LETTER

https://doi.org/10.1038/s41586-019-1076-8

Precise therapeutic gene correction by a simple nuclease-induced double-stranded break

Sukanya Iyer[1,9], Sneha Suresh[1,9], Dongsheng Guo[2,3], Katelyn Daman[2,3], Jennifer C. J. Chen[2,3,7], Pengpeng Liu[1], Marina Zieger[4,5], Kevin Luk[1], Benjamin P. Roscoe[1,8], Christian Mueller[4,5], Oliver D. King[2,3], Charles P. Emerson Jr[2,3,5]* & Scot A. Wolfe[1,5,6]*

Current programmable nuclease-based methods (for example, CRISPR–Cas9) for the precise correction of a disease-causing genetic mutation harness the homology-directed repair pathway. However, this repair process requires the co-delivery of an exogenous DNA donor to recode the sequence and can be inefficient in many cell types. Here we show that disease-causing frameshift mutations that result from microduplications can be efficiently reverted to the wild-type sequence simply by generating a DNA double-stranded break near the centre of the duplication. We demonstrate this in patient-derived cell lines for two diseases: limb-girdle muscular dystrophy type 2G (LGMD2G)[1] and Hermansky–Pudlak syndrome type 1 (HPS1)[2]. Clonal analysis of inducible pluripotent stem (iPS) cells from the LGMD2G cell line, which contains a mutation in *TCAP*, treated with the *Streptococcus pyogenes* Cas9 (*Sp*Cas9) nuclease revealed that about 80% contained at least one wild-type *TCAP* allele; this correction also restored *TCAP* expression in LGMD2G iPS cell-derived myotubes. *Sp*Cas9 also efficiently corrected the genotype of an HPS1 patient-derived B-lymphoblastoid cell line. Inhibition of polyADP-ribose polymerase 1 (PARP-1) suppressed the nuclease-mediated collapse of the microduplication to the wild-type sequence, confirming that precise correction is mediated by the microhomology-mediated end joining (MMEJ) pathway. Analysis of editing by *Sp*Cas9 and Lachnospiraceae bacterium ND2006 Cas12a (LbCas12a) at non-pathogenic 4–36-base-pair microduplications within the genome indicates that the correction strategy is broadly applicable to a wide range of microduplication lengths and can be initiated by a variety of nucleases. The simplicity, reliability and efficacy of this MMEJ-based therapeutic strategy should permit the development of nuclease-based gene correction therapies for a variety of diseases that are associated with microduplications.

MMEJ is an error-prone double-stranded break (DSB) DNA repair pathway that uses regions of microhomology (2–25 bp) on each side of the DSB to define the boundaries at which DNA segments are rejoined[3]. This mutagenic process generates deletions that result in the loss of one of the repeat sequences and the intervening region (Fig. 1a). A variety of cell types show hallmarks of MMEJ repair within DNA products produced by the editing of programmable nucleases. These sequence-templated deletion events, which produce products of a fixed length, can increase or decrease the efficiency of gene inactivation depending on the nature of the coding sequence disruption[4,5]. The MMEJ pathway has also been harnessed for the targeted insertion of exogenous donor DNAs in mammalian cells and zebrafish and frog embryos[6,7]. Here, we describe a nuclease-based therapeutic approach that harnesses the MMEJ pathway to precisely correct frameshift mutations resulting from microduplications (tandem duplications). We reasoned that MMEJ-based repair of a programmable nuclease-induced DSB near the centre of a disease-causing microduplication would achieve precise reversion to the wild-type genomic sequence. This strategy might be an effective alternative to homology-directed repair-based gene correction approaches and would not require co-delivery of a donor DNA. Furthermore, the reverted wild-type sequence would no longer be complementary to the single-guide RNA (sgRNA) targeting the microduplication, leading to stable correction even in the presence of Cas9 nuclease.

To evaluate the efficacy of our MMEJ-based correction strategy, we focused on LGMD2G and HPS1, two diseases that affect different human tissues and whose causes include pathogenic microduplications of different lengths. Both of these diseases are autosomal recessive disorders that are represented at modest frequencies in different human subpopulations and currently have no treatments. One of the disease alleles identified in patients with LGMD2G features an 8-bp duplication in exon 1 of *TCAP*, a mutation that is found in the East Asian population at a frequency of approximately 1 in 1,000 alleles. *TCAP* encodes the telethonin protein, a 19-kDa cardiac and striated muscle-specific structural protein located in the Z-disc of sarcomeres that links titin proteins to stabilize the contractile apparatus for muscle contraction[8]. Homozygous or compound heterozygous inactivating mutations in *TCAP* manifest as severe muscle atrophy and cardiomyopathy that typically develop during late adolescence into early adulthood[1,9].

We designed and tested an sgRNA for *Sp*Cas9 to generate a DSB one base pair away from the middle of the *TCAP* 8-bp microduplication (Fig. 1b). Purified *Sp*Cas9 protein was complexed with a synthetic sgRNA (RNP) and electroporated into iPS cells homozygous for the *TCAP* microduplication that were derived from a patient with LGMD2G. After four days, we used deep sequencing analysis to analyse the genomic region of interest for insertions and deletions (indels). We observed robust gene editing (about 80% indel rate), indicating that the *Sp*Cas9 RNP can efficiently generate DSBs at this site. Closer examination of the sequence variants revealed that on average about 57% of the alleles contained a precise 8-bp deletion corresponding to the wild-type allele (Fig. 1c, Extended Data Fig. 1a). Notably, when introduced into wild-type cells containing functional *TCAP*, the *Sp*Cas9 RNPs did not cause measurable editing at the *TCAP* allele, indicating that the corrected allele in the mutant cells is not subject to unintended damage following MMEJ-mediated reversion (Fig. 1c). In addition to the precise 8-bp deletion, we also observed that an additional approximately 17% of the alleles contained in-frame mutations, and therefore may encode hypomorphic alleles with some restoration of function (Extended Data Fig. 1a, c). Genotyping of 22 clones generated from a nuclease-treated LGMD2G iPS cell population revealed that 77% contained at least one wild-type allele, indicating that the majority of nuclease-treated cells would be therapeutically corrected (Fig. 1d, Supplementary Table 1). To independently verify the duplication collapse rates observed in edited iPS cells by Illumina short-read sequencing, we sequenced a 2-kb amplicon spanning the *TCAP* locus from a population of *Sp*Cas9-edited iPS cells using the Pacific Biosciences long-read sequencing platform

[1]Department of Molecular, Cell and Cancer Biology, University of Massachusetts Medical School, Worcester, MA, USA. [2]Department of Neurology, University of Massachusetts Medical School, Worcester, MA, USA. [3]Wellstone Muscular Dystrophy Program, University of Massachusetts Medical School, Worcester, MA, USA. [4]Horae Gene Therapy Center, University of Massachusetts Medical School, Worcester, MA, USA. [5]Li Weibo Institute for Rare Diseases Research, University of Massachusetts Medical School, Worcester, MA, USA. [6]Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA, USA. [7]Present address: Office of the Vice-Principal (Research), Queen's University, Kingston, Ontario, Canada. [8]Present address: COGEN Therapeutics, Cambridge, MA, USA. [9]These authors contributed equally: Sukanya Iyer, Sneha Suresh. *e-mail: charles.emersonjr@umassmed.edu; scot.wolfe@umassmed.edu

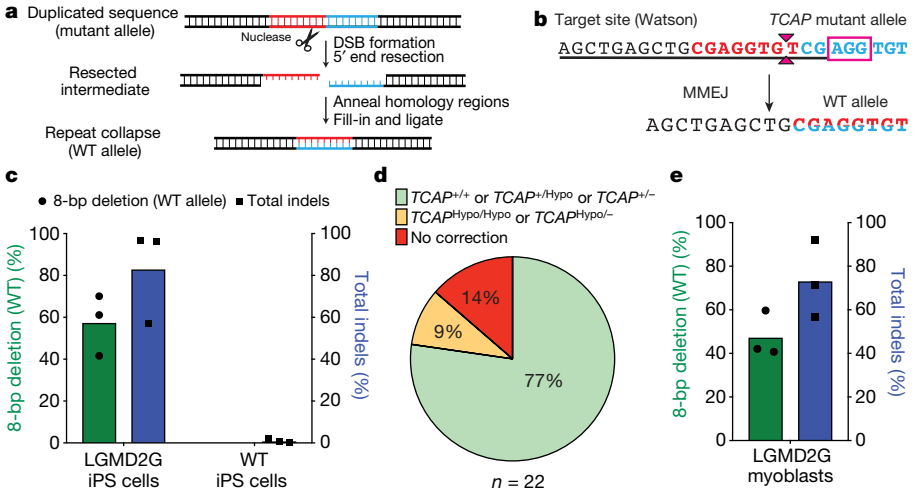


Fig. 1 | MMEJ-based repair efficiently and precisely corrects *TCAP* allele containing an 8-bp duplication. a, Schematic of MMEJ-based pathway for repair of a microduplication. A DSB at the centre of a microduplication (in red and blue) is expected to initiate 5′ end resection to expose the microhomologies on each side. These repeats anneal with each other and are repaired via the MMEJ pathway to yield the wild-type (WT) sequence. **b**, The pathogenic 8-bp microduplication within *TCAP* (bold red and blue text) with the *Sp*Cas9 protospacer-adjacent motif (PAM) sequence in the magenta box and the protospacer sequence underlined. A *Sp*Cas9-induced DSB (magenta carets) is expected to drive MMEJ repair to revert the mutant allele to the wild-type sequence (half red/half blue text). **c**, Percentage of 8-bp deletion (green bars) and total indels (blue bars) resulting from *Sp*Cas9 RNP treatment of LGMD2G iPS cells homozygous for the 8-bp microduplication or wild-type iPS cells. Bars denote mean and dots indicate individual data points. $n = 3$ biological replicates. **d**, Genotype analysis of 22 LGMD2G iPS cell clones after treatment with *Sp*Cas9 RNPs. Hypo, hypomorphic allele. **e**, Percentage of 8-bp deletion (green bars) and total indels (blue bars) resulting from *Sp*Cas9 treatment of myoblasts derived from patient-derived LGMD2G iPS cells. Bars denote mean and dots indicate individual data points. $n = 3$ biological replicates.

(PacBio). Analysis of these reads revealed that 67% of the edited alleles with insertions or deletions below 100 bp in length corresponded to the 8-bp collapse, which is similar to the 73% rate of 8-bp collapse determined by Illumina sequencing for this sample (Extended Data Fig. 2, Supplementary Table 2). Treatment of cells with Cas9 nuclease can produce large deletions (>100 bp) at the target locus at a modest frequency[10]. Consistent with these findings, our PacBio analysis revealed the presence of large deletions (100–1,000 bp) that would not have been detected by Illumina sequencing at a frequency of about 2% in bulk edited iPS cells. We also isolated a genotypically complex iPS cell colony that harboured two large deletions at the *TCAP* locus (Extended Data Fig. 3).

To demonstrate the translatability of our approach to muscle cell types, we differentiated the LGMD2G iPS cells into proliferative skeletal myoblasts that can be induced to terminally differentiate into myotubes[11]. iPS cell-derived myoblasts can repair damaged muscle in a similar way to myogenic satellite cells (one of the primary targets of gene therapy for myopathies). Myoblasts were electroporated with *Sp*Cas9 RNPs programmed to target the 8-bp microduplication. Following editing, about 45% of the alleles were precisely repaired back to the wild-type sequence (Fig. 1e, Extended Data Fig. 1b, d). Immunostaining of myotubes derived from corrected LGMD2G iPS cell clones with an anti-telethonin antibody showed that genetic correction restored telethonin expression (Extended Data Fig. 4). Collectively, these data show that introducing a DSB close to the centre of microduplication can efficiently achieve precise in vitro correction of the 8-bp microduplication associated with LGMD2G in iPS cells and in myoblasts that mimic cell populations that would be therapeutically targeted in vivo.

We further tested our approach on a 16-bp pathogenic microduplication in exon 15 of *HPS1*, which is associated with HPS1 and leads to the production of a truncated protein responsible for this autosomal recessive disease[12]. HPS1 has a high prevalence in the Puerto Rican population, with a carrier rate of approximately 1 in 21 in the northwest region[2]. HPS proteins are involved in the biogenesis of lysosome-related organelle complexes (BLOCs), which are necessary for the proper trafficking of cargo to melanosomes, dense granules and lysosomes[13]. Patients with HPS1 suffer from albinism, bleeding disorders, vision loss and progressive pulmonary fibrosis, which leads to premature death[14].

We tested the efficiency of gene correction in a patient-derived B lymphocyte cell line (B-LCL) homozygous for the 16-bp microduplication by electroporating these cells with *Sp*Cas9 RNPs programmed to cleave two base pairs away from the centre of the microduplication (target site 1; Fig. 2a). To accurately assess the observed editing rates, we added unique molecular identifiers (UMIs) to our PCR amplicons during Illumina library construction to allow the removal of any amplification bias[15]. We confirmed that this approach accurately captured the relative percentage of HPS1 microduplication and wild-type alleles present in a series of test populations (Extended Data Fig. 5). At *HPS1* target site 1, we observed editing at about 46% of the alleles with around 35% restored to the wild-type sequence (Fig. 2b, c, Supplementary Table 2). We further examined the effect of the position of the DSB within the microduplication on the efficiency of MMEJ-mediated repair by designing five additional sgRNAs that targeted the DSB to different positions relative to the centre of the microduplication (target sites 2–6, Fig. 2a). As the break site was shifted away from the centre, there was a decrease in the efficiency of achieving the precise 16-bp deletion (Fig. 2b, c). However, target sites 3 and 6 were notable exceptions to this trend. Target site 3 proved to be quite efficient at generating indels to the exclusion of the 16-bp deletion (Extended Data Fig. 6), probably because the wild-type sequence, once regenerated, can also be targeted by this sgRNA for further mutagenesis. On the other hand, target site 6 achieved efficient deletion of the 16-bp microduplication (more than 50% of the modified alleles), despite being the most distal of the cleavage sites (10 bp from the centre of the microduplication). Its efficiency may be due to the extended regions of homology that surround the cleavage site at this end of the microduplication (target site 6; Fig. 2a). Overall, these results demonstrate that the cleavage position within the microduplication and the presence of alternate regions of microhomology can influence the production of the desired wild-type end product (Fig. 2c).

To investigate whether nuclease-mediated collapse of a microduplication occurs via the MMEJ pathway, we inhibited a DNA repair factor (PARP-1) that regulates DSB flux through this pathway. PARP-1

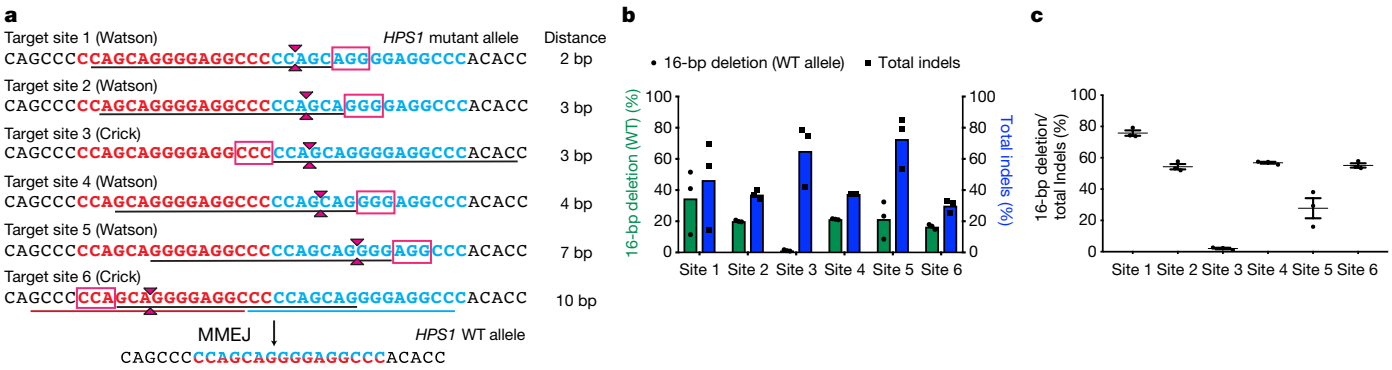


Fig. 2 | MMEJ-based repair efficiently and precisely corrects *HPS1* allele containing 16-bp microduplication. a, The 16-bp microduplication repeats are shown in bold red and blue text. For six *Sp*Cas9 guides targeting the microduplication, the PAM sequence is demarcated in the magenta box and the protospacer sequence is underlined. A DSB (magenta carets with distance from the repeat centre indicated) is expected to drive reversion to the wild-type sequence (half red/half blue text). Sequence underlined with red and blue bold lines in target site 6 indicates an alternate 16-bp microhomology within this repeat. **b**, Percentage of 16-bp deletions (green) and total indels (blue) for guides shown in **a** based on UMI-based Illumina sequencing. Bars denote mean and dots indicate individual data points. $n = 3$ biological replicates. **c**, Percentage of wild-type reverted alleles (16-bp deletion) among all alleles with insertions or deletions (indels) from **b**. Mean ± s.e.m., dots indicate individual data points. $n = 3$ biological replicates.

influences the repair of a DSB through resection-dependent DNA repair pathways, such as MMEJ[3,16], which are in competition with the non-homologous end joining pathway (NHEJ) for DSB repair[17] (Extended Data Fig. 7a). Inhibition of the catalytic activity of PARP-1 by rucaparib reduces DSB flux through the MMEJ pathway, resulting in a decrease in microhomology-based deletion products in the resulting repair events[18] (Extended Data Fig. 7a). Patient-derived *HPS1* B-LCL cells were treated with 10 μM or 20 μM rucaparib before and after treatment with *Sp*Cas9 RNP to suppress MMEJ-mediated repair of DSBs (Fig. 3a). We observed an overall reduction in editing rates at the *HPS1* locus upon rucaparib treatment (Fig. 3b). These lower editing rates were primarily the result of a reduction in the 16-bp deletion product, which decreased from about 50% in untreated cells to around 15% and 6% in cells treated with 10 μM and 20 μM rucaparib, respectively (Fig. 3b–d). We observed a similar reduction in microhomology-based deletions with *Sp*Cas9 RNP targeting the *AAVS1* locus in patient-derived *HPS1* B-LCL cells (Extended Data Fig. 7). Thus, the MMEJ pathway underlies the robust correction of the microduplications for LGMD2G and HPS1 in the presence of a targeted DSB.

To test the generality of this MMEJ-based repair approach and the range of sequence lengths over which duplication collapse is efficient, we evaluated the editing products generated by *Sp*Cas9 targeting endogenous microduplications within the human genome. We performed a bioinformatic analysis to identify non-pathogenic, unique endogenous microduplications ranging from 4 bp to 36 bp in length in the human genome (Fig. 4a). We examined the efficiency of microduplication collapse resulting from a *Sp*Cas9 produced DSB at the centre of the microduplications in HEK293T cells at these sites. Although the bulk editing rate varied across these target sites, we consistently found that duplication collapse was the major end product within the edited alleles (ranging from 45% to 93%), regardless of the microduplication length (Fig. 4b, Extended Data Fig. 8a). Consistent with the analysis at the *HPS1* locus, we observed a decrease in the duplication collapse efficiency for 24- and 27-bp-long microduplications as cut sites were moved away from the centre (Extended Data Fig. 8b–g).

Whereas *Sp*Cas9 generates blunt DSBs, the type V CRISPR–Cas nuclease Cas12a generates DSBs with 5′ overhangs[19]. We investigated whether LbCas12a-generated breaks might be preferentially repaired by a resection-dependent pathway such as MMEJ by comparing the efficiency of microduplication collapse engendered by *Sp*Cas9 and LbCas12a nucleases at three endogenous sites. Efficient repeat collapse (50–90% of edited alleles) could be achieved with LbCas12a at all three of these sites, with efficiencies similar to those of *Sp*Cas9 (Fig. 4c, Extended Data Fig. 8a). Overall, these data demonstrate that

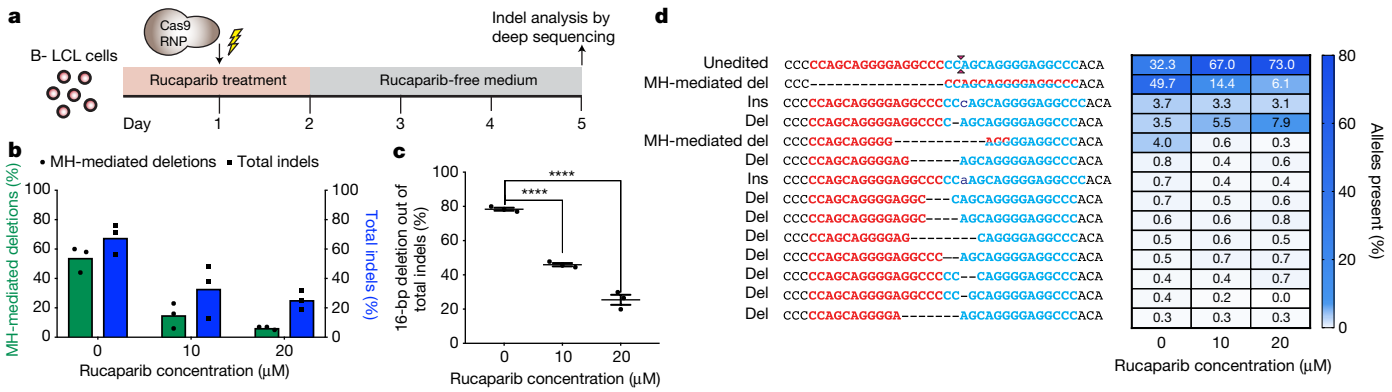


Fig. 3 | PARP-1 inhibition decreases efficiency of MMEJ-based repair. a, Experimental design. HPS1 B-LCL cells were treated with rucaparib 24 h before and after electroporation with *Sp*Cas9 RNPs targeting the *HPS1* locus and collected for subsequent UMI-based Illumina sequencing[15]. **b**, Percentage of microhomology (MH)-mediated deletion (green) and total indels (blue) in cells treated with *Sp*Cas9 in the presence of 0, 10 or 20 μM rucaparib, measured by UMI-based Illumina deep sequencing. Bars denote mean and dots indicate individual data points. $n = 3$ biological replicates. **c**, Percentage of microhomology-mediated deletion alleles among all alleles with indels from **b**. Mean ± s.e.m., dots indicate individual data points. $n = 3$ biological replicates. $^{****}P = 0.00003$, unpaired two tailed *t*-test (Supplementary Table 9). **d**, Left, alignment of resulting sequences observed by Illumina sequencing upon *Sp*Cas9 RNP treatment of *HPS1* B-LCL cells. Right, heat map showing percentage of alleles generated by *Sp*Cas9 for cells exposed to 0, 10 or 20 μM rucaparib. Gradient scale indicates the percentage occurrence of that sequence.

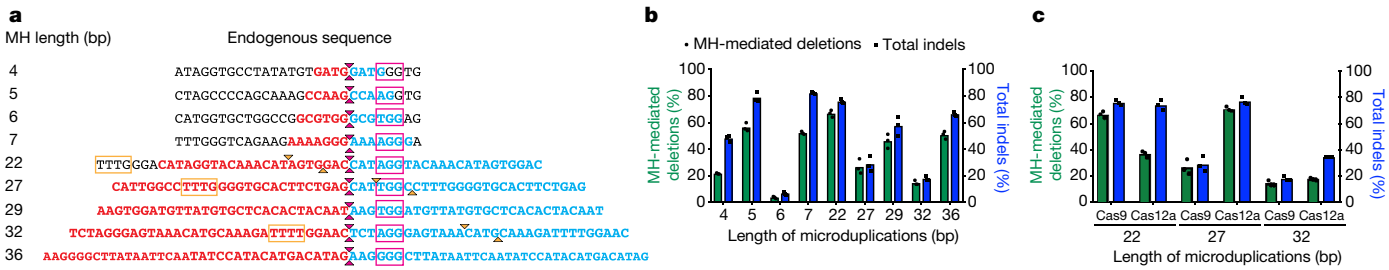


Fig. 4 | MMEJ-based approach efficiently achieves precise collapse of endogenous microduplications across various repeat lengths. a, Non-pathogenic endogenous microduplications ranging in size from 4 bp to 36 bp. Microduplication repeats are shown as bold red and blue text. The *Sp*Cas9 PAM sequence is shown in the magenta box and the LbCas12a PAM sequence is shown in the orange box. Anticipated DSBs produced by *Sp*Cas9 and LbCas12a are denoted by magenta and orange carets, respectively. **b**, Percentage of microhomology-mediated deletion (green) and total indels (blue) produced at each endogenous site following *Sp*Cas9 treatment, calculated using UMI-based Illumina sequencing. Bars denote mean and dots indicate individual data points. $n = 3$ biological replicates. **c**, Percentage of microhomology-mediated deletions (green) and total indels (blue) produced at three endogenous sites when treated with *Sp*Cas9 or LbCas12a. Bars denote mean and dots indicate individual data points. $n = 3$ biological replicates.

the MMEJ-based editing approach can be used to efficiently collapse microduplications up to lengths of at least 36 bp using either Cas9 or Cas12a programmable nucleases.

To investigate the potential of this MMEJ-based therapeutic strategy to be applied more broadly for correcting human genetic disorders, we performed a bioinformatic analysis to gauge the prevalence of disease-causing microduplications in human populations. The ClinVar database[20] includes about 4,700 duplications that are annotated as 'pathogenic' or 'pathogenic/likely pathogenic' (Extended Data Fig. 10a). We focused on duplications of lengths ranging from 2 to 40 bp because our data indicate that microhomologies within this range can be precisely repaired via the MMEJ pathway (Fig. 4). We also focused on 'simple' duplications—those for which the duplicated sequence is not part of a more complex repeat structure—to improve the odds that the primary homology-based collapse would result in the desired wild-type sequence. Finally, we examined all duplications in 'coding' regions (mainly exons plus 50 flanking bases) from the gnomAD exome and genome sequencing databases[21] to prioritize pathogenic duplications according to their frequencies in human populations (Extended Data Figs. 9, 10a). Our analysis yielded 143 likely disease-causing microduplications of lengths 2–40 bp that were observed at least once in gnomAD (Extended Data Fig. 10b), some of which occur in specific subpopulations at substantial frequencies (for example, Tay–Sachs disease, Supplementary Table 3).

To facilitate the utilization of our bioinformatics analysis, we have created an interactive, searchable webtool (https://rambutan.umassmed.edu/duplications/). This analysis also included the identification of potential Cas9 and Cas12a cleavage sites[22] within these microduplications (Supplementary Table 3 and expanded tables available at the website mentioned above). As shown by our guide 'tiling' data across the *HPS1* microduplication and endogenous microduplication sites, the position of the DSB break within the duplication, and the use of a guide design that avoids cleavage of the wild-type allele, are critical for efficient, stable collapse of microduplications. Rapid advances are being made in characterizing nucleases with alternate specificities[23,24] and in engineering nucleases with alternate or expanded recognition preferences[25–27], which will make correction of disease-causing microduplications using the MMEJ-based approach even more effective.

DSBs at most genomic sites (for example, *AAVS1*, Extended Data Fig. 7) are repaired primarily through the NHEJ pathway, which can produce small insertions or deletions during imprecise repair[28]. Our data, which span DSBs in 12 sequences, indicate that microduplications are preferentially repaired via the MMEJ pathway, which yields predictable and efficient collapse. For this class of pathogenic mutations, precise repair via the MMEJ pathway provides a favourable alternative to homology-directed repair, which is inefficient in many cell types[29]. Consistent with our findings, MMEJ-mediated repair was recently used to efficiently correct the pathogenic microduplication associated with *HPS1*[30]. Although the use of allele frequencies from gnomAD can help to prioritize potential targets for MMEJ-based repair, this underestimates the extent of genetic diseases—particularly dominant ones—caused by microduplications (see Supplementary Discussion). As genomic and phenotypic data for the human population continue to accrue, we are likely to discover new pathogenic microduplications that can be corrected using this MMEJ-based approach. Although the molecular mechanisms and cell-type-specific efficiencies of the MMEJ pathway remain to be completely elucidated, our findings lay the foundation for pursuing MMEJ-based therapeutic approaches for LGMD2G, HPS1 and a broader spectrum of microduplication-based diseases.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-019-1076-8.

Received: 28 May 2018; Accepted: 22 February 2019;
Published online 3 April 2019.

1. Moreira, E. S. et al. Limb-girdle muscular dystrophy type 2G is caused by mutations in the gene encoding the sarcomeric protein telethonin. *Nat. Genet.* **24**, 163–166 (2000).
2. El-Chemaly, S. & Young, L. R. Hermansky–Pudlak syndrome. *Clin. Chest Med.* **37**, 505–511 (2016).
3. Sfeir, A. & Symington, L. S. Microhomology-mediated end joining: a back-up survival mechanism or dedicated pathway? *Trends Biochem. Sci.* **40**, 701–714 (2015).
4. Bae, S., Kweon, J., Kim, H. S. & Kim, J.-S. Microhomology-based choice of Cas9 nuclease target sites. *Nat. Methods* **11**, 705–706 (2014).
5. Kim, S.-I. et al. Microhomology-assisted scarless genome editing in human iPSCs. *Nat. Commun.* **9**, 939 (2018).
6. Hisano, Y. et al. Precise in-frame integration of exogenous DNA mediated by CRISPR/Cas9 system in zebrafish. *Sci. Rep.* **5**, 8841 (2015).
7. Sakuma, T., Nakade, S., Sakane, Y., Suzuki, K. T. & Yamamoto, T. MMEJ-assisted gene knock-in using TALENs and CRISPR–Cas9 with the PITCh systems. *Nat. Protoc.* **11**, 118–133 (2016).
8. Bertz, M., Wilmanns, M. & Rief, M. The titin-telethonin complex is a directed, superstable molecular bond in the muscle Z-disk. *Proc. Natl Acad. Sci. USA* **106**, 13307–13310 (2009).
9. Nigro, V. & Savarese, M. Genetic basis of limb-girdle muscular dystrophies: the 2014 update. *Acta Myol.* **33**, 1–12 (2014).
10. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).
11. Caron, L. et al. A human pluripotent stem cell model of facioscapulohumeral muscular dystrophy-affected skeletal muscles. *Stem Cells Transl. Med.* **5**, 1145–1161 (2016).
12. Oh, J. et al. Positional cloning of a gene for Hermansky–Pudlak syndrome, a disorder of cytoplasmic organelles. *Nat. Genet.* **14**, 300–306 (1996).
13. Richmond, B. et al. Melanocytes derived from patients with Hermansky–Pudlak syndrome types 1, 2, and 3 have distinct defects in cargo trafficking. *J. Invest. Dermatol.* **124**, 420–427 (2005).
14. Brantly, M. et al. Pulmonary function and high-resolution CT findings in patients with an inherited form of pulmonary fibrosis, Hermansky–Pudlak syndrome, due to mutations in HPS-1. *Chest* **117**, 129–136 (2000).

15. Bolukbasi, M. F. et al. Orthogonal Cas9–Cas9 chimeras provide a versatile platform for genome editing. *Nat. Commun.* **9**, 4856 (2018).
16. Sharma, S. et al. Homology and enzymatic requirements of microhomology-dependent alternative end joining. *Cell Death Dis.* **6**, e1697 (2015).
17. Wang, M. et al. PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. *Nucleic Acids Res.* **34**, 6170–6182 (2006).
18. Dutta, A. et al. Microhomology-mediated end joining is activated in irradiated human cells due to phosphorylation-dependent formation of the XRCC1 repair complex. *Nucleic Acids Res.* **45**, 2585–2599 (2017).
19. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR–Cas system. *Cell* **163**, 759–771 (2015).
20. Landrum, M. J. et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–D1067 (2018).
21. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
22. Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell* **168**, 20–36 (2017).
23. Kim, E. et al. *In vivo* genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nat. Commun.* **8**, 14500 (2017).
24. Edraki, A. et al. A compact, high-accuracy Cas9 with a dinucleotide PAM for in vivo genome editing. *Mol. Cell* **73**, 714–726 (2019).
25. Kleinstiver, B. P. et al. Engineered CRISPR–Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).
26. Bolukbasi, M. F. et al. DNA-binding-domain fusions enhance the targeting range and precision of Cas9. *Nat. Methods* **12**, 1150–1156 (2015).
27. Hu, J. H. et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).
28. van Overbeek, M. et al. DNA repair profiling reveals nonrandom outcomes at Cas9-mediated breaks. *Mol. Cell* **63**, 633–646 (2016).
29. Suzuki, K. et al. *In vivo* genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **540**, 144–149 (2016).
30. Shen, M. W. et al. Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* **563**, 646–651 (2018).

Acknowledgements We thank E. Kittler and the UMass Medical School Deep Sequencing Core for sequencing; L. Hayward, L. Qin and D. McKenna-Yasek for coordinating patient enrolment and acquiring patient skin biopsies; Z. Matijasevic for generating LGMD2G iPS cell lines; and the Genome Aggregation Database (gnomAD) and the groups that provided exome and genome variant data to this resource. A full list of contributing groups can be found at http://gnomad.broadinstitute.org/about. This work was supported in part by the National Institutes of Health (R01DK098252, R01HL131471 and R01NS088689 (C.M.); R01AI117839, R01GM115911, R01HL093766 and U01HG007910 (S.A.W.); U54HD0060848 (C.P.E.); a SPARK award through UL1-TR001453) and the Worcester Foundation for Biomedical Research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Reviewer information *Nature* thanks Randall Platt and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions S.I. and S.S. performed and analysed the B-LCL and HEK293T editing experiments. D.G., J.C.J.C. and K.D. performed the iPS cell and myoblast editing experiments. S.I. and S.S. analysed the iPS cell and myoblast editing experiments. D.G. and J.C.J.C. generated the iPS cells and myoblast cell lines. D.G. and K.D. performed and analysed the flow cytometry and western blot data for detection of telethonin. B.P.R., P.L. and K.L. designed and purified the *Sp*Cas9 and LbCas12a proteins. P.L. and O.D.K. performed the bioinformatic analysis. P.L. analysed the deep sequencing data. C.M. and M.Z. contributed expertise to the HPS1 cell line editing and characterization. C.M., O.D.K., C.P.E. and S.A.W. directed the research. S.I., S.S., C.P.E. and S.A.W. wrote the manuscript with input from all of the other authors.

Competing interests The authors have filed patent applications related to genome engineering technologies. S.I., S.S., D.G., J.C.J.C., C.M., O.D.K., C.P.E. and S.A.W. have filed a patent application (62/667201) on this work. C.M. is a paid consultant for and with equity in Apic Bio. Apic Bio was not involved in the funding or any other aspects of this study. The authors have no other competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-019-1076-8.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1076-8.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to C.P.E. or S.A.W.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

METHODS

Human subjects. Cells for reprogramming *TCAP* iPS cell lines were recovered, with consent, from a skin biopsy from a patient with LGMD2G under a UMMS-IRB-approved protocol and assigned a de-identified ID number unlinked to the patient's medical record. The consent process included conditions for sharing de-identified samples and information with other investigators. No Protected Health Information will be shared at any time, per Health Insurance Portability and Accountability Act guidelines.

Cell culture. LGMD2G primary dermal fibroblasts were isolated from a skin biopsy from a patient with LGMD2G as described[31]. Fibroblasts were reprogrammed using the CytoTune 2.0 iPS Sendai Virus Reprogramming Kit (ThermoFisher Scientific) according to the manufacturer's directions. Clonal lines were expanded for 6–10 passages before banking. Immunostaining was performed to confirm the absence of Sendai virus and expression of OCT4. Human iPS cells were cultured in iPS-Brew XF medium (Miltenyi Biotec) and passaged every 3–5 days with Passaging Solution (Miltenyi Biotec) according to the manufacturer's directions.

Myoblasts were induced from iPS cells using a modification of the Genea Biocells protocol[11]. Following the generation of differentiated myotubes as described, cells were reseeded and cultured in human primary myoblast medium[32]. $CD56^+$ cells were purified by FACS using an anti-CD56–APC antibody (BD Biosciences) or MACS (Miltenyi Biotec) according to the manufacturer's directions. Myogenicity was confirmed by immunostaining myoblast and myotube cultures using the mouse monoclonal antibodies MyoD clone 5.8 (DAKO) and MF20 (DSHB) (data not shown).

A lymphoblastoid cell line from B lymphocytes (B-LCL) derived from a patient with HPS1 who was homozygous for the 16-bp microduplication was purchased from Coriell (GM14606). The cell line was cultured following the recommended procedure using RPM1 1640 with 2 mM L-glutamine, 15% FBS and 1% penicillin/streptomycin.

HEK293T cells were cultured following the recommended procedure using Dulbecco's modified Eagle's medium (DMEM), 10% FBS and 1% penicillin/streptomycin.

All cultures were maintained in a humidified incubator with 5% CO_2 at 37 °C. All cell lines were tested for mycoplasma contamination at regular intervals and the HEK293T cell line was authenticated by University of Arizona Genetics Core.

Purification of *Sp*Cas9 and LbCas12a. Protein purification for 3×NLS–*Sp*Cas9 and LbCas12a–2×NLS followed a common protocol. The generation and characterization of the 3×NLS–*Sp*Cas9 (Addgene 114365) and LbCas12a–2×NLS (Addgene 114366) constructs have been described[46,47]. The pET21a plasmid backbone (Novagen) is used to drive the expression of a hexa-His-tagged version of each protein. The plasmid expressing 3×NLS–*Sp*Cas9 (or LbCas12a–2×NLS) was transformed into *Escherichia coli* Rosetta (DE3)pLysS cells (EMD Millipore) for protein production. Cells were grown at 37 °C to an OD_{600} of ~0.2, then shifted to 18 °C and induced at an OD_{600} of ~0.4 for 16 h with isopropyl β-D-1-thiogalactopyranoside (IPTG, 1 mM final concentration). Following induction, cells were pelleted by centrifugation and then resuspended with Ni-NTA buffer (20 mM Tris pH 7.5, 1 M NaCl, 20 mM imidazole, 1 mM TCEP) supplemented with HALT Protease Inhibitor Cocktail, EDTA-Free (100×) (ThermoFisher Scientific) and lysed with M-110s Microfluidizer (Microfluidics) following the manufacturer's instructions. The protein was purified from the cell lysate using Ni-NTA resin, washed with five volumes of Ni-NTA buffer and then eluted with elution buffer (20 mM Tris, 500 mM NaCl, 500 mM imidazole, 10% glycerol, pH 7.5). The 3×NLS–*Sp*Cas9 (or LbCas12a protein) was dialysed overnight at 4 °C in 20 mM HEPES, 500 mM NaCl, 1 mM EDTA, 10% glycerol, pH 7.5. Subsequently, the protein was step dialysed from 500 mM NaCl to 200 mM NaCl (final dialysis buffer: 20 mM HEPES, 200 mM NaCl, 1 mM EDTA, 10% glycerol, pH 7.5). Next, the protein was purified by cation exchange chromatography (5 ml HiTrap-S column, buffer A: 20 mM HEPES pH 7.5, 1 mM TCEP; buffer B: 20 mM HEPES pH 7.5, 1 M NaCl, 1 mM TCEP; flow rate 5 ml/min, column volume 5 ml) followed by size-exclusion chromatography (SEC) on a Superdex-200 (16/60) column (isocratic size-exclusion running buffer: 20 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP for 3×NLS–*Sp*Cas9 or 20 mM HEPES pH 7.5, 300 mM NaCl, 1 mM TCEP for LbCpf1–2×NLS). The primary protein peak from the SEC was concentrated in an Ultra-15 Centrifugal Filters Ultracel-30K (Amicon) to a concentration around 100 μM based on absorbance at 280 nm. The purified protein quality was assessed by SDS–PAGE/Coomassie staining to be >95% pure and protein concentration was quantified with a Pierce BCA Protein Assay Kit (ThermoFisher Scientific). Protein was stored at −80 °C until further use.

In vitro transcription of guide RNAs. The DNA cassette containing the U6 promoter and the sgRNA framework for *Sp*Cas9 was cloned from pLKO1-puro vector[33] into pBluescript SK II+ backbone[47]. Plasmids expressing each guide RNA from the U6 promoter were constructed by annealing oligonucleotides encoding guide RNA and cloning it into BfuAI cleavage sites in this vector (Supplementary Table 4). Templates for in vitro transcription of *Sp*Cas9 guides were amplified from the cognate plasmids using NEB Q5 High-Fidelity DNA Polymerase for 30 cycles (98 °C, 15 s; 65 °C, 25 s; 72 °C, 20 s) using primer sets designed to include the T7 scaffold (Supplementary Table 5). To generate CRISPR RNA (crRNA) for LbCas12a, templates for in vitro transcription were generated by PCR amplification of oligonucleotides designed to include the T7 scaffold along with the guide RNA and a 15-mer overlap sequence to allow annealing between the oligos (Supplementary Table 8). The oligonucleotides encoded the full-length direct repeat crRNA sequence[47]. Thirty cycles of amplification were conducted using NEB Q5 High-Fidelity DNA polymerase (98 °C, 15 s; 60 °C, 25 s; 72 °C, 20 s). The PCR products were purified using Zymo DNA Clean & Concentrator Kit (Zymo D4005). In vitro transcription reactions were performed using the HiScribe T7 High Yield RNA Synthesis Kit using 300 ng of PCR product as template (NEB E2040S). After incubation for 16 h at 37 °C, samples were treated with DNase I for 40 min at 37 °C to remove any DNA contamination. Each guide RNA was purified using the Zymo RNA Clean and Concentrator Kit. Final RNA concentration was measured using Nanodrop and RNA was stored at −80 °C until further use.

Electroporation of cell lines with *Sp*Cas9 RNPs. 3×NLS–*Sp*Cas9 protein was precomplexed with sgRNAs either purchased from Synthego or made in-house by T7 transcription (Supplementary Table 6) and electroporated into cells using the Neon transfection system (ThermoFisher Scientific).

Electroporation of iPS cells. After washing with PBS, iPS cells were dissociated into single cells with 3:1 TrypLE:0.5 mM EDTA and neutralized with Ham's F10 and 20% FBS. To form RNP complexes, 20 pmol 3×NLS–*Sp*Cas9 protein and 25 pmol gRNA were combined in 10 μl Neon Buffer R and incubated for 10 min at room temperature. iPS cells (1×10^5) were resuspended in 10 μl RNP–Buffer R mix and then nucleofected as follows: pulse voltage 1,500 V, pulse width 20 ms, pulse number 1. After transfection, the cells were plated onto Matrigel-coated 24-well plates with iPS Brew XF supplemented with 10 μM Y27632 for expansion and grown in a humidified incubator at 37 °C, 5% CO_2, for 4 days before collecting them for analysis. iPS cell-derived myoblasts were electroporated using two pulses of 1,400 V and 20 ms width and plated onto a 24-well dish containing pre-warmed antibiotic-free human primary myoblast growth medium and cultured for 4–6 days before analysis.

Electroporation of HPS1 patient-derived B-LCL cells. In brief, 40 pmol of 3×NLS–*Sp*Cas9 protein was precomplexed with 50 pmol of sgRNA in buffer R for 10–20 min at room temperature in a final volume of 12 μl. Then, 300,000 cells per reaction were resuspended in 10 μl of RNP–buffer R mix and electroporated with 2 pulses at 1,700 V for 20 ms using the 10-μl tip. Cells were then plated in 24-well plates with 500 μl of pre-equilibrated antibiotic-free culture medium and grown in a humidified incubator at 37 °C and 5% CO_2 for 7 days before indel analysis.

For the PARP-1 inhibition experiments, 300,000 HPS1 patient-derived B-LCL cells were treated with 10 μM or 20 μM rucaparib camsylate (Sigma-Aldrich PZ0036) in standard growth medium for 24 h. Treated cells were electroporated with *Sp*Cas9 RNPs as described above. Following another 24 h incubation in rucaparib-containing medium, cells were resuspended in PARP-1 inhibitor-free medium and collected for analysis after 7 days.

Electroporation of HEK293T cells. In brief, 20 pmol of 3×NLS–*Sp*Cas9 protein and 25 pmol of in vitro transcribed sgRNA were pre-complexed in buffer R for 10–20 min at room temperature. Then, 100,000 cells per reaction were resuspended in 10 μl of RNP–buffer R mix and nucleofected with *Sp*Cas9 guide RNA complex using two pulses at 1,150 V for 20 ms using the 10-μl tip. Cells were then plated in 24-well plates with 500 μl of pre-equilibrated antibiotic-free culture medium and grown for 3 days before analysis. For Cas12a editing experiments at endogenous microduplications, 80 pmol of LbCas12a protein was pre-complexed with 100 pmol of in vitro transcribed crRNA and 100,000 cells per reaction were nucleofected as described above.

Indel analysis by TIDE. Genomic DNA was extracted from HEK293T cells using GenElute Mammalian Genomic DNA Miniprep Kit (Sigma Aldrich) according to the manufacturer's instructions. The DNA region containing the 24-bp microduplication was amplified using genomic DNA as template and primers listed in Supplementary Table 7 using NEB Q5 High-Fidelity DNA Polymerase (98 °C, 15 s; 67 °C, 25 s; 72 °C, 20 s) for 30 cycles. Subsequently, the PCR product was purified using the DNA Clean & Concentrator-5 kit (Zymo) and sequenced. Sanger sequencing trace data were analysed using the TIDE webtool at https://tide.nki.nl/ to infer the compositions of indels created at the sites of DSBs[34].

Library construction for Illumina deep sequencing. Library construction for deep sequencing was performed using a modified version of our previously described protocol[26]. In brief, iPS cells and myoblasts were collected following nuclease treatment and genomic DNA was extracted using the GenElute Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich G1N350). Genomic loci spanning the target sites were PCR amplified with locus-specific primers carrying tails complementary to the TruSeq adapters (Deepseq_TCAP_primer_fwd and Deepseq_TCAP_primer_rev; Supplementary Table 8). Subsequently, 50 ng of input genomic DNA was PCR amplified with Q5 High-Fidelity DNA Polymerase

(New England Biolabs): (98 °C, 15 s; 67 °C, 25 s; 72 °C, 20 s) for 30 cycles. Next, 0.1 μl of each PCR reaction was amplified with barcoded primers to reconstitute the TruSeq adaptors using Q5 High-Fidelity DNA Polymerase (New England Biolabs): (98 °C, 15 s; 67 °C, 25 s; 72 °C, 20 s) for 10 cycles. Products were qualitatively analysed by gel electrophoresis. Equal amounts of the products were pooled and gel-purified using QIAquick Gel Extraction Kit (Qiagen 28704). The purified library was deep sequenced using a paired-end 150-bp Illumina MiSeq run.

Illumina deep sequencing analysis. MiSeq data analysis was performed using Unix-based software tools. First, we used FastQC (version 0.11.3; http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to determine the quality of paired-end sequencing reads (R1 and R2 fastq files). Next, we used paired-end read merger (PEAR; version 0.9.8)[35] to pool raw paired-end reads and generate single merged high-quality full-length reads. Reads were then filtered according to quality via FASTQC[36] for a mean PHRED quality score above 30 and a minimum per base score above 24. After that, we used BWA (version 0.7.5) and SAMtools (version 0.1.19) to align each group of filtered reads to a corresponding reference sequence. To determine lesion type, frequency, size and distribution, all edited reads from each experimental replicate were combined and aligned, as described above. Lesion types and frequencies were then catalogued in a text output format at each base using bam-readcount. For each treatment group, the average background lesion frequencies (based on lesion type, position and frequency) of the triplicate negative control group were subtracted to obtain the nuclease-dependent lesion frequencies.

Library construction for UMI-based Illumina deep sequencing. The construction of the UMI-based library used a linear amplification step to incorporate UMIs within the amplicons from the target locus[15]. HPS1 B-LCL cells and HEK293T cells were collected following nuclease treatment for genomic DNA extraction using the GenElute Mammalian Genomic DNA Miniprep Kit (Sigma G1N350). Randomized unique molecular identifiers (UMIs) were incorporated within the 5′ locus-specific primers carrying tails complementary to TruSeq adaptors (Supplementary Table 8). In brief, 50 ng of input genomic DNA was linear amplified with NEB Q5 High-Fidelity DNA Polymerase (98 °C, 15 s; 67 °C, 25 s; 72 °C, 20 s) for 10 cycles using the 5′ locus-specific primer with TruSeq adaptor conjugated with a UMI sequence. Next a 5′ constant primer along with the 3′ locus-specific primer with TruSeq adaptor were added and further amplified for 30 cycles. Indexes were then incorporated using barcoded primers to diluted PCR products using NEB Q5 High-Fidelity DNA Polymerase (98 °C, 15 s; 67 °C, 25 s; 72 °C, 20 s) for 10 cycles. Products were qualitatively analysed by gel electrophoresis. Equal amounts of the products were pooled and gel-purified using QIAquick Gel Extraction Kit (Qiagen 28704) for DNA recovery. The purified library was deep sequenced using a paired-end 150-bp Illumina MiSeq run.

UMI-based deep sequencing analysis. We adapted the analysis of the UMI-tagged deep sequencing reads from our previous protocol[15]. Initially, BWA (version 0.7.5) and SAMtools (version 0.1.19) were used to align each group of filtered merged-read pairs to a corresponding reference sequence, ignoring the unique molecular barcodes. Next, we used a custom Python and PySAM script to process mapped reads into counts of UMI-labelled reads for each target. The mapped reads were filtered by requiring a mapping value (MAPQ) larger than 30. Alignments were categorized into different categories of indels using VarScan 2[37]. Next, we identified UMI duplicates and the minimal set of amplicons that can account for the full set of reads with unique UMIs. For each unique UMI, a minimum of five observations of the same sequence was required to consider the sequence to have a low likelihood of being an artefact (sequencing error in the UMI element). For sequences that met this threshold, all common sequences associated with the UMI were consolidated to one read for analysis of the distribution of sequence modifications that were present at a locus. The resulting UMI number tables, which describe the type of each sequence modification and its length, were concatenated and loaded into GraphPad Prism 7 for data visualization. Microsoft Excel version 16.21.1 was used for statistical analysis.

PacBio library preparation. Single molecule, real-time (SMRT) sequencing is modified from Pacific Biosciences (PacBio). Nuclease-treated patient-derived iPS cells were collected for genomic DNA extraction with GenElute Mammalian Genomic DNA Miniprep Kit (Sigma G1N350). In brief, regions that flanked the *TCAP* target site were PCR amplified using locus-specific primers (Supplementary Table 8). The forward primer was designed to have the barcode sequence followed by the UMI and locus-specific primer sequence. The reverse primer contains the barcode followed by the locus-specific primer sequence. Input DNA (25–50 ng) was PCR amplified with Phusion High Fidelity DNA Polymerase (New England Biolabs): (98 °C, 15 s; 65 °C, 25 s; 72 °C, 18 s) for 30 cycles. The products were qualitatively analysed by gel electrophoresis and subsequently gel purified with QIAquick Gel Extraction Kit (Qiagen 28704). The purified products were sent to the UMASS Medical School Deep Sequencing Core for SMRTbell Library Preparation and sequencing on the Pacific Biosciences Sequel Instrument.

PacBio sequencing data analysis. For PacBio sequencing data analysis, Minimap2 (version 2.14[38]) was used to align the raw Consensus_ROI (reads_of_insert.fastq) data to the 2-kb reference sequence. Alignment quality control and filtering were performed using custom Perl script to remove errors and filter out alignments with poor quality. For variation calling, a custom Python script was used to extract deletions or insertions larger than 5 bp for each read from the SAM files. Subsequently, deletions or insertions were classified into different groups on the basis of their length. IGV(version 2.4.16) was used for alignment visualization of the aligned reads using Quick consensus mode[39].

Clonal analysis of iPS cells. Following confirmation of MMEJ-mediated correction in the population of LGMD2G iPS cells, clonal analysis was performed. Cells from the corrected population were seeded into 96-well plates in the presence of Y27632 at a frequency of 0.8 cells per well. iPS cell clones were cultured for several weeks in iPS Brew XF (Miltenyi Biotec) before being collected for sequence analysis by deep-sequencing.

Myoblast differentiation and detection of telethonin expression. iPS cell-derived myoblasts were plated into 0.1% gelatin-coated 6-well plates at a density of 100,000 cells per well in myoblast expansion medium containing Ham's F10 (Cellgro) supplemented with 20% FBS (Hyclone, SH30071.03), 1.2 mM $CaCl_2$ (EMD OmniPur 3000) and 1% chick embryo extract isolated from day 12 SPF Premium Fertilized White Leghorn Chicken Eggs (Charles River). After 4 days of expansion, the cells were incubated with myotube differentiation medium including DMEM/F12 (ThermoFisher Scientific) supplemented with 1% N2 (ThermoFisher Scientific, 17502-048) and 1% insulin–transferrin–selenium (ThermoFisher Scientific, 41400045). After 10 days of differentiation, the cells were dissociated into single cells using TrypLE. Subsequently the cells were fixed with 2% PFA for 15 min and blocked with PBS including 2% BSA, 2% horse serum, 2% goat serum and 2% Triton X-100 for 20 min. The cells were then incubated with anti-telethonin antibody (Santa Cruz, sc-25327, 1:50) at 4 °C for 2 days and IgG goat anti-mouse secondary antibody labelled with Alexa 488 fluorophore (Invitrogen, A11017, 1:800) at room temperature for 1 h. The cells were suspended in flow buffer (PBS including 0.2% FBS) and flow cytometry was performed using a BD FACSAria IIu (UMMS Flow Cytometry Core Laboratory). Roughly 20,000 cells were included for analysis. FlowJo software (version 7.6) was used for data analysis.

Survey of microduplications in ClinVar and in human reference populations. Annotations of pathogenicity from ClinVar (ftp://ftp.ncbi.nlm.nih.gov/pub/clinvar/vcf_GRCh37/archive_2.0/2018/clinvar_20180128.vcf.gz)[20] were combined with annotations of allele-frequencies from gnomAD (https://console.cloud.google.com/storage/browser/gnomad-public/release/2.0.2/vcf)[21] and from the 1000 Genome Project (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/)[40] using the annotate function in bcftools[41] (1.9), after decomposition of multi-allelic sites and normalization of variants with vt[42] (v0.5772) against the reference genome http://www.broadinstitute.org/ftp/pub/seq/references/Homo_sapiens_assembly19.fasta. Most analyses were restricted to the intervals in ftp://ftp.broadinstitute.org/pub/ExAC_release/release1/resources/exome_calling_regions.v1.interval_list. Insertions were extracted using vt (view -h -f "VTYPE==INDEL&&DLEN>0"); then duplications were identified, repeat units counted, internal shift-symmetries determined, and flanking genomic regions extracted using a modified version of the vt function annotate_indels. Additional processing (filtering, finding maximal allele frequencies among different populations, scanning for PAM sites and so on) was performed using R (3.4.3), including the VariantAnnotation (1.24.5) package[43].

Exact tandem repeats in the reference genome were identified using the Tandem Repeats Finder program (4.09)[44] and checked for exact matches elsewhere in the genome with bwa fastmap (0.7.17)[45]. Examples of different lengths were manually selected to use for the tests of collapse of endogenous microduplications.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

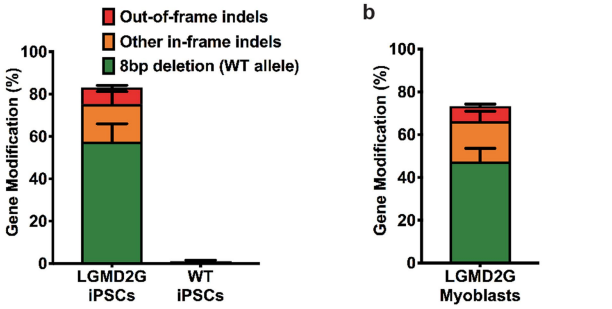
Data availability

Raw data associated with Extended Data Fig. 10 are reported in Supplementary Table 3. Raw script used for retrieving microduplication data listed in Supplementary Table 3 will be available upon request. Raw Illumina sequencing reads and PacBio data for this study have been deposited in the National Center for Biotechnology Information Short Read Archive under bioproject ID PRJNA517630.

Code availability

Data analysis used a combination of publicly available software and custom code, as detailed in the Methods. Custom Python (CRESA-lpp.py) and R (indel_background_filtering.R) scripts used in the Illumina data analysis and the shell script (Tcap_pacbio_analysis.sh) used for the analysis of the PacBio data are hosted on GitHub (https://github.com/locusliu/PCR_Amplicon_target_deep_seq). Scripts for the bioinformatic analysis of pathogenic microduplications are hosted at https://rambutan.umassmed.edu/duplications/.

31. Rittié, L. & Fisher, G. J. Isolation and culture of skin fibroblasts. *Methods Mol. Med*. **117**, 83–98 (2005).
32. Stadler, G. et al. Establishment of clonal myogenic cell lines from severely affected dystrophic muscles — CDK4 maintains the myogenic population. *Skelet. Muscle* **1**, 12 (2011).
33. Kearns, N. A. et al. Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. *Development* **141**, 219–223 (2014).
34. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res*. **42**, e168 (2014).
35. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).
36. Blankenberg, D. et al. Manipulation of FASTQ data with Galaxy. *Bioinformatics* **26**, 1783–1785 (2010).
37. Koboldt, D. C. et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. **22**, 568–576 (2012).
38. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
39. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol*. **29**, 24–26 (2011).
40. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
41. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
42. Tan, A., Abecasis, G. R. & Kang, H. M. Unified representation of genetic variants. *Bioinformatics* **31**, 2202–2204 (2015).
43. Obenchain, V. et al. VariantAnnotation: a Bioconductor package for exploration and annotation of genetic variants. *Bioinformatics* **30**, 2076–2078 (2014).
44. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. **27**, 573–580 (1999).
45. Li, H. Exploring single-sample SNP and INDEL calling with whole-genome de novo assembly. *Bioinformatics* **28**, 1838–1844 (2012).
46. Wu, Y. et al. Highly efficient therapeutic gene editing of human hematopoietic stem cells. *Nat. Med*. https://doi.org/10.1038/s41591-019-0401-y (2019).
47. Liu, P. et al. Enhanced Cas12a editing in mammalian cells and zebrafish. *Nucleic Acids Res*. https://doi.org/10.1093/nar/gkz184 (2019).



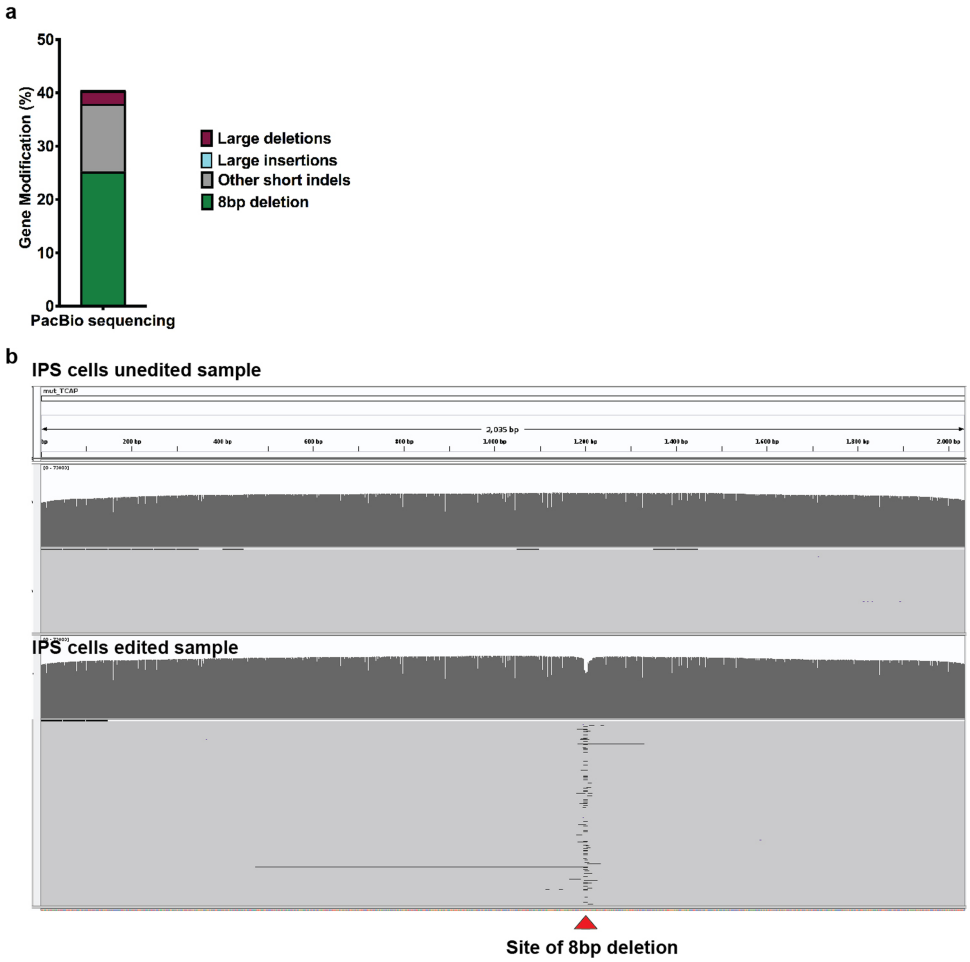
c

Indel observed	Sequence observed	Alleles present (%)
MH mediated del	AGCTGAGCTG--------CGAGGTGTCGGAGGAGAACTGTGA	67.70
del	AGCTGAGCTGCGAG--GTCGAGGTGTCGGAGGAGAACTGTGA	7.63
del	AGCTGAGCTGCGAGGT----------GAACTGTGA	6.09
Unedited	AGCTGAGCTGCGAGGTGTCGAGGTGTCGGAGGAGAACTGTGA	3.92
ins	AGCTGAGCTGCGAGGTGTtCGAGGTGTCGGAGGAGAACTGTGA	3.92
ins	AGCTGAGCTGCGAGGTGTgtCGAGGTGTCGGAGGAGAACTGTGA	3.75
del	AGCTGAGCTGCGAGGT-TCGAGGTGTCGGAGGAGAACTGTGA	2.86
del	AGCTGAGCTGCGAG---TCGAGGTGTCGGAGGAGAACTGTGA	1.27
del	AGCTGAGCTGCGAGGTGT-------TCGGAGGAGAACTGTGA	1.15
del	AGCTGAGCTGC------TCGAGGTGTCGGAGGAGAACTGTGA	0.98
del	AGCTGAGCTGCGA----------GGAGGAGAACTGTGA	0.74

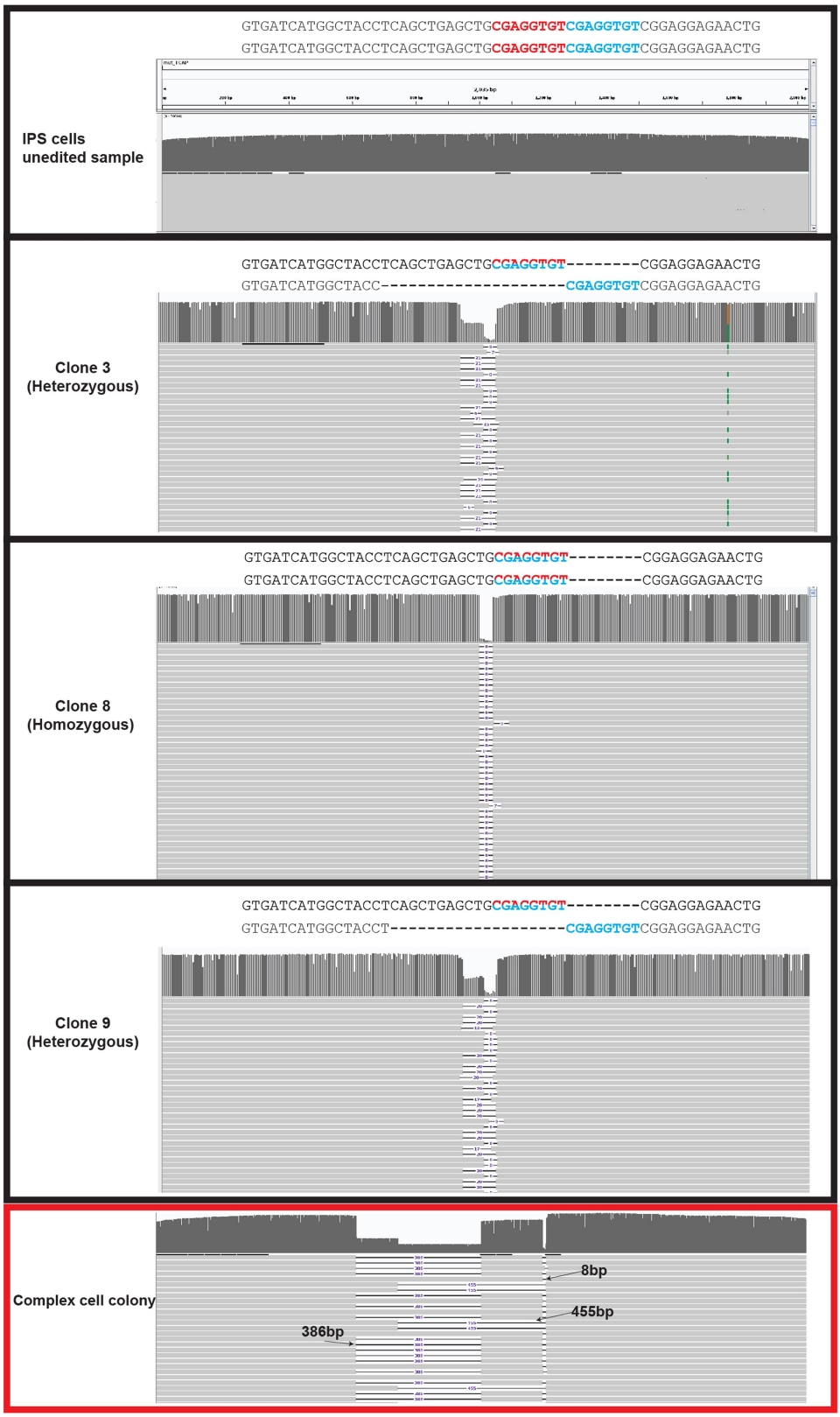
d

Indel observed	Sequence observed	Alleles present (%)
MH mediated del	AGCTGAGCTG--------CGAGGTGTCGGAGGAGAACTGTGA	45.11
Unedited	AGCTGAGCTGCGAGGTGTCGAGGTGTCGGAGGAGAACTGTGA	26.17
del	AGCTGAGCTGCGAGG--TCGAGGTGTCGGAGGAGAACTGTGA	7.05
ins	AGCTGAGCTGCGAGGTGTtCGAGGTGTCGGAGGAGAACTGTGA	6.23
del	AGCTGAGCTGCGAGGT----------GAACTGTGA	3.04
del	AGCTGAGCTGCGAGGTGT-------TCGGAGGAGAACTGTGA	2.67
ins	AGCTGAGCTGCGAGGTGTgtCGAGGTGTCGGAGGAGAACTGTGA	2.46
del	AGCTGAGCTGCGAGGT-TCGAGGTGTCGGAGGAGAACTGTGA	2.41
del	AGCTGAGCTGC------TCGAGGTGTCGGAGGAGAACTGTGA	2.34
del	AGCTGAGCTGCGA---GTCGAGGTGTCGGAGGAGAACTGTGA	1.70
del	AGCTGAGCTGCGA----------GGAGGAGAACTGTGA	0.80

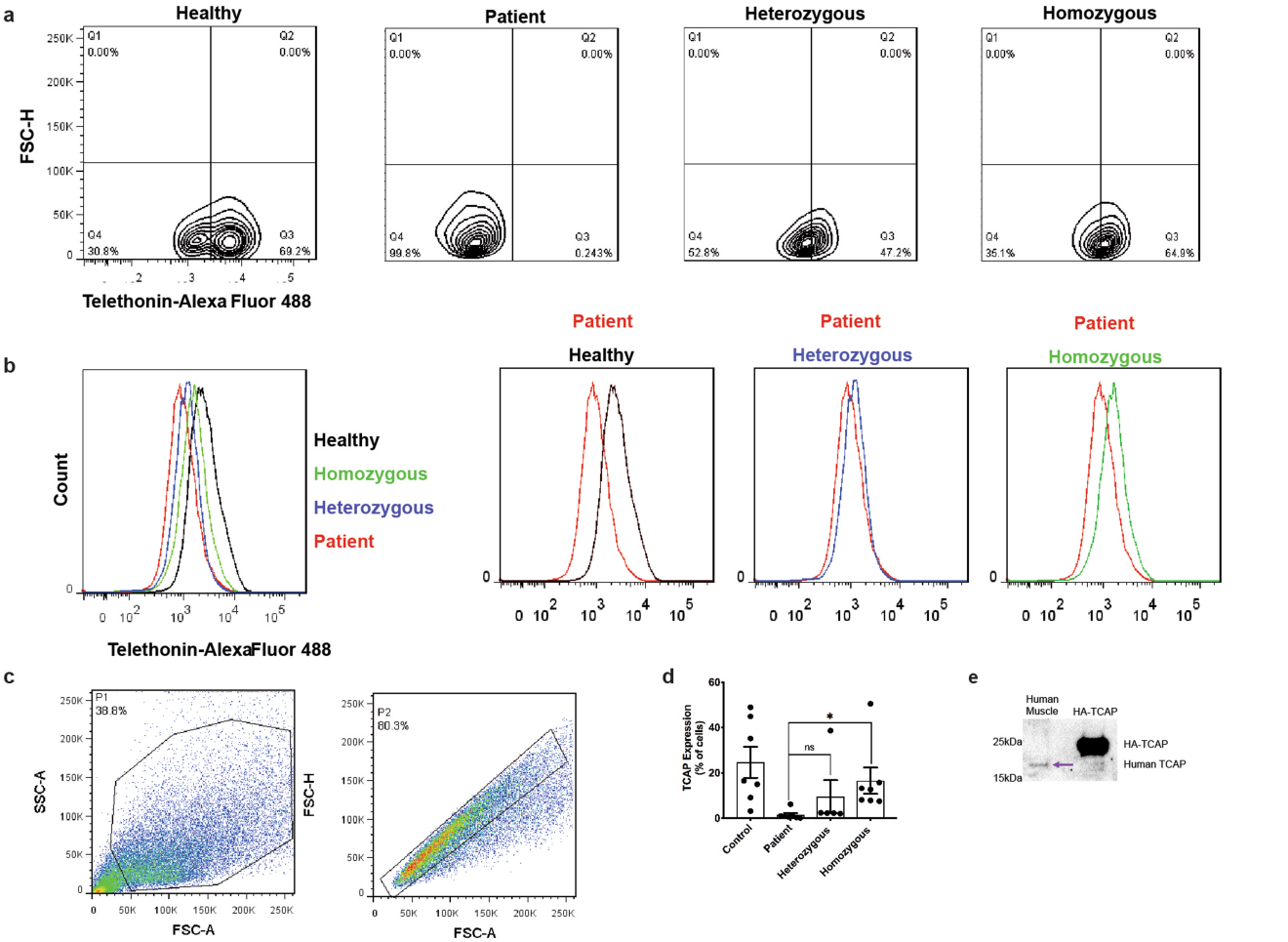
Extended Data Fig. 1 | Indel populations resulting from *Sp*Cas9 editing at the *TCAP* locus. a, Indel percentages resulting from *Sp*Cas9 RNP treatment in patient-derived iPS cells homozygous for the 8-bp microduplication or in wild-type iPS cells. Mean ± s.e.m. from three biological replicates. **b**, Breakdown of indel classes resulting from *Sp*Cas9 treatment of myoblasts derived from patient-derived LGMD2G iPS cells. Mean ± s.e.m. from three biological replicates. **c**, Sequence alignment of the edited alleles resulting from *Sp*Cas9 RNP treatment of LGMD2G iPS cells. Red and blue text indicates DNA repeats that constitute the microduplication, and collapse is indicated by half red and half blue text. Dashes indicate deleted bases and purple text indicates inserted bases. Data are from one biological replicate out of three independent biological replicates. **d**, Sequence alignment of the edited alleles resulting from *Sp*Cas9 RNP treatment of myoblasts derived from patient-derived LGMD2G iPS cells. Data are from one biological replicate out of three independent biological replicates.



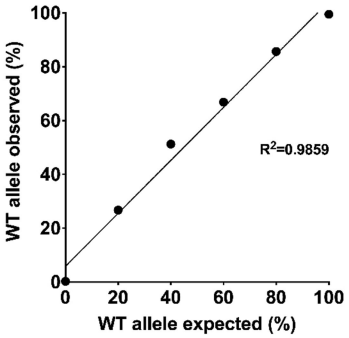
Extended Data Fig. 2 | PacBio long-read sequencing analysis for *Sp*Cas9-edited LGMD2G iPS cells at the *TCAP* locus. a, Percentage of gene modification observed from PacBio sequencing (one replicate from Fig. 1c out of three biological replicates). Green, alleles containing the 8-bp deletion; grey, other small indels($\leq$100 bp); blue, large insertions (0.14%, not visible on the graph); maroon, large deletions ($>$100 bp). **b**, IGV graphs depicting representative reads obtained for unedited (top) and edited (bottom) LGMD2G iPS cells, spanning a genomic region of about 2,035 bp surrounding the *TCAP* target site. Red caret indicates the 8-bp deletion site. Data represent one replicate out of three independent biological replicates.



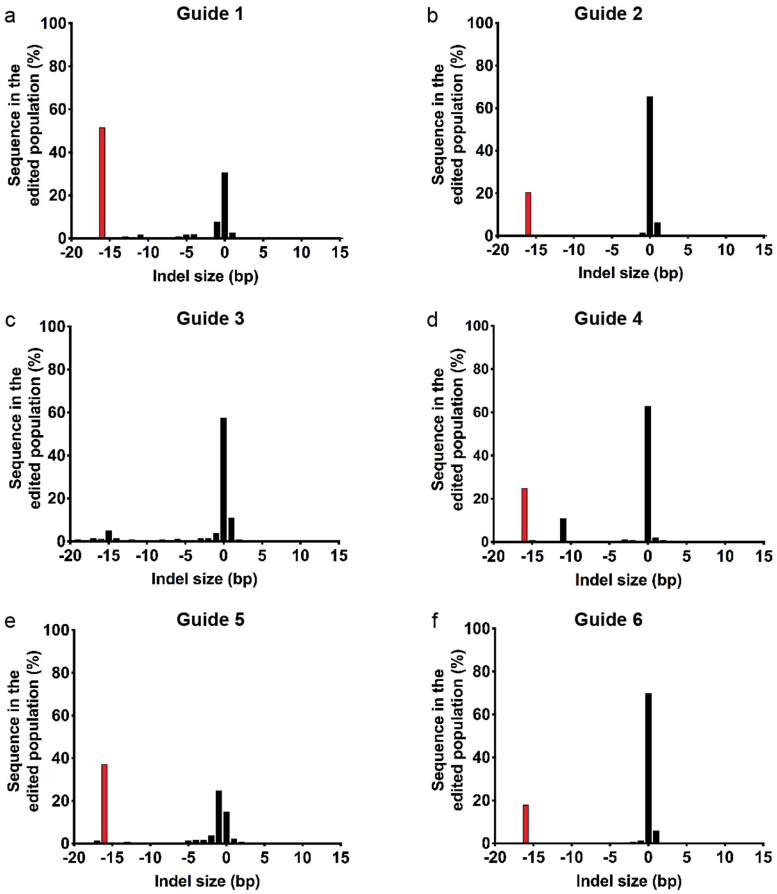
Extended Data Fig. 3 | PacBio long-read sequencing analysis of *Sp*Cas9-edited LGMD2G iPS cells clones and a complex colony at the *TCAP* locus. IGV graphs depicting representative reads obtained for clonal isolates of edited LGMD2G iPS cells (Fig. 1d), spanning a genomic region of about 2,035 bp surrounding the *TCAP* target site. The genotype of the clones (deduced by Illumina deep sequencing) is indicated beside an enlargement of the *TCAP* target region within the PacBio data. The sequences of the two alleles (listed above the IGV plot) obtained from sequencing are shown with repeats in red and blue. Alleles that reverted to wild-type as a result of collapse of microduplication are half red/half blue. Bottom, IGV plot for one complex iPS cell colony that appears to have been nucleated by more than one cell, with large deletions present in the genome (sizes indicated).



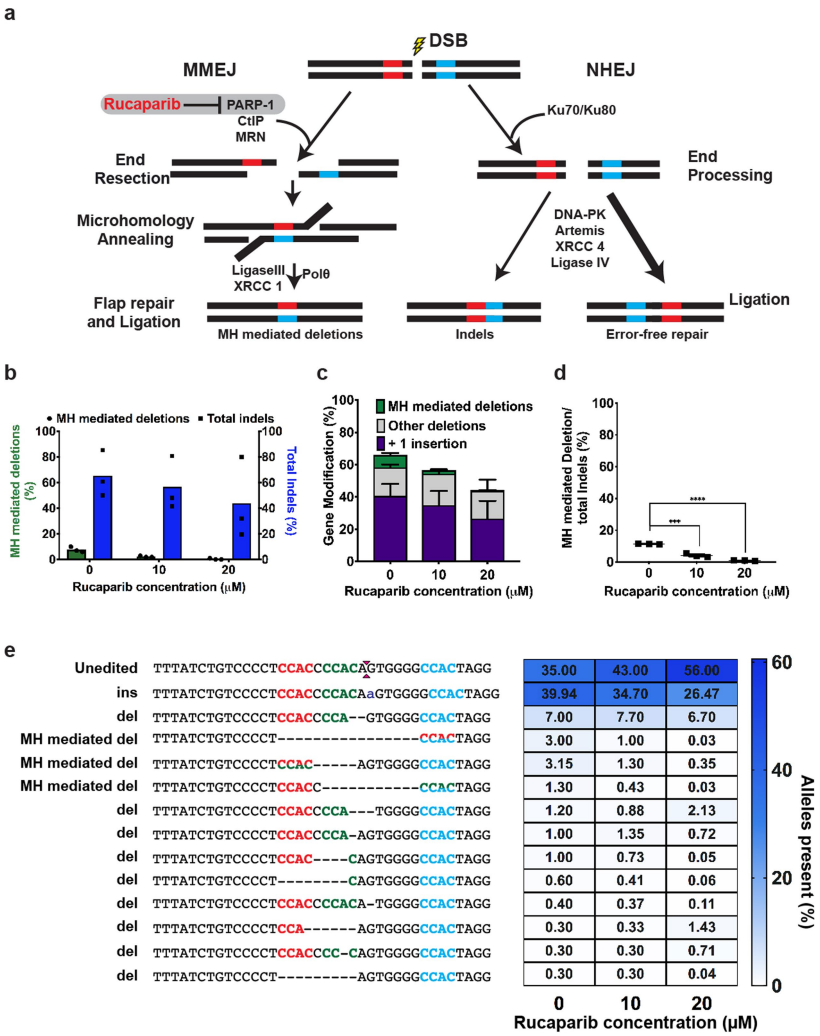
Extended Data Fig. 4 | Detection of telethonin expression by flow cytometry in patient-derived cells treated with *Sp*Cas9. a, Contour plots from a representative flow cytometry assay to detect telethonin expression in healthy control cells ($TCAP^{+/+}$), patient cells ($TCAP^{-/-}$), and *Sp*Cas9-treated homozygous and heterozygous iPS clone-derived myoblasts differentiated for 10 days in culture. Plots are representative of three independent replicates. **b**, Histograms from a representative flow cytometry assay to detect telethonin expression. Left, overlay of anti-telethonin antibody staining for four representative samples for different *TCAP* genotypes. Right, comparison between patient cells and healthy control cells, and *Sp*Cas9-treated homozygous and heterozygous iPS clone-derived myoblasts differentiated for 10 days in culture. Histograms are representative of three independent replicates. **c**, Cells were selected by removing cell debris first as shown by gate P1, and then single cells were selected from P1 by removing clustered cells as shown by gate P2. The cells in gate P2 were used for flow analysis. Plots are representative of one biological replicate. **d**, Average percentage of telethonin-expressing cells from two technical replicates of three biological replicates. Error bars indicate s.e.m ($n = 6$) and circles represent individual data points. *P* values ($P = 0.33$ for patient versus heterozygous and $^*P = 0.04$ for patient versus homozygous clones) were calculated by two-sided Student's *t*-test (Supplementary Table 9). ns, not significant. **e**, Western blot showing validation of anti-telethonin antibody (Santa Cruz Biotechnology). Human muscle lysate and lysate from HEK293T cells transfected with haemagglutinin-tagged telethonin expression construct were separated on an SDS 4–12% acrylamide gradient gel and the resulting blot was probed with anti-telethonin antibody. For gel source data, see Supplementary Fig. 1.



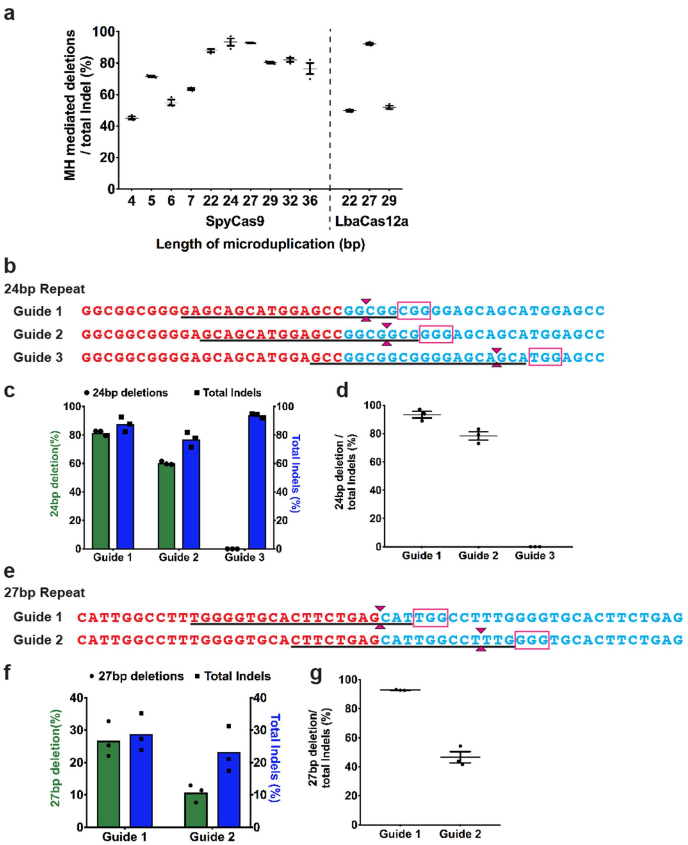
Extended Data Fig. 5 | Standard curve generated with genomic DNA of wild-type and *HPS1* mutant B-LCLs from UMI-based Illumina deep sequencing. Genomic DNA from wild-type cells and *HPS1* cells homozygous for the 16-bp microduplication were mixed at different ratios (*x* axis). These mixed DNAs were used for the construction of a UMI-based Illumina library to determine the ratio of the alleles through deep sequencing (*y* axis). These data are fitted to a regression line with the R^2 value reported. $n = 1$ biological replicate.



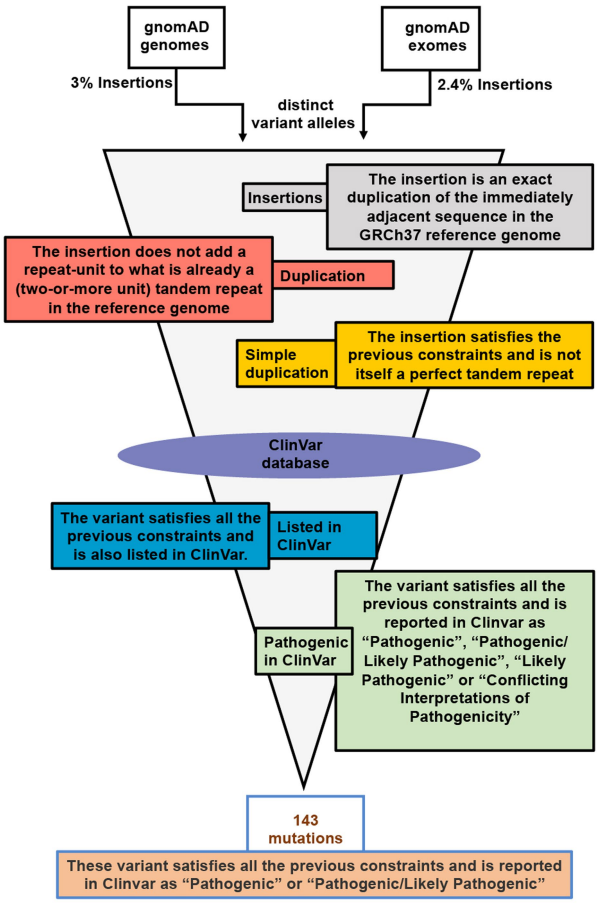
Extended Data Fig. 6 | Indel spectrum generated by *Sp*Cas9 editing at the *HPS1* locus in HPS1 B-LCL cells. Indel spectra of *Sp*Cas9 nuclease cells treated with different sgRNAs determined by UMI-based Illumina deep sequencing. **a**, Target site 1. **b**, Target site 2. **c**, Target site 3. **d**, Target site 4. **e**, Target site 5. **f**, Target site 6. Red bar indicates 16-bp deletion that corresponds to the deletion of one of the microduplication repeats. Data show indel spectra from one representative biological replicate out of three independent biological replicates.



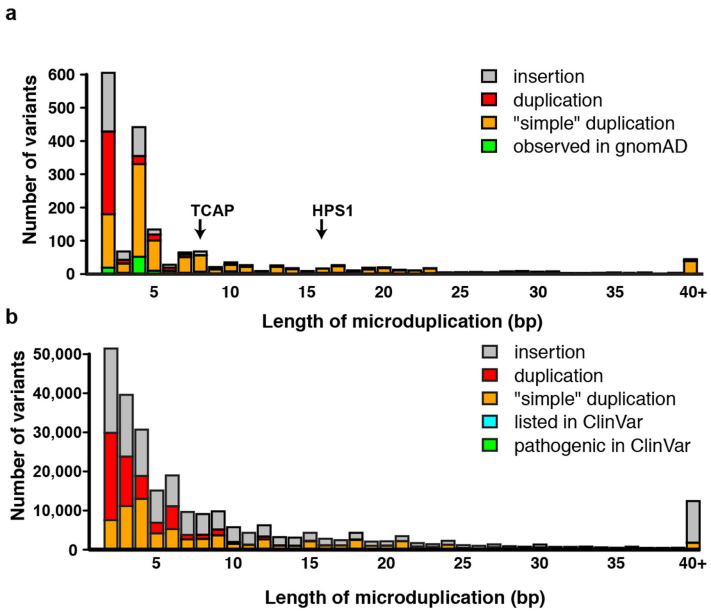
Extended Data Fig. 7 | Effect of rucaparib on the profile of microhomology-mediated deletion products at *AAVS1* locus in patient-derived HPS1 B-LCL cells. a, Schematic of two prominent DNA DSB repair pathways. A DSB can be repaired through various pathways that produce different DNA sequence end products. The NHEJ pathway is the dominant DSB repair pathway in most cells. The MMEJ pathway uses end-resection to discover small homologies on each side of the break that can be used to template the fusion of the broken ends. PARP-1 regulates DSB flux through the MMEJ pathway. Treatment of cells with rucaparib—an inhibitor of PARP-1—attenuates DSB flux down the MMEJ repair pathway. **b**, Percentage of microhomology-mediated deletions (green) and total indels (blue) resulting from *Sp*Cas9 treatment of cells in the presence of 0,10 and 20 μM rucaparib. Bars show mean and dots show individual data points from three biological replicates based on UMI-based Illumina deep sequencing. **c**, Percentage of 1-bp insertions (purple), microhomology mediated deletions (green) and other deletions (grey) produced by *Sp*Cas9 RNP with a sgRNA targeting the *AAVS1* locus with the addition of increasing amounts of rucaparib. Mean ± s.e.m. from three biological replicates based on UMI-based Illumina deep sequencing. **d**, Percentage of microhomology-mediated deletions out of total indels in cells treated with *Sp*Cas9 in the presence of rucaparib. Mean ± s.e.m., dots represent individual data points from three biological replicates. *P* values determined using two-tailed unpaired *t*-test (Supplementary Table 9). $^{***}P = 0.0004$, $^{****}P = 6.5 \times 10^{-7}$. **e**, Left, alignment of allele sequences obtained from deep sequencing analysis from samples treated with *Sp*Cas9 RNP in the presence of different rucaparib concentrations. Microhomologies present at the *AAVS1* locus are shown in by red, green and blue. Microhomology-mediated deletion is indicated by two-toned text. Magenta carets indicate site of DSB created by *Sp*Cas9. Inserted bases (ins) are shown in purple, deleted bases (del) are shown as black dashes. Right, heat map depicting the percentage of alleles generated after *Sp*Cas9 treatment of cells in the presence of different concentrations of rucaparib (0, 10 or 20 μM). The blue colour gradient scale indicates the percentage of occurrence of that sequence. Heat map represents mean values from a total of three independent biological replicates.



Extended Data Fig. 8 | Editing with *Sp*Cas9 and LbCas12a at endogenous microduplications. a, Percentage of microhomology-mediated deletions out of total indels at endogenous sites in cells treated with *Sp*Cas9 and LbCas12a. Mean ± s.e.m., dots represent individual data points from three biological replicates. **b**, Schematic of endogenous site containing a 24-bp microduplication for *Sp*Cas9 target sites 1–3. The 24-bp microduplication repeats are shown in bold red and blue. The PAM sequence is outlined in magenta and the protospacer sequence is underlined. Magenta carets indicate the site of DSB. **c**, Percentage of alleles with 24-bp deletion (green) and total indels (blue) for all three guides from TIDE analysis. Guide 3 produces primarily 23-bp deletions, but not 24-bp deletions, probably because it recuts the collapsed DNA sequence. Bars shows the mean from $n = 3$ biological repeats, individual data points are represented by dots. **d**, Proportion of the 24-bp deletion out of total indels as individual data points (dots), with mean ± s.e.m. $n = 3$ biological repeats. **e**, Schematic of endogenous site containing a 27-bp microduplication for *Sp*Cas9 target sites 1 and 2. **f**, Percentage of alleles with 27-bp deletion (green) and total indels (blue) for both guides from UMI-based Illumina deep sequencing. Bars show the mean from $n = 3$ biological repeats, individual data points are represented by dots. **g**, Proportion of the 27-bp deletion out of total indels as individual data points (dots) with mean ± s.e.m. $n = 3$ biological replicates.



Extended Data Fig. 9 | Bioinformatics pipeline for identification of disease alleles. Schematic shows the bioinformatics pipeline used to identify all microduplications amendable to efficient MMEJ-mediated collapse from the 'coding' regions (exome_calling_regions.v1; mainly exons plus 50 flanking bases) in the gnomAD genome and exome databases (version 2.0.2). Insertion variants observed in both databases were used for analysis (variants occurring in both databases were counted once). Insertions that do not add a repeat unit to an existing tandem repeat and are not themselves a perfect repeat were filtered to constrain only duplications that spanned 2–40 bp in length and are amendable to CRISPR–Cas9 targeting. This dataset was then cross-referenced against the ClinVar database (clinvar_20180225.vcf) to apply further filters for variants reported as pathogenic, which ultimately yielded 143 likely disease-causing microduplications.



Extended Data Fig. 10 | Pathogenic microduplications and their prevalence in human populations. a, Number of insertion variants of length >1 bp that are annotated as pathogenic or pathogenic/likely pathogenic in ClinVar. Variants are binned by length, with all those of length 40 bp or greater combined. The insertions (grey) are stratified into progressively finer categories: duplications (red); 'simple' duplications (described in text, orange); and the subset of these observed at least once in gnomAD exome/genome databases (green). **b**, Number of insertion variants of length >1 bp that are observed at least once in the 'coding' regions of the gnomAD exome/genome databases. As above, insertions (grey) are stratified into progressively finer categories: duplications (red); 'simple' duplications (orange); the subset of these listed in ClinVar (cyan); and the subset annotated as Pathogenic or Pathogenic/likely pathogenic in ClinVar (green). Cyan and green bars are not visible at this resolution.

nature research

Corresponding author(s): Scot Wolfe
Charles Emerson

Last updated by author(s): Feb 5, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Standard MiSeq data analysis used publicly available UNIX software described in the Methods section: FastQC v0.11.3; PEAR v0.9.8; BWA (version 0.7.5); SAMtools (version 0.1.19), and R (3.4.3). UMI-based MiSeq analysis used a custom Python and PySAM (v0.15.2) script to process mapped reads into counts of UMI-labeled reads for each target. Alignments were categorized into different categories of indels using VarScan v2.3.8. PacBio sequencing data analysis used Minimap2 (version 2.14). A custom python script was used to extract deletions or insertions larger than 5bp for each read from the SAM files. Data was visualized in IGV(version 2.4.16). Analysis of microduplications among catalogued variants used vt (v0.5772), and identification of endogenous microduplications used Tandem Repeats Finder (4.09) and bwa fastmap (0.7.17).
Data analysis	Sanger sequencing data analysis used publicly available TIDE webtool at https://tide.nki.nl/. Prizm 7.0 was used for data visualization. Microsoft Excel 16.21.1 was used to compute means, SEM and P-values (two tailed t test). Flow cytometry data was analyzed using FlowJo (version 7.6).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Illumina Sequencing data and PacBio data have been submitted to the Sequence Read Archive (Bioproject ID# PRJNA517630). Details on statistical calculations are found in Supplementary Table 9. The authors declare that all other data supporting the findings of this study are available within the paper and its supplementary

information files.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Three biological replicates were used in all the experiments. This number of replicates is necessary to calculate significance of each data point (standard deviation). No other statistical methods were used to determine sample size.
Data exclusions	No data point was excluded
Replication	Biological replicates confirm the reproducibility of our experimental findings and standard errors were within range of experimental error.
Randomization	No randomization was used as the data presented does not require randomization.
Blinding	The data presented does not require the use of blinding.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Anti-Telethonin antibody (Telethonin (G-11), Santa Cruz Biotechnology, sc-25327, 1:50)
Validation	The antibody was validated using western blots on human tissue expressing telethonin and in cell culture overexpressing epitope-tagged telethonin from a plasmid. The data is included in extended figure 4.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HPS1 B-LCL was obtained from Coriell institute, iPSC and myoblasts were derived from primary LGMD2G primary dermal fibroblasts isolated from the skin biopsy of human subject RB25199. Human Embryonic Kidney (HEK293T) cell line was a gift from our collaborator M. Green (at UMass Medical School, Worcester, Massachusetts, USA)
Authentication	Yes. HEK293T cells authenticated by University of Arizona Genetics Core.
Mycoplasma contamination	HEK293T cells were tested for mycoplasma contamination at regular intervals
Commonly misidentified lines (See ICLAC register)	None of the cell lines used in this study is categorized as a commonly misindentified cell line.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Cells for reprogramming TCAP iPSC lines were recovered from a skin biopsy of a consented LGMD 2G subject under a UMMS-IRB approved protocol and assigned a de-identified ID number unlinked to the subject's medical record. Consenting includes conditions for sharing de-identified samples and information with other investigators. No PHI will be shared at any time per HIPAA guidelines.
Recruitment	Subject agreed to undergo skin biopsy and provide a blood sample for genomic DNA analysis.
Ethics oversight	UMass Medical School Institutional Review Board

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Myotubes were differentiated from LGMD2G iPSCs. Cells were trypsined using TrpLE and fixed using PFA.
Instrument	BD FACSAria IIu
Software	Flowjo Software
Cell population abundance	N/A
Gating strategy	P1-FSC/SSC gates were applied to exclude cellular debris. P2 FSC-A/FSC-H gates were applied to isolate singlets. To detect protein expression, TCAP(+/+) and TCAP(-/-) derived samples were used as positive and negative controls respectively to gate for Alexa 488 stained cell population.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.