Supplementary Notes

Characteristic clinical features of FOP.

FOP is the most severe and disabling disorder of extra-skeletal (heterotopic) ossification in humans.[1] Heterotopic ossification in FOP begins in childhood and can be induced by trauma, or may occur without warning. Bone formation is episodic, progressive, and extensive, leading to the extra-articular ankylosis of all the major joints of the axial and appendicular skeleton, rendering movement impossible (Fig. 1a). Flareups of FOP arise and progress in a well-defined spatial pattern that result in ribbons, sheets, and plates of bone that fuse the joints of the axial and appendicular skeleton, entombing the patient in a "second skeleton" of heterotopic bone. One of the more readily recognized skeletal malformations in FOP patients are great toe malformations of metatarsals and proximal phalanges that can occur along with microdactyly, fused interphalangeal joints, and hallux valgus deviations at the metatarsophalangeal joints (Fig. 1b). The severe disability of FOP results in low reproductive fitness and few examples of inheritance of FOP are known. Death often results by the fifth decade from thoracic insufficiency syndrome.[2] There is no effective prevention or treatment.[3]

Descriptions of FOP families.

Our initial linkage analysis[4] used four families that showed autosomal dominant inheritance of FOP-type heterotopic ossification, although not all affected individuals in each pedigree had characteristic malformation of the great toes. With the experience of examining more patients over time, we were concerned that patients without both of these features could confound linkage

analysis due to locus heterogeneity or mosaicism. This clinical information prompted the decision to use only a subset of families in whom all affected individuals had unambiguous features of malformed toes and progressive heterotopic ossification (Fig. 1a, b) in our present linkage analysis (Fig. 1c). A combined multipoint lod plot for the markers in the chromosome 2 FOP linkage region is shown in **Supplementary Figure 1** online.

Investigation of the *ACVR1* c.617G>A (R206H) mutation in the five families used in the current linkage analysis show that all affected members have the mutation and none of the unaffected members available for examination carry the mutation. Of the four families used in the initial linkage study,[4] Family 1 had unambiguous features of FOP in all affected individuals and was used in the current study (Family A in Fig. 1c). (Family numbers are those used in Feldman et al. 2000; letters are used to identify families in the current study.) Family 2 showed ambiguous FOP features, with one member possessing only toe malformations without heterotopic ossification, while another had no toe malformation and mild heterotopic ossification that has not progressed.[5] This family was not used in the current linkage analysis since every member did not fulfill the most stringent diagnostic criteria for FOP. No *ACVR1* c.617G>A mutation was detected in any member of this family. In Family 3, FOP was inherited from mother to children and all had classic features of FOP (M. LeMerrer, personal communication). However, this family was not available for confirmational re-examination by us and was thus excluded from the current linkage analysis. Subsequent evaluation of Family 3 with chromosome 2 markers confirmed linkage to the FOP locus and all affected members of this family contain the *ACVR1* c.617G>A mutation. Family 4 had two affected members, one with classic features of FOP (daughter), while the other (father) showed only mild evidence of heterotopic ossification with no toe

malformation. The daughter was heterozygous for *ACVR1* c.617G>A, while the father does not carry a germline mutation. We suspect that the father may have been a somatic/germline mosaic for the mutation, however he is deceased and this possibility cannot be examined further.

An additional very recently identified family consists of a father with FOP (deceased, unavailable for analysis), an unaffected mother, and two affected children with classic FOP features. This family shows linkage to the chromosome 2 FOP locus and both children are heterozygous for *ACVR1* c.617G>A on the paternally inherited allele.

c.617G>A (R206H) mutations in the *ACVR1* gene in patients with FOP.

The chromosome 2q FOP critical genomic region (Fig. 2a) spans ~23.9 Mb between markers rs1020088 (centromeric) at 150,654,341 bp and D2S1238 (telomeric) at 174,505,230 bp as annotated by UCSC GenomeBrowser. The *ACVR1* gene spans ~138.6 kb of genomic DNA (chromosome 2: 158,418,469-158,557,131). *ACVR1* encodes a 509 amino acid protein that contains a ligand binding region, a transmembrane (TM) domain, a glycine-serine (GS) rich domain, and a protein kinase domain. The numbers above the protein representation in Figure 2a indicate the amino acids included in each identified domain. The schematics in this figure are drawn approximately to scale.

***ACVR1* gene structure.**

We obtained the intron-exon boundaries of the *ACVR1* gene through GenBank, Ensembl Human Genome Server, and the University of Santa Cruz. Transcript and exon information was obtained from Ensembl (Gene ID ENSG00000115170; transcript ID ENST00000263640) which reports

11 exons for *ACVR1* (exons 1 and 2 contain only 5'UTR; the protein start site is in exon 3), consistent with GenBank BC033867, full length cDNA clone. All databases for *ACVR1* are consistent for the sequence information for the nine exons containing protein-coding sequences, however, additional/alternate exons containing 5'UTRs are reported through the University of Santa Cruz Genome Browser (12 exons with the protein start in exon 4) and GenBank (10 exons with the protein start in exon 2). The R206H mutation occurs in nucleotide 617 of *ACVR1* cDNA (c.617G>A). [Notation follows standard nomenclature guidelines.[6]]

Protein structure predictions.

Structural protein homology modeling was used to determine possible biochemical consequences of the ACVR1 R206H mutation. While SWISS-MODEL analysis showed no deviation between proteins containing Arg206 or His206, both PredictProtein and CPHmodels predict a partial destabilization of the α-helix formed by ACVR1 amino acids 198-206 (Fig. 2e). These models reveal that Arg202 and Arg206 are spatially orientated on the same helical face (i, i+4). Previous studies have demonstrated that the electrostatic effects of charged ion pairs can have significant helix stabilizing interactions between side chains when the spacing between residues is close to the helical repeat of 3.6 residues per turn (i.e. i, i+4).[7,8] Additionally, polar side chains are often long, thus allowing their hydrophobic alkyl groups to interact favorably with nonpolar residues while keeping the polar parts free to interact with other polar groups. Side chains such as lysine and arginine can thus interact favorably with both polar and non-polar residues.[9] Therefore, the shorter side chain of the R206H mutant is expected to cause a partial destabilization of the α-helix altering the electrostatic potential of the ACVR1 protein (Fig. 2e).

Additionally the R206H mutation may impair protein-protein interactions with the GS domain. This 30 residue motif of the type I TGFβ receptor (TβR-I) kinase[10] has two regulatory functions: (1) tight control over the basal state with FKBP12 binding to the unphosphorylated GS domain and creating a inhibitory wedge that prevents interactions with other proteins and, (2) a catalytically "open" form that binds ATP leading to protein-protein interactions with the Smad2 MH2 domains.[10,11] Arginine-arginine pairs within a protein can stabilize complex formation between proteins or can stabilize regions of backbone structure through intramolecular interactions.[12]

The effect of the R206H mutation on the predicted protein structure of the ACVR1 α-helix, residues 198-206 is shown in Figure 2e. The homology model of wild-type ACVR1 shows that the most likely conformation (lowest scoring rotamer) of the arginine 206 side chain predicts that it interacts with the α-helix backbone to stabilize the protein. Homology model of mutant R206H ACVR1 shows that the most likely conformation (lowest scoring rotamer) of the histidine 206 side chain does not interact with the α-helix backbone which is predicted to result in partial destabilization of the protein.

The Arg > His amino acid change in codon 206 appears conservative in that one positively charged amino acid is substituted for another. (In fact, in human BMPRIA and BMPRIB, codon 206 is a lysine; **Supplementary Figure 2** online.) However, protein modeling predicts that the shorter histidine side chain will nevertheless alter protein structure and/or protein-protein interactions. Furthermore, we speculate that a non-conservative (non-positively charged) amino acid change in codon 206 may result in a lethal mutation.

Constitutively activating mutations in the GS domain of type I TGFβ receptors.

Type I TGFβ/BMP receptors contain a highly conserved 30 amino acid GS domain that is phosphorylated by ligand-bound type II receptors.[13] Amino acid substitutions in the GS domain (T204D) have been shown to lead constitutively activating forms of TβR-I (TGF-β type I receptor).[14] (Codon 204 in TβR-I is analogous to codon 207 in ACVR1.)

Recurrent mutations in human disease.

The FOP R206H ACVR1 mutation is one of the most specific codons in the human genome to be associated with a disease phenotype. We note two additional disorders that are mainly associated with mutations in a single codon.

Achondroplasia (Ach; OMIM 100800) has been ascribed mainly to mutations in codon 380 of the *FGFR3* gene. Gly380Arg (c.1138G>A) is the most common achondroplasia mutation, although glycine to cysteine substitution at the same nucleotide position (Gly380Arg; c.1138G>C) also occurs. A single rare case of achondroplasia that was associated with early lethality carried a Gly375Cys mutation. The dominant activating Gly308Arg mutation in the transmembrane domain of fibroblast growth factor receptor 3 (*FGFR3*; OMIM 134934) disrupts FGFR3 trafficking and results in accelerated chondrocyte maturation.[15] In both achondroplasia and FOP, the identified mutations are CpG dinucleotide changes, show autosomal dominant inheritance, and are fully penetrant.

Timothy Syndrome (TS; OMIM 601005) is another human disorder that is mainly caused by mutations in a single codon. Gly406Arg (c.1216G>A) mutations in the *CACNA1C* gene are

found in most patients, although two individuals with severe forms of TS were found to carry Gly402Ser (c.1204G>A) mutations.

References

1. Kaplan, F.S., Shore, E.M. & Connor, J.M. Fibrodysplasia Ossificans Progressiva. in *Connective Tissue and Its Heritable Disorders* (eds. Royce, P. & Steinmann, B.) 827-840 (Wiley-Liss, New York, 2002).
2. Kaplan, F.S. & Glaser, D.L. Thoracic insufficiency syndrome in patients with fibrodysplasia ossificans progressiva. *Clin. Rev. Bone Min. Metab.* **3**, 213-216 (2005).
3. Glaser, D.L. & Kaplan, F.S. Treatment considerations for the management of fibrodysplasia ossificans progressiva. *Clin. Rev. Bone Min. Metab.* **3**, 243-250 (2005).
4. Feldman, G. et al. Fibrodysplasia ossificans progressiva, a heritable disorder of severe heterotopic ossification, maps to human chromosome 4q27-31. *Am. J. Hum. Genet.* **66**, 128-135 (2000).
5. Connor, J.M., Skirton, H. & Lunt, P.W. A three generation family with fibrodysplasia ossificans progressiva. *J. Med. Genet.* **30**, 687-689 (1993).
6. den Dunnen, J.T. & Antonarakis, S.E. Nomenclature for the description of human sequence variations. *Hum. Genet.* **109**, 121-124 (2001).
7. Huyghues-Despointes, B.M., Klingler, T.M. & Baldwin, R.L. Measuring the strength of side-chain hydrogen bonds in peptide helices: the Gln.Asp (i, i + 4) interaction. *Biochemistry* **34**, 13267-13271 (1995).
8. Scholtz, J.M., Qian, H., Robbins, V.H. & Baldwin, R.L. The energetics of ion-pair and hydrogen-bonding interactions in a helical peptide. *Biochemistry* **32**, 9668-9676 (1993).
9. Andrew, C.D., Penel, S., Jones, G.R. & Doig, A.J. Stabilizing nonpolar/polar side-chain interactions in the alpha-helix. *Proteins: Structure, Function and Genetics* **45**, 449-455 (2001).
10. Huse, M. et al. The TGF beta receptor activation process: an inhibitor-to substrate-binding switch. *Molecular Cell* **8**, 671-682 (2001).
11. Huse, M., Chen, Y.G., Massagué, J. & Kuriyan, J. Crystal structure of the cytoplasmic domain of the type I TGF beta receptor in complex with FKBP12. *Cell* **96**, 425-436 (1999).
12. Magalhaes, A., Maigret, B., Hoflack, J., Gomes, J.N. & Scheraga, H.A. Contribution of unusual arginine-arginine short-range interactions to stabilization and recognition in proteins. *Journal of Protein Chemistry* **13**, 195-215 (1994).
13. Wrana, J.L., Attisano, L., Wieser, R., Ventura, F. & Massagué, J. Mechanism of activation of the TGF-beta receptor. *Nature* **370**, 341-347 (1994).
14. Wieser, R., Wrana, J.L. & Massagué, J. GS domain mutations that constitutively activate TβR-I, the downstream signaling component in the TGF-β receptor complex. *EMBO J.* **14**, 2199-2208 (1995).
15. Monsonego-Ornan, E., Adar, R., Feferman, T., Segev, O. & Yayon, A. The transmembrane mutation G380R in fibroblast growth factor receptor 3 uncouples ligand-mediated receptor activation from down-regulation. *Mol. Cell. Biol.* **20**, 516-522 (2000).