Differential Equation Modeling of Cell Population Dynamics in Skeletal Muscle Regeneration from Single-Cell Transcriptomic Data

by

Renad Al-Ghazawi

Thesis submitted to the University of Ottawa in partial fulfilment of the requirements for the degree of Master of Science in Cellular and Molecular Medicine with Specialization in Bioinformatics

Department of Cellular and Molecular Medicine
Faculty of Medicine
University of Ottawa

Abstract

Research in muscle stem cells (MuSCs) has consistently aimed to enhance our understanding of the complex dynamics within muscle tissues. However, traditional computational models for muscle regeneration often rely on aggregated data, potentially missing the subtle complexities of MuSCs and their cellular niches. The emergence of single-cell RNA sequencing (scRNA-seq) data offers a significant opportunity to refine these models and enhance our understanding of cellular interactions in muscle regeneration. In this study, we introduce a non-linear ordinary differential equation model, consisting of 10 variables and 22 parameters, to precisely capture the dynamics of key cell types involved in skeletal muscle repair. Focusing on myogenic lineage cells and immune cells, the model delineates critical cell fate decisions such as quiescence, activation, proliferation, differentiation, infiltration, apoptosis, and exfiltration in response to muscle damage and intercellular communication. We adapted this model to scRNA-seq data, from which we generated and refined population dynamic datasets. These datasets were further refined with supplementary experimental data, allowing for the determination of precise cell counts per tissue volume (cells/mm^3). The robustness of our model is validated by its ability to mirror the observed population dynamics. Additionally, a thorough sensitivity analysis was performed, involving several perturbations of parameters within a range of ±10% of their baseline values to highlight their influence and validate the reliability of our model.

Acknowledgements

First and foremost, I would like to express my deep appreciation to my supervisor, Dr. Theodore J. Perkins, and co-supervisor Dr. Xiaojian Shao for their continuous guidance, enthusiasm, moral and financial support. They provided me with the intellectual freedom to explore my research, supported my participation in various conferences, and consistently motivated me to maintain high standards in my work. I would also like to thank members of the Perkins lab, and OHRI members which includes but not limited to, Praveen Guttula, Daniel Gladish, Aarthie Senathirajah, Naveen Ahuja, Saber Saharkhiz, Enzhe Galimova, Madeeha Shaikh, Masoud Karimi, Soroush Fard, Satrajit Chatterjee, Rohan Ray, Gareth Palidwor, and Christopher Porter. Special thank you to Justin G. Chitpin for his warm welcome and for encouraging me to step outside of my academic routine. Our friendly games of pool and cafe work sessions were a welcome respite from navigating graduate school.

I extend my gratitude to thesis committee members, Dr. Mads Kaern and Dr. Jeffrey Dilworth, for taking an interest in my project and for their invaluable expertise which has helped improve the quality of my work, for which I am sincerely thankful.

To Adeeb, I owe a special note of gratitude. I am truly grateful for your genuine engagement in my work and your resolute belief in my potential. Your consistent support and affirmations have not only pushed me beyond my perceived limits but also instilled my academic journey with confidence.

There are no words that could adequately express how thankful I am for my family. The completion of this project is thanks to my dear parents, Zainah and Mustafa, for instilling in me the values of perseverance and integrity, and for their countless sacrifices that made my education possible.

My older siblings, Feras, Sawsan, Nebras, Wesam, and Ahmed, thank you for being my pillars of comfort and constant source of support. Mohammed, my younger brother, and my best friend, thank you especially, for being the person I can always count on and my biggest supporter.

Dedication

To

My dear father, in loving memory,

all my efforts, I always dedicate to you.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Activated Satellite Cells	ASCs
Agent-Based Models	ABMs
Batch Balanced K Nearest Neighbours	BBKNN
Cardiotoxin	CTX
Canonical Correlation Analysis	CCA
Checkpoint Kinases	CHKs
Complementary DNA	cDNA
Cyclin-Dependent Kinase 1	Cdk1
Cyclin-Dependent Kinases	Cdks
Cyclin-Dependent Kinase Inhibitor 1A	Cdkn1A
Cyclin-Dependent Kinase Inhibitor 1B	Cdkn1B
Cyclin-Dependent Kinase Inhibitor 1C	Cdkn1C
Days Post-Injury	DPI
Differentially Expressed Genes	DEGs
Deoxyribonucleic Acid	DNA
Duchenne Muscular Dystrophy	DMD
Extracellular Matrix	ECM
Fibro/Adipogenic Progenitors	FAPs
Fluorescence Activated Cell Sorting	FACS
Freshly Isolated Satellite Cells	FISCs
Hematopoietic stem cells	HSCs
Hepatocyte Growth Factor	HGF
Human Induced Pluripotent Stem Cells	hiPSCs
Immunofluorescence	IF
Immunomyoblasts	IMB
K-nearest Neighbor	KNN
Myogenic Precursor Cells	MPC
Myogenic Regulatory Factors	MRFs
Myoblast Determination Protein 1	Myod1
Myogenic Factor 5	Myf5
Myogenin	Myog
Myogenic Factor 6	Myf6
Muscle Satellite Cells	MuSCs
Muscular Dystrophy X-Linked	mdx
Myosin Heavy Chain	MYH
Myosin Light Chain 4	Myl4
Nonlinear Least Squares	NLLS
Notexin	NTX
Principal Component Analysis	PCA
Ordinary Differential Equations	ODE
Platelet-Derived Growth Factor	PDGF
Principal Component Analysis	PCA
Quiescent Satellite Cells	QSCs

Reactive oxygen species	ROS
Ribonucleic acid	RNA
Satellite Cells	SCs
Standard Deviation	SD
Self-renewing Quiescent Satellite Cells	SR-QSC
Shared Nearest Neighbor	SNN
Single Cell RNA Sequencing	scRNA-seq
Tibialis Anterior	TA
t-distributed Stochastic Neighbor Embedding	t-SNE
Total Sum of Squares	TSS
Uniform Manifold Approximation and Projection	UMAP
Vascular Endothelial Growth Factor	VEGF

Chapter 1

Introduction

1.1 Computational Biology in Muscle Stem Cell Research

The intersection of biology with computational and mathematical disciplines has significantly deepened our understanding of complex biological systems. Since the inception of modern computing in the 1960s, computational techniques were initially harnessed to interpret complex biochemical processes and manage growing experimental datasets[1]. As those techniques evolved, researchers have gained a more comprehensive perspective in specific areas such as the stem cells in skeletal muscle tissue. The muscle stem cell (MuSC) niche encapsulates a dynamic interplay between maintaining quiescence during tissue homeostasis while simultaneously facilitating regeneration[2,3]. More specifically, these cells engage in communication strategies with neighboring cell populations to allow the exchange of biological information, and the reciprocal dialogue among these cells is critical for the remodeling of tissue. Yet, various extraneous factors, ranging from the natural process of aging to specific pathologies in skeletal muscle diseases, can compromise the integrity of this niche.

These factors impact the regenerative capacity of MuSCs, leading to impaired muscle repair and function. Consequently, understanding the niche's response to such disruptions is necessary for developing therapeutic strategies for muscle-related disorders. Therefore, the complex nature of these interactions necessitates the need for innovative computational methodologies capable of integrating and analyzing such interactions these cellular dynamics.

Recent advances in single-cell technology provide us with the opportunity to deconvolute the heterogeneity of muscle tissue into individual cells enabling the characterization of dynamic interactions among cell populations at various development stages[4–6]. This technique enables the identification of rare or small cell types, their dynamic trajectories, and molecular programs governing sequential cell fates in skeletal muscle development[7]. While there is an abundance of information regarding muscle cells themselves, the precise modeling of muscle stem cells and their unique niches remains an area with much left to uncover. Current MuSC computational models have predominantly based their simulations and parameter estimations on datasets from multiple sources[8–11]. Such a practice, while advantageous in broadening the scope of the models, introduces inherent biases associated with pooling data. These biases can emerge from differences in experimental procedures, variations in sample preparation and handling, inconsistencies in data processing methodologies, and even divergences in the biological contexts from which the data originated. Moreover, when combining datasets, there is a risk of oversimplifying or generalizing unique cellular behaviors, thereby potentially missing out on specific nuances that would be revealed by a more unified dataset. Ideally, computational models would accurately capture the full spectrum of interactions and stages of muscle cell development, including infiltration,

differentiation, and exit. However, the inherent complexities of MuSC biological systems, specifically in describing their population dynamics remain considerably challenging. Therefore, to overcome these limitations, it is required to employ a refined modeling approach that accurately represents the behavior of the MuSC environment.

1.2 Utility of Mathematical Modeling in Cell Biology

With the growing accumulation of large volumes of biological data, there is an increasing necessity for mechanistic mathematical modeling, specifically targeting complex biological systems at the population level. Detailed dynamical models that describe the multifaceted nature of cellular behaviour, require computational methods to handle a significant number of undefined parameters including rate constants and saturation values from interacting cells. Even in cases where the parameters are known due to prior experiments or clinical data, variations stemming from different experimental designs often lead to computational deviations[14]. Hence, calibrating models based on direct empirical data need to be investigated to avoid these inaccuracies.

The mathematical frameworks of MuSCs and their dynamic behaviours found in the literature differ greatly in complexity, and hence the applicability of the models also varies. While complex models provide a detailed picture of a system, simpler models with reduced complexity enable for more direct comparisons with clinical data. Such simpler models allow for comparison with data which would have been inconclusive when based on the complex model[15]. This process of simplification is instrumental in identifying the specific data type required to differentiate among various intricate models. Consequently, the steps of model downsizing, refinement, and subsequent comparison against empirical data not only

emphasizes the model's relevance but also lays the groundwork for ensuing experimental pursuits to validate the foundational hypotheses.

In this dissertation, I introduce a mechanistic mathematical model specifically focused on the population dynamics of MuSCs and key cell types within their niche. The model is parameterized based on the analysis of time-series scRNA-seq data and its dynamics are related to experimental observations from the literature. It accounts for cellular processes such as exit from quiescence, stem cell proliferation, self-renewal, and differentiation. The focus of this study is to understand how these processes affect the temporal evolution of injured muscle cell population.

1.3 Problem Statement

In this thesis, I explore the dynamic behaviors in MuSCs and their interactions during muscle tissue regeneration. The aim is to construct a non-linear ordinary differential equation (ODE) model, constrained to scRNA-seq data offering a more comprehensive depiction of MuSC population dynamics. This approach is designed to provide a thorough and reliable quantitative analysis of MuSC behaviors and interactions, enhancing our understanding of the cellular processes that drive muscle repair and regeneration.

1.4 Hypothesis and Objectives

The hypothesis of this dissertation is that scRNA-seq can reveal detailed dynamics of MuSCs, which can be effectively modeled using advanced computational methods. The following objectives and contributions highlight the key aspects of this research.

1. **Data Acquisition, Annotation, and Single-Cell Analysis Techniques:**

    - **Objective:** To acquire and analyze multiple time-series scRNA-seq datasets from injured mice, focusing on the annotation of the cellular niche and calibration of population dynamics datasets.

    - **Contribution:** I have applied a combined annotation approach and sub-clustering to these datasets, leading to the identification of distinct cell types within regenerating muscle tissue.

2. **Calibration of Data:**

    - **Objective:** To recalibrate transcriptomic data for representing absolute cell counts per cubic meter, using supplemental experimental data as a reference to ensure outputs are grounded in real-world measurements.

    - **Contribution:** I constructed comprehensive population datasets which were recalibrated and aligned with supplemental measurements.

3. **Development and Optimization of the Mathematical Model:**

    - **Objective:** To develop a non-linear ODE model simulating post-injury cell type responses and optimize it using calibrated cell population data.

    - **Contribution:** I constructed a model comprising 10 non-linear ODEs with 22 parameters and optimized it with empirical scRNA-seq data for alignment with observed cellular dynamics.

4. **Validation through Sensitivity Analysis:**

    - **Objective:** To validate the model by replicating observed muscle tissue population dynamics and conduct a sensitivity analysis to identify influential parameters.

    - **Contribution:** The model's validity was demonstrated through successful replication of muscle tissue dynamics, complemented by a sensitivity analysis assessing model robustness and parameter impact.

1.5 Outline of the Thesis

Chapter 2 of the thesis conducts a thorough literature review, providing an overview of muscle stem cell biology, injury models and regeneration outcomes, applications of scRNA-seq within muscle regeneration, and the role of mathematical modeling in biological research. **Chapter 3** presents the re-analysis of the selected transcriptomic datasets, detailing the methodologies employed for data acquisition and processing, and emphasizing the identification and comparative analysis of unique cell types. **Chapter 4** introduces the development and implementation of a non-linear ODE model for simulating muscle tissue regeneration, covering the model's construction, parameterization, and optimization using empirical data. **Chapter 5** concludes the thesis by providing a summary of our findings from **Chapters 3** and **4**. This chapter also assesses the limitations and suggests potential avenues for future research in the field.

Chapter 2

Background

2.1 Skeletal Muscle Anatomy and Physiology

There are three distinct types of muscle tissue in the body: skeletal muscle, cardiac muscle, and smooth muscle. Each serves a unique set of functions; for instance, while cardiac muscle is essential for propelling blood through our veins, skeletal muscle allows our bodies to move and react to our environment[12]. Unique among its counterparts, skeletal muscle is under voluntary control and constitutes nearly half of our body's total mass[13,14] Governed by the neurons of the somatic nervous system, these muscles connect via tendons and other connective tissues, enabling us to deliberately initiate or stop movement[15,16]. Skeletal muscles are a highly organized tissue primarily made up of long cylindrical muscle fibers known as myofibers[12]. These fibers are grouped into bundles known as fascicles, and collectively, multiple fascicles form a whole muscle. Within these myofibers are densely packed myofibrils, formed from long proteins such as actin and myosin filaments, which systematically array to create the sarcomere, the primary unit responsible for muscle contraction (**Figure 2.1**). The

outer layers of the muscle fibers are encased in connective tissues where the endomysium surrounds individual myofibers, the perimysium wraps around fascicles, and the epimysium encases the entire muscle. For the muscle fibers to contract, they must first receive impulses from nerve cells. Each nerve penetrating the epimysium, is generally accompanied by a network of capillaries, an artery and at least one vein. These blood vessels and nerves closely follow the muscle's connective tissue components, ensuring efficient delivery of oxygen and nutrients essential for muscle function and facilitating the removal of metabolic waste. Notably, each myofiber houses several hundred nuclei, termed myonuclei, positioned at the periphery of the fiber between the myofibrils and the sarcolemma. MuSCs or satellite cells (SCs), have been recognized as the predominant source of myonuclei during muscle development and regeneration[17–19]. Every myonucleus originates from a satellite cell that generates myogenic progenitors following division, which subsequently fuse to form new myofibers[20]. The relevance of understanding skeletal muscle extends beyond its role in movement. A deep comprehension of the mechanisms of skeletal muscle regeneration, its niche and behavior are foundational to fully understand the complexities of various myopathies. These range from inherited diseases like Duchenne Muscular Dystrophy (DMD) to muscle-wasting diseases such as cachexia[21–23].

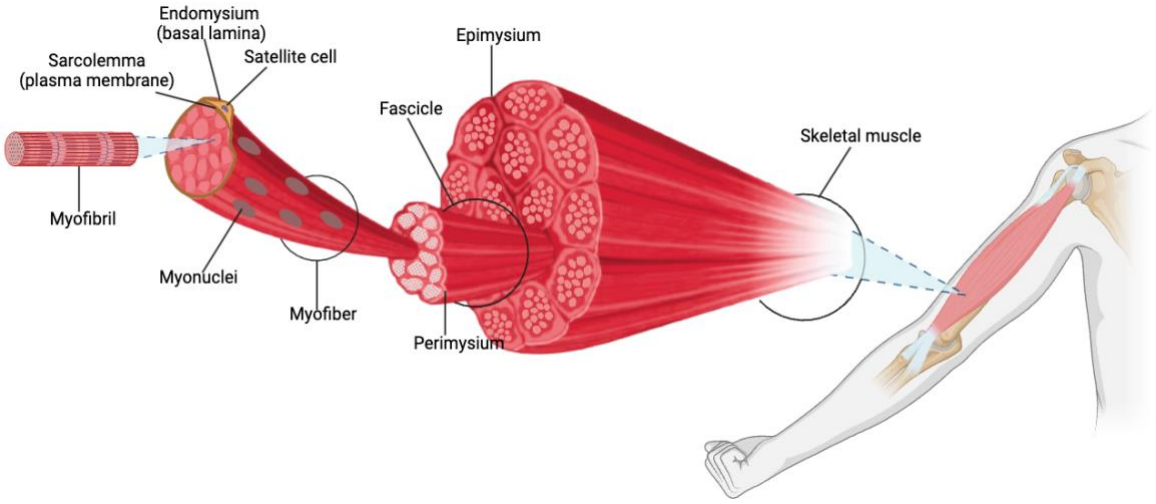


Figure 2.1. The Hierarchical Organization of Skeletal Muscle. The entire muscle is surrounded by the epimysium. Fascicles, or bundles of muscle fibers are encased by the perimysium. Each muscle cell, or myofiber, containing the myofibrillar contractile apparatus, is bounded by the sarcolemma or plasma membrane, and is further encased by the endomysium or basal lamina. MuSCs are positioned outside the myofiber, between the plasma membrane and the basal lamina. Created in BioRender.com.

2.2 Overview of Skeletal Muscle Satellite Cells

The regenerative potential of skeletal muscle tissue is observed across a diverse range of species and remains effective for a significant portion of an organism's life[24]. Initial evidence of this capability was first postulated in 1860[25], but it was not until a century later in 1961 where it was established that multi-nucleated myofibers arise through the fusion of mononucleated myoblasts[21]. This addressed earlier concerns regarding myofibers expanding in size and nuclear content without evident nuclear division within the fiber. However, it was through the use of electron microscopy, Alexander Mauro detected dormant cells positioned on the myofiber's surface, beneath its basement membrane, and provided the first description of the satellite cell[21]. Indeed, accumulated evidence pointed to these cells as the primary myogenic precursors responsible for muscle regeneration. While first discovered in frog muscle, the presence of satellite cells was soon confirmed in rats and is consistent across a variety of vertebrate species[20]. These cells are characterized by a distinct expression of specific

proteins, notably the paired-box transcriptional factor, Pax7 expression which additionally signifies the differences in cell commitment levels[18]. Specifically, high Pax7-expressing cells display reduced commitment relative to their low Pax7 counterparts[26–28]. Satellite cells also express a diverse range of markers that can be used to distinguish specific subpopulations, including Pax3, Myf5, c-Met, syndecan-3/4, Myod1, and many others, emphasizing their heterogeneous nature.

2.3 Myogenic Factors

In adult muscles, the MuSC pool is maintained in a quiescent state under homeostatic conditions. Once exposed to signals from the damaged environment, these cells harbor the potential to activate and differentiate into myocytes or self-renewing cells. Quiescence is generally characterized by the absence of cell cycling, defined as a G0 state[26,27]. This distinguishes them from differentiated and senescent cells, which exist in an irreversible G0 cell cycle state. QSCs express the transcription factor Pax7, yet they lack the expression of Myod1 or Myogenin (Myog). In addition, they are marked by low metabolic activities, protein, and mRNA production. This protective state ensures their longevity and readiness for tissue regeneration. The quiescent state is not merely an inactive phase for these cells but is dynamically regulated. Active maintenance of this state involves the suppression of genes related to the cell cycle such as cyclins, cyclin-dependent kinases (CDKs), and checkpoint kinases (CHKs). This is further supported by high expression levels of CDK inhibitors Cdkn1a, Cdkn1b, and Cdkn1c, which together, inhibit cell cycle progression. Many signaling pathways must be active to maintain a functional stem cell reservoir and tissue capacity. Therefore, disruptions in stem cell quiescence such as a mutation, can lead to the loss of the

stem cell pool and tissue regeneration capacities. In conditions such as DMD, mutations in the dystrophin gene trigger repetitive muscle degeneration-repair cycles, eventually depleting the satellite cell reserve. Over time, this reduces the muscle's ability to regenerate and results in more fibrotic tissue than healthy muscle.

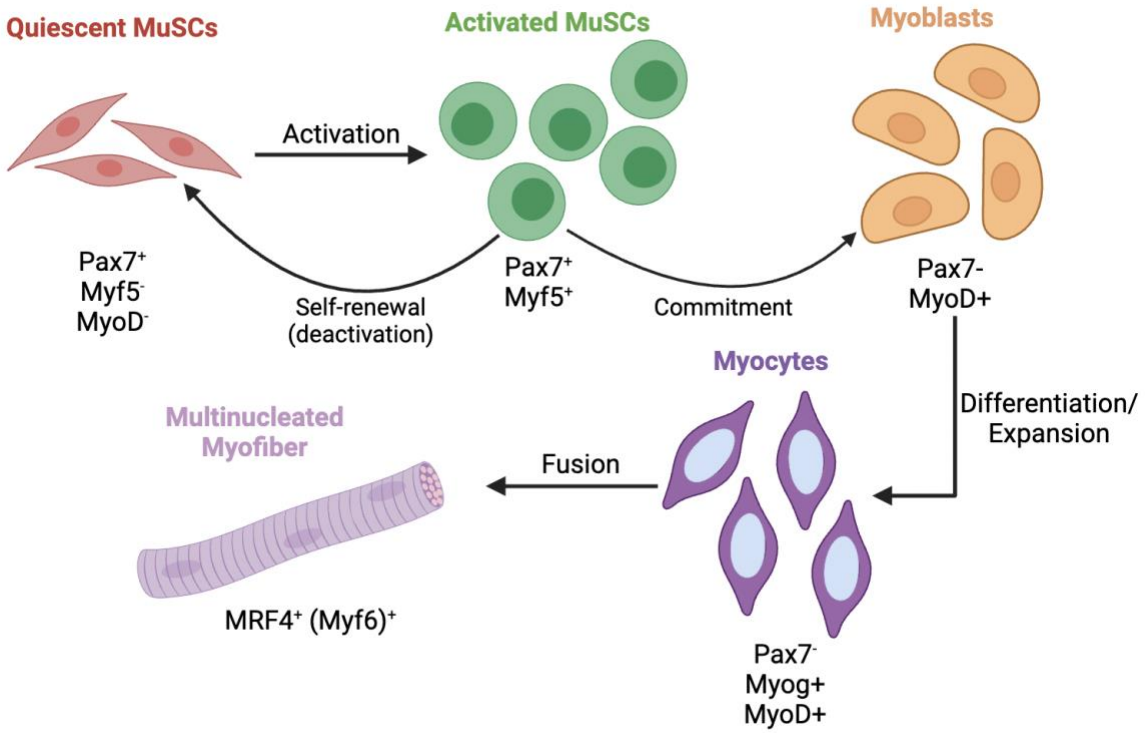


Figure 2.2. Representation of MuSCs Activation and Differentiation Pathway. The diagram illustrates the transition of MuSCs from a quiescent state through activation and self-renewal via asymmetric division, into myoblasts and subsequent expansion and commitment into myocytes, which then undergo fusion to form multinucleated myofibers. Created with BioRender.com.

Upon activation, satellite cells, referred to as ASCs, differentiate into myogenic precursor cells (MPC) or, alternatively, adult myoblasts. Their dynamic nature is evidenced by their high responsiveness to various signaling pathways. In fact, localized damage activates satellite cells along the entire muscle fiber, prompting them to move to the injury site. Their activities are further governed by the expression patterns of Myod1 and Myf5 **(Figure 2.2),** validated by several studies[29]. For instance, Myf5, as demonstrated by Ustanina et al., has a distinct role in

adult myogenesis, affecting muscle regeneration kinetics and myoblast proliferation[30]. More recently, Conerly et al. demonstrated that while Myf5 and Myod1 share genome-wide binding profiles, Myf5 primarily modifies chromatin structure for later transcriptional activation by Myod1 which directly activates myogenic genes[31]. Following proliferation, a significant portion of ASCs will begin differentiating. They either integrate with damaged muscle fibers or fuse to give rise to new muscle fibers. This differentiation phase is marked by the expression of Myog and Myf6 (MRF4), whereby Myf6 is primarily expressed in fully differentiated myofibers[32] **(Figure 2.2).**

2.4 Molecular Signatures and Regulations of MuSCs

2.4.1 Quiescent Satellite Cells

Much of our understanding of the quiescent state has emerged in the past two decades, particularly during the 2000s, which marked a significant increase in molecular and genetic studies that provided deeper insights into the regulators and signaling pathways involved in the maintenance of quiescent cells[26]. QSCs exhibit distinct characteristics that differentiate them from their activated progeny. Among these are genes that serve as negative regulators of the cell cycle and myogenic inhibitors, with roles in cell-cell adhesion, regulation of cell growth, extracellular matrix (ECM) formation, copper and iron homeostasis, and lipid transportation[3,33]. These differential expressions can be attributed to the QSCs' diminished cellular activity, smaller cell size, reduced mitochondrial content, and decreased levels of biomolecules [34]. Furthermore, they constitute approximately 2%-5% of all nuclei within a healthy adult skeletal muscle fiber, residing between the muscle fiber's plasma membrane and the ECM of the basal lamina[34,35].

Quiescence is actively maintained by cell-cell interactions, primarily via the Notch signaling pathway[36–39]. Any decline in Notch activity is correlated with MuSC activation, this was observed following targeted deletion of Rbpj[39]. Furthermore, deficiencies in Notch downstream targets, Hey1 and HeyL, have revealed a marked disturbance in maintaining quiescence[39,40]. One possible mechanism in which Notch maintains MuSC quiescence is its inhibition of myogenic differentiation: the Hey family transcription repressors are suggested to repress Myog gene expression. Signaling pathways like WNT, mTORC1 and fibroblast growth factor (FGF) have also been identified as potential regulators of QSC maintenance [41,42]. In addition to signalling from the niche, key regulators such as cyclin-dependent kinase inhibitors (CKIs) including p21cip1 (Cdkn1a) and p57kip2 (Cdkn1c) along with Rb family members, are elevated in QSCs [43]. These CDK inhibitors block G1 cyclins which drive cell cycle progression, however, this inhibition is not permanent and is different from the irreversible cell cycle arrest seen when these cells differentiate. The regulation of the balance between QSCs and entry into the cell cycle is influenced also by neighboring cells, including vessel-associated cells and myofibers, reinforcing the importance of cellular crosstalk in maintaining quiescence.

2.4.2 Activated Satellite Cells

Upon satellite cell activation, primarily due to muscle damage, exercise, or pathogenic conditions, the niche undergoes rapid alterations where the satellite cells will activate and proliferate. In 2014, Rodgers and colleagues showed that injury-related systemic signals could induce the QSCs to transit from G0 to G_{Alert}, defining an intermediate state. Ultimately, this questioned prior understanding of MuSCs strictly existing in either a quiescent or activated state[44]. Compared to G0 satellite cells, the G_{Alert} satellite cells have higher mitochondrial

activity and, when activated, rapidly enter the cell cycle[45,46]. The activation process is controlled by signals from the myofibers for effective regeneration. A good example is the temporal switch from Notch to canonical Wnt signaling required for proper differentiation[47]. Canonical Wnt signaling regulates the differentiation of MuSCs by directly antagonizing Notch signalling. Whereas non-canonical signals mediate the self-renewal of MuSCs and the growth of muscle fibers[28].

The sequential activation of key transcription factors involved during regeneration are the myogenic regulatory factors (MRFs), which are primarily characterized in developmental myogenesis through studies using transgenic knockout mice[48]. At this stage, ASCs begin to express markers such as Myod1 and migrate out of the basal lamina to the necrotic myofibers in the form of myoblasts. Which means that MuSCs transit to a $Pax7^{high}/Myod1^{high}$ state, allowing for proliferation and differentiation. As they commit to differentiation, they exhibit a $Pax7^{low}/Myog^{high}$ expression pattern afterwards. Research conducted by several studies have demonstrated these patterns before and after injury[29,37,49]. For example, an early study in 2004 by Halevy et al. found that $Pax7^{high}/Myod1^{low}$ cells represent self-renewing myoblasts[50]. Conversely, the frequency of $Pax7^{high}/Myod1^{high}$ cells (ASCs), declines in cultures as cells transition into a $Myog^{high}$ state. More recently, in 2018, Fujimaki et al. confirmed a reduction of $Pax7^{high}/Myod1^{low}$ cells in satellite cell-specific conditional knockout mice for Notch1/Notch2 compared to a control group after induced injury[37].

2.4.3 Self-Renewal

The ability to self-renew is a defining trait of stem cells and manifests in two primary division strategies, asymmetric and symmetric[51,52]. In asymmetric division, a single stem cell splits into two different daughter cells, one that remains a stem cell and another that differentiates. In

symmetric division, a single stem cell divides to yield two identical stem cells. Factors such as the orientation of the mitotic spindle during division play a pivotal role in influencing the choice between the two modes[53]. External stresses, including trauma or disease, can further affect this choice[51]. For instance, situations demanding rapid cell multiplication may shift the balance in favor of symmetric division.

The self-renewed MuSCs are presented as $Pax7^{high}/Myod1^{low}$ cells (**Figure 2.2**). One common signalling pathway is Delta-Notch signaling whereby Delta ligands from the niche activate Notch signaling in the MuSC[38]. Moreover, recent discoveries also highlight the importance of Spry1 in maintaining the homeostasis of self-renewing satellite cells primarily through an extracellular signal-regulated kinase (ERK)-dependent signaling cascade[54].

2.5 The Role of The Satellite Cell Niche in Regeneration

The stem cell niche is a dynamic microenvironment *in vivo* that supports stem cell survival and ensures their proper function. It is a unique spatial organization, providing not only its physical location for stem cells, but also the anatomical and functional interactions responsible for activation and repression of regulatory genes. This term was introduced by Schofield in 1978 to explicate bone marrow cell behavior in hematopoiesis[55]. This theory suggests that the niche is a constantly evolving environment, changing states to either maintain stem cell quiescence or activate them, depending on the demands of the local or systemic environment. Although the concept remains consistent, different tissues have specialized support systems, marked by distinctive cellular components. For instance, during regenerative myogenesis, satellite cells and their committed progeny partake in the remodeling of the niche predominantly regulated by extrinsic signals[29].

Over the last two decades, research has illuminated several ways through which myeloid cells, especially macrophages, play a role in muscle regeneration. Tissue-resident macrophages, originating from the bone marrow, differentiating in the injured tissue, have multifaceted roles including the engulfment of foreign materials, antigen presentation, release enzymes to fight off pathogens, and produce cytokines[30]. In addition, the activity of invading leukocytes including monocytes, neutrophils, and lymphocytes largely dictate the regenerative outcome[56]. Fibrogenic cells, encompassing fibroblasts and fibro/adipogenic progenitors (FAPs) also orchestrate the deposition and remodeling of the ECM[26]. Notably, the vascular landscape, consisting of endothelial cells and pericytes release growth factors which promote activation such as VEGF, PDGF, and HGF[57]. These various components of the MuSC niche promote the efficient progression of regeneration through signals sent by the microenvironment.

2.5.1 Overview of Skeletal Muscle Regeneration

Understanding the regeneration of skeletal muscles holds significance for functional recovery following injury or exercise. This highly orchestrated process, activated by injury-induced disruptions in local homeostasis, consists of several phases: degeneration, inflammation, regeneration, remodeling, and maturation.

2.5.2 Degeneration Phase

Prior to muscle tissue damage, the satellite cell niche is maintained in a quiescent state, where there is an architecturally sophisticated arrangement whereby the apical side engages with the ECM, and the basal side communicates through cell-to-cell interactions with the muscle fiber. The components of the ECM, including collagens, laminins, and proteoglycans, interact with

the cell surface receptors on MuSCs, such as integrin α7 and β1, and dystroglycan, which drive signal transmission from the ECM[3]. However, in response to injury, the ECM undergoes compositional modifications to preserve the gross integrity of the tissue until degenerated fibers have been cleared (through phagocytosis) and innervated young muscle fibers have been formed in the correct anatomical position[58]. Certain proteoglycans along with tissue inhibitors of metalloproteinases (TIMPs), contribute in preventing ECM degradation and maintaining the integrity of the niche[59].

2.5.3 Inflammation Phase

Injury typically initiates a rapid and sequential influx of inflammatory immune cells essential for the repair process. For these immune cells to effectively act within the muscle, they require signals from damaged or infected areas that attract and activate them. Post tissue repair or infection resolution, these immune cells either leave the tissue or undergo senescence to ensure uninterrupted regeneration[60]. The predominant immune cells involved in the regenerative sequence are neutrophils, macrophages, T cells, and mast cells. The injured sites become rich in pro-inflammatory molecules, including cytokines, reactive oxygen species (ROS), and lipid signaling mediators.

Following degeneration, neutrophils rapidly initiate chemotactic recruitment of circulating leukocytes at local sites of damage. Within hours, they will extravasate from circulation and enter the injured area, clean necrotic debris, and mount a pro-inflammatory response. Pro-inflammatory cytokines, such as IFN-γ, IL-1β, TNF-α, IL-1 and IL-8, are secreted from neutrophils and induce the infiltration of circulating monocytes to establish the early inflammatory microenvironment and the subsequent degradation of injured tissue[61]. This cascade of signals catalyzes the recruitment of circulating neutrophils, resulting in a swift

accumulation at the damaged site, with their numbers peaking between 6–24 hours post-injury and diminishing significantly by 72–96 hours[56,62].

Monocytes, guided by the interaction of signaling molecules such as Ccr2/Ccl2 and Cx3cr1/Cc3cl1, are recruited to the injury site by both innate and adaptive immune cells. They start to appear around 24 hours after injury, becoming more prominent two days later, coinciding with the decrease in neutrophil concentration[63,64]. The monocytes then differentiate into macrophages, specifically pro-inflammatory (M1) macrophages in the first 48 hours, which perform phagocytosis of necrotic debris and secrete cytokines and growth factors like IL-6, IL-1β, TNF-α, HGF, FGF, and VEGF, which all directly affect satellite cell activity and muscle regeneration[65].

2.5.4 Regeneration and Remodeling Phase

Macrophages play central roles in the regulation of the skeletal muscle regeneration. During the muscle regeneration process, these cells undergo two different stages of activation and are categorised to two major types: the classically activated macrophages M1 and the alternatively activated macrophages M2[61,62,66]. The M1 macrophages are proinflammatory, whereas the M2 macrophages are anti-inflammatory. At the early stage of regeneration, M1 macrophages are the most dominant macrophage type[65]. Their initial role is phagocytic, removing cellular debris produced by the trauma. However, it is important to note that beyond phagocytosis, M1 macrophages contribute significantly to angiogenesis and fibrosis, as well as maintaining the homeostasis of the muscle microenvironment.

M1 macrophages release a variety of cytokines such as TNF-α, IL-6, and IL-1β, which play essential roles in the subsequent stages of regeneration[65,66]. TNF-α, for instance, is crucial for muscle stem cell attraction to the injury site, as well as their proliferation. It achieves this

by triggering the nuclear factor-kappa B signaling pathway, which is necessary for differentiation. Any interference with TNF-α's function, such as through the use of anti-TNF-α or inhibition of the p38 kinase activity, results in a downregulation of muscle differentiation markers like Myod1, Myog, or Myosin[67].

As myogenic stem cells proliferate and differentiate, the local inflammatory milieu shifts. This transformation is characterized by the phenotypic conversion of macrophages from M1 to M2 subtypes, a process known as polarization. These M2 macrophages secrete anti-inflammatory cytokines such as IL-4, IL-10, and IL-13, mitigating localized inflammation post-injury[62,68]. Concurrently, these macrophages support MuSC differentiation into myotubes, thereby aiding myogenesis and regeneration.

2.6 Overview of Muscle Injury Models

Much of our current understanding about muscle regeneration and MuSC function originates from experimental injury models in mice and rats[69]. The most common models use protocols including freeze injury (FI), intramuscular injection of barium chloride (BaCl2) and myotoxins[70] (**Table 2.1**). Myotoxins, specifically, cardiotoxin (CTX) and notexin (NTX), are derived from snake venoms, which target and cause extensive destruction of muscle fibers, yet preserve the ECM and do not directly affect satellite cells[69–71]. This selective action facilitates a focused study of muscle regeneration processes without the confounding effects of structural damage to the muscle tissue. After CTX or NTX-induced damage, the muscle recovers to its pre-injured state within 10 days post-injury in mice. Although in all these models the muscle regenerates completely, the trajectories of the regenerative process vary considerably[70–74]. Therefore, one must recognize that the nature of the injury model should be chosen carefully depending on the desired outcome. For example, while FI models offer the opportunity to

examine the process of regeneration from a more severe form of injury, they come with certain limitations[70]. The non-specific nature of the damage caused by FI, affecting all cell types including satellite cells and vascular networks, can complicate studies focused on muscle-specific responses by introducing variability in injury extent and recovery patterns. In contrast, BaCl2 models are known for their specific induction of muscle membrane depolarization and fiber necrosis, making them suitable for studies focusing on membrane-related muscle injuries and the subsequent inflammatory response[73,75]. However, the use of BaCl2 can result in variable and sometimes non-uniform muscle damage, potentially leading to inconsistencies in experimental outcomes. Moreover, BaCl2 can also damage vascular structures within the muscle[75], which may not be ideal for studies concentrating solely on muscle fiber dynamics.

CTX and NTX are widely used myotoxic agents as their inoculation via intramuscular injection does not require a surgical procedure[76]. CTX is particularly effective for studies that require enhanced satellite cell proliferation[70,74,77]. Post-CTX injury, satellite cell counts initially diminish but may exhibit up to a threefold increase from baseline, maintaining elevated levels even after subsequent injuries[70,78]. This response, coupled with CTX-induced controlled inflammatory reaction where IL-6 levels remain elevated, contributes to an increase in satellite cell numbers[69,70]. The process may also lead to myofiber hypertrophy and increased muscle mass, reflecting CTX's role in augmenting macrophage and satellite cell activity for efficient tissue regeneration. In contrast, NTX exhibits a more toxic effect than CTX, with a high myolytic impact through the hydrolyzation of sarcolemma lipids[79]. It presents a less severe initial decrease in satellite cells compared to CTX, followed by a continuous increase during regeneration. This makes NTX valuable for studies focused on chronic inflammation and its impact on muscle regeneration. Despite their differing effects, both CTX and NTX are

effective in studying the trajectory of muscle regeneration, albeit with slight variations in satellite cell proportions and depletions.

In addition to the use of chemical agents for *in vivo* studies of muscle regeneration, there are also *in vitro* models that serve a complementary purpose[80–83]. These include traditional cell culture models, which primarily serve the purpose of elucidating basic cellular behaviors and molecular mechanisms under controlled conditions. Unlike *in vivo* models using CTX and NTX, which offer a more holistic representation of muscle injury and regeneration. Cell culture models, specifically two-dimensional (2D) systems, enable researchers to isolate and study specific cellular responses in a highly controlled environment, albeit with the limitation of not fully simulating the complex interactions among multiple cell types and the ECM[84]. However, more recently, three-dimensional (3D) culture models have been developed to better mimic physiological dynamics[85]. These systems come closer to representing the actual environment cells experience of the actual in vivo environment. For instance, in 2015, a study successfully bioengineered 3D human skeletal muscle cultures, termed 'myobundles,' using primary myogenic cells derived from human muscle biopsies[86]. These myobundles mimicked the structure and function of native muscle, featuring aligned, multinucleated, and striated myofibers, as well as a pool of Pax7+ satellite cells. They were capable of spontaneous and electrical stimulus-induced twitch and tetanic contractions, displaying a sophisticated level of contractile function and biochemical responsiveness. However, their implementation is limited due to high complexity and associated costs, the need for specialized technical expertise, and challenges in achieving consistent and reproducible results. Additionally, scaling these cultures for broader applications is difficult, and they may not fully replicate the full microenvironment.

Category	Injury Model	Description
Mechanical Injury Models	Chemical insults	Injuries caused by chemicals, often used to study regeneration.
	Crush/Contusion injuries	Physical injuries that may or may not cause open wounds.
	Freeze injuries	Injuries caused by freezing tissues.
	Excisions (VML model)	Large volume removal of muscle to simulate VML.
Ischemic Injury Models	Permanent Ligation	Permanent blockage of blood vessels, leading to muscle loss.
	Ischemia/Reperfusion (IR) injuries	Temporary blockage followed by restoration of blood flow.
VML Injury Models	VML Animal Models	Various models with differing experimental setups, focusing on substantial muscle mass removal.
In Vitro Cell Culture	3D Fibrin-matrigel® matrix model	Utilizes primary rat neonatal myogenic cells to study regeneration post cardiotoxin injury.
	Scratch/Stopper assays	Rapid modeling of cell proliferation and migration in isolation.
Myotoxin Models	Cardiotoxin (CTX) injection	A cobra-derived toxin that causes significant muscle damage but leaves the ECM intact.
	Myotoxin-induced injury	Induces massive myolysis and necrosis in muscles, preserving satellite cells for regeneration.

Table 2.1. Overview of Muscle Injury Models of in vivo and in vitro techniques.

2.7 Mathematical Modeling

2.7.1 Population Dynamics

Population dynamic models are among the oldest and most canonical in mathematical biology, from the early predictions of Malthus warning of exponential human growth outstripping resources to the more nuanced models of Verhulst and Gompertz which factor in environmental constraints and varying growth rates[87]. In 1874, Galton and Watson introduced the first systematic approach to explore social science phenomena using mathematical methods[88]. Because of the stochastic nature of biological systems, mathematical and

computational models have become a more acceptable method to describe complex cell mechanisms and tissue physiology[9,89,90].

The mathematical modeling of stem cells, particularly MuSCs, involves complex dynamics that are not yet fully understood[42]. While general stem cell models have been extensively studied, specific models for MuSCs are less common. One of the foundational concepts in stem cell biology, inspiring various mathematical and computational frameworks, is Waddington's epigenetic landscape which describes the dynamics of a cell from pluripotency to its differentiated state[91]. In this landscape, a pluripotent cell is represented as a ball positioned at the top of a hill. As the cell progressively loses its differentiation potential, it behaves similarly to a ball rolling downhill into a valley, which represents a differentiated state. More specifically, this analogy was formulated to guide the development of mathematical models to illustrate how genetic interactions and their products guide cells through various developmental pathways to establish specific phenotypes, which are often probabilistic in nature[92,93].

In biology, these models have been employed extensively, from modeling the spread of infectious diseases to understanding biological invasions and the dynamics of agriculture[94,95]. Similarly, mathematical models are widely applicable in interpreting the biological properties of stem cells. Notably, the first mathematical model of stem cells was conducted in 1964 using stochastic techniques to capture cell fate decisions, and since then has been extended to cover many other aspects of their behaviours such as gene regulation, cell cycle progression, and cell population dynamics based on experimental data[96]. These models compel us to explicitly define our knowledge about a system to test its adequacy in explaining the phenomena being investigated. This means not only describing cellular behaviors and

interactions but also critically evaluating if our current theoretical understanding is sufficient to account for the observed complexities in stem cell dynamics, such as differentiation, proliferation, and self-renewal processes. Indeed, when one can validate these models by experimental observations, the reciprocal benefit for experimentation can be insightful for biologists. For example, Johnston et al. developed an ODE model simulating the behavior of stem cell populations in a healthy colonic crypt and extends these models to consider the asynchronous division of cells[97]. Their model revealed that small variations in parameters such as cell renewal rates can lead to either exponential growth or decay of the cell population. Therefore, the authors introduced a feedback mechanism providing a potential explanation for the observed behavior after mutations occur. Building on this foundation, Mamis et al. developed a new model also analyzing cell population dynamics in colonic crypts, where their predictions closely matched the observed experimental increases in crypt area and the number of proliferating cells per crypt in mutant crypts[98]. This model along with several others, can then be implemented as a framework for testing hypotheses about cell dynamics and disease progression, potentially reducing the time and resources required for empirical studies.

2.7.2 Nonlinear Ordinary Differential Equations

The use of dynamical models to understand the function of biological systems has already been successfully illustrated in many biological systems, including signalling, metabolic and genetic regulatory networks[8,89,98–102]. Further, dynamic model-based approaches have also been used to identify possible ways of intervention or re-design, such as in optimal experimental design, metabolic engineering, and synthetic biology[103–106]. Other recent efforts have been focused on scaling-up, i.e. on the development and exploitation of large-scale (genome-scale) kinetic models, and ultimately, whole-cell models[107].

The simplest method in describing the changes in cell dynamics (i.e. kinetics) is to follow the total cell number along time. The natural associated mathematical formalism is the ODE, whereby the cell number dynamics is then described as a continuous function of time. In biology, ODEs are often used to model the change in population of a species, concentration of a molecule, or behavior of cells over time. For example:

$$\frac{dN(t)}{dt} = cN \tag{2.1}$$

Where N represents the population of a species, *t* is time, and *c* is the growth rate. This linear ODE tells us that the rate of change of the population is directly proportional to the current population. Using ODEs, the temporal variations in cell number can be depicted as continuous functions, simplifying the representation of complex biological processes like muscle regeneration. These mathematical constructs can be designed to account for variables including cell proliferation, apoptosis rates, and interaction coefficients, illustrating the mechanistic foundation of muscle regeneration processes.

While the use of ODEs to model MuSCs remain limited in the literature, their application in other stem cell systems has shown promising results[108–112]. One such example involves human induced pluripotent stem cells (hiPSCs)[113]. Hall et al. capture the population dynamics of hiPSCs seeded on melt electrospun scaffolds. By segmenting the cell aggregate into three distinct subpopulations, stem, progenitor, and differentiated cells, they quantitatively track the dynamics of their cellular processes under different conditions. Comparably, Rashidi et al. focuses on modeling hematopoietic stem cells (HSCs), which are responsible for the production of mature blood cells[114]. By combining an ODE model of human granulopoiesis with an agent-based model (ABM) of HSC organization, their model successfully describes the dynamics of bone marrow cell stages and circulating cells under various perturbations like

G-CSF treatment or chemotherapy. By incorporating specific variables and parameters such as cellular death rates, proliferation rates, differentiation rates, and the impact of extrinsic populations, ODE models allow biologists to interpret mathematical results across a range of clinical or experimental scenarios, such as growth factor applications and pharmacological treatments, leading to more informed and targeted experimental designs involving stem cell dynamics.

Mathematical descriptions of stem cell dynamics are often multi-compartmental and continuously structured. Each cell type within these models, represented as a distinct compartment, receives inputs from preceding stages and influences subsequent ones. When interactions between different cell stages are incorporated, these models become nonlinear, often necessitating the inclusion of negative regulatory feedback to ensure consistent cell production[115–117]. This differentiates nonlinear from linear equations, which often overlook cellular interactions. This is evident in previous models designed to maintain homeostasis in MuSCs, where feedback from the muscle compartment regulates both the symmetric division and proliferation rates of satellite cells[118]. Furthermore, the dynamics of cell proliferation and differentiation, governed by external signaling molecules (i.e. from neighboring immune cells), introduce additional nonlinearities into such models. In the context of our model, a nonlinear behavior might emerge when, for instance, the proliferation rate of active cells is not constant but rather depends on the number of cells itself. This could happen if resources become limited as the population grows, or if cells start inhibiting each other's growth beyond a certain density. For example, consider two cell types, *Cell Type A*, and *Cell Type B* and *Cell Type A* releases a specific signaling molecule that promotes the proliferation of *Cell Type B*. Then, the rate of proliferation of *Cell Type B* might be modeled as proportional to the product

of the number of *Cell Type A* cells and the existing number of *Cell Type B* cells, resulting in the nonlinear ODE as follows.

$$\frac{d[B](t)}{dt} = c[A][B] \tag{2.2}$$

Where, *c* represents a constant that describes the rate of *Cell Type B* proliferation in response to the signaling molecule from [A]. The term [A][B] is nonlinear, representing the product of the two cell population sizes. In other words, this term indicates that the growth rate of *Cell Type B* is influenced by both its own population size and that of *Cell Type A*. More generally, in defining any large system of cells, the general form of these equations can be given by a deterministic state-space model. Let's define $x(t,\theta)$ as the state vector capturing the various states of MuSCs over time. For instance, $x_1(t,\theta)$ could represent the quantity of QSCs, $x_2(t,\theta)$ as ASCs, $x_3(t,\theta)$ as differentiating cells, and so on. θ is the parameter vector that includes factors that influence the dynamics on proliferation rates, differentiation rates, etc. Therefore, the state-space model describing the system can be written as follows.

$$\frac{dx}{dt} = f(x, \theta) \tag{2.3}$$

For a given set of parameters θ, and an initial condition $x(0)$ representing the cell counts at time zero, the model can approximate a unique trajectory or behavior pattern over time, subject to specific conditions that pertain to the smoothness of the dynamics function *'f'*. While this trajectory can sometimes be solved mathematically, nonlinear functions typically necessitate the use of nonlinear ODE solvers available in various programming languages, such as Runge-Kutta, Adams-Moulton, or other numerical methods[119]. Finally, with empirical data on population sizes at different times, it becomes possible to seek parameters θ that align the model trajectory with observed data.

2.7.3 Parameter Estimation

Nonlinear Least Squares (NLLS) is a popular technique for parameter estimation, especially when dealing with systems that are fundamentally nonlinear. The objective in NLLS is to minimize the sum of the squared differences, between observed $y_i(t_i)$ and model-predicted values $x(t_i, \theta)$, where t_i represents observations times and where θ is a vector of parameters. The NLLS problem can be stated as an objective function F:

$$F(\theta)=\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}(t_i)-x_j(t_i,\theta))^2 \quad (2.4)$$

In this formulation, *n* represents the number of observation times, and *m* is the number of state variables in the model. The objective is to find the parameter vector θ that minimizes this sum of squared residuals. In solving this function, one must begin with a robust initial estimate either from prior knowledge, gathered from previous research or predicated upon established models like exponential decay, which are substantiated either by empirical observations or theoretical frameworks. Generally, this is followed by an optimization method to solve the NLLS such as Gauss-Newton, Levenberg-Marquardt, and Broyden–Fletcher–Goldfarb–Shanno (BFGS). Indeed, parameter estimation is a common challenge in dynamical models, as it represents the crucial step to obtaining predictions from computational models of biological systems[120]. This process is often referred to as the *inverse* problem, which stands in contrast to the *direct* problem of model simulation using known parameters[121]. The task is to deduce parameters from observed outputs rather than predicting outputs from given parameters. Two key challenges arise when dealing with the inverse problem, ill-conditioning, which results from actors like over-parametrization of the model, scarcity of experimental data, or significant measurement errors. Ill-conditioning can relate to the lack of identifiability due

to the model's structure or information-poor data. One aspect of ill-conditioning involves the identifiability of parameters, which can be influenced by the model's complexity or limited data availability. For instance, optimizing parameters to reflect experimental data on how some observables behave under various experimental conditions may result in a multimodal problem. In multimodal problems, there can be multiple sets of parameters that provide a good fit to experimental data, but not all these sets have meaningful biological interpretations. Several advanced optimization techniques have been developed to address this challenge. Methods such as gradient-based methods, like Conjugate Gradient Descent, seek to improve the speed and precision of convergence, which works by iteratively finding conjugate search directions and updating the solution in a way that minimizes the objective function[122]. However, one of the most prominent methods that balances both efficiency and robustness against local solutions is the quasi-newton method. For instance, the BFGS method is one of those algorithms and works by iteratively approximating the inverse hessian matrix using gradient evaluations, thus combining the rapid convergence properties of newton methods with reduced computational overhead.

2.7.4 Previous Efforts in Mathematical Modeling of Skeletal Muscles

By integrating computational modeling with experimental data, researchers and clinicians can predict muscle recovery strategies, assess the effectiveness of various therapies, and study the intricate interactions among cells, scaffolds, and pharmaceuticals during muscle regeneration. Various skeletal muscle models have been developed over the past decade. Agent-Based Models (ABMs) have been a popular choice due to their ability to capture the individual behaviors and interactions of cells during regeneration[123,124]. While earlier models were simplistic and one-dimensional, recent ones are more anatomically accurate and complex,

shifting from phenomenological representations to detailed mechanistic models. For example, Martin and colleagues propose a new agent-based computational model that focusses on inflammation and regeneration after acute muscle damage whereby they examine multiple cellular behaviours and their interconnectedness and intricate feedback loops[125]. Comparatively, Virgilio and colleagues provide an ABM computational model into muscle injury varying across healthy and muscular dystrophy X-linked (mdx) mice[123]. Although ABM models offer advantages, such as the ability to model heterogeneity, spatial dynamics, and individual-level interactions, ABMs often incorporate stochasticity to represent the variability in agent behaviors. While this can capture the inherent variability in biological systems, it also means that multiple runs might be necessary to obtain average behavior. ODE models, on the other hand, are deterministic by nature, providing consistent results for a given set of parameters. Their strength lies in efficiently modeling large-scale systems and capturing global trends with mathematical clarity and computational efficiency. This is particularly evident in fields like population dynamics and chemical kinetics, where ODEs can represent the complex interactions and growth patterns of species, or in epidemiology, where they enable streamlined modeling of disease spread and vaccination impacts across large populations. These models can be favorable in terms of ease of parameter estimation, robustness against missing data, and facilitating sensitivity analysis, making them highly effective for analyzing and predicting average behaviors in large, continuous systems.

Building upon previous work in the field, in 2014, Jarrah's and colleagues focused on the explorations of ODE modeling of DMD revealing a dynamic reminiscent of predator-prey interactions between the immune system and *mdx* tissue[11]. The model incorporates various cells and states including normal muscle fibers, damaged muscle fibers, regenerating fibers,

CD4+ T helper lymphocytes, macrophages, and Cd8+ cytotoxic T cells. The parameters, such as cell activation rates and degradation rates are estimated using pooled immune and histopathological time course data from soleus and tibialis anterior (TA) muscle types available within various sources of literature. While Jarrah's model focuses on the states of muscle fiber states during degeneration and regeneration, it does not delve into the dynamics of MuSCs and their different states. Despite their critical role in muscle repair and regeneration, a limited number of mathematical models profoundly address the nuances of MuSC dynamics. A subsequent study in 2021 by Kojouharov and his team adopted a more refined perspective, accentuating the significance of MuSCs in regeneration[126]. Their formulation distinctly delineates the three primary phases of muscle injury recovery: destructive, repair, and remodeling, while also emphasizing the distinct functions of pro-inflammatory and alternatively activated macrophages. Notably, the model relies on 21 distinct parameters, suggesting its comprehensive nature. Nevertheless, the derivation of these parameters was based on literature that employed diverse methodologies. Some parameters, such as decay rates, were more readily derived from established values like the half-life of certain species. The authors candidly concede that parameter estimation poses inherent challenges. Due to the natural variability and uncertainty in biological systems, there can be a significant range in parameter values, potentially impacting the model's predictive accuracy. Kojouharov's model had successfully captured the proportions, represented as log-linear plots, of immune cells and satellite cells involved in response to different intensities of damage. Similarly, Jarrah's model also simulated populations of immune cells but based on mdx mouse model of DMD. While such models offer valuable insights and generate hypotheses for future

experiments, their predictive power may significantly improve when aligned with empirical experimental data.

The objective of our modeling work, as outlined in **Chapter 4**, is to not only leverage a more uniformly acquired experimental dataset to focus on these specific satellite cell states and their transitions during regeneration but to also to identify parameters and understand their significant roles, particularly in reaction to variations and perturbations in these parameters.

While the strengths of using ODEs include a mathematically rigorous representation and the ability to tease out complex dependencies, one must note some limitations. As with any model, there is a balance between granularity and simplicity, and certain biological nuances might be streamlined for the sake of tractability. However, the integration of single-cell transcriptomic data with mathematical modeling, exemplified by Fischer et al.'s study and several others, has opened new pathways in understanding the complexities of biological systems, where instead of relying on conventional literature-derived parameters they employ a pseudodynamics framework to model the temporal progression of population density from progenitors to terminal fates in a transcriptome space[127].

2.8 Experimental Data

2.8.1 Computational Analysis of Single Cell RNA-Sequencing Data

Transcriptomics is a branch of genetics that focuses on the comprehensive study of messenger RNA (mRNA) levels across various cell types, tissues, species, and environmental conditions. The first transcriptome of a single cell was sequenced in 2009[128]. Prior to this, the prevalent method was bulk RNA sequencing, which only provided an averaged expression profile across thousands of cells. The first scRNA-seq study by Tang et al. quantified mRNA from an

individual cell within a mouse embryo, which revealed 75% more gene detection than previous microarray approaches[128]. Since then, other single-cell transcriptomics approaches were developed, making it possible to sequence tens of thousands of cells in parallel enabling for the characterization of rare cell types.

In this thesis, all analysis begins from the raw counts in the form of a gene-by-cell expression count matrix acquired from the Gene Expression Omnibus (GEO) database (**Figure 2.3**). To achieve this counts matrix, cells are first captured in a small reaction volume, such as individual droplets in oil-in-water emulsions or gravity to load cells into nanoliter chambers on microarrays. Following isolation, cells undergo lysis, then the cell's mRNA reverse transcribed such that its complementary DNA (cDNA) carries a unique cell barcode and a unique molecular identifier (UMI). These beads are then pooled, generating a first strand containing the full length of the mRNA transcript as well as the cell-specific barcode and transcript specific UMI. After reverse transcription, the cDNA product is amplified into a library compatible with next generation sequencing using paired-end mode allowing for identification of the cell barcode and UMI as well as the gene that was encoded on the transcript.

This results in a large cell by genes matrix, arranged with cells as columns and genes as rows, and the entries are integers that represent the number of counts of each transcript in each cell. The matrix is normalized by library size to account for differences in transcript recovery between individual cells and analyzed with a variety of unsupervised methods.

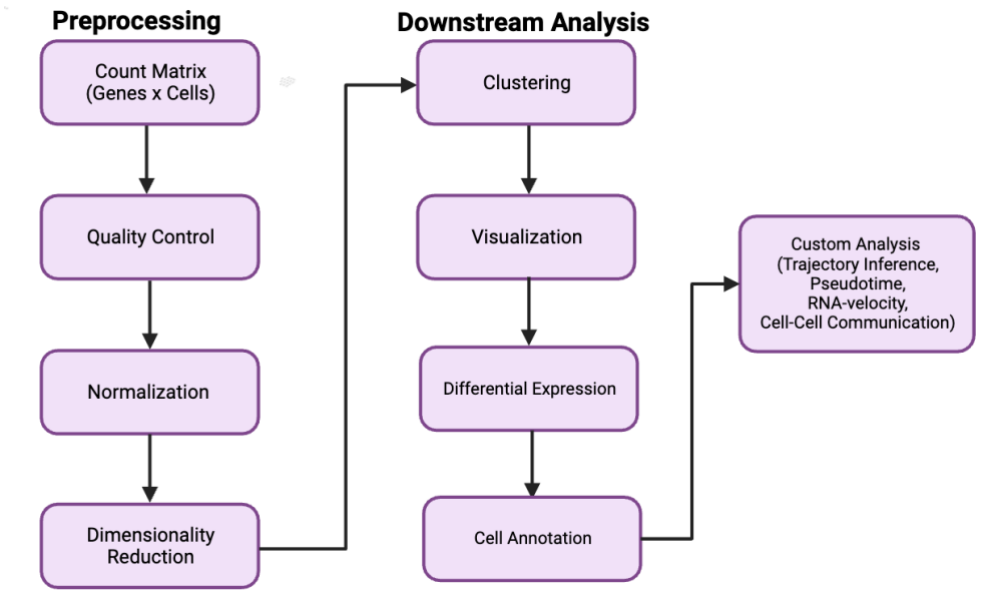


Figure 2.3. Common computational workflow in scRNA-seq data analysis. The preprocessing phase begins with the generation of a count matrix, followed by quality control measures and normalization to adjust for technical variability. Dimensionality reduction is then applied to compact the data from a high-dimensional onto a lower-dimensional space. The downstream analysis includes clustering to group similar expression profiles, visualization to interpret patterns, cell annotation based on marker genes, and custom analyses such as trajectory inference to elucidate developmental pathways.

Downstream analysis typically begins with noise reduction via principal component analysis (PCA), followed by the generation of two-dimensional representations through uniform manifold approximation and projection (UMAP). This facilitates the identification of cell clusters, each potentially representing a distinct cellular phenotype, using algorithms like Louvain or Leiden clustering. Differential expression analysis then facilitates the comparison of gene activity across these clusters or under different biological scenarios. Additional computational tools, including those for trajectory inference and pseudotemporal ordering, are applied depending on the specific biological questions being addressed.

2.8.2 Single-Cell RNA-seq in Stem Cell Biology

Advancements in scRNA-seq have transformed biology and medicine, enabling the quantification of gene expression levels across thousands of individual cells within a singular

controlled experiment[7,129–132]. This stands in contrast to bulk RNA-seq, which provides an averaged gene expression profile, obscuring heterogeneity by measuring average expression level for each gene across a large population of cells[133]. In stem cell biology, scRNA-seq has been used to expose the cellular heterogeneity and propose new boundaries in cell-state transitions during differentiation as well as to reconstruct developmental trajectories of entire organisms. For example, previous studies have recently revealed two primary MuSC subcategories[134]. MuSCs that are close to quiescence (cQ) exhibit elevated Pax7 expression alongside genes associated with cell cycle inhibition and stress resilience. In contrast, early activated (eA) MuSCs display gene patterns linked to ribosome formation and mRNA modification. Furthermore, Oprescu and colleagues identified a novel subpopulation of MuSCs termed immunomyoblasts (IMBs), enriched in immune cell activation genes (e.g. C1qa, C1qb, C1qc), major histocompatibility class II antigens (H2-Eb1, H2-Aa, H2-Ab1), and members of the cathepsin family (Ctsb, Ctss)[135]. This gene expression profile represents a previously undescribed MuSC fate whereby a significant proportion of MuSCs at early regeneration stages were classified within this sub-cluster. Similarly, Qin et al. identified an unknown cell population characterized by mitochondria-related musculoskeletal features[136]. Specifically, this was described as a subcluster of Col2a1+ populations, which they termed Col2-mito subcluster. This subcluster was shown to support the regeneration of the musculoskeletal system by providing energy for stress reaction, cell differentiation, and tissue remodeling. Overall, this scRNA-seq technology has enabled researchers with the opportunity to decipher the heterogeneity of muscle tissue into individual cells and characterize dynamic interactions between cell populations during differentiation. It also contributes to annotating

rare or small cell populations with distinct molecular features, their transitional states, and cell fate decisions.

2.8.3 Applications of scRNA-Seq in Skeletal Muscle Regeneration

Cho and Doles recently reported the first scRNAseq study performed on freshly isolated MuSCs and identified significant gene heterogeneity among freshly isolated MuSCs including the expression of muscle specific transcripts[137]. While this study highlights potential heterogeneity among MuSCs, the identification of novel sub-populations of MuSCs remains understudied due to the low cell numbers analysed despite isolation of FACS-sorted Pax7-tdTomato+ cells. The strength of single-cell resolution mapping to identify novel cell populations in skeletal muscle has been recently highlighted by Giordani and colleagues who combined scRNAseq with CyTOF, creating a transcriptomic atlas of over 12,551 muscle-resident cells and categorize between novel populations of cells including resident tenocyte-like cells and a myogenic subset of smooth muscle and mesenchymal marker cells[138]. This study clearly demonstrate that single cell analyses strategies are successful in deconstructing and identifying novel cell populations.

Dell'Orso and colleagues provided a more refined single cell description of MuSCs activation and differentiation by employing FACS to enrich and isolate MuSCs[134]. They were able to identify two distinct subpopulations within MuSC, the cQ MuSCs, which are enriched for genes such as Pax7, Hes1, and Gas. The eA MuSCs were found to be enriched for Myf5 and Myod, suggesting they're in a mitotically activated state. The study showcased a branched progression from the two subpopulations to either a population of cycling myoblasts or committed myocytes. The myoblasts expressed cell cycle regulators such as Ccnd1 and eIFs

(translation initiation), while the committed myocytes were enriched for Myog and cycling-dependant kinase inhibitors such as Cdkn1a.

Oprescu and colleagues created one of the most comprehensive datasets in characterizing muscle regeneration with 6 post-injury time points 0 to 21 days post-injury (DPI). Notably, this study identified a subset of MuSCs that showed enrichment for immune-related genes, termed immunomyoblasts (IMB). Through immunofluorescence (IF) and Fluorescence Activated Cell Sorting (FACS) methods, they confirmed that this subpopulation of MuSCs did express immune-related genes after injury. This particular MuSC population had not been distinctly defined before but does have overlaps with existing literature. Earlier works indicated potential roles of specific stem cell subsets and their contribution to muscle regeneration. There is also evidence of MuSCs undertaking immune roles, as shown by antigen presentation post-viral transduction. Additionally, immune cell populations were found to be the largest in number and displayed the most time-dependent dynamics. However, one limitation that was noted in this study was that it did not perform batch correction, which may be needed to account for variations and potential inconsistencies in the data that arise from using multiple biological samples (different mice) and conducting the experiment at different time points.

Micheli and colleagues have built an extensive atlas of muscle regeneration, analyzing over 34,000 cells, detailed 15 different cell types[139]. The authors observed specific gene expressions in QSCs, including Id3 and Btg2 whereby Btg2 had not been previously described in myogenic cells. Additionally, they provide detailed characterization and heterogenous expression of Sdc genes and their specific roles in MuSC cycling and self-renewal, as the mechanistic role of Sdcs in regulating MuSCs was previously a poorly understood area.

However, one notable limitation was that the analysis was based on samples gathered at four specific post-injury intervals: 0, 2, 5, and 7 DPI. This might overlook certain evolving changes between these periods. For example, Oprescu and colleagues identified a minor group of M2 macrophages at 3.5 DPI expressing genes such as C1qa, C1qb, C1qc, Ms4a6, and Ms4a7, which may elude studies with fewer time intervals.

Chapter 3

Temporal Dynamics of Muscle Cells in scRNA-seq

3.1 Objective

My first objective is to delineate the established organization of key cell populations and to quantify these cell types as a function of time throughout the muscle regeneration process. To do so, I downloaded and re-analyzed temporally resolved scRNA-seq datasets from three distinct studies, each documenting the cellular dynamics in response to a myotoxin-induced muscular injury. To ensure a comprehensive understanding of the transcriptional landscape and to maintain the intrinsic heterogeneity of the tissue context, I integrated all samples from each dataset for a consolidated analysis. To anchor the cell counts within a physiologically relevant framework, I normalized the data to cells per cubic millimeter using a scale factor derived from experimentally validated cell densities in the relevant literature. This normalization process was essential to translate the counts into a biological context that allows for direct comparison and validation against known empirical benchmarks.

3.2 Overview

Muscle regeneration is a multifaceted process involving the coordinated response of MuSCs and various stromal and immune cells. Building on the foundational knowledge of muscle biology and MuSC function outlined in **Chapter 2**, this chapter will delve into the application of scRNA-seq to capture the dynamic interplay between these cells *in vivo*. We will explore how individual cells transition through various states of activation, proliferation, and differentiation. The integration of multiple scRNA-seq datasets presents a unique opportunity to assess the consistency of cellular responses across different experimental models and to identify the underlying mechanisms that drive successful tissue regeneration. This analysis not only delineates the diverse cellular constituency of muscle tissue during repair but also quantitatively extracts cell proportions across various stages post-injury. These proportions are further scaled to reflect their relative abundances, for their subsequent application in mathematical modeling in **Chapter 4**.

3.3 Methods

3.3.1 Data Acquisition

To gather a comprehensive view of regeneration, I evaluated 27 scRNA-seq datasets **(Table S1)** and selected 3 published datasets[135,139,140] for individual re-analysis and subsequent integration, with a total of 38 mice samples, that vary across sex, age, injury model, and injury response time-points **(Table 3.1)**. To evaluate the dynamics of MuSCs, the selected datasets were sourced from studies focusing on the TA muscles of C57BL6 mice, each utilizing distinct protocols. De Micheli et al. generated a dataset from 10 scRNA libraries of samples taken at varied time points after inducing myofiber damage with NTX in 4–7 months old mice. Oprescu

et al. prepared 7 scRNA libraries, each combining cells from three 3-month-old C5Bl6/WT mice, sampled at 0-21 days post-injury (DPI) following a CTX injection. McKellar et al. provided a particularly comprehensive dataset, consisting of 21 scRNA libraries, derived from 20-month-old mice post-NTX injection. This dataset stands out due to its extensive sampling, with the most samples per time point and the most frequent post-injection time point intervals, along with the highest number of replicates compared to the other studies. All libraries across the studies were originally constructed using the Chromium 3' Library and Gel Bead Kits from 10X Genomics v3, USA. The sequencing platforms varied, with De Micheli and McKeller opting for the Illumina NextSeq 500 and Oprescu using the Illumina NovaSeq 6000. The processed UMI count matrix files for the external datasets were acquired from GEO, which housed raw scRNA-seq data that have been pre-processed using Cell Ranger. These files correspond to datasets with the accession numbers SRP241205, SRP225546, and SRP294168, originally deposited in the Sequence Read Archive (SRA).

3.3.2 Data pre-processing and downstream analysis

Raw read counts provided by a UMI count matrix for all samples were processed individually and combined (among other samples from the same source) using the Seurat (v4) within R v4.1.3[89]. Samples were first loaded into R using the '**Read10X'** function. Each samples data was then converted into a Seurat object using '**CreateSeuratObject'**. To filter out low-quality cells, genes that expressed less than 10 cells as well as cells with <1000 UMIs and <250 genes were removed from the gene expression matrix **(Figure S1).** In addition, I removed cells with >25% UMIs mapped to mitochondrial genes, as well as cell doublets with the **'DoubletFinder'** function. I log-normalized the data using the **'NormalizeData'** function in Seurat, setting a scaling factor of 10,000. Following this, I scaled the expression of each gene by applying the

'ScaleData' function, which regressed out the effects of both the number of UMIs and the percentage of mitochondrial gene expression in each cell. Variable features across the datasets were identified using the **'FindVariableFeatures'** function, setting the selection method to **'vst'** with a feature count of 2,000. Integration of samples for within their respective datasets was executed via the Canonical Correlation Analysis (CCA) approach, facilitated by Seurat's integration workflow. I first identified integration anchors using **'SelectIntegrationFeatures'** and **'FindIntegrationAnchors'** functions, focusing on shared features across datasets. Subsequently, the **'IntegrateData'** function was applied to merge the datasets into a single Seurat object, correcting for batch effects.

I performed PCA on the gene expression matrices and used the first 20 principal components (determined by **'ElbowPlot'** function) for clustering and visualization using **'RunPCA'** neighborhood graph construction using **'FindNeighbors',** and clustering via **'FindClusters'**. Within the FindClusters algorithm, K-nearest neighbor (KNN) was calculated within Seurat to detect communities of cells. These communities are clusters where cells are more connected to each other within a cluster than to cells in other clusters. The **'FindClusters'** function was executed with a resolution of 0.8, which was visualized and dimensionally reduced by the UMAP technique, using the **'RunUMAP'** function. Finally, differential expression analysis was achieved using the **'FindAllMarkers'** function using a likelihood ratio test that assumes the data follows a negative binomial distribution and only considering genes with >log2(0.25) fold-change and expressed in at least 25% of cells in the cluster.

3.3.3 Batch Effect Correction

To examine the complete heterogeneity and population data, all datasets were merged by employing Harmony algorithm for data integration[141]. Initially, four different approaches were examined for batch correction: CCA (Seurat), Harmony, Scanorama[142], BBKNN[143] **(Figure S5)**. The analysis was conducted using the Harmony package in R (4.2.2) using the **'RunHarmony'** function, which was applied to the PCA embeddings of each dataset with a resolution of 1.2. Harmony's iterative approach aligns these embeddings across different datasets, reducing their batch effects while preserving biological signals. The harmonized data were analyzed further using the clustering algorithms and dimensionality reduction techniques. Graph-based clustering was performed with Seurat, and UMAP was utilized for data visualization using the mentioned pipeline.

3.3.4 Cell Type Annotation and Sub-Clustering

The cell type annotation strategy for the data consisted of a dual approach for robust cell type labelling. This approach integrated both automated and manual annotation methods to capitalize on their respective strengths and mitigate their limitations. The initial step of the annotation process involved automated label transfer using a comprehensive reference atlas from a large-scale single cell/nuclei RNA-seq integration study[140]. This reference atlas was selected for its extensive and diverse cell type categorizations. Employing Seurat's **'FindTransferAnchors'** and **'TransferData'** functions, I aligned our dataset to the reference atlas. This step enabled the assignment of cell type labels to our data based on transcriptional similarity with the predefined cell labels in the atlas. This reference dataset was solely utilized as an annotation tool, and its cells were not integrated into our primary analysis. Dimensionality reduction techniques, such as t-distributed Stochastic Neighbor Embedding (t-

SNE) or UMAP, facilitated the visualization of cell clusters. Subsequent manual annotations were based on the expression patterns of well-established marker genes by referencing established literature for cell type identification. The final cell type labels were determined by comparing the outcomes of the automated and manual annotations (**Figure 3.1**).

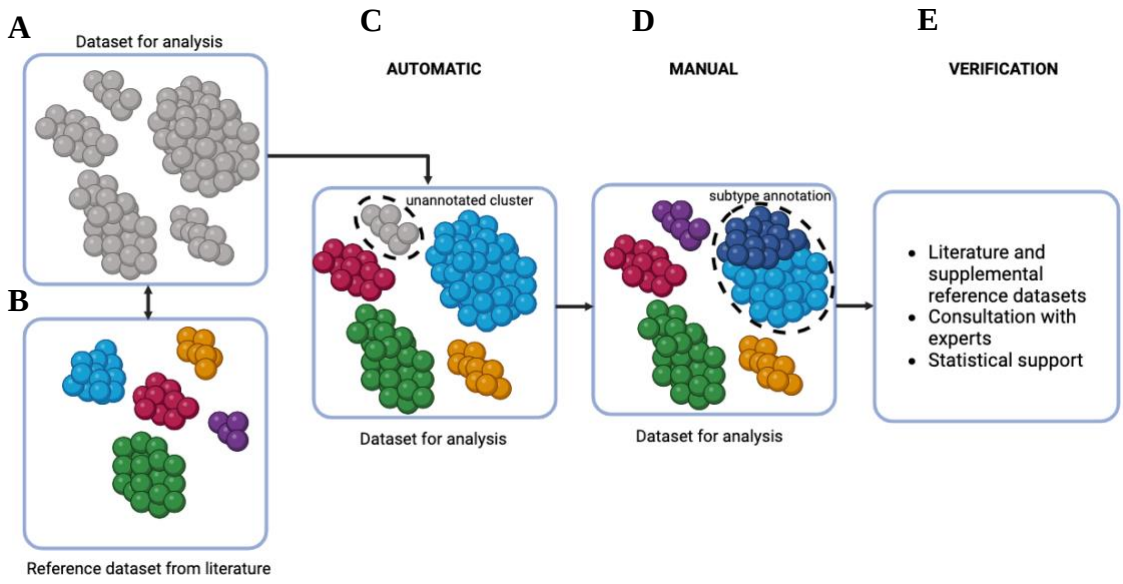


Figure 3.1 Schematic representation of the combined annotation approach for scRNA-seq data analysis. (A) Dataset for analysis: Initial unannotated clusters derived from the scRNA-seq dataset awaiting classification. **(B)** Reference dataset from literature: A pre-annotated dataset obtained from literature serves as a comparison standard for automatic annotation. **(C)** Automatic: The first annotation step involves an automated comparison of our dataset of interest with the reference dataset, resulting in preliminary cluster annotations. Unannotated clusters that do not match with the reference are highlighted. **(D)** Manual: Subsequently, manual curation is performed, where unannotated clusters undergo detailed analysis for subtype annotation and unannotated clusters. This step incorporates expert knowledge and may involve splitting or merging clusters based on gene expression profiles. **(E)** Verification: The final verification step includes a thorough review of the annotated clusters against the literature, supplemental reference datasets, and consultation with experts. Statistical support is used to confirm the robustness of the annotations (from differentially expressed gene analysis). Created with BioRender.com.

Upon completing the primary cell type annotation, I conducted high-resolution clustering on selected clusters for more in-depth analysis. For example, the MuSCs cluster was further divided using the **'FindClusters'** function with a higher resolution setting. This detailed clustering was guided by established marker genes, enabling the categorization of MuSCs into

different states such as quiescent, activated, and differentiating. These states were determined based on an extensive literature review. The specific outcomes and implications of these subclusters will be thoroughly discussed in **Section 3.4**.

3.3.5 Estimation and Calibration of Empirical Time-Series Cell Populations

scRNA-seq datasets naturally exhibit varying numbers of cells, which do not necessarily reflect actual or absolute differences in cell counts in the studied samples. scRNA-seq effectively reveals the fraction of cells in a sample that correspond to each identified type. While the mathematical model in **Chapter 4** could have been framed in terms of these cell fractions, such an approach, akin to the data itself, might misleadingly suggest increases or decreases in a cell population solely due to changes in other populations. My aim was to avoid this potential misinterpretation. Furthermore, a model based on cell fractions poses challenges in correlating with natural underlying biological processes like proliferation and differentiation rates.

The analyzed data provided counts of different cell types identified at each day post injury. These counts offer a snapshot of the cellular composition at specific moments (**Figures S2-S4)**. To standardize these counts for comparisons across samples, I converted them into proportions P_{cell} **(Tables S3-S6).** This was achieved by dividing the count of each cell type by the total cell count at each time point. To translate these proportions into a more biologically meaningful metric, I estimated the number of cells per cubic millimeter of skeletal muscle tissue:

$$S_{cell} = \frac{P_{cell}}{V_{cell}} \tag{3.1}$$

Where S represents the estimated cell count per unit volume $\left(\frac{cells}{mm3}\right)$ and $V_{cell}\left(\frac{mm3}{cell}\right)$ is the average volume of a cell in skeletal muscle. To enhance the accuracy of our cell number estimates, I incorporated external data on the density of mononuclear cells per unit volume in skeletal muscle. Proportions are converted to absolute cell numbers given by:

$$S_{cell} = X \cdot P_{cell} \tag{3.2}$$

The factor *X* in (Eq.3.2) represents the externally sourced estimate, providing a scaling factor to each observed value per day.

3.4 Results

The dataset by McKellar was selected for our model, outlined in **Chapter 4** due to its extensive cell counts and replicates within 7 DPI. This dataset's richness in data made it ideal for training our model. For model validation, we chose the dataset by De Micheli, as its time points aligned well with the dataset by McKellar. Although the dataset by Oprescu was thoroughly analyzed, it was not utilized in our model due to its fewer number of replicates and significant cell count reduction on day 2, possibly due to experimental variations. The raw cell counts averaged approximately 5600, 8000, and 6500 per sample, for McKellar, De Micheli, and Oprescu, respectively. Batch effect correction was conducted within each experimental dataset separately after QC and integration, using CCA and Harmony **(Figure 3.2)**.

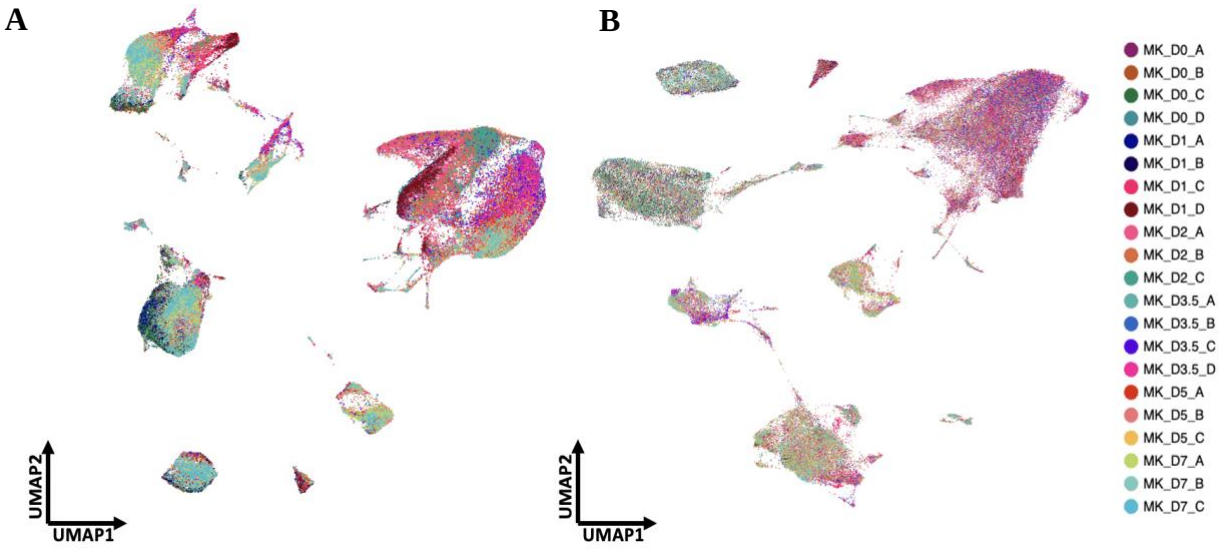


Figure 3.2. UMAP Visualizations Demonstrating the Impact of Batch Correction on scRNA-Seq Data from the McKellar Dataset. (A) The first UMAP plot represents the initial data distribution after performing PCA for dimensionality reduction, without any batch correction or clustering. **(B)** The second UMAP plot illustrates the data after applying batch correction techniques. The improved clustering of cells, indicated by a more cohesive and structured distribution, reflects the mitigation of batch effects, and highlights the underlying biological signals over technical noise. The same color scheme is maintained to represent different batches across distinct time-points.

In my initial exploration of dataset integration, I applied various batch effect correction methods, as depicted in **Figure S5**. Upon comparing techniques such as BBKNN, Scanorama, CCA, and Harmony, I observed that Harmony provided the most cohesive integration. In contrast, CCA, Scanorama, and BBKNN resulted in less overlap and the presence of multiple small, disparate clusters. These small clusters suggest incomplete integration, possibly due to the persistence of batch effects, where each method may have failed to fully reconcile inter-sample variability. For CCA, I utilized Seurat's **'RunCCA'** function. The integration of Scanorama and BBKNN was facilitated through their respective R wrappers, utilizing the '**reticulate'** package. Their packages **'Scanorama'** and **'BBKNN'** were then downloaded and applied. Harmony, however, outperformed these methods due to its faster runtime and discerning diverse cell populations. Harmony was implemented via the '**HarmonyMatrix'** function from the Harmony package. The other methods, while functional, left several small,

isolated clusters, suggesting less effective integration. In contrast, Harmony's performance in creating a more unified data representation suggests it was the most successful in mitigating batch effects while maintaining the biological integrity.

The initial dispersion of cells post-PCA, reflecting batch variation without correction is demonstrated as a UMAP plot (**Figure 3.2A)**. The second UMAP plot illustrated in **Figure 3.2** reveals a more uniform cell distribution post-batch correction. This uniformity is indicative of successful integration, where cells from different batches and time-points merge. The post-QC datasets contained about 65,000, 48,000, and 94,000 total cells respectively **(Table 3.1).** Additionally, the inclusion of multiple time points, from immediate response (D0) to later stages of repair, across all studies enabled a temporal analysis that reflects the progression and resolution of the regeneration process.

External Datasets	Age (mo.)	# raw cell counts	# cells after QC	# samples	# time points	Injury days	Injury model
Oprescu et al.	3	55,677	47,680	7	7	0, 0.5, 2, 3.5, 5, 10, 21	Cardiotoxin
McKellar et al.	20	135,823	93,954	21	6	0, 1, 2, 3.5, 5, 7	Notexin
De Micheli et al.	4-7	93,955	65,126	10	4	0, 2, 5, 7	Notexin

Table 3.1. Overview of External scRNA-Seq Datasets Used for Comparative Analysis of Muscle Regeneration.

3.4.1 Single-cell RNA-seq analysis of McKellar et al.

The selected dataset by McKellar was independently analyzed and annotated **(Figure 3.3)**. UMAP embeddings of each sample for each DPI was performed to highlight the progressive nature of skeletal muscle regeneration across time **(Figure S1, Figure S2)**. In examining the full atlas, the clusters broadly defined 22 distinct cell populations, and I further subclustered the MuSCs into 10 cell subpopulations **(Figure 3.3)**.

Temporal Dynamics of Muscle Cells in scRNA-seq

The dataset by McKellar et al. contained approximately 93,954 cells after QC, which were clustered into 22 cell types annotated with well-established marker genes **(Figure 3.3)**. Among these, cell types such as FAPS, immune cells, MuSCs, endothelial cells, smooth muscle cells, pericytes, and tenocytes**.** However, particular attention was given to MuSCs **(Figure 3.3B)** and immune cells M1 and M2 macrophages, neutrophils, and monocytes. Monocytes, for example, were identified by markers Cd14 and Ccr2[139,144,145], while M1 macrophages, associated with inflammation, were characterized by TNF and NOS2[66,140,145]. Conversely, M2 macrophages, implicated in tissue remodeling, expressed markers like Cd163 and Arg1[66]. Neutrophils were marked by specific genes like S100A8 and Cxcr2[140]. To validate the cell type annotations and the sub-clusters identified, the heatmap **(Figure 3.3A)** presents the top DEGs identified using the Wilcoxon rank sum test with a significance threshold set at a p-value of less than 0.05. Further validation of the cell types and their sub-clusters was conducted with identified cell type-specific gene expression profiles within the literature, demonstrated by the dot plot in **Figure 3.3B** where each dot's size indicates the proportion of cells expressing that gene, while the color intensity reflects the relative expression level. The marker genes for these annotations, along with their literature sources, are detailed in **Table S3** and **Figures 3.4A and 3.4B**.

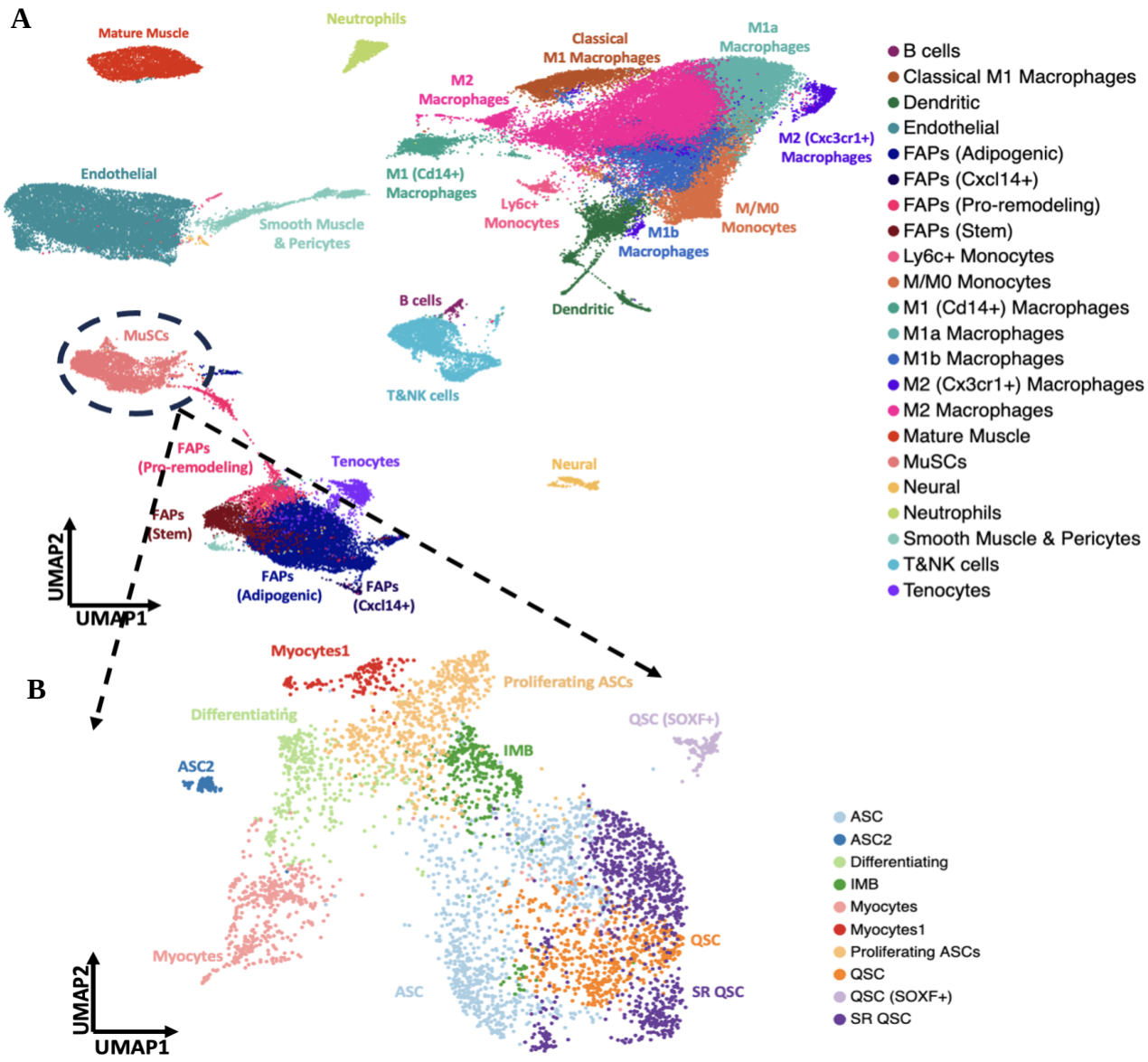


Figure 3.3. UMAP Visualization of Cellular Clusters from McKellar'a Dataset Across Muscle Regeneration Stages. (**A)** UMAP projection of 93,954 single cells from integrated samples from the dataset by McKellar et al., capturing the transcriptional diversity across identified cell populations. This analysis delineates 22 distinct clusters, each representing a unique cell population involved in muscle regeneration. Cells are color-coded based on cluster identity, illustrating the complex cellular ecosystem of the regenerative muscle tissue. Notable populations include various immune cell subsets (neutrophils, monocytes, pro-inflammatory, and anti-inflammatory macrophages), fibro/adipogenic progenitors (FAPs), and muscle stem cells (MuSCs), among others. **(B)** UMAP embedding of the MuSC subclusters displays the temporal evolution of MuSC subtypes integrated from all samples post-injury. Subtypes of MuSCs are differentiated by color, revealing the trajectory of MuSC activation, proliferation, and differentiation over the course of regeneration.

The stacked bar chart in **Figure 3.4C** illustrates the temporal changes of the cell populations, providing the relative abundance of each cell type at successive time points. Initially,

endothelial cells formed the majority at D0, constituting 50.15% of the cell populations, indicative of a rich vascular component. By D7, their proportion exhibited a biphasic pattern, decreasing to a base of 5.16% at D3.5 before returning to 20.65%, potentially reflecting a wave of vascular remodeling and adaptation to the post-intervention environment. The presence of mature muscle which initially accounted for 13.42%, is likely due to experimental procedures through tissue mincing, digestion and filtering out multinucleated myofibers during single cell isolation[135,139]. In contrast, FAPs displayed a notable pattern, starting from 20.82% and decreasing to 8.23% by D2, followed by a steady increase to 26.32% at D7. This trend suggests an initial engagement in tissue repair processes, with a subsequent expansion possibly due to their role in fibrosis or adipogenesis during later stages of tissue remodeling.

There was a transient proliferation in neutrophils at D1 (10.87%), rapidly declining thereafter, consistent with an immediate innate immune response post-injury. Monocytes showed a surge to 23.67% at D1, likely reflecting an acute inflammatory response, before declining across subsequent time points. M1 macrophages peaked at D2 (38.70%) and then gradually decreased to 5.04% by D7, while M2 macrophages showed a delayed peak at D3.5 (28.87%), which then diminished to 6.22% by D7, aligning with a transition from an inflammatory to a reparative tissue environment.

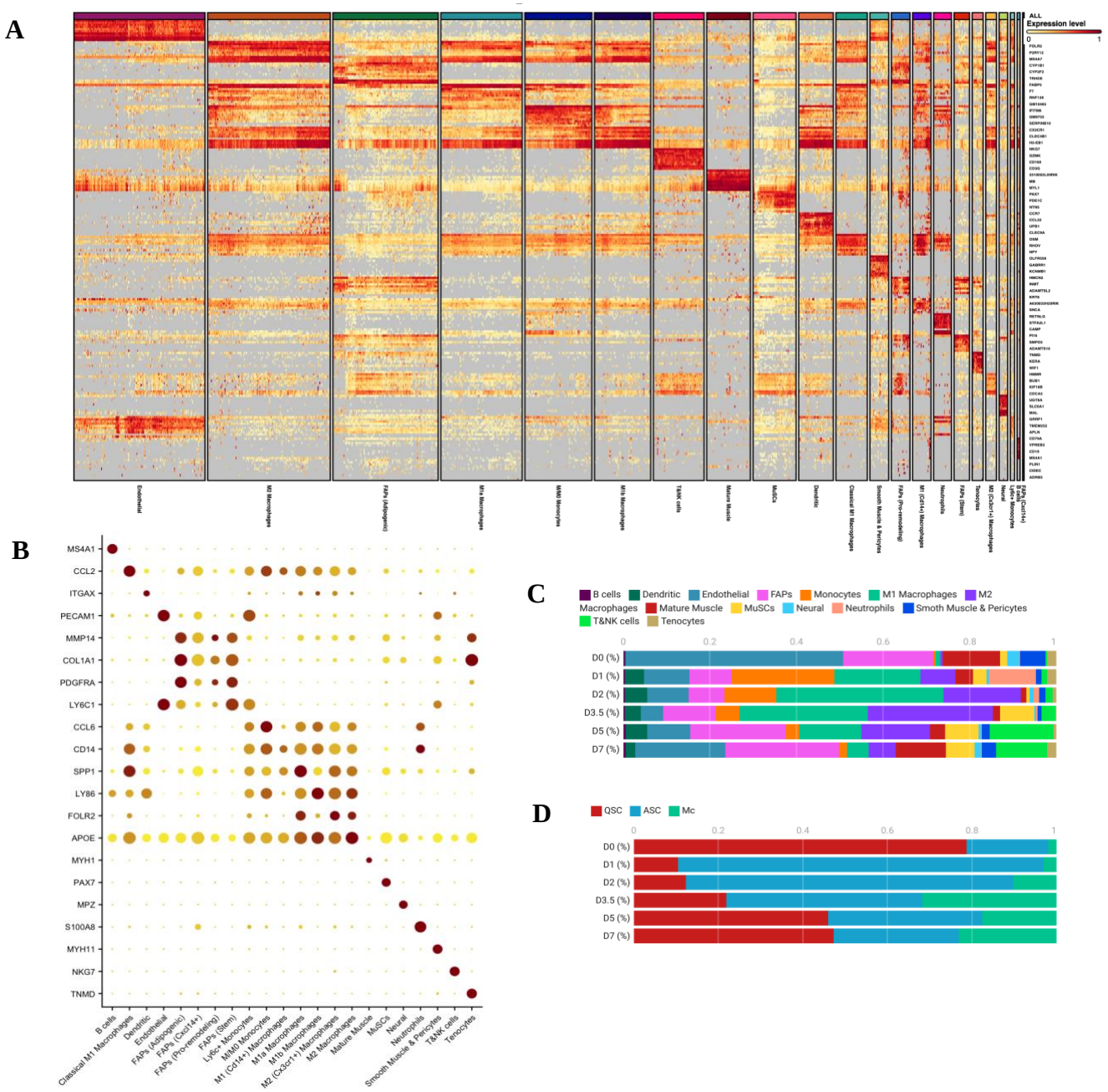


Figure 3.4 Comprehensive analysis of McKellar's dataset illustrating cell type dynamics during muscle regeneration. (A) Heatmap of the McKellar dataset displaying the top 10 differentially expressed genes (DEGs) across various cell types, with gene expression intensities represented by color gradations ranging from low (yellow) to high (red). **(B)** Dot plot validation of marker genes identified in the literature, overlaid on the McKellar dataset. Each dot's size reflects the percentage of cells expressing that gene, and its color intensity corresponds to the gene expression level. **(C)** Bar chart of cell count proportions calculated for each DPI in the McKellar dataset and **(D)** MuSC subcluster, with bars colored according to cell type.

MuSCs, containing 4,517 total cells, have been subclustered into quiescent, activated, and differentiating states, each manually annotated by distinct gene expressions (**Figures 3.5A &**

3.5B). QSCs expressed markers Chodl and HeyL[34]. Self-renewing QSC clusters, SR-QSC, were highly enriched in Notch2/3, Dpt, Calcr, Col15a1, and Spry1[135].One distinct cluster demonstrated immune-like genes including Ctsb, Ifit1, Ifit3, Isg15, and Cxcl[135]. ASCs were characterized by markers Myod1, Sdc4, Tcf7, Carm1, and Mapk[41,146,147]. Proliferating ASCs were particularly enriched in Ki67, Rac1, Top2a, Ezh2, Birc5 and Cdk1[135]. Myocytes and differentiating clusters were enriched in Acta1, Myog, Ttn, Myh and Myl4[140].

Initially, the vast majority of satellite cells are in a quiescent state (QSCs: 78.71%), indicative of a typical resting muscle prior to injury. It is likely that the trauma of injury at D0 rapidly induced a subset of QSCs to exit quiescence, thereby artificially lowering their apparent numbers when measured at the initial time point (**Figure 3.5C**). Notably, the proportion of QSCs recovers from an initial drop (to 10.38%) back towards baseline levels (47.13% by the final time point), suggesting a restoration of the quiescent cell pool due to self-renewal processes. The relative stability in the proportion of QSCs towards the latter days may reflect a balance being struck between cell renewal and differentiation, ensuring sustained muscle regeneration capability. Upon injury, there is a dramatic shift as QSCs rapidly activate (ASCs peaking on day 1-2), entering the cell cycle to contribute to tissue repair. As the repair process progresses, the proportion of ASCs decreases (to 29.70% by the final time point), with an increase in the fraction of cells that have differentiated into mature myocytes (from 1.94% to 23.17%). This trend signifies the differentiation of ASCs to replenish the pool of functional muscle fibers as part of the healing process.

The proportions of satellite cells in varying states, as captured by single-cell RNA-seq and with the annotation approach outlined in **Section 3.1**, align well with our current understanding of muscle regeneration. The dataset effectively illustrates the expected shifts in

satellite cell populations during the regenerative process. This simultaneous quantification of numerous cell types and states offers an ideal dataset for constructing and estimating a dynamic model of muscle repair and regeneration.

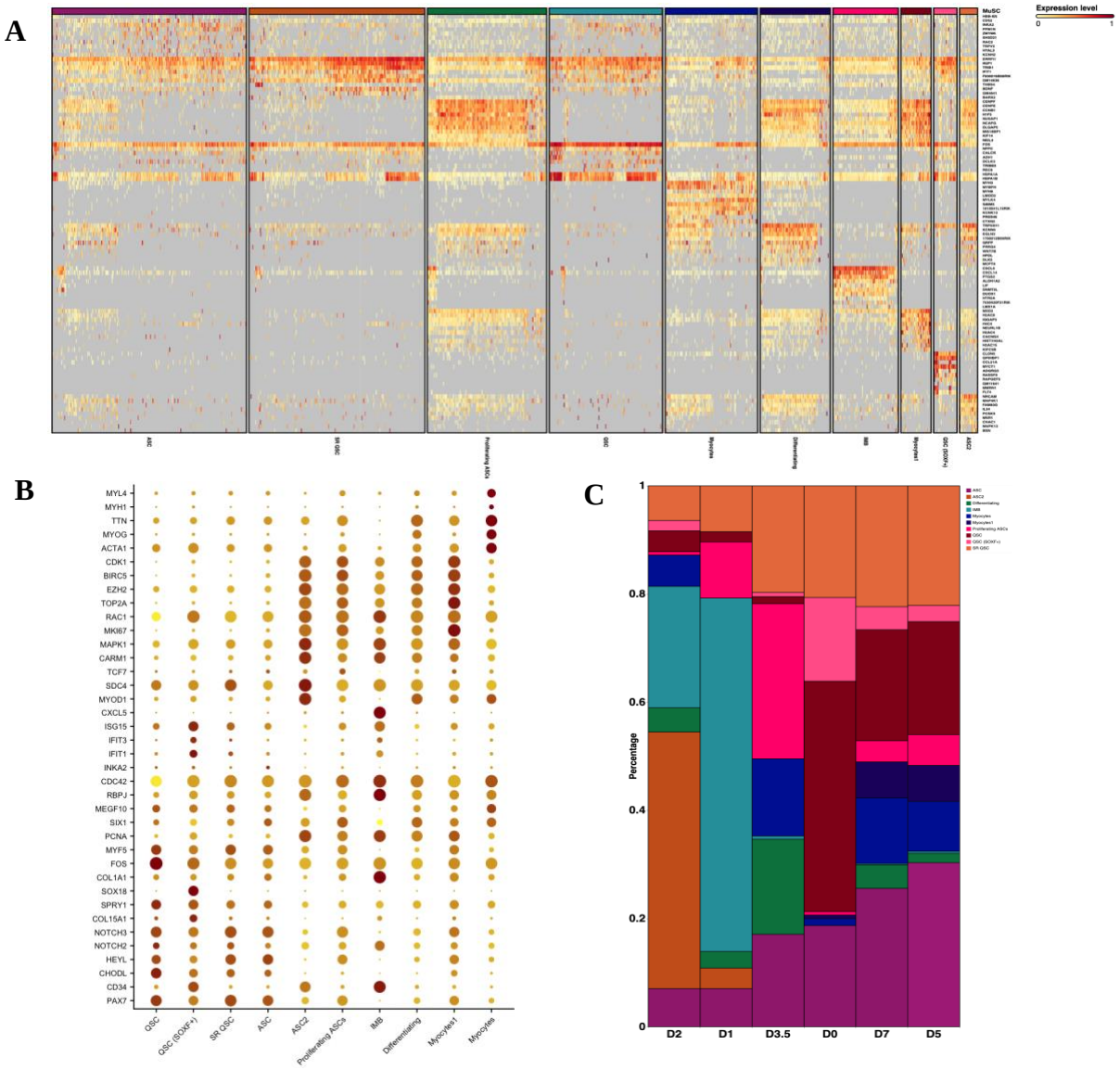


Figure 3.5. Comprehensive Analysis of MuSC Subclusters. (A) Heatmap illustrating the gene expression patterns across 10 MuSC sub-cell types, with each column representing a subcluster and rows representing genes. The color scale from yellow to red indicates low to high expression levels, respectively, with top DEGs selected based on the Wilcoxon rank sum test ($p < 0.05$). **(B)** Dot plot validating the sub-cluster-specific expression of marker genes, where the size of each dot corresponds to the proportion of cells expressing the gene, and the color intensity reflects expression level. **(C)** Proportional bar chart showing the distribution of cell counts within each MuSC subcluster across different time points post-injury, colored by sub-cell type.

3.4.2 Comparative Single-cell RNA-seq analysis of De Micheli et al.

While the dataset by McKellar et al. was used to train our model in the subsequent chapter, for the purpose of validation, I have extended the single-cell analysis to include independent time-series datasets from De Micheli et al. and Oprescu et al. The dataset by De Micheli et al., in particular, will be utilized to cross-validate that the model's predictions hold true across independent datasets (**Figure 3.6**). The results from the dataset by Oprescu et al., while not directly applied to model validation, is shown in **Figure S3** and **Figure S6** to complement our findings with additional experimental contexts. By applying the same analytical pipeline to these datasets, I aim to confirm the consistency of cellular responses and gene expression patterns across varied experimental models and injury settings.

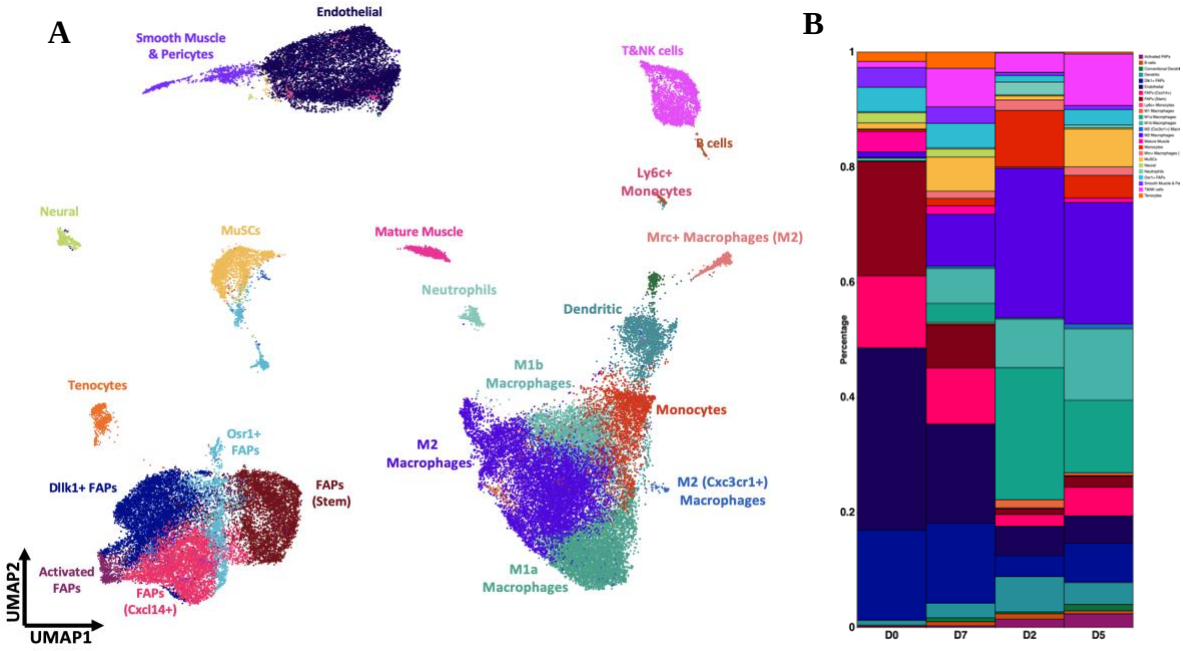


Figure 3.6. Comprehensive analysis of De Micheli's dataset illustrating cell type dynamics during muscle regeneration. (A) UMAP visualization of the dataset by De Micheli delineating 24 distinct clusters. Each cluster is color-coded to correspond with identified cell types. **(B)** Stacked bar chart showing the cell type proportions across four time points post-injury (days 0, 2, 5, 7).

In the dataset by De Micheli et al., there was a high initial percentage of FAPs at D0 (52.57%), which decreases significantly by D2 but then increases again by D7, suggesting a fluctuating role in the regenerative process. Monocytes initially comprised 0.39% of the cell population at D0, increased substantially to 9.98% by D2, and then decreased to 1.40% by D7. Similarly, this pattern suggests an early influx of monocytes to the site of injury. The M1 and M2 macrophage populations show a pronounced increase at D2, indicating an acute inflammatory response, with 32.82% and 28.02%, respectively, before decreasing (**Figure 3.6B**).

The results from both datasets corroborated our original findings, showing a relative consistency in the proportions of major cell populations such as FAPs, MuSCs, and some macrophage subsets across the different time points of muscle regeneration (**Tables S3-S5**). Notably, despite the inherent biological variability and potential differences in experimental conditions, the primary patterns of cell population dynamics remained similar. This was evidenced by the parallel shifts in populations such as endothelial cells, immune cells, and MuSCs in response to muscle injury and the subsequent repair process. These findings lend significant weight to the reproducibility of initial observations and highlight the potential universality of the cellular mechanisms governing muscle regeneration.

The immune cell dynamics within the dataset by McKellar et al. (**Figure 3.7A-D**), a pronounced peak was observed in neutrophil proportions (shown as red) at day 1 post-injury. Interestingly, the dataset by De Micheli et al. (**Figure 3.7A**) shows a peak for neutrophils at day 2, but the proportion was only half of that observed in the McKellar dataset. This discrepancy could stem from the absence of day 1 data in De Micheli's study, potentially obscuring the full neutrophil response. The trend for monocytes (shown as orange) followed a

similar pattern; however, the dataset by De Micheli et al. reported a significant 75% reduction by day 2, which might indicate an early depletion of monocytes or a possible underrepresentation in the sample. Consistency was more apparent in the proportions of M1 and M2 macrophages across both datasets, likely attributable to their sustained presence within the muscle regeneration milieu (**Figure 3.7C&D**). MuSCs also exhibited comparable dynamics, with both datasets revealing a similar pattern of depletion and subsequent partial recovery, though neither returned to the original baseline levels observed at day 0 (**Figure 3.7E**). This could be attributed to technical artifacts from the experimental protocol that may have inadvertently activated these cells, and a naturally longer timeline required for the stem cell pool to replenish.

ASCs (shown as blue) at day 2 presented comparable levels in both datasets. The lack of day 1 data in the De Micheli dataset (**Figure 3.7F**)**,** however, precludes us from confirming the rapid activation of ASCs that was evident in McKellar's data (**Figure 3.7F**)**.** For myocytes (shown as green), the dataset by McKellar et al. showed a peak at day 3.5 post-injury. In contrast, the dataset by De Micheli et al. displayed a peak at day 5, nearly 25% lower than McKellar's, suggesting that by day 5, a significant portion of myocytes might have already fused into new fibers, which is consistent with the muscle regeneration timeline. Based on these observations, it is plausible that the peak influx of myocytes occurs somewhere between days 3 and 4 post-injury.

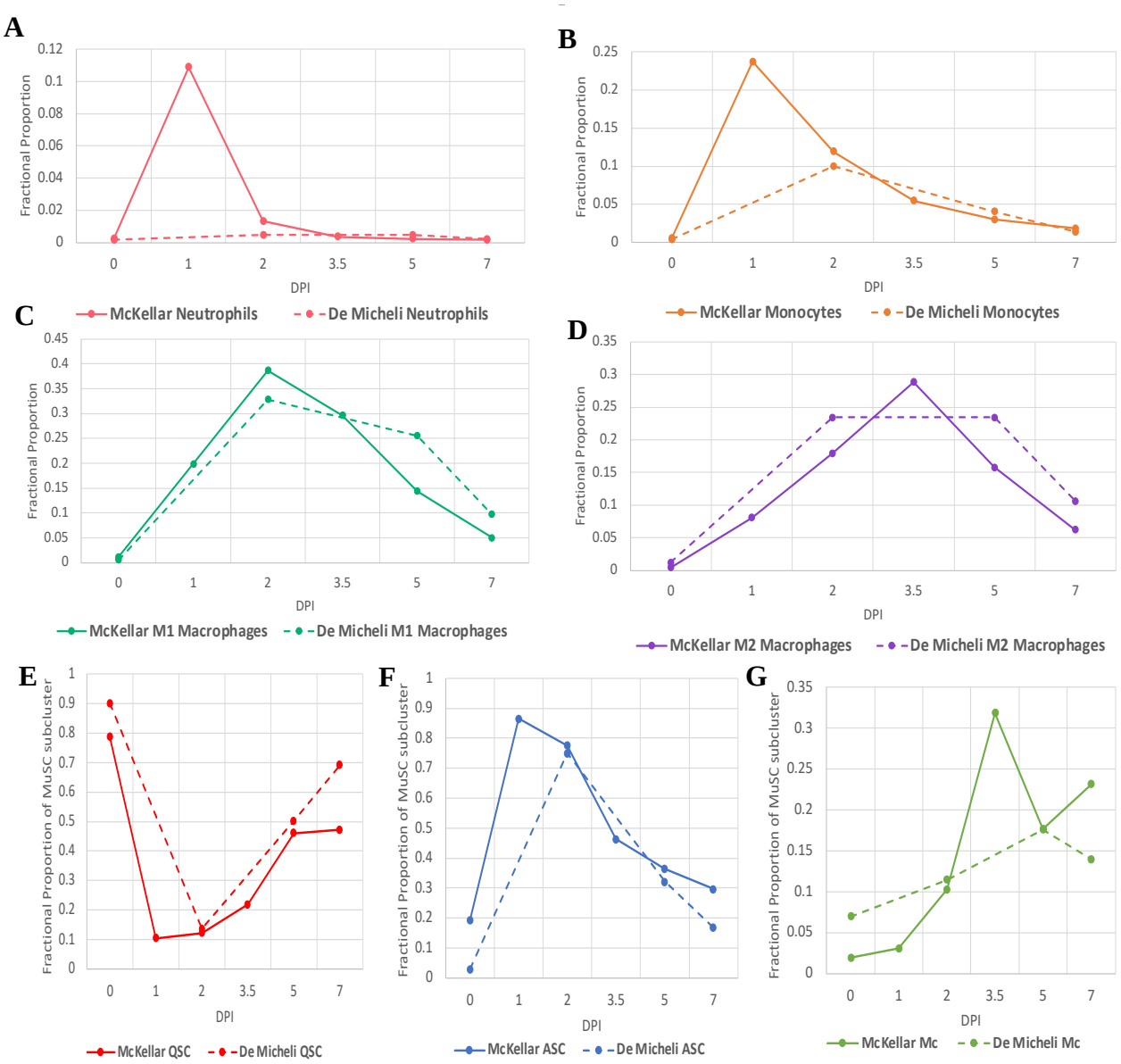
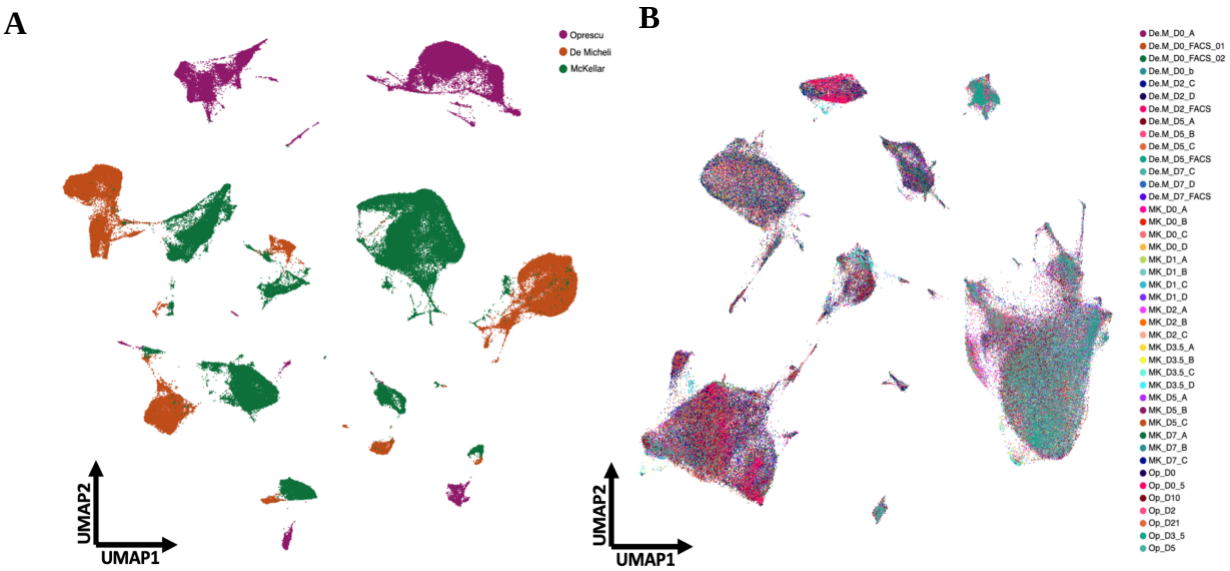


Figure 3.7. Comparative Analysis of Immune Cell Populations from De Micheli's dataset in Muscle Regeneration with McKellar's. Line graphs plotting the proportions of **(A)** neutrophils, **(B)** monocytes, **(C)** M1 macrophages, **(D)** M2 macrophages, across the time points for McKellar and De Micheli datasets, revealing similarities and differences in immune response patterns. Line graphs plotting the MuSC sub-populations of **(E)** QSCs **(F)** ASCs **(G)** Myocytes for both McKellar and De Micheli datasets, used to compare the regenerative patterns and cellular proportions between the two studies.

3.4.3 Preliminary Integration of Across scRNA-seq Sources and Challenges

To further explore the temporal patterns within an integrated dataset spanning multiple sources, I conducted a comprehensive analysis that began with the integration of scRNA-seq

data from Oprescu et al., De Micheli et al., and McKellar et al. (**Figure 3.8A**)**.** My goal was to understand if the temporal dynamics observed within individual datasets would hold when combined. Following the integration, I performed batch effect correction using the Harmony algorithm to align the datasets (**Figure 3.8B**). The corrected data showed a well-aligned distribution across the different sources, indicating a successful harmonization of the temporal patterns. With the datasets aligned, I then annotated the clusters (**Figure 3.8C**)**,** which were found to be in concordance with the previously established annotations, thereby reinforcing the reliability of our cell type identification across varied experimental conditions. Further analysis involved the creation of a heatmap (**Figure 3.8D**) to visualize the relative contribution of each cell type from the different source datasets. Notably, the Oprescu et al. dataset exhibited a pronounced representation of neutrophils, accounting for 54% of the dataset, likely due to the inclusion of samples from as early as 12 hours post-injury. Additionally, McKellar's dataset contributed more to the immune cell counts in general, which may reflect the intrinsic sample composition or nuances in sample collection and processing protocols.



D

Cell Type	McKellar	De Micheli	Oprescu
Activated FAPs	49%	23%	28%
B cells	52%	35%	14%
Conventional Dendritic Cell	39%	30%	32%
Dendritic	37%	31%	32%
Dlk1+ FAPs	31%	38%	31%
Endothelial	57%	41%	2%
FAPs (Cxcl14+)	29%	40%	31%
FAPs (Stem)	20%	57%	24%
Ly6c+ Monocytes	84%	12%	4%
M1 Macrophages	83%	9%	8%
M1a Macrophages	47%	28%	25%
M1b Macrophages	38%	30%	32%
M2 (Cxc3cr1+) Macrophages	69%	12%	19%
M2 Macrophages	44%	28%	28%
Mature Muscle	65%	14%	22%
Monocytes	52%	20%	28%
Mrc+ Macrophages (M2)	36%	32%	32%
MuSCs	56%	32%	12%
Neural	59%	39%	3%
Neutrophils	37%	8%	54%
Osr1+ FAPs	41%	37%	22%
Smooth Muscle & Pericytes	55%	35%	10%
T&NK cells	58%	37%	5%
Tenocytes	42%	30%	28%

Percentage of Cell Counts
0% 100%

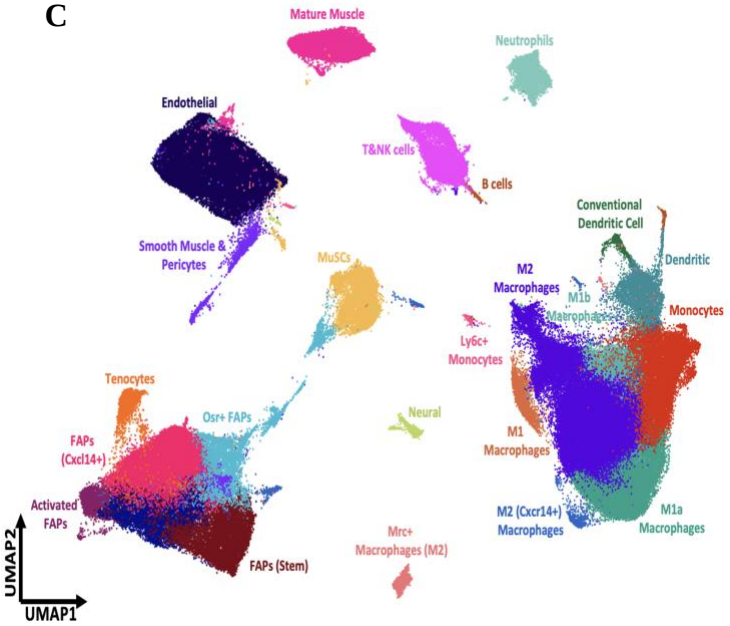


Figure 3.8. Integration and Annotation of scRNA-Seq Datasets. **(A)** UMAP representation showing the combined datasets from Oprescu, De Micheli, and McKellar before batch correction, with each dataset. **(B)** The post-batch correction UMAP visualization illustrating a harmonized distribution of cells from all three datasets, indicating successful mitigation of batch effects. **(C)** Annotated UMAP plot of the integrated dataset, where each cell type cluster is color-coded and labeled, revealing the diverse cellular landscape of muscle regeneration. **(D)** Heatmap displaying the percentage of each cell type originating from the respective source datasets, with color intensity ranging from white (fewer cells) to red (greater proportion of cells from a particular source), presenting the contribution of each dataset to the cell type composition within the integrated analysis.

3.4.4 Calibration of Cell Counts

To establish a consistent framework for comparing cell populations across time points, I calibrated the proportions of the reported datasets to standardize cell count measurements to a uniform unit of cells per cubic millimeter (cells/mm³), aligning with the three-dimensional structure of skeletal muscle tissue. **Table 3.2** displays the fractional proportions of immune cells and MuSCs relative to the total cell counts for each specific day post-injury from McKellar's single-cell data. This tabulation provides a focused view of the key cell populations central to our study's aims. For a comprehensive look at all cell type proportions from McKellar's dataset, readers are directed to **Table S3.**

Cell type	D0	D1	D2	D3.5	D5	D7
Neutrophils	0.00232	0.10873	0.01318	0.00360	0.00221	0.00160
Monocytes	0.0057	0.2367	0.1193	0.0547	0.0298	0.0184
M1 Macrophages	0.0118	0.1990	0.3870	0.2965	0.1441	0.0504
M2 Macrophages	0.0042	0.0807	0.1793	0.2887	0.1576	0.0622
QSC	0.0150	0.0056	0.0013	0.0223	0.0551	0.0458
ASC	0.0003	0.0255	0.0056	0.0324	0.0091	0.0058
Mc	0.0004	0.0006	0.0008	0.0241	0.0138	0.0147

Table 3.2. Proportions of MuSC and Immune Cells Across Time Points in McKellar's Dataset.

To translate the proportional data into absolute cell counts, literature-reported cell densities for various cell types at specific DPIs were used as a reference. To compensate for the absence of empirical data on certain DPIs, scale factors were established for each reported day based on corresponding cell counts from the literature (**Table 3.3**). These scale factors were determined for each cell type and then applied to their respective proportions as recorded in the datasets to estimate absolute cell numbers. Specifically, for each DPI with available data, the proportion of each cell type within the total cell population was multiplied by the appropriate scale factor,

reflective of literature-reported cell densities. This method allowed us to extrapolate the estimated cell counts for unreported time points, creating a continuous profile of cellular responses throughout the course of muscle regeneration. The process of scaling was applied across all cell types for each day, as evidenced by the detailed calculations presented in **Tables 3.4** and **3.5**.

DPI	Study Reference	Cell Type	Cell Count	Unit
0	Bouredji 2021, Keefe 2015, Hardy 2016, Novak 2009	SCs	2000, 2500, 2700, 5000	cells/mm^3
1	Dumont 2010, Pizza 2005	Neutrophils	3500-5500	cells/mm^3
2	Tidball 2007, Dumont 2010	M1 Macrophages	8000, 9500	cells/mm^3
3.5	Tidball 2007	Myocytes	900	cells/mm^3
5	Tonkin 2015	M2 Macrophages	7000	cells/mg
7	Bouredji 2021	M2 Macrophages	3000	cells/mm^3

Table 3.3. Literature-Based Cell Counts.

As some studies report cell densities in terms of cells per milligram (mg) of tissue rather than cells per cubic millimeter, I converted these values to a uniform volumetric density. This was accomplished by applying the established density of skeletal muscle tissue of 1.06 mg/mm³. Specifically, cell counts from Tonkin et al. regarding macrophage populations, initially provided in cells/mg, were recalculated into cells/mm³ using the formula:

$$cells/mm^3 = cells/mg \times muscle\ density\ (mg/mm^3) \tag{3.3}$$

Applying this formula, an initial count of 7000 M2 macrophages per milligram was translated into a range of approximately 7420 cells/mm³. Similarly, for two-dimensional measurements such as cells per square millimeter (cells/mm²), I converted these to cells/mm³ by considering the standard histological section thickness of 10 µm (0.01 mm):

$$cells/mm^3 = \frac{cells/mm^2}{SectionThickness(mm)} \tag{3.4}$$

Table 3.4 shows the scale factors that were assigned based on the literature-reported cell densities corresponding to each DPI. Therefore, as we established there are approximately 7,420 M2 macrophages per mm3 of tissue at day 5, and McKellar's dataset shows approximately 15.76% of this population, we estimate the total cell population for day 5 by assigning the following scale factor:

$$\text{Scale Factor} = \frac{7420\ \text{cells/mm3}}{0.1576}$$

The assigned scale factors for each DPI reflect the expected cell counts where QSCs with an estimated 2500 cells on day 0, neutrophils ranging between 3500 to 5500 cells at day 1, pro-inflammatory macrophages at approximately 8000 cells for day 2, and so forth. These values led to respective scale factors that were systematically assigned and are summarized in **Table 3.4**.

DPI	Scale Factor
0	167128
1	36788
2	20674
3.5	37303
5	47082
7	48257

Table 3.4. Assigned Scale Factors per DPI.

The results from this calibration process affirm the dynamic nature of muscle regeneration, with cell populations such as neutrophils and monocytes showing significant shifts in density that correspond to the phases of inflammation and repair. The calibrated cell counts presented in **Table 3.5**, which reflect the scaled data for each day post-injury, will serve as the foundational dataset for training our predictive model in **Chapter 4**. Furthermore, the same scale factors applied to the McKellar dataset will be used to adjust De Micheli's dataset,

providing a means to test the consistency of single-cell data population dynamics across independent studies. By doing so, I intend to validate whether the temporal patterns observed are consistent and whether they can be accurately captured within the constraints of the model. This step will not only be accompanied for model validation but also for reinforcing the reliability of single-cell data as a reflection of biological phenomena across various experimental conditions.

Cell Type	D0	D1	D2	D3.5	D5	D7
Neutrophils	389	4000	272	134	104	77
Monocytes	946	8709	2466	2040	1402	888
M1 Macrophages	1976	7321	8000	11059	6783	2432
M2 Macrophages	709	2970	3707	10769	7420	3000
QSC	2500	206	26	833	2594	2208
ASC	51	937	115	1207	429	279
Mc	68	22	16	900	651	711

Table 3.5. Scaled Cell Counts in McKellar's Dataset.

3.5 Discussion

In this chapter, I performed a comprehensive reanalysis of multiple scRNA-seq datasets to elucidate the temporal dynamics of muscle cells during regeneration. By employing a combined annotation approach, I identified distinct temporal gene expression patterns in MuSCs, delineating their states of quiescence, activation, proliferation, and differentiation. This analysis revealed significant shifts in the proportions of MuSCs and immune cells across various stages of muscle injury and repair, demonstrating the efficacy of high-throughput scRNA-seq in capturing the dynamic shifts of unique cell types within the regenerative milieu of muscle tissue. Our primary focus was on the detailed analysis of McKellar's dataset, which, with 93,954 cells post-quality control, stands as the most comprehensive dataset in our study.

This dataset not only corroborated established findings, such as the fluctuating presence of pro- and anti-inflammatory macrophages during the regeneration cycle, but also enabled the classification of MuSCs with further manual annotation. Notably, **Figures 3.5C** and **3.6E** reveal a predominantly quiescent state at the onset, followed by a rapid transition to activated and committed states. However, it should be noted that the observed proportion of 78% QSCs at day 0 might be influenced by the tissue processing methods used in single cell sequencing studies, potentially affecting the natural state of these cells.

Our comparative analysis with other datasets, like De Micheli's, revealed similar temporal patterns but with variations in the proportions of immune cells and MuSCs. This difference could be attributed to the sampling frequency and the exclusion of critical early regeneration stages in some datasets. For example, Oprescu's dataset validated the rapid influx of neutrophils at 0.5DPI, which cannot be captured in the other two datasets.

Despite the strengths of single-cell sequencing, a notable limitation lies in its inability to capture myonuclei. Future research integrating single-cell and single-nuclei sequencing could provide a more accurate understanding of both damaged and regenerated muscle. Additionally, spatial transcriptomics, which allows for the mapping of gene expression in the context of tissue architecture, has enabled the identification of transcriptionally distinct nuclei originating from the same myofiber[148–150]. Therefore, the integration of single-cell sequencing, nuclei sequencing, and spatial transcriptomics in future muscle regeneration studies could significantly enhance our understanding of the full tissue microenvironment.

Chapter 4

Mathematical Framework of Cellular Dynamics

4.1 Objective

The aim of this chapter is to construct a mathematical model of muscle regeneration informed and calibrated by scRNA-seq data. The specific focus is to develop a non-linear ODE model that accurately captures the behavior of critical cell populations involved in muscle repair, primarily myogenic lineage cells and immune cells. The main goal was to optimally fit the model to a single, normalized scRNA-seq dataset, ensuring a detailed portrayal of cellular dynamics during muscle regeneration. This method is chosen to eliminate the variability introduced by using multiple datasets mentioned in **Chapter 2**. In addition, sensitivity analysis is conducted to understand the influence of each parameter on the model, ensuring the reliability of the model against the chosen dataset.

4.2 Overview

The development of our model commenced with the formulation of a series of ODEs, reflecting the interactions between myogenic lineage cells and immune cells during muscle regeneration. This was achieved by assimilating biological knowledge from existing literature[77,151–159] with empirical data derived from scRNA-seq data annotations concerning cell types and their temporal dynamics. The equations in our model incorporate temporal changes and proportions of cell types as indicated by scRNA-seq data, which were used to adjust model parameters and their functional forms to reflect biological reality. The model parameters were initially estimated from literature sources and initial calculations of cell depletion, providing a basis for further refinement. The optimization of these parameters against scRNA-seq data was performed using NLLS approach, aimed at minimizing the difference between model predictions and observed data.

4.3 Methods

4.3.1 Biological Model Assumptions

The ODE system formulated here (**Figure 4.1**) encapsulates the interactions among ten primary cell types including damaged myonuclei (Md), healthy myonuclei (Mn), neutrophils (N), apoptotic neutrophils (Nd), monocytes (M), classically activated macrophages (M1), alternatively activated macrophages (M2), quiescent satellite cells (QSC), activated satellite cells (ASC), and myocytes (Mc). It is a representation of both cellular transformation and interaction processes, where each cell type can transform into another, or modulate the behavior of other cells. To simulate the onset of this biological system, specific initial conditions are established for the model. Considering the context of the model, where certain

cell types are not immediately present in significant numbers at the start of the simulation, the initial conditions for all immune cells (N, Nd, M, M1, M2) and MuSCs (ASC, Mc) are set to zero. This assumption is based on the premise that these cells do not infiltrate the system immediately or are present in negligible numbers initially. Conversely, the initial condition for QSCs is set to 2700 cells, representing their baseline presence in the muscle tissue. Additionally, 30000 Md are accounted for at day 0, indicating the starting state of muscle tissue prior to degeneration.

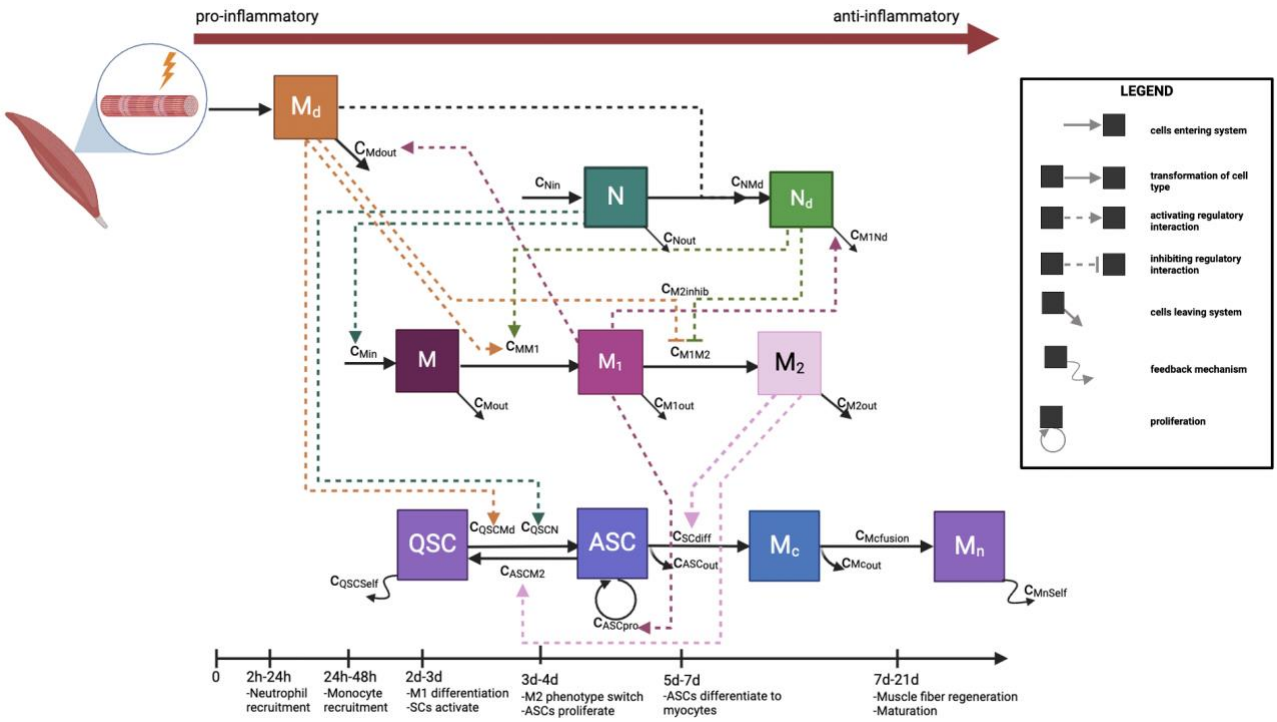


Figure 4.1 Schematic Representation of Cellular Dynamics in Muscle Regeneration. This flowchart delineates the interactions among various cell types involved in the muscle regeneration process following injury. The diagram illustrates the sequential recruitment and transformation of cells, starting from damaged myonuclei (Md) triggering an inflammatory response. The first responders, neutrophils (N), migrate to the injury site and transition into apoptotic neutrophils (Nd), which are subsequently cleared by classically activated macrophages (M1). Monocytes (M) are recruited by neutrophils and rapidly differentiate into M1 macrophages, which then transition into alternatively activated macrophages (M2) post-phagocytosis. Quiescent satellite cells (QSC) activate to activated satellite cells (ASC) in response to the damage, and under the influence of M2 macrophages, they can differentiate and fuse to form myocytes (Mc) or revert to a quiescent state, contributing to the pool of QSC. Healthy myonuclei (Mn) arise from the fusion of myocytes, completing the muscle repair process. Solid arrows represent transformations and transitions between cellular states, dashed arrows indicate regulatory influences, and squared arrows signify feedback mechanisms. The rough time points and corresponded major events after injury are indicated at the bottom of the timeline.

The system presented is based on the mass-action principle of chemical kinetics and is developed from a literature-based understanding of the biological context surrounding muscle stem cells and their niche[11,153–155,160–166]. Each equation's structure is reflective of the biological processes, such as chemotaxis, phagocytosis, differentiation, proliferation, apoptosis, and cell fusion. The core assumptions of our model are as follows.

Inflammatory Response Initiation: Upon muscle injury, the presence of Md is postulated to initiate an inflammatory response. This migration is driven by chemotactic factors, including pro-inflammatory cytokines, growth factors, and metabolites, released at a constant rate from the injury site, consistent with established chemotactic models [64,167,168].

Neutrophil recruitment and activity: Neutrophils are assumed to phagocytose the Md and subsequently secrete cytokines that trigger a cascade of cellular responses essential for orchestrating muscle regeneration[169]. The influx of neutrophils into the damaged area is modelled to be proportional to the severity of muscle damage, aligning with empirical observations of inflammation intensity corresponding to injury magnitude.

Monocyte to Macrophage Polarization: Following neutrophil recruitment, monocytes are mobilized and their transition to the M1 phenotype occurs typically within 24 hours[60]. The concentration of M1 macrophages peaks approximately two days post-injury, parallel with the decline of neutrophils, facilitating the clearance of remaining Md and Nd[170].

Macrophage-mediated regeneration: We incorporate the hypothesis that macrophages play a biphasic role in regeneration, where the M1 macrophages are assumed to foster the proliferation of ASCs while inhibiting their differentiation. Conversely, M2 macrophages are modeled to suppress proliferation but encourage differentiation and fusion[61,64,153,171].

Phenotypic Switching of Macrophages*:* Upon completion of phagocytosis and clearance of cellular debris, M1 macrophages undergo a phenotypic switch to alternatively activated M2 macrophages. This transition is assumed to occur *in situ* and facilitates the differentiation and fusion of satellite cells.

Satellite Cell Dynamics: In the absence of injury, satellite cells maintain a quiescent state within the muscle fiber. However, our model assumes that muscle damage triggers rapid satellite cell proliferation, a response that can occur independently but is amplified by the presence of M1 macrophages. The model also considers that myocytes, derived from ASCs, align and fuse under the influence of M2 macrophages to regenerate healthy muscle fibers.

Termination of the Regenerative Process: If no further damage occurs, M2 macrophages are assumed to exit the system by either dying or migrating to other regions, reflecting the self-limiting nature of the regenerative response. This assumption aligns with observations that the resolution of inflammation coincides with the cessation of active regeneration and the re-establishment of muscle homeostasis.

4.3.2 Mathematical Derivation of Equations

Each cell type population is represented by a variable and their change over time is expressed using an ODE. Rates of cell activities, such as their infiltration or coming into existence (influx), transforming into another type (differentiation/polarization), apoptosis or leaving (outflux), and affecting each other (signaling), are determined by rate constants represented by *'c'*, as well as numbers of cells performing or regulating the activity.

It should be noted that due to the limitations of scRNA-seq in capturing apoptotic neutrophils, damaged myonuclei, and regenerating myonuclei, empirical data for these variables were not available. Consequently, the data fitting process focused on cell types where empirical data were obtainable (N, M, M1, M2, QSC, ASC, and Mc). The fitting of these variables effectively informed the dynamics of the other model variables, for which empirical

data were not directly measured. By following the biological assumptions in **Section 4.3.1**, each step and equation in the ODE system is explained below.

Damaged Myonuclei Dynamics and Neutrophils (ODE 1&2):

The rate of change of Md (**Eq. 4.1**), which also signifying their clearance rate, is determined by the presence of M1 at the rate of $c_{Md_{out}} \cdot Md \cdot M1$, and the presence of neutrophils at the rate of $c_{NMd} \cdot Nd \cdot N$. To derive the rate of change of N, we consider three processes: the influx of neutrophils into the damaged area, their natural death or exit from the system, and their activity in phagocytosing Md (**Eq. 4.2**). The influx is assumed to be proportional to the amount of damage, represented by Md. The linear term $c_{N_{in}} \cdot Md$ represents the influx of neutrophils and the constant $c_{N_{in}}$encapsulates the rate at which neutrophils are recruited in response to chemotactic signals released by the damaged muscle. A first-order decay process is termed $c_{N_{out}} \cdot N$ which models the natural decay or exit of neutrophils from the system. The bilinear term $c_{NMd} \cdot Md \cdot N$ models the neutrophil activity in phagocytosis of Md. This synergistic interaction not only reduces the population of damaged cells but also reduces the neutrophil population itself. These three processes combine to generate the neutrophil dynamics equation:

$$\frac{\mathrm{dMd(t)}}{\mathrm{dt}} = -\,\mathrm{c_{Md_{out}}} \cdot \mathrm{Md} \cdot \mathrm{M1} - \mathrm{c_{NMd}} \cdot \mathrm{Md} \cdot \mathrm{N} \tag{4.1}$$

$$\frac{\mathrm{dN(t)}}{\mathrm{dt}} = \mathrm{c_{N_{in}}} \cdot \mathrm{Md} - \mathrm{c_{N_{out}}} \cdot \mathrm{N} - \mathrm{c_{NMd}} \cdot \mathrm{Md} \cdot \mathrm{N} \tag{4.2}$$

Apoptotic Neutrophil Dynamics (ODE 3):

To derive the rate of change for apoptotic neutrophils (**Eq. 4.3**), we consider their formation from N through phagocytosis and the clearance by M1 macrophages. This conversion, directly dependent on the phagocytosis activity, is modeled by the bilinear interaction in **Eq. 4.2**. Their

clearance by M1 macrophages is modeled with a second interaction, dependent on both their population and the population of M1, with $c_{M1Nd} \cdot M1 \cdot Nd$ representing the rate of this clearance process.

$$\frac{\mathrm{dNd(t)}}{\mathrm{dt}} = \mathrm{c_{NMd}} \cdot \mathrm{Md} \cdot \mathrm{N} - \mathrm{c_{M1Nd}} \cdot \mathrm{M1} \cdot \mathrm{Nd} \tag{4.3}$$

Monocyte Dynamics (ODE 4):

We consider three processes for monocyte dynamics: their recruitment to the injury site, their transformation into M1 macrophages, and natural exit from the system (**Eq. 4.4**). The recruitment to the site of injury is driven N, assuming a direct, proportional relationship, modeled by the linear term $c_{M_{in}} \cdot N$. They then polarize into M1 macrophages. This transformation is dependent on both the presence of Nd and/or Md, leading to the term $c_{MM1} \cdot M(Nd + Md)$ reflecting the influence of apoptotic neutrophils and damaged muscle on monocytes. The natural exit of monocytes from the system is given by $c_{M_{out}} \cdot M$.

$$\frac{\mathrm{dM(t)}}{\mathrm{dt}} = \mathrm{c_{M_{in}}} \cdot \mathrm{N} - \mathrm{c_{MM1}} \cdot \mathrm{M(Nd + Md)} - \mathrm{c_{M_{out}}} \cdot \mathrm{M} \tag{4.4}$$

Macrophage Dynamics (ODEs 5 & 6):

The dynamics of M1 and M2 macrophages are derived by considering their transformation from monocytes, transition between state, and exit from the system (**Eq's. 4.5** & **4.6**). M begin to differentiate into M1 influenced by signals from dead myonuclei or dead neutrophils. The progression from M1 to M2 macrophages incorporates Michaelis-Menten kinetics, indicating that this transition is modulated by the presence of signals from dead myonuclei or neutrophils and is subject to saturation effects or inhibitory mechanisms. M1 are subsequently set to exit the system at a rate of $c_{M1_{out}}$. Subsequent polarization to become M2 is represented by the

term $\frac{c_{M1M2} \bullet M1}{c_{M2_{inhib}} + Nd + Md}$ with an inhibition factor preventing the differentiation of M2 macrophages until clearance of Md and Nd is complete by M1. M2 macrophages exit the system at a rate of $c_{M2_{out}} \bullet M2$. This results in the following equations.

$$\frac{dM1(t)}{dt} = c_{MM1} \bullet M(Nd + Md) - c_{M1} c_{M1_{out}} \bullet M1 - \frac{c_{M1M2} \bullet M1}{c_{M2_{inhib}} + Nd + Md} \quad (4.5)$$

$$\frac{dM2(t)}{dt} = \frac{c_{M1M2} \bullet M1}{c_{M2_{inhib}} + Nd + Md} - c_{M2_{out}} \bullet M2 \quad (4.6)$$

Satellite Cell Dynamics (ODE 7 & 8):

The satellite cell dynamics incorporate both self-renewal and differentiation processes. The bilinear term $c_{QSCN} \bullet QSC \bullet N$ represents the conversion of QSCs as they activate to ASCs in response to the infiltration of N. The presence of Md also contributes to the depletion of QSCs represented by the rate $c_{QSCMd} \bullet Md \bullet QSC$. The process of QSC self-renewal must also be regulated to prevent overaccumulation of QSCs. This is achieved through a negative feedback loop where QSCs are reduced when exceeding a certain threshold, modeled by the quadratic term $c_{QSC_{self}} \bullet QSC(QSC - QSC_{max})^{+}$. The "+" signifies that this term is only active when their quantity exceeds the set threshold, QSC_{max}. The product $(QSC - QSC_{max})$ creates a condition where the self-renewal rate is proportional to the current QSC population but decreases as the number of QSCs approaches QSC_{max} (**Eq 4.7**). M2 macrophages are hypothesized to release signaling factors encouraging self-renewal and differentiation at a rate of $c_{ASCM2} \bullet M2 \bullet ASC$ and $c_{SC_{diff}} \bullet M2 \bullet ASC$, respectively. ASCs are converted in proportion to the QSC depletion; however, their subsequent proliferation is influenced in response to proinflammatory signals

by the rate of $c_{ASC_{pro}} \cdot M1 \cdot ASC$. ASC cells that were not subjected to differentiation or self-renewal will either exit or are subject to apoptosis by the rate $c_{ASC_{out}} \cdot ASC$ (**Eq. 4.8**).

$$\frac{\mathrm{dQSC(t)}}{\mathrm{dt}} = -\mathrm{c_{QSCN}} \cdot \mathrm{QSC} \cdot \mathrm{N} - \mathrm{c_{QSCMd}} \cdot \mathrm{Md} \cdot \mathrm{QSC} + \mathrm{c_{ASCM2}} \cdot \mathrm{M2} \cdot \mathrm{ASC} - \mathrm{c_{QSC_{self}}} \cdot \mathrm{QSC}\,(\mathrm{QSC} - \mathrm{QSC_{max}})^{+} \quad (4.7)$$

$$\frac{\mathrm{dASC(t)}}{\mathrm{dt}} = \mathrm{c_{QSCN}} \cdot \mathrm{QSC} \cdot \mathrm{N} + \mathrm{c_{QSCMd}} \cdot \mathrm{Md} \cdot \mathrm{QSC} - \mathrm{c_{ASCM2}} \cdot \mathrm{M2} \cdot \mathrm{ASC} + \mathrm{c_{ASC_{pro}}} \cdot \mathrm{M1} \cdot \mathrm{ASC} - \mathrm{c_{SC_{diff}}} \cdot \mathrm{M2} \cdot \mathrm{ASC} - \mathrm{c_{ASC_{out}}} \cdot \mathrm{ASC} \quad (4.8)$$

Myocyte and Myonuclei Dynamics (ODEs 9 & 10):

The myocyte (Mc) and myonuclei (Mn, Md) equations are derived considering the differentiation and fusion processes. The bilinear term $c_{SC_{diff}} \cdot M2 \cdot ASC$ models the differentiation of ASC into myocytes, and $c_{Mc_{fusion}} \cdot Mc$ captures the fusion of myocytes into healthy myonuclei (**Eq. 4.9**). The term $c_{Mn_{self}} \cdot Mn(Md - 3000)^{+}$ is a self-regulation feedback mechanism for the myonuclei population. $c_{Mn_{self}}$ is the rate constant for regulation and $(Mn - 30000)$ sets a threshold population level, beyond which the regulation kicks in to maintain tissue homeostasis (**Eq. 4.10**). The number 30,000 is based on biological data from the literature[158,159] indicating the upper limit of myonuclei in a given muscle volume to avoid hypertrophy[172]. Thus, when Mn is less than 30,000 this term leads to a negative value for the entire expression, implying that the population of myonuclei is below the threshold and there is room for growth or maintenance without risking hypertrophy. Conversely, when Mn exceeds 30,000, this term becomes positive, representing a regulatory mechanism that slows down or stops the production of myonuclei, thereby preventing excessive growth. This implies that at

precisely Mn = 30,000, the term becomes zero, suggesting that the population is at an ideal or balanced level, and no further self-regulation is needed.

$$\frac{dMc(t)}{dt} = c_{SC_{diff}} \cdot M2 \cdot ASC - c_{Mc_{fusion}} \cdot Mc - c_{Mc_{out}} \cdot Mc \quad (4.9)$$

$$\frac{dMn(t)}{dt} = c_{Mc_{fusion}} \cdot Mc - c_{Mn_{self}} \cdot Mn(Md - 3000)^{+} \quad (4.10)$$

4.3.3 Parameter Estimation

The initial values of our model's parameters are set based on a confluence of existing theoretical knowledge, empirical data, and established values from relevant literature[151,157,159]. Focusing on the kinetics of cell recruitment and decay in response to myonuclei damage, we specifically examine neutrophils as the primary infiltrating cells. For estimating parameters like $c_{N_{out}}$, which represent the natural exit rate of neutrophils independently of their interaction with Md, we apply an exponential decay model:

$$N(t) = N_0 e^{-kt} \quad (4.11)$$

$$t_k = \frac{\ln(2)}{k} \quad (4.12)$$

where N_0 is the initial quantity of *N*, *N*(*t*) is the quantity at time *t*, and *k* is the decay constant. By assuming that *N* reaches around 4000 day 1 and depletes to zero by approximately day 2, the decay constant using the half-life formula is calculated as follows.

$$\text{Half-life of N} = \frac{\text{Time for cell exit}}{2}, k = \frac{\ln(2)}{\text{Half-life of N}} \quad (4.13)$$

This approach yielded an initial estimate for $c_{N_{out}}$ at a value of around 1.4 per day. The same approach was applied to other parameters representing the rate under the influence of a single

variable, such as $c_{M1_{out}}, c_{M2_{inhib}}, c_{M_{out}}$, $c_{Mc_{out}}$, and $c_{ASC_{out}}$. Parameters representing regulatory interactions were initially estimated through manual iterative refinement, informed by expected outcomes guided from published experimental data. By systematically varying parameter values and observing the impact on model behavior, I identified a set of preliminary parameter estimates and their ranges that were consistent with the empirical data. The resulting initial parameter from estimates, which serve as a starting point for optimization, are as follows.

Parameter	Range of Initial Estimates
$\mathbf{c_{N_{in}}}$	[0.5 - 1.5]
$\mathbf{c_{N_{out}}}$	[1 - 3]
$\mathbf{c_{M1Nd}}$	[8.1e-05 - 1e-04]
$\mathbf{c_{M_{in}}}$	[15 - 30]
$\mathbf{c_{MM1}}$	[4e-05 - 6e-05]
$\mathbf{c_{M1_{out}}}$	[0.09 - 0.19]
$\mathbf{c_{M2_{inhib}}}$	[9e-17 - 1.1e-16]
$\mathbf{c_{M2_{out}}}$	[0.10 - 0.25]
$\mathbf{c_{QSCN}}$	[2.4e-06 -7.4e-06]
$\mathbf{c_{ASCM2}}$	[2.3e-05 - 3.3e-05]
$\mathbf{c_{ASC_{pro}}}$	[6.0e-05 - 9.5e05]
$\mathbf{c_{SC_{diff}}}$	[3e-04 - 5e-04]
$\mathbf{c_{Mc_{fusion}}}$	[2.5 - 3.5]
$\mathbf{c_{Md_{out}}}$	[1e-04 - 2e-04]
$\mathbf{c_{NMd}}$	[1e-05 - 2e-05]
$\mathbf{c_{M_{out}}}$	[5 - 10]
$\mathbf{c_{M1M2}}$	[450 - 550]
$\mathbf{c_{QSCMd}}$	[1e-04 - 3.5e-04]
$\mathbf{c_{QSC_{self}}}$	[0.08 - 0.1]
$\mathbf{c_{Mc_{out}}}$	[1.5e-05 - 5.5e-05]
$\mathbf{c_{Mn_{self}}}$	[9e-05 - 1e-04]
$\mathbf{c_{ASC_{out}}}$	[1e-04 - 2e-04]

Table 4.1. Range of Initial Parameter Estimates.

To refine these estimates, I utilize the L-BFGS-B algorithm implemented within a NLLS framework[119,173]. Our objective is to minimize the sum of squared errors (SSE) between observed data y_i^{obs} and model predictions $y_i^{mod}(\theta)$ normalized by the maximum observed value y_{max}^{obs}, where θ represents the vector of parameters, ensuring a dimensionless objective function, which aids in the comparability of residuals across different scales.

$$J(\theta) = \sum_{i=1}^{n} \left(\frac{y_i^{obs} - y_i^{mod}(\theta)}{y_{max}^{obs}} \right)^2 \text{ such that: } c_i \geq 0 \quad (4.14)$$

Lower bounds were set to 0 and upper bounds of infinity for all constants to maintain biological plausibility. Our optimization process utilized the '**optim'** function in R, which minimizes an objective function, **objective_func**, over the parameter space. The optimization step was an iterative process implemented through a loop structure. This loop executed the following steps for each set of initial parameter guesses.

1. **Parameter Feeding:** Initial parameter guesses are inputted into the optimization function.
2. **Model Fitting:** The **`optim`** function calls **`objective_func`**, which integrates the ODE model over the specified time sequence using the current set of parameters. This step generates model predictions for each state variable.
3. **Error computation:** **`objective_func`** then calculates the total error as the SSE between the model predictions and the observed data for all state variables.
4. **Result Evaluation:** Upon completion of each iteration, the total error is evaluated. If this error is lower than that from any previous iteration, the current set of parameters is considered the best fit so far.

5. **Iteration and output:** The loop tests new parameter guesses until convergence, finalizing the set with the lowest total error.

In summary, I define the initial parameter guesses, construct an objective function for SSE calculation, set parameter bounds, and employ the L-BFGS-B method for iterative parameter adjustment. The method is described in **Algorithm 1.**

Algorithm 1: Pseudo-code for parameter optimization.

```
Inputs:
 - initial_guesses: Initial parameter guesses from initialize_parameters()
 - observed_data: Loaded observed data
 - initial_state: Defined initial state of state variables
 - time_sequence: Defined time sequence for the model
 - bounds: Lower and upper bounds for each parameter

Function initialize_parameters()
  - Define and return initial parameter guesses
End Function

Function run_ODE_model(params, initial_state, time_sequence)
  - Simulate the model using the ODE solver 'lsoda' over the time sequence
  - Return model outputs for each state variable
End Function

Function calculate_normalized_SSE(model_output, observed_data)
  - Initialize normalized_SSE to 0
  - For each state variable:
     - Compute the squared difference between model output and observed data
     - Normalize the squared differences by the maximum observed value for the state variable
     - Accumulate normalized squared differences to normalized_SSE
  - Return normalized_SSE
End Function

Function objective_function(params)
  - Run the model with run_ODE_model() and store outputs
  - Calculate normalized SSE with calculate_normalized_SSE()
  - Return normalized_SSE
End Function

Function optimize_parameters(initial_guesses, bounds)
  - Initialize variables for the best result and best total error
  - For each initial guess in initial_guesses:
     - Call L-BFGS-B with objective_function, initial guess, and bounds
     - Compute total_error using objective_function with current parameters
     - If total_error is less than the best_total_error:
        - Update best_result and best_total_error with the current result and error
  - Return best_result.parameters and best_total_error
End Function

  #Main Execution:
  - Execute optimize_parameters() with initial_guesses and bounds
  - Output "Optimized Parameters: ", best_parameters
  - Output "Best Total Error: ", best_error
```

The optimized values are obtained using an R function that calculates the total error across variables with empirical data, including N, M, M1, M2, QSC, ASC, and Mc. The output parameters from the optimization were employed with the **'lsoda'** solver from the **'deSolve'** package. Data visualization was conducted using the **ggplot2** package.

To visually assess the model fit, I generated plots for each variable in the model. The simulated and empirical observations were first formatted into a long data structure in R where each row corresponded to a specific time point and variable. Residuals were then calculated by taking the difference between the simulated values and the empirical observations at the corresponding time points (days 0, 1, 2, 3.5, 5, 7). These residuals served as indicators of the model's accuracy, with smaller residuals implying a closer fit to the experimental data. To assess the uncertainty around our model's estimates, I calculated the standard deviation of the residuals for each variable across all time points (**Eq. 4.16**)

$$Residual = Simulated\ Value - Experimental\ Value \tag{4.15}$$

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{4.16}$$

where x_i is each residual, $\bar{x}$ is the mean of the residuals, and N is the number of data points. The SD is calculated for each variable that was fitted to the empirical data (N, M, M1, M2, QSC, ASC, Mc). To assess the model's prediction accuracy, a smaller SD indicates that the model predictions are consistently closer to the observed data, suggesting a better fit for that particular variable. The SD was then used to create a shaded band around the simulated values which represent the range within most of the observed data points are expected to fall, based on the model's predictions. The method is described in **Algorithm 2.**

Algorithm 2: Pseudo-code for residuals

```
Inputs:
  - initial_conditions: Starting state values for the ODE model
  - optimized_parameters: Parameter values optimized from the previous analysis
  - time_sequence: Time points at which the model output is evaluated
  - experimental_data: Observed data to compare against the model simulation

Function simulate_ODE_Model(initial_conditions, optimized_parameters, time_sequence)
  - Simulate the ODE model using the 'lsoda' solver over the time sequence with given parameters
  - Output: Return simulation results as a data frame
End Function

Function prepare_experimental_data(experimental_data)
  - Format the experimental data into a long format for comparative analysis
  - Output: Return the formatted experimental data
End Function

Function calculate_Residuals(simulation_results, formatted_experimental_data)
  - Merge simulation and experimental data on time and variable
  - Calculate the residual for each data point (simulated value minus experimental value)
  - Output: Return residuals as a data frame
End Function

Function compute_standard_deviation(residuals)
  - Group residuals by variable
  - Calculate the standard deviation of residuals for each variable
  - Output: Return the standard deviations as a data frame
End Function

Function create_plots(simulation_results, formatted_experimental_data, std_deviation)
  - For each variable, create a plot displaying both simulated and experimental values
  - Add variability ribbons to the plots using the standard deviation values
  - Output: Return a list of plots
End Function

Function display_plots(plots)
  - Combine and arrange the individual plots into a grid layout
  - Display the plot grid
End Function

#Main Execution:
- Call simulate_ODE_Model with initial conditions, optimized parameters, and time sequence
- Call prepare_experimental_data with experimental data
- Call calculate_Residuals with simulation results and formatted experimental data
- Call compute_standard_deviation with residuals
- Call create_plots with simulation results, formatted experimental data, and standard deviation
- Call display_plots with the list of plots
```

The individual plots for each variable were then combined into a comprehensive grid layout using the **'patchwork'** package[174]. The implementation of these visualization techniques was conducted using the **'tidyverse'**[175] and **'deSolve'** packages in R. These simulations and

visualizations were the main components of our model validation process, ensuring that our model accurately represents the underlying biological processes represented in the literature.

4.3.4 Sensitivity Analysis

To evaluate the robustness and responsiveness of our model, I performed a detailed sensitivity analysis. This analysis involved systematic perturbations of model parameters to understand their impact on the behavior of variables. First, I quantified the average sensitivity of each variable to ±10% perturbations in model parameters. Specifically, each parameter p ranged from $0.9 \times p \; to \; 1.1 \times \; p$. The ODE system was solved for both the perturbed parameter set $\theta_{perturbed}$ and the baseline parameter set, $\theta_{baseline}$.

$$Output_{perturbed} = ODEsystem(\theta_{perturbed}) \tag{4.17}$$

$$Output_{baseline} = ODEsystem(\theta_{baseline}) \tag{4.18}$$

Therefore, the sensitivity S of each variable v to the perturbation of parameter p was quantified using the mean of the absolute relative change:

$$S(v,p) = \; mean\left(\frac{|\,Output_{perturbed}(v) - Output_{baseline}(v)\,|}{|\,Output_{baseline}(v)\,|}\right) \tag{4.19}$$

The sensitivity values were visualized in a heatmap, with variables as rows, parameters as columns, and cell color intensity indicating the sensitivity level. To visualize the distribution of the relative changes in variables in response to ±10% perturbations, box plots were created for each variable to showcase the spread and variation of the relative changes.

To examine the dynamics of each variable's sensitivity to parameter perturbations over time, a similar method to the aggregate sensitivity heatmap was conducted, but sensitivity was assessed at each time point t in the model simulation. For each time point, the relative change due to parameter perturbation was calculated:

$$Relative\ Change\ (v,p,t) = \frac{Output_{perturbed}(v,t) - Output_{baseline}(v,t)}{Output_{baseline}(v,t)} \quad (4.20)$$

The relative changes were collated for each variable and a heatmap was generated for each variable, showcasing the sensitivity across time (x-axis) against different parameters (y-axis). Furthermore, I conducted specific evaluations focusing on individual parameters' effects on particular state variables. This was achieved by perturbing one parameter at a time and observing its impact on a chosen state variable across the simulation period. The perturbation levels were maintained within ±10% of the original parameter values, and the results were visualized using line plots to depict temporal dynamics. The method is described in **Algorithm 3.**

The sensitivity analysis was implemented in R, utilizing the **'deSolve'** package for solving the ODE system. Data reshaping and visualization were handled using **'reshape2'** and **'ggplot2'**, respectively. **'reshape2'** was utilized for reorganizing simulation data, making it suitable for subsequent analysis. **'ggplot2'** aided in creating visual outputs such as heatmaps and line plots. The analysis included both overall and time-specific sensitivity assessments, implemented through a custom **run_model** function in R. This function integrated differential equation solutions with parameter variations, enabling an examination of the model's response to changes in parameters over time.

Algorithm 3: Pseudo-code for Sensitivity Analysis.

```
Function aggregate_sensitivity_heatmap(parameters, initial_state)
  Initialize an empty data frame for storing aggregate sensitivity results.
  Define perturbation_range as [0.99, 1.01].
  For each parameter in 'parameters':
    Run the baseline model simulation with current parameters.
    For each perturbation factor in 'perturbation_range':
      Apply the perturbation to the current parameter.
      Run the model with the perturbed parameter.
      For each state variable in 'initial_state':
        Calculate the relative change from the baseline.
        Compute the mean sensitivity measure for the state variable.
        Store the mean sensitivity measure in the aggregate data frame.
  Reshape the aggregate sensitivity data for heatmap plotting.
  Plot the heatmap using ggplot2.
End Function

Function temporal_sensitivity_analysis(parameters, initial_state)
  Initialize an empty data frame for storing temporal sensitivity data.
  Define perturbation_range as [0.9, 1.1].
  For each parameter in 'parameters':
    Run the baseline model simulation with current parameters.
    For each perturbation factor in 'perturbation_range':
      Apply the perturbation to the current parameter.
      Run the model with the perturbed parameter.
      For each time point in the model output:
        Calculate the relative change at the time point.
        Store the relative change in the temporal sensitivity data frame.
  Reshape the temporal sensitivity data for plotting.
  For each state variable in 'initial_state':
    Filter and plot a heatmap for the state variable across time points.
End Function

Function parameter_specific_analysis(parameters, initial_state)
  For each state variable in 'initial_state':
    Initialize a list to store plots.
    For each parameter in 'parameters':
      Define a perturbation range from 0.9 to 1.1 with 0.01 increments.
      Initialize a data frame to store model output for plotting.
      For each perturbation factor in the range:
        Apply the perturbation to the current parameter.
        Run the model with the perturbed parameter.
        Store the model output in the plotting data frame.
      Reshape the plotting data for line plotting.
      Create a line plot showing the effect of the current parameter.
      Add the line plot to the plot list.
    Display all line plots using a grid arrangement.
End Function

#Main Execution:
-Call aggregate_sensitivity_heatmap with the list of parameters and initial state values.
-Call temporal_sensitivity_analysis with the list of parameters and initial state values.
-Call parameter_specific_analysis with the list of parameters and initial state values.
```

4.3.5 Validation

Validation of the model was performed using an independent dataset, by De Micheli et al. distinct from the one used for model calibration by McKellar et al. This additional dataset provided a separate set of observations to assess the model's generalizability and robustness. Each data point in the dataset represented an experimentally measured value associated with specific temporal dynamics of cellular processes during muscle regeneration at days 0, 2, 5 and 7. The model was simulated using the same optimized parameter values obtained during the calibration process. Residuals were computed for each data point by taking the difference between the observed values from the validation dataset and the corresponding predicted values generated by the model using the same steps indicated described in **Algorithm 2.**

4.4 Results

The development of the mathematical model was informed by McKellar's dataset, which identifies 22 distinct cell types during the muscle regeneration process. These cell types can be aggregated into 14 meta-clusters (**Table S4**), of which we have focused on population data acquired from 7 clusters (e.g., neutrophils, monocytes, M1 macrophages, M2 macrophages, QSCs, ASCs, and myocytes) (**Table 3.2**). Initial parameter guesses were optimized by minimizing the discrepancies between the observed values and the calibrated scRNA-seq dataset.

The set of ODEs developed for this model (explained in **Section 4.3.2**) are trained on the data collected at 6 time-points (days 0, 1, 2, 3.5, 5, 7), representing the critical period of muscle regeneration post-injury. Initial conditions were established based on a baseline scenario where all cell populations, with the exception of QSCs and Md, start at zero. This

assumption is predicated on the understanding that immediately following muscle injury, the presence of other cell types is minimal or undetectable. The initial number of QSCs is estimated at 2700 cells/mm³, and Md at approximately 30,000 cells/mm³, which reflects typical injury scenarios as documented in the literature[151,157,159].

The results of the model simulations are presented in **Figure 4.2,** where the model's predicted dynamics of cell populations were transposed against empirical single-cell data points collected over a seven-day period following muscle injury. The blue lines in the graphs represent the computational model's predictions for each cell type, including N, M, QSC, ASC, Mc, Nd, Md, Mn, M1 and M2 within the regenerating tissue. The shaded blue areas around the simulated values represents the range of one standard deviation from the simulated value. This area visualizes the typical error range of our model's predictions. The narrow bands suggest that the model parameters are well-calibrated and that the biological processes they represent are well-understood and captured by our model.

The parameter units in **Table 4.2** quantitively measure cell rates, where rates of cell entry are given in per day d^{-1}. An example is shown by parameter $c_{N_{in}}$, representing the maximum rate at which neutrophils enter the damaged tissue per day. Rates involving cell-cell interactions or cell-mediated processes are given in cubic millimeters per cell per day $\left(\frac{\text{mm}^3}{\text{cells}\cdot\text{d}}\right)$, describing a volume related process per cell per day, such as c_{M1Nd}, which is the rate at which lysed neutrophils leave the tissue, indicating the volume of tissue from which cells are removed per day. Parameters that describe the interaction between different cell populations and their effects on tissue volume, are measured by $\left(\frac{\text{cells}}{\text{mm}^3\cdot\text{d}}\right)$, like $c_{M2_{inhib}}$, which is the rate at which M1 infiltration is repressed by the presence of Md and Nd.

Parameter	Description	Value	Units
$c_{N_{in}}$	maximum rate at which neutrophils enter the damaged tissue	1.38E+00	d^{-1}
$c_{N_{out}}$	rate at which neutrophils die off and leave the system	3.10E+00	d^{-1}
c_{M1Nd}	rate at which lysed neutrophils leave the tissue	9.17E-05	$\frac{mm^3}{cells \cdot d}$
$c_{M_{in}}$	rate at which monocytes are recruited from the neutrophils signaling	2.81E+01	d^{-1}
c_{MM1}	rate of M to M1 conversion (proportional to Md and Nd)	5.81E-05	$\frac{mm^3}{cells \cdot d}$
$c_{M1_{out}}$	rate of which M1 leaves skeletal muscle tissue	4.29E-02	d^{-1}
$c_{M2_{inhib}}$	rate at which M1 infiltration is repressed by the presence of Md and Nd	4.58E-12	$\frac{cells}{mm^3 \cdot d}$
$c_{M2_{out}}$	rate at which M2 leaves the tissue	1.18E-01	d^{-1}
c_{QSCN}	rate of QSC to ASC conversion (from # of neutrophil influx) and already present SCs	5.22E-06	$\frac{mm^3}{cells \cdot d}$
c_{ASCM2}	rate of ASC deactivation back to QSC (self-renewal)	8.38E-03	$\frac{mm^3}{cells \cdot d}$
$c_{ASC_{pro}}$	rate of ASC proliferation	4.28E-05	$\frac{mm^3}{cells \cdot d}$
$c_{SC_{diff}}$	rate of ASC to Mc differentiation (influenced by M2 signaling factors)	7.24E-03	$\frac{mm^3}{cells \cdot d}$
$c_{Mc_{fusion}}$	rate at which myocytes fuse together to form new myotubes and myofibers	4.01E+00	d^{-1}
$c_{Md_{out}}$	rate at which damaged myonuclei (Md) get engulfed by M1 and neutrophils	9.81E-04	$\frac{mm^3}{cells \cdot d}$
c_{NMd}	rate at which neutrophils engulfs an Md and converts one N into an Nd	1.18E-05	$\frac{mm^3}{cells \cdot d}$
$c_{M_{out}}$	rate of remaining M's leaving the system	2.31E+01	d^{-1}
c_{M1M2}	rate of conversion from M1 to M2 from Md and Nd (inversely proportional)	5.48E+02	$\frac{cells}{mm^3 \cdot d}$
c_{QSCMd}	rate of QSC activation from Md damage/presence	2.09E-04	$\frac{mm^3}{cells \cdot d}$
$c_{QSC_{self}}$	QSC self-regulation / Heaviside function	9.63E-02	d^{-1}
$c_{Mc_{out}}$	rate of remaining Mc's leaving the system or not fusing into myotubes	3.73E-06	d^{-1}
$c_{Mn_{self}}$	Mn self-regulation / Heaviside function	9.39E-05	d^{-1}
$c_{ASC_{out}}$	rate of remaining ASC's leaving the system or apoptosis	1.55E-04	d^{-1}

Table 4.2. Optimized Parameter Values for ODE Model

4.4.1 Model Predictions

In this Section, I will present the results of the numerical simulations. Following optimization, the model parameters were able to capture the temporal profiles of cell populations. For instance, the population of neutrophils exhibited a peak early in the regeneration process, consistent with the initial inflammatory response observed in muscle healing[65]. The predicted dynamics of satellite cells and ASCs followed a trajectory that confirmed the phases of activation, proliferation, and subsequent self-renewal or differentiation, aligned with empirical observations and biological expectations. The model's predictive accuracy was assessed post-optimization by evaluating the proportion of data points that fell within the residual band (**Figure S7**). This approach was specifically designed to validate the optimization process and ensure a robust fit of the model to the empirical data. The SDs of these residuals established a band around the model's predictions, within which the majority of the observed data points were expected to lie if the model provided an accurate fit. By quantifying the percentage of data points within this band, we could directly assess the effectiveness of the optimization in aligning the model's simulations with the observed trends. Specifically, for the variables QSC, ASC, and Mc 83% of the data points were within the residual band, highlighting a strong agreement with the empirical data. When evaluating the overall model performance across all variables, the mean proportion of data points within the residual band was approximately 74%. Tighter intervals for M1, N, and QSC signified a robust prediction capability for these variables. Conversely, the broader intervals for ASCs and M2 macrophages indicated a higher degree of uncertainty in their predictions, potentially due to model limitations or inherent biological variability. For instance, the larger band for the ASC trajectory may reflect the complexity of the activation process, which is influenced by various signals and environmental

factors that might not be fully captured in the model. Such factors may include the influence of fibrogenic and vascular cells. To further assess our model fit, I calculated the Total Sum of Squares (TSS) for each variable, which represents the overall variance in the observed data. The TSS for a variable is calculated as the sum of the squared differences between each observed value and the mean of all observed values for that variable. The Total Sum of Squared Errors (SSE) was calculated from the model's optimization result, quantifying the discrepancy between the observed data and the model's predictions. It is obtained as the sum of squared differences between the observed values and the corresponding values predicted by the model, for each variable. The overall predictive performance of the model was then assessed by calculating the percentage of the total variance in the observed data that remains unexplained by the model. This is given by the ratio of the Total SSE to the Total TSS. The calculated percentage of unexplained variance was found to be approximately 1.38×10^{-5} %. A lower percentage of unexplained variance indicates a better fit of the model to the observed data, signifying that the model can explain a larger proportion of the variance observed in the empirical measurements.

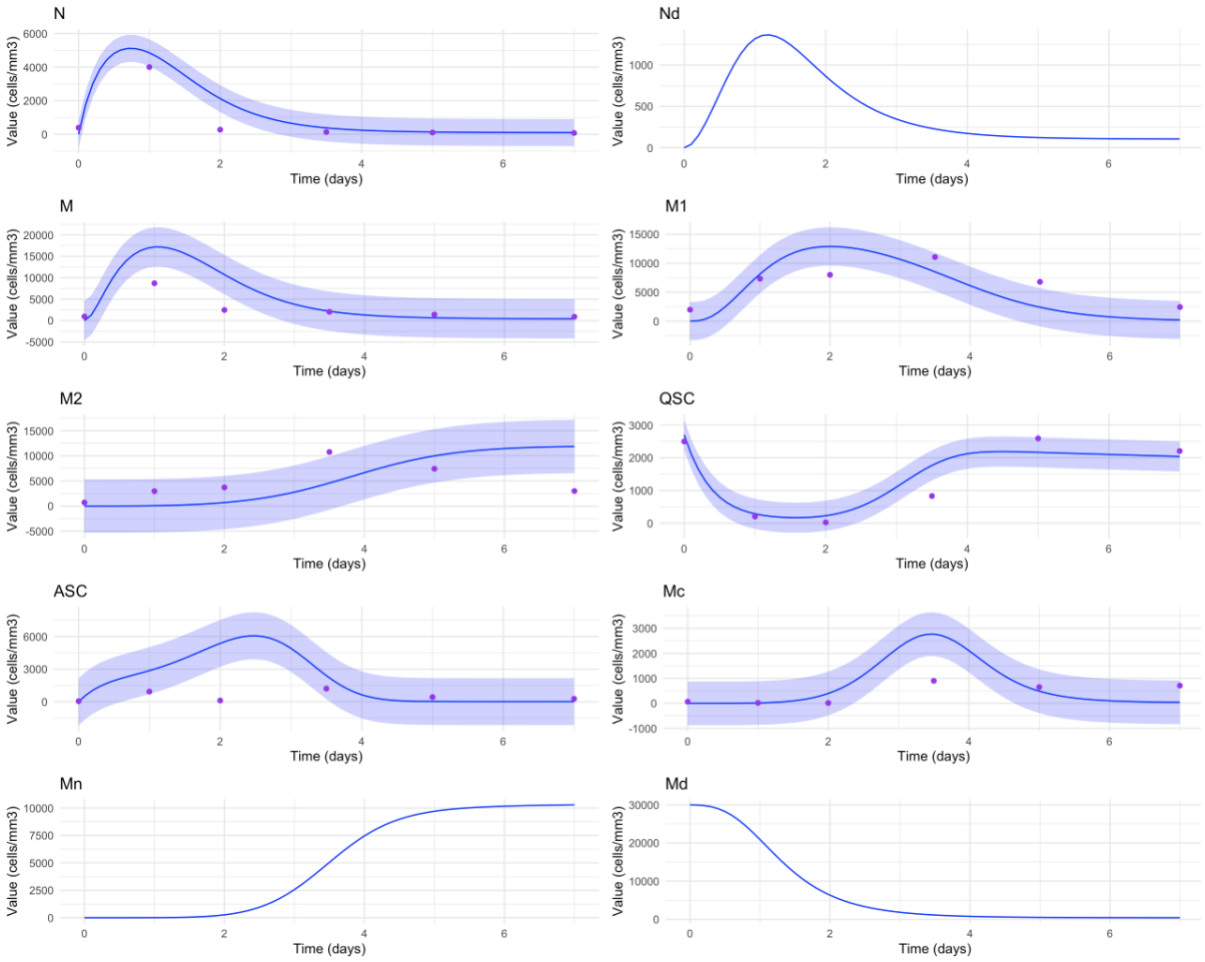


Figure 4.2. Model Simulation and Data Fitting of Cell Populations During Skeletal Muscle Regeneration. The plots illustrate the dynamics of cell populations within skeletal muscle over a time course of seven days post-injury. Blue lines represent the simulated trajectories of the cell populations based on the optimized parameters of the ODE model. Purple dots indicate empirical single-cell data points for corresponding cell types (N, M, M1, M2, QSC, ASC, and Mc) at discrete time points. Shaded blue areas around the simulated trajectories denote residual bands, reflecting the variability and precision of the model predictions.

4.4.2 Sensitivity Analysis of Model Parameters

The model's robustness and the influence of its parameters on the system's dynamics were assessed through sensitivity analysis. This process involved individually perturbing each parameter within a range of ±10% from their baseline values and observing the resultant effects on model outputs **(Figure 4.3).** This heatmap reveals that certain parameters exert a more pronounced influence on specific variables, as evidenced by the intensity of the coloration.

Darker shades indicate a higher average sensitivity, reflecting a greater propensity for the state variable to fluctuate in response to parameter perturbations.

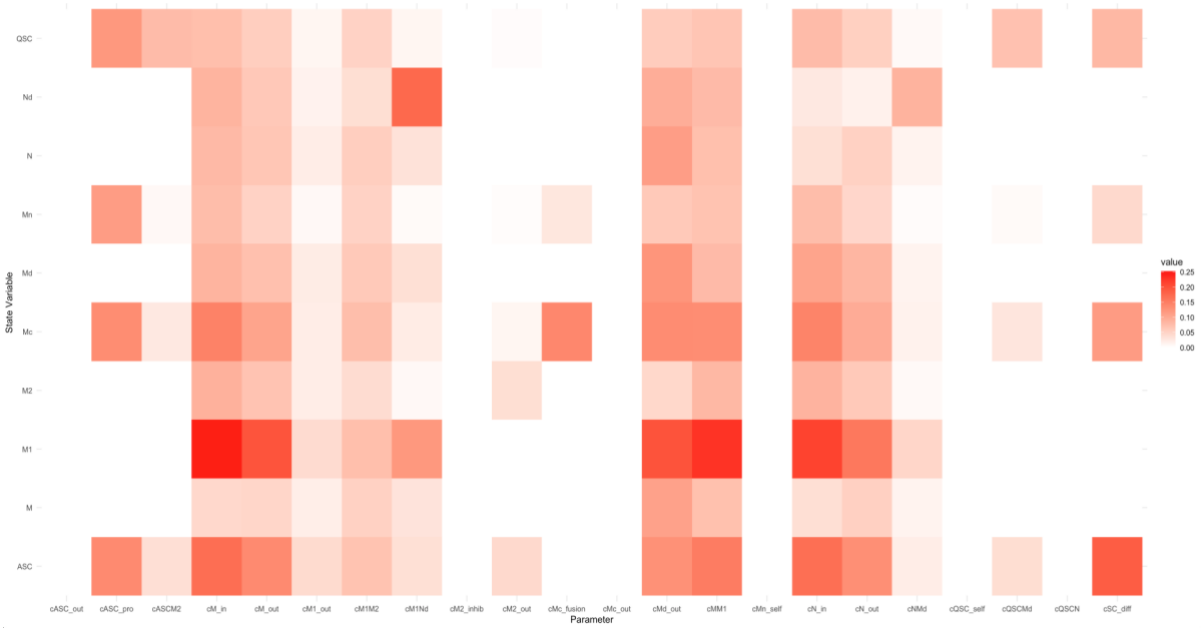


Figure 4.3. Heatmap of Average Sensitivity of Variables to ±10% Parameter Perturbations. This heatmap displays the average sensitivity of each variable (rows) to perturbations ranging from -10% to +10% in each model parameter (columns) in 1% intervals. Sensitivity is measured as the mean of absolute relative changes in variable values, reflecting the average impact of parameter variations on the system's behavior.

Notably, the parameter $c_{M_{in}}$ exhibits a significant impact on the state variable M1, indicating that the rate at which monocytes are recruited into the damaged tissue is a critical factor influencing the abundance of classically activated macrophages. Similarly, the parameters c_{MM1} and $c_{N_{in}}$, associated with the conversion rate of monocytes to M1 macrophages and the entry rate of neutrophils, respectively, also show high sensitivity for M1. In the context of myonuclei regeneration, $c_{ASC_{pro}}$, which represents the rate of activated satellite cell proliferation, is identified as the most sensitive parameter for Mn.

The temporal sensitivity analysis conducted across the model's parameters and state variables revealed dynamic interactions that evolve over time, as depicted in the series of

heatmaps presented in **Figure 4.4.** Each heatmap corresponds to a variable within the system, showing the degree of sensitivity to parameter perturbations over the simulation period. Notably, the variables Mc and Mn exhibit pronounced sensitivity to perturbations in almost all parameter perturbations. Interestingly, the most pronounced changes are evident between days 2 and 3.5, indicating a highly dynamic phase within the system, as characterized by an acute response to parameter fluctuations. The perturbed plots, **Figures S7 to S16,** for each variable can be found in the appendix section of this dissertation.

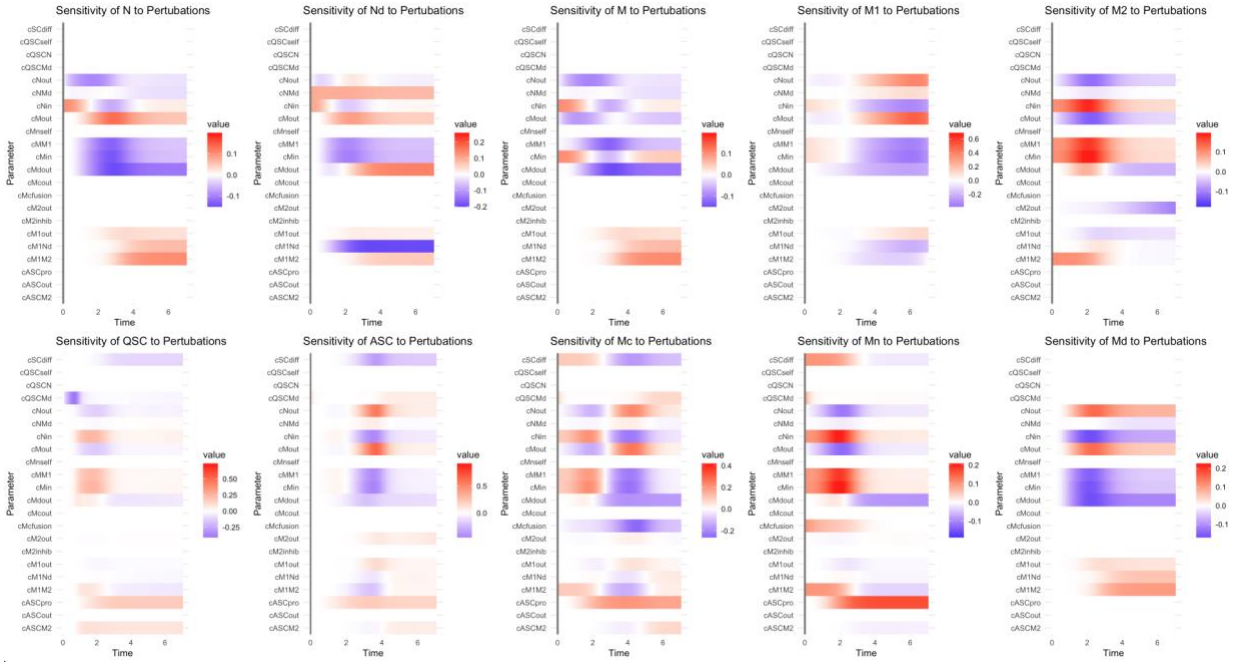


Figure 4.4. Heatmaps of Temporal Sensitivity of Variables to ±10% Parameter Perturbations. The heatmaps illustrate the time-dependent sensitivity of model variables to a range of parameter perturbations. Each row represents a parameter, each column a time point, and the color intensity the mean relative change in variable values due to parameter variation.

Upon analyzing the impact of ±10% parameter perturbations on state variables, a series of boxplots was generated to visualize the relative changes **Figure 4.5**. The overall results indicate a variable degree of model sensitivity to parameter changes. Some parameters showed a consistent influence across multiple variables, while others exhibited more variable-specific

effects such as $c_{Md_{out}}$ which displayed a pronounced influence on all variables, whereas $c_{M2_{inhib}}$, displayed no notable influence across all variables.

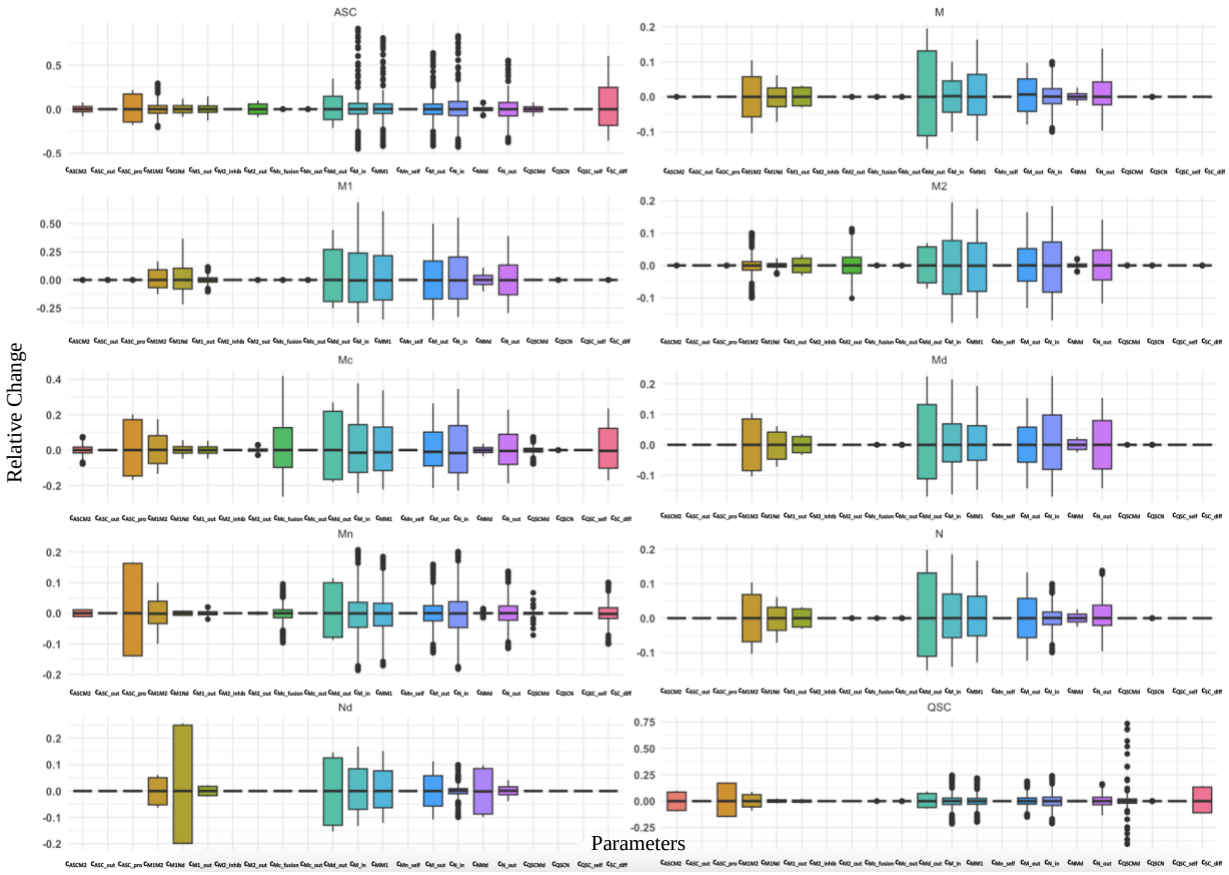


Figure 4.5. Boxplot Analysis of Relative Changes in State Variables Due to Parameter Perturbations. This series of boxplots demonstrates the distribution of relative changes in model state variables upon ±10% perturbation of each parameter. The y-axis represents the relative change, while the x-axis categorizes the different parameters perturbed. The box represents the interquartile range (25th to 75th percentile) of the relative change, with the median depicted by the line within the box. Whiskers extend to the furthest data points that lie within 1.5 times the interquartile range, and points outside this range are considered outliers.

4.5 Model Validation

Model validation was conducted using an independent dataset by De Micheli et al. previously analyzed and scaled in **Chapter 3 (Table S7)** to evaluate the predictive capability of the model beyond the conditions of the initial training data[176]. This assessment involved overlaying the

validation dataset onto the model's predicted trajectories derived from the training data and calculating the residuals using the same optimized parameters **(Figure 4.6)**. At days 0, 2, 5, and 7, the data points from Micheli's dataset generally align with the proportions observed in McKellar's dataset. Notably, the primary fluctuations within the model predictions occur between days 0 and 2. This observation raises the possibility that De Micheli's dataset might not fully encapsulate the early-stage dynamics of the system. Both datasets recorded lower ASC counts compared to the simulation predictions. Whereas, for Mc's, De Micheli's data showed an increased count of Mc at day 7, coupled with a later influx of M1 and M2 macrophages compared to McKellar's, suggesting a potential delay in the regenerative response. However, the model predictions seemed to capture M2 macrophages from De Micheli's time points relatively well. De Micheli's data aligns well with McKellar's dataset at day 2, but it is important to note that the model predicted earlier peaks for neutrophils and monocytes, indicating that De Micheli's dataset may not fully capture the early influx of those cells at day 2 as observed in the model. QSCs seemed to be captured well by our model as it remained in close proximity to the model simulation.

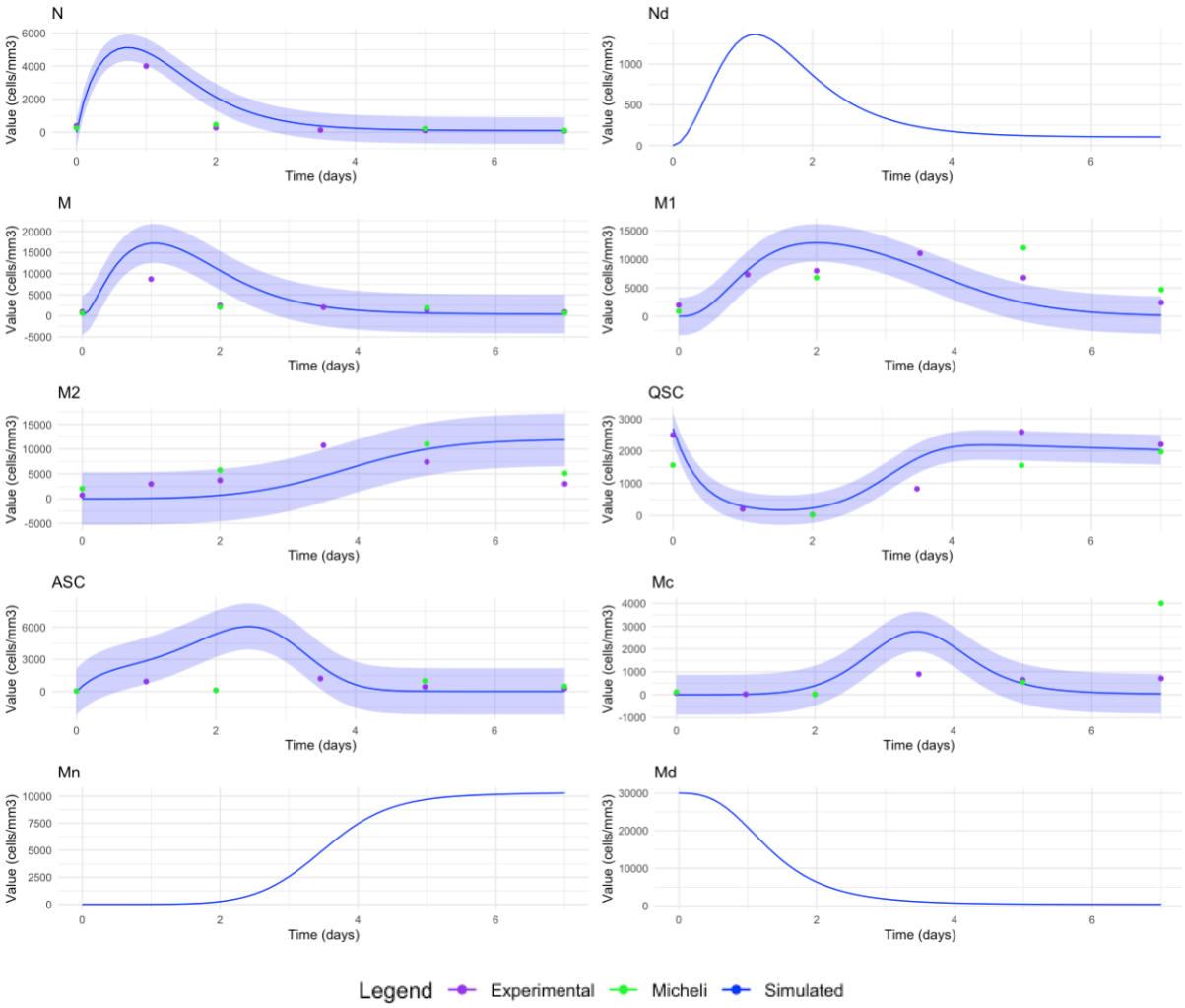


Figure 4.6. Comparative Analysis of Cell Populations Over Time. The graphs display the dynamic changes in cell populations measured in cells/mm³ over a 7-day period. Each subplot represents a different cell type. The blue line indicates the cell count trajectory as predicted by the model using the training dataset by McKellar, while the green dots denote the observed cell counts from the Micheli dataset. The graphs illustrate the model's ability to capture the trends of the experimental data, with deviations highlighting areas for further investigation and model refinement.

The residual band analysis, derived from the training dataset, highlights specific trends and discrepancies in model adherence across different cell types. Analyzing the De Micheli dataset against this residual band, we observe that for cell types like neutrophils, monocytes, M2 macrophages, ASC, and Mc's, approximately half of the data points align within the residual band. This indicates a reasonable model fit for these variables. However, QSCs, although visually exhibits the expected behavior as the model's simulation, only align with the band

around 33% of the time, and M1 macrophages show even less conformity, with only about 17% of data points falling within the residual band. This lower adherence rate for QSCs and M1 suggests that these aspects of the model may benefit from refinement. The variability in QSCs could be reflective of biological variability in their quiescent state, while the discrepancies in M1 macrophages might point to a need for a more nuanced model of their dynamics. The complexities in M1 behavior, possibly influenced by factors not fully incorporated in the current model, like interactions with fibro-adipogenic progenitors (FAPs) and mast cells, could be contributing to this variance. The sensitivity of M1 macrophages to a wide range of parameters, as shown **Figure 4.3**, validates their complex role in the system which may account for its variability.

4.6 Discussion

In this chapter, we introduce a mathematical model that consolidates signaling interactions, transformations, proliferation, and cell exit rates across 22 parameters and 10 cell types. It expands on previous models by detailing the dynamics of satellite cells in three distinct states and the interactions among various immune cell populations, including neutrophils, monocytes, and M1 and M2 macrophages. The selection of the 10 specific cell types and states, particularly, the differentiation of MuSCs into quiescent, activated, and committed states, was informed by their critical and well-established functions in the process of muscle regeneration. This decision allowed us to construct a model that is both detailed and focused, capturing essential aspects of muscle healing and renewal. Other cell types, such as FAPs, endothelial, and vascular cells, were not incorporated at this stage to streamline the model's complexity and ensure its analytical tractability. This approach was chosen to ensure a manageable scope

for the model, allowing for a targeted exploration of the regeneration process while laying a foundation for future expansions to include these and other cell types as our understanding and computational methods evolve.

A distinctive feature of our model lies in its utilization of a single dataset to portray the population dynamics within the muscle niche, a methodological choice that avoids the consequences of data pooling and thereby more accurately captures their dynamics. The model was trained with the empirical data revealing some correspondence for numerous data points within the residual band, affirming that the single cell data's accuracy in constraining the model. The model predictions suggest that the influx of immune cells, particularly types N and M, may reach peak levels that exceed those recorded in experimental data. This discrepancy suggests that complex interactions within the system are not fully captured by experimental methodologies, may be revealed by our ODE system. Such insights from the model could lead to the development of new hypotheses and guide the design of targeted experiments to explore these dynamics further, especially at critical early-stage time points. Our model indicates that these peaks may occur earlier than day 2, pointing to a substantial proportion of immune cell activity that remains unaccounted for in the model. However, the occurrence of data points that did not closely fit the model may also highlight the variability of biological systems, potential for measurement errors, and possible limitations in our modeling approach, including the exclusion of certain cellular interactions or external factors.

Significantly, our model distinguishes between QSCs and their activated progeny, a feature not commonly addressed in existing models. For example, while the model by Stephenson and Kojouharov differentiate between activated and committed states, they do not consider QSCs[8]. Moreover, Jarrah et al.'s model, though it accounts for macrophage states,

does not differentiate between pro- and anti-inflammatory macrophages, nor does it consider MuSCs, opting instead to represent regenerated muscle fibers as a percentage[11].

In the preliminary model formulation, I introduced a unique interaction term involving M2 macrophages influencing the rate of conversion of ASCs commitment to Mc. This addition was driven by our hypothesis that this interaction might have a significant influence on the commitment of ASCs to differentiation and their potential to self-renew back into QSCs. MuSC variables (QSC, ASC, and Mc) exhibited high sensitivity to parameters involved with M2 regulatory interactions such as the effect of c_{ASCM2} on QSC and ASCs. It is important to note that this aspect of the model had not been previously accounted for in existing models. While there were limited references in the literature supporting this concept, this interaction was incorporated because the model's functionality and capacity for self-renewal were contingent on its inclusion. An additional parameter that demonstrated high sensitivity to perturbations in regenerating myonuclei (Mn) is $c_{ASC_{pro}}$, which may emerge as a potential target for therapeutic intervention. Modulating ASC proliferation may fine-tune the regenerative process, suggesting that treatments aimed at amplifying M1 macrophage function or quantity could amplify ASC proliferation, potentially expediting muscle repair[177]. Analysis of **Figure 4.4** indicates that this influence is most pronounced starting from day 2. This is further validated by the high sensitivity of all MuSCs by pro-inflammatory immune cell parameters, particularly, $c_{M_{in}}$and c_{MM1}. Additionally, neutrophil's sensitivity to $c_{Md_{out}}$ as depicted in **Figures 4.3** and **4.5**, corroborates that neutrophil infiltration is directly linked to the extent of damage, with this interaction commencing rapidly within the first day post-injury (**Figure 4.4**). $c_{Md_{out}}$notably affects nearly all variables, indicating that the rest of the system's dynamics are substantially driven by the degree of damage as well. Nd's sensitivity is

predominantly towards negative perturbations in c_{M1Nd}, with this effect commencing between 0.5-to-1 DPI (**Figures 4.3 and 4.5**). M's overall sensitivity is relatively lower, but $c_{Md_{out}}$ appears to have the most significant effect on it, likely due to its correlation with N, which is in turn proportional to Md. M1 exhibits high sensitivity to perturbations in several parameters (**Figure 4.3**), reflecting its involvement in multiple system interactions. However, there is a minimal system impact observed for M1 around day 2 (**Figure 4.4**). On the other hand, M2 shows lower sensitivity to perturbations overall. Similarly, QSC's sensitivity to parameter changes is not pronounced, but it is most responsive to $c_{ASC_{pro}}$(**Figure 4.3**). **Figure 4.4** reveals a notable sensitivity of QSC to negative perturbations in c_{QSCMd}, at day 1, likely due to heightened cell activation during this phase. The proposed model enables predictions of cellular changes in response to parameter adjustments. For instance, modulating a single parameter can forecast shifts in the regeneration timeline. However, it is imperative to acknowledge that our model is a simplified representation, serving as a foundational platform for future expansion and refinement. Among the limitations is FAPs and vascular cells were not explicitly modeled despite their substantial influence on regeneration. Additionally, although Nd, Md, and Mn were modeled based on interactions and assumptions derived from empirical evidence, scRNA-seq technology is not effective in profiling myonuclei and apoptotic neutrophils. This limitation necessitated reliance on inferential modeling techniques to approximate the behaviors of these cell populations, which, while informative, highlights the need for a multifaceted approach in future modeling efforts that can integrate a broader spectrum of cell types and states.

Chapter 5

Conclusions and Future Work

In this thesis, I have presented a population dynamic model that quantitively represents the dynamics of skeletal muscle regeneration following injury. This work has expanded upon the mathematical assumptions informed by existing literature, adhering closely to a realistic portrayal of MuSC dynamics. There have been remarkably few population dynamic models that focus specifically on MuSCs and the intricacies of their niche. Prior models have typically not accounted for the distinct states of such as the quiescent state and the self-renewal process of MuSCs. To address this gap, this thesis introduces ten nonlinear ODEs, with 22 parameters that articulate the interactions across ten immune and muscle cell types, including the three primary states of MuSCs. **Chapter 2** outlines the mechanisms of muscle regeneration and the role of MuSCs, along with the mathematical modeling approaches employed to understand these dynamics. A key distinction of this work is its approach to data utilization. Our approach differs from conventional studies that typically pool data from multiple experiments for model fittings. In **Chapter 3**, we employed a re-analysis of scRNA-seq data, featuring extensive time points, to annotate cell types and quantify their proportions relative to tissue volume. This

method aimed to reduce biases from aggregating distinct datasets, providing a more cohesive basis for informing our mathematical model in **Chapter 4**, enhancing accuracy and avoiding cell type misrepresentation, as subtle populations can influence global dynamics. This is particularly evident within our system equations as modest perturbations of certain parameters has significantly influenced the dynamics of multiple cell types. The choice to train the model on a comprehensive scRNA-seq dataset, conducted from one experiment, represents a significant methodological advancement over existing MuSC population dynamic models, which often rely on multiple datasets and assumptions, due to the scarcity of comprehensive studies on cell types within the MuSC niche. Using only one calibrated dataset, consisting of 21 samples and 6 post-injury time points, yielded a well-aligned representation of a separate dataset and the biological insights from the literature. The population model presented aligns closely with experimental observations, offering an additional layer of validation.

While experiments such as cell cultures or FACS provide a substantial amount of information, they alone cannot encapsulate the entirety of the dynamics governing the niche and MuSCs. Additionally, a particularly valuable application of this model is its ability to evaluate the effects of minor perturbations at the individual cell level on the overall population dynamics. As research in scRNA-seq expands to explore dystrophic muscles, future work can leverage this model's framework to compare and contrast the cell populations in diseased and healthy muscles.

Notably, it was observed that the model accurately represented these biological processes only when including the interaction between M2 macrophages and ASCs, a factor governed by parameter c_{ASCM2}. It is well established that M2 macrophages promote differentiation of MuSCs by inhibiting over-proliferation of ASC through anti-inflammatory

regulatory processes. However, there has been limited evidence suggesting these macrophages may also play a role in suppressing ASC proliferation, potentially triggering their return to a self-renewal state[153,178]. Therefore, future research should explore the role of M2 macrophages in satellite cell self-renewal through targeted studies, such as gene knockouts.

Through biological interpretation, the model demonstrates a regenerative response marked by MuSC population peaks, reflecting the primary stages of muscle repair and remodeling. Despite its demonstrated utility, the model's current iteration does not encompass the spatial heterogeneity, or the full array of cell states that scRNA-seq could elucidate. Several challenges remain to improve the robustness of population dynamic modelling of MuSCs. Future studies would benefit from integrating diverse sequencing methods, such as single nuclei and spatial transcriptomics, extending the model's predictive capabilities across various injury models, thereby enhancing its scope and accuracy. Additionally, the model can be refined to include other influential cell types in the niche, including FAPs, or vascular cells. This limitation centers around the scope of cell types incorporated into the models. Our current model, while comprehensive in its current form, could be further refined by including additional influential cell types present in the MuSC niche, such as FAPs and vascular cells. The inclusion of these cells would not only enhance the model's biological accuracy but also offer a more holistic view of the complex interactions within the muscle regeneration process. This expansion would enable a deeper understanding of the nuanced interplay between different cell types during muscle repair and regeneration.

Appendix

Accession no.	# of samples	Tissue	Protocol	Cell isolation Procedure	Species
GSE126834	n=2	TA and EDL	10X Chromium v2	Whole muscle & FACS	Mouse
GSE147883	n=2	Hindlimb	SMART-Seq2	FACS (Stromal cells)	Mouse
M:GSE138707	H: n=4	Quadriceps & diaphragm	10X Chromium	FACS	Mouse
H:GSE130646	M: n=2	Quadriceps & diaphragm	10X Chromium	FACS	Human
GSE143476	n=16	Hindlimb	10X Chromium v2	FACS (MuSC)	Mouse
GSE130977	n=4	Vastus lateralis, mouse quadriceps, and mouse diaphragm	10X Chromium	FACS	Human & Mouse
GSE142480	n=4	Quadriceps and diaphragm	10X Chromium	FACS	Mouse
GSE159500	n=2	TA	10X Chromium v2 or v3	FACS (specific cell types)	Mouse
GSE143704	n=10	10 different whole muscles	10X Chromium v3	Dissociated whole muscle	Human
GSE168872	n=4	Plantaris	10X Chromium v3	FACS	Mouse
GSE143435	n=9	Hindlimb	10X Chromium v2	FACS (MuSC)	Mouse
GSE158987	n=6	Hindlimb: soleus muscle	10X Chromium	FACS	Mouse
GSE163446	n=15	Sagittal cross sections of mice hindlimb	10X Chromium v3	Dissociated whole muscle	Mouse
GSE144270	n=5	TA	10X Chromium v2	Dissociated whole muscle	Mouse
GSE113111	n=2	Hindlimb	10X Chromium v2	FACS (Macrophages)	Mouse
GSE142480	n=4	Quadriceps and diaphragm	10X Chromium v3	FACS	Mouse
GSE134540	n=3	TA & Soleus	10X Chromium v2	FACS (MuSC)	Mouse
GSE152467	n=6	TA	10X Chromium v2	Dissociated whole muscle	Mouse
GSE109774	n=2	TA	10X Chromium v2	Dissociated whole muscle	Mouse
GSE143636	n=2	Hindlimb, TA, GSN	10X Chromium v3	FACS	Mouse
GSE122209	n=2	TA, GSN	C1 High-Throughput IFC	Dissociated whole muscle	Mouse
GSE110037	n=5	Hindlimb	10X Chromium v2	FACS	Mouse
GSE138027	n=3	Quadriceps	10X Chromium v2	FACS	Mouse
GSE129057	n=2	Hindlimb	10X Chromium v2	FACS (Endo. & MuScs)	Mouse
GSE121469	n=3	TA	10X Chromium	FACS	Mouse
GSE122873	n=2	TA	10X Chromium v2	Dissociated whole muscle & FACS	Human
GSE133449	n=2	Meniscus	10X Chromium	FACS	Human

Table S1: Compilation of scRNA-seq studies reviewed for muscle regeneration post-injury, detailing sample counts, tissues analyzed, protocols employed, cell isolation procedures, and species.

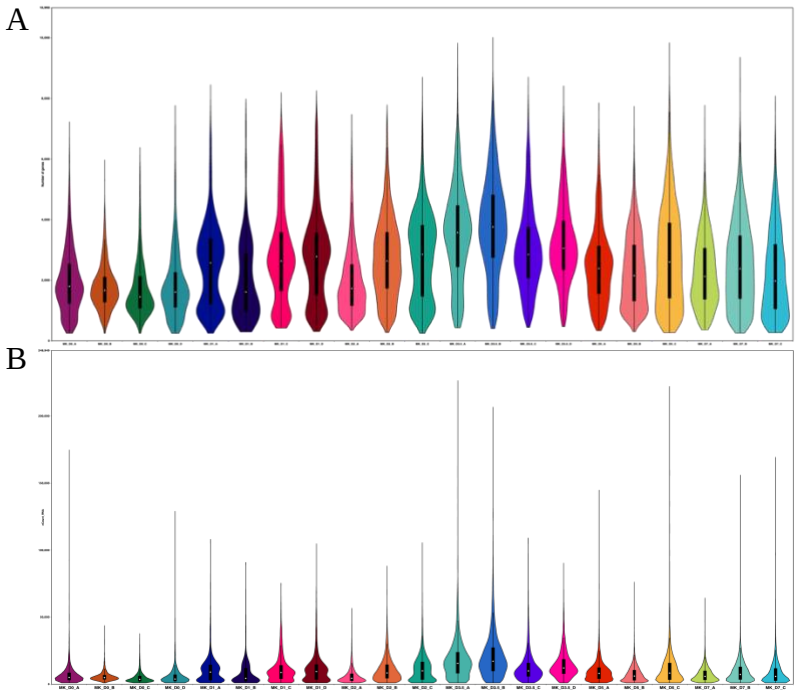


Figure S1. Quality control violin plots for McKellar's dataset. (A) Number of genes (n_Feature) and **(B)** UMIs (n_Count) per cell per batch after QC filtering. On average cells have more UMIs and genes detected for injured samples.

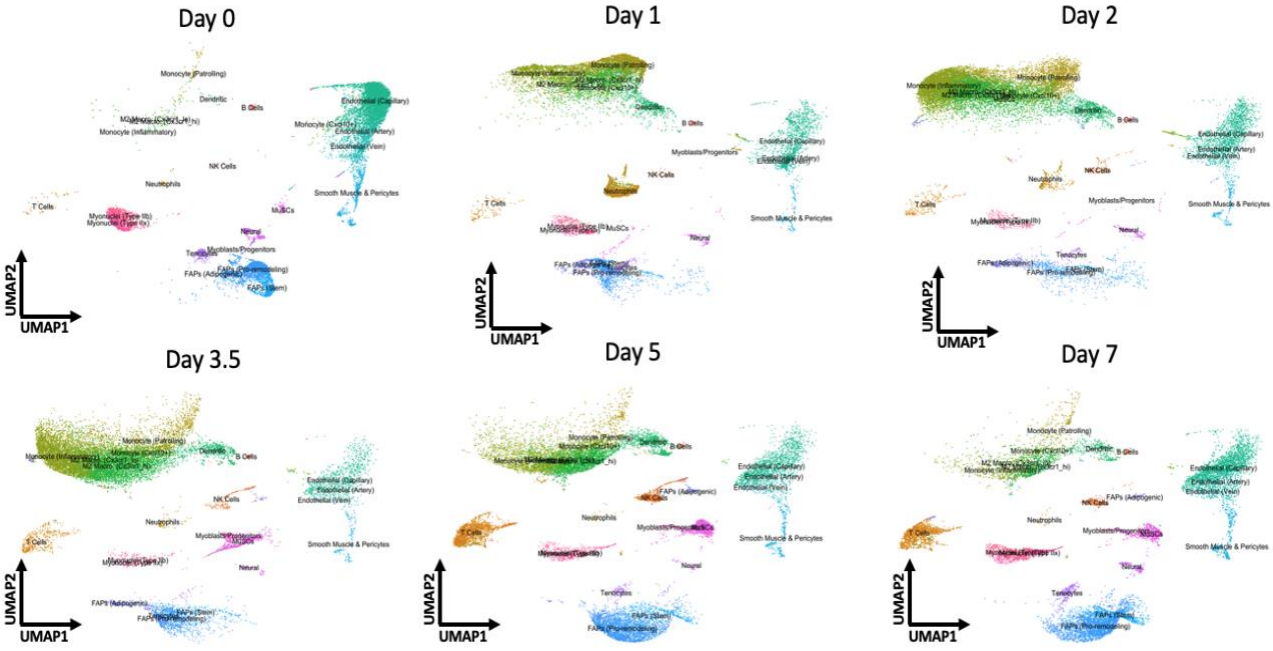


Figure S2. UMAP visualizations showing the temporal evolution of cell type clusters in McKellar's dataset across 0-7 DPI, with clusters annotated according to the reference study.

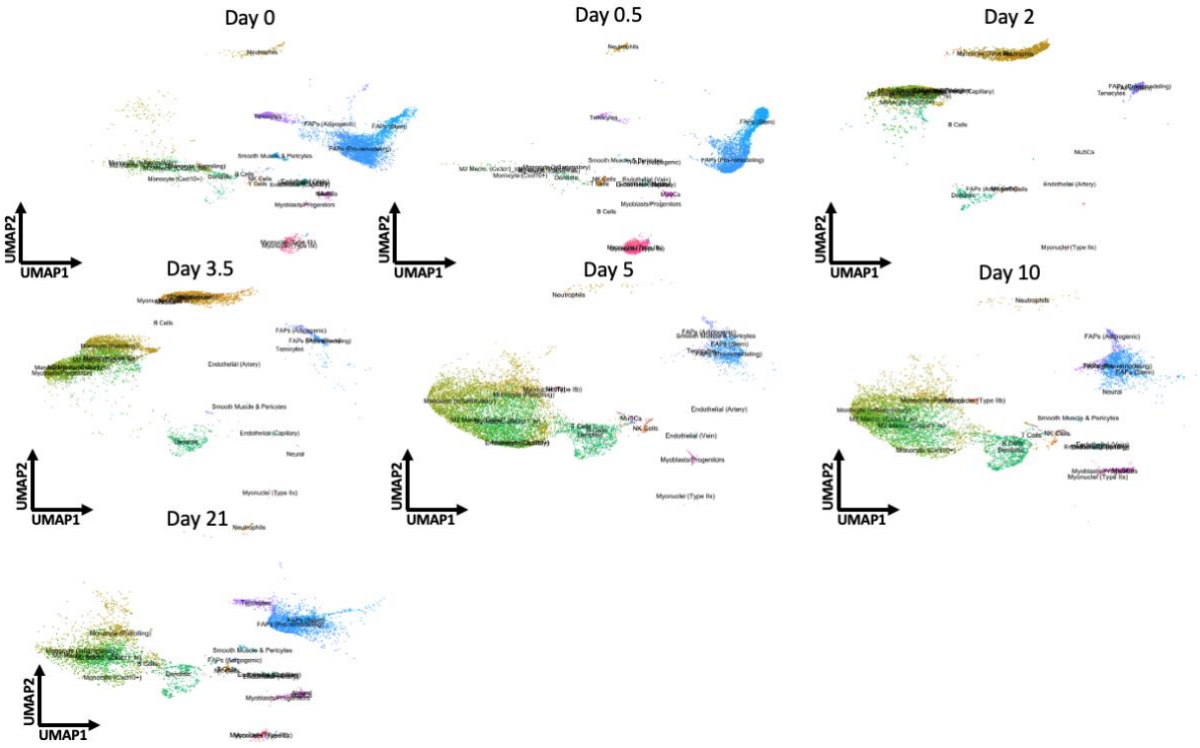


Figure S3. UMAP visualizations showing the temporal evolution of cell type clusters in Oprescu's dataset across 0-21 DPI, with clusters annotated according to the reference study.

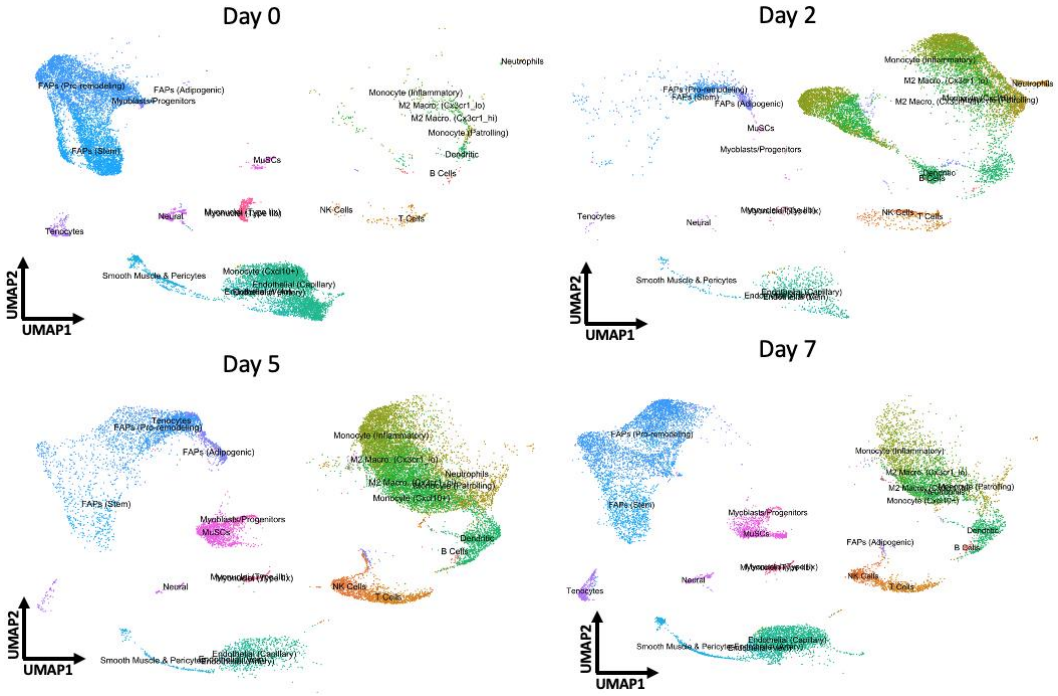


Figure S4. UMAP visualizations showing the temporal evolution of cell type clusters in De Micheli's dataset across 0-7 DPI, with clusters annotated according to the reference study.

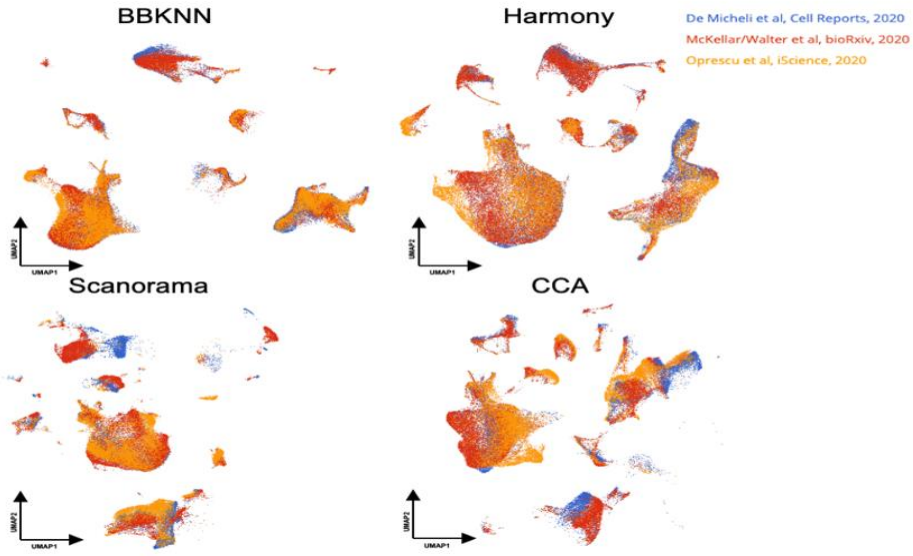


Figure S5. Comparative Analysis of Batch Correction Techniques in Integrated Single-Cell RNA-Seq Data. This figure presents UMAP visualizations of integrated single-cell RNA-seq data from the De Micheli et al., McKellar/Walter et al., and Oprescu et al. studies, following batch correction with four different algorithms: BBKNN, Harmony, Scanorama, and Canonical Correlation Analysis (CCA). Harmony batch correction yielded the most distinct and well-defined clusters, indicating its effectiveness in correcting batch effects while preserving the biological heterogeneity of cell populations. In contrast, BBKNN, Scanorama, and CCA showed varying degrees of integration, with some clusters appearing less distinct or more dispersed, suggesting potential over- or under-correction.

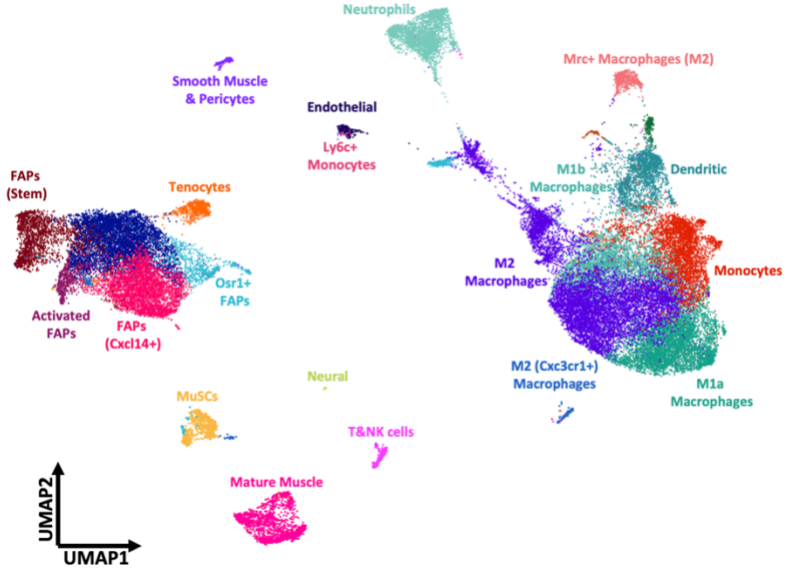


Figure S6. UMAP visualization of the dataset by Oprescu delineating 20 distinct clusters. Each cluster is color-coded to correspond with identified cell types.

Cell Type	Marker Genes	Reference
Endothelial	PECAM1 (CD31), CDH5, VEGFR2 (KDR)	Imoukhuede & Popel, 2012; Kalucka et al., 2020
Smooth	ACTA2, MYL9, MYH11	Muhl et al., 2022
Pericytes	MCAM, PDGFRB, NG2 (CSPG4)	Rippe et al., 2021
Tenocytes	TNMD, SCX, COL1A1	Gumucio et al., 2020
Mature	MYH1, MYH2, MYH4, ACTA1	Oprescu et al., 2020b; Williams et al., 2022; McKellar et al., 2021
FAPs (Adipogenic)	PPARG, CD36	Molina et al., 2021
FAPs (Stem)	PDGFRα, SCA1	McKellar et al., 2021
(Pro-remodeling)	TGFβ, MMP14	McKellar et al., 2021
FAPs (Cxc14+)	CXCL14	Oprescu et al., 2020b
Neutrophils	S100A8, S100A9, CXCR2, CCL3	McKellar et al., 2021; Reichel et al., 2009
M0/M Monocytes	CTSA, CTSL, CTSB, CCR2, CD14, ITGAM, LYZ2, S100A4, CCL2	De Micheli et al., 2020; Ingersoll et al., 2010; McKellar et al., 2021; Oprescu et al., 2020b
Ly6C+ Monocytes	CD177, LY61C, LY6C2C, IRF5	Corbin et al., 2020; Stroncek et al., 2004; Yang et al., 2021
Classical M1 Macrophages	CLL2, CCL6, CCL9, TNF, IL10, NOS2	Krasniewski et al., 2022a; McKellar et al., 2021; Corbin et al., 2020; Krasniewski et al., 2022b
M1a Macrophages	SPP1, IL17RA, CXCL3, CXCL16	Mao et al., 2022; Patsalos et al., 2022; Zhang et al., 2009
M1b Macrophages	CCR2, CXCL16, C1QC, IL6RA, CSF1R, AIF1	De Micheli et al., 2020; Krasniewski et al., 2022a
CD14+ Macrophages	CD14, SLFN4, CXCL2, FOSL2, ATF4, MDM2	De Micheli et al., 2020; Krasniewski et al., 2022a; McKellar et al., 2021; Oprescu et al., 2020b
CX3CR1+ Macrophages	CX3CR1, LY86, LPXN, DHRS3, DNMT3a	Wang et al., 2016
M2 Macrophages	ARG1, CD163, MRC1, CCL8, FOLR2, APOE, TREM2	De Micheli et al., 2020; Krasniewski et al., 2022a; McKellar et al., 2021; Oprescu et al., 2020b
T cells	CD3E, CD4, CCL5, CD8A	McKellar et al., 2021
NK cells	NKG7, KLRA7, NCAM1 (CD56)	McKellar et al., 2021
B cells	CD19, MS4A1 (CD20)	McKellar et al., 2021
Dendritic cells	CD1C, HLA-DR	McKellar et al., 2021

Table S2. Gene markers used for cell-type annotation.

Cell Type	D0 (%)	D1 (%)	D2 (%)	D3.5 (%)	D5 (%)	D7 (%)

B cells	0.25%	0.28%	0.32%	0.23%	0.45%	0.40%
Dendritic	0.21%	4.49%	5.22%	3.79%	5.08%	2.34%
Endothelial	50.15%	10.42%	9.47%	5.16%	9.89%	20.65%
FAPs	20.82%	9.86%	8.23%	12.13%	22.14%	26.32%
Monocytes	0.57%	23.67%	11.93%	5.47%	2.98%	1.84%
M1 Macrophages	1.18%	19.90%	38.70%	29.65%	14.41%	5.04%
M2 Macrophages	0.42%	8.07%	17.93%	28.87%	15.76%	6.22%
Mature Muscle	13.42%	3.94%	1.22%	1.70%	3.49%	11.62%
MuSCs	1.57%	3.17%	0.76%	7.88%	7.80%	6.63%
Neural	2.77%	0.61%	0.93%	0.34%	0.55%	1.47%
Neutrophils	0.23%	10.87%	1.32%	0.36%	0.22%	0.16%
Smooth Muscle & Pericytes	5.76%	1.23%	1.39%	0.94%	1.84%	3.42%
T&NK cells	0.40%	1.33%	1.69%	3.22%	14.75%	11.67%
Tenocytes	2.23%	2.16%	0.90%	0.27%	0.65%	2.23%
Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

Table S3. Proportions (%) of Total Cells Across Time Points (0,1,2,3.5,5,7) in McKellar's Dataset

Cell Type	D0	D0.5	D2	D3.5	D5	D10	D21

FAPs	46.00%	64.76%	5.49%	5.89%	15.50%	20.17%	39.11%
B cells	0.02%	0.04%	0.06%	0.29%	0.38%	0.75%	0.32%
Dendritic	1.23%	0.70%	1.86%	5.36%	8.96%	8.12%	4.88%
Endothelial	29.04%	0.63%	7.86%	1.62%	4.33%	0.84%	0.84%
Monocytes	0.52%	0.06%	10.51%	9.04%	4.44%	3.24%	2.46%
M1 Macrophages	1.36%	0.39%	40.66%	35.79%	25.06%	28.87%	16.75%
M2 Macrophages	2.59%	1.48%	24.22%	24.65%	21.78%	31.85%	24.51%
Mature Muscle	7.57%	22.84%	0.66%	1.45%	1.94%	0.23%	1.82%
MuSCs	1.47%	2.88%	0.75%	6.41%	6.15%	3.55%	2.92%
Neural	1.85%	0.15%	0.41%	0.36%	0.29%	0.04%	0.14%
Neutrophils	0.46%	1.90%	4.33%	5.60%	0.49%	0.61%	0.42%
Smooth Muscle & Pericytes	4.48%	0.24%	0.92%	0.75%	1.10%	0.19%	0.81%
T&NK cells	0.78%	1.89%	1.84%	2.45%	9.06%	1.10%	1.84%
Tenocytes	2.61%	2.05%	0.45%	0.34%	0.51%	0.45%	3.17%
Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	Total

Table S4. Proportions (%) of Total Cells Across Time Points (0,0.5,2,3.5,5,10,21) in Oprescu's Dataset

Cell Type	**D0**	**D2**	**D5**	**D7**

FAPs	52.57%	9.27%	18.74%	35.60%
B cells	0.17%	0.97%	0.54%	0.74%
Dendritic	0.86%	6.43%	4.90%	3.21%
Endothelial	31.65%	5.07%	4.77%	17.24%
Monocytes	0.39%	9.98%	4.03%	1.40%
M1 Macrophages	0.54%	32.82%	25.52%	9.76%
M2 Macrophages	1.21%	28.02%	23.43%	10.61%
Mature Muscle	3.53%	0.25%	0.80%	1.48%
MuSCs	1.04%	0.72%	6.59%	5.93%
Neural	1.79%	0.16%	0.24%	1.41%
Neutrophils	0.16%	2.25%	0.47%	0.21%
Smooth Muscle & Pericytes	3.43%	0.58%	0.71%	2.89%
T&NK cells	1.05%	3.36%	8.97%	6.70%
Tenocytes	1.62%	0.13%	0.31%	2.81%
Total	100.00%	100.00%	100.00%	100.00%

Table S5. Proportions (%) of Total Cells Across Time Points (0,2,5,7) in De Micheli's Dataset.

Cell type	D0	D0.5	D1	D2	D3.5	D5	D7	D10	D21
FAPs	45.82%	64.76%	10.29%	8.38%	8.83%	17.96%	31.35%	20.17%	39.11%
B cells	0.15%	0.04%	0.40%	0.66%	0.48%	0.67%	0.75%	0.75%	0.32%
Dendritic	1.01%	0.70%	3.97%	5.23%	3.62%	5.08%	2.59%	8.12%	4.88%
Endothelial	31.07%	0.63%	10.19%	6.96%	3.70%	5.72%	18.54%	0.84%	0.84%
Monocytes	0.55%	0.06%	20.63%	10.64%	9.71%	3.83%	1.51%	3.24%	2.46%
M1 Macrophages	1.18%	0.39%	18.14%	33.47%	31.12%	23.57%	7.21%	28.87%	16.75%
M2 Macrophages	2.25%	1.48%	13.87%	24.06%	25.98%	24.12%	9.35%	31.85%	24.51%
Mature Muscle	6.74%	22.84%	4.05%	0.80%	1.25%	1.69%	6.18%	0.23%	1.82%
MuSCs	1.33%	2.88%	2.35%	0.80%	5.87%	5.69%	6.23%	3.55%	2.92%
Neural	1.78%	0.15%	0.60%	0.56%	0.24%	0.31%	1.45%	0.04%	0.14%
Neutrophils	0.45%	1.90%	10.98%	4.75%	5.92%	0.47%	0.22%	0.61%	0.42%
Smooth Muscle & Pericytes	4.24%	0.24%	1.25%	0.96%	0.68%	0.98%	3.12%	0.19%	0.81%
T&NK cells	0.80%	1.89%	1.22%	2.21%	2.41%	9.51%	8.96%	1.10%	1.84%
Tenocytes	2.64%	2.05%	2.04%	0.52%	0.20%	0.40%	2.54%	0.45%	3.17%
Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

Table S6. Proportions (%) of Total Cells Across Time Points in Integrated Dataset of Three Single-Cell Studies.

Cell Type	D0	D2	D5	D7
QSC	1564	20	1558	1981
ASC	50	111	996	482
Mc	121	17	547	399
Neutrophils	262	465	222	104
Monocytes	646	2063	1897	674
M1	898	6786	12013	4708
M2	2017	5792	11030	5122

Table S7. Scaled cell counts of De Micheli's dataset.

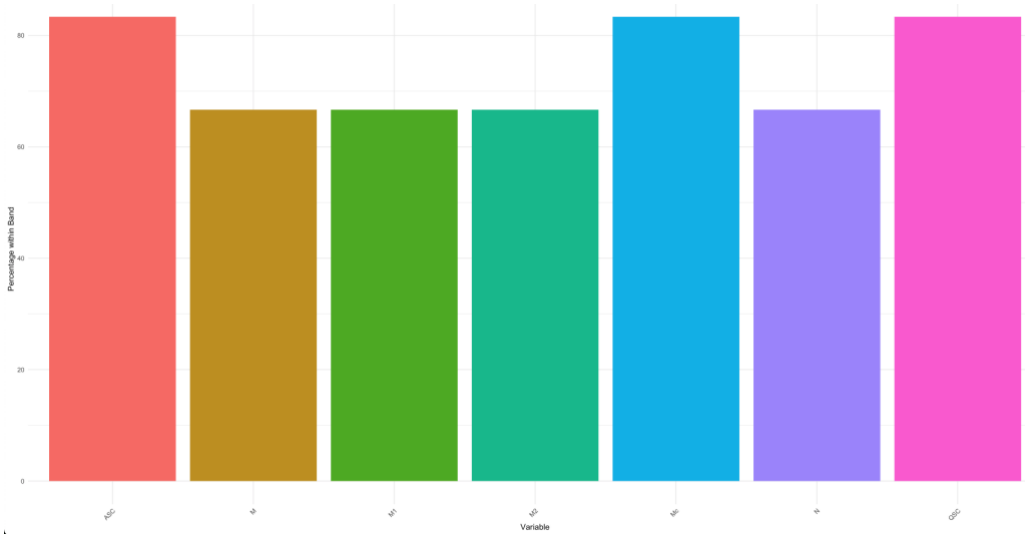


Figure S7. Percentage of residuals within band by variable in training dataset.

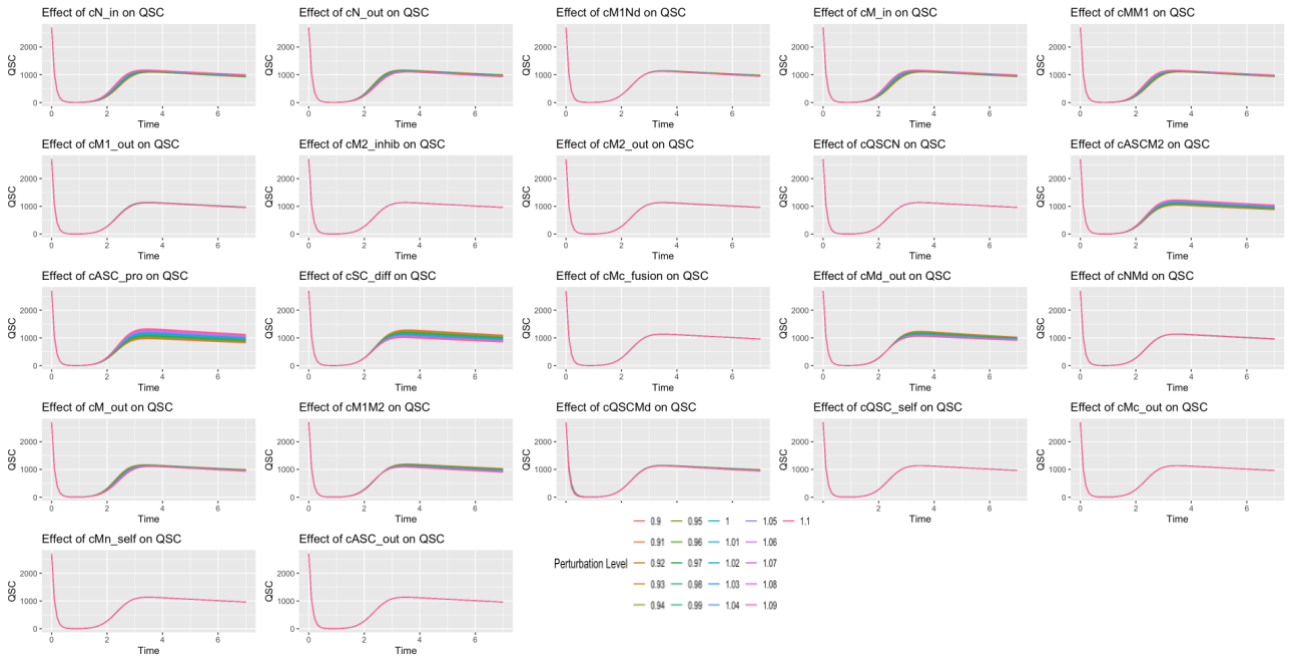


Figure S8. Perturbation Analysis of Model Parameters on within ±10% of its baseline value on the dynamics of QSCs Over 0-7 days.

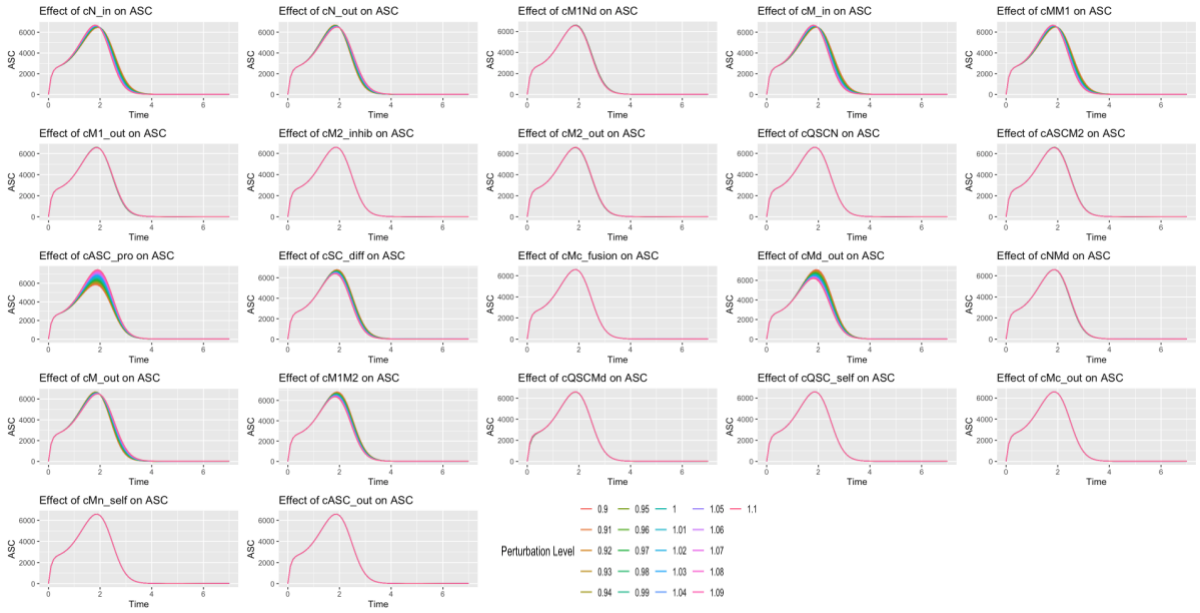


Figure S9. Perturbation Analysis of Model Parameters on within ±10% of its baseline value on the dynamics of ASCs over 0-7 days.

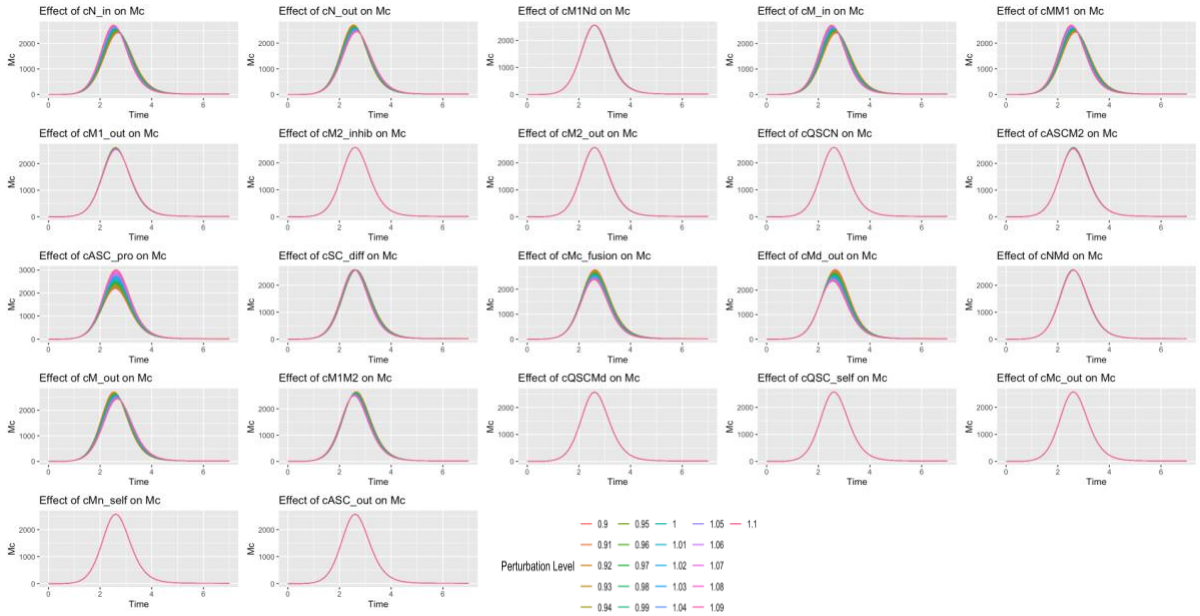


Figure S10. Perturbation Analysis of Model Parameters on within ±10% of its baseline value on the dynamics of Mc over 0-7 days.

Appendix

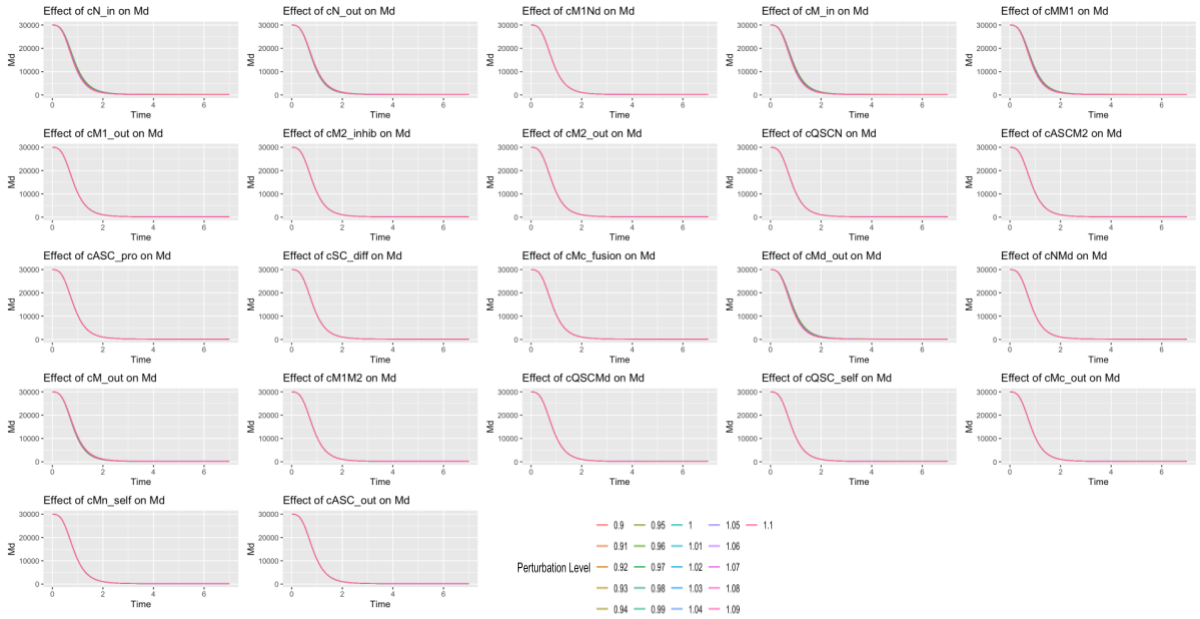


Figure S11. Perturbation Analysis of Model Parameters on within ±10% of its baseline value on the dynamics of Md over 0-7 days.

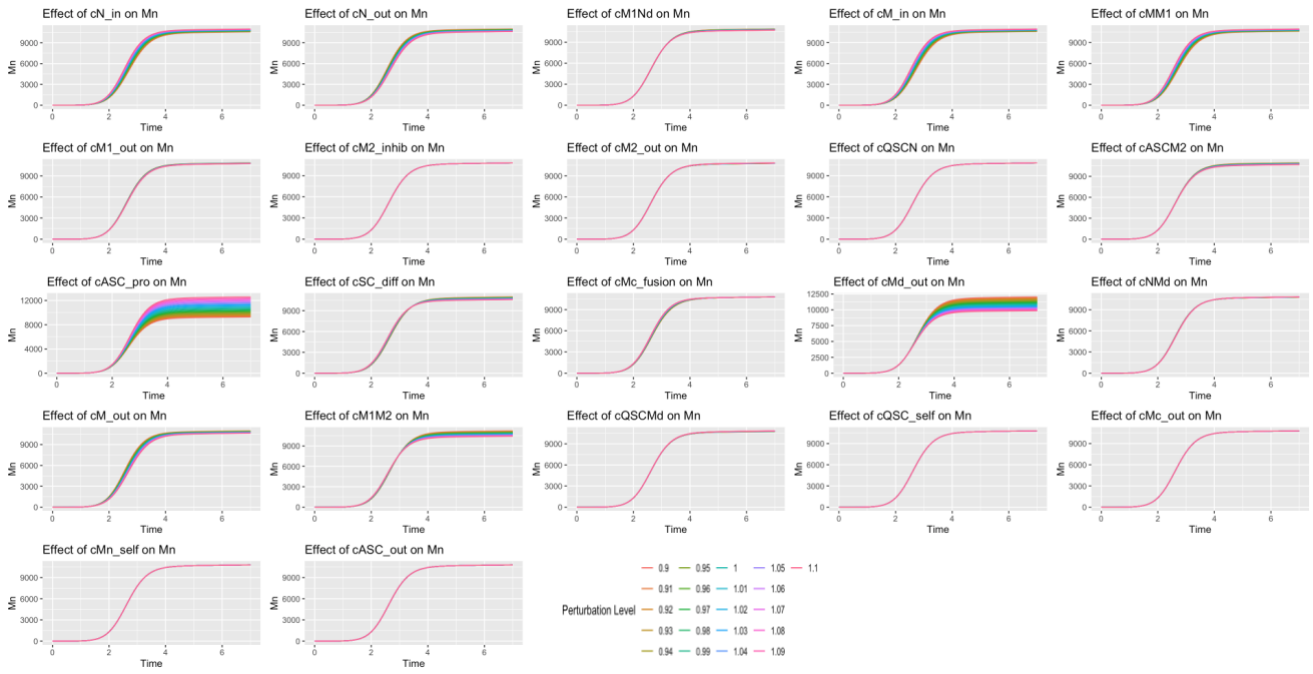


Figure S12. Perturbation Analysis of Model Parameters on within ±10% of its baseline value on the dynamics of Mn over 0-7 days.

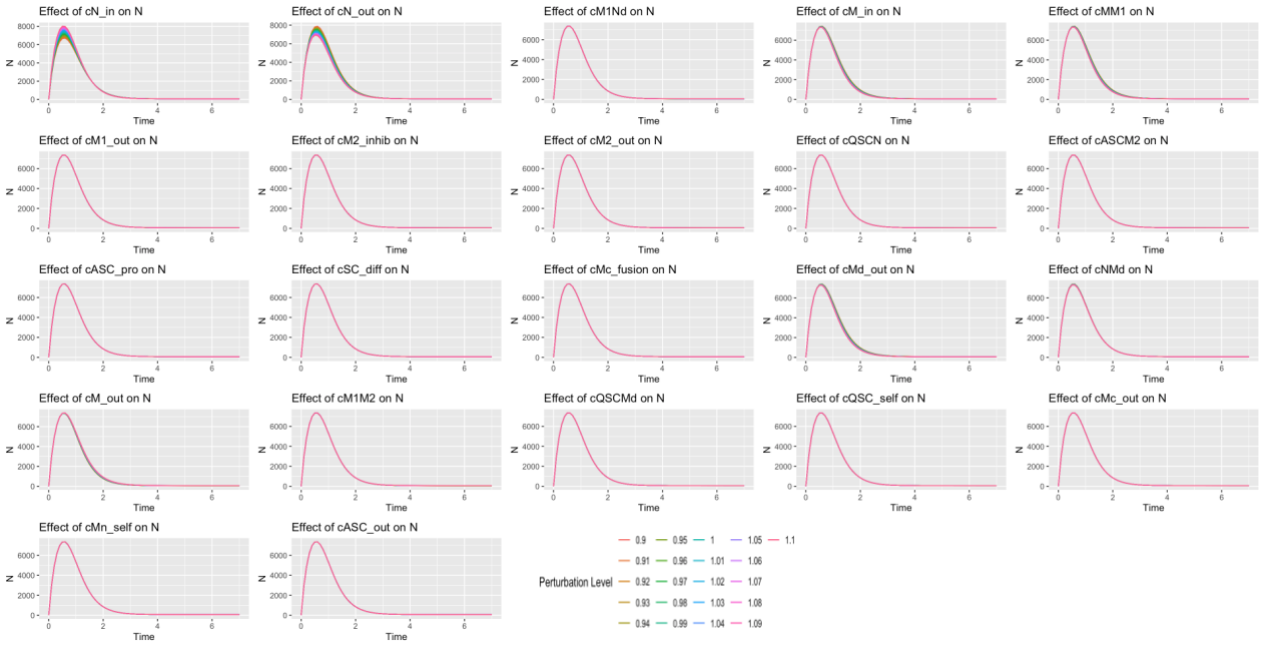


Figure S13. Perturbation Analysis of Model Parameters on within ±10% of its baseline value on the dynamics of QSCs Over 0-7 days.

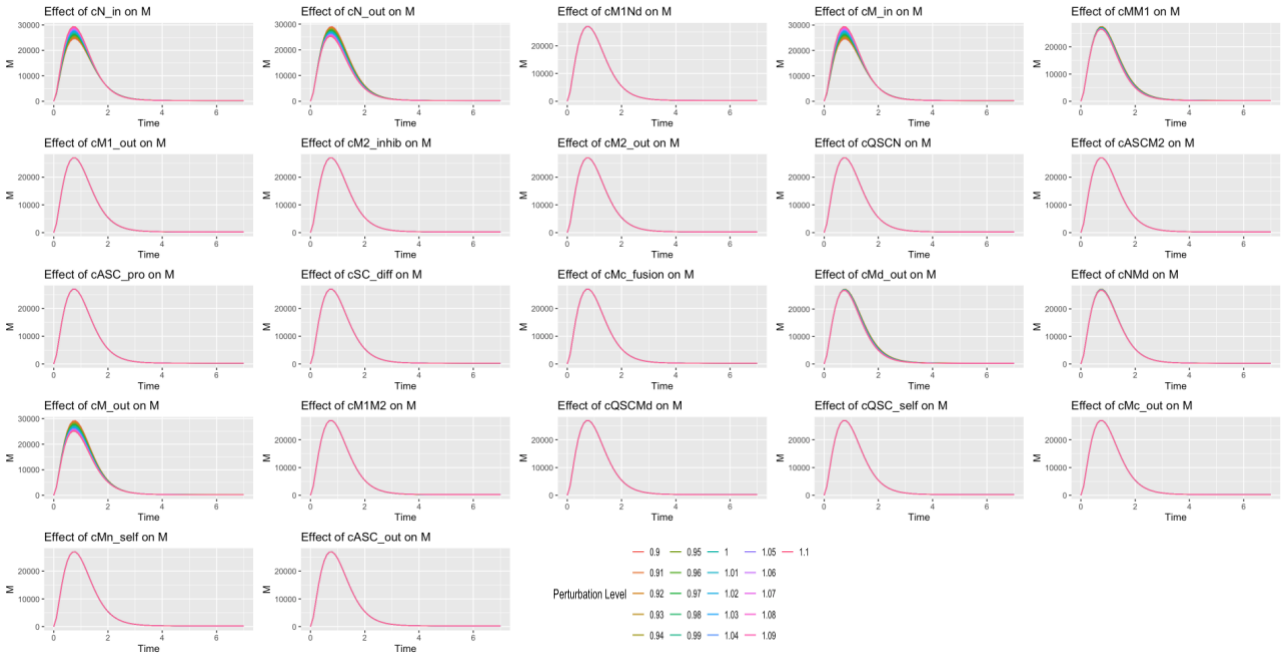


Figure S14. Perturbation Analysis of Model Parameters on within ±10% of its baseline value on the dynamics of Mover 0-7 days.

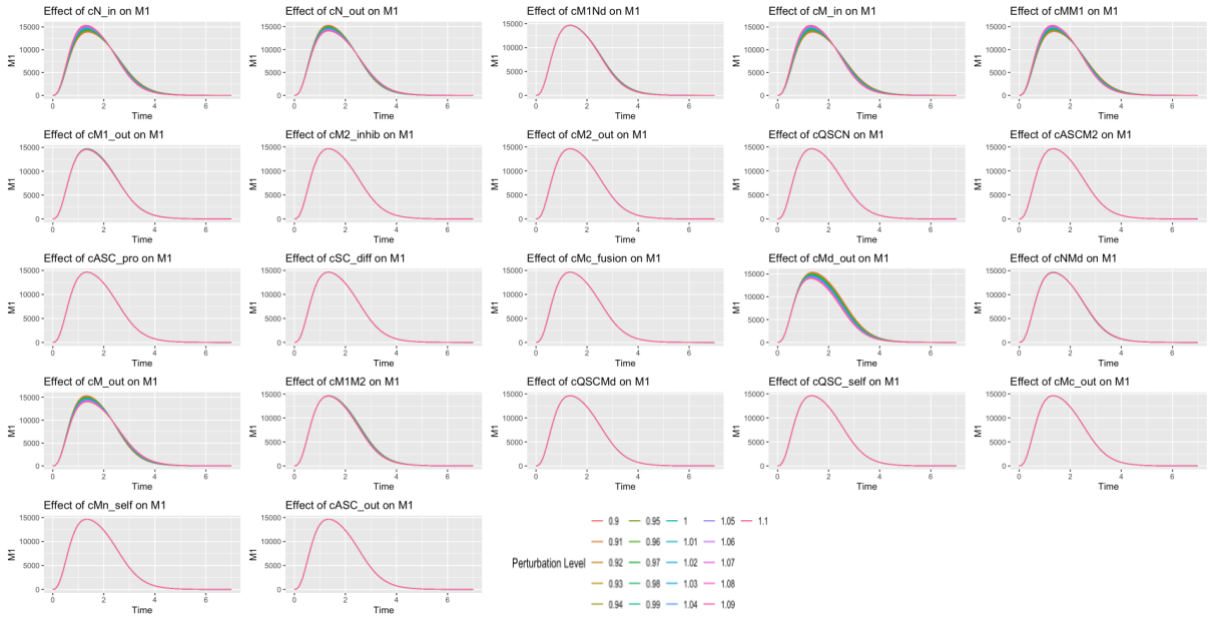


Figure S15. Perturbation Analysis of Model Parameters on within ±10% of its baseline value on the dynamics of M1 over 0-7 days.

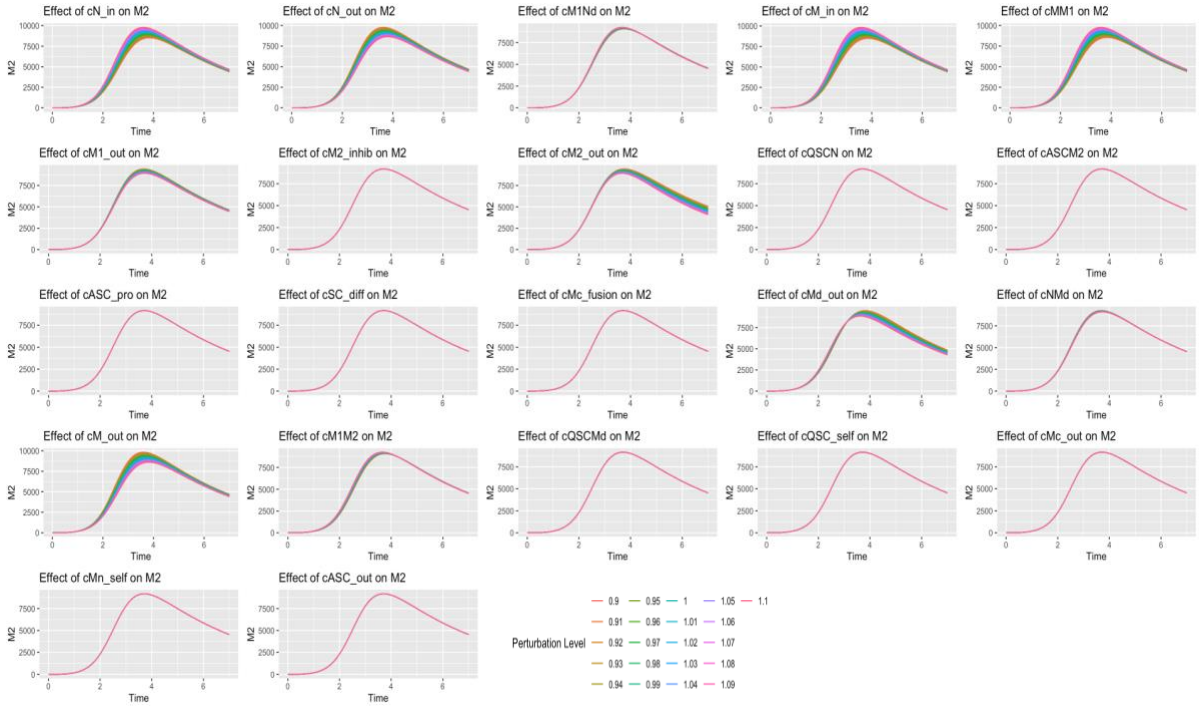


Figure S16. Perturbation Analysis of Model Parameters on within ±10% of its baseline value on the dynamics of M2 over 0-7 days.

References

1. Wood, C. C. The computational stance in biology. *Philos Trans R Soc Lond B Biol Sci* **374**, 20180380 (2019).
2. Yin, H., Price, F. & Rudnicki, M. A. Satellite Cells and the Muscle Stem Cell Niche. *Physiol Rev* **93**, 23–67 (2013).
3. Schüler, S. C. *et al.* Extracellular matrix: Brick and mortar in the skeletal muscle stem cell niche. *Front Cell Dev Biol* **10**, (2022).
4. Williams, K., Yokomori, K. & Mortazavi, A. Heterogeneous Skeletal Muscle Cell and Nucleus Populations Identified by Single-Cell and Single-Nucleus Resolution Transcriptome Assays. *Front Genet* **13**, (2022).
5. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).
6. Wen, L. & Tang, F. Recent advances in single-cell sequencing technologies. *Precis Clin Med* **5**, (2022).
7. Saber, J., Lin, A. Y. T. & Rudnicki, M. A. Single-cell analyses uncover granularity of muscle stem cells. *F1000Res* **9**, 31 (2020).
8. Stephenson, E. R. & Kojouharov, H. V. A mathematical model of skeletal muscle regeneration. *Math Methods Appl Sci* **41**, 8589–8602 (2018).
9. Dell'Acqua, G. & Castiglione, F. A Mathematical Model of Duchenne Muscular Dystrophy. in *Applied and Industrial Mathematics in Italy III* 311–322 (WORLD SCIENTIFIC, 2009). doi:10.1142/9789814280303_0028.
10. Kojouharov, H. V., Chen-Charpentier, B. M., Solis, F. J., Biguetti, C. & Brotto, M. A simple model of immune and muscle cell crosstalk during muscle regeneration. *Math Biosci* **333**, 108543 (2021).
11. Jarrah, A. S., Castiglione, F., Evans, N. P., Grange, R. W. & Laubenbacher, R. A Mathematical Model of Skeletal Muscle Disease and Immune Response in the *mdx* Mouse. *Biomed Res Int* **2014**, 1–11 (2014).
12. Mukund, K. & Subramaniam, S. Skeletal muscle: A review of molecular structure and function, in health and disease. *WIREs Systems Biology and Medicine* **12**, (2020).
13. Campbell, N. J. & Maani, C. V. *Histology, Muscle*. (2023).
14. Frontera, W. R. & Ochala, J. Skeletal Muscle: A Brief Review of Structure and Function. *Calcif Tissue Int* **96**, 183–195 (2015).
15. Betts, J. G. *et al. Anatomy & Physiology* . vol. 10.2 (OpenStax, 2013).
16. McCuller, C., Jessu, R. & Callahan, A. L. *Physiology, Skeletal Muscle*. (2023).
17. Mauro, A. SATELLITE CELL OF SKELETAL MUSCLE FIBERS. *J Cell Biol* **9**, 493–495 (1961).
18. Seale, P. *et al.* Pax7 Is Required for the Specification of Myogenic Satellite Cells. *Cell* **102**, 777–786 (2000).
19. Stuelsatz, P. *et al.* Extraocular muscle satellite cells are high performance myo-engines retaining efficient regenerative capacity in dystrophin deficiency. *Dev Biol* **397**, 31–44 (2015).
20. Shenkman, B. S., Turtikova, O. V, Nemirovskaya, T. L. & Grigoriev, A. I. Skeletal muscle activity and the fate of myonuclei. *Acta Naturae* **2**, 59–66 (2010).

21. Talbert, E. E. & Guttridge, D. C. Impaired regeneration: A role for the muscle microenvironment in cancer cachexia. *Semin Cell Dev Biol* **54**, 82–91 (2016).
22. Dumont, N. A. & Rudnicki, M. A. Targeting muscle stem cell intrinsic defects to treat Duchenne muscular dystrophy. *NPJ Regen Med* **1**, 16006 (2016).
23. Cappellari, O., Mantuano, P. & De Luca, A. "The Social Network" and Muscular Dystrophies: The Lesson Learnt about the Niche Environment as a Target for Therapeutic Strategies. *Cells* **9**, 1659 (2020).
24. Zullo, L. *et al.* The Diversity of Muscles and Their Regenerative Potential across Animals. *Cells* **9**, 1925 (2020).
25. Scharner, J. & Zammit, P. S. The muscle satellite cell at 50: the formative years. *Skelet Muscle* **1**, 28 (2011).
26. Johnson, A. L., Kamal, M. & Parise, G. The Role of Supporting Cell Populations in Satellite Cell Mediated Muscle Repair. *Cells* **12**, 1968 (2023).
27. Soleimani, V. D. *et al.* Transcriptional Dominance of Pax7 in Adult Myogenesis Is Due to High-Affinity Recognition of Homeodomain Motifs. *Dev Cell* **22**, 1208–1220 (2012).
28. von Maltzahn, J., Chang, N. C., Bentzinger, C. F. & Rudnicki, M. A. Wnt signaling in myogenesis. *Trends Cell Biol* **22**, 602–609 (2012).
29. Bentzinger, C. F., Wang, Y. X. & Rudnicki, M. A. Building Muscle: Molecular Regulation of Myogenesis. *Cold Spring Harb Perspect Biol* **4**, a008342–a008342 (2012).
30. Ustanina, S., Carvajal, J., Rigby, P. & Braun, T. The Myogenic Factor Myf5 Supports Efficient Skeletal Muscle Regeneration by Enabling Transient Myoblast Amplification. *Stem Cells* **25**, 2006–2016 (2007).
31. Conerly, M. L., Yao, Z., Zhong, J. W., Groudine, M. & Tapscott, S. J. Distinct Activities of Myf5 and MyoD Indicate Separate Roles in Skeletal Muscle Lineage Specification and Differentiation. *Dev Cell* **36**, 375–385 (2016).
32. Lazure, F. *et al.* Myf6/MRF4 is a myogenic niche regulator required for the maintenance of the muscle stem cell pool. *EMBO Rep* **21**, (2020).
33. Relaix, F. *et al.* Perspectives on skeletal muscle stem cells. *Nat Commun* **12**, 692 (2021).
34. Fukada, S. *et al.* Molecular Signature of Quiescent Satellite Cells in Adult Skeletal Muscle. *Stem Cells* **25**, 2448–2459 (2007).
35. Motohashi, N., Asakura, Y. & Asakura, A. Isolation, Culture, and Transplantation of Muscle Satellite Cells. *Journal of Visualized Experiments* (2014) doi:10.3791/50846.
36. Gioftsidi, S., Relaix, F. & Mourikis, P. The Notch signaling network in muscle stem cells during development, homeostasis, and disease. *Skelet Muscle* **12**, 9 (2022).
37. Fujimaki, S. *et al.* Notch1 and Notch2 Coordinately Regulate Stem Cell Function in the Quiescent and Activated States of Muscle Satellite Cells. *Stem Cells* **36**, 278–285 (2018).
38. Vargas-Franco, D. *et al.* The Notch signaling pathway in skeletal muscle health and disease. *Muscle Nerve* **66**, 530–544 (2022).
39. Gioftsidi, S., Relaix, F. & Mourikis, P. The Notch signaling network in muscle stem cells during development, homeostasis, and disease. *Skelet Muscle* **12**, 9 (2022).
40. Zhou, S., Han, L. & Wu, Z. A Long Journey before Cycling: Regulation of Quiescence Exit in Adult Muscle Satellite Cells. *Int J Mol Sci* **23**, 1748 (2022).

41. Segalés, J., Perdiguero, E. & Muñoz-Cánoves, P. Regulation of Muscle Stem Cell Functions: A Focus on the p38 MAPK Signaling Pathway. *Front Cell Dev Biol* **4**, (2016).
42. Krauss, R. S. & Kann, A. P. Muscle stem cells get a new look: Dynamic cellular projections as sensors of the stem cell niche. *BioEssays* **45**, (2023).
43. Mademtzoglou, D. & Relaix, F. From cyclins to CDKIs: Cell cycle regulation of skeletal muscle stem cell quiescence and activation. *Exp Cell Res* **420**, 113275 (2022).
44. Fu, X., Wang, H. & Hu, P. Stem cell activation in skeletal muscle regeneration. *Cellular and Molecular Life Sciences* **72**, 1663–1677 (2015).
45. Rodgers, J. T. *et al.* mTORC1 controls the adaptive transition of quiescent stem cells from G0 to GAlert. *Nature* **510**, 393–396 (2014).
46. Brun, C. E. *et al.* GLI3 regulates muscle stem cell entry into GAlert and self-renewal. *Nat Commun* **13**, 3961 (2022).
47. Schmidt, M., Schüler, S. C., Hüttner, S. S., von Eyss, B. & von Maltzahn, J. Adult stem cells at work: regenerating skeletal muscle. *Cellular and Molecular Life Sciences* **76**, 2559–2570 (2019).
48. Hernández-Hernández, J. M., García-González, E. G., Brun, C. E. & Rudnicki, M. A. The myogenic regulatory factors, determinants of muscle development, cell identity and regeneration. *Semin Cell Dev Biol* **72**, 10–18 (2017).
49. Madaro, L. *et al.* PKCθ signaling is required for myoblast fusion by regulating the expression of caveolin-3 and β1D integrin upstream focal adhesion kinase. *Mol Biol Cell* **22**, 1409–1419 (2011).
50. Halevy, O. *et al.* Pattern of Pax7 expression during myogenesis in the posthatch chicken establishes a model for satellite cell differentiation and renewal. *Developmental Dynamics* **231**, 489–502 (2004).
51. Blau, H. M., Cosgrove, B. D. & Ho, A. T. V. The central role of muscle stem cells in regenerative failure with aging. *Nat Med* **21**, 854–862 (2015).
52. Lin, H. Cell biology of stem cells: an enigma of asymmetry and self-renewal. *J Cell Biol* **180**, 257–260 (2008).
53. Feige, P., Brun, C. E., Ritso, M. & Rudnicki, M. A. Orienting Muscle Stem Cells for Regeneration in Homeostasis, Aging, and Disease. *Cell Stem Cell* **23**, 653–664 (2018).
54. Shea, K. L. *et al.* Sprouty1 Regulates Reversible Quiescence of a Self-Renewing Adult Muscle Stem Cell Pool during Regeneration. *Cell Stem Cell* **6**, 117–129 (2010).
55. Schofield, R. The relationship between the spleen colony-forming cell and the haemopoietic stem cell. *Blood Cells* **4**, 7–25 (1978).
56. Tidball, J. G. Regulation of muscle growth and regeneration by the immune system. *Nat Rev Immunol* **17**, 165–178 (2017).
57. Koike, H., Manabe, I. & Oishi, Y. Mechanisms of cooperative cell-cell interactions in skeletal muscle regeneration. *Inflamm Regen* **42**, 48 (2022).
58. Rahman, F. A., Angus, S. A., Stokes, K., Karpowicz, P. & Krause, M. P. Impaired ECM Remodeling and Macrophage Activity Define Necrosis and Regeneration Following Damage in Aged Skeletal Muscle. *Int J Mol Sci* **21**, 4575 (2020).
59. Lu, P., Takai, K., Weaver, V. M. & Werb, Z. Extracellular Matrix Degradation and Remodeling in Development and Disease. *Cold Spring Harb Perspect Biol* **3**, a005058–a005058 (2011).

60. Rigamonti, E., Zordan, P., Sciorati, C., Rovere-Querini, P. & Brunelli, S. Macrophage Plasticity in Skeletal Muscle Repair. *Biomed Res Int* **2014**, 1–9 (2014).
61. Dort, J., Fabre, P., Molina, T. & Dumont, N. A. Macrophages Are Key Regulators of Stem Cells during Skeletal Muscle Regeneration and Diseases. *Stem Cells Int* **2019**, 1–20 (2019).
62. Torres-Ruiz, J., Alcalá-Carmona, B., Alejandre-Aguilar, R. & Gómez-Martín, D. Inflammatory myopathies and beyond: The dual role of neutrophils in muscle damage and regeneration. *Front Immunol* **14**, (2023).
63. San Emeterio, C. L., Olingy, C. E., Chu, Y. & Botchwey, E. A. Selective recruitment of non-classical monocytes promotes skeletal muscle repair. *Biomaterials* **117**, 32–43 (2017).
64. Ziemkiewicz, N., Hilliard, G., Pullen, N. A. & Garg, K. The Role of Innate and Adaptive Immune Cells in Skeletal Muscle Regeneration. *Int J Mol Sci* **22**, 3265 (2021).
65. Yang, W. & Hu, P. Skeletal muscle regeneration is modulated by inflammation. *J Orthop Translat* **13**, 25–32 (2018).
66. Wang, X. & Zhou, L. The Many Roles of Macrophages in Skeletal Muscle Injury and Repair. *Front Cell Dev Biol* **10**, (2022).
67. Serrano, A. L., Baeza-Raja, B., Perdiguero, E., Jardí, M. & Muñoz-Cánoves, P. Interleukin-6 Is an Essential Regulator of Satellite Cell-Mediated Skeletal Muscle Hypertrophy. *Cell Metab* **7**, 33–44 (2008).
68. Li, Y. *et al.* Occurrences and Functions of Ly6Chi and Ly6Clo Macrophages in Health and Disease. *Front Immunol* **13**, (2022).
69. Forcina, L., Cosentino, M. & Musarò, A. Mechanisms Regulating Muscle Regeneration: Insights into the Interrelated and Time-Dependent Phases of Tissue Healing. *Cells* **9**, 1297 (2020).
70. Hardy, D. *et al.* Comparative Study of Injury Models for Studying Muscle Regeneration in Mice. *PLoS One* **11**, e0147198 (2016).
71. Sonavane, M. *et al.* Intramuscular Bleeding and Formation of Microthrombi during Skeletal Muscle Damage Caused by a Snake Venom Metalloprotease and a Cardiotoxin. *Toxins (Basel)* **15**, 530 (2023).
72. Sicherer, S. T., Venkatarama, R. S. & Grasman, J. M. Recent Trends in Injury Models to Study Skeletal Muscle Regeneration and Repair. *Bioengineering* **7**, 76 (2020).
73. Tierney, M. T. & Sacco, A. Inducing and Evaluating Skeletal Muscle Injury by Notexin and Barium Chloride. in 53–60 (2016). doi:10.1007/978-1-4939-3810-0_5.
74. Guardiola, O. *et al.* Induction of Acute Skeletal Muscle Regeneration by Cardiotoxin Injection. *Journal of Visualized Experiments* (2017) doi:10.3791/54515.
75. Morton, A. B. *et al.* Barium chloride injures myofibers through calcium-induced proteolysis with fragmentation of motor nerves and microvessels. *Skelet Muscle* **9**, 27 (2019).
76. Pereira, T. *et al.* Biomaterials and Stem Cell Therapies for Injuries Associated to Skeletal Muscular Tissues. in *Advances in Biomaterials Science and Biomedical Applications* (InTech, 2013). doi:10.5772/53335.
77. Bouredji, Z., Hamoudi, D., Marcadet, L., Argaw, A. & Frenette, J. Testing the efficacy of a human full-length OPG-Fc analog in a severe model of cardiotoxin-

induced skeletal muscle injury and repair. *Mol Ther Methods Clin Dev* **21**, 559–573 (2021).
78. Shelar, S. B. *et al.* Disruption of nuclear factor (erythroid-derived-2)-like 2 antioxidant signaling: a mechanism for impaired activation of stem cells and delayed regeneration of skeletal muscle. *FASEB J* **30**, 1865–79 (2016).
79. Plant, D. R., Colarossi, F. E. & Lynch, G. S. Notexin causes greater myotoxic damage and slower functional repair in mouse skeletal muscles than bupivacaine. *Muscle Nerve* **34**, 577–585 (2006).
80. Dessauge, F., Schleder, C., Perruchot, M.-H. & Rouger, K. 3D in vitro models of skeletal muscle: myopshere, myobundle and bioprinted muscle construct. *Vet Res* **52**, 72 (2021).
81. Kase, N., Kitagawa, Y., Ikenaka, A., Niwa, A. & Saito, M. K. A concise in vitro model for evaluating interactions between macrophage and skeletal muscle cells during muscle regeneration. *Front Cell Dev Biol* **11**, (2023).
82. Khodabukus, A., Prabhu, N., Wang, J. & Bursac, N. In Vitro Tissue-Engineered Skeletal Muscle Models for Studying Muscle Physiology and Disease. *Adv Healthc Mater* **7**, e1701498 (2018).
83. Brzeszczyńska, J. *et al.* Alterations in the *in vitro* and *in vivo* regulation of muscle regeneration in healthy ageing and the influence of sarcopenia. *J Cachexia Sarcopenia Muscle* **9**, 93–105 (2018).
84. Duval, K. *et al.* Modeling Physiological Events in 2D vs. 3D Cell Culture. *Physiology (Bethesda)* **32**, 266–277 (2017).
85. Habanjar, O., Diab-Assaf, M., Caldefie-Chezet, F. & Delort, L. 3D Cell Culture Systems: Tumor Application, Advantages, and Disadvantages. *Int J Mol Sci* **22**, 12200 (2021).
86. Madden, L., Juhas, M., Kraus, W. E., Truskey, G. A. & Bursac, N. Bioengineered human myobundles mimic clinical responses of skeletal muscle to drugs. *Elife* **4**, (2015).
87. Dunn, P. M. Thomas Malthus (1766-1834): population growth and birth control. *Arch Dis Child Fetal Neonatal Ed* **78**, F76–F77 (1998).
88. Watson, H. W. & Galton, F. On the Probability of the Extinction of Families. *The Journal of the Anthropological Institute of Great Britain and Ireland* **4**, 138 (1875).
89. Hao, W., Marsh, C. & Friedman, A. A Mathematical Model of Idiopathic Pulmonary Fibrosis. *PLoS One* **10**, e0135097 (2015).
90. Atwell, K., Dunn, S., Osborne, J. M., Kugler, H. & Hubbard, E. J. A. How computational models contribute to our understanding of the germ line. *Mol Reprod Dev* **83**, 944–957 (2016).
91. Waddington, C. H. The Strategy of the Genes. . *Google Scholar* (1957).
92. Wang, J., Zhang, K., Xu, L. & Wang, E. Quantifying the Waddington landscape and biological paths for development and differentiation. *Proceedings of the National Academy of Sciences* **108**, 8257–8262 (2011).
93. Guo, J., Lin, F., Zhang, X., Tanavde, V. & Zheng, J. NetLand: quantitative modeling and visualization of Waddington's epigenetic landscape using probabilistic potential. *Bioinformatics* **33**, 1583–1585 (2017).

94. Anwar, N., Ahmad, I., Kiani, A. K., Shoaib, M. & Raja, M. A. Z. Numerical treatment for mathematical model of farming awareness in crop pest management. *Front Appl Math Stat* **9**, (2023).
95. Sánchez-Romero, M. A. & Casadesús, J. Waddington's Landscapes in the Bacterial World. *Front Microbiol* **12**, (2021).
96. Till, J. E., McCulloch, E. A. & Siminovitch, L. A STOCHASTIC MODEL OF STEM CELL PROLIFERATION, BASED ON THE GROWTH OF SPLEEN COLONY-FORMING CELLS. *Proceedings of the National Academy of Sciences* **51**, 29–36 (1964).
97. Johnston, M. D., Edwards, C. M., Bodmer, W. F., Maini, P. K. & Chapman, S. J. Mathematical modeling of cell population dynamics in the colonic crypt and in colorectal cancer. *Proceedings of the National Academy of Sciences* **104**, 4008–4013 (2007).
98. Mamis, K., Zhang, R. & Bozic, I. Stochastic model for cell population dynamics quantifies homeostasis in colonic crypts and its disruption in early tumorigenesis. *bioRxiv* (2023).
99. Friedmann, E. PDE/ODE modeling and simulation to determine the role of diffusion in long-term and -range cellular signaling. *BMC Biophys* **8**, 10 (2015).
100. Kreutz, C. A New Approximation Approach for Transient Differential Equation Models. *Front Phys* **8**, (2020).
101. Kardynska, M., Kogut, D., Pacholczyk, M. & Smieja, J. Mathematical modeling of regulatory networks of intracellular processes – Aims and selected methods. *Comput Struct Biotechnol J* **21**, 1523–1532 (2023).
102. David S. Fischer *et al.* Beyond pseudotime: Following T-cell maturation in single-cell RNAseq time series. *bioRxiv* (2017).
103. Yamazaki, H., Tsuji, T., Doi, K. & Kawano, S. Mathematical model of the auditory nerve response to stimulation by a micro-machined cochlea. *Int J Numer Method Biomed Eng* **37**, (2021).
104. Gouzé, J.-L., Lasry, J.-M. & Changeux, J.-P. Selective stabilization of muscle innervation during development: A mathematical model. *Biol Cybern* **46**, 207–215 (1983).
105. Kim, O. D., Rocha, M. & Maia, P. A Review of Dynamic Modeling Approaches and Their Application in Computational Strain Optimization for Metabolic Engineering. *Front Microbiol* **9**, (2018).
106. Harischandra, N. *et al.* Evaluation of linear and non-linear activation dynamics models for insect muscle. *PLoS Comput Biol* **15**, e1007437 (2019).
107. Penas, D. R., González, P., Egea, J. A., Doallo, R. & Banga, J. R. Parameter estimation in large-scale systems biology models: a parallel and self-adaptive cooperative strategy. *BMC Bioinformatics* **18**, 52 (2017).
108. Waters, S. L., Schumacher, L. J. & El Haj, A. J. Regenerative medicine meets mathematical modelling: developing symbiotic relationships. *NPJ Regen Med* **6**, 24 (2021).
109. Paździorek, P. R. Mathematical Model of Stem Cell Differentiation and Tissue Regeneration with Stochastic Noise. *Bull Math Biol* **76**, 1642–1669 (2014).
110. Wadkin, L. E. *et al.* The recent advances in the mathematical modelling of human pluripotent stem cells. (2019).

111. Nguyen, T. N. T., Sasaki, K. & Kino-oka, M. Elucidation of human induced pluripotent stem cell behaviors in colonies based on a kinetic model. *J Biosci Bioeng* **127**, 625–632 (2019).
112. Deshpande, R. S. & Spector, A. A. Modeling Stem Cell Myogenic Differentiation. *Sci Rep* **7**, 40639 (2017).
113. Hall, M. E., Mohtaram, N. K., Willerth, S. M. & Edwards, R. Modeling the behavior of human induced pluripotent stem cells seeded on melt electrospun scaffolds. *J Biol Eng* **11**, 38 (2017).
114. Rashidi, N. M. *et al.* In vivo time-lapse imaging shows diverse niche engagement by quiescent and naturally activated hematopoietic stem cells. *Blood* **124**, 79–83 (2014).
115. Manicka, S., Johnson, K., Levin, M. & Murrugarra, D. The nonlinearity of regulation in biological networks. *NPJ Syst Biol Appl* **9**, 10 (2023).
116. Glass, D. S., Jin, X. & Riedel-Kruse, I. H. Nonlinear delay differential equations and their application to modeling biological network motifs. *Nat Commun* **12**, 1788 (2021).
117. Chulián, S. *et al.* Dynamical properties of feedback signalling in B lymphopoiesis: A mathematical modelling approach. *J Theor Biol* **522**, 110685 (2021).
118. Farhang-Sardroodi, S. & Wilkie, K. P. Mathematical Model of Muscle Wasting in Cancer Cachexia. *J Clin Med* **9**, 2029 (2020).
119. R Core Team. R: A Language and Environment for Statistical Computing. Preprint at (2023).
120. Gábor, A. & Banga, J. R. Robust and efficient parameter estimation in dynamic models of biological systems. *BMC Syst Biol* **9**, 74 (2015).
121. Aster, R. C., Borchers, B. & Thurber, C. H. *PARAMETER ESTIMATION AND INVERSE PROBLEMS*. (Elsevier Inc., 2019).
122. Riahi, M. K. & Qattan, I. A. On the convergence rate of Fletcher-Reeves nonlinear conjugate gradient methods satisfying strong Wolfe conditions: Application to parameter identification in problems governed by general dynamics. *Math Methods Appl Sci* **45**, 3644–3664 (2022).
123. Virgilio, K. M., Martin, K. S., Peirce, S. M. & Blemker, S. S. Agent-based model illustrates the role of the microenvironment in regeneration in healthy and *mdx* skeletal muscle. *J Appl Physiol* **125**, 1424–1439 (2018).
124. Westman, A. M., Peirce, S. M., Christ, G. J. & Blemker, S. S. Agent-based model provides insight into the mechanisms behind failed regeneration following volumetric muscle loss injury. *PLoS Comput Biol* **17**, e1008937 (2021).
125. Martin, K. S., Virgilio, K. M., Peirce, S. M. & Blemker, S. S. Computational Modeling of Muscle Regeneration and Adaptation to Advance Muscle Tissue Regeneration Strategies. *Cells Tissues Organs* **202**, 250–266 (2016).
126. Kojouharov, H. V., Chen-Charpentier, B. M., Solis, F. J., Biguetti, C. & Brotto, M. A simple model of immune and muscle cell crosstalk during muscle regeneration. *Math Biosci* **333**, 108543 (2021).
127. Fischer, D. S. *et al.* Inferring population dynamics from single-cell RNA-sequencing time series data. *Nat Biotechnol* **37**, (2019).
128. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* **6**, 377–382 (2009).

129. Rubenstein, A. B. *et al.* Single-cell transcriptional profiles in human skeletal muscle. *Sci Rep* **10**, 229 (2020).
130. Kalucka, J. *et al.* Single-Cell Transcriptome Atlas of Murine Endothelial Cells. *Cell* **180**, 764-779.e20 (2020).
131. Williams, K., Yokomori, K. & Mortazavi, A. Heterogeneous Skeletal Muscle Cell and Nucleus Populations Identified by Single-Cell and Single-Nucleus Resolution Transcriptome Assays. *Front Genet* **13**, (2022).
132. Mair, F. *et al.* A Targeted Multi-omic Analysis Approach Measures Protein Expression and Low-Abundance Transcripts on the Single-Cell Level. *Cell Rep* **31**, 107499 (2020).
133. Chen, M. & Zhou, X. Controlling for Confounding Effects in Single Cell RNA Sequencing Studies Using both Control and Target Genes. *Sci Rep* **7**, 13587 (2017).
134. Dell'Orso, S. *et al.* Single-cell analysis of adult skeletal muscle stem cells in homeostatic and regenerative conditions. *Development* (2019) doi:10.1242/dev.174177.
135. Oprescu, S. N., Yue, F., Qiu, J., Brito, L. F. & Kuang, S. Temporal Dynamics and Heterogeneity of Cell Populations during Skeletal Muscle Regeneration. *iScience* **23**, 100993 (2020).
136. Qin, T. *et al.* Single-cell RNA-seq reveals novel mitochondria-related musculoskeletal cell populations during adult axolotl limb regeneration process. *Cell Death Differ* **28**, 1110–1125 (2021).
137. Cho, D. S. & Doles, J. D. Single cell transcriptome analysis of muscle satellite cells reveals widespread transcriptional heterogeneity. *Gene* **636**, 54–63 (2017).
138. Giordani, L. *et al.* High-Dimensional Single-Cell Cartography Reveals Novel Skeletal Muscle-Resident Cell Populations. *Mol Cell* **74**, 609-621.e6 (2019).
139. De Micheli, A. J. *et al.* Single-Cell Analysis of the Muscle Stem Cell Hierarchy Identifies Heterotypic Communication Signals Involved in Skeletal Muscle Regeneration. *Cell Rep* **30**, (2020).
140. McKellar, D. W. *et al.* Large-scale integration of single-cell transcriptomic data captures transitional progenitor states in mouse skeletal muscle regeneration. *Commun Biol* **4**, 1280 (2021).
141. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289–1296 (2019).
142. Hie, B., Bryson, B. & Berger, B. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nat Biotechnol* **37**, 685–691 (2019).
143. Polański, K. *et al.* BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* **36**, 964–965 (2020).
144. Ingersoll, M. A. *et al.* Comparison of gene expression profiles between human and mouse monocyte subsets. *Blood* **115**, e10–e19 (2010).
145. Krasniewski, L. K. *et al.* Single-cell analysis of skeletal muscle macrophages reveals age-associated functional subpopulations. *Elife* **11**, (2022).
146. Jones, N. C. *et al.* The p38α/β MAPK functions as a molecular switch to activate the quiescent satellite cell. *J Cell Biol* **169**, 105–116 (2005).
147. Le Grand, F., Jones, A. E., Seale, V., Scimè, A. & Rudnicki, M. A. Wnt7a Activates the Planar Cell Polarity Pathway to Drive the Symmetric Expansion of Satellite Stem Cells. *Cell Stem Cell* **4**, 535–547 (2009).

148. Chau, J. *et al.* Relationship of *DUX4* and target gene expression in FSHD myocytes. *Hum Mutat* **42**, 421–433 (2021).
149. Kim, M. *et al.* Single-nucleus transcriptomics reveals functional compartmentalization in syncytial skeletal muscle cells. *Nat Commun* **11**, 6375 (2020).
150. Kann, A. P. & Krauss, R. S. Multiplexed RNAscope and immunofluorescence on whole-mount skeletal myofibers and their associated stem cells. *Development* (2019) doi:10.1242/dev.179259.
151. Hardy, D. *et al.* Comparative Study of Injury Models for Studying Muscle Regeneration in Mice. *PLoS One* **11**, e0147198 (2016).
152. Dumont, N. & Frenette, J. Macrophages Protect against Muscle Atrophy and Promote Muscle Recovery in Vivo and in Vitro. *Am J Pathol* **176**, 2228–2235 (2010).
153. Arnold, L. *et al.* Inflammatory monocytes recruited after skeletal muscle injury switch into antiinflammatory macrophages to support myogenesis. *J Exp Med* **204**, 1057–1069 (2007).
154. Tonkin, J. *et al.* Monocyte/Macrophage-derived IGF-1 Orchestrates Murine Skeletal Muscle Regeneration and Modulates Autocrine Polarization. *Molecular Therapy* **23**, 1189–1200 (2015).
155. Tidball, J. G. & Wehling-Henricks, M. Macrophages promote muscle membrane repair and muscle fibre growth and regeneration during modified muscle loading in mice *in vivo*. *J Physiol* **578**, 327–336 (2007).
156. Pizza, F. X., Peterson, J. M., Baas, J. H. & Koh, T. J. Neutrophils contribute to muscle injury and impair its resolution after lengthening contractions in mice. *J Physiol* **562**, 899–913 (2005).
157. Keefe, A. C. *et al.* Muscle stem cells contribute to myofibres in sedentary adult mice. *Nat Commun* **6**, 7087 (2015).
158. Schalbruch, H. & Hellhammer, U. [Satellite cells as myogenic stem cells and their incidence in normal human skeletal muscles]. *Verh Anat Ges* 1251–8 (1977).
159. Bruusgaard, J. C., Johansen, I. B., Egner, I. M., Rana, Z. A. & Gundersen, K. Myonuclei acquired by overload exercise precede hypertrophy and are not lost on detraining. *Proceedings of the National Academy of Sciences* **107**, 15111–15116 (2010).
160. Yin, H., Price, F. & Rudnicki, M. A. Satellite Cells and the Muscle Stem Cell Niche. *Physiol Rev* **93**, 23–67 (2013).
161. Domingues-Faria, C., Vasson, M.-P., Goncalves-Mendes, N., Boirie, Y. & Walrand, S. Skeletal muscle regeneration and impact of aging and nutrition. *Ageing Res Rev* **26**, 22–36 (2016).
162. Laumonier, T. & Menetrey, J. Muscle injuries and strategies for improving their repair. *J Exp Orthop* **3**, 15 (2016).
163. Domingues-Faria, C., Vasson, M.-P., Goncalves-Mendes, N., Boirie, Y. & Walrand, S. Skeletal muscle regeneration and impact of aging and nutrition. *Ageing Res Rev* **26**, 22–36 (2016).
164. Tidball, J. G. & Villalta, S. A. Regulatory interactions between muscle and the immune system during muscle regeneration. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **298**, R1173–R1187 (2010).
165. Novak, M. L., Weinheimer-Haus, E. M. & Koh, T. J. Macrophage activation and skeletal muscle healing following traumatic injury. *J Pathol* **232**, 344–355 (2014).

166. Musarò, A. The Basis of Muscle Regeneration. *Adv Biol* **2014**, 1–16 (2014).
167. Toumi, H., F'guyer, S. & Best, T. M. The role of neutrophils in injury and repair following muscle stretch. *J Anat* **208**, 459–470 (2006).
168. Wang, X., Zhao, W., Ransohoff, R. M. & Zhou, L. Infiltrating macrophages are broadly activated at the early stage to support acute skeletal muscle injury repair. *J Neuroimmunol* **317**, 55–66 (2018).
169. Yang, W. & Hu, P. Skeletal muscle regeneration is modulated by inflammation. *J Orthop Translat* **13**, 25–32 (2018).
170. Sloboda, D. D., Brown, L. A. & Brooks, S. V. Myeloid Cell Responses to Contraction-induced Injury Differ in Muscles of Young and Old Mice. *The Journals of Gerontology: Series A* **73**, 1581–1590 (2018).
171. Juhas, M. *et al.* Incorporation of macrophages into engineered skeletal muscle enables enhanced muscle regeneration. *Nat Biomed Eng* **2**, 942–954 (2018).
172. Buckley, K. H., Nestor-Kalinoski, A. L. & Pizza, F. X. Positional Context of Myonuclear Transcription During Injury-Induced Muscle Regeneration. *Front Physiol* **13**, (2022).
173. Liu, D. C. & Nocedal, J. On the limited memory BFGS method for large scale optimization. *Math Program* **45**, 503–528 (1989).
174. Lin Pedersen, T. patchwork: The Composer of Plots. . Preprint at (2023).
175. Wickham, H. *et al.* Welcome to the Tidyverse. *J Open Source Softw* **4**, 1686 (2019).
176. De Micheli, A. J., Spector, J. A., Elemento, O. & Cosgrove, B. D. A reference single-cell transcriptomic atlas of human skeletal muscle tissue reveals bifurcated muscle stem cell populations. *Skelet Muscle* **10**, (2020).
177. Scala, P. *et al.* Stem Cell and Macrophage Roles in Skeletal Muscle Regenerative Medicine. *Int J Mol Sci* **22**, 10867 (2021).
178. Ruffell, D. *et al.* A CREB-C/EBPβ cascade induces M2 macrophage-specific gene expression and promotes muscle injury repair. *Proceedings of the National Academy of Sciences* **106**, 17475–17480 (2009).