

The American Journal of Human Genetics, Volume 93

Supplemental Data

Mutations in *FAM111B* Cause Hereditary

Fibrosing Poikiloderma with Tendon Contracture,

Myopathy, and Pulmonary Fibrosis

Sandra Mercier, Sébastien Küry, Gasnat Shaboodien, Darren T. Houniet, Nonhlanhla P. Khumalo, Chantal Bou-Hanna, Nathalie Bodak, Valérie Cormier-Daire, Albert David, Laurence Faivre, Dominique Figarella-Branger, Romain K. Gherardi, Elise Glen, Antoine Hamel, Christian Laboisse, Cédric Le Caignec, Pierre Lindenbaum, Armelle Magot, Arnold Munnich, Jean-Marie Mussini, Komala Pillay, Thahira Rahman, Richard Redon, Emmanuelle Salort-Campana, Mauro Santibanez-Koref, Christel Thauvin, Sébastien Barbarot, Bernard Keavney, Stéphane Bézieau, and Bongani M. Mayosi

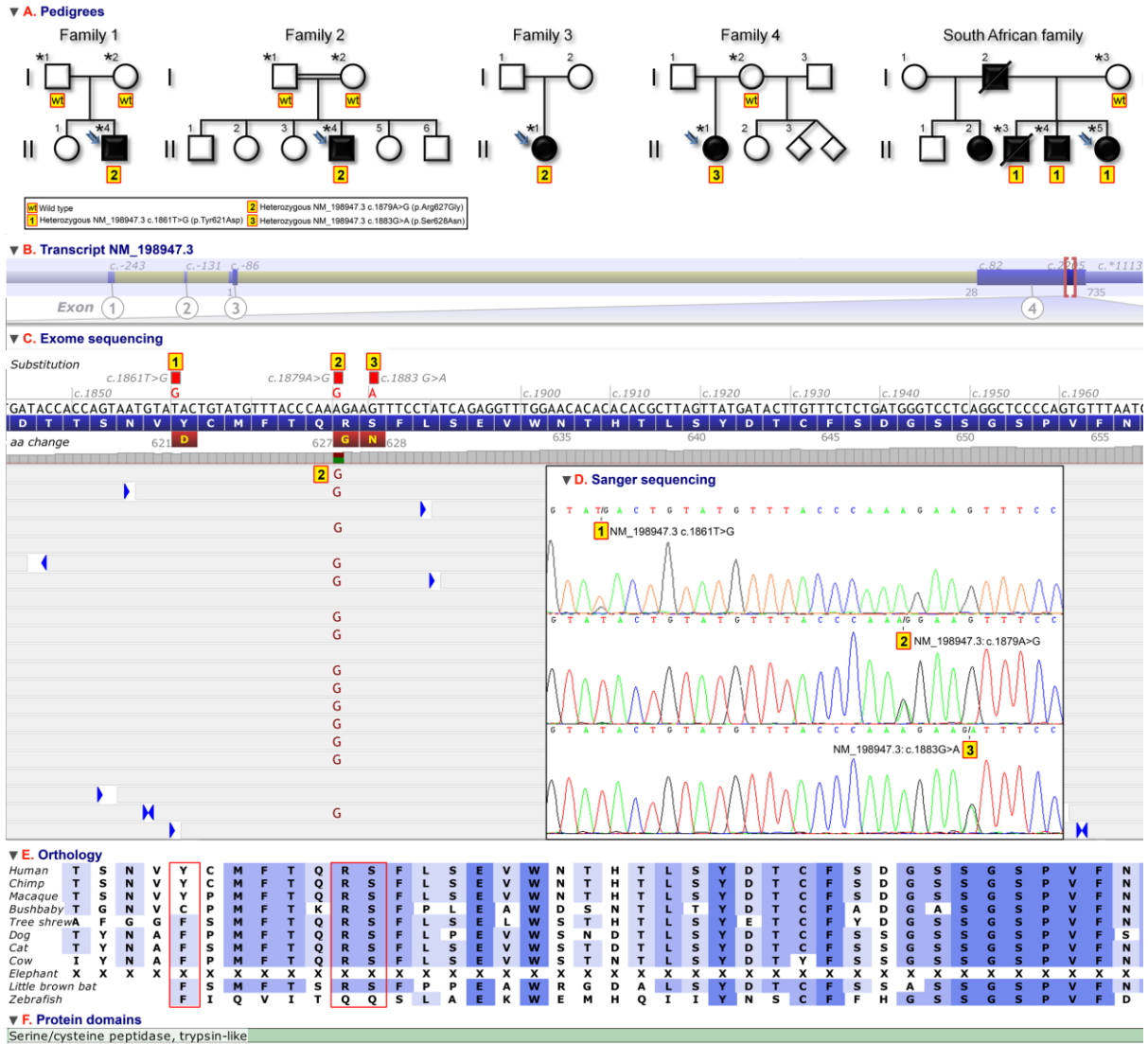


Figure S1. Genetic study. Part A shows the pedigrees related to the four affected individuals F1-II2, F2-II4, F3-II1 and F4-II1 described in the present report and to the three additional affected members FSA-II3, FSAII-4 and FSA-II5 from the South African family previously reported.⁴ Affected individuals are represented by solid square or circle, depending upon the gender, and index cases are indicated by arrows. Of note, the twins in family 4 are not born yet. Individuals tested by exome and/or Sanger sequencing are marked by asterisks and their genotypes are noted below in yellow rectangles: wt indicates a wild-type genotype, whereas the three different mutations observed in *FAM111B* (NM_198947.3:c.1861T>G (p.Tyr621Asp), c.1879A>G (p.Arg627Gly), and c.1861T>G (p.Tyr621Asp), are numbered 1 to 3. These mutations, located in a small region of exon 4 (within red brackets in part B), were found first by whole-exome sequencing (C), and confirmed by Sanger sequencing (D). The amino acids affected by substitutions 2 and 3 are well conserved across mammalian species, whereas the one altered by substitution 1 appeared conserved across primates only (E). All mutations are within a putative serine/cysteine peptidase, trypsin-like domain (F). Parts B, C, E and F were adapted from a screenshot of the *FAM111B* visualized with Alamut software (Interactive Biosoftware, Rouen, France).

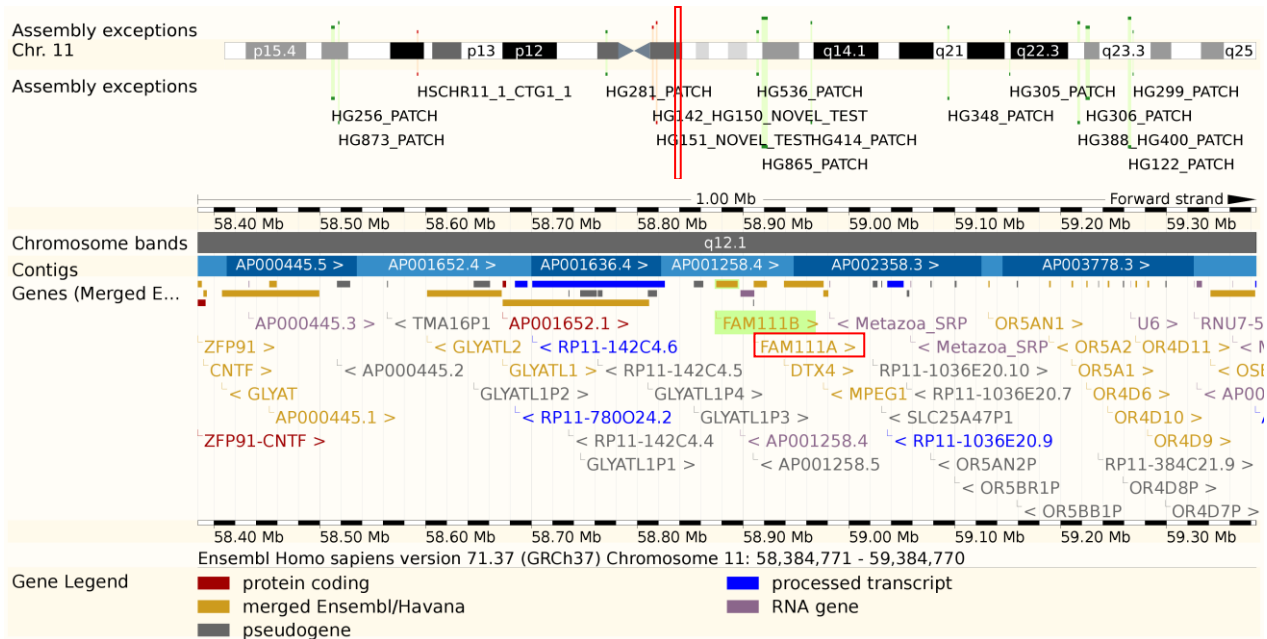


Figure S2. Chromosomal location of *FAM111B* (screenshot from Ensembl Genome Server; URL: <http://www.ensembl.org/index.html>). *FAM111B* (ENSG00000189057) belongs to a family of two members. It is located on the long arm of chromosome 11 at position 11q12.1, beside the other member of the family with sequence similarity 111, *FAM111A*.

ENSP00000341565

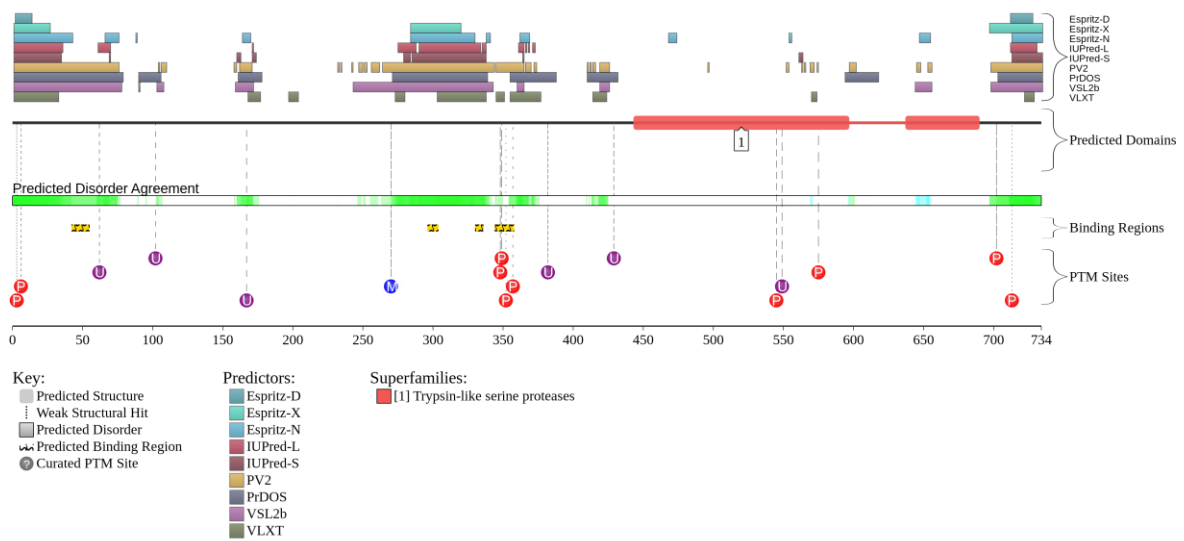


Figure S3. Relationship between structure disorder inferred for FAM111B from the Database of Disordered Protein Predictions (D²P²; URL: <http://d2p2.pro/>). According to D²P², the trypsin-like (cysteine/serine) protease domain of FAM111B (red rectangles) encompasses amino acids 423 to 717 and is split into two parts separated by a loop (red line). The three amino acid changes p.Tyr621Asp, p.Arg627Gly and p.Ser628Asn, observed in the patients reported in the study are precisely located in this small region, which is not predicted to be disordered. No significant homology with any known protein was found by Blast searches focused on this loop, the function of which remains unclear.

Score = 249 bits (636), Expect = 2e-85, Method: Compositional matrix adjust.

Identities = 124/275 (45%), Positives = 170/275 (62%), Gaps = 5/275 (2%)

```
FAM111B 431 FNIYKKDFGKMTANSVSVATCEQLTYYSKSVGFMQWDDNNGTGNATCFVFNNGYIFTCRH 490
F +++ FGK+T NS S+ + L S SVG++ WD+ TG ATCFVF G +I TCRH
FAM111A 327 FELHRTTTFGKVTKNS SSIKVVKLLVRLSDSVGYLEFWD-ATTGYATCFVFKGLFILTCRH 385

FAM111B 491 VVHLMVGNKTHPSLWPDIIISKCAKVTFTTYTEFCPTPDNWFSEIPEWLKVSNEENLDYAILKL 550
V+ +VG PS W II +C +VTF Y E N+F +EPW ++ NE LDYA+LKL
FAM111A 386 VIDSIVGDGIEPSKWATIIGQCVRVTFGYEELKDKETNYFFVEPWFEIHNEELDYAVLKL 513

FAM111B 551 KENGNAPPGLWRQISPPSTGLIYLIGHPEGQIKKIDGCTVIPLNERLKKYPNDCQDGL 610
KENG P L+ I+P P +GLI++IGHP G+ K+ID C VIP +R KK CQ+ +
FAM111A 514 KENGQQVPMELYNGITPVPLSGLIHIIGHYPYGEKKQIDACAVIPQGQRAKK----CQERV 501

FAM111B 611 VDLYDTSNVYCMFTQRSFLSEVWNTHTLSYDTCFSDGSSGSPVFNASGKLVALHTFGLF 670
+ M+TQRSF V N ++YDT F G+SGSPV++ G LVA+H G
FAM111A 502 QSKKAESPEVHMYTQRSFQKIVHNEDVITYDTEFFFGASGSPVFDKSGSLVAMHAAGFA 561

FAM111B 671 YQRFNVHALIEFGYSMDSILCDIKKTNESLYKSL 705
Y ++IEFG +M+SIL DIK+ ++ Y+ +
FAM111A 562 YTYQNETSIIIEFGSTMESILLDIKQRHKPWYEEV 596
```

Y = Amino acid of FAM111B affected by the South African HFP substitution

R and S = Amino acids of FAM111B affected by the HFP mutations observed in Patients 1 to 4

S, Y, P, D and R = Amino acids of FAM111A affected by KCS/OCS mutations

Figure S4. BLAST 2 sequences between FAM111A and FAM111B focused on the peptidase cysteine/serine, trypsin-like domain (URL: <http://blast.ncbi.nlm.nih.gov/>). The figure indicates that the catalytic domain predicted for FAM111B is shared by FAM111A with a homology of protein sequence about 45%. Amino acid changes of FAM111B involved in hereditary fibrosing poikiloderma (HFP) and in Kenny-Caffey syndrome/osteocarniostenosis (OCS)¹, are highlighted in yellow and blue, respectively.

>gi_39573718_ref_NP_945185.1_protein FAM111B isoform a [Homo sapiens]

MNSMKTEENKSF**S**A**M**E**D**DQ**R**T**R**P**E**V**S**K**D**T**V**M**K**Q**T**H**A**D**T**P**V**D**H**C**L**S**G**I**R**K**C**
S**S**T**F**K**L**K**S**E**V**N**K**H**E**T**A**L**E**M**Q**N**P**N**L**N**N**K**E**C**C**F**T**F**T**L**N**G**N**S**R**K**L**D**R**S**V**F**T**A**Y**
G**K**P**S**E**S**I**Y**S**A**L**S**A**N**D**Y**F**S**E**R**I**K**N**Q**F**N**K**N**I**I**V**Y**E**E**K**T**I**D**G**H**I**N**L**G**M**P**L**K**C**L**
P**S**D**S**H**F**K**I**T**F**G**Q**R**K**S**S**K**E**D**G**H**I**L**R**Q**C**E**N**P**N**M**E**C**I**L**F**H**V**V**A**I**G**R**T**R**K**K**I**V**K**
I**N**E**L**H**E**K**G**S**K**L**C**I**Y**A**L**K**G**E**T**I**E**G**A**L**C**K**D**G**R**F**R**S**D**I**G**E**F**E**W**K**L**K**E**G**H**K**K**I**Y**
G**K**Q**S**M**V**D**E**V**S**G**K**V**L**E**M**D**I**S**K**K**K**A**L**Q**Q**K**D**I**H**K**K**I**K**Q**N**E**S**A**T**D**E**I**N**H**Q**S**L**I**Q**
S**K**K**K**V**H**K**P**K**K**D**G**E**T**K**D**V**E**H**S**R**E**Q**I**L**P**P**Q**D**L**S**H**Y**I**K**D**K**T**R**Q**T**I**P**R**I**R**N**Y**Y**F**
C**S**L**P**R**K**Y**R**Q**I**N**S**Q**V**R**R**R**P**H**L**G**R**R**Y**A**I**N**L**D**V**Q**K**E**A**I**N**L**L**K**N**Y**Q**T**L**N**E**A**I**M**H**
Q**Y**P**N**F**K**E**E**A**Q**W**R**K**Y**F**R**E**E**Q**K**R**M**N**L**S**P**A**K**Q**F**N**I**Y**K**K**D**F**G**K**M**T**A**N**S**V**S**V**A**T
C**E**Q**L**T**Y**Y**S**K**S**V**G**F**M**Q**W**D**N**N**G**N**T**G**N**A**T**C**F**V**F**N**G**G**Y**I**F**T**C**R**H**V**V**H**L**M**V**G**K**N**T**
H**P**S**L**W**P**D**I**I**S**K**C**A**K**V**T**F**T**Y**T**E**F**C**P**T**P**D**N**W**F**S**I**E**P**W**L**K**V**S**N**E**N**L**D**Y**A**I**L**K**L**
K**E**N**G**N**A**F**P**P**G**L**W**R**Q**I**S**P**Q**P**S**T**G**L**I**Y**L**I**G**H**P**E**G**Q**I**K**K**I**D**G**C**T**V**I**P**L**N**E**R**L**K**
K**Y**P**N**D**C**Q**D**G**L**V**D**L**Y**D**T**T**S**N**V****Y**C**M**F**T**Q**R**S**F**L**S**E**V**W**N**T**H**T**L**S**Y**D**T**C**F**S**D**G**S**S
G**S**P**V**F**N**A**S**G**K**L**V**A**L**H**T**F**G**L**F**Y**Q**R**G**F**N**V**H**A**L**I**E**F**G**Y**S**M**D**S**I**L**C**D**I**K**K**T**N**E**S**
L**Y**K**S**L**N**D**E**K**L**E**T**Y**D**E**E**K**G**K**Q**E**S**S**L**Q**D**H**Q**I**E**P**M**E**C**

Figure S5. Two-dimensional structure of FAM111B predicted by GOR4 via Biology WorkBench 3.2 (URL: <http://workbench.sdsc.edu>). Alpha helices are colored in red, beta sheets in blue and random coils in black. The three amino acid changes identified in the study appear in larger font size. According to all other prediction programs tested, including PredictProtein (<https://www.predictprotein.org/>), and Biology WorkBench 3.2 (<http://workbench.sdsc.edu>), no transmembrane domains are predicted for *FAM111B*.

Query: MNSMKTEENKSFSAMEDDQTRPEVSKDT**VMKQ**THADTPV**DH**CLSGIR**KCS**STFKLKSEV
 Jpred: -----**HHH**-----**HH**-----**HH**-----
 Conf: 998744345565676777767530045400040146765110001400007700000000

Query: **NKH**ETALEMQPNLNN**KECCFT**FTLNNGNSRKLDRSV**FTAY**GKPSESI**YSALS**ANDY**FSE**R
 Jpred: **-HHH**-----**EEEEEE**-----**EEEE**-----**HHHHHH**-**HHHHHH**
 Conf: 0000000004677877506888843654546770588004787400000000099998

Query: **IKNQ**FNKNIIV**YEEK****TIDGH**INLGMP**LKCL**PSDS**HFKIT**FGQRKSS**KEDGH**IL**RQCEN**PN
 Jpred: **HHHH**-----**EEEE**--**HHHHHH**-----**EE**-----**EEEEEE**-----**HEEE**-----
 Conf: 75000875477743000000068750000588508988447777644000000005788

Query: **MEC**ILFH**VVA**IGR**TRKK**I**VKIN**EL**HEKGS**KL**CIYAL**K**GETI**E**GALC**KDGRFRSDIG**EFEW**
 Jpred: **-EEEEEEEEE**-----**EEEEEE****HHHHH**--**EEEEEE**--**HHHHHHH**-----**E**
 Conf: 7068899800072000000001203754788860675367666404664000077751

Query: **KLKE**GH**KIY**G**KQSM**VDEVSG**KVLE**MDIS**KKKAL**Q**QKDI**HKIK**QNESAT****DE**IN**HQS**LIQ
 Jpred: **EEE**-----**EE**--**EEEE**-----**EEEEEE**-----**HHHHHH**-----
 Conf: 5642564000560000001354378888605770076665677776530000000047

Query: SKKK**VHKPK**KD**GETK**D**VEHS**RE**QIL**PPQ**DL**SHYIK**DKTRQ****TIP**RI**RNY**F**CSL**PR**KYRQ**I
 Jpred: -----**HHHHHHHHH**-----**HHHHHH**--**HH**--**HHHHHHHHH**
 Conf: 764014554340008888003007777776645545400000000000000000000

Query: **NSQ**VR**RRPH**LGR**RYA**INLDVQ**KEA**IN**LK**NY**QTL**NEA**IMH**Q**YPNF****KEEA**Q**WVR**K**YF**REEQ
 Jpred: **H**-----**HH**--**HEEE**-----**HHHHHH**-**HHHHHHHHH**-----**HHHHHHHHHHHHHHH**
 Conf: 000000011100000006777525554100068999880068800689999999999999

Query: **KRM**NLSPAK**QFNI**Y**KKDF**G**KMT**ANSV**SAT**CE**QLTY**Y**SKS**V**GFM**Q**WD**NNG**NTGN**AT**CFV**F
 Jpred: **HHHH**-----**HHHHHHHHH**-----**HHHHHHHHHH**--**EEEEEE**-----**EEEEEE**
 Conf: 986158887001466541367577765178899880008806888615888500899984

Query: **NGY**I**FTCR**H**VVH**LMV**GKN**THPSL**WPD**I**ISK**AK**VT**FTY**TE**FC**PTD**N**WFS**IE**PWL**KVSN
 Jpred: --**EEEEEE****HHHHHHHHH**-----**HHHHHHHHHEEE**-----**EE**-----
 Conf: 570687000122133225777767000000000000367777763220566454457

Query: ENLD**YAIL**KL**KEN**NA**FP**PL**WRQ**ISP**QP**ST**GLI**Y**LIGH**PEG**QIK**KID**GCTV**I**PL**NER**LK**
 Jpred: ---**EEEEEE**-----**EEEEEE**-----**EE**-----**HH**
 Conf: 780058888406777777007777778750688850777740006700000775000

Query: **KYP**ND**CQ**DGLVDLYD**TTSNV****Y**CM**FTQ****RS**FLSEV**WN**TH**LSY**D**TCF**SDGSS**GS**PFV**NAS**GK
 Jpred: **HHHHHH**-----**EE****EEEE**--**EE**-----**EEEEEE**-----**E**
 Conf: 000000700077777735776100070767777760688001007777664157770

Query: **LVA**L**HTF**GL**FY**Q**R**GF**NV**HALIE**FGYS**M**D**S**IL**CD**IK**T**NES**LY**KSL**N**DE**K**LETY**D**E**E**K**G**KQ**
 Jpred: **EEEEEEEEEE**-----**EEEEEEEE****HHHHHHHHHHHHHHHHHHHHHHHHHHHHHH**-----
 Conf: 588753222000777700368875650267888750000799999870002145224320

Query: **ESS**L**QD**H**QI**E**P**M**E**C
 Jpred: -----
 Conf: 00111133677899

Figure S6. Two-dimensional structure of FAM111B predicted by JPRED 3 (<http://www.compbio.dundee.ac.uk/www-jpred>). Alpha helices (H) are colored in red, beta sheets (E) in blue and random coils (-) in black; conf means confidence. The three amino acid changes identified in the study appear in larger font size. Predictions of the two-dimensional structure suggest that the amino acid residue 621(Y) is included in a beta sheet structure, whereas residues 627 (R) and 628 (S) lie either in a beta sheet or in a random coil structure.

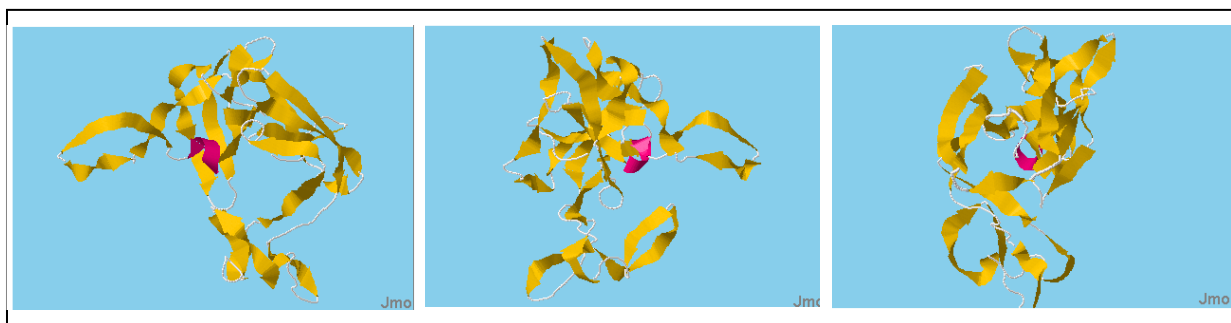


Figure S7. Three-dimensional structure of FAM111B predicted by RAPTORGX (URL: <http://raptorx.uchicago.edu/>). Although no significant homology was found between FAM111B and known proteins, a model was nonetheless proposed by Raptor X from templates 111j, 3b8j, 3lgv, 3mh5, 3nzi, 3tjn accessible through the ProteinDataBank (URL: <http://www.rcsb.org/pdb/home/>), which focused on the region of the protein harboring the putative protease domain

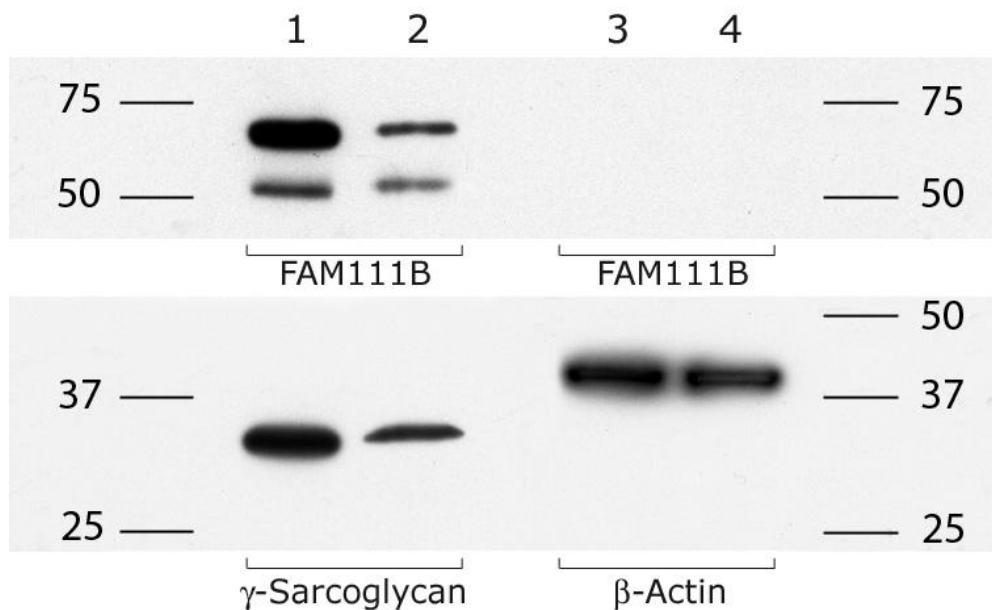


Figure S8. Western blot analysis of FAM111B levels in striated skeletal muscle and fibroblasts. Lane 1: normal control muscle; Lane 2: pathologic muscle (individual F1-II2); Lane 3: normal control fibroblasts; Lane 4: cultured fibroblasts from the skin of individual F1-II2. In order to determine FAM111B levels, immunoblot assays were performed on the muscle biopsy of individual F1-II2 as well as on a fibroblast culture from skin biopsy (upper panel). For total protein extraction, pellets of cultured cells and tissue samples from a normal control and individual F1-II2 were homogenized in a stringent SDS-containing RIPA buffer containing protease inhibitors as described previously.⁶ Protein concentration was determined using the Lowry assay (DC protein assay kit; Bio-Rad, Hercules, CA). The protein-containing lysate (20 μ g) was run on AnykD SDS-polyacrylamide precast gels (Bio-Rad) and electrotransferred onto nitrocellulose membranes (Bio-Rad). After overnight blocking (1% reagent; Roche Diagnostics), the membranes were probed with rabbit polyclonal antibodies directed to FAM111B (HPA038637, 1:200, Sigma-Aldrich, Co. LLC), followed by horseradish peroxidase-conjugated goat antirabbit antibody (1:20,000; Jackson ImmunoResearch, West Grove, PA). As a control, the membrane related to muscle biopsies was also probed with Novocastra mouse monoclonal antibody directed to SGCG (gamma-sarcoglycan; NCL-g-SARC, 35DAG/21B5, 1:150, Leica Biosystems Newcastle Ltd, UK), followed by horseradish peroxidase-conjugated goat antimouse antibody (1:2,000; sc-2005, Santa Cruz Biotechnology); the membrane related to fibroblasts cultures was probed with Sigma mouse monoclonal antibody directed to ACTB (beta-actin; clone AC-15, 1:20,000, Sigma-Aldrich, Co. LLC), followed by horseradish peroxidase-conjugated goat antimouse antibody (1:2,000; sc-2005, Santa Cruz Biotechnology). The immunoreactive proteins were detected on films using an enhanced chemiluminescence substrate according to the manufacturer's instructions (Roche Diagnostics). FAM111B levels were detected in striated skeletal muscle –expected size of the main product= 72 KDa-, but not in fibroblasts, thereby pointing to tissue-specific rather than to ubiquitous levels of FAM111B. In the pathologic muscle, both levels of FAM111B and of SGCG –a control specific to striated skeletal muscle of expected size about 35 KDa- were decreased in a proportional manner, compared to normal muscle. Given the tissue specificity of the loading control used, we can assess from that the apparent decreased levels of FAM111B in the pathologic muscle would rather be due to the infiltration of adipocytes in which FAM111B and SGCG are not detectable, than to a direct consequence of individual F1-II2's mutation on FAM111B levels in striated muscle.

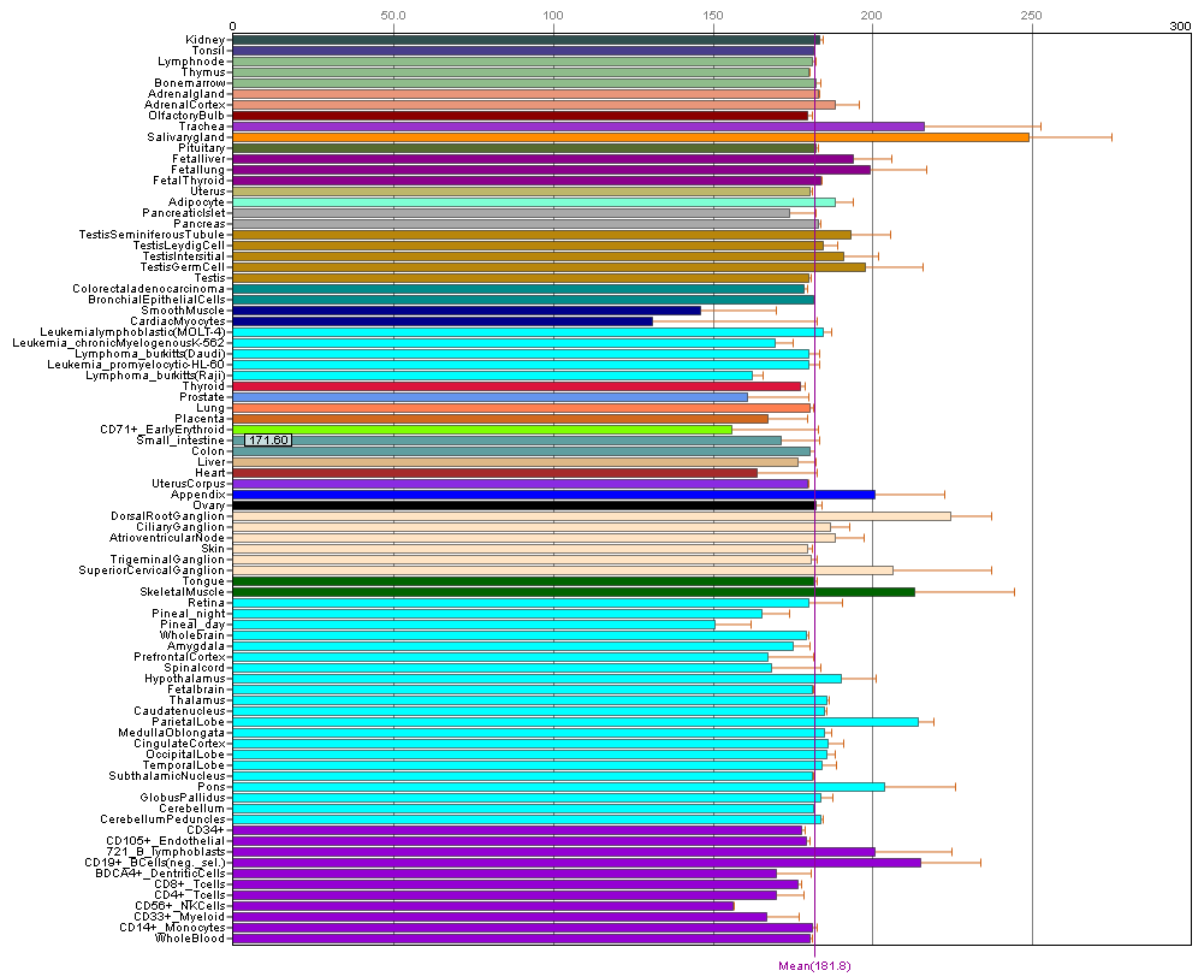


Figure S9. mRNA expression of *FAM111B* determined by microarray using dataset GeneAtlas U133A, gcrma and probeset gnf1h10646_at (results extracted from BioGPS website, URL: <http://biogps.org/#goto=genereport&id=374393>).

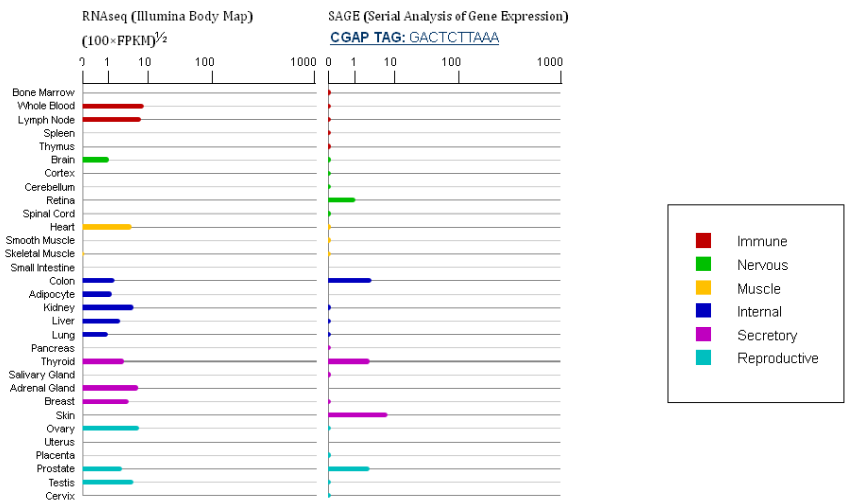


Figure S10. mRNA expression of *FAM111B* determined by RNAseq and Serial Analysis of Gene Expression (results extracted from Genecards website: <http://www.genecards.org/cgi-bin/carddisp.pl?gene=FAM111B&search=FAM111B>).

Table S1. Predictions of pathogenicity for the variants identified in HFP patients of the study.

Patient ID	Gene	Chromosomal location (assembly)	Transcript	Nucleotide Variant	Protein variant	Protein domain altered	Variant frequency in public databases	Grantham Distance	Align-GVGD class (Grantham Variation-Grantham Deviation)	SIFT prediction (weight-median)	PolyPhen-2 prediction (score)	Multivariate Analysis of Protein Polymorphism prediction (pValue)	Mutation Taster prediction (pValue)
Mother and sons of the SA Family	<i>FAM111B</i>	Chr11:58893431 (GRCh37)	NM_198947.3	c.1861T>G	p.Tyr621Asp	Serine/cysteine peptidase. trypsin-like	none (no rs id)	160	C15 (206.55-110.05)	Deleterious (0.01-3.25)	Benign (0.175)	bad (0.00002428)	polymorphism (1)
1.2.3	<i>FAM111B</i>	Chr11:58893449 (GRCh37)	NM_198947.3	c.1879A>G	p.Arg627Gly	Serine/cysteine peptidase. trypsin-like	none (no rs id)	125	C25 (42.81-86.56)	Deleterious (0.01-3.25)	Probably Damaging (0.999)	bad (0.0005459)	polymorphism (1)
4	<i>FAM111B</i>	Chr11:58893453 (GRCh37)	NM_198947.3	c.1883G>A	p.Ser628Asn	Serine/cysteine peptidase. trypsin-like	none (no rs id)	46	C0 (68.2-17.8)	Deleterious (0.05-3.25)	Probably Damaging (0.999)	good (0.01864)	polymorphism (1)

Supplemental References

1. Unger, S., Gorna, M.W., Le Behec, A., Do Vale-Pereira, S., Bedeschi, M.F., Geiberger, S., Grigelioniene, G., Horemuzova, E., Lalatta, F., Lausch, E., et al. (2013). FAM111A Mutations Result in Hypoparathyroidism and Impaired Skeletal Development. *Am J Hum Genet* <http://dx.doi.org/10.1016/j.ajhg.2013.04.020>.
2. Oates, M.E., Romero, P., Ishida, T., Ghalwash, M., Mizianty, M.J., Xue, B., Dosztanyi, Z., Uversky, V.N., Obradovic, Z., Kurgan, L., et al. (2013). D(2)P(2): database of disordered protein predictions. *Nucleic Acids Res* 41, D508-516.
3. Kallberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., and Xu, J. (2012). Template-based protein structure modeling using the RaptorX web server. *Nat Protoc* 7, 1511-1522.
4. Wu, C., Macleod, I., and Su, A.I. (2013). BioGPS and MyGene.info: organizing online, gene-centric information. *Nucleic Acids Res* 41, D561-565.
5. Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S., et al. (2010). Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol* 28, 1248-1250.
6. Charrier, L., Jarry, A., Toquet, C., Bou-Hanna, C., Chedorge, M., Denis, M., Vallette, G., and Laboisie, C.L. (2002). Growth phase-dependent expression of ICAD-L/DFF45 modulates the pattern of apoptosis in human colonic cancer cells. *Cancer Res* 62, 2169-2174.