performed using Hallmark gene sets with settings including 1000 permutations and no collapse[20]. GO term analysis was performed in the same manner. Heatmaps were scaled across the rows. Clustering results, when shown, were calculated using Euclidean distances. Copy number changes were defined as any segment with less than or more than 1 copy of either

allele in the UCLA cohort. The TCGA cohort values provided were log2(tumor/normal), and a copy number change was defined as absolute value >0.3, as described previously. In a sensitivity analysis, this definition was changed. The methods and results of the sensitivity analysis are described in the Appendix.

4.3.6 Statistical analysis

Correlation analyses were performed using a linear fit model. Pearson correlation coefficients are noted. Wilcoxon or chi square tests were used, as appropriate. Multiple hypothesis testing was performed using Bonferroni correction and noted where applicable.

4.3.7 Outlier definition

Principal component analyses and/or correlation plots were used to define outliers. When applicable, a Pearson mean correlation of <0.8 was used to identify outliers. Outliers defined by principal component analysis were defined by a panel of people who have extensive experience in PCA analyses and were blinded to the sample IDs and the study questions.

4.3.8 Sensitivity analysis

All transcriptomic analyses were performed on each dataset alone and all USARC samples together. In analyses where sequencing data from each dataset was analyzed separately, the immune cell infiltration groups (high versus low) were defined within each dataset separately. In

analyses where all sequencing data was analyzed together, the immune cell infiltration groups (high versus low) were defined with all in-silico immune deconvolution results together. TCGA copy number change cutoffs were changed and the results did not change (data not shown).

4.3.9 Software

The majority of statistical analyses and data visualizations were performed and generated in R (Version 4.0.5). Gene set enrichment analyses (GSEA) were performed using Broad Institute GSEA software (Version 4.2.3).

4.3.10 Approval

This study was approved by the UCLA Institutional Review Board (IRB #10-001857).

4.4 Results

4.4.1 Characteristics of study cohort

A total of 188 undifferentiated sarcoma (USARC) samples were included in this analysis. After exclusion of outliers there were 47, 42, 43, and 56 USARC samples in the EGA, SRA, TCGA, and UCLA cohorts, respectively (Table 4-1). The clinical characteristics of in-house samples are summarized in Table 4-2 and the results of the univariate analysis comparing samples with high and low immune cell infiltration is summarized in Table 4-3. The clinical characteristics of the

publicly available datasets are summarized in Tables 4-7 through 4-9, and corresponding univariate analysis are shown in Tables 4-11 through 4-13.

Of the in-house, untreated undifferentiated sarcoma samples, 34 (57%) were from male patients. The median age at the time of tissue acquisition (biopsy or surgery) was 62 years (range 19-101 years). The majority (n = 57, 95%) of samples were obtained from tumor resection while the remainder were obtained by incisional biopsy. The majority of tumors were located in the trunk/extremity (n = 45, 75%) or retroperitoneum/abdomen/pelvis (n=10, 17%), while a minority (n = 5, 8%) came from other sites, such as head or breast. The median size was 5.7cm (range 0.8-27.0). The majority of samples came from patients with primary disease (n = 51, 85%), while few came from recurrent or metastatic lesions (n = 5, 8% and n = 4, 6 %, respectively). Approximately half (n = 29, 48%) of patients had no evidence of disease (NED) at the time of follow up, and 21.7% (n = 13) were alive with disease (AWD), 2% (n = 1) had died of other causes (DOO), and 28% (n = 17) had died of disease (DOD) at the time of follow up. The median follow-up time was 5 years (range 0-11 years) (Table 4-2). There was no statistically significant difference in any of these characteristics when comparing samples with high versus low immune cell infiltration (Table 4-3).

4.4.2 Principal component analyses demonstrate a high degree of overlap of transcriptomic results when comparing undifferentiated sarcoma across various datasets

A principal component analysis of RNA-Seq data from multiple sarcoma subtypes and datasets is shown is Figure 4-1. As noted in chapter 3 the comparison of principal component 2 (PC2) and

PC3 demonstrates that UCLA USARC samples cluster with USARC samples from other publicly available datasets. When the data is restricted to USARC samples only, there is a significant amount of overlap across the four USARC datasets (EGA, SRA, TCGA, and UCLA) when comparing PC2 and PC3 (Figure 4-1C). Of these datasets, the EGA and SRA samples demonstrate the most overlap. TCGA and UCLA samples also overlap these samples in the same region, though there is most dispersion of these datasets across PC2 (TCGA) and PC3 (UCLA).

While principal component (PC) 2 and PC3 showed that our in-house USARC data clustered with other USARC subtypes (Figure 4-1), PC1 versus PC2 showed that our in-house samples clustered together (Figure 4-10), suggesting there are differences between the in-house RNA-Seq data and publicly available subtypes. These findings were consistent when including multiple sarcoma subtypes or USARC only.

4.4.3 Gene set enrichment analyses support that in-silico immune deconvolution accurately characterizes tumors with high versus low immune cell infiltration

Differential gene expression (DESeq) analysis was performed comparing tumors with high versus low immune cell infiltration (defined relative to the median of TIMER in-silico immune deconvolution tool results) in each of the USARC datasets (EGA, SRA, TCGA, and UCLA) separately as well as all the data together. Representative volcano plots depicting DESeq results are shown in Figure 4-14. Gene set enrichment analyses (GSEA) and gene ontology (GO) term analyses were performed on these results. The gene sets associated with samples in the high immune cell infiltration group, as defined by TIMER, were associated with gene sets associated

with the immune system. The GSEA results varied across the datasets, however, the theme overall was gene sets associated with the immune system, which was true when the datasets were analyzed separately or together (Figure 4-15). These results support the method of in-silico immune deconvolution that was used to define the samples with high versus low immune cell infiltration.

4.4.4 Rank-rank hypergeometric overlap demonstrates global similarities between tumors with low immune cell infiltration across datasets

Rank-rank hypergeometric overlap (RRHO) was performed to compare the gene lists derived from each of the DESeq comparisons, specifically comparing samples with high versus low immune cell infiltration in each of the USARC datasets (EGA, SRA, TCGA, UCLA) and in all the USARC datasets combined. The results of RRHO suggested that there was the highest degree of overlap with the genes associated with low immune cell infiltration across the datasets. The results from the UCLA dataset were the most different from the others, however, there was still a fair degree of overlap when examining the genes associated with low immune cell infiltration. The results were similar when compared to the DESeq results when all USARC samples were used (Figure 4-2).

4.4.5 Copy number alterations are associated with low immune cell infiltration in undifferentiated sarcoma

Copy number analyses demonstrated that samples with low immune cell infiltration demonstrated statistically significantly higher copy number alterations when compared to tumors with high immune cell infiltration. This finding was seen in both the TCGA and UCLA USARC cohorts (Figure 4-4). We also examined this association in the TCGA SARC cohort, and determined this association within each subtype of STS. Figure 4-5 demonstrates that low immune cell infiltration was associated with higher copy number changes in USARC and DDLPS samples. However, there was no association between immune cell infiltration and copy number change in the other sarcoma subtypes or in the STS cohort overall (Figure 4-5).

The GSEA and GO term results associated with the samples with low immune cell infiltration were frequently associated with cell division. Hallmark GSEA gene sets, such as "mitotic spindle" and "G2M checkpoint" as well as GO results related to chromosome or chromatid segregation were associated with low immune cell infiltrated samples (Figure 4-3B&C). This was true when all protein coding genes were included in the analysis or when only the cell lines genes were included. The results were similarly consistent when each dataset was analyzed separately and when all were analyzed together (Figure 4-15).

4.5 Discussion

Genomic instability is a hallmark of cancers[21]. Genomic instability in the context of DNA repair deficiencies drives an increased number of tumor associated antigens and/or neo-antigens. This process would be expected from a purely stochastic perspective to increase the likelihood of recognition by tumor-specific T cells. This was demonstrated to be the case in multiple cancers

including lung, melanoma, and colon or rectum, as TMB positively correlated with CD8 T cell infiltration and improved patient outcome[22–25].

However, higher order tumor genomic instability, including copy number alteration and aneuploidy, have been demonstrated to confer a selective advantage[26]. High broad CNA in HCC correlated with proliferation and immune evasion. Conversely, low broad CNA correlated with increased HLA-A expression and response to PD1 blockade[27]. Similarly, higher ploidy was found to correlate with lower adaptive immune signatures; including IFNg, cytotoxic CD8 T cells, NK cells, B cells, and chemokine receptor interactions; across BRCA, HSNC, CRC, LUAD, LUSC, OV, SKCM, STAD, UCEC, and KIRC. Moreover, the ratio of observed-to-expected neoantigens in tumors with higher ploidy suggests less immune editing, a finding echoed in TNBC[10,28].

Multiple mechanistic explanations for the link between genomic instability and immune subversion have been suggested. It may be that neoantigens are eliminated by successive rounds of immunoediting, LOH including antigen presentation machinery, hypermethylation and silencing of neoantigen promoters, or genomic imbalances leading to decreased frequency of strong antigens being loaded into MHC complexes[10,23,29]. One common manifestation of chromosomal instability is presence of intracellular micronuclei, which collapse to cause cytosolic leak of DNA. Aberrant DNA localization then activates cGAS-STING pathway driving IFNa signaling and immune activation[30]. However, chronic type I IFN signaling drives T cell exhaustion[31]. Experimental models of polyploid tumor cells may provide further insight. Induction of aneuploidy in human retinal pigment epithelial cells led to elimination by NK cells

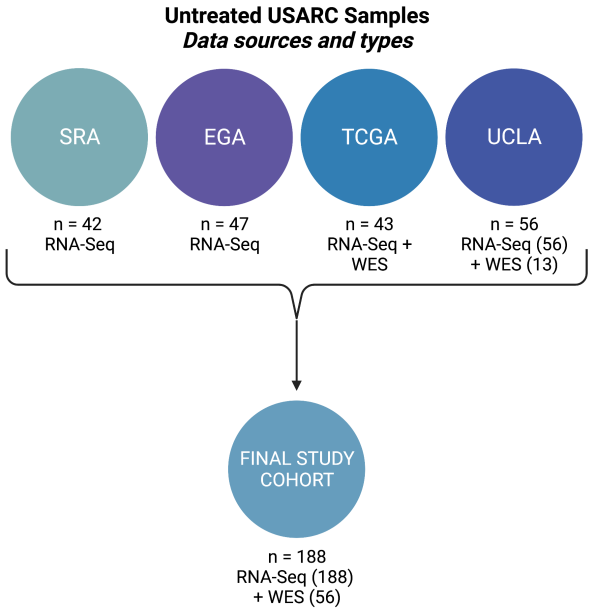
in co-culture, while euploid cell were spared[32]. Similarly, induced polyploid cancer cells grow unrestrained in immunocompromised mice but often fail to establish in those with an intact immune response, possibly through ER stress and subsequent calreticulin exposure[33]. Taken together, it may be that aneuploid cells are effectively eliminated in the early stages of tumorigenesis; however, once tumors become established, additional mechanisms sufficiently curtail immune-mediated selective pressure against chromosomal instability, enabling accrual of additional mutations and eventual selective advantage.

Detailed studies regarding the correlation between genomic instability and immune response in soft tissue sarcomas are lacking. One study reported an "immune-low" profile in UPS with higher rates of CNA[34]. Additionally, PDL1 copy number gain has been proposed as a mechanism of immune suppression in STS, correlating with worse survival in TCGA analysis[35]. Our findings demonstrate an association between high number of copy alterations and low immune cell infiltration in undifferentiated sarcoma. This association was seen in both transcriptomic analyses of multiple datasets (EGA, SRA, TCGA, and UCLA) as well as genomic studies of both the TCGA and UCLA cohorts. Interestingly, this association was not seen in all STS subtypes, rather only UPS and DDLPS. The underlying mechanisms for the association between copy number changes and immune cell infiltration in soft tissue sarcoma are not clear and warrant further study. Further, the unique association between CNA and immune cell infiltration in the DDLPS and UPS subtypes of STS warrants further study. Limitations of this study are described in detail in the discussion of Chapter 3.

4.6 Conclusion

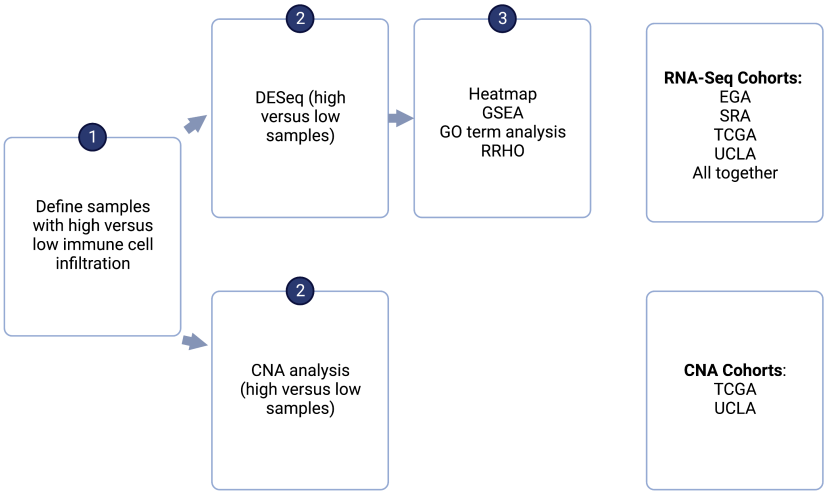
In this study, we aimed to define the transcriptomic and genomic drivers of immune cell infiltration in undifferentiated sarcoma. We found that increased copy number changes were associated with low immune cell infiltration in undifferentiated sarcoma. These findings were suggested in both transcriptomic and genomic analyses. Interestingly, this association between CNA and immune invasion were unique to the UPS and DDLPS subtypes of STS, but it was not seen in other subtypes of STS. The mechanisms underlying this association are not clear and warrant further study.

4.7 Figures



A

B



C

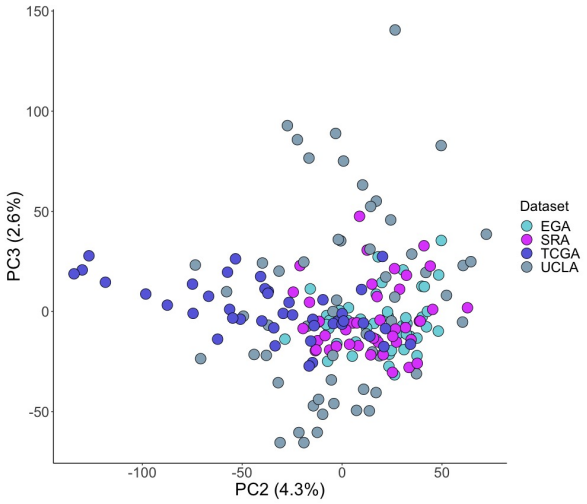


Figure 4-1: Overview of experimental design and study cohorts. (A) Summary of the data types and sources included in this study. Nucleic acid sequencing data was downloaded from publicly available sources (EGA, SRA, TCGA) or generated in-house (UCLA). (B) The analysis pipeline in this study. RNA-Seq data was available for all samples. Datasets were analyzed separately as well as all together. WES and copy number alteration data was analyzed when available (TCGA & UCLA). (C) PCA analyses of RNA-Seq data from untreated human USARC tumor samples in this study (EGA, SRA, TCGA, and UCLA). PCAs were performed on protein coding genes only and are colored by dataset. PCs and their contribution to the variance are noted.

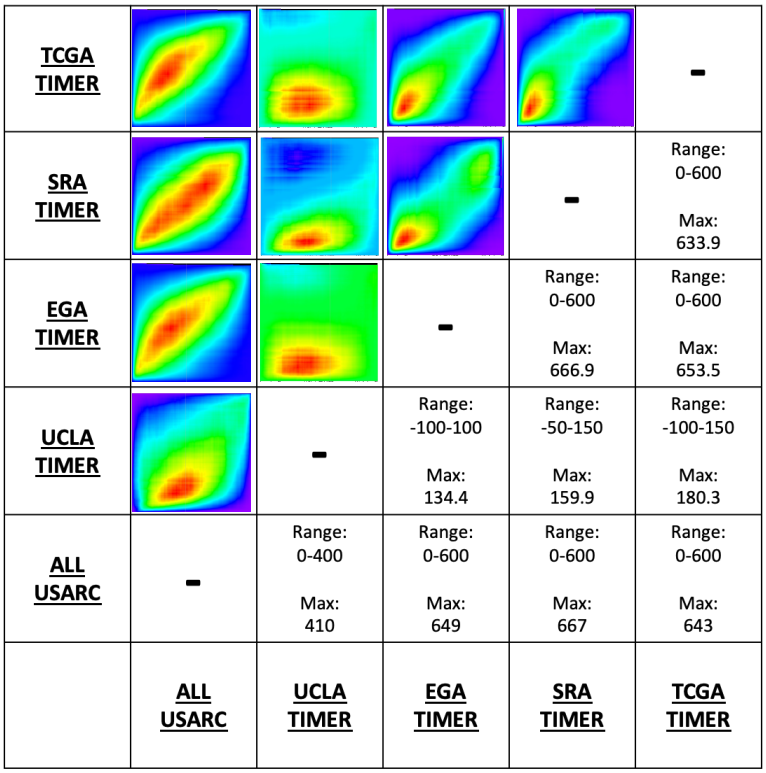
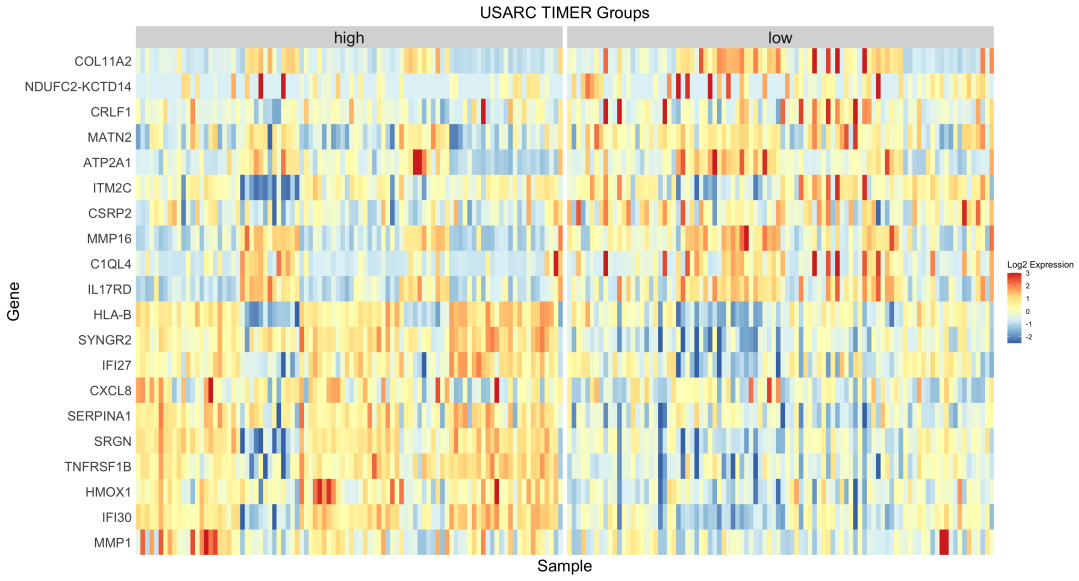
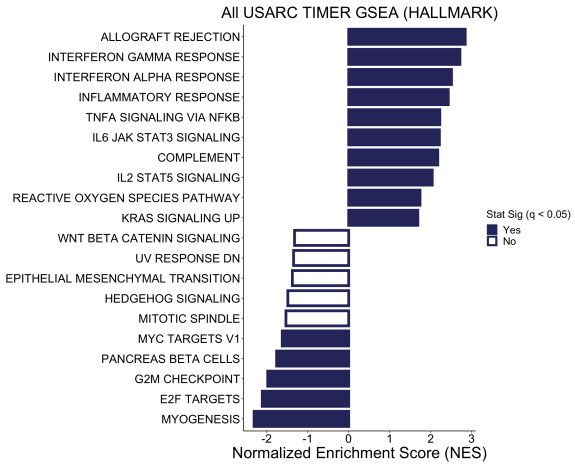


Figure 4-2: RRHO analysis comparing USARC tumor samples with high versus low immune cell infiltration. RRHO demonstrates global similarities between tumors with low immune cell infiltration across datasets.



A

B



C

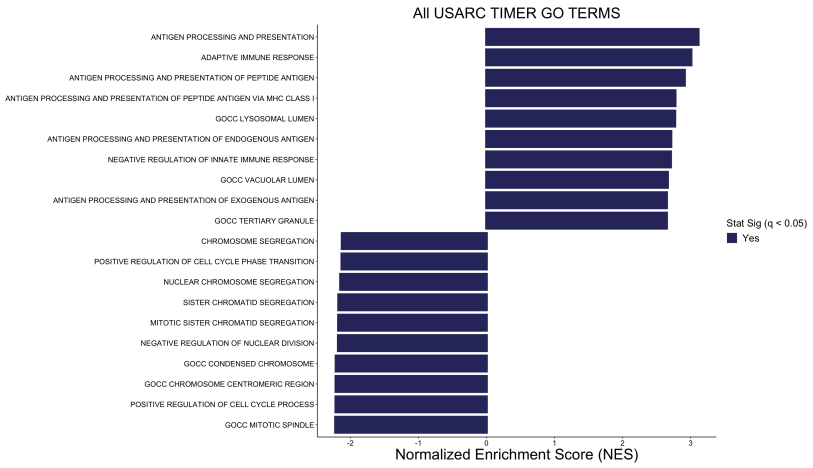


Figure 4-3: Transcriptomic analysis comparing USARC tumor samples with high versus low immune cell infiltration. Together, these results support that in-silico immune deconvolution accurately characterizes tumors with high versus low immune cell infiltration. It further suggests an association between gene sets related with cell division and tumors with low immune cell infiltration. (A) Heatmap of top 10 and bottom 10 genes on DESeq comparing samples with high versus low immune cell infiltration. (B&C) GSEA & GO term analyses (respectively) of results of DESeq comparing samples with high versus low immune cell infiltration.

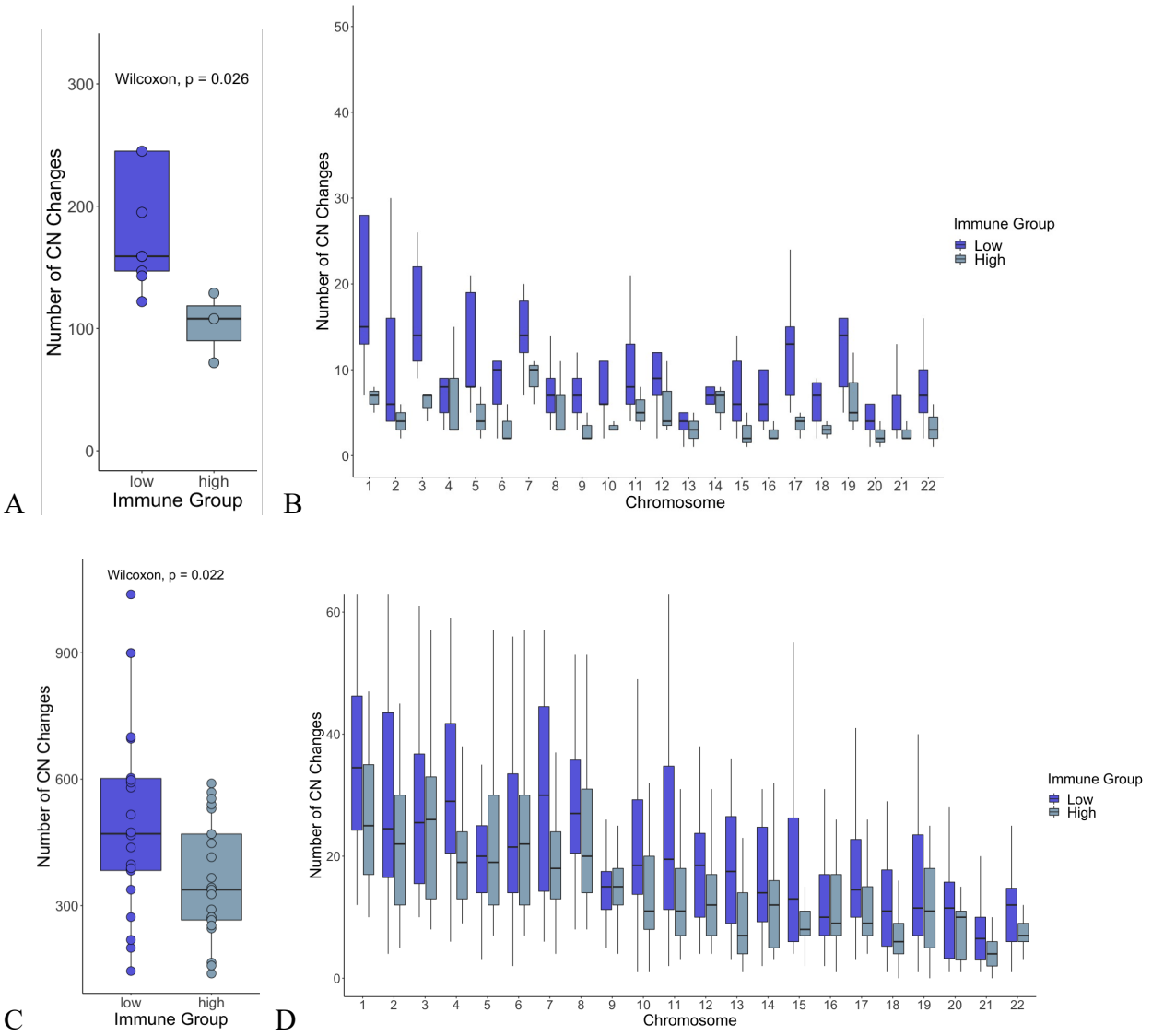


Figure 4-4: Increased copy number alterations are associated with lower immune cell infiltration in undifferentiated sarcoma. Comparisons of total number and per chromosome copy number segment changes are shown. Results were analyzed in the UCLA cohort (A&B) and in the TCGA cohort (C&D).

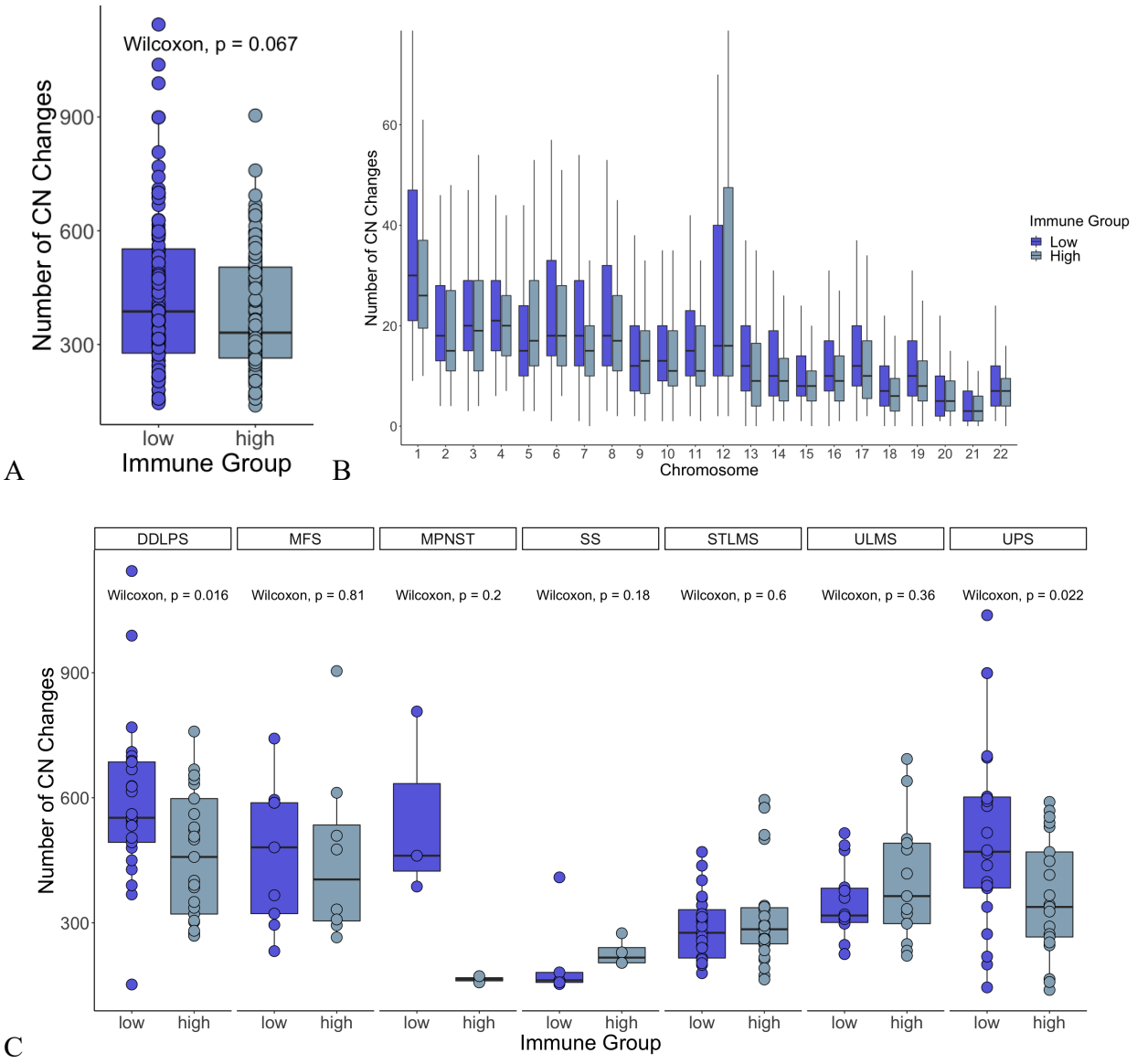


Figure 4-5: Increased copy number alterations are associated with lower immune cell infiltration in USARC and DDLPS, but not in other subtypes of STS and not in STS overall. (A&B) Bar graphs demonstrating the association between copy number changes and immune cell infiltration in all subtypes of STS from the TCGA. Comparisons were made examining total copy number changes (A) and total copy number changes within each chromosome (B). There was a trend toward increased copy number changes in the low immune cell infiltration group, though this did not reach statistical significance (p = 0.067). (C) Increased copy number alternations were

associated with low immune cell infiltration in USARC and DDLPS subtypes of STS, though there was no difference in copy number changes between high and low immune cell infiltration groups in the other subtypes of STS. Immune groups were defined relative to the median TIMER score within each histology.

4.8 Tables

Dataset	n
EGA	47
SRA	42
TCGA	43
UCLA	56
Total	188

Table 4-1: A summary of the data source and sample sizes of the untreated human tumor USARC samples included in this study. Sample sizes listed exclude outliers. Additional sarcoma subtypes and sample types were included in various analyses when noted. A complete list of all samples and data sources in included in Table 4-4. Outlier identification is summarized in Table 4-6.

Characteristic	Median (range) or n (%)
Sex	
Female	26 (43.3%)
Male	34 (56.7%)
Age	62 (19-101)
Tissue Type	
Incisional biopsy	3 (5.0%)
Resection	57 (95.0%)
Primary Site	
RP/Abdomen/Pelvis	10 (16.7%)
Trunk/Extremity	45 (75.0%)
Other	5 (8.3%)
Size (cm)	5.7 (0.8-27.0)
Disease Status	
Primary	51 (85.0%)
Recurrence	5 (8.3%)

Metastasis	4 (6.7%)
Status	
NED	29 (48.3%)
AWD	13 (21.7%)
DOO	1 (1.7%)
DOD	17 (28.3%)
Follow-Up Time (yrs)	5 (0-11)

Table 4-2: Summary of patient and tumor characteristics from 60 patients with untreated USARC tumors treated at UCLA from 2010-2020. Samples were selected as described in the CONSORT diagram in Figure 3-1 of Chapter 3. Follow up time listed is the years from the date of surgery to the date of last follow-up and excludes patients who died.

	Low		**High**		**p**
	n or median	*% or range*	*n or median*	*% or range*	
Total	28	50.0	28	50.0	
Sex					0.420
Female	14	56.0	11	44.0	
Male	14	45.2	17	54.8	
Age	62	29-88	62	19-101	0.512
Tissue Type					0.553
Incisional biopsy	2	66.7	1	33.3	
Resection	26	49.1	27	50.9	
Primary Site					0.497
RP/Abdomen/Pelvis	4	40.0	6	60.0	
Trunk/Extremity	21	50.0	21	50.0	
Other	3	75.0	1	25.0	
Size (cm)	5.2	1.4-18.3	5.9	0.8-27.0	0.855
Disease Status					0.189
Primary	21	44.7	26	55.3	
Recurrence	4	80.0	1	20.0	
Metastasis	3	75.0	1	25.0	
Status					0.579
NED	11	44.0	14	56.0	
AWD	7	53.8	6	46.2	
DOD	10	58.8	7	41.2	
DOO	0	0.0	1	100.0	
Follow-Up Time (yrs)	4	0-10	3	0-11	0.419

Table 4-3: Univariate analysis patient and tumor characteristics from 60 patients with untreated USARC tumors treated at UCLA from 2010-2020. Samples were selected as described in the CONSORT diagram in Figure 3-1 of Chapter 3. Characteristics are compared between tumors with high versus low immune cell infiltration, defined relative to the median TIMER total immune score. Chi square test or Wilcoxon test were used as appropriate. Follow up time listed is the years from the date of surgery to the date of last follow-up and excludes patients who died.

4.9 Appendix to Chapter 4

4.9.1 Summary of datasets included in this analysis

Sequencing data (primarily RNA-Seq and occasionally WES where noted) from multiple data sources was included in this study. The sources, histologies, and samples sizes of all the data aggregated in processed for this manuscript is summarized below. In most instances, human USARC samples were used in the analyses. When other histologies and/or sample types were included, this was specifically noted.

Dataset	Histology	n	Notes
Cell Lines	Multiple STS (4 USARC, 1 fibrosarcoma, 1 GSC)	12	6 cell lines 12 total RNA-Seq samples (pre/post BO112 for each cell line)
EGA	USARC	48	Normal samples were downloaded from EGA but not processed
PDX	Multiple sarcoma subtypes (DDLPS, STLMS, MFS, USARC, OS, SS, ULMS, other = epithelioid, Ewing's, fibrosarcoma, myxoid LPS, FDCS, RMS)	42	1 IM myxoma was dropped prior to processing

SRA	USARC	42	16 MFS were dropped prior to data processing
TARGET	OS	81	
TCGA	Multiple STS (DDLPS, MFS, MPNST, SS, STLMS, ULMS, UPS)	204	
UCLA	USARC	65	60 untreated samples, 1 sample redo (dropped), 2 paired samples from 2 patients (pre/post immunotherapy)
TOTAL	**Multiple sarcoma subtypes**	**494**	

Table 4-4: Summary of data sources, sarcoma subtypes, and sample sizes aggregated and processed for this study.

4.9.2 Sample selection based on tumor necrosis

One of the sample exclusion criteria for the in-house samples was necrosis >35% (see CONSORT diagram, Figure 3-1A). Prior studies used exclusion criteria of 40% (Lesluyes et al), 20-30% (TCGA), or 50% (Steele et al) necrosis when selecting tumor samples[12–14]. Based on these standards in the sarcoma literature, we selected a cut off of 35% necrosis for inclusion in our study. As a result, 14 samples were excluded from the final study population.

4.9.3 Summary of sequencing techniques from various data sources

The majority of the analyses in this manuscript focus on human tumor data from the following sources: EGA, TCGA, SRA, and UCLA. There was significant variation in the sample storage and sequencing techniques used in each of these datasets. Select available data is summarized in the table below.

Dataset	Tissue type	RNA Extraction Kit	Platform	Mode	Sequencing Length	Sequencing Depth
EGA	Frozen	Zymo Direct Zol RNA isolation kit	Illumina HiSeq 2000	Paired end	76bp	35M
SRA	Frozen and FFPE	TRIzol (Life Technologies) chloroform extraction [Frozen], Deparaffinization Solution and RNeasy FFPE kits (Qiagen) [FFPE]	Illumina HiSeq 2500	Paired end	100bp	60M
UCLA	FFPE	Covaris tNA Plus Kit	Illumina Hiseq 3000	Paired end	150bp	40M

Table 4-5: Summary of sequencing techniques for EGA, SRA, and UCLA. The sequencing techniques for TCGA vary depending on the submitting institution[36].

4.9.4 RNA-Seq fastqc analysis and in-house sample quality control

As described above, fastq files were downloaded or generated for the EGA, SRA, and UCLA datasets[11,13–15]. These files were then processed for downstream analyses using the Toil pipeline. The TCGA data was previously Toil processed and this processed data (rather than the fastq files) was downloaded from Xena browser. As such, the TCGA data is not included in this fastqc analysis.

The quality of each fastq file (n=155) was analyzed using the fastqc R package. Outputs from the fastqc package include analyses of adapter content, overrepresented sequences, per base N

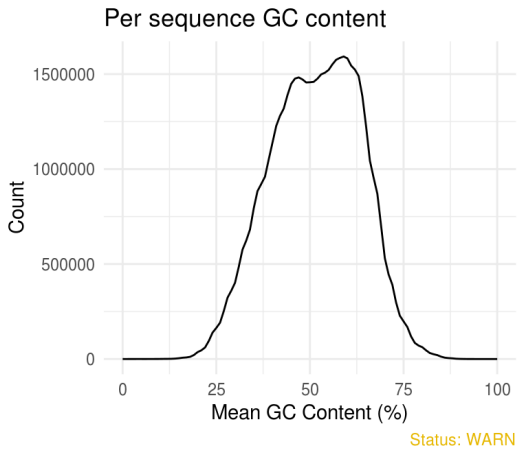
content, per base sequence content, per base sequence quality, per sequence GC content, per sequence quality scores, per tile sequence quality, sequence duplication levels, and sequence length distribution. Any outputs that had noted issues (i.e. labeled as "warn" or "fail") were examined. Many of the outputs noted no issues in any of the datasets. The two outputs that most frequently noted "warn" or "fail" for the fastq files in these datasets were the "per base sequence quality" and the "per sequence GC content" outputs.

In the SRA dataset, 13 samples were labeled as "warn" on the per sequence GC content. A representative figure is shown in Figure 4-6A. While these results did results in a "warn" error message, this graph demonstrates results of a GC content graph that are within normal limits. The typical mean GC-content in human genomes is 40% with a range from 35-60%[37–39]. The peaks of the graphs fall within this range, and generally show a normal distribution, as expected. In the SRA dataset, 3 samples were labeled as "fail" on the per base sequence quality analysis. A representative result is shown in Figure 4-6B. The decrease in per base sequence quality seen at the end corresponds to the adapter sequences, which are trimmed during Toil processing. In the EGA dataset, 11 samples were flagged as "warn." A representative figure is shown in Figure 4-6C. The notch seen on the left side of the graph likely corresponds to poly-A tail enrichment, and should not results in any issues with sample quality[40].

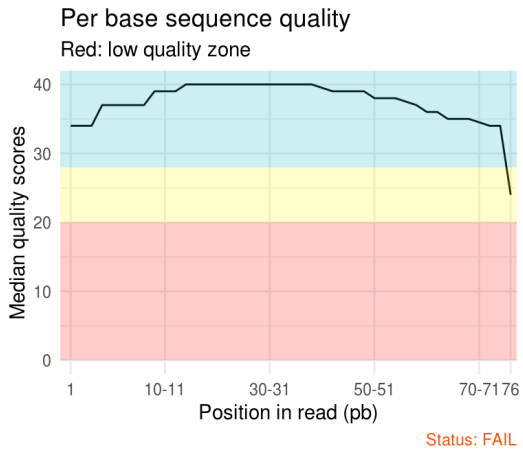
Representative paired samples of per base sequence quality analysis for the UCLA data are shown in Figure 4-6D (forward read) and 4-6E (reverse read). Most forward reads demonstrated high quality throughout, while the reverse reads typically demonstrated a decrease in quality toward the end of read. It is not atypical to have a decrease in quality of the per base sequence

quality in the reverse read compared to the forward read, however, this degree of decline was notable. Many additional downstream analyses and corrections, which are described in later sections, to attempt to account for and explain this decrease. Ultimately, after many analyses and discussions with experts in the field, the decision was made to not exclude these samples base on the results of the per base sequence quality on the reverse strand alone. Finally, there were 3 samples in the UCLA dataset that had per base sequence quality as shown in Figure 4-6F and 4-6G (forward and reverse strands, respectively). Although these samples were defined as "pass" in the fastqc results, it is abnormal to see no variation in the per base sequence quality across a read. The decision was made to drop these samples based on these results, though these samples were confirmed to be outliers on downstream analyses as well.

Three additional sarcoma RNA-Seq datasets (TARGET, PDX, Cell lines) were included in this analysis, as described above. These datasets were used in very few analyses throughout the project and were primarily used to elucidate overall trends across tumor types. The quality control performed on these datasets (n=135) was not as rigorous as that described for the human tumor USARC samples, which were the focus on this analysis.



A



B

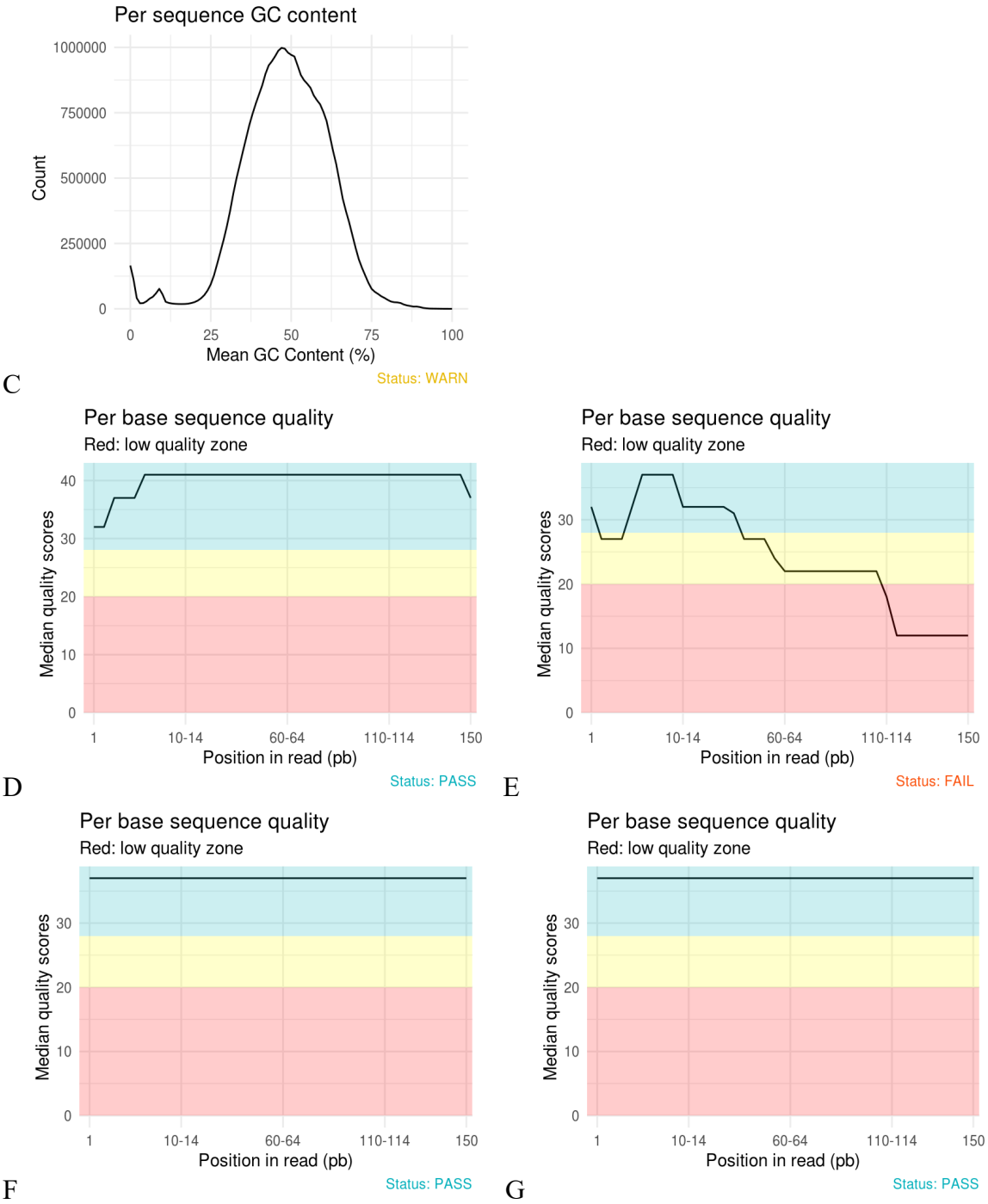


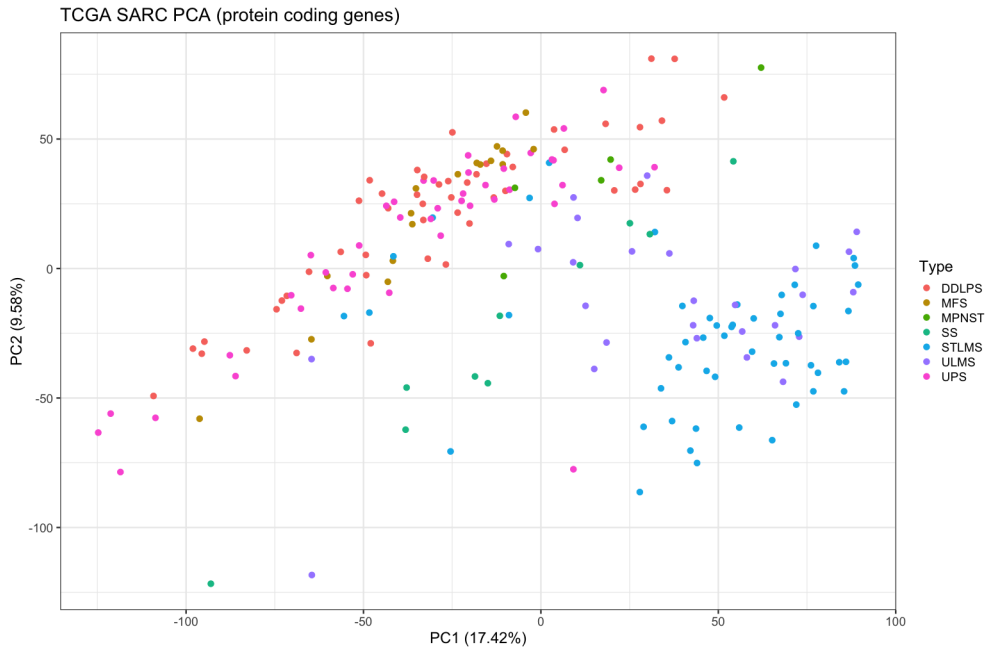
Figure 4-6: Summary of fastqc analysis results. All fastq files from EGA, SRA, and UCLA datasets were analyzed using the fastqc package. (A) A representative example of the per sequence GC content analysis resulting in an error in the SRA dataset. (B) A representative example of the per sequence base quality analysis resulting in an error in the SRA dataset. (C) A

representative example of the per sequence GC content analysis resulting in an error in the SRA dataset. (D&E) A representative example of the per sequence base quality analysis resulting in an error in the UCLA dataset. D demonstrates a typical forward read and E represents a typical reverse read. (F&G) An example of the per sequence base quality analysis resulting in an error in three samples in the UCLA dataset. F demonstrates a typical forward read and G represents a typical reverse read.

4.9.5 Identification of outliers

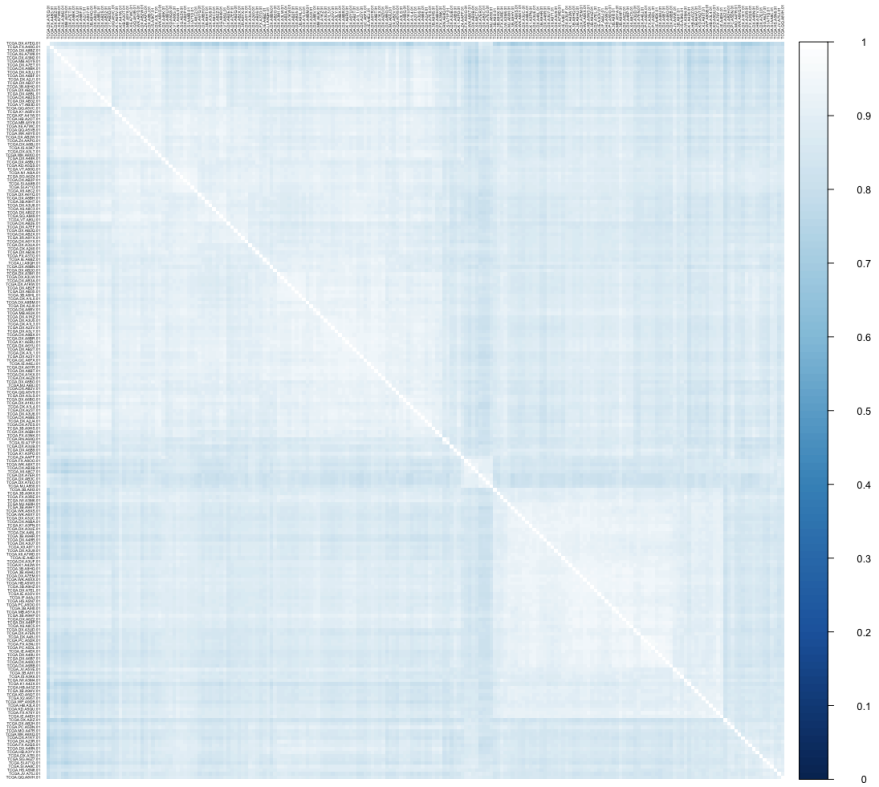
The figures and tables below note the outliers identified from each of the datasets. Outliers were defined as described in the methods section. Both PCAs and correlation plots were used to define outliers in the data. PCAs were explored up to PC1 vs PC10. Graphs depicting PC1 vs PC2 were shown below as higher PC graphs did not change the outlier analyses. Correlation plots demonstrate Pearson correlation values. Each of the datasets was analyzed separately when identifying outliers. Sample IDs were removed from all graphs that do not depict publicly available data. Figure 4-7 A&B demonstrates the results of the TCGA SARC outlier analysis. There was no clear outlier identified on PCA or correlation plot. Figure 4-7 C&D demonstrates the results of the TARGET OS outlier analysis. There was one clear outlier identified on both PCA and correlation plot. The sample is seen the top right corner of the PCA graph and is the dark line (lowest correlation) on the correlation plot. As such, one outlier was identified. Figure 4-7 E&F demonstrates the results of the EGA USARC outlier analysis. On PCA, there is a clear outlier. This sample corresponds to the sample with lowest correlation on the correlation plot. The decision was made to exclude this sample. Figure 4-7 G&H demonstrates the results of the

SRA USARC outlier analysis. There was no clear outlier identified on either the PCA or the correlation plot. Figure 4-7 I&J demonstrates the results of the sarcoma cell line outlier analysis. In this dataset, each of the cell lines were sequences before and after treatment with BO112. Both the correlation plot and the PCA depict a high correlation between paired samples from each cell lines. There was no clear outlier identified. Figure 4-2 K&L demonstrates the results of the sarcoma PDX outlier analysis. There was no clear outlier identified on either the PCA or the correlation plot. Figure 4-7 M&N demonstrates the results of the UCLA USARC outlier analysis. There are five samples (2 are duplicates of the same sample) that appear to be outliers on the PCA. These are seen on the lower left portion of the graph. These five samples correspond to the samples with the lowest correlation on the correlation plot and were determined to be outliers. The correlation between samples seen in the UCLA dataset was significantly lower than that seen between samples in each of the other datasets in this analysis. Possible explanations for this and additional batch effects are explored in the following sections. Table 4-6 summarizes the number of outliers and final sample sizes for each dataset.

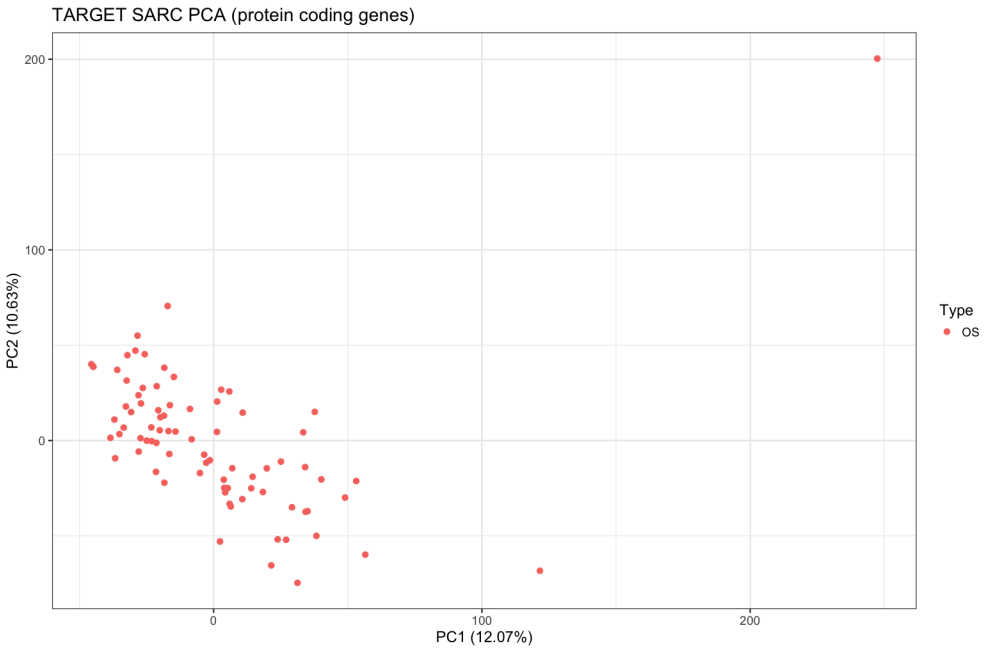


A

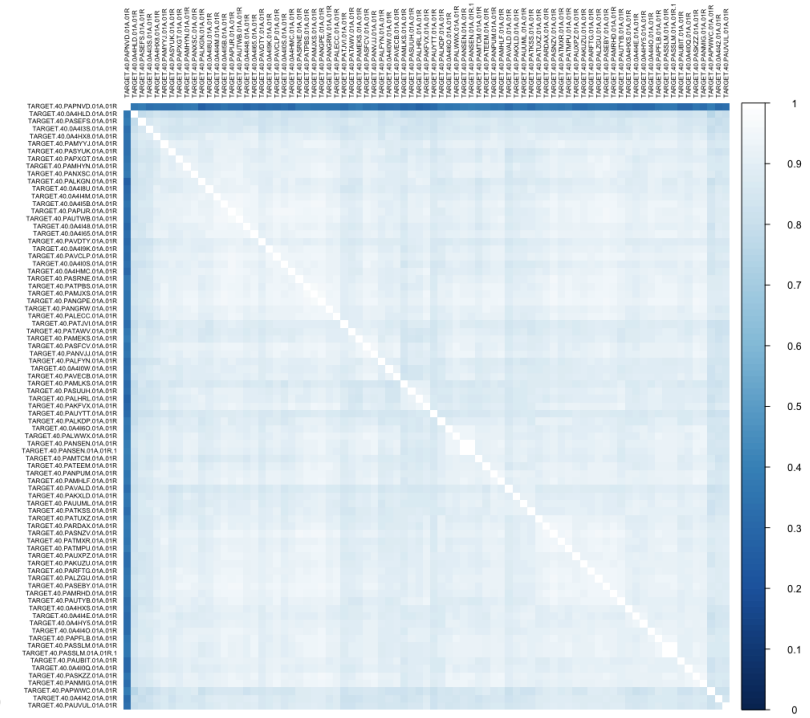
B


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

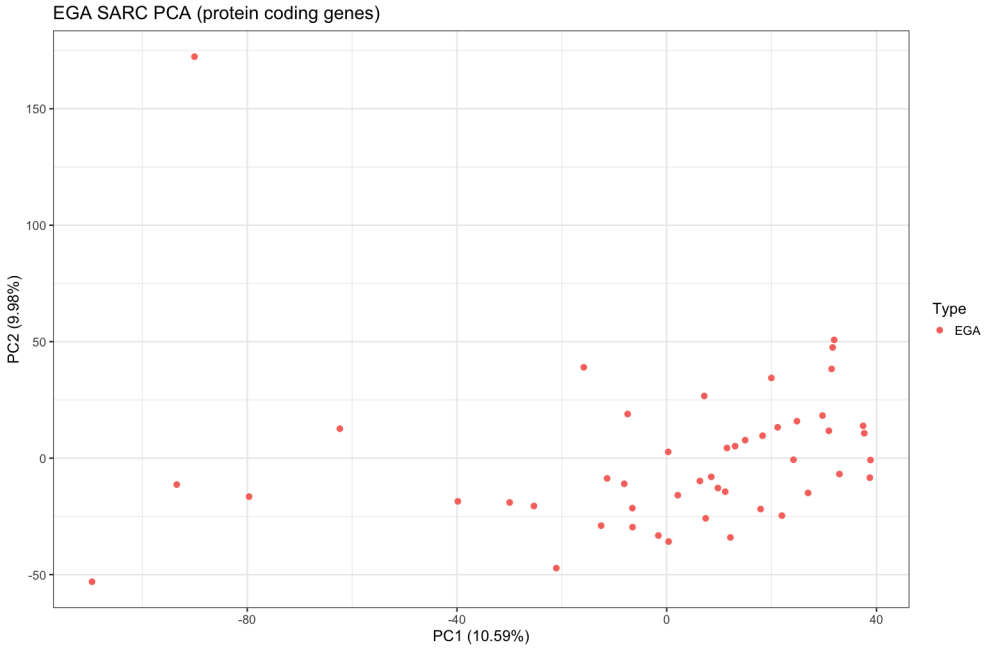
C


TARGET SARC PCA (protein coding genes)
200
100
0
PC2 (10.63%)
0
100
200
PC1 (12.07%)
Type
OS

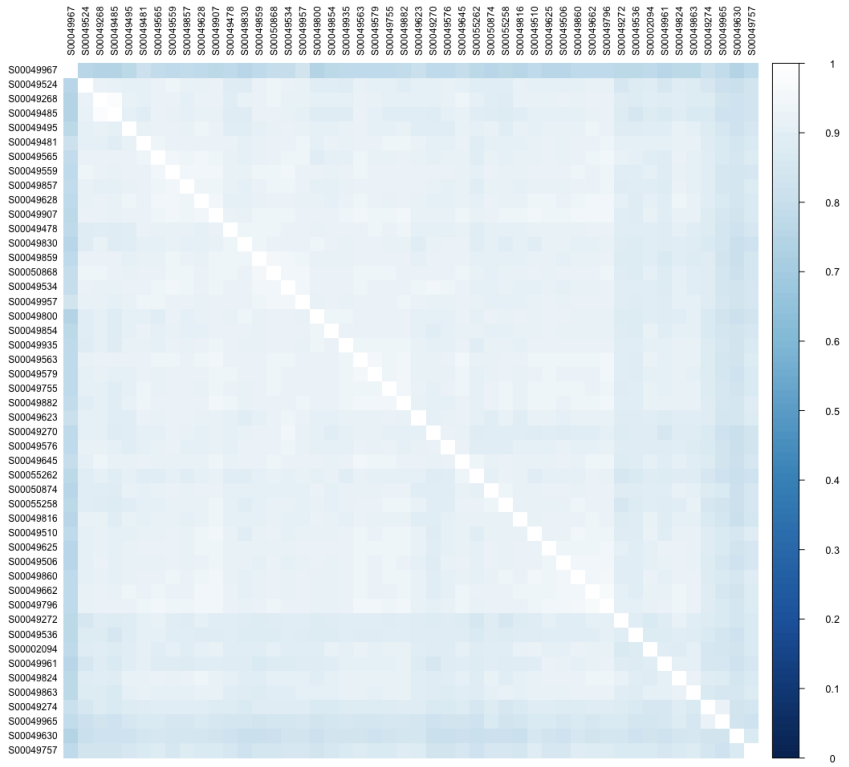
D


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

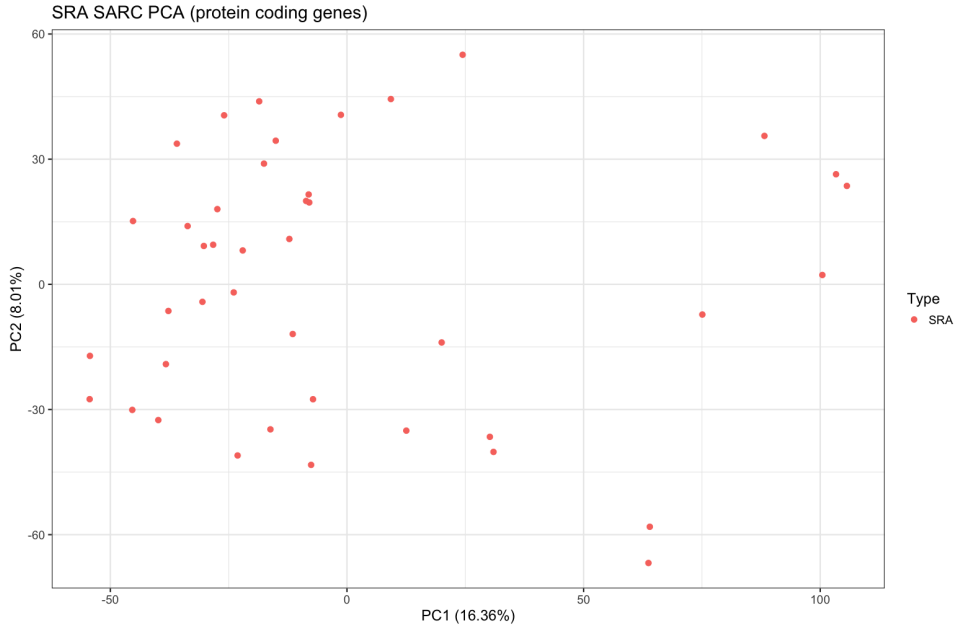
E


EGA SARC PCA (protein coding genes)
PC2 (9.98%)
PC1 (10.59%)
150
100
50
0
-50
-80
-40
0
40
Type
EGA

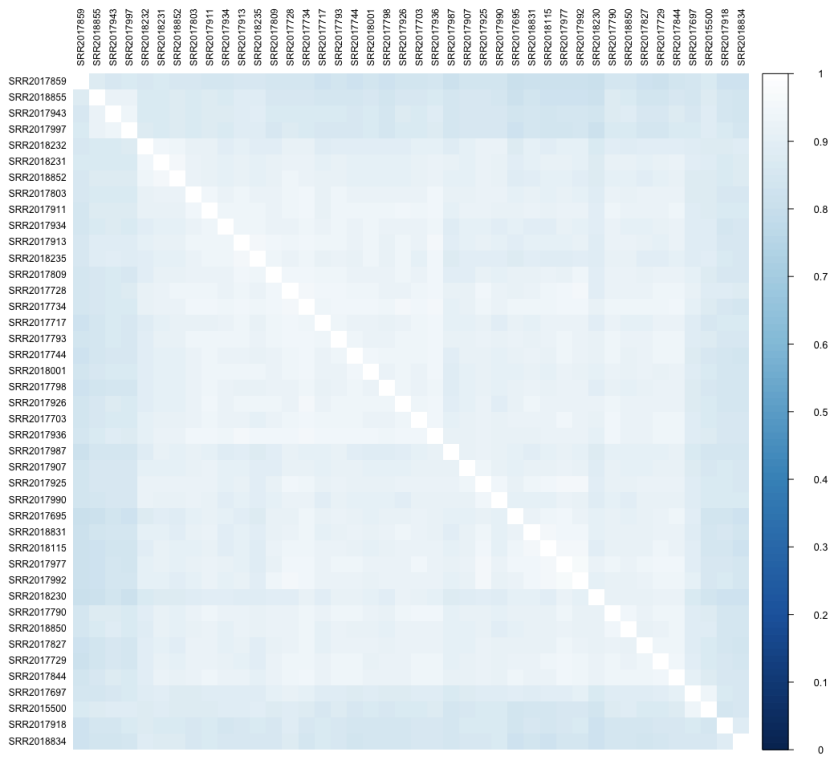
F


S00049967
S00049524
S00049268
S00049485
S00049495
S00049481
S00049565
S00049559
S00049857
S00049628
S00049907
S00049478
S00049830
S00049859
S00050868
S00049534
S00049957
S00049800
S00049854
S00049935
S00049563
S00049579
S00049755
S00049882
S00049623
S00049270
S00049576
S00049645
S00055262
S00050874
S00055258
S00049816
S00049510
S00049625
S00049506
S00049860
S00049662
S00049796
S00049272
S00049536
S00002094
S00049961
S00049824
S00049863
S00049274
S00049965
S00049630
S00049757
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

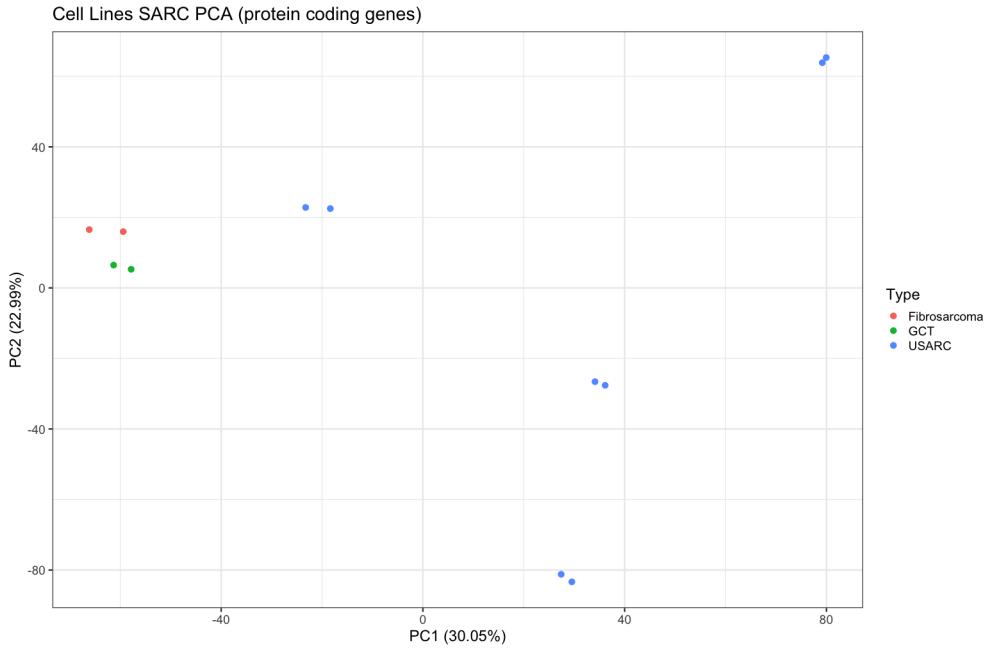
G


SRA SARC PCA (protein coding genes)
60
30
0
-30
-60
PC2 (8.01%)
-50
0
50
100
PC1 (16.36%)
Type
SRA

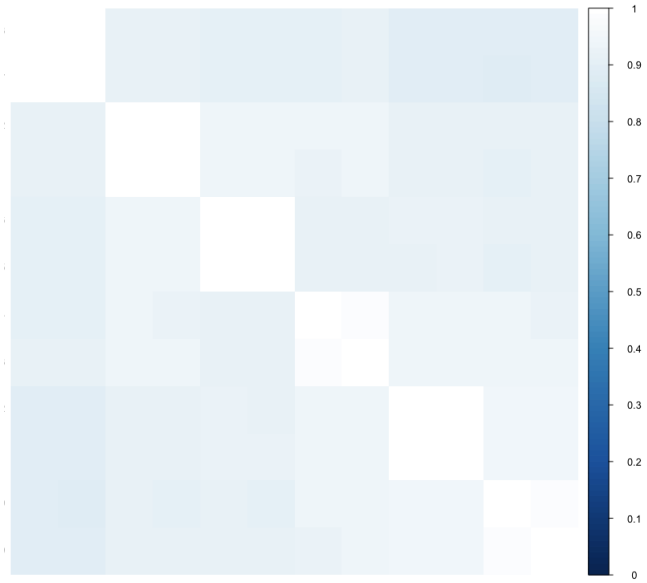
H


SRR2017859
SRR2018855
SRR2017943
SRR2017997
SRR2018232
SRR2018231
SRR2018852
SRR2017803
SRR2017911
SRR2017934
SRR2017913
SRR2018235
SRR2017809
SRR2017728
SRR2017734
SRR2017717
SRR2017793
SRR2017744
SRR2018001
SRR2017798
SRR2017926
SRR2017703
SRR2017936
SRR2017987
SRR2017907
SRR2017925
SRR2017990
SRR2017695
SRR2018831
SRR2018115
SRR2017977
SRR2017992
SRR2018230
SRR2017790
SRR2018850
SRR2017827
SRR2017729
SRR2017844
SRR2017697
SRR2015500
SRR2017918
SRR2018834
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

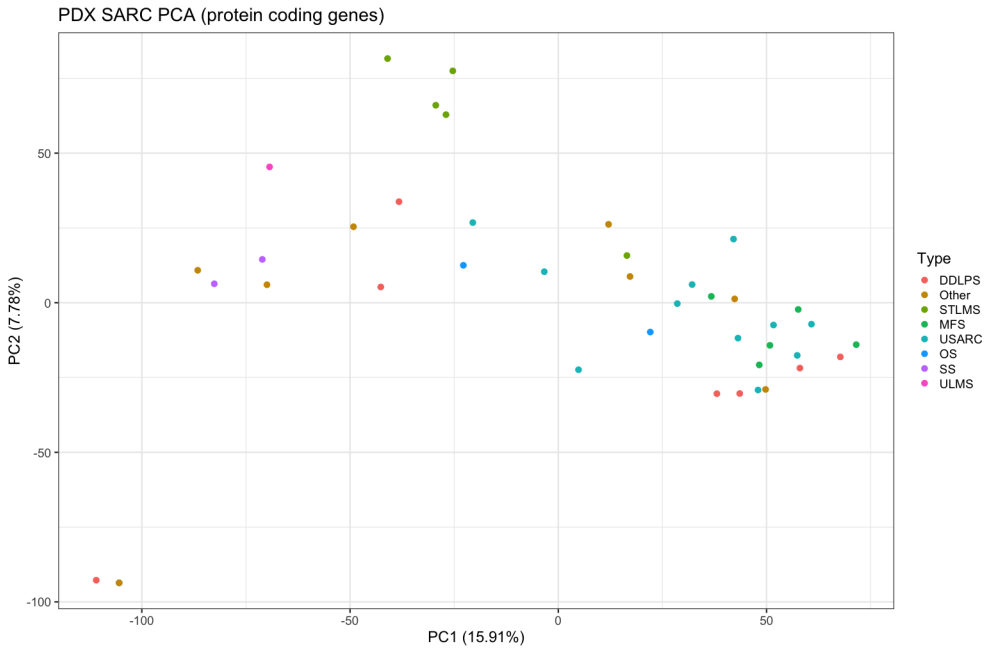
I


Cell Lines SARC PCA (protein coding genes)
PC2 (22.99%)
PC1 (30.05%)
40
0
-40
-80
-40
0
40
80
Type
Fibrosarcoma
GCT
USARC

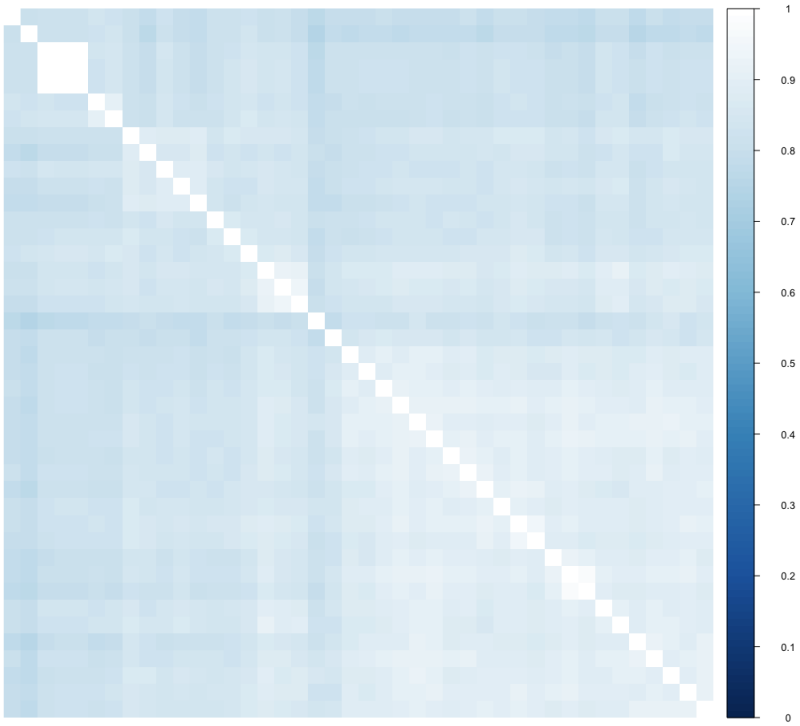
J


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

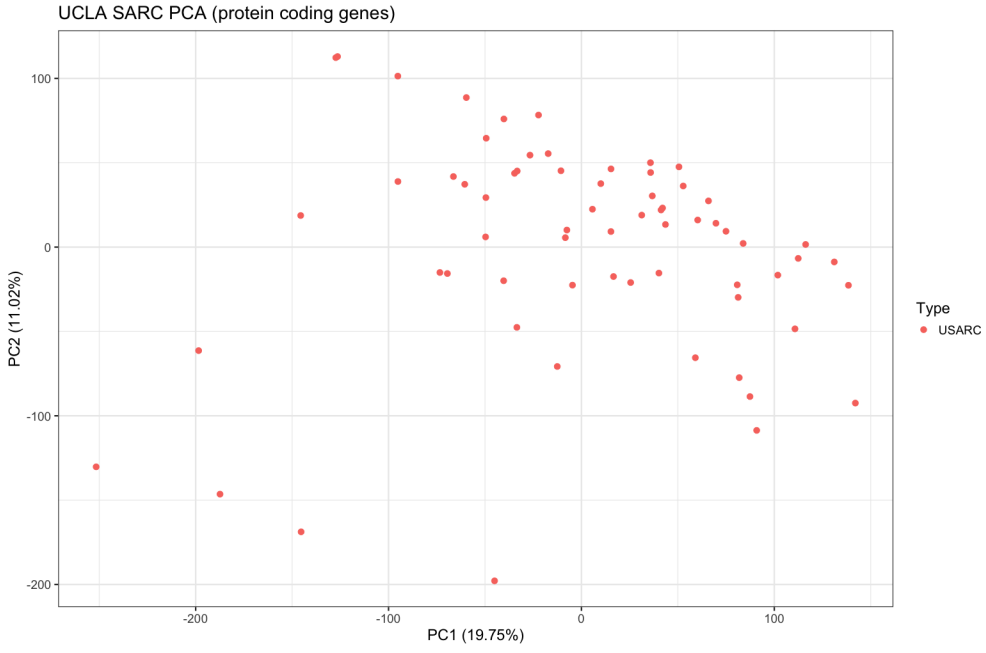
K


PDX SARC PCA (protein coding genes)
PC2 (7.78%)
PC1 (15.91%)
50
0
-50
-100
-100
-50
0
50
Type
DDLPS
Other
STLMS
MFS
USARC
OS
SS
ULMS

L


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

M


UCLA SARC PCA (protein coding genes)
100
0
-100
-200
PC2 (11.02%)
-200
-100
0
100
PC1 (19.75%)
Type
USARC

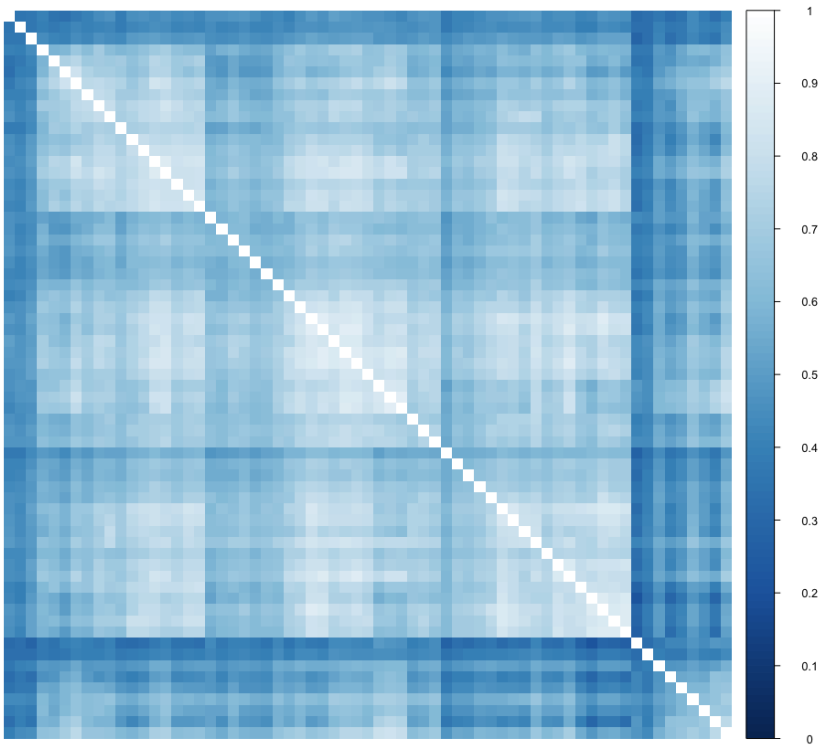
N 

Figure 4-7: Summary of outlier identification in this study. PCA and correlation plots were used to define outliers. PCAs were restricted to protein coding genes and colored by sarcoma subtype, when applicable. Graphs shown depict PC1 vs PC2. Correlation plots were restricted to protein coding genes and show pairwise Pearson correlation between samples in each dataset. (A-B) Outlier analysis of TCGA SARC RNA-Seq dataset. There was no outlier identified. (C-D) Outlier analysis of TARGET OS RNA-Seq dataset. There was one outlier identified. (E-F) Outlier analysis of EGA USARC RNA-Seq dataset. There was one outlier identified. (G-H) Outlier analysis of SRA USARC RNA-Seq dataset. There was no outlier identified. (I-J) Outlier analysis of sarcoma cell lines RNA-Seq dataset. There was no outlier identified. There are 6 cell lines, each of which were sequenced before and after treatment with BO112. The high degree of correlation between pairs of samples seen on the PCA and correlation plot represent the paired samples. (K-L) Outlier analysis of sarcoma PDX RNA-Seq dataset. There was no outlier

identified. (M-N) Outlier analysis of UCLA USARC RNA-Seq dataset. There were five outliers identified.

Dataset	Initial n	Outlier n	Final n
Cell Lines	12	0	12
EGA	48	1	47
PDX	42	0	42
SRA	42	0	42
TARGET	81	1	80
TCGA	204	0	204
UCLA	65	5	60
TOTAL	**494**	**7**	**487**

Table 4-6: Summary of outliers and final sample sizes.

4.9.6 Identifying batch effects in the data

PCAs using protein coding genes were performed between various combinations of the above datasets to explore batch effects. The PCAs below include the data from all sarcoma datasets in this analysis (Figure 4-8), all human tumor samples (Figure 4-9), and USARC tumors samples (Figure 4-10). The PCAs were colored by dataset or by histology, as noted. PC1 separates UCLA samples from the other samples. However, once graphs are comparing PC2 versus PC3 the samples cluster by histology rather than by dataset. This pattern was seen in the multiple analyses with various combinations of sample types. In Figure 4-11 PCAs were generated from publicly available RNA-Seq data from human sarcoma tumors from the TCGA, EGA, and SRA datasets. The UCLA USARC RNA-Seq samples were then projected onto these PCAs. Again, there was separation of the UCLA samples on PC1, though a closer clustering on PC2 versus PC3. The difference in the UCLA samples is also seen in the correlation plot in Figure 4-9.

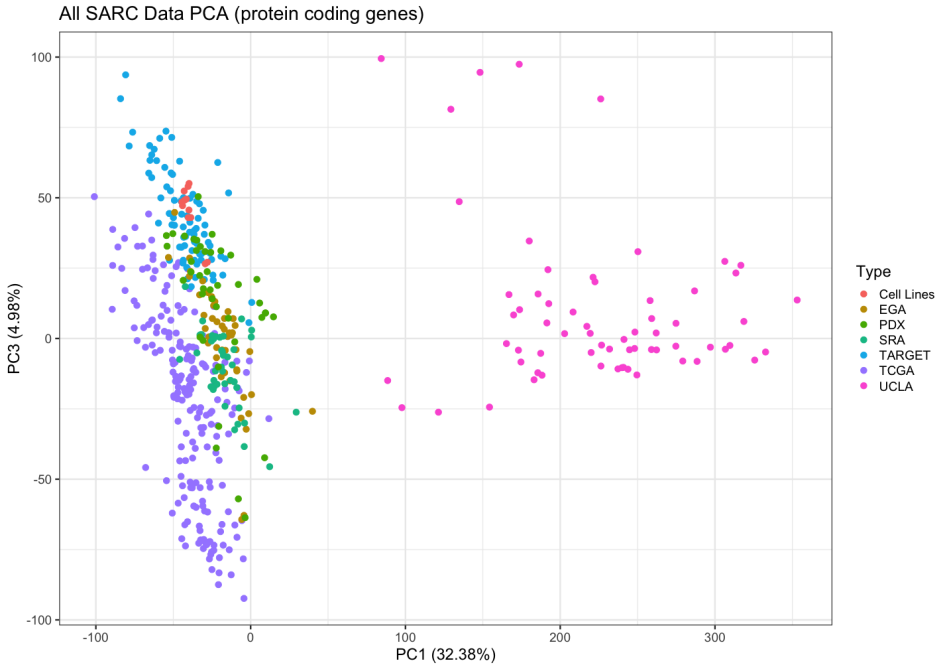
However, it is important to note that the UCLA samples also showed lower correlation between samples within the dataset than was seen between samples in each of the other datasets.

The reasons for this significant batch effect are not clear. It is possible that PC1 reflects FFPE samples and/or the degradation associated with samples over time. It is also possible that PC1 captures differences in nucleic acid extraction or sequencing techniques. In order to test this theory, the PCAs were re-created after being restricted to genes only in the library preparation that was used for the sequencing of the UCLA USARC samples. However, there were no changes seen on the PCAs (data not shown as there were no differences). Figure 4-12 explores the clustering UCLA USARC samples based on treatment and time of storage in FFPE. Neither variable showed cleared trends in the data.

We were ultimately unable to identify the reason for the batch effects seen in PC1 our data. However, it was reassuring that PC2 and PC3 did cluster according to histology as expected. Despite this, it was clear that care needed to be taken in the analyses to account for the batch effects seen. In order to account for the batch effects, multiple extensive sensitivity analyses were performed. All analysis pipelines in this study were performed on each dataset separately (primarily USARC samples from TCGA, EGA, SRA, and UCLA) and on all the USARC samples together. The results from all the analyses were compared and only results that were consistent across all datasets and all methods of analysis were considered reliable. Ultimately, despite the batch effects seen on PC1, the results of the analyses were fairly consistent across the datasets and various analysis methods.

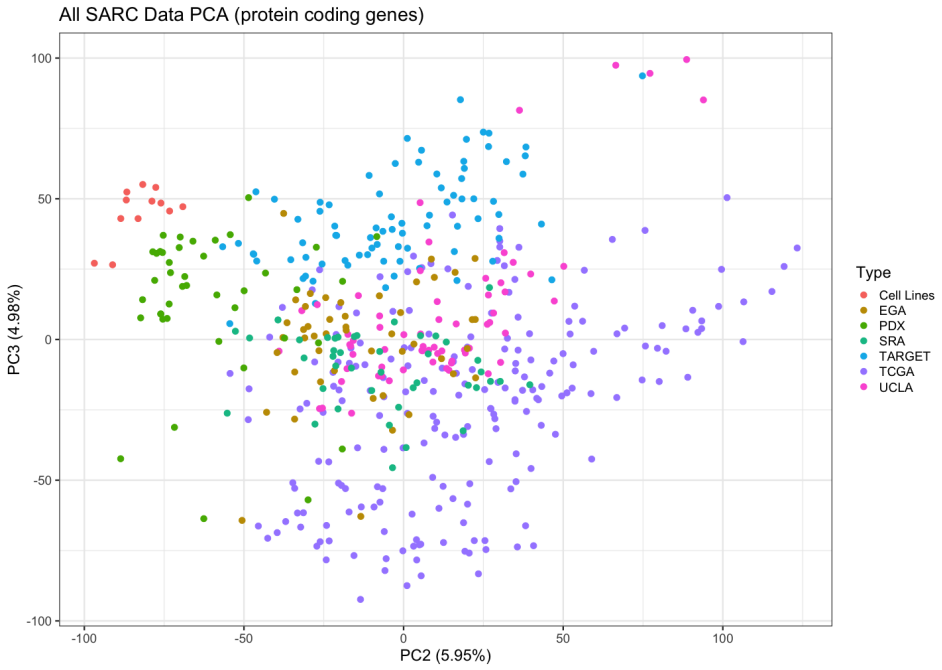

All SARC Data PCA (protein coding genes)
PC2 (5.95%)
PC1 (32.38%)
100
50
0
-50
-100
-100
0
100
200
300
Type
Cell Lines
EGA
PDX
SRA
TARGET
TCGA
UCLA

A


All SARC Data PCA (protein coding genes)
PC3 (4.98%)
PC1 (32.38%)
100
50
0
-50
-100
-100
0
100
200
300
Type
Cell Lines
EGA
PDX
SRA
TARGET
TCGA
UCLA

B

C



D

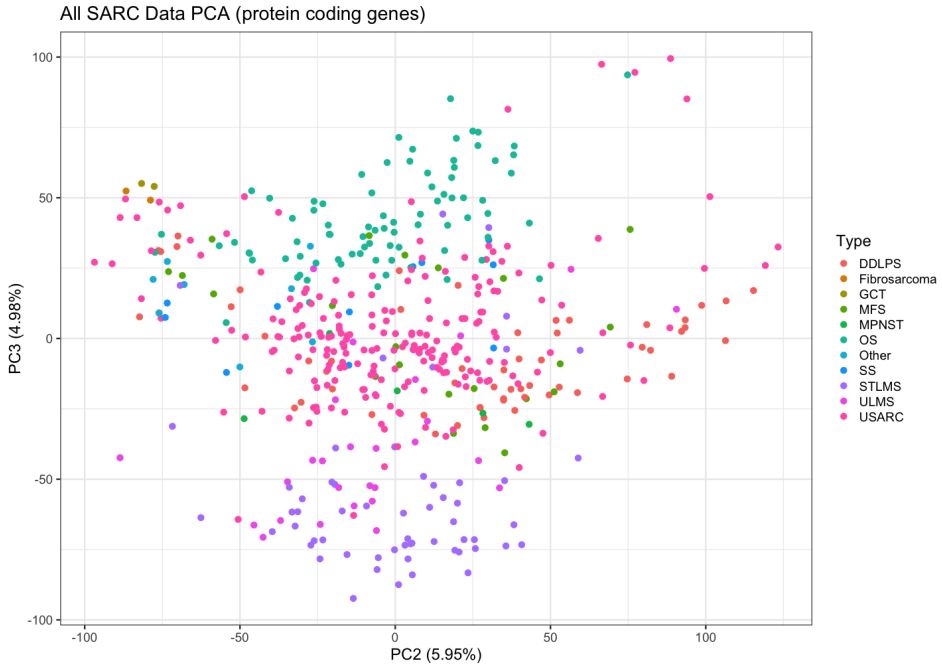
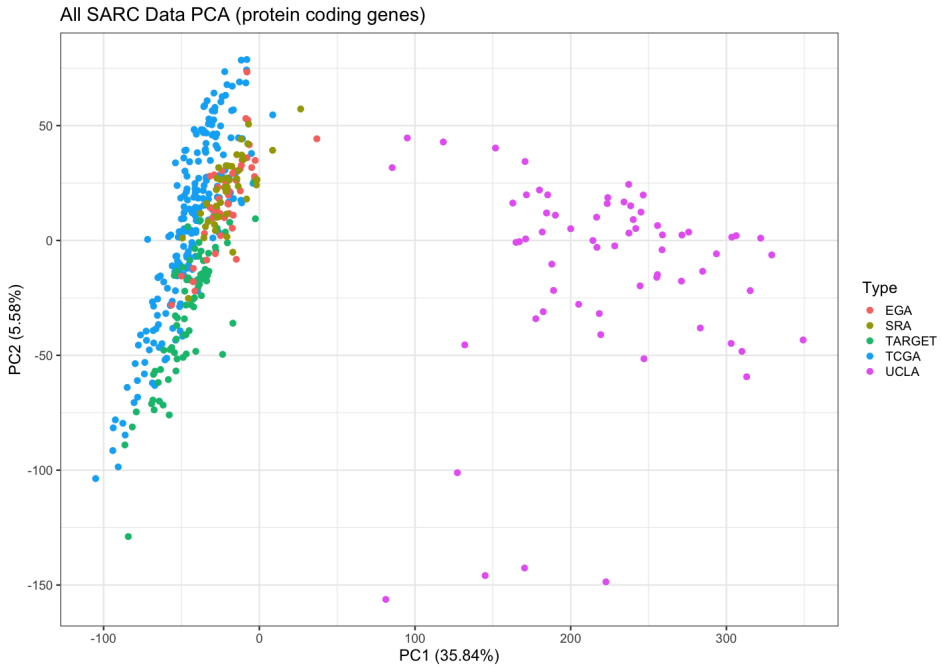
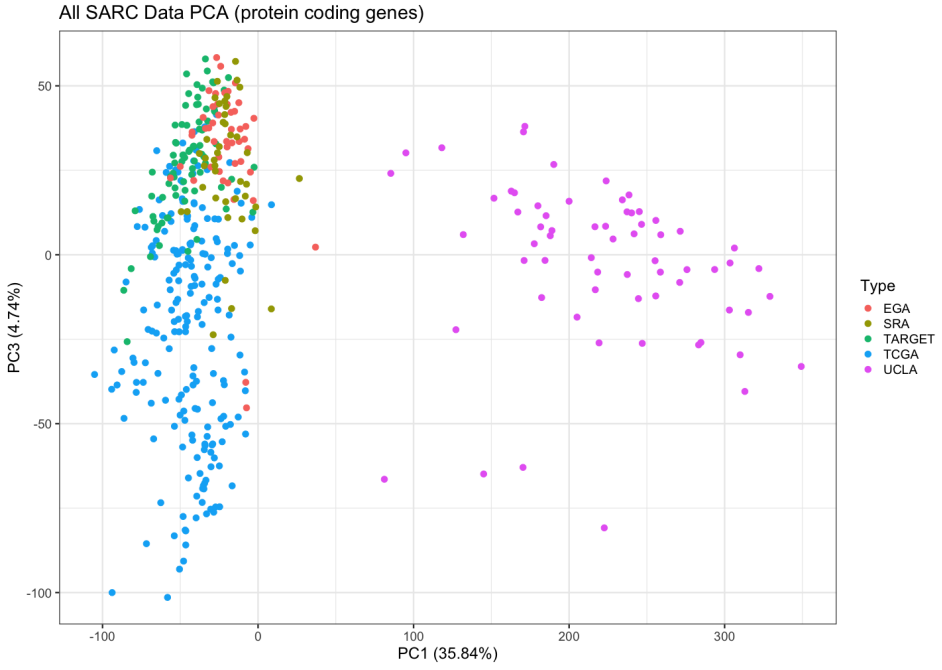


Figure 4-8: PCA analyses including all sarcoma datasets in this study (as listed above). PCAs were performed on protein coding genes only and are colored by dataset (A-C) or by histology (D), as indicated. PCs and their contribution to the variance are noted on each graph.

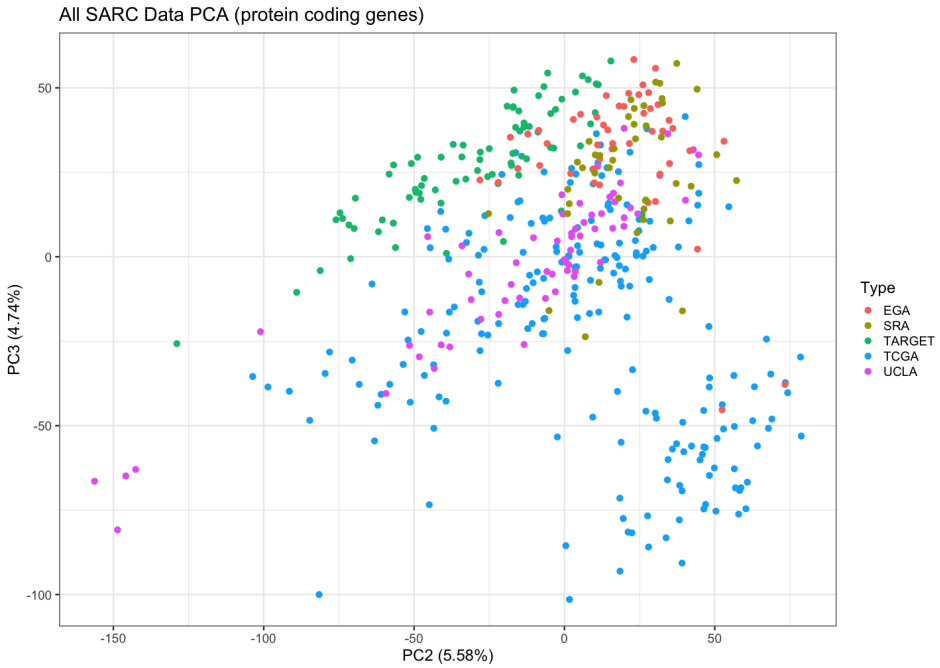

All SARC Data PCA (protein coding genes)
PC2 (5.58%)
PC1 (35.84%)
50
0
-50
-100
-150
-100
0
100
200
300
Type
EGA
SRA
TARGET
TCGA
UCLA

A

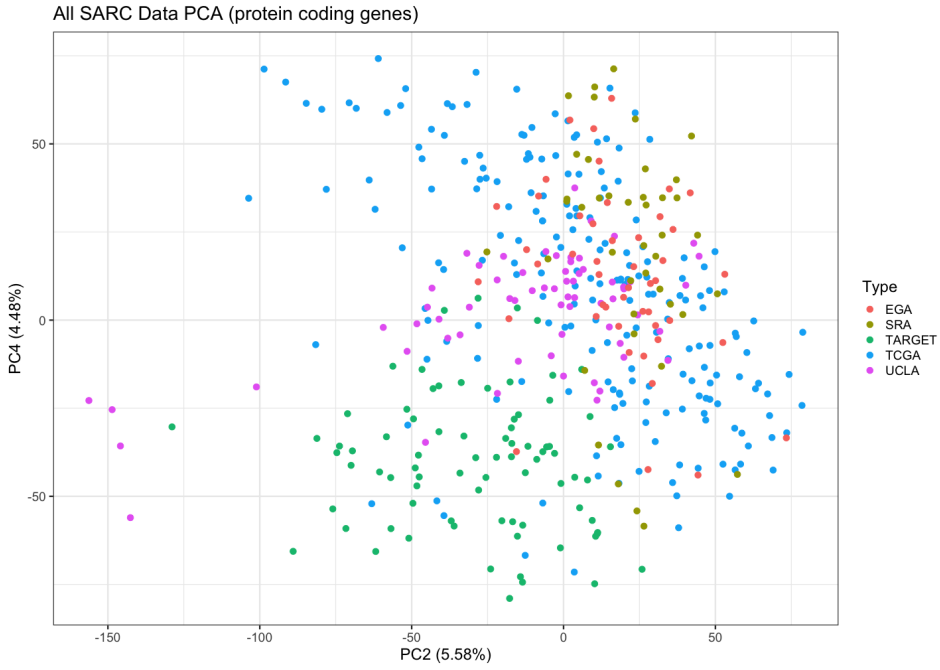

All SARC Data PCA (protein coding genes)
PC3 (4.74%)
PC1 (35.84%)
50
0
-50
-100
-100
0
100
200
300
Type
EGA
SRA
TARGET
TCGA
UCLA

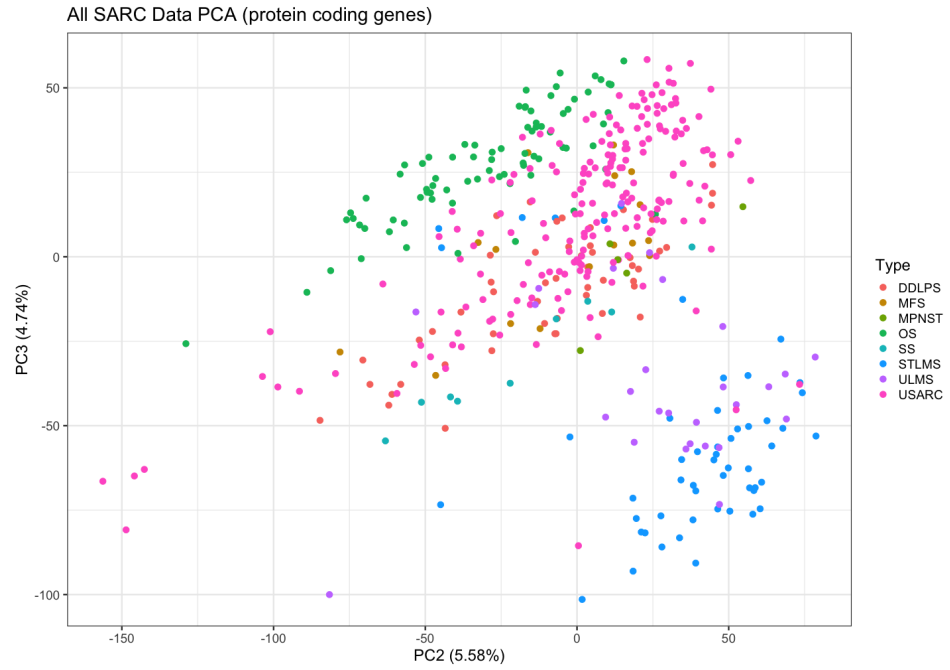
B

C

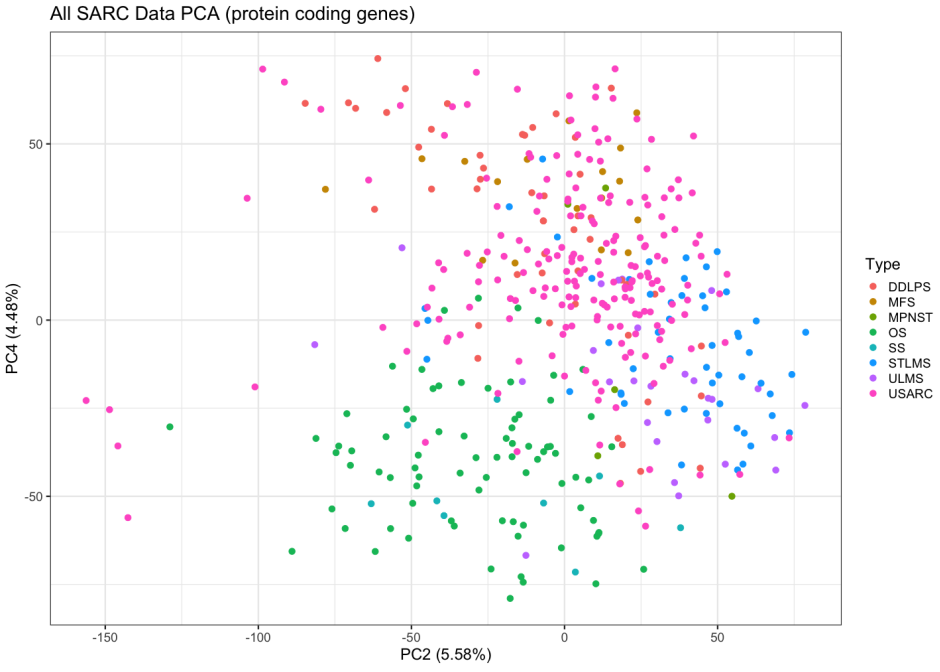

All SARC Data PCA (protein coding genes)
PC3 (4.74%)
PC2 (5.58%)
50
0
-50
-100
-150
-100
-50
0
50
Type
EGA
SRA
TARGET
TCGA
UCLA

D

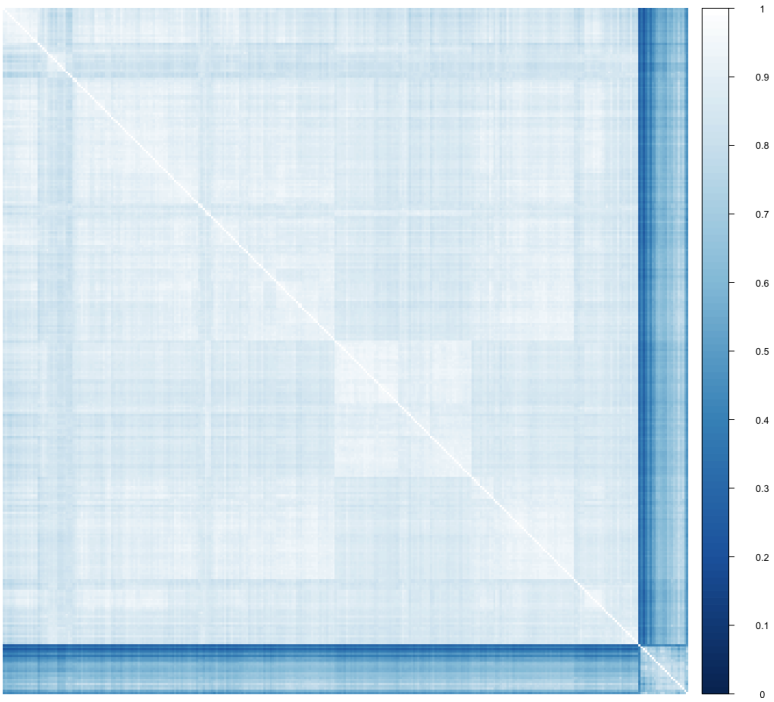

All SARC Data PCA (protein coding genes)
PC4 (4.48%)
PC2 (5.58%)
50
0
-50
-150
-100
-50
0
50
Type
EGA
SRA
TARGET
TCGA
UCLA


All SARC Data PCA (protein coding genes)
PC3 (4.74%)
PC2 (5.58%)
50
0
-50
-100
-150
-100
-50
0
50
Type
DDLPS
MFS
MPNST
OS
SS
STLMS
ULMS
USARC

E

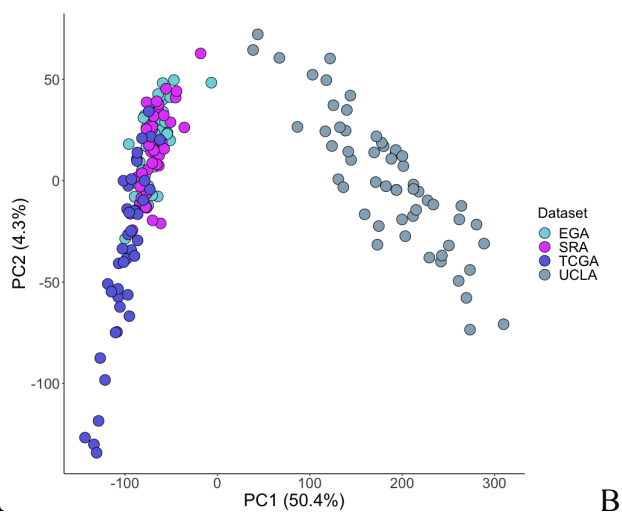

All SARC Data PCA (protein coding genes)
PC4 (4.48%)
PC2 (5.58%)
50
0
-50
-150
-100
-50
0
50
Type
DDLPS
MFS
MPNST
OS
SS
STLMS
ULMS
USARC

F

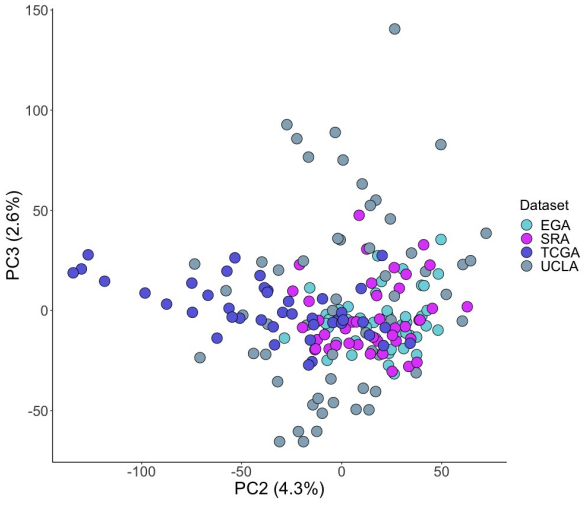


G

Figure 4-9: PCA analyses including all human sarcoma tumor datasets in this study (as listed above). PCAs were performed on protein coding genes only and are colored by dataset (A-D) or by histology (E-F), as indicated. PCs and their contribution to the variance are noted on each graph. (G) Correlation plot depicting the Pearson correlation between all human sarcoma tumor samples. Data is restricted to protein coding genes only.

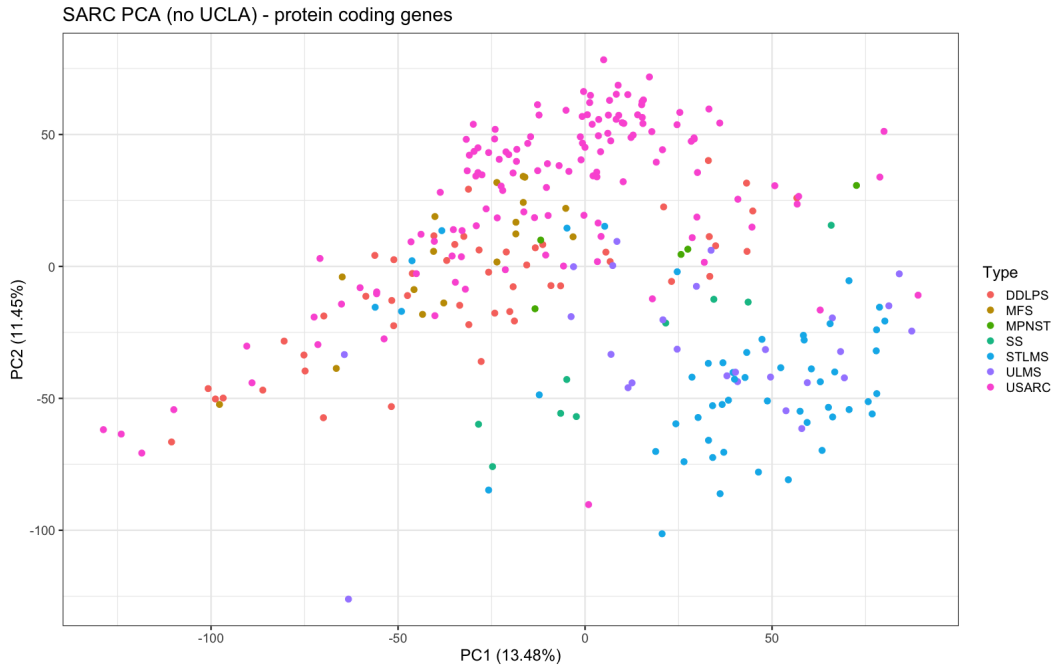


A

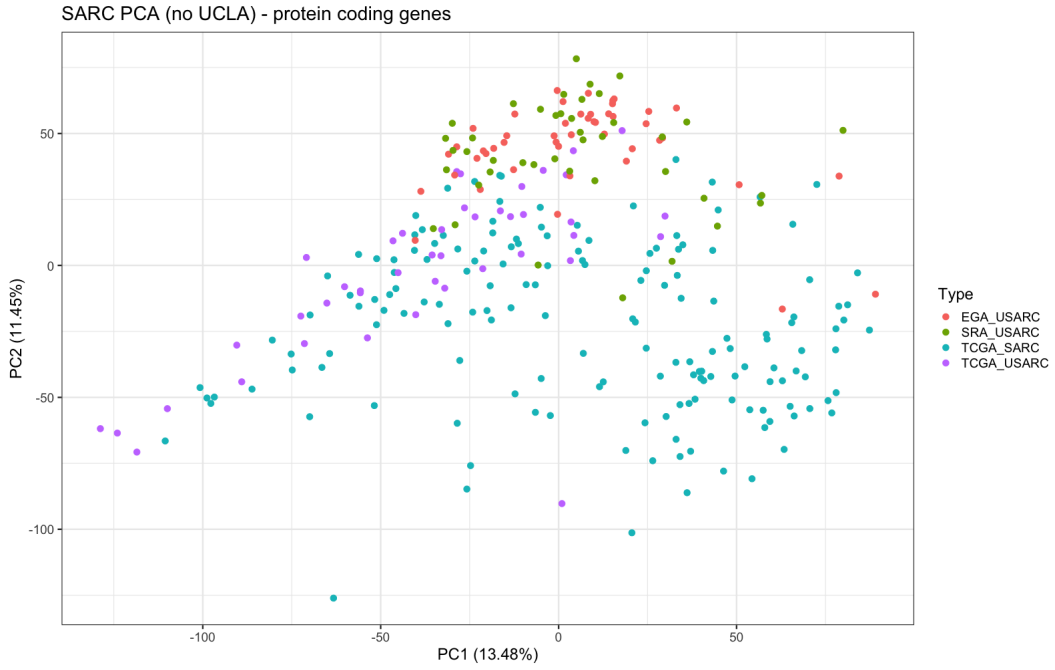


B

Figure 4-10: PCA analyses including human USARC RNA-Seq (EGA, SRA, TCGA, and UCLA). PCAs were performed on protein coding genes only and are colored by dataset. PCs and their contribution to the variance are noted on each graph.

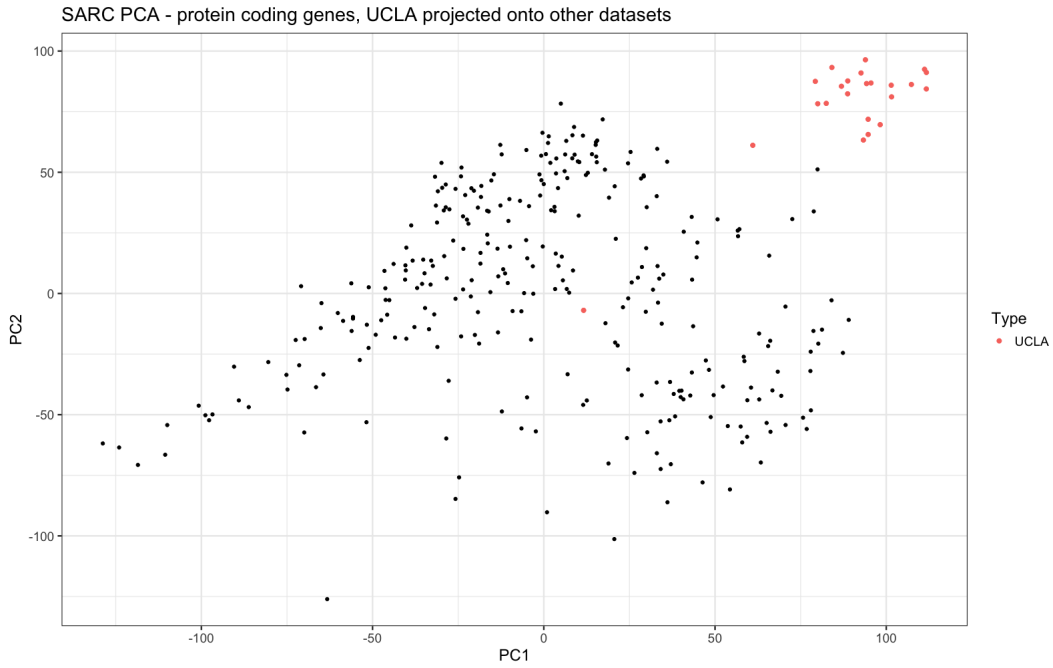


A

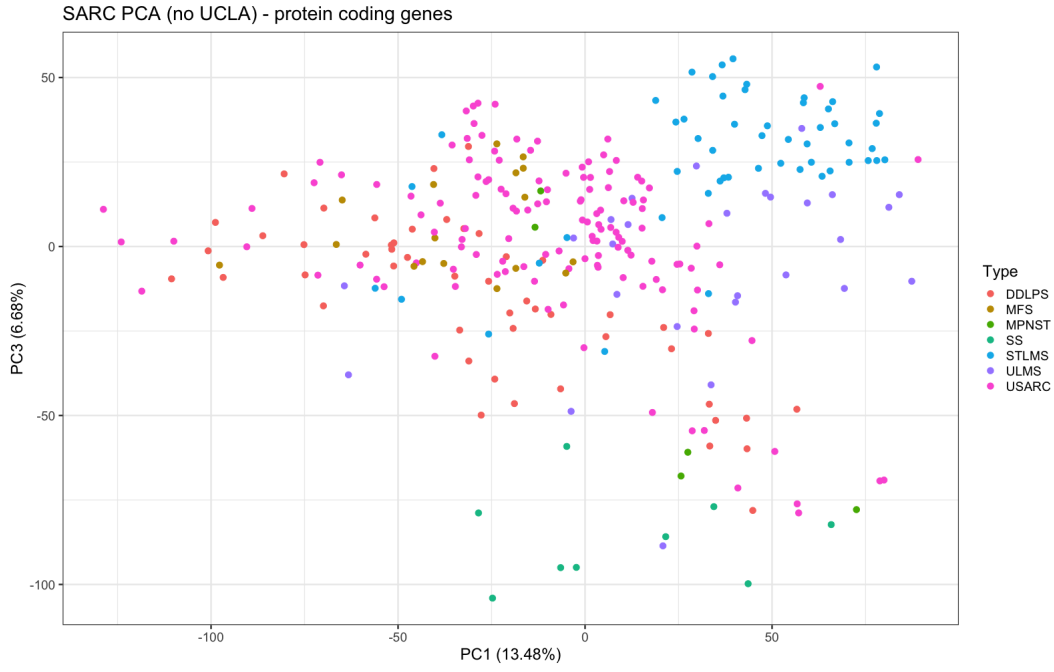


B

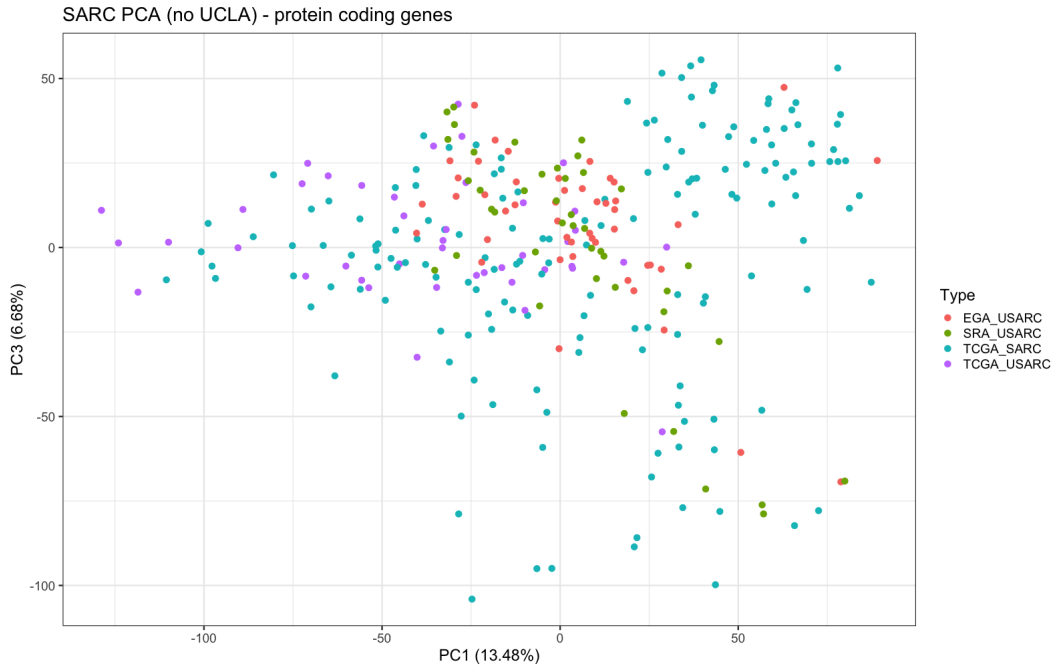
C


SARC PCA - protein coding genes, UCLA projected onto other datasets
PC2
PC1
100
50
0
-50
-100
-100
-50
0
50
100
Type
UCLA

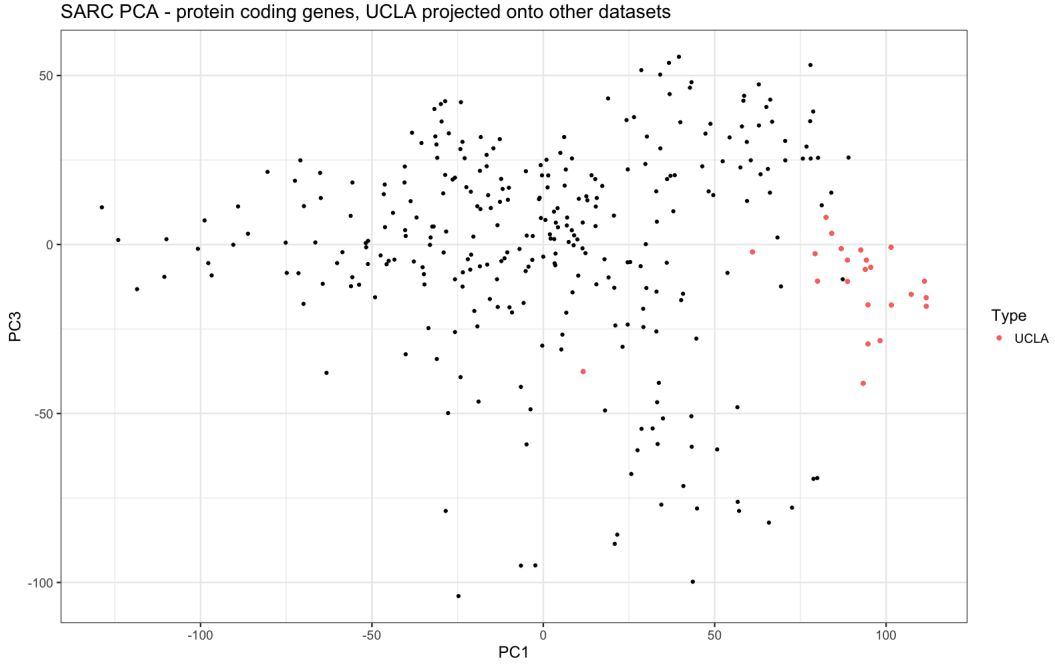
D


SARC PCA (no UCLA) - protein coding genes
PC3 (6.68%)
PC1 (13.48%)
50
0
-50
-100
-100
-50
0
50
Type
DDLPS
MFS
MPNST
SS
STLMS
ULMS
USARC

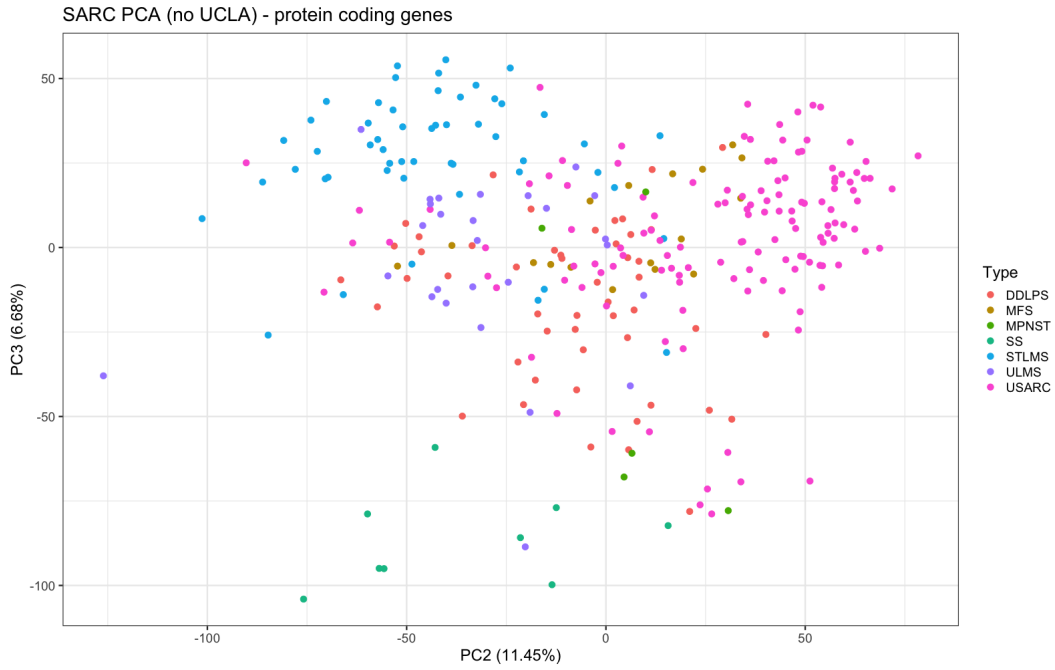
E


SARC PCA (no UCLA) - protein coding genes
PC3 (6.68%)
PC1 (13.48%)
50
0
-50
-100
-100
-50
0
50
Type
EGA_USARC
SRA_USARC
TCGA_SARC
TCGA_USARC

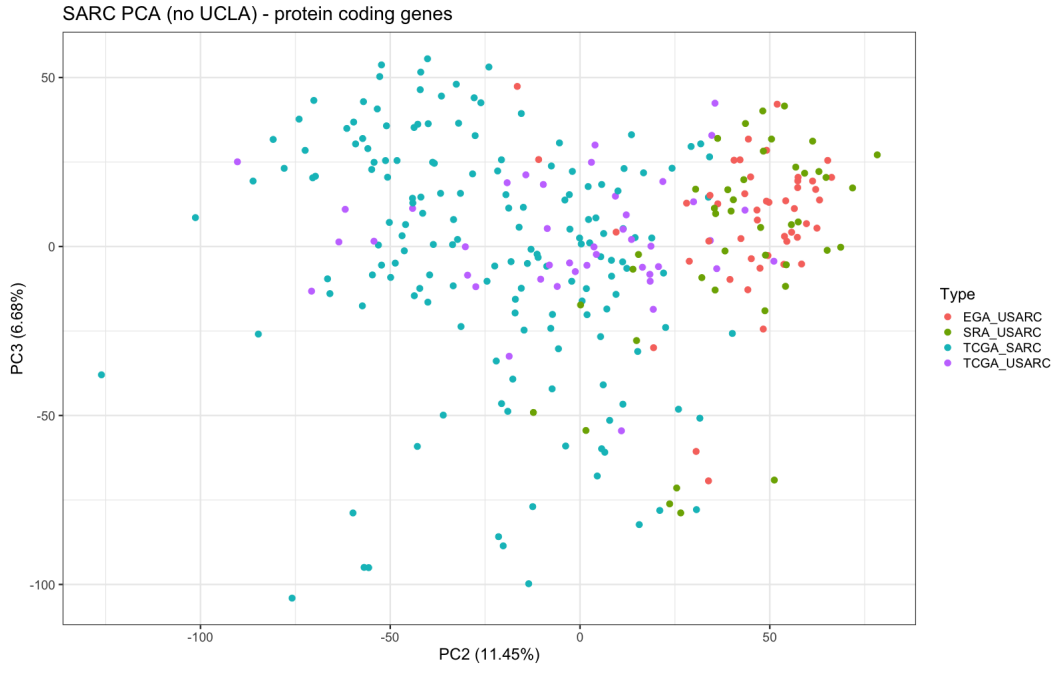
F

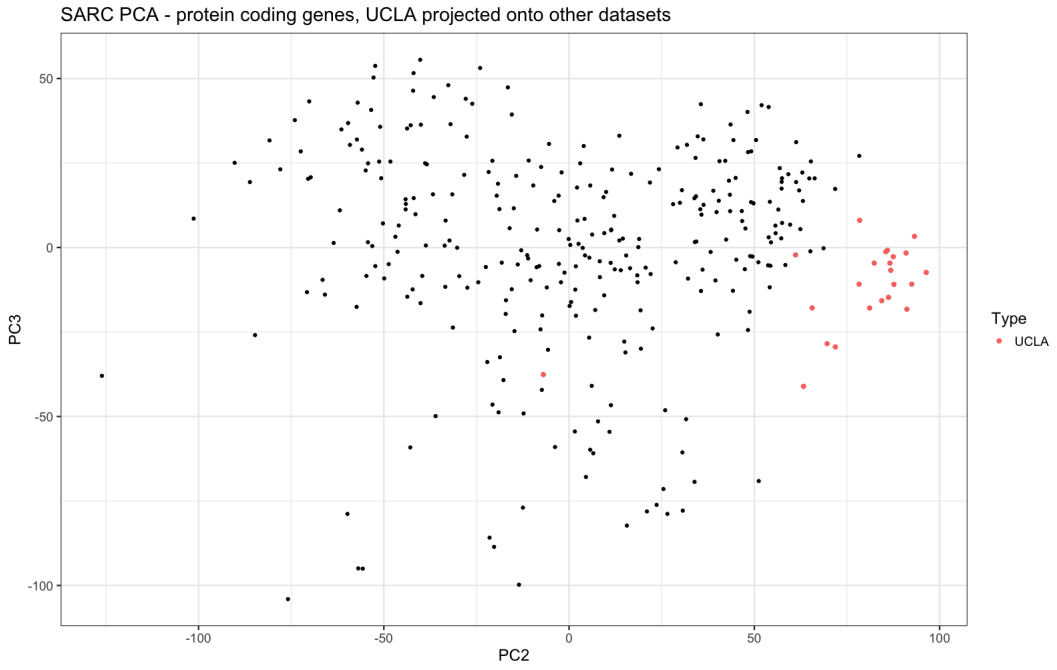

SARC PCA - protein coding genes, UCLA projected onto other datasets
PC3
PC1
50
0
-50
-100
-100
-50
0
50
100
Type
UCLA

G


SARC PCA (no UCLA) - protein coding genes
PC3 (6.68%)
PC2 (11.45%)
50
0
-50
-100
-100
-50
0
50
Type
DDLPS
MFS
MPNST
SS
STLMS
ULMS
USARC

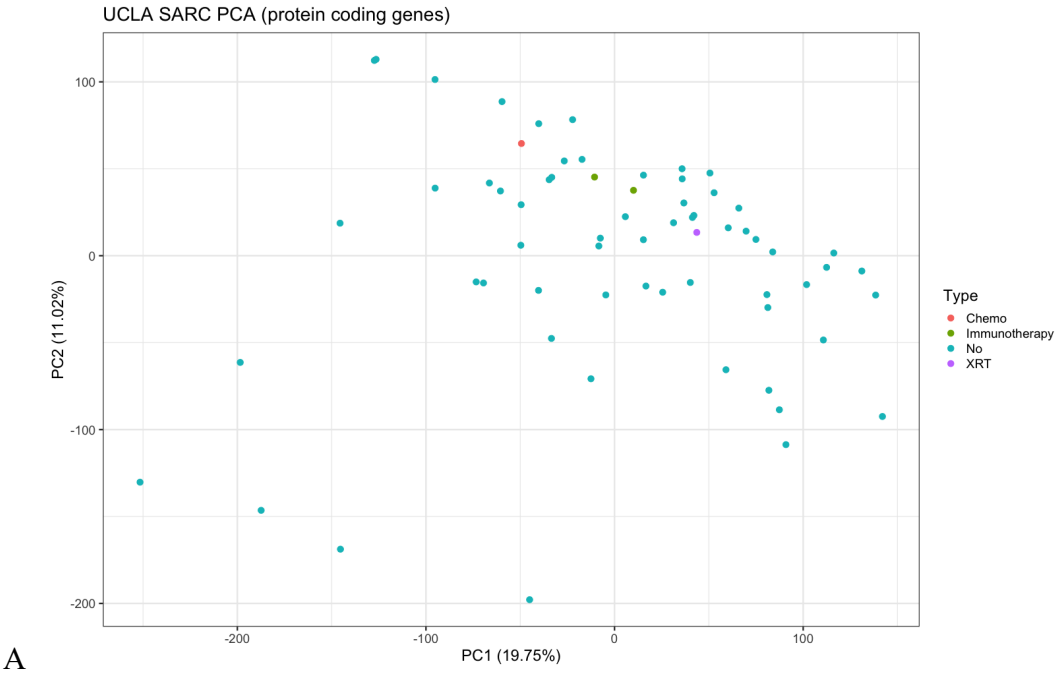
H


SARC PCA (no UCLA) - protein coding genes
PC3 (6.68%)
PC2 (11.45%)
50
0
-50
-100
-100
-50
0
50
Type
EGA_USARC
SRA_USARC
TCGA_SARC
TCGA_USARC

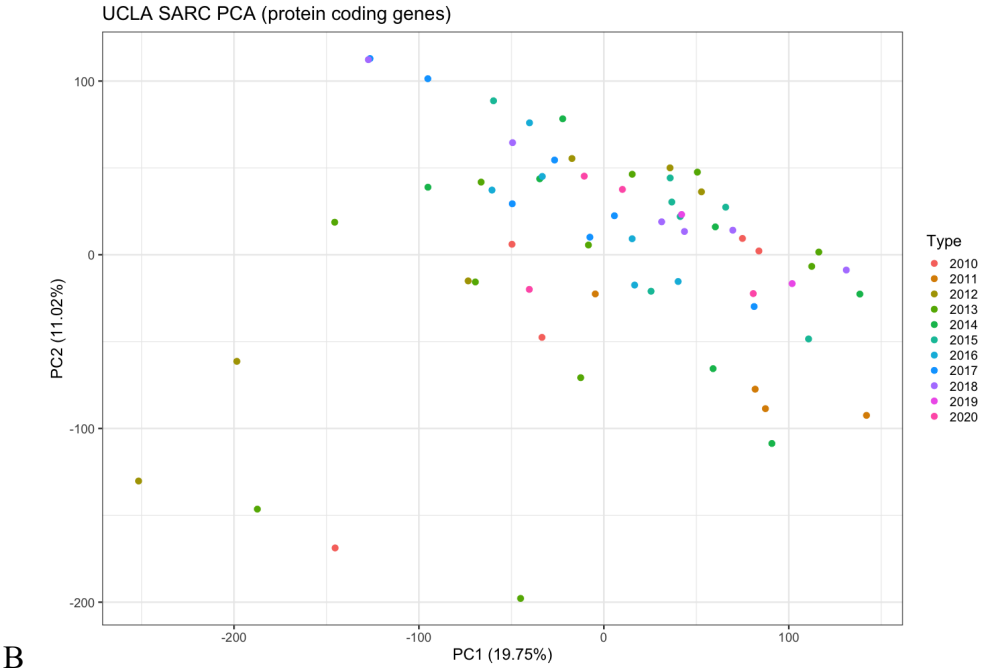


I

Figure 4-11: PCA analyses including all human sarcoma tumor datasets in this study (as listed above). Initial PCAs (A, B, D, E, G, H) included data from publicly available sarcoma datasets (EGA, SRA, TCGA). PCAs were performed on protein coding genes only and are colored by histology (A, D, G) or by dataset (B, E, H), as indicated. PCs and their contribution to the variance are noted on each graph. (C, F, I) PCAs depicting projection of UCLA USARC RNA-Seq data onto the aforementioned PCAs.



A



B

Figure 4-12: PCA analyses of UCLA USARC RNA-Seq samples. PCAs were performed on protein coding genes only and are colored (A) by treatment or (B) by year of sample resection (which corresponds to time of storage in FFPE). PCs and their contribution to the variance are noted on each graph. There were no clear trends in this unsupervised analysis.

4.9.7 Clinical analyses of publicly available datasets

As described above, the majority of the analyses in this manuscript utilized human tumor samples (USARC) from three publicly available datasets (EGA, SRA, and TCGA) and a dataset generated from patients treated in-house (UCLA). Specifically, these were the samples and datasets used to define the transcriptomic and genomic signatures associated with immune cell infiltration in USARC, which was the primary focus of this study. The corresponding patients and tumor characteristics from each of the publicly available datasets is summarized in the tables below. The patient and tumor characteristics from the UCLA samples was noted in Table 4-2 above. All UCLA tumor sites were characterized into the following categories: RP/abdomen/pelvis, trunk/extremity, or other. Additional tumor site information is summarized below. Further, the follow-up time (in years) in Table 4-2 was calculated excluding patients who died of disease or died of other causes (DOD or DOO). Median (range) was 5 (0-11) years. If all patients were included, the median (range) of follow-up was 4 (0-11) years.

Univariate analyses comparing the high and low immune cell infiltration groups were also performed. Immune cell infiltration was characterized relative to the median TIMER total immune score. Wilcoxon or chi squared tests were used as appropriate. The characteristics were similar across the groups. The only difference found was that in the TCGA cohort, size was correlated with immune group. Specifically, increased tumor size was associated with low immune cell infiltration.

Characteristic	**n**	**%**
Sex		
Female	21	44.7

Male	26	55.3

Table 4-7: Summary of the patient characteristics for the samples in the EGA dataset. This excludes outliers (n = 47).

Characteristic	**n or median**	**% or range**
Sex		
Female	20	47.6
Male	22	52.4
Age (years)	63	24-85
Tumor site		
Chest	3	7.1
Extremity	30	71.4
Head and neck	1	2.4
Trunk	8	19.0

Table 4-8: Summary of the patient and tumor characteristics for the samples in the SRA dataset. This excludes outliers (n = 42).

	All SARC		***USARC only***	
Characteristic	**n or median**	**% or range**	**n or median**	**% or range**
Total	204		43	
Histology				
DDLPS	50	24.5	NA	NA
Myxofibrosarcoma	17	8.3	NA	NA
MPNST	5	2.5	NA	NA
Synovial sarcoma	9	4.4	NA	NA
STLMS	53	26.0	NA	NA
ULMS	27	13.2	NA	NA
UPS	43	21.1	NA	NA
Age (years)	60	20-90	66	29-90
Sex				
Female	112	54.9	20	46.5
Male	92	45.1	23	53.5
History of other malignancy				
No	177	86.8	33	76.7
Yes	27	13.2	10	23.3
Tumor Site				
Chest	4	2.0	0	0.0
Extremity	60	29.4	28	65.1
Gynecological	23	11.3	0	0.0

Head and Neck	3	1.5	0	0.0
RP/Abd/Pelvis	97	47.5	9	20.9
Trunk	16	7.8	6	14.0
Tumor size (cm)	10.1	1.2-39.5	10.0	2.7-36.5
Depth				
Deep	173	84.8	33	76.7
Superficial	8	3.9	5	11.6
Residual tumor				
R0	126	61.8	26	60.5
R1	54	26.5	13	30.2
R2	8	3.9	2	4.7
RX	16	7.8	2	4.7
Adjuvant radiation				
No	139	68.1	17	39.5
Yes	57	27.9	25	58.1
Adjuvant systemic treatment				
No	150	73.5	34	79.1
Yes	46	22.5	8	18.6
Adjuvant chemo+XRT				
No	187	91.7	36	83.7
Yes	17	8.3	7	16.3
Tumor status (at follow-up)				
Tumor free	95	46.6	24	55.8
With tumor	96	47.1	17	39.5
OS status				
Alive	127	62.3	29	67.4
Dead	77	37.7	14	32.6
OS time (months)	32.0	0.5-170.6	22.5	0.5-106.2
RFS status				
No relapse	83	40.7	21	48.8
Relapse	121	59.3	22	51.2
RFS time (months)	17.5	0.2-151.7	13.3	0.2-106.2
Treatment outcome at final TCGA follow-up				
Complete response	69	33.8	20	46.5
Partial response	1	0.5	0	0.0
Progressive disease	44	21.6	4	9.3
Stable disease	6	2.9	5	11.6
Local recurrence				
Yes	46	22.5	7	16.3
No or NA	158	77.5	36	83.7
Time to local recurrence (months)	11.0	0.9-66.8	15.7	0.9-22.4
Distant recurrence				
Distant metastasis	71	34.8	15	34.9
New primary	6	2.9	1	2.3
Distant recurrence site				
Liver	7	3.4	0	0.0

Lung	50	24.5	13	30.2
RP/Abd/Pelvis	7	3.4	0	0.0
Other	11	5.4	3	7.0
Time to distant recurrence (months)	10.1	0.2-87.2	5.0	0.2-16.1

Table 4-9: Summary of the patient and tumor characteristics for the samples in the TCGA dataset. TCGA SARC and TCGA USARC are noted separately above. This excludes outliers (SARC n = 204; USARC n = 43).

Tumor Site	**Tumor Site Recode**
Abdomen	RP/Abdomen/Pelvis
Back	Trunk/extremity
Breast	Other
Chest wall	Extremity/Trunk
Extremity	Extremity/Trunk
Gluteus	Extremity/Trunk
Groin	Extremity/Trunk
Head	Other
Pelvis	RP/Abdomen/Pelvis
RP	RP/Abdomen/Pelvis
Trunk	Extremity/Trunk

Table 4-10: All UCLA tumor sites were characterized as the following categories: RP/abdomen/pelvis, trunk/extremity, or other. Additional information regarding the tumor sites and how they were recoded is summarized here.

	Low		**High**		**p**
	n	*%*	*n*	*%*	
Total	24	51.1	23	48.9	
Sex					0.671
Female	10	47.6	11	52.4	
Male	14	53.8	12	46.2	

Table 4-11: Univariate analysis patient and tumor characteristics from EGA USARC cohort. Characteristics are compared between tumors with high versus low immune cell infiltration, defined relative to the median TIMER total immune score. Chi square test or Wilcoxon test were

used as appropriate. Follow up time listed is the years from the date of surgery to the date of last follow-up and excludes patients who died.

	Low		High		p
	n or median	*% or range*	*n or median*	*% or range*	
Total	21	50.0	21	50.0	
Sex					1.000
Female	10	50.0	10	50.0	
Male	11	50.0	11	50.0	
Age (years)	63	24-85	63	36-83	0.980
Tumor site					0.210
Chest	2	66.7	1	33.3	
Extremity	12	40.0	18	60.0	
Head and neck	1	100.0	0	0.0	
Trunk	6	75.0	2	25.0	

Table 4-12: Univariate analysis patient and tumor characteristics from SRA USARC cohort. Characteristics are compared between tumors with high versus low immune cell infiltration, defined relative to the median TIMER total immune score. Chi square test or Wilcoxon test were used as appropriate. Follow up time listed is the years from the date of surgery to the date of last follow-up and excludes patients who died.

	Low		High		p
	n or median	*% or range*	*n or median*	*% or range*	
Total	22	51.2	21	48.8	
Age (years)	65	29-88	67	43-90	0.618
Sex					0.280
Female	12	60.0	8	40.0	
Male	10	43.5	13	56.5	
History of other malignancy					0.126
No	19	57.6	14	42.4	
Yes	3	30.0	7	70.0	
Tumor Site					0.638
Extremity	15	53.6	13	46.4	

RP/Abd/Pelvis	5	55.6	4	44.4	
Trunk	2	33.3	4	66.7	
Tumor size (cm)	12.2	7.0-36.5	7.5	2.7-27.5	0.003
Depth					0.816
Deep	17	51.5	16	48.9	
Superficial	2	40.0	3	60.0	
Unknown	3	60.0	2	40.0	
Residual tumor					0.561
R0	13	50.0	13	50.0	
R1	6	46.2	7	53.8	
R2	2	100.0	0	0.0	
RX	1	50.0	1	50.0	
Adjuvant radiation					0.569
No	8	47.1	9	52.9	
Yes	14	56.0	11	44.0	
Unknown	0	0.0	1	100.0	
Adjuvant systemic treatment					0.881
No	18	52.9	16	47.1	
Yes	4	50.0	4	50.0	
Unknown	0	0.0	1	100.0	
Adjuvant chemo+XRT					0.729
No	18	50.0	18	50.0	
Yes	4	57.1	3	42.9	
Tumor status (at follow-up)					0.853
Tumor free	12	50.0	12	50.0	
With tumor	9	52.9	8	47.1	
Unknown	1	50.0	1	50.0	
OS status					0.232
Alive	13	44.8	16	55.2	
Dead	9	64.3	5	35.7	
OS time (months)	20	1-106	23	3-102	0.585
RFS status					0.287
No relapse	9	42.9	12	57.1	
Relapse	13	59.1	9	40.9	
RFS time (months)	11	0-106	13	0-102	0.618
Treatment outcome at final TCGA f/u					0.921
Complete response	10	50.0	10	50.0	
Progressive disease	2	50.0	2	50.0	
Stable disease	2	40.0	3	60.0	
Unknown	8	57.1	6	42.9	
Local recurrence					0.257

Yes	5	71.4	2	28.6	
No	17	47.2	19	52.8	
Time to local recurrence (months)	16	1-22	13	10-15	0.847
Distant recurrence					0.242
Distant metastasis	9	60.0	6	40.0	
New primary	0	0.0	1	100.0	
Distant recurrence site					0.321
Bone	0	0.0	1	100.0	
Groin	1	100.0	0	0.0	
Inguinal LN	0	0.0	1	100.0	
Lung	8	61.5	5	38.5	
Time to distant recurrence (months)	5	1-16	4	0-13	0.916

Table 4-13: Univariate analysis patient and tumor characteristics from TCGA USARC cohort. Characteristics are compared between tumors with high versus low immune cell infiltration, defined relative to the median TIMER total immune score. Chi square test or Wilcoxon test were used as appropriate. Follow up time listed is the years from the date of surgery to the date of last follow-up and excludes patients who died.

4.9.8 Summary of USARC subtypes in the study cohort

Subtypes of USARC include spindled, epithelioid, pleomorphic, and ovoid[13]. An expert sarcoma pathologist re-reviewed all samples included in this study and noted the USARC subtype of each sample. A summary of the USARC subtypes is included below.

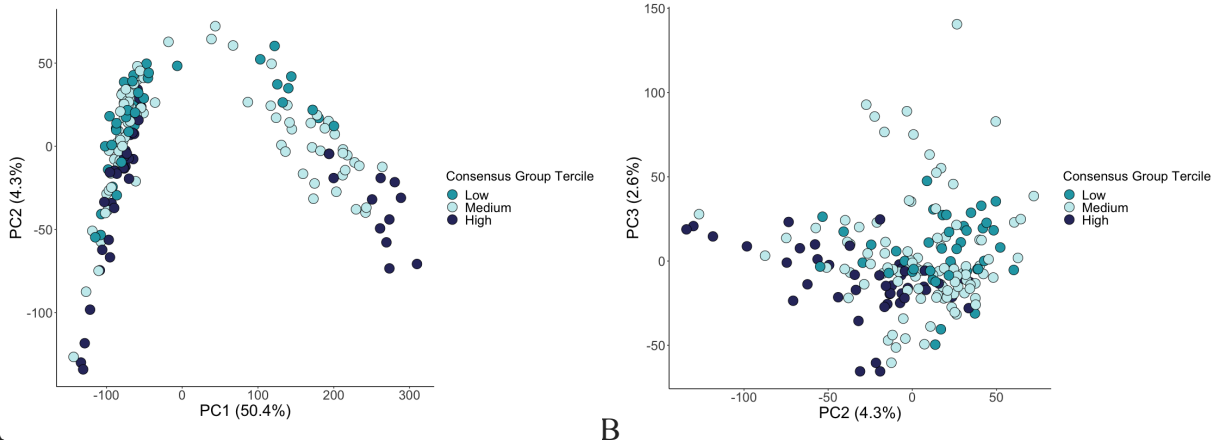
USARC Subtype	n
Spindled	9
Spindled/ovoid	2
Spindled/pleomorphic	24
Spindled/epithelioid	1
Ovoid	6

Epithelioid	2
Epithelioid/ovoid	1
Pleomorphic	3
Pleomorphic/ovoid	3
Epithelioid/ovoid/pleomorphic	2
Epithelioid/spindled/pleomorphic	2
Spindled/ovoid/pleomorphic	2
Spindled/pleomorphic/histiocytoid	1
Epithelioid/spindled/pleomorphic+chondroid	2

Table 4-14: Summary of USARC subtypes included in this study.

4.9.9 Unsupervised analysis of clustering of samples based on immune cell infiltration

PCAs were performed to explore the unsupervised clustering of USARC samples based on immune cell infiltration. The PCAs below are colored by immune cell infiltration category. Samples were characterized as high, medium or low (based on tercile) or high or low (based on median) immune cell infiltration according to Consensus Group (as defined in Chapter 3) or TIMER immune sum scores.



A B

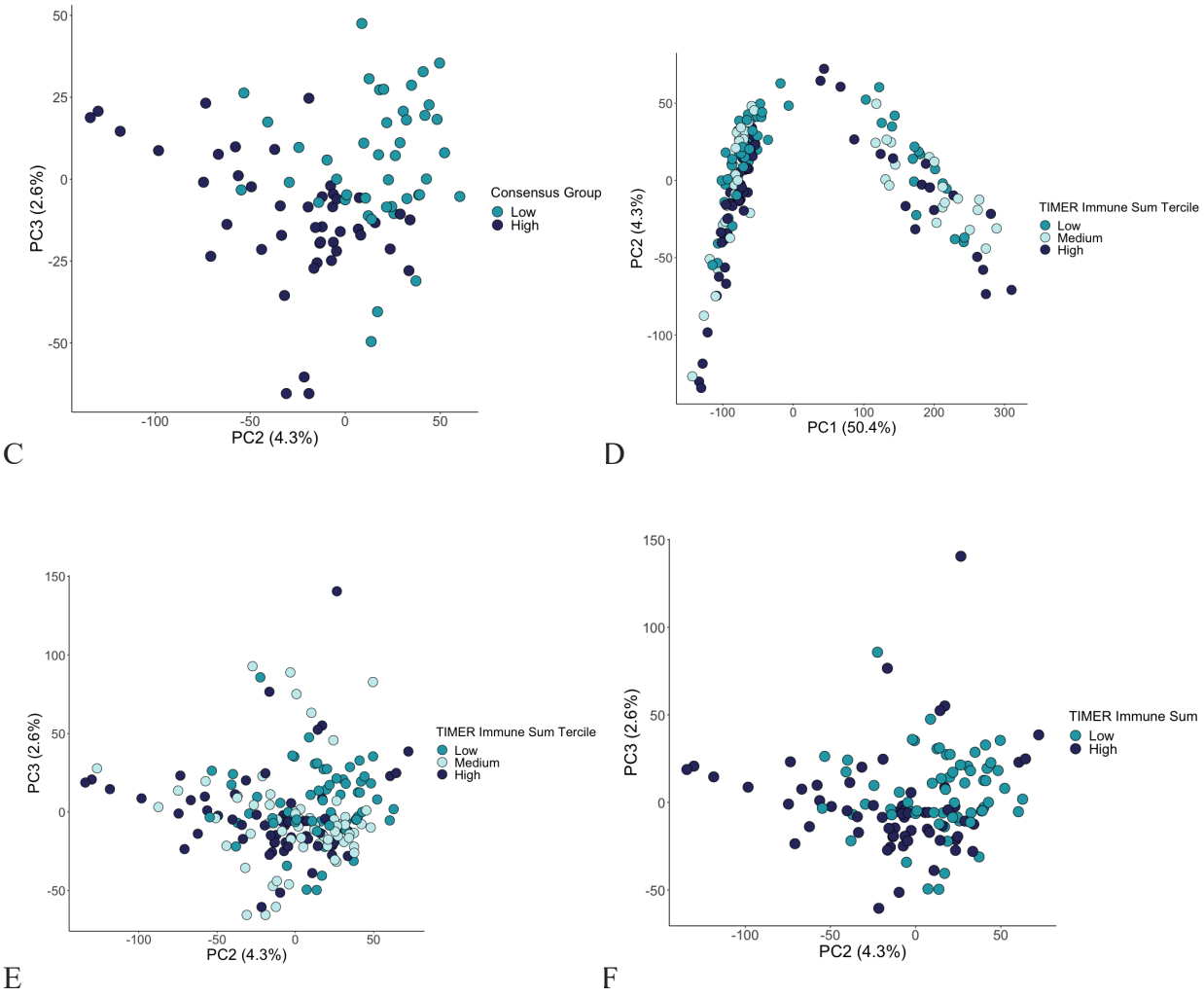


Figure 4-13: PCA analyses of human USARC RNA-Seq samples. PCAs were performed on protein coding genes only and are colored (A-C) by Consensus Group (as defined in Chapter 3) or (D-F) by TIMER immune sum scores. There was a general PCs and their contribution to the variance are noted on each graph. There were no clear trends in this unsupervised analysis.

4.9.10 Analysis of cell lines elucidates genes likely to be tumor-derived

RNA sequencing was performed on four undifferentiated sarcoma cell lines. Genes were restricted to protein coding genes only. Genes that had a raw count of greater than 10 in all four

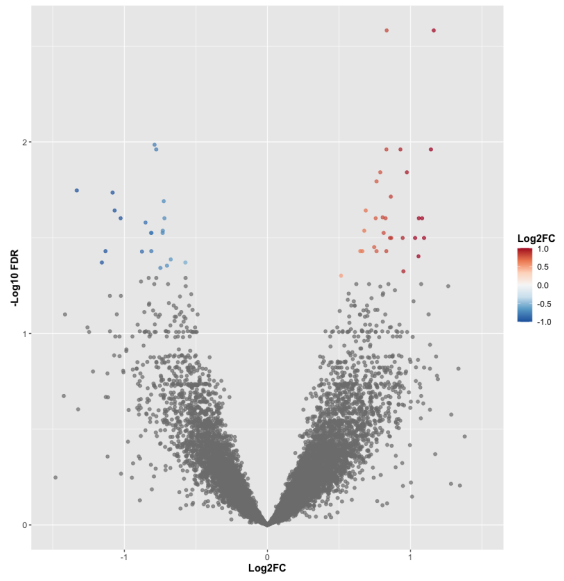
cell lines were included on the final list. A total of 9,948 genes meeting these criteria were identified (Table 4-15). These genes likely represent those that are tumor-derived, rather than derived from immune cells, as they are consistently present in cell lines, which should be devoid of immune cells. This gene list was used for additional sensitivity analyses in this study. For example, the PCAs including sarcoma genes and the GSEA analyses comparing tumor samples with high versus low immune cell infiltration were performed on all protein coding genes as well as on only the cell lines genes. There were no significant differences in the results (data not shown).

Row sum	**N of genes**	**Min (row sum)**	**1st quartile**	**Median**	**Mean**	**3rd quartile**	**Max**
Without any genes removed	60,448	0	0	1	2,130	99	1,327,961
Remove genes with row sum 0	31,366	1	5	76	4,106	2,741	1,327,961
Protein coding only + row sum > 0	16,327	1	109	2,037	7,393	6,943	1,327,961
Protein coding + row sum >10	14,352	10	463	2,920	8,410	7,921	1,327,961
>10 in 2 or more samples	12,776	NA	NA	NA	NA	NA	NA
>10 in all samples	9,948	NA	NA	NA	NA	NA	NA

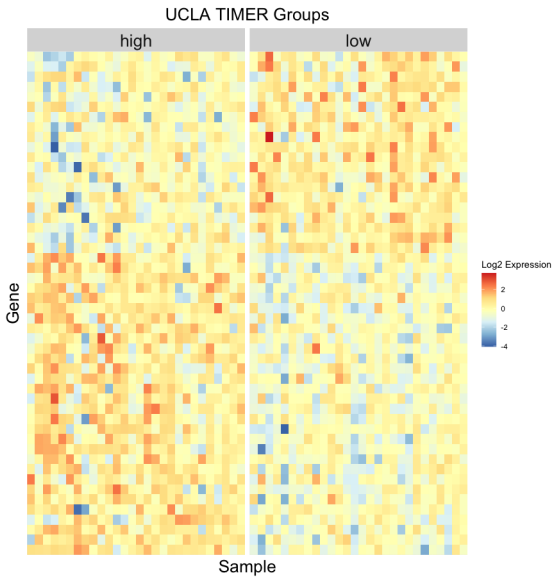
Table 4-15: Summary of analysis of genes in RNA-Seq data from four USARC cell lines.

4.9.11 Additional transcriptomic analysis comparing USARC samples with high versus low immune cell infiltration

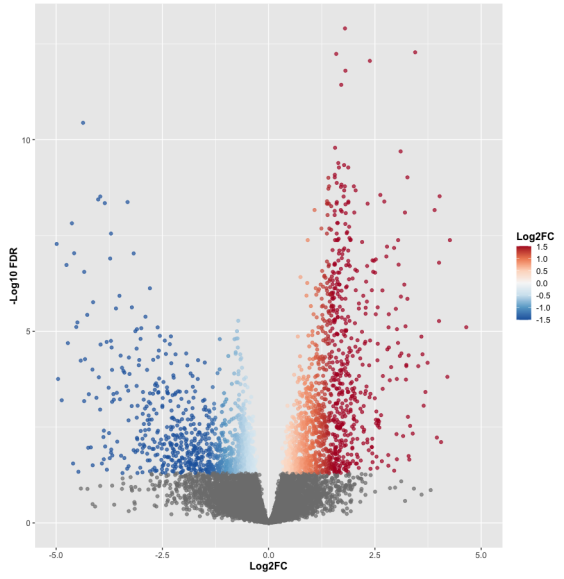
The transcriptomic analysis pipeline is summarized in Figure 4-1 and the results of this analysis are summarized below. All analyses were performed on each dataset (EGA, SRA, TCGA, and UCLA) separately as well as all datasets together and compare samples with high versus low immune cell infiltration. Volcano plots demonstrate DESeq analysis (Figure 4-14 A, C, E, G, I). Heatmaps summarizing the DESeq results demonstrate transcriptomic signatures associated with high versus low immune cell infiltration (Figure 4-14 B, D, F, H, J). GSEA using Hallmark gene sets was performed on the results of the DESeq analyses. Samples with high immune cell infiltration demonstrated an association with gene sets relating to immune function. These results suggest the characterization of samples with high versus low immune cell infiltration based on TIMER in-silico immune deconvolution were correct. Samples with low immune cell infiltration demonstrated an association with gene sets relating to cell division and growth, such as mitotic spindle and G2M checkpoint. These results were similar whether all protein coding genes were included or only genes highly expressed in cell line genes were included, as in the sensitivity analysis (Table 4-15, Figure 4-15). Finally, the transcriptomic signatures associated with high or low immune cell infiltration in each dataset were validated in the other datasets. Sum z scores of the top or bottom 50 genes were calculated to determine signature scores (Figure 4-16).


-Log10 FDR
Log2FC
Log2FC
1.0
0.5
0.0
-0.5
-1.0

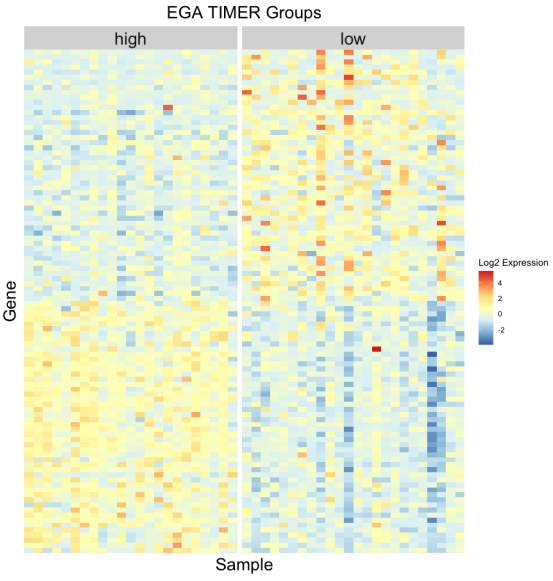
A


UCLA TIMER Groups
high
low
Gene
Sample
Log2 Expression

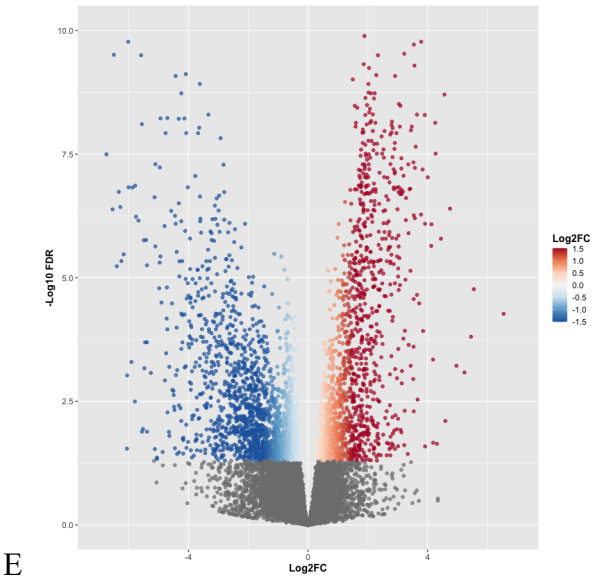
B


-Log10 FDR
Log2FC
Log2FC
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5

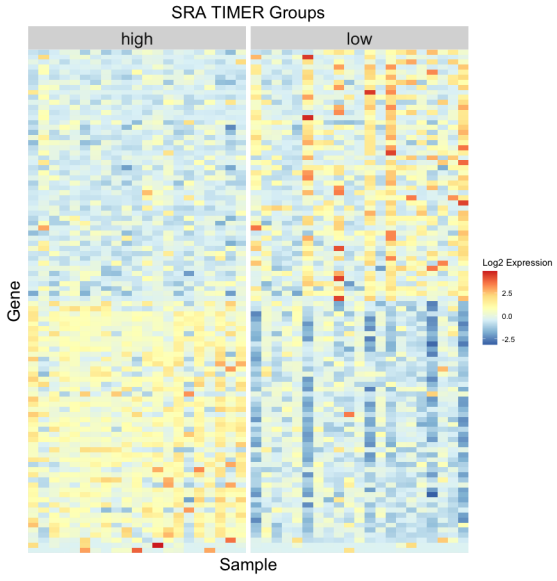
C


EGA TIMER Groups
high
low
Gene
Sample
Log2 Expression

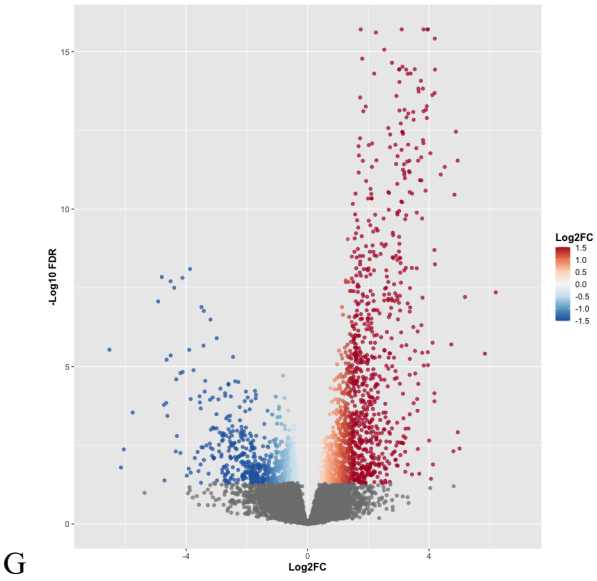
D


10.0
7.5
5.0
2.5
0.0
-Log10 FDR
-4
0
4
Log2FC
Log2FC
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5

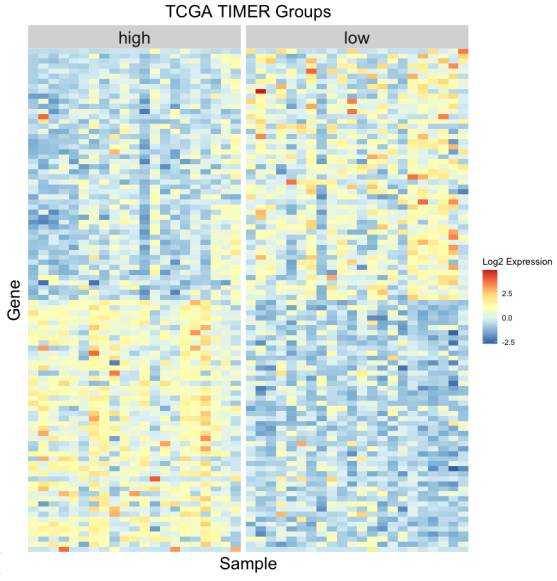
E


SRA TIMER Groups
high
low
Gene
Sample
Log2 Expression
2.5
0.0
-2.5

F


15
10
5
0
-Log10 FDR
-4
0
4
Log2FC
Log2FC
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5

G


TCGA TIMER Groups
high
low
Gene
Sample
Log2 Expression
2.5
0.0
-2.5

H

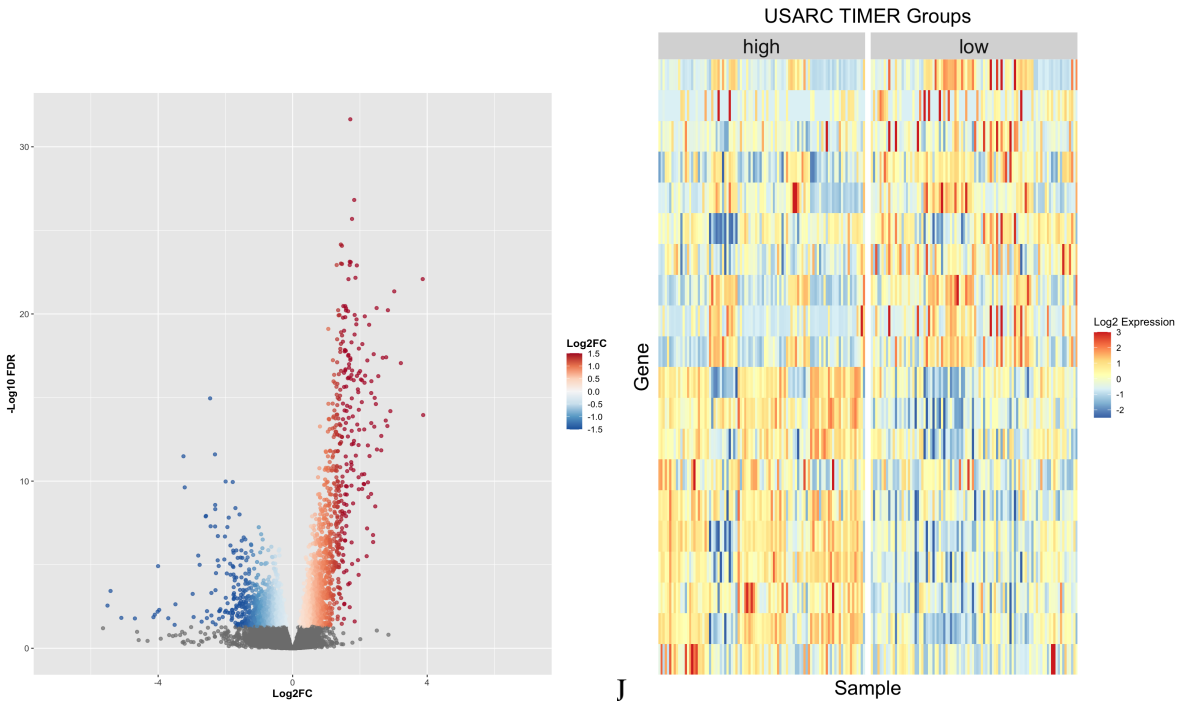
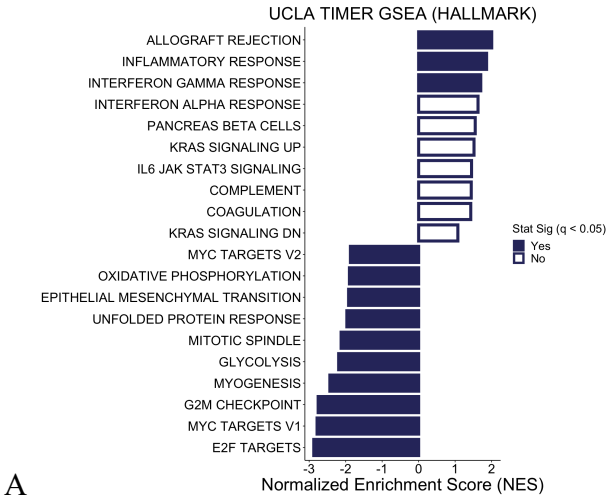
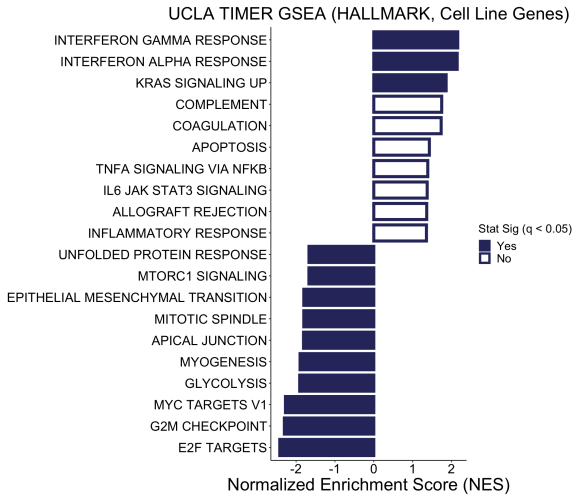


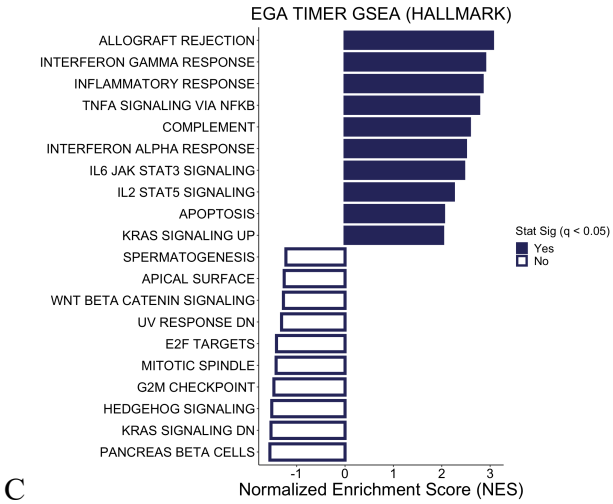
Figure 4-14: Transcriptomic analyses of human USARC tumors, comparing samples with high versus low immune cell infiltration. (A, C, E, G, I) Volcano plots demonstrate results of DESeq analysis comparing samples with high versus low immune cell infiltration. Colored dots represent statistically significant results. (B, D, F, H, J) Heatmaps showing the top genes associated with samples with high versus low immune cell infiltration on DESeq. Values are scaled across rows. Analyses were performed on each dataset (EGA, SRA, TCGA, UCLA) separately (A-H) as well as all datasets together (I-J).


UCLA TIMER GSEA (HALLMARK)
ALLOGRAFT REJECTION
INFLAMMATORY RESPONSE
INTERFERON GAMMA RESPONSE
INTERFERON ALPHA RESPONSE
PANCREAS BETA CELLS
KRAS SIGNALING UP
IL6 JAK STAT3 SIGNALING
COMPLEMENT
COAGULATION
KRAS SIGNALING DN
MYC TARGETS V2
OXIDATIVE PHOSPHORYLATION
EPITHELIAL MESENCHYMAL TRANSITION
UNFOLDED PROTEIN RESPONSE
MITOTIC SPINDLE
GLYCOLYSIS
MYOGENESIS
G2M CHECKPOINT
MYC TARGETS V1
E2F TARGETS
Stat Sig (q < 0.05)
Yes
No
-3 -2 -1 0 1 2
Normalized Enrichment Score (NES)

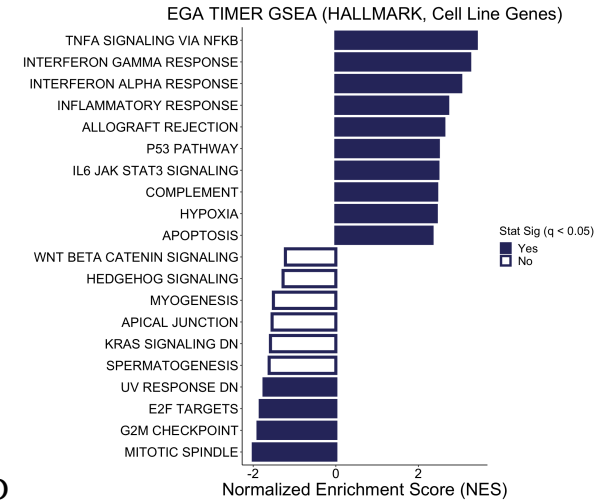
A


UCLA TIMER GSEA (HALLMARK, Cell Line Genes)
INTERFERON GAMMA RESPONSE
INTERFERON ALPHA RESPONSE
KRAS SIGNALING UP
COMPLEMENT
COAGULATION
APOPTOSIS
TNFA SIGNALING VIA NFKB
IL6 JAK STAT3 SIGNALING
ALLOGRAFT REJECTION
INFLAMMATORY RESPONSE
UNFOLDED PROTEIN RESPONSE
MTORC1 SIGNALING
EPITHELIAL MESENCHYMAL TRANSITION
MITOTIC SPINDLE
APICAL JUNCTION
MYOGENESIS
GLYCOLYSIS
MYC TARGETS V1
G2M CHECKPOINT
E2F TARGETS
Stat Sig (q < 0.05)
Yes
No
-2 -1 0 1 2
Normalized Enrichment Score (NES)

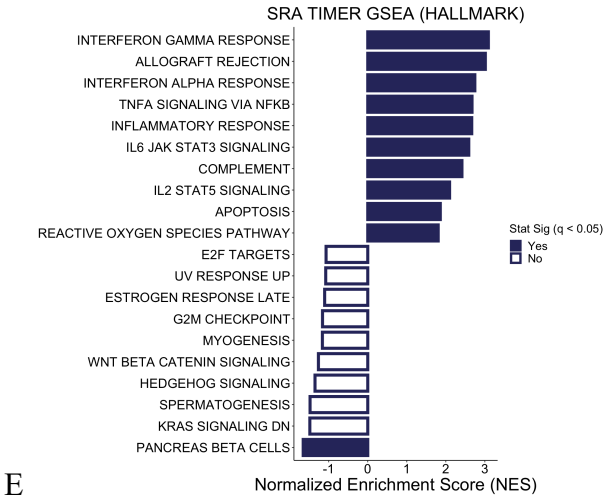
B


EGA TIMER GSEA (HALLMARK)
ALLOGRAFT REJECTION
INTERFERON GAMMA RESPONSE
INFLAMMATORY RESPONSE
TNFA SIGNALING VIA NFKB
COMPLEMENT
INTERFERON ALPHA RESPONSE
IL6 JAK STAT3 SIGNALING
IL2 STAT5 SIGNALING
APOPTOSIS
KRAS SIGNALING UP
SPERMATOGENESIS
APICAL SURFACE
WNT BETA CATENIN SIGNALING
UV RESPONSE DN
E2F TARGETS
MITOTIC SPINDLE
G2M CHECKPOINT
HEDGEHOG SIGNALING
KRAS SIGNALING DN
PANCREAS BETA CELLS
Stat Sig (q < 0.05)
Yes
No
-1 0 1 2 3
Normalized Enrichment Score (NES)

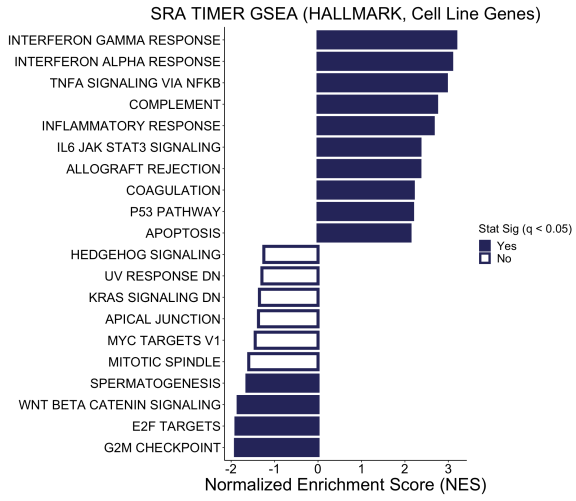
C


EGA TIMER GSEA (HALLMARK, Cell Line Genes)
TNFA SIGNALING VIA NFKB
INTERFERON GAMMA RESPONSE
INTERFERON ALPHA RESPONSE
INFLAMMATORY RESPONSE
ALLOGRAFT REJECTION
P53 PATHWAY
IL6 JAK STAT3 SIGNALING
COMPLEMENT
HYPOXIA
APOPTOSIS
WNT BETA CATENIN SIGNALING
HEDGEHOG SIGNALING
MYOGENESIS
APICAL JUNCTION
KRAS SIGNALING DN
SPERMATOGENESIS
UV RESPONSE DN
E2F TARGETS
G2M CHECKPOINT
MITOTIC SPINDLE
Stat Sig (q < 0.05)
Yes
No
-2 0 2
Normalized Enrichment Score (NES)

D


SRA TIMER GSEA (HALLMARK)
INTERFERON GAMMA RESPONSE
ALLOGRAFT REJECTION
INTERFERON ALPHA RESPONSE
TNFA SIGNALING VIA NFKB
INFLAMMATORY RESPONSE
IL6 JAK STAT3 SIGNALING
COMPLEMENT
IL2 STAT5 SIGNALING
APOPTOSIS
REACTIVE OXYGEN SPECIES PATHWAY
E2F TARGETS
UV RESPONSE UP
ESTROGEN RESPONSE LATE
G2M CHECKPOINT
MYOGENESIS
WNT BETA CATENIN SIGNALING
HEDGEHOG SIGNALING
SPERMATOGENESIS
KRAS SIGNALING DN
PANCREAS BETA CELLS
Stat Sig (q < 0.05)
Yes
No
-1 0 1 2 3
Normalized Enrichment Score (NES)

E


SRA TIMER GSEA (HALLMARK, Cell Line Genes)
INTERFERON GAMMA RESPONSE
INTERFERON ALPHA RESPONSE
TNFA SIGNALING VIA NFKB
COMPLEMENT
INFLAMMATORY RESPONSE
IL6 JAK STAT3 SIGNALING
ALLOGRAFT REJECTION
COAGULATION
P53 PATHWAY
APOPTOSIS
HEDGEHOG SIGNALING
UV RESPONSE DN
KRAS SIGNALING DN
APICAL JUNCTION
MYC TARGETS V1
MITOTIC SPINDLE
SPERMATOGENESIS
WNT BETA CATENIN SIGNALING
E2F TARGETS
G2M CHECKPOINT
Stat Sig (q < 0.05)
Yes
No
-2 -1 0 1 2 3
Normalized Enrichment Score (NES)

F

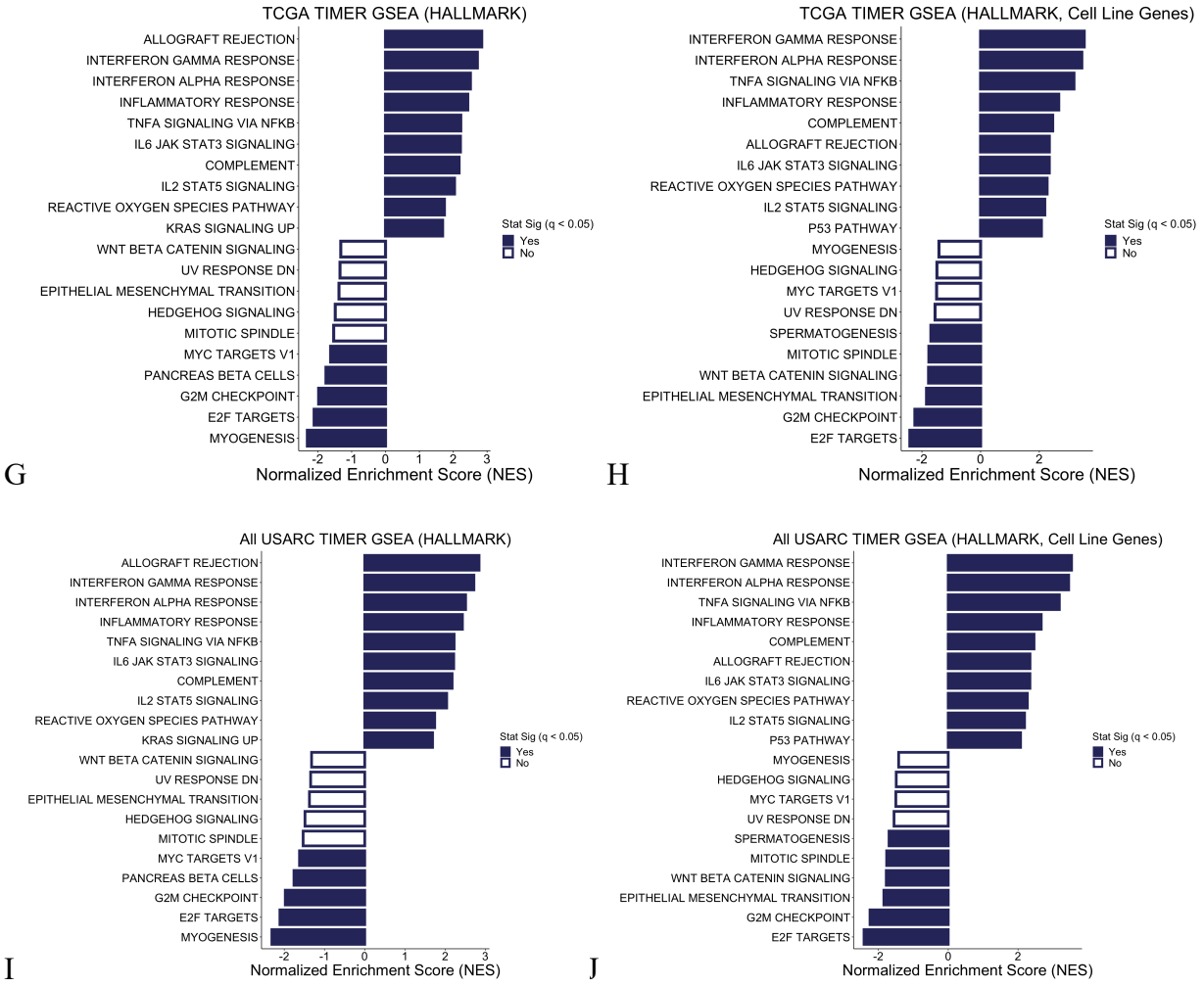
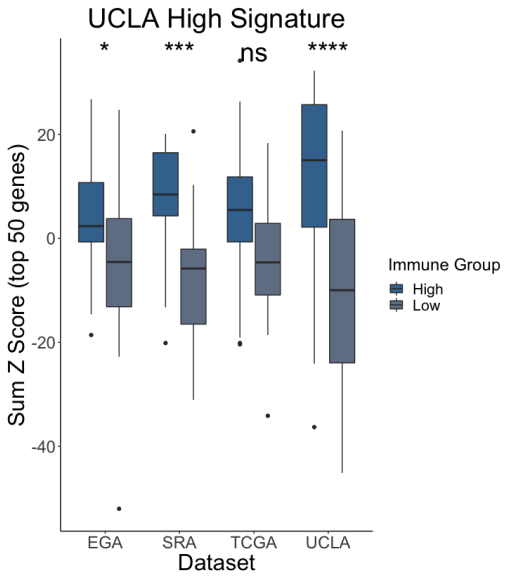
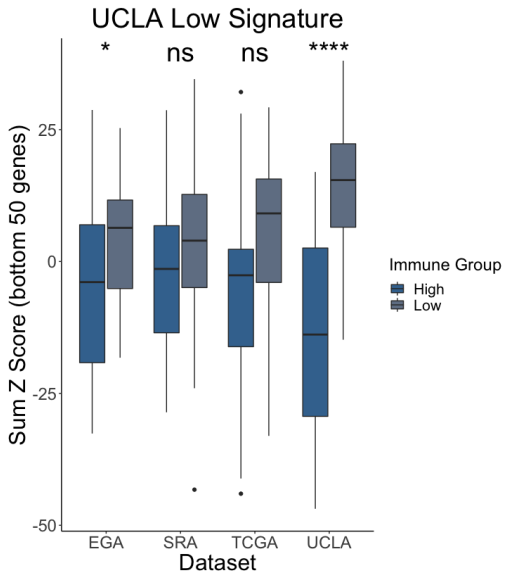


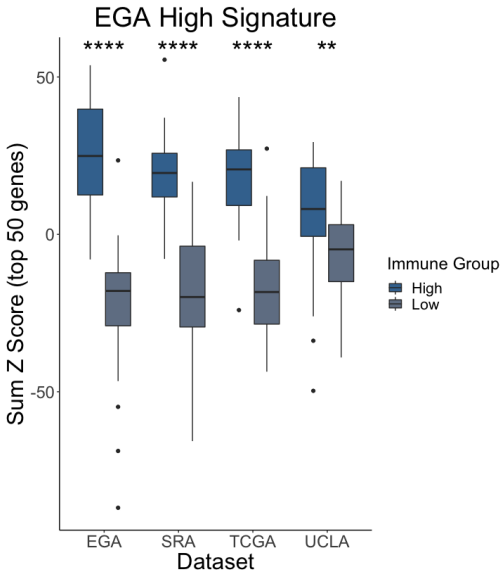
Figure 4-15: GSEA analyses, using Hallmark gene sets, of DESeq results of human USARC samples. Samples with high immune cell infiltration demonstrated an association with gene sets relating to immune function. Samples with low immune cell infiltration demonstrated an association with gene sets relating to cell division and growth, such as mitotic spindle and G2M checkpoint. These results were similar whether all protein coding genes were included (A, C, E, G, I) or only genes highly expressed in cell line genes were included (B, D, F, H, J). Analyses were performed on each dataset (EGA, SRA, TCGA, UCLA) separately (A-H) as well as all datasets together (I-J).


UCLA High Signature
*
***
ns
****
Sum Z Score (top 50 genes)
20
0
-20
-40
EGA
SRA
TCGA
UCLA
Dataset
Immune Group
High
Low

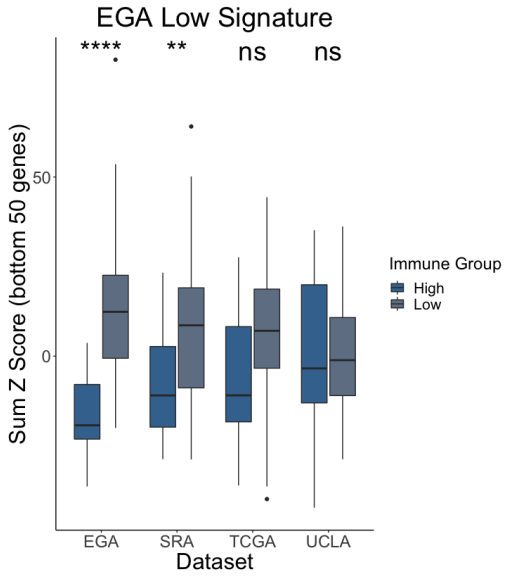
A


UCLA Low Signature
*
ns
ns
****
Sum Z Score (bottom 50 genes)
25
0
-25
-50
EGA
SRA
TCGA
UCLA
Dataset
Immune Group
High
Low

B


EGA High Signature
****
****
****
**
Sum Z Score (top 50 genes)
50
0
-50
EGA
SRA
TCGA
UCLA
Dataset
Immune Group
High
Low

C


EGA Low Signature
****
**
ns
ns
Sum Z Score (bottom 50 genes)
50
0
EGA
SRA
TCGA
UCLA
Dataset
Immune Group
High
Low

D

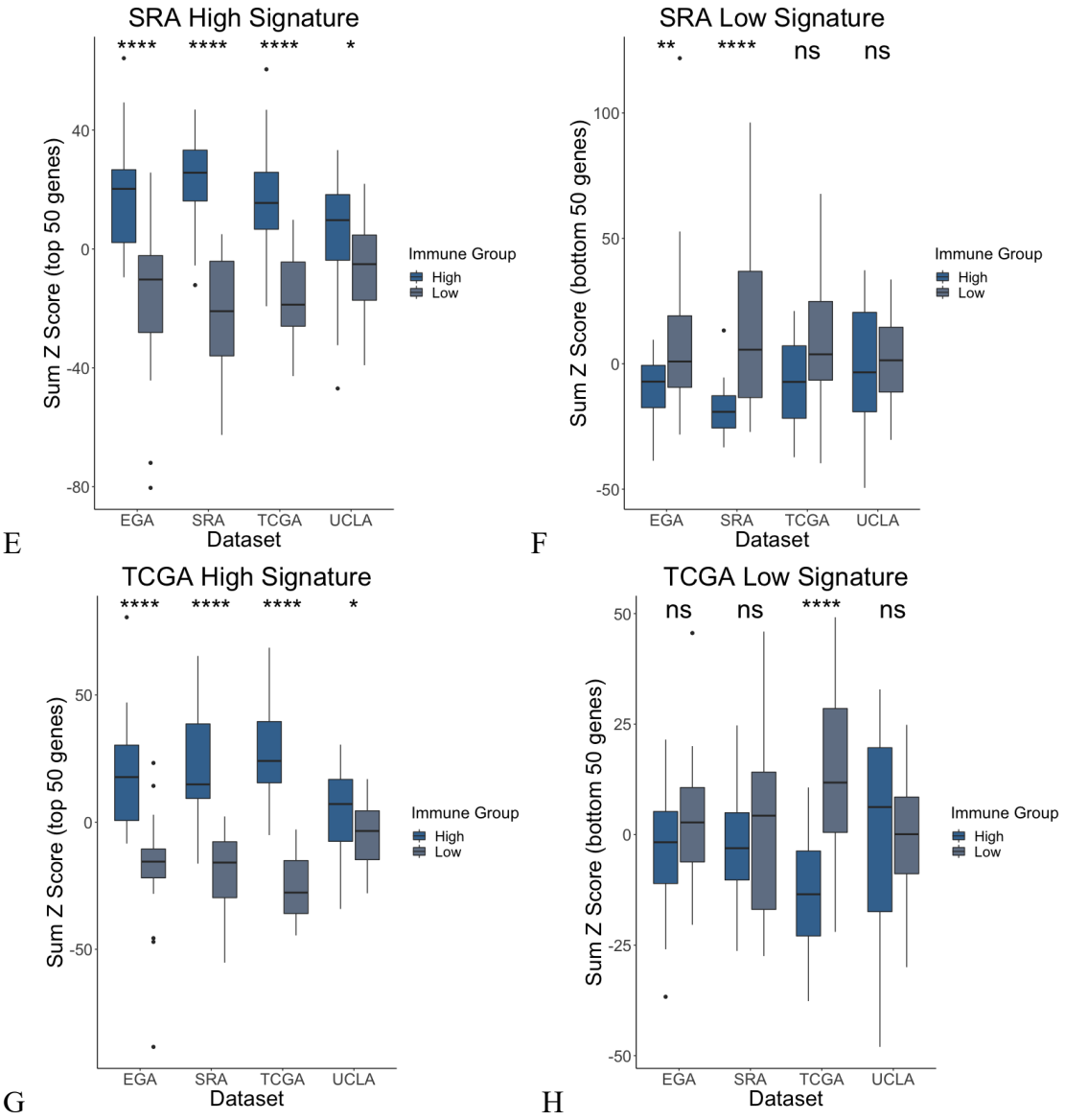
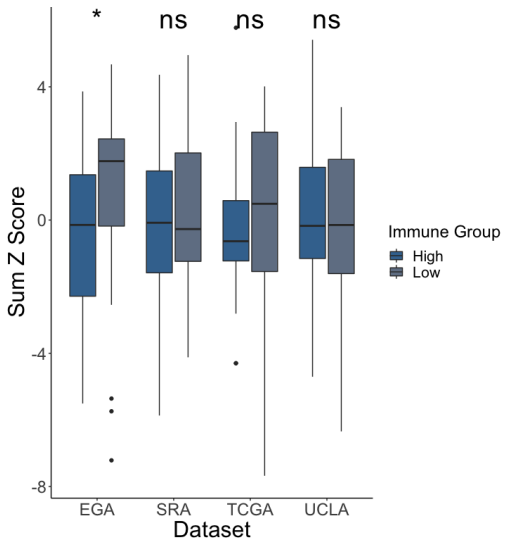


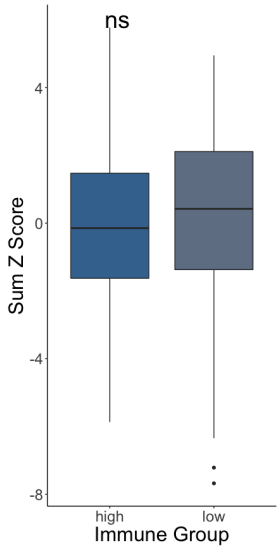
Figure 4-16: Transcriptomic signatures associated with high and low immune cell infiltration were determined using DESeq in each dataset (EGA, SRA, TCGA, UCLA). The signatures determined in each dataset were tested and validated in each of the other datasets in this study. Sum z scores of the top 50 or bottom 50 genes in each of these gene lists were calculated and compared.

4.9.12 Samples with low immune cell infiltration were associated with stem cell signatures

Mesenchymal stem cells, the presumed cell of origin of undifferentiated sarcoma, are known to play a role in immune evasion. Mesenchymal and pluripotent stem cell signatures were compared between samples with high and low immune cell infiltration[41–43]. Sum z scores of these genes were compared across samples. Samples with low immune cell infiltration showed a trend toward increased mesenchymal stem cell signatures across most of the datasets. This difference was statistically significant in the EGA dataset (Figure 4-17A). Samples with low immune cell infiltration generally showed a higher pluripotent stem cell score. This was statistically significant in the EGA and SRA datasets and when all samples were analyzed together (Figure 4-17C&D).



A



B

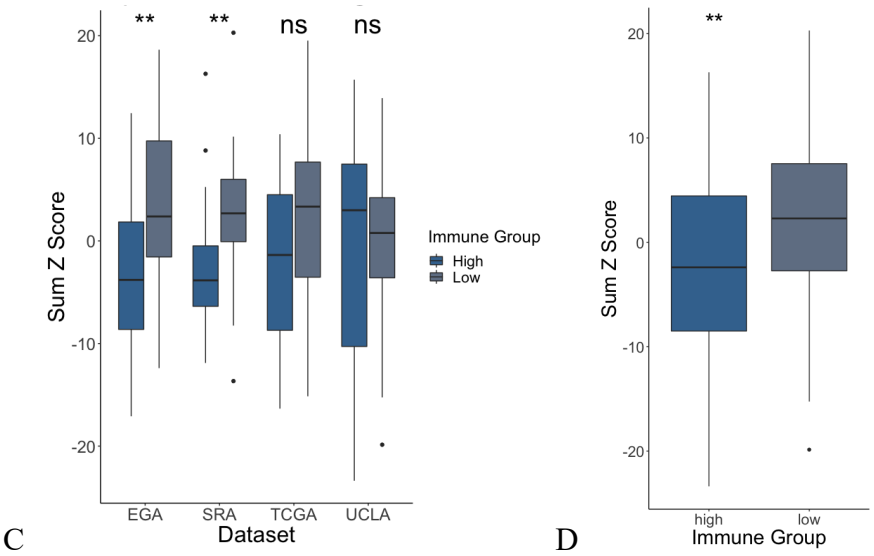


Figure 4-17: Summary of the association between stem cell gene signatures and immune cell infiltration in USARC tumor samples. Samples with low immune cell infiltration showed a trend toward increased mesenchymal stem cell signatures across most of the datasets. This difference was statistically significant in the EGA dataset (C&D). Samples with low immune cell infiltration generally showed a higher pluripotent stem cell score. This was statistically significant in the EGA and SRA datasets (C) and when all samples were analyzed together (D).

4.9.13 Additional copy number alteration analyses

The results below summarize additional analyses of copy number alterations in our study population. The UCLA and the TCGA datasets were analyzed separately. These copy number analyses quantified copy number changes by the number of copy number segment changes, unless otherwise noted. The total numbers of copy number changes are summarized in the histograms in Figure 4-18A&B. In addition to exploring the association between total immune cell infiltration and copy number changes, we examined the association between the infiltration

of various immune cell types and copy number changes. Specifically, we quantified total immune cell infiltration using TIMER, CD8 T cell infiltration using CIBERSORTx, macrophages using CIBERSORTx, and macrophage/CD8 T cell ratio using CIBERSORTx. These tools were selected based on the results summarized in Chapter 3 (Figure 3-4). Samples were characterized as high or low in each of these categories relative to the median value. The results of this categorization are summarized in Tables 4-16 & 4-17. We then examined the differences in copy number changes between these categories using the Wilcoxon test. In the UCLA cohort, there was an association between higher copy number changes and low overall immune cell infiltration and a trend toward high copy number changes and low CD8T cell infiltration (Figure 4-19). In the TCGA cohort, there was an association between higher copy number changes and low overall immune cell infiltration as well as low macrophages (Figure 4-20). In the TCGA cohort, we additionally analyzed the association between overall immune cell infiltration and copy number gain and losses separately (Figure 4-21). The results were similar when examining the association between overall immune cell infiltration and total copy number changes (Figure 4-4), copy number gains alone, and copy number losses alone (Figure 4-21). Figures representing the copy number changes seen in representative chromosomes in the TCGA dataset are shown in Figure 4-22. In addition to analyzing CN changes by copy number, the analysis was also performed by defining copy number changes according to % genome altered in the UCLA cohort. These results are summarized in Figure 4-23 and show a trend toward higher copy number changes in the low immune cell infiltration group, though this did not reach statistical significance. On the histogram in Figure 4-18, there is one sample with a particularly large number of segments with copy number changes, which may represent an outlier. Sensitivity analyses were also performed dropping the UCLA sample with the highest CN

changes as well as the two samples with the highest copy number changes. There was no meaningful difference in the results (data not shown).

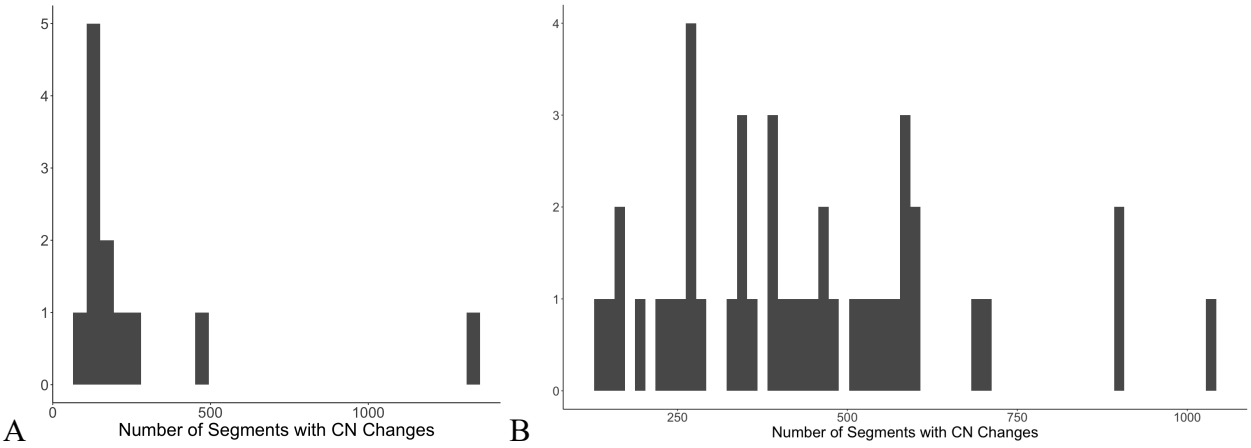


Figure 4-18: Histograms summarizing number of segments with copy number alteration in the (A) UCLA dataset and (B) TCGA dataset.

Cell Type	In-silico Immune Deconvolution Tool	High (n)	Low (n)
Total Immune Cell Infiltrates	TIMER	3	9
Macrophage	CIBERSORTx	4	8
CD8 T Cell	CIBERSORTx	6	6
Macrophage/CD8 T Cell Ratio	CIBERSORTx	6	6

Table 4-16: Summary of immune cell groups in the UCLA dataset.

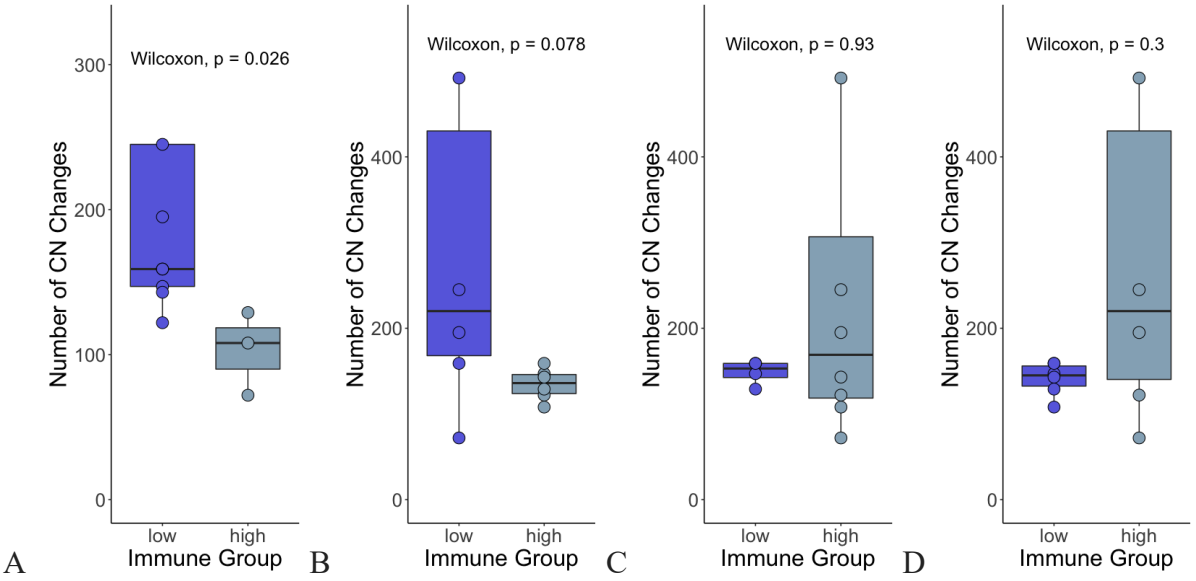


Figure 4-19: Summary of copy number changes associated with immune cell infiltration in the UCLA dataset. Data shown is number of CN segment changes. Comparisons were made between high versus low amounts of immune cell infiltration according to the median of the following in-silico immune deconvolution scores: (A) TIMER overall, (B) CIBERSORTx CD8 T cell, (C) CIBERSORTx macrophage, and (D) CIBERSORTx macrophage/CD8 T cell ratio. Wilcoxon test was used to compare groups.

Cell Type	In-silico Immune Deconvolution Tool	High (n)	Low (n)
Total Immune Cell Infiltrates	TIMER	21	22
Macrophage	CIBERSORTx	21	22
CD8 T Cell	CIBERSORTx	21	22
Macrophage/CD8 T Cell Ratio	CIBERSORTx	21	22

Table 4-17: Summary of immune cell groups in the TCGA dataset.

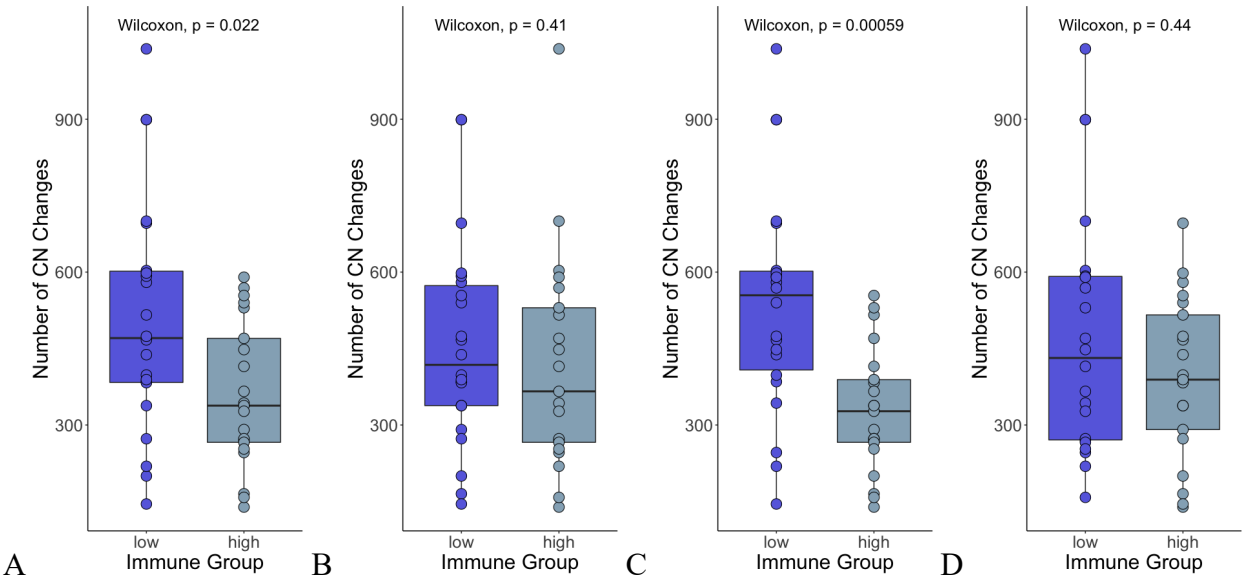
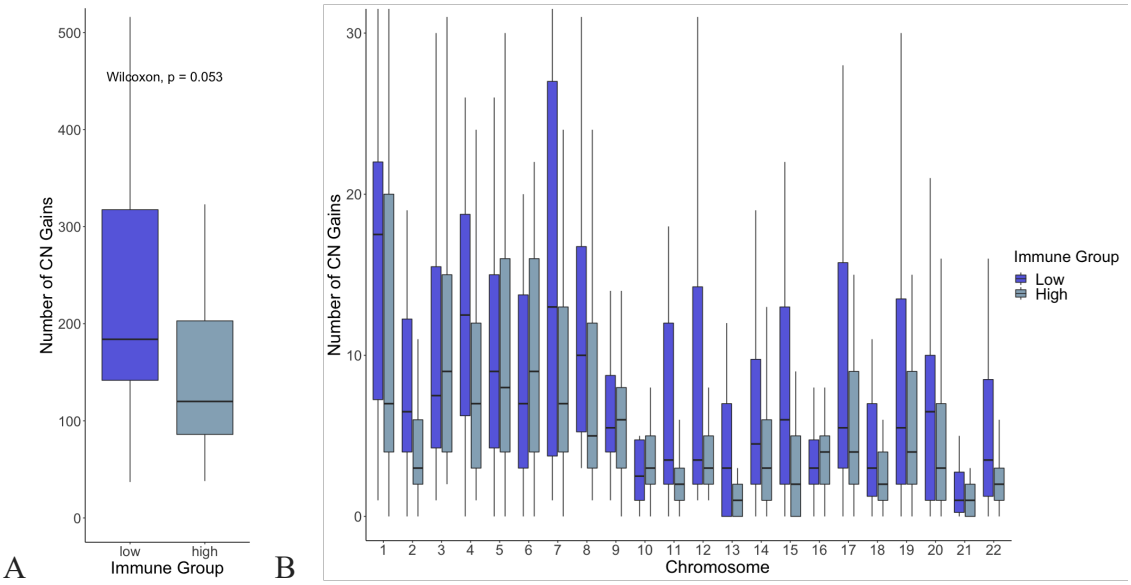


Figure 4-20: Summary of copy number changes associated with immune cell infiltration in the TCGA dataset. Data shown is number of CN segment changes. Comparisons were made between high versus low amounts of immune cell infiltration according to the median of the following in-silico immune deconvolution scores: (A) TIMER overall, (B) CIBERSORTx CD8 T cell, (C) CIBERSORTx macrophage, and (D) CIBERSORTx macrophage/CD8 T cell ratio. Wilcoxon test was used to compare groups.



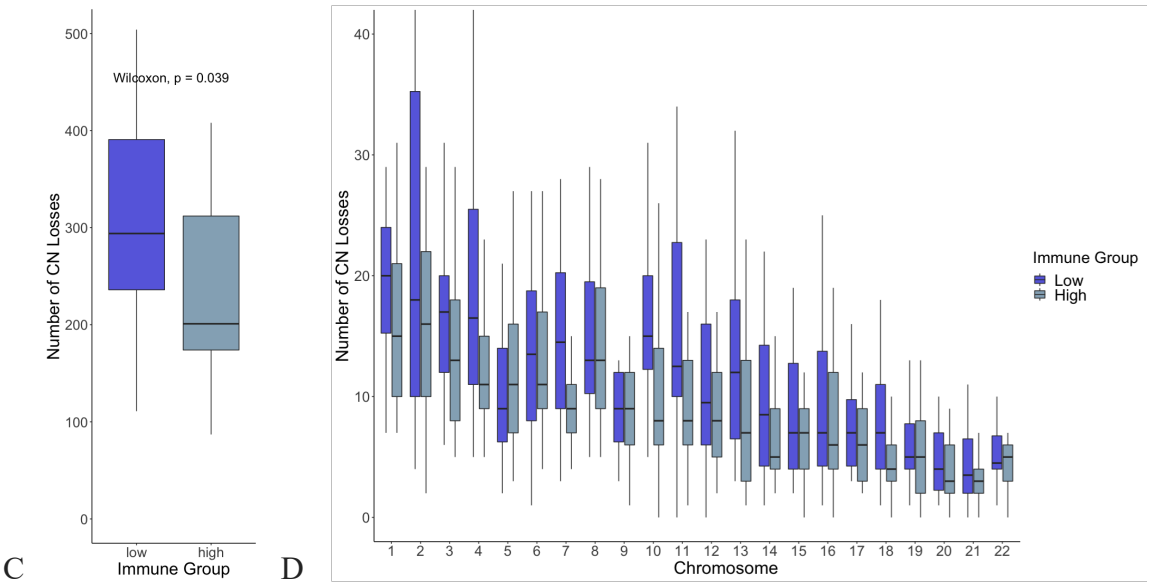


Figure 4-21: Increased copy number alterations are associated with lower immune cell infiltration in undifferentiated sarcoma in the TCGA cohort. Comparisons of total number and per chromosome copy number segment gains (A&B) and losses (C&D) are shown.

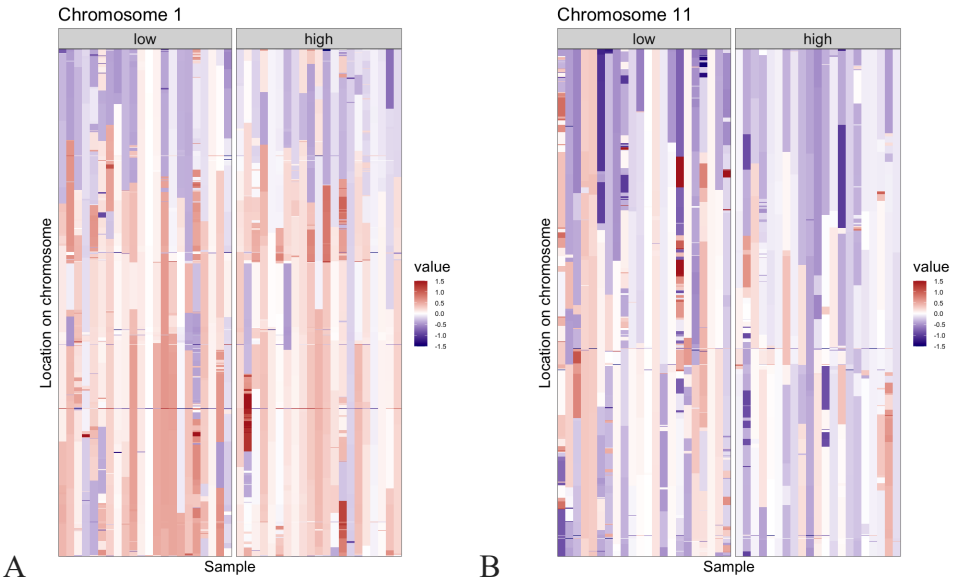


Figure 4-22: Representative images of copy number changes seen in select chromosomes in the TCGA dataset. Each column represents a distinct sample.

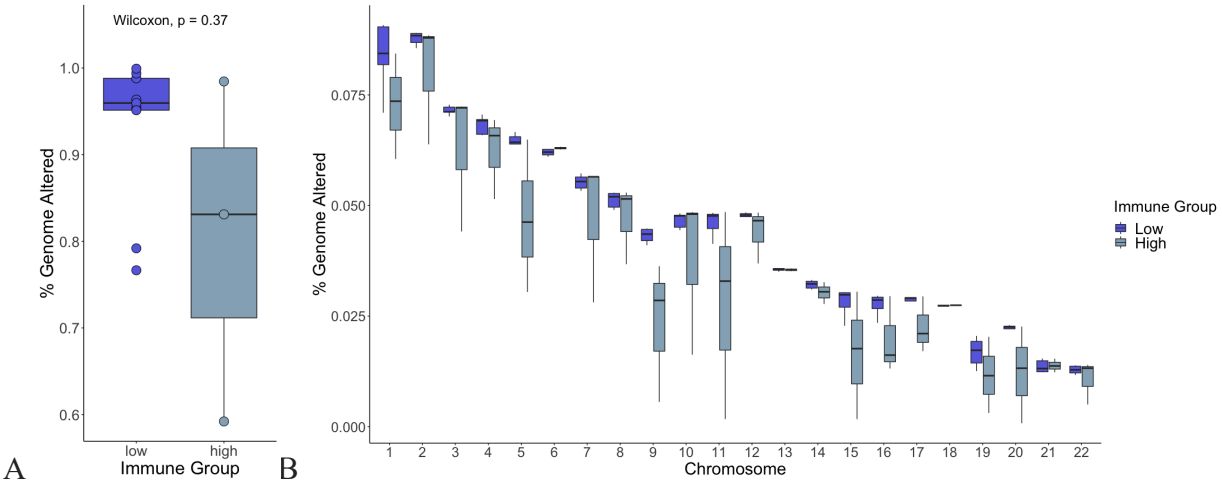


Figure 4-23: There is a trend toward increased copy number alterations associated with lower immune cell infiltration in undifferentiated sarcoma in the UCLA cohort. Comparisons of CN changes quantified by percent genome altered are shown. Comparisons were made across the whole exome (A) and in each chromosome (B).

4.9.14 Analysis of the association between CN changes and immune cell infiltration within various STS subtypes

Figure 4-5 above demonstrates the association between copy number changes and immune cell infiltration in each of the STS histologies. In the analysis above, the results demonstrate that low immune cell infiltration is associated with higher numbers of copy number changes in USARC and DDLPS subtypes, however, this association is not seen in the other STS subtypes nor is it seen in STS overall. Samples were defined as high versus low immune cell infiltration relative to the total immune cell infiltration defined by TIMER within each histology. A sensitivity analysis was performed defining samples as high versus low immune cell infiltration relative to the overall STS cohort, rather than within each histology. The analysis was then repeated. The

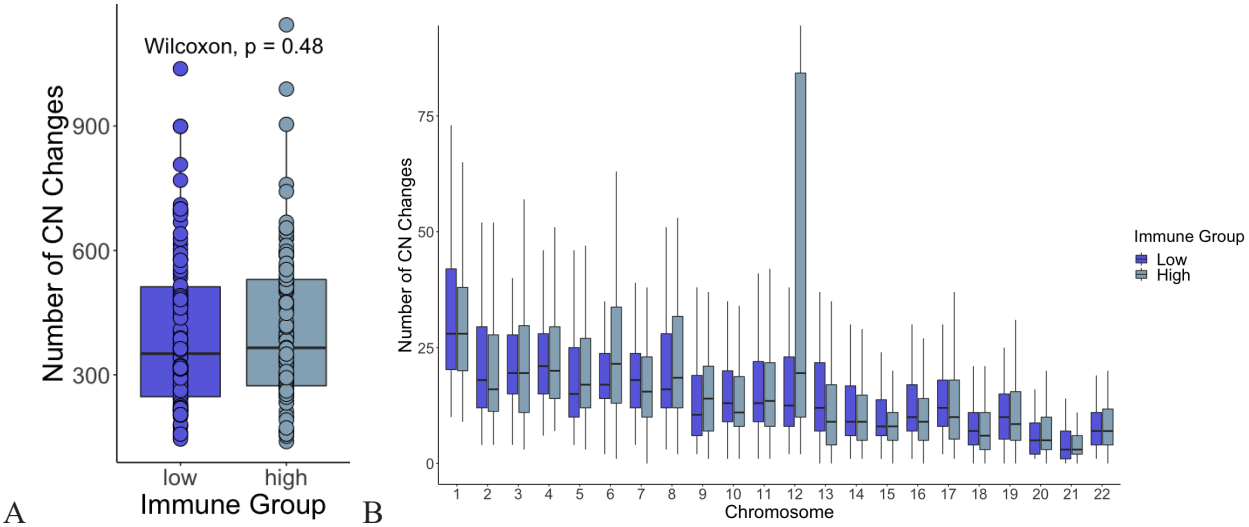
results did not change. Tables 4-18 & 4-19 show the number of each histology that was defined as high versus low immune cell infiltration using each of these definition strategies. Figure 4-24A&B show the analysis of overall copy number changes across samples with high versus low immune cell infiltration across all chromosomes (A) and within each chromosome (B). There was no difference in copy number changes within each group. Figure 4-24C compares the copy number changes across the high versus low immune cell infiltration groups within each histology. In USARC and DDLPS, higher copy number changes are associated with low immune cell infiltration. This association was not seen in the other STS histologies.

		Defined within histology	
		Low	**High**
Defined all together	**Low**	83	19
	High	22	80

Table 4-18: Summary of samples defined as high versus low immune cell infiltration in the TCGA SARC cohort. Samples were defined relative to the median TIMER total immune score across the entire cohort and within each histology. This table demonstrates the number of samples that switched from high to low immune cell infiltration groups as the definition changed.

	Defined all together				Defined within histology			
	Low		**High**		**Low**		**High**	
Histology	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*
Total	102	50.0	102	50.0	105	51.5	99	48.5
DDLPS	17	34.0	33	66.0	25	50.0	25	50.0
MFS	4	23.5	13	76.5	9	52.9	8	47.1
MPNST	3	60.0	2	40.0	3	60.0	2	40.0
SS	9	100.0	0	0.0	5	55.6	4	44.4
STLMS	35	66.0	18	34.0	27	50.9	26	49.1
ULMS	21	77.8	6	22.2	14	51.9	13	48.1
UPS	13	30.2	30	69.8	22	51.2	21	48.8

Table 4-19: Summary of samples defined as high versus low immune cell infiltration in the TCGA SARC cohort. This table shows the number of samples defined as high versus low immune cell infiltration within each histology. Samples were defined relative to the median TIMER total immune score across the entire cohort and within each histology.



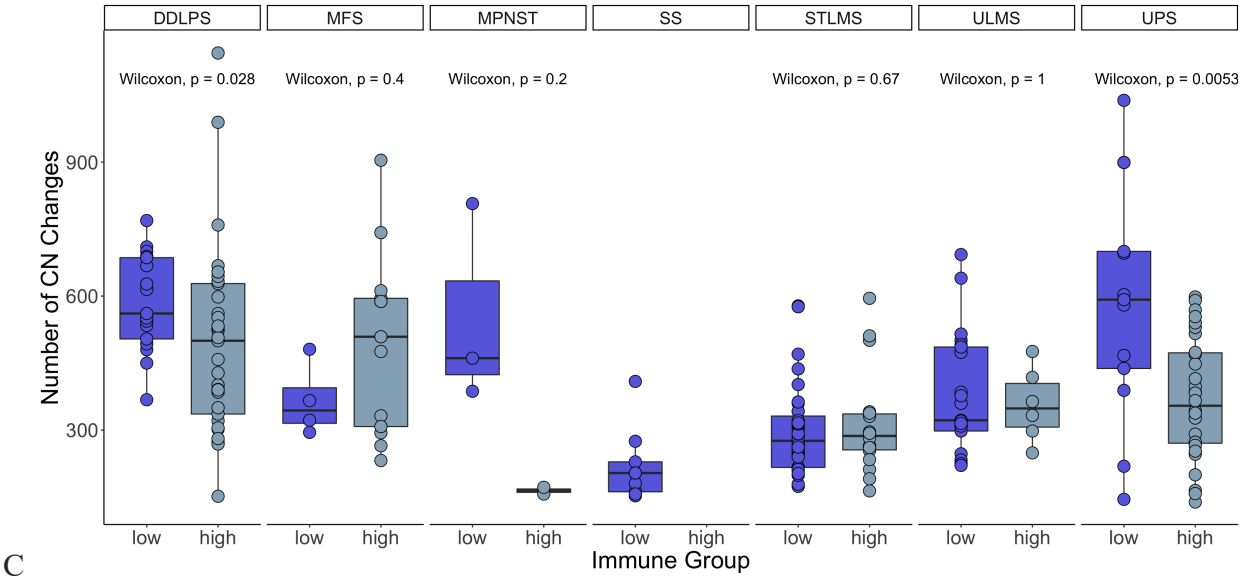


Figure 4-24: Increased copy number alterations are associated with lower immune cell infiltration in USARC and DDLPS, but not in other subtypes of STS and not in STS overall.

(A&B) Bar graphs demonstrating the association between copy number changes and immune cell infiltration in all subtypes of STS from the TCGA. Comparisons were made examining total copy number changes (A) and total copy number changes within each chromosome (B). There was a trend toward increased copy number changes in the low immune cell infiltration group, though this did not reach statistical significance ($p = 0.067$). (C) Increased copy number alternations were associated with low immune cell infiltration in USARC and DDLPS subtypes of STS, though there was no difference in copy number changes between high and low immune cell infiltration groups in the other subtypes of STS. Immune groups were defined relative to the median TIMER score in the STS cohort.

4.9.15 An analysis of the types of copy number changes seen across STS subtypes

Steele et al recently described copy number signatures across cancer types. There were 21 copy number signature groups defined. Here, we examined the copy number signatures seen in the various subtypes of soft tissue sarcoma. There is significant variability in the predominant copy number signature seen in each of the sarcoma subtypes (Figure 4-25A). We also examined the differences in copy number signatures between samples with high and low immune cell infiltration, and there was no significant difference seen (Figure 4-25B). In the primary analysis, scaled values were used for the copy number signatures assigned to each sample (Figures 4-25 A&B). In the sensitivity analyses, total counts of major and minor copy number signatures were tabulated and compared. The results did not change in any significant way (Figures 4-26 & 4-27).

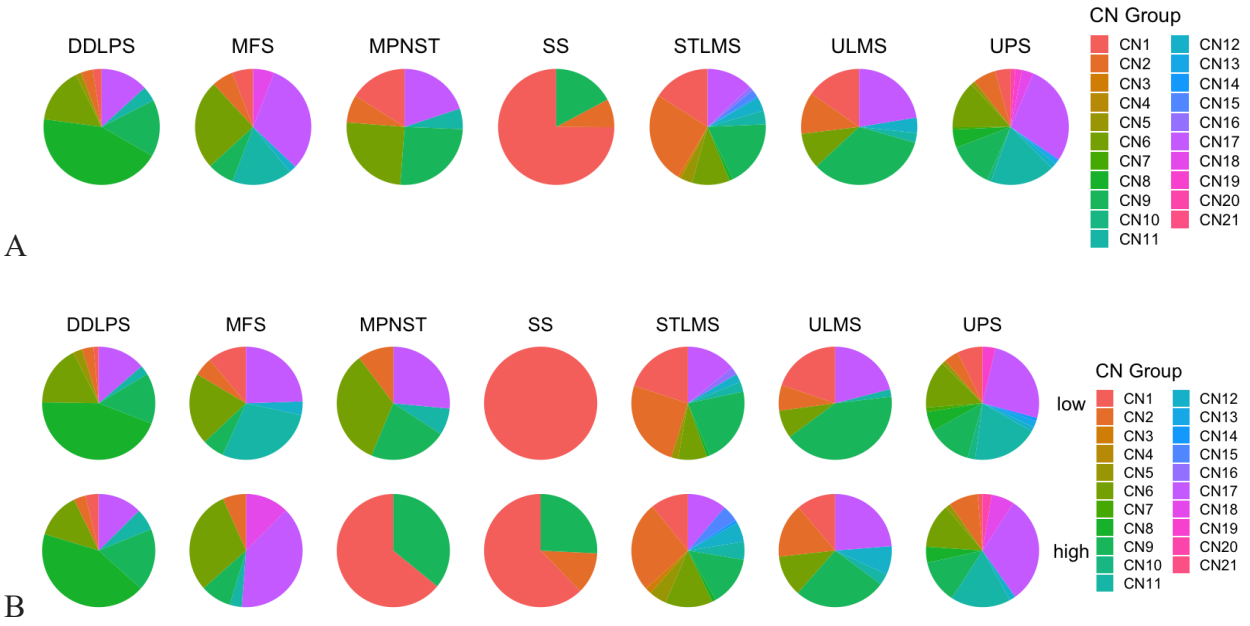
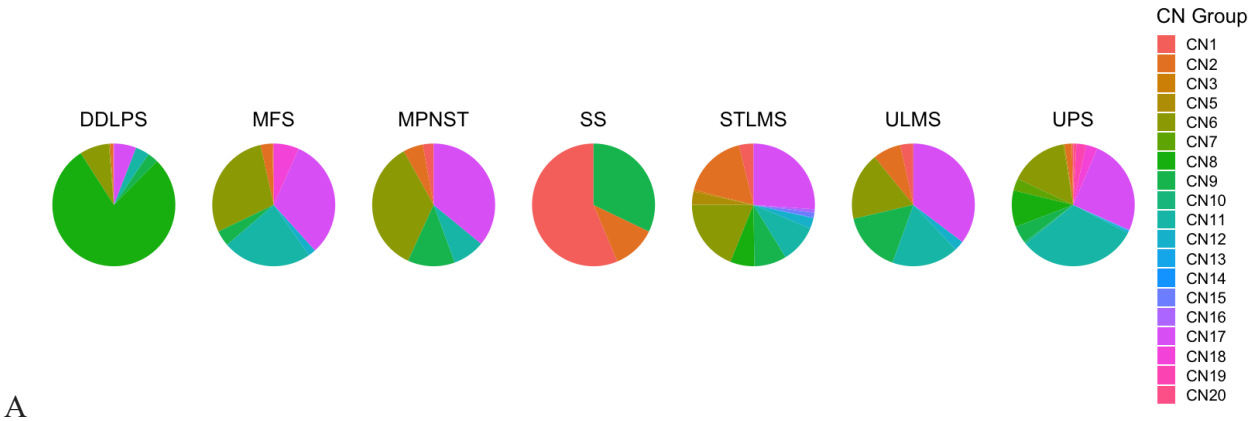
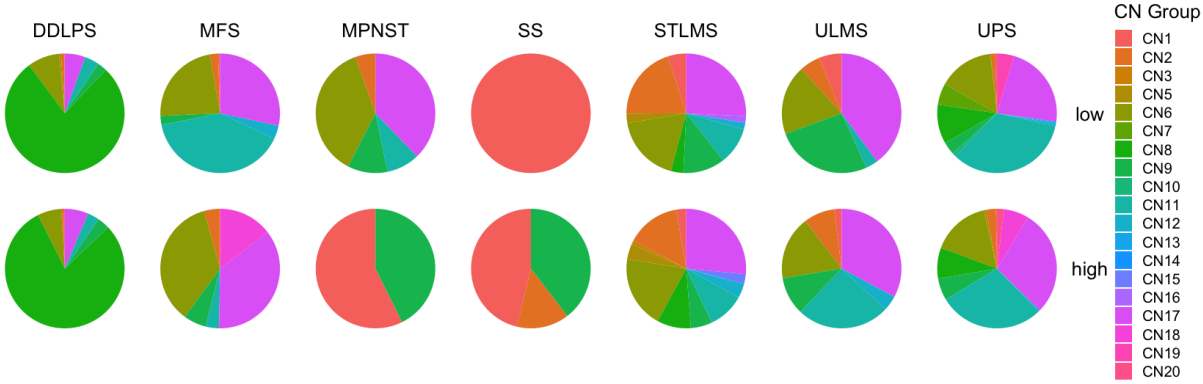


Figure 4-25: Analysis of copy number signatures across STS histologies (scaled values) (A) and between samples with high and low immune cell infiltration within each histology (B). Copy number signatures were defined and analyzed by Steele et al[44]. Values used are scaled copy number signature values assigned to each sample.

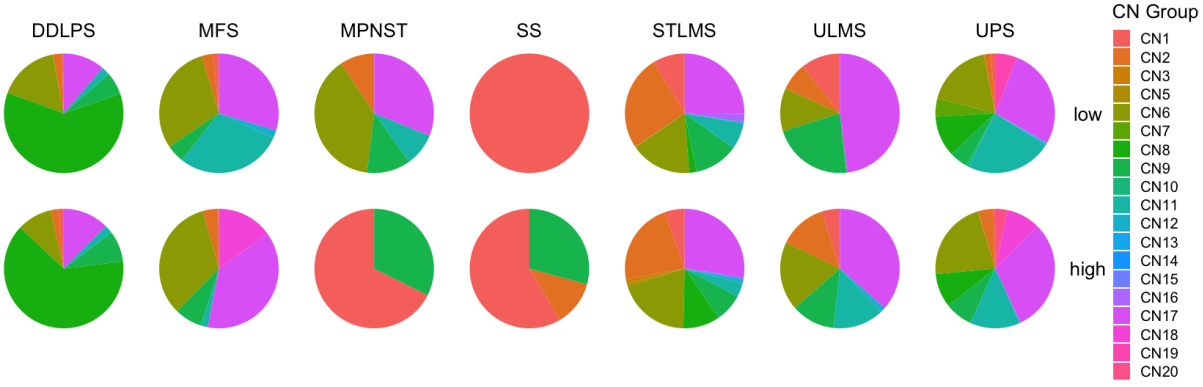


B

Figure 4-26: Analysis of copy number signatures across STS histologies (major and minor values). Copy number signatures were defined and analyzed by Steele et al[44]. Values used are total major (A) and minor (B) copy number signature values assigned to each sample.



A



B

Figure 4-27: Analysis of copy number signatures between samples with high and low immune

cell infiltration within each STS histology group (major and minor values). Copy number signatures were defined and analyzed by Steele et al[44]. Values used are total major (A) and minor (B) copy number signature values assigned to each sample.

4.9.16 Survival analyses

Survival analyses were performed to determine the association between immune cell infiltration and survival in patients with USARC. Tumors were classified as having "high" or "low" immune cell infiltration based on the median TIMER score, as previously described. Survival data was available for UCLA and TCGA cohorts. The cohorts were examined together and separately. There was a trend toward decreased survival associated with low immune cell infiltration, however, this difference was not statistically significant (Figure 4-28). We additionally examined the association between survival and the expression of select genes that were significantly differently expressed in all datasets analyzed together or separately. The genes analyzed were COL11A2 (associated with low immune cell infiltration), SYNGR2 (associated with high immune cell infiltration), and HMOX1 (associated with high immune cell infiltration). Samples were categorized based on the median expression of these genes. There was a trend toward improved survival with high SYNGR2 expression, no difference in survival with HMOX1 expression, and a trend toward improved survival with COL11A2 expression. Datasets were analyzed together (UCLA and TCGA) as well as separately. There was no meaningful difference in the results. The results of the datasets analyzed together is shown (Figure 4-29).

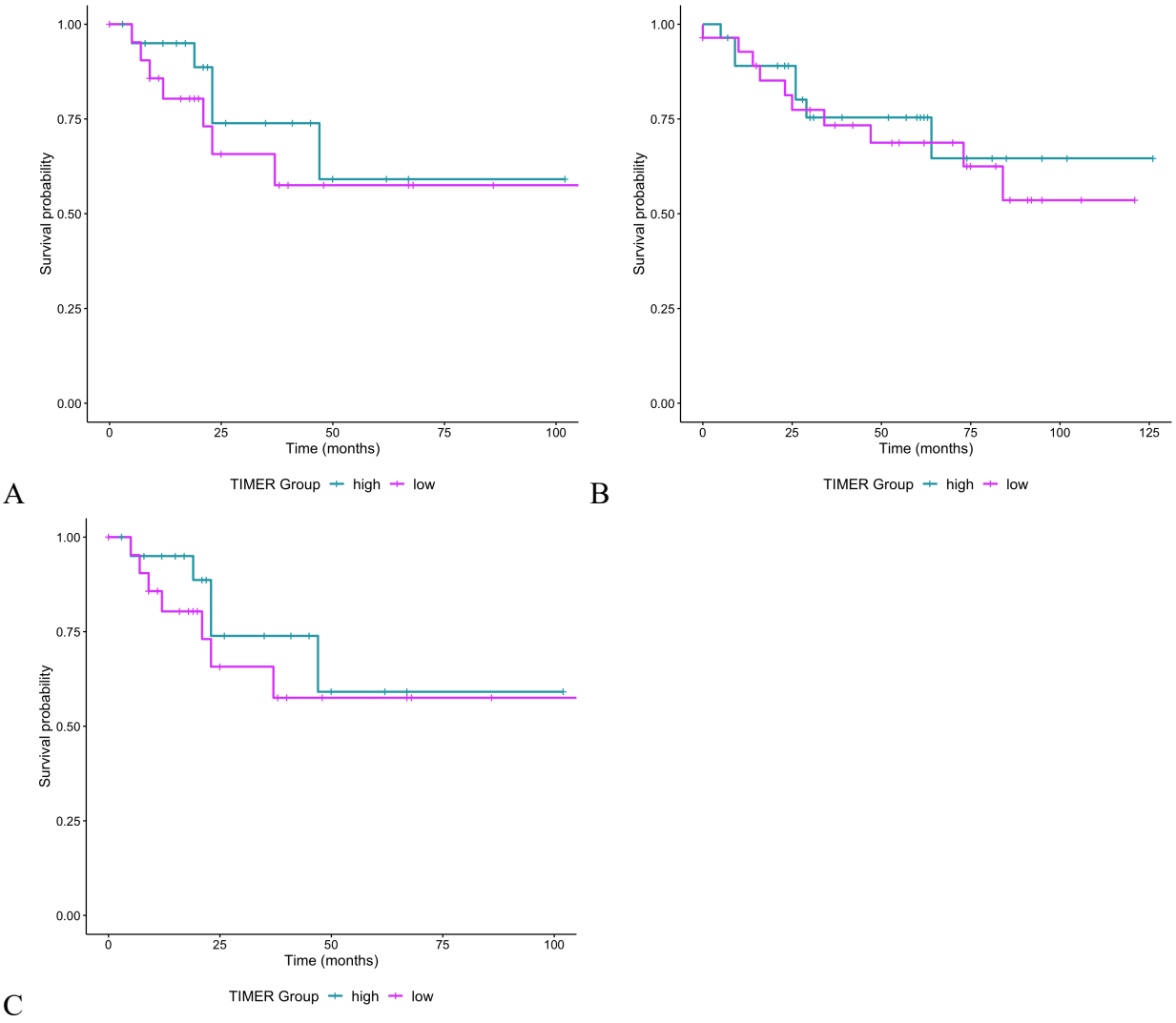


Figure 4-28: The association between immune cell infiltration and survival in patients with undifferentiated sarcoma. Kaplan Meier curves are shown. Samples were categorized as having "high" or "low" immune cell infiltration relative to the median overall TIMER score. Survival data was available for the UCLA and TCGA cohorts. Comparisons were made using (A) all samples, (B) UCLA samples only, and (C) TCGA samples only. There was a trend toward decreased survival associated with low immune cell infiltration, though this did not reach statistical significance ($p = 0.4$, $p = 0.7$, and $p = 0.5$, respectively).

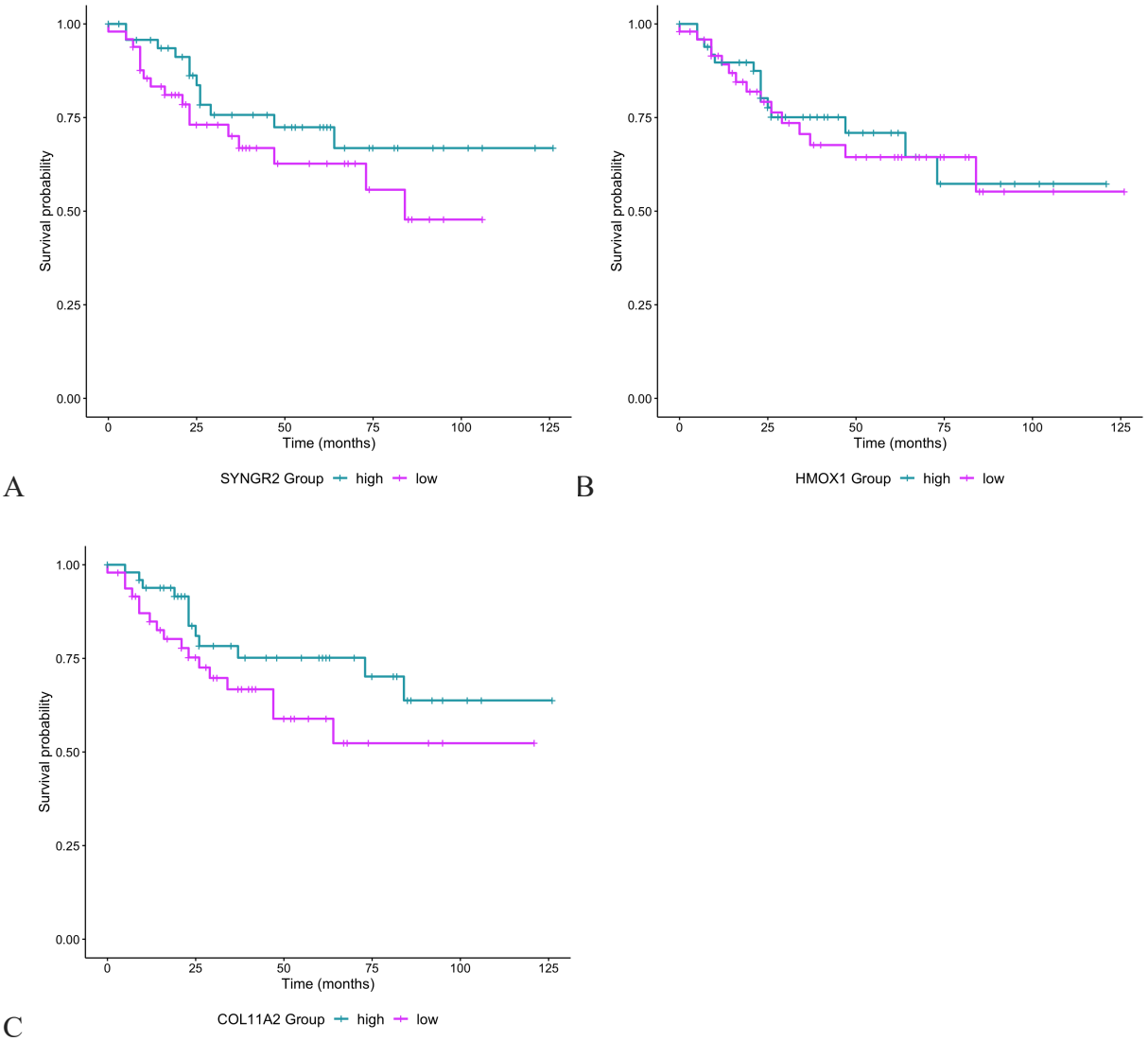


Figure 4-29: The association between expression of (A) SYNGR2, (B) HMOX1, and (C) COL11A2 and survival in patients with undifferentiated sarcoma. Kaplan Meier curves are shown. Samples were categorized as having "high" or "low" gene expression relative to the median. Survival data was available for the UCLA and TCGA cohorts. Comparisons shown were made using all samples. There was a trend toward improved survival with (A) SYNGR2 (p = 0.1) and (C) COL11A2 (p = 0.1) expression and no difference in survival with (B) HMOX1 (p = 0.8) expression.

4.9.17 An analysis of transcriptomic and genomic changes before and after treatment with immunotherapy

We explored the association between immunotherapy treatment and transcriptomic and genomic signatures. The data used in this analysis was RNA-Seq data from four USARC cell lines. Sequencing was performed both pre- and post-treatment with BO112. Additionally, we identified paired samples from patients treated at our institution. Samples from their tumors were available both from the patients before and after they were treated with immunotherapy. A summary of the clinical data is in Table 4-20. PCAs were performed on the cell lines and the tumors separately. In the cell line analysis, PC4 appears to identify genes associated with immunotherapy treatment. In the human tumor analysis, PC3 identifies a signature associated with immunotherapy treatment. Tables 4-21 & 4-22 show the top 50 genes that define each of these PCs. Genomic analyses of the human tumor samples show that the samples treated with immunotherapy demonstrated lower copy number segment changes than their paired samples prior to immunotherapy (Figure 4-29).

Pt #	Sex	Resection	Tumor Site	Size(cm)	Prior treatment	Primary
1	F	Resection	Lung	4.5	Immunotherapy	Metastasis
1	F	Resection	Sub-scapular	11.5	XRT	Primary
2	M	Resection	Abd/RP	12.0	Immunotherapy	Recurrence
2	M	Resection	Abd/RP	8.0	Chemo	Primary

Table 4-20: Summary of clinical characteristics from paired samples from patients treated with and without immunotherapy.

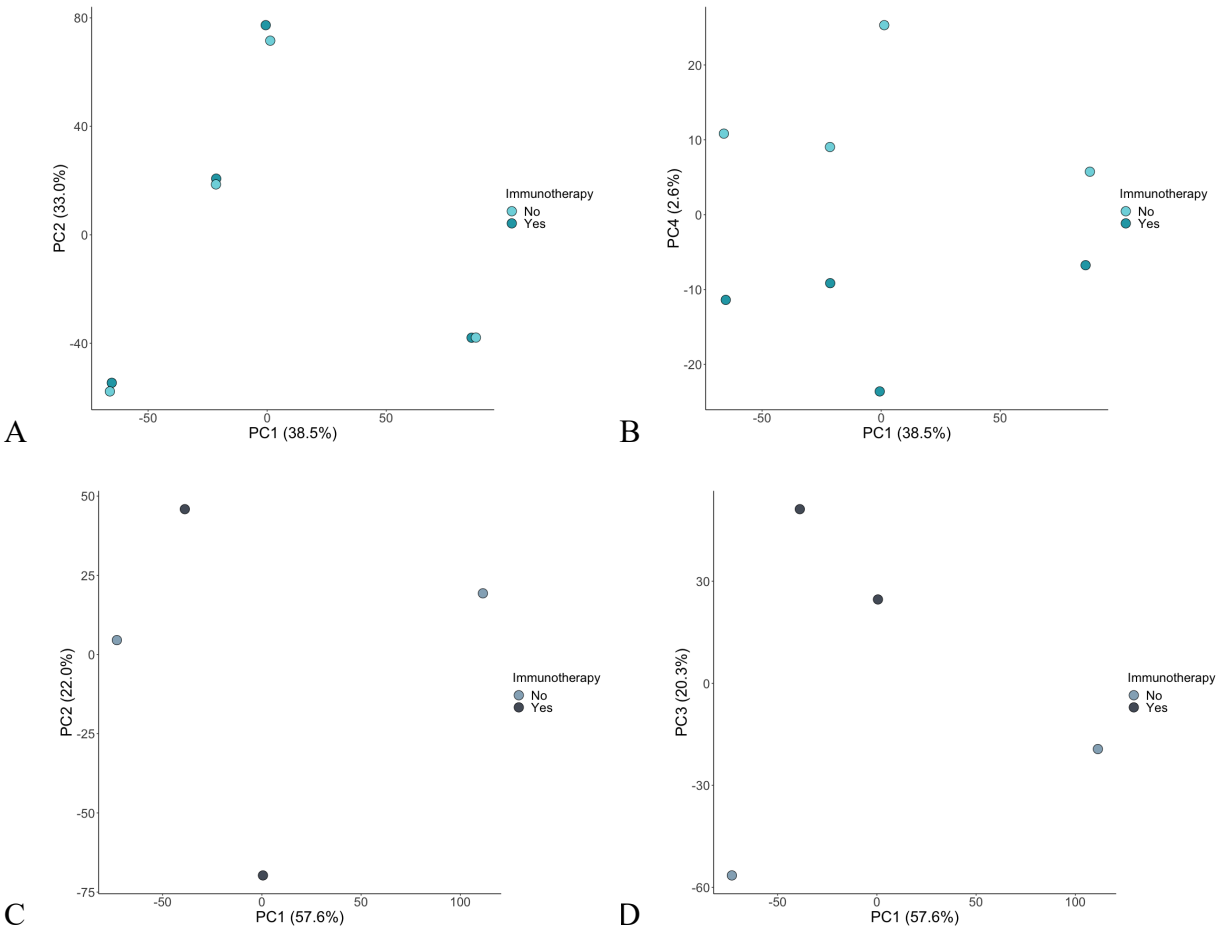


Figure 4-30: PCAs exploring the transcriptomic signatures associated with immunotherapy treatment in USARC. (A&B) PCA analyses of four USARC cell lines before and after treatment with BO112. PC4 identifies a signature associated with immunotherapy treatment. (C&D) PCA analyses of paired USARC samples from patients before and after treatment with immunotherapy. PC3 identifies a transcriptomic signature associated with immunotherapy treatment.

Top 50 Genes (PC4)

FAM156B	DIO2	GEM	CGB8	SLX1B
GAGE12C	ANKRD44	TOX	ZMAT3	DENND5B
GAGE12D	MICU3	FMN2	VGF	EPHX2
RGPD5	PDCD11	ZSCAN31	ATOH8	DLX2
SCHIP1	TLL1	TRIM2	GTF2IRD2	NDRG2

UBE3D	OXTR	CDNF	RDH10	ACTR3B
PYROXD2	FHIT	DSEL	CEP126	L3MBTL1
SPANXC	STK32A	EHHADH	ZNF573	KLF5
SMAD6	NOG	KRT80	TCF7L2	GPR87
EVI2A	ARL10	MBNL2	LURAP1L	ZFHX4

Bottom 50 Genes (PC4)				
IFIT2	GBP4	OAS1	APOL1	MX2
OASL	CXCL11	MX1	HERC5	APOL6
RSAD2	GBP5	USP18	CCL2	BATF2
IFIT1	IFIH1	CCL3L3	SAMD9	APOL2
CXCL10	OAS2	CCL3	CCL20	RAET1L
CCL5	GBP1	RTP4	ISG15	CFB
IFIT3	TRIM22	XAF1	CXCL8	ICAM1
IFNL1	ISG20	DHX58	PLEKHA4	THEMIS2
IFNB1	ZC3HAV1	HELZ2	GAGE12B	CMPK2
DDX58	SAMD9L	IDO1	TNFSF10	TRANK1

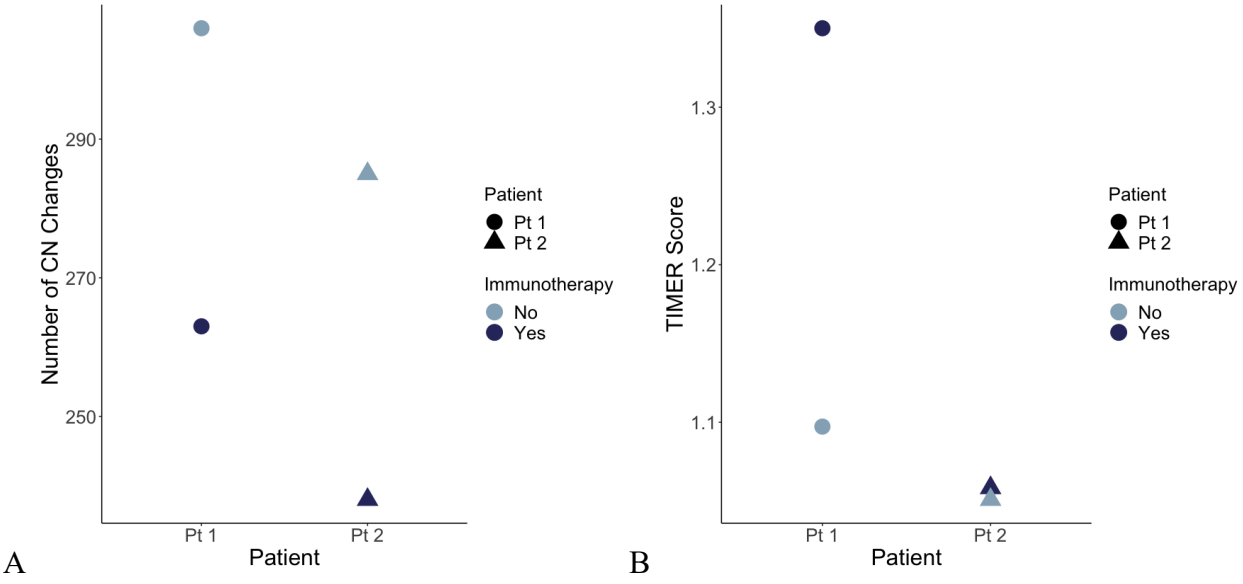
Table 4-21: Top 50 and bottom 50 genes on PC4 on cell line PCA. This appears to be a gene signature associated with BO112 treatment in USARC cell lines.

Top 50 Genes (PC3)				
SFTPB	HBB	CCL5	PIGR	FCN1
HBA1	JCHAIN	CST7	S100A12	VCY
CXCL13	CCL18	HBA2	S100A8	ITM2C
SFTPA2	SLC34A2	ECSCR	HCST	EGFL7
MMP13	RUNX3	CXCL9	C7	MPZL3
MT1A	GZMA	CCL17	ITGA1	CACNA2D2
MMP9	CD27	SCGB3A2	RHOF	SFTPA1
GZMK	UBD	CRTAM	CD3E	PDCD1
CXorf49	SFTPC	PGC	MT1M	SLAMF6
SCGB1A1	NKG7	TNFRSF9	GBP4	ABCA3

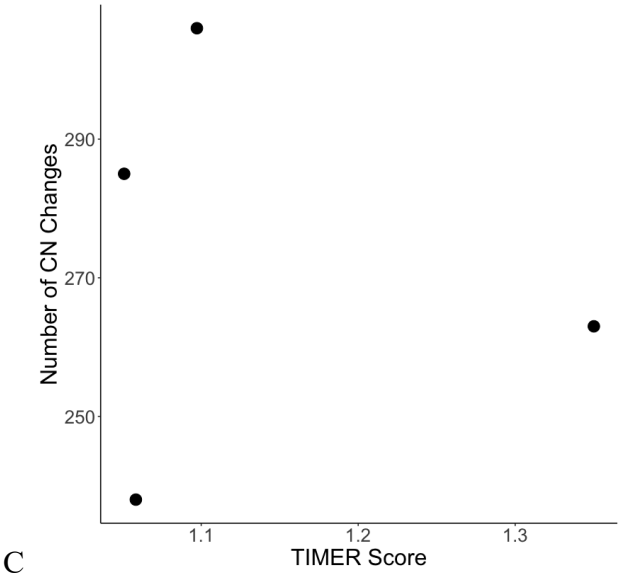
Bottom 50 Genes (PC3)				
TNNT3	ACTN2	TNNI1	SCN5A	NPIPA3
MYL1	ACTA1	NRAP	DNAH8	SYNC
MYH1	MYBPC1	MB	CGB8	STAC3
CKM	FUT5	MYLPF	MYBPC2	BNIP3

TNNC2	TTN	TRIM63	LUC7L2	SH2D5
TNNI2	COX6A2	CCDC169	CSRP3	GLIPR1L1
THBS4	TNNT1	TRPV3	ANKRD1	PAX8
DES	TMOD4	C1orf53	BPI	MT-ATP8
TBC1D3G	MYH2	MYBPH	S100A1	CDC25C
NEB	EEF1A2	IGFN1	HSPB7	C4orf51

Table 4-22: Top 50 and bottom 50 genes on PC3 on PCA of tumors treated with and without immunotherapy. This appears to be a gene signature associated with immunotherapy treatment in USARC tumors.



A B



C

Figure 4-31: An analysis of paired samples from two patients with and without treatment with immunotherapy. (A) Lower copy number segment changes are seen in post-immunotherapy samples. (B) Higher TIMER scores are seen in samples from patients previously treated with immunotherapy compared to samples from the same patient before they were treated with immunotherapy. (C) There is no clear correlation between TIMER score and CN segment changes.

4.10 Bibliography

1. Tawbi HA, Burgess M, Bolejack V, et al. Pembrolizumab in advanced soft-tissue sarcoma and bone sarcoma (SARC028): a multicentre, two-cohort, single-arm, open-label, phase 2 trial. *Lancet Oncol*. 2017;18(11):1493-1501. doi:10.1016/S1470-2045(17)30624-1
2. Pollack SM, Ingham M, Spraker MB, Schwartz GK. Emerging Targeted and Immune-Based Therapies in Sarcoma. *J Clin Oncol Off J Am Soc Clin Oncol*. 2018;36(2):125-135. doi:10.1200/JCO.2017.75.1610
3. Patnaik A, Kang SP, Rasco D, et al. Phase I Study of Pembrolizumab (MK-3475; Anti-PD-1 Monoclonal Antibody) in Patients with Advanced Solid Tumors. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2015;21(19):4286-4293. doi:10.1158/1078-0432.CCR-14-2607
4. Ben-Ami E, Barysauskas CM, Solomon S, et al. Immunotherapy with single agent nivolumab for advanced leiomyosarcoma of the uterus: Results of a phase 2 study. *Cancer*. 2017;123(17):3285-3290. doi:10.1002/cncr.30738
5. Cancer Genome Atlas Research Network. Electronic address: elizabeth.demicco@sinaihealthsystem.ca, Cancer Genome Atlas Research Network. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. *Cell*. 2017;171(4):950-965.e28. doi:10.1016/j.cell.2017.10.014
6. Am G, S K, L B, et al. Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. *Mol Cancer Ther*. 2017;16(11). doi:10.1158/1535-7163.MCT-17-0386
7. Tijhuis AE, Johnson SC, McClelland SE. The emerging links between chromosomal instability (CIN), metastasis, inflammation and tumour immunity. *Mol Cytogenet*. 2019;12(1):17. doi:10.1186/s13039-019-0429-1

8. Li J, Duran MA, Dhanota N, et al. Metastasis and Immune Evasion from Extracellular cGAMP Hydrolysis. *Cancer Discov*. 2021;11(5):1212-1227. doi:10.1158/2159-8290.CD-20-0387
9. Bakhoum SF, Ngo B, Laughney AM, et al. Chromosomal instability drives metastasis through a cytosolic DNA response. *Nature*. 2018;553(7689):467-472. doi:10.1038/nature25432
10. Davoli T, Uno H, Wooten EC, Elledge SJ. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science*. 2017;355(6322). doi:10.1126/science.aaf8399
11. Vivian J, Rao AA, Nothaft FA, et al. Toil enables reproducible, open source, big biomedical data analyses. *Nat Biotechnol*. 2017;35(4):314-316. doi:10.1038/nbt.3772
12. Abeshouse A, Adebamowo C, Adebamowo SN, et al. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. *Cell*. 2017;171(4):950-965.e28. doi:10.1016/j.cell.2017.10.014
13. Steele CD, Tarabichi M, Oukrif D, et al. Undifferentiated Sarcomas Develop through Distinct Evolutionary Pathways. *Cancer Cell*. 2019;35(3):441-456.e8. doi:10.1016/j.ccell.2019.02.002
14. Lesluyes T, Pérot G, Largeau MR, et al. RNA sequencing validation of the Complexity INdex in SARComas prognostic signature. *Eur J Cancer Oxf Engl 1990*. 2016;57:104-111. doi:10.1016/j.ejca.2015.12.027
15. UCSC Xena. Accessed September 30, 2022. https://xenabrowser.net/datapages/

16. Sturm G, Finotello F, Petitprez F, et al. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics*. 2019;35(14):i436-i445. doi:10.1093/bioinformatics/btz363
17. Li B, Severson E, Pignon JC, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol*. 2016;17(1):174. doi:10.1186/s13059-016-1028-7
18. Rank Rank Hypergeometric Overlap (RRHO) - Simple Version. Accessed September 30, 2022. https://systems.crump.ucla.edu/rankrank/rankranksimple.php
19. GSEA. Accessed September 30, 2022. https://www.gsea-msigdb.org/gsea/index.jsp
20. GSEA | MSigDB | Human MSigDB Collections. Accessed September 30, 2022. https://www.gsea-msigdb.org/gsea/msigdb/human/collections.jsp#H
21. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646-674. doi:10.1016/j.cell.2011.02.013
22. Richard C, Fumet JD, Chevrier S, et al. Exome Analysis Reveals Genomic Markers Associated with Better Efficacy of Nivolumab in Lung Cancer Patients. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2019;25(3):957-966. doi:10.1158/1078-0432.CCR-18-1940
23. Rosenthal R, Cadieux EL, Salgado R, et al. Neoantigen-directed immune escape in lung cancer evolution. *Nature*. 2019;567(7749):479-485. doi:10.1038/s41586-019-1032-7
24. Goodman AM, Kato S, Bazhenova L, et al. Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. *Mol Cancer Ther*. 2017;16(11):2598-2608. doi:10.1158/1535-7163.MCT-17-0386
25. Le DT, Uram JN, Wang H, et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med*. 2015;372(26):2509-2520. doi:10.1056/NEJMoa1500596

26. Selmecki AM, Maruvka YE, Richmond PA, et al. Polyploidy can drive rapid adaptation in yeast. *Nature*. 2015;519(7543):349-352. doi:10.1038/nature14187
27. Bassaganyas L, Pinyol R, Esteban-Fabró R, et al. Copy-Number Alteration Burden Differentially Impacts Immune Profiles and Molecular Features of Hepatocellular Carcinoma. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2020;26(23):6350-6361. doi:10.1158/1078-0432.CCR-20-1497
28. Gyorffy B, Bottai G, Nagy A, Pusztai L, Santarpia L. Immune gene signatures in triple-negative breast cancers characterized by varying levels of chromosomal instability. *J Clin Oncol*. 2017;35(15_suppl):1096-1096. doi:10.1200/JCO.2017.35.15_suppl.1096
29. Anagnostou V, Smith KN, Forde PM, et al. Evolution of Neoantigen Landscape during Immune Checkpoint Blockade in Non-Small Cell Lung Cancer. *Cancer Discov*. 2017;7(3):264-276. doi:10.1158/2159-8290.CD-16-0828
30. Woo SR, Fuertes MB, Corrales L, et al. STING-dependent cytosolic DNA sensing mediates innate immune recognition of immunogenic tumors. *Immunity*. 2014;41(5):830-842. doi:10.1016/j.immuni.2014.10.017
31. Wu T, Ji Y, Moseman EA, et al. The TCF1-Bcl6 axis counteracts type I interferon to repress exhaustion and maintain T cell stemness. *Sci Immunol*. 2016;1(6):eaai8593. doi:10.1126/sciimmunol.aai8593
32. Santaguida S, Richardson A, Iyer DR, et al. Chromosome Mis-segregation Generates Cell-Cycle-Arrested Cells with Complex Karyotypes that Are Eliminated by the Immune System. *Dev Cell*. 2017;41(6):638-651.e5. doi:10.1016/j.devcel.2017.05.022
33. Senovilla L, Vitale I, Martins I, et al. An immunosurveillance mechanism controls cancer cell ploidy. *Science*. 2012;337(6102):1678-1684. doi:10.1126/science.1224922

34. Toulmonde M, Lucchesi C, Verbeke S, et al. High throughput profiling of undifferentiated pleomorphic sarcomas identifies two main subgroups with distinct immune profile, clinical outcome and sensitivity to targeted therapies. *eBioMedicine*. 2020;62. doi:10.1016/j.ebiom.2020.103131
35. Budczies J, Denkert C, Győrffy B, Schirmacher P, Stenzinger A. Chromosome 9p copy number gains involving PD-L1 are associated with a specific proliferation and immune-modulating gene expression program active across major cancer types. *BMC Med Genomics*. 2017;10(1):74. doi:10.1186/s12920-017-0308-8
36. The Cancer Genome Atlas - Molecular Characterization Platforms - NCI. Published March 6, 2019. Accessed October 5, 2022. https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga/using-tcga/technology
37. Cohen N, Dagan T, Stone L, Graur D. GC Composition of the Human Genome: In Search of Isochores. *Mol Biol Evol*. 2005;22(5):1260-1272. doi:10.1093/molbev/msi115
38. Piovesan A, Pelleri MC, Antonaros F, Strippoli P, Caracausi M, Vitale L. On the length, weight and GC content of the human genome. *BMC Res Notes*. 2019;12(1):106. doi:10.1186/s13104-019-4137-z
39. Romiguier J, Roux C. Analytical Biases Associated with GC-Content in Molecular Evolution. *Front Genet*. 2017;8. Accessed September 30, 2022. https://www.frontiersin.org/articles/10.3389/fgene.2017.00016
40. Delhomme N, Mähler N, Schiffthaler B, Sundell D, Hvidsten TR, Street NR. Guidelines for RNA-Seq data analysis. :24.

41. Shelton M, Metz J, Liu J, et al. Derivation and Expansion of PAX7-Positive Muscle Progenitors from Human and Mouse Embryonic Stem Cells. *Stem Cell Rep*. 2014;3(3):516-529. doi:10.1016/j.stemcr.2014.07.001
42. Chal J, Oginuma M, Al Tanoury Z, et al. Differentiation of pluripotent stem cells to muscle fiber to model Duchenne muscular dystrophy. *Nat Biotechnol*. 2015;33(9):962-969. doi:10.1038/nbt.3297
43. Xi H, Fujiwara W, Gonzalez K, et al. In Vivo Human Somitogenesis Guides Somite Development from hPSCs. *Cell Rep*. 2017;18(6):1573-1585. doi:10.1016/j.celrep.2017.01.040
44. Steele CD, Abbasi A, Islam SMA, et al. Signatures of copy number alterations in human cancer. *Nature*. 2022;606(7916):984-991. doi:10.1038/s41586-022-04738-6

CHAPTER 5

Conclusion

5.1 Summary of Research

Soft tissue sarcoma (STS) is a rare and diverse group of malignancies with over 60 subtypes, presumed to be of mesenchymal origin. Patients with localized STS are managed with surgery with or without radiation[1–3]. Despite aggressive local therapy, up to 50% of patients with high-risk (≥ 5cm, high-grade) primary STS develop metastases, highlighting the critical need for more effective systemic therapies[4–11].

Immune-based therapies have dramatically changed the treatment of many cancers in recent decades. These include blockade of immune checkpoint molecules such as PD-1/PD-L1 and CTLA-4, cellular therapies using chimeric antigen receptor or T-cell receptor modified T-cells, and emerging therapies modulating tumor-associated macrophages. While early studies of immunotherapy in STS have not had broad success, there are signals of efficacy that warrant further exploration[12–15]. The SARC028 study evaluated the efficacy of the anti-PD1 antibody pembrolizumab in patients with advanced sarcoma, and identified the undifferentiated pleomorphic sarcoma (UPS) subtype as the most responsive to anti-PD1 immunotherapy, with 40% of patients demonstrating complete or partial response[14]. This was notable given that UPS, unlike most tumors that are most responsive to immune checkpoint blockade, have a low mutational burden.

Why a fraction of patients with UPS and other tumors with low mutational burden instigate an anti-tumor immune response is unclear. At the same time, 60% of patients with UPS in this study did not respond to anti-PD1 therapy, suggesting that within UPS there are drivers of immune evasion.

In order to stratify the patients with STS that may benefit from immunotherapy, we need alternate strategies to study sarcoma in a high throughput fashion. For STS, alternatives to categorizing patients based on "response" versus "non-response" to immunotherapy are needed, as this disease is extremely rare and large studies of immune checkpoint blockade are not readily available as in other disease[12–21]. In order to guide further studies and to develop optimal immunotherapy strategies for patients with STS, we must first understand the landscape of immune cell infiltration and the associated genomic changes associated with immune evasion in this disease.

In this study, I aimed to investigate the immunologic heterogeneity and identify transcriptomic and genomic correlates of immune evasion in undifferentiated sarcoma. In doing so, I determined the immune cell landscape, the optimal high-throughput tools, and the transcriptomic and genomic changes associated with high and low immune cell infiltration in STS.

The specific aims and primary findings were as follows:

Specific Aim 1: Characterize the landscape of immune cell infiltration in soft tissue sarcoma. My goal was to determine the overall immune cell landscape and the association between clinical and treatment factors and immune cell infiltration in STS. (Project 1, Chapter 2)

In this aim, we characterized the landscape of immune cell infiltration in soft tissue sarcoma and determined the association between clinical and treatment factors and immune cell infiltration in soft tissue sarcoma. We found that USARC tumors are characterized by a myeloid predominance and a relative abundance of suppressor cells, such as Treg cells and CD11b cells. We found similar trends when comparing STS subtypes with high levels of copy number alterations overall (DDLPS, USARC, and MFS; n=67) versus those that are characterized as having low levels of copy number alterations (MLS, LMS, SS, and GIST) as CNA-low (n=38)[22]. We additionally found that the immune composition of peripheral blood was associated with intratumoral leukocyte infiltration, and specifically that myeloid-predominant tumor and lymphocyte-predominant blood are mutually exclusive.

Specific Aim 2: Develop and apply an optimal in-silico immune deconvolution technique for sarcoma. My goal was to assess the concordance between various immune deconvolution techniques and compare their ability to recapitulate the results of immunohistochemistry on paired USARC specimens. (Project 2, Chapter 3)

In this aim, we determined the optimal in-silico immune deconvolution tool in undifferentiated sarcoma by determining the correlation between mIF and in-silico immune deconvolution scores. Based on our findings, we suggest the following practices when applying in-silico immune

deconvolution tools to undifferentiated sarcoma: (1) Use TIMER to define overall immune cell infiltration. (2) Use MCP counter to define monocyte infiltration or use CIBERSORTx, EPIC, quanTIseq, TIMER, or xCell to define macrophage infiltration. (3) Use caution when using in-silico immune deconvolution tools to define CD8+ T cell infiltration. CIBERSORTx most accurately defines CD8+ T cell immune infiltration, however, there are still many instances when tumors with high CD8+ T cell infiltration will be missed using this technique. (4) Avoid applying in-silico immune deconvolution results to define B cell or CD4+ T cell immune infiltration.

Specific Aim 3: Define transcriptomic signatures and genomic changes associated with immune cell infiltration in undifferentiated sarcoma. I aimed to determine the role of chromosomal instability in immune cell infiltration in this disease. (Project 3, Chapter 4)

In this aim, we found that increased copy number changes were associated with low immune cell infiltration in undifferentiated sarcoma. These findings were suggested in both transcriptomic and genomic analyses. Interestingly, this association between CNA and immune invasion were unique to the UPS and DDLPS subtypes of STS, but it was not seen in other subtypes of STS. The mechanisms underlying this association are not clear and warrant further study.

These insights provide necessary information to understand which patients may benefit from immunotherapy and guide future studies to further the treatment of STS. These studies provide the groundwork for further investigation in this study of immune cell infiltration in STS and provide insights into how we may be able to improve outcomes in this rare and devastating

disease. The mechanisms underlying these findings remain unclear and warrant further investigation. A deeper understanding of the drivers of immune cell infiltration, the unique tumor microenvironment in STS, and role that chromosomal instability plays in STS will hopefully ultimately lead to insights to new, and much-needed, treatments for this disease.

5.2 Bibliography

1. Cancer Facts & Figures 2017. Accessed May 5, 2021. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2017.html
2. Sarcomas, Soft Tissue - Statistics. Cancer.Net. Published June 25, 2012. Accessed May 5, 2021. https://www.cancer.net/cancer-types/sarcomas-soft-tissue/statistics
3. Ferguson PC, Deheshi BM, Chung P, et al. Soft tissue sarcoma presenting with metastatic disease: outcome with primary surgical resection. *Cancer*. 2011;117(2):372-379. doi:10.1002/cncr.25418
4. Brennan MF, Antonescu CR, Moraco N, Singer S. Lessons learned from the study of 10,000 patients with soft tissue sarcoma. *Ann Surg*. 2014;260(3):416-421; discussion 421-422. doi:10.1097/SLA.0000000000000869
5. Trovik CS, Bauer HC, Alvegård TA, et al. Surgical margins, local recurrence and metastasis in soft tissue sarcomas: 559 surgically-treated patients from the Scandinavian Sarcoma Group Register. *Eur J Cancer Oxf Engl 1990*. 2000;36(6):710-716. doi:10.1016/s0959-8049(99)00287-7
6. Rööser B, Gustafson P, Rydholm A. Is there no influence of local control on the rate of metastases in high-grade soft tissue sarcoma? *Cancer*. 1990;65(8):1727-1729. doi:10.1002/1097-0142(19900415)65:8<1727::aid-cncr2820650811>3.0.co;2-x
7. Gustafson P. Soft tissue sarcoma. Epidemiology and prognosis in 508 patients. *Acta Orthop Scand Suppl*. 1994;259:1-31.

8. Gustafson P, Rööser B, Rydholm A. Is local recurrence of minor importance for metastases in soft tissue sarcoma? *Cancer*. 1991;67(8):2083-2086. doi:10.1002/1097-0142(19910415)67:8<2083::aid-cncr2820670813>3.0.co;2-5
9. Alamanda VK, Crosby SN, Archer KR, Song Y, Schwartz HS, Holt GE. Predictors and clinical significance of local recurrence in extremity soft tissue sarcoma. *Acta Oncol Stockh Swed*. 2013;52(4):793-802. doi:10.3109/0284186X.2012.711953
10. Weitz J, Antonescu CR, Brennan MF. Localized extremity soft tissue sarcoma: improved knowledge with unchanged survival over time. *J Clin Oncol Off J Am Soc Clin Oncol*. 2003;21(14):2719-2725. doi:10.1200/JCO.2003.02.026
11. Zaidi MY, Ethun CG, Tran TB, et al. Assessing the Role of Neoadjuvant Chemotherapy in Primary High-Risk Truncal/Extremity Soft Tissue Sarcomas: An Analysis of the Multi-institutional U.S. Sarcoma Collaborative. *Ann Surg Oncol*. 2019;26(11):3542-3549. doi:10.1245/s10434-019-07639-7
12. Pollack SM, Ingham M, Spraker MB, Schwartz GK. Emerging Targeted and Immune-Based Therapies in Sarcoma. *J Clin Oncol Off J Am Soc Clin Oncol*. 2018;36(2):125-135. doi:10.1200/JCO.2017.75.1610
13. Patnaik A, Kang SP, Rasco D, et al. Phase I Study of Pembrolizumab (MK-3475; Anti-PD-1 Monoclonal Antibody) in Patients with Advanced Solid Tumors. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2015;21(19):4286-4293. doi:10.1158/1078-0432.CCR-14-2607
14. Tawbi HA, Burgess M, Bolejack V, et al. Pembrolizumab in advanced soft-tissue sarcoma and bone sarcoma (SARC028): a multicentre, two-cohort, single-arm, open-label, phase 2 trial. *Lancet Oncol*. 2017;18(11):1493-1501. doi:10.1016/S1470-2045(17)30624-1

15. Ben-Ami E, Barysauskas CM, Solomon S, et al. Immunotherapy with single agent nivolumab for advanced leiomyosarcoma of the uterus: Results of a phase 2 study. *Cancer*. 2017;123(17):3285-3290. doi:10.1002/cncr.30738
16. Gandhi L, Rodríguez-Abreu D, Gadgeel S, et al. Pembrolizumab plus Chemotherapy in Metastatic Non–Small-Cell Lung Cancer. *N Engl J Med*. 2018;378(22):2078-2092. doi:10.1056/NEJMoa1801005
17. Mok TSK, Wu YL, Kudaba I, et al. Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial. *The Lancet*. 2019;393(10183):1819-1830. doi:10.1016/S0140-6736(18)32409-7
18. Cramer-van der Welle CM, Verschueren MV, Tonn M, et al. Real-world outcomes versus clinical trial results of immunotherapy in stage IV non-small cell lung cancer (NSCLC) in the Netherlands. *Sci Rep*. 2021;11(1):6306. doi:10.1038/s41598-021-85696-3
19. Middleton G, Brock K, Savage J, et al. Pembrolizumab in patients with non-small-cell lung cancer of performance status 2 (PePS2): a single arm, phase 2 trial. *Lancet Respir Med*. 2020;8(9):895-904. doi:10.1016/S2213-2600(20)30033-3
20. Robert C, Ribas A, Schachter J, et al. Pembrolizumab versus ipilimumab in advanced melanoma (KEYNOTE-006): post-hoc 5-year results from an open-label, multicentre, randomised, controlled, phase 3 study. *Lancet Oncol*. 2019;20(9):1239-1251. doi:10.1016/S1470-2045(19)30388-2
21. Pembrolizumab versus Ipilimumab in Advanced Melanoma | NEJM. Accessed May 23, 2021. https://www.nejm.org/doi/full/10.1056/nejmoa1503093

22. Abeshouse A, Adebamowo C, Adebamowo SN, et al. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. *Cell*. 2017;171(4):950-965.e28. doi:10.1016/j.cell.2017.10.014