Supplemental Data

Supplemental Table 1: Conservation table for the 119th fibronectin 3 domain of *TTN*

Species	Reference sequence	Amino acid sequence
Human	CAD12456.1	PGPCGKLTVSR VTQEKCTLAWSLPQEDGGAEITHYIVERRETSRLNWVIVEGECPTLSYVVTRLIKNNEYIFRVRAVNKYGPGVPVESEPIVA
Chimp	ENSPTRP00000021693	PGPCGKLTVSR VTQEKCTLAWSLPQEDGGAEITHYIVERRETSRLNWVIVEGECPTLSY KRLIINNEYIFRVRAVNKYGPGVPVESQPIVA
Mouse	ENSMUSP00000107477.2	PGPCGKLTVSR VTEEKCTLAWSLPQEDGGAEITHYIVERRETSRLNWVIVEGECLTASYVVTRLIKNNEYTFRVRAVNKYGLGVPVESEPIVA
Cow	ENSBTAP00000053669	PGPCGKLTVSR VTEEKCTLAWSLPQEDGGAEITHYIVERRETSRLNWVIVEGECPTLSHVVTRLIKNNEYIFRVRAVNKYGPGVPVESEPIVA
Dog	ENSCAFP00000020728	PGPCGKLTVSR ITEEKCTLAWSLPQEDGGAEITHYIVERRETSRLNWVIVEAECPTLSYVVTRLIKNNEYIFRVRAVNKYGPGVPVESEPIVA
T devil	XP_003764065.1	PGPCGKLMISR VTEEKCTLAWTLPQEDGGAEITHYIVERRETSRLNWVIVEGECPTLSHVVTRLIKNNEYIFRVRAVNKYGPGVPVESEPIIA
Rabbit	ENSOCUP00000014521	PGPCGKLTVSR VTEEKCTLAWSLPQEDGGAEITHYIVERRETSRLNWVIVEGECPTLSYVVTRLIKNNEYIFRVRAVNKYGPGVPVESEPIVA
Rat	XP_575155.4	PGPPGGPIEFKTVTAEKITLLWRPFADDGGAKITHYIVEKRETSRVVWSMVAENLEECIITTTKIIKGNEYIFRVRAVNKYGIGEPLESEPVVA
Zebrafish	ENSDARP00000019957	PGPCQEITVSN VSEEKCTVSWKVPQEDGGDPITHYIVERRDTNRLNWVIMEAECKALTCEIRRLFKNNEYIFRVRGVNKYGPGVLVQSNPIFA

Figure legend: The corresponding positions to the human sequence for the mutations are indicated at p.30071 (blue), p.30091 (green), and p.30145 (yellow)

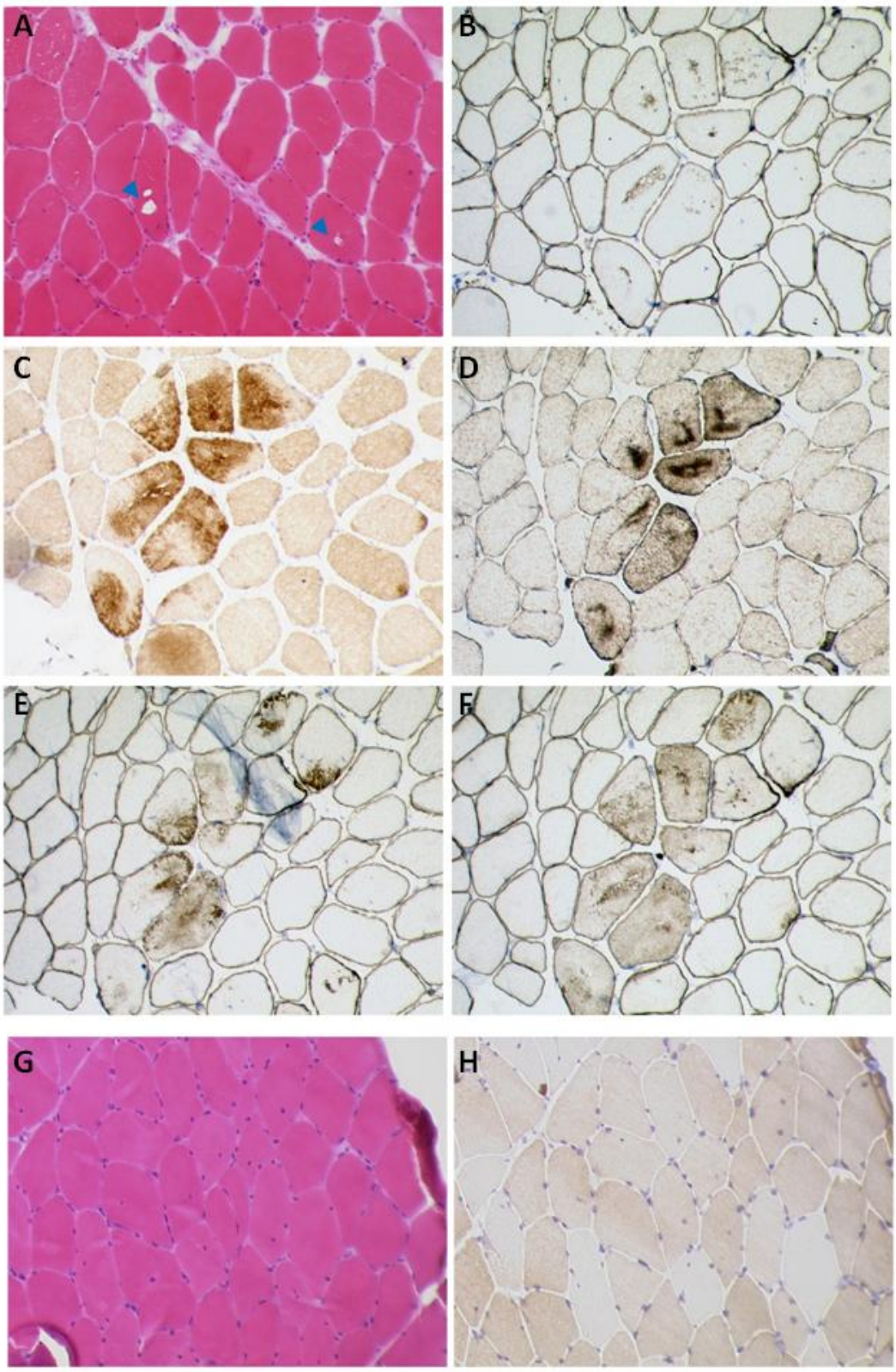
Supplemental Table 2: Novel coding variants in *TTN* from 239 exomes performed at our center

Domain	n aa	% aa	n variants	% variants
Z-DISC (1-1605)	1605	4.7	6	8.5
I-BAND (1606-14015)	12410	36.1	20	28.2
A-BAND (14016-31948)	17933	52.2	35	49.3
KINASE (31949-33817)	1869	5.4	7	9.9
M-LINE (33818-34350)	533	1.6	3	4.2
TOTAL (1-34350)	34350	100	71	100

Table legend: Parentheses in domain indicate the amino acid positions which constitute that domain, using CAD12456.1 as the reference sequence, and referring to domains outlined in Kontrogianni-Konstantopoulos A et al, Muscle giants: molecular scaffolds in sarcomerogenesis. *Physiol Rev* 89: 1217–1267, 2009 and Granzier and Labeit, Titin and its associated proteins, p. 88-119 in Squire JM and Parry DAD, Fibrous proteins: muscle and molecular motors. *Advances in Protein Chemistry and Structural Biology* Vol 71. Academic Press 2005. ISBN: 9780120342716

domain=the region of the titin gene of interest
n aa=number of amino acids in that domain
% aa=percentage of amino acids in this domain vs the length of the whole titin protein
n variants=number of novel sequence variants in that domain
% variants=percentage of novel sequence variants in that domain compared with all the novel titin variants

Supplemental Figure



Histological and Immunohistochemical findings for patients F.4A (A-F) and F.4B (G, H). A, G: H&E staining. Immunolabelling for spectrin (B), Myotilin (C,H), Desmin (D), delta-sarcoglycan (E), and dystrophin C-terminal (F). Note the presence rimmed vacuoles (arrow head) and intracitoplasmatic aggregates on labelling for delta-sarcoglycan and dystrophin in patient F.4A, and absence of major histopathological changes or protein aggregation in patient F.4B.

Supplemental Methods

Software and Scripts used.

The 100 genomes data was downloaded between positions 178404570 and 179734924 using **tabix** which is part of the **samtools** package (samtools.sourceforge.net). This was then converted to the IMPUTE format for phased data by **vcftools** (vcftools.sourceforge.net). Filtering this data by only including positions shared with our assays was performed using a custom written **python** script, and then the data was prepared for the programs **PHASE** (stephenslab.uchicago.edu/software.html) and **haploview** (www.broad.mit.edu/mpg/haploview/ using **R** scripts, selecting only those individuals from the CEU and GB populations that did not have parents in the data set, and by adding homozygote wild type alleles at the disease position for 1000 genomes data.

Sample information was obtained from the file phase1_samples_integrated_20101123.ped downloaded from the 1000 genomes data repository**.**

PHASE was used to phase haplotypes, using the known haplotypes option to use the 1000 genomes data as known phase data. The pairs output file was used to determine the relative posterior probability of the different haplotype reconstructions for these data. The statistical programming environment **R** was used for post processing of **PHASE** pairs output. The **haploview** program was used to investigate the haplotype structure of the population data and to produce plots.

A script containing the commands used for **tabix**, and **vcftools** and **R** and **python** scripts are available from http://www.staff.ncl.ac.uk/i.j.wilson/sharedhaplotypes.