AN ABSTRACT OF THE DISSERTATION OF

Arun Singh for the degree of Doctor of Philosophy in Pharmaceutical Sciences presented on December 5, 2017.

Title: Gene Networks of the Skeletal Muscle Cell

Abstract approved: _____

Chrissa Kioussi Theresa M. Filtz

Skeletal muscle is the largest organ in the body by mass, comprising roughly 40% of total bodyweight in adults. It plays diverse and unique roles that include movement, locomotion, and support for posture and internal organs, among others. The structural foundation for all skeletal muscle in adults is formed early in development, emphasizing the importance of understanding the mechanisms of skeletal muscle development. This is especially important since adult skeletal muscle is limited in its ability to regenerate, but the regeneration mechanism reactivates certain developmental pathways.

Skeletal muscle formation in the vertebrate forelimb occurs in distinct phases during embryogenesis. Beginning around embryonic day (E) 10.5 in mice, embryonic myogenic progenitor cells (EMPCs) express the gene *Pax3*, which triggers migration into the limb bud. Once settled, between E10.5 and E12.5, embryonic myoblasts fuse with each other to form embryonic myotubes. Between E12.5 and E17.5 fetal myoblasts fuse with both embryonic myotubes, and each other, to form fetal myofibers, which serve as the structural foundation of all skeletal muscle in the forelimb. Not much is known regarding the molecular mechanisms behind this process, except that they significantly overlap with the mechanisms responsible for skeletal muscle regeneration in adults. Knowledge gained about myogenesis can also be applied to muscle regeneration in adults, to both accelerate wound healing, or reverse muscle-wasting diseases, called myopathies. Two sequence specific transcription factors, *Pitx2* and *Pax3, are* fundamental to skeletal muscle development. Mice carrying locus specific alterations for both genes are used to molecularly dissect the skeletal muscle formation in time and space.

Pitx2 is required for the embryonic to fetal transition of skeletal muscle formation. ChIP-Seq (Chromatin-Immunoprecipitation followed by sequencing) approach was used to compare the chromatin state of E12.5 embryonic myoblasts from mice in which *Pitx2* was present or had been deleted. ChIP-seq data were integrated with previous gene expression profiling data of the forelimb transcriptome to identify changes in the chromatin state of embryonic myoblasts at Pitx2-target genes. We observed significant disruption in the chromatin state of genes related to neurogenesis and cytoskeletal organization, implying *Pitx2* regulates the cytoskeletal rearrangements during myogenesis.

Pax3 marks all skeletal muscle myoblasts as they migrate into the forelimb, beginning at E9.5 in the mouse. Whole-transcriptome profiling of pure forelimb isolated myoblasts was performed, via fluorescence activated cell sorting (FACS), from $Pax3^{Cre}|Rosa^{EGFP}$ mice. Myoblasts were isolated at 4 embryonic states (E11.5, E12.5, E13.5, E14.5), bracketing the embryonic to fetal transition during myogenesis. The increased expression of genes involved in cell-adhesion, angiogenesis, and immune system during fetal myogenesis, implying there is communication between different organ systems even when limited to what was thought to be a myogenic lineage. Additionally, coexpression network analysis revealed two distinct subnetworks present during all stages of myogenesis, but both expressed highest during fetal myogenesis. One network was enriched in genes that are involved in cell-adhesion, and the second was enriched in genes involving the immune-response, suggesting consistent interplay between the immune system and skeletal muscle. Our studies emphasize the complexity of myogenesis, with multiple different systems developing and communicating in parallel, and will serve as a base for future studies to explore the effect of specific perturbations during forelimb myogenesis. These perturbations will result in knowledge and techniques that can be used to enhance or reactivate skeletal muscle regeneration in mature muscle.

Gene Networks of the Skeletal Muscle Cell

by
Arun Singh

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented December 5, 2017
Commencement June 2018

Doctor of Philosophy dissertation of Arun Singh presented on December 5, 2017

APPROVED:

Major Professor, representing Pharmaceutical Sciences

Dean of the College of Pharmacy

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Arun Singh, Author

ACKNOWLEDGEMENTS

Before anything else I want to thank my co-mentors, Chrissa Kioussi and Theresa Filtz, for all the help and guidance they have provided to me during this program. They have been incredibly nurturing and patient with me, which allowed me to grow and become productive during these past four years. They helped me to become confident in my abilities as a scientist, and I am humbly grateful for the opportunities they gave me.

I also want to thank my girlfriend, Brenda, for all the support she has given me during my time in Corvallis. Ever she joined me in Corvallis two years ago, everything has been much easier to manage. She has supported me through difficulties in the program, and celebrated my accomplishments with me. I was and am lucky to have such an amazing partner by my side.

And I want to thank my committee, Drs. Stephen Ramsey, Barbara Taylor, and Andriy Morgun, for their support and feedback during this program. I especially want to thank Dr. Stephen Ramsey, for his thoughtful feedback and collaboration during this project, especially with regard to the computational aspects. It was tremendously helpful, and helped me to understand the link between biology and the computational aspects.

Last, I want to extend thanks to everyone else who was directly and indirectly involved in this project. A special thanks to my lab mate Vera, for her help with animal work, and with all aspects of my projects. I am happy to have worked with her as a colleague. To the College of Pharmacy, for funding and support. To Gary Miller for IT assistance. To the CGRB core infrastructure for computational materials. To the University of Oregon, and Maggie Weitzman, for use of their cell sorter, library preparation, and RNA-sequencing. And finally, to my fellow graduate students and friends in Corvallis, for their support throughout the program.

CONTRIBUTION OF AUTHORS

Chrissa Kioussi designed research, contributed reagents, contributed to data analysis, writing, and editing; Theresa Filtz contributed to data analysis, writing, and editing; Arun Singh designed research, performed research, data analysis, and writing. Michael Gross contributed reagents and data analytic tools. Stephen Ramsey contributed to data analysis and writing. Vera Chang and Hsiao-Yen Ma contributed to data.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

Page

LIST OF FIGURES

LIST OF TABLES

Differential Gene Regulatory Networks in Development and Disease

Chapter 1

Arun J. Singh, Stephen A. Ramsey, Theresa M. Filtz, Chrissa Kioussi

Cellular and Molecular Life Sciences
Published October 10, 2017
https://doi.org/10.1007/s00018-017-2679-6

Abstract

Gene regulatory networks, in which differential expression of regulator genes induce differential expression of their target genes, underlie diverse biological processes such as embryonic development, organ formation and disease pathogenesis. An archetypical systems biology approach to mapping these networks involves the combined application of (i) high-throughput sequencing-based transcriptome profiling (RNA-seq) of biopsies under diverse network perturbations and (ii) network inference based on gene-gene expression correlation analysis. The comparative analysis of such correlation networks across cell types or states, *differential correlation network analysis*, can identify specific molecular signatures and functional modules that underlie the state transition or have context-specific function. Here, we review the basic concepts of network biology and correlation network inference, and the prevailing methods for differential analysis of correlation networks. We discuss applications of gene expression network analysis in the context of embryonic development, cancer, and congenital diseases.

Introduction

One of the most fundamental scientific advancements of the early 21st century has been the sequencing of the human genome (Lander et al., 2001). Knowledge of the sequence and location of genes in the genome has revolutionized biomedical research and diagnostics. Genetic mutations responsible for congenital diseases can now be physically linked to specific positions in the genome, enabling the systematic mapping of the molecular basis of polygenic diseases and other traits. To fully understand the mechanistic basis of a polygenic trait or biological process, it is necessary to account for interactions such as gene expression correlation or genetic epistasis among the individual trait-associated genes. Gene networks are an intuitive and useful abstraction for representing the totality of such gene-gene interactions for a trait.

The advent of relatively low-cost quantitative transcriptome profiling (first by microarray hybridization and more recently, by high-throughput sequencing or RNA-seq (Lister et al., 2008; Mortazavi et al., 2008)) was key to enabling the systematic mapping of networks of genes whose expression levels are correlated across biological samples, i.e., gene expression correlation networks. Because eukaryotic gene regulation is thought to be hierarchically organized with regulated genes' protein products in turn regulating the expression of other genes, gene-gene correlation networks provided the first genome-scale view into gene regulation in diverse contexts such as embryogenesis, hematopoiesis, oncogenesis, and inflammation.

Although gene-gene interactions were the earliest biological network type to be analyzed on the whole-organism scale (Lee et al., 2002), network analysis has since found broad utility in understanding interactions among cellular molecular constituents of all types, with protein-metabolite networks and protein-protein interaction (PPI) networks being particularly widely used. The network abstraction has proved to be applicable across the scale of biological complexity; for example, on the organismal level, where each organ is a separate module that interacts with other organ(s) to form a functional organism, or organelles functioning together within a cell. The discovery that the structure of large-scale molecular interaction networks is functionally related to the networks' emergent properties such as robustness, which cannot be easily predicted based on knowledge of the functions of the network's individual components in isolation, led to the development of systems biology, a holistic and quantitative approach to biology. The heart of systems biology is the use-and refinement based on experimental challenge-of quantitative models that are grounded in knowledge of molecular interactions.

Another advantage of grappling with biological complexity from a network perspective includes

the condensation of a large amount of data into a simpler, visually intuitive format. For example, in a network, evolutionarily recurrent structural elements such as feedback loops (Alon et al., 1999), feed-forward loops (Bornholdt, 2005; Ideker et al., 2001; Kirschner, 2005; Kitano, 2002), and asymmetric positive feedback loops (Ratushny et al., 2012)are evident (Milo et al., 2002; Prill et al., 2005). On the larger scale, analysis of biological networks has provided insights into critical genes or molecules that are essential for function (Jeong et al., 2001) and evidence for selection for hierarchical modularity (Ravasz et al., 2002), and formed the basis for dynamical models of system function (Li et al., 2004). Comparative analysis of biological networks derived from different biological states (such as different environmental conditions, genetic backgrounds, or stages of development), i.e., differential network analysis, has proved particularly useful for uncovering mechanisms in diverse contexts such as cancer biology (Creixell et al., 2015; Grechkin et al., 2016) and organ development (Földy et al., 2016; Reyes-Bermudez et al., 2016).

Here, we review differential network analysis, with a particular emphasis on its application to gene expression correlation networks. In this context, differential correlation expression network analysis (DCENA) has been used to identify a three-gene combination able to diagnose glioma (Wu et al., 2016), and identify gene regulatory networks involved in mouse embryonic fibroblast development (Treutlein et al., 2016), among others.

Basic properties of biological networks

It is axiomatic that a complete understanding of a cell biological system requires a complete understanding of the molecular networks within a cell. Such networks can be modeled as a series of points, called nodes (which might be gene-specific types of RNA or protein, or other cellular molecular constituents), connected via edges (representing interactions) to each other. Each edge can be either directed or non-directed, meaning the biological information (or in some network contexts, substrate) flows from one node to the other. Directed networks imply co-dependence between connected nodes, whereas undirected networks assume that only one node is dependent on the other.

The discovery that biological networks followed a scale-free degree distribution invigorated the field of biological network analysis. A scale free degree distribution means that the probability that any node in the network, P(k), is attached to k other nodes, decays as a power law $P(k) \sim k^{-\lambda}$ (Barabási and Albert, 1999). Interestingly, this property is shared by many different types of

networks that are the product of organic growth rather than up-front complete design, including paper citations, social networks, and the internet (Barabási and Albert, 1999). Following the scale free property, most nodes have only a few edges while a small number of nodes, called hubs, are highly connected. These hubs are disassortative (i.e., they tend to not share the same neighbors), and are rarely directly connected to each other (Barabási and Albert, 1999). These features appear to endow biological networks with a robustness to disruptions of random nodes (Albert et al., 2000). On the other hand, they are more susceptible (compared to random networks that are not scale-free) to disruptions that specifically target hubs (Albert et al., 2000). The observed scale-free degree distribution of many biological networks appears to be a generic consequence of two assumptions regarding the growth of networks over evolutionary time, random growth with preferential attachment (Barabási and Albert, 1999).

One of the most powerful network analysis paradigms stems from the guilt by association principle, where nodes that are spatially close to each other often perform similar functions, or are related in some way (Hou et al., 2014). A node that is in close network proximity to another node that is directly associated with a disease is likely to be implicated in either the creation or maintenance of the disease state (Hou et al., 2014). Spatially proximate nodes also tend to cluster into highly connected clusters, where each cluster performs a specific function (Barabasi and Oltvai, 2004). These clusters, termed modules (Hartwell et al., 1999), are fundamental units of biological networks (Figure 1.1). Modules can identify novel functions and/or pathway associations of known genes and are often conserved between different biological systems (Mitra et al., 2013). Conserved modules are likely to represent biologically significant processes conserved by evolution and their identification can help answer fundamental questions of biological regulation (Mitra et al., 2013).

Module identification is one of the classic analysis tasks in network biology, in which genes of unknown function can be associated with the function of the module, assuming enough annotated genes are also present. At best, modules can identify novel gene-gene interactions that function as drug targets. As condition-specific global measurements of protein post-translational modifications, gene-level epigenomic profiling, and gene expression have become commonplace, particular emphasis has been placed on the identification of *active modules*, which are modules with a context-dependent function (Mitra et al., 2013). Unfortunately, there is no clearly defined optimal method for module identification (although MCODE (Bader and Hogue, 2003) is one of the most popular, likely due to its ease of use) and instead a variety heuristic approaches are used.

Here, we describe three standard classes of methods used to identify active modules. One such class of method is called *significant-area-search* (Mitra et al., 2013). Significant-area-search methods first annotate all edges and nodes with scores that represent their independent activity. Then aggregate scores are computed for different combinations of spatially close and connected nodes and assigned to each subnetwork. The highest scoring subnetworks are then identified as modules. While this method works, it has significant drawbacks. The algorithm is complex, the computations are time intensive, and it requires multiple user-defined thresholds that can complicate reproducibility and are a challenge for non-expert users. A second class of methods is known as *diffusion flow* (Enright et al., 2002). Diffusion flow-based methods are based on the concepts of fluid and heat transfer dynamics. On a basic level diffusion flow methods assume that the flow of biological information within a network diffuses from nodes of interest outwards along the edges of a network and accumulates in other nodes of interest. These methods are less computationally intensive, but are most often used in cancers or diseases where original genes of interest are known. The third type of methods used are *biclustering-based* methods (Reiss et al., 2006). Biclustering-based methods simultaneously cluster both interactions in the network and the biological conditions in which the interactions occur, called biclustering. This determines both the strength of network connectivity and the correlation between sample conditions. Biclustering based methods have the advantage that they can be applied to almost any type of biological data.

Despite the existence of several well-established algorithmic methods, in practice, module detection retains subjective characteristics. This is especially true for clustering-based methods, where some hard value threshold is needed to define module boundaries. Functional classification or analysis of modules is often accomplished using gene ontology (GO) term enrichment techniques. GO term enrichment analysis involves taking a list of genes (often from the identified module) and using publicly available annotated functions to search for enrichment of specific functions in the gene list. The premise of GO term enrichment analysis for module classification is essentially guilt-by-association, i.e., that if a module is enriched in genes from a specific biological function, any unknown genes also in the module are likely to be involved in the same process.

Despite its promise for elucidating biological mechanisms, large-scale biological network analysis has several inherent limitations. The most obvious is that it necessarily entails a dramatic simplification of biological systems, for example, ignoring constraints, kinetic rate constants, and context-dependent subcellular localization (Barabasi and Oltvai, 2004; Ideker and Krogan, 2012;

Mitra et al., 2013). Another key limitation is presented by the dynamic nature of living systems; accounting for dynamics in network modeling is more complex and can involve generating multiple static networks of the same biological system over different conditions. These considerations have led to the use of *differential network* techniques, as described below.

Differential networks

Differential network analysis is a powerful technique for identifying pathways or modules with a context- or condition-specific activity or function, for example, in diseased vs. healthy tissue, different time points in embryonic development, or in tissues with and without a specific molecular or genetic perturbation (e.g., a gene knockout). Differential networks are generated by comparing static networks from the same biological system over different conditions and are used to determine the parts of a network that are context-specific (Ideker and Krogan, 2012; Mitra et al., 2013) (Figure 1.2). Differential networks encode the changes in connections among nodes between the conditions or states. Nodes that change their connection between different network states are considered to be rewired, and are of significant biological interest (Hou et al., 2014). Under the principle that nodes that are functionally involved in the change of state rewire more frequently than uninvolved nodes, "guilt by rewiring" can be used to identify nodes or modules responsible for creating and/or maintaining different states (Hou et al., 2014).

For both gene co-expression networks and condition-specific PPI networks, differential network analysis serves as a powerful complement to other forms of quantitative expression analysis, specifically differential expression analysis (Ideker and Krogan, 2012). Differential expression analysis detects nodes that change between conditions, but does not reveal interactions between the nodes. Therefore, nodes with interaction but not expression changes will only be revealed in a differential network analysis (Ideker and Krogan, 2012). Similarly, node interactions that are present in both conditions can be assumed to be unaffected by the perturbation, even if the node is differentially expressed between conditions (Ideker and Krogan, 2012). Thus, differential expression analysis has the potential to be more sensitive and specific than standard node-based differential expression analysis for detecting context-specific molecular interactions.

Rewiring has different biological meanings depending on the components of the network. In PPI networks, rewiring represents the gain and loss of direct physical interactions between the proteins, which could occur due to changes in the protein post-translational modification state or

sub-cellular localization. In transcriptional networks, rewiring between nodes represents functional consequences of the change in cell or tissue state between conditions (Ideker and Krogan, 2012). These functional consequences could be changes in the co-expression patterns of genes caused by disruption of the regulator that mediates the genes' co-expression. Due to the ubiquity and archetypical role of transcriptome profiling as a systems approach, for the remainder of this review, we will discuss differential analysis of transcriptional co-expression networks.

Differential co-expression networks (DCEN)

Differential co-expression networks are correlation networks based on gene expression data such that each node represents a gene, and each edge represents co-expression or interactions between the genes, such as activation or repression (Dong et al., 2015). The premise of differential co-expression network analysis is that differentially expressed genes (DEGs) represent functional changes in the biology of the system, and functionally work together at some level to alter the system (Dong et al., 2015). DCENs have been shown to have (i) a scale-free degree distribution, (ii) low clustering coefficients, and (iii) an unexpectedly high average all-pairs-shortest-paths path length (Hsiao et al., 2016).

The power of the DCEN analysis approach is its ability to highlight genes that drive molecular mechanisms behind specific biological processes (Dong et al., 2015). Additionally, DCENs have been used to identify the potential mechanisms involved in biological pathways, the pathways that are functionally involved in the system being studied, and the key regulators of identified pathways or modules (Dong et al., 2015). These mechanisms and regulators are often identified as new biomarkers for specific processes or diseases, leading to more accurate diagnosis. Most software that is mentioned in this review is freely available and can be found online (Table 1.1).

An early example of differential correlation expression analysis was an applied form of hierarchical clustering analysis of time-course gene expression data from yeast *Saccharomyces cerevisiae* under a variety of culture perturbations and from serum-stimulated human fibroblasts (Eisen et al., 1998). This type of analysis clusters genes based on similar coexpression between the samples in an unsupervised manner. A weakness of this type of analysis is the poorly defined modules that it generates. Although clear patterns emerge, there is no hard threshold, and the identification of modules between the samples is highly sensitive to and based on the user-specified cutoff. A more significant weakness of this algorithm is that it assigns each gene to one

and only one module. This singular assignment is a poor representation of biological systems wherein genes frequently participate in multiple processes or occasionally act alone. This method is still prevalent with heat maps but is now more often combined with other methods when used for module identification.

The current, predominantly used methods for co-expression network inference fall into two broad categories, correlation-based and mutual information-based methods. The correlation-based methods for co-expression network inference involve the filtering of all-pairs Pearson correlation coefficients (PCC), nonparametric Spearman (SCC) correlation coefficients, or partial correlation coefficients to obtain the network adjacency matrix. The PCC and SCC are direct statistical measures of interdependence of two variables, whereas partial correlation coefficients are measures of interdependence of two variables' residual measurements after conditioning on one or more other explanatory variables (Butte and Kohane, 2000; Dong et al., 2015; de la Fuente et al., 2004). The related method of weighted gene co-expression network analysis (WGCNA) is based on correlation coefficients, but includes an additional rescaling before the filtering step (Langfelder and Horvath, 2008). Mutual information (MI)-based methods depend on higher-order moments of the joint pairwise distribution of gene expression measurements (rather than just the covariance), through the use of an information-theoretic measure of how informative one gene's expression level is for predicting the expression level of another gene (Basso et al., 2005; Margolin et al., 2006). MI-based methods have the benefit of being sensitive to potentially non-monotonic relationships between the expression level of a regulator gene and a target gene, but the disadvantage of being computationally expensive to estimate in comparison to purely correlation-based methods.

Correlation coefficient methods

Correlation coefficient (CC) based methods involve the use of one of several statistical correlation measures to determine the dependence of one variable on another. The PCC and SCC are both pairwise correlation coefficients, but they differ in that the PCC is most sensitive to a linear relationship between variables while the SCC allows for a non-linear relationship. This difference stems from the fact that the SCC is computed using the ranked values for each variable, rather than the gene expression measurement values themselves. Relevant to DCENs, the correlation coefficient represents a proxy for the extent of co-regulation between two variables, or genes (Figure 1.3).

Because CC-based methods involve the calculation of correlation coefficients for all pairs of genes, naïve application of such methods to all 20,000 genes in the human genome would require the calculation 2×10^8 correlation coefficients, which is both time- and memory- intensive. In practice, pairwise correlations are calculated for a subset of genes that are selected for probable relevance to the biological process being studied, such as selecting for differentially expressed genes, or known genes of interest. This reduces computational time and memory but at least for the "known genes" approach, it can also miss unexpected correlations. Due to their simplicity, CC-based methods have been a popular tool to analyze DCENs. One significant advantage of CC-based methods is that they can be used with small sample sizes (n = 16), although increasing n adds more statistical power and will uncover weaker correlations with higher confidence.

Frequently, CC-based methods are used to identify new putative biomarkers for different types of diseases. Recently, PCC-based methods were used to identify differentially correlated gene pairs in the prefrontal cortex between healthy patients and patients with Huntington's disease (Guitart et al., 2016). Focusing on the gene *ENT1* from a previous study, targeted analysis revealed that *ENT1* gains over 60 correlations in samples from patients with Huntington's disease, identifying *ENT1* as a potential biomarker and drug target for neurodegeneration in the early stages of Huntington's disease. Similarly, PCC-based methods were used to compare differentially correlated miRNA pairs from the plasma of healthy subjects and patients suffering from mild cognitive impairment (Kayano et al., 2016). The network of differentially correlated miRNA pairs was analyzed for known transcriptional regulatory interactions (based on the Ingenuity Pathways Analysis commercial database of gene regulatory interactions), suggesting that TP53 (p53) directly regulates all 11 correlated miRNAs in healthy patients. Additionally, in patients with mild cognitive impairment, gene expression changes of insulin-related genes (identified using Ingenuity Pathway Analysis) were associated with a loss in miRNA correlations, and could also serve as a new biomarker in early Huntington's disease.

CC-based methods have also been used in conjunction with gene ontology (GO)-based functional enrichment analysis. Southworth et al. used SCC-based methods to compare blood from old and young mice to examine the effect of aging on co-regulation of genes (Southworth et al., 2009). They found an overall decrease in gene co-expression in older mice, consistent with previous reports of transcriptional instability in older mice. When GO term enrichment was observed in differentially correlated modules, a stronger correlation existed among NFκβ direct target genes in young mice relative to old. It was concluded that old age may affect the stability of NFκβ expression and therefore affect its downstream targets. Similarly, proximal clustering on

chromosomes of enriched genes in the differentially correlated modules was observed, implying that chromosomal degradation could be a contributing factor to gene expression variance.

In the last decade, several CC-based methods (including methods for downstream analysis and module detection) have been implemented as documented software packages for use in the R statistical computing environment (Ihaka and Gentleman, 1996), whose free-software open-source approach has spurred methods development and resource sharing in bioinformatics. One of the first of such software packages is CoXpress (Watson, 2006) combines PCC-based methods along with hierarchical clustering to identify differentially correlated modules between samples or conditions. Because CoXpress uses hierarchical clustering, it identifies modules rather than differentially correlated genes. A second limitation of CoXpress is that each gene is only assigned to one module and it requires a hard-arbitrary cutoff for the clustering analysis, which in turn decreases reproducibility.

More recently, the DiffCorr (Fukushima, 2013) R package has been developed and released for DCEN analysis. It combines PCC-based methods with the Fisher Z-transformation to identify genes that are differentially correlated between conditions. DiffCorr is able to detect changes in sign or magnitude of the correlation coefficient for a gene pair between different conditions, and it can be applied to any form of normalized data, including metabolomics and lipidomics.

DiffCorr has been used to identify differentially correlated genes and modules between healthy and pathogen-infected Arabidopsis samples (Jiang et al., 2016). DiffCorr analysis uncovered (i) the gene *AT3G03440*, which previously had no known role in pathogen response, as a hub in the correlation network from pathogen-infected samples; and (ii) 36 pathways whose gene-gene correlations were differentially rewired. Additionally, no significant correlation was detected between differentially expressed genes and differentially correlated genes, and only 40% of sequence-specific transcription factors that were differentially correlated were differentially expressed. These findings support the viewpoint that DCEN analysis supplements–rather than replaces–traditional one-gene-at-a-time differential expression analysis.

A similar R package, R/Ebcoexpress, combines a CC-based approach with posterior probability calculations and a false discovery rate (FDR) cutoff (Dawson et al., 2012). R/Ebcoexpress has been used to detect differentially correlated genes and modules between human rectal adenocarcinoma and healthy tissue (Zuo et al., 2016), with the network analysis platform Cytoscape (Shannon et al., 2003) used for module detection. The network was constructed using nodes and edges from only differentially expressed genes calculated by edgeR (Robinson et al.,

2010) and only edges with a CC of at least 0.6 were retained. Six hub genes (defined by 8 or more edges) were identified in the network with most unsurprisingly involved in cell adhesion. MCODE identified three modules in the network enriched in cell adhesion-related GO terms. When the network was combined with drug-to-transcriptome-response data from the Connectivity Map database (Ecker et al., 2012), several small molecules including the histone deacetylase inhibitor scriptaid and antiallergenic drug spaglumic acid were predicted to interact with the network. Thus the analysis identified potential biomarkers and new therapeutic approaches for rectal adenocarcinoma.

Another differential correlation analysis R package, the Discordant method, uses PCC combined with binning to identify differentially correlated genes and modules (Siska et al., 2015). Using simulated data sets, the Discordant method outperformed R/Ebcoexpress in computational time and accuracy. When the Discordant method was applied to publicly available glioblastoma data sets, the hsa-mir-545 was identified as a candidate functional regulator in glioblastoma. The Discordant method assumes independence of the bivariate expression levels of gene pairs, which is biologically implausible but reduced the computational time required.

One of the more comprehensive R packages released recently, differential gene correlation analysis (DGCA), uses CCs transformed into normalized *Z*-scores to identify differentially correlated genes and modules while performing downstream analysis, including data visualization, GO enrichment, and network construction tools (McKenzie et al., 2016). This method is similar to Discordant and DiffCorr while also including an FDR cutoff to control the type I error rate. Module identification is performed using the network alignment algorithm MAGNA (Saraph and Milenković, 2014). DGCA was found to outperform two other methods (DICER (Amar et al., 2013) and DiffCoEx (Tesson et al., 2010)) in speed and accuracy on simulated data sets. Furthermore, for larger numbers of samples ($n > 50$), DGCA outperformed R/Ebcoexpress and Discordant.

A primary limitation of the DCEN analysis approach is the computational time and statistical power that it requires to apply to all genes in a higher eukaryote (~25,000 protein-coding genes). This limitation can be mitigated by performing DCEN on a collection of subsets of genes selected based on prior biological knowledge. These gene sets are essentially user-defined modules, and thus, the reduced computation time comes at the cost of introducing bias. One example of a tool using such an approach is DIfferential Network Analysis (DINA), which identifies differentially correlated gene sets between conditions (Gambardella et al., 2013). Publicly available data

consisting of sequence specific transcription factor (SSTF)-pathway associations were combined with datasets comprising thousands of microarrays from different tissue types to identify coregulated pathways between tissue types in both mouse and human. The DINA analysis identified 22 pathways that were coregulated between the tissue types, including pathways that were not differentially expressed between tissues. DINA identified the pathways as being regulated by nuclear receptor family transcription factors and agrees with the literature. Interestingly, YEATS2, a little-studied protein, was predicted to be a negative regulator of metabolic pathways in hepatocytes, and subsequently validated through targeted experiments.

Another challenge of DCEN analysis is identifying modules whose gene members have differential coexpression across more than two conditions. Some packages, such as DINA, are able to identify correlated modules, but require a gene list as input. A recent algorithm called Inference of Multiple Differential Modules (iMDM) was specifically created to work with more than two conditions (Ma et al., 2015). iMDM uses PCC-based methods to identify both unique and differentially correlated modules between multiple conditions. The algorithm was validated using RNA-seq transcriptome profiling data from mouse hearts in four different stages of hypertrophy, and identified multiple modules that are both unique and differentially correlated between conditions.

Weighted gene co-expression network analysis (WGCNA)

WGCNA, introduced above, is one of the most popular methods for CC-based DCEN analysis, owing in part to its availability as a point-and-click software application and an R software package (Langfelder and Horvath, 2008). A key benefit of WGCNA is that it identifies modules based on topological overlap and hierarchical clustering, and as such, does not require a hard threshold on the adjacency matrix for module detection (Langfelder and Horvath, 2008). While interpretation of the results of hierarchical clustering can be subjective, identified modules can be tested with GO term enrichment to verify whether the genes share a common function. WGCNA identifies differentially correlated genes and modules and performs subsequent downstream analysis including data visualization and GO term enrichment via other packages available in R.

Use of WGCNA has led to the implication of a variety of genes and modules in specific biological contexts. For example, WGCNA analysis of T cells implicated the genes *GABARAP* and *MPEG1* in asthma (Troy et al., 2016). Combined with upstream regulator analysis, the cytokine genes *IL2*

and *IL4* were identified as drivers of the asthma response. WGCNA was also used to investigate gene regulation during the various stages of coral development (Reyes-Bermudez et al., 2016). Early developmental stage-specific modules were detected that are highly enriched in long non-coding RNA (lncRNA), suggesting a lncRNA regulatory role in coral gastrulation. When WGCNA was applied to human brain development, significant rewiring of regulatory modules was detected between pre- and postnatal brain development [60], consistent with a model that brain development is organized into multiple stage-specific regulatory networks with little interstage overlap. Gene promoter sequence analysis identified specific SSTF binding sites uniquely enriched in each module, suggesting module-specific regulation. Finally, WGCNA analysis of skin transcriptomes from patients in a psoriasis study that were healthy or treated with a TNF-α inhibitor implicated a lncRNA-rich module via the guilt-by-rewiring principle (Ahn et al., 2016).

WGCNA has also been used to analyze transcriptome datasets spanning multiple tissue types, for example in a study of Huntington's disease (Scarpa et al., 2016) in which both human and mouse datasets were used. The analysis yielded multiple insights including the identification of an astrocyte-specific module and the identification of the transcription FOXO3 as a candidate regulator of the module.

Modified weighted gene correlation network analysis (mWGCNA)

Several network analysis algorithms in use are variants of WGCNA, of which the most popular and best known is DiffCoEx (Tesson et al., 2010). DiffCoEx follows the WGCNA method initially, but obtains the adjacency matrix as the difference of correlation matrices between the conditions. By design, DiffCoEx detects differentially correlated modules, not gene pairs. DiffCoEx's popularity is likely due in part to its inclusion of GO enrichment analysis functionality and the small number of tunable parameters required for its operation.

Klein et al. (Oros Klein et al., 2016) combined WGCNA with a topological overlap matrix (TOM) by constructing the TOM of the matrix of gene-gene correlation coefficients and then testing for differences in the TOM between sample groups. This approach has high sensitivity for identifying differentially correlated gene pairs and modules, but requires user-defined parameters, some of which the authors report must be experimentally determined. In their application of the method to a transcriptome study of ovarian, lung, breast, and skin cancer samples with different *TP53* (p53) mutations, Klein et al. detected differential correlation of *KIR3DL2,* identifying a novel gene in the

p53 pathway in cancer.

In their Differential Correlation Regulatory Analysis (DGCA) approach, Zuo et al. (Zuo et al., 2016) combined WGCNA with genome location data (SSTF → target gene) in order to detect differential activity of transcriptional regulatory mechanisms. DGCA application to glioma transcriptomes identified three SSTFs, ZNF423, AHR, and NFIL3, with differential regulatory signal across tumor grades, providing three potential new prognostic indicators.

Mutual information (MI)-based methods

A third form of DCEN analysis, used mainly to infer transcriptional regulatory relationships, utilizes mutual information (MI)-based methods. MI is a measure of dependence between two random variables, and like the correlation coefficient, is generally applicable to any type of multivariate biological measurement. As with correlation-based networks, MI is sensitive to indirect regulatory relationships, introducing spurious edges into the regulatory network. Similar to the method of partial correlation coefficients, the use of conditional mutual information (CMI) [64] mitigates this problem by using the mutual information between two variables conditioned on the value of a third variable, while also being able to detect direct or indirect relationships.

Among the widely-used CMI-based methods is Modulator Inference by Network Dynamics (MINDy) (Wang et al., 2009), which uses gene expression data from different conditions to identify SSTF-target interactions that change based on the expression of a third gene, the modulator. Wang et al. (Wang et al., 2009) applied MINDy to transcriptome data from B-cell lymphoma, identifying new modulators of the MYC protein. Like all MI-based methods, MINDy has the advantage of being sensitive to non-monotonic regulatory relationships. However, MINDy also has some limitations, including that it requires data from a large number of samples and is unable to detect if a SSTF switches from an activator to a repressor (or vice-versa). The R software package Driver-gene Inference by Genetic-Genomic Information Theory (DIGGIT) applies the MINDy algorithm to discover master regulator genes responsible for cell states or phenotypes (Alvarez et al., 2015). Since DIGGIT factors copy number variation in its calculations, it is best suited for cancer-related datasets. Another CMI-based algorithm is Conditional Inference of Network Dynamics (CINDy) (Giorgi et al., 2014). Similar to MINDy, CINDy calculates the CMI between SSTFs and their targets based on the expression of a signaling protein. In a side-by-side comparison with MINDy using microarray data, RNA-seq data, and a validated PPI network,

CINDy outperformed MINDy in terms of both recall and precision and it required half the memory (but double the computational time) that MINDy required. Application of CINDy to B cell lymphoma transcriptomes revealed a CDK2-HMG1 regulatory interaction. The same CMI-intrinsic limitations, described above for MINDy, also apply to CINDy. Another method, CMI2NI (Zhang et al., 2015) makes use of a path consistency algorithm which gradually removes redundant edges in a network based on conditional dependences. By design, CMI2NI infers an undirected network of probable regulatory interactions, in contrast to MINDy and CINDy, which infer directed regulatory networks.

One weakness of SSTF-gene CMI-based approaches (e.g., MINDy and CINDy) is that they require SSTFs and their target genes to be co-expressed across the full set of gene expression profiles, an assumption that does not hold for many SSTFs in many tissues and in various model species (Beer and Tavazoie, 2004). The algorithm Differential Multi Information (DMI) somewhat relaxes this biologically implausible requirement by assuming that the target genes of a TF are co-expressed if and only if the SSTF is expressed (Gambardella et al., 2015). DMI is computationally more efficient than MINDy, but also requires more input information. As such, DMI is a complementary tool to MINDy rather than a replacement.

One of the most powerful CMI based algorithms to be developed recently infers correlated gene pairs based on the expression of a modulator, similar to CINDy and MINDy. ModulAted Gene/gene set InteraCtion (MAGIC) measures the difference in MI between sample groups to identify SSTF-target pairs that have sample group-dependent interactions (Hsiao et al., 2016). MAGIC can optionally make use of user-supplied gene lists, and thus it can be operated in either a targeted or untargeted mode. When applied to gene expression data from estrogen receptor (ER)± breast cancer cell lines, MAGIC revealed a novel ER-mediated FGFR1-FOXP3 interaction. From an algorithmic standpoint, MAGIC, which is implemented in MATLAB, is most similar to DIGGIT, but requires less computational time. MAGIC can be applied to and integrate multiple different types of data such as expression, DNA-methylation, and gene sets.

Network analysis in single-cell RNA-seq

One of the newest applications of DCENA is analysis of single-cell RNA-seq data. For example, DCENA was applied to single-cell gene expression profiles from mouse hippocampal cells from three different developmental states (Földy et al., 2016). The network was limited to cellular

adhesion, exocytosis, RhoGAP and RhoGEF related genes, and constructed from pairwise PCCs. Interestingly, the analysis revealed two distinct subnetworks with minimal overlap, with the first subnetwork corresponding to early-developmental expression, and the second subgraph corresponding to later-stage expression. Both subnetworks and their independence were highly conserved across all cell types, implying that a conserved network is responsible for different stages of hippocampal-cell development. In another case, DCENA was applied to single cells from different stages of mouse-embryonic fibroblast-derived neuronal cells (Treutlein et al., 2016). When filtered to only transcription factors, the analysis revealed three different subnetworks, each with high intraconnectivity, and each composed of genes specific to a certain stage. In the network, expression of the gene *Ascl1* was highly correlated with transcription factors that were specific to the initiation and maturation subnetworks while strongly negatively correlated with those specific to the MEFs. This supports a previous discovery that *Ascl1* maintains a chromatin state that enables induced-neuronal cell maturation. Due to the significant increase in sample size (in this case, the number of individual cells profiled) for correlation estimation that is made possible by single-cell RNA-seq, the technique promises to significantly advance the field of regulatory network inference and enable the development of new network reconstruction approaches (Gawad et al., 2016; Kolodziejczyk et al., 2015)).

Conclusions and Future Directions

For all its strengths, DCN as an analytical approach has several challenges that will need to be overcome in order for the full potential of the approach to be realized. The most pressing of these involves the statistical methods used, both for correlations and module identification (Ideker and Krogan, 2012). While there exists an agreed upon set of statistics (FDR, PCC, SCC) that most algorithms use, the set will need to be expanded as networks become more complex. This is especially true in the context of network inference using multi-omics datasets (e.g., gene expression, PPI, metabolomic, protein-DNA, expression quantitative trait locus, or DNA methylation data). A second challenge is delineating the "noise" contributions of (i) inherent stochasticity of molecular abundances (e.g., cell-to-cell variation in transcript or protein abundance due to "gene expression noise" (Elowitz et al., 2002)), (ii) biological variation between conditions or genotypes, and (iii) variation due to measurement error. Distinguishing these different sources of variation will require advances in methods for "ground-truth" measurement, methods for normalization of datasets to account for batch effects, and more mechanistically-

based statistical models.

The abundant new insights derived from the studies summarized in this review underscore that the analysis of biological networks, and especially differential correlation networks, has significant potential to generate new biological knowledge. While an inferred network by itself is a useful abstraction whose qualitative analysis can lead to insights, complex biological networks are most useful in the context of quantitative analysis to identify key regulators, modules, and functions. Most importantly, the amount of biological data we are able to generate continually increases. As sequencing costs decline and the amount of starting material required for sequencing is decreasing, the field of biological network profiling may approach a point where the limiting step in DCN analysis is the implementation and validation of the network analysis algorithms rather than the experimental assays (Barabási, 2009).

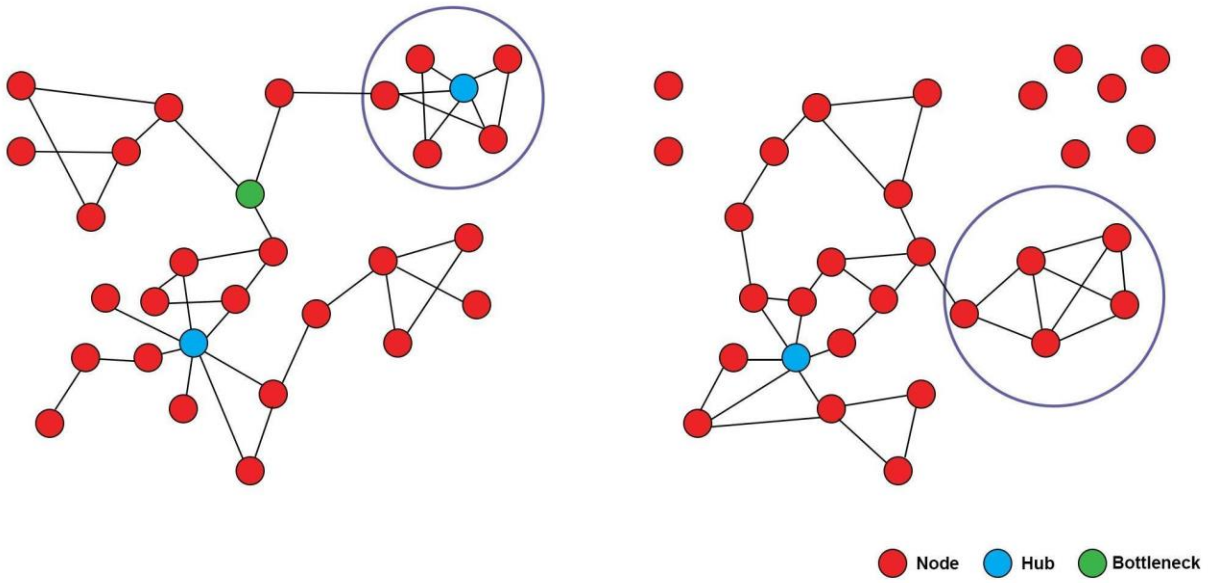


Figure 1.1 Basic biological networks.
Biological networks are visually represented as a series of nodes (red dots), connected by edges (lines). Typically, a network represents a snapshot in time, for example, a specific cell state. As cells develop, differentiate, or respond to disease, the cellular network changes structure. The two networks shown represent a change in cell state. Edges can be directed or undirected, to represent direction of flow of biological information. Each node (red dot) has a degree *k* that represents the number of nodes it is connected to. An undirected network has an average degree equal to twice the number of edges divided by the number of nodes. The average degree represents the average length of the shortest path between any two nodes in the network. Nodes that are highly connected are referred to as hubs (blue dot), and in many biological network contexts often correspond to essential or functionally significant nodes. Frequently in biological networks, nodes cluster together topologically into groups known as modules (e.g., blue circle), for example, a group of genes that work together to perform a specific biological function. Nodes that mediate communication between modules are referred to as "bottleneck" nodes (green dot).

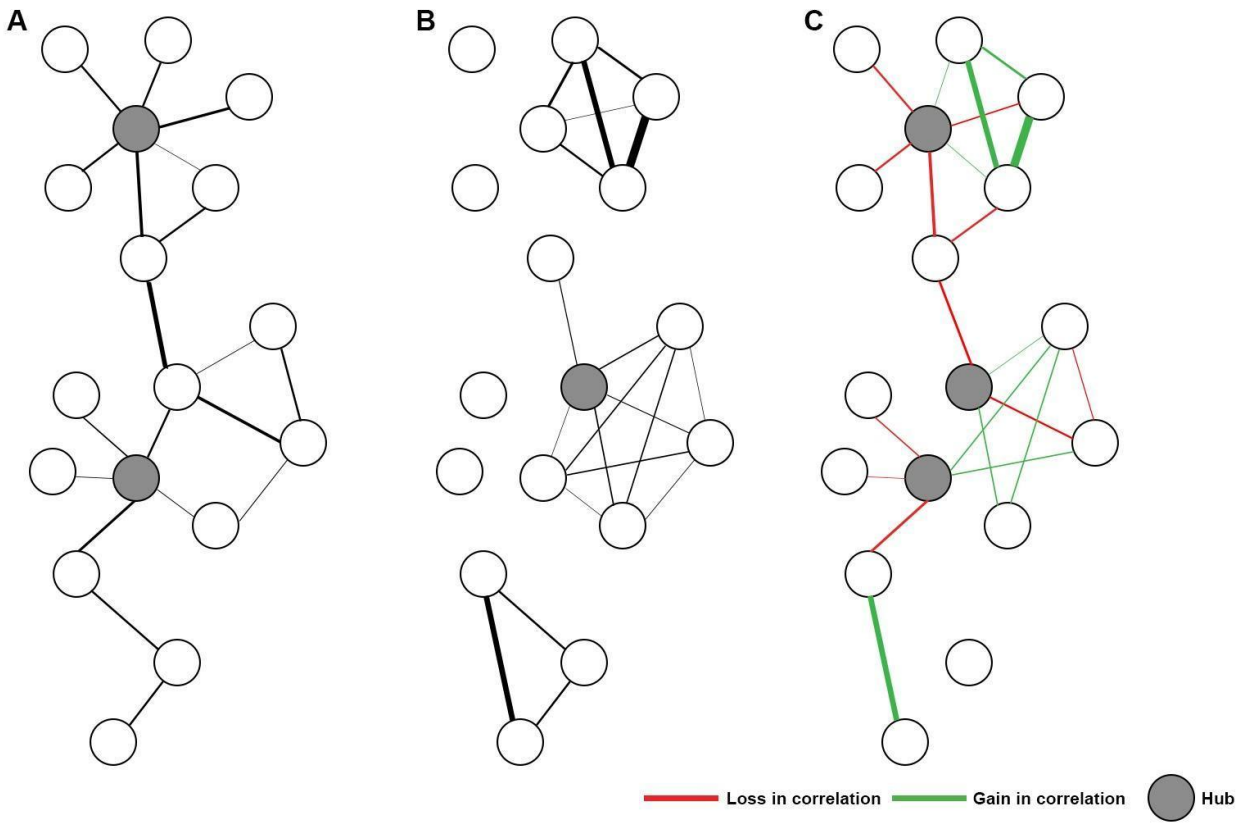


Figure 1.2 Basic differential network.
Differential networks are generated by "subtracting" two static networks from each other, for example, a "baseline" tissue (or cell) state from a "disease" tissue state.
(A) An example network from a control state. Nodes are shown as circles, and edges are lines, with weights representing how strong correlation between nodes is.
(B) An example network from a disease state. Nodes gain and lose edges, creating a substantially different network topology.
(C) Network B subtracted from network A. Red edges represent loss in correlation between two nodes, and green edges represent a gain in correlation. Differential networks allow simple visualization of correlation changes between nodes. Nodes that are "rewired" (i.e., that are connected in the differential network) are likely to be involved in the maintenance or creation of the change in cell state, a principle known as "guilt by rewiring".

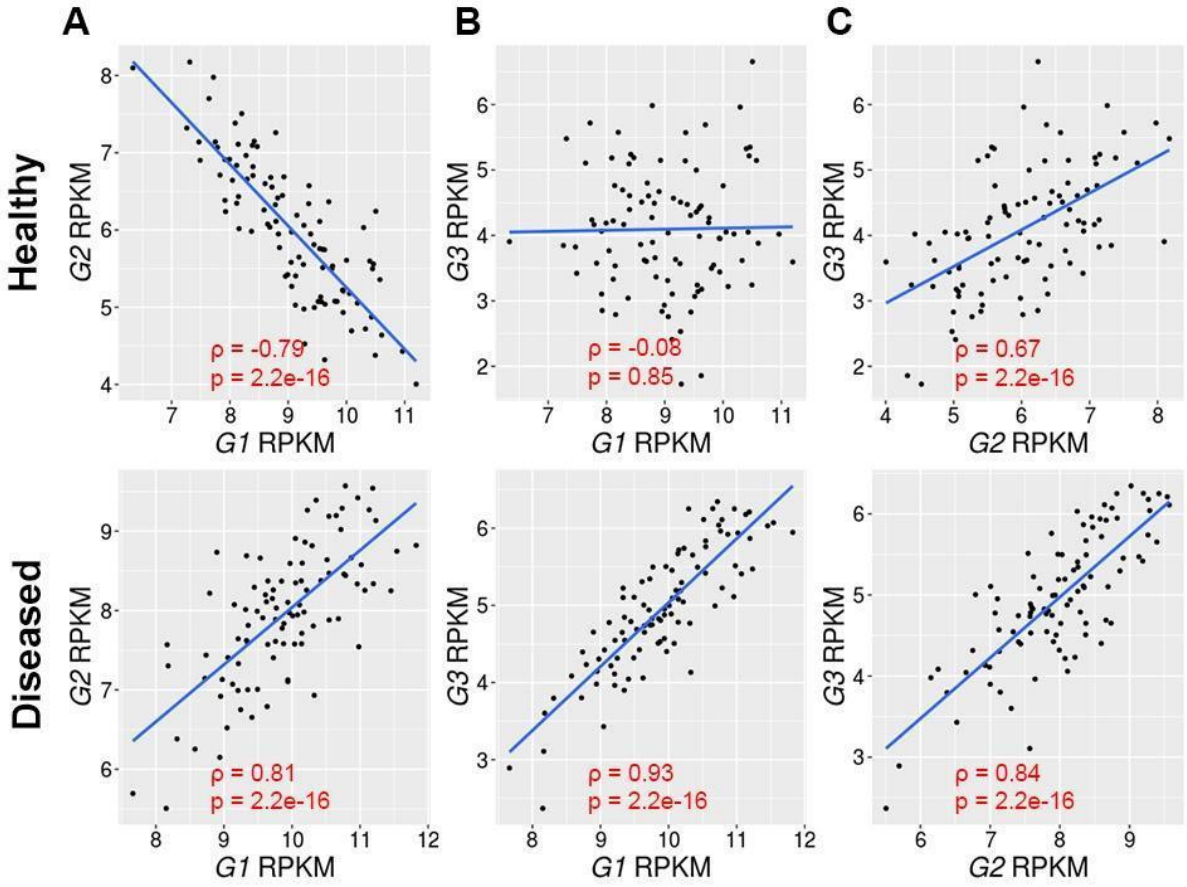


Figure 1.3 Differential correlation between samples.
Correlation between genes is determined by plotting normalized expression values of one gene vs. another across all biopsies within a biopsy group. Each dot represents a single biopsy and several biopsies represent a biopsy group. Correlation is determined by how well a line fits the plotted data, not by the slope of the line. SCC and p values (based on the Spearman test for independence based on 10,000 resamplings) for each correlation are shown on the plot.
(A) Gene G1 and G2 show a strong negative correlation in healthy biopsies, but in the disease state G1 and G2 correlation becomes positive. RPKM, reads (for this gene) per kilobase per million mapped reads, represents normalized gene expression data.
(B) Gene G1 and G3 show no correlation in healthy biopsies, and their correlation switches to strongly positive in the disease biopsies.
(C) Genes G2 and G3 show positive correlation in healthy biopsies, and additionally stronger positive correlation in disease biopsies. Genes that change in correlation between the cell states are likely to be involved in creating or maintaining the difference in cell states, making them candidates for therapeutics and diagnostics. Blue lines represent best-fit linear regressions and indicate positive or negative correlations.

Table 1.1 Software used for DCENA

Program	Method	Software	Comments	References
CoXpress	PCC	R, https://sourceforge.net/projects/coxpress/	Arbitrary cutoff for module ID	(Watson, 2006)
DiffCorr	PCC	R, https://cran.r-project.org/web/packages/DiffCorr/index.html	Used for gene expression profiling, metabolomics, etc	(Fukushima, 2013)
MCODE	CC	http://baderlab.org/Software/MCODE	Includes data visualization	(Bader and Hogue, 2003)
R/Ebcoexpress	CC	R, http://bioconductor.org/packages/release/bioc/html/EBcoexpress.html	Has built in data visualization	(Dawson et al., 2012)
DGCA	CC	R, https://github.com/andymckenzie/DGCA	Requires gene list as input (targeted)	(McKenzie et al., 2016)
DiNA	CC	http://dina.tigem.it/	Requires gene list as input (targeted)	(Gambardella et al., 2013)
WGCNA	CC	R, https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/	Has built in data visualization, compatible with other R packages	(Langfelder and Horvath, 2008)
iMDM	CC	Contact authors	Meant for multiple conditions	(Ma et al., 2015)
DiffCoEx	WGCNA	R, based of WGCNA, see additional file 1 in publication	Requires and uses WGCNA	(Tesson et al., 2010)
MINDy	CMI	http://wiki.c2b2.columbia.edu/workbench/index.php/Home		(Wang et al., 2009)
CINDy	CMI	http://califano.c2b2.columbia.edu/mindy2-cindy	Based on MINDy	(Giorgi et al., 2014)
CMI2NI	CMI	http://comp-sysbio.org/cmi2ni/index.html		(Zhang et al., 2015)
DMI	CMI	http://dmi.tigem.it/	Requires gene list as input (targeted)	(Gambardella et al., 2015)
MAGIC	CMI	MATLAB, https://github.com/chiuyc/MAGIC		(Hsiao et al., 2016)
DIGGIT	CMI	R, https://www.bioconductor.org/packages/release/bioc/html/diggit.html		(Alvarez et al., 2015)

REFERENCES

Ahn, R., Gupta, R., Lai, K., Chopra, N., Arron, S.T., and Liao, W. (2016). Network analysis of psoriasis reveals biological pathways and roles for coding and long non-coding RNAs. BMC Genomics *17*, 841.

Albert, R., Jeong, H., and Barabasi, A.L. (2000). Error and attack tolerance of complex networks. Nature *406*, 378–382.

Alon, U., Surette, M.G., Barkai, N., and Leibler, S. (1999). Robustness in bacterial chemotaxis. Nature *397*, 168–171.

Alvarez, M.J., Chen, J.C., and Califano, A. (2015). DIGGIT: a Bioconductor package to infer genetic variants driving cellular phenotypes. Bioinformatics *31*, 4032–4034.

Amar, D., Safer, H., and Shamir, R. (2013). Dissection of Regulatory Networks that Are Altered in Disease via Differential Co-expression. PLoS Comput. Biol. *9*, e1002955.

Bader, G.D., and Hogue, C.W.V. (2003). An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics *4*, 2.

Barabási, A.-L. (2009). Scale-free networks: a decade and beyond. Science *325*, 412–413.

Barabási, A.-L., and Albert, R. (1999). Emergence of Scaling in Random Networks. Science *286*, 509–512.

Barabasi, A.-L., and Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. Nat. Rev. Genet. *5*, 101–113.

Basso, K., Margolin, A.A., Stolovitzky, G., Klein, U., Dalla-Favera, R., and Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. Nat. Genet. *37*, 382–390.

Beer, M.A., and Tavazoie, S. (2004). Predicting gene expression from sequence. Cell *117*, 185–198.

Bornholdt, S. (2005). Systems biology. Less is more in modeling large genetic networks. Science *310*, 449–451.

Butte, A.J., and Kohane, I.S. (2000). Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. Pac. Symp. Biocomput. 418–429.

Creixell, P., Schoof, E.M., Simpson, C.D., Longden, J., Miller, C.J., Lou, H.J., Perryman, L., Cox, T.R., Zivanovic, N., Palmeri, A., et al. (2015). Kinome-wide decoding of network-attacking mutations rewiring cancer signaling. Cell *163*, 202–217.

Dawson, J.A., Ye, S., and Kendziorski, C. (2012). R/EBcoexpress: an empirical Bayesian framework for discovering differential co-expression. Bioinformatics *28*, 1939–1940.

Dong, X., Yambartsev, A., Ramsey, S.A., Thomas, L.D., Shulzhenko, N., and Morgun, A. (2015). Reverse enGENEering of Regulatory Networks from Big Data: A Roadmap for Biologists. Bioinform. Biol. Insights *9*, 61–74.

Ecker, J.R., Bickmore, W.A., Barroso, I., Pritchard, J.K., Gilad, Y., and Segal, E. (2012). Genomics: ENCODE explained. Nature *489*, 52–55.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. U. S. A. *95*, 14863–14868.

Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic Gene Expression in a Single Cell. Science *297*, 1183–1186.

Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. *30*, 1575–1584.

Földy, C., Darmanis, S., Aoto, J., Malenka, R.C., Quake, S.R., and Südhof, T.C. (2016). Single-cell RNAseq reveals cell adhesion molecule profiles in electrophysiologically defined neurons. Proc. Natl. Acad. Sci. U. S. A. *113*, E5222–E5231.

de la Fuente, A., Bing, N., Hoeschele, I., and Mendes, P. (2004). Discovery of meaningful associations in genomic data using partial correlation coefficients. Bioinformatics *20*, 3565–3574.

Fukushima, A. (2013). DiffCorr: an R package to analyze and visualize differential correlations in biological networks. Gene *518*, 209–214.

Gambardella, G., Moretti, M.N., de Cegli, R., Cardone, L., Peron, A., and di Bernardo, D. (2013). Differential network analysis for the identification of condition-specific pathway activity and regulation. Bioinformatics *29*, 1776–1785.

Gambardella, G., Peluso, I., Montefusco, S., Bansal, M., Medina, D.L., Lawrence, N., and di Bernardo, D. (2015). A reverse-engineering approach to dissect post-translational modulators of transcription factor's activity from transcriptional data. BMC Bioinformatics *16*, 279.

Gawad, C., Koh, W., and Quake, S.R. (2016). Single-cell genome sequencing: current state of the science. Nat. Rev. Genet. *17*, 175–188.

Giorgi, F.M., Lopez, G., Woo, J.H., Bisikirska, B., Califano, A., and Bansal, M. (2014). Inferring Protein Modulation from Gene Expression Data Using Conditional Mutual Information. PLoS One *9*, e109569.

Grechkin, M., Logsdon, B.A., Gentles, A.J., and Lee, S.-I. (2016). Identifying Network Perturbation in Cancer. PLoS Comput. Biol. *12*, e1004888.

Guitart, X., Bonaventura, J., Rea, W., Orrú, M., Cellai, L., Dettori, I., Pedata, F., Brugarolas, M., Cortés, A., Casadó, V., et al. (2016). Equilibrative nucleoside transporter ENT1 as a biomarker of Huntington disease. Neurobiol. Dis. *96*, 47–53.

Hartwell, L.H., Hopfield, J.J., Leibler, S., and Murray, A.W. (1999). From molecular to modular cell biology. Nature *402*, C47–C52.

Hou, L., Chen, M., Zhang, C.K., Cho, J., and Zhao, H. (2014). Guilt by rewiring: gene prioritization through network rewiring in genome wide association studies. Hum. Mol. Genet. *23*, 2780–2790.

Hsiao, T.-H., Chiu, Y.-C., Hsu, P.-Y., Lu, T.-P., Lai, L.-C., Tsai, M.-H., Huang, T.H.-M., Chuang, E.Y., and Chen, Y. (2016). Differential network analysis reveals the genome-wide landscape of

estrogen receptor modulation in hormonal cancers. Sci. Rep. *6*, 23035.

Ideker, T., and Krogan, N.J. (2012). Differential network biology. Mol. Syst. Biol. *8*, 565.

Ideker, T., Galitski, T., and Hood, L. (2001). A new approach to decoding life: systems biology. Annu. Rev. Genomics Hum. Genet. *2*, 343–372.

Ihaka, R., and Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. J. Comput. Graph. Stat. *5*, 299–314.

Jeong, H., Mason, S.P., Barabási, A.L., and Oltvai, Z.N. (2001). Lethality and centrality in protein networks. Nature *411*, 41–42.

Jiang, X., Zhang, H., and Quan, X. (2016). Differentially Coexpressed Disease Gene Identification Based on Gene Coexpression Network. Biomed Res. Int. *2016*, 3962761.

Kayano, M., Higaki, S., Satoh, J.-I., Matsumoto, K., Matsubara, E., Takikawa, O., and Niida, S. (2016). Plasma microRNA biomarker detection for mild cognitive impairment using differential correlation analysis. Biomark Res *4*, 22.

Kirschner, M.W. (2005). The meaning of systems biology. Cell *121*, 503–504.

Kitano, H. (2002). Systems biology: a brief overview. Science *295*, 1662–1664.

Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C., and Teichmann, S.A. (2015). The technology and biology of single-cell RNA sequencing. Mol. Cell *58*, 610–620.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., et al. (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. Science *298*, 799–804.

Li, F., Long, T., Lu, Y., Ouyang, Q., and Tang, C. (2004). The yeast cell-cycle network is robustly designed. Proc. Natl. Acad. Sci. U. S. A. *101*, 4781–4786.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell *133*, 523–536.

Ma, X., Gao, L., Karamanlidis, G., Gao, P., Lee, C.F., Garcia-Menendez, L., Tian, R., and Tan, K. (2015). Revealing Pathway Dynamics in Heart Diseases by Analyzing Multiple Differential Networks. PLoS Comput. Biol. *11*, e1004332.

Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., and Califano, A. (2006). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics *7 Suppl 1*, S7.

McKenzie, A.T., Katsyv, I., Song, W.-M., Wang, M., and Zhang, B. (2016). DGCA: A comprehensive R package for Differential Gene Correlation Analysis. BMC Syst. Biol. *10*, 106.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. Science *298*, 824–827.

Mitra, K., Carvunis, A.-R., Ramesh, S.K., and Ideker, T. (2013). Integrative approaches for finding modular structure in biological networks. Nat. Rev. Genet. *14*, 719–732.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods *5*, 621–628.

Oros Klein, K., Oualkacha, K., Lafond, M.-H., Bhatnagar, S., Tonin, P.N., and Greenwood, C.M.T. (2016). Gene Coexpression Analyses Differentiate Networks Associated with Diverse Cancers Harboring TP53 Missense or Null Mutations. Front. Genet. *7*, 137.

Prill, R.J., Iglesias, P.A., and Levchenko, A. (2005). Dynamic Properties of Network Motifs Contribute to Biological Network Organization. PLoS Biol. *3*, e343.

Ratushny, A.V., Saleem, R.A., Sitko, K., Ramsey, S.A., and Aitchison, J.D. (2012). Asymmetric positive feedback loops reliably control biological responses. Mol. Syst. Biol. *8*, 577.

Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., and Barabási, A.L. (2002). Hierarchical organization of modularity in metabolic networks. Science *297*, 1551–1555.

Reiss, D.J., Baliga, N.S., and Bonneau, R. (2006). Integrated biclustering of heterogeneous genome-wide datasets for the inference of global regulatory networks. BMC Bioinformatics *7*, 280.

Reyes-Bermudez, A., Villar-Briones, A., Ramirez-Portilla, C., Hidaka, M., and Mikheyev, A.S. (2016). Developmental Progression in the Coral Acropora digitifera Is Controlled by Differential Expression of Distinct Regulatory Gene Networks. Genome Biol. Evol. *8*, 851–870.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Saraph, V., and Milenković, T. (2014). MAGNA: Maximizing Accuracy in Global Network Alignment. Bioinformatics *30*, 2931–2940.

Scarpa, J.R., Jiang, P., Losic, B., Readhead, B., Gao, V.D., Dudley, J.T., Vitaterna, M.H., Turek, F.W., and Kasarskis, A. (2016). Systems Genetic Analyses Highlight a TGFβ-FOXO3 Dependent Striatal Astrocyte Network Conserved across Species and Associated with Stress, Sleep, and Huntington's Disease. PLoS Genet. *12*, e1006137.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Siska, C., Bowler, R., and Kechris, K. (2015). The Discordant Method: A Novel Approach for Differential Correlation. Bioinformatics.

Southworth, L.K., Owen, A.B., and Kim, S.K. (2009). Aging Mice Show a Decreasing Correlation of Gene Expression within Genetic Modules. PLoS Genet. *5*, e1000776.

Tesson, B.M., Breitling, R., and Jansen, R.C. (2010). DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. BMC Bioinformatics *11*, 497.

Treutlein, B., Lee, Q.Y., Camp, J.G., Mall, M., Koh, W., Shariati, S.A.M., Sim, S., Neff, N.F., Skotheim, J.M., Wernig, M., et al. (2016). Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. Nature *534*, 391–395.

Troy, N.M., Hollams, E.M., Holt, P.G., and Bosco, A. (2016). Differential gene network analysis for the identification of asthma-associated therapeutic targets in allergen-specific T-helper memory responses. BMC Med. Genomics *9*, 9.

Wang, K., Saito, M., Bisikirska, B.C., Alvarez, M.J., Lim, W.K., Rajbhandari, P., Shen, Q., Nemenman, I., Basso, K., Margolin, A.A., et al. (2009). Genome-wide identification of post-translational modulators of transcription factor activity in human B cells. Nat. Biotechnol. *27*, 829–839.

Watson, M. (2006). CoXpress: differential co-expression in gene expression data. BMC Bioinformatics *7*, 509.

Wu, S., Li, J., Cao, M., Yang, J., Li, Y.-X., and Li, Y.-Y. (2016). A novel integrated gene coexpression analysis approach reveals a prognostic three-transcription-factor signature for glioma molecular subtypes. BMC Syst. Biol. *10 Suppl 3*, 71.

Zhang, X., Zhao, J., Hao, J.-K., Zhao, X.-M., and Chen, L. (2015). Conditional mutual inclusive information enables accurate quantification of associations in gene regulatory networks. Nucleic Acids Res. *43*, e31.

Zuo, Z.-G., Zhang, X.-F., Ye, X.-Z., Zhou, Z.-H., Wu, X.-B., Ni, S.-C., and Song, H.-Y. (2016). Bioinformatic analysis of RNA-seq data unveiled critical genes in rectal adenocarcinoma. Eur. Rev. Med. Pharmacol. Sci. *20*, 3017–3025.

Mapping the Chromatin State Dynamics in Myoblasts

Chapter 2

Arun J. Singh, Michael K. Gross, Theresa M. Filtz, Chrissa Kioussi

Gene Reports
Published February 1, 2016
Gene Reports, 3, 5–13
https://doi.org/10.1016/j.genrep.2016.01.003

Abstract

Genome-wide mapping reveals chromatin landscapes unique to cell states. Histone marks of regulatory genes involved in cell specification and organ development provide a powerful tool to map regulatory sequences. H3K4me3 marks promoter regions; H3K27me3 marks repressed regions, and Pol II presence indicates active transcription. The presence of both H3K4me3 and H3K27me3 characterize poised sequences, a common characteristic of genes involved in pattern formation during organogenesis. We used genome-wide profiling for H3K27me3, H3K4me3, and Pol II to map chromatin states in mouse embryonic day 12 forelimbs in wild type (control) and *Pitx2*-null mutant mice. We compared these data with previous gene expression studies from forelimb Lbx1+migratory myoblasts and correlated *Pitx2*-dependent expression profiles and chromatin states. During forelimb development, several lineages including myoblast, osteoblast, neurons, angioblasts, etc., require synchronized growth to form a functional limb. We identified 125 genes in the developing forelimb that are *Pitx2*-dependent. Genes involved in muscle specification and cytoskeleton architecture were positively regulated, while genes involved in axonal pathfinding were poised. Our results have established histone modification profiles as a useful tool for identifying gene regulatory states in muscle development, and identified the role of Pitx2 in extending the time of myoblast progression, promoting formation of sarcomeric structures, and suppressing attachment of neuronal axons.

Introduction

Muscle formation during development is a tightly regulated process. Non-committed mesoderm-derived cells delaminate, migrate and fuse to form a mature muscle. This process is characterized by a series of developmental stages that are determined by a constellation of sequence specific transcription factors (SSTFs). Limb muscles originate from somites, the anatomical structures of the paraxial mesoderm that form in a rostral-caudal axis. The dermomyotome develops as a dorsal epithelium of the early somite and gives rise to skeletal muscles. The dermomyotome is subdivided into hypaxial and epaxial that give rise to the lateral trunk musculature and deep back muscles, respectively. Myogenic precursors in the dermomyotome express the paired homeodomain factors Pax3, Pax7 and the basic Helix-Loop-Helix transcription factor Myf5 (Goulding et al., 1994; Jostes et al., 1990; Kiefer and Hauschka, 2001). The $Pax3^{+}$ cells maintain their proliferative and undifferentiated status by external cues from the lateral plate mesoderm and surface ectoderm (Amthor et al., 1999). Limb myogenesis is initiated upon positive signaling by Wnt from the dorsal neural tube (Cossu et al., 1996) and Shh from the floor plate and notochord (Gustafsson et al., 2002). BMP signaling from the lateral plate mesoderm inhibits Myod while BMP signaling from the dorsal neural tube inhibits Myf5 (and thus myogenesis). In the limb, many cells co-express Pax3 and Myf5 as the progenitor cells that have entered the myogenic program need a more extended myoblast state before differentiation. Some of the $Pax3^{+}$ cells remain as reserve cells and subsequently express Pax7, Myod, and Myogenin (Shih et al., 2007a)**.** As cells enter the myogenic program, Pax3 regulates enhancer elements of Myf5. The transcription of Myf5 is regulated in a spatiotemporal manner by a large number of upstream enhancers distributed over 100kb (Carvajal et al., 2008). It has been shown that Pax3 regulates Myf5 expression from two known *cis* regulatory elements, one in the hypaxial somite at -110kb and another in the limbs at -57.5kb (Carvajal et al., 2008).

Pitx2 is a homeodomain transcription factor expressed in the lateral plate mesoderm, and in muscle anlagen during all stages of myogenic progression (Shih et al., 2007a, 2008). Pitx2 somatic null mutants die between embryonic day (E) 12.5 and E14.5 due to arrest of organ development (Gage et al., 1999; Kitamura et al., 1999; Lin et al., 1999; Lu et al., 1999). Pitx2 is a competence factor required for the temporally ordered and growth factor-dependent recruitment of a series of specific co-activator complexes that prove necessary for Ccnd2 gene induction (Kioussi et al., 2002). Pitx2 contributes to specification of the anatomical context that surrounds the muscle, including the jaw (Shih et al., 2007b) and the abdominal wall (Eng et al., 2012). Pitx2 regulates the relative amounts and types of cytoskeletal proteins that are produced as muscle

cells assemble into muscles and contributes to the higher order muscle assembly (Campbell et al., 2012a).

Protein-DNA interactions comprise the genetic regulatory networks (GRNs) that exist in any cell. These GRNs are responsible for maintaining cellular identity and managing differentiation in response to internal or external stimuli (Davidson et al., 2002). They consist of a multitude of SSTFs interacting directly with DNA and each other in a highly regulated manner throughout the entire genome. Understanding these GRNs is required to understand the drivers of biological changes in a cell, and is enabled by sequencing of the chromatin (Chip-seq), which allows for mapping specific protein-DNA interactions across the whole genome (Park, 2009).

Histone modifications control both gene repression and transcription, including CRM activity (Birney, 2012). The chromatin mark Histone H3 lysine 4 trimethylation (H3K4me3) is associated with open chromatin, promoter regions carried out by the Trithorax (trxG) activator protein complex, and frequently found in the presence of RNA polymerase II (Pol II). The modification Histone H3 lysine 27 trimethylation (H3K27me3) is associated with closed chromatin and gene repression, carried out by the Polycomb (PcG) repressor protein complex (Guenther et al., 2007). However, H3K4me3 and H3K27me3 often co-occur and bivalently mark chromatin, most often in promoter regions (Bernstein et al., 2006). Genes under bivalent promoters tend to show low levels of transcription, but are poised before being activated once the cell commits itself to a lineage (Mikkelsen et al., 2007a). Identification of bivalent chromatin is an important factor for decoding gene regulatory networks in development.

The chromatin states of mouse E12 forelimbs from control and *Pitx2* somatic-null mice were analyzed for H3K4me3, H3K27me3 and Pol II signatures. Correlation studies were performed for the sequences marked with different signatures. Differential enrichment of histone marks and Pol II were observed around a number of Pitx2 target genes identified in *Lbx1*$^{EGF/+}$ migratory myoblasts (Campbell et al., 2012a). Quantitative real-time PCR assays (qPCR) assessed gene expression and revealed that the H3K27me3 chromatin mark best predicted relative gene expression. Collectively, these studies suggest that Pitx2 dynamically regulates the chromatin state of genes involved in the myoblast proliferative state and axonal path finding in the developing forelimb.

Materials and Methods

Mice

All research was conducted according to the protocols reviewed and approved by the Oregon State University Institutional Animal Care and Use Committee. The *Pitx2*$^{+/Z}$ mouse line was maintained on an outcrossed ICR background. Noon on the day of a vaginal plug was considered embryonic day (E) 0.5. Yolk sacs of embryos were used for genotyping.

Gene Expression Arrays in Mouse Forelimb Migratory Myoblasts

Pitx2 target genes in mouse Lbx1^{+} migratory muscle cells from E12.5 *Lbx1*$^{EGFP/+}$*/Pitx2*$^{+/+}$ (WT) [27] and *Lbx1*$^{EGFP/+}$*/Pitx2*$^{Z/Z}$ (MT) mice (Campbell et al., 2012b) forelimbs, the dataset GSE31945 NCBI Gene Expression Omnibus and subsets of GSM791677, GSM791678, GSM791683, and GSM791684 were used (Table 2.S1). All data analysis was done in R using the Bioconductor package and its components (Gentleman et al., 2004). GEO data was first imported into R using the GEOquery package (Davis and Meltzer, 2007). The Simpleaffy package was used to normalize all datasets based on the RMA algorithm. Subsequent analysis and comparison was performed using the Limma Packageactive (Smyth, 2005). Relevant Pitx2 target sequences were identified as coding sequences with a significant fold-change difference between WT and MT of an adjusted P value < 0.1 (Benjamini-Hochberg FDR).

Chip-Seq Data Analysis

ChIP seq data was analyzed from dataset GSE49010 representing day E12.5 forelimbs from *Pitx2*$^{+/+}$ (WT) and *Pitx2*$^{Z/Z}$ (MT) mouse as described previously (Eng et al., 2014). WT data was obtained from our previous GSE49010 while MT data was newly generated and deposited in NCBI GEO under GSE71128 (Campbell et al., 2012a). Fastq files along with their inputs were aligned to the mouse genome (mm10/NCBI38) reference assembly using Bowtie2 version 2.2.3, with default parameters (Langmead and Salzberg, 2012). Samtools version 1.0 was used to convert the aligned.sam files into sorted.bam files (Li et al., 2009). The bedtools version 2.12.0 command bamToBed was used to convert the sorted.bam files into.bed files. The peak-calling algorithm MACS version 2.1.0 was used to identify regions of the mouse genome significantly enriched in the ChIP-seq samples over the controls (Zhang et al., 2008). MACS2 was run with the following parameters: “-f BAM -B --SPMR --broad --broad-cutoff 0.1 -g 1.87e9 –shiftsize 80.”

The function annotatepeaks.pl from Homer version 4.7 was used for annotation and comparison of called peaks (Heinz et al., 2010). Peaks were assigned to genes if the gene nearest to the peak was less than 2,000 bp away from the gene's TSS. The function mergepeaks.pl with "-d given" was used to determine overlapping peaks of different chromatin marks within each sample, and to identify genes with different chromatin states between WT and MT. The annotatepeaks.pl function was used separately with the -hist option to generate bins of ChIP fragment density centered around the transcription start sites of known transcripts. The output bed and bedgraph files were converted to bigbed and bigwig, respectfully. The resulting files were visualized with the UCSC genome browser (Karolchik et al., 2003). Raw sequences from MT forelimbs was deposited in the NCBI Gene Expression Omnibus GSE71128.

Results

Pitx2-dependent chromatin state of mouse forelimbs

Muscle forelimbs are distorted and fail to form higher order muscle assembly in the absence of Pitx2 (Campbell et al., 2012a). The altered expression profile of numerous gene families is the result of changes in the chromatin state due to the absence of Pitx2 functional protein in muscle cells. To correlate gene expression and chromatin state, ChIP-seq analyses for H3K4me3, H3K27me3 and Pol II in WT and MT E12 forelimbs were performed and their signatures compared. Mouse E12 forelimb harbors several muscle lineages and, to maximize fidelity, the current analysis is focused on the Pitx2 target sequences in the Lbx1 lineage (Campbell et al., 2012a, 2012b).

The R Bioconductor package and components were applied as described in the materials and methods section. The Pitx2-regulated 125 protein-coding transcripts identified by microarray (Campbell et al., 2012b), will be referred to as Pitx2 targets from herein. However, chromatin marks based on the peak-calling algorithm were exhibited only in the promoter region (0-2,000bp) or gene body domains (exons, introns, and 5' and 3' un-transcribed sequences), called loci from herein, of 101 Pitx2 target transcripts in WT and 103 in MT (Table 2.1). Pitx2 targets were grouped based on the combinations of H3K4me3, H3K27me3, and Pol II present at loci using Homer tools mergepeaks. No loci were marked with only Pol II, or with both H3K27me3 and Pol II. Summing loci for each mark resulted in 35% occupied by Pol II in WT (35/101) and 28% in MT (29/103). Likewise, 18% of the loci were occupied by only H3K4me3 in WT (18/101) and 22% were

occupied in MT (23/103), while only 13% (13/101) and 12% (12/103) were occupied only by H3K27me3 in WT and MT respectively. Pol II promoter occupancy is correlated with active gene expression in the presence of H3K4me3 (Guenther et al., 2007). The majority of Pitx2 targets were occupied by both H3K4me3 and H3K27me3, 36% in WT (36/101) and in 37% in MT (38/103), while the trivalent state, H3K4me3, H3K27me3 and Pol II occupancy, represented 13% of the targets in WT (13/101) and 7% in MT (7/103). Venn diagrams illustrate the overlap of chromatin marks (Pol II, H3K27me3, and H3K4me3) in WT and MT (Figure 2.1). Bivalently marked loci are found at higher frequency in embryonic stem cells (Bernstein et al., 2006) than in committed postmitotic cells (Mikkelsen et al., 2007b). The presence of numerous poised, inactive (bivalent) loci at E12 forelimbs, in both WT and MT mice, suggests that genes involved in patterning and organ development are stalled to allow the next set of genes to be activated and establish the next developmental state of the mitotic cells. The strong increase of solely H3K4me3 (22%) activated loci in combination with the decreased poised trivalent loci (54%) in MT suggests that Pitx2 negatively regulates the transcription of its targets.

Mapped read distribution around Refseq TSSs

The confirmed changes in chromatin state in Pitx2 target transcripts was followed by analysis of the patterning changes in the chromatin distribution. Homer tools v4.7 was used to generate tag density plots (Figure 2.2A-D) in different sequence sets. The detected tag density for all three marks were in accord with all sequences in the genome (Figure 2.2A), (Barski et al., 2007). The overall tag density was higher in the WT biopsies relative to the MT. This follows for all marks and persists despite the normalization of tag density for mapped reads. Examination of tag densities around the Pitx2 target transcripts (Figure 2.2B) showed a similar pattern. All curves appeared slightly less sharp due to the smaller sampling size (20K total transcripts vs. 125 Pitx2 target transcripts). The density of H3K27me3 tags became increasingly noisy, but the two-peak distribution remained visible. The tag density of H3K4me3 and Pol II was decreased, while tag density of H3K27me3 was increased in both WT and MT. These data imply that chromatin was more condensed around identified Pitx2 target transcripts, relative to all transcripts in both WT and MT. When compared to all identified Pitx2-target, non-SSTF transcripts (Figure 2.2C), an almost identical pattern emerged (Figure 2.2C). This pattern remained the same in the Pitx2 target SSTFs but was noisy due to the smaller transcript number (Figure 2.2D). In summary, embryonic forelimb chromatin for all genomic sequences was marked for H3K4me3 and Pol II, with dynamic

changes between WT and MT. Chromatin was rich for H3K4me3 marks, and for H3K4me3 and H3K27me3 for SSTF targets, suggesting that SSTF chromatin harbors active and inactive domains to determine distinct cell lineages as myogenesis progresses.

Chromatin state of Pitx2 target genes

To investigate the link between chromatin state and gene expression, the chromatin state around Pitx2 target genes in the Lbx1 lineage was visualized in detail (Table 2.1). Chromatin profiling for H3K4me3, H3K27me3 and Pol II in coding, and 5' and 3' non-translated sequences identified active and repressed domains (Figure 2.3). Analysis indicated that 23 transcripts (Table 2.1, indicated as bold) including Pitx2 exhibited differential chromatin marks. These genes are involved in organogenesis (Pitx2, Rfx8, Ebf1, Acta2), neurogenesis and axon path-finding (Astn1, Mab21l1, Tubb3, Dcx, Nrcam, Cadps, Lin7, Zswim5), transport (Ddah1, Slco3a1, Atp1b1), cell cycle (Csrnp3), cytoskeleton and organelle organization (Hook1, Cntn2, Crmp, Stmn2, Nol4, Lamp5), myogenesis (Fstl5) and osteogenesis (Skor1).

Genes were clustered in six groups based on the chromatin signature of the WT myoblasts. Genes that exhibited H3K4me3, H3K27me3 and Pol II presence, trivalent state (Figure 2.3A; Pitx2, Ebf1, Nol4, Slco3a1, Zswim5, Ddah1). Genes that exhibited H3K4me3 and Pol II presence, activated state (Figure 2.3B; Atp1b1, Mab21l1, Astn1, Csnrp3, Hook1). Genes that exhibited H3K4me3 and H3K27me3 presence, bivalent state (Figure 2.3C; Lin7). Genes that exhibited only H3K4me3 presence, open chromatin (Figure 2.3D; Hook1). Genes that exhibited only H3K27me3 presence, closed chromatin (Figure 2.3E; Scor1). Genes that did not exhibit any chromatin signature in the WT (Figure 2.3F; Fstl5).

Pitx2 mutants were generated by ablation of the homeodomain located between exon 4 and 5. ChIP-seq analysis indicated the expected absence of signature in the ablated sequences that was used as a technical control (Figure 2.3A). The Pitx2 locus was trivalent in WT while the Pol II signature was missing in MT due to the truncated protein. Ebf1, a helix-loop-helix transcription factor, is involved in B-cell lymphopoiesis (Hagman et al., 1993) in positioning the mesenchyme-derived ulna and radius and connective tissues surrounding tendons (Mella et al., 2004) and adipogenesis (Jimenez et al., 2007). While Elf1 was in the trivalent state in WT forelimbs, it was in an active state in the MT, suggesting that Pitx2 represses its activity towards the adipogenic and cartilage formation pathways.

The solute carrier organic anion transporter family member 3a1 (Slco3a1), associated with pathways that transport glucose, bile salts, organic acids, metal ions and amino acids, was inhibited by Pitx2 (Figure 2.3A). The dimethyl arginine dimethylaminohydrolase (Ddah1) plays a role in nitric oxide generation by regulating cellular concentrations of methylarginines, which in turn inhibit nitric oxide synthase activity, and is highly expressed during forelimb development (Breckenridge et al., 2010). While Ddah1 is in a trivalent state in WT, it is activated in the MT suggesting that Pitx2 acts as a repressor at this locus (Figure 2.3A).

The sodium/potassium-transporting ATPase subunit beta-1 (Atp1b1), an integral membrane protein responsible for establishing and maintaining the electrochemical gradients of Na and K ions across the plasma membrane, was active in WT with open chromatin in MT (Figure 2.3B), suggesting that Pitx2 regulates Na transport during limb development. Mab21l1 is a downstream target of transforming growth factor beta (TGFß) signaling (Mariani et al., 1999) and is involved in development of several organs including limb, neuronal and vascular systems (Heanue and Pachnis, 2006; Wong and Chow, 2002). Mab21l1 exhibited domains with a trivalent signature in MT while the H3K27me3 inactivation signature was missing in WT (Figure 2.3B). Mab21l1 is expressed in similar tissues with Pitx2, and Mab21l1 mutants are characterized by axial turning and abdominal malformations similar to Pitx2. The chromatin signature in the forelimb suggests that Pitx2 promotes Mab21l1 activation. Astrotactin, Astn1, is a neuronal adhesion molecule that acts as a ligand for glial-guided migration of neuroblasts (Adams et al., 2002). Astn1 domains enriched for H3K4me3 and Pol II were located close to the TSS (Figure 2.3B). The absence of the Pol II signature in MT suggests that its active transcription is Pitx2-regulated. Cysteine and Glycine-Rich Protein 3 (Csrnp3) promotes myogenic differentiation by acting as a cofactor of Myod, myogenin and MRF4 to increase their interactions with specific DNA regulatory elements (Kong et al., 1997). Csrnp3 also acts as a scaffold protein that promotes the assembly of macromolecular complexes along sarcomeres and cytoskeleton (Arber et al., 1997). The Csrnp3 locus was active in MT with H3K4me3 and Pol II occupancies, while in WT the Pol II signature was missing (Figure 2.3B). The Csrnp3 locus harbored active regulatory regions just downstream of the TSS suggesting that Pitx2 regulates Csrnp3 active regulatory regions and extends the myoblast state before their commitment to specified lineages.

Lin7, plays a role in establishing and maintaining the asymmetric distribution of channels and receptors at the plasma membrane of polarized cells, and in stabilizing the cell junctions (Bohl et

al., 2007). Lin7 was characterized by a bivalent chromatin state in WT and open chromatin in MT, suggesting that Pitx2 regulates cell polarization during limb development.

Hook homologue 1, Hook 1, is expressed in tubular endosomes and facilitates the directed recycling of clathrin-independent endocytosis cargo proteins, such as CD98 and CD147, through its interaction with microtubules and their cytoplasmic sequences on sorting endosomes (Maldonado-Báez et al., 2013). This sorting provides a means to monitor protein quality control of cell surface proteins. The Hook1 locus was active to transcription in MT, occupied by H3K4me3 and Pol II, while the Pol II mark was missing in WT (Figure 2.3B). This chromatin change suggests that Pitx2 influences the activity of this locus and the ability of myoblasts to segregate and rapidly recycle proteins, an activity that is carefully regulated in tissues during development.

The Ski family transcriptional corepressor 1, Skor1, interacts with Lbx1 and cooperatively represses transcription and acts as a transcriptional corepressor for Lbx1 in regulating cell fate determination in the spinal cord (Mizuhara et al., 2005). Skor1 interacts with general transcriptional corepressors, such as Hdac1, Ctbp and Grg1. Skor1 represses TGFß signaling through inhibition of transcriptional activity of Smad proteins and negative regulation of the bone morphogenetic proteins (BMPs) (Arndt et al., 2007). BMPs serve multiple functions in many cell and tissue types including proliferation, apoptosis, differentiation, chemotaxis, angiogenesis and matrix production during embryogenesis (Luo et al., 2004). The Skor1 locus was poised in MT and repressed in WT (Figure 2.3A) suggesting that Pitx2 represses the Skor1-mediated repression cascade involved in specification of the osteogenic lineage during development.

Folistatin like 5, Fstl5, a calcium binding protein, is involved in axonal guidance of DRG neurons during development (Masuda et al., 2009). The H3K4me3 occupancy in the Fstl5 locus was exhibited only in the MT (Figure 2.3F), suggesting that Pitx2 represses Fstl5 activity in myoblasts as they migrate along with peripheral neuronal axons.

Genes involved in cell growth, organogenesis, lineage specification and cytoskeleton organization were characterized with inactive chromatin domains in the absence of Pitx2. Genes involved in axonal pathfinding and muscle differentiation were characterized with active regulatory chromatin domains in the absence of Pitx2. Pitx2 might regulate chromatin regions involved in different areas of myoblast integrity, specification and muscle formation. Chromatin signatures indicated a complex gene network required at each developmental stage as several lineages intermingle to form a functional organ.

Discussion

Genome wide association studies like chromatin profiling are useful tools to map regulatory activity in the genome, and when combined with gene expression data, a much more precise model of regulatory networks can be developed. In this study, we combined gene expression data from lineage-specific *Lbx1*$^{EGFP/+}$ E12 mouse forelimb myoblasts with ChIP-seq data of histone modifications in E12 forelimbs to identify regulatory sequences influenced by Pitx2. Several transcripts were differentially expressed in WT and MT migratory myoblasts using new packages in R to re-analyze our previous micro-array data (Campbell et al., 2012b). Many of the identified transcripts with known functions play a role in the development of nervous system and/or cytoskeletal organization of the myofibers, further supporting previous observations (Campbell et al., 2012a, 2012b) that Pitx2 regulates higher order architectural assembly of the developing muscle.

The chromatin states around *Pitx2*-regulated transcripts exhibited a two-fold greater frequency of trivalent poised target transcripts in MT relative to WT (13% vs. 7%). This increase in inactive but poised target transcripts could imply that ablation of *Pitx2* at this stage in the embryonic forelimb created a time-shift,“stalled cells”, during development. Transcripts that are normally “repressed” were switched to “poised” due to deposition of excessive H3K4me3, thus priming them for later activation. The temporal shift is further supported by the 22% increase in transcripts marked solely with H3K4me3 in MT.

Even though the Pitx2 targeted transcripts were identified from Lbx1^{+} lineage myoblasts, changes in the chromatin state of many target transcripts in the whole forelimb were also identified by MACS2. The whole forelimbs contain vascular, bone, and other cell types, which generate noise in the data sets, likely to be responsible for disagreements between the expression profiling data and chromatin state. Additionally, Homer tools was used to create tag-density plots around different subsets of transcripts in the E12 mouse genome. The average H3K4me3 signal was decreased in both WT and MT around all target transcripts relative to all transcripts, especially the SSTFs, suggesting that at this developmental stage the target transcripts are less active than all other transcripts, on average. This observation was supported by the exact same trend in the Pol II occupancy data, and the large increase in H3K27me3 signal density. In general, the WT tag density was greater than the MT tag density despite normalization per million reads. This

persisted with the target transcripts for both the SSTFs and non-SSTFs data. An increase in H3K4me3 tag density around target transcripts in MT was expected, but the opposite was observed. Because this pattern is seen in every case, it is most likely an artifact of the normalization method. Homer tools does not have the option to generate tag-density plots relative to the input controls, which further exacerbates the difference. Increased WT signal strength is not significantly shown in the bigwig visualization with the UCSC genome browser, which takes the fold enrichment over input controls into account.

Conclusion

Genes regulated by Pitx2 in mouse embryonic forelimb are involved in neuronal, muscular and bone development. Pitx2 promotes myogenesis by extending the myoblast state while maintaining the availability of genes involved in axon pathfinding and osteogenesis in a poised state.

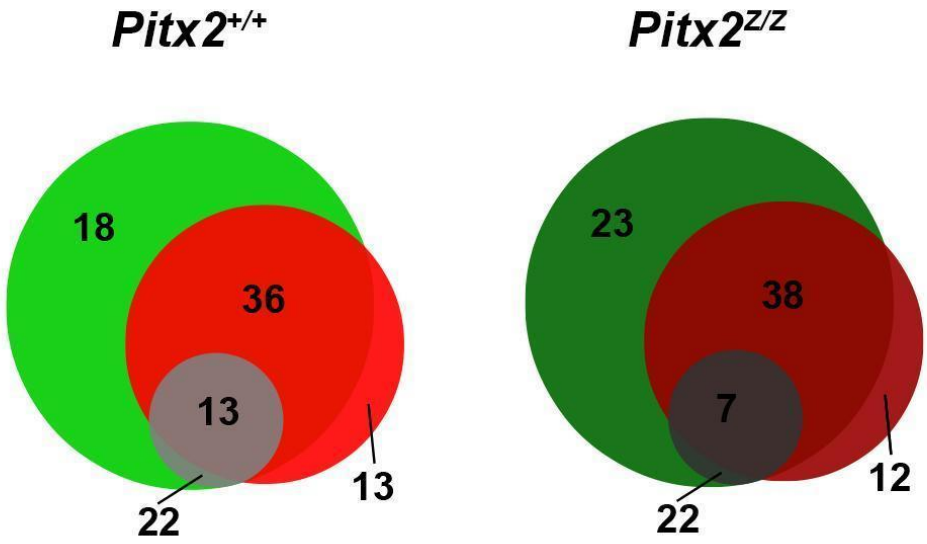


Figure 2.1 Chromatin state changes in Pitx2 Mutants
Venn diagrams illustrating the co-occurrence of enriched regions around identified Pitx2 target sequences in WT and *Pitx2* MT forelimbs, respectively. The MT data are represented by darker shades. Green represents H3K4me3; red represents H3K27me3; and gray represents Pol II in each sample. Homer v4.7 was used to match each identified target sequence to its nearest peaks.

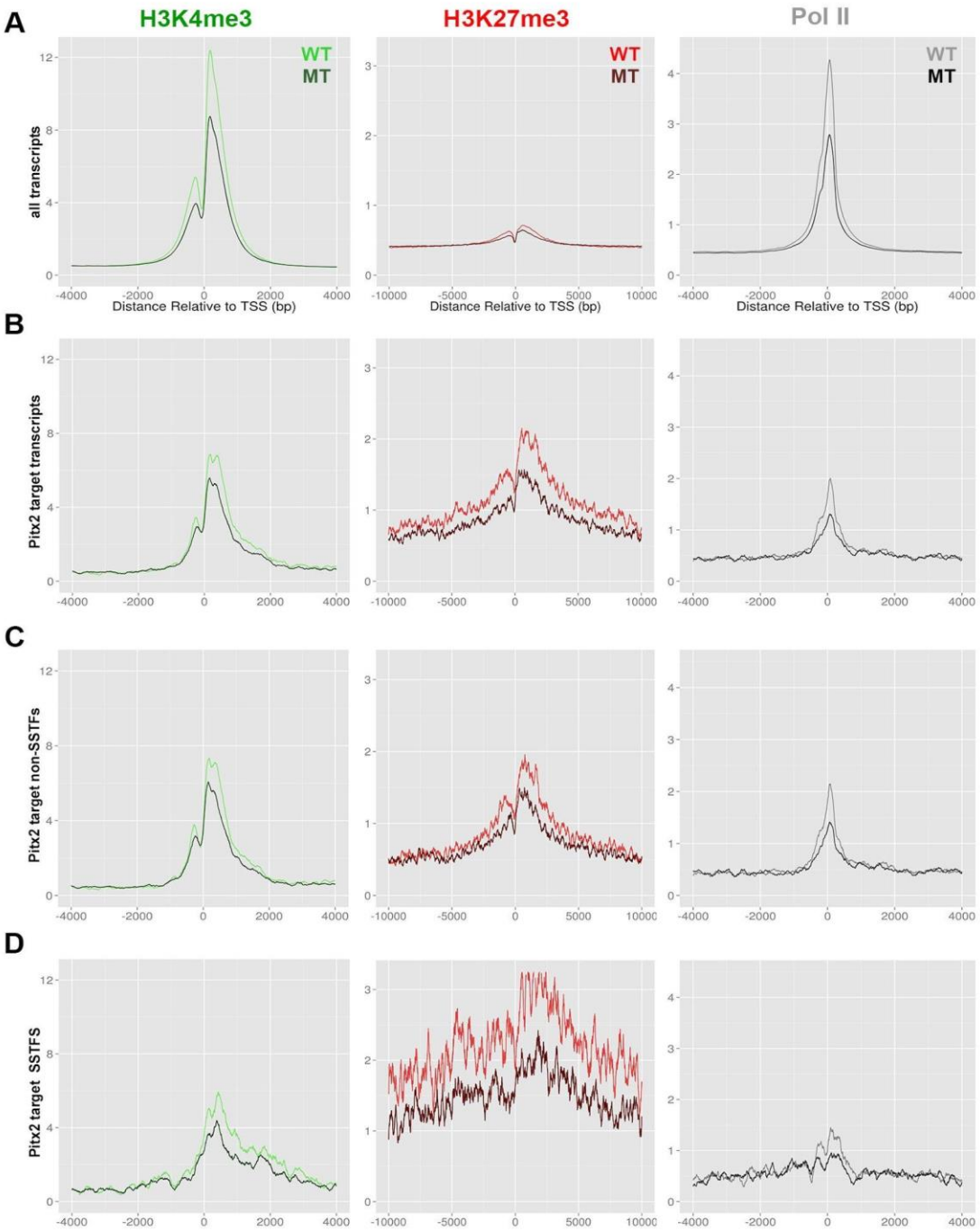

A
H3K4me3
H3K27me3
Pol II
WT
MT
all transcripts
Distance Relative to TSS (bp)
B
Pitx2 target transcripts
C
Pitx2 target non-SSTFs
D
Pitx2 target SSTFS

Figure 2.2

Figure 2.2 Constant tag density in Wild-Type and Pitx2 Mutants
(A) Normalized tag densities around each known Refseq TSS for all three marks in WT and MT. Homer v4.7 annotatePeaks.pl was used to count the number of aligned reads around Refseq TSSs by scanning in 5bp bins. The average number of reads per bin per sequence is plotted on the y axis against the distance of the bin from the TSS.
(B) Normalized tag densities around only the 125 Pitx2 target sequences identified.
(C) Normalized tag densities around only the subset of SSTFs from (B).
(D) Tag densities around all sequences not in (C).

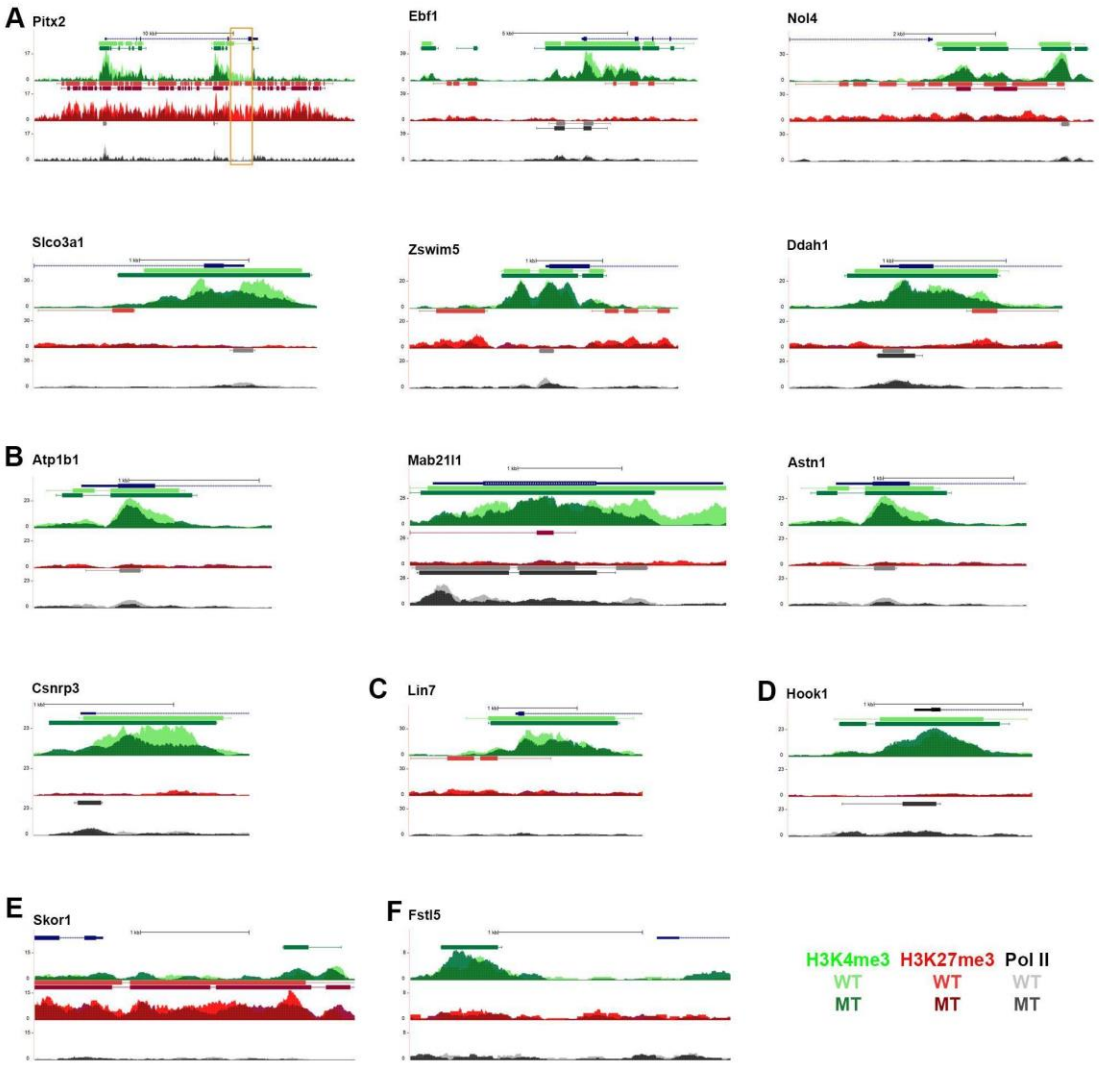

A
Pitx2
Ebf1
Nol4
Slco3a1
Zswim5
Ddah1
B
Atp1b1
Mab21l1
Astn1
Csnrp3
C
Lin7
D
Hook1
E
Skor1
F
Fstl5
H3K4me3
WT
MT
H3K27me3
WT
MT
Pol II
WT
MT

Figure 2.3

Figure 2.3. Pitx2-dependent chromatin state of mouse embryonic forelimbs
UCSC genome browser visualization of selected gene loci. Genes were selected based on strong H3K27me3 marked chromatin, and/or high relative fold changes determined from the GSE31945 array data. WT and MT tracks of the same mark are overlaid. H3K4me3 is represented by green, H3K27me3 red, and Pol II gray. MT tracks are illustrated by the darker shades.
(A) Trivalent, poised genes H3K4me3/H3K27me3/Pol II.
(B) Activated genes (H3K4me3/Pol II).
(C) Bivalent genes (H3K4me3/H3K27me3).
(D) Genes with open chromatin (H3K4me3).
(E) Genes with close chromatin (H3K27me3).
(F) Genes with no significant chromatin signature.

Table 2.1 Pitx2 Target Genes in Migratory Myoblasts

Gene name	Gene Symbol	Adjusted P. Val	FC MT/WT	Function	Chr	Chromatin State WT	Chromatin State MT
Paired-like homeodomain transcription factor 2	**Pitx2**	**6.21E-105**	**-29.06**	**organogenesis, cell cycle**	**3**	**K4|K27|Pol II**	**K4|K27**
T-box 1	Tbx1	1.72E-04	-2.17	organogenesis	16	K4|K27	K4|K27
LSM12 homolog (S. cerevisiae)	Lsm12	3.96E-04	-2.11	RNA processing	11	K4|Pol II	K4|Pol II
Fibroblast growth factor 15	Fgf15	3.92E-03	-1.94	organogenesis	7	K4|K27	K4|K27
Phosphatidylinositol 4-kinase, catalytic, beta polypeptide	Pi4kb	3.76E-03	-1.94	phospholipid metabolism	3	K4|Pol II	K4|Pol II
SRY-box containing gene 6	Sox6	3.92E-03	-1.94	cartilage development	7	NA	NA
Zinc finger protein 64	Zfp64	4.88E-03	-1.92	promotes osteogenesis	2	K4|Pol II	K4|Pol II
TRAF3 interacting protein 3	Traf3ip3	4.65E-03	-1.92	autophagy	1	NA	NA
DPH5 homolog (S. cerevisiae)	Dph5	5.00E-03	-1.91		3	K4|Pol II	K4|Pol II
Golgi transport 1 homolog B (S. cerevisiae)	Golt1b	1.12E-02	-1.85	vesicle transport	6	K4|Pol II	K4|Pol II
Polyadenylate binding protein-interacting protein 1	Paip1	1.12E-02	-1.85		13	K4|Pol II	K4|Pol II
Rho GTPase activating protein 29	Arhgap29	1.12E-02	-1.85	tubulogenesis	3	K4|Pol II	K4|Pol II
Potassium large conductance calcium-activated Channel, subfamily M, alpha member 1	Kcnma1	1.46E-02	-1.83	synaptic transmission,	14	K4|K27	K4|K27
Coiled-coil domain containing 141	Ccdc141	1.88E-02	-1.81	neural induction	2	K4	K4
Kelch-like 4 (Drosophila)	Klhl4	2.03E-02	-1.81		X	NA	NA
Cathepsin S	Ctss	3.05E-02	-1.77	endopeptidase activity	3	NA	NA
Endonuclease domain containing 1	Endod1	4.03E-02	-1.75	apoptosis – KEGG	9	K4|Pol II	K4|Pol II
Sushi-repeat-containing protein, X-linked 2	Srpx2	3.98E-02	-1.75	synaptogenesis	X	NA	NA
Actin, alpha 2	**Acta2**	**4.79E-02**	**-1.73**	**vasculogenesis**	**19**	**K4|Pol II**	**NA**
Regulatory factor X 8	**Rfx8**	**5.16E-02**	**-1.72**	**pancreas development**	**1**	**K4|Pol II**	**K4**
Reprimo	Rprm	5.68E-02	-1.71	cell cycle	2	K4	K4
Actinin alpha 3	Actn3	5.97E-02	-1.71	muscle contraction	19	K4|Pol II	K4|Pol II
Lymphatic vessel endothelial hyaluronan receptor 1	Lyve1	6.30E-02	-1.71	polysaccharide catabolism	7	NA	NA
Hydrolethalus syndrome 1	Hyls1	7.32E-02	-1.69	ciliogenesis	9	K4|Pol II	K4|Pol II
Complement component 3a receptor 1	C3ar1	8.99E-02	-1.67	locomotion	6	NA	NA
Myomesin 2	Myom2	9.36E-02	-1.67	muscle contraction	8	NA	NA
Zinc finger, SWIM domain containing 5	**Zswim5**	**9.68E-02**	**1.67**	**CNS**	**4**	**K4|K27|PolII**	**K4**
Lectin, galactoside binding-like	Lgalsl	9.02E-02	1.67	carbohydrate binding	11	K4|Pol II	K4|Pol II
Astrotactin 1	**Astn1**	**8.15E-02**	**1.68**	**neuron migration/adhesion**	**1**	**K4|Pol II**	**K4**
Homeo box D4	Hoxd4	8.57E-02	1.68	pattern specification	2	K4|K27	K4|K27
Lipid Phosphate Phosphatase-Related Protein Type 1	Lppr1	7.43E-02	1.69	neurite outgrowth	4	K4	K4
Bruno-like 5, RNA binding protein (Drosophila)	Celf5	7.43E-02	1.69	myotonic dystrophy	10	K4|K27	K4|K27
Dimethylarginine dimethylaminohydrolase 1	**Ddah1**	**7.48E-02**	**1.69**	**amino acid transport**	**3**	**K4|K27|PolII**	**K4|PolII**
RIKEN cDNA 4933400N17 gene	Rbfox1	7.72E-02	1.69	myotonic dystrophy	16	K4	K4
Fatty acid binding protein 7, brain	Fabp7	7.32E-02	1.69	Neuromuscular process	10	NA	NA
Family with sequence similarity 57, member B	Fam57b	7.32E-02	1.69	inhibition of adipogenesis	7	NA	NA
Transcription factor AP-2, alpha	Tfap2a	6.96E-02	1.7	direct regulator of Bmp2, 4	13	K4|K27|PolII	K4|K27|Pol II
Solute carrier family 32 member 1	Slc32a1	5.94E-02	1.71	neurotransmitter transport	2	K27	K27
Myosin, heavy polypeptide 7, cardiac muscle, beta	Myh7	5.97E-02	1.71	muscle contraction	14	NA	NA
Coxsackie virus and adenovirus receptor	Cxadr	4.79E-02	1.73	cell junction, myogenesis	16	K4|Pol II	K4|Pol II
Fasciculation and elongation protein zeta 1 (zygin I)	Fez1	4.77E-02	1.73	microtubule organizing	9	NA	NA
Peroxisome proliferative activated receptor, gamma	Ppargc1a	4.35E-02	1.74	insulin signaling – KEGG	5	K4	K4
CUGBP, Elav-like family member 4	Celf4	4.51E-02	1.74	spliceosome assembly	18	K4|K27	K4|K27
Hook homolog 1 (Drosophila)	**Hook1**	**4.13E-02**	**1.74**	**cytoskeleton**	**4**	**K4**	**K4|Pol II**
Microtubule-associated protein 2	Map2	4.35E-02	1.75	axonogenesis	1	K4	K4
RAS oncogene family member 3C	Rab3c	3.67E-02	1.75	exocytosis	13	K4	K4
Follistatin-like 5	**Fstl5**	**3.45E-02**	**1.76**	**muscle growth**	**3**	**NA**	**K4**
Homeo box D3 Opposite Strand	Hoxd3os1	3.36E-02	1.76		2	K4|K27	K4|K27
Early B-cell factor 1	**Ebf1**	**3.52E-02**	**1.76**	**adipogenesis**	**11**	**K4|K27|Pol II**	**K4|Pol II**
Calcium channel, voltage-dependent, alpha 2/delta Subunit 2; similar to Cacna2d2 protein	Cacna2d2	3.13E-02	1.77	muscle synaptic transmission	9	K4|K27	K4|K27
Transcription factor 21	Tcf21	3.11E-02	1.77	organogenesis	10	K27	K27
Homeo box B5 Opposite Strand	Hoxb5os	3.10E-02	1.77		11	K4|K27	K4|K27
Neural proliferation, differentiation and control gene 1	Npdc1	3.05E-02	1.77		2	K4|K27	K4|K27
Homeo box B2	Hoxb2	3.05E-02	1.77	pattern specification	11	K4|K27|PolII	K4|K27|PolII
Cysteine-serine-rich nuclear protein 3	**Csrnp3**	**3.09E-02**	**1.77**	**apoptosis**	**2**	**K4**	**K4|Pol II**
Solute carrier organic anion transporter 3a1	**Slco3a1**	**2.82E-02**	**1.78**	**organic anion transport**	**7**	**K4|K27|PolII**	**K4**
Runt-related transcription factor 1	Runx1t1	2.86E-02	1.78	fat cell differentiation	4	K4|Pol II	K4|Pol II
Basonuclin 1	Bnc1	2.41E-02	1.79		7	K4|K27	K4|K27
Meis homeobox 2	Meis2	2.58E-02	1.79	organogenesis	2	K4|K27|PolII	K4|K27|PolII
Mab-21-like 1 (C. elegans)	**Mab21l1**	**1.89E-02**	**1.81**	**neurons**	**3**	**K4|Pol II**	**K4|K27|PolII**
Protocadherin 8	**Pcdh8**	1.52E-02	1.83	cell adhesion, muscle	14	K4|K27	K4|K27
Neurexin III	Nrxn3	1.26E-02	1.84	synapse organization	12	NA	NA
lin-7 homolog A (C. elegans)	**Lin7a**	**1.24E-02**	**1.84**	**synaptic transmission**	**10**	**K4|K27**	**K4**
Synaptosomal-associated protein 91	Snap91	1.12E-02	1.85	immune system	9	K4|K27	K4|K27
Mab-21-like 2 (C. elegans)	Mab21l2	1.17E-02	1.85	neurons	3	K4|K27|Pol II	K4|K27|Pol II
Nucleolar protein 4	**Nol4**	**1.12E-02**	**1.85**	**organelle**	**18**	**K4|K27|Pol II**	**K4|K27**
Guanine nucleotide binding protein, gamma 3	Gng3	1.21E-02	1.85		19	NA	NA
Paired box gene 3	Pax3	2.23E-02	1.86		1	K4|K27	K4|K27
Doublecortin	**Dcx**	**3.45E-02**	**1.88**	**CNS development**	**X**	**K4**	**NA**

Bold genes represent the ones with different chromatin marks in WT and MT biopsies; K4, H3K4me3; K27, H3K27me3

Table 2.1 Continued

Pitx2 Target Genes in Migratory Myoblasts

Gene name	Gene Symbol	Adjusted P. Val	FC MT/WT	Function	Chr	Chromatin State WT	Chromatin State MT
Synaptosomal-associated protein 25	**Snap25**	**7.82E-03**	**1.88**	**neurotransmitter secretion**	**2**	**NA**	**K4**
Neuron-glia-CAM-related cell adhesion molecule	**Nrcam**	**8.46E-03**	**1.88**	**axon guidance, adhesion**	**12**	**K4**	**K4|K27**
RNA binding protein, fox-1 homolog 3	Rbfox3	8.00E-03	1.88		11	K4|K27	K4|K27
neurexin I	Nrxn1	8.46E-03	1.88	synaptogenesis	17	NA	NA
$Ca2^{+}$-dependent secretion activator	**Cadps**	**6.36E-03**	**1.9**	**catecholamine secretion**	**14**	**K4|K27**	**K4**
Inter-alpha (globulin) inhibitor H5	Itih5	4.65E-03	1.92	polysacharide metabolism	2	K4	K4
Solute carrier family 1 (glial high affinity glutamate transporter), member 2	Slc1a2	4.00E-03	1.93	glucose transport	2	K4|K27	K4|K27
SOGA family member 3	Soga3	3.76E-03	1.94		10	K4|K27	K4|K27
ATPase, Na^{+}/K^{+} transporting, beta 1 polypeptide	**Atp1b1**	**3.92E-03**	**1.94**	**ATP biosynthesis**	**1**	**K4|K27|Pol II**	**K4**
RIKEN cDNA C130071C03 gene	C130071C03Rik	1.48E-02	1.95		13	K27	K27
Contactin 2	**Cntn2**	**3.42E-03**	**1.95**	**neurons, cytoskeleton**	**1**	**K27**	**K4**
Cerebellin 2 precursor protein	Cbln2	2.87E-03	1.96		18	K4|K27	K4|K27
Contactin 1	Cntn1	2.74E-03	1.97	cell adhesion	15	K4	K4
Prior incubation determinant 1	Pid1	2.56E-03	1.97		1	K4	K4
Homeo box C8	Hoxc8	2.63E-03	1.97	pattern specification process	15	K4|K27	K4|K27
Claudin 9	Cldn9	2.28E-03	1.98	cell adhesion	17	NA	NA
UNC homeobox	Uncx	2.18E-03	1.98	cartilage condensation	5	K27	K27
Solute carrier family 17, member 6	Slc17a6	1.18E-03	2.02	synaptic transmission	7	K27	K27
Suppression of tumorigenicity 18	St18	1.18E-03	2.02		1	NA	NA
Kinesin family member 5C	Kif5c	1.18E-03	2.03	axonogenesis, guidance	2	K4|Pol II	K4|Pol II
Myocardial infarction associated transcript	Miat	1.05E-03	2.04	neuron differentiation	5	K4|K27	K4|K27
Myelin transcription factor 1	Myt1	9.91E-04	2.04	cell cycle	2	NA	NA
Secretogranin III	Scg3	8.78E-04	2.05		9	NA	NA
Homeo box B3	Hoxb3	1.87E-02	2.06	pattern specification	11	K4|K27	K4|K27
Fibroblast growth factor 5	Fgf5	7.94E-04	2.06	organogenesis	5	K4|K27	K4|K27
RUN and FYVE domain containing 3	Rufy3	1.20E-02	2.06	chromosome partitioning,	5	K4|Pol II	K4|Pol II
Hydroxyprostaglandin dehydrogenase 15 (NAD)	Hpgd	5.01E-04	2.09	fatty acid metabolic process	8	K4	K4
Neurogenic differentiation 6	Neurod6	4.83E-04	2.09	neurogenesis	6	NA	NA
Sorbin and SH3 domain containing 2	Sorbs2	4.34E-04	2.1		8	K4	K4
ELAV-like 2 (Hu antigen B)	Elavl2	1.44E-03	2.12		4	K4|K27|Pol II	K4|K27|Pol II/K27
Glutamic acid decarboxylase 1	Gad1	2.25E-04	2.15	synaptic transmission	2	K27	K27
Homeo box D3	Hoxd3	8.17E-05	2.21	pattern specification	2	K4|K27	K4|K27
Zinc finger protein of the cerebellum 1	Zic1	4.00E-04	2.22	pattern specification	9	K4|K27	K4|K27
ELAV-like 4 (Hu antigen D)	Elavl4	3.37E-05	2.31		4	K4|Pol II	K4|Pol II
Neuron navigator 2	Nav2	2.84E-06	2.43		7	K4|K27|PolII	K4|K27|PolII
Inhibitor of DNA binding 4	Id4	3.32E-02	2.44	neuroblast proliferation	13	K4|Pol II	K4|Pol II
LIM homeobox protein 1	Lhx1	2.85E-02	2.5	neuron specification	11	K4|K27	K4|K27
Stathmin-like 3	Stmn3	5.10E-04	2.52	cytoskeleton organization	2	K27	K27
Lysosomal-associated membrane protein family, 5	**Lamp5**	**7.03E-07**	**2.52**		**2**	**K4|K27**	**K27**
Collapsin response mediator protein 1	Crmp1	7.03E-07	2.52	neuron, cytoskeleton	5	K4|K27	K4|K27
Glycoprotein m6a	Gpm6a	1.54E-04	2.61		8	K4	K4
Cell adhesion molecule with homology to L1CAM	Chl1	5.46E-08	2.68	axonogenesis, guidance	6	K4|K27	K4|K27
Thrombospondin, type I, domain containing 7A	Thsd7a	1.35E-09	2.92		6	K4|Pol II	K4|Pol II
Roundabout homolog 3 (Drosophila)	Robo3	8.22E-10	2.96	neuron migration	9	K27	K27
Neurocan	Ncan	2.15E-06	3.06	adhesion, synapsis	8	K27	K27
Tubulin, beta 3; tubulin, beta 3, pseudogene 1	**Tubb3**	**9.40E-12**	**3.25**	**axon guidance**	**8**	**K4**	**K4|K27**
Reticulon 1	Rtn1	3.52E-12	3.32	inhibition of axonal growth	12	K4|K27	K4|K27
Nescient helix loop helix 2	Nhlh2	1.75E-12	3.36	physical activity behavior	3	K4|K27	K4|K27
Insulin-like growth factor binding protein-like 1	Igfbpl1	1.82E-13	3.52	regulation of cell growth	4	K4|K27	K4|K27
Homeo box B6	Hoxb6	3.63E-14	3.63	pattern specification process	11	K4|K27	K4|K27
Homeo box B9	Hoxb9	1.73E-19	4.45	pattern specification process	11	K27	K27
Stathmin-like 2	**Stmn2**	**2.12E-13**	**4.49**	**cytoskeleton organization**	**3**	**K27**	**NA**
Microtubule-associated protein tau	Mapt	1.12E-19	4.49	microtubule organization	11	K4|K27|Pol II	K4|K27|Pol II
Internexin neuronal intermediate filament protein, alpha	Ina	9.15E-14	4.93	cytoskeleton organization	19	K4|K27	K4|K27
Deleted in colorectal carcinoma	Dcc	8.20E-21	5.7	neuron migration, apoptosis	18	K4|K27	K4|K27
SKI family transcriptional co-repressor 1	**Skor1**	**3.77E-52**	**10.96**	**BMP inhibitor**	**9**	**K27**	**K4|K27**

Bold genes represent the ones with different chromatin marks in WT and MT biopsies; K4, H3K4me3; K27, H3K27me3

Supplementary Information

Table 2.S1 Data Sets Used for Analysis

GSE	GSM	Tissue, E12.5	Mouse Line	Assay	Sample	Reference
GSE31945	GSM791677	GFP^+ flow-sorted FL	$Lbx1^{EGFP}$	Microarray	RNA	Campbell, et al 2012
GSE31945	GSM791678	GFP^+ flow-sorted FL	$Lbx1^{EGFP}$	Microarray	RNA	Campbell, et al 2012
GSE31945	GSM791683	GFP^+ flow-sorted FL	$Lbx1^{EGFP}$|$Pitx2^{Z/Z}$	Microarray	RNA	Campbell, et al 2012
GSE31945	GSM791684	GFP^+ flow-sorted FL	$Lbx1^{EGFP}$|$Pitx2^{Z/Z}$	Microarray	RNA	Campbell, et al 2012
GSE49010	GSM1192095	Whole FL	ICR	ChIP-seq	H3K4me3	Eng et al., 2014
GSE49010	GSM1192096	Whole FL	ICR	ChIP-seq	H3K27me3	Eng et al., 2014
GSE49010	GSM1192097	Whole FL	ICR	ChIP-seq	IGG-Input	Eng et al., 2014
GSE49010	GSM1192099	Whole FL	ICR	ChIP-seq	RNA Pol II	Eng et al., 2014
GSE71128	GSM1827869	Whole FL	$Pitx2^{Z/Z}$	ChIP-seq	IGG-Input	Eng et al., 2014
GSE71128	GSM1827875	Whole FL	$Pitx2^{Z/Z}$	ChIP-seq	H3K4me3	Eng et al., 2014
GSE71128	GSM1827877	Whole FL	$Pitx2^{Z/Z}$	ChIP-seq	H3K27me3	Eng et al., 2014
GSE71128	GSM1827880	Whole FL	$Pitx2^{Z/Z}$	ChIP-seq	RNA Pol II	Eng et al., 2014

References

Adams, N.C., Tomoda, T., Cooper, M., Dietz, G., and Hatten, M.E. (2002). Mice that lack astrotactin have slowed neuronal migration. Development *129*, 965–972.

Amthor, H., Christ, B., and Patel, K. (1999). A molecular mechanism enabling continuous embryonic muscle growth - a balance between proliferation and differentiation. Development *126*, 1041–1053.

Arber, S., Hunter, J.J., Ross, J., Jr, Hongo, M., Sansig, G., Borg, J., Perriard, J.C., Chien, K.R., and Caroni, P. (1997). MLP-deficient mice exhibit a disruption of cardiac cytoarchitectural organization, dilated cardiomyopathy, and heart failure. Cell *88*, 393–403.

Arndt, S., Poser, I., Moser, M., and Bosserhoff, A.-K. (2007). Fussel-15, a novel Ski/Sno homolog protein, antagonizes BMP signaling. Mol. Cell. Neurosci. *34*, 603–611.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823–837.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315–326.

Birney, E. (2012). The making of ENCODE: Lessons for big-data projects. Nature *489*, 489049a.

Bohl, J., Brimer, N., Lyons, C., and Vande Pol, S.B. (2007). The stardust family protein MPP7 forms a tripartite complex with LIN7 and DLG1 that regulates the stability and localization of DLG1 to cell junctions. J. Biol. Chem. *282*, 9392–9400.

Breckenridge, R.A., Kelly, P., Nandi, M., Vallance, P.J., Ohun, T.J., and Leiper, J. (2010). A role for Dimethylarginine Dimethylaminohydrolase 1 (DDAH1) in mammalian development. Int. J. Dev. Biol. *54*, 215–220.

Campbell, A.L., Shih, H.-P., Xu, J., Gross, M.K., and Kioussi, C. (2012a). Regulation of motility of myogenic cells in filling limb muscle anlagen by Pitx2. PLoS One *7*, e35822.

Campbell, A.L., Eng, D., Gross, M.K., and Kioussi, C. (2012b). Prediction of gene network models in limb muscle precursors. Gene *509*, 16–23.

Carvajal, J.J., Keith, A., and Rigby, P.W.J. (2008). Global transcriptional regulation of the locus encoding the skeletal muscle determination genes Mrf4 and Myf5. Genes Dev. *22*, 265–276.

Cossu, G., Kelly, R., Tajbakhsh, S., Di Donna, S., Vivarelli, E., and Buckingham, M. (1996). Activation of different myogenic pathways: myf-5 is induced by the neural tube and MyoD by the dorsal ectoderm in mouse paraxial mesoderm. Development *122*, 429–437.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.-H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., et al. (2002). A genomic regulatory network for development. Science *295*, 1669–1678.

Davis, S., and Meltzer, P.S. (2007). GEOquery: a bridge between the Gene Expression Omnibus

(GEO) and BioConductor. Bioinformatics *23*, 1846–1847.

Eng, D., Ma, H.-Y., Xu, J., Shih, H.-P., Gross, M.K., Kioussi, C., and Kiouss, C. (2012). Loss of abdominal muscle in Pitx2 mutants associated with altered axial specification of lateral plate mesoderm. PLoS One *7*, e42228.

Eng, D., Vogel, W.K., Flann, N.S., Gross, M.K., and Kioussi, C. (2014). Genome-Wide Mapping of Chromatin State of Mouse Forelimbs. Open Access Bioinformatics *6*, 1–11.

Gage, P.J., Suh, H., and Camper, S.A. (1999). Dosage requirement of Pitx2 for development of multiple organs. Development *126*, 4643–4651.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. *5*, R80.

Goulding, M., Lumsden, A., and Paquette, A.J. (1994). Regulation of Pax-3 expression in the dermomyotome and its role in muscle development. Development *120*, 957–971.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77–88.

Gustafsson, M.K., Pan, H., Pinney, D.F., Liu, Y., Lewandowski, A., Epstein, D.J., and Emerson, C.P., Jr (2002). Myf5 is a direct target of long-range Shh signaling and Gli regulation for muscle specification. Genes Dev. *16*, 114–126.

Hagman, J., Belanger, C., Travis, A., Turck, C.W., and Grosschedl, R. (1993). Cloning and functional characterization of early B-cell factor, a regulator of lymphocyte-specific gene expression. Genes Dev. *7*, 760–773.

Heanue, T.A., and Pachnis, V. (2006). Expression profiling the developing mammalian enteric nervous system identifies marker and candidate Hirschsprung disease genes. Proc. Natl. Acad. Sci. U. S. A. *103*, 6919–6924.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Jimenez, M.A., Akerblad, P., Sigvardsson, M., and Rosen, E.D. (2007). Critical role for Ebf1 and Ebf2 in the adipogenic transcriptional cascade. Mol. Cell. Biol. *27*, 743–757.

Jostes, B., Walther, C., and Gruss, P. (1990). The murine paired box gene, Pax7, is expressed specifically during the development of the nervous and muscular system. Mech. Dev. *33*, 27–37.

Karolchik, D., Baertsch, R., Diekhans, M., Furey, T.S., Hinrichs, A., Lu, Y.T., Roskin, K.M., Schwartz, M., Sugnet, C.W., Thomas, D.J., et al. (2003). The UCSC Genome Browser Database. Nucleic Acids Res. *31*, 51–54.

Kiefer, J.C., and Hauschka, S.D. (2001). Myf-5 is transiently expressed in nonmuscle mesoderm and exhibits dynamic regional changes within the presegmented mesoderm and somites I-IV.

Dev. Biol. *232*, 77–90.

Kioussi, C., Briata, P., Baek, S.H., Rose, D.W., Hamblet, N.S., Herman, T., Ohgi, K.A., Lin, C., Gleiberman, A., Wang, J., et al. (2002). Identification of a Wnt/Dvl/beta-Catenin --> Pitx2 pathway mediating cell-type-specific proliferation during development. Cell *111*, 673–685.

Kitamura, K., Miura, H., Miyagawa-Tomita, S., Yanazawa, M., Katoh-Fukui, Y., Suzuki, R., Ohuchi, H., Suehiro, A., Motegi, Y., Nakahara, Y., et al. (1999). Mouse Pitx2 deficiency leads to anomalies of the ventral body wall, heart, extra- and periocular mesoderm and right pulmonary isomerism. Development *126*, 5749–5758.

Kong, Y., Flick, M.J., Kudla, A.J., and Konieczny, S.F. (1997). Muscle LIM protein promotes myogenesis by enhancing the activity of MyoD. Mol. Cell. Biol. *17*, 4750–4760.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Lin, C.R., Kioussi, C., O'Connell, S., Briata, P., Szeto, D., Liu, F., Izpisúa-Belmonte, J.C., and Rosenfeld, M.G. (1999). Pitx2 regulates lung asymmetry, cardiac positioning and pituitary and tooth morphogenesis. Nature *401*, 279–282.

Lu, M.F., Pressman, C., Dyer, R., Johnson, R.L., and Martin, J.F. (1999). Function of Rieger syndrome gene in left-right asymmetry and craniofacial development. Nature *401*, 276–278.

Luo, Q., Kang, Q., Si, W., Jiang, W., Park, J.K., Peng, Y., Li, X., Luu, H.H., Luo, J., Montag, A.G., et al. (2004). Connective tissue growth factor (CTGF) is regulated by Wnt and bone morphogenetic proteins signaling in osteoblast differentiation of mesenchymal stem cells. J. Biol. Chem. *279*, 55958–55968.

Maldonado-Báez, L., Cole, N.B., Krämer, H., and Donaldson, J.G. (2013). Microtubule-dependent endosomal sorting of clathrin-independent cargo by Hook1. J. Cell Biol. *201*, 233–247.

Mariani, M., Baldessari, D., Francisconi, S., Viggiano, L., Rocchi, M., Zappavigna, V., Malgaretti, N., and Consalez, G.G. (1999). Two murine and human homologs of mab-21, a cell fate determination gene involved in Caenorhabditis elegans neural development. Hum. Mol. Genet. *8*, 2397–2406.

Masuda, T., Yaginuma, H., Sakuma, C., and Ono, K. (2009). Netrin-1 signaling for sensory axons: Involvement in sensory axonal development and regeneration. Cell Adh. Migr. *3*, 171–173.

Mella, S., Soula, C., Morello, D., Crozatier, M., and Vincent, A. (2004). Expression patterns of the coe/ebf transcription factor genes during chicken and mouse limb development. Gene Expr. Patterns *4*, 537–542.

Mikkelsen, T.S., Wakefield, M.J., Aken, B., Amemiya, C.T., Chang, J.L., Duke, S., Garber, M., Gentles, A.J., Goodstadt, L., Heger, A., et al. (2007a). Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. Nature *447*, nature05805.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., et al. (2007b). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Mizuhara, E., Nakatani, T., Minaki, Y., Sakamoto, Y., and Ono, Y. (2005). Corl1, a novel neuronal lineage-specific transcriptional corepressor for the homeodomain transcription factor Lbx1. J. Biol. Chem. *280*, 3645–3655.

Park, P.J. (2009). ChIP-seq: advantages and challenges of a maturing technology. Nat. Rev. Genet. *10*, 669–680.

Shih, H.P., Gross, M.K., and Kioussi, C. (2007a). Expression pattern of the homeodomain transcription factor Pitx2 during muscle development. Gene Expr. Patterns *7*, 441–451.

Shih, H.P., Gross, M.K., and Kioussi, C. (2007b). Cranial muscle defects of Pitx2 mutants result from specification defects in the first branchial arch. Proc. Natl. Acad. Sci. U. S. A. *104*, 5907–5912.

Shih, H.P., Gross, M.K., and Kioussi, C. (2008). Muscle development: forming the head and trunk muscles. Acta Histochem. *110*, 97–108.

Smyth, G.K. (2005). limma: Linear Models for Microarray Data. In Bioinformatics and Computational Biology Solutions Using R and Bioconductor, (Springer New York), pp. 397–420.

Wong, Y.-M., and Chow, K.L. (2002). Expression of zebrafish mab21 genes marks the differentiating eye, midbrain and neural tube. Mech. Dev. *113*, 149–152.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Fluorescence Activated Cell Sorted Mouse Myoblasts

Chapter 3

Arun J. Singh and Chrissa Kioussi

Formatted for submission to Methods in Molecular Biology

Abstract

Fluorescence activated cell sorting (FACS) is a technology that has been used for more than 40 years to isolate specific cell populations, based on fluorescence or size parameters. In the context of developmental biology, coupling FACS with transgenic mice has been a boon. By choosing specific mouse lines with genetic mutations and perturbations marked with fluorescent proteins, it is possible to isolate hyper-specific populations for further downstream analysis. Here, we present a full method to isolate EGFP-positive myoblasts from embryonic mouse forelimbs.

1 Introduction

Fluorescence activated cell sorting (FACS) is a laboratory technique that has existed since the early 1970's, and is used to isolate specific cell populations based on physiological and fluorescent characteristics (Herzenberg et al., 1976). Single cells in a suspension are fed into the machine where they are isolated as individual droplets and interrogated by a laser. Size and fluorescence parameters are detected by sensors, and an electron gun charges the droplets based on user-defined threshold settings. As the droplets pass by electrically charged plates, they are separated and collected in individual containers that represent isolated cell populations (Herzenberg et al., 1976). Initially FACS was limited to only one color, but recent technological advances have enabled sorting cells with up to 12 (De Rosa et al., 2001), and even 17 colors (Perfetto et al., 2004) by using multiple lasers, sensors, and bandwidth filters.

Initially used to probe the biological relevance of subsets of cells in the blood and different organs (Herzenberg et al., 2002), the functionality and purposes of FACS have greatly expanded. Since its invention, the primary use of FACS has been for mostly immunological purposes. Single cell sorting based on specific parameters has enabled isolation and cloning of rare cell types, including hybridomas for monoclonal antibody production (Parks et al., 1979). Another early application of FACS was to determine the specific ratio of T lymphocyte populations in patient whole blood samples, which enabled a new method of testing for HIV infections (Bofill et al., 1992). Application of multi-color FACS to whole blood samples discovered that hundreds of different, unique cell types exist in the peripheral blood at any time (De Rosa and Roederer, 2001). More recent advances have used fluorescent nanoparticles to sort and purify specific bacterial sub-populations, despite their small size of 1-3 microns (Zahavy et al., 2012).

Outside the field of immunology, FACS has been applied to systems in developmental biology. One of the earliest applications was to isolate primordial germ cells for tissue culture applications from mouse embryos (Abe et al., 1996). Since these cells are viable, they can be used for downstream applications such as gene expression analysis. In this context, FACS has been applied to many different cell types as diverse as mouse neuroprogenitor cells (Abramova et al., 2005), human pancreatic cell (Dorrell et al., 2011), and mouse smooth muscle cells isolated from the colon (Peri et al., 2013). Here, we describe a complete protocol to isolate forelimb myoblasts from transgenic mice expressing extra-green fluorescent protein (EGFP), and extract high-quality RNA.

2 Materials

2.1. Equipment

1. Sony SH800 cell sorter
2. Fluorescent Dissection Microscope
3. Benchtop Centrifuge
4. Dissection tools (forceps, scissors)
5. 1.5 mL eppendorf tube heat block
6. 6 cm tissue culture plates
7. Micropipettes (10µL,100µL,1000µL)
8. Tips

2.2 Supplies

2.3 Media and Reagents

1. E11.5, E12.4, E13.5, E14.5 *Pax3*^{Cre}|*ROSA26*^{EGFP} transgenic mice
2. DMEM/F12 medium
3. HBSS without HEPES
4. 5 mM EDTA
5. 2 mg/mL Collagenase type I [Worthington Biochem] (0.2%)
6. 10 U DNAse I [Worthington Biochem] (added fresh)
7. PBS
8. Ethanol 100% and 70%
9. 6 cm cell culture dishes
10. Ice
11. 1.5 mL tubes
12. 15 mL tubes
13. 12x75 mm polystyrene tubes
14. 30 µm Nitex mesh
15. Glass syringe, or
16. Falcon 5mL round bottom tube with cell strainer cap [Product #352235]

3 Methods

3.1 Embryonic forelimb dissection

1. Dissect mouse embryos at E11.5, E12.5, E13.5, and E14.5.

2. Remove yolk sac, transfer embryos to 6 cm cell culture plate filled with PBS over ice.
3. Genotype embryos via fluorescence microscopy. Pax3 transgenic embryos will fluoresce green in brain, neural tube and somites.
4. Dissect and isolate forelimbs from both transgenic and non-transgenic embryos.
5. Pool all non-transgenic forelimbs by litter in a 1.5 mL tube (see Note 1).
6. Pool transgenic forelimbs by litter in 1.5 mL tubes, with a maximum of 12, 10, 6, and 4 forelimbs per 1 tube from E11.5, 12.5, E13.5, and 14.5, respectively (see Note 2).
7. Centrifuge all samples 3 min @ 2,300xg at 4^{o}C.
8. Remove supernatant from sample tubes.
9. Store samples over ice for a few hours until further processing.

3.2 Embryonic forelimb dissociation

1. Add 750 µL of dissociation buffer to the sample to be dissociated.
2. Place sample in heat block, incubate at 37C for 3 min.
3. Using 1mL tip from 1mL micropipette, gently pipette sample up and down ten times for E11.5 and E12.5 forelimbs, and fifteen times for E14.5 forelimbs (see Note 3)
4. Repeat steps ii and iii once if samples are E11.5, E12.5, and E13.5, and twice if they are E14.5 forelimbs.
5. Centrifuge 1 min @ 5,000 rpm at room temperature.
6. Remove the supernatant, re-suspend cells in 200-400 µL PBS (see Note 4)
7. Pass cells through glass syringe fitted with 35 µm mesh, filter into a 12x75 mm polystyrene tube.
8. Load tube into cell sorter.

3.3 Flow sorting of embryonic myoblasts

1. Load non-fluorescent cell sample into the cell sorter to set gates (see Note 5).
2. Set sample pressure to three, aim for targets events per second (EPS) between 1000-3000 and start the sort (see Note 5).
3. Create a forward scatter area (FSCa) vs side scatter area (SSCa) density plot
4. Adjust gain on FSC and SSC until the visible cell population is as spread out as possible, but still on screen (Figure 1A).

5. Create a gate (R1) containing the whole population minus the debris in the lower left hand corner (Figure 1A).
6. Create a FSCa vs FSC-height (FSCh) density plot gated on A.
7. Create a gate (B) containing only the top cluster of cells (Figure 1b), representing droplets with single cell.s
8. Create a Fluorescence II-area (FL2a) vs FSCa density plot, gated on B (Figure 1C) (see Note 6).
9. Create a FL2a histogram plot (Figure 1D).
10. Stop the sample flow, unload the sample, and load the first fluorescent sample.
11. Set sample pressure to three, aim for targets events per second (EPS) between 1000-3000 and start the sort.
12. Pause the sort between 20k-40k recorded events.
13. Set a gate (C) on the FL2a vs FSCa plot that contains the fluorescent cell population (see Note 7), (Figure 1C)
14. Load a 15 mL tube filled with 200µL PBS into the collection chamber of the cell sorter.
15. Set the sort logic from gate C into the collection tube.
16. Resume the flow and start the sort.
17. Monitor the sample volume, and sort until virtually no sample is left (see Note 8)
18. Stop the sort, unload the sample, then load a 15 mL tube filled with ddH2O.
19. Set sample pressure to ten and start the sample flow, to wash the sample line.
20. When the triggered EPS is consistent and close to zero, stop the flow and unload the water.

3.4 Purity check

1. Load the newly sorted sample into the machine, and run at sample pressure five for a purity check (see Note 9).
2. Stop flow rate after 1000 triggered events and record data as sort purity.

3.5 Spin down cells

1. Centrifuge sample at 3700 rpm for five minutes at 4^{o}C.
2. Remove supernatant and process cells (see Note 10).

4 Notes

1. These non-transgenic forelimbs will be used to set the gain on FSC and SSC. Each stage will have a slightly different gain based on cell size.
2. It is important not to overcrowd the eppendorf tube during the dissociation process. Too many forelimbs leads to excessive density and greatly reduces the efficiency of the dissociation.
3. While pipetting, be careful not to be too forceful and spill the sample
4. Re-suspend cells to a density between 1-10e6 cells/mL quench. A higher cell density run at a lower sample pressure results in a more efficient yield.
5. Events per second (EPS) is based on both the sample flow rate and density. Running at a lower EPS takes longer, but increases efficiency. It should be adjusted based on the needs at the time.
6. This plot will be used to check the extent of autofluorescence in the non-fluorescent population. It is normal for the autofluorescence to shift when samples are changed, but this plot serves as a good starting point.
7. This plot will be used ultimately to determine the population that is sorted and collected. Create a gate that encompasses the fluorescent cell population. There is a trade off of purity and yield that should be adjusted based on the needs of the user.
8. Be sure to monitor the sample closely. If the sample run when nothing is left in the tube air will get into the fluidics system and the sorter alignment will have to be redone.
9. The pressure may need to be adjusted based on the amount of cells sorted. Aim for 1000 events, it is sufficient to determine the purity of the sorted population.
10. If RNA-seq is the goal, cells should be lysed immediately and kept over ice. Extract the RNA from all samples in parallel. Freezing the samples after lysis results in dramatic decrease in RNA quality and yield.

Acknowledgments

We thank Maggie Weitzman at the Genomics and Cell Characterization Core Facility at the University of Oregon for her assistance with the flow sorter, Vera Chang and the LARC personnel

of the Oregon State University for the mouse husbandry. This work was supported by the College of Pharmacy at the Oregon State University and the Oregon State University.

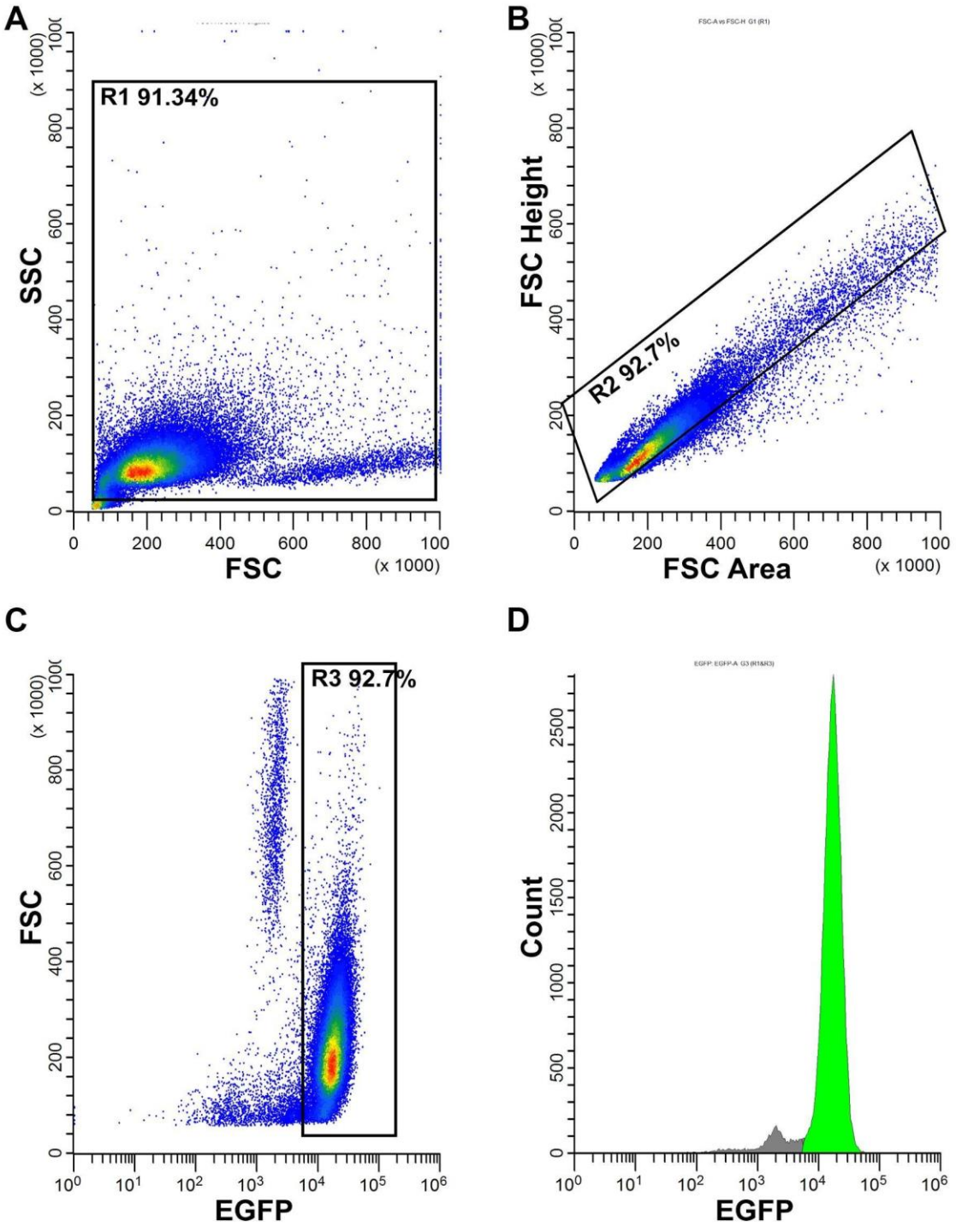

A
R1 91.34%
SSC
(x 1000)
FSC
(x 1000)
B
FSC Height
R2 92.7%
FSC Area
(x 1000)
C
R3 92.7%
FSC
EGFP
D
Count
EGFP

Figure 3.1

Figure 3.1 isolation of mouse embryonic myoblast by FACS
(A) First scatter plot showing all cells from E11.5 forelimbs, with forward scatter (FSC) on the x-axis, and side scatter (SSC) on the y-axis. The bright spot in the bottom left hand corner represents cellular debris and organelles from the dissociation process. Gate R1 covers all non-cellular debris.
(B) Second scatter plot, derived from gate R1, showing FSC-area on the x-axis and FSC-height on the y-axis. Plotting area vs height is used for "doublet discrimination," where droplets that contain two or more cells (doublets) are gated out. Doublets generate a signal with large area relative to height and should be filtered out. Gate R2 selects only singlets.
(C) Third scatter plot, derived from R2, showing EGFP fluorescence area (EGFP) on the x-axis vs FSC on the y-axis. Cells separate into two distinct population based on EGFP intensity. Gate R3 selects only EGFP-positive cells, which are then sorted into collection tubes.
(D) Count histogram derived from R2. EGFP intensity on the x-axis vs cell count on the y-axis. Complementary to **(C)**, it shows distinct cell populations based on fluorescence intensity. Histograms are used primarily as a visualization tool, and not to determine gates.

References

Abe, K., Hashiyama, M., Macgregor, G., and Yamamura, K. i. (1996). Purification of primordial germ cells from TNAPbeta-geo mouse embryos using FACS-gal. Dev. Biol. *180*, 468–472.

Abramova, N., Charniga, C., Goderie, S.K., and Temple, S. (2005). Stage-specific changes in gene expression in acutely isolated mouse CNS progenitor cells. Dev. Biol. *283*, 269–281.

Bofill, M., Janossy, G., Lee, C.A., MacDonald-Burns, D., Phillips, A.N., Sabin, C., Timms, A., Johnson, M.A., and Kernoff, P. (1992). Laboratory control values for CD4 and CD8 T lymphocytes. Implications for HIV-1 diagnosis. Clinical & Experimental Immunology *88*, 243–252.

De Rosa, S.C., and Roederer, M. (2001). Eleven-color flow cytometry. A powerful tool for elucidation of the complex immune system. Clin. Lab. Med. *21*, 697–712, vii.

De Rosa, S.C., Herzenberg, L.A., Herzenberg, L.A., and Roederer, M. (2001). 11-color, 13-parameter flow cytometry: identification of human naive T cells by phenotype, function, and T-cell receptor diversity. Nat. Med. *7*, 245–248.

Dorrell, C., Schug, J., Lin, C.F., Canaday, P.S., Fox, A.J., Smirnova, O., Bonnah, R., Streeter, P.R., Stoeckert, C.J., Jr, Kaestner, K.H., et al. (2011). Transcriptomes of the major human pancreatic cell types. Diabetologia *54*, 2832–2844.

Herzenberg, L.A., Sweet, R.G., and Herzenberg, L.A. (1976). Fluorescence-activated cell sorting. Sci. Am. *234*, 108–117.

Herzenberg, L.A., Parks, D., Sahaf, B., Perez, O., Roederer, M., and Herzenberg, L.A. (2002). The history and future of the fluorescence activated cell sorter and flow cytometry: a view from Stanford. Clin. Chem. *48*, 1819–1827.

Parks, D.R., Bryan, V.M., Oi, V.T., and Herzenberg, L.A. (1979). Antigen-specific identification and cloning of hybridomas with a fluorescence-activated cell sorter. Proc. Natl. Acad. Sci. U. S. A. *76*, 1962–1966.

Perfetto, S.P., Chattopadhyay, P.K., and Roederer, M. (2004). Seventeen-colour flow cytometry: unravelling the immune system. Nat. Rev. Immunol. *4*, 648–655.

Peri, L.E., Sanders, K.M., and Mutafova-Yambolieva, V.N. (2013). Differential expression of genes related to purinergic signaling in smooth muscle cells, PDGFRα-positive cells, and interstitial cells of Cajal in the murine colon. Neurogastroenterol. Motil. *25*, e609–e620.

Zahavy, E., Ber, R., Gur, D., Abramovich, H., Freeman, E., Maoz, S., and Yitzhaki, S. (2012). Application of nanoparticles for the detection and sorting of pathogenic bacteria by flow-cytometry. Adv. Exp. Med. Biol. *733*, 23–36.

Gene Expression Profiling During Embryonic and Fetal Myogenesis

Chapter 4

Arun J. Singh, Chih-Ning Chang, Hsiao-Yen Ma, Michael K. Gross, Stephen A. Ramsey, Theresa M. Filtz, and Chrissa Kioussi

ABSTRACT

Skeletal muscle is the largest organ in the body by mass, comprising roughly 40% of total body weight. Disruptions in skeletal muscle lead to muscle-wasting diseases that lead to muscle breakdown, and an overall loss in quality of life. Skeletal muscle in the forelimb develops in distinct phases during embryogenesis, including embryonic and fetal. During fetal myogenesis, embryonic myotubes fuse into both each other and fetal myotubes, to from fetal myofibers, which ultimately serve as the foundation for skeletal muscle that will continue to develop. Taking advantage of advances in DNA sequencing technologies, we performed whole transcriptome profiling via RNA-Seq of lineage-traced myoblasts during embryonic and fetal myogenesis. By isolating the same myogenic lineage at different developmental stages, we compared gene expression changes of a single cell population over time, and observed an upregulation of genes related to angiogenesis, cell adhesion, and the immune system, during fetal myogenesis. Coexpression analysis also revealed an immune-related gene subnetwork that exists during all stages of myogenesis, but is expressed at higher levels later in myogenesis. Our work will serve as a foundation for future studies that observe the effects of different perturbations on forelimb myogenesis.

INTRODUCTION

Skeletal muscles contract, cause movement and maintain homeostasis in the body. Forelimb muscles are derived from somites, which are anatomical structures derived from the paraxial mesoderm. Somites segment themselves into the myotome, sclerotome, and dermomyotome. The dermomyotome is divided into the epaxial and hypaxial dermomyotome, from the latter of which all skeletal muscle of the trunk and back derive (Burke and Nowicki, 2003; Christ and Ordahl, 1995). Around E10.5, embryonic myogenic precursor cells (EMPCs) express the homeobox sequence specific transcription factor (SSTF) Pax3, which triggers migration and delamination from the ventrolateral lip of the hypaxial dermomyotome into the limb bud (Bladt et al., 1995; Dietrich et al., 1999; Goulding et al., 1994; Hayashi and Ozawa, 1995). Ablation of Pax3 results in a forelimb deficient of skeletal muscle (Bober et al., 1994; Daston et al., 1996). Once colonized in the limb bud, skeletal muscle forms in distinct, successive stages (Tajbakhsh, 2005). Between E10.5 and E12.5, embryonic myoblasts fuse into embryonic myotubes. Between E12.5 and E14.5, fetal myoblasts fuse with both each other and embryonic myotubes to form fetal myofibers that serve as the foundation of future skeletal muscle. During this process, significant changes occur in gene expression (Biressi et al., 2007) and the underlying gene regulatory networks (Buckingham and Rigby, 2014; Messina et al., 2010), but not much is known about the driving molecular processes. Since skeletal muscle in the forelimb is derived from Pax3-positive progenitor cells, the Pax3 lineage offers an great tool to uncover the molecular processes during forelimb myogenesis.

Network analysis is a field that has existed for close to 20 years (Barabási and Albert, 1999), aiming to observe biological systems as individual parts working and interacting together (Barabasi and Oltvai, 2004; Kirschner, 2005). Recent advances in technology combined with a decrease in price for next-generation sequencing have resulted in frequent use and development of network analysis techniques, especially in development and disease (Singh et al., 2017). Nodes, representing genes, are connected to each other via edges, representing any sort of interaction. When applied to data on a large scale, it becomes possible to visualize complex interactions in an intuitive format. Coexpression networks are a type of biological network created from transcriptomics data and observe patterns of gene expression in biological systems (Dong et al., 2015). Coexpression networks have been used to identify changes in regulatory interaction responsible for cell-state phenotypes (Hsiao et al., 2016), and cell-type specific patterns of gene expression during development (Földy et al., 2016), among other uses. Applying coexpression analysis to Pax3 lineage traced myoblasts provides a model system to observe and decode the

mechanisms behind embryonic and fetal myogenesis, in the forelimb. In this study, we use next generation RNA sequencing (RNA-Seq) to perform differential expression and coexpression analysis during distinct stages of forelimb myogenesis. We discover the upregulation of vascularization and immune-related genes during fetal myogenesis, including a distinct immune-related subnetwork, implying an active role of the immune system in forelimb myogenesis.

MATERIALS AND METHODS

Forelimb Isolation and Dissociation

Female ICR mice were plugged on consecutive days by male *Pax3*^{Cre}|*Rosa*^{EGFP} mice. At 11.5, 12.5, 13.5, and 14.5 days post vaginal plug, the female mice were euthanized, and embryos collected in PBS over ice. Rapidly, embryos were genotyped using fluorescent microscopy. Forelimbs were dissected between the caudal edge of the shoulder and the lumbar region. Isolated forelimbs from each litter were pooled in Dulbecco's Modified Eagle Medium (DMEM) with 4.5 g/L glucose and no other additives, based on *Pax3*^{Cre}|*Rosa*^{EGFP} positive (G) and negative (W) genotypes; then pooled again based on both the father and time point. Dissociation of embryonic forelimbs was carried out as described previously (Campbell et al., 2012) with the following modifications. DMEM was removed, and dissociation buffer (HBSS without $CaCl_2$, $MgCl_2$, $MgSO_4$ [Gibco], 2 mg/mL Type I Collagenase [Worthington Biochem], 5mM EDTA) were added to pooled forelimbs at ~6 forelimbs per 1 mL buffer for E11.5 and E12.5, and ~2 forelimbs per 1 mL buffer at E13.5 and E14.5. Forelimbs were incubated for 3 minutes at 37°C. Forelimbs were then pipetted 10 times through a 1 mL pipette tip to promote dissociation. Forelimbs were incubated and pippeted once more at E11.5, E12.5, and E13.5, and twice more at E14.5. After the final dissociation step, each pooled sample was then centrifuged at 5000 rpm for one minute in a benchtop centrifuge. The media was aspirated off before re-suspending the cells in PBS by pipetting 15 times, to a final concentration between $1x10^6$ and $1x10^7$ cells/mL. Cell suspensions were transported to the flow cytometry facility and passed through a 35 µm nitex filter again, before they were sorted.

Fluorescence assisted cell sorting

Prepared cell suspensions were sorted using a Sony SH800 cell sorter [Sony Inc]. $EGFP^+$ (G)

cells were sorted directly into PBS. Once the full samples has been sorted, each tube (G) was spun at 3,800 rpm for 15 minutes at 4°C. PBS was aspirated off the cell pellets, and cell pellets were lysed with 350 µL Buffer RLT with added βMe [Qiagen]. Lysates were kept over ice until all samples were sorted.

RNA Preparation

RNA was extracted using RNAeasy mini kit [Qiagen] following the manufacturer's protocol. RNA was tested for quality and degradation using the AATI Fragment Analyzer [AATI]. RNA libraries were sequenced on a 100 bp single-end run on the Illumina Hiseq 4000 [Illumina]. Library preparation was done by trained technicians at the GC3F core facility using the Kapa Biosystems Stranded mRNA-seq Kit [Kapa]. 25 libraries were created and sequenced, corresponding to six, four, nine, and six biological replicates from each time point, respectively (E11.5, E12.5, E13.5, E14.5).

RNA Sequencing and Analysis

Primary Illumina data image analysis, base calling, and read-quality filtering were done using the Casava pipeline version 1.8.2 [Illumina]. Each sample was processed and analyzed with the same methods. After filtering low quality reads TopHat version 2.1.0 was used to align all reads to the mm10 genome with default parameters and to identify splice junctions (Kim et al., 2013; Trapnell et al., 2009). HTseq was used to create count tables from tophat2 aligned reads (Anders et al., 2015). DEseq2 was used to calculate differential gene expression between time points (Love et al., 2014) using an FDR adjusted cutoff of $p \leq 0.05$, with a fold change ≥ 1.5, between any two consecutive time points. Principal component analysis was performed using the prcomp function in R software (Ihaka and Gentleman, 1996). Heatmaps were generated using the pheatmap package in R software (Kolde, 2012). Signed difference ratios (SDR) were calculated similar to (Ramsey et al., 2008), except the average for each gene across all samples was subtracted from each sample.

Immunohistochemistry and Whole Mount Antibody Staining

Immunohistochemistry and PECAM and neurofilament staining was performed as done

previously (Ma et al., 2013).

Coexpression Network Construction and Analysis

Coexpression networks were constructed following the protocol from (Dong et al., 2015). Pairwise correlation coefficients were calculated between each of 4269 identified DEGs, in all samples, using an adjusted fdr cutoff of $p \leq 5e\text{-}15$. The coexpression network was visualized in Cytoscape (Shannon et al., 2003), and modules were identified via markov clustering (Enright et al., 2002)using the package MCL in R software. GO term enrichment in modules was determined by Panther GO (Mi et al., 2013, 2017).

RESULTS AND DISCUSSION

Isolation of Pax3-lineage traced myoblasts

In order to trace gene expression patterns during different stages of myogenesis in the forelimb, we developed a transgenic mouse model that combined a $Pax3^{Cre}$ driver (Engleka et al., 2005) with a ROSA26 EGFP tracer (Mao et al., 2001). Pax3 is a homeodomain SSTF that is known to mark all somite-derived skeletal muscle in the forelimb. In *Pax3*-null mice, myogenic progenitor cells fail to migrate and delaminate from the somite, which ultimately leads to little or no skeletal muscle in the forelimb (Bober et al., 1994; Daston et al., 1996). When both genotypes were combined into one mouse, EGFP expression was continuously induced in every cell that expressed Pax3 at one point, including any and all daughter cells (Figure 4.1A). This lineage tracer system enabled us to track the same myogenic population over time in the mouse forelimb as it developed and differentiated. We chose to profile the time points of E11.5, E12.5, E13.5, and E14.5, to trace development starting from the beginning of embryonic myogenesis, when the dermomyotome-derived cells had already entered the myogenic lineage, to the onset of fetal myogenesis, when the myoblasts started to form myotubes. Mouse embryos at each stage showed strong EGFP expression, especially noticeable in the forelimbs (Figure 4.1A) Individual digits and muscle groups developed in the forelimbs over time, seen clearly at E14.5. Fluorescence activated cell sorting (FACS) (Herzenberg et al., 1976) was used to isolate EGFP expressing cells ($Pax3^{EGFP}$ myoblasts), representing myoblasts during different stages of myogenesis. Density-based scatter plots that represent EGFP fluorescence intensity vs cell size,

revealed two distinct cell populations in each stage (Figure 4.1B). Visualization via histogram presented a more clear image of the two distinct cell populations present at each stage. The *Pax3*-derived myogenic lineage was represented by the EGFP-positive cells (highlighted in green), and the non-myogenic EGFP-negative population was displayed as the gray peaks (Figure 4.1C). Surprisingly, *Pax3*^{EGFP} myoblasts comprises 92% of the whole cell population of the forelimb at E11.5 and E12.5 (Figure 4.1B). This agreed with the strong EGFP-fluorescence seen by microscopy (Figure 4.1A). At E13.5, the *Pax3*^{EGFP} myoblast population dropped to 68% even though the fluorescence signal in the forelimb still looked strong (Figure 4.1A, B). While it was likely that other non-*Pax3*-derived cell populations were proliferating and differentiating, ultimately this drop was due to a decrease in dissociation efficiency. The onset of fetal myogenesis occurs between E12.5 and E13.5, when embryonic myofibers fuse with fetal myoblasts to form fetal myofibers that ultimately serve as the foundation for skeletal muscle. The cytoskeletal rearrangements that occurred between the cells imparted resistance to the enzo-mechanical dissociation process we used (see Materials and Methods). When the cells were passed through a 35 micron mesh, cell clumps were filtered out, including dense skeletal muscle that failed to dissociate. This was consistent with the decrease in *Pax3*^{EGFP} myoblasts observed from E13.5 to E14.5.

A more vigorous dissociation process would likely increase efficiency, but runs risk of changing gene expression in the cells. A previous study found few transcriptomics changes after dissociating and FACS of mouse adipocyte tissue (Richardson et al., 2015), which indicated the overall effect of dissociation and FACS may be relatively small. Our gene expression results (Figure 4.2B) were also similar to the results obtained from a previous study of sorted myoblasts, indicating replicability between the systems (Biressi et al., 2007). Recently, it has shown that adding the transcriptional inhibitor actinomycin d to each stage of dissociation and sorting of neuronal cells preserved the transcriptome of the cell state, despite the accumulated stress (Wu et al., 2017). This method would need to be verified in embryonic myoblasts first, but could potentially be used to increase dissociation efficiency without downstream side effects. Cell populations were sorted to a purity ranging from 97% to 99% (data not shown).

Transcriptomics analysis of Pax3-derived myoblasts

After sorting, total RNA from each sample was extracted, and tested with the Bioanalyzer for quality control. Only samples with an RIN above 7.0 were retained for library preparation and sequencing. Sequenced reads were aligned to me mm10 genome, and differentially expressed

genes were calculated between consecutive time points, using an FDR ≤ 0.05. Additional quality control was performed via principal component analysis (PCA) (Ringnér, 2008). The PCA plot (Figure 4.2A) showed distinct clustering of samples by stage. Interestingly, the samples followed a developmental trajectory. Samples from E11.5 clustered in the bottom left, and followed a horizontal parabola-like trajectory until E14.5, suggesting that time of conception is a significant factor in our analysis. Initially a surprise, the trajectory can be explained by the frequency of plug checks. Plugs were checked only once per day in the morning. Since vaginal plugs in mice are reported to last between 8-24 hours, each sample could be up to 12 hours apart, while still marked at the same stage. With more biological replicates and more stringent time points, in theory it would be possible to determine specific gene expression patterns that correspond to unique stages of myogenesis, at hourly intervals. The PCA plot also demonstrated the variability between biological replicates in our system, and emphasized the value of ample biological replicates for a study like this.

A heatmap was generated from the 4269 identified differentially expressed (DE) genes, based on the signed difference ratio from $\log_2$-normalized reads (see Materials and Methods) (Figure 4.2B). Genes formed distinct clusters based on their expression patterns in each stage. Of interest were the clusters marked with red, green, blue, and purple boxes, on the left. Genes from the red cluster were expressed specifically at E11.5, implying they are early embryonic myogenesis markers. Gene ontology (GO) term enrichment analysis of the red cluster revealed an overrepresentation of genes associated with pattern specification, CNS neuron differentiation, and digit morphogenesis (Figure 4.2C). Example genes in these categories were primarily homeodomain SSTFs represented by the Hox family. This agreed with previously reported studies that show the *Hox* gene family members, mostly the c and d, regulate patterning and digit formation in the embryonic limb (Martin, 1990; Pineault and Wellik, 2014; Raines et al., 2015).

Genes in the green cluster were expressed at E11.5 and E12.5, meaning they are markers of embryonic myogenesis. GO term enrichment analysis indicated that genes involved in epithelial tube morphogenesis, central nervous system (CNS) development, mesenchyme development, and cardiac outflow tract morphogenesis are overrepresented, suggesting that formation of the vascular system and CNS are both taking place during embryonic myogenesis. Since $Pax3^{EGFP}$ myoblasts represent the skeletal muscle lineage in the forelimb, it was surprising to find so many non-myogenic genes. To biologically validate these findings, we performed immunohistochemistry on sectioned $Pax3^{EGFP}$ forelimbs at E11.5 and E12.5, using antibodies against EGFP (green), Pitx2 (red), and Myog (blue) (Figures 4.2D, E). EGFP represented the

Pax3^{EGFP} lineage, Pix2 expressed in all muscel anlagen and Myog expressed in cells committed to the skeletal muscle lineage (Buckingham and Rigby, 2014). At both stages, there were clear pockets of green cells with no other colors overlayed, which represented non-myogenic populations within the *Pax3*^{EGFP} lineage. Previous studies have shown a small subset of *Pax3*^{EGFP} cells in the forelimb express Foxc2, which differentiate into vascular epithelial cells (Lagha et al., 2009). This could also explain the enrichment in vascular system related genes that is observed. To probe further, we performed neurofilament (Figures 4.2F, G) and PECAM staining (Figures 4.2H, I) on forelimbs at E11.5 and E12.5 to observe the developing nervous and vascular systems, respectively. Staining showed that both systems started to develop at E11.5 with development becaming much more pronounced at E12.5. There have been other studies that showed communication between the nervous system and skeletal muscle during development (Deries and Thorsteinsdóttir, 2016; Jostes et al., 1990). In addition, a population of smooth-muscle endothelial cells derive from Pax3-positive progenitor cells (Young et al., 2016). Immunohistochemistry using known angioblast, neuron, and myogenic markers could further elucidate the location of non-myogenic *Pax3*^{EFGP} cell populations in the forelimb. It is likely that other non-myogenic populations derived from the Pax3 lineage will continue to be discovered as more studies are performed.

Genes in the blue cluster showed high expression levels at E13.5 and E14.5 which coincideded with the onset of fetal myogenesis. They were enriched in the functions of angiogenesis and negative regulation of cell proliferation and differentiation (Figure 4.2C). Example genes included known angiogenesis markers such as *Angpt2*, *Anpep*, and negative markers of cell proliferation, *Ar* and *Dpt*. The presence of such markers suggests that *Pax3*^{EGFP} myoblasts stop proliferating while angiogenesis is continuing during fetal myogenesis. Multiple labs have demonstrated that expression of certain angiogenesis-related genes can increase the rate of muscle regeneration in adult skeletal muscle (Borselli et al., 2010; Mofarrahi et al., 2015). Though no studies focusing on the forelimb during development have been done, it has been shown that skeletal muscle regeneration in adults shares many of the same mechanisms as myogenesis including the activation of skeletal muscle-specific SSTFs (Yusuf and Brand-Saberi, 2012). Taken together, these imply that angiogenesis and myogenesis are interrelated in the forelimb, but studies have yet to prove so definitively.

The purple cluster of genes was expressed explicitly at E14.5 and unexpectedly, genes involved with the inflammatory response and immune system are overrepresented (Figure 4.2C). Example genes are interleukin receptors and CD antigens such as *Ccl6*, *Cd44*, *Il20rb*, and *Ciita*. Similar to

angiogenesis, there is little research on the interaction between skeletal muscle and the immune system during myogenesis. To bolster our results, we compared our differentially expressed genes with those from Biressi et al. (Biressi et al., 2007), and found a similar list of immune-related genes such as Anxa1, Cd44, and Myb, among others. This supports the expression of immune-related genes during fetal myogenesis.

An alternate explanation of the expression of immune-related genes during fetal myogenesis could be the non-specificity of GO terms. GO term enrichment analysis is inherently biased to some degree because it only takes into account the known and annotated functions of genes. Most genes have multiple, if not dozens of different biological functions, that can be context dependent based on tissue type or other variables. Some genes were only studied in one system with incomplete information in regard to their other functions. Additionally, certain GO terms, such as “immune response,” are contextually broad and poorly defined. Taken together, a GO term enrichment analysis could include the wrong context of one or more genes, and bias the results in a way that does not reflect the true underlying biology. More stringent biological validation such as immunohistochemistry with known lineage markers, or transgenic mouse KO studies is required to truly determine whether immune-related genes are expressed during fetal myogenesis.

Recent studies though have brought awareness to communication between the immune system and muscle regeneration Macrophage infiltration and inflammation occur during satellite-cell mediated skeletal muscle regeneration (Costamagna et al., 2015; Saclier et al., 2013). It should be noted that these genes were mostly expressed at the later stage E14.5 after the onset of fetal myogenesis, unlike the angiogenesis markers. Since cells were sorted to a final purity between 97-99% (data not shown), upregulation of these genes is unlikely to be caused by *non-Pax3*^{EGFP} cells. We cannot say if the enrichment is due to a sub-population of non-myogenic *Pax3*^{EGFP} cells, or is expressed by myogenic cells communicating with the immune system. As mentioned previously, a third explanation could be that the expressed, enriched immune-related genes could have a non-immune related function in the context of myogenesis. Immunohistochemistry using known immune-response markers combined with mouse KO models could help to answer this question.

Construction of coexpression network during myogenesis

To complete the DE analysis, we performed coexpression analysis, to identify coexpressed genes and modules during myogenesis. A single coexpression network was constructed from pairwise correlation coefficients between each of 4269 DE genes, using all samples (see Materials and Methods). We opted to construct a single network for all samples, rather than stage-wise individual networks, to increase the power of our analysis. This approach lead to slightly biased results, since the resulting significantly correlated gene pairs were coexpressed consistently across all stages. This resulted in a coexpression network that existed across all observed stages of myogenesis. We focused on genes that were differentially expressed between consecutive stages. This provided several advantages, (1) it reduced computational time and intensity and (2) it limited the results to genes that are likely biologically relevant during myogenesis. Upon calculating pearson correlation coefficients (PCC) in a pairwise manner, we needed to choose an FDR cutoff for significant correlation. Since the node-degree distribution of biological networks has been shown to closely follow a scale-free distribution (Barabási and Albert, 1999), the p-value choice needed to reflect that. When the p-value cutoff vs the R^2 of a best-fit power line was plotted for the resulting node-degree distribution (Figure 4.3A), a p-value cutoff of 5E-15 was chosen, and resulted in an R^2 value of 0.9, which indicated that the coexpression network followed a scale-free topology. The node degree vs the number of nodes with that degree was plotted (Figure 4.3B) and confirmed the fit. Ultimately, a network with 740 nodes and 4,481 edges was generated, with an average node degree of 12.11.

When the network was visualized with Cytoscape software (Shannon et al., 2003), we observed a single network composed of two mostly independent subnetworks (Figure 4.2C). Each node (circle) represented a transcript, and edges represented significant correlation between the transcripts. Node size was proportional to its degree. The top network was the larger network, consisting of 411 genes, compared to the smaller network with 196 genes. A quick GO term enrichment analysis revealed an overrepresentation of structural and cytoskeleton related genes in the top network, and immune response related genes in the bottom network. This was surprising, as little is known about the role of the immune system in myogenesis. It implies two different transcriptional coexpression networks co-exist during embryonic and fetal myogenesis with little interaction between them. As mentioned previously, a possible explanation could be the broad annotations of GO terms. Genes in the smaller subnetwork could have unannotated functions related to myogenesis. One other possible explanation could be that the immune system-related network reflects a sub-population of *Pax3*^{EGFP} cells that are not related to the myogenic lineage. The discovery of smooth-muscle epithelial cells from the *Pax3*^{EGFP} lineage

(Lagha et al., 2009) implies there may be other non-myogenic populations to be discovered. Another possibility could be that there are two separate networks expressed in the same cell type. Immunohistochemistry, and/or multi-color FACS, using markers for different hub genes in the smaller subnetwork, could be used in the future to determine which is the case. Similarly, conditional knock-out experiments using floxed transgenic mouse lines could be used to probe the function of identified immune-related genes in the context of forelimb myogenesis.

Modules are clusters of highly interconnected nodes that together perform a specific biological function (Barabasi and Oltvai, 2004). Using the MCL package in R statistical software we performed markov clustering to identify modules (Van Dongen, 2001). Markov clustering identifies modules by simulating flow in networks, and determining the clusters in which the most flow accumulates. The weakness of this method is that it assigns each gene to only a single module, which rarely reflects the true underlying biology.

Using a module size cutoff of nine, Markov clustering identified ten modules. Five modules (blue, orange, red, yellow, cyan) comprised the larger subnetwork, and four (green, gray, black, dark green) comprised the smaller subnetwork. The tenth module (purple) was observed to be a bottleneck, connecting the two distinct subnetworks (Figure 4.3C). Bottlenecks serve as bridges between seemingly unrelated genes and modules, imparting them with biological significance (Barabasi and Oltvai, 2004). GO term enrichment identified significant overrepresentation of extracellular matrix organization related genes in the orange module, and cell adhesion related genes in the blue module, implying that those modules are involved in the cytoskeletal rearrangements that occur during myogenesis.

The green module was enriched in immune response related genes, and the black, gray and dark green modules were all enriched in leukocyte GO term-related genes. No significant GO term enrichment, except for “unclassified,” was observed in any of the other modules. The purple module is interesting since it serves as a bottleneck between the two subnetworks, but showed no significant GO term enrichment. This supports the conclusion that the analysis is limited by the non-specific context of GO terms. Upon isolating only the ten identified modules, we observed high intra-connectivity within each module, and some modules were inter-connected to various degrees (Figure 4.3D). The blue and orange modules were the two largest, with 167 and 55 genes, respectively.

Identification of module function and expression

Following module identification, we observed the individual gene expression levels in each module over time, to identify stage specific module expression. To acquire a more clear picture of the individual changes happening in each module, a heatmap was generated based on SDR values for each gene, relative to E11.5 (Figure 4.4A). This enabled us to trace gene expression within modules over time. We observed in the yellow module that half of the genes were increased in expression while the other half decreased, as stage increases. Modules are color-coded with a bar on the right. The extracellular matrix organization and cell-adhesion related modules were expressed at E12.5, relative to E11.5, but strongest expression was observed at E13.5 and E14.5. This differed from the immune-response related modules, in which strongest expression occurred at E14.5, the later stage of fetal myogenesis. This difference suggested that the cell adhesion and extracellular matrix-related modules were active during the onset of fetal myogenesis, whereas the immune-response related modules were most active after the onset of fetal myogenesis. The results from the cell-adhesion related modules were expected, supported by significant cytoskeletal rearrangement that occur during the fusion processes of fetal myogenesis. The later expression of the immune-related modules demonstrated they were likely to be more involved in the progression of fetal myogenesis. Observing gene expression of *Pax3*^{EGFP} myoblasts at E15.5 and later stages could clarify this, but more informative would be the conditional knock-out of immune related genes in the *Pax3* lineage. These studies would further confirm whether an immmune-related *Pax3*-derived cell lineage exists in the forelimb, or whether the identified immune-related genes have an unknown function in the context of myogenesis.

Upon observing the late stage expression of all modules, we wanted to observe more specifically what was happening during early myogenesis. We cross referenced our list of 4,269 DE genes with a list of SSTFs, and created a heatmap based on the SDR values of the 302 DE SSTFs (Figure 4.4C). The genes were separated into four overlapping clusters, based on the stages in which they were expressed. SSTFs expressed at E11.5 were comprised mostly of homeodomain SSTFs, and enriched in genes related to the skeletal, nervous, and hematopoietic systems. The large number of homedomain SSTFs agreed with both similar studies (Biressi et al., 2007) and previous literature, showing expression of homedomain SSTFs early in development of the forelimb (Buckingham and Rigby, 2014). SSTFs expressed at E12.5 were similarly involved in development of the skeletal, muscular, and hematopoetic systems. This occured in late embryonic myogenesis, when embryonic myoblasts fused to create embryonic myotubes. PECAM staining at E12.5 (Figure 4.2H, I) confirmed significant development of the vasculature system during this period. SSTFs expressed at E13.5 pertained to the muscular and vascular systems, and are

involved in growth and metabolism. This agreed with what is known about fetal myogenesis, when fusion events lead to significant changes in cytoskeletal organization while fetal myofibers are formed. SSTFs upregulated at E14.5 were related to the vascular and immune systems. This agreed with the overall gene expression data and the coexpression network analysis, implicating possible involvement of the immune system during fetal myogenesis. Future studies should confirm whether the identified immune-related genes are involved in myogenesis, or if they are artifacts from the GO term analysis.

CONCLUSION

This study used RNA-Seq on *Pax3*^{EGFP} lineage-traced forelimb myoblasts isolated by FACS, to observe the transcriptional networks of embryonic and fetal myogenesis. Using differential expression analysis combined with GO enrichment analysis, we found stage-specific expression of gene groups. Patterning and CNS related genes are expressed during embryonic myogenesis, and angiogenesis, cytoskeletal, and immune-related genes are expressed during fetal myogenesis. Additionally, co-expression network analysis revealed two distinct subnetworks present during both embryonic and fetal myogenesis. The larger one is related to cell adhesion and extracellular matrix organization, and the smaller one is immune-response and leukocyte related. Both coexpression networks are present during both stages of myogenesis, but overall expression is higher during fetal myogenesis. Taken together, we have demonstrated that forelimb myogenesis is more complex than previously imagined. Even in the Pax3-lineage, which represents skeletal muscle in the forelimb, we have identified expression of non-skeletal muscle related genes, and subnetworks. More studies are needed to confirm the context specific function of the identified immune-related genes, in forelimb myogenesis. Our work will provide a foundational base for future studies to observe the communication between different cell types during forelim myogenesis.

ACKNOWLEDGEMENTS

We thank Maggie Weitzman at the Genomics and Cell Characterization Core Facility at the University of Oregon for her assistance with the flow sorter and the LARC personnel of the Oregon State University for the mouse husbandry. This work was supported by the College of Pharmacy at the Oregon State University and the Oregon State University.

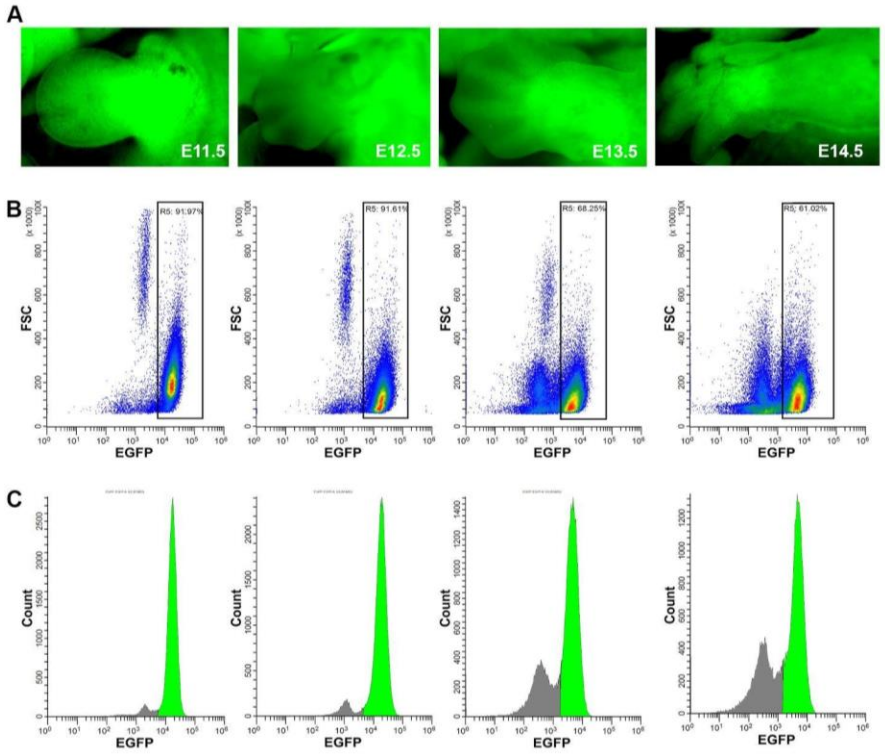
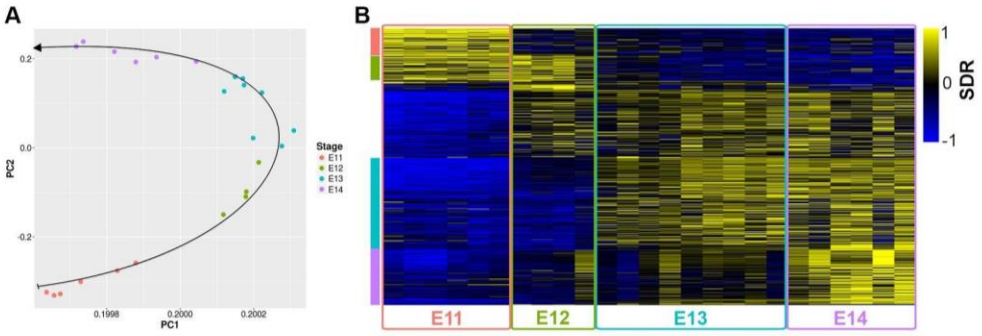


Figure 4.1. EGFP expression in mouse embryonic forelimbs
(A) Fluorescent microscopy showing EGFP expression based on a *Pax3*^{Cre}|*Rosa26*^{EGFP} driver at stages E11.5, E12.5, and E14.5.
(B) Scatter plots from FACS showing EGFP intensity on the x-axis, and forward scatter (FSC) on the y-axis. Gate R5 shows 92%, 92%, 68%, and 61% EFGP-positive cells in forelimbs at E11.5, E12.5, E13.5, and E14.5, respectively.
(C) Histograms showing EGFP intensity on the x-axis vs cell number (count) on the y-axis. Green peaks represent EFGP-positive populations based on gating from R5.



C

Category	P-value	Gene examples
anterior/posterior pattern specification	1.39E-06	Alx4, Cdx1, Hoxa9, Hoxa10, Hoxa11, Hoxc10, Hoxc11, Hoxc13, Hoxd4, Nkx3-1
proximal/distal pattern formation	1.79E-06	Hoxa9, Hoxa10, Hoxa11, Hoxc10, Hoxc11, Dlx2, Sp8
central nervous system neuron differentiation	8.88E-06	Cbln1, Cend1, Emx1, Evx1, Dicer1, Lhx1, Nrp2, Pou4f1, Sall1, Tfat2c, Uncx
embryonic digit morphogenesis	1.02E-04	Alx4, Bmp4, Grhl2, Mycn, Shh, Ttbk1, Wnt7a
epithelial tube morphogenesis	3.19E-05	Casr, Ccdc40, Dya1, Kif26b, Lbx1, Lhx2, Pbx1, Trim71, Zic2, Zic5
central nervous system development	4.31E-05	Draxin, Gdf7, Hmga2, Hydrin, Lbx1, Lhx2, Mnx1, Ncan, Tenm4
mesenchyme development	6.93E-05	Dli3, Eomes, Fgf15, Gsc, Hoxa5, Hnf1a, Six2, Six4
cardiac outflow tract morphogenesis	8.99E-05	Eya1, Fgf8, Gata4, Hand2, Hey1, Isl1, Lrp2, Pitx2, Six1, Sox11, Tbx3, Tgfb2
angiogenesis	1.01E-10	Angpt2, Angptl4, Anpep, Anxa2, Eng, Flgf, Flt4, Kdr, Plau, Plxnd1, Nos3
negative regulation of cell proliferation	1.39E-10	Agt, Ar, Bcl6, Btk, Cd109, Cpeb1, Dpt, Drd2, Foxa3, Gkn3 Lrrc32, Sparc, Spn
negative regulation of signal transduction	2.16E-10	Apod, Apoe, Cav3, Cmya5, Deptor, Gatsl3, Lect1, Notch4, Lpxn, Rgs6, Taok3
negative regulation of cell differentiation	7.49E-9	Abcc8, Atf5, Ccl9, Dab2, Ereg, Mkx, Ntrk4, Pdgfb, Zfpm2
inflammatory response	2.95E-10	Anxa1, Ccl6, Cd44, Cd300a, Cnr2, Il20rb, Mylk3, Pla2g7, Tgfb1, Unc13d
cell surface receptor signaling pathway	7.97E-08	Adgra1, Adgrf4, Cebpa, Dok2, Fes, Foxh1, Gas6, Myo1d, Plaur, Syk, Tspan32
positive regulation of immune response	2.66E-07	C3, Ccr2, Cd44, Cd177, Coch, Myb, Myd88, Prkcb, Stap1
innate immune response	5.78E-06	Card9, Ciita, Clec4n, Coro1a, Defb1, Fcnb, Fes, H2-Eb11, Il12rb1, Mx1

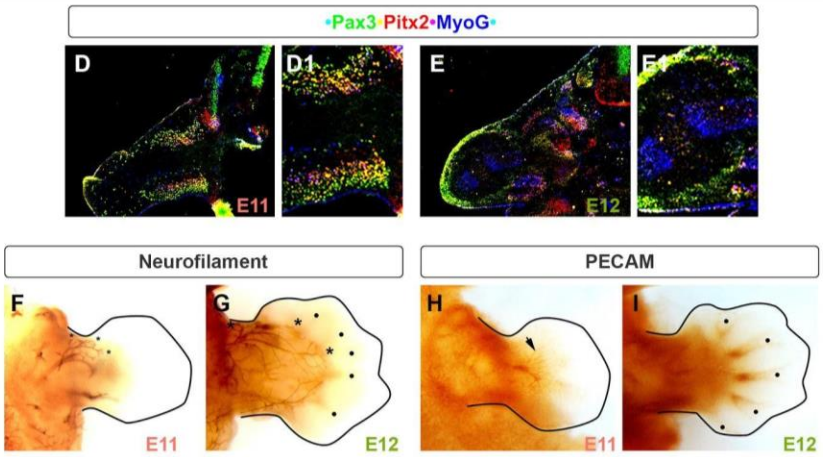


Figure 4.2

Figure 4.2. Differential expression (DE) and Gene Ontology (GO) term analysis of RNA-Seq data from sorted, EGFP-positive cells.
(A) Principal component analysis (PCA) and plot of all 25 samples. PCA shows good clustering of samples by biological time point, with variation between samples in the same group. Samples appear to follow a developmental trajectory as the stage increases.
(B) Heatmap of signed difference ratio (SDR) based on all 4,269 DE genes between any two consecutive stages. Columns represent samples, and each row represents one DE gene. Yellow indicates high expression and blue indicates low expression, relative to the average expression of each gene between all samples. Red, green, blue, and purple bars on the left indicate clusters of DE genes expressed at E11.5, E11.5 and E12.5, E13.5 and E14.5, and E14.5, respectively.
(C) Go term enrichment analysis of gene clusters shown in **(B**). GO term enrichment was calculated with Panther GO. The top four enriched "child" GO terms from each cluster, determined by p-value, are shown.
(D, E) Immunohistochemistry staining of E11.5 **(D)** and E12.5 **(E)** forelimbs from *Pax3*^{EGFP} embryos. EGFP is shown in green, Pitx2 is shown in red, and Myog is shown in blue. Note that not all green cells are blue, indicating non-myogenic cells that derive from the *Pax3*^{EGFP} lineage.
(F, G) Whole mount antibody Neurofilament staining in the forelimb at E11.5 **(F)** and E12.5 **(G)**. Neural development starts in the forelimb around E11.5 and increases at E12.5.
(H, I) Whole mount antibody PECAM staining of the forelimb at E11.5 **(H)** and E12.5 **(I)**, indicating blood vasculature development.

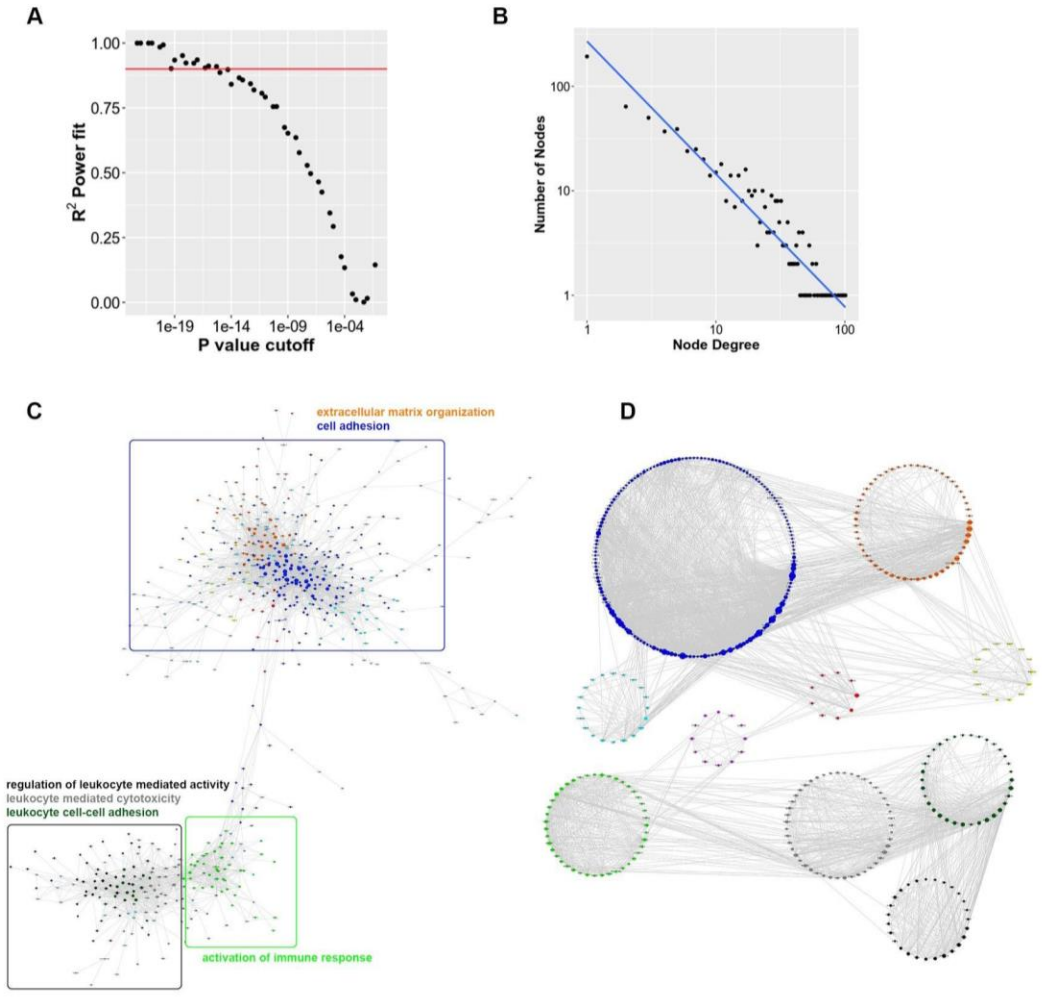

A
1.00
0.75
0.50
0.25
0.00
R^2 Power fit
1e-19
1e-14
1e-09
1e-04
P value cutoff
B
100
10
1
Number of Nodes
1
10
100
Node Degree
C
extracellular matrix organization
cell adhesion
regulation of leukocyte mediated activity
leukocyte mediated cytotoxicity
leukocyte cell-cell adhesion
activation of immune response
D

Figure 4.3

Figure 4.3. Coexpression network construction and module identification
(A) Plot showing p-value cutoff for significant pairwise pearson correlation coefficients (see materials and methods) vs R^2 of a power-fit line. Node degree distributions in biological networks are known to follow a power law, so a p-value cutoff of 5E-15 was chosen, giving an R^2 value of 0.9.
(B) Plot showing node degree on the x-axis vs the number of nodes with that degree on the y-axis, shown on a log10 scale. A line fits the data with an R^2 value of 0.9, showing that the generated network roughly follows a scale-free distribution.
(C) The generated coexpression network was visualized in Cytoscape software. Nodes (transcripts) are shown as circles, with size proportional to the degree of the node. Ten modules with at least nine nodes were identified via Markov clustering, and are color-coded accordingly. The full coexpression network is comprised of two, mostly-distinct subnetworks. The upper subnetworks contains more nodes and is enriched in genes involved in extracellular matrix organization and cell adhesion. The smaller, bottom subnetwork is enriched in genes involved in the activation of immune response, and leukocyte-related activities.
(D) Isolated identified modules, viewed in cytoscape. Modules show strong intra-connectivity, and additionally some interconnectivity with each other.

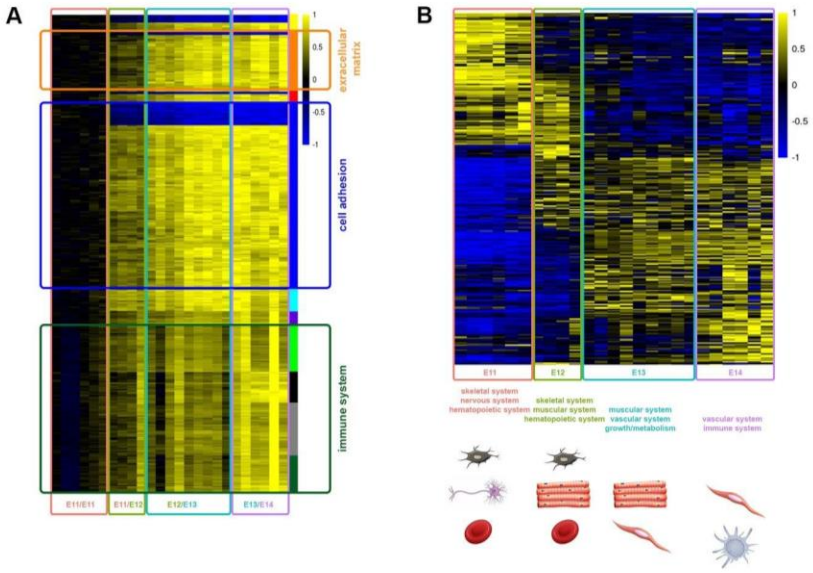


Figure 4.4. Expression of modules and SSTFs during myogenesis
(A) The heatmap of all genes in modules based on SDR value relative to E11.5, sorted by module. Each column is a sample and each row is a gene. The colored bar on the right represents which module the genes belong to. Modules separate based on their expression pattern, and are expressed higher overall at E13.5 and E14.5.
(B) The heatmap of only 302 differentially expressed SSTFs between any two time points, based on the SDR relative to all samples. Columns are samples, and rows represent individual SSTFs. SSTFs cluster distinctly, and show stage-specific expression patterns.

REFERENCES

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Barabási, A.-L., and Albert, R. (1999). Emergence of Scaling in Random Networks. Science *286*, 509–512.

Barabasi, A.-L., and Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. Nat. Rev. Genet. *5*, 101–113.

Biressi, S., Tagliafico, E., Lamorte, G., Monteverde, S., Tenedini, E., Roncaglia, E., Ferrari, S., Ferrari, S., Cusella-De Angelis, M.G., Tajbakhsh, S., et al. (2007). Intrinsic phenotypic diversity of embryonic and fetal myoblasts is revealed by genome-wide gene expression analysis on purified cells. Dev. Biol. *304*, 633–651.

Bladt, F., Riethmacher, D., Isenmann, S., Aguzzi, A., and Birchmeier, C. (1995). Essential role for the c-met receptor in the migration of myogenic precursor cells into the limb bud. Nature *376*, 376768a0.

Bober, E., Franz, T., Arnold, H.H., Gruss, P., and Tremblay, P. (1994). Pax-3 is required for the development of limb muscles: a possible role for the migration of dermomyotomal muscle progenitor cells. Development *120*, 603–612.

Borselli, C., Storrie, H., Benesch-Lee, F., Shvartsman, D., Cezar, C., Lichtman, J.W., Vandenburgh, H.H., and Mooney, D.J. (2010). Functional muscle regeneration with combined delivery of angiogenesis and myogenesis factors. Proc. Natl. Acad. Sci. U. S. A. *107*, 3287–3292.

Buckingham, M., and Rigby, P.W.J. (2014). Gene regulatory networks and transcriptional mechanisms that control myogenesis. Dev. Cell *28*, 225–238.

Burke, A.C., and Nowicki, J.L. (2003). A new view of patterning domains in the vertebrate mesoderm. Dev. Cell *4*, 159–165.

Campbell, A.L., Eng, D., Gross, M.K., and Kioussi, C. (2012). Prediction of gene network models in limb muscle precursors. Gene *509*, 16–23.

Christ, B., and Ordahl, C.P. (1995). Early stages of chick somite development. Anat. Embryol. *191*, 381–396.

Costamagna, D., Costelli, P., Sampaolesi, M., and Penna, F. (2015). Role of Inflammation in Muscle Homeostasis and Myogenesis. Mediators Inflamm. *2015*, 805172.

Daston, G., Lamar, E., Olivier, M., and Goulding, M. (1996). Pax-3 is necessary for migration but not differentiation of limb muscle precursors in the mouse. Development *122*, 1017–1027.

Deries, M., and Thorsteinsdóttir, S. (2016). Axial and limb muscle development: dialogue with the neighbourhood. Cell. Mol. Life Sci. *73*, 4415–4431.

Dietrich, S., Abou-Rebyeh, F., Brohmann, H., Bladt, F., Sonnenberg-Riethmacher, E., Yamaai, T., Lumsden, A., Brand-Saberi, B., and Birchmeier, C. (1999). The role of SF/HGF and c-Met in the development of skeletal muscle. Development *126*, 1621–1629.

Dong, X., Yambartsev, A., Ramsey, S.A., Thomas, L.D., Shulzhenko, N., and Morgun, A. (2015). Reverse enGENEering of Regulatory Networks from Big Data: A Roadmap for Biologists. Bioinform. Biol. Insights *9*, 61–74.

Engleka, K.A., Gitler, A.D., Zhang, M., Zhou, D.D., High, F.A., and Epstein, J.A. (2005). Insertion of Cre into the Pax3 locus creates a new allele of Splotch and identifies unexpected Pax3 derivatives. Dev. Biol. *280*, 396–406.

Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. *30*, 1575–1584.

Földy, C., Darmanis, S., Aoto, J., Malenka, R.C., Quake, S.R., and Südhof, T.C. (2016). Single-cell RNAseq reveals cell adhesion molecule profiles in electrophysiologically defined neurons. Proc. Natl. Acad. Sci. U. S. A. *113*, E5222–E5231.

Goulding, M., Lumsden, A., and Paquette, A.J. (1994). Regulation of Pax-3 expression in the dermomyotome and its role in muscle development. Development *120*, 957–971.

Hayashi, K., and Ozawa, E. (1995). Myogenic cell migration from somites is induced by tissue contact with medial region of the presumptive limb mesoderm in chick embryos. Development *121*, 661–669.

Herzenberg, L.A., Sweet, R.G., and Herzenberg, L.A. (1976). Fluorescence-activated cell sorting. Sci. Am. *234*, 108–117.

Hsiao, T.-H., Chiu, Y.-C., Hsu, P.-Y., Lu, T.-P., Lai, L.-C., Tsai, M.-H., Huang, T.H.-M., Chuang, E.Y., and Chen, Y. (2016). Differential network analysis reveals the genome-wide landscape of estrogen receptor modulation in hormonal cancers. Sci. Rep. *6*, 23035.

Ideker, T., and Krogan, N.J. (2012). Differential network biology. Mol. Syst. Biol. *8*, 565.

Ihaka, R., and Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. J. Comput. Graph. Stat. *5*, 299–314.

Jostes, B., Walther, C., and Gruss, P. (1990). The murine paired box gene, Pax7, is expressed specifically during the development of the nervous and muscular system. Mech. Dev. *33*, 27–37.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Kirschner, M.W. (2005). The meaning of systems biology. Cell *121*, 503–504.

Kolde, R. (2012). Pheatmap: pretty heatmaps. R Package Version *61*.

Lagha, M., Brunelli, S., Messina, G., Cumano, A., Kume, T., Relaix, F., and Buckingham, M.E. (2009). Pax3:Foxc2 reciprocal repression in the somite modulates muscular versus vascular cell fate choice in multipotent progenitors. Dev. Cell *17*, 892–899.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Ma, H.-Y., Xu, J., Eng, D., Gross, M.K., and Kioussi, C. (2013). Pitx2-mediated cardiac outflow tract remodeling. Dev. Dyn. *242*, 456–468.

Mao, X., Fujiwara, Y., Chapdelaine, A., Yang, H., and Orkin, S.H. (2001). Activation of EGFP expression by Cre-mediated excision in a new ROSA26 reporter mouse strain. Blood *97*, 324–326.

Martin, P. (1990). Tissue patterning in the developing mouse limb. Int. J. Dev. Biol. *34*, 323–336.

Mayeuf-Louchart, A., Montarras, D., Bodin, C., Kume, T., Vincent, S.D., and Buckingham, M. (2016). Endothelial cell specification in the somite is compromised in Pax3-positive progenitors of Foxc1/2 conditional mutants, with loss of forelimb myogenesis. Development *143*, 872–879.

Messina, G., Biressi, S., Monteverde, S., Magli, A., Cassano, M., Perani, L., Roncaglia, E., Tagliafico, E., Starnes, L., Campbell, C.E., et al. (2010). Nfix regulates fetal-specific transcription in developing skeletal muscle. Cell *140*, 554–566.

Mi, H., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2013). Large-scale gene function analysis with the PANTHER classification system. Nat. Protoc. *8*, 1551–1566.

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., and Thomas, P.D. (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. *45*, D183–D189.

Mofarrahi, M., McClung, J.M., Kontos, C.D., Davis, E.C., Tappuni, B., Moroz, N., Pickett, A.E., Huck, L., Harel, S., Danialou, G., et al. (2015). Angiopoietin-1 enhances skeletal muscle regeneration in mice. Am. J. Physiol. Regul. Integr. Comp. Physiol. *308*, R576–R589.

Pineault, K.M., and Wellik, D.M. (2014). Hox genes and limb musculoskeletal development. Curr. Osteoporos. Rep. *12*, 420–427.

Raines, A.M., Magella, B., Adam, M., and Potter, S.S. (2015). Key pathways regulated by HoxA9,10,11/HoxD9,10,11 during limb development. BMC Dev. Biol. *15*, 28.

Ramsey, S.A., Klemm, S.L., Zak, D.E., Kennedy, K.A., Thorsson, V., Li, B., Gilchrist, M., Gold, E.S., Johnson, C.D., Litvak, V., et al. (2008). Uncovering a Macrophage Transcriptional Program by Integrating Evidence from Motif Scanning and Expression Dynamics. PLoS Comput. Biol. *4*, e1000021.

Richardson, G.M., Lannigan, J., and Macara, I.G. (2015). Does FACS perturb gene expression? Cytometry A *87*, 166–175.

Ringnér, M. (2008). What is principal component analysis? Nat. Biotechnol. *26*, 303–304.

Saclier, M., Cuvellier, S., Magnan, M., Mounier, R., and Chazaud, B. (2013). Monocyte/macrophage interactions with myogenic precursor cells during skeletal muscle regeneration. FEBS J. *280*, 4118–4130.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski,

B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Singh, A.J., Ramsey, S.A., Filtz, T.M., and Kioussi, C. (2017). Differential gene regulatory networks in development and disease. Cell. Mol. Life Sci.

Tajbakhsh, S. (2005). Skeletal muscle stem and progenitor cells: reconciling genetics and lineage. Exp. Cell Res. *306*, 364–372.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Van Dongen, S.M. (2001). Graph clustering by flow simulation.

Wu, Y.E., Pan, L., Zuo, Y., Li, X., and Hong, W. (2017). Detecting Activated Cell Populations Using Single-Cell RNA-Seq. Neuron *96*, 313–329.e6.

Young, K., Krebs, L.T., Tweedie, E., Conley, B., Mancini, M., Arthur, H.M., Liaw, L., Gridley, T., and Vary, C.P.H. (2016). Endoglin is required in Pax3-derived cells for embryonic blood vessel formation. Dev. Biol. *409*, 95–105.

Yusuf, F., and Brand-Saberi, B. (2012). Myogenesis and muscle regeneration. Histochem. Cell Biol. *138*, 187–199.

General Conclusions and Future Directions
Chapter 5

Arun J Singh

CONCLUSIVE REMARKS

Organogenesis is well orchestrated spatiotemporal process that directs the formation and growth of organs during embryonic development. Genetic abnormalities at the molecular level manifest themselves as disorders with debilitating and/or lethal outcomes caused by malfunctions in developmental processes. Understanding how these processes affect organ development and formation can lead to the development of therapies and/or drugs that re-enable the developmental processes. We will be able to link developmental and regenerative processes by the common molecular mechanisms that they share. The emerging field of network analysis has exploited dynamic gene expression patterns to reveal functional modules, pathways and networks involved in organ development and disease.

Though all cells begin with identical genetic material, the epigenome determines the pattern of gene expression, imparting each cell with its distinct characteristics, functions, and behaviors. During embryonic development cells differentiate into multiple types, each with its own biological functions to fulfill its niche. Cis-regulatory modules (CRM) are key to this process; by using specific CRMs that are comprised of multiple sequence specific transcription factor (SSTF) binding sites, a cell can either stay in the cell cycle, or can exit the cell cycle and enter the post-mitotic differentiation state. CRMs can act as switches to determine “availability” of associated loci for expression, initiate SSTF and lineage-specific gene expression programs, and change epigenetic regulation to help stabilize unique expression patterns.

Forelimb myogenesis is a tightly regulated process, composed of multiple distinct phases, each with its own unique gene networks. In order to gain a full understanding of the molecular changes occurring during myogenesis, it is necessary to examine all properties of the cell, including the chromatin state and the whole transcriptome. Pitx2 is a homeodomain SSTF that plays a role in the transition from embryonic (E10.5-E12.5) to fetal myogenesis (E13.5-E17.5). In the embryonic forelimb, Pitx2 functions mostly as a repressor of its target genes (Chapter 2). Lack of Pitx2 in the forelimb at E12.5 results in an alteration of the chromatin state, especially around genes involved in neurogenesis and cytoskeletal adhesion (Figure 2.2). It is likely that this change in chromatin stage is responsible for the malformation of skeletal muscle anlagen in the forelimb found in Pitx2-null mice.

Pax3 is a paired-homeodomain SSTF that is responsible for the migration and delamination of embryonic myogenic precursor cells from the hypaxial dermomyotome into the limb bud. Transcriptional profiling of *Pax3* lineage-traced myoblasts in the embryonic forelimb revealed high

expression of neurogenesis related genes during embryonic myogenesis, implying simultaneous communication between the nervous system and skeletal muscle during early myogenesis (Chapter 4). Additionally, expression of angiogenesis related genes and unexpectedly, immune-response related genes during fetal myogenesis suggest that vascular and immune-system development are important during fetal myogenesis. Coexpression network analysis during all stages of myogenesis revealed two distinct subnetworks. GO term overrepresentation analysis revealed that the larger subnetwork was enriched with genes involved in cell adhesion, and extracellular matrix organization. The smaller network was enriched in genes related to the immune-response, indicating that an immune-related subnetwork exists during both embryonic and fetal myogenesis. Differential expression analysis confirmed that both subnetworks are expressed at higher levels during fetal myogenesis. The expression of the-cell adhesion related network increases at E13.5, which is the onset of fetal myogenesis. Expression of the immune-related subnetwork is highest at E14.5, alleging the immune system plays a role later in fetal myogenesis. Altogether, we believe that forelimb myogenesis is a complex process that requires communication of multiple different networks corresponding to unique cell types, such as myoblasts, neurons, angioblasts, and leukocytes, working together in a spatio-temporal manner.

As forelimb myoblasts continually differentiate and fuse together over time, drastic phenotypic changes occur on the cellular level. These phenotypic changes are ultimately the result of simultaneous changes in both the chromatin state, and gene coexpression patterns. Profiling the chromatin state of forelimb myoblasts at E12.5 revealed a stark change in the chromatin state around neurogenesis and cytoskeletal-related genes, caused by the ablation of a single gene, *Pitx2*. Focusing more specifically on gene expression patterns in forelimb myoblasts from the *Pax3* lineage, we observed significant expression of non-myogenesis related genes. These genes were associated with the immune and vascular systems, and expressed highest at E13.5 and E14.5, later in fetal myogenesis. This was perplexing, since *Pax3* is known to mark all skeletal muscle in the forelimb. Immunohistochemistry, using antibodies for known myogenic lineage markers, revealed the presence of non-myogenic cells in the forelimb at E11.5 and E12.5, further confirming the presence of at least one non-myogenic cell lineage derived from Pax3-positive embryonic myogenic progenitor cells. Taken together, our studies reveal insight into the complexity of mechanisms that overlap during embryonic and fetal myogenesis.

FUTURE DIRECTIONS

With these preliminary data, optimized dissociation and FACS protocol, and the available Cre mouse strains available, we will interrogate the effect of genetic perturbations on embryonic and fetal myogenesis. Our *Pax3*^{Cre} mouse is a knock-in that we could use to generate Pax3-null mice that replicate the *splotch* phenotype, a natural occured *Pax3* mutation. Following the same procedure, we will isolate the deformed myoblast populations from embryonic forelimbs at different stages, and perform differential expression and coexpression analysis. We will then compare the generated co-expression networks, to identify differentially correlated genes and modules that are necessary drivers of myogenesis. Similarly, using the *Pitx2*^{Z} mouse line available in our lab, we will cross *Pax3*^{Cre}|*Rosa*^{EGFP}|*Pitx2*$^{Z/+}$ male mice with *Pitx2*$^{Z/+}$ females, to generate *Pitx2*$^{Z/Z}$ mutants within the Pax3 lineage. Following the same protocol, we will observe the effect of *Pitx2* perturbation on gene expression and coexpression during myogenesis at E11.5, E12.5, E13.5, and E14.5. In these different perturbations we can also perform chromatin profiling via ChIP-Seq, or ATAC-Seq, to observe the chromatin state changes during myogenesis, and caused by perturbations. For Chip-Seq we would use antibodies for Histone H3 Lysine 27 acetylation (H3K27Ac), Mediator 1 (Med1), and p300, all representing *cis*-regulatory modules (CRM). This analysis would allow us to match gene expression changes with non-coding regions of the genome during different stages and perturbations of myogenesis.

Disruptions during myogenesis result in muscle-wasting diseases and myopathies after birth. By completing these studies, we will have a more full understanding of the regulatory mechanisms that occur during different stages of myogenesis. These regulatory mechanisms, identified by stage and perturbation, could be novel therapeutic targets either during myogenesis, or reactivated in adult skeletal muscle to enhance regeneration. The overall goal of this work is to be applied to regenerative medicine. As cells differentiate and become non-mitotic, their cell state changes. Each cell state has a unique signature of chromatin, and gene coexpression. By identifying these signatures in cell types, it becomes possible to apply reverse engineering, and force non-mitotic cells back to a mitotic state. This would ultimately be a boon for regenerative medicine, as embryonic myoblasts could be induced from a patient's own adult skeletal muscle and used to regenerate muscle where needed. Our research will likely have implications beyond muscle-wasting disease, because many disease-associated mutations and single nucleotide polymorphisms (SNPs) are located outside of protein-coding exons, and a large proportion of human genes display expression polymorphism. Our research on the epigenetics of muscle-wasting disease may shed light on other clinically relevant areas.

Chapter 6

Bibliography

Arun J. Singh

BIBLIOGRAPHY

Abe, K., Hashiyama, M., Macgregor, G., and Yamamura, K. i. (1996). Purification of primordial germ cells from TNAPbeta-geo mouse embryos using FACS-gal. Dev. Biol. 180, 468–472.

Abramova, N., Charniga, C., Goderie, S.K., and Temple, S. (2005). Stage-specific changes in gene expression in acutely isolated mouse CNS progenitor cells. Dev. Biol. 283, 269–281.

Adams, N.C., Tomoda, T., Cooper, M., Dietz, G., and Hatten, M.E. (2002). Mice that lack astrotactin have slowed neuronal migration. Development 129, 965–972.

Ahn, R., Gupta, R., Lai, K., Chopra, N., Arron, S.T., and Liao, W. (2016). Network analysis of psoriasis reveals biological pathways and roles for coding and long non-coding RNAs. BMC Genomics 17, 841.

Albert, R., Jeong, H., and Barabasi, A.L. (2000). Error and attack tolerance of complex networks. Nature 406, 378–382.

Alon, U., Surette, M.G., Barkai, N., and Leibler, S. (1999). Robustness in bacterial chemotaxis. Nature 397, 168–171.

Alvarez, M.J., Chen, J.C., and Califano, A. (2015). DIGGIT: a Bioconductor package to infer genetic variants driving cellular phenotypes. Bioinformatics 31, 4032–4034.

Amar, D., Safer, H., and Shamir, R. (2013). Dissection of Regulatory Networks that Are Altered in Disease via Differential Co-expression. PLoS Comput. Biol. 9, e1002955.

Amthor, H., Christ, B., and Patel, K. (1999). A molecular mechanism enabling continuous embryonic muscle growth - a balance between proliferation and differentiation. Development 126, 1041–1053.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169.

Arber, S., Hunter, J.J., Ross, J., Jr, Hongo, M., Sansig, G., Borg, J., Perriard, J.C., Chien, K.R., and Caroni, P. (1997). MLP-deficient mice exhibit a disruption of cardiac cytoarchitectural organization, dilated cardiomyopathy, and heart failure. Cell 88, 393–403.

Arndt, S., Poser, I., Moser, M., and Bosserhoff, A.-K. (2007). Fussel-15, a novel Ski/Sno homolog protein, antagonizes BMP signaling. Mol. Cell. Neurosci. 34, 603–611.

Bader, G.D., and Hogue, C.W.V. (2003). An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 4, 2.

Barabási, A.-L. (2009). Scale-free networks: a decade and beyond. Science 325, 412–413.

Barabási, A.-L., and Albert, R. (1999). Emergence of Scaling in Random Networks. Science 286, 509–512.

Barabasi, A.-L., and Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. Nat. Rev. Genet. 5, 101–113.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell 129, 823–837.

Basso, K., Margolin, A.A., Stolovitzky, G., Klein, U., Dalla-Favera, R., and Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. Nat. Genet. 37, 382–390.

Beer, M.A., and Tavazoie, S. (2004). Predicting gene expression from sequence. Cell 117, 185–198.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125, 315–326.

Biressi, S., Tagliafico, E., Lamorte, G., Monteverde, S., Tenedini, E., Roncaglia, E., Ferrari, S., Ferrari, S., Cusella-De Angelis, M.G., Tajbakhsh, S., et al. (2007). Intrinsic phenotypic diversity of embryonic and fetal myoblasts is revealed by genome-wide gene expression analysis on purified cells. Dev. Biol. 304, 633–651.

Birney, E. (2012). The making of ENCODE: Lessons for big-data projects. Nature 489, 489049a.

Bladt, F., Riethmacher, D., Isenmann, S., Aguzzi, A., and Birchmeier, C. (1995). Essential role for the c-met receptor in the migration of myogenic precursor cells into the limb bud. Nature 376, 376768a0.

Bober, E., Franz, T., Arnold, H.H., Gruss, P., and Tremblay, P. (1994). Pax-3 is required for the development of limb muscles: a possible role for the migration of dermomyotomal muscle progenitor cells. Development 120, 603–612.

Bofill, M., Janossy, G., Lee, C.A., MacDonald-Burns, D., Phillips, A.N., Sabin, C., Timms, A., Johnson, M.A., and Kernoff, P. (1992). Laboratory control values for CD4 and CD8 T lymphocytes. Implications for HIV-1 diagnosis. Clinical & Experimental Immunology 88, 243–252.

Bohl, J., Brimer, N., Lyons, C., and Vande Pol, S.B. (2007). The stardust family protein MPP7 forms a tripartite complex with LIN7 and DLG1 that regulates the stability and localization of DLG1 to cell junctions. J. Biol. Chem. 282, 9392–9400.

Bornholdt, S. (2005). Systems biology. Less is more in modeling large genetic networks. Science 310, 449–451.

Borselli, C., Storrie, H., Benesch-Lee, F., Shvartsman, D., Cezar, C., Lichtman, J.W., Vandenburgh, H.H., and Mooney, D.J. (2010). Functional muscle regeneration with combined delivery of angiogenesis and myogenesis factors. Proc. Natl. Acad. Sci. U. S. A. 107, 3287–3292.

Breckenridge, R.A., Kelly, P., Nandi, M., Vallance, P.J., Ohun, T.J., and Leiper, J. (2010). A role for Dimethylarginine Dimethylaminohydrolase 1 (DDAH1) in mammalian development. Int. J. Dev. Biol. 54, 215–220.

Buckingham, M., and Rigby, P.W.J. (2014). Gene regulatory networks and transcriptional mechanisms that control myogenesis. Dev. Cell 28, 225–238.

Burke, A.C., and Nowicki, J.L. (2003). A new view of patterning domains in the vertebrate mesoderm. Dev. Cell 4, 159–165.

Butte, A.J., and Kohane, I.S. (2000). Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. Pac. Symp. Biocomput. 418–429.

Campbell, A.L., Eng, D., Gross, M.K., and Kioussi, C. (2012). Prediction of gene network models in limb muscle precursors. Gene 509, 16–23.

Carvajal, J.J., Keith, A., and Rigby, P.W.J. (2008). Global transcriptional regulation of the locus encoding the skeletal muscle determination genes Mrf4 and Myf5. Genes Dev. 22, 265–276.

Christ, B., and Ordahl, C.P. (1995). Early stages of chick somite development. Anat. Embryol. 191, 381–396.

Cossu, G., Kelly, R., Tajbakhsh, S., Di Donna, S., Vivarelli, E., and Buckingham, M. (1996). Activation of different myogenic pathways: myf-5 is induced by the neural tube and MyoD by the dorsal ectoderm in mouse paraxial mesoderm. Development 122, 429–437.

Costamagna, D., Costelli, P., Sampaolesi, M., and Penna, F. (2015). Role of Inflammation in Muscle Homeostasis and Myogenesis. Mediators Inflamm. 2015, 805172.

Creixell, P., Schoof, E.M., Simpson, C.D., Longden, J., Miller, C.J., Lou, H.J., Perryman, L., Cox, T.R., Zivanovic, N., Palmeri, A., et al. (2015). Kinome-wide decoding of network-attacking mutations rewiring cancer signaling. Cell 163, 202–217.

Daston, G., Lamar, E., Olivier, M., and Goulding, M. (1996). Pax-3 is necessary for migration but not differentiation of limb muscle precursors in the mouse. Development 122, 1017–1027.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.-H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., et al. (2002). A genomic regulatory network for development. Science 295, 1669–1678.

Davis, S., and Meltzer, P.S. (2007). GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics 23, 1846–1847.

Dawson, J.A., Ye, S., and Kendziorski, C. (2012). R/EBcoexpress: an empirical Bayesian framework for discovering differential co-expression. Bioinformatics 28, 1939–1940.

de la Fuente, A., Bing, N., Hoeschele, I., and Mendes, P. (2004). Discovery of meaningful associations in genomic data using partial correlation coefficients. Bioinformatics 20, 3565–3574.

Deries, M., and Thorsteinsdóttir, S. (2016). Axial and limb muscle development: dialogue with the neighbourhood. Cell. Mol. Life Sci. 73, 4415–4431.

De Rosa, S.C., and Roederer, M. (2001). Eleven-color flow cytometry. A powerful tool for elucidation of the complex immune system. Clin. Lab. Med. 21, 697–712, vii.

De Rosa, S.C., Herzenberg, L.A., Herzenberg, L.A., and Roederer, M. (2001). 11-color, 13-parameter flow cytometry: identification of human naive T cells by phenotype, function, and T-cell receptor diversity. Nat. Med. 7, 245–248.

Dietrich, S., Abou-Rebyeh, F., Brohmann, H., Bladt, F., Sonnenberg-Riethmacher, E., Yamaai, T., Lumsden, A., Brand-Saberi, B., and Birchmeier, C. (1999). The role of SF/HGF and c-Met in the development of skeletal muscle. Development 126, 1621–1629.

Dong, X., Yambartsev, A., Ramsey, S.A., Thomas, L.D., Shulzhenko, N., and Morgun, A. (2015). Reverse enGENEering of Regulatory Networks from Big Data: A Roadmap for Biologists. Bioinform. Biol. Insights 9, 61–74.

Dorrell, C., Schug, J., Lin, C.F., Canaday, P.S., Fox, A.J., Smirnova, O., Bonnah, R., Streeter, P.R., Stoeckert, C.J., Jr, Kaestner, K.H., et al. (2011). Transcriptomes of the major human pancreatic cell types. Diabetologia 54, 2832–2844.

Ecker, J.R., Bickmore, W.A., Barroso, I., Pritchard, J.K., Gilad, Y., and Segal, E. (2012). Genomics: ENCODE explained. Nature 489, 52–55.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. U. S. A. 95, 14863–14868.

Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic Gene Expression in a Single Cell. Science 297, 1183–1186.

Eng, D., Ma, H.-Y., Xu, J., Shih, H.-P., Gross, M.K., Kioussi, C., and Kiouss, C. (2012). Loss of abdominal muscle in Pitx2 mutants associated with altered axial specification of lateral plate mesoderm. PLoS One 7, e42228.

Eng, D., Vogel, W.K., Flann, N.S., Gross, M.K., and Kioussi, C. (2014). Genome-Wide Mapping of Chromatin State of Mouse Forelimbs. Open Access Bioinformatics 6, 1–11.

Engleka, K.A., Gitler, A.D., Zhang, M., Zhou, D.D., High, F.A., and Epstein, J.A. (2005). Insertion of Cre into the Pax3 locus creates a new allele of Splotch and identifies unexpected Pax3 derivatives. Dev. Biol. 280, 396–406.

Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 30, 1575–1584.

Földy, C., Darmanis, S., Aoto, J., Malenka, R.C., Quake, S.R., and Südhof, T.C. (2016). Single-cell RNAseq reveals cell adhesion molecule profiles in electrophysiologically defined neurons. Proc. Natl. Acad. Sci. U. S. A. 113, E5222–E5231.

Fukushima, A. (2013). DiffCorr: an R package to analyze and visualize differential correlations in biological networks. Gene 518, 209–214.

Gage, P.J., Suh, H., and Camper, S.A. (1999). Dosage requirement of Pitx2 for development of multiple organs. Development 126, 4643–4651.

Gambardella, G., Moretti, M.N., de Cegli, R., Cardone, L., Peron, A., and di Bernardo, D. (2013). Differential network analysis for the identification of condition-specific pathway activity and regulation. Bioinformatics 29, 1776–1785.

Gambardella, G., Peluso, I., Montefusco, S., Bansal, M., Medina, D.L., Lawrence, N., and di Bernardo, D. (2015). A reverse-engineering approach to dissect post-translational modulators of transcription factor's activity from transcriptional data. BMC Bioinformatics 16, 279.

Gawad, C., Koh, W., and Quake, S.R. (2016). Single-cell genome sequencing: current state of the science. Nat. Rev. Genet. 17, 175–188.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5, R80.

Giorgi, F.M., Lopez, G., Woo, J.H., Bisikirska, B., Califano, A., and Bansal, M. (2014). Inferring Protein Modulation from Gene Expression Data Using Conditional Mutual Information. PLoS One 9, e109569.

Goulding, M., Lumsden, A., and Paquette, A.J. (1994). Regulation of Pax-3 expression in the dermomyotome and its role in muscle development. Development 120, 957–971.

Grechkin, M., Logsdon, B.A., Gentles, A.J., and Lee, S.-I. (2016). Identifying Network Perturbation in Cancer. PLoS Comput. Biol. 12, e1004888.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell 130, 77–88.

Guitart, X., Bonaventura, J., Rea, W., Orrú, M., Cellai, L., Dettori, I., Pedata, F., Brugarolas, M., Cortés, A., Casadó, V., et al. (2016). Equilibrative nucleoside transporter ENT1 as a biomarker of Huntington disease. Neurobiol. Dis. 96, 47–53.

Gustafsson, M.K., Pan, H., Pinney, D.F., Liu, Y., Lewandowski, A., Epstein, D.J., and Emerson, C.P., Jr (2002). Myf5 is a direct target of long-range Shh signaling and Gli regulation for muscle specification. Genes Dev. 16, 114–126.

Hagman, J., Belanger, C., Travis, A., Turck, C.W., and Grosschedl, R. (1993). Cloning and functional characterization of early B-cell factor, a regulator of lymphocyte-specific gene expression. Genes Dev. 7, 760–773.

Hartwell, L.H., Hopfield, J.J., Leibler, S., and Murray, A.W. (1999). From molecular to modular cell biology. Nature 402, C47–C52.

Hayashi, K., and Ozawa, E. (1995). Myogenic cell migration from somites is induced by tissue contact with medial region of the presumptive limb mesoderm in chick embryos. Development 121, 661–669.

Heanue, T.A., and Pachnis, V. (2006). Expression profiling the developing mammalian enteric nervous system identifies marker and candidate Hirschsprung disease genes. Proc. Natl. Acad. Sci. U. S. A. 103, 6919–6924.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589.

Herzenberg, L.A., Parks, D., Sahaf, B., Perez, O., Roederer, M., and Herzenberg, L.A. (2002). The history and future of the fluorescence activated cell sorter and flow cytometry: a view from Stanford. Clin. Chem. 48, 1819–1827.

Herzenberg, L.A., Sweet, R.G., and Herzenberg, L.A. (1976). Fluorescence-activated cell sorting. Sci. Am. 234, 108–117.

Hou, L., Chen, M., Zhang, C.K., Cho, J., and Zhao, H. (2014). Guilt by rewiring: gene prioritization through network rewiring in genome wide association studies. Hum. Mol. Genet. 23, 2780–2790.

Hsiao, T.-H., Chiu, Y.-C., Hsu, P.-Y., Lu, T.-P., Lai, L.-C., Tsai, M.-H., Huang, T.H.-M., Chuang, E.Y., and Chen, Y. (2016). Differential network analysis reveals the genome-wide landscape of estrogen receptor modulation in hormonal cancers. Sci. Rep. 6, 23035.

Ideker, T., and Krogan, N.J. (2012). Differential network biology. Mol. Syst. Biol. 8, 565.

Ideker, T., Galitski, T., and Hood, L. (2001). A new approach to decoding life: systems biology. Annu. Rev. Genomics Hum. Genet. 2, 343–372.

Ihaka, R., and Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. J. Comput. Graph. Stat. 5, 299–314.

Jeong, H., Mason, S.P., Barabási, A.L., and Oltvai, Z.N. (2001). Lethality and centrality in protein networks. Nature 411, 41–42.

Jiang, X., Zhang, H., and Quan, X. (2016). Differentially Coexpressed Disease Gene Identification Based on Gene Coexpression Network. Biomed Res. Int. 2016, 3962761.

Jimenez, M.A., Akerblad, P., Sigvardsson, M., and Rosen, E.D. (2007). Critical role for Ebf1 and Ebf2 in the adipogenic transcriptional cascade. Mol. Cell. Biol. 27, 743–757.

Jostes, B., Walther, C., and Gruss, P. (1990). The murine paired box gene, Pax7, is expressed specifically during the development of the nervous and muscular system. Mech. Dev. 33, 27–37.

Karolchik, D., Baertsch, R., Diekhans, M., Furey, T.S., Hinrichs, A., Lu, Y.T., Roskin, K.M., Schwartz, M., Sugnet, C.W., Thomas, D.J., et al. (2003). The UCSC Genome Browser Database. Nucleic Acids Res. 31, 51–54.

Kayano, M., Higaki, S., Satoh, J.-I., Matsumoto, K., Matsubara, E., Takikawa, O., and Niida, S. (2016). Plasma microRNA biomarker detection for mild cognitive impairment using differential correlation analysis. Biomark Res 4, 22.

Kiefer, J.C., and Hauschka, S.D. (2001). Myf-5 is transiently expressed in nonmuscle mesoderm and exhibits dynamic regional changes within the presegmented mesoderm and somites I-IV. Dev. Biol. 232, 77–90.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36.

Kioussi, C., Briata, P., Baek, S.H., Rose, D.W., Hamblet, N.S., Herman, T., Ohgi, K.A., Lin, C., Gleiberman, A., Wang, J., et al. (2002). Identification of a Wnt/Dvl/beta-Catenin --> Pitx2 pathway mediating cell-type-specific proliferation during development. Cell 111, 673–685.

Kirschner, M.W. (2005). The meaning of systems biology. Cell 121, 503–504.

Kitamura, K., Miura, H., Miyagawa-Tomita, S., Yanazawa, M., Katoh-Fukui, Y., Suzuki, R., Ohuchi, H., Suehiro, A., Motegi, Y., Nakahara, Y., et al. (1999). Mouse Pitx2 deficiency leads to anomalies of the ventral body wall, heart, extra- and periocular mesoderm and right pulmonary isomerism. Development 126, 5749–5758.

Kitano, H. (2002). Systems biology: a brief overview. Science 295, 1662–1664.

Kolde, R. (2012). Pheatmap: pretty heatmaps. R Package Version 61.

Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C., and Teichmann, S.A. (2015). The technology and biology of single-cell RNA sequencing. Mol. Cell 58, 610–620.

Kong, Y., Flick, M.J., Kudla, A.J., and Konieczny, S.F. (1997). Muscle LIM protein promotes myogenesis by enhancing the activity of MyoD. Mol. Cell. Biol. 17, 4750–4760.

Lagha, M., Brunelli, S., Messina, G., Cumano, A., Kume, T., Relaix, F., and Buckingham, M.E. (2009). Pax3:Foxc2 reciprocal repression in the somite modulates muscular versus vascular cell fate choice in multipotent progenitors. Dev. Cell 17, 892–899.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature 409, 860–921.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.

Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., et al. (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. Science 298, 799–804.

Li, F., Long, T., Lu, Y., Ouyang, Q., and Tang, C. (2004). The yeast cell-cycle network is robustly designed. Proc. Natl. Acad. Sci. U. S. A. 101, 4781–4786.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.

Lin, C.R., Kioussi, C., O'Connell, S., Briata, P., Szeto, D., Liu, F., Izpisúa-Belmonte, J.C., and Rosenfeld, M.G. (1999). Pitx2 regulates lung asymmetry, cardiac positioning and pituitary and tooth morphogenesis. Nature 401, 279–282.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133, 523–536.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550.

Lu, M.F., Pressman, C., Dyer, R., Johnson, R.L., and Martin, J.F. (1999). Function of Rieger syndrome gene in left-right asymmetry and craniofacial development. Nature 401, 276–278.

Luo, Q., Kang, Q., Si, W., Jiang, W., Park, J.K., Peng, Y., Li, X., Luu, H.H., Luo, J., Montag, A.G., et al. (2004). Connective tissue growth factor (CTGF) is regulated by Wnt and bone morphogenetic proteins signaling in osteoblast differentiation of mesenchymal stem cells. J. Biol. Chem. 279, 55958–55968.

Ma, H.-Y., Xu, J., Eng, D., Gross, M.K., and Kioussi, C. (2013). Pitx2-mediated cardiac outflow tract remodeling. Dev. Dyn. 242, 456–468.

Maldonado-Báez, L., Cole, N.B., Krämer, H., and Donaldson, J.G. (2013). Microtubule-dependent endosomal sorting of clathrin-independent cargo by Hook1. J. Cell Biol. 201, 233–247.

Mao, X., Fujiwara, Y., Chapdelaine, A., Yang, H., and Orkin, S.H. (2001). Activation of EGFP expression by Cre-mediated excision in a new ROSA26 reporter mouse strain. Blood 97, 324–326.

Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., and Califano, A. (2006). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics 7 Suppl 1, S7.

Mariani, M., Baldessari, D., Francisconi, S., Viggiano, L., Rocchi, M., Zappavigna, V., Malgaretti, N., and Consalez, G.G. (1999). Two murine and human homologs of mab-21, a cell fate determination gene involved in Caenorhabditis elegans neural development. Hum. Mol. Genet. 8, 2397–2406.

Martin, P. (1990). Tissue patterning in the developing mouse limb. Int. J. Dev. Biol. 34, 323–336.

Masuda, T., Yaginuma, H., Sakuma, C., and Ono, K. (2009). Netrin-1 signaling for sensory axons: Involvement in sensory axonal development and regeneration. Cell Adh. Migr. 3, 171–173.

Ma, X., Gao, L., Karamanlidis, G., Gao, P., Lee, C.F., Garcia-Menendez, L., Tian, R., and Tan, K. (2015). Revealing Pathway Dynamics in Heart Diseases by Analyzing Multiple Differential Networks. PLoS Comput. Biol. 11, e1004332.

Mayeuf-Louchart, A., Montarras, D., Bodin, C., Kume, T., Vincent, S.D., and Buckingham, M. (2016). Endothelial cell specification in the somite is compromised in Pax3-positive progenitors of Foxc1/2 conditional mutants, with loss of forelimb myogenesis. Development 143, 872–879.

McKenzie, A.T., Katsyv, I., Song, W.-M., Wang, M., and Zhang, B. (2016). DGCA: A comprehensive R package for Differential Gene Correlation Analysis. BMC Syst. Biol. 10, 106.

Mella, S., Soula, C., Morello, D., Crozatier, M., and Vincent, A. (2004). Expression patterns of the coe/ebf transcription factor genes during chicken and mouse limb development. Gene Expr. Patterns 4, 537–542.

Messina, G., Biressi, S., Monteverde, S., Magli, A., Cassano, M., Perani, L., Roncaglia, E., Tagliafico, E., Starnes, L., Campbell, C.E., et al. (2010). Nfix regulates fetal-specific transcription in developing skeletal muscle. Cell 140, 554–566.

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., and Thomas, P.D. (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. 45, D183–D189.

Mi, H., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2013). Large-scale gene function analysis with the PANTHER classification system. Nat. Protoc. 8, 1551–1566.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. Science 298, 824–827.

Mitra, K., Carvunis, A.-R., Ramesh, S.K., and Ideker, T. (2013). Integrative approaches for finding modular structure in biological networks. Nat. Rev. Genet. 14, 719–732.

Mizuhara, E., Nakatani, T., Minaki, Y., Sakamoto, Y., and Ono, Y. (2005). Corl1, a novel neuronal lineage-specific transcriptional corepressor for the homeodomain transcription factor Lbx1. J. Biol. Chem. 280, 3645–3655.

Mofarrahi, M., McClung, J.M., Kontos, C.D., Davis, E.C., Tappuni, B., Moroz, N., Pickett, A.E., Huck, L., Harel, S., Danialou, G., et al. (2015). Angiopoietin-1 enhances skeletal muscle regeneration in mice. Am. J. Physiol. Regul. Integr. Comp. Physiol. 308, R576–R589.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5, 621–628.

Oros Klein, K., Oualkacha, K., Lafond, M.-H., Bhatnagar, S., Tonin, P.N., and Greenwood, C.M.T. (2016). Gene Coexpression Analyses Differentiate Networks Associated with Diverse Cancers Harboring TP53 Missense or Null Mutations. Front. Genet. 7, 137.

Park, P.J. (2009). ChIP-seq: advantages and challenges of a maturing technology. Nat. Rev. Genet. 10, 669–680.

Parks, D.R., Bryan, V.M., Oi, V.T., and Herzenberg, L.A. (1979). Antigen-specific identification and cloning of hybridomas with a fluorescence-activated cell sorter. Proc. Natl. Acad. Sci. U. S. A. 76, 1962–1966.

Perfetto, S.P., Chattopadhyay, P.K., and Roederer, M. (2004). Seventeen-colour flow cytometry: unravelling the immune system. Nat. Rev. Immunol. 4, 648–655.

Peri, L.E., Sanders, K.M., and Mutafova-Yambolieva, V.N. (2013). Differential expression of genes related to purinergic signaling in smooth muscle cells, PDGFRα-positive cells, and interstitial cells of Cajal in the murine colon. Neurogastroenterol. Motil. 25, e609–e620.

Pineault, K.M., and Wellik, D.M. (2014). Hox genes and limb musculoskeletal development. Curr. Osteoporos. Rep. 12, 420–427.

Prill, R.J., Iglesias, P.A., and Levchenko, A. (2005). Dynamic Properties of Network Motifs Contribute to Biological Network Organization. PLoS Biol. 3, e343.

Raines, A.M., Magella, B., Adam, M., and Potter, S.S. (2015). Key pathways regulated by HoxA9,10,11/HoxD9,10,11 during limb development. BMC Dev. Biol. 15, 28.

Ramsey, S.A., Klemm, S.L., Zak, D.E., Kennedy, K.A., Thorsson, V., Li, B., Gilchrist, M., Gold, E.S., Johnson, C.D., Litvak, V., et al. (2008). Uncovering a Macrophage Transcriptional Program by Integrating Evidence from Motif Scanning and Expression Dynamics. PLoS Comput. Biol. 4, e1000021.

Ratushny, A.V., Saleem, R.A., Sitko, K., Ramsey, S.A., and Aitchison, J.D. (2012). Asymmetric positive feedback loops reliably control biological responses. Mol. Syst. Biol. 8, 577.

Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., and Barabási, A.L. (2002). Hierarchical organization of modularity in metabolic networks. Science 297, 1551–1555.

Reiss, D.J., Baliga, N.S., and Bonneau, R. (2006). Integrated biclustering of heterogeneous genome-wide datasets for the inference of global regulatory networks. BMC Bioinformatics 7, 280.

Reyes-Bermudez, A., Villar-Briones, A., Ramirez-Portilla, C., Hidaka, M., and Mikheyev, A.S. (2016). Developmental Progression in the Coral Acropora digitifera Is Controlled by Differential Expression of Distinct Regulatory Gene Networks. Genome Biol. Evol. 8, 851–870.

Richardson, G.M., Lannigan, J., and Macara, I.G. (2015). Does FACS perturb gene expression? Cytometry A 87, 166–175.

Ringnér, M. (2008). What is principal component analysis? Nat. Biotechnol. 26, 303–304.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

Saclier, M., Cuvellier, S., Magnan, M., Mounier, R., and Chazaud, B. (2013). Monocyte/macrophage interactions with myogenic precursor cells during skeletal muscle regeneration. FEBS J. 280, 4118–4130.

Saraph, V., and Milenković, T. (2014). MAGNA: Maximizing Accuracy in Global Network Alignment. Bioinformatics 30, 2931–2940.

Scarpa, J.R., Jiang, P., Losic, B., Readhead, B., Gao, V.D., Dudley, J.T., Vitaterna, M.H., Turek, F.W., and Kasarskis, A. (2016). Systems Genetic Analyses Highlight a TGFβ-FOXO3 Dependent Striatal Astrocyte Network Conserved across Species and Associated with Stress, Sleep, and Huntington's Disease. PLoS Genet. 12, e1006137.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504.

Shih, H.P., Gross, M.K., and Kioussi, C. (2008). Muscle development: forming the head and trunk muscles. Acta Histochem. 110, 97–108.

Singh, A.J., Ramsey, S.A., Filtz, T.M., and Kioussi, C. (2017). Differential gene regulatory networks in development and disease. Cell. Mol. Life Sci.

Siska, C., Bowler, R., and Kechris, K. (2015). The Discordant Method: A Novel Approach for Differential Correlation. Bioinformatics.

Smyth, G.K. (2005). limma: Linear Models for Microarray Data. In Bioinformatics and Computational Biology Solutions Using R and Bioconductor, (Springer New York), pp. 397–420.

Southworth, L.K., Owen, A.B., and Kim, S.K. (2009). Aging Mice Show a Decreasing Correlation of Gene Expression within Genetic Modules. PLoS Genet. 5, e1000776.

Tajbakhsh, S. (2005). Skeletal muscle stem and progenitor cells: reconciling genetics and lineage. Exp. Cell Res. 306, 364–372.

Tesson, B.M., Breitling, R., and Jansen, R.C. (2010). DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. BMC Bioinformatics 11, 497.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–1111.

Treutlein, B., Lee, Q.Y., Camp, J.G., Mall, M., Koh, W., Shariati, S.A.M., Sim, S., Neff, N.F., Skotheim, J.M., Wernig, M., et al. (2016). Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. Nature 534, 391–395.

Troy, N.M., Hollams, E.M., Holt, P.G., and Bosco, A. (2016). Differential gene network analysis for the identification of asthma-associated therapeutic targets in allergen-specific T-helper memory responses. BMC Med. Genomics 9, 9.

Van Dongen, S.M. (2001). Graph clustering by flow simulation.

Wang, K., Saito, M., Bisikirska, B.C., Alvarez, M.J., Lim, W.K., Rajbhandari, P., Shen, Q., Nemenman, I., Basso, K., Margolin, A.A., et al. (2009). Genome-wide identification of post-translational modulators of transcription factor activity in human B cells. Nat. Biotechnol. 27, 829–839.

Watson, M. (2006). CoXpress: differential co-expression in gene expression data. BMC Bioinformatics 7, 509.

Wong, Y.-M., and Chow, K.L. (2002). Expression of zebrafish mab21 genes marks the differentiating eye, midbrain and neural tube. Mech. Dev. 113, 149–152.

Wu, S., Li, J., Cao, M., Yang, J., Li, Y.-X., and Li, Y.-Y. (2016). A novel integrated gene coexpression analysis approach reveals a prognostic three-transcription-factor signature for glioma molecular subtypes. BMC Syst. Biol. 10 Suppl 3, 71.

Wu, Y.E., Pan, L., Zuo, Y., Li, X., and Hong, W. (2017). Detecting Activated Cell Populations Using Single-Cell RNA-Seq. Neuron 96, 313–329.e6.

Young, K., Krebs, L.T., Tweedie, E., Conley, B., Mancini, M., Arthur, H.M., Liaw, L., Gridley, T., and Vary, C.P.H. (2016). Endoglin is required in Pax3-derived cells for embryonic blood vessel formation. Dev. Biol. 409, 95–105.

Yusuf, F., and Brand-Saberi, B. (2012). Myogenesis and muscle regeneration. Histochem. Cell Biol. 138, 187–199.

Zahavy, E., Ber, R., Gur, D., Abramovich, H., Freeman, E., Maoz, S., and Yitzhaki, S. (2012). Application of nanoparticles for the detection and sorting of pathogenic bacteria by flow-cytometry. Adv. Exp. Med. Biol. 733, 23–36.

Zhang, X., Zhao, J., Hao, J.-K., Zhao, X.-M., and Chen, L. (2015). Conditional mutual inclusive information enables accurate quantification of associations in gene regulatory networks. Nucleic Acids Res. 43, e31.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137.

Zuo, Z.-G., Zhang, X.-F., Ye, X.-Z., Zhou, Z.-H., Wu, X.-B., Ni, S.-C., and Song, H.-Y. (2016). Bioinformatic analysis of RNA-seq data unveiled critical genes in rectal adenocarcinoma. Eur. Rev. Med. Pharmacol. Sci. 20, 3017–3025.