Review

Genomic Context and Mechanisms of the *ACVR1* Mutation in Fibrodysplasia Ossificans Progressiva

Roberto Ravazzolo [1,*,†] and Renata Bocciardi [1,2]

1   Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Sciences (DiNOGMI), University of Genova, 16100 Genova, Italy; bocciardi@unige.it
2   UOC Genetica Medica, IRCCS Giannina Gaslini, 16100 Genova, Italy
*   Correspondence: rravazzo@gmail.com
†   Retired.

Abstract: Basic research in Fibrodysplasia Ossificans Progressiva (FOP) was carried out in the various fields involved in the disease pathophysiology and was important for designing therapeutic approaches, some of which were already developed as ongoing or planned clinical trials. Genetic research was fundamental in identifying the FOP causative mutation, and the astonishing progress in technologies for genomic analysis, coupled to related computational methods, now make possible further research in this field. We present here a review of molecular and cellular factors which could explain why a single mutation, the R206H in the *ACVR1* gene, is absolutely prevalent in FOP patients. We also address the mechanisms by which FOP expressivity could be modulated by *cis*-acting variants in the *ACVR1* genomic region in human chromosome 2q. Finally, we also discuss the general issue of genetic modifiers in FOP.

Keywords: Fibrodysplasia Ossificans Progressiva; ACVR1; human genome; *cis*-regulatory elements; genetic modifiers; gene expression

1. Introduction

Research in rare diseases is of paramount importance to build knowledge for the design of potential therapeutic approaches. For Fibrodysplasia Ossificans Progressiva (FOP), the state of the art of research is highly consistent with such relationships between research findings and their application to clinics in term of therapeutic strategies. The International Fibrodysplasia Ossificans Progressiva Association (IFOPA, https://www.ifopa.org/ (accessed on 4 February 2021)) indicates a list of 16 therapeutic proposals, among them some already developed as ongoing or planned clinical trials (https://clinicaltrials.gov (accessed on 4 February 2021)). All of them derive from research on FOP pathophysiology and, because the cause of the disease is a mutation in the *ACVR1* gene, encoding the Alk2 receptor for Bone Morphogenetic Proteins (BMPs), which induces an anomalous activation of signaling, the majority of the proposed therapeutic strategies are focused on the inhibition of BMP signaling.

Most of the research directed to find ways to inhibit the BMP signaling took advantage of the availability of murine models. A very recent article published in the Special Issue "Fibrodysplasia Ossificans Progressiva: Studies on Disease Mechanism towards Novel Therapeutic Approaches" of Biomedicines [1] describes in detail several murine models utilized in studies on FOP. Besides mice, another animal model, zebrafish, has provided important information about the first inhibitor of BMP signaling [2] and, very recently, about the relationship between the *ACVR1* mutation and its ligand dependency [3]. FOP models were also generated in Drosophila [4]. Many cellular models have been utilized to investigate FOP disease mechanisms, which were useful for studying many peculiar aspects of FOP pathophysiology but had the inevitable limitations intrinsic to the model itself. However, a fundamental advancement in cell biology techniques, the induced Pluripotent

Stem Cell (iPSC) derived from FOP patients, proved to be suitable for application in the study of the disease and was used to demonstrate the role of Activin A in BMP signaling starting from the mutated ACVR1 receptor [5], also reported in an article that described the first conditional *ACVR1(R206H)* murine model [6]. The FOP iPSCs were subsequently used for further characterization of mechanisms and drug screening [7,8].

Important efforts have been devoted to research on pluripotent cell types which can be sensitive to inducing signals and differentiate into bone [9–12], and we think that future studies in this field will provide additional information on the contribution of different cell types in the FOP osteogenic process.

Among the determinants of FOP onset and progression, other topics in the FOP pathophysiology are important to understand the disease mechanisms and possibly become target of potential innovative therapeutic approaches. The local environment in which osteogenesis ultimately takes place is a complex assembly of multiple cell types and molecular components which play different roles in the FOP pathogenesis and, because of the complexity, much needs to be investigated and clarified. Local inflammation is present in all sites where new bone will be formed, sometimes with clear apparent signs in typical flare-ups, but also in an underhand manner, which is now possible to perceive thanks to modern diagnostic techniques like MRI coupled to PET-CT [13]. Pathological examination of early FOP lesions showed the presence of cells of innate immunity, lymphocytes, macrophages, and mast cells (reviewed in Shore, 2010) [14], and the large array of inflammatory molecules released in the local environment are described [15]. We think that many features of local inflammation in the FOP lesions have to be further investigated to clarify their role in the disease pathogenesis and, therefore, become potential targets of therapeutic intervention. Hypoxia associated with inflammation in the FOP lesion is one of the factors which has been described as an important contributor to heterotopic ossification (HO) [16] with potential application for therapeutic approaches. The fibroproliferative process described in the evolution of the FOP local lesion [14] is responsible for extracellular matrix production and tissue stiffness, which appears to be increased in FOP compared to the repair of muscle damage taking place in the absence of ACVR1 mutation [17,18] and to be another factor favoring HO in the local FOP lesion. We remark that all different cell types which play a role in the local environment where the HO develops in FOP are carrying the *ACVR1* mutation, and that such mutation is probably influencing to some extent their function, increasing the complexity of pathogenic factors that intervene in the onset and progression of the disease.

FOP, like many other genetic disorders caused by single gene mutation, shows variability in clinical manifestation, like age of onset, frequency and outcome of flareups, localization of heterotopic bone and others, which can depend on several reasons, among them genetic/genomic factors. In the recent past, DNA sequencing was a reliable procedure to find gene mutations causative of such diseases (see Shore, 2006 for FOP) [19]. Nowadays, genome-wide sequencing studies can be carried out with advanced technology and lower cost, which can allow us to investigate the genome searching for genetic factors that influence disease penetrance and expressivity.

In this article, we present a speculative analysis of two aspects of the FOP pathogenesis related to the *ACVR1* mutation itself: (a) why a single mutation has such a high prevalence in FOP patients; and (b) how the expression of *ACVR1* can be modulated by *cis*-acting variants in the surrounding genomic region of the gene in human chromosome 2q, thus affecting FOP expressivity. We also discuss the general issue of genetic modifiers in FOP.

2. The *ACVR1* Mutation in FOP

FOP is caused by mutations in the *ACVR1* gene, encoding the Alk2 Bone Morphogenetic Protein Type I receptor. More than 95% of affected individuals carry a heterozygous recurrent mutation, a G to A substitution at nucleotide c.617 that changes a CGC codon encoding arginine to CAC, encoding histidine (R206H) [19]. This residue is located in

the GS domain of the receptor protein. The remaining small percentage of patients carry different mutations in the same GS domain or in the kinase domain of the protein [20–22].

The very high recurrence rate of the c.617G>A mutation stimulates interest in understanding the reason of such recurrence. In most cases the mutation is a de novo event likely occurring in a germ cell, since very few cases are described as inherited from an affected parent with an autosomal dominant pattern [19]. A rare case of germ cell mosaicism is also described in two half sisters having the same mother and different fathers [23]; mosaicism caused by a *de novo* mutation in a post-zygotic phase cannot be excluded, although no evidence of such an event is available in the literature.

To explore the reasons for the high recurrence rate of the c.617G>A substitution, we have to consider the context in which this type of mutation occurs and its effect on the cell where it takes place. The base substitution occurs in a CpG dinucleotide in a coding exon of the gene devoid of CpG high density (CpG island). This base change is the most common among nucleotide substitutions [24], due to the high mutation rate at CpG dinucleotide where cytosine is frequently methylated (5-mC). 5-mC then undergoes spontaneous deamination that results in a C>T change and a G:T mismatch that, in the following replication step, is fixed as a G>A transition in the interested codon (Figure 1). Interestingly, this mutation was not exclusively found in humans. The same identical mutation in *ACVR1* was reported in two cats with FOP [25], confirming the general validity of this mutational mechanism.
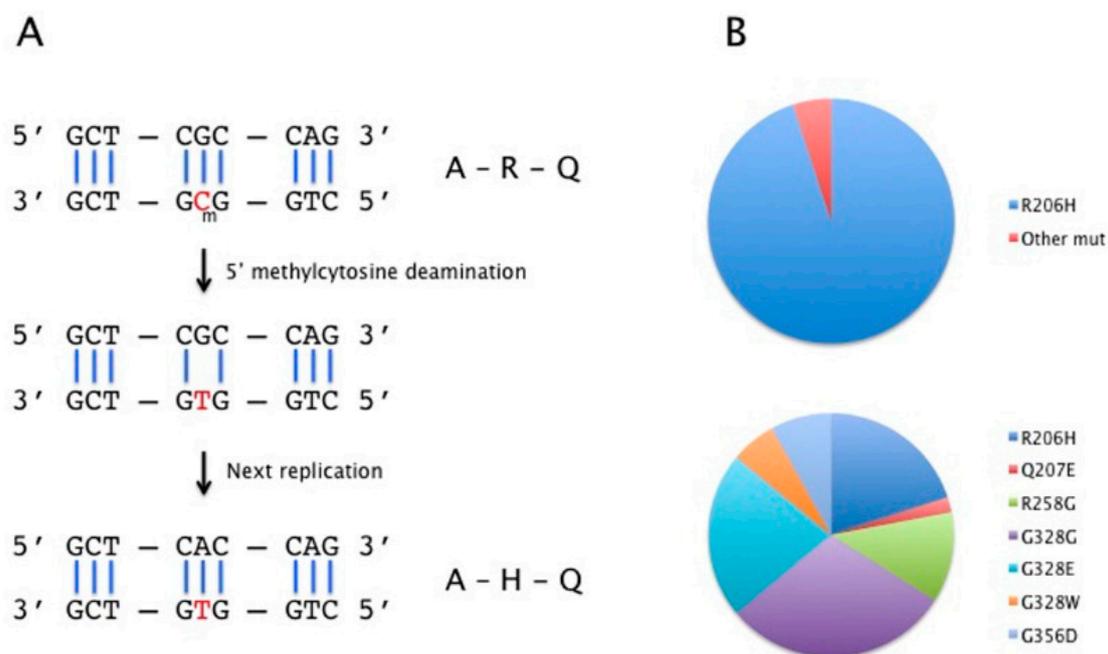

Figure 1. (**A**) Mechanism of the G>A mutagenesis; (**B**) Percentage of ACVR1 mutations in Fibrodysplasia Ossificans Progressiva (FOP) (upper) and in Diffuse Intrinsic Pontine Glioma (DIPG) (lower). The red color in part A indicates the interested nucleotide.

Besides FOP, mutations in the *ACVR1* gene have been described as somatic events in cancer tissues from 20–30% of patients affected with Diffuse Intrinsic Pontine Glioma (DIPG) [22,26,27]. It is interesting to observe that ACVR1 in DIPG is mutated in the same gene domains as in FOP but, unlike FOP, the very high frequency of c.617G>A is not detected in those somatic mutations (Figure 1), implying that the cell and the time when the mutational event occurs make a difference. The fundamental point is that the highly recurrent FOP mutation occurs in germ cells where mutation is a very rare event compared to somatic cells; the DNA methylation level is subjected to variation in the different phases of germ cell development and the processes of male and female germline development vary.

It is generally stated that the majority of germline de novo mutations have a paternal origin and that an increasing paternal age favors de novo mutations in the offspring (see Goldman, 2019, as a comprehensive review) [28], due to the continuous division of spermatogonial cells generating replication errors and failure to repair that is progressive with age. An article published in 1979, long before the discovery of the genetic cause of FOP, reported a paternal age effect [29]. Compared to spermatogenesis, oogenesis is concluded before birth and no further cell divisions take place after birth. However, maternal-derived germline de novo mutations also occur, particularly in some genomic regions [30,31], which would imply that, in those cases, the mechanism of mutation is different from replication errors. It is also reported that CpG sites in cells of the male germline are methylated at a higher rate than in the female germline, making the sites more prone to spontaneous deamination and C>T transitions [32].

In highly recurrent de novo mutations, like the c.617G>A causative of FOP, named "selfish" mutations, it has been proposed that they confer spermatogonial selection leading to clonal expansion of mutated spermatogonia. Examples of such mutations are found in *HRAS*, causing Costello syndrome, in *FGFR3*, causing Achondroplasia, and other genes [33]. Since this clonal expansion develops with time, this mechanism is better compatible with the occurrence of mutation in men with advancing age but a selective advantage in the maternal germline cannot be excluded.

Another intriguing general issue regards individual genetic background which may affect the ability of DNA repair, thus allowing differential efficiency of repair and individual differences in the chance of developing mutation [28].

Concerning the specific role of *ACVR1*/Alk2 in gametogenesis, experiments in the mouse demonstrated that Alk2$^{-/-}$ embryos cannot form Primordial Germ Cells (PGCs), and, in Alk2$^{+/-}$ heterozygous embryos, the number of PGCs was reduced when compared with wild-type littermates, demonstrating that Alk2 is necessary for PGC formation in vivo [34]. We are not aware of any evidence of how an activating mutation might affect this process and other steps of the gametogenesis.

Although we cannot define the precise mechanism underlying the FOP recurrent mutation, the significant correlation between the methylation level and the germline mutation rate at CpG sites suggests that this mechanism may be causative of the FOP G>A substitution.

In our experience, one FOP patient who carried a disease causative mutation different from the recurrent c.617G>A substitution, c.774G>C leading to the R258S aminoacid change in the kinase domain of the protein, was also heterozygous for the c.44C>G substitution, causing the A15G aminoacid change in the protein signal peptide. The c.44C>G is reported in the dbSNP (rs13406336) with a global frequency of the minor allele of 0.007 according to 1000 Genomes. This variant was inherited from the healthy father and the cDNA analysis showed that it resulted in *cis* configuration with the R258S causative mutation, thus showing that it occurred in the paternal allele [35]. Sequencing of trios (father, mother, FOP affected patient) could detect analogous events that might demonstrate the parental origin of the causative mutation. In particular, for patients carrying the c.617G>A substitution, the occurrence in *cis* with the signal peptide variant would be useful to attribute the parental origin of the *de novo* recurrent mutation. However, this variant is rare, and we are not aware of any other FOP affected individual with such occurrence. Interestingly, a synonymous variant in exon 4, c.270G>A (dbSNP: rs2227861), has frequency of the G allele of 0.35 and A 0.65 according to 1000 Genomes, which makes it suitable for trios analysis.

3. Regulatory Genetic Variants in *cis*

DNA sequencing is the most important tool available to diagnose genetic diseases caused by single gene mutations and has uncovered a great number of causative variants that can be analyzed to establish genotype–phenotype correlations and disease mechanisms. In the case of rare monogenic disorders, sequencing of coding regions of protein encoding

genes is mainly performed. Concerning FOP, sequencing is routinely carried out for diagnosis of FOP [36], which is required to confirm the clinical diagnosis.

In the previous section we describe the c.44C>G variant responsible for the A15G aminoacid change in the signal peptide of the *ACVR1* receptor protein, which is not a disease causing variant. Nevertheless, we cannot exclude that this rare variant in the signal peptide could affect the efficiency of transfer of the receptor protein to the plasma membrane; however, neither we nor others have done experiments to prove such a functional effect.

The study of the so called complex diseases, usually common conditions, has used other methods, such as Genome Wide Association Studies (GWAS), which have detected a huge amount of DNA variants, essentially single-nucleotide variants (SNVs), showing statistically significant association with complex disease phenotypic traits. The majority of these variants are common, having a frequency >1% in the population and, interestingly, are located outside of coding regions, likely involved in regulation of gene expression and, in few cases, proved to have a functional role [37]. Such a large mass of data required the creation of databases and analytical methods necessary to expand the initial essential information derived from the Human Genome Project. The Genotype-Tissue Expression (GTEx) project [38] is a public resource that provides important information on the regulation of gene expression in tissues also at the single gene level. GTEx has built a map of variants that play a role in gene regulation, named expression quantitative trait loci (eQTLs) [39]. In 2020, the GTEx Consortium published its final set of studies analyzing genotype data from almost 1000 post-mortem donors completely genotyped by whole genome sequencing and more than 17,000 RNA-seq samples across 54 tissue sites and 2 cell lines, reporting the detection of eQTLs in 48 tissues (GTEx Portal, see website). Although the aim of GTEx is mainly directed to the study of common diseases, the amount of information collected and publicly available is an invaluable resource for researchers who study human genetic diseases including rare genetic diseases caused by single-gene mutations. For these diseases, one important issue is how the concomitant occurrence of a pathogenic variant and a regulatory variant on the same allele may vary (increase or decrease) the expression of the mutated gene product, in particular when the pathogenic variant is present in heterozygosity. If such imbalance takes place, one can expect that the disease phenotype varies according to the amount of mutated protein. The variants in the GTEx portal can be viewed in dbSNP at NCBI (see website), the public domain archive for human single nucleotide variations, microsatellites, and small-scale insertions and deletions, that has a "genomic view" resource through which any single locus is placed in the human genome and can be further investigated by the UCSC genome browser. We then used the data of the ENCODE Consortium (see website) which has built a Registry of candidate *cis*-Regulatory Elements (CREs) based on signatures derived from DNase hypersensitivity, histone modifications (H3K4me3, H3K27ac), and CTCF ChIP-seq data. Currently, the Registry includes more than 900,000 CREs in the human genome classified either as distal enhancer-like or promoter-like or CTCF-only elements showed in different colors in the UCSF Genome Browser. Finally, we used the PROMO public resource (see website) to predict putative transcription factor binding sites (TFBS) in an unknown DNA sequence, which uses matrices constructed from TFBS defined in the TRANSFAC® database.

We applied the above online repositories to study how variants in *cis* might affect the FOP disease manifestation.

In FOP, the basis for this study is made more favorable by the more than 95% frequency of patients carrying the same mutation in the *ACVR1* gene (see above); then, other variants in *cis* exert their regulatory effect, if any, on the expression of the allele that carries the same genetic mutation. Moreover, for FOP, a sort of indirect evidence for a dependence of phenotypic features on the allelic composition was provided in a mouse model described by Lees Shepard et al. [12]. Heterozygous mice carrying the Acvr1 R206H mutation in which the wild type allele was deleted showed great increase in injury-induced HO compared to

the heterozygous ones. This finding suggests that the HO phenotype is dependent on a gene dosage effect.

Accessing the GTEx portal, we obtained information on eQTLs that affect the expression of the ACVR1 gene, which consists of a list of 62 variants. These variants, ordered from the most centromeric to the most telomeric, are located in an interval of the Chromosome 2q spanning from position 156,878,023 to 158,771,340 that contains all the eQTLs listed in GTEx. The SNVs list in the GTEx portal reports, for each of them, the tissue where the effect on expression was detected. However, we know that *ACVR1* is expressed in a large variety of tissues and cell types; then, we may assume that an eQTL exerts its effect in many or most tissues. We then localized these variants using the dbSNP "genomic view" resource and the UCSC genome browser where CREs made available by the ENCODE Consortium (see website) are annotated. The map of the 62 eQTLs for the *ACVR1* gene in this interval is reported in Figure 2. It is interesting to note that only a few of the 62 variants map in close vicinity to the ACVR1 gene or inside the gene region, while most of them are located relatively far, many of them arranged in two large clusters. Another consideration was that eight of the 62 SNVs are located inside a CRE (7 defined as distal enhancer-like and 1 as promoter-like) as reported in Table 1.
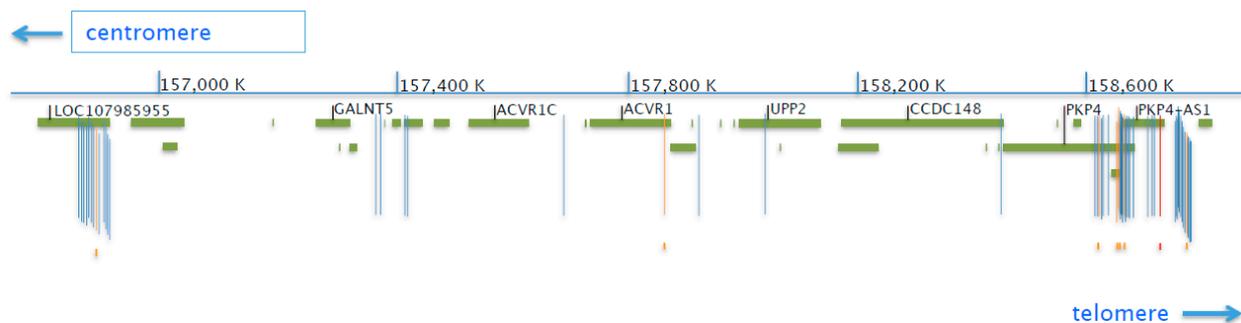

Figure 2. Map of the 62 eQTLs of the *ACVR1* gene (modified from the NCBI Sequence viewer). Green bars represent genes annotated in the NCBI sequence viewer at 30% magnification spanning a genomic region of around 2 Mb of the chromosome 2, from position 156,800,000 upstream of LOC107985955 to 158,816,000 in the *DAPL1* gene region. Vertical bars represent the position of each of the 62 single-nucleotide variants (SNVs) listed by the GTEx portal as affecting the *ACVR1* gene expression; colors are blue, orange for variants lying in a distal-like enhancer element, or red for the one lying in a promoter-like element. The small spots under the bars indicate the position of distal-like enhancer, orange color, or promoter elements, red color, according to ENCODE. Numbers above the map indicate the position in the chromosome according to the human genome GRCh38/hg38 release.

It is likely that, because of linkage disequilibrium, most of the mapped variants are associated with one or a few true functional variants that we cannot identify but which could also be included among those present in the map, which affect the expression of the gene. Therefore, we chose some criteria to indicate the ones which we might consider as possible candidate functional variants. The main criteria applied here were the location inside a regulatory element (according to ENCODE), conservation among species, and change of transcription factor binding (either abolition or introduction) dependent on the change of nucleotide sequence in the alternative alleles. For each of them, we analyzed TFBS by the PROMO resource, comparing the sequence carrying the major allele flanked by 50 nucleotides on each side with the same sequence carrying the alternative allele; and we found that for three of them, the alternative variant does not affect the binding of transcription factors. The indication of a variant as a candidate suggests that it should be tested by appropriate experimental methods to verify its functional role.

Table 1. List of the variants selected from the GTEx eQTLs affecting the expression of the *ACVR1* gene and located inside a candidate *cis*-Regulatory Region (CRE).

	dbSNP ID	Position	Alleles	Conservation	CRE	Effect on TF Binding
1	rs7586274	chr2:156896695 LOC107985955 intron	C/T 98/2	C: Homo, Chimp, Rhesus T: Mouse, Rat, Dog	EH38E2043429 distal enhancer-like	No effect
2	rs568810076	chr2:157866444 *ACVR1* intron 1	G repeat 7G/2G 97/35G/4G/3G	poor conservation	EH38E2043925 distal enhancer-like	Yes: alleles with 5G and 2G miss a WT1 bs compared to 7G
3	rs79473991	chr2:158621197 *PKP4* intron	A/G 98/2	A: Homo, Rhesus, Mouse, Dog	EH38E2044280 distal enhancer-like	No effect
4	rs59520356	chr2_158651353 *PKP4* intron	A/G 93/7	A: Homo, Rhesus, Mouse, Dog, Elephant	EH38E204429 distal enhancer-like	No effect
5	rs58540852	chr2:158658609 *PKP4* intron	A/T 93/7	A: Homo, Rhesus T: Mouse, Elephant	EH38E2044304 distal enhancer-like	Yes: the minor allele acquires a GATA1 bs and loses a STAT5A bs
6	rs73006932	chr2:158662549 *PKP4* intron	G/C 92/8	G: Homo, Rhesus, Mouse C: Dog, Elephant	EH38E2044306 distal enhancer-like	Yes: the minor allele loses ATF3, PR B, PR A bs
7	rs17231614	chr2:158724233 *PKP4*-AS1 intron	G/T 96/4	G: Homo, Rhesus, Mouse, Elephant	EH38E2044344 promoter-like	Yes: the minor allele loses ETF bs and acquires RXR-α bs
8	rs967380	chr2:158766477	T/C 20/80	T: Homo, Rhesus, Dog, Elephant	EH38E2044392 distal enhancer-like	No effect

dbSNP ID, variant accession number in the dbSNP database; Position reports the nucleotide position in chromosome 2q according to the human genome GRCh38/hg38 release; Alleles reports the frequency of the alternative alleles according to the 1000 Genomes Phase 3 dataset; Conservation reports conservation of the alleles across species as available in the UCSC Genome Browser; CRE reports the code of the *cis* Regulatory Element according to the ENCODE resource, annotated in the UCSC Genome Browser; the last column reports the effect on Transcription Factor binding sites analyzed by the PROMO resource at the ALGGEN server, comparing sequences carrying the alternative alleles. TF, Transcription Factor; bs, binding site.

A crucial point to evaluate the significance of a single variant in *cis* with the *ACVR1* causative mutation is the measure of allele-specific expression (ASE). This can be done through the analysis of RNA-seq reads that include heterozygous SNPs, allowing one to assign the read to an allele. Allelic imbalance occurs when the two alleles of a gene are differentially expressed and the magnitude of the imbalance can be quantified by allelic fold change (aFC) [40]. Although it is too simplistic to attribute an expression imbalance to a single variant, rather than to a cumulative effect of more variants on each allele, we cannot exclude that one or some of the variants within a haplotype have a stronger effect on ASE. Analysis of a specific allele requires the reconstruction of the phase of variants that are on the same chromosome, including the causative mutation. If a patient's parents are available for genotyping, phase reconstruction can be attempted, or long read sequencing by new sequencing technology can be utilized to obtain allele phasing, and interpretation of available data sets by informatics means can be of great help.

Modifier genetic variants, of course, can be located in *trans*-acting genes, and we expect that many of them have a role in phenotype variability dependent on the individual genetic background, as described in the following section.

The knowledge of the complex pattern of regulatory variants which can cause up- or down-regulation of the mutant allele could provide an explanation on one of the determinants, likely more than one, of phenotypic variability. Specifically for FOP, variability of disease manifestation is intrinsic to the unpredictable and intermittent disease course, and is therefore complex to evaluate, but the data collected by the natural history studies [41] and those expected from the registry organized by the IFOPA [42,43] will potentially allow one to better categorize the disease phenotypic features, which in turn can be related to etiological factors, among them genomic data underlying regulation of gene expression.

In diseases in which a mutation causes a gain of function effect, like the *ACVR1* mutation in FOP, down-regulation of the mutant allele could potentially reduce the pathogenic consequence of the activating variants. Approaches to target allele-specific gene expression to treat genetic disorders by modifying a gain of function disease mechanism are reported

for a number of conditions, such as Huntington disease and Facioscapulohumeral muscular dystrophy [44], and the advancements of genome editing technologies will certainly provide new strategies to treat genetic disorders by these means. On the other hand, considering the effect of the quantitative balance between wild type and mutated ACVR1, we can also hypothesize a strategy able to up-regulate the wild type allele to modify the wild type/mutant ratio, then the amount of mutant receptor available to the ligand compared to wild type. Examples of strategies to increase one gene copy by targeting endogenous regulatory elements, for instance, to rescue haplo insufficiency, are reported [44,45]. However, similar approaches to increase one gene copy aimed at changing the ratio between the two alleles, could be used to increase the amount of wild type Alk2 receptor. In the above cited article by Lees Shepard et al. [12], the authors propose a model in which wild-type *ACVR1* present in cells of heterozygous FOP individuals, although unable to transduce the signal, is able to bind activin A and functions as a competitive inhibitor for binding to the mutated ACVR1 receptor, thus limiting the activin-dependent osteogenic signaling through the mutated ACVR1. In this model, the wild-type/mutant ratio may modulate the amount of pathogenic signal; then, inducing an increase of wild-type versus mutant would increase the binding of ligand to wild-type and compete for binding to the mutated receptor.

4. A General Approach to Investigate Genetic Modifiers of FOP

We are well aware that the functional effect of variants acting in *cis* with the causative gene mutation is only a limited part of the phenotypic heterogeneity, since complex interactions between genes and between genes and other functional elements in the individual genome, commonly defined as "genetic background", contribute to affecting the disease manifestation. This issue is frequently underestimated because most efforts are focused on the causative gene mutation and its functional effects, limiting the understanding of the connection between a particular disease and many functional pathways. In the case of FOP, we think it would be theoretically possible to carry out a genetic interaction analysis looking for variants in genes that have functional relationships with ACVR1 and the BMP signaling pathways, although such analysis would be made difficult by some characteristics of the disease, such as its episodic and unpredictable course, and by the multiple functional roles of BMPs in a wide variety of organs and tissues.

Animal models can be of help also to search for genetic modifiers, provided that it is possible to model the phenotypic variation and that it can be attributed to genetic variation. As an example, we cite a recently published article in which the authors aimed to identify genetic determinants that regulate the severity of the phenotype of craniosynostosis, a disease caused by activating heterozygous mutation in the *FGFR2* gene [46]. The purpose of this study was to identify potential modifiers of craniosynostosis phenotype in mice models with the disease causative mutation on two different congenic background strains. By comparing transcriptomes of cranial tissues from craniosynostosis mouse models of the two different strains, they describe potential modifiers affecting the disease severity.

Another example of this type of study is provided by the use of genetically diverse Drosophila strains, known as the Drosophila Genetic Reference Panel (DGRP), to identify components of the genetic background that affect symptoms and severity of the rare *NGLY1* deficiency [47]. As already mentioned, a Drosophila FOP model has been reported [4]; thus, in principle it would be possible to apply a similar approach to search for genetic modifiers also for the FOP phenotype.

One major limitation in the study of modifiers of FOP in humans is the lack of large families where different affected individuals with minimal difference in genetic background and showing differential phenotypic features could be analyzed. On the other hand, the great majority of FOP affected individuals carry the same *ACVR1* mutation; then, their genome can be interrogated in search for differences that could be associated to some phenotypic features. However, grouping FOP individuals in distinguishable categories on the basis of phenotype has not reached sufficient accomplishment up to now. Hopefully, the data collected by the FOP Registry will allow some reasonable grouping in the future and a

method to analyze individuals from different groups could be designed. RNA-sequencing to analyze differentially expressed genes in individuals, carrying the R206H mutation, with different phenotypic features could be a useful approach, but the choice of cells from which RNA should be extracted and the time when this extraction should be performed might be a problem to avoid differences of gene expression caused by mechanisms not related to the effect of modifier genes. Cells collected from FOP patients at the time of diagnosis before any treatment is performed might be an initial approach. Fibroblasts from skin biopsies would be very useful, but, to use less invasive procedures, cells from hair bulb or from urine might be collected and treated to carry out an appropriate analysis of transcriptome to try to identify genes and pathways relevant for FOP pathophysiology related to phenotype diversity. Coupled to RNA-sequencing, whole genome sequencing of DNA extracted from the same cells could also provide information on variants associated with individuals in defined groups, possibly variants in non coding regulatory regions of genes involved in the wide spectrum of the functional pathways that sustain the FOP pathophysiology. The type of study we are conceiving here would require the availability of the above mentioned samples collected in an appropriate bio-bank, the use of appropriate computational methods to evaluate gene expression, variants in the genome, and significant interactions among gene functions and pathways. Therefore, presenting this sort of hypothetical program, we would foresee an interdisciplinary collaboration of the different types of researchers involved in the study of FOP, clinicians, geneticists, molecular biologists, immunologists, together with computational scientists.

5. Conclusions

In this work we aimed to discuss the issue of the mechanisms underlying the extremely high prevalence of the c.617G>A mutation in FOP patients and to trace genomic studies suitable for identifying variants which can modify some features of the disease phenotype. We think that the study of genetic modifiers in FOP is worthy to be undertaken in the FOP scientific community, not only to add more basic knowledge, but also to provide cues for translational approaches paving new avenues of investigation and innovative therapeutic strategies.

6. Methods

To study how variants in *cis* might affect the FOP disease manifestation, we retrieved data related to the tissue-specific expression of the *ACVR1* gene from the Genotype-Tissue Expression portal (GTEx). The data used for the analyses described in this manuscript were obtained from the GTEx Portal, then were combined with the dbSNP database at NCBI and the UCSC Genome Browser through which the available location of CREs from the ENCODE consortium can be visualized. Finally, we used the PROMO tool at the ALLGEN resource for the study of transcription factor binding sites in the DNA sequences containing the selected variants.

Cited Websites

- Genotype-Tissue Expression (GTEx) https://gtexportal.org/home/ (accessed on 4 February 2021)
- dbSNP at the NCBI https://www.ncbi.nlm.nih.gov/snp/ (accessed on 4 February 2021)
- ENCODE consortium https://www.genome.gov/Funded-Programs-Projects/ENCODE-Project-ENCyclopedia-Of-DNA-Elements (accessed on 4 February 2021)
- ENCODE at the UCSC Genome Browser https://genome.ucsc.edu/ENCODE/ (accessed on 4 February 2021)
- PROMO 3.0 http://alggen.lsi.upc.edu (accessed on 4 February 2021)

References

1. Cappato, S.; Gamberale, R.; Bocciardi, R.; Brunelli, S. Genetic and Acquired Heterotopic Ossification: A Translational Tale of Mice and Men. *Biomedicines* **2020**, *8*, 611. [CrossRef] [PubMed]
2. Yu, P.B.; Hong, C.C.; Sachidanandan, C.; Babitt, J.L.; Deng, D.Y.; Hoyng, S.A.; Lin, H.Y.; Bloch, K.D.; Peterson, R.T. Dorsomorphin inhibits BMP signals required for embryogenesis and iron metabolism. *Nat. Chem. Biol.* **2008**, *4*, 33–41. [CrossRef] [PubMed]
3. Allen, R.S.; Tajer, B.; Shore, E.M.; Mullins, M.C. Fibrodysplasia ossificans progressiva mutant ACVR1 signals by multiple modalities in the developing zebrafish. *Elife* **2020**, *9*, e53761. [CrossRef] [PubMed]
4. Le, V.; Anderson, E.; Akiyama, T.; Wharton, K.A. Drosophila models of FOP provide mechanistic insight. *Bone* **2018**, *109*, 192–200. [CrossRef]
5. Hino, K.; Ikeya, M.; Horigome, K.; Matsumoto, Y.; Ebise, H.; Nishio, M.; Sekiguchi, K.; Shibata, M.; Nagata, S.; Matsuda, S.; et al. Neofunction of ACVR1 in fibrodysplasia ossificans progressiva. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 15438–15443. [CrossRef] [PubMed]
6. Hatsell, S.J.; Idone, V.; Wolken, D.M.; Huang, L.; Kim, H.J.; Wang, L.; Wen, X.; Nannuru, K.C.; Jimenez, J.; Xie, L.; et al. ACVR1R206H receptor mutation causes fibrodysplasia ossificans progressiva by imparting responsiveness to activin A. *Sci. Transl. Med.* **2015**, *7*, 303ra137. [CrossRef] [PubMed]
7. Hino, K.; Horigome, K.; Nishio, M.; Komura, S.; Nagata, S.; Zhao, C.; Jin, Y.; Kawakami, K.; Yamada, Y.; Ohta, A.; et al. Activin-A enhances mTOR signaling to promote aberrant chondrogenesis in fibrodysplasia ossificans progressiva. *J. Clin. Investig.* **2017**, *127*, 3339–3352. [CrossRef]
8. Nakajima, T.; Ikeya, M. Insights into the biology of fibrodysplasia ossificans progressiva using patient-derived induced pluripotent stem cells. *Regen. Ther.* **2019**, *11*, 25–30. [CrossRef]
9. Medici, D.; Shore, E.M.; Lounev, V.Y.; Kaplan, F.S.; Kalluri, R.; Olsen, B.R. Conversion of vascular endothelial cells into multipotent stem-like cells. *Nat. Med.* **2010**, *16*, 1400–1406. [CrossRef]
10. Pignolo, R.J.; Shore, E.M. Circulating osteogenic precursor cells. *Crit. Rev. Eukaryot Gene Expr.* **2010**, *20*, 171–180. [CrossRef]
11. Dey, D.; Bagarova, J.; Hatsell, S.J.; Armstrong, K.A.; Huang, L.; Ermann, J.; Vonner, A.J.; Shen, Y.; Mohedas, A.H.; Lee, A.; et al. Two tissue-resident progenitor lineages drive distinct phenotypes of heterotopic ossification. *Sci. Transl. Med.* **2016**, *8*, 366ra163. [CrossRef]
12. Lees-Shepard, J.B.; Yamamoto, M.; Biswas, A.A.; Stoessel, S.J.; Nicholas, S.E.; Cogswell, C.A.; Devarakonda, P.M.; Schneider, M.J., Jr.; Cummins, S.M.; Legendre, N.P.; et al. Activin-dependent signaling in fibro/adipogenic progenitors causes fibrodysplasia ossificans progressiva. *Nat. Commun.* **2018**, *9*, 471–485. [CrossRef]
13. Botman, E.; Teunissen, B.P.; Raijmakers, P.; de Graaf, P.; Yaqub, M.; Treurniet, S.; Schoenmaker, T.; Bravenboer, N.; Micha, D.; Pals, G.; et al. Diagnostic Value of Magnetic Resonance Imaging in Fibrodysplasia Ossificans Progressiva. *JBMR Plus* **2020**, *4*, e10363. [CrossRef]
14. Shore, E.M.; Kaplan, F.S. Inherited human diseases of heterotopic bone formation. *Nat. Rev. Rheumatol.* **2010**, *6*, 518–527. [CrossRef]
15. Barruet, E.; Morales, B.M.; Cain, C.J.; Ton, A.N.; Wentworth, K.L.; Chan, T.V.; Moody, T.A.; Haks, M.C.; Ottenhoff, T.H.; Hellman, J.; et al. NF-κB/MAPK activation underlies ACVR1-mediated inflammation in human heterotopic ossification. *JCI Insight* **2018**, *3*, e122958. [CrossRef]
16. Wang, H.; Lindborg, C.; Lounev, V.; Kim, J.H.; McCarrick-Walmsley, R.; Xu, M.; Mangiavini, L.; Groppe, J.C.; Shore, E.M.; Schipani, E.; et al. Cellular Hypoxia Promotes Heterotopic Ossification by Amplifying BMP Signaling. *J. Bone Miner. Res.* **2016**, *31*, 1652–1665. [CrossRef]
17. Haupt, J.; Stanley, A.; McLeod, C.M.; Cosgrove, B.D.; Culbert, A.L.; Wang, L.; Mourkioti, F.; Mauck, R.L.; Shore, E.M. ACVR1(R206H) FOP mutation alters mechanosensing and tissue stiffness during heterotopic ossification. *Mol. Biol. Cell* **2019**, *30*, 17–29. [CrossRef]
18. Stanley, A.; Heo, S.J.; Mauck, R.L.; Mourkioti, F.; Shore, E.M. Elevated BMP and Mechanical Signaling Through YAP1/RhoA Poises FOP Mesenchymal Progenitors for Osteogenesis. *J. Bone Miner. Res.* **2019**, *34*, 1894–1909. [CrossRef]

19. 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. [CrossRef]

20. Pacifici, M.; Shore, E.M. Common mutations in ALK2/ACVR1, a multi-faceted receptor, have roles in distinct pediatric musculoskeletal and neural orphan disorders. *Cytokine Growth Factor Rev.* **2016**, *27*, 93–104. [CrossRef]

21. Katagiri, T.; Tsukamoto, S.; Nakachi, Y.; Kuratani, M. Recent Topics in Fibrodysplasia Ossificans Progressiva. *Endocrinol. Metab.* **2018**, *33*, 331–338. [CrossRef] [PubMed]

22. Valer, J.A.; Sánchez-de-Diego, C.; Pimenta-Lopes, C.; Rosa, J.L.; Ventura, F. ACVR1 Function in Health and Disease. *Cells* **2019**, *8*, 1366. [CrossRef] [PubMed]

23. Janoff, H.B.; Muenke, M.; Johnson, L.O.; Rosenberg, A.; Shore, E.M.; Okereke, E.; Zasloff, M.; Kaplan, F.S. Fibrodysplasia ossificans progressiva in two half-sisters: Evidence for maternal mosaicism. *Am. J. Med. Genet.* **1996**, *61*, 320–324. [CrossRef]

24. Youk, J.; An, Y.; Park, S.; Lee, J.K.; Ju, Y.S. The genome-wide landscape of C:G > T:A polymorphism at the CpG contexts in the human population. *BMC Genom.* **2020**, *21*, 270–281. [CrossRef]

25. Casal, M.L.; Engiles, J.B.; Zakošek Pipan, M.; Berkowitz, A.; Porat-Mosenco, Y.; Mai, W.; Wurzburg, K.; Xu, M.Q.; Allen, R.; ODonnell, P.A.; et al. Identification of the Identical Human Mutation in ACVR1 in 2 Cats With Fibrodysplasia Ossificans Progressiva. *Vet. Pathol.* **2019**, *56*, 614–618. [CrossRef]

26. Buczkowicz, P.; Hawkins, C. Pathology, Molecular Genetics, and Epigenetics of Diffuse Intrinsic Pontine Glioma. *Front. Oncol.* **2015**, *5*, 147–155. [CrossRef]

27. Han, H.J.; Jain, P.; Resnick, A.C. Shared ACVR1 mutations in FOP and DIPG: Opportunities and challenges in extending biological and clinical implications across rare diseases. *Bone* **2018**, *109*, 91–100. [CrossRef]

28. Goldmann, J.M.; Veltman, J.A.; Gilissen, C. De Novo Mutations Reflect Development and Aging of the Human Germline. *Trends Genet.* **2019**, *35*, 828–839. [CrossRef]

29. Rogers, J.G.; Chase, G.A. Paternal age effect in fibrodysplasia ossificans progressiva. *J. Med. Genet.* **1979**, *16*, 147–148. [CrossRef] [PubMed]

30. Jónsson, H.; Sulem, P.; Kehr, B.; Kristmundsdottir, S.; Zink, F.; Hjartarson, E.; Hardarson, M.T.; Hjorleifsson, K.E.; Eggertsson, H.P.; Gudjonsson, S.A.; et al. Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature* **2017**, *549*, 519–522. [CrossRef] [PubMed]

31. Goldmann, J.M.; Seplyarskiy, V.B.; Wong, W.S.W.; Vilboux, T.; Neerincx, P.B.; Bodian, D.L.; Solomon, B.D.; Veltman, J.A.; Deeken, J.F.; Gilissen, C.; et al. Germline de novo mutation clusters arise during oocyte aging in genomic regions with high double-strand-break incidence. *Nat. Genet.* **2018**, *50*, 487–492. [CrossRef]

32. Gao, Z.; Moorjani, P.; Sasani, T.A.; Pedersen, B.S.; Quinlan, A.R.; Jorde, L.B.; Amster, G.; Przeworski, M. Overlooked roles of DNA damage and maternal age in generating human germline mutations. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 9491–9500. [CrossRef]

33. Acuna-Hidalgo, R.; Veltman, J.A.; Hoischen, A. New insights into the generation and role of de novo mutations in health and disease. *Genome Biol.* **2016**, *17*, 241–260. [CrossRef] [PubMed]

34. de Sousa Lopes, S.M.; Roelen, B.A.; Monteiro, R.M.; Emmens, R.; Lin, H.Y.; Li, E.; Lawson, K.A.; Mummery, C.L. BMP signaling mediated by ALK2 in the visceral endoderm is necessary for the generation of primordial germ cells in the mouse embryo. *Genes Dev.* **2004**, *18*, 1838–1849. [CrossRef] [PubMed]

35. Bocciardi, R.; Bordo, D.; Di Duca, M.; Di Rocco, M.; Ravazzolo, R. Mutational analysis of the ACVR1 gene in Italian patients affected with fibrodysplasia ossificans progressiva: Confirmations and advancements. *Eur. J. Hum. Genet.* **2009**, *17*, 311–318. [CrossRef] [PubMed]

36. Matsuoka, M.; Tsukamoto, S.; Orihara, Y.; Kawamura, R.; Kuratani, M.; Haga, N.; Ikebuchi, K.; Katagiri, T. Design of primers for direct sequencing of nine coding exons in the human ACVR1 gene. *Bone* **2020**, *138*, 115469. [CrossRef] [PubMed]

37. Claussnitzer, M.; Dankel, S.N.; Kim, K.H.; Quon, G.; Meuleman, W.; Haugen, C.; Glunk, V.; Sousa, I.S.; Beaudry, J.L.; Puviindran, V.; et al. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N. Engl. J. Med.* **2015**, *373*, 895–907. [CrossRef]

38. The GTEx Consortium. *The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science* **2020**, *369*, 1318–1330. [CrossRef]

39. Gamazon, E.R.; Segrè, A.V.; van de Bunt, M.; Wen, X.; Xi, H.S.; Hormozdiari, F.; Ongen, H.; Konkashbaev, A.; Derks, E.M.; Aguet, F.; et al. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat. Genet.* **2018**, *50*, 956–967. [CrossRef]

40. Castel, S.E.; Aguet, F.; Mohammadi, P.; Ardlie, K.G.; Lappalainen, T. A vast resource of allelic expression data spanning human tissues. *Genome Biol.* **2020**, *21*, 234–246. [CrossRef]

41. Pignolo, R.J.; Baujat, G.; Brown, M.A.; De Cunto, C.; Di Rocco, M.; Hsiao, E.C.; Keen, R.; Al Mukaddam, M.; Sang, K.L.Q.; Wilson, A.; et al. Natural history of fibrodysplasia ossificans progressiva: Cross-sectional analysis of annotated baseline phenotypes. *Orphanet J. Rare Dis.* **2019**, *14*, 98–109. [CrossRef] [PubMed]

42. Peng, K.; Cheung, K.; Lee, A.; Sieberg, C.; Borsook, D.; Upadhyay, J. Longitudinal Evaluation of Pain, Flare-Up, and Emotional Health in Fibrodysplasia Ossificans Progressiva: Analyses of the International FOP Registry. *JBMR Plus* **2019**, *3*, e10181. [CrossRef] [PubMed]

43. Pignolo, R.J.; Cheung, K.; Kile, S.; Fitzpatrick, M.A.; De Cunto, C.; Al Mukaddam, M.; Hsiao, E.C.; Bauja, G.; Delai, P.; Eekhoff, E.M.W.; et al. Self-reported baseline phenotypes from the International Fibrodysplasia Ossificans Progressiva (FOP) Association Global Registry. *Bone* **2020**, *134*, 115274. [CrossRef] [PubMed]

44. Matharu, N.; Ahituv, N. Modulating gene regulation to treat genetic disorders. *Nat. Rev. Drug Discov.* **2020**, *19*, 757–775. [CrossRef]
45. Matharu, N.; Rattanasopha, S.; Tamura, S.; Maliskova, L.; Wang, Y.; Bernard, A.; Hardin, A.; Eckalbar, W.L.; Vaisse, C.; Ahituv, N. CRISPR-mediated activation of a promoter or enhancer rescues obesity caused by haploinsufficiency. *Science* **2019**, *363*, eaau0629. [CrossRef]
46. Dudakovic, A.; Nam, H.K.; Wijnen, A.J.V.; Hatch, N.E. Genetic background dependent modifiers of craniosynostosis severity. *J. Struct. Biol.* **2020**, *212*, 107629. [CrossRef]
47. Talsness, D.M.; Owings, K.G.; Coelho, E.; Mercenne, G.; Pleinis, J.M.; Partha, R.; Hope, K.A.; Zuberi, A.R.; Clark, N.L.; Lutz, C.M.; et al. A Drosophila screen identifies NKCC1 as a modifier of NGLY1 deficiency. *Elife* **2020**, *9*, e57831. [CrossRef]