Functional Genomics of Bone Metabolism

Novel Candidate Genes Identified by Studies in Chicken Models

CARL-JOHAN RUBIN
Dissertation presented at Uppsala University to be publicly examined in Enghoffsalen, Uppsala Akademiska Sjukhus, Ingårng 50, Uppsala, Friday, March 28, 2008 at 09:15 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English.

Abstract

Osteoporosis is a disease that leads to decreased bone mineral density (BMD), an altered bone micro-architecture and fragile bones. The disease is highly heritable and numerous genes are thought to be involved, making it difficult to identify the causative genetic elements.

Animal models, mainly intercrosses between laboratory strains of mice, have been successfully used to map genes affecting these traits, but may not mirror the multifactorial genetic etiology of highly complex traits such as osteoporosis.

Over the course of tens of thousand years humans have kept domestic animals whose phenotypic repertoires have been tailored to meet our needs. Wild-type red junglefowl (RJ) and domestic White Leghorn (WL) chicken differ for several bone traits.

In this thesis Quantitative Trait Loci (QTL) mapping was used to trace the inheritance of bone traits in two separate intercrosses between RJ and WL. In these studies we identified several QTL that contributed to differences in BMD, bone size and biomechanical strength of bone. In a comparison of QTL identified in the two intercrosses it was observed that nine QTL had overlapping genomic positions, implicating these loci as important to bone phenotypic variation in chicken.

In two separate studies, microarray technology was used to compare global gene expression in bone tissue from RJ and WL. In these studies, differential expression was observed for 779 and 560 genes, respectively. Many differentially expressed genes were co-localized with QTL, which implicates them as QTL-candidates.

Results presented in this thesis link several genomic regions and genes to variation in bone traits. Increased knowledge about these identified genes and regions will contribute to a better understanding of the mechanisms underlying inter-individual differences in bone metabolism, both in chicken and man.

Keywords: QTL, Bone, Osteoporosis, Gene expression, Microarray, Domestication

Carl-Johan Rubin, Department of Medical Sciences, Akademiska sjukhuset, Uppsala University, SE-75185 Uppsala, Sweden

© Carl-Johan Rubin 2008

ISSN 1651-6206
ISBN 978-91-554-7110-1
urn:nbn:se:uu:diva-8498 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-8498)
This thesis is based on the following papers, which will be referred to in the text by their roman numerals


*Paper I was reproduced from J Bone Miner Res 2007;22:375-384 with permission of the American Society for Bone and Mineral Research*
### ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-pb</td>
<td>Three-point bending</td>
</tr>
<tr>
<td>3’-UTR</td>
<td>Three-prime Untranslated Region of mRNA</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of Variance</td>
</tr>
<tr>
<td>BMD</td>
<td>Bone Mineral Density</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>Chr.</td>
<td>Chromosome</td>
</tr>
<tr>
<td>CGH</td>
<td>Comparative Genome Hybridization</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>CNV</td>
<td>Copy Number Variation/Variant</td>
</tr>
<tr>
<td>CNP</td>
<td>Copy Number Polymorphism</td>
</tr>
<tr>
<td>cM</td>
<td>centiMorgan</td>
</tr>
<tr>
<td>DE</td>
<td>Differential Expression/Differentially Expressed</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>DXA</td>
<td>Dual X-ray Absorptiometry</td>
</tr>
<tr>
<td>EST</td>
<td>Expressed Sequence Tag</td>
</tr>
<tr>
<td>eQTL</td>
<td>Expression Quantitative Trait Loci</td>
</tr>
<tr>
<td>L13</td>
<td>White Leghorn line L13 chicken</td>
</tr>
<tr>
<td>Mb</td>
<td>Mega bases (million bases of genomic DNA)</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger Ribonucleic Acid</td>
</tr>
<tr>
<td>miRNA</td>
<td>Micro RNA</td>
</tr>
<tr>
<td>OS</td>
<td>White Leghorn Obese strain chicken</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>pQCT</td>
<td>Peripheral Quantitative Computerized Tomography</td>
</tr>
<tr>
<td>qPCR</td>
<td>Quantitative Polymerase Chain Reaction</td>
</tr>
<tr>
<td>QTL</td>
<td>Quantitative Trait Locus/Loci</td>
</tr>
<tr>
<td>RJ</td>
<td>Red junglefowl chicken</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic Acid</td>
</tr>
<tr>
<td>siRNA</td>
<td>Small interfering RNA</td>
</tr>
<tr>
<td>SNP</td>
<td>Single Nucleotide Polymorphism</td>
</tr>
<tr>
<td>WGA</td>
<td>Whole Genome Association analysis</td>
</tr>
<tr>
<td>WL</td>
<td>White Leghorn chicken</td>
</tr>
</tbody>
</table>
“It is a dangerous business, Frodo, going out your door. You step onto the road, and if you do not keep your feet, there is no knowing where you might be swept off to”.

Gandalf on the topic of voyages in J.R.R. Tolkien's “The Fellowship of the Ring”

“It is like a voyage of discovery into unknown lands, seeking not for new territory but for new knowledge. It should appeal to those with a good sense of adventure”.

Frederick Sanger on the topic of scientific research on Dec. 10, 1980
SKELETAL TISSUE

The human skeleton consists of 206 bones of various sizes and shapes and functions. Bones provide a heavily mineralized rigid framework by which internal organs are protected and body posture is determined. They also act as leverages for muscles, thus enabling locomotion and precise motor functions. The mineralized component of bone is composed of calcium and phosphate, whose homeostasis is regulated by mechanisms coordinated by bone together with the kidneys and the parathyroid glands.

The skeleton harbors two main bone types; flat bones and long bones, differing in the processes by which they are formed. Flat bones include the bones in the skull, ribs and vertebrae whereas long bones are found in the arms and legs. Cortical bone makes up the outermost part of bones and is dense in structure whereas trabecular bone is made up of networks of interconnected bone and is spongier in structure. The relative proportion of cortical and trabecular bone is dependent on skeletal site. Some bones such as the vertebrae are mainly composed of trabecular bone while the diaphyses of long bones are composed almost exclusively of dense cortical bone. In birds, females develop medullary bone, a specialized bone which lines the inner surface of cortical bone and resides in close proximity to the trabecular bone.

BONE METABOLISM

All bones consist of living cells embedded in the mineralized organic matrix that makes up the tissue. The mineralized component is hydroxyapatite, $\text{Ca}_5(\text{PO}_4)_3(\text{OH})$, which comprises 50-70% of the bone mass. The major component of the organic part is collagen, made up of triple helices of two separate type I collagen fibrils, but other proteins are also present. Four cell types predominate in bone: osteocytes, bone lining cells, osteoblasts and osteoclasts, of which osteocytes are most abundant. Bone is constantly remodeled in a process involving interplay between osteoblasts and osteoclasts. Bone remodeling is initiated by resorption by osteoclasts and is followed by new bone matrix being laid down by osteoblasts where resorption took place and the newly formed bone matrix is subsequently mineralized.

Under normal conditions in adult individuals, bone remodeling is a balanced process, by which bone mass is sustained following each remodeling cycle.
OSTEOPOROSIS

Osteoporosis means porous bone and is a complex disease that constitutes a major public health problem in ageing human populations of the world. The disease is characterized by an altered bone metabolism which leads to a degeneration of bone tissue, a reduced bone mass, and consequently an increase in bone fragility and risk of fracture. According to recent estimates the annual incidence of osteoporotic fractures is roughly nine million worldwide, of which almost two million are hip fractures. In the early 1990’s the median cost of medical treatment the year following hip fracture was roughly 11,000 US $. These steep medical costs accompanying fractures make osteoporosis a large socioeconomic burden.

Women are more prone to develop osteoporosis and it is estimated that every other female, but also every fourth male will sustain an osteoporotic fracture in their lifetime. In females, the rate of bone loss is accelerated after menopause due to lowered circulating levels of estrogen. The average annual bone mass loss in post-menopausal women is 3%, compared to 0.5-1 % for similarly aged men. The most important positive predictor for risk of fracture is the presence of a previous fracture, which confers an approximately two-fold greater risk. However, among readily quantifiable traits, a low bone mineral density (BMD) is considered to be the prime predictor. BMD can be calculated by dividing the measured bone mineral content by the measured bone area in a densitometry study using Dual X-ray Absorptiometry (DXA). The diagnostic criterion for osteoporosis is a BMD 2.5 standard deviations lower than that of a gender-matched reference population measured at the age of peak BMD.
Bone mass changes during life; it is gradually added during adolescence and the velocity of bone acquisition is at its greatest at around 12.5 and 14 years of age for girls and boys, respectively. Bone acquisition continues until bone mass peaks at approximately 20-25 years of age, after which an initially slow degeneration of bone tissue commences in both sexes. Several independent routes can lead to osteoporosis, such as a normal or high peak BMD followed by an excessive rate of bone tissue decay, or a low peak BMD followed by a slow rate of bone degeneration. Early menopause is also considered a risk factor, as this will give more years spent in an estrogen deficient state.

Figure 2. Changes in bone mass during life. (A) Normal changes over life for men and women. (B) High, normal and low peak bone mass in females followed by similar rates of postmenopausal bone loss. (C) Both an early menopause and a rapid bone loss at menopause lead to the same amount of bone lost over time.
A trait is a distinguishing characteristic or feature of an individual (for example gender, bodyweight, state of disease or eye color) and the specific values of traits are referred to as phenotypes (e.g. male, 77 kg, diabetic or blue eyes). The phrase “nature or nurture” is often used to question whether a certain trait of an organism is dependent on heritable genetic factors or on the external milieu which the organism has been exposed to (environmental factors).

Traits and diseases that have so far been explained by specific genetic factors are primarily ones caused by mutations that interfere with the functions of single genes. Such monogenic traits and diseases have characteristic patterns of inheritance and the genetic components can therefore be traced in affected families. Examples of diseases caused by mutations of single genes include different types of the bone fragility disease Osteogenesis Imperfecta. In contrast to monogenic diseases, common diseases such as type II diabetes, obesity, osteoporosis, hypertension and vascular disease are termed complex because of their complex modes of inheritance. Complex diseases as well as inter-individual variation of complex traits such as height and weight are dependent on the combined effects of several genetic and/or environmental factors. This multi-factorial nature makes it difficult to elucidate the causative genetic variants and how these interact to establishment of disease or phenotype.
HERITABILITY OF BONE TRAITS

Heritability is the capacity of a trait to be passed on from one generation to the next. It has been seen that daughters of women suffering from osteoporosis have a reduced bone mass in comparison to women from the general population 16. The heritability of a trait can be assessed in studies of monozygotic twins, who have virtually identical genomes but for whom environmental variables can differ. Twin studies have estimated that genetic factors account for approximately 65-80% of the variance in BMD 17-20. Other twin studies have shown that heritable factors account for approximately 25-50% of the fracture risk, depending on bone compartment studied 21, 22. It has been shown that balance impairment may have heritable components 23, and since ninety percent of hip fractures occur due to falls 24 some of the genetic contribution to fracture may be orchestrated by genetic variants predisposing for falls. The lower heritability for fracture risk than for BMD can also likely be attributed to fracture occurrence being more heavily influenced by lifestyle and environmental factors.

It would be beneficial to identify the genetic factors responsible for inter-individual differences in bone traits such as BMD and fracture. Knowing in more detail the array of genes involved would further our understanding of the complex nature of osteoporosis and this knowledge could be used for designing novel pharmaceuticals to slow down or hinder the disease process. Certain variable genetic factors could also prove predictive for the risk of future fracture or low BMD, and could therefore be used in molecular screenings aimed at preventing the onset of disease at an early stage.
GENETIC VARIATION

The nitrogenous bases Adenine, Cytosine, Guanine and Thymine, commonly abbreviated A, C, G and T are the building blocks of DNA. DNA is composed of combined stretches of these four bases attached to a backbone of sugars and phosphate groups. In the genome two DNA strands together form anti-parallel double helices, chromosomes. Genomes are composed of several chromosomes that reside within the nucleus of almost every cell type in the body. In diploid organisms such as humans and chickens, each chromosome is present in two copies, one derived from each parent.

What separates each of us genetically from our fellow humans and to a large extent also from the apes, our closest relatives, are numerous inter-individual variations (polymorphisms) present in our genomes. The most common variations are single nucleotide polymorphisms (SNPs). A SNP is a substitution at a specific genomic position, where more than one nucleotide is represented in the population. In order to be defined as a SNP, the least abundant single nucleotide variant has to occur in the population at a frequency higher than 1%. In addition to SNPs, our genomes also harbor other variable elements, such as; microsatellites, short- and long interspersed nuclear elements (SINEs and LINEs), small insertions and deletions as well as copy number variations (CNV) of whole segments of DNA.

When two human chromosomes are compared, they differ due to a SNP on average once every 1000-2000 nucleotides. To this date approximately 12 million human SNPs have been reported, of which more than 6 million have been validated.

Genotyping is the process used to determine which genetic variant is present at a certain position in the genome. DNA sequencing can be used for this purpose, as can techniques only giving information about variable positions of the genome, e.g. SNPs, microsatellites etc. It has been observed that certain variable positions that are physically close on a chromosome have a tendency to co-occur. These co-occurrences (haplotypes) are beneficial because the genotyping of a limited number of polymorphisms can be sufficient to know the full repertoire of genetic variation in a genomic region. This information has been assembled in a haplotype map of the human genome.
 IDENTIFYING THE CAUSATIVE GENES

The complex genetic makeup of BMD and other bone traits complicate the identification of the causative genetic variants. Phenotypic variables can be subdivided into discrete and continuous, of which the former have a limited number of possible values, e.g. being affected by a disease or not. For traits such as BMD, the phenotypic values are continuously distributed in the general population. Both for continuous and discrete variables, statistical tests can be used to examine if genetic variation in certain genes or loci explain any of the phenotypic variation in a population. The methods within the field can be subdivided into two main groups:

- **Linkage analyses** – No prior hypothesis about gene: Phenotypes and genetic variants (SNPs or other variable elements) spread throughout the genome are documented in related individuals (human families or intercrosses between animal strains). Statistical tests are conducted to examine whether the inheritance of parental alleles at any locus is linked to phenotypic variation of the traits examined.

- **Association analyses** – Prior hypothesis about gene: Certain genes are examined for normal genetic variation such as SNPs present in a population. In statistical tests, such genetic variation is tested for association to phenotypic values observed in the population. Usually large numbers of unrelated individuals representing a cross-section of the population are tested.

STUDIES IN HUMAN FAMILIES AND COHORTS

Linkage studies

Numerous linkage studies have been performed in human families and linkage has been reported to all human chromosomes except the Y-chromosome. That bone traits have been linked to numerous genomic regions supports the common assumption that a multitude of genetic variants are involved in bone phenotypic variation. However, many of the identified loci have been linked to bone traits only in one study, suggesting either false positive linkage or
alternatively that these loci contribute to differing bone traits only in certain populations or under certain conditions. However, in several cases linkage has been replicated in one or several additional studies, suggesting presence of genetic variation relevant to bone phenotypic variation. Examples of replicated studies include a reported linkage to the q-arm of chr. 1 for peak BMD in females\textsuperscript{28, 29}. Another example of replicated linkage includes the linkage for bone size to a locus on chr. 17\textsuperscript{30, 31}, which harbors the candidate gene \textit{Collagen 1 alpha 1} (\textit{COL1A1}). In a large sample of Icelandic families, BMD as well as fracture incidence were linked to a locus on chr. 20, and follow up studied in a sample of similar ethnicity implicated one haplotype of the \textit{bone morphogenic protein 2} (\textit{BMP2}) gene as responsible for these effects\textsuperscript{32}.

Altogether, linkage studies have implicated several genomic regions as important to inter-individual variability in bone phenotypes. Important findings derived from these studies include those suggesting bone site specific as well as gender specific regulation of BMD\textsuperscript{33-35}. Furthermore, it appears as if the genetic variants that regulate peak BMD are different from those regulating BMD at old age\textsuperscript{35-37}.

Association analyses

Association analyses examine the contribution of genetic variants such as SNPs (or haplotypes) to the trait of interest. Most often the subjects to be tested are age-matched individuals chosen to represent a cross-section of the population. The genetic variations to be tested are often chosen based on previous knowledge or hypotheses about gene function. Previous knowledge could include genes that have rendered bone phenotypes in murine knock-out models, genes known to be responsible for monogenic human bone diseases or genes located in regions linked to bone traits.

Certain factors complicate the identification of the genetic variants involved in inter-individual differences in bone traits. These include gender-specific effects\textsuperscript{33}, gene-by-gene interactions (epistasis)\textsuperscript{38} and possibly also gene-by-environment interactions\textsuperscript{39}. It has also been shown that some loci controlling BMD have pleiotropic effects on lean body mass\textsuperscript{40}. In light of this, co-variables are usually applied in the statistic models used in quantitative genetics. Rigorous questionnaires are often used to gather information about exposure to environmental factors etc. Phenotypes such as gender, body weight and underlying diseases can therefore be applied as covariates in the statistical models used in linkage- and association analyses.

Cohort size is important with regard to statistical power, i.e. probability of being able to detect a genotype/phenotype association if there really is one. It is therefore crucial that large enough cohorts are used, in particular when studying traits such as BMD or fracture where many genes with individually small effects are involved. Many hitherto performed association analyses
have likely been under-powered and may therefore have reported false negative results. Conversely, spurious associations have likely been reported in cases where significance thresholds were too liberal, not adjusting for multiple tests having been conducted. Population stratification is another important parameter as this could both render false association and mask true effects. For example, individuals of African descent generally have a higher BMD than other ethnic groups, signifying the importance of comparing ethnically homogenous groups in association analyses. Considering these factors as well as the complexity of haplotype patterns present in human populations it becomes apparent that study design is crucial to association analyses. As the scientific community has become increasingly aware of this, cohorts comprising several thousand to tens of thousands individuals are commonly used in association analyses.

Repeated associations to traits such as BMD and fracture have been reported for some genes. One such example is the Vitamin D receptor (VDR) gene whose genetic make-up has been extensively studied in relation to both BMD and fracture since it was first found associated to BMD. Although some large-scale studies have reported absence of association between VDR polymorphisms and BMD as well as fracture, more than 100 reported associations to bone traits strongly suggest that genetic variation in VDR is important. It is however still not clear by which molecular mechanisms genetic variation in VDR confers these segregating traits. Some studies have reported findings suggesting that VDR polymorphisms may confer segregating bone phenotypes more dramatically in response to certain environmental factors such as physical exercise and diet. Another gene that has been extensively studied for association to bone traits is the Estrogen receptor alpha (ERA), which has been repeatedly associated to BMD. In a large-scale meta-analysis it was concluded that one haplotype of ERA was strongly associated to fracture risk and that this effects was independent of BMD. A specific polymorphism (Sp1) in the COL1A1 gene has been repeatedly associated both to BMD and fracture risk, and it has been shown that osteoblasts isolated from bearers of different variants of Sp1 differ in their ability to form mineralized bone nodules. Polymorphisms in several other genes have also repeatedly been found associated to bone traits, and examples of such genes are: Interleukin 6 (IL-6) and low-density lipoprotein receptor-related protein 5 (LRP5).

The functions of human genes, RNA-molecules, proteins and regulatory genetic elements are still largely unexplored. Such unexplored genetic elements may harbor genetic variation with profound effects on inter-individual variation in traits. This makes it conceivable and even likely that association analyses have been hampered by their candidate gene oriented nature. Novel genotyping techniques and their ever increasing coverage of variable elements in the human genome have made it possible to determine the genotype of virtually all known human SNPs simultaneously. This enables researchers
to perform hypothesis-free association analyses. Several such whole-genome association (WGA) analyses have been and are currently undertaken for various complex traits, including those of bone. One single report has so far been published for bone traits \(^6\). Traits analyzed in that study included various BMD measurements and geometrical properties of the hip. Results revealed that only less than 20 out of 40 SNPs receiving the highest statistical support for association to BMD were situated in the vicinity of known candidate genes. Although relatively small in scale with a cohort of 1141 individuals, this study signified the importance of a hypothesis-free approach when studying complex traits in association analyses.

**STUDIES IN ANIMAL MODELS**

In order to conduct genetic studies for osteoporosis in a controlled manner, animal models have been extensively utilized. In what is referred to as reverse genetics, specific genes can be manipulated in mice \(^6\text{3-6}^5\), either by gene over-expression or by removal of the gene from the genome (knock-out). This is followed by examination of how the phenotypic repertoires of these animals differ in comparison to their unaltered littermates. Reverse genetics is highly valuable when it comes to unraveling the functions of specific genes, but these studies do not examine the effects of naturally occurring genetic variation on complex traits.

Forward genetic approaches start out with a trait and examine the genetic contribution to that trait. In Quantitative Trait Loci (QTL) mapping, animal strains differing for traits of interest are intercrossed and the second generation offspring (F\(_2\)-individuals) are subjected to phenotyping as well as genotyping, the latter being used to construct a linkage map. In statistical analysis, phenotypic values of F\(_2\) individuals are regressed on the genotypes of genetic markers spaced throughout the genome. Together with the linkage map this enables the positional mapping of QTL, i.e. inherited loci where allelic variation contributes to phenotypic variation within the intercross. Advantageous features of QTL-mapping when conducted in animal models include large pedigrees, short generation times and the ability to control environmental factors. The use of animal models also enables investigation of phenotypes requiring invasive measurements, such as bone strength.

---

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantitative Trait</td>
<td>A biological trait that can be quantified</td>
</tr>
<tr>
<td>Locus</td>
<td>A region of a chromosome</td>
</tr>
<tr>
<td>Loci</td>
<td>Several distinct regions on one or more chromosomes</td>
</tr>
<tr>
<td>Quantitative Trait Loci</td>
<td>Loci contributing to differences in quantitative traits</td>
</tr>
</tbody>
</table>
Figure 3. Principle of Quantitative Trait Loci (QTL) mapping in an animal intercross. Two fictive populations (populations A and B) differing for one or several complex traits are intercrossed to generate \(F_1\)-individuals. \(F_1\)-individuals are then intercrossed to each other to generate a \(F_2\)-generation (hundreds to thousands of individuals). The phenotypes of interest are measured for each \(F_2\)-individual (grey dots in subplot A) and genotypes are determined for genetic markers (M1-M5) spaced throughout the genome (shown for four \(F_2\)-individuals in subplot B). At each position in their autosomal genome, \(F_2\)-individuals can bear genotypes AA, AB or BB. In regression analysis it is then tested if genotype at any genetic marker interval is statistically linked to differences in phenotypic values in the \(F_2\)-generation. As can be seen in subplot C, \(F_2\)-individuals carrying BB alleles at M2, M4 and particularly M3 had higher values for the phenotype than individuals having the other genotypes. Genetic variation in the vicinity of M3 appears to explain some of the phenotypic variation between populations A and B. This is visualized as a peak in statistical significance for linkage (LOD-score peak) at a map position near M3 (subplot D).
Quantitative Trait Loci mapping

The mouse has been the model of choice in many QTL-studies and several interesting findings have been published; linking loci on almost every murine chromosome to BMD. As has been observed in human linkage studies, some murine QTL have been identified as specifically affecting certain bone sites and others as sex-specific. An advantageous feature of animal models is that congenic lines can be created. Congenic animals differ from one of the parental lines only at certain loci, a feature enabling dissection of the effects of single QTL. Single QTL have in some cases been explained as multiple adjacent QTL, each with separate complex effects. This could have implications for the interpretation of human linkage studies, where more than one causative variant in linked regions may complicate repeatability of findings.

Fracture incidence, when tested in human cohorts or families is confounded by environmental factors other than the biomechanical properties of the bone. A powerful feature of QTL-analyses conducted in animals is the possibility to test invasive bone phenotypes such as bone strength, which can be tested ex vivo by bending and torsion. Several studies have employed such phenotyping techniques. In chicken, a QTL affecting bone strength has been identified in a two generation intercross population produced by mating lines that had been divergently selected for a bone index involving both bone strength and density. Several studies conducted in murine intercrosses have identified QTL for bone strength. In one of these studies it was concluded that three out of six bone strength QTL had effects that were independent of variation in BMD. In this study also found that epistatic interactions were attributable for half of the variance in bone strength in the F2-population.

For certain genes identified as causative of QTL in animal models, genetic variation in human orthologues or paralogues may also affect the trait. A murine QTL affecting BMD has been explained by genetic variation in the gene *Arachidonate 15-Lipoxygenase (ALOX15)*. Follow-up experiments led to the discovery that polymorphisms in *ALOX12*, another lipoxygenase gene, were associated to BMD in human subjects. This suggests that QTL-mapping for complex traits in animal models may serve an important hypothesis generating purpose. Performing QTL-analyses across an array of vertebrate species or breeds within species may also strengthen or decrease confidence intervals of already identified QTL.
Over the course of tens of thousand years humans have kept domestic animals whose phenotypic repertoires have been tailored to meet our needs. This long-term selection for desired traits has resulted in a wide variety of phenotypically divergent animal breeds which may be valuable models for genetic studies of complex traits \textsuperscript{77, 78}. Successful examples using domestic animal models include the explanation of a porcine QTL for muscle growth and fatness by a single nucleotide substitution in an intron of the \textit{Insulin growth factor 2} gene \textsuperscript{79}. In another study, a QTL for muscularity in sheep was explained by a polymorphic microRNA binding site in the \textit{Myostatin} gene \textsuperscript{80}.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{selection_diagram.png}
\caption{Principle of selection presented for a fictive population. In the upper plot, phenotypic values for a quantitative trait (X) are presented for all individuals (dots). Also shown is the same population with individuals stratified by genotype for two genes (A and B). Genotype for gene B affects the phenotypic value of X, whereas genotype for gene A does not. If a high X is beneficial for reproductive success or is artificially selected for, genotype bb will be enriched in subsequent generations. Effects of three different values of X as selection criteria on genotype frequencies observed in the subsequent generation are shown in diagrams below the plot.}
\end{figure}
The chicken as a model species

In addition to its importance in agriculture, chicken (*Gallus gallus*) is a well-established model organism commonly used in various branches of biological research\(^8^1\). The red junglefowl (RJ) is native to forests in south-east Asia and is regarded as the main ancestor of domestic chicken\(^8^2\). The first draft of the chicken genome sequence, featuring the genome of the RJ, was released in 2004\(^8^3\). In a parallel project to the genome sequencing, partial genome sequences were generated from three domestic chicken breeds (White Leghorn broiler and Silkie). When generated sequences from the three breeds were compared to each other and to the RJ genome sequence, 2.8 million SNPs were identified\(^8^4\). The size of the chicken genome is about one third of the human genome size, but only a slightly lower gene repertoire and consequently a higher gene density. The fraction of the chicken genome occupied by repetitive sequence is 9-11\%\(^8^3\), which is a substantially lower than for the murine and human genomes, where the corresponding fractions are 40-50\%\(^8^5, 8^6\). A wide variety of phenotypically divergent breeds are available, which makes chicken an attractive model for genetic studies of complex traits.

**Chicken bone biology**

A condition referred to as “cage layer fatigue” was reported in 1955 when fractures were found in 90\% of post-production egg-laying hens\(^8^5\). Recent studies have reported lifetime fracture incidences of ~30\% for hens held in egg-production facilities\(^8^6\). When the avian female become sexually mature, estrogen stimulates formation of medullary bone\(^3\), a labile type of non-structural woven bone\(^8^7\) which is more heavily calcified and metabolized at a much faster rate than cortical bone\(^8^8\). The medullary bone predominantly assembles in the medullary canal of long bones, where it functions as a labile source of calcium for eggshell formation. The mechanisms by which osteoporosis occurs in chicken have earlier been reviewed\(^8^9\). Briefly, the calcium used for deposition in eggshells can be derived both from the diet as well as from skeletal stores, but when dietary calcium is scarce; calcium is mainly mobilized from medullary bone. As opposed to medullary bone, remodeling of structural bone is limited during reproductive periods, leading to weakening of bones by a reduced cortical thickness as well as thinner trabeculae.
FUNCTIONAL GENOMICS

Linkage-, QTL- and association studies can implicate certain genomic regions or genes as important to phenotypic variation. However, these studies do not provide any molecular explanations to how genetic variation confers variability in phenotypic traits. Full or partial genome sequences have been determined for at least 41 animal species and this has given rise to the “omics” era of molecular biology. The term omics refers to the comprehensive analysis of biological systems and -omics is added as a suffix to describe the specific entity of a biological system that is analyzed in detail. Properties of DNA and genetic variation are studied in “genomics” and various attributes of gene products are studied in “proteomics” and “transcriptomics”.

In functional genomics, omics techniques are combined to study functions of and interactions between DNA, RNA and proteins in biological systems. Functional genomics focuses on dynamic aspects such as gene transcription and translation, and includes studies of polymorphisms and measures of molecular activities.

Gene expression

Save for occasional mutations the genome is static, identical in all somatic cells of an organism. Genomes contain information that is to be sent out of the nucleus and into the cellular cytoplasm or to other cells of an organism. The transmission of information is achieved by certain parts of DNA being copied into ribonucleic acid (RNA) in a process called transcription. Genomic regions that are transcribed into messenger RNA (mRNA) are called genes. The genetic code transmitted by genes via mRNA is deciphered at cytoplasmic complexes - ribosomes, where the code embedded in coding parts of mRNA (exons) is translated into building blocks of proteins, amino acids. The human genome is estimated to contain 20,000-25,000 protein coding genes and similar numbers are seen in the genomes of other mammals as well as non-mammalian vertebrates. The code embedded in genes can sometimes render more than one mRNA-variant formed from the same gene, thereby adding complexity to the DNA-code. Many proteins are subjected to post-translational modifications and the numbers of distinct proteins that can be formed are estimated to be several orders of magnitude higher than that of mRNAs.
Gene expression is a complex and tightly regulated process by which cells can respond to various stimuli in a dynamic manner. The transcription of genes is regulated by promoter regions that are located upstream of the coding region of genes. Promoters contain DNA sequences that serve as binding sites for specific proteins - transcription factors. When bound to response elements, transcription factors regulate transcriptional efficiency by recruiting the enzyme RNA polymerase II, which synthesizes mRNA. In addition to the transcriptional regulation of gene expression, several additional regulatory mechanisms control levels of mRNAs and proteins available post-transcription. Only a small fraction of the human genome is composed of protein coding genes. Intergenic regions as well as intronic regions of genes were long regarded as redundant remnants of evolution, serving no specific purpose. In recent years several classes of non-coding RNA have been found to be expressed from intergenic regions and introns, and some classes of such non-coding RNA have been attributed roles in the regulation and fine-tuning of gene expression. microRNAs (miRNAs) were discovered in 1993 and have since then been recognized as important regulators of gene expression. miRNAs bind to mRNAs that contain partially complementary sites in their 3'-untranslated regions (3'-UTRs), and this binding can affect translation efficiency as well as mRNA stability. Recent technical advances in RNA quantification have made it evident that RNA is widely expressed from non-coding regions of the murine as well as the human genome, suggesting that the extent and complexity of gene expression regulation by non-coding RNA has just begun to be unraveled with the identification of miRNAs.

Proteins have diverse functional repertoires in the body and are, together with other molecules, involved in the maintenance of cellular and systemic processes associated with life. Knowing the levels at which gene products are expressed in cells or tissues could therefore give important clues about ongoing cellular processes.

Quantifying gene expression

One of the first techniques that allowed for quantification of RNA abundance was Northern blotting. This technique takes advantage of the ability of an individual single-stranded nucleic-acid to form a double-stranded molecule (hybridize) upon interaction with another nucleic acid, showing a high degree of complementarity. In the mid 1990’s quantitative PCR (qPCR) technique was developed and enabled a more precise determination of RNA levels. Several other techniques have been used, such as Serial Analysis of Gene Expression (SAGE), a technique involving the sequencing of complementary DNA (cDNA) and then counting the number of specific RNA molecules observed. qPCR is still regarded as one of the most reliable techniques.
for RNA quantification, but like the other techniques mentioned above it does not enable simultaneous measurements of many RNAs. A breakthrough came in 1995 when cDNA-microarrays were developed. These also utilize the process of nucleic acid hybridization and comprise collections of many microscopic spots of cDNA clones linked to a solid surface. Each spot is referred to as a probe and contains cDNA clones complementary to a single gene. Originally these microarrays harboured probes derived from mRNAs expressed in certain cells or tissues. One limitation of the cDNA-microarray technology is that the probe repertoire only allows analysis of mRNAs that are available in the library from which the probes were created. Technical advances now enable chemical synthesis of oligonucleotides directly on glass surfaces. Several million probes can now be synthesized on these arrays, enabling quantification of mRNA levels for virtually all genes in a genome whose sequence has been determined.

Typically gene expression microarrays are used to compare mRNA levels in biological samples such as cells or tissues. Microarrays can be used to investigate up- and down regulation of gene expression in response to external stimuli such as drugs or proteins transfected to cell cultures or administered to animals. Additionally, microarrays can be used to analyse inter-individual differences in gene expression, for example by comparing cells or tissues derived from separate populations or species.

Gene expression can be influenced by both endogenous- and exogenous factors, which depending on experiments could introduce unwanted variability to data generated in microarray hybridizations. Also, microarray experiments involve numerous laboratory steps, each of which could potentially introduce variability in data derived from different hybridizations. Technical as well as biological replicates are often used to limit the impact of such unwanted variability on the list of genes identified as differentially expressed, and data derived from hybridizations need to be normalized before statistical analysis is carried out. Statistical methods used to infer differentially expressed genes from data generated by microarray hybridizations include: ANOVA, Student’s t-test and Bayesian methods, all of which assign probabilities (P-values) of differential expression to each gene on the microarray. Several thousand separate statistical tests are performed in statistical analyses of microarray data, making it necessary to adjust significance thresholds for differential expression. The Bonferroni method is commonly used in statistics to control for multiple testing, but may be too stringent for most microarray applications. Estimations of the False Discovery Rate (FDR) are commonly used to control the proportion of non-differentially expressed genes that will be identified as differentially expressed.
Figure 5. Simplified view of how cDNA-microarray technology can be used to identify genes differentially expressed between two populations. Fluorescently labeled RNA or cDNA samples are incubated on the surface of the microarray. cDNAs corresponding to specific genes will hybridize to complementary capture probes in different positions on the surface. After incubation, fluorescence intensities are measured for each capture probe position on the microarray. A labeled reference sample is often co-hybridized with each differently labeled individual sample, as this facilitates comparisons of gene expression levels in the individual samples.

Genetics of gene expression

Genetic variation can confer differential expression of genes. Such differential expression can be caused by cis-acting genetic variation, whereby polymorphisms in for example promoter regions or 3'-UTRs of genes can confer altered interactions with transcription factors, miRNAs or hitherto unknown regulators of gene expression. The differential expression of genes can also be secondary to genetic variation conferring altered functional properties or altered expression levels of trans-regulatory elements expressed from other loci. It has been suggested that non-coding genetic variation is important to a variety of quantitative traits. Differential expression of genes located within QTL-regions for complex traits could therefore implicate these genes as positional candidate genes of QTL.
search field, gene expression analysis enabled the identification of secreted frizzled-related protein 4 (SFRP4) as causative of a QTL affecting BMD in a murine intercross. Another BMD QTL was explained by regulatory genetic variation that conferred differential expression of ALOX15 in kidneys and osteoblasts from the two parental strains of mice.

Gene expression can be quantified and the genetic contribution to gene expression can therefore be mapped in expression QTL (eQTL) analyses. When combined with microarray expression analysis, eQTL mapping can be used for hypothesis-free mapping of QTL exerting differential regulation of gene expression. This is exemplified by microarray based eQTL-studies performed for tissues, as well as cells from rodent intercrosses.

**Figure 6.** An example of how SNPs could influence transcription through cis- and trans-regulatory effects. In the topmost example, individual 2 has a promoter SNP variant altering the interaction between a transcription factor (TF) and the promoter. This leads to decreased expression of gene A and the SNP exerts cis-regulatory effects on the transcription of gene A. In the lower example a SNP determines the expression level of the TF (Gene B), which is situated on another locus. This genetic variation exerts trans-acting regulation of Gene A expression.
Bioinformatics

Bioinformatics involves the analysis and management of biological information using computers and statistical techniques. As the amount of data generated by commonly used “-omics” techniques grows larger, there is a need to assemble this information into readily accessible and easily interpretable formats. Genome sequences are available for browsing in online databases such as Ensembl 90, which features genome sequences assembled into separate chromosomes. SNP genotype frequencies human populations as well as the validation status of these SNPs are available in the online SNP repository SNPdb (http://www.ncbi.nlm.nih.gov/SNP/), which also features information about SNPs identified in other species. Vast amounts of data have been generated since microarrays began to be utilized in molecular biology. In 2001, Minimum Information About a Microarray Experiment (MIAME) 115 was proposed as a standard for how microarray-based gene expression data should be recorded and reported in order for others to be able to interpret the results. Public repositories such as ArrayExpress 116 have been designed to accept, hold and distribute MIAME compliant microarray data.

The recent revolution in genome sequencing has, in combination with increased understanding of regulatory mechanisms governing gene expression, spurred the development of computer algorithms predicting regulatory elements in genomes. Online databases such as TRANSFAC 117 contain information about known transcription factors as well as their known and proposed binding sites in genomes. Several algorithms have been developed for predicting miRNA target sites within 3’-UTR regions of genes 118-120. Predicted target sites have been determined globally for several genomes, including human, chicken, mice and rat 98. Polymorphic miRNA binding sites can affect biological traits 80, 121, and polymorphisms that may affect the interaction between miRNAs and their targets have been determined globally for several species, including chicken 122.

What constitutes the highest biological relevance; the four-fold differential expression of one gene or the 0.25-fold differential expression of another gene? The answer is dependent on the functional properties of the two genes in the system studied and the hypothesis tested in the experiment. The Gene Ontology (GO) project is a collaborative effort with the aim of assigning consistent attributes to gene products 123. The GO-project strives to provide a relational vocabulary that describes various attributes of gene products in a species independent manner. Gene products are organized by three top-level GO-terms: cellular component, biological process and molecular function. Within each of these top-level ontologies, gene products are given hierarchical relationships to each other and each gene product can be associated to an unlimited number of GO-terms. The GO-annotation of gene products is normally carried out by curators in bioinformatics database resource groups, and evidence codes accompany each annotation. Evidence codes describe
the evidence that underlies each annotation, and experimentally verified annotations are considered most reliable, whereas those that have been solely annotated by computational approaches are the least reliable. Microarray experiments usually render long lists of genes identified as differentially expressed and further analysis using GO annotations can be useful to identify overrepresentation of certain classes of GO-terms among differentially expressed genes.

Biological systems are immensely complex and despite efforts such as the GO-project, our knowledge of the functional repertoires of gene products is still rudimentary. Furthermore, it is likely that the regulatory processes controlling levels of gene products have merely begun to be unravelled with the identification of miRNAs. Modern gene expression microarrays enable parallel quantification of transcript abundance for the full repertoire of mRNAs, and consequently allow hypothesis-free probing for perturbations in gene expression. In order to successfully unravel the mechanisms by which genetic variants govern complex traits, large-scale quantitative genetic studies will likely be complemented with functional genomic approaches, such as gene expression profiling in cells and tissues.
AIMS

The overall aim has been to identify genes involved in bone phenotypic variation. The aim of this thesis was to combine genetic and functional analyses to provide genetic explanations for observed differences in bone traits between wild-type red junglefowl (RJ) strains and domestic White Leghorn (WL) strains of chicken. Ultimately, such findings could prove important to bone phenotypic variation not only in chicken but also in man.

The specific aims were:

- To utilize Quantitative Trait Loci (QTL) mapping in order to examine the genetic contribution to variation in phenotypes such as bone mineral density, bone geometry and bone strength between domestic WL strain L13 (L13) and wild-type RJ chicken (Paper I)

- To perform QTL-mapping in an intercross between another WL-strain (OS) and a separate RJ-population than the one used previously. (Paper IV)

- To investigate the presence of overlapping bone traits QTL between the two intercrosses RJ x L13 and RJ x OS (Paper IV)

- To identify differentially expressed genes in femoral bone tissue from RJ and L13 and to determine which of these are located within QTL (Papers II and III)

- To identify functional pathways overrepresented among differentially expressed genes (Papers II and III)

- To identify genes and genomic regions important to bone phenotypic variation and bone metabolism in chicken (Papers I-IV)
MATERIALS AND METHODS

Chicken pedigrees and parental individuals

Red junglefowl x White Leghorn strain L13 (L13) pedigree (Paper I)

A three-generation pedigree was generated by crossing one red junglefowl male with three White Leghorn (WL) females from the SLU13 (L13) line as described in 125. The red junglefowl male was obtained from a Swedish zoo, and originated from a relatively closed European zoo population. Four male and 37 female F1-individuals were intercrossed to generate 851 F2-individuals, raised in six separate batches as described in 125. 337 F2 birds (159 female and 178 male) as well as ten White Leghorns and ten red junglefowls were sacrificed at 200 days of age. The birds were de-feathered and frozen immediately after death, and were then stored at –20°C until phenotypic measurements were performed.

Red junglefowl x White leghorn Obese strain (OS) pedigree (Paper IV)

A three generation intercross was generated by mating White Leghorn obese strain (OS) to the red junglefowl (RJ). The OS-line originates from a White Leghorn line and spontaneously develops autoimmune thyroiditis during the first weeks of life 126. Red junglefowl eggs were obtained from a breeding population of approximately 60 birds (~30 of each sex) maintained at Götala Research Station in Skara, Sweden. The population had been in captivity for 14 years, and were brought to a Swedish zoo (Frösö zoo) before being introduced to the research station; the background of this population is described in more detail in 127, 128. Approximately 1000 F2 individuals were hatched in five batches with approximately 200 individuals in each. Due to the hypothyroidism of the OS-line, thyroxin (T4) was supplemented to all but one batch of F2-individuals, starting at 12 weeks of age until 380 days of age, in the form of pulverized levothyroxin tablets (Levaxin, Nycomed AB, Stockholm, Sweden), which were added to the food (500 µg thyroxin/kg food). Immediately post mortem, right femoral bones were thoroughly stripped of soft tissue and were stored at -20°C until they were subjected to biomechanical tests (554 femurs) and to phenotyping with peripheral computerized tomography (pQCT) (543 femurs).
RNA preparation from parental L13 and RJ (II and III)

For use in study II, five individuals representing each sex and population from WL strain L13 and from red junglefowl were sacrificed at 40 weeks of age. In study III five female and four male individuals from each of the same two lines were sacrificed at 100 weeks of age. In addition, five female first generation RJ x L13 offspring (F1-individuals) were sacrificed at 200 days of age. The red junglefowls used for these studies originate from Thailand and their history is described in more detail in 127, 128.

Immediately post mortem, left and right femoral bones were thoroughly stripped of soft tissue and were snap frozen in liquid nitrogen and stored at -70°C until further use. From each left femur the mid-diaphysis was crushed to a fine powder in a mortar while submerged in liquid nitrogen. Bone powder was transferred to 1.5 ml tubes containing 1 ml Trizol reagent (Invitrogen), and RNA was subsequently prepared according to the manufacturer's instructions. RNA was purified using RNEasy spin column (Qiagen).

Phenotypic measurements

Dual X-Ray Absorptiometry (Paper I)

Each carcass was thawed and BMD of the total body was measured using a DXA scanner (Prodigy, Lunar Co, Madison, USA), with the small animal mode utilized. Two sequential total body measurements were performed on each animal, and the mean value of the two measurements were used in the subsequent analyses. Measurements of Bone Mineral Content (BMC) and areal Bone Mineral Density (aBMD) of the femur were performed ex vivo with a Norland pDEXA Sabre (Norland, Fort Atkinson, WI, USA).

Peripheral Quantitative Computerized Tomography (Paper I)

Computerized tomography (CT) was performed with the Stratec pQCT XCT Research M (Norland; v5.4B) operating at a resolution of 70 μm. Non-cortical BMD, which in the female bird reflects BMD of both trabecular and medullary bone, was determined ex vivo, with one metaphyseal pQCT scan of the region situated at 6% of bone length from the distal end of femur, and the non-cortical bone was defined by setting an inner threshold to density mode (400 mg/cm3). In addition to data for non-cortical bone, the metaphyseal scan was also used for derivation of data for total bone (including cortical, trabecular, and medullar bone). Cortical bone parameters were determined ex vivo with a mid-diaphyseal pQCT scan of the femur. Femurs from males and females were analyzed after replacing any air inside the medullary cavity with 70% ethanol. Because pQCT measurements of 70% ethanol alone gave values at 60 mg/cm³, it was concluded that values <60 mg/cm³
could not be correctly quantified. Because almost all males had values for non-cortical BMD that fell below the 60-mg/cm$^3$ threshold, all males were excluded from subsequent QTL analysis for this parameter.

**Peripheral Quantitative Computerized Tomography (Papers III and IV)**

Right femurs were phenotyped using the Stratec XCT - Research SA instrument (Stratec Medizintechnik, Germany). Two mid-diaphyseal scans as well as two distal metaphyseal scans at 6% of bone length were performed. Cortical bone was analyzed in the diaphyseal scans using the parameter CORT-MODE1 with density threshold $>$1000 mg/cm$^3$, which enabled analysis of cortical parameters specifically. In analyses of all other bone parameters an outer threshold of 280 mg/cm$^3$ was used to separate soft tissue from bone. For separation of cortical and non-cortical bone in the diaphysis, an inner threshold of 1000 mg/cm$^3$ was used in PEELMODE2.

For use in **Paper IV** a second round of data collection was performed, where an inner threshold of 150 mg/cm$^3$ was used in PEELMODE2. This lower threshold corresponds approximately to the density value obtained from scans of water alone and this round of data collection enabled calculations of parameters of the mineralized- and non-mineralized parts of bone.

![Figure 7. Images derived from pQCT-phenotyping of chicken bones and visualization of how this information can be used to subdivide bone compartments.](image-url)
Three-point bending (Papers I and IV)
The right femurs were tested for biomechanical strength in a three-point bending test on an electromechanical testing machine (Avalon technologies, Rochester, MN, USA). The specimens were placed with the posterior cortex resting against two end supports placed with a distance of 40 mm between them. The bones were placed in such a way that the load was applied 6 mm distal from the mid part of the femoral diaphysis with an antero-posterior direction. An axial load cell (Sensotec inc., Columbus, OH, USA) with the range 0-500 N was used to apply a load of one mm/sec to the bone. Values for load and displacement were collected 50 times per second until failure using software provided with the testing machine (Testware II).

Torsional strength tests (Paper I)
The left femurs were tested until failure in torsion. Both bone ends were placed in purpose made aluminum potting fixtures with a height of 15 mm each. The fixtures were filled with plastic padding (Plastic Padding Elastic, Sweden) while leaving 50-52 mm depending on femur free for testing. All specimens were tested to failure in torsion at room temperature on an electromechanical testing machine (Avalon technologies, Rochester, MN, USA) at a rate of four degrees per second, with no axial load during the testing. Values for angulation and torque were collected 50 times per second until failure using the software provided with the testing machine (Testware II).

Gene expression studies (Papers II and III)

Microarrays
cDNA Microarrays (Paper II)
The microarrays used in this study (KTH UniChicken 2x14k cDNAv1) were developed at the Royal Institute of Technology (RIT) in Stockholm, and comprise 13,907 Expressed Sequence Tags (ESTs) spotted in duplicate. Of the ESTs, 12,742 originated from red junglefowl and White Leghorn brain- and testis libraries manufactured at the Royal Institute of Technology in Stockholm 129. 1,136 probes originated from the BBSRC Gallus gallus EST database and these were included because of their biological functions. The remaining 29 probes were chosen for inclusion because they had previously been identified in a representational difference analysis (RDA). Details regarding cDNA amplification, purification and printing are available through the ArrayExpress microarray data repository using the array accession number A-MEXP-266.
Affymetrix chicken genome array (Paper III)
The Chicken Genome Array (Affymetrix) was used to study gene expression of 33,457 transcripts from more than 28,000 putative genes. Probes targeting the genomes of 17 avian viruses are also present on the array. Probes were designed from sequence information derived from GenBank, UniGene (Build 18; 15 May 2004), and Ensembl (version 1, released May 2004). Probe sets on the array were designed with 11 oligonucleotide pairs targeting each transcript.

Microarray laboratory procedures and flagging of spots (Paper II)
Purified RNA samples originating from all twenty individuals were mixed and aliquoted into 40 reference RNA samples. 20 μg RNA from each individual and 20 μg reference RNA were subjected to cDNA synthesis using reverse transcriptase (Superscript II, Invitrogen) for 2h at 42°C, during which deoxyribonucleic acids and aminoalyl-labeled Uracil were incorporated into cDNA-strands. Remaining RNA was removed from the cDNA sample by NaOH mediated hydrolysis (protocol available at www.ktharray.se, protocol SOP 002). Purified aminoalyl-labeled cDNA was purified on silica membrane spin colon (Qiagen, Hilden, Germany), after which one of the fluorophores Cy3 or Cy5 (Amersham biosciences) were attached to the aminoalyl at Uracil residues. Two subsequent silica membrane spin column purification steps were performed to remove unbound fluorophores from cDNA. In the following step, individual cDNA-samples were hybridized together with the differently labeled reference cDNA on the microarray. Each individual sample was subjected to two hybridizations (dye-swap hybridizations), i.e. the samples were all labeled with Cy3 in one experiment and with Cy5 in another experiment. The hybridizations were performed in five separate batches; each batch included samples from all groups of individuals i.e. (WL-females, WL-males, RJ-females and RJ-males) and also included an equal number of Cy3 and Cy5 labeled samples. Microarrays were subjected to 18 h of hybridization at 42°C and were subsequently scanned with a G2565BA DNA microarray scanner (Agilent technologies, U.S.A.). Scanning was performed with photo multiplier tube (PMT) settings that gave balanced signals from the two channels. Data-files containing raw fluorescence data (TIFF-files) were imported into the software GenePix (Molecular Devices Corp., Union City, CA, U.S.A.), in which spot identification, manual examination of the surface of the array and flagging of spots/regions with poor quality were all performed in GenePix.

Microarray laboratory procedures (Paper III)
From each individual five μg purified RNA was subjected to first- and second strand synthesis and subsequently cRNA synthesis according to the original Affymetrix protocol. A hybridization mastermix was generated ac-
According to the Affymetrix protocol and 29 μl of each fragmented cRNA sample was added to 271 μl of the hybridization mastermix. Each array was filled with cRNA hybridization mix originating from one individual. The arrays were incubated in a revolving hybridization oven at 45 °C for 16 hours. After completion of hybridization, the hybridization solution was removed and the chicken genome arrays were completely filled with 200 μl nonstringent buffer A. Washing and staining were carried out in a GeneChip FS-450 fluidics station (Affymetrix) according to the original Affymetrix protocol. Finally, the hybridized, washed, and stained GeneChips were scanned on a GeneChip Scanner 3000 (Affymetrix). For each array, filtration was performed on spots that had been flagged either manually or by the default settings in the GenePix software. Subsequent filtering was performed for the size of spots, background vs. foreground signal intensity, intensity ratio of fluorescence from the two channels and saturated spots.

Statistical methods

Paper II

All microarray analysis steps were conducted with the computer software R version 2.4.1 \(^{130}\) with the additional KTH-package, which can be retrieved from the website: [http://www.ktharray.se](http://www.ktharray.se). To normalize signal intensity over the surface of each individual slide and thereafter to normalize signal intensity between all slides, print-tip-lowess normalizations were applied \(^{131}\). Approximately 70% of all spots on the arrays were retained after filtration and normalization. Differentially expressed genes were identified using a B-statistic available in LIMMA \(^{132}\). The "B-value" assigned to each gene is the log posterior odds ratio of differential expression versus non-differential expression \(^{105, 106}\). Correlations between within-array replicate spots were integrated in the statistical model as described in \(^{133}\). In total, three contrasts between WL strain L13 and RJ were performed (male vs. male, female vs. female, and all L13 vs. all RJ). In each contrast, individuals belonging to the same group were treated as biological replicates. In the statistical test, probabilities for differential expression were determined for each probe, with these P-values then being adjusted for false discovery rate (FDR) \(^{108}\). Probes having obtained FDR-adjusted P-values (q-values) < 0.015 were considered differentially expressed.

Paper III

Affymetrix microarray image data were analyzed with Affymetrix Microarray Suite v5.0 using default parameters. All statistical analysis was conducted with the computer software R version 2.4.1 \(^{130}\). Normalization of microarray data was carried out using the RMA (robust multiarray average) option \(^{134}\) available through the affy-package \(^{135}\). Furthermore, for each
probeset the probability for presence/absence of the transcript was assessed with the MAS5 function \(^{136}\) which uses the Wilcoxon signed rank-based gene expression presence/absence detection algorithm as implemented in the affy-package. Only probesets for which at least one individual had _p_ < 0.05 for a detection call were retained for further analyses. Normalized expression values from each individual were assigned a group label (chicken breed and/or chicken breed and sex). Analysis of variance (ANOVA) tests were performed in order to identify differentially expressed (DE) genes in three contrasts (all L13 vs. all RJ; L13 males vs. RJ males; and L13 females vs. RJ females).

To address this multiple testing problem we determined, for each contrast, the significance threshold that corresponded to a false discovery rate (FDR) of 0.05. The FDR at a given significance threshold was estimated by performing ANOVA on each of 1000 randomly re-sampled datasets obtained by permuting group labels. FDRs for specific P-values were then calculated by: (median number of DE probesets in 1000 permutations) / (number of DE probesets in the real permutation). The significance thresholds for calling a probeset “DE” in the three contrasts between RJ and L13 were set at the P-values that corresponded to an FDR of 0.05.

Real time Polymerase chain reaction

_Papers II and III_

Quantitative PCR (qPCR) was performed for the MGP gene in an iCycler instrument (BioRad laboratories) with CybrGreen dye kit (Eurogentec). For analysis of all other transcripts Taq-man assays were used and the samples were analyzed on the ABI Prism 7900 Taqman instrument (Applied Biosystems) RNA-samples were reversely transcribed with the High Capacity cDNA reverse transcription kit (Applied Biosystems). Individual cDNA-samples were diluted 30-fold with nuclease-free water and a dilution series was generated for a reference chicken cDNA-sample. PCR reactions were performed with an initial polymerase activation step at 95° C for 10 minutes, followed by 40 cycles of 95° C for 15s and 60° C for 60s where fluorescence intensity was recorded during each annealing step. Diluted cDNA-samples from each individual were analyzed in triplicate (duplicates for MGP) and reference dilutions were analyzed in duplicate. Standard curve method was used for relative quantification of target abundance. _Glyceraldehyde 3-phosphate dehydrogenase (GAPDH)_ was chosen as reference gene and for each individual the relative expression level of a transcript was normalized relative to _GAPDH_-expression. Differential expression was assessed by Student’s t-tests.
Table 1. Primers and probes used in real time-PCR analyses

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>Reporter Probe</th>
</tr>
</thead>
<tbody>
<tr>
<td>WDR5</td>
<td>CTCTTAAGCTTTGGA</td>
<td>AGTACTCTCATTTGT</td>
<td>TCCCTGCCACTTTCC</td>
</tr>
<tr>
<td>PTK1</td>
<td>TCCGCCAGAGGATGG</td>
<td>GGGACCTACACGTT</td>
<td>CAAGGCAACCTCCTTA</td>
</tr>
<tr>
<td>TGOLN2</td>
<td>CGGGCGGGCCTACATG</td>
<td>GGCGACTCATCCACT</td>
<td>CCCTGGATCAAAGATC</td>
</tr>
<tr>
<td>WIFI</td>
<td>CGGATTCGAAAGGAC</td>
<td>TTCACCCAACATTTTC</td>
<td>AGTATGCATTATTGAT</td>
</tr>
<tr>
<td>GAPDH</td>
<td>GGAGTCCACTGGTCT</td>
<td>GTGAAGTATAACAGC</td>
<td>GCTGAGATGATAACAG</td>
</tr>
<tr>
<td>EDIL3</td>
<td>CTTACAGCTTTCCTACAG</td>
<td>AT</td>
<td>TTTCCCTTACACTCA</td>
</tr>
<tr>
<td>CDC8</td>
<td>GCTTCCATACCGGTCTT</td>
<td>AAAGGATCGTCAAAATTCA</td>
<td>CCAAGGATACGAT</td>
</tr>
<tr>
<td>LIMS1</td>
<td>GCTGGGTCTTAGAGG</td>
<td>ACCACACTCGTGGCAGCA</td>
<td>CCCAGACACATCC</td>
</tr>
<tr>
<td>NCE2</td>
<td>CATT</td>
<td>CAAAGGATCTGACCAATTCAG</td>
<td>TTT</td>
</tr>
<tr>
<td>NDUF7</td>
<td>CAGGCTCCTGACACT</td>
<td>GTGAGGTGTGGAGAGATC</td>
<td>CTGGAGGTCGCGCCC</td>
</tr>
<tr>
<td>MGP</td>
<td>AGAGAGAGGATG</td>
<td>GTTG</td>
<td>No reporter probe</td>
</tr>
</tbody>
</table>

Gene Ontology (GO) overrepresentation analysis (Paper II)

Clones on the microarray were attributed Gene Ontology (GO) terms by BLAST-analysis of all individual probe sequences against GO peptide sequences \(^{123}\). BLAST search hits with a similarity cut-off of greater than \(E = e^{-10}\) were retained for use in a modified version of the EASE-software \(^{137}\). In EASE, GO-term enrichment analysis was performed for probes differentially expressed between all WL and all RJ. In the analysis, each GO-term was assigned an EASE-score which is a conservative adjustment of Fisher's exact probability utilizing a jackknife approach. Enriched GO-terms (EASE-score <0.05) comprising six or more DE probes on the microarray were retained for further analysis.

QTL-mapping studies (Papers I and IV)

DNA-isolation, genotyping and Linkage Map Construction (Paper I)

DNA was isolated from blood samples collected from all F\(_2\) individuals, their parents (F\(_1\)), and grandparents (F\(_0\)). All animals were genotyped for 164 markers as specified in \(^{138}\). The sex-averaged map spanned 3356 centiMorgans (cM), with an average marker spacing of 20.98 cM. Linkage maps for 30 (29 autosomal and the Z chromosome) linkage groups were constructed using the CRIMAP software \(^{139}\).
Linkage Map Construction (Paper IV)

DNA was isolated from blood samples that had been collected from all F2-individuals, their parents (F1), and grandparents (F0). 356 SNP markers, of which 344 had information contents > 0.5, were genotyped in all F0, F1 and F2 individuals using the Illumina Golden Gate Assay (Illumina). The SNP markers covered 29 autosomes and the sex chromosome Z. The sex-averaged map was constructed in CRIMAP \(^{139}\) and spanned 2359 cM, giving an average marker spacing of 6 cM.

Statistical methods

QTL-mapping (Papers I and IV)

QTL-mapping and significance testing for the two separate F2 intercrosses were performed using a standard interval mapping method \(^{140}\) using least squares regression \(^{141}\), with these being implemented using the QTL Express software \(^{142}\). This package allows additive and dominance effects to be modeled. As well as such effects, sex interaction effects were also modeled, to detect QTL differentially expressed between the sexes. Fixed effects included in the model were rearing batch and sex, whilst body weight at 200 days was included as a growth covariate.

Significance thresholds for QTL were calculated by the permutation testing option available in QTL Express, with these values thereby tailored to the individual data sets. The threshold for significant QTL was set at a 5% genome wide significance level. A threshold of 20% genome-wide significance was used to infer suggestive QTL, with this being used in several studies previous to this \(^{143-145}\). QTL-mapping using Z-chromosome data was only performed in Paper I, using the Qxpak v.2.13 software \(^{146}\). The confidence intervals (CI) of QTL were defined as the region confined by two markers flanking a LOD-score drop of one at both sides of QTL peaks.
RESULTS AND DISCUSSION

Paper I

An intercross pedigree was created by mating red junglefowl (RJ) to White Leghorn strain L13 (L13). In phenotypic analysis of parental individuals, a 50% higher BMD measured by DXA was seen for L13 females compared to RJ females (Figure 8). QTL-mapping for femoral bone traits as well as whole body BMD revealed four significant QTL at the 5% genome-wide level. These QTL were distributed on chicken chromosomes 1 (n=2), 2 (n=1) and 20 (n=1) and had effects on both structural and biomechanical femoral bone traits. In addition to these four significant QTL, ten other QTL were significant at the 20% genome-wide level, indicating suggestive linkage to the measured traits. F-values of selected QTL are presented along chromosomes in Figure 9.

Figure 8. Phenotypic distributions of F2-individuals from the RJ x L13 intercross. Mean values for RJ and WL individuals (five of each sex and strain) are indicated by arrows.
Two of the identified QTL had what appeared to be female specific effects. QTL *nc-BMD1* on chr. 1 was only significant for non-cortical BMD of the femoral metaphysis, which represents a mix of trabecular- and female specific medullary bone. The domestic L13-allele at *nc-BMD1* appeared to act in a dominant manner, and gave female bearers lower BMD regardless of allele dosage. The sex-specific QTL *BMD1* was also female-specific and a higher BMD measured both by DXA and pQCT was observed for females carrying the domestic L13-allele, which acted in an additive manner. QTL *tors1* also had sex-dependent effects, but rather than having effects only in one sex, this QTL had effects going in opposite directions in the two sexes. In males the domestic allele gave more elastic femurs that were able to rotate more until the bone broke in a torsion strength test. In females *tors1* had opposite effects, with femurs of wild-type allele bearers exhibiting a higher degree of rotation until the femur failed. The fourth significant QTL (*ecirc1*) was located on chr. 1, and the domestic allele at this locus gave bearers a larger endosteal circumference of the femoral diaphysis in a sex-independent additive manner. This region had previously been shown to contain a RJ x L13 QTL affecting many traits \cite{143,144}, including bodyweight which was substantially higher in individuals carrying the domestic allele \cite{138}. This suggests pleiotropic effects of this locus, but QTL-mapping was conducted with adult bodyweight as a covariate in the analysis which suggests that the *ecirc1* locus had effects on endosteal circumference that were not entirely dependent on body weight. An additional locus, on chr. Z showed suggestive linkage for femoral BMD measured by DXA as well as for metaphyseal BMD measured by pQCT.

In conclusion, the identified QTL affected several important parameters of bone and the future identification of the causative genes will further our understanding of the genetic basis of bone phenotypic variation.
Figure 9. F-values along chromosomes where QTL were identified in the intercross between wild-type red junglefowl and domestic White leghorn strain L13. A) QTL identified with sex-interactions modeled, thereby allowing effect of QTL to differ between sexes. B) QTL identified in standard analysis.

In this study, cDNA-microarray technology was used to compare gene expression in femoral bone from RJ and L13. A total number of 837 probes, corresponding to 779 unique transcripts were identified as differentially expressed (DE) between RJ and L13. In the sex-independent comparison 604 microarray probes indicated DE. In comparison between males of the two strains, DE was observed for 410 probes and the corresponding female comparison revealed DE for 270 probes. Close to two thirds of the probes showing DE-between females of the two strains had lower expression in L13-females, which is a highly significant bias ($\chi^2$, d.f.=1, $P<0.001$). There was no apparent difference in the direction of differential expression between males, where 48 % of probes showed lower expression in L13. Female femurs were more heterogeneous than male ones, which could explain why more DE transcripts were observed between males of the two populations.

Nine DE genes were chosen for verification of DE by quantitative PCR (qPCR). qPCR-analysis revealed statistically significant DE ($P < 0.05$ in Student’s t-test) between L13 and RJ for all nine transcripts, thereby suggesting true DE of many genes on microarray (Figure 10).

No commonly used marker genes for bone cells were identified as differentially expressed between RJ and WL in the sex-independent contrast, nor in the two sex-specific contrasts. However, some genes for which differential expression was observed have previously been attributed roles in bone metabolism. For example WD-repeat containing protein 5 (WDR5), which has been found to accelerate osteoblast and chondrocyte differentiation and whose expression is induced by bone morphogenic protein 2 (BMP2) $^{147, 148}$, L13-individuals expressed WDR5 in lower amounts than did RJ-individuals.

The Gene Ontology (GO) terms “cytosolic large and small ribosomal subunits” were identified as overrepresented among DE genes in GO overrepresentation analysis. Upon examination of the genes corresponding to this overrepresentation it was seen that 15 separate ribosomal protein genes, were responsible for this overrepresentation and were all expressed in higher levels in RJ. This may be interpreted as a perturbation in protein biogenesis pathways between the two lines.

Of 779 unique DE transcripts, 57 were found to be localized within confidence intervals of previously identified RJ x L13 bone trait QTL. One of these genes, Wnt inhibitory factor 1 (WIF1) was identified as DE between females of the two strains, with L13 females having higher levels than RJ. Transcription of WIF1 is up-regulated during BMP2 induced osteoblastic differentiation of murine C2C12 and MC3T3 cells $^{149, 150}$, implicating the protein as important in the process of osteoblast differentiation. Furthermore, WIF1 has been reported to be expressed in trabecular but not cortical bone of mature mice $^{149}$. In chicken WIF1 is located within the confidence interval of a bone trait QTL-region originally identified for body weight and growth $^{138}$.
and also has pleiotropic effects on many bone traits\textsuperscript{151}. \textit{Peroxisomal D3, D2 -enoyl-CoA isomerase (PECI)}, a gene situated within the CI of QTL-region \textit{bmdl} on chr.2 was expressed in higher levels in RJ, as was \textit{Ribosomal protein L18A (RPL18A)} located to QTL-region \textit{ecirc1}. The GO-overrepresentation observed for ribosomal subunit proteins among DE genes may implicate \textit{RPL18A} as a potential QTL-candidate gene. One differentially expressed transcript was also highly similar to an avian endogenous retroviral insertion localized within the confidence interval of the pleiotropic QTL \textit{ecirc1}. It has been shown that several isolates of avian myeloblastosis-associated virus (MAV) and avian leukosis virus (ALV) can induce the bone disorder osteopetrosis and affect growth in chicken\textsuperscript{152}, making this viral insertion an interesting QTL-candidate.

Based on gene function as well as magnitude and statistical significance of DE, the best candidates among identified DE-genes included: \textit{WIF1}, \textit{PECI}, \textit{RPL18A} and the retroviral insertion. The numbers of probes on the microarray were substantially fewer than the 20,000-23,000 genes proposed to be present in the chicken genome\textsuperscript{83} and it is therefore possible that some genes important to bone metabolism may not have been represented on the microarray. The microarray data appears to be largely accurate but a potential bias may have been introduced by few numbers of probes present on the microarray.

![Figure 10](image)

Figure 10. Real-time PCR analyses verified differential expression of nine genes that were identified as differentially expressed in analysis of cDNA-microarray data.
Paper III

Prior to this study, the RJ x L13 intercross population studied in \(^{151}\) had been genotyped for an additional 351 SNP markers. QTL-mapping with this updated map had in a separate study enabled detection of several additional QTL (Wright et. al. manuscript).

Gene expression in femoral bone was compared between 100 week old RJ and L13 using high coverage oligonucleotide microarrays. In contrast to the cDNA-arrays previously used for this purpose \(^{153}\), these microarrays enable expression profiling of virtually all presumed chicken genes. In total 520 genes were identified as differentially expressed (DE), using a permutation based significance cut-off that was likely more stringent than the one used in our previous study \(^{153}\). An additional 40 genes were also considered DE as they had, in addition to P-values < 0.05, also mean expression levels at least twice as high in one population, rendering a total of 560 DE genes.

Only thirty-five genes were identified as DE both in this study and in the cDNA-microarray study \(^{153}\). Genes that obtained DE P-values < 0.10 (n=106) and < 0.05 (n=60) in both platforms had correlation coefficients of 0.63 and 0.66, respectively (Figure 11). It should be noted that the chickens analyzed on the two platforms differed in age, which could explain that only thirty-five genes were DE on both platforms. The high correlation coefficients between probes with P-values < 0.5 and < 0.10 suggest that a high proportion of these are DE between the two populations.

As had previously been observed in the cDNA-microarray study, WD-repeat containing protein 5 (WDR5) was identified as DE. It is possible that WDR5 is involved in the manifestation of bone phenotypic differences between the two lines, possibly by altering osteoblast differentiation. No QTL was found where WDR5 resides on chromosome 17, suggesting that WDR5 expression...
may be regulated by a *trans*-acting gene product expressed from a bone trait QTL-region.

The gene *bone morphogenic protein 2* (*BMP2*) is located on the border of the confidence interval of a QTL for whole body bone mineral density (*BMD2*) on chicken chromosome 3, and because of the known relationship between *BMP2* and *WDR5*, it is possible that genetic variation in proximity of *BMP2* may be causative of the QTL. *BMP2* was among the genes identified as DE and was like *WDR5* expressed in higher levels in RJ. Expression levels of *BMP2* and *WDR5* were highly correlated (R²=0.81), which supports the theory that DE of *WDR5* may be secondary to that of *BMP2*. The one-LOD drop confidence intervals that were used to define QTL may in this case have been too stringent, leaving *BMP2* outside of the confidence interval although it may be causative of the QTL.

Medullary bone lines the endosteal surfaces of long bones in the reproductive female chicken, but is not present in males. pQCT-analysis revealed that densities of medullary bone in right femurs of female birds were correlated with expression levels of *Wnt inhibitory factor 1* (*WIF1*) in the left femurs. The gene *Kelch-like 14* (*KLHL14*) was expressed at higher levels in female L13 than in female RJ, as had previously been observed in analysis using cDNA microarrays. When consulting phenotypic data it was noted that *KLHL14* levels, like those of *WIF1*, were correlated with the density of medullary bone, but appeared only to be so in L13 femurs. *KLHL14* is located in a RJ x L13 QTL-region (*BMD1*) for which the L13-allele gave females a higher BMD. *WIF1* is involved in wnt-signaling, a signaling pathway important to osteoblast maturation and bone development, and this may suggest that *KLHL14* is also involved in this pathway.

Seventy-two DE genes were found to be located to within confidence intervals (CIs) of QTL identified in analysis with the updated genetic map. Transcript levels of the *Matrix GLA Protein* (*MGP*) gene were several orders of magnitude higher in femoral bone from L13 than in RJ. The MGP protein acts as a potent inhibitor of calcification in bone as well as in several other tissues and *MGP* gene knock-out mice die within two months of birth due to arterial calcification. qPCR-analysis verified strong DE of *MGP*, with approximately seven-fold higher transcript levels observed in femurs of L13-individuals (Figure 12). *MGP* is located to the RJ x L13 QTL-region *ncBMD1*, possibly indicating that the observed DE of *MGP* is relevant to the variation in non-cortical BMD conferred by this locus.
Differentially expressed genes regarded as QTL-candidates due to previously described importance to bone metabolism include: *Matrix GLA Protein (MGP)*, *Trafficking protein particle complex 2 (TRAPPC2)*, *PDZ and LIM domain 7 (PDLIM7)*, *Peroxisomal biogenesis factor 7 (PEX7)*, genes involved in BMP2 signaling pathways and a gene presumed to encode an osteoblast membrane protein in chicken. However, other DE genes, such as *Kelch-like 14 (KLHL14)* and *Casein Kinase 1 epsilon (CSNK1E)* are also co-localized with QTL and may prove to have roles important to bone metabolism.

**Paper IV**

In this study, the genetic contribution to differences in femoral bone traits was examined in an intercross between RJ and the WL Obese Strain (OS). In QTL-mapping resulted in twelve separate QTL identified as significant on the 5% genome-wide level. In addition to these, two separate suggestive QTL were identified, rendering a total of 14 loci linked to femoral bone traits. Significant QTL were located to chromosomes: chr. 1 (n=3), chr. 2 (n=2), chr. 3 (n=2) and one each was identified on chromosomes 4, 5, 7, 12 and 28. Suggestive QTL generally clustered to significant QTL-regions but two separate loci on chromosomes 1 and 16 harbored only suggestive QTL. The identified QTL affected various bone traits, some in a sex-dependent manner. Identified significant QTL covered 147 Mb of genome sequence and the corresponding coverage was 269 Mb for all significant as well as all
suggestive QTL. These figures correspond approximately to 15 % and 28 % of the autosomal chicken genome, respectively.

\[ \text{PERIOSTEAL CIRCUMFERENCE (mm)} \]
\[ \begin{array}{ccc}
27 & 26 & 25 \\
RJ/RJ & RJ/OS & OS/OS \\
\end{array} \]

\[ \text{ENERGY TO FAILURE (N*mm)} \]
\[ \begin{array}{ccc}
1500 & 1000 & 500 \\
RJ/RJ & RJ/OS & OS/OS \\
\end{array} \]

\text{QTL7 GENOTYPE} \quad \text{QTL10 GENOTYPE}

\text{Figure 13.} Phenotypic values observed in male RJ x OS F_2-individuals for two femoral bone traits. The individuals (dots) are stratified by alleles (RJ or OS) inherited at two QTL (\textit{QTL7} on chr. 3 and \textit{QTL10} on chr. 7). The OS-allele at \textit{QTL7} gave bearers larger diaphyseal circumferences in an additive manner. Presence of an OS-allele at \textit{QTL10} gave weaker femurs, failing at lower energies than femurs of RJ-allele homozygotes, which indicates dominant action of \textit{QTL10}. For both QTL the same effects were seen for female femurs, but only males were included in graphs due to large phenotypic differences between sexes for these traits.

Interestingly, nine of 14 RJ x OS QTL had genomic confidence intervals (CIs) overlapping those of QTL previously identified in the RJ x L13 intercross. Nineteen QTL had previously been identified on autosomes in the RJ x L13 intercross (\textit{Wright et. al., manuscript}), and their combined CIs covered 16 % of the autosomal genome. Only for few QTL-overlaps identical phenotypic traits were responsible for QTL-signals in the two intercrosses. Partly, this could be explained by different methodologies having been used for phenotypic measurements in the two studies. It is also important to consider that phenotypic manifestation of genetic variants on traits such as those of bone can be dependent on the genetic background \textsuperscript{158} and also on the environment in which the variant is present \textsuperscript{39}. Therefore, the same gene or even the same segregating genetic variation could have shown linkage to different phenotypic measurements in the two intercrosses. Principal component analysis (PCA) of phenotypes in the RJ x OS intercross showed that several phenotypes were highly inter-correlated, suggesting that they are at least partly governed by mutual genetic variants. For example, strong positive correlations were observed between size traits such as bodyweight, femur
length, and area of the metaphysis. Diaphyseal size traits were negatively correlated with traits such as cortical thickness and diaphyseal BMD. This means that a QTL identified as significant for one measured trait is likely to give a signal also for other highly correlated traits.

Bearing this in mind, we sought to investigate if overlapping QTL affected inter-correlated traits in the two intercrosses. We concluded that similar phenotypes were affected by at least seven and possibly eight overlapping QTL (Figure 14). The directions of effects conferred by WL- and RJ-alleles appeared to be concordant for at least seven QTL.

![Figure 14. Overlaps between bone trait QTL identified in two intercrosses between domestic White Leghorn and wild-type red junglefowl chicken. Confidence intervals (CIs) of significant QTL are represented by solid lines and CIs of suggestive QTL are represented by dashed lines. Positions along the five chromosomes are presented as centiMorgan (cM) positions in the RJ x OS intercross. cM positions were translated to Megabase (Mb) positions and these were used for placement of RJ x L13 QTL. Only the peak is presented for the RJ x L13 epistatic QTL on chr. 5.](image)

The domestic lines OS and L13 both originate from early White Leghorns. White Leghorn L13-line has been specifically selected for egg-weight since the beginning of the 1970’s and compared to RJ females they lay more that twice as many eggs per week, each weighing more than twice as much as RJ-eggs. The female bird relies on calcium mobilized from skeletal stores for deposition in eggshells. Selection for egg-laying may therefore have taken molecular routes involving fixation of alleles conferring for example higher rates of bone turnover or larger skeletal stores.
The OS-strain was established by selection for spontaneous autoimmune thyroiditis, which occurred in the 1950’s in a White Leghorn population. All but one batch of F₂-individuals analyzed in study received dietary thyroxin supplementation to counteract the disease. Hypothyroidism and hyperthyroidism have both been associated with reduced bone mass in studies of human subjects and it is thus conceivable that bone phenotypes under study could have been affected by the disease or by the thyroxin supplementation. Level of T-cell infiltration in the thyroid gland, which is indicative of the autoimmune disease, was monitored in all F₂-individuals. It was found to be non-correlated with all bone traits analyzed, suggesting that bone traits were at least not heavily influenced by the autoimmunity.

The high degree of QTL-overlaps between studies indicates that these loci are important to bone phenotypic variation, independent of, or at least not heavily influenced by the autoimmune disease in the OS-line. Separate RJ populations were used in the establishment of these two intercrosses, which should have decreased the risk of deleterious RJ-alleles being responsible for overlapping QTL.

Genomic co-occurrences of QTL may indicate that the same segregating genetic variation is responsible for QTL-signal in both intercrosses, or could alternatively indicate that genes important to bone phenotypic variation in chicken are enriched to these loci. The former indication is strengthened by the high similarity between QTL-confidence intervals as well as by concordant phenotypic manifestations for at least seven overlapping QTL. We hypothesize that overlapping QTL are, at least in some cases, caused by alleles that were fixed during the mutual domestic history of the OS and L13-lines. At least 50 years of genetic recombination separates the two WL-lines, which should facilitate fine mapping of QTL caused by shared domestic haplotypes.
CONCLUSIONS

In the studies included in this thesis QTL-mapping and gene expression microarray analyses have been used to identify genomic regions and genes involved in the establishment of differences in bone traits in chicken intercrosses. Two separate intercrosses were used, both of which were generated by crossing wild-type red junglefowl (RJ) to domestic White Leghorn (WL) chicken. Several QTL for bone traits were identified in the intercross between RJ and WL strain L13 (L13). In two separate studies, gene expression in femoral bone was compared between RJ and L13. The gene expression analyses resulted in the identification of several interesting candidate genes and also indicated specific signaling pathways as possibly altered during the course of domestication. QTL-mapping in an intercross between RJ and the Obese strain (OS) of WL revealed that a high proportion of bone trait QTL overlapped those previously identified in the RJ x L13 intercross. The presence of such overlapping QTL indicates that these loci contain genetic variation important to bone phenotypic variation in chicken. The future aim is to limit the list of potential causative genes in chicken and to investigate in parallel if genetic variation in the human counterparts affects traits relevant to osteoporosis.
GENERAL DISCUSSION AND FUTURE PERSPECTIVES

It is well-established that heritable factors are important for bone phenotypic variation in humans, with effects on both bone mass and risk of fragility fractures. Elucidating the genes underlying these heritable traits can reveal so far unknown cellular pathways affecting bone, and may eventually lead to novel therapeutics for bone disease and/or to the possibility of earlier identification of individuals at risk of osteoporosis. Several genomic loci have been implicated as important to human inter-individual variation in bone traits, but the identification of specific genes responsible for these differences has so far proven difficult. Intercrosses between strains of laboratory animals, most often mice, have frequently been used to trace the inheritance of bone characteristics.

Domestic animals have generally been separated from their wild ancestors for thousands of years. Using domestic animal models in quantitative genetic studies may be very useful when dissecting complex traits. In the papers included in this thesis, chicken was studied as a model species. Humans and chickens have a similar bone metabolism, and share a high degree of gene synteny, apparent after the recent chicken genome sequencing. Chickens can have a large progeny, and have a short generation time. In addition, a significant proportion of commercially bred chicken develop osteoporosis, which makes chicken an even more attractive model for dissecting the genetics of bone phenotypes.

The results from the studies presented in this thesis show a number of significant and suggestive QTL for bone phenotypes. In addition, analyses of mRNA expression in femoral bone from parental animals revealed differential expression between animals with high bone mass and strong bones versus those with low bone mass and fragile bones. However, the numbers of genes in QTL are substantial, and mRNA expression analyses results in lists of genes ranked by statistical significance for differential expression between the populations. This means that we are left with quite a large number of potentially interesting candidate genes for effects on bone metabolism in chicken. The challenging task is to limit the list of potential candidate genes to a manageable number, and this will be addressed in a number of different ways. The ultimate aim is to translate the findings from our functional ge-
nomics studies in chicken into man, in order to assess effects of these genes on human bone phenotypes and bone metabolism.

Already mentioned in the thesis are analyses of the overlaps of QTL found in the RJ x WL and RJ x OS intercrosses, decreasing the genetic regions of interest further. Another resource is to compare the mRNA expression profiles of organs other than bone from the parental strains. The liver is a central organ for many metabolic processes and gene expression profiling in liver may aid in identifying metabolic pathways for which RJ and WL differ. Gene expression profiling in hypothalamus/thalamus may reveal perturbations in the central control of bone metabolism, as has been described in the case where the protein Leptin controls the extent of bone formation by modulating the proliferation of osteoblasts 164.

Differentially expressed genes residing in QTL-regions, whose DE is caused by cis-acting factors are prime candidates for being causative of QTL. Gene expression will be analyzed for selected genes in femoral bone from RJ x L13 F1- and F6-individuals, either by real-time PCR or by global expression analysis (e.g. microarrays). RNA has been isolated from femoral bone of 100 RJ x OS F2-individuals. This will enable eQTL mapping, where expression levels of certain genes identified in the microarray analyses will be used as phenotypes in QTL-analyses. Such analyses could be performed either on single genes by the use of quantitative PCR or globally through expression profiling with microarrays.

For overlapping QTL between RJ x OS and RJ x L13, candidate genes and immediately proximal regions to these will be examined for the presence of shared haplotypes between the two domestic lines. These two lines may be fixed for haplotypes that still segregate among, or are not present in wild-type chicken. DNA panels comprising divergent breeds of domestic chicken may also be used for this purpose as it is conceivable that other strains of White Leghorn and also other domestic breeds share selective sweeps. In light of recent advances in DNA-sequencing technology, it is not unlikely that the coming years will bring forth methods that will enable high throughput sequencing of whole genomes at low costs. This would immensely facilitate the discovery of selective sweeps.

Primary cultures of bone derived cells from man as well as from chicken will provide opportunities to study the effect of partial knockdown of these genes. In pilot studies small interfering RNAs (siRNAs) were successfully introduced into primary bone derived cells cultured from both human and chicken bone tissue. The effects of candidate gene silencing can be studied in assays where mineralization proliferation and viability are measured. Silenced genes in human cells so far include WIF1 and MGP. The global effects of these gene silencing experiments will be assessed by comparison of mRNA from siRNA transfected- and negative control transfected cells by gene expression microarrays.
Prime candidate genes derived from the chicken studies have been and are currently further examined for association to bone traits in our large population based cohorts of men and women using SNP association studies. Preliminary data indicate that candidate genes derived from the chicken model are associated to differences in bone phenotypes in human.

The results presented in this thesis show that functional genomic studies of domestic animals is a successful approach for identifying candidate genes for involvement in bone metabolism.
SAMMANFATTNING PÅ SVENSKA

Benskörhet
Benskörhet eller osteoporos som är den medicinska benämningen är en folk-
sjukdom som främst drabbar den äldre delen av vår befolkning och bland
dem främst kvinnor. Sjukdomen beror på att kroppens skelett bryts ned, får
en förändrad struktur och blir försvagat, vilket kan leda till benbrott. Det har
uppskattats att så många som varannan kvinna och var fjärde man under sin
livstid kommer att drabbas av ett benbrott orsakat av benskörhet. Efter kli-
makteriet sänks nivåerna av könshormonet östrogen, som i kroppen har en
beskyddande effekt på skelettet, vilket leder till en ökad nedbrytning av ben-
vävnad hos kvinnor efter denna fas i livet.

Vad bestämmer risken för benskörhet
Egenskaper som massa och geometri hos kroppens ben kan studeras med
speciella röntgenapparater som läser av skelettet längs med kroppen. I studi-
er där benets egenskaper undersöks hos enäggstvillingar har man sett att
ärftliga (genetiska) faktorer spelar en stor roll för vilken bentäthet man får.
Uppskattningsvis förklarar genetiska faktorer så mycket som 65-80 % av
skillnaden i bentäthet mellan två i övrigt friska personer av samma kön, me-
dan motsvarande siffror för att drabbas av benskörhetsrelaterat benbrott är 25-
50 %. Den viktigaste prediktiva faktorn för benbrottsrisk är tidigare benbrott.
Bland egenskaper som är mätbara för alla individer är dock benets täthet den
viktigaste, även om risken för benbrott även påverkas av icke ärftliga fakto-
er (miljöfaktorer), som t.ex. att falla. Förutom de ärftliga faktorerna påver-
kas benens täthet och struktur även av miljöfaktorer såsom kosthållning,
motion, vissa livsstilsfaktorer (t.ex. rökning) och kemiska ämnen som medi-
ciner, föroreningar etc. Skillnader i bentäthet mellan två i övrigt friska indi-
vider styrs av att vi har olika varianter av ett stort antal gener, av vilka
många ännu inte är kända. Det är den samlade repertoaren av dessa genvari-
anter (vår ärftliga uppsättning) som tillsammans med miljöfaktorer bestäm-
mer vilken bentäthet vi får.

Det är av stor vikt att specifikt identifiera de ärftliga faktorer som påver-
kar skelettets uppbryggnad respektive nedbrytning samt risken för benbrott.
Detta för att kunna identifiera och sätta in tidig förebyggande behandling av
personer i riskzonen för att utveckla benskörhet, men även för att kunna
utveckla nya effektiva läkemedel som motverkar sjukdomens uppkomst eller
förlopp. Vissa saker komplicerar sökandet efter dessa gener: vissa genvarianter påverkar benmassa endast för kvinnor, andra endast för män, somliga endast i särskilda etniska grupper och vissa andra endast när de förekommer i kombination med specifika varianter av andra gener eller tillsammans med specifika miljöfaktorer. Noggranna undersökningar av arvsmassa, benegen-skaper och livsstilsfaktorer för tusentals personer av samma kön, ålder och etnisk tillhörighet behövs för att statistiskt säkerställa effekten av enskilda genetiska varianter i människa.

**Kycklingen används som modell**

Mekanismerna som styr uppbyggnad och nedbrytning av benvävnad är liknande mellan människa och andra ryggradsdjur och vissa regioner av arvsmassan likaså. Höns får en stor avkomma, har korta generationstider och kan därmed korsas och avlas för att snabbt uppnå önskade egenskaper. Tamhöns härstammar från den röda djungelhönan som lever i fritt tillstånd i Sydostasien. Människan har i flera tusen år hållit höns i sin närhet för att kunna få tillgång till deras kött och ägg. Under dessa årtusenden har avel lett fram till en mångfald av nu levande hönsstammar, somliga specifikt framavlade för att lägga maximalt antal ägg, andra för att snabbt växa sig stora och muskulösa. Hos en del stammar av tamhöns har avel för dessa karakteristika gett upphov till stora förändringar i benmassa och benägenhet att få benskörhet.

**Målbeskrivning för avhandlingens arbeten**

Det översiktliga målet med de bedrivna studierna har varit att i arvsmassan hos höns hitta de genetiska regioner och i förlängningen de gener som bidrar till variation i benegen-skaper. Ett framtidiga mål är att sedan undersöka om dessa gener är viktiga också i människa.

- Genom att korsa röd djungelhöna med höns framavlade för äggläggning (Vit Leghorn) hade vi som mål att kartlägga de ärfliga faktorer som gett upphov till skillnaderna mellan arterna och dem emellan för egenskaper som täthet, geometri och hållfasthet av benen. (Arbeten I och IV).

- Undersökningar av lårben från röd djungelhöna och Vit Leghorn genomfördes med målet att identifiera gener vars aktivitet skiljer sig mellan de båda populationerna. (Arbeten II och III).

**Resultat och diskussion**

Med statistiska metoder undersöckes i arbete I om nedärvning av någon specifik region av arvsmassan påverkade vilka benegen-skaper 337 avkom-mor i korsningen mellan röd djungelhöna och Vit Leghorn fick. Resultaten visade att fyra regioner i arvsmassan var starkt kopplade till benegen-skaper och för ytterligare tio regioner påvisades en eventuell koppling. För vissa
av de identifierade regionerna har den mänskliga arvsmassans motsvarande regioner tidigare visat sig vara viktig för variation i benegenskaper.

**I arbete II** jämfördes genaktivitet i lårben för tio stycken 40 veckor gamla individer vardera från röd djungelhöna och Vit Leghorn. Resultaten visade att 779 geners aktivitet vara olika dem emellan. Av dessa gener låg 57 stycken i regioner som identifierats som viktiga i arbete I, och dessa gener föreslogs därför som eventuella kandidater för att orsaka skillnaderna i benegenskaper.

**I arbete III** använde vi oss av en ny teknik som till skillnad från tekniken från arbete II kunde mätta aktiviteten för i stort sett alla gener i arvsmassan hos höns. Totalt hittade vi förändrad aktivitet för 560 gener i jämförelser mellan 100 veckor gamla individer från stammarna röd djungelhöna och Vit Leghorn. Av dessa gener låg 72 stycken i regioner som visat sig vara viktiga för benegenskaper i en fortsättningsstudie av arbete I.

**I arbete IV** etablerades en ny hönskorsning som var snarlik den som användes i arbete I. Skillnaden var att en annan stam av Vit Leghorn korsades med en annan stam av röd djungelhöna. För andra generationens avkommar (554 stycken) från denna korsning bestämdes egenskaper som densitet, storlek, geometri och hållfasthet av lårben och dessutom spärrades den genetiska nedärvningen som skett. I arvsmassan identifierades 14 regioner som påverkade någon eller flera av de uppmätta benegenskaperna. Nio av dessa 14 regioner visade sig vara delade med regioner som tidigare identifierats i en fortsättningsstudie av arbete I. Dessutom var effekten av de delade regionerna i många fall liknande eller samma i de båda studierna. Sådana delade regioner kan betyda att det är samma variationer i arvsmassan som orsakar skilda benegenskaper i de båda korsningarna, vilket skulle underlätta identifieringen av de orsakande generna.

**Sammanfattning**

I två olika korsningar där ursprungliga höns korsats med tamhöns visade vi att variation i minst fyra (**Arbete I**) respektive 14 (**Arbete IV**) regioner av arvsmassan bidrog till skillnader i benegenskaper. En stor del av regionerna som identifierades i de båda korsningarna var överlappande, vilket sannolikt kommer underlätta uppspårningen av de gener där som är viktiga för variation i benens egenskaper. När genaktiviteten i lårbenet undersöckes i två ytterliggare studier kunde vi se att ett stort antal gener hade förändrad aktivitet mellan röda djungelhöns och tamhöns samt att en del av dessa låg mitt i de regioner av kycklingens arvsmassa som påverkar benegenskaper. Förändrad funktion hos gener är en sannolik orsak till förändrade benegenskaper och vi betraktar därför sådana gener, om de ligger i regioner som styr benegenskaper, som kandidatgener för benskörhet.

Vi har nu påbörjat genetiska analyser där tusentals människor undersöks, detta för att undersöka vilka effekter gener identifierade i höns har på variation i benegenskaper i människa.
ACKNOWLEDGEMENTS

The work included in this thesis was conducted at the Department of Medical Sciences, Faculty of Medicine at Uppsala University.

I have interacted with many helpful and friendly people over the past four years and it would be impossible to extend my gratitude to each and everyone. I therefore send out a sincere collective thank you to all whom in one way or another has contributed to this thesis. Especially I would like to thank:

Andreas Kindmark, my main supervisor, for giving me the opportunity to dive into the field of quantitative genetics as a PhD-student. I thank you for delicately balancing mentorship with independent thinking on my part. I will always remember you for your optimistic nature, for your generosity and for all our interesting scientific discussions.

Professor Östen Ljunggren, my co-supervisor, for your vast knowledge in the “bone field” and in what seems like just about any other discipline, be it Evert Taube songs, history or dentistry. Thank you also for sharing my fascination for RNA interference, I am looking forward to us applying it to cure OI once and for all.

Professor Leif Andersson, my co-supervisor, for always sharing your expertise in genetics. It has been a joy to be part of the “chicken group” and truly inspirational to witness and participate in the application of cutting edge genomic techniques up front.

Past and present members of the lab which has now become three; for the sake of brevity and ease of reading here referred to as the “bone lab”. I would like to thank: Majd Mirza, Elin Grundberg, Tijana Krajisnik, Jenny Svensson, Katarina Lindahl, Richard Marsell, Tobias Larsson, Eva Ribom, Kenneth Jonsson, Helena Brändström, Anna-Lena Johansson, Jessica Pettersson, Britt-Marie Andersson and Hannes Olausson. You have all made the days in the lab enjoyable. Special thanks to Majd for helping me out with formatting issues and to Elin for introducing me to Andreas and the lab.
Sune Larsson, Claes Ohlsson, Per Jensen, Hans Mallmin, Olle Kämpe, Joakim Lundeberg, Peter Savolainen, Finn Hallböök and Johan Lindberg for excellent scientific collaborations.

Colleagues from Research Departments 2 and 3 at UAS. Especially I would like to thank Annica Jacobsson for a never ending positive attitude and for our interesting discussions about all and nothing.

My thanks go out to past and previous members of the “Chicken group” at BMC. Even though I have merely been an “adoptive” member, showing up every other Tuesday or so, it has been a pleasure to work together with you. Dominic “the Dominator” Wright for your contributions to the QTL-analyses and for fun activities shared outside of the academic sphere. I thank my microarray comrades Carolyn Fitzsimmons and Sojeong Ka for valuable discussions regarding interpretation and analysis of data.

The OS-bunch: Anna-Stina Sahlqvist, Susanne Kerje, Dominic Wright, Brita Ardeşjö, Olov Ekvall and everyone not mentioned, whose company made it surprisingly easy to endure day after day of repetitive labor the weeks spent in Skara.

Everyone I had the privilege to work among at the Max Planck Institute for Evolutionary Anthropology in Leipzig. Especially Frank Albert who helped me get my bearings in the lab and who every now then dragged me out of there to show me that Leipzig was not all about science. Thanks to Svante Pääbo for making my trip possible and for letting me see “top notch” science up close.

All my friends who have helped me forget about thesis work whenever I have desperately needed to.

My extended family in Värmland and Västerås for the fun times spent together during weekends off from the lab.

My family in Skåne for being a cool and lovely bunch, for making every visit back home highly enjoyable and for always believing in me, despite that I moved from the ancestral soil ten years ago.

And finally I would like to thank Jenny, my beloved fiancée and skatt. It is amazing that you put up with me this winter and I am looking forward to us spending some quality time together again.
REFERENCES


26. National Center for Biotechnology Information, N. L. o. M.


43. Deng, H. W. Population admixture may appear to mask, change or reverse genetic effects of genes underlying complex traits. Genetics 159, 1319-23 (2001).


138. Kerje, S. et al. The twofold difference in adult size between the red junglefowl and White Leghorn chickens is largely explained by a limited number of QTLs. Anim Genet 34, 264-74 (2003).
144. Schutz, K. E. et al. Major growth QTLs in fowl are related to fearful behavior: possible genetic links between fear responses and production traits in a red junglefowl x white leghorn intercross. Behav Genet 34, 121-30 (2004).


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 314

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine”.)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-8498