Published in final edited form as:

Transl Res. 2017 March ; 181: 15–26. doi:10.1016/j.trsl.2016.10.009.

Using GWAS to identify novel therapeutic targets for osteoporosis

Olivia L. Sabik¹ and Charles R. Farber²

¹Center for Public Health Genomics and Department of Biochemistry and Molecular Genetics, School of Medicine, University of Virginia, Charlottesville, VA 22908

²Center for Public Health Genomics and Departments of Public Health Science and Biochemistry and Molecular Genetics, School of Medicine, University of Virginia, Charlottesville, VA 22908

Abstract

Osteoporosis is a common, increasingly prevalent, global health burden characterized by low bone mineral density (BMD) and increased risk of fracture. Despite its significant impact on human health, there is currently a lack of highly effective treatments free of side effects for osteoporosis. Therefore, a major goal in the field is to identify new drug targets. Genetic discovery has been shown to be effective in the unbiased identification of novel drug targets and genome-wide association studies (GWASs) have begun to provide insight into genetic basis of osteoporosis. Over the last decade, GWASs have led to the identification of ~100 loci associated with BMD and other bone traits related to risk of fracture. However, there have been limited efforts to identify the causal genes underlying the GWAS loci or the mechanisms by which GWAS loci alter bone physiology. In this review, we summarize the current state of the field and discuss strategies for causal gene discovery and the evidence that the novel genes underlying GWAS loci are likely to be a new source of drug targets.

Introduction

Osteoporosis is a disease of weakened bone, clinically characterized by low bone mineral density (BMD) and indicative of an increased risk for fracture¹. Fractures due to osteoporosis are a major public health burden^{2,3}. As a larger fraction of the population reaches old age, the annual rate of fracture and associated costs in the United States are projected to rise as much as 48% by 2025, resulting in approximately 3 million fractures and \$25.3 billion in health care costs⁴. This prognosis has led to an increased effort in developing more effective means of treating and preventing bone disease.

The majority of existing therapeutics for osteoporosis are antiresorptives, such as bisphosphonates⁵, which inhibit osteoclast-mediated bone resorption⁶. While these drugs are effective in halting bone loss and decreasing fracture risk, they are prescribed upon diagnosis, after significant bone loss has already occurred⁷. As a result, anabolic agents that build new bone are needed. Teriparatide, an injected peptide that targets the parathyroid

hormone receptor, has been shown to induce bone formation⁸. However, the need for a daily injection makes this a difficult course of treatment⁹. Additionally, side effects associated with existing therapeutics, such as atypical femoral fracture and osteonecrosis of the jaw^{10,11}, though rare, have lead to a marked decrease in preventative use¹². Given these many disincentives for using the current treatments, the identification of novel anabolic therapeutic targets that can be affected via orally active drugs is a major goal in the field.

Most of the current anti-osteoporotic therapies (Table 1) were identified using traditional molecular approaches and mouse knockout screens^{9,13,14}. Specific genes and pathways known to play a role in bone maintenance were tested for effects on bone cell function *in vitro* and bone mass *in vivo*¹⁵. While this method has been effective, in today's world of vast genomic tools there may be more efficient ways to identify novel drug targets. In fact, a recent retrospective analysis revealed that drugs targeting a wide range of diseases that had been implicated through genetic studies were almost twice as likely to succeed in the drug development pipeline than those not identified using genetic approaches¹⁶. The increased success rate of targets supported by genetics may be due to the fact that genetic studies provide a way to identify genes that, when modified, lead to an observable clinical effect not compensated for by other genes. Thus, genomic approaches provide an avenue for the unbiased discovery of novel drivers and regulators of specific biological processes and diseases. As we describe below, genome-wide association studies (GWASs) for osteoporosis-related traits have given us a wealth of potential new drug targets.

What are GWASs?

Before the development of GWAS, which we discuss below, disease-influencing loci were mapped using family-based linkage studies¹⁷. Linkage studies track the co-segregation of genetic markers with genetic variants that influence disease in families. The advantage of this approach is that large chromosomal regions are co-inherited in families, allowing disease alleles to be “tagged” using just a few hundred genetic markers spaced across the genome. This was critical because at the time, only a small number of genetic markers were known, and only a limited number of variants could be easily genotyped, so linkage was the state of the art. The disadvantage of linkage is that, due to the small number of recombinations breaking up chromosomes in families, the loci identified were large, containing hundreds of genes. Two critical advances allowed for the development of more high-resolution genetic approaches; 1) the completion of the Human Genome Project in 2003¹⁸ and the International HapMap Project¹⁹ in 2005, which provided a large list of reference single nucleotide polymorphisms (SNPs) that could be used as genetic markers²⁰ and 2) the development of massively parallel genotyping assays which allowed for the rapid and cost-effective collection of hundreds of thousands of SNPs in large numbers of samples²¹. These two advances made it possible to use GWASs to “scan” the genome for associations between millions of SNPs across the genome and phenotypes in large populations. In unrelated individuals, historical recombinations have “chopped” the genome into small blocks of co-inherited variants, and as a result, GWAS associations typically identify small genomic regions, implicating just a few genes. Thus, as genotyping large cohorts became technologically and economically feasible, GWASs have become the most common and effective means of investigating the genetic basis of common diseases.

GWASs are performed by genotyping hundreds of thousands of SNPs, often in cohorts of tens of thousands of individuals (Figure 1) ²². In addition to genotyped SNPs, non-genotyped SNPs are often imputed into a cohort ²³. Imputation is the process of predicting the genotype of untyped SNPs and is possible because we have detailed maps of how variants are co-inherited in specific populations ^{19,24}. Thus, it is common for a GWAS to test ~5–10 million common SNPs for disease associations ²⁵. GWASs are performed using either a case-control or quantitative trait design ²⁶. In the former, allele frequencies of SNPs are compared between genotyped cases, which are diagnosed with a disease, and controls. For disease-related quantitative traits, SNP genotypes are tested using regression-based approaches, which determine if SNP allele dosage is associated with a change in phenotype. There have been approximately 2,500 GWASs conducted to date, identifying over 16,000 loci associated with diseases and quantitative traits ²⁷.

The end result of a GWAS is a set of loci that harbor genetic variants that influence a disease. This is an important distinction as it is often assumed that GWAS identifies specific genes. In actuality, GWAS is just the first step in uncovering genes and variants contributing to a disease. GWAS associations pinpoint sets of SNPs in linkage disequilibrium (SNPs that are frequently co-inherited in a population). The resulting regions of association vary in size and typically implicate multiple genes, only a subset of which drives the observed phenotypic differences. Thus, the challenge of GWAS lies in identifying the causal genes and understanding the mechanism by which these genes affect the phenotype. Some of these associated SNPs are located in the exons, or coding regions, of genes and can impact protein function. It is often easy to identify the mechanism by which coding SNPs affect the phenotype, however the majority of loci identified through GWAS implicate non-coding SNPs ²⁸, suggesting they alter gene regulation ²⁹. Regulatory variation is particularly challenging to mechanistically characterize for several reasons. For instance, SNPs can affect regulation by altering transcription factor (TF) binding sites ^{30,31}, the 3-dimensional structure of the genome ³², or alternative splicing ³³. Additionally, the field has yet to develop robust approaches to identify regulatory sequences and methods to study the differences in function between two alternative regulatory sequences ^{34,35,36}. Thus, it remains challenging to dissect how GWAS associations impact disease and the functional entities that mediate the effects.

GWASs for osteoporosis-related traits

Since 2007, several GWASs have been performed for bone phenotypes. Most notably, bone mineral density (BMD) has been the trait of choice for GWAS, primarily due to its high heritability ($h^2 > 0.50$) ^{37,38}, association with fracture ³⁹ and availability in large cohorts. There have been over 30 GWASs conducted for BMD leading to the identification of nearly 100 independent loci ^{13,40}. The largest GWAS for BMD was conducted by GENetic Factors for Osteoporosis Consortium (GEFOS) ⁴¹. The GEFOSII study was a meta-analysis of lumbar spine (LS) and femoral neck (FN) BMD in 17 separate cohorts. GEFOSII used a two-stage design that included both a discovery and replication cohort. The discovery cohort consisted of ~32,000 individuals and the effects of the most significant SNPs were then replicated in ~50,000 individuals. Several important observations were made in the GEFOSII study. First, the effects on BMD of individual loci are small. In aggregate, the 64

loci explained less than 5% of the phenotypic variance in LSBMD and FNBMD. Because GEFOSII identified common variants with the largest effects and given the heritability of BMD is greater than 50%, this suggests that hundreds or possibly thousands of variants contribute to BMD. The complex genetic architecture of BMD is supported in mouse studies demonstrating that roughly 10% of random gene knockouts have BMD or other skeletal phenotypes⁴². It is also supported by GWAS for human height in >250,000 people which identified nearly 700 independent loci⁴³. Second, approximately half of the 64 BMD loci harbored genes known to be involved in BMD, while the rest harbored only genes not previously implicated in the regulation of BMD. As indicated above, we do not yet know which genes are truly causal, but for the loci harboring known genes we expect many of these to play a role. It is also possible that a subset of loci contain more than one causal gene. Generally, genes in GWAS loci that are known to play a role in BMD include: (1) members of the beta-catenin/Wnt signaling pathway which regulates osteoblastogenesis, osteoblast proliferation, and apoptosis of osteoblasts and osteoclasts^{44,45}, (2) the receptor activator of nuclear factor- κ B (RANK), RANK ligand (RANKL), and osteoprotegerin (OPG) pathway, which regulates the relationship between osteoblast and osteoclast activity in bone remodeling^{46,47}, and (3) developmental genes involved in the process of endochondral ossification, namely transcription factors which induce expression of key genes in the ossification process^{13,48}.

While these data promise to open new doors of investigation in the bone field and uncover novel therapeutic targets, it is important to point out that BMD is not an “ideal” osteoporosis phenotype. For example, some patients who suffer from low BMD do not experience osteoporotic fracture, and others who have normal BMD do experience fractures. Additionally, it has been shown that ~50% of the variance in bone strength, the main determinant of an individual’s risk of fracture, is due to BMD, whereas the other half of the variance is due to parameters such as bone size, geometry and tissue-level properties⁴⁹. Thus, there are a number of groups that are expanding to alternative phenotypes, such as trabecular and cortical microarchitecture defined by CT, to capture genetic influences on bone strength that are independent of BMD^{50,51}. It is also likely that case-control GWAS for osteoporotic fracture will help to fill the gap, though the small number of studies performed to date have identified few loci, likely because fracture is a noisy phenotype^{52,53}. Fracture is a noisy phenotype for GWAS because frequently, fractures from all sites are lumped together in a single analysis, introducing more noise. As mentioned above, sample sizes of tens of thousands of subjects are needed to identify the small effects of common genetic variants. One of the reasons that GWAS has been successful for BMD is the ability to assemble very large cohorts. Therefore, it is unclear how successful GWAS for traits other than BMD will be due to the difficulty in measuring most bone traits in the large number of subjects (N>10,000) needed for sufficient statistical power to detect genetic effects.

The magnitude of effects of genetic variants fall on a continuous spectrum ranging from single “Mendelian” mutations of large effect size that cause diseases like osteogenesis imperfecta⁵⁴, to the small effects identified by GWAS. Though still a contentious matter of debate, it is likely that most diseases and quantitative traits are influenced by variants along the entire spectrum from small to large. For example, rare, large-effect variants have been identified in WNT1 in individuals with very low BMD⁵⁵. Additionally, genome sequencing

and association studies in large cohorts have identified variants near the Engrailed Homeobox 1 (EN1), Leucine-Rich Repeat Containing G Protein-Coupled Receptor 4 (LGR4) and Collagen Type I Alpha 2 (COL1A2) genes that are rare and have relatively large effects on BMD⁵⁶⁻⁵⁸. However, the evidence is mounting that most of the genetic component of BMD is due to large number of common variants of small effects⁴¹.

GWAS is a potentially powerful approach to identify anti-osteoporotic therapeutics

The utility of genetics studies, and in particular GWASs, are often called into question, primarily because causal genes are not immediately revealed and the small effect sizes of common variants^{59,60}. In light of such criticism it is useful to define why GWAS is important. There are three general ways in which information from GWAS can provide important biological and clinical insight. First, genetic information can be used to “personalize” medicine. In theory once we define the genetic architecture of a disease, this information could be used to identify at-risk individuals. Information on variants that impact an individual’s response to treatment would also be invaluable as a clinical decision making tool. For instance, genetic diagnostics that identified individuals more likely to develop osteonecrosis of the jaw (ONJ) or atypical femoral fractures upon taking bisphosphonates would be of enormous clinical utility. Second, GWASs inform biology and identifying novel genes is important to develop a comprehensive understanding of osteoporosis and other diseases. The utility of GWAS for this purpose will only grow as we develop robust methods for moving from loci to genes to disease mechanisms. It is also important to highlight the fact GWAS differs from more traditional molecular gene discovery approaches in that it is unbiased. The importance of the unbiased nature of GWAS is underscored by the fact that over half of the GEFOSII BMD loci do not implicate any gene known to play a role in osteoporosis. Third, possibly the most important use of GWAS, is in the hunt for new anti-osteoporotic therapeutics. Current drug discovery paradigms have been hindered by the huge attrition rate in the development pipeline. Most of these targets have been identified by non-genetic studies. As described above, evidence suggests that drug targets implicated by GWAS are twice as likely to succeed in clinical trials¹⁶. Importantly, the success rate may be even higher for osteoporosis given that five out of the eight (63%) anti-osteoporosis therapeutics currently approved or in advanced clinical trials are supported by genetic data (Table 1)¹³. Thus, genetic and genomic approaches to identifying drug targets are not only feasible, but are likely more successful than other avenues of anti-osteoporotic target identification.

Causal gene discovery

Given the importance of translating GWAS data into biological knowledge, how does one go from locus to gene to mechanism? Below we outline state-of-the-art approaches and discuss how they can be used to inform GWASs.

“Direct” approaches – fine-mapping, annotating SNPs, eQTLs

There is not a standard “pipeline” one can use to go from GWAS loci to causal variants/genes (Figure 2). Instead, many different approaches may be taken depending on the disease, characteristics of the loci under investigation, and the resources available. In general though, it is desirable to both decrease the number of potentially causal variants and link those variants to either changes in protein function or, as is more often the case, an alteration in gene regulation.

Reducing the number of potential causal variants is referred to as fine-mapping. An example of fine-mapping is performing targeted complete re-sequencing of a GWAS locus in subset of study participants and then genotyping the identified variants in the entire cohort. By increasing the number of SNPs investigated in a region it may be possible to identify the truly causal variant that would be more statistically associated with the disease. It is also possible to fine-map a large number of loci at once by typing dense SNP sets using custom genotyping arrays⁶¹. It is also commonplace for the follow-up population to be larger than the original GWAS population, which allows for increased power to include or exclude individual variants as likely causal. Statistical fine-mapping is another approach that can be used to refine the list of potentially causal variants. These methods are similar in concept to “direct” fine-mapping, however, instead of targeted variant discovery by sequencing, resources such as the complete genome sequences generated by the 1000 Genomes project are used to identify a nearly comprehensive list of variants that can then be “imputed” statistically into the study population across previously identified loci and tested for association²⁴. No matter the method used, fine-mapping can be utilized in many cases to reduce the search space for causal SNPs and variants within a GWAS locus.

Once a high-confidence set of variants for a locus is identified the next approach typically taken to identify the causal entity in a GWAS locus is to annotate the SNPs within the locus. First, this will lead to the identification of non-synonymous (NS) SNPs that lie within genes. These altered gene sequences, and their protein products, are often strong causal candidates, especially if they are predicted to alter protein function using computational methods^{62,63}. Second, if coding variants are not implicated, which is usually the case, SNPs can be identified that overlap epigenetic marks often associated with poised or active regulatory elements, such as DNase I hypersensitivity sites and histone modifications. The ENCODE⁶⁴ and Epigenomics Roadmap⁶⁵ projects have generated epigenomics data on a wide-range of tissues and cell-types, including cell-types relevant to skeletal biology such as primary human osteoblasts, chondrocytes and mesenchymal stem cells. Coupled with these data are a number of computational approaches that can be used to inform GWAS, including the integration of information about pleiotropy⁶⁶, regulation⁶⁷, and epigenetics⁶⁸. For example, Bayesian approaches have been developed that rank SNPs based on functional annotation data^{69,70,71}.

Once the most likely causal variants have been identified, the next step is to link variants to their target gene(s). In the case of NS SNPs, this is immediately evident. For non-coding variants identifying their target is more difficult. This can be especially challenging in light of the observation that distal regulatory elements, such as enhancers, can be located up to 1 Mbp away from the gene promoter they act upon and it is often the case that a single

enhancer works to fine-tune the expression of more than one gene⁷². The most direct route of linking non-coding variants to their target gene(s) is to use population-scale expression data to identify expression quantitative trait loci (eQTL). eQTL are variants that regulate transcription or post-transcriptional processing (stability, splicing, etc.)^{73,74}. There are two types of eQTL, distal and local. Distal eQTL are variants that influence the expression of genes in trans, typically on different chromosomes⁷⁵. In contrast, local, or cis, eQTL influence transcript levels of genes in close proximity. In the context of GWAS, we are interested in identifying local eQTL for genes that may be causal for a particular locus. eQTL discovery consists of collecting and profiling disease-relevant tissues or cell-types in a population of densely genotyped individuals, using either gene expression microarrays⁷³ or RNA-seq⁷⁶. Ideally, these individuals would be a subset of the GWAS study population. Variants within a locus can then be tested for association with all the genes in proximity of the original locus. There are now resources, such as the data generated by the Gene Tissue Expression (GTEx) project⁷⁶, that have population-scale expression data and eQTL results for a large number of tissues that can be used to inform GWAS in the absence of disease-relevant samples from the disease GWAS. All of these methods of directly interrogating SNPs can aid in the prioritization of candidate genes and SNPs in the region, both through alteration of protein-coding sequence, and via regulatory mechanisms.

While these direct approaches to find the causal variants narrow down the list of candidates, they do not provide biological context for the potential effectors of the phenotype. Additionally, these approaches are generally based on the statistical significance of the association between the genotype of a particular SNP and the phenotype, which can be influenced by experimental design, and does not always reflect biology. In order to gain a mechanistic understanding of the drivers of these associations, genes and SNPs need to be prioritized based on biological information, rather than statistical ranking.

Network-based approaches

An additional framework for following up on GWAS associations is to use network-based approaches to biologically contextualize loci^{77,78}. Network-based strategies have been implemented in order to predict causal genes at GWAS loci, and to implicate network modules in disease (as examples^{79–84}). Especially in the case of regulatory variation, disease-associated SNPs may act via subtle changes that are propagated through entire cellular networks. Therefore, by approaching GWAS from a global, systems perspective, we can better connect associations with their physiological impacts.

One of the most widely used types of networks for informing GWAS are co-expression networks. Co-expression networks are modular, and each distinct module represents a group of highly co-expressed genes⁸⁵. These modules tend to contain genes involved in similar biological processes, e.g. the function of bone-forming osteoblasts or bone-resorbing osteoclasts⁸⁶. In practice, co-expression network analysis takes all the genes in the genome and, in a relatively unbiased manner, organizes them into functionally coherent groups. These properties are helpful for causal gene discovery because we know that complex traits are typically influenced by functionally similar genes. Therefore, by performing an unbiased, biologically driven grouping of genes and identifying modules that are enriched

for those identified by GWAS it is possible to prioritize which genes may be causal. Additionally, due to the functional similarity of genes within modules, the mechanism by which a novel gene affects disease can be inferred by the function of the other genes it is connected to within a module. Furthermore, networks have the property of being scale-free, meaning that they contain a small number of highly interconnected genes and an increasingly large number of less connected genes⁸⁷. Studies have demonstrated that in some modules highly interconnected “hub” genes are more likely to be key genes affecting a disease-related trait^{81,87,88}. As a result, analyzing GWAS data in the context of biological networks has the potential to identify causal genes and pathways relevant to disease.

One group, studying coronary artery disease (CAD), utilized both databases of known metabolic and signaling pathways and novel, tissue-specific gene co-expression networks generated from their own RNA-seq data⁸⁴. First, both “knowledge-driven pathways” and “data-driven modules” were used to generate a comprehensive list of gene sets potentially involved in CAD. Gene sets included “knowledge-based” biological pathways from the Reactome⁸⁹, Biocarta⁹⁰ and Kyoto Encyclopedia of Genes and Genomes (KEGG)⁹¹ and “data-driven” co-expression modules from ten different human and mouse co-expression studies of various CAD-related tissues, including adipose, blood, liver, muscle, heart, and kidney. Next, using gene expression data, eQTL were identified and correlated with results from the CARDIoGRAM GWAS⁹². Only SNPs that were both associated with CAD by the CARDIoGRAM GWAS and identified as influencing gene expression by eQTL analysis were included in downstream analyses. Using prioritized SNPs in conjunction with the comprehensive list of knowledge-based pathways and data-driven modules, specific gene sets that were enriched for GWAS/eQTL SNPs were identified. In order to identify the most influential genes, the resulting gene sets were then overlaid on a causal network derived from Bayesian network models of gene-gene interactions^{93,94}. Genes central to the gene-gene interaction network that were highly connected with CAD-associated genes and were identified in more than one gene set were termed “key drivers” of CAD. Finally, key drivers were perturbed using siRNA treatment, and their regulatory role was characterized. This work led to the association of novel genes with CAD, for example glyoxalase 1 (*GLO1*), which aids in defending against improperly glycosylated forms of proteins. This analysis was made possible by the large amount of available disease relevant and tissue-specific data, but could be applied to many other phenotypes with a wealth of high dimension data. As more and more data are generated in bone, this could become a feasible approach to study osteoporosis and other bone diseases.

Another type of network frequently used to identify genes central to regulatory processes are Gene Regulatory Networks (GRNs)⁹⁵. GRNs are models of molecular regulators that govern the transcriptional landscape of a cell. While many different processes impact the conversion from gene to gene product, the impact of transcription factor (TF) binding on gene expression is one of the most important. It has been observed that the transcription factors that drive disease initiation can be inferred by predicting the TF binding sites that are upstream of disease-associated genes⁹⁶. Thus, the identified TFs, which correspond to binding sites that are frequently perturbed by SNPs associated with disease using GWAS, could be early regulators of disease. These early regulators could be effectively used to diagnose diseases prior to the appearance of symptoms.

One group made use of this strategy in differentiating T-cells in order to study immune related diseases. The authors profiled the transcriptome and methylome of naive CD4+ T-cells differentiating into T_H1, T_H2, T_H17, and T_{regs}, four subsets of T-helper cells. The authors identified TFs that were differentially expressed between 6 and 24 hours for each type of T-helper cell. From this set of TFs, only those with binding sites that were enriched for GWAS-implicated SNPs were used to generate the GRN. The GRN was constructed using expression profiles from T_H1 and T_H2 cells and computationally predicted TF binding sites. Of the TFs identified as “hubs” within the GRN, those associated with disease via GWAS had the greatest number of targets, and three of the TFs, GATA binding protein 3 (GATA3), V-Maf Avian Musculoaponeurotic Fibrosarcoma Oncogene Homolog (MAF), and V-Myb Avian Myeloblastosis Viral Oncogene Homolog (MYB), were associated with 91% of the SNPs associated with any TF hub. Experimental investigation of the three TFs found differential expression of splice variants between symptomatic and asymptomatic patients, suggesting they play a role in disease progression⁸². This study highlights the power of integrating GRNs and GWAS data to identify genes that play a key role in disease.

Need for bone resources and –omics data

One of the obstacles hindering causal gene discovery for bone GWAS is the paucity of population-scale transcriptomic and epigenomic data from bone tissue or primary bone cells. There are a number of large repositories of expression and epigenetics data from particular tissue types; however, there is no such resource for the bone field yet. For example, the Gene Tissue Expression (GTEx) project is an NIH funded effort to generate RNA-seq expression profiles (and soon epigenetics and proteomics data) from multiple tissues (>40) in a large genotyped human cohort⁷⁶. The resulting resource is extraordinarily powerful and provides the opportunity to understand how genetic variation influences expression on a genome-wide basis. One of the primary efforts of GTEx to date is to provide the genetics and genomics community with eQTL results, allowing investigators to use these data to inform GWAS. Unfortunately, GTEx is not collecting bone tissues and primary bone cells. Our group and others are using the GTEx resource, along with data from two small bone eQTL studies^{97,98} to inform BMD GWAS; however, this only works for signals shared between bone and non-bone tissues or genetic effects on bone that arise from expression changes in non-bone tissues. Related resources, like the ENCODE project⁶⁴, which focuses on cataloging functional elements found in the genome, does contain histone modification and DNase I hypersensitivity site data on primary osteoblasts and related cells such as mesenchymal stem cells and chondrocytes, though not from other bone cells. Thus, there is a significant need in the bone field for generating -omics data on bone and primary bone cells that can be used for causal gene discovery and as an independent discovery platform.

Future Directions

As we have discussed above, one of the major bottlenecks in our understanding of how genetic variation leads to differences in traits, such as BMD, is identifying causal genes for existing GWAS loci. However, that does not imply that the utility of GWAS has been exhausted. To the contrary, there is much more, even for BMD, to discover, especially in light of the observation that the 64 loci identified in the GEFOSII meta-analysis explain only

~5% of the variance in BMD⁴¹. Heritability estimates for BMD are generally >50%, so to date GWASs have explained roughly 10% of its genetic component⁴¹. Additional GWAS studies are ongoing and there will no doubt be larger meta-analyses for BMD that will yield even more loci.

There is also a need for the investigation of phenotypes beyond BMD. In particular, GWASs for traits that account for aspects of bone strength independent of BMD would be ideal, for example bone size and microarchitecture. Though many of these traits are more difficult to measure than BMD, their interrogation would provide significant insight into the genetics of bone strength and fracture risk. Such studies are already starting to be performed in large-scale cohorts. For example, trabecular microarchitecture as measured by quantitative computed tomography (QCT) was recently investigated by GWAS in a cohort of ~15,000 individuals⁹⁹. We anticipate the trend of GWAS for a more diverse set of bone-strength related traits will increase in the coming years. In addition to more diverse phenotypes, a powerful application of GWAS would be to separate the genetic analysis of bone accrual and bone loss, especially in light of the observation that the genetic correlation between BMD in pre- and postmenopausal women is modest ($r=0.30$)¹⁰⁰. Bone is accrued until peak bone mass between the ages 20 and 25. GWAS in pediatric populations could therefore be used to identify loci specifically affecting the attainment of peak bone mass¹⁰¹. Small GWASs for BMD in pediatric populations have already started to provide insight¹⁰². We are also interested in developing a much better understanding of genetic loci affecting bone loss due to aging in both sexes, or more importantly, bone loss in females after menopause. The dramatic loss of bone after menopause in women is the primary reason why 80% of the 12 million Americans with osteoporosis are female⁷. To date, GWAS has not been used to study bone loss in postmenopausal women, although it is the single strongest determinant of poor bone health in women. Lastly, with the exception of GWAS studies for BMD in Asian populations (as examples^{103,104}), most GWAS for bone traits have been performed in individuals of European ancestry⁴¹. In order for GWAS results to inform drug discovery that is applicable to all populations of people it is imperative that GWASs for bone traits be performed in populations with diverse ethnic backgrounds.

As discussed above, in the context of causal gene discovery it is imperative that transcriptomic, and other -omics data, on bone and bone cells be collected in large genotyped populations. Without these resources it will be difficult to definitively identify causal variants and genes. However, in combination with these resources there is a need for better methods for defining causal variants and genes. Ideally such methods will include computational strategies for defining cis-acting regulatory sequences and accurately predicting the effects of genetic variation on these sequences¹⁰⁵. Furthermore, experimental methods are needed to query large number of variants on regulatory sequence function¹⁰⁶. Currently, most experimental strategies rely on reporter assays that assess regulatory function of variants using very artificial systems that investigate sequences outside of their native chromosomal context³⁴. There are, however, new methods, such as using lentiviral based reporter assays that integrate into the genome that are very promising¹⁰⁷. Additionally, CRISPR/Cas9 based genome-editing approaches that allow one to modify individual variants in human cells are becoming more efficient and will likely to play a large role in causal variant and gene discovery moving forward^{108,109}. Such experiments will

need to be performed in human cells, thus cell lines that mimic their *in vivo* counterparts are urgently needed. An attractive alternative to cell lines is bone cells derived from induced pluripotent stem cells (iPSCs). These would be the ideal resources for the necessary human cell studies and groups have already demonstrated that iPSCs can be used to derive osteoblasts¹¹⁰ and osteoclasts¹¹¹.

In summary, over the past decade GWAS has provided an unprecedented understanding of how genetic variation influences osteoporosis and fracture. We now know that most of the variation in BMD at the population level is due to hundreds, if not thousands, of variants with subtle effects on BMD and most of these variants exert their impact on bone by altering gene regulation. While the effects of variants are small, the discovery and characterization of causal genes has the potential to provide a much more detailed understanding of the etiology of fracture. While the initial GWAS results are promising, there is much to be done both in terms of comprehensively defining the genetic architecture of osteoporosis and fracture, and converting genetic data into biological knowledge. Along the way, we hope these data can be used to develop more effective and efficient therapies to treat and ultimately prevent bone loss and fracture.

Acknowledgments

The authors have read the journal's policy on disclosure of conflicts of interest and have none to declare. The authors have read the journal's authorship agreement and have reviewed and approved the manuscript. C.R.F is funded by grants R01-AR057759, R01-AR64790 and R01-AR068345 from the National Institute of Arthritis and Musculoskeletal and Skin Diseases. O.L.S. is funded by an institutional training grant T32-GM008136 at the University of Virginia from the National Institute of General Medical Sciences.

References

1. Kanis JA. Diagnosis of osteoporosis. *Osteoporos Int.* 1997; 7(3):108–116.
2. Cauley JA. Public Health Impact of Osteoporosis. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences.* 2013; 68(10):1243–1251.
3. Office of the Surgeon General (US). Bone Health and Osteoporosis: A Report of the Surgeon General. Rockville (MD): Office of the Surgeon General (US); 2004. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK45513/>
4. Burge R, Dawson-Hughes B, Solomon DH, et al. Incidence and Economic Burden of Osteoporosis-Related Fractures in the United States, 2005–2025. *Journal of Bone and Mineral Research.* 2006; 22(3):465–475.
5. Maraka S, Kennel KA. Bisphosphonates for the prevention and treatment of osteoporosis. *BMJ.* 2015; 351:h3783. [PubMed: 26333528]
6. Carano A, Teitelbaum SL, Konsek JD, et al. Bisphosphonates directly inhibit the bone resorption activity of isolated avian osteoclasts in vitro. *J Clin Invest.* 1990; 85(2):456–461. [PubMed: 2105340]
7. Black DM, Rosen CJ. Clinical Practice. Postmenopausal Osteoporosis. *N Engl J Med.* 2016; 374(3): 254–262. [PubMed: 26789873]
8. Neer RM, Arnaud CD, Zanchetta JR, et al. Effect of Parathyroid Hormone (1–34) on Fractures and Bone Mineral Density in Postmenopausal Women with Osteoporosis. *N Engl J Med.* 2001; 344(19): 1434–1441. [PubMed: 11346808]
9. Brommage R. Genetic Approaches To Identifying Novel Osteoporosis Drug Targets. *J Cell Biochem.* 2015; 116(10):2139–2145. [PubMed: 25833316]
10. Durie BGM, Katz M, Crowley J. Osteonecrosis of the jaw and bisphosphonates. *N Engl J Med.* 2005; 353(1) 99-102-discussion99-102.

11. Schilcher J, Koeppen V, Aspenberg P, et al. Risk of atypical femoral fracture during and after bisphosphonate use. *N Engl J Med*. 2014; 371(10):974–976. [PubMed: 25184886]
12. Kolata G. Fearing Drugs' Rare Side Effects, Millions Take Their Chances With Osteoporosis. *New York Times*. 2016 Jun 1.
13. Richards JB, Zheng H-F, Spector TD. Genetics of osteoporosis from genome-wide association studies: advances and challenges. *Nat Rev Gen*. 2012; 13(8):576–588.
14. Rissanen JP, Halleen JM. Models and screening assays for drug discovery in osteoporosis. *Expert Opin Drug Discov*. 2010; 5(12):1163–1174. [PubMed: 22822719]
15. Lacey DL, Boyle WJ, Simonet WS, et al. Bench to bedside: elucidation of the OPG-RANK-RANKL pathway and the development of denosumab. *Nat Rev Drug Discov*. 2012; 11(5):401–419. [PubMed: 22543469]
16. Nelson MR, Tipney H, Painter JL, et al. The support of human genetic evidence for approved drug indications. *Nature Genetics*. 2015; 47(8):856–860. [PubMed: 26121088]
17. Ralston SH, Uitterlinden AG. Genetics of Osteoporosis. *Endocrine Reviews*. 2010; 31(5):629–662. [PubMed: 20431112]
18. Auton A, Altshuler DM, Durbin RM, et al. A global reference for human genetic variation. *Nature*. 2015; 526(7571):68–74. [PubMed: 26432245]
19. Belmont JW, et al. The International HapMap Consortium. The International HapMap Project. *Nature*. 2003; 426(6968):789–796. [PubMed: 14685227]
20. Risch N, Merikangas K. The Future of Genetic Studies of Complex Human Diseases. *Science*. 1996; 273(5281):1516–1517. [PubMed: 8801636]
21. Shalon D, Smith SJ, Brown PO. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res*. 1996; 6(7):639–645. [PubMed: 8796352]
22. Altshuler D, Daly MJ, Lander ES. Genetic Mapping in Human Disease. *Science*. 2008; 322(5903):881–888. [PubMed: 18988837]
23. Porcu E, Sanna S, Fuchsberger C, et al. Genotype imputation in genome-wide association studies. *Curr Protoc Hum Genet*. 2013 Chapter 1:Unit1.25.
24. Altshuler DM, Durbin RM, Bentley DR, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; 491(7422):56–65. [PubMed: 23128226]
25. LaFramboise T. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Res*. 2009; 37(13):4181–4193. [PubMed: 19570852]
26. Pearson TA, Manolio TA. How to Interpret a Genome-wide Association Study. *JAMA*. 2008; 299(11):1335–1344. [PubMed: 18349094]
27. Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*. 2013; 42(D1):D1001–D1006. [PubMed: 24316577]
28. Maurano MT, Humbert R, Rynes E, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012; 337(6099):1190–1195. [PubMed: 22955828]
29. Freedman ML, Monteiro ANA, Gayther SA, et al. Principles for the post-GWAS functional characterization of cancer risk loci. *Nature Genetics*. 2011; 43(6):513–518. [PubMed: 21614091]
30. Hindorf LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA*. 2009; 106(23):9362–9367. [PubMed: 19474294]
31. Kilpinen H, Waszak SM, Gschwind AR, et al. Coordinated Effects of Sequence Variation on DNA Binding, Chromatin Structure, and Transcription. *Science*. 2013; 342(6159):744–747. [PubMed: 24136355]
32. Darrow EM, Huntley MH, Dudchenko O, et al. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proc Natl Acad Sci USA*. 2016 Jul.; 201609643.
33. Jakubczik F, Jones K, Nichols J, et al. A SNP in the Immunoregulatory Molecule CTLA-4 Controls mRNA Splicing In Vivo but Does Not Alter Diabetes Susceptibility in the NOD Mouse. *Diabetes*. 2016; 65(1):120–128. [PubMed: 26450994]

34. Arnold CD, Gerlach D, Stelzer C, et al. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science*. 2013; 339(6123):1074–1077. [PubMed: 23328393]
35. Peterson TA, Mort M, Cooper DN, et al. Regulatory Single Nucleotide Variant Predictor (RSVP) Increases Predictive Performance of Functional Regulatory Variants. *Human Mutation*. 2016 Jul.
36. Chen J, Tian W. Explaining the disease phenotype of intergenic SNP through predicted long range regulation. *Nucleic Acids Res*. 2016 Jun. gkw519.
37. Pocock NA, Eisman JA, Hopper JL, et al. Genetic determinants of bone mass in adults. A twin study. *J Clin Invest*. 1987; 80(3):706–710. [PubMed: 3624485]
38. Krall EA, Dawson-Hughes B. Heritable and life-style determinants of bone mineral density. *Journal of Bone and Mineral Research*. 1993; 8(1):1–9. [PubMed: 8427042]
39. Trémollières FA, Pouillès J-M, Drewniak N, et al. Fracture risk prediction using BMD and clinical risk factors in early postmenopausal women: Sensitivity of the WHO FRAX tool. *Journal of Bone and Mineral Research*. 2010; 25(5):1002–1009. [PubMed: 20200927]
40. Rivadeneira F, Styrkarsdottir U, Estrada K, et al. Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. *Nature Genetics*. 2009; 41(11): 1199–1206. [PubMed: 19801982]
41. Estrada K, Styrkarsdottir U, Evangelou E, et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nature Genetics*. 2012; 44(5):491–501. [PubMed: 22504420]
42. Brommage R, Liu J, Hansen GM, et al. High-throughput screening of mouse gene knockouts identifies established and novel skeletal phenotypes. *Bone Res*. 2014; 2:14034. [PubMed: 26273529]
43. Wood AR, Esko T, Yang J, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics*. 2014; 46(11):1173–1186. [PubMed: 25282103]
44. Krishnan V, Bryant HU, Macdougald OA. Regulation of bone mass by Wnt signaling. *J Clin Invest*. 2006; 116(5):1202–1209. [PubMed: 16670761]
45. Baron R, Kneissel M. WNT signaling in bone homeostasis and disease: from human mutations to treatments. *Nat Med*. 2013; 19(2):179–192. [PubMed: 23389618]
46. Boyce BF, Xing L. Functions of RANKL/RANK/OPG in bone modeling and remodeling. *Archives of Biochemistry and Biophysics*. 2008; 473(2):139–146. [PubMed: 18395508]
47. Wittrant Y, Théoleyre S, Chipoy C, et al. RANKL/RANK/OPG: new therapeutic targets in bone tumours and associated osteolysis. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*. 2004; 1704(2):49–57. [PubMed: 15363860]
48. Nishimura R, Hata K, Ono K, et al. Regulation of endochondral ossification by transcription factors. *Front Biosci (Landmark Ed)*. 2012; 17:2657–2666. [PubMed: 22652803]
49. Jepsen KJ. Functional Interactions Among Morphologic and Tissue Quality Traits Define Bone Quality. *Clin Orthop Relat Res*. 2010; 469(8):2150–2159.
50. Paternoster L, Lorentzon M, Lehtimäki T, et al. Richards JB. Genetic Determinants of Trabecular and Cortical Volumetric Bone Mineral Densities and Bone Microstructure. *PLoS Genet*. 2013; 9(2):e1003247. [PubMed: 23437003]
51. Levy R, Mott RF, Iraqi FA, et al. Collaborative cross mice in a genetic association study reveal new candidate genes for bone microarchitecture. *BMC Genomics*. 2015; 16(1):465. [PubMed: 26084559]
52. Guo Y, Tan L-J, Lei S-F, et al. Georges M. Genome-Wide Association Study Identifies ALDH7A1 as a Novel Susceptibility Gene for Osteoporosis. *PLoS Genet*. 2010; 6(1):e1000806. [PubMed: 20072603]
53. Hwang J-Y, Lee SH, Go MJ, et al. Meta-analysis identifies a MECOM gene as a novel predisposing factor of osteoporotic fracture. *Journal of medical genetics*. 2013; 50(4):212–219. [PubMed: 23349225]
54. Van Dijk FS, Sillence DO. Osteogenesis imperfecta: Clinical diagnosis, nomenclature and severity assessment. *American Journal of Medical Genetics Part a*. 2014; 164(6):1470–1481.
55. Laine CM, Joeng KS, Campeau PM, et al. WNT1 Mutations in Early-Onset Osteoporosis and Osteogenesis Imperfecta. *N Engl J Med*. 2013; 368(19):1809–1816. [PubMed: 23656646]

56. Styrkarsdottir U, Thorleifsson G, Sulem P, et al. Nonsense mutation in the LGR4 gene is associated with several human diseases and other traits. *Nature*. 2013; 497(7450):517–520. [PubMed: 23644456]
57. Styrkarsdottir U, Thorleifsson G, Eiriksdottir B, et al. Two Rare Mutations in the COL1A2 Gene Associate With Low Bone Mineral Density and Fractures in Iceland. *Journal of Bone and Mineral Research*. 2015; 31(1):173–179. [PubMed: 26235824]
58. Zheng H-F, Forgetta V, Hsu Y-H, et al. Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. *Nature*. 2015; 526(7571):112–117. [PubMed: 26367794]
59. McClellan J, King M-C. Genetic heterogeneity in human disease. *Cell*. 2010; 141(2):210–217. [PubMed: 20403315]
60. Visscher PM, Brown MA, McCarthy MI, et al. Five Years of GWAS Discovery. *The American Journal of Human Genetics*. 2012; 90(1):7–24. [PubMed: 22243964]
61. Törn C, Liu X, Hagopian W, et al. Complement gene variants in relation to autoantibodies to beta cell specific antigens and type 1 diabetes in the TEDDY Study. *Scientific Reports*. 2016; 6:27887. [PubMed: 27306948]
62. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009; 4(8):1073–1081. [PubMed: 19561590]
63. Adzhubei I, Jordan DM, Sunyaev SR. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. Hoboken, NJ, USA: John Wiley & Sons, Inc; 2001. p. 7.20.1–7.20.41.
64. Sloan CA, Chan ET, Davidson JM, et al. ENCODE data at the ENCODE portal. *Nucleic Acids Res*. 2016; 44(Database issue):D726–D732. [PubMed: 26527727]
65. Chadwick LH. The NIH Roadmap Epigenomics Program data resource. *Epigenomics*. 2012; 4(3):317–324. [PubMed: 22690667]
66. Chung D, Yang C, Li C, et al. Tang H. GPA: A Statistical Approach to Prioritizing GWAS Results by Integrating Pleiotropy and Annotation. *PLoS Genet*. 2014; 10(11):e1004787. [PubMed: 25393678]
67. Schaub MA, Boyle AP, Kundaje A, et al. Linking disease associations with regulatory information in the human genome. *Genome Res*. 2012; 22(9):1748–1759. [PubMed: 22955986]
68. Hardison RC. Genome-wide epigenetic data facilitate understanding of disease susceptibility association studies. *J Biol Chem*. 2012; 287(37):30932–30940. [PubMed: 22952232]
69. Spain SL, Barrett JC. Strategies for fine-mapping complex traits. *Hum Mol Genet*. 2015; 24(R1):R111–R119. [PubMed: 26157023]
70. Stephens M, Balding DJ. Bayesian statistical methods for genetic association studies. *Nature Publishing Group*. 2009; 10(10):681–690.
71. Maller JB, McVean G, et al. Wellcome Trust Case Control Consortium. Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nature Genetics*. 2012; 44(12):1294–1301. [PubMed: 23104008]
72. Carninci P, Sandelin A, Lenhard B, et al. Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genetics*. 2006; 38(6):626–635. [PubMed: 16645617]
73. Farber, CR., Lusis, AJ. Genetic Dissection of Complex Traits. Vol. 60. Elsevier; 2008. Integrating Global Gene Expression Analysis and Genetics; p. 571–601. *Advances in Genetics*
74. Rockman MV, Kruglyak L. Genetics of global gene expression. *Nat Rev Gen*. 2006; 7(11):862–872.
75. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. *Nat Rev Gen*. 2015; 16(4):197–212.
76. Melé M, Ferreira PG, Reverter F, et al. The human transcriptome across tissues and individuals. *Science*. 2015; 348(6235):660–665. [PubMed: 25954002]
77. Jia P, Zhao Z. Network-assisted analysis to prioritize GWAS results: principles, methods and perspectives. *Hum Genet*. 2013; 133(2):125–138.
78. Leiserson MDM, Eldridge JV, Ramachandran S, et al. Network analysis of GWAS data. *Current Opinion in Genetics & Development*. 2013; 23(6):602–610. [PubMed: 24287332]

79. Califano A, Butte AJ, Friend S, et al. Leveraging models of cell regulation and GWAS data in integrative network-based association studies. *Nature Genetics*. 2012; 44(8):841–847. [PubMed: 22836096]
80. Farber CR. Systems-Level Analysis of Genome-Wide Association Data. *G3: Genes|Genomes|Genetics*. 2013; 3(1):119–129. [PubMed: 23316444]
81. Farber CR. Identification of a gene module associated with BMD through the integration of network analysis and genome-wide association data. *Journal of Bone and Mineral Research*. 2010; 25(11):2359–2367. [PubMed: 20499364]
82. Gustafsson M, Gawel DR, Alfredsson L, et al. A validated gene regulatory network and GWAS identifies early regulators of T cell-associated diseases. *Science Translational Medicine*. 2015; 7(313):313ra178–313ra178.
83. Huan T, Meng Q, Saleh MA, et al. Integrative network analysis reveals molecular mechanisms of blood pressure regulation. *Molecular Systems Biology*. 2015; 11(4):799–799.
84. Mäkinen V-P, Civelek M, Meng Q, et al. Attie A. Integrative Genomics Reveals Novel Molecular Pathways and Gene Networks for Coronary Artery Disease. *PLoS Genet*. 2014; 10(7):e1004502. [PubMed: 25033284]
85. Horvath S, Dong J, Miyano S. Geometric Interpretation of Gene Coexpression Network Analysis. *PLOS Comput Biol*. 2008; 4(8):e1000117. [PubMed: 18704157]
86. Goh K-I, Cusick ME, Valle D, et al. The human disease network. *PNAS*. 2007; 104(21):8685–8690. [PubMed: 17502601]
87. Horvath, S. *Weighted Network Analysis*. New York, NY: Springer New York; 2011.
88. Horvath S, Zhang B, Carlson M, et al. Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *PNAS*. 2006; 103(46):17402–17407. [PubMed: 17090670]
89. Fabregat A, Sidiropoulos K, Garapati P, et al. The Reactome pathway Knowledgebase. *Nucleic Acids Res*. 2016; 44(D1):D481–D487. [PubMed: 26656494]
90. Nishimura D. *BioCarta. Biotech Software & Internet Report*. 2001; 2(3):117–120.
91. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000; 28(1):27–30. [PubMed: 10592173]
92. Preuss M, König IR, Thompson JR, et al. Design of the Coronary ARtery Disease Genome-Wide Replication And Meta-Analysis (CARDIoGRAM) Study: A Genome-Wide Association Meta-analysis Involving More Than 22 000 Cases and 60 000 Controls. *Circulation: Cardiovascular Genetics*. 2010; 3(5):475–483. [PubMed: 20923989]
93. Zhu J, Wiener MC, Zhang C, et al. Increasing the Power to Detect Causal Associations by Combining Genotypic and Expression Data in Segregating Populations. *PLOS Comput Biol*. 2007; 3(4):e69. [PubMed: 17432931]
94. Zhu J, Zhang B, Smith EN, et al. Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nature Genetics*. 2008; 40(7):854–861. [PubMed: 18552845]
95. Karlebach G, Shamir R. Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol*. 2008; 9(10):770–780. [PubMed: 18797474]
96. Chen C-Y, Chang I-S, Hsiung CA, et al. On the identification of potential regulatory variants within genome wide association candidate SNP sets. *BMC Med Genomics*. 2014; 7(1):D1047.
97. Grundberg E, Kwan T, Ge B, et al. Population genomics in a disease targeted primary cell model. *Genome Res*. 2009; 19(11):1942–1952. [PubMed: 19654370]
98. Grundberg E, Adoue V, Kwan T, et al. Gibson G. Global Analysis of the Impact of Environmental Perturbation on cis-Regulation of Gene Expression. *PLoS Genet*. 2011; 7(1):e1001279. [PubMed: 21283786]
99. Nielson CM, Liu C-T, Smith AV, et al. Novel Genetic Variants are Associated With Increased Vertebral Volumetric BMD, Reduced Vertebral Fracture Risk, and Increased Expression of SCL1A3 and EPHB2. *Journal of Bone and Mineral Research*. 2016 Aug.
100. Shaffer, J., Kammerer, C., Dressen, A., et al. [Accessed August 16, 2016] Different Genes Contribute to Variation in Peak Bone Density and Bone Loss. Annual Meeting of the American Society for Bone and Mineral Research. <http://www.asbmr.org/Itinerary/PresentationDetail.aspx?id=68c5c92f-1060-4d1b-a177-88b7126301e1>. Published 2014

101. Kemp JP, Medina-Gomez C, Tobias JH, et al. The case for genome-wide association studies of bone acquisition in paediatric and adolescent populations. *BoneKEy Rep.* 2016; 5:796. [PubMed: 27257477]
102. Chesi A, Mitchell JA, Kalkwarf HJ, et al. A trans-ethnic genome-wide association study identifies gender-specific loci influencing pediatric aBMD and BMC at the distal radius. *Hum Mol Genet.* 2015; 24(17):5053–5059. [PubMed: 26041818]
103. Cho YS, Go MJ, Kim YJ, et al. A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nature Genetics.* 2009; 41(5):527–534. [PubMed: 19396169]
104. Choi HJ, Park H, Zhang L, et al. Genome-wide association study in East Asians suggests UHMK1 as a novel bone mineral density susceptibility gene. *Bone.* 2016; 91:113–121. [PubMed: 27424934]
105. Knight JC. Approaches for establishing the function of regulatory genetic variants involved in disease. *Genome Medicine.* 2014; 6(10):1748.
106. Dailey L. High throughput technologies for the functional discovery of mammalian enhancers: New approaches for understanding transcriptional regulatory network dynamics. *Genomics.* 2015; 106(3):151–158. [PubMed: 26072436]
107. Inoue F, Kircher M, Martin B, et al. A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *bioRxiv.* 2016:061606 Jun.
108. Soldner F, Stelzer Y, Shivalila CS, et al. Parkinson-associated risk variant in distal enhancer of α -synuclein modulates target gene expression. *Nature.* 2016; 533(7601):95–99. [PubMed: 27096366]
109. Raghavan, A., Wang, X., Rogov, P., et al. High-Throughput Screening and CRISPR-Cas9 Modeling of Causal Lipid-Associated Expression Quantitative Trait Locus Variants. *Cold Spring Harbor Labs Journals*; 2016. p. 056820
110. Bilousova G, Hyun JD, King KB, et al. Osteoblasts derived from Induced Pluripotent Stem Cells form Calcified Structures in Scaffolds both in vitro and in vivo. *Stem Cells.* 2011; 29(2):206–216. [PubMed: 21732479]
111. Jeon OH, Panicker LM, Lu Q, et al. Human iPSC-derived osteoblasts and osteoclasts together promote bone regeneration in 3D biomaterials. *Scientific Reports.* 2016; 6:26761. [PubMed: 27225733]
112. Cummings SR, Martin JS, McClung MR, et al. Denosumab for Prevention of Fractures in Postmenopausal Women with Osteoporosis. *N Engl J Med.* 2009; 361(8):756–765. [PubMed: 19671655]
113. Lewiecki EM. Sclerostin: a novel target for intervention in the treatment of osteoporosis. *Discov Med.* 2011; 12(65):263–273. [PubMed: 22031665]
114. Ettinger B, Black DM, Mitlak BH, et al. Reduction of Vertebral Fracture Risk in Postmenopausal Women With Osteoporosis Treated With Raloxifene: Results From a 3-Year Randomized Clinical Trial. *JAMA.* 1999; 282(7):637–645. [PubMed: 10517716]
115. Greenspan SL, Bone HG, Ettinger MP, et al. Effect of Recombinant Human Parathyroid Hormone (1–84) on Vertebral Fracture and Bone Mineral Density in Postmenopausal Women with Osteoporosis: A Randomized Trial. *Ann Intern Med.* 2007; 146(5):326–339. [PubMed: 17339618]
116. Liberman UA, Weiss SR, Bröll J, et al. Effect of Oral Alendronate on Bone Mineral Density and the Incidence of Fractures in Postmenopausal Osteoporosis. *N Engl J Med.* 1995; 333(22):1437–1444. [PubMed: 7477143]
117. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. *JAMA.* 2002; 288(3):321–333. [PubMed: 12117397]
118. Gauthier JY, Chauret N, Cromlish W, et al. The discovery of odanacatib (MK-0822), a selective inhibitor of cathepsin K. *Bioorganic & Medicinal Chemistry Letters.* 2008; 18(3):923–928. [PubMed: 18226527]
119. Canalis E. Update in new anabolic therapies for osteoporosis. *The Journal of Clinical Endocrinology & Metabolism.* 2010; 95(4):1496–1504. [PubMed: 20375217]

Genome Wide Association Studies

Figure 1.

Workflow for genome wide association studies (GWASs). A cohort of subjects, either cases, who are diagnosed with disease, and controls, who are healthy, or a group of people who vary in a quantitative trait, are genotyped at a large number of SNPs. Next, for a quantitative trait, a regression analysis is used to identify differences in phenotype as a function of genotype for millions of genetic variants. In case-control studies, the allele frequencies of variants between disease-diagnosed and disease-free subjects are compared. Of these comparisons, those that are statistically significant are called associations; these represent

genetic loci that are associated with changes in the quantitative trait or the disease phenotype. Finally, the genes within these associated regions are studied in order to identify the causal genes that impact the phenotype. This process is outlined in Fig. 2.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Causal Gene Discovery Pipeline

Figure 2.

Workflow for causal gene discovery. As described in Fig. 1, GWASs result in the identification of genomic loci associated with a trait or disease. Both the significantly associated SNPs and the genes within the region can be identified. If a SNP is identified within the coding region of a gene, computational techniques can be employed to determine the likelihood that the SNP causes a functional change in the protein it produces. If the SNP lies in a non-coding region, it could play a role in regulating gene expression. This hypothesis can be supported by relating the SNP to gene expression using expression quantitative trait locus (eQTL) analysis, or by SNP colocalization with regulatory epigenomic marks. The genes within the region can also be integrated with network information to identify potentially causal genes.

Table 1

Anti-osteoporotic drug targets that have been linked to changes in BMD by GWAS.

Drug Class	Drug Target	Target locus IDed through GWAS	Refs
Denosumab	RANKL	RANKL	112
Sclerostin inhibitors	Sclerostin (SOST)	SOST	113
Selective oestrogen receptor modulators	Oestrogen receptor	ESR1	114
Parathyroid hormone analogues	Parathyroid hormone (PTH) receptor	Not identified, but pathway highlighted by PTH-like hormone and PTH-related protein	115,8
Bisphosphonates	Farnesyl pyrophosphate	Not identified	116
Oestrogen	Oestrogen receptor	ESR1	117
Cathepsin K inhibitors	Cathepsin K	Not identified	118
Dickkopf 1 (DKK1) inhibitors	DKK1	DKK1	119

Table adapted from¹³.