


# Design of efficacious somatic cell genome editing strategies for recessive and polygenic diseases

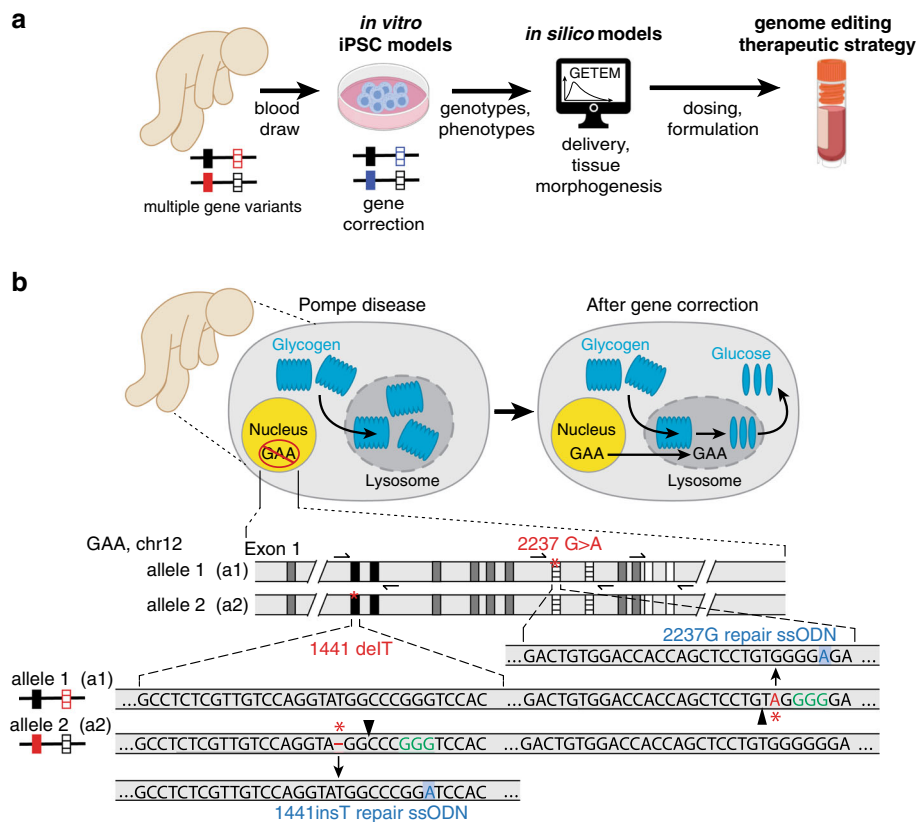
Jared Carlson-Stevermer<sup>1,2,6</sup>, Amritava Das<sup>1,3,6</sup>, Amr A. Abdeen<sup>1</sup>, David Fflis<sup>1,2</sup>, Benjamin I Grindel<sup>1,2</sup>, Shivani Saxena<sup>1,2</sup>, Tugce Akcan<sup>4</sup>, Tausif Alam<sup>4</sup>, Heidi Kletzien<sup>2</sup>, Lucille Kohlenberg<sup>1</sup>, Madelyn Goedland<sup>1,2</sup>, Micah J. Dombroe<sup>1</sup> & Krishanu Saha <sup>1,2,5</sup>✉

Compound heterozygous recessive or polygenic diseases could be addressed through gene correction of multiple alleles. However, targeting of multiple alleles using genome editors could lead to mixed genotypes and adverse events that amplify during tissue morphogenesis. Here we demonstrate that Cas9-ribonucleoprotein-based genome editors can correct two distinct mutant alleles within a single human cell precisely. Gene-corrected cells in an induced pluripotent stem cell model of Pompe disease expressed the corrected transcript from both corrected alleles, leading to enzymatic cross-correction of diseased cells. Using a quantitative in silico model for the in vivo delivery of genome editors into the developing human infant liver, we identify progenitor targeting, delivery efficiencies, and suppression of imprecise editing outcomes at the on-target site as key design parameters that control the efficacy of various therapeutic strategies. This work establishes that precise gene editing to correct multiple distinct gene variants could be highly efficacious if designed appropriately.

<sup>1</sup>Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, WI, USA. <sup>2</sup>Department of Biomedical Engineering, University of Wisconsin-Madison, Madison, WI, USA. <sup>3</sup>Morgridge Institute for Research, Madison, WI, USA. <sup>4</sup>Department of Surgery, University of Wisconsin-Madison, Madison, WI, USA. <sup>5</sup>Retina Research Foundation Kathryn and Latimer Murfee Chair, Madison, WI, USA. <sup>6</sup>These authors contributed equally: Jared Carlson-Stevermer, Amritava Das. ✉email: [ksaha@wisc.edu](mailto:ksaha@wisc.edu)

Gene therapies typically involve the editing of a single allele<sup>1</sup>, or delivery of exogenous genetic material (through nucleic acid delivery, viruses, or ex vivo engineered cells) to overexpress the gene of interest or suppress translation of the defective allele<sup>1</sup>. The development of these strategies traditionally starts with studies in animal models, however, such studies are frequently insufficient for genetic disorders that are polygenic—involving different mutations in different alleles that exacerbate a diseased phenotype. Animal models for polygenic diseases are challenging to generate, and the number of mutations implicated in a particular disease precludes the generation of animal models for every mutation (e.g., >400 for *GAA*, the causative gene for Pompe disease<sup>2</sup>). Animal models for genetic disorders have different genomes from that of humans and thus interrogating human gene therapy in animal backgrounds may leave some questions unanswered, especially those regarding off-target effects of editing strategies. Therefore, as delivery systems improve<sup>3,4</sup> and genome editors become more precise<sup>5–17</sup>, new platforms and approaches (Fig. 1a) are needed to fully understand the genotypic and phenotypic implications of editing multiple alleles within a person or a human cell.

Genome editors are routinely evaluated against a single allele within a pool of cells. Such studies have produced robust methods to understand the genomic changes, downstream gene expression, and phenotypic changes from editing a single allele, as has been demonstrated in several prior studies<sup>18–23</sup>. Some of these studies, however, reveal that unintended genomic deletions and translocations are potential outcomes at the on-target allele<sup>24</sup>, and such unintended outcomes are predicted to be exacerbated as when making multiple cuts in the genome through the delivery of two or more different genome editors<sup>25</sup>. In addition, the adeno-associated viral vectors (AAVs), commonly used to deliver genome editors, can integrate into 5% of targeted alleles<sup>26</sup>. Finally, clonal analysis of genome editing outcomes demonstrates that precise editing of a single mutant allele can generate unintended mutations in other alleles<sup>27</sup>. Thus, the genomic integrity and allelic composition of cells could be variable when attempting to edit multiple alleles, leading to variable expression of multiple alleles with a single cell. To date, precise correction of multiple alleles and an associated phenotypic rescue has not been demonstrated within a single human cell.



**Fig. 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 (GETEM model, Fig. 5). 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 *SpCas9* 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*.

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

Patient-derived induced pluripotent stem cells (iPSCs) harbor the exact mutations to be targeted by editing strategies and can recapitulate some aspects of cell and tissue pathology within affected patients<sup>36–39</sup>. Mathematical and computational tools can provide insight on somatic processes at scales larger than cells and small tissue constructs<sup>40</sup>, such as genome editor delivery, tissue morphogenesis, and physiological responses (Fig. 1a). Many genetic diseases do not require the correction of *all* diseased cells: for example, 44% of a live donor pancreas<sup>41</sup> could treat a patient with type 1 diabetes (with event-free survival of both the donor and the recipient<sup>42</sup>), and targeted correction of <1% of tissue progenitors may be sufficient to address various diseases<sup>21,22</sup>. If precise editing of multiple alleles is feasible within a single cell, then a combined in vitro–in silico approach (Fig. 1a) could help answer open translational questions regarding the efficacy of allele targeting and genome editor delivery/dosing strategies.

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

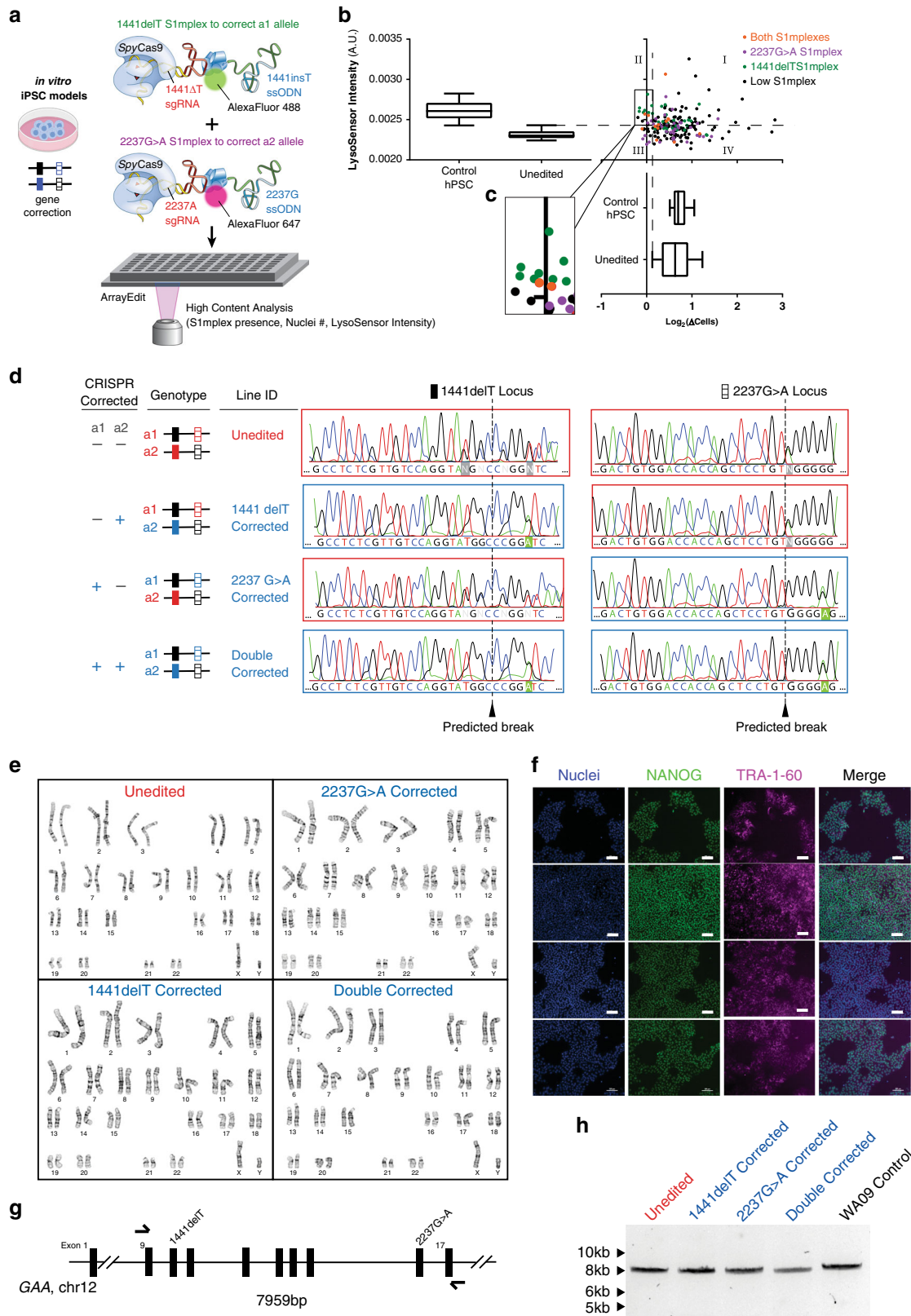
## Results

### Correction of two distinct diseased alleles within Pompe iPSCs.

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

Using a combination of S1mplex<sup>19</sup> and ArrayEdit<sup>49</sup> technologies developed by our lab, we enriched for properly-edited iPSCs after delivery of the two genome editors by tracking the presence of genome editors within the nucleus (Fig. 2b). Next, by using high-content analysis imaging of the iPSC clones during culture post-delivery of the editors, we tracked the growth rate of clones, as well as screening the pH of the lysosome<sup>47</sup> using a LysoSensor dye. LysoSensor is sensitive to the buildup of glycogen in the diseased lysosome of mutant *GAA* cells, as high glycogen phosphorolysis neutralizes this otherwise acidic organelle<sup>47</sup> (Supplementary Fig. 1 and Fig. 2a, b). Individual plotted colonies were also assayed for the presence of either genome editor (Fig. 2c, represented in either purple or green), both genome editors (Fig. 2c, red) and low amounts of genome editors (Fig. 2c, black). Colonies of interest for subsequent analysis were identified as those with high genome editor expression and lower growth rates, presumably arising from the stress of genome editing.

We isolated cell lines that were corrected at the 1441delT allele and the 2237G>A allele individually (termed ‘single-corrected’). Single corrected clones remain identical to the unedited line at the unedited locus and contained PAM wobble on the corrected allele (Fig. 2d). Sequencing chromatograms do not show evidence of undesired NHEJ products. Wobble A bases in the corrected lines are highlighted in Fig. 2d to indicate repair from the ssODN. We also isolated a clone corrected at both *GAA*:c.[1441delT];[2237G>A] alleles (Fig. 2d, termed ‘double-corrected’; clone ‘c1’). The double-corrected line contained PAM wobble at both loci. (Supplementary Fig. 2 for characterization of the corrected lines. After karyotyping each of the isolated lines, we observed no large transversions or inversions (Fig. 2e) and verified that all gene-corrected lines remained pluripotent (Fig. 2f). Because genome editing can create large indel mutations<sup>50</sup>, we also conducted an 8 kb PCR on *GAA* that included both sgRNA target sites and observed no genomic deletions between the sgRNA target sites (Fig. 2g, h). Sequencing of these large PCR amplicons confirmed that both alleles were present, and no other sequence abnormalities were detected at the edited loci (Supplementary Fig. 2). Finally, chromatograms from Sanger sequencing at the top ten off-target sites for each sgRNA matched the untransfected, patient-derived cell line, indicating that none of the top ten off-target regions were modified by our editing strategy (Supplementary Figs. 3 and 5; Supplementary Table 4). To ensure that our observations were not specific to one clone, five additional double-corrected lines (Supplementary Fig. 7; clones c21, c28,

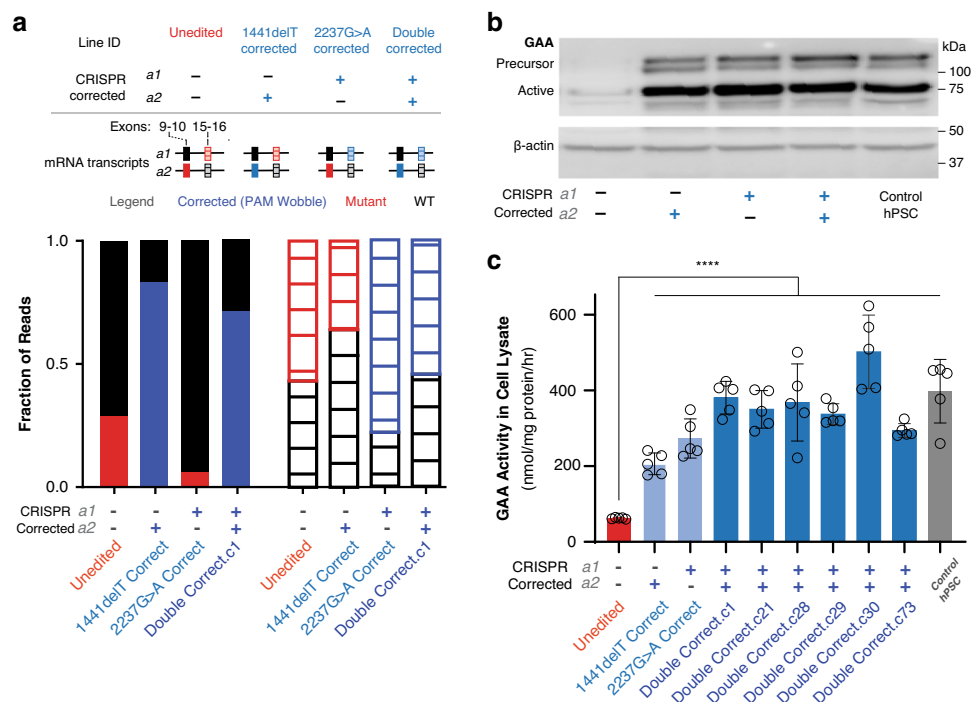


c29, c30, and c73) were generated using the S1plexes shown in Fig. 2c or by a transient puromycin treatment protocol<sup>27</sup>.

Quantitative RT-PCR (qRT-PCR) at the 3' end of the *GAA* mRNA transcript, as well as around each edited locus (Fig. 3a, Supplementary Fig. 6), indicated that the corrected loci were correctly expressed. We observed that the unedited line expressed

the lowest levels of *GAA* transcripts when compared to internal *GAPDH* levels (Supplementary Fig. 6), despite the presence of full-length, mature mRNA that could be used to express the protein. Both the single- and double-corrected lines also expressed mature *GAA* transcripts. By looking for the presence of disease variants and protoadjacent motif (PAM) wobbles

**Fig. 2 Gene correction of two distinct mutant alleles within a single human cell.** **a** ArrayEdit-based isolation of iPSC clones corrected at either or both loci. CRISPR S1mplex design for the gene correction of compound heterozygous mutations. S1mplexes targeting 1441delT mutant were labeled with an AlexaFluor488 compound while S1mplexes targeting the 2237G>A mutation were labeled with an AlexaFluor647. These ribonucleoprotein genome editors were mixed prior to transfecting into cells and subsequently plated on the ArrayEdit platform to conduct high-content analysis (see Supplementary Fig. 1). **b** Left: LysoSensor quantification per  $\mu$ Feature of two mock transfections after 7 days of growth. Normal control hPSCs were significantly more intense than unedited, Pompe diseased iPSCs on ArrayEdit. Bottom right: the growth rate of unedited and control hPSCs following a mock transfection to establish a baseline for growth. Growth rates were calculated by measuring the per-day change in the number of cells of the  $\mu$ Feature. Top right: LysoSensor intensity was plotted against growth rate per  $\mu$ Feature to identify edited colonies. Dashed lines indicate regions of interest. ( $n = 145$  independent cell lines). **c** Magnification of quadrant II from (b).  $\mu$ Features in this region were selected for genomic analysis to isolate edited clones. ( $n = 17$  independent cell lines). **d** Sanger sequencing traces of corrected cell lines. The unedited line contains mutations at both alleles: 1441delT mutation causes a breakdown of sequence trace, whereas a single point mutation demonstrates a heterozygosity 2237G>A locus. *SpyCas9* cut site is denoted by a dotted line. **e** Karyotypes of all isolated gene-corrected lines as well as unedited cells. No abnormalities were detected at a band resolution of 500. **f** Immunocytochemistry of pluripotency markers in gene-corrected lines. All lines were positive for pluripotency markers NANOG and TRA-1-60 (scale bar: 100  $\mu$ m). **g** Schematic of long PCR covering both *SpyCas9* cut sites. Arrows denote primers. The expected PCR amplicon is 7959 bp in length. **h** Gel analysis of long-range PCR described in (g) in each isolated cell line. No significant deviations from the expected length were detected, and no other notable bands were observed. WA09 control cells are hPSCs.



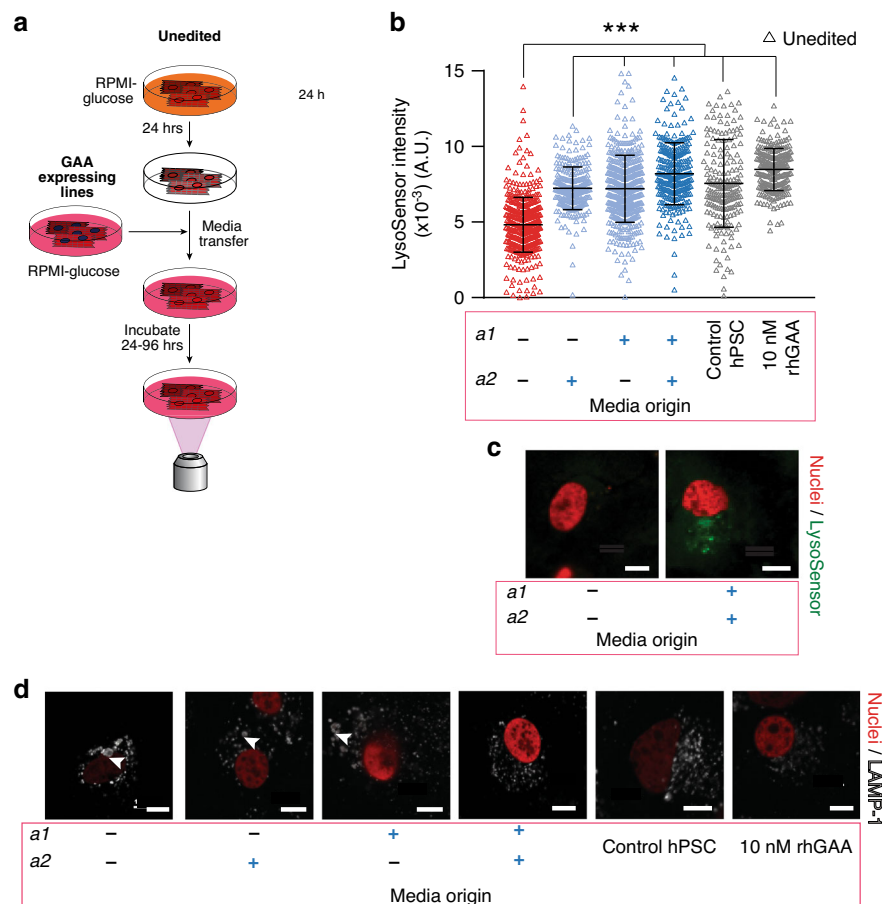
introduced by the ssODN (Fig. 1b) via deep sequencing on endpoint PCR samples of mRNA, we observe that both alleles are expressed individually at higher levels (3–5 fold increase) than unedited cells (Supplementary Fig. 6). Each allele is expressed similarly to the corresponding single corrected line (Supplementary Fig. 6b, Supplementary Information). These findings suggest nonsense-mediated decay of the mutant

transcript<sup>51</sup> or cellular compensation<sup>52</sup> to overcome the mutant allele within the single corrected lines. We detected active GAA protein using a Western blot (Fig. 3b) at levels comparable to a control hPSC line. We were also able to identify precursor polypeptides, which are important for protein secretion<sup>53</sup>, showing the GAA transcripts from the edited alleles are correctly translated and processed within cells. Notably, we were able to

detect only small amounts of GAA protein and precursor polypeptides in the unedited iPSCs. All edited cell lines (Supplementary Fig. 7) were able to produce (Fig. 3c) and secrete active GAA (Fig. 3d and Supplementary Fig. 6d).

**Enzymatic cross-correction by gene-corrected cells.** Detection of active GAA secretion by the edited cells led us to test the potential of edited cells to enzymatically cross-correct diseased cells (Fig. 4a). Because Pompe disease has a significant effect on cardiac tissue in infants, we differentiated iPSC lines to cardiomyocytes (Pompe iPSC-CMs) using a previously described small-molecule inhibitor protocol<sup>54</sup> (Fig. 4a). For all differentiations, we observed spontaneous contraction (Supplementary Videos) and confirmed the expression of  $\alpha$ -actinin, a marker of cardiac lineage commitment (Supplementary Fig. 8a). Similar to results seen in the iPSC state, differentiated corrected lines still expressed and secreted active GAA, as indicated in a 4-MUG cleavage assay on cardiomyocyte protein lysates and spent culture media (Supplementary Fig. 8b–e, Supplementary Information). It has previously

been demonstrated that by culturing in medium devoid of glucose, Pompe iPSC-CMs display an accumulation of glycogen within the lysosome<sup>47</sup>. We performed a medium exchange experiment wherein we took a partially spent, glucose-free medium from each corrected line (putatively containing secreted active GAA) and used it to replace glucose-free medium on unedited Pompe iPSC-CMs (Fig. 4a). One day after this media exchange, cells were stained with LysoSensor, and subsequent confocal microscopy was used to measure lysosome acidity as a proxy for glycogen clearance. As a control, we added rhGAA to unedited Pompe iPSC-CMs to simulate ERT. When unedited cardiomyocytes were supplemented with 10 nM rhGAA (ERT), LysoSensor intensity increased, indicating clearance of glycogen from the lysosome. Media from all edited cells were able to recover the lysosomal pH at 96 h (Fig. 4b), and this clearance is expected to continue until normal levels of glycogen were reached<sup>55</sup>. Within these cultures, we qualitatively observed lysosomal size of unedited Pompe iPSC-CMs in GAA-positive media through visualization of Lysosomal Associated Membrane Protein 1, (LAMP-1). Detailed quantification of lysosome size was not



**Fig. 4 Enzymatic cross-correction of unedited cardiomyocytes by gene-corrected cardiomyocytes.** **a** Schematic of enzymatic cross-correction experiments using gene-corrected cardiomyocytes. Unedited iPSC-CMs (red) were supplied media without glucose for 24 h (orange). After 24 h, media was replaced with media (pink) that had previously been exposed to corrected cell lines (blue) or supplemented with rhGAA. 24–96 hours after replacement, unedited cells were stained with LysoSensor and imaged using confocal microscopy for dye intensity. **b** Quantification of LysoSensor intensity in cross-corrected lines 96 h post media exchange. Each triangle represents a corrected cell identified using CellProfiler. After 96 h of daily media changes or supplementing with rhGAA, all conditions had a significant increase in dye intensity over control conditions. Unedited cells were modified to express histone 2B (H2B)-mCherry to facilitate imaging of the nuclei in these assays. See also Supplementary Fig. 7. ( $***p < 10^{-15}$ ,  $n = 262, 201, 261, 249, 135,$  and  $302$  independent lysosomes respectively per condition, as detailed in Supplementary Table 7, two-tailed  $t$ -test,  $\alpha = 0.05$ , heteroscedastic; mean  $\pm$  s.d.). **c** Representative images of unedited iPSC-CMs stained with LysoSensor in media from unedited and double-corrected iPSC-CMs. **d** Representative images of LAMP1 staining in unedited, single corrected, double-corrected cells and control PSC-CM and unedited iPSC-CM treated with rhGAA. (scale bars: 10  $\mu$ m). Select enlarged LAMP1-positive lysosomes are identified by white arrowheads.

possible, as the pH-sensitive dye utilized in this method is prone to photobleaching as noted previously<sup>56</sup>. In media from unedited cells, lysosomes were enlarged, consistent with the buildup of glycogen (Fig. 4d). In comparison, when media was taken from double-corrected cells or supplemented with rhGAA, lysosomes appeared as punctae. Samples from single corrected cells fell between these two extremes. Taken together, the single- and double-corrected cells enzymatically cross-correct diseased cardiomyocytes quickly and effectively.

### Establishing a Pompe disease gene therapy efficacy model.

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

First, we simulated standard ERT to ground our model with empirical clinical data gathered from infants and children treated with ERT<sup>58</sup>. This grounding enables our Pompe GETEM to account for enzymatic cross-correction (Fig. 5a) properly. Despite high levels of GAA in the liver on ERT, GAA transport to and uptake within the distal muscle can be low<sup>59</sup>, and this is accounted for in Pompe GETEM through a loss factor (Supplementary Information). Liver progenitors in this model are proliferative and give rise to non-proliferative mature cells<sup>60</sup>. Both progenitor and mature cells can absorb extracellular GAA to be enzymatically cross-corrected from a diseased to a normal state, and subsequently reverted to the diseased state as GAA degrades. In distal cardiac and skeletal muscle tissues, the tissue absorbs GAA for enzymatic cross-correction to a normal phenotype (Supplementary Information). Biweekly intravenous ERT doses<sup>58</sup> lead to an oscillatory percentage of phenotypically normal cells in the heart, skeletal muscle, and liver within our model (Fig. 5b, Supplementary Fig. 9), reaching 32.5% normal cells in the heart on average after one year, similar to levels observed with ERT, enough for significant heart glycogen clearance<sup>59</sup>.

### Empirical genome editing data to ground and validate GETEM.

Because genome editing data of the human GAA within patients *in vivo* or within any animal model are not available, we use empirical data from previously published genome editing experiments targeting a single endogenous gene with mice. These studies provide quantitative preclinical data that ground the delivery mechanisms to the liver within GETEM. While several studies have performed precise gene correction studies in the mouse livers, they have typically delivered their editor of choice using a plasmid<sup>46,61</sup>, a virus<sup>62</sup> or an encapsulated mRNA (in conjunction with a virus for the donor correction template)<sup>63</sup>. In GETEM, we consider all these delivery strategies to produce RNPs that ultimately produce precise and imprecise genome edits

within a cell (Fig. 5c). We then utilized the published empirical data from plasmid DNA, mRNA and viral delivery to train GETEM ( $n = 18$  mice for training, three different genome editors), and subsequently validated GETEM using data from RNP delivery into the liver<sup>64</sup> and four more liver editing studies<sup>65–68</sup> ( $n = 29$  mice for validation; Fig. 5d; details on mouse studies in Supplementary Table 8).

To account for the different delivery strategies, a previously reported model of the central dogma<sup>69</sup> was employed to calculate the amounts of transcribed and translated RNP. Either (1) plasmid DNA transcription and mRNA translation, (2) mRNA translation, or (3) AAV-mediated transcription and translation led to the formation of editor proteins. In comparison to sgRNA transcription, Cas9 transcription or translation was assumed to be rate limiting for the formation of the RNP when Cas9 is delivered as an mRNA or encoded on DNA. AAV doses were converted to an effective viral DNA dose based on the published vector copies per cell data provided<sup>62</sup>. The degradation of the plasmid or viral DNA occurs constitutively as previously measured with mouse macrophages<sup>69</sup>. For studies that targeted the *Fah* allele, a selection pressure enriches for edited cells Fig. 5e. The dynamics for the enrichment has been well characterized through prior experimental studies<sup>70</sup>. For these studies, the unedited cells die quickly unless the mouse is continuously fed 2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione (NTBC). Once corrected, the liver cells can survive without NTBC supplementation. Within our model, delivery of genome editors to livers within *Fah*<sup>-/-</sup> mice can target a mutant fumarylacetoacetate hydroxylase allele to correct a mutated stop codon. Both *Fah*<sup>-</sup> progenitor and mature hepatocytes can be edited to form *Fah*<sup>+</sup> hepatocytes, and imprecisely edited *Fah*<sup>-</sup> hepatocytes. The *Fah*<sup>-</sup> hepatocytes are subject to a death rate which is modified by the presence of NTBC, and thus this model incorporates the growth advantage of precisely edited cells.

GETEM considers the rate of genome editing,  $\frac{dE}{dt}$ , to be a second-order mass action rate equation dependent on both the number of RNPs,  $N_{RNP}$ , and the number of cells,  $N_{Cell}$  (Eq. (1)), where  $k$  is the genome editing rate constant:

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

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

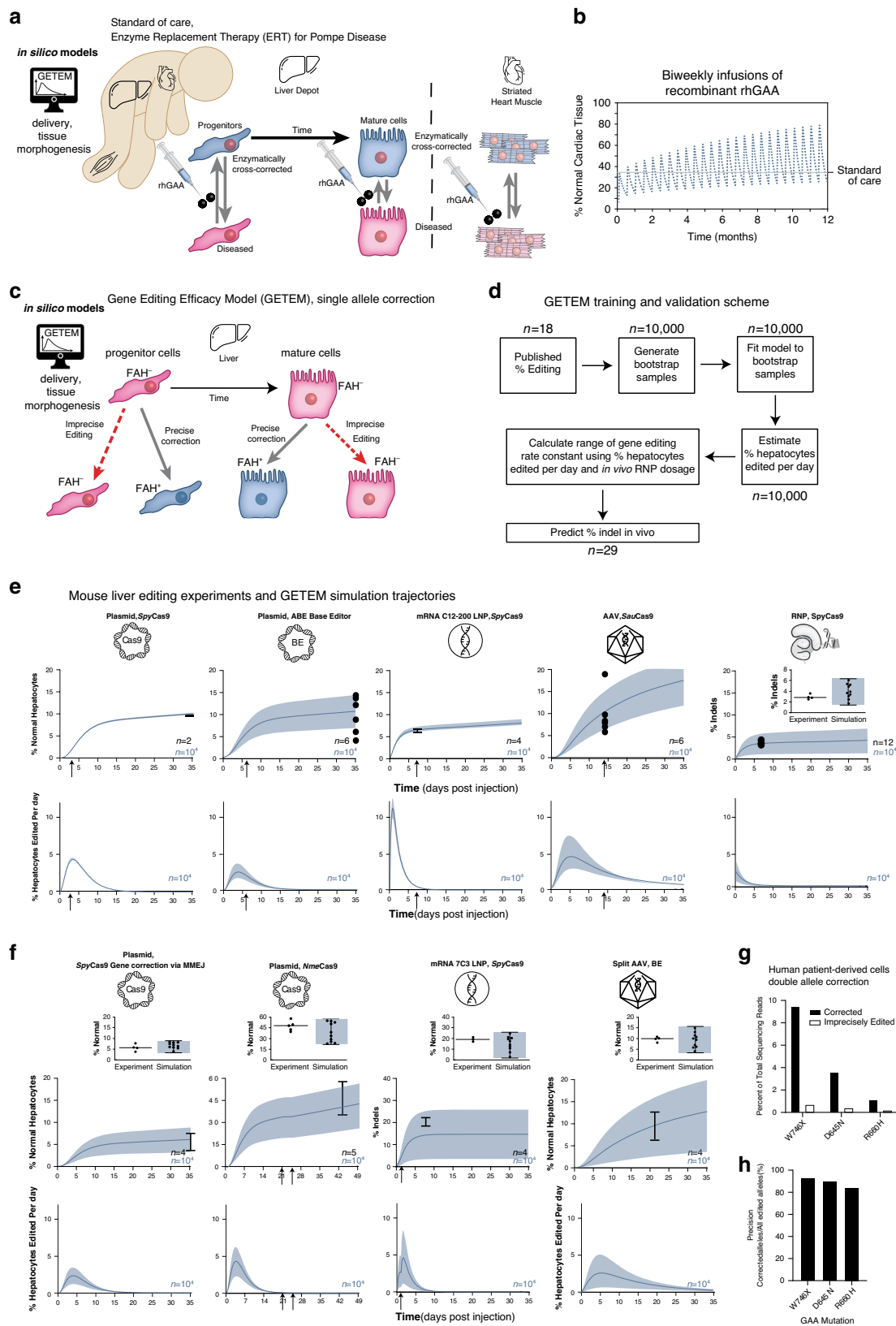
$$\frac{dE}{dt} \sim \frac{\Delta E}{\Delta t} = \frac{E_n - E_{n-1}}{t_n - t_{n-1}} \quad (2)$$

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

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

For each published study in the training dataset ( $n = 18$  mice), random sampling of the experimental data by bootstrapping<sup>71</sup> produced ten thousand samples that constituted a surrogate “bootstrapped” data set. For each of these bootstrapped samples, the average rate of liver genome change,  $\overline{\frac{dE}{dt}}$ , was calculated individually (Fig. 5d). The mean and standard deviation of this set of bootstrapped  $\overline{\frac{dE}{dt}}$  values was utilized in GETEM to simulate the range of outcomes for the mice in each study in the training set (Fig. 5e, left four panels).

Training studies employed three different genome editors delivered using four different methods: plasmid delivery of *SpyCas9*<sup>61</sup>, gRNA, and repair template in a *Fah*<sup>-/-</sup> mouse model;



plasmid delivery of base editor and gRNA in a *Fah*<sup>-/-</sup> mouse model<sup>46</sup>; mRNA delivery of *SpyCas9* (C12-200 lipid nanoparticle) and gRNA and AAV delivery of repair template in a *Fah*<sup>-/-</sup> mouse model<sup>63</sup>; and, AAV delivery of *SpyCas9*, gRNA, and repair template to edit Ornithine transcarboxylase (*Otc*)<sup>62</sup>. Ten thousand simulated samples were generated from the empirical

data for each mouse via bootstrapping<sup>71</sup>, and the GETEM was fit to each of these bootstrapped samples to calculate mean ± 1 s.d. for the gene-editing rate constant via Eqs. (1)–(3). The trained GETEM shows strong concordance of the simulation with experimental data at the assay timepoints for each study (black experimental data is within the blue range of simulation



**Fig. 5 Derivation, empirical training, and validation of in silico gene therapy efficacy model (GETEM) for Pompe disease correction in a developing infant.** **a** Schematic showing enzymatic cross-correction of somatic tissues by injected rhGAA (enzyme replacement therapy, ERT). For a given developing liver of a patient, progenitor cells proliferate and can asymmetrically divide into mature cells. Injected rhGAA is absorbed by the liver as well as striated muscle tissue (both heart and skeletal muscle). **b** Percentage of normal cardiac tissue within a heart of a Pompe diseased infant after one year of biweekly ERT at 20 mg/kg. Percentages of normal skeletal muscle tissue and hepatocytes are in Supplementary Fig. 9. **c** Schematic for in silico gene editing model for previously published precise correction of a single base pair in mouse disease models, **d** Flowchart for training and validating the GETEM model using previously published studies of in vivo somatic cell genome editing of a mouse liver by intravenous injection. **e** (top) Absolute change of edited cells over time predicted by the GETEM for the studies utilized in (**d**). For studies that employed selection in the liver, arrows indicate the time at which the selective pressure favoring edited cells was applied (for *Fah*<sup>-/-</sup> models, this indicates the removal of NTBC supplementation in the diet; and, for *Otc* editing, this indicates the induction of a high protein diet.). **f** Validation of the model using additional published editing strategies. Arrows on days 21 and 24 indicate removal of selection pressure on the treated mice as per published experimental protocol. Arrow on day 2 in the mRNA LNP indicates redosing of the LNP. (*n* represents biological replicates as previously reported in the literature, mean  $\pm$  s.d.) **g** Percentage of total sequencing reads from primary fibroblasts of Pompe diseased patients that were treated with S1plexes targeting W746X D645N or R660H GAA mutations. Results indicate gene correction and imprecise editing for three different mutations. **h** Percentage of edited alleles that are precisely edited in (**g**).

outcomes). The lines in Fig. 5e indicate the mean and standard deviation of edited alleles in the mice cohorts, while dots signify the percentage edited alleles of individual mice for studies in which individual replicate values were reported.

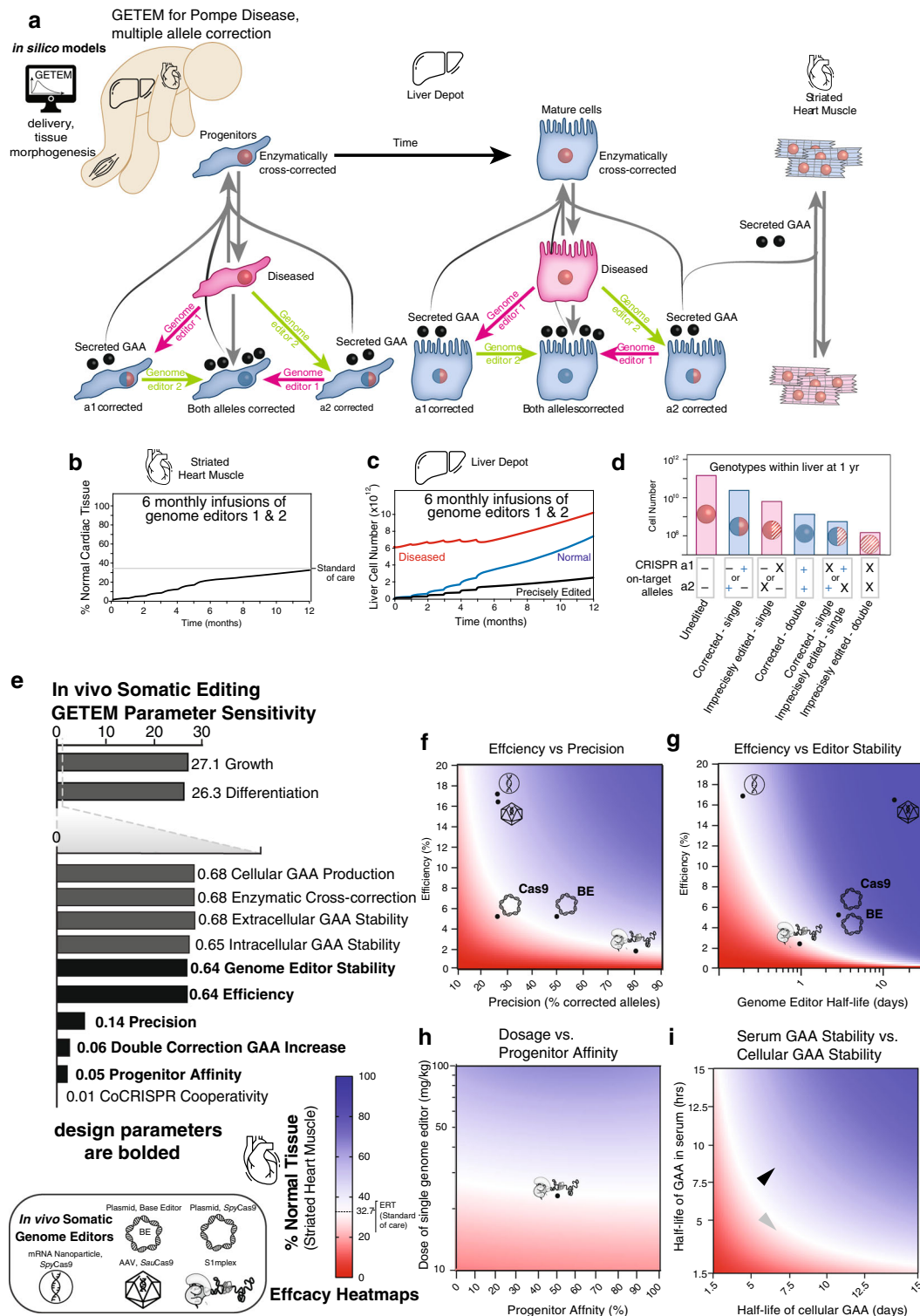
To validate the model, the bootstrapped  $\frac{dE}{dt}$  values from all the training studies were combined. The mean and one standard deviation of the combined set of surrogate  $\frac{dE}{dt}$  values were used to simulate the genome editing outcomes for RNP doses employed by a separate study<sup>64</sup> outside of the training dataset (Fig. 5e, right). There is strong concordance of the experimental data with the results of the simulation at the assay timepoint, day seven (Fig. 5e, inset), as the percent of edited alleles at the assay timepoint (black) nearly all fall within the range of predicted percent of edited alleles (blue). Lower panels in Fig. 5e, f indicate the rate of liver genome editing for each individual bootstrapped sample (*n* = 10,000 per panel) over time for the various mRNA, plasmid, and AAV studies.

To further validate GETEM, we modeled other somatic cell genome editing strategies evaluated in mouse livers—microhomology based gene correction using *SpyCas9*<sup>65</sup>, *Neisseria Meningitidis* Cas9 (*NmeCas9*) based editing<sup>66</sup>, a different type of lipid nanoparticle (7C3 mRNA Lipid Nanoparticle) delivering *SpyCas9* mRNA<sup>67</sup> and AAV delivery of split base editors BE3<sup>68</sup>. The simulation results for all these strategies showed concordance with the experimental data (Fig. 5f). In total, the validation data set includes 29 mice across four different delivery strategies with four different editing strategies (Fig. 5d–f).

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

In order to extend GETEM—already validated for single allele editing—to biallelic gene correction for Pompe disease, additional training data is required on the precision of biallelic genome editing and their cellular phenotypic consequences. We performed additional in vitro experiments with four different pathological mutations in GAA in Pompe patient-derived cells (Fig. 5g, h). Precise gene correction with S1plexes occurred approximately in 84–93% of the edited sequence reads at the target locus as assayed by deep sequencing of genomic DNA. The remaining are imprecise, unintended modification to the on-target site, which could destroy the PAM or modify the on-target site for subsequent editing of these alleles. According to the nomenclature of Shen et al.<sup>72</sup> to describe this ratio of gene correction to other editing outcomes, S1plexes are approximately “precise-80.” In comparison, normal *SpyCas9* is precise-50, and prime editors<sup>44</sup> are precise-90. Next, experimental measurements of functional outcomes from this distribution of edited cells ground protein levels in the Pompe GETEM. Cells corrected at both alleles a1 and a2 have been modeled to secrete the same amount of GAA than those edited at a single allele (Fig. 3d) at first approximation. We include a parameter in our cell therapy GETEM to explore the effects of higher secretion arising from double-corrected cells. In summary, high-resolution measurements of alleles after genome editing, as well as quantitative measures of protein levels and activity post-editing, provide key empirical rate constants for the training of Pompe GETEM.

**Design of in vivo treatment informed by Pompe GETEM.** The power of building a computational model is that many different doses can be quickly simulated, providing insights into the spectrum of genotypes with the tissues over time as well as with the effects of this gene-corrected tissue on other parts of the body. Building on the model in Fig. 5, we generated a model that considers ERT along with gene editing outcomes after the administration of genome editors targeting multiple alleles (Fig. 6a). First, we simulate doses of genome editors required to achieve or exceed efficacy equivalent to ERT standard of care. For S1plexes, the dosages of RNP to reach comparable efficacy as ERT were evaluated using the editing rate established in Fig. 5. The S1plex dose was evaluated to be 23.9 mg/kg/allele to reach equivalent healing in the liver using 6 monthly doses via intrahepatic injections, starting at birth, scaled using the 50th percentile growth of a male infant. This dose represents the amount of S1plex that reaches the hepatocytes, and not the amount that may need to be injected systemically. At one year after the first injection, the normal cells in the heart reach 32.5%, the standard calculated from the ERT simulation (Fig. 6b), and the normal cells in the liver consist of 34%



genotypically corrected cells (corrected at alleles a1, a2 or both, Fig. 6c), and 66% enzymatically cross-corrected cells (Fig. 6c). Genome editing post addition of genome editor is a relatively quick process, and the percentage of edited cells stabilizes within five days (Supplementary Fig. 10). However, enzymatic cross-correction via secreted and absorbed GAA is a slower process, and continues beyond the five day gene-editing period (Supplementary Fig. 10). Single-corrected cells constitute the most prevalent edited genotype (Fig. 6d).

Sensitivity analysis (Fig. 6e) indicated that tissue morphogenesis factors and cell/tissue biology factors also control efficacy (Supplementary Information). These intrinsic biological properties of each patient may offer opportunities to tailor therapies to each infant. The sensitivity analysis helps clarify trade-offs required for gene therapy decisions. For instance, when considering increasing the magnitude of benefit by editing both alleles, or editing one allele with higher efficiency—a lower sensitivity is observed for increasing GAA production within a

**Fig. 6 In vivo somatic cell gene correction strategies involve tradeoffs between efficiency, precision, progenitor affinity, and editor stability.** **a** In silico Gene Therapy Efficacy Model (GETEM) for Pompe disease correction in a developing infant. Schematic showing gene correction for two diseased alleles in a liver indicating correction of alleles, a1 and a2, by genome editors 1 and 2 to form gene-corrected cells capable of secreting GAA to enzymatically correct other unedited cells. Secreted GAA is also absorbed by striated muscle tissue (both heart and skeletal muscle). **b** Percentage of normal cardiac tissue within a developing heart of a Pompe diseased infant after the administration of six doses of genome editors at 23.9 mg/kg. **c** Cell numbers indicating growth of diseased, normal, and precisely-edited cells in the gene-edited liver depot after the administration of 6 doses of genome editors at 23.9 mg/kg. **d** Distribution of genotypes in the gene-edited liver depot after the administration of 6 doses of genome editors at 23.9 mg/kg. **e** Sensitivity analysis of the model indicating the absolute values of parameter sensitivity of tissue morphogenesis factors, genome editor factors, and cell/tissue biology intrinsic factors. **f** Tradeoff between genome editor efficiency and genome editor stability, focusing on the percentage of enzymatically cross-corrected heart tissue. Heatmap indicates that lower efficiency editors could be efficacious if the extracellular editor stability increases. **g** Tradeoff between genome editor efficiency and precision, focusing on the percentage of enzymatically cross-corrected heart tissue. Heatmap indicates that lower efficiency editors can be efficacious if higher precision editors are used. **h** Tradeoff between increasing genome editor dose and progenitor affinity, focusing on the percentage of enzymatically cross-corrected heart tissue. **i** Using GETEM, heatmap indicating tradeoff in heart muscle correction in the developing infant between the degradation rates of GAA in the serum and cellular GAA, indicating that stabilization of GAA in the serum can improve clinical outcome (gray arrowhead indicates pre-stabilization, black arrowhead indicates post-stabilization). Simulation results for liver and skeletal muscle are shown in Supplementary Figs. 11 and 12.

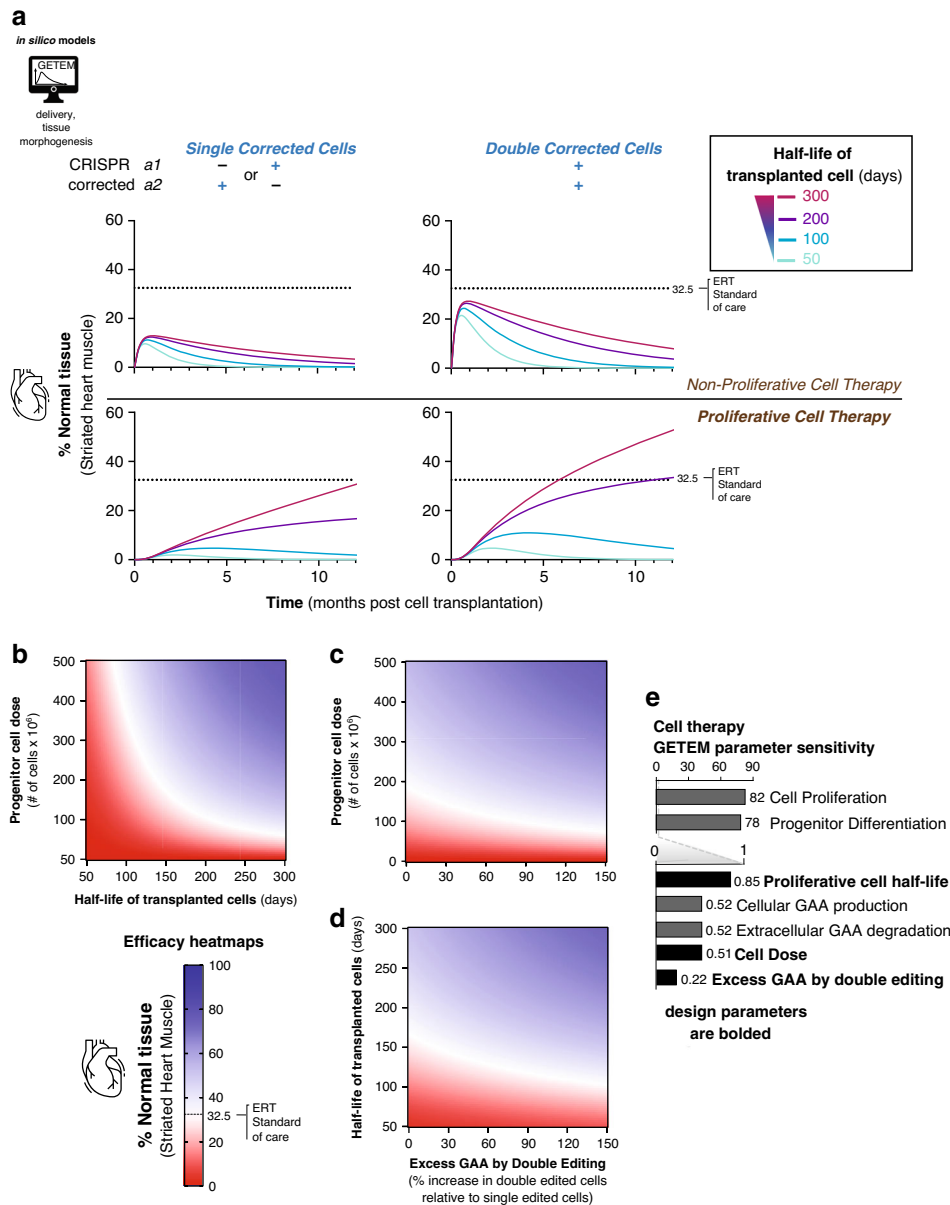
double-corrected cell (0.06) relative to a single-corrected cells than editing efficiency or precision (~0.64). This lower sensitivity indicates that editing a single allele at higher efficiency is likely to be more efficacious, as a sensitivity of 0.06 implies that a 1% increase in GAA production caused by biallele editing would lead to a 0.06% increase in percentage of normal cardiac tissue, whereas, the higher sensitivity of editing efficiency and precision of 0.64 implies that a 1% increase in either editing efficiency or precision would increase the percentage of normal cardiac tissue by 0.64%. Therefore, for S1plexes, there is higher expected benefit in editing a single allele more efficiently, than in editing both alleles at a lower efficiency. In addition to growth and differentiation rates of the edited cells, outcomes are highly sensitive to the efficiency, stability, precision, and progenitor affinity of the genome editor. These later parameters relevant to the design of the somatic cell genome editors are bolded (Fig. 6e).

Systematic in silico investigation of thousands of genome editing strategies are summarized in several heatmaps (Fig. 6f–i and Supplementary Figs. 11 and 12). First, we observe that imprecise editors require much higher efficiencies to achieve similar therapeutic levels in the heart (32.7% normal; Fig. 6f and Supplementary Figs. 11 and 12). Using a precise-80 editor similar to the S1plex, even though it has an efficiency of 1%, can produce similar outcomes as *SpyCas9*, which is almost six times more efficient, but only precise-25. Editing strategies used for model validation in Fig. 5e are also displayed on the heatmap, and these include *SpyCas9* delivered as a plasmid<sup>61</sup>, *SpyCas9* delivered as an encapsulated mRNA with an AAV HDR template<sup>63</sup>, *SauCas9* delivered via AAV<sup>62</sup> and RA6.3 base editor delivered as a plasmid<sup>46</sup>. Second, we observe a negative correlation between editor efficiency and genome editor half-life required to reach therapeutic levels in the heart (Fig. 6g and Supplementary Figs. 11 and 12). The parameter for efficiency in our model lumps diverse processes of cellular uptake of the editors, intracellular trafficking, nuclease activity and DNA repair at the target site because few empirical studies have been able to distinguish these processes in vivo. The efficiency of S1plex editors at this dose level was calculated to be 0.29–2% based on the bootstrap sampling based method presented in Fig. 5. Decay rate refers to the extracellular degradation or binding by serum proteins or other cells/biomolecules in the extracellular space that prevent the activity of a genome editor. Therefore, strategies that stabilize the editor in the extracellular space—even with low (~1%) efficiency editors<sup>73</sup>—could be one strategy to increase efficacy. Third, lower dosing with delivery strategies targeting progenitors could be a potent feasible strategy with highly-precise, yet low-efficiency strategies (e.g., base editors<sup>45</sup> and S1plexes<sup>74</sup>): increasing the

progenitor affinity reduces the combined genome editor dose by 50% for comparable efficacy (Fig. 6h and Supplementary Figs. 11 and 12). Progenitor targeting can, therefore, increase the potency of a genome editing-based gene therapy approach. Fourth, interventions that affect progenitor growth rate, enzymatic cross-correction, and serum GAA stability (e.g., adjuvants like Duvoglustat<sup>75</sup>) could promote efficacy. Increasing serum GAA stability by altering serum GAA half-life from 4 h (Fig. 6i and Supplementary Figs. 11 and 12, gray arrowhead) to 8 h (Fig. 6i and Supplementary Figs. 11 and 12, black arrowhead) could increase the percentage of corrected cardiac tissue by 50%. Therefore, gene therapy with GAA stabilizing adjuvants may also provide a way to boost efficacy with low-efficiency editors. Finally, we recognized that there is variation within clones in their ability to secrete GAA (Fig. 3c), and therefore modeled the therapeutic levels in the liver (Supplementary Fig. 12b), assuming that double-corrected liver cells produce between 0 and 150% excess GAA than single corrected cells. In these simulations, there is a minimal difference due to the excess secretion of GAA, due to the small number of double-corrected cells (Fig. 6d).

Because single-edited genotypes outnumber double-edited genotypes in our GETEM results, dosing time and dosing amount are not critical in the design of therapy, as we observe limited differences in potency between a single large dose or multiple smaller doses (Supplementary Fig. 13, Supplementary Information). Since genome editors can trigger an immune response<sup>76–79</sup>, which likely scale with the dose<sup>80,81</sup>, we have performed our simulations using multiple smaller doses of genome editors rather than a large single dose. The sensitivity analysis allowed us to consider tradeoffs in genome editor selection, progenitor targeting, genome editor delivery, and GAA stabilization via an adjuvant.

**Design of gene-corrected cell therapy with GETEM.** Autologous cell therapies involving gene-corrected cells avoid exposure of the body to genome editors in vivo and, therefore, may be a preferred strategy over in vivo editing, if an immune response to the editor and off-tissue or off-target effects in the body present serious safety issues. Editing both loci could provide additional therapeutic benefit, but the therapeutic benefit could depend on whether cells engraft well and retain similar function after transplantation into the body. GETEM simulations for autologous cell therapy was implemented for various cell therapeutic strategies: single-corrected progenitor cells (proliferative), double-corrected progenitor cells (proliferative), as well as single-corrected mature cells (non-proliferative) and double-corrected



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

成熟细胞 (非增殖性)。移植功能细胞的半衰期被系统地变化从 50 天到 300 天，典型的半衰期为肝细胞<sup>82</sup>。对于成熟细胞治疗，没有剂量能达到与 ERT (Fig. 7a) 相当的疗效。然而，使用增殖性祖细胞的治疗可能产生持久、有效的反应，如果移植的祖细胞半衰期超过 100 天。虽然存在

可检测到的治疗，对于移植并存活较短时间的细胞，没有持久的反应，可能需要重新给药。双重校正的细胞提供了一种类似剂量的额外益处 (Fig. 7a, 假设双重校正的细胞分泌 150% 过量的 GAA 比单校正的细胞)。为了了解细胞治疗的给药策略，我们进行了数千次模拟，变化增殖性祖细胞的

thousands of simulations varying the proliferative progenitor cell

dose, half-life, and GAA production (Fig. 7b–d). Sensitivity analysis on cell therapy GETEM indicates that tissue morphogenesis factors were much more sensitive (Fig. 7e) in comparison to cell therapy factors (dose, lifespan, and secretory phenotype of engrafted cells), or cell/tissue intrinsic factors (extracellular degradation or single allele GAA production rate). Even if the dose is raised to  $500 \times 10^6$  proliferative progenitors, if the lifespan is less than 100 days, a therapeutic outcome matching ERT is not achieved (Fig. 7b). Similarly, even if the dosed cells have native hepatocyte lifespans, as few as  $75 \times 10^6$  cells are required to observe a therapeutic effect matching ERT (Fig. 7b). If double-corrected cells engrafted well and have 75% higher production of GAA compared to single corrected cells, the dose required to reach the ERT therapeutic threshold is halved (Fig. 7c, assuming a 250-day half-life of the transplanted cells). Finally, engrafting cells with higher GAA secretion will match the standard of care, even if the lifespan of cells is lower, establishing a tradeoff between high GAA secretion and persistence of transplanted cells (Fig. 7d). Similar effects were observed on the striated skeletal muscle tissue (Supplementary Fig. 14) and the liver depot (Supplementary Fig. 15). These results combining our experimental data with GETEM support the conclusion that ex vivo cell therapy could be efficacious in addressing pathogenic polygenic mutations. Therefore, while in vivo dosing of genome editors is expected to deliver therapeutic effect largely through the involvement of single-corrected cells (Fig. 6d), a double correction strategy would be preferable for ex vivo cell therapy as it would reduce the dose of cells required. Finally, precisely edited cells in these models could also represent cells transduced in vivo with a gene augmentation therapy. Overall, we believe GETEM could be easily adapted for a wide variety of somatic tissues treated with various cell and gene therapies.

## Discussion

Many polygenic diseases have no animal models available for preclinical studies<sup>18</sup>, because many animal models arise from a single gene disruption. Moreover, it is currently infeasible to generate unique transgenic animal models for every patient's specific mutation, and it is difficult, if not impossible, to recapitulate the human genetic background in animal models. In contrast, human iPSCs provide an important biological substrate to evaluate genome editing approaches, as data from iPSC studies have been used to inform human clinical trials for gene therapies<sup>83,84</sup>. Using allele-specific genome editors, we demonstrate biallelic gene correction of two distinct mutations in a single iPSC. Many of the common Pompe disease mutations can be targeted in an allele-specific manner using *SpyCas9* strategies (Supplementary Table 6). We observe that transcriptional regulation is driven by the endogenous promoter, potentially correcting many different isoforms for GAA<sup>85,86</sup>. The tissue targets of genome editors, therefore, could expand from the traditional foci of liver and muscle to other tissues that may use alternate GAA isoforms. In contrast, in standard gene augmentation approaches, all cells must process a single isoform. Further, silencing from synthetic or viral elements has been observed for gene therapies, and in our hands with targeted knock-in strategies that overexpress a transgene via a synthetic promoter (Supplementary Fig. 8f). Transgene silencing raises concerns about the durability of viral gene therapies and proposed cell therapies where GAA is overexpressed from a safe harbor locus<sup>35</sup>. In our gene correction strategy, post-translational processing of the enzyme also appears to be intact, as the distribution of processed GAA is identical to healthy controls. In contrast, GAA overexpression in mammalian cells can cause cellular stress, leading to differential trafficking and processing of the nascent translated

peptide<sup>53</sup>. Furthermore, an mRNA- or RNP-based gene correction strategy reduces insertional oncogenesis by using a non-viral approach for the delivery of the genome editor.

Our analysis indicates that the precision of genome editing is a key intrinsic parameter controlling efficacy for in vivo somatic editing approaches. Guide RNA design for CRISPR genome editors has already been identified as critical in the design of genome editors, and the genome editing community now routinely employs design tools integrating empirical and computational approaches<sup>72,87–89</sup>. Our approach complements these guide RNA design tools to consider mechanisms in vivo involving delivery, cell targeting, and stability of editors as well as cell biological processing of the edited protein product. Mechanistic modeling of the genotypic outcomes of gene correction within a growing, differentiating tissue is necessary to understand the somatic impact of genome editing. For animal studies and clinical studies, GETEM provides insight into the distribution of edited genotypes within cells in the body and could be tailored to an individual animal or patient, creating “digital twins” of specific bodies undergoing somatic cell genome editing. Our in vitro and in silico results come together to inform cell therapeutic approaches, in which patient cells are ex vivo engineered to produce GAA from the native locus, as opposed to a safe harbor locus or viral vector<sup>90</sup>. Both single and double gene correction are efficacious for in vivo somatic cell editing strategies. While the predicted benefits from correcting both alleles are small within simulations of in vivo editing, these benefits are greater for the design of cell therapies. GETEM establishes a quantitative basis for making tradeoffs between GAA secretion, persistence, and dose of engrafted cells while considering therapeutic effects in three different organs. Thus, further development of GETEM could facilitate rapid in silico evaluation of different strategies for both gene therapies and cell therapies at an organismal level. Such an expanded toolkit may ultimately reduce the number of pre-clinical studies needed to establish efficacy before embarking on first-in-human clinical trials for genome editing therapeutics.

For gene correction in the liver, we evaluated a variety of strategies using plasmid delivery using hydrodynamic injection (an approach not easily translatable to larger animals<sup>91</sup>) viral methods (delivering both the editing machinery or just the correction DNA template), which carry significant risks of viral integration<sup>26</sup>, mRNA delivery and ribonucleoprotein delivery. Our simulations were able to design efficacious dosing and formulation for Pompe disease patients for all of these strategies. When analyzing base editor editing outcomes in vitro<sup>46</sup> (with no selection for edited outcomes, RA6.3 base editor) at the on-target site, 45% of the on-target adenines were converted to guanines, but, 55% of the adenines upstream were also converted to guanines. Therefore, the on-target precision is about 50%, or precise-50 using the Shen et al.<sup>72</sup> nomenclature, as all the other adenine to guanine edits are on-target, but imprecise. Prime editing<sup>44</sup> addresses this issue, and future development with in vivo somatic delivery of base and prime editors has strong potential even at low delivery efficiencies.

The GETEM framework is versatile and can simulate a variety of animal studies involving a wide range of growth characteristics and selection pressures for a wide array of editing strategies. For example, in their demonstration of *NmeCas9* to introduce indels to knock out *Hpd* within a mouse Hereditary Tyrosinemia Type 1 model<sup>66</sup>, the authors observe that *Hpd* indel cells are metabolically reconditioned such that they survive dietary tyrosine despite the knock-out of *Fah*. Additionally, the mice used in this experiment lost 20% of their body weight in 15 days and were between 15 and 20 weeks of age. The growth and differentiation rate constants in our GETEM were easily modified to reflect slow growth in these mice. The authors also reported that they fed

mice with NTBC from days 21 through 24 of the experiment, which can be modeled in our GETEM, and it has effects on the rate at which the fraction of the edited hepatocytes in the liver grows (Fig. 5f). Finally, the plasmid DNA dose was calculated using the length of *NmeCas9* + sgRNA plasmid (4790 bp), which is significantly shorter than the modified pX330 plasmid used in other *SpyCas9* studies (8484 bp). The breadth of studies successfully simulated by our GETEM approach indicates that it can likely be modified easily to accommodate selection, enzymatic cross-correction, and biallelic editing for many different future gene therapy applications.

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

Current limitations of our approach arise from incomplete knowledge of several phenomena, including immune response, adverse effects linked to off-target modifications, transport of editors and edited protein products, and tissue morphogenesis. First, an immune response to the editor, delivery vector, or the edited protein product is possible. Continuous constitutive expression of *SauCas9* in hepatocytes elicits an innate immune response, which accelerates the death of the edited hepatocytes<sup>92</sup>. Using strategies with a shorter-term expression of genome editor (mRNA) may prevent adverse events in the edited cells. For Pompe disease, patients already exhibit a variable response to the ERT, leading to variable efficacy. Both transient and long-term immunosuppressive strategies can be employed in conjunction with in vivo somatic editing and cell therapeutic approaches to mitigate these immune responses, and future work to expand GETEM could leverage empirical studies on the immune response both at the cellular level in vitro or in vivo. Second, we did not isolate any iPSC lines that harbored detectable off-target events in the top ten predicted sites, but others in the field have noted cells to undergo phenotypic changes induced by cutting the genome, such as cell cycle arrest<sup>93</sup> and upregulation of innate immune transcriptional programs<sup>94,95</sup>. Further, more complex events like large deletions and translocations could occur at the on- and off-target site that would be missed by our Sanger sequencing approach<sup>50</sup>. The chromatin state of iPSCs may also alter the ability of off-target sites from being edited. However, all of the methods used to characterize off-target modifications are inherently incomplete, as it is not feasible to non-destructively achieve whole-genome sequencing of every single edited cell<sup>96</sup>. Therefore, additional studies in applying the genome editor to the target tissue of interest could be necessary to identify and model additional off-target events. Further characterization of more clones and within differentiated cells<sup>95</sup> could shed light on these low-frequency events, which become crucial for cell therapeutic approaches where upwards of  $10^8$  cells have now been implanted

into patients<sup>97</sup>. Third, correction efficiencies may also only be lower in post-mitotic cells with some genome editors, and dispersion within tissues may be inhomogeneous, leading to spatial patterns of edited cells/progeny that could have variable phenotypic effects. Finally, additional empirical studies of the pharmacokinetics of editors and protein products, liver growth rates during development, glycogenolysis, and engraftment of edited cells could help us refine several assumptions made in GETEM (see Supplemental Information). Overall, these limitations provide insights into the types of measurements that are needed—detailed genomic analysis, cellular analysis, transport measurements and morphogenesis measurements—to enable a more predictive platform.

Overall, our results indicate that with appropriate engineering and design, recent advances in base editing<sup>46</sup>, delivery vectors<sup>98</sup>, nuclease decoration<sup>99</sup>, and nuclease engineering offer additional possibilities to establish efficacious gene correction therapies in a streamlined fashion with reduced dependence on animal models. We, therefore, anticipate that multiple dosing with safe and precise genome editors can be developed for a greatly expanded set of targets in diseases with compound heterozygous and complex polygenic origins.

## Methods

**Cell culture.** All human pluripotent stem cells (hPSCs) were maintained in mTeSR1 medium on Matrigel (WCell) coated tissue culture polystyrene plates (BD Falcon). Cells were passaged every 4–5 days at a ratio of 1:8 using Versene solution (Life Technologies). Patient-derived iPSC line, Pompe GM04192, was a gift from the T. Kamp and M. Suzuki (UW-Madison). Cardiomyocytes derived from hPSC and iPSC cultures were maintained in RPMI/B27 on Matrigel (WCell) coated polystyrene plates (BD Falcon). Patient-derived fibroblast lines were obtained from Coriell Institute with different GAA mutations (W746X mutation was from Coriell ID: GM04912; D645N mutation was from Coriell ID: GM20090; R660H was from Coriell ID: GM13522) and cultured in DMEM supplemented with 10% FBS and 1% Penicillin/Streptomycin. All cells were maintained at 37 °C in 5% CO<sub>2</sub>, and tested monthly for possible mycoplasma contamination. Samples obtained from the Coriell Institute have been approved by the Working Group of the Human Genetic Cell Repository.

**Cardiomyocyte differentiation.** hPSCs and iPSCs were differentiated into cardiomyocytes using a small molecule-directed differentiation protocol in a 12-well plate format<sup>54</sup>. Briefly, all adherent hPSCs and iPSCs were dissociated in TrypLE solution (Life Technologies), counted with a hemocytometer, and centrifuged at  $200 \times g$  for 5 min. Cells were plated at a density between  $0.5\text{--}1 \times 10^6$  cells/well depending on cell line. Once tissue culture plate wells reached 100% confluency (day 0), medium in each well was replaced with a solution containing ml RPMI/B27-Insulin (Life Technologies), 12  $\mu\text{M}$  CHIR99021 (BioGems 25917), and 1  $\mu\text{g/ml}$  Insulin solution (Sigma-Aldrich I9278). Exactly 24 h later (day 1) medium in each well was removed and replaced with RPMI/B27-insulin. Exactly 48 h after (day 3) half of the spent medium was collected. To this, an equal volume of fresh RPMI/B27-Insulin was mixed. This combined media was then supplemented with 7.5  $\mu\text{M}$  IWP2 (BioGems 75844). Two days later (day 5) medium in each well was replaced with RPMI/B27-Insulin. Two days (day 7) later and every three days following, spent medium was replaced with RPMI/B27. Spontaneous contraction was generally observed between days 12–16 of differentiation.

**Creation of S1m-sgRNAs to correct both mutant alleles.** S1m-sgRNAs were designed by combining the sgRNA scaffold with the S1m aptamer sequence<sup>74</sup>. S1m gBlocks were annealed with Phusion polymerase (New England Biolabs) under the following thermocycler conditions: 98 °C for 30 s followed by 30 cycles at 98 °C for 10 s, and 72 °C for 15 s with a final extension at 72 °C for 10 min. S1m cDNA was annealed with Phusion polymerase (New England Biolabs) under the following thermocycler conditions: 98 °C for 30 s followed by 30 cycles at 98 °C for 10 s, 60 °C for 10 s, and 72 °C for 15 s with a final extension at 72 °C for 10 min. In vitro transcription was performed with the MEGAShortscript T7 Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. For guide RNAs for fibroblast transfection, in vitro transcription was performed using HiScribe T7 RNA synthesis Kit (New England Biolabs) according to the manufacturer's instructions.

**Genome editor delivery.** All hPSC transfections were performed using the 4D-Nucleofector System (Lonza) as per the manufacturer's instructions. Briefly, hPSCs were harvested using TrypLE (Life Technologies) and counted.  $2 \times 10^5$  cells per transfection were then centrifuged at  $100 \times g$  for 3 min. Excess media was aspirated

and cells were resuspended using 20  $\mu$ l of RNP solution per condition and then electroporated using protocol CA-137. 50 pmol Cas9, 60 pmol sgRNA, 50 pmol streptavidin, and 60 pmol ssODN were used to form particles per ssODN-S1mplex as described previously<sup>74</sup>. Cells were then harvested using TrypLE (Life Technologies) and counted.  $2 \times 10^5$  cells per transfection were then centrifuged at  $100 \times g$  for 3 min. Excess media was aspirated and cells were resuspended using 20  $\mu$ l of RNP solution per condition. After nucleofection, samples were incubated in nucleocuvettes at room temperature for 15 min prior to plating into  $2 \times 10^4$  cells per well on ArrayEdit in mTeSR media + 10  $\mu$ M ROCK inhibitor. Media was changed 24 h post transfection and replaced with mTeSR1 medium. Fibroblast transfections were performed in 24 well plates using 50,000 cells/well using 2  $\mu$ l Lipofectamine 2000/well (0.5  $\mu$ g Cas9/well and sgRNA, streptavidin and ssODN at a 1:1:1:1 molar ratio).

**Synthesis of ArrayEdit platform.** Microcontact printing was performed to modify the wells of a standard tissue culture plate<sup>100</sup>. The surface modification involved printing of an alkanethiol initiator to nucleate the polymerization of hydrophilic poly(ethylene glycol) (PEG) chains. Briefly, double sided-adhesive was attached to the bottom of a standard tissue culture plate, after which a laser cutter was used to cut out the well bottoms. Patterns were transferred to gold-coated glass via a polydimethylsiloxane stamp after which the glass was submerged in a PEG solution overnight to build PEG chains surrounding  $\mu$ Features<sup>101</sup>. Standard tissue culture plates with well bottoms cut out were then fastened to processed sheets using a custom-made alignment device.

**High-content analysis.** Automated microscopy was performed using a Nikon Eclipse Ti epifluorescent scope. A  $15 \times 15$  grid with one  $\mu$ Feature per image was established and maintained so that each feature imaged was consistent each day. Nikon Perfect Focus was used to ensure that all colonies were in the same Z-plane and LysoSensor intensity was measured accurately. Images were processed using CellProfiler<sup>102</sup> to count the number of nuclei and quantify LysoSensor intensity.

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

**Off-target analysis.** Candidate off-target sites were identified and ranked for both of the sgRNAs used for gene correction using a previously established algorithm<sup>103</sup>. This algorithm was validated via unbiased genome-wide off-target analysis by Tsai et al.<sup>104</sup>, and the top scoring sites are listed in Supplementary Table 4. PCR on genomic DNA from established clonal lines was performed with primers listed in Supplementary Table 3 flanking each of the top ten scoring candidate off-target sites. These PCR products were analyzed using Sanger sequencing by the University of Wisconsin-Madison Biotechnology Center and Genewiz.

**RT- and qPCR.** RNA was isolated from cells using QuickExtract RNA Extraction Solution (Epicentre) following the manufacturer's protocol. 100 ng of extracted RNA was reverse transcribed using Superscript IV Reverse Transcriptase (Invitrogen). Endpoint PCR amplification of cDNA product was performed using 1  $\mu$ l of cDNA product and AccuPrime HiFi Taq (Life Technologies) following the manufacturer's instructions. Efficacy of the endpoint PCR was evaluated via gel electrophoresis of PCR product in a 1% agarose gel.

qPCR reaction was performed in triplicates for each cell line and sequence (GAPDH, 1441delT, 2237G>A, and GAA), by mixing 10  $\mu$ l iTaq Universal SYBR Green Supermix (Bio-Rad), 0.5  $\mu$ l sequence specific forward primer, 0.5  $\mu$ l sequence specific reverse primer, 1  $\mu$ l cDNA product, and 8  $\mu$ l water. qPCR analysis was performed in a CFX96 Real Time PCR System under the following thermocycling conditions: 95 °C for 30 s followed by 35 cycles of 95 °C for 5 s, and 60 °C for 30 s.

**Next generation sequencing analysis.** Cas-Analyzer<sup>105</sup> was used to perform sequence analysis. For each sample, sequences with frequency of less than 1000 were filtered from the data. Sequences in which the reads matched with primer and reverse complement subsequences classified as "target sequences". Target sequences were aligned with corresponding wildtype sequence using global pairwise sequence alignment and analyzed manually for mutants and PAM wobbles.

**Western blotting.** Protein levels of GAA and  $\beta$ -Actin were determined in each cell line. Following cell lysis in ice-cold RIPA buffer supplemented with protease and phosphatase inhibitors and EDTA (5 mM), protein concentration was determined

(DC Protein Assay, BioRad). 40  $\mu$ g of protein from each cell line was loaded into a 4–12% Bis-Tris precast gel (Criterion XT, BioRad) and gel electrophoresis was performed. Proteins were then transferred to a nitrocellulose membrane and blocked in filtered 5% nonfat dry milk in TBS-T (Tris-buffered saline, 0.15% Tween20) for 1 h at room temperature. The membrane was then incubated overnight at 4 °C with GAA (Abcam ab137068, 1:1000) and  $\beta$ -Actin (Millipore, MAB1501, 1:40,000) primary antibodies. Following the incubation period, the membrane was washed in TBS-T and incubated with appropriate horseradish peroxidase secondary antibodies (Goat Anti-Rabbit IgG, Abcam ab205718, 1:2000; Anti-Mouse IgG, Cell Signaling Technologies 7076 1:20,000) for 1 h. The membrane was washed again in TBS-T, and then developed (SuperSignal West Pico Plus Chemiluminescent Substrate, Thermo Scientific) for 5 min using a ChemiDoc-It2 Imaging System (UVP) and imaged.

**GAA Activity assay.** Acid glucosidase activity was measured by hydrolysis of 4-methylumbelliferyl-D-glucoside (4-MUG, Sigma M-9766) at pH 4 to release the fluorophore 4-methylumbelliferone (4-MU)<sup>47</sup>. Briefly, 4-MUG was incubated with 10  $\mu$ l protein lysate in 0.2 M sodium acetate buffer (pH 4.3) for 90 min at 37 °C. Fluorescence from 4-MU was then measured using a Gloomax plate reader (Promega) and activity was calculated using a standard curve.

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

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

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

**Media exchange.** Cardiomyocytes were cultured in RPMI/B27 + insulin and media was exchanged every 2 days. As a normal media exchange, diseased and corrected cells were introduced to RPMI/B27 +insulin/–glucose. 24 h post change, cells were stained with LysoSensor as described above to determine a baseline fluorescent intensity. After staining, media was replaced with media from either corrected or healthy lines and cultured for an additional 24 h. After incubation, cells were again stained with LysoSensor and imaged using confocal microscopy.

**In silico modeling: gene therapy efficacy model (GETEM).** Preliminary model design was performed using COPASI 4.21<sup>106</sup>, and the final model construction was performed using MATLAB R2020a using the SimBiology<sup>®</sup> package. Ten thousand replicates<sup>107</sup> from published data were generated via the MATLAB (R2020a) bootstrap function using default options in the MATLAB software. The SimBiology sbproj files used for the simulations (Supplementary Table 9) has been provided in the Supplementary Files hosted at Zenodo—<https://tinyurl.com/GETEMZenodo> (contains Supplementary Videos, Modeling Code, Sequencing Data, and Western blot image).

**Statistics and reproducibility.** For Fig. 2f, three technical replicates performed. For Fig. 2h, PCR was performed once. For Fig. 3b, western blotting was performed once with three technical replicates. For Fig. 4c, d, three technical replicates performed. For Supplementary Fig. 1c, d, three technical replicates performed. For Supplementary Fig. 8, three technical replicates were performed for all conditions.

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

## Data availability

These authors declare that all essential data supporting the conclusion of the study as well as detailed assay protocols, analytical algorithms, and customized computational codes are within the paper and Supplementary materials. Data are present in Zenodo repository (<https://doi.org/10.5281/zenodo.3910777>, <https://tinyurl.com/GETEMZenodo>). Code is available at <https://github.com/ada586/PompeGetem.git>. Raw reads and traces from sequencing are available at NCBI Bioproject PRJNA675893 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA675893>). Any additional relevant information can be obtained from the authors upon reasonable request. Source data are provided with this paper.

Received: 9 October 2020; Accepted: 13 November 2020;

Published online: 08 December 2020

## References

- Cox, D. B. T., Platt, R. J. & Zhang, F. Therapeutic genome editing: prospects and challenges. *Nat. Med.* **21**, 121–131 (2015).
- Lim, J.-A., Li, L. & Raben, N. Pompe disease: from pathophysiology to therapy and back again. *Front. Aging Neurosci.* **6**, 177 (2014).
- Yin, H., Kauffman, K. J. & Anderson, D. G. Delivery technologies for genome editing. *Nat. Rev. Drug Discov.* **16**, 387–399 (2017).
- Rui, Y., Wilson, D. R. & Green, J. J. Non-viral delivery to enable genome editing. *Trends Biotechnol.* **37**, 281–293 (2019).
- 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).
- 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).
- Ran, F. A. et al. Double Nicking by RNA-Guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–1389 (2013).
- Kleinstiver, B. P. et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).
- Kleinstiver, B. P. et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).
- Slaymaker, I. M. et al. Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2015).
- 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).
- Yu, C. et al. Small molecules enhance CRISPR genome editing in pluripotent stem cells. *Cell Stem Cell* **16**, 142–147 (2015).
- Yang, D. et al. Enrichment of G2/M cell cycle phase in human pluripotent stem cells enhances HDR-mediated gene repair with customizable endonucleases. *Sci. Rep.* **6**, 21264 (2016).
- Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* **3**, e04766 (2014).
- Gutschner, T., Haemmerle, M., Genovese, G., Draetta, G. F. & Chin, L. Post-translational regulation of Cas9 during G1 enhances homology-directed repair. *Cell Rep.* **14**, 1555–1566 (2016).
- Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34**, 339–344 (2016).
- Liang, X., Potter, J., Kumar, S., Ravinder, N. & Chesnut, J. D. Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. *J. Biotechnol.* **241**, 136–146 (2017).
- Prakash, V., Moore, M. & Yáñez-Muñoz, R. J. Current progress in therapeutic gene editing for monogenic diseases. *Mol. Ther.* **24**, 465–474 (2016).
- Firth, A. L. et al. Functional gene correction for cystic fibrosis in lung epithelial cells generated from patient iPSCs. *Cell Rep.* **12**, 1385–1390 (2015).
- Ousterout, D. G. et al. Correction of dystrophin expression in cells from duchenne muscular dystrophy patients through genomic excision of exon 51 by zinc finger nucleases. *Mol. Ther.* **23**, 523–532 (2015).
- Ribeil, J.-A. et al. Gene therapy in a patient with sickle cell disease. *N. Engl. J. Med.* **376**, 848–855 (2017).
- Li, H. et al. In vivo genome editing restores haemostasis in a mouse model of haemophilia. *Nature* **475**, 217–221 (2011).
- Ravin, S. S. D. et al. CRISPR-Cas9 gene repair of hematopoietic stem cells from patients with X-linked chronic granulomatous disease. *Sci. Transl. Med.* **9**, eaah3480–eaah3480 (2017).
- Maeder, M. L. et al. Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. *Nat. Med.* **25**, 229–233 (2019).
- Stadtmauer, E. A. et al. CRISPR-engineered T cells in patients with refractory cancer. *Science* **367**, eaba7365 (2020).
- Nelson, C. E. et al. Long-term evaluation of AAV-CRISPR genome editing for Duchenne muscular dystrophy. *Nat. Med.* **25**, 427–432 (2019).
- Steyer, B. et al. Scarless genome editing of human pluripotent stem cells via transient puromycin selection. *Stem Cell Rep.* **10**, 642–654 (2018).
- Kishnani, P. S. & Howell, R. R. Pompe disease in infants and children. *J. Pediatr.* **144**, S35–S43 (2004).
- Bodamer, O. A., Scott, C. R. & Giugliani, R., Group, on behalf of the P. D. N. S. W. Newborn screening for Pompe disease. *Pediatrics* **140**, S4–S13 (2017).
- Han, S. et al. Low-dose liver-targeted gene therapy for Pompe disease enhances therapeutic efficacy of ert via immune tolerance induction. *Mol. Ther. Methods Clin. Dev.* **4**, 126–136 (2017).
- Martin-Touaux, E. et al. Muscle as a putative producer of acid  $\alpha$ -glucosidase for glycogenosis type II gene therapy. *Hum. Mol. Genet.* **11**, 1637–1645 (2002).
- Ding, E. Y. et al. Long-term efficacy after [E1-, polymerase-] adenovirus-mediated transfer of human acid- $\alpha$ -glucosidase gene into glycogen storage disease type II knockout mice. *Hum. Gene Ther.* **12**, 955–965 (2001).
- Xu, F. et al. Glycogen storage in multiple muscles of old GSD-II mice can be rapidly cleared after a single intravenous injection with a modified adenoviral vector expressing hGAA. *J. Gene Med.* **7**, 171–178 (2005).
- van der Wal, E., Bergsma, A. J., Pijnenburg, J. M., van der Ploeg, A. T. & Pijnappel, W. W. M. P. Antisense oligonucleotides promote exon inclusion and correct the common c.-32-13T>G GAA splicing variant in pompe disease. *Mol. Ther. Nucleic Acids* **7**, 90–100 (2017).
- van der Wal, E. et al. Large-scale expansion of human iPSC-derived skeletal muscle cells for disease modeling and cell-based therapeutic strategies. *Stem Cell Rep.* **10**, 1975–1990 (2018).
- Hendriks, W. T., Warren, C. R. & Cowan, C. A. Genome editing in human pluripotent stem cells: approaches, pitfalls, and solutions. *Cell Stem Cell* **18**, 53–65 (2016).
- Warren, C. R. et al. Induced pluripotent stem cell differentiation enables functional validation of GWAS variants in metabolic disease. *Cell Stem Cell* **20**, 547–557.e7 (2017).
- Saha, K. & Jaenisch, R. Technical challenges in using human induced pluripotent stem cells to model disease. *Cell Stem Cell* **5**, 584–595 (2009).
- Chen, I. Y., Matsa, E. & Wu, J. C. Induced pluripotent stem cells: at the heart of cardiovascular precision medicine. *Nat. Rev. Cardiol.* **13**, 333–349 (2016).
- Torii, R., Velliou, R.-I., Hodgson, D. & Mudera, V. Modelling multi-scale cell-tissue interaction of tissue-engineered muscle constructs. *J. Tissue Eng.* **9**, 2041731418787141 (2018).
- Matsumoto, S., Tanaka, K., Strong, D. M. & Reems, J. A. Efficacy of human islet isolation from the tail section of the pancreas for the possibility of living donor islet transplantation. *Transplantation* **78**, 839–843 (2004).
- Matsumoto, S. et al. Insulin independence after living-donor distal pancreatectomy and islet allotransplantation. *Lancet* **365**, 1642–1644 (2005).
- Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**, 290–296 (2020).
- Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
- Zafra, M. P. et al. Optimized base editors enable efficient editing in cells, organoids and mice. *Nat. Biotechnol.* **36**, 888–893 (2018).
- Song, C.-Q. et al. Adenine base editing in an adult mouse model of tyrosinaemia. *Nat. Biomed. Eng.* **4**, 125–130 (2019).
- Raval, K. K. et al. Pompe disease results in a golgi-based glycosylation deficit in human induced pluripotent stem cell-derived cardiomyocytes. *J. Biol. Chem.* **290**, 3121–3136 (2015).
- Pattanayak, V. et al. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31**, 839–843 (2013).
- Carlson-Stevermer, J. et al. High-content analysis of CRISPR-Cas9 gene-edited human embryonic stem cells. *Stem Cell Rep.* **6**, 109–120 (2016).
- Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).
- Popp, M. W. & Maquat, L. E. Leveraging rules of nonsense-mediated mRNA decay for genome engineering and personalized medicine. *Cell* **165**, 1319–1322 (2016).
- El-Brolosy, M. A. & Stainier, D. Y. R. Genetic compensation: A phenomenon in search of mechanisms. *PLoS Genet.* **13**, e1006780 (2017).
- Migani, D., Smales, C. M. & Bracewell, D. G. Effects of lysosomal biotherapeutic recombinant protein expression on cell stress and protease and general host cell protein release in Chinese hamster ovary cells. *Biotechnol. Prog.* **33**, 666–676 (2017).
- Lian, X. et al. Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. *Proc. Natl Acad. Sci.* **109**, E1848–E1857 (2012).
- Yang, H. W. et al. Recombinant human acid  $\alpha$ -glucosidase Corrects Acid- $\alpha$ -glucosidase-deficient human fibroblasts, quail fibroblasts, and quail myoblasts. *Pediatr. Res.* **43**, 374–380 (1998).
- Ma, L., Ouyang, Q., Werthmann, G. C., Thompson, H. M. & Morrow, E. M. Live-cell microscopy and fluorescence-based measurement of luminal pH in intracellular organelles. *Front. Cell Dev. Biol.* **5**, 71 (2017).
- Bond, J. E., Kishnani, P. S. & Koeberl, D. D. Immunomodulatory, liver depot gene therapy for Pompe disease. *Cell. Immunol.* **342**, 103737 (2019).



58. Kishnani, P. S. et al. Recombinant human acid  $\alpha$ -glucosidase major clinical benefits in infantile-onset Pompe disease. *Neurology* **68**, 99–109 (2007).
59. Raben, N. et al. Enzyme replacement therapy in the mouse model of Pompe disease. *Mol. Genet. Metab.* **80**, 159–169 (2003).
60. Wang, B., Zhao, L., Fish, M., Logan, C. Y. & Nusse, R. Self-renewing diploid Axin2+ cells fuel homeostatic renewal of the liver. *Nature* **524**, 180–185 (2015).
61. Yin, H. et al. Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nat. Biotechnol.* **32**, 551–553 (2014).
62. Yang, Y. et al. A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice. *Nat. Biotechnol.* **34**, 334–338 (2016).
63. Yin, H. et al. Therapeutic genome editing by combined viral and non-viral delivery of CRISPR system components in vivo. *Nat. Biotechnol.* **34**, 328–333 (2016).
64. Wei, T., Cheng, Q., Min, Y.-L., Olson, E. N. & Siegwart, D. J. Systemic nanoparticle delivery of CRISPR-Cas9 ribonucleoproteins for effective tissue specific genome editing. *Nat. Commun.* **11**, 3232 (2020).
65. Shin, J. H., Jung, S., Ramakrishna, S., Kim, H. H. & Lee, J. In vivo gene correction with targeted sequence substitution through microhomology-mediated end joining. *Biochem. Biophys. Res. Commun.* **502**, 116–122 (2018).
66. Ibrahim, R. et al. All-in-one adeno-associated virus delivery and genome editing by Neisseria meningitidis Cas9 in vivo. *Genome Biol.* **19**, 137 (2018).
67. Sago, C. D. et al. High-throughput in vivo screen of functional mRNA delivery identifies nanoparticles for endothelial cell gene editing. *Proc. Natl Acad. Sci. USA* **115**, E9944–E9952 (2018).
68. Villiger, L. et al. Treatment of a metabolic liver disease by in vivo genome base editing in adult mice. *Nat. Med.* **24**, 1519–1525 (2018).
69. Albayrak, C. et al. Digital quantification of proteins and mRNA in single mammalian cells. *Mol. Cell* **61**, 914–924 (2016).
70. Azuma, H. et al. Robust expansion of human hepatocytes in Fah<sup>-/-</sup>/Rag2<sup>-/-</sup>/Il2rg<sup>-/-</sup> mice. *Nat. Biotechnol.* **25**, 903–910 (2007).
71. Kulesa, A., Krzywinski, M., Blainey, P. & Altman, N. Sampling distributions and the bootstrap. *Nat. Methods* **12**, 477–478 (2015).
72. Shen, M. W. et al. Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* **563**, 646 (2018).
73. Harrington, L. B. et al. A thermostable Cas9 with increased lifetime in human plasma. *Nat. Commun.* **8**, 1424 (2017).
74. 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).
75. 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).
76. Charlesworth, C. T. et al. Identification of preexisting adaptive immunity to Cas9 proteins in humans. *Nat. Med.* **25**, 249–254 (2019).
77. 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).
78. Chew, W. L. et al. A multifunctional AAV-CRISPR-Cas9 and its host response. *Nat. Methods* **13**, 868–874 (2016).
79. Wagner, D. L. et al. High prevalence of Streptococcus pyogenes Cas9-reactive T cells within the adult human population. *Nat. Med.* **25**, 242 (2019).
80. 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. Vaccin. Immunol.* **17**, 2029–2032 (2010).
81. Katare, Y. K., Muthukumar, 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).
82. Seeger, C. & Mason, W. S. Hepatitis B virus biology. *Microbiol. Mol. Biol. Rev.* **64**, 51–68 (2000).
83. Duong, T. T. et al. Use of induced pluripotent stem cell models to probe the pathogenesis of Choroideremia and to develop a potential treatment. *Stem Cell Res.* **27**, 140–150 (2018).
84. Torriano, S. et al. The effect of PTC124 on choroideremia fibroblasts and iPSC-derived RPE raises considerations for therapy. *Sci. Rep.* **8**, 8234 (2018).
85. Herzog, A. et al. A cross-sectional single-centre study on the spectrum of Pompe disease, German patients: molecular analysis of the GAA gene, manifestation and genotype-phenotype correlations. *Orphanet J. Rare Dis.* **7**, 35 (2012).
86. Fukuhara, Y. et al. A molecular analysis of the GAA gene and clinical spectrum in 38 patients with Pompe disease in Japan. *Mol. Genet. Metab. Rep.* **14**, 3–9 (2018).
87. Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
88. Cloney, R. The oracle of inDelphi predicts Cas9 repair outcomes. *Nat. Rev. Genet.* **20**, 4 (2019).
89. Allen, F. et al. Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat. Biotechnol.* **37**, 64–72 (2018).
90. Ronzitti, G., Collaud, F., Laforet, P. & Mingozi, F. Progress and challenges of gene therapy for Pompe disease. *Ann. Transl. Med.* **7**, 287 (2019).
91. Sendra, L., Herrero, M. J. & Aliño, S. F. Translational advances of hydrofection by hydrodynamic injection. *Genes* **9**, 136 (2018).
92. Li, A. et al. AAV-CRISPR Gene editing is negated by pre-existing immunity to Cas9. *Mol. Ther.* **28**, 1432–1441 (2020).
93. Haapaniemä, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* <https://doi.org/10.1038/s41591-018-0049-z> (2018).
94. Cromer, M. K. et al. Global Transcriptional Response to CRISPR/Cas9-AAV6-Based Genome Editing in CD34+ Hematopoietic Stem and Progenitor Cells. *Mol. Ther.* **26**, 2431–2442 (2018).
95. Sinha, D. et al. Human iPSC modeling reveals mutation-specific responses to gene therapy in a genotypically diverse dominant maculopathy. *Am. J. Hum. Genet.* **107**, 278–292 (2020).
96. Temple, A. Somatic Cell Gene Editing. *Somatic Cell Gene Editing* <https://scge.mcw.edu/>.
97. Liu, E. et al. Use of CAR-transduced natural killer cells in CD19-positive lymphoid tumors. *N. Engl. J. Med.* **382**, 545–553 (2020).
98. Zhu, H. et al. Spatial control of in vivo CRISPR-Cas9 genome editing via nanomagnets. *Nat. Biomed. Eng.* **3**, 126 (2019).
99. Lee, B. et al. Nanoparticle delivery of CRISPR into the brain rescues a mouse model of fragile X syndrome from exaggerated repetitive behaviours. *Nat. Biomed. Eng.* **2**, 497 (2018).
100. Harkness, T. et al. High-content imaging with micropatterned multiwell plates reveals influence of cell geometry and cytoskeleton on chromatin dynamics. *Biotechnol. J.* **10**, 1555–1567 (2015).
101. Sha, J., Lippmann, E. S., McNulty, J., Ma, Y. & Ashton, R. S. Sequential nucleophilic substitutions permit orthogonal click functionalization of multicomponent PEG brushes. *Biomacromolecules* **14**, 3294–3303 (2013).
102. Carpenter, A. E. et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).
103. Ran, F. et al. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
104. Tsai, S. Q. et al. GUIDE-Seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).
105. Park, J., Lim, K., Kim, J.-S. & Bae, S. Cas-analyzer: an online tool for assessing genome editing results using NGS data. *Bioinformatics* **33**, 286–288 (2017).
106. Hoops, S. et al. COPASI—a COMplex PATHway Simulator. *Bioinforma. Oxf. Engl.* **22**, 3067–3074 (2006).
107. 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).

## Acknowledgements

We thank members of the Saha lab for helpful discussion and comments on the manuscript, Masatoshi Suzuki and Tim Kamp for sharing Pompe iPSC lines, the University of Wisconsin Biotechnology Center for the use of the equipment and technical support, ArtforScience for assistance with figure art, and Aldevron for technical support with Cas9 proteins. We acknowledge generous financial support from the National Science Foundation (CBET-1350178, CBET-1645123), National Institute for Health (1R35GM119644-01), Environmental Protection Agency (EPA-G2013-STAR-L1), University of Wisconsin Carbone Cancer Center Support Grant P30 CA014520, Wisconsin Alumni Research Foundation, and the Wisconsin Institute for Discovery.

## Author contributions

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

## Competing interests

J.C.-S., A.A., L.K., and K.S. have filed a patent application on the S1mplex technology. Patent Applicant: Institution - WISCONSIN ALUMNI RESEARCH FOUNDATION (MADISON, WI, US) Name of Inventors: Carlson-stevermer, Jared Matthew (Madison, WI, US) Saha, Krishanu (Middleton, WI, US) Abdeen, Amr Ashraf (Madison, WI, US) Kohlenberg, Lucille Katherine (Madison, WI, US) Application number: United States Patent Application 20180362971 Status of Application: Filed Aspect of Manuscript: Generation of single edited clones, and double-edited clones c1, c28, c29, and c30. The authors declare no other competing interests.

**Additional information**

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41467-020-20065-8>.

**Correspondence** and requests for materials should be addressed to K.S.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020