

1   **Design of efficacious somatic cell genome editing strategies for recessive and polygenic diseases**  
2   Jared Carlson-Stevermer<sup>1,2\*</sup>, Amritava Das<sup>1,3\*</sup>, Amr A. Abdeen<sup>1</sup>, David Fiflis<sup>1,2</sup>, Benjamin Isaac Grindel<sup>1,2</sup>,  
3   Tugce Akcan<sup>4</sup>, Tausif Alam<sup>4</sup>, Heidi Kletzein<sup>2</sup>, Lucille Kohlenberg<sup>1</sup>, Madelyn Goedland<sup>1,2</sup>, Micah J.  
4   Dombroe<sup>1</sup>, Krishanu Saha<sup>1,2,5</sup>

5   <sup>1</sup> Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, WI, USA

6   <sup>2</sup> Department of Biomedical Engineering, University of Wisconsin-Madison, Madison, WI, USA

7   <sup>3</sup> Morgridge Institute for Research, Madison, WI, USA

8   <sup>4</sup> Department of Surgery, University of Wisconsin, Madison WI, USA

9   <sup>5</sup> Retina Research Foundation Kathryn and Latimer Murfee Chair, Madison WI, USA

10   \* indicates equal contribution

11  
12   **Correspondence:** Dr. Krishanu Saha, Department of Biomedical Engineering and Wisconsin Institute for  
13   Discovery, University of Wisconsin-Madison, Madison, WI 53715, USA; **E-mail:** ksaha@wisc.edu

14   **Subject Terms:** CRISPR-Cas9, somatic cell genome editing, homology-directed repair, human pluripotent  
15   stem cells, computational modeling, biomedical engineering, scarless editing, non-viral delivery, glycogen  
16   storage disorder, autosomal recessive, iPSCs, ERT, gene therapy

17  
18   **ABSTRACT**  
19   **Gene correction of multiple alleles for compound heterozygous recessive or polygenic diseases is a**  
20   **promising therapeutic strategy. However, the targeting of multiple alleles using genome editors in a**  
21   **single cell could lead to mixed genotypes and adverse events that amplify during tissue**  
22   **morphogenesis. Here we demonstrate that SpyCas9-based S1mplex genome editors can be designed**  
23   **and developed to correct two distinct mutant alleles within a single human cell precisely. Gene-**  
24   **corrected cells in a patient-derived, induced pluripotent stem cell (iPSC) model of Pompe disease**  
25   **robustly expressed the corrected transcript from both corrected alleles. The translated protein from**  
26   **the gene-corrected cells was properly processed after translation and was able to enzymatically cross-**

27 **correct diseased cells at levels equivalent to standard-of-care, enzyme replacement therapy (ERT).**  
28 **Using a novel *in silico* model for the *in vivo* delivery of these and many other genome editors into a**  
29 **developing liver of a human infant, we identify progenitor cell targeting, delivery efficiencies, and**  
30 **suppression of imprecise editing outcomes at the on-target site as key design parameters controlling**  
31 **the potency and efficacy of *in vivo* somatic cell genome editing. Both single and double gene correction**  
32 **are efficacious for *in vivo* somatic cell editing strategies, while double gene correction is more effective**  
33 **than single-gene correction for autologous cell therapy with *ex vivo* gene-corrected cells.** This work  
34 **establishes that precise gene correction using genome editors to correct multiple distinct gene**  
35 **variants could be efficacious in the treatment of recessive and polygenic disorders.**

36

37 **INTRODUCTION**

38 Gene therapies typically involve the editing of a single allele<sup>1</sup>, or delivery of exogenous genetic  
39 material (through nucleic acid delivery, viruses, or *ex vivo* engineered cells) to over-express the gene of  
40 interest or suppress translation of the defective allele<sup>1</sup>. The development of these strategies traditionally  
41 starts with studies in animal models, however such studies are frequently insufficient to test genetic  
42 disorders that are polygenic – involving different mutations in different alleles that exacerbate a diseased  
43 phenotype. Animal models for polygenic diseases are extremely challenging to generate, and the number  
44 of mutations implicated in a particular disease precludes the generation of animal models for every mutation  
45 (e.g., >200 for *GAA*, the causative gene for Pompe disease<sup>2</sup>). Animal models for genetic disorders have  
46 different genomes from that of humans and thus interrogating human gene therapy in animal backgrounds  
47 may leave some questions unanswered, especially those regarding off-target effects of editing strategies.  
48 Therefore, as delivery systems improve<sup>3,4</sup> and genome editors become more precise<sup>5–17</sup>, new platforms and  
49 approaches (**Fig. 1a**, see page 6) are needed to fully understand the genotypic and phenotypic implications  
50 of editing multiple alleles within a person or a human cell.

51 Genome editors are routinely evaluated against a single allele within a pool of cells. Such studies  
52 have produced robust methods to understand the genomic changes, downstream gene expression, and

53 phenotypic changes from editing a single allele, as has been demonstrated in several prior studies<sup>18–23</sup>. Some  
54 of these studies, however, reveal that unintended genomic deletions and translocations are potential  
55 outcomes at the on-target allele<sup>24</sup>, and such unintended outcomes are predicted to be exacerbated as when  
56 making multiple cuts in the genome through the delivery of two or more different genome editors<sup>25</sup>. In  
57 addition, the adeno-associated viral vectors (AAVs), commonly used to deliver genome editors, can  
58 integrate into 30% of targeted alleles<sup>26</sup>. Finally, clonal analysis of genome editing outcomes demonstrates  
59 that precise editing of a single mutant allele can generate unintended mutations in other alleles<sup>27</sup>. Thus, the  
60 genomic integrity and allelic composition of cells could be variable when attempting to edit multiple alleles,  
61 leading to variable expression of multiple alleles with a single cell. To date, [precise correction of multiple](#)  
62 [alleles and an associated phenotypic rescue](#) has not been demonstrated within a single human cell.

63 Here we determine the [phenotypic consequences of correcting single or multiple pathogenic](#)  
64 [mutations within a single patient-derived cell](#). We focus on infantile-onset Pompe disease, an autosomal  
65 recessive glycogen storage disorder caused by multiple mutations in the acid- $\alpha$ -glucosidase (*GAA*) gene.  
66 *GAA* encodes an enzyme that breaks down glycogen within the lysosome<sup>28</sup> (**Fig. 1b**, see page 6). Over 400  
67 different *GAA* mutations have been noted within ClinVar, and detailed case studies indicate a buildup of  
68 glycogen, leading to clinical complications, most prominently in cardiac and muscle tissues<sup>2</sup>. Left untreated,  
69 patients with infantile-onset Pompe disease typically die within the first year of life, and Pompe disease is  
70 now frequently included within newborn screening panels<sup>29</sup>. While enzyme replacement therapy (ERT)  
71 using recombinant human GAA (rhGAA) and other gene and cell therapies are in development for Pompe  
72 disease<sup>30–35</sup>, none of these approaches retain endogenous *GAA* regulation nor have corrected the underlying  
73 *GAA* mutations (**Supplementary Information**). However, once some of the consequences of gene  
74 correction of multiple alleles within a single cell are characterized, important questions remain regarding  
75 how to design translational studies with gene correction strategies, both for *in vivo* somatic gene editing  
76 strategy or for *ex vivo* for cell therapy with autologous gene-corrected cells.

77 Patient-derived induced pluripotent stem cells (iPSCs) harbor the exact mutations to be targeted by  
78 editing strategies and can recapitulate some aspects of cell and tissue pathology within affected patients<sup>36–</sup>

79       <sup>39</sup>. Mathematical and computational tools can provide insight on somatic processes at scales larger than  
80    cells and small tissue constructs<sup>40</sup>, such as genome editor delivery, tissue morphogenesis, and physiological  
81    responses (**Fig. 1a**, see page 6). Many genetic diseases do not require the correction of *all* diseased cells:  
82    for example, 44% of a live donor pancreas<sup>41</sup> could treat a patient with type 1 diabetes (with event-free  
83    survival of both the donor and the recipient<sup>42</sup>), and targeted correction of <1% of tissue progenitors may be  
84    sufficient to address various diseases<sup>21,22</sup>. If precise editing of multiple alleles is feasible within a single  
85    cell, then a combined *in vitro*- *in silico* approach (**Fig. 1a**, see page 6) could help answer open translational  
86    questions regarding the efficacy of allele targeting and genome editor delivery/dosing strategies.

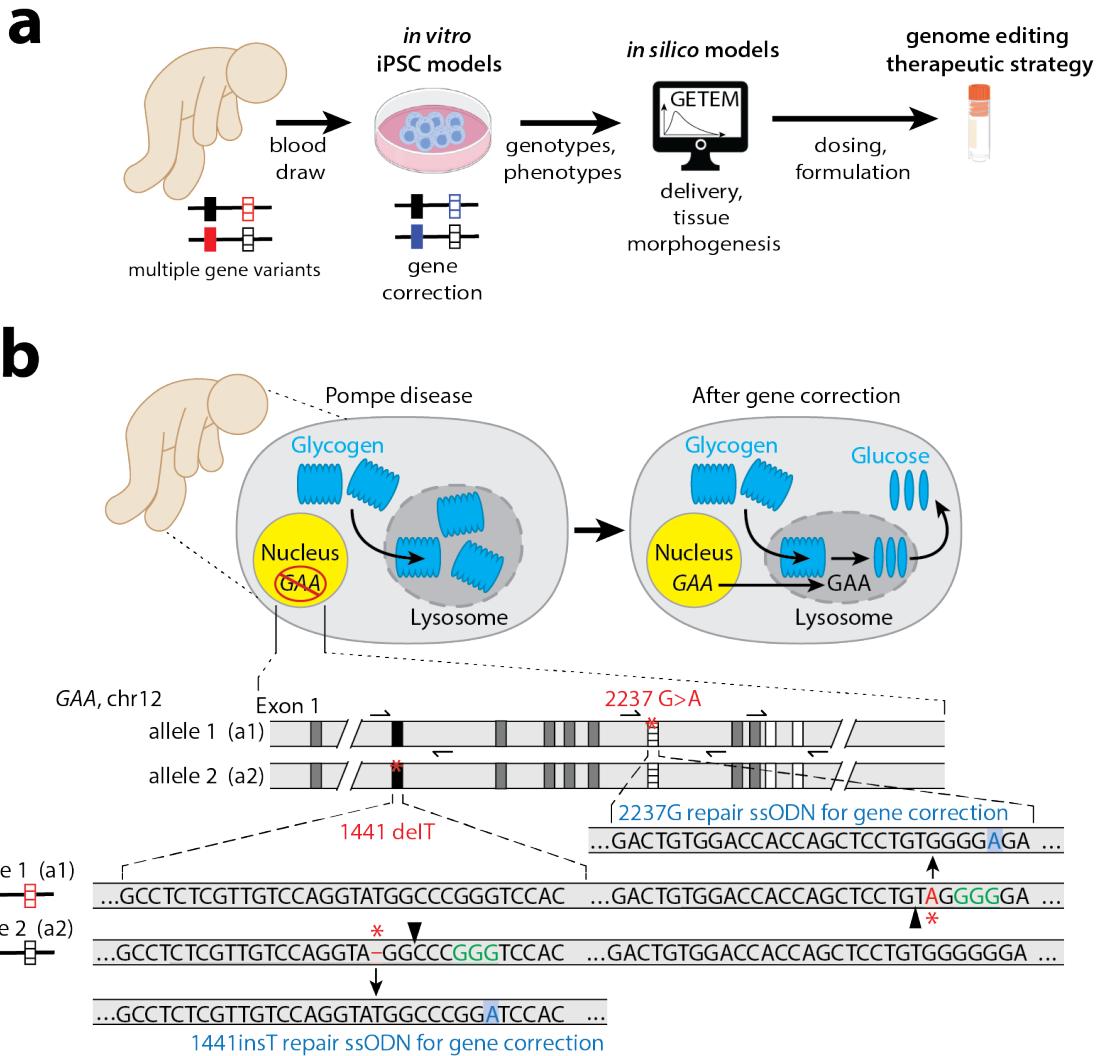
87           Using CRISPR-Cas9 genome editors, we first characterize the phenotypic consequence of genome  
88    correction of two alleles within the same cell and phenotype rescue of the disease phenotype in both iPSC  
89    cells and in a differentiated therapeutically relevant tissue. We then elucidate the second step of a platform  
90    for polygenic diseases (**Fig. 1a**) by generating a novel quantitative *in silico* model to determine an optimal  
91    therapeutic strategy by considering several mechanisms: differing rates of progenitor and mature cell  
92    proliferation, precise and imprecise editing at two alleles, and enzymatic cross-correction mechanisms. **We**  
93    **trained and validated our *in silico* model by considering results from 47 mice across nine studies utilizing**  
94    **four gene therapy delivery strategies and six different genome editors.** We then adapted this model to  
95    examine cell therapy using *ex vivo* engineered cells, and the therapeutic phenotype and persistence of the  
96    engrafted cells. Our analysis indicates that correcting multiple alleles *in vitro* is possible and can correct  
97    phenotypic defects via both single and double gene correction. For cell therapy, the phenotype of the *ex*  
98    *vivo* engineered cells and their persistence relative to native tissue determines the efficacy of *ex vivo*  
99    genome-engineered cell therapy. Additionally, in diseases where non-cell-autonomous mechanisms can be  
100   therapeutic (e.g., storage disorders with enzymatic cross-correction), precise correction strategies can be  
101   highly efficacious. For these conditions, several new genome editing strategies<sup>43,44</sup> could be designed with  
102   new delivery systems<sup>45,46</sup> and higher fidelity genome editors to correct multiple mutant alleles within a  
103   particular patient.

104

105 **RESULTS**

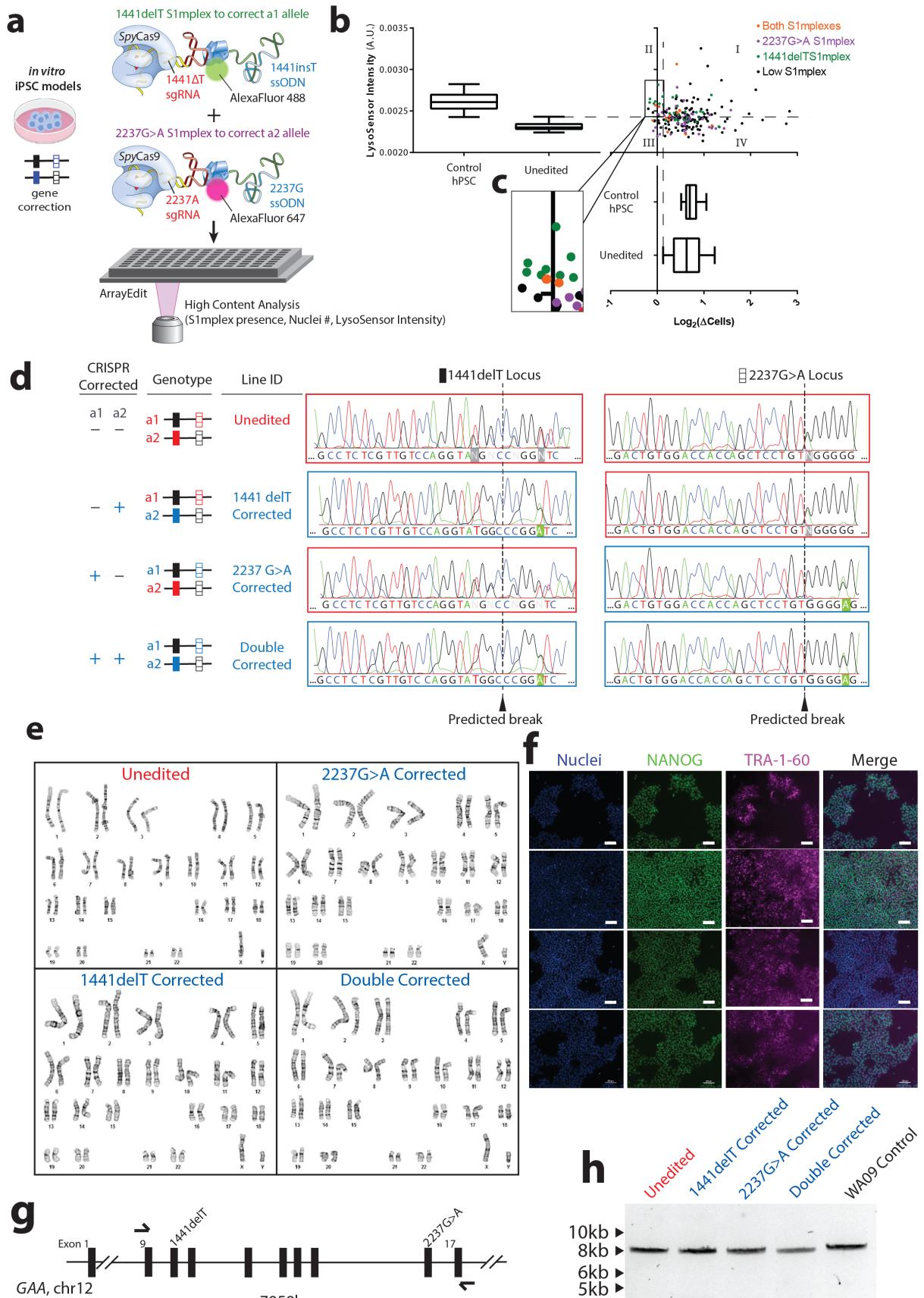
106 **Correction of two distinct diseased alleles within Pompe iPSCs.** To correct two endogenous alleles  
107 within the same cell, several clonal isogenic iPSC lines were generated by CRISPR-Cas9 gene editing of  
108 an iPSC line derived from a patient with infantile-onset Pompe disease<sup>47</sup>. In this line, compound  
109 heterozygous *GAA* mutations responsible for the disease phenotype are a deletion of a thymidine nucleotide  
110 at position 1441 (GAA:c.[1441delT], “1441delT”) causing a frameshift, and premature stop codon on one  
111 allele, and a G>A conversion at nucleotide 2237 (GAA:c.[2237G>A], “2237G>A”), forming an immediate  
112 stop codon on the other (**Fig. 1b**). The mutations within *GAA* in this patient are ~6.1 kb apart, and hence  
113 using a single DNA double-strand break (DSB) with homology directed repair from a long plasmid or viral  
114 donor would likely be inefficient<sup>17</sup>. We, therefore, used a strategy utilizing two distinct SpyCas9  
115 ribonucleoproteins (RNPs) with accompanying single-stranded oligonucleotide (ssODN) templates  
116 encoding the gene correction (**Fig. 1b and 2a, Supplementary Tables 1-2**). The use of a transient RNP-  
117 based strategy lowers the lifetime of the editor within the cells and therefore reduces the chance of off-  
118 target and adverse events<sup>48</sup>.

119 Using a combination of S1mplex<sup>19</sup> and ArrayEdit<sup>49</sup> technologies developed by our lab, we enriched  
120 for properly-edited iPSCs after delivery of the two genome editors by tracking the presence of genome  
121 editors within the nucleus (**Fig. 2b**, see page 8). Next, by using high-content analysis imaging of the iPSC  
122 clones during culture post-delivery of the editors, we tracked the growth rate of clones, as well as screening  
123 the pH of the lysosome<sup>47</sup> using a Lysosensor dye. Lysosensor is sensitive to the buildup of glycogen in the  
124 diseased lysosome of mutant *GAA* cells, as high glycogen phosphorolysis neutralizes this otherwise acidic  
125 organelle<sup>47</sup> (**Fig. S1, Fig. 2a,b**).



**Figure 1 | A combined *in vitro* and *in silico* strategy to evaluate the efficacy of different gene correction therapeutic strategies.** **a**, Schematic indicating the modeling approach in which samples from patients are collected *ex vivo* are then genome-edited *in vitro*. Genotypes and phenotype outcomes from the *in vitro* studies are inputs for an *in silico* model that simulates the delivery of the therapeutic *in vivo* as well as tissue morphogenesis. The results of the *in silico* model can ultimately guide dosing and formulation decisions. **b**, Editing strategy for gene correction of Pompe-diseased induced pluripotent stem cells (iPSCs). Pompe disease is caused by two defective copies of the acid- $\alpha$ -glucosidase (*GAA*) gene. This enzyme is responsible for breakdown of glycogen within lysosomes inside cells. Without *GAA*, glycogen build up can cause downstream health issues. After correction, *GAA* expresses a functional protein leading to a reduction in glycogen. The schematic indicates the editing locations within *GAA* locus and CRISPR gene correction strategy. In the Pompe patient-derived line, cells harbor compound heterozygous mutations in *GAA*. Allele one, a1, contains a point mutation that causes a premature stop codon (*GAA:c.[1441=;2237G>A]*) while allele two, a2, carries a one base pair deletion (*GAA:c.[1441delT;2237=]*). For the CRISPR gene correction strategy, single guide RNAs (sgRNAs; the predicted DNA double-strand break by SpyCas9 is denoted by the arrowhead) were designed to be specific to only the diseased allele by containing the mutant bases (red) within the seed region. Single-stranded oligonucleotides (ssODNs) used for genomic repair contained the wildtype sequence at the mutation site (blue) as well as a silent mutation “wobble” to remove the PAM site (green) to prevent re-cutting of the corrected allele while preserving the amino acid sequence of *GAA*.

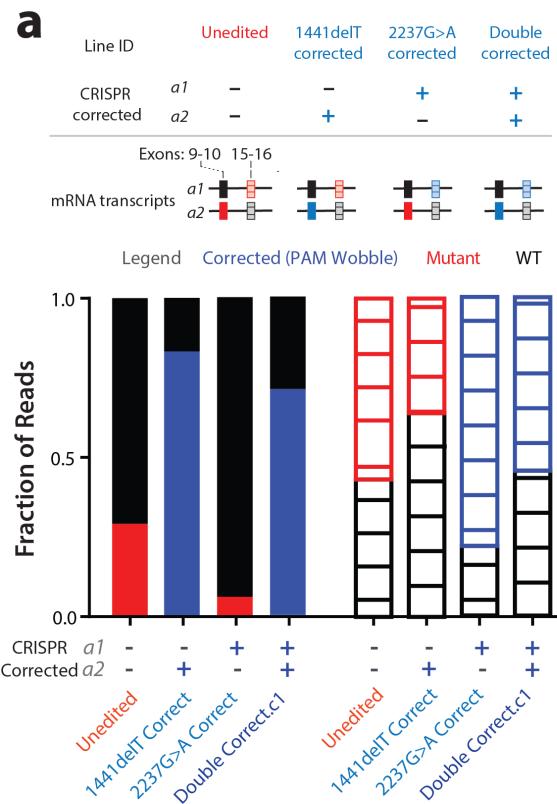
146 We isolated cell lines that were corrected at the 1441delT allele and the 2237G>A allele  
147 individually (**Fig. 2c**, termed ‘single-corrected’). We also isolated a clone corrected at both  
148 GAA:c.[1441delT];[2237G>A] alleles (**Fig. 2d**, termed ‘double-corrected’; clone ‘c1’). After karyotyping  
149 each of the isolated lines, we observed no large transversions or inversions (**Fig. 2e**) and verified that all  
150 gene-corrected lines remained pluripotent (**Fig. 2f**). Because genome editing can create large indel  
151 mutations<sup>50</sup>, we also conducted an 8 kb PCR on *GAA* that included both sgRNA target sites and observed  
152 no genomic deletions between the sgRNA target sites (**Fig. 2g,h**). Sequencing of these large PCR amplicons  
153 confirmed that both alleles were present, and no other sequence abnormalities were detected at the edited  
154 loci (**Fig. S2**). Finally, chromatograms from Sanger sequencing at the top ten off-target sites for each  
155 sgRNA matched the untransfected, patient-derived cell line, indicating [that none of the top ten off-target](#)  
156 [regions were modified by our editing strategy \(Fig. S3-S5; Supplementary Table 4\)](#). To ensure that our  
157 observations were not specific to one clone, five additional double-corrected lines (**Fig. S7**; clones c21,  
158 c28, c29, c30, and c73) were generated using the S1mplices shown in **Fig. 2c** or by a transient puromycin  
159 treatment protocol<sup>27</sup>.



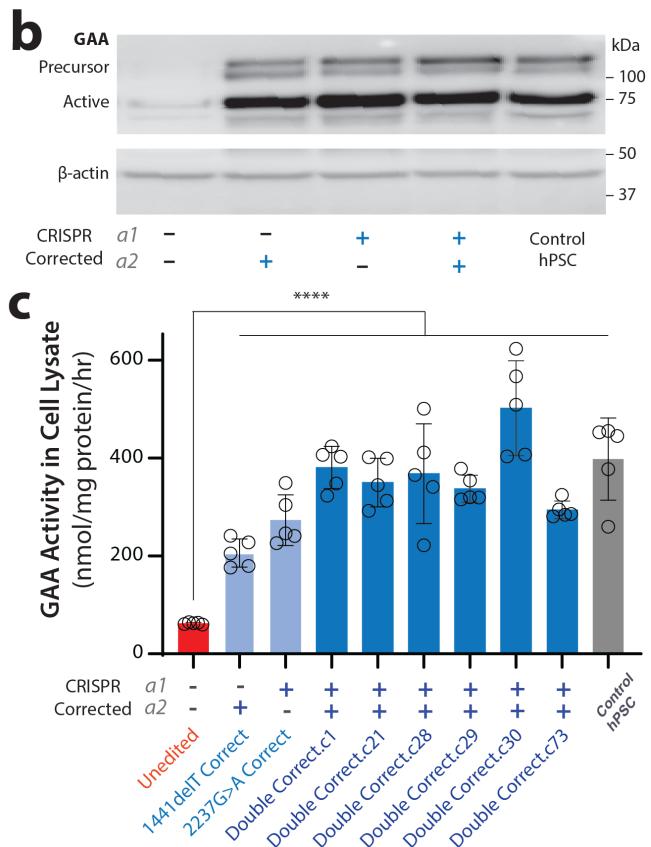
161 **Figure 2| Gene correction of two distinct mutant alleles within a single human cell.** **a**, ArrayEdit-based  
162 isolation of iPSC clones corrected at either or both loci. CRISPR S1mplex design for the gene correction  
163 of compound heterozygous mutations. S1mplexes targeting 1441delT mutant were labeled with an  
164 AlexaFluor488 compound while S1mplexes targeting the 2237G>A mutation were labeled with an  
165 AlexaFluor647. These ribonucleoprotein (RNP) genome editors were mixed prior to transfecting into cells  
166 and subsequently plated on the ArrayEdit platform to conduct high-content analysis (see Fig. S1). **b, Left:**  
167 LysoSensor quantification per  $\mu$ Feature of two mock transfactions after seven days of growth. Normal  
168 control hPSCs were significantly more intense than unedited, Pompe diseased iPSCs on ArrayEdit. **Bottom**  
169 **Right:** The growth rate of unedited and control hPSCs following a mock transfection to establish a baseline  
170 for growth. Growth rates were calculated by measuring the per-day change in the number of cells of the  
171  $\mu$ Feature. Features were graphed as an average of these per day changes. **Top Right:** LysoSensor intensity  
172 was plotted against growth rate per  $\mu$ Feature to identify edited colonies. Individual plotted colonies were  
173 also assayed for the presence of either genome editor (represented in either purple or green), both genome  
174 editors (red) and low amounts of genome editors (black). Colonies of interest are identified as those with  
175 high genome editor expression and lower growth rates, presumably arising from the stress of genome  
176 editing. Dashed lines indicate regions of interest. **c**, Magnification of quadrant II from panel b.  $\mu$ Features  
177 in this region were selected for genomic analysis to isolate edited clones. **d**, Sanger sequencing traces of  
178 corrected cell lines. The unedited line contains mutations at both alleles: 1441delT mutation causes a  
179 breakdown of sequence trace, whereas a single point mutation demonstrates a heterozygosity 2237G>A  
180 locus. Single corrected clones remain identical to the unedited line at the unedited locus and contained PAM  
181 wobble on the corrected allele. The double-corrected line contained PAM wobble at both loci. Wobble A  
182 bases in the corrected lines are highlighted to indicate repair from the ssODN. SpyCas9 cut site is denoted  
183 by a dotted line, and sequencing chromatograms do not show evidence of undesired NHEJ products. See  
184 Fig. S2 for characterization of the corrected lines. **e**, Karyotypes of all isolated gene-corrected lines as well  
185 as unedited cells. No abnormalities were detected at a band resolution of 500. **f**, Immunocytochemistry of  
186 pluripotency markers in gene-corrected lines. All lines were positive for pluripotency markers NANOG  
187 and TRA-1-60 (scale bar: 100 $\mu$ m). **g**, Schematic of long PCR covering both SpyCas9 cut sites. Arrows  
188 denote primers. The expected PCR amplicon is 7959 bp in length. **h**, Gel analysis of long-range PCR  
189 described in panel g in each isolated cell line. No significant deviances from the expected length were  
190 detected, and no other notable bands were observed. WA09 control cells are hPSCs.  
191

192 Quantitative RT-PCR (qRT-PCR) at the 3' end of the GAA mRNA transcript, as well as around  
193 each edited locus (**Fig. 3a**), indicated that the corrected loci were correctly expressed. We observed that the  
194 unedited line expressed the lowest levels of *GAA* transcripts when compared to internal *GAPDH* levels  
195 (**Fig. 3a**, **Fig. S6**), despite the presence of full-length, mature mRNA that could be used to express the  
196 protein (**Fig. 3a**). Both the single- and double-corrected lines also expressed mature *GAA* transcripts. By  
197 looking for the presence of disease variants and protoadjacent motif (PAM) wobbles introduced by the  
198 ssODN (**Fig. 1b**) via deep sequencing on endpoint PCR samples of mRNA, we observe that both alleles  
199 are expressed individually at higher levels (3-5 fold increase) than unedited cells (**Fig. 3a**). Each allele is  
200 expressed similarly to the corresponding single corrected line (**Fig. 3a**, **Fig. S6b**, **Supplementary**  
201 **Information**). These findings suggest nonsense-mediated decay of the mutant transcript<sup>51</sup> or cellular

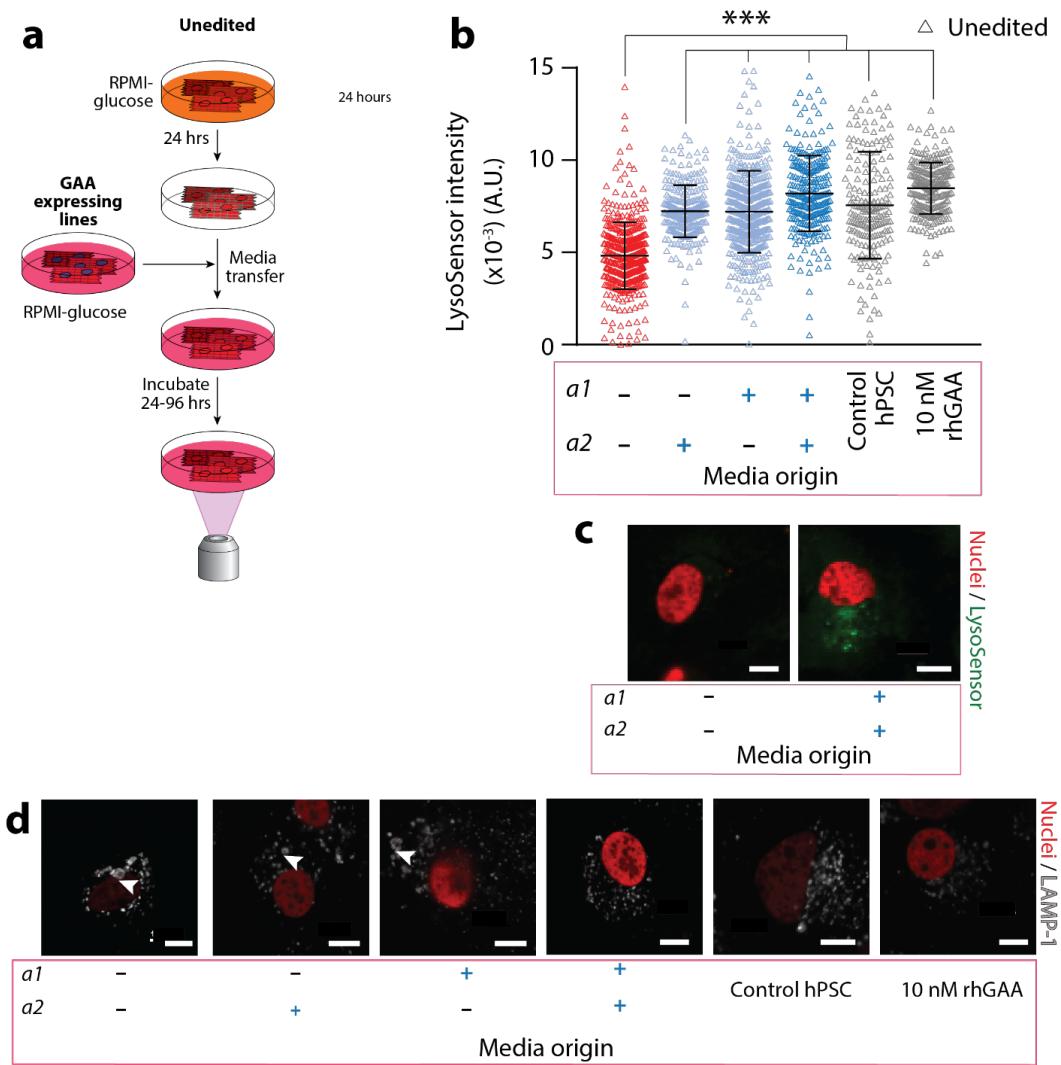
202 compensation<sup>52</sup> to overcome the mutant allele within the single corrected lines. We detected active GAA  
203 protein using a Western blot (**Fig. 3b**) at levels comparable to a control hPSC line. We were also able to  
204 identify precursor polypeptides, which are important for protein secretion<sup>53</sup>, showing the *GAA* transcripts  
205 from the edited alleles are correctly translated and processed within cells. Notably, we were able to detect  
206 only small amounts of GAA protein and precursor polypeptides in the unedited iPSCs. All edited cell lines  
207 (**Fig. S7**) were able to produce (**Fig 3c**) and secrete active GAA (**Fig. 3d, Fig. S6d**).  
208



211 **Figure 3 | Transcription and protein levels within gene-corrected cells.** **a**, Schematic indicating the  
212 genotypes of the iPSC lines generated and mRNA transcripts collected. Deep sequencing from qRT-PCR  
213 around both diseased loci in isolated cell lines. Reads were classified as WT, mutant, or corrected (PAM  
214 wobble) and mapped to either allele 1 (top bar) or allele 2 (bottom bar). When neither allele was corrected,  
215 both alleles were expressed at approximately the same rate. However, when either mutation was corrected,  
216 the corresponding allele was expressed at a higher rate than the one that still possessed a mutation. When  
217 both alleles were corrected, the fraction of reads making up the population was evenly distributed.  
218 Observations at individual alleles were consistent across both assayed loci. **b**, Western blot for GAA  
219 protein. Each of the corrected lines expressed high levels of active protein as well as detectable levels of  
220 the precursor protein. Unedited cells expressed significantly lower levels of GAA protein, but the level was  
221 still above the limit of detection. **c**, GAA activity in cell lysate as measured by 4-MUG cleavage in acidic  
222 conditions. Unedited cells have significantly lower activity showing there was little to no active protein.  
223 Corrected cell lines include Double Correct.c1 (isolated via ArrayEdit), Double Correct.c21, Double  
224 Correct.c28, Double Correct.c29, Double Correct.c30 (isolated via sequential correction of 2237G>A allele  
225 followed by S1mplex electroporation correction of 1440delT) and Double Correct.c73 (isolated via  
226 sequential correction of 2237G>A allele followed by using transient puromycin based correction of  
227 1440delT). All corrected lines had significantly higher activity than unedited cells but were  
228 indistinguishable from each other ( $n=5$  technical replicates,  $****p<0.0001$ , two-tailed t-test,  $\alpha = 0.05$ ,  
229 heteroscedastic; Bar graphs plotted with standard deviation).



231 **Enzymatic cross-correction by gene-corrected cells.** Detection of active GAA secretion by the edited  
232 cells led us to test the potential of edited cells to enzymatically cross-correct diseased cells (**Fig. 4a**).  
233 Because Pompe disease has a significant effect on cardiac tissue in infants, we differentiated iPSC lines to  
234 cardiomyocytes (Pompe iPSC-CMs) using a previously described small-molecule inhibitor protocol<sup>54</sup> (**Fig.**  
235 **4a**). For all differentiations, we observed spontaneous contraction (**Supplementary Videos**) and confirmed  
236 the expression of  $\alpha$ -actinin, a marker of cardiac lineage commitment (**Fig. S8a**). Similar to results seen in  
237 the iPSC state, differentiated corrected lines still expressed and secreted active GAA, as indicated in a 4-  
238 MUG cleavage assay on cardiomyocyte protein lysates and spent culture media (**Fig. S8b-e**,  
239 **Supplementary Information**). It has previously been demonstrated that by culturing in medium devoid of  
240 glucose, Pompe iPSC-CMs display an accumulation of glycogen within the lysosome<sup>47</sup>. We performed a  
241 medium exchange experiment wherein we took a partially spent, glucose-free medium from each corrected  
242 line (putatively containing secreted active GAA) and used it to replace glucose-free medium on unedited  
243 Pompe iPSC-CMs (**Fig. 4a**). One day after this media exchange, cells were stained with LysoSensor, and  
244 subsequent confocal microscopy was used to measure lysosome acidity as a proxy for glycogen clearance.  
245 As a control, we added rhGAA to unedited Pompe iPSC-CMs to simulate ERT. When unedited  
246 cardiomyocytes were supplemented with 10 nM rhGAA (ERT), LysoSensor intensity increased, indicating  
247 clearance of glycogen from the lysosome. Media from all edited cells were able to recover the lysosomal  
248 pH at 96 hr (**Fig. 4b**), and this clearance is expected to continue until normal levels of glycogen were  
249 reached<sup>55</sup>. Within these cultures, we qualitatively observed lysosomal size of unedited Pompe iPSC-CMs  
250 in GAA-positive media through visualization of Lysosomal Associated Membrane Protein 1, (LAMP-1).  
251 **Detailed quantification of lysosome size was not possible, as the pH sensitive dye utilized in this method is**  
252 **prone to photobleaching as noted previously**<sup>56</sup>. In media from unedited cells, lysosomes were enlarged,  
253 consistent with the buildup of glycogen (**Fig. 4d**). In comparison, when media was taken from double-  
254 corrected cells or supplemented with rhGAA, lysosomes appeared as punctae. Samples from single  
255 corrected cells fell between these two extremes. Taken together, the single- and double-corrected cells  
256 enzymatically cross-correct diseased cardiomyocytes quickly and effectively.



259 **Figure 4 | Enzymatic cross-correction of unedited cardiomyocytes by gene-corrected cardiomyocytes.**  
260 a, Schematic of enzymatic cross-correction experiments using gene-corrected cardiomyocytes. Unedited  
261 iPSC-CMs (red) were supplied media without glucose for 24 hours (orange). After 24 hours, media was  
262 replaced with media (pink) that had previously been exposed to corrected cell lines (blue) or supplemented  
263 with rhGAA. Ninety-six hours after replacement, unedited cells were stained with LysoSensor and imaged  
264 using confocal microscopy for dye intensity. b, Quantification of LysoSensor intensity in cross-corrected  
265 lines 96 hours post media exchange. Each triangle represents a corrected cell identified using CellProfiler.  
266 After 96 hours of daily media changes or supplementing with rhGAA, all conditions had a significant  
267 increase in dye intensity over control conditions. Unedited cells were modified to express histone 2B  
268 (H2B)-mCherry to facilitate imaging of the nuclei in these assays. See also Fig. S7. (\*\* $p < 0.005$ ,  $n > 134$ ,  
269 details per condition in Supplementary Table 7, two-tailed t-test,  $\alpha = 0.05$ , heteroscedastic; bar graphs  
270 plotted with standard deviation). c, Representative images of unedited iPSC-CMs stained with LysoSensor  
271 in media from unedited and double-corrected iPSC-CMs. d, Representative images of LAMP1 staining in  
272 unedited, single corrected, double corrected cells and control PSC-CM and unedited iPSC-CM treated with  
273 rhGAA. (scale bars:  $10\mu m$ ). Select enlarged LAMP1-positive lysosomes are identified by white  
274 arrowheads.

275 **Establishing a Pompe disease gene therapy efficacy model (Pompe GETEM).** Having observed  
276 phenotypic rescue with single- and double-correction in both iPSCs and disease-relevant, differentiated  
277 cardiomyocytes, we next sought to identify potential strategies to design translational studies for polygenic  
278 diseases. For instance, determining the minimum effective dose of genome editors for infants with Pompe  
279 Disease involves the consideration of tissue morphogenesis, delivery to various tissues, edited protein  
280 levels, as well as the spectrum of various genomic outcomes. Whenever possible, we look to patient data  
281 to gain information about these processes and accordingly have created a novel *in silico* approach to model  
282 somatic cell genome editing. Specifically, to investigate the dynamics of gene correction approaches *in*  
283 *vivo*, we developed an *in silico* GEne Therapy Efficacy Model (GETEM) simulating genome editing of two  
284 alleles within a developing liver. The gene-corrected liver within Pompe diseased patients would act as a  
285 depot for GAA dissemination to distal organs (e.g., the heart and skeletal muscle<sup>30</sup>), based on mechanisms  
286 described in prior gene augmentation studies in animal models and currently in clinical trials<sup>57</sup>.

287 First, we simulated standard ERT to ground our model with empirical clinical data gathered from  
288 infants and children treated with ERT<sup>58</sup>. This grounding enables our Pompe GETEM to account for  
289 enzymatic cross-correction (**Fig. 5a**, see page 19) properly. Despite high levels of GAA in the liver on ERT,  
290 GAA transport to and uptake within the distal muscle can be low<sup>59</sup>, and this is accounted for in Pompe  
291 GETEM through a loss factor (**Supplementary Text**). Liver progenitors in this model are proliferative and  
292 give rise to non-proliferative mature cells<sup>60</sup>. Both progenitor and mature cells can absorb extracellular GAA  
293 to be enzymatically cross-corrected from a diseased to a normal state, and subsequently reverted to the  
294 diseased state as GAA degrades. In distal cardiac and skeletal muscle tissues, the tissue absorbs GAA for  
295 enzymatic cross-correction to a normal phenotype (**Supplementary Information**). Biweekly intravenous  
296 ERT doses<sup>58</sup> lead to an oscillatory percentage of phenotypically normal cells in the heart, skeletal muscle  
297 and liver within our model (**Fig. 5b**, see page 19; **Fig. S9**), reaching 32.5% normal cells in the heart on  
298 average after one year, similar to levels observed with ERT, enough for significant heart glycogen  
299 clearance<sup>59</sup>.

300

301 **Empirical genome editing data to ground and validate GETEM.** Because genome editing data of the  
302 human *GAA* within patients *in vivo* or within any animal model are not available, we use empirical data  
303 from [previously-published](#) genome editing experiments targeting a single endogenous gene with mice.  
304 These studies provide quantitative preclinical data that ground the delivery mechanisms to the liver within  
305 GETEM. While several studies have performed precise gene correction studies in the mouse livers, they  
306 have typically delivered their editor of choice using a plasmid<sup>46,61</sup>, a virus<sup>62</sup> or an encapsulated mRNA (in  
307 conjunction with a virus for the donor correction template)<sup>63</sup>. In GETEM, we consider all these delivery  
308 strategies to produce RNPs that ultimately produce precise and imprecise genome edits within a cell (**Fig.**  
309 **5c**, see page 19). [We then utilized the published empirical data from plasmid DNA, mRNA and viral](#)  
310 [delivery to train GETEM \(n=18 mice for training, three different genome editors\), and subsequently](#)  
311 [validated GETEM using data from RNP delivery into the liver<sup>64</sup> and four more liver editing studies<sup>65–68</sup>](#)  
312 (*n*=29 mice for validation; **Fig. 5d**; details on mouse studies in **Supplementary Table 8**).

313 To account for the different delivery strategies, a previously reported model of the central dogma<sup>69</sup>  
314 was employed to calculate the amounts of transcribed and translated RNP. Either 1) plasmid DNA  
315 transcription and translation, 2) mRNA translation, or 3) AAV-mediated transcription and translation led  
316 to the formation of editor proteins. [In comparison to sgRNA transcription, Cas9 transcription or translation](#)  
317 [was assumed to be rate limiting for the formation of the RNP when Cas9 is delivered as an mRNA or](#)  
318 [encoded on DNA.](#) AAV doses were converted to an effective viral DNA dose based on the published vector  
319 copies per cell data provided<sup>62</sup>. The degradation of the plasmid or viral DNA occurs constitutively as  
320 previously measured with mouse macrophages<sup>62</sup>. For studies that targeted the *Fah* allele, a selection  
321 pressure enriches for edited cells through the experiments as delineated by the dashed lines in **Fig. 5e**. [The](#)  
322 [dynamics for the enrichment has been well characterized through prior experimental studies<sup>70</sup>.](#) For these  
323 studies, the unedited cells die quickly unless the mouse is continuously fed 2-(2-nitro-4-  
324 trifluoromethylbenzoyl)-1,3-cyclohexanedione (NTBC). Once corrected, the liver cells can survive without  
325 NTBC supplementation.

326 GTEM considers the rate of genome editing,  $dE/dt$ , to be a second-order mass action rate  
327 equation dependent on both the number of RNPs,  $N_{RNP}$ , and the number of cells,  $N_{Cell}$  (Equation 1), where  
328  $k$  is the genome editing rate constant:

$$\frac{dE}{dt} = k \cdot N_{Cell} \cdot N_{RNP} \quad (1)$$

329 The change of edited cells against time,  $\frac{\Delta E}{\Delta t}$ , was calculated using numerical differentiation for each time  
330 step (Equation 2):

$$\frac{dE}{dt} \sim \frac{\Delta E}{\Delta t} = \frac{E_n - E_{n-1}}{t_n - t_{n-1}} \quad (2)$$

331 To account for tissue growth, this time derivative was normalized for the total number of cells in the mouse  
332 at the same time point and averaged over the experiment duration (Equation 3):

$$\overline{\frac{dE}{dt}} = \frac{1}{N_{Cell}} \overline{\frac{\Delta E}{\Delta t}} \quad (3)$$

333 For each published study in the training dataset ( $n = 18$  mice), random sampling of the experimental data  
334 by bootstrapping<sup>71</sup> produced ten thousand samples that constituted a surrogate “bootstrapped” data set. For  
335 each of these bootstrapped samples, the average rate of liver genome change,  $\overline{\frac{dE}{dt}}$  was calculated individually  
336 (**Fig. 5d**). The mean and standard deviation of this set of bootstrapped  $\overline{\frac{dE}{dt}}$  values was utilized in GTEM to  
337 simulate the range of outcomes for the mice in each study in the training set (**Fig. 5e**, left four panels). The  
338 trained GTEM shows strong concordance of the simulation with experimental data at the assay timepoints  
339 for each study (black experimental data is within the blue range of simulation outcomes).

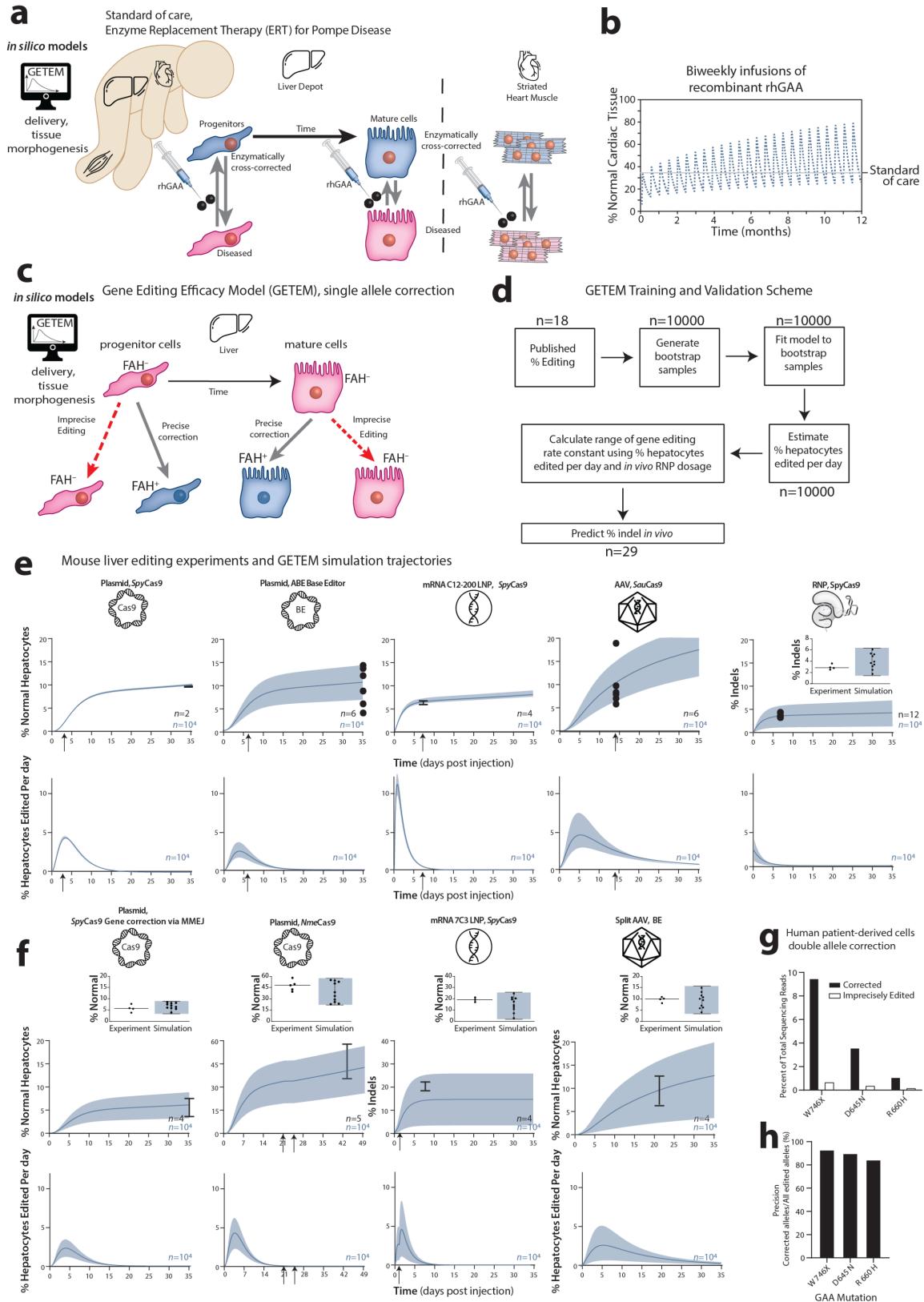
340 To validate the model, the bootstrapped  $\overline{\frac{dE}{dt}}$  values from all the training studies were combined. The  
341 mean and one standard deviation of the combined set of surrogate  $\overline{\frac{dE}{dt}}$  values were used to simulate the  
342 genome editing outcomes for RNP doses employed by a separate study<sup>64</sup> outside of the training dataset  
343 (**Fig. 5e**, right). There is strong concordance of the experimental data with the results of the simulation at  
344 the assay timepoint, day seven (**Fig. 5e, inset**). To further validate GTEM, we modeled other somatic cell

345 genome editing strategies evaluated in mouse livers – microhomology based gene correction using  
346 *SpyCas9*<sup>65</sup>, *Neisseria Meningitidis Cas9* (*NmeCas9*) based editing<sup>66</sup>, a different type of lipid nanoparticle  
347 delivering *SpyCas9* mRNA and AAV delivery of split base editors<sup>68</sup>. The simulation results for all these  
348 strategies showed concordance with the experimental data (**Fig. 5f**). In total, the validation data set includes  
349 29 mice across four different delivery strategies with four different editing strategies (**Fig. 5d-f**).

350 The increase in the percentage of edited cells is dependent on both genome editor activity and  
351 genome editor independent processes (growth dynamics, selection of edited phenotypes). However, *in vivo*,  
352 these processes can be difficult to analyze separately. Using the validated GETEM approach, we can  
353 analyze the processes of liver genome change and genome editor activity separately. The top panels (**Fig.**  
354 **5e and Fig. 5f**) indicate the cumulative *change* in the liver genome over time. The lower panels (**Fig. 5e**  
355 **and Fig. 5f**) exhibit the *rate* of the genome editing in the target cells, hepatocytes. During the duration of  
356 the simulation (top panels, **Fig. 5e and Fig. 5f**), the percentage of the liver genome that is changed continues  
357 to rise even after genome editor activity declines (bottom panels, **Fig. 5e and Fig. 5f**). Genome editor  
358 activity increases initially for all genome editors except RNP. This is driven by transcription and translation  
359 of the delivered payload to generate genome editors within cells *in situ*. For dosing with RNP genome  
360 editors, there is no *in situ* generation of the editor within the liver, and only a decline in activity is observed.  
361 Thus, the prolonged increase of editing at the tissue-level highlights the role of genome editor independent  
362 processes (e.g., attributed to growth or selection of edited phenotypes) in driving efficacy.

363 In order to extend GETEM - already validated for single allele editing - to biallelic gene correction  
364 for Pompe disease, additional training data is required on the precision of biallelic genome editing and their  
365 cellular phenotypic consequences. We performed additional *in vitro* experiments with four different  
366 pathological mutations in *GAA* in Pompe patient-derived cells (**Fig. 5g,h**). Precise gene correction with  
367 S1mplexes occurred approximately in 84 – 93% of the edited sequence reads at the target locus as assayed  
368 by deep sequencing of genomic DNA. The remaining are imprecise, unintended modification to the on-  
369 target site, which could destroy the PAM or modify the on-target site for subsequent editing of these alleles.  
370 According to the nomenclature of Shen et al<sup>73</sup> to describe this ratio of gene correction to other editing

371 outcomes, S1mplexes are approximately “precise-80.” In comparison, normal *SpyCas9* is precise-50, and  
372 prime editors<sup>44</sup> are precise-90. Next, experimental measurements of functional outcomes from this  
373 distribution of edited cells ground protein levels in the Pompe GETEM. Cells corrected at both alleles a1  
374 and a2 have been modeled to secrete the same amount of GAA than those edited at a single allele (**Fig. 3d**)  
375 at first approximation. We include a parameter in our cell therapy GETEM to explore the effects of higher  
376 secretion arising from double-corrected cells. In summary, high-resolution measurements of alleles after  
377 genome editing, as well as quantitative measures of protein levels and activity post-editing, provide key  
378 empirical rate constants for the training of Pompe GETEM.



380 **Figure 5| Derivation, empirical training and validation of *in silico* Gene Therapy Efficacy Model**  
381 (**GETEM**) for Pompe disease correction in a developing infant. **a**, Schematic showing enzymatic cross-  
382 correction of somatic tissues by injected rhGAA (Enzyme Replacement Therapy, ERT). For a given  
383 developing liver of a patient, progenitor cells proliferate and can asymmetrically divide into mature cells.  
384 Injected rhGAA is absorbed by the liver as well as striated muscle tissue (both heart and skeletal muscle).  
385 **b**, Percentage of normal cardiac tissue within a heart of a Pompe diseased infant after one year of biweekly  
386 ERT at 20 mg/kg. Percentages of normal skeletal muscle tissue and hepatocytes are in Fig. S9. **c**, Schematic  
387 for *in silico* gene editing model for previously published precise correction of a single base pair in mouse  
388 disease models. Delivery of genome editors to livers within *Fah*<sup>-/-</sup> mice can target a mutant  
389 fumarylacetoacetate hydroxylase allele to correct a mutated stop codon. Both *Fah*<sup>-/-</sup> progenitor and mature  
390 hepatocytes can be edited to form *Fah*<sup>+/+</sup> hepatocytes, and imprecisely edited *Fah*<sup>-/-</sup> hepatocytes. The *Fah*  
391 hepatocytes are subject to a death rate which is modified by the presence of 2-(2-nitro-4-  
392 trifluoromethylbenzoyl)-1,3-cyclohexanedione (NTBC). NTBC treatment matches the net growth rate of  
393 *Fah*<sup>-/-</sup> and *Fah*<sup>+/+</sup> hepatocytes, while in the absence of NTBC, *Fah*<sup>-/-</sup> hepatocytes have a net negative growth  
394 rate. This model incorporates the growth advantage of precisely edited cells, **d**, Flowchart for training and  
395 validating the GETEM model using previously published studies of *in vivo* somatic cell genome editing of  
396 a mouse liver by intravenous injection. Studies employed three different genome editors delivered using  
397 four different methods: plasmid delivery of SpyCas9, gRNA and repair template in a *Fah*<sup>-/-</sup> mouse model;  
398 plasmid delivery of base editor and gRNA in a *Fah*<sup>-/-</sup> mouse model; mRNA delivery of SpyCas9 (C12-200  
399 lipid nanoparticle) and gRNA and AAV delivery of repair template in a *Fah*<sup>-/-</sup> mouse model; and, AAV  
400 delivery of SpyCas9, gRNA, and repair template to edit Ornithine transcarboxylase (*Otc*). Ten thousand  
401 simulated samples were generated from the empirical data for each mouse via bootstrapping<sup>71</sup>, and the  
402 GETEM was fit to each of these bootstrapped samples to calculate mean  $\pm$  1 s.d. for the gene editing rate  
403 constant via Equations 1 – 3 in the text. The range of editing outcomes was generated and validated against  
404 the published data for intravenous delivery of SpyCas9 RNP targeting *Pten* in the liver. **e**, (top) Absolute  
405 change of edited cells over time predicted by the GETEM for the studies described in d. For studies that  
406 employed selection in the liver, arrows indicate the time at which the selective pressure favoring edited  
407 cells was applied (for *Fah*<sup>-/-</sup> models, this indicates the removal of NTBC supplementation in the diet; and,  
408 for *Otc* editing, this indicates the induction of a high protein diet.). The lines indicate the mean and standard  
409 deviation of edited alleles in the mice cohorts, while dots signify the percentage edited alleles of individual  
410 mice for studies in which individual replicate values were reported. Inset: Empirical studies with RNP  
411 delivery validate the models grounded in separate empirical training data based on mRNA, plasmid and  
412 AAV studies. The percent of edited alleles at the assay timepoint (black) fall within the range of predicted  
413 percent of edited alleles (blue). (bottom) the rate of liver genome editing for each individual bootstrapped  
414 sample (n=10000 per panel) over time for the studies described in d. **f**, Validation of the model using  
415 additional published editing strategies including gene correction using microhomology based gene  
416 correction, metabolic pathway reconditioning using gene editing by *NmeCas9*, 7C3 mRNA lipid  
417 nanoparticle (LNP) based gene knockout using SpyCas9 and gene correction using AAV delivered intein  
418 split BE3. Arrows on days 21 and 24 indicate removal of selection pressure on the treated mice as per  
419 published experimental protocol. Arrow on day 2 in the mRNA LNP indicates redosing of the LNP. **g**,  
420 Percentage of total sequencing reads from primary fibroblasts of Pompe diseased patients that were treated  
421 with S1mplices targeting W746X D645N or R660H GAA mutations. Results indicate gene correction and  
422 imprecise editing for three different mutations. **h**, Percentage of edited alleles that are precisely edited in g.  
423

424 **Design of *in vivo* treatment informed by Pompe GETEM.** The power of building a computational model  
425 is that many different doses can be quickly simulated, providing insights into the spectrum of genotypes  
426 with the tissues over time as well as with the effects of this gene-corrected tissue on other parts of the body.

427 Building on the model in **Fig. 5a**, we generated a model that considers ERT along with gene editing  
428 outcomes after the administration of genome editors targeting multiple alleles (**Fig. 6a**, see page 24). First,  
429 we simulate doses of genome editors required to achieve or exceed efficacy equivalent to ERT standard of  
430 care. For S1mplexes, the dosages of RNP to reach comparable efficacy as ERT were evaluated using the  
431 editing rate established in **Fig. 5**. The S1mplex dose was evaluated to be 23.9 mg/kg/allele to reach  
432 equivalent healing in the liver using six monthly doses via intrahepatic injections, starting at birth, scaled  
433 using the 50<sup>th</sup> percentile growth of a male infant. This dose represents the amount of S1mplex that reaches  
434 the hepatocytes, and not the amount that may need to be injected systemically. At one year after the first  
435 injection, the normal cells in the heart reach 32.5%, the standard calculated from the ERT simulation (**Fig.**  
436 **6b**, see page 21), and the normal cells in the liver consist of 34% genotypically corrected cells (corrected  
437 at alleles a1, a2 or both, **Fig. 6c**), and 66% enzymatically cross-corrected cells (**Fig. 6c**). **Genome editing**  
438 post addition of genome editor is a relatively quick process, and the percentage of edited cells stabilizes  
439 within five days (**Fig. S10**). However, enzymatic cross-correction via secreted and absorbed GAA is a  
440 slower process, and continues beyond the five day gene editing period (**Fig. S10**). Single-corrected cells  
441 constitute the most prevalent edited genotype (**Fig. 6d**).

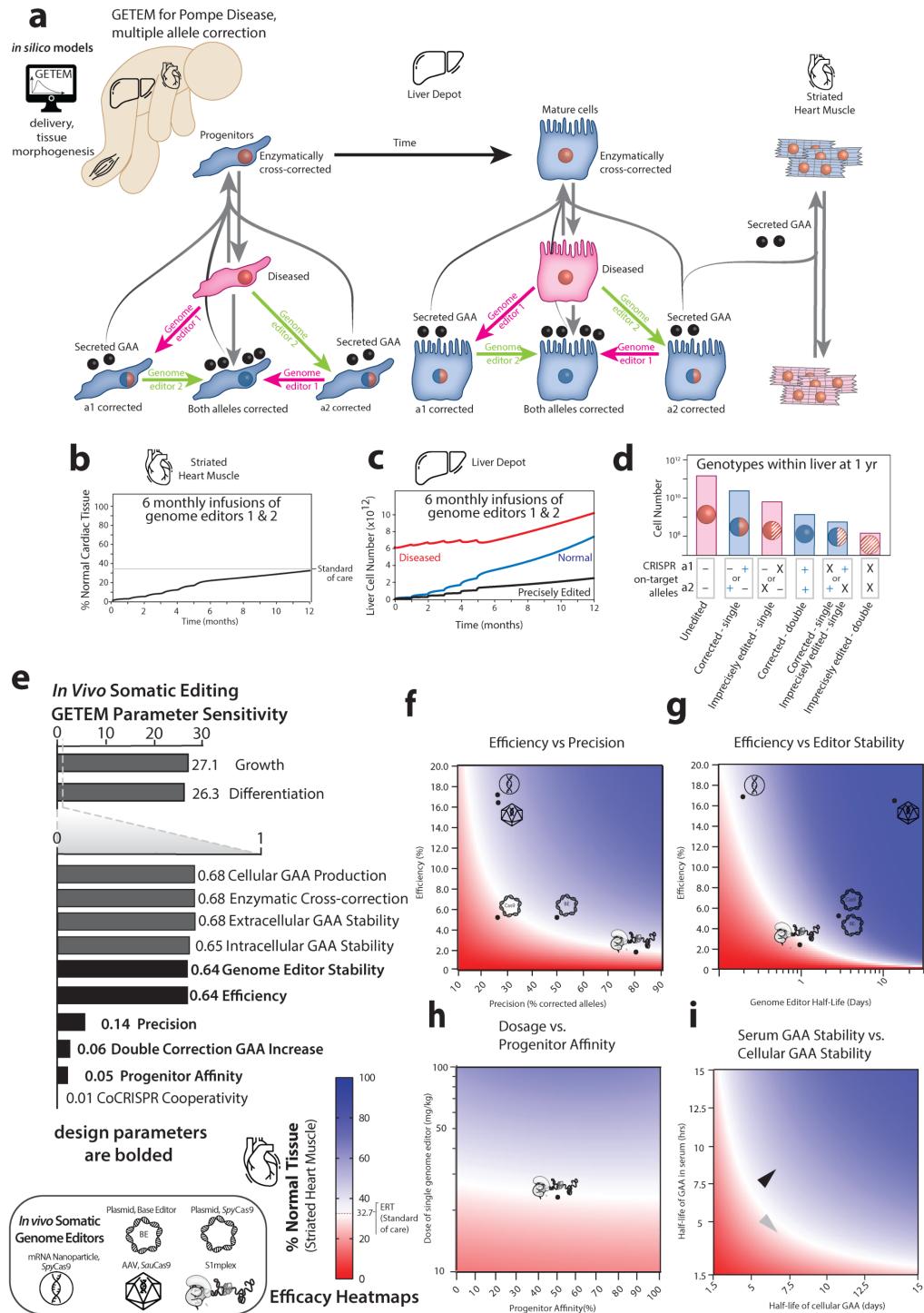
442 Sensitivity analysis (**Fig. 6e**) indicated that tissue morphogenesis factors and cell/tissue biology  
443 factors also control efficacy (**Supplementary Information**). These intrinsic biological properties of each  
444 patient may offer opportunities to tailor therapies to each infant. **The sensitivity analysis helps clarify trade-**  
445 **offs required for gene therapy decisions.** For instance, when considering increasing the magnitude of benefit  
446 by editing both alleles, or editing one allele with higher efficiency – a lower sensitivity is observed for  
447 increasing GAA production within a double corrected cell (0.06) relative to a single-corrected cells than  
448 editing efficiency or precision (~0.64). This lower sensitivity indicates that editing a single allele at higher  
449 efficiency is likely to be more efficacious, as a sensitivity of 0.06 implies that a 1% increase in GAA  
450 production caused by biallele editing would lead to a 0.06% increase in percentage of normal cardiac tissue,  
451 whereas, the higher sensitivity of editing efficiency and precision of 0.64 implies that a 1% increase in  
452 either editing efficiency or precision would increase the percentage of normal cardiac tissue by 0.64%.

453 Therefore, for S1mplices, there is higher expected benefit in editing a single allele more efficiently, than  
454 in editing both alleles at a lower efficiency.

455 Systematic *in silico* investigation of thousands of genome editing strategies are summarized in  
456 several heatmaps (**Fig. 6f-i**, see page 21; **Figs. S11-S12**). First, we observe that imprecise editors require  
457 much higher efficiencies to achieve similar therapeutic levels in the heart (32.7% normal; **Fig. 6f and Fig.**  
458 **S10-S11**). Using a precise-80 editor similar to the S1mplex, even though it has an efficiency of 1%, can  
459 produce similar outcomes as *SpyCas9*, which is almost six times more efficient, but only precise-25.  
460 Second, we observe a negative correlation between editor efficiency and genome editor half-life required  
461 to reach therapeutic levels in the heart (**Fig. 6g and Fig. S11-S12**). The parameter for efficiency in our  
462 model lumps diverse processes of cellular uptake of the editors, intracellular trafficking, nuclease activity  
463 and DNA repair at the target site because few empirical studies have been able to distinguish these processes  
464 *in vivo*. The efficiency of S1mplex editors at this dose level was calculated to be 0.29 – 2% based on the  
465 bootstrap sampling based method presented in **Fig. 5**. Decay rate refers to the extracellular degradation or  
466 binding by serum proteins or other cells/biomolecules in the extracellular space that prevent the activity of  
467 a genome editor. Therefore, strategies that stabilize the editor in the extracellular space – even with low  
468 (~1%) efficiency editors<sup>74</sup> – could be one strategy to increase efficacy. Third, lower dosing with delivery  
469 strategies targeting progenitors could be a potent feasible strategy with highly-precise, yet low-efficiency  
470 strategies (e.g., base editors<sup>45</sup> and S1mplices<sup>75</sup>): increasing the progenitor affinity reduces the combined  
471 genome editor dose by 50% for comparable efficacy (**Fig. 6h and Fig. S11-S12**). Progenitor targeting can,  
472 therefore, increase the potency of a genome editing-based gene therapy approach. Fourth, interventions that  
473 affect progenitor growth rate, enzymatic cross-correction, and serum GAA stability (e.g., adjuvants like  
474 Duvoglustat<sup>76</sup>) could promote efficacy. Increasing serum GAA stability by altering serum GAA half-life  
475 from 4 hours (**Fig. 6i and Fig. S11-S12, grey arrowhead**) to 8 hours (**Fig. 6i and Fig. S11-S12, black**  
476 **arrowhead**) could increase the percentage of corrected cardiac tissue by 50%. Therefore, gene therapy with  
477 GAA stabilizing adjuvants may also provide a way to boost efficacy with low-efficiency editors. Finally,  
478 we recognized that there is variation within clones in their ability to secrete GAA (**Fig. 3d**), and therefore

479 modeled the therapeutic levels in the heart, assuming that double corrected liver cells produce between 0  
480 and 150% excess GAA than single corrected cells. In these simulations, there is a minimal difference due  
481 to the excess secretion of GAA, due to the small number of double-corrected cells (**Fig. 6d**).

482 Because single-edited genotypes outnumber double-edited genotypes in our GETEM results,  
483 dosing time and dosing amount are not critical in the design of therapy, as we observe limited differences  
484 in potency between a single large dose or multiple smaller doses (**Fig. S13, Supplementary Information**).  
485 Since genome editors can trigger an immune response<sup>77–80</sup>, which likely scale with the dose<sup>81,82</sup>, we have  
486 performed our simulations using multiple smaller doses of genome editors rather than a large single dose.  
487 The sensitivity analysis allowed us to consider tradeoffs in genome editor selection, progenitor targeting,  
488 genome editor delivery, and GAA stabilization via an adjuvant.



490 **Figure 6 | In vivo somatic cell gene correction strategies involve tradeoffs between efficiency,**  
491 **precision, progenitor affinity, and editor stability.** **a**, *In silico* Gene Therapy Efficacy Model (GETEM)  
492 for Pompe disease correction in a developing infant. Schematic showing gene correction for two diseased  
493 alleles in a liver indicating correction of alleles, a1 and a2, by genome editors 1 and 2 to form gene-corrected  
494 cells capable of secreting GAA to enzymatically correct other unedited cells. Secreted GAA is also  
495 absorbed by striated muscle tissue (both heart and skeletal muscle). **b**, Percentage of normal cardiac tissue  
496 within a developing heart of a Pompe diseased infant after the administration of six doses of genome editors  
497 at 23.9 mg/kg. **c**, Cell numbers indicating growth of diseased, normal, and precisely-edited cells in the gene-  
498 edited liver depot after the administration of 6 doses of genome editors at 23.9 mg/kg. **d**, Distribution of  
499 genotypes in the gene-edited liver depot after the administration of 6 doses of genome editors at 23.9 mg/kg.  
500 **e**, Sensitivity analysis of the model indicating the absolute values of parameter sensitivity of tissue  
501 morphogenesis factors, genome editor factors, and cell/tissue biology intrinsic factors. In addition to growth  
502 and differentiation rates of the edited cells, outcomes are highly sensitive to the efficiency, stability,  
503 precision, and progenitor affinity of the genome editor. These later parameters relevant to the design of the  
504 somatic cell genome editors are bolded. **f**, Tradeoff between genome editor efficiency and genome editor  
505 stability, focusing on the percentage of enzymatically cross-corrected heart tissue. Heatmap indicates that  
506 lower efficiency editors could be efficacious if the extracellular editor stability increases. **g**, Tradeoff  
507 between genome editor efficiency and precision, focusing on the percentage of enzymatically cross-  
508 corrected heart tissue. Heatmap indicates that lower efficiencies can be efficacious if higher precision  
509 editors are used, and also shows the efficiency and precision using *SpyCas9* delivered as a plasmid, *SpyCas9*  
510 delivered as an encapsulated mRNA with an AAV HDR template, *SauCas9* delivered via AAV and RA6.3  
511 base editor delivered as a plasmid **h**, Tradeoff between increasing genome editor dose and progenitor  
512 affinity, focusing on the percentage of enzymatically cross-corrected heart tissue. Heatmap indicates that  
513 lower doses can be used with delivery strategies targeting preferentially targeting progenitors. **i**, Using  
514 GETEM, heatmap indicating tradeoff in heart muscle correction in the developing infant between the  
515 degradation rates of GAA in the serum and cellular GAA, indicating that stabilization of GAA in the serum  
516 to increase its half-life by even 4 hours can improve clinical outcome (grey arrowhead indicates pre-  
517 stabilization, black arrowhead indicates post-stabilization). Results for liver and skeletal muscle are shown  
518 in Figs. S11-S12.

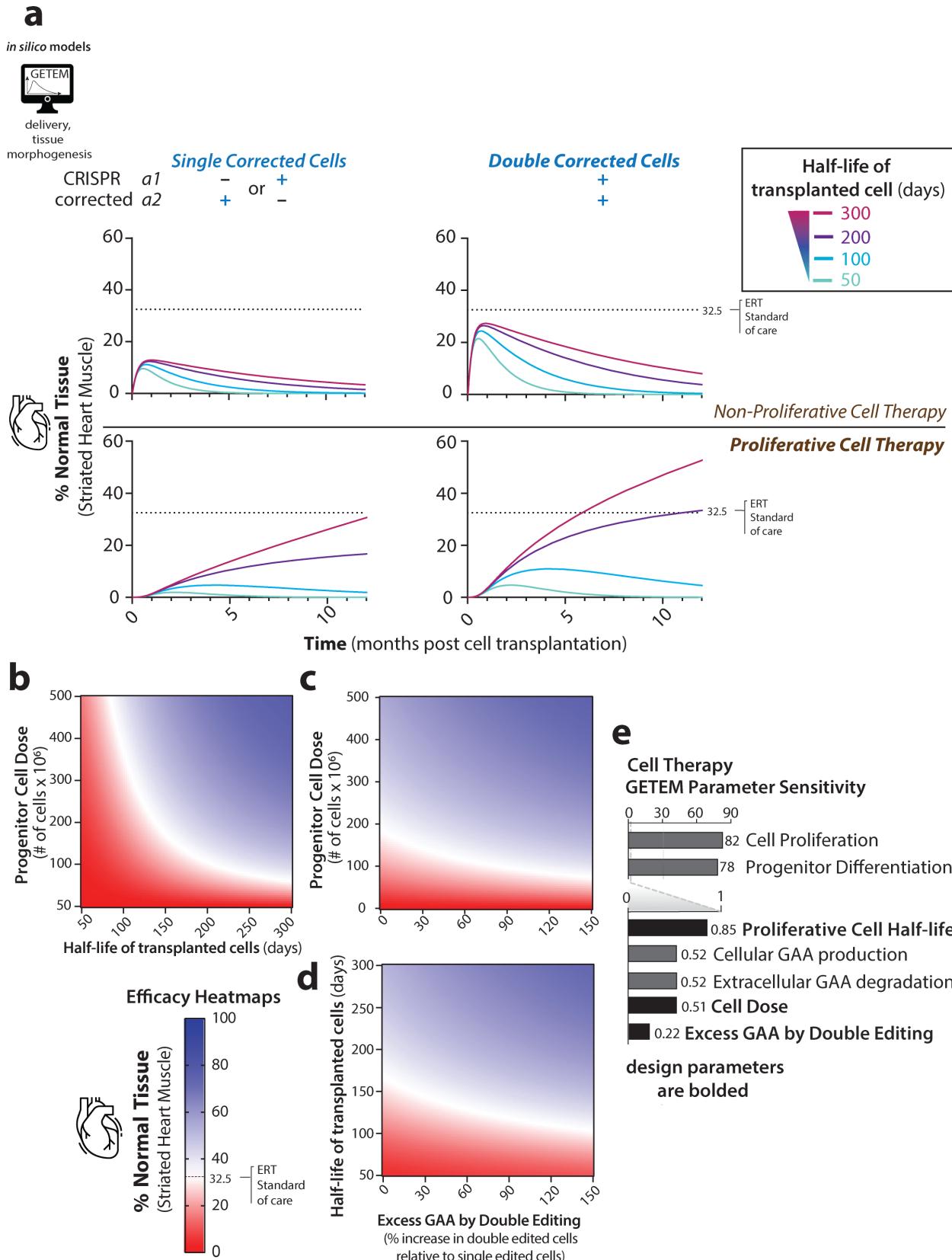
519

520 **Design of gene-corrected cell therapy with GETEM.** Autologous cell therapies involving gene-corrected  
521 cells avoid exposure of the body to genome editors *in vivo* and therefore, may be a preferred strategy over  
522 *in vivo* editing if an immune response to the editor and off-tissue or off-target effects in the body present  
523 serious safety issues. Editing both loci could provide additional therapeutic benefit, but the therapeutic  
524 benefit could depend on whether cells engraft well and retain similar function after transplantation into the  
525 body. GETEM simulations for autologous cell therapy was implemented for various cell therapeutic  
526 strategies: single-corrected progenitor cells (proliferative), double-corrected progenitor cells (proliferative),  
527 as well as single-corrected mature cells (non-proliferative) and double-corrected mature cells (non-  
528 proliferative). The half-life of engrafted functional cells was systematically varied from 50 days to 300  
529 days, the typical half-life of a hepatocyte<sup>83</sup>. For cell therapy with mature cells, no doses up to  $10 \times 10^9$  cells

530 could achieve efficacy comparable to ERT (**Fig. 7a**, see page 24). However, cell therapies with proliferative  
531 progenitor cells could result in durable, efficacious responses for a single  $0.1 \times 10^9$  cell dose if the half-life  
532 of an engrafted progenitor cell exceeds 100 days. While there is detectable therapy for cells that engraft and  
533 live for shorter periods of time, there is not a durable response, and redosing would likely be necessary.  
534 Double-corrected cells provide an added benefit for a similar dose of cells (**Fig. 7a**, assuming that double  
535 corrected cells secrete 150% excess GAA than single corrected cells) versus the same dose of single  
536 corrected cells.

537 To understand dosing strategies for cell therapy, we performed thousands of simulations varying  
538 the proliferative progenitor cell dose, half-life, and GAA production (**Fig. 7b-d**, see page 24). Sensitivity  
539 analysis on cell therapy GETEM indicates that tissue morphogenesis factors were much more sensitive  
540 (**Fig. 7e**) in comparison to cell therapy factors (dose, lifespan, and secretory phenotype of engrafted cells),  
541 or cell/tissue intrinsic factors (extracellular degradation or single allele GAA production rate). Even if the  
542 dose is raised to  $500 \times 10^6$  proliferative progenitors, if the lifespan is less than 100 days, a therapeutic  
543 outcome matching ERT is not achieved (**Fig. 7b**). Similarly, even if the dosed cells have native hepatocyte  
544 lifespans, as few as  $75 \times 10^6$  cells are required to observe a therapeutic effect matching ERT (**Fig. 7b**). If  
545 double-corrected cells engrafted well and have 150% higher production of GAA compared to single  
546 corrected cells, the dose required to reach the ERT therapeutic threshold is halved (**Fig. 7c**, assuming a 250-  
547 day half-life of the transplanted cells). Finally, engrafting cells with higher GAA secretion will match the  
548 standard of care, even if the lifespan of cells is lower, establishing a tradeoff between high GAA secretion  
549 and persistence of transplanted cells (**Fig. 7d**). Similar effects were observed on the striated skeletal muscle  
550 tissue (**Fig. S14**) and the liver depot (**Fig. S15**). These results combining our experimental data with  
551 GETEM support the conclusion that *ex vivo* cell therapy could be efficacious in addressing pathogenic  
552 polygenic mutations. Therefore, while *in vivo* dosing of genome editors is expected to deliver therapeutic  
553 effect largely through the involvement of single-corrected cells (**Fig. 6d**), a double correction strategy  
554 would be preferable for *ex vivo* cell therapy as it would reduce the dose of cells required. Finally, precisely  
555 edited cells in these models could also represent cells transduced *in vivo* with a gene augmentation therapy.

556 Overall, we believe GETEM could be easily adapted for a wide variety of somatic tissues treated with  
557 various cell and gene therapies.



559 **Figure 7 | Durable gene-corrected cell therapy for Pompe disease requires persistent proliferative**  
560 **cells** **a**, Plots showing the degree of phenotypically normal cardiac tissue when either 10 billion non-  
561 proliferative single- or double-corrected cells are dosed or 0.1 billion proliferative single- or double-  
562 corrected cells are dosed. For these plots, the GAA production for double-corrected cells is 150% excess  
563 that of single corrected cells for four different half-lives of dosed cells. For long-lasting correction,  
564 proliferative progenitor cells need to have a half-life exceeding 100 days. **b**, Heatmap showing the  
565 proliferative cell dose (assuming double corrected cells with 75% excess GAA production compared to  
566 single-corrected cells) against the half-life of dosed cells. If the dosed cells have a half-life similar to  
567 endogenous hepatocytes, a 75 million cell dose is sufficient for matching ERT. **c**, Heatmap showing the  
568 proliferative cell dose (assuming that the dosed cells have 250 day half-life) against the GAA production  
569 rate of dosed cells, demonstrating that dose has a higher effect on therapeutic efficacy than GAA production  
570 gained from double correction. **d**, Heatmap showing that the GAA production against the half-life of the  
571 dosed cells (assuming that 250 million cells were dosed), indicating that the half-life of the dosed cells has  
572 a higher effect on therapeutic efficacy than the GAA production gained from double correction. **e**,  
573 Sensitivity analysis for model parameters relevant to cell therapy. In addition to growth and differentiation  
574 rates of the transplanted cells, outcomes are highly sensitive to the proliferative cell half-life, cell dose and  
575 excess GAA produced by double gene correction. These later parameters, relevant to the design of therapy  
576 with autologous gene-corrected cells, are bolded.  
577

## 578 DISCUSSION

579 Many polygenic diseases have no animal models available for preclinical studies<sup>18</sup>, because many  
580 animal models arise from a single gene disruption. Moreover, it is currently infeasible to generate unique  
581 transgenic animal models for every patient's specific mutation, and it is difficult, if not impossible, to  
582 recapitulate the human genetic background in animal models. In contrast, human iPSCs provide an  
583 important biological substrate to evaluate genome editing approaches, as data from iPSC studies have been  
584 used to inform human clinical trials for gene therapies<sup>84,85</sup>. Using allele-specific genome editors, we  
585 demonstrate biallelic gene correction of two distinct mutations in a single iPSC. Many of the common  
586 Pompe disease mutations can be targeted in an allele-specific manner using *SpyCas9* strategies  
587 (**Supplementary Table 6**). We observe that transcriptional regulation is driven by the endogenous  
588 promoter, potentially correcting many different isoforms for *GAA*<sup>86,87</sup>. The tissue targets of genome editors,  
589 therefore, could expand from the traditional foci of liver and muscle to other tissues that may use alternate  
590 *GAA* isoforms. In contrast, in standard gene augmentation approaches, all cells must process a single  
591 isoform. Further, silencing from synthetic or viral elements has been observed for gene therapies, and in  
592 our hands with targeted knock-in strategies that overexpress a transgene via a synthetic promoter (**Fig. S8f**).  
593 Transgene silencing raises concerns about the durability of viral gene therapies and proposed cell therapies

594 where *GAA* is overexpressed from a safe harbor locus<sup>88</sup>. In our strategy, post-translational processing of the  
595 enzyme also appears to be intact, as the distribution of processed GAA is identical to healthy controls. In  
596 contrast, *GAA* overexpression in mammalian cells can cause cellular stress, leading to differential  
597 trafficking and processing of the nascent translated peptide<sup>53</sup>. Furthermore, an mRNA- or RNP-based gene  
598 correction strategy reduces insertional oncogenesis by using a non-viral approach for the delivery of the  
599 genome editor.

600 Our analysis indicates that the precision of genome editing is a key intrinsic parameter controlling  
601 efficacy for *in vivo* somatic editing approaches. Guide RNA design for CRISPR genome editors has already  
602 been identified as critical in the design of genome editors, and the genome editing community now routinely  
603 employs design tools integrating empirical and computational approaches<sup>73,89–91</sup>. Our approach  
604 complements these guide RNA design tools to consider mechanisms *in vivo* involving delivery, cell  
605 targeting, and stability of editors as well as cell biological processing of the edited protein product.  
606 Mechanistic modeling genotypic outcomes of gene correction within a growing, differentiating tissue is  
607 novel and is necessary to understand the somatic impact of genome editing. For animal studies and clinical  
608 studies, GETEM provides insight into the distribution of edited genotypes within cells in the body and  
609 could be tailored to an individual animal or patient-basis, creating “digital twins” of specific bodies  
610 undergoing somatic cell genome editing. Our *in vitro* and *in silico* results come together to inform cell  
611 therapeutic approaches, in which patient cells are *ex vivo* engineered to produce GAA from the native locus,  
612 as opposed to a safe harbor locus or viral vector<sup>92</sup>. While the predicted benefits from correcting both alleles  
613 are small within simulations of *in vivo* editing, these benefits are apparent for the design of cell therapies.  
614 GETEM establishes a quantitative basis for making tradeoffs between GAA secretion, persistence, and dose  
615 of engrafted cells while considering therapeutic effects in three different organs. Thus, further development  
616 of GETEM could facilitate rapid *in silico* evaluation of different strategies for both gene therapies and cell  
617 therapies at an organismal level. Such an expanded toolkit may ultimately reduce the number of preclinical  
618 studies needed to establish efficacy before embarking on first-in-human clinical trials for genome editing  
619 therapeutics.

620 For gene correction in the liver, we evaluated a variety of strategies using plasmid delivery using  
621 hydrodynamic injection (an approach not easily translatable to larger animals<sup>93</sup>) viral methods (delivering  
622 both the editing machinery or just the correction DNA template), which carry significant risks of viral  
623 integration<sup>26</sup>, mRNA delivery and ribonucleoprotein delivery. Our simulations were able to design  
624 efficacious dosing and formulation for Pompe disease patients for all of these strategies. When analyzing  
625 base editor editing outcomes *in vitro*<sup>46</sup> (with no selection for edited outcomes, RA6.3 base editor) at the on-  
626 target site, 45% of the on-target adenines were converted to guanines, but, 55% of the adenines upstream  
627 were also converted to guanines. Therefore, the on-target precision is about 50%, or precise-50 using the  
628 Shen et al<sup>73</sup> nomenclature, as all the other adenine to guanine edits are on-target, but imprecise. Prime  
629 editing<sup>44</sup> addresses this issue, and future development with *in vivo* somatic delivery of base and prime  
630 editors has strong potential even at low delivery efficiencies.

631 The GETEM framework is versatile and can simulate a variety animal studies involving a wide  
632 range of growth characteristics and selection pressures for a wide array of editing strategies. For example,  
633 in their demonstration of *NmeCas9* to introduce indels to knock out *Hpd* within a mouse Fah model<sup>66</sup>, the  
634 authors observe that *Hpd* indel cells are metabolically reconditioned such that they survive dietary tyrosine  
635 despite the knock-out of *Fah*. Additionally, the mice used in this experiment lost 20% of their body weight  
636 in 15 days and were between 15 – 20 weeks of age. The growth and differentiation rate constants in our  
637 GETEM could easily be modified to reflect slow growth in these mice. The authors also reported that they  
638 fed mice with NTBC from days 21 through 24 of the experiment, which can be modeled in our GETEM,  
639 and it has effects on the rate at which the fraction of the edited hepatocytes in the liver grows (**Fig. 5f**).  
640 Finally, the plasmid DNA dose was calculated using the length of *NmeCas9* + sgRNA plasmid (4790 bp),  
641 which is significantly shorter than the modified pX330 plasmid used in other *SpyCas9* studies (8484 bp).  
642 The breadth of studies successfully simulated by our GETEM approach indicates that it can likely be  
643 modified easily to accommodate selection, enzymatic cross-correction and biallelic editing for many  
644 different future gene therapy applications.

645        Multiple characteristics of a particular somatic cell genome editing therapeutic strategy are  
646    analyzed in GETEM: the enzymatic editing rate of the editor itself, the editor's precision, the delivery of  
647    the editor, and the selection of edited and unedited outcomes. When reporting the results of *in vitro* and *in*  
648    *vivo* experiments, these results are coupled and are reported as a single value indicating either the percentage  
649    of alleles edited, or the percentage of cells demonstrating phenotypic correction. Our GETEM approach  
650    decouples these constituent characteristics, enabling individual modeling of these characteristics or  
651    selectively grouped modeling of these characteristics to discover synergistic effects within these processes.  
652    It enables us to analyze how each of these interact with each other and affect the overall therapeutic outcome  
653    and suggest trade-offs for gene therapy improvement. [The monotonous growth of the percentage of edited](#)  
654    [cells in cases with and without selection of the edited phenotype \(Fig. 5e and 5f, top plots\)](#), contrasts with  
655    [the dramatic rise and fall of the percentage of the liver undergoing editing \(Fig. 5e and 5f, bottom plots\)](#).  
656    [This contrast highlights the editor-agnostic processes driving the growth of the edited phenotype in animal](#)  
657    [models.](#)

658        Current limitations of our approach arise from incomplete knowledge of several phenomena,  
659    including immune response, adverse effects linked to off-target modifications, transport of editors and  
660    edited protein products, and tissue morphogenesis. First, an immune response to the editor, delivery vector,  
661    or the edited protein product is possible. Continuous constitutive expression of *SauCas9* in hepatocytes  
662    elicits an innate immune response, which accelerates the death of the edited hepatocytes<sup>94</sup>. Using strategies  
663    with a shorter-term expression of genome editor (mRNA, or RNP) may prevent adverse events in the edited  
664    cells. For Pompe disease, patients already exhibit a variable response to the ERT, leading to variable  
665    efficacy. Both transient and long-term immunosuppressive strategies can be employed in conjunction with  
666    *in vivo* somatic editing and cell therapeutic approaches to mitigate these immune responses, and future work  
667    to expand GETEM could leverage empirical studies on the immune response both at the cellular level *in*  
668    *vitro* or *in vivo*. Second, we did not isolate any iPSC lines that harbored detectable off-target events in the  
669    top ten predicted sites, but others in the field have noted cells to undergo phenotypic changes induced by  
670    cutting the genome, such as cell cycle arrest<sup>95</sup> and upregulation of innate immune transcriptional

671 programs<sup>96,97</sup>. Further, more complex events like large deletions and translocations could occur at the on-  
672 and off-target site that would be missed by our Sanger sequencing approach<sup>50</sup>. The chromatin state of iPSCs  
673 may also alter the ability of off-target sites from being edited. However, all of the methods used to  
674 characterize off-target modifications are inherently incomplete, as it is not feasible to non-destructively  
675 achieve whole-genome sequencing of every single edited cell<sup>98</sup>. Therefore, additional studies in applying  
676 the genome editor to the target tissue of interest could be necessary to identify and model additional off-  
677 target events. Further characterization of more clones and within differentiated cells<sup>97</sup> could shed light on  
678 these low-frequency events, which become crucial for cell therapeutic approaches where upwards of 10<sup>8</sup>  
679 cells have now been implanted into patients<sup>99</sup>. Third, correction efficiencies may also only be lower in post-  
680 mitotic cells with some genome editors, and dispersion within tissues may be inhomogeneous, leading to  
681 spatial patterns of edited cells/progeny that could have variable phenotypic effects. Finally, additional  
682 empirical studies of the pharmacokinetics of editors and protein products, liver growth rates during  
683 development, glycogenolysis, and engraftment of edited cells could help us refine several assumptions  
684 made in GETEM (see **Supplemental Information**). Overall, these limitations provide insights into the  
685 types of measurements that are needed – detailed genomic analysis, cellular analysis, transport  
686 measurements and morphogenesis measurements - to enable a more predictive platform.

687 Overall, our results indicate that with appropriate engineering and design, recent advances in base  
688 editing<sup>46</sup>, delivery vectors<sup>100</sup>, nuclease decoration<sup>101</sup>, and nuclease engineering offer new possibilities to  
689 establish efficacious gene correction therapies in a streamlined fashion with reduced dependence on animal  
690 models. We, therefore, anticipate that multiple dosing with safe and precise genome editors can be  
691 developed for a greatly expanded set of targets in diseases with compound heterozygous and complex  
692 polygenic origins.

693

#### 694 **Acknowledgments**

695 We thank members of the Saha lab for helpful discussion and comments on the manuscript, Masatoshi  
696 Suzuki and Tim Kamp for sharing Pompe iPSC lines, the University of Wisconsin Biotechnology Center

697 for the use of the equipment and technical support, ArtforScience for assistance with figure art, and  
698 Aldevron for technical support with Cas9 proteins. We acknowledge generous financial support from the  
699 National Science Foundation (CBET-1350178, CBET-1645123), National Institute for Health  
700 (1R35GM119644-01), Environmental Protection Agency (EPA-G2013 –STAR-L1), University of  
701 Wisconsin Carbone Cancer Center Support Grant P30 CA014520, Wisconsin Alumni Research  
702 Foundation, and the Wisconsin Institute for Discovery.

703

#### 704 **Contributions**

705 These authors contributed equally: J.C-S. and A.D. J.C-S. and A.D planned research and analyzed data.  
706 J.C-S., A.D., and K.S. designed experiments. J.C-S., A.D., D.F., A.A., B.G. and H.K. performed  
707 experiments. J.C-S and A.D. developed the *in silico* modeling framework. T. Alam, T. Akcan and S.S.  
708 provided key insights into design of dosing and delivery for GETEM simulations. J.C-S., A.D. and K.S.  
709 wrote the manuscript with input from all authors. K.S. supervised research.

710

#### 711 **Competing interests**

712 J.C-S, A.A., L.K., and K.S. have filed a patent application on the S1mplex technology.

713

#### 714 **METHODS**

##### 715 **Cell culture**

716 All human pluripotent stem cells (hPSCs) were maintained in mTeSR1 medium on Matrigel (WiCell)  
717 coated tissue culture polystyrene plates (BD Falcon). Cells were passaged every 4-5 days at a ratio of 1:8  
718 using Versene solution (Life Technologies). Patient-derived iPSC line, Pompe GM04192, was a gift from  
719 the T. Kamp and M. Suzuki (UW-Madison) labs. Cardiomyocytes derived from hPSC and iPSC cultures  
720 were maintained in RPMI/B27 on Matrigel (WiCell) coated polystyrene plates (BD Falcon). Patient derived  
721 fibroblast lines were obtained from Coriell Institute with different GAA mutations (W746X mutation was  
722 from Coriell ID: GM04912. D645N mutation was from Coriell ID: GM20090. R660H was from Coriell

723 ID: GM13522) and cultured in DMEM supplemented with 10% FBS and 1% Penicillin/Streptomycin. All  
724 cells were maintained at 37°C in 5% CO<sub>2</sub>, and tested monthly for possible mycoplasma contamination.

725 **Cardiomyocyte Differentiation**

726 hPSCs and iPSCs were differentiated into cardiomyocytes using a small molecule-directed differentiation  
727 protocol in a 12-well plate format as previously described<sup>54</sup>. Briefly, all adherent hPSCs and iPSCs were  
728 dissociated in TrypLE solution (Life Technologies), counted with a hemocytometer, and centrifuged at 200  
729 x g for 5 minutes. Cells were plated at a density between 0.5-1x10<sup>6</sup> cells/well depending on cell line. Once  
730 tissue culture plate wells reached 100% confluency (day 0), medium in each well was replaced with a  
731 solution containing ml RPMI/B27-Insulin (Life Technologies), 12μM CHIR99021 (BioGems 25917), and  
732 1 μg/ml Insulin solution (Sigma-Aldrich I9278). Exactly 24 hours later (day 1) medium in each well was  
733 removed and replaced with RPMI/B27-insulin. Exactly 48 hours after (day 3) half of the spent medium was  
734 collected. To this, an equal volume of fresh RPMI/B27-Insulin was mixed. This combined media was then  
735 supplemented with 7.5μM IWP2 (BioGems 75844). Two days later (day 5) medium in each well was  
736 replaced with RPMI/B27-Insulin. Two days (day 7) later and every three days following, spent medium  
737 was replaced with RPMI/B27. Spontaneous contraction was generally observed between days 12-16 of  
738 differentiation.

739 **Creation of S1m-sgRNAs to correct both mutant alleles**

740 S1m-sgRNAs were synthesized as previously described<sup>75</sup>. S1m gBlocks were annealed with Phusion  
741 polymerase (New England Biolabs) under the following thermocycler conditions: 98°C for 30 sec followed  
742 by 30 cycles at 98°C for 10 s, and 72°C for 15 s with a final extension at 72°C for 10 minutes. S1m cDNA  
743 was annealed with Phusion polymerase (New England Biolabs) under the following thermocycler  
744 conditions: 98°C for 30 sec followed by 30 cycles at 98°C for 10 s, 60°C for 10 s, and 72°C for 15 s with  
745 a final extension at 72°C for 10 minutes. *In vitro* transcription was performed with the MEGAShortscript  
746 T7 Kit (Thermo Fisher Scientific) according to manufacturer's instructions. For guide RNAs for fibroblast  
747 transfection, *in vitro* transcription was performed using HiScribe T7 RNA synthesis Kit (New England  
748 Biolabs)

749 **Genome Editor Delivery**

750 All hPSC transfections were performed using the 4D-Nucleofector System (Lonza) in P3 solution using  
751 protocol CA-137. 50 pmol Cas9, 60 pmol sgRNA, 50 pmol streptavidin, and 60 pmol ssODN were used to  
752 form particles per ssODN-S1mplex as described previously<sup>75</sup>. Cells were then harvested using TrypLE  
753 (Life Technologies) and counted. 2x10<sup>5</sup> cells per transfection were then centrifuged at 100xg for 3 minutes.  
754 Excess media was aspirated and cells were resuspended using 20 µL of RNP solution per condition. After  
755 nucleofection, samples were incubated in nucleocuvettes at room temperature for 15 minutes prior to plating  
756 into 2x10<sup>4</sup> cells per well on ArrayEdit in mTeSR media+10 µM ROCK inhibitor. Media was changed 24  
757 hours post transfection and replaced with mTeSR1 medium. Fibroblast transfections were performed in 24  
758 well plates using 50,000 cells/well using 2µl Lipofectamine 2000/well (0.5µg Cas9/well and sgRNA,  
759 streptavidin and ssODN at a 1:1:1:1 molar ratio).

760 **Synthesis of ArrayEdit Platform**

761 µCP was performed using previously described methods<sup>102</sup>. The surface modification involved printing of  
762 an alkanethiol initiator to nucleate the polymerization of hydrophilic poly(ethylene glycol) (PEG) chains.  
763 Briefly, double sided-adhesive was attached to the bottom of a standard tissue culture plate, after which a  
764 laser cutter was used to cut out the well bottoms. Using previously described chemistry<sup>103</sup>, patterns were  
765 transferred to gold-coated glass via a polydimethylsiloxane stamp after which the glass was submerged in  
766 a poly(ethylene glycol) (PEG) solution overnight to build PEG chains surrounding µFeatures. Standard  
767 tissue culture plates with well bottoms cut out were then fastened to processed sheets using a custom-made  
768 alignment device.

769 **High-Content Analysis**

770 Automated microscopy was performed using a Nikon Eclipse Ti epifluorescent scope. A 15x15 grid with  
771 one µFeature per image was established and maintained so that each feature imaged was consistent each  
772 day. Nikon Perfect Focus was used to ensure that all colonies were in the same Z-plane and Lysosensor  
773 intensity was measured accurately. Images were processed using CellProfiler<sup>104</sup> to count the number of  
774 nuclei and quantify Lysosensor intensity.

775 **DNA Sequencing**

776 DNA was isolated from cells using QuickExtract DNA Extraction Solution (Epicentre) following TrypLE  
777 treatment and centrifugation. Extracted DNA was incubated at 65°C for 15 min, 68°C for 15 min, and 98°C  
778 for 10 min. Genomic PCR was performed using AccuPrime HiFi Taq (Life Technologies) or Q5 polymerase  
779 (New England Biolabs) and 500 ng of genomic DNA according to manufacturer's instructions. Long (8kb)  
780 PCR reactions were thermocycled using an extension step of 10 minutes. Genomic PCR products were then  
781 submitted to the University of Wisconsin-Madison Biotechnology Center for DNA sequencing or analyzed  
782 on an Illumina Miniseq instrument.

783 **Off-Target Analysis**

784 Candidate off-target sites were identified and ranked for both of the sgRNAs used for gene correction using  
785 a previously established algorithm<sup>105</sup>. This algorithm was validated via unbiased genome wide off-target  
786 analysis by Tsai et al.<sup>106</sup>, and the top scoring sites are listed in **Supplementary Table 4**. PCR on genomic  
787 DNA from established clonal lines was performed with primers listed in **Supplementary Table 5** flanking  
788 each of the top ten scoring candidate off-target sites. These PCR products were analyzed using Sanger  
789 sequencing by the University of Wisconsin-Madison Biotechnology Center and Genewiz.

790 **RT- and qPCR**

791 RNA was isolated from cells using QuickExtract RNA Extraction Solution (Epicentre) following  
792 manufacturer's protocol. 100 ng of extracted RNA was reverse transcribed using Superscript IV Reverse  
793 Transcriptase (Invitrogen). Endpoint PCR amplification of cDNA product was performed using 1 µL of  
794 cDNA Product and AccuPrime HiFi Taq (Life Technologies) following manufacturer's instructions.  
795 Efficacy of the endpoint PCR was evaluated via gel electrophoresis of PCR product in a 1% agarose gel.  
796 qPCR reaction was performed in triplicates for each cell line and sequence (GAPDH, 1441delT, 2237G>A,  
797 and GAA), by mixing 10 µL iTaq Universal SYBR Green Supermix (Bio-Rad), 0.5 µL sequence specific  
798 forward primer, 0.5 µL sequence specific reverse primer, 1 µL cDNA product, and 8 µL water. qPCR  
799 analysis was performed in a CFX96 Real Time PCR System under the following thermocycling conditions:  
800 95°C for 30 s followed by 35 cycles of 95°C for 5 s, and 60°C for 30 s.

801 **Next Generation Sequencing Analysis**

802 A custom python script was developed to perform sequence analysis. For each sample, sequences with  
803 frequency of less than 1000 were filtered from the data. Sequences in which the reads matched with primer  
804 and reverse complement subsequences classified as “target sequences”. Target sequences were aligned with  
805 corresponding wildtype sequence using global pairwise sequence alignment and analyzed manually for  
806 mutants and PAM wobbles. For primary fibroblast NGS data analysis, Cas-Analyzer<sup>107</sup> was used.

807 **Western Blotting**

808 Protein expression of GAA and β-Actin was determined in each cell line. Following cell lysis in ice-cold  
809 RIPA buffer supplemented with protease and phosphatase inhibitors and EDTA (5mM), protein  
810 concentration was determined (DC Protein Assay, BioRad). 40 µg of protein from each cell line was loaded  
811 into a 4-12% Bis-Tris precast gel (Criterion XT, BioRad) and gel electrophoresis performed. Proteins were  
812 then transferred to a nitrocellulose membrane and blocked in filtered 5% nonfat dry milk in TBS-T (Tris  
813 buffered saline, 0.15% Tween20) for 1 hour at room temperature. The membrane was then incubated  
814 overnight at 4°C with GAA (Abcam ab137068, 1:1000) and β-Actin (Millipore, MAB1501, 1:40,000)  
815 primary antibodies. Following the incubation period, the membrane was washed in TBS-T and incubated  
816 with appropriate horseradish peroxidase secondary antibodies (Goat Anti-Rabbit IgG, Abcam ab205718,  
817 1:2000; Anti-Mouse IgG, Cell Signaling Technologies 7076 1:20,000) for 1 hour. The membrane was  
818 washed again in TBS-T, and then developed (SuperSignal West Pico Plus Chemiluminescent Substrate,  
819 Thermo Scientific) for 5 minutes using a ChemiDoc-It2 Imaging System (UVP) and imaged.

820 **GAA Activity Assay**

821 Acid glucosidase activity was measured by hydrolysis of 4-methylumbelliferyl-D-glucoside (4-MUG,  
822 Sigma M-9766) at pH 4 to release the fluorophore 4-methylumbelliferone (4-MU) as previously  
823 described<sup>47</sup>. Briefly, 4-MUG was incubated with 10 µL protein lysate in 0.2M sodium acetate for one hour  
824 at 37 °C. Fluorescence from 4-MU was then measured using a Glomax plate reader (Promega) and activity  
825 was calculated using a standard curve.

826 **Immunocytochemistry**

827 Live cell imaging of lysosome intensity was done using LysoSensor Green (Life Technologies L7535). Dye  
828 was mixed in culture media at a 1:1000 dilution prior to adding media to wells. Cells were then incubated  
829 for 5 minutes in LysoSensor solution. Media was then aspirated and cells were washed 2x with PBS. All  
830 imaging was done within one hour of staining.

831 To assay for pluripotency markers, hPSC cultures were fixed using 4% PFA and incubated at room  
832 temperature for 10 minutes. Cells were then permeabilized using 0.05% Triton X-100 and incubated for 10  
833 minutes. Following two washes with 5% goat serum, NANOG antibody (R&D Systems AF1997, 1:200)  
834 and TRA-1-60 antibody (Millipore MAB5360, 1:150), was added to cells and incubated overnight at 4 °C.  
835 The next day, cells were rinsed twice with 5% goat serum and then incubated with a donkey anti-goat  
836 secondary antibody (Life Technologies A11055 1:500) for one hour at room temperature. Cells were then  
837 washed twice with PBS and mounted for imaging.

838 Cardiomyocyte cultures were processed in the same manner as above. After permeabilization cells were  
839 incubated with anti-sarcomeric alpha-actinin (Abcam ab68167 1:250) overnight at 4 °C. The next day, cells  
840 were rinsed twice with 5% goat serum and then incubated with a goat anti-rabbit secondary antibody (Santa  
841 Cruz Biotech sc-362262, 1:500)

#### 842 **Media Exchange**

843 Cardiomyocytes were cultured in RPMI/B27+insulin and media was exchanged every two days. As a  
844 normal media exchange, diseased and corrected cells were introduced to RPMI/B27+insulin/-glucose. 24  
845 hours post change, cells were stained with LysoSensor as described above to determine a baseline  
846 fluorescent intensity. After staining, media was replaced with media from either corrected or healthy lines  
847 and cultured for an additional 24 hours. After incubation, cells were again stained with LysoSensor and  
848 imaged using confocal microscopy.

#### 849 ***In silico* modeling: GEne Therapy Efficacy Model (GETEM)**

850 Preliminary model design was performed using COPASI 4.21<sup>108</sup>, and the final model construction was  
851 performed using MATLAB R2020a using the SimBiology© package. Ten thousand replicates<sup>72</sup> from  
852 published data were generated via the MATLAB (R2020a) bootstrap function using default options in the

853    **MATLAB software.** The SimBiology sbproj files used for the simulations (**Supplementary Table 9**) has  
854    been provided in the Supplementary Files hosted at Zenodo - <https://tinyurl.com/GETEMZenodo>  
855    (contains Supplementary Videos, Modeling Code, Sequencing Data, and Western Blot image).

856

857    **Data Availability**

858    These authors declare that all essential data supporting the conclusion of the study as well as detailed assay  
859    protocols, analytical algorithms, and customized computational codes are within the paper and  
860    supplementary materials.

861

862 REFERENCES

- 863 1. Cox, D. B. T., Platt, R. J. & Zhang, F. Therapeutic genome editing: prospects and challenges. *Nat. Med.* **21**, 121–131 (2015).
- 864 2. Lim, J.-A., Li, L. & Raben, N. Pompe disease: from pathophysiology to therapy and back again. *Front. Aging Neurosci.* **6**, 177 (2014).
- 865 3. Yin, H., Kauffman, K. J. & Anderson, D. G. Delivery technologies for genome editing. *Nat. Rev. Drug Discov.* **16**, 387–399 (2017).
- 866 4. Rui, Y., Wilson, D. R. & Green, J. J. Non-Viral Delivery To Enable Genome Editing. *Trends Biotechnol.* **37**, 281–293 (2019).
- 867 5. Davis, K. M., Pattanayak, V., Thompson, D. B., Zuris, J. A. & Liu, D. R. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat. Chem. Biol.* **11**, 316–318 (2015).
- 868 6. Hemphill, J., Borchardt, E. K., Brown, K., Asokan, A. & Deiters, A. Optical Control of CRISPR/Cas9 Gene Editing. *J. Am. Chem. Soc.* **137**, 5642–5645 (2015).
- 869 7. Ran, F. A. *et al.* Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* **154**, 1380–1389 (2013).
- 870 8. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).
- 871 9. Kleinstiver, B. P. *et al.* High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).
- 872 10. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* (2015) doi:10.1126/science.aad5227.
- 873 11. Chu, V. T. *et al.* Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* **33**, 543–548 (2015).
- 874 12. Yu, C. *et al.* Small Molecules Enhance CRISPR Genome Editing in Pluripotent Stem Cells. *Cell Stem Cell* **16**, 142–147 (2015).

- 887 13. Yang, D. *et al.* Enrichment of G2/M cell cycle phase in human pluripotent stem cells enhances HDR-  
888 mediated gene repair with customizable endonucleases. *Sci. Rep.* **6**, 21264 (2016).
- 889 14. Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome  
890 engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* **3**, e04766 (2014).
- 891 15. Gutschner, T., Haemmerle, M., Genovese, G., Draetta, G. F. & Chin, L. Post-translational Regulation  
892 of Cas9 during G1 Enhances Homology-Directed Repair. *Cell Rep.* **14**, 1555–1566 (2016).
- 893 16. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed  
894 genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat.*  
895 *Biotechnol.* **34**, 339–344 (2016).
- 896 17. Liang, X., Potter, J., Kumar, S., Ravinder, N. & Chesnut, J. D. Enhanced CRISPR/Cas9-mediated  
897 precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA.  
898 *J. Biotechnol.* **241**, 136–146 (2017).
- 899 18. Prakash, V., Moore, M. & Yáñez-Muñoz, R. J. Current Progress in Therapeutic Gene Editing for  
900 Monogenic Diseases. *Mol. Ther.* **24**, 465–474 (2016).
- 901 19. Firth, A. L. *et al.* Functional Gene Correction for Cystic Fibrosis in Lung Epithelial Cells Generated  
902 from Patient iPSCs. *Cell Rep.* doi:10.1016/j.celrep.2015.07.062.
- 903 20. Ousterout, D. G. *et al.* Correction of Dystrophin Expression in Cells From Duchenne Muscular  
904 Dystrophy Patients Through Genomic Excision of Exon 51 by Zinc Finger Nucleases. *Mol. Ther.*  
905 (2015) doi:10.1038/mt.2014.234.
- 906 21. Ribeil, J.-A. *et al.* Gene Therapy in a Patient with Sickle Cell Disease. *N. Engl. J. Med.* **376**, 848–855  
907 (2017).
- 908 22. Li, H. *et al.* In vivo genome editing restores haemostasis in a mouse model of haemophilia. *Nature*  
909 **475**, 217–221 (2011).
- 910 23. Ravin, S. S. D. *et al.* CRISPR-Cas9 gene repair of hematopoietic stem cells from patients with X-  
911 linked chronic granulomatous disease. *Sci. Transl. Med.* **9**, eaah3480–eaah3480 (2017).

- 912 24. Maeder, M. L. *et al.* Development of a gene-editing approach to restore vision loss in Leber congenital  
913 amaurosis type 10. *Nat. Med.* **25**, 229–233 (2019).
- 914 25. Stadtmauer, E. A. *et al.* CRISPR-engineered T cells in patients with refractory cancer. *Science* **367**,  
915 (2020).
- 916 26. Nelson, C. E. *et al.* Long-term evaluation of AAV-CRISPR genome editing for Duchenne muscular  
917 dystrophy. *Nat. Med.* **25**, 427–432 (2019).
- 918 27. Steyer, B. *et al.* Scarless Genome Editing of Human Pluripotent Stem Cells via Transient Puromycin  
919 Selection. *Stem Cell Rep.* **10**, 642–654 (2018).
- 920 28. Kishnani, P. S. & Howell, R. R. Pompe disease in infants and children. *J. Pediatr.* **144**, S35–S43  
921 (2004).
- 922 29. Bodamer, O. A., Scott, C. R., Giugliani, R. & Group, on behalf of the P. D. N. S. W. Newborn  
923 Screening for Pompe Disease. *Pediatrics* **140**, S4–S13 (2017).
- 924 30. Han, S. *et al.* Low-Dose Liver-Targeted Gene Therapy for Pompe Disease Enhances Therapeutic  
925 Efficacy of ERT via Immune Tolerance Induction. *Mol. Ther. - Methods Clin. Dev.* **4**, 126–136  
926 (2017).
- 927 31. Martin-Touaux, E. *et al.* Muscle as a putative producer of acid  $\alpha$ -glucosidase for glycogenosis type II  
928 gene therapy. *Hum. Mol. Genet.* **11**, 1637–1645 (2002).
- 929 32. Ding, E. y. *et al.* Long-Term Efficacy after [E1-, polymerase-] Adenovirus-Mediated Transfer of  
930 Human Acid- $\alpha$ -Glucosidase Gene into Glycogen Storage Disease Type II Knockout Mice. *Hum. Gene*  
931 *Ther.* **12**, 955–965 (2001).
- 932 33. Xu, F. *et al.* Glycogen storage in multiple muscles of old GSD-II mice can be rapidly cleared after a  
933 single intravenous injection with a modified adenoviral vector expressing hGAA. *J. Gene Med.* **7**,  
934 171–178 (2005).
- 935 34. van der Wal, E., Bergsma, A. J., Pijnenburg, J. M., van der Ploeg, A. T. & Pijnappel, W. W. M. P.  
936 Antisense Oligonucleotides Promote Exon Inclusion and Correct the Common c.-32-13T>G GAA  
937 Splicing Variant in Pompe Disease. *Mol. Ther. Nucleic Acids* **7**, 90–100 (2017).

- 938 35. van der Wal, E. *et al.* Large-Scale Expansion of Human iPSC-Derived Skeletal Muscle Cells for  
939 Disease Modeling and Cell-Based Therapeutic Strategies. *Stem Cell Rep.* **10**, 1975–1990 (2018).
- 940 36. Hendriks, W. T., Warren, C. R. & Cowan, C. A. Genome Editing in Human Pluripotent Stem Cells:  
941 Approaches, Pitfalls, and Solutions. *Cell Stem Cell* **18**, 53–65 (2016).
- 942 37. Warren, C. R. *et al.* Induced Pluripotent Stem Cell Differentiation Enables Functional Validation of  
943 GWAS Variants in Metabolic Disease. *Cell Stem Cell* **20**, 547-557.e7 (2017).
- 944 38. Saha, K. & Jaenisch, R. Technical Challenges in Using Human Induced Pluripotent Stem Cells to  
945 Model Disease. *Cell Stem Cell* **5**, 584–595 (2009).
- 946 39. Chen, I. Y., Matsa, E. & Wu, J. C. Induced pluripotent stem cells: at the heart of cardiovascular  
947 precision medicine. *Nat. Rev. Cardiol.* **13**, 333–349 (2016).
- 948 40. Torii, R., Velliou, R.-I., Hodgson, D. & Mudera, V. Modelling multi-scale cell–tissue interaction of  
949 tissue-engineered muscle constructs. *J. Tissue Eng.* **9**, (2018).
- 950 41. Matsumoto, S., Tanaka, K., Strong, D. M. & Reems, J. A. Efficacy of human islet isolation from the  
951 tail section of the pancreas for the possibility of living donor islet transplantation. *Transplantation* **78**,  
952 839–843 (2004).
- 953 42. Matsumoto, S. *et al.* Insulin independence after living-donor distal pancreatectomy and islet  
954 allotransplantation. *The Lancet* **365**, 1642–1644 (2005).
- 955 43. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting  
956 with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**, 290–296 (2020).
- 957 44. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor  
958 DNA. *Nature* (2019) doi:10.1038/s41586-019-1711-4.
- 959 45. Zafra, M. P. *et al.* Optimized base editors enable efficient editing in cells, organoids and mice. *Nat.*  
960 *Biotechnol.* **36**, 888–893 (2018).
- 961 46. Song, C.-Q. *et al.* Adenine base editing in an adult mouse model of tyrosinaemia. *Nat. Biomed. Eng.*  
962 1 (2019) doi:10.1038/s41551-019-0357-8.

- 963 47. Raval, K. K. *et al.* Pompe Disease Results in a Golgi-based Glycosylation Deficit in Human Induced  
964 Pluripotent Stem Cell-derived Cardiomyocytes. *J. Biol. Chem.* **290**, 3121–3136 (2015).
- 965 48. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-  
966 programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31**, 839–843 (2013).
- 967 49. Carlson-Steversmer, J. *et al.* High-Content Analysis of CRISPR-Cas9 Gene-Edited Human Embryonic  
968 Stem Cells. *Stem Cell Rep.* **6**, 109–120 (2016).
- 969 50. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9  
970 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* (2018) doi:10.1038/nbt.4192.
- 971 51. Popp, M. W. & Maquat, L. E. Leveraging Rules of Nonsense-Mediated mRNA Decay for Genome  
972 Engineering and Personalized Medicine. *Cell* **165**, 1319–1322 (2016).
- 973 52. El-Brolosy, M. A. & Stainier, D. Y. R. Genetic compensation: A phenomenon in search of  
974 mechanisms. *PLoS Genet.* **13**, (2017).
- 975 53. Migani, D., Smales, C. M. & Bracewell, D. G. Effects of lysosomal biotherapeutic recombinant  
976 protein expression on cell stress and protease and general host cell protein release in Chinese hamster  
977 ovary cells. *Biotechnol. Prog.* **33**, 666–676.
- 978 54. Lian, X. *et al.* Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal  
979 modulation of canonical Wnt signaling. *Proc. Natl. Acad. Sci.* (2012) doi:10.1073/pnas.1200250109.
- 980 55. Yang, H. W. *et al.* Recombinant Human Acid  $\alpha$ -Glucosidase Corrects Acid $\alpha$ -Glucosidase-Deficient  
981 Human Fibroblasts, Quail Fibroblasts, and Quail Myoblasts. *Pediatr. Res.* **43**, 374–380 (1998).
- 982 56. Ma, L., Ouyang, Q., Werthmann, G. C., Thompson, H. M. & Morrow, E. M. Live-cell Microscopy  
983 and Fluorescence-based Measurement of Luminal pH in Intracellular Organelles. *Front. Cell Dev.*  
984 *Biol.* **5**, (2017).
- 985 57. Bond, J. E., Kishnani, P. S. & Koeberl, D. D. Immunomodulatory, liver depot gene therapy for Pompe  
986 disease. *Cell. Immunol.* **342**, 103737 (2019).
- 987 58. Kishnani, P. S. *et al.* Recombinant human acid  $\alpha$ -glucosidase Major clinical benefits in infantile-onset  
988 Pompe disease. *Neurology* **68**, 99–109 (2007).

- 989 59. Raben, N. *et al.* Enzyme replacement therapy in the mouse model of Pompe disease. *Mol. Genet.*  
990 *Metab.* **80**, 159–169 (2003).
- 991 60. Wang, B., Zhao, L., Fish, M., Logan, C. Y. & Nusse, R. Self-renewing diploid Axin2<sup>+</sup> cells fuel  
992 homeostatic renewal of the liver. *Nature* **524**, 180–185 (2015).
- 993 61. Yin, H. *et al.* Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype.  
994 *Nat. Biotechnol.* **32**, 551–553 (2014).
- 995 62. Yang, Y. *et al.* A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease  
996 in newborn mice. *Nat. Biotechnol.* **34**, 334–338 (2016).
- 997 63. Yin, H. *et al.* Therapeutic genome editing by combined viral and non-viral delivery of CRISPR system  
998 components in vivo. *Nat. Biotechnol. advance on*, (2016).
- 999 64. Wei, T., Cheng, Q., Min, Y.-L., Olson, E. N. & Siegwart, D. J. Systemic nanoparticle delivery of  
1000 CRISPR-Cas9 ribonucleoproteins for effective tissue specific genome editing. *Nat. Commun.* **11**,  
1001 3232 (2020).
- 1002 65. Shin, J. H., Jung, S., Ramakrishna, S., Kim, H. H. & Lee, J. In vivo gene correction with targeted  
1003 sequence substitution through microhomology-mediated end joining. *Biochem. Biophys. Res.  
1004 Commun.* **502**, 116–122 (2018).
- 1005 66. Ibraheim, R. *et al.* All-in-one adeno-associated virus delivery and genome editing by Neisseria  
1006 meningitidis Cas9 in vivo. *Genome Biol.* **19**, 137 (2018).
- 1007 67. Sago, C. D. *et al.* High-throughput in vivo screen of functional mRNA delivery identifies  
1008 nanoparticles for endothelial cell gene editing. *Proc. Natl. Acad. Sci.* **115**, E9944–E9952 (2018).
- 1009 68. Villiger, L. *et al.* Treatment of a metabolic liver disease by in vivo genome base editing in adult mice.  
1010 *Nat. Med.* **24**, 1519–1525 (2018).
- 1011 69. Albayrak, C. *et al.* Digital Quantification of Proteins and mRNA in Single Mammalian Cells. *Mol.*  
1012 *Cell* **61**, 914–924 (2016).
- 1013 70. Azuma, H. *et al.* Robust expansion of human hepatocytes in Fah<sup>-/-</sup>/Rag2<sup>-/-</sup>/Il2rg<sup>-/-</sup> mice. *Nat.*  
1014 *Biotechnol.* **25**, 903–910 (2007).

- 1015 71. Kulesa, A., Krzywinski, M., Blainey, P. & Altman, N. Sampling distributions and the bootstrap. *Nat. Methods* **12**, 477–478 (2015).
- 1016  
1017 72. Cui, M. *et al.* Combining Nordtest method and bootstrap resampling for measurement uncertainty estimation of hematology analytes in a medical laboratory. *Clin. Biochem.* **50**, 1067–1072 (2017).
- 1018  
1019 73. Shen, M. W. *et al.* Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* **563**, 646 (2018).
- 1020  
1021 74. Harrington, L. B. *et al.* A thermostable Cas9 with increased lifetime in human plasma. *Nat. Commun.* **8**, 1424 (2017).
- 1022  
1023 75. Carlson-Stevermer, J. *et al.* Assembly of CRISPR ribonucleoproteins with biotinylated oligonucleotides via an RNA aptamer for precise gene editing. *Nat. Commun.* **8**, 1711 (2017).
- 1024  
1025 76. Kishnani, P. *et al.* Duvoglustat HCl Increases Systemic and Tissue Exposure of Active Acid  $\alpha$ -Glucosidase in Pompe Patients Co-administered with Alglucosidase  $\alpha$ . *Mol. Ther.* **25**, 1199–1208 (2017).
- 1026  
1027  
1028 77. Charlesworth, C. T. *et al.* Identification of preexisting adaptive immunity to Cas9 proteins in humans. *Nat. Med.* (2019) doi:10.1038/s41591-018-0326-x.
- 1029  
1030 78. Simhadri, V. L. *et al.* Prevalence of Pre-existing Antibodies to CRISPR-Associated Nuclease Cas9 in the USA Population. *Mol. Ther. - Methods Clin. Dev.* **10**, 105–112 (2018).
- 1031  
1032 79. Chew, W. L. *et al.* A multifunctional AAV-CRISPR-Cas9 and its host response. *Nat. Methods* **13**, 868–874 (2016).
- 1033  
1034 80. Wagner, D. L. *et al.* High prevalence of Streptococcus pyogenes Cas9-reactive T cells within the adult human population. *Nat. Med.* **25**, 242 (2019).
- 1035  
1036 81. Youm, J. W. *et al.* Antibody Responses in Mice Stimulated by Various Doses of the Potato-Derived Major Surface Antigen of Hepatitis B Virus. *Clin Vaccine Immunol* **17**, 2029–2032 (2010).
- 1037  
1038 82. Katare, Y. K., Muthukumaran, T. & Panda, A. K. Influence of particle size, antigen load, dose and additional adjuvant on the immune response from antigen loaded PLA microparticles. *Int. J. Pharm.* **301**, 149–160 (2005).
- 1039  
1040

- 1041 83. Seeger, C. & Mason, W. S. Hepatitis B Virus Biology. *Microbiol. Mol. Biol. Rev.* **64**, 51–68 (2000).
- 1042 84. Duong, T. T. *et al.* Use of induced pluripotent stem cell models to probe the pathogenesis of  
1043 Choroideremia and to develop a potential treatment. *Stem Cell Res.* **27**, 140–150 (2018).
- 1044 85. Torriano, S. *et al.* The effect of PTC124 on choroideremia fibroblasts and iPSC-derived RPE raises  
1045 considerations for therapy. *Sci. Rep.* **8**, 8234 (2018).
- 1046 86. Herzog, A. *et al.* A cross-sectional single-centre study on the spectrum of Pompe disease, German  
1047 patients: molecular analysis of the GAA gene, manifestation and genotype-phenotype correlations.  
1048 *Orphanet J. Rare Dis.* **7**, 35 (2012).
- 1049 87. Fukuhara, Y. *et al.* A molecular analysis of the GAA gene and clinical spectrum in 38 patients with  
1050 Pompe disease in Japan. *Mol. Genet. Metab. Rep.* **14**, 3–9 (2018).
- 1051 88. van der Wal, E. *et al.* Large-Scale Expansion of Human iPSC-Derived Skeletal Muscle Cells for  
1052 Disease Modeling and Cell-Based Therapeutic Strategies. *Stem Cell Rep.*  
1053 doi:10.1016/j.stemcr.2018.04.002.
- 1054 89. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of  
1055 CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
- 1056 90. Cloney, R. The oracle of inDelphi predicts Cas9 repair outcomes. *Nat. Rev. Genet.* **20**, 4 (2019).
- 1057 91. Allen, F. *et al.* Predicting the mutations generated by repair of Cas9-induced double-strand breaks.  
1058 *Nat. Biotechnol.* (2018) doi:10.1038/nbt.4317.
- 1059 92. Ronzitti, G., Collaud, F., Laforet, P. & Mingozi, F. Progress and challenges of gene therapy for  
1060 Pompe disease. *Ann. Transl. Med.* **7**, (2019).
- 1061 93. Sendra, L., Herrero, M. J. & Aliño, S. F. Translational Advances of Hydrofection by Hydrodynamic  
1062 Injection. *Genes* **9**, (2018).
- 1063 94. Li, A. *et al.* AAV-CRISPR Gene Editing Is Negated by Pre-existing Immunity to Cas9. *Mol. Ther.* **0**,  
1064 (2020).

- 1065 95. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR–Cas9 genome editing  
1066 induces a p53-mediated DNA damage response. *Nat. Med.* **1** (2018) doi:10.1038/s41591-018-0049-  
1067 z.
- 1068 96. Cromer, M. K. *et al.* Global Transcriptional Response to CRISPR/Cas9-AAV6-Based Genome  
1069 Editing in CD34+ Hematopoietic Stem and Progenitor Cells. *Mol. Ther.* **26**, 2431–2442 (2018).
- 1070 97. Sinha, D. *et al.* Human iPSC modeling reveals mutation-specific responses to gene therapy in a  
1071 genotypically diverse dominant maculopathy. *Am. J. Hum. Genet.* **in press**, (2020).
- 1072 98. Temple, A. Somatic Cell Gene Editing. *Somatic Cell Gene Editing* <https://scge.mcw.edu/>.
- 1073 99. Kondo, R. Clinical trial implementation for congenital urea cycle disorder, using human ES cells  
1074 —The world’s first successful transplantation of human ES cell-derived hepatocytes into humans.  
1075 (2020).
- 1076 100. Zhu, H. *et al.* Spatial control of in vivo CRISPR–Cas9 genome editing via nanomagnets. *Nat. Biomed.*  
1077 *Eng.* **3**, 126 (2019).
- 1078 101. Lee, B. *et al.* Nanoparticle delivery of CRISPR into the brain rescues a mouse model of fragile X  
1079 syndrome from exaggerated repetitive behaviours. *Nat. Biomed. Eng.* **2**, 497 (2018).
- 1080 102. Harkness, T. *et al.* High-content imaging with micropatterned multiwell plates reveals influence of  
1081 cell geometry and cytoskeleton on chromatin dynamics. *Biotechnol. J.* **10**, 1555–1567 (2015).
- 1082 103. Sha, J., Lippmann, E. S., McNulty, J., Ma, Y. & Ashton, R. S. Sequential Nucleophilic Substitutions  
1083 Permit Orthogonal Click Functionalization of Multicomponent PEG Brushes. *Biomacromolecules* **14**,  
1084 3294–3303 (2013).
- 1085 104. Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell  
1086 phenotypes. *Genome Biol.* **7**, R100 (2006).
- 1087 105. Ran, F. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
- 1088 106. Tsai, S. Q. *et al.* GUIDE-Seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas  
1089 nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).

- 1090 107. Park, J., Lim, K., Kim, J.-S. & Bae, S. Cas-analyzer: an online tool for assessing genome editing  
1091 results using NGS data. *Bioinformatics* **33**, 286–288 (2017).
- 1092 108. Hoops, S. *et al.* COPASI--a COmplex PAthway SImulator. *Bioinforma. Oxf. Engl.* **22**, 3067–3074  
1093 (2006).

1094