12-2012

IDENTIFYING GENETIC VARIANTS AND CHARACTERIZING THEIR ROLE IN CLUBFOOT

Katelyn S. Weymouth

IDENTIFYING GENETIC VARIANTS AND CHARACTERIZING THEIR ROLE IN CLUBFOOT

By

Katelyn S. Weymouth, B.S., M.S.

APPROVED:

_____
Jacqueline T. Hecht, Ph.D.
Supervisory Professor

_____
Joseph Alcorn, Ph.D.

_____
Rebecca Berdeaux, Ph.D.

_____
Gilbert Cote, Ph.D.

_____
Stephen P. Daiger, Ph.D.

APPROVED:

_____
Dean, The University of Texas
Graduate School of Biomedical Sciences

_____
Dean, The University of Texas
Graduate School of Biomedical Sciences

IDENTIFYING GENETIC VARIANTS AND CHARACTERIZING THEIR ROLE IN CLUBFOOT

A
DISSERTATION

Presented to the Faculty of
The University of Texas Health Science Center at Houston
and
The University of Texas M. D. Anderson Cancer Center
Graduate School of Biomedical Sciences

In Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

by

Katelyn S. Weymouth, B.S., M.S.

Houston, Texas
December, 2012

DEDICATION

I would like to dedicate this work to my parents for their continued love and support they have provided me through my years. In addition, this work is dedicated to Dr. Jacqueline T. Hecht for her continued support and guidance in molding me into the scientist I am today. Thank you!

ACKNOWLEDGEMENTS

I have so many people to thank that aided in my success and completion of my PhD, along with being where I am today. First and for most, I must thank my wonderful parents. I cannot remember a day in which they were not there for me. Their continued love, support and guidance have molded me into the person I am today. My only wish in life is to continue to make them proud in all my future endeavors. In addition, I want to thank my extended family, my aunts, uncles and cousins, especially my Aunt Marlene for being a true model of living one's life to the fullest and enjoying every moment of it. Thank you to my grandma and grandpa, Alice and Harold Songin for providing me with such a loving atmosphere filled with many great stories and family traditions that I hope to continue on for years to come. And I can't forget to thank my grandpa for passing on those stubborn Pollack genes to me. They have come in handy, now if I could only get him to stop mowing the lawn at the age of 83.

Dr. Jacqueline T. Hecht, thank you for all your guidance and being a role model to me as a woman in science and seeing the true potential of a career in science. I know at times you may have wanted to ring my neck but I hope at the end of the day I have made you proud. I, also, hope as being your last PhD graduate student (even though I know you will have a hard time saying no to other potential students) that you were able to go out on a high note even though as you said "you have to pull me out of the garbage" from time to time. Well, I hope I ended up being a diamond in the rough. In addition, I would like to thank my committee members (Dr. Alcorn, Dr. Berdeaux, Dr. Cote and Dr. Daiger) for their continued guidance and support, along with Dr. Blanton for all her statistical knowledge.

To the Hecht lab both past and present, thank you for everything you bestowed upon me. I know you will miss my wonderful singing and dancing but at least now with the lack of my baked goods you can all not worry about weight gain. Also, thank you all for acting like second, third, fourth, etc. mothers to me (Shea, Wei, Peiman, Francoise, Miss P). To Brett, my older/younger brother, thank you for the Friday Starbuck mornings, the consistent self-esteem checks, the reality television reviews, finding my stories to always be entertaining and of course always answering those silly questions I typically ask and not making me feel AS silly for asking them. To Nikki, thank you for always being real with me and being your true self, along with your constant energy. To Tamar and Candace, thank you for your many great lab discussions, instant dance parties and many ridiculous memories with hopefully many more to come. To Bryan or as I know him better, Overachiever, thank you for our lunch dates filled with great discussions and allowing me to feel like part of your family with your amazing wife, Kristin and adorable daughter, Addy. Of course, I cannot forget many great Baylor ball memories especially with Tiana, my fatty partner. To John (Sunshine), thank you for your constant "sunny" disposition and answering all my continuous experimental questions no matter how ridiculous they may have been.

And last but not least I cannot forget my friends from back home that I have known since I was in kindergarten. It has been very hard to be so far away from them but it has made visits back home mean that much more to me. To my girls, Jules and Bren, I would be lost without the two of you and look forward to many more years of memories and watching your families grow. Melissa, I would have never been able to return to UConn without you. You have an absolutely wonderful family and I look forward to watching Samantha and Alexa grow up to be beautiful, success women like their mother. Alissa, my sports partner

in crime, thank you for always being up for whatever adventure I throw at you especially when it was sports related. I look forward to many more adventures at whatever locale we choose. Brooke, I love how we may not talk for months at a time but are able to pick right up where we left off as if no time has passed and that I truly cherish. And finally to my boys, especially Jeff, Mezzoni, Brif, Daniel and Seanathan, thanks for always looking out for me and being like brothers to me. And to all those that I was unable to mention by name, I have not forgotten about you, nor will I ever and I truly thank you for being a part of my life and aiding in my ability to "live my life to the fullest" as I have so successful done thus far and will continue to do both personally and scientifically.

Identifying Genetic Variants and Characterizing Their Role in Clubfoot

Katelyn S. Weymouth

Advisor: Jacqueline T. Hecht, Ph.D.

Clubfoot is a common, complex birth defect affecting 4,000 newborns in the United States and 135,000 world-wide each year. The clubfoot deformity is characterized by inward and rigid downward displacement of one or both feet, along with persistent calf muscle hypoplasia. Despite strong evidence for a genetic liability, there is a limited understanding of the genetic and environmental factors contributing to the etiology of clubfoot. *The studies described in this dissertation were performed to identify variants and/or genes associated with clubfoot.* Genome-wide linkage scan performed on ten multiplex clubfoot families identified seven new chromosomal regions that provide new areas to search for clubfoot genes. Troponin C (*TNNC2*) the strongest candidate gene, located in 20q12-q13.11, is involved in muscle contraction. Exon sequencing of *TNNC2* did not identify any novel coding variants. Interrogation of fifteen muscle contraction genes found strong associations with SNPs located in potential regulatory regions of *TPM1* (rs4075583 and rs3805965), *TPM2* (rs2025126 and rs2145925) and *TNNC2* (rs383112 and rs437122). In previous studies, a strong association was found with rs3801776 located in the basal promoter of *HOXA9*, a gene also involved in muscle development and patterning. Altogether, this data suggests that SNPs located in potential regulatory regions of genes involved in muscle development and function could alter transcription factor binding leading to changes in gene expression. Functional analysis of 3801776/*HOXA9,* rs2025126/*TPM2* and rs2145925/*TPM2* showed altered protein binding, which significantly influenced promoter activity. Although the ancestral allele (G) of rs4075583/*TPM1* creates a DNA-

protein complex, it did not affect *TPM1* promoter activity. However and importantly, in the context of a haplotype, rs4075583/G significantly decreased *TPM1* promoter activity. These results suggest dysregulation of multiple skeletal muscle genes, *TPM1*, *TPM2*, *TNNC2* and *HOXA9,* working in concert may contribute to clubfoot. However, specific allelic combinations involving these four regulatory SNPs did not confer a significantly higher risk for clubfoot. Other combinations of these variants are being evaluated. Moreover, these variants may interact with yet to be discovered variants in other genes to confer a higher clubfoot risk. Collectively, we show novel evidence for the role of skeletal muscle genes in clubfoot indicating that there are multiple genetic factors contributing to this complex birth defect.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

Actin, alpha 1, skeletal muscle ACTA1

Alternate ..... ALT

Ancestral ..... ANC

Anterior-posterior ..... AP

Apical ectodermal ridge ..... AER

Apoptosis-related cysteine peptidase ..... CASP

Apoptotic peptidase activating factor 1 ..... APAF1

Association in the presence of linkage ..... APL

Base pair ..... BP

BCL2-antagonist/killer ..... BAK

BCL2-associated X protein ..... BAX

BH3 interacting domain death agonist ..... BID

Bone morphogenetic protein ..... BMP

BRCA1 interacting protein C-terminal helicase 1 ..... BRIP1

CASP8 and FADD-like apoptosis regulator ..... CFLAR

Chromodomain helicase DNA binding protein ..... CHD

Chromosome ..... CHR

Chromosomal microarray analysis ..... CMA

Comparative genomic hybridization ..... CGH

Confidence interval CI

Congenital talipes equinovarus ..... CTEV

Copy number variant ..... CNV

Cytochrome P450, family 1, subfamily A, polypeptide 1 ..... CYP1A1

Cytochrome P450, family 2, subfamily A, polypeptide 6 ..... CYP2A6

Deoxyribonucleic acid ..... DNA

Differentiated ..... DIF

Distal Arthrogyrposis ..... DA

Dorsal-ventral ..... DV

Engrailed homeobox 1 ..... EN1

Family History ..... FH

Fas-associated via death domain ..... FADD

Fibroblast growth factor ..... FGF

Generalized estimating equations ..... GEE

Genome wide association scan ..... GWAS

Genotype pedigree disequilibrium test ..... GENO-PDT

GLI family zinc finger 3 ..... GLI3

Graphical overview of linkage disequilibrium ..... GOLD

Gremlin 1 ..... GREM1

Hardy-Weinberg equilibrium ..... HWE

Homeobox gene ..... HOX

Institutional review board ..... IRB

Insulin-like growth factor ..... IGF

Insulin-like growth factor binding protein ..... IGFBP

Interdigital necrotic zone ..... INZ

Lateral plate mesoderm ..... LPM

Likelihood ratio test ..... LRT

LIM homeobox transcription factor 1-beta LMXB1

Linkage disequilibrium ..... LD

Mexican Ancestry ..... MXL

Multipoint Engine for Rapid Likelihood Inferences ..... MERLIN

Mothers against DPP homolog ..... SMAD

Muscle progenitor cells..... MPC

Myogenic differentiation ..... MYOD

Myogenic factor..... MYF

Myogenic regulatory transcription factor ..... MRF

Myogenin ..... MYOG

Myosin binding protein C..... MYBPC

Myosin binding protein H..... MYBPH

Myosin heavy chain ..... MYH

Myosin, light polypeptide 1 ..... MYL1

N-acetyltransferase 2 ..... NAT2

Nascent polypeptide-associated complex alpha subunit 2..... NACA2

National Center for Biotechnology Information..... NCBI

Neutralized homolog 2 ..... NEURL2

nonHispanic white ..... NHW

Opaque patch ..... OP

Paired box ..... PAX

Paired-like homeodomain factor..... PITX

Pedigree disequilibrium test ..... PDT

Peroneal muscular atrophy ..... PMA

Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2..... PLOD2

Polymerase chain reaction ..... PCR

Programmed cell death ..... PCD

Progressive zone ..... PZ

Proximal-distal ..... PD

Relative Risk ..... RR

Retinoic acid ..... RA

Retinoic acid receptor ..... RAR

Retinoid X receptor ..... RXR

Scatter factor/hepatocyte growth factor ..... SF/HGF

Sex determining region Y box 14 ..... SOX14

Short tandem repeat ..... STR

Single nucleotide polymorphism ..... SNP

Sonic hedgehog ..... SHH

Standard deviation ..... SD

Statistical Analysis Software ..... SAS

T-box transcription factor ..... TBX

Transmission disequilibrium test ..... TDT

Tropomyosin ..... TPM

Troponin C TNNC2

Troponin I ..... TNNI

Troponin T ..... TNNT

Ubiquitously transcribed X chromosome tetratricopeptide repeat protein ..... UTX

Undifferentiated ..... UNDIF

Upstream regulatory element ..... URE

Wingless-type MMTV integration site family ..... WNT

Zone of polarizing activity ..... ZPA

CHAPTER 1: Introduction

1.0 Introduction

Idiopathic congential talipes equinovarus, more commonly referred to as clubfoot, is one of the most common, serious congenital musculoskeletal anomalies. Clubfoot is characterized by the inward and rigid downward displacement of one or both feet[1]. In addition, calf muscles in the affected leg(s) are underdeveloped and remain small even after corrective treatment[2,3]. The etiology of clubfoot is complex involving both genetic and environmental factors. The genetic variation(s) contributing to this birth defect is largely unknown. Hindlimb and muscle development is a multifaceted well-orchestrated process involving cell migration, proliferation, patterning and apoptosis[4-9]. These processes are tightly regulated by multiple factors such as growth factors, signaling molecules and transcription factors[4-9]. Dysregulation of genes involved in limb and muscle development could contribute to clubfoot. This chapter reviews hindlimb and muscle development and provides a comprehensive discussion of putative etiological mechanisms contributing to the etiology of clubfoot.

1.1 Hindlimb development

Hindlimb development is an intricate process involving cell proliferation, migration, patterning and programmed cell death, all being regulated by multiple factors such as growth factors, transcription factors and signaling molecules[4-9]. Hindlimb development begins during week three of gestation with swelling along the lateral plate mesoderm (LPM) initiated by fibroblast growth factor 8 (*FGF8*) which in turn initiates the expression of fibroblast growth factor 10 (*FGF10*) creating a *FGF8-FGF10* positive feedback loop[10,11]. In addition, T-box transcription factor 4, *TBX4,* and paired-like homeodomain factor, *PITX1,*

are important for hindlimb induction[6-9]. By the middle of the fourth week gestation, the limb bud is composed of mesenchyme derived from the LPM (source of skeletal components) and from the myotomes of the paraxial mesoderm (source of muscular components). The processes that transform the limb bud to a mature limb are commonly defined by three major axes: proximal-distal (PD, running from knee to digits), anterior-posterior (AP, running from the big toe to the little toe) and dorsal-ventral (DV, running from the top of the foot to the bottom of the foot)(Figure 1.1). Three key signaling centers are required for proper growth and patterning along the three axes: the apical ectodermal ridge (AER), the zone of polarizing activity (ZPA) and the dorsal ectoderm[10].

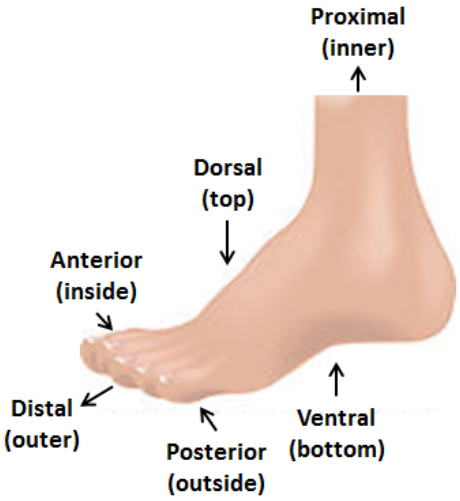

Figure 1.1. Axes of the foot
The three major axes of the foot are shown: Anterior-Posterior, Dorsal-Ventral and Proximal-Distal

1.1.1 Proximodistal patterning

Once the limb bud is formed, limb outgrowth is regulated along the proximal-distal axis by the AER. Thickening of the ectoderm at the distal end of the limb bud forms the AER containing proliferating, undifferentiated cells expressing four *FGF*s: *FGF4, 8, 9* and *17* [6,12,13]. Currently, three models are proposed to explain proximodistal patterning during limb ougrowth[10]. The 'progress zone' (PZ) model suggests that proximal-distal patterning develops progressively, whereby more distal structures are patterned sequentially (Figure 1.3A)[14]. The first mesenchymal cells that leave the PZ will form the more proximal

structures, while the later mesenchymal cells will form distal structures. The 'early-specification model' proposes that cells are predetermined to form the stylopod (femur), zeugopod (tibia and fibula) and autopod (tarsals, metatarsals and phalanges)[10]. The 'two-signal model' suggests that limb mesenchymal cells are initially exposed to a proximal signal from the lateral mesoderm such as retinoic acid (RA) that initiates expression of Meis homeobox transcription factors which influence proximal development (Figure 1.2B)[15-17]. The AER expresses an opposing signal such as FGFs to establish a distal domain and these opposing cellular domains form the zeugopod of the hindlimb[18,19].

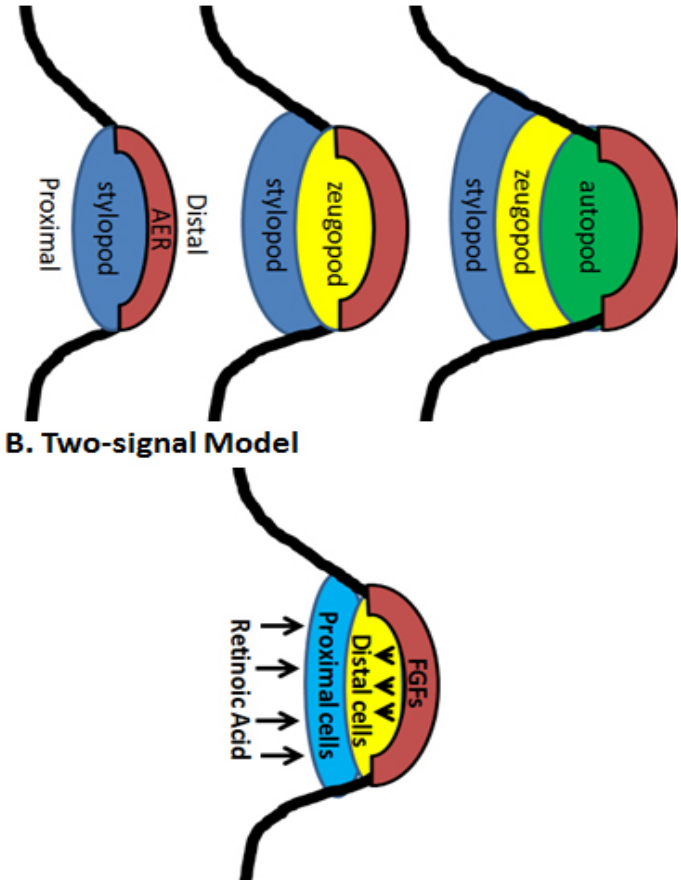

Figure 1.2. Proximodistal Models
A. In the progess zone (PZ) model, initial cells to leave the PZ will form the more proximal structures, while later cells form the distal structures B. In the two-signal model, proximal and distal domains are established through RA and FGF and determines mesenchymal cell destination. In the early-specification model, cells are predetermined to form the various hindlimb elements (not pictured).

1.1.2 Anteroposterior patterning

Anteroposterior patterning is regulated by a group of mesodermal cells located at the posterior region of the limb bud known as the ZPA[10]. Sonic hedgehog (*SHH*), a morphogen, is produced in the ZPA and controls number and type of digits. Posteriorly expressed homeobox genes, *HOXD10-13,* initiate *SHH* expression; later in limb development these homoebox box genes become *SHH*-dependent[20,21]. Three different models have been suggested to explain the patterning mechanism along the anterior-posterior axis of the limb[12,22,23]. The classic morphogen model suggests *SHH* creates a posterior-anterior gradient, where cells acquire a positional value that changes over time (Figure 1.3A)[10,11]. As a consequence of the established *SHH* gradient, *GLI3* is expressed[10,11]. The repressor activity of *GLI3* causes a polarizing activity with *SHH*[10,11]. The temporal expansion model incorporates *SHH* exposure time and concentration (Figure 1.4B)[10,11]. More posterior digital features have longer exposure to higher concentrations of *SHH*, whereas low concentrations of *SHH* for a shorter period of

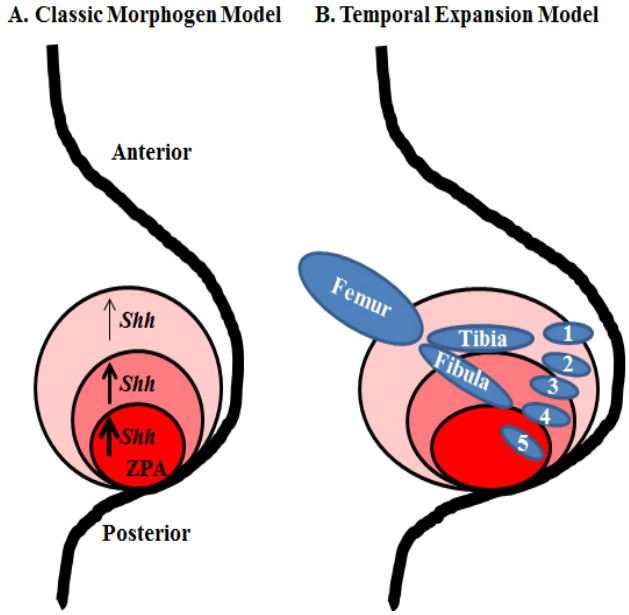

Figure 1.3. Anteroposterior Models
A. In the classic morphogen model, *SHH* concentration gradient determines A-P limb development. B. In the temporal expansion model, *SHH* concentration and time determines A-P limb development. In the biphasic model, early limb development requires the *SHH* concentration gradient, while later digit development only requires *SHH* as a mitogen (not pictured).
1,2,3,4,5 indicates digit number

time promote more anterior digital features. The third model, the biphasic model, proposes that digit specification occurs at very early stages of limb development by a *SHH* concentration gradient, while later limb development only requires *SHH* as a mitogen for digit formation[10,11].

1.1.3 Dorsoventral patterning

Dorsoventral patterning is controlled by dorsal and ventral ectoderm derived signaling[6]. In the dorsal ectoderm, *WNT7A* signals to the dorsal mesenchyme regulating a LIM-homeodomain transcription factor, *LMX1B,* to establish dorsal patterning[24-26]. While in the ventral ectoderm, members of the BMP family regulate *EN1,* a homeobox-containing transcription factor to establish ventral patterning[27].

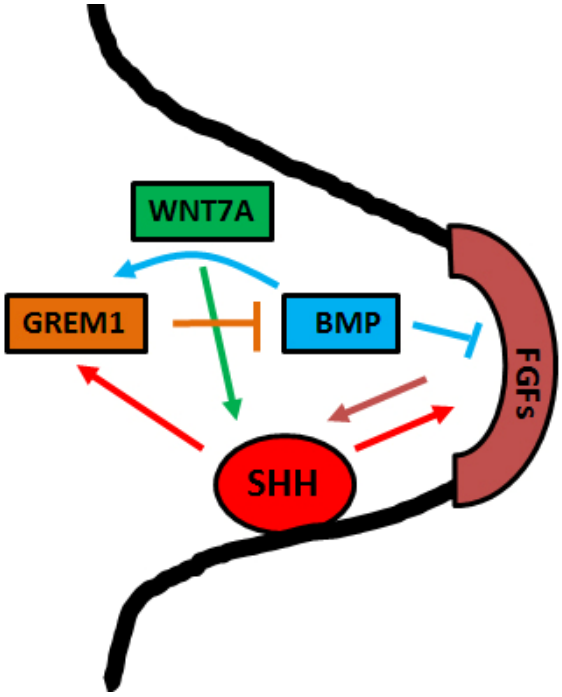

Figure 1.4. Limb development signaling Interaction of the factors that regulate the three axes is important for correct limb

1.1.4 Coordination of the three axes

Multifaceted maintenance of each individual axis is required for limb formation. As shown in Figure 1.4, coordination and interaction between the three core signaling centers of the three axes allow for correct limb morphogenesis. *SHH* is required to regulate *FGF8* expression and maintenance of *FGF4, 9* and *17*[6]. In turn, *FGF4* in the posterior portion of AER and *WNT7A* in the dorsal ectoderm sustain *SHH* expression in the ZPA[10]. *SHH* positively regulates *GREM1*, a BMP antagonist that blocks BMP inhibitory action on AER-FGF expression[10]. BMP signals modulate AER-

FGF activity by negatively regulating FGF gene expression in the AER. These precisely coordinated interactions allow for accurate hindlimb development and dysregulation of any of these key signaling factors could result in a hindlimb malformation.

1.1.5 Programmed cell death (PCD)

Programmed cell death, also called apoptosis, is the process in which unwanted cells are actively removed[4]. Apoptosis plays an important role in limb development including establishment of prechondrogenic condensation, removal of ectoderm along the AER, joint formation, establishment of axon pathways and remodeling of the vascular pattern[4].

Apoptosis is involved in the formation of the zeugopod elements, digit formation and AER ectoderm, allowing for correct limb shape and skeletal patterning[4]. Skeletal formation of the tibia and fibula is initiated by apoptosis in the opaque patch (OP), the central mesenchyme of the limb bud[28]. Massive apoptosis along the interdigital mesoderm known as the interdigital necrotic zone (INZ) is needed for digit formation and shape[29,30]. In addition to apoptosis of mesodermal cells, cell death occurs along the ectoderm of the AER and inhibition or delay of apoptosis in the AER results in polydactyly[31].

Common mechanisms and pathways known to regulate proliferation and differentiation are involved in programmed cell death. BMPs, bone morphogentic proteins, in particular BMP2, 4 and 7, are known to trigger apoptosis in both mesodermal cells and in the AER ectoderm[5]. In addition, BMPs are involved in chondrogenic differentiation mediated through the BMP type Ib receptor, while initiation of activated BMP apoptosis remains unknown[4,32,33]. Multiple pathways have been identified which regulate BMP-initiated apoptosis. For example, BMP activation can be initiated through SMAD1, SMAD5 and SMAD8; however, *SMAD1* and *SMAD8* are only expressed in the interdigital tissue[10].

Phosphorylated SMAD levels coincide with the onset of programmed cell death, thus cell death in the INZ may be mediated through BMP-SMAD signaling[4]. Conversely, BMP activation can initiate a protective response through FGF signaling in interdigital mesenchyme[10]. Decreases in *FGF8* correlate with premature cell death via BMP signaling[10]. However, inactivation of BMP signaling leads to increased expression of *FGF8* in the AER[10]. Multiple factors are also known to regulate BMP-signaling that modulate intensity and/or spatial distribution. For example, *GREM1*, a BMP antagonist, expresses in a pattern to protect undifferentiated mesoderm from apoptosis[4].

In addition to BMP-regulated apoptosis, RA is involved in regulating apoptosis[10]. RA, a derivative of retinol (vitamin A) metabolism, is mediated by two types of nuclear receptors, RAR and RXR[34,35]. RA induced apoptosis is mediated by BMPs[36]. In contrast, inhibition of RA leads to the failure of interdigital cell death[36]. Thus, RA most likely activates the cascade that leads to interdigital cell death.

Final processing of programmed cell death involves the activation of executioner caspases through two main pathways, intrinsic and extrinsic (Figure 1.5)[10]. The extrinsic pathway is activated by death receptors such as FADD, Fas-associated via death domain[37,38]. With the activation of death receptors, CFLAR, an apoptotic antagonist, isreleased, causing an activation of caspase 8 (CASP8) and 10 (CASP10)[39]. CASP8 or 10 then initiates the release of cytochrome *c* from the mitochondria[39]. In the intrinsic pathway, the release of cytochrome *c* is initiated by pro-apoptotic proteins, BAX, BAK and BID[4]. Once cytochrome *c* is released from the mitochondria, the intrinsic and extrinsic pathways are congruent. Cytochrome *c* interacts with APAF1, activating CASP9 which, in turn, activates the executioner caspases such as CASP3 to carry out the final stages of cell death[40].

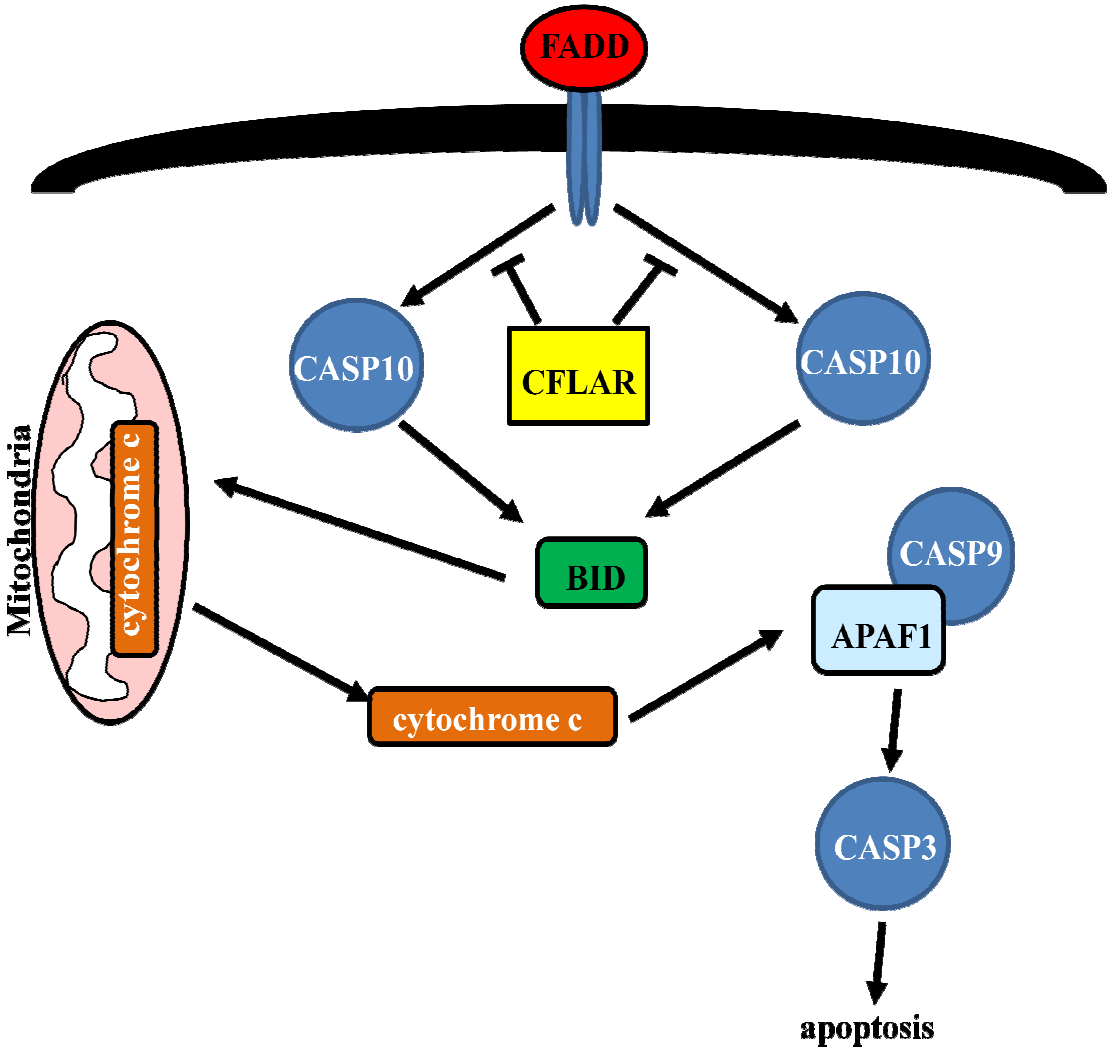

Figure 1.5. Intrinsic and extrinsic programmed cell death pathway
The extrinsic pathway is activated by death receptors causing a cascade of events beginning with the activation of *CASP8* and *CASP10* and in turn releasing cytochrome c from the mitochondria. The extrinsic and intrinsic pathway converge with the release of cytochrome c leading to the final activation of *CASP3* carrying out the final stages of cell death.

1.2 Hindlimb muscle development

All body muscle, except for head musculature, is derived from the condensation of paraxial mesoderm into epithelial structures called somites, epithelial balls of cells. The dorso-lateral region of the somite forms the dermomyotome which will give rise to dermis and muscle progenitors cells (MPC)[41]. The hypaxial (dorso-lateral lip) part of the

dermomytome generates the limb, tongue, diaphragm and ventral wall musculature[41]. Beginning at embryonic day 11.0 (E11), thousands of muscle progenitors cells originating from the hypaxial part of the dermomytome migrate to the hindlimb bud[41]. Migration of these progenitor cells are regulated by *PAX3* and *LBX1*, both homeobox transcription factors[42,43]. At this stage, the cells are undifferentiated muscle cells also known as myoblasts (mononucleated). Initially, as the myoblasts enter the limb bud, they begin to form dorsal and ventral muscle masses, largely dependent upon scatter factor/hepatocyte growth factor (*SF/HGF*) expressed in the limb mesenchyme[44,45]. Once in the actual limb bud, multiple factors such as BMPs, WNTs, FGFs, SHH, RA and apoptosis further refine and position these premuscle masses[46]. Myogenic differentiation, the formation of skeletal muscle cells also known as myotubes, begins with the expression of myogenic regulatory transcription factors (*MRFs*)[41,47,48]. *MYF5* (myogenic factor 5) and *MYOD1* (myogenic differentiation 1) are the first factors to designate myogenic commitment, followed by *MRF4* and *MYOG* (myogenin)[41,47].

Terminal differentiation requires the formation of highly specific and well-organized slow and fast muscle fiber types. Primarily, the hindlimb is composed of slow- and fast-twitch muscle fibers characterized by the expression of specific myosin heavy chains (MYH) and distinct metabolic activities. Fast-twitch muscle fibers are characterized by quick, force generating contractions that fatigue easily[41]. Slow-twitch muscle fibers use oxidative metabolism to generate slower, longer contractions[41]. Specification of when and where myogenic cells form slow- or fast-twitch muscle fibers has been an ongoing debate. One model suggests that limb myogenic cells are predisposed to either a slow or fast fate[41,49-53]. Studies have shown that the first premyogenic cells to enter the limb bud form

the proximal slow-twitch muscles, while the second migratory wave of cells form the fast-twitch muscles[54]. Other studies, however, propose a model that premyogenic cells are unspecified and have the ability to generate slow- or fast-twitch muscle fibers and formation of either fiber is influenced by limb bud factors[55-58]. A combination of these two models has also been suggested, where the myogenic cells are initially biased towards a slow or fast fate, however once in the limb bud, factors within the limb bud overrule this bias towards a fiber fate specified by the local limb bud factors[49,59].

1.3 Clubfoot

Clubfoot (congenital talipes equinovarus, CTEV) is one of the most common birth defects involving the musculoskeletal system. Positional deformation of the foot responds to minimum manipulation and therefore is not considered to be clubfoot[60]. A true clubfoot is a rigid malformation. While there is no universally accepted clubfoot classification scheme, two classification systems, one developed by Dimeglio *et* al. and the other by Pirani *et* al. are most commonly used[61,62]. Both systems apply a point score (0-20) based on a number of different physical findings, which when totaled is used to predict clubfoot severity with the higher score equaling greater severity. As shown in Figure 1.6, clubfoot consists of four components: equinus (downward), hindfoot varus (inward), forefoot adductus (inward) and cavus (high

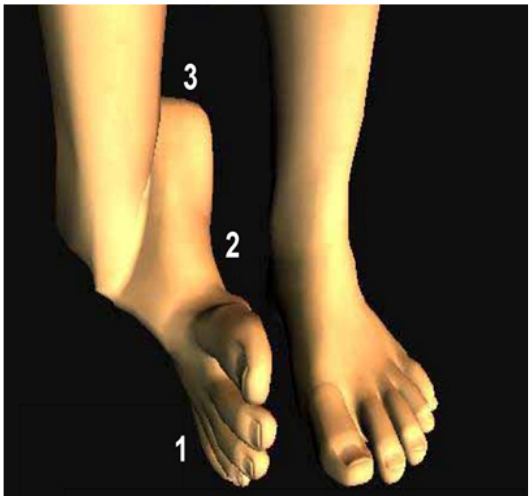

Figure 1.6. Clubfoot
Clubfoot is characterized by forefoot adduction (1), midfoot cavus (2) and hindfoot varus and equinus (3).

Adapted from www.fpnotebook.com with permission

arch)[1]. Clubfoot mimics the embryonic foot position during the second month of embryonic development. In addition, to structural malformation of the foot, individuals with clubfoot have persistent calf muscle hypoplasia.

1.3.1 Isolated (nonsyndromic) clubfoot

Twenty to twenty-five percent of clubfoot cases are associated with a syndrome such as distal arthrogryposis (DA), congenital myotonic dystrophy, myelomeningocele, amniotic band sequence, Trisomy 18 and Chromosome 22q11 deletion[63-65]. The remaining 75-80% of cases are isolated (idiopathic, no other malformations). Isolated clubfoot is a common birth defect, although the birth prevalence varies across ethnicities with a high of 1/150 in Polynesians, a low of 1/2500 in African Americans and a worldwide average rate of approximately 1/1000 live births[66-71]. Half of all cases have both feet affected (bilateral), while those unilateral cases involve the right foot being affected more often than the left[69,72]. Males are affected twice as often as females[72,73].

1.4 Treatment

Untreated clubfoot results in a rigid deformity causing gait disturbance, callus formation and skin/bone infections. This leads to substantial limitations in mobility and employment opportunities, particularly in third world countries. Fortunately, multiple treatment options are available. Early treatment methods relied heavily on surgery as the first choice in corrective treatment for clubfoot[74,75]. One of two surgical methods were typically used, the 'a la carte' or 'one-size fits all'. In 1979, Turco described the 'one-size fits all' approach as each foot undergoing the same surgical procedure[74]. In 1987, the 'a la carte' approach described by Bensahel *et* al. involved the release of all structures until full correction of the foot was obtained[75]. Because of often observed side effects after surgical

treatment such as persistent painful feet, residual deformities and relapses, more conservative approaches were developed. The two most common nonsurgical methods now used are Ponseti and Montpellier (French or functional), both promoting the progressive stretching of the muscles and tendons to avoid the use of a surgical soft tissue release procedure[76,77]. For both methods, treatment beginning within the first few weeks of life is ideal. The Ponseti method involves serial manipulation, casting and bracing, along with cutting of the Achilles tendon (when necessary) to obtain a corrective foot[75,77-79]. In contrast the Montpellier method involves daily manipulations of the clubfoot and uses adhesive taping to maintain the correction achieved with stretching[75,79]. The Montpellier method requires more parental participation but allows more flexibility in the foot, whereas the Ponseti method is less time consuming but more restrictive with foot movement because of the casting and bracing. However, while conservative treatment methods have improved outcomes, they do not always work and surgery is needed to correct the foot to normal plantar position. Even after correction, the calf muscle of the affected leg(s) remains underdeveloped.

1.5 Clubfoot etiology

Clubfoot is a complex birth defect that does not follow a classic Mendelian inheritance pattern. Studies suggest a multifactorial model with both genetic and environmental factors, separately and in combination, contributing to the etiology of clubfoot. In addition, multiple theories have been suggested as causes for clubfoot and will be discussed in the following sections.

1.5.1 Etiologic theories for clubfoot

As early as 400 BC, Hippocrates described the first clubfoot and he believed that the pressure from uterine molding on the developing foot caused deformity[80]. The restriction of fetal foot movement by the uterus as a cause for clubfoot became a widely held hypothesis[81-83]. Multiple births, reduced amniotic fluid volume, prolonged gestation, younger maternal age and increased birth weight have also been suggested to cause decreased fetal movement[82,84,85]. Amniotic leakage due to early amniocentesis has been associated with clubfoot, which would further suggest decreased fetal movement as a cause for clubfoot[84]. However, multiple studies have shown no association with lack of fetal movement as a cause for clubfoot[69,86-89] In addition, clubfoot is detected by ultrasound early in the second trimester, which is well before uterine constraint would affect the developing fetus. Furthermore, Wynne-Davies found concordance for clubfoot was the same between dizygotic twins and single births providing evidence against decreased limb movement as a cause for clubfoot[89]. Other mechanisms including bone, connective tissue, vascular and neurological abnormalities have been suggested; none have been proven[82,84,89-94].

The developmental arrest theory suggested by Hueter and von Volkmann has become more widely accepted[1]. Beginning at the second month of development the hind-foot resembles a clubfoot that normally rotates out to the normal plantar grade position[95]. However, developmental arrest at this critical time point (weeks 9-11) could result in a clubfoot[1,96].

1.5.2 Environmental causes of clubfoot

Environmental factors have been suggested to contribute to clubfoot. Seasonal variation has been observed in some studies with higher birth prevalence incidence in winter

(December-March), while other studies have found no correlation[85,97-100]. Teratogen exposure such as sulfonamides and abortifacient agents has been suggested but no solid epidemiological evidence supports these assertions[37,38]. While maternal folic acid supplementation has shown a decrease in birth prevalence of neural tube defects, only a small reduction in isolated clubfoot has been found on a population basis[40,71].

Only maternal smoking is consistently associated with clubfoot. Mothers who smoke during pregnancy have an increased risk of having a child with clubfoot and this risk increases in a dose-dependent pattern (Relative Risk: 1.3 to 2.2)[93,101,102]. Maternal smoking during pregnancy and a positive family history of clubfoot increases the risk of clubfoot 20-fold[101].

1.5.3 Genetic causes of clubfoot

Evidence for a genetic etiology for clubfoot comes from studies showing (1) aggregation of clubfoot in families, (2) increased risk with number of affected siblings, (3) heritability for clubfoot of 72% and (4) higher concordance in monozygotic twins than dizygotic twins (32.5% vs. 2.9%)[88,103-105]. Segregation analyses of various populations support a multifactorial/oligogenic model[67,71,81,89,103,106,107]. This model also accounts for gender differences observed in multiple population studies[108]. Males are affected more often than females, thus females require more susceptibility loci than males[108]. This is known as the Carter effect and is part of the multifactorial model[108]. It is now well accepted that genetic variation contributes to clubfoot, however the mechanism(s) are not well understood.

1.6 Approaches to candidate gene identification

Given that there is a genetic component to clubfoot, studies have focused on identifying the genes[64,69,71,73,104,109,110]. Currently, there are many approaches to identify

candidate genes for clubfoot such as genome-wide scans, animal models, syndromes, developmental pathways and chromosomal abnormalities. A gene is identified as a potential candidate for clubfoot because of its biological relevance; typically the gene plays a role in hindlimb development. Once a candidate gene is identified, polymorphisms such as single nucleotide polymorphisms (SNPs) and short tandem repeats (STRs) that span the gene of interest are identified and genotyped within an affected population. There are various analytic methodologies to test for association and/or linkage depending on study design. Linkage analysis utilizes the tendency for certain alleles to be inherited together even after recombination and are typically performed on families with multiple affected individuals[111]. Association studies test whether the sequence variant has a frequency in an affected population significantly different from the frequency in the general population[111]. Association studies work well with case-control studies, but can also be performed on datasets composed of families and trios[111]. The following sections will discuss the genetic variants identified through these approaches that contribute to the genetic etiology of clubfoot.

1.6.1 Genome scans

Genome-wide scans utilize polymorphic markers such as SNPs and copy number variants (CNVs) that span the whole genome to identify linked and/or associated regions with a phenotype of interest. Currently, only one genome-wide linkage scan has been reported on one five generation clubfoot family incorporating 13 family members[112]. Linkage to chromosome 5q31 with a LOD score of 3.3 was found. *PITX1*, a transcription factor that is required for hindlimb expression, is located within this region. A single missense mutation (E310K) in *PITX1* segregated with clubfoot and was not present in 500

controls[112]. Many of the affected individuals with the mutation had other skeletal malformations in addition to clubfoot indicating that the affected individuals did not have isolated clubfoot[112]. Interestingly, a genome-wide CNV array on forty clubfoot probands identified one individual with a 241 kb chromosome 5q31 microdeletion involving *PITX1*[113]. The deletion segregated in an autosomal dominant pattern over three generations[113]. In addition, a recent study reported haploinsufficiency of *Pitx1* in mice having a clubfoot-like phenotype with 8.9% penetrance (20 out of 225 mice)[113]. The haploinsufficient *Pitx1* mouse slightly differs from human clubfoot in some phenotypic characteristics such as females being affected more often than males, peroneal artery hypoplasia and reduced tibial and fibular bone volumes. These findings suggest that the E310K *PITX1* mutation plays a role in syndromic clubfoot, while *PITX1* haploinsufficiency contributes to isolated clubfoot.

A 17q23.1q23.2 microduplication in three families with clubfoot segregating in an autosomal dominant pattern with reduced penetrance was identified in the genome-wide CNV array[114]. The 17q23.1q23.2 microduplication contains the T-box transcription factor 4, *TBX4*, which is specifically expressed in the hindlimb and is important for muscle patterning[115,116]. *TBX4* is a direct transcriptional target of *PITX1*[117]. The authors suggest that the 17q23.1q23.2 microduplication is a common cause for familial isolated clubfoot and that the *TBX4-PITX1* pathway could play a role in isolated clubfoot. However, the three individuals identified to have the microduplication also had other skeletal malformations such as short wide feet and toes suggesting variable expressivity. Studies performed in Dr. Jacqueline T. Hecht's laboratory (Hecht lab) identified only one multiplex family with this microduplication[118]. This microduplication segregated in a large multigeneration family

with clubfoot and short, wide feet, once again supporting variable expressivity with this microduplication[118]. Interestingly, no association was found for *TBX4* and clubfoot[118]. Thus, the role of *TBX4* in the etiology of isolated clubfoot is still being investigated.

1.6.2 Developmental genes

Hindlimb and muscle development involves the interaction of multiple pathways involving growth factors, transcription factors, receptors and many other factors (Section 1.1). Limb initiation, patterning, outgrowth and development are all required for correct limb formation, in which perturbation of any of these genes could contribute to a clubfoot. Thus, there are many potential candidate genes for clubfoot in these developmental processes.

1.6.3 *Chromosomal abnormalities*

Chromosomal deletions, duplications and translocations that cause syndromes with multiple malformations have been shown to harbor genes that contribute to isolated birth defects. One of the first studies to identify chromosomal abnormalities associated with syndromic clubfoot was reported by Brewer *et* al[65,119]. This study compiled individuals who had a variety of phenotypic characteristics caused by chromosomal deletions and duplications. Phenotypic abnormalities were listed by specific diagnosis and chromosomal detections. For clubfoot, this included six large chromosomal deletion regions on 2q31-33, 3q23-24, 4p16-14, 7p22, 13q33-34 and 18q22-23 and two duplication regions on 6q21-27 and 10p15-11[65,119]. Our lab (the Hecht lab) systematically interrogated each region for genes associated with clubfoot[120,121]. Beginning with the 2q31-33 deletion region, two short tandem repeats (STRs), GATA149B10 and D2S1371, were found to be associated with clubfoot[120]. GATA149B10 was located near three apoptotic genes, *CASP8, CASP10* and

CFLAR. These genes are involved in the mitochondrial-mediated apoptotic pathway which is consistent with the key role that apoptosis plays in limb and muscle development (see Sections 1.1.5 and 1.2). Interrogation of the seven genes (*CASP9, CASP10, CASP8, CASP3, APAF1, BCL2* and *BID*) involved in this pathway identified suggestive associations with a SNP in each gene and clubfoot[121]. Further analysis of the 2q31-33 deletion identified another candidate gene, Homeobox gene cluster D (*HOXD)*. The *HOXD* gene cluster directs limb and muscle patterning during development and is functionally redundant with the Homeobox A gene cluster (*HOXA*) located on chromosome 7p15[122-124]. Mutations in both *HOXA* and *HOXD* have been associated with syndromes that involve limb abnormalities but not clubfoot[125,126]. Interrogation of these genes in our clubfoot nonHispanic white (NHW) and Hispanic multiplex and simplex families discovery group found associations with SNPs in both gene clusters and clubfoot[127]. One SNP, rs3801776, located in the basal promoter of *HOXA9*, gave the strongest association with clubfoot in both discovery and validation clubfoot groups[127]. These results suggest that perturbation in genes involved in limb and muscle development play a role in clubfoot.

1.6.4 Animal models of clubfoot

Animal models have proven to be useful tools in identifying candidate genes for many different diseases and birth defects[128,129]. Several mechanisms have been used to create animal models that harbor a phenotype of interest. Suggestive clubfoot-like animal models have been generated, but many show differences from human clubfoot. RA, a teratogen previously discussed in section 1.1.5 has been used in mice and rats to induce a clubfoot-like phenotype[130-132]. High levels of RA can lead to birth defects, while lack of RA can prevent normal embryonic development[28]. These studies utilizing RA administered

above normal levels, thus creating an unnatural environment for the pregnant mother, thus suggesting a nonviable clubfoot animal model. Some studies have suggested early amniocentesis as a cause for some sporadic cases[84,133]. A study performing amniotic sac punctures in mice observed a clubfoot phenotype[133]. However, the amniotic sac puncture animal model had multiple malformations thus causing a syndromic form of clubfoot and not isolated clubfoot. One other mouse model is the *pma* mutant mouse, which lacks the peroneal nerve[134-136]. This mouse model closely recapitulates the human form of clubfoot however the peroneal nerve is present and usually normal in most cases of human clubfoot[136-138]. Currently, the only plausible clubfoot model is the *PITX1* haploinsufficient mouse discussed in section 1.6.1. Though the haploinsufficient *Pitx1* mouse differs slightly in phenotypic characteristics compared to the human form, this genetic model of clubfoot may provide important clues about the genetic underpinnings of clubfoot.

1.6.5 Syndromes

Clubfoot can occur as a phenotypic manifestation of a syndrome that follows a Mendelian mode of inheritance and has a known genetic cause. It has been hypothesized that gene mutations causing syndromic clubfoot may harbor variants (common or rare) that contribute to isolated clubfoot. Distal Arthrogyrposis (DA) is a syndrome characterized by congenital joint contractures including clubfoot[139]. Currently, there are nine different types of DA and clubfoot occurs in four types. Mutations in muscle contraction genes, *MYH3, TNNT3, TNNI2, TPM2, MYBPC1* and *TPM8,* cause DA1, DA2A, DA2B and DA7[139-148]. Coding exons of *MYH3, TNNT3* and *TPM2* from 20 clubfoot probands were sequenced in a recent study[149]. Rare variants were identified but none segregated with the disease phenotype. The authors concluded that these genes do not contribute to the genetic etiology

of clubfoot[149]. Since the promoter regions of these genes were not assed, these genes remain candidates for clubfoot and studies focusing on these genes are presented in Chapter 5.

1.7 Significance

Clubfoot is a common birth defect affecting 135,000 newborns annually worldwide. Even with corrective treatment, foot and leg abnormalities and discomfort often persist. The complex genetic heterogeneity underlying clubfoot has only recently been appreciated and needs to be further investigated. Identification of these genetic factors will aide in identifying at-risk genotypes and this information may be translated into better clinical care and genetic counseling.

The goal of this project was to identify genetic variants that contribute to clubfoot in our well-characterized NHW and Hispanic multiplex and simplex clubfoot families. Interrogation of candidate genes from an animal model and a genome-wide linkage scan are discussed in Chapters 3 and 4. Muscle contraction genes implicated to cause syndromic clubfoot are assessed in Chapter 5. Functional implications of associated regulatory SNPs in *TPM1, TPM2, TNNC2* and *HOXA9* are described in Chapter 6. Overall, this project utilized multiple approaches and identified regulatory SNPs associated with clubfoot suggesting a common mechanism, gene regulation, as playing a key role in the genetic etiology of clubfoot.

CHAPTER 2: Materials and methods

2.1 Dataset

Institutional review board (IRB) approval for the use of these human samples was obtained from the Committee for the Protection of Human Subjects at the University of Texas Health Sciences Center at Houston (HSC-MS-04-239). Clubfoot families were recruited from clinics at Shriners Hospitals for Children in Houston, Los Angeles and Shreveport, and Texas Scottish Rite Hospital for Children of Dallas and the University of British Columbia. All probands and family members underwent clinical and radiographic examinations to exclude syndromic cases of clubfoot. Ethnicity was self-reported. Two generation pedigrees were obtained for all probands and expanded based on a positive family history. Pedigrees were subdivided into simplex trios (those without affected relatives = no family history) and multiplex families (those with multiple affected individuals = family history) (Figure 2.1). Our primary (discovery) dataset was comprised of 600 clubfoot families (1,923 individuals). Clinical information about the discovery dataset is listed in Table 2.1.

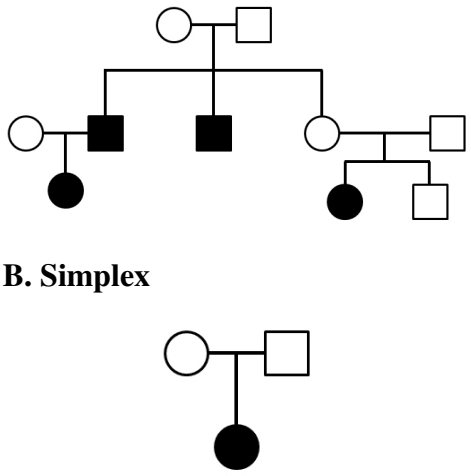

Figure 2.1. Multiplex and simplex family structures
Figure represents structure of multiplex (A) and simplex families (B). Circles represent females and squares represent males. Solid shapes indicate affected status.

After informed consent was obtained, blood and/or saliva samples were collected from patients and family members. In the presence of a family history, all family members' saliva were collected either in person or mail. DNA was extracted from blood using the

Roche DNA Isolation Kit for Mammalian Blood (Roche, Switzerland) and from saliva using the Oragene Purifier for Saliva (DNA Genotek, Inc., Ontario, Canada) following

Table 2.1. Clubfoot Discovery Dataset

	NHW			Hispanics			Total		
	+FH	**-FH**	**Total**	**+FH**	**-FH**	**Total**	**+FH**	**-FH**	**Total**
Families	149	144	293	91	216	307	240	360	600
Individuals	400	637	1037	326	560	886	726	1197	1923
Male	334	204	538	156	279	435	490	483	973
Affected	186	98	284	83	152	235	269	250	519
Sidedness									
Left side	28	18	46	17	38	55	45	56	101
Right side	29	23	52	15	35	50	44	58	102
Unilateral*	4	0	4	0	0	0	4	0	4
Bilateral	109	56	165	49	77	126	158	133	291
CF Unknown	16	1	17	2	2	4	18	3	21
Female	303	196	499	170	281	451	473	477	950
Affected	90	46	136	45	64	109	135	110	245
Sidedness									
Left side	13	12	25	7	18	25	20	30	50
Right side	13	11	24	8	18	26	21	29	50
Unilateral*	3	0	3	0	0	0	3	0	3
Bilateral	56	23	79	26	28	54	82	51	133
CF Unknown	5	0	5	4	0	4	9	0	9

*CF side not known
FH: Family History

manufacturer's protocol. DNA samples were quantified using the Nanodrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA) and checked for degradation on 2% agarose gels. DNA was stored at -20°C.

A secondary (validation) dataset was ascertained and characterized in the Orthopedic Clinic in the Department of Orthopedics at Washington University in St. Louis, MO. The validation dataset consisted of 142 NHW simplex trios. The validation dataset was used to confirm positive results in the discovery dataset.

A case-control validation dataset was composed of DNA from de-identified isolated clubfoot cases and matched control newborn blood spots from the Texas Birth Registry. Controls were matched to the cases by sex, maternal ethnicity, county of maternal residence, and birth ±8 weeks of the case's date of birth. This case-control dataset consisted of 616 NHW (308 cases and 308 controls) and 752 Hispanics (376 cases and 376 controls). DNA was extracted from the dried blood spots using the Qiagen DNeasy blood and tissue kit (Qiagen, Valencia, CA) and amplified using the Qiagen REPLI-g kit (Qiagen,Valencia, CA) following the manufacturer's protocol.

2.2 SNP selection and genotyping

Single nucleotide polymorphisms (SNPs) were identified using the National Center for Bioinformatics (NCBI, http://www.ncbi.nlm.nih.gov) and HapMap (www.hapmap.org) websites. SNPs were chosen based on (1) heterozygosity greater than or equal to 0.3 in NHW population (HapMap CEU dataset), (2) minor allele frequency greater than 0.2 in NHW population, (3) exonic and coding SNPs given higher priority over intronic SNPs, (4) "tagging" ability of SNPs, which provide genotyping information for multiple SNPs in linkage disequilibrium and (5) coverage of the gene 1 SNP every 10 kilobases on average including upstream and downstream of the gene. "Tagging" SNPs were identified using Haploview[150]. Genotyping was performed using either TaqMan® SNP Genotyping Assays or SNPlex™ (Applied Biosystems, Foster City, CA). TaqMan® results were analyzed on the 7900HT using SDS 2.1 (Applied Biosystems, Foster City, CA). SNPlex™ results were analyzed on a 3730 using GeneMapper® 4.0 (Applied Biosystems, Foster City, CA). Manufacturer's protocols were used to perform both genotyping assays. Genotype calls were imported into Progeny Lab database management software (Progeny Software, Delray

Beach, FL). PedCheck software was used to check for Mendelian inconsistencies and discrepancies[151].

2.3 Statistical Analyses

Analyses were performed on the dataset as a whole (all) and then stratified based on the presence (multiplex) or absence (simplex) of family history. Allele frequency and Hardy-Weinberg equilibrium (HWE) was calculated by SAS (Statistical Analysis Software) v9.1. For the majority of SNPs, the allele frequency significantly differed ($p<0.05$) between NHW and Hispanics, thus analyses were performed on each ethnicity separately for all tests. GOLD (Graphical Overview of Linkage Disequilibrium) was used to calculate pairwise linkage disequilibrium (LD) and displayed using Haploview[150,152].

Multiple methods to assess linkage and/or association were used to obtain the maximum amount of information from our datasets. Parametric and non-parametric linkage analyses were conducted using Merlin (Multipoint Engine for Rapid Likelihood Inferences)[153]. Two parametric models based on segregation analysis were tested on the multiplex families[107,109,154]. Model I used penetrances of 0.000, 0.020, 0.494 and 0.0, 0.008, 0.358 while model II used 0.0, 0.002, 0.067 and 0.0, 0.008, 0.358 for males and females, respectively.

Single SNP association with clubfoot was tested using three different analyses: PDT (pedigree disequilibrium test), geno-PDT (genotype PDT) and APL (association in the presence of linkage). PDT utilizes the traditional transmission disequilibrium test (TDT) while incorporating extended family members in addition to the nuclear family[155]. The geno-PDT is an expansion of PDT that focuses on the association of a genotype and clubfoot and is more accurate when looking at dominant and recessive models[156]. APL allows for the

incorporation of families in which a parental genotype is missing[157]. In addition to single SNP analysis, 2-SNP haplotype association was analyzed using APL[157]. Generalized estimating equations (GEE) was used to detect gene-gene interaction; assessing whether SNPs in two different genes is associated with clubfoot[158].

2.4 In silico analysis to assess transcription factor binding

In silico analyses were performed to assess potential function of associated regulatory SNPs ($p<0.05$). Using three online binding site prediction programs, Alibaba2, Patch and TESS, these programs predict whether the ancestral or alternate allele creates, eliminates and/or alters transcription factor binding[159-161]. The ancestral allele is defined by NCBI as the allele originating from the chimpanzee genome. Sequences for the ancestral and alternate allele were obtained from NCBI (www.ncbi.nlm.nih.gov) and inputted into the three online programs. Predicted outputs were compared across all three programs.

CHAPTER 3: Variation in IGFBP3 contributes to clubfoot

Note: The information presented in this chapter was published in 2009, in which I was co-first author: "Altered transmission of HOX and apoptotic SNPs identify a potential common pathway for clubfoot." *Am J Med Genet A* **149A(12)**, 2745-52 (2009). Permission for use of the article in this dissertation has been obtained from the American Journal of Medical Genetics Part A.

3.1 Introduction

The candidate gene approach has been successful in gene discovery in human genetics and has provided valuable information on the genetic etiology of clubfoot (see discussion in Section 1.6). Various approaches have been utilized to identify potential candidate genes for clubfoot. One approach utilizes chromosomal abnormalities such as duplications and deletions that cause syndromic clubfoot. Brewer *et* al. identified six chromosomal deletion regions and two duplication regions specific for syndromic clubfoot[65,119]. Evaluation of these chromosomal regions identified the *HOXD* gene cluster in the 2q31-q33 deletion region as biologically relevant transcription factor genes for isolated clubfoot because they direct limb and muscle patterning and differentiation [122,123]. In addition, the *HOXA* gene cluster located on 7p15-p14 has a redundant function with *HOXD*[122,123]. Mutations in the *HOXA* and *HOXD* gene clusters cause limb abnormalities including synpolydactyly and brachydactyly[21,125,126,162-165]. *HOXA* and *D* genes are important because they are spatially and temporally regulated during embryogenesis and specify limb patterning[122,123]. Twenty SNPs spanning the *HOXA* and *HOXD* genes were interrogated in our clubfoot discovery dataset. Strong association was found for one SNP, rs3801776, located in the regulatory region of *HOXA9,* in both our discovery and validation datasets[127]. In addition, multiple gene interactions between the *HOXA* and *D* clusters and the mitochondrial-mediated apoptotic genes were associated with clubfoot[127]. These results suggested that perturbation of the *HOXA* and *D* gene clusters and apoptotic genes may play a role in clubfoot.

Interestingly, conditional knockout of *Hoxa13* in mice causes loss of digits and decreased apoptosis[166]. These mice show decreased expression of insulin-like growth factor

binding protein (*IGFBP3*) in their hind limbs. *IGFBP3* belongs to a family of six genes that are involved in modulating insulin-like growth factors (*IGF*s), which regulate cell proliferation and apoptosis[167-169]. Previous studies performed in our lab identified associations with mitochondrial-mediated apoptotic genes and clubfoot (see Section 1.6.2)[121]. Furthermore, subcutaneous injections of the IGF-I/IGFBP3 protein complex slows the onset of muscle atrophy in immobilized hind limbs of rats, which is relevant because individuals with clubfoot show calf muscle hypoplasia[170]. The involvement of *IGFBP3* in multiple processes of limb development such as apoptosis and muscle formation suggest that *IGFBP3* is a biologically important candidate gene for clubfoot[121]. Thus, *IGFBP3* was interrogated in our clu

3.2 Materials and Methods

Information pertaining to our clubfoot discovery and validation datasets, sample collection, DNA extraction, SNP selection, genotyping and analysis (PDT, geno-PDT, APL, GEE and transcription factor binding) are described in Chapter 2.

For this analysis, 598 families including 179 multiplex (122 NHW and 57 Hispanic) and 331 simplex (130 NHW and 201 Hispanic) families from the discovery dataset and 142 NHW trios in the validation dataset were used. Twelve *IGFBP3* SNPs were genotyped in our discovery dataset (Table 3.1).

Table 3.1. *IGFBP3* SNPs

dbSNP	Chr:bp[a]	Location	Alleles[b]	MAF	HMAF[c]
rs2132571	7:45928199	U	A/G	0.306	0.218
rs2132572	7:45928070	U	A/G	0.227	0.478
rs2854744	7:45927600	U	A/C	0.472	0.316
rs2854746	7:45927170	E1 (missense)	C/G	0.415	0.264
rs2854747	7:45926442	I1	C/T	0.415	0.631
rs3793345	7:45924203	I1	C/T	0.199	0.15
rs2471551	7:45923580	I1	C/G	0.192	0.154
rs3110697	7:45921554	I3	A/G	0.422	0.629
rs10255707	7:45921217	I3	C/T	0.244	0.545
rs2453839	7:45920098	I4	C/T	0.197	0.13
rs6670	7:45918779	3' UTR	A/T	0.216	0.129
rs13223993	7:45917755	D	A/G	0.248	0.289

MAF: minor allele frequency; HMAF: Hispanic minor allele frequency;
U: upstream; I: intron; E: exon; UTR: untranslated region; D: downstream
[a]SNP data source; NCBI map – genome build 36.3
[b]Major allele listed first based on NCBI CEU listing
[c]Hispanic corresponding minor allele frequency
Cited with permission from *Am J Med Genet A*, Ester, Weymouth *et* al., 2009

3.3 Results

All SNPs were in HWE except for rs10255707 which was removed from further analysis. Parametric and nonparametric multipoint linkage analysis detected suggestive evidence (maximum LOD>1.5) in the Hispanic discovery dataset (Table 3.2).

The strongest association was with rs13223993, located downstream of *IGFBP3*. This SNP was significantly associated with all three statistical tests in the NHW simplex discovery dataset (Table 3.3). The remaining associations were marginal for both single SNP and 2-SNP haplotype associations in the discovery and the validation dataset. No single SNP association was found in the Hispanic discovery dataset (data not shown). However, it is of interest that many of these marginal associations involve SNPs located in

potential regulatory regions (Table 3.3 and 3.4). One of these SNPs, rs2132571, located in a potential regulatory region, was confirmed in our validation dataset (p=0.04).

Because of decreased *IGFBP3* expression in the conditional *HOXA13* knockout mouse and its role in apoptosis, potential gene interactions between *IGFBP3* and the *HOXA* and *D* genes and the mitochondrial-mediated pathway genes were evaluated. Interactions were found ($p<0.05$) in both the NHW and Hispanic discovery datasets (Tables 3.5 and 3.6, $p<0.01$ only shown). The most significant interactions in the NHW dataset were between rs3801776/*HOXA9*, a promoter SNP, and rs13223993/*IGFBP3,* located downstream

Table 3.2. *IGFBP3* multipoint linkage results for Hispanics

	Nonparametric		Parametric
dbSNP	**LOD**	**p-value**	**HLOD[a]**
rs2132571	2.06	0.001	1.93
rs2854746	2.13	0.0009	2.01
rs2453839	2.22	0.0007	2.12
rs6670	2.29	0.0006	2.19
rs13223993	2.29	0.0006	2.19

[a]All families linked under Modell II
(See Section 2.3 for model design)
Cited with permission from *Am J Med Genet A*, Ester, Weymouth *et* al., 2009

(p=0.0001), and rs2132572/*IGFBP3*, a potential regulatory SNP and rs1049210/*CASP3*, a missense mutation. The most significant interactions in the Hispanics were rs13223993/*IGFBP3* and rs741610/*HOXD* (p=0.002) and rs2453839/*IGFBP3* and rs2278361/*APAF1* (p=0.002).

Table 3.3. Single SNP Associations in NHW*

	ALL			Multiplex			Simplex		
dbSNP	**PDT**	**geno-PDT**	**APL**	**PDT**	**geno-PDT**	**APL**	**PDT**	**geno-PDT**	**APL**
rs2132571	0.206	0.378	0.100	0.365	0.691	0.086	0.362	0.360	0.605
rs2132572	0.105	0.217	0.154	0.571	0.570	0.860	**0.027**	**0.026**	0.053
rs2854744	0.673	0.317	0.660	0.408	0.161	0.229	0.593	0.878	0.503
rs2854746	0.319	0.453	0.595	0.260	0.154	0.352	1.000	0.523	0.827
rs2854747	0.404	0.650	0.671	0.941	0.504	0.238	0.128	0.209	0.547
rs3793345	0.210	0.326	**0.015**	0.302	0.427	0.066	0.475	0.721	0.101
rs2471551	0.307	0.509	**0.032**	0.254	0.461	**0.041**	0.882	0.979	0.311
rs3110697	0.847	0.952	0.289	0.657	0.592	0.095	0.241	0.224	0.739
rs2453839	0.553	0.763	0.799	0.556	0.778	0.356	0.866	0.976	0.548
rs6670	0.050	0.111	0.871	0.051	0.100	0.852	0.537	0.766	0.943
rs13223993	0.549	0.734	**0.047**	0.108	0.240	0.852	**0.003**	**0.004**	**0.003**

*p<0.05 in bold

Cited with permission from *Am J Med Genet A*, Ester, Weymouth *et* al., 2009

Table 3.4. 2-SNP Haplotypes*

		NHW	Hisp
SNP 1	**SNP 2**	**p-value**	**p-value**
rs2854747	rs2132571	0.243	**0.046**
rs2854747	rs2854744	0.440	**0.046**
rs3793345	rs2132572	**0.041**	0.396
rs3793345	rs2854744	**0.036**	0.247
rs3793345	rs2471551	**0.041**	0.177
rs3793345	rs13223993	**0.012**	0.580
rs2471551	rs3110697	**0.036**	0.826
rs2471551	rs13223993	**0.021**	0.695

*Haplotypes with p-value<0.05 in bold.

Cited with permission from *Am J Med Genet A*, Ester, Weymouth *et* al., 2009

Table 3.5. Gene interactions for *HOX* and *IGFBP3* SNPs*

				NHW	Hisp
Gene 1	**SNP 1**	**Gene 2**	**SNP 2**	**p-value**	**p-value**
IGFBP3	rs2854747	*HOXA*	rs6968828	**0.003**	0.978
IGFBP3	rs3110697	*HOXA*	rs6968828	**0.001**	0.943
IGFBP3	rs13223993	*HOXA*	rs3801776	**0.0001**	0.890
IGFBP3	rs3793345	*HOXD*	rs2113563	**0.008**	0.832
IGFBP3	rs3793345	*HOXD*	rs2592394	**0.003**	0.921
IGFBP3	rs2471551	*HOXD*	rs2113563	**0.007**	0.687
IGFBP3	rs2471551	*HOXD*	rs2592394	**0.001**	0.627
IGFBP3	rs13223993	*HOXD*	rs741610	0.415	**0.002**
IGFBP3	rs13223993	*HOXD*	rs711812	0.272	**0.003**

*p-value<0.01 only shown and in bold
Cited with permission from *Am J Med Genet A*, Ester, Weymouth *et* al., 2009

Table 3.6. Gene interactions for *IGFBP3* and mitochondrial-mediated apoptotic SNPs*

				NHW	Hisp
Gene 1	**SNP 1**	**Gene 2**	**SNP 2**	**p-value**	**p-value**
APAF1	rs2278361	*IGFBP3*	rs2132572	**0.008**	0.888
APAF1	rs2278361	*IGFBP3*	rs2453839	0.507	**0.002**
APAF1	rs7968661	*IGFBP3*	rs2132571	0.891	**0.007**
APAF1	rs7968661	*IGFBP3*	rs2453839	0.769	**0.006**
APAF1	rs3900115	*IGFBP3*	rs2854747	**0.008**	0.750
CASP3	rs1049210	*IGFBP3*	rs2854746	**0.005**	□?
CASP3	rs1049210	*IGFBP3*	rs2854744	**0.005**	□?
CASP3	rs1049210	*IGFBP3*	rs2132572	**<.0001**	□?

*p-value<0.01 only shown and in bold
– value not available
Cited with permission from *Am J Med Genet A*, Ester, Weymouth *et* al., 2009

3.4. Discussion

In this study, we assessed whether *IGFBP3* has an etiologic role in clubfoot. Twelve *IGFBP3* SNPs were genotyped and association was tested in our discovery and validation clubfoot datasets. The strongest association was found for rs13223993 located downstream of *IGFBP3* and for rs2132572, in a potential regulatory region. Other SNPs providing only suggestive evidence for single SNP and 2-SNP haplotype associations were also in potential regulatory regions (Table 3.3). These results suggest that variants in noncoding regions, such as the promoter or enhancer regions have a role in the etiology of clubfoot by perturbing gene expression.

Clubfoot is a complex disorder caused by multiple genes and environmental factors[104,108,110]. This assumes that multiple variants, not just one single variant in a gene can alter gene function and/or expression. To evaluate this possibility, we tested for potential interactions with variants in other genes previously shown to be associated with clubfoot[120,121,127]. We found evidence for significant gene interactions with *HOXA, HOXD* and three mitochondrial-mediated apoptotic genes, for example *CASP3* (Table 3.5 and 3.6). Apoptosis plays a key role in limb and muscle development as one study has shown that *CASP3* plays a role in muscle and tendon shaping later in limb development[46]. Even though mitochondrial-mediated apoptotic genes provided only suggestive evidence for a role in the etiology of clubfoot, we show significant gene interactions with *CASP3* and *IGFBP3* (Table 3.6)[121]. These interactions involved rs1049210/*CASP3* and three different *IGFBP3* regulatory SNPs (Table 3.6). Alteration in CASP3 protein function in combination with *IGFBP3* misexpression could contribute to clubfoot. Many of the associated gene interactions incorporate regulatory SNPs suggesting alteration of expression of multiple limb and/or muscle genes could be a key mechanism for clubfoot.

The association results (Table 3.3) in most cases did not attain significance after Bonferroni correction ($p<0.004$) thus only providing suggestive evidence for association. However, it is important to note that the majority of the suggestive associations incorporated regulatory SNPs, in particular three SNPs, rs2132572, rs3793345 and rs2471551. In addition, the regulatory SNP, rs2132572, was associated in our validation dataset. Regulatory SNPs can alter transcription factor binding and affect gene expression. *In silico* analysis of the three associated regulatory SNPs predict DNA binding to be altered depending on the presence of the ancestral or alternate allele (Table 3.7). These three SNPs will need to be further evaluated through functional assays. Even though these results do not meet Bonferroni correction significance, multiple associations incorporating regulatory SNPs and *in silco* analysis predictions further support the importance of regulation of gene expression as a key mechanism for clubfoot.

Table 3.7. Predicted transcription factor binding sites for associated *IGFBP3* SNPs

	Alibaba		**Patch**		**TESS**	
	Alleles		**Alleles**		**Alleles**	
SNP	**Ancestral**	**Alternate**	**Ancestral**	**Alternate**	**Ancestral**	**Alternate**
rs2132572	NFKB, Sp1	None	None	None	None	None
rs3793345	Oct-1	C/EBPalp	None	None	None	None
rs2471551	None	None	None	GATA-1	None	None

Abbreviations:
NFKB: nuclear factor of kappa light polypeptide gene enhancer in B-cells, Sp1: simian virus 40 protein 1, Oct-1: octomer-binding transcription factor 1, C/EBPalp: CCAAT Enhancer Binding Protein alpha, GATA-1: GATA binding protein 1

The results of this study provide a new way of looking at association data from which we have developed a model to explain the underlying etiology of clubfoot. We found evidence in our single SNP association and interaction analyses for the role of numerous regulatory variants in different genes in the development of clubfoot. All of the genes

included in this study are those known to play a role in limb and/or muscle development. This suggests a model wherein perturbation of gene expression in one gene is not sufficient to cause clubfoot. Rather, it is variation/perturbation of a number of genes that are both necessary and sufficient to disturb gene expression thereby affecting limb development. As we show in Figure 3.1, this clubfoot model incorporates genes involved in key aspects of limb patterning, apoptosis and muscle development as well as undefined environmental influences. The model suggests that phenotypic expression depends on having a specific genetic liability signature; the level of phenotypic severity is likely dependent upon the total number of variants across multiple genes.

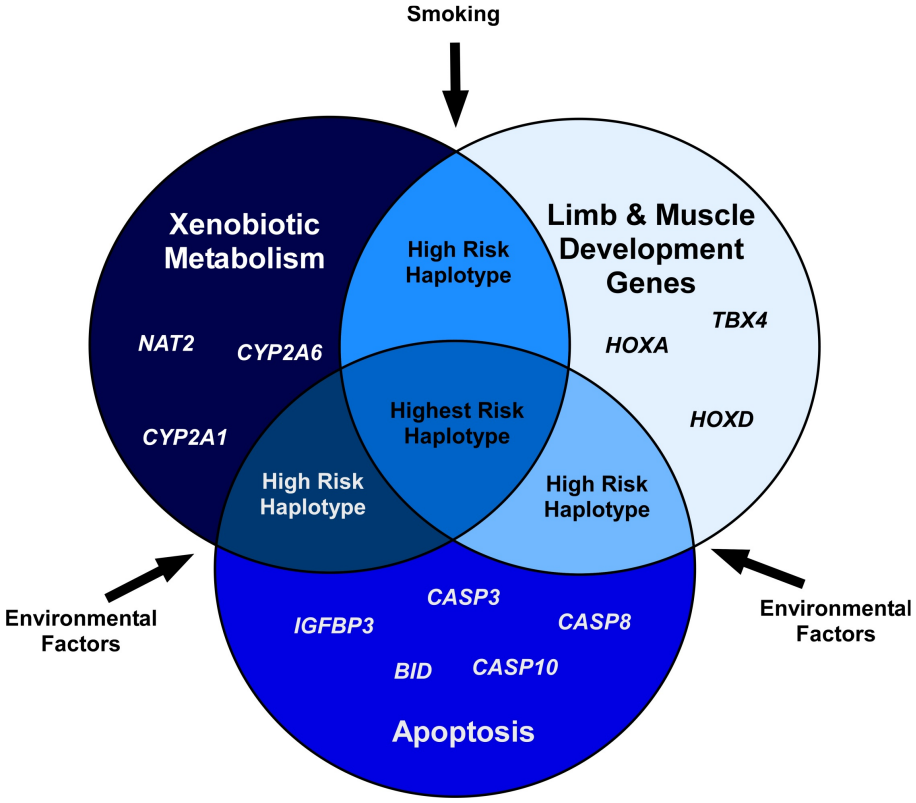

Figure 3.1. Clubfoot working model. This model incorporates associated genes involved in key aspects of limb patterning, apoptosis and muscle development along with potential environmental influences that could contribute to clubfoot.
Modified with permission from *Am J Med Genet A*, Ester, Weymouth *et* al., 2009

CHAPTER 4:

Genome-wide scan identifies candidate regions for clubfoot

4.1 Introduction

Nonsyndromic clubfoot, a common, isolated orthopedic birth defect, is characterized by equinus, hindfoot varus, forefoot adductus and cavus abnormalities[1]. In addition to these structural malformations of the foot, calf muscle hypoplasia is present at birth and persists even after corrective treatment. Clubfoot occurs sporadically in the majority of families; however, 20-25% of families have at least one other affected individual, but the pattern of familial aggregation does not follow a Mendelian mode of inheritance[69,71,107,109,110,154,171]. Concordance in monozygotic twins and segregation analyses provide evidence for a genetic etiology[69,88,103,104,171,172]. The multifactorial threshold model was developed to explain the non-Mendelian inheritance pattern for common birth defects, including clubfoot[173-176]. Recurrence risk estimates for clubfoot families are empiric, derived from family studies and incorporate laterality and gender[177]. Thus, there is an important need to identify the genetic variation underlying clubfoot in order to provide personalized genetic risk assessment.

Towards this goal, multiple approaches for gene identification have been applied to clubfoot families, including interrogation of candidate genes based on biological function or location in chromosomal duplication/deletion regions of syndromic clubfoot, genome scans, and, more recently, chromosomal microarray analysis (CMA) to identify chromosomal copy number variation[112-114,120,127,149,178-180]. Thus far, these studies have uncovered only a small percentage of the genetic variation underlying clubfoot [112-114,120,121,127,179,181]. Variation in two genes, *TBX4* and *PITX1*, has been reported in different families with clubfoot; interestingly, *TBX4* is a direct transcriptional target of *PITX1*[112-114]. A chromosome 17q23.1q23.2 microduplication involving *TBX4* was identified in four clubfoot families that had other skeletal anomalies, suggesting that this is a cause of syndromic clubfoot[114]. This

represents a rare cause of clubfoot, as more than 600 families were screened and the microduplication was identified in only four families. *PITX1* was initially implicated in a multigenerational clubfoot family, with the identification of a missense mutation (E130K)[112]. This family had additional skeletal anomalies suggesting this too is a cause of syndromic clubfoot. However, a 241 kb microdeletion involving *PITX1* was identified in one multiplex clubfoot family with no other anomalies. Taken together, these data suggest that variation in *PITX1* can cause both syndromic and non-syndromic forms of clubfoot[113]. *PITX1* is a particularly interesting candidate since approximately 9% of *Pitx1* haploinsufficient mice have a clubfoot-like phenotype[113]. Altogether, the data suggest that variation in *PITX1* and *TBX4* account for less than 1% of familial cases[112-114]. Lastly, family studies have found suggestive evidence for associations between smoking and apoptotic and muscle contracture genes[120,121,127,179,182]. Thus, while we have some insight into the causes of clubfoot, the vast majority of the underlying genetic liability for nonsyndromic clubfoot remains to be identified. Towards this goal, we submitted ten of our largest nonHispanic white (NHW) multiplex clubfoot families for a 6K genome wide linkage screen to identify potential candidate regions for clubfoot.

4.2 Materials and methods

4.2.1 IRB approval

This study was approved by the Committee for the Protection of Human Subjects at the University of Texas Health Sciences Center at Houston.

4.2.2 Study population and sample preparation

Genotyping was performed on 35 affected and 57 unaffected individuals from ten NHW multiplex families (Figure. 4.1). These families were selected based on having 2 or more

affected individuals and a sufficient quantity of DNA for the study. The two largest families, F1 and F2, have 11 and 5 affected individuals, respectively, available for genotyping, while F3-F10 each have 2-4 affected individuals available. Probands were recruited from outpatient clinics in Shriners Hospitals for Children in Houston, Los Angeles and Shreveport and Texas Scottish Rite Hospital for Children of Dallas and the University of British Columbia. A standard three-generation pedigree was obtained for all probands and extended to include all affected relatives. All probands and family members underwent clinical and radiographic examinations to exclude syndromic cases of clubfoot. Ethnicity was self-reported. Blood and/or saliva were collected from patients and relevant family members.

DNA was extracted from blood using the Roche DNA Isolation Kit for Mammalian Blood (Roche, Switzerland) and from saliva using the Oragene Purifier for Saliva (DNA Genotex, Inc., Ontario, Canada) following manufacturer's protocol. DNA samples were quantified using the Nanodrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA) and checked for degradation by running on a 2% agarose gel.

4.2.3 Genome scan

DNA samples from 92 individuals (Fig. 4.1) were subjected to the 6K Illumina Linkage IVb mapping panel consisting of 6,008 SNPs using a BeadStation system (Illumina Inc., San Diego, CA). Allele detection and genotype calling were performed using the BeadStudio software (Illumina Inc., San Diego, CA).

Genotyping results from the genome scan were subjected to nonparametric and parametric linkage analyses using MERLIN (Multipoint Engine for Rapid Likelihood Inferences)[153]. Two parametric models based on segregation analyses were tested (0.000,

0.020, 0.494 and 0.0, 0.008, 0.358 (Model I) and 0.0, 0.002, 0.067 and 0.0, 0.008, 0.033 (Model II)) for males and females respectively[107,109,154]. The two large families (F1 and F2) were analyzed individually and the remaining families (F3-F10) were analyzed in aggregate, allowing for heterogeneity.

4.2.4 TNNC2 sequencing

The troponin C2 (*TNNC2)* gene was sequenced in two clubfoot individuals and one unaffected family member from F2, as well as a laboratory control. Primer sets were designed to capture the six exons and the exon/intron junctions (Table 4.1). PCRs were performed in a thermal cycler with 30 cycles of 94ºC for 30 seconds, T_mºC (Table 4.1) for 1 minute and 72ºC for 30 seconds. PCR products were purified according to manufacturer's protocol (Qiagen, Valencia, CA). Sequencing results were analyzed using Sequencher v4.8 (Gene Codes, Ann Arbor, MI).

Table 4.1. ***TNNC2*** **primers**

Exon	Forward Primer	Reverse Primer	bp	T_m (°C)
1	GTAATGGGTTCAGACTGTGGG	GAAAAGTCAAAGGCCTTCCTCC	366	59
2	TGGGAAGAATGGCTTTGAGGTGA	CCACCCTGCCTAGAGGCCACC	201	58
3	GGTAGGTGTGAGGCTGACAGT	AGCAGGTGGCAGACTGAGCCTGA	384	62
4	GAGGAGGTGGATGAGGACGGTG	CACTCCCAACACGGGGAAGCTTC	308	61
5	GAGGGGCTTAGCAGTCAGAAC	GTCGTGGAGCGCTTCTATACC	441	60
6	GTCATCCCCTCTGTGTGGC	CCAGCTCCCTTCCACATCC	397	59

bp: base pair

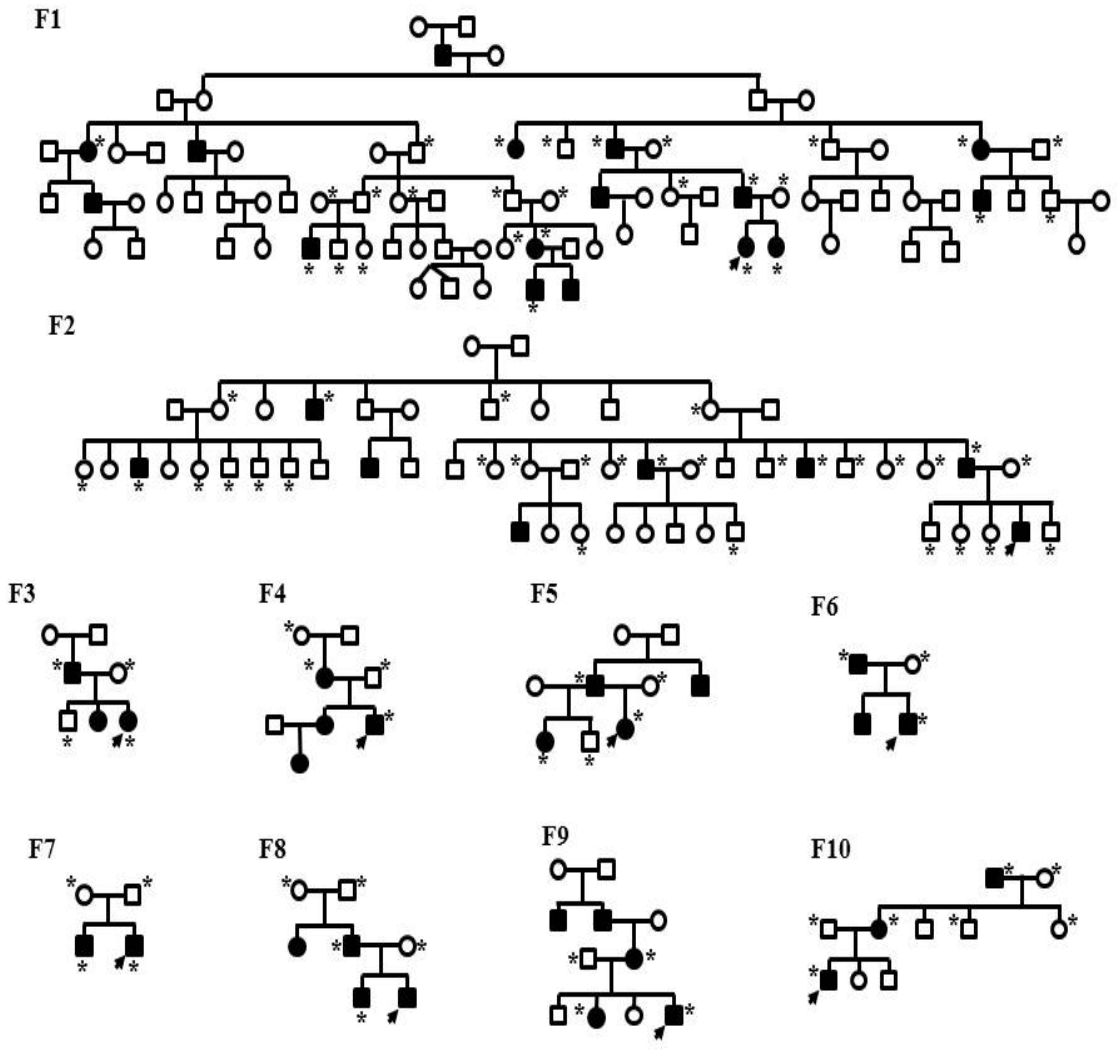

Figure 4.1. Pedigrees of ten NHW multiplex clubfoot families
Pedigrees of the ten NHW multiplex clubfoot families (F1-F10) are depicted with shaded shapes indicating an individual with clubfoot and an arrow indicating the proband. Individuals subjected to the 6 K genome scan are represented with an asterisk.

4.3 Results

A maximum multipoint LOD score of 2.54 was obtained for the 17q23.2 region for family F1, and a maximum multipoint LOD of 2.36 was obtained for the 3q22.1-3q24 for F2 (Tables 4.2 and 4.3). Additionally, there was suggestive evidence for linkage to 4p14, 4q32.1, 14q13.1 and 20q13.12-20q13.13 for family F2 (Table 4.3). *TNNC2*, a member of the muscle contractile apparatus, located in 20q13.12-q13.13, is a good candidate gene for clubfoot. However, no novel variants were found after sequencing the coding region of *TNNC2* in two affected individuals from F2 (data not shown).

Suggestive evidence for linkage was found for 14q32.12 (LOD=2.27) and 17q21.33-q22 (LOD=1.71) (Table 4.4) for families F3-F10.

Table 4.2 Linkage results for F1*

Chr	dbSNP	base pair[a]	Model I
17q23.2	rs725900	59674303	**2.54**
17q23.3	rs4353	61570422	**2.34**
17q23.3	rs1043127	61791210	**2.32**
17q24.1	rs7591	63525082	**2.08**
17q24.1	rs2107654	63633073	**2.07**
17q24.2	rs908150	64536433	**1.95**

*LOD>1.5 only shown, Chr: chromosomal region
[a]Base pair position based on NCBI genome build 37.3

Table 4.3. Linkage results for F2*

Chr	dbSNP	base pair[a]	Model I
3q22.1	rs1402455	132714125	1.57
3q22.1	rs1355776	133007112	1.84
3q22.1	rs12595	133496553	2.09
3q22.2	rs1984630	134414219	2.36
3q22.2	rs36178	134647169	2.36
3q22.2	rs750543	135026476	2.36
3q22.2	rs1502186	135244629	2.36
3q22.3	rs1052620	136574521	2.36
3q22.3	rs930984	136940786	2.36
3q22.3	rs768496	137237541	2.36
3q22.3	rs751357	138078155	2.36
3q22.3	rs531577	138401299	2.36
3q23	rs1426054	139618657	2.35
3q23	rs868534	140275652	2.35
3q23	rs1863868	141087623	2.35
3q23	rs1709	141331565	2.35
3q23	rs1479137	142623798	2.34
3q24	rs765695	143653877	2.32
3q24	rs1405597	144314276	2.30
3q24	rs1527732	145299384	2.23
3q24	rs1398775	145547356	2.18
3q24	rs1024080	145709954	2.13
3q24	rs1707465	145849864	2.02
4p14	rs278973	40396560	2.30
4p14	rs2035383	40688736	2.37
4p14	rs790142	40730503	2.37
4p14	rs951149	41024951	2.36
4p11	rs1350123	48999842	1.93
4q32.1	rs716428	156506869	1.59
14q13.1	rs10147920	34003009	2.01
14q13.1	rs2027338	34586663	2.30
20q13.12	rs244123	43189278	2.34
20q13.12	rs244099	43225466	2.34
20q13.12	rs1080026	43413198	2.35
20q13.12	rs1003855	43715188	2.35
20q13.12	rs1981431	43975451	2.35
20q13.12	rs411945	44418471	2.35
20q13.12	rs9074	44688665	2.35
20q13.12	rs1010310	44835044	2.35
20q13.12	rs460067	45125293	2.35
20q13.12	rs1046661	45817149	2.35
20q13.12	rs1537304	46093212	2.35
20q13.13	rs1547429	46549142	2.35
20q13.13	rs761272	47315581	2.34
20q13.13	rs911411	47718431	2.33
20q13.13	rs756529	48011008	2.30
20q13.13	rs119416	48035597	2.30
20q13.13	rs718630	49111256	2.05

*LOD>1.5 only shown, Chr: chromosomal region
[a]Base pair position based on NCBI genome build 37.3

Table 4.4. Linkage results for F3-10*

Chr	dbSNP	base pair[a]	Model I
14q32.12	rs1242119	93182466	**2.07**
14q32.12	rs1740696	93498149	**2.16**
14q32.12	rs882023	93532014	**2.27**
17q21.33	rs1063647	48187884	**1.55**
17q21.33	rs1124281	49849063	**1.71**
17q22	rs2033108	53823872	**1.57**

* LOD>1.5 only shown, Chr: chromosomal region
[a]Base pair position based on NCBI genome build 37.3

4.4 Discussion

In this study, we subjected 10 multiplex clubfoot families to an Illumina 6K genome-wide linkage analysis in order to identify chromosomal regions/genes contributing to clubfoot. Eight chromosomal regions with a LOD score ≥ 1.5 were identified: 3q22.1-q24, 4p14-p11, 4q32.1, 14q13.1, 14q32.12, 17q21.33-q22, 17q23.2-q24.2 and 20q13.12. Only one of these regions, 17q23.2-q24.2, has previously been implicated in clubfoot[114,118]. The remaining seven chromosomal regions are novel and need to be further evaluated for candidate clubfoot genes.

Linkage analysis in F1 identified the 17q23.2-q24.2 region as potentially harboring a clubfoot locus. Subsequently, using copy number and oligonucleotide array CGH testing modalities, we identified a 350 kb microduplication in this family[118]. As previously reported, the microduplication included the complete duplication of *TBX4* and *NACA2* and partial duplication of *BRIP1*. Notably, screening all 605 probands in our dataset did not identify any other families with a *TBX4* duplication[118].

Linkage analysis from F2 identified five novel regions: 3q22.1-q24, 4p14-p11, 4q32.1, 14q13.1 and 20q13.12 (Table 4.3). The 20q13.12 region is of interest because it contains *TNNC2*, a biologically interesting gene because it initiates muscle contraction after binding to Ca^{2+}[183]. Previously, we reported evidence for association between nonsyndromic clubfoot and variation in *TNNC2*[179]. While no novel variants were identified after sequencing the coding exons, promoter variants cannot be excluded and should be evaluated in future studies.

The regions identified on chromosomes 3 and 20 contain several biologically interesting genes, *SOX14* (3q22-q23), *PLOD2* (3q24) and *NEURL2* (20q13.12),. *NEURL2* (neutralized homolog 2 also known as Ozz-E3) is a muscle-specific ubiquitin ligase active during myogenesis[184]. *NEURL2* knockout mice develop maturation defects of the sarcomeric apparatus which is important for muscle contraction[184]. This is relevant as human clubfoot is sometimes characterized as a contracture abnormality[3,185,186]. Mutations in *PLOD2* (procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2) cause Bruck syndrome, a rare condition with clinical features of distal arthrogryposis 1 and osteogenesis imperfecta[187]. Clubfoot has been reported in some affected individuals thus suggesting that this gene could play a role in clubfoot[187,188]. Lastly, *SOX14* is a member of the *SOX* gene family, which is expressed during embryogenesis in a wide range of tissues. *SOX14* is expressed in the apical epidermal ridge of the hindlimb, which is important for limb outgrowth and patterning[189]. In addition, *SOX14* has been implicated as a candidate gene for limb defects in association with blepharophimosis, ptosis and epicanthus inversus and Mobius syndromes[189,190]. The limb defects in these conditions include brachydactyly,

clinodactyly, syndactyly, metacarpal abnormalities, hypoplasia of the lower legs, and clubfoot[189-191]. These regions and genes will be considered in future studies.

Finally, two additional novel regions, 14q32.12 and 17q21.33, were identified in the remaining eight families and need to be studied further (Table 4.4). The 17q21.33 region is upstream of the 17q23.1q23.2 microduplication and the linkage results clearly separate the regions[118]. The 14q32.12 and 17q21.33 regions are large and contain many genes and also need to be evaluated further.

While two genome-wide scans have identified 5q31 as potentially containing clubfoot genes, this region was not identified in the current study[112,113]. This may be due to a number of factors, including the small number of families tested and locus heterogeneity. However, the 5q31 region remains interesting because it contains *PITX1*, a hindlimb specific transcription factor, which has been implicated because of mutational events segregating with clubfoot in two families and increased frequency of clubfoot in *Pitx1* haploinsufficient mice[112,113]. However, one of these families also had additional skeletal anomalies, suggesting that this is a syndromic form of clubfoot and thus would not likely be detected in our nonsyndromic clubfoot families[112]. Interestingly, this data may also suggest that gene regulation may more directly affect foot development with regard to nonsyndromic clubfoot. Therefore, evaluation of the regulatory region of *PITX1* is warranted in future studies.

In this study, seven new potential clubfoot regions were identified and provide new candidate areas for future studies. Confirmation in additional familial clubfoot datasets is needed to validate the results and narrow these regions in order to identify the genetic variation contributing to nonsyndromic clubfoot. This will then allow for more personalized risk counseling for clubfoot families.

CHAPTER 5: Variation in muscle contraction genes contribute to clubfoot

Note: The information presented in this chapter was published in 2011, in which I was first author: "Variants in genes that encode muscle contractile proteins influence risk for isolated clubfoot." *Am J Med Genet A* **155A**, 2170-9 (2011). Permission for use of the article in this dissertation has been obtained from the American Journal of Medical Genetics Part A.

5.1 Introduction

As previously discussed in Section 1.6.4, a syndromic approach is a viable resource in identifying candidate genes. The hypothesis is that mutations in genes that cause rare multiple malformation syndromes that include the phenotype of interest may harbor common variants that contribute to the isolated condition. Distal Arthrogryposis (DA) syndromes are a group of rare autosomal dominant disorders characterized by multiple congenital joint contractures, including clubfoot and muscle hypoplasia[140-148]. The feet are generally more severely affected than the upper extremities. Nine different types of DA have been delineated and clubfoot is a common characteristic of several of these, including DA1, DA2A, and DA2B[139]. Mutations in muscle contraction genes, *MYH3, TNNT3, TNNI2, TPM2* and *MYBPC1,* have been identified as causes for these DAs with a clubfoot phenotype[140-148]. The calf muscles of individuals with clubfoot have consistently been reported to show a variety of abnormalities including disorganization of muscle fibers, increased number of Type I fibers (slow-twitch) and a decrease in Type II fibers (fast-twitch)[3,185,186]. With these observations, muscle contraction genes are plausible candidates for isolated clubfoot. Therefore, interrogation of fifteen muscle contraction genes was performed in our clubfoot dataset to assess whether variants in these genes influence the risk of clubfoot.

5.2 Materials and Methods

Information pertaining to our clubfoot discovery and validation datasets, sample collection, DNA extraction, SNP selection, genotyping and analysis (PDT, geno-PDT, APL, GEE and transcription factor binding) are described in Chapter 2.

For this analysis, multiple datasets were used. The discovery dataset consisted of 224 multiplex families (137 NHW and 87 Hispanic) and 357 simplex families (139 NHW and 218 Hispanic), a validation dataset of 142 NHW simplex trios and a case-control validation dataset consisting of 616 NHW (308 cases and 308 controls) and 752 Hispanics (376 cases and 376 controls). Seventy-four SNPs spanning fifteen muscle contraction genes were selected for evaluation based upon their expression and role in muscle contraction were genotyped (Figure 5.1 and Table 5.1).

> Log-linear regression models were used to evaluate the independent effects of maternal and inherited (child) genotypes for the *TNNC2* SNPs that were out of HWE in the NHW families[192-194]. Specifically, only one triad was selected per family consisting of the affected proband and their parents. For each SNP, the likelihood ratio test was used to compare the full model, which included parameters for both maternal and inherited genotypes, with reduced models, which included parameters for only the maternal or the inherited genotype. In addition, estimates of genotype relative risks and their associated 95% confidence intervals were estimated. All log-linear models assumed a log-additive model of inheritance. (cited from Weymouth, *et.* al, Am J Med Genet Part A, 2011)

Table 5.1. Muscle Contraction Gene SNPs

Gene[a]	SNP[b]	Position (bp)[c]	Alleles[d]	Location[e]
MYBPH 1q32.1	rs4950926	201403289	G/A	D
	rs2642531	201410348	C/G	E3 (missense)
	rs884209	201413912	A/G	U
ACTA1 1q42.13	rs728614	227630740	G/A	D
	rs506388	227637684	A/C	U
MYL1 2q33-q34	rs867342	210860950	T/C	D
	rs2136457	210865694	T/C	I5
	rs12469767	210876591	A/C	I1/U
	rs1074158	210883288	A/G	I1/U
	rs925274	210891742	C/T	U
TPM2 9p13.2-p13.1	rs3750431	35670337	C/G	D
	rs1998308	35673882	T/A	I8
	rs2145925	35679373	C/T	I1
	rs2025126	35686625	G/A	U
TNNI2 11p15.5	rs2292474	1815148	C/T	U
	rs1877444	1817801	C/A	I2
TNNT3 11p15.5	rs909116	1898522	T/C	I1/U
	rs2734510	1905537	T/C	I6/I5
	rs2734495	1915572	T/C	I14/I13
	rs7395920	1920888	C/T	D
TPM1 15q22.1	rs3809565	61120672	A/G	U
	rs4075583	61127280	A/G	U/I2
	rs4238371	61134456	C/G	I1/I2
	rs12148828	61142392	T/C	I8/I7
	rs1972041	61147900	G/A	I8/D/I7

Table 5.1. Muscle Contraction Gene SNPs (continued)

Gene[a]	SNP[b]	Position (bp)[c]	Alleles[d]	Location[e]
MYH13 17p13	rs1984620	10141073	C/T	D
	rs3744550	10147320	T/C	E3 (missense)
	rs11868948	10154767	A/G	I8
	rs17690195	10159838	C/T	E13 (missense)
	rs2074877	10164439	C/T	E17 (missense)
	rs1859999	10169540	G/A	I19
	rs2240579	10177190	A/G	E23 (syn.)
	rs11869897	10186410	C/T	I16
	rs11651414	10192536	A/G	I12
	rs4791980	10200165	C/T	I8
	rs12936065	10210239	C/T	I2
	rs7213488	10220668	G/T	U
	rs9906430	10228548	T/C	U
MYH8 17p13.1	rs9906430	10228548	T/C	D
	rs2270056	1023622	T/C	I50
	rs7211175	10237747	A/C	I47
	rs3744552	10244986	A/G	E32 (syn.)
	rs12601552	10255100	G/A	I14
	rs2277648	10265705	C/T	I2
	rs11078846	10269685	A/T	U
MYH4 17p13.1	rs11654423	10286056	C/T	D
	rs2058101	10295699	T/C	I27
	rs2058099	10303471	A/G	I14
	rs2011488	10311370	C/A	I2
MYH1 17p13.1	rs8077200	10331008	A/G	U
	rs3744563	10340622	A/G	I33
	rs2320950	10348281	A/G	I22
	rs8082669	10358592	G/A	I6
	rs9916035	10364565	T/C	U

Table 5.1. Muscle Contraction Gene SNPs (continued)

Gene[a]	SNP[b]	Position (bp)[c]	Alleles[d]	Location[e]
MYH1 17p13.1	rs8077200	10331008	A/G	U
	rs3744563	10340622	A/G	I33
	rs2320950	10348281	A/G	I22
	rs8082669	10358592	G/A	I6
	rs9916035	10364565	T/C	U
MYH2 17p13.1	rs9916035	10364565	T/C	D
	rs7223755	10367068	T/C	I39
	rs2277651	10373363	T/C	I25
	rs2277653	10383702	T/C	I12
	rs3760431	10393038	A/G	I2
	rs4239117	10396857	G/T	U
MYH3 17p13.1	rs2285475	10483196	C/A	E25 (syn.)
	rs876657	10485141	A/C	E19 (syn.)
	rs2239933	10489909	T/C	I11
	rs201622	10518759	C/G	U
MYBPC2 19q13.33	rs12462762	55633501	G/A	I7
	rs10405793	55640362	A/T	I11
	rs25665	55649209	G/A	E17 (missense)
	rs25667	55659452	G/A	E27 (missense)
	rs1274597	55665071	G/A	D
TNNC2 20q12-q13.11	rs3848711	43879507	T/C	D
	rs8860	43885308	G/A	3' UTR
	rs4629	43886104	G/T	E5 (syn.)
	rs437122	43888385	C/T	I1
	rs373018	43889466	C/T	U
	rs380397	43890062	T/G	U
	rs383112	43890756	C/T	U

U, upstream; D, downstream; I, intron; E, exon; S, synonymous; M, missense

[a]Gene name and chromosomal location

[b]SNP data source; NCBI map – genome build 36.3

[c]Base pair position

[d]Major allele listed first based upon NCBI listing

[e]SNP gene location

Cited with permission from *Am J Med Genet Part A*, Weymouth, *et.* al, 2011

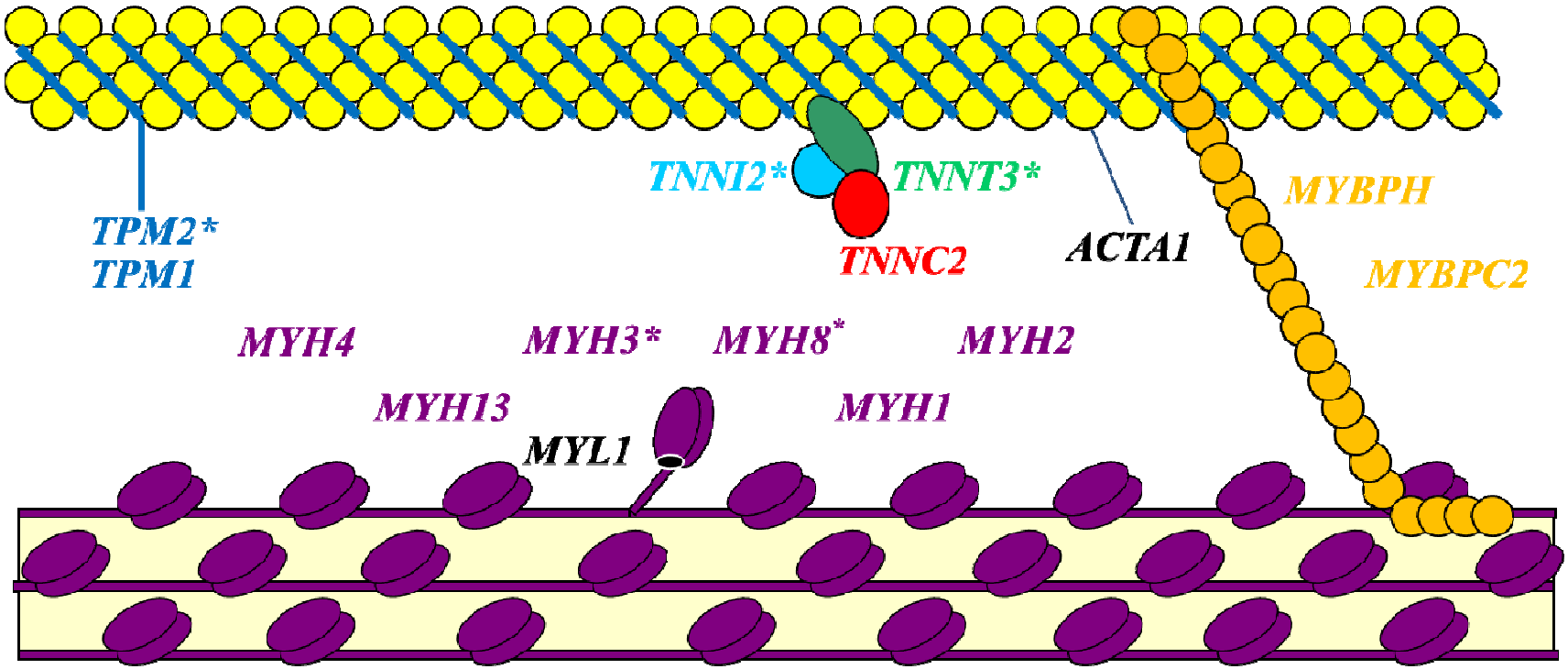

Figure 5.1. The muscle contractile apparatus

The fifteen muscle contraction genes interrogated in our clubfoot dataset are involved in various aspects of muscle contraction and are represented here. Gene name color correlates with location in the apparatus. Asterisks depict genes implicated in a type of Distal Arthrogryposis with a clubfoot phenotype.

5.3 Results

None of the SNPs in *TNNC2* were in HWE in the NHW discovery dataset and were removed from the association analyses; all remaining SNPs in the NHW were in HWE. All *TNNC2* SNPs were in HWE in the Hispanic dataset and were therefore included in the association analyses. Only rs2074877 in *MYH13* was out of HWE in the Hispanic discovery dataset and was removed from analyses. Allele frequencies differed significantly between the NHW and Hispanic groups for SNPs in 14 of the 15 examined genes. Therefore, the data were stratified by ethnicity. Parametric and nonparametric linkage analysis found no evidence for linkage (data not shown).

Overall, nominal evidence for association was found for SNPs in 12 of 15 genes in the discovery datasets ($p<0.05$; Table 5.2). For the NHW dataset, evidence for association was seen for SNPs in six genes: *MYBPH, TPM2, TNNT3, TPM1, MYH13,* and *MYH3* (Table 5.2A). Three SNPs in *MYH3* had altered transmission primarily in the NHW multiplex subset. All other associations involved a single SNP in each of the five other genes. In the Hispanic dataset, there was evidence for altered transmission in 11 genes (Table 5.2B). Five of these genes, *MYBPH*, *TPM2*, *TNNT3*, *TPM1*, and *MYH13*, also had SNPs with altered transmission in the NHW dataset; only one SNP was common to both datasets (*MYH13*/rs17690195). In addition, several genes had multiple SNPs with altered transmission [*MYL1*(3), *TNNT3*(3), *MYH8*(4), *MYH4*(3), *MYH1*(2), and *MYH2*(2)].

When 2-SNP haplotypes were considered, altered transmission was found for five genes in the NHW group ($p<0.01$; Table 5.3A). Two of these genes, *ACTA1* and *MYH8*, did not have individually altered transmitted SNPs. Three different *MYH13* haplotypes had altered transmission; none of the haplotypes included the individual SNPs with altered transmission (Table 5.3A). The two *TPM2* haplotypes both contained rs1998303, which had altered transmission in the single SNP analyses. In the Hispanic discovery dataset, three *MYH13* haplotypes had altered transmission (Table 5.3B); only one contained rs17690195, which had altered transmission in the single SNP analysis (Table 5.2B). There was no overlap between the NHW *MYH13* haplotypes and the Hispanic *MYH13* haplotypes, and only one SNP (*MYH13*/rs2240579) was common to both ethnicities.

Numerous potential gene interactions were identified in both the NHW and Hispanic discovery datasets ($p<0.01$; Table 5.4). The only gene interaction present in both datasets was *TPM1* and *MYH13*, although the same SNPs were not involved in the two datasets. SNPs in *ACTA1*, *MYH1*, *MYH13*, *MYH2*, *MYH4*, *MYH3*, *MYH8*, *MYL1*, *TNNT3*, *TPM1*, and *TPM2* were involved in interactions in both ethnic groups.

Three genes (*TNNI2*, *MYBPC2*, and *TNNC2*) did not have any SNPs meeting our criteria for follow-up in the validation datasets. In the family-based validation dataset, only two SNPs in the single SNP analyses demonstrated any evidence for altered transmission, *TNNT3*/rs2734495 ($p=0.04$) and *TPM1*/rs1972041 ($p=0.000074$; data not shown). The *TPM1* result is supported by the 2-SNP analyses in the validation dataset where only

TPM1 haplotypes had altered transmission (Table 5.5). All four of the significant haplotypes contained rs1972041. In the case-control dataset, only nominal evidence for association was seen with rs1248828 in *TPM1* (p=0.04) in the Hispanic subset; there were no associations in the NHW subset (data not shown).

Further examination of the NHW maternal, paternal, and proband *TNNC2* genotype frequencies revealed that only the maternal genotypes deviated from HWE, suggesting the presence of a maternal genetic effect. Table 5.6 summarizes the results of log-linear models assessing maternal and inherited genotypic effects. For rs383112, significant associations were observed with both the maternal and inherited genotypes (p=0.02 and 0.03, respectively). The maternal genotype for rs393112 was associated with 1.38-fold increased risk (CT vs. CC; 95% CI: 1.13-1.72) of clubfoot in offspring, while a protective inherited genotypic effect was conferred with a relative risk of 0.77 (CT vs. CC; 95% CI: 0.50-0.99). In addition, a significant protective inherited genotypic effect (p=0.02), with a relative risk of 0.74 (TG vs. TT; 95% CI: 0.48-0.97), was found for rs4629. (cited from Weymouth, *et.* al, Am J Med Genet Part A, 2011)

Table 5.2. Single SNP Association by Ethnicity[a,b]

A. NHW

		ALL			Multiplex			Simplex		
Gene	**SNP**	**APL**	**PDT**	**GENO-PDT**	**APL**	**PDT**	**GENO-PDT**	**APL**	**PDT**	**GENO-PDT**
MYBPH	rs4950926	0.149	0.477	0.733	**0.021**	0.128	0.179	0.812	0.413	0.447
TPM2	rs1998308	**0.003**	0.065	0.056	0.090	0.322	0.228	**0.009**	**0.027**	0.091
TNNT3	rs2734495	**0.019**	**0.043**	0.088	0.220	0.176	0.397	0.062	0.096	0.113
TPM1	rs4075583	**0.014**	0.519	0.700	0.221	0.694	0.723	**0.031**	**0.027**	**0.028**
MYH13	rs17690195	0.065	0.256	0.250	**0.039**	0.144	0.216	0.674	0.873	0.749
MYH3	rs2285475	0.442	0.091	0.242	**0.042**	**0.020**	0.081	0.364	0.696	0.861
MYH3	rs876657	0.399	**0.039**	0.109	**0.021**	**0.006**	**0.020**	0.345	0.696	0.926
MYH3	rs223993	0.320	0.104	0.211	**0.030**	0.058	0.161	0.705	0.884	0.848

B. Hispanic

		ALL			Multiplex			Simplex		
Gene	**SNP**	**APL**	**PDT**	**GENO-PDT**	**APL**	**PDT**	**GENO-PDT**	**APL**	**PDT**	**GENO-PDT**
MYBPH	rs884209	**0.045**	1.000	0.886	0.388	0.564	0.282	0.068	0.612	0.544
ACTA1	rs728614	0.299	0.053	0.227	0.095	0.398	0.737	0.812	**0.024**	0.050
MYL1	rs867342	0.059	**0.016**	0.069	0.637	0.196	0.413	0.062	**0.024**	0.122
MYL1	rs2136457	**0.034**	**0.021**	0.099	0.198	0.168	0.439	0.115	**0.047**	0.187
MYL1	rs12469767	0.108	**0.020**	0.083	—	0.206	0.454	0.113	**0.048**	0.156
TPM2	rs3750431	0.089	0.147	0.070	0.094	0.084	0.145	0.350	0.806	**0.008**
TNNT3	rs909116	0.628	1.000	0.138	**0.018**	0.527	0.062	0.284	0.292	0.440
TNNT3	rs2734510	0.143	0.697	0.361	**0.016**	0.607	0.208	0.746	0.912	0.876
TNNT3	rs7395920	**0.023**	**0.006**	**0.024**	0.110	**0.031**	0.069	0.095	0.085	0.184
TPM1	rs1972041	**0.017**	0.167	0.387	0.249	0.691	0.846	**0.039**	0.149	0.337

Table 5.2. Single SNP Association by Ethnicity (continued)[a,b]

B. Hispanic

		ALL			Multiplex			Simplex		
Gene	**SNP**	**APL**	**PDT**	**GENO-PDT**	**APL**	**PDT**	**GENO-PDT**	**APL**	**PDT**	**GENO-PDT**
MYH13	rs17690195	**0.038**	**0.003**	**0.010**	N/A	**0.043**	0.084	0.141	**0.029**	0.063
MYH13/ MYH8	rs9906430	0.300	0.116	0.352	0.804	0.814	0.664	0.151	**0.005**	**0.014**
MYH8	rs2270056	0.157	0.401	0.156	0.763	0.898	0.880	0.150	0.174	**0.042**
MYH8	rs12601552	0.229	**0.015**	0.103	—	0.353	0.395	0.261	**0.016**	**0.021**
MYH8	rs2277648	**0.012**	**0.028**	0.089	0.124	0.132	0.225	0.056	0.107	0.312
MYH8	rs11078846	0.174	0.052	0.076	0.445	0.385	0.734	0.301	0.059	**0.031**
MYH4	rs11654423	0.206	**0.016**	0.112	—	0.103	0.232	0.293	0.077	0.155
MYH4	rs2058099	0.070	**0.027**	0.121	0.376	0.431	0.613	0.144	**0.020**	**0.025**
MYH4	rs2011488	0.161	**0.018**	0.055	0.169	0.276	0.634	0.478	**0.024**	**0.030**
MYH1	rs8077200	0.872	0.677	0.732	0.106	0.194	0.152	0.079	**0.010**	**0.026**
MYH1	rs3744563	0.052	0.050	0.153	0.683	0.884	0.617	0.061	**0.015**	**0.018**
MYH2	rs2277651	0.092	**0.038**	0.121	0.891	0.362	0.386	0.064	0.050	0.193
MYH2	rs3760431	0.223	**0.037**	0.145	—	0.327	0.584	0.128	0.056	0.080

NHW, nonHispanic White; —, no value because of low APL variance

[a]SNPs with p<0.05 shown in bold

[b]p-values uncorrected for multiple testing.

Cited with permission from *Am J Med Genet Part A*, Weymouth, *et.* al, 2011

Table 5.3. 2-SNP Haplotype Transmission – Discovery Population[a,b]

A. NHW

Gene	SNP A	SNP B	p-value
ACTA1	rs728614	rs506388	0.008
MYH8	rs2270056	rs3744552	0.008
MYH13	rs11868948	rs1859999	0.007
MYH13	rs3744550	rs1859999	0.004
MYH13	rs3744550	rs2240579	0.00004
TPM2	rs1998308	rs2145925	0.006
TPM2	rs1998308	rs2025126	0.006
TNNT3	rs2734495	rs2734510	0.002

B. Hispanic

Gene	SNP A	SNP B	p-value
MYH13	rs1984620	rs4791980	0.0006
MYH13	rs2240579	rs7213488	0.008
MYH13	rs17690195	rs7213488	0.007

[a]p-values not corrected for multiple testing

[b]only p-values<0.01 shown

Cited with permission from *Am J Med Genet Part A*, Weymouth, *et.* al, 2011

Table 5.4. Gene Interactions Between SNPs in Different Muscle Contraction Genes[a,b]

A. NHW

Gene A	SNP 1	Gene B	SNP 2	p-value
ACTA1	rs506388	*MYBPC2*	rs1274597	0.007
ACTA1	rs728614	*MYH1*	rs2320950	0.004
ACTA1	rs506388	*MYH13*	rs2074877	0.008
ACTA1	rs728614	*MYH13*	rs1859999	0.009
MYL1	rs867342	*MYH1*	rs3744563	0.009
MYL1	rs867342	*MYH8*	rs2270056	0.009
MYL1	rs867342	*MYH8*	rs11078846	0.006
MYH4	rs2058101	*MYH3*	rs201622	0.007
MYH8	rs3744552	*MYH1*	rs8077200	0.007
MYH8	rs3744552	*MYH1*	rs3744563	0.008
MYH8	rs3744552	*MYH4*	rs2058099	0.004
TPM1	rs1972041	*MYH1*	rs2320950	0.003
TPM1	rs12148828	*MYH13*	rs1984620	0.007
TPM2	rs1998308	*MYH2*	rs2277651	0.003
TPM2	rs1998308	*MYH2*	rs3760431	0.008
TPM2	rs1998308	*MYH4*	rs2058101	0.006
TPM2	rs1998308	*MYH4*	rs2011488	0.006
TNNT3	rs2734495	*MYH4*	rs2058099	0.003
TNNT3	rs7395920	*TPM1*	rs3809565	0.008

[a]p-values not corrected for multiple testing

[b]only p-values<0.01 shown

Cited with permission from *Am J Med Genet Part A*, Weymouth, *et.* al, 2011

Table 5.4. Gene Interactions Between SNPs in Different Muscle Contraction Genes (continued)[a,b]

B. Hispanic

Gene A	SNP 1	Gene B	SNP 2	p-value
ACTA1	rs728614	*MYL1*	rs1074158	0.002
MYBPH	rs4950926	*TNNI2*	rs1877444	0.005
MYH1/MYH2	rs9916035	*MYH3*	rs2239933	0.004
MYH1/MYH2	rs9916035	*MYH3*	rs2285475	0.009
MYH13	rs17690195	*TPM1*	rs12148828	0.006
MYH13	rs1859999	*TPM1*	rs3809565	0.004
MYH13	rs1859999	*MYH2*	rs7223755	0.006
MYH13	rs2240579	*MYH3*	rs2285475	0.007
MYH13	rs12936065	*TPM2*	rs2025126	0.007
MYH13	rs12936065	*TNNT3*	rs909116	0.002
TNNC2	rs4629	*MYBPH*	rs2642531	0.002
TNNC2	rs4629	*TPM2*	rs2025126	0.002
TNNC2	rs4629	*TPM2*	rs3750431	0.002
TNNC2	rs4629	*MYBPC2*	rs25665	0.006
TNNC2	rs3848711	*MYBPH*	rs2642531	0.002
TNNC2	rs3848711	*MYBPC2*	rs25665	0.006
TNNC2	rs3848711	*TNNT3*	rs2734510	0.006
TNNC2	rs383112	*MYBPH*	rs2642531	0.003
TNNC2	rs383112	*MYBPC2*	rs25665	0.006
TNNC2	rs383112	*MYBPC2*	rs25667	0.006
TNNC2	rs383112	*MYH4*	rs2058099	0.008
TNNC2	rs437122	*MYH1*	rs8077200	0.009
TNNC2	rs437122	*MYH8*	rs2270056	0.009

[a]p-values not corrected for multiple testing

[b]only p-values<0.01 shown

Cited with permission from *Am J Med Genet Part A*, Weymouth, *et.* al, 2011

Table 5.5. 2-SNP Haplotype Transmission – Validation Population[a,b]

Gene	SNP A	SNP B	p-value
TPM1	rs1972041	rs3809565	<0.00000
TPM1	rs1972041	rs4075583	0.000009
TPM1	rs1972041	rs4238371	0.0002
TPM1	rs1972041	rs12148828	0.0002

[a]p-values not corrected for multiple testing

[b]only p-values<0.01 shown

Cited with permission from *Am J Med Genet Part A*, Weymouth, *et.* al, 2011

Table 5.6. Results of Log-Linear Modeling for *TNNC2* in the NHW Case-Parent Triads[a,b]

SNP	RR Child (95% CI)	RR Mom (95% CI)	LRT Child p-value	LRT Mom p-value
rs4629	**0.74 (0.48-0.97)**	**1.27 (1.03-1.61)**	**0.02**	0.11
rs8860	0.80 (0.94-1.53)	1.20 (0.54-1.04)	0.08	0.22
rs380397	1.24 (1.00-1.53)	0.81 (0.48-1.09)	0.11	0.18
rs383112	**0.77 (0.50-0.99)**	**1.38 (1.13-1.72)**	**0.03**	**0.02**
rs437122	0.79 (0.52-1.03)	1.23 (0.97-1.58)	0.08	0.17
rs3848711	0.80 (0.54-1.03)	1.23 (0.97-1.56)	0.07	0.17

CI, confidence interval; RR, relative risk; LRT, likelihood ratio test

[a]p-value<0.05 and significant CI in bold

[b]Relative risk of the heterozygote compared to the common homozygotes

Cited with permission from *Am J Med Genet Part A*, Weymouth, *et.* al, 2011

Table 5.7. Predicted Transcription Factor Binding Sites

		Alibaba 2		Patch		TESS	
SNP	**Gene**	**Ancestral**	**Alternate**	**Ancestral**	**Alternate**	**Ancesteral**	**Alternate**
rs3809565	***TPM1***	None	None	None	c-myb	RC2	None
rs4075583	***TPM1***	None	None	Lef-1, RUNX2	c-myc, c-myb	None	c-myc
rs9906430	***MYH13***	None	None	None	HIF1A	NF-E	NF-E
rs383112	***TNNC2***	None	AP-2, Sp1, NF-1	None	None	None	None
rs437122	***TNNC2***	NF-1	NF-1	c-FOS, AP-1, CRE-BP1, CREB	NF-E, PKNOX1	CREB	NF-E, GAL4
rs2025126	***TPM2***	MT2A, c-jun	None	HNF1-A	None	NF-1, CP2, CEBPZ	None
rs2145925	***TPM2***	NF-1	SP-1	ETV4	None	NF-1	None
rs909116	***TNNT3***	None	None	LXR-alpha, LXR-beta, RXR-alpha	None	AP-1, ER, ER-alpha	RAF

RUNX2, runt-related transcription factor 2; LEF1, lymphoid enhancer-binding factor 1; c-myc, v-myc myelocytomatosis viral oncogene homolog (avian); c-myb, v-myb myeloblastosis viral oncogene homolog (avian); HIF1A, hypoxia inducible factor 1, alpha subunit; NF-E, nuclear factor E; AP-1/2, activating enhancer binding protein 1/2; NF-1, neurofibromin1; Sp1, simian virus 40 protein 1; MT2A, metallothionein 2A; c-jun, jun proto-oncogene; CP2, ceruloplasmin; CEBPZ, CCAAT/enhancer binding protein (C/EBP), zeta; HNF1A, HNF1 homeobox A; ETVA, ets variant 4; LXR-alpha, liver X receptor, alpha; LXR-beta, liver X receptor, beta; RXR-alpha, Retinoid X receptor alpha; ER, estrogen receptor; ER-alpha, estrogen receptor, alpha; RAF, rapidly accelerated fibrosarcoma; c-FOS, FBJ murine osteosarcoma viral oncogene homology; PKNOX1, Pbx/knotted 1 homeobox; GAL4, galactoside-binding, soluble 4

5.4 Discussion

The focus of this study was to assess whether variants within muscle contraction genes have an etiologic role in clubfoot. We report the first evidence for maternal and inherited genotypic effects involving two *TNNC2* SNPs, rs4629 and rs383112, in the NHW group. Strong evidence for association was found for SNPs located in *TPM1* and *TPM2* for the NHW group (Table 5.2A). In the Hispanic group, the strongest association was with *MYH13* and *TNNT3* SNPs. Interestingly, as seen previously with *IGFBP3* (Chapter 3), multiple associations incorporated regulatory SNPs further suggesting regulation of gene expression as a key mechanism for clubfoot. These results suggest common variants within muscle contraction genes, in particular *TNNC2*, *TPM1* and *TPM2,* play an etiological role in clubfoot.

Evidence for genotypic effects was observed with 2 *TNNC2* SNPs, rs383112 and rs4629, in the NHW group (Table 5.6). A deleterious maternal effect was found for rs383112, while a protective effect was observed for rs4629. *TNNC2* encodes troponin C which plays a key role in initiating muscle contraction in fast-twitch muscles by binding Ca^{2+} causing a conformational change that releases troponin T's inhibition and allowing for actin and myosin to interact[183,195]. The SNP, rs4629 located in exon 5, causes a synonymous change with the alternate allele. Synonymous changes can affect translation rate and cause changes to protein structure and function[196,197]. The deleterious maternal effect SNP, rs383112, is located 1.5 kb upstream of the start site. *In silico* analysis of rs383112 predicts alteration in transcription factor binding depending on the presence of the ancestral or alternate allele (Table 5.7). The genotypic effects of each of these SNPs could alter protein

function/or expression. Though our genotypic effects are very interesting, additional independent datasets need to be assessed for these genotypic effects.

In our NHW discovery group, evidence for association was found with SNPs in *TPM1* and *TPM2*, which encode members of the tropomyosin family, a key component in regulating the interaction of actin and myosin during muscle contraction[183,195]. *TPM1* is expressed in fast-twitch muscle fibers, while *TPM2* is expressed in slow-twitch muscle fibers. An intronic SNP, rs1998308 in *TPM2,* showed suggestive evidence for association in our discovery dataset (p=0.003). In addition, 2-SNP haplotype analysis identified two associated haplotypes incorporating rs1998308 with two potential regulatory SNPs, rs2025126 and rs2145925 that are predicted to alter DNA binding (Table 5.7). *TPM1* SNPs were associated in all three datasets, discovery, validation and case-control, albeit with different SNPs. One of these associated SNPs, rs4075583, is located in a potential regulatory region and predicted to alter DNA binding (Table 5.7), while rs1972041 and rs12148828 are either intronic or downstream depending on the *TPM1* isoforms. Multiple *TPM1* isoforms are produced through alternate splicing and expression is tissue specific[198]. Three *TPM1* regulatory SNPs, including our associated rs4075583, are associated with Metabolic Syndrome and were evaluated for their effect on the expression of the short *TPM1* isoform[198,199]. The presence of the ancestral allele (G) of rs4075583 (the risk allele in our NHW group) decreased gene expression in human embryonic kidney cells (HEK293)[199]. Altered gene expression could affect muscle contraction and needs to be further evaluated in biologically relevant cells such as muscle cells.

In our Hispanic discovery group, the strongest evidence for association was with SNPs in *TNNT3* and *MYH13*. Troponin T type 3 (*TNNT3*) is expressed in fast-twitch

skeletal muscle fibers and is the subunit of the troponin complex that binds the complex to tropomyosin which regulates muscle contraction[183]. Myosin heavy chain 13 (*MYH13)* is a key component of the thick filament that interacts with actin to allow for muscle contraction. Two *MYH13* SNPs, rs9906430, located in a potential regulatory region and predicted to alter DNA binding, and rs17690195, an exonic SNP that creates a missense mutation, were associated in the discovery group (Table 5.7). Three *TNNT3* SNPs were associated including one potential regulatory SNP, rs909116, that is predicted to alter DNA binding (Table 5.7). Interestingly, mutations in *TNNT3* have been implicated in DA2B, which has a clubfoot phenotype[145]. However, a previous study evaluating the coding regions of three skeletal muscle contraction genes, *TNNT3, TPM2* and *MYH3* identified no coding mutations but the regulatory regions of these three genes were not evaluated[149].

As shown in Figure 5.1, muscle contraction is a well-orchestrated process involving multiple proteins[183]. The combination of variants in multiple muscle contraction genes could alter muscle contraction and contribute to clubfoot. Multiple potential variant interactions were associated in both the NHW and Hispanic discovery dataset (Table 5.4A and B). These interactions could not be confirmed in our validation datasets possibly due to small sample size. Interestingly, many of these associated interactions incorporated at least one regulatory SNP. Risk variants in multiple muscle contraction genes could lead to alteration in gene function and/or expression contributing to susceptibility of clubfoot.

This study is the first to identify genotypic effects with SNPs in *TNNC2* for clubfoot. In addition, we showed that a syndromic approach is a valuable technique to use in identifying candidate genes for isolated conditions. For example, mutations in *TNNT3* and *TPM2* have been identified to cause syndromic clubfoot associated with DAs and we found

common variants within these genes that are associated with isolated clubfoot. Utilizing additional genes involved in muscle contraction, we were able to identify another novel gene, *TPM1,* to be associated with isolated clubfoot. Once again, multiple associations involved potential regulatory SNPs; these SNPs need to be further evaluated through functional assays to assess their implications on expression so we can begin to understand their role in the etiology of clubfoot. Furthermore, this study supports the importance of genes involved in muscle contraction and development as key factors to the genetic etiology of clubfoot and should be a focus for future association studies[179].

CHAPTER 6: Analysis of potential regulatory SNPs in HOXA9, TPM1, TPM2 and TNNC2

6.1 Introduction

Clubfoot is a common complex birth defect characterized by the inward posturing of the foot in a rigid, downward position with significant calf muscle hypoplasia[1-3]. Both genetic and environmental factors have been suggested to play an etiologic role in this orthopedic birth defect. However, maternal smoking is the only exogeneous factor that has consistently been associated with clubfoot[93,102,200]. How maternal smoking affects foot development remains to be determined. There are numerous lines of evidence supporting a role for genes including (1) twin studies showing a higher concordance in monozygotic (33%) than dizygotic twins (3%), (2) increased relative risk of a child being affected when a parent and sibling are affected (10-20%), (3) 72% heritability and (4) segregation analyses[88,103,108,172]. Multifactorial inheritance is widely accepted as a model for clubfoot inheritance[73,88,103,108,172,178,181,201].

Candidate gene and GWAS approaches have been applied to clubfoot datasets and large multiplex families with varying levels of success[64,112-114,118,120,121,125,127,149,179,181,182,202]. The most interesting results to date involve *TBX4* and *PITX1*, two genes that play significant limb patterning roles[115-117]. These genes were identified in studies scanning families for associated genes. A microduplication located on 17q23 has been identified in two independent studies and is particularly interesting since it contains *TBX4*, a transcription factor expressed specifically in the hindlimb and plays an important role in muscle patterning[115,116]. The four clubfoot families with the 17q23 microduplication also had family members with other skeletal anomalies such as short wide feet, enlargement of the distal fibular head and shortened calcaneus[114,118]. However, association analysis of eleven *TBX4* SNPs in a large clubfoot dataset consisting of families with and without a family

history of clubfoot found no significant association[118]. Thus, this microduplication is a rare cause for familial clubfoot and is more suggestive of a syndromic form of clubfoot/foot anomalies that are not typically reported with isolated clubfoot.

Interestingly, *TBX4* is a direct transcriptional target of *PITX1*[117]. *PITX1,* a transcription factor required for hindlimb development has also been implicated in clubfoot[6,9,112,113]. A genome-wide study found linkage to 5q31 in one five-generation family in which clubfoot segregated with incomplete penetrance[112]. *PITX1,* in the 5q31 linkage region, was sequenced and a missense mutation (E130K) in a highly conserved homeodomain was found and segregated with the disease phenotype in the family[112]. Decrease in *PITX1* expression was observed in a dose-dependent manner, thus suggested to have a dominant-negative effect on transcription[112]. A microdeletion involving the same 5q31 region containing *PITX1* was found in a second family with clubfoot. The pattern of phenotypic findings suggests that alteration of *PITX1* expression contributes to syndromic clubfoot rather than the nonsyndromic type[113]. However, in support of the role of *PITX1* in nonsyndromic clubfoot, 8.9% of *Pitx1* heterozygote knockout mice had a clubfoot-like phenotype[113].

A recent study by Alvarado *et* al. (2012) evaluated copy number variants (CNVs) in 413 clubfoot probands and compared them to 759 controls. Although there was no difference between number of CNVs in cases and controls, 9 new CNVs segregated with the disease phenotype. Using a gene expression array on E12.5 mouse hindlimb bud, four biologically relevant genes, *HOXC13, RIPPLY2, CHD* and *UTX,* were identified in the new CNV regions. *HOXC13* is a member of the homeobox C gene cluster and is differentially expressed in the hindlimb, while *CHD* and *UTX* have a functional role in regulating the

HOXC13 genes[203-205]. *RIPPLY2* regulates T-box transcription factors during embryogenesis[206]. These interesting results suggest a potential hindlimb transcriptional regulatory pathway that may contribute to clubfoot.

In previous studies, we investigated genes involved in hindlimb skeletal muscle patterning, development and contraction[127,179]. Mutations in these genes cause a range of syndromes that have limb anomalies such as brachydactyly to Distal Arthrogryposis with clubfoot[125,126,139-148,165,207]. The *HOXA* and *D* gene clusters were interrogated because of their roles in limb and muscle patterning and development. Mutations in these gene clusters cause limb anomalies; for example *HOXD13* mutations can cause synpolydactyly and brachydactyly[165,207]. In addition, mutations in muscle contraction genes (*MYH3, TNNT3, TNNI2, TPM2* and *MYBPC1*) cause a spectrum of Distal Arthrogryposis disorders many of which have muscle hypoplasia and clubfoot as part of the phenotype[22,37-46]. Association analysis of these muscle patterning and contraction genes found positive associations with SNPs in the regulatory regions of four genes, rs3801776/*HOXA9*, rs3809565/*TPM1,* rs4075583/*TPM1,* rs2025126/*TPM2*, rs2145925/*TPM2,* rs383112/*TNNC2* and rs437122/*TNNC2*[127,179]. *In silico* analyses of these SNPs predict allele-dependent transcription factor binding sites that could affect gene expression. We hypothesized that variation in expression of one or more of these genes involved in skeletal muscle patterning and function might contribute to nonsyndromic clubfoot. In these studies, we performed functional analysis of seven SNPs to better understand the biologic role of SNP variation in the regulatory regions of *TPM1, TPM2, TNNC2* and *HOXA9*.

6.2 Materials and methods

6.2.1 Electrophoretic mobility shift assays (EMSA)

Genomic sequences were obtained from NCBI (http://www.ncbi.nlm.nih.gov) and the oligonucleotide probes were designed to incorporate approximately 10 base pairs upstream and downstream from the SNP location to create 20 nucleotide DNA probes (20mers). Oligonucleotides were synthesized by Integrated DNA Technologies (Coralville, IA) (Table 6.1). The double stranded DNA 20mer probes were generated by annealing complementary oligonucleotides and end-labeling with corresponding radiolabeled α-^{32}P nucleotides on the forward strand (Perkin Elmer, Waltham, MA). Nuclear extracts from EMSAs were carried out by incubating 4.5 µg (undifferentiated) or 5.00 µg (differentiated) nuclear extract with radiolabeled probe in a 20 µl incubation mixture containing 20mM Tris pH 7.5, 50mM KCl, 10% glycerol, 0.5 mM EDTA, 0.5mM DTT, 0.05% NP-40, 1mM PMSF and 1 µL of dG/dC (Sigma-Aldrich, St. Louis, MO). After incubation at 4ºC for 1 hour, the samples were loaded on 5% polyacrylamide gels that had been prerun for 30 min at 150V in 1X TBE. After electrophoresis (150V) for 2½ hours at room temperature, the gels were dried and radioactive signals were visualized by exposure to a radiograph film at -80°C for 18-36 hrs.

6.2.2 Cell culture technique

C2C12 mouse muscle cells (ATCC# CRL-1772) were obtained from the American Type Culture Collection (ATCC, Manassas, VA). For the undifferentiated cells, the cells were expanded using standard techniques in Gibco® DMEM High Glucose medium (Life Technologies, Grand Island, NY) supplemented with 10% fetal bovine serum. To initiate differentiation, cells were first washed with PBS and media was changed to Gibco® DMEM

High Glucose medium (Life Technologies, Grand Island, NY) supplemented with 5% horse serum.

6.2.3 Generation of TPM1, TPM2, TNNC2 and HOXA9 promoter constructs

Promoter constructs were designed for *TPM1, TPM2, TNNC2* and *HOXA9* to evaluate promoter activity. Since promoter regions for *TPM1* and *TPM2* are not well-defined, we designed constructs to incorporate approximately 500 base pairs upstream from the transcriptional start site since most of the promoter elements are found in this region[208,209]. For *TPM1,* the skeletal muscle isoform was targeted (NM_001018005.1) and the promoter region of *TNNC2* is well characterized and includes 1,625 base pair region upstream of the transcriptional start site which includes a necessary upstream regulatory element (URE) for activity[210]. Primers were designed to incorporate 5'-Xho I and 3'- Bgl II cut sites for ligation into the double digested pGL4.10 luciferase basic vector (pGL4.10, Promega, Madison, WI) using the In-Fusion® HD Cloning System (Clontech, Mountain View, CA). BAC clones containing *TPM1* (RP11-244F12) and *TPM2* (RP-112J3) were amplified to create the ~500 bp inserts using *TPM1* primer set (500 bp, T_m: 55ºC) forward primer, gctcgctagcctcgagCGCGCTCTCCCGGCCTCCGGC and reverse primer, cgccgaggccagatctGGTGGCGGCGGCGAGGGGCC and *TPM2* primer set (461 bp, T_m: 64ºC) forward primer, gctcgctagcctcgagCGGTCCCTGCGCCCGGGCAGC and reverse primer, cgccgaggccagatctGTGAGGACCGGACGGACTGGGCTGGG following standard PCR conditions with the modification of 7-deaza-2'-deoxyguanosine 5'-triphosphate (dc^7GTP) being added. For *TNNC2,* genomic DNA from a control sample was amplified using the primer set (1625 bp, T_m: 65ºC), forward primer, gctcgctagcctcgagCCCTCACCCTTTGGCACCCTG and reverser primer,

cgccgaggccagatctGGTGACCGGGACTCCTCTGTTG following standard PCR conditions. All constructs were sequenced and compared to the NCBI consensus sequence (http://www.ncbi.nlm.nih.gov) using Sequencher v4.8 (Gene Codes, Ann Arbor, MI). The 400 bp *HOXA9* promoter construct was obtained from Dr. Chandrashekhar V. Patel and construct design is described in Trivedi *et* al., 2008[211].

6.2.4 Generation of regulatory SNP with promoter constructs

Potential regulatory SNPs in *TPM1, TPM2, TNNC2* and *HOXA9* that differed in allele-dependent DNA-binding ability through the EMSAs were further evaluated to determine their effect on promoter activity. The ancestral and alternate 20mers for each SNP used in the EMSA were redesigned to incorporate 5'-KpnI and 3'-XhoI cut site overhangs to allow for direct ligation into the front of double digested corresponding promoter vector (described above). The ancestral allele of rs3801776 was within the 400 bp *HOXA9* construct. Site-directed mutagenesis was used to create the *HOXA9* alternate allele construct using QuikChange® II following manufacturer's protocol (Agilent Technologies, Santa Clara, CA).

6.2.5 Evaluation of common TPM1 haplotypes on skeletal muscle TPM1 isoform promoter activity

6.2.5.1 Identification of common TPM1 haplotypes

A 1,774 base pair region containing eight *TPM1* SNPS was obtained from Savill *et* al.[199] To establish *TPM1* haplotype frequency in the general population (control), genotype data was obtained from the 1000 genome project (http://browser.1000genomes.org/index.html) for each of the eight *TPM1* SNPs in the 1,774 base pair region. Individuals of Mexican Ancestry (MXL), obtained from the 1000 genome

project, were used as controls for our Hispanic cases that were also of Mexican ancestry. To establish *TPM1* haplotype frequencies in our clubfoot dataset, DNA from 64 nonHispanic White (NHW) probands (28 multiplex (+FH) and 36 simplex (─FH)) and 73 Hispanic probands (21 multiplex (+FH) and 52 simplex (─FH)) were sequenced. Two primers sets (set 1 (940 bp, T_m: 63°C): forward primer, ACTCACCTGAAACTGACCTTCCCA; reverse primer, AAGTCACGCAGCAGGAAACTAGGA; set 2: (1,281 bp, T_m: 56°C): forward primer, ATGGGCCTCAGCCTGACTCTTAAA; reverse primer, AACGGGTGGTGTTGAGAAGGTTCT) were designed to amplify the ~1.7 kb *TPM1* region using standard PCR conditions. Sequences were compared to the NCBI consensus sequence (http://www.ncbi.nlm.nih.gov, NM_001018005.1). Genotype data was inputted into Haploview to generate the common haplotypes in the case and control populations[150].

6.2.5.2 Evaluation of TPM1 haplotypes on promoter activity

Previously, Savill *et* al. found that 3 SNP haplotypes within a ~1.7 kb region upstream of the short *TPM1* isoform differentially affected expression[199]. We evaluated the four common *TPM1* haplotypes for effect on promoter activity of the skeletal muscle *TPM1* isoform. Common haplotype 2-4 inserts were obtained through XhoI and SacI double digestion of the constructs of Savill *et.* al using standard techniques (see Savill *et.* al, 2010 for construct designs)[199]. Inserts were ligated into the double digested skeletal *TPM1* promoter construct (Figure 6.2A). Site-specific mutagenesis was used to create the *TPM1* haplotype 1 (Table 6.2) using QuikChange® II following the manufacturer's protocol (Agilent Technologies, Santa Clara, CA).

6.2.6 Luciferase assays

C2C12 cells (100,000 cells/well) were seeded in 12-well plates for 24 hrs before transfection. For transfection, 1.12 μg of luciferase reporter construct, .048 μg of *Renilla* internal control and Opti-MEM were incubated with FuGENE HD (Promega, Madison, WI) following the manufacturer's protocol. All experiments were performed in triplicate and repeated three independent times. For undifferentiated C2C12 cells, luciferase activities were determined 48 hrs after transfection using the dual-luciferase system (Promega, Madison, WI). For differentiated C2C12 cells, 48 hours after transfection the media was replaced with DMEM medium supplemented with 2% horse serum (Life Technologies, Grand Island, NY) and five days later luciferase activities were determined. Unpaired t-tests were used to compare luciferase expression between constructs.

6.2.7 Comparison of allelic variation (genetic signature) across genes

Genotype data for rs3801776/*HOXA9*, rs2025126/*TPM2*, rs2145925/*TPM2* and rs4075583/*TPM1* from 258 NHW and 288 Hispanic clubfoot probands and controls including 174 NHW (CEU+GBR) and 66 Hispanic (MXL) individuals was entered into Haploview[150]. The custom analysis tool in Haploview was used to generate the common genetic signatures. Associations between clubfoot and each gene, on its own and along with the other genes (as factors within an additive model), were assessed using logistic regression models. Odds ratios and confidence intervals were calculated separately for NHW and Hispanics. Additionally, the allelic burden within each study subject was evaluated by summing the alleles (0 for homozygous ancestral allele, 1 for homozygous alternate allele and 2 for heterozygous allele). Mann-Whitney U tests were used to compare the allelic sums between case and controls (stratified by ethnicity).

Table 6.1. Probes used in EMSAs

Gene	SNP	Forward oligo*	Reverse oligo*
TPM1	rs4075583	ATTCTTGC[G/A]GTTGGGATCA	CCTGATCCCAAC[C/T]GCAAGAAT
TPM1	rs4075584	CTCTGGCTCC[T/C]GGGCATG	CCCATGCCC[A/G]GGAGCCAGAG
TPM1	rs73431508	AGAAGTGG[A/G]AGCCAGAGC	GGGCTCTGGCT[T/C]CCACTTCT
TPM1	rs79854225	GCCAGACACCC[G/A]GTTCCC	GGGGGAAC[C/T]GGGTGTCTGGC
TPM1	rs111470259	GGACAGCCGCGG[C/T]AGCCG	CCCGGCT[G/A]CCGCGGCTGTCC
TPM1	rs4075047	GGAAGTCGC[G/A]GCCTCCAG	CCCTGGAGGC[C/T]GCGACTTCC
TPM1	rs76273871	GAGCCCCAGGGAG[G/A]CTGGC	GGCCAG[C/T]CTCCCTGGGGCTC
TPM1	rs57645645	AACCTGA[G/A]GAGAAAAAGC	GGGCTTTTTCTC[C/T]TCAGGTT
TPM1	rs3809565	CTATTAAC[A/G]AGACCCTCA	GGTGAGGGTCT[T/C]GTTAATAG
TPM2	rs2025126	GGACAGAGT[G/A]GCTGGATG	CCCATCCAGC[C/T]ACTCTGTCC
TPM2	rs2145925	GGCTACTGGGA[T/C]GGAAGC	CCGCTTCC[A/G]TCCCAGTAGCC
TNNC2	rs383112	GCACTGGGGAG[T/C]AGGCAA	CCTTGCCT[A/G]CTCCCCAGTGC
TNNC2	rs437122	GTCGGAGGC[T/C]GTCAGCTT	GGAAGCTGAC[A/G]GCCTCCGAC

*Ancestral allele listed first based on NCBI CEU listing

6.3 Results

6.3.1 The allele-specific transcription factor binding site of rs4075583/TPM1 does not influence TPM1 promoter activity.

In previous studies, we described associations between four *TPM1* SNPs and clubfoot[179]. Two of these associated SNPs, rs3809565 and rs4075583 are located upstream or within the first intron of *TPM1,* respectively*,* which are potential regulatory regions. *In silico* analysis predicted allele-specific transcription factor binding affinity for rs4075583 and rs3809565, this could lead to changes in transcription. Using undifferentiated and differentiated C2C12 (mouse muscle cells) nuclear extracts, EMSAs were used to test the allele-dependent presence of different transcription factor binding sites for each SNP. As shown in Figure 6.1B, the ancestral allele of rs4075583/*TPM1* produced a DNA-protein complex in both the undifferentiated and differentiated nuclear extracts, while rs3805965/*TPM1* did not show allele-specific binding (data not shown).

Luciferase assays were used to assess whether the allele-specific binding of the 20mers containing the ancestral and alternate alleles of rs4075583/*TPM1* could influence promoter activity in either undifferentiated or differentiated C2C12 cells. As shown in Figure 6.1C and D, although a significant increase in promoter activity was found when the rs4075583/*TPM1* 20mers were upstream of the *TPM1* promoter construct, there was no significant difference in activity between the ancestral and alternate alleles. These results suggest that the genomic region containing rs4075583 influences promoter activity but the allele-dependent transcription factor binding site does not significantly alter promoter activity.

6.3.2 The allele-specific transcription factor binding site of rs4075583 in the context of haplotypes significantly decreases TPM1 promoter activity.

Savill *et* al. (2010) previously reported that three SNPs, rs4075583, rs4075584 and rs4075047, in a ~1.7 kb 5' upstream region of the short form of *TPM1* show differential expression dependent on cell type and specific haplotype[199]. Although the ancestral and alternate forms of rs4075583/*TPM1* showed no significant difference in promoter activity, we evaluated the ~1.7 kb region that contains rs4075583 and seven additional SNPs in the context of a haplotype for effect on promoter activity of the skeletal muscle *TPM1* isoform in C2C12 mouse muscle cells. We identified the most common *TPM1* haplotypes in control and clubfoot datasets. As shown in Table 6.2, haplotype 1 was the most common in both ethnicities for both cases and controls. The second most common haplotype was haplotype 2 in NHWs and haplotype 3 in Hispanics. This supports the importance of treating each ethnicity individually. Before assessing the impact of these four haplotypes on the promoter activity of the skeletal muscle *TPM1* isoform, EMSAs were performed on the seven additional *TPM1* SNPs to evaluate for allele-dependent transcription factor binding sites. No DNA-protein complexes were observed with any of these SNPs (data not shown). However as shown in Figure 6.2B and C, the haplotype constructs with all eight SNPS significantly affected promoter activity. Haplotype 1 with the alternative form (A) of rs4075583/*TPM1* produced the greatest expression. In contrast, haplotypes 2-4 with the ancestral allele (G) of rs4075583 showed significant decreased promoter activity (Figure 6.2B and C). Interestingly, haplotype 3, which produces the least promoter activity, has the most allelic variation with five different variants compared to haplotype 1 which was the most common.

These results suggest that evaluation of individual SNP expression does not provide a complete functional assessment.

6.3.3 The allele-specific transcription factor binding affinity for rs2025126 and rs2145925 influences TPM2 promoter activity.

Previously, we reported associations between five *TPM2* SNPs and clubfoot[179]. Two SNPs are located in potential regulatory regions, rs2025126, upstream and rs2145925 in intron 2 of *TPM2* *In silico* analysis predicted the presence of allele-dependent transcription factor binding sites for each SNP. As shown in Figures 6.3B and 6.4B, only the ancestral allele of rs2025126/*TPM2* created a DNA-protein complex, while rs2145925 created two DNA-protein complexes associated with the presence of the alternate allele. Figures 6.3 and 6.4 show that the 20mers of rs2025126 and rs2145925 increase *TPM2* promoter activity, suggesting that these SNPs have a regulatory function. While the transcription factor binding affinity of the ancestral allele of rs2025126 increased promoter activity in differentiated cells, no significant effect on promoter activity was found in undifferentiated cells (Figure 6.3C and D). Interestingly, the binding affinity of the alternate allele of rs2145925/*TPM2* that creates two DNA-protein complexes caused different effects on promoter activity during the different myogenesis stages used in this study (Figure 6.4C and D). The transcription factors binding associated with the alternate allele of rs2145925/*TPM2* caused a decrease in promoter activity with the undifferentiated cells, while an increase in activity was found with differentiated cells (Figure 6.4C and D). These results for *TPM2* SNPs suggest that the allele-dependent presence of different transcription factor binding sites could play key roles during specific stages of myogenesis.

6.3.4 The allele-dependent DNA-binding ability of rs437122/TNNC2 creates a DNA-protein complex that does not influence promoter activity.

Previously, we reported a genotypic effect with rs383112/*TNNC2* and gene interactions with rs437112/*TNNC2* and clubfoot[179]. *In silico* analysis of these SNPs suggested that allele-specific transcription factor binding sites differed between alleles for each SNP. As shown in Figure 6.5B, a DNA-protein complex was generated with the ancestral allele of rs437122/*TNNC2*. None were formed with rs383112/*TNNC2* (data not shown). As observed previously with the other SNPs, inclusion of the ancestral and alternate 20mers upstream of the promoter produced significant increased activity (Figure 6.5C). However, even with allele-specific DNA-protein complex observed in the EMSA (Figure 6.5B), no significant difference in promoter activity was found between the alleles of the SNP (Figure 6.5B).

6.3.5 The allele-dependent transcription factor binding of rs3801776/HOXA9 influences promoter activity.

Previously, we found that rs3801776, a SNP located in the basal promoter of *HOXA9,* was significantly associated with clubfoot [127]. This is important because *HOXA9* plays a role in muscle patterning and development[124]. As shown in Figure 6.6B, the ancestral allele of rs3801776/*TNNC2* created a DNA-protein complex that was not present with the alternate allele. A four hundred base pair *HOXA9* construct containing the basal promoter of *HOXA9* where rs3801776 is located was used to evaluate rs3801776 effect on promoter activity. This construct is described in Trivedi *et al.*[211]. As shown in Figure 6.6C and D, for both undifferentiated and differentiated muscle cells a significant decrease in expression was associated with the alternate allele-dependent non-transcription factor

binding site. These results suggest that the differing allele-dependent transcription factor binding of rs3801776/*HOXA9* has a regulatory function that affects *HOXA9* expression.

6.3.6 Genetic signature across genes is shared by both cases and controls.

We found that allelic variation in three SNPs, rs3801776/*HOX9*, rs2025126/*TPM2* and rs2145925/*TPM2* altered promoter activity. In addition, the ancestral allele of rs4075583/*TPM1* altered promoter activity in the context of a haplotype. We asked whether different allelic combinations (genetic signature) of these four SNPs differed between cases and controls. As shown in Table 6.3, no difference was found between the case and control groups for either ethnicity. However, there were differences in frequency between the NHW and Hispanics for the genetic signatures providing additional support for the importance of evaluating each ethnicity separately.

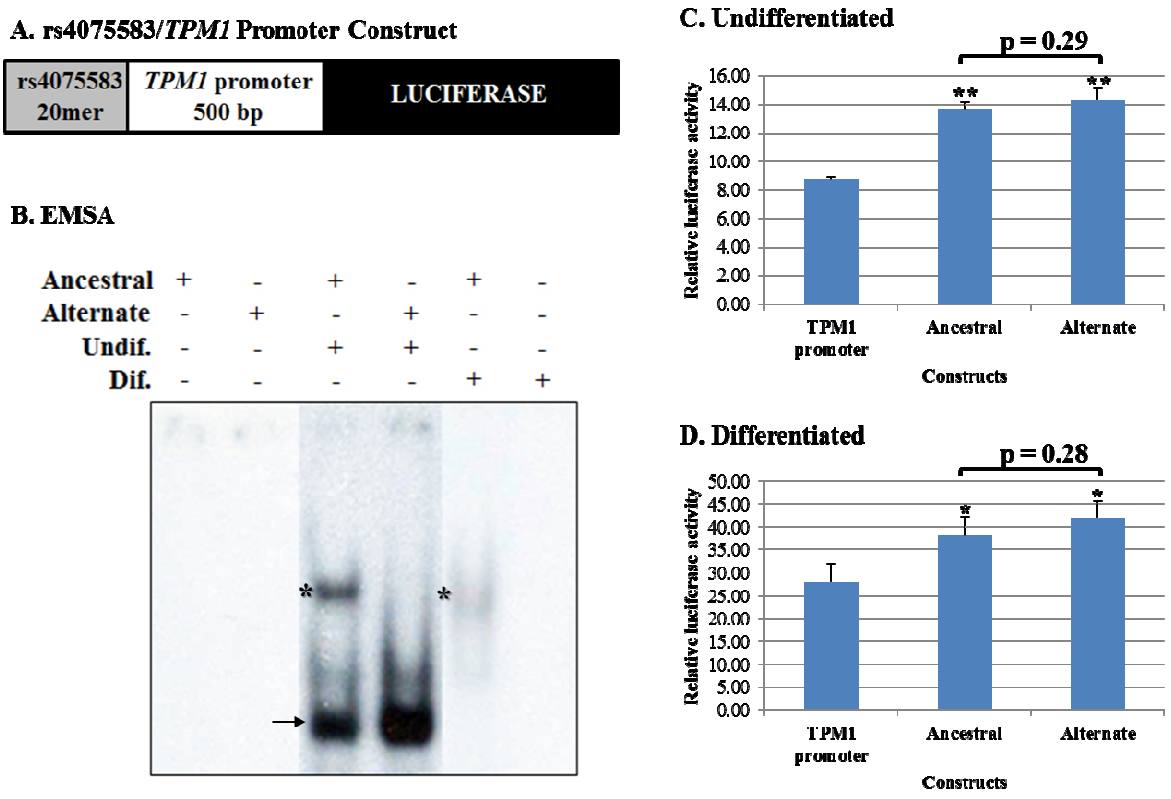

Figure 6.1. Functional analysis of rs4075583/*TPM1*. (A) Luciferase constructs were designed to incorporate 500 base pairs upstream of the transcriptional start site of the *TPM1* skeletal muscle isoform (promoter construct). The ancestral and alternate luciferase constructs were generated by using the twenty base pair double stranded oligonucleotides that incorporated either the ancestral or alternate allele used in the EMSA and ligating them in front of the *TPM1* promoter construct. (B) ^{32}P-labeled double stranded oligonucleotides incorporating either he ancestral or alternate allele of rs4075583 were incubated with undifferentiated (undif.) (4.5 µg) and differentiated (dif.) (5.00 µg) C2C12 nuclear extract. Asterisks indicate DNA-protein complexes. Arrow shows nonspecific band in EMSAs performed with undifferentiated nuclear extract. (C and D) Luciferase test constructs were co-transfected with a a Renilla reporter construct (pGL4.73) into C2C12 undifferentiated (B) and differentiated (C) cells. Luciferase and Renilla activities were determined and luciferase activities were normalized to the Renilla activity to correct for variation in transfection efficiencies. The data represents the mean values ±SD from three independent experiments done in triplicate. Unpaired t-tests were used to compare luciferase expression between the ancestral and alternate constructs. While ancestral allele creates a DNA-protein complex, no significant difference in luciferase expression was observed in either undifferentiated or differentiated muscle cells.
*: 0.01<p<0.05, **: p<0.01: Unpaired t-tests results for comparison to promoter construct
Undif: undifferentiated; Dif: differentiated.

Table 6.2. Frequency of *TPM1* haplotypes

A. nonHispanic Whites

	SNP genotype								Frequency			
									Controls	Cases		
Haplotype	rs79854225	rs57645645	rs76273871	rs4075583	rs4075584	rs4075047	rs73431508	rs111470259	ALL (n=85)	ALL (n=64)	+ FH (n=28)	− FH (n=36)
1	G	G	G	A	T	A	T	C	0.641	0.687	0.642	0.722
2	G	G	G	G	C	A	T	C	0.194	0.140	0.178	0.111
3	A	A	G	G	T	G	C	T	0.082	0.055	0.054	0.056
4	G	G	A	G	T	A	T	C	0.059	0.055	0.018	0.083

B. Hispanics

	SNP genotype								Frequency			
									Controls	Cases		
Haplotype	rs79854225	rs57645645	rs76273871	rs4075583	rs4075584	rs4075047	rs73431508	rs111470259	ALL (n=66)	ALL (n=73)	+ FH (n=21)	− FH (n=52)
1	G	G	G	A	T	A	T	C	0.621	0.699	0.667	0.712
2	G	G	G	G	C	A	T	C	0.106	0.048	0.048	0.048
3	A	A	G	G	T	G	C	T	0.220	0.171	0.190	0.163
4	G	G	A	G	T	A	T	C	0.030	0.014	0.000	0.019

FH: family history; ALL includes both +FH and − FH

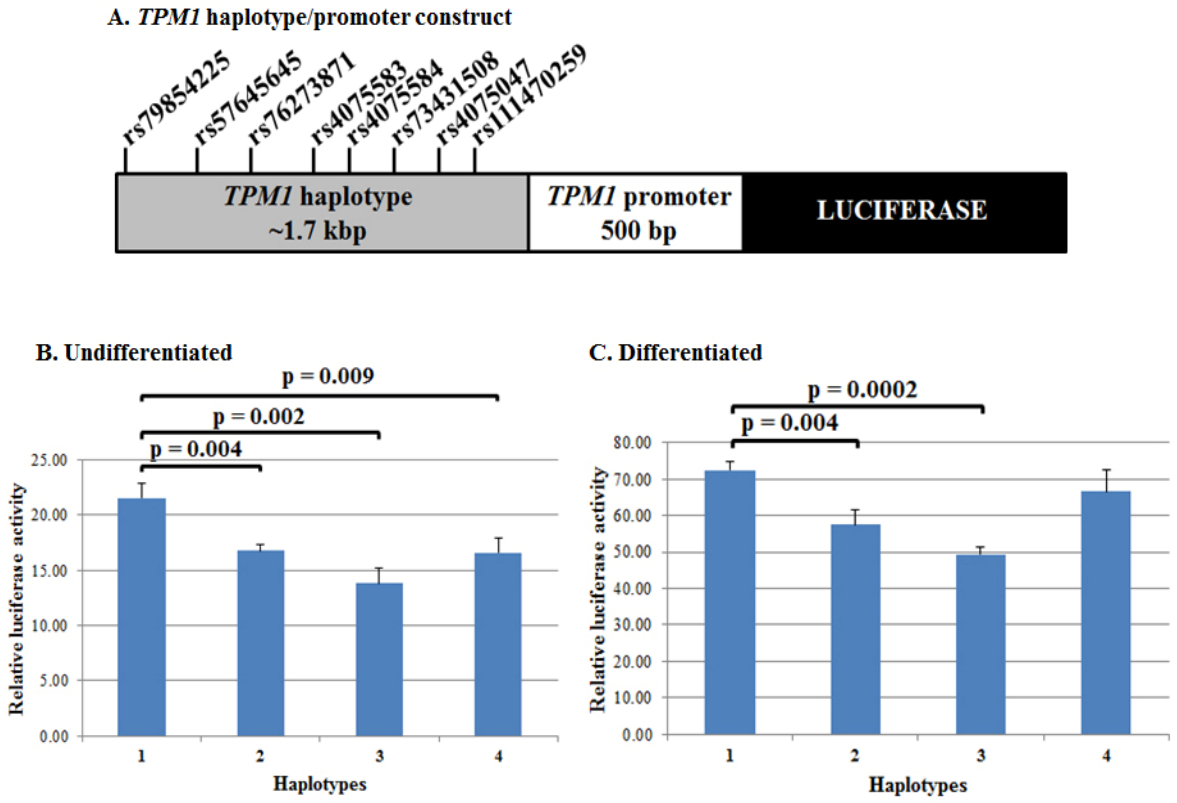

Figure 6.2. Functional analysis of four common *TPM1* haplotypes. (A) A 1,774 bp *TPM1* fragment was inserted in front of the 500 bp *TPM1* promoter construct. The haplotype constructs were co-transfected with a Renilla reporter construct (pGL4.73) into C2C12 undifferentiated (B) and differentiated (C) cells. Luciferase and Renilla activities of cell extracts were determined and luciferase activities were normalized to the Renilla activity to correct for variation in transfection efficiencies. The data represents the mean values ±SD from three independent experiments done in triplicate. Unpaired t-tests were used to compare luciferase expression between haplotype constructs. Decrease in luciferase expression observed when the ancestral allele of rs4075583 is present in the haplotype (2-4).

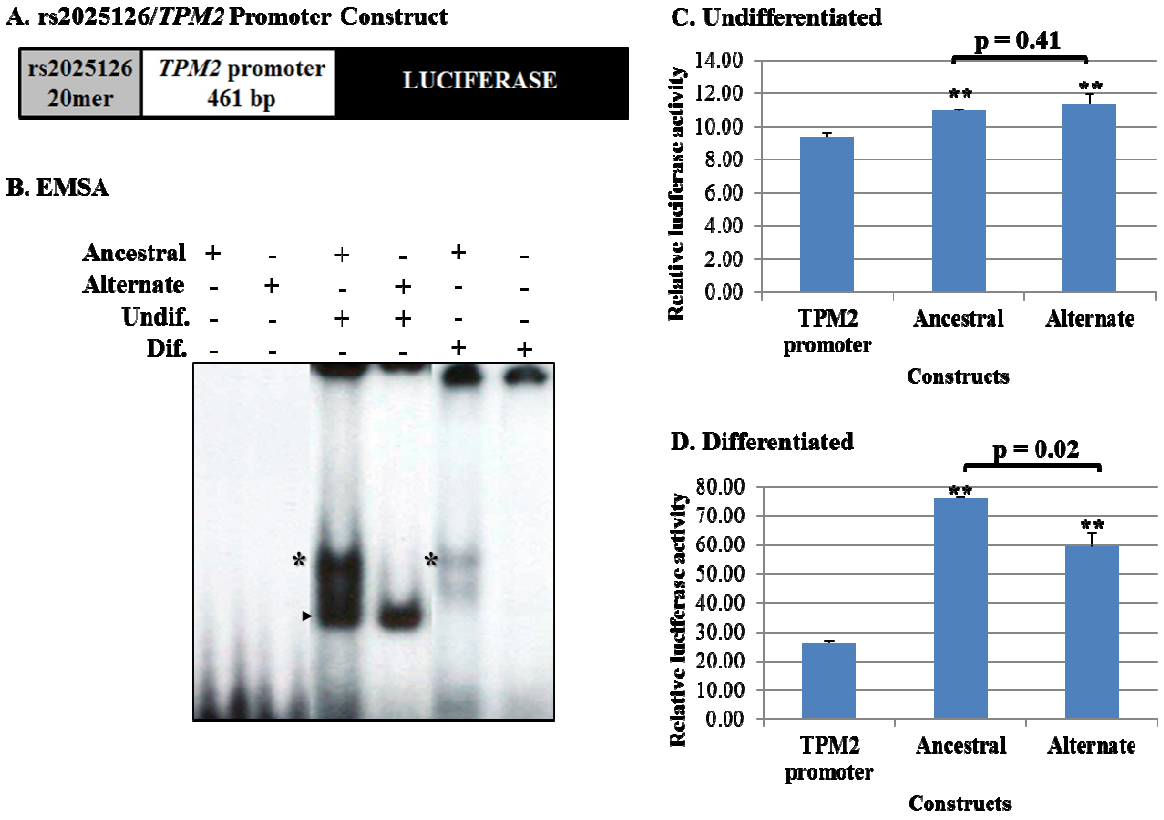

Figure 6.3. Functional analysis of rs2025126/*TPM2*. (A) Luciferase constructs were designed to incorporate 500 base pairs upstream of the transcriptional start site of the *TPM1* skeletal muscle isoform (promoter construct). The ancestral and alternate luciferase constructs were generated by using the twenty base pair double stranded oligonucleotides that incorporated either the ancestral or alternate allele used in the EMSA and ligating them in front of the *TPM1* promoter construct. (B) ^{32}P-labeled double stranded oligonucleotides incorporating either he ancestral or alternate allele of rs4075583 were incubated with undifferentiated (undif.) (4.5 µg) and differentiated (dif.) (5.00 µg) C2C12 nuclear extract. Asterisks indicate DNA-protein complexes. Arrow shows nonspecific band in EMSAs performed with undifferentiated nuclear extract. (C and D) Luciferase test constructs were co-transfected with a a Renilla reporter construct (pGL4.73) into C2C12 undifferentiated (B) and differentiated (C) cells. Luciferase and Renilla activities were determined and luciferase activities were normalized to the Renilla activity to correct for variation in transfection efficiencies. The data represents the mean values ±SD from three independent experiments done in triplicate. Unpaired t-tests were used to compare luciferase expression between the ancestral and alternate constructs. While the ancestral allele creates a DNA-protein complex with both undifferentiated and differentiated nuclear extract, significant decrease in luciferase expression was only observed in the differentiated muscle with the alternate allele. *: 0.01<p<0.05, **: p<0.01: Unpaired t-tests results for comparison to promoter construct Undif: undifferentiated; Dif: differentiated.

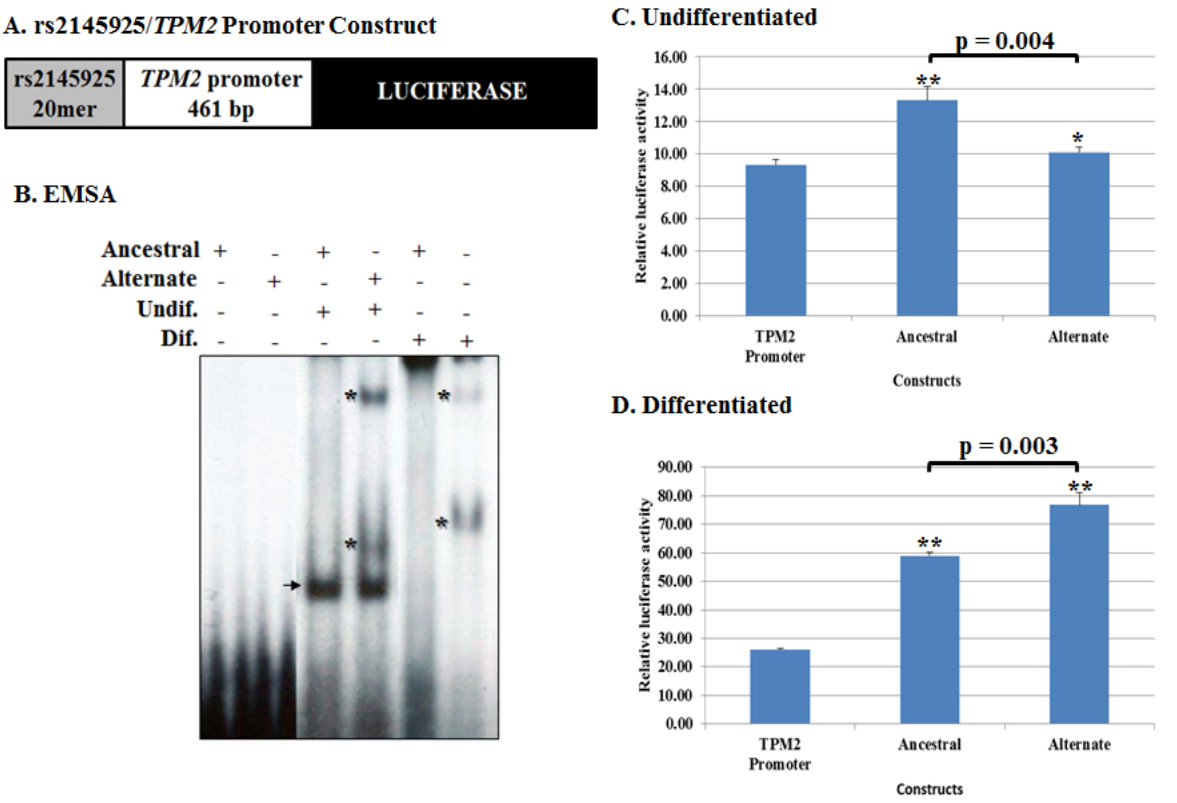

Figure 6.4. Functional analysis of rs2145925/*TPM2*. (A) Luciferase constructs were designed to incorporate 500 base pairs upstream of the transcriptional start site of the *TPM1* skeletal muscle isoform (promoter construct). The ancestral and alternate luciferase constructs were generated by using the twenty base pair double stranded oligonucleotides that incorporated either the ancestral or alternate allele used in the EMSA and ligating them in front of the *TPM1* promoter construct. (B) ^{32}P-labeled double stranded oligonucleotides incorporating either he ancestral or alternate allele of rs4075583 were incubated with undifferentiated (undif.) (4.5 µg) and differentiated (dif.) (5.00 µg) C2C12 nuclear extract. Asterisks indicate DNA-protein complexes. Arrow shows nonspecific band in EMSAs performed with undifferentiated nuclear extract. (C and D) Luciferase test constructs were co-transfected with a Renilla reporter construct (pGL4.73) into C2C12 undifferentiated (B) and differentiated (C) cells. Luciferase and Renilla activities were determined and luciferase activities were normalized to the Renilla activity to correct for variation in transfection efficiencies. The data represents the mean values ±SD from three independent experiments done in triplicate. Unpaired t-tests were used to compare luciferase expression between the ancestral and alternate constructs. The alternate allele creates two DNA-protein complexes that significantly affect luciferase expression in both undifferentiated and differentiated muscle cells.
*: 0.01<p<0.05, **: p<0.01: Unpaired t-test results for comparison to promoter construct
Undif: undifferentiated; Dif: differentiated.

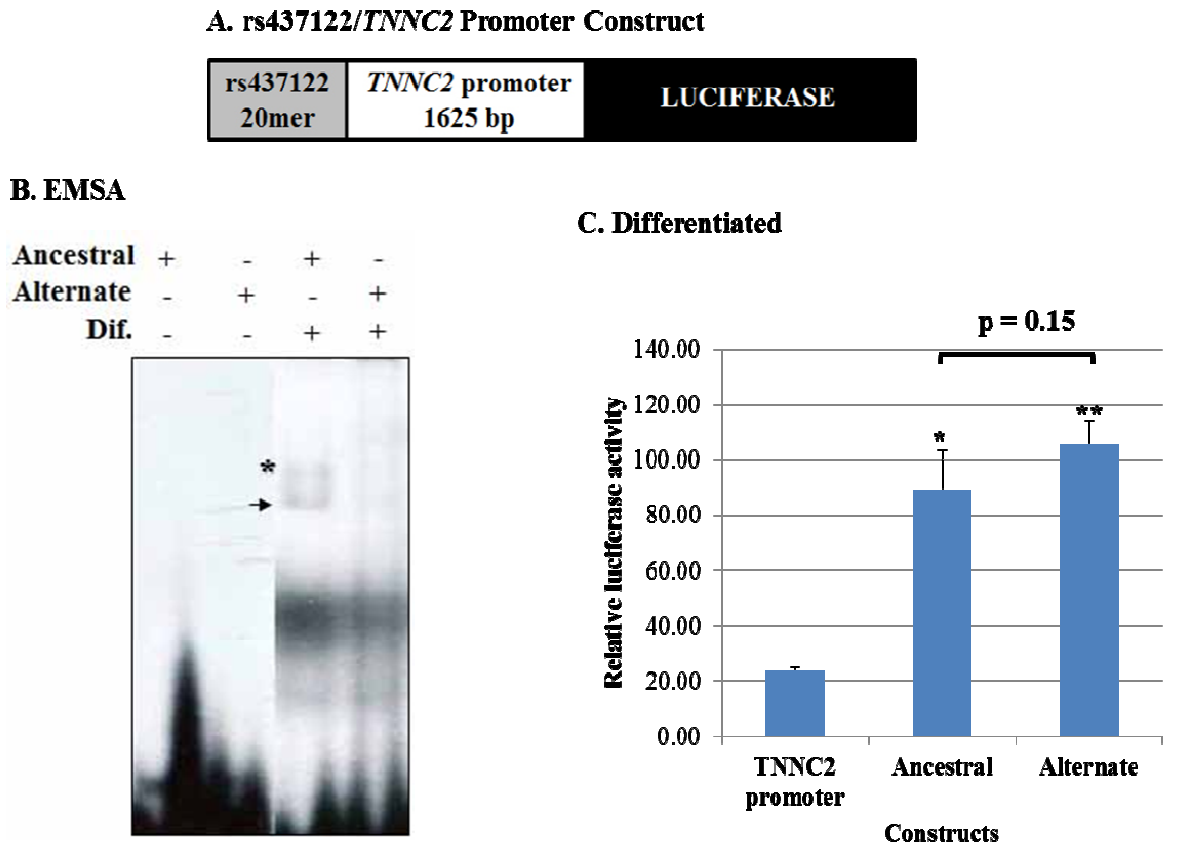

Figure 6.5. Functional Analysis of rs437122/*TNNC2*. (A) Luciferase constructs were designed to incorporate 500 base pairs upstream of the transcriptional start site of the *TPM1* skeletal muscle isoform (promoter construct). The ancestral and alternate luciferase constructs were generated by using the twenty base pair double stranded oligonucleotides that incorporated either the ancestral or alternate allele used in the EMSA and ligating them in front of the *TPM1* promoter construct. (B) ^{32}P-labeled double stranded oligonucleotides incorporating either the ancestral or alternate allele of rs437122 were incubated with differentiated (5.00 µg) C2C12 nuclear extract. Asterisk indicates DNA-protein complex. Arrow depicts nonspecific band. (C) Luciferase test constructs were co-transfected with a Renilla reporter construct (pGL4.73) into C2C12 differentiated cells. Luciferase and Renilla activities were determined and luciferase activities were normalized to the Renilla activity to correct for variation in transfection efficiencies. The data represents the mean values ±SD from three independent experiments done in triplicate. Unpaired t-tests were used to compare luciferase expression between the ancestral and alternate constructs. While the alternate allele creates a DNA-protein complex, no significant difference in luciferase expression was observed in differentiated muscle cells.
*: 0.01<p<0.05, **: p<0.01: Unpaired t-test results for comparison to promoter construct
Dif: differentiated.

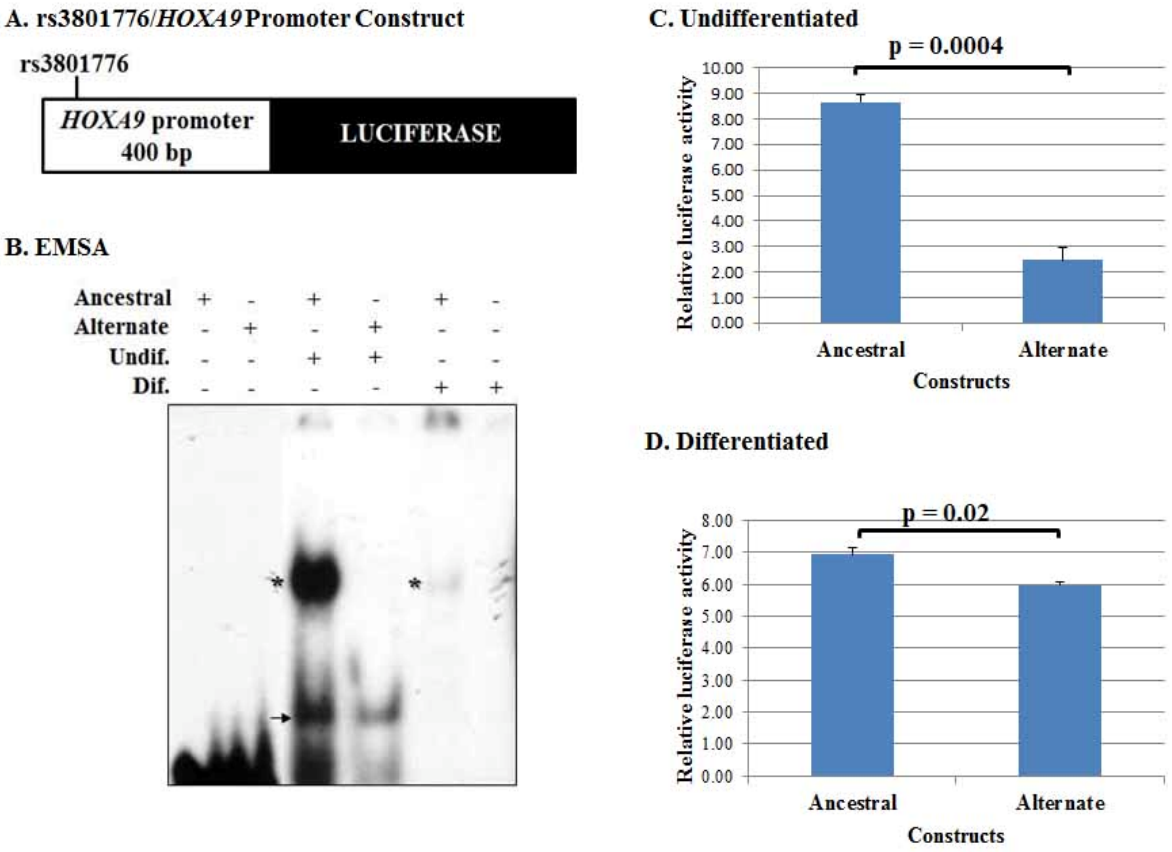

Figure 6.6. Functional analysis of rs3801776/*HOXA9*. (A) The *HOXA9* luciferase construct was obtained from the Patel lab and incorporated rs3801776 within the 400 base pair promoter region. Site-specific mutagenesis was utilized to create the alternate allele luciferase construct. (B) ^{32}P-labeled double stranded oligonucleotides incorporating either the ancestral or alternate allele of rs3801776 were incubated with undifferentiated (4.5 µg) and differentiated (5.00 µg) C2C12 nuclear extract. Asterisks indicate DNA-protein complexes. Arrow depicts nonspecific band in gel shift assays performed with undifferentiated nuclear extract. (C and D) The *HOXA9* basal promoter construct incorporates the ancestral allele of rs3801779. Site-specific mutagenesis was used to create the alternate allele construct. The test constructs were co-transfected with a Renilla reporter construct into C2C12 undifferentiated (B) and differentiated (C) cells. Luciferase and Renilla activities were determined and luciferase activities were normalized to the Renilla activity to correct for variation in transfection efficiencies. The data represents the mean values ±SD from three independent experiments done in triplicate. Unpaired t-tests were used to compare luciferase expression between the ancestral and alternate constructs. The ancestral allele creates a DNA-protein complex that significantly decreases luciferase expression in both undifferentiated and differentiated muscle cells.
Undif: undifferentiated; Dif: differentiated.

Table 6.3. Genetic Signatures*

				nonHispanic Whites				Hispanics			
					Cases				Cases		
rs3801776 (*HOXA9*)	**rs2025126 (*TPM2*)**	**rs2145925 (*TPM2*)**	**rs4075583 (*TPM1*)**	**Control (N=174)**	**ALL (N=258)**	**FH (N=129)**	**No FH (N=129)**	**Control (N=66)**	**All (N=288)**	**FH (N=86)**	**No FH (N=202)**
G	G	C	A	**0.212**	**0.219**	**0.254**	**0.190**	**0.141**	**0.160**	**0.180**	**0.150**
G	A	T	A	**0.138**	**0.156**	**0.168**	**0.140**	**0.114**	**0.130**	**0.106**	**0.143**
G	G	C	G	**0.131**	**0.124**	**0.142**	**0.104**	0.098	0.056	0.044	0.061
G	G	T	A	**0.107**	**0.104**	**0.084**	**0.121**	**0.184**	**0.173**	**0.153**	**0.183**
A	G	T	A	0.067	**0.101**	**0.101**	**0.104**	0.073	**0.118**	**0.109**	**0.116**
G	A	T	G	0.053	0.065	0.036	0.090	**0.103**	0.053	0.064	0.049

*frequency greater than 10% in bold

6.4 Discussion

Clubfoot is characterized by the inward posturing of the foot in a rigid, downward position, almost as if the foot is constrained in a contracted state[1]. Calf muscle hypoplasia is generally universal and is persistent even after corrective treatment[2,3]. This constellation of anomalies suggested a role for limb patterning and skeletal muscle genes in clubfoot. In previous studies, *HOXA* and *HOXD* gene clusters and fifteen muscle contraction genes were interrogated leading to the identification of multiple associations[127,179]. The majority of these associated SNPs were located in potential regulatory regions, which could lead to allele-dependent transcription factor binding sites that could affect gene expression. To determine if any of these SNPs have regulatory function, we functionally assessed seven potential regulatory SNPs in four of our strongest associated genes, *TPM1, TPM2, TNNC2* and *HOXA9*.

Five SNPs, rs3801776/*HOXA9*, rs2025126/*TPM2,* rs2145925/*TPM2*, rs437122/*TNNC2* and rs4075583/*TPM1,* showed allele-dependent transcription factor binding sites that differed between ancestral and alternate alleles by EMSA. To determine effect on promoter activity, each SNP, in the context of a 20mer, was tested by ligating it in front of its corresponding promoter construct and expressed in biologically relevant C2C12 muscle cells. Because muscle development occurs in two stages with cells beginning as myoblasts (undifferentiated) and developing into myotubes (differentiated), requiring different transcription factors during each stage, these experiments were performed during both muscle stages[41]. The incorporation of each 20mer caused an increase in luciferase activity suggesting that these genomic regions have regulatory function.

Previously, we found associations with SNPs in the *HOXA* gene cluster and clubfoot[127]. The strongest association was with rs3801776, a SNP located in the basal promoter of *HOXA9*[127]. The ancestral allele of rs3801776 creates a DNA-protein complex that lead to an increase in promoter activity (Figure 6.6B-D). *HOXA9* is a member of the homeobox A gene cluster, a group of transcription factors expressed in fore- and hindlimb muscles that are involved in patterning and differentiation of muscles in both embryonic limbs and adult limbs during muscle repair[124]. The *HOXA* genes are also known to regulate the synchronized development of muscles, tendons and cartilages[124,212]. Mutations in these genes cause mammalian limb abnormalities[162,164]. In mice, *Hoxa13* mutations cause hypodactyly, while deletion of *Hoxa13* causes absence of the autopod[162,213]. Therefore, enhanced promoter activity associated with rs3801776 may alter muscle patterning by altering slow and fast-twitch muscle designations, ultimately causing changes in muscle function.

Previously, we found genotypic effects with rs383112/*TNNC2* and significant gene interactions with rs437122/*TNNC2*, both SNPs located in potential *TNNC2* regulatory regions[179]. *TNNC2* encodes troponin C and plays a role in initiating muscle contraction in fast-twitch muscle fibers binding Ca^{2+}[183]. This causes a conformational change in troponin I, which releases inhibition of troponin T causing tropomyosin to allow actin-myosin interactions[183,195]. No allele-dependent transcription factor binding sites were observed for rs383112, located upstream of *TNNC2*. Although rs437122 located in intron 1 of *TNNC2* showed an allele-dependent transcription factor binding site with the ancestral allele, however, this binding site did not affect promoter activity (Figure 6.5B and C). While

neither SNP showed altered gene expression, they may tag for other SNPs in the region that could affect the regulation of *TNNC2*.

Previously, we found associations with four SNPs in two members of the tropomyosin family, *TPM1* (fast-twitch muscle) (rs3805965 and rs4075583) and *TPM2* (slow-twitch muscle) (rs2025126 and rs2145925)[179]. Tropomyosin functions with the troponin complex to regulate muscle contraction by restricting myosin from binding to actin[183]. The upstream SNP, rs3805965/*TPM1* did not incorporate a transcription factor binding site for either allele (data not shown) suggesting that this associated allele could be tagging for a causative allele, thus investigating the region that is in linkage disequilibrium with this allele needs to be assessed for other potential regulatory SNPs.

Although functional analysis of rs3805965 yielded no potential regulatory effects, rs4075583, located in intron 1 of the *TPM1* skeletal muscle isoform, showed evidence for a role in regulation of *TPM1* promoter expression (Figure 6.1 and 6.2). A previous study by Savill *et.* al (2010) investigating the cause of Metabolic syndrome, described a ~1.7 kb region containing three SNPs, rs4075583, rs4075584 and rs4075047, that influenced the expression of *TPM1*[199]. Expression varied by haplotype and cell type with those studies focusing primarily on the cytoskeletal *TPM1* isoform in HEK293 and THP-1 cells[199]. Interestingly, different isoforms of *TPM1* are produced by alternate splicing and are expressed only in specific cell types. Our studies focused on the skeletal muscle *TPM1* skeletal isoform, which requires a different promoter for expression. We first assessed whether rs4075583 as a 20mer could influence the promoter activity. Although the an allele-dependent transcription factor binding site was found with the ancestral allele, no effect on promoter activity was found. Based on Savill's finding of specific haplotypes

influencing expression and that the ~1.7 kb region contains rs4075583, and seven additional SNPs being in a conserved region, we assessed whether rs4075583 in the context of a haplotype could affect promoter activity[199]. The four common haplotypes evaluated differed in effect on promoter activity, even though the seven additional SNPs did not incorporate transcription factor binding sites, thus these SNPs may influence the binding stringency of the transcription factor associated with the ancestral allele of rs4075583. The most common haplotype that contains the alternate allele of rs4075583 eliminates a transcription factor binding site and increases promoter activity suggesting that through evolution this allele may have become important for skeletal muscle function in humans (Figure 6.1B and 6.2). Indeed, the ancestral allele (G), which was invariant in haplotypes 2-4, significantly decreased promoter activity. Interestingly, haplotype 3 incorporates the alternate allele for five of the eight SNPs resulting in the greatest decrease in promoter activity (Table 6.2 and Figure 6.2B and C). These findings suggest the importance of this region for *TPM1* gene regulation, in particular the conservation of these seven additional *TPM1* SNPs and the necessity of the evolved alternate allele of rs4075583 to obtain the required expression of *TPM1* for muscle function. However, to begin to delineate the SNPs contributing to gene regulation, individual analysis of each SNP is needed and will be the focus of future studies.

Interestingly, evaluation of two SNPs in *TPM2,* rs2025126 and rs2145925, suggests the importance of these SNPs for gene regulation at different stages of myogenesis. rs2025126, located upstream of *TPM2,* creates a DNA-protein complex with both undifferentiated and differentiated nuclear extracts (Figure 6.3B), however, promoter activity was only increased in differentiated cells (Figure 6.3D). As was found with rs2145925, the alternate allele incorporated two DNA-protein complexes that increased

activity in undifferentiated cells (Figure 6.4B and C), while decreasing in differentiated cells (Figure 6.4D). Myogenesis requires different transcription factors during different stages of development such as when myoblast fate (undifferentiated) determination is occurring and during subsequent muscle function (differentiated). Specification of when and where myogenic cells form slow- or fast-twitch muscle fibers has been an ongoing debate, with some models suggesting the cells are predisposed to a slow or fast fate[41,49-58]. Utilizing this model, it can be postulated that one of the factors binding with the alternate allele of rs245925 is involved in regulating the specification of the myoblast (undifferentiated) to become a slow-twitch muscle fiber. Once at the differentiated state when muscle functions such as contraction occur, the factors associated with rs2025126 and rs2145925 that caused increase promoter activity could be required for regulating muscle contraction and the decrease in activity with the alternate alleles could lead to dysregulation of muscle contraction. This dysregulation could lead to the foot being stuck in a contracted state as observed with clubfoot.

Muscle development and function are multi-faceted processes involving various proteins. Dysregulation during either process could cause a cascade of effects resulting in muscle anomalies such as muscle hypoplasia, which is observed with clubfoot. For example, *HOX* genes regulate *LBX1*, a transcription factor that is important in the migration of muscle precursor cells into the developing limb bud (Figure 6.7A)[214]. *HOXA9* misexpression could affect *LBX1* expression causing a decrease in cell migration leading to muscle hypoplasia as seen in clubfoot individuals (Figure 6.7B). In addition, it has been suggested that the first muscle precursor cells to enter the limb bud are intended to form the slow-twitch muscle fibers with the fast-twitch muscle fibers being secondary[215].

Misexpression of *HOXA9* could affect the muscle composition of the limb, which has been observed with some clubfoot individuals having disorganization of muscle fibers. With the disorganization of the muscle fibers in combination with misexpression of *TPM1* and *TPM2,* genes important in regulating muscle contraction, a persistent contracted state of the foot could occur as characterized with clubfoot (Figure 6.7C). With these hypothetical pathways contributing to the key characteristics of clubfoot, persistent contracted state and calf muscle hypoplasia, we asked whether there were differences in allelic combinations (genetic signature) of these four SNPs/genes in cases compared to controls. This is based on the multifactorial model which states that variation in multiple genes contribute to quantitative traits and common birth defects[93,108,110]. This approach has been successfully applied to the folate pathway genes and has been used in spina bifida to identify genetic risk signatures for different ethnicities [216]. Using a less stringent analysis (Haploview) than the one employed in the spina bifida study, we identified genetic signatures for our case and controls groups by ethnicity (Table 6.3). However, the frequency of the most common genetic signatures did not differ between cases and controls for either ethnicity (Table 6.3). While these specific genetic signatures did not appear to contribute to clubfoot, other regulatory variants within these genes cannot be excluded from playing a role.

The goal of this study was to assess the expression of potential muscle-specific regulatory variants that had previously been associated with clubfoot. We found that five SNPs, rs3801776/*HOXA9,* rs2025126/*TPM2*, rs2145925/*TPM2,* rs4075583/*TPM1* and rs437122/*TNNC2,* showed allele-dependent transcription factor binding sites but only the first three also altered promoter activity. However, rs4075583/*TPM1,* in the context of the different DNA sequence/haplotypes, had varying effects on promoter activity. Although

delineating the individual *TPM1* SNPs within the ~1.7 kb region that regulate promoter activity is needed, our results suggest that future expression analyses of potential promoter variants should be performed in the context of the *in vivo* DNA sequence also. Finally, although a genetic signature incorporating these four SNPs was not identified, this approach of analyzing variants within genes in a common pathway such as muscle development, patterning and/or function could begin to identify genetic signatures for clubfoot.

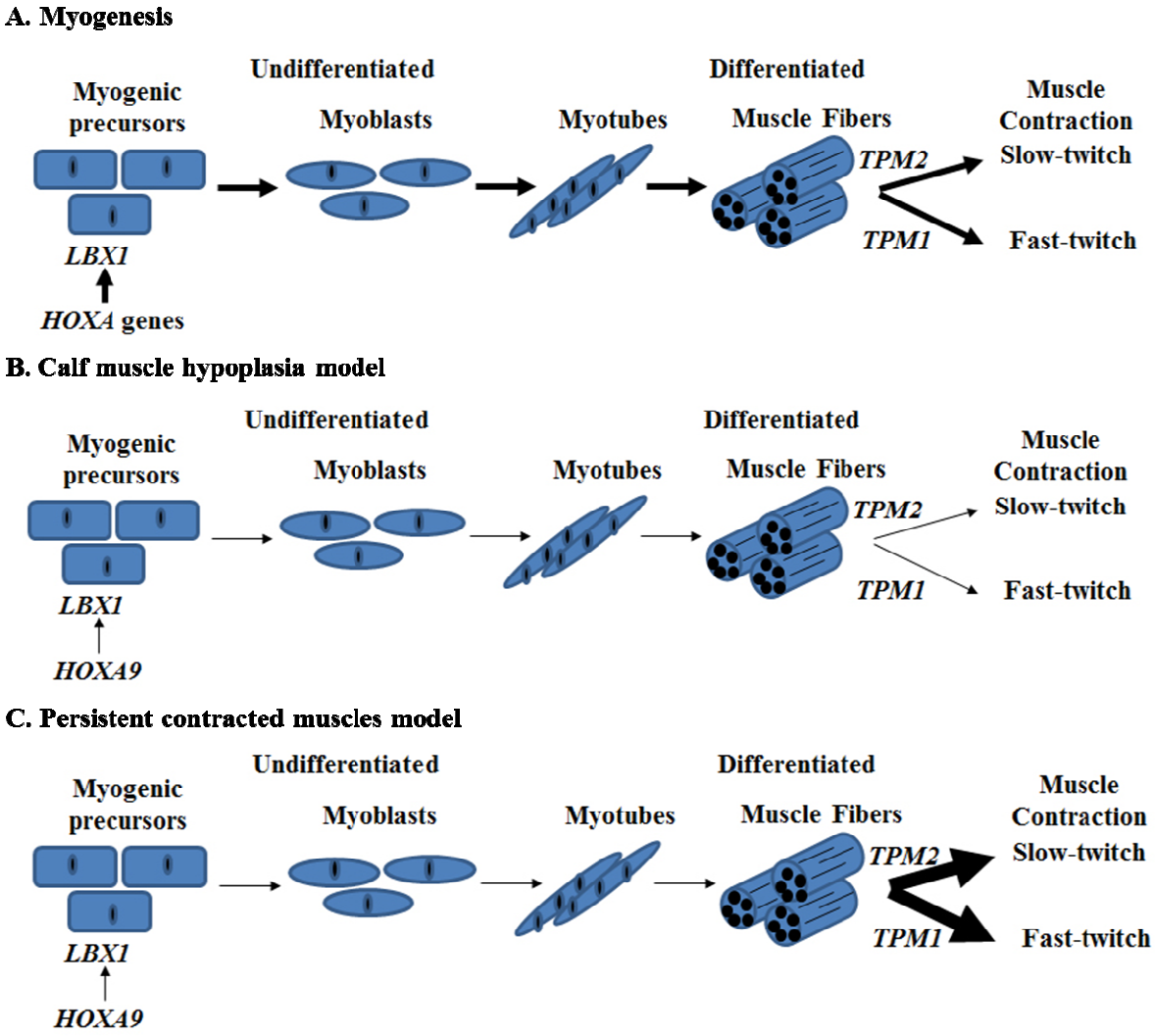

Figure 6.7 Models for the dysregulation of muscle development and function contributing to clubfoot. (A) Myogenesis is a multi-step process involving two key stages of development, undifferentiated and differentiated. During the undifferentiated state of development, myogenic precursor cells migrate to the limb bud under the control of *LBX1*, a transcription factor regulated by *HOXA* genes. Once in the limb bud, these cells transform into myoblasts. At the differentiated stage, myoblasts form myotubes that mature into slow- and fast-twitch muscle fibers that can actively contract. (B) *HOXA* misexpression causes a decrease in migration of the myogenic precursor cells into the limb bud causing calf muscle hypoplasia. (C) Decreased cell myogenic precursor cell migration into the limb bud leads to a disorganization of muscle fibers favoring a higher proportion of slow-twitch muscles combined with the misexpression of *TPM1* and *TPM2* causes the foot to be stuck in a persistent contracted state.

Chapter 7: Summary and future studies

7.1 Summary and future studies

Clubfoot is a common birth defect affecting more than 135,000 newborns world-wide each year[66-71,107]. Even with corrective treatment during the first few years of life, residual foot and leg abnormalities persist[77,78]. A genetic etiology is suggested based on familial recurrences and segregation analyses, although Mendelian inheritance does not explain these recurrences. Environmental factors have also been implicated with maternal smoking having the most consistent association. Based on these findings clubfoot has been posited to fit a multifactorial model with both genetic and environmental factors contributing to causation [69,71,101,106,110]. Identifying the genetic factors has been a challenge. The overall goal of these studies was to identify the genetic variation contributing to clubfoot. We describe the use of a candidate gene approach wherein we interrogated limb and muscle-specific genes and found associations with regulatory variants in *HOXA9, TPM1, TPM2* and *TNNC2*. We extended those studies by performing functional analyses on these SNPs to determine whether the allelic forms of these SNPs alter promoter expression. We hypothesized that variation in these SNPs can perturb gene expression; individually they may not affect foot development but in aggregate they may alter the final outcome. We also describe the use of whole genome linkage and association analyses on ten multiplex clubfoot families and our success in identifying new regions to interrogate in future studies.

Chapter 1 provides an overview of clubfoot from a clinical description to approaches currently being employed to identify genes contributing to clubfoot. Chapter 3 discusses the results of studies interrogating limb patterning genes, *HOXA* and *HOXD* gene clusters, and *IGFBP3*[121,127]. As shown in Figure 3.1, the results of these studies allowed us

to begin filling in a model wherein functional perturbation of these genes contributes to clubfoot. The work on *HOXA9* continued and is discussed in Chapter 6.

Chapter 4 presents the results of our genome-wide linkage scan performed on 10 multiplex clubfoot families. Eight regions were identified: 3q22.1-q24, 4p14, 4q32.1, 14q13.1, 14q32.12, 17q21.33-q22, 17q23.2-q24.2 and 20q13.12. Only one of these regions, 17q23.2-q24.2, has been previously implicated in clubfoot[118]. *TNNC2* is in 20q13.12 and is a candidate gene because of its role in muscle contraction. Sequencing of the coding region did not identify any mutations or rare variants likely to cause disease. This gene still remains of interest because the promoter region has not been completely interrogated and needs to be assessed further. These new regions provide areas to data mine for clubfoot genes. Further validation of these regions in our complete clubfoot dataset needs to be performed in order to narrow down these regions.

Chapter 5 describes the interrogation of muscle contraction genes that cause Distal Arthrogryposis syndromes. Many of these conditions have clubfoot as an associated finding with multiple congenital joint contractures. Strong associations were found for SNPs in *TPM1* and *TPM2* in single SNP, haplotype and gene interaction analyses and were the first to identify genotypic effect with *TNNC2* SNPs. Interestingly, these associated SNPs were located in potential regulatory regions suggesting gene regulation may play an important role in the etiology of clubfoot. Based on these findings, future studies could focus on the additional genes known to play a role in muscle contraction through interrogation in our large clubfoot dataset.

Chapter 6 presents the functional assessment of seven associated SNPs identified in studies described in Chapter 3 and 5. All of these SNPs have a potential regulatory function

because they are located either in the promoter or enhancer/suppressor regions of *HOXA9, TPM1, TPM2* and *TNNC2*. Although rs3801776/*HOXA9,* rs2025126/*TPM2* and rs2145925/*TPM2* altered promoter activity, the specific binding protein needs to be identified which might lead to the identification of a common pathway and/or mechanism. Interestingly, one SNP, rs4075583/*TPM1,* altered promoter expression but only when in the context of the surrounding DNA environment containing seven other variants. Therefore, in order to obtain a better functional assessment, all potential regulatory SNPs should be evaluated individually and should not be excluded without a haplotype functional analysis.

The results of these studies present new areas for future clubfoot gene hunting and a paradigm for assessing regulatory variants. Our results and a recent study suggest that gene regulation in limb development genes play a role in clubfoot[180]. It also begins to define gene pathways, may allow us to identify at-risk genetic signatures and ultimately should be useful in population-based genetic screening for at-risk genotypes that predispose to clubfoot. This would translate to improved genetic counseling for clubfoot families.

References

1. Bohm M. Embryologic origin of clubfoot. Journal of Bone & Joint Surgery 1929;11:229-59.
2. Irani RN, Sherman MS. The pathological anatomy of idiopathic clubfoot. Clin Orthop Relat Res 1972;84:14-20.
3. Isaacs H, Handelsman JE, Badenhorst M, Pickering A. The muscles in club foot--a histological histochemical and electron microscopic study. J Bone Joint Surg Br 1977;59-B:465-72.
4. Zuzarte-Luis V, Hurle JM. Programmed cell death in the developing limb. Int J Dev Biol 2002;46:871-6.
5. Capdevila J, Izpisua Belmonte JC. Patterning mechanisms controlling vertebrate limb development. Annu Rev Cell Dev Biol 2001;17:87-132.
6. Duboc V, Logan MP. Building limb morphology through integration of signalling modules. Curr Opin Genet Dev 2009;19:497-503.
7. Logan M. Finger or toe: the molecular basis of limb identity. Development 2003;130:6401-10.
8. Yang Y. Growth and patterning in the limb: signaling gradients make the decision. Sci Signal 2009;2:pe3.
9. Duboc V, Logan MP. Regulation of limb bud initiation and limb-type morphology. Dev Dyn 2011;240:1017-27.
10. Abarca-Buis RF CR, Rios-Flores AJ, Diaz-Hernandez M, Lozano-Flores C, Bustamante M and Chimal-Monroy J. Limb development. In: Chimal-Monroy J, ed. Topics in Animal and Plant Development: From Cell Differentiation to Morphogenesis: Transworld Research Network; 2011:113-34.

11. Tickle C. Patterning systems--from one end of the limb to the other. Dev Cell 2003;4:449-58.

12. Towers M, Tickle C. Growing models of vertebrate limb development. Development 2009;136:179-90.

13. Mariani FV, Ahn CP, Martin GR. Genetic evidence that FGFs have an instructive role in limb proximal-distal patterning. Nature 2008;453:401-5.

14. Summerbell D, Lewis JH, Wolpert L. Positional information in chick limb morphogenesis. Nature 1973;244:492-6.

15. Mercader N, Leonardo E, Piedra ME, Martinez AC, Ros MA, Torres M. Opposing RA and FGF signals control proximodistal vertebrate limb development through regulation of Meis genes. Development 2000;127:3961-70.

16. Tabin C, Wolpert L. Rethinking the proximodistal axis of the vertebrate limb in the molecular era. Genes Dev 2007;21:1433-42.

17. Mercader N. LE, Azpiazu, N., Serrano, A., Morata, G., Martinez, C., Torres, M. Conserved regulation of proximodistal limb axis development by Meis1/Hth. Nature 1999;402:425-9.

18. Maden M. Intercalary regeneration in the amphibian limb and the rule of distal transformation. J Embryol Exp Morphol 1980;56:201-9.

19. Stocum DL. The urodele limb regeneration blastema. Determination and organization of the morphogenetic field. Differentiation 1984;27:13-28.

20. Tarchini B, Duboule D. Control of Hoxd genes' collinearity during early limb development. Dev Cell 2006;10:93-103.

21. Tarchini B, Duboule D, Kmita M. Regulatory constraints in the evolution of the tetrapod limb anterior-posterior polarity. Nature 2006;443:985-8.

22. McGlinn E, Tabin CJ. Mechanistic insight into how Shh patterns the vertebrate limb. Curr Opin Genet Dev 2006;16:426-32.

23. Tabin CJ, McMahon AP. Developmental biology. Grasping limb patterning. Science 2008;321:350-2.

24. Hill TP, Taketo MM, Birchmeier W, Hartmann C. Multiple roles of mesenchymal beta-catenin during murine limb patterning. Development 2006;133:1219-29.

25. Chen H, Lun Y, Ovchinnikov D, Kokubo H, Oberg K, Pepicelli C, Can L, Lee B, Johnson R. Limb and kidney defects in Lmx1b mutant mice suggest an involvement of LMX1B in human nail patella syndrome. Nat Genet 1998;19:51-5.

26. Riddle RD, Ensini M, Nelson C, Tsuchida T, Jessell TM, Tabin C. Induction of the LIM homeobox gene Lmx1 by WNT7a establishes dorsoventral pattern in the vertebrate limb. Cell 1995;83:631-40.

27. Loomis CA, Harris E, Michaud J, Wurst W, Hanks M, Joyner AL. The mouse Engrailed-1 gene and ventral limb patterning. Nature 1996;382:360-3.

28. Alles AJ, Sulik KK. Retinoic-acid-induced limb-reduction defects: perturbation of zones of programmed cell death as a pathogenetic mechanism. Teratology 1989;40:163-71.

29. Saunders JW, Jr. Death in embryonic systems. Science 1966;154:604-12.

30. Hurle JM, Ros MA, Climent V, Garcia-Martinez V. Morphology and significance of programmed cell death in the developing limb bud of the vertebrate embryo. Microsc Res Tech 1996;34:236-46.

31. Naruse I, Kameyama Y. Prevention of polydactyly manifestation in Polydactyly Nagoya (Pdn) mice by administration of cytosine arabinoside during pregnancy. Teratology 1986;34:283-9.

32. Zou H, Choe KM, Lu Y, Massague J, Niswander L. BMP signaling and vertebrate limb development. Cold Spring Harb Symp Quant Biol 1997;62:269-72.

33. Macias D, Ganan Y, Sampath TK, Piedra ME, Ros MA, Hurle JM. Role of BMP-2 and OP-1 (BMP-7) in programmed cell death and skeletogenesis during chick limb development. Development 1997;124:1109-17.

34. Zuzarte-Luis V, Montero JA, Kawakami Y, Izpisua-Belmonte JC, Hurle JM. Lysosomal cathepsins in embryonic programmed cell death. Dev Biol 2007;301:205-17.

35. Zuzarte-Luis V, Montero JA, Torre-Perez N, Garcia-Porrero JA, Hurle JM. Cathepsin D gene expression outlines the areas of physiological cell death during embryonic development. Dev Dyn 2007;236:880-5.

36. Rodriguez-Leon J, Merino R, Macias D, Ganan Y, Santesteban E, Hurle JM. Retinoic acid regulates programmed cell death through BMP signalling. Nat Cell Biol 1999;1:125-6.

37. Heydanus R, Ledward RS. Failed mifepristone termination of pregnancy and talipes equinovares: Coincidence? J Obstet Gynaecol 1995;15:211.

38. Czeizel AE, Puho E, Sorensen HT, Olsen J. Possible association between different congenital abnormalities and use of different sulfonamides during pregnancy. Congenit Anom (Kyoto) 2004;44:79-86.

39. Kurokawa M, Kornbluth S. Caspases and kinases in a death grip. Cell 2009;138:838-54.

40. Ulrich M, Kristoffersen K, Rolschau J, Grinsted P, Schaumburg E, Foged N. The influence of folic acid supplement on the outcome of pregnancies in the county of Funen in Denmark. Part III. Congenital anomalies. An observational study. Eur J Obstet Gynecol Reprod Biol 1999;87:115-8; discussion 03-4.

41. Francis-West PH, Antoni L, Anakwe K. Regulation of myogenic differentiation in the developing limb bud. J Anat 2003;202:69-81.

42. Gross MK, Moran-Rivard L, Velasquez T, Nakatsu MN, Jagla K, Goulding M. Lbx1 is required for muscle precursor migration along a lateral pathway into the limb. Development 2000;127:413-24.

43. Daston G, Lamar E, Olivier M, Goulding M. Pax-3 is necessary for migration but not differentiation of limb muscle precursors in the mouse. Development 1996;122:1017-27.

44. Brand-Saberi B, Muller TS, Wilting J, Christ B, Birchmeier C. Scatter factor/hepatocyte growth factor (SF/HGF) induces emigration of myogenic cells at interlimb level in vivo. Dev Biol 1996;179:303-8.

45. Heymann S, Koudrova M, Arnold H, Koster M, Braun T. Regulation and function of SF/HGF during migration of limb muscle precursor cells in chicken. Dev Biol 1996;180:566-78.

46. Rodriguez-Guzman M, Montero JA, Santesteban E, Ganan Y, Macias D, Hurle JM. Tendon-muscle crosstalk controls muscle bellies morphogenesis, which is mediated by cell death and retinoic acid signaling. Dev Biol 2007;302:267-80.

47. Tajbakhsh S, Cossu G. Establishing myogenic identity during somitogenesis. Curr Opin Genet Dev 1997;7:634-41.

48. Tajbakhsh S, Rocancourt D, Cossu G, Buckingham M. Redefining the genetic hierarchies controlling skeletal myogenesis: Pax-3 and Myf-5 act upstream of MyoD. Cell 1997;89:127-38.

49. Bismuth K, Relaix F. Genetic regulation of skeletal muscle development. Exp Cell Res 2010;316:3081-6.

50. DiMario JX, Fernyak SE, Stockdale FE. Myoblasts transferred to the limbs of embryos are committed to specific fibre fates. Nature 1993;362:165-7.

51. Miller JB, Stockdale FE. Developmental origins of skeletal muscle fibers: clonal analysis of myogenic cell lineages based on expression of fast and slow myosin heavy chains. Proc Natl Acad Sci U S A 1986;83:3860-4.

52. Miller JB, Stockdale FE. Developmental regulation of the multiple myogenic cell lineages of the avian embryo. J Cell Biol 1986;103:2197-208.

53. Stockdale FE. Myoblast diversity and the formation of the early limb musculature. Ann N Y Acad Sci 1990;599:111-8.

54. Van Swearingen J, Lance-Jones C. Slow and fast muscle fibers are preferentially derived from myoblasts migrating into the chick limb bud at different developmental times. Dev Biol 1995;170:321-37.

55. Kardon G, Campbell JK, Tabin CJ. Local extrinsic signals determine muscle and endothelial cell fate and patterning in the vertebrate limb. Dev Cell 2002;3:533-45.

56. Hughes SM, Blau HM. Muscle fiber pattern is independent of cell lineage in postnatal rodent development. Cell 1992;68:659-71.

57. Robson LG, Hughes SM. Local signals in the chick limb bud can override myoblast lineage commitment: induction of slow myosin heavy chain in fast myoblasts. Mech Dev 1999;85:59-71.

58. DiMario JX, Stockdale FE. Both myoblast lineage and innervation determine fiber type and are required for expression of the slow myosin heavy chain 2 gene. Dev Biol 1997;188:167-80.

59. Roy S, Wolff C, Ingham PW. The u-boot mutation identifies a Hedgehog-regulated myogenic switch for fiber-type diversification in the zebrafish embryo. Genes Dev 2001;15:1563-76.

60. Loren GJ, Karpinski NC, Mubarak SJ. Clinical implications of clubfoot histopathology. J Pediatr Orthop 1998;18:765-9.

61. Dimeglio A, Bensahel H, Souchet P, Mazeau P, Bonnet F. Classification of clubfoot. J Pediatr Orthop B 1995;4:129-36.

62. Pirani S. A reliable and valid method of assessing the amount of deformity in the congenital clubfoot. In: Pediatric Orthopaedic Society of North America. St Louis, MO; 2004.

63. Gurnett CA, Boehm S, Connolly A, Reimschisel T, Dobbs MB. Impact of congenital talipes equinovarus etiology on treatment outcomes. Dev Med Child Neurol 2008;50:498-502.

64. Gurnett MBDaCA. Genetics of clubfoot. J Pediatr Orthop B 2011;21:7-9.

65. Brewer C, Holloway S, Zawalnyski P, Schinzel A, FitzPatrick D. A chromosomal deletion map of human malformations. Am J Hum Genet 1998;63:1153-9.

66. Carey M, Bower C, Mylvaganam A, Rouse I. Talipes equinovarus in Western Australia. Paediatr Perinat Epidemiol 2003;17:187-94.

67. Ching GH, Chung CS, Nemechek RW. Genetic and epidemiological studies of clubfoot in Hawaii: ascertainment and incidence. Am J Hum Genet 1969;21:566-80.

68. Chung CS, Nemechek RW, Larsen IJ, Ching GH. Genetic and epidemiological studies of clubfoot in Hawaii. General and medical considerations. Hum Hered 1969;19:321-42.

69. Lochmiller C, Johnston D, Scott A, Risman M, Hecht JT. Genetic epidemiology study of idiopathic talipes equinovarus. Am J Med Genet 1998;79:90-6.

70. Beals RK. Club foot in the Maori: a genetic study of 50 kindreds. N Z Med J 1978;88:144-6.

71. Moorthi RN, Hashmi SS, Langois P, Canfield M, Waller DK, Hecht JT. Idiopathic talipes equinovarus (ITEV) (clubfeet) in Texas. Am J Med Genet A 2005;132:376-80.

72. Cardy AH, Sharp L, Torrance N, Hennekam RC, Miedzybrodzka Z. Is there evidence for aetiologically distinct subgroups of idiopathic congenital talipes equinovarus? A case-only study and pedigree analysis. PLoS One 2011;6:e17895.

73. Dietz F. The genetics of idiopathic clubfoot. Clin Orthop Relat Res 2002:39-48.

74. Turco VJ. Resistant congenital club foot--one-stage posteromedial release with internal fixation. A follow-up report of a fifteen-year experience. J Bone Joint Surg Am 1979;61:805-14.

75. Bensahel H, Csukonyi Z, Desgrippes Y, Chaumien JP. Surgery in residual clubfoot: one-stage medioposterior release "a la carte". J Pediatr Orthop 1987;7:145-8.

76. Hulme A. The management of congenital talipes equinovarus. Early Hum Dev 2005;81:797-802.

77. Laaveg SJ, Ponseti IV. Long-term results of treatment of congenital club foot. J Bone Joint Surg Am 1980;62:23-31.

78. Ponseti IV. Treatment of congenital club foot. J Bone Joint Surg Am 1992;74:448-54.

79. Bensahel H, Catterall A, Dimeglio A. Practical applications in idiopathic clubfoot: a retrospective multicentric study in EPOS. J Pediatr Orthop 1990;10:186-8.

80. Hippocrates. Loeb Classical. London & New York: Heinemann & Putnam; 1927.

81. Yang HY, Chung CS, Nemechek RW. A genetic analysis of clubfoot in Hawaii. Genet Epidemiol 1987;4:299-306.

82. Miedzybrodzka Z. Congenital talipes equinovarus (clubfoot): a disorder of the foot but not the hand. J Anat 2003;202:37-42.

83. Hoffa A. Lehrbuch der orthopa□dischen Chirurgie: Stuttgart: Ferdinand Enke; 1902.

84. Farrell SA, Summers AM, Dallaire L, Singer J, Johnson JA, Wilson RD. Club foot, an adverse outcome of early amniocentesis: disruption or deformation? CEMAT. Canadian Early and Mid-Trimester Amniocentesis Trial. J Med Genet 1999;36:843-6.

85. Barker SL, Macnicol MF. Seasonal distribution of idiopathic congenital talipes equinovarus in Scotland. J Pediatr Orthop B 2002;11:129-33.

86. Mufti M. Prime factors in the etiology of congenital dislocation of the hip and talipes equinovarus in Riyadh. Ann Saudi Med 1988;8.

87. Pagnotta G, Maffulli N, Aureli S, Maggi E, Mariani M, Yip KM. Antenatal sonographic diagnosis of clubfoot: a six-year experience. J Foot Ankle Surg 1996;35:67-71.

88. Idelberger K. Die Ergebnisse der Zwillingsforschung beim angeborenen Klumpfuss. Verh Deutsch Orthopdd Ges 1939;33:272-6.

89. Wynne-Davies R. Family Studies and the Cause of Congenital Club Foot. Talipes Equinovarus, Talipes Calcaneo-Valgus and Metatarsus Varus. J Bone Joint Surg Br 1964;46:445-63.

90. Fritsch H, Eggers R. Ossification of the calcaneus in the normal fetal foot and in clubfoot. J Pediatr Orthop 1999;19:22-6.

91. Atlas S, Menacho LC, Ures S. Some new aspects in the pathology of clubfoot. Clin Orthop Relat Res 1980:224-8.

92. Ippolito E, Ponseti IV. Congenital club foot in the human fetus. A histological study. J Bone Joint Surg Am 1980;62:8-22.

93. Honein MA, Paulozzi LJ, Moore CA. Family history, maternal smoking, and clubfoot: an indication of a gene-environment interaction. Am J Epidemiol 2000;152:658-65.

94. Nadeem RD, Brown JK, Lawson G, Macnicol MF. Somatosensory evoked potentials as a means of assessing neurological abnormality in congenital talipes equinovarus. Dev Med Child Neurol 2000;42:525-30.

95. Bareither D. Prenatal development of the foot and ankle. J Am Podiatr Med Assoc 1995;85:753-64.

96. Kawashima T, Uhthoff HK. Development of the foot in prenatal life in relation to idiopathic club foot. J Pediatr Orthop 1990;10:232-7.

97. Loder RT, Drvaric DM, Carney B, Hamby Z, Barker S, Chesney D, Maffulli N. Lack of seasonal variation in idiopathic talipes equinovarus. J Bone Joint Surg Am 2006;88:496-502.

98. Pryor GA, Villar RN, Ronen A, Scott PM. Seasonal variation in the incidence of congenital talipes equinovarus. J Bone Joint Surg Br 1991;73:632-4.

99. Wallander H, Hovelius L, Michaelsson K. Incidence of congenital clubfoot in Sweden. Acta Orthop 2006;77:847-52.

100. Pavone V, Bianca S, Grosso G, Pavone P, Mistretta A, Longo MR, Marino S, Sessa G. Congenital talipes equinovarus: an epidemiological study in Sicily. Acta Orthop 2012.

101. Skelly AC, Holt VL, Mosca VS, Alderman BW. Talipes equinovarus and maternal smoking: a population-based case-control study in Washington state. Teratology 2002;66:91-100.

102. Dickinson KC, Meyer RE, Kotch J. Maternal smoking and the risk for clubfoot in infants. Birth Defects Res A Clin Mol Teratol 2008;82:86-91.

103. Barker S, Chesney D, Miedzybrodzka Z, Maffulli N. Genetics and epidemiology of idiopathic congenital talipes equinovarus. J Pediatr Orthop 2003;23:265-72.

104. Yamamoto H. A clinical, genetic and epidemiologic study of congenital club foot. Jinrui Idengaku Zasshi 1979;24:37-44.

105. Strach E. Club-foot through the centuries. Prog Pediatr Surg 1986;20:215-37.

106. Wynne-Davies R. Family studies and aetiology of club foot. J Med Genet 1965;2:227-32.

107. de Andrade M, Barnholtz JS, Amos CI, Lochmiller C, Scoot A, risman M, Hecht JT. Segregation analysis of idiopathic talipes equinovarus in a Texan population. Am J Med Genet 1998;79:97-102.

108. Kruse LM, Dobbs MB, Gurnett CA. Polygenic threshold model with sex dimorphism in clubfoot inheritance: the Carter effect. J Bone Joint Surg Am 2008;90:2688-94.

109. Wang JH, Palmer RM, Chung CS. The role of major gene in clubfoot. Am J Hum Genet 1988;42:772-6.

110. Wynne-Davies R. Genetic and environmental factors in the etiology of talipes equinovarus. Clin Orthop Relat Res 1972;84:9-13.

111. Collins A. Approaches to the identification of susceptibility genes. Parasite Immunol 2009;31:225-33.

112. Gurnett CA, Alaee F, Kruse LM, Desruisseau DM, Hecht JT, Wise CA, Bowcock AM, Dobbs MB. Asymmetric lower-limb malformations in individuals with homeobox PITX1 gene mutation. Am J Hum Genet 2008;83:616-22.

113. Alvarado DM, McCall K, Aferol H, Silva MJ, Garbow JR, Spees WM, Patel T, Siegel M, Dobbs MB, Gurnett CA. Pitx1 haploinsufficiency causes clubfoot in humans and a clubfoot-like phenotype in mice. Hum Mol Genet 2011;20:3943-52.

114. Alvarado DM, Aferol H, McCall K, Huang JB, Techy M, Buchan J, Cady J, Gonzales PR, Dobbs MB, Gurnett CA. Familial isolated clubfoot is associated with recurrent chromosome 17q23.1q23.2 microduplications containing TBX4. Am J Hum Genet 2010;87:154-60.

115. Rodriguez-Esteban C, Tsukui T, Yonei S, Magallon J, Tamura K, Izpisua Belmonte JC. The T-box genes Tbx4 and Tbx5 regulate limb outgrowth and identity. Nature 1999;398:814-8.

116. Hasson P, DeLaurier A, Bennett M, Grigorieva E, Naiche LA, Papaioannou VE, Mohun TJ, Logan MP. Tbx4 and tbx5 acting in connective tissue are required for limb muscle and tendon patterning. Dev Cell 2010;18:148-56.

117. Logan M, Tabin CJ. Role of Pitx1 upstream of Tbx4 in specification of hindlimb identity. Science 1999;283:1736-9.

118. Lu W, Bacino CA, Richards BS, Alvarez C, VanderMeer JE, Vella M, Ahituv N, Sikka N, Dietz FR, Blanton SH, Hecht JT. Studies of TBX4 and chromosome 17q23.1q23.2: An uncommon cause of nonsyndromic clubfoot. Am J Med Genet A 2012;158A:1620-7.

119. Brewer C, Holloway S, Zawalnyski P, Schinzel A, FitzPatrick D. A chromosomal duplication map of malformations: regions of suspected haplo- and triplolethality--and tolerance of segmental aneuploidy--in humans. Am J Hum Genet 1999;64:1702-8.

120. Heck AL, Bray MS, Scott A, Blanton SH, Hecht JT. Variation in CASP10 gene is associated with idiopathic talipes equinovarus. J Pediatr Orthop 2005;25:598-602.

121. Ester AR, Tyerman G, Wise CA, Blanton SH, Hecht JT. Apoptotic gene analysis in idiopathic talipes equinovarus (clubfoot). Clin Orthop Relat Res 2007;462:32-7.

122. McGinnis W, Krumlauf R. Homeobox genes and axial patterning. Cell 1992;68:283-302.

123. Mark M, Rijli FM, Chambon P. Homeobox genes in embryogenesis and pathogenesis. Pediatr Res 1997;42:421-9.

124. Houghton L, Rosenthal N. Regulation of a muscle-specific transgene by persistent expression of Hox genes in postnatal murine limb muscle. Dev Dyn 1999;216:385-97.

125. Dobbs MB, Gurnett CA, Pierce B, Exner GU, Robarge J, Morcuende JA, Cole WG, Templeton PA, Foster B, Bowcock AM. HOXD10 M319K mutation in a family with isolated congenital vertical talus. J Orthop Res 2006;24:448-53.

126. Shrimpton AE, Levinsohn EM, Yozawitz JM, Packard DS Jr, Cady RB, Middleton FA, Persico AM, Hootnick DR. A HOX gene mutation in a family with isolated congenital vertical talus and Charcot-Marie-Tooth disease. Am J Hum Genet 2004;75:92-6.

127. Ester AR, Weymouth KS, Burt A, Wise CA, Scott A, Gurnett CA, Dobbs MB, Blanton SH, Hecht JT. Altered transmission of HOX and apoptotic SNPs identify a potential common pathway for clubfoot. Am J Med Genet A 2009;149A:2745-52.

128. Johnston MC, Bronsky PT. Animal models for human craniofacial malformations. J Craniofac Genet Dev Biol 1991;11:277-91.

129. Nagai T, Ibi D, Yamada K. Animal model for schizophrenia that reflects gene-environment interactions. Biol Pharm Bull 2011;34:1364-8.

130. Li ZG, Ji H, Fu WN, Zhao YY, Jin CL, Ji SJ, Sun KL. [Proteomic analysis of the ankle joint bone, ankle joint tissue and spinal cord of clubfoot-like deformity in rat fetuses]. Zhonghua Yi Xue Yi Chuan Xue Za Zhi 2007;24:52-8.

131. Guo L, Zhao YY, Ji SJ, Sun KL, Dong LY. [Effect of the regulation of IGFs system components in retinoic acid-induced congenital clubfoot]. Yi Chuan Xue Bao 2004;31:766-70.

132. Delgado-Baeza E, Santos-Alvarez I, Martos-Rodriguez A. Retinoic acid-induced clubfoot-like deformity: pathoanatomy in rat fetuses. J Pediatr Orthop B 1999;8:12-8.

133. Kaufman MH, Chang HH. Studies of the mechanism of amniotic sac puncture-induced limb abnormalities in mice. Int J Dev Biol 2000;44:161-75.

134. Ashby PR, Pincon-Raymond M, Harris AJ. Regulation of myogenesis in paralyzed muscles in the mouse mutants peroneal muscular atrophy and muscular dysgenesis. Dev Biol 1993;156:529-36.

135. Nonaka I, Kikuchi A, Suzuki T, Esaki K. Hereditary peroneal muscular atrophy in the mouse: an experimental model for congenital contractures (arthrogryposis). Exp Neurol 1986;91:571-9.

136. Duce S, Madrigal L, Schmidt K, Cunningham C, Liu G, Barker S, Tennant G, Tickle C, Chudek S, Miedzybrodzka Z. Micro-magnetic resonance imaging and embryological analysis of wild-type and pma mutant mice with clubfoot. J Anat 2010;216:108-20.

137. Hootnick DR, Levinsohn EM, Crider RJ, Packard DS, Jr. Congenital arterial malformations associated with clubfoot. A report of two cases. Clin Orthop Relat Res 1982:160-3.

138. Kruse L, Gurnett CA, Hootnick D, Dobbs MB. Magnetic resonance angiography in clubfoot and vertical talus: a feasibility study. Clin Orthop Relat Res 2009;467:1250-5.

139. Bamshad M, Jorde LB, Carey JC. A revised and extended classification of the distal arthrogryposes. Am J Med Genet 1996;65:277-81.

140. Gurnett CA, Desruisseau DM, McCall K, Choi R, Meyer ZI, Talerico M, Miller SE, Ju JS, Pestronk A, Connolly AM, Druley TE, Weihl CC, Dobbs MB. Myosin binding protein C1: a novel gene for autosomal dominant distal arthrogryposis type 1. Hum Mol Genet 2010;19:1165-73.

141. Bamshad M, Bohnsack JF, Jorde LB, Carey JC. Distal arthrogryposis type 1: clinical analysis of a large kindred. Am J Med Genet 1996;65:282-5.

142. Stevenson DA, Carey JC, Palumbos J, Rutherford A, Dolcourt J, Bamshad MJ. Clinical characteristics and natural history of Freeman-Sheldon syndrome. Pediatrics 2006;117:754-62.

143. Stevenson DA, Swoboda KJ, Sanders RK, Bamshad M. A new distal arthrogryposis syndrome characterized by plantar flexion contractures. Am J Med Genet A 2006;140:2797-801.

144. Sung SS, Brassington AM, Grannatt K, Rutherford A, Whitby FG, Krakowiak PA, Jorde LB, Carey JC, Bamshad M. Mutations in genes encoding fast-twitch contractile proteins cause distal arthrogryposis syndromes. Am J Hum Genet 2003;72:681-90.

145. Sung SS, Brassington AM, Krakowiak PA, Carey JC, Jorde LB, Bamshad M. Mutations in TNNT3 cause multiple congenital contractures: a second locus for distal arthrogryposis type 2B. Am J Hum Genet 2003;73:212-4.

146. Toydemir RM, Bamshad MJ. Sheldon-Hall syndrome. Orphanet J Rare Dis 2009;4:11.

147. Toydemir RM, Rutherford A, Whitby FG, Jorde LB, Carey JC, Bamshad MJ. Mutations in embryonic myosin heavy chain (MYH3) cause Freeman-Sheldon syndrome and Sheldon-Hall syndrome. Nat Genet 2006;38:561-5.

148. Veugelers M, Bressan M, McDermott DA, Weremowicz S, Morton CC, Mabry CC, Lefaivre JF, Zunamon A, Destree A, Chaudron JM, Basson CT. Mutation of perinatal myosin heavy chain associated with a Carney complex variant. N Engl J Med 2004;351:460-9.

149. Gurnett CA, Alaee F, Desruisseau D, Boehm S, Dobbs MB. Skeletal muscle contractile gene (TNNT3, MYH3, TPM2) mutations not found in vertical talus or clubfoot. Clin Orthop Relat Res 2009;467:1195-200.

150. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263-5.

151. O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet 1998;63:259-66.

152. Abecasis GR, Cookson WO. GOLD--graphical overview of linkage disequilibrium. Bioinformatics 2000;16:182-3.

153. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 2002;30:97-101.

154. Rebbeck TR, Dietz FR, Murray JC, Buetow KH. A single-gene explanation for the probability of having idiopathic talipes equinovarus. Am J Hum Genet 1993;53:1051-63.

155. Martin ER, Monks SA, Warren LL, Kaplan NL. A test for linkage and association in general pedigrees: the pedigree disequilibrium test. Am J Hum Genet 2000;67:146-54.

156. Martin ER, Bass MP, Gilbert JR, Pericak-Vance MA, Hauser ER. Genotype-based association test for general pedigrees: the genotype-PDT. Genet Epidemiol 2003;25:203-13.

157. Chung RH, Hauser ER, Martin ER. The APL test: extension to general nuclear families and haplotypes and examination of its robustness. Hum Hered 2006;61:189-99.

158. Hancock DB, Martin ER, Li YJ, Scott WK. Methods for interaction analyses using family-based case-control data: conditional logistic regression versus generalized estimating equations. Genet Epidemiol 2007;31:883-93.

159. Grabe N. AliBaba2: context specific identification of transcription factor binding sites. In Silico Biol 2002;2:S1-15.

160. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE, Wingendeer E. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res 2006;34:D108-10.

161. Schug J. Using TESS to predict transcription factor binding sites in DNA sequence. Curr Protoc Bioinformatics 2008;Chapter 2:Unit 2 6.

162. Zakany J, Duboule D. The role of Hox genes during vertebrate limb development. Curr Opin Genet Dev 2007;17:359-66.

163. Carpenter EM, Goddard JM, Davis AP, Nguyen TP, Capecchi MR. Targeted disruption of Hoxd-10 affects mouse hindlimb development. Development 1997;124:4505-14.

164. Kmita M, Tarchini B, Zakany J, Logan M, Tabin CJ, Duboule D. Early developmental arrest of mammalian limbs lacking HoxA/HoxD gene function. Nature 2005;435:1113-6.

165. Akarsu AN, Stoilov I, Yilmaz E, Sayli BS, Sarfarazi M. Genomic structure of HOXD13 gene: a nine polyalanine duplication causes synpolydactyly in two unrelated families. Hum Mol Genet 1996;5:945-52.

166. Fromental-Ramain C, Warot X, Messadecq N, LeMeur M, Dolle P, Chambon P. Hoxa-13 and Hoxd-13 play a crucial role in the patterning of the limb autopod. Development 1996;122:2997-3011.

167. Baxter RC. The insulin-like growth factors and their binding proteins. Comp Biochem Physiol B 1988;91:229-35.

168. Baxter RC, Martin JL. Binding proteins for the insulin-like growth factors: structure, regulation and function. Prog Growth Factor Res 1989;1:49-68.

169. Butt AJ, Firth SM, Baxter RC. The IGF axis and programmed cell death. Immunol Cell Biol 1999;77:256-62.

170. Zdanowicz MM, Teichberg S. Effects of insulin-like growth factor-1/binding protein-3 complex on muscle atrophy in rats. Exp Biol Med (Maywood) 2003;228:891-7.

171. Chapman C, Stott NS, Port RV, Nicol RO. Genetics of club foot in Maori and Pacific people. J Med Genet 2000;37:680-3.

172. Engell V, Damborg F, Andersen M, Kyvik KO, Thomsen K. Club foot: a twin study. J Bone Joint Surg Br 2006;88:374-6.

173. Carter CO, Evans KA. Inheritance of congenital pyloric stenosis. J Med Genet 1969;6:233-54.

174. Happle R, Schnyder UW. Evidence for the Carter effect in atopy. Int Arch Allergy Appl Immunol 1982;68:90-2.

175. Happle R, Traupe H. [Polygenic inheritance of familial malignant melanoma]. Hautarzt 1982;33:106-11.

176. Kantarci OH, Barcellos LF, Atkinson EJ, Ramsay PP, Lincoln R, Achenbach SJ, De Andrade M, Hauser SL, Weinshenker BG. Men transmit MS more often to their children vs women: the Carter effect. Neurology 2006;67:305-10.

177. Nussbaum RL, McInnes, R.R. & Willard, H.F. Thompson & Thompson Genetics in Medicine (7th Ed). Philadelphia, PA: Saunders Elsevier; 2007.

178. Dietz FR, Cole WG, Tosi LL, Carroll NC, Werner RD, Comstock D, Murray JC. A search for the gene(s) predisposing to idiopathic clubfoot. Clin Genet 2005;67:361-2.

179. Weymouth KS, Blanton SH, Bamshad MJ, Beck AE, Alvarez C, Richards S, Gurnett CA, Dobbs MB, Barnes D, Mitchell LE, Hecht JT. Variants in genes that encode muscle contractile proteins influence risk for isolated clubfoot. Am J Med Genet A 2011;155A:2170-9.

180. Alvarado DM, Buchan JG, Frick SL, Herzenberg JE, Dobbs MB, Gurnett CA. Copy number analysis of 413 isolated talipes equinovarus patients suggests role for transcriptional regulators of early limb development. Eur J Hum Genet 2012.

181. Dobbs MB, Gurnett CA. Genetics of clubfoot. J Pediatr Orthop B 2012;21:7-9.

182. Hecht JT, Ester A, Scott A, Wise CA, Iovannisci DM, Lammer EJ, Langlois PH, Blanton SH. NAT2 variation and idiopathic talipes equinovarus (clubfoot). Am J Med Genet A 2007;143A:2285-91.

183. Gordon AM, Homsher E, Regnier M. Regulation of contraction in striated muscle. Physiol Rev 2000;80:853-924.

184. Nastasi T, Bongiovanni A, Campos Y, Mann L, Toy JN, Bostrom J, Rottier R, Hahn C, Conaway JW, Harris AJ, D'Azzo A. Ozz-E3, a muscle-specific ubiquitin ligase, regulates beta-catenin degradation during myogenesis. Dev Cell 2004;6:269-82.

185. Fukuhara K, Schollmeier G, Uhthoff HK. The pathogenesis of club foot. A histomorphometric and immunohistochemical study of fetuses. J Bone Joint Surg Br 1994;76:450-7.

186. Handelsman JE, Isaacs H. Proceedings: Aetiology of club foot. J Bone Joint Surg Br 1975;57:262.

187. Ha-Vinh R, Alanay Y, Bank RA, Campos-Xavier AB, Zankl A, Superti-Furga A, Bonafe´ L. Phenotypic and molecular characterization of Bruck syndrome (osteogenesis imperfecta with contractures of the large joints) caused by a recessive mutation in PLOD2. Am J Med Genet A 2004;131:115-20.

188. McPherson E, Clemens M. Bruck syndrome (osteogenesis imperfecta with congenital joint contractures): review and report on the first North American case. Am J Med Genet 1997;70:28-31.

189. Wilmore HP, Smith MJ, Wilcox SA, Bell KM, Sinclair AH. SOX14 is a candidate gene for limb defects associated with BPES and Mobius syndrome. Hum Genet 2000;106:269-76.

190. Wittebol-Post D, Hennekam RC. Blepharophimosis, ptosis, polythelia and brachydactyly (BPPB): a new autosomal dominant syndrome? Clin Dysmorphol 1993;2:346-50.

191. Verzijl HT, van der Zwaag B, Cruysberg JR, Padberg GW. Mobius syndrome redefined: a syndrome of rhombencephalic maldevelopment. Neurology 2003;61:327-33.

192. Weinberg CR, Wilcox AJ, Lie RT. A log-linear approach to case-parent-triad data: assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. Am J Hum Genet 1998;62:969-78.

193. Wilcox AJ, Weinberg CR, Lie RT. Distinguishing the effects of maternal and offspring genes through studies of "case-parent triads". Am J Epidemiol 1998;148:893-901.

194. van Den Oord EJ, Vermunt JK. Testing for linkage disequilibrium, maternal effects, and imprinting with (In)complete case-parent triads, by use of the computer program LEM. Am J Hum Genet 2000;66:335-8.

195. Schiaffino S, Reggiani C. Molecular diversity of myofibrillar proteins: gene regulation and functional significance. Physiol Rev 1996;76:371-423.

196. Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV, Gottesman MM. A "silent" polymorphism in the MDR1 gene changes substrate specificity. Science 2007;315:525-8.

197. Komar AA. Silent SNPs: impact on gene function and phenotype. Pharmacogenomics 2007;8:1075-80.

198. Perry SV. Vertebrate tropomyosin: distribution, properties and function. J Muscle Res Cell Motil 2001;22:5-49.

199. Savill SA, Leitch HF, Daly AK, Harvey JN, Thomas TH. Polymorphisms in the tropomyosin TPM1 short isoform promoter alter gene expression and are associated with increased risk of metabolic syndrome. Am J Hypertens 2010;23:399-404.

200. Burton PR, Tobin MD, Hopper JL. Key concepts in genetic epidemiology. Lancet 2005;366:941-51.

201. Dobbs MB, Gurnett CA. Update on clubfoot: etiology and treatment. Clin Orthop Relat Res 2009;467:1146-53.

202. Sommer A, Blanton SH, Weymouth K, Alvarez C, Richards BS, Barnes D, Hecht JT. Smoking, the xenobiotic pathway, and clubfoot. Birth Defects Res A Clin Mol Teratol 2011;91:20-8.

203. Tao Y, Xi S, Briones V, Muegge K. Lsh mediated RNA polymerase II stalling at HoxC6 and HoxC8 involves DNA methylation. PLoS One 2010;5:e9163.

204. Lan F, Bayliss PE, Rinn JL, Whetstine JR, Wang JK, Chen S, Iwase S, Alpatov R, Issaeva I, Canaani E, Roberts TM, Chang HY, Shi Y. A histone H3 lysine 27 demethylase regulates animal posterior development. Nature 2007;449:689-94.

205. Nelson CE, Morgan BA, Burke AC, Laufer E, DiMambro E, Murtaugh LC, Gonzales E, Tessarollo L, Parada LF, Tabin C. Analysis of Hox gene expression in the chick limb bud. Development 1996;122:1449-66.

206. Kawamura A, Koshida S, Takada S. Activator-to-repressor conversion of T-box transcription factors by the Ripply family of Groucho/TLE-associated mediators. Mol Cell Biol 2008;28:3236-44.

207. Muragaki Y, Mundlos S, Upton J, Olsen BR. Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13. Science 1996;272:548-51.

208. Khambata-Ford S, Liu Y, Gleason C, Dickson M, Altman RB, Batzoglou S, Myers RM. Identification of promoter regions in the human genome by using a retroviral plasmid library-based functional reporter gene assay. Genome Res 2003;13:1765-74.

209. Maston GA, Evans SK, Green MR. Transcriptional regulatory elements in the human genome. Annu Rev Genomics Hum Genet 2006;7:29-59.

210. Gahlmann R, Kedes L. Tissue-specific restriction of skeletal muscle troponin C gene expression. Gene Expr 1993;3:11-25.

211. Trivedi CM, Patel RC, Patel CV. Differential regulation of HOXA9 expression by nuclear factor kappa B (NF-kappaB) and HOXA9. Gene 2008;408:187-95.

212. Hall BK, Miyake T. All for one and one for all: condensations and the initiation of skeletal development. Bioessays 2000;22:138-47.

213. Stadler HS, Higgins KM, Capecchi MR. Loss of Eph-receptor expression correlates with loss of cell adhesion and chondrogenic capacity in Hoxa13 mutant limbs. Development 2001;128:4177-88.

214. Brohmann H, Jagla K, Birchmeier C. The role of Lbx1 in migration of muscle precursor cells. Development 2000;127:437-45.

215. Bryson-Richardson RJ, Currie PD. The genetics of vertebrate myogenesis. Nat Rev Genet 2008;9:632-46.

216. Marini NJ, Hoffmann TJ, Lammer EJ, Hardin J, Lazaruk K, Stein JB, Gilbert DA, Wright C, Lipzen A, Pennacchio LA, Carmichael SL, Witte JS, Shaw GM, Rine J. A genetic signature of spina bifida risk from pathway-informed comprehensive gene-variant analysis. PLoS One 2011;6:e28408.

VITA

Katelyn Songin Weymouth was born in Southbridge, MA on March 25, 1981 to Karin M. Songin and Russell F. Weymouth, Jr.. Three years later, she welcomed her younger sister, Emily Ann to the family. Katelyn grew up in Charlton, MA where she attended Shepherd Hill Regional High School in Dudley, MA. As a high school student, she was involved in student council, peer helpers, the national honor society and a member of the 1998 Division II Women's Soccer Championship team. Katelyn graduated from The Hill on June 6, 1999 and went on to enroll as a Molecular and Cell Biology major at the University of Connecticut in Storrs, CT. However, three weeks into her freshman year, she contracted bacterial meningitis causing a long stay in multiple hospitals and intense rehabilitation to overcome the loss of both feet. One year later, August 2000, she returned to UConn and graduated with her Bachelor of Science in Molecular and Cell Biology in May 2004. During her time at UConn, she was strongly involved in advocating for bacterial meningitis and vaccination awareness. Katelyn played a strong role in the passage of bills that now require all students living on college campuses in CT and MA to be vaccinated for bacterial meningitis. In continuing with her studies, she enrolled in the Applied Genomics Master's program at UConn, while maintaining a full time job as a Diagnostic and Serology Technician. In 2006, she graduated and left the comfort of her New England home to pursue a PhD degree at the University of Texas-Houston, Graduate School of Biomedical Sciences. However, she did not leave behind her Massachusetts roots neither her love for her hometown teams, the Red Sox, Patriots, Celtics, Bruins and the UConn Huskies. In October of 2012, she successfully completed and obtained her PhD in Human and Molecular Genetics. Katelyn now looks forward to a successful future in genetic research.