EARLY DISRUPTION OF THE EXTRACELLULAR MATRIX IN MURINE

MUCOPOLYSACCHARIDOSIS I

by

JONATHAN MICHAEL HEPPNER

B.Sc., The University of Alberta, 2009

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Medical Genetics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

July 2015

© Jonathan Michael Heppner, 2015

Abstract

Progressive skeletal and connective tissue disease is a major clinical burden in Mucopolysaccharidosis type I (MPS I). Although enzyme replacement therapies are available and improve some aspects of the disease, bone and joint disease is recalcitrant. The underlying pathogenic mechanisms of MPS I skeletal and connective tissue disease, and the basis of the recalcitrance to therapy, remain unknown. The classical view of MPS I describes somatic disease as the direct result of glycosaminoglycan (GAG) accumulation; however, it is now clear that many lysosomal storage disorders involve more complex pathogenic mechanisms than simple GAG storage. In order to understand the pathogenic mechanisms underlying skeletal and connective tissue disease in MPS I, I have used proteomic and genome wide expression studies of the femoral head growth plate cartilage, and functional studies of the murine MPS I model knee joint to identify early pathogenic events. Three and five-week-old mice were used; thus these studies represent a previously-unexamined time point at which underlying pathogenic mechanisms may be discovered.

Unbiased iTRAQ differential proteomic and multiple reaction monitoring mass spectrometry approaches identified significant decreases in six key structural and signalling extracellular matrix proteins (biglycan, type I collagen, fibromodulin, lactotransferrin, proline/arginine-rich end leucine-rich repeat protein, and SERPINF1). Genome-wide expression studies in five-week growth plate cartilage revealed fourteen significantly deregulated mRNAs (*Adamts4, asporin, chondroadherin, type II collagen, type IX collagen, hyaluronan and proteoglycan link protein, lumican, matrillin 1, matrix metalloproteinase 3, osteoglycin,*

ii

osteomodulin, prolyl 4-hydroxylase, alpha polypeptide II, proline/arginine-rich end leucine-rich repeat protein, and member RAS oncogene family 32). The involvement of members of the small leucine repeat proteoglycan family (asporin, chondroadherin, osteoglycin, osteomodulin, and proline/arginine-rich end leucine-rich repeat protein) in MPS I disease pathogenesis is novel and intriguing, as these proteins are associated with the pathogenesis of osteoarthritis. Functional studies of the MPS I mouse knee joint suggested that early disruption of the extracellular matrix may predispose skeletal and connective tissues to late-stage degeneration. These results imply that biomechanical failure of chondro-osseous tissue may underlie skeletal and joint disease in MPS I. This represents a novel finding which has clear therapeutic implications.

Preface

This dissertation is the product of research performed by me during my Doctor of Philosophy in Medical Genetics program. All experiments and data analysis were performed by me except the following. **Chapter 1:** The initial experimental project was designed by Dr. Lorne Clarke and further expanded by Dr. Clarke and myself. **Chapter 3:** The iTRAQ sample labeling and analysis by LC-MS was carried out by the staff at the University of Victoria Genome BC Proteomic Facility. **Chapter 4:** While the development of the multiple reaction monitoring protein validation assay was carried out by me with the assistance of Drs. Gregg Morin and Annie Moradian, the final assay was performed by the University of Victoria staff. The immunohistochemical staining was performed in the lab of Dr. Frank Zaucke, who also provided assistance with interpretation of the results. **Chapter 5:** The Illumina genome-wide RNA expression analysis was performed by staff at the Center for Molecular Medicine and Therapeutics Gene Expression core facility.

Parts of this dissertation in **Chapters 2, 3, 4, 5, and 6** have been published where noted in:

Heppner JM, Zaucke F, Clarke LA. *Extracellular matrix disruption is an early event in the pathogenesis of skeletal disease in mucopolysaccharidosis I.* Mol Genet Metab 114 (2015) 146-155.

Table of Contents

Abstract	ii
Preface	iii
Table of Contents	iv
List of Tables	vii
List of Figures	ix
Acknowledgments	xi
1. Introduction	1
1.1. Mucopolysaccharidosis Type I (MPS I) Clinical Essentials	1
1.2. Animal Models of MPS I	
1.3. Glycosaminoglycans: The Primary Storage Material	6
1.4. Chondro-osseous Disease in the Mucopolysaccharidoses (MPSs)	8
1.5. Understanding Chondro-osseous System Development and Biomechanics	13
1.6. Previously Reported Pathogenic Mechanisms of MPS Chondro-osseous Disease	23
1.7. Therapeutic Intervention in MPS I and Other MPSs	27
1.8. Rationale and Research Questions	29
2. Methods	33
2.1. iTRAQ Proteomic Procedures	33
2.1.1. iTRAQ Proteomics	33
2.1.2. iTRAQ Limitations	38
2.1.3. Protein Extraction from Bone Tissues	39
2.1.4. Sample collection: Femoral Head Growth Plate Collection	40
2.1.5. Protein Extraction	41
2.1.6. Sample Replicate Number and Pooling	42
2.1.7. Sex Ratios in Pooling	49
2.1.8. iTRAQ LC-MS	51
2.1.9. ProteinPilot Software	52
2.1.10. The iTRAQ Ratio p-value	54
2.1.11. Background Correction	56
2.1.12. Bias Correction	57
2.1.13. Extreme iTRAQ Ratios	58
2.1.14. False Discover Rate (FDR) Analysis of iTRAQ Data	59
2.2. Pathway Analysis of Candidates	59
2.3. Western Blot Analysis	61

2.4. Immunohistochemical Analysis	
2.5. MRM Panel	
2.6. Modulation of MPS I Chondro-osseous Phenotype by Exercise	
2.6.1. Mouse Exercise	
2.6.2. Measurement of Knee Joint Arthropathy	
2.7. RNA Expression Studies	
2.7.1. RNA Extraction	
2.7.2. Illumina Hybridization Chip	
2.7.3. qPCR	
3. Proteomics Results Analysis	
3.1. Introduction	
3.2. Protein Confidence and False Discovery Rate (FDR) Analysis	
3.3. Bias Correction	
3.4. Summary of Primary Analysis	
3.5. Overall iTRAQ Ratios Between Protein Extracts	
3.6. Candidate Lists	
3.7. Pathway Analysis of Candidates	
3.8. Summary	
4. Proteomics Candidate Validation and Functional Studies	
4.1. Introduction	
4.2. iTRAQ Validation	
4.2.1. Western Blots of Bone Protein Lysates	
4.2.2. Western Analysis of MPS I Murine Serum	
4.3. MRM Assay	
4.3.1. Peptide Selection	
4.3.2. Peptide Transition Selection and Optimization	
4.3.3. MRM Validation of iTRAQ Results	
4.4. Immunohistochemical Femoral Head Analysis	
4.5. Modulation of MPS I Chondro-osseous Phenotype by Exercise	
4.6. Summary	
5. RNA Expression Studies	
5.1. Introduction	
5.1.1. Genome Wide Expression Arrays	
5.1.2. RNA Integrity	

5.1.3. qPCR Validation
5.2. Illumina Array Data Analysis
5.3. Illumina Genome Array Candidates
5.4. qPCR Validation
5.5. Analysis of RNA Results
5.6. Summary
6. Discussion and Conclusions
6.1. Discussion
6.1.1. Biomechanical Failure Underlying Chondro-osseous Disease in MPS I 139
6.1.2. Biomechanical Failure Model: Linking the pathogenesis of MPS skeletal disease to forms of Arthritis and other Genetic Skeletal Dysplasias
6.1.3. Proteomic and Transcriptomic Analysis of MPS I Murine Model Chondro-osseous Disease
6.1.4. Relationship between the Proteomic and Transcriptomic Data
6.2. Limitations and Future Studies
6.3. Conclusions
Works Cited
Appendix 1 – R Code for Illumina Array Analysis

List of Tables

Table 1.1: Disaccharide composition of glycosaminoglycans
Table 1.2: Summary of chondro-osseous clinical phenotypic traits seen in MPS patients 10
Table 1.3: Important chondro-osseous system proteins relevant to MPS I, and their known roles and disease phenotypes. 14
Table 2.1: Analysis of 7 +/+ mouse micro-dissected femoral head growth plate protein extracts (measured by BCA assay) to assess variation between individuals
Table 2.2: Pooling volumes and statistics used to create the 5 week iTRAQ proteomic pools for the femoral head growth plates. 47
Table 2.3: Pooling volumes and statistics used to create the 5 week iTRAQ proteomic pools for the cranial plates. 48
Table 2.4: Final average total protein concentration of 5 and 3 week sample pools using the Micro-BCA assay, and volumes used for iTRAQ labelling to normalize for equal protein 49
Table 2.5: Sex composition of protein 5 and 3 week pools used for iTRAQ analysis. "FH"denotes femoral head samples while "C" denotes cranial plate samples
Table 2.6: qPCR gene targets and SABiosciences cataloged primer pairs
Table 3.1: Lists of proteins from 3 and 5 week MS runs which meet the confidence and fold change cut off values. 73
Table 3.2: Summary of differentially-expressed proteins between the 5 week MPS I and wildtype mouse growth plate samples. 77
Table 3.3: Final protein candidate list from 5 week cranial plate analysis based on peptide ratio agreement. 79
Table 3.4: Final protein candidate list from 3 week growth plate and cranial plate analyses based on peptide ratio agreement. 80
Table 3.5: Query proteins ranked within the top 100 protein hits with strong evidence of co-regulation from SLRP family MEM analysis.86
Table 4.1: List of protein candidates that overlap with those identified with high confidence in the HUPO Plasma Proteome
Table 4.2: List of proteins and corresponding selected peptides for which the MRM assay was developed. 99

Table 4.3: Transitions and LC retention times (RT) for all peptides in the MRM panel. Naming isaccording to parent protein.peptide abbreviation. peptide sequence.102
Table 4.4: Selected MRM proteins not from iTRAQ candidates. 113
Table 4.5: Summary of MRM quantitation results in the 5 week MPS I growth plate compared to wildtype indicating significant protein abundance changes. 115
Table 4.6: Summary of calculated GAG stain intensity (grey value) and articular cartilage area values averaged across serial sections of the knee joints in four areas from wildtype and MPS I mice that were exercised for 5 weeks compared to un-exercised controls
Table 5.1: Average ratio of signal intensity for 15,658 genes between biological and technical replicates. 131
Table 5.2: List of GenomeStudio® candidate loci with significant DiffScores detected in only MPS I samples. 132
Table 5.3 Summary of qPCR results for the 5 week MPS I compared to wildtype growth plate. Significant p-values are reported (after adjusting for multiple hypothesis testing) for the ratio (fold change) of RNA abundance in MPS I:wildtype
Table 5.4: Full qPCR array results. Significant (after adjusting for multiple hypothesis testing) p-values are reported. 135
Table 5.5: RNA expression analysis candidates consistent between analyses. Negative fold change indicated a decrease in the MPS I Samples. 137

List of Figures

Figure 1.1: Example disaccharide sequences for heparan sulfate and dermatan sulfate glycosaminoglycans
Figure 1.2: a) Example of the growth plate cartilage in a mouse femoral head stained with H&E and conceptual femoral head growth plate illustrating key component locations and the portion of tissue used in this dissertation. 20
Figure 1.3: Flow chart of experimental approaches used towards identifying a model of chondro- osseous disease pathogenesis in early MPS I development
Figure 2.1: Diagram of iTRAQ tag composition
Figure 2.2: Femoral head growth plate micro-dissection method showing progressive isolation of the growth plate zone (including a portion of the resting cartilage, proliferative zone, hypertrophic zone, apoptotic zone, and trabecular bone)
Figure 2.3: Pooling strategy for individual mice and protein extracts, for femoral head growth plate and cranial plate samples
Figure 3.1: FDR analysis report generated by ProteinPilot for the 5 week growth plate experiment
Figure 3.2: Distribution of fold change in 842 proteins from the 5 week growth plate iTRAQ MS run comparing MPS I to wildtype with and without Background Correction
Figure 3.3: Workflow of bioinformatics tools used to generate hypotheses for key altered signalling and regulatory networks based on the iTRAQ proteomics candidates
Figure 3.4: Examples of RSAT feature maps for SLRP protein family member promoters regions, and SLRP and other protein candidate predicted binding sites for SOX5 and SOX9 transcription factors
Figure 4.1: Images of Western blots for Perlecan, COMP, and PRELP from individual mouse MPS I or wildtype grwoth plate protein extracts, and densitometry results
Figure 4.2: Images of Western blots for Perlecan, COMP, and PRELP from individual mouse MPS I or wildtype serum protein, and densitometry results
Figure 4.3: Example of MRM quantitation of two proteins in total protein lysates for peptides of Thrombospondin 3 and Fibronectin 1
Figure 4.4: MRM transition areas for significantly decreased proteins comparing the 5 week MPS I to wildtype growth plate

Figure 4.5: Representative immunohistochemical staining images of key ECM proteins in the 5 week MPS I and wildtype femoral head growth plate
Figure 4.6: Representative frontal knee sections stained with Safranin O and Fast Green from exercised and control wildtype and MPS I mice
Figure 5.1: Venn diagram of Illumina expression array GenomeStudio and custom R analysis candidate intersection
Figure 5.2: Scatter plot of average probe intensity per gene in MPS I and wildtype samples for all genes identified using GenomeStudio®
Figure 5.3: Scatter plot of log probe intensity per gene in MPS I (-/-) and wildtype (+/+) samples for all genes identified using custom R analysis
Figure 6.1: Hypothetical chronology and relationships between chondro-osseous metabolic activities and pathogenic changes in the MPSs
Figure 6.2: The biomechanical failure model describes 3 contiguous phases of chondro-osseous pathogenic progression (Setup, Establishment, and Culmination) beginning in early prenatal development and progressing through to maturity

Acknowledgments

I would like to thank my supervisor, Dr. Lorne Clarke for all of his advice, support, and guidance throughout my degree program. He always pushed me to expand my boundaries, and allowed me to learn a tremendously diverse set of research skills. I'd also like to thank Dr. Annie Moradian for her hours of guidance and work towards developing the MRM assay, and for her intellectual help and support. I learned a great deal about mass spectrometry through her expertise. Derek Smith at the University of Victoria Genome BC Proteomics Facility was also very helpful in his services regarding the iTRAQ and final MRM data acquisition, as well as his intellectual input in analysis of the data. Thank you to my advisory committee members, for your always helpful suggestions and thoughts about my project, and for your mentorship throughout my training. In particular, Dr. Wyeth Wasserman and his lab played a pivotal role in educating me about bioinformatic analysis. Finally, a huge "thank you" to my parents, who have always (and continue to be), the most supportive people in my life and incredible role models. I could not have made it to this point without them.

1. Introduction

1.1. Mucopolysaccharidosis Type I (MPS I) Clinical Essentials

Mucopolysaccharidosis type I (MPS I; Hurler, Hurler/Sheie, Sheie syndromes) (OMIM #607014 and 607016) is a lysosomal storage disease caused by deficiency of α-L-iduronidase (IDUA gene product) and has a worldwide population prevalence of 1:100,000 (1). Deficiency of this lysosomal hydrolase leads to partially or completely defective glycosaminoglycan (GAG) degradation, directly affecting the degradation of heparan sulfate (HS) and dermatan sulfate (DS). Incompletely processed GAG fragments accumulate in the lysosome leading to cellular vacuolation. The clinical phenotype of MPS I is characterized by progressive multi-system disease including CNS involvement (leading in some cases to mental retardation), communicating high-pressure hydrocephalus, corneal clouding, glaucoma, retinal degeneration, coarse facial features (including thickening of the tongue and lips), organomegaly (primarily the spleen and liver) and associated abdominal protuberance, Eustachian tube dysfunction, chronic rhinitis, mitral and aortic valve thickening (leading to regurgitation), inguinal hernias, obstructive airway disease (including dysmorphic tracheal rings leading to a narrow airway and pulmonary hypertension), and progressive skeletal disease (discussed below) (2).

MPS I has been historically classified as severe (Hurler syndrome), attenuated (Hurler/Sheie syndrome), or mild (Sheie syndrome) based on relatively subjective assessment of age of onset and symptom severity. Patients with severe disease typically have early onset, with diagnosis during the first postnatal year, and rapidly progressive symptoms, usually surviving only until age 10 in the absence of therapeutic intervention (2). Patients with less severe disease (i.e. attenuated) may not present with symptoms until adulthood and may show very slowly progressive symptoms thereafter (2). Progressive cognitive impairment is seen in only severe patients, while skeletal disease is prevalent in all severe and attenuated MPS I patients. Current estimates suggest that 80% of MPS I cases are severe, though bias may be present in this estimate, whereby missed diagnoses of mild attenuated cases would result in underestimation of the attenuated disease incidence (2).

In reality, it may be more realistic to consider MPS I a continuous spectrum of disease from severe to attenuated. Biochemical studies in a small number of patient cell lines has demonstrated that as little as 1% residual α -L-iduronidase activity may lead to an attenuated MPS I phenotype (3, 4). However, enzyme activity alone is not sufficient to predict severity of disease in individual patients unless the alleles in question have previously been associated with a particular phenotypic outcome (2). Accordingly, current nomenclature discussions have suggested that a more appropriate classification of MPS I would be the binary designation of either severe or attenuated disease, with the former designated as Hurler syndrome (2).

The observation that small amounts of residual enzyme activity drastically alters the natural history of disease makes MPS I an excellent target for genetic therapies. Currently, available therapies include enzyme replacement therapy using intravenous recombinant enzyme or hematopoietic stem cell transplantation. Both approaches have been shown to reverse or prevent some disease manifestations of MPS I including coarse facial features, organomegaly, and in some cases neurological disease (2, 3, 5), although it does not reverse skeletal disease or cardiac valve disease (2, 6) as will be discussed in further detail below.

Skeletal disease and the associated symptoms represent one of the greatest disease burdens affecting quality of life and disability in both severe and attenuated patients (7). In addition, the relative recalcitrance of skeletal disease to therapy represents a critical issue in the management of MPS I patients. Although the clinical phenotype of MPS I and its response to

enzyme replacement therapy have been extensively studied, the pathogenic mechanisms underlying the skeletal phenotype and reasons for therapeutic shortcomings remain unknown.

While MPS I has been classically considered a "storage disorder," direct lysosomal and cellular GAG accumulation only partially explains phenotypic features (e.g. organomegaly and coarseness of the face); thus it is now clear that many of the clinical disease features cannot be explained solely and directly on the basis of GAG storage. Therefore additional pathogenic mechanisms must be involved. The concept of pathogenic cascades underlying disease pathogenesis in other lysosomal diseases has been well demonstrated, but such observations have only recently been made in the MPSs. These pathogenic cascades identified thus far in the literature are described in detail in section **1.6**.

1.2. Animal Models of MPS I

There are several animal models described in the literature that recapitulate the phenotypic traits seen in human MPS I. The first published MPS I murine model was used for this dissertation for the reasons discussed below.

Mutation of the *Idua* gene locus has been noted in canine (8), feline (9), and murine species (10, 11). In addition, a common nonsense allele analogous to the human W402X mutation has been knocked into the mouse *Idua* locus to produce a specific mouse model (12). The canine model of MPS I was established from a naturally-occurring G to A transition splice site mutation in intron I (13), leading to α -L-iduronidase deficiency in a beagle. This canine model had a phenotype consistent with attenuated MPS I (displaying stunted growth, enlarged tongue, joint degeneration, skeletal effusions, and cardiac valve thickening (8)). The feline model of MPS I was also naturally occurring, discovered initially in a single domestic short hair animal displaying progressive mobility limitation, broad facial structure, corneal clouding, and

thickened skin consistent with the clinical phenotype of human MPS I (9). The causative lesion was identified as a 3 base pair deletion in either nucleotides 1107-1109 or 1108-1110 (14).

The first murine model of MPS I was created in vitro by insertion of a neomycin resistance gene cassette into the *Idua* exon VI (10) and was established on the C57BL/6J strain background. Clarke *et al* note that interruption of exon VI was chosen so as to avoid disruption of the *Sat1* locus, which is within intron II of the *Idua* locus (10). Insertion of the neomycin resistance cassette effectively blocked transcription of full-length *Idua* mRNA as shown by qPCR, and thus produced a true *Idua* null model (10). Accordingly this murine model of MPS I has a -/- genotype when homozygous for the interrupted *Idua* locus and largely mimics the human phenotype of MPS. Mice homozygous for the null allele showed evidence of gross pathology as early as 4 weeks of age, with a broadened face and thick, short digits and paws (10). Thickened ribs were also apparent radiographically by 4 weeks of age (10). Urinary GAG excretion was 2-fold increased over wild type mice, as it is also increased in patients (10). The phenotype was progressive, with obvious lysosomal storage of GAGs by 8 weeks of age, worsening skeletal symptoms by 15 weeks of age (10).

The second murine model of MPS I was created by another group through insertion of a neomycin resistance cassette into the same location in exon VI of the *Idua* locus on the C57BL/6J background (11), and is identical to the first murine model described above. An additional MPS I mouse model has been described, produced by knocking-in the common human severe MPS I mutation W402X (W392X in mice) into the murine *Idua* locus (12). Both the W402X mutation in humans and W392X mutation in mice are null alleles resulting in premature transcriptional termination and no expression of α -L-iduronidase protein, as confirmed by lack of

detectable enzyme activity or mRNA in the C57BL/6J homozygous knock-in mice (12). Homozygous W392X mice showed phenotypic, histologic, and urinary GAG features consistent with the two knockout murine models and human MPS I (12). As all three mouse models have been established in the C57BL/6J genetic background, it is unknown what effect genetic background would have on the MPS I phenotype in murine models.

The genetic background in the canine and feline strains is more heterogeneous than the murine models, due to extensive inbreeding in the murine models. Inbreeding of the murine MPS I models is beneficial for proteomic and transcriptomic studies because it can prevent background genetic differences (i.e. changes not related to Idua mutation) from obscuring identification of MPS I pathogenic mechanisms. The inbred strain may have alleles which modify the MPS I phenotype but are uncommon or not present in humans, and such modifiers may also bias identification of pathogenic mechanisms in the murine model. However, the benefit of using the inbred murine model of MPS I for the types of studies carried out here outweighs this risk. Furthermore, existing protein databases are primarily populated by proteins identified in mice and humans. Although other species are represented, the mouse subset of proteins in current databases is more comprehensive than for other models, allowing for more comprehensive protein matching during a proteomic analysis; thus mice are arguably better models for database matching protein identification. Rapid breeding of mice can offset some of the tissue sample size limitations compared to larger animals. The robust history of the first MPS I murine model and established breeding colony in the Clarke lab made this model a clear choice

1.3. Glycosaminoglycans: The Primary Storage Material

The primary effect of α -L-iduronidase deficiency is defective glycosaminoglycan (GAG) catabolism, resulting in GAG accumulation. Specifically, α -L-iduronidase cleaves terminal iduronate residues (shown in **Figure 1.1**) during the stepwise catabolism of GAGs. Catabolism of DS and HS can include the activities of iduronate sulfatase, α -L-iduronidase, acetylgalactosamine 4-sulfatase, hexosaminidase, and β -glucuronidase, depending on the exact sequence of the saccharide chain and modifications which have been added. The examples in **Figure 1.1** show a sulfated iduronate residue in HS and a sulfated glucosamine residue in HS. Catabolism of the DS terminal end shown would require the activity of iduronate sulfatase before α -L-iduronidase could act. Catabolism of the HS terminal end shown would require acetylgalactosamine 4-sulfatase activity before α -L-iduronidase. A deficiency of α -L-iduronidase blocks further progression of GAG catabolism at iduronate residues, thus leading to accumulation of GAG fragments with non-reducing end terminal iduronate residues.

GAGs are a heterogeneous group of linear heteropolysaccharide molecules, composed of specific repeating disaccharide units with a typical molecule length of 50 to 150 disaccharides. The individual disaccharide units define the GAG subtype (**Table 1.1**). A representation of disaccharides of the GAGs HS and DS are shown in **Figure 1.1**. In addition to the core

Table 1.1: Composition of glycosaminoglycans

GAG	Saccharide U	Modifications	
	Glucosamine/galactosamine	Hexuronic acid	
Heparan sulfate (HS)	glucosamine	iduronate or glucuronate	N-acetylation, N- sulfation, O-sulfation, epimerization
Dermatan sulfate (DS)	galactosamine	iduronate or glucuronate	N-acetylation, N- sulfation, O-sulfation, epimerization
Chondroitin sulfate (CS)	galactosamine	glucuronate	N-acetylation, N- sulfation, O-sulfation
Keratan sulfate (KS)	glucosamine and galactose	-	N-acetylation, N- sulfation, O-sulfation, epimerization
Hyaluronic acid (HA)	glucuronate and glucosamine	-	-

disaccharide units, GAG chains (with the exception of HA) have N-acetylation, N-sulphation, Osulphation modification, as well as epimerisation of glucuronate to iduronate. GAGs can serve an incredibly diverse range of biological roles based on the specific disaccharide sequence and functional modifications. With the exception of HA, GAGs normally exist covalently linked to a core protein through serine or asparagine residues to produce a proteoglycan. Functional activity of proteoglycans is largely determined by the GAG species they are modified by, and the GAGs themselves also provide functions which can be independent of the core protein. Thus GAGs are considered highly bio-active molecules in light of their diverse functional capacity to modulate protein activity (6).

a) An example heparan sulfate disaccharide



b) An example dermatan sulfate disaccharide



Figure 1.1: Example disaccharide sequences for heparan sulfate and dermatan sulfate glycosaminoglycans. These two disaccharide sequences are for illustrative purposes only and do not exhaustively describe HS and DS saccharide composition.

1.4. Chondro-osseous Disease in the Mucopolysaccharidoses (MPSs)

Despite having unique genetic enzyme deficiencies (and thus different primary catabolic defects and storage material), the MPSs lead to partially convergent chondro-osseous features. The "chondro-osseous" system can be defined as tissues which are comprised of, or derived from, cartilage or bone; including skeletal elements, articular cartilage, ligaments, tendons, tracheal rings, spinal discs, knee meniscii, the synovium, teeth, and cardiac valves. This dissertation focuses on the skeletal aspects of the chondro-osseous system in MPS I, most

specifically the growth plate cartilage. However, it is expected that discoveries made in the growth plate will have wider implications for the other chondro-osseous tissues with which the growth plate shares its developmental origins, including the articular cartilage.

Clinical similarity of chondro-osseous phenotypic traits between the MPSs may reflect related mechanisms of pathogenesis; thus discussion of chondro-osseous disease in MPS I should include observations in the other MPSs as well. The clinical phenotype of MPS chondro-osseous disease has been thoroughly characterized in both patient populations and animal models. The reader is directed to the following sources for the most comprehensive review of clinical features: covering MPS I (Hurler syndrome) MPS II (Hunter syndrome), MPS III (Sanfilippo syndrome), MPS IV (Morquio syndrome), MPS VI (Maroteaux-Lamy syndrome), MPS VII (Sly syndrome) (15, 16, 17), and MPS IX (18). Primary similarities in several of the MPSs include reduction of longitudinal bone growth, generalized abnormal bone morphology (dysostosis multiplex), arthropathy, and alteration of joint mobility (**Table 1.2**). MPS I and II have similar phenotypes, while MPS VI and VII share a different set of phenotypic traits. MPS III, IV, and IX have some unique phenotypic traits (**Table 1.2**). Shared phenotypic relationships suggest the possibility of overlap in pathogenic mechanisms.

				MPS			
Phenotype	Ι	П		IV	VI	VII	IX
Short stature	•	•	• *	• **	•	•	•
Restricted ROM	•	•	• *	0	•	•	0
Trigger digits	•	•	•	•	•	•	0
Kyphoscoliosis	•	•	•	•	•	٠	0
Acetabular dysplasia	•	•	•	•	•	٠	0
Arthropathy	•	•	• *	•	•	٠	•
Dysostosis multiplex	•	•	0	●**	•	•	0
Carpal tunnel syndrome	•	•	0	•	•	•	0
Odontid deformity	•	•	0	•	•	•	0
Coxa valga	•	•	0	•	•	•	0
Genu valgum	•	•	0	•	•	•	0
Atlanto-axial instability	•	•	0	•**	•	•	0
Cardiac valve dysfunction	•	•	0	•	•	•	0
Tracheal deformity	•	•	0	0	0	•	0
Corneal opacity	•	0	0	•	•	•	0
Pectus carnatum	0	0	0	•	•	•	0
Joint hypermobility	0	0	0	•	0	0	0
Pes planus	0	0	0	•	0	0	0
Joint soft tissue masses	0	0	0	0	0	0	•
Acetabular erosions	0	0	0	0	0	0	•

Table 1.2: Summary of chondro-osseous clinical phenotypic traits seen in MPS patients. • present, \circ absent, *mild, and **severe. Note that MPS V and VIII are not used as classifications for the MPSs.

Skeletal involvement (termed "dysostosis multiplex") in the MPSs can manifest as increased cortical bone thickness, oar shaped ribs, joint malalignment (genu valgum, coxa valga), scoliosis, "beaked" vertebrae, acetabular dysplasia, and atlanto-axial instability in most of the MPSs (18, 15, 19, 17, 20) (**Table 1.2**). The term dysostosis multiplex is poorly defined and non-specific; conferring no information regarding classification of disease severity or diagnosis despite its widespread use in the literature. Short stature and skeletal deformity appear to be permanent and irreversible, due to the limited existence of the growth plate in long bones. There is also evidence supporting reduction of bone remodeling in the MPSs (as will be discussed later), which would contribute to the permanence of skeletal pathology.

Patients with MPS I, II, VI, and VII develop progressive joint stiffness which can severely limit mobility and function. Extensively restricted range of motion (ROM) of joints has been observed for each of the MPSs and has been used as a biometric measure in therapeutic trials, in combination with the 6 minute walk test and growth measurements (21, 22). Joints are complex systems of interdependent tissues, wherein mobility is dependent on the function of ligaments, the synovial membrane, synovial fluid, articular cartilage, and the subchondral bone structure. This inter-dependent relationship has been well-characterized in the studies of osteoarthritis, where the combination of underlying bone deformity, articular surface degeneration, and inflammatory changes collectively act to limit joint function (23). Neither clinical nor animal model studies in the MPSs have characterized in what ways specific pathology in each of these distinct joint tissues contributes to joint disease. Ligaments, for example, have not been studied in MPS patients or models despite their important role, although carpal tunnel syndrome is a common feature in MPS I, II, IV, VI and VII (approaching 100% prevalence in MPS II (16, 24)). Each tissue in a joint will most likely respond differently to a defect of GAG catabolism, based on the diverse biomechanical properties and requirements these tissues entail. Measurement of joint ROM and walk tests, however, generalize disease burden. Studies in arthritis utilizing MRI and ultrasound imaging (25, 26) suggest potential non-invasive measures to better characterize chondro-osseous disease progression in the MPSs than joint ROM. Recent studies of bone dysplasia in the murine models of MPS I, III, IV, and VII using micro-CT imaging (27) and dual energy x-ray absorptiometry in MPS I, II, III, IV, and VI patients further suggest the potential for utilization of more specific and reproducible imaging measures of chondro-osseous disease in the MPSs.

Despite the progression of chondro-osseous disease (leading to arthropathy, bone deformity, spinal mal-alignment, and joint ROM restriction in the MPSs) observed clinically, lack of systematic measures and reporting has prevented a detailed natural history of chondroosseous disease from being defined. MPS chondro-osseous pathology is progressive, with an early age of onset for certain traits such as skeletal deformity and later onset for other traits such as arthropathy and carpal tunnel syndrome. Deformity of the hip and ribs has been observed radiographically at birth in severe cases (2), though it would also likely be visible in the third trimester if this were to be examined radiographically. MPS I and II patients have been observed to have above average body lengths at birth, but stature below the 3rd percentile by 3 years of age in MPS I or by 5 years of age in MPS II (29), indicating that reduced growth requires a longer period of time to to become apparent. Data from the MPS I Registry (an international database of MPS I patient information maintained by Genzyme Corp.) reveals that the median age of carpal tunnel syndrome in attenuated MPS I is 13 years of age (30) (later than other phenotypic features in attenuated cases), indicating that carpal tunnel syndrome also has delayed onset. Chondro-osseous disease is prevalent in attenuated as well as in severe patients, and represents the largest burden to quality of life and functional capacity in attenuated MPS I patients (7).

A critical facet of MPS disease characterization missing from the literature is the examination of earlier disease events, which will likely be the most informative of potential pathogenic drivers. This is especially relevant to chondro-osseous disease in the MPSs, as the early developmental processes giving rise to these tissues have long ceased at later ages, and there is a limited regenerative capacity.

Most MPS animal model studies have focussed on characterizing late disease, where the articular cartilage and joints of mice in excess of 12 months of age have been studied

histologically and by RNA profiling. Neonatal and pubertal progression of chondro-osseous disease has received little attention, while prenatal development has been almost completely overlooked. Identification of pathogenic mechanisms in the MPSs will require care to be taken to distinguish and understand early drivers of disease. Identification of mechanisms occurring later in disease natural history will be confounded by tertiary, generalized damage responses like inflammation, and the subtle early drivers of phenotype may be obscured by the background of these non-specific activities. This concept of early vs. late mechanisms is especially relevant to therapeutics directed to chondro-osseous disease, as strategies targeting later mechanisms will be both ineffective at stopping early symptom drivers and ineffective at impacting irreversible changes. Although studying earlier drivers in the MPSs is challenging, it is essential to understanding the relationship between the primary genetic defect, storage material, and the clinical phenotype. Unbiased pathogenic mechanism approaches using proteomic and transcriptomic strategies have not been utilized widely in the MPSs, though these approaches hold great potential for discovery of these mechanisms. Although the MPSs are monogenic, I hypothesize that ultimately the mechanisms underlying chondro-osseous disease pathogenesis will be numerous and multifactorial.

1.5. Understanding Chondro-osseous System Development and Biomechanics

Understanding the stages of skeletal development and cell types involved in skeletal development is essential to the discovery and characterization of pathogenic mechanisms underlying chondro-osseous disease in the MPSs. The pathways and cells discussed below and in **Table 1.3** are the basis by which the MPS I chondro-osseous system must be understood and will

be the foundation of analyzing candidates from the unbiased candidate analyses in Chapters 3

and 5.

|--|

				Associated		Common Disease	
Cell Type	Protein	Gene	Key Role	Diseases	Phenotype	Variants	Ref.
Chondrocyte	Aggrecan	ACAN	Structural organization and integrity of cartilage	Familiar osteochondriti s dissicans, spondyloepiph yseal dysplasia (Kimberley type)	Detachment of articular cartilage from subchondral bone, short stature, osteoarthritis (familial osteochonritis dissicans). Severe short stature, brachydactyly, osteoarthritis	-	(31) (32)
	Asporin	ASPN	Potential TGFb regulator in chondrocyte s	-	-	Osteoarthritis, intravertebral disc disease	(33)
	Chondroadherin	CHAD	Potential role in chondrocyte adhesion	-	-	-	(34)
	Collagen I	COL1A 1, COL1A 2	Key structural component of ECM in bone	Caffey disease, Ehlers-Danlos syndrome, osteogenesis imperfecta	Swelling of soft tissues, pain, and excessive hyperostosis (Caffey disease). Joint hypermobility, fragile skin (Ehrlers-Danlos syndrome). Severe bone fragility, blue sclerae, short stature, hearing loss, respiratory problems, and tooth development disorder (Osteogenesis imperfecta)	Osteoporosis, osteoarthritis	(35) (36) (37)

					Common	
Protoin	Gana	Koy Polo	Associated	Phonotypo	Disease	Pof
Collagen II	COL12 A1, COL2A 2	Key structural component of ECM in cartilage	Achondroplasi a, Czech dysplasia, hypochondrog enesis, Kniest dysplasia, plattysponylic skeletal dysplasia (Torrance type), spondyloepim etaphyseal dysplasia (Strudwick type), spondyloepiph yseal dysplasia congenita, spondyloperip heral dysplasia, Stickler syndrome	Short stature and limbs, umbilical and inguinal hernia (achondrogenesis). Platyspondyly, kyphoscoliosis, hearing loss (Czech dysplasia). Short limbs, limited cranial and spinal ossification, flat rounded face (hypochondrogenesis). Short stature and limbs, enlarged joints, joint mobility restriction, arthritis flat rounded face, tracheal instability (Kniest dysplasia). Short stature and limbs, Platyspondyly, lordosis (platyspondylic skeletal dysplasia). Short stature and limbs, lordosis or scoliosis, Platyspondyly, pectus carinatum, coxa vara, clubfoot, and arthritis (spondyloepimetaph yseal dysplasia). Short stature and limbs, kyphoscoliosis or lordosis, Platyspondyly, pectus carinatum, coxa vara, clubfoot, and arthritis (spondyloepimetaph yseal dysplasia). Short stature and limbs, kyphoscoliosis or lordosis, Platyspondyly, pectus carinatum, coxa vara, clubfoot, and arthritis (spondyloepiphyseal dysplasia congenita). Short stature and limbs, lordosis, brachdactyly, clubfoot, platyspnodyly (spondyloperipheral	Avascular necrosis of the femoral head, osteoarthritis	(38) (39) (40) (41) (42) (43) (44) (45) (46)

						Common	
Cell Type	Protein	Gene	Key Role	Associated	Phenotyne	Disease Variants	Ref
Cen Type			in the second seco		dysplasia).		nen
	Collagen X	COL10 A1	Potential role in mineralizati on	Schmid type metaphyseal chondrodyspla sia	Mild short stature, coxa vara, sclerosis of the ribs, flaring of the metaphyses, wide irregular growth plate.	-	(47) (46)
	Collagen IX	COL9A 1, COL9A 2, COL9A 3	Bridge between collagen II and ECM proteoglyca ns	Multiple epiphyseal dysplasia, Stickler syndrome	Short stature, arthritis, scoliosis, clinodactyly, clubfoot (multiple epiphyseal dysplasia).	-	(48) (46)
	Collagen XI	COL11 A1, COL11 A2	Reinforce and organize collagen II fibrils	Stickler syndrome, Marshall syndrome	Pierre Robin sequence, hearing loss, joint hypermobility, arthritis, Platyspondyly, kyphoscoliosis (Marshall syndrome).	Osteoarthritis	(46) (49)
	СОМР	COMP (THBS5)	Potential role in proliferation and apoptosis	Multiple epiphyseal dysplasia, pseudoachond roplasia	Short stature and limbs, wadling gait, osteoarthirits, joint hypermobility, valgus or varus deformity, kyphosis or scoliosis (pseudoachondroplas ia).	-	(47)
	Decorin	DCN	Regulation of collagen fibril organization and TGFb activity	Congenital stromal corneal dystrophy	Corneal clouding and dysmorphology		(50)
	Epiphycan	EPYC	Regulation of collagen fibril organization	-	-	-	(51)
	Fibronectin	FN1	Essential for mineralizati on, regulation of collagen I	-	-	-	(52)

				Associated		Common	
Cell Type	Protein	Gene	Kev Role	Diseases	Phenotype	Variants	Ref.
		Conc	deposition	2.56066	. nenetype	Fundatio	
	Fibromodulin	FMOD	Potential role in collagen I and II fibril organization and TGFb regulation	Kniest dysplasia, hypochondrog enesis	Short limbs, limited cranial and spinal ossification, flat rounded face (hypochondrogenesis).	-	(41) (40)
	Lubricin	PRG4	Boundary lubricant and absorptive properties of synovial fluid	Camptodactyly -arthropathy- coxa vara- pericarditis syndrome	Progressive joint failure associated with noninflammatory synoviocyte hyperplasia and subintimal fibrosis	-	(53)
	Lumican	LUM	Potential role in collagen fibril organization	-	-	-	(54)
	Matrillin 1	MATN1	Potential role in organizing collagen and proteoglyca ns in cartilage	Relapsing polychondritis	Inflammation of ECM tissues in the ear, nose, or eye, arthritis, mitral and aortic valve regurgitation.	-	(55)
	Matrillin 3	MATN3	Potential role in organizing collagen and proteoglyca ns in cartilage	Multiple epiphyseal dysplasia	Short stature, arthritis, scoliosis, clinodactyly, clubfoot	-	(47)
	Osteonectin	SPARC	Essential for collagen I mineralizati on	-	-	-	(56)
	Perlecan	HSPG2	Structural integrity of basement membrane and vascular ECM	Schwartz- Jampel syndrome, Dyssegmental dysplasia Silverman- Handmaker type	Short stature, kyphoscoliosis, bowing of the diaphyses, irregular epiphyses (Schwartz- Jampel syndrome). Anisospondyly, micromelia, flat face, micrognathia, reduced joint mobility, short	-	(57) (58)

				Associated		Common Disease	
Cell Type	Protein	Gene	Key Role	Diseases	Phenotype	Variants	Ref.
					growth plate (Dyssegmental dysplasia (Silverman- Handmaker type)).		
Osteoblast	Alkaline phosphatase	ALPL	Essential for calcium utilization in mineralizati on	Hypophosphat asia	Short limbs, hypercalcemia, odontohypophosphat asia	-	(59)
	Biglycan	BGN	Potential role in collagen fibril organization	-	-	-	(60)
	Periostin	POSTN	Potential role in mineralizati on	-	-	-	(61)
Osteoclast	Matrix metallo- proteinase 13	MMP1 3	Cleaves collagen II (and I to a lesser extent)	Spondyloepim etaphyseal dysplasia Missouri type, Metaphyseal anadysplasia 1	Bowing of the femora and/or tibiae, coxa vara, genu varum and pear-shaped vertebrae (Spondyloepimetaph yseal dysplasia Missouri type). Early short stature and varus deformity with progressive spontaneous resolution (Metaphyseal anadysplasia 1).	-	(62) (63)
	Osteopontin	OPN (SPP1)	Osteoclast attachment to mineralized matrix	-	-	-	(64)

Mammalian bone development involves two distinct processes: endochondral and intramembranous ossification. Intramembranous ossification takes place in flat bones such as the cranio-facial elements as well as during local fracture healing of long bones (65, 66) and entails direct differentiation of mesenchymal progenitor cells into osteoblasts followed by rapid ossification. Intramembranous bone growth does not involve a growth plate, but rather growth occurs through proliferation of the mesenchymal progenitor cells enlarging the template ahead of the front of ossification.

Endochondral ossification of long bones begins with aggregation of mesenchymal progenitor cells, which then differentiate into chondrocytes to produce a cartilaginous bone template. A subset of the chondrocytes in the middle of the cartilage template are calcified to form a bone collar, and a front of ossification then proceeds longitudinally towards the ends of the template while secondary zones of ossification develop in the epiphyses. The two ossification fronts leave a region of chondrocytes between the epiphysis and the metaphysis delineating the growth plate cartilage (**Figure 1.2**). Linear bone growth occurs exclusively in the growth plate as the direct result of deposition of new cartilage within the growth plate and proliferation of the chondrocytes in vertical stacks pushing the epiphyses outwards.

The primary tissue of interest in this dissertation is the femoral head growth plate (shown in **Figure 1.2**). The growth plate consists of resting chondrocytes above the proliferative zone, proliferative chondrocyte stacks within the proliferative zone, hypertrophic chondrocytes at the distal end of the stacks in the hypertrophic zone, and osteoblasts and osteoclasts at the end of the hypertrophic zone (**Figure 1.2**). During sample collection, the growth plate was excised from the femoral head as per **Figure 1.2** (discussed in detail in section **2.1.4**.) such that chondrocytes, osteoblasts, and osteoclasts were included. Accordingly, understanding the protein content of these samples requires understanding of the developmental origin and role of these cell types, as will be discussed below. Furthermore, the signaling pathways and cellular activities associated with these cell types may be important targets of pathogenesis in MPS I chondro-osseous

disease. Since longitudinal bone growth is abnormal in MPS I, understanding the normal function of the growth plate and processes involved in bone growth is important.





Figure 1.2: a) Example of the growth plate cartilage in a mouse femoral head stained with H&E. The stacks of chondrocytes can be seen in the purple cartilage band through the middle of the femoral head. b) Conceptual femoral head growth plate illustrating key component locations and the portion of tissue used in this dissertation highlighted in red. The zones have been exaggerated for illustrative purposes and are not to scale.

The chondro-osseous system is established through activity of five cell types:

mesenchymal progenitor cells, chondrocytes, osteoblasts, osteocytes, and osteoclasts. Each of

these cells has a specialized task relating to chondro-osseous system development. Mesenchymal progenitor cells have relatively limited anabolic and catabolic activities focussed on cell adhesion to facilitate aggregation into cartilage condensations as templates for bone development (67). Mesenchymal progenitor cells differentiate into chondrocytes, osteoblasts, and osteocytes through the direction of several fibroblast growth factor (FGF) and bone morphogenic protein (BMP) ligand signals (67). Thus, mesenchymal progenitor cells serve primarily a stem cell-like role during early chondrocyte-osseous development.

The sry-box 9 (SOX9) transcription factor is the master regulator of chondrocyte fate, and is required to direct differentiation of mesenchymal progenitor cells into chondrocytes as well as directly driving expression of the key cartilage proteins collagen II, collagen XI, and aggrecan (68). The expression of BMP ligand may also be downstream of SOX9 regulation (69). Differentiated chondrocytes in the active growth plate are highly proliferative and productive of the key extracellular matrix (ECM) proteinaceous components listed in **Table 1.3** (among others) which directly drive skeletal growth through mechanical elongation of the bone epiphyses. The signalling factor environment in the growth plate stimulates chondrocytes to proliferate rapidly in linear stacks, with the organization of these cell stacks being of paramount importance to successful linear bone growth.

Cooperation between parathyroid hormone-related protein (PTHrP) and indian hedgehog (IHH) ligand signalling defines the height of the growth plate by regulating chondrocyte stack proliferative rate (70, 71). IHH promotes expression of PTHrP in the proliferative zone of the growth plate, while PTHrP feedback-inhibition of IHH in the pre-hypertrophic zone of the stacks allows chondrocyte hypertrophy (72). Hypertrophied chondrocytes are resorbed by osteoclasts and replaced by osteoblasts (discussed further below). Chondrocytes outside of the growth plate

are more sparsely distributed and primarily responsible for maintenance of cartilaginous tissues such as articular surfaces, vertebral discs, cardiac valves, tracheal rings, etc. where they are largely metabolically static; slowly synthesizing and maintaining proteinaceous components of the ECM.

Osteoblasts are derived from mesenchymal progenitor cells, similarly to chondrocytes, within the mesenchymal condensations (73). Osteoblast differentiation is promoted by IHH signaling, while fibroblast growth factor (FGF18) and BMP3 ligands block differentiation. IHH works in combination with runt-related transcription factor 2 (RUNX2) (71), which is the master regulator of osteoblast phenotype (74). Adherence of osteoblasts at the hypertrophic zone of the growth plate is mediated by ECM proteins such as periostin, and adhered osteoblasts deposit new ECM which can be mineralized in the presence of alkaline phosphatase activity (75). Osteoblast-produced ECM is mineralized to create primary spongiosa (calcified cartilage), and the process of ECM turnover and replacement is repeated once more to form the final mature lamellar bone matrix. After producing mature bone, osteoblasts differentiate into osteocytes entombed within the lamellar bone structure where they serve a maintenance role, becoming metabolically active in response to mechanical stimuli during bone damage or hormonal signals to facilitate local remodelling (76).

Osteoclast progenitors are of monocytic lineage and migrate to the growth plate from bone marrow. Once at the hypertrophic zone of the growth plate, the progenitors differentiate into osteoclasts and resorb cartilage ECM proteoglycans, collagen matrix, and hypertrophic chondrocytes (77) heavily dependent on cathepsin K (CTSK) enzyme activity and endosomal H+ secretion (75). Thus osteoclasts serve a purely catabolic role similar to the macrophage (78). Osteoclast differentiation in the growth plate is positively regulated by receptor activator of

nuclear factor kappa-B ligand (RANKL) along with macrophage colony stimulating factor (M-CSF), ligands which drive differentiation, and osteoprotegerin (OPG), which inhibits differentiation (78). Osteoclasts, like osteocytes, are likely activated by mechanical damage or hormonal signalling in bone for remodelling outside of normal bone growth (79).

1.6. Previously Reported Pathogenic Mechanisms of MPS Chondro-osseous Disease

Recent studies have begun to suggest that pathophysiology of chondro-osseous disease in the MPSs is more complex than simple GAG storage. It is unlikely that there will be any major single mechanism that explains all MPS chondro-osseous disease features, given the diversity of mechanisms summarized below. This makes it important to begin considering this potential network of pathogenic changes in a systems manner rather than each mechanism working in isolation. Individual pathogenic mechanisms, including those discussed below, likely represent parts of a more broad and dynamic network of pathogenic changes.

Differences in collagen fibril diameter and organization have been shown in the cornea of a single MPS III patient (80), while other studies have hinted at increases in collagen fibril diameter in MPS I, IV, and VI (80). Though data are presently limited, the implications of collagen fibril abnormalities are enormous for chondro-osseous tissue biomechanical function. Growth plate and articular cartilage is composed of 90-96% collagen II, while mature bone organic matrix is 95% collagen I (81). In these tissues, the collagen fibril network forms the foundation for other ECM proteins, acts as a scaffold for mineral crystal deposition, and defines mechanical stability. Further studies are needed to elaborate the details of collagen fibril abnormalities in the MPSs.

Oxidative stress markers in the plasma of MPS II patients have been shown to be elevated prior to ERT and reduced after 3 months of treatment (82). Filippon *et al.* predicted that increased oxidative stress may further destabilize lysosomes and lead to inflammation in the central nervous system and joints (82). Indications of oxidative stress have also been shown in the MPS I mouse model (83) and other lysosomal disease models.

GAGs are highly bioactive molecules known to interact with and mediate receptor-ligand interactions in the ECM (6) (84). Chondroitin sulfate (CS) has been demonstrated to modulate the activity of BMP and TGF β signaling *in vivo* (85), and HS-containing proteoglycans are known modulators of BMP ligand-receptor interaction (86). It has also been suggested indirectly that GAGs with abnormal sulfation patterns can disrupt growth factor signals (such as FGF ligands and cytokines) through inhibition of hematopoiesis *in vitro* (87). It is not yet clear how GAGs modulate these signalling pathways in the context of the MPSs (17), but the data compel further investigation into whether these signalling differences may be enacted through free GAG fragments, altered proteoglycans, or both. CS (chondroitin-4-sulfate specifically), has been suggested to reduce the activities of leukemia inhibitory factor (LIF) and Janus kinase 1 and 2 (JAK1 and JAK2), leading to reduced chondrocyte proliferation in the growth plate of MPS VII mice (88)

Osteoclasts are dependent on proteoglycan turnover for their normal function. Thus osteoclasts should conceptually be heavily-impacted by defective GAG catabolism. While the osteoblastic differentiation capacity of bone marrow stromal cells from MPS I patients has been shown to be normal, these progenitor cells have a higher than normal osteoclastic differentiation rate, which may be related to abnormal skeletal development (89). In the MPS VI feline model osteoclastic activity was observed as decreased (90). MPS VII osteoclasts have been shown to
attach poorly to the ECM in the mouse model, which Monroy *et al.* suggest may be based on some intrinsic property of the MPS VII osteoclasts rather than the ECM they adhere to (91). Increased osteoclastic differentiation potential may be compensatory in light of these observations.

A high prevalence of acetabular dysplasia and osteonecrosis of the femoral head in MPS I and III patients (92) (20) suggests a defect in both ossification and bone growth. As many as 66% of MPS III patients with acetabular dysplasia may also have osteonecrosis (20). Thus, the under-developed bone (resulting from defective ossification) may also be structurally compromised through improper maturation (leading to osteonecrosis). Furthermore, bone remodeling must be defective in the MPSs. It has been suggested that skeletal lesions which are not repaired in the MPSs (such as irregular skeletal morphology and cartilage inclusions within the mature bone matrix) indicate a lower capacity for remodelling in MPS I (17) (93). Normally, local bone remodelling would reduce or eliminate these lesions in healthy bone. Further support of abnormal ossification in the MPSs has been shown in MPS I, VI, and VII models (17, 89).

Activation of inflammatory cascades has been demonstrated in the articular cartilage and synovium of MPS VI and VII models, and has been postulated to involve stimulation of the toll like receptor 4 (TLR4) receptor pathway with subsequent increases of MMP1 and MMP13 leading to cartilage degradation (94). The synovial membrane and fluid have been studied in MPS VII feline, canine, and murine models, revealing prominent clear cells (GAG storage laden) as well as indications of inflammation such as infiltration of white cells and inflammatory signalling marker increases (tumor necrosis factor α (TNF α), receptor activator of nuclear factor ligand (RANKL), interleukin 1 β (IL-1 β), and cluster of differentiation 14 (CD14)) (95). Articular cartilage and underlying bone have also been grossly examined in MPS VII models and

have shown similar changes as seen in the synovium (16). Inflammation is clearly a factor in late-stage joint disease in the MPS VI and VII models. It is also clear, however, that inflammation is non-specific to arthropathy in the MPSs, as it is a late stage manifestation of many degenerative diseases. Critical studies in patients and earlier developmental time points may help to further clarify the role of inflammation in MPS chondro-osseous disease.

The mechanisms above represent the first evidence supporting complex pathogenic mechanisms occurring secondarily to defective GAG catabolism in the MPSs and provide rationale for further investigation. Clearly the effects of abnormal GAG catabolism are diverse, affecting cell signalling, bone growth and development, the extracellular space, and inflammation. However, the putative mechanisms discussed above, and those sought in this dissertation, are not likely pleiotropic effects of the mutated IDUA locus. Rather, they highlight the potential for downstream effects of defective GAG catabolism on processes regulating cellular homeostasis. These effects may be cell-specific but do not suggest direct pleiotropic effects from a single locus mutation. The lysosomal GAG catabolic role of the α -L-iduronidase enzyme is the same regardless of cell type.

This dissertation did not attempt to confirm the relevance of changes in collagen fibril diameter, increased oxidative stress markers, bone remodeling or late-stage inflammation, because these mechanisms have been identified in older animals. The early age studied here is unprecedented, and it is hypothesized that early chondro-osseous pathogenic mechanisms will be distinct from those activated during late disease.

1.7. Therapeutic Intervention in MPS I and Other MPSs

Enzyme replacement therapy (ERT) and hematopoietic stem cell transplant (HSCT) have a relatively limited impact on chondro-osseous disease in the MPSs (6). These approaches successfully mitigate and reverse organomegaly, coarse facial features, and possibly neurological defects. Skeletal deformities have been shown to continue to worsen in one study of MPS I patients who have undergone HSCT; with no apparent benefit to hip dysplasia, genu valgum, spine abnormalities, hand abnormalities, and joint range of motion (97). A review of 23 post-HSCT patients also noted that surgical intervention for joint mobility problems, acetabular dysplasia, and genu valgum was still necessary despite transplantation (98), indicating that HSCT was not having a significant impact on disease course.

However, five studies have shown positive clinical responses to HSCT and ERT: a one year examination of 45 MPS I patients treated with recombinant enzyme (Aldurazyme®, Genzyme) showed improvement of walking capacity (99). An additional 4 year follow up to this cohort showed increased shoulder flexion in 77% of patients (96). Some researchers suggest that joint mobility improves in most MPS I patients post HSCT (5). Improved joint ROM in response to ERT and HSCT has been further demonstrated recently in MPS I (21) and VI patients (22). These improvements ranged from increased joint flexion angle, to increased performance on a 12 minute walk test after HSCT and 18 months of intravenous ERT (22).

Assessment of joint improvement following HSCT or ERT in the MPSs is challenging when using measures such as ROM or a walk test. This is because these measures are subjective (21) and have large error margins. Furthermore, they do not reveal any specific insight into disease burden on specific tissues within the joint (such as cartilage, joint capsule, or ligaments). The standard urinary GAG measure is correlated with organ tissue response to enzyme but not

skeletal or joint response. MPS I patients undergoing ERT showed some improvement in the walk test and joint flexion as mentioned above, though this effect plateaued after 2 years despite continued ERT administration and urinary GAG stabilization (96). The studies discussed above have shown that, although HSCT and ERT can clearly have some positive impact on functional measurements of chondro-osseous disease in the MPSs, the effect is not robust.

ERT and HSCT intervention in MPS I, II, and VI patients leads to improved height profiles and growth rates approaching normal values when intervention is initiated early (as soon as within the first year of life) (21, 100, 101). One study has shown significant increases in growth rate, which exceeded average normal values, with ERT beginning between ages 5-11 in MPS I patients (102). It has also been shown that initiating ERT prior to the pubertal growth spurt has a greater impact on growth in MPS I patients than beginning ERT in adulthood (103), as would be expected. Early administration does not, however, allow for catch-up growth, thus initiating ERT as early as possible is now thought to be important for enabling nearly-normal growth rate during the limited window of time before growth naturally halts. Although growth resumes its natural rate for ~70% of MPS I patients after receiving ERT, final stature remains under average (96) due to lack of catch-up growth.

Population studies in humans have shown that the greatest bone growth rate is seen during embryonic development and immediate postnatal years (104) (105). Estimates from fetal growth measurements and the Center for Disease Control (CDC) growth chart slopes reveal that the growth rate during embryonic developmental week 12-19 is approximately 30% higher (~1.2cm/week) than that of children age 1-1.5 years (0.9cm/week). Shortly after the perinatal period the growth rate has declined significantly, before briefly spiking during puberty and rapid growth plate fusion (closure) in late adolescence (104, 105, 106). These observations highlight the need and importance of early study and initiation of therapy in MPS patients.

1.8. Rationale and Research Questions

Although GAG accumulation represents the primary event in the MPSs, the exact mechanisms by which this primary defect leads to chondro-osseous disease symptoms remain largely unknown. There have been no mechanisms identified which directly link defective GAG catabolism with the chondro-osseous phenotype. Evidence for secondary effects of defective GAG catabolism on other homeostatic pathways is now emerging in the MPSs (including oxidative stress, collagen fibril disorganization chondrocyte proliferation, inflammation, bone remodelling, and ossification discussed in section **1.6.**), indicating the role of these downstream effects as direct mediators of disease symptoms and therapeutic responsiveness (6). Studies of disease pathogenesis in other lysosomal storage disorders have also demonstrated the importance of secondary pathogenic cascades as direct mediators of disease symptoms. This important shift in understanding of the MPSs as complex disorders rather than simple "storage" diseases will have important implications for understanding the responsiveness of symptoms to therapeutics, and the identification of alternative therapeutic targets directed to key chondro-osseous disease features. The central hypothesis of this dissertation is that protein changes identified in the MPS I murine model growth plate will be indicative of pathogenic changes underlying chondro-osseous disease in MPS I.

It is not surprising that there have been no direct mechanisms discovered thus far, as pathogenesis must involve a complex network of changes which coordinately contribute to the disease phenotype. Such changes are difficult to discover by traditional biochemical and molecular methods used in MPS I research so far, and may require a systems approach rather

than targeted pathway analysis. Discovery of these mechanisms also require investigation of the chondro-osseous system at an early time point in development, because these tissues have very limited regenerative capacity as discussed in **1.3.** However previous studies of MPS I chondro-osseous disease have examined only late-stage disease, and thus have missed early events more proximate to disease pathogenesis.

The research objectives of this dissertation are as follows:

- Identify candidate protein changes in the MPS I mouse model growth plate which were hypothesized to drive or contribute to chondro-osseous disease.
- 2) Validate candidate protein changes and characterize functional implications.
- Understand the mechanisms relating observed changes to the MPS I chondro-osseous phenotype.

In order to discover the pathways underlying chondro-osseous disease in the murine model of MPS I, an unbiased proteomic approach was used first (as noted in the flow chart summarizing the experimental approach of this dissertation (Figure 1.3)). Any proteins which were observed to be significantly increased or decreased relative to the wildtype control mice were then re-examined through validation in new samples using targeted approaches. In this way, the candidate proteins identified would not be *a priori* biased or restricted to specific pathways or mechanisms. Protein candidates identified in the discovery phase also dictated proteins and pathways to be examined more closely for functional relation to MPS I pathogenesis.



Figure 1.3: Flow chart of experimental approaches used towards identifying a model of chondro-osseous disease pathogenesis in early MPS I development. Phases of the project are arranged in chronologic order from sample preparation, discovery of protein and RNA candidates, validation of candidates, and integration of protein and RNA results to propose pathogenic mechanisms occurring.

In addition to the protein studies, analysis of genome-wide mRNA expression followed by targeted qPCR enabled comparison of mRNA and protein level for examination of transcriptional vs translational control of protein deregulation. By studying RNA and protein changes, the strategy used here facilitated greater understanding of cellular disruption in response to the MPS I catabolic defect.

This dissertation explores the pathogenic mechanisms occurring early in the developing MPS I mouse model growth plate, and generates directions for future investigations. In light of the similarities and potential for overlapping pathogenic mechanisms between chondro-osseous disease in MPS I, II, IV, VI, VII, and IX (1.2.), these results will also be applicable to the other MPSs.

2. Methods

2.1. iTRAQ Proteomic Procedures 2.1.1. iTRAQ Proteomics

In order to identify potential candidate proteins representative of the pathways involved in the pathogenesis of MPS I skeletal disease, I used an unbiased proteome-wide differential mass spectrometry (MS) analysis of the developing growth plate as described in Section **1.8**.

Comparative quantitative proteomic techniques can be categorized as either labelled or label-free. Label-free quantitation involves acquiring MS/MS data for two or more samples using sequential runs, then comparing the spectral profiles of shared peptides between samples to look for relative increases or decreases (107). This can either be achieved by measuring the peak areas of precursor peptide liquid chromatography (LC) profiles for a given protein in both samples (with the MS fragment ion spectra being used only for identification) or by counting the fragment ion spectra for each peptide of a protein in both samples (also known as spectral counting) (107).

A study comparing the dynamic range (highest protein concentration to lowest protein concentration) of label-free to labeling quantitative proteomic strategies showed that label-free approaches can quantitate a much broader range of protein concentrations (2 times broader than labeling approaches) (107). Label-free approaches are also simpler to set up and less expensive than labelling approaches (107). However, there remain limitations which make label-free approaches less appropriate for applications such as this dissertation. Accuracy of quantitation in LC profile label-free approaches is lower due to the unpredictable linearity of concentration and run-to-run variability in peptide retention. In addition, elution profiles between LC runs are difficult to empirically correct (107). Technical replicates can be performed to amelioarate run-

to-run variability, but the number of runs required may quickly escalate to become infeasible. Furthermore, quantitation by LC profile requires high MS sampling density within each LC peak for better accuracy, at the cost of identifying fewer proteins (and thus fewer proteins quantitated) due to machine time allocation limitations (107). The spectral counting approach sacrifices accuracy of quantitation for increased proteome coverage over the LC profile approach through the requirement for an exclusion list (107). Detected peptides must be added to an exclusion list which prevents the MS from sampling them for a certain time interval to prevent peptides from more abundant proteins being over-represented due to their higher proportion in the sample.

Labelling approaches involve chemically modifying proteins or peptides in samples prior to analysis by LC-MS so that samples can be distinguished from each other and analyzed in a single MS run in order to eliminate the run-to-run variation (which is problematic in label-free approaches as discussed above). The technical variability present when running replicates of the same samples would of course remain (if replicates were done here), so the benefit of running all samples simultaneously is that relative quantitation is done between samples within a single run to reduce the run-to-run noise. While the setup for labelling approaches is more involved and costly than for label-free, there are benefits as discussed below.

Peptide labelling can be metabolic (occurring through endogenous enzymatic addition in live cells), chemical (addition of an artificial 'tag' to proteins or peptides), or via inclusion of "heavy" synthetic versions of the peptide(s) of interest (typically using ¹³C or ²H) for absolute quantitation in a single sample (107). Metabolic labelling is most commonly achieved by <u>S</u>table <u>I</u>sotopic <u>L</u>abelling by <u>A</u>mino acids in <u>C</u>ell culture (SILAC) wherein ¹³C-labelled amino acids are introduced to the culture medium of one sample such that all synthesized proteins in that sample can be distinguished on the basis of their predictably higher mass (107). For the most part,

SILAC has been used on relatively simple systems such as cell lines or yeast (107), as the primary limitation of SILAC is the need for expensive and time-consuming cell culture. Cell culture may also identify misleading proteome changes related to the artificial culture environment rather than the disease condition. The inability to multiplex several samples using SILAC or other metabolic labelling approaches may also be a limitation for some studies, such as this dissertation.

There are a number of options for chemical labelling, wherein a stable isotopic tag is enzymatically or chemically added in vitro to the peptides before or after protein digestion (107). The tags can be targeted to specific amino acids (such as derivatization of cysteine – <u>I</u>sotope-<u>Coded Affinity Tag (ICAT)</u>) or more general chemical groups (like primary amines of peptides – <u>Isotopic Tags for Relative and Absolute Quantification (iTRAQ) (1)</u>. There exist many variations on these two themes for which companies have made commercial options (107, 108), though iTRAQ from Applied Biosystems was the most widely used at the time of this work, with 1117 citations in PubMed at the time of writing. Importantly, all labelling methods allow multiplexing (from three samples with SILAC to as many as eight samples with iTRAQ), meaning that quantitation of multiple samples is done within a single MS run.

Absolute quantitation can be achieved by using a concentration vs. signal standard curve of heavy synthetic peptide equivalents, and comparing the real peptide quantitation result to the standard curve. This is known as <u>Single Reaction Monitoring (SRM) or Multiple Reaction</u> <u>Monitoring (MRM), depending on whether one or several peptides are quantitated (107).</u> Absolute quantities of the peptides of interest can then be compared between samples. Perhaps the greatest benefit of absolute quantitation is the precision and accuracy with which quantitation is measured, though the need for labelled standards for this very targeted and potentially

expensive approach is not well suited for interrogation of the whole proteome. iTRAQ tags have also been used for absolute quantitation, though synthetic peptide standards are still required (where the standard is labelled with a distinct tag and spiked into the sample) (109). In light of the major strengths and weaknesses of quantitative MS proteomic analysis approaches discussed above, iTRAQ labelling was selected for the present work, because of its high proteome coverage, high multiplex analysis capacity, and robust history of use.

iTRAQ tags are comprised of a reporter group based on *N*-methylpiperazine (109) with masses of 113, 114, 115, 116, 117, 118, 119, or 121 Da (m/z), and a balancer group with masses of 184, 186, 187, 188, 189, 190, 191, or 192 Da (m/z) (**Figure 2.1**). The balancer group is required to ensure that tags are isobaric when labelling peptides, so that a peptide labelled with the 113 tag and, the same peptide in another sample labelled with the 121 tag, will both perform similarly in the MS because the same mass (tag and reporter) has been added to both peptides before fragmentation. In such an example, the 113 Da tag would have a 192 Da balancer while the 121 Da tag would have a 184 Da balancer. An amine-reactive ester binds to the N-terminal amine group of peptides (**Figure 2.1**) (110).

It has been shown that the reported tag *N*-methylpiperazine masses are detected in a low complexity region of the fragment ion spectrum for typical proteomic MS experiments because these masses are not common products of peptide fragmentation (109). This means that quantitation of tags is more free from interference, and therefore be more accurate, than quantitation of peptide fragments directly. Quantitation of the reporter tag serves as a proxy for quantitation of the peptide because it is more straightforward and less likely to be obscured by interference. The original description of the iTRAQ method indicated that the balancer group was a carbonyl group whose mass was adjusted by incorporation of ¹³C, ¹⁵N, and ¹⁸O (109);

however ABI does not explicitly state how the balancer and reporter groups in modern iTRAQ tags are created. The reporter and balancer groups are combined so as to ensure peptides remain isobaric (i.e. have the same mass).



Figure 2.1: Diagram of iTRAQ tag composition from (110).

Upon collision-induced dissociation (CID), the reporter, balancer, and peptide fragments are separated so that the reporter mass can be read as an identifier for sample and the peptide fragment ions can be read to assign peptide identity (110). In this fragmentation, the charge shared between the balancer and the reporter goes to the reporter, so the neutral balancer is not detected by the MS (109). The amine reactive group of iTRAO tags binds to primary or secondary amines by an amide linkage reaction (109) including the N-termini of peptide products generated by trypsin digestion, side groups like the hydroxyl group on Lysine, and primary amines that may be in the solution (such as guanidinium, ammonium ions, and tris base) (109, 110). However, reaction with the amine group of Tyrosine has been shown to occur at a rate of <3%, while labelling of the hydroxyl group of Serine and Threonine was not seen under optimized conditions (109). As part of the labelling workflow, reduction and alkylation of proteins is used to prevent Cysteine reaction (109), and there is a Cysteine blocking reagent in the iTRAQ workflow (111). Thus, only N-terminal amines of peptides should be labelled under normal iTRAQ conditions, making iTRAQ reagents very predictable and specific. Care must be taken to exclude amines from preparatory workflows to ensure correct labelling of peptides. The

protein extraction method I used (112) included guanidinium (from guanidine-HCl) and tris base (from Tris-HCl), so the guanidine-HCl had to be removed prior to iTRAQ labelling through ethanol and acetone precipitation.

2.1.2. iTRAQ Limitations

iTRAQ quantitation provides excellent proteome coverage and depth because of the nonspecific amine labelling approach discussed above. iTRAQ tags can be added to any N-terminal tryptic peptide end such that potentially any protein could be measured. The primary limitations of iTRAQ relating to proteome coverage and depth are dynamic range of sampling in a complex mixture of proteins, and the sampling rate of the mass spectrometer.

The "dynamic range" of a proteomic sample refers to the range of low to high protein species concentrations. Some proteins are more ubiquitous and highly expressed than others, and this distribution will change according to tissue type. For example, in growth plate and articular cartilage, 70% of the total protein content is type II collagen, while >90% of the protein content of mature bone is type I collagen (113). Transcription factors and cytokines, on the other hand, are typically expressed at very low levels and may be hundreds of fold-lower than other proteins. Modern mass spectrometers are still only able to sample a dynamic range spanning 4 orders of magnitude (113), thus proteins on the high or low ends of expression may not be accurately sampled, limiting proteome coverage. Sample fractionation by liquid chromatography (LC) is a common practice for mass spectrometry proteomic experiments (LC-MS) for this reason, as using multiple fractions or elution windows can ameliorate the dynamic range issue, though at the cost of increased preparation and machine time requirements. Another common practice for avoiding dynamic range issues is depletion of high-abundance proteins (either by immunological methods or affinity binding), and is often seen in blood serum proteomic experiments (114), for example. Depletion is not an ideal strategy if high-abundance proteins could also be candidates of functional interest, as was the case in this dissertation, where the highly abundant collagens were important proteins to assay.

Mass spectrometers also have a finite capacity for sampling due to the time required for acquisition of peptides, isolation of the highest intensity species, and measurement of the selected target. The time between one acquisition and the next, referred to as "cycle time", has a large impact on proteome coverage – longer cycle times allow more sensitive measurement of acquired peptides but also mean that fewer total peptides can be measured. This issue is not specific to iTRAQ and applies to any proteomic analysis utilizing mass spectrometry. For these reasons, the iTRAQ experiments in this dissertation were not considered good measures of actual fold change in protein level between samples, but rather simply a discovery phase for generating candidates.

2.1.3. Protein Extraction from Bone Tissues

Extraction of high quality protein from chondro-osseous tissues is challenging due to the hard, relatively protein-sparse outer inorganic matrix (periosteum) which reduces exposure of cells to extraction buffers, the limited number of cells within these tissues, and dilution of the important metabolically active tissues (i.e. the growth plate) by less active tissues which are far more abundant. Specific examination of the growth plate was important to this dissertation, given this tissue's key role in defining the mature skeletal system, though access to this tissue was complicated by its location deep within the bone structure.

A report by Jiang et al described a proteomic analysis of bone tissue using a novel fourbuffer method of protein extraction (including an initial decalcification step followed by three distinct lysis buffers), in which they showed that this method efficiently extracts proteins suitable

for proteomic analyses (112). As there were few protein extraction protocols specialized towards chondro-osseous tissues in the literature, and in light of the successful identification of a large number of proteins by iTRAQ analysis using samples generated with Jiang *et al*'s 4 buffer approach (112), I elected to use this method. The strengths of this method included comprehensive homogenization and extraction of diverse protein species from difficult chondroosseous tissues and proven compatibility with iTRAQ analysis. Disadvantages of this method included the very long extraction time (taking several days per sample) and the complexity introduced by multiple extraction solutions (discussed further below in terms of protein distribution between extracts not being exclusive). In the original description of the 4 buffer approach (112), mechanical homogenization was not used; however this is a common procedure in protein extraction. In the protein extractions I performed, mechanical homogenization was introduced by growth plate micro-dissection prior to extraction (**2.1.4.**) and sonication during the 4-buffer extraction method.

2.1.4. Sample collection: Femoral Head Growth Plate Collection

After sacrifice of the mice, the hind limbs were dislocated at the femoral head joint by pulling the femoral head out of the socket at a 90° angle. This frequently resulted in a clean separation of the hind limb from the pelvic socket (acetabulum) although occasionally (particularly in older mice) the ligamentum teres separated the femoral head along the growth plate line into two pieces, which destroyed the ability to isolate un-contaminated growth plate tissue as it came into contact with blood, muscle, and synovial fluid. This was avoided by cutting the ligamentum teres with an ophthalmic scalpel. Femoral heads were embedded in Cryomatrix (Shandon), frozen on dry ice, and sectioned at 50µm (approximately 20 sections per femoral head on average). Growth plate cartilage was carefully dissected from each section on a glass

microscope slide using an ophthalmic scalpel under dissecting microscope and placed into the first protein extraction buffer (see section **2.1.5.**). A ~150µm region above and below the growth plate cartilage was also trimmed and included so as to sample the entire active growth plate region (**Figure 2.2**). All sections of both femoral heads were combined. The total time spent above 0°C for the femoral heads was 5 minutes, in addition to 7 minutes per section during mounting and micro-dissection.



Figure 2.2: Femoral head growth plate micro-dissection method showing progressive isolation of the growth plate zone (including a portion of the resting cartilage, proliferative zone, hypertrophic zone, apoptotic zone, and trabecular bone). The growth plate cartilage proper is false colored in yellow and the articular cartilage is false colored in red.

2.1.5. Protein Extraction

The femoral heads of 18 mice (9 MPS I and 9 wildtype) at each time point of three and five weeks were used for total protein extraction following the 4-buffer approach described by Jiang *et al.* (112). The buffer compositions and protocols specified by Jiang *et al.* were used: Buffer 1 - 0.18M HCl, 1 Roche Complete-Mini protease inhibitor cocktail tablet, Buffer 2 - 6M guanidine-HCl, 100mM Tris-HCl, one Roche Complete-Mini protease inhibitor cocktail tablet, pH7.4, Buffer 3 - 6M guanidine-HCl, 100mM HEPES, 0.5M tetrasodium EDTA, one Roche Complete-Mini protease inhibitor cocktail tablet, pH7.4, Buffer 4 - 6M HCl, one Roche Complete-Mini protease inhibitor cocktail tablet, pH7.4, Buffer 4 - 6M HCl, one Roche Complete-Mini protease inhibitor cocktail tablet, pH7.4, Buffer 4 - 6M HCl, one Roche Complete-Mini protease inhibitor cocktail tablet, pH7.4, Buffer 4 - 6M HCl, one Roche Complete-Mini protease inhibitor cocktail tablet, pH7.4, Buffer 4 - 6M HCl, one Roche Complete-Mini protease inhibitor cocktail tablet, pH7.4, Buffer 4 - 6M HCl, one Roche Complete-Mini protease inhibitor cocktail tablet, pH7.4, Buffer 4 - 6M HCl, one Roche Complete-Mini protease inhibitor cocktail tablet, pH7.4, Buffer 4 - 6M HCl, one Roche Complete-Mini protease inhibitor cocktail tablet, pH-1.2. The original method included a wash step of the residue after demineralization (112), which was excluded here to avoid any potential

sample loss. 1-2 hours after the addition of the second extraction buffer, the samples were pulse sonicated using a Fisher Scientific Sonic Dismembrator 500 system; 10 cycles of 7 seconds sonication and 30 seconds incubation on ice. Samples were then incubated at 4°C for the remainder of the 72 hour period.

2.1.6. Sample Replicate Number and Pooling

The minimum of 120µg total protein per sample required by the Genome BC Proteomics Facility for an iTRAQ proteomic experiment reflects the labelling efficiency of iTRAQ tags. Protein hydrolysis and amine labelling reactions are occurring simultaneously during iTRAQ sample preparation. Thus, if there is not enough protein, there will not be enough reactive peptide amines, and iTRAQ labels will be hydrolyzed, resulting in low efficiency (111). According to the manufacturer, 5-100µg protein is required, with the caveat that highercomplexity samples will require 100µg (111). Thus the minimum of 120µg allows for errors in concentration assessment and handling. Based on preliminary trials of protein extraction from the micro-dissected growth plate and cranial plate samples of interest, it was not possible to obtain enough protein from a single mouse to meet these thresholds, thus the lysates from 9 and 9 ++ mice were pooled together for iTRAQ labelling according to the strategy outlined in **Figure 2.3**.

Pooling has drawbacks, such as eliminating the possibility to assess individual variation (and the associated statistics which can be calculated from this analysis), as has been noted (115). It would have been ideal to use multiple smaller pools rather than a single larger one (115); however this was not possible. There have been no studies specifically investigating the differences between quantitating proteins in pools versus individual samples, though it has been shown that measuring the average mRNA expression quantity in a pool does yield a different

candidate list of differentially regulated genes when compared to the same samples analyzed individually (116) (117). The issue of losing individual variance data by pooling could potentially be addressed by follow-up validation experiments utilizing single samples rather than the pools (section **4.3.3.**).

Several options for normalizing the contribution of each individual sample were considered, but the decision was complicated by the four extracts produced by the extraction protocol (section 2.1.5). The first option was to pool all four extracts separately, resulting in four growth plate pools and four cranial plate pools for MPS I and wildtype mice (four pools for analysis per time point). The second option was to pool all 4 extracts per mouse but not pool the individual mice together (18 pools for analysis per time point). The final option was to combine the three lower complexity extracts (#1, 3, and 4) while making a separate pool for the more complex extract #2. It was expected that each extract would be largely mutually exclusive for the proteins species within due to differences in buffer composition as Jiang et al. claim (112), and so extracts were pooled separately (Figure 2.3). The logistics of analyzing ~18 samples per age group and the inability to achieve enough protein from the growth plates of a single mouse made the second option impossible. For the growth plate samples, option 1 did not allow each extract pool to meet the 300µg total protein target, but pooling extracts 1, 3, 4 together fulfilled all requirements and was the approach used. Option 1 was compatible with the cranial plate samples (due to their greater protein yields per extract), and so was used for this tissue as a result of the lower protein complexity than was seen in the growth plate samples (data not shown). Accordingly the growth plate samples were analyzed by a 4-plex iTRAQ run (MPS I extract 2, MPS I extract 134, +/+ extract 2, and +/+ extract 134) and the cranial plate samples by 8-plex

(MPS I extract 1, MPS I extract 2, MPS I extract 3, MPS I extract 4, +/+ extract 1, +/+ extract 2, +/+ extract 3, and +/+ extract 4) (Figure 2.3).



Figure 2.3: Pooling strategy for individual mice and protein extracts, for femoral head growth plate and cranial plate samples. Protein candidates were investigated by comparing like pools between mutant and WT groups. Individual mouse samples were extracted independently and pooled as depicted.

Prior to proteomic analysis, extract 2 was kept separate from 1, 3 and 4 but was pooled with extract 2 from the other mice of the same genotype. Extracts 1, 3, and 4 were also pooled together from mice of the same genotype. This resulted in 4 total sample pools for proteomic analysis. To ensure that individual mice contributed equally to the pool, a normalization factor had to be used. Typical pooling normalization factors used in the literature include total protein concentration, weight of tissue or cell count, and volume. However, for micro-dissected femoral head growth plates, none of these options was satisfactory. Weighing was impractical and difficult given the small size and the inability to reliably weigh an intact femoral head (which is

not a good measure of growth plate tissue regardless, due to its small size relative to the whole femoral head). Cell count normalization also was not possible because the micro-dissected tissue sections were too thick to count all cells, and the sections were not stained. It would have been possible to normalize by protein concentration if the concentration of every extract could have been measured prior to pooling, though in attempting this approach it was found that too much sample was consumed by the protein quantitation assay (Pierce Micro-BCA). The most biologically relevant normalizer was two femoral head units, because the variation between growth plate size (and thus relative contribution) between mice is expected to be smaller than the variation in results of the protein assay. Furthermore, the precision and specificity of the micro-dissection method used (**2.1.4.**) ensured that the whole growth plate was included and that variations in surrounding tissues did not alter the resulting protein. Thus the femoral head unit was an appropriate normalizer.

In this way, equal volumes of each extract from each individual were pooled as each sample was derived from two femoral heads. In principle this is the same as normalizing by volume, since the volume of extraction buffer used was identical for all samples. Volumes taken from extracts 1, 3, and 4 were equal between samples according to their relative protein content (estimated in **Table 2.2**) so that each extract contributed a similar quantity of protein. This was also the case for the volume per sample pooled for extract 2 (**Table 2.2**). Total protein concentration was measured in seven practice +/+ samples (**Table 2.2**) to estimate the expected protein yield from the 4-buffer method so that \geq 300 µg quantities of protein could be achieved in each pool for iTRAQ. Volumes used for pooling were established by estimating the minimum concentration of each extract (calculated from the data in **Table 2.2**) and calculating the volume

needed to include 1/9 of 300 μg from each. This volume was 76.6 μL per extract #2 and 139.7

 μ L per extract #1/3/4 (**Table 2.2**).

	Femoral Head Extracts			
	Estimated Concentration (µg/mL)			
	E1	E2	E3	E4
1	48.2	598.8	-	163.8
2	54.8	620.0	-	116.0
3	93.4	685.3	-	259.5
4	110.6	534.7	-	258.2
5	72.3	409.5	-	45.0
6	21.4	391.5	-	100.34
7	76.1	489.0	-	72.2
Average	68.1	532.7		145.0
Standard				
Deviation	29.6	109.785		86.0
% Difference	44	21		59

Table 2.1: Analysis of 7 +/+ mouse micro-dissected femoral head growth plate protein extracts (measured by BCA assay) to assess variation between individuals.

	F	emoral H	lead Extr	acts	
	Actual	Volume	Added to	Pool	(µL)
	E1	E2	E3	E4	
M 01	139.7	76.6	139.7		139.7
M 02	139.7	76.6	139.7		139.7
M 03	139.7	76.6	130		137
M 04	139.7	76.6	139.7		139.7
M 05	139.7	76.6	139.7		139.7
M 06	139.7	76.6	139.7		139.7
M 07	139.7	76.6	139.7		139.7
M 08	139.7	76.6	139.7		139.7
M 09	139.7	76.6	139.7		139.7
WT 01	85.0	76.6	85.0		85.0
WT 02	85.0	76.6	85.0		85.0
WT 03	85.0	76.6	85.0		85.0
WT 04	85.0	76.6	85.0		85.0
WT 05	139.7	76.6	139.7		139.7
WT 06	139.7	76.6	139.7		139.7
WT 07	139.7	76.6	139.7		139.7
WT 08	139.7	76.6	139.7		139.7
WT 09	139.7	76.6	139.7		139.7

 Table 2.2: Pooling volumes used to create the 5 week iTRAQ proteomic pools for the femoral head growth plates. E1, E2, E3, E4 indicate extract 1, extract 2, etc. M denotes MPS I mice while WT denotes wildtype mice.

In the case of the cranial plate samples, normalization by total protein concentration was the only option given the relatively crude isolation of the tissues compared to growth plate micro-dissection. The much higher protein yields from cranial plate samples also made sample loss per extract for the protein quantitation assay acceptable. Extracts 1, 2, 3, and 4 were kept as separate pools unlike the growth plate samples. For each cranial plate extract pool, the volume required to contribute 1/9 of 300µg (33.33µg) was calculated based on the protein concentration results of each sample (data not shown). The total protein concentration of each pool was assayed in triplicate to ensure precise quantitation for normalization of amounts for iTRAQ labelling. The final average concentration results and volumes used for labelling are shown in

Table 2.3.

r				
		Cranial Pla	ate Extracts	5
	Actua	I Volume A	dded to Po	ool (µL)
	E1	E2	E3	E4
M 01	31.8	13.1	38.6	29.8
M 02	28.1	10.8	36.2	26.6
M 03	43.1	18.7	30.6	23.6
M 04	48.6	16.4	35.6	23.6
M 05	80.8	12.4	23.6	35.7
M 06	79.2	26.8	64.0	37.3
M 07	84.6	15.8	47.0	36.4
M 08	28.8	27.3	25.1	28.7
M 09	29.3	12.2	22.9	23.9
WT 01	79.6	25.8	42.2	48.7
WT 02	33.1	29.9	40.7	54.0
WT 03	141.9	22.8	55.0	22.5
WT 04	94.9	21.0	88.0	21.6
WT 05	54.7	26.8	90.0	29.8
WT 06	100.0	25.0	51.1	56.4
WT 07	100.0	41.2	46.1	62.3
WT 08	98.3	24.1	90.0	63.9
WT 09	48.8	42.2	87.0	41.4

Table 2.3: Pooling volumes used to create the 5 week iTRAQ proteomic pools for the cranial plates.

Table 2.4: Final average total protein concentration of 5 and 3 week sample pools using the Micro-BCA assay, and volumes used for iTRAQ labelling to normalize for equal protein. "FH" denotes femoral head samples while "C" denotes cranial plate samples. E2 denotes the extract 2 pool, while E134 denotes the pool combining extracts 1, 3, and 4.

5 Week			
Sample	Concentration	Volume used	
MPS I FH E2	0.68ug/µL	146.1µL	
Wildtype FH E2	0.45ug/µL	224.6µL	
MPS I FH E134	0.40 ug/μL	250.6µL	
Wildtype FH E134	0.20ug/µL	496.9µL	
MPS I C E1	0.32ug/µL	316.6µL	
Wildtype C E1	0.35ug/µL	286.9µL	
MPS I C E2	0.65ug/µL	154.5μL	
Wildtype C E2	0.48ug/µL	207.6µL	
MPS I C E3	0.82ug/µL	122.0µL	
Wildtype C E3	0.96ug/µl	104.1µL	
MPS I C E4	1.1ug/µL	88.3µL	
Wildtype C E4	1.6ug/µL	62.2μL	
3 Week			
Sample	Concentration	Volume	
		used	
MPSIFH	0.68ug/µL	146.1µL	
Wildtype FH	0.45ug/µL	224.6µL	
MPSIC	0.40 ug/µL	250.6µL	
Wildtype C	0.20ug/µL	496.9µL	

2.1.7. Sex Ratios in Pooling

Initially balancing the ratio of male to female samples in the iTRAQ proteomic pools was not a concern, as there were no apparent differences between the femoral head growth plates of males and females at 3 and 5 weeks of age. It is known that the rate of skeletal growth and development is sexually-dimorphic, with female skeletal growth being slower and continuing for a shorter period of time due to the anti-chondroproliferative effect of estrogen (118, 119). It has also been shown that low level estrogens may have a positive effect on the early developing growth plate, but peak levels during puberty in females signal epiphyseal fusion (120). Puberty typically begins between six to eight weeks of age in female mice and six to seven weeks in male mice (121), thus at 3-5 weeks of age sexual dimorphism in growth plate development should not be a concern. Furthermore, no differences in growth plate morphology (i.e. height) were noted that would indicate sexual differences. Due to limitations of mouse numbers during sample collection, a neutral sex ratio of 1.0 was not possible, and certain pools were biased towards one sex or the other (**Table 2.5**). There was evidence that hinted at detectable differences in the level of cartilage oligomeric protein (COMP) and proline/arginine-rich end leucine-rich repeat protein (PRELP) between males and females used in the 5 week group (discussed in section **4.2.1.**). While these results suggested there may be differences in protein expression between male and female MPS I mice, the trend did not reach statistical significance. Due to the possibility of sexual dimorphism in protein expression, all validation experiments following iTRAQ used single sex pools (females only) to avoid any issues stemming from potential effects on protein concentration by sexual dimorphism.

	5 Week	
Sample Pool	Male Mice	Female Mice
MPS I FH	8	1
Wildtype FH	5	4
MPSIC	8	1
Wildtype C	5	4
	3 Week	
Sample Pool		
Sumple 1001		Female Mice
MPS I FH	5	Female Mice 4
MPS I FH Wildtype FH	5 7	Female Mice 4 2
MPS I FH Wildtype FH MPS I C	5 7 5	Female Mice 4 2 4

Table 2.5: Sex composition of protein 5 and 3 week pools used for iTRAQ analysis. "FH" denotes femoral head samples while "C" denotes cranial plate samples.

2.1.8. iTRAQ LC-MS

LC-MS/MS analysis was performed using an integrated Famos autosampler, SwitchosII switching pump, and UltiMate micro pump (LC Packings) system with an Hybrid Quadrupole-TOF LC/MS/MS Mass Spectrometer (QStar Pulsar i) equipped with a nano-electrospray ionization source (Proxeon) and fitted with a 10µm fused-silica emitter tip (New Objective, Woburn, MA). Chromatographic separation was achieved on a 75µm x 15 cm C18 PepMap Nano LC column (3µm, 100Å, LC Packings) and a 300µm x 5mm C18 PepMap guard column (5µm, 100Å, LC Packings) in place. The mobile phase (solvent A) consisted of water/acetonitrile (98:2 (v/v)) with 0.05% formic acid for sample injection and equilibration on the guard column at a flow rate of 100µL/min. A linear gradient was created upon switching the trapping column inline by mixing with solvent B which consisted of acetonitrile/water (98:2 (v/v) with 0.05% formic acid and the flow rate was reduced to 200nL/min for high resolution chromatography and introduction into the mass spectrometer. Samples were brought up to $100\mu L$ with 5% acetonitrile (ACN) and 3% formic acid (FA) and transferred to autosampler vials (LC Packings). 20uL of sample were injected in 95% solvent A and allowed to equilibrate on the trapping column for 10 minutes. Upon switching in line with the MS, a linear gradient from 95% to 40% solvent A was developed for 100 minutes, and in the following 5 minutes the composition of mobile phase was decreased to 20% A before increasing to 95% A for a 15 minute column re-equilibration prior to the next sample injection. MS data were acquired automatically using Analyst QS 1.0 software Service Pack 8(ABI MDS SCIEX).

An information-dependent acquisition method consisting of a 1 second TOFMS survey scan of mass range 400-1200 amu and two 2.5 second product ion scans of mass range 100-1500 amu. was used. The two most intense peaks over 20 counts with charge state 2-5 were selected

for fragmentation, and a 6 amu window was used to prevent the peaks from the same isotopic cluster from being fragmented again. Once an ion was selected for MS/MS fragmentation it was put on an exclude list for 180 seconds. Curtain gas was set at 23, nitrogen was used as the collision gas and the ionization tip voltage used was 2700V.

If the observed A_{214} was greater than 0.1 for any fraction collected during the strong cation exchange (SCX), a 2.5 hour gradient (95-50% solvent A) was used to compensate for the higher peptide concentration in that fraction.

2.1.9. ProteinPilot Software

ProteinPilot Version 4.0.8085 from AB-SCIEX was utilized for the primary visualization of the raw MS data. This software was chosen for its widespread usage with regards to MS data from ABI MS machines, and because it is part of the Genome BC Proteomics Facility's standard iTRAQ data analysis pipeline. MS file interpretation, *in silico* database spectra generation, database matching (by the Paragon and ProGroup algorithms within Protein Pilot[™] respectively), and peptide quantitation were carried out within ProteinPilot version 3.0 at the Genome BC facility, while a read-only version of ProteinPilot 4.0 was used to view the analysis result files.

Although the specifics of the Paragon and ProGroup algorithms are beyond the scope of this thesis, it is worth noting one feature about how the experimental peptides are matched to the *in silico* database. There are a number of peptide modifications and missed trypsin cleavage events that could increase the complexity of peptide spectral results. A single peptide should normally be matched to a single peptide in the database, but if the peptide had a modification that altered its MS profile this may not be the case unless these variables are accounted for. The Paragon algorithm does account for many common modifications and missed cleavage events

but only applies these modifications under limited circumstances (122), to allow for modifications while maintaining reasonable computational complexity. ProteinPilot also, by default, corrects for biases introduced by label-label contamination and isotopic interferences, which have been shown to be problematic (123).

When using the database matching approach for peptide identification (such as Paragon), the protein database used for identifying peptides and proteins from MS data is important in that the resulting analysis can only be as comprehensive as the database is. There are several available large-scale protein databases that can be utilized for this purpose, each with largely similar but some unique content. The Celera Discovery System (124), Mouse Genome Database (125), and Universal Protein Resource (UniProt) (126) are a few examples. For this dissertation, the UniProt database was used for all matching purposes due to its superiority in terms of total proteins contained. At the time of primary analysis and matching, the UniProt TrEMBL database used contained 15,833,688 protein entries (for all species).

The UniProt database is divided into two sections; TrEMBL and SwissProt. The SwissProt is a collection of exclusively manually-curated proteins that generally have the most detailed and well-validated information, while the TrEMBL dataset is a combination of both curated proteins from SwissProt and those that have been added by users but have not been manually reviewed (126). Although the SwissProt dataset is more accurate, in the interest of comprehensiveness the TrEMBL dataset was used for all analyses in this work. In this way, the larger dataset of TrEMBL can be used for maximum identification while the higher-quality data of SwissProt can be manually applied to identify candidates during candidate analysis. Since the TrEMBL dataset is so large, peptide assignment and protein identification can sometimes be more challenging for ProteinPilot and can lead to inappropriate matching results (assigning a

mouse protein as a homologous drosophila protein, for example, or confusing two homologous mouse proteins).

Despite these challenges introduced by using a non-curated dataset like TrEMBL, the increased coverage of TrEMBL makes it a better choice for discovery projects. For example, it was common in these results to see proteins identified as belonging to a non-mouse species. While these proteins typically had homologous versions in the mouse, clearly the level of homology between species across the discovered peptides is sufficiently high so as to make species distinction impossible based on the few peptides identified in by iTRAQ alone. These proteins also often had few peptides assigned to them, further increasing identification difficulty. In such cases, if the these proteins were included as candidates, the identified proteins were inferred to have originated from the mouse tissues, and the mouse homolog entry in the UniProt database was used for later analyses. There were also cases of proteins being identified as human or chimpanzee (such as keratin). These may be similar cases of homology as above, but are also potentially human contamination from sample processing, as this is a well-known issue in MS experiments.

Another group of potentially ambiguous proteins in this dataset are those which match to protein fragments from large scale proteomics experiments and/or predicted peptides from *in silico* genome analyses. In some cases, these peptides and proteins cannot be matched to any known mouse proteins and thus were excluded from analysis.

2.1.10. The iTRAQ Ratio p-value

The iTRAQ p-value calculated by ProteinPilot is designed to be the statistical measure of significance for changes reported in proteins between samples. Functionally, it tests the hypothesis that the ratio of a protein in sample 1 to sample 2 is not 1.0 according to the equation:

Weighted Average of Log Ratios =
$$\frac{\sum_{i=1}^{n} w_i \times x_i}{\sum_{i=1}^{n} w_i}$$
(127)

The average ratio for each protein (where average refers to the ratios between component peptides) is used to establish the Student t-value (127). Even if the data are background corrected by ProteinPilot (discussed in **2.1.11.**), the p-value calculation remains the same except that the standard error of the average log ratios is then dictated by the number of peptides for the protein rather than determined empirically. For example, if the protein has <4 peptides, the standard error is set as 0.17 (127). If there are >4 peptides, the standard error is automatically calculated (127) and will likely be <0.17. Thus in background corrected data the p-value becomes most strongly influenced by the number of peptides for each protein, where more peptides lead to a lower (more significant) p-value because the standard error will be lower. In this scenario, the capacity to determine significance by p-value is dependent on the number of peptides.

The p-value is not affected by including Bias Correction (discussed in **3.3.**) because all ratios are changed by the same factor independently of peptide number. Since the p-value calculation is done for hundreds or thousands of proteins, multiple hypothesis testing must also be implemented to adjust the cut-off for p-value significance as well. This was achieved by the Bonferroni correction (128).

An important factor which was considered when using the iTRAQ p-value is ratio agreement between peptides of the same protein. For example, if a protein has ten peptides identified, and five peptide ratios indicate a decrease between samples while the other five indicate an increase, this should increase (reduce the statistical significance) of the p-value of the ratio This consideration is intrinsically included in the p-value calculation because it is calculated from the average ratio across peptides of a protein, so greater variance of peptides within a protein predictably leads to a higher (less significant) p-value due to the uncertainty.

However, proteins with significant iTRAQ p-values may still have peptides which disagree, and these were assessed further while refining the results. ProteinPilot does not account for the potential effect protein isoforms may have on assignment of peptides to proteins and the iTRAQ ratio p-value. For example, if different isoforms of a protein are differentially changed between the mutant and control samples, then the isoform-specific peptides would report a different ratio than those which are shared by all isoforms (129). While this is a possibility with potentially very important and relevant implications for understanding disease mechanisms, it remains difficult to identify this phenomenon in proteomics results due to lack of isoform functional knowledge in many cases. This difficulty is exacerbated by challenges in assigning shared peptides to one isoform or another, and the limited isoform information contained in modern protein and genomic databases in most cases.

It has been noted that the iTRAQ p-value calculated by Protein Pilot is "not appropriate in all experimental designs" (127) although the circumstances in which iTRAQ p-value may not be appropriate have not been clearly described. Most studies simply adopt the standard $p \le 0.05$ (or a multiple hypothesis testing corrected variation in a few cases) to determine the statistical significance of an iTRAQ result. In this thesis, iTRAQ p-value was used as one selection criterion for identifying potential candidates, but it was not considered strong statistical evidence in light of this uncertainty regarding when the ProteinPilot p-value is suitable.

2.1.11. Background Correction

Background correction is a function within ProteinPilot which attempts to control for background ion noise in the MS results, which can occur when >1 labelled peptide is fragmented

at once because it has a very similar m/z as another peptide. Since these "background" peptides are expected to occur uniformly between samples of the same source (e.g. the 5 week growth plates of MPS I and wildtype mice), the presence of background peptides will tend to "bias ratios towards unity" (127). Background peptides are also expected to be at lower intensity than the target peptides. Background Correction "models the distribution of the background signal across multiple spectra from the same protein and determines the parameters of this distribution in order to correct each individual spectrum." (127) Once a spectrum has been identified by the software as "background", it is removed from all spectra assigned to each protein. Since establishment of a spectrum as background requires multiple data points to show that it occurs consistently, at least 6 spectral measurements for a peptide are required for this correction factor to be accurate (127). Furthermore, since all protein candidates examined must first pass the FDR cut off value filter (section 3.3.), they satisfy this requirement by having an Unused ProtScore >4 (equating to at least 2 unique peptides per protein) with >3 observations per peptide. Use of background correction in iTRAQ experiments is the standard protocol because of its utility in complex samples containing many proteins, and was used for all iTRAQ protein data reported in this dissertation.

2.1.12. Bias Correction

Bias correction within ProteinPilot is intended to control for differences in iTRAQ labelling efficiency between samples (primarily from technical errors during reagent use) (127). Bias correction hypothesizes that most proteins in a differential proteomic experiment will not change between samples such that the overall average ratio between samples should be close to 1.0. The algorithm calculates the mean of all protein ratios in the experiment and attempts to adjust the result to 1.0 by dividing all protein ratios by the appropriate factor (127). Bias correction was used for my data.

2.1.13. Extreme iTRAQ Ratios

Extreme iTRAQ ratios are defined as peptides where the reported value is 100 or 0.01. These values are not actually informative of the true fold change of the protein or peptide associated with it, as they are artificial cut offs imposed by ProteinPilot so that very large and very small ratios (which may be more than 100 or less than 0.01) can still be used in calculating the average protein fold change (127). Cases of peptide absence in one sample would also be reported as an extreme ratio (127). This is partially because iTRAQ quantitation is less accurate with increasing fold change such that the accuracy of >10-fold values is low. There were a number of cases in the iTRAQ data where peptides had extreme ratios (>100 or <0.01), and it was considered whether they should be manually excluded from protein average ratio calculation due to their inaccuracy and potential to bias the protein average.

Instances of extreme ratios occurred in two forms; either 1-2 peptides of several peptides identified for a protein showed these ratios, or most of the peptides for a protein showed these ratios. After examining the protein average fold change before and after manually excluding extreme ratio peptides, in both cases removing peptides was not deemed useful. This was because excluding one extreme peptide from a protein with many non-extreme peptides changed the protein ratio very little. However, many extreme ratios within a protein could represent biologically-meaningful information about the protein. Thus for all ratios reported in this work, extreme ratio peptides were included in the average protein ratios. These extreme ratios were examined more closely on a case-by-case basis during later candidate list filtering (section **3.6.**).

2.1.14. False Discover Rate (FDR) Analysis of iTRAQ Data

Confidence of protein identity within the ProteinPilot software is reported through three metrics: Protein confidence percentage, ProtScore, and Unused ProtScore. The protein confidence is the error margin of protein identity based on the peptide information available from the iTRAQ results. A greater number of high-confidence peptides identified for a protein will increase its confidence score. The ProtScore is a logarithmic transformation of the percent confidence according to ProtScore = $-\log(1 - (Percent Confidence/100))$ (127). The Unused ProtScore is calculated in the same manner but is distinct from the ProtScore as the Unused ProtScore considers only peptides matching a single protein while the ProtScore allows peptides shared between proteins to count towards protein confidence (127). Thus protein percent confidence is equivalent to the ProtScore, while the Unused ProtScore is a more stringent measure of protein confidence. For example, in cases where a protein is supported only by evidence from peptides shared with other proteins, it cannot be confidently stated that this protein was detected, thus the Unused ProtScore may be a more accurate account of identified proteins (127). Shared peptides are assigned to the single most confidently identified protein and are excluded from subsequent proteins under the Unused ProtScore analysis.

False discovery rate analysis was carried out by the ProteinPilot software and is useful for establishing cut-off values in protein identity confidence.

2.2. Pathway Analysis of Candidates

Commonly cited online bioinformatics software solutions are available to perform these network analyses on input proteins or genes of interest. One of these tools is <u>R</u>egulatory

<u>Sequence Analysis Tools (RSAT) (130)</u>, which searches for potential regulatory sequence motifs in genomic sequences. Another, TFSearch (131), identifies short sequences matching transcription factor consensus binding sequences. <u>Multi Experiment Matrix (MEM) (132)</u> discovers evidence of transcriptional co-regulation between genes based on microarray results compiled in the NCBI Gene Expression Omnibus (GEO). The workflow I used for pathway analysis employed each of these tools. The goal of this workflow was to generate hypotheses about how the candidates identified by iTRAQ may fit into networks relevant to pathogenic mechanisms in MPS I chondro-osseous disease.

MEM analysis was used first to identify potential co-regulatory relationships between candidate genes. In MEM analysis, genes for the proteins of interest were used as queries against all experiments in the NCBI GEO database using the Affymetrix mouse 430 2.0 bead array chip platform. At the time of analysis, this included >900 experiments. MEM systematically examines all included array datasets for genes that show evidence of co-regulation (i.e. change together under experimental conditions) (132). P-values for significance of matching co-expression were calculated automatically according to how strong the evidence of co-regulation was, and a manual p-value threshold of 1.0×10^{-9} was applied to allow for p < 0.01 after Bonferroni Correction.

Once genes with evidence of co-regulation from MEM were identified, RSAT was used to search the genomic sequences around these genes for shared potential regulatory sequences. RSAT analysis was performed on genomic sequence from 10KB upstream of the transcriptional start site and 10KB downstream of the transcriptional termination site of each candidate gene from MEM analysis. Genomic sequences were retrieved from the NCBI GenBank mouse reference genome (build 37.1). Sequences were compiled into a FASTA batch file for "oligo-
analysis" RSAT input. Oligo length was set to 4, 5, 6, 7, and 8 nt and manual expected frequency tables were created for each gene as there was no pre-existing frequency table for mouse sequences available. The "feature map" was used for each query gene to discover shared sequence patterns.

After investigation of potential shared regulatory sequences by RSAT in the candidate genes from MEM, these sequences were examined for matching to known transcription factor binding consensus sites. Transcription factor binding site analysis was performed on *all* identified enriched sequences found in candidate genes from RSAT analysis. TFSearch was used to identify transcription factors which may bind to these putative regulatory sequences based on known consensus transcription factor binding sites. The JASPAR database of transcription factors was also manually searched for matching consensus sequences.

2.3. Western Blot Analysis

Total protein extracted from the mouse femoral head growth plate samples described above was analyzed by Western blot as individual mouse and extract samples, unlike the pools used for iTRAQ analysis. 5-20µg (depending on the remaining amount after iTRAQ and the concentration of the extract) of each sample was separated on pre-cast 7.5% polyacrylamide gels (Bio-Rad) and transferred onto the polyvinyl diflouride (PVDF) membrane (Millipore Imobilon® P). Detection was achieved using primary polyclonal antibodies against COMP (Thbs5), PRELP (Prelp), and perlecan (Hspg2) (Santa Cruz) with appropriate secondary antibodies conjugated to horseradish peroxidase (HRP) (Santa Cruz). HRP activity was detected with a chemiluminescent substrate (Pierce SuperSignal West Pico chemiluminescent reagent kit). Blots were photographed and analyzed using ImageJ (NIH) to quantify intensity.

2.4. Immunohistochemical Analysis

Five week old mouse femoral heads were fixed in 4% paraformaldehyde (PFA) for 24 hours and decalcified in 10% EDTA (pH7) for 7 days before dehydration, embedding, and sectioning at 5 µm. Sections were taken at 40 µm intervals through the entire femoral head. Immunohistochemical interrogation of several key extracellular matrix proteins (perlecan, fibronectin 1, hyaluronan and proteoglycan link protein, matrillin 1, matrillin 3, thrombospondin 1, cartilage oligomeric protein, PRELP, and chondroadherin) was done on sagittal sections of MPS I and wildtype femoral heads. DAPI staining was used as a control for cellularity (i.e. to establish that the plane of sectioning and amount of cellular content was similar between MPS I and wildtype comparative samples). Primary antibodies were applied to sections individually and labeled with fluorescent tagged secondary antibodies.

2.5. MRM Panel

As noted in the original publication of this method (133), proteotypic peptides were selected from the iTRAQ data for MRM analysis. Peptides with extreme iTRAQ ratios (i.e. 0.01 or 100) were excluded. Synthetic peptides for each were used to empirically select 5-10 transitions per peptide and determine optimal declustering potential (DP), collision energy (CE), and escape potential (CXP) voltage settings of the Triple Quad LC/MS 6490 MS (Agilent) for each peptide. A pool of all synthetic peptides and a trial protein sample from micro-dissected wildtype femoral head growth plate cartilage were used to confirm detection and optimization of all peptides. All protein extracts from the 4 buffer method (section **2.1.5.**) (#1, 2, 3, and 4) were pooled together prior to analysis. Samples were precipitated in 80% acetone overnight at -20°C before reduction, alkylation, and tryptic digestion.

2.6. Modulation of MPS I Chondro-osseous Phenotype by Exercise

2.6.1. Mouse Exercise

34wk old wildtype and MPS I male mice were acclimated to running on a Model 80800A Mouse Exercise Walking Wheel System (Lafayette Instruments) at a slow speed increasing gently over 7 days from 30 to 8 seconds per revolution daily for 1hr. The mice then ran daily for 1hr at 8 sec/rev for 30 days before sacrifice 2 days after the last run. Whole knee joints were collected and fixed for histologic analysis (section **2.6.2**).

2.6.2. Measurement of Knee Joint Arthropathy

Mouse hind limbs were detached at the hip, muscle tissue trimmed from both knee joint capsules, and whole joints placed in 1mL of PFA. Whole knees joints were fixed for 24hrs at 4°C before being transferred to 1mL of 10% EDTA (pH 7.5) and demineralized for 7 days. Samples were dehydrated and embedded in paraffin and knee joints serially sectioned frontally at 5µm in 100µm intervals through the entire tissue. Slides were heated at 50°C for 10min before deparaffinising and rehydrating, followed by staining with H&E or Safranin O & Fast Green. Stained serial sections were imaged on a Zeiss Axiovert 200M microscope at 500X magnification.

The area of the tibial and femoral articular surfaces was defined by hand and the area and grey value (stain density) calculated using ImageJ® software. The values were calculated in 4 quadrants of the knee sections to reflect the distinct regions of the joint: medial tibia, lateral tibia, medial femur, and lateral femur. Comparisons between animals and between knees of the same animal were only made between like quadrants (medial: medial and lateral: lateral) to prevent joint loading confounding factors. Average grey value and area were averaged across all sections

of a single knee joint, and then averaged between knee joints of each mouse. The p-value of the difference between averages for each mouse was calculated using a standard T-test.

2.7. RNA Expression Studies

2.7.1. RNA Extraction

Mouse femoral heads were collected after sacrifice and immediately embedded in OCT medium and frozen by suspension on liquid nitrogen. 50µm cryosections were attached to Superfrost Plus microscope slides. Slides were then quickly re-frozen on dry ice and stored until growth plate micro-dissection. Micro-dissection was as per section **2.1.4.**, but sections were thawed and dissected in a drop of RNAlater (Ambion) solution.

Micro-dissected growth plates for both femoral heads per mouse were collected in TRIzol on ice and disrupted by grinding with Pellet Pestles (Kontes) for three min. RNA was extracted according to the manufacturer's instructions. The aqueous phase was loaded onto RNeasy (Qiagen) columns for clean-up and DNase digestion. RNA was eluted in 70µL RNase-free water (minimum required volume) twice and evaporated by vacuum centrifuge down to 30µL.

2.7.2. Illumina Hybridization Chip

Purified RNA from six MPS I and six wildtype five week old mice was pooled into 4 samples (three mice per pool) and analyzed by Agilent RNA Nano chip BioAnalyzer to ensure integrity (RIN >7.0). RNA concentration was measured by NanoDrop (Thermo Scientific) for normalization of RNA loading. 200ng of total RNA from each pool was reverse-transcribed into cDNA using the RiboGreen® kit. The concentration of cDNA was measured by NanoDrop and used to normalize loading of the chip. Loading and analysis of the chips were according to the

manufacturer's instructions (134). The arrays were scanned on an HiScanSQ (Illumina) and data processed using GenomeStudio (Illumina) software (average normalization, subtract background, Illumina Custom error model, Benjamini-Hochberg multiple hypothesis testing correction). The exported probe profile tables were also interrogated with a custom R code and compared to the GenomeStudio results.

2.7.3. qPCR

531ng of purified RNA from each of four MPS I and four wildtype micro-dissected femoral head growth plate pairs was used for cDNA synthesis using the RT^2 First Strand kit (Qiagen). A custom 96-well array of qPCR primer pairs for 48 target genes and two housekeeping gene controls (*Acntb* and *Gapdh*) was produced through SABiosciences (Qiagen) RT^2 Profiler system (**Table 5.4**). Quantitation of SYBR signal was carried out on an ABI7500 (Applied BioSystems) thermal cycler and the $\Delta\Delta$ Ct method for normalization and fold-change calculation according to the manufacturer's instructions (135).

Gene	SABiosciences RT ²
Symbol	Catalog Number
Adam12	PPM28766
Adamts12	PPM34009
Adamts4	PPM26398
Alpl	PPM03155
Aspn	PPM35565
Bgn	PPM03187
Chad	PPM25471
Col10a1	PPM05135
Col11a1	PPM05136
Col11a2	PPM26047
Col12a1	PPM05127
Col1a1	PPM03845
Col1a2	PPM04448
Col2a1	PPM03184
Col9a1	PPM03161
Col9a2	PPM05124
Comp	PPM29133

Table 2.6: qPCI	R gene	targets and	SABiosciences	cataloged	primer	nairs
1 abic 2.0. qi Ci	v sene	tai sets and	Diosciences	catalogeu	primer	pans

Gene	SABiosciences RT ²		
Symbol	Catalog Number		
Dcn	PPM03172		
Ehd1	PPM27752		
Ерус	PPM31628		
Fmod	PPM37524		
HapIn4	PPM31708		
lhh	PPM24820		
Lum	PPM25962		
Matn1	PPM36440		
Matn3	PPM28975		
Mmp13	PPM03675		
Mmp3	PPM03673		
Nfkb1	PPM02930		
Ogn	PPM25080		
Omd	PPM31464		
P4ha1	PPM05622		
P4ha2	PPM25006		
Prelp	PPM33480		
Rab32	PPM27884		
Tnfrsf11a	PPM03749		
Tnfsf11	PPM03047		
Smad3	PPM04461		
Smad5	PPM03070		
Sox9	PPM05134		
Sparc	PPM03651		
Timp1	PPM03693		
Tlr4	PPM04207		
Gapdh	PPM02946		
Actb	PPM02945		

3. Proteomics Results Analysis

3.1. Introduction

Evidence of secondary effects downstream of defective GAG catabolism on cellular homeostatic pathways is now emerging in the MPSs (discussed in section **1.6.**). Such changes are difficult to discover by traditional biochemical and molecular methods used in MPS I research so far, and may require a systems approach rather than targeted pathway analysis. In order to discover the pathways underlying chondro-osseous disease in the murine model of MPS I, an unbiased proteomic approach was used first (as noted in the flow chart summarizing the experimental approach of this dissertation (**Figure 1.3**)). In this way, the candidate proteins identified would not be *a priori* biased or restricted to specific pathways or mechanisms. Protein candidates identified in the discovery phase also dictated proteins and pathways to be examined more closely for a functional relation to MPS I pathogenesis.

Mass spectrometry (MS) data analysis of iTRAQ results entailed 3 primary objectives: *1*. Read and graphically present the fragmentation ion spectra from peptides, *2*. Assign peptide identity based on fragment ion spectra, and *3*. determine the quantity of the parent peptide by comparing to the same peptide in another sample (relative quantitation).

Fragmentation ion spectra analysis entails interpretation of the MS data output files as graphical spectra which can be viewed by the user and queried by software to obtain quantitative information. The parent peptide quantity is determined by integrating the area under the curve of the fragment ion spectra. It is beyond the scope of this thesis to discuss the algorithms used to achieve such interpretation in detail, though the MS output file format will vary depending on the MS machine with which it was produced. Often these formats will dictate the software that can be used for analysis. In this case, ProteinPilot software was required for data interpretation.

The second objective entailed automated matching of spectra profiles obtained from the MS data to predicted MS spectra from all peptides in a protein database. Possible peptides are generated *in silico* from the database by using the known trypsin cleavage sites (where trypsin cleaves C-teminal to K or R residues in an amino acid chain). The predicted mass and charge of the in silico peptides are calculated based on the mass and charge of each amino acid in the peptide. Easily detected fragment ions under collision-induced dissociation (CID) for each peptide are predicted, which usually entails doubly charged fragments, as single and triplely charged fragments are not easily identified by currently available MS instruments. The experimental MS spectra are then compared, one at a time, against every spectrum in the in silico database to find the best match. Most often the spectral match is required to be highly similar, meaning that many experimental spectra may not be matched even if their equivalent peptides are actually in the protein database. Non-matching can result from covalent modifications, loss of chemical groups, low quality spectra, or interference to the peptide spectra by other peptides or contaminants that occupy the same mass to charge ratio (m/z), to name a few possibilities. Peptides may match several parent proteins, so the best match must be chosen and the peptide sequence and parent protein information assigned to the experimental peptide.

Quantitation of peptide intensity (ions detected per unit time) typically involves combining the counts of each fragment ion (distinct m/z value) for each peptide to calculate a total intensity score for the peptide. Each peptide intensity score can then be integrated with others from a parent protein to estimate the quantity of the protein in the sample. In the case of iTRAQ labelling, as mentioned in section **2.1.1.**, quantitation of peptides uses the iTRAQ tag as a proxy for peptide concentration. In this case, the parent peptide will have one fragment ion with the iTRAQ tag so only that ion is used for quantitation. However, a well-represented protein

will have multiple peptides with iTRAQ tags, and each can typically be used to calculate the average protein intensity reported by all of the peptides of that protein. Comparison of intensities between samples typically occurs at the peptide level and is then summarized at the protein level, depending on the analysis approach taken.

3.2. Protein Confidence and False Discovery Rate (FDR) Analysis

False discovery rate analysis is carried out by the ProteinPilot software and is useful for establishing cut-off values in protein identity confidence. For example, in order to generate a list of proteins that has a 5% FDR overall (global), the reported protein confidence score cut-off for each protein would be >99.023% (corresponding to an Unused ProtScore of >2 and 620 proteins identified in the growth plate 5 week sample) (**Figure 3.1**). If a more stringent list of protein will be >99.998% (corresponding to an Unused ProtScore of >3.7 and 423 identified proteins in the growth plate at 5 weeks. To be more stringent, FDR analysis dictated that only reported proteins with an Unused ProtScore of \geq 4.0 would be considered as being true for this experiment. Similar examination of the FDR analysis reports for the 5 week cranial plate and 3 week growth plate/cranial plate experiments resulted in Unused ProtScore cut-offs of \geq 4.0 and \geq 2.7 respectively.





3.3. Bias Correction

Use of Bias Correction for this iTRAQ experiment was considered, although examination of the uncorrected data showed a large number of proteins reported as differentially regulated between wildtype and MPS I samples (**Figure 3.2**). It was not clear if these changes were real or due to technical error. A technical iTRAQ labelling error would bias all ratios toward a reported increase in one sample uniformly, but this trend was not observed in this iTRAQ data. Many proteins were both increased, and many others were decreased between the samples, making it unlikely that such a labelling error occurred. Additionally, there were no indications from the protein concentration assay prior to iTRAQ labelling of unequal protein loading between samples.

If most of the uncorrected ratios of proteins are real and not the effect of unequal iTRAQ labelling, then applying the Bias Correction would reduce the fold change reported, potentially causing some to drop below fold-change thresholds. This effect is clearly observed in **Figure 3.2**, where a number of proteins in the corrected dataset are no longer >2-fold change increased or decreased.

In most differential proteomic international experiments at the time of this writing, the large majority of proteins reported had ratios close to 1, so that a very small Bias Correction factor would be applied (unless a labelling issue had occurred), making use of this algorithm appropriate. This was, however, not the case with the present data (**Figure 3.2**). Two factors led to the decision to apply Bias Correction in ProteinPilot despite this issue; the ProteinPilot software prevents the use of Background Correction in the absence of Bias Correction, and applying it would increase the stringency of analysis by reducing lower fold changes and thus highlighting the larger changes. However, as this was a hypothesis-generating experiment, bias correction may be unnecessarily stringent and it would have been ideal to not have to apply it.



Figure 3.2: Distribution of fold change in 842 proteins from the 5 week growth plate iTRAQ MS run comparing MPS I to wildtype a) without and b) with Background and Bias Correction factors applied. Negative fold change indicated a decrease in the MPS I Samples compared to wild type; positive fold changes indicated an increase in MPSI samples compared to wild type

3.4. Summary of Primary Analysis

From **Table 3.1**, it is clear that the depth of proteome sampling in these experiments was strong based on the number of proteins identified (when compared to current skeletal proteomic literature datasets where 121 (136) to 1231 (137) proteins have been identified in cartilage or bone tissue lysates), despite this study using a pure tissue sample. The numbers in **Table 3.1** are also much more stringent than others have reported where, for example, a global rather than local protein confidence of 95% is a more typical threshold (which would change the number of identified proteins in the 5 week growth plate to 1050 instead of a conservative 409). The percentage of used spectra (**Table 3.1**) was typical.

Table 3.1: Lists of proteins from 3 and 5 week MS runs that meet the confidence and fold change cut off values.

5 week growth plate proteins identified (>99% Conf.)	409
Distinct Peptides	4829
Spectra used (% of 27712)	47.2
Proteins increased by ≥2-fold in MPS I	95
Proteins decreased by ≥2-fold in MPS I	125
5 week cranial plate proteins identified (>99% Conf.)	377
Distinct Peptides	4709
Spectra used (% of 28191)	52.8
Proteins increased by ≥2-fold in MPS I	173
Proteins decreased by ≥2-fold in MPS I	68
3 week growth plate proteins identified (>99% Conf.)	110
8· p· · · · · · · · · · · ·	410
Distinct Peptides	410 4951
Distinct Peptides Spectra used (% of 17786)	4951 70.5
Distinct Peptides Spectra used (% of 17786) Proteins increased by ≥2-fold in MPS I	4951 70.5 6
Distinct Peptides Spectra used (% of 17786) Proteins increased by ≥2-fold in MPS I Proteins decreased by ≥2-fold in MPS I	418 4951 70.5 6 7
Distinct Peptides Spectra used (% of 17786) Proteins increased by ≥2-fold in MPS I Proteins decreased by ≥2-fold in MPS I 3 week cranial plate proteins identified (>99% Conf.)	418 4951 70.5 6 7 418
Distinct Peptides Spectra used (% of 17786) Proteins increased by ≥2-fold in MPS I Proteins decreased by ≥2-fold in MPS I 3 week cranial plate proteins identified (>99% Conf.) Distinct Peptides	418 4951 70.5 6 7 418 4951
Distinct Peptides Spectra used (% of 17786) Proteins increased by ≥2-fold in MPS I Proteins decreased by ≥2-fold in MPS I 3 week cranial plate proteins identified (>99% Conf.) Distinct Peptides Spectra used (% of 17786)	418 4951 70.5 6 7 418 4951 70.5
Distinct Peptides Spectra used (% of 17786) Proteins increased by ≥2-fold in MPS I Proteins decreased by ≥2-fold in MPS I 3 week cranial plate proteins identified (>99% Conf.) Distinct Peptides Spectra used (% of 17786) Proteins increased by ≥2-fold in MPS I	418 4951 70.5 6 7 418 4951 70.5 20

This strong sampling depth is likely a result of the intensive protein extraction protocol used, the sample fractionation prior to and during the LC run, and the comprehensiveness of the UniProt database used for peptide matching. It was initially thought that keeping separate the distinct extracts yielded by the 4-buffer protein extraction method would aid in sample fractionation. However, all of the proteins identified in **Table 3.1** were represented in extract 2, and most were represented in the extract 1, 3, and 4 pool as well. Due to the lack of mutual exclusivity between extracts observed and the majority of protein content being present in extract #2, analyzing extracts individually likely only unnecessarily complicated later analyses. This made any gain in sample fractionation not worthwhile contrary to Jiang *et al*'s claims in the original report of the extraction method (112).

3.5. Overall iTRAQ Ratios Between Protein Extracts

Analysis of the iTRAQ ratios given for protein candidates identified by iTRAQ analysis is complicated by the protein extracts not being mutually exclusive (noted above). Ultimately, I decided the most appropriate approach to resolve this issue was to sum the ratios for a specific protein across each extract to estimate the overall ratio. While summing, each protein ratio in each extract pool would be weighted inversely according to the iTRAQ p-value associated that protein, to prevent protein ratios based on fewer or lower quality peptides from excessively biasing the final protein ratio. When using this summing approach, the primary iTRAQ ratio was considered to be the one from extract #2 because it contained the highest proportion of protein quantity and in all cases also contained the most peptides for each protein. For this reason, the total iTRAQ ratio was calculated by adding or subtracting the other ratios from this primary one as per the equation:

$$Ratio_{rotal} = Ratio_1 + \sum_{i=1}^{n} Ratio_i (1 - p \ val_i)$$

In most cases, applying this equation led to a small change in the overall iTRAQ ratio as the pvalue of one or more of the other extracts was poor and/or the fold change was similar to the primary ratio.

To be considered a candidate, a protein needed to have an overall ratio of \geq 2-fold change as a stringent criterion to account for variances (including reporter quantitation accuracy, experimental handling, technical variance, and biological variance). Variance in relative protein quantitation in typical iTRAQ experiments has been estimated (based on proteomic and transcriptomic dataset overlap) to be approximately 58% (23% for experimental variation and 25% for biological variation) (138). It is expected, however, that biological variance should not contribute as much here, as pooling of individual mouse samples prior to analysis would partially normalize for this (138).

3.6. Candidate Lists

From the primary dataset of proteins established after iTRAQ data filtering as described above, several sub lists were created in order to prioritize protein candidates for validation and further investigation. These were created on the basis of iTRAQ ratio p-values, peptide agreement on fold-change for each protein (both within and between extracts), and number of peptides identified for each protein. The final candidate list for further investigation and validation was based on peptide agreement within each protein, as this was the most stringent and biologically-relevant filtering criteria. Proteins were considered candidates if they fulfilled the criteria described above and that \geq 88% of peptides for each protein showed similar significant fold-changes in their parent protein between wildtype and MPS I samples. While the peptide agreement cut off of 88% is somewhat arbitrary, it allowed for "reasonable" disagreement within a protein so as to be neither too restrictive nor too inclusive. For my application, this cut off was the most logical. The final candidate proteins meeting these criteria are shown in **Tables 3.2, 3.3, and 3.4**.

Table 3.2: Summary of differentially-expressed proteins between the 5 week MPS I and wildtype mouse growth plate samples. iTRAQ ratios (fold change) and p-values are for the ratio of protein abundance in MPS I:wildtype in protein extract 2. Proteins marked with * indicate iTRAQ candidates selected for the MRM panel, while proteins marked with \diamond indicate candidates selected for the qPCR array. Negative fold change indicated a decrease in the MPS I Samples.

≥2-fold Increased in Mutant			
Protein (Gene Name)	UniProt	Peptides	Fold Difference
Structural			
Tenascin-W (Tnn)	Q80Z71	3	22.1
* Annexin A1, Anxa1 protein (Anxa1)	<u>P10107</u>	13	3.5
Integrin alpha M (Itgam)	P05555	3	3.2
Lamin A (Lmna)	P48678	14	2.4
Matricellular			
* Periostin, osteoblast specific factor (Postn)	Q62009	8	32.5
*◊ Asporin, Aspn protein (Aspn)	Q99MQ4	4	15.4
* Osteoglycin (Ogn)	Q62000	4	5.6
*◊ Lumican (Lum)	P51885	5	5.3
* Serine (Or cysteine) proteinase inhibitor f1 (Serpinf1)	P97298	10	2.9
Enzyme			
Catalase (Cat)	P24270	2	14.7
Creatine kinase, muscle (Ckm)	P07310	5	11.2
Hexosaminidase A (Hexa)	P29416	6	3.3
Other			
Myosin, heavy polypeptide 1, skeletal muscle, adult			
(Myh1)	Q5SX40	17	24.7
Lysosomal-associated membrane protein 2 (Lamp2)	P17047	4	7.4
* Lactotransferrin (Ltf)	P08071	9	5.7
Myeloid bactenecin (F1) (Ngp)	008692	8	5.2
Fibrinogen, alpha polypeptide (Fga)	Q99K47	3	4.5
Non-muscle myosin heavy polypeptide 9 (MYH9)	Q8VDD5	6	2.4
≥2-fold Decreased in Mutant			
Protein (Gene Name)	UniProt	Peptides	Fold Difference
Structural		-	
* Perlecan (Heparan sulfate proteoglycan 2) (Hspg2)	Q05793	31	-17.1
* Fibronectin 1 (Fn1)	P11276	39	-4.7
Hyaluronan and proteoglycan link protein 1 (HapIn1)	Q9QUP5	12	-3.7
* Cartilage oligomeric matrix protein, Comp protein			
(Comp)	Q9R0G6	9	-2.8
Matricellular			
*◊ Chondroadherin (Chad)	055226	11	-5.7
* Milk fat globule-EGF factor 8 protein (Mfge8)	P21956	10	-3.9
Leukocyte cell-derived chemotaxin 2 (Lect2)	<u>088803</u>	2	-3.8

♦ Matrilin 1, cartilage matrix protein (Matn1)	P51942	12	-3.6
* Thrombospondin 1 (Thbs1)	P35441	13	-3.5
*◊ Matrilin 3 (Matn3)	035701	9	-3.2
Serine (or cysteine) peptidase inhibitor, clade E, member			
2 (Serpine2)	Q07235	7	-2.8
*◊ Proline arginine-rich end leucine-rich repeat (Prelp)	<u>Q9JK53</u>	10	-2.5
*◊ Epiphycan (Epyc)	P70186	4	-2.1
Enzyme			
♦ Alkaline phosphatase (Alpl)	P09242	7	-7.9
Glyoxalase I (Glo1)	Q9CPU0	2	-5.6
Matrix metallopeptidase 13 (Mmp13)	<u>P33435</u>	12	-3.3
Procollagen lysine, 2-oxoglutarate 5-dioxygenase 2			
(Plod2)	Q9R0B9	5	-3.1
L-lactate dehydrogenase (Ldha)	P06151	14	-2.1
Pyrophosphatase (inorganic) 1 (Ppa1)	Q9D819	10	-4.5
Enolase (Eno1)	P17182	12	-4.1
Other			
Nucleolar pre-ribosomal-associated protein 1 (URB)	Q571H0	5	-6.9
Nidogen 2 protein (Nid2)	<u>088322</u>	8	-4.7
Alpha S1 casein (Csn1s1)	P19228	3	-3.5
Apoe protein (Apoe)	<u>P08226</u>	11	-2.5

 Table 3.3: Final protein candidate list from 5 week cranial plate analysis based on peptide ratio agreement. Negative fold change indicated a decrease in the MPS I Samples.

Cranial Plate

≥2-fold Increased in Mutant			
Protein (Gene Name)		Peptides	Fold Difference
S100 calcium binding protein A8 (calgranulin A) (S100a8)	<u>P27005</u>	2	99.1
Nucleophosmin 1 (Npm1)	Q61937	3	27.5
S100 calcium binding protein A9 (calgranulin B) (S100a9)	<u>P31725</u>	4	22.7
Cbx3 protein (Cbx3)	P23198	2	14.5
histone cluster 1, H1c (Hist1h1c)	<u>P15864</u>	6	9.3
lipocalin 2 (Fragment) (Lcn2)	P11672	2	8.4
Elongation factor 1-alpha (Eef1a1)	<u>P10126</u>	5	6.7
Hexosaminidase A (Hexa)	P29416	3	3.8
myeloperoxidase , Mpo protein (Mpo)	<u>P11247</u>	5	3.4
lactotransferrin (Ltf)	P08071	8	2.5
MCG2706 (Txn1)	<u>P10639</u>	2	2.1
≥2-fold Decreased in Mutant			
Protein (Gene Name)		Peptides	Fold Difference
			I old Difference
riotein (Gene Name)	<u>Q99MQ</u>		
asporin, Aspn protein (Aspn)	<u>Q99MQ</u> <u>4</u>	7	-14.7
asporin, Aspn protein (Aspn) Kappa-casein (Fragment) (Csn3)	<u>Q99MQ</u> <u>4</u> P06796	7 2	-14.7 -7.8
asporin, Aspn protein (Aspn) Kappa-casein (Fragment) (Csn3) serine (or cysteine) peptidase inhibitor, clade H, member 1	<u>Q99MQ</u> <u>4</u> <u>P06796</u>	7 2	-14.7 -7.8
asporin, Aspn protein (Aspn) Kappa-casein (Fragment) (Csn3) serine (or cysteine) peptidase inhibitor, clade H, member 1 (Serpinh1)	<u>Q99MQ</u> <u>4</u> <u>P06796</u> <u>P19324</u>	7 2 13	-14.7 -7.8 -6.8
asporin, Aspn protein (Aspn) Kappa-casein (Fragment) (Csn3) serine (or cysteine) peptidase inhibitor, clade H, member 1 (Serpinh1) Triosephosphate isomerase (Tpi1)	<u>Q99MQ</u> <u>4</u> <u>P06796</u> <u>P19324</u> <u>P17751</u>	7 2 13 2	-14.7 -7.8 -6.8 -4.0
asporin, Aspn protein (Aspn) Kappa-casein (Fragment) (Csn3) serine (or cysteine) peptidase inhibitor, clade H, member 1 (Serpinh1) Triosephosphate isomerase (Tpi1) Osteopontin (Spp1)	Q99MQ 4 P06796 P19324 P17751 P10923	7 2 13 2 3	-14.7 -7.8 -6.8 -4.0 -3.9
asporin, Aspn protein (Aspn) Kappa-casein (Fragment) (Csn3) serine (or cysteine) peptidase inhibitor, clade H, member 1 (Serpinh1) Triosephosphate isomerase (Tpi1) Osteopontin (Spp1) Heat shock protein 90, beta (Grp94), member 1 (Hsp90b1)	Q99MQ 4 P06796 P19324 P17751 P10923 P08113	7 2 13 2 3 5	-14.7 -7.8 -6.8 -4.0 -3.9 -3.6
asporin, Aspn protein (Aspn) Kappa-casein (Fragment) (Csn3) serine (or cysteine) peptidase inhibitor, clade H, member 1 (Serpinh1) Triosephosphate isomerase (Tpi1) Osteopontin (Spp1) Heat shock protein 90, beta (Grp94), member 1 (Hsp90b1) 40S ribosomal protein S12 (Rps12)	Q99MQ 4 P06796 P19324 P17751 P10923 P08113 P63323	7 2 13 2 3 5 2	-14.7 -7.8 -6.8 -4.0 -3.9 -3.6 -3.0
asporin, Aspn protein (Aspn) Kappa-casein (Fragment) (Csn3) serine (or cysteine) peptidase inhibitor, clade H, member 1 (Serpinh1) Triosephosphate isomerase (Tpi1) Osteopontin (Spp1) Heat shock protein 90, beta (Grp94), member 1 (Hsp90b1) 40S ribosomal protein S12 (Rps12) osteoglycin (Ogn)	Q99MQ <u>4</u> P06796 P19324 P17751 P10923 P08113 P63323 Q62000	7 2 13 2 3 5 2 5	-14.7 -7.8 -6.8 -4.0 -3.9 -3.6 -3.0 -2.4

Table 3.4: Final protein candidate list from 3 week growth plate and cranial plate analyses based on peptide ratio agreement. Negative fold change indicated a decrease in the MPS I Samples.

Growth Plate

>2 fold Increased in Mutant			
22-rold increased in Mutant	UniDrot		
Protein (Gene Name)		Pentides	Fold Difference
Serine (or cysteine) peptidase inhibitor, clade H, member 1	, teen	reptides	
(Serpinh1)	P19324	12	2.7
Lysosomal-associated membrane protein 2 (Lamp2)	P17047	3	3.0
Osteomodulin (Omd)	035103	5	2.2
Osteoglycin (Ogn)	Q62000	3	2.0
≥2-fold Decreased in Mutant			
Protein (Gene Name)		Peptides	Fold Difference
-	-	-	-
Cranial Plate			
≥2-fold Increased in Mutant			
Protein (Gene Name)		Peptides	Fold Difference
Lysosomal-associated membrane protein 2 (Lamp2)	P17047	3	5.6
Hexosaminidase A (Hexa)	P29416	3	5.0
Chitinase 3-like-3 (Chi3l3)	035744	3	2.1
Annexin A1 (Anxa1)	<u>P10107</u>	4	1.6
Protein disulfide isomerase family A, member 6 (Pdia6)	<u>Q922R8</u>	5	2.7
Annexin A11 (Anxa11)	<u>P97384</u>	2	1.8
Myeloperoxidase (Mpo)	<u>P11247</u>	2	2.2
Lectin, galactose binding, soluble 3 (Lgals3)	<u>P16110</u>	2	1.9
S100 calcium binding protein A9 (S100a9)	<u>P31725</u>	2	1.6
Ubiquitin B (Ubb)	P0CG49	6	1.7
lymphocyte cytosolic protein 1 (L-plastin) (Lcp1)	<u>Q61233</u>	8	1.5
≥2-fold Decreased in Mutant			
Protein (Gene Name)		Peptides	Fold Difference
Biglycan (Bgn)	<u>P28653</u>	12	-6.7
Decorin (Dcn)	<u>P28654</u>	9	-2.1
Matrix metalloproteinase 13 (Mmp13)	<u>P33435</u>	6	-2.2
Alkaline phosphatase (Alpl)	<u>P09242</u>	5	-2.7
Serine (Or cysteine) proteinase inhibitor, clade F, member 1			4 -
(Serpinf1)	<u>P97298</u>	10	-1.5
Fructose-bisphosphate aldolase (Aldoa)	<u>P05064</u>	7	-6.2
Putative uncharacterized protein (Ckm)	<u>P07310</u>	3	-1.7
Nidogen-2 (Nid2)	088322	7	-5.4
Talin 1 (Fragment) (Tin1)	<u>P26039</u>	9	-1.9
AHNAK (Fragment) (Ahnak)	<u>Q6UL10</u>	3	-2.5

3.7. Pathway Analysis of Candidates

After establishing final iTRAQ candidate lists (section **3.6.**), the biological relevance of these proteins to each other, and to MPS I skeletal disease, was investigated through bioinformatic network analysis of protein interactions, gene co-regulation, and transcription factor activity. In particular, key transcription factors involved with chondro-osseous system development (described in section **1.5.**) were targeted, as it was hypothesized that changes in one or more of these transcription factors could coordinately cause alteration of several protein candidates identified by iTRAQ. These hypothesized changes in transcription factor levels were unlikely to be revealed by iTRAQ analysis given the low level and transient nature of transcription factor expression.

Multi Experiment Matrix (MEM) (132) discovers evidence of transcriptional coregulation between genes based on microarray results compiled in the NCBI Gene Expression Omnibus (GEO). As a starting point (**Figure 3.3**), MEM analysis was carried out for fourteen proteins of interest (asporin, biglycan, chondroadherin, Decorin, epiphycan, fibromodulin, lactotransferrin, lectin GBS3, lumican, milk fat globule EGF factor 8 (MFGE8), osteoglycin, osteomodulin, periostin, proline/arginine-rich end leucine-rich repeat protein (PRELP), and osteonectin (SPARC)) from the iTRAQ 5 week growth plate dataset (**Table 3.2**). These proteins were selected on the basis of perceived biological interest to MPS I and chondro-osseous system function (**Table 1.3**). The small leucine repeat proteoglycan proteins (SLRPs) were largely represented in this list. The SLRPs are matricellular proteins, which are extracellular matrix components involved with both signaling and structural roles during chondro-osseous development and repair (139). Established roles of these proteins include cellular proliferation, modulation of extracellular matrix organization, growth factor modulation, and matrix adhesion. The resulting ranked list of ~35,000 loci (according to co-regulation p-value) for each query gene was cut off at the first arbitrarily 100 genes (**Table 3.4** shows an example of this for select proteins). The list of the top 100 genes for each query gene was then compared to the other query gene result lists.

Once genes with evidence of co-regulation from MEM were identified, RSAT (130) was used to search the genomic sequences around these genes for shared potential regulatory sequences. RSAT analysis was performed on genomic sequence from 10KB upstream of the transcriptional start site and 10KB downstream of the transcriptional termination site of each candidate gene from MEM analysis was retrieved from the NCBI GenBank. A list of 20 potentially co-regulated genes was ultimately established from the MEM results analysis (collagen I, collagen III, collagen V (a1, a2, a3), collagen VI (a1, a2, a3, a4, a6), osteonectin, SERPIN F1, periostin, matrillin 2, fibronectin, procollagen C, and procollagen C2, asporin, biglycan, chondroadherin, decorin, epiphycan, fibromodulin, lumican, osteoglycin, and PRELP).

After investigation of shared potential regulatory sequences in the candidate genes from MEM and RSAT, these sequences were examined for matching to known transcription factor binding consensus sites. Transcription factor binding site analysis was performed on 50 identified enriched sequences found in candidate genes from RSAT analysis. This identified shared patterns enriched in these genes including AAAACAAA, ACACACAC, AAAATAT, ACAAAACA, CACACACA, and CAAAAAAA above the expected genomic frequency defined by RSAT. Analysis of these sequences by TFSearch

(http://www.rwcp.or.jp/lab/pdappl/papia.html) identified them as potential SOX/SRY

transcription factor binding sites. Data in the JASPAR database confirms that the consensus binding sites for SOX5 and SOX9 include the sequence AACAAT in most cases, while the SRY consensus binding site often includes ACAAT. TFSearch was used to identify transcription factors which may bind to these putative regulatory sequences based on known consensus transcription factor binding sites.

The JASPAR database of transcription factors (140) was also manually searched for matching consensus sequences. This identified eight potential transcription factors (having putative consensus sequence binding sites near genes coding for proteins of interest from MEM analysis (SOX/SRY, RUNX1, E-BOX, Osterix, SOX5, SOX9, SOX10, MSX1, and MSX2). It was hypothesized that these transcription factors might be coordinately regulating several protein candidates deregulated in the MPS I growth plate. SOX/SRY, RUNX1, E-BOX, Osterix, SOX5, SOX9, SOX10, MSX1, and MSX2, SOX9, SOX10, MSX1, and MSX2 consensus binding sites were then mapped 10KB upstream of the transcriptional start site and 10KB downstream of the transcriptional termination site of each candidate gene (**Figure 3.4**).

The series of analyses described above collectively revealed evidence of potential coregulation between several members of the SLRP protein family as indicated by their enrichment in the top 100 ranked co-expressed genes after MEM analysis. Further evidence is provided by SLRP genes being co-regulated with other genes relevant to skeletal development such as collagen I, collagen III, collagen VI, pro-collagen C, osteonectin, periostin, and serpin F1 (**Table 3.5**).



Figure 3.3: Workflow of bioinformatics tools used to generate hypotheses for key altered signalling and regulatory networks based on the iTRAQ proteomics candidates. The particular software tool(s) used for each stage is indicated beside.

When these consensus sequences were mapped on the promoter regions of the genes in **Table 3.5**, there were numerous sites (more than expected by chance) in the ± 10 kb non-coding region of the SLRP family genes, as well as in the genes suggested to be co-regulated (**Figure 3.4**). This sequence clustering provided further evidence for potential co-regulation of epiphycan, PRELP, osteoglycin, osteomodulin, and lumican by SOX9 and SOX5, and co-regulation of biglycan, chondroadherin, decorin, epiphycan, osteoglycin, and lumican by MSX1 and MSX2. Despite OSE2 and RUNX2 being well-known skeletal regulatory transcription factors, there was no evidence of consensus binding sequence enrichment in the genes queried.

3.8. Summary

Comparative iTRAQ proteomic interrogation of the five week MPS I and wildtype femoral head growth plate identified 409 distinct proteins. 19 of these proteins were significantly increased in the MPS I growth plate, while 24 proteins were significantly decreased (**Table 3.2**). Among these proteins, most the SLRP family proteins were shown to be deregulated in MPS I. After establishing final iTRAQ candidate lists, the biological relevance of these proteins to each other, and to MPS I skeletal disease, was investigated through bioinformatic network analysis of protein interactions, gene co-regulation, and transcription factor activity. This series of analyses collectively revealed evidence for co-regulation of epiphycan, PRELP, osteoglycin, osteomodulin, and lumican by SOX9 and SOX5, and co-regulation of biglycan, chondroadherin, decorin, epiphycan, osteoglycin, and lumican by MSX1 and MSX2. This potential co-regulation of SLRP genes may explain their coordinate changes observed in the MPS I growth plate (**Table 3.2**).

Query Gene	Asporin	Fibromodulin	Biglycan	Decorin	Lumican	PRELP	Osteoglycin	Osteomodulin
Result								
Genes	Osteoglycin	Procollagen C	Collagen I	Lumican	Decorin	Biglycan	Decorin	Osteoglycin
	Periostin	Osteoglycin	Collagen III	Collagen I	Collagen III	Decorin	Asporin	Asporin
	Collagen V	Collagen I	Collagen V	Collagen III	Collagen I	Osteoglycin	Collagen III	Serpin F1
	Lumican	Biglycan	SPARC	Biglycan	Collagen V	Serpin F1	Collagen I	Fibromodulin
	Collagen III	Collagen III	Collagen VI Procollagen	Collagen V	Periostin	Procollagen C	Procollagen C	Decorin
	Collagen I	Collagen VI	С	Osteoglycin	Collagen VI	SPARC	Lumican	Procollagen C
	Collagen VI	Collagen V	Collagen IV	SPARC	Osteoglycin	Collagen III	Collagen VI	Collagen I
	Procollagen C	Serpin F1	Serpin F1	Serpin F1	Biglycan Procollagen	Collagen VI	Fibromodulin	Lumican
	Decorin	SPARC	Serpin H1	Collagen VI	С	Fibromodulin	Biglycan	Collagen III
	HSP b11	Decorin	Periostin Fibronectin	Periostin	Serpin F1	Lumican	Collagen IV	Collagen XI
	Matrilin 2	Collagen XVI	1	Procollagen C	Serpin H1	Collagen I	Collagen V	Collagen VI
	Biglycan	Lumican	Decorin	LGALS 1	SPARC		Periostin	
	Collagen XII	Asporin	Lumican		Asporin		SPARC	
	SPARC	Collagen VIII			Collagen XIV			
	Collagen VIII				LGALS 1			
	Collagen IV				Matrilin 2			
	Fibromodulin							

Table 3.5: Query proteins ranked within the top 100 protein hits with strong evidence of co-regulation from SLRP family MEM analysis.

	Asporin	-2000	-1000 199 9 9 9 9999	Legend 	.00	
	Biglycan	99	٩			
	Chondroadher	in [°]	٩			
	Deserie	1 9 9	199 1999 P			
	Decorin	· · · · ·	299 99 9			
	Epiphycan	<u> </u>	99 9			
	Fibromodulin					
	PRELP		<u>, ү</u>			
	Osteoglycin	000000000	000000000000000000000000000000000000000	<u>>>></u>		
	Osteomodulin	100000	000000000000000000000000000000000000000			
	Lumican	222 22999	99999 9			
gi 149288871: 94787584-94797584-		IN[AC0][AC0] 1.00 00 AIN[AC0][AC0]0 1.0		-2000	-1000	Legend AACAAT 1.00 (ATJAACAAT 1.00
Collagen I	000000000000000000000000000000000000000		Collagen I	4/8/584-94/9/584 99999	999	
si 1149290852:45350303-45360303-			Collagen III	5358383-45368383		
811 149288869:c76171537-76161537	<u> </u>		Collagen VI a	6171537-76161537	~~~~~	
stil 1492088691:76058501-76048501	<u> </u>		Collagen VI a	6058501-76048501 12		
gi 149288852:092663435-92653435			Collagen VI a	2663435-92653435 → 1 3		
gi 149361524:c105892649-105882649			Collagen VI a	892649-105882649 		
giiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii	00000000000		collagen VI a	591747-105581747 	220	
gi1149338249:27731945-27741945			^{gi 149338249:2} Collagen V a	7731945-27741945 		
gi 14928852:c45431176-45421176	89 92229		gi 149288852:c4	5431176-45421176	P29	
si 149361524:c20574494-20564494	9 <mark>9</mark> 9		gi 149361524:c2	∠ 0574494-20564494 →	220	
Collagen V a3	2000 99		Collagen V a	3 5208003-55198003→	8	
SPARC	P		SPARC			
g11149208971ro75223531-75213531 Serpin f1	99		Serpin f1	5223531-75213531	000	
gi 149352351:54155047-54165047 Periostin	<u> </u>		Periostin	4155047-54165047	<u>P</u>	
gi 149301884:34226436-34236436	92220 99		matrilin 2	4226436-34236436	229	
sil1492888521c71632097-71622097	88		Fibronectin 1	1632097-71622097		
^{g111493542241c138046335-138036335}	<u>م</u>		Procollagen	046335-138036335		
g1 149361524:95528047-95538047	89999		^{gi 149361524:9} Procallagen (5528047-95538047→	<u>89</u>	

Figure 3.4: Examples of RSAT feature maps for a) SLRP protein family member promoters regions, and SLRP and other protein candidate predicted binding sites for SOX5 and SOX9 (b and c respectively) transcription factors. The first 2KB from the transcriptional start site are shown.

4. Proteomics Candidate Validation and Functional Studies

4.1. Introduction

In order to discover the pathways underlying chondro-osseous disease in the murine model of MPS I, an unbiased proteomic approach was used first (as noted in the flow chart summarizing the experimental approach of this dissertation (Figure 1.3)). Any proteins which were observed to be significantly increased or decreased relative to the wildtype control mice were then re-examined through validation in new samples using targeted approaches such as Western blotting, immunohistochemistry, and targeted mass spectrometric analysis.

Quantitative proteomic experiments such as iTRAQ reveal broad results spanning potentially thousands of proteins to generate hypotheses in an unbiased way. Validation and further study of candidates is essential to ascertain the relevance of candidates. Historically, validation of quantitative proteomic experiments has involved Western blotting (or similar antibody-mediated detection methods such as immunohistochemical staining), RNA expression studies, and sometimes very specific characterization of candidate proteins such as crystal structure determination. Western blots are still the most sensitive method of protein quantitation; however, in the case of this work, validation of protein candidates by Western blot was challenging due to the lack of multiplexing capacity, the very limited quantity of protein obtained from bone tissue, and the large number of protein candidates for which validation was sought.

<u>Multiple reaction monitoring (MRM)</u> mass spectrometry has gained popularity as a means of quantifying specific proteins of interest with high accuracy. Rather than discovering and quantifying peptides in an unbiased approach, the MS in an MRM experiment is focused only on one or a few peptides from certain proteins in a complex sample. In this way, MRM

experiments can provide more accurate quantification of protein between samples, because representative peptides can be selected for each protein that will perform optimally and predictably, while avoiding confusion introduced by problematic peptides. MRM experiments are commonly performed on triple-quadrupole (QQQ) mass spectrometers because of the technical approach these machines use (141). The first and third quadrupole modules within the QQQ MS are capable of scanning for specific m/z values to identify the parent peptide and its fragmentation ions. In an MRM experiment performed on a QOQ MS, the peptides of interest are identified in Q1 and fragmented before being sent to Q3 (141). Predefined fragment ions are selected in Q3 and guided towards the detector for quantitation (141). For each peptide of interest, so-called 'transitions' must also be chosen and applied as a selective filter in the second MS scan, where each transition represents the m/z values to detect the parent and specific fragment ions (141). These two filters mean that MRM is extraordinarily specific, as each peptide is identified by its unique m/z value and then by the m/z values for at least 2 of its fragmentation ions (141). Transitions (i.e., the parent ion and fragment ions) are necessary because quantitation of just the parent ion from Q1 would be obscured by the complexity of the background. Quantitation of the fragment ions rather than the parent peptide allows for more precision, though it decreases sensitivity because not all fragment ions of a parent peptide are quantitated. Theoretically, the sum of quantitation results from all transitions should be similar to the original quantity of the parent peptide. Therefore, multiple transitions are required to adequately identify and quantitate the parent peptide.

The sensitivity of MRM is greater compared to non-targeted MS approaches because target filtering reduces equipment and computational demand within any given sampling window (141). This makes ratio values obtained by the MRM assay theoretically more accurate than the

iTRAQ ratios. Most MRM experiments utilize labelled synthetic versions of each targeted peptide which are heavier than their physiological counterparts due to incorporation of ¹³C, ¹⁵N, or ²H. These standards can then be spiked into the samples in known concentrations and compared to an empirically established standard curve for each peptide to obtain concentration data. By comparing the physiological peptide with the spiked-in synthetic version and a standard curve, the absolute concentration of the peptide (and thus the parent protein) in the sample can be determined. Heavy labeled standards were not used, as only relative quantitation was sought.

MRM is ideal for multiplex validation of primary quantitative proteomic results because it addresses the issue of limited samples and multiple targets (142). Thus the primary validation of iTRAQ candidates I used was a custom MRM panel for 27 proteins of interest on freshly prepared five week mouse micro-dissected growth plate protein to compliment the Western blots and immunohistochemistry results. Technical replication of initial iTRAQ data on the same MS is considered ideal but is not always feasible due to cost and limiting samples. Previous analyses of iTRAQ technical variation have suggested it to be 10-12% at the protein ratio level (138). If these estimations are correct, then applying a fold-change cut off filter of >100% (>2-fold, **3.6.**) during iTRAQ candidate filtering should prevent technical variation from biasing results in most cases.

4.2. iTRAQ Validation

4.2.1. Western Blots of Bone Protein Lysates

Western blots were done for 5 of the candidates in **Table 3.2**: COMP, PRELP, chondroadherin, thrombospondin 1, and perlecan. These candidates were based on interest in the SLRP family of proteins and the thrombospondins because of their key roles in skeletal development and the large magnitude of reduction these proteins showed in MPS I samples. As

the proteomic samples were produced by pooling individual protein lysates from 9 animals (section **2.3.**), Western analysis of individual mouse samples left over from pooling enabled me to investigate if the iTRAQ ratios for these proteins were being biased by one mouse for a reason not related to the MPS I genotype (e.g., technical or biological variation between individuals). Western analyses were performed on each protein extract (i.e., 1, 2, 3, and 4 separately) and each individual mouse rather than pooling. The Western results (**Figure 4.1**) did not indicate the presence of a biasing outlier individual for the tested proteins and mice.

Despite loading up to 20µg of protein from each sample, there was no detectable signal from thrombospondin 1 or chondroadherin. Western analysis of COMP and PRELP supported the iTRAQ results (decreased in the MPS I samples relative to the wildtype samples), while the perlecan Western results were ambiguous (**Figure 4.1**). The large size (>200kDa) of perlecan made gel electrophoresis difficult, potentially preventing accurate quantitation of this protein by Western analysis. While the trend of decreased levels of COMP and PRELP in the MPS I samples by Western analysis is clear, validation of the other protein candidates from iTRAQ analysis was not attempted due to the limitations of this method (i.e. the inaccuracies and subjectivity of densitometry, and the lack of multiplexing capacity). Fold change validation by MRM was used rather than Western analysis for further validation of iTRAQ candidates (section **4.3.**).







Figure 4.1: Summary of Western blot results. a). images of blots for Perlecan, COMP, and PRELP from individual mouse (M (MPS I) or WT (wildtype)) protein extracts (A, B, C, and D). b) densitometry results of blots normalized to total µg protein loaded (WT in blue and mutant in red). All mice were 5 weeks of age and the sex of each is indicated below the sample.

There was suggestive evidence of sexual dimorphism of expression in COMP and PRELP proteins. The decrease in expression in MPS I samples was only significant in the Western data when the MPS I and wildtype mice compared were of the same sex (**Figure 4.1a**). The blots for COMP and PRELP (**Figure 4.1**) suggest that these proteins may be expressed at a lower level in 5 week old females compared to males based on the male to male comparison of wildtype to MPS I showing a significant (p < 0.05) decrease but the male MPS I to female wildtype showing an insignificant increase in MPS I. Despite the male/female ratio in the wildtype pools being lower than the MPS I pool (**2.1.7.**), this would not have necessarily biased the resulting candidates. Most of the identified proteins showed decreased expression between MPS I to wildtype, which is the opposite of what would be expected if the proteins were expressed at a lower level in females. Since there are 33% more females in the wildtype pools than MPS I, this could be corrected for in the iTRAQ ratio data by imposing a fold change cut-off that is greater than 33% or 0.33. With the initial fold change cut-off of ≥ 2 fold this difference would be considered background and should not bias the reported candidates.

4.2.2. Western Analysis of MPS I Murine Serum

There remains a lack of biomarkers for MPS I that directly correlate with disease burden. An exception is heparin cofactor II-thrombin (HCII-T) complex recently characterized by another iTRAQ proteomic analysis (114). However, neither this biomarker nor others have been shown to correlate or reflect chondro-osseous disease burden in MPS I. Some of the protein candidates discovered by iTRAQ may hold potential as biomarkers due to their important roles in chondro-osseous homeostasis. For example, COMP is a well-known key structural component of the extracellular matrix, while PRELP is a member of the SLRP protein family, which are regulators of the collagen network and extracellular matrix growth factor signaling (143, 144, 145, 146, 147). COMP may also be a biomarker for osteoarthritis (OA), as suggested by elevation of COMP fragments in OA patient serum (148). While the proteins identified here originated from the growth plate cartilage, sampling of this tissue is not feasible for traditional biomarker analysis due to the inaccessibility of growth plate cartilage in patients (without surgery). Serum biomarkers, such as HCII-T, are much more popular and successful in the clinic due to the ease of obtaining samples.

COMP, and nine other candidates identified in this iTRAQ analysis have been identified in the HUPO Plasma Proteome (**Table 4.1**), thus they are potentially feasible biomarker targets due to their measurability in plasma. In order to determine if the changes seen in the growth plate were represented in the serum levels of these proteins, Western blots for COMP and PRELP were done for four MPS I and four wildtype mouse serum samples. There were no significant differences between these proteins in MPS I and wildtype samples (**Figure 4.2**).

 Table 4.1: List of protein candidates that overlap with those identified with high confidence in the HUPO Plasma

 Proteome (149)

Protein

apolipoprotein A-I (Apoa1)
lactotransferrin (Ltf)
nucleolin (Ncl)
Prostaglandin I2 (Prostacyclin) synthase (Ptgis)
fibronectin 1 (Fn1)
albumin (Alb)
Thrombospondin 1 (Thbs1)
cartilage oligomeric matrix protein, Comp protein (Fragment) (Comp)
F10 protein (F10)
secreted acidic cysteine rich glycoprotein (Sparc)


b)



Figure 4.2: Summary of Western blot results. a). images of blots for COMP and PRELP from individual mouse (M (MPS I) or WT (wildtype)) serum samples (A, B, C, and D). b) densitometry results of blots normalized to total µg protein loaded (WT average density in blue and MPS I average density in red). All mice were 5 weeks of age and the sex of each is indicated below the sample. COMP band at ~115kDa, PRELP band at ~32kDa.

4.3. MRM Assay

As discussed above, the large number of candidate proteins to be validated and low amount of protein yielded from micro-dissected mouse femoral head growth plate cartilage made Western analysis challenging for comprehensive iTRAO validation. A MRM assay made validation of 27 candidates (**Table 4.2**) feasible, but it also laid the groundwork for a very sensitive, specific, multiplexed, and rapid assay that could be used on low concentration protein samples in the future to establish either the relative or absolute quantity of these proteins of interest. Selected proteins of interest for candidate validation were based on > 88% of iTRAO peptide ratios of each protein suggesting the same direction of change between mutant and wildtype (i.e., >88% peptides suggest the protein is either increased or decreased) (133). In the five week growth plate data, this identified 43 proteins of interest, and 6 proteins of interest in the three week growth plate (133) (**Tables 3.2 and 3.4**). As the published report of these data states, 27 proteins were selected for MRM (133); Proteins were included in the MRM panel based on the iTRAQ dataset (16 proteins marked with an asterisk in Table 3.2) or proteins that were related to the iTRAQ dataset and/or skeletal disease (11 proteins in Table 1.3) (133). Fresh micro-dissected 5 week growth plate samples were prepared for MRM analysis and assayed as individual mice (6 MPS I and 5 wildtype mice) rather than in pools (133).

The final list of proteins and selected peptides used in the MRM assay is shown in **Table 4.2**.

Table 4 2. I ist of such that and a			MDM assessed and developed
Table 4.2.2 List of broteins and co	srresnonaing selectea i	nentiaes for which the	P VIR VI ASSAV WAS DEVELODED.
I WOLD IN LIST OF PLOTOINS WING CO	si i esponanic selectea		

Protein	UniProt Accession Number	Selected MRM Peptide Sequence
Aggrecan	P16112	ARPNCGGNLLGVR
		YQCTEGFVQR
Annexin A1	P10107	GVDEATIIDILTK
		TPAQFDADELR
Asporin	Q99MQ4	LYLSHNQLSEIPLNLPK
		YWEIQPATFR
Biglycan	P28653	NHLVEIPPNLPSSLVELR
		EISPDTTLLDLQNNDISELR
Chondroadherin	055226	SIPDNAFQSFGR
		YLETLWLDNTNLEK
Collagen I α1	P02452	SAGVSVPGPMGPSGPR
		GETGPAGPAGPIGPAGAR
Collagen II α1	P02458	GAQGPPGATGFPGAAGR
		GFPGLPGPSGEPGK
COMP	Q9R0G6	DTDLDGFPDEK
		ELQETNAALQDVR
Decorin	P28654	ASYSAVSLYGNPVR
		YWEIFPNTFR
Epiphycan	P70186	IDLTSNLISEIDEDAFR
		QLPELPNTLTFIDISNNR
Fibromodulin	P50608	IPPVNTNLENLYLQGNR
		SLILLDLSYNHLR
Fibronectin	P11276	FLTTTPNSLLVSWQAPR
		TFYQIGDSWEK
Lactotransferrin	P08071	LRPVAAEVYGTK
		STNDKEEAIWELLR
Lect2	O88803	HHPGVDVLCS DGSVVYAPFTGK
LGALS 3	P16110	- IQVLVEADHFK
		GNDVAFHFNPR
Lumican	P51885	ISNIPDEYFK
		NNQIDHIDEK
Matrilin 3	035701	ASGIELYAVGVDR
		VAVVNYASTVK
MFGE8	P21956	INAWTAQSNSAK
		VSGVMTQGASR
Osteoglycin	Q62000	RLDFTGNLIEDIEDGTFSK
<u> </u>		LDFTGNLIEDIEDGTFSK
Osteomodulin	035103	IHSIYYGEQR
		-

Protein	UniProt Accession Number	Selected MRM Peptide Sequence
Periostin	Q62009	GFEPGVTNILK
		IIDGVPVEITEK
Perlecan	Q05793	LVSEDPINDGEWHR
		SLPEVPETIEFEVR
PRELP	Q9JK53	LENLLLDLQHNR
		NQLEEVPSALPR
Serpin F1	P97298	DTDTGALLFIGR
		LAAAVSNFGYDLYR
SPARC	P07214	LEAGDHPVELLAR
		NVLVTLYERDEGNNLLTEK
Thrombospondin 1	P35441	AGTLDLSLSLPGK
		FVFGTTPEDILR
Thrombospondin 3	Q05895	QTEQTYWQATPFR
		FYVVMWK

4.3.1. Peptide Selection

All peptides identified in the iTRAQ dataset for each of the 27 proteins were examined as potential MRM peptides based on a confidence score >99%, unique to only a single protein, not containing M, C, KK, KR, or RK amino acids (missed tryptic cleavage sites) (141), and including P amino acids when possible (known to have a strong and unique mass profile). Of the peptides matching these criteria, two that were also listed in the MRM peptide database of the Global Proteomics Machine Database (GPMDb) (150) were ultimately selected for inclusion in the MRM assay. By referring to the GPMDb, selected peptides had a higher likelihood of being successfully fragmented and interpretable in the MRM experiment, because only peptides used at least once in a real MRM experiment are included in this database. Individual peptide iTRAQ ratio information was not a major selection criterion; however peptides with extreme ratios (ratio of 0.01 or 100) were not chosen.

By selecting peptides from the iTRAQ results, it was known that they would be present and detectable by MS in mouse growth plate protein samples. All selected peptides were detected in both extract pools (#1, 3, 4, and #2). In cases where no peptides could be selected from the iTRAQ results, peptides were selected from the GPMDb database (on the basis of being the most commonly used in MRM experiments within the GPMDb). The list of peptides was then analyzed for cases of very hydrophilic or hydrophobic peptides (which were avoided to prevent potential chromatography issues). Only one selected peptide had a hydrophobicity score outside of the range $-0.6 \le x \le 0.6$. All peptides were of a measurable length (7-30 amino acids) for the 4000Q MS (ABI) used (141). Two peptides were chosen for all proteins except Osteomodulin (as there was only one satisfactory peptide from the iTRAQ results and no GPMDb entry at the time of selection). Modifications on the selected peptides were not considered, as there was no evidence of potential modifications on these.

4.3.2. Peptide Transition Selection and Optimization

As discussed in section **4.3.**, typically at least two transitions per peptide are selected to improve precision of quantitation and account for potentially more than one charge state per fragment ion (142). I used 5-10 transitions for each peptide (**Table 4.3**) to have increased precision of measurement. Synthetic peptide equivalents of the peptides listed in **Table 4.2** were analyzed by MS/MS to identify all fragment ions produced. The three fragment ions with the highest intensity (in both predicted and acquired data) were selected as transitions. In most cases only the +2 charge state was detected reliably but occasionally the +3 state fragment ions also had a strong signal. In such cases, 6 transitions were selected (3 from both +2 and +3 parent ion charge states). +1 charge state peptides were infrequent and not used due to their poor fragmentation. The final list of transitions used is given in **Table 4.3**.

Selected peptides were synthesized by PEPTIDE 2.0 (www.peptide2.com/) as crude samples. Each peptide was then desalted on SepPak C18 light columns (Waters) and resuspended in an appropriate solution to facilitate solubility (either 2% formic acid, 2% NH₄OH, 2% acetonitrile, or 30% acetonitrile and 0.1% formic acid). The peptides were then combined into pools of 4-9 to be more efficient for optimization on the mass spectrometer. Each peptide pool was first analyzed by MS/MS to identity each peptide in the pool. The MASCOT results were analyzed by MultiQuant to determine the three best transitions for each peptide to be queried in the MRM experiment. Then a second run was used to establish optimal voltages for the declustering potential (DP), and a third run was used to determine optimal collision energy (CE) and escape potential (CXP) by static infusion of the synthetic peptide mix.

The optimized parameters were then tested on another pool of all 46 peptides together to obtain LC retention time windows for each peptide. Retention times were used to establish a time scheduling MRM method. The peptides were tested for detection using the time scheduled MRM method when spiked into an arbitrary precipitated total protein from HEK 293 cell lysate with each peptide at 1, 8, 100, or 450 fmol per LC injection. This helped to establish the likely minimum detection range for each peptide. Finally, the method was tested on practice samples from real growth plate and cranial plate lysates before interrogation of the experimental samples.

Name	Transition	Target RT
Acan1.Acan 1.ARPNCGGNLLGVR.b10++.light	442.9 -> 498.8	15.3
Acan1.Acan 1.ARPNCGGNLLGVR.b2.light	442.9 -> 228.1	15.3
Acan1.Acan 1.ARPNCGGNLLGVR.b8.light	442.9 -> 770.3	15.3
Acan1.Acan 1.ARPNCGGNLLGVR.y3.light	442.9 -> 331.2	15.3
Acan1.Acan 1.ARPNCGGNLLGVR.y4++.light	442.9 -> 222.7	15.3
Acan2.Acan 2.YQCTEGFVQR.b9++.light	615.8 -> 528.7	21.468

Table 4.3: Transitions and LC retention times (RT) for all peptides in the MRM panel. Naming is according to parent protein.peptide abbreviation. peptide sequence. b or y ion. charge. light or heavy lable standard.

Name	Transition	Target RT
Acan2.Acan 2.YQCTEGFVQR.y5.light	615.8 -> 606.3	21.468
Acan2.Acan 2.YQCTEGFVQR.y6.light	615.8 -> 735.4	21.468
Acan2.Acan 2.YQCTEGFVQR.y7.light	615.8 -> 836.4	21.468
Acan2.Acan 2.YQCTEGFVQR.y7++.light	615.8 -> 418.7	21.468
Annexin A1.Anxa1 1.GVDEATIIDILTK.b3.light	694.4 -> 272.1	36.941
Annexin A1.Anxa1 1.GVDEATIIDILTK.y2.light	694.4 -> 248.2	36.941
Annexin A1.Anxa1 1.GVDEATIIDILTK.y6.light	694.4 -> 702.4	36.941
Annexin A1.Anxa1 1.GVDEATIIDILTK.y7.light	694.4 -> 815.5	36.941
Annexin A1.Anxa1 1.GVDEATIIDILTK.y8.light	694.4 -> 916.6	36.941
Annexin A1.Anxa1 2.TPAQFDADELR.y10++.light	631.8 -> 581.3	19.21
Annexin A1.Anxa1 2.TPAQFDADELR.y2.light	631.8 -> 288.2	19.21
Annexin A1.Anxa1 2.TPAQFDADELR.y5.light	631.8 -> 603.3	19.21
Annexin A1.Anxa1 2.TPAQFDADELR.y6.light	631.8 -> 718.3	19.21
Annexin A1.Anxa1 2.TPAQFDADELR.y7.light	631.8 -> 865.4	19.21
Asporin.Aspn 1.LYLSHNQLSEIPLNLPK.b10++.light	660.4 -> 593.3	35.229
Asporin.Aspn 1.LYLSHNQLSEIPLNLPK.b11++.light	660.4 -> 649.8	35.229
Asporin.Aspn 1.LYLSHNQLSEIPLNLPK.y2.light	660.4 -> 244.2	35.229
Asporin.Aspn 1.LYLSHNQLSEIPLNLPK.y6.light	660.4 -> 681.4	35.229
Asporin.Aspn 1.LYLSHNQLSEIPLNLPKy6++.light	660.4 -> 341.2	35.229
Asporin.Aspn 2.YWEIQPATFR.b2.light	655.8 -> 350.1	31.67
Asporin.Aspn 2.YWEIQPATFR.b3.light	655.8 -> 479.2	31.67
Asporin.Aspn 2.YWEIQPATFR.y2.light	655.8 -> 322.2	31.67
Asporin.Aspn 2.YWEIQPATFR.y7.light	655.8 -> 832.5	31.67
Asporin.Aspn 2.YWEIQPATFR.y8.light	655.8 -> 961.5	31.67
Biglycan.Bgn 1.NHLVEIPPNLPSSLVELR.b6.light	676.4 -> 706.4	35.877
Biglycan.Bgn 1.NHLVEIPPNLPSSLVELR.y10++.light	676.4 -> 564.3	35.877
Biglycan.Bgn 1.NHLVEIPPNLPSSLVELR.y12++.light	676.4 -> 661.4	35.877
Biglycan.Bgn 1.NHLVEIPPNLPSSLVELR.y8.light	676.4 -> 900.5	35.877
Biglycan.Bgn 1.NHLVEIPPNLPSSLVELR.y8++.light	676.4 -> 450.8	35.877
Biglycan.Bgn 2.EISPDTTLLDLQNNDISELR.y16+++.light	762.7 -> 620.6	36.039
Biglycan.Bgn 2.EISPDTTLLDLQNNDISELR.y17+++.light	762.7 -> 653.0	36.039
Biglycan.Bgn 2.EISPDTTLLDLQNNDISELR.y2.light	762.7 -> 288.2	36.039
Biglycan.Bgn 2.EISPDTTLLDLQNNDISELR.y4.light	762.7 -> 504.3	36.039
Biglycan.Bgn 2.EISPDTTLLDLQNNDISELR.y8.light	762.7 -> 960.5	36.039
Chondroadherin.Chad 1.SIPDNAFQSFGR.b2.light	669.8 -> 201.1	27.333
Chondroadherin.Chad 1.SIPDNAFQSFGR.y10++.light	669.8 -> 569.8	27.333
Chondroadherin.Chad 1.SIPDNAFQSFGR.y4.light	669.8 -> 466.2	27.333
Chondroadherin.Chad 1.SIPDNAFQSFGR.y6.light	669.8 -> 741.4	27.333
Chondroadherin.Chad 1.SIPDNAFQSFGR.y8.light	669.8 -> 926.4	27.333
Chondroadherin.Chad 2.YLETLWLDNTNLEK.b2.light	876.4 -> 277.2	35.738

Name	Transition	Target RT
Chondroadherin.Chad 2.YLETLWLDNTNLEK.b3.light	876.4 -> 406.2	35.738
Chondroadherin.Chad 2.YLETLWLDNTNLEK.y7.light	876.4 -> 833.4	35.738
Chondroadherin.Chad 2.YLETLWLDNTNLEK.y8.light	876.4 -> 946.5	35.738
Chondroadherin.Chad 2.YLETLWLDNTNLEK.y9.light	876.4 -> 1132.6	35.738
Col1a11.Col1a1 1.SAGVSVPGPMGPSGPR.b3.light	726.9 -> 216.1	20.494
Col1a11.Col1a1 1.SAGVSVPGPMGPSGPR.b4.light	726.9 -> 315.2	20.494
Col1a11.Col1a1 1.SAGVSVPGPMGPSGPR.b6.light	726.9 -> 501.3	20.494
Col1a11.Col1a1 1.SAGVSVPGPMGPSGPR.y10.light	726.9 -> 952.5	20.494
Col1a11.Col1a1 1.SAGVSVPGPMGPSGPR.y10++.light	726.9 -> 476.7	20.494
Col1a12.Col1a1 2.GETGPAGPAGPIGPAGAR.y11.light	766.9 -> 963.5	15.203
Col1a12.Col1a1 2.GETGPAGPAGPIGPAGAR.y12.light	766.9 -> 1020.6	15.203
Col1a12.Col1a1 2.GETGPAGPAGPIGPAGAR.y14++.light	766.9 -> 594.8	15.203
Col1a12.Col1a1 2.GETGPAGPAGPIGPAGAR.y8.light	766.9 -> 738.4	15.203
Col1a12.Col1a1 2.GETGPAGPAGPIGPAGAR.y9.light	766.9 -> 795.4	15.203
Col2a11.Col2a1 1.GAQGPPGATGFPGAAGR.y12.light	734.9 -> 1058.5	16.16
Col2a11.Col2a1 1.GAQGPPGATGFPGAAGR.y13++.light	734.9 -> 578.3	16.16
Col2a11.Col2a1 1.GAQGPPGATGFPGAAGR.y14++.light	734.9 -> 606.8	16.16
Col2a11.Col2a1 1.GAQGPPGATGFPGAAGR.y6.light	734.9 -> 528.3	16.16
Col2a11.Col2a1 1.GAQGPPGATGFPGAAGR.y9.light	734.9 -> 833.4	16.16
Col2a12.Col2a1 2.GFPGLPGPSGEPGK.b12++.light	648.8 -> 547.3	25.379
Col2a12.Col2a1 2.GFPGLPGPSGEPGK.b5.light	648.8 -> 472.3	25.379
Col2a12.Col2a1 2.GFPGLPGPSGEPGK.y12++.light	648.8 -> 546.8	25.379
Col2a12.Col2a1 2.GFPGLPGPSGEPGK.y3.light	648.8 -> 301.2	25.379
Col2a12.Col2a1 2.GFPGLPGPSGEPGK.2.y9.light	648.8 -> 825.4	25.379
COMP.Thbs 5 1.DTDLDGFPDEK.b10++.light	626.3 -> 553.2	19.033
COMP.Thbs 5 1.DTDLDGFPDEK.b4.light	626.3 -> 445.2	19.033
COMP.Thbs 5 1.DTDLDGFPDEK.b4++.light	626.3 -> 223.1	19.033
COMP.Thbs 5 1.DTDLDGFPDEK.b8++.light	626.3 -> 431.2	19.033
COMP.Thbs 5 1.DTDLDGFPDEK.y2.light	626.3 -> 276.2	19.033
COMP.Thbs 5 2.ELQETNAALQDVR.y2.light	743.9 -> 274.2	17.581
COMP.Thbs 5 2.ELQETNAALQDVR.y4.light	743.9 -> 517.3	17.581
COMP.Thbs 5 2.ELQETNAALQDVR.y6.light	743.9 -> 701.4	17.581
COMP.Thbs 5 2.ELQETNAALQDVR.y8.light	743.9 -> 886.5	17.581
COMP.Thbs 5 2.ELQETNAALQDVR.y9.light	743.9 -> 987.5	17.581
Decorin.Dcn 1.ASYSAVSLYGNPVR.y3.light	742.4 -> 371.2	24.943
Decorin.Dcn 1.ASYSAVSLYGNPVR.y5.light	742.4 -> 542.3	24.943
Decorin.Dcn 1.ASYSAVSLYGNPVR.y6.light	742.4 -> 705.4	24.943
Decorin.Dcn 1.ASYSAVSLYGNPVR.y8.light	742.4 -> 905.5	24.943
Decorin.Dcn 1.ASYSAVSLYGNPVR.y9.light	742.4 -> 1004.6	24.943
Decorin.Dcn 2.YWEIFPNTFR.b2.light	686.8 -> 350.1	36.409

Name	Transition	Target RT
Decorin.Dcn 2.YWEIFPNTFR.b3.light	686.8 -> 479.2	36.409
Decorin.Dcn 2.YWEIFPNTFR.y2.light	686.8 -> 322.2	36.409
Decorin.Dcn 2.YWEIFPNTFR.y6.light	686.8 -> 781.4	36.409
Decorin.Dcn 2.YWEIFPNTFR.y8.light	686.8 -> 1023.5	36.409
Epiphycan.Epyc 1.IDLTSNLISEIDEDAFR.y10.light	976.0 -> 1194.6	37.398
Epiphycan.Epyc 1.IDLTSNLISEIDEDAFR.y3.light	976.0 -> 393.2	37.398
Epiphycan.Epyc 1.IDLTSNLISEIDEDAFR.y5.light	976.0 -> 637.3	37.398
Epiphycan.Epyc 1.IDLTSNLISEIDEDAFR.y6.light	976.0 -> 752.3	37.398
Epiphycan.Epyc 1.IDLTSNLISEIDEDAFR.y9.light	976.0 -> 1081.5	37.398
Epiphycan.Epyc 2.QLPELPNTLTFIDISNNR.b12++.light	695.7 -> 684.4	36.999
Epiphycan.Epyc 2.QLPELPNTLTFIDISNNR.y16++.light	695.7 -> 922.5	36.999
Epiphycan.Epyc 2.QLPELPNTLTFIDISNNR.y17+++.light	695.7 -> 653.0	36.999
Epiphycan.Epyc 2.QLPELPNTLTFIDISNNR.y5.light	695.7 -> 603.3	36.999
Epiphycan.Epyc 2.QLPELPNTLTFIDISNNR.y6.light	695.7 -> 718.3	36.999
Fibromodulin.Fmod 1.IPPVNTNLENLYLQGNR.b2.light	978.0 -> 211.1	33.894
Fibromodulin.Fmod 1.IPPVNTNLENLYLQGNR.y15++.light	978.0 -> 873.0	33.894
Fibromodulin.Fmod 1.IPPVNTNLENLYLQGNR.y16++.light	978.0 -> 921.5	33.894
Fibromodulin.Fmod 1.IPPVNTNLENLYLQGNR.y3.light	978.0 -> 346.2	33.894
Fibromodulin.Fmod 1.IPPVNTNLENLYLQGNR.y5++.light	978.0 -> 294.2	33.894
Fibromodulin.Fmod 2.SLILLDLSYNHLR.b2.light	519.6 -> 201.1	35.962
Fibromodulin.Fmod 2.SLILLDLSYNHLR.b3.light	519.6 -> 314.2	35.962
Fibromodulin.Fmod 2.SLILLDLSYNHLR.y10++.light	519.6 -> 622.3	35.962
Fibromodulin.Fmod 2.SLILLDLSYNHLR.y11++.light	519.6 -> 678.9	35.962
Fibromodulin.Fmod 2.SLILLDLSYNHLR.y9++.light	519.6 -> 565.8	35.962
Fibronectin 1.Fn 2.TFYQIGDSWEK.b2.light	687.3 -> 249.1	28.454
Fibronectin 1.Fn 2.TFYQIGDSWEK.y6.light	687.3 -> 721.3	28.454
Fibronectin 1.Fn 2.TFYQIGDSWEK.y7.light	687.3 -> 834.4	28.454
Fibronectin 1.Fn 2.TFYQIGDSWEK.y9.light	687.3 -> 1125.5	28.454
Fibronectin 1.Fn 2.TFYQIGDSWEK.y9++.light	687.3 -> 563.3	28.454
Fibronectin.Fn 1.FLTTTPNSLLVSWQAPR.b2.light	966.0 -> 261.2	36.449
Fibronectin.Fn 1.FLTTTPNSLLVSWQAPR.b3.light	966.0 -> 362.2	36.449
Fibronectin.Fn 1.FLTTTPNSLLVSWQAPR.b5.light	966.0 -> 564.3	36.449
Fibronectin.Fn 1.FLTTTPNSLLVSWQAPR.y6.light	966.0 -> 744.4	36.449
Fibronectin.Fn 1.FLTTTPNSLLVSWQAPR.y8.light	966.0 -> 956.5	36.449
Lactotransferrin.Ltf 1.LRPVAAEVYGTK.b5.light	435.3 -> 537.4	13.885
Lactotransferrin.Ltf 1.LRPVAAEVYGTK.b7.light	435.3 -> 737.4	13.885
Lactotransferrin.Ltf 1.LRPVAAEVYGTK.b7++.light	435.3 -> 369.2	13.885
Lactotransferrin.Ltf 1.LRPVAAEVYGTK.y5.light	435.3 -> 567.3	13.885
Lactotransferrin.Ltf 1.LRPVAAEVYGTK.y6.light	435.3 -> 696.4	13.885
Lactotransferrin.Ltf 2.STNDKEEAIWELLR.b10++.light	568.6 -> 587.8	35.422

Name	Transition	Target RT
Lactotransferrin.Ltf 2.STNDKEEAIWELLR.y4.light	568.6 -> 530.3	35.422
Lactotransferrin.Ltf 2.STNDKEEAIWELLR.y5.light	568.6 -> 716.4	35.422
Lactotransferrin.Ltf 2.STNDKEEAIWELLR.y5++.light	568.6 -> 358.7	35.422
Lactotransferrin.Ltf 2.STNDKEEAIWELLR.y6++.light	568.6 -> 415.3	35.422
Lect2.Lect2 1.HHPGVDVLCSDGSVVYAPFTGK.b14++.light	762.4 -> 702.3	32.538
Lect2.Lect2 1.HHPGVDVLCSDGSVVYAPFTGK.b15++.light	762.4 -> 751.9	32.538
Lect2.Lect2 1.HHPGVDVLCSDGSVVYAPFTGK.y5++.light	762.4 -> 275.2	32.538
Lect2.Lect2 1.HHPGVDVLCSDGSVVYAPFTGK.y7.light	762.4 -> 783.4	32.538
Lect2.Lect2 1.HHPGVDVLCSDGSVVYAPFTGK.y8.light	762.4 -> 882.5	32.538
LGALS 3.Lgals3 1.IQVLVEADHFK.b2.light	433.6 -> 242.1	24.452
LGALS 3.Lgals3 1.IQVLVEADHFK.y10++.light	433.6 -> 593.3	24.452
LGALS 3.Lgals3 1.IQVLVEADHFK.y6.light	433.6 -> 746.3	24.452
LGALS 3.Lgals3 1.IQVLVEADHFK.y7.light	433.6 -> 845.4	24.452
LGALS 3.Lgals3 1.IQVLVEADHFK.y9++.light	433.6 -> 529.3	24.452
LGALS 3.Lgals3 2.GNDVAFHFNPR.y4.light	425.2 -> 533.3	19.11
LGALS 3.Lgals3 2.GNDVAFHFNPR.y6++.light	425.2 -> 409.2	19.11
LGALS 3.Lgals3 2.GNDVAFHFNPR.y7++.light	425.2 -> 444.7	19.11
LGALS 3.Lgals3 2.GNDVAFHFNPR.y8++.light	425.2 -> 494.3	19.11
LGALS 3.Lgals3 2.GNDVAFHFNPR.y9++.light	425.2 -> 551.8	19.11
Lumican.Lum 1.ISNIPDEYFK.b2.light	613.3 -> 201.1	25.506
Lumican.Lum 1.ISNIPDEYFK.b3.light	613.3 -> 315.2	25.506
Lumican.Lum 1.ISNIPDEYFK.b4.light	613.3 -> 428.3	25.506
Lumican.Lum 1.ISNIPDEYFK.y6.light	613.3 -> 798.4	25.506
Lumican.Lum 1.ISNIPDEYFK.y8.light	613.3 -> 1025.5	25.506
Lumican.Lum 2.NNQIDHIDEK.y2.light	409.2 -> 276.2	13.117
Lumican.Lum 2.NNQIDHIDEK.y3.light	409.2 -> 391.2	13.117
Lumican.Lum 2.NNQIDHIDEK.y6++.light	409.2 -> 378.7	13.117
Lumican.Lum 2.NNQIDHIDEK.y7++.light	409.2 -> 435.2	13.117
Lumican.Lum 2.NNQIDHIDEK.y9++.light	409.2 -> 556.3	13.117
Matrilin 3.Matn3 1.ASGIELYAVGVDR.b4.light	675.4 -> 329.2	28.637
Matrilin 3.Matn3 1.ASGIELYAVGVDR.y5.light	675.4 -> 545.3	28.637
Matrilin 3.Matn3 1.ASGIELYAVGVDR.y6.light	675.4 -> 616.3	28.637
Matrilin 3.Matn3 1.ASGIELYAVGVDR.y7.light	675.4 -> 779.4	28.637
Matrilin 3.Matn3 1.ASGIELYAVGVDR.y8.light	675.4 -> 892.5	28.637
Matrilin 3.Matn3 2.VAVVNYASTVK.b3.light	575.8 -> 270.2	17.613
Matrilin 3.Matn3 2.VAVVNYASTVK.y7.light	575.8 -> 782.4	17.613
Matrilin 3.Matn3 2.VAVVNYASTVK.y8.light	575.8 -> 881.5	17.613
Matrilin 3.Matn3 2.VAVVNYASTVK.y9.light	575.8 -> 980.5	17.613
Matrilin 3.Matn3 2.VAVVNYASTVK.y9++.light	575.8 -> 490.8	17.613
MFGE8.Mfge8 1.INAWTAQSNSAK.b2.light	645.8 -> 228.1	10.243

Name	Transition	Target RT
MFGE8.Mfge8 1.INAWTAQSNSAK.b3.light	645.8 -> 299.2	10.243
MFGE8.Mfge8 1.INAWTAQSNSAK.y7.light	645.8 -> 705.4	10.243
MFGE8.Mfge8 1.INAWTAQSNSAK.y8.light	645.8 -> 806.4	10.243
MFGE8.Mfge8 1.INAWTAQSNSAK.y9.light	645.8 -> 992.5	10.243
MFGE8.Mfge8 2.VSGVMTQGASR.b3.light	546.8 -> 244.1	6.221
MFGE8.Mfge8 2.VSGVMTQGASR.b4.light	546.8 -> 343.2	6.221
MFGE8.Mfge8 2.VSGVMTQGASR.y6.light	546.8 -> 619.3	6.221
MFGE8.Mfge8 2.VSGVMTQGASR.y7.light	546.8 -> 750.4	6.221
MFGE8.Mfge8 2.VSGVMTQGASR.y9.light	546.8 -> 906.4	6.221
Osteoglycin.Ogn 1.LDFTGNLIEDIEDGTFSK.b7.light	672.0 -> 761.4	37.779
Osteoglycin.Ogn 1.LDFTGNLIEDIEDGTFSK.y10++.light	672.0 -> 570.8	37.779
Osteoglycin.Ogn 1.LDFTGNLIEDIEDGTFSK.y5.light	672.0 -> 539.3	37.779
Osteoglycin.Ogn 1.LDFTGNLIEDIEDGTFSK.y6.light	672.0 -> 654.3	37.779
Osteoglycin.Ogn 1.LDFTGNLIEDIEDGTFSK.y7.light	672.0 -> 783.4	37.779
Osteoglycin.Ogn 1.RLDFTGNLIEDIEDGTFSK.b3.light	724.0 -> 385.2	37.056
Osteoglycin.Ogn 1.RLDFTGNLIEDIEDGTFSK.b7.light	724.0 -> 804.4	37.056
Osteoglycin.Ogn 1.RLDFTGNLIEDIEDGTFSK.b8.light	724.0 -> 917.5	37.056
Osteoglycin.Ogn 1.RLDFTGNLIEDIEDGTFSK.b8++.light	724.0 -> 459.2	37.056
Osteoglycin.Ogn 1.RLDFTGNLIEDIEDGTFSK.b9++.light	724.0 -> 515.8	37.056
Osteomodulin.Omd 1.IHSIYYGEQR.b2.light	422.5 -> 251.2	10.148
Osteomodulin.Omd 1.IHSIYYGEQR.y2.light	422.5 -> 303.2	10.148
Osteomodulin.Omd 1.IHSIYYGEQR.y3++.light	422.5 -> 216.6	10.148
Osteomodulin.Omd 1.IHSIYYGEQR.y4.light	422.5 -> 489.2	10.148
Osteomodulin.Omd 1.IHSIYYGEQR.y5.light	422.5 -> 652.3	10.148
Periostin.Postn 1.GFEPGVTNILK.b2.light	587.8 -> 205.1	29.781
Periostin.Postn 1.GFEPGVTNILK.b3.light	587.8 -> 334.1	29.781
Periostin.Postn 1.GFEPGVTNILK.b4++.light	587.8 -> 216.1	29.781
Periostin.Postn 1.GFEPGVTNILK.y8.light	587.8 -> 841.5	29.781
Periostin.Postn 1.GFEPGVTNILK.y8++.light	587.8 -> 421.3	29.781
Periostin.Postn 2.IIDGVPVEITEK.b2.light	656.9 -> 227.2	28.107
Periostin.Postn 2.IIDGVPVEITEK.b4.light	656.9 -> 399.2	28.107
Periostin.Postn 2.IIDGVPVEITEK.b5.light	656.9 -> 498.3	28.107
Periostin.Postn 2.IIDGVPVEITEK.y3.light	656.9 -> 377.2	28.107
Periostin.Postn 2.IIDGVPVEITEK.y7.light	656.9 -> 815.5	28.107
Perlecan.Hspg2 1.SLPEVPETIEFEVR.b2.light	549.0 -> 201.1	35.543
Perlecan.Hspg2 1.SLPEVPETIEFEVR.b4++.light	549.0 -> 214.1	35.543
Perlecan.Hspg2 1.SLPEVPETIEFEVR.b6++.light	549.0 -> 312.2	35.543
Perlecan.Hspg2 1.SLPEVPETIEFEVR.y4.light	549.0 -> 550.3	35.543
Perlecan.Hspg2 1.SLPEVPETIEFEVR.y5+++.light	549.0 -> 227.1	35.543
Perlecan.Hspg2 2.LVSEDPINDGEWHR.b2.light	556.3 -> 213.2	31.268

Name	Transition	Target RT
Perlecan.Hspg2 2.LVSEDPINDGEWHR.b4++.light	556.3 -> 215.1	31.268
Perlecan.Hspg2 2.LVSEDPINDGEWHR.b6.light	556.3 -> 641.3	31.268
Perlecan.Hspg2 2.LVSEDPINDGEWHR.y4+++.light	556.3 -> 209.8	31.268
Perlecan.Hspg2 2.LVSEDPINDGEWHR.y6+++.light	556.3 -> 267.1	31.268
PRELP.Prelp 1.LENLLLDLQHNR.b2.light	531.0 -> 243.1	35.13
PRELP.Prelp 1.LENLLLLDLQHNR.y11++.light	531.0 -> 674.9	35.13
PRELP.Prelp 1.LENLLLLDLQHNR.y12++.light	531.0 -> 739.4	35.13
PRELP.Prelp 1.LENLLLLDLQHNR.y8++.light	531.0 -> 504.8	35.13
PRELP.Prelp 1.LENLLLLDLQHNR.y9++.light	531.0 -> 561.3	35.13
PRELP.Prelp 2.NQLEEVPSALPR.b4++.light	676.9 -> 243.1	24.301
PRELP.Prelp 2.NQLEEVPSALPR.y10.light	676.9 -> 1110.6	24.301
PRELP.Prelp 2.NQLEEVPSALPR.y4++.light	676.9 -> 228.7	24.301
PRELP.Prelp 2.NQLEEVPSALPR.y6.light	676.9 -> 640.4	24.301
PRELP.Prelp 2.NQLEEVPSALPR.y9.light	676.9 -> 997.5	24.301
Serpin F1.Serpinf1 1.DTDTGALLFIGR.b8.light	639.8 -> 787.4	34.76
Serpin F1.Serpinf1 1.DTDTGALLFIGR.y3.light	639.8 -> 345.2	34.76
Serpin F1.Serpinf1 1.DTDTGALLFIGR.y4.light	639.8 -> 492.3	34.76
Serpin F1.Serpinf1 1.DTDTGALLFIGR.y5.light	639.8 -> 605.4	34.76
Serpin F1.Serpinf1 1.DTDTGALLFIGR.y6++.light	639.8 -> 359.7	34.76
Serpin F1.Serpinf1 2.LAAAVSNFGYDLYR.b4.light	780.4 -> 327.2	32.832
Serpin F1.Serpinf1 2.LAAAVSNFGYDLYR.y2.light	780.4 -> 338.2	32.832
Serpin F1.Serpinf1 2.LAAAVSNFGYDLYR.y6.light	780.4 -> 786.4	32.832
Serpin F1.Serpinf1 2.LAAAVSNFGYDLYR.y7.light	780.4 -> 933.4	32.832
Serpin F1.Serpinf1 2.LAAAVSNFGYDLYR.y9.light	780.4 -> 1134.5	32.832
SPARC.Sparc 1.LEAGDHPVELLAR.y10++.light	473.9 -> 553.8	21.777
SPARC.Sparc 1.LEAGDHPVELLAR.y11++.light	473.9 -> 589.3	21.777
SPARC.Sparc 1.LEAGDHPVELLAR.y12++.light	473.9 -> 653.8	21.777
SPARC.Sparc 1.LEAGDHPVELLAR.y7.light	473.9 -> 797.5	21.777
SPARC.Sparc 1.LEAGDHPVELLAR.y8++.light	473.9 -> 467.8	21.777
SPARC.Sparc 2.NVLVTLYERDEGNNLLTEK.b3.light	740.7 -> 327.2	33.551
SPARC.Sparc 2.NVLVTLYERDEGNNLLTEK.y13++.light	740.7 -> 790.9	33.551
SPARC.Sparc 2.NVLVTLYERDEGNNLLTEK.y15++.light	740.7 -> 897.9	33.551
SPARC.Sparc 2.NVLVTLYERDEGNNLLTEK.y16++.light	740.7 -> 947.5	33.551
SPARC.Sparc 2.NVLVTLYERDEGNNLLTEK.y17++.light	740.7 -> 1004.0	33.551
Thrombospondin 1.Thbs 1 2.FVFGTTPEDILR.b2.light	697.9 -> 247.1	35.87
Thrombospondin 1.Thbs 1 2.FVFGTTPEDILR.y10.light	697.9 -> 1148.6	35.87
Thrombospondin 1.Thbs 1 2.FVFGTTPEDILR.y10++.light	697.9 -> 574.8	35.87
Thrombospondin 1.Thbs 1 2.FVFGTTPEDILR.y6.light	697.9 -> 742.4	35.87
Thrombospondin 1.Thbs 1 2.FVFGTTPEDILR.y9.light	697.9 -> 1001.5	35.87
Thrombospondin 1.Thbs1 1.AGTLDLSLSLPGK.y2.light	636.4 -> 204.1	33.811

Name	Transition	Target RT
Thrombospondin 1.Thbs1 1.AGTLDLSLSLPGK.y3.light	636.4 -> 301.2	33.811
Thrombospondin 1.Thbs1 1.AGTLDLSLSLPGK.y5.light	636.4 -> 501.3	33.811
Thrombospondin 1.Thbs1 1.AGTLDLSLSLPGK.y7.light	636.4 -> 701.4	33.811
Thrombospondin 1.Thbs1 1.AGTLDLSLSLPGK.y9.light	636.4 -> 929.5	33.811
Thrombospondin 3.Thbs3 1.QTEQTYWQATPFR.y3.light	828.4 -> 419.2	28.438
Thrombospondin 3.Thbs3 1.QTEQTYWQATPFR.y4.light	828.4 -> 520.3	28.438
Thrombospondin 3.Thbs3 1.QTEQTYWQATPFR.y6.light	828.4 -> 719.4	28.438
Thrombospondin 3.Thbs3 1.QTEQTYWQATPFR.y7.light	828.4 -> 905.5	28.438
Thrombospondin 3.Thbs3 1.QTEQTYWQATPFR.y8.light	828.4 -> 1068.5	28.438
Thrombospondin 3.Thbs3 2.FYVVMWK.b2.light	486.8 -> 311.1	33.707
Thrombospondin 3.Thbs3 2.FYVVMWK.y2.light	486.8 -> 333.2	33.707
Thrombospondin 3.Thbs3 2.FYVVMWK.y3.light	486.8 -> 464.2	33.707
Thrombospondin 3.Thbs3 2.FYVVMWK.y4.light	486.8 -> 563.3	33.707
Thrombospondin 3.Thbs3 2.FYVVMWK.y5.light	486.8 -> 662.4	33.707

The total number of transitions in the original MRM assay was 159. The mass spectrometer scanned for all 53 parent ions in Q1 and all 159 daughter ions in MS Q3. This large scanning demand meant that less time could be spent sampling the distribution of each transition/peptide as it eluted from the LC; this is known as the dwell time (141). For 159 transitions, the dwell time was ~10ms, which, while acceptable, reduced the sensitivity of quantification as the shape of the distribution is less accurate with fewer data points (141). To avoid this issue, 'time scheduling' can be applied to the MRM assay parameters wherein retention time data for each of the peptides is used to specify discrete time windows in which the MS will scan for specific transitions. This reduces the number of transitions scanned at any given time, and means that the dwell time can be lower and sensitivity higher (though the absolute dwell time will be determined in real-time based on the number of peptides that appear in a window (141)). The retention time of each peptide was obtained from a mix of the synthetic peptides and used to center a ±4 minute scan window for each peptide in the real samples. The scan window was intentionally large to allow for elution differences between real samples with a complex background compared to the pure synthetic peptides.

While the above parameters were used to set up and analyze the original iTRAQ protein samples, another version of the MRM assay was developed at the University of Victoria Genome BC Proteomics Facility. This second and final version used for validation of iTRAQ candidates was performed at the Genome BC Proteomics Facility and used 5-10 transitions rather than the 3-6 transitions per peptide used in the initial MRM panel described above. The optimization data from the first MRM panel was the basis of the version utilized at the Genome BC facility. Both versions yielded usable data from growth plate protein samples, however issues with the machine used for the first version required the method to be moved to the Genome BC facility before validation could be completed. The improved technical capabilities of the Genome BC facility enabled the increased transition monitoring.

4.3.3. MRM Validation of iTRAQ Results

The quality of data obtained in an MRM experiment can be estimated by trends in the resulting data, which ultimately showed that the MRM data obtained in this project were robust. As discussed in section **4.3.**, transitions from a parent peptide in an MRM experiment should sum to represent the original concentration of the parent peptide. These transitions should stack upon each other around the retention time of the parent peptide as shown in **Figure 4.3**. In addition, each transition when assayed in several samples should be approximately equivalent proportionally to each other. For example, transition 1 of peptide A should have an approximately equal intensity ratio to transition 2 of peptide A in all samples. If transitions do not show equivalent relative intensities between samples, there may have been an issue with sample loading, sample preparation, or transition reliability.

a) Total Peptide Mix



Figure 4.3: Example of MRM quantitation of two proteins in total protein lysates. a) elution profile of all peptides in the sample and extracted peaks for peptides of Thrombospondin 3 and Fibronectin 1 (b and c respectively). Quantification of each transition is achieved by measuring the area under the curve for each transition (shaded in blue) (d and e). Note that transition peaks stack on the same elution time because they come from the same parent peptide.

As **Figure 4.4** shows, the relative intensities of each transition for each peptide is approximately proportional between samples, indicating robustness of data and performance of transitions. Furthermore, peptide quantitation variance within the MPS I and wildtype sample groups was acceptable (p<0.05) (**Figure 4.4**), also suggesting robustness of data for the significant peptides discussed. Peptides not represented in **Figure 4.4** either did not show a significant difference between MPS I and wildtype quantity or did not have acceptable intragroup variance.

After ascertaining the quality of the MRM data, trends of relative increased or decreased peptide quantity between the MPS I and wildtype samples were examined across all transitions. This identified six proteins that were significantly altered (p<0.05) in the MPS I growth plate as shown in **Figure 4.4**. These results validated the decreased protein levels identified by iTRAQ analysis for the SLRPs biglycan, fibromodulin, and PRELP. The MRM results also identified decreased levels of collagen I and SERPIN F1 that were not shown by iTRAQ (133). Due to the higher sensitivity and precision of quantitation by MRM compared to iTRAQ, the decreases in collagen I and SERPIN F1 was considered sufficient evidence to include these proteins as candidates (**Table 4.4**). The MRM panel did not identify any significant changes in the 3 week femoral head growth plate between MPS I and wildtype samples (data not shown) (133).

Table 4.4:	Selected MRN	1 proteins not from	iTRAQ candidates.
		1	

Protein (Gene Name)	UniProt	Function
Aggrecan	P16112	Structural
Collagen 1a1	P02452	Structural
Collagen 2a1	P02458	Structural
Asporin	Q9BXN1	Matricellular
Biglycan	P21810	Matricellular
Decorin	P07585	Matricellular
Fibromodulin	Q06828	Matricellular
Osteomodulin	Q99983	Matricellular
SPARC	P09486	Matricellular
Thrombospondin 3	P49746	Matricellular
LGALS3	<u>P17931</u>	Lectin











Figure 4.4: MRM transition areas for significantly decreased proteins comparing the 5 week MPS I to wildtype growth plate. Both peptides for each protein are shown, along with all transition areas for each individual mouse in the MPS I (M1...M6) and wildtype (WT1...WT5) groups. Each colored bar represents a single transition. The transition area is measured in (counts)*(minutes) units. Abbreviations: Biglycan (Bgn), collagen 1a1 (Col1a1), fibromodulin (Fmod), lactotransferrin (Ltf), PRELP (Prelp), SERPIN F1 (Serpinf1).

Table 4.5: Summary of MRM quantitation results in the 5 week MPS I growth plate compared to wildtype indicating significant protein abundance changes. Ratios (fold change) and p-values are for the ratio of protein abundance in MPS I:wildtype.

Protein	Peptides	Fold	P-val
Biglycan	Bgn1	-1.68	4.10E-06
	Bgn2	-1.44	0.00067
Collagen 1	Col1a1 1	-1.42	0.0011
	Col1a1 2	-1.34	6.70E-09
Fibromodulin	Fmod 1	-1.01	0.21
	Fmod 2	-1.39	6.00E-07
Lactotransferrin	Ltf 1	-1.26	5.10E-06
	Ltf 2	-1.54	1.80E-08
PRELP	Prelp1	-1.21	1.40E-03
	Prelp2	-1.32	5.60E-03
Serpin F1	Serpinf1 1	-1.64	1.80E-05
	Serpinf1 2	-2.03	2.90E-07

4.4. Immunohistochemical Femoral Head Analysis

Immunohistochemical (IHC) analysis utilizes primary antibodies raised against proteins (or certain peptides thereof) coupled with anti-primary antibodies to detect and localize visually the proteins' location in a histological tissue section. Primary antibodies are raised by injecting the protein or peptide of interest into one animal species and purifying them from the serum, while the non-specific secondary antibodies are derived from a different species and will bind to the primary antibodies. Upon binding of the primary antibody to its target protein, the secondary antibody covalently bound to an enzyme (horse radish peroxidase – HRP, for example), is used to metabolize a chromogenic or chemiluminescent substrate for visual protein localization. Alternatively, the secondary antibody can be labeled with a fluorescent tag for protein localization and quantitation. The use of secondary antibodies has financial and practical benefits compared to using only primary antibodies. In addition to reducing the cost of IHC by using lower volumes of expensive primary antibodies and greater volumes of more affordable secondary antibodies, use of labeled secondary antibodies also enables amplification of the reporter signal because multiple secondary antibodies can bind to a single primary antibody.

While MRM validation of iTRAQ candidates provided accurate estimation of targeted protein fold change, further analysis by more traditional IHC staining was expected to provide additional support and reveal insight into the localization of protein changes within the growth plate. Obtaining objective quantitation from IHC images can be challenging, thus the more precise protein fold change reported by MRM was used exclusively for quantitative values, whereas the IHC results served as a more visual and qualitative representation to support the quantitative data and provide biological context.

IHC analysis was undertaken for chondroadherin, PRELP, cartilage oligomeric protein (COMP), fibronectin 1, hyaluronan and proteoglycan link protein, perlecan, matrillin 1, matrillin 3, and thrombospondin 1 proteins in the MPS I and wildtype 5 week femoral head by Dr. Frank Zaucke (133). DAPI staining was used to assess the cellularity of the femoral head sections and confirmed that the MPS I and wildtype growth plate had similar morphology (**Figure 4.5**). In comparing the immunologic stain intensity of each protein between the MPS I and wildtype femoral heads, PRELP was shown to be clearly decreased in the MPS I femoral head growth plate (**Figure 4.5**) (133), which corroborates the decrease in PRELP protein suggested by iTRAQ and MRM. While the other proteins investigated by IHC did not show clear results, there were trends of decreased fibronectin 1 and thrombospondin 1 protein, and increased perlecan protein in the 5 week growth plate that may or may not be meaningful in this limited dataset (**Figure 4.5**). Further sample analysis would be required to ascertain the significance of changes between MPS I and wildtype for fibronectin 1, thrombospondin 1, and perlecan by IHC.



Figure 4.5: Representative immunohistochemical staining images of key ECM proteins in the 5 week MPS I and wildtype femoral head growth plate.

4.5. Modulation of MPS I Chondro-osseous Phenotype by Exercise

Clinical presentation of MPS I in patients often includes arthropathy similar to osteoarthritis (discussed in section **1.4.**), with mild inflammation of the joint capsule tissues, painful joint movement, and degeneration of the articular surface. In order to examine what (if any) role joint movement has on the progression of arthritic disease in MPS I, wildtype and MPS I mice were exposed to daily exercise on a motorized wheel for 5 weeks, after which knee joints were examined using standardized histopathology practices (151). The main objective of this experiment was to test the mechanical integrity of MPS I cartilage.

Previous studies in animal models of the MPSs (MPS I, II, VI, VII, and IX primarily) have supported the concept of osteoarthritic disease through evidence of synovial inflammation, GAG depletion, and erosion of the articular cartilage (18) (88) (94) (95). However, classification and grading of osteoarthritis is complicated by the three dimensional structure and function of a joint, as well as the complex interplay of the various tissue systems involved in joint movement. It has also been well established in the arthritis literature that joint pressure loading is not evenly distributed across a joint (169) (170), thus examination of arthritic joint tissues requires assessment of the entire joint rather than single histological slices, as well as drawing comparisons between similarly loaded joint regions (151). Unlike in the MPS literature, a robust history of osteo- and rheumatoid arthritis has led to standardization of histopathology practices which address these complexities by serially examining the entire joint and compiling measurements (14). This grading approach was used on the MPS I murine model for the first time.

The MPS I mouse model used here has thus far not been shown to develop arthritic-like changes in joint tissues even at advanced ages (>20 weeks of age). While this absence may

reflect model-specific differences, it was hypothesized that degenerative changes may not be occurring due to the largely sedentary lifestyle of a laboratory mouse, with relatively little room and motivation for natural exercise. Although joint tissues, including the articular cartilage, may be mechanically predisposed to damage as a result of changes to their biomechanical integrity resulting from MPS I metabolic changes, these changes alone may not be sufficient to induce degenerative changes.



Figure 4.6: Representative frontal knee sections stained with Safranin O and Fast Green from exercised and control wildtype and MPS I mice. The femur is oriented to the right of each slide. The left knee of the control MPS I mouse was lost during processing. "WT" denotes wildtype and "MU" denotes MPS I mice.

Microscopic examination of serial whole knee sections revealed a trend of higher Safranin O stain (which labels charged molecules such as GAGs) intensity in both MPS I exercised and MPS I control mice compared to their wildtype counterparts (**Figure 4.6**). This was expected in MPS I tissue in light of GAG accumulation. As the MPS I mice have increased Safranin O staining due to the GAG storage, it was necessary to compare changes between wildtype mice that ran to wildtype mice that did not run, and MPS I mice that ran to MPS I mice that did not run, rather than a simple wildtype run to MPS I run. Comparison of the measured average grey values (color density measured in ImageJ) showed decreased GAG staining in the lateral tibia (1.8%) and femur (1.9%) in MPS I mice that were exercised, whereas GAG staining in the lateral tibia and femur was increased (2.8%) in wildtype mice that were exercised (**Table 4.6**). No other knee articular surface quadrants showed statistically significant changes in GAG staining.

Comparison of the articular cartilage area, averaged across all sections of each knee joint per mouse, showed reduced articular cartilage area in the lateral tibia (16%) and increased articular cartilage in the medial tibia (16%) and femur (6%) in wildtype mice that ran (**Table 4.6**). The MPS I mice that ran also showed reduced articular cartilage relative to control animals, in the lateral femur and medial tibia (12% and 19% decreased respectively) (**Table 4.6**). Similar to the GAG stain intensity results, the magnitude of change between mice that exercised compared to those that did not was greater in the MPS I group than the wildtype group. Table 4.6: Summary of calculated GAG stain intensity (grey value) and articular cartilage area values averaged across serial sections of the knee joints in four areas from wildtype and MPS I mice that were exercised for 5 weeks compared to un-exercised controls. Lat (lateral), Med (medial), Tib (tibia), Fem (femur). The T-test p-value is shown for each comparison. Negative numbers indicate decreased GAG stain intensity or articular cartilage area between comparisons.

GAG Stain Intensity								
	WT Run							
	Lat Tib	р	Lat Fem	р	Med Tib	р	Med Fem	р
WT Control	3.641	0.031	3.834	0.004	-0.120	0.469	0.172	0.289
	MU Run							
	Lat Tib	р	Lat Fem	р	Med Tib	р	Med Fem	р
MU Control	-5.539	0.036	-3.124	0.284	4.232	0.052	2.443	0.244
Articular Cartilage Area								
	WT Run							
	Lat Tib	р	Lat Fem	р	Med Tib	р	Med Fem	р
WT Control	-1.341	0.002	-0.303	0.126	0.835	0.015	0.343	0.049
	MU Run							
	Lat Tib	р	Lat Fem	р	Med Tib	р	Med Fem	р
MU Control	0.562	0.095	-0.878	0.004	-1.504	0.003	-0.402	0.056

4.6. Summary

Proteins which were observed to be significantly increased or decreased in the MPS I growth plate were then re-examined through validation in new samples using targeted approaches such as Western blotting, immunohistochemistry, and targeted mass spectrometric analysis. The primary validation of iTRAQ candidates used was a custom MRM panel for 27 proteins of interest on freshly prepared five week mouse micro-dissected growth plate protein to compliment the Western blots and immunohistochemistry results. Western analysis of COMP and PRELP supported the iTRAQ results. MRM analysis identified six proteins which were significantly altered in the MPS I growth plate (**Table 4.5**) and validated the decreased protein levels identified by iTRAQ analysis for the SLRPs biglycan, fibromodulin, and PRELP. The MRM results also identified decreased levels of collagen I and SERPIN F1 that were not shown

by iTRAQ. Immunohistochemical analysis of the five week femoral head growth plate showed PRELP to be decreased in the MPS I femoral head growth plate (**Figure 4.5**), which further supports the decrease in PRELP protein suggested by iTRAQ and MRM. In order to examine the role of joint movement on the progression of arthritic disease in MPS I, and test for biomechanical instability in MPS I cartilage, wildtype and MPS I mice were exposed to daily exercise for 5 weeks. Wildtype mice that were exercised showed reduced articular cartilage area in the lateral tibia, and increased articular cartilage in the medial tibia and femur (**Table 4.6**). MPS I mice that were exercised also showed reduced articular cartilage, in the lateral femur and medial tibia, but the loss of cartilage area was greater than in for the wildtype mice (**Table 4.6**). The more extensive reduction of articular cartilage area in MPS I mice knees suggests that MPS I cartilage is more vulnerable to degeneration upon use in the murine model.

5. RNA Expression Studies

5.1. Introduction

In addition to the protein studies (**Chapters 4 and 5**), analysis of genome-wide mRNA expression followed by targeted qPCR enabled comparison of mRNA and protein level for examination of transcriptional vs translational control of protein deregulation. By studying RNA and protein changes, the strategy used here facilitated greater understanding of cellular disruption in response to the MPS I catabolic defect.

To further examine protein candidates identified by iTRAQ and validated by MRM (**Table 4.4**), investigation of gene expression was undertaken by both unbiased gene expression array and targeted quantitative PCR (qPCR) analyses. While relative quantitative protein analyses were undertaken initially to identify changes occurring in the MPS I growth plate, it was critical to understand the factors responsible for the observed protein changes – transcriptional, translational, or post-translational. mRNA studies were performed to explore the relationship of transcription to these protein changes. Despite the obvious relationship between DNA transcription and protein level, the two are imperfectly correlated. Previous genome-wide RNA and protein expression surveys estimate an average genome-wide correlation of 0.3 - 0.70 between mRNA expression and protein translation, indicating that post-transcriptional and post-translational regulation of proteins often has a large role in regulating protein abundance (152, 153).

Several platform options for genome-wide profiling of mRNA levels were available at the time of these studies, including RNA sequencing, oligo hybridization chips, and qPCR arrays. The most commonly used approach has traditionally been oligonucleotide array chips due to their relatively-lower cost, commercial availability, ease of use, and multiplexing capacity.

However large custom qPCR arrays are now capable of measuring hundreds of genes in a multiplex assay, making broad expression analyses increasingly possible by this approach as well. RNA sequencing (RNAseq), a more recently developed method that leverages the increasing throughput of DNA sequencing machines by directly sequencing cDNA produced from genomic RNA transcripts (154), and is currently the least biased genome-wide expression profiling option. At the time of this writing however, RNAseq was still considerably more expensive and involved a more complicated bioinformatic analysis compared to hybridization arrays. The multiplex capacity and affordability of oligonucleotide chips makes them an excellent primary method for discovery-phase experiments.

5.1.1. Genome Wide Expression Arrays

Oligonucleotide hybridization chips are offered by several companies, including Affymetrix, Agilent, and Illumina, with each offering a similar a similar array product for mRNA expression analysis (the primary difference being the non-"gene" probes included (see below)). As the Illumina MouseWG-6 v2.0 chip was used here, further discussion will be restricted to this platform. Oligonucleotide chips are considered 'genome-wide' but are not truly comprehensive by nature of their hybridization to only predefined locus targets (155). For detection of mRNA expression level, RNA is amplified by reverse-transcribing into cDNA, which is then amplified. The cDNA is subsequently re-transcribed back into RNA in-vitro by priming from the poly-A sequences (156). The amplified RNA (cRNA) is then purified and allowed to hybridize with their complementary sequence oligonucleotide probes arrayed on a chip. Quantitation of cRNA hybridization to beads is measured by Cy3 fluorescent dye bound to streptavidin (which binds to the biotin labels that were incorporated into the cRNA during *in vitro* transcription (156)). Thus, in order to detect an mRNA, the locus it is transcribed from must

be known *a priori* so that it is represented on the array. The majority of probes in the array are designed from the ~30,000 annotated and confirmed loci in the RefSeq dataset, while the remainder are derived from predicted, un-annotated, or non-coding RNAs. The MouseWG-6 v2.0 chip probes, for example, are derived from 26,768 coding loci, and 18,515 predicted from other sources (including the RIKEN FANTOM 2, UniGene, and MEEBO databases) for a total of 45,283 loci covered (156). While these loci cover a majority of the genome, there will be transcripts missing and thus unreadable by the assay.

5.1.2. RNA Integrity

Isolation of high-quality RNA is paramount to successful RNA expression studies; however, obtaining sufficient RNA from chondro-osseous tissues like the mouse femoral head growth plate was challenging, given the difficulty of tissue disruption and lysis, as well as the low cell number. In order to sufficiently disrupt the micro-dissected growth plates, a combination of TRIzol reagent and mechanical grinding was empirically found to release the most RNA from the cartilage. While TRIzol is generally accepted to isolate pure RNA free from DNA and protein contamination (157), further purification and DNase digestion was also used. The integrity of extracted RNA is potentially the most important factor for the success of genomewide expression analyses, as degradation leads to biased and erroneous hybridization and quantitation results. The standard measure of RNA integrity is the <u>RNA Int</u>egrity score (RIN score), reported by analysis on a BioAnalyzer (Agilent) (158). A small aliquot of the RNA samples was used for microfluidic gel electrophoresis, and the migration of ubiquitous 18S and 28S ribosomal RNA (rRNA) was used to assess RNA degradation (158). If there is RNA degradation, the 18S and 28S rRNA bands show lower intensity and more diffuse bands. Thus,

rRNA measurement is used as a proxy for the integrity of overall genomic RNA molecules in the samples. Typically a RIN score or \geq 7.0 is required for a successful RNA expression array study.

It was found that RNA degradation occurred during micro-dissection of the growth plate over the ~2min period that sections were thawed, resulting in RIN scores <7.0. This issue was corrected by thawing the femoral head sections under a drop of RNAlater solution (Ambion) and micro-dissecting the growth plate within the drop to protect the RNA from degradation. Doing so resulted in excellent quality RNA (RIN 7.5-8.0). It does not seem that detectable RNA degradation occured during femoral head collection or adherence of cryosections to slides.

5.1.3. qPCR Validation

Validation of genome-wide expression array candidates is most commonly achieved through targeted analyses such as quantitative real-time PCR (qPCR) of specific mRNA levels with high precision. qPCR is a well-established method (reviewed in (159)), providing quantification of transcript levels by first reverse transcribing poly-A mRNA into cDNA with reverse transcriptase, and then using primers for the target exonic sequence of interest to PCR amplify the cDNA. Detection and quantification of the amplification can be carried out using fluorogenic hydrolysis probes (TaqMan system, Applied Biosystems) or fluorogenic SYBR Green intercalation (RT² Profiler Assay system, Qiagen) to name just two of the more frequently cited options (160). The primary advantage of a hydrolysis probe system is the capacity for multiplex reactions, allowing each single reaction solution to assay multiple targets (159), whereas the non-specific SYBR Green system can only quantify one target per reaction (159). The SYBR Green system is, however, generally lower-cost and equally sensitive.

Here the RT² Profiler Array SYBR Green system from Qiagen was used. Each of the primer pairs designed by Qiagen was tested to ensure similar PCR efficiency (enabling multiplex analysis of several gene targets at once) and was shown to have minimal primer-primer dimers (a common problem with qPCR primer design) by melt curve analysis. The reaction solution mixes were all optimized for the ABI7500 (ABI) thermal cycler to reduce experimental confounding factors.

The custom array included two control housekeeping genes for normalization, as well as a ribosomal rRNA control, PCR check control, and a test for residual genomic DNA (161). Selection of normalizing genes is a significant problem for qPCR analysis, as no one control will work well for all tissues nor be unaffected by all experimental conditions (159). Frequently glyceraldehyde dehydrogenase (GAPDH) and beta-actin (ACTNB) are used, given their ubiquitous and high expression, and these two were used here in combination in an attempt for accurate normalization. Relative quantification was carried out using the $\Delta\Delta$ Ct method (detailed in (159)) after normalization to compare the MPS I and wildtype samples.

5.2. Illumina Array Data Analysis

The Illumina chip data were interpreted by two complementary approaches: using Illumina's GenomeStudio software and a custom algorithm designed in the R (162) statistical environment. Illumina's software package is commonly used as it has been designed specifically for Illumina chip products; however, independent analysis within R allows much greater user manipulation and customization of analysis. GenomeStudio software focuses on strictly genelevel analyses for identifying candidates. Because there are multiple probes for each gene on the chip, this approach requires averaging the quantification signal recorded for every probe

assigned to a gene before calculating the overall expression value. Averaging probes, however, may lead to imprecision if the values for certain probes for a gene are outliers or do not perform as expected during the analysis, as they cannot be removed manually in GenomeStudio (even if there is evidence suggesting it would be appropriate to remove them).

GenomeStudio analysis utilizes an "average" normalization factor whereby the intensity values for all probes are scaled according to the ratio of Average_{real values}:Average_{virtual values} (163). The virtual probe values are a simulated dataset based on the distribution of probe values observed in the real samples that is intended to correct for non-biological variables that may affect probe performance (163). The Benjamini-Hochberg multiple hypothesis testing correction used in GenomeStudio allows for more false positives than false negatives to avoid being overly conservative, as this was a discovery phase experiment similar to the intent behind iTRAQ analysis. The DiffScore is a GenomeStudio value intended to show the magnitude and direction of fold-change between samples for expression of a locus. It was calculated using the Illumina Custom method, which assumes that probe intensity will be normally distributed between samples in the control and experimental groups (163). GenomeStudio then transforms the ratio of average intensity for each gene in sample group 1 vs. sample group 2 according to the equation (163):

$DiffScore = [10Sgn(Intensity_{Experimental} - Intensity_{Control})Log_{10}p]$

(Sgn is a function returning the sign of an integer as either 1, 0, or -1, while intensity is the quantitation value reported from a probe on the array). According to this equation, the DiffScore indicates fold change of the gene between MPS I and wildtype samples as well as the p-value associated with that difference where:

DiffScore $\pm 13 = p < 0.05$ $\pm 20 = p < 0.01$ $\pm 30 = p < 0.001$

To complement the GenomeStudio dataset, the R analysis focused on individual probes rather than genes. R analysis first filtered out any probes with a non-significant p-value of detection (>0.05) and then normalized all significant probe intensities. The full R-code used can be found in Appendix 1. Normalization was done by dividing each probe intensity in a sample by the average intensity of that probe in all samples of the Illumina experiment, and was done to convert all intensity values to numbers closer to 1.0, so as to avoid detection of only large foldchange candidates. Significant changes in probe intensity between wildtype and MPS I samples were identified by fitting the normalized probe intensities to a linear curve and calculating the slope. A positive slope indicated an increase in the MPS I samples while a negative slope indicated a decrease. The p-value of the difference was calculated from the tightness of fit to the trend line using an Empirical Bayes approach. Probe p-values were adjusted for multiple hypothesis testing using the limma package (164). Probes were classified as significant candidates if they showed a >1.5 fold change between MPS I and wildtype (to stay within the statistical power of the experiment as it was designed) and had an adjusted p-value <0.05.

5.4. Illumina Genome Array Candidates

One MPS I and one wildtype pool were run in duplicate on the same chip to serve as technical replicates, which confirmed that technical variability was low (**Table 5.1**). Similarly, comparison of biological replicates on the chip indicated that inter-sample variation was also low (**Table 5.1**).

Table 5.1: Average ratio of signal intensity for 15,658 genes between three biological and one technical replicates. "M" denotes MPS I samples while "WT" denotes wildtype samples.

Biological Replicates		Technical Replicates		AverageM/AverageWT
M1/Average(M2,M2)	WT1/Average(WT2,WT2)	M2/M2	WT1/WT1	
1.10	0.98	1.01	0.98	1.08

The Illumina MouseRef 6 v.2 bead array assayed probes for 45,281 genomic loci and included redundant probes and beads for technical error correction. Analysis of the data using GenomeStudio identified 15,805 transcripts in the wildtype and 15,560 transcripts in the MPS I five week growth plate samples (133). Most of the identified transcripts did not show significant differences in expression between MPS I and wildtype samples (<1.5 fold change)(133). Of the identified transcripts, 1077 were detected only in wildtype and 1064 only in MPS I samples (133).

GenomeStudio analysis of the Illumina data identified 196 loci as significantly increased and 52 genes as significantly decreased (>1.5-fold and p<0.05) in the 5 week MPS I femoral head growth plate (14) (p<0.05 after Bonferroni multiple hypothesis correction) (**Figures 5.1 and 5.2**).



Figure 5.1: Venn diagram of Illumina expression array GenomeStudio and custom R analysis candidate intersection.

As discussed above, GenomeStudio analyzes candidates at the gene level, while the R analysis I used analyzes candidates at the probe level. This complimentary R analysis identified 239 probes as significantly increased and 79 probes as significantly decreased in the 5 week MPS I growth plate (>1.5-fold change, corrected p <0.05) (**Figures 5.1 and 5.3**). 97 genes were identified as significant candidates by both GenomeStudio and R analysis (20 decreased genes and 77 increased genes) (**Figure 5.1**).

SYMBOL	MPS I AVG Signal	Detection Pval	DiffScore
lghg	121.8	0.00126	171.278
Alox12	179.9	0.01506	114.185
Epb4.9	189.8	0.03493	49.046
Mki67	576.8	0.0245	39.852
Stx11	176.2	0.00347	33.042
LOC100046690	112.6	0.0439	32.206
IGKV8-31_AJ235957_Ig_kappa_variable_8-31_3	29.6	0.00107	31.435
LOC385291	92	< 0.001	29.665
LOC384413	27.7	0.00214	17.21
LOC672339	40.4	0.00107	16.684
LOC672342	108.6	<0.001	15.533

Table 5.2: List of GenomeStudio® candidate loci with significant DiffScores detected in only MPS I samples.


Figure 5.2: Scatter plot of average probe intensity per gene in MPS I (-/-) and wildtype (+/+) samples for all genes identified using GenomeStudio®. Those which were identified as being differentially expressed (DiffScore >13) are indicated in red.



Figure 5.3: Scatter plot of log probe intensity per gene in MPS I (-/-) and wildtype (+/+) samples for all genes identified using custom R analysis. Those which were identified as being differentially expressed (adjusted p-value <0.05) are indicated in red.

5.5. qPCR Validation

A 96 well plate array assaying 48 genes by qPCR was developed to simultaneously validate select candidates from the Illumina whole genome array dataset, to investigate genes of interest from the proteomics datasets (\Diamond in **Table 3.2**) and certain key genes for chondro-osseous development (Table 1.3) (133). Eleven genes for qPCR analysis were selected for validation from the Illumina array datasets (133). None of these eleven genes showed significant changes in mRNA concentration by qPCR, whether the genes had been selected by analysis through GenomeStudio or the custom R analysis. (133). However, as noted in the original publication of these data, "The qPCR results identified 14 transcripts as being significantly altered in the MPS I samples (8 increased and 4 decreased) \geq 1.3-fold (p < 0.05 after multiple hypothesis correction) (Table 5.3)." (133). The statistical power of the qPCR analysis allowed the fold cut off to be lower (1.35 fold) than for the Illumina array analysis. Stronger weighting of significance was given to the qPCR results than the Illumina array results as "the qPCR experiment was more sensitive and appropriately powered to detect small changes in mRNA." (133). As stated above, the Illumina array was intended as a discovery phase experiment to generate candidates and hypotheses for further study. As such, final mRNA candidates were chosen from significant qPCR results (Table 5.3). "It was not specifically required that the iTRAQ or MRM results show the same trend as the RNA results, as *a priori* the two need not be perfectly correlated. As some extracellular matrix proteins have very long half-lives, measurable changes in mRNA level may precede changes in protein level by a considerable amount of time." (133)

Table 5.3: Summary of qPCR results for the 5 week MPS I compared to wildtype growth plate. Significant p-values are reported (after adjusting for multiple hypothesis testing) for the ratio (fold change) of RNA abundance in MPS I:wildtype. Genes indicated with a * showed consistent changes in the Illumina expression array dataset. Negative fold change indicated a decrease in the MPS I Samples.

Gene	2^-(ΔΔCt)(fold change)	p-value
Adamts4	1.74	0.045
Aspn	2.74	0.007
Chad	-2.35	0.0004
Col2a1	-1.30	0.047
Col9a1	-1.45	0.003
Hapln4*	-1.94	0.024
Lum	1.45	0.0004
Matn1	1.39	0.019
Mmp3*	3.19	0.001
Ogn	1.69	0.0004
Omd	1.45	0.008
P4ha2	1.28	0.006
Prelp	-1.31	0.012
Rab32*	1.26	0.023

 Table 5.4: Full qPCR array results. Significant (after adjusting for multiple hypothesis testing) p-values are reported.

 Negative fold change indicated a decrease in the MPS I Samples.

	Average MU	Average WT		2^-(ΔΔCt)	
Gene	Value	Value	MU/WT	(fold change)	p-value
Adam12	4.27	4.18	1.02	-1.06	
Adamts12	6.15	6.32	-1.03	1.12	
Adamts4	7.95	8.75	-1.10	1.74	0.04548
Alpl	1.25	1.39	-1.11	1.10	
Aspn	4.79	6.24	-1.30	2.74	0.00772
Bgn	0.23	0.50	-2.17	1.21	
Chad	1.37	0.13	10.19	-2.35	0.00039
Col10a1	0.82	0.38	2.15	-1.35	0.37373
Col11a1	-0.39	-0.38	1.03	1.01	
col11a2	1.18	1.08	1.09	-1.07	
Col12a1	5.09	5.19	-1.02	1.07	
Col1a1	-4.84	-4.96	-1.02	-1.08	
Col1a2	-5.54	-5.69	-1.03	-1.11	
Col2a1	-2.66	-3.04	-1.14	-1.30	0.04713
Col9a1	2.30	1.76	1.31	-1.45	0.00308
Col9a2	4.98	4.93	1.01	-1.03	
Comp	1.13	1.07	1.05	-1.04	
Dcn	-0.37	-0.32	1.16	1.04	

	Average MU	Average WT		2^-(ΔΔCt)	
Gene	Value	Value	MU/WT	(fold change)	p-value
Ehd1	3.84	3.75	1.02	-1.06	
Ерус	5.74	5.30	1.08	-1.35	0.13757
Fmod	3.10	3.32	-1.07	1.16	
Hapln4	7.83	6.88	1.14	-1.94	0.02402
Ihh	6.22	5.90	1.05	-1.25	
Lum	1.06	1.60	-1.51	1.45	0.00038
Matn1	4.76	5.24	-1.10	1.39	0.01922
Matn3	2.26	2.04	1.11	-1.17	
Mmp13	-0.88	-0.77	1.14	1.08	
Mmp3	5.28	6.96	-1.32	3.19	0.00098
Nfkb1	4.29	4.23	-0.98	0.95	
Ogn	4.06	4.82	-1.19	1.69	0.00037
Omd	3.87	4.41	-1.14	1.45	0.00763
P4ha1	3.29	3.52	-1.07	1.17	
P4ha2	4.36	4.72	-1.08	1.28	0.00577
Prelp	3.96	3.58	1.11	-1.31	0.01179
Rab32	5.94	6.28	-1.06	1.26	0.02313
Smad3	5.01	4.89	1.03	-1.09	
Smad5	4.34	4.62	-1.06	1.21	
Sox9	4.91	4.68	1.05	-1.17	
Sparc	-3.38	-3.57	-1.05	-1.14	
Timp1	3.92	3.88	1.01	-1.03	
Tlr4	5.67	5.78	-1.02	1.08	
Tnfrsf11	4.47	4.51	0.99	1.02	
Tnfrsf11a	8.69	8.87	-1.02	1.13	

5.6. Analysis of RNA Results

As the Illumina expression array results could not be validated, only genes which showed consistent results (trends of increase or decrease) between the qPCR data and at least one of the Illumina datasets (derived by GenomeStudio or R) were considered strong RNA candidates (**Table 5.3**, full results in **Table 5.4**). As the qPCR data set was the most sensitive and appropriately powered RNA experiment used here, candidates had to first be significant within

the qPCR dataset. In all cases where iTRAQ data were available for these candidate genes, the trend of change in MPS I samples was conserved across RNA and protein levels (**Table 5.5**).

			Chip Fold	Chip	Chip Fold	Chip R Adj.	iTRAQ Fold
Gene	2^-(∆∆Ct)	p-value	Change	DiffScore	Change	p-value	Change
Chad	-2.35	0.00039	-1.41	-9.67	-	-	-5.65
Hapln4	-1.94	0.024	-1.96	-54.35	-1.95	0.01	-
Col9a1	-1.45	0.0031	-1.88	-6.53	-	-	-2.12
Prelp	-1.31	0.012	-1.26	-7.84	-	-	-2.54
Col2a1	-1.30	0.047	-1.42	-29.41	-	-	-1.79
Rab32	1.26	0.023	35.11	1.32	1.32	0.02	-
Mmp3	3.19	0.00098	17.89	1.40	1.38	0.03	-

 Table 5.5: RNA expression analysis candidates consistent between analyses. Negative fold change indicated a decrease in the MPS I Samples.

5.7. Summary

To to explore the relationship of transcription to the protein candidates identified by iTRAQ and validated by MRM (**Table 4.4**), investigation of gene expression was undertaken by both unbiased gene expression array and targeted quantitative PCR (qPCR) analyses. The Illumina MouseWG-6 v2.0 chip was used to measure transcriptome-wide expression of mRNA, and identified 15,805 transcripts in the wildtype and 15,560 transcripts in the MPS I five week growth plate samples. Two complimentary data analysis options were used to interrogate the Illumina chip data – GenomeStudio and a custom R script. GenomeStudio analysis of the Illumina data identified 196 loci as significantly increased and 52 genes as significantly decreased in the 5 week MPS I femoral head growth plate (**Figures 5.1 and 5.2**). R analysis identified 239 probes as significantly increased and 79 probes as significantly decreased in the 5 week MPS I growth plate (**Figures 5.1 and 5.3**). A 96 well plate array assaying 48 genes by qPCR was developed to simultaneously validate eleven candidates from the Illumina chip dataset, and to investigate genes of interest from the proteomics datasets (\diamond in **Table 3.2**) and key genes for chondro-osseous development (**Table 1.3**). None of the eleven candidates were validated by the qPCR panel, however it identified 14 transcripts as being significantly altered in the MPS I samples (8 increased and 4 decreased) (**Table 5.3**). Genes which showed consistent trends of increase or decrease between the qPCR data and at least one of the Illumina datasets (derived by GenomeStudio or R) were considered RNA candidates. By these criteria, seven mRNAs were identified as being significantly altered in the MPS I murine model growth plate (**Table 5.5**).

6. Discussion and Conclusions

6.1. Discussion

6.1.1. Biomechanical Failure Underlying Chondro-osseous Disease in MPS I

This dissertation is an unprecedented description of early pathogenic mechanisms for chondro-osseous disease in MPS I discovered through unbiased proteomic and transcriptomic techniques. These approaches identified six proteins (Table 4.5) and fourteen mRNAs (Table 5.3) that were significantly deregulated in the MPS I murine model femoral head growth plate cartilage. The proteins and mRNAs identified are key constituents or regulators of the extracellular matrix, each vital for the normal structural integrity of connective tissue. The early alterations of extracellular matrix proteins in MPS I suggests that biomechanical changes to chondro-osseous tissue function likely represent very early events in the pathogenesis of disease. These observations evoke a "biomechanical failure model" for skeletal and connective tissue disease in the MPSs. (Figures 6.1 and 6.2). The model predicts that early pathogenic changes set the stage for later disease by establishment of tissues that are prone to biomechanical failure. Under mechanical loading, these compromised tissues activate multiple pathways of degeneration (e.g., inflammation), which subsequently lead to further damage. An increasingly complex network of early and late pathogenic cascades culminates in profound damage to compromised tissues and irreversible disability (Figure 6.2). Thus, this biomechanical failure model can accommodate later disease pathways that have been previously identified in other studies. This model implies that complex pathways, secondary to the primary catabolic defect in the MPSs, will be contributors to disease pathogenesis and the recalcitrance of chondro-osseous disease to therapy.



Figure 6.1: Hypothetical chronology and relationships between chondro-osseous metabolic activities and pathogenic changes in the MPSs. Speculative disease mechanisms related to broad clinical phenotypic features are indicated in the red circles.



Figure 6.2: The biomechanical failure model describes 3 contiguous phases of chondro-osseous pathogenic progression (Setup, Establishment, and Culmination) potentially beginning in early prenatal development and progressing through to maturity. The primary defect of GAG catabolism creates a cellular environment that enables establishment of a biomechanically-compromised tissue, which together culminate in degenerative changes to these tissues with increasing disability. The complexity of cellular pathogenic changes will initially be low, beginning with the primary GAG catabolic defect, but increase rapidly with time culminating in a very complex network (illustrated by the hypothetical networks shown) of pathogenic cascades that will be difficult to impact with therapeutics.

Importantly, the biomechanical failure model implies that therapeutic intervention

directed towards the primary metabolic block and applied during later phases of disease may

stabilize disease but would be unlikely to reverse damage and re-establish extracellular matrix

function. As later stage disease is characterized by activation of multiple pathogenic pathways, multiple therapeutics would be required to effectively address late chondro-osseous disease. According to this model, earlier stages of disease are less complex and potentially more efficient therapeutic targets.

6.1.2. Biomechanical Failure Model: Linking the pathogenesis of MPS skeletal disease to forms of Arthritis and other Genetic Skeletal Dysplasias

The hypothesis that alterations of the extracellular matrix occurring early in the natural history of disease lead to biomechanical failure of chondro-osseous tissue in the MPSs is strengthened by observations in other kinds of arthritis and genetic skeletal dysplasias. Extensive studies have demonstrated that these diseases have similar mechanisms of pathogenesis involving early biomechanical failure. OA is a late-onset multifactorial disease characterized by stages of articular cartilage degeneration, pain and inflammation of the joints, and structural changes in the subchondral bone. A clear genetic and epigenetic predisposition to development of OA is emerging, with the identification of several risk loci and genes - hypoxiainducible factor- 2α (HIF- 2α), matrix metalloproteinase 13 (MMP13), discoidin domain receptor (DDR-2), syndecan 4, TGF^β, the alarmins S100A8 and A9,miRNA9, miRNA98, hypomethylated receptor activator of nuclear factor κB ligand (RANKL), osteoprotegrin (OPG), and the rs143383 allele of the growth and differentiation factor 5 (GDF5) locus (23, 165, 166) as examples. Manifestations of OA typically involve the interplay of numerous genetic and environmental factors (injury, for example, is often a major environmental contributor (167)), which over time lead to articular cartilage structural changes and degeneration (168). Genetic and environmental factors are interconnected such that mechanical loading impacts the metabolic state and proteinaceous content of cartilage and bone through induction of A Disintegrin And Metalloproteinase with Thrombospondin Motifs (ADAMTS), and matrix metalloproteinase

(MMP) expression leading to structural extracellular matrix changes related to aggrecan and collagen proteolysis (169, 170).

There are currently more than 180 skeletal dysplasias described (of which the MPSs are a major subset (171)) with a similar number of identified underlying genes. Many skeletal dysplasias are monogenic (172), indicating that the developing growth plate is very sensitive to changes in even a single protein component of the diverse network of structural and signalling components. For example, mutations in cartilage oligomeric protein or matrillin 3 can lead to both pseudoachondroplasia and multiple epiphyseal dysplasia, which present with skeletal dysplasia, tendon laxity, and early-onset OA (173). The diverse phenotypic effects of cartilage oligomeric protein or matrillin 3 are one example of how extracellular matrix protein changes (such as the ones described in this dissertation) can lead to broad chondro-osseous disease. This example also demonstrates mechanical and biochemical interdependence between bone and cartilage; where changes in bone structure can destabilize the joint and erode articular cartilage over time. The phenotype of many skeletal dysplasias also includes arthropathy as a result of changes in subchondral bone structure, which shifts loading on the articular surface (167) (172). Dysostosis multiplex and joint malalignment will similarly be contributing factors of arthropathy in the MPSs by altering gait and shifting zones of pressure loading within joints.

6.1.3. Proteomic and Transcriptomic Analysis of MPS I Murine Model Chondroosseous Disease

The following section contains information published in a research journal (133).

I have shown reduction of several SLRP family members, including biglycan, fibromodulin, and proline/arginine-rich end leucine-rich repeat protein (PRELP), in the fiveweek MPS I murine model growth plate. I have also shown reduction of collagen I and SERPIN F1. Based on the known biological roles of these proteins (discussed below), their involvement in MPS I supports the biomechanical failure model.

The involvement of the SLRPs in MPS I pathogenesis suggested by this dissertation is novel and intriguing. Biglycan, fibromodulin, and PRELP have been demonstrated as important for cartilage development and function. These proteins are classified as 'matricellular proteins,' describing their functional role as regulating both structural integrity and signalling activity of the extracellular matrix during development and repair (174). Matricellular proteins have been shown to regulate extracellular matrix organization, adhesion, cellular proliferation, and growth factor sequestration (primarily bone morphogenic protein (BMP) and transforming growth factor beta (TGF β) ligands). Biglycan and fibromodulin have been shown to bind to and modulate signalling of BMP2, BMP4, TLR4, TGF β , TGF β 2, TGF β 3, and wingless integration (Wnt) ligands (143, 144, 145, 146, 147). Binding of these ligands occurs through both the GAG chains of biglycan and fibromodulin (175), as well as by interaction with the core proteins (147), to enable sequestration of the ligands in the extracellular matrix (176).

The relevance of the SLRPs to pathogenesis of chondro-osseous disease in MPS I is strengthened by studies of genetic knockout models demonstrating that lack of biglycan and fribromodulin leads to age-dependant cartilage mechanical failure and osteoarthritis (177) (178) (179). Furthermore, as biglycan is able to modulate the signalling activity of BMPs controlling osteoclast development, this protein is clearly important for bone mineralization (143). It has also been shown to induce osteogenic factors through toll-like receptor 2 (TLR2) in aortic valve interstitial cells, which is thought to explain cardiac valvular calcification leading to aortic valve stenosis (180). Studies of biglycan and fibromodulin knockout mouse models have also shown disruption of collagen I and II fibrils in tendon and bone, resulting in joint destabilization and

osteoarthritis (181, 182), indicating a role for the SLRPs in collagen regulation. PRELP is another SLRP that has been shown to interact directly with collagen I and II fibrils, although the importance of this binding is not yet clear (183). It has been suggested that PRELP binding to collagen I and II promotes structural integrity through its co-binding of perlecan (183). PRELP has also been shown serve a chondroprotective role through inhibition of innate immunemediated cartilage degradation, where PRELP binds to and inhibits complement factor 9 and NFkB (184). The NFkB inhibitory activity of PRELP also inhibits osteoclast resorption of bone (185).

Decreased collagen type I and SERPIN F1 protein in the MPS I growth plate likely has structural implications. Collagen type I is the primary structural scaffold of mature bone (>90% of the organic mass). The importance of collagen I to structural integrity is shown through genetic deficiency of collagen I leading to skeletal dysplasias such as Caffey disease, Ehlers-Danlos syndrome, and osteogenesis imperfecta (OI) type VI (35, 36, 37). Mutation of the SERPINF1 locus in humans has been shown to cause OI type VI as well, though this mechanism is not directly related to defective collagen processing (186) as might have been expected. The effects of collagen I and SERPIN F1 deficiency in MPS I shown in this dissertation may be compounded by the effects of deficient biglycan, fibromodulin, and PRELP (given their roles in the collagen network as well).

The structural integrity of the extracellular matrix may also be compromised in MPS I through increased ADAMTS4, MMP3, and decreased collagen IX expression shown by qPCR. ADAMTS4 promotes articular cartilage degradation though cleavage of the aggrecan protein leading to structural fragility and osteoarthritis (187). MMP3 is also known to cleave aggrecan, as well as decorin, and perlecan (188, 189). MMP3 has also been proposed as a prognostic serum

biomarker of arthritis in mouse models and humans (190). Collagen IX has been identified as a genetic risk factor of osteoarthritis based on genetic knockout of collagen IX, leading to lateonset articular cartilage degeneration (191), and based on human linkage studies identifying alleles of the COL9A1 locus as predisposition factors to osteoarthritis (192, 193).

The proteomic and transcriptomic studies described by this dissertation suggest that cellular changes occur in the MPS I murine model femoral head growth plate as early as five weeks of age. Proteomic studies of the three week growth plate cartilage did not reveal a significant disruption of protein homeostasis, unlike at the five week time point (section **3.5.**). This does not indicate that changes are not happening prior to five weeks of age, but rather that such changes may be more subtle and challenging to detect with the experimental approaches used here. While subtle changes in mRNA expression can generally be more readily detected than protein changes, it is not yet clear whether the most proximate pathogenic mechanisms begin at the transcriptional, translational, or protein level based solely on these studies. The changes discussed above represent the earliest observed to occur in response to the MPS I catabolic defect so far.

In the context of previous studies of MPS chondro-osseous disease, inflammation has been well defined and clearly has a role in late disease degeneration (94, 195, 196). The studies in this dissertation showed no evidence of inflammatory pathway activation early in disease natural history. However, the biomechanical failure model describes how early extracellular matrix changes, such as those demonstrated by this dissertation, can lead to inflammation and other late pathological changes. Late-stage mechanisms occur in response to damaged chondroosseous tissues, which were predisposed to damage beginning from an early age as a result of key extracellular matrix functional protein deregulation, such as those described in **Table 3.2**.

6.1.4. Relationship between the Proteomic and Transcriptomic Data

Quantitative protein analyses were undertaken initially to identify changes occurring in the MPS I growth plate. In order to understand the factors responsible for the resulting protein changes, mRNA studies were used. Despite the obvious relationship between DNA transcription and protein level, the two are imperfectly correlated. Genome wide RNA and protein expression surveys have estimated an average genome-wide correlation of 0.3 - 0.70 between mRNA expression and protein translation, indicating that post-transcriptional regulation of proteins often has a large role in regulating protein abundance (151, 152). In this dissertation, mRNA candidates discovered by transcriptomic analysis (validated by qPCR) were compared and contrasted to protein candidates identified by proteomic analysis (validated by MRM). Five proteins identified as being significantly decreased in the 5 week MPS I growth plate were also identified as having significantly lower mRNA expression (Table 5.3). None of the other proteins in **Table 3.2** showed similar correlation of significantly changing protein and mRNA levels. It is not clear why chondroadherin, collagen IX, collagen II, and PRELP would show coordinate down-regulation of mRNA and protein level, while other candidates did not show this. The effect which small changes in mRNA levels will have on the corresponding protein level is difficult to predict. What effect those protein changes may have on cellular physiology is also challenging to define. For example, it is difficult to predict whether a 1.5-fold or lower change in mRNA (as reported in **Tables 4.5 and 5.5**) would have a biologically relevant impact on proteins in vivo. In cases of lowly-expressed genes, smaller fold changes are expected to have a larger impact. In the data analysis of this dissertation, a fold-change cut-off of ± 1.3 for significant mRNA candidates was used to remain within the statistical power of detection of the qPCR array experiment. Combined with the required additional support of the genome-wide

expression array data, the candidates in **Table 5.5** are high confidence, though further studies will be required to elucidate their biological implications. In the case of chondroadherin, hyaluronan and proteoglycan link protein 4, collagen IX, PRELP, and collagen II genes, the protein data indicated that changes as low as a 1.3-fold decrease in mRNA did lead to a significant and measurable decrease in protein level (**Table 5.5**). With the exception of collagen type II, which is among the most highly expressed genes in cartilage, these proteins are expressed at a relatively low level, which may contribute to these low fold-changes leading to measurable changes in protein.

Although ADAMTS4 and MMP3 mRNAs were shown to be significantly increased in the MPS I growth plate (discussed in section **6.1.2.**), their common cleavage substrate, aggrecan protein, was not shown to be decreased in the growth plate by proteomic techniques. It is not yet clear if a modest increase in ADAMTS4 and MMP3 mRNA would immediately lead to aggrecan depletion. It is also not known if MMP3/ADAMTS4 mediated aggrecan depletion would be a mechanism more active in the articular cartilage than in the growth plate cartilage. I propose (in line with the biomechanical failure model) "that gradual weakening of extracellular matrix biomechanical integrity through diminishing aggrecan and extracellular matrix proteins contributes to cartilage biomechanical instability over time." (133).

6.2. Limitations and Future Studies

The objectives of this dissertation were to 1) identify candidate protein changes in the MPS I mouse model growth plate which may drive or contribute to chondro-osseous disease, 2) validate candidate protein changes and characterize functional implications, and 3) understand the mechanisms relating observed changes to the MPS I chondro-osseous phenotype. The studies described within this dissertation have resulted in the identification of six proteins and 14

mRNAs that were significantly deregulated in the five week murine MPS I femoral head growth plate. These results led to development of the biomechanical failure model as an explanation of chondro-osseous disease in MPS I (section **6.1.2.**). The outcome of this dissertation would suggest that further studies examining joint stability and biomechanical properties should be performed.

The techniques used to discover candidates (i.e., iTRAQ proteomic and Illumina transcriptomic analyses) have inherent limitations. For example, as discussed in section **2.1.2.** not all proteins can be identified by a single experiment, nor can all identified proteins be accurately quantitated. This also applies to identification and relative quantitation of mRNA candidates. These broad approaches were used as discovery phase experiments, and successfully suggested candidates for further study. Proteins and mRNAs not identified in this dissertation may also be relevant candidates and factors in early pathogenesis of MPS I chondro-osseous disease. Replication of these studies may identify new candidates, particularly if different experimental approaches are utilized. This may also be facilitated through the MRM assay to study these proteins of interest in other models, ages, and tissues.

The primary tissue studied in this dissertation was the femoral head growth plate. This tissue was selected as the most feasible source of pure cartilage tissue from the mouse, and also represents other hyaline cartilages in the body (including articular cartilage, tracheal rings, intervertebral discs, and knee menisci). The developmental origins of these tissues are similar, though their ultimate roles differ. The chondrocytes in the growth plate are rapidly proliferative to enable bone growth, while articular cartilage chondrocytes proliferate very little, to name one example of these differences. I propose that candidates and putative mechanisms discovered in the growth plate will be relevant in describing pathology occurring the articular cartilage

(relating to the osteoarthritic phenotype in MPS I patients). Specific studies of these candidates in other chondro-osseous tissues (such as articular cartilage) will be required to assess this possibility. I furthermore propose that these candidates and mechanisms are applicable to chondro-osseous disease in the other MPSs, which will similarly require specific future studies in animal models and patients of the other MPSs.

Functional examination of *in vivo* chondro-osseous tissues is needed to test the biomechanical failure model suggested by this dissertation. Functional testing (especially of the articular cartilage) is required to extrapolate results obtained in the growth plate cartilage to other chondro-osseous tissues. To address the need for functional testing, a small cohort of mice was made to run daily for five weeks. Analysis of whole knee joints from these mice showed that MPS I mice which ran lost 19% more articular cartilage volume and had 1.47-fold more GAG staining compared to wild type mice (section **4.5.**). These results suggest that the articular cartilage in MPS I mice is biomechanically compromised and susceptible to degenerative changes. This preliminary functional testing of the biomechanical failure model suggested that the MPS I murine joint is predisposed to arthritic damage upon use. Longer studies should be undertaken to ascertain the predisposition of MPS I articular cartilage to damage. Chronological examination of the pathologic changes occurring in the MPS I joint would also help to strengthen the model; especially since this type of chronological examination has never been done.

While the direct link between deregulation of biglycan, type I collagen, fibromodulin, lactotransferrin, proline/arginine-rich end leucine-rich repeat protein (PRELP), and SERPINF1 to defective GAG catabolism remains unclear, studies of how these proteins respond to therapeutic intervention (ERT and/or HSCT) in the murine model would be beneficial to support the role of these candidates in disease pathogenesis. Such therapeutic studies could also begin to

suggest the relevance of these candidates as biomarkers of chondro-osseous disease in MPS I. Those studies may also provide a more direct rationale for early initiation of therapeutic intervention, as the very early changes described in this dissertation suggest logical rationale for earlier intervention.

The early disruption of extracellular matrix proteins may be indicative of potential biomarkers reflective of extracellular matrix maintenance. Only one protein biomarker of MPS I has been successfully translated from an animal model to patients in the clinic (the heparin cofactor-II-thrombin complex). Biglycan, type I collagen, fibromodulin, lactotransferrin, PRELP, and SERPINF1 may represent further potential biomarkers as discussed in **4.2.2.**, although human serum studies of these proteins in MPS I have not been done. Examination of these proteins in the human patient population may support their validity as biomarkers, and clarify the potential differences between chondro-osseous disease mechanisms in the murine model and human patients.

6.3. Conclusions

This dissertation has shown that early events in the pathogenesis of bone and joint disease in MPS I involve alterations in extracellular matrix components, which I suggest leads to altered cartilage biomechanical properties. The early pathogenic changes revealed by the experiments in this dissertation and the biomechanical failure model that is evoked lead the conceptualization of MPS I pathogenesis away from a storage-centric model. Rather than GAG accumulation directly driving disease, modest changes in key extracellular matrix components occurring very early in disease natural history may set the stage for biomechanical failure and degeneration later in life. The clinical features of the MPSs have traditionally been explained solely by considering the direct effect of GAG accumulation i.e., "classical storage diseases". However, the data here

shows that it is more likely that the diverse presentations of the MPSs and other lysosomal diseases have complex underlying pathogenic mechanisms that are downstream and nonlinearlyrelated to the primary catabolic defect (i.e., the relationship between one or more pathogenic mechanisms may not be directly downstream of GAG catabolism or GAG accumulation). Previous studies of chondro-osseous disease in the MPSs have concentrated on examining late time points in disease natural history and have identified secondary mechanisms acting in the MPSs (described in section **1.6.** including oxidative stress, collagen fibril disorganization chondrocyte proliferation, inflammation, bone remodelling, and ossification). These studies represent the beginning of a shift in how "storage disorders" are conceptualized and describe the end results of the biomechanical failure model. The biomechanical failure model provides an explanation of the events preceding these mechanisms and further supports the shift in storage disorder conceptualization.

Although the mechanisms by which the primary disruption of GAG catabolism in MPS I lead to secondary extracellular matrix alterations described by this dissertation are not yet clear, the implications of these findings are profound. The mechanisms noted here would provide a potential explanation for the observation that current enzyme replacement strategies, via either intravenous ERT or hematopoietic stem cell transplant in MPS I, have a limited impact on chondro-osseous disease because the changes begin earlier than previously thought. Furthermore, these data suggest stronger rationale for early therapeutic intervention in patients. The overlap of chondro-osseous phenotypes between the MPSs suggests shared underlying pathogenic mechanisms, making the pathogenic changes discovered here and the biomechanical failure model relevant to understanding chondro-osseous disease more broadly in the MPSs.

Although the MPSs are monogenic, this dissertation suggests that the mechanisms underlying their pathogenesis are numerous and multifactorial.

Works Cited

1. Meikle PJ, Hopwood JJ, Clague AE, Carey WF. Prevalence of lysosomal storage disorders. JAMA 281 (1999) 249-254.

2. **Clarke LA, Heppner JM.** *Mucopolysaccharidosis Type I. 2002 Oct 31 [Updated 2011 Jul 21].* [Internet] [ed.] Adam MP, Ardinger HH, *et al.* Pagon RA. Seattle, WA : University of Washington, 2011. GeneReviews® [Internet] < http://www.ncbi.nlm.nih.gov/books/NBK1162/>.

3. Clarke, LA. *The mucopolysaccharidoses: a success of molecular medicine*. Epert Rev Mol Med 18 (2008) e1.

4. Ashton LJ, Brooks DA, McCourt PA, Muller VJ, Clements PR, Hopwood JJ. Immunoquantification and enzyme kinetics of alpha-L-iduronidase in cultured fibroblasts from normal controls and mucopolysaccharidosis type I patients. Am J Hum Genet 50 (1992) 787-794.

5. **Prasad VK, Kurtzberg J.** *Transplant Outcomes in Mucopolysaccharidoses*. Semin Hematol, 47 (2010) 59-69.

6. **Clarke, LA.** *Pathogenesis of skeletal and connective tissue involvement in the mucopolysaccharidoses: glycosaminoglycan storage is merely the instigator.* Rheumatology 50 (2011) 13-18.

7. Vijay S, Wraith JE. Clinical presentation and follow-up of patients with the attenuated phenotype of mucopolysaccharidosis type I. Acta Paediatr 94 (2005) 872-877.

8. Shull RM, Munger RJ, Spellacy E, Hall CW, Constantopoulos G, Neufeld EF. Canine alpha-L-iduronidase deficiency. A model of mucopolysaccharidosis I. Am J Pathol 109 (1982) 244-248.

9. Haskins ME, Jezyk PF, Desnick RJ, McDonough SK, Patterson DF. *Alpha-L-iduronidase deficiency in a cat: a model of mucopolysaccharidosis I.* Nature 13 (1979) 1294-1297.

10. Clarke LA, Russell CS, Pownall S, Warrington CL, Borowski A, Dimmick JE, Toone J, Jirik FR. *Murine mucopolysaccharidosis type I: targeted disruption of the murine alpha-L-iduronidase gene*. Hum Mol Genet 6 (1997) 503-511.

11. Ohmi K, Greenberg DS, Rajavel KS, Ryazantsev S, Li HH, Neufeld EF. Activated microglia in cortex of mouse models of mucopolysaccharidoses I and IIIB. 4, 2003, Proc Natl Acad Sci 100 (2003) 1902-1907.

12. Wang D, Shukla C, Liu X, Schoeb TR, Clarke LA, Bedwell DM, Keeling KM. *Characterization of an MPS I-H knock-in mouse that carries a nonsense mutation analogous to the human IDUA-W402X mutation.* Mol Genet Metab 99 (2010) 62-67.

13. Menon KP, Tieu PT, Neufeld EF. Architecture of the canine IDUA gene and mutation underlying canine mucopolysaccharidosis I. Genomics 14 (1992) 763-8.

14. He X, Chi-Ming L, Simonaro CM, Wan Q, Haskins ME, Desnick RJ, Schuchman EH. *Identification and CHaracterization of the Molecular Lesion Casuing Mucoolysccharidosis Type I in Cats.* Mol Genet Metab 67 (1999) 106-112.

15. **KK, White.** *Orthopaedic aspects of mucopolysaccharidoses.* Suppl 5, 2011, Rheumatology 50 (2011) v26-33.

16. **Opoka-Winiarska V, Jurecka A, Emeryk A, Tylki-Szymańska A.** Osteoimmunology in mucopolysaccharidoses type I, II, VI and VII. Immunological regulation of the osteoarticular system in the course of metabolic inflammation. 2013, Osteoarthritis Cartilage 12 (2013) 1813-1823.

17. Oussoren E, Brands MM, Ruijter GJ, der Ploeg AT, Reuser AJ. Bone, joint and tooth development in mucopolysaccharidoses: relevance to therapeutic options. 11, 2011, Biochim Biophys Acta 1812 (2011) 1542-1556.

18.

Natowicz MR, Short MP, Wang Y, Dickersin GR, Gebhardt MC, Rosenthal DI, Sims KB, Rosenberg AE.

Clinical and biochemical manifestations of hyaluronidase deficiency. N Engl J Med 114 (1996) 1029-1033.

19. Morishita K, Petty RE. *Musculoskeletal manifestations of mucopolysaccharidoses*. Rheumatology 50 (2011) 19-25.

20. de Ruijter J, Maas M, Janssen A, Wijburg FA. High prevalence of femoral head necrosis in Mucopolysaccharidosis type III (Sanfilippo disease) a national, observational, cross-sectional study. Mol Genet Metab 109 (2013) 49-53.

21. Laraway S, Breen C, Mercer J, Jones S, Wraith JE. Does early use of enzyme replacement therapy alter the natural history of mucopolysaccharidosis I Experience in three siblings. Mol Genet Metab 109 (2013) 316-316.

22. Sohn YB, Park SW, Kim SH, Cho SY, Ji ST, Kwon EK, Han SJ, Oh SJ, Park YJ, Ko AR, Paik KH, Lee J, Lee DH, Jin DK. *Enzyme replacement therapy improves joint motion and outcome of the 12-min walk test in a mucopolysaccharidosis type VI patient previously treated with bone marrow.* Am J Med Genet A 158A (2012) 1158-1163.

23. **WB., van den Berg.** *Osteoarthritis year 2010 in review: pathomechanisms*. Osteoarthritis Cartilage 19 (2011) 338-41.

24. Kwon JY, Ko K, Sohn YB, Kim SJ, Park SW, Kim SH, Cho SY, Jin DK. *High Prevalence of Carpal Tunnel Syndrome in Mucopolysaccharidosis Type II (Hunter Syndrome)*. Am J Med Genet A 155A (2011) 1329-1335.

25. Keen HI, Mease PJ, Bingham CO 3rd, Giles JT, Kaeley G, Conaghan PG. Systematic review of MRI, ultrasound, and scintigraphy as outcome measures for structural pathology in interventional therapeutic studies of knee arthritis: focus on responsiveness. The Journal of Rheumatology 38 (2011) 142-154.

26. Kubassova O1, Boesen M, Peloschek P, Langs G, Cimmino MA, Bliddal H, Torp-Pedersen S. *Quantifying disease activity and damage by imaging in rheumatoid arthritis and osteoarthritis*. Annals of the New York Academy of Sciences 1154 (2009) 207-238.

27. Rowan DJ, Tomatsu S, Grubb JH, Montaño AM, Sly WS. Assessment of bone dysplasia by micro-CT and glycosaminoglycan levels in mouse models for mucopolysaccharidosis type I, IIIA, IVA, and VII. J Inherit Metab Dis 36 (2013) 235-246.

28. Lin HY, Shih SC, Chuang CK, Chen MR, Niu DM, Lin SP. Assessment of bone mineral density by dual energy x-ray absorptiometry in patients with mucopolysaccharidoses. Orphanet Journal of Rare Diseases 11 (2013) 71.

29. Różdżyńska-Świątkowska A, Jurecka A, Cieślik J, Tylki-Szymańska A. Growth patterns in children with mucopolysaccharidosis I and II. World J Pediatr (2014) [Epub ahead of print].

30. Thomas JA, Beck M, Clarke JT, Cox GF. Childhood onset of Scheie syndrome, the attenuated form of mucopolysaccharidosis I. J Inherit Metab Dis 33 (2010) 421-427.

31. **Stattin EL, Tegner Y, Domellof M, Dhal N.** *Familial osteochondritis dissecans associated with early osteoarthritis and disproportionate short stature.* Osteoarthritis and Cartilage 16 (2008) 890-896.

32. Thompson SW, Merriman B, Funari VA, Feresquet M, Lachman RS, Rimoin DL, Nelson SF, Briggs MD, Cohn DH, Krakow D. *A recessive skeletal dysplasia, SEMD aggrecan type, results from a missense mutation affecting the C-type lectin domain of aggrecan.* 1, 2009, American Journal of Human Genetics 84 (2009) 72-79.

33. Kizawa H, Kou I, Iida A, Sudo A, Miyamoto Y, Fukuda A, Mabuchi A, Kotani A, Kawakami A, Yamamoto S, Uchida A, Nakamura K, Notoya K, Nakamura T Ikegawa S. *An aspartic acid repeat polymorphism in asporin inhibits chondrogenesis and increases susceptibility to osteoarthritis.* Nature Genetics 37 (2005) 138 - 144.

34. Hessle L, Stordalen GA, Wenglén C, Petzold C, Tanner EK, Brorson S, Baekkevold ES, Onnerfjord P, Reinholt FP, Heinegard D. *The Skeletal Phenotype of Chondroadherin Deficient Mice*. PLoS One 8 (2013) [E pub].

35. Suphapeetiporn, K, Tongkobpetch, S, Mahayosnond, A, Shotelersuk, V. Expanding the phenotypic spectrum of Caffey disease. Clinical Genetics 71 (2007) 280–284.

36. Beighton P, De Paepe A, Steinmann B, Tsipouras P, Wenstrup RJ. Ehlers-Danlos syndromes: revised nosology, Villefranche, 1997. Ehlers-Danlos National Foundation (USA) and Ehlers-Danlos Support Group (UK). Am J of Med Gen 77 (1998) 31-37.

37. Van Dijk FS, Sillence DO. Osteogenesis imperfecta: clinical, biochemical and molecular findings. Am J of Med Gen 164 (2014) 1470-1481.

38. Aederich MB, Horton WA. Achondroplasia: pathogenesis and implications for future treatment. Cur Oppin in Ped 22 (2010) 516-523.

39. Hoornaert KP, Marik I, Kozlowski K, Cole T, Le Merrer M, Leroy JG, Coucke PJ, Sillence D, Mortier GR. *Czech dysplasia metatarsal type: another type II collagen disorder*. European Journal of Human Genetics 15 (2007) 1269-1275.

40. Körkkö J, Cohn DH, Ala-Kokko L, Krakow D, Prockop DJ. Widely distributed mutations in the COL2A1 gene produce achondrogenesis type II/hypochondrogenesis. Am J of Med Gen 92 (2000) 95-100.

41. Wilkin DJ, Artz AS, South S, Lachman RS, Rimoin DL, Wilcox WR, McKusick VA, Stratakis CA, Francomano CA, Cohn DH. *Small deletions in the type II collagen triple helix produce kniest dysplasia.* Am J of Med Gen 85 (1999) 105-12.

42. Nishimura G, Nakashima E, Mabuchi A, Shimamoto K, Shimamoto T, Shimao Y, Nagai T, Yamaguchi T, Kosaki R, Ohashi H, Makita Y, Ikegawa S. *Identification of COL2A1 mutations in platyspondylic skeletal dysplasia, Torrance type.* J Med Genet 41 (2004) 75-79.

43. Tiller GE, Polumbo PA, Weis MA, Bogaert R, Lachman RS, Cohn DH, Rimoin DL, Eyre DR. Dominant mutations in the type II collagen gene, COL2A1, produce spondyloepimetaphyseal dysplasia, Strudwick type. Nat Genet 11 (1995) 87-89.

44. **Spranger J, Winterpacht A, Zabel B.** *The type II collagenopathies: a spectrum of chondrodysplasias.* Eur J Ped 153 (1994) 56-65.

45. Zankl A, Zabel B, Hilbert K, Wildhardt G, Cuenot S, Xavier B, Ha-Vinh R, Bonafé L, Spranger J, Superti-Furga A. Spondyloperipheral dysplasia is caused by truncating mutations in the C-propertide of COL2A1. Am J Med Genet A 129A (2004) 144-148.

46. **Snead MP, Yates JR.** *Clinical and Molecular genetics of Stickler syndrome.* Am J Med Genet 36 (1999) 353-359.

47. Mäkitie O, Susic M, Ward L, Barclay C, Glorieux FH, Cole WG. Schmid type of *metaphyseal chondrodysplasia and COL10A1 mutations--findings in 10 patients*. Am J Med Genet A 137A (2005) 241-248.

48. **Briggs MD, Chapman KL.** *Pseudoachondroplasia and multiple epiphyseal dysplasia: mutation review, molecular interactions, and genotype to phenotype correlations.* Hum Mutat 19 (2002) 465-478.

49. Griffith AJ, Sprunger LK, Sirko-Osadsa DA, Tiller GE, Meisler MH, Warman ML. *Marshall syndrome associated with a splicing defect at the COL11A1 locus.* Am J Med Genet 62 (1998) 816-823.

50. Bredrup C, Knappskog PM, Majewski J, Rødahl E, Boman H. Congenital stromal dystrophy of the cornea caused by a mutation in the decorin gene. 2, 2005, Investigative Ophthal Vis Sci 46 (2005) 420-426.

51. Johnson HJ, Rosenberg L, Choi HU, Garza S, Höök M, Neame PJ. Characterization of Epiphycan, a Small Proteoglycan with a Leucine-rich Repeat Core Protein. J Biol Chem 272 (1997) 18709-18717.

52. Brunner M, Millon-Frémillon A, Chevalier G, Nakchbandi IA, Mosher D, Block MR, Albigès-Rizo C, Bouvard D. Osteoblast mineralization requires betal integrin/ICAP-1dependent fibronectin deposition. J Biol Chem 194 (2007) 307-322.

53. Akawi NA, Ali BR, Al-Gazali L. *A novel mutation in PRG4 gene underlying camptodactyly-arthropathy-coxa vara-pericarditis syndrome with the possible expansion of the phenotype to include congenital cataract*. Birth Defects Res A Clin Mol Teratol 94 (2012) 553-556.

54. Neame PJ, Kay CJ, McQuillan DJ, Beales MP, Hassell JR. Independent modulation of collagen fibrillogenesis by decorin and lumican. Cell Mol Life Sci 57 (2000) 859-863.

55. Buckner JH, Wu JJ, Reife RA, Terato K, Eyre DR. Autoreactivity against matrilin-1 in a patient with relapsing polychondritis. Arthritis Rheum 43 (2000) 939-943.

56. **Pataquiva-Mateus AY, Wu HC, Lucchesi C, Ferraz MP, Monteiro FJ, Spector M.** *Supplementation of collagen scaffolds with SPARC to facilitate mineralization.* J Biomed Mater Res B Appl Biomater 100 (2012) 862-870.

57. Stum M, Davoine CS, Vicart S, Guillot-Noël L, Topaloglu H, Carod-Artal FJ, Kayserili H, Hentati F, Merlini L, Urtizberea JA, Hammouda el-H, Quan PC, Fontaine B, Nicole S. Spectrum of HSPG2 (Perlecan) mutations in patients with Schwartz-Jampel syndrome. 11, 2006, Hum Mutat 27 (2006) 1082-1091.

58. Arikawa-Hirasawa E, Wilcox WR, Le AH, Silverman N, Govindraj P, Hassell JR, Yamada Y. Dyssegmental dysplasia, Silverman-Handmaker type, is caused by functional null mutations of the perlecan gene. Nat Genet 27 (2001) 431-434.

59. **E**, **Mornet**. *Hypophosphatasia: the mutations in the tissue-nonspecific alkaline phosphatase gene*. Hum Mutat 1 (2000) 309-315.

60. Schönherr E, Witsch-Prehm P, Harrach B, Robenek H, Rauterberg J, Kresse H. *Interaction of biglycan with type I collagen.* J Biol Chem 270 (1995) 2776-2783.

61. Rios H, Koushik SV, Wang H, Wang J, Zhou HM, Lindsley A, Rogers R, Chen Z, Maeda M, Kruzynska-Frejtag A, Feng JQ, Conway SJ. Periostin Null Mice Exhibit Dwarfism, Incisor Enamel Defects, and an Early-Onset Periodontal Disease-Like Phenotype. Mol Cell Biol 25 (2005) 11131-1144.

62. Kennedy AM, Inada M, Krane SM, Christie PT, Harding B, López-Otín C, Sánchez LM, Pannett AA, Dearlove A, Hartley C, Byrne MH. *MMP13 mutation causes spondyloepimetaphyseal dysplasia, Missouri type (SEMDMO).* J Clin Invest 115 (2005) 2832-2842.

63. Lausch E, Keppler R, Hilbert K, Cormier-Daire V, Nikkel S, Nishimura G, Unger S, Spranger J, Superti-Furga A, Zabel B. *Mutations in MMP9 and MMP13 determine the mode of inheritance and the clinical spectrum of metaphyseal anadysplasia*. Am J Hum Genet 85 (2009) 168-178.

64. Rosenthal AK, Gohr CM, Uzuki M, Masuda I. Osteopontin promotes pathologic mineralization in articular cartilage. Matrix Biology 26 (2007) 96-105.

65. **Thompson Z, Miclau T, Hu D, Helms JA.** A model for intramembranous ossification during fracture healing. J Orthop Res 20 (2002) 1091–1098.

66. Xian CJ, Zhou FH, McCarty RC, Foster BK. Intramembranous ossification mechanism for bone bridge formation at the growth plate cartilage injury site. J Orthop Res 22 (2004) 417-426.

67. Hall BK, Miyake T. *All for one and one for all: condensations and the initiation of skeletal development.* Bioessays 22 (2000) 138–147.

68. Bi W, Deng JM, Zhang Z, Behringer RR, Crombrugghe B. Sox9 is required for cartilage formation. Nat Genet 22 (1999) 85-89.

69. Yoon, BS and Lyons, KM. *Multiple functions of BMPs in chondrogenesis*. J Cell Biochem 93 (2004) 93-103.

70. Karsenty G, Kronenberg HM, Settembre, C. *Genetic control of bone formation*. Annu Rev Cell Dev Biol 25 (2009) 629-648.

71. **HM, Kronenberg.** *Developmental regulation of the growth plate.* Nature 423 (2003) 332-336.

72. Karp SJ, Schipani E, St-Jauques B, Hunzelman J, Kronenberg H, McMahon AP. *Ihh and PTHrP in endochondral bone development*. Development 127 (2000) 543-548.

73. Wagner EF, Karsenty G. *Genetic control of skeletal development*. Curr Opin Genet Dev 11 (2001) 527-532.

74. **GA, Harada S and Rodan.** *Control of osteoblast function and regulation of bone mass.* Nature 423 (2003) 349-355.

75. HC, Blair. How the osteoclast degrades bone. Bioessays 20 (1998) 837-846.

76. Crockett JC, Rogers MJ, Coxon FP, Hocking LJ, Helfrich MH. Bone remodelling at a glance. J Cell Sci 124 (2011) 991-998.

77. Xing L, Schwarz EM, Boyce BF. Osteoclast precursors, RANKL/RANK, and immunology. Immun Rev 208 (2005) 19-29.

78. SL, Teitelbaum. Bone Resorption by Osteoclasts. Science 289 (2000) 1504-1508.

79. Kreja L, Liedert A, Hasni S, Claes L, Ignatius A. *Mechanical regulation of osteoclastic genes in human osteoblasts*. Biochem Biophys Res Commun 368 (2008) 582-587.

80. Young RD, Liskova P, Pinali C, Palka BP, Palos M, Jirsova K, Hrdlickova E, Tesarova M, Elleder M, Zeman J, Meek KM, Knupp C, Quantock AJ. Large proteoglycan complexes and disturbed collagen architecture in the corneal extracellular matrix of mucopolysaccharidosis type VII (Sly syndrome). Invest Ophthalmol Vis Sci 52 (2011) 6720-6728.

81. von der Mark, K. Structure, biosynthesis, and gene regulation of collagens in cartilage and bone. [book auth.] Robins SP, Bilezikian JP Seibel MJ. [ed.] 2. *Dynamics of Bone and Cartilage Metabolism: Principles and Clinical Applications*. Burlington : Academic Press, 2006, pp. 3-40.

82. Filippon L, Vanzin CS, Biancini GB, Pereira IN, Manfredini V, Sitta A, Peralba Mdo C, Schwartz IV, Giugliani R, Vargas CR. Oxidative stress in patients with mucopolysaccharidosis type II before and during enzyme replacement therapy. Mol Genet Metab103 (2011) 121-127.

83. Reolon GK, Reinke A, de Oliveira MR, Braga LM, Camassola M, Andrades ME, Moreira JC, Nardi NB, Roesler R, Dal-Pizzol F. *Alterations in oxidative markers in the cerebellum and peripheral organs in MPS I mice*. Cell Mol Neurobiol 29 (2009) 443-448.

84. Alliston T. Chondroitin sulfate and growth factor signaling in the skeleton: Possible links to MPS VI. J Pediatr Rehabil Med 3 (2010) 129-138.

85. Klüppel M, Wight TN, Chan C, Hinek A, Wrana JL. *Maintenance of chondroitin sulfation balance by chondroitin-4-sulfotransferase 1 is required for chondrocyte development and growth factor signaling during cartilage morphogenesis.* 17, 2005, Development 132 (2005) 3989-4003.

86. **Alliston A, Piek E, Derynck R.** TGF-β family signaling in skeletal development, maintenance, and disease. [book auth.] Miyazono K Deryck R. *The TGF-b Family*. Woodbury : Cold Spring Harbor Press, 2008, pp. 667-723.

87. Gupta P, Oegema TR Jr, Brazil JJ, Dudek AZ, Slungaard A, Verfaillie CM. Structurally specific heparan sulfates support primitive human hematopoiesis by formation of a multimolecular stem cell niche. Blood 92 (1998) 4641-4651.

88. Metcalf JA, Zhang Y, Hilton MJ, Long F, Ponder KP. *Mechanism of shortened bones in mucopolysaccharidosis VII*. Mol Genet Metab 7 (2009) 202-211.

89. Gatto F, Redaelli D, Salvadè A, Marzorati S, Sacchetti B, Ferina C, Roobrouck VD, Bertola F, Romano M, Villani G, Antolini L, Rovelli A, Verfaillie CM, Biondi A, Riminucci M, Bianco P, Serafini M. Hurler disease bone marrow stromal cells exhibit altered ability to support osteoclast formation. Stem Cells Dev 21 (2012) 1466-1477.

90. Nuttall JD, Brumfield LK, Fazzalari NL, Hopwood JJ, Byers S. *Histomorphometric* analysis of the tibial growth plate in a feline model of mucopolysaccharidosis type VI. Calcif Tissue Int 65 (1999) 47-52.

91. Monroy MA, Ross FP, Teitelbaum SL, Sands MS. *Abnormal osteoclast morphology and bone remodelling in a murine model of a lysosomal storage disease*. Bone 30 (2002) 352-9.

92. Thawrani DP, Walker K, Polgreen LE, Tolar J, Orchard PJ. *Hip dysplasia in patients with Hurler syndrome (mucopolysaccharidosis type 1H)*. J Pediatr Orthop 33 (2013) 635-43.

93. Russell C, Hendson G, Jevon G, Matlock T, Yu J, Aklujkar M, Ng KY, Clarke LA. *Murine MPS I: insights into the pathogenesis of Hurler syndrome.* Clin Genet 53 (1998) 349-361.

94. Schuchman EH, Ge Y, Lai A, Borisov Y, Faillace M, Eliyahu E, He X, Iatridis J, Vlassara H, Striker G, Simonaro CM. *Pentosan polysulfate: a novel therapy for the mucopolysaccharidoses*. PLoS One 8 (2013) [E pub].

95. Simonaro CM, D'Angelo M, He X, Eliyahu E, Shtraizent N, Haskins ME, Schuchman EH. *Mechanism of Glycosaminoglycan-Mediated Bone and Joint Disease*. Am J Pathol 172 (2008) 112-122.

96. Clarke LA, Wraith JE, Beck M, Kolodny EH, Pastores GM, Muenzer J, Rapoport DM, Berger KI, Sidman M, Kakkis ED, Cox GF. Long-term efficacy and safety of laronidase in the treatment of mucopolysaccharidosis I. Pediatrics 123 (2009) 229-240.

97. Weisstein JS, Delgado E, Steinbach LS, Hart K, Packman S. *Musculoskeletal manifestations of Hurler syndrome: long-term follow-up after bone marrow transplantation.* J Pediatr Orthop 24 (2004) 97-101.

98. Taylor C, Brady P, O'Meara A, Moore D, Dowling F, Fogarty E. *Mobility in Hurler syndrome*. J Pediatr Orthop 28 (2008) pp. 163-168.

99. Wraith EJ, Hopwood JJ, Fuller M, Meikle PJ, Brooks DA. Laronidase treatment of *mucopolysaccharidosis I.* 2005, BioDrugs 19 (2005) 1-7.

100. van der Linden MH, Kruyt MC, Sakkers RJ, de Koning TJ, Oner FC, Castelein RM. Orthopaedic management of Hurler's disease after hematopoietic stem cell transplantation a systematic review. J Inherit Metab Dis 34 (2011) 657-669.

101. Schulze-Frenking G, Jones SA, Roberts J, Beck M, Wraith JE. *Effects of enzyme replacement therapy on growth in patients with mucopolysaccharidosis type II.* J Inherit Metab Dis 34 (2011) 203-208.

102. Kakkis ED, Muenzer J, Tiller GE, Waber L, Belmont J, Passage M, Izykowski B, Phillips J, Doroshow R. *Enzyme-Replacement Therapy in Mucopolysaccharidosis I*. NEMJ 344 (2000) 182-188.

103. Sifuentes M, Doroshow R, Hoft R, Mason G, Walot I, Diament M, Okazaki S, Huff K, Cox GF, Swiedler SJ, Kakkis ED. *A follow-up study of MPS I patients treated with laronidase enzyme replascement therapy for 6 years.* Mol Genet Metab 90 (2007) 171-180.

104. Hansen K, Sung CJ, Huang C, Pinar H, Singer DB, Oyer CE. *Reference values for second trimester fetal and neonatal organ weights and measurements*. Pediatr Dev Pathol 6 (2003) 160-167.

105. Kuczmarski RJ, Ogden CL, Guo SS, Grummer-Strawn LM, Flegal KM, Mei Z, Wei R, Curtin LR, Roche AF, Johnson CL. 2000 CDC Growth Charts for the United States: Methods and Development. Vital Health Stat 11 (2000) 1-190.

106. **Rogol AD, Clark PA, Roemmich JN.** *Growth and pubertal development in children and adolescents: effects of diet and physical activity.* Am J Clin Nutr 72 (2000) 521-528.

107. Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. *Quantitative mass spectrometry in proteomics: a critical review.* Anal Bioanal Chem 389 (2007) 1017-1031.

108. Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature 422 (2003) 198-207.

109. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A, Pappin DJ. *Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents*. Mol Cell Proteomics 3 (2004) 1154-1169.

110. **Applied Biosystems, AB-SCIEX.** Multiplex Protein Quantitation using iTRAQ Reagents – 8plex Applied Biosystems (2011) <

http://www.google.ca/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCMQFjAA&url =http%3A%2F%2Ftools.lifetechnologies.com%2Fcontent%2Fsfs%2Fmanuals%2Fcms_049786. pdf&ei=IGj3VNaXGcWsogSpvYDoBw&usg=AFQjCNG0xcv9Wg2jkmwIl2TYJhAolP50BA&s ig2=Q8_N6OIAIziZYDrhqnIGvA&bvm=bv.87519884,d.cGU&cad=rja>

111. **Applied BioSystems, AB-SCIEX.** Applied Biosystems iTRAQ[™] Reagents, Chemistry Reference Guide. Applied BioSystems, (2004) <

http://www.google.ca/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved= 0CB4QFjAA&url=http%3A%2F%2Fwww3.appliedbiosystems.com%2Fcms%2Fgroups%2Fps m_marketing%2Fdocuments%2Fgeneraldocuments%2Fcms_041463.pdf&ei=Zmj3VJjOJ4zxoA TnhYKQCg&usg=AFQjCNEeYFptdtEk6_DdEyO5wdmfiJyc0Q&sig2=nB6JKTAjcNfp3kDakQucg&bvm=bv.87519884,d.cGU>.

112. Jiang X, Ye M, Jiang X, Liu G, Feng S, Cui L, Zou H. Method Development of Efficient Protein Extraction in Bone Tissue for Proteomic Analysis. J Proteome Res 6 (2007) 2287-2294.

113. **RA**, **Zubarev**. *The challenge of the proteome dynamic range and its implications*. Proteomics 13 (2013) 723-726.

114. **Randall DR, Sinclair GB, Colobong, KE, Hetty E, Clarke LA.** *Heparin cofactor IIthrombin complex in MPS I: A biomarker of MPS disease.* Mol. Genet. Metab 88 (2006) 235-243.

115. **Oberg AL, Vitek O.** *Statistical design of quantitative mass spectrometry-based proteomic experiments.* J Proteome Res 8 (2009) 2144-2156.

116. Shih JH, Michalowska AM, Dobbin K, Ye Y, Qiu TH, Green JE. *Effects of pooling*. Bioinformatics 20 (2004) 3318-3325.

117. Kendziorski C, Irizarry RA, Chen KS, Haag JD, Gould MN. On the utility of pooling biological samples in microarray experiments. Proc Natl Acad Sci USA 102 (2005) 4252-4257.

118. Weise M, De-Levi S, Barnes KM, Gafni RI, Abad V, Baron J. *Effects of estrogen on growth plate senescence and epiphyseal fusion*. Proc Natl Acad Sci USA 98 (2001) 6871-6876.

119. Chagin AS, Lindberg MK, Andersson N, Moverare S, Gustafsson JA, Sävendahl L, Ohlsson C. Estrogen receptor-beta inhibits skeletal growth and has the capacity to mediate growth plate fusion in female mice. J Bone Min Res 19 (2004) 72-77.

120. Börjesson AE, Lagerquist MK, Liu C, Shao R, Windahl SH, Karlsson C, Sjögren K, Movérare-Skrtic S, Antal MC, Krust A, Mohan S, Chambon P, Sävendahl L, Ohlsson C. *The role of estrogen receptor* α *in growth plate cartilage for longitudinal bone growth*. J Bone Min Res 25 (2010) 2690-2700.

121. Whittingham, D. G., and Wood, M. J. Reproductive Physiology. [book auth.] H. L., Small, J. D., and Fox, J. G. Foster. *The Mouse in Biomedical Research*. N.Y. : Academic Press, 1983, 137-164.

122. **Higgins L, Jagtap P.** Introduction to Protein Pilot Data Analysis. Minneapolis, Minnisota, USA : Regents of the University of Minnesota, December 1, 2009 < http://www.google.ca/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved= 0CB4QFjAA&url=http%3A%2F%2Fstatic.msi.umn.edu%2Ftutorial%2Flifescience%2FProteinP ilot_Intro_120109.pdf&ei=fGr3VPi3E4y3ogT5wYHYCQ&usg=AFQjCNGRjVqWOGKRj7UB 1RkJypyXOUcIVA&sig2=oNLZ6yvf7bRrBclvA96eOQ&bvm=bv.87519884,d.cGU>.

123. Ow SY, Salim M, Noirel J, Evans C, Rehman I, Wright PC. *iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly".* J Proteome Res 8 (2009) 5347-5355.

124. Kerlavage A, Bonazzi V, di Tommaso M, Lawrence C, Li P, Mayberry F, Mural R, Nodell M, Yandell M, Zhang J, Thomas P. *The Celera Discovery System*. Nucleic Acids Res 30 (2002) 129-136.

125. Bult CJ, Eppig JT, Kadin JA, Richardson JE, Blake JA and Group, and the members of the Mouse Genome Database. *The Mouse Genome Database (MGD): mouse biology and model systems*. Nucleic Acids Res 36 (2008) D724-728.

126. The UniProt Consortium. *The Universal Protein Resource (UniProt) in 2010*. Nucleic Acids Res 38 (2010) 142-148.

127. **Applied Biosystems, AB-SCIEX.** ProteinPilot Software Online Help. Applied Bisosystems, 2011 [software help files].

128. P., Shaffer J. Multiple Hypothesis Testing. Ann. Rev. Psych 46 (1995) 561-584.

129. Nesvizhskii AI, Vitek O, Aebersold R. Analysis and validation of proteomic data generated by tandem mass spectrometry. Nat Methods 4 (2007) 787-797.

130. Thomas-Chollier M, Defrance M, Medina-Rivera A, Sand O, Herrmann C, Thieffry D, van Helden J. *RSAT 2011: regulatory sequence analysis tools.* 2011, Nucleic Acids Res 39 (2011) W86-W91.

131. Heinemeyer T, Wingender E, Reuter I, Hermjakob H, Kel AE, Kel OV, Ignatieva EV, Ananko EA, Podkolodnaya OA, Kolpakov FA, Podkolodny NL, Kolchanov NA. Databases

on transcriptional regulation: TRANSFAC, TRRD and COMPEL. Nucleic Acids Res 26 (1998) 262-267.

132. Adler P, Kolde R, Kull M, Tkachenko A, Peterson H, Reimand J, Vilo J. *Mining for coexpression across hundreds of datasets using novel rank aggregation and visualisation methods*. Genome Biol 10 (2009) [E pub].

133. Heppner JM, Zaucke F, Clarke LA. Extracellular matrix disruption is an early event in the pathogenesis of skeletal disease in mucopolysaccharidosis I. Mol Genet Metab 114 (2015) 146-155.

134. **Illumina.** *Whole-Genome Gene Expression Direct Hybridization Assay Guide* < http://www.google.ca/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved= 0CCMQFjAA&url=http%3A%2F%2Fsupport.illumina.com%2Fcontent%2Fdam%2Filluminasupport%2Fdocuments%2Fmyillumina%2F3466bf71-78bd-4842-8bfc-393a45d11874%2Fwggex_direct_hybridization_assay_guide_11322355_a.pdf&ei=VWz3VKfMZDhoATK5IKIBQ&usg=AFQjCNHF6fAwg8qA0qDBD_BjOmxOFOxUTw&sig2=n6Hb1C SUGc4-Nnyf7d11zg&bvm=bv.87519884,d.cGU>.

135. **SABiosciences.** *RT2 Profiler PCR Array Handbook* < *http://www.qiagen.com/us/resources/resourcedetail?id=6161ebc1-f60f-4487-8c9e-9ce0c5bc3070>*.

136. Önnerfjord P, Khabut A, Reinholt FP, Svensson O, Heinegård D. Quantitative proteomic analysis of eight cartilaginous tissues reveals characteristic differences as well as similarities between subgroups. J Biol Chem 287 (2012) 18913-18924.

137. Bell PA, Wagener R, Zaucke F, Koch M, Selley J, Warwood S, Knight D, Boot-Handford RP, Thornton DJ, Briggs MD. Analysis of the cartilage proteome from three different mouse models of genetic skeletal diseases reveals common and discrete disease signatures. Biol. Open 2 (2013) 802-811.

138. Gan CS, Chong PK, Pham TK, Wright PC. *Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ)*. J Proteome Resh 6 (2007) 821-827.

139. Bornstein P, Sage EH. *Matricellular proteins: extracellular mediators of cell function*. Curr Opin Cell Biol 14 (2002) 608-616.

140. Sandelin A, Alkema W, Engstrom P, Wasserman, WW, Lenhard B. *JASPAR: an openaccess database for eukaryotic transcription factor binding profiles*. Suppl 1, 2004, Nucl Acid Res 32 (2004) D91-94.

141. Lange V, Picotti P, Domon B, Aebersold R. Selected reaction monitoring for quantitative proteomics: a tutorial. Mol Sys Biol 4 (2008) [E pub].

142. Kitteringham NR, Jenkins RE, Lane CS, Elliott VL, Park BK. *Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics*. J Anal Chrom B 877 (2009) 1229-1239.

143. **Miguez PA, Terajima M, Nagaoka H, Ferreira JA, Braswell K, Ko CC, Yamauchi M.** *Recombinant biglycan promotes bone morphogenetic protein-induced osteogenesis.* J Dent Res 93 (2014) 406-411.

144. Halper J. Proteoglycans and diseases of soft tissues. Adv Exp Med Biol 802 (2014) 49-58.

145. Genovese F, Barascuk N, Larsen L, Larsen MR, Nawrocki A, Li Y, Zheng Q, Wang J, Veidal SS, Leeming DJ, Karsdal MA. *Biglycan fragmentation in pathologies associated with extracellular matrix remodeling by matrix metalloproteinases*. Fibrogenesis Tissue Repair 6 (2013) [E pub].

146. Berendsen AD, Fisher LW, Kilts TM, Owens RT, Robey PG, Gutkind JS, Young MF. *Modulation of canonical Wnt signaling by the extracellular matrix component biglycan*. Proc Nat Acad Sci USA 108 (2011) 17022-17027.

147. Burton-Wurster N, Liu W, Matthews GL, Lust G, Roughley PJ, Glant TT, Cs-Szabó G. *TGF beta 1 and biglycan, decorin, and fibromodulin metabolism in canine cartilage.* Osteoarthritis Cartilage 11(2003) 167-176.

148. Lotz M, Martel-Pelletier J, Christiansen C, Brandi ML, Bruyère O, Chapurlat R, Collette J, Cooper C, Giacovelli G, Kanis JA, Karsdal MA, Kraus V, Lems WF, Meulenbelt I, Pelletier JP, Raynauld JP, Reiter-Niesert S, Rizzoli R, Sandell LJ, Van Spil WE, Regin. Value of biomarkers in osteoarthritis: current status and perspectives. Ann Rheum Dis 72 (2013) 1756-1763.

149. Farrah T, Deutsch EW, Omenn GS, Campbell DS, Sun Z, Bletz JA, Mallick P, Katz JE, Malmström J, Ossola R, Watts JD, Lin B, Zhang H, Moritz RL, Aebersold R. *A High-Confidence Human Plasma Proteome Reference Set with Estimated Concentrations in PeptideAtlas*. Mol Cell Proteomics 10 (2011) [E pub].

150. Craig R, Cortens JP, Beavis RC. Open source system for analyzing, validating, and storing protein identification data. J Proteome Res 3 (2004) 1234-1242.

151. Glasson SS, Chambers MG, Van Den Berg WB, Little CB. *The OARSI histopathology initiative - recommendations for histological assessments of osteoarthritis in the mouse*. Osteoarthritis Cartilage 18 (2010) 17-23.

152. **Gry M, Rimini R, Stromberg S, Asplund A, Ponten F, Uhlen M, Nilsson**. *Correlations between RNA and protein expression profiles in 23 human cell lines*. BMC Genomics 365 (2009) [E pub].

153. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses Nat Review Genet 13 (2012) 227-232.

154. Wang Z, Gerstein M, Snyder M. *RNA-Seq: a revolutionary tool for transcriptomics*. Nat Review Genet 10 (2009) 57-63.

155. Rogler CE, Tchaikovskaya T, Norel R, Massimi A, Plescia C, Rubashevsky E, Siebert P, Rogler LE. *RNA expression microarrays (REMs), a high-throughput method to measure differences in gene expression in diverse biological samples* Nuleic Acids Res 32 (2004) e120 [E pub].

156. **Illumina Inc.** *Array Based Gene Expression Analysis*. Illumina < http://www.google.ca/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&uact=8&ved= 0CC0QFjAB&url=http%3A%2F%2Fwww.illumina.com%2Fdocuments%2Fproducts%2Fdatash eets%2Fdatasheet_gene_exp_analysis.pdf&ei=l7b4VNPnAtDSoATBioCYCQ&usg=AFQjCNH Vxx-K5pv8ffNvtjjRdGGM4eD8Pg&sig2=vVMqf9IuqPJHYWkmOm1Cw&bvm=bv.87611401,d.cGU>.

157. Rio DC, Ares M Jr, Hannon GJ, Nilsen TW. *Purification of RNA using TRIzol (TRI reagent)*. 2010, Cold Spring Harb Protoc 6 (2010) [E pub].

158. Aglient Technologies Inc. Aglient 2100 Bioanalyzer 2100 Expert User's Guide. Waldbronn, Germany (2003) < http://www.google.ca/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved= 0CCEQFjAA&url=http%3A%2F%2Farboretum.harvard.edu%2Fwpcontent%2Fuploads%2FAgilent-2100-Bioanalyzer-Software-Guide.pdf&ei=-Lb4VPKzNdSwogSUw4LgCg&usg=AFQjCNHfJEeNmp_CsKsOF0XlcKywS92fA&sig2=owh0SXkuAdhSN7_8y4nG6Q&bvm=bv.87611401,d.cGU>.

159. Kubista M, Andrade JM, Bengtsson M, Forootan A, Jonák J, Lind K, Sindelka R, Sjöback R, Sjögreen B, Strömbom L, Ståhlberg A, Zoric N. *The real-time polymerase chain reaction*. Mol Aspects Med 27 (2006) 95-125.

160. Cao H, Shockey JM. Comparison of TaqMan and SYBR Green qPCR methods for quantitative gene expression in tung tree tissues. J Agric Food Chem (2012) 12296-122303.

161. **SABiosciences.** *RT2 Profiler*TM*PCR Arrays: Pathway-Focused Gene Expression Profiling with qRT-PCR. SABiosciences* (\leq

 $\label{eq:http://www.google.ca/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0CCMQFjAA&url=http%3A%2F%2Fwww.sabiosciences.com%2Fmanuals%2Fpcrarraywhitepaper.pdf&ei=XLj4VIOMDIzXoASbn4CADw&usg=AFQjCNF40PiOUVNvV_uPVIvdQ-17ETbmkA&sig2=QHba367v6cqo1SIE7k87zA&bvm=bv.87611401,d.cGU>.$

162. **R Development Core Team.** *R: A language and environment forstatistical computing, reference index version 2.x.x.* R Foundation for Statistical Computing (2005) < http://www.google.ca/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved= 0CCEQFjAA&url=http%3A%2F%2Fweb.mit.edu%2Fr_v3.0.1%2Ffullrefman.pdf&ei=x7j4VN G8H8XgoATczYLQCw&usg=AFQjCNGmPXwWrqvpnl7awxnDRBeOHUhaQ&sig2=Fm1FeuEZ1G9sMi8pgi05AQ&bvm=bv.87611401,d.cGU>.

163. **Illumina Inc.** *GenomeStudio Gene Expression Module v1.0 User Guide*. Illumina < http://support.illumina.com/downloads/genomestudio_gx_module_v10_ug_%2811319121_a%2 9.html>.

164. **Smyth GK.** *Linear models and empirical Bayes methods for assessing dierential expression in microarray experiments.* Stat Appl Genet Mol Biol (2014) [E pub].

165. Vrtačnik P, Marc J, Ostanek B. *Epigenetic mechanisms in bone*. Clin Chem Lab Med 52 (2014) 589-608.

166. yddall CM, Reynard LN, Young DA, Loughlin J. The identification of trans-acting factors that regulate the expression of GDF5 via the osteoarthritis susceptibility SNP rs143383. PLoS Genetics 9 (2013) [E pub].

167. **G, Nuki.** Role of Mechanical Factors in the Aetiology, Pathogenesis and Progression of Osteoarthritis. [book auth.] Pelletier J-P, Martel-Pelletier J, Henrotin Y, Crasborn L Reginster J-Y. *Osteoarthritis: Clinical and Experimental Aspects.* Springer Berlin Heidelberg, 1999 101-114.

168. **Beasley J.** Osteoarthritis and Rheumatoid Arthritis: Conservative Therapeutic Management. J Hand Ther 5 (2012) 163-171.

169. **Sun HB.** *Mechanical loading, cartilage degradation, and arthritis.* Ann N Y Acad Sci 1211 (2010) 37-50.

170. Griffin TM, Guilak F. *The Role of Mechanical Loading in the Onset and Progression of Osteoarthritis.* Exerc Sport Sci Rev 33 (2005) 195-200.

171. Cui Y, Zhao H, Liu Z, Liu C, Luan J, Zhou X, Han J. *A systematic review of genetic skeletal disorders reported in Chinese biomedical journals between 1978 and 2012*. Orphanet J Rare Dis 7 (2012) [E pub].

172. Krakow D, Rimoin DL. The skeletal dysplasias. Genet Med 12 (2010) 327-341.

173. Jackson GC, Mittaz-Crettol L, Taylor JA, Mortier GR, Spranger J, Zabel B, Le Merrer M, Cormier-Daire V, Hall CM, Offiah A, Wright MJ, Savarirayan R, Nishimura G, Ramsden SC, Elles R, Bonafe L, Superti-Furga A, Unger S, Zankl A, Briggs MD.

Pseudoachondroplasia and multiple epiphyseal dysplasia: a 7-year comprehensive analysis of the known disease genes identify novel and recurrent mutations and provides an accurate assessment of their relative contribution. Human Mutation 33 (2012) 144-157.
174. **Bornstein P, Sage EH.** *Matricellular proteins: extracellular modulators of cell function.* Curr Opin Cell Biol 14 (2002) 608-616.

175. Ye Y, Hu W, Guo F, Zhang W, Wang J, Chen A. *Glycosaminoglycan chains of biglycan promote bone morphogenetic protein-4-induced osteoblast differentiation*. Int J Mol Med 30, (2012) 1075-1080.

176. Merline R, Schaefer RM, Schaefer L. *The matricellular functions of small leucine-rich proteoglycans (SLRPs)*. J Cell Commun Signal 3 (2009) 323-335.

177. Chen J, Gupta T, Barasz JA, Kalajzic Z, Yeh WC, Drissi H, Hand AR, Wadhwa S. *Analysis of microarchitectural changes in a mouse temporomandibular joint osteoarthritis model*. Arch Oral Biol 54 (2009) 1091-1098.

178. Embree MC, Kilts TM, Ono M, Inkson CA, Syed-Picard F, Karsdal MA, Oldberg A, Bi Y, Young MF. *Biglycan and fibromodulin have essential roles in regulating chondrogenesis and extracellular matrix turnover in temporomandibular joint osteoarthritis.* Am J Pathol 176 (2010) 812-826.

179. Young MF, Bi Y, Ameye L, Chen XD. *Biglycan knockout mice: new models for musculoskeletal diseases.* Glycoconj J 19 (2002) 257-262.

180. Song R, Zeng Q, Ao L, Yu JA, Cleveland JC, Zhao KS, Fullerton DA, Meng X. *Biglycan induces the expression of osteogenic factors in human aortic valve interstitial cells via Toll-like receptor-2.* Arterioscler Thromb Vasc Biol 32 (2012) 2711-2720.

181. Juneja SC, Veillette C. Defects in tendon, ligament, and enthesis in response to genetic alterations in key proteoglycans and glycoproteins: a review. Arthritis (2013) [E pub].

182. Young MF, Fallon JR. *Biglycan: a promising new therapeutic for neuromuscular and musculoskeletal diseases.* Curr Opin Genet Dev 22 (2012) 398-400.

183. Bengtsson E, Mörgelin M, Sasaki T, Timpl R, Heinegård D, Aspberg A. *The leucinerich repeat protein PRELP binds perlecan and collagens and may function as a basement membrane anchor.* J Biol Chem 277 (2002) 15061-15068.

184. **Tillgren V, Onnerfjord P, Haglund L, Heinegård D.** *The tyrosine sulfate-rich domains of the LRR proteins fibromodulin and osteoadherin bind motifs of basic clusters in a variety of heparin-binding proteins, including bioactive factors.* J Biol Chem 284 (2009) 28543-2854353.

185. Happonen KE, Fürst CM, Saxne T, Heinegård D, Blom AM. *PRELP protein inhibits the formation of the complement membrane attack complex*. J Biol Chem 287 (2012) 8092-8100.

186. Homan EP, Rauch F, Grafe I, Lietman C, Doll JA, Dawson B, Bertin T, Napierala D, Morello R, Gibbs R, White L, Miki R, Cohn DH, Crawford S, Travers R, Glorieux FH, Lee

B. *Mutations in SERPINF1 cause osteogenesis imperfecta type VI.* J Bone MinRes 26 (2011) 2798-2803.

187. Tortorella MD, Malfait AM, Deccico C, Arner E. *The role of ADAM-TS4 (aggrecanase-1) and ADAM-TS5 (aggrecanase-2) in a model of cartilage degradation*. Osteoartritis Cartilage 9 (2001) 539-552.

188. Fosang AJ, Last K, Maciewicz RA. *Aggrecan is degraded by matrix metalloproteinases in human arthritis. Evidence that matrix metalloproteinase and aggrecanase activities can be independent.* J Clin Invest 98 (1996) 2292-2299.

189. McCawley LJ, Matrisian LM. *Matrix metalloproteinases: they're not just for matrix anymore!* Cur Opin Cell Biol 13 (2001) 534-540.

190. Lee A, Choi SJ, Park K, Park JW, Kim K, Choi K, Yoon SY, Youn I. Detection of active matrix metalloproteinase-3 in serum and fibroblast-like synoviocytes of collagen-induced arthritis mice. Bioconjug Chem 24 (2013) 1068-1074.

191. Fässler R, Schnegelsberg PN, Dausman J, Shinya T, Muragaki Y, McCarthy MT, Olsen BR, Jaenisch R. Mice lacking alpha 1 (IX) collagen develop noninflammatory degenerative joint disease. PNAS 91 (1994) 5070-5074.

192. Alizadeh BZ, Njajou OT, Bijkerk C, Meulenbelt I, Wildt SC, Hofman A, Pols HA, Slagboom PE, Duijn CM. Evidence for a role of the genomic region of the gene encoding for the alpha1 chain of type IX collagen (COL9A1) in hip osteoarthritis: A population-based study. Arthritis and Rhuem 2 (2005) 1437-1442.

193. Loughlin J, Mustafa Z, Dowling B, Southam L, Marcelline L, Räinä SS, Ala-Kokko L, Chapman K. *Finer linkage mapping of a primary hip osteoarthritis susceptibility locus on chromosome 6*. Eur J Hum Genet 10 (2002) 562-568.

195. Simonaro CM, Ge Y, Eliyahu E, He X, Jepsen KJ, Schuchman EH. Involvement of the Toll-like receptor 4 pathway and use of TNF-alpha antagonists for treatment of the mucopolysaccharidoses. PNAS 107 (2010) 222-227.

196. Eliyahu E, Wolfson T, Ge Y, Jepsen KJ, Schuchman EH, Simonaro CM. *Anti-TNF-alpha therapy enhances the effects of enzyme replacement therapy in rats with mucopolysaccharidosis type V.* PLoS One 6 (2011) [E pub].

Appendix 1 - R Code for Illumina Array Analysis

library(limma) library(gplots)

without any filtering etc, just naive data ### Work with P-values detection, delete >0.5

IamHere<-paste("") DataHere<-paste(IamHere,sep="") Samples<c("X6903016064_A.AVG_Signal","X6903016064_B.AVG_Signal","X6903016064_C.AVG_Si gnal","X6903016064_D.AVG_Signal","X6903016064_E.AVG_Signal","X6903016064_F.AVG _Signal") setwd(DataHere) ReadNames<-as.matrix(read.table("Samples_Probes.txt",nrow=1,header=T,fill=T,sep="\t"))

toReadPval<-which(colnames(ReadNames)%in%Samples)+1 toReadExp<-which(colnames(ReadNames)%in%Samples) toReadPr<-which(colnames(ReadNames)%in%"PROBE_ID") toReadSym<-which(colnames(ReadNames)%in%"SYMBOL") toRead<-c(toReadPr,toReadExp,toReadSym)

ReadTable<-read.columns("Samples_Probes.txt", required.col=toReadPval) dim(ReadTable) ##45283 x 6

d1<-which(ReadTable[,1]>0.05) d2<-which(ReadTable[,2]>0.05) d3<-which(ReadTable[,3]>0.05) d4<-which(ReadTable[,4]>0.05) d5<-which(ReadTable[,5]>0.05) d6<-which(ReadTable[,6]>0.05) del<-unique(c(d1,d2,d3,d4,d5,d6)) length(del) ##26927

ReadTable<-read.columns("Samples_Probes.txt", required.col=toRead) ReadTable<-ReadTable[-del,] dim(ReadTable)## 18356

M1="6903016064_A.AVG_Signal" M2a="6903016064_B.AVG_Signal" M2b="6903016064_C.AVG_Signal" WT1a="6903016064_D.AVG_Signal" WT1b="6903016064_E.AVG_Signal" WT2="6903016064_F.AVG_Signal"

nSamples<-

```
c("6903016064_A.AVG_Signal","6903016064_B.AVG_Signal","6903016064_C.AVG_Signal",
"6903016064_D.AVG_Signal","6903016064_E.AVG_Signal","6903016064_F.AVG_Signal")
```

```
Notations<-c("M1","M2a","M2b","WT1a","WT1b","WT2")
need=array()
```

```
for(i in 1:length(nSamples)){
```

```
need[i]<-which(colnames(ReadTable)==nSamples[i])
```

}

```
Data<-ReadTable[,need]
SYMBOL<-as.matrix(ReadTable[,which(colnames(ReadTable)=="SYMBOL")])
DataN<-cbind(Data,SYMBOL)
colnames(DataN)<-c(Notations,"SYMBOL")
```

```
rownames(Data) <- c(1:dim(Data)[1])
```

Data<-as.matrix(CutMatn)

ind<-which(pv<0.05) ## you can change this value, pv is a vector of p-values length(ind) ## N of samples

genes<-unique(as.matrix(DataN[ind,7]))
length(genes) ## N of genes
write(sort(genes),paste(IamHere,"ResultingGenes.txt",sep="")) ##</pre>

length(DataN[which(DataN[,7]%in%genes),7]) ## - all the possible copies

MatWithAllPossibleCopiesOfGoodGenes<-DataN[which(DataN[,7]%in%genes),] ## matrix of all samples which has the same name as our chosen differently regulated genes

pdf(paste(IamHere,"Result.pdf",sep="")) ##

MatrixToPlot<-as.matrix(DataN[which(DataN[,7]%in%genes),1:6]) ## matrix to plot RowNames<-as.matrix(DataN[which(DataN[,7]%in%genes),7]) ## names of the rows heatmap(MatrixToPlot,cexRow=0.3,labRow=RowNames)

dev.off() ##

ReadJon<-as.matrix(read.table(paste(DataHere,"Result.txt",sep=""))) dim(ReadJon) ##520

length(which(genes%in%ReadJon)) ## N of genes in new result which are in Illumina result

no<-unique(ReadJon[which(ReadJon%in%genes==F)]) ## genes from Illumina which are not in new result length(no) ##

ClustM<-as.matrix(DataN[ind,1:6]) rownames(ClustM)<-as.matrix(DataN[ind,7]) ## d<-dist(ClustM, method = "euclidean") ## you can change method to "euclidean", "maximum", "manhattan", "canberra", "binary" or "minkowski" hc<-hclust(d, method = "complete", members=NULL) ## you can change method to "ward", "single", "complete", "average", "mcquitty", "median" or "centroid". I am usually use average

pdf(paste(IamHere,"ResultCluster2.pdf",sep="")) ## plot(hc,cex=0.25,main="Corrected Genes",xlab="X parameters genes") ## cex means size of letters in the plot dev.off() ##

######## ResultingMatrix<-as.matrix(Data[ind,1:6])

```
ResultingMatrixGenes<-as.matrix(DataN[ind,7])
mean1=array()
mean2=array()
for(i in 1:dim(ResultingMatrix)[1]){
    mean1[i]<-mean(ResultingMatrix[i,1:3])
    mean2[i]<-mean(ResultingMatrix[i,4:6])
}
```

```
M.up_WT.down<-unique(ResultingMatrixGenes[which(mean1>=mean2)])
M.down_WT.up<-unique(ResultingMatrixGenes[which(mean1<mean2)])
```

```
M.up_WT.down[M.up_WT.down%in%M.down_WT.up]
```

```
write(sort(M.up_WT.down),paste(IamHere,"M_up.txt",sep="")) ##
write(sort(M.down_WT.up),paste(IamHere,"M_down.txt",sep="")) ##
```

```
fit <- lmFit(Data,design=design.matrix)
fit1 <- contrasts.fit(fit,contrasts=contrast.matrix)
fit1 <- eBayes(fit1)</pre>
```

th=0.05 ## threshold

Fit1<-topTable(fit1, coef=1,number=dim(fit1)[1]) length(which(Fit1[,which(colnames(Fit1)=="adj.P.Val")]<th)) #

```
pdf(paste(IamHere, "C_I.pdf"))
p.v.a<-Fit1[,6]
plot(p.v.a, cex=0.2,xlab="Genes",ylab="p-values")
lines(c(0,(length(p.v.a)+1)),c(0.05,0.05),lty="dashed")
```

dev.off()

Fit_cut1<-Fit1[which(Fit1[,which(colnames(Fit1)=="adj.P.Val")]<th),]

```
i1<-which(DataN[,7]%in%"Ehd1") ## <------
Fit1[which(Fit1[,1]%in%i1),] ## <------
```

dim(Fit_cut1) ##
indTOtake<-Fit_cut1[,1]</pre>

MatCut<-DataN[as.numeric(indTOtake),] ## Good matrix dim(MatCut) ## genes<-unique(as.matrix(MatCut[,7])) length(genes) ##

p.val.ad<-which(colnames(Fit1)=="adj.P.Val")

means.Mutant<-rowMeans(resMatCut[,1:3]) means.WT<-rowMeans(resMatCut[,4:6]) adj.p.val<-Fit_cut1[,p.val.ad] Ratio<-means.Mutant/means.WT

resMatCut <-cbind(MatCut,adj.p.val,means.Mutant,means.WT,Ratio)

```
write(sort(genes),paste(IamHere,"result2_removeP-values.txt",sep=""))
```

pdf(paste(IamHere,"Result_new.pdf",sep="")) ##

MatrixToPlot<-as.matrix(MatCut[which(MatCut[,7]%in%genes),1:6]) ## matrix to plot RowNames<-as.matrix(MatCut[which(MatCut[,7]%in%genes),7]) ## names of the rows heatmap(MatrixToPlot,Rowv=NA,Colv=NA, cexRow=0.3,labRow=RowNames)

dev.off() ## use this always after pdf() function

write.table(resMatCut,paste(IamHere,"Expressionts_Genes_pval Table.txt",sep=""),row.names=F)

JonathanList<-as.matrix(read.table(paste(IamHere,"JonathanList.txt",sep=""),header=T)) write.table(List[which(List%in%as.matrix(MatCut)[,7])],paste(IamHere,"OverlappingGenes.txt", sep=""),row.names=F,col.names=F) ### clustering

ClustNew<-as.matrix(MatCut[,1:6]) rownames(ClustNew)<-as.matrix(MatCut[,7]) d<-dist(ClustNew, method = "euclidean") hc<-hclust(d, method = "complete", members=NULL) plot(hc)

pdf(paste(IamHere,"ResultCluster22.pdf",sep="")) ## plot(hc,cex=0.2,main="Bayes Corrected Genes",xlab="X parameters genes") ## dev.off() ##

up-down regulated

```
ResultingMatrix.2<-as.matrix(MatCut[,1:6])
ResultingMatrixGenes.2<-as.matrix(MatCut[,7])
mean1.2=array()
for(i in 1:dim(ResultingMatrix.2)[1]){
    mean1.2[i]<-mean(ResultingMatrix.2[i,1:3])
    mean2.2[i]<-mean(ResultingMatrix.2[i,4:6])
}
```

M.up_WT.down.2<-unique(ResultingMatrixGenes.2[which(mean1.2>=mean2.2)]) M.down WT.up.2<-unique(ResultingMatrixGenes.2[which(mean1.2<mean2.2)])

M.up_WT.down.2 ## mutant up - WT down M.down_WT.up.2 ## WT up , mutant down

write(sort(M.up_WT.down.2),paste(IamHere,"M_up2.txt",sep="")) ##
write(sort(M.down_WT.up.2),paste(IamHere,"M_down2.txt",sep="")) ##