

Review

## The Revolution in Human Monogenic Disease Mapping

Emma Duncan <sup>1,2</sup>, Matthew Brown <sup>1</sup> and Eileen M. Shore <sup>3,4,\*</sup>

<sup>1</sup> University of Queensland Diamantina Institute, Translational Research Institute, Princess Alexandra Hospital, 37 Kent Road, Woolloongabba, Brisbane 4102, Queensland, Australia; E-Mails: emma.duncan@uq.edu.au (E.D.); matt.brown@uq.edu.au (M.B.)

<sup>2</sup> Department of Endocrinology and Diabetes, Royal Brisbane and Women's Hospital, Herston 4029, Queensland, Australia

<sup>3</sup> Center for Research in FOP and Related Disorders, Department of Orthopaedic Surgery, Perelman School of Medicine, University of Pennsylvania, 3450 Hamilton Walk, Philadelphia, PA 19104, USA

<sup>4</sup> Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Curie Boulevard, Philadelphia, PA 19104, USA

\* Author to whom correspondence should be addressed; E-Mail: shore@mail.med.upenn.edu; Tel.: +1-215-898-2331; Fax: +1-215-573-2133.

Received: 8 August 2014; in revised form: 29 August 2014 / Accepted: 1 September 2014 /

Published: 5 September 2014

---

**Abstract:** The successful completion of the Human Genome Project (HGP) was an unprecedented scientific advance that has become an invaluable resource in the search for genes that cause monogenic and common (polygenic) diseases. Prior to the HGP, linkage analysis had successfully mapped many disease genes for monogenic disorders; however, the limitations of this approach were particularly evident for identifying causative genes in rare genetic disorders affecting lifespan and/or reproductive fitness, such as skeletal dysplasias. In this review, we illustrate the challenges of mapping disease genes in such conditions through the ultra-rare disorder fibrodysplasia ossificans progressiva (FOP) and we discuss the advances that are being made through current massively parallel (“next generation”) sequencing (MPS) technologies.

**Keywords:** Human Genome Project; fibrodysplasia ossificans progressiva; monogenic diseases; disease gene discovery; NGS

---

## 1. Introduction

Until very recently, mapping the causative gene for monogenic diseases depended on finding families with demonstrable Mendelian inheritance of the disease, preferably in multiple generations. Linkage approaches in such families were successful in identifying mutations responsible for many of the more frequent monogenic diseases and traits [1]. However, identifying the underlying mutations in rare genetic diseases, especially those associated with low reproductive fitness, late onset diseases, or diseases with early lethality, proved much more challenging.

The Human Genome Project (HGP) set the stage for success in meeting these challenges. The ultra-rare disorder fibrodysplasia ossificans progressiva (FOP), with a frequency of one in two million, is an example of such a success. Linkage in FOP families identified chromosomal regions of interest; Human Genome Project databases then facilitated the identification of candidate genes within the linkage region and permitted the efficient identification of altered DNA sequences. FOP was found to be caused by a recurrent single nucleotide substitution occurring in >95% of patients. Despite the ultimate success in mapping the FOP gene, this discovery took decades of effort to identify families with inheritance of the disease (linkage analysis was eventually accomplished with just five two-generation families) followed by the time-consuming tasks of conducting genome-wide linkage analysis, and subsequent identification and re-sequencing of candidate genes within the linkage intervals. Knowing the genetic mutation in FOP has led quickly to better understanding of the underlying pathology and directed strategies for treatments, neither of which were possible before molecular aetiology was determined.

Building on the technology, computation, and scientific information generated through the HGP, the continued advances in mapping disease genes have been extraordinarily rapid. Faced today with the challenge of identifying a rare gene mutation in a disease like FOP, high-throughput exome and/or whole genome sequencing approaches would identify the genetic mutation rapidly. Moreover, gene identification may require sequencing of only a small number of unrelated cases, small families, or, in some cases, even a single proband. These breakthroughs are leading to an upsurge in disease gene discoveries with their associated benefits [2].

After mapping a disease-causing gene, many challenges remain in understanding additional genetic contributions to disease onset and progression. In FOP, for example, although the main disease characteristics are unique and readily recognized there is significant variability in the age of disease onset, the rate of disease progression, and the severity of disease, likely to arise from as yet un-identified genetic causes. Similar issues exist with many common heritable diseases such as osteogenesis imperfecta and Marfan's syndrome. With further development of genome technologies, the ability to understand phenotypic variability and the participation of genetic modifiers is becoming a reality.

## 2. Linkage Mapping

### 2.1. History of Linkage Mapping

At the turn of the twentieth century, Archibald Garrod coined the term “inborn errors of metabolism” to explain the increased incidence of alkaptonuria (and, subsequently, also cystinuria, pentosuria, and albinism) in consanguineous families compared with the general population, and suggested that these

conditions were caused by transmissible elements within families [3]. Since this time, the challenge has been to identify these errors.

For most of the twentieth century, linkage mapping was the standard means of identifying the gene/s underlying an inherited disorder. Linkage is the co-segregation of a genetic region with disease phenotype within a family. Markers, such as restriction fragment length polymorphisms, microsatellites, or single nucleotide polymorphisms (SNPs), are genotyped in family members. Markers close to a disease-causing mutation will be co-inherited with the disease-causing mutation, unless separated by a meiotic event—the closer the marker to the disease-causing gene, the less likely it will be separated at meiosis. An area of linkage within a family may thus extend a considerable genetic distance. Whole genome linkage scans, in which approximately 300–400 microsatellite markers are genotyped in family members, usually result in identification of regions of linkage spanning 10–20 cM (on average ~10–20 million DNA bases). Such a region may harbor many hundreds of genes, and fine mapping (by further marker genotyping and/or sequencing of candidate genes) will usually be necessary to identify the exact causative gene.

## 2.2. Weaknesses of the Linkage Approach

Traditional “parametric” linkage analysis compares the likelihood of the observed transmission of genetic markers in relation to the trait or disease, in the context of a specified model of inheritance. Non-parametric methods not requiring knowledge of mode of inheritance can be used, though are less powerful when the correct mode of inheritance is known. Diseases with late onset of clinical features or with incomplete penetrance are harder to map by linkage due to possible incorrect attribution of disease status among family members. Diseases with significant gene/environment interaction present similar issues, unless all family members are exposed to the same environmental stressors. The ability to map a gene also depends on the number of informative meioses within a pedigree. Thus, large families with many affected individuals—especially distantly-related individuals who will have a high number of meioses and recombination events between them—are the most useful for linkage mapping. Diseases that affect reproductive fitness (such as skeletal dysplasias) are less likely to have such large informative pedigrees. One solution to this problem is to use many families affected by the same condition in order to identify a common linkage region shared by affected persons within each family. This strategy requires that the disease be caused by mutations in the same single common locus in all families, although the causative mutation within this common locus may be unique in each individual family. The approach will, therefore, not be successful for diseases with “phenocopies”, in which a clinical phenotype may arise from mutations in more than one gene. (For example, phenocopies might result if mutations in various genes along a biological pathway resulted in a common phenotypic endpoint.) “Pooling” linkage information from families with disparate underlying causes would not be a successful strategy. Extremely rare diseases are, by definition, extremely rare; obtaining sufficient number of families to pool their linkage information will usually require international collaboration. Lastly, novel/spontaneous mutations (those newly occurring within an individual) cannot be mapped by linkage.

### 2.3. Successes in Linkage Mapping: Monogenic vs. Complex Diseases

Despite these limitations, mapping monogenic diseases by linkage has been quite successful, even for rare diseases, with well over 1000 monogenic (Mendelian) disorders mapped by the turn of the century [1]. Approximately two-thirds of the 400 or so recognized skeletal dysplasias were mapped by linkage or similar approaches by 2010 [4]. In contrast, mapping complex diseases by linkage was much less successful. Complex genetic diseases are typically polygenic in nature: affected individuals have different variants in multiple, but overlapping, sets of genes, with each variant contributing only a small part to the final overall phenotype. Before the era of high throughput microarray genotyping and the advent of genome-wide association studies (GWAS), only a handful of genes had been identified for complex diseases using linkage [5].

### 3. Mapping the FOP Gene Highlights the Challenges

Fibrodysplasia ossificans progressiva (FOP; MIM 135100) is a severe disorder of progressive and extensive extra-skeletal ossification. Heterotopic ossification in FOP begins in childhood within connective tissues, such as skeletal muscle, tendon, and ligament. Onset of bone formation can be induced by trauma, or may occur spontaneously. Bone formation is episodic, leading to cumulative disability and shortened lifespan [6].

Reproductive fitness is low in FOP, resulting in infrequent inheritance and a population frequency (about one per two million) that reflects the rate of new mutations. When the search for the FOP gene began, only very few cases of familial inheritance of FOP had been reported, with most known cases occurring *de novo* in families (reviewed in [7,8]). Although these few family pedigrees suggested an autosomal dominant mode of inheritance, the paucity of families with transmission of FOP made genome-wide linkage analysis, the state-of-the-art approach at the time, an impractical strategy for gene identification.

Eventually, four small pedigrees with autosomal dominant inheritance of FOP were identified, although some had ambiguous clinical features. An initial genome-wide linkage analysis using 240 microsatellite markers (spaced  $\leq 25$  cM) was conducted, although it was recognized that there was incomplete/uneven marker coverage across the genome and many markers lacked sufficient informativity [9]. Whilst the initial linkage analysis focused attention on a chromosome 4 locus [9], further analysis revealed additional linked loci on chromosomes 2 and 6. However, subsequent sequencing analysis of numerous candidate genes in the linkage regions revealed no identifiable mutations. The limited information available at this time about the human genome, including gene locations and sequence, made this process much more challenging.

As genome analysis technologies continued to develop and additional families with autosomal dominant transmission of FOP were identified, further genome-wide linkage studies were performed. These used both a higher density SNP marker panel as well as more dense and informative microsatellite panels, which were combined in a single analysis. The international team involved focused their studies on five two-generation families with stringent and unambiguous phenotypic features of FOP in all affected family members (congenital malformation of the great toes and progressive heterotopic ossification in characteristic anatomic patterns). Consistent linkage was then

demonstrated with the chromosome 2q23–24 interval in all five families (LOD score 2.3) [8], with no other locus segregating with the disease in all pedigrees. Using better characterized and annotated human DNA sequences generated through the Human Genome Project (HGP), we selected candidate genes within the 16.5 Mb linkage interval for sequencing and mutation identification. The interval contained more than 40 known genes, however, concurrent investigations had identified the BMP signaling pathway as altered in FOP [10–13], and genes in this pathway were given high priority. One such gene was the *Activin A type I receptor* gene (*ACVRI*; OMIM 102576; also known as *Alk2* or *ActRI*), encoding a receptor for bone morphogenetic protein (BMP).

DNA sequence analysis of all *ACVRI* protein-coding exons and splice junctions identified a heterozygous c.617G>A (Arg206His; CGC ≥ CAC) mutation present in all affected members in these FOP families, with the same mutation present in multiple sporadic cases of FOP [8].

Cumulative data over the past eight years shows that FOP is caused by this recurrent single nucleotide substitution in >95% of patients ([14–16]). In exceptional cases of FOP, mainly those whose phenotype varies slightly from the description above, affected individuals have mutations at other amino acid positions in *ACVRI* in the glycine-serine (GS) or protein kinase domains [14,17]. Thus far, all patients clinically diagnosed with the “classical” FOP phenotype have *ACVRI* mutations, and these mutations are fully penetrant.

#### 4. Massive Parallel Sequencing: A New Era

The mapping of rare monogenic disorders has been completely transformed by the advent of massive parallel sequencing (MPS), also known as “next-generation” sequencing. In the last few years, the causative genes for dozens of monogenic disorders have been identified using MPS, and the rate of discovery has been exponential. To illustrate this latter point, we recently published a review of MPS in skeletal dysplasias; at the time of paper submission (April 2013) 26 papers had been published using MPS to identify the causative gene for 22 skeletal dysplasias; by the time of paper acceptance (July 2013) a further ten papers had added another six novel skeletal dysplasia genes to the list [2]. Since then, further causative genes for skeletal dysplasias, as well as a host of other Mendelian disorders, have been identified, and it is likely that many of the remaining unmapped monogenic diseases will prove tractable to mapping by MPS.

##### 4.1. MPS Technologies

Disease gene identification by MPS became possible because of three main developments: the technology of sequencing multiple genetic regions simultaneously; the success of the Human Genome Project in providing a complete and reliable reference genome for comparison with the test sequence data; and the availability of large databases of genomic information from healthy individuals, and increasingly in patients with disease, to assist in assessment of variants observed.

The pivotal technological breakthrough for MPS was the development of technologies and platforms for simultaneous sequencing of multiple regions of fragmented DNA or RNA in a single experiment. It is beyond the scope of this paper to provide a comprehensive discussion of these technologies. However, briefly: DNA is fragmented and common adapters are ligated to the fragment ends. The fragments are subsequently amplified by PCR, followed by sequencing-by-synthesis, the

common adapters providing uniform starting templates for both the amplification and sequencing reactions for all fragments (more recent technological developments, so-called “third-generation” technologies, involve sequencing without this step of any preceding amplification, improving both accuracy and speed of MPS). Sequencing-by-synthesis involves addition of bases to a growing strand: as each base is added, a signal is generated and “read” by the software, thus generating the sequence of each fragment. The sequence of each fragment is then mapped against the human genome, allowing identification of genetic variants present in the sequenced individual(s).

Large databases of genetic variation (such as The HapMap project, UK10K, 1000genomes, Human Variome Project, and NCBI dbSNP) are used in interpreting sequence data obtained through MPS: identified variants can be characterized as part of the “normal” variability seen in the population, or as novel or rare variants and thus more likely to be pathogenic. Of note, population genetic variability differs among populations of different ethnicities; the sequence data of an individual should be compared against an ethnically-matched reference genome sequence. Once the sequence data have been generated and compared with the appropriate reference human genome, the data can be analyzed empirically based on the observed inheritance and population prevalence of the condition under examination.

Although MPS was developed for whole-genome sequencing (WGS), targeted sequencing proved more cost-effective and efficient initially. Thus, prior to amplification, a library of fragments containing regions of particular interest may be selected (for example, by using probes that anneal to specific regions of interest, allowing their subsequent identification and isolation for PCR amplification and sequencing). Whole exome sequencing (WES) with capture and sequencing limited to gene exons may be particularly suited for rare Mendelian disorders since prior to the advent of MPS 85% of monogenic diseases were predicted to arise from protein-coding mutations [18]—whether this will hold true as more causative mutations are identified is as of yet unknown.

The power of MPS methodology is illustrated by the many causative genetic mutations identified since its advent, especially since they are frequently mapped through sequencing of remarkably few individuals. Indeed, some disease genes have been identified through sequencing of a single proband [2], although confirmation of pathogenicity requires subsequent validation, such as genetic evidence in other affected individuals and/or functional studies of the candidate gene.

#### *4.2. Mapping Strategies for Monogenic Diseases Using MPS*

The experimental design for mapping a monogenic disease by WES does not necessarily require any prior linkage or association data. Rather it depends on the population frequency of disease, the mode of inheritance (including penetrance), and the presence or absence of consanguinity of the affected individuals. These parameters then determine an appropriately parsimonious experimental design—how many and which individuals should be sequenced and what empirical approach should be adopted for analysis of the sequence data. For example, a rare autosomal recessive condition in a non-consanguineous family is likely to arise from compound heterozygosity; identification of two novel (or very rare) damaging variants in a single gene provides strong evidence of likely causality even if only a single affected individual from the family is sequenced. In contrast, mapping an autosomal recessive condition in a consanguineous family is more difficult. In this circumstance, the disease is more likely to arise from homozygous carriage of a novel (or rare) variant carried by both

parents. However, a high number of homozygous rare variants would be expected due to consanguinity anyway; determining which of these is most likely pathogenic can be difficult. For an autosomal dominant condition, the most parsimonious experimental design is to sequence distantly-related affected individuals with a large number of meioses (and, by implication, recombination events) separating the affected cases—with  $n$  meioses between the individuals, the chance of a dominant variant segregating with affection status by chance is  $1/2^n$ . It is also possible to map *de novo* dominant conditions by sequencing unrelated individuals and analyzing the data either for a single variant carried by all affected individuals or for unique variants carried in a common gene by each individual [19]. These last analysis strategies depend crucially upon correct clinical phenotyping of the unrelated affected individuals. The inclusion of phenocopies in the analysis would decrease the success of mapping the causative gene—unless the stringent parameters of analysis are relaxed to allow for their possible presence. For example, one can search for a common shared gene amongst only a proportion of affected individuals rather than requiring a variant to be present in the same gene in every sequenced case. An alternative approach includes pathway analysis (searching for variants in a common pathway amongst individuals with a common phenotype). From a clinical viewpoint, large online databases cataloguing observed variants/mutations in patients with common conditions are also useful in identifying likely disease-causing mutation(s) (as examples, the Leiden Open Variation Database and the Genome Medicine Database of Japan).

#### 4.3. MPS Limitations

Like all technologies, MPS has limitations. Good sequencing data depend upon sufficient capture of the causative gene by coverage of sufficient depth of sequencing to call homozygous or heterozygous variants (typically, 10-fold coverage is required for calling a homozygous variant and 15-fold for a heterozygous variant). WES off-the-shelf target platforms vary in their coverage of the “whole exome” [20,21], which may result in a gene of interest failing to be sequenced. For example, we (and others) failed to identify the disease-causing mutation in OI type V despite sequencing several families with the condition; the causative mutation was identified in the 5' UTR of *IFITM5* [22,23], a region not captured with the whole exome capture platform we had employed. Less-than-complete whole exome capture can arise for several additional reasons, including new gene annotation subsequent to platform development and production [24]; a manufacturing decision to target only the main transcript for a gene rather than all known transcripts of a gene; and the technical challenges of capturing GC-rich sequences (which are common in the first exons of many genes [25]). Conversely, there are some regions that amplify excessively: if duplicates are not removed, strand-specific PCR-introduced errors may result, skewing variant allele frequencies with consequent effects on variant detection sensitivity and specificity [26]. Regions of genomic sequence similarity may result in non-specificity of target fragment selection—for example, a probe may anneal not only to the desired target exon but also to an unwanted region of high homology which, when incidentally captured and sequenced, results in apparently novel heterozygous variants at those points of difference between the two selected regions (a phenomenon known as multi-mapping [27]).

However, despite these limitations, faced today with the challenge of identifying a rare gene mutation in a disease like FOP, WES of a small number of FOP patients would likely rapidly reveal

the recurrent c.617G>A (R206H) *ACVR1* mutation, leading to quick recognition of *ACVR1* as the disease-causing gene. If MPS technologies had been available when the search for the FOP gene began, the answer could have been found in 15 weeks, not 15 years. Certainly, this has proved to be the case in many other skeletal dysplasias where researchers are faced with similar issues of small families afflicted with diseases having a detrimental effect upon reproduction and lifespan [2,4].

#### 4.4. Rare Variants as a Cause of Complex Diseases?

Although we focused in this paper on the use of new mapping approaches for Mendelian disorders, MPS has also been used for mapping rare variants that contribute to the heritability of complex diseases. The contributions of rare variation in loci that also harbor common susceptibility alleles, or in genomic regions without common susceptibility alleles, are still the subject of active research. Whilst many examples exist of rare variants contributing to loci harboring common variant associations, these are few by comparison with the total number of common variant associations identified to date. Indeed, targeted sequencing of 25 loci associated with autoimmune disorders in nearly 25,000 individuals with six autoimmune phenotypes and just over 17,000 controls failed to identify any rare variants contributing significantly to immune-mediated disease susceptibility [28]. Styrkarsdottir *et al.* used whole genome scanning in the Icelandic population to identify a rare variant in *LGR4* associated with both bone mineral density and fracture risk [29]. WES identified mutations in *WNT1* as the cause of autosomal dominant early-onset osteoporosis in some families; however, as mutations in *WNT1* were also identified in consanguineous families with autosomal recessive osteogenesis imperfecta it would perhaps be more correct to regard the families diagnosed with AD osteoporosis as having a subtle form of OI and/or a monogenic skeletal dysplasia rather than the common polygenic disease of osteoporosis [30,31]. Studies that conducted dense rare-variant genotyping, such as ImmunoChip-based analyses of immune-mediated diseases, have had little success in identifying novel rare variant associations, despite large sample sizes [29,32,33].

## 5. Conclusions and Challenges

After mapping a disease-causing gene, many challenges remain, and many of these challenges are likely to be met through the resources and technologies that continue to build on the foundation of the Human Genome Project.

One important consideration is in understanding the multiple genetic contributions to disease onset and progression in addition to the primary causative gene in monogenic disorders. In FOP, for example, although the main disease characteristics are unique and readily recognized, variability in the age of disease onset, the rate of disease progression, and the severity of disease can be high, even in the context of the same *ACVR1* c.617G>A mutation. Such phenotypic variability is likely influenced by underlying genetic causes. Identification of genetic modifiers that “protect” an individual with an FOP mutation, for example by directing a late onset or less aggressive disease progression, would provide new therapeutic targets and strategies for treating the disease. Similar issues exist for many common heritable diseases, such as osteogenesis imperfecta and Marfan’s syndrome. With further development of genome technologies, the ability to understand phenotypic variability and the participation of genetic modifiers is becoming a reality.



The ultimate challenge is to elucidate the functions of the target gene and the consequences of its mutant forms, and, most importantly, the translation of this knowledge to treatments. Identification of the specific mutation in *ACVR1* has clear and important diagnostic value, providing a means to confirm suspicion of FOP based on toe malformations or, in cases of potential inheritance (and when sequencing in early life becomes more commonplace), a means to diagnose the condition before irreversible clinical manifestations occur allowing for early intervention. Identification of the target pathway and the specific mutation mechanism in FOP has opened up therapeutic strategies for this disease. Additionally, although the roles of BMP signaling in a wide range of tissue development and homeostasis functions had been well established and the signaling pathway and its components were well defined prior to identifying the FOP mutation, the roles of ACVR1/ALK2 in these processes were unrecognized and poorly understood. FOP identified ACVR1 as a key regulator of skeletal development and bone formation, providing an important new focus for skeletal biology and regenerative medicine. This has been the case for many skeletal dysplasias mapped to date, in which the causative gene was often not known to be involved in bone prior to its identification [34]. There are many examples of such findings providing important insights into musculoskeletal development and pathology, and many current treatments have been developed based on genetic discoveries—for example the development of anti-sclerostin antibodies based on the discovery that the high bone mass conditions of sclerosteosis and van Buchem's disease arise from mutations in the gene for sclerostin [35,36]. The power of MPS to map disease-associated mutations will thus benefit not only affected individuals and families, but also lead to a dramatic expansion in our understanding of human diseases. This will inform development of new therapies not only for rare monogenic disorders but also for diseases common in the general population.

### Acknowledgments

The authors acknowledge the help of Murray Hargrave in editing the manuscript. This work was supported in part by the International Fibrodysplasia Ossificans Progressiva Association, the Center for Research in FOP and Related Disorders, the Cali-Weldon Professorship in FOP (to EMS), and the National Institutes of Health (NIH R01-AR41916).

### Author Contributions

Emma Duncan, Matthew Brown and Eileen M. Shore all wrote the manuscript.

### Conflicts of Interest

The authors declare no conflict of interest.

### References

1. Peltonen, L.; McKusick, V.A. Genomics and medicine. Dissecting human disease in the postgenomic era. *Science* **2001**, *291*, 1224–1229.
2. Lazarus, S.; Zankl, A.; Duncan, E.L. Next-generation sequencing: A frameshift in skeletal dysplasia gene discovery. *Osteoporos. Int.* **2014**, *25*, 407–422.

3. Garrod, A.E. The Croonian Lectures on inborn errors of metabolism. *Lancet* **1908**, *172*, 214–220.
4. Warman, M.L.; Cormier-Daire, V.; Hall, C.; Krakow, D.; Lachman, R.; LeMerrer, M.; Mortier, G.; Mundlos, S.; Nishimura, G.; Rimoin, D.L.; *et al.* Nosology and classification of genetic skeletal disorders: 2010 revision. *Am. J. Med. Genet. A* **2011**, *155A*, 943–968.
5. Glazier, A.M.; Nadeau, J.H.; Aitman, T.J. Finding genes that underlie complex traits. *Science* **2002**, *298*, 2345–2349.
6. Shore, E.M.; Kaplan, F.S. Inherited human diseases of heterotopic bone formation. *Nat. Rev. Rheumatol.* **2010**, *6*, 518–527.
7. Shore, E.M.; Feldman, G.J.; Xu, M.; Kaplan, F.S. The genetics of fibrodysplasia ossificans progressiva. *Clin. Rev. Bone Miner. Metab.* **2005**, *3*, 201–204.
8. Shore, E.M.; Xu, M.; Feldman, G.J.; Fenstermacher, D.A.; Cho, T.-J.; Choi, I.H.; Connor, J.M.; Delai, P.; Glaser, D.L.; LeMerrer, M.; *et al.* A recurrent mutation in the BMP type I receptor ACVR1 causes inherited and sporadic fibrodysplasia ossificans progressiva. *Nat. Genet.* **2006**, *38*, 525–527.
9. Feldman, G.; Li, M.; Martin, S.; Urbanek, M.; Urtizberea, J.A.; Fardeau, M.; LeMerrer, M.; Connor, J.M.; Triffitt, J.; Smith, R.; *et al.* Fibrodysplasia ossificans progressiva, a heritable disorder of severe heterotopic ossification, maps to human chromosome 4q27–31. *Am. J. Hum. Genet.* **2000**, *66*, 128–135.
10. Gannon, F.H.; Kaplan, F.S.; Olmsted, E.; Finkel, G.C.; Zasloff, M.A.; Shore, E. Bone morphogenetic protein 2/4 in early fibromatous lesions of fibrodysplasia ossificans progressiva. *Hum. Pathol.* **1997**, *28*, 339–343.
11. Viridi, A.S.; Shore, E.M.; Oreffo, R.O.; Li, M.; Connor, J.M.; Smith, R.; Kaplan, F.S.; Triffitt, J.T. Phenotypic and molecular heterogeneity in fibrodysplasia ossificans progressiva. *Calcif. Tissue Int.* **1999**, *65*, 250–255.
12. Fiori, J.L.; Billings, P.C.; Serrano de la Pena, L.S.; Kaplan, F.S.; Shore, E.M. Dysregulation of the BMP-p38 MAPK signaling pathway in cells from patients with fibrodysplasia ossificans progressiva (FOP). *J. Bone Miner. Res.* **2006**, *21*, 902–909.
13. Serrano de la Pena, L.S.; Billings, P.C.; Fiori, J.L.; Ahn, J.; Kaplan, F.S.; Shore, E.M. Fibrodysplasia ossificans progressiva (FOP), a disorder of ectopic osteogenesis, misregulates cell surface expression and trafficking of BMPRIA. *J. Bone Miner. Res.* **2005**, *20*, 1168–1176.
14. Kaplan, F.S.; Xu, M.; Seemann, P.; Connor, J.M.; Glaser, D.L.; Carroll, L.; Delai, P.; Fastnacht-Urban, E.; Forman, S.J.; Gillesen-Kaesbach, G.; *et al.* Classic and atypical fibrodysplasia ossificans progressiva (FOP) phenotypes are caused by mutations in the bone morphogenetic protein (BMP) type I receptor ACVR1. *Hum. Mutat.* **2009**, *30*, 379–390.
15. Zhang, W.; Zhang, K.; Song, L.; Pang, J.; Ma, H.; Shore, E.M.; Kaplan, F.S.; Wang, P. The phenotype and genotype of fibrodysplasia ossificans progressiva in China: A report of 72 cases. *Bone* **2013**, *57*, 386–391.
16. Shore, E.M. Department of Orthopaedic Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. Unpublished work, 2014.

17. Culbert, A.L.; Chakkalakal, S.A.; Convente, M.R.; Lounev, V.Y.; Kaplan, F.S.; Shore, E.M. Fibrodysplasia (myositis) ossificans progressiva. In *Genetics of Bone Biology and Skeletal Disease*; Thakker, R.V., Whyte, M.P., Eisman, J.A., Igarashi, T., Eds.; Academic Press: New York, NY, USA, 2013; pp. 375–393.
18. Cooper, D.N.; Krawczak, M.; Antonorakis, S.E. The nature and mechanisms of human gene mutation. In *The Metabolic and Molecular Bases of Inherited Disease*; Scriver, C., Beaudet, A.L., Sly, W.S., Valle, D., Eds.; McGraw-Hill: New York, NY, USA, 1995; pp. 259–291.
19. Zankl, A.; Duncan, E.L.; Leo, P.J.; Clark, G.R.; Glazov, E.A.; Addor, M.C.; Herlin, T.; Kim, C.A.; Leheup, B.P.; McGill, J.; *et al.* Multicentric carpotarsal osteolysis is caused by mutations clustering in the amino-terminal transcriptional activation domain of MAFB. *Am. J. Hum. Genet.* **2012**, *90*, 494–501.
20. McInerney-Leo, A.M.; Marshall, M.S.; Gardiner, B.; Benn, D.E.; McFarlane, J.; Robinson, B.G.; Brown, M.A.; Leo, P.J.; Clifton-Bligh, R.J.; Duncan, E.L. Whole exome sequencing is an efficient and sensitive method for detection of germline mutations in patients with pheochromocytomas and paragangliomas. *Clin. Endocrinol.* **2014**, *80*, 25–33.
21. McInerney-Leo, A.M.; Marshall, M.S.; Gardiner, B.; Coucke, P.J.; van Laer, L.; Loeys, B.L.; Summers, K.M.; Symoens, S.; West, J.A.; West, M.J.; *et al.* Whole exome sequencing is an efficient, sensitive and specific method of mutation detection in osteogenesis imperfecta and Marfan syndrome. *Bonekey Rep.* **2013**, *2*, 456.
22. Cho, T.J.; Lee, K.E.; Lee, S.K.; Song, S.J.; Kim, K.J.; Jeon, D.; Lee, G.; Kim, H.N.; Lee, H.R.; Eom, H.H.; *et al.* A single recurrent mutation in the 5'-UTR of IFITM5 causes osteogenesis imperfecta type V. *Am. J. Hum. Genet.* **2012**, *91*, 343–348.
23. Semler, O.; Garbes, L.; Keupp, K.; Swan, D.; Zimmermann, K.; Becker, J.; Iden, S.; Wirth, B.; Eysel, P.; Koerber, F.; *et al.* A mutation in the 5'-UTR of IFITM5 creates an in-frame start codon and causes autosomal-dominant osteogenesis imperfecta type V with hyperplastic callus. *Am. J. Hum. Genet.* **2012**, *91*, 349–357.
24. Brunham, L.R.; Hayden, M.R. Hunting human disease genes: Lessons from the past, challenges for the future. *Hum. Genet.* **2013**, *132*, 603–617.
25. Hoischen, A.; Gilissen, C.; Arts, P.; Wieskamp, N.; van der Vliet, W.; Vermeer, S.; Steehouwer, M.; de Vries, P.; Meijer, R.; Seiquer, J.; *et al.* Massively parallel sequencing of ataxia genes after array-based enrichment. *Hum. Mutat.* **2010**, *31*, 494–499.
26. Koboldt, D.C.; Ding, L.; Mardis, E.R.; Wilson, R.K. Challenges of sequencing human genomes. *Brief Bioinform.* **2010**, *11*, 484–498.
27. Hashimoto, T.; de Hoon, M.J.; Grimmond, S.M.; Daub, C.O.; Hayashizaki, Y.; Faulkner, G.J. Probabilistic resolution of multi-mapping reads in massively parallel sequencing data using MuMRescueLite. *Bioinformatics* **2009**, *25*, 2613–2614.
28. Hunt, K.A.; Mistry, V.; Bockett, N.A.; Ahmad, T.; Ban, M.; Barker, J.N.; Barrett, J.C.; Blackburn, H.; Brand, O.; Burren, O.; *et al.* Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. *Nature* **2013**, *498*, 232–235.

29. Styrkarsdottir, U.; Thorleifsson, G.; Sulem, P.; Gudbjartsson, D.F.; Sigurdsson, A.; Jonasdottir, A.; Jonasdottir, A.; Oddsson, A.; Helgason, A.; Magnusson, O.Y.; *et al.* Nonsense mutation in the LGR4 gene is associated with several human diseases and other traits. *Nature* **2013**, *497*, 517–520.
30. Laine, C.M.; Joeng, K.S.; Campeau, P.M.; Kiviranta, R.; Tarkkonen, K.; Grover, M.; Lu, J.T.; Pekkinen, M.; Wessman, M.; Heino, T.J.; *et al.* WNT1 mutations in early-onset osteoporosis and osteogenesis imperfecta. *N. Engl. J. Med.* **2013**, *368*, 1809–1816.
31. Pyott, S.M.; Tran, T.T.; Leistriz, D.F.; Pepin, M.G.; Mendelsohn, N.J.; Temme, R.T.; Fernandez, B.A.; Elsayed, S.M.; Elsobky, E.; Verma, I.; *et al.* WNT1 mutations in families affected by moderately severe and progressive recessive osteogenesis imperfecta. *Am. J. Hum. Genet.* **2013**, *92*, 590–597.
32. Parkes, M.; Cortes, A.; van Heel, D.A.; Brown, M.A. Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat. Rev. Genet.* **2013**, *14*, 661–673.
33. Keupp, K.; Beleggia, F.; Kayserili, H.; Barnes, A.M.; Steiner, M.; Semler, O.; Fischer, B.; Yigit, G.; Janda, C.Y.; Becker, J.; *et al.* Mutations in WNT1 cause different forms of bone fragility. *Am. J. Hum. Genet.* **2013**, *92*, 565–574.
34. Glazov, E.A.; Beleggia, F.; Kayserili, H.; Barnes, A.M.; Steiner, M.; Semler, O.; Fischer, B.; Yigit, G.; Janda, C.Y.; Becker, J.; *et al.* Whole-exome re-sequencing in a family quartet identifies POP1 mutations as the cause of a novel skeletal dysplasia. *PLoS Genet.* **2011**, *7*, e1002027.
35. Balemans, W.; Patel, N.; Ebeling, M.; van Hul, E.; Wuyts, W.; Lacza, C.; Dioszegi, M.; Dikkers, F.G.; Hildering, P.; Willems, P.J.; *et al.* Identification of a 52 kb deletion downstream of the SOST gene in patients with van Buchem disease. *J. Med. Genet.* **2002**, *39*, 91–97.
36. Van Hul, W.; Balemans, W.; van Hul, E.; Dikkers, F.G.; Obee, H.; Stokroos, R.J.; Hildering, P.; Vanhoenacker, F.; van Camp, G.; Willems, P.J. Van Buchem disease (hyperostosis corticalis generalisata) maps to chromosome 17q12–q21. *Am. J. Hum. Genet.* **1998**, *62*, 391–399.