UNIVERSIDAD POLITÉCNICA DE MADRID
Escuela Técnica Superior de Ingenieros de Telecomunicación

Network modelling of the regulatory pathways involved in organ regeneration

DOCTORAL THESIS

Submitted for the degree of Doctor by:

Inés Rivero García
BSc Biochemistry and MSc Molecular Techniques in Life Science

Madrid, 2024

UNIVERSIDAD POLITÉCNICA DE MADRID
Escuela Técnica Superior de Ingenieros de Telecomunicación

Doctoral Degree in Biomedical Engineering

Network modelling of the regulatory pathways involved in organ regeneration

DOCTORAL THESIS

Submitted for the degree of Doctor by:

Inés Rivero García
BSc Biochemistry and MSc Molecular Techniques in Life Science

Under the supervision of:
Dr. Miguel Torres Sánchez

Madrid, 2024

Title: Network modelling of the regulatory pathways involved in organ regeneration
Author: Inés Rivero García
Doctoral Programme: Biomedical Engineering

Thesis Supervision:

Dr. Miguel Torres Sánchez, Full Professor, Centro Nacional de Investigaciones Cardiovasculares (CNIC)(Supervisor)

External Reviewers:

Thesis Defense Committee:

Thesis Defense Date:

This thesis has been partially supported by a fellowship from the ”la Caixa” Foundation (ID 100010434), with code LCF/BQ/DR20/11790019 and by the European Union’s Framework Programme for Research and Innovation “Horizon 2020” under grant agreement No. 874764.

"If the only tool you have is a hammer,
you tend to see every problem as a nail"
Abraham Maslow

Acknowledgement

I am beyond grateful to many individuals whose contributions and support made this thesis possible. Firstly, I want to deeply thank **Miguel Torres** and **Fátima Sánchez Cabo** for their mentorship. Miguel, thank you for giving me invaluable opportunities, the freedom to explore, and your constant availability when I needed help. Fátima, I cannot thank you enough for being a real co-supervisor despite the unfair lack of recognition from the University. Thank you for encouraging me to take on new challenges and dream big. I have learnt a lot from both of you and I aspire to be the trusted and inspiring leaders that you are.

I am incredibly grateful for having been part of two cohesive and wonderful groups at CNIC. I want to thank present and past members of **LAB MT** for nurturing an empowering team. This thesis would have not been possible without **Cristina Villa del Campo** and **Jorge Peña**, who put their hard work and talent to my disposal and took on the SLAM-seq challenge. Thank you Cris for trusting me to explore your data, I never imagined my thesis would lead to a patent application! I also want to thank **Irene Delgado** for supervising my very first internship in the lab and making me want to come back. I even managed to include limbs in a "heart" thesis. Thank you **Gloria** and **Consuelo** for your altruism; **Morena**, **Jyothi** and **Cova** for your energy; and **Miquel**, **Jorge** and **Óscar** for your moral-boosting sense of humor. I am also grateful for **Sandra**, **Vane**, **Susana** and **Lorena D.** because they are fundamental pillars for the lab, and not even a computational thesis would be possible without them. Thank you **Rocío**, **Lorena E.**, **Ester**, **Lin** and **Isaac** for your support during early times in my PhD, and thank you **Dima**, **Cielo Aparna**, and **Francesca** for your support during the final stages of it.

I am beyond grateful to the CNIC **BioIT Unit**. Having been surrounded by such a professional, kindhearted, and open team makes me feel privileged. Thank you **Carlos** and **Fernando** for sharing all your knowledge and solving all of my computer-related problems. Thank you **Dani** for your uplifting spirits and for being a great teacher. **Jorge**, I won't miss the powerball, but I will miss your jokes. **Juan Carlos** and **Manuel**, I have learnt a lot from you. To all the Juniors, I could have not asked for better company. **Lucía**, **Carlos Relaño**, **Sara**, **Jon**, **Álvaro** and **Juan**: thank you for making everyday fun. You really made a difference. Special thanks to **Diego** for being a supportive desk mate and revealing the solution to all problems: NMF. I am also thankful for the encouragement provided by **Adrián**, **Marta**, **Roxana**, **Jonatan** and **Mario** during the last months of my PhD. Lastly, I want to thank **Víctor** for being the SLAM-seq trigger.

Much of the work contained in this thesis would have not been possible without **Elly Tanaka**, who opened the doors of her lab and fully included me into her team. I am truly grateful to **Elad Bassat** and **Anastasia Polikarpova**, who shared their knowledge and projects with me and provided invaluable advice. I am also thankful for having learned from the axolotl bioinformatic genies **Jingkui Wang** and **Tobias Gerber**. The **Tanaka lab** alone could fill a whole research institute, and yet made me feel at home. Thank you for your insightful feedback, the fun beer hours, and for the memories discovering Vienna. **VRC Badminton** is also responsible for my fantastic time in this city. Special thanks to **Michael** and **Stefanie Wagner** for welcoming me into the team and to **Johannes Schöll** for putting me in touch with - in his words - einen Verein in dem Spaß kommt im Normalfall nicht zu kurz.

I am grateful for having being part of the **REANIMA** consortium, which allowed me to discover cardiobiology and other cultures. Special mention to **Marie Nehring** and **Benedetta Coppe** for sparkling up these meetings.

I feel grateful that I could do my PhD at **CNIC** with the support of **"la Caixa" Foundation**, as they provided me with support to develop my career.

Estoy muy agradecida de contar con mis amigos y familia, que me han apoyado desde mucho antes del comienzo de este viaje. **Laura**, **Marta**, **Marcos**, **Inés**, **Gonzalo**, **Antonio**, **Blanca**, **Leti**, **Maca** and **César**: estoy muy orgullosa de contar con vuestra amistad y deseo ver cuáles serán las próximas aventuras que compartamos. Especialmente, **Laura** y **Marta** han sido un pilar constante estos cuatro años y me han ayudado a desconectar y poner las cosas en perspectiva. **Inés**, gracias por acompañarme en tantas y variadas aventuras. Voy a echar mucho de menos al ente de "las Ineses". I also want to thank the **Ampayitos**, specially **Miren**, **Amparo**, **Sofia** and **Astrid**, as they kept me sane while applying for fellowships. Thank you **Kavan** for your inspirational curiosity and always lending a helping hand.

Por último, quiero agradecer a mi familia su apoyo incondicional. A mis padres, **Carmen** y **Alejandro**, por siempre ayudarme a luchar por mis metas, guiarme con buenas decisiones (por ejemplo, la bioinformática), dotarme de oportunidades que vosotros no tuvisteis y ser una red de seguridad infalible. A mi hermana, **Raquel**, porque admiro tu valentía y porque somos un equipo. Muchas gracias, **Rebaño**, por vuestro apoyo constante. Finally, thank you **Manu** for bridging thousands of kilometers with your care, encouragement and support. Thank you for making easy what's difficult and always believing in me.

Abstract

Motivation: Fibrotic diseases cause long-term organ damage and are a heavy burden on health systems worldwide. While adult mammals usually generate a fibrotic scar in response to injury, neonatal mammals and other vertebrate species can fully regenerate certain body structures through the dedifferentiation and proliferation of preexisting, terminally differentiated cells. Complex biological processes such as rewiring of cellular identity require a holistic approximation, as the study of isolated components is not enough to obtain a realistic overview of the underlying mechanism in action. Gene regulatory networks model transcriptional relationships between transcription factors and their target genes and are the main underlying determinants of cell identity.

Objectives: In this thesis we hypothesize that physiological regenerative mechanisms are controlled by latent gene regulatory networks at least partially conserved between lower vertebrates and mammals. We used state-of-the-art computational approaches on bulk and single-cell transcriptomic data to model and perturb the gene regulatory networks that control regeneration in heart and bone, in order to identify new candidate genes with therapeutic potential.

Findings: This thesis presents three sets of findings related to the understanding of gene regulatory networks in regeneration. (1) We inferred the gene regulatory network controlling axolotl cardiomyocyte sub-population transitions upon cryoinjury. By systematic *in silico* perturbations of the network, we identified the repressive role of *Hey2* for the initiation of the injury response in border zone cardiomyocytes. Independently, footprint-based analysis of a mouse scRNA-seq data set with annotated cardiomyocyte number of nuclei suggested *Hey2* as a positive regulator of cardiomyocyte polyploidization, which blocks regeneration. (2) Comparison of the transcriptional footprints caused by overexpression of *caErbb2*, *Myc* and *Yap5sa* in neonatal mouse cardiomyocytes identified two pro-regenerative gene expression programs, associated with cardiomyocyte proliferation and dedifferentiation. Boolean models of a core gene regulatory network controlling the emergence of these transcriptional states suggested the robust nature of the maturating cardiomyocyte gene regulatory network and identified a handful of perturbations with a potential therapeutic application. (3) We compared the gene regulatory networks between regenerating and fibrotic injuries in axolotl and identified an insufficient rewiring in the fibrotic network. We predicted partial antagonism in the function of the TCF/LEF transcription factors *Lef1* and *Tcf7l2* that might contribute to understand why fibrotic bone injuries fail to regenerate.

Conclusions: By leveraging different types of transcriptomic data and computational and experimental approaches, we were able to model the cardiomyocyte and fibroblast gene regulatory networks in regenerative and fibrotic conditions. Furthermore, we identified candidate genes for their validation as therapeutic pro-regenerative agents and provided mechanistic hypotheses to understand the differences between regenerative and fibrotic injury responses.

Resumen

Motivación: Las respuestas fibróticas dañan los órganos a largo plazo y representan un gran problema de salud a nivel mundial. Pese a que los mamíferos adultos responden con fibrosis ante la mayoría de las lesiones, los mamíferos neonatales y otras especies de vertebrados pueden regenerar completamente ciertas estructuras corporales mediante la desdiferenciación y proliferación de células preexistentes totalmente diferenciadas. Estos complejos procesos que implican cambios en la identidad celular requieren ser investigados de manera holística, ya que el estudio de componentes individuales no es suficiente para obtener una representación realista de los mecanismos de acción subyacentes. Las redes de regulación génica modelan las relaciones entre factores de transcripción y sus genes diana y son las principales determinantes de la identidad celular.

Objetivos: En esta tesis doctoral se hipotetiza que los mecanismos de regeneración endógena están controlados por redes de regulación génica latentes que están al menos parcialmente conservadas entre vertebrados inferiores y mamíferos. Centrándonos en el estudio de la regeneración de miocardio y hueso, empleamos aproximaciones computacionales novedosas y datos de secuenciación masiva de ARN de tejidos homogeneizados y células individuales para modelar y perturbar las redes de regulación génica que controlan la regeneración en estos dos tejidos, e identificar nuevos candidatos con potencial terapéutico.

Hallazgos: En esta tesis se presentan tres proyectos distintos, todos ellos enfocados en mejorar nuestra comprensión de las redes de regulación génica que subyacen a los procesos fibróticos y regenerativos. (1) Se modeló la red de regulación génica que controla las transiciones entre subpoblaciones de cardiomiocitos de ajolote tras una criolesión. La perturbación sistemática *in silico* de esta red, identificó a *Hey2* como represor de la iniciación de la respuesta a herida en los cardiomiocitos. De manera independiente, datos scRNA-seq de cardiomiocitos murinos con anotación del número de núcleos se analizaron empleando huellas transcriptómicas y se identificó el posible rol de *Hey2* para inducir poliploidización, barrera para la proliferación de cardiomiocitos. (2) Se compararon las huellas transcriptómicas causadas por la sobre-expresión de *caErbb2*, *Myc* y *Yap5sa* en cardiomiocitos murinos neonatales. Se identificaron dos programas de expresión génica pro-regenerativos, asociados con la proliferación y desdiferenciación de cardiomiocitos, respectivamente. Mediante modelos Booleanos de la red de regulación génica que controla la aparición de estos estados transcripcionales, se predijo la robustez de la red de regulación génica que promueve la maduración de cardiomiocitos neonatales e identificaron un grupo de perturbaciones con potencial pro-regenerativo. (3) Se compararon las redes de regulación génica de lesiones regenerativas y fibróticas en el ajolote. Se identificó una remodelación insuficiente en la red de regulación génica fibrótica y se predijo el antagonismo funcional parcial entre los factores de transcripción *Lef1* y *Tcf7l2*, de la familia TCF/LEF, como causa contribuyente a la respuesta fibrótica en ciertas lesiones óseas.

Conclusiones: Mediante el uso de datos transcriptómicos y diferentes estrategias computacionales y experimentales, en esta se tesis modelan las redes de regulación génica en cardiomiocitos y fibroblastos en condiciones pro- y anti-regenerativas. Además, en esta tesis doctoral se identifican genes candidatos para su validación como agentes terapéuticos pro-regenerativos y se proporcionan hipótesis mecanísticas para entender las diferencias entre respuestas regenerativas y fibróticas a lesiones.

Table of Contents

List of Figures

List of Tables

Abbreviations and acronyms

AAV	Adeno-asociated viral vector
AEC	Apical ectodermal cap
AMI	Acute myocardial infarction
AUROC	Area under the receiver operating characteristic curve
bHLH	Basic helix-loop-helix
caErbb2	Constitutively active *Erbb2*
cDNA	Complementary DNA
CM	Cardiomyocyte
CNIC	Centro Nacional de Investigaciones Cardiovasculares
CNN	Convolutional neural network
CNNC	Convolutional Neural Network for Coexpression
CSD	Critical size defect
CT	Connective tissue
dpi	Days post-injury
EdU	5-Ethynyl-2’-deoxyuridine
GoM	Grade-of-membership
GRN	Gene regulatory network
GSEA	Gene Set Enrichment Analysis
HF	Heart failure
I/R	Ischemia-reperfusion injury
IAA	Iodoacetamide
IMP	Institute of Molecular Pathology
IPA	Ingenuity Pathway Analysis

IS	Injury-specific
KO	Knockdown
LAD	Left anterior descending coronary artery
MAP	Mode of the posterior
mGFP	Membrane GFP
MNDCM	Mononucleated diploid cardiomyocyte
MOI	Multiplicity of infection
MOR	Mode of regulation
OCT	Optimal Cutting Temperature
ODE	Ordinary differential equation
OFT	Outflow tract
ORA	Over-representation analysis
PN	Post-natal day N
PC	Principal component
PCA	Principal component analysis
PCP	Planar cell polarity
pHH3	Phospho-histone 3
PKN	Prior knowledge network
Prol	Proliferative
PS	Perturbation score
RA	Retinoic acid
rcf	Relative centrifugal force
RTK	Receptor tyrosine kinase
s4U	4-thiouridine
scRNA-seq	Single-cell RNA sequencing

SLAM-seq	Thiol(SH)-linked alkylation for the metabolic sequencing of RNA
snRNA-seq	Single-nuclei RNA sequencing
TAD	Topologically associating domain
TCF/LEF	T-cell factor/Lymphoid enhancer factor
TE	Transfer entropy
TF	Transcription factor
TSS	Transcription start site
UMAP	Uniform manifold approximation and projection
UPM	Universidad Politécnica de Madrid
VST	Variance stabilizing transformation
WE	Wound epidermis
WM	Working myocardium
WT	Wild-type
Yap5sa	Constitutively active *Yap1*

1 | Introduction

1.1 Motivation

Regenerative medicine aims to restore damaged tissues and organs to their functional state [1]. This ambitious goal of functional restoration, instead of symptomatic management, has the potential to revolutionize healthcare by providing more sustainable and long-lasting solutions for a wide range of diseases. Adult mammals generally respond to injury with fibrotic scarring and impairment of organ function, but some lower vertebrates and neonatal mammals can fully regenerate certain body structures [2–4]. These observations support the possibility of reactivating endogenous regenerative pathways that are dormant in adult humans. This doctoral thesis is focused on understanding endogenous mechanisms for heart and bone regeneration.

Cardiovascular disease is the main cause of death globally, accounting for over 30% of annual global deaths [5]. An acute myocardial infarction (AMI) occurs when an occlusion of the coronary arteries deprives the myocardium from oxygen and nutrients, causing extensive cardiomyocyte (CM) death within hours. Between 3 to 5 days after the AMI, the lost CMs are replaced by a fibrotic scar produced by activated fibroblasts and, over time, the pathological tissue remodeling is extended and leads to heart failure (HF). To date, pharmacological therapies to reduce HF mortality exist [6], but the only cure against HF is heart transplantation, which has a limited applicability [7]. It is therefore crucial to find new therapeutic avenues that achieve a high functional improvement and can be applied to a wide spectrum of patients. Basic research has shown that lower vertebrates and neonatal mouse pups can regenerate their myocardium after an ischemic injury through the dedifferentiation and proliferation of pre-existing CMs [2, 4, 8], which suggests that the regenerative capacity of the adult heart might not irreversibly lost, but rather insufficiently activated.

In spite of their lower mortality, large bone fractures constitute a public health issue due to their negative impact on quality of life and health-care costs [9]. In particular, bone critical size defects (CSD) are caused by severe trauma, osteoporosis, or osteosarcoma and affect more than 30% of the bone length, which makes them too large for spontaneous healing [10]. Traditional casting or fixation is not sufficient to fully stabilize these fractures, which normally lead to fibrotic deposits and bone non-union defects, which cause severe pain, deformities, and instability [10]. While mammalian embryos can achieve limb regeneration, this ability is progressively lost with age, being neonatal mice only able to regenerate digital tips after a distal injury [11, 12]. The axolotl can scarlessly regenerate fully patterned and functional limbs after amputation, even at adult stages of its life cycle, thanks to the formation of a highly proliferative pool of progenitors known as blastema [13].

Compared to heart regeneration, blastema formation requires more extensive dedifferentiation and recapitulation of embryonic gene expression programs [14].

Regardless of the biological context, regenerative processes require at least partial changes in identity, encoded by a complex gene regulatory network (GRN) in which transcription factors (TFs) regulate the expression of target genes [15]. Cellular identity is therefore not a simple combination of TF activity levels, but rather emergent from these and the regulatory interactions between them [15]. GRNs can be modeled as networks in which the nodes represent genes and edges represent transcriptional regulation between them, which may be directed, signed and/or weighted [16]. GRNs have been traditionally inferred from microarrays or bulk RNA-seq data [17–19], but the availability of single-cell and single-nuclei RNAseq (scRNA-seq and snRNA-seq, respectively) has truly transformed the field. A wide variety of methodologies have been developed for the complicated process of GRN inference [20–22], providing useful models to generate molecular mechanistic hypotheses and identify candidate regulators with therapeutic potential [23].

1.2 Objectives

Throughout this thesis we hypothesize that endogenous regenerative mechanisms are controlled by latent GRNs at least partially conserved between lower vertebrates and mammals. Their reactivation by targeting key regulators might provide an effective therapeutic avenue to regenerate damaged tissue. We use a variety of methodological approaches and types of transcriptomic data to infer, model, and understand the GRNs controlling myocardial and bone regeneration, defining the following specific objectives:

1. Model the CM GRN that allows the regeneration of the injured axolotl heart and identify the transcriptional hubs that orchestrate this process. This objective is addressed in Chapter 3
2. Identify new therapeutic hubs to activate endogenous regenerative mechanisms in the adult mammalian myocardium:
    - 2.1 Identify molecular drivers of mammalian CM polyploidization and explore their potential to induce CM proliferation. This objective is addressed in Chapter 3.
    - 2.2 Infer the GRNs that control the pro-regenerative gene expression programs induced by *Erbb2*, *Yap1* and *Myc* and find new transcriptional hubs with similar footprints. This objective is addressed in Chapter 4
3. Model the GRN in regenerative and fibrotic axolotl bone injuries and propose candidate molecular mechanisms to transform the fibrotic response into a regenerative response. This objective is addressed in Chapter 5.

1.3 Chapter outline

Chapter 2 provides a general introduction to the fields of endogenous regeneration and GRN inference, focusing on the cellular and molecular mechanisms behind heart and limb regeneration, as well as the state-of-the-art methodologies and challenges of GRN inference using bulk and sc/snRNA-seq. Then, the three research projects that compose this doctoral thesis are structured as follows:

- **Chapter 3: *Hey2* as a roadblock for vertebrate heart regeneration**. In this chapter we infer the GRN controlling the CM response to a ventricular cryoinjury in axolotl and identify its key transcriptional regulators. Additionally, we integrate scRNA-seq and a manually curated database of literature findings to identify TF that regulate the polyploidization of mammalian CM.
- **Chapter 4: Network modeling of the pro-regenerative effect of *caErbb2*, *Myc* and *Yap5sa* on CMs**. We overexpress the pro-regenerative genes *Yap5sa*, *caErbb2* and *Myc* in neonatal mouse CMs and analyze their transcriptional footprints using SLAM-seq and grade-of-membership (GoM) models. We use Boolean models to predict the status of the core GRN orchestrating these footprints, obtained through the intregration of scRNA-seq and a database of transcriptional regulation. We simulate *in silico* perturbations to find new regulators with similar regenerative potential.
- **Chapter 5: Inference of the gene regulatory network of axolotl bone regeneration**. We profile a regenerative (blastema) and non-regenerative (CSD) injury in axolotl hindlimbs using scRNA-seq. We infer transcriptome-wide GRNs controlling both injury responses and use network topology analysis to find structural differences between them. We construct a candidate prioritization pipeline to select potentially pro-regenerative TF and predict the effect of their perturbation using *in silico* experimentation.

The three projects are tied together in **Chapter 6**, in which we provide a general discussion and reflect on the limitations of this work. Lastly, in **Chapters 7** (English) and **8** (Spanish) the conclusions that can be drawn from this doctoral work are presented.

1.4 Contributions

- Torres M, Villa del Campo C, **Rivero García I**, inventores; Centro Nacional de Investigaciones Cardiovasculares (CNIC), titular. *Hey2* inhibition for the use in the treatment of cardiac diseases in mammals. European Patent application number EP 24 382 533.8. 17 mayo 2024.
- **Rivero-García I**, Torres M, Sánchez-Cabo F. Deep generative models in single-cell omics. Comput Biol Med. 2024;176:108561. doi: 10.1016/j.compbiomed.2024.108561 [24].
- Mañanes D, **Rivero-García I**, Relaño C, Torres M, Sancho D, Jimenez-Carretero D, Torroja C, Sánchez-Cabo F. SpatialDDLS: an R package to deconvolute spatial transcriptomics data using neural networks. Bioinformatics 2024;40(2):btae072. doi: 10.1093/bioinformatics/btae072 [25].
- **Rivero-García I**, Villa del Campo C, Peña-Peña J, Sánchez-Cabo F, Torres M. Network modelling of the regulatory pathways involved in cardiac regeneration. *Manuscript in preparation.*
- Villa del Campo C, **Rivero-Garcia I**, Sierra R, Calleja S, Dopazo A, Torres M. Single cell transcriptomics of adult mononucleated cardiomyocytes reveals a single transcription factor that regulates CM polyploidization in mammalian hearts. *Manuscript in preparation.*
- Polikarpova A, **Rivero-García I**, Gerber T, Wang J, Torres M, Sánchez-Cabo F, Tanaka EM. Cellular and molecular profiling of critical bone fractures in axolotl. *Manuscript in preparation.*
- Bassat E, Wang J, Peña-Peña J, **Rivero-Garcia I**, Piszczek A, Falcon-Chavez FJ, Sanchez-Cabo F, Torres M, Tanaka E. Spatial transcriptomic analysis of axolotl cardiac regeneration reveals AXL as a pro-regenerative factor. *Manuscript in preparation.*

2 | Background

2.1 Regeneration

Regenerative medicine aims to address unmet medical challenges involving loss of organ function due to fibrotic healing through the development of technologies that can revert tissue damage [1, 26]. A large variety of therapeutic strategies are being explored, including *in situ* regeneration, injection of stem cells and transplantation of tissues engineered *in vitro* [27]. *In situ* tissue regeneration relies on reprogramming adult and functional resident cell populations through the rewiring of their GRNs and is based on the observed ability of other species, including vertebrate and neonatal mammals, to successfully regenerate damaged body parts [4, 13, 28–31]. Despite requiring a deeper understanding of the molecular underpinnings of regeneration than other strategies, *in situ* regeneration might provide a more physiological, less immunogenic, and more affordable therapeutic strategy as no exogenous cells are introduced into the body and biopharmaceuticals could be enough to re-activate dormant GRNs.

Regeneration occurs to a different extent in a large variety of taxonomic groups along the animal kingdom [32, 33] (Fig. 2.1). The most extreme type of regeneration is whole body regeneration, observed in the fresh water hydrozoan Hydra and the bilaterally symmetrical planaria, which are able to grow a complete organism from a fragment of it [34]. Structural regeneration is observed in salamander limb or zebrafish fin regeneration, where a multicellular structure involving 3D growth of complex tissues is formed [34]. More modest regenerative processes are organ regeneration, in which an organ is restored after injury (e.g., heart regeneration in zebrafish, lens regeneration in salamander. . .), and tissue regeneration, which closes gaps in homogeneous cell populations (e.g., epidermis and gut lining regeneration in Drosophila) [34]. Lastly, cellular regeneration, such as the restoration of severed axons, is the simplest form of regeneration that has been observed [35]. In most circumstances, mammals respond to injury with the formation of a fibrotic scar that inhibits regenerative processes and normal tissue function [36]. However, mammals have some potential for regeneration. Adult mammals can, to some extent, regenerate small bone fractures, skeletal muscle [37], peripheral nerves [38], liver [31], and hippocampus [39] among others. Adult deer and moose fully regenerate their antlers annually, showcasing an exceptional regenerative capacity for mammals [40]. Within the mammalian class, the spiny mouse (*genus Acomys*) might be the best performer, surpassing other mouse strains in skin and ear punch wound regeneration [29, 30]. Importantly, age exerts a negative influence on mammalian regeneration: unlike their adult counterparts, neonatal mice can regenerate the heart during the first postnatal days of life [4]. Understanding these differences in the success of mammalian regeneration is key for the successful translation of *in situ* regeneration to the clinics.

Incrementing the complexity of the regeneration problem, different structures follow different mechanisms for regeneration [36]. In the case of the liver, the proliferation of parenchymal cells restores the tissue. Other tissues, such as skeletal muscle, rely on tissue-specific progenitor cells for regeneration [36]. In this doctoral work we focus on more complicated regenerative processes in which adult, functional cells must dedifferentiate and proliferate to recover the damaged tissue. Two examples of structures following this regenerative avenue are the heart and the limb [4, 13].

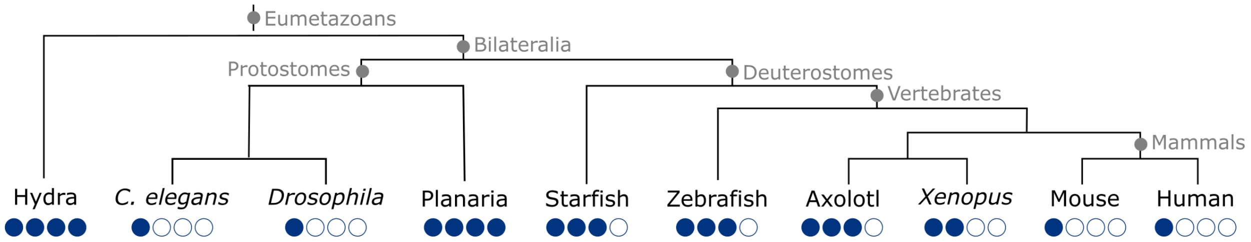


Figure 2.1: Animal models commonly used in regenerative studies. For each model, their regenerative capacity is indicated by the colored circles as very high (●●●●), high (●●●○), medium (●●○○), or low (●○○○).

2.1.1 Model organisms to study regeneration

One of the historic hindrances in the study of regeneration is the lack of regenerative abilities in common model organisms for which the molecular toolbox is highly developed. This has raised the popularity of other model organisms for regenerative studies [41, 42]. Here we briefly describe the model organisms relevant for this doctoral thesis.

Zebrafish: *Danio rerio*, commonly known as zebrafish, is a small fresh water teleost able to regenerate multiple organs throughout life, including its heart, caudal fin, retina and brain [2, 43–45]. Its small size, easy maintenance, and the availability of genetic, -omics, live imaging, and injury models make the zebrafish a highly accessible vertebrate model for regenerative studies [46]. While not being the focus of this thesis, the zebrafish is an essential model in the heart regeneration field to which the mouse and axolotl are compared to.

Axolotl: the axolotl (*Ambystoma mexicanum*) is a type of urodele amphibian with remarkable regeneration abilities, being capable of regrowing tissues, organs, and whole-body structures [47]. Axolotls are fully aquatic and remain neotenic, retaining larval features such as external gills throughout their life and living in warm waters with a constant temperature of around 20°C [47, 48]. However, their husbandry is complicated by their large size, which can reach 30 cm at adulthood, and by their long life cycle, as axolotls reach sexual maturity at 1 year of age [48]. Additionally, their genome spreads over 32Gb and is highly repetitive [48]. Despite these drawbacks, the axolotl is the most used salamander model in research because its genome has been annotated, genetic tools have been developed and the semi-transparent *d/d* line lacking pigmented cells in the skin has been generated [47, 48]. The axolotl has served as a classical system for limb regeneration studies [14, 49–53], and has recently been in the focus of molecular studies underpinning cardiac regeneration [54], although this observation was initially described fifty years ago [8].

Mouse: the mouse (*Mus musculus*) is the model organism by excellence and has been widely employed to investigate developmental and pathological processes. However, it's regenerative ability is limited to certain structures and rapidly decreases with age. During embryonic stages, the mouse can fully regenerate structures as complicated as the skin [55], limb buds [11] and the heart, being only the latter scarlessly regenerated during the first postnatal days of life [4]. Neonatal regeneration in mouse is mostly limited to simpler structures, such as hair follicles [56] or hair cells of the inner ear [57], and becomes abolished after the first week(s) of post-natal life. Regeneration in the adult mouse has only been reported to occur in skeletal muscle [58], digital tips after a distal injury [59], and liver [60], being the development of a fibrotic scar the most common wound response in this organism and stage [61]. The genus *Acomys*, also known as spiny mouse, is a newcomer in the field of regenerative biology and showcases remarkable regenerative abilities for a mammal, as it fully regenerates skin [30], ear punches [29], kidney [62] and spinal cord injuries [63] among others.

2.1.2 Heart regeneration

The lack of CM renewal transforms an AMI into chronic HF

The heart is a central organ for body function and constitutes the first organ to be formed during embryonic development [64]. Owing to the contractile force generated by the billions of CMs in the adult human heart, it pumps blood to provide all tissues and organs with nutrients and oxygen. Alongside CMs, the heart contains a myriad of other cell types, such as fibroblasts, immune cells, adipocytes, neuronal and endothelial cells [64]. While these cell types are essential for heart function, the focus of this thesis lies in CMs.

An AMI is caused by a block in the coronary arteries that depletes the myocardial muscle from oxygen and nutrients, causing extensive necrosis. Despite border zone CMs slightly increasing their proliferation after myocardial ischemia, cytokinesis does not occur frequently enough to regenerate the lost myocardium [65]. Instead, the lost muscle tissue is replaced by a fibrotic scar that impairs heart contractility, induces pathological remodeling and arrhythmia, and leads to end-stage HF [64–67] (Fig. 2.2). Improved clinical interventions, including percutaneous coronary intervention, and pharmacological treatments, such as ACE inhibitors and β-blockers, have increased the initial chances of survival and delayed the disease progression, but the myocardial loss is only addressed by heart transplantation, which is not applicable to all affected patients [68]. Therapeutic approaches that aim to regenerate the lost myocardium could be a safe, efficient, and accessible way to prevent chronic HF, but they must address the following challenges:

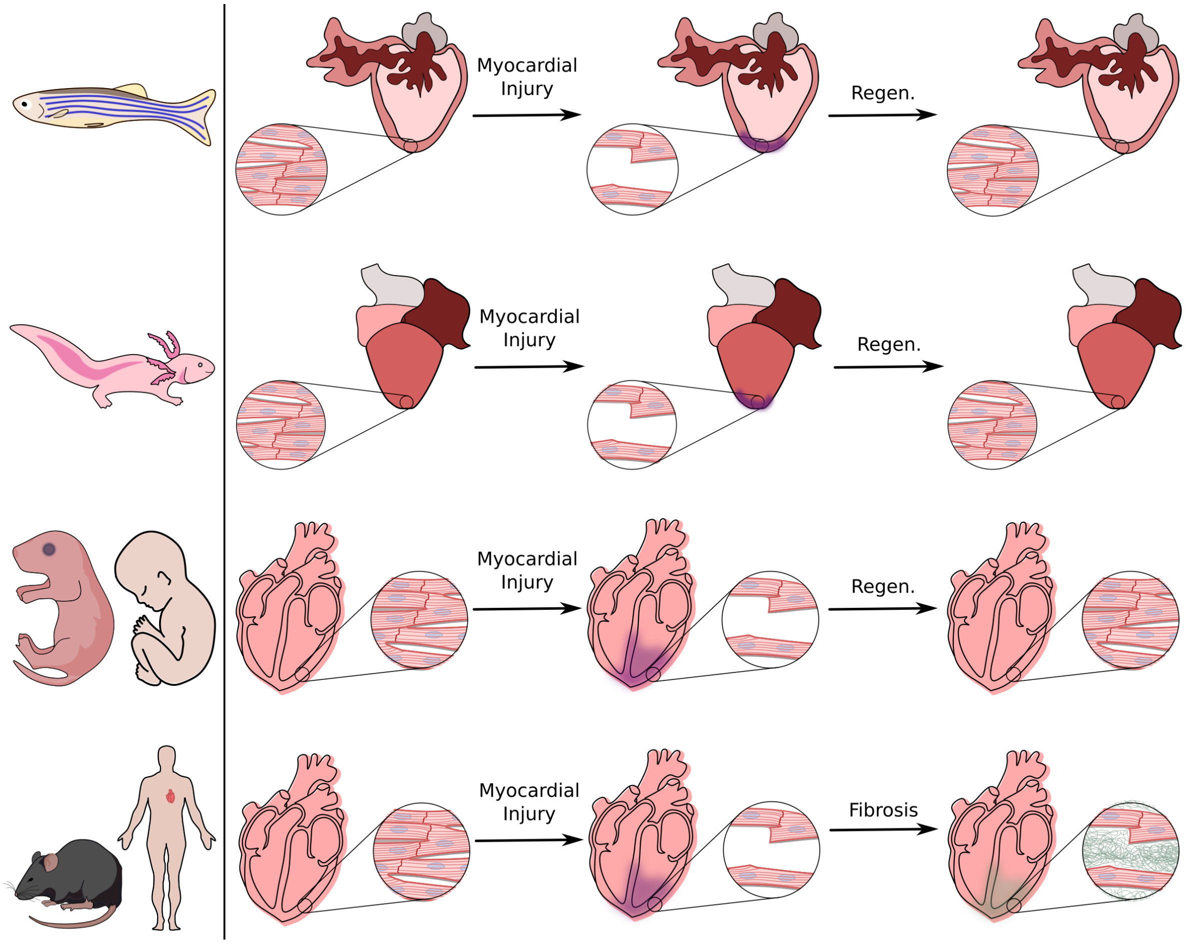


Figure 2.2: Response to heart injury in regenerative and non-regenerative vertebrates. The zebrafish heart contains one ventricle, one atrium and the bulbus arteriosus. A myocardial injury induces extensive CM death but activates the proliferation of preexisting CMs, regenerating the lost myocardium. The axolotl heart mounts a similar response to myocardial injury, in spite of its larger size and more complex anatomy, with two atria and one ventricle. The mammalian heart presents two ventricles and two atria. While neonates can activate the regenerative mechanisms that recover the lost myocardium, adult mammals respond to injury with the generation of a fibrotic scar rich in collagen fibers that causes long term HF. Regen. stands for regeneration.

- Mature CMs fail to complete the cell cycle: the heart grows by hyperplasia and hypertrophy during embryonic and postnatal development, respectively [65]. The adult mammalian CM cell cycle dynamics abruptly change soon after birth, causing cell cycle arrest at G2/M and increasing CM ploidy through endoreplication [69]. As a result, the population of mononucleated diploid CMs (MNDCMs) is largely replaced by polyploid CMs with predominantly one (human) or two (mouse) nuclei [64]. Despite polyploidy not being a general block for cell division, a study by González-Rosa *et al.* showed that the zebrafish myocardium can only regenerate when CMs are MNDCM, and that increasing the proportion of polyploid CMs hindered regeneration [70].

- Maturation of the CM sarcomeric machinery: the postnatal increase in ventricular pressure shifts sarcomeric size and protein composition to achieve increased contractility [71]. These complex sarcomeres are difficult to disassemble and interfere with cell proliferation.

- Maturation of the CM metabolism: coincidental with this increased load is an increased exposure to oxygen and higher energetic demands, driving a glycolytic-to-oxidative switch in the CM metabolism. The resulting increase in oxidative stress and consequent DNA damage might contribute to the CM cell cycle arrest [65].

In vivo models of heart injury

Different animal models of myocardial injury and ischemia have been developed to mimic the human pathology (Table 2.1).

Table 2.1: Animal models used in the study of heart regeneration.

Injury	Description	Organisms
Apical resection	The heart tip is cut in a simple surgery, but it does not involve resolution of necrosis.	Zebrafish, axolotl, neonatal mouse.
Genetic ablation	A cell type is deleted due to the cell-type specific expression of a cytotoxic protein. It clarifies cell-type specific contributions to regeneration and involves the resolution of necrosis.	Zebrafish, mouse embryo.
Cryoinjury	A flash-frozen metal probe is used to damage the heart, generating extensive necrosis with a simple surgery.	Zebrafish, axolotl, mouse, pig.
LAD ligation	The left anterior descending coronary artery (LAD) is ligated, more realistically mimicking the characteristic hypoxia and necrosis of an AMI.	Mouse, pig.
Ischemia-reperfusion injury (I/R)	The temporal LAD ligation is followed by a reperfusion phase. It is the most physiologically accurate model.	Mouse, pig.

Evidence to support adult human heart regeneration

Despite the inability of the adult mammalian heart to regenerate, heart regeneration has been described in other vertebrates and younger mammals, including zebrafish, axolotl, and neonatal (until post-natal day P7) mice [2, 4, 8]. While the exact response to injury and time to regenerate the myocardium depends on the organism and injury models, there are some key differences with the adult mammalian response. Firstly, while these organisms also form a scar, it is mainly composed by fibrin fibers, with scarce collagen deposition, which might facilitate its clearance. Secondly, while adult mammalian border zone CMs enter the cell cycle, they do not proliferate enough to replace the lost myocardium and, in most cases, grow by hypertrophy and endoreplication for functional compensation. On the other hand, regenerative organisms can induce massive CM proliferation to fully replace the lost tissue [66, 72]. Lineage tracing studies in these models have shown that pre-existing CM, and not stem cell populations, generate the new myocardium through their dedifferentiation and proliferation [73] (Fig. 2.2). Bergmann and collages leveraged the pulse-chase-like atmospheric presence of 14C during the cold war to date cells, showing that human adult MNDCMs retain some proliferative capacity and undergo annual age-dependent renewal, with 1% and 0.4% new CM per year at ages 20 and 75, respectively [74]. Therefore, this evidence shows that the adult human heart retains low levels of CM proliferation and supports the hypothesis that mechanisms able to regenerate the myocardium are not lost in humans, but rather insufficiently activated. In fact, there is clinical evidence of human heart regeneration in newborn children with severe AMI that, despite significant early damage, fully restored heart function [75, 76].

Strategies to induce adult human heart regeneration

***In situ* regeneration through the activation of endogenous pro-regenerative signaling pathways:** previous studies have identified endogenous pathways that can regenerate the adult mammalian heart upon activation. Three of them play a central role in this thesis:

- The Hippo pathway and *Yap1*: the Hippo pathway is a mechanosensing kinase cascade that controls organ size [72]. At the end of this cascade are the transcriptional co-activators YAP/TAZ, which become ubiquitinated and proteasomally degraded upon phosphorylation. When the Hippo pathway is inactive, YAP/TAZ proteins translocate to the nucleus and activate the expression of proliferative genes through the activation of the TEAD TFs [77] (Fig. 2.3a). *Yap1* expression progressively decreases during heart development, while its phosphorylation quickly increases between P2 and P10 [78]. Mice with abruptive Hippo signaling (*Sav1* knockdown (KO)) showed enhanced CM proliferation and reduced scar size following AMI [78]. Overexpression of the constitutively active form of *Yap1* (*Yap5sa*) reprograms the CM transcriptome into a fetal-like state and significantly increases CM proliferation [79].

- *Nrg1* signaling through *Erbb2/4*: *Nrg1* is an epidermal growth factor with a major role in cardiac development [80, 81]. In zebrafish, myocardial injury induces *Nrg1* expression in the outer heart wall, which binds the ERBB2/4 receptor tyrosine kinases. This induces the heterodimerization and autophosphorylation of the ERBB receptors, which in turn activates an intracellular cascade that increases CM survival and proliferation [82] (Fig. 2.3b). D'Uva *et al.* showed that *Erbb2* expression decreases after birth, and that transient induction of a constitutively active *Erbb2* (*caErbb2*) reactivates CM proliferation and achieves heart regeneration in adult mice [83].

- *Myc*: the family of *Myc* proto-oncogenes (*Myc, Mycn and Mycl*) regulates G1-to-S cell cycle progression and ribosome biosynthesis processes [84] (Fig. 2.3c). *Mycn* is exclusively responsible for CM proliferation during heart development, but its function can be completely replaced by

Myc [85]. In neonatal and adult mice, *Myc* overexpression increases CM proliferation, but only if the positive transcription elongation factor component *Cctn1* is available [86]. Transient and coordinated overexpression of these two genes increases CM proliferation in the adult heart and recovers heart function post-AMI [87].

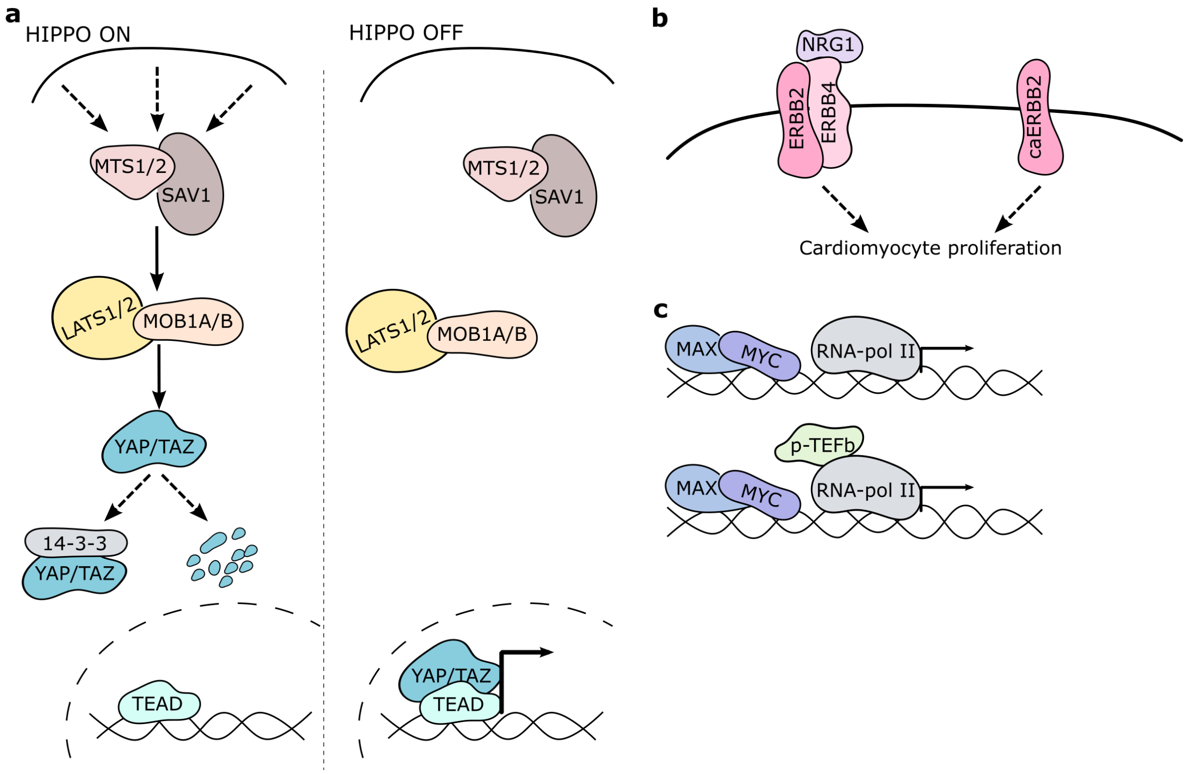


Figure 2.3: Signaling pathways that induce mammalian heart regeneration. **a)** Hippo signaling. Hippo activity leads to the degradation or cytoplasmic sequestration of the transcriptional co-activator YAP/TAZ. When the Hippo pathway is inactive, YAP/TAZ translocates to the nucleus and activates gene expression though the TEAD TFs. **b)** ERBB2 dimerizes with ERBB4 to mediate NRG1 signaling and induce CM proliferation and survival. The constitutively active form of ERBB2 (caERBB2) can activate this pathway without the need of NRG1 presence. **c)** The MYC TF activate gene expression through the heterodimerization with MAX. Presence of the p-TEFb complex potentiates MYC transcriptional activation.

Injection of pluripotent or multipotent progenitor cells: as an alternative to endogenous regeneration, several studies have explored the potential of progenitor cells to regenerate the failing heart [88–90]. In contrast with these efforts, a long list of comprehensive lineage tracing studies has shown that pre-existing CM are the only source of new CM in the postnatal myocardium [73, 91, 92]. The injection of several sorts of pluri- and multipotent progenitor cells has even reached clinical trials (e.g., NCT00474461, NCT00893360), but the reason behind their arguably positive effect on heart function is paracrine immunomodulation rather than generation of new CM [93].

Tissue engineering: another alternative currently being explored is the *in vitro* generation of myocardial tissue using iPSCs as a CM source. While these myocardial patches integrate with the resident myocardium and contribute to its contractive function [94, 95], concerns regarding their maturation, their ability to electrically synchronize with the host myocardium and the feasibility of scaling the production methods to meet the demands of an adult human heart must be addressed [69].

2.1.3 Limb regeneration in axolotl

Axolotl limb regeneration is a classically studied example of epimorphic regeneration, a healing process that involves the generation of a faithful copy of the lost structure through the formation of a blastema [13]. Limb regeneration in axolotl consists of 5 main phases [96] (Fig. 2.4):

1. Wound response and epidermal closure: within hours after wounding, coagulation stops the bleeding and the stump is covered by a thin layer of migrating keratinocytes coming from the adjacent epidermis [97]. Signaling from the severed nerves (e.g., *Fgf2*, *Fgf8* and *Bmp2* [96]) induces dedifferentiation of the epidermal keratinocytes, which transform into the wound epidermis (WE) [97].

2. Formation of the apical ectodermal cap (AEC): a couple of days after injury, the tip of the WE thickens and transforms into the AEC, a specialized signaling center that induces and maintains blastema cell proliferation.

3. Blastema formation: signals from the nerves and AEC dedifferentiate stump cells, which migrate beneath the WE to form a blastema [96]. The blastema is a mass of progenitor cells that will proliferate, pattern, and differentiate into the missing limb structures [96]. Blastema formation is a conserved mechanism required for the regeneration of complex structures, including planaria whole-body regeneration. Blastemas are independent units that contain enough positional information to give rise to fully patterned limbs even when transplanted to other body regions [97].

4. Proliferation and patterning: despite blastema initiation being substantially different to limb bud initiation, later blastema stages replicate developmental processes [14, 96]. Multiple cell types contribute to the blastema, which can be roughly classified into pattern-forming and pattern-following. Fibroblasts are pattern-forming cells that lead the growth of new limb structures and can give rise to their own cell type or others, e.g., cartilage. In contrast, pattern-following cells can only give rise to their own cell type e.g., muscle.

5. Differentiation, morphogenesis, and growth: proximal blastema cells start to differentiate and organize into complex structures, therefore growing the regenerating limb. Newly generated tissues integrate with mature structures and only structures more distal to original ones are generated. The new limb, a perfectly functional and structural replica of the original, is completely formed a few months after injury [96].

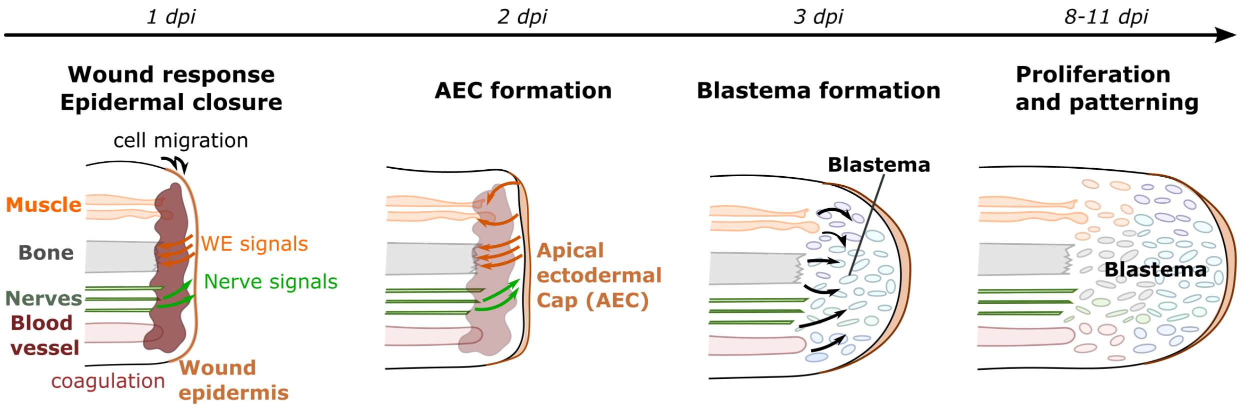


Figure 2.4: Blastema formation during axolotl limb regeneration. Within 24 hours after injury, coagulation stops bleeding and the stump is covered by a thin layer of epidermal cells that transforms into a specialized WE after the exposure to signals from the severed nerves. The WE thickens and transforms into the AEC, a specialized signaling center responsible of providing morphogenetic cues to support the dedifferentiation and migration of mature cells to form the blastema. The blastema is a mass of dedifferentiated progenitor cells that slowly gives rise to the lost limb structures through proliferation and patterning.

Wnt singaling during axolotl limb regeneration

Pathways involved in limb regeneration are shared with those used in development, highlighting the role of FGFs, BMPs, WNT, TGF-β and SHH (see [97] for a longer list). The WNT proteins are a large family of secreted glycoproteins with important functions in embryonic development, carcinogenesis and stem cell maintenance [98–100]. The WNT proteins form morphogenetic gradients tightly regulated by their diffusion and the presence of secreted inhibitory proteins, such as the DKK family, sFRP family and WIF [99]. Activation of Wnt signaling happens when extracellular WNT binds to the Frizzled receptors and activates one of the three possible intracellular signaling pathways (Fig. 2.5):

- Canonical Wnt signaling: WNT proteins bind the heterodimeric complex formed by a Frizzled receptor and the LRP5/6 co-receptor. This recruits AXIN and subsequently Dishevelled to the intracellular domain of LRP5/6. This stabilizes intracellular β-catenin levels through the inhibition of the β-catenin degrading complex [99]. Instead of undergoing phosphorylation and degradation through the proteasomal pathway, β-catenin remains unphosphorylated and travels to the nucleus, where it acts as a co-activator of the T-cell factor/Lymphoid enhancer factor (TCF/LEF) TFs [99]. Canonical Wnt signaling regulates cell fate decisions during early embryogenesis, neural patterning, stem cell renewal and cell proliferation [99].
- Planar cell polarity (PCP) pathway: The WNT-Frizzled binding event recruits Dishevelled to the cellular membrane and activates the RHO and RAC GTPases, leading to cytoskeletal reorganization through ROCK and JNK. PCP signaling plays an essential role in providing polarity information during gastrulation, limb development and hair follicle orientation [99].

- Wnt-Ca2+ signaling: The WNT-Frizzled complex recruits G-proteins and Dishevelled to the cellular membrane, leading to Ca2+ release through protein kinase C and calcium/calmodulin-dependent kinase II. The Wnt-Ca2+ signaling pathway regulates gastrulation, dorso-ventral axis formation, and heart development, and influences canonical Wnt and PCP [99].

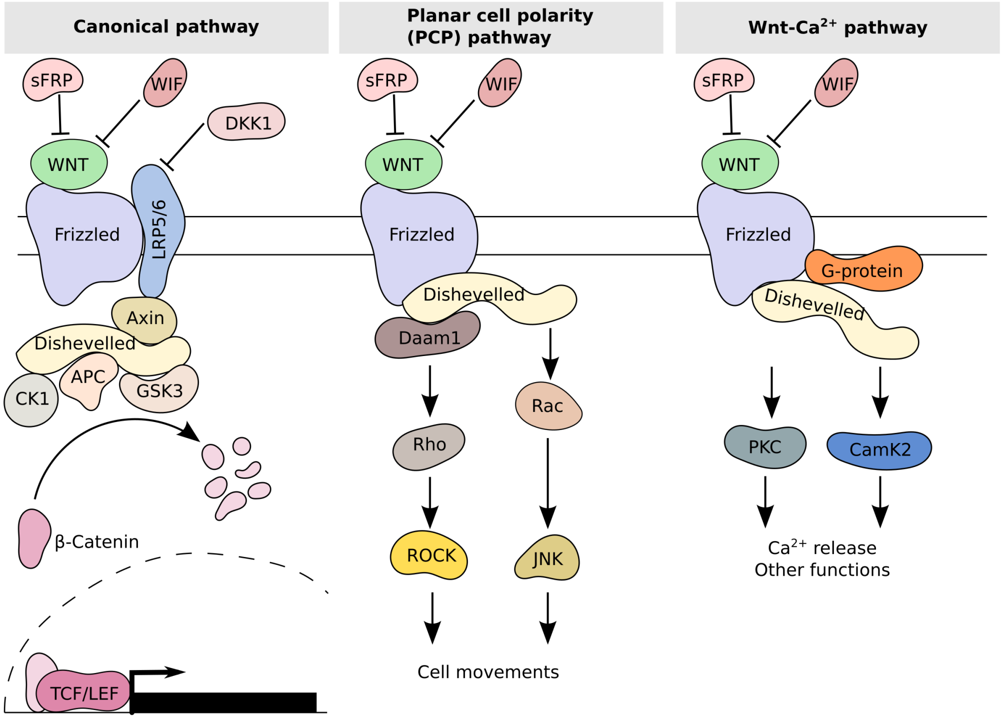


Figure 2.5: Schematic representation of Wnt signaling pathways.

Despite this categorization, the Wnt signaling pathway is very complex due to the high number of genes that participate in the pathway. The human and axolotl genomes contain 19 and 21 *Wnt* genes, respectively [99]. These can activate 10 Frizzled receptors that act through 3 Dishevelled isoforms. How the different WNT-Frizzled combinations are channeled through the different signaling pathway remains unknown, although some WNT isoforms have been proposed to have pathway preferences: WNT3 has been associated with both canonical and PCP signaling, while WNT5A predominantly activates PCP signaling [101].

Wnt signaling has been the focus of previous studies in the context of axolotl limb development and regeneration. Ghosh *et al.* have reported expression of *Wnt3a*, *Wnt5a*, *Wnt5b* and *Wnt7* in axolotl early limb buds and blastema, despite their absence in mature limbs [50]. *Wnt5a/b* is predominantly expressed in the distal blastema region and can be detected as early as 3 days post-injury (dpi), time in which the stump cells are dedifferentiating. Contrary to *Wnt3*, which is mostly provided by the AEC, *Wnt5a/b* are almost uniquely expressed by the dedifferentiated mesenchyme [50]. Reducing Wnt signaling though injection of *Axin1* and *Dkk1* leads to AEC formation defects and abrupted regeneration [102]. Regarding the type of Wnt signaling, the canonical and PCP pathways have been implicated with important functions during limb development and regeneration: canonical

Wnt signaling and *Lef1* activity are required for successful zebrafish fin regeneration [103]; and inhibition of Wnt signaling via the WNT secretion inhibitor C59 disrupts blastema outgrowth and the expression of patterning genes, reflecting essential PCP activity [104].

2.2 Gene regulatory network inference

Biological systems are characterized by their emergent properties, resulting from the abundance and interactions between single components. Network representations are useful models to understand and predict the consequences of these interactions. Even though a comprehensive network model of the cell requires different layers that are interconnected to reflect epigenetic, transcriptional, post-transcriptional and metabolic regulation, GRNs are arguably the underlying cause behind cellular phenotypes [16, 21, 105, 106]. GRNs are representations of the interactions through which TFs regulate the expression levels of target genes [105]. Biologically, TFs regulate the expression of target genes by binding to conserved short DNA sequences, called motifs, located in the cis regulatory regions of the target gene [105]. Binding of a TF to a cis regulatory region can have multiple outcomes, depending not only on the main mode of regulation (MOR) (activation or inhibition) characteristic of the TF, but also on the context, which includes the presence of co-regulators and the chromatin state. TFs are key regulators for the cellular phenotype and represent 5-10% of the protein coding genes in *E. coli* (7%), *S. cerevisiae* (5%), *Drosophila* (5%) and humans (8%) [105]. In fact, the human genome is estimated to contain 1,600 genes encoding TFs, classified into 30 families according to their binding motif [105].

Inferring the GRNs behind physiological and pathological processes has been a long-lasting goal of molecular biology and it remains a challenge [105]. GRN inference is the process of reverse engineering the transcriptional regulatory relationships from observed gene expression data, aiming to obtain a GRN representation that approximates the real, underlying GRN or is a useful simplification of it [107] (Fig. 2.6). Mathematically, GRN models are represented as graphs in which nodes are genes and edges represent regulatory relationships between them. GRN models are by nature directed, and the edges might have weights that reflect the strength of the interactions.

2.2.1 Transcriptomic profiling for GRN inference

GRN inference relies on transcriptome profiling, nowadays performed through bulk RNA-seq, scRNA-seq or snRNA-seq. Bulk RNA-seq profiles the transcriptome of pooled cells through next generation sequencing technologies. It provides high-quality measurements but is restricted to an average quantification of the gene expression in the sample. sc/snRNA-seq measure the transcriptome of thousands of individual cells or nuclei at once, isolated through the use of microfluidic (e.g., 10xGenomics) or multi-well (e.g., SMART-seq2) technologies. Despite their increased resolution, these technologies detect less genes per sample, are noisier, and suffer from dropouts manifested as a zero-inflation of the gene expression matrix (Fig. 2.7). Recently, multi-omic profiling of chromatin

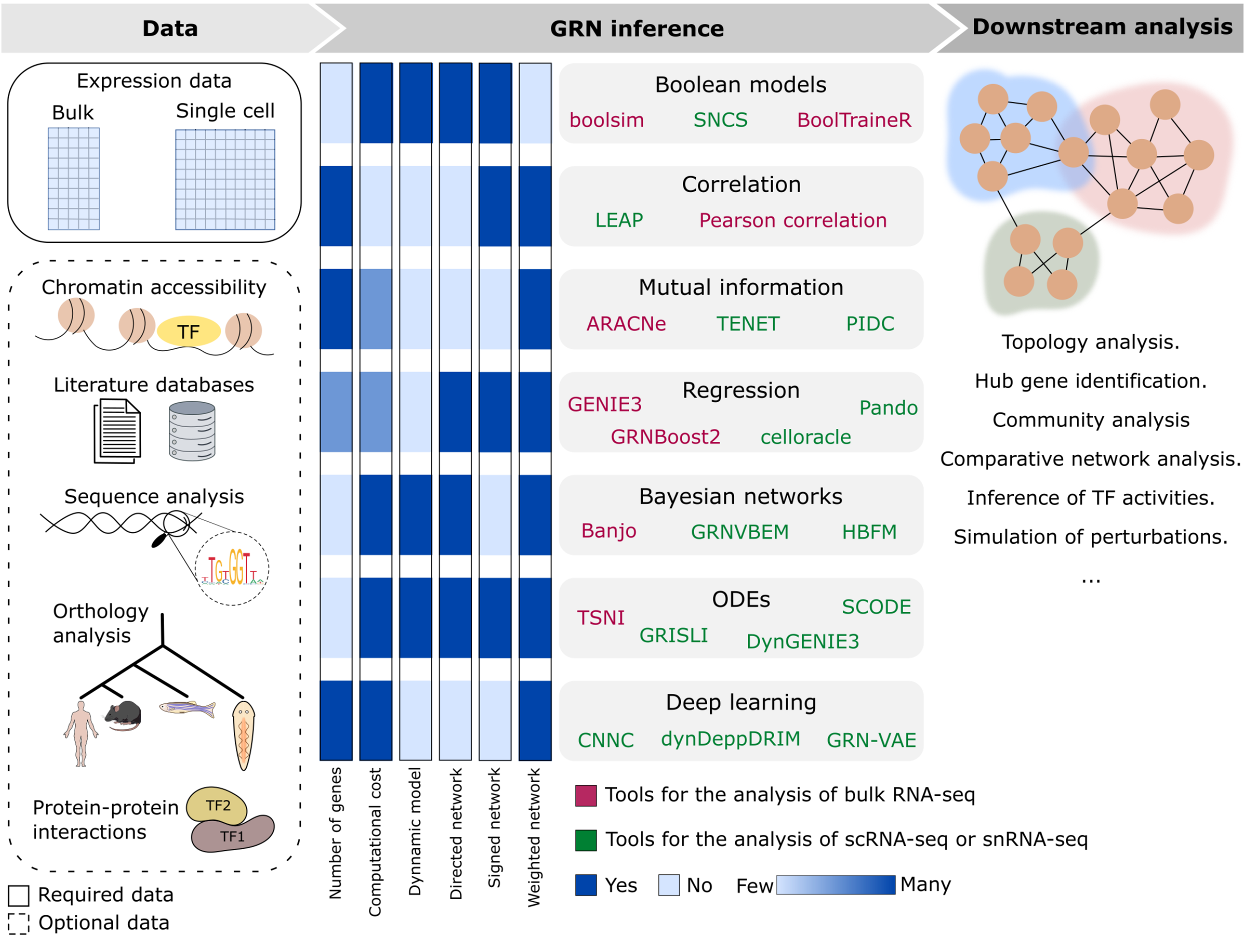


Figure 2.6: Main steps in the GRN inference process. A non-comprehensive list of representative tools for each algorithmic solution is included.

accessibility (ATAC-seq) and gene expression (RNA-seq) in the same nuclei has emerged as a useful tool for GRN inference.

Regardless of the profiling modality, transcriptomic data is represented as a matrix in which the expression of N genes is measured in M samples, cells, or nuclei, being x_{ij} the normalized expression of gene i in sample/cell j. In bulk RNA-seq, the so-called "skinny" gene expression matrix has a rectangular shape, containing information about 20,000 genes in just a few samples. Single-cell and single-nuclei experiments make this “skinny matrix” squared, as they contain gene expression measurements in thousands of cells. In both cases, dealing with these large and complex data sets requires efficient algorithmic solutions.

Algorithmic approaches for GRN inference

GRN inference can solely rely on the gene expression data or leverage a network scaffold, either from other data-types (e.g., ATAC-seq) or from previous knowledge (PKN, prior knowledge network).

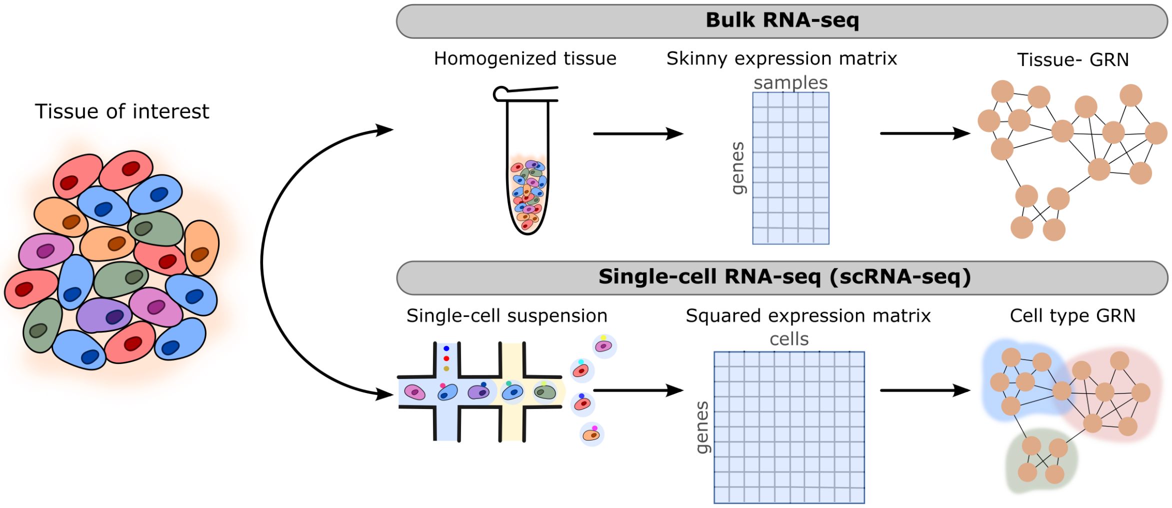


Figure 2.7: Differences between bulk and single-cell RNA-seq for GRN inference.

Using a network scaffold helps the inference process by reducing the number of relationships to model, decreasing the false positive rate and speeding up computation [108]. The most common sources of GRN scaffolds are:

- Literature databases: PKN scaffolds are constructed through manual or text mining curation of existing literature in which the relationships between genes has been experimentally evaluated. They provide a reliable starting point for GRN inference [109] and allow to perform footprint analysis, in which the activity of key TFs is estimated from the changes of gene expression of their targets [110]. Ingenuity Knowledge Base or CollecTRI are high-quality databases of transcriptional regulation derived from the literature [111, 112].
- Chromatin accessibility: ATAC-seq profiles are obtained through DNA tagmentation with the Tn5 hyperactive transposase and give useful condition-specific insights into open chromatin regions to which TF might bind to regulate gene expression. By purifying DNA – protein complexes, ChIP-seq adds a refinement on the chromatin accessibility scaffold as it detects which regulatory regions are being bound by a TF of interest.
- Sequence analysis: TF binding motifs can be identified through analysis of the genomic DNA sequence. Usually, regulatory regions located at a certain distance from the gene promoter are scanned to find the presence of known TF binding motifs. Comprehensive TF – motif collections are stored as position weight matrices or hidden Markov models in databases such as JASPAR or TRANSFAC [113–115]. While the presence of a TF motif alone is not a reliable indicator of a regulatory relationship, sequence analysis is an efficient way to obtain a network scaffold that can be contextualized using the transcriptome data. Tools such as iRegulon [116], i-cisTarget [117] and HOMER [118] can identify TF motifs in genomic regions.

- Orthology analysis: a GRN inferred for one species can be translated for another. While this approach might miss important regulatory interactions or introduce false positives, it could be informative if the two species are closely related and there is a good ortholog translation between them [105].

- Protein-protein interaction networks: protein-protein interaction networks might provide useful information of TF complexes necessary for their activity and might complement to a network scaffold obtained through other approaches [105].

Regarding the mathematical approximation to the GRN inference problem, different strategies can be used based on the question to answer and on the trade-off between the number of nodes/genes to model and the number of samples available [105]. Here we review the most frequently used modeling strategies and give examples of algorithms that apply them:

- Boolean networks: Boolean networks model TF activation as a dichotomous variable through binarization of its expression profile [119]. The activity status of each TF is defined by an activation function that describes the logical interactions between network components [120]. Boolean networks are useful as descriptive and predictive models of static and dynamic systems and do not contain a large number of parameters. If the network topology is known, Boolean models can be used to find stable network states, also called attractors, that associate with cellular phenotypes [18]. If the network topology is not clear and enough data with temporal resolution is available, Boolean models can decipher the logical combination of TF that explains the transition between transcriptomic states [121]. In these cases, the probabilistic implementation of Boolean models proofs useful, as it provides insights into the most likely activation function to govern the expression of each gene. Despite binarization making these models more robust to dropouts and reducing the number of model parameters, it impedes modeling dose-response relationships and reduces the accuracy of the data [119]. Another drawback of Boolean models is their large computational cost, as they rely on brute force exploration of all possible network states [105]. Therefore, Boolean models can only be fit after a small number of relevant TFs has been preselected. Additionally, Boolean networks fail to appropriately model negative feedback loops affecting a single TF and can only answer qualitative questions [105]. Several algorithms to fit GRN Boolean models have been developed, including boolsim [18] and BoolTraineR [122] for bulk RNA-seq and SCNS [123], LogicNet [124] and BooleanPseudotime [125] for scRNA-seq data.

- Correlation: Pearson or Spearman correlation is a simple and useful way to construct co-expression networks, even when the number of genes is very large. Correlation-based algorithms usually calculate a fully connected network that is filtered to keep edges that pass a correlation or significance threshold [16]. They are frequently used to construct whole transcriptome GRNs in which master regulators and network communities can be identified. The drawbacks of this approach are that it does not return directed networks, unless time-lagged correlation in time series experiments is used [126], and that a large number of false positive interactions is found. WGCNA is the most popular correlation-based algorithm, usually applied on bulk RNA-seq data to find co-expressed gene modules with a putative functional relationship [127]. PPCOR [128] and LEAP [129] are other examples of correlation-based algorithms optimized to calculate partial correlations and deal with scRNA-seq data, in which the abundance of dropouts complicates the calculation of correlation coefficients [130].

- Mutual information: mutual information measures the dependency between two variables based on their Shannon entropy: $M_{ij} = H_i + H_j - H_{ij}$, where i and j represent two genes and H is calculated as $H_i = \sum p(x_i) * log(p(x_i))$. Similarly to correlation, mutual information can infer undirected whole-transcriptome networks that reveal master regulators [119]. Recent benchmarks show that mutual information usually outperforms correlation but implies an increased computational complexity [131]. ARACNe is the most popular algorithm for mutual information-based GRN inference using bulk RNA-seq [132]. scRNA-seq has motivated the development of algorithms that modify the definition of mutual information to incorporate pseudotime information (e.g., TENET [133], Scribe [134]).

- Regression: linear or non-linear regression can be used to model the expression of each gene as a function of the level of expression of its regulators. Regression-based approaches can infer transcriptome-wide GRNs or be applied on a restricted set of genes and interactions contained in a network scaffold. Linear regression models as implemented by Pando [22] or celloracle [21] are specifically designed for scRNA-seq data and can efficiently simulate network perturbations. Other tools use regression models based on random forests and are applicable to both bulk and single-cell data sets (e.g., GENIE3 [19], GRNBoost2 [135] and SCENIC [20]).

- Bayesian networks: they represent GRNs as directed acyclic graphs in which each node is a random variable and model the conditional dependences between genes by estimating the joint probability distribution of all interactions [105]. They provide highly-detailed and interpretable models, as they output the posterior distribution for each edge. Additionally, their Bayesian nature makes them a convenient platform for the integration of multiple data modalities. Apart from their computational complexity, their main drawback is that they cannot model feedback loops, which limits their applicability [16]. Dynamic Bayesian networks overcome this limitation by modeling the GRN as a sequence of network states, with an increase in the computational cost of the inference problem [119]. Banjo is a Bayesian algorithm for the inference of GRNs using bulk RNA-seq [136]. GRNVBEM and HBFM [137] are Bayesian-based algorithms specifically designed for scRNA-seq data. Additionally, dsReg uses a bayesian approach to model the regulation of alternative splicing, adding an additional layer of complexity to the GRN inference problem [138].

- Ordinary differential equations (ODEs): ODEs model the rate of change in the expression of a gene with respect to the changes in the expression of its regulators. While they can be used with static or dynamic data, ODE models are most suitable to model dynamic systems and can therefore deeply benefit from pseudotime trajectories calculated on scRNA-seq data [119].

ODEs can provide more realistic and causal GRN representations but must be limited to a small number of genes due to the high number of parameters that must be fit. Algorithms such as SCODE [139], GRISLI [140] and DynGENIE3 [141] apply ODEs for GRN inference using scRNA-seq data, while TSNI [142] was specifically developed for bulk RNA-seq.

- Deep learning: algorithms leveraging deep learning architectures have recently entered the GRN inference arena and have been mostly applied to scRNA-seq data due to the large sample sizes required for training [143]. CNNC leverages convolutional neural networks to predict regulatory relationships between genes from images of 2-dimensional gene expression histograms [144]. DeepDRIM extends this framework to reduce the number of false positives resulting from indirect regulation [145], and dynDeepDRIM expands this algorithm to include pseudotime information [146]. More recently, tools based on deep generative models have been developed [24]. DeepSEM and GRN-VAE use variational autoencoders with a GRN layer to identify regulatory relationships between genes [147, 148]. Lastly, scGPT is a foundational model trained to learn gene and cell embeddings [149].

The large diversity of GRN inference algorithms makes selecting the most appropriate tool a daunting task. The decision should mostly be based on the biological question and the available data, taking into account that there is an inverse relationship between the granularity of the model and the data and computational power required to fit it [120]. In theory, algorithms developed for bulk RNA-seq should work on scRNA-seq, although the higher data sparsity and noise limits their performance and asks for specialized solutions [105]. Additionally, the availability of a network scaffold determines whether a small number of candidate genes should be modeled in detail, or whether a noisy whole-transcriptome GRN should be inferred with exploratory purposes. It is difficult to accurately compare algorithm performance due to the lack of curated real GRNs, and while synthetic data might be useful for computational power comparisons, a recent benchmark showed that better performance on synthetic data does not reflect better performance on real data [131]. Additionally, Pratapa *et al.* showed that methods that do not require pseudotime information usually outperform those which do, probably due to errors in the trajectory inference process. Regarding accuracy, regression algorithms in which the number of relationships is restricted to a high-quality network scaffold are the best performers [131].

2.2.2 Downstream analysis: GRN topology and emergent properties

Once a GRN model has been fitted, it can be exploited for different downstream analyses:

- Network topology analysis and identification of hub nodes: GRN models can contain thousands of nodes and edges, which calls for a birds-eye-view analysis of the GRN properties. The degree distribution of biological GRNs follows a power law, indicating that while most genes have few connections, a small number of highly connected hubs maintain the network architecture. Additionally, GRNs are highly hierarchical [150] and hub genes can be identified using a variety of centrality metrics [106, 151]. Here we define the two centrality metrics used throughout this thesis:

  - Out-degree: quantifies the number of out-going edges that originate from a given node in the network. Nodes with a high out-degree are hubs because they directly influence a large number of nodes in the network.
  - Betweenness centrality: quantifies the number of shortest paths in a network that go through a given node of it. Eq. (2.1) shows how betweenness centrality is calculated, where v, s and t refer to nodes in the network and d represents the number of shortest paths between two nodes. Nodes with high betweenness are hubs because they play a key role in the flow of information through the network.

$$Betwenness(v) = \sum_{s \neq v \neq t} \frac{d_{s,t}(v)}{d_{s,t}} \tag{2.1}$$

- Identification of functional modules: the hierarchy in GRNs causes the appearance of network communities. Communities are groups of genes that are highly connected to each other, but minimally connected to other regions of the network. Algorithms such as fast greedy clustering or Infomap approach the community identification problem using hierarchical clustering or random walkers to maximize the network modularity [152, 153]. After annotation using enrichment analysis, communities can give insights into the most important functions controlled by a GRN.
- Comparative analysis: GRNs inferred for two conditions can be compared through direct pairwise comparison of TF – target relationships. In spite of its simplicity, direct network comparison often fails to provide informative results due to the noisy and sparse nature of GRNs [106]. Ranking-based approaches to compare the hub TFs in each GRN are more robust to identify candidate genes that can explain the biological differences between the two conditions.
- Inference of TF activities: because the expression of a TF is not necessarily a good readout of its activity, enrichment tools applied on the GRN-derived set of targets for each TF can be used to estimate TF activity [106].
- Prediction of perturbation effects: GRN models can be used to predict how changes in activity of a TF impact the rest of the network. Depending on the modeling strategy, these analyses provide more or less granular information about the consequences of the perturbation. For example, celloracle in combination with pseudotime trajectory information has been used to predict how the perturbation of a TF affects the natural course of differentiation [21].
- Understanding the molecular mechanisms that regulate a biological process: small-scale but high-quality GRNs can be used to gain quantitative insights into gene regulation [107]. As

an example, ODE models have shed light into how short and long-term IFN-β exposure can differentially induce the expression of innate immunity of pro-inflammatory gene sets [154].

Lastly, experimental validation of the inferred GRNs and their properties is important. Experimental validations might aim to interrogate whether TF – target regulation occurs through direct or indirect molecular mechanisms [107], or whether the predicted effect of perturbing a TF matches experimental observations. By reiteratively inferring GRN models and validating the most interesting properties for the biological question at hand, the models can be improved to better represent reality [155].

2.3 Modeling GRNs to understand and promote regeneration

Previous studies have inferred the GRNs behind heart regeneration, at least to some extent. Providing single-cell granularity, Wang *et al.* and Nomura *et al.* combined scRNA-seq with scATAC-seq and H3K27ac ChIP-seq to identify TF with differentially accessible motifs between control and pathologic conditions (AMI and HF, respectively) [156, 157]. Kuppe and colleages generated the most comprehensive atlas of human AMI to date, integrating snRNAseq, snATACseq and spatial transcriptomics of control and hypoxic, fibrotic, border zone and remote regions of infarcted hearts. Among others, this data was used to infer CM and fibroblast GRNs. The authors leveraged the single-nuclei data to find putative regulators based on their accesible motifs and the Pearson correlation between the accessibility of regulatory regions and the expression level of target genes. This analysis provided interesting insights into the hub TFs regulating CM response to AMI [158]. However, none of the aforementioned studies provides mechanistic GRN models that are amenable to the simulation of perturbations. Bulk RNA-seq data has also been employed for GRN inference of heart regeneration. Meta-analyses that integrate bulk RNA-seq from different studies have been the most popular approach, as they partially ameliorate the data dimensionality problem. This approach has been used to identify TFs driving zebrafish heart regeneration [159], hiPSC differentiation towards CMs [127], congenital heart disease [160] and CM maturation [161]. Lastly, an approach based on the construction of network models purely from literature has modeled the CM cell cycle regulation pathway [162].

In the case of axolotl limb regeneration, several studies have identified hub TFs involved in the regenerative response. Gerber *et al.* provided important insights into the molecular rewiring of connective tissue (CT) cells necessary for blastema formation, but did neither infer GRNs nor master regulators of this process [14]. Lin *et al.* compared axolotl and frog CT, finding that frog CT cells are unable to re-express the embryonic gene expression program necessary for blastema formation, but did not interrogate which TFs drive this process [163]. Lastly, Ye *et al.* compared the integrated GRN of neotenic and metamorphosed axolotl using scRNA-seq of over 15 tissues, but did not focus on limbs [164].

Altogether, the majority of previous research has used a correlation-based approach to identify master regulators of regenerative or fibrotic responses, but have failed to provide causal and predictive models for their *in silico* perturbation. An additional limitation specific to the heart studies is the fact that single-nuclei profiling does not distinguish between diploid and polyploid nuclei or cells, failing to identify the mechanism through which mammalian CM polyploidization influences the regenerative avility. To address these limitations, we focus on regenerative models, namely the axolotl and the P1 mouse heart, to model the GRNs behind myocardial and bone regeneration. Using a wide range of methodologies and data, we infer the GRNs governing myocardial regeneration in vertebrates and give special attention to the identification of molecular drivers of mammalian CM polyploidization. Additionally, we propose new therapeutic hubs to activate endogenous regeneration of the adult mammalian myocardium. Lastly, we provide the first single-cell atlas of axolotl CSD and compare the GRNs in regenerative and fibrotic axolotl bone injuries, identifying candidate molecular mechanisms to transform the fibrotic response into a regenerative response.

3 | *Hey2* as a roadblock for vertebrate heart regeneration

The capacity for heart regeneration is variable among vertebrates: while the axolotl (*Ambystoma mexicanum*) and zebrafish (*Danio Rerio*) can regenerate their hearts throughout life, mammals can only achieve heart regeneration until the first postnatal days of life. Coincidental with this loss in regenerative capacity is the polyploidization of mammalian CMs that hinders their proliferation. This chapter explores these two processes through the inference of the GRN controlling CM identity and injury response in axolotl and the GRN controlling CM polyploidization in mouse. Both GRNs, inferred using different data and methodologies, identified *Hey2* as a common key regulator. *In silico* simulations and preliminary *in vivo* experimentation support the requirement of *Hey2* inactivation to induce axolotl border zone CM proliferation upon injury. Additionally, preliminary in vivo tracing of the *Hey2*$^{+}$ lineage in mouse shows a positive correlation between CM ploidy and *Hey2* activity. Altogether, these results suggest a novel role for *Hey2* as a regulator of mammalian CM ploidy, provide the first genetic markers to facilitate the study of the mouse MNDCMs and support the role of *Hey2* as a roadblock for heart regeneration.

3.1 Materials and methods

3.1.1 Methods related to the axolotl data set

Experimental procedures

All experimental procedures on axolotl were performed by Elad Bassat, PhD and are briefly described here to provide the necessary background for the *in silico* analysis of the gathered data. The scripts used for GRN inference will be available at `https://github.com/iriverog`.

Axolotl husbandry: 8-10 cm nose-to-tail juvenile axolotls were used in this project. The sex of the animals was unknown, as they were not genotyped, and sexual characteristics are not present at this stage. Animals were bred and maintained in the animal facilities of the Institute of Molecular Pathology (IMP) in Vienna, Austria, as described previously [165]. Animals were kept individually, and experiments were conducted with approval from the Magistrate of Vienna (GZ:MA 58-105623-2020-26).

Axolotl line tgScel(Myl7:TFP-NLS-T2a-HEY2): the axolotl line tgScel(Myl7:TFP-NLS-T2a-HEY2), abbreviated MYL7:TFP-HEY2_OE, overexpresses the axolotl *Hey2* coding sequence under the control of the CM-specific promoter *myl7* of *Xenopus laevis*. The TFP reporter has a nuclear localization signal and can be used as a tracker of exogenous *Hey2* expression. In this study, F0 mosaic animals with variable numbers of *Hey2*$^{+}$ cells were used.

Cryoinjury: cryoinjury was used as a model of ischemic myocardial damage. Briefly, axolotls were anesthetized using 0.015% benzocaine. Anesthetized animals were placed in a supine position on a wet paper towel during the surgical procedure. After skin disinfection, a lateral thoracotomy was performed, and the chest cartilage plates were retracted to achieve direct access to the heart pericardium. A small incision on the pericardium exposed the ventricle and then a liquid nitrogen chilled probe was placed on the heart apex for 5 seconds and then flushed with PBS containing antibiotics. A white area remained where the probe had been placed, serving as a visual confirmation of the injury. Afterwards, animals were closed and allowed to recover on a wet paper towel for 10-20 minutes. After surgery, animals were housed in individual tanks with clean tap water and observed daily. For the first 72 hours, animals were treated with 0.5mg/L butorphanol, 50 U/mL penicillin, 50µg/mL streptomycin and 20% holftrors solition. Between 72h and 1 week, the animals received only antibiotics and 20% holftrors solition. Animals were sacrificed by benzocaine overdose 1, 4, 7, and 14 dpi. Hearts from uninjured and healthy animals were included as a physiological control (day 0) (Fig. 3.1).

Quantification of proliferation in the border zone: To assess the proliferation levels of *Hey2*-overexpressing vs wild-type (WT) cells, a single 5-Ethynyl-2'-deoxyuridine (EdU) injection at a concentration of 400 µM and at a dosage of 20 µL/g was given i.p. and animals were left to recover overnight (approximately 16h). Animals were sacrificed and their hearts were harvested for analysis.

Tissue dissection and cell isolation: immediately after euthanasia, hearts were harvested for multi-omics (snRNA-seq and snATAC-seq on the same cell). Hearts were flash frozen in liquid nitrogen and moved to -80°C until the day of extraction. On the day of processing, hearts were allowed to equilibrate on wet ice in the 4°C cold room. 10 mL of nuclei pure lysis buffer (Sigma Aldrich) was supplemented with 1 complete Mini Protease Inhibitor Cocktail (Roche). 2mL of this lysis buffer were supplemented with 2 µL DTT and 20 µL 10% Triton X100 (working lysis buffer). The working lysis buffer was placed in a 2mL dounce homogenizer. 6 hearts per timepoint were finely chopped and added to the lysis buffer. Using Pastel A, the tissue was dounced until resistance was gone, then it was filtered with a 40 µm mesh into a new 15 mL tube and then resuspended in 1mL Nuclei Pure storage buffer (Sigma Aldrich) with 20 µL RNase-OUT (Thermo-Fischer). Nuclei were centrifugated again and most of the liquid was removed, leaving approximately 40 µL which were used for analysis.

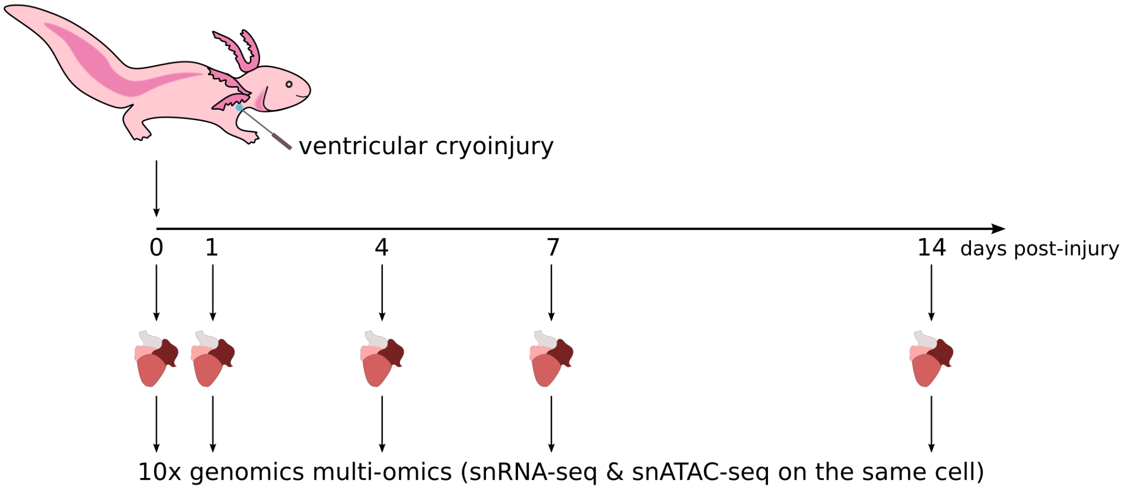


Figure 3.1: Experimental design. Juvenile axolotls were subjected to ventricular cryoinjury. The hearts were collected at 0 (non-injured controls), 1, 4, 7, and 14 dpi and samples were processed for 10xGenomics multi-omics (simultaneously profiles snRNA-seq and snATAC-seq).

Multi-omics sequencing

Simultaneous snRNA-seq and snATAC-seq on the same cell (multi-omics) was performed in a single batch using the 10xGenomics platform in collaboration with the NGS Vienna Biocenter Campus facility. Libraries were prepared and sequenced on a NovaSeqS4 sequencing platform following manufacturer recommendations. For snRNA-seq and Visium data, the raw fastq files containing reads for each library were processed using the same pipelines as in Lust *et al.* [166]. Briefly, the raw reads were mapped and quantified on the latest axolotl genome release (v6.0) and transcriptome annotation (v47) by running kallisto (v0.46.0) [167] and bustools (v0.40.0) [168] with default parameters. Both reads from introns and exons were included in this step. The snATAC-seq data was aligned and quantified using CellRanger-ARC (v2.0.0, 10xGenomics), also as previously described [166]. These pre-processing steps were performed by Jingkui Wang, PhD and the NGS Vienna Biocenter Campus facility.

snRNA-seq and snATAC-seq analysis

These steps were performed by Jingkui Wang, PhD and Elad Bassat, PhD. Multi-omics pre-processing and clustering was done using the R packages Seurat v4.3.0 [169] and Signac v1.9.0 [170]. For snRNA-seq data, cells with (1) less than 200 or more than 10,000 detected features and (2) more than 60% of mitochondrial reads were filtered out. After log-normalization, principal component analysis (PCA) was computed using the 8,000 most variable genes. To visualize the data, dimensionality reduction by uniform manifold approximation and projection (UMAP) was calculated using 30 principal components (PCs). Marker genes of each Louvain cluster were calculated with the Seurat function FindAllMarkers() using Wilcoxon Rank Sum test. Clusters were annotated to their corresponding broad cell types using canonical marker genes (Table 3.1).

Table 3.1: Marker genes used for broad cell type annotation. Marker genes were described in [164, 171, 172].

Cell type	Marker genes
B cell	*Abcc4, Bank1, Bcl2l11, Blnk*
Cardiomyocyte	*Cacna1g, Myh6, Myh7, Myoz2, Nppa, Ryr2, Tnnt2, Ttn*
Endothelial cell	*Cdh5, Clu, Eng, Kdr, Pecam1, Tagln2, Vwf*
Erythrocyte	*Alas2, Hbg1, Hbg2*
Fibroblast	*Col1a1, Col3a1, Dcn, Krt19, Pdgfra, Postn*
Megakaryocytes	*Fli1, Gp9, Itga2b, Mpl, Plec, Selp, Srgn, Tubb3, Zfpm1*
Monocyte / Macrophage	*Ccr5, Cltc, Ctsb, Lyve1*
Neuron	*Cacna1a, Cacna1e, Cacna2d1, Cacnb1, Cacng3, Kcnj12, Kcnj13, Snap25*
Neutrophil	*Adgre1, Arg1, Csf3r, Itgam, Spi1, Itgam*
T cell	*Bach2, Ccr7, Il7r, Stat4*

Prior to the snATAC-seq analysis, all peaks called by CellRanger-ARC for each time point were combined and quantified only in cells found in the snRNA-seq. Cells with less than 200 or more than 1,000,000 features, and cells with nucleosome_signal > 6 or TSS_enrichment < 1 were discarded. Because all samples were prepared and sequenced in the same batch, the two modalities were combined without additional integration. The cell type annotations of snRNA-seq were transferred to snATAC-seq. The UMAP embedding of snRNA-seq was used for data visualization.

Ventricular CMs were computationally isolated based on the cell type assignments described above and the deconvolution of a spatial transcriptomics dataset of axolotl heart regeneration at 4 dpi.

Briefly, axolotl heart tissue was quickly washed in PBS, patted dry and placed in Optimal Cutting Temperature (OCT) (Takara) solution in a cryomold. The cryomold was frozen in isopentane chilled in liquid nitrogen until OCT became opaque. The blocks were cryosectioned and placed on histology slides for visualization and evaluation of the damage area. When the region of interest containing large damage localized to the apex was identified, a single section was placed on the Visium slide for further processing according to manufacturer's protocol. The Visium count matrix was analyzed in the same way as the snRNA-seq, except for SCT normalization of gene expression. RCTD was used for deconvolution of the spot gene expression into cell type proportions [173]. RCTD was run with default parameters using doublet mode. Regarding CM trajectory inference, SCANPY v1.91 was used [174].

Inference of a large-scale GRN for ventricular CMs

A large-scale GRN for ventricular CMs was reconstructed using Pando, which integrates single-nuclei multi-omic measurements with the genomic localization of TF binding motifs [22]. The Pando framework consists of three steps (Fig. 3.2):

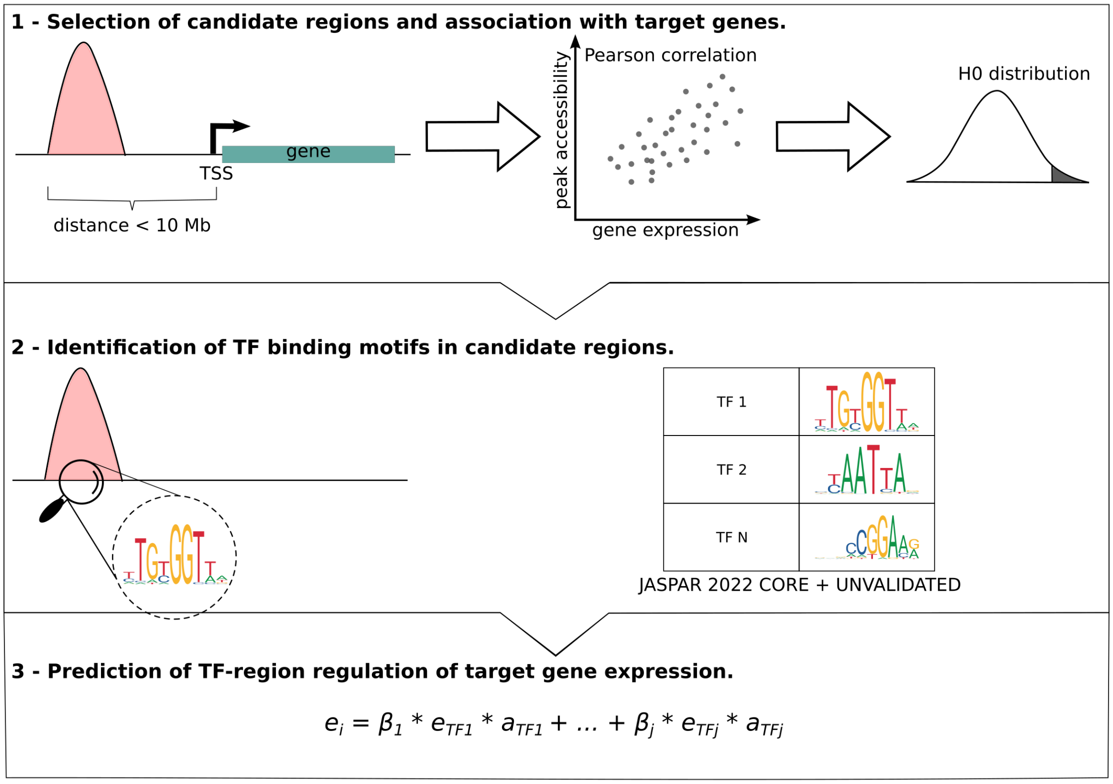


Figure 3.2: GRN inference with Pando. The Pando workflow consists of three steps: (1) selection of candidate regions and association with target genes, (2) identification of TF binding motifs in candidate regions and (3) prediction of TF-region regulation of target gene expression. In the equation i represents the target gene and j are its transcriptional regulators. e stands for expression level, a stands for accessibility and β are the coefficients.

1. Selection of candidate regions and association with target genes: candidate regions refer to regions of accessible chromatin to which TFs can bind to regulate the expression of target genes. Candidate regions were identified using the Signac function LinkedPeaks(), which calculates the Pearson correlation between the expression of a gene and the chromatin accessibility of all peaks within a given distance from its transcription start site (TSS) [170, 175]. This maximum distance was set to 10 Mb, as axolotl topologically associating domains (TADs) are on average 7 to 10 times larger than their human counterparts [176, 177] and all gene regulatory interactions are expected to happen within TADs [178]. Correlations were only calculated for the 10,000 most variable genes and for pairs of peaks and genes detected in at least 10 cells. Absolute correlation coefficients larger than 0.05 were retained and transformed into z-scores for which empirical p-values were calculated using a null distribution constructed by sampling 200 random peaks. Peak-gene interactions with $p\text{-values} < 0.05$ were retained for the next step.

2. Identification of TF binding motifs in candidate regions: candidate regions were scanned for TF binding motifs to predict causal TF – target gene regulatory relationships. This step was performed using the Pando find_motifs() function [22]. TF binding motif information was obtained from the JASPAR CORE and UNVALIDATED collections (2022 release) [113, 115]. The latest axolotl genome sequence release (amexG_v6.0) was used as a reference [179].

3. Prediction of TF-region regulation of target gene expression: regulatory relationships between TFs and target genes were inferred using regression models with region-TF as independent variables and the expression of the target gene as the dependent variable. Generalized linear models as implemented by the Pando function infer_grn() were used for GRN reconstruction [22]. These linear models predict the log-normalized expression of target gene i (e_i) by combining the log-normalized expression of a TF and the accessibility of its associated region (e_j and a_j for each TF j, respectively) (Eq. 3.1). The fitted coefficients (β_j) are a measure of the interaction strength between the region-TF pair and the target gene.

$$e_i = \sum_j \beta_j e_j a_j + \epsilon \tag{3.1}$$

The resulting network contains directed, weighted, and signed edges, where the sign of the coefficient indicates the MOR (activation or inhibition). Edges with an FDR-adjusted $p\text{-value} < 0.05$ were retained. Network edges were collapsed at the gene level to obtain a TF – gene GRN.

Network topology analysis was performed using Cytoscape v3.9.0 Network Analyzer [180, 181]. Network communities (locally dense connected subgraphs in the network [182]) were detected using fast greedy clustering as implemented in the Cytoscape clusterMaker plug-in [152]. This algorithm hypothesizes that the optimal community structure in a network is that with the maximum modularity (Eq. 3.2):

$$M = \sum_{c=1}^{n_c} \left[\frac{L_c}{L} - \left(\frac{k_c}{2L} \right)^2 \right] \tag{3.2}$$

where L is the total number of edges in the network, L_c is the number of edges in community c, k_c is the total degree of nodes in this community and n_c is the total number of communities in the network [182]. The fast greedy algorithm starts by assigning each node in the network to its own community. Then, it calculates the change in modularity obtained by merging two nodes into the same community and performs the merge that achieves the highest network modularity. These steps are repeated until the whole network is assigned into a single community. As the modularity was recorded for each merge, the optimal network partition with maximal modularity is selected (Fig. 3.3). Network communities were annotated with biological functions by performing an over-representation analysis (ORA) of the community genes on the GO Biological Process database [183, 184], as implemented by the fgsea R package [185].

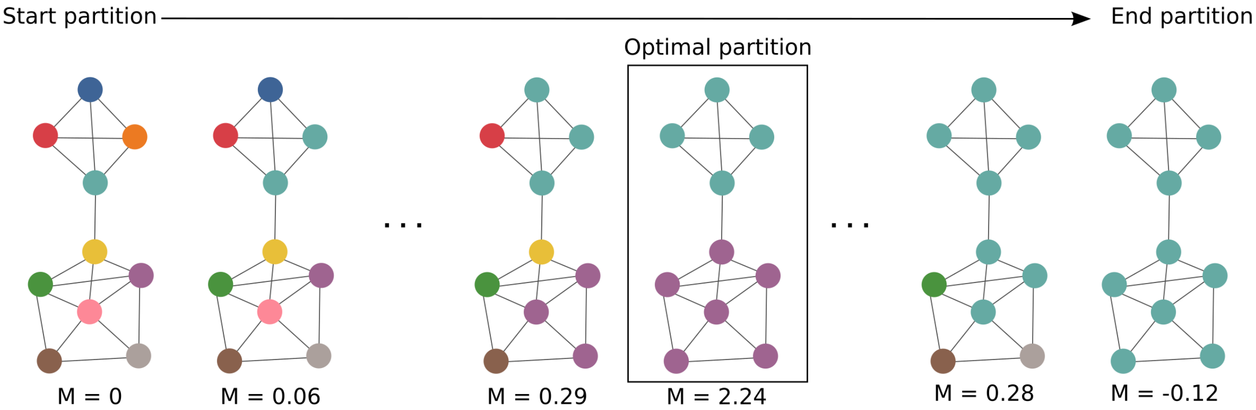


Figure 3.3: Fast greedy clustering algorithm for network community detection. The algorithm starts by assigning each node in the network to its own community. Then, it iteratively merges single nodes into existing communities until the whole network becomes a single community. The network modularity of each partition is recorded, and the network partition with maximal modularity is selected as optimal.

Systematic *in silico* perturbation of TFs in the ventricular CM GRN

The effect of TF perturbations on the ventricular CM transcriptome was predicted using celloracle (version 0.14), available at `https://morris-lab.github.io/CellOracle.documentation/#` [21]. Celloracle uses linear GRN models to predict shifts in cellular identity following the perturbation of one or more TFs. The celloracle pipeline is divided into three blocks: first, cluster-specific GRNs are inferred using regularized linear regression models. Then, these GRNs are used to model the changes in the cellular transcriptome upon a perturbation. Lastly, these changes are compared with the transcriptome changes induced by the differentiation trajectory (Fig. 3.4).

GRN inference with celloracle requires two inputs. One is a scRNA-seq data set in which cell type clusters have been identified and, optionally, a differentiation trajectory has been inferred. For this analysis, the purified ventricular CMs in the injury trajectory branch (*Cav3.1*$^+$, IS and Prol-IS clusters) were used. The second input is a base GRN with unweighted and directional edges between TFs and target genes. This base GRN reduces the number of potential edges that must be modeled and assigns a causality to the transcriptional regulation, which balances the simplicity of regularized linear models used to fit the GRN. As a base GRN, the ventricular CM GRN inferred with Pando containing 1,853 genes and 5,691 edges was used. These two pieces of information were used to fit cluster specific GRN models using regularized linear regression (Eq. 3.3):

$$y_i = \sum_{j=1}^{n} b_{i,j} x_j + c_{i,j} \tag{3.3}$$

where y_i is the expression of the target gene, $x_j \in \{x_0, x_1, ..., x_n\}$ are the expression levels of the regulators of gene i in the base GRN, $b_{i,j}$ is the coefficient and takes the value of 0 when genes i and j are the same, and $c_{i,j}$ is the intercept. L2 regularization was applied using Bayesian ridge to control the magnitude of the coefficients and to identify informative connections. Lastly, one-sided t-tests were used to calculate p-values for the coefficients. Robust connections were identified by filtering Bonferroni-corrected p-values using a significance threshold of 0.05.

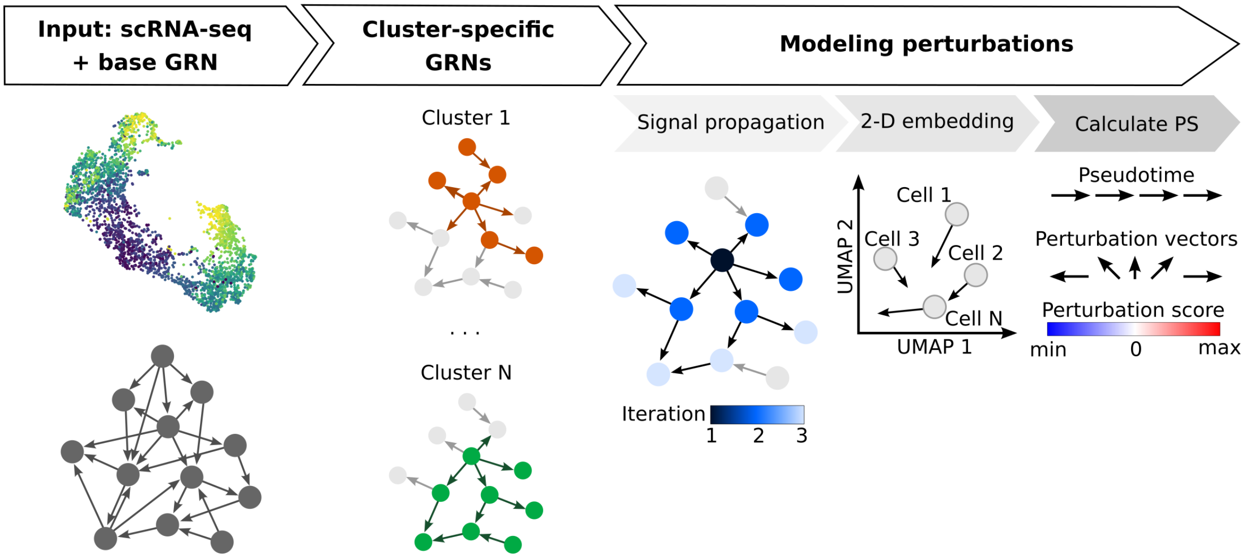


Figure 3.4: GRN inference and perturbation with celloracle. Celloracle [21] takes two inputs: (1) a scRNA-seq data set with annotated cell clusters and a differentiation trajectory and (2) a base GRN. These are used to fit GRN models using cluster-specific regularized linear regression. The fitted models can be used to predict how TF perturbations shift the cellular transcriptome using network propagation and by calculating the inner product between the perturbation and differentiation vectors in the 2-dimensional cellular embedding. A summarized version of the inner product called PS reports the effect and magnitude of the perturbation.

The second step in the celloracle pipeline is to use the linear GRN models to predict the effect of TF perturbations. The effect of a TF perturbation was estimated by propagating the shifts in gene expression through the GRN for three iterations. KO simulations assumed absolute TF expression of

0, while overexpression simulations utilized the maximum expression value observed in the dataset. The linear nature of the model makes these computations simple, as the derivatives (change in the expression of gene i as the expression of TF j changes) are the coefficients of the linear models (Eq. 3.4).

$$\Delta y_i = \frac{\delta y_i}{\delta x_j}\Delta x_j = b_{i,j}\Delta x_j \tag{3.4}$$

In each iteration, negative expression values were changed to zeros and controls checked whether the gene expression values remained within biologically plausible ranges. The resulting perturbation vectors were projected into the 2-dimensional UMAP embedding by calculating the Pearson correlation between the perturbation vectors and the gene expression differences between each cell and its k-nearest neighbors on the PCA space, with k optimized to 61 for this data set. The correlations were normalized using the $SoftMax$ function to represent the probability of transition between adjacent cell-states (Eq. 3.5):

$$p_{i,j} = \left(\frac{e^{cor(r_{i,j},d_i)/T)}}{\sum_{j\in G} e^{cor(r_{i,j},d_i/T)}}\right) \tag{3.5}$$

where d_i refers to the simulated gene expression shift vector for cell i, $r_{i,j}$ is the subtraction between the original gene expression vectors for cells i and $j \in G$ and T refers to a default temperature parameter of 0.05. These probabilities were transformed into 2-dimensional transition vectors (Eq. 3.6):

$$v_{i,simulated} = \sum_{j\in G} p_{i,j} v_{i,j} \tag{3.6}$$

where $v_{i,j}$ is the vector obtained by substracting the coordinates in the 2D space of cell i and cell j.

To facilitate the interpretation of results, the 2-dimensional transition vectors were summarized by grouping individual vectors by grid point using Gausian Kernel smoothing on a grid with 40 points. To compare the perturbation vectors with the differentiation trajectories, polynomial regression was applied to project the pseudotime onto the 2-dimensional grid. These projections were transformed into 2-dimensional differentiation vectors using the numpy.gradient() function. Then, a perturbation score (PS) that compares the perturbation and differentiation vectors was calculated through their inner product. The PS takes a value of 1 if the two vectors have the same direction, a value of 0 if they are perpendicular to each other, and a value of -1 if they have opposite directions. The PSs were summarized per cluster by adding all positive or negative PSs in grid points belonging to said cluster. Statistical significance of the PSs was empirically tested by obtaining a null distribution of the PSs calculated with a randomized GRN model. The obtained p-values were Bonferroni-corrected to account for multiple testing.

To select candidates for further experimental exploration, the TFs were ranked by these three criteria:

1. Statistical filter: Bonferroni-adjusted p-value < 0.05 for either the negative PS and/or positive PS.
2. Magnitude filter: PS larger than the percentile 99 of random PS.
3. Coherence filter: in order to select the perturbations that have a homogeneous effect, a final PS that summarizes the negative and positive PS was calculated (Eq. 3.7). We focused on the TFs with the largest summarized PS.

$$PS_{total} = -PS_{PS<0} + PS_{PS>0} \tag{3.7}$$

3.1.2 Methods related to the mouse data set

Experimental procedures

All experimental procedures were performed by Cristina Villa del Campo, PhD and Rocío Sierra and are briefly described here to provide the necessary background for the *in silico* analysis of the gathered data. The scripts used for GRN inference will be available at `https://github.com/iriverog`.

Mouse husbandry: P7 and adult (8-10 weeks old) BHP//B6.Cg-Tg(HIST1H2BB/EGFP)1Pa/J male and female mice were used in this project. This mouse line expresses the EGFP nuclear reporter, which allowed the classification of CMs as mono- or binucleated before scRNA-seq. Animals were bred and maintained in the animal facilities of the Centro Nacional de Investigaciones Cardiovasculares (CNIC) in Madrid, Spain. All experiments were conducted in accordance with the CNIC Ethics Committee, Spanish laws, and the EU Directive 2010/63/EU for the use of animals in research. All mouse experiments were approved by the CNIC and the area of “Protección Animal” of the Community of Madrid with reference PROEX 144.1/21.

Tissue dissection, cell isolation and nuclei annotation: WT mice were sacrificed by cervical dislocation at P7 and adulthood (8-10 weeks). After euthanasia, hearts were harvested, and ventricular CMs were purified by Langendorff perfusion with liberase (Roche Applied Science). In preparation for scRNA-seq with the SMART-seq2 technology, single ventricular CMs were visually inspected, annotated as “mononucleated” or “binucleated” thanks to the EGFP nuclear reporter, and isolated in single wells of 96-well plates. Mononucleated ventricular CMs were enriched with the purpose of having comparable sample sizes for statistical testing.

Mouse line Hey2^{CreERT2}: the mouse line Hey2^{CreERT2};R26R^{mTmG} (Hey2^{CreERT2}, in short) [186] was used to track the *Hey2*-expressing lineage *in vivo*. Upon tamoxifen administration, this mouse line expresses the Cre recombinase under the control of the *Hey2* promoter and Cre activity recombines the R26R^{mTmG} allele to inactivate the membrane-Tomato reporter and activate the membrane-GFP (mGFP) fluorescent reporter.

Correlation between the *Hey2* lineage and number of CM nuclei: Cre expression and subsequent mGFP expression in cells with active *Hey2* was induced by oral tamoxifen administration to Hey2$^{CreERT2/+}$ adult mice. Afterwards, the hearts were harvested, and ventricular CMs were isolated using Langendorff perfusion with liberase (Roche Applied Science). Purified ventricular CMs were plated, and the number of mononucleated ventricular CMs in the *Hey2*$^-$ (mGFP$^-$) and *Hey2*$^+$ (mGFP$^+$) lineages was counted.

scRNA-seq

scRNA-seq was performed using the SMART-seq2 platform in collaboration with the CNIC Genomics unit. Complementary DNA (cDNA) generation, pre-amplification and library preparation were performed using the SMART-Seq v4 3'DE kit (Takara Bio) following manufacturer instructions. Libraries were sequenced on an HiSeq2500 or NextSeq sequencing platform (Illumina) with read lengths of 49 nucleotides (R1) and 10 nucleotides (R2). Fastq files containing reads for each cell were obtained by demultiplexing raw sequencing files using Nextera indexes and bcl2fastq2 (Illumina) followed by demultiplexing using R2 indexes and the cutadapt software [187]. Reads were mapped and quantified on the mouse genome GRCm38 with ensemble gene build version 91 using the RSEM software [188]. These pre-processing steps were performed by Carlos Torroja, PhD and the CNIC Genomics and Bioinformatics units.

scRNA-seq analysis

These steps were performed by Carlos Torroja, PhD, from the CNIC Bioinformatics Unit.

scRNA-seq pre-processing and clustering was done using the R packages Seurat v2 [169] and Scater [189]. The following three criteria were used to minimize the number of low-quality cells and improve the subsequent analysis: (1) cells with normalized counts greater than 1,700,000 or less than 150,000, (2) cells expressing less than 4,000 genes, and (3) a mitochondrial transcript content above 75% were excluded. Genes detected in at least 6 cells were considered for further analysis.

Analysis of adult ventricular CM: The sequencing batches were integrated using the CCA method [190] on the SCT-transformed data. Briefly, the SCT transformation [191] was calculated for the 500 most variable genes, and 200 of those were used for calculating the integration anchors using 10 PCA dimensions and 5 nearest neighbors and considering 60 nearest neighbors when weighting anchors. For visualization purposes, the integrated dataset was subjected to UMAP dimensionality reduction using 10 PCA PCs. Louvain clusters were estimated with a resolution of 1 and 10 PCs. Marker genes of each cluster were calculated with the Seurat function FindAllMarkers() using Wilcox Sum Ranked test on the SCT-transformed data. Only positive log fold change entries greater than 0.25, with an adjusted p-value < 0.01 and detected in at least 30% of cells were reported as cluster markers. Based on the manual exploration these markers, cluster C3 was further divided into three subclusters: C3a, C3b and C3c. Marker genes of the manually corrected clusters were calculated as before. Clusters were annotated to the cardiac conduction system or ventricular working myocardium (WM) using marker genes (Table 3.2) obtained from Goodryer *et al.* [192].

Analysis of P7 ventricular CM: P7 ventricular CM were analyzed as described above, with a slight modification to the number of nearest neighbors when weighting anchors, which was set to 30 for this data. In this case no manual refinement of cluster assignment was necessary.

Table 3.2: Marker genes for the identification of CM subtypes. Obtained from [192].

CM subtype	**Marker genes**
Cardiac conduction	*Acta1, Actc1, Adprhl1, Aspscr1, Atp1b1, Bex4, Casq2, Cited1, Clu, Cst3, Cxcl12, Eid1, Fbn2, Fxyd1, Hspb1, Igf2r, Kcne1, Mlip, Myh6, Myl7, Nppa, Pam, Pgam2, Ramp1, Rbpms, Rrad, Sh3bgr, Slit2, Sorbs2, Sulf2, Tcap, Tgm2, Tnnc1, Tnni3, Tnnt1, Tsc22d3, Trdn*
Atrio-ventricular node	*Acta1, Actc1, Adprhl1, Aspscr1, Bex4, Clu, Fxyd1, Hspb1, Kcne1, Mlip, Nppa, Pam, Pgam2, Rbpms, Rrad, Sh3bgr, Sorbs2, Tnnc1, Tnni3, Tnnt1*
Purkinje fiber	*Atp1a1, Atp1b1, Casq2, Cited1, Cst3, Cxcl12, Eid1, Fbn2, Fxyd1, Igf2r, Myh6, Myl7, Pam, Ramp1, Slit2, Sulf2, Tcap, Tgm2, Tsc22d3, Trdn*
Sino-atrial node	*Aspscr1, Atp1a1, Atp1b1, Bmp2, Cacna2d2, Csrp2, Fbxo32, Igfbp5, Lbh, Myl2, Pln, Ramp1, Shox2, Smoc2, Snap91, Sorbs2, Tcap, Tmod1, Tpm1, Vsnl*
Atrial myocardium	*Gja1*$^-$, *Gja5*$^+$, *Hcn4*$^-$, *Myh7*$^+$, *Nkx2-5*$^+$, *Scn5a*$^+$
Ventricular WM	*Gja5*$^-$, *Gja1*$^+$, *Hcn4*$^-$, *Myh6*$^+$, *Myh7*$^-$, *Myl2*$^+$, *Nkx2-5*$^+$

Identification of TFs driving binucleation

Identification of mononucleated CM markers in adult WM: adult WM CMs (clusters C0, C1 and C2) were virtually isolated. The Seurat function FindMarkers() using the MAST test [193] was used to identify differentially expressed genes between mono- and binucleated CMs, regardless of their cluster membership. All genes detected in more than 10% of cells were tested and their differential expression p-values were corrected using FDR.

Identification of TFs controlling polyploidization: the differentially expressed genes (FDR-corrected p-value < 0.05) were used as input to predict which TFs can potentially explain the observed differences in gene expression between mono- and binucleated CMs. This analysis was performed using Ingenuity Pathway Analysis (IPA) [111] and the Ingenuity® Knowledge Base (QIAGEN Inc., `https://digitalinsights.quage.com/IPA`), a literature-based and human-curated database of over 24 million relationships between molecules, diseases and phenotypes.

IPA finds candidate TFs to explain the observed gene expression differences by (1) examining how many known targets of a TF are present in the data set and by (2) comparing their direction of change (Log_2FC) to what would be expected from the literature. It provides 2 statistical measures of TF relevance:

1. Overlap p-value: p-value of the overlap between known targets for a TF and the differentially expressed genes, as calculated by the right-tailed Fisher's Exact Test and adjusted for multiple testing using Bonferroni's correction (Fig. 3.5a).

2. Activation z-score: determines the activation state of a TF (active or inactive) from the known MOR for each target (Fig. 3.5b). Considering a given TF in the Ingenuity® Knowledge Base that is connected to at least one gene in the observed data, these directional relationships are annotated with literature findings that report consistent or ambiguous MORs. IPA predicts the activation state of the TF by defining a z-score that quantifies whether the TF has more activating predictions than inhibiting predictions ($z>0$) or vice versa ($z<0$). The z-score is calculated as follows (Eq. 3.8).

$$z = \frac{\sum_{i=1}^{N} w_i x_i}{\sqrt{\sum_{i=1}^{N} w_i^2}} \tag{3.8}$$

   This equation is independently applied to each TF, where i refers to each of its observed targets and N is the total number of observed targets in the data set. The variable x, which quantifies the number of activating and inhibiting predictions, is calculated as (Eq. 3.9)

$$x = \sum_{i=1}^{N} x_i = \frac{N^+ - N^-}{\sqrt{N}} \tag{3.9}$$

   and has variance $\sigma^2 = 1$. N^+ and N^- are the number of activated and inhibited targets, respectively. The variable w quantifies the consistency in the MOR for each TF – target relationship (Eq. 3.10):

$$w_i = \frac{|M_{act} - M_{inh}|}{M_{act} + M_{inh} + 1} \tag{3.10}$$

   where M_{act} and M_{inh} refer to the number of literature findings supporting an activating or inhibiting TF – target relationship, respectively.

TFs with an adjusted overlap p-value < 0.05 were ranked by their absolute activation z-score to select the most likely candidate(s) to explain the difference in gene expression between mono- and binucleated CMs. To account for the fact that the detected TFs are not necessarily independent of each other (e.g., the effect of TF_1 in the data set is relayed through TF_2, with both TF_1 and TF_2 being detected as independent significant regulators), the Mechanistic Networks algorithm (QIAGEN Inc.) was applied to find causal GRNs [111]. This algorithm assumes that, if the causal effect of TF_1 on some gene g is transmitted through TF_2, there should be an increased occurrence of the $TF_1 \rightarrow g$, $TF_2 \rightarrow g$ and $TF_1 \rightarrow TF_2$ edges in the network. This enrichment can be quantified by the Fisher's Exact Test p-value of the overlap of the regulated gene sets $O(TF_1) \cap O(TF_2)$ in the data set, using the common target genes of TF_1 and TF_2 as universe. By calculating this p-value for all significant TFs, the Mechanistic Network algorithm predicts the causal hierarchy of events that lead to the observed gene expression changes (Fig. 3.5c). The resulting network was visualized using Cytoscape v.3.9.0.

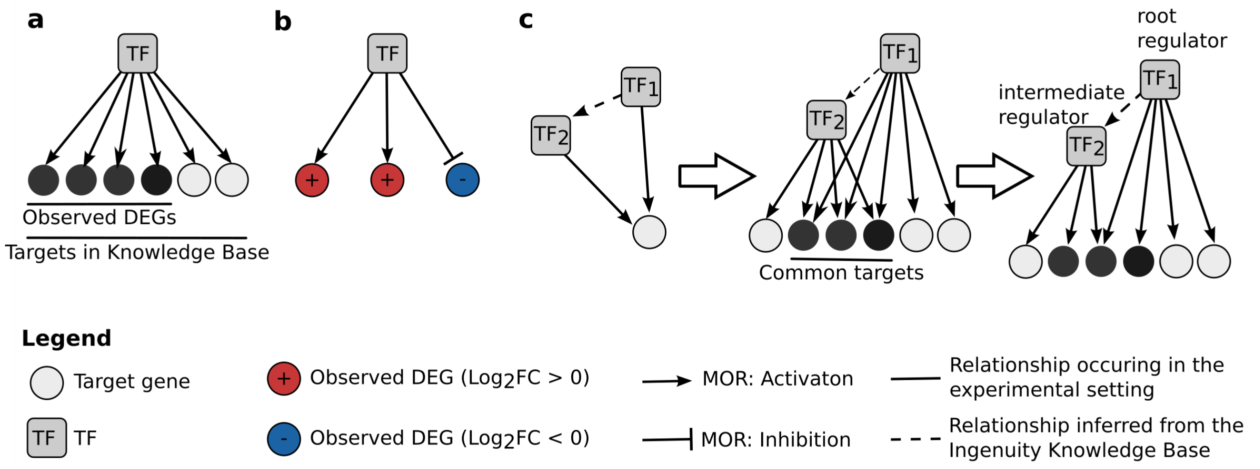


Figure 3.5: Identification of relevant TFs causing the observed changes in gene expression using IPA (QIAGEN Inc.). Panels **a-b)** show the two statistic measures used by IPA to find likely relevant TFs to explain the gene expression differences: **a)** the overlap between known targets of a TF and differentially expressed genes in the data set, expressed as the p-value of the right-tailed Fisher's Exact Test; and **b)** the predicted TF activation state, as expressed by the activation z-score (see text). **c)** Scheme of the Mechanistic Networks algorithm, used to find the causal GRN behind the data. First, "causal transitive triangles" are identified using the Fisher's Exact Test p-value of enrichment between shared TF_1 and TF_2 targets in the observed data. These causal dependencies are pruned using the Ingenuity® Knowledge Base to find the causal chain of regulatory events behind the data. In all panels, circles represent target genes and squares represent TFs. The target gene color represents overlapping targets (black), targets with an observed $\log_2$FC > 0 (red) or targets with an observed $\log_2$FC < 0 (blue). Pointed arrows represent activating MOR and blunted arrow ends represent inhibiting MOR. Continuous lines represent relationships supported by the expression data and discontinuous lines represent relationships inferred only from the previous knowledge.

3.2 Results

3.2.1 The dynamic landscape of axolotl ventricular CMs upon cryoinjury

The molecular mechanisms that allow the axolotl to regenerate its heart are unknown. To answer this question, we generated a comprehensive multi-omic atlas of intact and injured axolotl hearts at 0, 1, 4, 7 and 14 dpi (Fig. 3.1). Through computational integration of chromatin accessibility and gene expression on 38,280 nuclei and unbiased clustering, we identified 41 molecularly distinct clusters that were aggregated into 7 broad cell type classes using the expression of known marker genes (Fig. 3.6a, b). These included CMs (e.g., *Myh6*$^+$, *Myh7*$^+$, *Nppa*$^+$, *Ryr2*$^+$, *Tnnt2*$^+$), endothelial cells (e.g., *Cdh5*$^+$, *Eng*$^+$, *Kdr*$^+$, *Pecam1*$^+$, *Vwf*$^+$), erythrocytes (e.g., *Alas2*$^+$, *Hbg1*$^+$, *Hbg2*$^+$), fibroblasts (e.g., *Col1a1*$^+$, *Dcn*$^+$, *Krt19*$^+$, *Pdgfra*$^+$, *Postn*$^+$), immune cells (e.g., *Adgre1*$^+$, *Blnk*$^+$, *Ccr5*$^+$, *Itgam*$^+$, *Lyve1* $^+$), neurons (e.g., *Cacna1a*$^+$, *Cacnb1+*, *Cacng3+*, *Kcnj12*$^+$, *Snap25*$^+$) and platelets (e.g., *Fli1*$^+$, *Gp9*$^+$, *Plec*$^+$, *Selp*$^+$, *Zfpm1*$^+$). By leveraging a spatial transcriptomics data set of axolotl heart at 4 dpi, we could explore the spatial distribution of each cell type and, importantly, annotate the different CM clusters as located to the ventricle, atria, or outflow tract (OFT) (Fig. 3.6c).

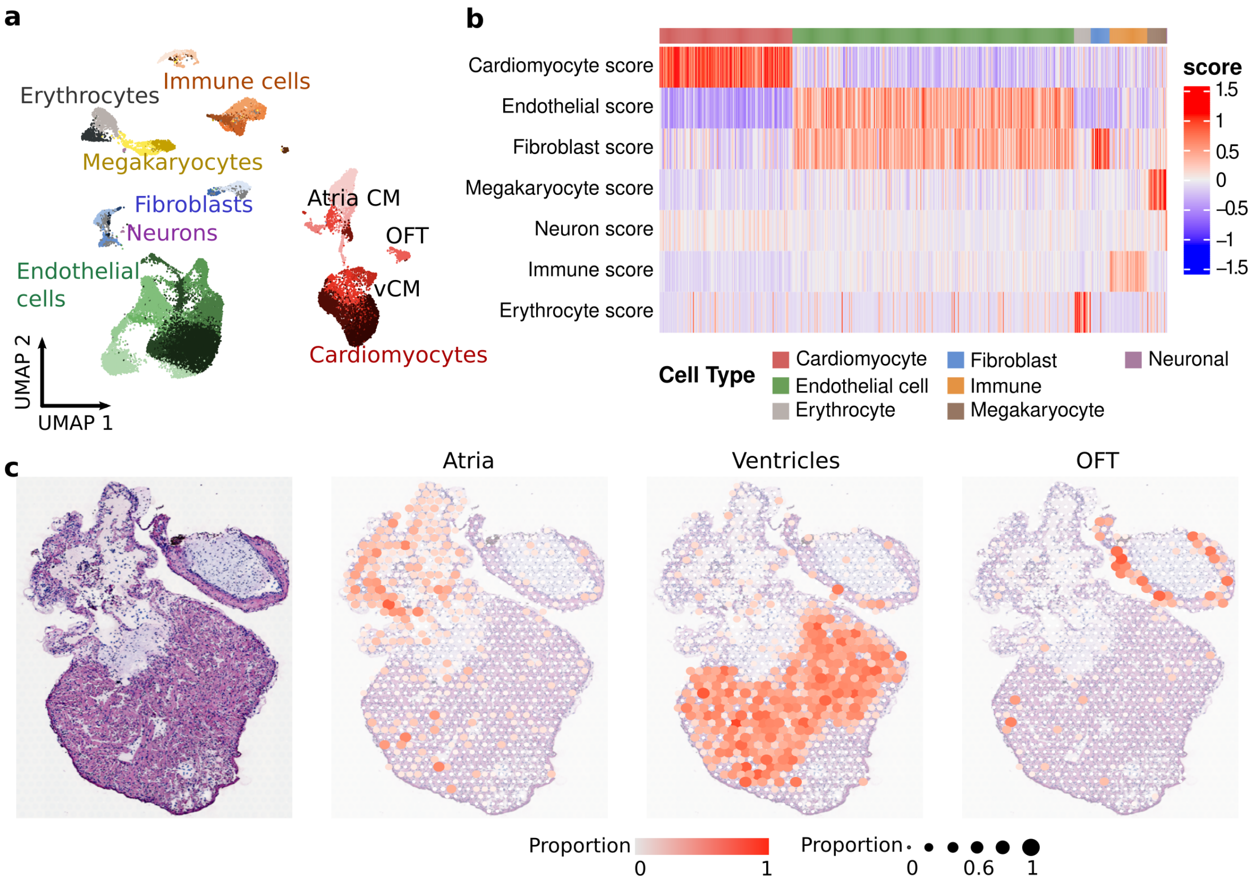


Figure 3.6: A single-nuclei multi-omic atlas of axolotl heart regeneration. a) UMAP visualization of the integrated snRNA-seq and snATAC-seq. Different colors identify the major cell types in the dataset. Different shades identify the distinct molecular clusters for each cell type. **b)** Heatmap of cell type scores calculated from the expression of known cell type marker genes (Table 3.1). Only 500 randomly sampled cells are shown. **c)** Proportions of atrial, ventricular and OFT CMs on a Visium spatial transcriptomics data set of axolotl heart at 4dpi. The red color gradient and spot diameter indicate higher abundance.

We focused on the ventricular CMs, as they are directly affected by the injury and responsible to generate the lost myocardium [4, 28]. We found 6 molecularly distinct ventricular CM clusters with differences in spatial localization, injury dynamics and cell cycle status: *Cav3.1*$^+$, *Robo2*$^+$, injury-specific (IS), proliferative IS (Prol-IS), proliferative in S phase (Prol-S) and proliferative in G2/M phase (Prol-G2M) (Fig. 3.7a). Regarding their localization, the IS and Prol-IS populations are mostly restricted to the border zone, the *Cav3.1*$^+$ CM are located to the ventricular periphery and *Robo2*$^+$ CM are found in the inner myocardium mass (Fig. 3.7b). As for their injury dynamics, the IS and Prol-IS populations appear only after injury, while the four other clusters are present in uninjured hearts. Interestingly, while the IS population is detectable throughout the examined timepoints, the Prol-IS population is transient and only detected in the samples collected at 4 and 7 dpi (Fig. 3.7c). Regarding their cell cycle state, *Cav3.1*$^+$, *Robo2*$^+$ and IS CMs are mostly in interphase and the two basal proliferation clusters can be associated with the S (Prol-S) and G2/M

(Prol-G2M) phases. The Prol-IS cluster contains a mixture of the three phases, with a similar and predominant contribution from G1 and S cells (Fig. 3.7e). Given the temporally dynamic nature of the data, we explored the dynamics between different ventricular CM clusters using trajectory inference with SCANPY [174]. Two separate trajectories were found: one that predicts the transformation of *Cav3.1*$^+$ CMs into IS and subsequently Prol-IS (Fig. 3.7d, injury trajectory, in red), and another that predicts the cell cycle progression of *Robo2*$^+$ cells as they transition into Prol-S and Prol-G2M CMs (Fig. 3.7d, proliferation trajectory, in blue). Altogether, this data shows the existence of six ventricular CM subpopulations in the injured axolotl heart, four of which appear in basal conditions and two being triggered by injury and restricted to the border zone. Trajectory analysis predicts the *Cav3.1*$^+$ CMs as the main contributor to the injury-specific CM populations.

3.2.2 Distinct communities in the GRN regulate axolotl ventricular CM identity

To understand what TFs are regulating the ventricular CM identities and their transitions, we inferred their GRN using Pando, which leverages the snATAC-seq data to find meaningful TF – region – target gene relationships that are fit as linear models using the snRNA-seq data [22]. The GRN inference was restricted to the 10,000 most variable genes in the data in order to find a balance between the computational and biological complexity of the problem. The resulting GRN contains 1,853 genes and 5,691 edges, organized into 87 communities (Fig. 3.8 and Table 3.3). The top hub TFs in this network, based on betweenness centrality, are *Smad3*, *Ets1*, *Lef1*, *Glis3*, *Hey2*, *E2f1*, *E2f8*, *Runx1*, *Tbx20* and *Mybl1* (Fig. 3.8a).

Table 3.3: Ventricular CM GRN topology analysis.

Metric	CM GRN	Interpretation
Number of nodes	1,853	Number of genes in the network
Number of edges	5,691	Number of connections in the network
Average number of neighbors	4	Indicates the average connectivity of a node
Network diameter	5	Largest distance between two nodes in the network
Characteristic path length	1.963	Average distance between two nodes in the network
Number of communities	87	Community: group of genes more connected to each other than to the rest of the network

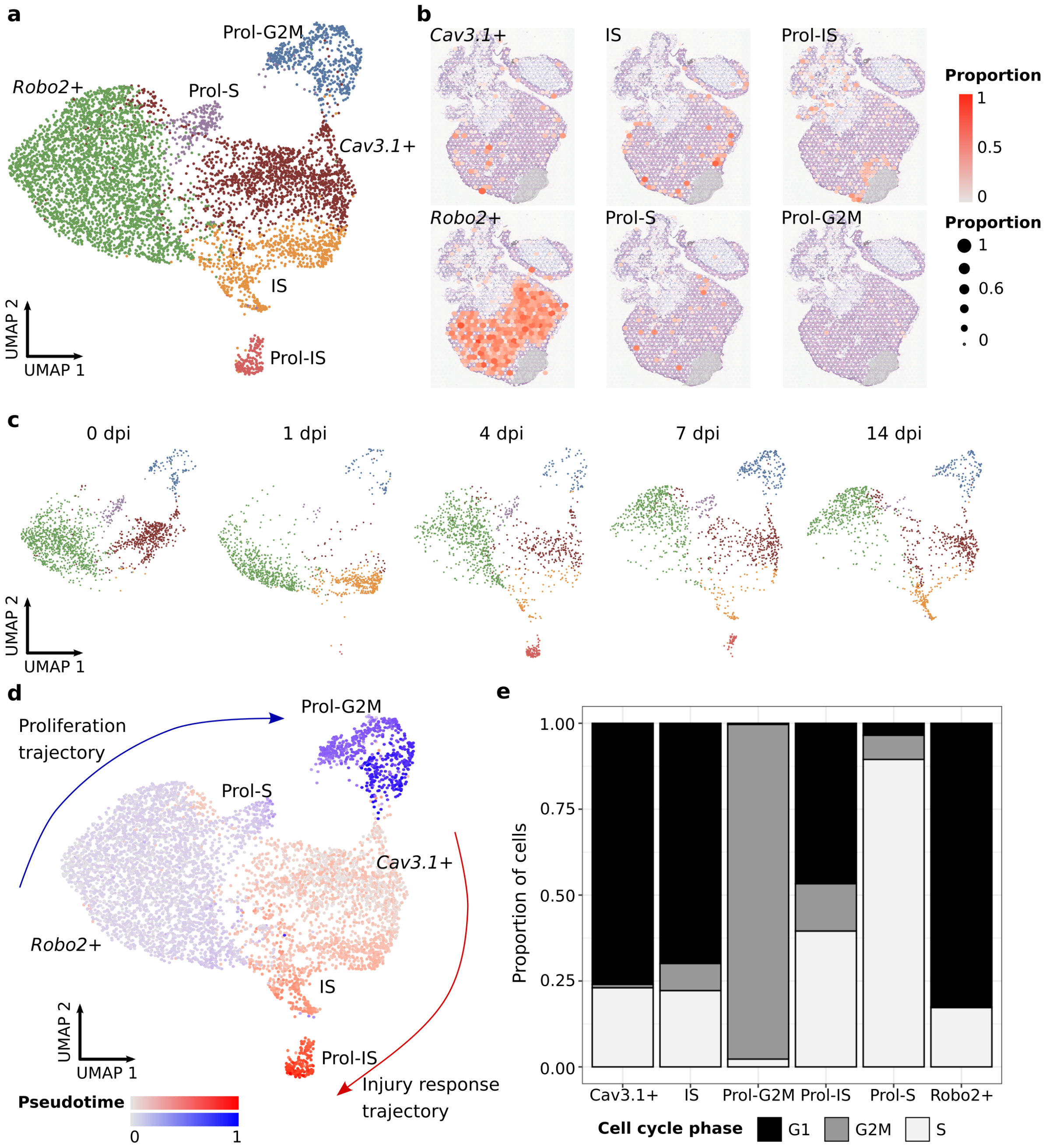


Figure 3.7: Dynamic changes in ventricular CM populations upon cryoinjury. a) UMAP visualization of the integrated snRNA-seq and snATAC-seq for ventricular CMs, colored by cluster membership. **b)** Proportions of each ventricular CM subtype on a Visium spatial transcriptomics data set of axolotl heart at 4dpi. The red color gradient and spot diameter indicate higher abundance. **c)** Timeline of changes in the CM clusters abundance at 0, 1, 4, 7 and 14 dpi. **d)** UMAP visualization of the integrated snRNA-seq and snATAC-seq for ventricular CM, colored by the two independent trajectories found by SCANPY. The injury trajectory (red) shows a transformation of *Cav3.1*$^{+}$ ventricular CM into IS and subsequently Prol-IS ventricular CM. The proliferation trajectory (blue) shows a transformation of *Robo2*$^{+}$ ventricular CM into Prol-S and Prol-G2M ventricular CM. **e)** Proportion of cells in each cell cycle phase (G1, S or G2/M) for each ventricular CM cluster.

Communities facilitate the crosstalk between genes, and hence genes participating in similar processes tend to be members of the same community [194–196]. We annotated the biological function of each community by performing an ORA of the community genes in the GO Biological Process database [183, 184]. We could associate 11 (12.64%) of the communities with biological functions (Table 3.4). Interestingly, the three largest communities (7, 15 and 24) are associated with heart maturation and contraction (community 7, red in Fig. 3.8a), cell cycle (community 15, purple in Fig. 3.8a) and wound healing (community 24, pink in Fig. 3.8a). These three functions are a reminiscence of the ventricular CM clusters, leading us to hypothesize that predominant usage of specific communities determines CM subtypes. By visualizing the $\log_2$FC of ventricular CM cluster markers in the GRN, we identified a preference for the "heart and ventricle maturation and contraction community" in *Cav3.1*$^+$ and *Robo2*$^+$ subpopulations (Fig. 3.8b, e). As cells advance in the injury trajectory, the main GRN is switched towards the "wound healing, collagen-containing ECM and heart development" community (Fig. 3.8c, d). Regarding the proliferation trajectory, cell cycle progression leads to an activation of the "cell cycle" community (Fig. 3.8f, g). Collectively, these results show that the ventricular CM GRN contains several communities with different functionalities. Basal CM function seems to be governed by a single community dominated by the hub TFs *Tbx20* and *Hey2*. Cell cycle progression temporally switches the CM gene regulation towards a "cell cycle" program with *E2f1* and *E2f8* as hub TFs. Lastly, ventricular cryoinjury activates a "wound healing" community in which TF such as *Lef1* and *Ets1* have a main regulatory role.

Table 3.4: Biological function and hub genes of the communities in the ventricular CM GRN. Only those communities with at least one statistically significant association (FDR-adjusted p-value < 0.05) appear in the table. Hub genes were identified based on their out-degree due to the large number of ties in betweenness centrality for most genes in the GRN.

Community	**Biological function**	**Top hubs**
1	Myofibril assembly, CM differentiation	*Cux1, Creb5, Mef2d*
2	Mesenchyme development	*Znf664, Tfap2b, Nr1h4*
5	Histone H3K9 methylation	*Myb, Pax5, Tcaf2*
7	Heart and ventricle maturation, heart contraction	*Tbx20, Zeb1, Hey2*
9	DNA methylation	*Egr1, Irf1, Fosb*
15	Cell cycle	*E2f1, Mybl1, E2f8*
16	Epithelial cell development, cell death, estrogen gene expression	*Foxa1, Bhlha15, Nhlh2*
24	Wound healing, collagen-containing ECM, heart development	*Lef1, Ets1, Klf13*
50	Muscle contraction, musculoskeletal development	*Tbx15, Foxf1, Sox17*
77	Notch signaling	*Lhx9, Hey1, Hes5*
78	Interferon response	*Stat1, Irf3, Irf8;Irf9*

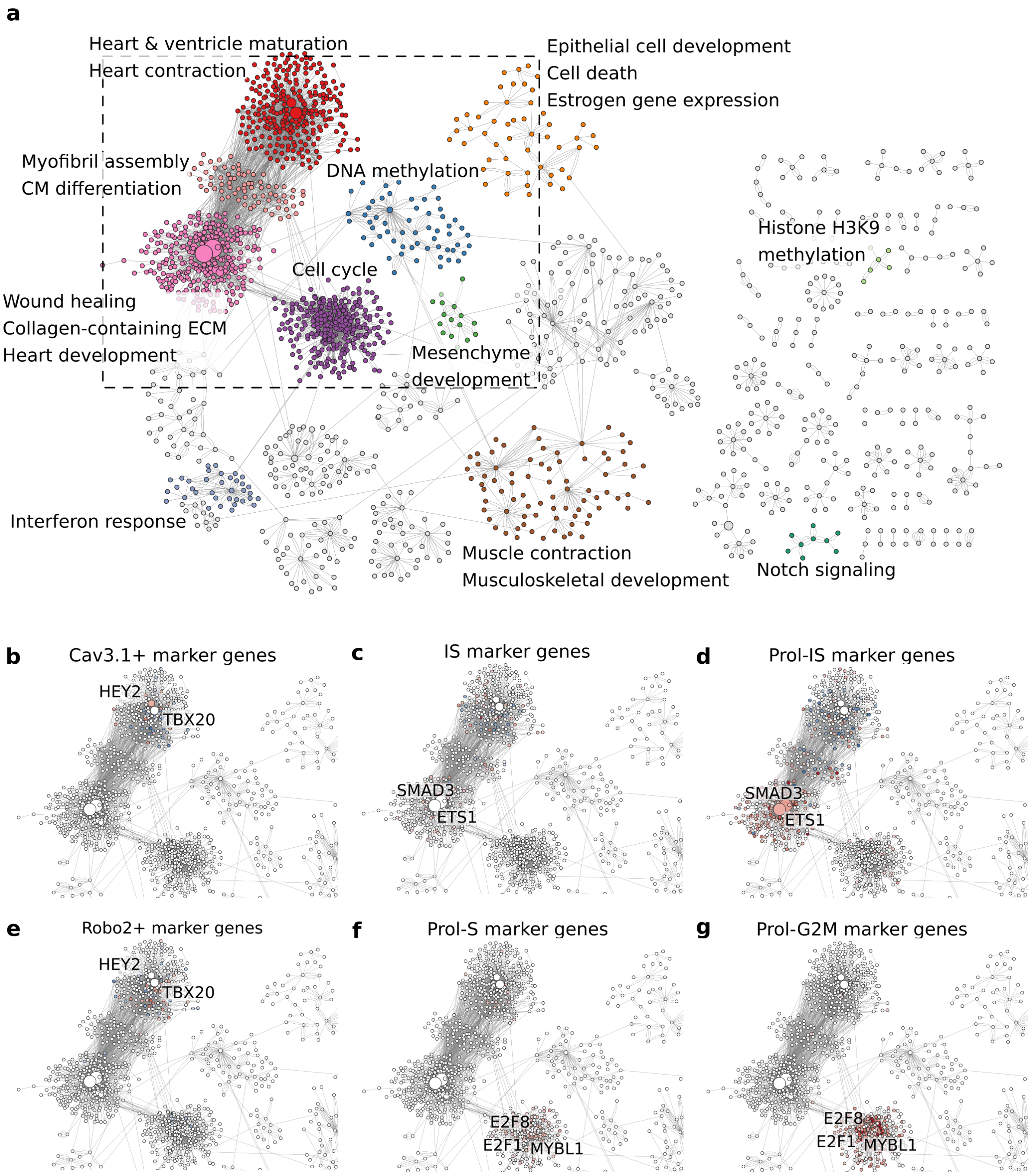


Figure 3.8: Different communities in the GRN regulate axolotl ventricular CM identity. a) CM GRN inferred with Pando [22] Genes are colored by their community membership. Communities for which at least one statistically significant (FDR adjusted p-value < 0.05) biological function was found are highlighted. Gene size represents betweenness centrality. **b-g)** Expression of marker genes for the **b)** *Cav3.1*$^+$ CM, **c)** IS CM, **d)** Prol-IS CM, **e)** *Robo2*$^+$ CM, **f)** Prol-S CM and **g)** Prol-G2M CM. The section of the network shown in these visualizations is marked in panel **a)** by a discontinuous line. In panels **b-g)**, gene size represents betweenness centrality. Gene color represents $\log_2$FC, with red representing positive markers and blue representing negative markers for a given cluster.

3.2.3 Systematic *in silico* TF perturbations discover the antagonistic role of *Hey2* during regeneration

Having seen that cryoinjury switches the ventricular CM GRN towards a wound healing program, we wondered which changes in TF activity are responsible for this. Since GRNs based on linear models can be used to simulate the effect of perturbations (e.g., TF KO or overexpression) on the cellular transcriptome, we used celloracle implementation of *in silico* perturbations to simulate the effect of knocking-down or overexpressing the 268 TF present in the GRN. For this analysis, we focused on the 3 CM populations that participate in the injury-response trajectory: *Cav3.1*$^+$, IS and Prol-IS CM (Fig. 3.9a, b). We measured the impact of each perturbation by calculating the inner product between the perturbation vectors and trajectory vectors, named PS. This PS takes positive values if the perturbation and trajectory vectors follow the same direction, and negative values if the perturbation and trajectory vectors oppose each other.

The predicted effect of all perturbations along with the names of top candidates are summarized in Fig. 3.9c (see Sup. Fig. S3.1 for the consequences of the perturbations on a randomized version of the GRN). Knocking-down TFs in red could promote the injury trajectory progression (e.g., *Hey2*, *Zeb1*, *Mef2d*), while knocking-down TFs in blue could have the opposite effect (e.g., *Klf13*, *Smad3*, *Ets1*). The candidate with the largest predicted effect was *Hey2*, a basic helix-loop-helix (bHLH) transcriptional repressor with an essential role in cardiovascular development [197]. In fact, *Hey2* is a positive marker of the *Cav3.1*$^+$ subpopulation of CMs ($\log_2$FC = 0.5, adjusted p-value = 4.4e-207) and its expression is downregulated upon injury (Fig. 3.9d). Looking at the predicted effect of *Hey2* perturbations in more detail, the *Hey2* KO seems to promote the transition from *Cav3.1*$^+$ to IS populations, while having a limited effect on the Prol-IS CMs (Fig. 3.9e). As expected, the overexpression of *Hey2* is predicted to have the opposite effect and hinder this transition (Fig. 3.9f). Aiming to validate these findings, we generated a transgenic axolotl line that overexpresses TFP-*Hey2* under the control of a minimal *myl7* promoter. F0 mosaic animals were subjected to ventricular cryoinjury and the proliferation levels of *Hey2*$^+$ (TFP$^+$) and *Hey2*$^-$ (TFP$^-$) border zone CMs were assessed by EdU incorporation at 7 dpi. Preliminary results suggest a decrease in the abundance of EdU$^+$ CMs at 7 dpi in *Hey2* overexpressing cells, while no differences are observed in uninjured controls. Altogether, these results support the critical role of *Hey2* for the maturation and function of ventricular CMs and suggest that its inactivation is necessary to initiate a regenerative response in the axolotl heart (Fig. 3.9g).

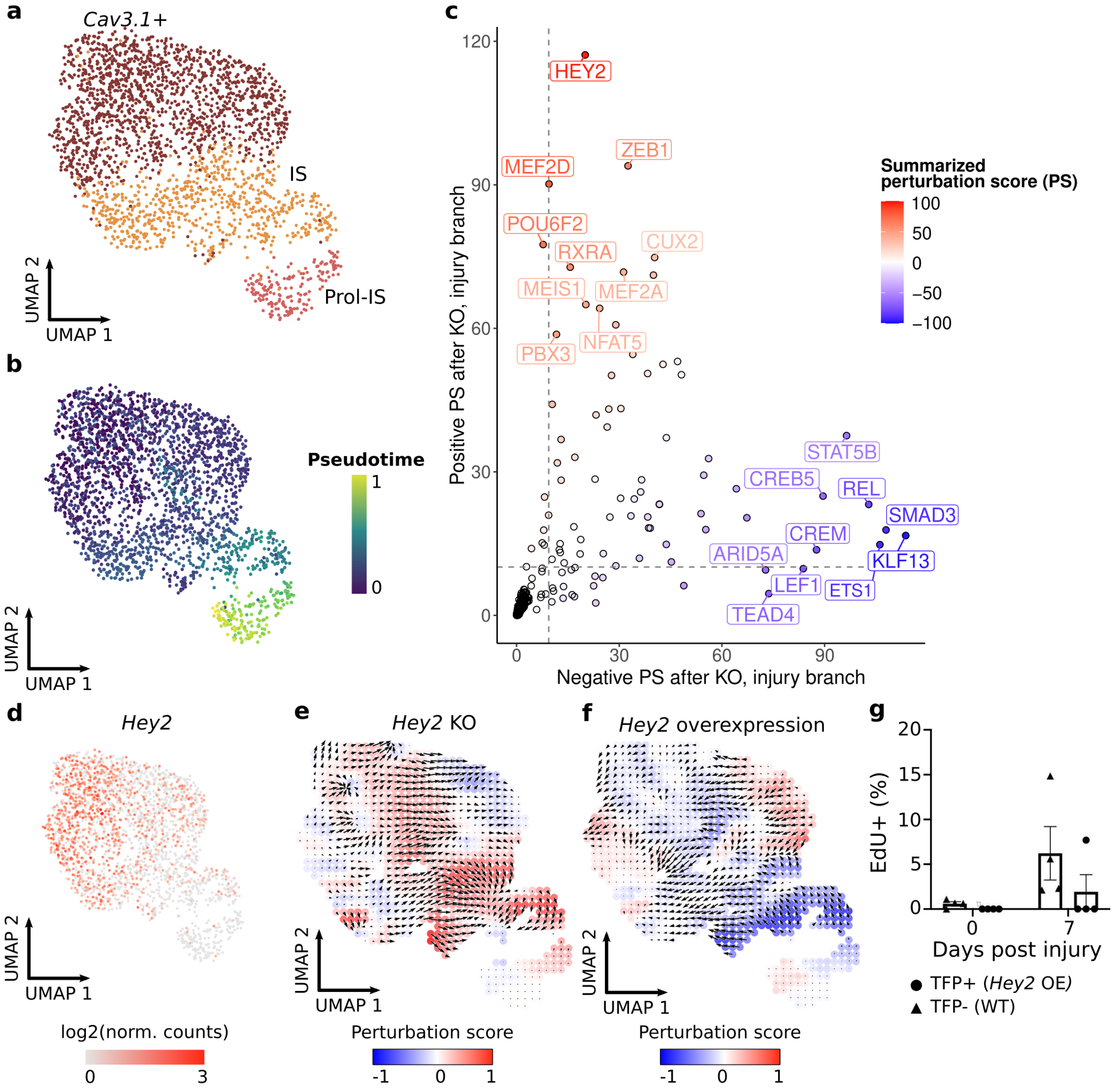


Figure 3.9: Systematic *in silico* TF perturbation discover *Hey2* as an antagonist for regeneration. **a)** UMAP visualization of the injury branch, which includes the *Cav3.1*$^{+}$, IS and Prol-IS CM clusters. Cells are colored based on cluster membership. **b)** UMAP visualization of the injury branch, with cells colored by pseudotime. **c)** Scatterplot of the predicted effect of TF KO on the injury branch trajectory. The effect of TF KO on the trajectory is quantified by a PS (See Methods). Negative PS (blue) means that the TF KO hinders the trajectory from progressing. Positive PS (red) means that the TF KO promotes the progression of the trajectory. The discontinuous lines indicate the 99 percentile of random PS. **d)** UMAP visualization of the injury branch, with cells colored by their expression level of *Hey2*. **e)** UMAP visualization of the injury branch showing the predicted effect of *Hey2* KO. The colors represent whether the *Hey2* KO promotes (red) or hinders (blue) the progression of the trajectory. The arrows indicate the direction towards which cells will transform as a consequence of the KO. **f)** UMAP visualization of the injury branch showing the predicted effect of *Hey2* overexpression. The colors represent whether the *Hey2* overexpression promotes (red) or hinders (blue) the progression of the trajectory. The arrows indicate the direction towards which cells will transform as a consequence of the overexpression. **g)** Percentage of TFP^{+} (*Hey2* overexpressing) and TFP^{-} EdU^{+} CM measured at 0 and 7 dpi. Each dot indicates a biological replicate.

3.2.4 Molecular profiling of mononucleated diploid and binucleated polyploid adult mouse CMs

The CMs of the vertebrate myocardium have a strict spatial organization and different developmental origins and functions, mainly being atrial, ventricular, or conduction system cells [192]. Within the mammalian ventricle, most CM undergo polyploidization by endoreplication, predominantly undergoing karyokinesis in mouse (2x2n) but not in human (1x4n) [198]. In the mouse, all embryonic CM are mononucleated diploid (1x2n, MNDCM) but they massively polyploidize (2x2n) during the first postnatal week of life [199], and only an average of 6% of all mouse ventricular CM remain as MNDCM in adulthood [200]. Interestingly, the MNDCMs have a higher proliferative capacity [73] and there is causal demonstration that regeneration is blocked when zebrafish ventricular CM are forced to polyploidize [70]. As the MNDCMs have been suggested to be the source for new CMs upon injury [200], finding a unique MNDCM transcriptomic signature would be key to understanding their function and increasing their frequency to, potentially, induce regeneration.

In this project we generated a high-quality scRNA-seq of CMs for which the number of nuclei were manually annotated through visual inspection after isolation by Langendorff perfusion. As mononucleated CMs in the mouse are almost uniquely diploid [200], we used their number of nuclei as a readout of their ploidy. Unbiased clustering revealed 7 molecularly different clusters, classified as WM or conduction system based on previously known markers (Fig. 3.10a, e, Table 3.2). The proportion of MNDCMs is homogeneous in all clusters, with the exception of clusters C3b and C4, in which more than 90% of cells are mononucleated (Fig. 3.10b, c). Based on their transcriptomic signature, we could annotate these clusters as Purkinje cells, known to be mostly mononucleated (Fig. 3.10d) [192]. These results support that our scRNA-seq approach provided a high-quality transcriptomic characterization of mouse ventricular CM subtypes and could not identify a specific cluster of MDCM in the WM, but rather three molecularly different WM clusters with a mixture of diploid and polyploid CM.

3.2.5 MNDCMs of the WM share a unique transcriptomic footprint

We hypothesized that, even if there are other reasons that explain the existence of different CM clusters in the WM, the high resolution of our data should allow the identification of a WM MNCDM-specific gene signature, provided it exists. We found 104 differentially expressed genes (FDR-adjusted p-value < 0.05) between MNDCM and binucleated CMs of the WM (Fig. 3.11a, b). Interestingly, several genes characteristic of embryonic CMs (e.g., *Myl7*, *Myl4* and *Nppa* [161]) and genes associated with decreased contractility and calcium cycling (E.g., *Sln* [201]) were overexpressed in MNDCM (Fig. 3.11c). These results suggest that the WM MNDCM population has a more immature transcriptional footprint than their binucleated counterparts.

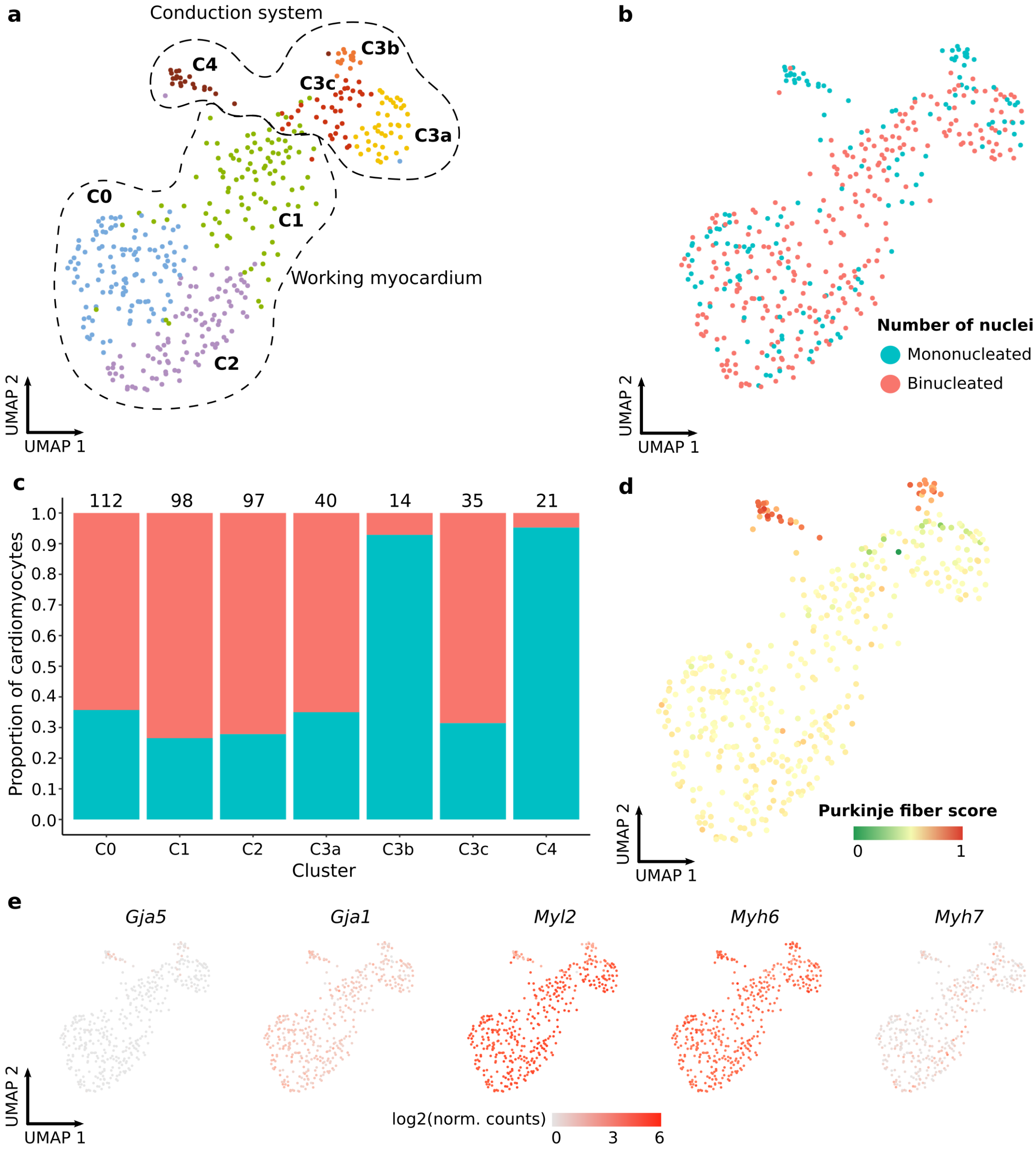


Figure 3.10: Ploidy does not define the molecularly distinct subpopulations of CMs in the adult mouse WM. a) UMAP visualization of adult mouse ventricular CMs colored by cluster. **b)** UMAP visualization of adult mouse ventricular CMs colored by number of nuclei. **c)** Bar plot of the proportion of mononucleated and binucleated CMs per cluster. The numbers on top of the bar indicate the total number of cells in each cluster. **d)** UMAP visualization of adult mouse ventricular CMs colored by conduction system score (Table 3.2). **e)** UMAP visualization of adult mouse CMs colored by expression of ventricular and atrial marker genes (Table 3.2).

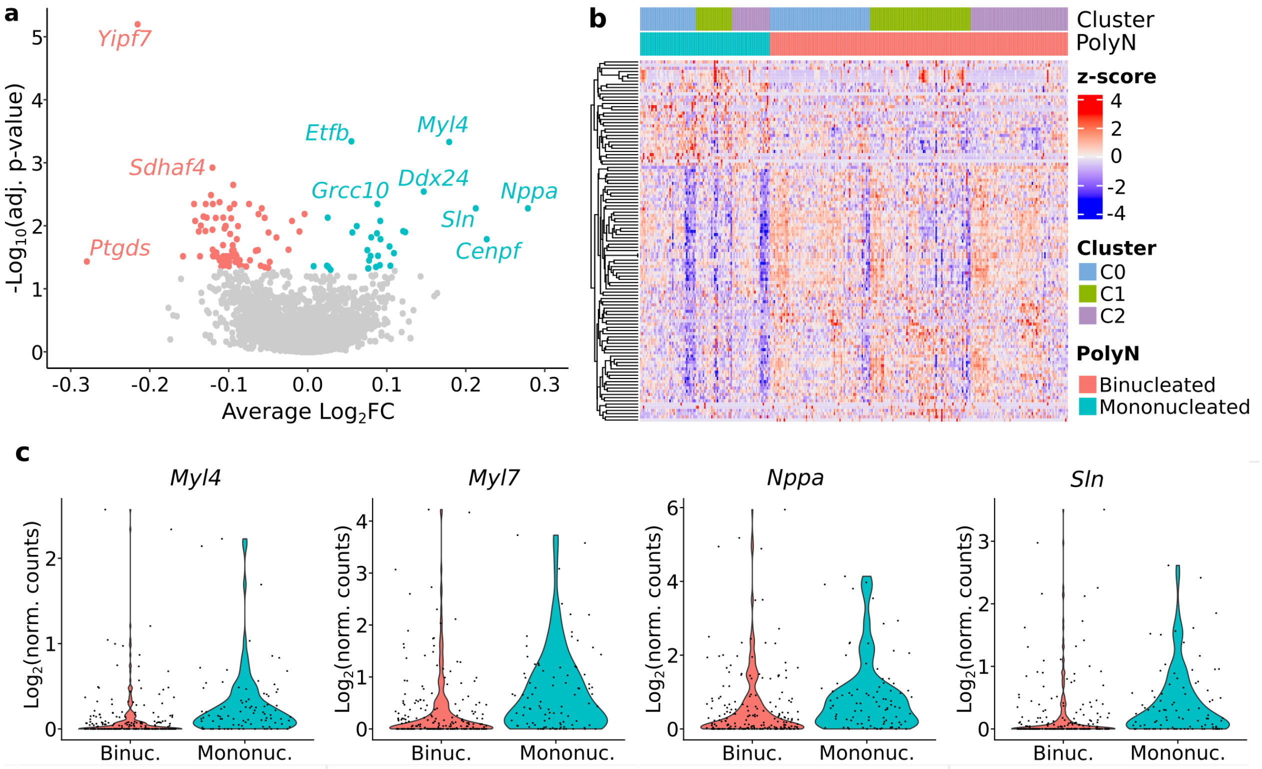


Figure 3.11: Mononucleated and binucleated CMs of the adult WM have distinct expression profiles. **a)** Volcano plot showing the differential expression between mononucleated and binucleated CMs of the adult WM. **b)** Heatmap visualization of the scaled expression of DEGs in adult WM CMs. **c)** Violin plot of the normalized expression of four mononucleated markers in mono- and binucleated WM CMs.

3.2.6 *Hey2* is a negative regulator of MNDCMs in the adult mouse WM

While the MNDCM signature genes might serve as useful markers, they are effector genes and hence might not allow to modify the CM identity. We leveraged IPA [111] and the Ingenuity® Knowledge Base (QIAGEN Inc.) to predict which TF(s) could regulate these signature genes and potentially determine the CM ploidy. We identified 8 significant TFs (Bonferroni-adjusted p-value < 0.05) (Fig. 3.12a). From these, *Hey2* was the top candidate, with an absolute z-score twice as large as that of the other candidates and an adjusted p-value 3 orders of magnitude smaller than that of the other candidates (z-score = -1.994, adj. p-value = 1.94e-06). We applied the Mechanistic Network algorithm (QIAGEN Inc.) to find a causal GRN that provides mechanistic hypotheses about how *Hey2* causes the observed gene expression changes. A single causal GRN in which *Hey2* inhibits the expression of *Sln*, *Myl7*, *Myl4* and *Nppa* directly and through the inhibition of *Gata4* was predicted (Fig. 3.12b). These results show how beneficial the incorporation of prior biological knowledge for GRN reconstruction is, as *Hey2* expression was not detected in the data (Fig. 3.12c).

To validate the association of *Hey2* expression with CM polyploidy, we used the Hey2^{CreERT2} mouse line in which, upon tamoxifen induction, the *Hey2*$^{+}$ CM lineage expresses the mEGFP reporter, while the *Hey2*$^{-}$ lineage expresses mTomato. We assessed the number of nuclei of isolated adult ventricular CMs in these two groups and found a fifteen-fold reduction in the percentage of mononucleated ventricular CMs in the *Hey2*$^+$ lineage, with only 1 mononucleated CM being identified in the three biological replicates examined. Altogether, these results demonstrate the association between postnatal *Hey2* expression and CM polyploidization (Fig. 3.12d-f).

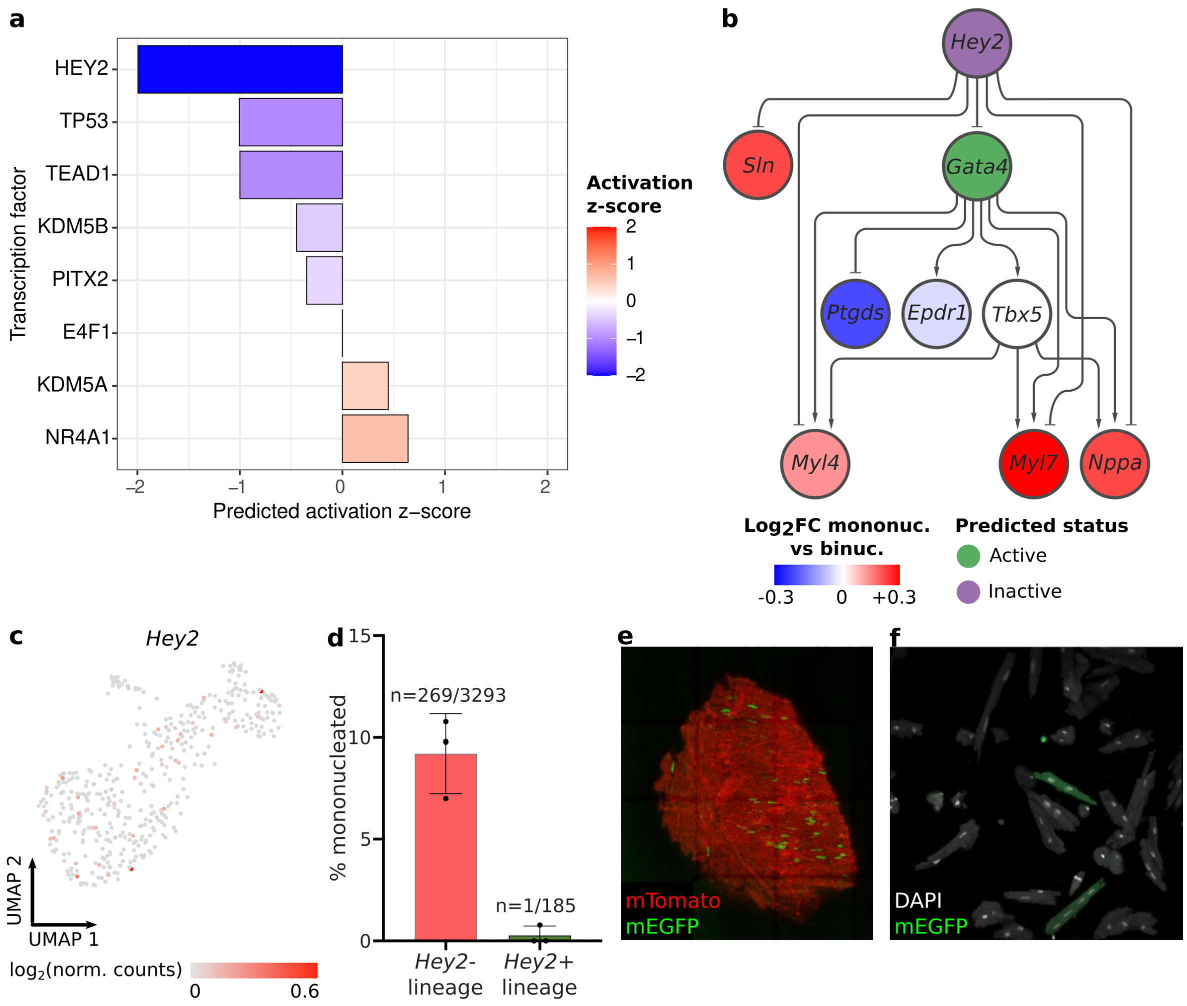


Figure 3.12: *Hey2* is a negative regulator of mononucleated CMs in the adult WM. a) Barplot of the TF predicted activity, expressed as z-scores. Only TFs with a Bonferroni-adjusted p-value < 0.05 are shown. **b)** Predicted GRN controlled by *Hey2*. Node color indicates either expression $\log_2$FC in mononucleated vs. binucleated WM CMs (blue-red scale) or the predicted activation z-score (purple-green scale). **c)** UMAP visualization of the expression level of *Hey2* in adult mouse CMs. **d)** Bar plot of the percentage of mononucleated ventricular CMs in the *Hey2*$^-$ (mTomato, red) and *Hey2*$^+$(mEGFP, green) lineages. The number of mononucleated CMs and total number of quantified CMs are shown on top of the bars. Each dot represents a biological replicate. **e)** Whole mount image of the adult heart of Hey2^{CreERT2} mice upon tamoxifen administration. **f)** Fluorescence image of purified CMs of an adult Hey2^{CreERT2} mouse heart after tamoxifen administration.

3.2.7 The MNDCM signature is not yet detectable at P7

CM polyploidization is known to occur during the first 10 postnatal days of life [199]. To explore whether the MNDCM signature is detectable during this time, the same scRNA-seq experiment was performed on P7 ventricular CMs. Through unbiased clustering, 5 molecularly distinct CM subtypes were identified (Fig. 3.13a, e). All clusters had a homogeneous contribution of mononucleated CMs (Fig. 3.13b, d) and no cardiac conduction system cluster could be identified based on their genetic signatures, probably because it was not enriched during mononucleated CM selection due to the higher percentage of mononuclear CM at P7 (Fig. 3.13c). Differential expression analysis found 14 differentially expressed genes (FDR-adjusted p-value < 0.05), but their absolute $\log_2$FC was very small in all cases (Fig. 3.13f). When analyzing the MNDCM marker genes in detail, we could not see any differences in the expression of *Hey2*, *Myl4*, *Myl7*, *Nppa*, nor *Sln* between mono- and binucleated CMs (Fig. 3.13g). These results suggest that, at P7, the gene expression profile between MNDCM and CM in the process of polyploidization is virtually identical.

3.3 Discussion

This chapter provides evidence of the role of *Hey2* as a roadblock for regeneration in two different vertebrate model species. First, we leveraged an axolotl heart atlas with single-nuclei resolution to reconstruct the GRN controlling ventricular CM function, proliferation, and response to injury. By systematic perturbation of all TFs in this GRN, we discovered *Hey2* as a critical antagonist for the initiation of the wound response in axolotl CMs. Secondly, we interrogated the transcriptomic footprint of mono- and binucleated mouse ventricular CMs at P7 and adulthood using a high-quality scRNA-seq in which the number of nuclei per cell was manually recorded. We identified a MNDCM-specific gene expression program in the WM and suggested four candidate markers for the identification and tracking of MNDCMs. By using a data base of manually curated transcriptional regulation, we identified *Hey2* as the potential master regulator of CM polyploidization and validated the anti-correlation between *Hey2* expression and abundance of mononucleated CMs *in vivo*. The two lines of evidence that gave rise to these results were independently followed, and in both cases *Hey2* was independently found as the top candidate. This potential evolutionary conservation reinforces the significance of these results and the role of *Hey2* as a key player in CM renewal (Fig. 3.14).

MNDCMs have been shown to be the main source of new CMs after injury [73], and the correlation between MNDCM abundance and positive response to injury supports this hypothesis [198]. Studies in which CM ploidy is experimentally manipulated *in vivo* are strong evidence of the need for a high diploid:polyploid CM ratio to achieve a better functional recovery [70]. While several hypotheses about the cause of polyploidization (e.g, thyroid hormones, oxidative stress) and genes associated with it (e.g., *Tnni3k*) have been proposed [200, 202], a mechanistic understanding how genetic information is translated into polyploidization needs transcriptional regulators. *Hey2* is a bHLH

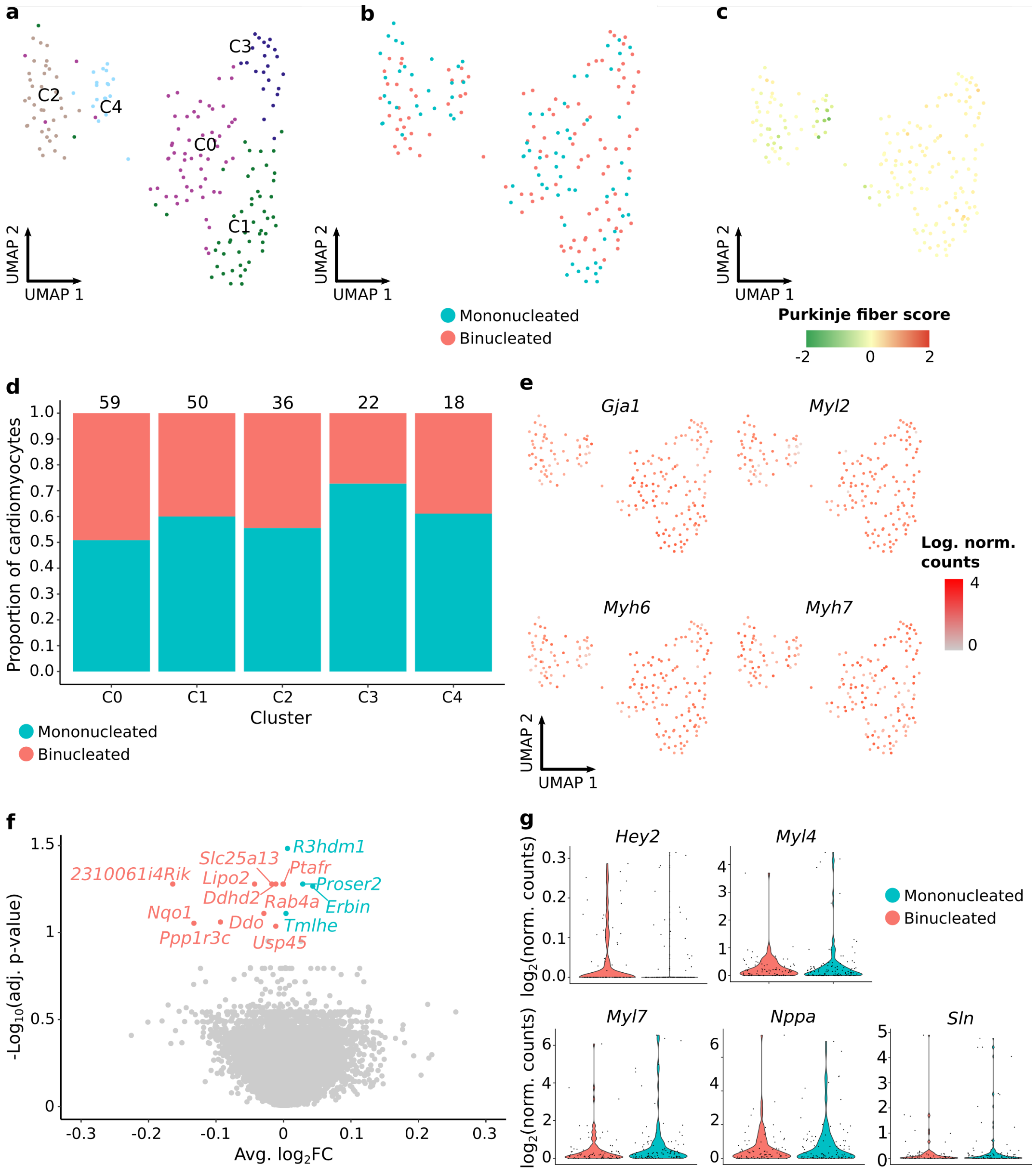


Figure 3.13: P7 ventricular CMs do not show differences in the activity level of *Hey2*. a) UMAP visualization of P7 mouse ventricular CMs colored by cluster. **b)** UMAP visualization of P7 mouse ventricular CMs colored by number of nuclei. **c)** UMAP visualization of P7 mouse ventricular CMs colored by conduction system score (Table 3.2). **d)** Barplot of the proportion of mononucleated and binucleated CMs per cluster. The numbers on top of the bar indicate the total number of cells in each cluster. **e)** UMAP visualization of P7 mouse CM colored by expression of ventricular and atrial marker genes. **f)** Volcano plot showing the differential expression between mononucleated and binucleated P7 ventricular CM. **g)** Violin plot of the normalized expression of the four *Hey2* target genes in mono- and binucleated P7 ventricular CM.

transcriptional repressor and a direct target of the Notch pathway that regulates expansion of the cardiac progenitor pool, atrioventricular boundary formation and myocardial compaction [203, 204]. Lack of *Hey2* function is associated with a myriad of heart defects, also in humans [205], and *Hey2* loss-of-function results in abnormal expression of atrial, or embryonic, genes in the ventricle [203] and leads to perinatal lethality in homozygosity [197]. Additionally, overexpression of the zebrafish *Hey2* homolog *grl* impairs regeneration, although whether this affects CM ploidy remains to be investigated [206]. The fact that *Hey2* was independently identified in several vertebrate species is promising, because it suggests that, despite the variability in polyploidization mechanisms observed in different mammals [198], *Hey2* could be a conserved upstream regulator of this process.

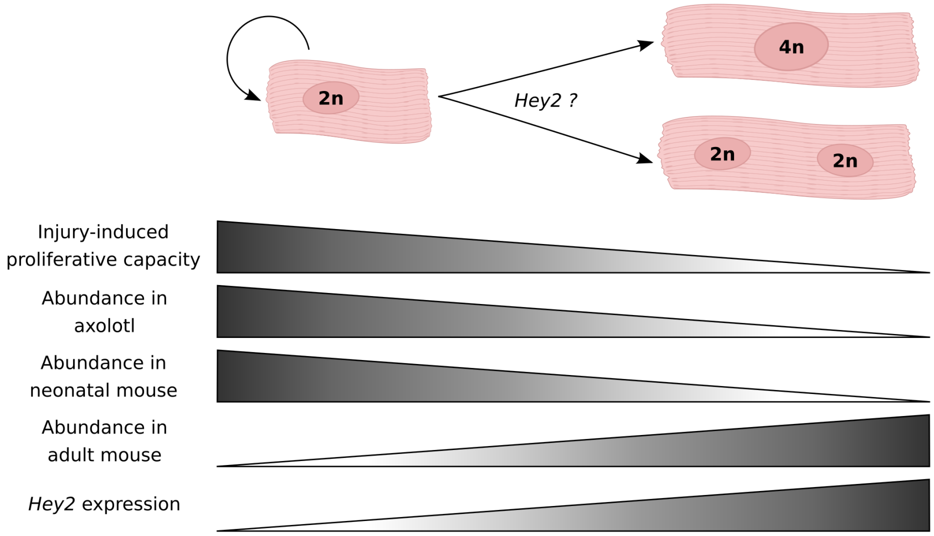


Figure 3.14: Relationship between CM ploidy, *Hey2* expression and regenerative capacity. Vertebrates able to regenerate their hearts, such as the zebrafish or neonatal mouse, have ventricles almost exclusively composed of MNDCMs. On the other hand, non-regenerative mammals have high levels of CM polyploidization. MNDCMs have a higher proliferative capacity upon injury and could give rise to new MNDCMs or polyploid CMs with two diploid or one plyploid nucleus. *Hey2* is a transcriptional repressor with a predicted higher activity in mouse binucleated CMs compared to their diploid counterparts, and its inactivation may be necessary for axolotl border zone CMs to undergo proliferation.

This discovery was only possible thanks to methodological innovations and careful experimental design. Regarding the axolotl data, it is the first single-nuclei multi-omics and spatial atlas of the axolotl heart in basal conditions and at 4 timepoints in the regenerative process. This chapter only leveraged the tip of the data iceberg, which should be further interrogated to uncover inter-cellular relationships and how they influence and change during regeneration. In this regard, deconvolution tools play a critical role in decoding spatial transcriptomics data in a cheap, fast, and reliable manner, as most spatial technologies do not yet reach single-cell resolution and those which do are still expensive and of limited access. Particularly, the use of deep neural networks to solve the deconvolution problem is a computationally feasible way to utilize the non-linear relationships in gene expression data and has shown a good performance compared with other state-of-the-art alternatives [25]. Regarding the mouse data set, previous attempts at identifying gene markers of MNDCMs have failed due to a suboptimal experimental design. Previous approaches [207, 208] have used FACS or microfluidic devices to obtain single-cell suspensions, which can damage large cells like CMs, and nuclear dyes for nuclei quantification, which interfere with the scRNA-seq library preparation procedures. Another evidence of the limitations of these approaches is that none of them identified a Purkinje-fiber genetic signature, despite the presence of those cells in the ventricle. The manual CM isolation and manual nuclei identification based on the expression of a fluorescent nuclear reporter combined with the higher sequencing quality provided by the SMART-seq2 platform have been indispensable for the success of this project and have allowed to identify for the first time genetic markers that distinguish diploid and polyploid mouse CMs, although these should be extensively validated *in vitro* and *in vivo*. Regarding GRN inference, two very different approaches were used in this chapter: a data-driven approach for the axolotl data and a literature-driven approach for the mouse data. The key reasons behind this decision were (1) availability of snATAC-seq data in the axolotl and (2) larger sample size (number of cells) in the axolotl data. The axolotl GRN is a good example of the extreme power of single cell multi-omics to uncover gene regulatory relationships, even in organisms with a complicated genomic architecture. However, manually curated GRN scaffolds should not be undervalued, as they provide an accurate starting point for GRN inference [109], provided that enough information is available for the species of interest and most likely TFs to be involved in the biological of process under study.

Many open questions remain to be answered. In the axolotl data, a particularly intriguing question is why the *Cav3.1*$^+$ population is the only contributor to the injury trajectory. The lack of contribution from *Robo2*$^+$ CMs to the injury lineage might be explained by the cryoinjury model, as it only affects the outer myocardial layer that is enriched in *Cav3.1*$^+$ CMs. Whether these two populations arise from common or different progenitors can only be truthfully answered with lineage-tracing experiments. Another point for further exploration is the consequence of *Hey2* overexpression in axolotl, as we currently don't know whether *Hey2*-overexpressing CMs completely fail to enter the cell cycle or suffer arrest at any of its points. In combination with *Hey2*-silencing studies in mouse, this data can shed light onto whether *Hey2* is able to revert CM ploidy, or whether its effect in polyploidization is limited to the first postnatal week in mice. Experimental work to better understand the role of *Hey2* in physiological and regenerative conditions is ongoing: larger sample sizes and a complete and multi-angular evaluation of heart function (e.g., ejection fraction, scar area, CM ploidy and nucleation…) are fundamental to understand the therapeutic potential of this candidate and whether its inactivation could have a pro-regenerative or protective effect.

4 Network modeling of the pro-regenerative effect of *Erbb2*, *Myc* and *Yap1* on CMs

The adult mammalian heart lacks endogenous regenerative ability, but lower vertebrate species and at least some newborn mammals can heal their injured hearts by activating endogenous pro-regenerative regulators. These, that include *Yap1*, *Erbb2* and *Myc*, rejuvenate the phenotype of resident CMs by inducing the disassembly of the contractile machinery and cell cycle re-entry, without irreversibly affecting contraction. While extensive efforts have been put into characterizing these factors and their ability to induce regeneration in the adult mammalian heart, there is a lack of understanding of the regulatory wiring between them. This chapter uses transcriptomics coupled with metabolic labelling, GoM models and Boolean network models to obtain high-quality transcriptomic footprints activated by these factors, identify shared and unique gene expression changes and model the core GRN controlling these programs. Our results show that, while all three perturbations activate proliferative and dedifferentiation responses, the effect of *Erbb2* and *Myc* is highly similar. The Boolean models identified a core GRN well correlated with a CM maturation gene expression profile and identified candidate pro-regenerative factors. Ultimately, our models provide insight into how CMs regulate their proliferation and captures mechanistic relationships between key transcriptional regulators of CM identity.

4.1 Materials and methods

4.1.1 Experimental procedures

All experimental procedures were performed by Cristina Villa del Campo, PhD and Jorge Peña and are briefly described here to provide the necessary background for the in silico analysis of the gathered data. The scripts used for data analysis are available at: `https://github.com/iriverog/SLAMseq-GRN`.

Mouse husbandry

P1 C57BL/6 male and female mice were used in this project. Animals were bred and maintained in the animal facilities of the CNIC in Madrid, Spain. All experiments were conducted in accordance with the CNIC Ethics Committee, Spanish laws, and the EU Directive 2010/63/EU for the use of animals in research. All mouse experiments were approved by the CNIC and the area of “Protección Animal” of the Community of Madrid with reference PROEX 144.1/21.

Isolation of neonatal mouse CMs

WT mice were sacrificed by decapitation at P1. After euthanasia, hearts were harvested, atria and excess blood were removed, and ventricular CMs were purified by gentleMACS Octo Dissociator with heaters (Miltenyi Biotec) following manufacturer instructions. The digested tissue was filtered to remove cellular debris using 70µm (Fisherbrand) and resuspended in P/S media supplemented with 5% inactive FBS, 2 mM of L-glutamine and 1 µM vitamin B12. Contaminant erythrocytes were further removed using RBC lysis buffer (Thermofisher Scientific). The resulting cell suspension was pre-plated on plastic-coated cell culture plates to achieve high CM purity, as these cells do not easily attach on plastic surfaces and remain in the supernatant. The CM-enriched supernatant was split between p24 corning primaria plates (Corning), used for SLAM-seq, and p18 µSlide ibiTreat (Ibidi) plates, used for validation using immunocytofluorescence.

Activation of pro-regenerative genes by transduction with adenoviral vectors

CM-enriched cultures with approximately 0.2 x 10^6 cells per well were transduced with AAV9-vectors carrying the pro-regenerative gene of interest (Table 4.1) using a MOI (multiplicity of infection) of 5 x 10^5. Transductions were carried out for 44h and plates for SLAM-seq and microscopy controls were simultaneously transduced. Three replicates were used per condition.

Adeno-associated viruses (AAVs) are small, ssDNA viruses that depend on co-infection with helper viruses to complete their life cycle [209]. AAVs are commonly used as viral vectors for gene delivery because they achieve long-term transgene expression in non-dividing cells, their genome can be easily manipulated using standard plasmid cloning and other molecular biology techniques, and they present low immunogenicity and cytotoxicity [209]. Different AAV serotypes recognize different cell

Table 4.1: Information about gene expression cassettes used for each pro-regenerative treatment.

Condition	Information about the transgene
caErbb2	It is a constitutively active form of the ERBB2 receptor. Transient *caErbb2* overexpression after AMI in juvenile and adult hearts promotes regeneration [83]. The construct for *caErbb2* overexpression was designed by Dr. Miguel Torres' group.
Gfp	*AcGfp* was used as control in all transduction experiments. The construct was provided by the CNIC Viral Vectors Unit.
Myc	The $iMOS^{T1\text{-}Myc}$ construct designed by Claveria *et al.* [210] was used. This construct achieves a *Myc* expression level similar to that of a wild type allele. This construct was designed in the group of Dr. Miguel Torres.
Yap5sa	*Yap5sa* is a constitutively active form of the Hippo pathway effector *Yap1* that contains 5 point mutations that prevent its phosphorylation and subsequent degradation [211]. *Yap5sa* overexpression has been shown to induce CM proliferation and rejuvenation in the adult heart [79]. The construct for *Yap5sa* overexpression was kindly provided by Prof. Eldad Tzahor.

receptors and therefore display tissue and cell-type tropism, with serotypes AAV-6 and AAV-9 being highly specific for CMs [209]. While empirical observations show that AAV-6 and AAV-9 respectively have higher efficiency for *in vitro* and *in vivo* CM transduction, in this project we transduced in vitro CM cultures with AAV-9 vectors to facilitate future *in vivo* translation of results. All AAV-9 vectors used in this project were synthesized by the CNIC Viral Vectors Unit. Their genome contains a simple gene expression cassette with the CMV promoter, the sequence of the transgene of interest and the BGH or SV40 polyadenylation signal as termination signal. No fluorescent reporters were introduced due to the size limit of 5 kb that applies to AAV-9 genomes [209]. Table 4.1 summarizes the pro-regenerative genes that were tested in this project.

The effect of the transductions on CM proliferation were validated by immunocytofluorescence. Briefly, the CM cultures were fixated in PFA 2% for 20 minutes at RT. Samples were then washed with PBS and permeabilized with 0.1% PBS-TritonX for 10 minutes at RT. Samples were blocked with TNB during 1 hour at RT and incubated with primary antibodies in TNB overnight at 4°C. After three PBS washes, samples were incubated with secondary antibodies in 0.1% PBS-TritonX and DAPI.

4.1.2 SLAM-seq

SLAM-seq (thiol(SH)-linked alkylation for the metabolic sequencing of RNA) is a modification of standard RNA-seq in which the incorporation of the nucleotide analog 4-thiouridine (s4U) in mRNA molecules is bioinformatically detected to quantify the gene expression before and during the s4U pulse [212]. This detection is based on a change in the base pairing capacity of s4U upon thiol-specific alkylation with iodoacetamide (IAA), that introduces T-to-C and A-to-G mismatches during the library preparation RT-PCR step (Fig. 4.1). These mismatches can be bioinformatically detected and quantified to estimate the transcripts synthesized during the s4U pulse and distinguish them from pre-existing mRNAs.

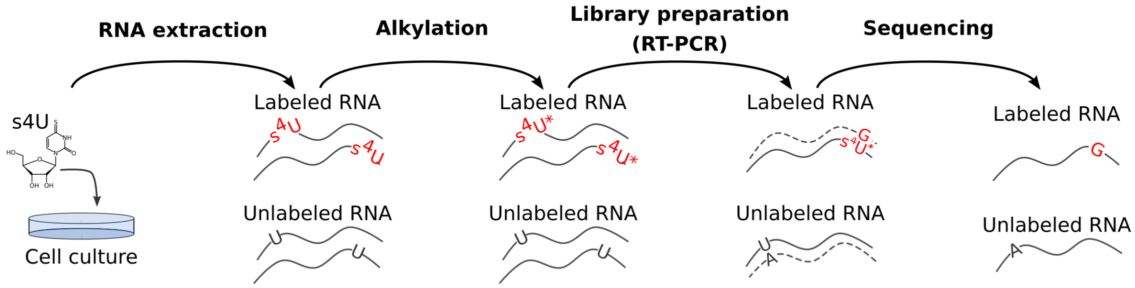


Figure 4.1: Molecular basis of SLAM-seq. The nucleotide analog s4U is uptaken by the cell and incorporated in newly synthesized RNA. After purification, RNA is subjected to thiol-specific alkylation with IAA. The modified s4U (s4U*) changes its base pairing capacity during the RT-PCR step of library preparation: while U normally pairs with A, s4U* pairs with G. This introduces T-to-C and A-to-G mismatches that are bioinformatically detected and quantified.

Fig. 4.2 shows the complete experimental pipeline used in this project and how AAV-9 perturbations were coupled with SLAM-seq to distinguish the transcriptional footprint of the perturbations from pre-existing mRNAs.

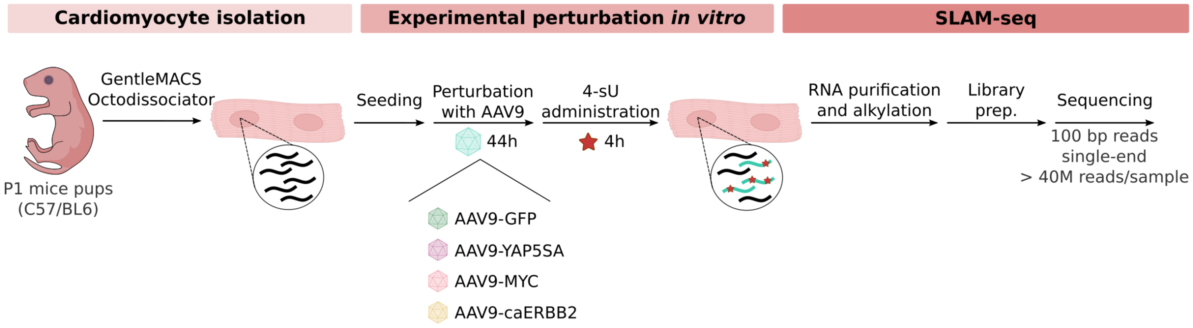


Figure 4.2: Experimental pipeline to combine AAV-9-based CM perturbations with SLAM-seq. Neonatal mouse CMs were extracted from P1 C57BL/6 hearts. After seeding, the pro-regenerative gene of interest was overexpressed by transducing the cells with an AAV-9 vector for 44h. Then, new media containing s4U was administered. Afterwards, the RNA was purified and subjected to thiol-specific alkylation with IAA before undergoing standard library preparation and sequencing.

s4U administration: After 44h of transduction, the media of CM-enriched cultures was changed with new media containing 100 µM s4U. The s4U treatment lasted 4h, with media replacement after 2h to ensure that enough mismatches were introduced for bioinformatic detection of newly synthesized mRNA. This protocol was based on a SLAM-seq protocol on mESCs [213] and optimized for CM. After the s4U treatment, cells were washed with PBS and mechanically collected.

RNA purification: RNA purification was performed using Qiagen RNeasy kit (Qiagen) following manufacturer instructions.

Thiol-specific alkylation with IAA: s4U alkylation with IAA was performed as previously described [213].

Library preparation and sequencing: cDNA generation, pre-amplification and library preparation were performed following manufacturer instructions. Libraries were sequenced on an NextSeq 2000 sequencing platform (Illumina) using single-end 100 nucleotide long reads and aiming for a total of 40 million reads per sample. Fastq files containing reads for each sample were obtained by demultiplexing raw sequencing files using bcl2fastq2 (Illumina). These steps were performed by the CNIC Genomics unit.

4.1.3 Transcriptomic analysis of SLAM-seq

Figure 4.3 summarizes the bioinformatic pipeline used for pre-processing, alignment, and quantification of total and newly synthesized gene expression.

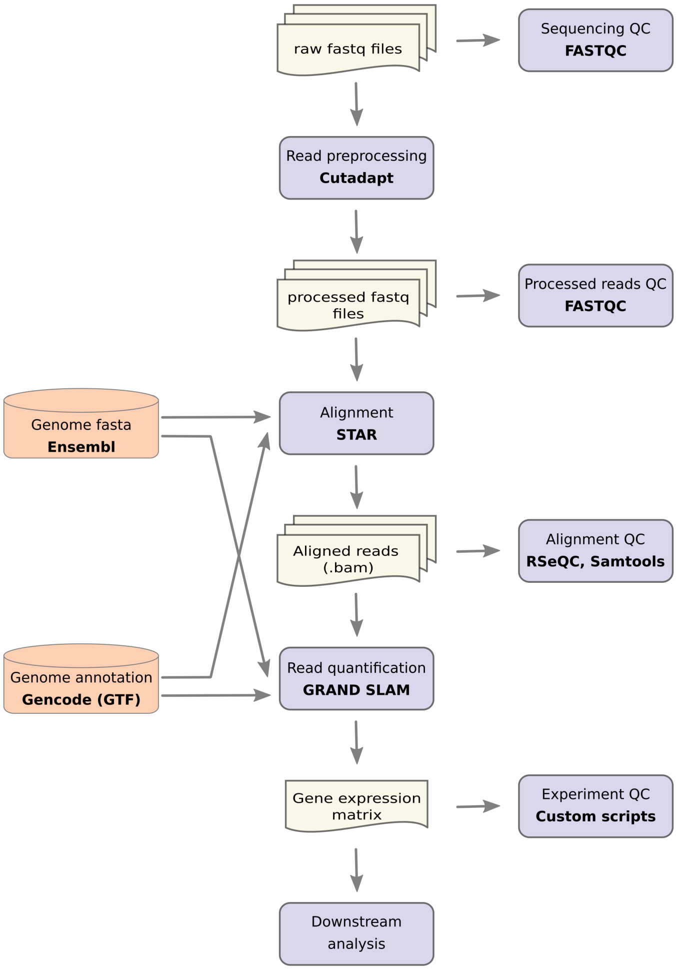


Figure 4.3: Bioinformatic pipeline for the pre-processing and quantification of gene expression from SLAM-seq reads. Light yellow boxes indicate files, lilac boxes indicate algorithms and orange cylinders indicate databases.

Quality control and read alignment

Quality control of raw and trimmed reads was performed using FastQC v0.11.9. 3' adapters matching "AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC" or "GATCGGAAGAGCACACGTCTGAACTCCAGTCAC" were trimmed using cutadapt v3.5 [187]. After adapter trimming, reads shorter than 30 nucleotides were removed. Reads were mapped to the mouse genome version GRCm39 with Ensembl106 annotation using STAR v2.7.10a [214]. STAR was run with default settings, adding –outSAMattributes "NM" "MD" "NH" and –alignEndsType "EndToEnd" as they are required by GRAND-SLAM [215] to estimate the fraction of newly transcribed mRNA. These parameters record the number of mismatches in each read (NM), encode the detected mismatches in the SAM file (MD), record the number of mappings found for each read (NH) and align the reads without soft-clipping (–alignEndsType "EndToEnd" flag). Samtools v1.15 [216] was used to index the alignment files and RSeQC v4.0.0 [217] was used to visualize the distribution of mismatches along the read length and distribution of reads along the gene body.

Estimation of the fraction of newly transcribed mRNA

The quantification of gene expression is further complicated in SLAM-seq compared to standard RNA-seq, as the relatively infrequent incorporation events and significant sequencing error rates make it challenging to differentiate pre-existing mRNAs from newly transcribed mRNAs carrying s4U-induced mismatches. Even in long (24h) labeling experiments with high s4U concentrations (100 µM), no more than 1 in 40 (2.5%) U are substituted by s4U, according to empirical observations [212]. Despite current NGS error rates having dropped below 0.1% [215], it is difficult to decide with certainty whether an individual read with T-to-C mismatch(es) comes from a new or old mRNA. We used the Bayesian approach implemented by GRAND-SLAM v2.0.5f (available at `https://erhard-lab.de/software`) to estimate the fraction of newly transcribed mRNA for each gene and quantify the uncertainty in the estimation [215] (Fig. 4.4).

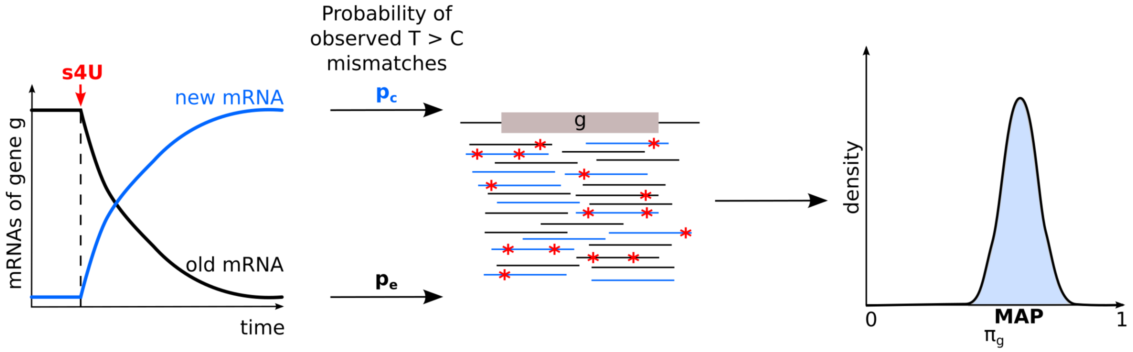


Figure 4.4: Estimation of the fraction of newly transcribed mRNA for each gene using GRAND-SLAM. GRAND-SLAM assumes that, after incubation with s4U, observed T-to-C conversions can be either due to s4U incorporation (occurring with probability p_c) or due to sequencing errors (occurring with probability p_e). GRAND-SLAM uses binomial models to model this process, and Bayesian statistics to estimate the posterior of the fraction of newly transcribed mRNA for each gene (π_g) [215]

GRAND-SLAM uses binomial mixture models to model the observed T-to-C mismatches as a consequence of sequencing errors and/or mismatches introduced during SLAM-seq. Therefore, for a genomic region containing n T within gene g, the number of observed T-to-C mismatches (y) depends on the average mismatch rate on unlabeled RNA (p_e), the average mismatch rate on labeled RNA (p_c), the fraction of newly transcribed RNA for gene g (π_g) and the number of T in the considered genomic region (n) (Eq. 4.1).

$$B(y; p_e, p_c, \pi_g, n) = (1 - \pi_g)B(y : p_e, n) + \pi_g B(y; p_c, n) \tag{4.1}$$

GRAND-SLAM uses a Bayesian framework to estimate gene-specific π_g given the observed T-to-C mismatches and the maximum T-to-C mismatches that could have occurred for each read (Eq. 4.2):

$$f(\pi_g; y, n, p_e, p_c) = \frac{\prod_i P(y; p_c, p_e, \pi_g, n) \cdot b(\pi_g; \alpha, \beta)}{P(y)} \tag{4.2}$$

The prior for π_g is assumed to be a uniform beta prior with hyperparameters $\alpha = \beta = 1$. In order to solve Eq. 4.2, gene-specific p_e and p_c values are estimated as follows:

- While p_e can be estimated from spike-ins or using an unlabeled sample, these estimates are costly (spike-ins) or lead to inaccurate estimations (unlabeled sample) due to the high variability of p_e observed in unlabeled samples [215]. Instead, GRAND-SLAM trains a linear regression model to predict the basal T-to-C error rate from the T-to-A mismatch rate in the same sample, as they have been found to be linearly related [215].
- Regarding p_c, the model assumes that $p_e < p_c$. Under this assumption, there must exist a certain k for each n for which less than 1% of observed reads with $\geq k$ T-to-C mismatches originates from the p_e component. The k value for each n is computed using Eq. (4.3) and (k, n) pairs are only retained if $e_{k,n} < 0.01a_{k,n}$, where $a_{k,n}$ is the number of reads mapped to a genomic region within gene g containing n T and k observed T-to-Cmismatches. The remaining $a_{k,n}$ are used to estimate p_c using an EM algorithm that treats the excluded (k, n) as missing data (Eq.4.4 for the E step and Eq. 4.5 for the M step).

$$e_{k,n} = B(k; n, p_e) \cdot \sum_{k'} a_{k',n} \tag{4.3}$$

$$a_{k,n}^{(t+1)} = \frac{\sum_{(k',n)\notin X} B\left(k; n, p_c^t\right) \cdot a_{k',n}}{\sum_{(k',n)\notin X} B\left(k'; n, p_c^t\right)} \tag{4.4}$$

$$p_c^{(t+1)} = \frac{\sum_{k,n} k a_{k,n}^{(t+1)}}{\sum_{k,n} n a_{k,n}^{(t+1)}} \tag{4.5}$$

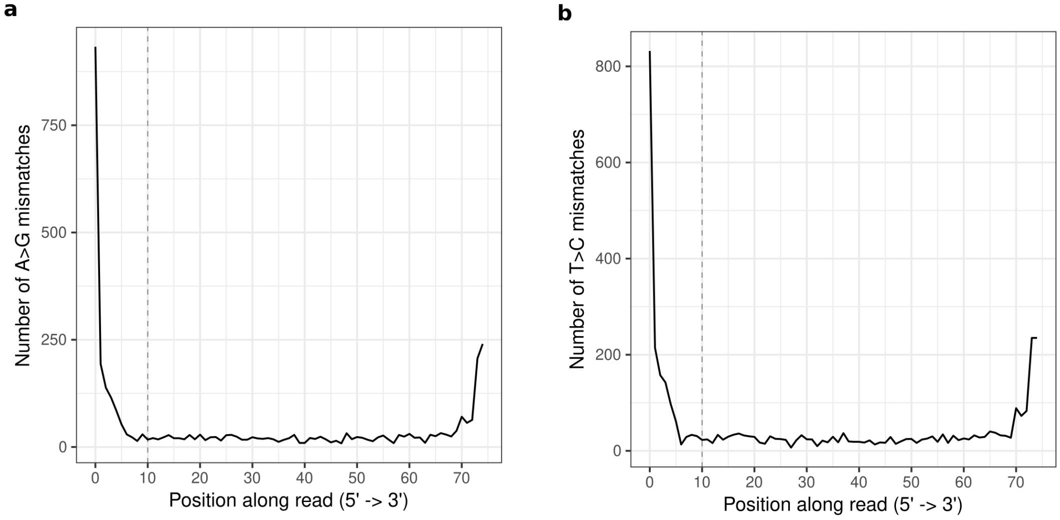


Figure 4.5: Distribution of a) A-to-G and b) T-to-C mismatches along the read length of a sample not treated with s4U. The 5' end of the read accumulates higher number of mismatches that can affect the p_e and p_c estimates calculated by GRAND-SLAM. Therefore, the 10 5'-most nucleotides (indicated by the dashed line) are not considered for estimation.

The output of GRAND-SLAM is the gene-specific π_g posterior, which takes values between 0 and 1. We use the mode of the posterior (MAP) as the fraction of newly synthesized mRNA (π_g) for each gene. The newly synthesized gene expression was obtained by multiplying the total counts recorded for each gene by its π_g MAP (Eq. 4.6).

$$NewExpression_g = TotalCounts_g \cdot MAP(\pi_g) \tag{4.6}$$

GRAND-SLAM was run using Ensembl 106 genome annotation, masking mismatches appearing in more than 30% reads covering those positions as SNPs, and by trimming the 10 5'-terminal nucleotides of each read before π_g. This trim was performed to reduce the error in the estimations, as the 5'-end of RNA-seq reads are frequently noise and accumulate more mismatches than the rest of the read (Fig. 4.5).

Differential expression analysis

Differentially expression analysis was performed using DESeq2 [218]. Before differential expression analysis, genes with less than 10 counts in less than 3 samples were prefiltered out. Scaling factors were calculated for each experimental perturbation and their corresponding GFP controls using the design formula: $\sim Condition$. Independent filtering to optimize the number of significant adjusted p-values was used to filter out non-informative genes and increase the detection power. P-value adjustment was done using FDR correction, and genes with an adjusted p-value < 0.1 were called significant. LFC shrinkage estimates calculated using the apeglm method were used to improve gene ranking and visualization [219]. Normalization of gene expression data for visualization was performed using the variance stabilizing transformation (VST), which produces log2-transformed data normalized with respect to library size [220].

Gene Set Enrichment Analysis (GSEA) for each experimental condition was performed using the ranked $\log_2$FC, the GO Biological Process database [183, 184] and the GSEA implementation in the fgsea R package [185].

Detection of gene expression programs using GoM models

GoM models are a generalization of cluster methods that allow each sample to have a membership in more than one cluster. These models allow to cluster samples based on their gene expression and to identify sets of genes (gene topics) that work together or are co-regulated.

GoMs as implemented by the R package fasttopics v0.6-147 were used in this project [221]. In an experiment in which c_{gn} is the number of reads for gene g in sample n, these models assume that each sample has a proportion of its reads q_{kn} that come from topic k. Each topic is characterized by the vector θ_k that contains the relative expression of each gene g in that topic k (Eq. 4.7, 4.8). These computations are implemented using Poisson negative matrix factorization [222] (Fig. 4.6).

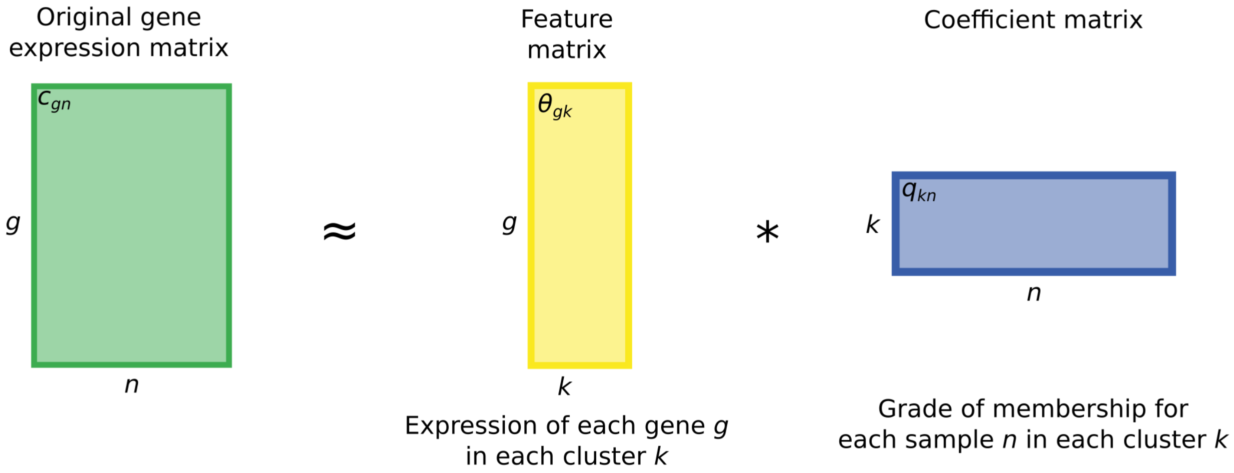


Figure 4.6: Non-negative matrix factorization implementation of GoM models. The original gene expression matrix contains raw gene counts for gene g in sample n (c_{gn}). This matrix is decomposed into two matrices: a feature matrix containing the expression values of each gene in each cluster (θ_{gk}) and a coefficient matrix containing the GoM of each sample in each cluster (q_{kn}).

$$(c_{1n},\ c_{2n},\ ...,\ c_{Gn})\ \sim Multinomial\,(c_n,\ p_{1n},\ p_{2n}, ..., p_{Gn}) \tag{4.7}$$

$$p_{gn}\ :\ = \sum_{k=1}^{k} q_{kn} \cdot \theta_{gk} \tag{4.8}$$

As a result, the GoM estimates a membership proportion for each sample in each topic (q_{kn}, normalized between 0 and 1) and an expression value for each gene and topic (θ_{gk}). The gene topics were biologically interpreted by finding the genes with the largest difference in expression compared to other topics. This was quantified as the θ_{gk} logFC between a given topic and the topic with the most similar expression value for that gene. Genes with s-value < 0.05 were called statistically significant [222]. As the gene contributions to each topic can only take positive values [222], the gene topics were annotated using the ORA implemented by fgsea [185] on the GO Biological Process [183, 184] database using all measured genes in the experiment as universe.

Boolean GRN models and identification of pro-regenerative genes

Fitting GRN models to small sample sizes is complicated due to the imbalance between the number of parameters that must be estimated and the available observations. This problem was approached by reducing the parameter space to a small set of key TF with the largest influence in CM maturation (Fig. 4.7). For this, we combined the CollecTRI database of transcriptional regulation with the scRNA-seq of adult and P7 CMs introduced in Chapter 3. CollecTRI is a PKN of transcriptional regulation created by integrating manually-curated TF – target pairs using literature text-mining, public databases, and manual literature curation [112]. As of May 2024, it contains 1,201 TFs, 5,734 target genes and 64,495 edges. For each TF – target pair, CollecTRI provides a MOR obtained from the literature consensus, or general MOR of the TF, for those TF – targets lacking a MOR in the literature. Therefore, each edge in the CollecTRI PKN is directed and signed, two indispensable elements for fitting Boolean models. The CM scRNA-seq was used to contextualize the CollecTRI PKN to obtain a CM-specific GRN scaffold that focuses on genes and edges relevant for CMs. This contextualization was done using GENIE3 and the scRNA-seq of adult and P7 mouse CMs introduced in Chapter 3. GENIE3 predicts target gene expression pattern from the expression pattern of regulatory genes using random forests [19]. For each target gene, an ensemble of 1,000 trees using the TFs regulating that gene in the CollecTRI database as predictor variables were fit to the data. Each TF – target pair was assigned a weight between 0 and 1, representative of the amount of the target expression variance explained by the expression of each TF. The algorithm was run 100 times with default parameters and the average weight per edge was selected. To identify statistically significant edges, a null distribution of TF – target edge weights was obtained by running GENIE3 on 100 randomizations of the network. An empirical p-value was calculated for each observed weight, and those edges with p-value < 0.05 were retained. The GRN scaffold was visualized and analyzed using Cytoscape v3.9.0 [180]. The fast greedy algorithm of the Cytoscape clusterMaker plug-in [152] was used to find communities within the GRN, which were associated with biological processes using fgsea [185] ORA on the GO Biological Process database [183, 184], using all genes in the GO Biological Process database as universe. For further modeling, the TFs within the network largest connected component showing an absolute Log_2FC larger than 0.25 in at least one experiment were selected. This rendered a GRN consisting of 29 TF and 77 edges.

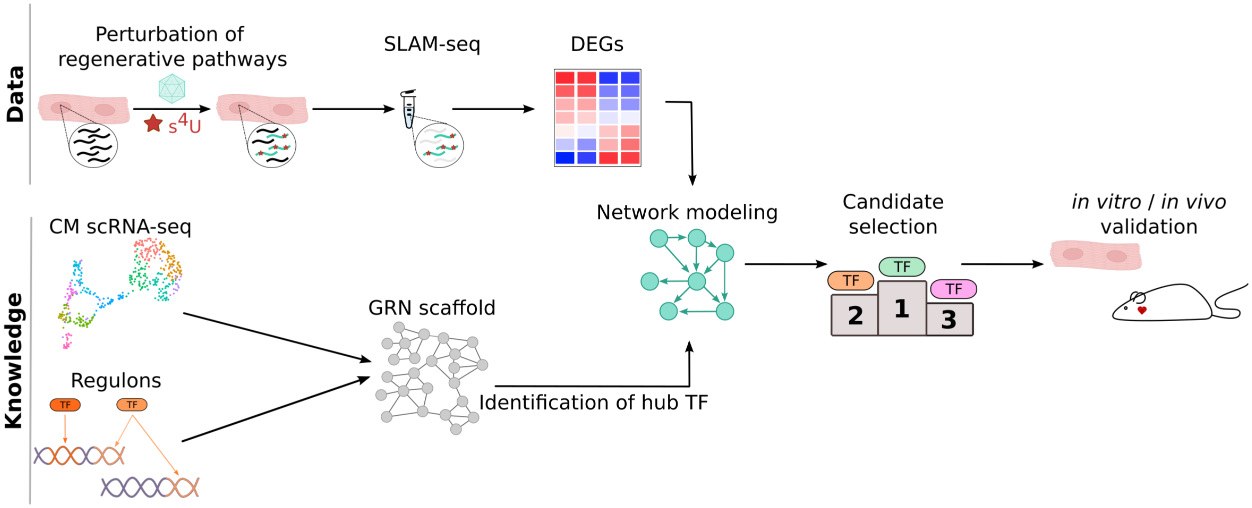


Figure 4.7: Pipeline for network modeling and identification of pro-regenerative candidates. A PKN obtained from the CollecTRI regulons [112] was contextualized using scRNA-seq data and random forests to obtain a GRN scaffold from which the key TF in the CM GRN were selected. These were subjected to Boolean modeling to identify network attractors that were compared with the SLAM-seq data. *In silico* systematic network perturbations were performed to rank the TFs based on their pro-regenerative potential for *in vitro* and/or *in vivo* validation.

Due to the lack of a temporal dimension in the data, a Boolean approach was used to qualitatively model the GRN and identify pro-regenerative candidates. The boolsim algorithm (v1.2.0) [18] was used for modeling. In a Boolean GRN model, each gene can be in an active (1) or inactive (0) state. Each gene has a Boolean function associated to it that summarizes the activating and inhibiting inputs (edges in the network) affecting that node. The Boolean function is a logic combination of the input node states, mainly characterized by AND, NOT and OR relationships. For modeling, the GRN is randomly initialized, and the logic rules are applied to update the current GRN status. This is repeated several times to simulate the network behavior over time. The update can be synchronous (all genes are simultaneously updated) or asynchronous (in each iteration a random gene is updated, with all genes having the same probability of being updated) (Fig. 4.8). Here we used asynchronous updating as it better reflects the stochasticity of biological systems [18]. As a result of this process, network attractors are found. Attractors are a stable state or set of states reached by a Boolean network after a certain number of iterations. While being theoretical, attractors represent potential stable behaviors or phenotypes of the biological system being studied [18]. Overexpression and KO perturbations were simulated on the Boolean GRN model by fixing the expression of the perturbed gene to 1 or 0, respectively. A total of 58 individual TF perturbations as well as 4 double perturbations (*Myc* + *Mycn* and *Junb* + *Fos* KO and overexpression) were simulated. For each perturbation, boolsim finds the attractors of the unperturbed (basal) and perturbed network. Then, it analyzes the forward reachability from basal network attractors to perturbed network attractors.

The basal and perturbed network attractors were associated to the gene topics by dichotomizing the topic Log_2FC for each gene (1 if $\text{Log}_2\text{FC} > 0$, 0 otherwise) and calculating the area under the receiver operating characteristic curve (AUROC). Pro-regenerative candidates were selected by identifying the GRN perturbations that caused a transition from a “CM maturation” attractor to a “cell cycle” or “CM dedifferentiation” attractor. Cytoscape v3.9.0 was used for network visualization [180].

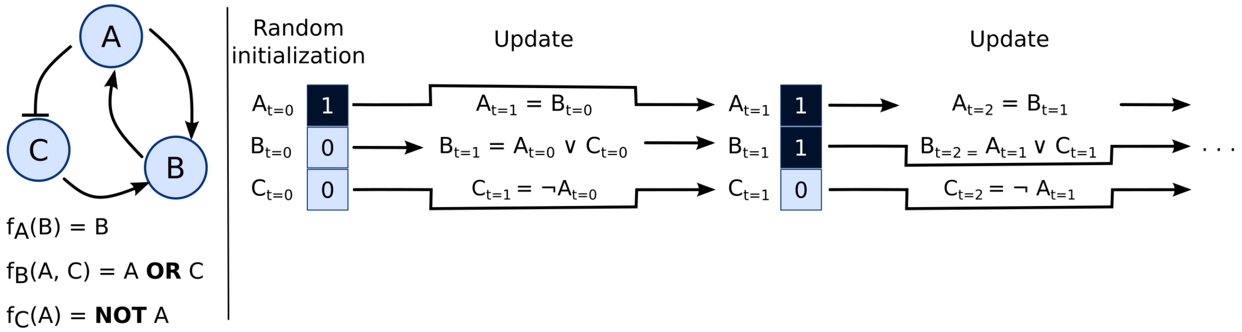


Figure 4.8: Example of a GRN Boolean model. Each gene has a Boolean function containing the logic rules for its activation. During modeling, the network is randomly initialized. The network is updated for a number of iterations, and in each of these, one randomly selected gene is updated according to its activation function, while the rest of genes keep their current state. After a series of iterations, a stable or set of stable network states, called an attractor, is reached. $\neg$ is the logic symbol for NOT. $\wedge$ is the logic symbol for AND. $\vee$ is the logic symbol for OR.

4.2 Results

4.2.1 SLAM-seq provides a clearer transcriptomic footprint than RNA-seq

Perturbations of biological pathways cause changes in the cellular transcriptome that ultimately shape the cellular phenotype and function. Reconstructing GRNs is crucial to understand these mechanisms but remains a challenge due to the lack of predictive accuracy of binding site presence in the genomic regulatory regions and the noise introduced by the presence of pre-existing mRNAs, which can hide significant changes in expression [109, 223]. Metabolic labeling coupled with NGS can be used to separate mRNAs synthesized before and after a labeling pulse with a nucleotide analog (Fig. 4.2). SLAM-seq is a particularly advantageous technology, as it computationally purifies the newly transcribed mRNAs and therefore requires less laborious and time-consuming experimental work than other metabolic labeling approaches [212]. One key aspect of GRN reconstruction is the identification of TF that regulate the transcriptional changes produced by a perturbation. Therefore, we evaluated whether SLAM-seq provides a competitive advantage compared to RNA-seq for the identification of TF from transcriptional footprints. For this analysis we used publicly available data generated by Muhar *et al.* in which a modified version of the K562 cell line containing an AID-tag in the endogenous *Myc* locus is used, which causes rapid degradation of AID-tagged MYC protein upon treatment with IAA [223]. In this experiment, cells were treated with s4U for 60 minutes after incubation with IAA or DMSO (control). As SLAM-seq labeled (s4U containing) reads are bioinformatically purified, the same experiment can be analyzed as SLAM-seq (using total reads ·

fraction of newly transcribed mRNA, see Methods) or as if it had been a conventional RNA-seq experiment (using total reads) (Fig. 4.9a).

A bird's eye view examination of the data showed that gene expression quantified from SLAM-seq and RNA-seq achieved a good clustering of the samples and separated the control and *Myc* KO cells into two clear groups (Fig. 4.9b, c). When looked at in more detail, the Euclidean distance between samples of different groups was larger for SLAM-seq than for RNA-seq, while the distance between samples from the same experimental group remained unchanged (Fig. 4.9d). This supports the idea that the metabolic labeling performed in SLAM-seq removes transcriptional noise arising from pre-existing mRNAs. We calculated the differentially expressed genes between Myc KO and control cells using DESEq2 [218]. While the total number of differentially expressed genes (FDR-adjusted p.value < 0.05) was similar for both techniques (1,785 for SLAM-seq and 1,513 for RNA-seq), the asymmetry of the distributions was striking: RNA-seq detected roughly the same number of up- and down-regulated genes (696 and 817, respectively), while 92% of SLAM-seq differentially expressed genes were down-regulated (141 upregulated and 1,644 downregulated genes) (Fig. 4.9e, f). We hypothesize this could reflect the predominantly activating role of MYC as a TF [86]. Since one of the main goals of this project is to detect TF orchestrating a transcriptional response to a perturbation, we evaluated the ability of SLAM-seq and RNA-seq gene expression to predict the TF causing the observed transcriptional changes. For this, we performed a GSEA using the CollecTRI database of transcriptional regulation [112]. Fig. 4.9g shows the normalized enrichment scores and adjusted p-values for the statistically significant TFs in SLAM-seq and RNA-seq data. SLAM-seq detected MYC as a differentially active TF, which was expected. However, RNA-seq failed to detect MYC as a significant regulator (adj. p-value $= 0.684$) and instead detected FOXA1 and AR. Lastly, we investigated the transcriptional status of known MYC target genes, as annotated in CollecTRI [112]. From the 567 MYC targets for which a $\log_2$FC could be calculated using both techniques, SLAM-seq detects a higher number of downregulated genes, as happened in the whole-transcriptome footprint (Fig. 4.9h). Additionally, we used a logistic regression model to predict the MOR (stimulation or inhibition) from the observed $\log_2$FC. The model fit using SLAM-seq data (p-value = 2.36e-05, AIC = 321.35) performed better than the model fit with RNA-seq data (p-value = 0.02, AIC = 341.65). Overall, these results suggest that the distinction between pre-existing mRNAs and those synthesized after a perturbation achieved by SLAM-seq reduces the noise in the transcriptomic data and is beneficial for the identification of upstream regulators.

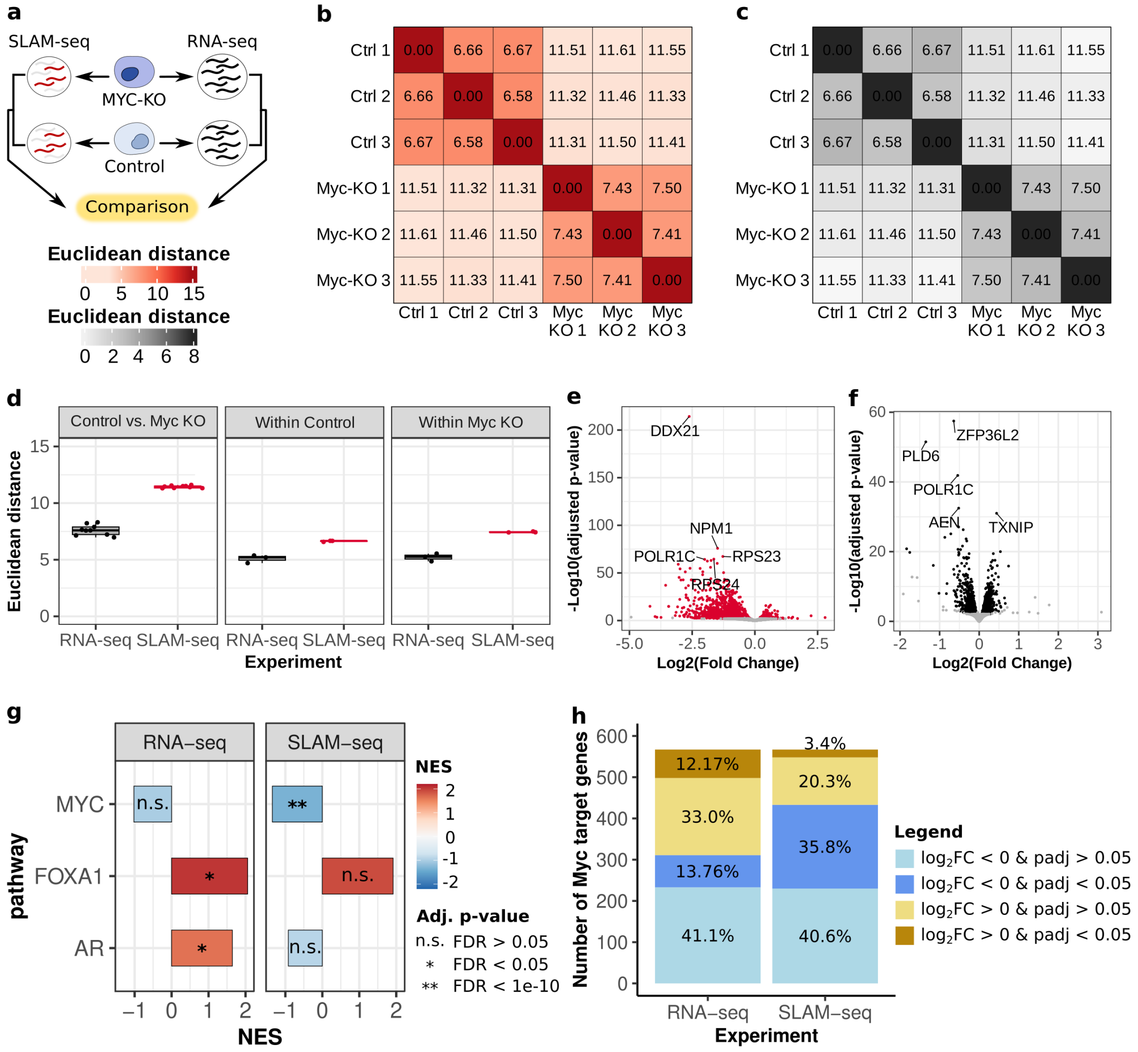


Figure 4.9: Comparison between RNA-seq and SLAM-seq for the identification of upstream regulators of transcriptional changes. **a)** Experimental design. **b)** Heatmap of the Euclidean distance between samples calculated using SLAM-seq reads. **c)** Heatmap of the Euclidean distance between samples calculated using RNA-seq reads. **d)** Boxplot of the Euclidean distances between samples of different groups (Control vs. *Myc* KO) or from the same group (Within control and Within *Myc* KO) measured using the RNA-seq or SLAM-seq gene expression. Wilcox paired test with Bonferroni's correction of p-values was used for comparison. **e)** Volcano plot of the differentially expressed genes calculated with SLAM-seq reads. Highlighted genes have a FDR-adjusted p-value < 0.05. **f)** Volcano plot of the differentially expressed genes calculated with RNA-seq. Highlighted genes have a FDR-adjusted p-value < 0.05. **g)** Statistically significant (FDR-adjusted p-value < 0.05) upstream regulators detected using GSEA using RNA-seq or SLAM-seq reads on the CollecTRI database. The bars are colored by the normalized enrichment score (NES). The numbers on the bars indicate the FDR-adjusted p-value of the enrichment. **h)** Classification of *Myc* direct targets described by the CollecTRI database as not significantly downregulated (light blue), significantly downregulated (dark blue), not significantly upregulated (light ocher) and significantly upregulated (dark ocher) using RNA-seq and SLAM-seq reads.

4.2.2 SLAM-seq protocol optimization in cultures of neonatal mouse CMs

To set up the SLAM-seq protocol in primary CM cultures we tested three s4U incubation times (6h, 12h and 24h) along with an untreated control (Fig. 4.10a). First, we explored whether the s4U treatment introduced too many mismatches that would compromise read mapping and downstream analyses (Table 4.2). We did not find differences in the percentage of uniquely mapped reads between treated and untreated samples. However, the mismatch rate per base was increased in treated samples with respect to the control, indicating that s4U was being incorporated into the mRNA and that the T-to-C and A-to-G mismatches were being produced during the library preparation steps. This increased mismatch rate did not have a big effect in the percentage of multimapping reads or unmapped reads due to a high number of mismatches.

We validated that the increased number of mismatches were a direct consequence of s4U incorporation by visualizing the distribution of mismatches along the read length (Fig. 4.10b). The T-to-C and A-to-G mismatches introduced by SLAM-seq were only higher than the other possible mismatches in the treated samples and were evenly distributed along the read length. Interestingly, increasing the s4U incubation time did not increase the number of mismatches. Lastly, we asked what was the percentage of newly synthesized mRNA detected at the different incubation times. To answer this question, we estimated the average proportion of newly synthesized mRNA per sample and fitted an asymptotic model to the data (p-value = 0.03). The shortest incubation time labeled 50% of the cellular mRNAs (Fig. 4.10c). As this proportion seemed too high to benefit from the higher granularity provided by SLAM-seq, we decided to limit the s4U incubation time to 4h for further experiments. Overall, these results show that SLAM-seq can be performed in neonatal CMs and that the analysis pipeline can detect the incorporation of s4U as T-to-C and A-to-G mismatches.

Table 4.2: Summary of read alignment quality metrics.

Metric	No s4U	6h s4U	12h s4U	24h s4U
Uniquely mapped reads (%)	86.98	88.19	87.98	89.45
Mismatch rate per base (%)	0.27	1.53	1.39	1.17
Reads mapped to multiple loci (%)	7.21	8.76	8.98	7.79
Unmapped reads due to mismatches (%)	0.29	0.74	0.60	0.46

4.2.3 *Yap5sa*, *Myc* or *caErbb2* overexpression rejuvenate the CM transcriptome

To integrate and compare the changes in the CM transcriptome induced by each pro-regenerative gene, we set up a standardized protocol for CM transduction, s4U administration and SLAM-seq (Fig. 4.2). The pro-regenerative genes *Yap5sa*, *Myc* and *caErbb2* were tested alongside *Gfp* controls. Neonatal (P1) CM were highly enriched using the gentleMACS Octo Dissociator with heaters and a pre-plating step to reduce the abundance of fibroblasts and other non-myocyte cells. The cultures were transduced with AAV9 vectors carrying the gene of interest for 44h. Then, 4sU was administered during 4h with medium replacement after 2h and samples were processed for SLAM-seq. As shown above, all samples showed a high alignment quality, enrichment of SLAM-seq-introduced mismatches and approximately 25% of newly transcribed mRNA (Sup. Fig. S4.1).

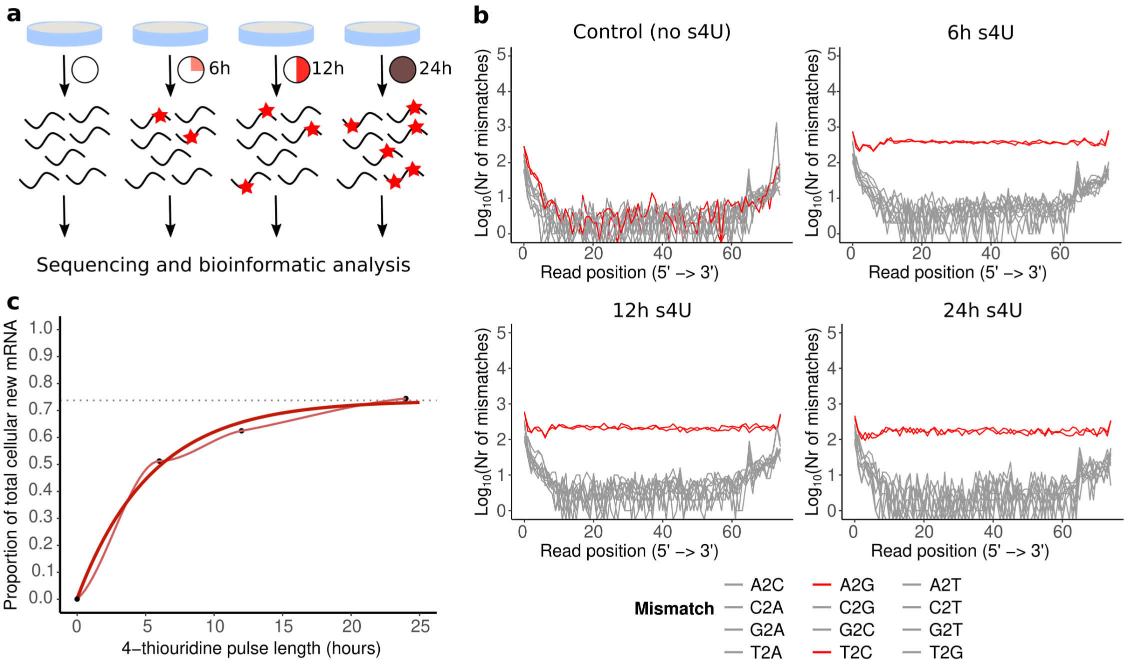


Figure 4.10: Setting-up SLAM-seq in neonatal mouse CMs. a) Experimental design: CM cultures were treated with s4U for 6, 12 or 24h and underwent SLAM-seq. An untreated sample was included as control. **b)** Distribution of mismatches along the read length. Red lines represent the T-to-C and A-to-G mismatches introduced by SLAM-seq. Grey lines represent the other 10 possible mismatches that can arise due to sequencing errors. **c)** Relationship between length of the s4U pulse and the proportion of newly synthesized mRNA. The dark line indicates the asymptotic model fit to the data (p-value = 0.03).

We started by comparing the newly synthesized mRNA in each one of the perturbations (*Yap5sa*, *Myc* or *caErbb2* overexpression) with their respective *Gfp* controls to validate that the activation of the pro-regenerative gene recapitulated previously reported effects [79, 83, 86]. In the case of *Yap5sa*, 139 upregulated and 111 downregulated genes (adj. p-value < 0.1) were detected and clearly separated the *Yap5sa* and *Gfp* samples when visualized as a heatmap (Fig. 4.11a). A GSEA using the GO Biological Process database [183, 184] showed that functions related to protein translation and ribosomal biogenesis were enriched, while functions related to CM contraction, cardiac conduction and aerobic respiration were depleted (Fig. 4.11b). Looking at known YAP1 targets and genes involved in CM maturation and function, we saw a significant increase of the YAP1 targets *Vgll2* ($\log_2$FC = 9.16, adj. p-value = 1.4e-09) and *Wwc1* ($\log_2$FC = 2.4, adj. p-value = 3.2e-06) while the calcium release mediator *Ryr2* ($\log_2$FC = -0.7, adj. p-value = 0.08) and the mature CM myosin *Myh6* ($\log_2$FC = -1.1, adj. p-value = 0.01) were downregulated (Fig. 4.11c). The increase in CM proliferation was validated *in vitro* by performing a phospho-histone 3 (pHH3) staining (Fig. 4.11d).

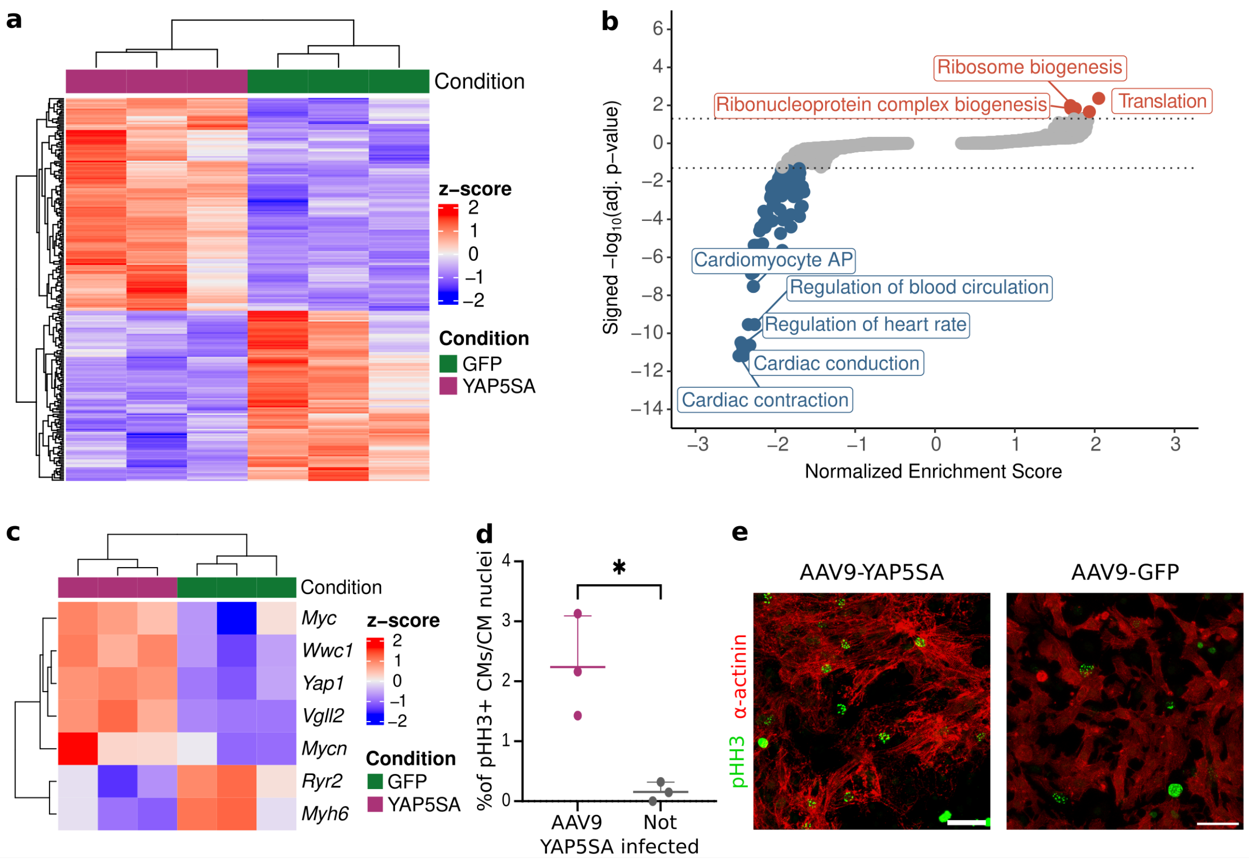


Figure 4.11: Effect of *Yap5sa* overexpression in the neonatal CM transcriptome. a) Heatmap of the differentially expressed genes between AAV9-*Yap5sa* and AAV9-*Gfp*-treated CMs. **b)** Scatterplot of the GSEA results. Dots colored in red (enriched) or blue (depleted) indicate GO Biological Process terms with an adjusted p-value < 0.05. **c)** Heatmap of the scaled expression of selected genes in the AAV9-*Yap5sa* and AAV9-*Gfp*-treated CMs. **d)** Quantification of the percentage of pHH3$^+$ CM nuclei after transduction with AAV9-*Yap5sa* vs non-infected CMs. Each point represents the average proportion of positive CMs in a single technical replicate. A two-tailed t-test was used to compare the samples. The asterisk represents a p-value $<$ 0.05. **e)** A representative image of CM cultures transduced with AAV9-*Yap5sa* and of a non-infected CM culture is shown. In this experiment, CMs were transduced for 72h before pHH3 immunohistochemistry.

Regarding *Myc* overexpression, we detected 254 upregulated and 278 downregulated genes (adj. p-value < 0.1) (Fig. 4.12a). A GSEA using the GO Biological Process database [183, 184] showed that functions related to protein translation, ribosomal biogenesis, DNA replication and cell cycle were enriched, while functions related to CM contraction, cardiac conduction and electron transport chain were depleted (Fig. 4.12b). Looking at specific genes, we saw a significant increase of the MYC targets *Ddx21* ($\log_2$FC = 0.38, adj. p-value = 0.079) and *Rrp9* (log2FC = 0.4, adj. p-value = 0.07) and some cell cycle-related genes such as *Aurka* ($\log_2$FC = 0.4, adj. p-value = 0.1), while in this case the CM maturation genes *Ryr2* ($\log_2$FC = -0.1, adj. p-value = 0.41) and *Myh6* ($\log_2$FC = -0.11, adj. p-value = 0.53) were not significantly downregulated (Fig. 4.12c). The increase in CM proliferation was validated *in vitro* by performing a pHH3 staining (Fig. 4.12d).

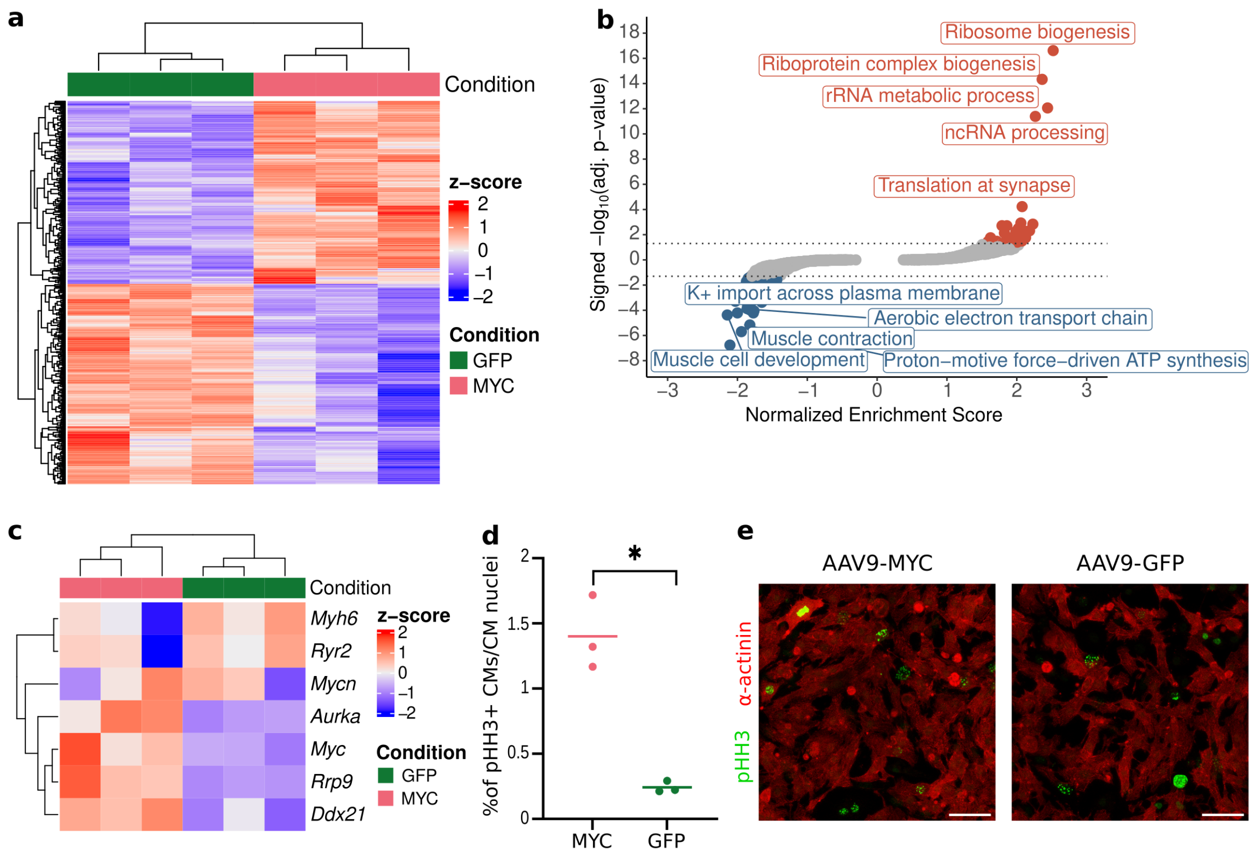


Figure 4.12: Effect of *Myc* overexpression in the neonatal CM transcriptome. **a)** Heatmap of the differentially expressed genes between AAV9-*Myc* and AAV9-*Gfp*-treated CMs. **b)** Scatterplot of the GSEA results. Dots colored in red (enriched) or blue (depleted) indicate GO Biological Process terms with an adjusted p-value < 0.05. **c)** Heatmap of the scaled expression of selected genes in the AAV9-*Myc* and AAV9-*Gfp*-treated CMs. **d)** Quantification of the percentage of pHH3$^+$ CM nuclei after transduction with AAV9-*Myc* or AAV9-*Gfp*. Each point represents the average proportion of positive CMs in three technical replicates. A two-tailed t-test with Welch's correction was used to compare the samples. The asterisk represents a p-value < 0.05. **e)** A representative image of CM cultures transduced with AAV9-*Myc* or AAV9-*Gfp* is shown. In this experiment, the transduction conditions were identical to those used for SLAM-seq.

Lastly, overexpression of *caErbb2* in CMs had the mildest effect. We detected 34 upregulated and 36 downregulated genes (adj. p-value < 0.1) (Fig. 4.13a). A GSEA using the GO Biological Process database [183, 184] showed enrichment of functions related to protein translation and ribosomal biogenesis, while functions related to CM maturation, contraction and cardiac conduction were depleted (Fig. 4.13b). Looking at specific genes, we saw a significant increase of ERBB2-induced genes such as *Rps14* ($\log_2$FC = 0.62, adj. p-value = 0.04), *Rps23* ($\log_2$FC = 0.46, adj. p-value = 0.07) or *Timp1* ($\log_2$FC = 0.46, adj. p-value =. 0.06), while the CM maturation gene *Myh6* ($\log_2$FC = -0.36, adj. p-value = 0.09) but not *Ryr2* ($\log_2$FC = -0.11, adj. p-value = 0.25) was significantly downregulated (Fig. 4.13c). The increase in CM proliferation was explored *in vitro* by performing a pHH3 staining, although we did not observe a significant increase probably due to the high variability observed in the AAV9-*caErbb2* group (Fig. 4.13d).

Overall, the overexpression of *Yap5sa*, *Myc*, and *caErbb2* had the expected rejuvenating effect on the CM transcriptome, disassembling sarcomeres and increasing proliferation. While the observed variability in the magnitude of changes could be due to biological or technical reasons, we hypothesize that all three genes seem to regulate similar pathways.

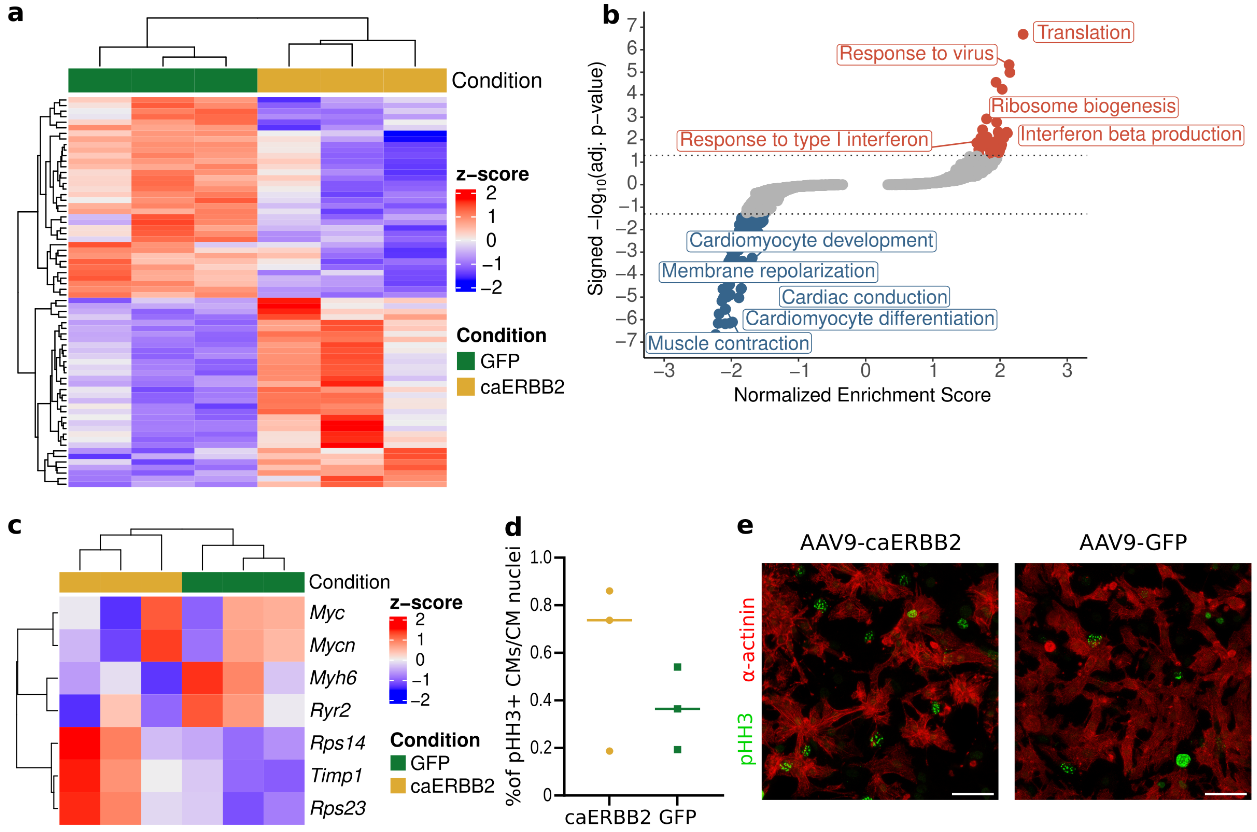


Figure 4.13: Effect of *caErbb2* overexpression in the neonatal CM transcriptome. a) Heatmap of the differentially expressed genes between AAV9-*caErbb2* and AAV9-*Gfp*-treated CMs. **b)** Scatterplot of the GSEA results. Dots colored in red (enriched) or blue (depleted) indicate GO Biological Process terms with an adjusted p-value < 0.05. **c)** Heatmap of the scaled expression of selected genes in the AAV9-*caErbb2* and AAV9-*Gfp*-treated CMs. **d)** Quantification of the percentage of pHH3$^+$ CM nuclei after transduction with AAV9-*caErbb2* or AAV9-*Gfp*. Each point represents the average proportion of positive CMs in three technical replicates. A two-tailed t-test with Welch's correction was used to compare the samples. **e)** A representative image of CM cultures transduced with AAV9-*caErbb2* or AAV9-*Gfp* is shown. In this experiment, the transduction conditions were identical to those used for SLAM-seq.

4.2.4 GoM models show a common effect between *Myc* and *caErbb2*

Clustering methods are useful to detect shared patterns between samples or conditions. However, clustering methods such as k-means or hierarchical clustering have the problem that each sample, or gene, is assigned into a single cluster. The reality of gene expression profiles is more complex, with multiple processes being active at the same time. GoM models allow to identify sets of genes (gene topics) that share a common expression pattern in the samples and assign a grade of membership for each sample in each topic, therefore quantifying the contribution of each gene expression program in the sample. Focusing on the differentially expressed genes in the SLAM-seq data, a GoM model consisting of three topics was fit to the data (Fig. 4.14a). Despite each sample being a combination of all three topics (Fig. 4.14b), we detected an enrichment of Topic1 gene expression in *caErbb2* and *Myc*-treated CMs (Bonferroni-corrected one-way ANOVA p-value = 3.42e-03), an enrichment of Topic2 gene expression in *Yap5sa*-treated samples (Bonferroni-corrected one-way ANOVA p-value = 2.66e-03) and an enrichment of Topic3 gene expression in *Gfp*-treated samples (Bonferroni-corrected one-way ANOVA p-value = 7.73e-03) (Fig. 4.14c).

Using an ORA, we associated the gene topics to their corresponding biological functions: Topic 1 is related to cell cycle and ribosomal biogenesis, Topic 2 is involved in muscle development and proliferation and Topic 3 contains genes that participate in CM contraction, cardiac conduction, and sarcomere organization (Fig. 4.14d). Interestingly, when we repeated the analysis using the total gene expression as quantified by RNA-seq, the best fitting GoM model contained four topics: a cell proliferation and protein synthesis topic predominant in the *caErbb2* and the *Myc*-treated samples, a muscle proliferation topic highly associated with the *Yap5sa*-treated samples and two topics associated with the *Gfp* samples. While both topics contained genes involved in CM maturation and sarcomere organization, they separated the *Gfp* samples based on their experimental batch (Sup. Fig. S4.2). Overall, the GoM model found functional gene expression programs involved in rejuvenation of the CM transcriptome –potentially required for regeneration– as well as a basal CM maturation program. It also detected a shared transcriptional response by *caErbb2* and *Myc* mainly inducing cell proliferation. Regarding the comparison between GoM models fit with SLAM-seq and RNA-seq, the RNA-seq fitted model captures batch effect variability, while the SLAM-seq model captures biologically relevant variability, once more supporting the use of this technique for transcriptomic profiling after a perturbation.

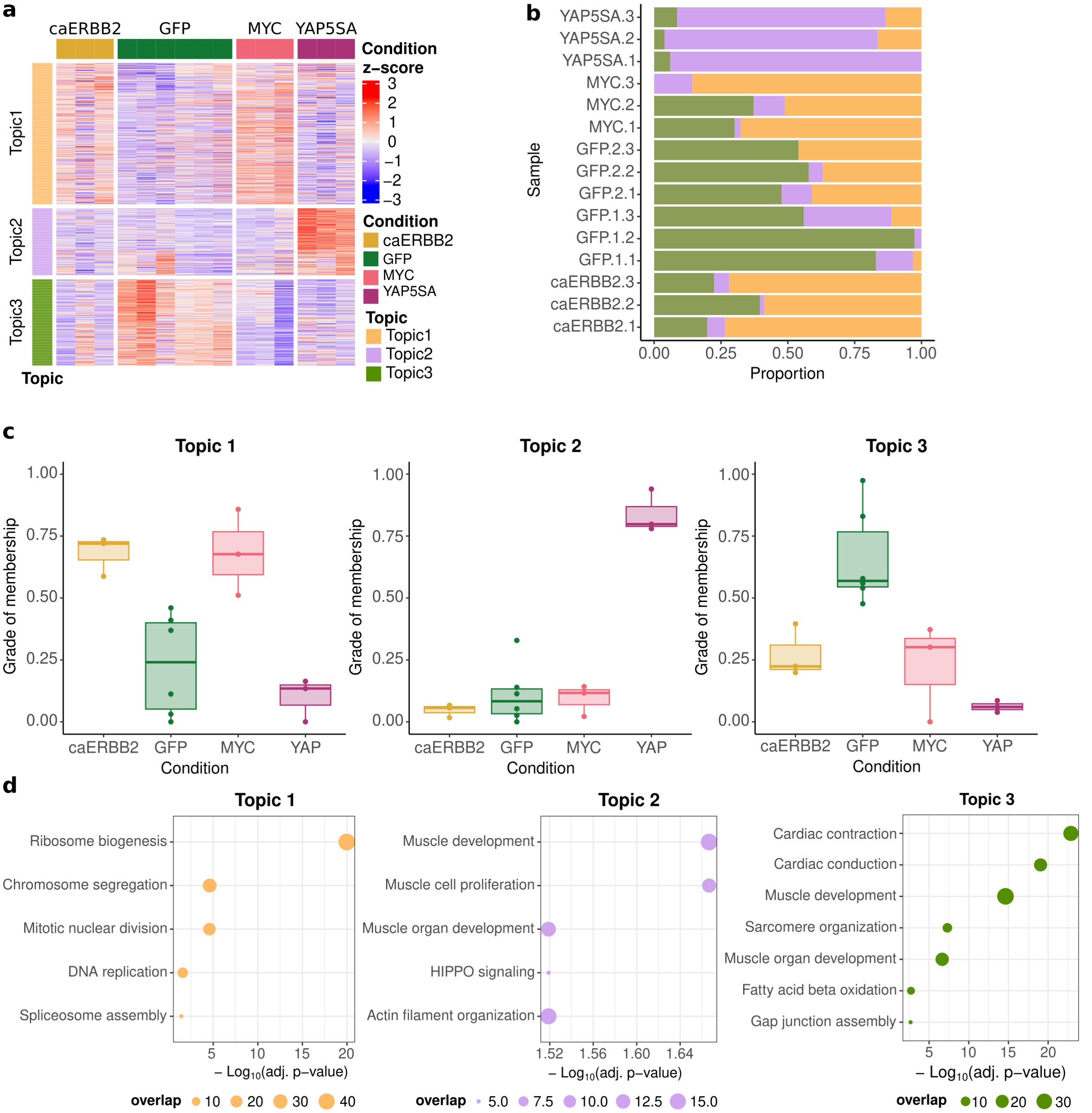


Figure 4.14: GoM models find a common gene expression program induced by *caErbb2* and *Myc*. **a)** Heatmap of scaled gene expression. Genes are classified into topics based on their maximum topic $\log_2$FC. Each column indicates a biological replicate, and samples are grouped based on their experimental condition. **b)** Barplot of the degree of membership of each sample in each topic. **c)** Boxplots of the grade of membership of each condition on each topic. **d)** Most significant biological functions found for each topic using an ORA. The size of the dot indicates the number of overlapping genes between the pathway and the topic. In panel **b**, the experimental batch of the GFP samples is indicated in their name. Batch 1 consisted on AAV9-*Yap5sa* and AAV9-*Gfp*.1 treated cultures. Batch 2 included AAV9-*Myc*, AAV9-*caErbb2* and AAV9-*Gfp*.2 treated cultures

4.2.5 Integrating literature findings and scRNA-seq into a CM GRN scaffold

Whole-transcriptome GRN inference is challenging due to the skinny-matrix nature of bulk transcriptomics data, in which the number of genes is highly superior to the number of observations. Databases of transcriptional regulation are a useful starting point to reduce the complexity of the GRN model and focus on the TF – target gene relationships that are more likely to happen in the given context. Aiming to find a small set of TFs with a key function controlling CM maturation, we contextualized the CollecTRI PKN using the CM scRNA-seq data described in Chapter 3 and random forests [19]. The resulting GRN contains 2,097 genes and 2,169 edges, organized into 127 independent components (Fig. 4.15). The largest of them contains 1,637 genes and is organized into 36 communities, 25 of which could be associated to biological processes through an ORA on the GO Biological Process database (Table S4.1).

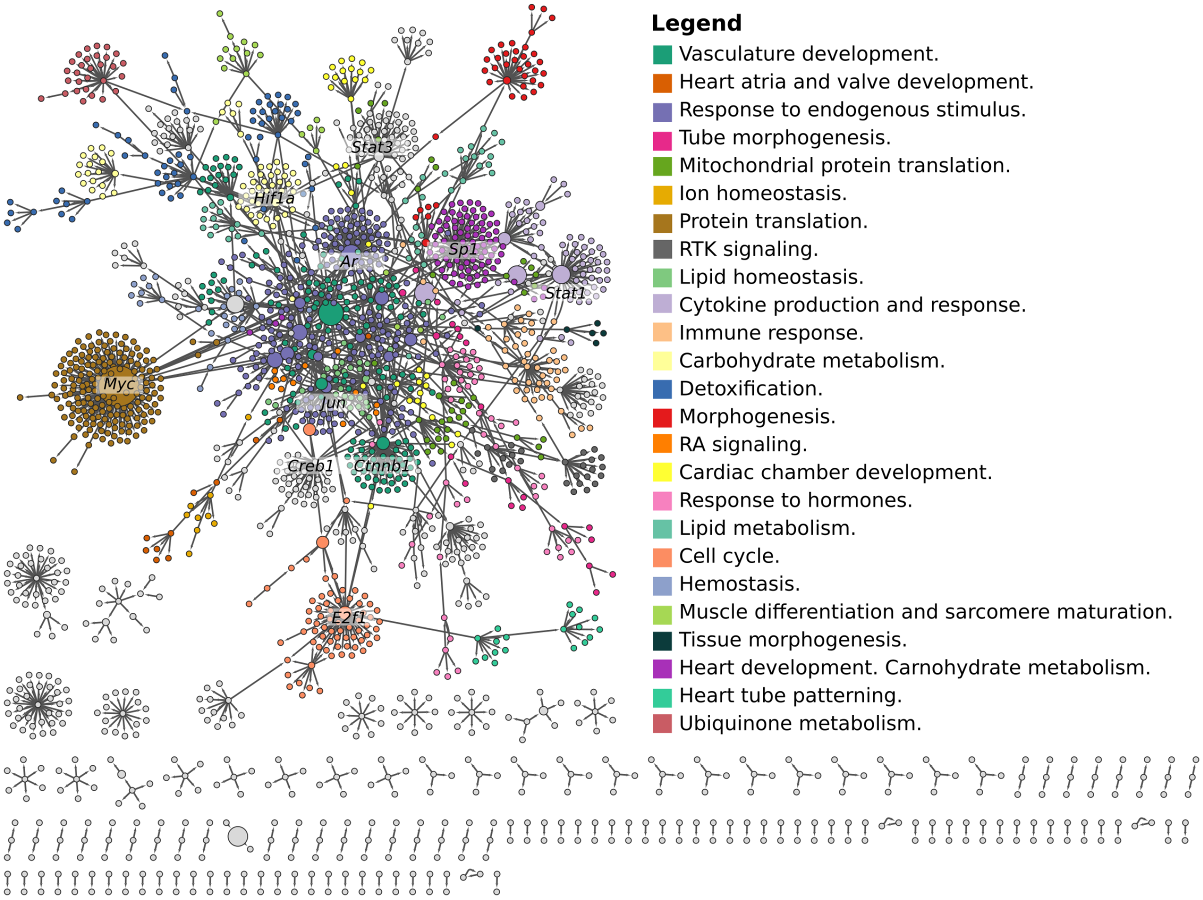
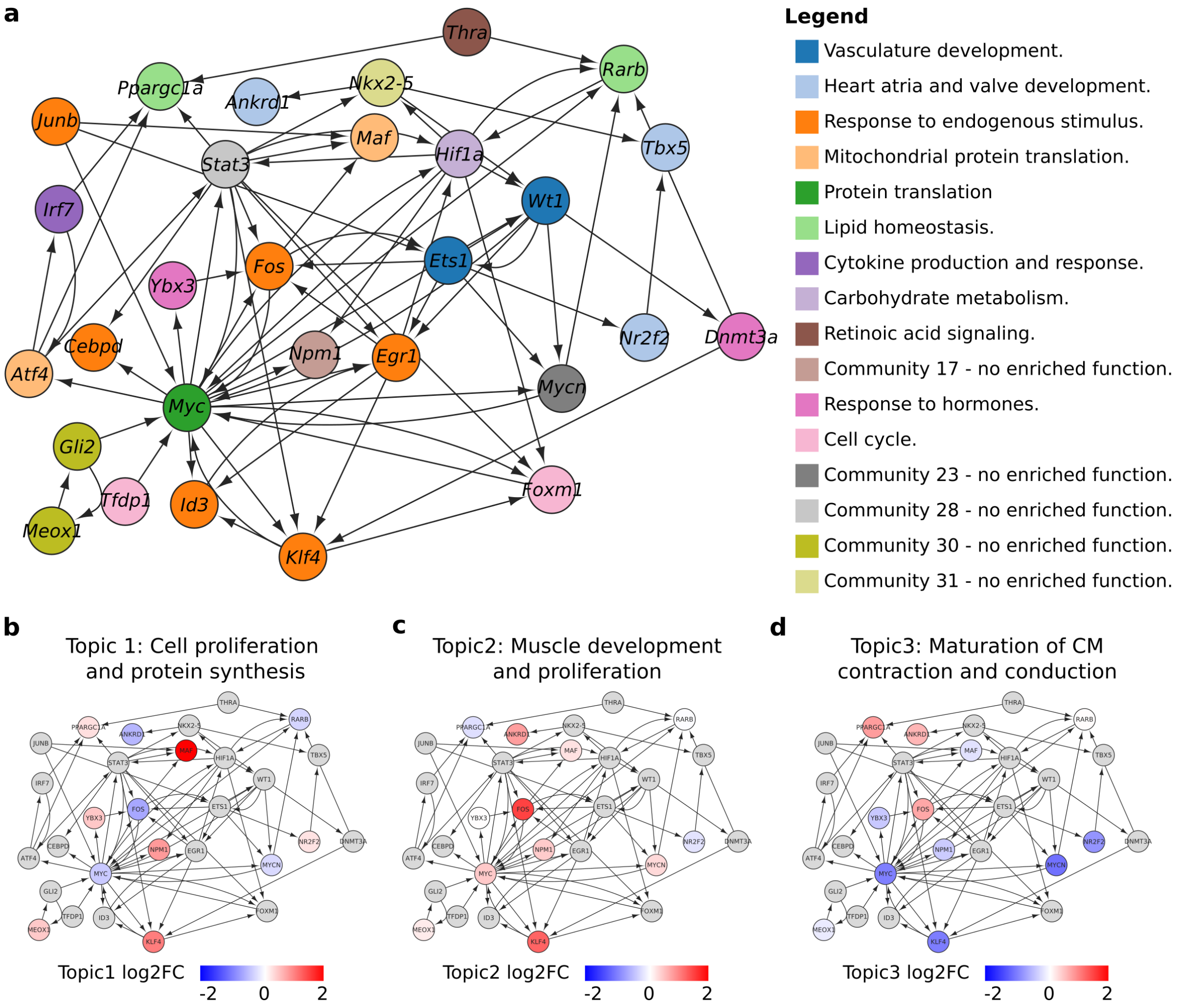


Figure 4.15: Contextualization of the CollecTRI database using CM scRNA-seq data. Genes are colored by their community membership. Communities for which at least one statistically significant (FDR adjusted p-value < 0.05) biological function was found are highlighted. Gene size represents betweenness centrality.

Because transcriptional regulation is ultimately defined by interactions between TF and most of them are located in the largest network component, we focus on it for further modeling. We reduced this component to a core set of TFs by considering only regulators (out degree > 0) with a putative function in the tested perturbations ($|\log_2 \text{FC}| > 0.25$ in at least a single experiment). The resulting core GRN consisted in 29 TF and 77 edges (Fig. 4.16a and Table S4.2). Most of these have been previously studied in the context of heart development, congenital heart disease or heart regeneration. These TF represent 16 out of the 36 communities detected in the GRN largest component and nearly half of them (12/29) are overexpressed in at least one gene topic (Fig. 4.16b-d). This core GRN will be the base for the Boolean models and perturbation simulations described in the following sections.



b Topic 1: Cell proliferation and protein synthesis

c Topic2: Muscle development and proliferation

d Topic3: Maturation of CM contraction and conduction

Figure 4.16: Core CM GRN. In panels **a-d)** the set of core TF in the CM GRN is represented. Each gene is colored by **a)** the community to which it belongs. **b)** the $\log_2$FC for Topic1 (cell cycle and ribosomal biogenesis), **c)** the $\log_2$FC for topic 2 (muscle development and proliferation) and **d)** the $\log_2$FC for topic 3 (CM contraction, cardiac conduction, and sarcomere organization).

4.2.6 The CM maturation GRN is controlled by two attractors

Boolean modeling with boolsim [18] was applied to the core GRN. First, the GRN topology was examined to identify potential attractors or stable network states that could be representative of cellular phenotypes [18]. The core GRN contains two basal attractors in which all genes have a clear ON or OFF status except *Atf4* and *Irf7*, whose activation state was flexible and could be either ON or OFF (Fig. 4.17a,b). In order to find which gene topic in Fig. 4.14a could be explained by these attractors, we measured the extent to which the dichotomized topic $\log_2$FC could predict the gene activation status in the attractors. The AUROC showed that both attractors are correlated with the CM maturation topic (Fig. 4.17c). This result was expected, as the core GRN was obtained from basal CM gene expression and two different timepoints. This basal state serves as a good starting point to find new pro-regenerative candidates that could potentially change the GRN behavior.

4.2.7 Systematic *in silico* perturbations show transitions from maturation to proliferation attractors

Starting from the two basal attractors, we simulated KO and overexpression perturbations and performed a reachability analysis to see what new attractors could be reached if each TF in the core GRN was individually perturbed (Fig. 4.18). 76% of the perturbations did not change the basal attractors, 18% of the perturbations generated new attractors that were associated to the gene expression topics as described above and 6% of the perturbations generated new attractors for which a correspondence with the gene topics could not be found. Interestingly, only two perturbations were able to change both basal attractors into new attractors potentially involved in cell proliferation: overexpression of *Myc* or overexpression of *Ybx3*. In the case of *Myc* overexpression, this was concordant with our AAV9-Myc CM transduction data (Fig. 4.12d). In the case of *Ybx3*, it is a TF with a known role in regulating the expression of amino acid transporters [224], but its function in the heart remains unknown. Overall, these results provide a potential model to understand how a core set of TFs can regulate CM maturation and which perturbations can modify that behavior. Our simulations show that CM maturation and basal function is a strongly wired phenotype difficult to change but highlight a few candidates worth of further exploration (Table 4.3).

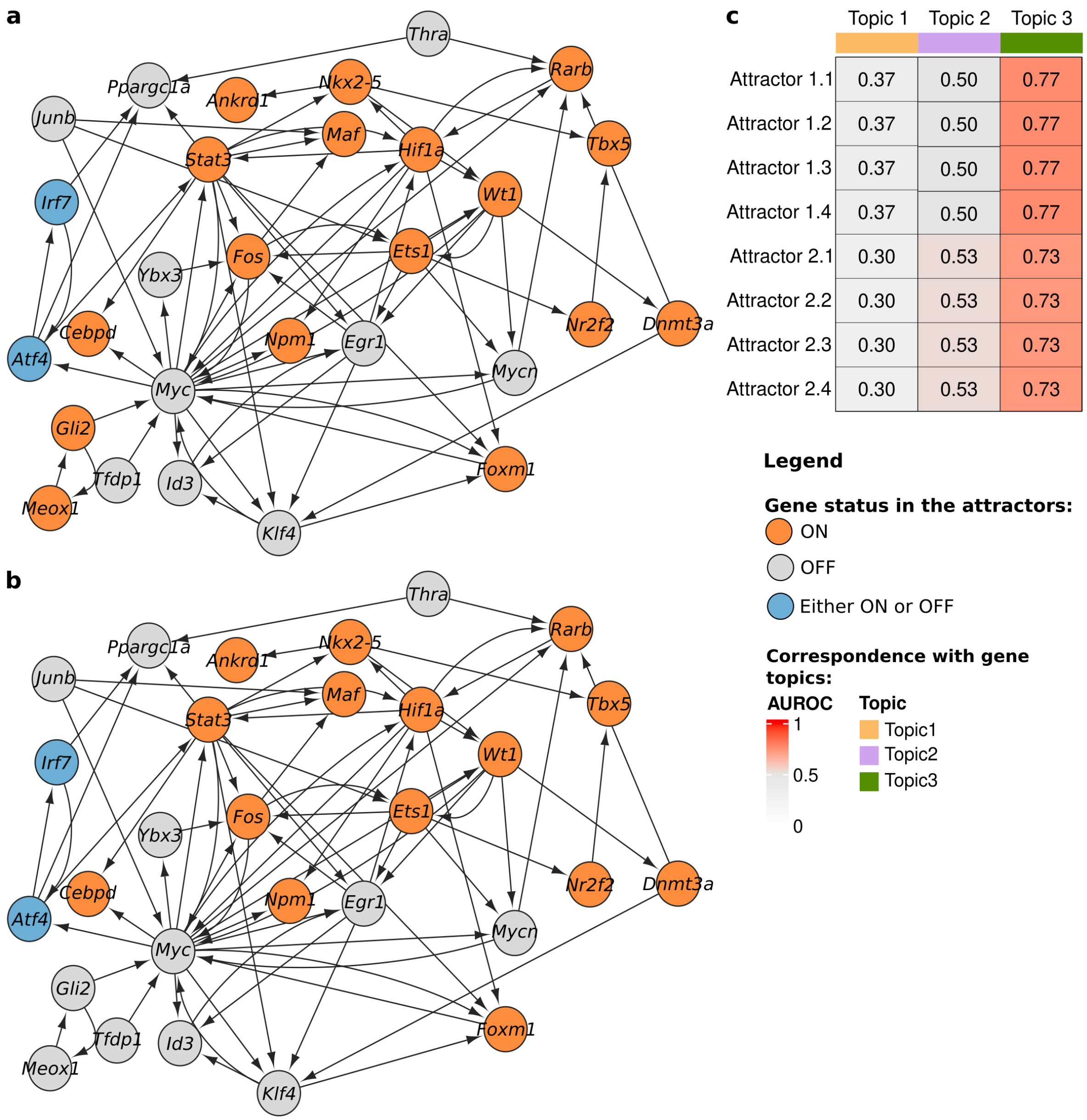


Figure 4.17: The core CM GRN is characterized by two basal attractors that correspond to the CM maturation topic (topic 3). a-b) Core CM GRN with the genes colored by their activation status in basal **a)** attractor 1 or **b)** attractor 2. In both networks, orange represents activation, grey represents inactivation and blue represents that both active and inactive states are possible. **c)** Heatmap of the AUROC between dichotomized topic $\log_2$FC and gene activation status. The basal attractors 1 and 2 are subdivided into 4 subattractors due to the presence of two genes that can be indistinctively active or inactive.

Table 4.3: Candidates proposed by the Boolean models and perturbation simulations. Previous descriptions refer to publications in which the perturbation of the TF was evaluated in the context of heart regeneration. The asterisk indicates that *Wt1* has been implicated in the role of epicardium during regeneration, but its cell autonomous pro-regenerative function in CMs is yet to be explored. Whether the new attractor is reachable from a single ((a1), (a2)) or both basal attractors is indicated.

TF	**Experiment**	**Prediction**	**Reachable attractor**	**Known effect**
Ybx3	Overexpression	Cell cycle	Yes (both)	No
Rarb	KO	Cell cycle	Yes (a1)	No
Wt1	KO	Dedifferentiation	No	Yes* [225, 226]
Irf7	KO	Cell cycle	No	No
Myc	Overexpression	Cell cycle	Yes (both)	Yes [86, 87]
Fos	KO	Cell cycle	Yes (a1)	Yes [227]
Atf4	KO	Cell cycle	No	Yes [228, 229]
Stat3	KO	Cell cycle	No	Yes [230, 231]

4.3 Discussion

This chapter explored the transcriptional footprints induced by the expression of the pro-regenerative genes *caErbb2*, *Myc* and *Yap5sa* and modeled the GRN controlling them. We used a high-throughput metabolic labeling approach (SLAM-seq) to reduce the noise in the transcriptome measurement of neonatal mouse CM following transduction with AAV9 vectors. By leveraging GoM models, we found three gene expression programs characteristic of *caErbb2* and *Myc* overexpressing CM (cell cycle and protein translation), *Yap5sa* overexpressing CM (muscle dedifferentiation) and *Gfp* expressing controls (CM maturation, cardiac conduction, and sarcomere organization). The mouse scRNA-seq data set introduced in Chapter 3 was used to contextualize a database of literature-based knowledge of transcriptional regulation using random forest models. The resulting GRN contained a single large connected component in which communities associated to developmental and homeostatic CM functions were found. This GRN was reduced to a core set of 29 key regulators subjected to Boolean modeling. The Boolean models found two possible attractor states of the GRN, that correlate with the “CM maturation” gene expression program. By systematic *in silico* experiments, we found 10 perturbations with the potential to drive the core GRN towards a “cell proliferation” or “muscle dedifferentiation” state. Altogether, these results provide insights into the molecular dynamics underlying the pro-regenerative response initiated by *caErbb2*, *Myc* and *Yap5sa* expression and suggest candidate TFs for their experimental perturbation (Fig. 4.19).

To our knowledge, this project constitutes the first use of SLAM-seq combined with AAV9-based perturbations. Combining these two advanced experimental methodologies required extensive testing and, despite their strengths, is not extent of limitations. Particularly, the transcriptional changes induced by *caErbb2* were smaller than expected, given its profound effect on CM dedifferentiation [83, 232]. We hypothesize that, since it is the only membrane protein tested in this project, it might need longer transduction times than *Myc* or *Yap5sa* to exert a transcriptome rewiring of the same magnitude. Another point to raise is that neonatal mouse CM do not behave like adult CM in their potential to enter the cell cycle [4]. This is clearly showcased by *Myc* overexpression, which can only induce CM proliferation when *Ccnt1* is available: while neonatal CM express enough *Ccnt1*, adult CM have lost its expression [86]. This is of important consideration when looking into translation, as more than one simultaneous perturbation might be necessary in some cases.

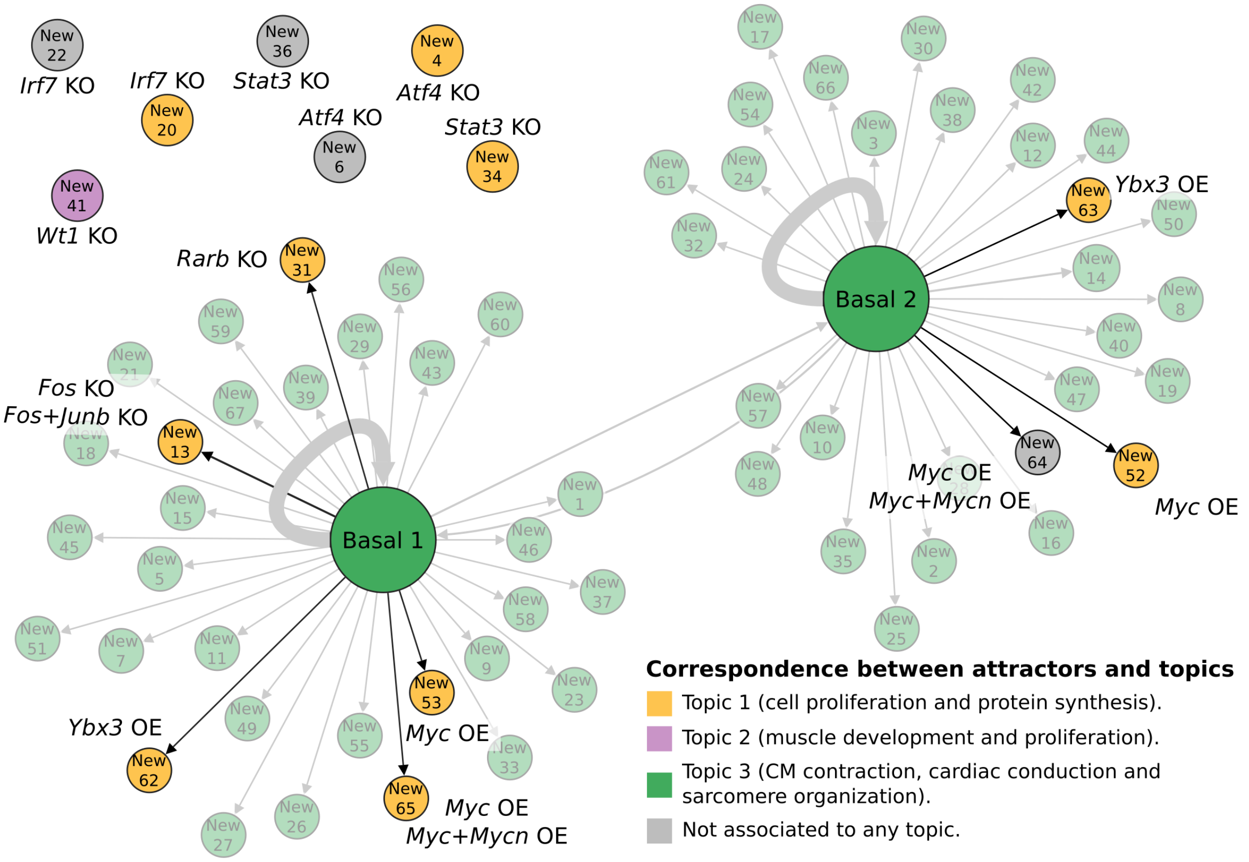


Figure 4.18: Attractor reachability caused by the *in silico* perturbations. In this network, each node represents an attractor, and each edge represents a potential transition between attractors. Node size represents the number of times the attractor was observed. Node color represents the gene topic to which the attractor corresponds. Edge width represents the number of times a given transition between attractors was observed. Basal attractors and new attractors that correspond with the pro-regenerative gene topics (topic 1 and 2) are highlighted. The perturbations that give rise to these new pro-regenerative attractors are indicated. KO stands for knockdown and OE stands for overexpression.

Regarding SLAM-seq, it is a technique with a lot of potential that, in combination with GoM models, allowed us to obtain batch-free gene expression programs. In fact, classical hierarchical clustering with or without batch correction was not able to identify batch-free and condition-wide gene expression programs. It was the combination of SLAM-seq data with GoM models that could obtain a batch-free representation of the transcriptomic changes. The identified gene topics represent a useful baseline to which future pro-regenerative candidate treatments can be compared.

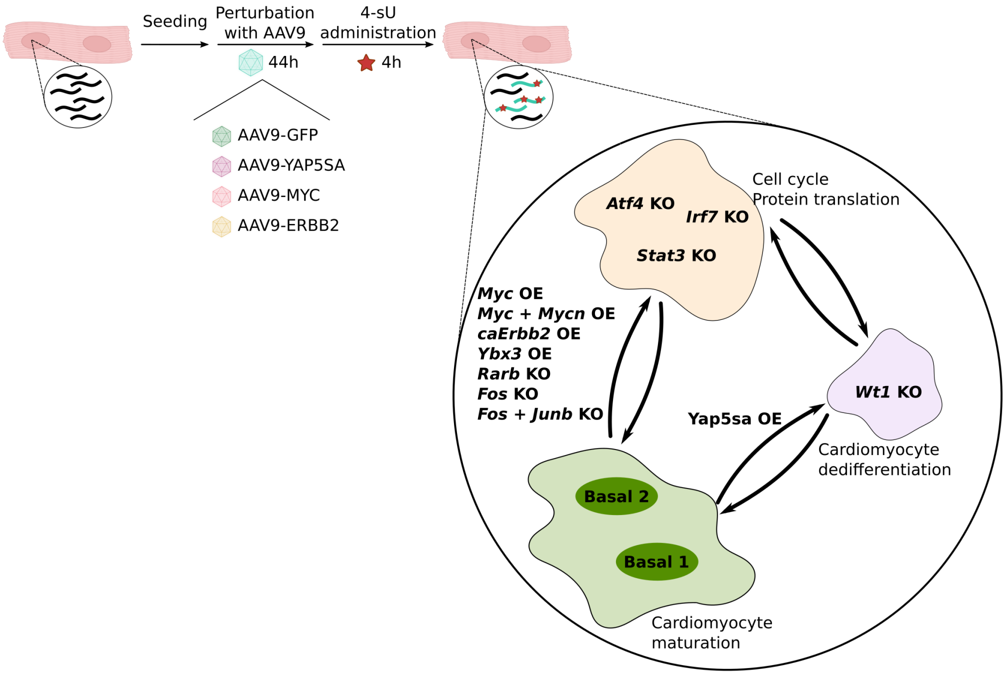


Figure 4.19: Summary of the project. SLAM-seq and GoM models found three gene topics in CM transduced with the pro-regenerative genes *caErbb2*, *Myc*, *Yap5sa* or *Gfp*. A topic mostly expressed in the *caErbb2* and *Myc* treated samples is associated with cell cycle and protein translation. A gene topic mostly associated with the *Yap5sa* treated sample associates with muscle dedifferentiation and proliferation. A topic predominant in the *Gfp* group is associated with CM maturation and sarcomere assembly. We found two core GRN attractors that correspond to the CM maturation gene expression program, and through *in silico* perturbations we propose candidate treatments to induce CM proliferation and potentially heart regeneration.

Regarding the computational models, the small sample sizes and lack of time-series information were the largest hurdle to overcome. Boolean models are a useful tool in these situations, as the gene expression is dichotomized, and the network update iterations do not represent real time. However, Boolean models of large networks are computationally expensive, as they exhaustively enumerate all possible trajectories and network states, which exponentially grow as more nodes are included in the network. This computational limitation called for a stringent network filtering criteria and the selection of a small set of core TFs representative of the whole GRN. While the selection of core TF was systematic and driven by available data and previous knowledge, the network model might be oversimplistic and relevant TF might be missing. If time-series data was available, a more satisfactory approach would have been to combine logical and quantitative methods such as ODEs to fully exploit data and available knowledge.

Lastly, the model interpretation leads us to conclude that the CM maturation GRN state is a robust network state difficult to exit. This is supported by the fact that only a handful of genetic perturbations achieve a significant increase in CM proliferation or have measured functional recovery after MI [79, 83, 87, 200, 230, 233–235]. The main candidate coming out of this project is the fairly unknown *Ybx3*. Its function as regulator of amino acid transporters [224] hints towards its potential implication in the regulation of metabolism, which is substantially different between neonatal (proliferative) and adult (quiescent) CMs. Future experiments should address its expression levels as the heart matures and in response to injury, as well as its ability to induce CM cell cycle activity. If its proliferative role is confirmed, functional studies in an injury context should be performed to address its regenerative potential and transcriptomic studies in basal and injury context should be performed to validate the similarity between the *Myc*-induced transcriptional changes and the transcriptional changes induced by *Ybx3*.

5 Inference of the gene regulatory network of axolotl bone regeneration

Axolotl (*Ambystoma mexicanum*) limb regeneration is a classic system to study vertebrate regeneration. Contrary to the full regeneration achieved after an amputation, axolotl bone CSD fails to regenerate. This chapter delves into the molecular differences between blastema and CSD CT to unravel their differential regenerative potential. By inferring and comparing the CT GRNs in both injuries, we identified two TCF/LEF family TFs, *Lef1* and *Tcf7l2*, having potentially divergent roles. This finding was further strengthened by the identification of genes uniquely associated with each TF and through simulations of *Lef1* and *Tcf7l2* perturbations on the CT GRN. These results collectively suggest different and eventually antagonistic functions of *Lef1* and *Tcf7l2* during axolotl limb regeneration and propose a new level of complexity in Wnt signaling that could contribute to the difference in regenerative competence observed in different tissues.

5.1 Materials and methods

5.1.1 Experimental procedures

All experimental procedures were performed by Anastasia Polikarpova, PhD and are briefly described here to provide the necessary background for the *in silico* analysis of the gathered data. The scripts used for data analysis are available at: `https://github.com/iriverog/AxolotlLimbGRN`.

Axolotl husbandry

10-11 cm nose-to-tail Prrx1-CreER x CAGGS: LP-STOP-LP-Cherry axolotls were used in this project. Upon Cre activation by bathing the animals in a 1-2 μM 4-hydrotamoxifen water solution, these animals express the Cherry fluorescent protein in *Prrx1*$^+$ limb bud CT cells and their progeny. The sex of the animals was unknown, as they were not genotyped, and sexual characteristics are not present at this stage. Animals were bred and maintained in the animal facilities of the IMP in Vienna, Austria, as described previously [165]. Animals were kept individually, and experiments were conducted with approval from the Magistrate of Vienna (GZ:MA 58-65248-2021-26).

Blastema and CSD injury models

Two injury models were used in this study: proximo-distal amputation (blastema) and CSD (Fig. 5.1).

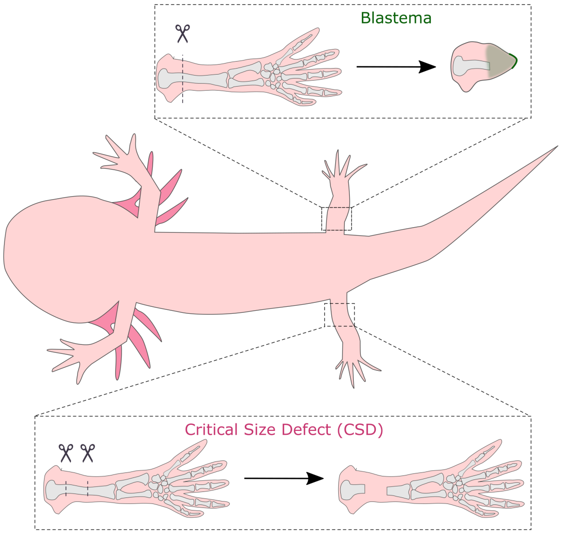


Figure 5.1: Axolotl bone injury models. Axolotl hindlimbs were subject to either amputation to generate a regenerative blastema (green, top box) or a non-regenerative CSD (magenta, lower box). In both cases the injury affects the femur. In the case of blastema, all tissues (e.g., bone, muscle, nerves, epidermis, etc.) are affected by the injury. In the case of CSD, only a bone segment is removed, with the remaining limb tissues being minimally disturbed.

Blastema: Hindlimb amputation for the development of a regenerative blastema was performed as previously described [236]. Briefly, axolotls were anesthetized by bathing in a 0.03% benzocaine solution (Sigma-Aldrich, Germany) for 15-20 minutes until a good muscle relaxation and lack of reflexive movement upon limb touching was achieved. Anesthetized animals were placed on a paper towel bathed in a 0.03% benzocaine solution during the surgical procedure. Axolotl hindlimbs were amputated at the femur diaphysis midshaft using a disposable scalpel. The extruding bone was trimmed using iridectomy scissors to match the level of the surrounding tissue and prevent irregularities in blastema shape.

CSD: Animals were anesthetized and prepared for surgery as described above. A hindlimb was stretched and an external polyolefin fixator tube was attached by size 7.0 nylon sutures (Optilene, BBraun, Germany) to ensure bone alignment and preservation of the CSD gap size. A longitudinal incision was made through the surgical window of the fixator above the femur using iridectomy scissors (FST, Germany). The nerves, blood vessels and muscles were carefully displaced from the surgery site without cutting using a fine surgical forceps (FST, Germany). The femur was lifted and exposed for surgery using a forceps and cut twice in the diaphysis with iridectomy scissors to create a CSD of approximately 30% of the bone length. As axolotl wounds close rapidly within 6-8 hours, no stitches were applied on the skin cut.

After surgery, animals were housed in individual tanks with clean tap water and observed daily. 50 U/mL penicillin and 20 µg/mL streptomycin (Gibco, 15140-122) were added the first 3 days after amputation. Butorphanol (0.5 mg/L water) was used to ensure analgesia. Animals were sacrificed by bathing in 0.03% benzocaine solution for 20-30 minutes 3, 5, 6, 7, 8, and 11 days post injury. A hindlimb sample from an uninjured and healthy animal was included as physiological control (day 0). Between 12 and 24 animals were used per time point and injury (Fig. 5.2).

Tissue dissection and cell isolation for scRNA-seq

Immediately after euthanasia, limbs were harvested and the blastema or tissue within CSD were extracted using sterile fine forceps and a disposable scalpel. The tissue was finely minced using a scalpel blade and digested with 5-10 mL LiberaseTM (0.26 WU/mL, Roche) in 0.7x PBS (without Mg^{2+}/Ca^{2+}) at room temperature and rotation for 45 minutes with shaking every 10-15 minutes. The digestion was blocked using 2 mL serum-free AMEM (125 mL MEM, 2mL insulin (1 mg/mL, Sigma-Aldrich), 2 mL glutamine (200 mM, Gibco), 2mL Pen/Strep (10000U/mL, Sigma-Aldrich) and 50 mL ddH20) and a single-cell suspension was generated by filtering the samples through a 30 µm diameter strainer. For Batch 1, the obtained suspension was centrifuged at 300 relative centrifugal force (rfc) for 5 minutes and re-suspended in 200 µL AMEM. For Batch 2, red blood cells were depleted by adding 2 mL 5% Ficoll/APBS to the single-cell suspension and placing it on top of a Ficoll gradient (2 mL of 15%, 17%, 19% and 25% Ficoll/APBS). After centrifugation for 30 minutes at 500 rcf at 4°C, cells were washed with APBS and re-suspended in 10 μL 2% BSA in APBS (Fig. 5.2).

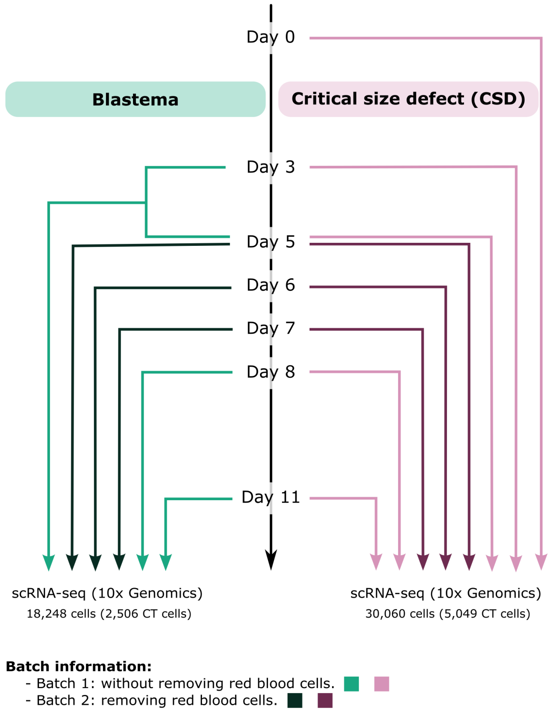


Figure 5.2: Time-points and batches used in the experiment. One replicate with cells extracted from 12-24 limbs was used for each timepoint and batch. A single uninjured sample (day 0, cells were extracted from 12 limbs) was included with the CSD samples during library preparation. Blastema samples day 3 and day 5 (batch 1) were merged before library preparation due to low cell numbers. Samples belonging to batch 1 (light color) did not undergo erythrocyte removal before library preparation. Samples belonging to batch 2 (darker color) underwent erythrocyte removal before library preparation

.

qPCR

Tissue dissociation and cell sorting: The CSD gap tissue or blastema tissue was collected at 11 dpi, chopped to small pieces, and dissociated for 45 minutes in digesting solution (0.26WU/mL Liberase TM in 0.7x PBS) on a shaker at room temperature and mixed by shaking every 10-15 minutes. This reaction was stopped by adding 2x volume serum-free AMEM. The samples were filtered through 70 µm MACS Smart-Strainers (Miltenyi Biotec, 130-098-458) and the cell suspension was pelleted using centrifugation at 300 rcf for 5 minutes at room temperature, then resuspended in serum-free medium and kept on ice. To purify the CT (Cherry$^+$) cells, the cell digestion was filtered through a 45 µm filter immediately prior to sorting using FACSAriall with a 100 µm nozzle size directly into 250 µL Trizol (ThermoFischer Scientific). 3 limbs were used per replicate.

RNA isolation, RT, and qPCR: RNA was isolated per manufacturer's recommendations. Briefly, 100 µL Chloroform (Sigma-Aldrich) was added to the samples, which were then mixed by vortexing, and centrifugated at 21,000g and 6°C. for 15 minutes. The aqueous phase was transferred to new 1.5mL DNA Lo-Bind tubes (Eppendorf), and then 5 µg of RNase-free glycogen and 250 µL 2-propanol were added. The samples were mixed by inverting the tubes, and centrifugated at 21,000g and 6 for

15 minutes to precipitate the RNA. The pellets were washed with 70% ethanol and centrifugated at 21,000g and 6°C for 10 minutes. Lastly, the ethanol was removed, and the pellets were dried at room temperature, and finally resuspended in 10-15 µL RNAse-free water. The RNA concentration was measured using a NanoDropTM Spectrophotometer and 1 µg RNA was used as a template for the RT reaction (LunaScript® kit, New England Biolabs). The cDNA was diluted 1:10 with nuclease-free water and used for qPCR (Luna Universal qPCR kit, with 0.25 μM qPCR primers and 40 cycles, New England Biolabs), run on Bio-Rad CFX96 Touch Real-Time PCR qPCR system (Biorad) using Hard-Shell® 96-Well PCR Plates (Biorad). The primers were designed using the latest axolotl genome with transcriptome annotation version 4.7 (Sup. Table S5.1).

5.1.2 scRNA sequencing

scRNA-seq was performed using the 10xGenomics platform in collaboration with the NGS Vienna Biocenter Campus facility. Cell encapsulation, cDNA generation, pre-amplification and library preparation were performed using the Chromium Single Cell 3’ v3 Reagent Kit (10xGenomics) following standard manufacturer instructions. Sequencing libraries were sequenced on an Illumina NovaSeq S4 PE150XP sequencing platform following 10xGenomics recommendations. Fastq files containing reads for each library were obtained by running Cellranger v6 (10xGenomics) on the raw sequencing bcl files with default parameters. Reads were mapped and quantified on the latest axolotl genome release (v6.0) [179]. These pre-processing steps were performed by Tobias Gerber, PhD and the NGS Vienna Biocenter Campus facility.

5.1.3 scRNA-seq analysis

These steps were performed in collaboration with Tobias Gerber, PhD.

Cell quality control, dimensionality reduction, clustering, and cell type identification

Single cell pre-processing and clustering was done using the R package Seurat v4 [169]. The following two criteria were used to minimize the number of low-quality cells and improve the subsequent analysis: (1) cells with total counts greater than 50,000 or less than 1,000, and (2) cells with a mitochondrial transcript content above 10% were excluded. Blastema and CSD data sets were separately analyzed with the same parameters. Briefly, raw counts were log-normalized and scaled with regressing out differences in RNA counts, followed by PCA computed using all genes. The two batches (Fig. 5.2) were integrated using Harmony [237] with default parameters and 100 PCs as input. For visualization purposes, the integrated dataset was subjected to dimensionality reduction by UMAP using 100 Harmony corrected PCs as input. Louvain clusters were estimated with a resolution of 0.25 and 100 Harmony corrected PCs. Marker genes of each cluster were calculated with the Seurat function FindAllMarkers() using Wilcoxon Rank Sum test. Only positive entries with $\log_2$FC > 0.5 were reported. Clusters were annotated to their corresponding broad cell types using canonical marker genes (Table 5.1).

Table 5.1: Marker genes used for broad cell type and CT subtype annotation. Marker genes were extracted from [14, 49, 163, 171, 172].

Cell type	**Marker genes**
B cell	*Bank1, Blnk, Pax5*
Connective tissue cell	*Col1a2, Col3a1, Dpt, Igfbp3, Lum, Mfap5, Prrx1*
Endothelial cell	*Cdh5, Eng, Lyve1, Pecam1, Tie1, Vwf*
Eosinophil	*C3ar1, Cebpe, Il5ra*
Epidermal cell	*Cdh1, Krt5, Krt17*
Erythrocyte	*Alas2, Hbg1*
Hematopoietic progenitor	*Exd2, Flt3, Gfi1*
Natural killer cell	*Gzma, Gzmb, Itgam, Itgax*
Macrophage	*Adgre1, Csf1r, Itgal, Itgam, Siglec1*
Mast cell	*Kit, Lmo4*
Neutrophil	*Arg1, Cebpe, Csf3r, Itgam, Mmp1*
Schwann cell	*Lamc1, Mpz*
Skeletal muscle cell	*Myl1, Tnni2, Mybph*
T cell	*Ccr7, Gata3, Il7r*
Thrombocyte	*Gp1ba, Gp5, Gp9, Itga2b*
Chondrocyte	*Agc1, Col2a1, Sox9*
Dermal fibroblast	*Twist2, Ptgds*
Interstitial fibroblast	*Col4a2, Igfbp3, Mfap5*
Osteoblast	*Runs2, Tbx2, Tbx3, Tbx5*
Osteoclast	*Acp5, Calcr, Ctr, Mmp9, Runx2b*
Periskeletal cell	*Col8a2*
Tenocyte	*Tnmd*

Virtual purification of CT, cell type identification, and trajectory analysis

CT cells were isolated in the blastema and CSD datasets, respectively, based on the cell type assignments described above. The two datasets were merged and subsequently reprocessed as done before, except for a slightly increased lambda value of 5 during integration with Harmony [237]. UMAP embedding and Louvain clustering were performed as previously detailed, except for an increased resolution of 1. This revealed a few low-quality cells that were removed from the analysis and the embedding and clustering was repeated afterwards. These steps were performed by Tobias Gerber, PhD. Marker genes of each cluster were calculated as previously stated. Following sub-clustering, all blastema CT cells and CSD interstitial fibroblast sub-populations were kept for all subsequent analyses. We refer to these cells as CT cells. These populations were selected based on manual inspection of known marker genes (Table 5.1).

Trajectory inference was applied to CT cells. Trajectory inference was performed using the R package destiny [238]. Pseudo-time estimates for CT cells were calculated in each injury (blastema and CSD) separately and cell number normalized ranks were afterwards added to the integrated CT object introduced before. Trajectory inference was performed by Tobias Gerber, PhD.

5.1.4 GRN inference

Whole-transcriptome GRN inference

Large-scale GRNs were independently reconstructed for blastema and CSD CT cells using TENET (version 2.4), available at `https://github.com/neocaleb/TENET` [133]. TENET measures the causal dependency between pairs of genes using transfer entropy (TE), a non-linear metric based on mutual information [239]. Mutual information symmetrically quantifies the overlap in information content of two variables, without considering the dynamics and directionality of the interaction (Eq. 5.1) [239]. While a time-lagged version of mutual information that incorporates the time component exists, it does not explicitly provide the direction of the relationship. TE solves this problem by measuring the extent to which information about the past state of one variable x helps to predict the future of another variable y, beyond what can be predicted by the past of y alone (Eq. 5.2). This explicitly captures the direction of the information flow and indicates which variable influences the other.

$$I_{x,y} = \sum_{x \in X} \sum_{y \in Y} p(x,y) \cdot log \frac{p(x,y)}{p(x)p(y)} = H(X) - H(X|Y) = H(Y) - H(Y|X) \tag{5.1}$$

$$TE_{x,y} = H(y_t|y_{t-1:t-L}) - H(y_t|y_{t-1:t-L}, x_{t-1:t-L}) \tag{5.2}$$

In both equations, H represents Shannon's entropy (Eq. 5.3 and L refers to the history length or number of past events considered.

$$H(X) = -\sum_i p(x_i) \cdot log_2(p(x_i)) \tag{5.3}$$

TENET quantifies causal transcriptional relationships between pairs of genes using TE and assumes that causality in gene regulation is time-dependent, since an effect (target gene) cannot occur (be expressed) before its cause (regulator) [133]. GRNs were independently reconstructed for blastema and CSD following these steps (Fig. 5.3):

1. Ordering cells along time: blastema and CSD CT cells were ordered based on their collection time points (3, 5, 6, 7, 8, and 11 dpi). To avoid confounding effects between time and perturbation, the uninjured control (day 0) was not considered for network reconstruction.

2. Selecting relevant genes: following TENET guidelines, genes with high expression (log-normalized counts > 1 in 2% of cells) and variability (deviance $>$ quartile 3 for each injury) were selected for GRN reconstruction [133, 240, 241]. Following previous recommendations [242], 2% of total cells was chosen as a threshold because it is the strictest percentage that still allows to include all CT clusters in at least one of the GRNs (Table 5.2).

3. Calculating TE between gene pairs using TENET with default history length ($L = 1$).

4. Trimming indirect edges: potentially indirect relationships in the shape of fast feedforward loops were eliminated if $TE_{x,z} < min(TE_{x,y}, TEy, z)$.

5. Filtering by statistical significance: known targets have significantly higher TE values than random targets [133]. Therefore, a statistically significance filter of FDR < 0.01 calculated on one-sided t-tests on standardized TE values was applied.

The resulting networks contain directed and weighted edges, where higher TE values indicate more probable regulatory interactions. However, the edges are unsigned as TE does not provide information about the MOR [133].

Table 5.2: Number of cells per cluster and injury in the cartilage-contributing CT populations. A total of 1,506 blastema cells and 1,748 CSD cells were used for GRN reconstruction.

Cluster	Blastema cells	CSD cells	GRN containing the cluster
Blastema cartilage	342	13	Blastema
CT cluster 5	44	2	Blastema
Blastema mesenchyme	903	47	Blastema
Early injury fibroblasts	182	423	Blastema and CSD
CSD fibroblasts I	9	796	CSD
CSD fibroblasts II	1	403	CSD
CT cluster 6	25	64	CSD

Topology and power-law analysis

Network topology analysis was performed using Cytoscape v3.9.0 Network Analyzer tool and the igraph R package v1.4.1 in R v4.0.3. Power-law distributions were fit using the igraph R package v1.4.1 and the poweRlaw R package version v0.70.6 in R v4.0.3. Statistical testing of the power-law goodness-of-fit was done using the Kolmogorov-Smirnov test. All network visualizations were done with Cytoscape v3.9.0.

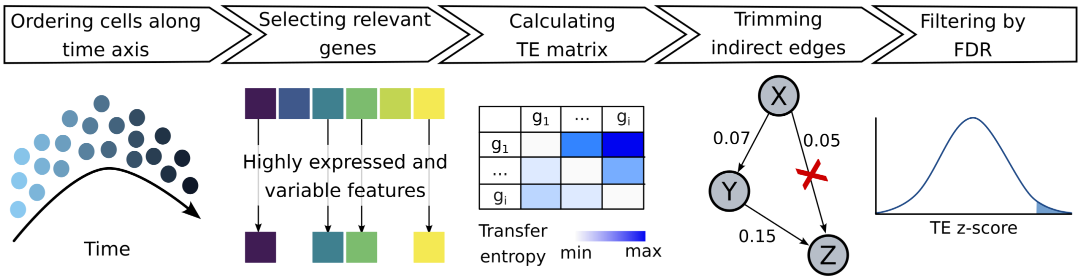


Figure 5.3: Workflow for GRN reconstruction with TENET [133]. Before GRN reconstruction cells are ordered along time and relevant features are selected. Then, TENET is used to calculate the TE matrix that will give rise to the directed, weighted, and unsigned GRN. The last two steps are related to edge filtering and consist of the removal of indirect relationships and the selection of statistically significant relationship.

Candidate prioritization

Network and gene expression information were integrated to generate a ranking of candidate genes for their perturbation (Fig. 5.4). For each GRN, genes were ranked based on their betweenness centrality (Eq. 2.1), which measures how often a gene acts as a bridge between any other two genes. Genes absent in one network were added at the end of the ranking with rank $N+1$, being N the number of genes in that network. Differences in betweenness were calculated by comparing the ranking of each gene in the two GRNs (Eq. 5.4, where $R(Betweenness)$ represents ranked betweenness) and standardized. The top 25% genes with largest (positive and negative) standardized differences were pre-selected as candidates.

$$\Delta Betweenness = R(Betweenness)_{CSD} - R(Betweenness)_{BL} \tag{5.4}$$

Gene expression information was incorporated into the prioritization pipeline by calculating the differential gene expression between early injury blastema and CSD CT cells using the Wilcox Rank Summed test in the FindMarkers() function from the Seurat v4 R package [169]. Genes exhibiting a significant differential expression (Bonferroni corrected p-value < 0.05), an absolute $\log_2$FC $>$ than 0.5 and included among the topologically pre-selected candidates were selected for further analysis. Finally, genes annotated as “Transcription factor” or “Cofactor” in AnimalTFDB v4.0 [243] were prioritized for further investigation.

Wnt pathway modeling and prediction of *Lef1* and *Tcf7l2* target genes

We used Convolutional Neural Network for Co-expression (CNNC), available at `https://github.com/xiaoyeye/CNNC` [144], with the aims of (1) validating the differential status of the Wnt signaling network in both injuries and (2) finding TF-specific targets that could explain the different regener-

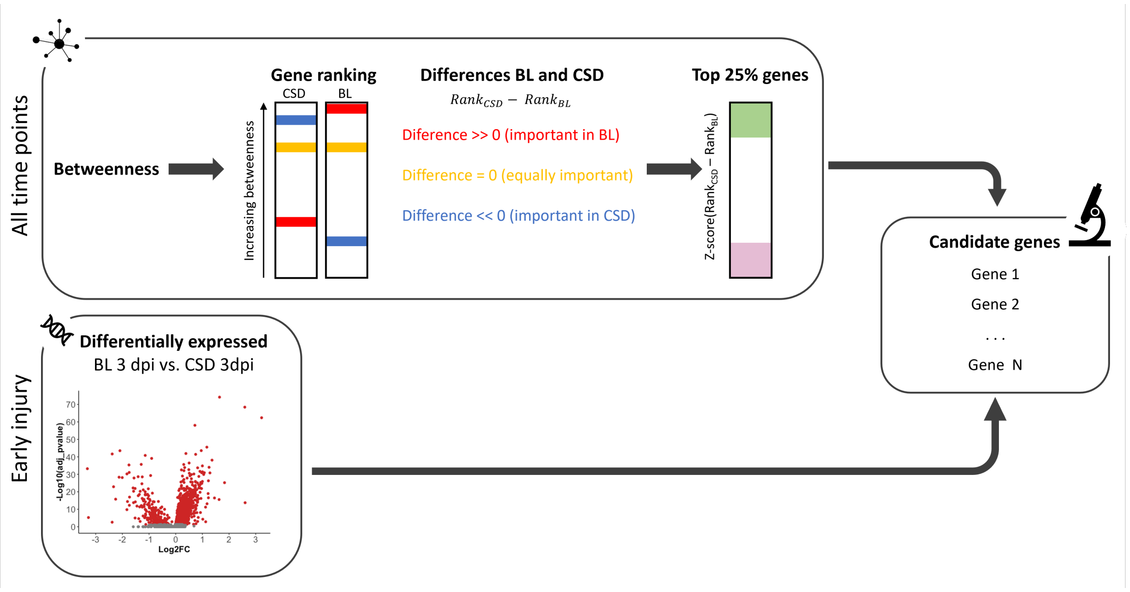


Figure 5.4: Pipeline for candidate selection. Genes were independently ranked by betweenness centrality in the blastema and CSD GRNs. Genes absent in a network received the ranking $N + 1$, with N being the number of genes in the network. The two rankings were compared to find the genes with the largest difference in relative betweenness between the two GRNs. Top 25% genes for each network were pre-selected as candidates. This list was pruned to keep only those genes with a significant differential expression (Bonferroni-adjusted p-value < 0.05) and absolute log2FC > 0.5 between blastema and CSD at 3 dpi.

ative potential of both injuries. CNNC predicts gene relationships by leveraging the information contained in scRNA-seq and the way deep learning discovers intricate patterns in large datasets. Specifically, deep convolutional neural networks (CNN) excel at processing images and have proved themselves useful for supervised classification problems [244]. The large number of cells profiled in scRNA-seq experiments allows to encode gene co-expression as a 2-dimensional co-occurrence histogram in which the expression levels of genes a and b are summarized together. This data transformation scheme fully exploits the scRNA-seq information and the ability of CNN to process spatial information (Fig. 5.5).

CNNC was applied in this project to predict gene regulatory relationships. Because this constitutes a supervised classification problem, training labels were obtained from Kawaguchi *et al.*, who combined ATAC-seq and scRNA-seq data of axolotl upper limb CT cells to build a PKN of transcriptional regulation during axolotl limb regeneration [53]. Briefly, FACS-sorted CT cells of upper limb 5, 9 and 13 dpi blastemas were subjected to chromatin profiling by ATAC-seq. After trimming and alignment of paired-end ATAC-seq reads to the axolotl genome, narrow peaks were called using macs2 [245] and annotated to genes using ChIPseeker (v1.22.1) [246, 247]. ATAC-seq promoter (-2,000 to +300 bp from the annotated TSS) and enhancer peaks that were assigned to genes expressed in the axolotl regenerating limb were scanned for TF motifs using FIMO [248]. Only relevant TFs for axolotl limb regeneration, selected based on their detection in scRNA-seq [14], were used for analysis. The JASPAR2022 CORE and UNVALIDATED motifs was used as a reference [113]. The resulting

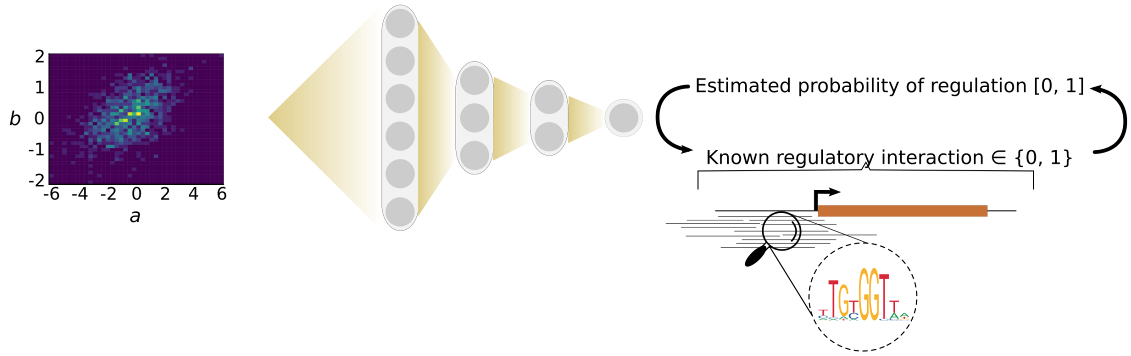


Figure 5.5: Summary of CNNC training for the prediction of transcriptional regulation [144]. 2-dimensional co-occurrence histograms are generated for each training pair using the expression levels in scRNA-seq. A binary label for each training pair is obtained from predicted transcriptional regulatory relationships obtained from the presence of TF motifs in open regulatory regions of ATAC-seq data. These images together with their labels are used to train a CNN. Once trained, the model can be used to predict regulatory relationships between new gene pairs.

PKN contained 197 TFs and 19,431 target genes. From this network, TF-target pairs for which both genes have at least 1 log-normalized counts in more than 5% of cells were selected. Negative training examples were generated by randomly selecting pairs of non-interacting genes that pass this expression threshold. A total of 121,098 training cases were used, with a balance between positive and negative training examples.

2-dimensional co-occurrence histograms for the training pairs were built from the blastema 8 and 11 dpi scRNA-seq data generated in this project following the CNNC pipeline. Briefly, normalized gene expression values for each gene were uniformly divided into 32 bins. For each pair of genes, 2-dimensional co-occurrence histograms were generated by assigning each cell to an entry in the matrix and counting the number of samples for each entry. To mitigate the dropout effect, a pseudocount was added to all entries and the matrix was log-transformed. These matrices were visualized as a 32 x 32 histogram and used to train a CNN with the following architecture: one 32 x 32 input layer, 10 intermediate layers (convolutional and maxpooling), 1 flatten layer and a final sigmoid layer. All layers used ReLu as an activation function except the output sigmoid layer. The network was trained using 200 epochs, minibatches of size 1,024 and 3-fold cross-validation. Once trained, the model was used for two analyses. Firstly, it was used to predict regulatory relationships on 32 x 32 2-dimensional histograms of 3 dpi blastema or CSD scRNA-seq calculated for all pairs of genes involved in the “Wnt signaling pathway” as described in the KEGG Pathway Database (entry reference map04310) [249, 250], if the genes had at least 1 log-normalized count in at least 5% of blastema or CSD 3 dpi CT cells. Secondly, it was used to predict *Lef1* and *Tcf7l2* target genes by generating 2-dimensional histograms of the co-expression between *Lef1* or *Tcf7l2* and all other genes with more than 1 log-normalized count in at least 5% of blastema or CSD 3-dpi CT cells, respectively. In both cases, interactions with a probability > 0.5 were considered positive. Enrichment of *Lef1*-specific and *Tcf7l2*-specific target genes in blastema and CSD was calculated with GSEA on the $\log_2$FC between blastema and CSD at 3dpi as implemented in the R package fgsea v1.29.1 [185].

In silico perturbation of candidate TFs

The effect of TF perturbations on the CT transcriptome was predicted using celloracle (version 0.14), available at `https://morris-lab.github.io/CellOracle.documentation/#` [21]. A detailed description of the GRN inference algorithm implemented by celloracle can be found in Chapter 3. For this analysis, the integrated blastema and CSD CT data set and an integrated version of the blastema and CSD CT GRNs inferred with TENET [133] containing 2,785 genes and 6,518 edges were used as input. After fitting the linear models, robust connections were identified by filtering Bonferroni-corrected p-values using a significance threshold of 0.05.

An optimized k value of 81 was used to quantify the effect of the perturbations by correlating the perturbation vector with the gene expression differences between each cell and its 81 k-nearest neighbors. The perturbation vectors were compared with the blastema and CSD trajectories to calculate a PS. Their statistical significance was empirically tested by obtaining a null distribution of the PS calculated with a randomized GRN model. The obtained p-values were Bonferroni-corrected to account for multiple testing.

5.2 Results

5.2.1 Cellular landscapes of blastema and CSD

We generated a comprehensive atlas of the cell types present in blastema (regenerative) and CSD (non-regenerative) axolotl limbs at various time-points following injury. For this, we used high-throughput droplet based scRNA-seq (10xGenomics) of blastema and CSD limbs collected 3, 5, 6, 7, 8, and 11 dpi, along with cells from intact limbs as a physiological control (day 0). Through computational integration of 48,308 cells and unbiased clustering, we identified 27 molecularly distinct clusters that were manually aggregated into 8 broad cell types using the expression of known marker genes (Fig. 5.6a). These cell types included CT (e.g., *Col1a2*$^+$, *Lum*$^+$, and *Prrx1*$^+$ cells), endothelium (e.g., *Cdh5*$^+$, *Eng*$^+$, *Lyve1*$^+$ cells), epidermis (e.g., *Cdh1*$^+$, *Krt5*$^+$, and *Krt17*$^+$ cells), immune cells (including macrophages (*Adgre1*$^+$, *Itgal*$^+$, *Itgam*$^+$) and neutrophils (*Arg1*$^+$, *Cebpe*$^+$, *Csf3r*$^+$)), skeletal muscle (e.g., *Myl1*$^+$, *Tnni2*$^+$, and *Mybph*$^+$ cells), erythrocytes (e.g., *Alas2*$^+$ and *Hbg1*$^+$ cells), thrombocytes (e.g., *Gp1ba*$^+$, *Gp5*$^+$, *Gp9*$^+$ cells) and Schwann cells (e.g., *Lamc1*$^+$, and *Mpz*$^+$ cells) (Fig. 5.6a, b). Both injuries and batches were homogeneously integrated (Fig. 5.6c, d), indicating a shared cellular landscape with substantial overlap in composition. Notably, the CT, epidermis and macrophages emerged as the most distinct cell types between injuries (Fig. 5.6c), suggesting their involvement in the observed regenerative disparity.

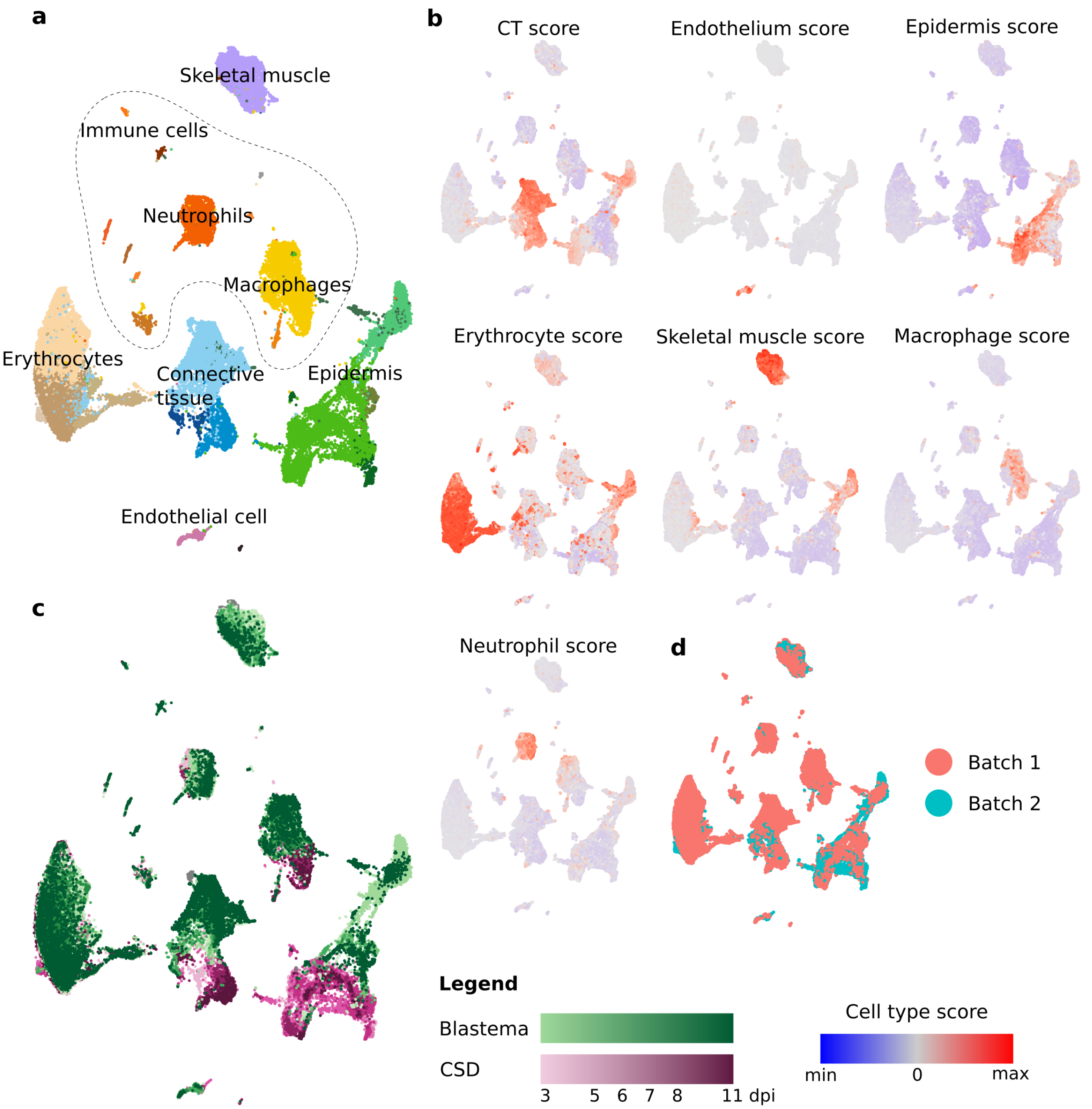


Figure 5.6: Blastema and CSD have a similar cellular landscape. a) UMAP of integrated blastema and CSD cells colored by cell type. Different shades indicate the molecularly different clusters. **b)** UMAP of integrated blastema and CSD cells colored by injury and time-point. **c-i)** UMAPs of integrated blastema and CSD cells colored by cell type scores calculated from the expression of known cell type marker genes. **j)** UMAP of integrated blastema and CSD cells colored by batch.

5.2.2 Blastema and CSD CT progressively diverge with time

Previous studies have determined that pre-existing CT dedifferentiates, proliferates, and guides the patterning of the regenerating limb [14]. We examined the molecular differences between blastema and CSD CT by computationally isolating CT cells based on the expression of known marker genes (Table 5.1). Sub-clustering of CT cells identified 12 molecularly different populations, from which we focused on blastema cells and CSD interstitial fibroblasts, as they could potentially migrate and promote bone healing [251]. We refer to these populations as CT cells (Fig. 5.7a). These clusters were annotated as blastema mesenchyme (e.g., *Kazald1*$^+$ and *Hoxd10*$^+$ cells), blastema cartilage (e.g., *Kazald1*$^+$ and *Runx2*$^+$ cells), early injury fibroblasts (e.g., *Lrrc32*$^+$ cells), and two clusters of CSD fibroblasts (e.g., *Tnxb*$^+$ and *Lats2*$^+$ cells) (Fig. 5.7d). Notably, blastema and CSD CT formed a homogeneous cluster at 3 dpi but progressively diverged over time (Fig. 5.7b). Trajectory inference and pseudotime analysis were independently performed for each injury and aligned well with real-time information (Fig. 5.7c). Furthermore, cell cycle analysis, based on the expression of cell cycle-dependent genes, revealed that blastema cells were predominantly in S or G2/M phases, with few cells in G1. In contrast, CSD cells transitioned from S to G2/M to G1 phases as time post-injury increased (Fig. 5.7e). These findings lead us to hypothesize that different GRNs govern blastema and CSD CT identities and might be responsible for their differential behavior and regenerative potential.

5.2.3 Topological differences suggest a loss of blastema hubs in the CSD GRN

Bone regeneration is a complex phenomenon that depends on the orchestrated activity of numerous components that interact with each other. Aiming to find the molecular drivers that could explain the differences between blastema and CSD CT, we inferred injury-specific GRNs using a method based on TE [133]. The GRNs for blastema and CSD consisted of 989 and 2,524 genes, respectively, of which 728 were present in both networks (Fig. 5.8a, b, Table 5.3).

Table 5.3: Network topology analysis.

Metric	Blastema GRN	CSD GRN	Interpretation
Number of nodes	989	2,524	Number of genes in the network.
Number of edges	2,390	4,128	Number of connections in the network.
Average number of neighbors	4.698	3.214	Indicates the average connectivity of a node.
Network diameter	7	8	Largest distance between two nodes in the network.
Characteristic path length	3.816	4.170	Average distance between two nodes in the network.

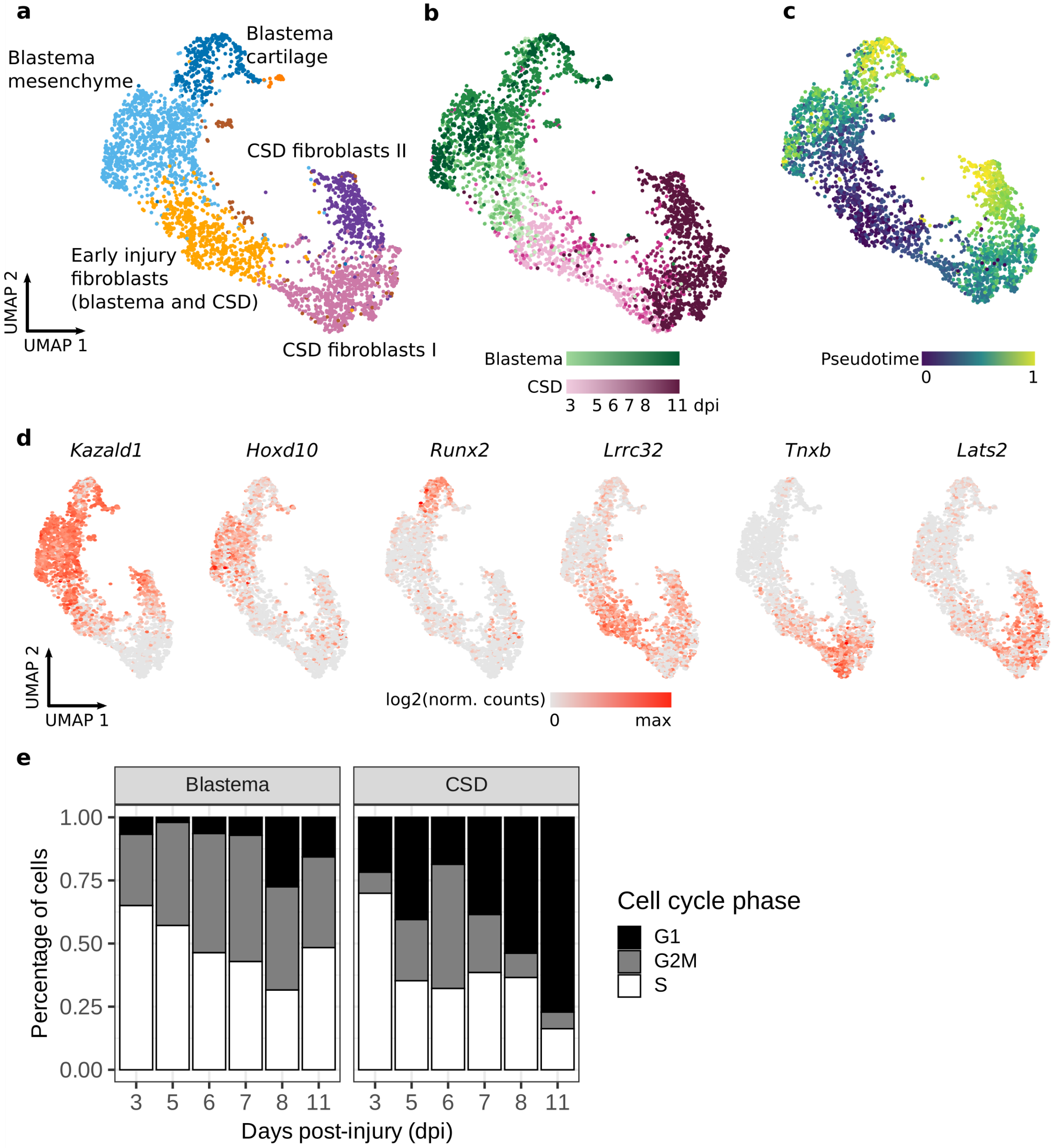


Figure 5.7: Blastema and CSD CT become progressively different as time post-injury passes. **a)** UMAP of integrated blastema and CSD CT cells colored by cluster. **b)** UMAP of integrated blastema and CSD CT cells colored by injury and time-point. **c)** UMAP of integrated blastema and CSD CT cells colored by pseudotime estimates. **d)** UMAP plot of the expression of markers for the clusters in panel a. **e)** Classification of blastema or CSD CT cells in the different phases of the cell cycle (G1, S or G2/M) at each time point.

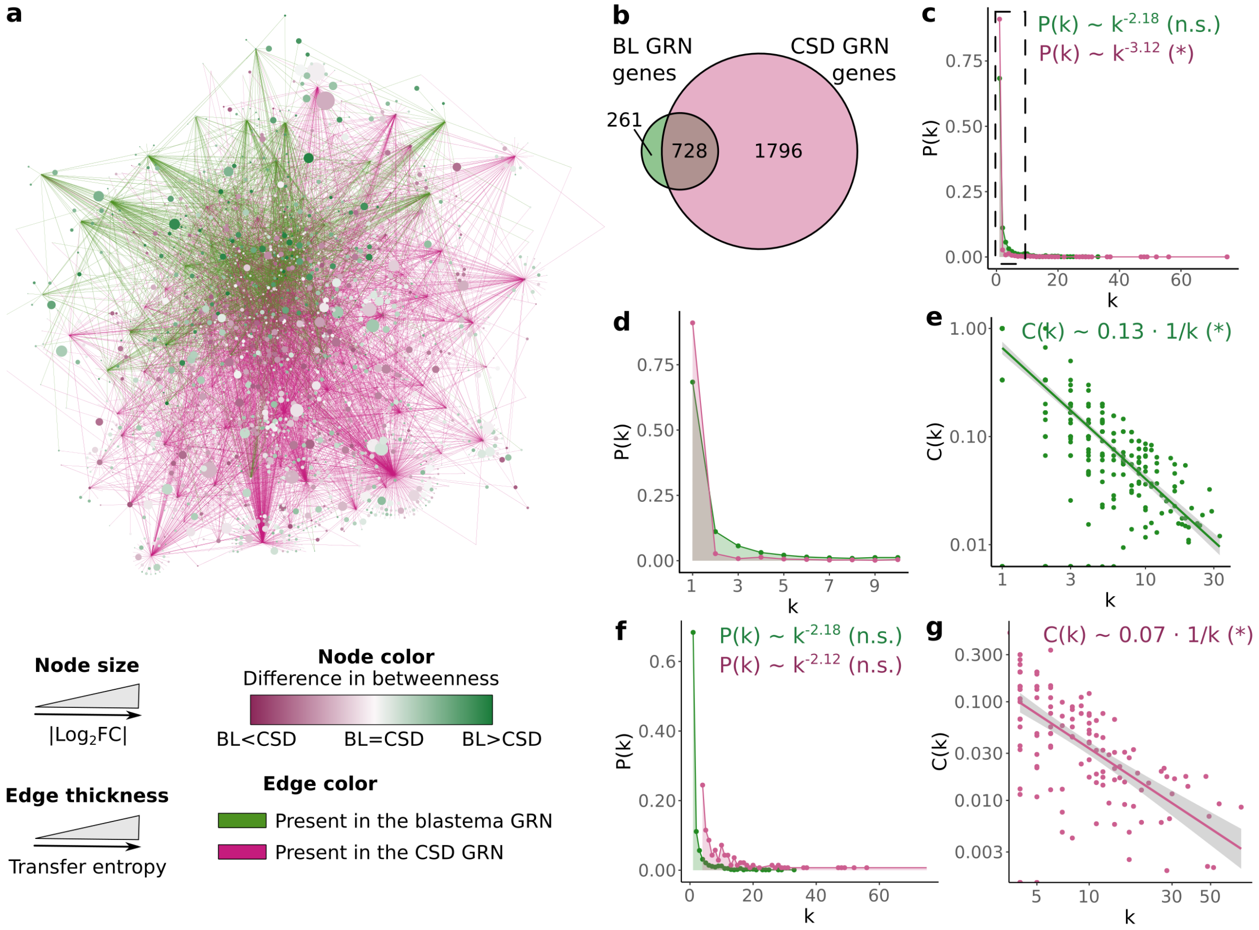


Figure 5.8: The blastema and CSD CT GRNs have different hub genes. **a)** Integrated visualization of blastema and CSD CT GRNs. The node size indicates Log2FC between blastema and CSD CT cells at 3 dpi. The node color indicates the difference in ranked betweenness between the blastema and CSD GRNs. The edge color indicates whether it is present in the blastema or CSD CT GRN. The edge thickness represents the TE of each edge. **b)** Venn diagram of the number of genes in blastema and CSD CT GRNs. **c)** Power-law fit of blastema (green) and CSD (magenta) CT GRN out-degree. The p-value corresponds to a Kolmogórov-Smirnov goodness-of-fit test. A p-value > 0.05 (n.s.) means that the observed distribution fits the power-law distribution, while a p-value < 0.05 (*) means that the observed distribution cannot be properly described by a power-law distribution. **d)** Zoomed-in visualization of the marked area in panel **c**. **e)** Scatter plot of the relationship between clustering coefficient and node out-degree for the blastema CT GRN. The asterisk indicates a slope p-value < 0.05. **f)** Power-law fit of blastema (green) and CSD genes with out-degree $k \geq 4$ (magenta) CT GRN out-degree. The p-value corresponds to a Kolmogórov-Smirnov goodness-of-fit test. A p-value > 0.05 (n.s.) means that the observed distribution fits the power-law distribution, while a p-value < 0.05 (*) means that the observed distribution cannot be properly described by a power-law distribution. **g)** Scatterplot of the relationship between clustering coefficient and out-degree for nodes with an out-degree ≥ 4 in the CSD CT GRN. The asterisk indicates a slope p-value < 0.05.

Beyond network size differences, we explored whether the two GRNs had topological differences that could influence cell behavior. Empirical evidence has shown that the architecture of natural networks is governed by a few simple laws [252]. Most cellular networks have been described to follow a scale-free topology in which a few highly connected "hub" nodes coexist with numerous nodes having relatively few connections. Scale-free networks can be easily identified because their degree distribution (probability that a chosen node has exactly k edges) follows a power law $P(k) \sim k^{-\gamma}$, where $2 < \gamma < 3$ [253]. Additionally, GRNs have a modular structure in which sparsely connected genes are members of highly connected modules, communicated between each other by a few hub TFs. The most characteristic signature of hierarchical networks is the dependence of the clustering coefficient on the node degree: $C(k) \sim k^{-1}$ [150]. The blastema out-degree distribution followed a power-law with degree exponent $\gamma_{Blastema} = 2.17$ (KS p-value = 0.79) and had a clear inverse relationship between $C(k)$ and k (Fig. 5.8c, d). The CSD out-degree distribution, on the other hand, could not be properly described by a power-law ($\gamma_{CSD} = 3.12$ but KS p-value = 3.9e-9) (Fig. 5.8e), but it could be approximated by a power-law ($\gamma_{CSD} = 2.12$ and KS p-value = 0.16) if nodes with an out-degree < 4 were excluded, while retaining the inverse relationship between the clustering coefficient and the out-degree (Fig. 5.8f, g). As the scale-free topology necessarily depends on the presence of hubs, these results suggest a loss of blastema hubs in the CSD GRN.

Centrality analysis identifies key network elements, and several centrality metrics have been created to quantify the relevance of each gene in a complex network [254]. Genes were ranked by their betweenness centrality, which quantifies the number of shortest paths between any two nodes that pass through a third node in the network [255]. Betweenness centrality is higher in essential genes than in non-essential genes [256] and allows to identify two types of hubs: those with a large degree and those that act as bridges between modules or regions in the network [254, 257]. Absolute betweenness centrality should not be directly compared between networks of different size [258], and therefore we used a ranked approach to compare the betweenness of each gene between the two networks. Genes with high differences in betweenness centrality and differential expression between blastema and CSD at the earliest time-point (3 dpi) were of particular interest, as modifying the activity of regeneration initiators could have the largest effect. Among the 227 genes that passed the filtering criteria, 21 were TFs (Fig. 5.9a). From these, the Wnt effectors *Lef1* and *Tcf7l2* caught our attention because the network analysis predicted an opposite centrality despite their known structural and functional similariy [259]: *Lef1* is a hub in the blastema GRN while *Tcf7l2* is a hub in CSD GRN. The scRNA-seq shows an injury-specific expression of these TFs, with *Lef1* being present only in blastema and *Tcf7l2* being predominant in CSD (Fig. 5.9b, c). This effect was validated using qPCR on FACS-purified blastema or CSD CT collected at 11 dpi (Fig. 5.9d). Collectively, these results suggest that *Lef1* and *Tcf7l2* play important roles in the blastema and CSD GRNs, respectively, and their differential abundance might contribute to the differential regenerative behavior of both injuries.

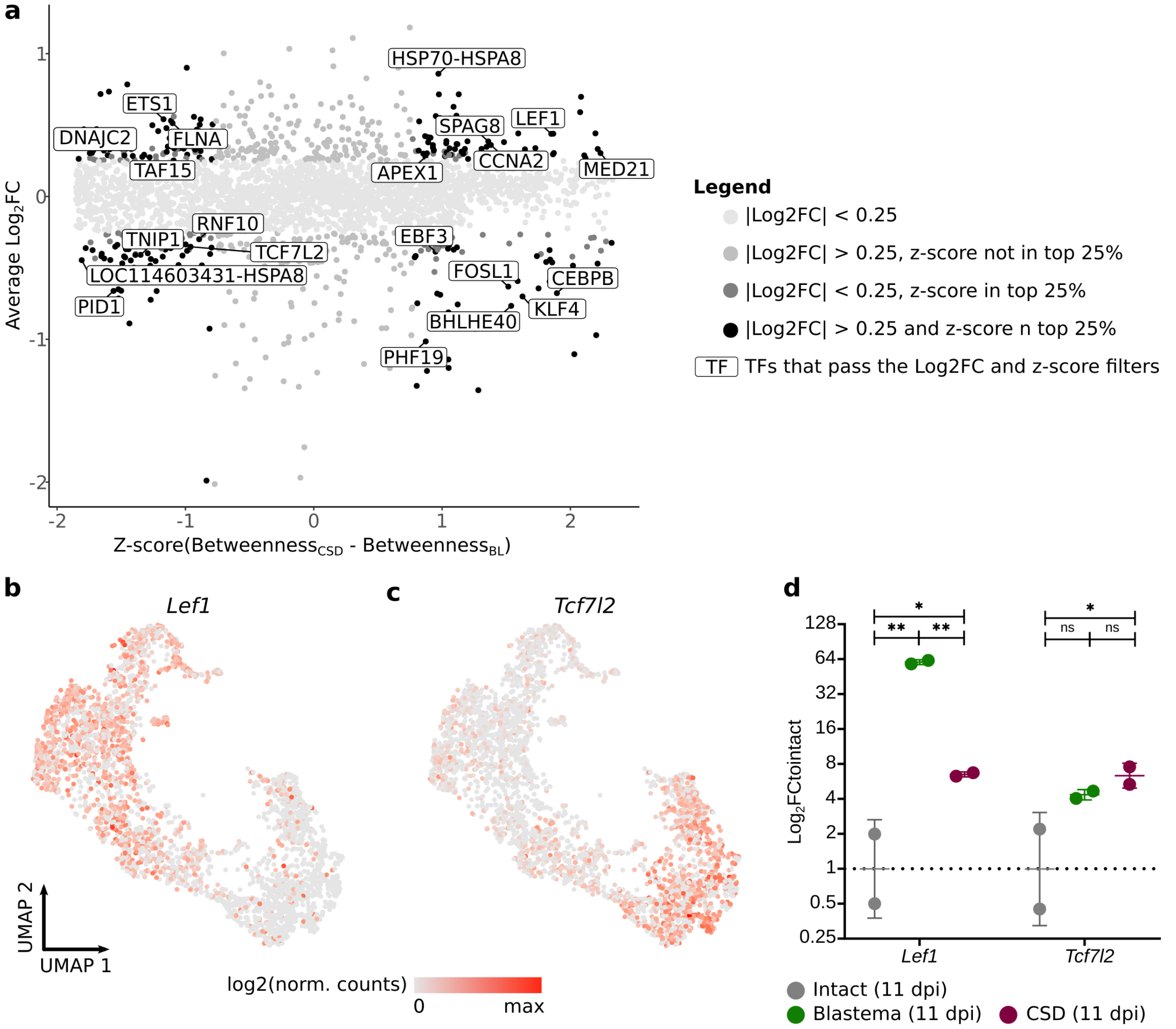


Figure 5.9: *Lef1* and *Tcf7l2* show injury-specific expression. **a)** Scatterplot of the difference in ranked betweenness in the blastema and CSD CT GRNs expressed as a z-score and the $\log_2$FC between blastema and CSD CT at 3 dpi. Genes are colored by the filtering thresholds that they passed during the process of candidate selection. The names of top candidate TFs are indicated. **b)** UMAP plot of the expression of *Lef1* in blastema and CSD CT. **c)** UMAP of the expression of *Tcf7l2* in blastema and CSD CT. **d)** qPCR of *Lef1* and *Tcf7l2* in intact CT (0 dpi), blastema CT (11 dpi) and CSD CT (11 dpi). A two-way ANOVA with Turkey's multiple comparison tests were used to asses statistical significance (Adjusted p-values: * indicates < 0.05, ** indicates < 0.0001, and ns indicates not significant).

5.2.4 The Wnt effectors *Lef1* and *Tcf7l2* regulate different gene programs in blastema and CSD

Lef1 and *Tcf7l2* belong to the TCF/LEF family, which act as nuclear effectors of the canonical Wnt pathway (Fig. 2.5) [260]. Hypothesizing that *Lef1* and *Tcf7l2* could have differences in functionality that impact bone regeneration in axolotl, we explored whether Wnt signaling is being differentially employed and whether this activates different gene expression programs. We focused on the first time point (3 dpi) because, despite belonging to the early injury CT cluster, blastema and CSD cells already show some segregation (Fig. 5.7a, b). To predict gene regulation probabilities at this single time-point we used a deep learning-based methodology [144] trained with 121,098 TF – target gene pairs from axolotl blastema scRNA-seq and bulk ATAC-seq (accuracy = 0.63, F1-score = 0.68 and AUROC = 0.67).

Firstly, we investigated whether the two injuries show differences in the Wnt pathway by predicting the probability of regulation between genes participating in it. Both injuries showed different regulation of the Wnt signaling pathway at various levels (Fig. 5.10). Specifically, in terms of ligand-receptor interactions, blastema showed activation of the WNT5A-ROR2 interaction, known to trigger the PCP pathway, essential for limb morphogenesis [261]. In CSD, Wnt signaling appeared to be weaker, due to the higher expression of the Wnt inhibitors sFRPs and DKKs [260]. Regarding transcriptional regulation, blastema seemed to show a preference for LEF1 whereas CSD exhibited predominant TCF7L2 activity (Fig. 5.10).

Secondly, we investigated the extent to which *Lef1* and *Tcf7l2* regulate different transcriptional programs. We used the same methodology to predict potential transcriptional targets of LEF1 in blastema and TCF7L2 in CSD. Despite a large overlap in their gene sets, we found 102 genes uniquely associated to LEF1 and 1,245 genes uniquely associated to TCF7L2 (Fig. 5.11a). Furthermore, their expression was found to be enriched in blastema and CSD at 3dpi, respectively (Fig. 5.11b). By employing a PKN derived from blastema ATAC-seq, we identified putative direct targets based on the presence of LEF1 or TCF7L2 binding sites in the gene promoters. Notably, LEF1-specific targets included matrix metalloproteases (e.g., *Mmp11*, *Mmp13*) and cell cycle regulators (e.g., *Aurka*, *Cenpf*) (Fig. 5.11c-f). In contrast, TF7L2-specific targets encompassed autophagy modulators (e.g., *Dram2*), Hippo-signaling proteins (e.g., *Sav1*), Wnt inhibitors (e.g., *Dkk1*) and stimulators of cell proliferation (e.g., *Gas6*) (Fig. 5.11g-j). Taken together, these results support the hypothesis that Wnt signaling might be differentially active and govern different processes in the two injuries, with a probable downregulation of pro-regenerative behaviors characteristic of blastema CT cells in CSD fibroblasts.

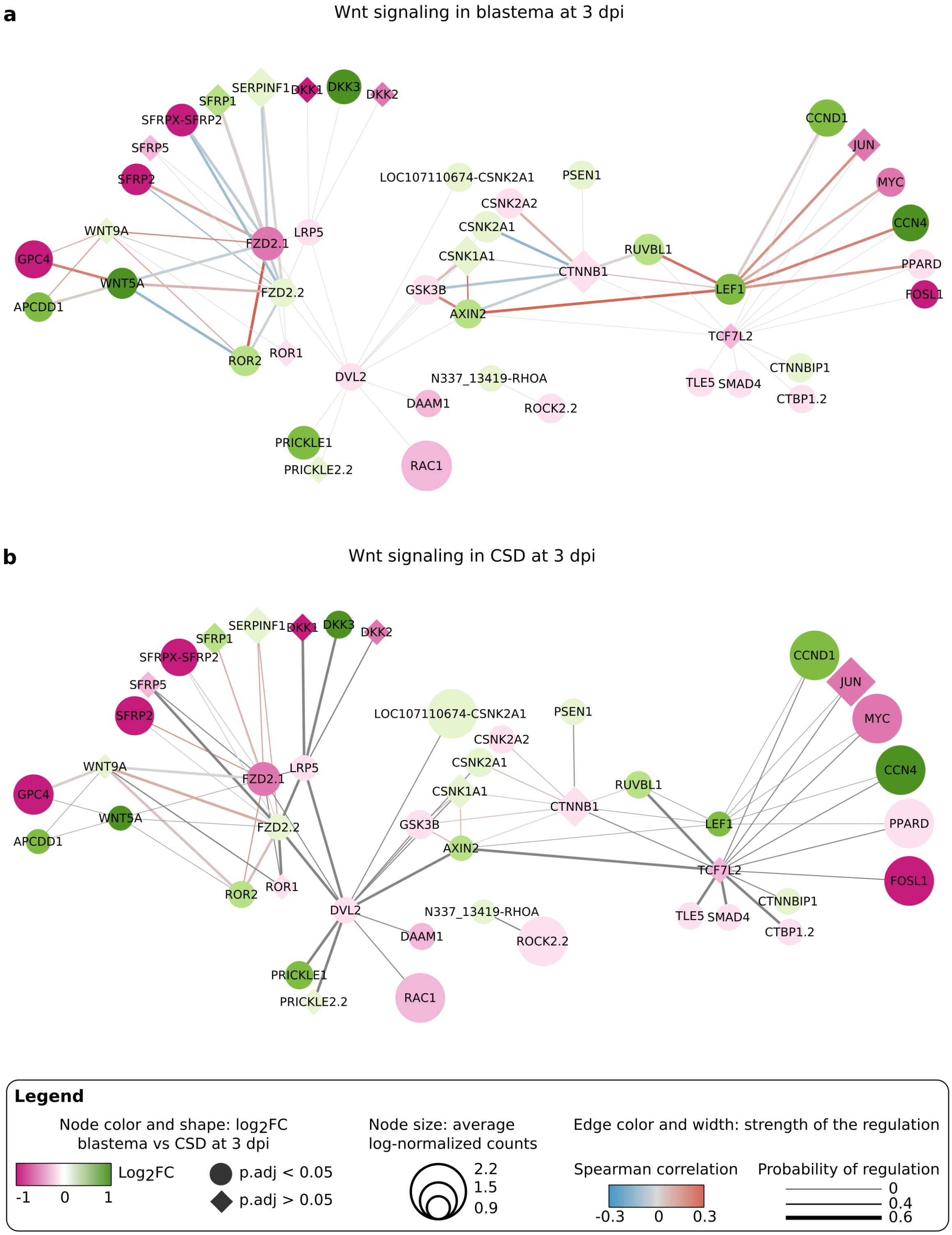


Figure 5.10: Differential status of the Wnt signaling pathway in **a)** blastema and **b)** CSD CT, based on network reconstruction using CNNC [144]. Nodes are colored based on the $\log_2$FC between blastema and CSD CT at 3dpi. Node shape indicates whether the differential expression is significant (Bonferroni-adjusted p-value < 0.05) or not. Edge width indicates the predicted probability of regulation. Edge color indicates the spearman correlation of gene expression.

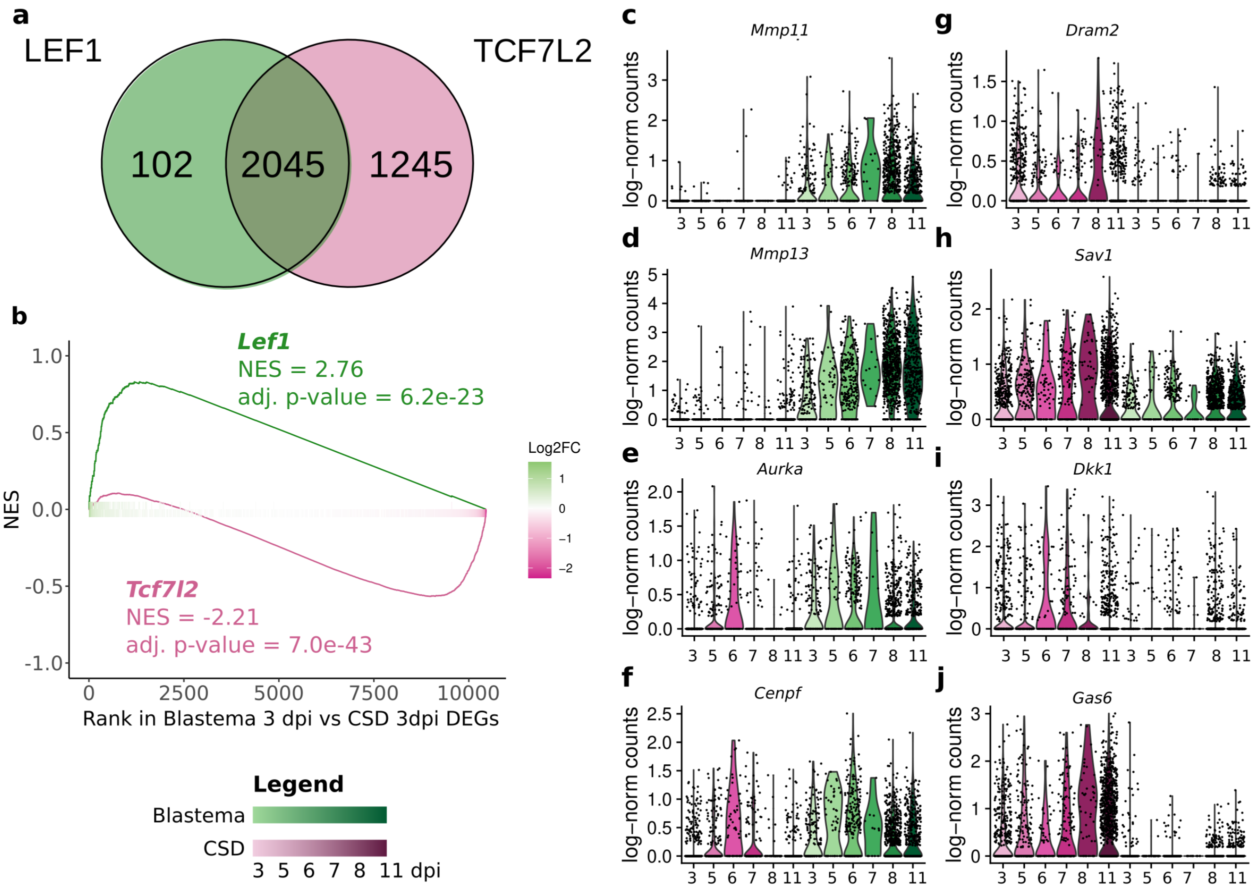


Figure 5.11: *Lef1* and *Tcf7l2* regulate different genetic programs in blastema and CSD. a) Venn diagram showing the number of shared and unique genes associated with *Lef1* and *Tcf7l2*. **b)** Enrichment plot for the Lef1- and *Tcf7l2*-specific targets calculated on the log2FC between blastema and CSD at 3 dpi. **c-j)** Violin plots showing the changes in gene expression in the two injuries over time for **c)** *Mmp11*, **d)** *Mmp13*, **e)** *Aurka*, **f)** *Cenpf*, **g)** *Dram2*, **h)** *Sav1*, **i)** *Dkk1* and **j)** *Gas6*.

5.2.5 *In silico Lef1* and *Tcf7l2* perturbations predict heterogeneous effects on the trajectories of CT subpopulations

Our previous findings led us to hypothesize that *Lef1* and *Tcf7l2* could have antagonistic functions that promote or hinder regeneration, respectively. GRNs based on linear models can be used to predict the effect of perturbations (e.g., TF KO or overexpression) on the transcriptome. We used celloracle [21] to transform a merged version of the blastema and CSD GRNs inferred using TENET (Fig. 5.8) into a GRN in which gene regulation follows a linear relationship and simulate the effect of TF KO or overexpression. The impact of perturbing each TF was estimated by a PS, which compares the perturbation vectors and the 2-dimensional trajectory vectors using their inner product.

Overall, the impact of *Tcf7l2* perturbations was more pronounced than that of *Lef1* (Fig. 5.12. Table 5.4 and Sup. Fig. S5.1). We focused on the effect of early injury fibroblasts and CSD fibroblasts I, as the perturbations yielded mixed effects on the other clusters. Notably, none of the tested perturbations significantly affected early blastema fibroblasts (Fig. 5.12 and Table 5.4). Regarding early CSD fibroblasts, the *Lef1* KO appeared to promote the CSD trajectory (summarized PS = 0.99, adj. p-value = 2.3e-12, Fig. 5.12a) while the *Tcf7l2* KO seemed to oppose it (summarized PS = -1.09, adj. p-value = 7.3e-08, Fig. 5.12c). Interestingly, *Lef1* and *Tcf7l2* perturbations affected different sub-regions of the cluster: *Lef1* KO positively impacted the right-most region of the cluster, while showing no effect on cells near the early fibroblasts (summarized PS = 2.05 with adjusted p-value = 1.3e-8, Fig. 5.12a). Conversely, the effect of *Tcf7l2* KO was limited to CSD fibroblasts I closer to the early injury cluster, with minimal impact on the region affected by *Lef1* perturbations (summarized PS = -3.65, adj. p-value = 9.3e-19, Fig. 5.12c). The effect of *Lef1* or *Tcf7l2* overexpression was exactly the opposite of its KO, as expected (Fig. 5.12b, d).

Given their apparent antagonism, we wondered whether simultaneously perturbing both TFs would potentiate the outcome. Interestingly, double perturbations potentiated or cancelled each other, depending on the UMAP region (Fig. 5.12e, f). The predicted effect of combining *Lef1* KO and *Tcf7l2* overexpression closely resembled that of *Tcf7l2* overexpression alone (Fig. 5.12e and Table 5.4). To induce blastema properties and potentially regeneration, the most interesting perturbation would be to simultaneously overexpress *Lef1* and KO *Tcf7l2* (Fig. 5.12f and Table 5.4). These two perturbations promoted the retention of the early CSD cell phenotype, while having cancelling effects on a subset of CSD fibroblasts I. Overall, these results support the hypothesis that *Lef1* and *Tcf7l2* might have different and occasionally antagonistic roles in early blastema and CSD development, heterogeneously influencing different subpopulations.

Table 5.4: PS for *Lef1* and *Tcf7l2* in early injury and CSD fibroblasts I. FB stands for fibroblasts, KO stands for knockdown, OE stands for overexpression and * indicates adjusted p-value < 0.05.

Perturbation	**Early CSD FB.**	**Early blastema FB.**	**CSD FB. I**
Lef1 KO	PS = 0.99*	PS = -0.04	PS = 2.05*
Lef1 OE	PS = -0.92*	PS = 0	PS = -1.92*
Tcf7l2 KO	PS = - 1.09*	PS = 0.18	PS = -3.65*
Tcf7l2 OE	PS = 1.16*	PS = -0.18	PS = 3.81*
Lef1 KO + *Tcf7l2* OE	PS = 1.32*	PS = 0	PS = 3.95*
Lef1 OE + *Tcf7l2* KO	PS = -1.15*	PS = 0	PS = -1.71*

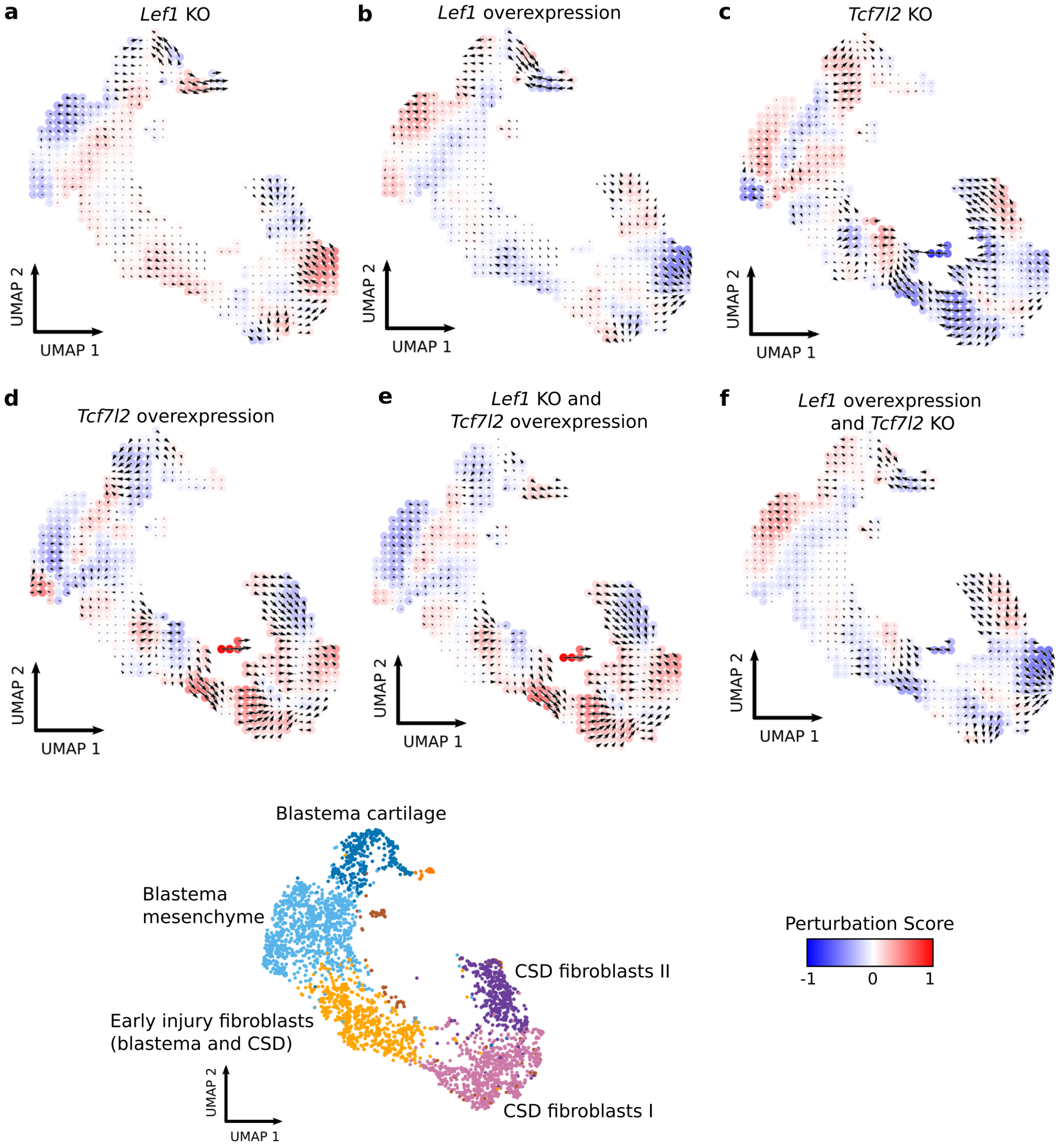


Figure 5.12: Effect of *Lef1* and *Tcf7l2* *in silico* perturbations on the blastema and CSD CT trajectories. Panels **a-f)** show a gridded UMAP visualization of integrated blastema and CSD CT cells. Colors and vectors represent the effect of **a)** *Lef1* KO, **b)** *Lef1* overexpression, **c)** *Tcf7l2* KO, **d)** *Tcf7l2* overexpression, **e)** simultaneous *Lef1* KO and *Tcf7l2* overexpression and **f)** simultaneous *Lef1* overexpression and *Tcf7l2* KO. Red colors indicate that the perturbation promotes the differentiation trajectory of the cells, while blue colors indicate that the perturbation opposes the differentiation trajectory. Vector arrows indicate the magnitude and direction of the transcriptome changes upon perturbation. For reference, the UMAP visualization of integrated blastema and CSD CT cells as described in Fig. 5.7a is included.

5.3 Discussion

This chapter explored the differences between blastema and CSD CT to understand their differential regenerative potential. We used a systems biology approach based on whole-transcriptome GRN inference on a scRNA-seq time series that enabled us to narrow down the scope of potential molecular regulators with differential functions between both injuries to two TFs: *Lef1* and *Tcf7l2*. These TFs belong to the TCF/LEF family of TFs and share a high identity level. Their differential function was further supported by the identification of genes uniquely associated to *Lef1* or *Tcf7l2* and by the simulation of *Lef1* and *Tcf7l2* perturbations on the CT GRN. Overall, these results provide insights into the molecular dynamics underlying the differential regenerative potential of blastema and CSD and suggest these two candidate TFs for their experimental perturbation (Fig. 5.13).

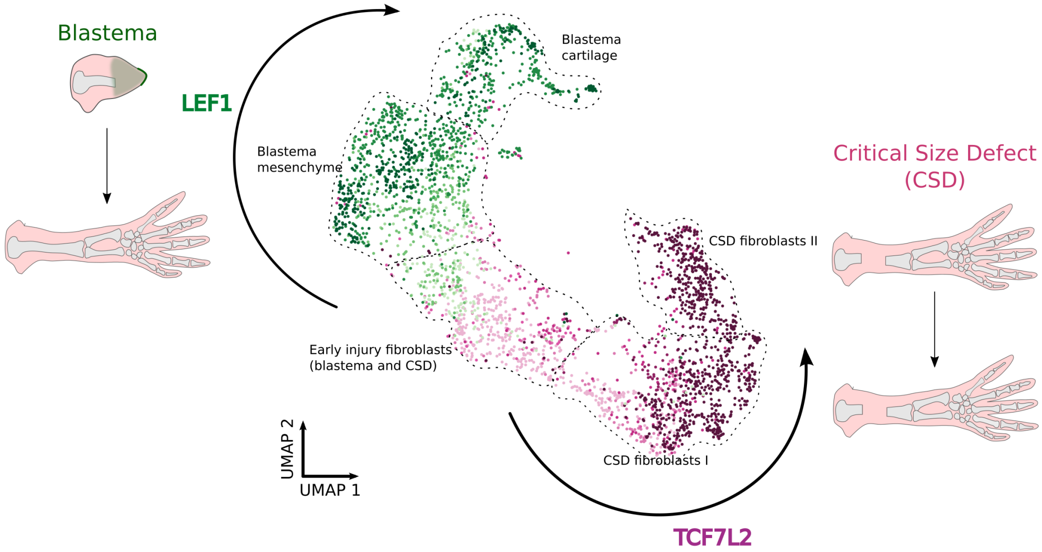


Figure 5.13: *Lef1* and *Tcf7l2* play a different and potentially antagonistic role in mediating the pro- or anti-regenerative response of injuries in the axolotl limb.

Regarding the GRN, we used a data-driven transcriptome-wide approach with no additional information due to the complexity of the axolotl genome. The axolotl genome is over ten-times larger than the human one and has higher levels of duplication [179]. This has caused the appearance of multiple orthologs for certain human genes and the extension of the regulatory domains, as axolotl regulatory elements can be as far as 10 Mb from the gene promoter [176]. Due to these evolutionary differences, the axolotl genome contains axolotl-specific genes that could be innovations of the salamander lineage or have been lost during tetrapod development and which functional relevance remains largely unknown [97]. Therefore, we opted for using an unbiased yet potentially noisy GRN inference approach that can be further filtered by incorporating previous knowledge from axolotl or other species. In this chapter we focused on the role of TFs as regulators, but our GRN could also be used to unveil the regulatory role of molecules such as ligands or receptors.

The datasets generated in this project constitute the first scRNA-seq profiling of axolotl CSD.

The comparison with the blastema cellular landscape reinforces the known importance of CT and epidermis for epimorphic limb regeneration [96] and motivates to further investigate the function of immune cells during axolotl limb regeneration, which is starting to be unveiled [51, 52]. Diving deeper into CT sub-types revealed that, contrary to blastema, CSD CT does not show molecular signs of dedifferentiation and might undergo a single round of cell proliferation. Additionally, assuming that CT cells are not polyploid and that the cell cycle rate would be invariant between injuries, our results could indicate that blastema and CSD have preferential cell cycle checkpoints, G2/M and G1/S, respectively. Regarding the selected times for cellular profiling, we focused on 3-11 dpi because this window corresponds to blastema establishment and CT dedifferentiation [14]. Molecular profiling of earlier stages (1-2 dpi) is technically unfeasible due to the absence of a visible blastema before 3 dpi and the abundance of erythrocytes. Understanding how limb CT cells respond within the first 48 hours post-injury remains an open question and could explain the early differential expression of *Lef1* and *Tcf7l2*.

Regarding the implication of Wnt signaling during epimorphic regeneration, its importance has previously been described in several species [102, 103, 262]. Our model could capture the activity of PCP and canonical Wnt signaling, known to orchestrate developmental limb outgrowth and regeneration, but has the limitation of only considering mesenchymal gene expression without spatial constrains. This might be less problematic for PCP signaling, as *Wnt5a* is the main trigger of this cascade and is expressed by CT fibroblasts [50]. In the case of canonical Wnt signaling, it is mostly initiated by *Wnt3a* expressed by the WE [262]. A more realistic model of Wnt signaling during axolotl limb regeneration should consider the gene expression duality between epidermal and mesenchymal cells, as well as their spatial organization, which physically constrains which cells can respond to extracellular signals.

Our *in silico* work highlighted *Lef1* and *Tcf7l2* as potential pro- and anti-regenerative regulators, respectively. To our knowledge, this constitutes the first evidence of their different and partially antagonistic role during axolotl limb regeneration. Our hypothesis is supported by previous work that unveiled their different function in mediating canonical Wnt signaling during *Xenopus laevis* limb development, where *Lef1* activates and *Tcf7l2* represses the expression of Wnt target genes [263]. We provided different pieces of evidence, obtained through different computational methods, that support this hypothesis. While the qPCR validation could only support the difference in expression for *Lef1*, we hypothesize that primers for *Tcf7l2* might have non-specifically hybridized with other gene products, due to the large similarity between the TCF/LEF family members or more likely failed to detect some transcript isoform, as these TFs give rise to a large variety of transcripts through alternative splicing [259]. Surprisingly, despite being differentially expressed at 3 dpi, the results of the *in-silico* perturbations show little effect on the early injury fibroblasts, affecting later populations to a larger extent, which suggests that Wnt engagement might start at later times post-injury. In this regard, *in vivo* KO and overexpression of these TFs using localized CRISPR-Cas9 and baculovirus injections will be necessary to gain a deeper understanding of their role during axolotl limb regeneration. These perturbations might need to be accompanied by localized Wnt injections in CSD to break the wound symmetry and mimic the Wnt gradient present in blastema. Additionally, the heterogeneous response of CSD fibroblasts to the perturbations suggests that this cluster contains different populations. Increasing the clustering granularity achieves a separation of CSD fibroblasts I between those responsive to *Tcf7l2* perturbations (cluster A, in blue) from those responding to *Lef1* perturbations (cluster B, in orange) (Fig. 5.14). Further molecular exploration combined with *in silico* and *in vivo* experiments should be performed to properly characterize these populations and understand the effects of *Lef1* and *Tcf7l2* during bone regeneration.

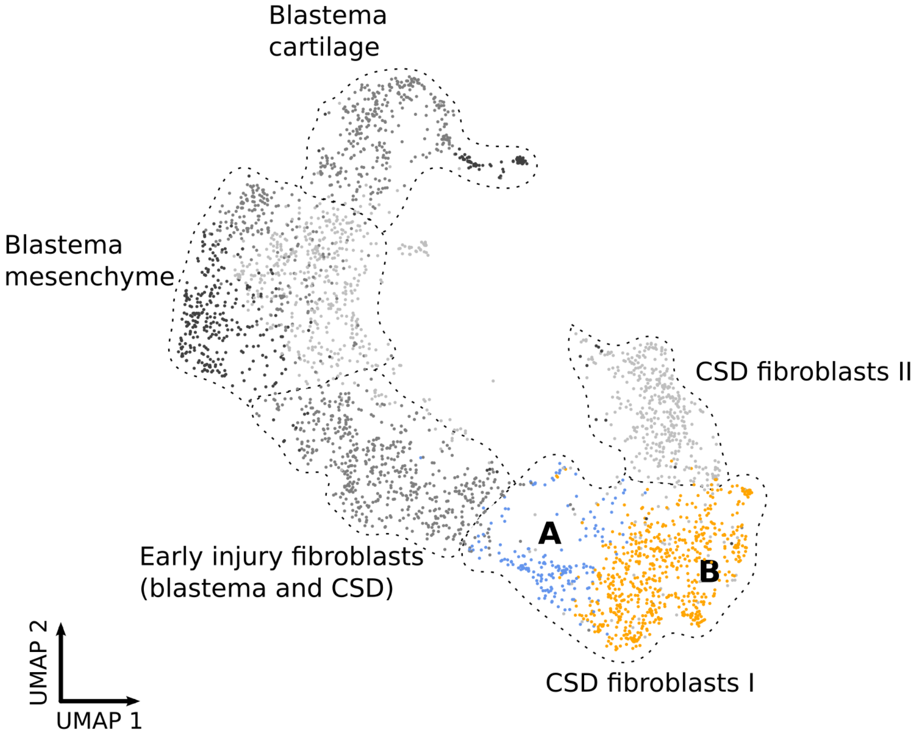


Figure 5.14: Heterogeneity within the CSD fibroblasts I might explain the differential response to TCF/LEF perturbations. UMAP visualization of integrated blastema and CSD CT cells. Dotted lines indicate the clusters described in Fig. 5.7a. Shades of grey indicate high-resolution clusters. The CSD fibroblasts I contain two subsclusters, responsive to *Tcf7l2* (A, blue) or *Lef1* (B, orange) perturbations.

6 | General discussion

This doctoral thesis aims to understand endogenous regeneration, hypothesizing that physiological regenerative mechanisms are controlled by latent GRNs that are at least partially conserved between lower vertebrates and mammals. We used a variety of computational approaches to model the GRNs controlling myocardial and bone regeneration using multi-omics data, at the bulk and single-cell level. We set the objectives to (1) model the CM GRN that allows the regeneration of the injured axolotl heart and identify the transcriptional hubs that orchestrate this process; (2) identify new therapeutic hubs to activate endogenous regenerative mechanisms in the adult mammalian myocardium, with a specific focus on (2.1) the regulators that induce mammalian CM polyploidization and (2.2) regulators that potentially regulate the pro-regenerative gene expression programs induced by *Erbb2*, *Yap1* and *Myc*; and (3) model the GRN in regenerative and fibrotic axolotl bone injuries and propose candidate molecular mechanisms to transform the fibrotic response into a regenerative response. In the presented findings, we identified *Hey2* as a roadblock for the injury response initiation in axolotl CM and as a potential driver of mammalian CM polyploidization. Additionally, we identified two pro-regenerative gene expression programs and estimated their contribution to the pro-regenerative footprints downstream of *Erbb2*, *Myc* and *Yap1*. Lastly, we predicted an insufficient GRN rewiring in axolotl CSD and identify partial antagonism between the TCF/LEF TFs *Lef1* and *Tcf7l2* that might partially explain why CSD fails to regenerate. Although these results must be considered in the context of biological and computational limitations, they represent exciting challenges for further studies.

From the perspective of computational biology, this work showcases the importance of letting the type and amount of data drive algorithmic choices, being the number of sampled data points arguably the most critical determinant. The popularity of scRNA-seq has switched the focus of algorithmic development towards accelerating computations in large datasets and incorporating temporal information, whether real or inferred. Additionally, the large-scale single-cell profiling of perturbation effects provided by Perturb-seq is an exciting avenue for causal GRN modeling [264]. However, this does not render bulk RNA-seq and its methods obsolete, as they can provide higher detail into the molecular mechanisms that regulate a process. Another important data-driven decision is whether to limit GRN inference to a network scaffold and, if so, which one to choose. When single-cell multi-omic data is available, the data-driven ATAC-seq scaffold is the best choice. When only scRNA-seq data is available, highly-curated literature databases of transcriptional regulation also provide a useuful network scaffold to aid the inference process. A clear example of this is shown in Chapter 3, in which we leverage prior information to infer upstream regulators using transcriptomic information from just 307 cells. However, the use of a literature-derived scaffold is only realistic when the organism is well studied. As shown in Chapter 5, information-agnostic and purely data-driven approaches could be useful when studying non-model organisms.

Regarding algorithms for GRN inference, extensive benchmark work shows that both linear and non-linear approaches can achieve similar accuracy, and highlight the importance of using a good network scaffold for accurate inference [131, 265]. Marbach and colleagues suggest to use a "wisdom of crowds" approximation in which the results provided by different algorithms are integrated. While this approach might render more accurate results, it implies longer run times, increased computational complexity and raises the problem of how should the different inference results be integrated [265]. Even when such an approach is followed, the resulting GRNs should be interpreted with caution. Smaller, higher-quality models might be able to represent molecular mechanisms to a certain degree, but they are overly simplistic cellular representations. Whole-transcriptome GRNs represent better the whole picture, but they should only be used to analyze broad network properties, as there is an inverse relationship between the network size and the detail with which individual edges are modeled. The accuracy limitations of current GRN inference algorithms need coordinated algorithmic and biotechnological developments. Most importantly, efforts into developing multi-omic profiling in single cells are imperative. Particularly, simultaneous transcriptome and proteome profiling in single-cells might significantly increase GRN accuracy, as they would allow to use TF protein abundance as a proxy of its activity. Lastly, computational developments to integrate the different molecular layers determining cell identity are required. Particularly, autoencoders and their variations provide a useful framework to solve this problem and are expected to provide exciting insights in the future [24, 266].

In spite of the relevance of the findings, this work has three main limitations. Firstly, all projects were centered on a single cell type. Supporting this decision is the fact that the lost tissue cannot be regenerated without proliferation of its main cell type. However, the heart and the limb are composed by a myriad of cell types that influence each other. Taking the heart as an example, Liu and colleages demonstrated that secretion of *Reln* by lymphatic endothelial cells affects CM proliferation and survival, and that local *Reln* administration using collagen pathches amelliorates heart function after MI [267]. In the case of the limb, signals from the epidermis and nerves are essential for CT dedifferentiation, and their absence completely abrogates regeneration [13]. Therefore, it is crucial to understand how the spatial milieu affects cell proliferation and whether it can be therapeutically targeted to promote functional recovery, and computational approaches that link ligand-receptor networks to GRNs must be developed to faithfully answer these questions. An additional challenge is understanding how stimuli arising from biomechanical forces influence cell dedifferentation and proliferation, specially in the context of the beating heart. Secondly, the results presented here only focus on transcriptional regulation and reduce cell phenotypes to their transcriptomic profile. Despite the transcriptome being arguably the main contributor to the cell identity, several other molecular layers (e.g., alternative splicing, post-translational modifications) are crucial in determining the cellular phenotype. Lastly, our results only provide computationally-generated hypotheses that must be thoroughly experimentally validated before a potential therapeutic translation. We must understand the molecular mechanisms controlled by the new regenerative candidates here identified and we must analyze the effect caused by their administration, both short and long-term.

7 | Conclusions

1. The axolotl ventricular CM GRN contains a modular structure made up by communities with different functionalities.
2. *Hey2* inactivation is critical for the initiation of axolotl ventricular CM injury response and its overexpression *in vivo* blunts the proliferative capacity of axolotl border zone CM at 7 dpi.
3. The adult WM MNDCM population has a more immature transcriptional footprint than their binucleated counterparts.
4. *Hey2* expression in adult mammalian ventricular CMs is anti-correlated with the abundance of mononucleated CMs *in vivo* and might repress the expression of MNDCM marker genes.
5. *caErbb2* and *Myc* overexpression mainly induce a shared transcriptional footprint related with cell proliferation, while *Yap5sa* overexpression mainly activates CM dedifferentiation.
6. The integration of previous knowledge with CM scRNA-seq data identifies a set of TFs that constitute the core of the mouse CM GRN. This network has two possible attractors that promote a CM maturation gene expression program. These attractors are robust and only a few perturbations can potentially transform them to induce cell proliferation.
7. The CSD CT GRN lacks a hierarchical topology characteristic of real transcriptional networks due to the loss of hub TFs present in the blastema network.
8. *Lef1* and *Tcf7l2* show injury-specific expression and behavior: *Lef1* is associated with pro-regenerative gene expression characteristic of blastema, while *Tcf7l2* fails to trigger this regenerative response in CSD CT fibroblasts.

8 | Conclusiones

1. La red de regulación génica de los cardiomiocitos ventriculares del ajolote presenta una estructura modular compuesta por comunidades con diferente funcionalidad.
2. La inactivación de *Hey2* es esencial para la iniciación de la respuesta a herida que montan los cardiomiocitos ventriculares de ajolote. La sobre-expresión de *Hey2 in vivo* obstaculiza la capacidad de proliferación de los cardimiocitos observada 7 días tras la criolesión.
3. Los cardiomiocitos mononucleados diploides del miocardio de trabajo adulto presentan una huella transcripcional más inmadura que la de los cardiomiocitos binucleados.
4. La expresión de *Hey2* en cardiomiocitos ventriculares de ratones adultos está negativamente correlacionada con la abundancia de cardiomiocitos mononucleados observada *in vivo*, y puede reprimir la expresión de genes marcadores de los cardiomiocitos mononucleados diploides.
5. La sobre-expresión de *caErbb2* y *Myc* comparten una huella transcripcional relacionada con proliferación celular, mientras que la sobre-expresión de *Yap5sa* principalmente activa la desdiferenciación del cardiomiocito.
6. La integración de conocimiento previo con datos de scRNA-seq identifican un grupo de factores de transcripción que constituyen el núcleo de la red de regulación génica de los cardiomiocitos murinos. Esta red presenta dos posibles atractores que regulan un programa de expresión génico relacionada con la maduración del cardiomiocito. Estos atractores son robustos y solo unas pocas perturbaciones son capaces de potencialmente transformarlos en atractores relacionados con proliferación celular.
7. La red de regulación génica del tejido conectivo de CSD carece de la topología jerárquica que caracteriza a las redes transcripcionales reales debido a la pérdida de factores de transcripción centrales presentes en la red de regulación génica de blastema.
8. *Lef1* y *Tcf7l2* muestran una expresión y función específicas de lesión: *Lef1* se asocia con expresión génica pro-regenerativa característica de blastema, mientras que *Tcf7l2* no es capaz de iniciar esta respuesta regenerativa en el tejido conectivo de CSD.

References

1. Maienschein J. Regenerative medicine's historical roots in regeneration, transplantation, and translation. Dev Biol. 2011; 358:278–84. DOI: `10.1016/j.ydbio.2010.06.014`
2. Poss KD, Wilson LG, and Keating MT. Heart regeneration in zebrafish. Science. 2002; 298:2188–90. DOI: `10.1126/science.1077857`
3. Ye L, D'Agostino G, Loo SJ, Wang CX, Su LP, Tan SH, et al. Early regenerative capacity in the porcine heart. Circulation. 2018; 138:2798–808. DOI: `10.1161/CIRCULATIONAHA.117.031542`
4. Porrello E, Mahmoud AI, Simpson E, Hill JA, Richardson JA, Olson EN, et al. Transient regenerative potential of the neonatal mouse heart. Science. 2011; 331:1078–80. DOI: `10.1126/science.1200708`
5. World Health Organization. Cardiovascular diseases (CVDs). `https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)`. Accessed: 2024-05-09. 2021
6. Nabel EG and Braunwald E. A tale of coronary artery disease and myocardial infarction. N Engl J Med. 2012; 366:54–63. DOI: `10.1056/NEJMra1112570`
7. Cahill TJ and Kharbanda RK. Heart failure after myocardial infarction in the era of primary percutaneous coronary intervention: Mechanisms, incidence and identification of patients at risk. World J Cardiol. 2017; 9:407–15. DOI: `10.4330/wjc.v9.i5.407`
8. Oberpriller JO and Oberpriller JC. Response of the adult newt ventricle to injury. J Exp Zool. 1974; 187:249–59. DOI: `10.1002/jez.1401870208`
9. Wu AM, Bisignano C, James SL, Abady GG, Abedi A, Abu-Gharbieh E, et al. Global, regional, and national burden of bone fractures in 204 countries and territories, 1990–2019: a systematic analysis from the Global Burden of Disease Study 2019. Lancet Healthy Longev. 2021; 2(9):e580–e592. DOI: `10.1016/S2666-7568(21)00172-0`
10. Kengelbach-Weigand A, Thielen C, Bäuerle T, Götzl R, Gerber T, Körner C, et al. Personalized medicine for reconstruction of critical-size bone defects – a translational approach with customizable vascularized bone tissue. NPJ Regen Med. 2021; 6:49. DOI: `10.1038/s41536-021-00158-8`
11. Chan WY, Lee KKH, and Tam PPL. Regenerative capacity of forelimb buds after amputation in mouse embryos at the early-organogenesis stage. J Exp Zool. 1991; 260(1):74–83. DOI: `10.1002/jez.1402600110`
12. Han M, Yang X, Lee J, Allan CH, and Muneoka K. Development and regeneration of the neonatal digit tip in mice. Dev Biol. 2008; 315(1):125–35. DOI: `10.1016/j.ydbio.2007.12.025`

13. Bassat E and Tanaka EM. The cellular and signaling dynamics of salamander limb regeneration. Curr Opin Cell Biol. 2021; 73:117–23. DOI: `10.1016/J.CEB.2021.07.010`
14. Gerber T, Murawala P, Knapp D, Masselink W, Schuez M, Hermann S, et al. Single-cell analysis uncovers convergence of cell identities during axolotl limb regeneration. Science. 2018; 362:eaaq0681. DOI: `10.1126/science.aaq0681`
15. Barbosa S, Niebel B, Wolf S, Mauch K, and Takors R. A guide to gene regulatory network inference for obtaining predictive solutions: Underlying assumptions and fundamental biological and data constraints. BioSystems. 2018; 174:37–48. DOI: `10.1016/j.biosystems.2018.10.008`
16. Akers K and Murali TM. Gene regulatory network inference in single-cell biology. Curr Opin Syst Biol. 2021; 26:87–97. DOI: `10.1016/j.coisb.2021.04.007`
17. van Dam S, Võsa U, van der Graaf A, Franke L, and de Magalhães JP. Gene co-expression analysis for functional classification and gene–disease predictions. Brief Bioinform. 2018; 19:575–92. DOI: `10.1093/BIB/BBW139`
18. Garg A, Di Cara A, Xenarios I, Mendoza L, and De Micheli G. Synchronous versus asynchronous modeling of gene regulatory networks. Bioinformatics. 2008; 24:1917–25. DOI: `10.1093/bioinformatics/btn336`
19. Huynh-Thu VA, Irrthum A, Wehenkel L, and Geurts P. Inferring Regulatory Networks from Expression Data Using Tree-Based Methods. PLoS One. 2010; (9):e12776. DOI: `10.1371/JOURNAL.PONE.0012776`
20. Aibar S, Bravo González-Blas C, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. SCENIC: single-cell regulatory network inference and clustering. Nat Methods. 2017; 14:1083–6. DOI: `10.1038/nmeth.4463`
21. Kamimoto K, Stringa B, Hoffmann CM, Jindal K, Solnica-Krezel L, and Morris SA. Dissecting cell identity via network inference and in silico gene perturbation. Nature. 2023; 614:742–51. DOI: `10.1038/s41586-022-05688-9`
22. Fleck JS, Jansen SMJ, Wollny D, Zenk F, Seimiya M, Jain A, et al. Inferring and perturbing cell fate regulomes in human brain organoids. Nature. 2022; 621:365–72. DOI: `10.1038/s41586-022-05279-8`
23. Emmert-Streib F, Dehmer M, and Haibe-Kains B. Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. Front Cell Dev Biol. 2014; 2:38. DOI: `10.3389/fcell.2014.00038`
24. Rivero-Garcia I, Torres M, and Sánchez-Cabo F. Deep generative models in single-cell omics. Comput Biol Med. 2024; 176:108561. DOI: `10.1016/j.compbiomed.2024.108561`
25. Mañanes D, Rivero-García I, Relaño C, Torres M, Sancho D, Jimenez-Carretero D, et al. SpatialDDLS: an R package to deconvolute spatial transcriptomics data using neural networks. Bioinformatics. 2024; 40(2):btae072. DOI: `10.1093/bioinformatics/btae072`
26. Smith DK and Zhang CL. Regeneration through reprogramming adult cell identity *in vivo*. Am J Pathol. 2015; 185:2619–28. DOI: `10.1016/j.ajpath.2015.02.025`

27. Clarke G, Harley P, Hubber EL, Manea T, Manuelli L, Read E, et al. Bench to bedside: Current advances in regenerative medicine. Curr Opin Cell Biol. 2018; 55:59–66. DOI: `10.1016/j.ceb.2018.05.006`
28. Jopling C, Sleep E, Raya M, Martí M, Raya A, and Izpisúa Belmonte JC. Zebrafish heart regeneration occurs by cardiomyocyte dedifferentiation and proliferation. Nature. 2010; 464:606–9. DOI: `10.1038/nature08899`
29. Santos DM, Rita AM, Casanellas I, Ova AB, Araújo IM, Power D, et al. Ear wound regeneration in the African spiny mouse *Acomys cahirinus*. Regeneration. 2016; 3:52–61. DOI: `10.1002/reg2.50`
30. Seifert AW, Kiama SG, Seifert MG, Goheen JR, Palmer TM, and Maden M. Skin shedding and tissue regeneration in African spiny mice (*Acomys*). Nature. 2012; 489:561–5. DOI: `10.1038/nature11499`
31. Wei Y, Wang YG, Jia Y, Li L, Yoon J, Zhang S, et al. Liver homeostasis is maintained by midlobular zone 2 hepatocytes. Science. 2021; 371:eabb1625. DOI: `10.1126/science.abb1625`
32. Bely AE and Nyberg KG. Evolution of animal regeneration: re-emergence of a field. Trends Ecol Evol. 2010; 25:161–70. DOI: `10.1016/j.tree.2009.08.005`
33. Sánchez Alvarado A and Tsonis PA. Bridging the regeneration gap: genetic insights from diverse animal models. Nat Rev Genet. 2006; 7:873–84. DOI: `10.1038/nrg1923`
34. Mehta AS and Singh A. Insights into regeneration tool box: An animal model approach. Dev Biol. 2019; 453:111–29. DOI: `10.1016/j.ydbio.2019.04.006`
35. David S and Aguayo AJ. Axonal elongation into peripheral nervous system "bridges" after central nervous system injury in adult rats. Science. 1981; (4523):931–3. DOI: `10.1126/science.6171034`
36. Carlson BM. Some principles of regeneration in mammalian systems. Anat Rec B New Anat. 2005; 287B:4–13. DOI: `10.1002/ar.b.20079`
37. Lipton BH and Schultz E. Developmental fate of skeletal muscle satellite cells. Science. 1979; 205:1292–4. DOI: `10.1126/science.472747`
38. Fu SY and Gordon T. The cellular and molecular basis of peripheral nerve regeneration. Mol Neurobiol. 1997; 14:67–116. DOI: `10.1007/BF02740621`
39. Moreno-Jiménez EP, Flor-García M, Terreros-Roncal J, Rábano A, Cafini F, Pallas-Bazarra N, et al. Adult hippocampal neurogenesis is abundant in neurologically healthy subjects and drops sharply in patients with Alzheimer's disease. Nat Med. 2019; 25(4):554–60. DOI: `10.1038/s41591-019-0375-9`
40. Zhang W, Ke CH, Guo HH, and Xiao L. Antler stem cells and their potential in wound healing and bone regeneration. World J Stem Cells. 2021; 13:1049–57. DOI: `10.4252/wjsc.v13.i8.1049`
41. Vogg MC, Galliot B, and Tsiairis CD. Model systems for regeneration: Hydra. Development. 2019; 146(21):dev177212. DOI: `10.1242/dev.177212`

42. Ivankovic M, Haneckova R, Thommen A, Grohme MA, Vila-Farré M, Werner S, et al. Model systems for regeneration: planarians. Development. 2019; 146(17):dev167684. DOI: `10.1242/dev.167684`
43. Nechiporuk A and Keating MT. A proliferation gradient between proximal and msxb - expressing distal blastema directs zebrafish fin regeneration. Development. 2002; 129:2607–17. DOI: `10.1242/dev.129.11.2607`
44. White JA, Boffa MB, Jones B, and Petkovich M. A zebrafish retinoic acid receptor expressed in the regenerating caudal fin. Development. 120:1861–72. DOI: `10.1242/dev.120.7.1861`
45. Vihtelic TS and Hyde DR. Light-induced rod and cone cell death and regeneration in the adult albino zebrafish (*Danio rerio*) retina. J Neurobiol. 2000; 44:289–307. DOI: `10.1002/1097-4695(20000905)44:3<289::AID-NEU1>3.0.CO;2-H`
46. Marques IJ, Lupi E, and Mercader N. Model systems for regeneration: zebrafish. Development. 2019; 146:dev167692. DOI: `10.1242/dev.167692`
47. Joven A, Elewa A, and Simon A. Model systems for regeneration: salamanders. Development. 2019; 146:dev167700. DOI: `10.1242/dev.167700`
48. Sandoval-Guzmán T. The axolotl. Nat Methods. 2023; 20(8):1117–9. DOI: `10.1038/s41592-023-01961-5`
49. Rodgers AK, Smith JJ, and Voss SR. Identification of immune and non-immune cells in regenerating axolotl limbs by single-cell sequencing. Exp Cell Res. 2020; 394:112149. DOI: `10.1016/j.yexcr.2020.112149`
50. Ghosh S, Roy S, Séguin C, Bryant SV, and Gardiner DM. Analysis of the expression and function of Wnt-5a and Wnt-5b in developing and regenerating axolotl (*Ambystoma mexicanum*) limbs. Dev Growth Differ. 2008; 50:289–97. DOI: `10.1111/j.1440-169X.2008.01000.x`
51. Godwin JW, Pinto AR, and Rosenthal NA. Macrophages are required for adult salamander limb regeneration. Proc Natl Acad Sci U S A. 2013; 110:9415–20. DOI: `10.1073/pnas.1300290110`
52. Tsai SL, Baselga-Garriga C, and Melton DA. Blastemal progenitors modulate immune signaling during early limb regeneration. Development. 2019; 146:dev169128. DOI: `10.1242/dev.169128`
53. Kawaguchi A, Wang J, Knapp D, Murawala P, Nowoshilow S, Masselink W, et al. A chromatin code for limb segment identity in axolotl limb regeneration. Dev Cell. 2024. DOI: `10.1016/j.devcel.2024.05.002`
54. Godwin JW, Debuque R, Salimova E, and Rosenthal NA. Heart regeneration in the salamander relies on macrophage-mediated control of fibroblast activation and the extracellular landscape. NPJ Regen Med. 2017; 2:22. DOI: `10.1038/s41536-017-0027-y`
55. Yates CC, Hebda P, and Wells A. Skin wound healing and scarring: fetal wounds and regenerative restitution. Birth Defects Res C Embryo Today. 2012; 96(4):325–33. DOI: `10.1002/bdrc.21024`
56. Ito M, Yang Z, Andl T, Cui C, Kim N, Millar SE, et al. Wnt-dependent de novo hair follicle regeneration in adult mouse skin after wounding. Nature. 2007; 447:316–20. DOI: `10.1038/nature05766`

57. Montero AM and Huang AH. The regenerative capacity of neonatal tissues. Development. 2022; 149(12):dev199819. DOI: `10.1242/dev.199819`
58. Garry GA, Antony ML, and Garry DJ. Cardiotoxin induced injury and skeletal muscle regeneration. Skeletal Muscle Regeneration in the Mouse. Springer New York, NY, 2016 :61–71
59. Fernando WA, Leininger E, Simkin J, Li N, Malcom CA, Sathyamoorthi S, et al. Wound healing and blastema formation in regenerating digit tips of adult mice. Dev Biol. 2011; 350:301–10. DOI: `10.1016/j.ydbio.2010.11.035`
60. Fausto N, Campbell JS, and Riehle KJ. Liver regeneration. J Hepatol. 2012; 57:692–4. DOI: `10.1016/j.jhep.2012.04.016`
61. Okamura DM, Nguyen ED, Collins SJ, Yoon K, Gere JB, Weiser-Evans MCM, et al. Mammalian organ regeneration in spiny mice. J Muscle Res Cell Motil. 2023; 44:39–52. DOI: `10.1007/s10974-022-09631-3`
62. Okamura DM, Brewer CM, Wakenight P, Bahrami N, Bernardi K, Tran A, et al. Spiny mice activate unique transcriptional programs after severe kidney injury regenerating organ function without fibrosis. iScience. 2021; 24:103269. DOI: `10.1016/j.isci.2021.103269`
63. Streeter KA, Sunshine MD, Brant JO, Sandoval AGW, Maden M, and Fuller DD. Molecular and histologic outcomes following spinal cord injury in spiny mice, Acomys cahirinus. J Comp Neurol. 2020; 528:1535–47. DOI: `10.1002/cne.24836`
64. Garbern JC and Lee RT. Heart regeneration: 20 years of progress and renewed optimism. Dev Cell. 2022; 57:424–39. DOI: `10.1016/j.devcel.2022.01.012`
65. Vujic A, Natarajan N, and Lee RT. Molecular mechanisms of heart regeneration. Semin Cell Dev Biol. 2020; 100:20–8. DOI: `10.1016/j.semcdb.2019.09.003`
66. Laflamme MA and Murry CE. Heart regeneration. Nature. 2011; 473:326–35. DOI: `10.1038/nature10147`
67. Price EL, Vieira JM, and Riley PR. Model organisms at the heart of regeneration. Dis Model Mech. 2019; 12:dmm040691. DOI: `10.1242/dmm.040691`
68. Broughton KM and Sussman MA. Myocardial Regeneration for Humans — Modifying Biology and Manipulating Evolution. Circ J. 2017; 81:142–8. DOI: `10.1253/circj.CJ-16-1228`
69. Tzahor E and Poss K. Cardiac regeneration strategies: Staying young at heart. Science. 2017; 356:1035–9. DOI: `10.1126/science.aam5894`
70. González-Rosa JM, Sharpe M, Field D, Soonpaa MH, Field LJ, Burns CE, et al. Myocardial polyploidization creates a barrier to heart regeneration in zebrafish. Dev Cell. 2018; 44:433–446.e7. DOI: `10.1016/j.devcel.2018.01.021`
71. Yin HM, Burns CG, and Burns CE. Innate mechanisms of heart regeneration. Cold Spring Harb Perspect Biol. 2021; 13:a040766. DOI: `10.1101/cshperspect.a040766`
72. Uygur A and Lee RT. Mechanisms of cardiac regeneration. Dev Cell. 2016; 36:362–74. DOI: `10.1016/j.devcel.2016.01.018`

73. Senyo SE, Steinhauser ML, Pizzimenti CL, Yang VK, Cai L, Wang M, et al. Mammalian heart renewal by pre-existing cardiomyocytes. Nature. 2013; 493:433–6. DOI: `10.1038/nature11682`

74. Bergmann O, Zdunek S, Felker A, Salehpour M, Alkass K, Bernard S, et al. Dynamics of cell generation and turnover in the human heart. Cell. 2015; 161:1566–75. DOI: `10.1016/j.cell.2015.05.026`

75. Haubner BJ, Schneider J, Schweigmann U, Schuetz T, Dichtl W, Velik-Salchner C, et al. Functional recovery of a human neonatal heart after severe myocardial infarction. Circ Res. 2016; 118:216–21. DOI: `10.1161/CIRCRESAHA.115.307017`

76. Saker D, Walsh-Sukys M, Spector M, and Zahka K. Cardiac recovery and survival after neonatal myocardial infarction. Pediatric Cardiology 1997; 18:139–42. DOI: `10.1007/s002469900133`

77. Zhao B, Tumaneng K, and Guan KL. The Hippo pathway in organ size control, tissue regeneration and stem cell self-renewal. Nat Cell Biol. 2011; 13:877–83. DOI: `10.1038/ncb2303`

78. Heallen T, Morikawa Y, Leach J, Tao G, Willerson JT, Johnson RL, et al. Hippo signaling impedes adult heart regeneration. Development. 2013; 140:4683–90. DOI: `10.1242/dev.102798`

79. Monroe TO, Hill MC, Morikawa Y, Leach JP, Heallen T, Cao S, et al. YAP Partially Reprograms Chromatin Accessibility to Directly Induce Adult Cardiogenesis In Vivo. Dev Cell. 2019; 48:765–779.e7. DOI: `10.1016/j.devcel.2019.01.017`

80. Hertig CM, Kubalak SW, Wang Y, and Chien KR. Synergistic roles of neuregulin-1 and insulin-like growth factor-I in activation of the phosphatidylinositol 3-kinase pathway and cardiac chamber morphogenesis. J Biol Chem. 1999; 274:37362–9. DOI: `10.1074/jbc.274.52.37362`

81. Reischauer S, Arnaout R, Ramadass R, and Stainier DYR. Actin binding GFP allows 4D *in vivo* imaging of myofilament dynamics in the zebrafish heart and the identification of Erbb2 signaling as a remodeling factor of myofibril architecture. Circ Res. 2014; 115:845–56. DOI: `10.1161/CIRCRESAHA.115.304356`

82. D'Amato V, Raimondo L, Formisano L, Giuliano M, De Placido S, Rosa R, et al. Mechanisms of lapatinib resistance in HER2-driven breast cancer. Cancer Treat Rev. 2015; 10:877–83. DOI: `10.1016/j.ctrv.2015.08.001`

83. D'Uva G, Aharonov A, Lauriola M, Kain D, Yahalom-Ronen Y, Carvalho S, et al. ERBB2 triggers mammalian heart regeneration by promoting cardiomyocyte dedifferentiation and proliferation. Nat Cell Biol. 2015; 17:627–38. DOI: `10.1038/ncb3149`

84. Dang CV. c-Myc Target Genes Involved in Cell Growth, Apoptosis, and Metabolism. Mol Cell Biol. 1999; 19:1–11. DOI: `10.1128/MCB.19.1.1`

85. Muñoz-Martín N, Sierra R, Schimmang T, Villa del Campo C, and Torres M. Myc is dispensable for cardiomyocyte development but rescues Mycn-deficient hearts through functional replacement and cell competition. Development. 2019; 146:dev170753. DOI: `10.1242/dev.170753`

86. Bywater MJ, Burkhart DL, Straube J, Sabò A, Pendino V, Hudson JE, et al. Reactivation of Myc transcription in the mouse heart unlocks its proliferative capacity. Nat Commun. 2020; 11:1827. DOI: `10.1038/s41467-020-15552-x`

87. Boikova A, Quaife-Ryan GA, Batho CAP, Lawrence E, Robinson H, Ascanelli C, et al. A transient modified mRNA encoding Myc and Cyclin T1 induces cardiac regeneration and improves cardiac function after myocardial injury. bioRxiv. 2023. DOI: `10.1101/2023.08.02.551469`
88. Amado LC, Saliaris AP, Schuleri KH, John MS, Xie JS, Cattaneo S, et al. Cardiac repair with intramyocardial injection of allogeneic mesenchymal stem cells after myocardial infarction. Proc Natl Acad Sci U S A. 2005; 102:11474–9. DOI: `10.1073/pnas.0504388102`
89. Beltrami AP, Barlucchi L, Torella D, Baker M, Limana F, Chimenti S, et al. Adult cardiac stem cells are multipotent and support myocardial regeneration. Cell. 2003; 114:763–76. DOI: `10.1016/S0092-8674(03)00687-1`
90. Hare JM, Fishman JE, Gerstenblith G, DiFede Velazquez DL, Zambrano JP, Suncion VY, et al. Comparison of allogeneic vs autologous bone marrow–derived mesenchymal stem cells delivered by transendocardial injection in patients with ischemic cardiomyopathy. JAMA. 2012; 308:2369. DOI: `10.1001/jama.2012.25321`
91. Sultana N, Zhang L, Yan J, Chen J, Cai W, Razzaque S, et al. Resident c-kit+ cells in the heart are not cardiac stem cells. Nat Commun. 2015; 6:8701. DOI: `10.1038/ncomms9701`
92. van Berlo JH, Kanisicak O, Maillet M, Vagnozzi RJ, Karch J, Lin SCJ, et al. c-kit+ cells minimally contribute cardiomyocytes to the heart. Nature. 2014; 509(7500):337–41. DOI: `10.1038/nature13309`
93. Vagnozzi RJ, Maillet M, Sargent MA, Khalil H, Johansen AKZ, Schwanekamp JA, et al. An acute immune response underlies the benefit of cardiac stem cell therapy. Nature. 2020; 577:405–9. DOI: `10.1038/s41586-019-1802-2`
94. Stüdemann T, Rössinger J, Manthey C, Geertz B, Srikantharajah R, von Bibra C, et al. Contractile force of transplanted cardiomyocytes actively supports heart function after injury. Circulation. 2022; 146:1159–69. DOI: `10.1161/CIRCULATIONAHA.122.060124`
95. Weinberger F, Breckwoldt K, Pecha S, Kelly A, Geertz B, Starbatty J, et al. Cardiac repair in guinea pigs with human engineered heart tissue from induced pluripotent stem cells. Sci Transl Med. 2016; 8:363ra148. DOI: `10.1126/scitranslmed.aaf8781`
96. McCusker C, Bryant SV, and Gardiner DM. The axolotl limb blastema: cellular and molecular mechanisms driving blastema formation and limb regeneration in tetrapods. Regeneration. 2015; 2:54–71. DOI: `10.1002/reg2.32`
97. Haas BJ and Whited JL. Advances in Decoding Axolotl Limb Regeneration. Trends Genet. 2017; 33:553–65. DOI: `10.1016/j.tig.2017.05.006`
98. Grainger S and Willert K. Mechanisms of Wnt signaling and control. Wiley Interdiscip Rev Syst Biol Med. 2018; 10:e1422. DOI: `10.1002/wsbm.1422`
99. Komiya Y and Habas R. Wnt signal transduction pathways. Organogenesis. 2008; 4:68–75. DOI: `10.4161/org.4.2.5851`
100. Zhan T, Rindtorff N, and Boutros M. Wnt signaling in cancer. Oncogene. 2017; 36:1461–73. DOI: `10.1038/onc.2016.304`

101. Habas R and Dawid IB. Dishevelled and Wnt signaling: Is the nucleus the final frontier? J Biol. 2005; 4:2. DOI: `10.1186/jbiol22`

102. Kawakami Y, Esteban CR, Raya M, Kawakami H, Martí M, Dubova I, et al. Wnt/beta-catenin signaling regulates vertebrate limb regeneration. Genes Dev. 2006; 20:3232–7. DOI: `10.1101/gad.1475106`

103. Poss KD, Shen J, and Keating MT. Induction of lef1 during zebrafish fin regeneration. Dev Dynam. 2000; 219:282–6. DOI: `10.1002/1097-0177(2000)9999:9999<::AID-DVDY1045>3.0.CO;2-C`

104. Lovely AM, Duerr TJ, Qiu Q, Galvan S, Voss R, and Monaghan JR. Wnt Signaling Coordinates the Expression of Limb Patterning Genes During Axolotl Forelimb Development and Regeneration. Front Cell Dev Biol. 2022; 10:814250. DOI: `10.3389/fcell.2022.814250`

105. Mercatelli D, Scalambra L, Triboli L, Ray F, and Giorgi FM. Gene regulatory network inference resources: A practical overview. Biochim Biophys Acta. 2019; 1863:194430. DOI: `10.1016/j.bbagrm.2019.194430`

106. Badia-i-Mompel P, Wessels L, Müller-Dott S, Trimbour R, Ramirez Flores RO, Argelaguet R, et al. Gene regulatory network inference in the era of single-cell multi-omics. Nat Rev Genet. 2023; 24:739–54. DOI: `10.1038/s41576-023-00618-5`

107. Bansal M, Della Gatta G, and di Bernardo D. Inference of gene regulatory networks and compound mode of action from time course gene expression profiles. Bioinformatics. 2006; 22:815–22. DOI: `10.1093/bioinformatics/btl003`

108. Mochida K, Koda S, Inoue K, and Nishii R. Statistical and machine learning approaches to predict gene regulatory networks from transcriptome datasets. Front Plant Sci. 2018; 9:1770. DOI: `10.3389/fpls.2018.01770`

109. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, and Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res. 2019; 29:1363–75. DOI: `10.1101/GR.240663.118`

110. Dugourd A and Saez-Rodriguez J. Footprint-based functional analysis of multiomic data. Curr Opin Syst Biol. 2019; 15:82–90. DOI: `10.1016/J.COISB.2019.04.002`

111. Krämer A, Green J, Pollard J, and Tugendreich S. Causal analysis approaches in ingenuity pathway analysis. Bioinformatics. 2014; 30(4):523–30. DOI: `10.1093/bioinformatics/btt703`

112. Müller-Dott S, Tsirvouli E, Vazquez M, Ramirez Flores RO, Badia-i-Mompel P, Fallegger R, et al. Expanding the coverage of regulons from high-confidence prior knowledge for accurate estimation of transcription factor activities. Nucleic Acids Res. 2023; 51:10934–49. DOI: `10.1093/nar/gkad841`

113. Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, Lemma RB, Turchi L, Blanc-Mathieu R, et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 2022; 50:D165–D173. DOI: `10.1093/nar/gkab1113`

114. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, et al. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res. 2006; 34:D108–D110. DOI: `10.1093/nar/gkj143`

115. Sandelin A, Alkema W, Engström P, Wasserman WW, and Lenhard B. JASPAR: An open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res. 2004; 32:D91–D94. DOI: `10.1093/nar/gkh012`

116. Janky R, Verfaillie A, Imrichová H, Van de Sande B, Standaert L, Christiaens V, et al. iRegulon: from a gene list to a gene regulatory network using large motif and track collections. PLoS Comput Biol. 2014; 10:e1003731. DOI: `10.1371/journal.pcbi.1003731`

117. Herrmann C, Van de Sande B, Potier D, and Aerts S. i-cisTarget: an integrative genomics method for the prediction of regulatory features and cis-regulatory modules. Nucleic Acids Res. 2012; 40:e114. DOI: `10.1093/nar/gks543`

118. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010; 38:576–89. DOI: `10.1016/j.molcel.2010.05.004`

119. Fiers MWEJ, Minnoye L, Aibar S, Bravo González-Blas C, Kalender Atak Z, and Aerts S. Mapping gene regulatory networks from single-cell omics data. Brief Funct Genomics. 2018; 17:246–54. DOI: `10.1093/BFGP/ELX046`

120. Karlebach G and Shamir R. Modelling and analysis of gene regulatory networks. Nat Rev Mol Cell Biol. 2008; 9:770–80. DOI: `10.1038/nrm2503`

121. Schwab JD, Kühlwein SD, Ikonomi N, Kühl M, and Kestler HA. Concepts in Boolean network modeling: What do they all mean? Comput Struct Biotechnol J. 2020; 18:571–82. DOI: `10.1016/j.csbj.2020.03.001`

122. Lim CY, Wang H, Woodhouse S, Piterman N, Wernisch L, Fisher J, et al. BTR: Training asynchronous Boolean models using single-cell expression data. BMC Bioinformatics. 2016; 17:355. DOI: `10.1186/s12859-016-1235-y`

123. Woodhouse S, Piterman N, Wintersteiger CM, Göttgens B, and Fisher J. SCNS: A graphical tool for reconstructing executable regulatory networks from single-cell genomic data. BMC Syst Biol. 2018; 12:59. DOI: `10.1186/s12918-018-0581-y`

124. Malekpour SA, Alizad-Rahvar AR, and Sadeghi M. LogicNet: Probabilistic continuous logics in reconstructing gene regulatory networks. BMC Bioinformatics. 2020; 21:318. DOI: `10.1186/s12859-020-03651-x`

125. Hamey FK, Nestorowa S, Kinston SJ, Kent DG, Wilson NK, and Gottgens B. Reconstructing blood stem cell regulatory network models from single-cell molecular profiles. Proc Natl Acad Sci U S A. 2017; 114:5822–9. DOI: `10.1073/PNAS.1610609114`

126. Bogner-Strauss JG, Prokesch A, Sanchez-Cabo F, Rieder D, Hackl H, Duszka K, et al. Reconstruction of gene association network reveals a transmembrane protein required for adipogenesis and targeted by PPARγ. Cell Mol Life Sci. 2010; 67:4049–64. DOI: `10.1007/s00018-010-0424-5`

127. Cha J and Lee I. Single-cell network biology for resolving cellular heterogeneity in human diseases. Exp Mol Med. 2020; 52:1798–808. DOI: 10.1038/s12276-020-00528-0

128. Kim S. ppcor: an R package for a fast calculation to semi-partial correlation coefficients. Commun Stat Appl Methods. 2015; 22:665–74. DOI: 10.5351/csam.2015.22.6.665

129. Specht AT and Li J. LEAP: constructing gene co-expression networks for single-cell RNA-sequencing data using pseudotime ordering. Bioinformatics. 2017; 33(5):764–6. DOI: 10.1093/BIOINFORMATICS/BTW729

130. Angerer P, Simon L, Tritschler S, Wolf FA, Fischer D, and Theis FJ. Single cells make big data: New challenges and opportunities in transcriptomics. Curr Opin Syst Biol. 2017; 4:85–91. DOI: 10.1016/j.coisb.2017.07.004

131. Pratapa A, Jalihal AP, Law JN, Bharadwaj A, and Murali TM. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. Nat Methods. 2020; 17:147–54. DOI: 10.1038/s41592-019-0690-6

132. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Favera RD, et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics. 2006; 7:S7. DOI: 10.1186/1471-2105-7-S1-S7

133. Kim J, Jakobsen ST, Natarajan KN, and Won KJ. TENET: gene network reconstruction using transfer entropy reveals key regulatory factors from single cell transcriptomic data. Nucleic Acids Res. 2021; 49:e1. DOI: 10.1093/nar/gkaa1014

134. Qiu X, Rahimzamani A, Wang L, Ren B, Mao Q, Durham T, et al. Inferring causal gene regulatory networks from coupled single-cell expression dynamics using Scribe. Cell Syst. 2020; 10:265–274.e11. DOI: 10.1016/j.cels.2020.02.003

135. Moerman T, Aibar Santos S, Bravo González-Blas C, Simm J, Moreau Y, Aerts J, et al. GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. Bioinformatics. 2019; 35:2159–61. DOI: 10.1093/BIOINFORMATICS/BTY916

136. Sanchez-Castillo M, Blanco D, Tienda-Luna IM, Carrion MC, and Huang Y. A Bayesian framework for the inference of gene regulatory networks from time and pseudo-time series data. Bioinformatics. 2018; 34:964–70. DOI: 10.1093/bioinformatics/btx605

137. Sekula M, Gaskins J, and Datta S. A sparse Bayesian factor model for the construction of gene co-expression networks from single-cell RNA sequencing count data. BMC Bioinformatics. 2020; 21:361. DOI: 10.1186/s12859-020-03707-y

138. Martí-Gómez C, Lara-Pezzi E, and Sánchez-Cabo F. dSreg: a Bayesian model to integrate changes in splicing and RNA-binding protein activity. Bioinformatics. 2020; 36:2134–41. DOI: 10.1093/bioinformatics/btz915

139. Matsumoto H, Kiryu H, Furusawa C, Ko MSH, Ko SBH, Gouda N, et al. SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. Bioinformatics. 2017; 33:2314–21. DOI: 10.1093/bioinformatics/btx194

140. Aubin-Frankowski PC and Vert JP. Gene regulation inference from single-cell RNA-seq data with linear differential equations and velocity inference. Bioinformatics. 2020; 36:4774–80. DOI: `10.1093/bioinformatics/btaa576`
141. Huynh-Thu VA and Geurts P. dynGENIE3: dynamical GENIE3 for the inference of gene networks from time series expression data OPEN. Sci Rep. 2018; 8:3384. DOI: `10.1038/s41598-018-21715-0`
142. Bansal M, Belcastro V, Ambesi-Impiombato A, and di Bernardo D. How to infer gene networks from expression profiles. Mol Syst Biol. 2007; 3:78. DOI: `10.1038/msb4100120`
143. van der Sande M, Frölich S, and van Heeringen SJ. Computational approaches to understand transcription regulation in development. Biochem Soc Trans. 2023; 51:1–12. DOI: `10.1042/BST20210145`
144. Yuan Y and Bar-Joseph Z. Deep learning for inferring gene relationships from single-cell expression data. Proc Natl Acad Sci U S A. 2019; 116:27151–8. DOI: `10.1073/pnas.1911536116`
145. Chen J, Cheong CW, Lan L, Zhou X, Liu J, Lyu A, et al. DeepDRIM: a deep neural network to reconstruct cell-type-specific gene regulatory network using single-cell RNA-seq data. Brief Bioinform. 2021; 22:bbab325. DOI: `10.1093/BIB/BBAB325`
146. Xu Y, Chen J, Lyu A, Cheung WK, and Zhang L. dynDeepDRIM: a dynamic deep learning model to infer direct regulatory interactions using time-course single-cell gene expression data. Brief Bioinform. 2022; 23:bbac424. DOI: `10.1093/bib/bbac424`
147. Shu H, Zhou J, Lian Q, Li H, Zhao D, Zeng J, et al. Modeling gene regulatory networks using neural network architectures. Nat Comput Sci. 2021; 1:491–501. DOI: `10.1038/s43588-021-00099-8`
148. Zhu H and Slonim DK. Improving gene regulatory network inference using dropout augmentation. bioRxiv. 2023. DOI: `10.1101/2023.01.26.525733`
149. Cui H, Wang C, Maan H, Pang K, Luo F, and Wang B. scGPT: Towards building a foundation model for single-cell multi-omics using generative AI. Nat Methods. DOI: `10.1038/s41592-024-02201-0`
150. Ravasz E and Barabási AL. Hierarchical organization in complex networks. Phys Rev E. 2003; 67:026112. DOI: `10.1103/PhysRevE.67.026112`
151. Marku M and Pancaldi V. From time-series transcriptomics to gene regulatory networks: A review on inference methods. PLoS Comput Biol. 2023; 19:e1011254. DOI: `10.1371/journal.pcbi.1011254`
152. Morris JH, Apeltsin L, Newman AM, Baumbach J, Wittkop T, Su G, et al. ClusterMaker: A multi-algorithm clustering plugin for Cytoscape. BMC Bioinformatics. 2011; 12:436. DOI: `10.1186/1471-2105-12-436`
153. Rosvall M, Axelsson D, and Bergstrom CT. The map equation. Eur Phys J Spec Top. 2009; 178:13–23. DOI: `10.1140/epjst/e2010-01179-1`

154. Wilder CL, Lefaudeux D, Mathenge R, Kishimoto K, Munoz AZ, Nguyen MA, et al. A stimulus-contingent positive feedback loop enables IFN-beta dose-dependent activation of pro-inflammatory genes. Mol Syst Biol. 2023; 19:e11294. DOI: `10.15252/msb.202211294`

155. Palsson BØ. Basics of systems biology. Systems Biology: Properties of reconstructed networks. Cambridge University Press, 2006 :12–28

156. Wang Z, Cui M, Shah AM, Tan W, Liu N, Bassel-Duby R, et al. Cell-type-specific gene regulatory networks underlying murine neonatal heart regeneration at single-cell resolution. Cell Rep. 2020; 33:108472. DOI: `10.1016/j.celrep.2020.108472`

157. Nomura S, Satoh M, Fujita T, Higo T, Sumida T, Ko T, et al. Cardiomyocyte gene programs encoding morphological and functional signatures in cardiac hypertrophy and failure. Nat Commun. 2018; 9:4435. DOI: `10.1038/s41467-018-06639-7`

158. Kuppe C, Ramirez Flores RO, Li Z, Hayat S, Levinson RT, Liao X, et al. Spatial multi-omic map of human myocardial infarction. Nature. 2022; 608:766–77. DOI: `10.1038/s41586-022-05060-x`

159. Reconstruction of regulatory network predicts transcription factors driving the dynamics of zebrafish heart regeneration. Gene. 2022; 819:146242. DOI: `10.1016/j.gene.2022.146242`

160. Chen H and VanBuren V. A provisional gene regulatory atlas for mouse heart development. PLoS One. 2014; 9:e83364. DOI: `10.1371/journal.pone.0083364`

161. Uosaki H, Cahan P, Lee DI, Wang S, Miyamoto M, Fernandez L, et al. Transcriptional landscape of cardiomyocyte maturation. Cell Rep. 2015; 13:1705–16. DOI: `10.1016/j.celrep.2015.10.032`

162. Harris BN, Woo LA, Perry RN, Civelek M, Wolf MJ, and Saucerman JJ. Dynamic map illuminates Hippo to cMyc module crosstalk driving cardiomyocyte proliferation. bioRxiv. 2022. DOI: `10.1101/2022.10.11.511763`

163. Lin TY, Gerber T, Taniguchi-Sugiura Y, Murawala P, Hermann S, Grosser L, et al. Fibroblast dedifferentiation as a determinant of successful regeneration. Dev Cell. 2021; 56:1541–1551.e6. DOI: `10.1016/j.devcel.2021.04.016`

164. Ye F, Zhang G, Weigao E, Chen H, Yu C, Yang L, et al. Construction of the axolotl cell landscape using combinatorial hybridization sequencing at single-cell resolution. Nat Commun. 2022; 13:4228. DOI: `10.1038/s41467-022-31879-z`

165. Khattak S, Murawala P, Andreas H, Kappert V, Schuez M, Sandoval-Guzman T, et al. Optimized axolotl (*Ambystoma mexicanum*) husbandry, breeding, metamorphosis, transgenesis and tamoxifen-mediated recombination. Nat Protoc. 2014; 9:529–40. DOI: `0.1038/nprot.2014.040`

166. Lust K, Maynard A, Gomes T, Fleck JS, Camp JG, Tanaka EM, et al. Single-cell analyses of axolotl telencephalon organization, neurogenesis, and regeneration. Science. 2022; 377:eabp9262. DOI: `10.1126/science.abp9262`

167. Bray NL, Pimentel H, Melsted P, and Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016; 34:525–7. DOI: `10.1038/nbt.3519`

168. Melsted P, Booeshaghi AS, Liu L, Gao F, Lu L, Min KH, et al. Modular, efficient and constant-memory single-cell RNA-seq preprocessing. Nat Biotechnol. 2021; 39:813–8. DOI: `10.1038/s41587-021-00870-2`

169. Hao Y, Hao S, Andersen-Nissen E, III WMM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021; 184:P3573–3587.E29. DOI: `10.1016/j.cell.2021.04.048`

170. Stuart T, Srivastava A, Madad S, Lareau CA, and Satija R. Single-cell chromatin state analysis with Signac. Nat. Methods 2021; 18:1333–41. DOI: `10.1038/s41592-021-01282-5`

171. Hu C, Li T, Xu Y, Zhang X, Li F, Bai J, et al. CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. Nucleic Acids Res. 2023; 51:D870–D876. DOI: `10.1093/nar/gkac947`

172. Franzén O, Gan LM, and Björkegren JLM. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. Database. 2019; 2019:baz046. DOI: `10.1093/database/baz046`

173. Cable DM, Murray E, Zou LS, Goeva A, Macosko EZ, Chen F, et al. Robust decomposition of cell type mixtures in spatial transcriptomics. Nat Biotechnol. 2022; 40:517–26. DOI: `10.1038/s41587-021-00830-w`

174. Wolf FA, Angerer P, and Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018; 19:15. DOI: `10.1186/s13059-017-1382-0`

175. Ma S, Zhang B, LaFave LM, Earl AS, Chiang Z, Hu Y, et al. Chromatin potential identified by shared single-cell profiling of RNA and chromatin. Cell. 2020; 183:1103–1116.e20. DOI: `10.1016/j.cell.2020.09.056`

176. Schloissnig S, Kawaguchi A, Nowoshilow S, Falcon F, Otsuki L, Tardivo P, et al. The giant axolotl genome uncovers the evolution, scaling, and transcriptional control of complex gene loci. Proc Natl Acad Sci U S A. 2021; 118:e2017176118. DOI: `10.1073/pnas.2017176118`

177. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012; 485:376–80. DOI: `10.1038/nature11082`

178. Beagan JA and Phillips-Cremins JE. On the existence and functionality of topologically associating domains. Nat Genet. 2020; 52:8–16. DOI: `10.1038/s41588-019-0561-1`

179. Nowoshilow S, Schloissnig S, Fei JF, Dahl A, Pang AWC, Pippel M, et al. The axolotl genome and the evolution of key tissue formation regulators. Nature. 2018; 554:50–5. DOI: `10.1038/nature25458`

180. Otasek D, Morris JH, Bouças J, Pico AR, and Demchak B. Cytoscape Automation: Empowering workflow-based network analysis. Genome Biol. 2019; 20:185. DOI: `10.1186/s13059-019-1758-4`

181. Assenov Y, Ramírez F, Schelhorn SE, Lengauer T, and Albrecht M. Computing topological parameters of biological networks. Bioinformatics. 2008; 24:282–4. DOI: `10.1093/bioinformatics/btm554`

182. Barabási AL. Communities. Network Science. Cambridge University Press, 2016 :321–76
183. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nat Genet. 2000; 25:25–9. DOI: `10.1038/75556`
184. The Gene Ontology Consortium, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, et al. The Gene Ontology knowledgebase in 2023. Genetics. 2023; 224:iyad031. DOI: `10.1093/genetics/iyad031`
185. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, and Sergushichev A. Fast gene set enrichment analysis. bioRxiv. 2021. DOI: `10.1101/060012`
186. Zhang L, Tian X, Li Y, He L, Zhang H, Huang X, et al. Identification of a hybrid myocardial zone in the mammalian heart after birth. Nat Commun. 2017; 8:87. DOI: `10.1038/s41467-017-00118-1`
187. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011; 17:10–12. DOI: `10.14806/ej.17.1.200`
188. Li B and Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011; 12:323. DOI: `10.1186/1471-2105-12-323`
189. McCarthy DJ, Campbell KR, Lun ATL, and Wills QF. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics. 2017; 33:1179–86. DOI: `10.1093/bioinformatics/btw777`
190. Butler A, Hoffman P, Smibert P, Papalexi E, and Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018; 36:411–20. DOI: `10.1038/nbt.4096`
191. Hafemeister C and Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genom Biol. 2019; 20:296. DOI: `10.1186/s13059-019-1874-1`
192. Goodyer WR, Beyersdorf BM, Paik DT, Tian L, Li G, Buikema JW, et al. Transcriptomic profiling of the developing cardiac conduction system at single-cell resolution. Circ Res. 2019; 125:379–97. DOI: `10.1161/CIRCRESAHA.118.314578`
193. Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, et al. MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol. 2015; 16:278. DOI: `10.1186/s13059-015-0844-5`
194. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, and Barabási AL. The human disease network. Proc Natl Acad Sci U S A. 2007; 104:8685–90. DOI: `10.1073/pnas.0701361104`
195. Barabási AL, Gulbahce N, and Loscalzo J. Network medicine: A network-based approach to human disease. Nat Rev Genet. 2011; 12:56–68. DOI: `10.1038/nrg2918`
196. Hartwell LH, Hopfield JJ, Leibler S, and Murray AW. From molecular to modular cell biology. Nature. 1999; 402:C47–C52. DOI: `10.1038/35011540`

197. Fischer A and Gessler M. Hey genes in cardiovascular development. Trends Cardiovasc Med. 2003; 13:221–6. DOI: `10.1016/S1050-1738(03)00082-3`

198. Gan P, Patterson M, and Sucov HM. Cardiomyocyte polyploidy and implications for heart regeneration. Annu Rev Physiol. 2020; 82:45–61. DOI: `10.1146/annurev-physiol-021119-034618`

199. Soonpaa MH, Kim KK, Pajak L, Franklin MT, and Field LJ. Cardiomyocyte DNA synthesis and binucleation during murine development. Am J Physiol Heart Circ Physiol. 1996; 271:H2183–H2189. DOI: `10.1152/ajpheart.1996.271.5.H2183`

200. Patterson M, Barske L, Van Handel B, Rau CD, Gan P, Sharma A, et al. Frequency of mononuclear diploid cardiomyocytes underlies natural variation in heart regeneration. Nat Genet. 2017; 49:1346–53. DOI: `10.1038/ng.3929`

201. Gramolini AO, Trivieri MG, Oudit GY, Kislinger T, Li W, Patel MM, et al. Cardiac-specific overexpression of sarcolipin in phospholamban null mice impairs myocyte function that is restored by phosphorylation. Proc Natl Acad Sci U S A. 2006; 103:2446–51. DOI: `10.1073/pnas.0510883103`

202. Hirose K, Payumo AY, Cutie S, Hoang A, Zhang H, Guyot R, et al. Evidence for hormonal control of heart regenerative capacity during endothermy acquisition. Science. 2019; 364:184–8. DOI: `10.1126/science.aar2038`

203. Xin M, Small EM, Van Rooij E, Qi X, Richardson JA, Srivastava D, et al. Essential roles of the bHLH transcription factor Hrt2 in repression of atrial gene expression and maintenance of postnatal cardiac function. Proc Natl Acad Sci U S A. 2007; 104:7975–80. DOI: `10.1073/pnas.0702447104`

204. Gibb N, Lazic S, Yuan X, Deshwar AR, Leslie M, Wilson MD, et al. Hey2 regulates the size of the cardiac progenitor pool during vertebrate heart development. Development. 2018; 145:dev167510. DOI: `10.1242/dev.167510`

205. van Walree ES, Dombrowsky G, Jansen IE, Mirkov MU, Zwart R, Ilgun A, et al. Germline variants in HEY2 functional domains lead to congenital heart defects and thoracic aortic aneurysms. Genet Med. 2021; 23:103–10. DOI: `10.1038/s41436-020-00939-4`

206. She P, Zhang H, Peng X, Sun J, Gao B, Zhou Y, et al. The Gridlock transcriptional repressor impedes vertebrate heart regeneration by restricting expression of lysine methyltransferase. Development. 2020; 147:dev190678. DOI: `10.1242/dev.190678`

207. Windmueller R, Leach JP, Babu A, Zhou S, Morley MP, Wakabayashi A, et al. Direct Comparison of Mononucleated and Binucleated Cardiomyocytes Reveals Molecular Mechanisms Underlying Distinct Proliferative Competencies. Cell Rep. 2020; 30:3105–3116.e4. DOI: `10.1016/j.celrep.2020.02.034`

208. Yekelchyk M, Guenther S, Preussner J, and Braun T. Mono- and multi-nucleated ventricular cardiomyocytes constitute a transcriptionally homogenous cell population. Basic Res Cardiol. 2019; 114:36. DOI: `10.1007/s00395-019-0744-z`

209. Wang D, Tai PW, and Gao G. Adeno-associated virus vector as a platform for gene therapy delivery. 2019. DOI: `10.1038/s41573-019-0012-9`

210. Clavería C, Giovinazzo G, Sierra R, and Torres M. Myc-driven endogenous cell competition in the early mammalian embryo. Nature. 2013; 500:39–44. DOI: `10.1038/nature12389`

211. Zhao B, Li L, Tumaneng K, Wang CY, and Guan KL. A coordinated phosphorylation by Lats and CK1 regulates YAP stability through SCF(beta-TRCP). Genes Dev. 2010; 24:72–85. DOI: `10.1101/gad.1843810`

212. Herzog VA, Reichholf B, Neumann T, Rescheneder P, Bhat P, Burkard TR, et al. Thiol-linked alkylation of RNA to assess expression dynamics. Nat Methods. 2017; 14:1198–204. DOI: `10.1038/nmeth.4435`

213. Herzog VA, Fasching N, and Ameres SL. Determining mRNA stability by metabolic RNA labeling and chemical nucleoside conversion. Methods in Molecular Biology. Vol. 2062. 2020 :169–89

214. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013; 29:15–21. DOI: `10.1093/bioinformatics/bts635`

215. Jürges C, Dölken L, and Erhard F. Dissecting newly transcribed and old RNA using GRAND-SLAM. Bioinformatics. 2018; 34:i218–i226. DOI: `10.1093/bioinformatics/bty256`

216. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021; 10:giab008. DOI: `10.1093/gigascience/giab008`

217. Wang L, Wang S, and Li W. RSeQC: quality control of RNA-seq experiments. Bioinformatics. 2012; 28(16):2184–5. DOI: `10.1093/bioinformatics/bts356`

218. Love MI, Huber W, and Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15:550. DOI: `10.1186/s13059-014-0550-8`

219. Zhu A, Ibrahim JG, and Love MI. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics. 2019; 35:2084–92. DOI: `10.1093/bioinformatics/bty895`

220. Anders S and Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010; 11:R106. DOI: `10.1186/gb-2010-11-10-r106`

221. Dey KK, Hsiao CJ, and Stephens M. Visualizing the structure of RNA-seq expression data using grade of membership models. PLoS Genet. 2017; 13:e1006759. DOI: `10.1371/journal.pgen.1006759`

222. Carbonetto P, Luo K, Sarkar A, Hung A, Tayeb K, Pott S, et al. GoM DE: interpreting structure in sequence count data with differential expression analysis allowing for grades of membership. Genome Biol. 2023; 24:236. DOI: `10.1186/s13059-023-03067-9`

223. Muhar M, Ebert A, Neumann T, Umkehrer C, Jude J, Wieshofer C, et al. SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis. Science. 2018; 360:800–5. DOI: `10.1126/science.aao2793`

224. Cooke A, Schwarzl T, Huppertz I, Kramer G, Mantas P, Alleaume AM, et al. The RNA-Binding Protein YBX3 Controls Amino Acid Levels by Regulating SLC mRNA Abundance. Cell Rep. 2019; 27:3097–3106.e5. DOI: `10.1016/j.celrep.2019.05.039`
225. Cao J and Poss KD. The epicardium as a hub for heart regeneration. Nat Rev Cardiol. 2018; 15:631–47. DOI: `10.1038/s41569-018-0046-4`
226. Sanz-Morejón A, García-Redondo AB, Reuter H, Marques IJ, Bates T, Galardi-Castilla M, et al. Wilms tumor 1b expression defines a pro-regenerative macrophage subtype and is required for organ regeneration in the zebrafish. Cell Rep. 2019; 28:1296–1306.e6. DOI: `10.1016/j.celrep.2019.06.091`
227. Cao Y, Xia Y, Balowski JJ, Ou J, Song L, Safi A, et al. Identification of enhancer regulatory elements that direct epicardial gene expression during zebrafish heart regeneration. Development. 2022; 149:dev200133. DOI: `10.1242/dev.200133`
228. Felkin LE, Narita T, Germack R, Shintani Y, Takahashi K, Sarathchandra P, et al. Calcineurin splicing variant calcineurin Aβ1 improves cardiac function after myocardial infarction without inducing hypertrophy. Circulation. 2011; 123:2838–47. DOI: `10.1161/CIRCULATIONAHA.110.012211`
229. Gao F, Liang T, Lu YW, Pu L, Fu X, Dong X, et al. Reduced mitochondrial protein translation promotes cardiomyocyte proliferation and heart regeneration. Circulation. 2023; 148:1887–906. DOI: `10.1161/CIRCULATIONAHA.122.061192`
230. Fang Y, Gupta V, Karra R, Holdway JE, Kikuchi K, and Poss KD. Translational profiling of cardiomyocytes identifies an early Jak1/Stat3 injury response required for zebrafish heart regeneration. Proc Natl Acad Sci U S A. 2013; 110:13416–21. DOI: `10.1073/pnas.1309810110`
231. O'Meara CC, Wamstad JA, Gladstone RA, Fomovsky GM, Butty VL, Shrikumar A, et al. Transcriptional reversion of cardiac myocyte fate during mammalian cardiac regeneration. Circ Res 2015; 116:804–15. DOI: `10.1161/CIRCRESAHA.116.304269`
232. Paltzer WG, Aballo TJ, Bae J, Flynn CGK, Wanless KN, Hubert KA, et al. mTORC1 regulates the metabolic switch of postnatal cardiomyocytes during regeneration. J Mol Cell Cardiol. 2024; 187:15–25. DOI: `10.1016/j.yjmcc.2023.12.004`
233. Eulalio A, Mano M, Ferro MD, Zentilin L, Sinagra G, Zacchigna S, et al. Functional screening identifies miRNAs inducing cardiac regeneration. Nature. 2012; 492:376–81. DOI: `10.1038/nature11739`
234. Singh BN, Koyano-Nakagawa N, Gong W, Moskowitz IP, Weaver CV, Braunlin E, et al. A conserved HH-Gli1-Mycn network regulates heart regeneration from newt to human. Nat Commun. 2018; 9:4237. DOI: `10.1038/s41467-018-06617-z`
235. Wu HY, Zhou YM, Liao ZQ, Zhong JW, Liu YB, Zhao H, et al. Fosl1 is vital to heart regeneration upon apex resection in adult Xenopus tropicalis. NPJ Regen Med. 2021; 6:36. DOI: `10.1038/s41536-021-00146-y`

236. Polikarpova A, Ellinghaus A, Schmidt-Bleek O, Grosser L, Bucher CH, Duda GN, et al. The specialist in regeneration—the Axolotl—a suitable model to study bone healing? NPJ Regen Med. 2022; 7:35. DOI: `10.1038/s41536-022-00229-4`
237. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019; 16:1289–96. DOI: `10.1038/s41592-019-0619-0`
238. Angerer P, Haghverdi L, Büttner M, Theis FJ, Marr C, and Buettner F. destiny: diffusion maps for large-scale single-cell data in R. Bioinformatics. 2019; 32:1241–3. DOI: `10.1093/bioinformatics/btv715`
239. Schreiber T. Measuring information transfer. Phys Rev Lett. 2000; 85:461. DOI: `10.1103/PhysRevLett.85.461`
240. Germain PL, Sonrel A, and Robinson MD. pipeComp, a general framework for the evaluation of computational pipelines, reveals performant single cell RNA-seq preprocessing toolsR. Genome Biol. 2020; 21:227. DOI: `10.1186/s13059-020-02136-7`
241. Townes FW, Hicks SC, Aryee MJ, and Irizarry RA. Feature selection and dimension reduction for single-cell RNA-Seq based on a multinomial model. Genome Biol. 2019; 20:295. DOI: `10.1186/s13059-019-1861-6`
242. Van de Sande B, Flerin C, Davie K, Waegeneer MD, Hulselmans G, Aibar S, et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nat Protoc. 2020; 15:2247–76. DOI: `10.1038/s41596-020-0336-2`
243. Shen WK, Chen SY, Gan ZQ, Zhang YZ, Yue T, Chen MM, et al. AnimalTFDB 4.0: a comprehensive animal transcription factor database updated with variation and expression annotations. Nucleic Acids Res. 2023; 51:D39–D45. DOI: `10.1093/nar/gkac907`
244. LeCun Y, Bengio Y, and Hinton G. Deep learning. Nature. 2015; 521:436–44. DOI: `10.1038/nature14539`
245. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008; 9:R137. DOI: `10.1186/gb-2008-9-9-r137`
246. Yu G, Wang LG, and He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics. 2015; 31:2382–3. DOI: `10.1093/bioinformatics/btv145`
247. Wang Q, Li M, Wu T, Zhan L, Li L, Chen M, et al. Exploring epigenomic datasets by ChIPseeker. Curr Protoc. 2022; 2:e585. DOI: `10.1002/cpz1.585`
248. Grant CE, Bailey TL, and Noble WS. FIMO: scanning for occurrences of a given motif. Bioinformatics. 2011; 27:1017–8. DOI: `10.1093/bioinformatics/btr064`
249. Kanehisa M and Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000; 28:27–30. DOI: `10.1093/nar/28.1.27`
250. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. Protein Sci. 2019; 28:1947–51. DOI: `10.1002/pro.3715`

251. Hutchison C, Pilote M, and Roy S. The axolotl limb: a model for bone development, regeneration and fracture healing. Bone. 2007; 40:46–56. DOI: `10.1016/j.bone.2006.07.005`

252. Barabási AL and Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet. 2004; 5:101–13. DOI: `10.1038/nrg1272`

253. Barabási AL and Albert R. Emergence of scaling in random networks. Science. 1999; 286:509–12. DOI: `10.1126/science.286.5439.509`

254. Koschützki D and Schreiber F. Centrality analysis methods for biological networks and their application to gene regulatory networks. Gene Regul Syst Bio. 2008; 2:193–201. DOI: `10.4137/grsb.s702`

255. Freeman LC. A set of measures of centrality based on betweenness. Sociometry. 1977; 40:35–41. DOI: `10.2307/3033543`

256. Hahn MW and Kern AD. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Mol Biol Evol. 2005; 22:803–6. DOI: `10.1093/molbev/msi072`

257. Potapov AP, Voss N, Sasse N, and Wingender E. Topology of mammalian transcription networks. Genomics Inform. 2005; 16:270–8. DOI: `10.11234/gi1990.16.2_270`

258. Smith A, Calder CA, and Browning CR. Empirical reference distributions for networks of different size. Soc Networks. 2016; 47:24–37. DOI: `10.1016/j.socnet.2016.03.004`

259. Cadigan KM and Waterman ML. TCF/LEFs and Wnt signaling in the nucleus. Cold Spring Harb Perspect Biol. 2012; 4:a007906. DOI: `10.1101/cshperspect.a007906`

260. Clevers H and Nusse R. Wnt/beta-catenin signaling and disease. Cell. 2012; 149:1192–205. DOI: `10.1016/j.cell.2012.05.012`

261. Gao B, Ajima R, Yang W, Li C, Song H, Anderson MJ, et al. Coordinated directional outgrowth and pattern formation by integration of Wnt5a and Fgf signaling in planar cell polarity. Development. 2018; 145:dev163824. DOI: `10.1242/dev.163824`

262. Glotzer GL, Tardivo P, and Tanaka EM. Canonical Wnt signaling and the regulation of divergent mesenchymal Fgf8 expression in axolotl limb development and regeneration. Elife. 2022; 11:e79762. DOI: `10.7554/eLife.79762`

263. Liu F, van den Broek O, Destrée O, and Hoppler S. Distinct roles for Xenopus Tcf/Lef genes in mediating specific responses to Wnt/beta-catenin signalling in mesoderm development. Development. 2005; 132:5375–85. DOI: `10.1242/dev.02152`

264. Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, et al. Perturb-seq: Dissecting molecular circuits with scalable single cell RNA profiling of pooled genetic screens. Cell. 2017; 167:1853–1866.e17. DOI: `10.1016/j.cell.2016.11.038`

265. Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, Camacho DM, et al. Wisdom of crowds for robust gene network inference. Nat Methods. 2012; 9:796–804. DOI: `10.1038/nmeth.2016`

266. Feuerriegel S, Frauen D, Melnychuk V, Schweisthal J, Hess K, Curth A, et al. Causal machine learning for predicting treatment outcomes. Nat Med. 2024; 30:958–68. DOI: `10.1038/s41591-024-02902-1`

267. Liu X, De la Cruz E, Gu X, Balint L, Oxendine-Burns M, Terrones T, et al. Lymphoangiocrine signals promote cardiac growth and repair. Nature. 2020; 588:705–11. DOI: `10.1038/s41586-020-2998-x`

268. Xie R, Yuan S, Hu G, Zhan J, Jin K, Tang Y, et al. Nuclear AGO2 promotes myocardial remodeling by activating ANKRD1 transcription in failing hearts. Mol Ther. 2024; 32:1578–94. DOI: `10.1016/j.ymthe.2024.03.018`

269. van der Pijl RJ, van den Berg M, van de Locht M, Shen S, Bogaards SJ, Conijn S, et al. Muscle ankyrin repeat protein 1 (MARP1) locks titin to the sarcomeric thin filament and is a passive force regulator. J Gen Physiol. 2021; 153:e202112925. DOI: `10.1085/jgp.202112925`

270. Arimura T, Bos JM, Sato A, Kubo T, Okamoto H, Nishi H, et al. Cardiac Ankyrin Repeat Protein Gene (ANKRD1) Mutations in Hypertrophic Cardiomyopathy. J Am Coll Cardiol. 2009; 54:334–42. DOI: `10.1016/j.jacc.2008.12.082`

271. Duboscq-Bidot L, Charron P, Ruppert V, Fauchier L, Richter A, Tavazzi L, et al. Mutations in the ANKRD1 gene encoding CARP are responsible for human dilated cardiomyopathy. Eur Heart J. 2009; 30:2128–36. DOI: `10.1093/eurheartj/ehp225`

272. Xu Z, Lu D, Yuan J, Wang L, Wang J, Lei Z, et al. Storax attenuates cardiac fibrosis following acute myocardial infarction in rats via suppression of AT1R–Ankrd1–P53 signaling pathway. Int J Mol Sci. 2022; 23:13161. DOI: `10.3390/ijms232113161`

273. Boskovic S, Juez RM, Stamenkovic N, Radojkovic D, Stainier DYR, and Kojic S. The stress responsive gene ankrd1a is dynamically regulated during skeletal muscle development and upregulated following cardiac injury in border zone cardiomyocytes in adult zebrafish. Gene. 2021; 792:145725. DOI: `10.1016/j.gene.2021.145725`

274. Guimarães-Camboa N, Stowe J, Aneas I, Sakabe N, Cattaneo P, Henderson L, et al. HIF1α represses cell stress pathways to allow proliferation of hypoxic fetal cardiomyocytes. Dev Cell. 2015; 33:507–21. DOI: `10.1016/j.devcel.2015.04.021`

275. Zhao K, Huang X, Zhao W, Lu B, and Yang Z. LONP1-mediated mitochondrial quality control safeguards metabolic shifts in heart development. Development. 2022; 149:dev200458. DOI: `10.1242/dev.200458`

276. Nabeebaccus AA, Zoccarato A, Hafstad AD, Santos CXC, Aasum E, Brewer AC, et al. Nox4 reprograms cardiac substrate metabolism via protein O-GlcNAcylation to enhance stress adaptation. JCI Insight. 2017; 2:e96184. DOI: `10.1172/jci.insight.96184`

277. Sun F, Ou J, Shoffner AR, Luan Y, Yang H, Song L, et al. Enhancer selection dictates gene expression responses in remote organs during tissue regeneration. Nat Cell Biol. 2022; 24:685–96. DOI: `10.1038/s41556-022-00906-y`

278. Zhang J, Chang L, Chen C, Zhang M, Luo Y, Hamblin M, et al. Rad GTPase inhibits cardiac fibrosis through connective tissue growth factor. Cardiovasc Res. 2011; 91:90–8. DOI: `10.1093/cvr/cvr068`

279. De Pauw A, Andre E, Sekkali B, Bouzin C, Esfahani H, Barbier N, et al. Dnmt3a-mediated inhibition of Wnt in cardiac progenitor cells improves differentiation and remote remodeling after infarction. JCI Insight. 2017; 2(12):e91810. DOI: `10.1172/jci.insight.91810`

280. Fang X, Poulsen RR, Wang-Hu J, Shi O, Calvo NS, Simmons CS, et al. Knockdown of DNA methyltransferase 3a alters gene expression and inhibits function of embryonic cardiomyocytes. FASEB J. 2016; 30:3238–55. DOI: `10.1096/fj.201600346R`

281. Han P, Li W, Yang J, Shang C, Lin CH, Cheng W, et al. Epigenetic response to environmental stress: Assembly of BRG1–G9a/GLP–DNMT3 repressive chromatin complex on Myh6 promoter in pathologically stressed hearts. Biochim Biophys Acta. 2016; 1863:1772–81. DOI: `10.1016/j.bbamcr.2016.03.002`

282. Aslan GS, Polat F, Eren SN, Yucel D, Arbatli S, Cumbul A, et al. Identification of novel and potent modulators involved in neonatal cardiac regeneration. Pediatr Cardiol. 2021; 42:1554–66. DOI: `10.1007/s00246-021-02640-y`

283. Suda M, Shimizu I, Yoshida Y, Hayashi Y, Ikegami R, Katsuumi G, et al. Inhibition of dipeptidyl peptidase-4 ameliorates cardiac ischemia and systolic dysfunction by up-regulating the FGF-2/EGR-1 pathway. PLoS One. 2017; 12:e0182422. DOI: `10.1371/journal.pone.0182422`

284. Zhou J, Yao Y, Zhang J, Wang Z, Zheng T, Lu Y, et al. JNK-dependent phosphorylation and nuclear translocation of EGR-1 promotes cardiomyocyte apoptosis. Apoptosis. 2022; 27:246–60. DOI: `10.1007/s10495-022-01714-3`

285. Zhang Y, Liao H, Zhong S, Gao F, Chen Y, Huang Z, et al. Effect of N-n-butyl haloperidol iodide on ROS/JNK/Egr-1 signaling in H9c2 cells after hypoxia/reoxygenation. Sci Rep. 2015; 5:11809. DOI: `10.1038/srep11809`

286. Feng Y, Desjardins CA, Cooper O, Kontor A, Nocco SE, and Naya FJ. EGR1 functions as a potent repressor of MEF2 transcriptional activity. PLoS One. 2015; 10:e0127641. DOI: `10.1371/journal.pone.0127641`

287. Grossfeld P. ETS1 and HLHS: implications for the role of the endocardium. J Cardiovasc Dev Dis. 2022; 9:219. DOI: `10.3390/jcdd9070219`

288. Ruan H, Liao Y, Ren Z, Mao L, Yao F, Yu P, et al. Single-cell reconstruction of differentiation trajectory reveals a critical role of ETS1 in human cardiac lineage commitment. BMC Biol. 2019; 17:89. DOI: `10.1186/s12915-019-0709-6`

289. Zhou P, Zhang Y, Sethi I, Ye L, Trembley MA, Cao Y, et al. GATA4 regulates developing endocardium through interaction with ETS1. Circ Res. 2022; 131:e152–e168. DOI: `10.1161/CIRCRESAHA.120.318102`

290. Wang L, Lin L, Qi H, Chen J, and Grossfeld P. Endothelial loss of ETS1 impairs coronary vascular development and leads to ventricular non-compaction. Circ Res. 2022; 131:371–87. DOI: `10.1161/CIRCRESAHA.121.319955`

291. Gawdzik JC, Yue MS, Martin NR, Elemans LMH, Lanham KA, Heideman W, et al. sox9b is required in cardiomyocytes for cardiac morphogenesis and function. Sci Rep. 2018; 8:13906. DOI: `10.1038/s41598-018-32125-7`

292. Liu Q, Jiang C, Xu J, Zhao MT, Van Bortle K, Cheng X, et al. Genome-wide temporal profiling of transcriptome and open chromatin of early cardiomyocyte differentiation derived from hiPSCs and hESCs. Circ Res. 2017; 121:376–91. DOI: `10.1161/CIRCRESAHA.116.310456`

293. Palomer X, Capdevila-Busquets E, Botteri G, Davidson MM, Rodríguez C, Martínez-González J, et al. miR-146a targets c-Fos expression in human cardiac cells. Dis Model Mech. 2015; 8:1081–91. DOI: `10.1242/dmm.020768`

294. Kato R, Nomura A, Sakamoto A, Yasuda Y, Amatani K, Nagai S, et al. Hydrogen gas attenuates embryonic gene expression and prevents left ventricular remodeling induced by intermittent hypoxia in cardiomyopathic hamsters. Am J Physiol Heart Circ Physiol. 2014; 307:H1626–H1633. DOI: `10.1152/ajpheart.00228.2014`

295. Jeong MY, Kinugawa K, Vinson C, and Long CS. AFos dissociates cardiac myocyte hypertrophy and expression of the pathological gene program. Circulation. 2005; 111:1645–51. DOI: `10.1161/01.CIR.0000160367.99928.87`

296. Babu GJ, Lalli JM, Sussman MA, Sadoshima Ji, and Periasamy M. Phosphorylation of Elk-1 by MEK/ERK pathway is necessary for c-fos gene activation during cardiac myocyte hypertrophy. J Mol Cell Cardiol. 2000; 32:1447–57. DOI: `10.1006/jmcc.2000.1185`

297. Calderon MJ, Ploegman AG, Bailey B, Jung DO, Navratil AM, and Ellsworth BS. Loss of Foxm1 results in reduced somatotrope cell number during mouse embryogenesis. PLoS One. 2015; 10:e0128942. DOI: `10.1371/journal.pone.0128942`

298. Bolte C, Zhang Y, Wang IC, Kalin TV, Molkentin JD, and Kalinichenko VV. Expression of Foxm1 transcription factor in cardiomyocytes is required for myocardial development. PLoS One. 2011; 6:e22217. DOI: `10.1371/journal.pone.0022217`

299. Sengupta A, Kalinichenko VV, and Yutzey KE. FoxO1 and FoxM1 transcription factors have antagonistic functions in neonatal cardiomyocyte cell-cycle withdrawal and IGF1 gene regulation. Circ Res. 2013; 112:267–77. DOI: `10.1161/CIRCRESAHA.112.277442`

300. Smeitink JAM, Elpeleg O, Antonicka H, Diepstra H, Saada A, Smits P, et al. Distinct clinical phenotypes associated with a mutation in the mitochondrial translation elongation factor EFTs. Am J Hum Genet. 2006; 79:869–77. DOI: `10.1086/508434`

301. Zuppo DA, Missinato MA, Santana-Santos L, Li G, Benos PV, and Tsang M. Foxm1 regulates cardiomyocyte proliferation in adult zebrafish after cardiac injury. Development. 2023; 150:dev201163. DOI: `10.1242/dev.201163`

302. Voronova A, Al Madhoun A, Fischer A, Shelton M, Karamboulas C, and Skerjanc IS. Gli2 and MEF2C activate each other's expression and function synergistically during cardiomyogenesis in vitro. Nucleic Acids Res. 2012; 40:3329–47. DOI: `10.1093/nar/gkr1232`

303. Gianakopoulos PJ and Skerjanc IS. Hedgehog signaling induces cardiomyogenesis in P19 cells. J Biol Chem. 2005; 280:21022–8. DOI: `10.1074/jbc.M502977200`

304. Fair JV, Voronova A, Bosiljcic N, Rajgara R, Blais A, and Skerjanc IS. BRG1 interacts with GLI2 and binds Mef2c gene in a hedgehog signalling dependent manner during in vitro cardiomyogenesis. BMC Dev Biol. 2016; 16:27. DOI: `10.1186/s12861-016-0127-8`

305. Knutson AK, Williams AL, Boisvert WA, and Shohet RV. HIF in the heart: development, metabolism, ischemia, and atherosclerosis. J Clin Invest. 2021; 131:e137557. DOI: `10.1172/JCI137557`

306. Menendez-Montes I, Escobar B, Gomez MJ, Albendea-Gomez T, Palacios B, Bonzon-Kulichenko E, et al. Activation of amino acid metabolic program in cardiac HIF1-alpha-deficient mice. iScience. 2021; 24:102124. DOI: `10.1016/j.isci.2021.102124`

307. Yokura-Yamada Y, Araki M, and Maeda M. Ectopic expression of Id1 or Id3 inhibits transcription of the GATA-4 gene in P19CL6 cells under differentiation condition. Drug Discov Ther. 2021; 15(4):189–96. DOI: `10.5582/ddt.2021.01069`

308. Du M, Jiang H, Liu H, Zhao X, Zhou Y, Zhou F, et al. Single-cell RNA sequencing reveals that BMPR2 mutation regulates right ventricular function via ID genes. Eur Respir J. 2022; 60:2100327. DOI: `10.1183/13993003.00327-2021`

309. Becher PM, Hinrichs S, Fluschnik N, Hennigs JK, Klingel K, Blankenberg S, et al. Role of Toll-like receptors and interferon regulatory factors in different experimental heart failure models of diverse etiology: IRF7 as novel cardiovascular stress-inducible factor. PLoS One. 2018; 13:e0193844. DOI: `10.1371/journal.pone.0193844`

310. Long H, Steimle JD, Grisanti Canozo FJ, Kim JH, Li X, Morikawa Y, et al. Endothelial cells adopt a pro-reparative immune responsive signature during cardiac injury. Life Sci Alliance. 2024; 7:e202201870. DOI: `10.26508/lsa.202201870`

311. Velecela V, Lettice LA, Chau YY, Slight J, Berry RL, Thornburn A, et al. WT1 regulates the expression of inhibitory chemokines during heart development. Hum Mol Genet. 2013; 22:5083–95. DOI: `10.1093/hmg/ddt358`

312. Fang Y, Wang S, Lv J, Zhao Z, Guo N, Wu G, et al. Slc39a2-Mediated Zinc Homeostasis Modulates Innate Immune Signaling in Phenylephrine-Induced Cardiomyocyte Hypertrophy. Front Cardiovasc Med. 2021; 8:736911. DOI: `10.3389/fcvm.2021.736911`

313. Jiang DS, Liu Y, Zhou H, Zhang Y, Zhang XD, Zhang XF, et al. Interferon regulatory factor 7 functions as a novel negative regulator of pathological cardiac hypertrophy. Hypertension. 2014; 63:713–22. DOI: `10.1161/HYPERTENSIONAHA.113.02653`

314. Hilfiker-Kleiner D, Hilfiker A, Kaminski K, Schaefer A, Park JK, Michel K, et al. Lack of JunD promotes pressure overload–induced apoptosis, hypertrophic growth, and angiogenesis in the heart. Circulation. 2005; 112:1470–7. DOI: `10.1161/CIRCULATIONAHA.104.518472`

315. Akhmedov A, Montecucco F, Costantino S, Vdovenko D, Schaub Clerigué A, Gaul DS, et al. Cardiomyocyte-specific JunD overexpression increases infarct size following ischemia/reperfusion cardiac injury by downregulating Sirt3. Thromb Haemost. 2020; 120:168–80. DOI: `10.1055/s-0039-3400299`

316. Mani SK, Egan EA, Addy BK, Grimm M, Kasiganesan H, Thiyagarajan T, et al. β-Adrenergic receptor stimulated Ncx1 upregulation is mediated via a CaMKII/AP-1 signaling pathway in adult cardiomyocytes. J Mol Cell Cardiol. 2010; 48:342–51. DOI: `10.1016/j.yjmcc.2009.11.007`

317. Schneiders D, Heger J, Best P, Michaelpiper H, and Taimor G. SMAD proteins are involved in apoptosis induction in ventricular cardiomyocytes. Cardiovasc Res. 2005; 67:87–96. DOI: `10.1016/j.cardiores.2005.02.021`

318. Meder B, Just S, Vogel B, Rudloff J, Gärtner L, Dahme T, et al. JunB-CBFβ signaling is essential to maintain sarcomeric Z-disc structure and when defective leads to heart failure. J Cell Sci. 2010; 123:2613–20. DOI: `10.1242/jcs.067967`

319. Yan M, Yang S, Meng F, Zhao Z, Tian Z, and Yang P. MicroRNA 199a-5p induces apoptosis by targeting JunB. Sci Rep. 2018; 8:6699. DOI: `10.1038/s41598-018-24932-9`

320. Zhu Y, Ackers-Johnson M, Shanmugam MK, Pakkiri LS, Drum CL, Chen Y, et al. Asparagine synthetase marks a distinct dependency threshold for cardiomyocyte dedifferentiation. Circulation. 2024; 149:1833–51. DOI: `10.1161/CIRCULATIONAHA.123.063965`

321. Chen Y, Lüttmann FF, Schoger E, Schöler HR, Zelarayán LC, Kim KP, et al. Reversible reprogramming of cardiomyocytes to a fetal state drives heart regeneration in mice. Science. 2021; 373:1537–40. DOI: `10.1126/science.abg5159`

322. Kisby T, de Lázaro I, Stylianou M, Cossu G, and Kostarelos K. Transient reprogramming of postnatal cardiomyocytes to a dedifferentiated state. PLoS One. 2021; 16:e0251054. DOI: `10.1371/journal.pone.0251054`

323. Yan H, Yi S, Zhuang H, Wu L, Wang D, and Jiang J. Sphingosine-1-phosphate ameliorates the cardiac hypertrophic response through inhibiting the activity of histone deacetylase-2. Int J Mol Med. 2017; 41:1704–14. DOI: `10.3892/ijmm.2017.3325`

324. Yoshida T, Yamashita M, Horimai C, and Hayashi M. Kruppel-like factor 4 protein regulates isoproterenol-induced cardiac hypertrophy by modulating myocardin expression and activity. J Biol Chem. 2014; 289:26107–18. DOI: `10.1074/jbc.M114.582809`

325. Liang X, Yadav SP, Batz ZA, Nellissery J, and Swaroop A. Protein kinase CK2 modulates the activity of Maf-family bZIP transcription factor NRL in rod photoreceptors of mammalian retina. Hum Mol Genet. 2023; 32:948–58. DOI: `10.1093/hmg/ddac256`

326. Kennedy KAM, Porter T, Mehta V, Ryan SD, Price F, Peshdary V, et al. Retinoic acid enhances skeletal muscle progenitor formation and bypasses inhibition by bone morphogenetic protein 4 but not dominant negative β-catenin. BMC Biol. 2009; 7:67. DOI: `10.1186/1741-7007-7-67`

327. Liu X, Yin K, Chen L, Chen W, Li W, Zhang T, et al. Lineage-specific regulatory changes in hypertrophic cardiomyopathy unraveled by single-nucleus RNA-seq and spatial transcriptomics. Cell Discov. 2023; 9:6. DOI: `10.1038/s41421-022-00490-3`

328. Tani H, Sadahiro T, Yamada Y, Isomi M, Yamakawa H, Fujita R, et al. Direct reprogramming improves cardiac function and reverses fibrosis in chronic myocardial infarction. Circulation. 2023; 147:223–38. DOI: `10.1161/CIRCULATIONAHA.121.058655`

329. Villa del Campo C, Lioux G, Carmona R, Sierra R, Muñoz-Chápuli R, Clavería C, et al. Myc overexpression enhances epicardial contribution to the developing heart and promotes extensive expansion of the cardiomyocyte population. Sci Rep. 2016; 6:37880. DOI: `10.1038/srep37880`

330. Villa del Campo C, Clavería C, Sierra R, and Torres M. Cell competition promotes phenotypically silent cardiomyocyte replacement in the mammalian heart. Cell Rep. 2014; 8:1741–51. DOI: `10.1016/j.celrep.2014.08.005`

331. Lee Hg, Chen Q, Wolfram JA, Richardson SL, Liner A, Siedlak SL, et al. Cell cycle re-entry and mitochondrial defects in Myc-mediated hypertrophic cardiomyopathy and heart failure. Plos One. 2009; 4:e7172. DOI: `10.1371/journal.pone.0007172`

332. Sendra M, de Dios Hourcade J, Temiño S, Sarabia AJ, Ocaña OH, Domínguez JN, et al. Cre recombinase microinjection for single-cell tracing and localised gene targeting. Development. 2023; 150:dev201206. DOI: `10.1242/dev.201206`

333. Harmelink C, Peng Y, DeBenedittis P, Chen H, Shou W, and Jiao K. Myocardial Mycn is essential for mouse ventricular wall morphogenesis. Dev Biol. 2013; 373:53–63. DOI: `10.1016/j.ydbio.2012.10.005`

334. Cao C, Li L, Zhang Q, Li H, Wang Z, Wang A, et al. Nkx2.5: a crucial regulator of cardiac development, regeneration and diseases. Front Cardiovasc Med. 2023; 10:1270951. DOI: `10.3389/fcvm.2023.1270951`

335. Douglas M, O'Loughlin C, Lynch AT, Prialgauskaite R, Adamson AD, Dibb KM, et al. The generation and validation of two NKX2-5 fluorescent reporter human embryonic stem cell lines: UMANe002-A-1 and UMANe002-A-2. Stem Cell Res. 2024; 74:103262. DOI: `10.1016/j.scr.2023.103262`

336. Koss M, Bolze A, Brendolan A, Saggese M, Capellini TD, Bojilova E, et al. Congenital asplenia in mice and humans with mutations in a Pbx/Nkx2-5/p15 module. Dev Cell. 2012; 22:913–26. DOI: `10.1016/j.devcel.2012.02.009`

337. de Sena-Tomás C, Aleman AG, Ford C, Varshney A, Yao D, Harrington JK, et al. Activation of Nkx2.5 transcriptional program is required for adult myocardial repair. Nat Commun. 2022; 13:2970. DOI: `10.1038/s41467-022-30468-4`

338. Avitabile D, Bailey B, Cottage CT, Sundararaman B, Joyo A, McGregor M, et al. Nucleolar stress is an early response to myocardial damage involving nucleolar proteins nucleostemin and nucleophosmin. Proc Natl Acad Sci U S A. 2011; 108:6145–50. DOI: `10.1073/pnas.1017935108`

339. Wang T, Wang Z, Fabritus L de, Tao J, Saied EM, Lee HJ, et al. 1-deoxysphingolipids bind to COUP-TF to modulate lymphatic and cardiac cell development. Dev Cell. 2021; 56:3128–3145.e15. DOI: `10.1016/j.devcel.2021.10.018`

340. Al Turki S, Manickaraj AK, Mercer CL, Gerety SS, Hitz MP, Lindsay S, et al. Rare variants in NR2F2 cause congenital heart defects in humans. Am J Hum Genet. 2014; 94:574–85. DOI: `10.1016/j.ajhg.2014.03.007`

341. Wu SP, Kao CY, Wang L, Creighton CJ, Yang J, Donti TR, et al. Increased COUP-TFII expression in adult hearts induces mitochondrial dysfunction resulting in heart failure. Nat Commun. 2015; 6:8245. DOI: `10.1038/ncomms9245`

342. Maréchal L, Sicotte B, Caron V, Brochu M, and Tremblay A. Fetal cardiac lipid sensing triggers an early and sex-related metabolic energy switch in intrauterine growth restriction. J Clin Endocrinol Metab. 2021; 106:3295–311. DOI: `10.1210/clinem/dgab496`

343. Dörmann N, Hammer E, Struckmann K, Rüdebusch J, Bartels K, Wenzel K, et al. Metabolic remodeling in cardiac hypertrophy and heart failure with reduced ejection fraction occurs independent of transcription factor EB in mice. Front Cardiovasc Med. 2024; 10:1323760. DOI: `10.3389/fcvm.2023.1323760`

344. Wiesinger A, Boink GJ, Christoffels VM, and Devalla HD. Retinoic acid signaling in heart development: Application in the differentiation of cardiovascular lineages from human pluripotent stem cells. Stem Cell Reports. 2021; 16:2589–606. DOI: `10.1016/j.stemcr.2021.09.010`

345. Da Silva F, Jian Motamedi F, Weerasinghe Arachchige LC, Tison A, Bradford ST, Lefebvre J, et al. Retinoic acid signaling is directly activated in cardiomyocytes and protects mouse hearts from apoptosis after myocardial infarction. Elife. 2021; 10:e68280. DOI: `10.7554/eLife.68280`

346. Harhous Z, Booz GW, Ovize M, Bidaux G, and Kurdi M. An update on the multifaceted roles of STAT3 in the heart. Front Cardiovasc Med. 2019; 6:150. DOI: `10.3389/fcvm.2019.00150`

347. Jacoby JJ, Kalinowski A, Liu MG, Zhang SSM, Gao Q, Chai GX, et al. Cardiomyocyte-restricted knockout of STAT3 results in higher sensitivity to inflammation, cardiac fibrosis, and heart failure with advanced age. Proc Natl Acad Sci U S A. 2003; 100:12929–34. DOI: `10.1073/pnas.2134694100`

348. Haghikia A, Ricke-Hoch M, Stapel B, Gorst I, and Hilfiker-Kleiner D. STAT3, a key regulator of cell-to-cell communication in the heart. Cardiovasc Res. 2014; 102:281–9. DOI: `10.1093/cvr/cvu034`

349. He Z, Sun H, Yu T, Yang Y, Zhang X, and Yang Y. Generation of a human embryonic stem cell line (SKLRMe004-A) carrying NKX2.5-EGFP and TBX5-Tdtomato dual fluorescent reporters. Stem Cell Res. 2023; 67:103015. DOI: `10.1016/j.scr.2023.103015`

350. Sweat ME, Cao Y, Zhang X, Burnicka-Turek O, Perez-Cervantes C, Kulandaisamy A, et al. Tbx5 maintains atrial identity in postnatal cardiomyocytes by regulating an atrial-specific enhancer network. Nat Cardiovasc Res. 2023; 2:881–98. DOI: `10.1038/s44161-023-00334-7`

351. Steimle JD and Moskowitz IP. TBX5: A key regulator of heart development. Curr Top Dev Biol. 2017; 122:195–221. DOI: `10.1016/bs.ctdb.2016.08.008`

352. van Ouwerkerk AF, Bosada FM, van Duijvenboden K, Houweling AC, Scholman KT, Wakker V, et al. Patient-specific TBX5-G125R variant induces profound transcriptional deregulation and atrial dysfunction. Circulation. 2022; 145:606–19. DOI: `10.1161/CIRCULATIONAHA.121.054347`

353. Helin K, Wu CL, Fattaey AR, Lees JA, Dynlacht BD, Ngwu C, et al. Heterodimerization of the transcription factors E2F-1 and DP-1 leads to cooperative trans-activation. Genes Dev. 1993; 7:1850–61. DOI: `10.1101/gad.7.10.1850`
354. Cutie S, Payumo AY, Lunn D, and Huang GN. In vitro and in vivo roles of glucocorticoid and vitamin D receptors in the control of neonatal cardiomyocyte proliferative potential. J Mol Cell Cardiol. 2020; 142:126–34. DOI: `10.1016/j.yjmcc.2020.04.013`
355. Placoná Diniz G, Antunes Lino C, Castilho Guedes E, Nascimento Moreira L do, and Barreto-Chaves MLM. Cardiac microRNA-133 is down-regulated in thyroid hormone-mediated cardiac hypertrophy partially via Type 1 Angiotensin II receptor. Basic Res Cardiol. 2015; 110:49. DOI: `10.1007/s00395-015-0504-7`
356. Rawal S, Nagesh PT, Coffey S, Van Hout I, Galvin IF, Bunton RW, et al. Early dysregulation of cardiac-specific microRNA-208a is linked to maladaptive cardiac remodelling in diabetic myocardium. Cardiovasc Diabetol. 2019; 18:13. DOI: `10.1186/s12933-019-0814-4`
357. Poon E, Keung W, Liang Y, Ramalingam R, Yan B, Zhang S, et al. Proteomic analysis of human pluripotent stem cell–derived, fetal, and adult ventricular cardiomyocytes reveals pathways crucial for cardiac metabolism and maturation. Circ Cardiovasc Genet. 2015; 8:427–36. DOI: `10.1161/CIRCGENETICS.114.000918`
358. Díaz del Moral S, Benaouicha M, Villa del Campo C, Torres M, Wagner N, Wagner KD, et al. Cardiomyocyte-specific Wt1 is involved in cardiac metabolism and response to damage. J Cardiovasc Dev Dis. 2023; 10(5):211. DOI: `10.3390/jcdd10050211`
359. Marques IJ, Ernst A, Arora P, Vianin A, Hetke T, Sanz-Morejón A, et al. Wt1 transcription factor impairs cardiomyocyte specification and drives a phenotypic switch from myocardium to epicardium. Development. 2022; 149:dev200375. DOI: `10.1242/dev.200375`
360. Awad S, Skipper W, Vostrejs W, Ozorowski K, Min K, Pfuhler L, et al. The YBX3 RNA-binding protein posttranscriptionally controls SLC1A5 mRNA in proliferating and differentiating skeletal muscle cells. J Biol Chem. 2024; 300:105602. DOI: `10.1016/j.jbc.2023.105602`
361. Cao X, Zhu N, Zhang Y, Chen Y, Zhang J, Li J, et al. Y-box protein 1 promotes hypoxia/reoxygenation- or ischemia/reperfusion-induced cardiomyocyte apoptosis via SHP-1-dependent STAT3 inactivation. J Cell Physiol. 2020; 235(11):8187–98. DOI: `10.1002/jcp.29474`
362. Heger J, Partsch S, Harjung C, Varga ZV, Baranyai T, Weiß J, et al. YB-1 Is a novel target for the inhibition of α-adrenergic-induced hypertrophy. Int J Mol Sci. 2023; 25:401. DOI: `10.3390/ijms25010401`

Supplementary Material

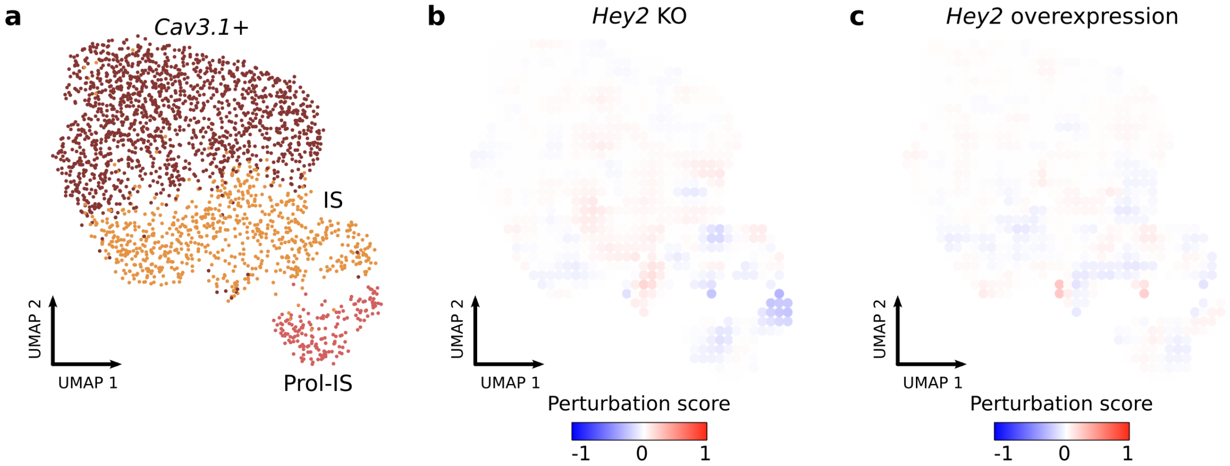


Figure S3.1: Effect of *Hey2 in silico* perturbations on the ventricular CM injury trajectory using a randomized version of the GRN. a) UMAP visualization of the ventricular CM populations involved in the injury trajectory. Cells are colored by cluster. **b-c)** Gridded UMAP visualization of the ventricular CM involved in the injury trajectory. Colors represent the effect of **b)** *Hey2* KO and **c)** *Hey2* overexpression calculated on a randomized version of the ventricular CM GRN. Red colors indicate that the perturbation promotes the differentiation trajectory of the cells, while blue colors indicate that the perturbation opposes the differentiation trajectory. White colors indicate that the perturbation had no effect.

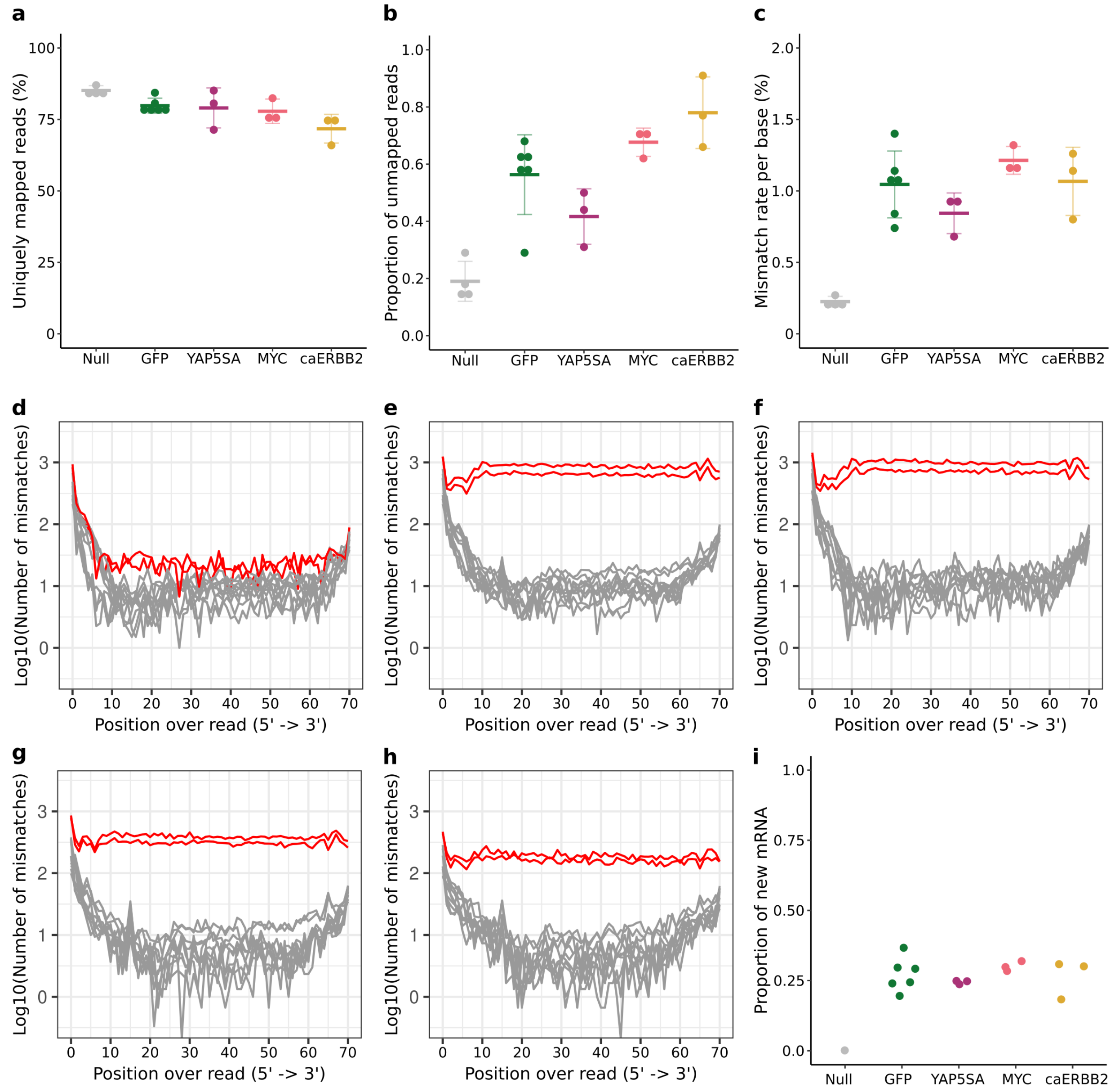


Figure S4.1: Quality control of SLAM-seq in transduced CM. a) Dotplot of the percentage of uniquely mapped read for each sample. **b)** Dotplot of the percentage of unmapped reads due to a high number of mismatches per sample. **c)** Dotplot of the mismatch rate per nucleotide per sample. **d-h)** Distribution of mismatches along the read length for **d)** samples without s4U, **e)** samples transduced with AAV9-*Gfp* and treated with s4U, **f)** samples transduced with AAV9-*Yap5sa* and treated with s4U, **g)** samples transduced with AAV9-*Myc* and treated with s4U, **h)** samples transduced with AAV9-*caErbb2* and treated with s4U. In all cases, the transduction lasted 44h and the s4U treatment lasted 4h. Red lines indicate mismatches introduced by SLAM-seq (A-to-G and T-to-C) while grey lines represent the other 10 possible mismatches. **i)** Dotplot of the percentage of newly transcribed mRNA in each sample, as estimated by GRAND-SLAM [215]

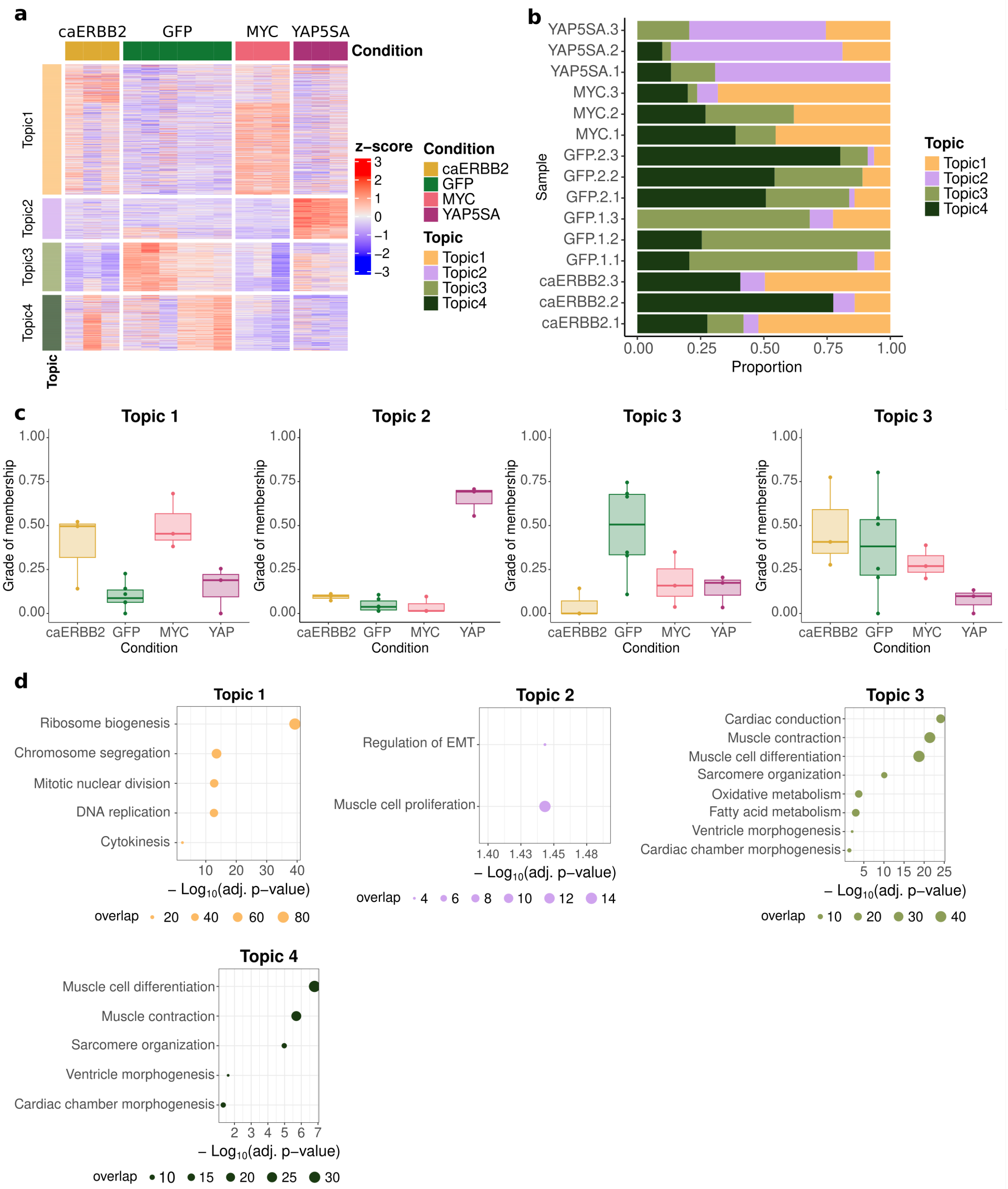


Figure S4.2: GoM models fit using RNA-seq data. a) Heatmap of scaled gene expression. Genes are classified into topics based on their maximum topic $\log_2$FC. Each column indicates a biological replicate, and samples are grouped based on their experimental condition. **b)** Barplot of the degree of membership of each sample in each topic. **c)** Boxplots of the grade of membership of each condition on each topic. **d)** Most significant biological functions found for each topic using an ORA. The size of the dot indicates the number of overlapping genes between the pathway and the topic. In panel **b**, the experimental batch of the GFP samples is indicated in their name. Batch 1 consisted on AAV9-*Yap5sa* and AAV9-*Gfp*.1 treated cultures. Batch 2 included AAV9-*Myc*, AAV9-*caErbb2* and AAV9-*Gfp*.2 treated cultures

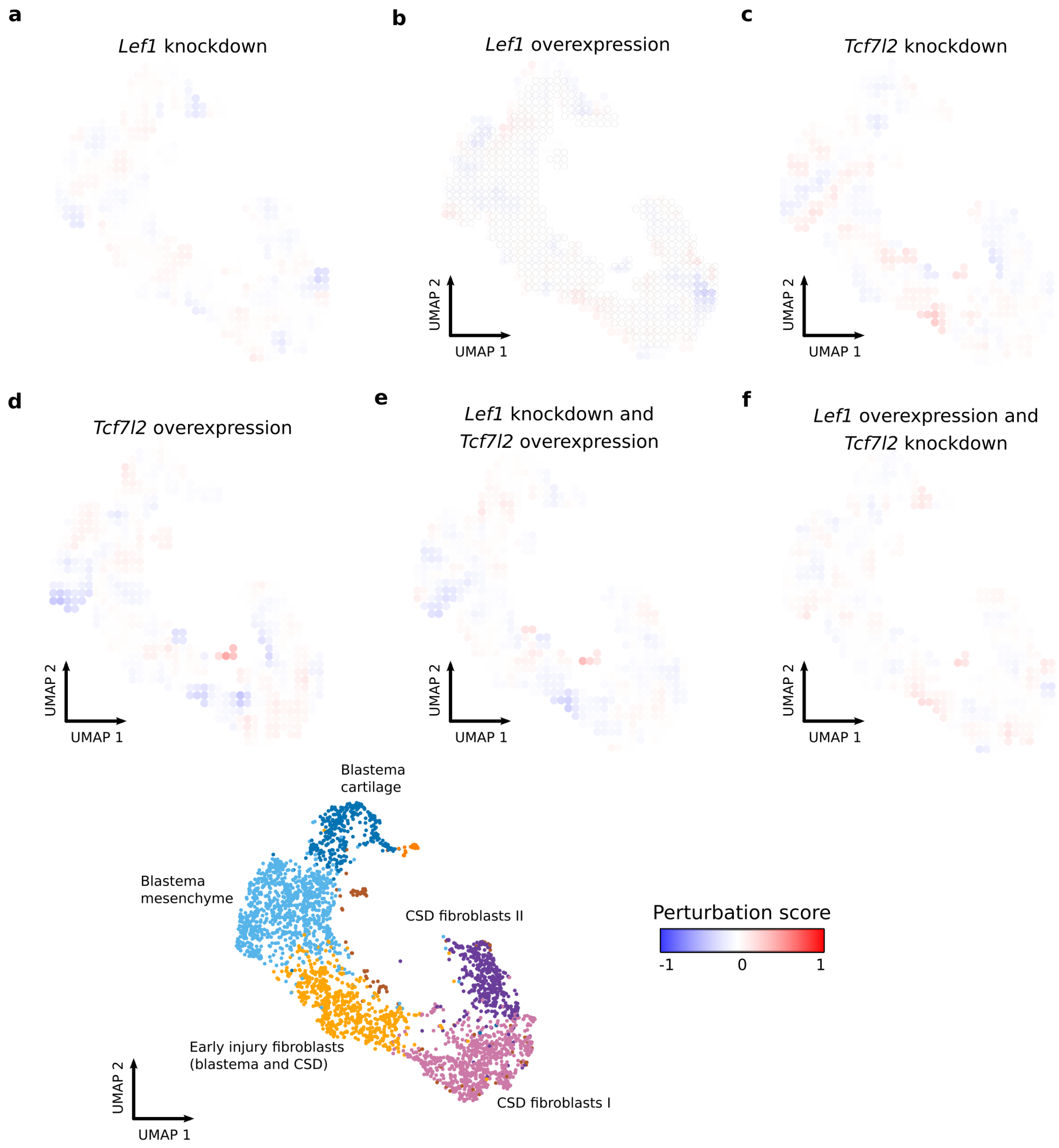


Figure S5.1: Effect of randomized *Lef1* and *Tcf7l2 in silico* perturbations on the blastema and CSD CT trajectories using a randomized version of the GRN. Panels **a-f)** show a gridded UMAP visualization of integrated blastema and CSD CT cells. Colors and vectors represent the effect of **a)** *Lef1* KO, **b)** *Lef1* overexpression, **c)** *Tcf7l2* KO, **d)** *Tcf7l2* overexpression, **e)** simultaneous *Lef1* KO and *Tcf7l2* overexpression and **f)** simultaneous *Lef1* overexpression and *Tcf7l2* KO. Perturbation simulations were calculated on a randomized version of the CT GRN. Red colors indicate that the perturbation promotes the differentiation trajectory of the cells, while blue colors indicate that the perturbation opposes the differentiation trajectory. White colors indicate that the perturbation had no effect. For reference, the UMAP visualization of integrated blastema and CSD CT cells as described in Fig. 5.7a is included.

Table S4.1: Communities in the CM GRN with a significant biological function. Only communities with at least one statistically significant association (FDR-adjusted p-value < 0.05) are included. Hub genes were identified based on their out-degree.

Community	**Biological function**	**Top hubs (out-degree)**
2	Vasculature development	*Ctnnb1, Ets1, Rela*
3	Heart atria and valve development	*Tbx5, Nr2f2, Ankrd1*
4	Response to endogenous stimulus	*Ar, Jun, Egr1*
5	Tube morphogenesis	*Sox9, Tead1, Srf*
6	Mitochondrial protein translation	*Atf4, Tcf3, Kat7*
7	Ion homeostasis	*Mtf1, Nr0b2*
8	Protein translation	*Myc, Nfatc3, Rfx1*
9	Receptor tyrosine kinase (RTK) signaling	*Runx1, Foxo3, Arid5b*
10	Lipid homeostasis	*Pparg, Smad3, Rarb*
11	Cytokine production and response	*Stat1, Irf, Stat2*
12	Immune response	*Sp1, Cebpb, Elf1*
13	Carbohydrate metabolism	*Hif1a, Ppara, Arnt*
14	Detoxification	*Nfe2l2, Tfap2, Kmt2a*
15	Morphogenesis	*Gli1, Hes1, Gli3*
16	Retinoic acid (RA) signaling	*Thra, Rara*
18	Cardiac chamber development	*Rbpj, Ewsr1, Smad4*
20	Response to hormones	*Nr3c1, Ybx1, Ybx3*
21	Lipid metabolism	*Mef2a, Srebf1, Cebpg*
22	Cell cycle	*E2f1, Foxm1, Tfdp1*
27	Hemostasis	*Sp3, Klf2, Pura*
32	Muscle differentiation and sarcomere maturation	*Mef2c, Mecp2, Id2*
33	Tissue morphogenesis	*Pbx1*
34	Heart development and carbohydrate metabolism	*Sp1, Nfatc2, Zbtb7a*
35	Heart tuve patterning	*Nkx2-5, Pitx2*
36	Ubiquinone metabolism	*Yy1, Prdm2*

Table S4.2: TF in the core CM maturation GRN.

TF	Function and relationship with CM biology
Ankrd1	Repressive transcriptional cofactor with a dynamic nucleo-sarcomeric localization involved in titin-based mechanotransduction and regulation of myocardial adaptation and remodeling [268, 269]. It has been associated with multiple cardiomyopathies [270, 271] and has been shown to induce apoptosis through P53 activation [272]. In a cryoinjury context, *Ankrd1* expression becomes elevated in the border zone CM [273].
Atf4	Participates in CM maturation by promoting the switch towards an oxidative metabolism and its expression is induced upon loss of *Hif1a* [274, 275]. It increases fatty acid oxidation and regulates the mitochondrial response to stress [229, 276].
Cebpd	bZIP TF involved in the transcriptional regulation of immune and inflammatory responses [277]. It has been shown to activate the fibrotic mediator *Ctgf* in CM [278].
Dnmt3a	*De novo* DNA methyltransferase that regulates CM gene expression through methylation [279]. It is critical for correct heart contraction and conduction during development [280]. Its expression increases upon pathological stress [281] and decreases in the injured neonatal mouse heart [282].
Egr1	Regulator of inflammation, apoptosis, autophagy, angiogenesis, and cardiac remodeling [283, 284]. I/R generation of ROS-induced oxidative stress induces CM apoptosis through the ROS/JNK/EGR1 signaling pathway [285]. *Egr1* can regulate CM contractility through repression of *Mef2a* [286].
Ets1	Regulates cardiac cell fate determination and is crucial for endocardial development, repressing the CM identity [287]. It becomes activated as a result of the crosstalk between cardiac progenitors and endodermal cells [288]. It competes with *Nkx2-5* to guide *Gata4* to endocardial-specific promoters instead of CM-specific promoters during heart development [289]. Loss of *Ets1* is the cause of congenital heart defects [290].
Fos	Leucine zipper TF that forms the AP-1 TF complex after dimerization with JUN proteins. It has been found to be necessary for CM progenitor differentiation [291] and CM maturation [292]. It activates metalloproteases involved in cardiac remodeling [293] and its expression is increased upon intermittent hypoxia [294]. Fos is an early gene that becomes activated in cardiac hypertrophy [295], and activation of AP-1 is necessary for the induction of the pathological/fetal gene program [296].

Continuation of Table S4.2

TF	**Function and relationship with CM biology**
Foxm1	Expressed by CM during developmental stages and is essential for CM proliferation and myocardial growth [297–299]. While embryonic *Foxm1* deletion is lethal due to heart malformations, ventricular hypoplasia and reduced CM proliferation [297, 298], postnatal deletion of *Foxm1* causes cardiac hypertrophy and fibrosis in aged mice [300]. In zebrafish border-zone CM, *Foxm1* has an increased expression, regulates cell cycle genes and is required for CM proliferation [301].
Gli2	C2H2-type zinc finger protein that mediates hedgehog signaling [302]. *Shh* functions through *Gli2* to specify mesodermal cells into the CM lineage by inducing the expression of *Mef2c*, *Nkx2-5*, and *Tbx5* in cardiac progenitors [303, 304].
Hif1a	bHLH TF considered to be the master transcriptional regulator of the response to hypoxia [305]. Under normoxic conditions, HIF1A protein is degraded via VHL-mediated ubiquitination. During hypoxia, HIF1A degradation is prevented, and the protein accumulates [305]. During heart development, *Hif1a* plays an important role in trabeculation, septation, and metabolic compartmentalization of the embryonic heart, promoting glycolytic and mitochondrial programs in the compact and trabecular myocardium, respectively [305, 306]. During regeneration, functional *Hif1a* is required for CM dedifferentiation and proliferation in the hypoxic zebrafish border-zone [305].
Id3	bHLH TF that lacks the DNA binding region and acts as a transcriptional repressor through heterodimerization with other of bHLH TF [307]. It is involved in myogenesis by preventing CM differentiation and promoting CM proliferation through sequestration of *Gata4* [307]. It is a downstream effector of *Bmpr2* signaling in cardiac mesoderm progenitors, in which regulates CM differentiation [308].
Irf7	Regulates type I interferon response and is induced upon cardiovascular stress [309]. It is expressed by endothelial cells and might regulate the immune response after AMI [310]. Epicardial *Wt1* expression activates *Irf7* to decrease epicardium *Cxcl10* and *Ccl5* levels and promote CM proliferation [311]. *Irf7* is also involved in the regulation of CM hypertrophy, especially in stress conditions such as fibrosis and pressure overload [312, 313].

Continuation of Table S4.2

TF	**Function and relationship with CM biology**
Junb	Member of the JUN protein family that constitutes the AP-1 TF complex together with FOS TFs [314]. Together with *Fosl1*, promotes the expression of the cell cycle regulator *Ccnt1* during zebrafish heart regeneration [235]. Additionally, regenerative zebrafish CM change their chromatin accessibility landscape to activate AP-1 binding sites, a process necessary for CM proliferation and protrusion into the injury area. *Junb* and *Fosl1* are key mediators of this process [235]. In mammals, I/R and chronic β-AR activation induce *Junb* expression, followed by replacement with *Jund* [315, 316]. *Junb* plays a key role in regulating CM hypertrophy, sarcomeric stability and function, and preventing apoptosis [317–319].
Klf4	Zinc-finger TF that, together with *Oct4*, *Sox2* and *Myc*, constitutes the Yamanaka TF cocktail for cell reprogramming [320]. Overexpression of these four TF cause CM dedifferentiation and permit adult mouse heart regeneration [321]. *Klf4* expression alone induces CM dedifferentiation and proliferation [322]. *Klf4* is also a mediator of the anti-hypertrophic transcriptional response and its expression is induced by hypertrophic stimuli [323, 324].
Maf	Leucine zipper TF that regulates eye development [325].
Meox1	Homeobox mesodermal TF involved in somitogenesis and development of the sclerotome. It expressed by muscle cell progenitors [326] and is a central regulator of fibroblast activation upon heart injury [327, 328].
Myc	bHLH TF that acts as a master regulator of cell growth and proliferation [329]. *Myc* can functionally replace *Mycn*, the main *Myc* isoform in CM [85]. The total *Myc* and *Mycn* levels guide a CM competition mechanism during development and adulthood [85, 330]. *Myc* is not normally expressed in the adult heart, but is induced upon hypertrophic stimuli [331]. Overexpression of *Myc* and *Ccnt1* leads to regeneration in the adult mouse heart [86, 87].
Mycn	Essential cell-autonomous factor for CM proliferation and heart development [332]. While *Mycn* is exclusively responsible for developmental CM proliferation, its function can be recovered by *Myc* by elimination of *Mycn*-deficient cells through cell competition [85]. *Mycn* mutations are associated with congenital heart defects [333]. In the context of heart regeneration, *Mycn* has been found to participate in the conserved HH-GLI1-MYCN signaling network [234].

Continuation of Table S4.2

TF	**Function and relationship with CM biology**
Nkx2-5	Homeobox TF essential for heart development and CM lineage commitment and differentiation [334, 335]. Mutations in *Nkx2-5* cause congenital heart defects [336]. Zebrafish with a loss of *Nkx2-5* have impaired regeneration [337].
Npm1	Cytoprotective agent against nucleolar stress, an early event associated with cardiac stress [338].
Nr2f2	Orphan receptor critical for heart development [300, 339]. *Nr2f2* promotes atrial chamber differentiation and maintenance and mutations in *Nr2f2* are associated with human defects in atrioventricular septation [340]. Elevated *Nr2f2* levels lead to dilated cardiomyopathy, reduced mitochondrial function and increased levels of reactive oxygen species [341].
Ppargc1a	Transcriptional regulator of cellular energetics that induces mitochondrial biogenesis([342]. It mediates the metabolic switch from fatty acid oxidation to glucose oxidation seen in cardiac hypertrophy and HFrEF [343].
Rarb	RA plays a crucial role during heart development by establishing anteroposterior polarity, inflow and outflow tract progenitors, and inducing growth of the ventricular compact wall [344]. RA also has been found to be protective against MI [345]
Stat3	TF activated by several cytokines and growth factors. It generally plays a protective role in CMs, activating antioxidative and antiapoptotic gene programs while suppressing antiinflammatory and antifibrotic responses [346, 347]. Additionally, it is a key modulator of the cardiac microenvironment and the communication between the CM and endothelial cells [348]
Tbx5	Key regulator of heart development and CM lineage commitment [349]. *Tbx5* is involved in formation of the septum and specifying and maintaining atrial identity [350]. During development, it is expressed throughout the myocardium and septum, regressing towards the atria in adulthood, and has also been implicated in the development of the conduction system [351]. Loss of function of *Tbx5* has been associated with Holt-Oram syndrome, causing heart abnormalities and atrial fibrillation problems [352].
Tfdp1	Heterodimerizes with E2F proteins to enhance their DNA binding activity. Promotes G1-to-S transition [353].

Continuation of Table S4.2

TF	Function and relationship with CM biology
Thra	Nuclear receptor for thyroid hormone. It regulates the proliferative potential of CMs [354], causing cell cycle arrest and polyploidization [202]. It has been involved as a regulator of cardiac hypertrophy [355, 356] and as a regulator of CM metabolic maturation [357].
Wt1	C2H2-type zinc-finger TF involved in embryogenesis and required for normal heart development [358]. *Wt1* is involved in the development of the epicardium and promotes epicardial-like properties when overexpressed in CM [359]. *Wt1* represses the epicardial expression of chemokines *Cxcl10* and *Ccl5* and might therefore be relevant to positively respond to injury [311]
Ybx3	TF able to regulate the expression of amino acid transporters [224]. While it has not been widely studied in the heart, it is expressed at detectable levels [360]. *Ybx1*, from the same family of TFs, is a known regulator of CM hypertrophy that becomes upregulated upon stress [361, 362].

Table S5.1: Primer sequences used for qPCR detection of *Lef1* and *Tcf7l2* in axolotl CT.

Gene	Forward primer	Reverse primer
Lef1	CGCTCTCAGTCGACAACTCC	TGTTCCTTCCTCTGGGTCGTG
Tcf7l2	TTAGAGGAAGCCGCCAAGAG	ATGACTACCCGGACTGCAACA
Rp4	TGAAGAACTTGAGGGTCATGG	CTTGGCGTCTGCAGATTTTTT