Identification of molecular-genetic causes for osteogenesis imperfecta, interdigital hyperplasia and ribosomopathies in cattle

Dissertation

to obtain the Ph. D. degree
in the International Ph. D. Program for Agricultural Sciences in Göttingen (IPAG)
at the Faculty of Agricultural Sciences,
Georg-August-University Göttingen, Germany

presented by

Xuying Zhang

born in Shanxi, P.R.China

Göttingen, September, 2019
1. Name of supervisor: Prof. Dr. Dr. Bertram Brenig

2. Name of co-supervisor: Prof. Dr. Jens Tetens

Date of dissertation: 10. September 2019
To my family
Table of Contents

Table of Contents                              i
List of Figures                                ii
List of Tables                                 iii
List of Publications                           iv
Abstract                                       v
Abbreviations                                 vii

CHAPTER 1                                     General Introduction 1
  1 Osteogenesis imperfecta                    2
     1.1 Complexity and dynamic nature of bone tissue 2
     1.2 Phenotypic aspect of osteogenesis imperfecta 3
     1.3 Molecular dissection of osteogenesis imperfecta 5
  2 Interdigital hyperplasia                   13
     2.1 Overview of lameness in dairy cattle      13
     2.2 Research progress on interdigital hyperplasia 14
  3 Ribosomopathy                              18
     3.1 Ribosome biogenesis                      18
     3.2 Research progress on ribosomopathies     18

CHAPTER 2                                     Osteogenesis imperfecta in an embryo transfer Holstein calf 45

CHAPTER 3                                     Interdigital hyperplasia in Holstein Friesian cattle is associated with a missense mutation in the signal peptide region of the tyrosine-protein kinase transmembrane receptor gene 71

CHAPTER 4                                     Processed pseudogene confounding the presence of a putative lethal recessive deletion in the bovine 60S ribosomal protein L11 gene (uL5) 95

CHAPTER 5                                     General Discussion 103
  1 Significance of the research study          104
  2 Evolutionary genetic dissection technologies 104
     2.1 Genome-wide association study           104
     2.2 NGS-based analysis                      105
     2.3 Functional effect validation of novel variants 106
  3 Cattle as an animal model to study claw disorders 106

Conclusions and Outlook 108

Acknowledgments                               x
Curriculum Vitae                              xii
# List of Figures

## Chapter 1

<table>
<thead>
<tr>
<th>Fig.</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Molecular overview of osteogenesis imperfecta</td>
<td>6</td>
</tr>
</tbody>
</table>

## Chapter 2

<table>
<thead>
<tr>
<th>Fig.</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pedigree of the affected calf</td>
<td>53</td>
</tr>
<tr>
<td>2</td>
<td>Post mortom examination of left hind limb</td>
<td>54</td>
</tr>
<tr>
<td>3</td>
<td>Histological analysis of bone of the affected calf</td>
<td>55</td>
</tr>
<tr>
<td>4</td>
<td>Manhattan plot of genome-wide association analysis for osteogenesis imperfecta</td>
<td>56</td>
</tr>
</tbody>
</table>

## Chapter 3

<table>
<thead>
<tr>
<th>Fig.</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Clinic signs of bovine interdigital hyperplasia</td>
<td>78</td>
</tr>
<tr>
<td>2</td>
<td>Manhattan plot of genome-wide association study for bovine interdigital hyperplasia</td>
<td>79</td>
</tr>
<tr>
<td>3</td>
<td>Detection of <em>ROR2</em> transcripts and isoforms in different bovine organs and tissues</td>
<td>83</td>
</tr>
<tr>
<td>4</td>
<td>Comparison of <em>ROR2</em> expression in hyperplastic interdigital skin tissue</td>
<td>84</td>
</tr>
<tr>
<td>5</td>
<td>Quantification of ROR2 protein in hyperplastic interdigital skin tissue</td>
<td>85</td>
</tr>
</tbody>
</table>

## Chapter 4

<table>
<thead>
<tr>
<th>Fig.</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Alignment and DNA sequencing of a putative 2 bp deletion in <em>uL5</em> and comparison with a processed <em>uL5</em> pseudogene (LOC112442327)</td>
<td>100</td>
</tr>
</tbody>
</table>
List of Tables

Chapter 1
Table 1  Summary of reported ribosomopathies  20

Chapter 2
Table 1  Detection of DNA variants in osteogenesis imperfecta candidate genes  53
Table 2  Iterative determination of genotype frequency of OI causative variants in 2,612 random samples of Holstein cattle  58
Table S1  Primers for genotyping of functional variants  65
Table S2  NGS variant analysis  69

Chapter 3
Table 1  Results of clinical inspections during two farm visits  75
Table 2  Genotype frequencies of ROR2 variants rs377953295 (exon 1) and rs43572154 (exon 9) in type A affected and IH free (= healthy) cattle  80
Table 3  Statistical evaluation of SNP rs377953295 (exon 1) as causative variant for type A and type B interdigital hyperplasia  81
Table S1  ROR2 primers for PCR and Sanger sequencing  94
List of Publications

This thesis is based on the first three papers.


Abstract


Musculoskeletal anomalies are a heterogeneous group of fairly common, often with severe clinical symptoms, but mostly fatal disorders in human and animals. By far, complex vertebral malformation (CVM) and Brachyspina are the most important lethal osteodysplasias described in Holstein Friesian (HF) cattle. First aim of this thesis was to elucidate the genetic cause of a lethal case of osteogenesis imperfecta (OI) in a newborn HF calf. The molecular mechanism underlying OI in humans has been clearly explored, with a series of candidate genes reported. However, none of these hitherto known genes harboured any causative variant in the affected OI calf. To identify genomic regions associated with OI, a genome wide association study was performed. A total of 6 significantly associated SNP loci were detected on 1, 5, 6, 17, 18, and 24 (p < 0.05). In addition, a whole genome re-sequencing of the affected calf and its parents as well as gene prioritization using CANDID was performed. From these analyses 16 candidate genes with functional variants remained. Validation was performed in 2612 randomly chosen healthy German HF using high-resolution melting curve analysis. Finally, a haplotype of 4 functional variants in 3 candidate genes (ABCA13: ENSBTAT00000061018.2:c.12553+1A>G, p.Gln4393Arg; QRFP: p.Arg412Ter; IFITM5: p.Ala30Ser) was determined explaining the development of OI in the calf. This was in agreement with the results of the GWAS indicating a so far unknown oligogenic origin of OI.

Bovine interdigital hyperplasia (IH) is a typical claw disorder affecting cattle welfare and causing serious economic losses. IH shows the highest estimated heritabilities in comparison to other feet/claw disorders. Second aim of this thesis was to elucidate the molecular genetic cause of IH in HF. To investigate the pathogenesis, the IH status was inspected during hoof trimming of first-lactation dairy cows. To study the clinical course, the cows were examined seven times every three months. From herds with highest IH prevalence, cows were selected to perform a genome-wide association study. Two significantly associated chromosomal positions were revealed with both located in the tyrosine-protein kinase transmembrane receptor ROR2 gene. Due to its previously reported function, ROR2 was a reasonable candidate for IH. A comparative sequencing between cases and controls was conducted and 2 missense mutations were identified. Large-scale genotyping showed that one variant ROR2 p.Trp9Arg was highly significantly associated with IH (p < 0.0001). Quantification of mRNA and protein levels proved significant reduction of ROR2 expression due to this variant.

Ribosome biogenesis is the process of generating ribosomes which serve as the site of biological protein synthesis in all living cells. Impaired or decreased ribosome biosynthesis is associated with a group of diseases called ribosomopathies. Due to the lack of functional ribosomes, these diseases broadly fall under the category of cellular hypo-proliferation phenotypes. Before supportive care was developed, such hypo-proliferation phenotypes were severe even lethal in the past. Molecular dissection of ribosomopathies in humans has revealed
a list of causal genes, however, ribosomopathies have not been reported in cattle so far. Third aim of this thesis was to address the question whether lethal variants in the causal genes of ribosomopathies exist in cattle. A 2bp deletion has been deposited into bovine genomic databases in bovine $uL5$ gene on chromosome 2, which would result in a frameshift and a premature stop codon. The deletion causes a truncation of bovine $uL5$. A probe for this variant had been included as expert-selected marker in the custom add-on part of the Illumina BovineLD BeadChip and can therefore be monitored during routine genotyping. To determine the frequency of the variant, we genotyped 370,527 cattle, including 18 different dairy and beef cattle breeds. 299,218 homozygous wild type and 71,249 heterozygous cattle were called from the bead chip. Cattle harboring the homozygous mutant genotype were not detected. According to Hardy-Weinberg equilibrium around 4,241 homozygous individuals carrying the deletion should have been present in the cohort. Hence, the data apparently indicated that the homozygous 2 bp deletion in $uL5$ was a lethal variant interfering with ribosome biogenesis and resulting in embryonic death. However, using Sanger sequencing we could not detect the putative deleterious variant in randomly selected heterozygous cattle identified by the BeadChip genotyping. An alignment of the BeadChip probe to the bovine genome showed a perfect match to a processed $uL5$ pseudogene on bovine chromosome 18 mimicking the assumed exonic deletion. Subsequently, an examination by Sanger sequencing showed that the previously detected heterozygous cattle were homozygous for the 2 bp deletion within the pseudogene. The BeadChip genotyping results were clearly confounded by the erroneous detection of this variant. Therefore, special care should be taken when designing probes and/or primers for a specific assay in high-throughput genotyping platforms to avoid severe misinterpretations led by pseudogenes. We not yet identify a lethal variant, however, our strategy can be used to farther explore potentially recessive lethal variants causing ribosomopathies in cattle.
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACSL5</td>
<td>Acyl-CoA synthetase long chain family member 5</td>
</tr>
<tr>
<td>ARMS</td>
<td>Amplification refractory mutation system</td>
</tr>
<tr>
<td>ATF6</td>
<td>Activating transcription factor 6</td>
</tr>
<tr>
<td>BLUP</td>
<td>Best linear unbiased predictions</td>
</tr>
<tr>
<td>BMP</td>
<td>Bone morphogenetic protein</td>
</tr>
<tr>
<td>BMP1</td>
<td>Bone morphogenetic protein 1</td>
</tr>
<tr>
<td>BMU</td>
<td>Basic multicellular unit</td>
</tr>
<tr>
<td>CNTN1</td>
<td>Contactin 1</td>
</tr>
<tr>
<td>COL1A1</td>
<td>Collagen type I alpha 1 chain</td>
</tr>
<tr>
<td>COL1A2</td>
<td>Collagen type I alpha 2 chain</td>
</tr>
<tr>
<td>Creb3l1</td>
<td>CAMP Responsive Element Binding Protein 3 Like 1</td>
</tr>
<tr>
<td>CRTAP</td>
<td>Cartilage associated protein</td>
</tr>
<tr>
<td>CXCR4</td>
<td>C-X-C motif chemokine receptor 4</td>
</tr>
<tr>
<td>CyPB</td>
<td>Cyclophilin B</td>
</tr>
<tr>
<td>DKC1</td>
<td>Dyskerin pseudouridine synthase 1</td>
</tr>
<tr>
<td>DUSP2</td>
<td>Dual specificity phosphatase 2</td>
</tr>
<tr>
<td>EBV</td>
<td>Estimated breeding value</td>
</tr>
<tr>
<td>ECM</td>
<td>Extracellular matrix</td>
</tr>
<tr>
<td>EMG1</td>
<td>EMG1 N1-specific pseudouridine methyltransferase</td>
</tr>
<tr>
<td>ESCO2</td>
<td>Establishment of sister chromatid cohesion N-acetyltransferase 2</td>
</tr>
<tr>
<td>FABP3</td>
<td>Fatty acid binding protein 3</td>
</tr>
<tr>
<td>FGF</td>
<td>Fibroblast growth factor</td>
</tr>
<tr>
<td>FKBP10</td>
<td>FKBP prolyl isomerase 10</td>
</tr>
<tr>
<td>FKBP65</td>
<td>65-kDa FK506-binding protein</td>
</tr>
<tr>
<td>FRET</td>
<td>Fluorescence resonance energy transfer</td>
</tr>
<tr>
<td>Fz-LRP</td>
<td>Frizzled-low density lipoprotein receptor-related protein</td>
</tr>
<tr>
<td>GATA1</td>
<td>GATA binding protein 1</td>
</tr>
<tr>
<td>GWAS</td>
<td>Genome-wide association study</td>
</tr>
<tr>
<td>HDAd</td>
<td>Helper-dependent adenoviral vector</td>
</tr>
<tr>
<td>HES1</td>
<td>Hes family bHLH transcription factor 1</td>
</tr>
<tr>
<td>HF</td>
<td>Holstein Friesian</td>
</tr>
<tr>
<td>HSP47</td>
<td>Heat shock protein 47</td>
</tr>
<tr>
<td>IARS</td>
<td>Isoleucyl-tRNA synthetase</td>
</tr>
<tr>
<td>IFITM5</td>
<td>Interferon induced transmembrane protein 5</td>
</tr>
<tr>
<td>IGF</td>
<td>Insulin-like growth factors</td>
</tr>
<tr>
<td>IH</td>
<td>Interdigital hyperplasia</td>
</tr>
<tr>
<td>InDel</td>
<td>Insertion or deletion</td>
</tr>
<tr>
<td>KO</td>
<td>Knockout</td>
</tr>
<tr>
<td>LEPRE1</td>
<td>leucine and proline-enriched proteoglycan 1</td>
</tr>
<tr>
<td>LH1</td>
<td>Lysyl hydroxylase 1</td>
</tr>
<tr>
<td>LH2</td>
<td>Lysyl hydroxylase 2</td>
</tr>
<tr>
<td>LH3</td>
<td>Lysyl hydroxylase 3</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>LncRNA</td>
<td>Long non-coding RNA</td>
</tr>
<tr>
<td>LRP5</td>
<td>Lipoprotein related protein 5</td>
</tr>
<tr>
<td>LSU</td>
<td>Large subunit</td>
</tr>
<tr>
<td>MAS</td>
<td>Marker-assisted selection</td>
</tr>
<tr>
<td>MATN1</td>
<td>Matrilin 1</td>
</tr>
<tr>
<td>MBTPS1</td>
<td>Membrane bound transcription factor peptidase, Site 1</td>
</tr>
<tr>
<td>MBTPS2</td>
<td>Membrane bound transcription factor peptidase, Site 2</td>
</tr>
<tr>
<td>MiR</td>
<td>MicroRNA</td>
</tr>
<tr>
<td>mTLD</td>
<td>Mammalian tolloid protein</td>
</tr>
<tr>
<td>mTLL1</td>
<td>Mammalian tolloid-like protein 1</td>
</tr>
<tr>
<td>mTLL2</td>
<td>Mammalian tolloid-like protein 2</td>
</tr>
<tr>
<td>NGS</td>
<td>Next-generation sequencing</td>
</tr>
<tr>
<td>NHP2</td>
<td>NHP2 ribonucleoprotein</td>
</tr>
<tr>
<td>NODAL</td>
<td>Nodal growth differentiation factor</td>
</tr>
<tr>
<td>NOP10</td>
<td>NOP10 ribonucleoprotein</td>
</tr>
<tr>
<td>nt</td>
<td>Nucleotide</td>
</tr>
<tr>
<td>OASIS</td>
<td>Old astrocyte specifically induced substance</td>
</tr>
<tr>
<td>OGN</td>
<td>Osteoglycin</td>
</tr>
<tr>
<td>OI</td>
<td>Osteogenesis imperfecta</td>
</tr>
<tr>
<td>OMD</td>
<td>Osteomodulin</td>
</tr>
<tr>
<td>Osx</td>
<td>Osterix</td>
</tr>
<tr>
<td>P3H1</td>
<td>Prolyl 3-hydroxylase 1</td>
</tr>
<tr>
<td>PCBD1</td>
<td>Pterin-4 alpha-carbinolamine dehydratase 1</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PDCD4</td>
<td>Programmed cell death 4</td>
</tr>
<tr>
<td>PDGF</td>
<td>Platelet-derived growth factors</td>
</tr>
<tr>
<td>PEDF</td>
<td>Pigment epithelium-derived factor</td>
</tr>
<tr>
<td>PLOD1</td>
<td>Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 1</td>
</tr>
<tr>
<td>PLOD2</td>
<td>Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2</td>
</tr>
<tr>
<td>PLOD3</td>
<td>Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 3</td>
</tr>
<tr>
<td>PPIB</td>
<td>Peptidylprolyl isomerase B</td>
</tr>
<tr>
<td>RB1CC1</td>
<td>RB1 inducible coiled-coil 1</td>
</tr>
<tr>
<td>rDNA</td>
<td>Ribosomal DNA</td>
</tr>
<tr>
<td>RFLP</td>
<td>Restriction fragment length polymorphism</td>
</tr>
<tr>
<td>RMRP</td>
<td>RNA component of mitochondrial RNA processing endoribonuclease</td>
</tr>
<tr>
<td>ROR2</td>
<td>Receptor tyrosine kinase like orphan receptor 2</td>
</tr>
<tr>
<td>RP</td>
<td>Ribosomal protein</td>
</tr>
<tr>
<td>RPL-</td>
<td>Ribosomal protein L-</td>
</tr>
<tr>
<td>r-protein</td>
<td>Ribosomal protein</td>
</tr>
<tr>
<td>RPS-</td>
<td>Ribosomal Protein S-</td>
</tr>
<tr>
<td>rRNA</td>
<td>Ribosomal RNA</td>
</tr>
<tr>
<td>RTK</td>
<td>Receptor tyrosine kinase</td>
</tr>
<tr>
<td>RUNX2</td>
<td>Runt-related transcription factor 2</td>
</tr>
<tr>
<td>SBDS</td>
<td>SBDS ribosome maturation factor</td>
</tr>
</tbody>
</table>
SDC3  Syndecan 3
SFRPs  Secreted frizzled related proteins
SERPINF1  Serpin family F member 1
SERPINH1  Serpin family H member 1
SGS  Second-generation sequencing
SHOC2  SHOC2, leucine rich repeat scaffold protein
SIAH1  Siah E3 ubiquitin protein ligase 1
snoRNA  Small nucleolar RNA
snoRNP  Small nucleolar ribonucleoprotein complexes
SNP  Single nucleotide polymorphism
SP7  Sp7 transcription factor
SREBP  Sterol regulatory element binding protein
SSU  Small subunit
STAT1  Signal transducer and activator of transcription 1
STAT3  Signal transducer and activator of transcription 3
SV  Structural variant
TCAB1  WD repeat containing antisense to TP53
TCF7L2  Transcription factor 7 like 2
TCOF1  Treacle ribosome biogenesis factor 1
TECTB  Tectorin beta
TERC  Telomerase RNA component
TERT  Telomerase reverse transcriptase
TGF-β  Transforming growth factor-beta
TINF2  TERF1 interacting nuclear factor 2
TMEM38B  Transmembrane protein 38B
TNFα  Tumor necrosis factor alpha
Tric-A  Trimeric intracellular cation channel subtype A
Tric-B  Trimeric intracellular cation channel subtype B
TSR2  TSR2 ribosome maturation factor
USP15  Ubiquitin specific peptidase 15
UTP4  UTP4 small subunit processome component
UTR  Untranslated region
VTI1A  Vesicle transport through interaction with t-SNAREs 1A
WGS  Whole genome sequencing
WIF1  Wnt inhibitory factor 1
WNT1  Wnt family member 1
Wnt5A  Wnt family member 5A
ZDHHC6  Zinc finger DHHC-type containing 6
CHAPTER 1

General Introduction
1 Osteogenesis imperfecta

1.1 Complexity and dynamic nature of bone tissue

Bone is a mineralized tissue, exerting important functions in the body, such as structural support for the body, protection of vital organs, harboring of bone marrow, and minerals storage. Bone comprises two main tissue types on the macroscopic level: cortical bone and trabecular bone [1]. Cortical bone tissue also called compact bone, is a dense, hard and rigid outer layer of bone. In contrast, the inner layer is light and spongy-like, called cancellous or trabecular bone. There are four types of cells in bones, i.e. osteoblasts, osteoclasts, bone lining cells and osteocytes, responsible for bone formation, maintenance, modeling as well as remodelling [1].

1.1.1 Osteogenesis and bone growth

Osteogenesis is the process generating new bone substance. Bone is formed mainly in two ways [1]. Intramembranous ossification involves the replacement of mesenchymal tissue with bone in embryogenic processes, resulting in the formation of skull and some facial bones [2]. The formation of the majority of other bone tissue, such as femur, humerus and many other kinds of long bones originates from endochondral ossification. It involves the initial generation of a hyaline cartilage intermediate, which is afterwards converted into bone by bone forming cells - osteoblasts [2].

The majority of limb bones feature long bone characteristics [3]. Each long bone comprises two extremities and one shaft, with length wider than width. Long bone contains compact bone substance in outside layer, medullary cavity insides, as well as spongy bone and an epiphyseal line at ends. During the whole childhood and adolescence, long bone keeps growing in width and length. Increase in length relies on hyaline cartilage in epiphyseal line, which continues forming bone at the ends of long bones in endochondral ossification model [4]. In contrast, accumulation in width of long bones is attained by forming new bone substance on the outside of the compact bone type.

1.1.2 Bone remodeling

Once bone formation is completed in adulthood, the only mode to replace old bone substance with new tissue is via bone remodeling. This complex mode remains active until death to maintain mineral homeostasis and bone strength [3]. This also enables self-heal processing of bone microdamages and skeleton recovery from mechanical use [5]. One bone remodeling cycle consists of three phases: (i) initial removal of mineral matrix of old bone by activated osteoclasts, (ii) reversal process from resorption to new-synthesized proteinaceous matrix, (iii) matrix mineralization to generate new bone substance by osteoblasts [6]. This cycle relies on close cooperation of osteoclasts, osteoblasts, as well as bone lining cells and osteocytes, which are organized in a temporary anatomical body, known as basic multicellular unit (BMU) [7]. Bone mass and integrity depend on maintenance of a delicate balance between bone resorption by osteoclasts and bone formation by osteoblasts [8]. Osteocytes act as orchestrators in bone remodeling process through regulation of both osteoclast and osteoblast activity [9]. Bone lining cells play an essential role in coupling bone resorption to bone formation by removal of bone collagen left by osteoclasts in Howship's lacunae [10].
An imbalance of bone resorption and formation leads to various bone diseases. For instance, defects in osteoclast formation and function cause increased bone mass and osteopetrosis [11]. Whereas excessive bone resorption by osteoclasts contributes to osteoporosis and corticosteroid-induced bone loss [12-14]. In addition, emerging evidences in human patients reveal the relevance of abnormal genes/proteins important for osteoblast differentiation and mineralization with osteogenesis imperfecta [15].

1.2 Phenotypic aspect of osteogenesis imperfecta

1.2.1 Clinical introduction

Osteogenesis imperfecta (OI), also called “brittle bone disease”, is a genetically heterogeneous skeletal dysplasia, affecting about 1 in 10,000 to 20,000 live births in humans [16, 17]. OI patients are characterized by a prominent skeletal phenotype, with widely varying clinical severity. Also, joint laxity, scoliosis, dentinogenesis imperfecta and craniofacial abnormalities may be present in patients with OI; other extra-skeletal manifestations include hearing impairment, blue or gray sclerae, lung abnormalities, and hypercalciuria for instance [17-20].

The earliest OI case reported was dated to circa 1000 BC. After reconstruction of the skull of the Egyptian mummy, paleopathologists were convinced of an OI affected infant, because of its deformity of vertical axis and transverse axis bones as well as deformed dentition [21]. The first scientific description about OI was supplied by Olaus Jakob Ekman in 1788, in the ‘congenital osteomalacia’ thesis [22]. The hereditary fragile bone disease occurring in three generations of a family was introduced to medical expert knowledge [22]. Afterwards, various terms were used to characterize familial skeletal fragility. Until 1849, the phrase ‘osteogenesis imperfecta’ (in Dutch: gebrekkige beenwording) was first-time introduced by Willem Vrolik, a professor from Amsterdam University [23]. By stating that the primary impairment of ossification already existed in the specimen, he realized that various skeletal dysplasias may be caused by insufficient intrinsic ‘generative energy’, but not the result of a postnatally acquired disorder [23].

In 1974, by scanning electron microscopy of bone collagen in 3 congenital OI patients, abnormalities of bone collagen aggregation from thin collagen fibers into extracellularly large collagen fiber bundles of bone were described [24]. The organizational alterations may partially explain the changes in physical properties leading to the frequent fractures featuring this disease [24]. Using pepsin digests of skin of OI patients an increased ratio of the alpha 1 (III) to alpha 1 (I) chains from two kinds of collagen type III and type I was identified [25]. This increased ratio may be due to a reduction of type I collagen [25]. These findings set the basis to clarify molecular and pathogenic mechanisms of OI.

1.2.2 Classification

Increasing OI cases have clarified in detail that OI represents a bone disease with clinical severity remarkably varying from perinatal lethality to slightly increased fracture frequency. In 1906, the first OI classification was proposed by Looser [26]. After studying OI in Victoria,
Australia, ‘Silence classification’ was introduced in 1979 to further classify this disease and became the basis of current classification and severity verification of OI [19]. 180 OI cases were verified and studied in epidemiological and genetic aspects. According to their clinical phenotypes and inheritance modes the OI cohort was classified into at least four different syndromes. The first group, with autosomal dominant inheritance, showed osteoporosis and consequently fractures as well as blue sclerae [19]. The second group, with perinatal lethality, showed beaded ribs as well as broad and crumpled femora. Some cases in this group showed autosomal recessive inheritance [19]. The third group, comprised sporadic OI cases, 2/3 of them showed fractures at birth, and had severe progressive deformity of spine and limbs. The inheritance pattern in this group could not be clarified. But it was stated that this group could be heterogeneous with both recessive and dominant inheritance [19]. In the fourth group, patients were characterized by dominantly inherited osteoporosis resulting in fractures, with variable deformities in long bones but normal sclerae [19]. In 1984, Sillence et al. subdivided the second group (OI type II) into three groups (A, B, C) based on the differences of the clinical and radiological findings [27]. In the cases of group A, ribs were short and thick with continuous beading, femora appeared broad and crumpled, and tibiae were angulated [27]. In group B, long bones and skull were similar in appearance to those of group A patients, but ribs were normal or showed incomplete beading [27]. In group C, cases had slender beaded ribs. The shafts of the long bones showed acute angulation deformities with multiple fractures [27].

In 2004, Rauch and Glorieux expanded the original Sillence classification to 7 distinct types [18]. Until then, mutations in COL1A1 and COL1A2 genes, which encode α1 and α2 chains of the collagen type I respectively, were found in all the four types of OI [18]. OI types V–VII were added without defining specific genetic reasons [18]. However, OI type V was presumed to be inherited in an autosomal dominant pattern, while type VI and VII were supposed to be autosomal recessively inherited, with significantly different typical features [18].

In 2006, the first genetic reason of recessive OI type was reported, i.e. CRTAP mutations [28]. Since then, more and more genetic causes of OI have been reported, although COL1A1 or COL1A2 mutations were still responsible for the majority of OI cases. There was some debate in literatures on how to incorporate this newly identified OI heterogeneity into the classification system. Some researchers proposed to designate the OI caused by recessive mutations in PPIB as OI type IX [29], SERPINH1 missense variants as OI type X, and FKBP10 frameshift variants as OI type XI [30]. However, another viewpoint was proposed by van Dijk and his colleagues, since the newly defined OI types base on different causative gene mutations resulted in a confusing OI classification [31]. They proposed to exclude type VII and VIII [31]. These types were added due to genetic criteria, but their clinical and radiological features were not distinguishable from types II-IV [31]. Besides, Bruck syndrome manifested clinically with a combination of classical osteogenesis imperfecta and multiple congenital joint contractures, which aroused the debate about whether this syndrome should be classified as a subtype of OI [32].

Finally in 2011, the International Nomenclature group for Constitutional disorders of the Skeleton (INCDS) published a decision in the 25th group and classified osteogenesis imperfecta
into 5 types [33]. It retained the original 4 types from the Sillence classification and amended by the addition of the OI type V, defined as osteogenesis imperfecta with calcification of the interosseous membranes and/or hypertrophic callus [33]. The importance of the various genetic reasons of the OI types was acknowledged through encapsulating them separately as subtypes of the OI type I to V [33]. Bruck syndrome was not classified into any of OI types I-V [33].

Forty years ago, Sillence proposed the first systematic classification based on clinical features. Today it is clear that OI is a disorder with a greater heterogeneity than previously thought. It is necessary to consider correlations between the proposed clinical types and the genetic/molecular findings.

1.3 Molecular dissection of osteogenesis imperfecta

1.3.1 Genetic heterogeneity and mechanisms in humans

Collagen synthesis

Almost 90% of OI cases are caused by dominant mutations in either COLIA1 or COLIA2, encoding the α1 and α2 chains of type I collagen, respectively [18, 34]. Collagen is a fibrous protein that is part of connective tissues. Collagen I is a heterotrimer composed of two α1 and one α2 polypeptide chains, which form a triple helix. Each chain contains hundreds of uninterrupted repeats of the Gly-X-Y triplet, where X and Y can be any amino acid but are often proline and lysine. The latter residues are commonly hydroxylated by specific proline or lysine hydroxylases, which protects the triple-helical structure from melting at normal body temperatures [35]. The conserved glycine occurs in every third position, since it is the only amino acid residue that fits sterically in the core of the triple helix without disrupting the structure.

There are mainly two types of collagen I variants resulting in OI, one type resulting in a normal structure but an altered quantity of type I procollagen, another type leading to a structurally aberrant protein [34]. Quantitative defects are usually the result of nonsense and frameshift variants throughout most part of the COLIA1/COLIA2 gene causing premature termination of translation and hence a significant reduction in the amount of transcript from the mutant allele [36]. Those result in haploinsufficiency of type I collagen and are commonly seen in mild or moderate forms of OI [36, 37]. Qualitative defects are commonly caused by variants impairing triple helix formation, e.g. splice site variants causing in-frame exon skipping resulting in shorter alpha chains [34]. Qualitatively altered collagens are commonly seen in more severe or lethal forms of OI [34, 38].

The functional importance of type I collagen gene COLIA1 and COLIA2 for osteogenesis have also been proved in mouse models. The COL1A1\textsuperscript{Mov13} mouse has been generated with a transcription blockage in the murine COL1A1 gene [39, 40]. Mice embryos carrying this variant in a homozygous manner (Col1a1\textsuperscript{Mov13/Mov13}) display degenerated development and prenatal lethality [39, 40]. The Aga2 (abnormal gait 2) mice have been created to carry a frameshift mutation in the C-terminal end of COL1A1 gene [41]. Aga2 heterozygous animals (Aga2\textsuperscript{+/-}) have obviously increased bone turnover and a disrupted collagen network [41]. Also, the mice display lowered bone mass, fractures as well as perinatal death due to cardiac and pulmonary
disfunctions [41, 42]. The BrtlIV mice carry a heterozygous substitution from glycine to cysteine at amino acid p349 of COL1A1 [43]. BrtlIV mice display a phenotypical variability in severity ranging from moderate forms to perinatal lethality [43]. They also exhibit the classical OI phenotype, i.e. deformity, fragility, osteoporosis as well as disrupted trabecular structure [43]. The Oim mice carry a G deletion at nucleotide 3983 of the COLIA2 gene, resulting in the sequence alteration of the last 48 amino acids [44]. The mutation results in accumulation of α(I) homotrimeric collagen in skin and bone tissues, and a deficiency of the pro-α2(I) collagen chain [44]. Phenotypes observed in the homozygous mice (Oim−/−) include susceptibility to skeletal fractures from a young age, limb deformities, generalized severe osteopenia and small body size [44]. Histologically, their bones show decreased enzymatic cross-links and increased nonenzymatic cross-links of collagen, smaller mineral crystals with more variable orientation in cortical bone, and altered lacunar and vascular porosity [45-47].

Fig. 1 Molecular overview of osteogenesis imperfecta
(Modified based on the figure by Dr. Namrata Chhabra [48])

Collagen processing

Bone morphogenetic protein 1, also called mammalian tolloid protein, is an astacin metalloprotease encoded by BMP1 gene [49, 50]. BMP1 was initially demonstrated as the type I procollagen C-proteinase [51]. It cleaves the COOH-propeptides from procollagens I–III to generate the major fibrous components in extracellular matrix [51]. Proteolytic processing of
N- and C-propeptides of procollagen is a crucial step in collagen fibril formation [52]. Especially, the removal of the C-propeptides triggers the self-assembly of collagen monomers into fibrils [52]. Type I procollagen C-propeptide cleavage site variants were identified in two OI patients with high bone mineralization density [53]. The report demonstrates that procollagen C-propeptide cleavage plays a crucial role in normal bone mineralization [53]. In humans, homozygous or heterozygous BMP1 variants have been identified as causes of OI with bone fragility and frequent fractures [15, 54-58]. Most of these BMP1 variants have been proved to be associated with impaired procollagen type I C-propeptide cleavage [55-57].

There are four secreted mammalian BMP1-like proteinases, i.e. BMP1, mTLD, mTLL1 and mTLL2, involved in processing of procollagen I C-terminal propeptide with varying cleavage efficiency. BMP1 has the highest protease activity in early developmental stages [59, 60]. Lysyl oxidase catalyzes the oxidative deamination of lysine and hydroxylysine residues in collagen and peptidyl-lysine in elastin precursors to form semialdehyde residues. Then the aldehydes undergo nonenzymatic reactions leading to the cross-linkages known to be critical in the generation of elastin and collagens in extracellular matrix biosynthesis [61]. BMP1/TLD-like proteases have also been reported to be responsible for the main processing resulting in the activation of lysyl oxidase [62].

**Prolyl 3-hydroxylase complex**

Proteins CRTAP, P3H1 and CypB assemble in a 1:1:1 ratio to form the endoplasmic reticulum-localized collagen prolyl 3-hydroxylation complex. The variants disrupting the expression of any protein in this complex have been reported to cause recessive OI [63-65]. Analysis on the OI cases has discovered the 3-hydroxylation function of the prolyl 3-hydroxylation complex at proline residues in the helical region of type I procollagen, Pro707 of chain α2(I) and Pro986 of chain α1(I) [66]. But the biological importance of this collagen 3-hydroxylation modification is not clear so far. In addition, the P3H1-CRTAP-CypB complex acts as a collagen chaperone, by interacting with triple helical collagen [67].

CRTAP acts as the helper protein of the complex and is mutually stabilized by P3H1 in the collagen prolyl 3-hydroxylation complex [68]. CRTAP is expressed in the skeleton by growth-plate proliferating chondrocytes, cells joining cartilage and bone, osteoblasts and -clasts, as well as other cells to varying degrees [69]. In addition to the function in the ER complex, this protein can be secreted from cells into matrix and exerts multiple functions [68, 70]. The Crtap-null mice showed multiple abnormalities of connective tissue, decreased mechanical integrity of skin as well as renal and lung defects [71]. Emanating from these functional insights in murine models, CRTAP deficiency in humans is associated with recessive OI ranging from type II (neonatal lethal cases) to type VII (mild phenotype) [63]. Most of the reported CRTAP variants resulted in frameshifts or alternative splicing leading to nonsense-mediated decay and absence of CRTAP protein [72]. Lack or reduction of α1(I) Proline-986 hydroxylation near the carboxyl end of collagen have been revealed in lethal or severe OI patients with recessive CRTAP deficiency [28]. They do not have a primary collagen defect but have excess post-translational modification of collagen [28]. This suggests that prolyl 3-hydroxylation of type I collagen is of importance for bone formation [28].
P3H1, is encoded by the LEPRE1 gene and provides the enzymatic component of the complex. It is the only component containing a Lys-Asp-Glu-Leu (KDEL) ER-retrieval sequence at the carboxyl terminus [73]. P3H1 was first isolated as a novel matrix chondroitin sulfate proteoglycan [74]. Tissue specific prolyl 3-hydroxylase gene expression in embryonic and adult tissues indicated its developmental role [75]. The most common LEPRE1 variant for recessive OI is a West African founder mutation introduced through the Atlantic slave trade [76]. Most of the reported LEPRE1 variants led to premature termination codon causing RNA decay. A novel variant resulting in a defect in KDEL caused a non-lethal OI demonstrating the KDEL ER-retrieval sequence is of importance for P3H1 functionality in vivo [77].

CypB, the third protein of the collagen 3-hydroxylation complex, is encoded by the PPIB gene. Peptidyl prolyl cis-trans-isomerization is the rate limiting step in the folding of the collagen triple helix, and CypB was thought to be the unique collagen peptidylprolyl isomerase for a long time [78, 79]. CypB levels are independent of null mutations in either CRTAP or LEPRE1 [68]. Heterozygous variants in PPIB lead to OI type IX [80]. Collagen-directed biochemical studies of the few reported patients with PPIB defects showed inconsistent results. Two cases with PPIB variants causing moderately severe OI have normal levels of collagen α1(I) Pro986 3-hydroxylation [29]. Two lethal cases of CypB deficiency had 30% α1(I)Pro986 3-hydroxylation with the overmodification of the collagen helix [65, 81]. A PPIB knockout mouse model clarified this inconsistencies in humans and revealed the major functions of protein CypB in the regulation of collagen biosynthesis and post-translational modification [79]. Specifically, the absence of CypB leads to cell- and tissue-specific dysregulation of collagen hydroxylation and glycosylation, which is independent of impaired collagen folding [79]. The reduced hydroxylation of specific helical lysine residues in collagen influences intermolecular crosslinking and fibrillogenesis, which helps to maintain the mechanical properties of the bone tissues [79].

Lysyl-hydroxylase complex and collagen chaperones

Lysyl hydroxylase 1-3, encoded by PLOD1-3, respectively, catalyze the conversion of lysine to 5-hydroxyllysine residues in the α-chains of procollagen. Some of the hydroxyllysine residues are used as substrates by the lysyl oxidases family to form certain covalent cross links. Additionally, hydroxyllysine can be used as an attachment site for carbohydrates, leading to galactosylhydroxylysine or glucosylgalactosylhydroxylysine [82]. Lysyl hydroxylases require the formation of dimers for activity. For Lysyl hydroxylase 2, FKBP65 contributes to the dimerization [82].

LH2 contains a short splice form LH2a and a long splice form LH2b. LH2b contains an additional exon 13A encoding 21 amino acids and is mainly present in tissues rich in fibrillar collagens, while LH2a mostly does not exist in these tissues [82]. LH2b is a telopeptide lysyl hydroxylase and plays an important role in the generation of collagen telopeptide-derived pyridinoline cross-links [82]. In humans, variants in PLOD2 have been reported to cause recessive form of OI as well as Bruck syndrome type 2 [83, 84]. The variants are predominantly located in the C-terminal part and between exon 12 and 19 [85].
FKBP65, encoded by FKBPI0, affects lysyl hydroxylation indirectly. FKBP65 belongs to the family of FKBPI-type peptidyl-prolyl cis-trans isomerases and is a rough endoplasmic reticulum resident protein. FKBP65 participates in folding of type I procollagen as an important molecular chaperone-like protein. Some recessive OI cases and Bruck syndrome type 1 are caused by variants of FKBPI0 [86-89]. Interestingly, all variants of FKBPI0 are associated with a phenotype that overlaps with Bruck syndrome type 2 [90]. Both syndromes display a similar biochemical defect that is an intense underhydroxylation of the telopeptides of collagen type I [82]. Specifically, FKBPI0 variants indirectly affect procollagen maturation, by ablating the support of FKBP65 for collagen telopeptide hydroxylation by LH2, thereby reducing collagen cross-linking formation in tendon and bone matrix [87, 91]. Contractures and bone fractures are a variable manifestation of null/missense FKBPI0 variants [87, 91].

As reported, FKBP65 does not form complexes with LH1 or LH3, but only with LH2. Both depletion of FKBP65 and inhibition of FKBP65 PPIase activity does not affect the binding of monomeric LH2 to procollagen Iα1, but decrease the dimeric form of LH2. Since monomeric LH2 is inactive and only dimeric form of LH2 shows lysyl hydroxylase activity, this explains why variants in FKBPI5 have such a significant effect on LH2 [92]. Also, CypB can interact with LH2, as well as with LH1, LH3 and the putative LH2 chaperone FKBP65 [93]. CypB regulates collagen cross-linking by differentially influencing the lysine hydroxylation in the telopeptidyl and helical domains of tendon collagen type I [93].

HSP47, another chaperone for collagen, is a member of the serpin superfamily of serine proteinase inhibitors encoded by SERPINH1 gene. The expression of this protein is induced by heat shock. It preferentially binds to the N-terminus of the triple helical procollagen, stabilizes the folded collagen in the endoplasmic reticulum and shuttles properly folded collagen into cis-Golgi [94]. Hsp47−/− mouse embryos die before 11.5 days post coitus, showing ruptured blood vessels and aberrant epithelial tissue due to defects in collagen biosynthesis [95]. Hsp47−/− cells have high aggregation of type I collagen in the endoplasmic reticulum, delayed secretion, deficiency in N-propeptide processing of type I collagen and abnormal fibrillogenesis [96]. The first SERPINH1 variant (p.L326P) was identified in Dachshunds with OI [97]. Another case with severe recessive OI in humans was caused by the homozygous missense variant of SERPINH1 (p.Leu78Pro) [98].

Collagen secretion and endoplasmic reticulum stress

Proper conformational change in endoplasmic reticulum is necessary for a newly synthesized protein before its secretion. Excessive protein synthesis or protein misfolding can give rise to endoplasmic reticulum-stress and trigger the unfolded protein response. A few recessive OI genes, CREB3L1, MBTPS2 and TMEM38B, have been reported probably involved in endoplasmic reticulum stress and unfolded protein response [15].

CREB3L1 encodes the endoplasmic reticulum-stress transducer OASIS, which is a basic leucine zipper transcription factor of the CREB/ATF family [99-104]. To cope with ER stress, it is processed by adjusted intramembrane proteolysis. This protein plays a critical role in bone formation, through regulating the COL1A1 transcription and secretion of bone matrix proteins
[105]. A homozygous genomic deletion of CREB3L1 has been identified in a human family with severe recessive OI [106]. The knockout mice (OASIS<sup>−/−</sup>) exhibit severe osteopenia, spontaneous fractures, decreased cortical bones, calcaneus fracture as well as significantly reduced bone formation rate which are reminiscent of severe human OI [107, 108].

**MBTPS2** encodes a membrane-embedded zinc metalloprotease which is located in the Golgi membrane and important in development. It cleaves regulatory proteins including OASIS, ATF6 and SREBP which are transported from the endoplasmic reticulum membrane and involed in the endoplasmic reticulum-stress response. Two MBTPS2 missense variants at highly conserved residues cause moderate/severe recessive form of OI in two independent kindreds in humans [109]. The mutant site-2 metalloprotease leads to impaired operating in regulated intramembrane proteolysis of OASIS, ATF6 and SREBP substrates and decreased hydroxylation of the collagen lysine residue in the proband bone tissue [109].

**TMEM38B** encodes the endoplasmic reticulum membrane Tric-B protein, that acts as a monovalent cation channel. Tric-B is ubiquitously present in most mammalian tissues [110]. Tric channels, including Tric-A and Tric-B subtypes, regulate Ca<sup>2+</sup> release from intracellular stores [110, 111]. Tric-B knockout mice (Tric-B<sup>−/−</sup>) die shortly after birth because of respiratory failure [112, 113]. This demonstrates that Tric-B subtype plays an essential role in mouse perinatal development, in particular in Ca<sup>2+</sup> processing of the alveolar type II epithelial cells [113]. Also, Tric-B knockout mice display deficient bone ossification [112]. Bone histological examination indicates that collagen matrix deposition is decreased in Tric-b knockout mice [112]. Osteoblasts of the knockout mice display lowered Ca<sup>2+</sup> release and increased Ca<sup>2+</sup> content in endoplasmic reticulum [112]. They are associated with the swelling of endoplasmic reticulum and also the sign of the collagen trafficking defect through endoplasmic reticulum [112]. Tric-B function is indispensable to support the formation and normal release of large amounts of collagen by osteoblasts [112]. In humans only a few autosomal recessive OI cases have been reported with functional variants in TMEM38B [114-118].

**Osteoblast differentiation and mineralization**

Osteoblasts, terminally specialized and differentiated products of mesenchymal lineage cells, are responsible for synthesizing bone in groups of connected cells [119]. Aberrant proliferation, differentiation and function of osteoblasts significantly affect bone quality and quantity. After collagen is secreted into the extracellular matrix by mature osteoblasts, a variety of factors regulate the mineralization process [15]. Molecular biological studies on human OI provide evidence that few genes, *i.e.* WNT1, SP7, SERPINFI1 and IFITM5, play important roles in osteoblast differentiation and mineralization [15].

**IFITM5** (bone restricted Ifitm-like protein) is involved in bone mineralization and expressed most prominently in osteoblasts [120, 121]. In 2012, a heterozygous *de novo* variant (c.-14C>T) in the 5'-UTR of IFITM5 has been confirmed to be the causative variant for OI type V featuring hyperplastic callus development [122]. Transgenic mice overexpressing this OI type V mutant IFITM5 exhibited perinatal lethality, delayed or abnormal mineralization, skeletal malformation and *in utero* fractures [123]. Furthermore, a patient with the heterozygous *de*
novel variant (c.119C>T) resulting in an amino acid exchange p.Ser40Leu did not have typical signs of OI type V but exhibited limb shortening at prenatal period [124]. Analysis on Ifitm5-deficient mice suggested that that Ifitm5 deficiency might have a greater impact on prenatal bone development, since the effect of Ifitm5 deficiency on bone formation was more obvious in neonates than in young and adult mice [121]. In the study, FKBP11 was identified to interact with IFITM5 by liquid chromatography mass spectrometry [121]. However, it is not clear whether IFITM5 also participated in protein folding and immunosuppressant binding after linking to FKBP11 [121].

SERPINF1 encodes PEDF a member of the serpin family. PEDF does not exhibit serine protease inhibitory activity displayed by many other serpin proteins. It is a secreted neurotrophic protein participating in neuronal differentiation in retinoblastoma cells and also functions as potent inhibitor of angiogenesis. In bone, PEDF has been reported to inhibit osteoclast function by regulating osteoprotegerin expression [125]. Therefore, PEDF plays a role in the maintenance of bone homeostasis [125]. Loss of function and truncating mutations in SERPINF1 are reported in humans with type VI and severe OI, respectively [126, 127]. Recessive OI type VI is uniquely characterized with an increased amount of unmineralized osteoid. Studies identify that PEDF is involved in bone mineralization [126]. Pedf null mice (Serpinf1−/−) show increased mineral deposition, decreased trabecular bone volume and brittle bones recapitulating key structural and biochemical bone features observed in human OI type VI [128]. PEDF is absent in the sera of OI type VI patients, but restored serum PEDF level by using HDAd vector to express SERPINF1 in liver is not able to rescue the bone phenotype observed in Serpinf1−/− mice [129]. This suggests that PEDF might function in bone homeostasis in a context-dependent and paracrine manner [129]. However, intraperitoneal injection of PEDF-containing microspheres raises bone volume fraction and enhances biomechanical parameters of bone plasticity in PEDF-KO mice [130]. The report also indicates that a conserved motif on PEDF accentuates blockade of the Wnt receptor LRP6 to allow for terminal osteoblast differentiation [130].

SP7/Osx encodes Osterix protein, a member of the subfamily of Sp or XKLF transcription factors. It represents a bone specific transcription factor and plays an essential role in osteoblast differentiation [131]. A homozygous deletion resulting in a frameshift mutation in Osx is reported in an Egyptian child with recessive OI, featuring with delayed tooth growth, mild bone deformities and recurrent fractures [132]. In Osx-null mice, no endochondral or intramembranous bone formation occur. Studies on the embryos reveal that Cbfa1 is needed for Osx expression [133]. In osteoblast differentiation pathway, Osx works downstream of Cbfa1 [133]. The inactivation of Sp7 gene in mature bone causes a functional defect in osteoblasts and lowered bone formation [134]. The study shows that Osx is necessary to maintain osteoblast function following adult bone maintenance [134].

WNT1 is a member of the WNT gene family, which is composed of structurally related genes encoding secreted signaling proteins. WNT signaling has been implicated in bone homeostasis [135]. Studies in humans and mice have proved that the canonical Wnt/b-catenin signaling is indispensable for skeletal development and bone formation after childbirth by regulating the differentiation and activity of osteoclasts and osteoblasts [136-147]. Although the mechanism
by which WNT signaling adjusts bone formation has been well documented, the critical WNT ligands involved in this process remain unclear. Interestingly, several studies demonstrate that functional variants in WNT1 cause recessively inherited OI, also heterozygous WNT1 variants result in early onset osteoporosis in humans [148-151]. Furthermore, mice carrying homozygous null alleles of Wnt1 exhibit embryonic lethality [152]. The Swaying mouse (Wnt1<sup>sw/sw</sup>) carries a single nucleotide deletion in the exon of WNT1 resulting in a premature termination [153]. It exhibits spontaneous fractures, severe osteopenia, lowered bone strength, compromised collagen and mineral properties as well as declined osteoblast activity [153]. Therefore, these reports suggest that WNT1 could be a critical ligand involved in the WNT signaling adjusting bone homeostasis.

1.3.2 Genetics research progress in domestic animals

OI has also been documented in domestic animals, including Red Angus calves [154], Fleckvieh cattle [155], Charolais cattle [156], lambs [157], cats [158-160] and dogs [97, 161-163].

Around 50 newborn New Zealand Romney lambs in a flock were affected with OI and skin fragility [157]. They also showed multiple intrauterine bone fractures, soft bones, aberrant long bones, brachynathia inferior, obvious joint laxity, blue sclera, small pink teeth and subcutaneous oedema [157]. The defective collagen production by fibroblasts was confirmed by histopathologic and ultrastructural analyses [157]. The disease seemed to be caused by a de novo mutation in the germ line of a ram of the flock and was transmitted as an autosomal dominant disorder [157].

In six newborn half-sib Charolais calves, a hereditary bone disease was described [156]. In many aspects, the findings from the clinical, pathological, radiological, and genetic studies were similar to the congenital OI reported in humans and sheep [156]. However, the genetic reason of this case has not been clarified so far. Recently, a de novo mutation in COL1A1 (p.A1049_P1050delinsS, insertion-deletion) was described in OI type 2 in Fleckvieh cattle [155]. Another de novo mutation in COL1A1 (p.Gly355Ser) was identified as the causative variant of OI in two Red Angus calves [154]. The variant led to the substitution of a glycine residue with serine in the triple helical domain of the protein. In this region, glycine normally occupies every third position since it is critical for correct formation of the Type I collagen molecule.

In cat, the first described OI case was dated back to 1961 [160]. 30 years later, an OI-like syndrome was reported in a 12-year-old cat by radiographical and histological examinations and serum-based biochemical profile [158]. The third feline OI case reported in a 4.5-month-old male domestic shorthair cat exhibited bilateral femoral fractures after falling from a low position [159]. Also, decreased radio-opacity and thin cortices of all long bones were identified [159]. Based on post mortem examinations the kitten was diagnosed with OI [159].

Few forms of canine OI have been clarified on a molecular level. A 12-weeks-old male Golden Retriever was diagnosed with dentinogenesis imperfecta and osteogenesis imperfecta [161].
This dog carried a heterozygous missense mutation in \textit{COL1A1} (p.Gly208Ala), leading to an abnormal Gly-X-Y pattern of the collagen triple helix and resulting in over-hydroxylated alpha chains and delayed procollagen processing [161]. A heterozygous canine \textit{COL1A2} mutation due to replacement of "CTAG" at nucleotides 3991-3994 position with "TGTCATTGG" was reported leading to a C-terminal truncation in pro-alpha2(I) [162]. This genetic alteration resulted in increased density of pC-alpha2(I) and overhydroxylated alpha-chains [162]. The corresponding phenotype of the Beagle puppy presented with fractures was consistent with OI type III [162]. Furthermore, a missense mutation (p.L326P) in an evolutionary conserved domain of \textit{SERPINH1} was associated with a recessive form of OI in Dachshunds [97]. Therefore, in respects of clinical manifestations, etiology, and pathogenesis, canine and human OI appears to be similar [161].

2 Interdigital hyperplasia

2.1 Overview of lameness in dairy cattle

Foot disorders and the resulting lameness represent promising and economically highly relevant research topics in dairy cattle health, due to high incidence rates, severity as well as rather long-term duration processes [164-167]. The lesions that result in lameness in affected cattle cause severe pain and are a major issue in animal welfare. Lameness also has an economic impact, including losses from reduced milk productions, costs for culling, treatment costs, prolonged intervals of calving, as well as possibly nursing and caring labor [167]. It is expensive to rear replacement heifers and newly replaced cows are not initially as productive as mature ones [168]. A lame cow is less active in struggling for feed and therefore is more likely to be culled or die early. In addition, cows suffering from lameness before breeding show a decreased ability to conceive and cystic ovaries occur much more often in these animals [169, 170]. Lame cows are more reluctant to use an automatic milking system than healthy cows and when forced to do so, they show signs of stress [171].

From the perspectives of economic and animal welfare, foot diseases are an important health problem in dairy cows. Considerable funding is invested in the research relevant to bovine lameness. Numerous studies focused on risk factors and possible intervention strategies about foot disorders [172-175]. Some risk factors have been proved to be highly significantly associated with probabilities of lameness and foot lesions, including long-time standing on concrete, standing in wet slurry, factors which facilitating claw trauma, management failures especially in poor claw condition and inadequate foot caring [175]. Lameness control programmes on dairy farms have been developed and tested. However intervention remained widely ineffective, mainly due to inadequate implementation of programmes into practice. Primarily, the compliance in the program by the farmer and the veterinarian was insufficient [175]. Although lameness data are mainly collected by hoof trimmers, veterinarians and researchers should be familiar with this data information in order to play a key role in management of bovine lameness. More importantly, final action of dairy farmers are critical to improve dairy cow foot health. Therefore, awareness raising in farmers is most important. Farmers need to be trained to estimate foot health conditions and to realize correctly the advantages of taking action earlier and more thoroughly. Also, more insight information into
General Introduction

the economic losses and consequences of foot disorders needs to be conveyed. Furthermore, databases detailing bovine lameness are expected to gain availability as management tools of foot disorders.

There are various types of foot disorders distinguished. Interdigital phlegmon, also called interdigital necrobacillosis, footrot or foul-of-the-foot, is an acute or subacute necrotizing inflammation of the dermal layers of the interdigital skin and adjacent coronary band, causing sudden and severe lameness [176, 177]. The condition seems to be a mixed infection of multiple anaerobic pathogens with *Fusobacterium necrophorum* as the primary pathogen [176, 178]. Digital dermatitis, also called Mortellaro’s disease or foot warts, is a contagious disease of cattle characterized by ulcerative and necrotizing foot lesions [179, 180]. It is found worldwide as a leading reason of lameness in the cattle industry. Bovine interdigital dermatitis is an epidermatitis of the interdigital skin extending to the dermis. This is likely caused by the infection with *Bacteroides nodosus* and *Fusobacterium necrophorum* [181]. If the infection is chronic, heel horn erosion becomes its consequence [181]. Sole hemorrhage, is a common noninfectious lesion in lame cows, characterized with blood stains in the sole of the claw. It is predominantly caused by physical or metabolic disturbances because of overloading and pressure on the claws. White line disease is mainly caused by physical damage. It occurs when the sole separates from the side wall of the claw which allows foreign matter to penetrate or infect this region, resulting in hemorrhages and lesions in the white line. Sole ulcer, also called *Pododermatitis circumscripta*, can be applied to all the ulcerations in the sole, toe, and heel zones [182]. It is a very painful type of noninfectious claw lesion, arising when soft tissue inside the sole gets damaged and horn can not be formed normally. Also, thin sole, scissor claws, and many other foot disorders have been detected in dairy cows. The objective of the second research project was to elucidate the molecular cause of bovine interdigital hyperplasia (IH).

2.2 Research progress on interdigital hyperplasia

2.2.1 Clinical introduction

Interdigital hyperplasia, often referred to as tyloima or fibroma, is a firm and fibrous mass that protrudes from the interdigital space of the bovine claw. It may arise in one or multiple feet of an animal [183]. The lesion can gradually enlarge and become easily injured. Subsequently the affected feet provoke severe lameness. Skin injuries at the lesion site that are not treated properly and early enough, trigger foot infections. This secondary pathology very commonly invades deeper tissues, especially synovial structures, tendons, ligaments as well as bones [184, 185]. It was observed that wet conditions and accumulation of slurry also impaired the integrity of the external structure of the feet. Thereby, ulceration and local or deep foot infection frequently occur [183, 186].

Most intriguingly, it is noticed that in dairy cows, IH predominantly affects the hind feet. In contrast, beef cattle commonly display a IH localization at the fore feet [187]. Recently, IH was also confirmed in the right fore limb of an eight years old male buffalo after case history, clinical and histopathological examinations [188]. Furthermore, lameness in sheep and goat is also very common, and its causes include claw disorders like footrot, interdigital dermatitis, interdigital hyperplasia, as well as non-foot problems like penetrating foreign bodies, injury, joint-ill, and
muscular dystrophy [189]. In sheep and goat, lameness is one of the most important welfare issues and affects the industry and production, like decreased fertility, reduced growth rates of lambs and increased neonatal losses [189].

2.2.2 Prevalence and incidence

Claw disorders are common in dairy cattle with a rather high prevalence. Frequently, more than 70% of the cows have one or more claw lesions or abnormal shapes in front or hind claws [190-193]. Prevalences of different distinguished claw disorders vary widely from each other. Interdigital dermatitis, acute laminitis, heel-horn erosion and sole hemorrhage are often identified with the highest prevalence rates [190, 192-194].

In 1992 the prevalence of IH was reported to be 8.8% in dairy cows housed all year round in free stalls in the Netherlands [193]. 22% of the affected animals displayed IH on both claws [193]. Also on Dutch dairy farms, its prevalence in different housing systems was identified to differ from 9% (pasture period) to 14.2% (housing period) [195]. More recent Dutch IH studies observed lower prevalence rates (5% - 5.9%) [190, 192]. Possible reasons for this could be that their research subjects were collected from different large-scale farms with varying breeds, age ranges and farm environmental factors [183, 193, 196-200].

The prevalence rate of IH is relatively high in association with some particular breeds, such as Hereford in UK and Holstein in Germany [197, 201]. Most cases of IH arise at the age of 1-2 years. The frequency declines at around 6 years. So far no animals have been identified with a first appearance of IH at an age of 9 years or older [198, 199, 201]. Besides, the incidence of IH is commonly greater in males than in females [201].

2.2.3 Etiology and pathogenesis

IH has been defined as the proliferative formation of firm tissue mass in the interdigital region because of dermatogenous reactions following persistent irritation and inflammation of the interdigital skin [202, 203]. The etiology of IH is multifactorial. Infectious processes seem to be implicated in the pathogenesis of IH. IH affected tissues exhibit decreased microbial richness and diversity of the outer skin bacterial flora compared to healthy skin [204]. A tendency of an increased presence of Porphyromonas spp. and Treponema spp. is observed in IH samples [204]. Also, other microorganisms seem to be implicated in the etiopathogenesis of serial lesions related to IH, including Dichelobacter nodosus, Fusobacterium necrophorum and Spirochetes [203]. Cattle with IH also often exhibit common lesions of infectious etiology, like digital dermatitis and interdigital phlegmon [183, 194, 203-205].

Despite being a prevalent lesion, the etiopathogenesis of IH is not fully elucidated. The following factors have been speculated to predispose the development of IH. A chronic environmental irritation provoked by poor hygiene, grazing stubble or rocky pastures, as well as the long-term irritation caused by inadequate hoof trimming can lead to a predisposition to IH [183, 200, 206]. Another perspective are endogenous factors. These include: overweight-dependent excessive splaying of the toes, horn overgrowth causing manure trapped in the interdigital space, stretching or irritation of the interdigital skin due to unbalanced claws [183].
Underdeveloped interdigital ligaments, weak connective tissue, insufficient mineralization of the bone tissues causing irritation of the periostium, as well as excess interdigital fat have been reported as additional predisposing factors [200, 207].

Histologically, IH lesion represents epidermal thickening of the interdigital skin [203]. It involves multiplex papilliferous epidermal ridges bridged and covered by large quantity of keratin [203]. IH is characterized by the thickening of varous epidermal layers, i.e. Stratum granulosum, spinosum and corneum [183, 203]. The skin from IH lesions is hyperkeratotic or parakeratotic [183]. There is also evidence of chronic inflammation in the lesions [183]. Furthermore, histopathological examination in the male buffalo with IH identified proliferating spindle-shape fibroblasts with streaming and interlacing bundles [188]. These bundles were running in criss-cross directions leading to the occasional formation of whorls [188].

2.2.4 Genetic background

Holstein cattle are at higher risk of developing claw diseases than several other breeds, e.g. Meuse Rhine Yssel, Ayshire and Norwegian cattle [208-210]. The breed differences indicate a potential genetic background for claw disorders.

Heritabilities, repeatabilities and genetic correlations

Heritabilities of feet disorders are generally low [190, 192, 211-215]. Studies in Holstein dairy farms in Eastern Germany identified the highest estimated heritabilities for IH ($h^2 = 0.115$) in comparison to other common types of foot disorders, like digital dermatitis, sole ulceration and wall disorder [196]. In French HF dairy farms similar heritabilities of 0.14 were observed. Higher heritability of 0.43 was identified in trimmed cows [216]. However, heritabilities showed a tendency to decline when a larger number of cows were trimmed [216]. Heritability for IH was also found to be higher in younger animals with affection of more than one foot [200]. There seems to be a predisposition in some bull lineages [200, 206, 217].

The highest repeatability (0.33) was reported for IH compared to other claw disorders. This means that once a cow is affected with IH, it tends to be affected at subsequent occasions [216].

In 1964 IH has been reported to be significantly correlated with blood group factors [206]. A moderately strong genetic correlation between IH and dermatitis-erosion ((inter-)digital dermatitis and heel horn erosion) was identified (0.66 ± 0.08) [216]. The report indicated a possible similar genetic background for both diseases [216]. A negative genetic correlation was identified between IH and sole hemorrhage indicating a different genetic background [216]. Genetic correlations between trimming status and claw diseases were generally moderate to high in Holstein dairy (0.34 ± 0.07 with IH) [216].

GWAS reports

Locomotion, lameness as well as feet and leg conformation are useful indicator traits for claw health. Routinely, they are recorded in most breeding schemes. The indicator traits including rear leg rear view, rear leg side view, hock quality, bone structure, foot angle and locomotion score have been shown to be correlated with claw disorders [190, 192, 218, 219]. Previous
General Introduction

studies have detected several QTLs for these indicator traits in different dairy cattle breeds, like US Holstein cattle, Dutch Holstein cattle, German Holstein cattle, Danish Holstein Cattle, as well as in 3 French dairy cattle breeds (French Holstein, Normande, and Montbeliarde), although there was little coincidence among the QTLs reported [220-225].

Until 2015 in HF cattle significant (p < 0.05) and suggestive (p < 0.20) SNPs were identified for feet/claw disorders and trimming status [226]. The 10 significant SNPs, mainly located on BTA8, were associated with sole ulcer [226]. For IH, 17 suggestive associations were detected, located on chromosomes BTA6, BTA7, BTA8, BTA9, BTA11, BTA14, BTA20, BTA24, and BTA26 [226]. Interestingly among these, the SNP on BTA14 ARS-BFGL-NGS-4929 (64.0 Mbp) co-localized with markers associated with foot angle, BMS1899- BM4513 (0–76 cM) [220, 226]. The SNP on BTA6 BTA-77057-no-rs (87.4 Mbp) associated with IH was close to marker FBN14 (88 cM) associated with foot angle [223, 226]. However, so far no genes/markers/variants with major effects on claw disorders or IH have been reported.

**Genetic parameters comparison in different parities, lactation and frequency**

Calving and the onset of lactation were reported to increase the susceptibility to feet disorders [227, 228]. In HF cattle, van der Spek et al. found that the frequency of interdigital hyperplasia (from 5.7% to 11.1%) and white line separation increased dramatically (from 10.8% to 20.5%) from the first to later parities [229]. Harder et al. identified a higher heritability in the first parity in comparison of first and later parities together for a combined claw and leg disorder trait in German Holstein dairy cattle [214]. Conversely, in Dutch dairy cattle, van der Linde et al. found that heritabilities of parity 1 and parities ≥2 were quite comparable for the majority of claw traits [190]. However, the largest differences in heritabilities (parities ≥2 minus parity 1) were identified for IH (+0.06), interdigital dermatitis (+0.03), and sole ulcer (+0.04) [190]. The report also showed that the genetic correlation between 1 and ≥2 parities for digital dermatitis, sole ulcer and IH were significantly different from unity [190]. Repeatabilities of parity 1 and parities ≥2 were generally low for most claw traits, but IH showed the highest estimates (0.42 in first parity and 0.62 in later parities) [190]. Also, van der Spek et al. found that in HF cattle, IH had an increase in heritability from first to later parities [229]. But they reported that the genetic correlation between first and later parities was not significantly different from one for IH, double sole, sole ulcer, white line separation, and trimming status [229].

Van der Spek et al. showed that frequencies of feet disorders in HF cattle were generally similar from early to late lactation [229]. But a large difference in frequencies was found for trimming status, which was genetically correlated to IH [216, 229]. Germand et al. found that genetic correlations for same claw disorders between different DIMs (days in milk) were high for adjacent test days in Holstein cows in Thuringia, but substantially dropped close to zero for distant test days [230]. They also reported that heritabilities for the same claw disorders were relatively stable during the course of lactation (DIM 50 - DIM 300), but tended to increase at early and late lactation [230]. Van der Spek et al. identified the largest absolute difference of heritabilities from early to late lactation for IH (from 0.07 to 0.09), white line separation (from 0.04 to 0.02) and sole ulcer (from 0.05 to 0.07) [229]. In addition, they reported that the genetic
correlation between early and late lactation did not significantly differ from one for double sole, IH, sole hemorrhage and sole ulcer [229]. It was argued that the heritabilities for disease resistance were higher under some specific environmental conditions in which animals could express resistance to disease [231]. In high-frequency herds, cows are also more likely to express resistance to disease [231]. In high-frequency herds, cows are also more likely to express resistance against claw diseases. It has been reported that for bovine tuberculosis susceptibility genetic correlations between environments that differed in herd prevalence of the disease were all different from one [232]. For most claw traits, heritabilities were similar in herd groups with low and high frequencies of claw diseases [229]. However, IH and sole ulcer showed a lower heritability in low-frequency group (both 0.04) in comparison to high-frequency group (0.14 and 0.08, respectively) [229]. For sole ulcer a genetic correlation of $(0.75 \pm 0.14)$ between low-frequency group and high-frequency group was found significantly different from one [229]. But for IH and many other claw traits, the genetic correlations between low-frequency group and high-frequency group did not significantly differ from one [229].

3 Ribosomopathy

3.1 Ribosome biogenesis

Ribosome biogenesis is the process of forming ribosomes which occurs in all cells. Ribosomes are complex macromolecular machines that serve as the site of biological protein synthesis either linked to endoplasmic reticulum or in cytosol. The eukaryotic ribosome contains a small 40S (SSU) and large 60S (LSU) subunit. Ribosome biogenesis starts with the transcription of 47S precursor rRNA by RNA polymerase I, and 5S rRNA by RNA polymerase III. The 47S rRNA precursor is afterwards modified, processed and cleaved into 18S, 28S, and 5.8S rRNAs. The 18S rRNA (1874 nt) together with 33 small subunit r-proteins forms the SSU. The 5S rRNA (120 nt), 28S rRNA (4718 nt), and 5.8S rRNA (160 nt) along with 47 large subunit r-proteins comprise the LSU. The SSU contains mRNA entry and exit sites, the path along which mRNA progresses for decoding of the genetic instructions. The LSU takes responsibility for peptide bond formation and contains a polypeptide exit tunnel. Both subunits are assembled into the translationally active mature 80S ribosome for protein synthesis. The assembly process requires the coordinated synthesis of 4 rRNA, 80 core r-proteins, more than 150 associated proteins as well as approximately 70 snoRNAs [233, 234]. The complex ribosome biogenesis is one of the most energy-consuming processes in eukaryotic cells. There are approximately 10 million ribosomes in a rapidly growing mammalian cell. An active HeLa cell synthesizes about 7500 ribosomal subunits per minute [235]. Ribosome biogenesis is integral to all cell growth, proliferation and differentiation due to the roles of ribosomes in regulating the quality and quantity of proteins in a cell [236]. Perturbation of any step of ribosome biogenesis process can consequently lead to disorders in embryonic development or adult homeostasis.

3.2 Research progress on ribosomopathies

Ribosomopathies compose a collection of diseases, caused by mutations in the genes of ribosome biogenesis factors leading to an impaired or a decreased rate of ribosome biosynthesis. So far naturally occurring ribosomopathies have only been described in humans,
but not in cattle. A summary of reported causative mutations of ribosomopathies, existing animal models and corresponding clinical defects are presented in Table 1. Diamond-Blackfan anemia (DBA) is a congenital erythroid dysplasia characterized by anemia caused by selective decrease or absence of erythroid precursors, macrocytosis and reticulocytopenia [237]. The estimated incidence is around 5 cases per million live births. The majority of patients can be diagnosed in the first year of life, with lethargy and pallor as the most common symptoms. Other possible clinical characteristics include craniofacial defects, thumb abnormalities, cardiac defects, and short stature [237, 238]. In 1999, RPS19 was first identified as a DBA gene, and mutations in it accounted for approximately one-fourth of DBA cases [239, 240]. Over the past decades, a substantial progress has been made for improved explanation of the genetic causes of DBA. Changes in a large number of r-proteins in both SSU and LSU have been identified in DBA patients or families, including RPS17, RPS24, RPL5, RPL11, RPL35A, RPS7, RPS15, RPS27A, RPL36, RPS8, RPS14, RPL15, RPL18, RPS28, RPS26 and its binding partner TSR2 [241-248]. However, not all DBA cases can be attributed to r-protein dysfunction. A study has reported mutations in the hematopoietic transcription factor GATA1 [249]. Numerous animal models of DBA have been developed. Rpl11 homozygous deletion in adult mice leads to lethality within a few weeks [250]. Rpl11 heterozygous deletion in adult mice leads to anemia related with reduced erythroid progenitors and damaged erythroid maturation [250]. Rpl11-deficient zebrafish embryos shows hematopoietic failure [251]. Zebrafish models of RPS27 and RPL27 mutations show impaired erythrocyte production and damaged tail and/or brain development [252]. The findings in zebrafish and mouse DBA models were consistent with the genetic reports in humans and the models were also used for further mechanism studies.

5q-syndrome, first described in 1974 as a refractory anemia with a distinct karyotype, shares clinical and pathologic features with DBA [253]. It is an independent subtype of myelodysplastic syndrome (MDS) and a de novo deletion of the region between bands q21 and q32 on chromosome 5 is its sole cytogenetic abnormality [254]. Clinically, 5q-syndrome presents mainly among females and a relatively low risk of progression to acute myeloid leukemia compared with other types of MDS. Blood findings include a severe macrocytic anemia, normal/elevated quantities of platelets with hypolobulated micromegakaryocytes and few blasts in both peripheral blood and bone marrow [255]. RPS14 was determined as a 5q-syndrome gene in the RNA interference screen of each gene within the 5q-deletion region [256]. The haploinsufficient expression of RPS14 has been confirmed in patients with 5q-syndrome [256, 257]. RPS14 deficiency damages the processing of the 18S rRNA and the SSU formation [256]. A mouse model for human 5q-syndrome has been developed. With haploinsufficiency of the Cd74-Nid67 region including RPS14, it recapitulates the main features of the disease including macrocytic anemia, prominent dyserythropoiesis and monolobulated megakaryocytes [258]. Therefore, RPS14 haploinsufficiency is critical to the pathogenesis of 5q-syndrome [259].
Table 1. Summary of reported ribosomopathies

<table>
<thead>
<tr>
<th>Ribosomopathy</th>
<th>Causal Gene</th>
<th>Clinical Feature</th>
<th>Mouse Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diamond-Blackfan anemia (DBA)</td>
<td>RPS19, RPS24, RPS17, RPL35A, RPL5, RPL11, RPS7, RPL36, RPS15, RPS27A, RPS26, TSR2, RPS28, RPS8, RPS4, RPL15, RPL18, GATA1</td>
<td>-anemia&lt;br&gt;-congenital abnormalities&lt;br&gt;-pallor&lt;br&gt;-lethargy&lt;br&gt;-growth retardation</td>
<td>-lethality within a few weeks&lt;br&gt;-anemia&lt;br&gt;-reduced erythroid progenitors&lt;br&gt;-delayed erythrocyte development&lt;br&gt;-decreased red blood cell</td>
</tr>
<tr>
<td>5q-syndrome</td>
<td>RPS14</td>
<td>-mainly among females&lt;br&gt;-normal/elevated quantities of platelets&lt;br&gt;-blasts in peripheral blood and bone marrow</td>
<td>-decreased hematopoietic progenitor cells&lt;br&gt;-macrocytic anemia&lt;br&gt;-prominent dyserythropoiesis&lt;br&gt;-monolobulated megakaryocytes</td>
</tr>
<tr>
<td>Shwachman-diamond syndrome (SDS)</td>
<td>SBDS</td>
<td>-exocrine pancreatic insufficiency&lt;br&gt;-hematologic abnormalities&lt;br&gt;-increased risk of leukemia transformation&lt;br&gt;-skeletal and neurocognitive abnormalities</td>
<td>-embryonic-lethal&lt;br&gt;-bone marrow dysfunction</td>
</tr>
<tr>
<td>Treacher-Collins syndrome (TCS)</td>
<td>TCOFI</td>
<td>-craniofacial malformation&lt;br&gt;-microtia&lt;br&gt;-malar and mandibular hypoplasia&lt;br&gt;-developmental brain abnormalities</td>
<td>-severe cranioskeletal defects</td>
</tr>
<tr>
<td>Cartilage-hair hypoplasia (CHH)</td>
<td>RMRP</td>
<td>-short stature&lt;br&gt;-immunodeficiency&lt;br&gt;-increased risk of malignancies&lt;br&gt;-hematologic abnormalities</td>
<td>-viable homozygous RMRP null mice not available</td>
</tr>
<tr>
<td>Dyskeratosis congenita (DC)</td>
<td>TINF2, DRC1, TERC, TERT, NOP10, NHP2, TCAB1</td>
<td>-nail dystrophy&lt;br&gt;-mucosal leukoplakia&lt;br&gt;-reticular skin pigmentation&lt;br&gt;-somatic abnormalities&lt;br&gt;-increased risk of immune deficiency, pulmonary complications and malignancy</td>
<td>-shortened telomeres&lt;br&gt;-significant anemia&lt;br&gt;-decreased erythroblasts&lt;br&gt;-reduced hematopoietic stem cell populations</td>
</tr>
<tr>
<td>Roberts syndrome (RBS)</td>
<td>ESCO2</td>
<td>-growth retardation&lt;br&gt;-bilateral symmetric limb reduction&lt;br&gt;-craniofacial defects</td>
<td>-termination of mouse development</td>
</tr>
<tr>
<td>North American Indian Childhood Cirrhosis (NAIC)</td>
<td>UTP4</td>
<td>-cholestasis&lt;br&gt;-metabolic defects</td>
<td>-Cirrhin highly and predominantly expressed in liver</td>
</tr>
<tr>
<td>Bowen-Conradi syndrome (BCS)</td>
<td>EMG1</td>
<td>-growth retardation&lt;br&gt;-macrocephaly&lt;br&gt;-psychomotor delay&lt;br&gt;-knee and hip contractures&lt;br&gt;-abnormal feet</td>
<td>-defects in early cell lineage specification and in nucleologenesis&lt;br&gt;-pre-implantation lethality</td>
</tr>
</tbody>
</table>

Shwachman-diamond syndrome (SDS) is clinically characterized by exocrine pancreatic insufficiency, hematologic abnormalities, an increased risk of leukemia transformation as well as skeletal and neurocognitive abnormalities [260, 261]. The incidence of SDS is estimated at
1 in 50,000 births. Approximately 90% of SDS patients carry biallelic mutations in \textit{SBDS} gene [262]. Mouse models have suggested the roles of \textit{SBDS} in both hematopoietic and stromal cells of the bone marrow and ribosomal subunit maturation [263, 264].

Treacher-Collins syndrome (TCS) is a congenital disorder characterized by craniofacial malformation, microtia, malar and mandibular hypoplasia. Patients with TCS may present abnormalities of swallowing, airway, hearing and brain development [265, 266]. TCS occurs with an estimated frequency of 1 in 50,000 live births, primarily due to mutations in \textit{TCOF1} gene [267]. The nucleolar phosphoprotein Treacle, encoded by \textit{TCOF1}, colocalizes with upstream binding factor and RNA polymerase I. It plays an essential role in both rDNA transcription and rRNA methylation [268]. Mice haploinsufficient of \textit{TCOF1} show severe cranioskeletal defects, because the deficiency of the ribosome biogenesis regulator Treacle disrupts neural crest cell formation and proliferation [269]. Consistently, fish with \textit{TCOF1} loss-of-function also displays craniofacial malformation [270]. However, unlike other ribosomopathies, TCS was not associated with hematologic abnormalities, suggested a different ribosomal based pathophysiology.

Cartilage-hair hypoplasia (CHH) most commonly occurs in Old Order Amish and Finnish populations. It is clinically characterized by short stature resulting from skeletal dysplasia, hypoplastic hair, immunodeficiency, and increased risk of malignancies, particularly non-Hodgkin lymphoma and basal cell carcinoma [271-273]. Additionally, the hematologic abnormalities may include macrocytic anemia and lymphopenia. CHH is an autosomal recessive disorder arising from mutations in the untranslated \textit{RMRP} gene. \textit{RMRP} encodes the RNA component of the RNase mitochondrial RNA-processing (MRP) complex. RNase MRP RNA is classified as a snoRNA and is mainly located in the nucleolus. snoRNAs assemble small nucleolar ribonucleoprotein complexes (snoRNPs) and play roles in various steps in the rRNA synthesis [274]. The viable homozygous \textit{RMRP} null mice is not available, indicating that \textit{RMRP} is indispensable for early embryonic development [275].

Dyskeratosis congenita (DC) is an inherited bone marrow failure disease. It is clinically characterized by nail dystrophy, mucosal leukoplakia and reticular skin pigmentation [278]. In addition, patients often show somatic abnormalities and increased risk of immune deficiency, pulmonary complications and malignancy [278]. The syndrome is genetically heterogeneous. In addition to the most common X-linked form, autosomal dominant and recessive forms occur. In the past decades, around 50% DC patients have been assigned to mutations in the following genes, \textit{i.e. TERC, TERT, NOP10, NHP2, TINF2, TCAB1} [276-278]. This reveals that DC is a disorder caused mainly by defects in telomere maintenance. Since all the reported causative mutations are in genes whose products function in telomerase assembly or activity or in telomere integrity [278]. Studies on \textit{TERT}-deficient mice establish a direct link between the loss of TERT activity, telomere shortening and damaged erythropoiesis [279].

Mutations in \textit{ESCO2}, \textit{UTP4} and \textit{EMG1} have been detected in patients with Roberts syndrome (RBS), North American Indian Childhood Cirrhosis (NAIC) and Bowen-Conradi syndrome (BCS), respectively [280, 282, 284]. Their relevant functions in ribosome biogenesis have also been explored in mouse models [281, 283, 285].
References


General Introduction


Ashwell MS, Da Y, Van Tassell CP, Vanraden PM, Miller RH, Rexroad CE. Detection of putative loci affecting milk production and composition, health, and type traits in a United


General Introduction


42


CHAPTER 2

Osteogenesis imperfecta in an embryo transfer Holstein calf
Osteogenesis imperfecta in an embryo transfer Holstein calf

Xuying Zhang*, Marc Hirschfeld*,†, Julia Beck†, Alexandra Kupke§, Kernt Köhler**, Ekkehard Schütz*, Bertram Brenig*

*University of Goettingen, Institute of Veterinary Medicine, 37077 Göttingen, Germany
†University Medical Center Freiburg, Department of Obstetrics and Gynecology, 79110 Freiburg, Germany
‡Chronix Biomedical, 37077 Göttingen, Germany
§Philipps University Marburg, Institute of Virology, 35037 Marburg, Germany
**Justus Liebig University Gießen, Institute of Veterinary Pathology, 35392 Gießen, Germany
Short running title: Bovine osteogenesis imperfecta

Key words: Osteodysplasia, osteodystrophy, osteogenesis imperfecta, brittle bone disease, congenital bone disorder

Corresponding author:
Prof. Bertram Brenig
University of Goettingen
Institute of Veterinary Medicine
Burckhardtweg 2
D-37077 Goettingen
Phone +49-551-39-28383
Fax +49-551-39-33392
Email bbrenig@gwdg.de
Abstract
Musculoskeletal anomalies generally in combination with severe clinical symptoms, comprise a heterogeneous group of fairly common and mostly fatal disorders in man and animals. In Holstein cattle, complex vertebral malformation and Brachyspina are the most important lethal osteodysplasias. Here we describe the molecular analysis of a case of osteogenesis imperfecta in an embryo transfer Holstein calf from clinically healthy parents. Causes of human osteogenesis imperfecta have been clearly explored with a number of candidate genes reported. In a comparative sequence analysis of the affected calf and healthy control cattle none of these hitherto known genes harbored any causative variants. Therefore, to identify associated genomic regions, a genome-wide association study was performed. A total of six significantly associated SNP loci above a Bonferroni threshold of $-\log_{10}P= 5.7 (p = 0.05)$ were detected on bovine chromosomes 1, 5, 6, 17, 18, and 24. As the association analysis did not give an unequivocal evidence on the genetic cause, a whole genome re-sequencing of the affected calf and its parents was conducted. From this analysis 16 candidate genes with potential functional variants were selected. To determine the frequency of the variants and to analyze whether they were unique for the defect, 2,612 randomly chosen healthy Holstein cattle were genotyped. The analysis suggests that the defect was most likely due to a detrimental additive oligogenic effect.

Introduction
Development and maintenance of the skeletal system requires complex processes, which are characterized by continuous modelling and re-modelling providing structural and reservoir functions for the body throughout life. Hence, it seems comprehensible that a variety of pathological conditions interfere with a physiological development of the skeleton including genetic defects, nutritional deficiencies and hormonal disorders. This prompted a group of researchers in the early 1970s to establish a nosology and classification of human skeletal disorders in order to facilitate definitive diagnosis and help to delineate variant or newly recognized conditions [1]. Currently, this system includes 436 genetic disorders of the skeleton classified in 42 groups involving 364 genes [2]. Osteogenesis imperfecta (OI) together with other decreased bone density disorders form group 25 in this classification [2]. Osteogenesis imperfecta, also known as brittle bone disease, affects about 1/10,000 – 1/20,000 live births in humans [3, 4]. OI patients exhibit prominent skeletal abnormalities causing bone fragility and deformity, with widely varying clinical severity. In addition, joint laxity, scoliosis, kyphosis, dentinogenesis imperfecta as well as craniofacial abnormalities may represent concomitant phenomena in OI patients. Prominent extra-skeletal accessory manifestations of OI comprise blue/gray sclerae, hearing impairment, lung abnormalities and hypercalciuria [3, 5-7].

In humans, more than 90% of OI cases are caused by autosomal dominant mutations in the genes encoding type I collagen (COLIA1, collagen type I alpha 1 chain; COLIA2, collagen type I alpha 2 chain) [5]. A minority of OI cases occur as a result of recessive mutations in various genes, leading to different functional defects from structural to enzymatic as well as from intracellular transport to chaperones [3]. Most of these genes are involved in collagen metabolism, e.g. BMP1 (bone morphogenetic protein 1), CRTAP (cartilage associated protein), PLOD2 (procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2), FKBP65 (65-kDa FK506-
binding protein), \textit{HSP47} (heat shock protein 47), \textit{IFITM5} (interferon induced transmembrane protein 5), \textit{TRIC-B} (trimeric intracellular cation channel subtype B), \textit{WNT1} (Wnt family member 1) and \textit{SP7/OSX} (SP7 transcription factor/osterix) [8-16]. Depending on the causative gene and mutation, 13 different types of autosomal recessive and dominant OI are currently described in humans that are classified into five different groups based on their phenotypes [17].

Cases of OI in Holstein (HF) cattle have been reported since 1983 [18]. However, no causative mutations have been identified to date [18, 19]. A study with focus on non-collagenous proteins in OI affected Texan HF cattle showed a severe depletion of osteonectin and bone proteoglycan in bones and phosphophoryn in teeth [20]. OI has also been described in other cattle breeds (Charolais), sheep, cats and dogs [21-25]. However, only two forms of canine OI have been elucidated on a molecular level so far [26-29] and only recently \textit{COLIA1 de novo} mutations have been described in Fleckvieh and Red Angus [30, 31].

**Material and Methods**

**Ethical Statement**

Blood samples were collected during routine diagnostic parentage control with written owner consent. Blood samples were drawn exclusively by local veterinarians. The collection of samples was approved by the Lower Saxony State Office for Consumer Protection and Food Safety (33.19-42502-05-17A196) according to §8a Abs. 1 Nr. 2 of the German Animal Protection Law.

**Tissue preparation and histology**

Tissues were fixed in 10 % formalin. Thin bone sections were prepared with a diamond band saw and decalcified in EDTA at 37 °C for up to 14 days, rinsed in tap water over night and routinely embedded in paraffin. Sections of 4 µm thickness were dewaxed in xylene, rehydrated in a decreasing ethanol series and H&E stained in an autostainer (Microm, HMS 740).

**DNA isolation and Sanger sequencing**

DNA isolation from blood and liver tissue (OI calf) samples was performed using MagNa Pure LC DNA Isolation Kit I (Roche Diagnostics, Germany) or a modified salting out procedure [32].

Using NCBI/Primer-BLAST, PCR primer pairs were designed (Table S1) [33]. A total volume of 25 µl was prepared for each PCR reaction, including 1× PCR reaction buffer plus 20 mM MgCl₂, 0.5 µl of 10 mM dNTPs, 1 U FastStart Taq Polymerase (Roche), 1 µl of 10 pmol/µl each primer (Sigma-Aldrich, Germany), and 1µl of 20 ng/µl DNA. Reaction conditions included 95 °C for 10 min, then 30 cycles of 95 °C for 30 s, primer-specific annealing temperatures (Table S1) for 30 s and 72 °C for 45 s, followed by final elongation at 72 °C for 7 min.

Purification of PCR products was performed using Rapid PCR Cleanup Enzyme Set (New England Biolabs GmbH, Germany). Amplicons were sequenced using the BigDye™ Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Thermo Fisher Scientific GmbH, Germany) and fragments were separated on an ABI PRISM 3130xl Genetic Analyzer (Life
Osteogenesis imperfecta in an embryo transfer Holstein calf

Technologies, USA) following the manufacturers’ protocols. DNA sequences were aligned using SeqMan Pro software (DNASTAR Lasergene).

**Genome-wide association analysis (GWAS)**

For GWAS a total of 970 unaffected randomly chosen HF cattle (= controls) and the OI calf (= case) were genotyped using the Illumina BovineSNP50 BeadChip (56,919 SNP). Chips were processed on a HiScan SQ System (Illumina, USA) and raw data were converted using GenomeStudio 2.0.4 Software. Final reports were imported into SNP & Variation Suite 8.8.3 (GoldenHelix, USA). Prior to GWAS data were filtered using a call rate <0.81, number of alleles >2, minor allele frequency (MAF) <0.05, and Fisher’s HWE <0.001 (based on controls) as marker dropping criteria. The low call rate of <0.81 had to be used due to the results of the OI calf. Heterozygosity rates were calculated for each individual based on autosomes and outliers were removed. Finally, LD pruning was applied using a window size of 100bp, window increment of 5 bp, $r^2 = 0.5$ (LD threshold) and CHM (LD method). After filtering, 24,714 markers and 926 samples (925 controls, 1 case) remained for further analysis and association was calculated using an additive model and the Cochran-Armitage trend as described [34]. Family based association was analyzed using PBAT [35]. Associations were regarded as statistically significant above a Bonferroni threshold of $-\log_{10}P = 5.7$ ($p = 0.05$). Associations of markers ($-\log_{10}P$-value, y-axis) were plotted against their chromosomal positions (UMD3.1.1, x-axis).

**Next generation sequencing and data analysis**

Paired-end sequencing was performed using the TG NextSeq 500/550 High Output Kit v2 (Illumina) on a NextSeq500 platform. Sequence arithmetics (average depth of coverage, per base coverage) were calculated using bedtools [36, 37]. For the OI calf, its mother and father an average coverage of 36.3x, 27.2x and 38.3x was obtained, respectively. Sequences were aligned to the latest bovine genome build ARS-UCD1.2 using BWA-MEM with alignment parameters $t = 8$, $A = 1$, $B = 4$, $O = 6$, $E = 1$ and $k = 31$. For variant calling sorted bam-files were converted to vcf using samtools and bcftools [38]. Vcfs were imported into SNP & Variation Suite 8.8.3 (GoldenHelix, USA) for calling of SNPs and indels. Sequences were aligned to 3,093 samples of Run 7 of the 1000 bulls genome project [39]. SVDetect and delly was used for identification of larger genomic structural variations [40, 41]. CANDID v1.1 was used for gene prioritization [42, 43].

**Genotyping of variants**

Unless otherwise specified all positions refer to bovine genome assembly ARS-UCD1.2. Genotyping of $ABCA13$ variants, i.e. rs381405831 (4:g.7398705T>C) and rs110593220 (4:g.7324346T>C) was realized by multiplex-fluorescence resonance energy transfer (FRET)-PCR [44]. $ABCA13\_rs381405831$ primers and $ABCA13\_rs110593220$ primers (Table S1) were designed using NCBI/Primer-BLAST [33]. Probe/anchor, rs381405831_6-FAM/ROX and rs110593220_6-FAM/Cyanine 5 (Table S1), were designed using MeltCalc [45, 46]. Multiplex PCR was performed on LightCycler 480 (Roche) in a total volume of 15μl using FastStart Taq DNA Polymerase, dNTPack (Roche). One reaction mix included 0.75 U Faststart Taq DNA Polymerase, 3 nmol dNTPs, 6 pmol of each primer and probe, 1x Q-solution (Qiagen,
Osteogenesis imperfecta in an embryo transfer Holstein calf

Hilden, Germany), 1x PCR reaction buffer plus 20 mM MgCl₂, and 20 ng of DNA. Cycling conditions were 95°C for 10 min, followed by 35 cycles of 95°C for 30 s, 61°C for 30 s and 72°C for 30 s. The final elongation step was 72°C for 7 min. Melting analysis was performed using the appropriate filter set (483-610 filter comb for rs381405831; 483-670 filter comb for rs110593220) and the following program: 95°C for 30 s, 37°C for 30 s, 95°C continuous acquisition mode (2 acqui./°C), ramp rate 0.14°C/s, followed by 37°C for 10 s.

Amplification refractory mutation system (ARMS) [47] was designed to genotype QRFPR variant rs209556962 (6:g.4198178C>T). Primers, rs209556962_WTfwd/rev and rs209556962_MUTfwd/rev (Table S1), were designed using NCBI/Primer-BLAST [33], to amplify the wild type or mutant specific fragment with an internal amplification control sequence in a single reaction. The ARMS-PCR was performed on LightCycler 480 (Roche) in a total volume of 25 μl. One reaction mix included 1 U Faststart Taq DNA Polymerase (Roche), 5 nmol dNTPs (Roche), 10 pmol of each primer, 1x PCR reaction buffer plus 20 mM MgCl₂ (Roche), 1x EvaGreen (Jena Bioscience, Germany) and 10 ng of DNA. Cycling conditions were 95°C for 10 min, followed by 28 cycles of 95°C for 30 s, 66°C for 30 s and 72°C for 25 s. Final elongation step was 72°C for 7 min. Melting analysis was performed using the appropriate filter set and the following program: 95°C for 1 min, 40°C for 1 min, 75°C for 1 s, 90°C continuous acquisition mode (25 acqui./°C), ramp rate 0.02°C/s.

Statistical analysis

Hardy-Weinberg equilibrium was calculated as described [48]. \( \chi^2 \) scores were converted to \( p \)-values using R [49].

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. Supplemental tables (Table S1-S4) have been uploaded to figshare.

Results

Clinical and histological analysis

The affected OI calf was produced by embryo transfer and delivered four weeks prematurely. There were no reports of a previous transmission or case of osteogenesis imperfecta for both parents in their ancestry which excluded a dominant trait. As several sires in the pedigree have been widely used in Holstein breeding a recessive inheritance was also unlikely, however was not excluded completely. As shown in Figure 1 the pedigree reflects a normal Holstein cattle
family structure with no unexpected increased relatedness among animals due to inbreeding [50]. It is evident that sires Ro and especially Be appear more often in the paternal as well as maternal ancestral line.

Due to sucking weakness, bronchopneumonia and hypogammaglobulinemia, the OI calf was hospitalized shortly after birth. Four days after birth the OI calf developed a comminuted fracture of the right tibia distal of the knee joint and two days later a similar fracture on the left side without any obvious traumatic influence. The OI calf had to be euthanized and was dissected for diagnosis. Post mortem examination revealed comminuted fractures of both tibias approximately 2-5 mm distal of the epiphyseal plate (Figure 2). Fractures were accompanied by hemorrhages in the adjacent tissue. Additional fractures and disruptions of the articular cavities of the scapulae were determined. Other organs were without special findings. Bone tissue showed marked osteopenia characterized by thin cortical bone and thin trabeculae of the spongiosa, often consisting of woven instead of lamellar bone. Only few osteoclasts were present throughout the sections. Multifocally, trabeculae were widely separated by abundant mucinous matter, replacing bone marrow cells. In some regions, a direct transition between columnar cartilage and fibrous connective tissue was observed (Figure 3 A-D). From the postmortal and histological examinations an osteogenesis imperfecta (OI) was diagnosed.

**Analysis of known candidate genes causative for osteogenesis imperfecta**

In a first attempt a comparative sequencing of all potential OI candidate genes reported in humans, *i.e.* COL1A1, COL1A2, IFITM5, SERPINF1, CRTAP, LEPRE1, PPIB, SERPINH1, FKBP10, SP7, BMP1, TMEM38B and WNT1, was performed in a cohort comprising the affected OI calf, its parents and randomly chosen DNA of healthy Holstein cattle as controls. Only within IFITM5 and CRTAP non-synonymous variants were detected (Table 1).

In IFITM5 a homozygous missense variant (NC_037338.1:g.82966108C>A) was identified in the OI calf resulting in an amino acid exchange ENSBTAP00000041562.3:p.Ala30Ser. This variant was heterozygous in the parents and would therefore correspond to a recessive inheritance of the disorder. However, the homozygous alternative genotype was also detected in a healthy control excluding this variant as single causative candidate. Moreover, this variant seems to be tolerated as predicted by SIFT (0.42) [51].

In CRTAP a heterozygous inframe 9bp-deletion (NC_037349.1:g.7459319_7459327del) resulting in a truncation of three amino acids (ENSBTAP00000028588.4:p.Val18_Ala20del) was identified in the OI calf reflecting a dominant defect if causative for the disorder. The deletion was inherited by the mother and was also detected in one healthy control excluding this variant as causative.

The analysis of hitherto known candidate genes associated with human OI unexpectedly did not result in the identification of any single causative mutation in the OI calf. Therefore, a genome-wide association analysis was performed.
Table 1. Detection of DNA variants in osteogenesis imperfecta candidate genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Position</th>
<th>Variant</th>
<th>Genotypes (OI-family)</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>OI calf</td>
<td>My(a)</td>
</tr>
<tr>
<td>IFT5M5</td>
<td>11:g.82966108</td>
<td>rs209568970</td>
<td>Alt/Alt</td>
<td>Ref/Alt</td>
</tr>
<tr>
<td>CRTAP</td>
<td>22:g.7459319</td>
<td>rs466604499</td>
<td>Ref/Alt(c)</td>
<td>Ref/Alt</td>
</tr>
</tbody>
</table>

a) My: Mother (see also Fig. 1); b) S: Father (see also Fig. 1); c) Ref: Reference allele, Alt: Alternative allele.

Fig. 1 Pedigree of the affected calf

The pedigree depicts the six generations family structure of the affected calf. Selected individuals are indicated with letters. The pedigree was generated using the java webstart application Pedigree Chart Designer v2.0 (PCD).
Figure 2 Post mortal examination of left hind limb

The picture shows the medial view of the left hind limb including the proximal part of the metatarsus, ankle and distal part of the tibia. A comminuted fracture of the distal tibia is accompanied by hemorrhages in the adjacent tissue.
Fig. 3 Histological analysis of bone of the affected calf

Tissue sections (4 µm) were prepared and hematoxylin/eosin stained as described in Materials and Methods. (A-D) Affected calf showing (A) mucinous matter (Mc) replacing bone marrow (Bm), (B) direct transition of columnar proliferative cartilage (cpCh) to fibrous connective tissue, (C) not clearly separated metaphyseal bone (Mpb) from the hypertrophic chondrocytes (hCh) and (D) woven instead of lamellar bone. (E, F) Age-matched control calf with trabecular structured bone (Tbn) and distinct zones of epiphyseal growth.

Genome-wide association analysis

The OI calf was compared with 925 control cattle including parents and seven ancestors. Different genetic models were applied (dominant, recessive, basic allelic, genotypic, additive), however only the additive model resulted in associations above a genome-wide Bonferroni significance level ($-\log_{10} P = 5.7, p = 0.05$). As shown in Figure 4 single SNPs with significant
associations were detected on bovine chromosome 1 (ARS-BFGL-NGS-93288, \(-\log_{10}P = 5.88\)), 5 (BTA-00226457, \(-\log_{10}P = 6.48\)), 6 (HapMap49852-BTA-107572, \(-\log_{10}P = 7.33\)), 17 (BTB-01327818, \(-\log_{10}P = 5.77\)), 18 (ARS-BFGL-NGS-64158, \(-\log_{10}P = 8.57\)) and 24 (BTB-01542431, \(-\log_{10}P = 5.85\)). Besides OI candidate genes SP7 and WNT1 which are located on BTA5 upstream the respective associated SNP (BTA-00226457), none of the other associated SNPs were located in an OI candidate gene harboring chromosomal region. To analyze whether the associated chromosomal regions contained any potential so far unknown candidate genes, regions of 500kb up- and downstream of the associated SNPs were investigated. All genes in the corresponding regions were analyzed in silico regarding their function in bone, skeletal or cartilage development. According to the available databases (OMIM, MGI, OMIA, PubMed) none of the genes located in these regions have been described to be involved in any of the above traits. These results finally prompted us to perform a whole genome re-sequencing of the OI calf and its parents.

![Manhattan plot of genome-wide association analysis for osteogenesis imperfecta](image)

**Fig. 4** Manhattan plot of genome-wide association analysis for osteogenesis imperfecta

Marker associations are plotted as negative log-transformed \(P\) values against the position in the bovine genome (UMD3.1.1). Eight markers exceed the genome-wide significance threshold of \(-\log_{10}P = 5.7\) (black line).

**Next generation sequencing of OI calf and parents**

5’- and 3’-UTRs, splice sites and coding regions of the OI calf were analyzed for the presence of recessive homozygous polymorphisms, *de novo* mutations, compound heterozygous polymorphisms and larger genomic structural variations [52]. No structural variations, *de novo* mutations or compound heterozygous polymorphisms were identified qualifying as candidates. A total of 10,995,619 recessive homozygous variants were identified in the OI calf. 374 functional variants (1 deletion, 347 nonsynonymous variants, 4 nonsense mutations, 22 splice sites or promoter variants) remained after filtering. To further narrow potential OI causative variants, gene prioritization was performed using osteogenesis, bone fracture, loose joint, brittle...
teeth, white sclera and bone deformity as keywords. This resulted in a list of 1,038 prioritized genes. The two candidate gene lists were compared and sixteen genes were present in both lists, i.e. *QRFPR*, *ABCA13*, *MMAB*, *CCDC137*, *PAG21*, *KRT20*, *SOX8*, *GTPBP4*, *FAT4*, *SPOCK2*, *CACNA1I*, *COL6A6*, *AGER*, *MXRA5*, *CRTAC1* and *CALCR*. Validation of all detected functional variants was performed in these 16 genes, including 2 stop gain variants, 2 start lost mutations, 2 splice sites variants, 16 missense mutations, 1 inframe deletion and 1 frameshift insertion. Furthermore, in 8 functionally unknown genes 7 stop gain variants and 1 frameshift insertion were analyzed. All variants were comparatively sequenced in the affected calf, its parents and healthy control cattle (Table S2). None of the filtered and selected variants proved to be the only causative mutation.

**Discussion**

The molecular mechanisms underlying OI in humans have been analyzed in detail with a number of candidate genes reported. Although cases of OI have been reported in Charolais and Holstein cattle already decades ago, only recently different causative *COL1A1* de novo mutations have been described in Fleckvieh and Red Angus [18, 30, 31, 53]. In other domestic species, *i.e.* sheep, cat and dog, cases of OI have also been published, however, only in dogs causative mutations in *COL1A1*, *COL1A2* and *SERPINH1* have been elucidated [21, 22, 26, 27, 29]. Similar to humans, where OI is classified as orphan disease, OI also seems to be a rare disorder with low prevalence in domestic animals as estimated from the number of specific cases reported so far [54]. Hence, it was a fortunate coincidence that such a rare case of an OI calf had been reported to us and liver tissue was made available for further molecular analysis. In an initial attempt to identify causative mutations, all known OI associated genes that have been identified in humans were comparatively sequenced between the affected calf and its parents. Surprisingly, only two functional variants in *IFITM5* and *CRTAP* were identified. The inframe 9bp-deletion in *CRTAP*, however, had to be excluded as candidate, as it was heterozygous in the OI calf and transmitted by the healthy mother, also heterozygous for this variant. An additional proof for the exclusion of this variant came from the identification of an unrelated heterozygous healthy control cattle. Although the remaining variant in *IFITM5* was homozygous in the OI calf corresponding to a recessive inheritance, it was also rather unlikely to be the only cause as the same homozygous genotype was detected in a healthy control cattle. From the direct comparative sequencing of known OI candidate genes it was evident that the defect of the calf must have been the result of other and/or additional mutations in so far unassociated genes. To determine which further loci and genes could be involved, a GWAS was performed. For the GWAS it was possible to include seven relatives of the OI calf in the analysis also allowing a family-based association test using PBAT [35]. The results indicated that the defect was of oligogenic origin with at least six genome-wide significantly associated chromosomal regions. Furthermore, it was evident that the defect was presumably due to an additive genetic effect. However, a search in the flanking chromosomal regions up- and downstream of the associated SNPs did not lead to any apparent potential causative genes. Hence, a whole-genome re-sequencing of the OI calf and its parents was conducted and sequences examined regarding functional and structural variants. In combination with a gene prioritization it was finally possible to obtain a short list of 16 functional relevant candidate genes and eight additional loci with loss of function due to nonsense or frameshift mutations.
(Table S2). All variants were genotyped in the OI family and unrelated healthy control cattle. Variants in *AGER*, *SOX8*, *PAG21* and two loci (ENSBTAG000000313, NSBTAG000000303) were excluded because the OI calf was either homozygous wild type or heterozygous. For the remaining variants, the OI calf carried the homozygous mutated genotype, however, this genotype was also detected in several healthy control cattle. These results again supported the previous findings that the defect was most likely of oligogenic origin. Although, most of the OI cases in humans and domestic animals reported so far are caused by distinct monogenic mutations, an oligogenic additive genetic effect was not a complete surprise as combined heterozygous mutations in human OI have recently been described [55]. Therefore, different combinations of the remaining variants were tested in respect to their occurrence in the cattle population (2,612 random samples). Finally, the variants in *ABCA13* (rs381405831, rs110593220), *QRFPR* (rs209556962) and *IFITM5* (rs209568970) together were uniquely present in the OI calf (Table 2). *ABCA13* is a member of the ATP-binding cassette subfamily A and has a basic function as transmembrane transporter [56]. *ABCA13* is highly expressed in bone marrow stromal cells which can differentiate into bone, cartilage, adipocytes and hematopoietic supporting tissue [57]. A highly significant deviation from HWE of the homozygous carriers of the alternative allele (rs110593220) was detected for within the 2,612 samples with $\chi^2 = 252.97$ ($p = 1.2e-55$). For *QRFPR* a direct involvement in bone formation has already been demonstrated and *QRFPR*−/− mice show thin osteochondral growth plates, thickened trabecular branches and reduced osteoclast numbers. As the stop gain variant in *QRFPR* had been included on the bovine BeadChip (DEU_QRFPR4152699_4152699, DEU_QRFPR4152699_4152699_r) it was possible to determine the frequency of the genotypes in a larger cohort of 139,364 cattle showing a significant under-representation of the affected homozygous genotype ($\chi^2 = 9.7$, $p = 0.008$). It is therefore conceivable that variants in both genes together with *IFITM5* could contribute to the development of OI.

Table 2. Iterative determination of genotype frequency of OI causative variants in 2,612 random samples of Holstein cattle

<table>
<thead>
<tr>
<th>Gene</th>
<th>Position</th>
<th>Variant</th>
<th>Ref&gt;Alt</th>
<th>Ref_Ref</th>
<th>Ref_Alt</th>
<th>Alt_Alt</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>ABCA13</em></td>
<td>4:g7398705</td>
<td>rs381405831</td>
<td>T&gt;C</td>
<td>1479</td>
<td>978</td>
<td>155</td>
</tr>
<tr>
<td><em>ABCA13</em></td>
<td>4:g7324346</td>
<td>rs110593220</td>
<td>T&gt;C</td>
<td>1104</td>
<td>1442</td>
<td>66</td>
</tr>
<tr>
<td><em>QRFPR</em></td>
<td>6:g4198178</td>
<td>rs209556962</td>
<td>C&gt;T</td>
<td>19</td>
<td>28</td>
<td>16</td>
</tr>
<tr>
<td><em>IFITM5</em></td>
<td>11:g82966108</td>
<td>rs209568970</td>
<td>C&gt;A</td>
<td>6</td>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

a) *ABCA13* genotypes were determined in 2,612 samples. Samples harboring the homozygous alternative alleles were genotyped for the stop gain variant in *QRFPR*. The remaining 16 samples harboring the homozygous *QRFPR* variant were genotyped for the *IFITM5* variant.
Conclusion

In conclusion, the analysis of the OI case revealed an oligogenic origin of the disease attributable to additive effects of three candidate genes, i.e. \textit{ABCA13}, \textit{QRFPR}, and \textit{IFTIM5}. All three genes are directly or indirectly involved in bone development and formation or have been described as causative for human OI previously. The elucidation of an additive genetic effect demonstrates the complexity of the disease. The pathogenesis of the OI case presented here may be unique, however, demonstrates that cross-species genetic analyses might not always be straightforward.

Acknowledgments

The authors are grateful to D. Niehaus, S. Bierau, S. Balzer, L. Jüttner, M. Scharfenstein and S. Pach for expert technical assistance. C. Breidenbach, R. Fillies, J. Geibel, K. Götz, C. Otto, M. Westerhold and V. Zillgens are thanked for help with sequencing of OI candidate genes. The Association for Bioeconomy Research (FBF), the German Livestock Association (BRS) and the IT-Solution for Animal Production (vit) are thanked for support. We are grateful to A. Wehrend for clinical examinations and W. Wemheuer for organization of semen samples and contact with breeders. Thanks to W. Hermanns for providing histological bone sections of healthy control cattle. We acknowledge support by the German Research Foundation and the Open Access Publication Funds of the University of Goettingen. X. Zhang is a fellow of the Chinese Scholarship Council (CSC).

References


Table S1 Primers for genotyping of functional variants

<table>
<thead>
<tr>
<th>Primer name</th>
<th>5' - 3'</th>
<th>$T_a$ (°C)</th>
<th>Product size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFITM5_exon1 primers</td>
<td>CTCGAAGCAGGAGTCCGAG</td>
<td>61</td>
<td>276</td>
</tr>
<tr>
<td>IFITM5_Ex1fwd</td>
<td>CATGAAATGCTCCTGGGCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRTAP_exon1 primers</td>
<td>TCTAAAGGAGCGATCACGGG</td>
<td>60</td>
<td>564</td>
</tr>
<tr>
<td>CRTAP_Ex1.1fwd</td>
<td>CCAAAGAGGCGAACTCCG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FKBPI0_exon4 primers</td>
<td>GAGACAAGGAGCTGGTGAT</td>
<td>60</td>
<td>560</td>
</tr>
<tr>
<td>FKBPI0_Ex4fwd</td>
<td>GGTGGAATGGAGTGTCAGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RB1CC1_exon17 primers</td>
<td>AAGTGGAGAGAAGGGGAGGTG</td>
<td>60</td>
<td>305</td>
</tr>
<tr>
<td>RB1CC1_Ex17fwd</td>
<td>TGCTTCACTGCGATACAGGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OGN_exon2 primers</td>
<td>GGAGAGTGGAAAGACTGTCCA</td>
<td>60</td>
<td>422</td>
</tr>
<tr>
<td>OGN_Ex2fwd</td>
<td>ACGTGGAAAGGCAAATGGTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MATN1_exon2 primers</td>
<td>CTAAGTAGGCCCTGATGCC</td>
<td>60</td>
<td>583</td>
</tr>
<tr>
<td>MATN1_Ex2fwd</td>
<td>AGCCAGAGGGGAGCGACATTA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QRFPR Rs209556962 primers</td>
<td>GGCATGGAATTCAGGATTACG</td>
<td>60</td>
<td>204</td>
</tr>
<tr>
<td>rs209556962fwd</td>
<td>TGCCTAAACGAGATTCTCCGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QRFPR Rs208471741 primers</td>
<td>GGCATGGAAATTCAGGATTACG</td>
<td>60</td>
<td>204</td>
</tr>
<tr>
<td>rs208471741fwd</td>
<td>TGCTTCACTGCGATACAGGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CALCR Rs42358343 primers</td>
<td>TCTACATCTGACCAGAG</td>
<td>56</td>
<td>236</td>
</tr>
<tr>
<td>rs42358343fwd</td>
<td>GATCCTACCAGGACGCTCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRTAC1_r110294381 primers</td>
<td>TGTGTTCTCCCTCCGTACCA</td>
<td>56</td>
<td>141</td>
</tr>
<tr>
<td>rs110294381fwd</td>
<td>CCGTCAATGATCCTGAGT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MXRA5_r5382394831 primers</td>
<td>TCCACTTCCCGGTCTTTGAGG</td>
<td>60</td>
<td>228</td>
</tr>
<tr>
<td>rs382394831fwd</td>
<td>AGGCCGAAGTGAGATG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MXRA5_r538275669 primers</td>
<td>GATCCGGCATGGGAGATGGA</td>
<td>60</td>
<td>602</td>
</tr>
<tr>
<td>rs38275669fwd</td>
<td>AGTTGCACCAACGGTCTTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MXRA5_r53209853968 primers</td>
<td>GATCCGGCATGGGAGATGGA</td>
<td>60</td>
<td>602</td>
</tr>
<tr>
<td>rs209853968fwd</td>
<td>AGTTGCACCAACGGTCTTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGER Rs110735388 primers</td>
<td>TCTCCCTCATCTCCTCTGTT</td>
<td>61</td>
<td>208</td>
</tr>
<tr>
<td>rs110735388fwd</td>
<td>GAAGTGGAGAGACGAGGCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COL6A6 Rs110043748 primers</td>
<td>TGATGTGATGCTTTGCAAG</td>
<td>60</td>
<td>237</td>
</tr>
<tr>
<td>rs110043748fwd</td>
<td>GGAAGAGGAGTGCCACATCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene</td>
<td>Primer ID</td>
<td>ID</td>
<td>First Base</td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
<td>----------</td>
<td>------------</td>
</tr>
<tr>
<td>CACNA1I</td>
<td>rs135349128</td>
<td>65</td>
<td>585</td>
</tr>
<tr>
<td></td>
<td>rs135349128fwd</td>
<td></td>
<td>AGAAGACAGAGGCAGGGAA</td>
</tr>
<tr>
<td></td>
<td>rs135349128rev</td>
<td></td>
<td>CAGAGTCAGGTCAAGGAGA</td>
</tr>
<tr>
<td>SPOCK2</td>
<td>rs13689096r</td>
<td>65</td>
<td>389</td>
</tr>
<tr>
<td></td>
<td>rs13689096fwd</td>
<td></td>
<td>ATCCCTGTGGCCCTCCTATCC</td>
</tr>
<tr>
<td></td>
<td>rs13689096rev</td>
<td></td>
<td>TAGTACCGGCTCCTCCAG</td>
</tr>
<tr>
<td>FAT4</td>
<td>rs20778317f</td>
<td>61</td>
<td>155</td>
</tr>
<tr>
<td></td>
<td>rs20778317r</td>
<td></td>
<td>AGCCCCCTTCTCAAGTCTT</td>
</tr>
<tr>
<td></td>
<td>rs20778317r</td>
<td></td>
<td>TCTTGTCTTTCTTCCTCAG</td>
</tr>
<tr>
<td>FAT4</td>
<td>rs20889361f</td>
<td>61</td>
<td>152</td>
</tr>
<tr>
<td></td>
<td>rs20889361r</td>
<td></td>
<td>TGTCGTTGAGATGCCGCTA</td>
</tr>
<tr>
<td></td>
<td>rs20889361r</td>
<td></td>
<td>TCAACCTCCAGACACAGGA</td>
</tr>
<tr>
<td>GTPBP4</td>
<td>rs380469336f</td>
<td>64</td>
<td>229</td>
</tr>
<tr>
<td></td>
<td>rs380469336fwd</td>
<td></td>
<td>GGAAGAGGAAGCGGGAAGAC</td>
</tr>
<tr>
<td></td>
<td>rs380469336r</td>
<td></td>
<td>CAGCAGCTAGAACAGGCCAC</td>
</tr>
<tr>
<td>SOX8</td>
<td>rs438171974f</td>
<td>64</td>
<td>309</td>
</tr>
<tr>
<td></td>
<td>rs438171974fwd</td>
<td></td>
<td>CGGGAAACAAAAAGCCGTTGA</td>
</tr>
<tr>
<td></td>
<td>rs438171974r</td>
<td></td>
<td>GGTGTCCATGTGTCGAGATGA</td>
</tr>
<tr>
<td>KRT20</td>
<td>rs211283960f</td>
<td>61</td>
<td>308</td>
</tr>
<tr>
<td></td>
<td>rs211283960fwd</td>
<td></td>
<td>ACGGACTCTGCTCCTCATCCAT</td>
</tr>
<tr>
<td></td>
<td>rs211283960r</td>
<td></td>
<td>AGCTTGCTAGGCGGTCATT</td>
</tr>
<tr>
<td>PAG21</td>
<td>rs137216482f</td>
<td>61</td>
<td>305</td>
</tr>
<tr>
<td></td>
<td>rs137216482fwd</td>
<td></td>
<td>TACACACACGGATCTCCTCG</td>
</tr>
<tr>
<td></td>
<td>rs137216482r</td>
<td></td>
<td>TGACTGGATTTGTCGAGGC</td>
</tr>
<tr>
<td>CCDC137</td>
<td>rs110483301f</td>
<td>61</td>
<td>173</td>
</tr>
<tr>
<td></td>
<td>rs110483301fwd</td>
<td></td>
<td>GCCTACAGGGGCTAAAGAGA</td>
</tr>
<tr>
<td></td>
<td>rs110483301r</td>
<td></td>
<td>AAGGACGCTCAACTGTAGGC</td>
</tr>
<tr>
<td>CCDC137</td>
<td>rs41577609f</td>
<td>61</td>
<td>173</td>
</tr>
<tr>
<td></td>
<td>rs41577609fwd</td>
<td></td>
<td>GCCTACAGGGGCTAAAGAGA</td>
</tr>
<tr>
<td></td>
<td>rs41577609r</td>
<td></td>
<td>AAGGACGCTCAACTGTAGGC</td>
</tr>
<tr>
<td>CCDC137</td>
<td>rs207551614f</td>
<td>61</td>
<td>283</td>
</tr>
<tr>
<td></td>
<td>rs207551614fwd</td>
<td></td>
<td>CTTCCGAGTCCAGTCAACC</td>
</tr>
<tr>
<td></td>
<td>rs207551614r</td>
<td></td>
<td>TCACCGCAACTTCACCCTGTC</td>
</tr>
<tr>
<td>CCDC137</td>
<td>rs109405883f</td>
<td>61</td>
<td>278</td>
</tr>
<tr>
<td></td>
<td>rs109405883fwd</td>
<td></td>
<td>TGCCCTTCATCTCCTCCTCCT</td>
</tr>
<tr>
<td></td>
<td>rs109405883r</td>
<td></td>
<td>GCTCCCACTCAGGCTTTTT</td>
</tr>
<tr>
<td>MMAB</td>
<td>rs110898391f</td>
<td>60</td>
<td>337</td>
</tr>
<tr>
<td></td>
<td>rs110898391fwd</td>
<td></td>
<td>AGGTTGAGAATGCAGGGAGG</td>
</tr>
<tr>
<td></td>
<td>rs110898391r</td>
<td></td>
<td>TTCATAGACGCTCCAGCTTTTAG</td>
</tr>
<tr>
<td>ABCA13</td>
<td>rs381405831f</td>
<td>61</td>
<td>124</td>
</tr>
<tr>
<td></td>
<td>rs381405831fwd</td>
<td></td>
<td>CTTTACAGATGCCAAGGAGG</td>
</tr>
<tr>
<td></td>
<td>rs381405831r</td>
<td></td>
<td>GCCTGCACTAGGAGTGAGA</td>
</tr>
<tr>
<td></td>
<td>rs381405831l-6-FAM™/ROX™</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABCA13</td>
<td>rs831 probe</td>
<td></td>
<td>ggaaactggggaactag-6-FAM™</td>
</tr>
<tr>
<td></td>
<td>rs831 probe</td>
<td></td>
<td>ggaaactggggaactag-6-FAM™</td>
</tr>
<tr>
<td>Primers</td>
<td>Primer Sequence</td>
<td>ABI Probe</td>
<td>NCBI SNP ID</td>
</tr>
<tr>
<td>---------</td>
<td>----------------</td>
<td>-----------</td>
<td>-------------</td>
</tr>
<tr>
<td>rs110593220_fwd</td>
<td>ACAGGAAGTGTTGAGGCAG</td>
<td>ROX™-gcaggtctcatgattaaccagtga-Phosphate</td>
<td>61</td>
</tr>
<tr>
<td>rs110593220_rev</td>
<td>CCATCCTCTCTCGTCCATC</td>
<td>Cyanine 5</td>
<td></td>
</tr>
<tr>
<td>rs110593220-6-FAM™/Cyanine 5</td>
<td>gctgagcctttgccccg-6-FAM™</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs110593220_fwd</td>
<td>ACAGGAAGTGTTGAGGCAG</td>
<td>ROX™-gcaggtctcatgattaaccagtga-Phosphate</td>
<td>61</td>
</tr>
<tr>
<td>rs110593220_rev</td>
<td>CCATCCTCTCTCGTCCATC</td>
<td>Cyanine 5</td>
<td></td>
</tr>
<tr>
<td>rs110593220-6-FAM™/Cyanine 5</td>
<td>gctgagcctttgccccg-6-FAM™</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs110593220_fwd</td>
<td>ACAGGAAGTGTTGAGGCAG</td>
<td>ROX™-gcaggtctcatgattaaccagtga-Phosphate</td>
<td>61</td>
</tr>
<tr>
<td>rs110593220_rev</td>
<td>CCATCCTCTCTCGTCCATC</td>
<td>Cyanine 5</td>
<td></td>
</tr>
<tr>
<td>rs110593220-6-FAM™/Cyanine 5</td>
<td>gctgagcctttgccccg-6-FAM™</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs110593220_fwd</td>
<td>ACAGGAAGTGTTGAGGCAG</td>
<td>ROX™-gcaggtctcatgattaaccagtga-Phosphate</td>
<td>61</td>
</tr>
<tr>
<td>rs110593220_rev</td>
<td>CCATCCTCTCTCGTCCATC</td>
<td>Cyanine 5</td>
<td></td>
</tr>
<tr>
<td>rs110593220-6-FAM™/Cyanine 5</td>
<td>gctgagcctttgccccg-6-FAM™</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs110593220_fwd</td>
<td>ACAGGAAGTGTTGAGGCAG</td>
<td>ROX™-gcaggtctcatgattaaccagtga-Phosphate</td>
<td>61</td>
</tr>
<tr>
<td>rs110593220_rev</td>
<td>CCATCCTCTCTCGTCCATC</td>
<td>Cyanine 5</td>
<td></td>
</tr>
<tr>
<td>rs110593220-6-FAM™/Cyanine 5</td>
<td>gctgagcctttgccccg-6-FAM™</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs209556962_MUTrev</td>
<td>CAGAATTAAGAGGTAAACGTTCAGT</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------------</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td>GTPBP4_SDM-RFLP primers</td>
<td>rs380469336_SDM-RFLPfwd</td>
<td>GGAAGAGGAAGCGGGAGAAC</td>
<td>91</td>
</tr>
<tr>
<td>rs380469336_SDM-RFLPrev</td>
<td>CGGAGACCAGAGACATCAGCTGG</td>
<td>AGCTCGA</td>
<td></td>
</tr>
</tbody>
</table>
### Table S2 NGS variant analysis

<table>
<thead>
<tr>
<th>Position</th>
<th>Gene</th>
<th>Effect</th>
<th>rs number</th>
<th>Genotypes of OI trio</th>
<th>Genotypes of healthy control cattle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>O calf</td>
<td>Father</td>
</tr>
<tr>
<td>AC_000163.1:g.4152699C&gt;T</td>
<td>QRFPR</td>
<td>stop gained</td>
<td>rs209556962</td>
<td>ALT_ALT</td>
<td>REF_ALT</td>
</tr>
<tr>
<td>AC_000163.1:g.4152598G&gt;A</td>
<td>QRFPR</td>
<td>Missense</td>
<td>rs208471741</td>
<td>ALT_ALT</td>
<td>REF_ALT</td>
</tr>
<tr>
<td>AC_000161.1:g.10615176T&gt;C</td>
<td>CALC</td>
<td>Missense</td>
<td>rs42358343</td>
<td>ALT_ALT</td>
<td>REF_ALT</td>
</tr>
<tr>
<td>AC_000183.1:g.18900025C&gt;T</td>
<td>CRTAC1</td>
<td>Missense</td>
<td>rs110294381</td>
<td>ALT_ALT</td>
<td>REF_ALT</td>
</tr>
<tr>
<td>AC_000187.1:g.13882706G&gt;A</td>
<td>MXRA5</td>
<td>Missense</td>
<td>rs382394831</td>
<td>ALT_ALT</td>
<td>REF_ALT</td>
</tr>
<tr>
<td>AC_000187.1:g.138885931_138885933del</td>
<td>MXRA5</td>
<td>Inframe</td>
<td>rs382875669</td>
<td>ALT_ALT</td>
<td>REF_ALT</td>
</tr>
<tr>
<td>AC_000187.1:g.138886370G&gt;C</td>
<td>MXRA5</td>
<td>Missense</td>
<td>rs209853968</td>
<td>ALT_ALT</td>
<td>REF_ALT</td>
</tr>
<tr>
<td>AC_000180.1:g.27009084T&gt;C</td>
<td>AGER</td>
<td>Start lost</td>
<td>rs110735388</td>
<td>ALT_ALT</td>
<td>REF_ALT</td>
</tr>
<tr>
<td>AC_000158.1:g.153331589A&gt;G</td>
<td>COL6A6</td>
<td>Missense</td>
<td>rs110043748</td>
<td>ALT_ALT</td>
<td>REF_ALT</td>
</tr>
<tr>
<td>AC_000162.1:g.111601472_111601475dup</td>
<td>CACNA1I</td>
<td>Frameshift</td>
<td>rs135349128</td>
<td>ALT_ALT</td>
<td>REF_ALT</td>
</tr>
<tr>
<td>AC_000185.1:g.28314738T&gt;C</td>
<td>SPOCK2</td>
<td>Missense</td>
<td>rs136889096</td>
<td>ALT_ALT</td>
<td>REF_ALT</td>
</tr>
<tr>
<td>AC_000174.1:g.32713454C&gt;A</td>
<td>FAT4</td>
<td>Missense</td>
<td>rs207788317</td>
<td>ALT_ALT</td>
<td>REF_ALT</td>
</tr>
<tr>
<td>AC_000174.1:g.32887994T&gt;C</td>
<td>FAT4</td>
<td>Missense</td>
<td>rs208893611</td>
<td>ALT_ALT</td>
<td>REF_ALT</td>
</tr>
<tr>
<td>AC_000170.1:g.46690834G&gt;A</td>
<td>GTPBP4</td>
<td>Missense</td>
<td>rs380469336</td>
<td>ALT_ALT</td>
<td>REF_ALT</td>
</tr>
<tr>
<td>AC_000182.1:g.790262A&gt;C</td>
<td>SOX8</td>
<td>Missense</td>
<td>rs43817974</td>
<td>REF_REF</td>
<td>REF_REF</td>
</tr>
<tr>
<td>AC_000186.1:g.39052073A&gt;T</td>
<td>PAG21</td>
<td>stop gained</td>
<td>rs137216482</td>
<td>REF_REF</td>
<td>REF_REF</td>
</tr>
<tr>
<td>AC_000176.1:g.51758959A&gt;G</td>
<td>CCDC137</td>
<td>Missense</td>
<td>rs110483301</td>
<td>ALT_ALT</td>
<td>REF_ALT</td>
</tr>
<tr>
<td>AC_000176.1:g.51759001G&gt;A</td>
<td>CCDC137</td>
<td>Missense</td>
<td>rs41577609</td>
<td>ALT_ALT</td>
<td>REF_ALT</td>
</tr>
<tr>
<td>AC_000176.1:g.51759668T&gt;C</td>
<td>CCDC137</td>
<td>Missense</td>
<td>rs207551614</td>
<td>ALT_ALT</td>
<td>REF_ALT</td>
</tr>
<tr>
<td>AC_000176.1:g.51761251C&gt;T</td>
<td>CCDC137</td>
<td>Missense</td>
<td>rs109405883</td>
<td>ALT_ALT</td>
<td>REF_ALT</td>
</tr>
<tr>
<td>AC_000161.1:g.7309608T&gt;C</td>
<td>ABCA13</td>
<td>splice donor</td>
<td>rs381405831</td>
<td>ALT_ALT</td>
<td>REF_ALT</td>
</tr>
<tr>
<td>AC_000161.1:g.7237183T&gt;C</td>
<td>ABCA13</td>
<td>Missense</td>
<td>rs110593220</td>
<td>ALT_ALT</td>
<td>REF_ALT</td>
</tr>
<tr>
<td>AC_000176.1:g.41731744C&gt;T</td>
<td>KRT20</td>
<td>start lost</td>
<td>rs211283960</td>
<td>ALT_ALT</td>
<td>REF_ALT</td>
</tr>
<tr>
<td>AC_000174.1:g.65887778A&gt;C</td>
<td>MMA</td>
<td>splice acceptor</td>
<td>rs110898391</td>
<td>ALT_ALT</td>
<td>REF_ALT</td>
</tr>
</tbody>
</table>
Osteogenesis imperfecta in an embryo transfer Holstein calf

| AC_000158.1:g.119962272G>A | ENSBTAG00000048150 | stop gained | rs136954177 | ALT_ALT | REF_ALT | REF_ALT | 5 | 1 | 2 | 2 |
| AC_000186.1:g.38843505dup | ENSBTAG00000040140 | frameshift | rs382569063 | ALT_ALT | REF_ALT | REF_ALT | 7 | 0 | 5 | 2 |
| AC_000182.1:g.39309083G>T | ENSBTAG00000047781 | stop gained | rs109822444 | ALT_ALT | REF_ALT | REF_ALT | 7 | 4 | 0 | 3 |
| AC_000186.1:g.38405455G>C | ENSBTAG00000047145 | stop gained | rs134136884 | ALT_ALT | REF_ALT | REF_ALT | 7 | 5 | 0 | 2 |
| AC_000176.1:g.13526380C>T | ENSBTAG0000000560 | stop gained | rs207627937 | ALT_ALT | REF_ALT | REF_ALT | 30 | 21 | 8 | 1 |
| AC_000178.1:g.20157841C>A | ENSBTAG00000040580 | stop gained | rs211629862 | ALT_ALT | REF_ALT | REF_ALT | 6 | 2 | 2 | 2 |
| AC_000170.1:g.63327326C>T | ENSBTAG00000031361 | stop gained | rs209803099 | REF_ALT | REF_ALT | REF_ALT | 7 | 1 | 6 | 0 |
| AC_000160.1:g.103500366G>T | ENSBTAG00000030386 | stop gained | rs468298771 | REF_REF | REF_REF | REF_ALT | 6 | 6 | 0 | 0 |

a) Position refer to UMD3.1.1
CHAPTER 3

Interdigital hyperplasia in Holstein Friesian cattle is associated with a missense mutation in the signal peptide region of the tyrosine-protein kinase transmembrane receptor gene
Interdigital hyperplasia in Holstein cattle is associated with a missense mutation in the signal peptide region of the tyrosine-protein kinase transmembrane receptor gene

Xuying Zhang¹, Hermann H. Swalve², René Pijl³, Frank Rosner², Monika Wensch-Dorendorf², Bertram Brenig¹*

¹ University of Goettingen, Institute of Veterinary Medicine, 37077 Göttingen, Germany
² Animal Breeding, Institute of Agricultural and Nutritional Sciences, Martin-Luther-University Halle-Wittenberg, Halle/Saale, Germany
³ Fischershaeuser 1, 26441 Jever, Germany

XZ: xuying.zhang@stud.uni-goettingen.de
HHS: hermann.swalve@landw.uni-halle.de
RP: r.pijl@t-online.de
FR: frank.rosner@landw.uni-halle.de
MWD: monika.dorendorf@landw.uni-halle.de
BB: bbrenig@gwdg.de

*Corresponding author
Prof. Bertram Brenig
Institute of Veterinary Medicine
Burckhardtweg 2
D-37077 Göttingen
Germany
Phone +49-551-39-28383
Fax +49-551-39-33392
Email bbrenig@gwdg.de
Interdigital hyperplasia in Holstein cattle is associated with a missense mutation in the signal peptide region of the tyrosine-protein kinase transmembrane receptor gene

Abstract

Bovine interdigital hyperplasia (IH) is a typical disease of the foot with varying prevalence depending on age, breed and environmental factors resulting in different degrees of lameness. In studies based on assessments of claw health status at time of hoof trimming and applying genetic-statistical models to analyze this data, IH consistently exhibits high estimates of heritability in the range of 0.30 to 0.40. Although some studies have identified chromosomal regions that could possibly harbor causative genes, a clear identification of molecular causes for IH is lacking so far. While analyzing the large database of claw health status as documented at time of hoof trimming, we identified one herd with extreme prevalence of IH of > 50% of affected Holstein dairy cows. This herd subsequently was chosen as the object of a detailed study. A total of n = 91 cows was assessed and revealed a prevalence of 59.3% and 38.5% for IH cases, documented as ‘at least one-sided’ or ‘two-sided’, respectively. Cows were genotyped using the BovineSNP50 BeadChip. A genome wide association study revealed two significantly associated chromosomal positions (\(-\log_{10}P = 5.57\)) on bovine chromosome 8 (BTA8) located in intron 5 and downstream of the receptor tyrosine kinase-like orphan receptor 2 (ROR2) gene. As ROR2 plays a key role in ossification of the distal limbs and is associated with brachydactylies in humans, it was a reasonable candidate for IH. A comparative sequencing of the ROR2 gene between cases and controls revealed two missense variants in exon 1 (NC_037335.1:g.85,905,534T>A, ARS-UCD1.2) and exon 9 (NC_037335.1:g.86,140,379A>G, ARS-UCD1.2), respectively. Genotyping of both variants in the cohort of 91 cattle showed that the exon 1 variant (rs377953295) remained significantly associated with IH (p < 0.0001) as a risk factor of the disease. This variant resulted in an amino acid exchange (ENSBTAP00000053765.2:p.Trp9Arg) in the N-terminal region of the ROR2 signal peptide which is necessary for proper topology of the polypeptide during translocation. Quantification of ROR2 mRNA and ROR2 protein showed that the variant resulted in a significant suppression of ROR2 expression in homozygous affected compared to wild type and carrier cows.

Key words: Claw disease, cattle, interdigital hyperplasia, limax, tyloma, ROR2
Introduction

Foot diseases are a major and increasing health problem in dairy cattle and show rather high prevalence [1]. Often over 70% cows in dairy herds experience one or more front or hind claw lesions or deformities during their life [2-5]. Due to milk production losses and veterinary costs as well as culling of severe cases, feet and leg are of economical importance [6, 7]. Besides disorders like double sole, sole ulcer, sole hemorrhage, white line separation, and digital dermatitis, interdigital hyperplasia (IH) plays an important role with a prevalence of around 5-14.2% [2, 4, 5, 8]. Normally IH begins with a small and painless protrusion of the interdigital skin which sometimes can already be diagnosed at calf age. In the final stage clinical signs of IH are firm tumor like masses found in the interdigital space with skin lesions that can result in deep necrosis and phlegmonous inflammations. Infectious processes are possibly implicated in the pathogenesis of IH. Tissues affected with IH exhibit decreased microbial richness and diversity compared to healthy skin [9]. An increased susceptibility for IH has been described depending on the number of parities and lactations. From first parity to later parities, IH frequency and estimates of heritability increase [10]. In a longitudinal study in which cows that had been affected with IH were analyzed with respect to the point in time when they were first diagnosed with IH, it was found that about 50% of all cows that were susceptible were affected in their first lactation, another 25% in their second lactation and another 10 to 15% in third lactation while the remaining cows only exhibited IH very late in life [11]. Already in very early literature, it has been hypothesized that a genetic predisposition is the main cause for the development of IH [12]. In more recent studies using data collected at time of hoof trimming, rather elevated estimates of heritabilities are found [1, 16, 17] and these even amount to magnitudes of 0.30 to 0.40 when threshold models are used or estimates from a linear estimation are converted to the underlying scale [13-16]. In a study estimating the odds ratio for cows to be affected by IH when comparing the status of female ancestors, it was found that the risk to be affected increases 8.5-fold when comparing cows with affected dams and grandsams vs. cows with non-affected ancestors [17]. Up to recently, evidence for a genetic predisposition on a molecular level has been scarce and inconclusive [1, 16, 18]. In a very recent study, a number of chromosomal regions putatively involved in the etiology of IH have been identified [19].

Aim of the present study was to identify genes that might play a role in the development of IH by means of a genome-wide association study based on a case-control design. As a further step, genes in the identified regions were subject to sequencing for detection of single nucleotide polymorphisms. Finally, polymorphic sites were further analyzed with respect to differences in gene expression. A herd of dairy cows exhibiting an extreme prevalence of IH served as a basis for the study.

Materials and Methods

Ethical statement

Clinical inspections and sampling of cattle were done during routine hoof trimming on farm with written owner consent. Samples were taken exclusively by local veterinarians. The
Interdigital hyperplasia in Holstein cattle is associated with a missense mutation in the signal peptide region of the tyrosine-protein kinase transmembrane receptor gene.

A collection of samples was approved by the Lower Saxony State Office for Consumer Protection and Food Safety (33.19-42502-05-17A196) according to §8a Abs. 1 Nr. 2 of the German Animal Protection Law.

Clinical investigations and sample collection

Clinical investigations were completely connected to the hoof trimming routine as practiced on the study farm. The hoof trimming routine consisted on visits of the hoof trimmer at intervals of approximately three months. At each visit, around half of the cows of the herd were subject to hoof trimming, i.e. two consecutive visits were supposed to cover the entire herd. The herd therefore was inspected at two consecutive hoof trimming events in October, 2016, and January, 2017. Another hoof trimming event in January 2018 was used to take 2mm fine needle biopsies (Tru-Punch Sterile Disposable Biopsy Punch, Sklar Instruments, VWR, Germany) from hyperplastic interdigital skin of eight affected cows for RNA and protein analysis. Healthy control samples were collected from an abattoir.

A total of 110 cattle was assessed at visits 1 (V1) and 2 (V2) (Table 1). Excluding animals that had been presented twice, i.e. at the first and second visit, n = 91 animals remained. During the visits, individuals were inspected visually, phenotypes were recorded and pictures were taken for documentation. Blood samples were taken for DNA extraction during visits 1 and 2 (n = 91).

Table 1 Results of clinical inspections during two farm visits

<table>
<thead>
<tr>
<th>Clin. inspection</th>
<th>Cattle (n =)</th>
<th>Interdigital hyperplasia affected (n =)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Type A (%)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>V1</td>
<td>58</td>
<td>34 (59)</td>
</tr>
<tr>
<td>V2</td>
<td>52</td>
<td>37 (71)</td>
</tr>
<tr>
<td>Total</td>
<td>110</td>
<td>71 (65)</td>
</tr>
</tbody>
</table>

<sup>a</sup> V1: First farm visit in October 2016, V2: Second farm visit in January 2017; <sup>b</sup> IH type A: At least one affected hind leg; <sup>c</sup> IH type B: Both hind legs affected.

Genome-wide Association Study (GWAS) and statistical analysis

Genotyping of the 91 samples was performed using the Illumina BovineSNP50 BeadChip. Raw data were processed using GenomeStudio V2011.1 (Illumina, San Diego, USA). Final reports were imported into SVS 8.8.3 (Golden Helix, Bozeman, USA) and low quality SNPs were filtered if call rates < 90 %, MAF < 0.01 and Fisher’s HWE p < 0.0001 (based on controls). LD pruning was performed using default parameters. Samples were filtered with call rates < 0.95. Mitochondrial DNA and sex chromosomes were excluded from the analysis. After filtering 70 samples and 45,232 SNPs remained in the analysis. Genome-wide associations were calculated under an additive and dominant model [20]. Associations were regarded as genome-wide statistically significant above a threshold of -log<sub>10</sub> P = 5.47 (p = 0.05) [21]. Associations of markers (-log<sub>10</sub> P-value, y-axis) were plotted against their chromosomal positions (UMD3.1.1, x-axis).

All n = 91 cows from V1 and V2 were genotyped for rs43572154 and rs377953295. Genotype counts, genotype and allele frequencies were calculated for both variants. Hardy-Weinberg test
was calculated according to Rodriguez et al. [22]. Frequency distribution for alleles and genotypes vs. types of definitions of disease status (Type A IH, Type B IH) was calculated using both Chi-squared test and Fisher’s Exact Test.

**Sanger sequencing and genotyping**

DNA was extracted from blood samples using MagNa Pure LC DNA Isolation Kit I (Roche Diagnostics, Mannheim, Germany) or a modified salting out procedure [23]. For Sanger sequencing of ROR2, primers were designed to amplify exons and intron-exon boundaries (Supplementary Table S1) [24]. PCR products were purified with Rapid PCR Cleanup Enzyme Set (New England Biolabs GmbH, Frankfurt am Main, Germany) and sequenced using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Fisher Scientific GmbH, Schwerte, Germany) on an ABI PRISM 3130xl Genetic Analyzer (Life Technologies, Foster City, USA) according to the manufacturers’ protocols. The same primers were used for sequencing PCR products. SeqMan Pro software (version 15.0.0 (160) Intel, DNASTAR) was used for sequence alignments. Genotyping of variants NC_037335.1:g.85,905,534T>A and NC_037335.1:g.86,140,379A>G (ARS-UCD1.2) was done by diagnostic sequencing.

**RNA and protein preparation**

Harvested tissue was immediately immersed in RNAlater (Ambion, ThermoFisher Scientific, Dreieich, Germany) and stored at 4°C. Tissue samples were homogenized in the Qiazol Lysis Reagent (Qiagen, Hilden, Germany). According to the user guide of TRIzol Reagent, total RNA was extracted and eluted in 30 μl RNase-free water and the concentration and purity were measured on a NanoDrop (ThermoFisher Scientific, Dreieich, Germany). 0.2 μg of RNA was converted to cDNA using Maxima H Minus First Strand cDNA Synthesis Kit with dsDNase (ThermoFisher Scientific, Dreieich, Germany). Protein was extracted and solubilized in an optimized EDTA lysis buffer [25]. Protein concentrations were measured using the Bradford assay [26]. RNA and protein samples were stored at –80°C until further use.

**ROR2 isoform detection and semi-quantitative real time RT-PCR**

Potential ROR2 isoforms were analyzed in tissue samples of the interdigital region, lung and spleen of healthy cattle using primer pairs ROR2_cDNA_1/2_fwd (5´-CAGCCCTGTCCAACCTCTGA-3´), ROR2_cDNA_1/2_rev (5´-CCGTATTCCGTCTTGCGGA-3´), ROR2_cDNA_2_fwd (5´-GGCATGGAGTACCTGTCCAG-3´), and ROR2_cDNA_2_rev (5´-GGCCAGGTCTTTGTGGACCA-3´). Real-time quantitative PCR was performed using cDNA synthesized from RNA isolated from fine needle biopsies (FNB) of IH affected cattle and controls. Statistical testing for mRNA expression normalized to GAPDH (glyceraldehyde 3-phosphate dehydrogenase) and β-actin was determined by the \(2^{-\Delta\Delta CT}\) method using Microsoft Excel for Mac 2011 [27]. GAPDH was amplified using primers GAPDH_cDNA_fwd (5´-CCACTCCCAACGTGTCTGT-3´) and GAPDH_cDNA_rev (5´-GCTTCACCACCTTCTTCTCATC-3´) [28]. β-actin was amplified using primers ACTB_cDNA_fwd (5´-GTCATCCACCACGGCAATGAG-3´) and
Interdigital hyperplasia in Holstein cattle is associated with a missense mutation in the signal peptide region of the tyrosine-protein kinase transmembrane receptor gene

ACTB_cDNA_rev (5’-AATGCCGCAGGATTCCATG-3’) [29]. Relative ROR2 expression (ΔCT) was analyzed in 10 tissue samples of a healthy Holstein cow using GAPDH as normalizer.

Western blot analysis

Protein samples (10 μl) were mixed with 4X Bolt LDS Sample Buffer (Novex, ThermoFisher Scientific, Dreieich, Germany) supplemented with 10% 2-mercaptoethanol, and incubated at 70°C for 10 min. Electrophoresis was performed on 8% Bolt Bis-Tris Plus Gels (Novex, Thermo Fisher Scientific) in Bolt MES SDS Running Buffer (Novex, ThermoFisher Scientific, Dreieich, Germany) at 165 V for 57 mins. Proteins were transferred to nitrocellulose membrane (ThermoFisher Scientific, Dreieich, Germany) at 15 V for 1 h in transfer buffer. After blocking overnight at 4°C, immunoblots were incubated with primary anti-ROR2 antibody (1:500, ABIN2706970, Cohesion Biosciences, Aachen, Germany) and anti-β-actin (1:10,000, A5441, Sigma Aldrich, Darmstadt, Germany) at room temperature for 1 h. Incubation with the secondary antibodies (1:5000 for #1706515, 1:10,000 for #1706516, Bio-Rad, Munich, Germany) was done at room temperature for 1 h. Immunoblots were developed with Western ECL (GERPN2109, Sigma Aldrich, Darmstadt, Germany). Images from Western blots were captured and quantification was performed with ImageJ software [30].

Results

Identification of an IH associated region on bovine chromosome 8 (BTA8) using a genome-wide association study

During V1 and V2 a total of 110 cattle was clinically inspected (Fig. 1) and blood samples drawn (Table 1). After removing duplicates an initial set of 91 samples remained. After quality filtering of genotyping data as described in Materials and Methods 70 cattle were used for the genome-wide association study. Depending on the type of IH 41 cases/29 controls (type A) or 29 cases/41 controls (type B) were analyzed using an additive and/or dominant genetic model. For both IH types and genetic models markers on BTA8, i.e. ARS-BFGL-NGS-64395, ARS-BFGL-NGS-69582, showed highest -log10P values above a chromosome-wide significance threshold of -log10P = 5.47 (p = 0.05) (Fig. 2A, B). ARS-BFGL-NGS-64395 is located in intron 5 and ARS-BFGL-NGS-69582 downstream of the ROR2 gene (tyrosine-protein kinase transmembrane receptor) (Fig. 2B). Genes flanking ROR2, i.e. SPTLC1 (long-chain base subunit 1 of serine palmitoyltransferase), NFIL3 (interleukin 3-regulated nuclear factor) and AUH (3-methylglutacetyl-CoA hydratase) were excluded as potential candidates as they have been reported in humans or mice to be causative for hereditary sensory and autonomic neuropathy (type 1A), susceptibility to inflammatory bowel disease and 3-methylglutaconic aciduria (type 1), respectively [31-36]. On the other hand, as ROR2 had been associated with terminal limb malformations in humans including cutaneous syndactyly, it seemed to be a reasonable candidate for further analysis [37-44].

To identify IH associated variants, ROR2 was comparatively sequenced using 8 IH affected and 8 healthy control cattle. DNA sequence comparison revealed two variants in the coding region of ROR2 in exon 1 (NC_037335.1g.85905534T>A; rs377953295; ARS-UCD1.2) and exon 9
Interdigital hyperplasia in Holstein cattle is associated with a missense mutation in the signal peptide region of the tyrosine-protein kinase transmembrane receptor gene (NC_037335.1.g.86140379A>G; rs43572154; ARS-UCD1.2). The variant in exon 1 resulted in an amino acid exchange from tryptophan to arginin (ENSBTAP00000053765.2:p.Trp9Arg) located in the N-terminal signal peptide. The variant in exon 9 caused an amino acid exchange from methionine to valine (ENSBTAP00000053765.2:p.Met901Val) at the C-terminal intracellular end outside of any functional domain [45].

Fig. 1 Clinic signs of bovine interdigital hyperplasia

During hoof trimming cattle were clinically inspected. From all affected animals pictures were taken for documentation. A)-D) show different developmental stages with increasing size of dermal hyperplasia found in the interdigital space with skin final skin lesions (D). D) Interdigital hyperplasia after hoof trimming.
Interdigital hyperplasia in Holstein cattle is associated with a missense mutation in the signal peptide region of the tyrosine-protein kinase transmembrane receptor gene.

Fig. 2 Manhattan plot of genome-wide association study for bovine interdigital hyperplasia

A) Marker associations are plotted as negative log-transformed $P$ values against the position in the bovine genome (UMD3.1.1). Two markers ARS-BFGL-NGS-64395 and ARS-BFGL-NGS-69582 exceed the genome-wide significance threshold of $-\log_{10} P = 5.47$ (black line).

B) Enlargement of the associated chromosomal region on BTA8 flanking markers ARS-BFGL-NGS-64395 and ARS-BFGL-NGS-69582. Genes located within this region, *i.e.* $SPTLC1$, $ROR2$, $NFIL3$ and $AUH$, are shown with black bars at their approximate positions.
Interdigital hyperplasia in Holstein cattle is associated with a missense mutation in the signal peptide region of the tyrosine-protein kinase transmembrane receptor gene

Genotyping and association analysis of ROR2 variants

To determine the genotype frequencies of rs377953295 (exon 1) and rs43572154 (exon 9) in IH affected (IHA) and control cattle (IHF) both variants were genotyped in all clinically inspected cattle. Results are summarized in Table 2. Although for both SNPs the analyzed cohorts were in Hardy-Weinberg equilibrium, it was striking that the A_A-genotype (rs377953295) was not detected in healthy individuals. Fisher’s exact test using a 2x3 contingency table showed that the genotype distribution for this SNP was significantly deviating from expectation (p = 4.9e-4) whereas the distribution of genotypes for rs43572154 (exon 9) did not deviate. SNP rs43572154 (exon 9) was therefore excluded as potential causative variant. Association of rs377953295 (exon 1) was further investigated comparing the frequency distribution of the three genotypes with the different clinical IH types in a total of 94 cattle. As shown in Table 3 the presence of the A_A-genotype significantly associated with the different clinical forms of IH. Differences between type A and B IH are detectable in the distribution of genotypes T_T and A_T. Furthermore, it is evident that heterozygous individuals are more often affected which suggests that the A-allele at rs377953295 (exon 1) seems to be associated with an increased risk to develop IH.

In summary, the genome-wide association study identified a significantly associated chromosomal region on BTA8 harboring a potential candidate gene with a missense variant that is significantly overrepresented in IH affected cattle. As the variant rs377953295 (exon 1) results in an amino acid exchange at position nine of the ROR2 signal peptide it was hypothesized that this could influence expression and finally an insufficient translocation of ROR2 into the plasma membrane.

Table 2 Genotype frequencies of ROR2 variants rs377953295 (exon 1) and rs43572154 (exon 9) in type A affected and IH free (= healthy) cattle

<table>
<thead>
<tr>
<th></th>
<th>rs377953295 (exon 1)</th>
<th>rs43572154 (exon 9)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T_T</td>
<td>A_T</td>
</tr>
<tr>
<td>IHF&lt;sup&gt;a)&lt;/sup&gt;</td>
<td>32</td>
<td>5</td>
</tr>
<tr>
<td>IHA&lt;sup&gt;b)&lt;/sup&gt;</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>Total</td>
<td>55</td>
<td>28</td>
</tr>
<tr>
<td>HWE&lt;sup&gt;c)&lt;/sup&gt; (χ²)</td>
<td>1.53</td>
<td></td>
</tr>
</tbody>
</table>

a) IHF: Interdigital hyperplasia free; b) IHA: Interdigital hyperplasia affected; c) HWE: Hardy-Weinberg equilibrium of total cohort.
Interdigital hyperplasia in Holstein cattle is associated with a missense mutation in the signal peptide region of the tyrosine-protein kinase transmembrane receptor gene.

Table 3 Statistical evaluation of SNP rs377953295 (exon 1) as causative variant for type A and type B interdigital hyperplasia

<table>
<thead>
<tr>
<th></th>
<th>Type A IH&lt;sup&gt;a)&lt;/sup&gt;</th>
<th>Type B IH&lt;sup&gt;b)&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T_T</td>
<td>A_T</td>
</tr>
<tr>
<td>IHF&lt;sup&gt;c)&lt;/sup&gt;</td>
<td>21</td>
<td>4</td>
</tr>
<tr>
<td>IHA&lt;sup&gt;d)&lt;/sup&gt;</td>
<td>38</td>
<td>24</td>
</tr>
<tr>
<td>Total</td>
<td>59</td>
<td>28</td>
</tr>
<tr>
<td>F-statistic</td>
<td>7.16</td>
<td></td>
</tr>
<tr>
<td>P-values&lt;sup&gt;f)&lt;/sup&gt; (χ², FET&lt;sup&gt;e)&lt;/sup&gt;</td>
<td>0.0279, 0.0026</td>
<td>0.0002, &lt; 0.0001</td>
</tr>
</tbody>
</table>

<sup>a)</sup> In type A IH: IHF = no IH, IHA = at least one IH at one hind leg; <sup>b)</sup> In type B IH: IHF = no IH or only one IH, IHA = IH at both hind legs; <sup>c)</sup> IHF: Interdigital hyperplasia free; <sup>d)</sup> IHA: Interdigital hyperplasia affected; <sup>e)</sup> FET: Fisher’s exact test.

Expression of ROR2 in the interdigital skin is significantly downregulated in affected homozygous animals harboring the A-allele at rs377953295

ROR2 is a classical housekeeping gene harboring a GC-rich and TATA-less promoter [46]. Therefore, it was expected that ROR2 is widely expressed as shown in humans [47]. However, as ROR2 has also been shown to be mainly expressed in early and expression abates during later embryonal development, it was first of all important to demonstrate ROR2 expression in tissues of adult cattle [48, 49]. Therefore, 10 different tissues/organs, i.e. interdigital skin, lung, spleen, brain, heart, liver, stomach, intestine, urinary bladder and bone marrow, were collected at an abattoir and analyzed. As shown in Fig. 3A expression was detected at different levels in all analyzed samples. Importantly, ROR2 was expressed in the skin of the interdigital region further supporting the genomic data that it could be involved in the development of IH. In addition, it has been reported that two isoforms of ROR2 (ROR2-201, ROR2-202) exist differing in the usage of two alternative first exons. Only ROR2-201 harbors the exon with variant rs377953295 and therefore it was necessary to prove that this isoform was expressed in the interdigital skin. Using an isoform specific PCR it was possible to show that both isoforms are expressed in the interdigital skin (Fig. 3B). Finally, the association of the three genotypes of variant rs377953295 with the expression of ROR2 in the interdigital skin was tested. For this purpose samples of the hyperplastic interdigital skin of cattle harboring the T_T-, A_T- and A_A-genotypes at rs377953295 were compared using interdigital skin of healthy cattle as normalizer. As depicted in Fig. 4 the different genotypes within the IH affected samples showed significant reductions between genotype T_T vs A_A as well as A_T vs A_A. The difference between T_T vs A_T was not significant. The differences in ROR2 mRNA levels were also reflected in the ROR2 protein levels (Fig. 5). Note, that for the protein extraction of the T_T genotype only one hyperplastic skin sample was available and therefore a statistical verification was omitted. For reasons of animal welfare we refrained from additional sample collections. However, the ROR2 protein amount in the T_T sample was clearly above the other two genotypes.
Interdigital hyperplasia in Holstein cattle is associated with a missense mutation in the signal peptide region of the tyrosine-protein kinase transmembrane receptor gene.

In summary, the functional analyses are in agreement with the genomic data that a reduction of ROR2 expression on RNA and protein level seems to be associated with the risk to develop IH and is highest in cattle with the A_A genotype at rs377953295.
Interdigital hyperplasia in Holstein cattle is associated with a missense mutation in the signal peptide region of the tyrosine-protein kinase transmembrane receptor gene.

Fig. 3 Detection of ROR2 transcripts and isoforms in different bovine organs and tissues

A) Organ and/or tissue samples (De = interdigital dermis, HE = heart, Br = brain, Li = liver, Int = intestine, Ur = urinary bladder, Bo = bone marrow, Lu = lung, Sp = spleen, St = stomach) were collected from healthy cattle at an abattoir and RNA extracted. ROR2 transcripts were amplified as described in Material and Methods. Relative expression levels were calculated using the average ΔCt-values of three biological replicates with GAPDH as normalizer.

B) ROR2 isoform specific primers (see Materials and Methods) were used to amplify RNA from IntD = interdigital dermis, Lu = lung and Sp = spleen (NTC = non template control). Two amplicons corresponding to isoform ROR2-202 and isoforms ROR2-201/202 together can be seen. Amplicons were separated on a 2% agarose gel and visualized using Ethidium bromide staining.
Interdigital hyperplasia in Holstein cattle is associated with a missense mutation in the signal peptide region of the tyrosine-protein kinase transmembrane receptor gene.

Fig. 4 Comparison of ROR2 expression in hyperplastic interdigital skin tissue

Interdigital skin tissues were collected from 3 A_A, 3 A_T, 2 T_T cattle during clinical inspection at the farm using fine needle biopsies. Genotypes A_A, A_T and T_T correspond to the missense variant in exon 1 (rs377953295). Two samples were taken from healthy control cattle at an abattoir as reference. RNA was extracted and expression of ROR2 and internal control was done by real-time quantitative PCR using primers listed in Materials and Methods. Calculation of ROR2 expression fold change in the IH affected cattle was done using the $2^{\Delta\Delta Ct}$ method with β-actin as normalizer and the average expression of ROR2 in the healthy controls as reference [50]. ROR2 expression fold change of the three genotypes is shown as box-and-whisker plot. Significance was calculated using a one-tailed t-test with $p < 0.05$. 

\[ p = 0.0057 \]
\[ \text{ns} \]
\[ p = 0.045 \]
Interdigital hyperplasia in Holstein cattle is associated with a missense mutation in the signal peptide region of the tyrosine-protein kinase transmembrane receptor gene

Fig. 5 Quantification of ROR2 protein in hyperplastic interdigital skin tissue

Interdigital skin tissues were collected from 3 A_A, 3 A_T, 1 T_T cattle during clinical inspection at the farm using fine needle biopsies (see also Fig. 4). Genotypes A_A, A_T and T_T correspond to the missense variant in exon 1 (rs377953295). Two samples were taken from healthy control cattle at an abattoir as reference. Protein extracts were separated on 8% Bolt Bis-Tris Plus gels and transferred to nitrocellulose membranes. Membranes were incubated with primary anti-ROR2 antibody and anti-β-actin. Blots were developed with Western ECL. Images were captured and intensity quantification was performed with ImageJ software [30]. Significance was calculated using a one-tailed t-test with p < 0.05. As only one samples for genotype T_T were available statistical significance is not depicted.
Interdigital hyperplasia in Holstein cattle is associated with a missense mutation in the signal peptide region of the tyrosine-protein kinase transmembrane receptor gene

**Discussion**

Interdigital hyperplasia is a serious health issue in cattle production. Besides the clinical and animal welfare relevance it also has tremendous effects on the general performance of the animal. Therefore, it is of great importance to elucidate the molecular genetics of IH and find potential causative variants that could be used for selection in future breeding programs.

In the study described here a herd with a high prevalence of IH was identified which was followed over a longer period of time allowing a repeated thorough clinical inspection and sampling during professional hoof trimming. A GWAS was performed showing two genome-wide significantly associated SNP loci at around 87.5Mb on BTA8. In a previous study, 17 suggestive associations (p < 0.20) spreading across the bovine genome had been detected for IH in Holstein cattle [51]. Although five of them were also located on BTA8 (8Mb, 24Mb, 25Mb, 43Mb), the IH associated region in this study was located further downstream. Fortunately, only four genes were located in the direct chromosomal proximity of the associated SNPs. Three of the genes, *i.e.* SPTLC1, NFIL3 and AUH, were excluded from further analysis because they had been reported to be causative for other disorders with rather different detrimental phenotypes in humans and mice [31-36]. On the other hand, the remaining gene, *i.e.* ROR2, seemed to be the most promising positional candidate due to the fact that one of the associated SNPs was directly located within the gene (intron 5). A DNA sequence comparison of ROR2 in IH affected and healthy cattle revealed two missense variants in exon 1 and exon 9, respectively. However, only the variant in exon 1 (rs377953295) resulting in an amino acid exchange in the signal peptide of ROR2 (ENSBTAP00000053765.2:p.Trp9Arg) remained significantly associated with IH after screening the cohort of cattle at the farm. So far ROR2 variants have not been associated with any other disorder or trait in cattle and therefore the identification of the missense variant was a novel finding.

But also regarding its biological function, ROR2 was obviously an interesting candidate. ROR2 (Receptor Tyrosine Kinase-like Orphan Receptor 2) belongs to the receptor tyrosine kinases (RTK), a large superfamily of transmembrane glycoproteins. Previous studies have shown that ROR2 is important for the formation of the distal limbs [52, 53]. Molecular genetic analyses in humans have revealed that mutations in ROR2 cause dominant Brachydactyly type B and recessive Robinow Syndrome, with terminal limb malformations as common symptoms [37-44, 54]. Defects of the distal limbs have also been observed in ROR2<sup>−/−</sup> mice, mainly due to abnormal cartilage and growth plate development as well as ossification [53, 55]. Hence, the reported functional role, determined positional association and identification of a missense variant in ROR2 in IH affected cattle were convincing enough to extend the analyses.

Therefore, ROR2 expression was analyzed in the interdigital skin. Although earlier studies indicated that ROR2 expression in adults was restricted to parathyroid, testis and uterus, it is now known that its expression may be more widespread than originally thought [47, 56]. During embryogenesis ROR2 expression was identified in heart, lungs, face and limbs [52]. Our data show that different amounts of ROR2 mRNA can be detected in interdigital skin, lung, spleen, brain, heart, liver, stomach, intestine, urinary bladder and bone marrow of adult cattle. In addition, two ROR2 isoforms (ROR2-201, ROR2-202) were detected differing in the usage of an alternative first exon. Especially isoform ROR2-201 harboring the exon with the missense variant was detectable in the interdigital skin. Although compared to other organs and/or tissues
expression of ROR2 in the interdigital skin was intermediate, the identification of ROR2 transcripts was important to support its assumed role in IH development. Furthermore, the relative amount of ROR2 transcripts significantly differed between the three genotypes in correlation with the IH status. Homozygous A_A IH affected cattle showed significantly reduced ROR2 expression in the hyperplastic interdigital skin which is also reflected in the amount of ROR2 protein. In addition, ROR2 expression in hyperplastic compared to normal interdigital skin showed a significant reduction which explains the observation that hyperplastic skin alterations were also present in T_T or A_T cattle. The down-regulation of ROR2 could be explained by a general mechanism of transcriptional quality control. For instance, when mutant signal sequences fail to bind to the signal recognition particle at the ribosome exit site, the nascent chain instead contacts Argonaute2 and the mutant mRNAs are specifically degraded \[57\]. Specific mRNA degradation preemptively regulates aberrant protein production \[57\]. However, other mechanisms resulting in the general down-regulation of ROR2 in hyperplastic interdigital skin have to be taken into account \[58, 59\].

In the etiology of IH, the role of ROR2 in signal transduction seems to be important. The main histological alterations in hyperplastic interdigital skin are proliferating fibroblasts, multiplex papilliferous epidermal ridges as well as increased cellularity in the stratum granulosum and stratum spinosum \[60, 61\]. In this respect, it is important to note that Wnt signaling pathways play crucial roles in the regulation of skin development and epidermal stem cells behavior \[62, 63\]. Sustained epidermal activation of Wnt/β-catenin signaling not only stimulates fibroblast proliferation, but also causes structural remodeling of the entire dermis \[64\]. Binding of Wnt isoforms to Fz-LRP complex (frizzled-low density lipoprotein receptor-related protein) generates β-catenin signaling, whereas binding to the atypical receptor tyrosine kinase ROR2 can inhibit this activity \[65\]. The significant down-regulation of ROR2 could affect the inhibition of the canonical Wnt pathway, resulting in the abnormal cellular processes related to aberrant epidermal development. This potential biochemical effect is consistent with our finding that ROR2 seems to be an IH suppressor and that decreased expression level of ROR2 leads to IH development. Noteworthy, disease suppressive relationships of ROR2 have been described in hepatocellular cancer, colon cancer and hematological malignancies \[58, 59, 66, 67\]. The canonical Wnt pathway has a pro-tumorigenic effect, leading to a series of cellular processes including proliferation, differentiation, polarity, migration, invasion, adhesion and survival \[68, 69\]. Through inhibiting this canonical Wnt signaling as a gatekeeper, ROR2 has been proposed to play a role in tumor suppression \[68\]. However, further studies will be required to clarify the exact molecular mechanism caused by down-regulation of ROR2 expression in IH development.

**Acknowledgements**

The authors are grateful to M. Hirschfeld, D. Niehaus, S. Bierau, S. Balzer, L. Jüttnner, M. Scharfenstein and S. Pach for expert technical assistance. The Association for Bioeconomy Research (Bonn), German Livestock Association (E. Feddersen) and VIT (F. Reinhardt) are thanked for support. X. Zhang is a fellow of the Chinese Scholarship Council (CSC). We acknowledge support by the German Research Foundation and the Open Access Publications of the University of Goettingen.
Interdigital hyperplasia in Holstein cattle is associated with a missense mutation in the signal peptide region of the tyrosine-protein kinase transmembrane receptor gene

References

Interdigital hyperplasia in Holstein cattle is associated with a missense mutation in the signal peptide region of the tyrosine-protein kinase transmembrane receptor gene

15. Swalve HH, Alkhoder H, Pijl R, editors. Genetic background of disorders of the bovine hoof from data collected at hoof trimming. 8th Conference on Lameness in Ruminants; 2011 28.02.-03.03.2011; New Zealand.
17. Pijl R, Swalve HH, editors. Incidence Rates For Claw Disorders In Relation To Status Of Female Ancestors Based On Assessments At Routine Hoof Trimming In German Holstein Dairy Cattle. Proceedings of the 19th International Symposium and 11th Conference Lameness in Ruminants; 2017 07.-09.09-2017; Munich, Germany.
Interdigital hyperplasia in Holstein cattle is associated with a missense mutation in the signal peptide region of the tyrosine-protein kinase transmembrane receptor gene


Interdigital hyperplasia in Holstein cattle is associated with a missense mutation in the signal peptide region of the tyrosine-protein kinase transmembrane receptor gene.
Interdigital hyperplasia in Holstein cattle is associated with a missense mutation in the signal peptide region of the tyrosine-protein kinase transmembrane receptor gene.

Interdigital hyperplasia in Holstein cattle is associated with a missense mutation in the signal peptide region of the tyrosine-protein kinase transmembrane receptor gene.

### Table S1 ROR2 primers for PCR and Sanger sequencing

<table>
<thead>
<tr>
<th>Primer name</th>
<th>5´- 3´</th>
<th>Product size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROR2_Ex1fwd</td>
<td>CGAAAGGGACCAACTTGCCA</td>
<td>282</td>
</tr>
<tr>
<td>ROR2_Ex1rev</td>
<td>CATCGCGAGGGAGGGACG</td>
<td></td>
</tr>
<tr>
<td>ROR2_Ex2fwd</td>
<td>GGGCAGTTTTGCTAAACACTAT</td>
<td>300</td>
</tr>
<tr>
<td>ROR2_Ex2rev</td>
<td>GGGAGACCCCTGACCACATCAT</td>
<td></td>
</tr>
<tr>
<td>ROR2_Ex3fwd</td>
<td>GTGGTGAGGGCAACATTCTA</td>
<td>617</td>
</tr>
<tr>
<td>ROR2_Ex3rev</td>
<td>TCATGCACAATGGGAAAGGC</td>
<td></td>
</tr>
<tr>
<td>ROR2_Ex4fwd</td>
<td>CAGAGAGCACCCTTCCATC</td>
<td>370</td>
</tr>
<tr>
<td>ROR2_Ex4rev</td>
<td>TGCCATCCCTGTGTGAAGTC</td>
<td></td>
</tr>
<tr>
<td>ROR2_Ex5fwd</td>
<td>GCCAAGGGACAAGATGGCTA</td>
<td>330</td>
</tr>
<tr>
<td>ROR2_Ex5rev</td>
<td>GGGACAAAAATACACAAAATGAGACTG</td>
<td></td>
</tr>
<tr>
<td>ROR2_Ex6fwd</td>
<td>CAGTTGCAAATCTGGGCGG</td>
<td>668</td>
</tr>
<tr>
<td>ROR2_Ex6rev</td>
<td>GAATGGAGCAGGGTCTGTG</td>
<td></td>
</tr>
<tr>
<td>ROR2_Ex7fwd</td>
<td>TGGCGAGGTTGTTGTTAT</td>
<td>479</td>
</tr>
<tr>
<td>ROR2_Ex7rev</td>
<td>ATGGTGGTGAACACGGTG</td>
<td></td>
</tr>
<tr>
<td>ROR2_Ex8fwd</td>
<td>AGTTTGAGGTGGAGTG</td>
<td>346</td>
</tr>
<tr>
<td>ROR2_Ex8rev</td>
<td>GTAGAGGTTAAGCTGGGG</td>
<td></td>
</tr>
<tr>
<td>ROR2_Ex9.1fwd</td>
<td>CGGCCCAACCCCTTCTCC</td>
<td>552</td>
</tr>
<tr>
<td>ROR2_Ex9.1rev</td>
<td>GCCCAGGTCTTTGTGGAACA</td>
<td></td>
</tr>
<tr>
<td>ROR2_Ex9.2fwd</td>
<td>GGCATGGAGTACCTGTCAG</td>
<td>965</td>
</tr>
<tr>
<td>ROR2_Ex9.2rev</td>
<td>GTCTCAGGGACTGAGCGG</td>
<td></td>
</tr>
<tr>
<td>ROR2_Ex9.3fwd</td>
<td>CTCCCATCACAGCGGCG</td>
<td>399</td>
</tr>
<tr>
<td>ROR2_Ex9.3rev</td>
<td>AGGTGGGCACACCGGCT</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 4

Processed pseudogene confounding the presence of a putative lethal recessive deletion in the bovine 60S ribosomal protein L11 gene \((uL5)\)
Processed pseudogene confounding the presence of a putative lethal recessive deletion in the bovine 60S ribosomal protein L11 gene (uL5)

Xuying Zhang, Christin Wacker, Ekkehard Schütz, Bertram Brenig*

University of Goettingen, Institute of Veterinary Medicine, Burckhardtweg 2, 37077 Göttingen, Germany

XZ: xuying.zhang@stud.uni-goettingen.de
CW: christin.wehrhahn@agr.uni-goettingen.de
ES: eschuet@gwdg.de
BB: bbrenig@gwdg.de

*Corresponding author:
Prof. Bertram Brenig
Institute of Veterinary Medicine
Burckhardtweg 2
D-37077 Göttingen
Germany

Phone +49-551-39-28383
Fax +49-551-39-33392
Email bbrenig@gwdg.de
Processed pseudogene confounding the presence of a putative lethal recessive deletion in the bovine 60S ribosomal protein L11 gene (uL5)

Summary

Ribosomal protein L11 is a member of the 80S core ribosomal proteins encoded by the 60S ribosomal protein L11 gene. Due to its role in ribosome biogenesis and that it has been associated with fatal ribosomopathies, we addressed the question whether lethal uL5 variants exist in cattle. Several deleterious variants have been deposited including a two base pair deletion (rs381576999) resulting in a frameshift and premature stop. A probe for this variant had been included as expert-selected marker in the custom add-on part of the BovineLD BeadChip and can therefore be monitored during routine genotyping. To determine the frequency of the variant, we genotyped 370,527 cattle, including 18 different dairy and beef cattle breeds. 299,218 homozygous wild type and 71,249 heterozygous cattle were called from the bead chip. Cattle harboring the homozygous mutant genotype were not detected. According to Hardy-Weinberg equilibrium 4,241 homozygous individuals carrying the deletion should have been present in the cohort. Hence, the data apparently indicated that the homozygous 2 bp deletion in uL5 is a lethal variant interfering with ribosome biogenesis. An attempt to verify the deletion in randomly chosen cattle with heterozygous calls using Sanger sequencing failed to detect the putative deletion. An alignment of the BeadChip probe to the bovine genome showed a perfect match to a processed uL5 pseudogene on bovine chromosome 18 mimicking the assumed exonic deletion. The analysis shows how processed pseudogenes may confound quantitative results depending on the placement of probes in a particular assay design resulting in false results.

Key words: Ribosome biogenesis, ribosomal protein, RPL11, uL5, processed pseudogene, ribosomopathy, Diamond-Blackfan anemia, embryonic lethal

Ribosomes are large RNA-protein complexes that catalyze protein synthesis either linked to the endoplasmic reticulum or in the cytosol. Eukaryotic ribosomes consist of a large 60S (LSU) and a small 40S (SSU) subunit. For protein synthesis both subunits are assembled into the final 80S ribosome containing more than 5500 nucleotides of rRNA, 80 core ribosomal proteins (RP) and more than 150 associated proteins in higher eukaryotes [1, 2]. Several external and genetic factors result in ribosomal stress and impairment of ribosome biogenesis leading to accumulation of ribosome-free forms of RPs [3]. Impaired ribosome biogenesis and function can result in specific clinical phenotypes the so-called ribosomopathies [2]. In humans at least six ribosomopathies have been described, i.e. Diamond-Blackfan anemia (DBA), 5q-syndrome, Shwachman-Diamond syndrome, X-linked dyskeratosis congenita, Treacher Collins syndrome and cartilage hair hypoplasia, most of them also associated with an increased cancer risk [2, 4]. DBA belongs to a rare group of disorders known as inherited bone marrow failure syndromes [5]. It is a genetically heterogenous disorder caused by variants in 20 different ribosomal genes resulting in distinct clinical phenotypes (DBA1-20) [2]. In humans, DBA7 is caused by variants of the 60S ribosomal protein L11 gene (uL5, formerly RPL11) and is inherited mainly autosomally although autosomal recessive inheritance has also been reported [6-8]. Patients with uL5 deficiency show cleft lip and/or palate and isolated thumb abnormalities [2, 9, 10]. So far naturally occurring uL5 defects have only been described in humans. An inducible Rpl11-null allele has been introduced in mice demonstrating that a single gene dose is
insufficient to support embryonic development. Homozygous deletion of Rpl11 is lethal within a few weeks due to bone marrow aplasia and intestinal atrophy [11]. Rpl11 deficiency has also been studied in a zebrafish model for DBA [12, 13].

Despite the obvious importance of ribosome biogenesis for viability, ribosomopathies have not been described in cattle so far. Therefore, we were interested whether lethal variants exist in bovine ribosomal protein genes. Similar to the identification of the lethal haplotypes in several important cattle breeds, we hypothesized that the complete absence or statistically significant underrepresentation of a homozygous mutant genotype in the population is indicative for a lethal defect [14-17]. Depending on the genotype frequency of homozygous wild type and healthy heterozygous carriers this defect will actually be embryonic lethal. Due to the severe effects of \( uL5 \) mutations in humans and lethal effect in mice we decided to analyze bovine \( uL5 \).

In exon 3/4 of the bovine \( uL5 \) gene located on chromosome 2 a deletion of two base pairs has been deposited into bovine genomic databases, which would result in a frameshift and a premature stop codon (ENSBTAG00000020905:g.129,195,922_129,195,923del; ARS-UCD1.2; rs381576999). The deletion causes a truncation of \( uL5 \). To monitor this variant a probe located on BTA2 between position 129,195,924 to 129,195,973 (ARS-UCD1.2) was included on the BovineLD BeadChip (Illumina) (Fig. 1). High-throughput genotyping was performed on a HiScanSQ and/or iScan System (Illumina). Raw data were converted using GenomeStudio 2.0.4 Software (Illumina). Cattle blood samples were drawn by local veterinarians. The collection of samples was approved by the Lower Saxony State Office for Consumer Protection and Food Safety (33.19-42502-05-17A196) according to §8a Abs. 1 Nr. 2 of the German Animal Protection Law.

The presence of the deletion was validated in a total of 370,527 cattle. 299,218 homozygous wild type and 71,249 heterozygous cattle were detected. No homozygous carriers were identified resulting in highly significant deviation from HWE (\( \chi^2 = 4,193.4 \)) [18]. The observed cluster separation and GC-scores of the \( uL5 \) SNP were sufficiently high to exclude any technical bias by the chip-based genotyping method as confounding factor. If the deletion would not have a lethal effect approx. 4,241 homozygous carriers would have been expected according to HWE. In contrast to the DBA mouse model heterozygous carriers seemed to be healthy, as we did not have any reports on phenotypical abnormalities in cattle of the heterozygous cohort. As the deletion was detected in at least 18 different dairy and beef cattle breeds that were within the genotyped cohort, \textit{i.e.} British Angus, Charolais, Braunvieh, Belted Galloway, Simmental, Dexter, German Black Pied cattle, Gelbvieh, Hereford, Limousin, Red Holstein, German Red cattle, Holstein, Shorthorn, Uckermärker, Wagyu, Welsh Black cattle and Belgian Blue, it must have been introduced into the cattle population early in domestication. In respect to the deduced rather high allele frequency of 0.11, it was quite astonishing that abnormal cattle have never been noticed so far. We therefore decided to verify the deletion by Sanger sequencing in cattle that had been genotyped as being heterozygous. For PCR and Sanger sequencing primers were designed based on the bovine reference sequence ARS-UCD1.2 using Primer3 [19] to specifically amplify the genomic region of rs381576999. PCR was performed in a total volume of 25 μl, including 1× PCR reaction buffer plus 20 mM MgCl2, 0.5 μl of 10 mM dNTPs, 1 U FastStart Taq Polymerase (Roche), 1 μl of 10 pmol/μl each primer (\( uL5\text{-Ex4}_\text{for}: 5´-AGAGCCGGAGTGGGAGACTCGGT-3´; \) \( uL5\text{-Ex4}_\text{rev}: 5´-TGGGGGTCTTGAGTGTTGCAGGCT-3´ \), and 1μl of 20 ng/μl DNA. Reaction conditions
Processed pseudogene confounding the presence of a putative lethal recessive deletion in the bovine 60S ribosomal protein L11 gene (uL5)

included 95 °C for 10 min, then 30 cycles of 95 °C for 30 s, primer-specific annealing temperatures for 30 s and 72 °C for 45 s, followed by final elongation at 72 °C for 7 min. The purification of the PCR products was performed using Rapid PCR Cleanup Enzyme Set (New England Biolabs GmbH, Germany). Subsequently sequencing was realized using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Thermo Fisher Scientific GmbH, Germany) and products were analyzed on an ABI PRISM 3130xl Genetic Analyzer (Life Technologies, USA) following the manufacturer’s protocols. Sequencing primers were identical to PCR primers. Identification of nucleotide polymorphisms via sequence alignments was realized using Sequencher 5.4.6 (Gene Codes, USA). None of the randomly selected heterozygous cattle identified by genotyping using the BeadChip were proved to be carriers of the putative deletion (Fig. 1). An alignment of the BeadChip probe to the bovine genome revealed an almost perfect match (49/50 nucleotides) to a processed uL5 pseudogene on BTA18 (LOC112442347; position 54,982,088 to 54,982,687) harboring the reported 2 bp deletion in uL5. We therefore designed amplification primers for the putative binding region of the probe on the pseudogene and sequenced the previously detected heterozygous cattle. As shown in Fig. 1 the analyzed cattle were homozygous for the 2 bp deletion within the pseudogene. The genotyping results were clearly confounded by the erroneous detection of this variant. Our data demonstrate how pseudogenes can lead to severe misinterpretations when using high throughput genotyping platforms without knowing the assay design. Therefore, special care should be taken when designing probes and/or primers for a specific assay to accomplish accurate results that can be used in breeding programs.
Processed pseudogene confounding the presence of a putative lethal recessive deletion in the bovine 60S ribosomal protein L11 gene (*uL5*)

Fig. 1 Alignment and DNA sequencing of a putative 2 bp deletion in *uL5* and comparison with a processed *uL5* pseudogene (LOC112442327)

The figure depicts the alignment of *uL5* (BTA2) with the BovineLD BeadChip probe and the processed *uL5* pseudogene (LOC112442327) (BTA18) between the indicated positions (ARS-UCD1.2). The putative 2 bp deletion (rs381576999) is underlined and bold. The upper and lower chromatograms show the local sequences flanking rs381576999 (*uL5*) and the homologous region on BTA18 of the same individual. At both locations the individual is homozygous wild type resulting in an erroneous heterozygous genotype at *uL5*.

Acknowledgements

D. Niehaus is thanked for expert technical assistance. The authors are also grateful to the Association for Bioeconomy Research (FBF), the German Livestock Association BRS (E. Feddersen), VIT (F. Reinhardt), IFN Schönow e. V. (M. Jung) for support. X. Zhang is a fellow of the China Scholarship Council (CSC).

References


CHAPTER 5
General Discussion
In this thesis analyses were performed to elucidate potential pathological genetic aberrations of osteogenesis imperfecta and interdigital hyperplasia in HF cattle, as well as ribosomopathies in 18 different dairy and beef cattle breeds.

1 Significance of the research study

The first question underlying this dissertation can be stated as “What is the genetic cause of the lethal case of OI in HF?” In humans the OI-related molecular mechanism has been clearly explored (see General Introduction 1.3.1). Cases of OI in HF cattle have also been reported since 1983 [1, 2]. The identification of their causative mutations would be of great importance for cattle healthiness and breeding improvement. However, it has been challenging in the past decades to elucidate the causative mutation of OI in HF, which might imply a more complicated OI genetic cause in cattle than in humans.

The second question in this dissertation comprises “What is the molecular cause of IH in HF?” and “Does a major-effect gene exist in IH development?” Claw disorders are recognized as an important animal welfare problem at dairy farms causing serious economic losses. Hoof health has been included as an important trait in dairy cattle breeding. The recently updated breeding value evaluation for Holsteins covers the whole range of important traits including mastitis resistance, reproduction, metabolic stability, health as a total trait as well as hoof health [3]. IH as a common foot disease shows a rather high estimated heritability (see General Introduction 2.2.4). Elucidation of the potential genetic cause of IH risk is of great importance contributing to future practical breeding for better hoof health.

The third question in this dissertation was “Do lethal variants exist in the causal genes of ribosomopathies in cattle?”. Ribosome biogenesis plays fundamental roles in cellular development. Its functional impairment contributes to embryonic lethality and tumorigenesis [4, 5]. In humans a list of ribosomopathies have been described, with the causative mutations reported (see General Introduction 3.2). The identification of developmentally lethal recessive variants in these causal genes has potential significance to reduce the incidence rate of embryonic loss/lethality in cattle breeding.

2 Evolutionary genetic dissection technologies

The rapid development of technologies, e.g. DNA chip, NGS, advanced statistical models, has not only driven discoveries in monogenic disorders but also accelerated genetic dissections of complex traits [6]. However, every technology carries advantages and disadvantages. It is of importance to know their features, limitations and the fitness for their potential applications.

2.1 Genome-wide association study

GWAS is a powerful and popular tool for the dissection of genetic diseases. It provides a hypothesis-free way that systematically tests hundreds of thousands or even more variants in the whole genome without priorly knowing the position of the causal gene/variant [7]. To date, thousands of loci for common diseases have been identified by GWAS. However, whether
GWAS can produce meaningful results of positive associations depending on a plethora of parameters, e.g. selection of cases/controls, phenotype, and available sample size [6]. In Chapter 3 of this thesis, a herd with extreme prevalence of IH was genotyped using the BovineSNP50 BeadChip and subsequent GWAS successfully identified a significantly IH associated chromosomal region. In Chapter 2 of this thesis, GWAS detected six significantly associated SNP loci, implying a more complicated molecular background of OI in Hostein cattle than the monogenic OI in humans. In addition, GWAS is designed to detect small variants like SNPs and small InDels, but it is difficult to identify an associated region, if the causative mutation is a large structural variant. Besides, BeadChip can be used for high throughput genotyping purpose. In Chapter 4 of this thesis, our data demonstrate how pseudogenes can lead to severe misinterpretations when using high throughput genotyping platforms without knowing the assay design in the BeadChip. Therefore, it is very important to take special care when designing probes and/or primers for a specific assay to accomplish accurate genotyping results. Compared to custom BeadChips, a sequence-called genotyping assay would be more reliable to accomplish accurate results.

2.2 NGS-based analysis

Efficiency of DNA sequencing has developed tremendously in recent years [8]. NGS technology with massively parallel processing can be applied flexibly for whole genome sequencing from small genomes like viruses or bacteria to complex genomes like animals and plants. Its application ranges from exome and metagenome, to transcriptome, also to analysis of DNA-protein-interactions and methylation sequencing. On a genomic level, mutations can be divided into two major categories, single nucleotide polymorphisms (SNPs) and structural variants (SVs).

2.2.1 Single nucleotide polymorphisms

A comprehensive and efficient SNP-based analysis of NGS data contains three mapping strategies, i.e. rare recessive homozygous polymorphisms, de novo candidate and compound heterozygous polymorphisms, to identify causal genes/SNPs of monogenic disorders. Large-scale genotyping is crucial to validate potential causative SNPs. High-resolution melting curve analysis can be applied efficiently for large-scale studies including FRET, TaqMan assay, ARMS, RFLP and so on. For instance, in Chapter 2 of this thesis, PCR-RFLP, ARMS and multiplex-FRET-PCR were performed to validate the variants rs209568970 in IFITM5, rs209556962 in QRFPR, as well as rs381405831 and rs110593220 in ABCA13, respectively, in a large-scale cohort of healthy control animals. In Chapter 4, FRET was also designed to test the 2bp-deletion (rs381576999) in RPL11 in 370,527 cattle. These techniques were proven to be very time saving.

2.2.2 Structural variants

The combination of NGS and sophisticated computational tools have been developed to detect somatic SVs from the massive amount of raw sequencing data. Each computational method has its own unique characteristics, e.g. focusing on certain types of SVs, read mapping or clustering strategy, as well as the use of split reads and/or discordant read pairs [9]. In the OI study of this thesis, SV detection has been applied to NGS data using DELLY and SVDetect [10, 11]. Five
most basic SV types were detected, \textit{i.e.} deletion, insertion, duplication, inversion and translocation. However, validation by Sanger sequencing showed that the predictions were misled by repetitive elements from different chromosomes. Therefore, whether repetitive elements exist should be considered to interpret SV prediction results. In addition, the following aspects should be considered for accurate and complete detection of SV \cite{9}. Longer read-length sequence data with lower error rates can improve the capacity in SV detection. Uniform coverage across the genome is also important, since non-uniform coverage can cause substantial difficulties for SV detection in regions with low coverage. Diverse combinations of the same or different SV types lead to more complicated chromosomal rearrangement events, which can be hard to detect and validate. It is necessary and crucial to develop rules and standards to compare the results from different programs. Besides, dosage-variant DNA copy number variation, a subtype of SVs, also needs to be considered \cite{9}.

2.3 Functional effect validation of novel variants

When a potential causative mutation is identified, it is crucial to validate its functional effect. Functional experiments are also effective to exclude nonfunctional variants. To substantiate the pathogenicity of a novel variant, both \textit{in vitro} and \textit{in vivo} experiments are robust approaches. Compared to \textit{in vivo} test in regard to expenditure of time, \textit{in vitro} characterization of a novel variant should be preferably undertaken to verify its impact on protein stability, structure, and function. Deleterious exonic variants could induce protein conformational changes, affecting protein folding, protein interactions with other proteins as well as stability or solubility of protein molecules. Exonic variants, especially variants in signal peptides, could also lead to changes in gene expression level and mRNA stability. Deleterious variants in transcriptional regulatory elements, \textit{i.e.} promoter, activator, enhancer and silencer, could result in altered gene expression level. Additionally, the production of animal models can be used. However, both methods are time-consuming, labor-intensive, technically difficult and expensive. Therefore, prior to functional validation, prediction of the potential functional effect of a novel variant is necessary using computational tools.

3 Cattle as an animal model to study claw disorders

Mice are popular as an animal model to study genetic disorders like osteogenesis imperfecta and ribosomopathies, because of their genomes similar to that of humans, low cost, little size, fast reproduction rate and ease of handling \cite{12}. Other common animal models include zebra fish, fruit flies, worms, and chicken \cite{12}. Claw disorders mainly cause detrimental impacts on large livestock. In additional to cattle, they have also been reported in buffalo, sheep and goat. Foot diseases would better studied in dairy cattle as the more suitable animal model, compared to these little size animals. Early in 1989, Harper et al. employed bovine models to study citrullinemia, a urea cycle genetic disorder characterized by hyperammonemia \cite{13, 14}. A distinct advantage of bovine models is that they manifest well the clinical signs of claw disorders, since livestock have the largest body size. Also, claw disorders are common in dairy cattle with a rather high prevalence, which easily brings them to the attention of veterinary and research communities. In addition, the availability of extensive recording of phenotype, well-developed NGS technologies and high-density SNP chips offers an unprecedented opportunity
to quickly dissect the genetic architecture in cattle [15]. Most importantly, body size, body weight, herd size, poor hygiene, rocky pastures etc. are crucial factors to trigger a genetic predisposition in IH development [16, 17]. While traditional lab-based animal models are likely too little and too light for this trigger, even carrying the variant with an increased risk of developing IH.
Conclusions and Outlook

From the results of the three subprojects on cattle genetic diseases (osteogenesis imperfect, interdigital hyperplasia and ribosomopathies), the following interpretations and conclusions can be derived for future investigation:

Study 1: Osteogenesis imperfecta in an embryo transfer Holstein calf

- OI has severe and even lethal impacts in humans and animals. The OI-related molecular mechanism in humans has been clearly explored, with a series of candidate genes reported. OI cases in HF cattle have been reported since 1983. However, no causative mutations have been identified to date.
- An oligogenic inheritance should also be considered for the OI development in HF.

Study 2: Interdigital hyperplasia in Holstein cattle is associated with a missense mutation in the signal peptide region of the tyrosine-protein kinase transmembrane receptor gene

- Interdigital hyperplasia very likely only causes harmful impacts on large livestock. Hoof health has been included as an important trait in dairy cattle breeding, from the perspectives of both economic and animal welfare.
- Genetically, interdigital hyperplasia is significantly associated with the missense mutation in the signal peptide region of ROR2 (NC_037335.1g.85905534T>A; rs377953295; ARS-UCD1.2). Genotyping and association analysis suggest that the A-allele at rs377953295 could be detrimental and associated with an increased risk to develop IH.
- This variant results in an amino acid exchange (ENSBTAP00000053765.2:p.Trp9Arg) in the N-terminal region of the ROR2 signal peptide which is necessary for proper topology of the polypeptide during translocation. Quantification of ROR2 mRNA and ROR2 protein shows that the variant results in a significant suppression of ROR2 expression in homozygous affected compared to wild type and carrier cows.
- ROR2 expression in hyperplastic compared to normal interdigital skin also shows a significant reduction which explains the observation that hyperplastic skin alterations are also present in T_T or A_T cattle.
- The biological function of ROR2 in the molecular mechanism of IH development is still unclear. The down-regulation of ROR2 in mutated cattle might be explained by a general mechanism of transcriptional quality control. Previous reports also gave some clues to what could cause the down-regulation of ROR2 in hyperplastic interdigital skin, especially in wild type cattle. Also, further studies will be required to clarify the molecular mechanism caused by the down-regulation of ROR2 expression in IH development, and Wnt signaling pathways might be a clue.
• Cattle should be considered as an animal model to study claw disorders, instead of other animal models.

Study 3: Processed pseudogene confounding the presence of a putative lethal recessive deletion in the bovine 60S ribosomal protein L11 gene (uL5)

• So far no lethal recessive variants of ribosomopathies have been detected in cattle. The putative lethal recessive deletion in the DBA causal gene RPL11, was confounded by the processed pseudogene and actually does not exist.

• By identifying haplotypes for which homozygotes are not present but would be expected based on genotype frequencies and HWE, developmentally lethal recessive variants can be determined.

• Compared to custom BeadChips, a sequence-called genotyping assay would be more reliable. Pseudogenes can lead to severe misinterpretations when using high throughput genotyping platforms without knowing the assay design.

• In practical breeding, through targeted matings with non-carriers, a potential identified detrimental allele should be excluded to mitigate/remove the defects or embryonic loss/lethality of ribosomopathies.
References


Acknowledgments

Completion of this doctoral dissertation was possible with the help and support of many people. I would like to express my deep and sincere gratitude to all of them.

First of all, I am extremely grateful to my supervisor Prof. Dr. Dr. Bertram Brenig. He offered me such a great opportunity to join the projects. He guided me during all the time of my research and this thesis writing with his immense knowledge, great patience and positive research attitude. His encouragement and continuous support was so important for me during the hard time of my study. I could not have imagined having a better supervisor for my Ph.D study. I am also deeply thankful to my co-supervisor Prof. Dr. Jens Tetens for his great support and invaluable advice. Many sincere thanks are due to my dissertation committee member Prof. Dr. Jürgen Hummel for his insightful comments, encouragement and help.

Some colleagues in the Institute have been very kind enough to extend their help at various phases of this research, I do hereby acknowledge all of them. I am deeply thankful to Dr. rer. nat. Marc Hirschfeld for his generous, effective, professional suggestions and help whenever I approached him. Many sincere thanks also go to Dr. Julia Beck and Christin Wacker, they supported me so much with their great experience and professional knowledge. I am thankful to Delia Niehaus, for the effective technical support, especially for the stressful period we were working together before deadlines. I also thank Fangzheng Xu and Shuwen Shan for their help in conducting Western blot. In addition, Sabrina Pach created a good experimental environment and also helped with reagent order.

I am also grateful to our project collaborator Prof. Dr. Hermann H. Swałe and his research team from Martin-Luther-University Halle-Wittenberg, without their precious support it would not be possible to conduct this research. Thanks are also due to Chinese Scholarship Council (CSC) for the financial support for my Ph.D study.

I am thankful to my friends, Guojing Luo, Nan Nan and Yang Xu, for the hard time under study pressure I could phone them and get great spiritual support from them, also for the nice travelling we once had together.
Last but not least, I would like to express my great gratitude to my family. My mother Haofeng Li, father Ensheng Zhang, brother Yubin Zhang, for their warm love, great patience, and endless support throughout my study time and my life in general.
Curriculum Vitae

Personal Details

Family Name: Zhang  
Given Name: Xuying  
Gender: Female  
Date of Birth: 23.11.1986  
Place of Birth: Yuncheng, P.R.China  
Marital Status: Single  
Parents' Name: Ensheng Zhang and Haofeng Li  
Nationality: Chinese

Education

2015 - Current  
Doctoral Studies  
Department of Molecular Biology of Livestock, Georg-August-Universität Göttingen, Germany

2011 - 2014  
Master Studies  
Department of Animal Genetics, Breeding and Reproduction, Huazhong Agricultural University, Wuhan, P.R.China  
Graduation Degree: Master of Agriculture

2007 - 2011  
Bachelor Studies  
Department of Biological Science on Animals, Shanxi Agricultural University, Taigu, P.R.China  
Graduation Degree: Bachelor of Science