Applying Next Generation Sequencing to Skeletal Development and Disease

Citation
Bowen, Margot Elizabeth. 2014. Applying Next Generation Sequencing to Skeletal Development and Disease. Doctoral dissertation, Harvard University.

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:11745698

Terms of Use

Share Your Story

Applying Next Generation Sequencing to
Skeletal Development and Disease

A dissertation presented

by

Margot Elizabeth Bowen

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Genetics

Harvard University

Cambridge, Massachusetts

November 2013

Dissertation Advisor: Professor Matthew L. Warman   Margot Elizabeth Bowen

Applying Next Generation Sequencing to Skeletal Development and Disease

ABSTRACT

Next Generation Sequencing (NGS) technologies have dramatically increased the throughput and lowered the cost of DNA sequencing. In this thesis, I apply these technologies to unresolved questions in skeletal development and disease. Firstly, I use targeted re-sequencing of genomic DNA to identify the genetic cause of the cartilage tumor syndrome, metachondromatosis (MC). I show that the majority of MC patients carry heterozygous loss-of-function mutations in the *PTPN11* gene, which encodes a phosphatase, SHP2, involved in many signaling pathways. Furthermore, I show that cartilage lesions in MC patients likely arise following somatic second-hit mutations in *PTPN11*. Secondly, I use RNA-seq to identify gene expression changes that occur following genetic inactivation of *Ptpn11* in mouse chondrocyte cultures. I show that chondrocytes lacking *Ptpn11* fail to properly undergo terminal differentiation and instead continue to express genes associated with earlier stages of chondrocyte maturation. I validate these findings *in vivo* by examining markers of specific chondrocyte maturation stages in the vertebral growth plates of mice following postnatal mosaic inactivation of *Ptpn11*. Together, my results provide insight into the molecular mechanisms underlying the initiation and growth of cartilage tumors. In the third component of my thesis, I develop a method to map and clone zebrafish mutations by performing whole genome sequencing on pooled DNA. I apply this method to zebrafish mutants identified in a mutagenesis screen for adult phenotypes, including skeletal phenotypes, and determine that a nonsense mutation in *bmp1a* underlies the craniofacial phenotype in the *wdd* mutant. In summary, I show that NGS technologies can be successfully utilized to firstly identify the genetic cause of a human skeletal disorder, secondly investigate the molecular mechanisms regulating the maturation of skeletal cells, and thirdly expedite the process of mapping and cloning zebrafish mutants with skeletal phenotypes. Altogether, my research provides insight into the pathways and processes regulating skeletal development and disease.

Table of Contents

Acknowledgements

First and foremost I would like to express my appreciation for my dissertation advisor, Dr. Matthew Warman. His constant optimism and enthusiasm for science and his thoughtful guidance allowed me to grow as a research scientist. I would like to thank Dr. Kyle Kurek for being a supportive and encouraging mentor during my first few years in the lab. I am grateful towards Dr. Matthew Harris, Dr. Katrin Henke and Dr. Eric Boyden for their discussions and guidance during our collaborations, as well as to all members of the Warman lab for their assistance and friendship over the years. I would like to thank the members of my dissertation advisory committee, Dr. Andrew Lassar, Dr. Vicki Rosen and Dr. Karen Cichowski, for their guidance and support. Finally, I would like to thank my friends, family and boyfriend for their continued support during my time at graduate school.

Chapter **1**

General Introduction

1.1   Skeletal development

The vertebrate skeleton

The vertebrate skeleton is a complex organ system that provides structural support, allows for controlled movements, protects internal organs, stores minerals, and houses hematopoiesis [1]. Disrupting the normal process of skeletal development and homeostasis can adversely affect survival and quality of life. This is reflected by the large number of inherited disorders of the skeleton [2], as well as age related disorders such as osteoarthritis, a degenerative disease of the joints [3], and osteoporosis, a weakening of the bones [4]. Two main connective tissues that comprise the vertebrate skeleton are cartilage and bone [1]. Cartilage is an avascular tissue that contains chondrocytes embedded in a stiff yet flexible extracellular matrix. Bone is a rigid tissue that contains osteoblasts, which lay down a mineralized matrix, and osteoclasts, that degrade matrix to facilitate tissue remodeling. In contrast, invertebrates do not have an internal skeleton made of bone or cartilage. Instead, invertebrates can have fluid-filled hydrostatic skeletons (eg. jelly fish), or endoskeletons or exoskeletons made of the carbohydrate chitin (eg. insects), calcium carbonate (eg. mollusks), or silica (eg. sponges) [5].

Endochondral and intramembranous ossification

The development of the vertebrate skeleton is an elaborate and complex task, with over 200 bones that must each attain their specific size, shape, location and tissue composition [1]. Skeletogenesis is initiated when cells from two of the three primary germ layers in the early embryo, the mesoderm and ectoderm, become multipotent mesenchymal cells and migrate to sites of skeletogenesis throughout the body. These cells are the source of chondrocytes and osteoblasts. Most bones in the skeleton, including the long, short and irregular bones, form through a process called endochondral ossification [1]. In this process, the primary skeleton forms entirely from cartilage, which is capable of growing rapidly as the embryo develops, and is gradually replaced by bone (Figure 1.1) [6]. In contrast, flat bones form by intramembranous ossification, whereby mesenchymal cells differentiate directly into osteoblasts, without first forming a cartilage template [1].

The growth plate

Bones that form through the process of endochondral ossification contain cartilaginous growth
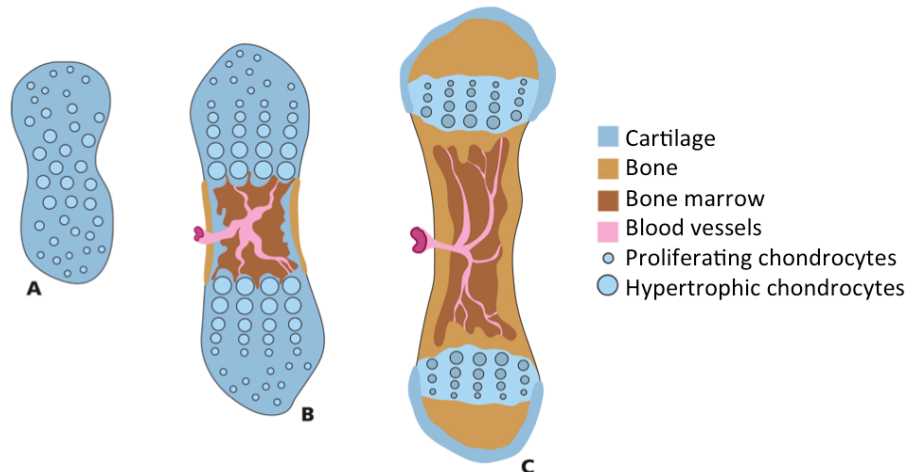
Figure 1.1. Formation of bones through endochondral ossification. A: Mesenchymal cells condense and differentiate into chondrocytes to form a cartilaginous model of the bone. Chondrocytes in the center of the model undergo hypertrophy and calcify the matrix. **B:** The primary ossification center forms in the center of the cartilaginous model when hypertrophic chondrocytes undergo apoptosis and promote vascular invasion. Osteoblasts are recruited to lay down the new bone matrix. **C:** Secondary ossification centers form at the ends of the cartilaginous models. The only cartilage that remains postnatally is articular cartilage, which covers the ends of bones, and the growth plate cartilage, which allows for the longitudinal growth of bones. Figure adapted from Teixeira *et al* (2008).

plates that allow for longitudinal bone growth. Growth plates consist of chondrocytes arranged in distinct zones of maturation (Figure 1.2) [7,8]. The resting zone contains scattered chondrocytes in a quiescent state, which serve as a source of cells for the second layer, the proliferative zone. Chondrocytes in the proliferative zone divide rapidly and arrange themselves into columns, after which they exit the cell cycle to become pre-hypertrophic chondrocytes. A dramatic increase in cell size takes place as pre-hypertrophic chondrocytes enter the early-hypertrophic zone. The final zone is the terminal or late-hypertrophic zone in which chondrocytes secrete factors for matrix degradation, calcification and vascular recruitment, and ultimately undergo programmed cell death. Invading blood vessels bring in osteoclasts / chondroclasts and osteoblasts, which progressively erode and replace the remaining calcified cartilage matrix with a new bone matrix. This cycle of chondrocyte proliferation, maturation and replacement by bone is what drives longitudinal bone growth.

The perichondrium

The growth plate is surrounded by the perichondrium, which consists of an outer fibroblast layer, and an inner layer that contains osteoprogenitor and osteo-chondroprogenitor cells. The osteoprogenitor cells give rise to osteoblasts that form the bone collar, and can migrate into the bone cavity to contribute
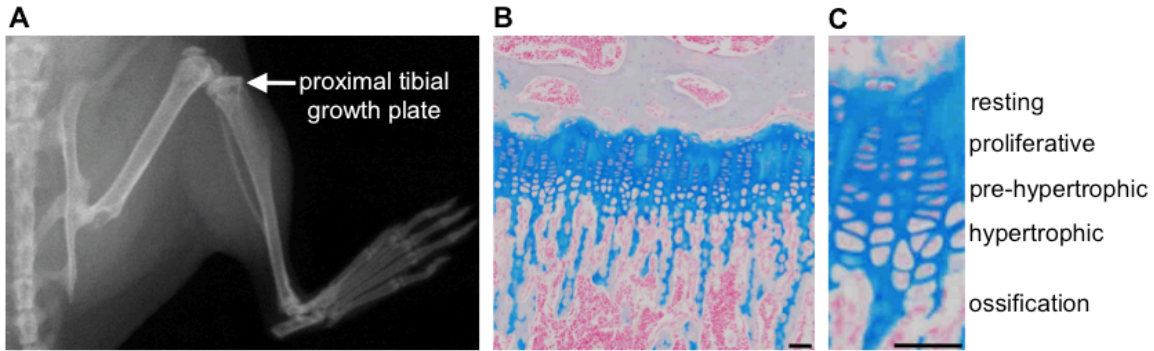
Figure 1.2. Growth plates contain chondrocytes that proliferate, increase in size, and subsequently get degraded and replaced by bone. A: Radiograph of a mouse hind limb. The location of the proximal tibial growth plate is indicated. **B,C**: Histological images of the proximal tibial growth plate of a 7 week-old mouse. The stages of chondrocyte maturation are indicated. Tissue sections are stained with Alcian Blue and Nuclear Fast Red. Scale bars represent 50 microns.

to the pool of intramedullary osteoblasts [10]. The osteo-chondroprogenitor cells can migrate to the site of injury and differentiate into chondrocytes or osteoblasts during fracture repair [11].

## 1.2   Signaling pathways regulating the growth plate

The rate of chondrocyte proliferation, maturation and terminal differentiation is tightly regulated by a number of paracrine and endocrine hormones and growth factors, to ensure that each skeletal element reaches its appropriate size. Several of the key signaling pathways that have been studied in the growth plate are outlined here.

IHH/PTHrP negative feedback loop regulates the length of proliferating columns

Indian hedgehog (IHH) is a morphogen secreted by pre-hypertrophic chondrocytes [12,13]. IHH induces the production of parathyroid hormone related protein (PTHrP) in periarticular chondrocytes. PTHrP signals through its receptor PTHR1 to proliferative and pre-hypertrophic chondrocytes to delay their maturation. This well-characterized feedback loop regulates the rate at which chondrocytes exit the proliferative zone (Figure 1.3) [12,13]. IHH also acts independently of PTHrP to promote the proliferation of chondrocytes by inducing expression of the cell cycle regulator, cyclin D1 [14]. IHH also promotes the differentiation of osteoblasts in the perichondrium to form the bone collar [15].

BMP signaling promotes chondrocyte proliferation and maturation

Bone morphogenetic proteins (BMPs), acting via type I and type II serine/threonine kinase receptors, induce phosphorylation of the receptor-regulated SMADs (R-SMAD1, R-SMAD5 and R-
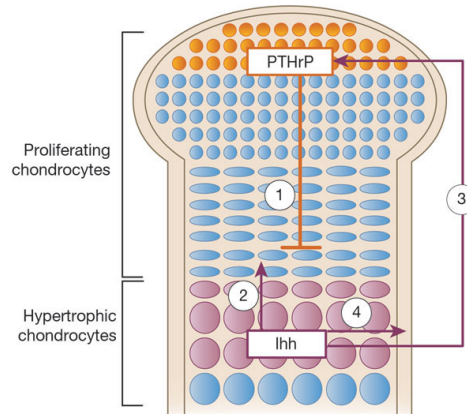
Figure 1.3: IHH/PTHrP negative feedback loop regulates the rate of chondrocyte maturation. Periarticular chondrocytes secrete the PTHrP ligand, which signals to proliferating chondrocytes to delay their maturation (1). Chondrocytes at the lower end of the proliferating zone that are sufficiently far from the PTHrP source, start to secrete IHH and undergo maturation. IHH promotes chondrocyte proliferation (2), promotes PTHrP production in the peri-articular region (3), and promotes the formation of the bone collar (4). Figure adapted from Kronenberg *et al* (2003).

SMAD8), which bind SMAD4 and enter the nucleus to regulate gene expression. Inactivation of *Bmpr1a* and *Bmpr1b*, which encode BMP receptors, indicates that BMP signaling promotes chondrocyte proliferation, promotes the onset of hypertrophy, and is required for the columnar organization of proliferative chondrocytes [16]. Conflicting reports suggest the BMP signaling either a promotes or inhibits the terminal differentiation of late-hypertrophic chondrocytes [16,17].

FGF signaling antagonizes IHH and BMP signaling and inhibits chondrocyte proliferation and maturation

There are four fibroblast growth factor (FGF) receptors (FGFR1-4) and 22 FGF ligands in mammals. The ligands FGF9 and FGF18 are expressed in the perichondrium, and during early stages of chondrogenesis, they activate FGFR3 to promote chondrocyte proliferation, and activate FGFR1 to promote the initiation of hypertrophy [18–20]. In contrast, at late embryonic and postnatal stages FGF18 acts through FGFR3 to inhibit chondrocyte proliferation and maturation, and inhibit *Ihh* and *Bmp4* expression (Figure 1.4) [21–24].

IGF1 promotes chondrocyte proliferation and maturation

Growth Hormone (GH) is produced in the pituitary gland and stimulates the secretion of insulin-like growth factor 1 (IGF1), both in the liver and in growth plate chondrocytes themselves. IGF1 signals
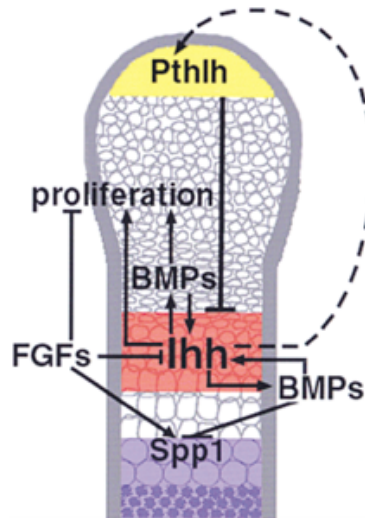
Figure 1.4: Interactions between BMP, FGF and IHH signaling in the growth plate. A positive feedback loop exists between BMP and IHH signaling. In contrast, FGF signaling antagonizes BMP and IHH signaling. Figure adapted from Minina *et al* (2002).

via IGF1R, which is expressed in proliferating and prehypertrophic chondrocytes [25]. Loss of *Igf1* or *Igf1r* in mice indicates that IGF1/IGF1R signaling promotes chondrocyte proliferation and maturation [26,27].

Signaling through EGFR promotes the removal of hypertrophic chondrocytes

The epidermal growth factor (EGF) receptor, EGFR, is a member of the ErbB family of receptor tyrosine kinases, and can be activated by several growth factors, including EGF and TGFalpha. Inactivation of *Egfr* or *Tgfalpha*, indicates that signaling through TGFalpha and EGFR positively regulates the removal of hypertrophic chondrocytes, by regulating matrix degradation and osteoclast recruitment [28–30]. While inactivation of *Egfr* alone does not affect chondrocyte proliferation [31], expression of an inhibitor of all ErbB receptors inhibits chondrocyte proliferation [32].

Canonical Wnt signaling inhibits chondrogenesis but promotes chondrocyte proliferation and maturation

The canonical Wnt pathway inhibits ectopic cartilage formation in the perichondrium, and promotes chondrocyte proliferation and maturation. This has been identified by the conditional inactivation of beta-catenin, the mediator of canonical Wnt signaling. Inactivation of beta-catenin in mesenchymal progenitor cells blocks osteoblast differentiation, and causes osteoblast precusor cells in the perichondrium to differentiate into chondrocytes instead of osteoblasts [33,34]. Activation of beta-catenin in chondrocytes accelerates the onset of hypertrophy while inactivation of beta-catenin in

chondrocytes delays hypertrophy, and delays the final maturation and removal of hypertrophic chondrocytes [35,36]. Inactivation of beta-catenin also leads to decreased chondrocyte proliferation [37].

1.3   Metachondromatosis and related cartilage tumor syndromes

### Benign cartilage tumors: enchondromas and exostoses

The two most common benign cartilaginous tumors are exostoses/osteochondromas, which form on the surface of the bone, and enchondromas, which form within bones (Figure 1.5) [38]. These tumors normally arise during childhood near the growth plate of bones formed by endochondral ossification, and, although they are benign tumors, malignant transformation to chondrosarcoma can occur [39]. Most often, exostoses and enchondromas occur as single sporadic lesions in an individual, but less frequently, multiple lesions can occur in an individual as part of a syndrome. Multiple enchondromas are present in individuals affected with an enchondromatosis syndrome, the most common of which are Ollier disease and Maffucci syndrome [40], while multiple exostoses can occur in patients with the autosomal dominant disorder multiple hereditary exostoses (MHE) (also known as multiple osteochondromas (MO)) [41]. A third type of cartilage tumor syndrome, metachondromatosis (MC), which is the focus of this thesis, is unique in that it is associated with both exostoses and enchondromas.
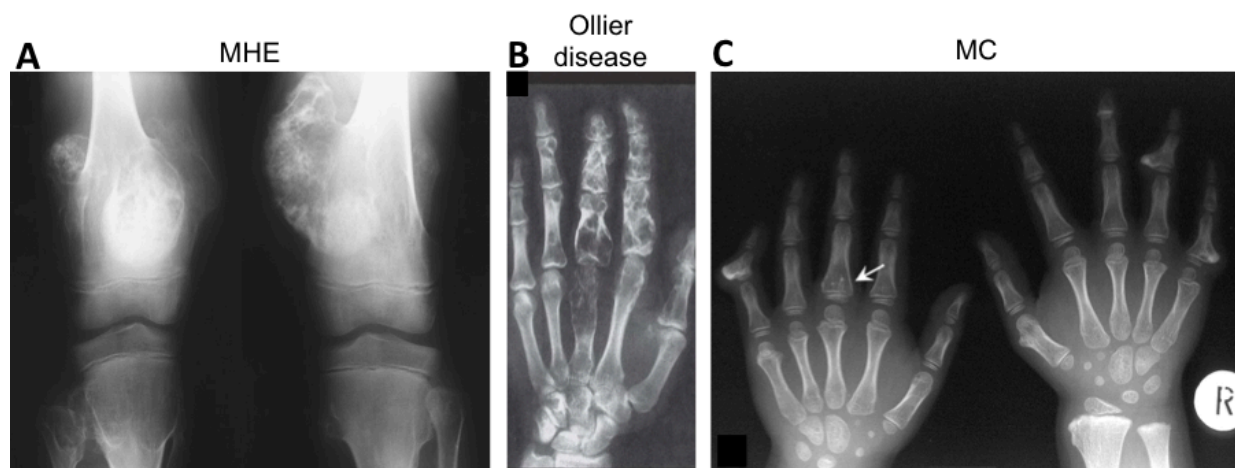
Figure 1.5. Radiographs of exostoses and/or enchondromas in patients with different cartilage tumor syndromes. A: Radiograph showing exostoses on the femurs and tibiae of an individual with MHE. Image adapted from Pannier *et al* (2008). **B:** Radiograph showing enchondromas in the hand of an individual with Ollier disease. Image adapted from Horvai *et al* (2006). **C:** Radiograph showing exostoses and an enchondroma (arrow) in the hands on an individual with MC. Image adapted from Bovee *et al* (2006).

Genetic basis of the enchondromatosis syndromes, Ollier disease and Maffucci syndrome

It has recently been reported that somatic mosaicism for a missense mutation in *Isocitrate Dehydrogenase 1* (*IDH1*) or *IDH2* underlies Ollier disease and Maffucci syndrome [42–44]. The phenotypic differences between Ollier disease and Maffucci syndrome are postulated to be due differences in the tissue distribution and/or proportion of cells that carry a somatic *IDH1/IDH2* mutation, and the lack of genetic transmission of these disorders suggests that heterozygosity for an *IDH1/IDH2* mutation in the germline or embryo is lethal [45]. Missense mutations in *IDH1* and *IDH2* have also been detected in other neoplasms, and lead to neomorphic enzyme activity, resulting in 2-hydroxyglutarate production, which induces global DNA hypermethylation [46]. Indeed, an increase in DNA methylation was observed in enchondromas carrying *IDH1* mutations, but it is not yet known whether or how altered methylation contributes to enchondroma formation [44]. Missense mutations in *PTHR1* have also been found in a small percentage of patients with Ollier disease [47]. One such mutation was shown to enhance hedgehog signaling, and mice over-expressing a mutant *PTHR1* allele, or the Hedgehog transcriptional regulator, *Gli2*, in chondrocytes, have impaired chondrocyte maturation and develop enchondroma-like lesions [47] (Figure 1.6). This supports the hypothesis that enchondromas can arise from growth plate chondrocytes that fail to undergo terminal differentiation and be replaced by bone. An alternative hypothesis for enchondroma formation is that they arise from mesenchymal precursor cells in the bone cavity or perichondrium/periostium that inappropriately differentiate into chondrocytes [48]. Further studies will be needed to determine how mutations in *IDH1* and *IDH2* affect chondrocyte differentiation or maturation and lead to enchondroma formation.
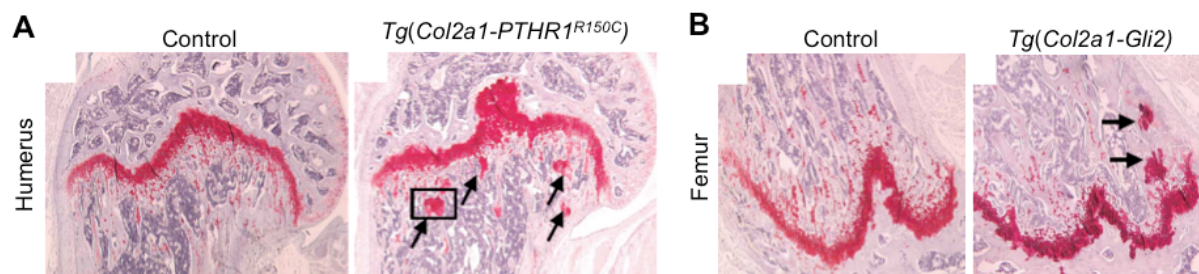

Figure 1.6. Disrupting the IHH/PTHrP feedback loop can lead to enchondroma formation in mice. Enchondromas arise in mice expressing a mutant *PTHR1* transgene **(A)** or over-expressing the Hedgehog-activated transcription factor, *Gli2* **(B).** Images adapted from Hopyan et al (2002).

Genetic basis of the exostosis syndrome, MHE

Inactivating mutations in *Exostosin*-1 or -2 (*EXT1* or *EXT2*) cause MHE [41]. While patients with MHE are heterozygous for a mutation in *EXT1* or *EXT2*, their exostoses are believed to arise from cells in which a somatic second hit mutation has inactivated the remaining wild-type *EXT1* or *EXT2* allele [49]. The EXT1/EXT2 protein complex catalyzes the polymerization of sugars to form heparan sulfate proteoglycans (HSPGs), which act as receptors or co-receptors in signaling pathways and interact with a variety of morphogens, growth factors and extracellular matrix components [50–52]. Studies in mice suggest that a deficiency for heparan sulfate in either chondrocytes or perichondrial cells in the growth plate results in ectopic BMP signaling in the perichondrium, which induces perichondrial cells to inappropriately differentiate into cartilage, resulting in exostosis formation [53–55] (Figure 1.7).

Unique features of MC: a combined exostosis and enchondroma syndrome

Patients with the rare (< 50 cases reported in the literature) autosomal dominant genetic disorder, metachondromatosis (MC), develop both enchondromas and exostoses [56–60] (Table 1.1). The enchondromas associated with MC most frequently occur in the pelvis and long bones of the lower extremities, in contrast to the enchondromas associated with Ollier disease or Maffucci syndrome, which occur more frequently in the hands and feet [40]. Only one case of malignant transformation has been reported for a MC enchondroma [61]. The exostoses associated with MC normally occur on the bones of
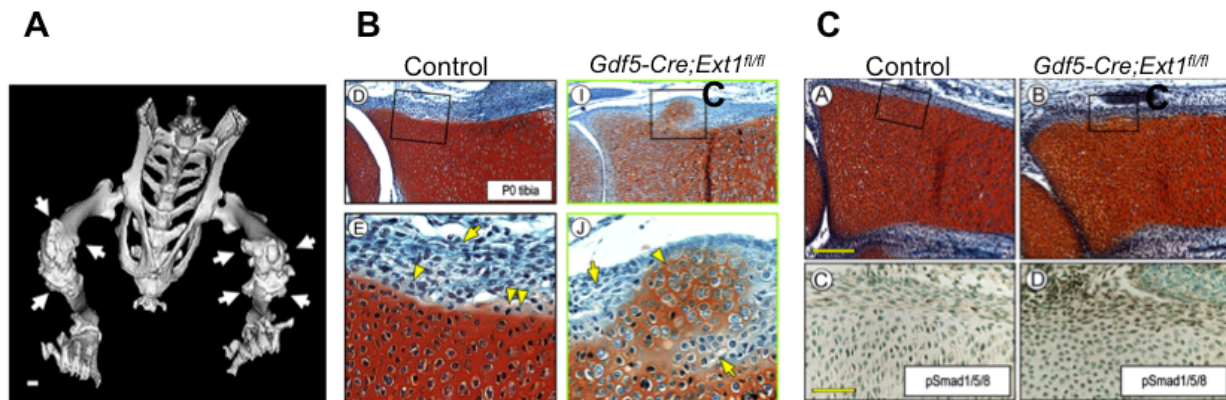
Figure 1.7. Conditional inactivation of *Ext1* in mice leads to exostosis formation and is associated with ectopic BMP signaling and chondrogenesis in the perichondrium. A: 3D microCT image of exostoses that arise in mice following postnatal mosaic inactivation of *Ext1* in chondrocytes. Image adapted from Jones *et al* (2010). **B-C:** Histological images of mice in which *Ext1* was inactivated in joint progenitor cells and the perichondrium. Note the formation of ectopic cartilage in the perichondrium at P0 (B) and ectopic BMP signaling, as assessed by phospho-smad1/5/8 immunostaining, in the perichondrium at E15.5 (C). Image adapted from Huegel *et al* (2013).

Table 1.1: Comparison of metachondromatosis and related cartilage tumor syndromes

	MHE	MC	Ollier disease	Maffucci syndrome
Tumor types	Exostoses	Exostoses & enchondromas	Enchondromas	Enchondromas & vascular lesions
Incidence	1/50,000	~50 case reports	1/100,000	~200 case reports
Most common location of tumors:				
Exostoses	Long bones of the limbs	Hands & feet	n/a	
Enchondromas	n/a	Pelvis & long bones	Hands & feet	
Inheritance	Autosomal dominant	Autosomal dominant	Sporadic	
Affected gene(s)	*EXT1* or *EXT2*	*PTPN11*	*IDH1* or *IDH2*	

the hands and feet, point towards the affected joint, and may regress spontaneously, while exostoses associated with MHE normally occur on long bones, point away from the affected joint, and only rarely regress [62,63]. The phenotypic differences between MC and MHE, Ollier disease and Maffucci syndrome suggests that MC is associated with a unique underling genetic defect; indeed, mutations in *EXT1/EXT2* and *IDH1/IDH2* were not detected in MC patients [44,57]. As part of this thesis, I identify mutations in the *PTPN11* gene in MC patients (Chapter 2). Thus, MC does appear to have a unique genetic cause; however, the underlying molecular pathways and developmental processes involved in tumor formation in MC patients may be related to those involved in tumor formation in other cartilage tumor syndromes. Therefore, the research on *PTPN11* presented in this thesis (Chapter 2 and 3) is likely to not only further our understanding of MC, but to also further our understanding of the mechanisms underlying all types of benign cartilage tumors.

1.4   Role of *PTPN11*/SHP2 in cartilage tumor formation

Inactivation of *Ptpn11* in mice causes skeletal defects and cartilage tumors

In this thesis, I show that loss-of-function mutations in *PTPN11* underlie the human cartilage tumor syndrome, metachondromatosis (Chapter 2). Concurrent with my work, other investigators

have shown that cartilage tumors can be induced in mice by conditionally inactivating *Ptpn11* (Figure 1.8; Table 1.2). A requirement for *Ptpn11* during mouse skeletal development was first revealed after studying mice in which *Ptpn11* had been inactivated postnatally in all cell types [64]. These mice developed scoliosis, and were found to have an increase in bone mineral content, an increase in trabecular bone, disorganized cortical bone, impaired osteoclastogenesis, and disorganized and expanded growth plates. Furthermore, these mice formed ectopic cartilage growths on the surface of long bones, resembling exostoses, and ectopic cartilage islands in the vertebrae, resembling enchondromas. To determine the cell type that contributed to these skeletal defects, *Ptpn11* was inactivated postnatally using the *Col2a1-CreER* driver, which is expressed in chondrocytes and some perichondrial cells [65,66]. These mice developed kyphosis and scoliosis, as well as exostoses on the bones of the hands, feet, wrists and ankles, enchondromas in the ribs and vertebrae, and osteophyte-like outgrowths on the vertebrae [65,66].
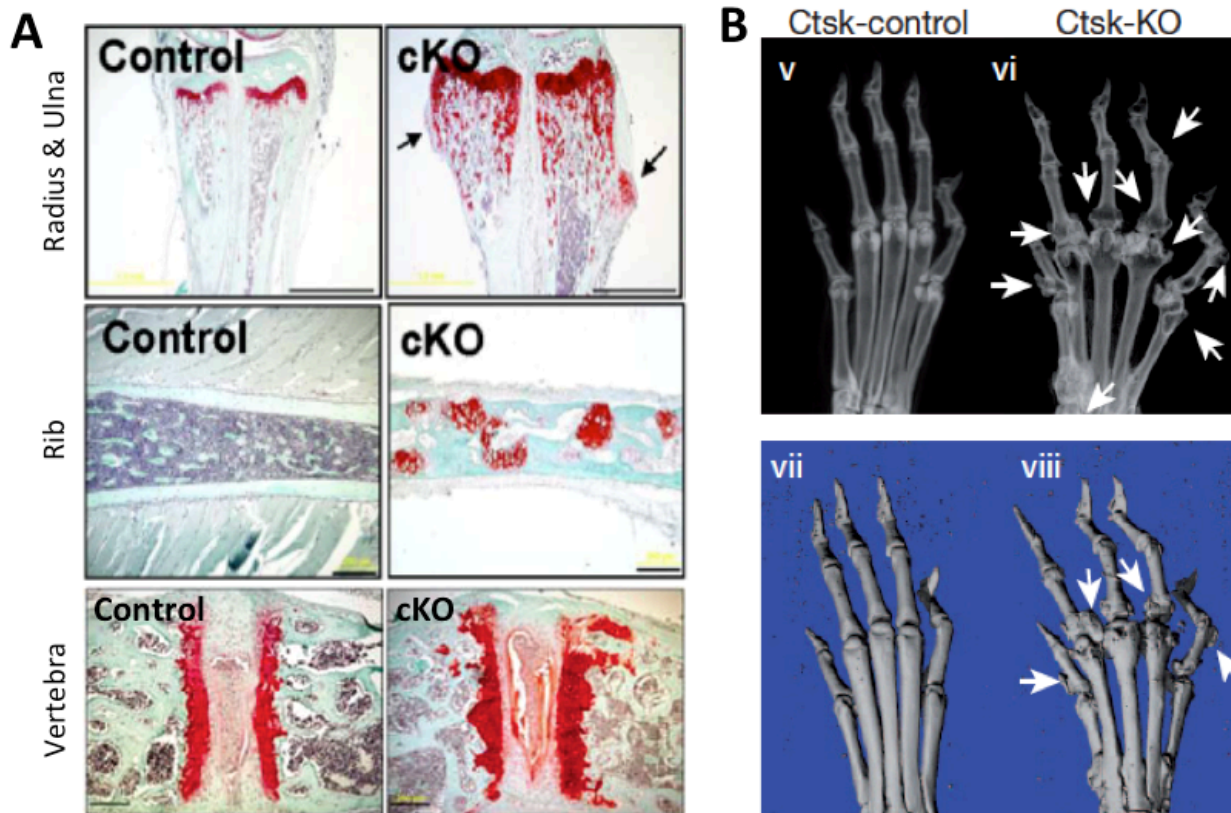
Figure 1.8. Exostoses and enchondromas arise following conditional inactivation of *Ptpn11* in mice. **A**: Histological images of exostoses in the radius and ulna, and enchondromas in the ribs and vertebrae, following postnatal inactivation of *Ptpn11* using the *Col2a1-CreERT2* driver. Images adapted from Kim *et al* (2013) **B:** Radiographs and microCT images of exostoses arising in the hands following inactivation of *Ptpn11* using the *Ctsk-Cre* driver. Images adapted from Yang *et al* (2013).

Table 1.2: Comparison of published mouse models of metachondromatosis

	Bauler et al. (2011)	Kim et al. (2013)	Yang et al. (2013)
Genetic inactivation of *Ptpn11*			
Cell type	All cell types	Chondrocytes & perichondrium	Osteoclasts & perichondrium
Time of tamoxifen administration	Postnatal	Postnatal	–
Cre or *CreERT2* driver	*Ubc-CreERT2*	*Col2a1-CreERT2*	*Ctsk-Cre*
Exostoses*			
Knee	Y	n/d	Y
Ankles & wrists	n/d	Y	Y
Hands & feet	n/d	Y	Y
Enchondromas*			
Ribs	n/d	Y	n/d
Vertebrae	Y	Y	n/d
Long bones	n/d	n/d	Y
Other skeletal phenotypes*			
Disorganized growth plates	Y	Y	n/d
Scoliosis & kyphosis	Y	Y	Y
Increased bone mineral content	Y	n/d	Y

n/d = not determined

It was postulated that the exostoses arose from perichondrial cells, while the rib enchondromas were remnants of growth plate cartilage [65]. A perichondrial-origin for exostoses was supported by the observation that exostoses arose when *Ptpn11* was inactivated using the *Ctsk-Cre* driver, which is expressed in perichondrial cells in addition to its expression in osteoclasts [67]. Together, these studies suggest that *Ptpn11* is essential for normal skeletal development and for suppressing the formation of benign cartilage tumors.

PTPN11/SHP2 regulates many signaling pathways and developmental processes

PTPN11 is ubiquitously expressed and encodes SHP2, a nonreceptor protein tyrosine phosphatase (PTP) [68]. In addition to its role in skeletal development, SHP2 has been shown to play a

role in many different tissue types. Complete loss of *Ptpn11* in mice leads to peri-implantation lethality [69], and tissue specific inactivation of *Ptpn11* demonstrates that SHP2 is required for normal development of many tissues, including the heart, pancreas, liver, mammary gland, T-cells, CNS, and neural-crest derived tissues [68]. SHP2 contains two tandem Src homology 2 (SH2) domains, a PTP domain and a C-terminal tail with tyrosyl phosphorylation sites. In response to extracellular stimuli, the SH2 domains of SHP2 bind to phosphotyrosine residues on cytokine receptors, receptor tyrosine kinases (RTKs) and/or docking proteins that bind to RTKs [68]. Although the physiological substrates of SHP2 are not clear, SHP2 has been shown to regulate a variety of signaling pathways [68]. SHP2 is required for full activation of the Ras/ERK cascade, although the exact mechanism is not clear [70]. SHP2 has also been implicated in regulating signaling through PI3K/Akt [71], Wnt/β-catenin [72], Src family kinases [73], JAK/Stat [74], RhoA [75], JNK MAPK [76], p38 MAPK [77], NFAT [78], NF-kB [79], FAK [80]. Studies indicate that SHP2 regulates different signaling pathways in different cell types and in response to different stimuli [68]. It is not yet known which of these signaling pathways are regulated by SHP2 in chondrocytes. Thus, one aim of this thesis was to investigate the molecular consequences of SHP2 depletion specifically in chondrocytes, and to determine which SHP2-dependent signaling pathways may contribute to cartilage tumor formation (Chapter 3).

Mutations in *PTPN11* cause skeletal defects associated with Noonan Syndrome and LEOPARD syndrome

A key role for *PTPN11* in human skeletal development first became apparent when mutations in *PTPN11* were identified in patients with Noonan Syndrome (50%) and LEOPARD syndrome (80%) [81]. These syndromes belong to a group called the RASopathies, which are associated with a variety of development defects, including skeletal defects such as short stature, osteopenia, osteoporosis, scoliosis, kyphosis, hand and anterior chest wall anomalies, and increased fracture rates [82]. The *PTPN11* mutations associated with Noonan syndrome increase the phosphatase activity and alter the binding specificity of SHP2 [83], while mutations associated with LEOPARD syndrome impair SHP2 phosphatase activity but enhance interactions with some binding partners [84]. Mutations in components of the Ras/MAPK pathway (*SOS1*, *RAF1*, *KRAS, SHOC2)* have also been found in patients with Noonan Syndrome, and mutations in *RAF1* have been found in LEOPARD syndrome patients, suggesting that

altered Ras/MAPK signaling underlies these disorders [81]. Indeed, enhanced Ras/ERK signaling has been shown to contribute to Noonan Syndrome [84]. On the other hand, enhanced AKT/mTOR signaling plays a larger role in LEOPARD Syndrome [84]. Based on the role of SHP2 in many signaling pathways, it is likely that the dysregulation of multiple pathways contributes to these disorders. Thus, the research on *PTPN11* presented in this thesis (Chapter 2 and 3) may also further our understanding of the skeletal defects associated with the RASopathies.

1.6   Next Generation Sequencing Technologies

### First generation DNA sequencing: the Sanger sequencing method

Following the discovery of the double helix in 1953, the first DNA sequencing methods emerged in the late 1960s and 1970s [85]. In 1977, Fred Sanger published his "dideoxy method" for DNA sequencing. This method, also referred to as Sanger sequencing, became the gold standard method for DNA sequencing, and was the basis for the development of the first automatic sequencing machine by Applied Biosystems in 1986. Further enhancements of this technology over the years enabled the eventual completion of the first draft of the human genome assembly in 2001. In the Sanger method, a DNA synthesis reaction using DNA polymerase is performed using both deoxynucleotide triphosphates (dNTPs), which are the building blocks of DNA, and fluorescently labeled dideoxynucleotide triphosphates (ddNTPs), which lack a 3' OH group. When a dNTP is incorporated into the growing DNA strand, DNA synthesis can continue; however, when a ddNTP is incorporated, DNA synthesis terminates. The DNA sequence can be deduced by separating the resulting DNA fragments using electrophoresis and detecting their fluorescent label. Automated sequencing machines using the Sanger method are generally capable of running 96 reactions in parallel, with each reaction producing between 500 and 1,000 bases of highly accurate sequence. [85]

Next Generation Sequencing

While Sanger sequencing is accurate, it is also low-throughput and labor intensive. Over the past decade, new sequencing technologies have emerged that have significantly higher throughput than Sanger sequencing. These methods are referred to as massively parallel sequencing or Next Generation Sequencing (NGS) [86]. While these methods are currently less accurate and have shorter read lengths

than Sanger sequencing, the combined data that is assembled from many individual reads is highly accurate due to the high sequence coverage obtainable. The first commercially available NGS platform was introduced in 2005 by 454 Life Sciences, which was followed by several additional platforms in subsequent years, such as the first Solexa sequencer, the Genome Analyzer, in 2006 [86]. Further improvements to these NGS technologies over the years has led to a dramatic reduction in the cost of sequencing DNA (Figure 1.9) [87].
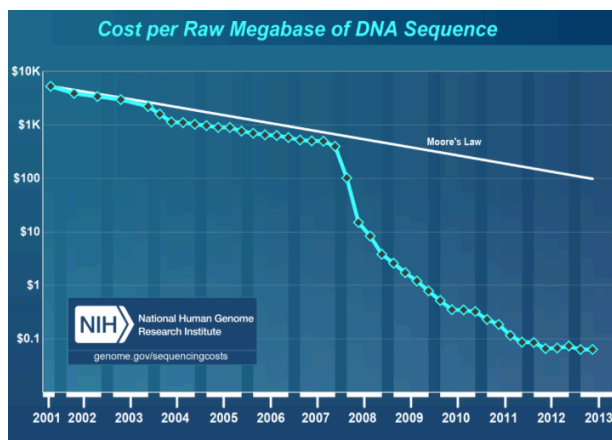
Figure 1.9. The cost of DNA sequencing has substantially decreased since the introduction of NGS. Graph showing the cost per megabase of DNA sequence over time. The sharp drop in 2008 is due to the switch from Sanger sequencing to NGS in sequencing centers. Since 2008, the cost of DNA sequencing has decreased at a greater rate than predicted by Moore's Law (grey line). Moore's Law was based on the observation that, in the computer industry, the number of transistors per square inch (a measure of computational power) doubles every year. Image adapted from *http://www.genome.gov/sequencingcosts/*

Illumina sequencing method

One of the leading NGS platforms is the Illumina/Solexa platform, which was used for the work in this thesis. This platform utilizes a sequencing-by-synthesis approach to sequence millions of clusters bound to the surface of a glass slide known as a flow cell [85]. To prepare a DNA library for sequencing, the DNA is first sheared to fragments that are less than a few hundred bp in size, and adapters are ligated to the ends of these DNA fragments. After PCR amplification and purification, the final DNA library is then denatured to create single stranded DNA fragments, which are added to the flow cell and allowed to bind to primers that are attached to the flow cell surface. Each fragment is amplified through a process known as bridge amplification, to obtain millions of dense clusters of single stranded DNA that have sequencing primers attached (Figure 1.10) [88]. The sequencing reaction is then performed using fluorescently labeled nucleotides that act as reversible chain terminators, and the fluorescence is
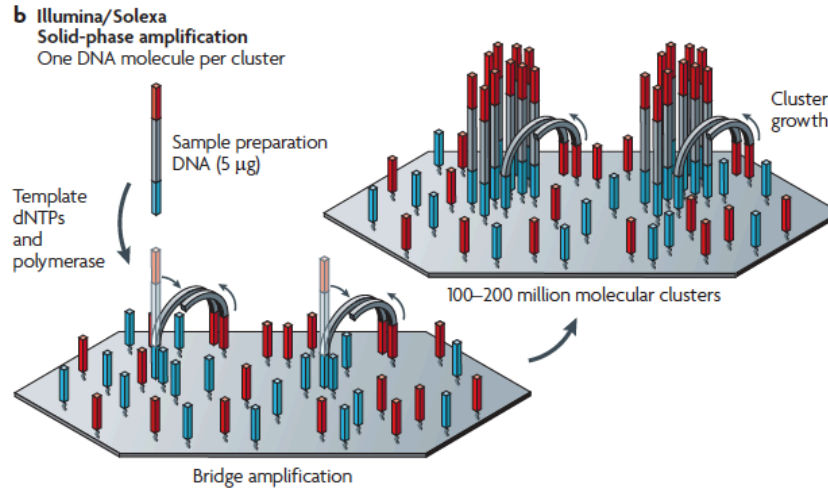
Figure 1.10. Generating clusters of DNA on Illumina flow cells. The DNA fragments in a library are denatured and allowed to bind to primers that are attached to the surface of the flow cell. Bridge PCR amplification enables the generation of millions of clusters on the flow cell surface. Image adapted from Metzker et al (2010)

measured after each new base is incorporated, to determine the sequence of each cluster (Figure 1.11) [88].

Applications of Next Generation Sequencing

In this thesis, three applications of NGS are explored: whole genome sequencing, targeted re-sequencing, and RNA-seq [89]. For whole genome sequencing, a library is prepared from genomic DNA and is then sequenced by NGS. The resulting DNA fragments are then aligned to a reference genome, and variants are identified by comparing the reference genome to the aligned sequence. This allows one to obtain sequence information for the entire genome, with the exception of regions that are difficult to sequence (such as CG rich regions) or regions that are difficult to align short reads to (such as repetitive elements).

If one is only interested in variants that lie within the coding regions of the genome, one can perform targeted re-sequencing. In this method, only a subset of the genome is sequenced, with the advantage being that higher coverage can be obtained for the same cost. One method for targeted re-sequencing is to perform an array-based target capture (Figure 1.12). DNA libraries are hybridized to an array containing DNA probes corresponding to a particular subset of the genome, such as all coding exons. This allows one to obtain a library that is enriched for fragments from the region-of-interest.
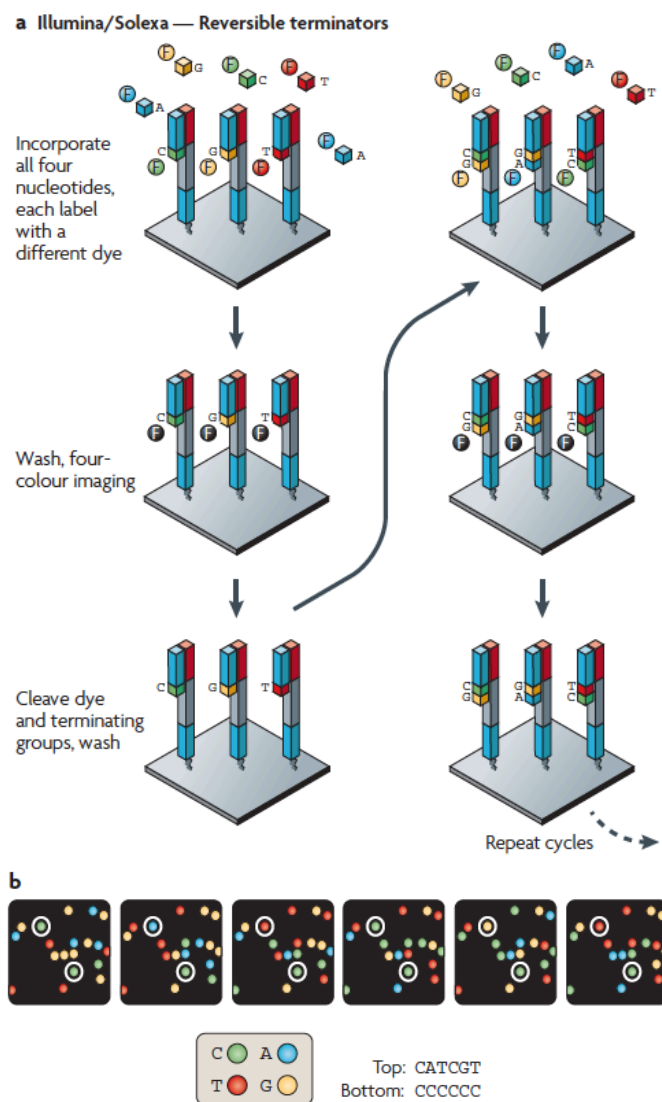
Figure 1.11. Illumina sequencing method using reversible chain terminators. A: Fluorescently labeled chain terminating nucleotides are added to the flow cell. After one nucleotide is added to each growing DNA strand, the excess nucleotides are washed away, and the flow cell is imaged. The fluorescent label is then removed to leave an unblocked 3' end, which is now able to bind to a new nucleotide. **B:** Images are taken after each round of nucleotide incorporation, enabling the sequence of each cluster to be determined. Image adapted from Metzker et al (2010).

A third application of NGS is RNA-seq. For this method, mRNA is converted to cDNA, which is then used to construct a library for NGS. This allows one to identify variants in the transcribed regions of the genome, and also allows one to determine the relative abundance of each transcript, based on the number of reads that mapped to it. Thus, RNA-seq can be used in place of microarrays to identify differentially expressed genes.

For all of the applications described here, multiplexing can be used to sequence more than one

sample on the same sequencing machine. This is achieved by adding a unique index or barcode to each

library, and then pooling multiple libraries (Figure 1.12). This allows one to obtain sequence for a greater

number of samples for the same cost, albeit fewer sequencing reads per sample.

1.6 Studying skeletal development using zebrafish forward genetics

Mutagenesis screens identify zebrafish mutants with skeletal defects

The zebrafish is a powerful model organism for the study of skeletal development [90]. While

differences exist between the morphology of the skeleton in fish compared to mammals, many of the

underlying developmental pathways are conserved. A number of large scale genetic screens have been

performed in zebrafish to identify mutants with skeletal phenotypes, such as defective patterning of the

craniofacial skeleton, or impaired cartilage differentiation and morphogenesis [91]. Many of these mutants

have furthered our understanding of human craniofacial or skeletal disorders. For example, the *dak*

mutant, which has defective cartilage morphogenesis and chondrocyte maturation, was found to contain a
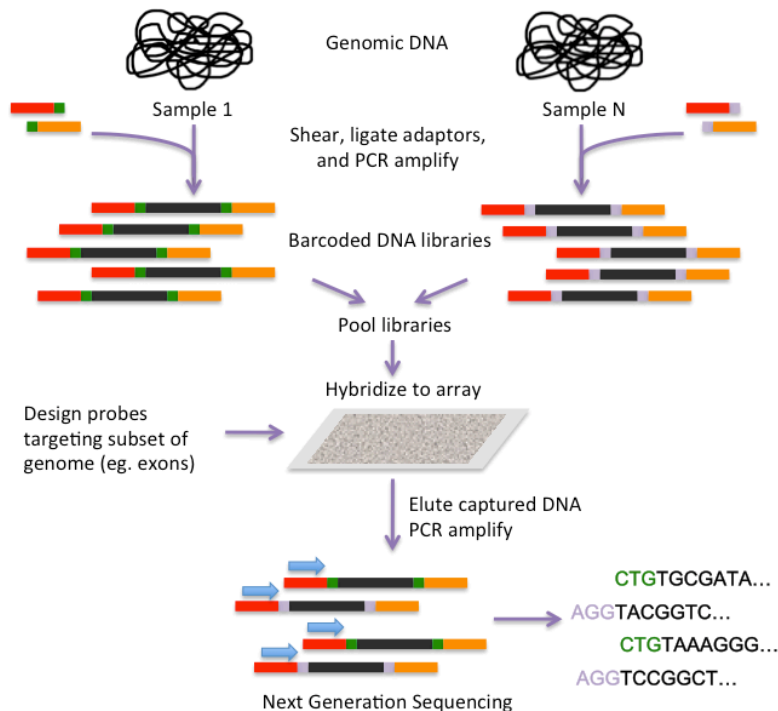
Figure 1.12. Multiplexed targeted re-sequencing by using a capture array to enrich for a subset of the genome. Genomic DNA libraries are prepared for each individual, using a unique barcode for each library. Libraries are pooled together and hybridized to an array that contains probes to target a subset of the genome, such as all exons, or only the sequence within a linkage interval. The captured DNA is then eluted from the array, PCR amplified, and submitted for NGS. Once the sequence data is obtained, each sequencing read can be assigned to an individual based on the barcode. In this illustration, the barcode sequence corresponds to the first three bases of each read.

mutation in the ortholog to the human *EXT2* gene, which underlies the MHE cartilage tumor syndrome disused above [92].

Traditional methods for mapping and cloning zebrafish mutants

The current bottleneck in zebrafish forward genetics is the identification of the causal mutation underlying the phenotype of a mutant. Mapping is traditionally performed by analyzing a predefined set of markers, such as microsatellite repeats that can be assessed by gel electrophoresis, in DNA pooled from multiple mutant fish, to find regions that co-segregate with the mutant phenotype [93,94]. The leads to a rough linkage position, which is then further refined by analyzing additional microsatellite repeats or single nucleotide polymorphisms (SNPs) in individual fish. Subsequently, candidate genes in the interval are sequenced, with the hope of identifying the mutation underlying the phenotype. This method for mapping and cloning zebrafish mutations is time consuming and costly [95]. Often, a large number of fish must be analyzed to narrow down the linkage interval. Furthermore, the predefined markers are not always polymorphic in the particular genetic background used and are therefore uninformative. In addition, selecting candidate genes for sequencing is biased, making one more likely to clone a mutation that lies in a gene with a known role in development, and less likely to successfully clone mutations in uncharacterized genes.

Using Next Generation Sequencing for mapping in model organisms

Following the development of NGS technologies, the possibility of identifying causal mutations by directly sequencing the genome, without the need for prior mapping data, has become a reality. For model organisms such as *C. elegans* and *A. thaliana*, it has been demonstrated that WGS can be used to simultaneously identify polymorphisms that can be used for mapping, as well as identify variants that are potential causative mutations [96–101]. These approaches rely on mapping recessive mutations based on homozygosity-by-descent (Figure 1.13). A mapping-cross is first performed to introduce genetic diversity. Multiple offspring that have the mutant phenotype are then pooled and sequenced. Regions tightly linked to the causative mutation will lack diversity and appear as homogenous sequence. In contrast, unlinked regions will contain diversity and appear as heterogeneous sequence. Therefore, by scanning the genome for regions of homogeneity, the genomic regions linked to the mutation can be identified. Finally, the actual sequence in the linked regions can be analyzed to identify novel variants that

may be the phenotype-causing mutation.

This approach has been successful applied to *C. elegans* and *A. thaliana* [96–101]. The advantage of these organisms is that their genomes are relatively small and common polymorphisms are well characterized. Importantly, due to their small genome size, it is not cost-prohibitive to obtain high coverage sequence of the entire genome, thereby facilitating accurate variant detection. Only recently has it become affordable to obtain low coverage sequence of the entire zebrafish genome, which is ten-fold larger than that of *C. elegans* or *A. thaliana*. However, with low coverage sequence data, it is difficult to distinguish true polymorphisms from sequencing errors, especially when dealing with the highly polymorphic zebrafish genome [102,103] for which a well-characterized database of all common polymorphisms does not exist. Thus, I chose to first

Figure 1.13. Mapping recessive mutants in model organisms based on homozygosity-by-descent. A mapping cross is performed to introduce genetic diversity. DNA from multiple mutants from the F2 generation is pooled and sequenced. Genomic regions tightly linked to the causative will appear homogeneous. As distance from the causative mutation increases, diversity increases due to recombination, and the sequence will appear heterogeneous.

establish a comprehensive list of zebrafish polymorphisms, by performing WGS on several wild-type strains. Secondly, I incorporated this list of polymorphisms into a pipeline I developed to map and clone zebrafish mutants based on low coverage WGS data (Chapter 4).

1.7   Scope of this thesis

In this thesis, I design and apply massively parallel sequencing strategies to address open questions in skeletal biology. In **chapter 2**, I perform targeted re-sequencing on multiple patients with metachondromatosis, and find that loss-of-function mutations in *PTPN11* underlie this cartilage tumor syndrome. In **chapter 3**, I further our understanding of the molecular mechanisms involved in cartilage tumor initiation and growth by using RNA-seq to identify gene expression changes that occur after the loss of *Ptpn11* in chondrocytes *in vitro*. I confirm these findings *in vivo* by studying mice with a conditional

Ptpn11 deletion. Finally, in **chapter 4**, I develop a whole-genome sequencing data analysis pipeline for mapping and cloning mutations in zebrafish, and apply this pipeline to map mutants with skeletal phenotypes.

1.8   Literature cited

1.   Lefebvre V, Bhattaram P: **Vertebrate skeletogenesis**. *Curr. Top. Dev. Biol.* 2010, **90**:291–317.

2.   Warman ML, Cormier-Daire V, Hall C, Krakow D, Lachman R, LeMerrer M, Mortier G, Mundlos S, Nishimura G, Rimoin DL, et al.: **Nosology and classification of genetic skeletal disorders: 2010 revision**. *Am. J. Med. Genet. A* 2011, **155A**:943–968.

3.   Maldonado M, Nam J: **The Role of Changes in Extracellular Matrix of Cartilage in the Presence of Inflammation on the Pathology of Osteoarthritis**. *Biomed Res Int* 2013, **2013**:284873.

4.   Pietschmann P, Rauner M, Sipos W, Kerschan-Schindl K: **Osteoporosis: an age-related and gender-specific disease--a mini-review**. *Gerontology* 2009, **55**:3–12.

5.   Wilt FH, Killian CE, Livingston BT: **Development of calcareous skeletal elements in invertebrates**. *Differentiation* 2003, **71**:237–250.

6.   Teixeira CC, Agoston H, Beier F: **Nitric oxide, C-type natriuretic peptide and cGMP as regulators of endochondral ossification**. *Dev. Biol.* 2008, **319**:171–178.

7.   Wuelling M, Vortkamp A: **Chondrocyte proliferation and differentiation**. *Endocr Dev* 2011, **21**:1–11.

8.   Mackie EJ, Ahmed YA, Tatarczuch L, Chen K-S, Mirams M: **Endochondral ossification: How cartilage is converted into bone in the developing skeleton**. *The International Journal of Biochemistry & Cell Biology* 2008, **40**:46–62.

9.   Tenenbaum HC, Palangio KG, Holmyard DP, Pritzker KP: **An ultrastructural study of osteogenesis in chick periosteum in vitro**. *Bone* 1986, **7**:295–302.

10. Maes C, Kobayashi T, Selig MK, Torrekens S, Roth SI, Mackem S, Carmeliet G, Kronenberg HM: **Osteoblast Precursors, but Not Mature Osteoblasts, Move into Developing and Fractured Bones along with Invading Blood Vessels**. *Developmental Cell* 2010, **19**:329–344.

11. Kawanami A, Matsushita T, Chan YY, Murakami S: **Mice expressing GFP and CreER in osteochondro progenitor cells in the periosteum**. *Biochemical and Biophysical Research Communications* 2009, **386**:477–482.

12. Lai LP, Mitchell J: **Indian hedgehog: Its roles and regulation in endochondral bone development**. *Journal of Cellular Biochemistry* 2005, **96**:1163–1173.

13. Kronenberg HM: **Developmental regulation of the growth plate**. *Nature* 2003, **423**:332–336.

14. Long F, Zhang XM, Karp S, Yang Y, McMahon AP: **Genetic manipulation of hedgehog signaling in the endochondral skeleton reveals a direct role in the regulation of chondrocyte proliferation**. *Development* 2001, **128**:5099–5108.

15. Chung U, Schipani E, McMahon AP, Kronenberg HM: **Indian hedgehog couples chondrogenesis to osteogenesis in endochondral bone development**. *Journal of Clinical Investigation* 2001, **107**:295–304.

16. Yoon BS, Pogue R, Ovchinnikov DA, Yoshii I, Mishina Y, Behringer RR, Lyons KM: **BMPs regulate multiple aspects of growth-plate chondrogenesis through opposing actions on FGF pathways**. *Development* 2006, **133**:4667–4678.

17. Minina E, Wenzel HM, Kreschel C, Karp S, Gaffield W, McMahon AP, Vortkamp A: **BMP and Ihh/PTHrP signaling interact to coordinate chondrocyte proliferation and differentiation**. *Development* 2001, **128**:4523–4534.

18. Hung IH, Yu K, Lavine KJ, Ornitz DM: **FGF9 regulates early hypertrophic chondrocyte differentiation and skeletal vascularization in the developing stylopod**. *Dev. Biol.* 2007, **307**:300–313.

19. Liu Z, Lavine KJ, Hung IH, Ornitz DM: **FGF18 is required for early chondrocyte proliferation, hypertrophy and vascular invasion of the growth plate**. *Dev. Biol.* 2007, **302**:80–91.

20. Iwata T, Chen L, Li C, Ovchinnikov DA, Behringer RR, Francomano CA, Deng CX: **A neonatal lethal mutation in FGFR3 uncouples proliferation and differentiation of growth plate chondrocytes in embryos**. *Hum. Mol. Genet.* 2000, **9**:1603–1613.

21. Colvin JS, Bohne BA, Harding GW, McEwen DG, Ornitz DM: **Skeletal overgrowth and deafness in mice lacking fibroblast growth factor receptor 3**. *Nat. Genet.* 1996, **12**:390–397.

22. Naski MC, Colvin JS, Coffin JD, Ornitz DM: **Repression of hedgehog signaling and BMP4 expression in growth plate cartilage by fibroblast growth factor receptor 3**. *Development* 1998, **125**:4977–4988.

23. Liu Z: **Coordination of chondrogenesis and osteogenesis by fibroblast growth factor 18**. *Genes & Development* 2002, **16**:859–869.

24. Minina E, Kreschel C, Naski MC, Ornitz DM, Vortkamp A: **Interaction of FGF, Ihh/Pthlh, and BMP signaling integrates chondrocyte proliferation and hypertrophic differentiation**. *Dev. Cell* 2002, **3**:439–449.

25. Parker EA, Hegde A, Buckley M, Barnes KM, Baron J, Nilsson O: **Spatial and temporal regulation of GH-IGF-related gene expression in growth plate cartilage**. *J. Endocrinol.* 2007, **194**:31–40.

26. Wang Y: **Insulin-Like Growth Factor-I Is Essential for Embryonic Bone Development**. *Endocrinology* 2006, **147**:4753–4761.

27. Liu JP, Baker J, Perkins AS, Robertson EJ, Efstratiadis A: **Mice carrying null mutations of the genes encoding insulin-like growth factor I (Igf-1) and type 1 IGF receptor (Igf1r)**. *Cell* 1993, **75**:59–72.

28. Zhang X, Siclari VA, Lan S, Zhu J, Koyama E, Dupuis HL, Enomoto-Iwamoto M, Beier F, Qin L: **The critical role of the epidermal growth factor receptor in endochondral ossification**. *J. Bone Miner. Res.* 2011, **26**:2622–2633.

29. Usmani SE, Pest MA, Kim G, Ohora SN, Qin L, Beier F: **Transforming growth factor alpha controls the transition from hypertrophic cartilage to bone during endochondral bone growth**. *Bone* 2012, **51**:131–141.

30. Wang K, Yamamoto H, Chin JR, Werb Z, Vu TH: **Epidermal growth factor receptor-deficient mice have delayed primary endochondral ossification because of defective osteoclast recruitment**. *J. Biol. Chem.* 2004, **279**:53848–53856.

31. Sibilia M, Wagner B, Hoebertz A, Elliott C, Marino S, Jochum W, Wagner EF: **Mice humanised for the EGF receptor display hypomorphic phenotypes in skin, bone and heart**. *Development* 2003, **130**:4515–4525.

32. Fisher MC, Clinton GM, Maihle NJ, Dealy CN: **Requirement for ErbB2/ErbB signaling in developing cartilage and bone**. *Dev. Growth Differ.* 2007, **49**:503–513.

33. Day T, Guo X, Garrettbeal L, Yang Y: **Wnt/β-Catenin Signaling in Mesenchymal Progenitors Controls Osteoblast and Chondrocyte Differentiation during Vertebrate Skeletogenesis**. *Developmental Cell* 2005, **8**:739–750.

34. Hill T, Spater D, Taketo M, Birchmeier W, Hartmann C: **Canonical Wnt/β-Catenin Signaling Prevents Osteoblasts from Differentiating into Chondrocytes**. *Developmental Cell* 2005, **8**:727–738.

35. Dao DY, Jonason JH, Zhang Y, Hsu W, Chen D, Hilton MJ, O'Keefe RJ: **Cartilage-specific β-catenin signaling regulates chondrocyte maturation, generation of ossification centers, and perichondrial bone formation during skeletal development**. *Journal of Bone and Mineral Research* 2012, **27**:1680–1694.

36. Guo X, Mak KK, Taketo MM, Yang Y: **The Wnt/β-Catenin Pathway Interacts Differentially with PTHrP Signaling to Control Chondrocyte Hypertrophy and Final Maturation**. *PLoS ONE* 2009, **4**:e6067.

37. Hu H: **Sequential roles of Hedgehog and Wnt signaling in osteoblast development**. *Development* 2004, **132**:49–60.

38. Bovée JVMG, Hogendoorn PCW, Wunder JS, Alman BA: **Cartilage tumours and bone development: molecular pathology and possible therapeutic targets**. *Nature Reviews Cancer* 2010, **10**:481–488.

39. Horvai A, Unni KK: **Premalignant conditions of bone**. *J Orthop Sci* 2006, **11**:412–423.

40. Pansuriya TC, Kroon HM, Bovée JVMG: **Enchondromatosis: insights on the different subtypes**. *Int J Clin Exp Pathol* 2010, **3**:557–569.

41. Jennes I, Pedrini E, Zuntini M, Mordenti M, Balkassmi S, Asteggiano CG, Casey B, Bakker B, Sangiorgi L, Wuyts W: **Multiple osteochondromas: mutation update and description of the multiple osteochondromas mutation database (MOdb)**. *Hum. Mutat.* 2009, **30**:1620–1627.

42. Amary MF, Bacsi K, Maggiani F, Damato S, Halai D, Berisha F, Pollock R, O'Donnell P, Grigoriadis A, Diss T, et al.: **IDH1 and IDH2 mutations are frequent events in central chondrosarcoma and central and periosteal chondromas but not in other mesenchymal tumours**. *The Journal of Pathology* 2011, **224**:334–343.

43. Amary MF, Damato S, Halai D, Eskandarpour M, Berisha F, Bonar F, McCarthy S, Fantin VR, Straley KS, Lobo S, et al.: **Ollier disease and Maffucci syndrome are caused by somatic mosaic mutations of IDH1 and IDH2**. *Nature Genetics* 2011, **43**:1262–1265.

44. Pansuriya TC, van Eijk R, d' Adamo P, van Ruler MAJH, Kuijjer ML, Oosting J, Cleton-Jansen A-M, van Oosterwijk JG, Verbeke SLJ, Meijer D, et al.: **Somatic mosaic IDH1 and IDH2 mutations are**

associated with enchondroma and spindle cell hemangioma in Ollier disease and Maffucci syndrome. *Nature Genetics* 2011, **43**:1256–1261.

45. Superti-Furga A, Spranger J, Nishimura G: **Enchondromatosis revisited: new classification with molecular basis**. *Am J Med Genet C Semin Med Genet* 2012, **160C**:154–164.

46. Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A, Li Y, Bhagwat N, Vasanthakumar A, Fernandez HF, et al.: **Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation**. *Cancer Cell* 2010, **18**:553–567.

47. Hopyan S, Gokgoz N, Poon R, Gensure RC, Yu C, Cole WG, Bell RS, Jüppner H, Andrulis IL, Wunder JS, et al.: **A mutant PTH/PTHrP type I receptor in enchondromatosis**. *Nature Genetics* 2002, **30**:306–310.

48. Douis H, Davies AM, James SL, Kindblom LG, Grimer RJ, Johnson KJ: **Can MR imaging challenge the commonly accepted theory of the pathogenesis of solitary enchondroma of long bone?** *Skeletal Radiology* 2012, **41**:1537–1542.

49. Bovee JVMG: **EXTra hit for mouse osteochondroma**. *Proceedings of the National Academy of Sciences* 2010, **107**:1813–1814.

50. Dreyfuss JL, Regatieri CV, Jarrouge TR, Cavalheiro RP, Sampaio LO, Nader HB: **Heparan sulfate proteoglycans: structure, protein interactions and cell signaling**. *An. Acad. Bras. Cienc.* 2009, **81**:409–429.

51. Senay C, Lind T, Muguruma K, Tone Y, Kitagawa H, Sugahara K, Lidholt K, Lindahl U, Kusche-Gullberg M: **The EXT1/EXT2 tumor suppressors: catalytic activities and role in heparan sulfate biosynthesis**. *EMBO Rep.* 2000, **1**:282–286.

52. McCormick C, Duncan G, Goutsos KT, Tufaro F: **The putative tumor suppressors EXT1 and EXT2 form a stable complex that accumulates in the Golgi apparatus and catalyzes the synthesis of heparan sulfate**. *Proc. Natl. Acad. Sci. U.S.A.* 2000, **97**:668–673.

53. Huegel J, Mundy C, Sgariglia F, Nygren P, Billings PC, Yamaguchi Y, Koyama E, Pacifici M: **Perichondrium phenotype and border function are regulated by Ext1 and heparan sulfate in developing long bones: A mechanism likely deranged in Hereditary Multiple Exostoses**. *Dev. Biol.* 2013, doi:10.1016/j.ydbio.2013.02.008.

54. Jones KB, Piombo V, Searby C, Kurriger G, Yang B, Grabellus F, Roughley PJ, Morcuende JA, Buckwalter JA, Capecchi MR, et al.: **A mouse model of osteochondromagenesis from clonal inactivation of Ext1 in chondrocytes**. *Proc. Natl. Acad. Sci. U.S.A.* 2010, **107**:2054–2059.

55. Matsumoto K, Irie F, Mackem S, Yamaguchi Y: **A mouse model of chondrocyte-specific somatic mutation reveals a role for Ext1 loss of heterozygosity in multiple hereditary exostoses**. *Proceedings of the National Academy of Sciences* 2010, **107**:10932–10937.

56. Bassett GS, Cowell HR: **Metachondromatosis. Report of four cases**. *J Bone Joint Surg Am* 1985, **67**:811–814.

57. Bovée JVMG, Hameetman L, Kroon HM, Aigner T, Hogendoorn PCW: **EXT-related pathways are not involved in the pathogenesis of dysplasia epiphysealis hemimelica and metachondromatosis**. *J. Pathol.* 2006, **209**:411–419.

58. Herman TE, Chines A, McAlister WH, Gottesman GS, Eddy MC, Whyte MP: **Metachondromatosis: report of a family with facial features mildly resembling trichorhinophalangeal syndromePediatr Radiol 1997 Nov;27(11):864**. *Pediatr Radiol* 1997, **27**:436–441.

59. Hunter AG, Kozlowski K, Hochberger O: **Metachondromatosis**. *Can Assoc Radiol J* 1995, **46**:202–208.

60. Ikegawa S, Nagano A, Matsushita T, Nakamura K: **Metachondromatosis: a report of two cases in a family**. *Nippon Seikeigeka Gakkai Zasshi* 1992, **66**:460–466.

61. Mavrogenis AF, Skarpidi E, Papakonstantinou O, Papagelopoulos PJ: **Chondrosarcoma in metachondromatosis: a case report**. *J Bone Joint Surg Am* 2010, **92**:1507–1513.

62. Pannier S, Legeai-Mallet L: **Hereditary multiple exostoses and enchondromatosis**. *Best Pract Res Clin Rheumatol* 2008, **22**:45–54.

63. Deprez FC, Beltrán-Marín M, Malghem J, Menten R, Clapuyt R: **Solitary osteochondroma: spontaneous regression**. *JBR-BTR* 2011, **94**:217.

64. Bauler TJ, Kamiya N, Lapinski PE, Langewisch E, Mishina Y, Wilkinson JE, Feng G-S, King PD: **Development of severe skeletal defects in induced SHP-2-deficient adult mice: a model of skeletal malformation in humans with SHP-2 mutations**. *Dis Model Mech* 2011, **4**:228–239.

65. Kim HKW, Feng G-S, Chen D, King PD, Kamiya N: **Targeted disruption of Shp2 in chondrocytes leads to metachondromatosis with multiple cartilaginous protrusions**. *J. Bone Miner. Res.* 2013, doi:10.1002/jbmr.2062.

66. Kim HKW, Aruwajoye O, Sucato D, Richards BS, Feng G-S, Chen D, King P, Kamiya N: **Induction of SHP2-deficiency in chondrocytes causes severe scoliosis and kyphosis in mice**. *Spine* 2013, doi:10.1097/BRS.0b013e3182a3d370.

67. Yang W, Wang J, Moore DC, Liang H, Dooner M, Wu Q, Terek R, Chen Q, Ehrlich MG, Quesenberry PJ, et al.: **Ptpn11 deletion in a novel progenitor causes metachondromatosis by inducing hedgehog signalling**. *Nature* 2013, **499**:491–495.

68. Grossmann KS, Rosário M, Birchmeier C, Birchmeier W: **The tyrosine phosphatase Shp2 in development and cancer**. *Adv. Cancer Res.* 2010, **106**:53–89.

69. Yang W, Klaman LD, Chen B, Araki T, Harada H, Thomas SM, George EL, Neel BG: **An Shp2/SFK/Ras/Erk Signaling Pathway Controls Trophoblast Stem Cell Survival**. *Developmental Cell* 2006, **10**:317–327.

70. Dance M, Montagner A, Salles J-P, Yart A, Raynal P: **The molecular functions of Shp2 in the Ras/Mitogen-activated protein kinase (ERK1/2) pathway**. *Cell. Signal.* 2008, **20**:453–459.

71. Edouard T, Combier JP, Nedelec A, Bel-Vialar S, Metrich M, Conte-Auriol F, Lyonnet S, Parfait B, Tauber M, Salles JP, et al.: **Functional Effects of PTPN11 (SHP2) Mutations Causing LEOPARD Syndrome on Epidermal Growth Factor-Induced Phosphoinositide 3-Kinase/AKT/Glycogen Synthase Kinase 3 Signaling**. *Molecular and Cellular Biology* 2010, **30**:2498–2507.

72. Takahashi A, Tsutsumi R, Kikuchi I, Obuse C, Saito Y, Seidi A, Karisch R, Fernandez M, Cho T, Ohnishi N, et al.: **SHP2 tyrosine phosphatase converts parafibromin/Cdc73 from a tumor suppressor to an oncogenic driver**. *Mol. Cell* 2011, **43**:45–56.

73. Zhang SQ, Yang W, Kontaridis MI, Bivona TG, Wen G, Araki T, Luo J, Thompson JA, Schraven BL, Philips MR, et al.: **Shp2 Regulates Src Family Kinase Activity and Ras/Erk Activation by Controlling Csk Recruitment**. *Molecular Cell* 2004, **13**:341–355.

74. Dittrich A, Quaiser T, Khouri C, Görtz D, Mönnigmann M, Schaper F: **Model-driven experimental analysis of the function of SHP-2 in IL-6-induced Jak/STAT signaling**. *Molecular BioSystems* 2012, **8**:2119.

75. Schoenwaelder SM, Petch LA, Williamson D, Shen R, Feng G-S, Burridge K: **The protein tyrosine phosphatase Shp-2 regulates RhoA activity**. *Current Biology* 2000, **10**:1523–1526.

76. Yu M: **The Scaffolding Adapter Gab2, via Shp-2, Regulates Kit-evoked Mast Cell Proliferation by Activating the Rac/JNK Pathway**. *Journal of Biological Chemistry* 2006, **281**:28615–28626.

77. He Z, Zhu HH, Bauler TJ, Wang J, Ciaraldi T, Alderson N, Li S, Raquil M-A, Ji K, Wang S, et al.: **Nonreceptor tyrosine phosphatase Shp2 promotes adipogenesis through inhibition of p38 MAP kinase**. *Proc. Natl. Acad. Sci. U.S.A.* 2013, **110**:E79–88.

78. Fornaro M, Burch PM, Yang W, Zhang L, Hamilton CE, Kim JH, Neel BG, Bennett AM: **SHP-2 activates signaling of the nuclear factor of activated T cells to promote skeletal muscle growth**. *The Journal of Cell Biology* 2006, **175**:87–97.

79. You M, Flick LM, Yu D, Feng GS: **Modulation of the nuclear factor kappa B pathway by Shp-2 tyrosine phosphatase in mediating the induction of interleukin (IL)-6 by IL-1 or tumor necrosis factor**. *J. Exp. Med.* 2001, **193**:101–110.

80. Tsutsumi R, Takahashi A, Azuma T, Higashi H, Hatakeyama M: **Focal Adhesion Kinase Is a Substrate and Downstream Effector of SHP-2 Complexed with Helicobacter pylori CagA**. *Molecular and Cellular Biology* 2005, **26**:261–276.

81. Tidyman WE, Rauen KA: **The RASopathies: developmental syndromes of Ras/MAPK pathway dysregulation**. *Curr. Opin. Genet. Dev.* 2009, **19**:230–236.

82. Stevenson DA, Yang F-C: **The musculoskeletal phenotype of the RASopathies**. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* 2011, **157**:90–103.

83. Keilhack H: **Diverse Biochemical Properties of Shp2 Mutants: IMPLICATIONS FOR DISEASE PHENOTYPES**. *Journal of Biological Chemistry* 2005, **280**:30984–30993.

84. Lauriol J, Kontaridis MI: **PTPN11-Associated Mutations in the Heart: Has LEOPARD Changed Its RASpots?** *Trends in Cardiovascular Medicine* 2011, **21**:97–104.

85. Hutchison CA 3rd: **DNA sequencing: bench to bedside and beyond**. *Nucleic Acids Res.* 2007, **35**:6227–6237.

86. Pareek CS, Smoczynski R, Tretyn A: **Sequencing technologies and genome sequencing**. *J. Appl. Genet.* 2011, **52**:413–435.

87. **http://www.genome.gov/sequencingcosts/**. [date unknown], [no volume].

88. Metzker ML: **Sequencing technologies - the next generation**. *Nat. Rev. Genet.* 2010, **11**:31–46.

89. Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M: **Comparison of next-generation sequencing systems**. *J. Biomed. Biotechnol.* 2012, **2012**:251364.

90. Javidan Y, Schilling TF: **Development of cartilage and bone**. *Methods Cell Biol.* 2004, **76**:415–436.

91. Yelick PC, Schilling TF: **Molecular dissection of craniofacial development using zebrafish**. *Crit. Rev. Oral Biol. Med.* 2002, **13**:308–322.

92. Clément A, Wiweger M, von der Hardt S, Rusch MA, Selleck SB, Chien C-B, Roehl HH: **Regulation of zebrafish skeletogenesis by ext2/dackel and papst1/pinscher**. *PLoS Genet.* 2008, **4**:e1000136.

93. Geisler R, Rauch G-J, Geiger-Rudolph S, Albrecht A, van Bebber F, Berger A, Busch-Nentwich E, Dahm R, Dekens MPS, Dooley C, et al.: **Large-scale mapping of mutations affecting zebrafish development**. *BMC Genomics* 2007, **8**:11.

94. Zhou Y, Zon LI: **The zon laboratory guide to positional cloning in zebrafish**. *Methods Cell Biol.* 2011, **104**:287–309.

95. Henke K, Bowen ME, Harris MP: **Perspectives for identification of mutations in the zebrafish: Making use of next-generation sequencing technologies for forward genetic approaches**. *Methods* 2013, **62**:185–196.

96. Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen AH, Nielsen KL, Jørgensen J-E, Weigel D, Andersen SU: **SHOREmap: simultaneous mapping and mutation identification by deep sequencing**. *Nat. Methods* 2009, **6**:550–551.

97. Cuperus JT, Montgomery TA, Fahlgren N, Burke RT, Townsend T, Sullivan CM, Carrington JC: **Identification of MIR390a precursor processing-defective mutants in Arabidopsis by direct genome sequencing**. *Proc. Natl. Acad. Sci. U.S.A.* 2010, **107**:466–471.

98. Doitsidou M, Poole RJ, Sarin S, Bigelow H, Hobert O: **C. elegans mutant identification with a one-step whole-genome-sequencing and SNP mapping strategy**. *PLoS ONE* 2010, **5**:e15435.

99. Zuryn S, Le Gras S, Jamet K, Jarriault S: **A strategy for direct mapping and identification of mutations by whole-genome sequencing**. *Genetics* 2010, **186**:427–430.

100. Austin RS, Vidaurre D, Stamatiou G, Breit R, Provart NJ, Bonetta D, Zhang J, Fung P, Gong Y, Wang PW, et al.: **Next-generation mapping of Arabidopsis genes**. *Plant J.* 2011, **67**:715–725.

101. Uchida N, Sakamoto T, Kurata T, Tasaka M: **Identification of EMS-induced causal mutations in a non-reference Arabidopsis thaliana accession by whole genome sequencing**. *Plant Cell Physiol.* 2011, **52**:716–722.

102. Bradley KM, Elmore JB, Breyer JP, Yaspan BL, Jessen JR, Knapik EW, Smith JR: **A major zebrafish polymorphism resource for genetic mapping**. *Genome Biol.* 2007, **8**:R55.

103. Guryev V, Koudijs MJ, Berezikov E, Johnson SL, Plasterk RHA, van Eeden FJM, Cuppen E: **Genetic variation in the zebrafish**. *Genome Res.* 2006, **16**:491–497.

Chapter **2**

Loss-of-function mutations in *PTPN11* cause metachondromatosis, but not Ollier disease or Maffucci syndrome

2.1 Abstract

Metachondromatosis (MC) is a rare, autosomal dominant, incompletely penetrant combined exostosis and enchondromatosis tumor syndrome. MC is clinically distinct from other multiple exostosis or multiple enchondromatosis syndromes and is unlinked to *EXT1* and *EXT2*, the genes responsible for autosomal dominant multiple hereditary exostoses (MHE). To identify a gene for MC, we performed linkage analysis with high-density SNP arrays in a single family, used a targeted array to capture exons and promoter sequences from the linked interval in 16 participants from 11 MC families, and sequenced the captured DNA using high-throughput parallel sequencing technologies. DNA capture and parallel sequencing identified heterozygous putative loss-of-function mutations in *PTPN11* in 4 of the 11 families. Sanger sequence analysis of *PTPN11* coding regions in a total of 17 MC families identified mutations in 10 of them (5 frameshift, 2 nonsense, and 3 splice-site mutations). Copy number analysis of sequencing reads from a second targeted capture that included the entire *PTPN11* gene identified an additional family with a 15 kb deletion spanning exon 7 of *PTPN11*. Microdissected MC lesions from two patients with *PTPN11* mutations demonstrated loss-of-heterozygosity for the wild-type allele. We next sequenced *PTPN11* in DNA samples from 54 patients with the multiple enchondromatosis disorders Ollier disease or Maffucci syndrome, but found no coding sequence *PTPN11* mutations. We conclude that heterozygous loss-of-function mutations in *PTPN11* are a frequent cause of MC, that lesions in patients with MC appear to arise following a ''second hit,'' that MC may be locus heterogeneous since 1 familial and 5 sporadically occurring cases lacked obvious disease-causing *PTPN11* mutations, and that *PTPN11* mutations are not a common cause of Ollier disease or Maffucci syndrome.

2.2   Contributing authors

Margot E. Bowen[1,2,3*], Eric D. Boyden[1,2,3*], Ingrid A. Holm[4,5], Belinda Campos-Xavier[6], Luisa Bonafé[6], Andrea Superti-Furga[6], Shiro Ikegawa[7], Valerie Cormier-Daire[8], Judith V. Bovée[9], Twinkal C. Pansuriya[9], Sérgio B. de Sousa[10], Ravi Savarirayan[11,12], Elena Andreucci[11,12,13], Miikka Vikkula[14], Livia Garavelli[15], Caroline Pottinger[16], Toshihiko Ogino[17], Akinori Sakai[18], Bianca M. Regazzoni[19], Wim Wuyts[20], Luca Sangiorgi[21], Elena Pedrini[21], Mei Zhu[2,3], Harry P. Kozakewich[22], James R. Kasser[1], Jon G. Seidman[2,3], Kyle C. Kurek[1,22], Matthew L. Warman[1,2,3]

1 Department of Orthopaedic Surgery, Children's Hospital Boston and Harvard Medical School, Boston, Massachusetts, United States of America, **2** Howard Hughes Medical Institute, Boston, Massachusetts, United States of America, **3** Department of Genetics, Harvard Medical School, Boston, Massachusetts, United States of America, **4** Division of Genetics, Program in Genomics, and The Manton Center for Orphan Disease Research, Children's Hospital Boston, Boston, Massachusetts, United States of America, **5** Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, United States of America, **6** Division of Molecular Pediatrics, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland, **7** Laboratory for Bone and Joint Diseases, Center for Genomic Medicine, RIKEN, Tokyo, Japan, **8** Department of Medical Genetics, Paris Descartes University, INSERM U781, Hospital Necker Enfants Malades, Paris, France, **9** Department of Pathology, Leiden University Medical Centre, Leiden, The Netherlands, 10 Department of Medical Genetics, Hospital Pedia´trico de Coimbra, Coimbra, Portugal, **11** Victorian Clinical Genetics Services, Murdoch Childrens Research Institute, Melbourne, Australia, **12** Department of Pediatrics, University of Melbourne, Melbourne, Australia, **13** Department of Clinical Pathophysiology, University of Florence and Meyer Children's Hospital Genetics Unit, Florence, Italy, **14** de Duve Institute, Universite´ Catholique de Louvain, Brussels, Belgium, **15** Department of Clinical Genetics, Arcispedale S. Maria Nuova, Reggio Emilia, Italy, **16** Merseyside and Chesire Regional Genetics Service, Alder Hey Hospital, Liverpool, United Kingdom, **17** Department of Orthopaedic Surgery, Yamagata University Faculty of Medicine, Yamagata, Japan, **18** Department of Orthopaedic Surgery, University of Occupational and Environmental Health, Kitakyushu, Japan, **19** Department of Pediatrics, S. Anna Hospital, Lugano, Switzerland, **20** Department of Medical Genetics, University of Antwerp, Antwerp, Belgium, **21** Department of Medical Genetics, Rizzoli Orthopaedic Institute, Bologna, Italy, **22** Department of Pathology, Children's Hospital Boston and Harvard Medical School, Boston, Massachusetts, United States of America, * These authors contributed equally.

2.3   Specific contributions

Candidate's contributions

Agilent Array design; DNA extraction and library preparation; Next Generation Sequencing data analysis; Sanger sequencing; Laser Capture Microdissection; Multiplex Ligation-Dependent Probe Amplification (MLPA); SHP2 Immunodetection.

Other authors' contributions

Patient diagnosis and selection; Histological analysis of lesions; Collection of human blood and tissue samples; Genome-wide linkage analysis.

2.4   Introduction

Cartilage tumor syndromes are characterized by multiple cartilaginous bone tumors that develop in childhood, often causing significant morbidity and predisposing to chondrosarcoma. Tumors can form as exostoses (on the surface of bone), as in the autosomal dominant, multiple hereditary exostoses syndrome (MHE; MIM 133700 and 133701), or as endosteal tumors (within bone), as in the sporadically occurring multiple enchondromatosis disorders (MIM 166000) Ollier disease and Maffucci syndrome. In MHE, mutations in *EXT1* or *EXT2*, which encode heparan sulfate glycosyltransferases, affect chondrocyte orientation in the growth plate [1]. A small percentage of patients with Ollier syndrome have mutations in *PTH1R*, which encodes the receptor for parathyroid hormone and parathyroid hormone-related protein, causing altered chondrocyte differentiation in the growth plate [2]. The cause of Maffucci syndrome is unknown [3]. Patients with MHE do not develop endosteal tumors, and patients with Ollier disease or Maffucci syndrome do not develop exostotic tumors [1,3,4]. Patients with metachondromatosis (MC; MIM 156250) form exostotic and endosteal tumors (Figure 2.1). Fewer than 50 cases of MC have been published since Maroteaux's initial description in 1971 [5]. Exostotic lesions in MC occur frequently in the digits, involve metaphyses and epiphyses, and tend to grow toward the joint; in contrast, exostotic lesions in MHE occur frequently in the long bones, involve only the metaphyses, and tend to grow away from the joint [6–11]. MC exostotic lesions can also spontaneously decrease in size and completely regress [6,7,9,12]. Endosteal lesions in MC are common in the metaphyses of long bones and in the pelvis [7–11]. Avascular necrosis of the femoral head, due to endosteal tumors, has been a frequent complication in patients with MC [7,8,13–15]. Hand deformity due to endosteal tumors is uncommon in patients with MC, whereas it is often a significant problem for patients with Ollier disease and Maffucci syndrome [3]. Finally, malignant transformation has only been reported in one patient with MC, whereas it has been more frequently reported in patients with MHE, Ollier disease, and Maffucci syndrome [3,4,16]. The distinct distribution and clinical behavior of lesions in patients with MC, suggest that MC is pathophysiologically distinct from these other cartilage tumor syndromes. We therefore sought to better characterize MC and to determine its genetic basis.
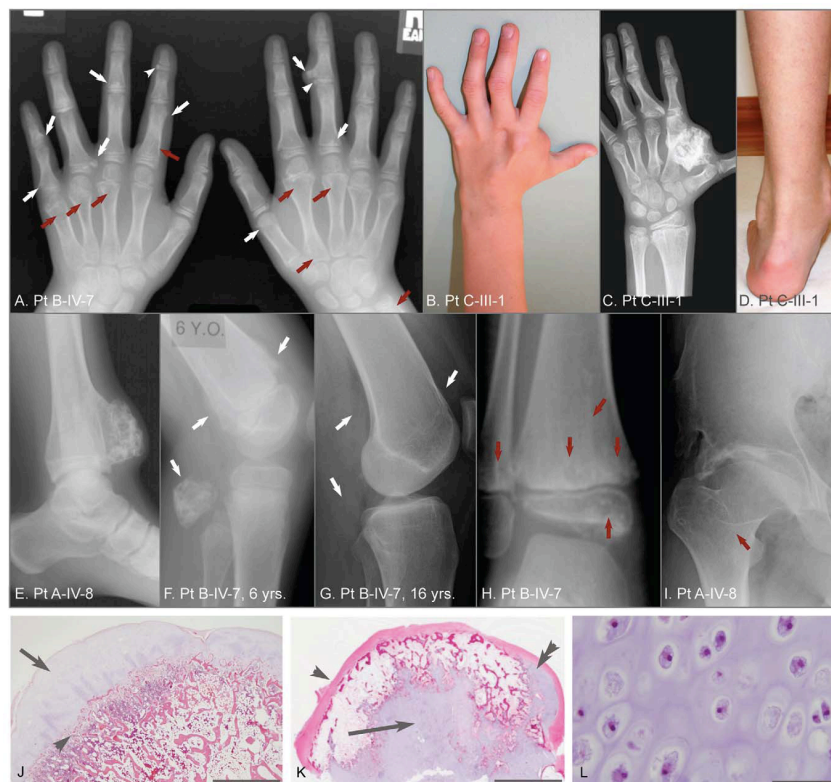
Figure 2.1. Clinical, radiographic, and histologic features of metachondromatosis. (A) Hand radiographs of participant B-IV-7, taken when 8-years-old. Exostotic lesions (white arrows) are present in the phalanges and metacarpals, and arise from the metaphysis (arrows) or epiphysis (arrowheads). Exostoses tend to point toward the adjacent joint. Endosteal lesions (red arrows) cause metaphyseal expansion. (B, C, D) Hand photograph and radiograph, and foot photograph of participant C-III-1 taken when 9-years-old, depicting mild shortening and deformity of the digits, a large exostotic lesion arising from the second metacarpal bone in the hand, and ankle enlargement superior to the malleoli due to exostoses of the tibia and fibula. (E) Ankle radiograph of participant A-IV-8 taken when 19-years-old depicting a recurrence of a previously excised exostotic lesion of the distal tibia that spans the physis. (F, G) Lateral knee radiographs of participant B-IV-7, taken at 6 years and 16 years, respectively. Note that multiple exostotic lesions of the distal femur and proximal fibula (white arrows) seen when 6-years-old (F) have regressed in the absence of surgical intervention by 16-years of age (G). (H) Ankle radiograph of participant B-IV-7, taken when 5-years-old, demonstrating radiolucency associated with endosteal lesions (red arrows) in the tibia and fibula, and mild metaphyseal flaring. Despite combined metaphyseal and epiphyseal involvement, this individual's linear growth was not affected. (I) Hip radiograph of participant A-IV-8 taken when 22-years-old depicting an endosteal lesion of the femoral neck (arrow) that has caused degeneration of the femoral head and spurring of the acetabulum. (J) Low power photomicrograph of an hematoxylin and eosin (H&E) stained section through an exostosis that had been excised from a patient with hereditary multiple exostoses. Note this exostosis is a typical osteochondroma, having a well-developed surface cartilaginous cap (arrow) and endochondral bone immediately below (arrowhead). The scale bar represents 0.15 cm (K) Photomicrograph of an H&E stained exostotic lesion excised from participant A-IV-5 when 5-years-old. This lesion is predominantly covered by a fibrous cap (arrow) and has only a small, eccentric cartilaginous cap (double arrowhead). The majority of cartilage in this, and in 14 other exostoses from patients with MC that have been analyzed, is found within a central core (arrow) and has bone formation occurring at the periphery of this cartilage core. The scale bar represents 0.5 cm. (L) High-power image of the central cartilage core shows chondrocytes with prominent cytoplasm no organization typical of a growth plate. The scale bar represents 100 μm

2.5 Results

Patient selection

We diagnosed participants as having MC based upon the presence of both multiple exostotic and endosteal cartilaginous lesions as previously described [5–10,15]. We excluded from analysis participants with solitary lesions, contiguous endosteal lesions suggestive of Ollier disease, soft tissue lesions suggestive of Maffucci syndrome, or radiographs suggestive of MHE. We included participants who had clinical and radiographic features of MC, even if they lacked a positive family history. For each patient, the clinical history and radiographs were reviewed by at least 3 authors. MC patients from 17 unrelated families from 9 countries were identified (Table 2.1). All participants gave their informed consent following the guidelines of each referring institution. In 10 families disease segregation is consistent with autosomal dominant inheritance. In 7 families, the disease is suspected to have arisen de novo. Immediate family members of patients with sporadically occurring MC were interviewed and examined, although detailed imaging was not performed. For 8 familial cases, blood or DNA was available from additional family members.

Clinical and pathologic features of metachondromatosis

Figure 2.1 depicts features seen in affected participants with MC. No phenotypic differences could be found between sporadic or familial cases of MC. Radiographs identify exostotic and endosteal lesions of the digits (Figure 2.1A–1C) and long bones (Figure 2.1E– 1H), along with degenerative hip disease secondary to endosteal lesions in the femoral neck (Figure 2.1I). Spontaneous regression of exostotic lesions is seen in radiographs obtained 10 years apart in the same patient (Figure 2.1F,1G). Also depicted in Figure 2.1 are histopathologic features that distinguish exostoses in patients with MHE from those in patients with MC, based upon a comparison of 30 exostoses excised from children with MHE and 15 exostotic lesions excised from 3 affected individuals with MC. Exostoses in children with MHE have cartilage caps with endochondral bone growth immediately beneath the cap (Figure 2.1J). In contrast, exostoses in children with MC have a predominantly fibrous cap and a core of disorganized cartilage.

Table 2.1. Summary of metachondromatosis families and *PTPN11* mutations

	Family	Familial/ sporadic	# Affecteds collected	Referral site	DNA change (chr12, hg19 build)	cDNA change	Predicted protein change	Mutation type	Mutation site	Method Identified	Included in Next Gen[1]	Included in aCGH[1]
1	A	F	10	USA	g.112,891,075 del5	c.409_413del5	p.V137RfsX17	frameshift	exon 4	Next Gen 1	2A, 1U (III-9,10, IV-5)	2A
2	B	F	2	USA	g.112,891,124 del11ins24[2]	c.458_468 del11ins24	p.T153KfsX8	frameshift	exon 4	Next Gen 1	2A, 1U (III-5, 6, IV-7)	1A
3	C	F	4	Switzerland	g.112,891,019 del2	c.353_354del2	p.S118WfsX10	frameshift	exon 4	Next Gen 1	2A (II-5, III-1)	1A
4	I	F	2	UK	g.112,888,279 A>T	c.295A>T	p.K99X	nonsense	exon 3	Sanger		
5	D	F	2	Japan	g.112,924,369 del1	c.1315del1	p.L439WfsX33	frameshift	exon 11	Sanger	1A	1A
6	E[3]	F	3	Japan	g.112,926,896 C>T	c.1516C>T	p.Q506X	nonsense	exon 13	Next Gen 1	1A	1A
7	J	S	1	Japan							1A	
8	K	S	1	Japan							1A	
9	L	S	1	Australia							1A	1A
10	M	S	1	Australia							1A	
11	F	F	3	France	g.112,893,752 A>C	c.643-2A>C	unknown	splice site	intron 5	Sanger	1A	
12	G	F	1	France	g.112,919,877 G>T	c.1093-1G>T	unknown	splice site	intron 9	Sanger		
13	S	S	1	Italy	g.112,897,487del146 29ins2		p.T253LfsX54	frameshift	exon 7	Next Gen 2	1A	
14	T	S	1	Italy								
15	H	F	1	Belgium	g.112,893,791 del4	c.680_683del4	p.E227AfsX6	frameshift	exon 10	Sanger		
16	O[4]	S	1	Netherlands	g.112,924,278G>A	c.1225-1G>A	unknown	splice site	intron 10	Sanger		
17	P[4]	F	1	Netherlands								

[1]A= Affected, U= Unaffected.
[2]Insertion of the following 24 bp sequence: AAGAACACAGGGGAGAGCAATGAC.
[3]Patients previously published in Reference 8.
[4]Patients previously published in Reference 10.

surrounded by trabecular bone (Figure 2.1K, 1L). In all MC cases, the lesions were bilateral and not obviously confined to a single body segment as in Ollier or Maffucci patients.

Linkage analysis.

We performed linkage analysis in the largest family (Family A, Figure 2.2A) to identify a genetic locus for MC. Raw genotype data were generated using Affymetrix 6.0 SNP arrays and multipoint parametric linkage analysis of the autosomal genome was performed using MERLIN [17]. Because non-penetrance and non-ascertainment are potential confounding factors in the diagnosis of MC, we analyzed only founders and affected individuals (Figure 2.2A). Although this limited the maximum attainable LOD score to 2.7, which is lower than the genome-wide significance threshold of 3.3, the dense marker set ensured a reasonable probability that only one large interval that achieved the maximum LOD score would be observed, with the remainder of the genome being excluded. We identified a single interval on chromosome 12, from 121.5 to 132.3 cM, that attained the maximum LOD score of 2.7 (Figure 2.2B). No other autosomal interval >1 cM yielded a peak LOD score >–1.9. Several intervals <1 cM attained LOD scores >0 (Figure 2.3), but we considered these unlikely to be candidate intervals and instead assumed they represented either unfiltered genotyping errors or short ancestral population haplotypes, rather than familial haplotypes inherited from a common ancestor.

Multiplexed targeted genomic capture and sequencing of linkage interval

We next performed array-based capture, followed by Illumina GAII sequencing, using barcoded DNA libraries created from 16 individuals in 11 families. We sequenced 1 affected individual in 8 families, 2 affected individuals in 1 family, and 2 affected and 1 unaffected individual in 2 families (Table 2.1). We prepared barcoded genomic DNA libraries, having an average insert size of 150 bp, using sheared DNA from each of the 16 individuals (Figure 2.4). We performed array-based targeted capture by pooling each DNA library and hybridizing the pooled sample to an Agilent Technologies 1M SureSelect DNA capture array containing 973,952 probes targeting 844,339 bp within the 8.6 Mb candidate interval, including 88.4% and 98.6% of UCSC exons and CCDS coding sequence, respectively. After hybridization and elution, the captured DNA was PCR amplified, purified to remove primer dimers, and sequenced on two lanes of an Illumina GA II. We obtained 50 million, 80 bp, single-end reads. Novobarcode software was used to sort the reads according to their 3 bp barcode, and Novoalign was used to align the reads to
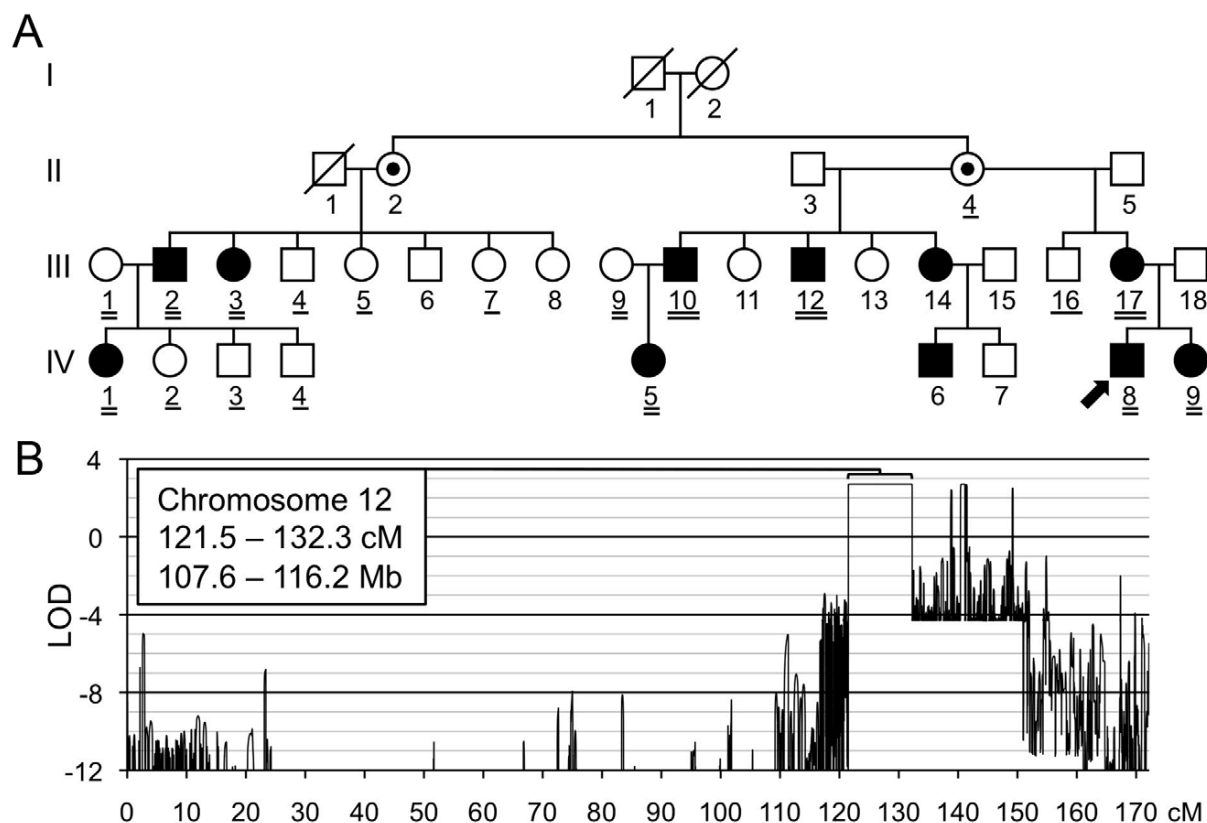
Figure 2.2. Linkage mapping of metachondromatosis to chromosome 12q. (A) Pedigree of the family (Family A) used to define the metachondromatosis candidate interval. Affected individuals have filled symbols. Individuals who were not examined but who were assumed to be obligate carriers have interior dots. An arrow identifies the proband. Participants whose DNA was used for linkage analysis are double underlined. Those participants with a single or double underlines were tested for the *PTPN11* mutation, after the mutation had been identified in participants AIII-10 and A-IV-5 by targeted array capture and Illumina sequencing. (B) LOD score plot of chromosome 12. Only one interval, larger than 1 cM, in the entire genome attained the maximum LOD score of 2.7. Several autosomal intervals, each smaller than 1 cM, also achieved maximum or positive LOD scores (Figure 2.3). These likely represent either genotyping errors or short ancestral haplotypes that are coincidental with linkage. The physical coordinates shown are derived from GRCh37/hg19.
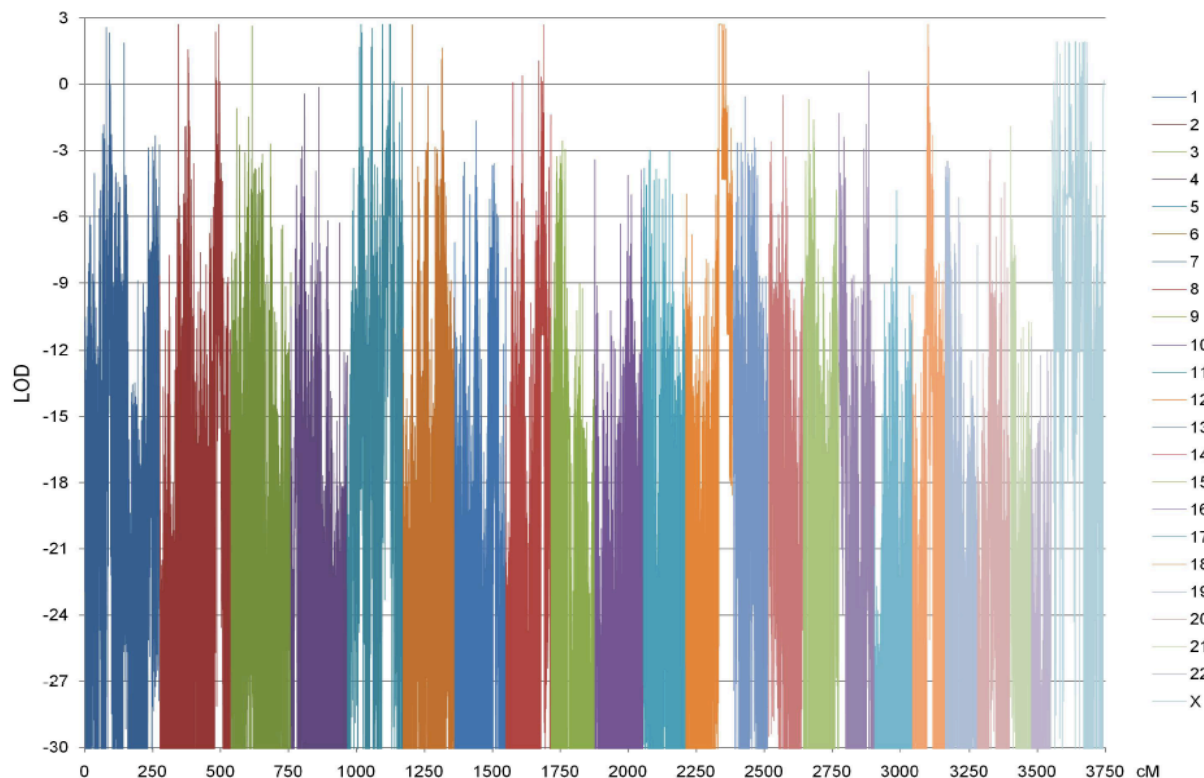
Figure 2.3. Linkage mapping of metachondromatosis to chromosome 12q. A whole genome LOD score plot is depicted. The chromosomes are color-coded according to the key on the right. Raw genotype data were generated using Affymetrix 6.0 SNP arrays and multipoint parametric linkage analysis was performed using MERLIN with Affymetrix Caucasian allele frequencies and deCODE Genetics genetic map positions. The disease allele frequency was estimated at 1E-7, and phenocopies and non-penetrance were not permitted (affectation probability 0/1/1). Error checking identified and removed unlikely genotypes during initial analysis, and the cleaned data were then reanalyzed. Because non-penetrance and non-ascertainment are potential confounding factors in the diagnosis of MC, we analyzed only founders and affected individuals. Our analysis identified one interval on chromosome 12 (orange above), from 121.5 to 132.3 cM, that attained the maximum LOD score of 2.7. The minimum linked interval is bounded by rs1861693 (LOD score 2.1; GRCh37/hg19 coordinate 107603157) and rs1520173 (2.3; 116182525), and is flanked proximally by rs12422243 (−5.8; 107595567) and distally by rs2460488 (−4.3; 116187660), for a maximum linked interval of 8,592,092 bp. We did not observe any other autosomal interval >1 cM with a peak LOD score >−1.9. Several autosomal intervals <1 cM attained LOD scores >0, but we considered these unlikely to be candidate intervals and instead assumed they represented either unfiltered genotyping errors or short ancestral population haplotypes, rather than familial haplotypes inherited from a common ancestor. Although MC is well documented as autosomal dominant and many of our own pedigrees show male-to-male inheritance, family A does not; therefore we also examined the X chromosome. Four X-linked intervals >1 cM attained LOD scores >−2, but the largest was 3.4 cM and the highest LOD score was 0.3; we therefore assumed these intervals also represented either unfiltered genotyping errors or short ancestral population haplotypes, rather than familial haplotypes inherited from a common ancestor.
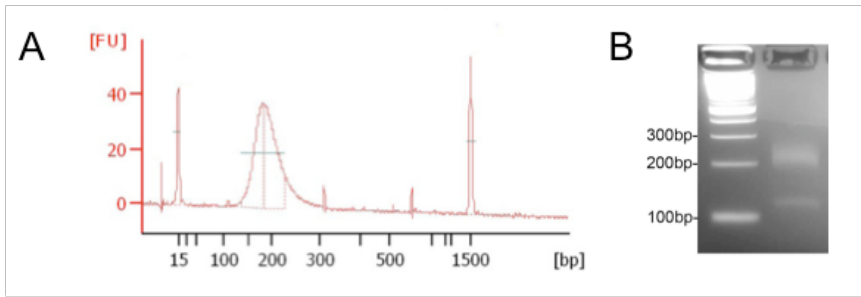
Figure 2.4. Amplified libraries before and after array capture. (A) Bioanalyzer plot of a representative amplified genomic DNA library before targeted array capture. Peaks at 15 bp and 1500 bp represent size standards. The broad peak between 100 and 300 bp comprises the barcoded and amplified DNA library. (B) Photograph of an Ethidium Bromide stained 4% agarose gel containing the pooled library after targeted array capture and amplification, but before E-gel purification. Note the broad band of amplimer between 200 bp and 300 bp in size. These sized amplimers were used to obtain Illumina GA II sequence.

Table 2.2. Novel coding variants identified in three metachondromatosis families

Family	A		B		C	
Individual	III-10	IV-5	IV-7	III-5	III-1	II-5
Total variants*	529	388	536	1499	480	709
Coding	45	40	50	200	37	51
Not listed in dbSNP	6	12	10	167	5	7
Shared	3		1		1	
Not present in unaffected family member	1		1		n/a	
Genes affected and predicted protein changes	*PTPN11* p.V137fs		*PTPN11* p.T153fs		*PTPN11* p.S118fs	

*filtered to remove low confidence variants.
doi:10.1371/journal.pgen.1002050.t001
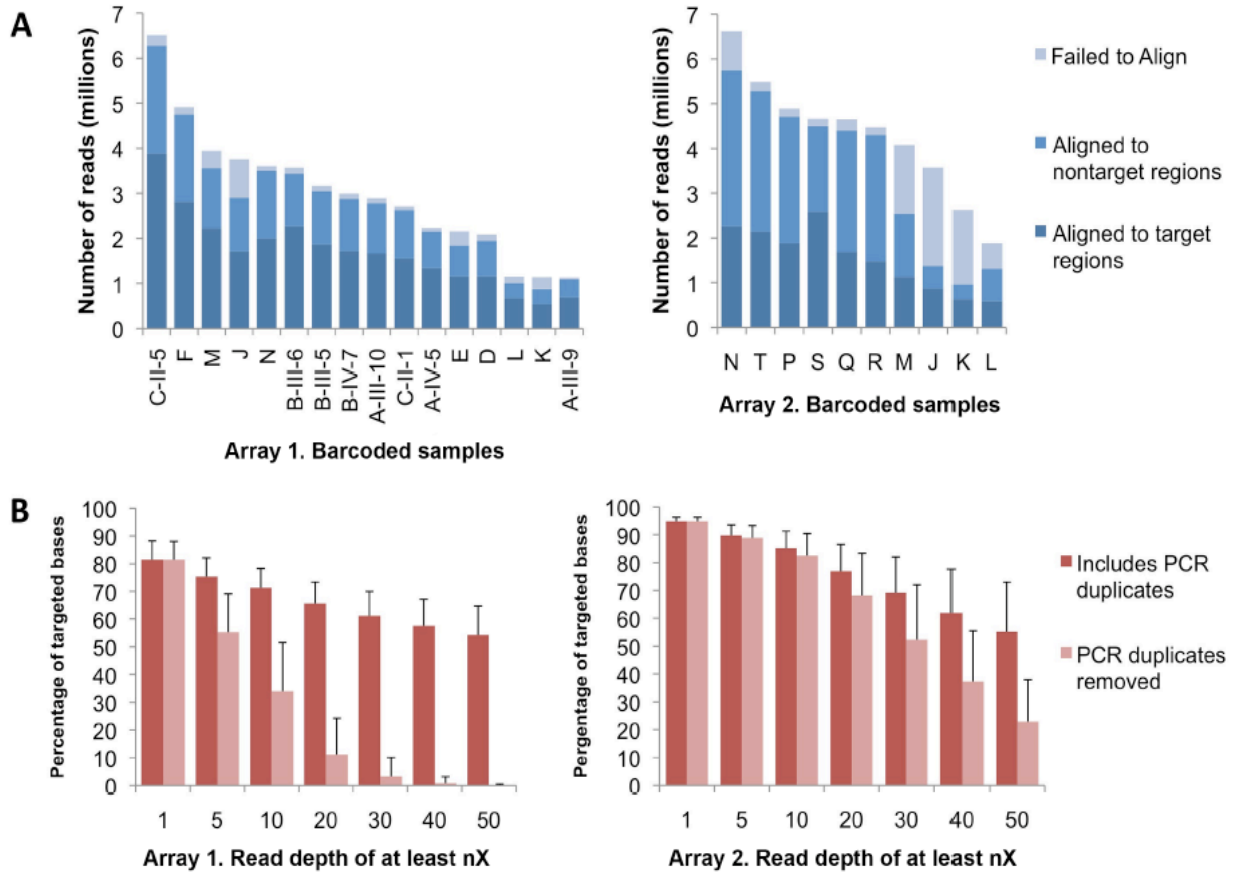
Figure 2.5. Specificity of target capture and read depth of targeted bases. (A) Graph depicting the number of filtered Illumina GA II reads per barcoded sample that aligned to the targeted region (dark blue), aligned elsewhere in the human genome (medium blue), or failed to align to the reference genome (light blue), for both the first (left) and second (right) capture array experiment. Each letter refers to a family and if more than one member of the family was sequenced then the specific family member is indicated. For each array, of the 50 million reads obtained from two lanes of GA II sequencing, between 1 and 6 million reads were obtained for each individual. The percentage of each individual's reads that mapped to the targeted region was ~60% for the first array and ~40% for the second array. (B) Percentage of bases targeted by the array having read depths greater than or equal to a specific fold-coverage. Graph depicting the mean of the percent of targeted bases captured in each individual at greater than or equal to 1, 5, 10, 20, 30, 40, and 50-fold coverage (nX). Values are shown before (dark red) and after (light red) removal of PCR duplicates. Error bars indicate 1 SD.

the reference genome (hg19). We obtained between 1 and 6 million reads for each individual. Among individuals, an average of 61% (±2%) of the aligned reads mapped to regions targeted by the capture array (Figure 2.5A). Of the bases targeted by the capture array, 75% (±7%) had a read depth of at least 5x, which diminished to 55% (±14%) after the removal of PCR duplicates (Figure 2.5B). Twenty percent of targeted bases were not captured. In the 3 families for which pairs of affected family members were sequenced, total filtered variants in the candidate interval (388–1499 per individual) were analyzed to find variants shared by both affected individuals from the same family (Table 2.2). In all 3 families, frameshift mutations in exon 4 of *PTPN11* were the only novel coding variants present in both affected family members and, for Families A and B, absent in the unaffected individual. Family A had a 5 bp deletion, Family B had a more complex deletion/insertion, and Family C had a 2 bp deletion (Table 2.1). In the remaining 8 families for which only 1 affected individual per family was sequenced, there were 18 novel coding variants present in ≥3 reads, one of which was a nonsense mutation in exon 13 of *PTPN11* (p.Q506X) (Figure 2.6). We used Sanger sequence analysis of PCR amplimers to demonstrate that affected family members from these 4 families had *PTPN11* mutations, and that unaffected family members lacked *PTPN11* mutations. Sanger sequence analysis of the 15 coding exons of *PTPN11* in the 7 families for whom we had not found mutations by array capture and Illumina-sequencing detected a 1 bp deletion in exon 11 in 1 of the 7 families (Family D). This deletion was within a 98 bp segment that had been targeted but not captured in any of the DNA samples. Another family (F) had a splice-acceptor site mutation (AG>CG) in intron 5 in 2 affected siblings, but not in either parent. The siblings' mother was clinically affected with MC, although less severely than her children. The mother was the first in the family to have MC and was the only member of the family who was included in the Illumina sequencing. The site of the splice-site mutation identified in her children was covered 256 in her DNA sequence and was always wild-type, as were her Sanger sequence results. These data suggest the mother is mosaic for a *PTPN11* mutation and that the family's mutation would have been found by Illumina sequencing had we initially sequenced her children's DNA

We subsequently collected DNA from an additional 6 MC families. Sanger sequence analysis revealed a nonsense mutation involving exon 3 (p.K99X) in one family (I), and a splice site mutation in intron 9 (c.1093-1G>T) in another family (G). In total, we found *PTPN11* mutations in 10 of 17 families.
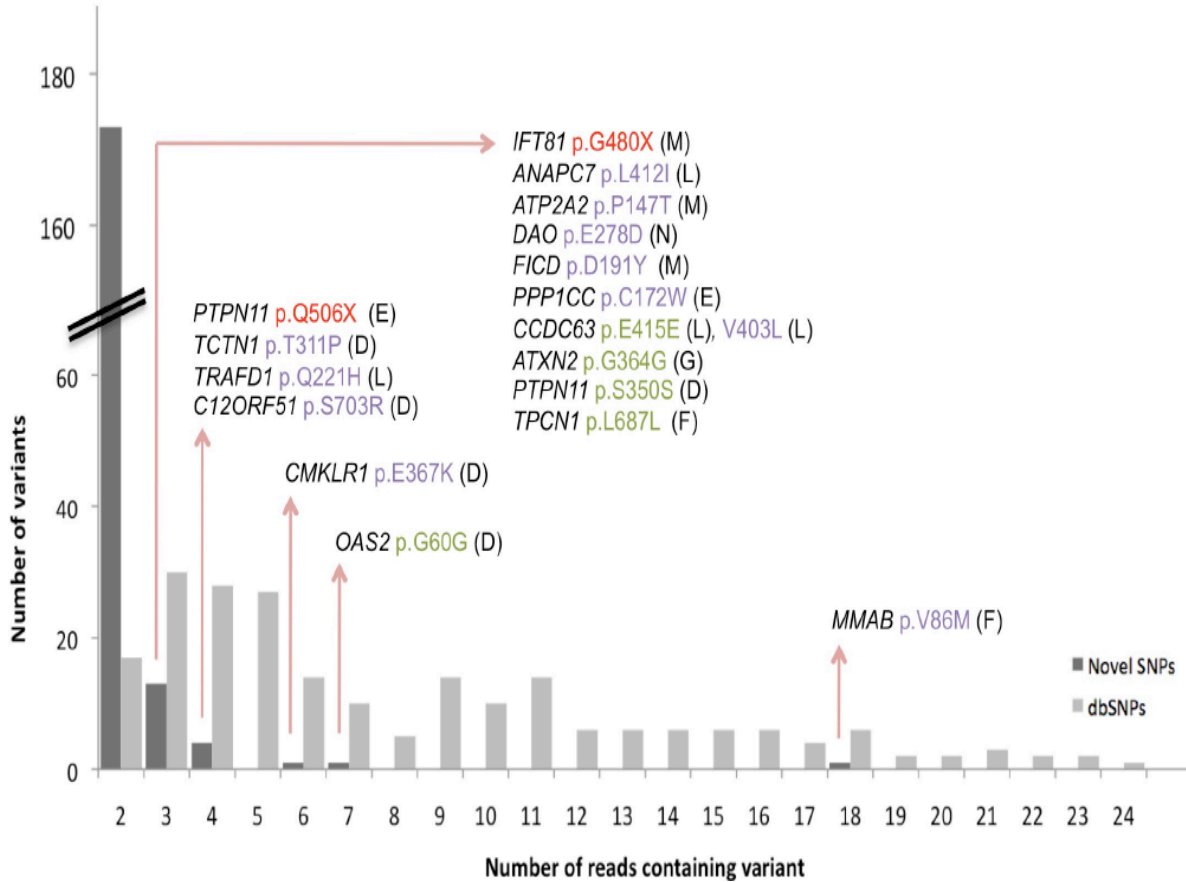
Figure 2.6. Identified coding variants in unrelated individuals from the first capture array. Total coding variants identified in the 8 unrelated individuals who were the only members of their families included in the first array capture experiment. The graph depicts the number of variants and the number of reads containing each variant in an individual. Novel (dark gray) and previously characterized (light gray) SNPs are separated into different bars. Note that the ratio between novel and known SNPs decreases as the number of reads containing that SNP increases. These data suggest that most novel SNPs seen in 3 or fewer reads represent sequence artifacts rather than real variants. Shown above the graph are novel variants that were observed in three or more reads in single individuals, with the gene names and predicted amino acid changes indicated. Nonsense mutations are indicated in red, nonsynonomous mutations are indicated in purple, and synonomous mutations are indicated in green. No insertion or deletion variants were seen in more than 2 reads. The family in which the novel variant was identified is indicated in parentheses. Two mutations were identified in *PTPN11*, including a nonsense mutation and a synonomous mutation. The nonsense mutation was confirmed by Sanger sequence analysis; however, the synonomous mutation could not be confirmed and therefore likely represents a false positive result.
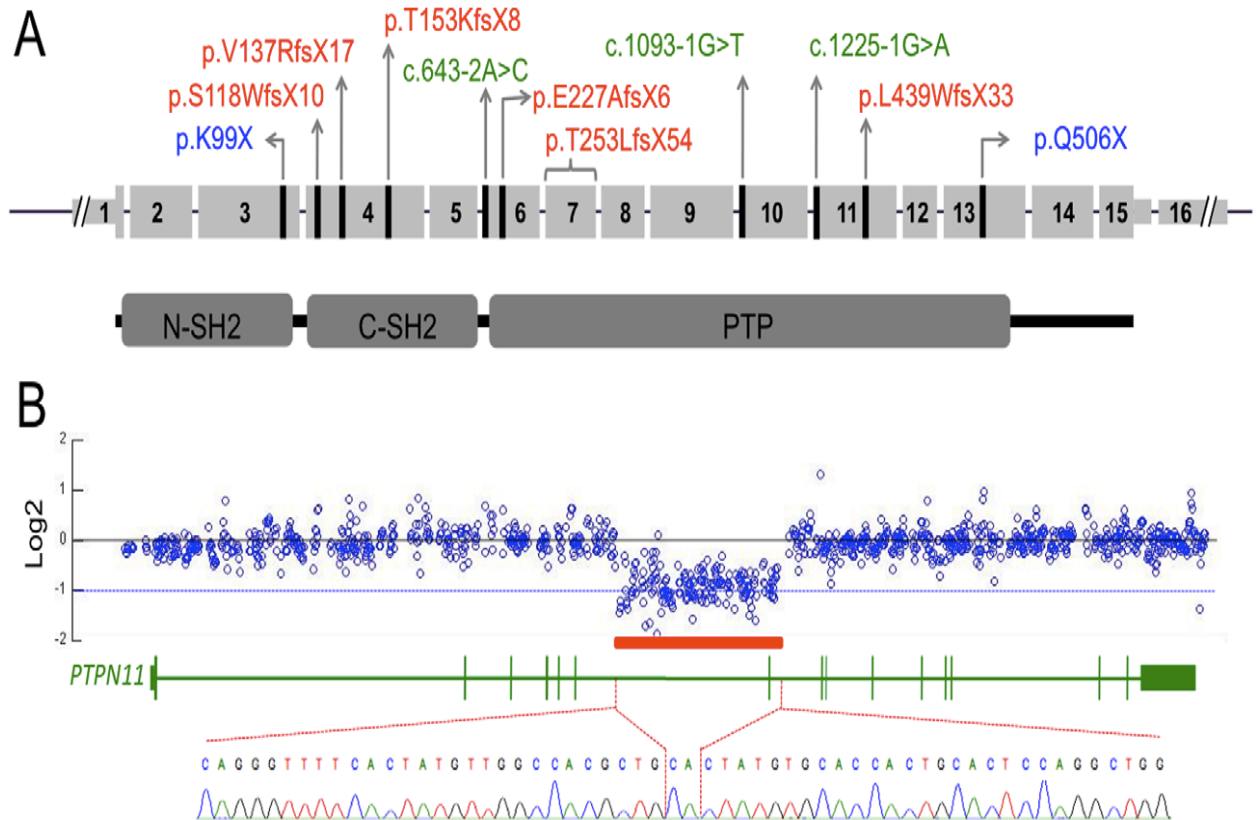
Figure 2.7. *PTPN11* mutations identified in MC participants. (A) Schematic of the exonic structure of *PTPN11* (above) and the corresponding protein structure of SHP2 (below). The locations of mutations that were identified in MC are indicated with black lines. Predicted protein changes are indicated for the nonsense (blue) and frameshift (red) mutations, while the cDNA designation is indicated for the splice-site mutations (green). (B) Log2 values of the number of Illumina reads obtained per 50 bp window in participant S, divided by the average number of reads obtained in other participants whose DNA was captured simultaneously using the second capture array. Shown are all 50 bp windows spanning regions of *PTPN11* targeted by the array, with the corresponding exonic structure of *PTPN11* shown below. The red bar indicates a region spanning exon 7, in which the average log2 value is approximately −1, suggesting a heterozygous deletion. PCR amplification and sequencing of the breakpoint, using primers on either end of the deletion, indicate that 14,629 bp of sequence have been deleted and replaced with a single CA dinucleotide.

Five mutations were frameshift, 2 were nonsense, and 3 disrupted a splice-acceptor site (Figure 2.7A, Table 2.1). Each family had a different mutation and mutations were scattered across the gene (Figure 2.7A). In two families without mutations we had performed aCGH and did not find evidence of *PTPN11* intragenic deletions or duplications (Table 2.1). We did not have other family members' DNA samples from the one familial MC patient who lacked a *PTPN11* mutation to be able to test for locus heterogeneity by linkage analysis.

Multiplexed targeted genomic capture and sequencing of *PTPN11* and associated genes

To test whether the patients without *PTPN11* coding mutations had noncoding mutations in *PTPN11* or had mutations in other genes, we designed a second Agilent 1M capture array. Firstly, we included probes to target the entire *PTPN11* gene, excluding Alu repeats. Secondly, we included probes targeting the exons of 74 genes that function in the same pathways as *PTPN11*, including the Ras/MAPK and PI3K/Akt pathways (Table 2.3). Thirdly, we included probes targeting the exons of the MHE genes *EXT1* and *EXT2*, to determine whether any of our patients lacking mutations and classic radiographic features of MC had been misdiagnosed. Barcoded genomic libraries for an individual from each of the 7 MC families without *PTPN11* coding mutations and from 2 families originally referred with MC, but whose radiographic features were more consistent with MHE, were pooled and hybridized to the capture array. The captured DNA was then sequenced using two lanes of Illumina GAII 42 bp single end sequencing. For each barcoded sample, 4.3 (±1.2) million reads were obtained, of which 37% (±7%) mapped to regions targeted by the array (Figure 2.5A). Of the bases targeted by the array, 85% (±6%) were covered by a read depth of at least 10x, which dropped to 83% (±8%) after the removal of PCR duplicate reads (Figure 2.5B). Identified variants with a quality score >20 were filtered to remove SNPs listed in the SNP

Table 2.3. Genes included in second capture array

Category	Gene
Ras/MAPK pathway	*KRAS, HRAS, NRAS, ARAF, BRAF, RAF1, MAPK1, MAPK3, MAP2K1, MAP2K2, SOS1, SOS2, GRB2, SHOC2, GAB1, GAB2, GAB3, FRS2, IRS1, SHC1, SIRPA, PXN, MVP, MPZL1, CRK*
PI3K/AKT pathway	*AKT1, AKT2, AKT3, PIK3R1, PIK3R2, PIK3R3, PIK3CA, PIK3CB, PIK3C2A, PIK3C2B, PIK3C2G, MTOR, RPS6KB1, PDPK1, ILK, PRKCD, PRKCZ, PTEN*
Receptor Tyrosine Kinases	*FGFR1, FGFR2, FGFR3, EGFR, PDGFRA, PDGFRB, IGF1R, ERBB2, ERBB3, ERBB4, ROR1, ROR2, MUSK, AXL, MERTK, DDR1, DDR2*
Other	*EXT1, EXT2, RHOA, CDC42, RAC1, PTH1R, SRC, PTK2, YES1, FYN, LYN, FGR, PLCG1, PLCG2, HSPA4, INPP5D*

database (version 132) and the 1000 genomes project (Nov. 2010 release). No *PTPN11* coding

mutations were found. We then analyzed the noncoding regions and identified one 3' UTR mutation, and

7 intronic mutations, of which 6 were in LINE elements or other repetitive regions (Table 2.4). All intronic

mutations were at least 300 bp from the nearest exon and, using an online splice prediction tool

(http://genes.mit.edu/GENSCAN.html), were not predicted to alter splicing. We then analyzed variants

identified in the exons of the 76 other genes included in the capture array (Table 2.3). No nonsense,

frameshift or splice site mutations were identified. Of the missense mutations that were present in more

than 4 independent sequencing reads, 4 were nonsynonomous (*ERBB2* p.S1050L, *MTOR* p.P1408S,

MVP p.R49S, *SOS2* p.D952N) and 4 were synonomous (*PIK3C2B* p.D478D, *RAF1* p.L351L, *MAP2K2*

p.D140D, *MVP* p.T199T). Further experiments will be needed to determine if any of the novel noncoding

PTPN11 mutations or novel variants in the other genes are disease causing.

We next analyzed the sequencing read depth across the *PTPN11* locus to detect deletions or

duplications. In one individual (Patient S), we identified a ~15 kb region spanning exon 7 that contained

half as many reads as would be expected based upon the read depths of the other patients included in

the capture array (Figure 2.7B). As expected, PCR primers that flank this 15 kb region failed to produce

amplimers when wild-type genomic DNA was used as template. However, PCR amplification using

genomic DNA from Patient S yielded a 700 bp PCR product and Sanger sequence analysis of this

product indicated that 14,629 bp of genomic sequence (chr12:112,897,487–112,912,115) had been

Table 2.4. Mutations identified in second capture array

Location	Patient	Mutation
PTPN11		**Chr12 genomic change**
3' UTR		
	M	g.112945402G>A
Intronic		
	J	g.112857243G>T, g.112869736del1, g.112924937C>T
	M	g.112862223C>T, g.112914741C>G, g.112933866A>G, g.112945402G>A
Coding mutations		**Gene name and predicted protein change**
	J	*MTOR* p.P1408S, *MVP* p.R49S
	K	*ERBB2* p.S1050L, *MAP2K2* p.D140D
	L	*MVP* p.T199T
	M	*PIK3C2B* p.D478D
	P	*SOS2* p.D952N, *RAF1* p.L351L
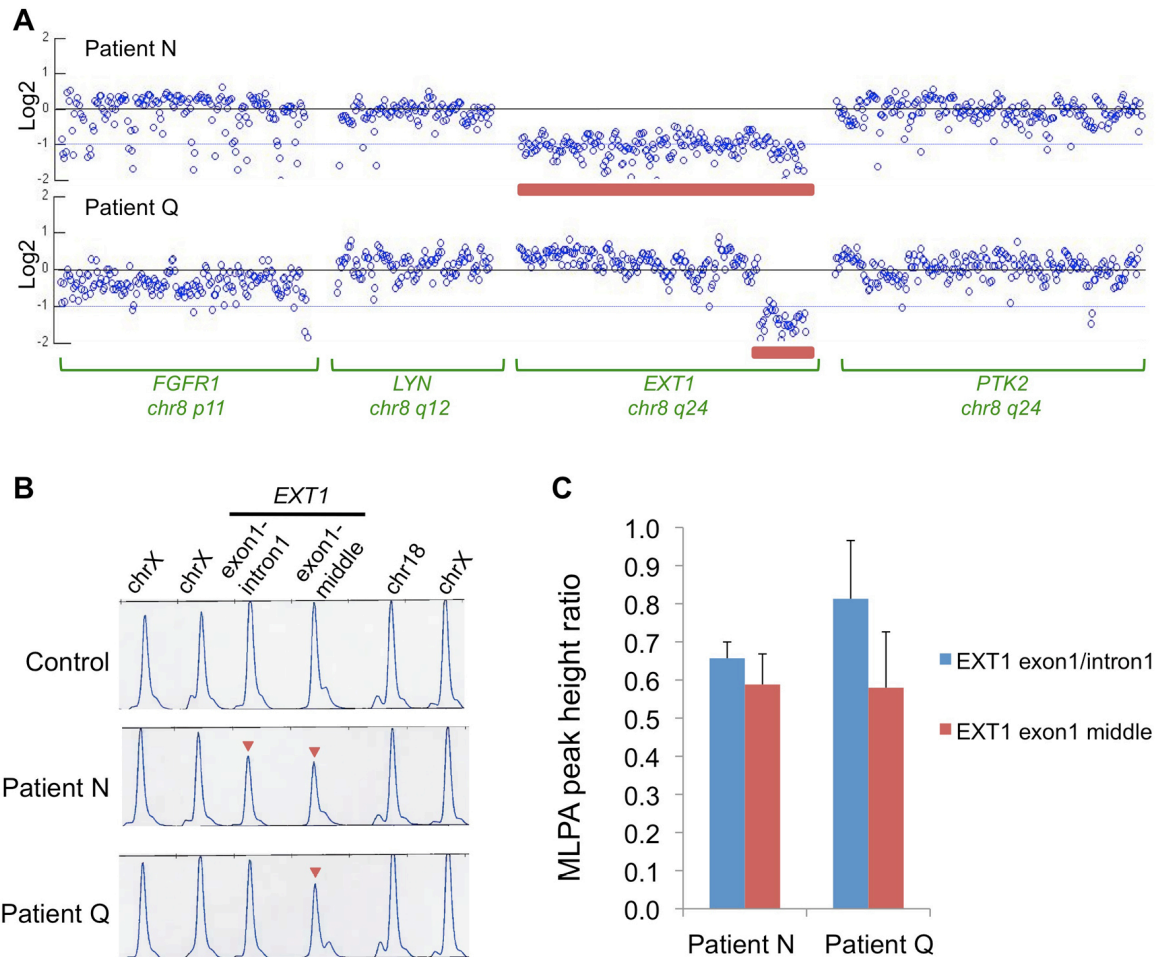
Figure 2.8. Heterozygous deletions of *EXT1* identified in two participants. (A) Log2 values of the number of Illumina reads obtained per 50 bp window in participants N and Q, divided by the average number of reads obtained in other participants whose DNA was captured simultaneously using the second capture array. Shown are all 50 bp windows spanning the exons of four genes (*FGFR1, LYN, EXT1, PTK2*) located on chromosome 8 that were targeted by the capture array. The red bars indicate a region spanning all *EXT1* exons in participant N, and exon 1 of *EXT1* in participant Q, where consecutive windows have log 2 values of approximately −1, suggesting heterozygous deletions of these regions. (B) MLPA amplification products separated by electrophoresis. The amplification products of the MLPA probes targeting both the middle and the boundary of *EXT1* exon 1, and the MLPA probes targeting the middle of *EXT1* exon 1, have reduced peak heights in participants N and Q respectively (arrowheads). (C) Ratio of each *EXT1* peak height to the average height of the four control peaks. The ratios were normalized based on the ratios obtained in control individuals, such that a ratio of 1.0 indicates a copy number of 2, while a ratio less than the threshold of 0.8 suggests a copy number of 1. In both participants N and Q, the ratio for the probe corresponding to the middle of *EXT1* exon 1 is below 0.8, suggesting a heterozygous deletion of this exon. MLPA was performed twice for each participant. Error bars indicate standard deviation.

replaced with a single CA dinucleotide (Figure 2.7B). In addition, PCR amplification and sequencing of *PTPN11* in peripheral blood cDNA from this patient, using a forward primer in exon 6 and a reverse primer in exon 8, detected a mutant cDNA that lacked exon 7 (data not shown). The loss of exon 7 results in a frameshift with introduction of a premature stop codon (T253LfsX54). The read depth of the remaining 76 genes targeted by the array was also analyzed to detect deletions or duplications. Two patients initially included in the study, but on radiographic review were felt more likely to have MHE than MC, were found to have deletions involving *EXT1* (Figure 2.8). In one patient (Q), the first exon of *EXT1* contained half as many reads over its 1.8 kb as expected (Figure 2.8A). In a second patient (N), all exons of *EXT1* had half as many reads as expected (Figure 2.8A). In additional to skeletal lesions, this patient has developmental delay, microcephaly and mild dysmorphism, suggesting a possible contiguous gene deletion syndrome. *EXT1* deletions were confirmed in both patients by multiplex ligation dependent probe amplification (MLPA) (Figure 2.8B,C).

PTPN11 loss-of-function mutations in metachondromatosis

Our finding of nonsense, frameshift, and splice-site mutations in multiple exons, as well as a large deletion, suggests that MC causing *PTPN11* alleles are loss-of-function. We tested this hypothesis by performing Western blots on whole protein extracts from white blood cells and from an excised exostotic lesion in a patient (B-IV-7) with a *PTPN11* frameshift mutation in exon 4. An anti-SHP2 antibody that recognizes an epitope amino-terminal of the polypeptide encoded by the frameshifted exon detected only full-length, wild-type SHP2 protein (Figure 2.9).

Loss of *PTPN11* wild-type alleles in the cartilage cores of MC exostoses

We next determined whether MC exostoses arise from a ''second hit'', similar to what has been observed in autosomal dominant MHE [18]. We looked for a second hit in cells of the cartilage core of an MC lesion (e.g., Figure 2.1K) by performing microdissection, PCR amplifying the mutation containing exon, and Sanger sequencing the amplimers. In tumors from two different patients (A-IV-5, A-IV-8), with a 5 bp frameshift mutation in exon 4, we observed a clear excess of mutant sequence versus wild-type sequence in the tumors' cartilage cores, as compared to the patients' peripheral blood and bone/marrow from the lesion (Figure 2.10A). We quantified the amount of mutant versus wild-type sequence, by extracting DNA from the cartilage core of patient A-IV-8, PCR amplifying exon 4, and subcloning
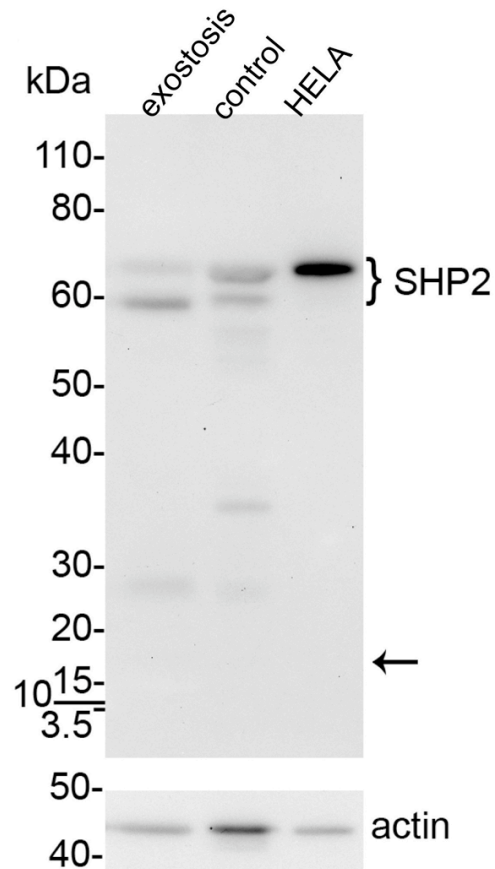
Figure 2.9. Truncated SHP2 protein is not detected. Immunoblots of protein extracted from an exostosis from participant B-IV-7 with a truncating frameshift mutation, rib cartilage from an unaffected individual (cartilage expression tissue control), and HELA cells (antibody control) that have been separated on a 4–12% SDS-PAGE gel and transferred to Immobilon-P. Upper panel: immunodetection using mouse-anti-SHP2 that detects an epitope N-terminal of the frameshift mutation. Brackets indicate wild-type SHP2 that is differentially phosphorylated. Wild-type protein is present in the exostosis, likely from the bone marrow and fibrous elements since the lesional cartilage was not selectively isolated. Note absence of a unique band 17 kD in size, that would have been expected if truncated protein was produced (arrow). Lower panel: immunodetection of the same blot using a mouse-anti-actin antibody as a loading control.
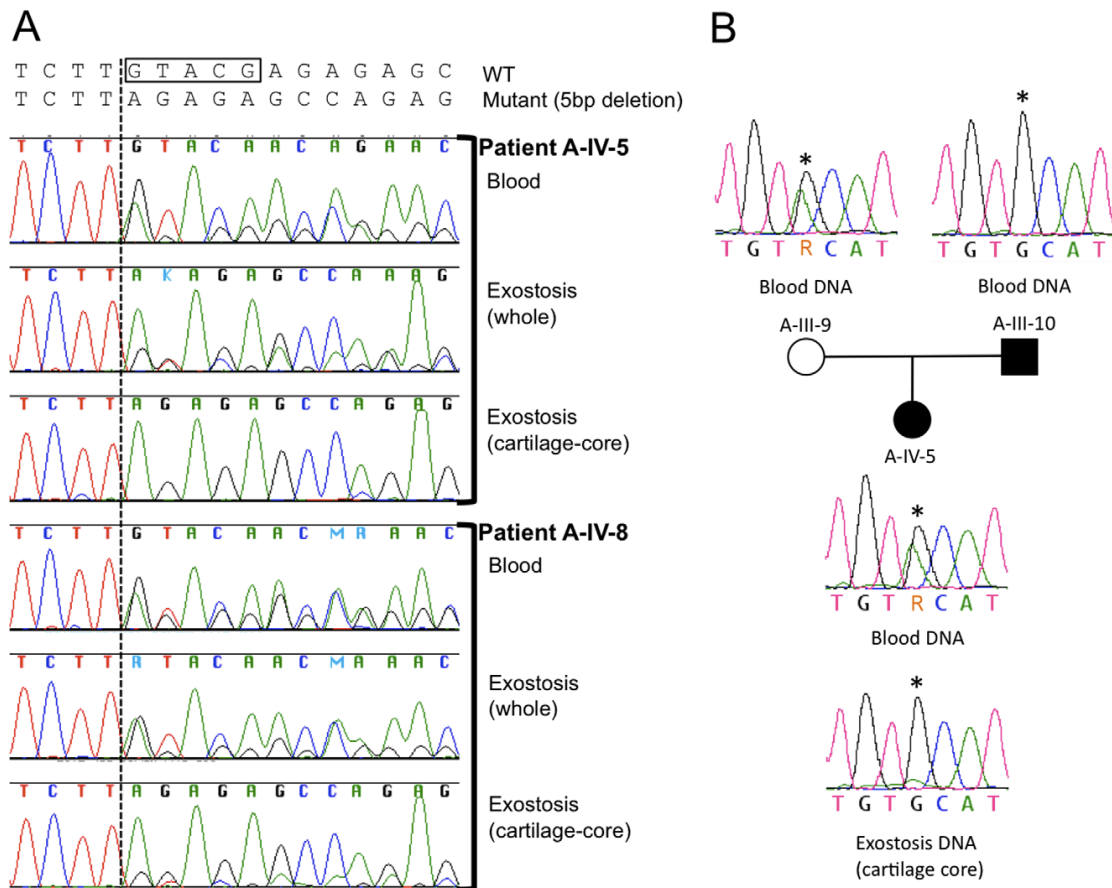
Figure 2.10. Loss of the wild-type *PTPN11* allele in the cartilage of two exostoses. (A) Electropherograms of PCR amplified template DNA that had been extracted from whole blood, a section of an exostosis, or the cartilage core of the same exostosis. Exostoses were available from patients A-IV-5 and A-IV-8. The site of the 5 bp deletion in exon 4 of *PTPN11* in both patients is indicated with a box. Note that that the heights of the peaks corresponding to the mutant sequence are markedly reduced in amplimers from the cartilage-core compared to amplimers from blood or from a section that contains cartilage, bone and fibrous tissue. This is consistent with loss-of-heterozygosity in the cartilage component of the exostoses. (B) Electropherograms of PCR amplified template DNA extracted from blood from participants A-III-9, A-III-10 and A-IV-5, as well as DNA extracted from the cartilage core of the exostosis from participant A-IV-5 shown in (A). Blood DNA electropherograms indicate that participants A-III-9 and A-IV-10 are heterozygous at a position (asterisk) in intron 11 of *PTPN11*. This is the site of a known common polymorphism (rs41279092). Exostosis cartilage DNA electropherograms have a reduced adenine peak height at this position. This suggests that the wild-type *PTPN11* allele inherited from the unaffected parent (A-III-9), which carries an adenine at this position, has been lost in cells that contribute to formation of the exostosis' cartilage core.

amplimers to determine the percent that contained the mutant allele. Forty-four of 52 individual subclones contained the mutant allele, which is significantly higher (p<0.001) than expected for a heterozygous mutation. In contrast, 58% of subclones (34/59) from adjacent unaffected bone/bone marrow contained the mutant allele, which is not significantly different from the expected value of 50% (p=0.24). These data are consistent with an MC exostosis arising from a second hit (loss of the wild-type allele) within a cell that ultimately contributes to the lesion's cartilage core. Because the mutant allele is 5 bp shorter than wild-type *PTPN11* in these two patients, we tested for loss of heterozygosity at a second polymorphic site in *PTPN11* to control for potential PCR bias in amplifying the exon with the deletion. In their peripheral blood DNA, patient A-IV-5 and her unaffected mother are heterozygous for a benign polymorphism in intron 11 of *PTPN11* (rs41279092). The abundance of this SNP, which is in the wild-type *PTPN11* allele, was markedly reduced in the lesion's cartilage core, again consistent with LOH occurring in the cell that drives formation of the cartilage core (Figure 2.10B).

PTPN11 mutations are not found in other cartilaginous tumor syndromes

We finally asked whether mutations in *PTPN11* are associated with other cartilaginous tumor syndromes. We sequenced the coding exons of *PTPN11* in 38 lesions excised from patients with Ollier disease, 2 peripheral blood samples from patients with Ollier disease, 15 lesions excised from patients with Maffucci syndrome, 4 solitary enchondromas, 9 chondrosarcomas (1 polyostotic), and 3 osteochondromas without *EXT1* or *EXT2* mutations. We did not find *PTPN11* coding sequence mutations in any patient sample. In 24 percent of the samples we observed heterozygosity for noncoding SNPs that are known common variants, suggesting that large *PTPN11* gene deletions and other causes of LOH are not frequently associated with these other cartilaginous tumors.

2.6   Discussion

We identified 17 unrelated families with MC. Clinical features were similar to previously published cases [5–10,15]. The exostoses of MC had been assumed to be identical to the osteochondromas of MHE; however, we demonstrate that they are histologically unique lesions with a large cartilaginous core (Figure 2.1J–1L). We combined linkage analysis in a single MC family with DNA capture and parallel sequencing of bar-coded DNAs from several MC families to identify mutations in *PTPN11* as a cause of MC. In MC patients without *PTPN11* coding sequence and splice site mutations, we generated and

pooled barcoded DNAs, and performed a second targeted capture that included the entire *PTPN11* gene and exons from 76 other genes. This led us to detect a ~15 kb deletion in a patient by analyzing the depth of sequencing reads. In total, we found likely disease-causing *PTPN11* mutations in 11 of 17 families.

Concurrent with our studies of MC, Sobreira *et al.* (2010) reported *PTPN11* mutations in 2 MC families [11]. They performed whole-genome sequencing (WGS) in a single affected individual who was a member of family in which MC was segregating. This approach also required these investigators to include linkage data to reduce the number of novel potentially disease-causing heterozygous changes that are identified by WGS [19,20]. The investigators next identified an independently arising *PTPN11* mutation in an unrelated patient to strengthen the evidence for causality. However, having only studied genomic DNA and finding frameshift mutations in the same exon (exon 4) in their two unrelated patients, Sobreira *et al.* (2010) could not definitively determine the mechanism by which *PTPN11* mutations cause MC.

Missense mutations in *PTPN11* have previously been identified in patients with Noonan, Noonan-like, and LEOPARD syndromes, as well as in juvenile myelomonocytic leukemia [21]. In these disorders, the mutations are gain-of-function and/or dominant negative for SHP2, which is the *PTPN11* protein product [22,23]. SHP2 is a protein tyrosine phosphatase and an important intracellular signaling molecule linking several growth factor receptors to the Ras/MAPK and other signaling pathways [24]. Therefore, frameshift mutations in exon 4 might also create an abnormal protein product by altering *PTPN11* mRNA splicing. Alternatively, the frameshift mutations might result in loss-of-function because of nonsense mediated mRNA decay or rapid degradation of a truncated SHP2 polypeptide. Our finding of nonsense, frameshift, and splice-site mutations in multiple exons, as well as a whole-exon deletion, suggests that MC-causing *PTPN11* alleles are loss-of-function. We tested this hypothesis by performing Western blots on whole protein extracts from white blood cells and from an excised lesion containing affected and unaffected tissue, and detected only fulllength, wild-type SHP2 protein (Figure 2.9), confirming that the mutant alleles are loss-of-function.

Exostoses in MHE originate from the "second hit" mutations [18]. Mice with conditional alleles at the *EXT1* locus demonstrate that only a few cells having two mutant alleles are sufficient to cause other cells to become misdirected and form an exostosis [25]. By performing microdissection, we found

evidence for loss of the wild-type *PTPN11* alleles in the majority of cells within the cartilage cores of exostoses from two MC patients (Figure 2.10), consistent with a "second hit." Recently, Bauler *et al.* used a ubiquitously expressed Ert2-Cre driver in 6–8 week-old *Ptpn11* floxed mice to generate mice that were *Ptpn11*-null in multiple tissues. Among the consequences of completely deleting SHP2 was the appearance of ectopic cartilage islands in the animals' metaphyseal trabecular bone and growth plates [26]. These findings are consistent with the distribution of endosteal tumors and exostoses seen in patients with MC. The findings in mice with homozygous deletion of *Ptpn11* contrast with the absence of skeletal lesions in mice that have heterozygous loss-of-function mutations [27]. We suspect that mice with heterozygous mutations have a much lower incidence of noticeable "second hits" compared to humans because they have fewer skeletal cells and shorter lifespans. Homozygous inactivation of *Ptpn11* solely in mouse chondrocytes may be required to enable a detailed understanding of how SHP2 deficiency leads to tumorigenesis.

We did not detect *PTPN11* mutations in 6 of 17 patients with MC phenotypes, including 1 patient with a family history of MC and 5 patients who are the first affected members in their families. DNA is not available from other affected family members of the familial case to determine whether MC exhibits locus heterogeneity. Two patients with de novo disease did have DNA variants found in the 3' UTR and/or in introns. None of these variants are in likely regulatory regions or in regions important for mRNA splicing; however, we cannot conclude they are benign. Furthermore, we cannot exclude the possibility that patients with de novo disease are somatic cell mosaics for *PTPN11* mutations that are not present in white blood cell DNA, similar to the mildly affected mother in Family F who had two affected children. Despite these caveats, MC could be locus heterogeneous, similar to Noonan syndrome, which can be caused by mutations in *PTPN11* or in other components of the Ras/MAPK pathway [28]; however, our targeted capture and sequencing of 74 genes that included most of the Ras/MAPK and PI3K/Akt signaling pathways did not find an obvious mutation in another gene in any of the 6 *PTPN11* mutation-negative MC patients.

We found no evidence of *PTPN11* coding mutations in other cartilage tumor syndromes, including Ollier disease and Mafucci syndrome. Although sequencing was performed on lesional tissue rather than whole blood, it is possible that we may have missed causative mutations that are present in only a subset

of cells within the *lesion*. We may also have missed mutations in the 5' and 3' untranslated regions of *PTPN11* contained within exons 1, 15, and 16, that we did not sequence in these patients. Based on our finding heterozygosity for noncoding SNPs in many of these samples, it is unlikely that large *PTPN11* gene deletions or other causes of LOH are common in these syndromes. Despite the aforementioned limitations of our mutation detection method, our data are consistent with the separation of MC from the other cartilage tumor syndromes based on clinical and pathologic features.

In conclusion, we combined linkage analysis in a single family with DNA capture and parallel sequencing of barcoded DNAs from several families to identify mutations in *PTPN11* as a cause of MC. The advantages of this approach are its ability to identify a region of interest, then simultaneously sequence affected individuals from multiple unrelated families, and then focus on genes for which novel SNPs or other mutations are seen in more than one family, all at reasonable cost (<$10,000 in consumables). In patients with MC and *PTPN11* mutations, we conclude that the mutations are loss-of-function since the mutant protein is not expressed, and that the loss of the remaining wild-type allele via a ''second hit'' is responsible for the formation of the exostoses. Since we did not detect *PTPN11* mutations in all MC families, MC may be locus heterogeneous, although we have not found evidence after sequencing more than 70 genes that function in related pathways. Finally, precisely how mutations in *PTPN11* give rise to the exostoses and endosteal tumors in patients with MC is not yet known. However, this question can now be addressed since mice with alleles of *Ptpn11* that can be conditionally inactivated in temporal and site-specific manner are available [26,29].

2.7   Materials and Methods

Ethics

Informed consent was obtained through a Children's Hospital Boston IRB approved protocol. Specimens and/or DNA received from external institutions were collected under IRB approved protocols at host institutions and received coded without identifying information.

Linkage analysis

Raw genotype data were generated for multiple members of family A using Affymetrix 6.0 SNP arrays, and genotypes were called using Affymetrix Genotyping Console with the Birdseed v2 algorithm

and a confidence threshold of 0.02. SNPs with 100% sample call rate or pedigree minor allele frequency of 0 were removed, then multipoint parametric linkage analysis of the remaining 421,922 autosomal and 17,169 X-linked SNPs was performed using MERLIN and its derivative MINX, respectively, with Affymetrix Caucasian allele frequencies and deCODE Genetics genetic map positions. The disease allele frequency was estimated at 1E-7, and phenocopies and non-penetrance were not permitted (affectation probability 0/1/1). Because non-penetrance and non-ascertainment are potential confounding factors in MC, we analyzed only founders and affected individuals.

Barcoded genomic libraries

To generate genomic libraries for each individual, 2 µg of genomic DNA were first sheared to 100 bp–200 bp using Adaptive Focused Acoustics following the manufacturer's protocol (Covaris, Inc). Blunt-ended fragments were generated using an End-it DNA End-Repair kit (Epicenter), purified using Agencourt AMPure XP magnetic beads (Beckman Coulter), and eluted in 10 mM Tris Acetate, pH 8.0. The fragments were A-tailed using the Klenow fragment (NEB), purified, eluted in 1x Quick Ligase Buffer (Quick Ligation kit, NEB), and incubated with Quick T4 DNA ligase and 100 µM barcoded-adapters (Table 2.5) to create a library of adapter-ligated fragments. A different barcoded adapter was used for each genomic DNA library. Each library was again purified using Agencourt AMPure XP magnetic beads and eluted in 40 µl of 10 mM Tris Acetate, pH 8.0. Libraries used for hybridization to the first capture array were amplified according to two strategies: 3 µl amplified for 18 cycles in four 50 ml PCR reactions (Phusion High-Fidelity DNA polymerase, Finnzymes), or 2 µl amplified for 11 cycles in one 50 ml PCR reaction that was then purified and amplified for 17 cycles in ten 50 µl PCR reactions (FastStart Taq DNA polymerase, Roche). For the libraries used for hybridization to the second capture array, 13 µl was amplified for 15 cycles in ten 50 µl PCR reactions (Phusion High-Fidelity DNA polymerase, Finnzymes). Primers are provided in Table 2.6. Sizes of amplified libraries were confirmed to be between 200–300 bp necessary for Illumina GA II sequencing prior to hybridization (Figure 2.4).

Capture array design

To enrich regions of interest in the linked interval for sequencing, we used an Agilent Technologies 1M SureSelect DNA capture array. Target regions were defined using the UCSC Genome Browser and included: the union of exons from multiple GRCh37/hg19 gene, mRNA, and Alt Events

Table 2.5. Barcoded library adapters

AAC	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT**AAC***T-3'
	5'-/5Phos/**GTT**AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'
AGC	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT**AGC***T-3'
	5'-/5Phos/**GCT**AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'
AGG	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT**AGG***T-3'
	5'-/5Phos/**CCT**AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'
CAG	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT**CAG***T-3'
	5'-/5Phos/**CTG**AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'
CGA	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT**CGA***T-3'
	5'-/5Phos/**TCG**AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'
CGT	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT**CGT***T-3'
	5'-/5Phos/**ACG**AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'
CTC	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT**CTC***T-3'
	5'-/5Phos/**GAG**AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'
GAC	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT**GAC***T-3'
	5'-/5Phos/**GTC**AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'
GCA	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT**GCA***T-3'
	5'-/5Phos/**TGC**AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'
GGA	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT**GGA***T-3'
	5'-/5Phos/**TCC**AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'
GTC	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT**GTC***T-3'
	5'-/5Phos/**GAC**AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'
GTG	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT**GTG***T-3'
	5'-/5Phos/**CAC**AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'
TCC	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT**TCC***T-3'
	5'-/5Phos/**GGA**AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'
TCG	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT**TCG***T-3'
	5'-/5Phos/**CGA**AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'
TGA	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT**TGA***T-3'
	5'-/5Phos/**TCA**AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'
TGG	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT**TGG***T-3'
	5'-/5Phos/**CCA**AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'

Asterisk (*) indicates a phosphorothioate bond to prevent degradation. Adapter sequences are Illumina paired end adapters with a 3bp identifying barcode indicated in bold type.

tracks; 30 bp of proximal and distal intronic flanking sequence; and 1000 bp of upstream promoter sequence. Targets were padded with 60 bp of additional proximal and distal flanking sequence to promote uniform capture coverage, for a total size of 1,187,477 bp. Probes were designed against NCBI36/hg18 using the Agilent eArray software (https://earray.chem.agilent.com/earray/) and translated coordinates, with 60-nt length, 3-nt spacing, and repetitive elements masked. The resulting 243,488 probes spanned 844,339 bp (GRCh37/hg19), and included 71.1%, 72.2%, 88.4%, and 98.6% of the padded target, unpadded target, UCSC exons, and CCDS coding sequence, respectively. The probes and their reverse complements were each applied in duplicate to the capture array for a total of 973,952 probes. For the second 1M SureSelect DNA capture array, Biomart (http://uswest.ensembl.org/biomart/) was used to obtain the Ensembl NCBI37/hg19 coordinates for the exons of 76 genes (Table 2.3). Exons were padded with 90 bp to define a 718,566 bp target region. eArray was used to design 568,634 probes to target the repeat masked sequences of this region (91%). The target region for *PTPN11* was defined as 93,180 bp spanning 1 kb upstream to 2 kb downstream of the gene. Repeat masker (http://www.repeatmasker.org/) was used to mask only Alu repeats (37% of the region) resulting in a target region of 61,365 bp, for which 55,205 probes were designed using eArray. All probes were 60-nt in length and spaced every 1-nt. The capture array was designed to include all probes (623,839 total), as well as the reverse complement of every 2nd probe and every 17th probe, for a total of 972,455 probes.

Array hybridization

For the first capture array, 1.4 μg of each of the 16 amplified libraries was pooled and hybridized to the array following Agilent's SureSelect DNA Capture Array protocol version 1.0. Different blocking oligonucleotides (Table 2.6) were added to the hybridization. After elution from the array, half of the captured library was amplified in five 50 ml PCR reactions for 18 cycles using Phusion High-Fidelity DNA polymerase (Finnzymes) and post-capture primer pair (Table 2.6), purified using an E-Gel CloneWell (Invitrogen) to remove primer dimers, and re-amplified using fifteen PCR cycles with the same primer pair. The amplified library was again purified to eliminate primer dimers using E-Gel. For the second capture array, 2 μg of each of the 12 amplified libraries was pooled and hybridized to the array. After elution, half of the captured library was amplified in five 50 ml PCR reactions for 15 cycles and purified using Agencourt AMPure XP magnetic beads.

Table 2.6. PCR primers for library amplification and blocking oligonucleotides

Pre-capture Primers

Pre-capture Fw	5'-ACACTCTTTCCCTACACGACGCT-3'
Pre-capture Rev	5'-GGTCTCGGCATTCCTGCTGAA-3'

Post-capture Primers

Post-capture Fw[1]	5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T-3'
Post-capture Rev[2]	5'-CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC* T-3'

Blocking Oligos

BO_A (Adapter sequence 1)	5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'
Rev. comp. BO_A	5'-AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
BO_B (Adapter sequence 2)	5'-GGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT-3'
Rev. comp. BO_B	5'-AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACC-3'

[1] PCR primer © 2006 Illumina, Inc. All rights reserved.
[2] Bentley *et al.* (2008) [32].

Illumina data analysis

The amplified targeted array captured libraries were sequenced using two lanes of an Illumina GA II, using single end 80 bp (first array) or 42 bp (second array) sequencing. Novobarcode software (http://novocraft.com/) was used to sort the reads according to their 3 bp barcode. For the first capture array, Novoalign was used to align the reads to the reference genome (hg19), using default settings but including adapter trimming. For the second capture array, the BWA aligner (http://bio-bwa.sourceforge.net/) was used. PCR duplicates were removed by the MarkDuplicates command in Picard (http://picard.sourceforge.net/index.shtml). Variants were called using SAMtools and low confidence variants were removed by the SAMtools varFilter command and by removing all insertions and deletions present in only one read. Annovar (http://www.openbioinformatics.org/annovar/) was used to identify coding variants, classify SNPs as synonomous or nonsynonomous, and classify SNPs as either novel or previously described in dbSNP version 132. For each individual, the percentage of reads that overlapped with regions targeted by the capture array was calculated using Galaxy (http://main.g2.bx.psu.edu/). The read depth for each base targeted by the array was calculated using the

DepthOfCoverage command in the Genome Analysis Toolkit (http://www.broadinstitute.org/gsa/wiki/index.php/ The_Genome_Analysis_Toolkit).

Copy number analysis based on Illumina read depth

The regions targeted by the capture array were divided into non-overlapping 50 bp windows. For each individual, the read depth for each window was normalized based on the total number of reads obtained for that individual. For each window in an individual, the normalized read depth was divided by the average read depth for that window in all 10 individuals, and the log2 value was calculated. Windows with an average read depth of less than 5X were excluded from the analysis. A perl script was written to find regions of consecutive windows that had log2 ratios below -0.7 or above 0.5, to identify deletions and duplications, respectively.

Sanger sequencing of *PTPN11*

All coding *PTPN11* exons were PCR amplified using primers containing an M13 sequencing tag (Table 2.7). PCR Primers were located between 26 and 160 bp away from the intron/exon boundary. For fragmented DNA extracted from formalin-fixed paraffin-embedded (FFPE) tumor samples, primers closer to the intron/exon boundary were used to allow for smaller PCR products (Table 2.7). PCR amplification was performed using FastStart Taq DNA Polymerase (Roche) at 95 ºC for 4 min, 35 to 40 cycles of 95 ºC for 30 sec, 60 ºC for 30 sec and 72 ºC at 50 sec, followed by 5 min at 72 ºC. PCR products were sent for purification and sequencing in both forward and reverse directions, using M13 sequencing primers, at the Molecular Genetics Core Facility at Children's Hospital Boston.

DNA extraction from exostoses and subcloning of PCR products

DNA was extracted from several sections of a formalin fixed, paraffin embedded exostosis excised from participant A-IV-5, using a QIAamp DNA FFPE Tissue Kit (Qiagen). To isolate DNA from the cartilage core of this lesion, Laser Capture Microdissection (LCM) was performed. 7-micron sections were placed on PEN membrane glass slides (MDS Analytical Technologies). Sections were stained using Toluidine Blue. The cartilage cores of several sections were isolated using LCM, pooled, and DNA was extracted using a PicoPure DNA Extraction Kit (Applied Biosystems) with an overnight incubation at 65 ºC. A second exostosis, that had been removed from participant A-IV-8, was frozen and embedded in several blocks of OCT. Hematoxylin & Eosin staining of sections of this exostosis was used to identify a block

Table 2.7. Primers used for amplification of *PTPN11* coding exons

Exon	Forward primer		Reverse primer		PCR product size (bp)
	Sequence[1]	Position[2]	Sequence[3]	Position[2]	
1	GGGCAGCTGCACA GTCTC	-178 to -160	GGGGACGAGGAG GGAAC	+143 to +127	334
1S[4]	CGGAGCCTGAGCA AGGAG	-54 to -35	CTTCCGGACGGGG CTAAC	+71 to +56	139
2	AAGGGACAGGGAA GGTCTTG	-137 to -118	AACCCTGAAGGCA GCCC	+133 to +117	393
2S[4]	AGTGCTGACAGTGT CTTGTTTT	-52 to -30	CTCTCAGGATCCTC TCTTTTCA	+33 to +12	208
3	TGTGGTTATTTCAC CCATCG	-163 to -144	AAAGGTACTCTGAA AATAATTTGATGC	+100 to +121	539
3S[4]	AACTCTTTATTTGTC CCCTTGC	-47 to -26	AGGGAGCAGCAGA CTTTGTG	+40 to +21	284
4	TGATCAATCCCTTG GAGGAA	-134 to -115	GAAAAATCACCCAA AGGTAACA	+56 to +35	383
4B	TGAAAGAACAACAT GAACCCATA	-61 to -38	GTGCGCACAGAAA GAACAAA	-68 to -87	187
4S1[4]	TGAAAGAACAACAT GAACCCATA	-61 to -39	GTGCGCACAGAAA GAACAAA	c.451 to c.471	187
4S2[4]	CTTGTACGAGAGAG CCAGAGC	c.406 to c.427	TCTTGCCAGACCC ATTTTTC	+35 to +15	154
5	CTGCAGTGAACATG AGAGTGC	-83 to -63	TTCTGTCTACCTCT GTATGTTGC	+89 to +66	289
5S[4]	TGAGAGTGCTTGAA AACACTAATGTAA	-71 to -45	TTTAAGTTTTTAAG GAGAAAGCTTGAG	+44 to +18	232
6	TTTGCATTAACACC GTTTTCTG	-64 to -43	CAAGAGCAACTTCT CCTCCA	+71 to +52	249
6S[4]	CATCAACTGCTGTA CTCGATCA	-23 to -6	CAGGTCCAATTCCA AACACA	+89 to +70	226
7	CCCAGATGAACATT CTTGTAGC	-145 to -126	CAATTAAAGAAGCT AAGGCCTCC	+124 to +102	366
7S[4]	TTTTTCTGTGACTC TTTGACACG	-52 to -30	GGTGCTAGGAATC AAAATCTCC	+38 to +59	208
8&9	TGGACTAGGCTGG GGAGTAA	-87 to -68	TTCCTAAACATGGC CAATCTG	+55 to +76	527
8S[4]	TGGACTAGGCTGG GGAGTAA	-68 to -87	TTTCAGGACATGAG GAAGGA	+83 to +64	250
9S[4]	TCCTTCCTCATGTC CTGAAAG	-63 to -43	TTCCTAAACATGGC CAATCTG	+75 to +55	297
10	CTAACAGATGCGAA ACAGGC	-133 to -114	GGAAGCTAATAAAT GGCTACTGAATC	+137 to +112	402
10S[4]	CAGAGTTCACAGAA TTAACTTTCTTTT	-44 to -18	TCCCAAAGTCTCTA GAATACGACA	+34 to +11	210
11	GGACCTTTCAGTGG AACCTG	-128 to -109	CAATGCAGTTGCTC TATGCC	+139 to +120	423
11S[4]	CGGGTGATTCCTCA ACCTCT	-64 to -45	CGCTAGGAGACAG GGACAGG	+52 to +33	271
12	CTGACTCCAAAGCC CTATGC	-173 to -154	TCTGTTCTCAGGCA GACACAC	+154 to +134	395

(Table 2.7 continued)

Exon	Forward primer		Reverse primer		PCR product size (bp)
	Sequence[1]	Position[2]	Sequence[3]	Position[2]	
12S[4]	GAGTCTGAAACCCC CATGAA	-59 to -40	CCAGACTGTTTTCG TGAGCA	+79 to +59	205
13	CTGTAGCCATTGCA ACATGC	-140 to -121	TGGGAGTTTGTTTC CCATTC	+130 to +111	422
13S[4]	CCTGGCTCTGCAGT TTCTCT	-47 to -28	AGAGGCCTAGCAA GAGAATGA	+51 to +31	250
14	TTTTTGCTTTTTATC CCCTTAAA	-64 to -43	AAACCACCATGAC CAGGTAAA	+46 to +26	223
15	CAGGAACTGTGTCT GTACCATATC	-121 to -102	CCTGGTTTGTAAAA CTCAATTATGG	+120 to +96	347
15S[4]	GCTTAAACAGCGTG GTCTACA	-57 to -37	TGCCAATTATATAA ACAATTAACCAA	+48 to +23	207

[1] The forward M13 sequencing tag (5'-GTAAAACGACGGCCAGT-3') was added on the 5' end of all forward primers.
[2] Number of nucleotides upstream of the 5' end of the exon (-) or downstream of the 3' end of the exon (+). For exons 1 and 15, the end of the UTR was used rather than the end of the exon.
[3] The reverse M13 sequencing tag (5'- AACAGCTATGACCATG-3') was added on the 5' end of all reverse primers.
[4] Primers closer to the intron/exon boundary were used on some formalin-fixed paraffin-embedded tumor samples for which the amplification of PCR products greater than ~300bp failed, which suggested fragmentation of DNA.

consisting of >90% cartilage cells, and similarly to identify a block consisting of a majority of bone and fibrous tissue. For both blocks, several sections were pooled and DNA was extracted using the PicoPure DNA Extraction Kit (Arcturus) with a 3 hr incubation at 65 °C. PCR amplification of DNA from both participant's exostoses was performed using 2 to 6 µl of extracted DNA, in a 20 µl PCR reaction, using PCR conditions described above, and primer pair 4S1 (Table 2.7), or, to amplify the SNP in intron 11 of *PTPN11,* the primers 5'-GACCCTACAGCACTGCCATT-3' and 5'-CCCCACCACCTCAATACCTA-3'.

For the exostosis from participant A-IV-8, PCR amplimers from the cartilage or the bone and fibrous tissue sections were subcloned into a PCR-4-TOPO Vector using a TOPO TA cloning kit (Invitrogen). Individual colonies were selected and resuspended in a PCR mix, and amplified using primer pair 4S1 (Table 2.7). To distinguish between mutant and wild-type subclones, the PCR amplimers were restriction

digested by combining 0.2 μl of *DdeI*, 1.0 μl of Buffer 3 (New England Biolabs), and 9 μl of each PCR

reaction and then incubating at 37 °C for 2 hours. Digested products were separated by electrophoresis

on a 4% agarose gel. The 5 bp deletion creates a novel DdeI restriction site.

Multipex Ligation-Dependent Probe Amplification

MLPA was performed using the SALSA MLPA EK1 Kit (MRC-Holland). A custom probemix was

used, including two probes targeting exon 1 of *EXT1* [30,31] and four control probes (Table 2.8). The

resulting labeled PCR products were separated according to size (Applied Biosystems 3730 DNA

Analyzer) and the height of each peak was calculated. For each *EXT1* peak, a ratio was calculated by

dividing the height of the *EXT1* peak by the average height of the non-*EXT1* peaks. For each participant,

this ratio was divided by the average ratio obtained in the control samples, to obtain a corrected ratio,

such that a value of 1.0 represents a copy number of 2, and a value below the threshold of 0.80

represents a deletion.

Table 2.8. Probes used for MLPA

Gene	Chr	Amplicon size (bp)	Upstream hybridizing sequence	Downstream hybridizing sequence
EXT1 exon1 middle	8	116	5'_GGACACATGCAGGCCA AAAAACGCTA_3'	5'_TTTCATCCTGCTCTCAGC TGGCTCTTGTCTCGCCCTTT TGTTTTCATG_3'
EXT1 exon1-intron1	8	112	5'_GCATGGCAAAGACTGG CAAAAGCACAAGGAT_3'	5'_TCTCGCTGTGACAGAGA CAACACCGAGTATGAGAAG TAA_3'
RGAG1	X	102	5'_TGACTGTGAATTCTACA ACATATGGAGCGC_3'	5'_CTACCTTGTGCAAGGCA TTGTGGGTAAGTC_3'
AMMEC R1	X	106	5'_GCTCCCGGAAGATGGT GGTGTCAGCAGAGATG_3'	5'_TGCTGCTTTTGCTTCGAT GTGCTCTACTGTCA_3'
SERPIN B2	18	122	5'_CATGACTCCAGAGAACT TTACCAGCTGTGGGTTCAT GCA_3'	5'_GCAGATCCAGAAGGGTA GTTATCCTGATGCGATTTTG CAGG_3'
CHRDL 1	X	126	5'_CCAATGGAAAGACCTAT TCTCATGGCGAGTCCTGG CACCCAA_3'	5'_ACCTCCGGGCATTTGGC ATTGTGGAGTGTGTGCTAT GTACTT_3'

Microarray-based comparative genomic hybridization analysis

To determine if some patients with MC had deletions or duplications within the candidate interval on chromosome 12, we performed microarray-based Comparative Genomic Hybridization (aCGH) on 8 affected individuals from 7 families (Table 2.1). We designed an 8x60K Agilent SurePrint G3 array using eArray to include probes spanning the linkage interval, with an average probe spacing of 160 bp. We found 7 potential small duplications or deletions, all of which were less than < 1.5 kb in size. One of the duplications spanned a coding region in *IFT81*, but we could not verify this duplication using a PCR-based assay.

Immunodetection of SHP2

An excised exostotic lesion from participant B-IV-7 and rib cartilage from an unaffected individual were homogenized in RIPA buffer (25 mM Tris pH7.4, 150 mM NaCl, 5 mM EDTA, 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholic acid). HELA cells were lysed directly in NuPAGE LDS sample buffer (Invitrogen). Protein extracts were separated on a NuPAGE 4-12% Bis-Tris Gel (Invitrogen) and transferred to an Invitrolon PVDF membrane (Invitrogen). Immunodetection of SHP2 was performed using a WesternBreeze Chemiluminescent Immunodetection Kit (Invitrogen) and a 1:1000 dilution of mouse-anti-SHP2 (610621, BD Biosciences). As a loading control, immunodetection using mouse-anti-actin antibody (ab3280, Abcam, 1:600 dilution), was performed on the same blot.

2.7   Literature Cited

1. Bovée JVMG: **Multiple osteochondromas**. *Orphanet J Rare Dis* 2008, **3**:3.

2. Couvineau A, Wouters V, Bertrand G, Rouyer C, Gerard B, Boon LM, Grandchamp B, Vikkula M, Silve C: **PTHR1 mutations associated with Ollier disease result in receptor loss of function**. *Human Molecular Genetics* 2008, **17**:2766–2775.

3. Pansuriya TC, Kroon HM, Bovée JVMG: **Enchondromatosis: insights on the different subtypes**. *Int J Clin Exp Pathol* 2010, **3**:557–569.

4. Silve C, Jüppner H: **Ollier disease**. *Orphanet J Rare Dis* 2006, **1**:37.

5. Maroteaux P: **[Metachondromatosis]**. *Z Kinderheilkd* 1971, **109**:246–261.

6. Kennedy LA: **Metachondromatosis**. *Radiology* 1983, **148**:117–118.

7. Bassett GS, Cowell HR: **Metachondromatosis. Report of four cases**. *J Bone Joint Surg Am* 1985, **67**:811–814.

8. Ikegawa S, Nagano A, Matsushita T, Nakamura K: **Metachondromatosis: a report of two cases in a family**. *Nippon Seikeigeka Gakkai Zasshi* 1992, **66**:460–466.

9. Herman TE, Chines A, McAlister WH, Gottesman GS, Eddy MC, Whyte MP: **Metachondromatosis: report of a family with facial features mildly resembling trichorhinophalangeal syndromePediatr Radiol 1997 Nov;27(11):864**. *Pediatr Radiol* 1997, **27**:436–441.

10. Bovée JVMG, Hameetman L, Kroon HM, Aigner T, Hogendoorn PCW: **EXT-related pathways are not involved in the pathogenesis of dysplasia epiphysealis hemimelica and metachondromatosis**. *J. Pathol.* 2006, **209**:411–419.

11. Sobreira NLM, Cirulli ET, Avramopoulos D, Wohler E, Oswald GL, Stevens EL, Ge D, Shianna KV, Smith JP, Maia JM, et al.: **Whole-genome sequencing of a single proband together with linkage analysis identifies a Mendelian disease gene**. *PLoS Genet.* 2010, **6**:e1000991.

12. Shaw BA: **Resolving neonatal osteochondroma: a case report and literature review**. *Am J. Orthop.* 1996, **25**:226–228.

13. Keret D, Bassett GS: **Avascular necrosis of the capital femoral epiphysis in metachondromatosis**. *J Pediatr Orthop* 1990, **10**:658–661.

14. Wenger DR, Birch J, Rathjen K, Tobin R, Billman G: **Metachondromatosis and avascular necrosis of the femoral head: a radiographic and histologic correlation**. *J Pediatr Orthop* 1991, **11**:294–300.

15. Hunter AG, Kozlowski K, Hochberger O: **Metachondromatosis**. *Can Assoc Radiol J* 1995, **46**:202–208.

16. Mavrogenis AF, Skarpidi E, Papakonstantinou O, Papagelopoulos PJ: **Chondrosarcoma in metachondromatosis: a case report**. *J Bone Joint Surg Am* 2010, **92**:1507–1513.

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

18. Bovee JV, Clenton-Jansen AM, Wuyts W, Caethoven G, Taminiau AH, Bakker E, Van Hul W, Cornelisse CJ, Hogendoorn PC: **EXT-mutation analysis and loss of heterozygosity in sporadic and hereditary osteochondromas and secondary chondrosarcomas**. *American Journal of Human Genetics* 1999, **65**:689–698.

19. Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DCY, Nazareth L, Bainbridge M, Dinh H, Jing C, Wheeler DA, et al.: **Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy**. *N. Engl. J. Med.* 2010, **362**:1181–1191.

20. Roach JC, Glusman G, Smit AFA, Huff CD, Hubley R, Shannon PT, Rowen L, Pant KP, Goodman N, Bamshad M, et al.: **Analysis of genetic inheritance in a family quartet by whole-genome sequencing**. *Science* 2010, **328**:636–639.

21. Jorge AAL, Malaquias AC, Arnhold IJP, Mendonca BB: **Noonan syndrome and related disorders: a review of clinical features and mutations in genes of the RAS/MAPK pathway**. *Horm. Res.* 2009, **71**:185–193.

22. Kontaridis MI, Swanson KD, David FS, Barford D, Neel BG: **PTPN11 (Shp2) mutations in LEOPARD syndrome have dominant negative, not activating, effects**. *J. Biol. Chem.* 2006, **281**:6785–6792.

23. Tartaglia M, Zampino G, Gelb BD: **Noonan syndrome: clinical aspects and molecular pathogenesis**. *Mol Syndromol* 2010, **1**:2–26.

24. Grossmann KS, Rosário M, Birchmeier C, Birchmeier W: **The tyrosine phosphatase Shp2 in development and cancer**. *Adv. Cancer Res.* 2010, **106**:53–89.

25. Jones KB, Piombo V, Searby C, Kurriger G, Yang B, Grabellus F, Roughley PJ, Morcuende JA, Buckwalter JA, Capecchi MR, et al.: **A mouse model of osteochondromagenesis from clonal inactivation of Ext1 in chondrocytes**. *Proc. Natl. Acad. Sci. U.S.A.* 2010, **107**:2054–2059.

26. Bauler TJ, Kamiya N, Lapinski PE, Langewisch E, Mishina Y, Wilkinson JE, Feng G-S, King PD: **Development of severe skeletal defects in induced SHP-2-deficient adult mice: a model of skeletal malformation in humans with SHP-2 mutations**. *Dis Model Mech* 2011, **4**:228–239.

27. Saxton TM, Henkemeyer M, Gasca S, Shen R, Rossi DJ, Shalaby F, Feng GS, Pawson T: **Abnormal mesoderm patterning in mouse embryos mutant for the SH2 tyrosine phosphatase Shp-2**. *EMBO J.* 1997, **16**:2352–2364.

28. Denayer E, Devriendt K, de Ravel T, Van Buggenhout G, Smeets E, Francois I, Sznajer Y, Craen M, Leventopoulos G, Mutesa L, et al.: **Tumor spectrum in children with Noonan syndrome and SOS1 or RAF1 mutations**. *Genes Chromosomes Cancer* 2010, **49**:242–252.

29. Grossmann KS, Wende H, Paul FE, Cheret C, Garratt AN, Zurborg S, Feinberg K, Besser D, Schulz H, Peles E, et al.: **The tyrosine phosphatase Shp2 (PTPN11) directs Neuregulin-1/ErbB signaling throughout Schwann cell development**. *Proc. Natl. Acad. Sci. U.S.A.* 2009, **106**:16704–16709.

30. Jennes I, Entius MM, Van Hul E, Parra A, Sangiorgi L, Wuyts W: **Mutation screening of EXT1 and EXT2 by denaturing high-performance liquid chromatography, direct sequencing analysis, fluorescence in situ hybridization, and a new multiplex ligation-dependent probe amplification probe set in patients with multiple osteochondromas**. *J Mol Diagn* 2008, **10**:85–92.

31. White SJ, Vink GR, Kriek M, Wuyts W, Schouten J, Bakker B, Breuning MH, den Dunnen JT: **Two-color multiplex ligation-dependent probe amplification: detecting genomic rearrangements in hereditary multiple exostoses**. *Hum. Mutat.* 2004, **24**:86–92.

32. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, et al.: **Accurate whole human genome sequencing using reversible terminator chemistry**. *Nature* 2008, **456**:53–59.

Chapter **3**

Loss of SHP2 delays chondrocyte terminal differentiation and disrupts vertebral growth plate architecture

3.1   Abstract

Loss of SHP2 in mice or in human metachondromatosis (MC) patients causes benign cartilage tumors on the bone surface (exostoses) and within bones (enchondromas). To further our understanding of the mechanisms underlying cartilage tumor formation, we investigated the role of SHP2 in chondrocyte maturation *in vitro* and in *vivo*. We performed massively parallel mRNA sequencing (RNA-seq) to define the transcriptional consequences of SHP2 depletion during the maturation of chondrocytes in primary pellet cultures. We found that SHP2 depletion delays terminal differentiation to the late-hypertrophic stage and maintains gene expression associated with earlier maturation stages (proliferative, pre-hypertrophic and early-hypertrophic stage). We observed similar transcriptional changes in pellets treated with a MEK1/2 inhibitor, suggesting that SHP2 signals via the ERK1/2 pathway to delay terminal differentiation. We next analyzed chondrocyte maturation *in vivo* following postnatal mosaic inactivation of *Ptpn11*, which encodes SHP2. We observed that *Ptpn11*-cKO vertebral growth plates are highly disorganized and have an expanded domain of hypertrophic chondrocytes that have not yet terminally differentiated to the late-hypertrophic stage. Furthermore, we found that enchondroma-like lesions in *Ptpn11* cKO vertebrae arise from chondrocytes displaced from the growth plate due to non-uniform rates of maturation and ossification, and that they contain both SHP2 deficient and wild-type chondrocytes. These data suggest that mosaic loss of *Ptpn11* alters paracrine signaling within growth plates, consistent with our observations that maturation zones are similarly disorganized in the cartilage core of exostoses from human MC patients. In summary, these data suggest that in patients with MC, who are heterozygous for inherited *PTPN11* loss-of-function mutations, second-hit mutations in *PTPN11* produce enchondromas by causing disorganized growth, delayed terminal differentiation, and abnormal paracrine signaling from SHP2-deficient chondrocytes.

3.2   Contributing authors

Bowen ME[1], Ayturk UM[1], Kurek KC[1], Yang W[2], Warman ML[1].

1. Orthopaedic Research Laboratories, Children's Hospital Boston, MA 02115.

2. Department of Orthopaedics, Brown University, Providence, RI 02903

3.3   Specific contributions

Candidate's contributions

      Primary chondrocyte isolation and establishment of pellet cultures; RNA-seq library construction and data analysis; Mouse breeding and tamoxifen administration; Histology; Immunohistochemistry; q-RT-PCR.

Other author's contributions

      Assisting with mouse breeding; Assisting with RNA-seq data analysis; Providing human samples.

3.4   Introduction

Enchondromas in bone cavities and exostoses on bone surfaces are the two most common types of benign cartilage tumors [1]. Both tumors occur in patients with the autosomal dominant heritable disorder metachondromatosis (MC) [2]. Families with MC segregate heterozygous loss-of-function mutations in *PTPN11,* and their cartilage tumors are thought to arise from cells that have second-hit somatic *PTPN11* mutations [2,3]. *PTPN11* encodes the tyrosine phosphatase, SHP2, which has cell type-dependent and context-dependent roles in many signaling pathways [4]. MC is unique in that it is associated with both enchondromas and exostoses. Other genetic conditions cause enchondromas or exostoses, but not both. Patients with Ollier disease and Maffucci syndrome, due to somatic gain-of-function mutations in *IDH1* or *IDH2,* have enchondromas [5–7]. Patients with heterozygous loss-of-function mutations in *EXT1* or *EXT2* segregate multiple hereditary exostoses (MHE) [8]. The mechanism by which cartilage tumors arise in patients with MC is incompletely understood. Also unknown is whether tumor formation in patients with MC results from perturbation of the same biologic pathway(s) that are affected in patients with Ollier disease, Maffucci syndrome, and MHE.

Mice have been used to model tumor formation caused by SHP2 deficiency. Unlike humans, cartilage tumors have not been observed in mice that are heterozygous for *Ptpn11* loss-of-function mutations, and complete absence of *Ptpn11* in mice causes embryonic lethality [9]. However, cartilage tumors can be induced by a conditional biallelic inactivation of *Ptpn11* [10–12]. Mosaic inactivation of *Ptpn11* in chondrocytes and/or perichondrial cells produces enchondromas in the vertebrae and ribs, and exostoses in the limbs [11–13]. These data suggest that exostoses arise from *Ptpn11*-null perichondrial cells that inappropriately differentiate into chondrocytes [11,12], and that enchondromas arise from *Ptpn11*-null growth plate chondrocytes that fail to be replaced by bone [12]. Thus SHP2 appears to participate in the specification, proliferation, and maturation of chondrocytes.

One downstream signaling pathway likely altered in SHP2 deficient chondrocytes is the ERK1/2 pathway. Exostoses in *Ptpn11*-mutant mice have reduced levels of phospho-ERK1/2 [11,12]. While *in vivo* studies have not directly demonstrated that inactivation of the ERK1/2 pathway is sufficient to produce cartilage tumors, studies have shown that inactivation of *Erk1* and *Erk2* in mesenchymal cells leads to ectopic cartilage clusters in the perichondrium of mouse embryos [14]. Therefore, it is possible

that these ectopic clusters could develop into exostoses had the mice survived into adulthood. Furthermore, inactivation of *Erk1* and *Erk2* in chondrocytes disrupts the growth plate's columnar-organization, causes chondrocytes to hypertrophy prematurely, and delays the replacement of cartilage with bone [14,15]. Thus reduced ERK1/2 signaling following SHP2 depletion may mediate some of the phenotypic features in patients with MC

Because the molecular consequences of SHP2 depletion or ERK1/2 inactivation in chondrocytes are incompletely understood, we investigated the transcriptional response to either SHP2 depletion or MEK1/2 inhibition in chondrocyte pellet cultures by performing massively parallel mRNA sequencing (RNA-seq). We also conditionally inactivated *Ptpn11 in vivo* and examined the effect of SHP2 depletion on chondrocyte maturation, and we examined chondrocyte maturation in exostosis lesions from MC patients. Herein, we report that SHP2 and the ERK1/2 pathway have overlapping effects on gene expression during chondrocyte maturation in pellet culture, and we suggest a mechanism by which enchondromas arise in MC patients that involves disrupted growth plate architecture and inclusion of normal cells within enchondromatous lesions. Lastly, in addition to providing insights into the roles of SHP2 and ERK1/2 during chondrocyte maturation, the RNA-seq method and transcriptome database we developed should be useful to other scientists who are interested in growth plate biology and chondrocyte maturation.

3.5 Results

RNA-seq analysis of pellet cultures identifies both previously known and novel transcriptional changes that occur as chondrocytes mature from the proliferative through to the late-hypertrophic stage

To establish a system in which the molecular consequences of SHP2 depletion in chondrocytes could be studied, we utilize mouse primary chondrocyte pellet cultures [16,17]. In the growth plate, chondrocytes proceed through distinct zones of maturation before their cartilage is degraded and replaced by bone (Figure 3.1A). The proliferative zone (PZ) contains small, rapidly dividing chondrocytes that form columns, while the hypertrophic zone (HZ) contains chondrocytes that have exited the cell cycle and dramatically increased in size (Figure 3.1A) [18]. In chondrocyte pellet cultures, the transition from
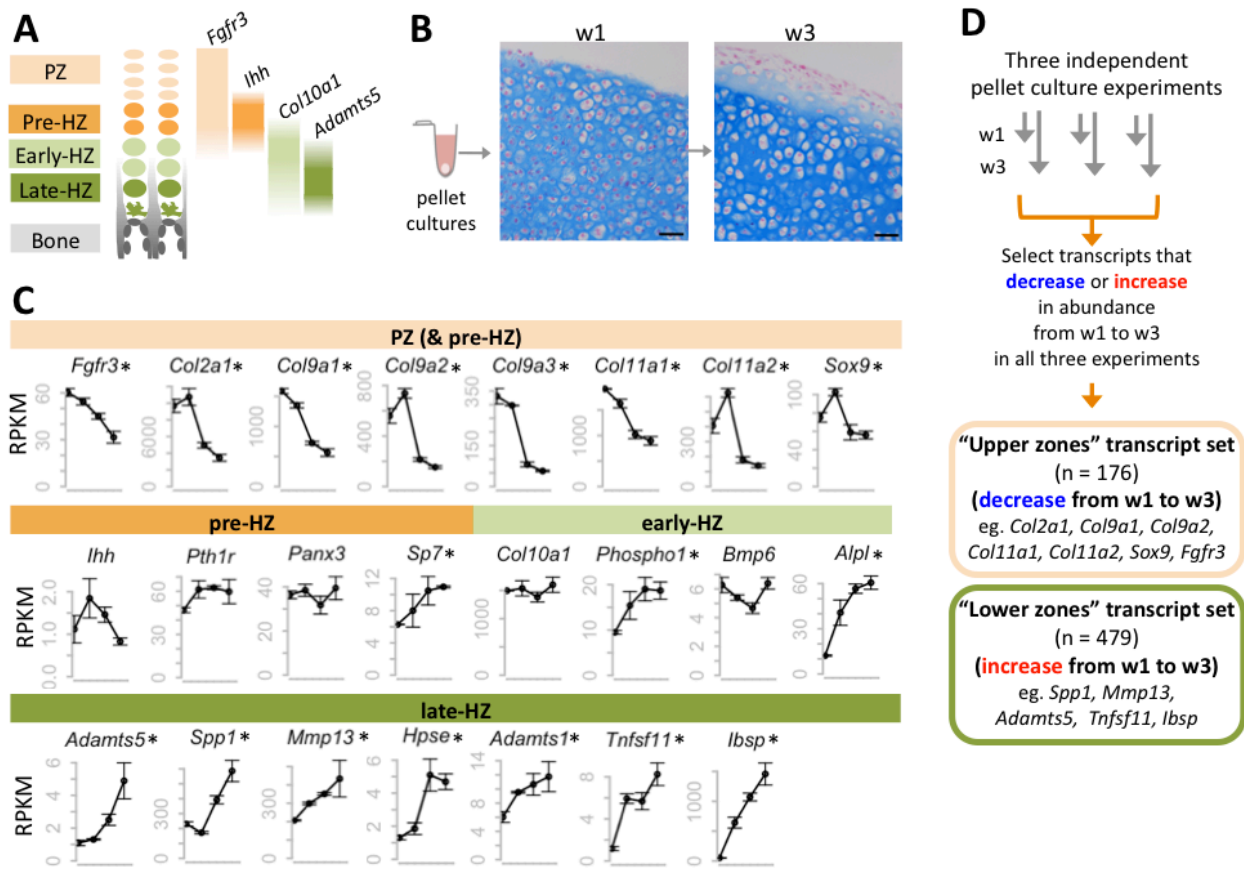
Figure 3.1. Identifying transcripts associated with the maturation of wild-type chondrocytes in pellet cultures. A: Schematic representation of the zones of maturation in the growth plate. Examples of four transcripts that are associated with specific maturation states are indicated. PZ: proliferative zone. HZ: hypertrophic zone. **B:** Photomicrographs of Alcian Blue and Nuclear Fast Red stained sections through chondrocyte pellets after 1 week (w1) and 3 weeks (w3) in culture (scale bar = 50 μm). Note the overall increase in chondrocyte size between w1 and w3, consistent with chondrocytes undergoing hypertrophy. **C**: Graphs showing the abundance (RPKM) of selected transcripts in wild-type chondrocytes after 4 days, or 1, 2 or 3 weeks in pellet culture. Above each graph, the growth plate zones in which transcripts have previously been shown to be most abundant are indicated. RNA-seq was performed on 3 pellets per time point, and the average RPKM at each time point is indicated (± 1 SD). Transcripts that significantly (p<0.05) increase or decrease in abundance over time are indicated with an asterisk. **D**: Flow Chart showing the method used to define a set of transcripts that consistently decreased or increased in abundance over time in three independent experiments.

the proliferative to hypertrophic state can be modeled (Figure 3.1B). We evaluated the process of chondrocyte maturation at the transcriptional level by performing RNA-seq on pellet cultures.

To verify that RNA-seq is a reproducible method for measuring transcript abundance in chondrocytes undergoing maturation in pellet cultures, we first performed RNA-seq on wild-type chondrocytes after 4 days, or 1, 2 or 3 weeks in culture. We sequenced 3 libraries per time point, and obtained, on average, 13 million 50-bp paired-end reads per library, of which >95% could be mapped to

the mouse genome (Table 3.1). To determine the relative abundance of transcripts at each time point, we counted the number of reads mapping to each transcript, and normalized this based on the length of the transcript and the number of reads per library, to obtain an RPKM value [19] (Supp Data File 1). We found that the abundance of transcripts was similar across libraries from the same time-point ($R^2 > 0.99$) (Figure 3.2), and that the most abundant transcripts corresponded to genes known to be highly expressed in chondrocytes (e.g., *Col2a1*, *Col9a1*, *Sparc*) [20]. Furthermore, while transcripts associated with intermediate stages of chondrocyte maturation (pre-hypertrophic and early-hypertrophic) did not consistently change in abundance over time, many transcripts known to be associated with proliferative chondrocytes decreased in abundance and transcripts associated with late-hypertrophic/terminally differentiated chondrocytes increased in abundance (Figure 3.1C; Table 3.2). This indicates that as pellet cultures mature, the proportion of proliferative chondrocytes decreases, while the proportion of terminally-differentiated chondrocytes increases. Thus, our data suggest that RNA-seq is a reliable method for detecting changes in transcript abundance that occur as chondrocytes mature from the proliferative through to the late-hypertrophic stage.

To further define transcriptional changes associated with chondrocyte maturation, we sought to identify a reproducible set of transcripts that change in abundance over time. First, we selected all transcripts that significantly increased or decreased in abundance over time in our wild-type pellet cultures. Only transcripts with an RPKM > 3 that showed at least a 2-fold change and a *p* value < 0.05 were selected. Second, to reduce false positives, we incorporated data from 2 additional independent RNA-seq data sets. These were from pellets that were used as vehicle controls in our experiments described in further detail below. We justified incorporating data from vehicle-treated pellets by showing that vehicle-treatment alone did not affect known chondrocyte maturation transcripts (Figure 3.3) and that the correlation coefficient between untreated and vehicle-treated libraries at the same time point was high ($R^2 > 0.98$) (Figure 3.2). We found that 81% of all transcripts that increased or decreased in abundance over time in untreated pellets also increased or decreased in abundance by at least 1.5 fold in both sets of vehicle-treated pellets. By focusing only on the transcripts that reached statistical significance and exhibited at least a 2-fold change in all three experiments, we identified a high-confidence set of 176 transcripts that consistently decreased in abundance over time ("upper zones" transcript set) and a set of

**Table 3.1: Number of reads and mapping statistics for each library**

Mouse litter	Genotype & time point	Replicate	Total # reads (50bp paired end)	% mapped
A	**Day 4** Wild-type	1	24,958,041	96%
		2	10,908,661	97%
		3	13,805,624	97%
	Week 1 Wild-type	1	13,367,197	95%
		2	13,300,200	98%
		3	9,673,123	97%
	Week 2 Wild-type	1	13,234,707	96%
		2	9,378,957	96%
		3	10,341,598	98%
	Week 3 Wild-type	1	11,077,264	96%
		2	10,969,002	97%
		3	13,946,829	97%
B	**Week 1**, 4-OHT Wild-type	1	16,566,070	96%
		2	17,246,647	96%
		3	16,205,375	96%
	Week 1, untreated Wild-type	1	17,057,605	97%
		2	23,336,815	97%
		3	25,856,434	97%
	Week 3, 4-OHT Wild-type	1	20,776,870	95%
		2	16,113,528	96%
		3	14,954,606	85%
	Week 3, untreated Wild-type	1	15,769,721	97%
		2	14,294,479	95%
		3	20,438,780	97%
C	**Week 1**, 4-OHT *Ptpn11 –/fl; CMV-CreERT2*	1	12,110,000	93%
		2	9,657,059	94%
		3	9,507,599	96%
	Week 1, untreated *Ptpn11 –/fl; CMV-CreERT2*	1	25,954,789	95%
		2	12,766,566	97%
		3	7,648,448	86%
	Week 3, 4-OHT *Ptpn11 –/fl; CMV-CreERT2*	1	31,478,241	97%
		2	40,187,879	96%
		3	31,181,967	96%
	Week 3, untreated *Ptpn11 –/fl; CMV-CreERT2*	1	26,929,091	96%
		2	18,316,729	93%
		3	21,570,699	94%
D	**Week 1** U0126	1	14,138,829	97%
		2	12,634,668	96%
		3	13,823,998	96%
	Week 1 DMSO control	1	12,165,140	97%
		2	11,748,157	96%
		3	12,558,375	97%
	Week 3 U0126	1	12,765,004	93%
		2	11,862,249	97%
		3	13,849,488	95%
	Week 3 DMSO control	1	13,018,137	94%
		2	12,150,097	93%
		3	12,601,300	96%

Table 3.2: Previous studies identifying transcripts that are associated with a specific chondrocyte maturation stage.

Growth plate zone	Transcript	Reference
PZ	*Col2a1*	[21]
	Col9a1	[21]
	Col9a2	[21]
	Col9a3	[21]
	Col11a1	[21]
	Col11a2	[21]
	Sox9	[22]
	Fgfr3	[23]
Pre-HZ	*Pth1r*	[24]
	Ihh	[24]
	Panx3	[25]
	Sp7	[26]
Early-HZ	*Col10a1*	[22]
	Phospho1	[27]
	Bmp6	[28]
	Alpl	[29]
Late-HZ	*Spp1*	[22]
	Mmp13	[22]
	Hpse	[30]
	Adamts5	[31]
	Adamts1	[32]
	Tnfsf11	[33]
	Ibsp	[34]
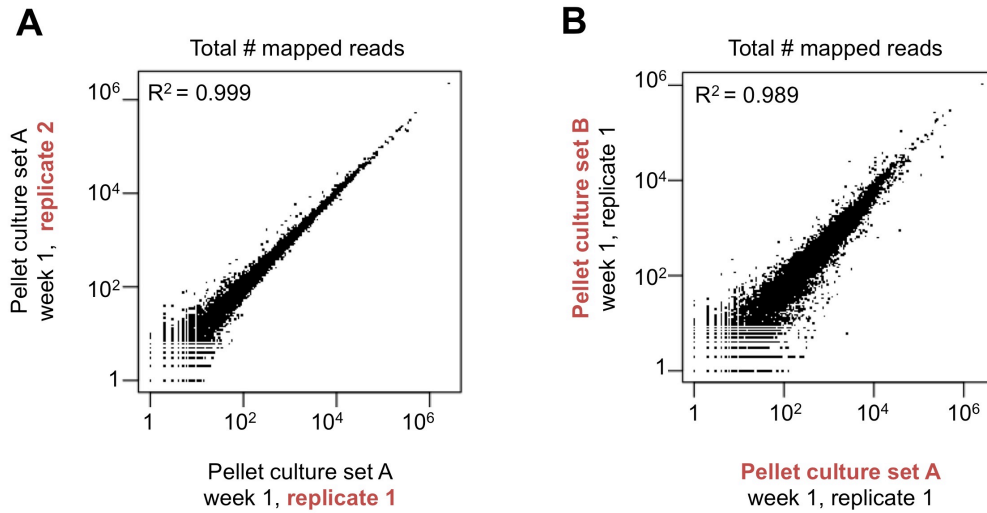
Figure 3.2: Correlation between libraries from the same time point. Scatterplots comparing the number of reads mapping to each gene in RNA-seq libraries prepared from pellets harvested after 1 week in culture. Each circle represents an individual gene. **A:** Comparison of two representative RNA-seq libraries prepared from pellets that were established at the same time from the same litter of mice. **B:** Comparison of two representative RNA-seq libraries prepared on different days from pellets that were independently established from different mouse litters. Pellets in set A were untreated. Pellets in set B were treated with 0.1% DMSO and were used as vehicle control pellets in our MEK1/2-ihnibition experiment. The Pearson correlation coefficient (R^2), calculated using the trimmed-mean method, is indicated on each plot.
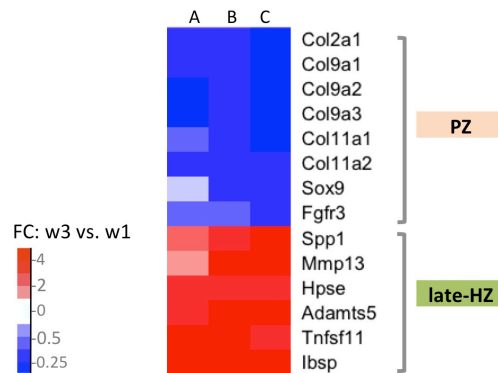
Figure 3.3: Vehicle treatment alone does not significantly alter chondrocyte maturation in pellet cultures. Heatmap showing the fold change in transcript abundance over time (from w1 to w3) in three independent pellet culture experiments, for a set of transcripts associated with the proliferative zone (PZ) or late-hypertrophic zone (late-HZ). A = untreated pellets. B = pellets treated with 0.1% DMSO as a vehicle control in our MEK1/2-inhibition experiments. C = pellets treated with 0.1% EtOH as a vehicle control in our 4-OHT-treatment experiments.

479 transcripts that consistently increased in abundance ("lower zones" transcript set) (Figure 3.1D, Supp Data File 1). Known markers of proliferative or hypertrophic chondrocytes were identified in these transcript sets, respectively, thus validating our approach (Figure 3.1D). Conversely, many of the transcripts in these data sets have not yet been studied in the growth plate and likely represent novel factors involved in chondrocyte maturation.

## SHP2 depletion increases the abundance of transcripts associated with early maturation stages, and decreases the abundance of transcripts associated with terminal differentiation

We performed RNA-seq on pellets carrying a conditional *Ptpn11* allele and a tamoxifen-dependent Cre recombinase, that were treated daily with 4-hydroxytamoxifen (4-OHT), to inactivate *Ptpn11* (*Ptpn11* cKO), or with vehicle control (0.1% v/v EtOH). One week of 4-OHT treatment induced recombination in >90% of chondrocytes (Figure 3.4) and reduced SHP2 levels (Figure 3.5A). Two weeks of 4-OHT treatment, followed by an additional week in culture, almost completely depleted SHP2 (Figure 3.5A). Since it was not yet known whether long-term 4-OHT treatment itself would affect gene expression, we also performed RNA-seq on similarly treated wild-type pellets (Figure 3.5B, Supp Data File 1). Of all transcripts that changed in abundance in *Ptpn11* cKO pellets, 18 (5%) at w1 and 59 (15%) at w3 also changed in wild-type pellets treated with 4-OHT. We therefore excluded these 77 transcripts from subsequent analyses (Figure 3.5B, Figure 3.6). There was significant (p < 0.001) overlap between
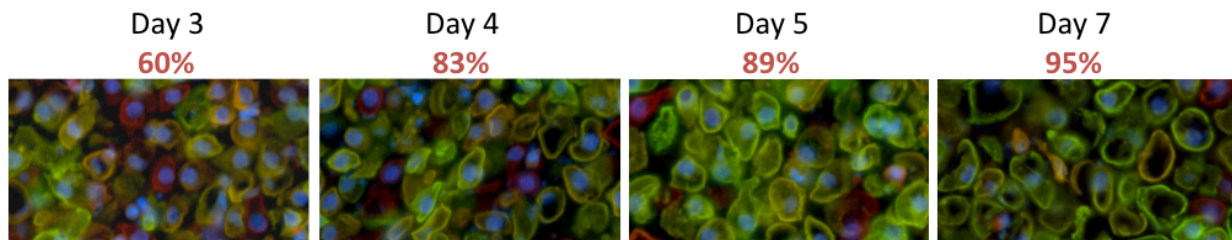
Figure 3.4: Using the *mTmG* reporter to monitor CreER mediated recombination in pellet cultures. Merged images of GFP, RFP and DAPI channels after daily 4-OHT treatment of chondrocytes pellet cultures established from mice carrying the *mTmG* reporter allele. The *mTmG* reporter initially drives expression of *Tomato Fluorescent Protein* (*TFP*), but after CreER-mediated recombination, it drives the expression of *Green Fluorescent Protein* (*GFP*). The number of green cells (recombined) and red cells (un-recombined) were counted in multiple tissue sections from pellets cultured for 3, 4, 5 or 7 days. The average percentage of recombined cells at each time point is indicated. Based on the half-life of the TFP protein, some TFP will still be present in cells with a relatively recent recombination event, and these cells will appear yellow.
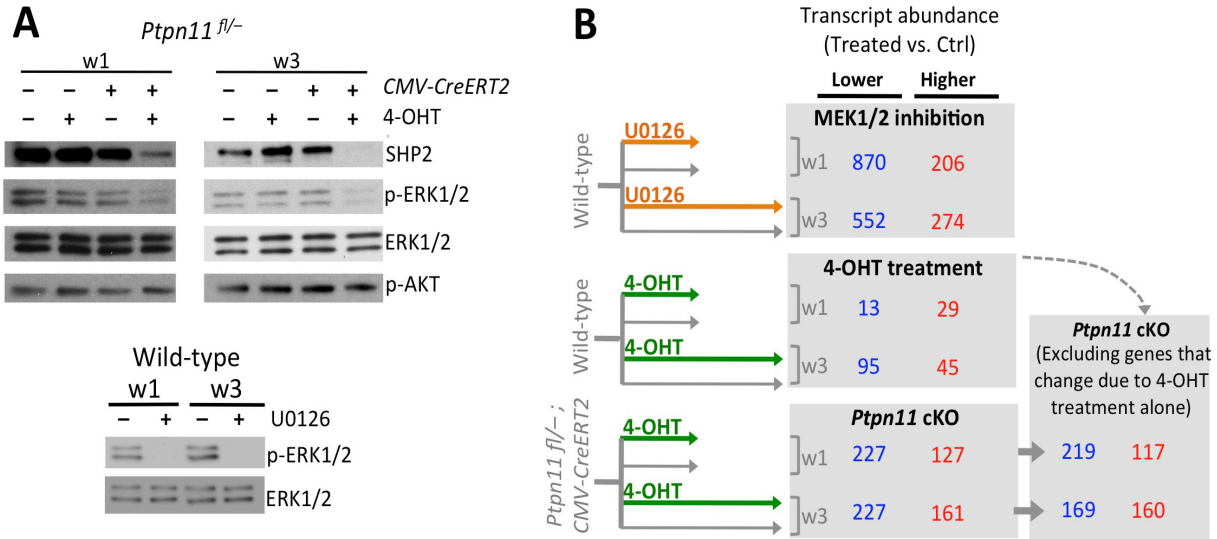
Figure 3.5: Identifying transcripts that change in abundance following SHP2-depletion or MEK1/2-inhibition in pellet cultures. A: Immunodetection of SHP2, phospho-ERK1/2, total ERK1/2, and phospho-AKT in chondrocyte pellet cultures. Top: pellets from *Tg*(*CMV-CreERT2); Ptpn11*^fl/−^ mice were treated daily with 4-OHT or vehicle control (EtOH) an harvested after 1 and 3 weeks. Bottom: pellets from wild-type mice were treated daily with a MEK1/2 inhibitor (U0126), or vehicle control (DMSO), and harvested after 1 and 3 weeks. **B**: Number of transcripts that had significantly lower (blue) or higher (red) abundance in treated pellets compared to vehicle treated pellets, either at w1 or w3. Transcripts that changed in abundance in *Ptpn11* cKO pellets were excluded if they exhibited similar changes in wild-type pellets treated with 4-OHT.
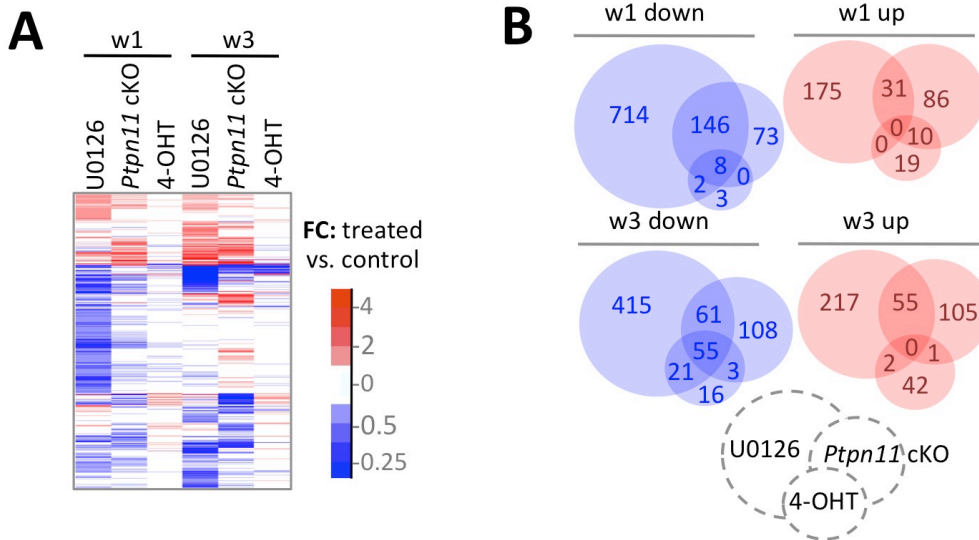
Figure 3.6: Overlap between transcripts that change in abundance after SHP2-depletion, 4-OHT treatment, or U0126-treatment. **A:** Heat map of all transcripts that changed in abundance after SHP2-depletion, 4-OHT-treatment or U0126-treatment, at either w1 or w3. **B:** Venn diagrams showing the overlap of transcripts that decrease (blue) or increase (red) in abundance, at w1 or w3, in U0126-treated (left circle), *Ptpn11*-cKO (right circle) and 4-OHT-treated (lower circle) pellets.

transcripts whose abundance changed in *Ptpn11* cKO pellets at w1 and w3 (Figure 3.7). However, there were also differences between the w1 and w3 data sets. For example, transcripts encoding protein translation components were decreased in w1 but not w3 in *Ptpn11* cKO pellets (Figure 3.8). Thus, during chondrocyte maturation the transcriptional response to short term versus sustained SHP2 depletion is similar but not identical

To assess the effect of SHP2 depletion on chondrocyte maturation, we compared transcript abundance in *Ptpn11* cKO pellets to that of vehicle pellets at w3. Transcripts expressed by proliferative through early hypertrophic chondrocytes were increased and transcripts expressed by late stage hypertrophic chondrocytes were decreased in *Ptpn11* cKO pellets (Table 3.3; Figure 3.9A). The greatest change in transcript abundance following SHP2 depletion was in the pre-hypertrophic chondrocyte

A *Ptpn11* cKO*

* excluding genes that were also differentially expressed after 4-OHT treatment alone

w1 w3

Up 90 **27** 133

Alpl	Gm10222	mt-Co3	Nog
Col27a1	Gm8113	mt-Cytb	Panx3
Cyb5r2	Gramd2	mt-Nd3	Slc17a9
D830030K20Rik	Grm1	mt-Nd4	Slc7a3
Dppa5a	Ifit3	mt-Nd4l	Tcfl5
Drp2	mt-Co1	mt-Nd5	Zfp185
Fam180a	mt-Co2	mt-Nd6	

Down 178 **41** 128

Acaa2	Cebpa	G0s2	Nrn1
Acaca	Cfd	Gbe1	Pcx
Adipoq	Col15a1	Gm17722	Pparg
Adipor2	Ddah1	Gm20425	Rarres2
Alox5ap	Dgat2	Gpam	Selenbp1
Apcdd1	Egln3	Gsn	Serpine1
Apoe	Errfi1	Hp	Sik2
C3	Etfb	Il1r1	Trf
Car3	F3	Loxl1	
Car9	Fasn	Mt2	
Cav1	Fzd4	Nid1	

Ptpn11 cKO
w1 w3

FC: treated vs. control
-4
-2
-0
-0.5
-0.25

B U0126

w1 w3

Up 167 **39** 235

2810055G20Rik	Gls	mt-Co3	Pak3
2810055G20Rik	Gm10222	mt-Cytb	Pcgf6
9230110C19Rik	Gm10925	mt-Nd3	Peg10
Amy1	Id4	mt-Nd4	Prkaa2
BC030870	Ifitm1	mt-Nd4l	Ptbp2
Clk4	Luc7l3	mt-Nd5	Rnpc3
Crebzf	mt-Atp6	mt-Nd6	Zcchc7
D4Wsu53e	mt-Atp8	Npr3	Zfp101
Dusp8	mt-Co1	Ogt	Zfp185
Fam36a	mt-Co2	Olfr613	

Down 742 **128** 424

1110007C09Rik	Crip2	Incenp	Plau
1700019G06Rik	Crlf1	Inf2	Pld3
5031439G07Rik	Cx3cl1	Isg15	Por
Aebp1	Cyba	Lgals1	Ppap2b
Ank	Dlk1	Lmna	Ptpn1
Ap1b1	Dnmt1	Loxl3	Rgs16
Apoe	Dok1	Ltbp2	Rnh1
Apol7d	Ecm1	Ltbp3	Rpl36
Aprt	Efhd2	Ly6e	Sbno2
Aqp1	Emp3	Lyz1	Serpina3n
Arhgdia	Fcer1g	Lyz2	Serping1
Arhgdib	Flot2	Mcm3	Sertad1
Arpc1b	Fn1	Mfap2	Sh3bgrl3
Arpc3	Fosl1	Mfge8	Slc27a1
B4galt5	Ftl1	Mmp19	Slc29a1
Bhlhe40	Gm10382	Mmp2	Slpi
Bst2	Gm10393	Mmp23	Socs3
C1rb	Gm16627	Mnda	Sod3
C230081A13Rik	Gm16721	Mt1	Spp1
C4b	Gm17329	Mt2	Tagln2
Capg	Gm17507	Mvp	Taldo1
Cd248	Gm20425	Myh9	Thbs4
Cdca4	Gpnmb	Nab2	Thy1
Chi3l1	Grn	Nbl1	Timp1
Ckb	H2-D1	Ncapd2	Tln1
Clip2	Hmga1	Nid1	Tmem189
Clu	Hp	Nt5dc2	Trf
Col16a1	Icam1	Oaf	Tspo
Col4a1	Ifi27l2a	P4ha3	Ttyh3
Coro1c	Ifitm2	Pcolce	Twf2
Cotl1	Ifitm3	Pdgfra	Tyrobp
Crip1	Igfbp6	Pdpn	Zyx

U0126
w1 w3

FC: treated vs. control
-4
-2
-0
-0.5
-0.25

Figure 3.7: Overlap between transcripts that changed in abundance at w1 and w3. Venn diagrams showing the overlap between transcripts that increased (red) or decreased (blue) in abundance after SHP2-depletion (**A**) or U1026 treatment (**B**) at w1 and w3.

Figure 3.8: Gene Process Networks enriched in differentially expressed genes. List of enriched Gene Process Networks, identified using the MetaCore software suite from GeneGo Inc, in the list of transcripts that increased (red) or decreased (blue) in abundance after U0126-treatment, SHP2 depletion, or 4-OHT treatment, at w1 or w3. The list of transcripts included in each network is indicated. Only networks with a MetaCore p value less than 1×10^{-6} are shown.

Table 3.3: Effect of SHP2 depletion or MEK1/2 inhibition on markers of chondrocyte maturation zones

GP zone	Gene	MEK1/2 inhibition		SHP2 depletion		4-OHT treatment	
		FC*	P value	FC*	P value	FC*	P value
PZ	Sox9	2.81	2.4E-03	1.55	7.7E-01	1.19	1.0E+00
	Col9a1	2.26	3.3E-02	1.73	5.2E-01	0.98	1.0E+00
	Fgfr3	4.29	1.4E-05	3.44	7.7E-03	0.75	1.0E+00
	Acan	3.60	1.4E-04	1.49	7.7E-01	1.09	1.0E+00
Pre-HZ	Pth1r	2.88	5.0E-03	4.00	1.5E-03	0.75	1.0E+00
	Ihh	7.18	3.1E-09	14.13	7.6E-11	0.61	4.5E-01
	Panx3	5.90	1.6E-08	10.35	5.5E-09	0.68	7.0E-01
	Sp7	2.80	4.4E-03	2.06	2.5E-01	0.88	1.0E+00
Early-HZ	Col10a1	2.14	5.3E-02	2.07	2.5E-01	0.56	1.3E-01
	Phospho1	4.52	8.3E-06	3.44	8.5E-03	0.90	1.0E+00
	Bmp6	3.56	2.4E-04	2.83	4.1E-02	0.71	1.0E+00
	Alpl	2.49	1.9E-02	4.60	3.2E-04	0.81	1.0E+00
Late-HZ	Adamts1	0.61	3.5E-01	0.30	1.6E-02	1.42	8.8E-01
	Adamts5	0.18	1.1E-07	0.34	2.9E-02	1.14	1.0E+00
	Spp1	0.05	1.5E-22	0.83	1.0E+00	0.62	3.4E-01
	Hpse	0.21	1.8E-06	0.26	4.5E-03	0.30	2.9E-06

* FC = Fold change when comparing treated pellets to control pellets at w3

expressed protein IHH (14-fold increase). However, not all well-characterized stage-specific transcripts were altered by SHP2 depletion. Therefore, we took a non-biased approach by examining the effect of SHP2 depletion on the "upper zones" and "lower zones" transcript sets we had previously defined in wild-type pellet cultures (Figure 3.1D). Compared to control pellets at w3, *Ptpn11* cKO pellets had 44% of the "upper zones" transcripts increase in abundance by > 1.5 fold, and 34% of the "lower zones" transcripts decrease in abundance by > 1.5 fold. When considering *Ptpn11* cKO transcripts whose change in abundance satisfied our significance criteria (2-fold change, $p < 0.05$) 20 "upper zones" transcripts were increased and 54 "lower zones" transcripts were decreased (Figure 3.9B; Figure 3.10). These data are consistent with SHP2 depletion delaying terminal differentiation while maintaining chondrocytes in the proliferative, pre-hypertrophic, or early-hypertrophic stage.
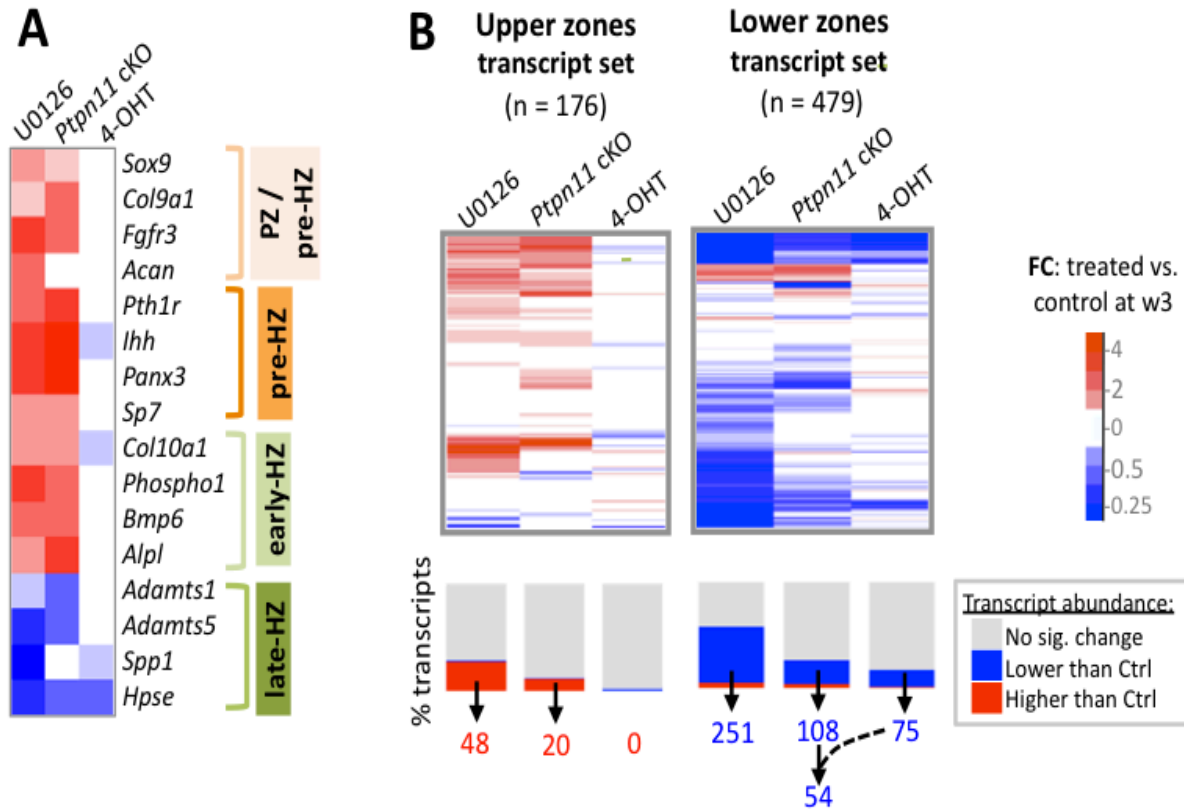
Figure 3.9: SHP2 depletion or MEK1/2 inhibition increases the abundance of PZ, pre-HZ and early-HZ transcripts, and decreases the abundance of late-HZ transcripts. A: Heat map showing the fold change in abundance (treated vs. control pellets at w3) of selected transcripts that are associated with a specific stage of chondrocyte maturation. **B**: Heat maps showing the fold change in abundance (treated vs. control pellets at w3) for all transcripts that we included in the "upper zones" and "lower zones" transcript sets. The graph below the heat map indicates the percentage of transcripts for which the increase (red) or decrease (blue) in abundance was significant (p<0.05). The number of "upper zones" transcripts that increased in abundance in treated pellets (red), and the number of "lower zones" transcripts that decreased in abundance in treated pellets (blue), is indicated below. For *Ptpn11* cKO pellets, of the 108 "lower zones" transcripts that significantly decreased in abundance, 54 did not significantly change in wild-type pellets treated with 4-OHT.
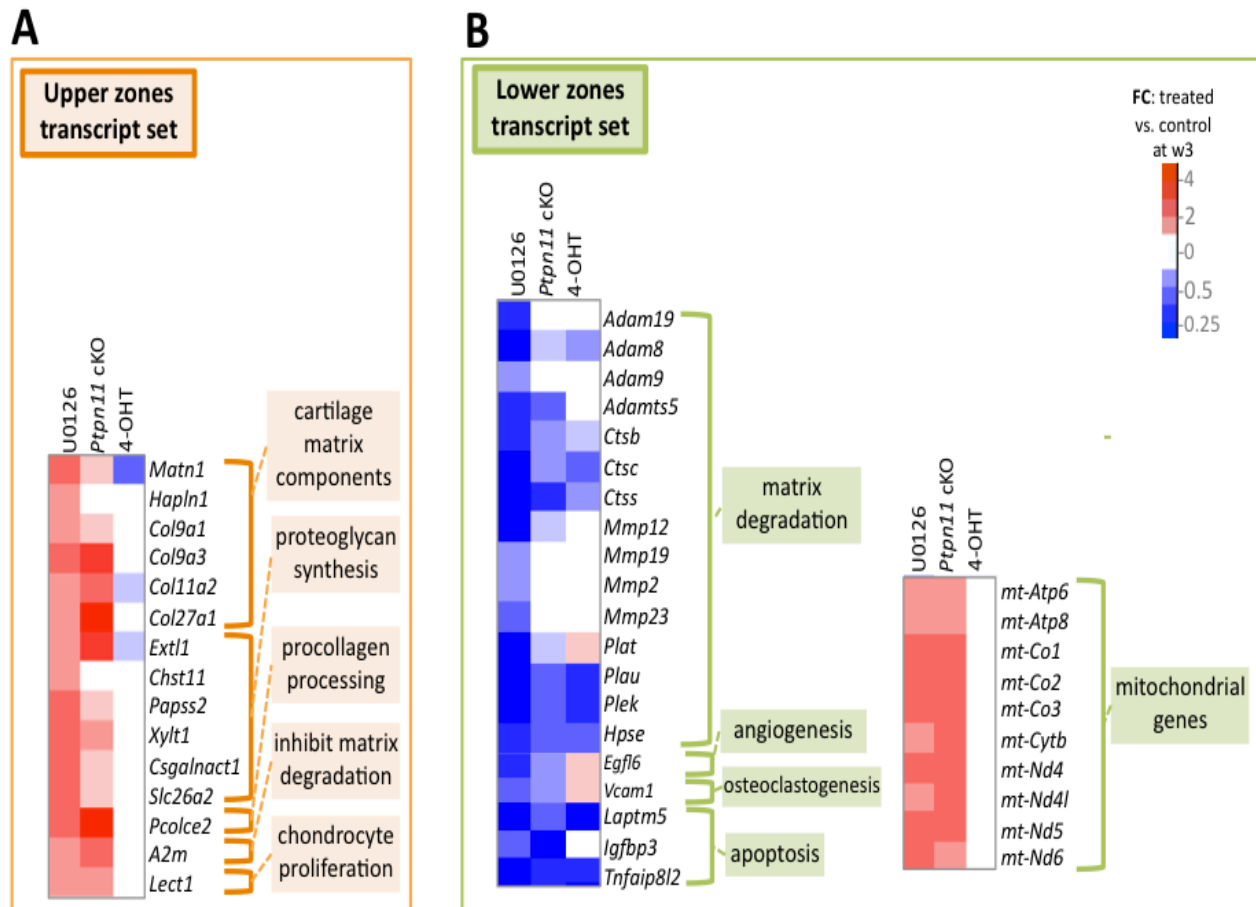
Figure 3.10: Identification of transcripts associated with chondrocyte maturation that changed in abundance after MEK1/2-inhibition or SHP2-depleiton. Heatmaps for a subset of transcripts identified in our "upper zones" (A) or "lower zones" (B) transcript sets that significantly changed in abundance at w3 after either MEK1/2-inhibition or SHP2-depletion. Transcripts were selected based on their known roles in processes that are relevant to proliferative chondrocytes (eg. matrix synthesis) or late-hypertrophic chondrocytes (eg. matrix degradation). Some, but not all, of the "lower zones" transcripts that decreased in abundance after SHP2-depletion also decreased in abundance after 4-OHT treatment of wild-type pellets. The increase in the abundance of transcripts encoded by the mitochondrial genome may be a result of the increase in the number of mitochondria per cell due to the increase in chondrocyte cell size following SHP2 depletion or MEK1/2 inhibition.

MEK1/2 inhibition perturbs chondrocyte maturation in a similar manner to SHP2 depletion

SHP2 has been proposed to positively regulate the ERK1/2 pathway in chondrocytes [11,12]. Indeed, we observed reduced phospho-ERK1/2 levels after SHP2 depletion in pellet cultures (Figure 3.5A). To determine whether this reduction in ERK1/2 pathway activity mediates the transcriptional changes that occur following SHP2 depletion, we performed RNA-seq on pellets cultured for 1 or 3 weeks in the presence of either a vehicle control (0.1% v/v DMSO) or a MEK1/2 inhibitor (U0126) (Figure 3.5A, B). When considering the transcripts that changed in abundance after SHP2 depletion (excluding transcripts that changed after 4-OHT treatment alone), we found that 177 (53%) and 116 (35%) had similar changes after U0126-treatment at w1 and w3, respectively (Figure 3.5B, Figure 3.11). The observed overlap is significantly greater than expected by chance alone (p<0.001). As with SHP2 depletion, we found that U0126-treatment increased the abundance of transcripts associated with the PZ, pre-HZ and early-HZ, with the most substantial changes involving the pre-HZ transcripts, *Ihh* and *Panx3* (> 6-fold increase) (Table 3.3; Figure 3.9A,B). Furthermore, we found that U0126-treatment decreased the abundance of late-HZ transcripts associated with matrix degradation, apoptosis, angiogenesis and osteoclastogenesis (Figure 3.9A,B; Figure 3.10). In particular, a gene ontology analysis revealed that transcripts with reduced abundance in U0126-treated pellets were significantly enriched for proteins involved in matrix degradation and remodeling (Figure 3.8). We used quantitative reverse transcription PCR (qRT-PCR) to confirm that 3 weeks of U0126-treatment increased the abundance of pre-HZ and early-HZ markers (*Ihh*, *Phospho1*) and decreased the abundance of a late-HZ marker (*Spp1*) (Figure 3.12). Altogether, these data indicate there is substantial overlap between the genes regulated by SHP2 and MEK1/2 in chondrocytes, and that MEK1/2 inhibition and SHP2 depletion have similar effects on chondrocyte maturation.

A Week 1

U0126   *Ptpn11* cKO*

Up 175 **31** 86

Down 724 **146** 73

2810055G20Rik	Gm10925	mt-Co1	mt-Nd4l
2810055G20Rik	Gm2897	mt-Co2	mt-Nd5
A930018M24Rik	Gm3020	mt-Co3	mt-Nd6
Amy1	Gm8348	mt-Cytb	Npr3
Ccng1	Id4	mt-Nd1	Thoc1
Ddx26b	Il18	mt-Nd2	Trp53inp1
Gm10222	mt-Atp6	mt-Nd3	Zfp185
Gm10409	mt-Atp8	mt-Nd4	

1700019G06Rik	Egr1	Map1lc3a	Rpl35
2810422J05Rik	Eln	Mbd3	Rpl37a
9430023L20Rik	Emilin1	Mfap2	Rpl8
Aacs	Eno1	Mif	Rplp0
Acaa2	Etfb	Mospd3	Rplp1
Adam15	Extl3	Mrpl12	Rplp2-ps1
Aes	Fasn	Mt1	Rps15
Agt	Fgfr1	Mt2	Rps26
Aldoa	Flot2	Mvp	Rps29
Aldoart1	Fosl1	Nab2	Rps5
Apcdd1	Furin	Nacc1	Rps8
Apln	Fxyd1	Naglu	Rps9
Apoe	Gaa	Nbl1	Ruvbl2
Aprt	Galk1	Ndufa11	Rxra
Atp5d	Gapdh	Ndufa4l2	Sbno2
Bckdha	Gm16721	Ndufv3	Scarb1
Bcl9l	Gm20425	Nid1	Sik2
Bhlhe40	Gm5506	Nt5dc2	Slc16a3
C3	Gm6981	Nxph4	Slc25a39
C4b	Gm8730	P4ha2	Slc27a1
Cd248	Gnb2	Pafah1b3	Slc2a1
Cdc42ep4	Grhpr	Pcx	Snai1
Cfd	Gsn	Pdgfra	Socs3
Chadl	H2-Ab1	Pex16	Spp1
Chchd6	Hmga1	Pfkl	Srebf1
Col15a1	Hp	Pkm2	Taldo1
Col6a2	Hspg2	Plod1	Timm13
Crip2	Igfbp6	Polr2f	Tomm40
Csnk1g2	Igsf8	Por	Trappc6a
Dag1	Isoc2a	Psmg3	Trf
Dapk3	Isyna1	Ptgis	Tspo
Dctpp1	Klk8	Ptprs	Uba52
Ddt	Lrp1	Rabac1	Uqcr11
Dgat2	Lrp5	Rcn3	Wdr38
Dlk1	Lrrc32	Rgma	Zyx
Ecm1	Ltbp3	Rpl18	
Eef2	Man2b2	Rpl28	

Week 1

U0126   Ptpn11 cKO

B Week 3

U0126   *Ptpn11* cKO*

Up 219 **55** 105

Down 491 **61** 108

A2m	Fabp3	mt-Nd3	Serinc5
Abi3	Fam180a	mt-Nd4	Slc13a5
Adamts18	Fgf21	mt-Nd4l	Slc7a3
Alpl	Fgfr3	mt-Nd5	Slc8a3
Angpt4	Foxd1	mt-Nd6	Smpd3
Bambi	Gm10222	Nog	Sost
Cadm1	Gm8113	Nrbp2	Tcfl5
Cdsn	Gramd2	Panx3	Thbs3
Col11a2	Ifitm1	Pcolce2	Tmie
Col27a1	Mfi2	Phospho1	Vit
Col9a3	mt-Co1	Pth1r	Wnk4
Cyb5r2	mt-Co2	Ptprz1	Wscd2
Drp2	mt-Co3	Ramp1	Zfp185
Extl1	mt-Cytb	Rcor2	

1100001G20Rik	Egln3	Lamb1	Thbd
6330406I15Rik	Emilin2	Loxl1	Thy1
Adamts5	Errfi1	Maff	Tnmd
Alas1	Etv5	Mfap4	Trf
Apoe	F3	Mt2	Ugt1a1
Car9	Fbn1	Nid1	Ugt1a10
Cd36	Fgf10	Penk	Ugt1a2
Cdc42ep2	Gcnt4	Ptn	Ugt1a5
Clec4a1	Gm10393	Rnf150	Ugt1a6b
Col14a1	Gm20425	Rspo2	Ugt1a7c
Col18a1	Gng2	S100a4	Ugt1a8
Col3a1	Hgsnat	S1pr3	Ugt1a9
Col4a1	Hp	Sdpr	Xdh
Cxcl16	Igfbp3	Serpine1	
Cyp1b1	Il1r1	Sfrp1	
Ddah1	Ism1	Sod3	

Week 3

U0126   Ptpn11 cKO

* excluding genes that are also differentially expressed after 4-OHT treatment alone

Figure 3.11: Overlap between transcripts that changed in abundance following SHP2 depletion and MEK1/2 inhibition. Venn diagrams showing the overlap between transcripts that increased (red) or decreased (blue) in abundance after SHP2-depletion or U1026 treatment at w1 (**A**) and w3 (**B**).

Relative expression

Col9a3   Ihh *   Phospho1 *

Spp1 *   Adamts5

Hmga2 *   Egr1 *   Fosl1 *

Figure 3.12: pRT-PCR validation of transcripts that changed in abundance following MEK1/2 inhibition at w3. Graph showing expression, relative to beta-actin, of selected genes in vehicle treated pellets (black) and U0126-treated pellets (grey). For each gene, the expression levels were normalized such that the expression level in vehicle treated pellets equals 1. Error bars indicate standard deviation. Asterisk indicates a significant ($p < 0.05$) difference between Ctrl and U0126-treated pellets.

RNA-seq identifies transcription factors with known roles in chondrocyte maturation, and suggests novel transcription factors that may be involved in this process

To identify the transcription factors that may mediate the perturbation of chondrocyte maturation observed in U0126-treated and SHP2-depleted pellets, we selected the 57 transcription factors that had increased or decreased abundance in U0126-treated or SHP2-depleted pellets at w3, and were unaffected in wild-type pellets treated with 4-OHT (Figure 3.13). We identified several transcription factors that may be responsible for maintaining the expression of genes associated with the pre-hypertrophic and early-hypertrophic stages, and for repressing the expression of genes associated with the late-hypertrophic stage, in U1026-treated or SHP2-depleted pellets. In particular, we observed increased expression of transcription factors known to promote hypertrophy and/or inhibit terminal differentiation (Mef2c, Tcf7, Dlx5, Msx2, Hey1, Sox9), and decreased expression of transcription factors known to inhibit hypertrophy and/or promote terminal differentiation (Foxp1, Rbpj) (Figure 3.14A, Table 3.4). In addition, we identified several transcription factors whose known target genes were also affected by MEK1/2 inhibition or SHP2 depletion, such as *Egr1* and its target genes *Spp1* and *Hpse*, which all decreased in abundance following MEK1/2 inhibition (Figure 3.14B, Table 3.5).   We selected 3 transcription factors for qRT-PCR validation (*Hmga2, Fols1, Egr1*) and verified that they were less abundant following U0126-treatment (Figure 3.12). Thus, we have identified multiple transcription factors with known roles in the growth plate that may act downstream of SHP2 and ERK1/2 during the maturation and terminal differentiation of chondrocytes. This suggests that other transcription factors identified by RNA-seq (Figure 3.13) also have roles in chondrocyte maturation or terminal differentiation.
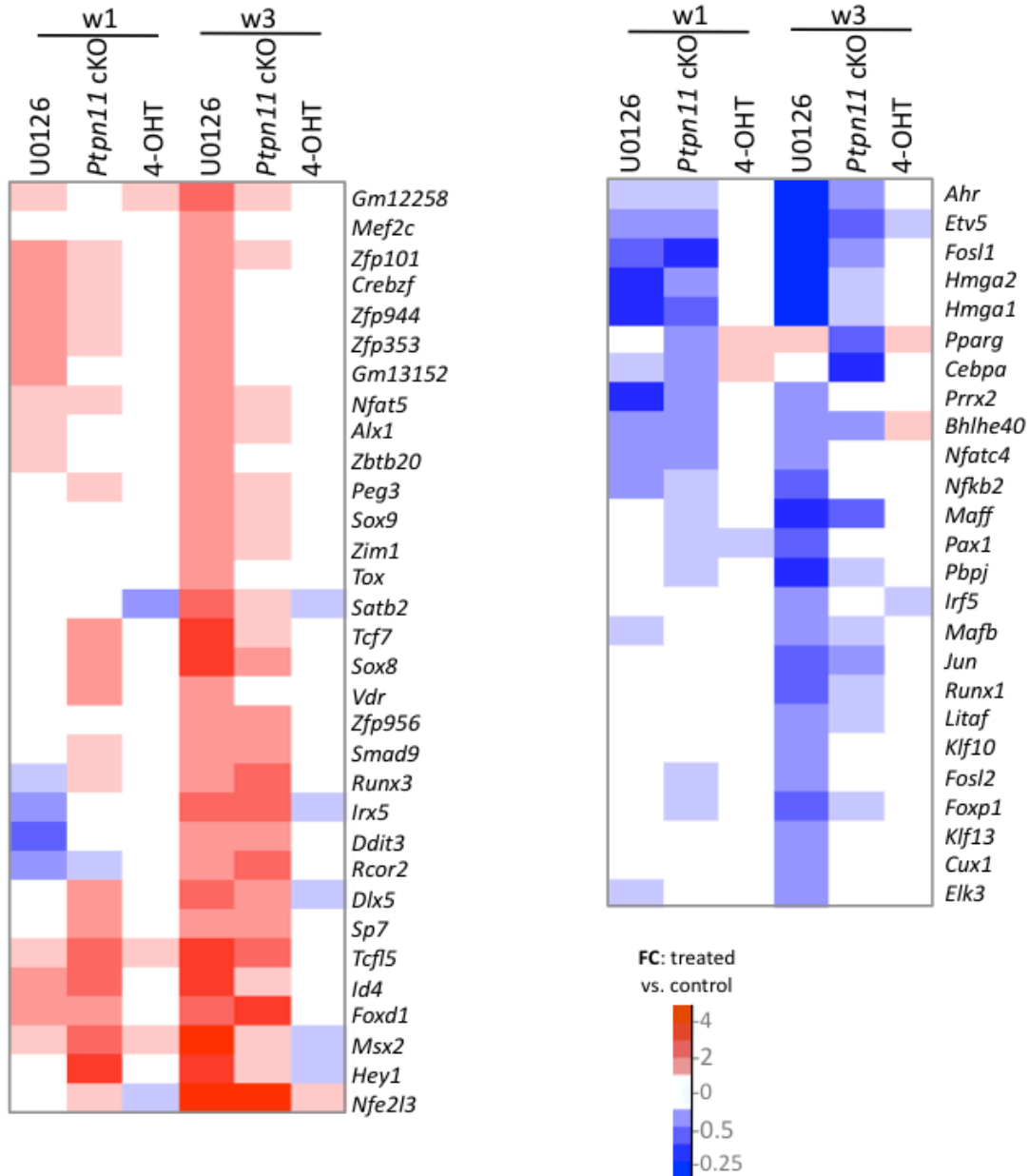
Figure 3.13: MEK1/2- and SHP2-dependent transcription factors. Heatmap showing all 57 transcription factors that were differentially expressed at w3 in *Ptpn11* cKO or U0126-treated pellets. For comparison purposes, the fold change at w1 is also shown, and the fold change in 4-OHT treated pellets is shown.
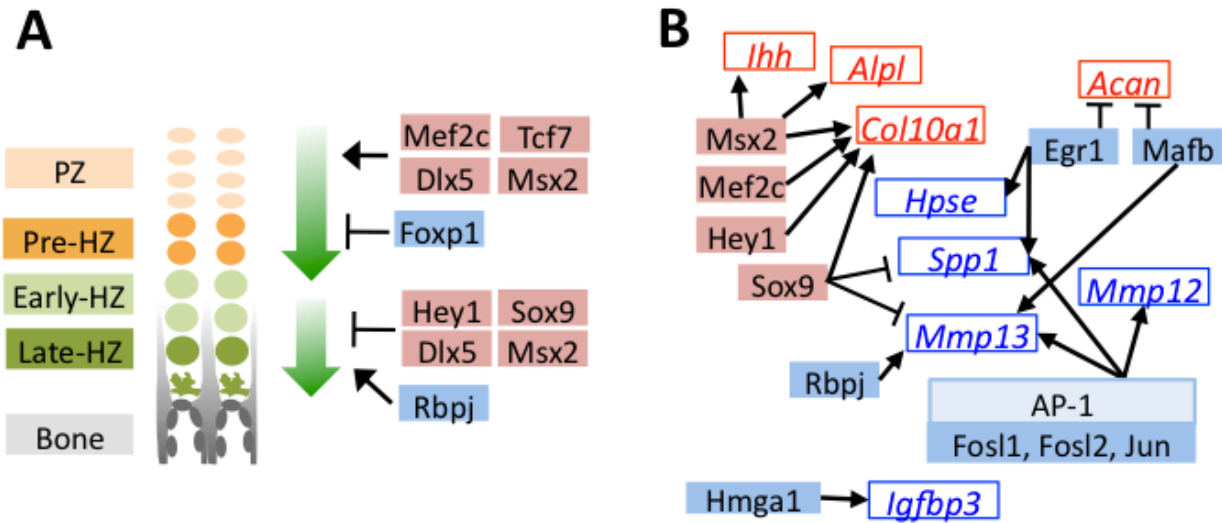
Figure 3.14: MEK1/2 inhibition or SHP2 depletion alters the expression of transcription factors with known roles in chondrocyte maturation. Schematic diagrams of genes encoding transcription factors (solid boxes) or their known target genes (empty boxes) that significantly decreased in abundance (blue) or increased in abundance (red) in U0126-treated pellets at w3. All transcription factors showed similar increases or decreases in *Ptpn11* cKO pellets at w3, but not all reached statistical significance. **A:** Transcription factors with known roles in promoting chondrocyte hypertrophy (upper green arrow), or in promoting the terminal differentiation of chondrocytes and/or their replacement by bone (lower green arrow). **B:** Transcription factors whose known target genes also significantly changed in abundance in U0126-treated pellets. Arrows indicate transcriptional activation while bar-headed lines indicate transcriptional repression.

Table 3.4: Previous studies identifying transcription factors involved in chondrocyte maturation and/or terminal differentiation.

| Transcription factor	Role in the growth plate	Reference
Mef2c		[35]
Tcf7	Promotes hypertrophy	[36]
Dlx5		[37]
Msx2		[38]
Foxp1	Inhibits hypertrophy	[39]
Hey1		[40]
Sox9	Inhibits terminal differentiation or	[22]
Dlx5	cartilage resorption	[37]
Msx2		[38]
Rbpj	Promotes terminal differentiation or cartilage resorption	[41]

**Table 3.5: Genes regulated by MEK1/2-dependent transcription factors**

Transcription factor	Effect on gene expression	Target gene	Reference
Msx2	positive	*Ihh*	[38]
		Alpl	
		Col10a1	
Mef2c	positive	*Col10a1*	[35]
Hey1	positive	*Col10a1*	[40]
Sox9	positive	*Col10a1*	[42]
	negative	*Spp1*	[22]
		Mmp13	
Egr1	positive	*Hpse*	[43]
		Spp1	[44]
	negative	*Acan*	[45]
Mafb	positive	*Mmp13*	[46]
	negative	*Acan*	
AP-1 complex	positive	*Mmp13*	[47]
		Mmp12	[48]
		Spp1	[49]
Rbpj	positive	*Mmp13*	[41]

Mosaic inactivation of *Ptpn11 in vivo* alters chondrocyte maturation, disrupts the normal organization of growth plates, and leads to enchondroma-like lesions that contain mixtures of *Ptpn11* mutant and non-mutant chondrocytes

Since SHP2 depletion perturbed chondrocyte maturation in pellet cultures, we inactivated *Ptpn11 in vivo* and determined the effect on chondrocyte maturation. It has recently been reported that postnatal inactivation of *Ptpn11* in chondrocytes alters the thickness and organization of vertebral growth plates [13], but it was not shown on a molecular level whether the process of chondrocyte maturation was altered. We used a tamoxifen-inducible chondrocyte *Cre* driver (*Col2a1-CreER*) to inactivate *Ptpn11* postnatally. Mice were treated with tamoxifen for 5 days, starting at week 1, and analyzed at weeks 3, 7 and 10. A clear expansion of the vertebral growth plates, osteophyte-like lateral cartilaginous outgrowths, enchondroma-like cartilage islands below the growth plate, and an increase in chondrocyte cell size were observed by weeks 7 and 10 (Figure 3.15A-E). The severity of the phenotype was variable between mice (Figure 3.16). Areas of ectopic ossification in upper regions of some growth plates were observed (Figure
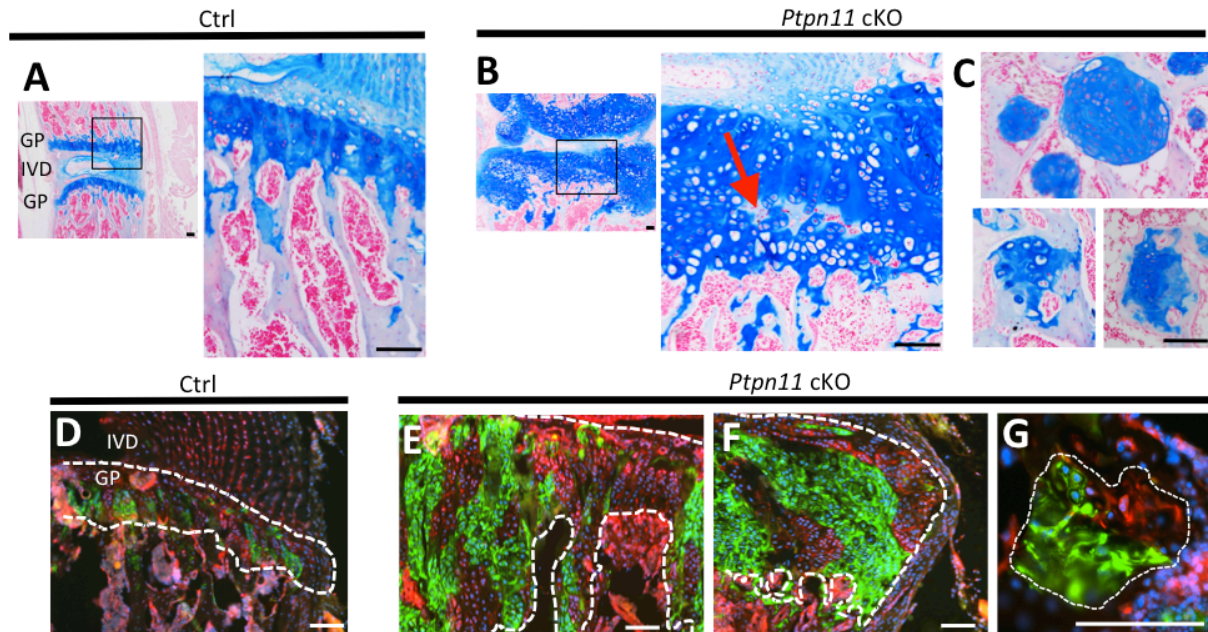
Figure 3.15: Lineage tracing of SHP2-depleted chondrocytes in mouse vertebral growth plates following mosaic postnatal *Ptpn11* inactivation. Tissue sections of lumbar vertebral growth plates (GP) and intervertebral disks (IVD) from *Tg*(*Col2a1-CreER); Rosa26^mTmG* ;*Ptpn11^fll/+^* or *Ptpn11^fll/fl^* mice that had been administered tamoxifen for 5 days, starting at w1, and sacrificed at w7. **A-C**: Tissue sections stained with Alcian Blue and Nuclear Fast Red. In *Ptpn11 cKO* mice, the vertebral growth plates are enlarged and disorganized, with ectopic areas of ossification inside the growth plate (red arrow in B) and enchondroma-like lesions below the growth plate (C). **D-G**: Merged fluorescent images showing cells in which the *mTmG* reporter allele has (green fluorescence) or has not (red fluorescence) been recombined by Cre recombinanse. Note that approximately 50% of growth plate chondrocytes fluoresce green, and in *Ptpn11* cKO mice, the expanded regions of the growth palate (E), lateral outgrowths (F) and enchondroma-like lesions (G) contain both green and red fluorescing cells. Dashed-white lines outline the cartilage. Scale bars = 100 µm.
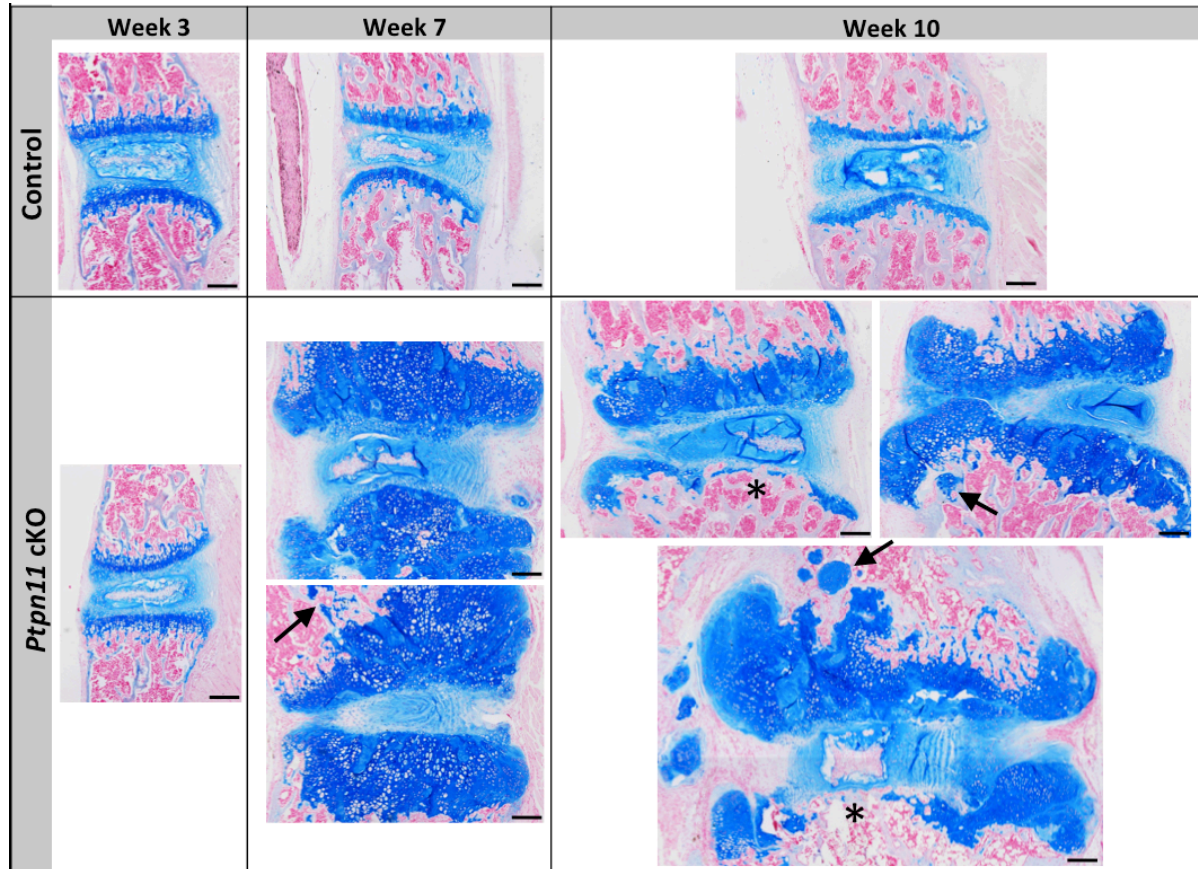
Figure 3.16: Phenotypic range observed in *Ptpn11*-cKO vertebral growth plates. Tissue sections of lumbar vertebral growth plates from *Col2a1-CreER; mTmG* mice that were heterozygous (Ctrl mice) or homozygous (*Ptpn11* cKO mice) for a conditional *Ptpn11* allele and had been administered tamoxifen for 5 days, starting at w1, and sacrificed at week 3, 7 or 10. Tissue sections were stained with Alcian Blue and Nuclear Fast Red. In *Ptpn11 cKO* mice, the vertebral growth plates are substantially enlarged and disorganized at week 7 and 10, with osteophyte-like lateral outgrowths, and ectopic cartilage islands below the growth plate (arrows). By week 10, some growth plates had been completely resorbed (asterisks). Scale bar = 250 µm.

3.15D), and by week 10 some growth plates had been completely resorbed while other growth plates remained expanded (Figure 3.16). Analysis of the *mTmG* reporter allele indicated that 5 days of tamoxifen treatment had induced recombination in ~50% of chondrocytes (Figure 3.15F). Importantly, not all chondrocytes in the cartilage outgrowths and enchondroma-like lesions had evidence of Cre-mediated recombination, suggesting that the lesions contained mixtures of *Ptpn11* mutant and non-mutant chondrocytes (Figure 3.15G,H). Thus, our histological and lineage tracing results indicate that mosaic inactivation of *Ptpn11* in growth plates affects mutant and wild-type chondrocytes; the latter most likely due to altered paracrine signaling.

To determine whether mosaic *Ptpn11* inactivation alters the normal organization of chondrocytes into distinct maturation zones we performed immunohistochemistry (IHC) for type 10 collagen (COLX), which is produced by early-hypertrophic chondrocytes [18]. As expected, in control mice, COLX was localized to the matrix surrounding only the lowest ~3 cells in each chondrocyte column (Figure 3.17A). In contrast, the distribution of COLX was altered in *Ptpn11* cKO mice. Their growth plates had regions in which the height of the COLX-positive zone was expanded to >10 cells (Figure 3.17B). Furthermore, the organization of the COLX-positive cells was altered, with ectopic COLX-positive clusters in upper growth plate zones, ectopic COLX-negative clusters in lower growth plate zones, and COLX-positive chondrocytes in the lateral outgrowths that were arranged into columns oriented perpendicular to the main axis of the growth plate (Figure 3.17B). Both COLX-positive and COLX-negative chondroyctes were detected in the enchondroma-like lesions (Figure 3.17C). These data indicate that the organization of chondrocytes into distinct growth plate zones is disrupted, and that the lateral outgrowths and enchondroma-like lesions contain both mature and immature chondrocytes.

To further characterize the process of chondrocyte maturation, we examined the distribution of phospho-ERK1/2, which is normally detected in the hypertrophic zone [50], and FRA1, a transcription factor encoded by *Fosl1*, which decreased in abundance following SHP2 depletion in pellet cultures (Figure 3.14B). FRA1 and p-ERK1/2 were detected in hypertrophic chondrocytes in Ctrl mice (Figure 3.17A), but were distributed in a scattered, disorganized fashion in the hypertrophic zone of *Ptpn11* cKO mice (Figure 3.17B). In addition, we observed minimal staining for the terminal differentiation marker osteopontin (SPP1) (Figure 3.17D,E) [18], suggesting that the hypertrophic zone was expanded due to an
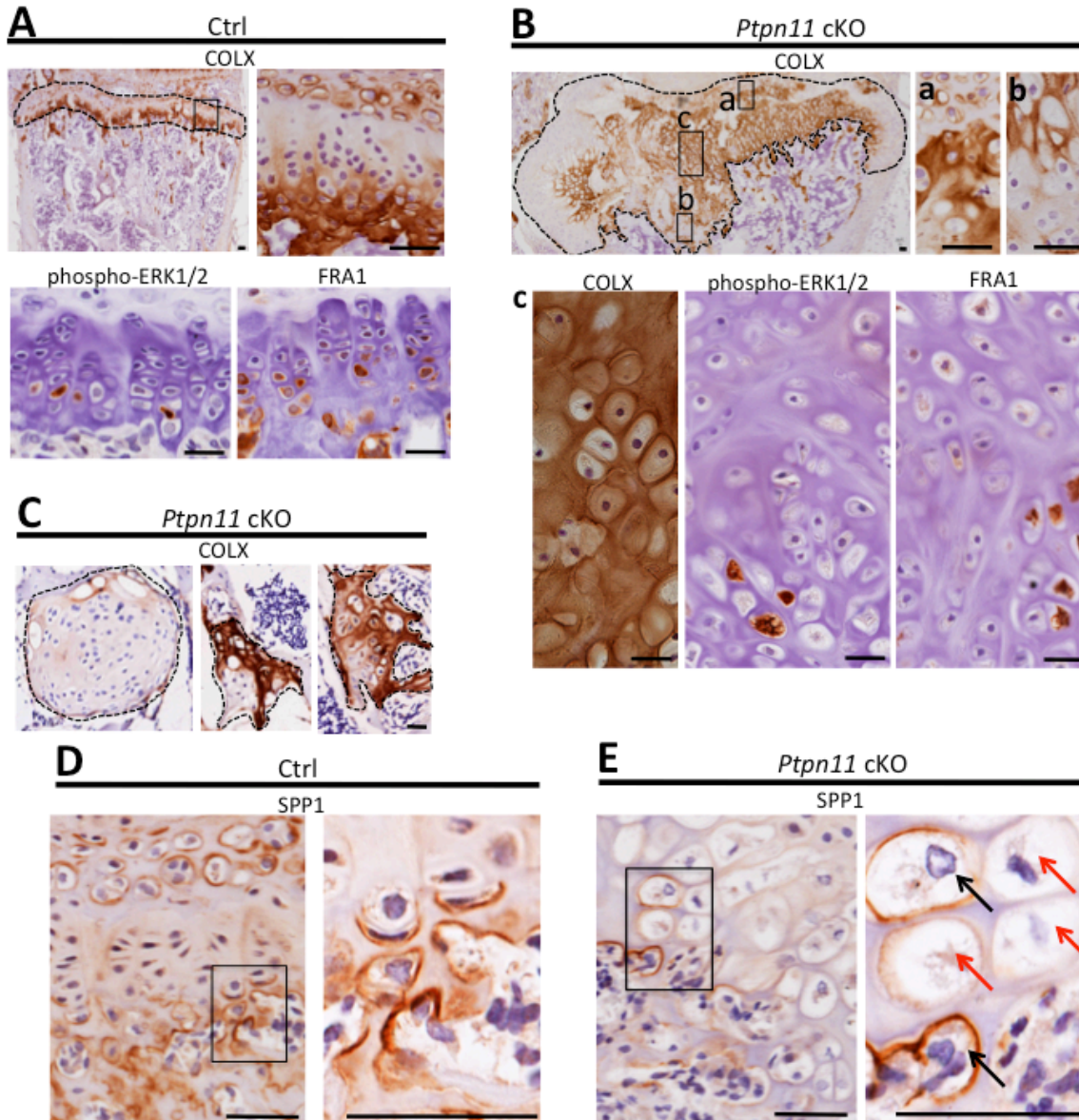
Figure 3.17: Disordered maturation zones and delayed terminal differentiation in mouse vertebral growth plates following mosaic postnatal inactivation of *Ptpn11*. Immunohistochemistry performed on tissue sections from the vertebral growth plates of Ctrl mice (A,D) or the growth plates (B,E) or enchondroma-like lesions (C) in *Ptpn11* cKO mice. **A**: Ctrl growth plate showing COLX, p-ERK1/2 and FRA1 immunoreactivity in the hypertrophic zone. **B**: *Ptpn11* cKO growth plate showing ectopic COLX-positive clusters at the top of the growth plate (a), ectopic COLX-negative clusters at the bottom of the growth plate (b), and disorganized, scattered p-ERK1/2 and FRA1 immunoreactivity in hypertrophic chondrocytes (c). **C**: Enchondroma-like lesions contain both COLX-positive and COLX-negative chondrocytes. **D**: Ctrl growth plate showing SPP1 immunoreactivity at the bottom of the hypertrophic zone. **E**: In *Ptpn11* cKO growth plates, expanded regions of SPP1-negative hypertrophic chondrocytes are observed (red arrows) and only a few scattered chondrocytes hypertrophic chondrocytes are SPP1-positive (black arrows). IHC tissue sections were counterstained with hematoxylin. Scale bars = 25 μm.

increase in the number of chondrocytes that had not yet reached the late-hypertrophic stage. In summary, the normal distribution pattern of stage-specific matrix components and signaling pathway components is altered in *Ptpn11*-cKO growth plates, and the observed delay in the terminal differentiation of hypertrophic chondrocytes is consistent with our pellet culture data.

Disorganized chondrocyte maturation zones are also observed in excised human MC exostoses

Finally, we determined whether chondrocyte maturation is altered in cartilage lesions from MC patients who carry mutations in *PTPN11*. We observed a variable and disorganized distribution of COLX in MC exostoses. In particular, we observed some regions of intense COLX immunoreactivity throughout the matrix (Figure 3.18A-C), while other regions had minimal and scattered COLX immunoreactivity (Figure 3.18D). We compared this to MHE exostoses, which have cartilage caps that display relatively normal growth plate architecture [1]. We found that MHE exostoses have a narrow zone of COLX immunoreactivity at the base of the cartilage cap, as previously reported [51] (Figure 3.18E). Thus, in contrast to MHE exostoses, lesions from MC patients display disorganized chondrocyte maturation zones.
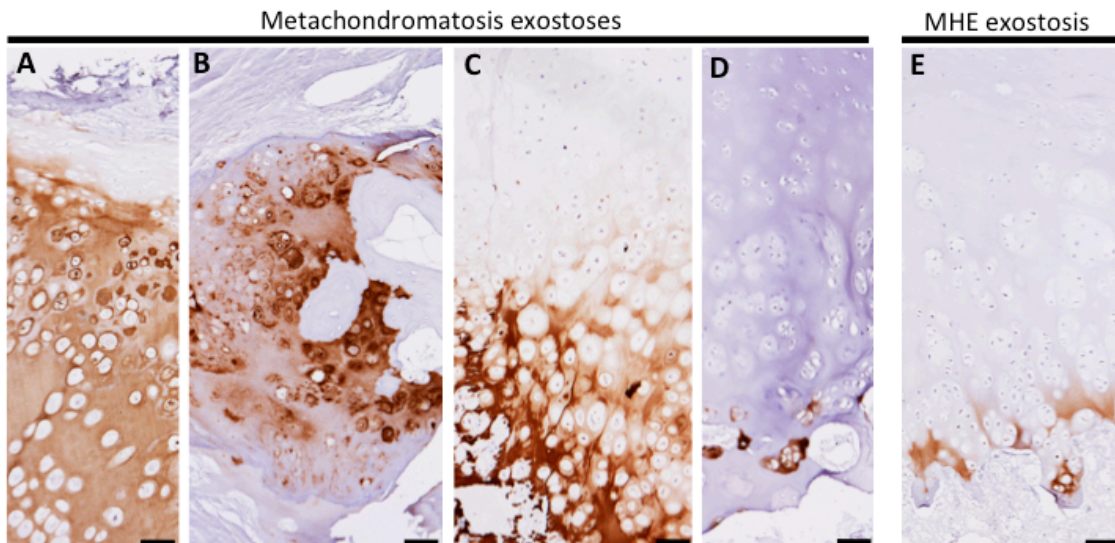
Figure 3.18: Disorganized distribution of COLX in exostoses from MC patients. Immunohistochemical staining of COLX in exostoses from MC patients (**A-D**) or an MHE patient (**E**). Note that the cartilage cap of an MHE exostosis shows a thin layer of COLX staining near the bone-cartilage junction (E). In contrast, in MC exostoses, the distribution of COLX staining is variable and disorganized, with some exostoses showing staining directly below the outer fibrous layer (A,B) or within the center of the lesion (C), while other exostoses showing minimal staining (D). Scale bar = 100 µm

3.6   Discussion

Loss of *PTPN11*/SHP2 causes cartilage tumors in humans and mice [2,11,12]. However, the molecular mechanisms by which SHP2 depletion produces enchondromas and exostoses are incompletely understood. We found that SHP2 depletion reduced ERK1/2 signaling and delayed chondrocyte terminal differentiation. Furthermore, we observed disorganized growth plate architecture in mice and humans that are mosaic for chondrocytes lacking SHP2. Finally, we found that cartilage lesions in mice contain both wild-type and SHP2-deficient cells.

Pellet cultures have been used to study chondrocyte biology at the molecular, biochemical, and histologic levels [16,17,52]. We applied massively parallel RNA-seq technology to comprehensively evaluate the transcriptome of wild-type, SHP2-depleted, and MEK1/2-inhibited pellet cultures. In wild-type cultures we confirmed previously reported changes in gene expression, and we identified new transcripts that likely distinguish proliferative from late-hypertrophic chondrocytes (Figure 3.1). We also determined the effect of 4-hydroxytamoxifen (4-OHT) administration on wild-type transcriptomes, which is an important control experiment since 4-OHT is frequently administered to conditionally alter gene expression *in vitro* and *in vivo* (Figure 3.5). When we depleted SHP2 in pellet cultures, we found that transcripts associated with pre-hypertrophic and early-hypertrophic chondrocytes increased and transcripts associated with late-hypertrophic chondrocytes decreased (Figure 3.9). SHP2 depletion also reduced ERK1/2 phosphorylation, which led us to determine whether the MEK1/2 inhibitor U0126 would have similar effects on the transcriptome of wild-type pellet cultures; it did (Figure 3.5; Figure 3.9). Taken together, our data suggest that SHP2 depletion, acting in part via ERK1/2 signaling, delays the terminal differentiation of hypertrophic chondrocytes (Figure 3.19A). We provided further support for this hypothesis *in vivo* by showing increased numbers of hypertrophic chondrocytes in growth plates of SHP2 mutant mice (Figure 3.17) and in human MC lesions (Figure 3.18). Other studies have also observed accumulation of hypertrophic chondrocytes following *Ptpn11* inactivation in mice [10,12,53].

In addition to regulating terminal differentiation, we found that mosaic inactivation of SHP2 disrupted vertebral growth plate architecture (Figure 3.15; Figure 3.19B). After we inactivated SHP2 in ~50% of growth plate chondrocytes, we observed non-uniform rates of chondrocyte maturation and clusters of chondrocytes surrounded by bone, which resemble enchondromas seen in human MC
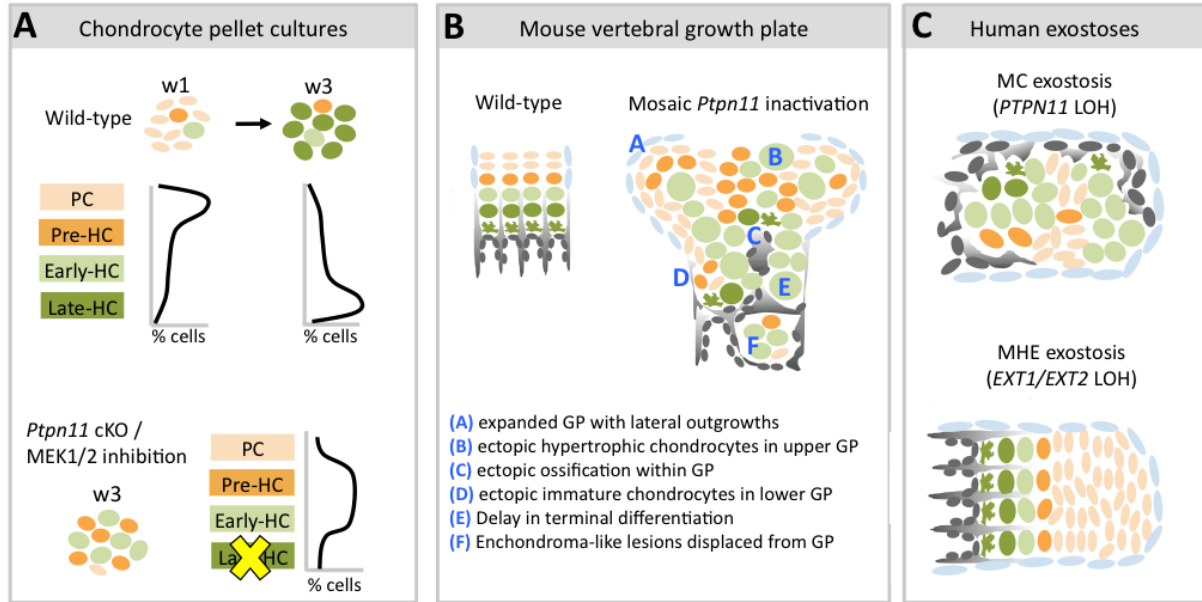
Figure 3.19: Model of the consequences of SHP2 depletion in pellet cultures, mouse vertebral growth plates and human MC exostoses. Schematic diagrams showing chondrocytes at different stages of maturation (PC = proliferative chondrocyte. HC = hypertrophic chondrocyte). **A:** In wild-type pellets, most chondrocytes are in the proliferative stage at w1, and transition to the late-hypertrophic stage by w3. *Ptpn11* cKO or U0126-treated pellets have more pre-HC and early-HC transcripts and fewer late-HC transcripts compared to control pellets at w3. This suggests that SHP2 depletion or MEK1/2 inhibition delays the entry of chondrocytes into the late-hypertrophic stage, thus leading to an accumulation of chondrocytes in the pre-hypertrophic and early-hypertrophic stages. **B:** Mosaic postnatal inactivation of *Ptpn11* in vertebral growth plates disrupts the normal organization of chondrocytes into distinct zones, leads to lateral outgrowths and enchondroma-like lesions. The scarcity of chondrocytes in the late-hypertrophic stage and the abundance of chondrocytes in the early-hypertrophic stage suggest that loss of SHP2 delays the transition from the early to late-hypertrophic stage, thereby delaying terminal differentiation. **C**: Hypertrophic chondrocytes are variably distributed in MC exostoses. This is in contrast to MHE exostoses, which have a cartilage cap that resembles a normal growth plate.

patients. In addition, we observed lateral-outgrowths with abnormal chondrocyte column orientation, which resemble vertebral osteophytes that have also been reported in MC patients [54]. We determined whether the mouse lesions solely contained SHP2 deficient cells by using a reporter allele (*mTmG*) to lineage trace cells that had undergone Cre-mediated inactivation of SHP2. We observed that lesions contained both Cre-recombined and non-recombined cells (Figure 3.15). These data imply that disturbed growth plate architecture is not only due to cell autonomous effects of SHP2 depletion, but also due to altered paracrine signaling between mutant and wild-type cells.

Interestingly, in humans with MC, enchondromas are most often found in the appendicular skeleton [55], whereas in mice, mosaic postnatal inactivation of *Ptpn11* in chondrocytes led to

enchondroma-like lesions in the axial skeleton but not long bones [12,13] (Figure 3.15). Whether this difference is *Col2a1-CreER*-specific, due to differences in the timing of SHP2 loss between humans and mice, or reflective of differences in mechanical forces or signaling pathways in growth plates between mice and humans, remains to be determined.

In pellet cultures SHP2 depletion reduced ERK1/2 phosphorylation (Figure 3.5), and nearly half of all transcripts affected by SHP2 depletion were similarly affected by MEK1/2 inhibition (Figure 3.11). These data suggest ERK1/2 is an important pathway downstream of SHP2 in growth plate chondrocytes. The striking overlap between the effects of SHP2 depletion and MEK1/2 inhibition on transcripts associated with chondrocyte maturation (Figure 3.9) is consistent with previous observations that inactivating *Erk1* and *Erk2* in primary chondrocytes increased the expression of an early-hypertrophic marker (*Col10a1*) and decreased the expression of a late-hypertrophic marker (*Tnfsf11*/RANKL) [14]. Similarly, inactivating *Erk1* and *Erk2 in vivo* delayed cartilage resorption [14,15,50]. These results are consistent with our proposed role for SHP2 in promoting the transition from the early-hypertrophic to the late-hypertrophic stage via the ERK1/2 pathway.

We do not know which receptor tyrosine kinases act upstream of SHP2 and MEK1/2 during chondrocyte terminal differentiation; however, EGFR1, FGFR1 and FGFR3 are likely candidates since inactivation of any one of these receptor leads to an expansion of the hypertrophic zone in mice [56–59]. Furthermore, transcripts encoding several transcription factors had altered abundance after SHP2 depletion or MEK1/2 inhibition (Figure 3.14). Under normal circumstances, these transcription factors likely regulate terminal differentiation in response to SHP2 and ERK1/2 signaling. Of particular interest are Hey1, Msx2, Dlx5 and Sox9, which have all been shown to delay the removal of hypertrophic chondrocytes *in vivo* (Figure 3.14) [22,37,38,40]. Since these four transcription factors are induced by BMP signaling [60–62], and the BMP pathway has been shown to inhibit terminal differentiation [63], it is possible that increased BMP signaling contributes to the delay in terminal differentiation following MEK1/2 inhibition or SHP2 depletion.

Our data, and those of other investigators, implicate aberrant IHH signaling in contributing to the growth plate abnormalities. IHH is a potent secreted regulator of chondrocyte proliferation and differentiation [64]. We observed significantly increased *Ihh* transcript abundance in SHP2-depleted and

MEK1/2-inhibited pellet cultures (Figure 3.9). Similarly, it has previously been reported that knockdown of SHP2 in a chondrogenic cell line increases *Ihh* expression, and exostoses induced by SHP2 depletion have higher levels of *Ihh* expression than normal growth plates [11]. Also, an inhibitor of IHH signaling reduced the growth of cartilage tumors in mosaic SHP2 mutant mice [11]. However, we cannot preclude additional paracrine factors also contributing to tumor formation since transcripts for other secreted growth regulators (SOST, BMP7, BMP8A, WNT5B, WNT10B, and FGF21) exhibited increased abundance in SHP2-depleted pellet cultures. Thus, the pathogenic mechanism(s) that produce cartilage tumors in patients with MC may be complex.

Our data suggest that the exostoses in patients with MC develop via a fundamentally different mechanism than the exostoses in patients with MHE. In both disorders the lesions likely arise from perichondrial cells that have inappropriately differentiated into growth plate chondrocytes [11,12,65]. However, the growth of an MC exostosis is highly disordered and there is little recognizable growth plate morphology (Figure 3.17). In contrast, MHE lesions have well defined cartilage caps and well organized, but aberrantly directed, growth plates [1]. Therefore, *EXT1* and *EXT2* mutations likely predispose to misdirected growth by disorienting otherwise normally functioning chondrocytes, while SHP2 mutations cause misdirected and disorganized growth by creating dysfunctional chondrocytes that signal abnormally (Figure 3.19C). These differences in growth mechanism are consistent with the observation that MHE lesions do not regress but stop growing at puberty when growth plates normally close [1]. Exostoses in MC patients have been reported to regress or disappear over time [1], consistent with our observation that terminal differentiation is delayed rather than completely blocked. Therefore, when SHP2 mutant chondrocytes finally terminally differentiate they will no longer drive abnormal growth. In this regard, it will be interesting to determine whether restoration of SHP2 function will induce terminal differentiation of mutant chondrocytes and hasten regression of these lesions.

Whereas the mechanism of exostosis growth differs between MC and MHE, we do not yet know whether enchondroma formation that occurs in MC, Ollier disease, and Maffucci Syndrome have biologic overlap. We found that mosaic inactivation of SHP2 disrupted the normal organization of maturation zones. This resulted in ectopic ossification within the growth plate, which led to displacement below the growth plate of chondrocytes containing hypertrophic and non-hypertrophic chondrocytes, and SHP2-

deficient and wild-type chondrocytes (Figure 3.15). This suggests MC enchondromas are the result of disrupted growth plate architecture, rather than defective cartilage resorption or neoplastic growth alone. It is not yet known whether mutations in *IDH1* or *IDH2*, which cause Ollier disease and Maffucci syndrome affect growth plate architecture in a similar manner to SHP2 depletion [7]. However, it has been shown that enchondromas in patients with *IDH1* mutations also appear to contain both wild-type and *IDH1*-mutant chondrocytes [7]. Maffucci syndrome patients also develop vascular lesions, suggesting that in vasculature the consequence of SHP2 depletion and *IDH1*/*IDH2* mutation are different.

The RNA-seq datasets we identified in chondrocyte pellet cultures should be useful to other investigators interested in endochondral growth and the *ex vivo* modeling of this process.   For investigators that use tamoxifen to induce Cre-mediated recombination of floxed alleles, the dataset of transcripts whose abundance changed following tamoxifen administration will also be helpful. In conclusion, by coupling *ex vivo* RNA-seq with *in vivo* conditional mouse models, we have added to our understanding of how cartilage tumors arise in patients with MC, as well as the pathways that are regulated by SHP2 during chondrocyte maturation.

3.7   Methods

Mouse Husbandry

The Boston Children's Hospital Institutional Animal Care and Use Committee approved these studies. The *Rosa26^{mTmG}* reporter allele [66] and the *Tg(Col2a1-CreER)* driver [67] were purchased from The Jackson Laboratory. The *Ptpn11* floxed (*Ptpn11$^{fl/fl}$*) and *Tg(CMV-CreERT2)* mice have been previously described [68,69]. For mosaic postnatal inactivation of *Ptpn11* in chondrocytes, 1-week-old *Tg(Col2a1-CreER);Rosa26$^{mTmG/+}$;Ptpn11$^{fl/fl}$* mice were given daily intra-peritoneal injections of 50 mg/kg tamoxifen (Sigma) suspended in corn oil for 5 days.   *Tg(Col2a1-CreER);Rosa26$^{mTmG/+}$;Ptpn11$^{fl/+}$* and *Rosa26$^{mTmG/+}$;Ptpn11$^{fl/fl}$* mice that were similarly treated with tamoxifen served as controls.

Chondrocyte pellet culture

Chondrocytes were harvested from the rib cages of newborn mouse pups as previously described [70]. Pups were either wild-type (mixed 129/FVB background) or, for SHP2 depletion experiments, *Tg(CMV-CreERT2);Ptpn11$^{fl/-}$* on a mixed 129/FVB background. Chondrocytes from 4 to 7

pups from the same litter and with the same genotype were pooled, which provided sufficient cells for all the experimental and control pellets of an individual experiment. Chondrocytes were re-suspended in DMEM/F12 with L-Glutamine (Invitrogen), supplemented with ascorbic acid (50 μg/ml), 1% Penicillin-Streptomycin solution (Life Technologies) and 10% Fetal Bovine Serum (Omega Scientific). Five x 10^5 cells were aliquoted into 0.5 mL eppendorf tubes, centrifuged at 500g for 5 min, and then cultured in 400 μL of medium at 37°C and 5% CO_2. Medium was changed every 1 to 2 days. Stock solutions of 10 mM U0126 (Sigma) dissolved in DMSO, and 1 mM 4-OHT (Sigma) dissolved in 100% EtOH were aliquoted and stored at -20°C. For MEK1/2 inhibition experiments, starting at day 2 of culture, pellets were switched to medium containing 10 μM U0126 or 0.1% DMSO (vehicle control), which was changed daily, and pellets were harvested after 1 or 3 weeks. For SHP2 depletion or 4-OHT-treatment experiments, starting at day 1 pellets were placed in medium containing 1 μM 4-OHT or 0.1% EtOH (vehicle control), and either harvested after 1 week, or treated for an additional week, switched to untreated medium for 1 week, and then harvested (after 3 weeks total in culture).

RNA sequencing library construction and data analysis

Two pellet cultures were pooled prior to performing RNA extraction with the RNeasy mini kit (Qiagen) following the manufacturer's recommendation. RNA extracts, which contained 200 to 1,000 ng RNA were used to prepare one barcoded cDNA library, using the TruSeq RNA Library Preparation Kit (Illumina). Three different barcoded cDNA libraries were made for each genotype, treatment group, and time point. Eight to 12 separate libraries were pooled and used to obtain 50-basepair paired-end reads per single lane of an Illumina HiSeq. Reads were aligned to the mouse genome (mm9) using Tophat [71]. The number of reads mapping to the exons of each gene were calculated using the GenomicFeatures R package [72] and custom scripts (available on request). A trimmed mean method (excluding the 0.5% most and least abundant transcripts) was used to calculate the Pearson's Correlation Coefficient between libraries [73]. Tests for differential gene expression was performed with edgeR [74] and *p* values were adjusted for multiple hypothesis testing. For identification of differentially expressed gene sets, only genes with an RPKM > 3, a fold change of at least 2, and an adjusted *p* value < 0.05 were included. A Chi-squared test was performed to determine whether the overlap between differentially

expressed transcripts at w1 and w3, and in *Ptpn11* cKO pellets and U1026 treated pellets, was significant.

Histology, fluorescence microscopy, and immunohistochemistry

Chondrocyte pellet cultures were fixed in 4% paraformaldehyde (PFA) for 1 hour at room temperature. Mouse vertebrae and long bones were fixed in 4% PFA overnight at 4°C and then decalcified for 14 days in 0.5 M EDTA pH 8 (Apex BioResearch Products). All specimens were then either frozen in OCT embedding media (Tissue Tek) and stored at -80°C, or were stored in 70% EtOH at 4°C and then dehydrated in graded ethanol, cleared in xylene, and embedded in paraffin following standard methods. Paraffin-embedded exostoses were obtained from individuals with MC, who were previously described (Patients A-IV-8, A-IV-5 and B-IV-7 in Bowen *et al* (2011)). For histological analysis, 6 µm paraffin sections were stained with 0.1% Alcian Blue in 0.1 M HCl for 30 minutes and then counterstained with Nuclear Fast Red (Sigma) for 15 minutes. For fluorescence microscopy, frozen sections were mounted in DAPI-Fluoromount (SouthernBiotech). Immunohistochemistry was performed on 6 µm paraffin sections using the Vectastain Elite ABC kit (Vector Labs) following the manufacturer's instructions and using the hybridization buffers, secondary antibodies and visualization reagents provided in the kit. To unmask antigens, tissue sections were first incubated in Citrate Buffer (10 mM citric acid, 0.05% Tween 20, pH 6.0) for 20 min at 90°C. To detect the extracellular proteins Type 10 collagen (COLX), Indian Hedgehog (IHH) and Secreted phosphoprotein 1 (SPP1), sections were also treated with 1mg/ml Hyaluronidase (Sigma) for 1 hr at 37°C and Pepsin Solution (Diagnostic Biosystems) for 15 min at 37°C. The following antibodies and dilutions were used: mouse anti-COLX (X53; Quartett) at 1:100, rabbit anti-FRA1 (D80B4; Cell Signaling) at 1:400; rabbit anti-IHH (N-term; Abgent) at 1:100; rabbit anti-C3aR (H-300; Santa Cruz) at 1:300; rabbit anti-SPP1 (from the laboratory of Dr. Dick Heinegard, Lund University) at 1:3,000; rabbit anti-ERK1/2 (137F5; Cell Signaling) at 1:400; and rabbit anti-phospho-SMAD1/5/8 (Cell Signaling Technology) at 1:400. Primary antibodies were incubated with the tissue sections overnight at 4°C.

Quantitative reverse-transcriptase PCR (qRT-PCR)

Total RNA (500 ng) extracted from the pellet cultures was used to synthesize cDNA with the QuantiTect Reverse Transcription Kit (Qiagen) in a 20 µl reaction volume. qRT-PCR was performed using

SYBR Green JumpStart Taq ReadyMix (Sigma) and 1 µl of cDNA per reaction. Reactions were performed in triplicate. Changes in transcript abundance were calculated using the ddCt method [75] with β-actin (*Actb*) as the reference transcript. *Actb* was selected as a reference transcript because our RNA-seq data indicated that its abundance was not significantly affected by SHP2 depletion, 4-OHT treatment or U0126 treatment. Gene-specific primer pairs are listed in Table 3.6.

Table 3.6: Transcript-specific primers used for q-RT-PCR.

Transcript	Forward Primer	Reverse Primer
Col9a3	AGGTACACCTTGAGGCCCTT	TCCCACTGGGGAACTAGGAG
Ihh	TCAAGGACGAGGAGAACACG	GAGTGATGGCCATCTTCATCC
Phospho1	CTGTTGGTCTCCAGCTGTCA	GCAGAAGCACATCATCCACA
Spp1	TGGCTATAGGATCTGGGTGC	TTGGCAGTGATTTGCTTTTG
Adamts5	GTGTCACATGAATGATGCCC	CGACCCTCAAGAACTTTTGC
Hmga2	TCTCCTGAGCAGGCTTCTTC	AAGGCAGCAAAAACAAGAGC
Egr1	GAGTCGTTTGGCTGGGATAA	CCTTCAATCCTCAAGGGGAG
Fosl1	CTCTTCCTCCTCTGGGCTG	TGGCCTATCCCCAGTACAGT
Actb	CAGCTTCTTTGCAGCTCCTT	AGGAGTCCTTCTGACCCATTC

Immunoblotting

Individual chondrocyte pellet cultures were homogenized in 60 µL of Pierce IP Lysis Buffer (Thermo Scientific) with freshly added Halt Protease and Phosphatase Inhibitor (Thermo Scientific). Twenty µL of lysate was added to 7.5 µL of NuPAGE LDS Sample Buffer and 3 µl NuPAGE Reducing Agent, heated to 70°C for 10 minutes, and loaded onto a 4–20% NuPAGE Bis-Tris Mini Gel (Life Technologies). Following transfer to Invitrolon PVDF membranes using standard procedures, immunodetection was performed using the Western Breeze Chemiluminescent Immunodetection kit (Life Technologies). The following primary antibodies were used: mouse anti-SHP2 (M163; Abcam), rabbit anti-phospho-ERK1/2 (Thr202/Tyr204; D13.14.4E; Cell Signaling), rabbit anti-ERK1/2 (137F5; Cell Signaling), and rabbit anti-phospho-AKT (Ser473;D9E; Cell Signaling). All primary antibodies were diluted to 1:1000 and incubated overnight at 4°C.

3.8   Literature cited

1. Romeo S, Hogendoorn PCW, Dei Tos AP: **Benign cartilaginous tumors of bone: from morphology to somatic and germ-line genetics**. *Adv Anat Pathol* 2009, **16**:307–315.

2. Bowen ME, Boyden ED, Holm IA, Campos-Xavier B, Bonafé L, Superti-Furga A, Ikegawa S, Cormier-Daire V, Bovée JV, Pansuriya TC, et al.: **Loss-of-function mutations in PTPN11 cause metachondromatosis, but not Ollier disease or Maffucci syndrome**. *PLoS Genet.* 2011, **7**:e1002050.

3. Sobreira NLM, Cirulli ET, Avramopoulos D, Wohler E, Oswald GL, Stevens EL, Ge D, Shianna KV, Smith JP, Maia JM, et al.: **Whole-genome sequencing of a single proband together with linkage analysis identifies a Mendelian disease gene**. *PLoS Genet.* 2010, **6**:e1000991.

4. Grossmann KS, Rosário M, Birchmeier C, Birchmeier W: **The tyrosine phosphatase Shp2 in development and cancer**. *Adv. Cancer Res.* 2010, **106**:53–89.

5. Amary MF, Bacsi K, Maggiani F, Damato S, Halai D, Berisha F, Pollock R, O'Donnell P, Grigoriadis A, Diss T, et al.: **IDH1 and IDH2 mutations are frequent events in central chondrosarcoma and central and periosteal chondromas but not in other mesenchymal tumours**. *The Journal of Pathology* 2011, **224**:334–343.

6. Amary MF, Damato S, Halai D, Eskandarpour M, Berisha F, Bonar F, McCarthy S, Fantin VR, Straley KS, Lobo S, et al.: **Ollier disease and Maffucci syndrome are caused by somatic mosaic mutations of IDH1 and IDH2**. *Nature Genetics* 2011, **43**:1262–1265.

7. Pansuriya TC, van Eijk R, d' Adamo P, van Ruler MAJH, Kuijjer ML, Oosting J, Cleton-Jansen A-M, van Oosterwijk JG, Verbeke SLJ, Meijer D, et al.: **Somatic mosaic IDH1 and IDH2 mutations are associated with enchondroma and spindle cell hemangioma in Ollier disease and Maffucci syndrome**. *Nature Genetics* 2011, **43**:1256–1261.

8. Jennes I, Pedrini E, Zuntini M, Mordenti M, Balkassmi S, Asteggiano CG, Casey B, Bakker B, Sangiorgi L, Wuyts W: **Multiple osteochondromas: mutation update and description of the multiple osteochondromas mutation database (MOdb)**. *Hum. Mutat.* 2009, **30**:1620–1627.

9. Yang W, Klaman LD, Chen B, Araki T, Harada H, Thomas SM, George EL, Neel BG: **An Shp2/SFK/Ras/Erk Signaling Pathway Controls Trophoblast Stem Cell Survival**. *Developmental Cell* 2006, **10**:317–327.

10. Bauler TJ, Kamiya N, Lapinski PE, Langewisch E, Mishina Y, Wilkinson JE, Feng G-S, King PD: **Development of severe skeletal defects in induced SHP-2-deficient adult mice: a model of skeletal malformation in humans with SHP-2 mutations**. *Dis Model Mech* 2011, **4**:228–239.

11. Yang W, Wang J, Moore DC, Liang H, Dooner M, Wu Q, Terek R, Chen Q, Ehrlich MG, Quesenberry PJ, et al.: **Ptpn11 deletion in a novel progenitor causes metachondromatosis by inducing hedgehog signalling**. *Nature* 2013, **499**:491–495.

12. Kim HKW, Feng G-S, Chen D, King PD, Kamiya N: **Targeted disruption of Shp2 in chondrocytes leads to metachondromatosis with multiple cartilaginous protrusions**. *J. Bone Miner. Res.* 2013, doi:10.1002/jbmr.2062.

13. Kim HKW, Aruwajoye O, Sucato D, Richards BS, Feng G-S, Chen D, King P, Kamiya N: **Induction of SHP2-deficiency in chondrocytes causes severe scoliosis and kyphosis in mice**. *Spine* 2013, doi:10.1097/BRS.0b013e3182a3d370.

14. Matsushita T, Chan YY, Kawanami A, Balmes G, Landreth GE, Murakami S: **Extracellular Signal-Regulated Kinase 1 (ERK1) and ERK2 Play Essential Roles in Osteoblast Differentiation and in Supporting Osteoclastogenesis**. *Molecular and Cellular Biology* 2009, **29**:5843–5857.

15. Sebastian A, Matsushita T, Kawanami A, Mackem S, Landreth GE, Murakami S: **Genetic inactivation of ERK1 and ERK2 in chondrocytes promotes bone growth and enlarges the spinal canal**. *Journal of Orthopaedic Research* 2011, **29**:375–379.

16. Kato Y, Iwamoto M, Koike T, Suzuki F, Takano Y: **Terminal differentiation and calcification in rabbit chondrocyte cultures grown in centrifuge tubes: regulation by transforming growth factor beta and serum factors**. *Proc. Natl. Acad. Sci. U.S.A.* 1988, **85**:9552–9556.

17. Chen K-S, Tatarczuch L, Ahmed Y, Huang HH, Mirams M, Pagel CN, Mackie EJ: **Identification of light and dark hypertrophic chondrocytes in mouse and rat chondrocyte pellet cultures**. *Tissue and Cell* 2010, **42**:121–128.

18. Nishimura R, Hata K, Ono K, Amano K, Takigawa Y, Wakabayashi M, Takashima R, Yoneda T: **Regulation of endochondral ossification by transcription factors**. *Front. Biosci.* 2012, **17**:2657–2666.

19. Oshlack A, Robinson MD, Young MD: **From RNA-seq reads to differential expression results**. *Genome Biol.* 2010, **11**:220.

20. Zhang H: **Profiling genes expressed in human fetal cartilage using 13,155 expressed sequence tags**. *Osteoarthritis and Cartilage* 2003, **11**:309–319.

21. Keene DR, Oxford JT, Morris NP: **Ultrastructural localization of collagen types II, IX, and XI in the growth plate of human rib and fetal bovine epiphyseal cartilage: type XI collagen is restricted to thin fibrils.** *Journal of Histochemistry & Cytochemistry* 1995, **43**:967–979.

22. Hattori T, Muller C, Gebhard S, Bauer E, Pausch F, Schlund B, Bosl MR, Hess A, Surmann-Schmitt C, von der Mark H, et al.: **SOX9 is a major negative regulator of cartilage vascularization, bone marrow formation and endochondral ossification**. *Development* 2010, **137**:901–911.

23. Naski MC, Colvin JS, Coffin JD, Ornitz DM: **Repression of hedgehog signaling and BMP4 expression in growth plate cartilage by fibroblast growth factor receptor 3**. *Development* 1998, **125**:4977–4988.

24. Shung C-Y, Ota S, Zhou Z-Q, Keene DR, Hurlin PJ: **Disruption of a Sox9- -catenin circuit by mutant Fgfr3 in thanatophoric dysplasia type II**. *Human Molecular Genetics* 2012, **21**:4628–4644.

25. Iwamoto T, Nakamura T, Doyle A, Ishikawa M, de Vega S, Fukumoto S, Yamada Y: **Pannexin 3 Regulates Intracellular ATP/cAMP Levels and Promotes Chondrocyte Differentiation**. *Journal of Biological Chemistry* 2010, **285**:18948–18958.

26. Nishimura R, Wakabayashi M, Hata K, Matsubara T, Honma S, Wakisaka S, Kiyonari H, Shioi G, Yamaguchi A, Tsumaki N, et al.: **Osterix Regulates Calcification and Degradation of Chondrogenic Matrices through Matrix Metalloproteinase 13 (MMP13) Expression in Association with Transcription Factor Runx2 during Endochondral Ossification**. *Journal of Biological Chemistry* 2012, **287**:33179–33190.

27. Houston B, Stewart AJ, Farquharson C: **PHOSPHO1-A novel phosphatase specifically expressed at sites of mineralisation in bone and cartilage**. *Bone* 2004, **34**:629–637.

28. Vu TH, Shipley JM, Bergers G, Berger JE, Helms JA, Hanahan D, Shapiro SD, Senior RM, Werb Z: **MMP-9/gelatinase B is a key regulator of growth plate angiogenesis and apoptosis of hypertrophic chondrocytes**. *Cell* 1998, **93**:411–422.

29. Hu H: **Sequential roles of Hedgehog and Wnt signaling in osteoblast development**. *Development* 2004, **132**:49–60.

30. Brown AJ, Alicknavitch M, D'Souza SS, Daikoku T, Kirn-Safran CB, Marchetti D, Carson DD, Farach-Carson MC: **Heparanase expression and activity influences chondrogenic and osteogenic processes during endochondral bone formation**. *Bone* 2008, **43**:689–699.

31. Makihira S: **Thyroid Hormone Enhances Aggrecanase-2/ADAM-TS5 Expression and Proteoglycan Degradation in Growth Plate Cartilage**. *Endocrinology* 2003, **144**:2480–2488.

32. Little CB, Mittaz L, Belluoccio D, Rogerson FM, Campbell IK, Meeker CT, Bateman JF, Pritchard MA, Fosang AJ: **ADAMTS-1-knockout mice do not exhibit abnormalities in aggrecan turnover in vitro or in vivo**. *Arthritis Rheum.* 2005, **52**:1461–1472.

33. Silvestrini G, Ballanti P, Patacchioli F, Leopizzi M, Gualtieri N, Monnazzi P, Tremante E, Sardella D, Bonucci E: **Detection of osteoprotegerin (OPG) and its ligand (RANKL) mRNA and protein in femur and tibia of the rat**. *J. Mol. Histol.* 2005, **36**:59–67.

34. Chen J, Shapiro HS, Sodek J: **Developmental expression of bone sialoprotein mRNA in rat mineralized connective tissues**. *Journal of Bone and Mineral Research* 2009, **7**:987–997.

35. Arnold MA, Kim Y, Czubryt MP, Phan D, McAnally J, Qi X, Shelton JM, Richardson JA, Bassel-Duby R, Olson EN: **MEF2C transcription factor controls chondrocyte hypertrophy and bone development**. *Dev. Cell* 2007, **12**:377–389.

36. Mikasa M, Rokutanda S, Komori H, Ito K, Tsang YS, Date Y, Yoshida CA, Komori T: **Regulation of Tcf7 by Runx2 in chondrocyte maturation and proliferation**. *Journal of Bone and Mineral Metabolism* 2010, **29**:291–299.

37. Ferrari D, Kosher RA: **Dlx5 Is a Positive Regulator of Chondrocyte Differentiation during Endochondral Ossification**. *Developmental Biology* 2002, **252**:257–270.

38. Amano K, Ichida F, Sugita A, Hata K, Wada M, Takigawa Y, Nakanishi M, Kogo M, Nishimura R, Yoneda T: **Msx2 Stimulates Chondrocyte Maturation by Controlling Ihh Expression**. *Journal of Biological Chemistry* 2008, **283**:29513–29521.

39. Guo X, Zhao H: **Foxp1/2/4, New Transcriptional Regulators for the Chondrocyte Hypertrophy and Osteoblast Differentiation during Skeletal Ossification**. 2012.

40. Salie R, Kneissel M, Vukevic M, Zamurovic N, Kramer I, Evans G, Gerwin N, Mueller M, Kinzel B, Susa M: **Ubiquitous overexpression of Hey1 transcription factor leads to osteopenia and chondrocyte hypertrophy in bone**. *Bone* 2010, **46**:680–694.

41. Kohn A, Dong Y, Mirando AJ, Jesse AM, Honjo T, Zuscik MJ, O'Keefe RJ, Hilton MJ: **Cartilage-specific RBPj -dependent and -independent Notch signals regulate cartilage and bone development**. *Development* 2012, **139**:1198–1212.

42. Dy P, Wang W, Bhattaram P, Wang Q, Wang L, Ballock RT, Lefebvre V: **Sox9 Directs Hypertrophic Maturation and Blocks Osteoblast Differentiation of Growth Plate Chondrocytes**. *Developmental Cell* 2012, **22**:597–609.

43. de Mestre AM, Rao S, Hornby JR, Soe-Htwe T, Khachigian LM, Hulett MD: **Early growth response gene 1 (EGR1) regulates heparanase gene transcription in tumor cells**. *J. Biol. Chem.* 2005, **280**:35136–35147.

44. Liu Q-F, Yu H-W, Liu G-N: **Egr-1 upregulates OPN through direct binding to its promoter and OPN upregulates Egr-1 via the ERK pathway**. *Mol. Cell. Biochem.* 2009, **332**:77–84.

45. Rockel JS, Bernier SM, Leask A: **Egr-1 inhibits the expression of extracellular matrix genes in chondrocytes by TNFalpha-induced MEK/ERK signalling**. *Arthritis Res. Ther.* 2009, **11**:R8.

46. Zhang Y, Ross AC: **Retinoic acid and the transcription factor MafB act together and differentially to regulate aggrecan and matrix metalloproteinase gene expression in neonatal chondrocytes**. *Journal of Cellular Biochemistry* 2013, **114**:471–479.

47. Ahmad R, Sylvester J, Zafarullah M: **MyD88, IRAK1 and TRAF6 knockdown in human chondrocytes inhibits interleukin-1-induced matrix metalloproteinase-13 gene expression and promoter activity by impairing MAP kinase activation**. *Cell. Signal.* 2007, **19**:2549–2557.

48. Wu L, Tanimoto A, Murata Y, Fan J, Sasaguri Y, Watanabe T: **Induction of human matrix metalloproteinase-12 gene transcriptional activity by GM-CSF requires the AP-1 binding site in human U937 monocytic cells**. *Biochem. Biophys. Res. Commun.* 2001, **285**:300–307.

49. Sharma P, Kumar S, Kundu GC: **Transcriptional regulation of human osteopontin promoter by histone deacetylase inhibitor, trichostatin A in cervical cancer cells**. *Mol. Cancer* 2010, **9**:178.

50. Miedlich SU, Zalutskaya A, Zhu ED, Demay MB: **Phosphate-induced Apoptosis of Hypertrophic Chondrocytes Is Associated with a Decrease in Mitochondrial Membrane Potential and Is Dependent upon Erk1/2 Phosphorylation**. *Journal of Biological Chemistry* 2010, **285**:18270–18275.

51. de Andrea CE, Wiweger MI, Bovée JVMG, Romeo S, Hogendoorn PCW: **Peripheral chondrosarcoma progression is associated with increased type X collagen and vascularisation**. *Virchows Arch.* 2012, **460**:95–102.

52. Okubo Y, Reddi AH: **Thyroxine downregulates Sox9 and promotes chondrocyte hypertrophy**. *Biochem. Biophys. Res. Commun.* 2003, **306**:186–190.

53. Lapinski PE, Meyer MF, Feng G-S, Kamiya N, King PD: **Deletion of SHP-2 in mesenchymal stem cells causes growth retardation, limb and chest deformity and calvarial defects in mice**. *Dis Model Mech* 2013, doi:10.1242/dmm.012849.

54. Herman TE, Chines A, McAlister WH, Gottesman GS, Eddy MC, Whyte MP: **Metachondromatosis: report of a family with facial features mildly resembling trichorhinophalangeal syndromePediatr Radiol 1997 Nov;27(11):864**. *Pediatr Radiol* 1997, **27**:436–441.

55. Kennedy LA: **Metachondromatosis**. *Radiology* 1983, **148**:117–118.

56. Jacob AL, Smith C, Partanen J, Ornitz DM: **Fibroblast growth factor receptor 1 signaling in the osteo-chondrogenic cell lineage regulates sequential steps of osteoblast maturation**. *Dev. Biol.* 2006, **296**:315–328.

57. Wang K, Yamamoto H, Chin JR, Werb Z, Vu TH: **Epidermal growth factor receptor-deficient mice have delayed primary endochondral ossification because of defective osteoclast recruitment**. *J. Biol. Chem.* 2004, **279**:53848–53856.

58. Colvin JS, Bohne BA, Harding GW, McEwen DG, Ornitz DM: **Skeletal overgrowth and deafness in mice lacking fibroblast growth factor receptor 3**. *Nat. Genet.* 1996, **12**:390–397.

59. Deng C, Wynshaw-Boris A, Zhou F, Kuo A, Leder P: **Fibroblast growth factor receptor 3 is a negative regulator of bone growth**. *Cell* 1996, **84**:911–921.

60. Liu T, Gao Y, Sakamoto K, Minamizato T, Furukawa K, Tsukazaki T, Shibata Y, Bessho K, Komori T, Yamaguchi A: **BMP-2 promotes differentiation of osteoblasts and chondroblasts inRunx2-deficient cell lines**. *Journal of Cellular Physiology* 2007, **211**:728–735.

61. Pan Q, Yu Y, Chen Q, Li C, Wu H, Wan Y, Ma J, Sun F: **Sox9, a key transcription factor of bone morphogenetic protein-2-induced chondrogenesis, is activated through BMP pathway and a CCAAT box in the proximal promoter**. *J. Cell. Physiol.* 2008, **217**:228–241.

62. Holleville N, Quilhac A, Bontoux M, Monsoro-Burq AH: **BMP signals regulate Dlx5 during early avian skull development**. *Dev. Biol.* 2003, **257**:177–189.

63. Minina E, Wenzel HM, Kreschel C, Karp S, Gaffield W, McMahon AP, Vortkamp A: **BMP and Ihh/PTHrP signaling interact to coordinate chondrocyte proliferation and differentiation**. *Development* 2001, **128**:4523–4534.

64. Lai LP, Mitchell J: **Indian hedgehog: Its roles and regulation in endochondral bone development**. *Journal of Cellular Biochemistry* 2005, **96**:1163–1173.

65. Huegel J, Mundy C, Sgariglia F, Nygren P, Billings PC, Yamaguchi Y, Koyama E, Pacifici M: **Perichondrium phenotype and border function are regulated by Ext1 and heparan sulfate in developing long bones: A mechanism likely deranged in Hereditary Multiple Exostoses**. *Dev. Biol.* 2013, doi:10.1016/j.ydbio.2013.02.008.

66. Muzumdar MD, Tasic B, Miyamichi K, Li L, Luo L: **A global double-fluorescent Cre reporter mouse**. *Genesis* 2007, **45**:593–605.

67. Nakamura E, Nguyen M-T, Mackem S: **Kinetics of tamoxifen-regulated Cre activity in mice using a cartilage-specific CreER(T) to assay temporal activity windows along the proximodistal limb skeleton**. *Dev. Dyn.* 2006, **235**:2603–2612.

68. Zhang SQ, Yang W, Kontaridis MI, Bivona TG, Wen G, Araki T, Luo J, Thompson JA, Schraven BL, Philips MR, et al.: **Shp2 Regulates Src Family Kinase Activity and Ras/Erk Activation by Controlling Csk Recruitment**. *Molecular Cell* 2004, **13**:341–355.

69. Hayashi S, McMahon AP: **Efficient recombination in diverse tissues by a tamoxifen-inducible form of Cre: a tool for temporally regulated gene activation/inactivation in the mouse**. *Dev. Biol.* 2002, **244**:305–318.

70. Gosset M, Berenbaum F, Thirion S, Jacques C: **Primary culture and phenotyping of murine chondrocytes**. *Nature Protocols* 2008, **3**:1253–1260.

71. Trapnell C, Pachter L, Salzberg SL: **TopHat: discovering splice junctions with RNA-Seq**. *Bioinformatics* 2009, **25**:1105–1111.

72. Carlson M, Pages H, Aboyoun S, Falcon S, Morgan M, Sarkar D, Lawrence M: *GenomicFeatures: Tools for making and manipulating transcript centric annotations. R package version 1.8.2.* [date unknown].

73. Ayturk UM, Jacobsen CM, Christodoulou DC, Gorham J, Seidman JG, Seidman CE, Robling AG, Warman ML: **An RNA-seq protocol to identify mRNA expression changes in mouse diaphyseal bone: Applications in mice with bone property altering Lrp5 mutations**. *Journal of Bone and Mineral Research* 2013, doi:10.1002/jbmr.1946.

74. Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data**. *Bioinformatics* 2010, **26**:139–140.

75. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method**. *Methods* 2001, **25**:402–408.

Chapter **4**

Efficient mapping and cloning of mutations in zebrafish by low-coverage whole-genome sequencing

4.1 Abstract

The generation and analysis of mutants in zebrafish has been instrumental in defining the genetic regulation of vertebrate development, physiology, and disease. However, identifying the genetic changes that underlie mutant phenotypes remains a significant bottleneck in the analysis of mutants. Whole-genome sequencing has recently emerged as a fast and efficient approach for identifying mutations in nonvertebrate model organisms. However, this approach has not been applied to zebrafish due to the complicating factors of having a large genome and lack of fully inbred lines. Here we provide a method for efficiently mapping and detecting mutations in zebrafish using these new parallel sequencing technologies. This method utilizes an extensive reference SNP database to define regions of homozygosity-by-descent by low coverage, whole-genome sequencing of pooled DNA from only a limited number of mutant F_2 fish. With this approach we mapped each of the five different zebrafish mutants we sequenced and identified likely causative nonsense mutations in two and candidate mutations in the remainder. Furthermore, we provide evidence that one of the identified mutations, a nonsense mutation in *bmp1a*, underlies the *welded* mutant phenotype.

4.2   Contributing authors

Bowen ME[1,2,*], Henke K[1,2,*], Siegfried KR[1], Warman ML[1,2], Harris MP[1,2].

1. Orthopedic Research Laboratories, Children's Hospital, Boston, Massachusetts 02115

2. Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115

* These authors contributed equally

4.3   Specific contributions

Candidate's contributions

Genomic DNA library preparation; Development of scripts and pipeline for analyzing whole genome sequence data.

Other author's contributions

Zebrafish mutagenesis screen; Maintenance of zebrafish mutants; DNA extraction; Microsattelite-based linkage analysis; Morpholino Knockdown; Sanger sequencing to confirm variants.

4.4   Introduction

A major strength of the zebrafish (*Danio rerio*) model is the feasibility of performing large-scale genetic screens as a means to isolate mutants to study gene function. Such forward genetic screens have led to the identification of a large collection of mutants defective in a variety of biological processes. The standard approach for identifying the responsible mutation underlying a mutant phenotype is to perform bulked segregant analysis with simple sequence length polymorphisms (SSLPs) [1], followed by fine mapping using individual fish to define the region in which the mutation lies. Candidate genes within the mapped interval are then screened for the presence of mutations, typically by sequencing cDNA or genomic DNA. This approach is time and labor intensive, requiring large numbers of mutant fish and often years to successfully clone a mutant. To date this is a major limitation in zebrafish research, and large numbers of mutants have not yet been mapped or cloned.

Whole-genome sequencing (WGS) has the potential to expedite the process of mutation detection in zebrafish. In *Caenorhabditis elegans* and *Arabidopsis thaliana*, multiple studies have shown that, by pooling from 10 to 500 recombinant progeny and sequencing to a relatively high depth, a linked region between 0.5 and 5 Mb in size as well as the responsible mutation can be identified [2–7]. Similarly, WGS of individual mutant mice [8] or human patients with genetic disorders [9] has led to the identification of causative mutations. However, in the case of mice and humans, prior knowledge of linkage was necessary to determine which of the many sequence variants identified in the genome were associated with the phenotype.

Mapping mutants by performing WGS has not yet been applied to zebrafish. One prohibitive factor has been the high cost of sequencing an entire zebrafish genome (~1.5 Gb, compared to 100 Mb for *C. elegans* and 120 Mb for *A. thaliana*). However, new sequencing platforms have increased the throughput of sequencing and reduced its cost, now making it practical to obtain low-coverage sequence data of an entire zebrafish genome. A second prohibitive factor for applying WGS for mutation detection in zebrafish is the high level of inter- and intrastrain variation [10–13] and the absence of a well-annotated catalog of natural variation; consequently, this makes it more difficult to determine whether a novel homozygous variant is a causative mutation or a low-frequency single-nucleotide polymorphism (SNP). This contrasts with the inbred organisms for which WGS has been successfully applied [2–7]. Here, we

describe the establishment of an extensive zebrafish SNP database. Using this database in combination with low-coverage (~3×) WGS, we developed a rapid and inexpensive method to efficiently map, and frequently clone, recessive mutations in zebrafish. Furthermore, the methodology described here can be used to identify genetic loci in other model organisms with larger and highly polymorphic genomes that have annotated genomes, such as rats, mice, dogs, chickens, and other fish species.

4.5   Results

Sequencing libraries generated from pooled DNA

We performed WGS on five previously uncharacterized mutants isolated in an ENU mutagenesis screen for adult phenotypes (ZF-MODELS; Tübingen, Germany, 2004). These recessive mutants, generated in the Tü background, were outcrossed to a polymorphic mapping strain (WIK or TLF) (Figure 4.1); progeny from F_1 intercrosses were phenotyped and frozen for analysis. We pooled DNA from 20 affected F_2 fish from each mutant, mixing, when possible, individuals from several independent F_1 intercrosses. The F_2 fish used often stemmed from either one or two parental (P_0) crosses for a particular mutant, thus limiting the total genomic variation within a pool. Whole-genome sequencing was performed on genomic DNA libraries constructed from each mutant pool, resulting in between 60 and 83 million 100-bp reads per library (Table 4.1). We obtained between 2.6× and 4.1× coverage of the genome per mutant, after excluding 2–9% of the reads that were potential PCR duplicates (reads with identical 5′-end coordinates) and ~25% of reads that failed to map to unique locations in the reference genome (Zv9).

We also sequenced the genomes of four routinely used wild-type strains to establish a database of existing SNP variation. This information enabled us to predict the parental origin of SNP alleles in our mutant pools. Tü and WIK strains are commonly used in laboratories around the world. To assess the diversity among parental strains, we generated WGS from lines maintained at Children's Hospital Boston (TüB and WIKB) and the Max Planck Institute in Tübingen, Germany (TüG and WIKG). DNA libraries, constructed from pooled DNA from 20 fish for each of the TüB, WIKB, TüG, WIKG, TLF, and AB lines, were sequenced and 3.8× to 5.1× average genome coverage was obtained (Table 4.1).
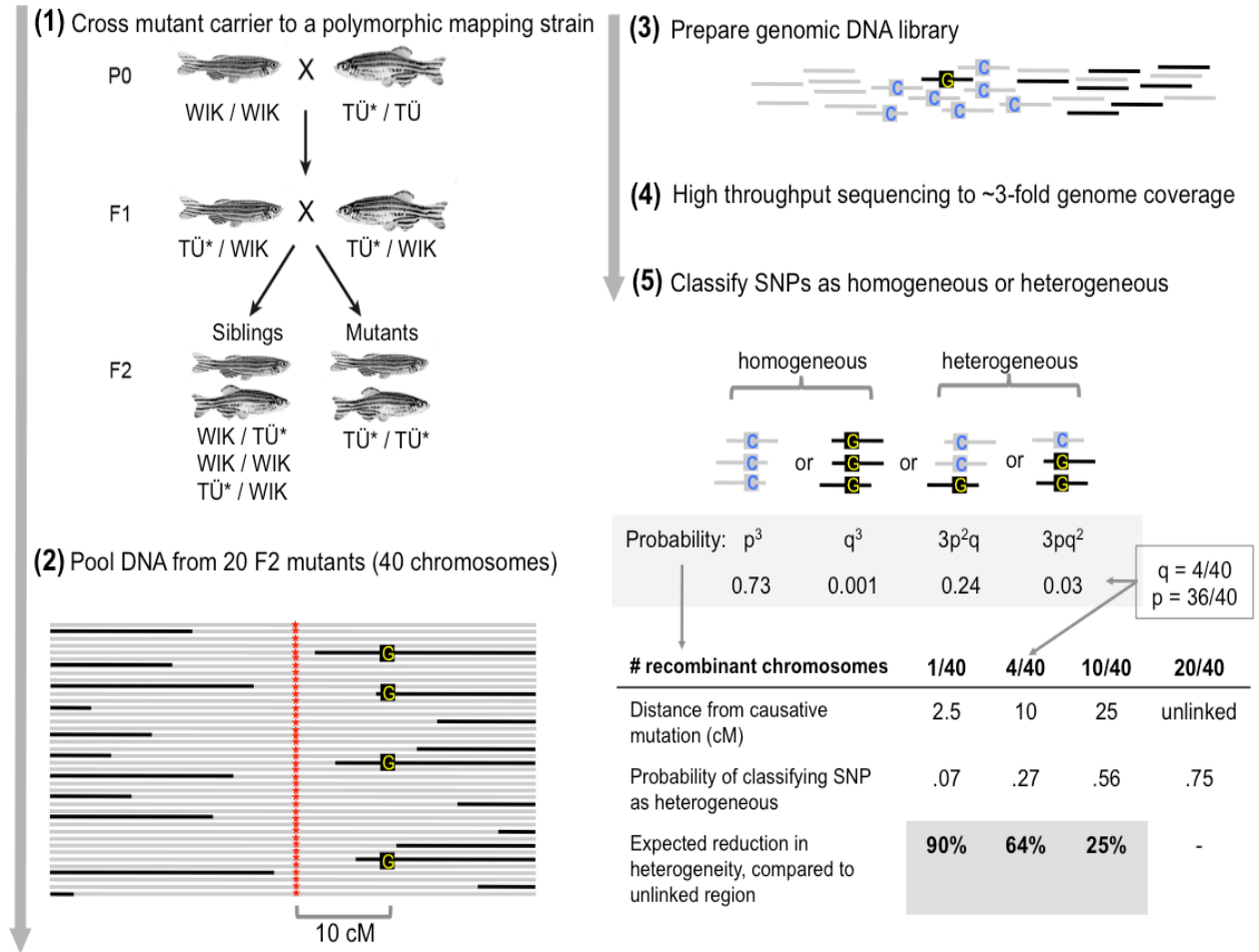
Figure 4.1: Outline and predicted sensitivity of the approach. (**1**) The generation of fish used in mapping is accomplished by crossing identified mutants carrying a recessive ENU-induced mutation (*) within the Tü background, to a polymorphic mapping strain, (e.g., WIK). Mutant carriers (Tü*/WIK) of the F1 generation are then intercrossed to generate F2 progeny. These F2 fish are sorted based on the presence or absence of the mutant phenotype. (**2**) DNA is prepared from 20 F2 mutant progeny (Tü*/Tü*) and pooled in equal quantities. The diagram depicts the 40 chromosomes containing a phenotype-causing ENU-induced mutation (red asterisk) among the 20 mutant fish. The mutation is linked to genomic sequence originating from the Tü strain used for mutagenesis (grey fragments). Recombinants having sequence originating from the outcross strain (black fragments) can be observed at different distances from the causative mutation as a result of meiotic recombination during meiosis in the F1 generation. In a SNP located ~10 cM from the causative mutation, we expect by definition, 4 of the 40 mutation-containing chromosomes to show a mapping strain allele (G in WIK; black square) as a result of meiotic recombination. (**3**) Physical fractionation of DNA from the 20 mutant fish produces DNA fragments, that contain the aforementioned SNP (boxed C for the Tü and G for the WIK alleles), (**4**) Whole genome sequencing of the fragmented DNA library is performed on a single lane of an Illumina HiSeq platform resulting in ~3x genome coverage. (**5**) Probability for detecting a SNP as being homogenous or heterogeneous in pooled DNA from 20 mutant fish sequenced to 3x coverage. SNPs are classified as homogeneous if all 3 reads covering the SNP represent the same allele (probability = $p^3 + q^3$) and as heterogeneous if both alleles are represented (probability = $3p^2q + 3pq^2$). In an unlinked region, where both alleles are equally represented (q = 0.5; p = 0.5), the probability of a SNP being detected as heterogeneous is 0.75. Likewise, in regions where 10% of the chromosomes are recombinant, as in our

(continued: Figure 4.1) example, statistically 4 out of 40 (q = 0.1) reads would show the mapping-strain allele (G), while 36 out of 40 (p = 0.9) would show the reference allele (C; Tü). Thus the probability of detecting the SNP in a heterogeneous state is 0.27. Therefore, the number of heterogeneous SNPs identified in such a region is expected to be ~64% lower than in an unlinked region (0.27/0.75 = 64%). Similarly, a ~90% reduction in heterogeneity is expected for regions containing 1 recombinant chromosome, while a ~25% reduction is expected for regions with 10 recombinant chromosomes. According to this analysis using low genome coverage, it would be of no added benefit to pool larger numbers of fish to increase the resolution of mapping. As the probability of detecting a single recombinant in, for example, 40 fish (80 chromosomes) would be lower than the level of detecting false positive heterogeneous SNPs and thus indistinguishable from noise.

Table 4.1: Whole genome sequencing of pooled DNA from wild-type zebrafish strains

Wild-type pool	# reads[a] (millions)	Genome coverage	# SNPs[b] (per kb)	% het[c]
Tü[B]	97	5.1x	2.6	88
Tü[G]	81	4.1x	2.5	89
WIK[B]	96	4.1x	2.9	64
WIK[G]	81	4.0x	3.5	73
AB	90	4.6x	3.3	79
TLF	91	3.8x	2.9	56

[a]Number of 100 bp reads obtained by Illumina single end sequencing.
[b]Number of positions at which at least one read representing an alternate allele was observed. Only positions at the 7.6 million SNP sites identified in this study were considered.
[d]Percentage of SNPs that were heterogeneous (i.e., both reference-genome and alternate alleles were represented)

Establishment of a reference SNP database

With low-coverage sequencing of pooled DNA, it is challenging to distinguish true SNPs from sequencing errors as many variants are represented by only a single sequencing read. However, if the same variant is observed in more than one strain, it is more likely to be a real SNP than a sequencing error. Therefore, to enhance the accuracy of SNP detection we combined the WGS data from all wild-type strains and mutants, resulting in 50× genome coverage, and then selected only the variants that were present in at least three reads for inclusion in our SNP database (see *Materials and Methods* for filtering criteria). Although variants present in only one or two reads in the combined data could also represent real SNPs, many are likely to represent sequencing errors or alignment artifacts and therefore were not included in the database. In total, we identified a set of 7.6 million SNPs (http://www.fishyskeleton.com),

which is substantially greater than the 0.7 million zebrafish SNPs currently annotated in publically available databases. Of the SNPs in public databases, 85% were detected in at least one read in our sequence data, and 45% had been included in our SNP database since they met all filtering criteria (such as being present in at least three reads). Importantly, 7.3 million of the SNPs we identified were not annotated, thus vastly expanding our knowledge of genetic variation in zebrafish. Using the individual WGS data from pooled DNA for each mutant and wild-type strain, we were then able to classify each SNP within that sequence as being either heterogeneous (at least one read representing each SNP allele was observed) or homogeneous (all reads represented the same allele). In each pool, an average of ~2 heterogeneous and ~3 homogeneous SNPs were observed per kilobase of genomic sequence (Table 4.2).

Table 4.2: Classifying SNPs identified by whole genome sequencing of pooled DNA from zebrafish mutants

Mutant pool	SNP genotype[a] (average per kb)				Parental origin of alleles[b] (average per kb)	
	Het	Hom		n/d	Mapping strain allele	Reference genome allele
		Non-ref	Ref			
moto	1.6	0.9	1.8	1.3	0.6	0.6
wdd	1.0	0.4	2.1	2.2	0.3	0.3
hlw	1.7	0.4	2.4	1.0	0.4	0.8
frnt	2.1	0.6	2.1	0.7	0.6	0.7
sump	1.8	1.1	2.0	0.6	0.7	0.6

[a]Calculated for the 7.6 million SNP sites identified in this study. SNPs were defined as heterogeneous (Het) for sites at which both a reference genome allele (Zv9) and an alternate allele were observed in the WGS of pooled DNA. SNPs were defined as homogeneous (Hom) for sites at which only the alternate allele (non-ref) or the reference genome allele (ref) were observed; sites that were covered by less than 2 sequencing reads were deemed uninformative (n/d).
[b]Calculated for all SNP sites at which an alternate allele was present in the TLF or WIK mapping strain, but not in the Tü strain (0.7 million and 1.2 million sites respectively); Mapping strain allele = at least one read representing the alternate allele was observed in a mutant pool; Reference genome allele = all reads in a mutant pool represented the reference genome allele. Note that the reference genome is based on the Tü strain.

Identification of strain-specific diversity

To allow us to predict the parental origin of alleles in mutant pools, which facilitates mapping based on homozygosity-by-descent, we identified alleles that differed between parental strains. In the 7.6 million total SNPs identified, an alternate allele (with respect to the Zv9 reference genome, which is based on the Tü strain) was observed at 3–4 million sites in each wild-type line (Table 4.1). Consistent with previous reports noting a high degree of variation within each zebrafish strain [10–13], the vast majority of these sites were heterogeneous (*i.e.*, had reads representing both the reference and the alternate alleles) (Table 4.1). Thus, to identify SNPs that differed between lines, we selected SNPs at which all reads represented the reference allele in one line, while the other line had at least one read representing an alternate allele. When only the SNPs with sequence coverage in all six lines (5.2 million) were considered, any two lines differed at ~40% of loci (Figure 4.2A), which is in agreement with previous estimates of interstrain diversity [10]. The majority (72%) of SNPs were shared by at least three lines, while only 11% were unique to a single line (Figure 4.2B). For use in our mapping studies, we selected all sites at which alternate alleles were present in the strain used for outcrossing (TLF or WIK), but not in the strain used for mutagenesis (Tü). These alternate alleles were referred to as "mapping strain alleles" and consisted of 0.74 million and 1.2 million alleles for the TLF and WIK strains, respectively (Figure 4.2C-D). In each mutant pool, these sites were analyzed for the presence or absence of the mapping strain allele (Table 4.2).

Mapping mutants using homozygosity-by-descent

We next mapped each mutant on the basis of homozygosity-by-descent. For each mutant pool, we scanned the WGS data for regions with two characteristics: having a reduced level of heterogeneity and a reduced level of SNPs originating from the outcrossed strain, relative to the genome-wide averages of these measures. To quantify these characteristics, we designed an algorithm that produced a "mapping score," using sliding windows throughout the genome (Figure 4.3). Since we expected a characteristic footprint to span at least 10 cM on either side of the causative mutation (Figure 4.1), a window size of 20 cM, tiled at 0.25-cM intervals, was utilized. We based the window size on genetic distance (centimorgans) rather than physical distance (megabases), to take local recombination rates into account. This makes the analysis more accurate in regions close to centromeres and telomeres. Once regions with high
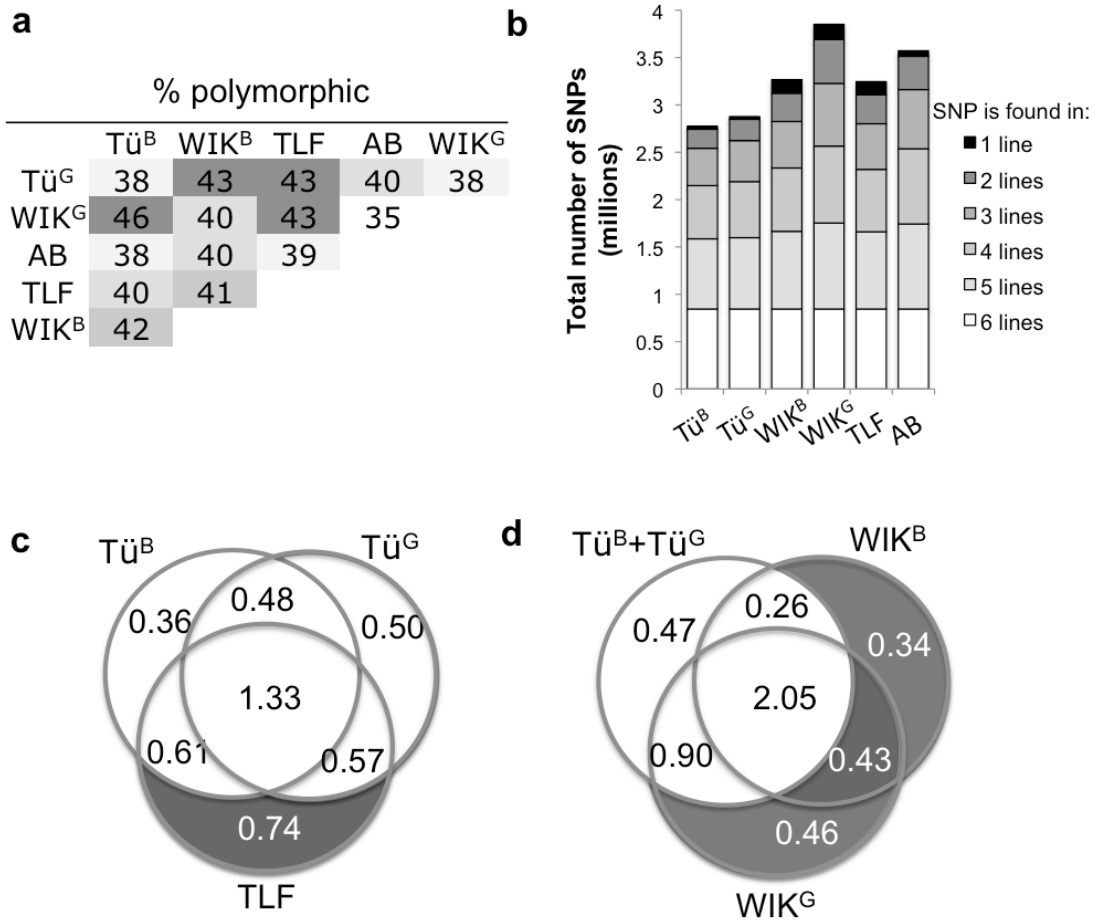
Figure 4.2: Genetic variation in zebrafish strains detected in low coverage WGS data of pooled DNA from 20 fish. (**a**) Pairwise comparison of SNP genotypes between parental lines showing the percentages of SNPs that are polymorphic and thus could be used to predict the parental origin for mapping based on homozygosity-by-descent. SNPs were classified as polymorphic if only the reference genome allele was observed in one line, while at least one alternate allele was observed in the second line. (**b**) Graph showing the number of polymorphic SNPs (in millions) identified in each parental line, as well as the number of lines with which these SNPs are shared. For (a) and (b), only sites with sequence coverage in all lines were considered (5.2 million sites out of the 7.6 million total SNP sites). (**c, d**) Venn diagrams showing the SNPs that were classified as mapping strain SNPs. This includes 0.74 million SNPs found in TLF but not in either of the Tü lines, and 1.2 million SNPs found in either of the WIK lines but not in either of the Tü lines. For (c) the two Tü lines are shown separately, while in (d) the data from these two lines were combined. Interestingly, due to high levels of intra-strain variation, there is a high number of SNPs that are not shared between the two Tü lines (c), and a similar number of SNPs that are not shared between the two WIK lines (d).

mapping scores were identified for a particular mutant, we independently tested linkage to these regions by the use of SSLP or SNP markers on DNA pools as well as in individual progeny (Table 4.3). In each of the five mutants analyzed, we confirmed that the region with the highest mapping score was linked to the mutation (Figure 4.3). In some cases, other unlinked areas exhibited relatively high mapping scores. We postulate that these regions represent haplotype blocks that were, by chance, shared by the two parental fish used for the initial mapping cross. We asked whether these shared haplotype blocks could have been predicted on the basis of the WGS of the parental strains, but found that each block occurred in a region in which heterogeneous SNPs (and therefore more than one haplotype) were observed in each of the parental strains. Furthermore, these blocks occurred in different locations in each mutant analyzed. Thus, if multiple regions with a high mapping score are obtained, independent tests for linkage will be needed to distinguish shared haplotype blocks from the region linked to the causative mutation. The presence of multiple high mapping scores in the genome could also represent second-site modifiers of the phenotype. These regions could then be analyzed for sequence variants that alter the expressivity of the mutant phenotype.

Our approach has two major differences from those previously used to map *C. elegans* and *A. thaliana* mutants [2–7]. First, our analysis is based on genetic rather than physical distance. Second, we combine the levels of homogeneity and strain-specific SNP signatures to map the locus. We find that this analytical method provides a robust and reliable means to correctly map the region linked to the mutation in zebrafish (Figure 4.4 and Figure 4.5).

The genetic architecture of linked regions

We further refined the linked interval by identifying an area of homogeneity within the broader region defined by our mapping algorithm. Since 20 fish were pooled for each mutant, we expected the region of homogeneity to span, on average, 2.5 cM on either side of the causative mutation (one recombinant per 40 meioses). Because of the low resolution of the genetic map, we utilized 100-kb windows (rather than centimorgans) to facilitate fine mapping of the interval. Assuming random sampling of alleles with only ~3× coverage, we expected and confirmed that linked regions containing two recombination events had an ~81% reduction in heterogeneity compared to unlinked regions, while
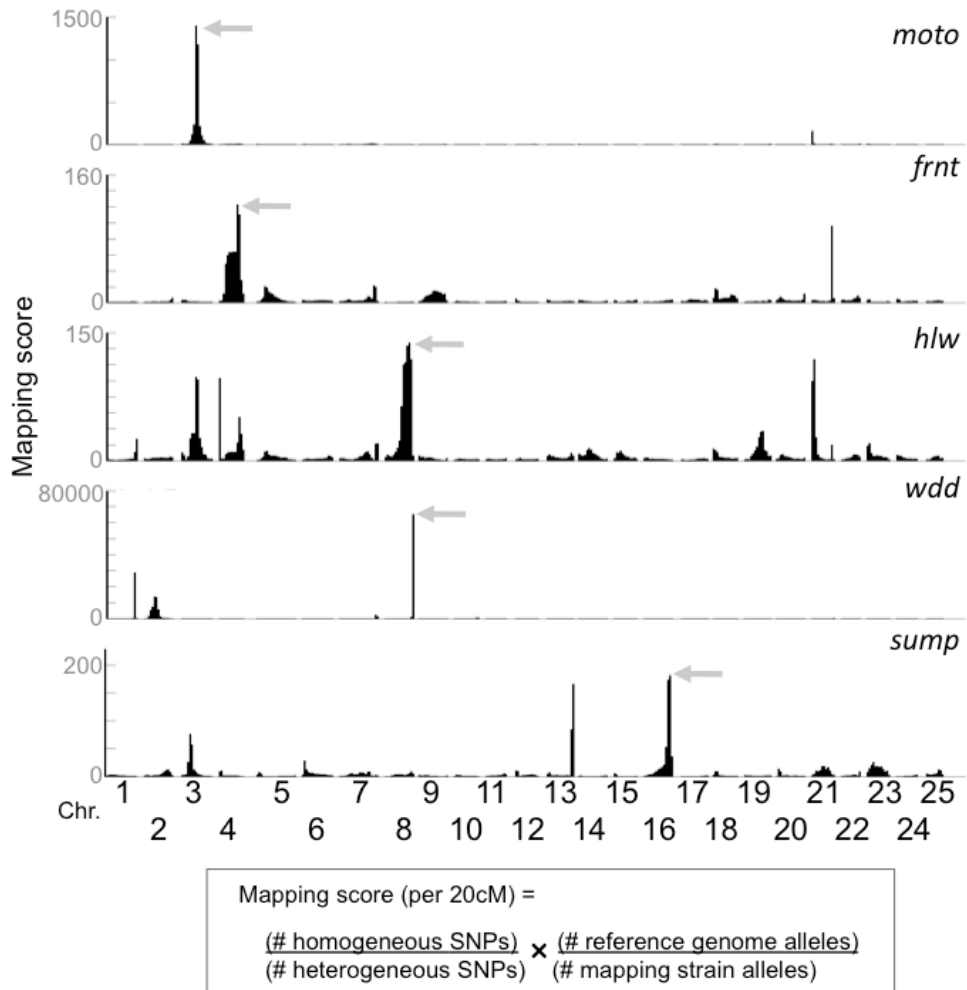
Figure 4.3: Mapping zebrafish mutants based on homozygosity-by-descent. Individual graphs depicting the mapping scores along each chromosome for the five different mutants (*moto, frnt, hlw, wdd, sump).* The mapping score is calculated as the ratio of homogeneous to heterogeneous SNPs, multiplied by the ratio of reference alleles to mapping strain alleles in sliding windows. The size of the sliding windows is 20 cM with an overlap of 19.75 cM between adjacent windows. Physical distances were converted to genetic distances using markers from the MGH meiotic map that have been mapped onto the Zv9 reference genome. In each of the five mutants, the region with the highest mapping score in the genome (gray arrows) was subsequently confirmed as containing the linked interval using SSLP or SNP markers.

Table 4.3: SSLP and SNP marker linkage data

| Zebrafish mutant	Chr[a]	Region of homogeneity (Mb)[b]	Marker[c]	Position[d] (Mb)	Recombinants/meiosis[e]
moto	3	19-36.6	z9964	34.3	2/90
wdd	8	51.8-55	*bmp1a*	53.5	0/66
hlw	8	43-47	z25210	43	2/44
			z7130	45.6	1/46
frnt	4	15-27.8	z11538	23.1	3/44
			z20450	24.4	2/44
sump	16	41.8-51.2	z8819	41.3	2/56
			z15739	43.9	0/56
			z4670	50.8	1/40
			z6854	50.6	1/40

[a]Chromosome showing the highest mapping score.
[b]Interval on the chromosome showing homogeneity.
[c]Marker used to confirm linkage.
[d]Marker position on the Chromosome in Mb.
[e]Recombinants identified for each marker in the number of meiosis tested.
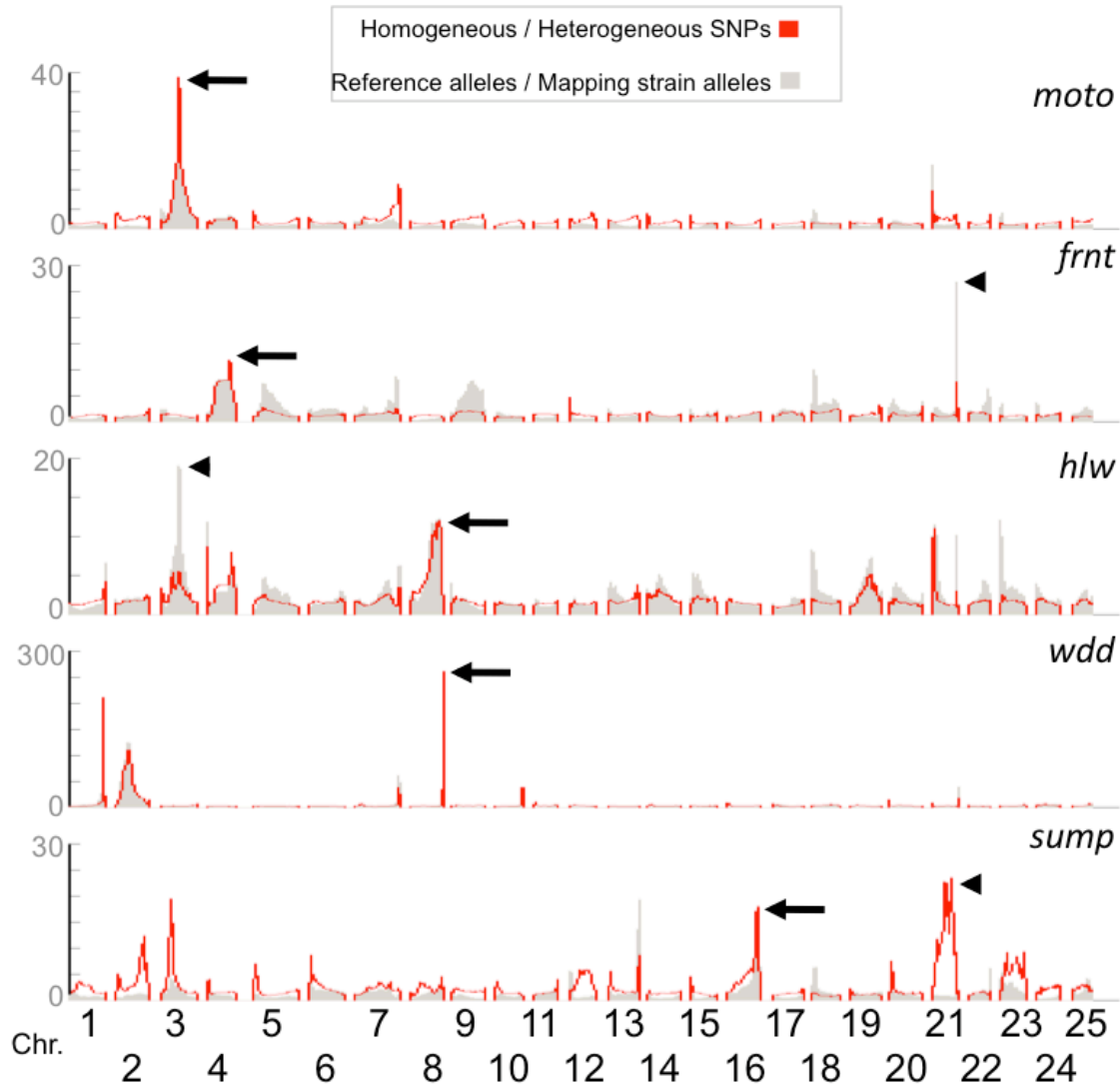
Figure 4.4: Mapping by combining both the frequency of heterogeneous SNPs and the frequency of mapping strain SNPs helps to eliminate false positives. Graphs show the comparison between the ratio of homogeneous to heterogeneous SNPs (red lines) to the ratio of reference genome alleles to mapping strain alleles (gray bars), for the five mutants analyzed (*moto, frnt, hlw, wdd, sump*). Ratios were calculated for all 25 chromosomes using sliding windows of 20 cM in size, with an overlap of 19.75 cM between adjacent windows. Genetic distances were defined by the MGH meiotic map. The arrow indicates the linked region for each mutant. For three mutants (*moto, frnt, wdd*), both approaches independently predict the linked region as the region in the genome with the highest ratio. In the *hlw* and *frnt* mutants, other regions show the highest ratio of reference alleles to mapping strain alleles (arrowheads). These regions do not have a high ratio of homogeneous to heterogeneous SNPs. Similarly, for the *sump* mutant, region on Chr21 shows the highest ratio of homogeneous to heterogeneous SNPs, but this region does not have a high ratio of reference alleles to mapping strain alleles (arrowhead). Accordingly, these false positive regions would result in a lower mapping score in our combined analysis and thus would be ranked as less likely to be linked to the mutation.
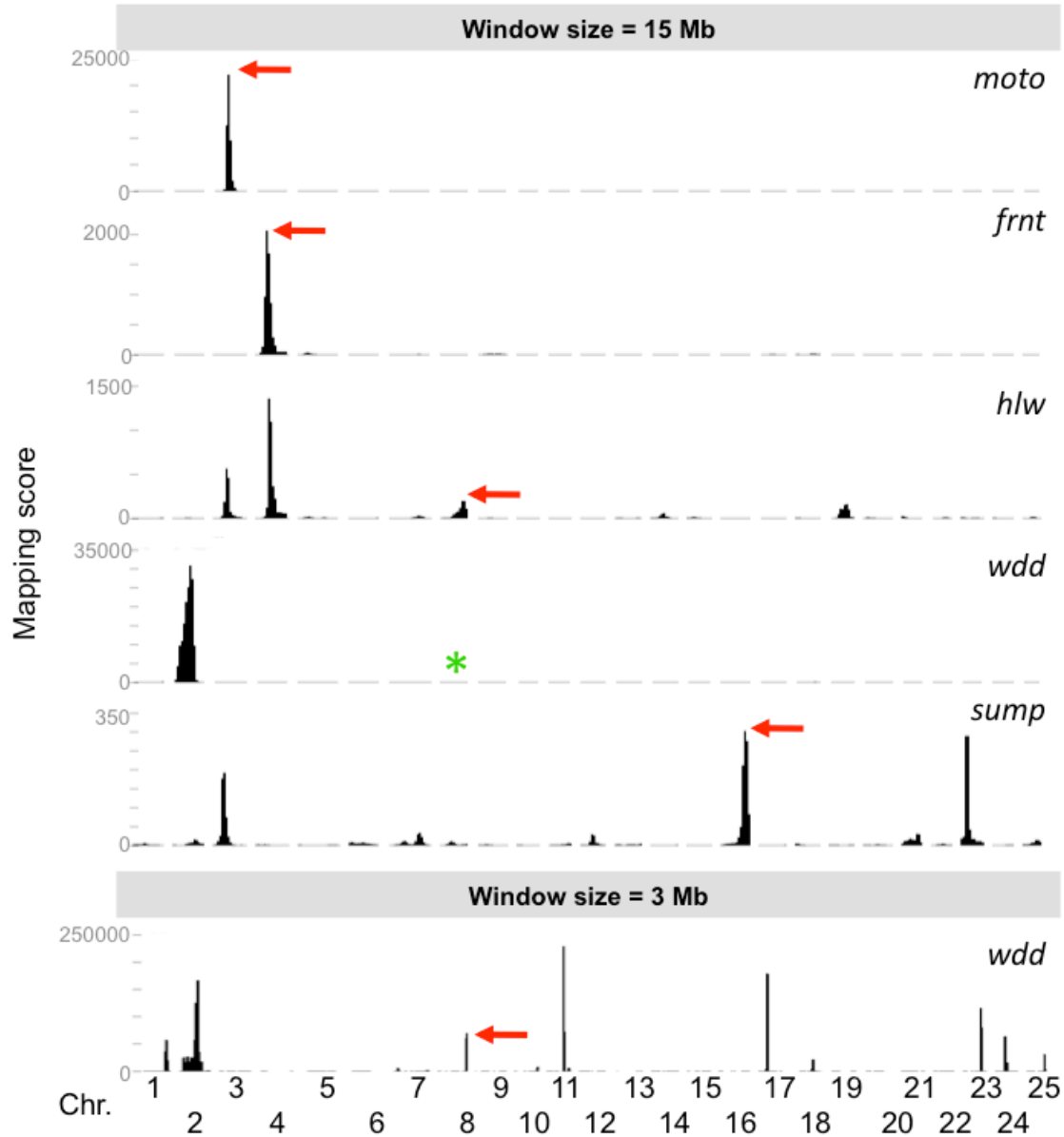
Figure 4.5: The sensitivity and specificity of mapping is affected by the size of the window used to calculate the mapping score. Graphs showing the genome-wide mapping scores using a sliding window of 15 Mb in size for all mutants (above), or a sliding window of 3 Mb in size for *wdd* (below), rather than the 20 cM windows used in our analysis. When 15 Mb sliding windows are used, in only three of the five mutants (*moto, frnt, sump*) the linked region is contained within the window with the highest mapping score in the genome (red arrows). In *hlw* the linked region is contained within the peak with the third highest mapping score (red arrow). In *wdd,* the linked region is not detected by an increase in the mapping score (asterisk), because the linked interval on Chr8 spans only 4 Mb. When a 3 Mb window was used for *wdd*, which should be small enough to detect the linked region, a mapping score peak appears at the linked interval (red arrow), but it is only the 5th highest peak.

regions containing one recombination event had a reduction in heterogeneity of ~90% (Figure 4.6 and Figure 4.1). We found that regions without recombination events were almost, but not completely, homogeneous, likely due to false positive variants resulting from sequencing errors or alignment artifacts. Therefore, we defined a candidate region of homogeneity as having a reduction in heterogeneity >90%. This approach allowed us to narrow down the candidate interval in each mutant to a region between 4 and 19 Mb in size (Table 4.4).

Identifying candidate phenotype-causing mutations within linked intervals

One of the powerful aspects of WGS is that it provides a large amount of sequence information throughout the candidate interval, allowing for the exclusion of much of the sequence in the interval as harboring the causative mutation. Additionally, the sequence allows the potential to identify the causative change. In the five mutants analyzed, between 76% and 92% of the coding sequence within the candidate interval was covered by at least two sequencing reads (Table 4.4). We identified hundreds to thousands of homogeneous variants in each candidate interval, of which between 4 and 136 were predicted to be nonsynonymous. However, we could exclude most of these variants as being causative for the phenotype since we also observed them in the WGS from the other unaffected strains (Table 4.4). In two of the five mutants we identified the likely causative mutation as a nonsynonymous change covered by at least two reads; these particular changes are predicted to encode nonsense alleles. In the three other mutants, unique nonsynonymous changes covered by two or more reads were not detected, but between 7 and 22 nonsynonymous changes were present in sequences covered by one read (Table 4.4). Further studies will be required to determine whether these single-read variants represent sequencing errors, normal variation, or phenotype-causing mutations.

A benefit of having performed WGS is, apart from being able to map the mutation in all mutants analyzed and to identify candidate coding mutations, that >87% of the coding sequence within the interval could be excluded because it did not differ from the reference data set. A second benefit of having performed WGS is that homogeneous SNPs identified in the candidate interval can serve as markers to test for linkage in additional F_2 fish, which will allow one to further refine the candidate interval.

The nonsynonymous changes we identified in the *welded* (*wddt31169*) and the *minamoto* (*motot31533*) mutants exemplify the value of the WGS method. For the *moto* mutant, characterized by
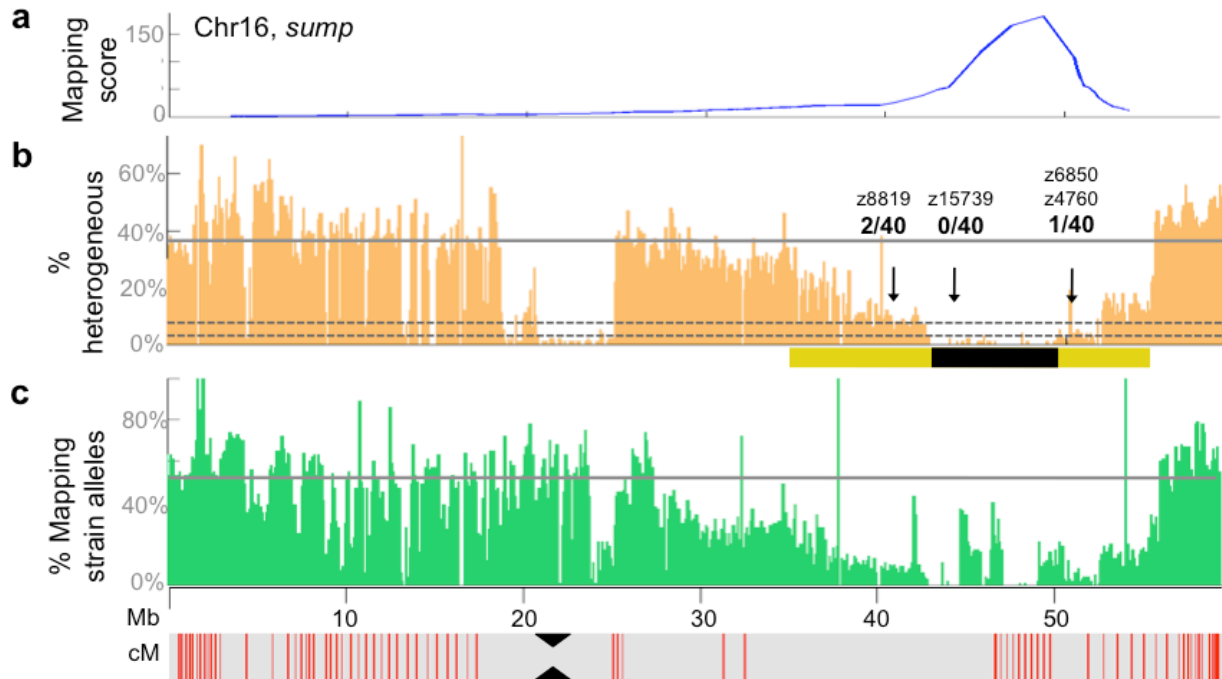
Figure 4.6: Genetic architecture of SNP diversity at a linked interval. (A) Graph of the mapping score across chromosome 16 in the *sump* mutant. This chromosome contained the highest mapping score in the genome. (B) Graph depicting the percentage of SNPs that were classified as heterogeneous in non-overlapping 100 kb windows along Chr16. The solid gray line indicates the genome wide average for SNPs classified as heterogeneous. Dotted gray lines indicate reductions in SNP heterogeneity of 90% (lower) and 81% (upper), respectively, as compared to the genome-wide average. The yellow bar demarcates the region with a reduction in heterogeneity of at least 30%, while the black bar demarcates the candidate region, defined by a reduction in heterogeneity of greater than 90%. Black arrows indicate the locations of SSLP markers used to confirmed linkage to this interval by individually genotyping the 20 *sump* mutants that had been pooled for WGS. The fraction of recombination events per 40 meioses for each SSLP marker is indicated. (C) Graph showing the percentage of sites containing mapping strain alleles, in non-overlapping 100 kb windows along the chromosome. This percentage is only calculated for sites at which the strain used for mapping (WIK) showed an allele that was not observed in the strain used for mutagenesis (Tü). The gray line indicates the genome wide average of the percentage of sites containing mapping-strain alleles. Physical distances in Mb along Chr16 are indicated. The red vertical lines in the grey bar below the graphs indicate genetic distances, with lines spaced at ~1 cM intervals. The position of the centromere is indicated by black triangles.

Table 4.4: Characteristics of linked intervals identified by whole genome sequencing

Zebrafish mutant[a]	*moto*	*wdd*	*hlw*	*frnt*	*sump*
Whole genome sequencing					
# reads (10^6)	61	81	60	79	83
Genome coverage	2.6x	2.7x	2.8x	3.9x	4.1x
Size of the linked interval[b]					
Region of reduced heterogeneity[c]					
Physical size (Mb)	35	4	26	46	19
Genetic size (cM)	<25	<30	<25	<18	<18
Region of homogeneity[d]					
Physical Size (Mb)	19	4	5	5	8
Genetic size (cM)	<14	<30	<4	<3	<11
Coding sequence coverage in linked interval[e]					
≥1 read	93%	92%	87%	95%	88%
≥ 2 reads	83%	77%	76%	92%	85%
Homogeneous SNPs in linked interval[f]					
Total	7640	225	1071	7783	9518
Not in dbSNP[g]	5984	223	1023	6431	9110
Noncoding	5663	207	997	6303	8687
Synonymous	229	12	17	89	287
Nonsynonymous	92	4	9	39	136
Unique[h]					
Noncoding	38	12	21	79	106
Synonymous	0	0	0	0	2
Nonsynonymous	2	1	0	0	0
	(63)[i]	(119)	(22)	(4)	(7)

[a]*minamoto* (*moto[t31533]*), *welded* (*wdd[t31169]*), *hollow* (*hlw[t3373]*), *fruehrentner* (*frnt[t31786]*) and *schrumpfkopf* (*sump[t3625]*). [b]Intervals identified by a high mapping score; linkage was confirmed by analysis of SSLP or SNP markers. [c]Region with a reduction in heterogeneity of at least 30% compared to the genome-wide average. [d]Defined as a region with a greater than 90% reduction in heterogeneity compared to the genome wide average. [e]Coding sequence of RefSeq and Ensembl genes. [f]Detected homogeneous variants covered by at least 2 sequencing reads. [g]Homogeneous variants not present in the publicly available SNP database (dbSNP) downloaded from Ensemble. [h]Homogeneous variants not present in the reference strain database established in this study. [i]The number of unique nonsynonymous mutations covered by only one sequencing read.

defective spermatogenesis, two nonsynonymous mutations (Table 4.4), covered by two and three reads respectively, were identified within the linked interval, one missense and one nonsense mutation. The nonsense mutation was confirmed to be homozygous in all 20 fish sequenced. This mutation lies within a novel gene (*ENSDARG00000090664*) that is conserved in vertebrates. Consistent with the observed spermatogenesis defects in the mutants, this gene is expressed in testes among vertebrates (http://www.ncbi.nlm.nih.gov/unigene) and thus is a strong candidate for causing the *moto* phenotype. For the *wdd* mutant, characterized by its adult craniofacial phenotype, only one nonsynonymous change, which was supported by eight reads, was detected in the linked interval (Table 4.4). This change creates a nonsense mutation (p.R227X) in the gene encoding Bone morphogenetic protein 1a (Bmp1a). PCR amplification and capillary dideoxy sequencing of the genomic region in individual F_2 mutants and siblings confirmed the mutation and linkage to the mutant phenotype (0 recombinants in 66 meioses). It previously had been shown that morpholino-mediated reduction of Bmp1a function in zebrafish impairs larval development, leading to a wavy fin fold phenotype [14]. We detected a similar larval fin phenotype in *wdd* mutants and confirmed that this phenotype occurs in wild-type embryos injected with a morpholino targeting the translation initiation site of *bmp1a* (Figure 4.7). Thus we show that the nonsense mutation in *bmp1a* is the likely causative mutation underlying the *wdd* phenotype. With a causative mutation in hand, it is now possible to investigate the mechanistic basis of this skeletal phenotype.

Minimum genome coverage needed for mapping

Our analysis showed that ~3× genome coverage was sufficient to correctly map each mutant to a defined interval, to cover >87% of coding sequence within the candidate interval, and to identify a manageable number of variants as being potential causative mutations. To determine whether lower genome coverage would be sufficient for mapping and mutation detection, we applied the same mapping algorithm to randomly selected subsets of the total sequence reads obtained for each mutant. Utilizing only 5 million reads, which is equivalent to ~0.2× genome coverage, we could still reliably identify the linked regions (Figure 4.8 and Figure 4.9). However, with 0.2× genome coverage, only 5% of coding sequence in the linked interval was covered by ≥2 reads, and 73% was not sequenced at all (Figure 4.8). Thus, using this method, it is feasible to map multiple mutants simultaneously by barcoding ~14 mutant
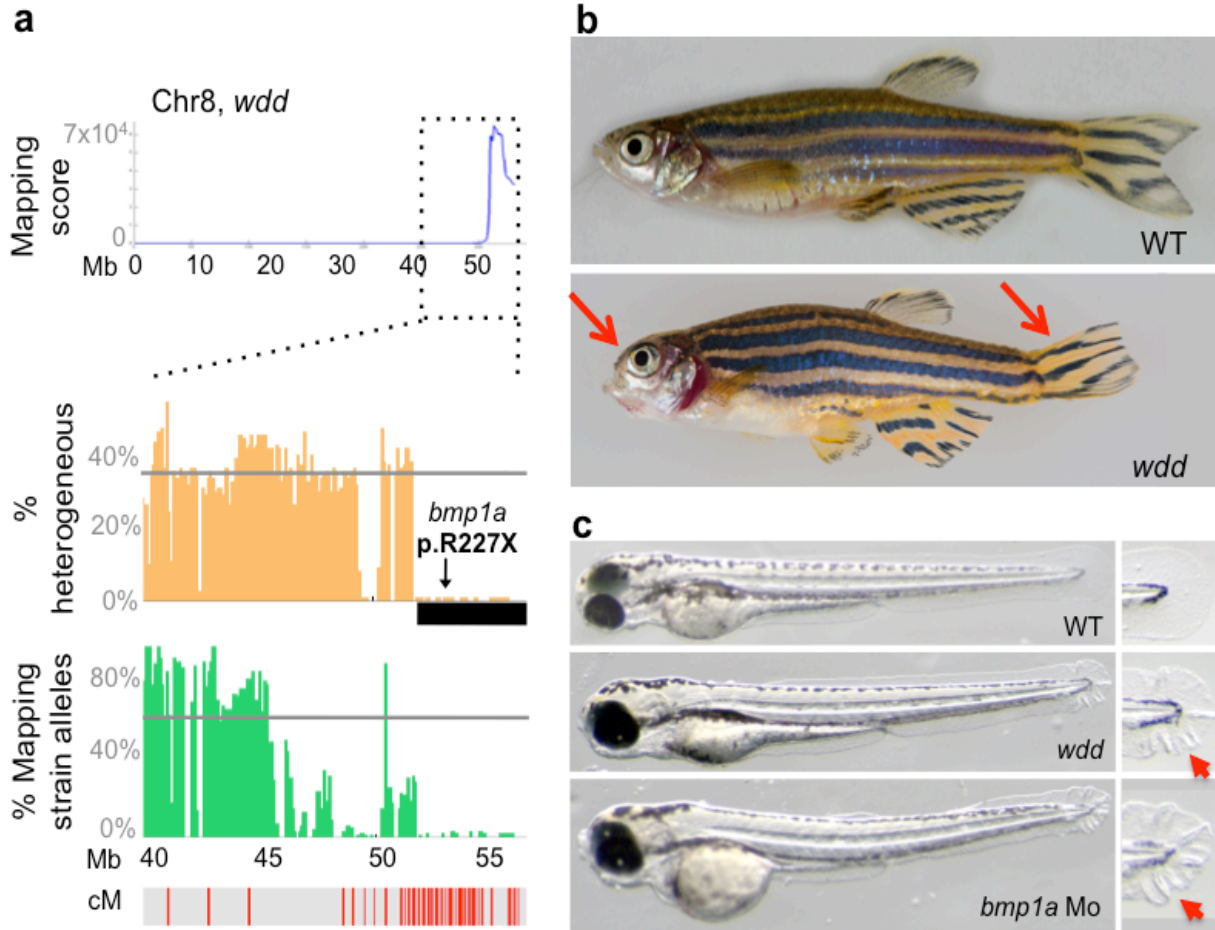
Figure 4.7: Identification of a loss-of-function allele of Bmp1a that underlies the *wdd* mutant phenotype. (A) The mapping score plot for *wdd* is shown for Chr8, which contained the highest mapping score in the genome. In the graphs below, the genetic architecture of the linked region is shown. Annotation is similar to Figure 2. The location of the nonsense mutation within *bmp1a* that lies in the candidate interval is indicated (arrow). (B) Lateral-view of adult wild-type, and homozygous *wdd* mutant fish. Mutant fish are characterized by frontonasal shortening of the skull and deformed tailfins (red arrows). (C) Lateral- view of wild-type, *wdd* mutant, and *bmp1a* morpholino-injected larvae. Mutant and morphant larvae show a similar characteristic wavy appearance of their fin folds (red arrowheads) at 3 days post fertilization (dpf), which is not observed in wild-type larvae. Insets show a higher magnification of the distal part of the finfold.
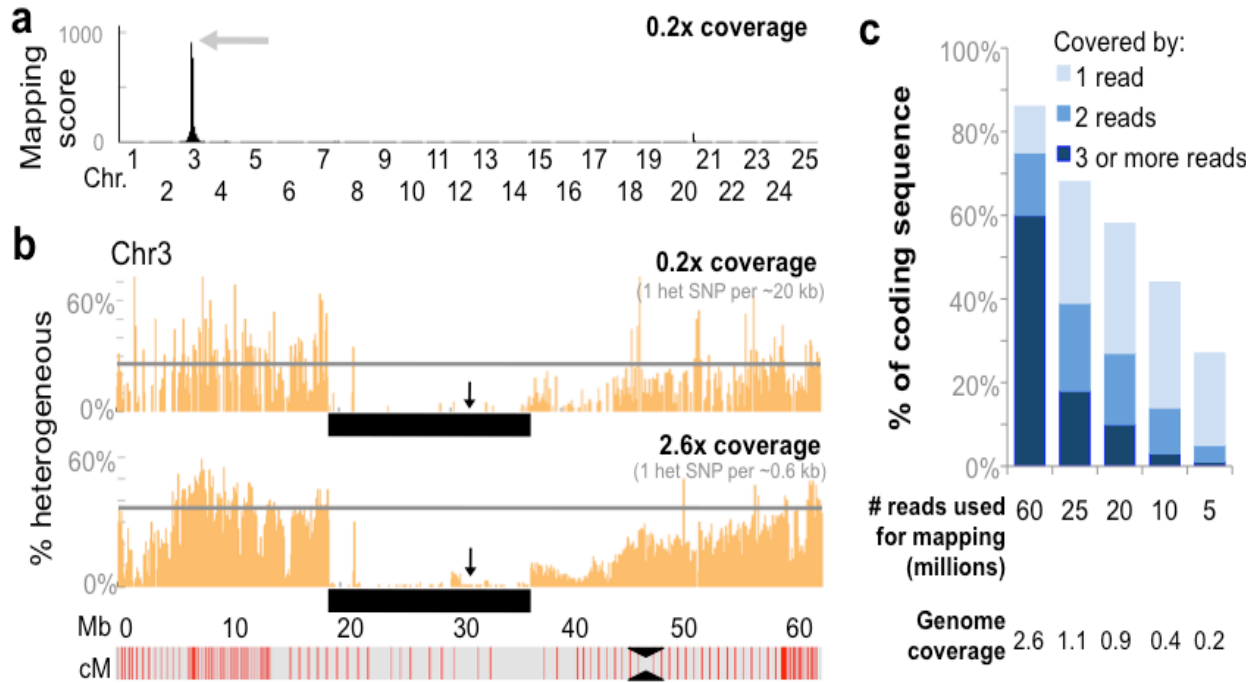
Figure 4.8: Mapping of mutants using only ~0.2x genome coverage. (A) Graph depicting the genome-wide mapping score plot for the *moto* mutant generated with a randomly selected subset (5 million) of the total sequencing reads, which results in a genome coverage of 0.2x. (B) Graph depicting the percentage of SNPs that were classified as heterogeneous in non-overlapping 100 kb windows along Chr3. The arrow indicates the location of a SNP marker that was used to confirm linkage (0 recombinants in 40 meioses). While the overall number of detectable heterogeneous SNPs is reduced with a genome-wide coverage of only 0.2x, the boundaries of the linked interval can be identified just as well as with 2.6x coverage. Black bar underlies the region of homogeneity (C) Graph depicting the loss in coverage of coding sequence that occurs as genome-wide coverage decreases.
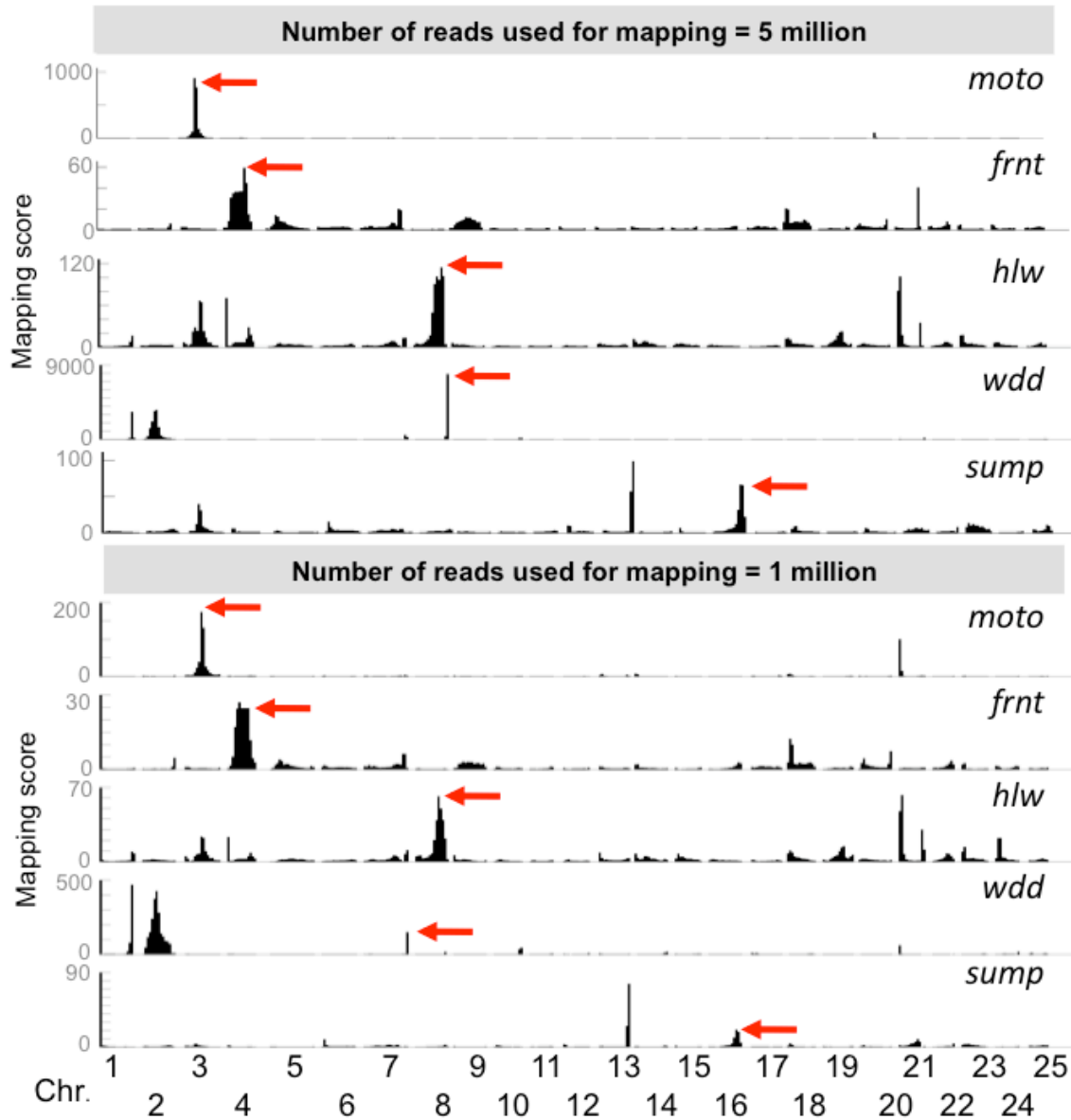
Figure 4.9: Minimum coverage needed for efficient mapping of zebrafish mutants. Graphs depicting the genome-wide mapping scores calculated for each mutant in 20 cM sliding windows, using either only 5 million (top) or 1 million (bottom) randomly selected Illumina sequencing reads. The actual map positions for each mutant are indicated (red arrows). When 5 million reads are used, the mapping score plots are not significantly different from those generated using all reads (>60 million) (Figure 1). The only exception is that, for *sump*, the linked region has only the 2nd highest mapping score. Even when only 1 million reads are used, in three mutants (*moto, frnt, hlw*) the linked region has the highest mapping score in the genome. For two other mutants (*wdd, sump*), the relative heights of the false positive peaks are significantly increased.

DNA libraries and then sequencing a pool of these libraries on a single lane of an Illumina HiSeq apparatus. However, with this "bulk mapping" approach it would be unlikely to identify the causative mutations using the generated sequence alone.

4.6   Discussion

We show that recessive zebrafish mutations can be efficiently mapped and cloned using low-coverage WGS of only 20 pooled mutant progeny. While WGS has been used in other experimental models, such as *C. elegans* and *A. thaliana* [2–7], the size and polymorphic diversity of the zebrafish genome posed unique challenges. By constructing an extensive SNP database using WGS from six different wild-type lines, we increased the accuracy of mapping as well as the ability to distinguish phenotype-causing mutations from previously unannotated SNPs. This newly identified SNP database, containing millions of SNPs, is an order of magnitude larger than the SNPs previously annotated within publically available databases. This large database allowed us to identify strain-specific SNP signatures, which facilitated our detection of intervals that were homozygous-by-descent.

An alternative strategy of mapping mutants using WGS would be to separately sequence pools of mutants and unaffected siblings, rather than using a comparison to wild-type strains. With the limited recombination rate within the 20 fish sequenced, both strategies would provide similar resolution of the mapping interval. Using a sequence data set representing ~50× coverage, we increase the accuracy of identifying SNPs within the mapping interval without the need for low-coverage sequence data from siblings. Additionally, analysis of the siblings for each mutant would double the cost per mutant analyzed. We think that the strain-specific and reference SNP databases we created provide a more efficient means of analyzing sequence data from multiple mutants in parallel. This SNP data set can be utilized by a large number of researchers to facilitate mapping of mutants (data and scripts available at http://www.fishyskeleton.com).

It is important to note that the detection of candidate mutations depends not only on the genome coverage obtained by WGS, but also on the quality and extent of the genome assembly that is used as a reference; in regions with poor genome assembly, lack of detection of a causative mutation will not be remedied by higher sequencing depth. Further improvements in assembly of the zebrafish genome, in the SNP database, and in massively parallel sequencing will enhance the sensitivity and specificity of our

mapping approach. At present, low-coverage WGS using pooled DNA samples provides a fast and efficient means for mapping and identifying recessive mutations in zebrafish, allowing for more timely determination of altered gene function and systematic analysis of genetic regulation of vertebrate development and physiology.

4.7   Materials and methods

Zebrafish husbandry and strains

Zebrafish were raised and maintained as described [15]. Mutants were identified in the 2004 ZF-MODELS screen performed at the Max-Planck Institute for Developmental Biology (MPI-EB) in Tübingen, Germany. Mutant and wild-type strains were obtained from stocks at Children's Hospital, Boston (TüB, WIKB, AB, and TLF) and at the MPI-EB (TüG and WIKG). The *minamoto* (*motot31533*), *welded* (*wddt31169*), *hollow* (*hlwt3373*), *fruehrentner* (*frntt31786*), and *schrumpfkopf* (*sumpt3625*) mutants were generated in the Tü background. For mapping, the majority of mutants were outcrossed to the WIK line, except for *sump*, which was crossed to TLF. F_2 crosses were screened for the phenotype and identified mutants and siblings were frozen.

Linkage analysis

Linkage was assessed by analysis of microsatellites (SSLPs) and SNP markers on genomic DNA from single fish using standard PCR amplification and, in the case of SNPs, analysis by dideoxy capillary sequencing.

Morpholino injections

A morpholino directed against the translation initiation site of *bmp1a* (MO1) [14] was injected at a concentration of 0.3 mM into one-cell stage Tü embryos. The phenotype was assessed at 3 days postfertilization (dpf).

Genomic DNA library construction and Illumina sequencing

For each mutant or parental strain, genomic DNA from 20 adult fish was pooled (150–250 ng from each fish—also easily obtainable from larvae), and 3–5 µg was sheared to an average size of 200 bp, using Adaptive Focused Acoustics following the manufacturer's protocol (Covaris). For three samples (*wdd*, *sump*, and *frnt*), the shearing step was omitted, since the genomic DNA appeared degraded, with

most fragments being <250 bp in size as assessed by electrophoresis on a 4% agarose gel. To construct DNA libraries, the DNA fragments were blunt-ended, 5′ phosphorylated, A-tailed, and ligated to adaptors as previously described [16] , with the exception that adaptors did not have a 3-bp barcode sequence, and the volume of AMPure XP beads used for purification was 1.4× rather than 3.0×. Phusion High-Fidelity DNA polymerase (Finnzymes) was then used to amplify 12 µl (30%) of each library, in a total of four 50-µl PCR reactions, using the "postcapture" primers described in Bowen *et al*. (2011) [16]. Eight cycles of PCR were used for *wdd* and six cycles for all other samples. Each amplified library was sequenced on one lane of an Illumina HiSeq2000, using 100-bp single-end sequencing. Since the number of reads obtained for *frnt*, TüG, WIKG, and *sump* was lower than expected, one lane of GAII 100-bp single-end sequencing was also performed for each of these samples.

Illumina data analysis

Illumina sequence reads were aligned to the reference genome (version Zv9/danRer7), using Novoalign software (**http://www.novocraft.com/main/index.php**) with default settings and including 3′-adaptor trimming. PCR duplicates were removed using the MarkDuplicates command in Picard (**http://picard.sourceforge.net/**). Multisample variant calling was performed for each chromosome on all samples simultaneously, using SAMtools and BCFtools. The SNPs were then filtered using the GATK VariantFiltrationWalker to exclude the following variants: (1) SNPs lying in low-complexity sequences or interspersed repeats, classified by RepeatMasker; (2) SNPs lying within 10 bp of an indel; (3) SNPs lying in a cluster of ≥3 SNPs per 10 bp; (3) SNPs with a quality score <30; (4) SNPs with a root-mean-square mapping quality of covering reads <40; and (5) SNPs with a total read depth <15 or >120. A perl script was used to exclude variants seen in <3 reads, variants not seen in both the forward and the reverse direction, variants with a tail bias <0.05, and variants that were not biallelic. Only the 7.6 million SNPs that passed these filtration steps were used for downstream analyses; of these, only a small percentage (0.25%, 18,978 SNPs) were found solely in one mutant and may represent ENU-induced variation. A perl script was written to classify the genotype of each mutant or reference strain at each of the 7.6 million "pass filter" SNP sites. Genotypes were classified as heterogeneous or homogeneous on the basis of the "BCFtools phred scaled genotype likelihood score." Sites covered by <2 reads were considered uninformative.

To determine the physical size of the 20-cM windows used to calculate the mapping score, the MGH mapping panel was downloaded from ZFIN (**http://zfin.org/zf_info/downloads.html#marker**). A script was written to obtain the Zv9 genomic coordinate of each marker from Ensembl (**http://useast.ensembl.org/index.html**). A genomic coordinate could be obtained for 2100 of the 3845 markers. Seventy markers mapped to more than one location and were excluded from the analysis. In addition, a BLAST search was performed to find the coordinates of some markers that did not have genomic coordinates listed in Ensembl. These markers were then used to approximate the genomic coordinate of sliding 20-cM windows throughout the genome, with a new window starting every 0.25 cM.

Annovar (**http://www.openbioinformatics.org/annovar/**) was used to classify variants as noncoding, synonymous, or nonsynonymous and to determine whether variants were listed in the publically available SNP database, downloaded from Ensembl (**http://useast.ensembl.org/info/data/ftp/index.html**). To identify variants present in only one read (which would not have been identified using SAMtools/BCFtools multisample variant calling), the SAMtools mpileup command was performed on all mutants and reference strains, for all coding exons, and a perl script was used to select variants unique to each mutant. All perl scripts, as well as aligned sequence files for each wild-type strain, are available online at **http://www.fishyskeleton.com**.

4.8   Literature Cited

1. Geisler R, Rauch G-J, Geiger-Rudolph S, Albrecht A, van Bebber F, Berger A, Busch-Nentwich E, Dahm R, Dekens MPS, Dooley C, et al.: **Large-scale mapping of mutations affecting zebrafish development**. *BMC Genomics* 2007, **8**:11.

2. Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen AH, Nielsen KL, Jørgensen J-E, Weigel D, Andersen SU: **SHOREmap: simultaneous mapping and mutation identification by deep sequencing**. *Nat. Methods* 2009, **6**:550–551.

3. Cuperus JT, Montgomery TA, Fahlgren N, Burke RT, Townsend T, Sullivan CM, Carrington JC: **Identification of MIR390a precursor processing-defective mutants in Arabidopsis by direct genome sequencing**. *Proc. Natl. Acad. Sci. U.S.A.* 2010, **107**:466–471.

4. Doitsidou M, Poole RJ, Sarin S, Bigelow H, Hobert O: **C. elegans mutant identification with a one-step whole-genome-sequencing and SNP mapping strategy**. *PLoS ONE* 2010, **5**:e15435.

5. Zuryn S, Le Gras S, Jamet K, Jarriault S: **A strategy for direct mapping and identification of mutations by whole-genome sequencing**. *Genetics* 2010, **186**:427–430.

6. Austin RS, Vidaurre D, Stamatiou G, Breit R, Provart NJ, Bonetta D, Zhang J, Fung P, Gong Y, Wang PW, et al.: **Next-generation mapping of Arabidopsis genes**. *Plant J.* 2011, **67**:715–725.

7. Uchida N, Sakamoto T, Kurata T, Tasaka M: **Identification of EMS-induced causal mutations in a non-reference Arabidopsis thaliana accession by whole genome sequencing**. *Plant Cell Physiol.* 2011, **52**:716–722.

8. Arnold CN, Xia Y, Lin P, Ross C, Schwander M, Smart NG, Müller U, Beutler B: **Rapid identification of a disease allele in mouse through whole genome sequencing and bulk segregation analysis**. *Genetics* 2011, **187**:633–641.

9. Sobreira NLM, Cirulli ET, Avramopoulos D, Wohler E, Oswald GL, Stevens EL, Ge D, Shianna KV, Smith JP, Maia JM, et al.: **Whole-genome sequencing of a single proband together with linkage analysis identifies a Mendelian disease gene**. *PLoS Genet.* 2010, **6**:e1000991.

10. Stickney HL, Schmutz J, Woods IG, Holtzer CC, Dickson MC, Kelly PD, Myers RM, Talbot WS: **Rapid mapping of zebrafish mutations with SNPs and oligonucleotide microarrays**. *Genome Res.* 2002, **12**:1929–1934.

11. Guryev V, Koudijs MJ, Berezikov E, Johnson SL, Plasterk RHA, van Eeden FJM, Cuppen E: **Genetic variation in the zebrafish**. *Genome Res.* 2006, **16**:491–497.

12. Bradley KM, Elmore JB, Breyer JP, Yaspan BL, Jessen JR, Knapik EW, Smith JR: **A major zebrafish polymorphism resource for genetic mapping**. *Genome Biol.* 2007, **8**:R55.

13. Coe TS, Hamilton PB, Griffiths AM, Hodgson DJ, Wahab MA, Tyler CR: **Genetic variation in strains of zebrafish (Danio rerio) and the implications for ecotoxicology studies**. *Ecotoxicology* 2009, **18**:144–150.

14. Jasuja R, Voss N, Ge G, Hoffman GG, Lyman-Gingerich J, Pelegri F, Greenspan DS: **bmp1 and mini fin are functionally redundant in regulating formation of the zebrafish dorsoventral axis**. *Mech. Dev.* 2006, **123**:548–558.

15. Nusslein-Volhard C, Dahm R: *Zebrafish: A Practical Approach*. University Press; 2002.

16. Bowen ME, Boyden ED, Holm IA, Campos-Xavier B, Bonafé L, Superti-Furga A, Ikegawa S, Cormier-Daire V, Bovée JV, Pansuriya TC, et al.: **Loss-of-function mutations in PTPN11 cause metachondromatosis, but not Ollier disease or Maffucci syndrome**. *PLoS Genet.* 2011, **7**:e1002050.

Chapter **5**

Discussion and Future Directions

5.1   Discussion and Future Directions

Next Generation Sequencing (NGS) technologies enable the parallel sequencing of millions of DNA fragments, making it feasible to sequence entire genomes, exomes and transcriptomes. In this thesis, I used NGS to further our understanding of skeletal development and disease in both humans and model organisms. I explored three applications of NGS: targeted re-sequencing (**Chapter 2**), RNA-seq (**Chapter 3**) and whole genome sequencing (WGS) (**Chapter 4**). Firstly, by performing targeted re-sequencing of genes in an interval identified by genome-wide linkage analysis, I identified heterozygous loss-of-function mutations in *PTPN11* as the genetic cause for the cartilage tumor syndrome, metachondromatosis (MC) (**Chapter 2**). Secondly, I performed RNA-seq on chondrocytes undergoing maturation *in vitro*, and found that loss of *Ptpn11* delayed terminal differentiation and maintained the expression of genes associated with earlier maturation stages (**Chapter 3**). I confirmed this *in vivo* by showing that terminal differentiation was delayed in mouse vertebral growth plates following mosaic inactivation of *Ptpn11* (**Chapter 3**). Furthermore, I found that mosaic inactivation of *Ptpn11* disrupted the organization of maturation zones in mouse vertebral growth plates and led to enchondroma-like lesions that contained both wild-type and mutant chondrocytes, and that cartilage tumors from MC patients also showed disorganized maturation zones (**Chapter 3**). Thirdly, I developed a data analysis pipeline to map and clone zebrafish mutations based on low-coverage WGS data, and applied this to map mutations that affect skeletal development (**Chapter 4**). Altogether, my results provide insight into the regulation of skeletal development and cartilage tumor formation in humans and model organisms, and demonstrate the utility of applying NGS technologies to skeletal biology.

Genetic basis of cartilage tumor syndromes

Patients with the rare, autosomal dominant cartilage tumor syndrome, metachondromatosis (MC), develop cartilage tumors on the surface of their bones (exostoses) and inside their bones (enchondromas) [1]. Other cartilage tumor syndromes are associated with either exostoses (MHE) or enchondromas (Ollier disease, Maffucci syndrome), but not both [2]. When the work in this thesis was initiated in 2009, the genetic and molecular mechanisms underlying cartilage tumor formation were poorly understood (Figure 5.1). While it was known that heterozygous loss-of-function mutations in *EXT1* or
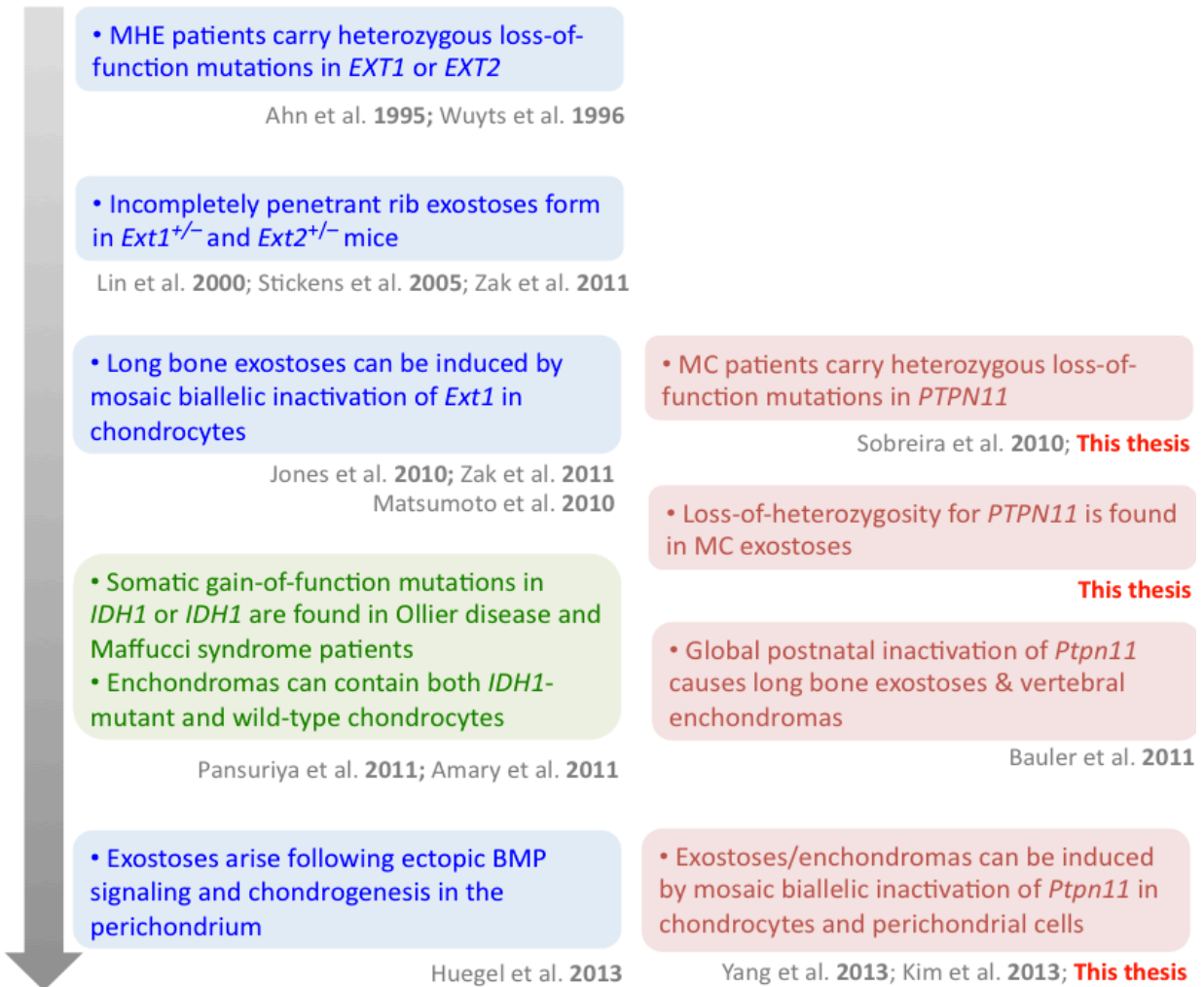
Figure 5.1: Timeline of research on the genetic causes of cartilage tumor syndromes. Research involving multiple hereditary exostoses (blue), Ollier disease and Mafficci syndrome (green) and metachondromatosis (red) is shown.

EXT2 cause MHE [3,4], the genetic cause(s) of MC, Ollier disease and Maffucci syndrome were not known. In particular, it was not known whether each cartilage tumor syndrome had its own unique genetic cause, or whether they were caused by mutations that affected the same genes or the same pathways. In this thesis, I identified heterozygous loss-of-function mutations in *PTPN11* in 11 out of 17 MC individuals or families (**Chapter 2**)*,* consistent with the independent identification of *PTPN11* mutations in two MC families by Sobeira *et al* (2010) [5]. Concurrently, it was shown that somatic gain-of-function mutations in *IDH1* or *IDH2* cause Ollier Disease and Maffucci Syndrome [6,7]. Thus, although MC shares phenotypic features with other cartilage tumor syndromes, it has a unique genetic cause. Furthermore, as discussed

in more detail below, loss of *Ptpn11* in mice appears to have different effects on chondrocyte maturation than loss of *Ext1/Ext2* or mutations in *Idh1*, suggesting that the cartilage lesions in MC patients arise due to unique underlying molecular and cellular mechanisms.

In MC patients who carry heterozygous *PTPN11* mutations, cartilage lesions could arise due to haploinsufficiency for *PTPN11* or due to somatic second-hit *PTPN11* mutations. To test these hypotheses, I sequenced DNA extracted from the cartilage core of MC exostoses. I found evidence for loss-of-heterozygosity (LOH) for *PTPN11* in two MC exostoses, thus supporting the two-hit hypothesis (**Chapter 2**). Furthermore, cartilage tumors do not form in *Ptpn11*$^{+/-}$ mice but can be induced by mosaic bi-allelic inactivation of *Ptpn11* [8–10] (**Chapter 3**). Together, these data support the hypothesis that tumors in MC patients form when both *PTPN11* alleles are lost. As suggested by lineage tracing in mice, both *Ptpn11*-deficient chondrocytes and wild-type chondrocytes can be incorporated into lesions (**Chapter 3**). Exostoses from MHE patients and enchondromas from Ollier disease patients are also thought to contain both mutant and non-mutant chondrocytes [7,11]. Together, these data suggest that enchondromas and exostoses may arise in part due to altered paracrine signaling that affects neighboring wild-type cells, rather than simply due to the neoplastic growth of mutant cells themselves. The extent to which neighboring *PTPN11*-heterozygous cells can be incorporated into lesions in human MC patients remains to be determined. One approach that could be used to quantify the relative proportion of *PTPN11*-heterozygous cells compared to *PTPN11*-null cells within a lesion would be to use digital droplet PCR [12].

We failed to detect any mutations in *PTPN11* or in 74 genes that act in the same signaling pathways as *PTPN11* in 6 out of 17 MC patients (**Chapter 2**). This included five individuals who were the first in their families to be affected with MC and one individual with a family history of MC. For these patients, only DNA from white blood cells was sequenced, since lesional tissue was not available. It is possible that these individuals have mutations in a second gene or in regulatory regions of *PTPN11* that were not detected by our sequencing approach. Alternatively, the 5 individuals with sporadic MC could carry somatic *PTPN11* mutations that are present only in only a small percentage of their white blood cells and therefore escaped detection by our sequencing approaches. Indeed, in one of the 11 MC families in which a *PTPN11* mutation was detected, the mutation was detected by sequencing blood DNA

from two affected children, but was not detected in blood DNA from the affected mother who was the first in the family to have MC (**Chapter 2**). This demonstrates that *PTPN11* mutations in individuals with sporadic MC can be present at low enough levels in blood DNA to escape detection by Sanger sequencing or low coverage NGS. Furthermore, in individuals who are somatic mosaics, there may be negative selection against *PTPN11*-heterozygous cells in the blood, as suggested by mouse studies indicating that *Ptpn11* heterozygous hematopoietic stem cells (HSCs) have reduced survival compared to wild-type HSCs [13]. Thus, to determine whether individuals with sporadic MC have mutations in *PTPN11* in their peripheral blood, one could perform very high coverage NGS of *PTPN11* to determine whether mutations are present in a small fraction of cells.

Malignant progression of enchondromas and exostoses has been observed in Ollier disease, Maffucci syndrome and MHE patients, and has been documented for one MC patient [14,15]. Interestingly, it was shown that chondrosarcomas arising from exostoses with *EXT1*/*EXT2* mutations arise from cells with functional *EXT1*/*EXT2*, rather than the *EXT1/EXT2*-deficient cells [16]. Thus, bi-allelic loss of *EXT1*/*EXT2* is sufficient to induce the formation of an exostosis, which creates a microenvironment that supports the malignant progression of cells with functional *EXT1*/*EXT2*. It will be of interest to determine whether this scenario is also the case for the chondrosarcoma that was reported to arise from an MC enchondroma [15]. Since *PTPN11* can act as an oncogene [17], it is plausible that cells with functional *PTPN11* would be more likely to become malignant than cells lacking *PTPN11*.

Cellular mechanisms of enchondroma formation

Enchondromas have been suggested to arise either from growth plate chondrocytes or from progenitor cells in the marrow cavity that inappropriately differentiate into chondrocytes [18,19]. Previous mouse models of enchondroma formation (mutant *PTHR1* or over-expression of *Gli2*) indicated that enchondromas in the long bones can arise from growth plate chondrocytes [20]. Inactivation of *Ptpn11* in chondrocytes was found to induce multiple enchondromas in the ribs and vertebrae [9,21] (**Chapter 3**), but only rarely affected the long bones – I observed an enchondroma in the tibia in only 1 out of 7 mice examined (data not shown). It is not known whether long bone enchondromas would have formed more frequently if *Ptpn11* had been inactivated at an earlier time point or in a greater percentage of chondrocytes, or if the mice had been analyzed beyond 12 weeks of age; nevertheless, these data

indicate that enchondromas can arise from growth plate chondrocytes.

I found that loss of *Ptpn11 in vitro* delayed the terminal differentiation of hypertrophic chondrocytes, and that mosaic postnatal inactivation of *Ptpn11* in chondrocytes *in vivo* led to non-uniform rates of proliferation, maturation, terminal differentiation and ossification in vertebral growth plates (**Chapter 3**). Importantly, ossification centers formed within vertebral growth plates, thereby causing the chondrocytes below them to become displaced from the growth plate. These displaced chondrocytes formed enchondroma-like lesions and contained both immature and mature chondrocytes (as assessed by COLX immunohistochemistry), and both *Ptpn11*-deficient and wild-type chondrocytes (as assessed by analysis of a reporter allele which marks cells in which Cre recombinase has acted) (**Chapter 3**). Together, these data suggest that MC enchondromas may arise due to both disrupted growth plate organization and delayed terminal differentiation, rather than simply due to neoplastic growth or delayed cartilage resorption alone (Figure 5.2).
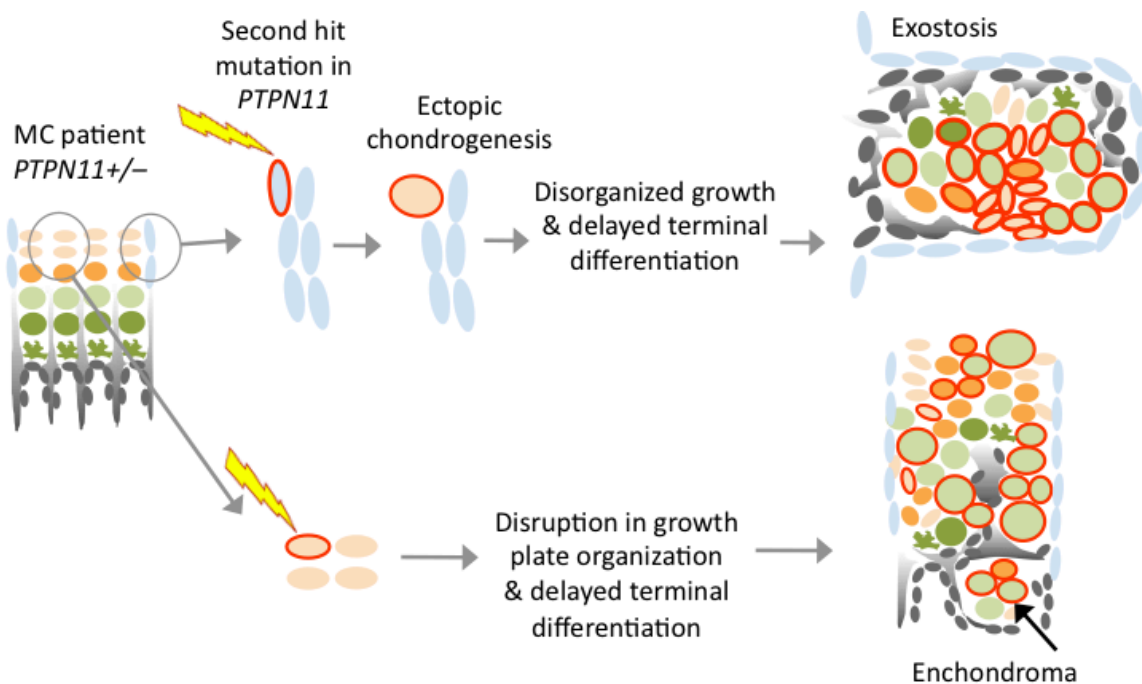
Figure 5.2: Model for exostosis and enchondroma formation in MC patients. **A:** Inactivation of *PTPN11* in perichondrial cells causes ectopic chondrogenesis. The resulting chondrocytes proliferate in a disorganized fashion to form an exostosis. Inactivation of *PTPN11* in growth plate chondrocytes disrupts the architecture of the growth plate and delays terminal differentiation, thereby leading to enchondroma formation. Enchondromas and exostoses likely contain both *PTPN11*-deficient (red outline) and *PTPN11*[+/−] chondrocytes.

The observation that vertebral enchondromas can arise from the growth plate does not rule out the possibility that enchondromas could also arise from progenitor cells in the marrow cavity. It would be of interest to inactivate *Ptpn11* in mesenchymal stem cells (MSCs) in the marrow cavity and determine whether this leads to ectopic cartilage formation. This could be achieved by using the *Nestin-CreER* or *Mx1-Cre drivers*. The *Nestin-CreER* driver is expressed in nerve cells, but has recently been shown to also be expressed in a population of perivascular MSCs in the marrow cavity that can give rise to osteoblasts and chondrocytes [22]. The *Mx1-Cre* driver can be used to label a population of osteogenic stem cells in the marrow cavity that partially overlaps with the Nestin+ MSC population [23]. These Cre drivers could be used to determine whether inactivation of *Ptpn11* in progenitor cells in the marrow cavity can cause them to differentiate into chondrocytes instead of osteoblasts. Alternatively, local administration of adenoviral Cre could be performed by injection into the marrow cavity.

Interestingly, an enchondroma in the proximal femur was observed when *Ptpn11* was inactivated in the perichondrium using *Ctsk-Cre* [10]. This suggests that enchondromas could also arise from perichondrial cells that inappropriately undergo chondrogenesis and proliferate inwards towards the marrow cavity to form an enchondroma. However, since a fraction of *Ctsk-Cre*-positive cells were shown to migrate from the perichondrium into the growth plate and articular cartilage during normal development [10], it is also possible that the enchondroma observed in the *Ctsk-Cre;Ptpn11*$^{fl/fl}$ mouse arose from the growth plate. Further analysis of *Ctsk-Cre;Ptpn11*$^{fl/fl}$ mice should be performed to determine how frequently and in which locations enchondromas form, and earlier stages of the disease should be examined to determine whether the enchondromas arise from perichondrial cells or growth plate chondrocytes.

It is not yet known whether the mechanism of enchondroma formation in MC, Ollier disease and Maffucci syndrome patients is the same. Ollier disease and Maffucci syndrome are caused by somatic missense mutations in *IDH1* or *IDH2,* the most common of which is the R132C mutation in *IDH1* [6,7,24]. While these mutations have been shown to lead to increased DNA methylation [25], it is not yet know how they affect the growth plate and lead to enchondroma formation. An inducible *IDH1*(*R132H*) knock-in mouse has recently been generated [26]. Preliminary results indicate that expression of mutant *IDH1* in chondrocytes induces enchondroma formation and reduces the height of the hypertrophic zone (Benjamin

Alman, personal communication, Nov 2012). It will be of interest to characterize the enchondromas in these mice and determine whether their location, frequency, and histology are similar to or different from enchondromas in *Ptpn11* cKO mice. Interestingly a shorter hypertrophic zone was also observed in other mouse models of enchondroma formation (mutant *PTHR1* or over-expression of *Gli2*) [20], in contrast to the expanded hypertrophic zone observed following *Ptpn11* inactivation (**Chapter 3**). This suggests that mutations in *Ptpn11* and *IDH1/IDH2* have different effects on chondrocyte maturation.

Molecular mechanisms of enchondroma formation

PTPN11 encodes a phosphatase, SHP2, involved in many signaling events [17]. The signaling pathways that contribute to the disruption in vertebral growth plate organization, delay in terminal differentiation and enchondroma formation that we observed following SHP2 depletion *in vivo* (**Chapter 3**) are not yet known. One likely candidate pathway for contributing to both the disrupted organization and delay in terminal differentiation is the ERK1/2 pathway. Inactivation of *Ptpn11* in chondrocytes decreases phospho-ERK1/2 immunoreactivity [9,10,27] and inhibiting the ERK1/2 pathway in chondrocyte pellet cultures delays terminal differentiation in a similar manner to SHP2 depletion (**Chapter 3**). Furthermore, there is a striking overlap in the phenotypes of mice with an osteo-chondroprogenitor specific deletion of *Ptpn11* or of *Erk1* and *Erk2* – in both cases the mice develop severe chondrodysplasia and an accumulation of hypertrophic chondrocytes [27,28]. Similarly, mice with a chondrocyte-specific deletion of *Ptpn11* or of *Erk1* and *Erk2* both develop an expanded hypertrophic zone in the long bones and a severe
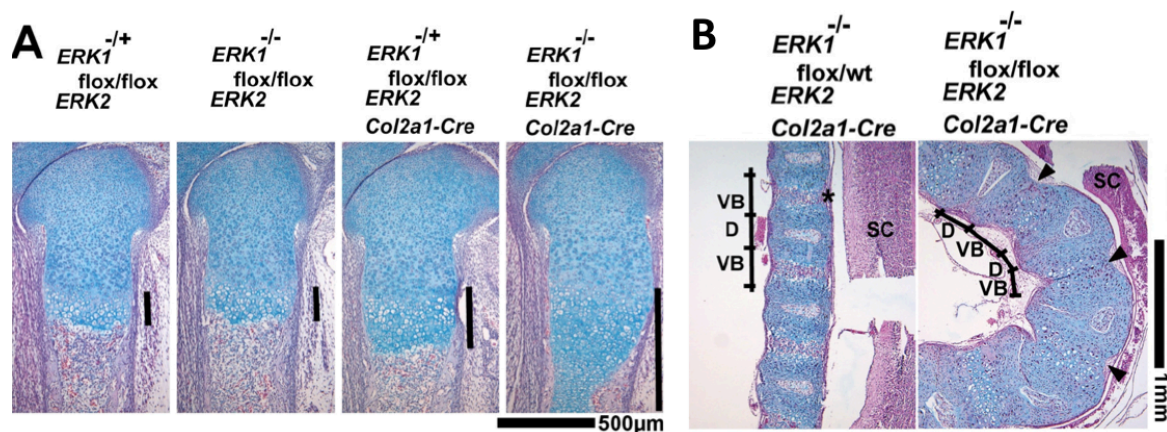
**Figure 5.3: Inactivation of *Erk1* and *Erk2* in chondrocytes leads to an accumulation of hypertrophic chondrocytes in the long bones and an expansion and disorganization of vertebral growth plates**. Hemotoxylin, eosin and alcian blue staining of the E16.5 femur (A) and E18.5 spine (B) in Ctrl mice or mice lacking *Erk1* and *Erk2* in chondrocytes. Images adapted from Matsushita et al (2009).

expansion and disorganization of the vertebral growth plates [8,21,28] (**Chapter 3**) (Figure 5.3). Together, these studies indicate that ERK1/2 is a likely candidate for acting downstream of SHP2 in the growth plate. It will be of interest to determine whether inactivation of *Erk1* and *Erk2* is sufficient to induce enchondroma formation. The vertebral growth plates of mice lacking *Erk1* and *Erk2* were only studied at embryonic stages [28]; thus, an area of future research would be to inactivate *Erk1* and *Erk2* postnataly in chondrocytes and determine whether enchondromas arise.

SHP2 is thought to act downstream of most, if not all, receptor tyrosine kinases (RTKs) [17]. However, there are 58 known RTKs [29] and it is not yet known which of these receptors SHP2 modulates specifically in chondrocytes. Our finding that loss of SHP2 delays terminal differentiation suggests that SHP2 acts downstream of EGFR, FGFR1 or FGFR3, since each one of these receptors has been suggested to promote terminal differentiation [30–32]. FGFR3 is a particularly strong candidate, since *Fgfr3* knockout mice develop enchondroma-like lesions in some long bones [33] (Figure 5.4). Furthermore, *Fgfr3* knockout mice have an increased rate of chondrocyte proliferation and develop expanded and disorganized vertebral growth plates [33,34] (Figure 5.4). The vertebral phenotype is similar to what is observed following *Ptpn11* inactivation (**Chapter 3**). However, since the phenotype in *Fgfr3* knockout mice is less severe than in *Ptpn11* cKO mice, loss of signaling through FGFR3 cannot be solely responsible for the phenotypes that arise following SHP2 depletion in chondrocytes. Thus, the consequences of SHP2 depletion in chondrocytes are likely to be complex.

In the growth plate, a positive feedback loop exists between IHH and BMP signaling, whereas
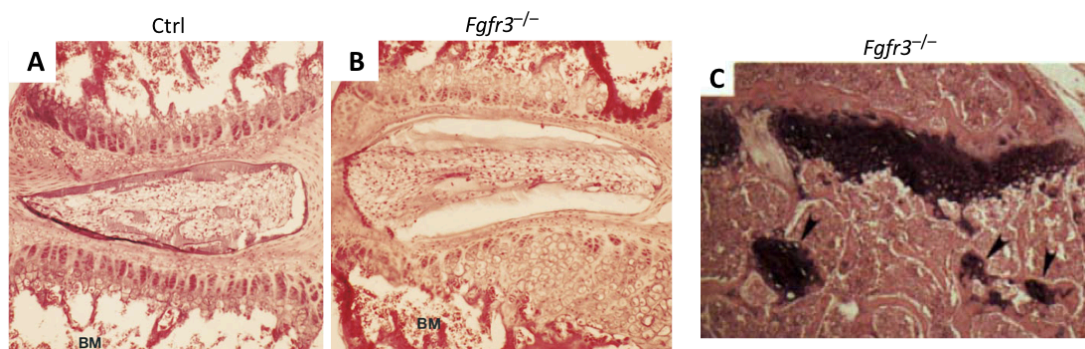
Figure 5.4: Inactivation of *Fgfr3* leads to expanded and disorganized vertebral growth plates, and enchondroma-like lesions in the long bones. **A-B:** Sections of a vertebral growth plate from a P20 wild-type mouse (A) or *Fgfr3* knockout mouse (B). Image adapted from Deng *et al* (1996) **C:** Section of the distal femur of a 6.3 month-old *Fgfr3* knockout mouse. Note the enchondroma-like lesions below the growth plate (arrows). Image adapted from Colvin *et al* (1996).

FGF signaling antagonizes IHH and BMP signaling [32,35]. Since SHP2 and ERK1/2 can act downstream of FGF growth factors [36], it would be expected that loss of SHP2 or ERK1/2 signaling would promote IHH and BMP signaling. Indeed, I found that *Ihh* was one of the transcripts with the most substantial increases in abundance after SHP2 depletion or ERK1/2 inhibition in pellet cultures (**Chapter 3**). Similarly, I also observed an increase in the abundance of transcripts encoding transcription factors known to both be induced by BMP signaling and to inhibit terminal differentiation (*Hey1, Msx2, Dlx5, Sox9*) [37–39][40–43] (**Chapter 3**). Together these results suggest that loss of SHP2 leads to a decrease in ERK1/2 signaling, which may increase IHH and BMP signaling. IHH and BMP both promote chondrocyte proliferation [44] and BMP signaling has been suggested to delay the terminal differentiation of hypertrophic chondrocytes [32]. Therefore, increased IHH and BMP signaling could be factors contributing to the expansion of the vertebral growth plates observed in *Ptpn11* cKO mice (**Chapter 3**). Interestingly, overexpression of *Ihh* accelerates the differentiation of periarticular chondrocytes into proliferative chondrocytes [45]. Thus, it is possible that increased IHH levels in *Ptpn11* cKO vertebral growth plates accelerates the maturation of resting zone chondrocytes, which would explain why some growth plates were found to be completely resorbed by week 10 (ie their resting zones had been depleted) (**Chapter 3**). It would be of interest to determine whether the vertebral growth plate phenotype that results from *Ptpn11* inactivation could be rescued by decreasing IHH or BMP signaling, for example by using mice that are heterozygous for loss-of-function mutations in IHH signaling components, *Ihh* or *Smo*, or BMP pathway components, *Smad1*, *Smad5* and *Smad8*.

It is not known whether chondrocytes in the vertebral growth plates of *Ptpn11* cKO mice are disorganized simply due to unequal rates of proliferation between *Ptpn11*-deficient and wild-type clusters of chondrocytes, or whether loss of *Ptpn11* has a direct effect on chondrocyte polarity and columnar organization. In normal growth plates, chondrocyte orientation and polarity are tightly controlled [46]. The cell division plane of chondrocytes in the proliferative zone is always oriented such that daughter cells are displaced lateral to the axis of the column. Daughter cells then organize into columns through cell intercalation [46]. An area of future research would be to determine whether loss of *Ptpn11* affects either the orientation of the cell division plane or the intercalation of chondrocytes to form columns, using the methods described in Ahrens et al (2009) [47].

In summary, many signaling pathways are likely to be affected by SHP2 depletion and contribute to enchondroma formation. It will be of interest to perform *in situ* hybridization or immunohistochemistry on *Ptpn11* cKO growth plates to determine the activation status of pathways with known roles in the growth plate. In addition, further research into the signaling pathways regulating vertebral growth plates versus long bone growth plates may further our understanding of why loss of SHP2 has a more severe effect in the vertebrae than the long bones.

Cellular mechanisms of exostosis formation

Exostoses have been hypothesized to arise either from perichondrial cells that inappropriately undergo chondrogenesis, or from chondrocytes that escape the normal boundaries of the growth plate [48]. Mouse data support a perichondrial origin for MC exostoses. Firstly, exostoses can be induced by inactivating *Ptpn11* using *Ctsk-Cre*, which is expressed in the perichondrium and not in growth plate chondrocytes [10]. Secondly, when *Ptpn11* is inactivated using *Col2a1-CreERT2,* which is expressed in chondrocytes and a few perichondrial cells, clusters of ectopic chondrocytes in the perichondrium/periostium are observed in the metacarpals after the growth plate has already closed [9]. These studies suggest that loss of *Ptpn11* causes perichondrial cells to differentiate into chondrocytes, which then proliferate in a disorganized manner to form an exostosis (Figure 5.2). Interestingly, while exostoses formed on all long bones when *Ptpn11* was inactivated using *Ctsk-Cre* [10], exostoses formed only on some bones (eg. metacarpals) and not others (eg. femur and tibia) when the *Col2a1-CreERT2* driver was used [9]. It is not known whether this is due to differences in efficiency of *Col2a1-CreERT2* and *Ctsk-Cre* expression in the perichondrium in different bone types or due to differences in the timing of *Ptpn11* inactivation. Further studies using a reporter allele could be performed to compare the perichondrial expression of *Ctsk-Cre* and *Col2a1-CreERT2* in different bone types and at different time points.

It has been debated whether exostoses induced by loss of *Ext1* originate from growth plate chondrocytes or perichondrial cells [11,49,50]. One hypothesis that is consistent with all published mouse data is that bi-allelic loss of *Ext1* in either chondrocytes or perichondrial cells is sufficient to induce chondrogenesis in the perichondrium, and that once these ectopic chondrocytes start to proliferate, neighboring growth plate chondrocytes can be co-opted into the growing lesions. Interestingly,

compound heterozygous *Ext1*[+/−]*; Ext2*[+/−] mice also develop exostoses [51], which suggests that a local reduction in the levels of extracellular heparan sulfate is sufficient to affect paracrine signaling in the perichondrium, thereby inducing ectopic chondrogenesis. Thus, it is likely that exostoses from both MC patients and MHE patients arise due to inappropriate chondrogenesis in the perichondrium. However, the way in which the ectopic clusters of chondrocytes expand to form an exostosis likely differs between MC and MHE. This statement is based on observations that MHE and MC lesions differ at the histological level. MHE exostoses have a cartilage cap that displays relatively normal growth plate architecture with a zone of hypertrophic (COLX-positive) chondrocytes at its base. In contrast, we found that MC exostoses have a cartilage core with a disorganized distribution of COLX-positive chondrocytes (**Chapter 2 & 3**). Similarly, when comparing the histology of exostoses in the radius and ulna in mice, exostoses arising due to *Ext1* inactivation contain chondrocyte columns oriented perpendicular to the main axis of the growth plate [52], while exostoses arising due to *Ptpn11* inactivation lack organization [9]. Furthermore, while postnatal inactivation of *Ext1* in chondrocytes decreases the height of the hypertrophic zone [52], an expanded hypertrophic zone is observed following *Ptpn11* inactivation (**Chapter 3**). Thus, although loss of *Ptpn11* and loss of *Ext1* each induces chondrogenesis in the perichondrium, the consequences on chondrocyte organization and maturation are different.

Molecular mechanisms of exostosis formation

One pathway that may act downstream of SHP2 in perichondrial cells is the ERK1/2 pathway. Loss of SHP2 in the perichondrium reduces phospho-ERK1/2 immunoreactivity [10]. Furthermore, inactivating *Erk1* and *Erk2* in osteo-chondroprogenitor cells (*Prx1-Cre*) was found to induce ectopic chondrogenesis in the perichondrium; however, due to the severity of the skeletal defects, these mice were only examined at embryonic and early postnatal stages [28] (Figure 5.5). Therefore, it is not known whether the ectopic cartilage clusters would have progressed to exostoses had the mice survived until adulthood. An area of future research would be to inactivate *Erk1* and *Erk2* in perichondrial cells (*Ctsk-Cre* [10]) or in a mosaic manner in osteo-chondroprogenitors (*Prx1-CreER*), and examine adult mice to determine whether exostoses develop. An alternate approach would be to constitutively activate the ERK1/2 pathway and determine whether this rescues exostosis formation in *Ptpn11* cKO mice. This could be achieved using a constitutively active *MEK1* transgene that is under the control of the *Prx1* promoter

[28]. Together, these experiments will determine whether the reduction in ERK1/2 signaling following loss of SHP2 is responsible for exostosis formation.

Other signaling pathways implicated in regulating the cell fate of perichondrial cells are the WNT, BMP and IHH pathways. Inactivation of beta-catenin in osteo-chondroprogenitors (*Prx1-Cre*) induces ectopic cartilage formation in the perichondrium [53,54], and postnatal inactivation of *beta-catenin* in chondrocytes induces exostosis-like outgrowths [55]. In
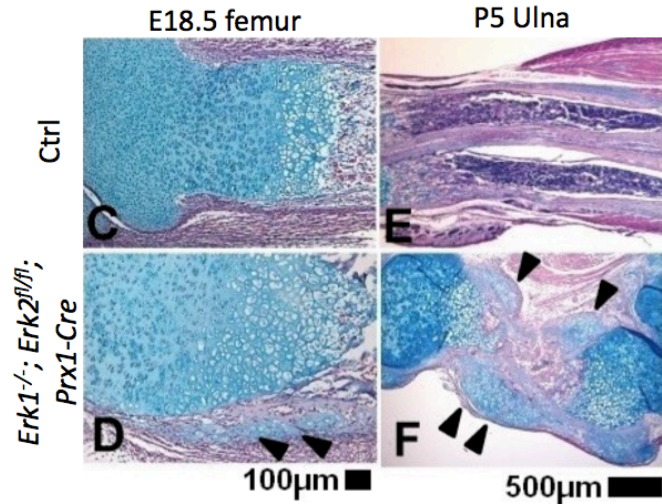
Figure 5.5: Ectopic chondrogenesis in the perichondrium following inactivation of *Erk1* and *Erk2*. Tissue sections of the femur and ulna of Ctrl mice or mice lacking *Erk1* and *Erk2* in osteo-chondroprogenitor cells. Arrows indicate ectopic cartilage in the perichondrium. Images adapted from Matsushita *et al* (2009).

metatarsal organ cultures, activating BMP signaling while inhibiting HH signaling induces ectopic cartilage in the perichondrium [56]. Similarly, inactivating HH signaling by inactivation of *Smoothened* in perichondrial cells induces ectopic cartilage formation and prevents bone collar formation in long bones [57]. In contrast, it was also suggested that excessive IHH signaling in the perichondrium of synchondroses, caused by deletion of *Kif3a* in chondrocytes, drives ectopic cartilage formation [58]. In summary, inactivating WNT signaling or activating BMP signaling promotes ectopic chondrogenesis, while conflicting data exist on the role of HH signaling in promoting ectopic chondrogenesis. It will be of interest to determine whether loss of SHP2 in the perichondrium decreases WNT signaling or increases BMP signaling, and whether modulating either of these pathways can decrease the rate of exostosis formation in *Ptpn11* cKO mice.

Can metachondromatosis be treated?

The formation of exostoses and enchondromas in MC patients can lead to nerve or blood vessel compression, restricted motion of joints and deformity of the affected bones [59]. Exostoses can be surgically removed, but may re-grow after surgery. Therefore, it would be of great benefit to MC patients if a method to prevent lesion formation or promote lesion regression could be identified. It was found that

treating *Ptpn11* cKO mice with an inhibitor of HH signaling reduced the severity of exostosis formation [10]. This is likely due to the role of IHH in promoting chondrocyte proliferation [44]. It will be of interest to determine whether inhibiting HH signaling can also reduce enchondroma formation or reduce the expansion of the vertebral growth plates observed after *Ptpn11* inactivation. Also of importance will be to determine whether inhibiting HH signaling is only beneficial if treatment is started at a young age, or whether inhibiting HH signaling in adult mice could induce the regression of exostoses or enchondromas that have already formed. Finally, careful studies will be needed to determine whether inhibiting HH signaling in children would negatively effect skeletal development. In one study it was found that treating young mice with a HH inhibitor led to permanent skeletal defects [60]. In humans, inhibiting HH signaling is being used to treat cancer in adults, but has not yet been tested on children [61]. It could also be tested whether an inhibitor of HH signaling could be administered locally, by injection into the lesional tissue, rather than systemic administration.

Our observation that loss of *Ptpn11* delays chondrocyte terminal differentiation (**Chapter 3**) raises the possibility that inducing terminal differentiation may promote the regression of lesions. An area of future research would be to induce exostoses and enchondromas in mice by inactivating *Ptpn11*, and then restoring *Ptpn11* by using a reversible gene trap allele and determining whether this induces chondrocytes to terminally differentiation, thereby promoting lesion regression. Alternatively, activation of pathways known to promote terminal differentiation, such as the ERK1/2 pathway, may promote the regression of lesions.

Developing more efficient methods for mapping zebrafish mutants

In **chapter 4**, I focused on the use of zebrafish as a model organism for studying the genetic regulation of skeletal development. The laboratory of Dr. Matthew Harris isolated a variety of recessive zebrafish mutants with adult phenotypes, including defects in the skeleton. Since the traditional method for mapping and cloning zebrafish mutations is time consuming and costly, I developed a method for mapping and cloning zebrafish mutations using low-coverage whole genome sequencing (WGS). Since the existing genetic variation in zebrafish was not well characterized, I first performed WGS on several wild-type laboratory strains and created an extensive database of common polymorphisms. I then incorporated this SNP database into a data analysis pipeline that I developed to map mutants based on

homozygosity-by-descent. For each mutant, I performed WGS on pooled DNA from ~20 mutant offspring from a mapping cross. I developed an algorithm to scan the genome in sliding windows, searching for regions of reduced diversity that were the expected genetic size (20 cM). Regions of reduced diversity smaller than 20 cM were excluded and may represent haplotype blocks that were shared in the parental strains. I showed that this method could be used to successfully map five different zebrafish mutants, which was confirmed with independent SSLP mapping data (**Chapter 4**).

Other investigators have also recently developed methods to use NGS to map zebrafish mutations [62–66]. All methods rely on the basic principle of sequencing pooled DNA, obtained from >20 mutant offspring from an outcross, and performing homozygosity mapping using SNPs identified in the sequence data. However, some aspects of the experimental design and data analysis differ between these published methods (Figure 5.6). Here, I firstly discuss the two main differences between these published methods, and secondly provide suggestions for future work that may improve the accuracy and lower the cost of mapping.

The first key difference between published methods is the use of WGS [62–64] versus RNA-seq [65,66] (Figure 5.6). Each sequencing method has advantages and drawbacks. With RNA-seq, only the

Bowen *et al* 2012 (This thesis)	Leshchiner *et al* 2012	Obholzer *et al* 2012	Voz *et al* 2012	Hill *et al* 2013	Miller *et al* 2013
1. Perform a mapping cross					
2. Pool DNA/RNA from >20 mutant offspring					
3. Perform whole genome sequencing (WGS) or RNA-seq					
WGS	WGS	WGS	WGS	RNA-seq	RNA-seq
4. To facilitate accurate SNP identification, compare sequence data to:					
WGS from wild-type strains	WGS from unaffected siblings	WGS from wild-type strains	n/a	n/a	n/a
5. Scan the genome for regions of reduced diversity, using window sizes based on:					
Genetic distance (cM)	Physical distance (bp)	Genetic distance (cM)	SNP density	Genetic distance (cM)	SNP density

Figure 5.6. Comparison of published methods for mapping zebrafish mutations using Next Generation Sequencing.

transcriptome is sequenced, rather than the entire genome. Thus, greater fold coverage of the coding sequence can be obtained for a lower cost, thereby enabling more accurate variant detection. In addition, RNA-seq would be capable of identifying causal mutations that affected splicing or expression levels. However, a disadvantage of RNA-seq is that the coverage is not distributed evenly and will be biased towards genes that are highly expressed at the particular time-point or in the particular tissue type that was used for RNA extraction. This will lower the likelihood of identifying the causal mutation if it lies in a gene that is not abundantly expressed. Another disadvantage of RNA-seq is that there will be far fewer polymorphisms available for mapping, since the majority of polymorphisms lie in non-coding regions of the genome. Thus, mapping may be difficult if the causative mutation lies in a gene desert, or in a region of low diversity due to shared haplotype blocks between the parental strains.

The second difference between published methods is the use of genetic distances (cM) versus physical distances (bp) (Figure 5.6). In each of the published methods for mapping zebrafish mutations, the data analysis pipeline involves subdividing the genome into bins, and searching for the bin with the lowest level of heterogeneity. Using a large bin size allows one to eliminate false positive regions that have reduced heterogeneity due to short haplotype blocks that were shared by the parental strains. What differs between methods is whether these bins are defined based on a physical distance [62], genetic distance [63,65] (this thesis), or number of SNPs [64,66]. We believe that defining the bin size based on genetic distance is most accurate. Recombination rates are not equal throughout the genome, with high rates of recombination in telomeres and low rates of recombination in centromeres. If a causative mutation lay close to the telomere, the sequence linked to the mutation would span a relatively small physical distance, and therefore this region of reduced heterogeneity might escape detection if the bin size was based on physical distance. Nevertheless, despite the differences between published methods, each method was shown to be capable of mapping at least one zebrafish mutant. Therefore, any one of these methods is likely to be a good starting point for investigators who wish to perform mapping based on NGS.

The drawback of NGS is that data analysis can be complicated and often requires a significant time investment. Therefore, for NGS to be widely adopted by zebrafish labs, an automated and easy-to-use data analysis pipeline needs to be made available. Four such pipelines have recently been

developed: SNPtrack [63], RNAmapper [66], MMAPPR [65], and MegaMapper [63]. These pipelines provide an excellent starting point for labs that do not have expertise in analyzing NGS data. However, these pipelines could be improved by incorporating a more extensive database of known common polymorphisms. This would increase the accuracy of mapping when dealing with low coverage sequence data. In regions of low coverage, if a variant is observed in only one or two sequencing reads, it is difficult to know whether it represents a sequencing error or a true polymorphism. Noise from sequencing errors can be significantly reduced by excluding all variants that do not occur at sites of known common polymorphisms. Therefore, the more comprehensive the list of common polymorphisms, the more accurate this method of mapping is likely to be. We compiled a database of 7.6 million common polymorphisms that we identified in our WGS data from several wild-type strains (**Chapter 4**). This database could now be improved by incorporating additional WGS data from other zebrafish wild-type strains and mutants that have subsequently been sequenced [62–66]. Not only will this allow for more accurate mapping, but, once the linked interval has been identified, it will help narrow down the list of potential causal mutations (since variants that are known common polymorphisms can be excluded as being causal).

Future improvements in the quality of the zebrafish genome assembly will also make mapping more accurate. As noted by Leshchiner *et al* (2012), accurate mapping can be compromised by incorrectly placed scaffolds in the genome assembly [62]. Furthermore, some scaffolds exist in the Zv9 version of the zebrafish genome that have not yet been placed on a chromosome. Mapping a mutation that lies on one of these unassigned scaffolds is not likely to be successful. The accuracy of mapping is likely to be improved when the next version of the genome (Zv10) is released in 2014 (http://www.sanger.ac.uk). Future improvements in the annotation of the zebrafish genome will also be beneficial, since if a causal mutation lay in an unannotated gene, it is likely that it would be overlooked.

It will also be of interest to determine whether the cost of mapping could be reduced by multiplexing samples and performing targeted sequencing of the exome instead of WGS or RNA-seq. Since the exome is two orders of magnitude smaller than the genome, sufficient coverage could be obtained even if multiple samples were sequenced per lane. Targeted sequencing of the exome may be a more robust method than RNA-seq, since one is likely to obtain more even coverage across all exons.

This is in contrast to RNA-seq, which provides extremely high coverage for abundant transcripts, but low or no coverage for low abundance transcripts. Agilent has only recently developed a zebrafish exome capture array (www.agilent.com). Using this capture array, Ryan *et al* (2013) showed that zebrafish mutants can be mapped by obtaining high coverage of the exome [67]. Thus, an area of future research would be to sequence multiple libraries per lane to obtain low coverage of the exome, and determine whether this is sufficient for mapping.

Finally, it will also be of interest to develop tools for using NGS to map mutations in other model organisms. Whole genome or exome sequencing has been used to identify mutations in mice, but studies normally rely on prior linkage data, obtained using SNP arrays, to determine which of the many mutations identified is the causative mutation [68,69]. Leshchiner *et al* (2012) demonstrated that the WGS approach they developed to map zebrafish mutations could also be used to map a mouse ENU-induced mutation [62]. It will be of interest to determine whether these methods can also be applied to map traits in other organisms with large and polymorphic genomes, such as chickens, pigs, rabbits and other fish species.

5.2 Conclusions

In the first component of this thesis, I showed that *PTPN11*/SHP2 acts as a suppressor of benign cartilage tumors and regulates the proliferation, maturation and organization of growth plate chondrocytes, which may be mediated in part by the ERK1/2 pathway. The work in this thesis demonstrates that the cartilage tumors in MC patients may form due to different genetic, cellular and molecular mechanisms to the cartilage tumors in MHE, Ollier disease and Maffucci Syndrome. Furthermore, while gain-of-function mutations in *PTPN11* are known to promote childhood leukemia and developmental defects associated with Noonan Syndrome, the work in this thesis is the first to associate loss of *PTPN11* with a human disorder. Future studies may reveal why different tissue types have different sensitivities to either gain or loss of SHP2 activity.

In the second component of this thesis, I demonstrate that low-coverage whole genome sequencing can be used to map zebrafish mutations based on homozygosity-by-descent. The data analysis pipelines I developed, together with computational tools developed by other investigators, are likely to expedite the process of mapping mutations in zebrafish. This will increase the utility of zebrafish as a model organism to study the genetic regulation of skeletal development. Now that mapping using

NGS has been shown to be possible, future work will be needed to firstly more fully characterize the existing genetic variation in laboratory zebrafish strains, secondly determine to what extent the cost of mapping can be reduced by multiplexing samples, and thirdly develop robust and easy-to-use software for investigators without expertise in the analysis of NGS data.

Finally, the novel NGS approaches I developed have been utilized to address additional questions related to skeletal development in disease that were beyond the scope of this thesis. For example, we used the array that I designed to target genes acting in the same pathway as *PTPN11* to identify mutations in *PIK3CA* as the genetic cause for the overgrowth syndrome, CLOVES [70], and I applied the WGS mapping pipeline I developed to a craniofacial trait that differs between the DBA/1J and DBA/2J mouse strains [71]. As sequencing technologies continue to improve, with longer read lengths, higher accuracy and greater throughput, approaches using NGS will continue to expedite the process of mutation identification, mapping and gene expression profiling in humans and model organisms.

5.3   Literature Cited

1. Bassett GS, Cowell HR: **Metachondromatosis. Report of four cases**. *J Bone Joint Surg Am* 1985, **67**:811–814.

2. Romeo S, Hogendoorn PCW, Dei Tos AP: **Benign cartilaginous tumors of bone: from morphology to somatic and germ-line genetics**. *Adv Anat Pathol* 2009, **16**:307–315.

3. Ahn J, Lüdecke HJ, Lindow S, Horton WA, Lee B, Wagner MJ, Horsthemke B, Wells DE: **Cloning of the putative tumour suppressor gene for hereditary multiple exostoses (EXT1)**. *Nat. Genet.* 1995, **11**:137–143.

4. Stickens D, Clines G, Burbee D, Ramos P, Thomas S, Hogue D, Hecht JT, Lovett M, Evans GA: **The EXT2 multiple exostoses gene defines a family of putative tumour suppressor genes**. *Nat. Genet.* 1996, **14**:25–32.

5. Sobreira NLM, Cirulli ET, Avramopoulos D, Wohler E, Oswald GL, Stevens EL, Ge D, Shianna KV, Smith JP, Maia JM, et al.: **Whole-genome sequencing of a single proband together with linkage analysis identifies a Mendelian disease gene**. *PLoS Genet.* 2010, **6**:e1000991.

6. Amary MF, Damato S, Halai D, Eskandarpour M, Berisha F, Bonar F, McCarthy S, Fantin VR, Straley KS, Lobo S, et al.: **Ollier disease and Maffucci syndrome are caused by somatic mosaic mutations of IDH1 and IDH2**. *Nature Genetics* 2011, **43**:1262–1265.

7. Pansuriya TC, van Eijk R, d' Adamo P, van Ruler MAJH, Kuijjer ML, Oosting J, Cleton-Jansen A-M, van Oosterwijk JG, Verbeke SLJ, Meijer D, et al.: **Somatic mosaic IDH1 and IDH2 mutations are associated with enchondroma and spindle cell hemangioma in Ollier disease and Maffucci syndrome**. *Nature Genetics* 2011, **43**:1256–1261.

8. Bauler TJ, Kamiya N, Lapinski PE, Langewisch E, Mishina Y, Wilkinson JE, Feng G-S, King PD: **Development of severe skeletal defects in induced SHP-2-deficient adult mice: a model of skeletal malformation in humans with SHP-2 mutations**. *Dis Model Mech* 2011, **4**:228–239.

9. Kim HKW, Feng G-S, Chen D, King PD, Kamiya N: **Targeted disruption of Shp2 in chondrocytes leads to metachondromatosis with multiple cartilaginous protrusions**. *J. Bone Miner. Res.* 2013, doi:10.1002/jbmr.2062.

10. Yang W, Wang J, Moore DC, Liang H, Dooner M, Wu Q, Terek R, Chen Q, Ehrlich MG, Quesenberry PJ, et al.: **Ptpn11 deletion in a novel progenitor causes metachondromatosis by inducing hedgehog signalling**. *Nature* 2013, **499**:491–495.

11. Matsumoto K, Irie F, Mackem S, Yamaguchi Y: **A mouse model of chondrocyte-specific somatic mutation reveals a role for Ext1 loss of heterozygosity in multiple hereditary exostoses**. *Proceedings of the National Academy of Sciences* 2010, **107**:10932–10937.

12. Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, Makarewicz AJ, Bright IJ, Lucero MY, Hiddessen AL, Legler TC, et al.: **High-throughput droplet digital PCR system for absolute quantitation of DNA copy number**. *Anal. Chem.* 2011, **83**:8604–8610.

13. Chan G, Cheung LS, Yang W, Milyavsky M, Sanders AD, Gu S, Hong WX, Liu AX, Wang X, Barbara M, et al.: **Essential role for Ptpn11 in survival of hematopoietic stem and progenitor cells**. *Blood* 2011, **117**:4253–4261.

14. Pannier S, Legeai-Mallet L: **Hereditary multiple exostoses and enchondromatosis**. *Best Pract Res Clin Rheumatol* 2008, **22**:45–54.

15. Mavrogenis AF, Skarpidi E, Papakonstantinou O, Papagelopoulos PJ: **Chondrosarcoma in metachondromatosis: a case report**. *J Bone Joint Surg Am* 2010, **92**:1507–1513.

16. de Andrea CE, Reijnders CMA, Kroon HM, de Jong D, Hogendoorn PCW, Szuhai K, Bovée JVMG: **Secondary peripheral chondrosarcoma evolving from osteochondroma as a result of outgrowth of cells with functional EXT**. *Oncogene* 2012, **31**:1095–1104.

17. Grossmann KS, Rosário M, Birchmeier C, Birchmeier W: **The tyrosine phosphatase Shp2 in development and cancer**. *Adv. Cancer Res.* 2010, **106**:53–89.

18. Aigner T: **Towards a new understanding and classification of chondrogenic neoplasias of the skeleton--biochemistry and cell biology of chondrosarcoma and its variants**. *Virchows Arch.* 2002, **441**:219–230.

19. Douis H, Davies AM, James SL, Kindblom LG, Grimer RJ, Johnson KJ: **Can MR imaging challenge the commonly accepted theory of the pathogenesis of solitary enchondroma of long bone?** *Skeletal Radiology* 2012, **41**:1537–1542.

20. Hopyan S, Gokgoz N, Poon R, Gensure RC, Yu C, Cole WG, Bell RS, Jüppner H, Andrulis IL, Wunder JS, et al.: **A mutant PTH/PTHrP type I receptor in enchondromatosis**. *Nature Genetics* 2002, **30**:306–310.

21. Kim HKW, Aruwajoye O, Sucato D, Richards BS, Feng G-S, Chen D, King P, Kamiya N: **Induction of SHP2-deficiency in chondrocytes causes severe scoliosis and kyphosis in mice**. *Spine* 2013, doi:10.1097/BRS.0b013e3182a3d370.

22. Méndez-Ferrer S, Michurina TV, Ferraro F, Mazloom AR, MacArthur BD, Lira SA, Scadden DT, Ma'ayan A, Enikolopov GN, Frenette PS: **Mesenchymal and haematopoietic stem cells form a unique bone marrow niche**. *Nature* 2010, **466**:829–834.

23. Park D, Spencer JA, Koh BI, Kobayashi T, Fujisaki J, Clemens TL, Lin CP, Kronenberg HM, Scadden DT: **Endogenous Bone Marrow MSCs Are Dynamic, Fate-Restricted Participants in Bone Maintenance and Regeneration**. *Cell Stem Cell* 2012, **10**:259–272.

24. Amary MF, Bacsi K, Maggiani F, Damato S, Halai D, Berisha F, Pollock R, O'Donnell P, Grigoriadis A, Diss T, et al.: **IDH1 and IDH2 mutations are frequent events in central chondrosarcoma and central and periosteal chondromas but not in other mesenchymal tumours**. *The Journal of Pathology* 2011, **224**:334–343.

25. Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A, Li Y, Bhagwat N, Vasanthakumar A, Fernandez HF, et al.: **Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation**. *Cancer Cell* 2010, **18**:553–567.

26. Sasaki M, Knobbe CB, Munger JC, Lind EF, Brenner D, Brüstle A, Harris IS, Holmes R, Wakeham A, Haight J, et al.: **IDH1(R132H) mutation increases murine haematopoietic progenitors and alters epigenetics**. *Nature* 2012, **488**:656–659.

27. Lapinski PE, Meyer MF, Feng G-S, Kamiya N, King PD: **Deletion of SHP-2 in mesenchymal stem cells causes growth retardation, limb and chest deformity and calvarial defects in mice**. *Dis Model Mech* 2013, doi:10.1242/dmm.012849.

28. Matsushita T, Chan YY, Kawanami A, Balmes G, Landreth GE, Murakami S: **Extracellular Signal-Regulated Kinase 1 (ERK1) and ERK2 Play Essential Roles in Osteoblast Differentiation and in Supporting Osteoclastogenesis**. *Molecular and Cellular Biology* 2009, **29**:5843–5857.

29. Lemmon MA, Schlessinger J: **Cell signaling by receptor tyrosine kinases**. *Cell* 2010, **141**:1117–1134.

30. Wang K, Yamamoto H, Chin JR, Werb Z, Vu TH: **Epidermal growth factor receptor-deficient mice have delayed primary endochondral ossification because of defective osteoclast recruitment**. *J. Biol. Chem.* 2004, **279**:53848–53856.

31. Jacob AL, Smith C, Partanen J, Ornitz DM: **Fibroblast growth factor receptor 1 signaling in the osteo-chondrogenic cell lineage regulates sequential steps of osteoblast maturation**. *Dev. Biol.* 2006, **296**:315–328.

32. Minina E, Wenzel HM, Kreschel C, Karp S, Gaffield W, McMahon AP, Vortkamp A: **BMP and Ihh/PTHrP signaling interact to coordinate chondrocyte proliferation and differentiation**. *Development* 2001, **128**:4523–4534.

33. Colvin JS, Bohne BA, Harding GW, McEwen DG, Ornitz DM: **Skeletal overgrowth and deafness in mice lacking fibroblast growth factor receptor 3**. *Nat. Genet.* 1996, **12**:390–397.

34. Deng C, Wynshaw-Boris A, Zhou F, Kuo A, Leder P: **Fibroblast growth factor receptor 3 is a negative regulator of bone growth**. *Cell* 1996, **84**:911–921.

35. Minina E, Kreschel C, Naski MC, Ornitz DM, Vortkamp A: **Interaction of FGF, Ihh/Pthlh, and BMP signaling integrates chondrocyte proliferation and hypertrophic differentiation**. *Dev. Cell* 2002, **3**:439–449.

36. Krejci P, Masri B, Salazar L, Farrington-Rock C, Prats H, Thompson LM, Wilcox WR: **Bisindolylmaleimide I suppresses fibroblast growth factor-mediated activation of Erk MAP kinase in chondrocytes by preventing Shp2 association with the Frs2 and Gab1 adaptor proteins**. *J. Biol. Chem.* 2007, **282**:2929–2936.

37. Liu T, Gao Y, Sakamoto K, Minamizato T, Furukawa K, Tsukazaki T, Shibata Y, Bessho K, Komori T, Yamaguchi A: **BMP-2 promotes differentiation of osteoblasts and chondroblasts inRunx2-deficient cell lines**. *Journal of Cellular Physiology* 2007, **211**:728–735.

38. Pan Q, Yu Y, Chen Q, Li C, Wu H, Wan Y, Ma J, Sun F: **Sox9, a key transcription factor of bone morphogenetic protein-2-induced chondrogenesis, is activated through BMP pathway and a CCAAT box in the proximal promoter**. *J. Cell. Physiol.* 2008, **217**:228–241.

39. Holleville N, Quilhac A, Bontoux M, Monsoro-Burq AH: **BMP signals regulate Dlx5 during early avian skull development**. *Dev. Biol.* 2003, **257**:177–189.

40. Hattori T, Muller C, Gebhard S, Bauer E, Pausch F, Schlund B, Bosl MR, Hess A, Surmann-Schmitt C, von der Mark H, et al.: **SOX9 is a major negative regulator of cartilage vascularization, bone marrow formation and endochondral ossification**. *Development* 2010, **137**:901–911.

41. Salie R, Kneissel M, Vukevic M, Zamurovic N, Kramer I, Evans G, Gerwin N, Mueller M, Kinzel B, Susa M: **Ubiquitous overexpression of Hey1 transcription factor leads to osteopenia and chondrocyte hypertrophy in bone**. *Bone* 2010, **46**:680–694.

42. Amano K, Ichida F, Sugita A, Hata K, Wada M, Takigawa Y, Nakanishi M, Kogo M, Nishimura R, Yoneda T: **Msx2 Stimulates Chondrocyte Maturation by Controlling Ihh Expression**. *Journal of Biological Chemistry* 2008, **283**:29513–29521.

43. Ferrari D, Kosher RA: **Dlx5 Is a Positive Regulator of Chondrocyte Differentiation during Endochondral Ossification**. *Developmental Biology* 2002, **252**:257–270.

44. Lai LP, Mitchell J: **Indian hedgehog: Its roles and regulation in endochondral bone development**. *Journal of Cellular Biochemistry* 2005, **96**:1163–1173.

45. Kobayashi T, Soegiarto DW, Yang Y, Lanske B, Schipani E, McMahon AP, Kronenberg HM: **Indian hedgehog stimulates periarticular chondrocyte differentiation to regulate growth plate length independently of PTHrP**. *J. Clin. Invest.* 2005, **115**:1734–1742.

46. Li Y, Dudley AT: **Noncanonical frizzled signaling regulates cell polarity of growth plate chondrocytes**. *Development* 2009, **136**:1083–1092.

47. Ahrens MJ, Li Y, Jiang H, Dudley AT: **Convergent extension movements in growth plate chondrocytes require gpi-anchored cell surface proteins**. *Development* 2009, **136**:3463–3474.

48. Porter DE, Simpson AH: **The neoplastic pathogenesis of solitary and multiple osteochondromas**. *J. Pathol.* 1999, **188**:119–125.

49. Huegel J, Mundy C, Sgariglia F, Nygren P, Billings PC, Yamaguchi Y, Koyama E, Pacifici M: **Perichondrium phenotype and border function are regulated by Ext1 and heparan sulfate in developing long bones: A mechanism likely deranged in Hereditary Multiple Exostoses**. *Dev. Biol.* 2013, doi:10.1016/j.ydbio.2013.02.008.

50. Jones KB, Piombo V, Searby C, Kurriger G, Yang B, Grabellus F, Roughley PJ, Morcuende JA, Buckwalter JA, Capecchi MR, et al.: **A mouse model of osteochondromagenesis from clonal inactivation of Ext1 in chondrocytes**. *Proc. Natl. Acad. Sci. U.S.A.* 2010, **107**:2054–2059.

51. Zak BM, Schuksz M, Koyama E, Mundy C, Wells DE, Yamaguchi Y, Pacifici M, Esko JD: **Compound heterozygous loss of Ext1 and Ext2 is sufficient for formation of multiple exostoses in mouse ribs and long bones**. *Bone* 2011, **48**:979–987.

52. Sgariglia F, Candela ME, Huegel J, Jacenko O, Koyama E, Yamaguchi Y, Pacifici M, Enomoto-Iwamoto M: **Epiphyseal abnormalities, trabecular bone loss and articular chondrocyte hypertrophy develop in the long bones of postnatal Ext1-deficient mice**. *Bone* 2013, **57**:220–231.

53. Day T, Guo X, Garrettbeal L, Yang Y: **Wnt/β-Catenin Signaling in Mesenchymal Progenitors Controls Osteoblast and Chondrocyte Differentiation during Vertebrate Skeletogenesis**. *Developmental Cell* 2005, **8**:739–750.

54. Hill T, Spater D, Taketo M, Birchmeier W, Hartmann C: **Canonical Wnt/β-Catenin Signaling Prevents Osteoblasts from Differentiating into Chondrocytes**. *Developmental Cell* 2005, **8**:727–738.

55. Cantley L, Saunders C, Guttenberg M, Candela ME, Ohta Y, Yasuhara R, Kondo N, Sgariglia F, Asai S, Zhang X, et al.: **Loss of β-catenin induces multifocal periosteal chondroma-like masses in mice**. *Am. J. Pathol.* 2013, **182**:917–927.

56. Hojo H, Ohba S, Taniguchi K, Shirai M, Yano F, Saito T, Ikeda T, Nakajima K, Komiyama Y, Nakagata N, et al.: **Hedgehog-Gli activators direct osteo-chondrogenic function of bone morphogenetic protein toward osteogenesis in the perichondrium**. *J. Biol. Chem.* 2013, **288**:9924–9932.

57. Long F, Chung U, Ohba S, McMahon J, Kronenberg HM, McMahon AP: **Ihh signaling is directly required for the osteoblast lineage in the endochondral skeleton**. *Development* 2004, **131**:1309–1318.

58. Koyama E, Young B, Nagayama M, Shibukawa Y, Enomoto-Iwamoto M, Iwamoto M, Maeda Y, Lanske B, Song B, Serra R, et al.: **Conditional Kif3a ablation causes abnormal hedgehog signaling topography, growth plate dysfunction, and excessive bone and cartilage formation during mouse skeletogenesis**. *Development* 2007, **134**:2159–2169.

59. Pansuriya TC, Kroon HM, Bovée JVMG: **Enchondromatosis: insights on the different subtypes**. *Int J Clin Exp Pathol* 2010, **3**:557–569.

60. Kimura H, Ng JMY, Curran T: **Transient inhibition of the Hedgehog pathway in young mice causes permanent defects in bone structure**. *Cancer Cell* 2008, **13**:249–260.

61. Lin TL, Matsui W: **Hedgehog pathway as a drug target: Smoothened inhibitors in development**. *Onco Targets Ther* 2012, **5**:47–58.

62. Leshchiner I, Alexa K, Kelsey P, Adzhubei I, Austin-Tse CA, Cooney JD, Anderson H, King MJ, Stottmann RW, Garnaas MK, et al.: **Mutation mapping and identification by whole-genome sequencing**. *Genome Res.* 2012, **22**:1541–1548.

63. Obholzer N, Swinburne IA, Schwab E, Nechiporuk AV, Nicolson T, Megason SG: **Rapid positional cloning of zebrafish mutations by linkage and homozygosity mapping using whole-genome sequencing**. *Development* 2012, **139**:4280–4290.

64. Voz ML, Coppieters W, Manfroid I, Baudhuin A, Von Berg V, Charlier C, Meyer D, Driever W, Martial JA, Peers B: **Fast homozygosity mapping and identification of a zebrafish ENU-induced mutation by whole-genome sequencing**. *PLoS ONE* 2012, **7**:e34671.

65. Hill JT, Demarest BL, Bisgrove BW, Gorsi B, Su Y-C, Yost HJ: **MMAPPR: mutation mapping analysis pipeline for pooled RNA-seq**. *Genome Res.* 2013, **23**:687–697.

66. Miller AC, Obholzer ND, Shah AN, Megason SG, Moens CB: **RNA-seq-based mapping and candidate identification of mutations from forward genetic screens**. *Genome Res.* 2013, **23**:679–686.

67. Ryan S, Willer J, Marjoram L, Bagwell J, Mankiewicz J, Leshchiner I, Goessling W, Bagnat M, Katsanis N: **Rapid identification of kidney cyst mutations by whole exome sequencing in zebrafish**. *Development* 2013, **140**:4445–4451.

68. Arnold CN, Xia Y, Lin P, Ross C, Schwander M, Smart NG, Müller U, Beutler B: **Rapid identification of a disease allele in mouse through whole genome sequencing and bulk segregation analysis**. *Genetics* 2011, **187**:633–641.

69. Caruana G, Farlie PG, Hart AH, Bagheri-Fam S, Wallace MJ, Dobbie MS, Gordon CT, Miller KA, Whittle B, Abud HE, et al.: **Genome-wide ENU mutagenesis in combination with high density SNP analysis and exome sequencing provides rapid identification of novel mouse models of developmental disease**. *PLoS ONE* 2013, **8**:e55429.

70. Kurek KC, Luks VL, Ayturk UM, Alomari AI, Fishman SJ, Spencer SA, Mulliken JB, Bowen ME, Yamamoto GL, Kozakewich HPW, et al.: **Somatic Mosaic Activating Mutations in PIK3CA Cause CLOVES Syndrome**. *The American Journal of Human Genetics* 2012, **90**:1108–1115.

71. Adams A, McBratney-Owen B, Newby B, Bowen ME, Olsen BR, Warman ML: **Presphenoidal synchondrosis fusion in DBA/2J mice**. *Mamm. Genome* 2013, **24**:54–62.