Human genome-edited hematopoietic stem cells phenotypically correct Mucopolysaccharidosis type I

Natalia Gomez-Ospina[1*], Sam Glynne Scharenberg[2], Nathalie Mostrel[2], Rasmus O. Bak[3], Sruthi Mantri[2], Rolen M. Quadros[4], Channabasavaiah B. Gurumurthy[4,5], Ciaran Lee[6], Gang Bao[6], Laure Aurelian[7,8], Matthew H. Porteus[2*].

[1]Division of Medical Genetics, Stanford University School of Medicine, Stanford, CA, USA.

[2]Division of Stem Cell Transplantation and Regenerative Medicine, Stanford University School of Medicine, Palo Alto, CA, USA.

[3]Department of Biomedicine, Aarhus University, DK-8000 Aarhus C., Denmark.

[4]Mouse Genome Engineering Core Facility, Vice Chancellor for Research Office, University of Nebraska Medical Center, Omaha, Nebraska

[4,5]Developmental Neuroscience, Munroe Meyer Institute for Genetics and Rehabilitation, of University of Nebraska Medical Center, Omaha, Nebraska

[6]Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, Georgia, USA.

[7]Stanford University School of Medicine, Stanford, CA 94305

[8]University of Maryland School of Medicine, Baltimore, MD 21201

* Corresponding author: mporteus@stanford.edu and gomezosp@stanford.edu

1   **Summary**

2   Lysosomal enzyme deficiencies comprise a large group of genetic disorders that generally lack

3   effective treatments. A potential treatment approach is to engineer the patient's own hematopoietic

4   system to express high levels of the deficient enzyme, thereby correcting the biochemical defect

5   and halting disease progression. Here, we present an efficient ex *vivo* genome editing approach

6   using CRISPR/Cas9 that targets the lysosomal enzyme iduronidase to the *CCR5* safe harbor locus

7   in human CD34+ hematopoietic stem and progenitor cells. The modified cells secrete supra-

8   endogenous enzyme levels, maintain long-term repopulation and multi-lineage differentiation

9   potential, and can correct biochemical and phenotypic abnormalities in an immunocompromised

10   mouse model of Mucopolysaccharidosis type I. Our studies provide support for the development

11   of human, genome-edited CD34+ hematopoietic stem and progenitor cells for the treatment of a

12   multi-systemic lysosomal storage disorder. Our safe harbor approach constitutes a flexible

13   platform for the expression of lysosomal enzymes, exemplifying a potential new paradigm for the

14   treatment of these diseases.

15   **Introduction**

16   Lysosomal storage diseases (LSDs) comprise a large group of genetic disorders caused by

17   deficiencies in lysosomal proteins; many lack effective treatments. Mucopolysaccharidosis type I

18   (MPSI) is a common LSD caused by insufficient iduronidase (IDUA) activity that results in

19   glycosaminoglycan (GAG) accumulation and progressive multi-systemic deterioration that

20   severely affects the neurological and musculoskeletal systems[1]. Current interventions for MPSI

21   include enzyme replacement therapy (ERT) and allogeneic hematopoietic stem cell transplantation

22   (allo-HSCT); both have limited efficacy. ERT does not cross the blood-brain barrier, requires

23   costly life-long infusions, and inhibitory antibodies can further decrease enzyme bioavailability[2].

24   Allo-HSCT results in better outcomes than ERT by providing a persistent source of enzyme and

25   tissue macrophages that can migrate into affected organs, including the brain, to deliver local

26   enzyme[3-5]. However, allo-HSCT has significant limitations, including the uncertain availability of

27   suitable donors, delay in treatment (allowing for irreversible progression), and transplant-

28   associated morbidity and mortality such as graft-versus-host disease and drug-induced

29   immunosuppression.

30   Human and animal studies in MPSI have shown that the therapeutic efficacy of HSCT can be

31   enhanced by increasing the levels of circulating IDUA. In humans, patients transplanted with non-

32   carrier donors had better clinical responses than patients transplanted with HSPCs from MPSI

33   heterozygotes with decreased enzyme expression[6]. In mice, transplantation of virally transduced

34   murine hematopoietic stem and progenitor cells (HSPCs) expressing supra-normal enzyme

35   levels[7,8] dramatically corrected the phenotype. Based on this, autologous transplantation of

36   lentivirus-transduced HSPCs overexpressing lysosomal enzymes is being explored in human trials

37   for LSDs[9] (ClinicalTrials.gov, NCT03488394). This autologous approach eliminates the need to

38   find immunologically matched donors and minimizes many of the potential complications from

39   allogeneic transplants. However, concerns remain about the potential for tumorigenicity associated

40   with random insertion of the viral genomes[10,11], carry-over of infectious particles[12], the immune

41   response to some of the vectors, and variable transgene expression[13].

42   Recently developed genome editing tools combine precise gene addition with genetic alterations

43   that can add therapeutic benefit[14]. Among these, Clustered Regularly Interspaced Short

44   Palindromic Repeats-associated protein-9 nuclease (CRISPR/Cas9) is the simplest to engineer and

45   has been used to successfully modify HSPCs in culture[15]. The system was repurposed for editing

46   eukaryotic cells by delivering the Cas9 nuclease, and a short guide RNA (sgRNA). When targeted

47   to the sequence determined by the sgRNA, Cas9 creates a double-stranded DNA break, thereby

48   stimulating homologous recombination with a designed donor DNA template that contains the

49   desired genetic modification embedded between homology arms centered at the break site. This

50   process, termed "homologous recombination-mediated genome editing" (HR-GE) is most often

51   used for in-situ gene correction and has been hailed as a tool to treat monogenic diseases. Although

52   its therapeutic potential in LSDs is unknown, to maximize therapeutic correction by autologous

53   transplantation of genetically modified HSPCs in LSDs, functional enzymes must be expressed at

54   higher-than-endogenous levels. This can be achieved by inserting an expression cassette

55   (exogenous promoter-gene of interest) into non-essential genomic region (or "safe harbor"). A safe

56   harbor provides a platform that is independent of specific patient mutations, is easily adaptable to

57   various lysosomal enzymes and, compared to lentiviral transduction, ensures more predictable and

58   consistent transgene expression because the insertion sites are restricted (up to 2 in autosomes).

3

59   Herein, we describe the development of such an approach for MPSI. We use *CCR5* as the target

60   safe harbor to insert an expression cassette to overexpress IDUA in human CD34+ HPSCs and

61   their progeny. *CCR5* is considered a non-essential gene because bi-allelic inactivation of *CCR5*

62   (CCR5Δ32) has no general detrimental impact on human health and the only known phenotypes

63   of CCR5 loss are resistance to HIV-1 infection and increased susceptibility to West Nile virus[16].

64   We report that human HSPCs modified using genome editing to express IDUA from the *CCR5*

65   locus engraft and correct the biochemical, visceral, musculoskeletal, and neurologic manifestations

66   of the disease in a new immunocompromised model of MSPI.

67

68   **Results**

69   **Efficient targeting of IDUA into the *CCR5* locus in human HSPCs**

70   To generate human CD34[+] HPSCs overexpressing IDUA, we used sgRNA/Cas9 ribonucleoprotein

71   (RNP) and adeno-associated viral vector serotype six (AAV6) delivery of the homologous

72   templates[17]. RNP complexes consisting of 2′-*O*-methyl 3′phosphorothioate-modified *CCR5*

73   sgRNA[18] and Cas9 protein were electroporated into cord blood-derived (CB) and adult peripheral

74   blood-derived HSPCs (PB). The efficiency of double-strand DNA break (DSB) generation by our

75   *CCR5* RNP complex was estimated by measuring the frequency of insertions/deletions (Indel) at

76   the predicted cut site. The mean Indel frequencies were 83% ± in CB-HSPCs and 76% ± 8 in PB-

77   HSPCs, consistent with a highly active sgRNA**.** The predominant Indel was a single A/T insertion

78   that abrogated CCR5 protein expression (**Extended Data Fig. 1**).

79   To achieve precise genetic modification, the templates for homologous recombination were made

80   by inserting IDUA expression cassettes driven by the spleen focus-forming virus (SFFV) or the

81   phosphoglycerate kinase (PGK) promoter, followed by a yellow fluorescent protein (YFP)

82   downstream of the self-cleaving P2A peptide into the AAV vector genome. A third expression

83   cassette containing IDUA driven by PGK but without a selection marker was also made (**Fig. 1a).**

84   Following electroporation, CB and PB cells transduced with the SFFV-IDUA-YFP and PGK-

85   IDUA-YFP viruses were examined for YFP fluorescence to quantify the efficiency of

86   modification. As shown in **Figure 1b**, RNP electroporation followed by AAV6 transduction lead

4

87   to a marked increase in the median fluorescence intensity of the cells. In CB-derived HSPCs the

88   mean fraction of YFP-positive cells, was 34% ± 7 and 32% ± 8 with SFFV and PGK-driven

89   expression cassettes respectively. In PB-HSPCs, the frequencies were 21% ± 5, and 24% ± 5 for

90   the same AAV6 donors (**Fig. 1c).** AAV6 transduction alone showed <2% YFP positive cells, while

91   mock cells that underwent electroporation but not AAV transduction had no detectable

92   fluorescence. We measured the efficiency of modification in CB and PB cells transduced with the

93   PGK-IDUA virus lacking the reporter (PGK-IDUA) by genotyping single cell-derived colonies

94   from colony formation assays (CFAs) (**Extended Data Fig. 2a, b)**. In these cells, the frequencies

95   of modification were 54% ± 10, and 44% ± 7 in CB and PB-HSPCs, considerably higher than the

96   larger, YFP-containing cassettes, suggesting that efficiency is dependent on insert size (**Fig. 1c)**.

97   Based on these targeting frequencies we conclude that our genome editing protocol is highly

98   efficient and reproducible for human CB and PB-derived HSPCs.

99   We also characterized the genomic modifications at the *CCR5* loci, by quantifying the fraction of

100   targeted alleles in bulk DNA preparations using droplet digital PCR (ddPCR) (**Extended Data**

101   **Fig. 2c, d)**. This data allowed us to estimate the distribution of cells with one (mono-allelic) or two

102   (bi-allelic) alleles targeted and indicated that for the YFP constructs, 65% to 100% of the cells had

103   mono-allelic modification **(Supplementary Data 1)**. Consistent with this, genotyping of YFP-

104   positive colonies in CFAs showed an average mono-allelic modification frequency of 80% ± 7.5

105   **(Fig. 1d).**

106

Enhanced IDUA secretion from edited HSPCs and HPSC-derived macrophages

108   A central concept in our approach is that HSPCs and their progeny will secrete stable, supra-

109   endogenous IDUA levels that can cross-correct the lysosomal defect in affected cells. Examination

110   of modified HSPCs in culture showed that 3 days post-modification, three distinct cell populations

111   could be discerned based on YFP expression: high/medium/low (**Fig. 2a)**. YFP-high cells

112   exhibited persistent fluorescence in culture for at least 30 days, demonstrating stable integration

113   of the cassettes. YFP-negative cells had no detectable YFP expression at the time of selection,

114   though 1% of cells eventually became positive. Most cells with intermediate fluorescence

115   converted to YFP-high (80%) (**Fig. 2b)**. In these cultures with mixed YFP-positive and negative

116 cells, grown under expansion conditions, the fraction of YFP-positive cells remained stable for 30

117 days, suggesting that neither the modification, nor the overexpression of the enzyme, nor the

118 reporter impacted the cells' proliferative potential.

119 When compared to mock-treated cells expressing endogenous IDUA levels, YFP-high cells

120 secreted 250-fold and 25-fold more enzyme for the SFFV and PGK-driven cassettes respectively,

121 while cell lysates expressed 600 and 50-fold more enzymatic activity (**Fig. 2c).** When YFP-high

122 IDUA-HSPCs were co-cultured with patient-derived MPSI fibroblasts, they led to a decrease in

123 the average area of lysosomal-associated membrane protein 1 (LAMP-1) positive specks,

124 consistent with reduced lysosomal compartment size and cross-correction of the cellular phenotype

125 (**Fig. 2d**). These data confirm that IDUA-HSPCs secrete supra-physiological IDUA levels and that

126 the secreted IDUA has the post-translational modifications required for uptake into MPSI cells,

127 thereby biochemically cross-correcting the patient derived enzyme deficient cells.

128 For IDUA-HSPCs to successfully correct biochemical abnormalities in the organs affected in

129 MPSI, they must differentiate into monocytes that will migrate to and differentiate into tissue-

130 resident macrophages such as microglia (brain), Kupffer cells (liver), osteoclasts (bone), and

131 splenic macrophages to deliver the enzyme and cross-correct enzyme-deficient cells. To confirm

132 that IDUA-HSPC could generate macrophages and that these cells can continue to produce IDUA,

133 we differentiated these cells in culture and assayed for IDUA activity (**Supplementary Data 2**

134 **and Fig. 2e).** These IDUA-HPSC-derived macrophages secreted 182-fold and 69-fold more IDUA

135 for the SFFV and PGK-driven cassettes respectively than mock-cell-derived macrophages.

136 Likewise, lysates exhibited 75-fold and 24-fold more IDUA activity (**Fig. 2f).** These data

137 established that IDUA-HPSC can reconstitute monocyte/macrophages *in vitro* and that IDUA-

138 HPSC-derived macrophages also exhibit enhanced IDUA expression.

139

140 **Preserved repopulation and differentiation potential in IDUA-HSPCs**

141 To determine if HSPCs that have undergone genome editing can engraft *in vivo,* we performed

142 serial engraftment studies into NOD-scid-gamma (NSG) mice. We first tested cells modified with

143 the SFFV and PGK constructs expressing YFP, which allowed us to identify the modified cells *in*

144   *vivo*. Equal numbers of CB and PB-derived mock, YFP-negative (YFP-), and YFP-positive

145   (YFP+) cells were transplanted intra-femorally into sub-lethally irradiated 6 to 8-week-old mice.

146   Primary human engraftment was measured 16 weeks-post-transplantation by establishing the

147   percent of bone marrow (BM) cells expressing both human CD45 and human leukocyte antigens

148   (HLA-ABC) out of total mouse and human CD45+ cells (**Extended Data Figure 3b and Fig. 3a**).

149   For the PGK-driven constructs, the median frequencies of hCD45$^+$/HLA$^+$ cells in BM were as

150   follows: Mock 76.25% (min-max: 46.4-95.4%), YFP- 21.5% (0.06-89.5%), YFP+ 4.3% (0.06-

151   96%) (**Fig. 3b**). This showed a 5-fold drop in repopulation capacity in cells that underwent HR-

152   GE (YFP+) compared to cells that did not but were also exposed to RNP, AAV transduction, and

153   sorting (YFP-). The median frequency of human cells expressing YFP was 0.6% (0-18.5%) and

154   95.8% (1-100%) for YFP- and YFP+ transplants respectively, confirming that edited cells had

155   engrafted in these mice (**Fig. 3c**). Human cells were also found in the peripheral blood with

156   frequencies of 31/3.1/1.1% in mock, YFP-, and YFP+ cells respectively (**Fig. 3b)**.

157   The apparent engraftment advantage of cells that had not undergone HR-GE was also examined

158   by transplanting bulk populations of HSPCs modified with the cassette without YFP. In two

159   independent experiments, an initial fraction of targeted alleles of 28% (43% modified cells)

160   declined to 5.2% and 6.5% in the engrafted cells (8 and 10% modified cells) despite big differences

161   in human chimerism **(Fig. 3c, d).** This corresponded to a 5-fold drop in donor 1 and 4-fold drop in

162   donor 2. Interestingly, this fall in targeted alleles showed significant variation in individual mice

163   (2 to 10-fold). This data re-demonstrated the observed loss in engraftment potency after

164   modification and supports the idea that the HR-GE cell population has fewer clones with long-

165   term repopulation potential.

166   Serial transplantation is considered a gold standard to assess self-renewal capacity of HSCs. For

167   secondary transplants, we isolated human CD34$^+$ cells from the bone marrow of primary mice and

168   transplanted into secondary mice. YFP+ engrafted mice showed 3.9% (0.8-9.7%) median human

169   cell chimerism, while YFP- mice showed 30.4% (7.7-48.2%) (**Fig. 3e**). YFP expression in the

170   engrafted human cells was 0.27% (0-1.35) for YFP- cells, and 41.9% (20.8-100) for YFP+ cells

171   (**Fig. 3f**). Similar levels of human cell chimerism were observed for the SFFV-driven constructs

172   in serial transplants (**Extended data Fig. 4**). Collectively, the presence of YFP expressing cells at

173   16- and 32-weeks post-modification demonstrates that cells with long-term repopulation potential

174   can be edited, albeit at lower frequencies than cells that did not undergo HR-GE.

175   To establish the modified cells' ability to differentiate into multiple hematopoietic lineages, we

176   looked *in vitro* using colony formation assays (CFAs) and *in vivo* after engraftment in NSG mice.

177   In CFAs, CB-derived and PB-derived YFP-expressing cells gave rise to all progenitor cells at the

178   same frequencies as mock treated and YFP- cells, indicating that IDUA-HSPCs can proliferate

179   and differentiate into multiple lineage progenitors in response to appropriate growth factors

180   (**Extended data Fig. 5a, b**). *In vivo*, B, T and myeloid cells were identified using the human CD19,

181   CD3, and CD33 markers. Compared to mock cells that demonstrated a roughly equal distribution

182   of B and myeloid cells (1:1, CD19:CD33) 16 post-transplantation, YFP+ and YFP- cells showed

183   skewing towards myeloid differentiation (YFP+ = 1:16, and YFP- = 1:5) (**Extended data Fig. 5c**).

184   Subgrouping the mice based on human-cell chimerism showed that in mice with low human

185   engraftment (< 1%), the cells exhibited myeloid skewing (72% of these mice had human myeloid

186   cell fractions >90%). In contrast, mice with chimerism greater than 20% displayed a mean CD33+

187   fraction of 55% (**Extended Data Fig. 5d, e)**. This myeloid bias was not observed in circulating

188   cells in the peripheral blood or in secondary transplants (**Extended data Fig. 5f, g**). These data

189   suggest that myeloid skewing is tightly and inversely correlated with the degree of human cell

190   engraftment, and that neither the genome editing process, nor IDUA expression, affects the

191   modified cell's capacity to differentiate into multiple hematopoietic lineages *in vitro* or *in vivo*.

192

193   **Biochemical correction in NSG-IDUA$^{X/X}$ mice by human IDUA-HSPCs**

194   To determine the potential of human IDUA-HSPCs to correct the metabolic abnormalities of

195   MPSI, we established a new mouse model of the disease capable of engrafting human cells. NSG-

196   IDUA$^{X/X}$ mice replicated the phenotype of patients affected with MPSI[1] and previously described

197   immunocompetent[19,20] and immunocompromised[21] MPSI mice (**Supplementary Data 3,**

198   **Extended data Fig. 6 and 7)**. We focused the correction experiments on cells expressing IDUA

199   under the PGK promoter, as this promoter has better translational potential because it has

200   decreased enhancer-like activity compared to SFFV[22]. As a first series of experiments, we

201   examined PB-derived cells in which the modification did not include a selection marker. In bulk

202 transplants, the median human cell chimerism in the bone marrow was 62.2% (min=39.2,
203 max=96.7%) and no statistically significant differences in human engraftment were observed
204 between NSG-IDUA$^{X/X}$ and NSG-IDUA$^{W/X}$ mice (**Fig. 4a**). The editing frequencies before
205 transplantation were 30% of *CCR5* alleles (by ddPCR) and 46% of cells (as measured by CFA).
206 GAG urinary excretion was measured at 4, 8, and 18 weeks post-transplantation in NSG-IDUA$^{X/X}$
207 and IDUA$^{W/X}$ mice. Biochemical correction was detectable after 4 weeks, and the trend towards
208 normalization increased over time (**Fig. 4b**). These kinetics are consistent with the time lag needed
209 for the genetically engineered human HSPCs to engraft, expand, and migrate to affected tissues
210 and "cross correct" diseased cells. At 18 weeks, NSG-IDUA$^{X/X}$ mice that had been transplanted
211 with IDUA-HSPCs (X/X Tx) excreted 65% less GAGs in the urine compared to sham-treated
212 NSG-IDUA$^{X/X}$ mice (X/X sham) (median Tx= 387.2 μg/mg of creatinine, sham=1,122 μg/mg)
213 though the levels had not normalized (W/X sham =155 μg/mg) (**Fig. 4b**). Transplantation of
214 IDUA-HSPCs also resulted in increased IDUA activity to 11.3%, 50.1%, 167.5%, and 6.8% of
215 normal in serum, liver, spleen, and brain respectively (compared to undetectable in X/X sham) and
216 resulted in normalization of tissue GAGs in liver and spleen (**Fig. 4c, d**). Supra-endogenous levels
217 of activity were detected consistently in the spleen and occasionally in the liver of some mice. This
218 can be attributed to robust human cell engraftment in these organs, as demonstrated by increases
219 in liver and spleen size in transplanted mice regardless of genotype (**Extended Data Fig. 8b-e**).

220 Because we could not discount the contribution of unmodified cells to the observed correction in
221 bulk transplants, we then examined the effect of HSPCs expressing IDUA and YFP under the PGK
222 promoter after FACS. Of 15 NSG-IDUA$^{X/X}$ and 5 NSG-IDUA$^{W/X}$ 13/15 and 5/5 were deemed to
223 have engrafted (human chimerism in the bone marrow >0.1%). The median percent human
224 chimerism was 4.2% (W/X) and 9.9% (X/X). IDUA-YFP-HSCPs increased IDUA tissue activity
225 to 2.9%, 7.4%, 25.5%, and 1.3% of normal in serum, liver, spleen, and brain respectively (**Fig.
226 4a**). Tissue and urine GAGs were also significantly reduced (**Fig. 4f**). Together, this data indicates
227 that IDUA-HSPCs can correct the metabolic abnormalities in MPSI and suggest that the degree of
228 correction correlates with human cell chimerism.

229

230 **Phenotypic correction in NSG-IDUA$^{IDUAX/X}$ mice by human IDUA-HSPCs**

231 To investigate the effect of IDUA-HSPCs on the skeletal and neurological manifestations of MPSI,

232 sham-treated and transplanted mice also underwent whole body micro-CT and neurobehavioral

233 studies 18 weeks after transplantation. The effect of transplantation on the skeletal system was

234 measured on the skull parietal and zygomatic bone thickness and the cortical thickness and length

235 of femoral bones. In experiments where the mice were transplanted using unselected cells and

236 where human cell chimerism was high (**Fig. 4d**), we observed almost complete normalization of

237 bone parameters by visual inspection and on CT scan measurements (**Fig. 5a, b**). Mice transplanted

238 with cells that had undergone selection showed partial but statistically significant reduction in the

239 thickness of the zygomatic, parietal bones, and femur (**Fig. 5c**).

240 We also examined the open field behavior, passive inhibitory avoidance, and marble-burying

241 behavior of sham-treated and transplanted mice. Transplantation of bulk-edited cells resulted in

242 marked reduction in locomotor activity and long-term memory, regardless of genotype **(Extended**

243 **Data Fig. 9a-c**). We suspected that high human-cell chimerism was detrimental for the overall

244 health of the mice. Consistent with this, we observed growth restriction following human cell

245 transplantation in both homozygous and heterozygous mice (**Extended Data Fig. 8a)**. This likely

246 represents a toxicity artifact of this xenogeneic transplant model and could be explained in part by

247 the defective erythropoiesis seen in these xenograft models[23] when human cell chimerism is high.

248 In contrast, NSG-IDUA$^{X/X}$ mice transplanted with YFP-selected cells in which human cell

249 chimerism was low exhibited locomotor activity indistinguishable from their sham-treated

250 heterozygous littermates, and markedly higher that the sham-treated knock-out mice (**Fig. 5d**).

251 These mice also had increased vertical counts at all time points and demonstrated the same

252 exploratory behavior as sham heterozygous mice (**Fig. 5e**). Transplantation of IDUA-HSPCS in

253 NSG-IDUA$^{X/X}$ also enhanced performance in the passive inhibitory avoidance test 24 hours later

254 (**Fig. 5f**). Digging and marble burying behavior also improved but did not completely normalize

255 (**Fig. 5g**).

256

Safety of our genome editing strategy

258 To assess genotoxicity and characterize the off-target repertoire of our *CCR5* guide, we used the

259 bioinformatics-based tool COSMID (CRISPR Off-target Sites with Mismatches, Insertions, and

260   Deletions)[24]. Off-target activity at a total of 67 predicted loci was measured by deep sequencing

261   in two biological replicates of CB-derived HSPCs. In each replicate we compared the percent

262   Indels measured in mock and cells electroporated with RNP with either wild-type (WT) Cas9 or a

263   higher fidelity (HiFi) Cas9[25]. Five of the 67 sites were located within repetitive elements and Indel

264   rates could not be assigned to specific loci in this group **(Extended Data Fig. 10)**. For the

265   remaining 62 genomic locations, sites were deemed true off-targets if: 1) the percent of indels at

266   the site was > 0.1% (limit of detection), 2) off-target activity was present in both biological

267   samples, and 3) indels were higher in the RNP compared with the mock samples. Given these

268   criteria only 4 sites were deemed to be true off-targets **(Figure 6a, b)**. For all of these sites the

269   frequency of Indels was < 0.5% and the use of the HiFi Cas9 abolished off-target activity entirely

270   while maintaining on-target efficiency. Only one exonic site was found in the *SUOX* gene (sulfite

271   oxidase). The highest off-target activity measured at this site was 0.128%, which was reduced

272   below the limit of detection with HiFi Cas9. These data suggest that our *CCR5* sgRNA combined

273   with either WT Cas9 or especially HiFi Cas9 has negligible off-target activity on a large screen of

274   bioinformatically predicted sites.

275   Collectively, we performed 191 autopsies (101 mice used in primary engraftment, 50 in secondary

276   engraftment, and 40 in NSG-IDUA$^{X/X}$ correction studies) in which no gross tumors were found.

277   Three tumor-like masses were evaluated by histology and confirmed to be abscesses. These 191

278   mice were transplanted with a combined dose of 90 million human cells that underwent our

279   genome editing protocol. Considering that the median age for HSCT in MSPI patients is around

280   one year[26], and that an average one year-old is 10 Kg, the total number of modified cells used in

281   this study is roughly equivalent to two clinical doses of 4.5×10^6 CD34 HSPCs/kg. We conclude

282   that the apparent lack of tumorigenicity and the low off-target activity of the *CCR5* sgRNA provide

283   evidence for the safety our modification strategy.

284

Discussion

285

286   We describe an efficient application of RNP and AAV6-mediated template delivery to overexpress

287   IDUA from a safe harbor locus in human CD34+ HSPCs. The suitability for *CCR5* to be a safe

288   harbor for the insertion and expression of therapeutic genes has been described [22,27]. For LSDs like

289   MPSI, the use of the safe harbor has several advantages compared to genetic correction of the

290   affected locus: 1) it enhances therapeutic potential, as it allows for supra-endogenous expression,

291   2) it circumvents design for specific mutations, 3) the coding sequences can be engineered with

292   enhanced therapeutic properties, e.g., crossing the blood brain barrier[28], and 4) it is versatile and

293   easily adaptable to other LSDs.

294   Our approach attempts to commandeer the patient's own hematopoietic system to express and

295   deliver lysosomal enzymes. Because of the unique ability of this system to generate tissue

296   macrophages that can migrate into affected tissues to deliver the enzyme[29], an HSCT-based

297   approach will likely be more effective that other potential enzyme depots[29] for harder-to-treat

298   organs like the CNS and bone. The autologous source for this approach also improves on safety,

299   by eliminating the morbidity of graft rejection, graft-versus-host disease, and immunosuppression,

300   and can lead to earlier intervention by obviating the need for donor matching.

301   We studied the self-renewal and multi-lineage differentiation capacity of the modified cells to

302   establish the potential of genome-edited HSPCs as one-time therapy for MPSI. Our data

303   demonstrates that this approach can modify cells with long-term repopulation potential and

304   preserves multi-lineage differentiation capacity *in vivo* and *in vitro*. In experiments comparing

305   engraftment potential of the YFP- and YFP+ cells, as well as in bulk transplantation experiments,

306   cells that underwent HR-GE had approximately a 5-fold lower long-term engraftment capacity.

307   This is not entirely surprising, as HR efficiencies are higher in cycling cells[30,31] and therefore

308   would be expected to be lower in stem than progenitor cells. The lower engraftment could also

309   represent a negative effect of expression of a foreign fluorescent protein in HSCs[32] as previously

310   substituting a truncated form of the low-affinity nerve growth factor receptor resulted in higher

311   engraftment frequencies than using a fluorescent protein to mark HR-GE cells[15]. As observed in

312   allo-HSCT, this engraftment challenge might be circumvented by using larger doses of genome

313   edited cells. This can be achieved through selection followed by *in vitro* expansion in optimized

12

314   culturing conditions that could help maintain self-renewal capacity, perhaps including recently

315   discovered small molecules such as UM171[33] and SR1[34].

316   *Ex vivo* manipulation of the HSPCs allows for a thorough examination of the genotoxicity and the

317   magnitude of biochemical potency of the cells before delivering the engineered cell drug product

318   to patients. Through a bioinformatics-guided strategy we identified four potential off-target sites

319   with minimal off-target activity. Fortunately, all were reversed by using a higher fidelity

320   nuclease[25]. The conclusion that our genome editing strategy is safe is also supported by the lack

321   of tumorigenicity in 191 mice transplanted with 90 million edited cells.

322   We examined the therapeutic potential of the edited HSPCs in a new model of MPSI capable of

323   robust human cell engraftment. Engraftment of the IDUA-HPSCs led to reconstitution of enzyme

324   activity and decreased GAG storage in multiple organs. Notably, small changes in circulating and

325   tissue IDUA lead to significant phenotypic improvements. This is not surprising as even a small

326   fraction of normal IDUA activity can dramatically improve the physical manifestations of MPSI.

327   Mean IDUA activity in fibroblasts from patients with severe MPSI is 0.18% (range 0–0.6), while

328   0.79% residual activity (range 0.3–1.8) results in mild disease (minimal neurological involvement

329   and the possibility of a normal life span)[35]. In fact, healthy individuals can be found with enzymatic

330   activity as low as 4%[36]. Our data constitutes the first study to show symptomatic correction of an

331   LSD with human genome-edited HSPCs and provides support for the further development of this

332   strategy for the treatment of the visceral, skeletal, and neurological manifestations in MPSI.

333   In sum, these pre-clinical studies provide proof-of-concept evidence of the safety and efficacy of

334   using genome-edited human HSPCs modified to express a lysosomal enzyme to correct the

335   biochemical, structural, and behavioral phenotype of a mouse model of MPSI, a canonical

336   lysosomal storage disease. Moreover, this work provides specific evidence of safety and efficacy

337   to support the optimization and development of this strategy into a clinical protocol to treat patients

338   with MPSI and a platform approach to potentially treat other lysosomal storage disorders.

339   **Figure Legends**

340   **Fig. 1 | Efficient CRIPR/Cas9-mediated integration of IDUA overexpression cassettes into**

341   **the *CCR5* locus in human CD34+ HSPCs. a,** Schematic of targeted integration of IDUA and

342   expression cassettes. The AAV6 genome was constructed to have 500bp arms of homology

343   centered on the cut site, and the IDUA sequence placed under the control of the SFFV or the PGK

344   promoter. In two DNA templates, YFP was expressed downstream of IDUA using the self-

345   cleaving P2A peptide. Analysis was performed 3-days post-modification **b,** FACs and histogram

346   plots of mock and human HSPCs that underwent RNP and AAV6 exposure with YFP-containing

347   expression cassettes. **c**, Targeting frequencies in CB (red) and PB (blue)-derived HSPCs read by

348   percent fluorescent cells in YFP expressing cassettes and percent colonies with targeted *CCR5*

349   alleles by single cell-derived colony genotyping in cassettes without the reporter. RNP+AAV6

350   conditions with YFP templates, CB=20, PB=11. For the template without selection CB=6, PB=6.

351   Lines indicate mean and SD. **d**, Distribution of wildtype (WT), mono and bi-allelically modified

352   cells in YFP-positive HSPCs (n=400, 3 human donors).

353   **Fig. 2 | Enhanced IDUA expression by IDUA-HSPCs and derived macrophages. a,** FACS plot

354   shows distinct populations based on YFP expression 3 days post-modification. **b,** Persistent YFP

355   expression up to 30 days cultures. **c**, Fold increase in IDUA secretion and intracellular expression

356   by YFP-high, YFP-low, and YFP-negative populations compared to mock cells. **d**, Average

357   LAMP-1+ area in MPSI fibroblasts co-cultured with IDUA-HSPCs. Each dot represents a cell. **e**,

358   Human CD34, CD14, and CD11b marker expression in HSPC-derived macrophages and human

359   monocyte-derived macrophages after *in vitro* differentiation compared to undifferentiated cells

360   (CD34+ HSPCs). Macrophage morphology and YFP expression after differentiation. **f**, Fold

361   increase in IDUA secretion and intracellular expression in HSPC-macrophages modified with

362   SFFV and PGK expression cassettes. **c, e, and f**, Each dot represents average of triplicates in a

363   human cell donor. All data expressed as mean ± SD, *** p < .001 in two-sided unpaired t-test.

364   **Fig. 3 | IDUA-HSPCs maintain long-term repopulation capacity. a,** Schematic and

365   representative FACS plots showing phenotyping by flow of human, myeloid, B-cell, and targeted

366   cells after engraftment. **b,** Percent human cell chimerism in bone marrow (BM) and peripheral

367   blood (PM) of mice 16-weeks post-transplant with CB (blue) and PB (red)-derived HSPCs targeted

14

368   with PGK cassettes; mock (n=11), YFP- (n=21), and YFP+ (n=36). Each point represents a mouse.

369   **c**, Percent human, YFP+ cells in BM of mice in BM 16-weeks post-transplant. **d,** Percent human

370   cell chimerism in BM in mice transplanted with bulk cells without selection with two different

371   human cell donors. **e,** Percent modified alleles in engrafted cells by ddPCR. 30% was the starting

372   allele modification frequency for both human donors. **f,** Percent human cell chimerism in BM of

373   mice in secondary transplants 32 weeks after genome editing; YFP- (n=10), and YFP+ (n=10). **e,**

374   Percent human, YFP+ cells in BM of mice in secondary transplants.

375   **Fig. 4 | Biochemical correction in NSG-IDUA$^{X/X}$ mice by human IDUA-HSPCs.** IDUA activity

376   and GAG accumulation in heterozygous sham-treated (W/X sham- clear), heterozygous

377   transplanted (W/X Tx- black), homozygous sham-treated (X/X sham- blue), and homozygous

378   transplanted (X/X Tx- red) mice. **a**, Percent human and YFP+ cells in BM in experiments using

379   bulk and sorted cells. **b,** Urinary GAGs at 4,8, and 18 weeks in experiments using bulk cells (n=5

380   mice per cohort, two measurements per mouse). **c,** Serum and tissue IDUA activity in experiments

381   using bulk cells (n=5 per cohort). **d**, Fold GAG storage in liver and spleen (normalized by W/X

382   sham, n=5 per cohort). **e**, Serum and tissue IDUA activity in experiments using sorted cells (n=5

383   for W/X Tx and sham mice, and n=13 for X/X Tx and sham mice). **f**, Fold GAG urinary excretion

384   and tissue storage in experiments using sorted cells (normalized by W/X sham). Median values in

385   all scatter plots. **d** and **f** show box plots with whiskers at the 5-95th percentiles. ****: p < .0001 in

386   one-way ANOVA test. Post hoc comparisons were made with the Tukey's multiple comparisons

387   test.

388   **Fig. 5 | Phenotypic restitution in NSG-IDUA$^{X/X}$ mice by human IDUA-HSPCs.** Behavioral and

389   skeletal assessment in: W/X sham (clear or gray, n=11), X/X sham (blue, n=10), and X/X Tx (red,

390   n=11). **a,** Representative photos showing facial features in mice transplanted with bulk cells. **b,**

391   Bony features in mice transplanted with bulk and **c,** sorted cells. Box plots with whiskers show

392   min and max. **d**, Ambulatory distance in mice transplanted with sorted cells. W/X sham vs. X/X

393   sham: **; W/X sham vs. X/X Tx: n.s.; X/X sham vs. X/X sham: *. **e,** Vertical rearing in mice

394   transplanted with sorted cells. W/X sham vs. X/X sham: *; W/X sham vs. X/X Tx: n.s.; X/X sham

395   vs. X/X sham: *. **f**, Memory retention in mice transplanted with sorted cells. **g**, Quantification of

396   digging behavior in mice transplanted with sorted cells. Data shown as mean ± SEM. **b-g**,

397   Comparisons between groups were performed using one-way ANOVA test and post-hoc

398   comparisons were made with the Tukey's multiple comparisons test. *: p < .05, **: p < .01, ***:

399   p < .001, and ****: p < .0001. Open field testing and vertical rearings were analyzed using within-

400   subject modeling by calculating the are under the curve for each mouse within the first five minutes

401   and comparing between groups with one-way ANOVA.

402   **Fig. 6 | OFF-target analysis of the *CCR5* sgRNA. a,** Percent reads with Indels at 62 off-target

403   sites predicted using COSMID. For each site, red dots indicate samples treated with WT Cas9 and

404   blue dots indicate samples treated with HiFi Cas9. The limit of detection for NGS is 0.1% and is

405   indicated on the graph by a dashed line. **b,** Table summarizing five bona fide off-target sites. For

406   all of these sites the percent of Indels was < 0.5%. For 4 of these sites, the use of the HiFi Cas9

407   abolished off-target activity.

408

409   **Extended Data**

410   **Extended Data Fig. 1 | Characterization of the *CCR5* sgRNA. a,** Indel frequency in CB and

411   PB-derived cells by the RNP complex. **b**, Representative indel distribution from next generation

412   sequencing reads. **c,** Histogram of CCR5 protein expression in mock-treated and RNP-treated cells

413   showing an 80% reduction in protein expression after indel induction. **d,** Sample sequence traces

414   around the CCR5 sgRNA sequence (gray box. PAM in red) in mock samples and RNP-treated CB-

415   derived HSPCs showing predominant single A insertion. **e,** Summary of indels and respective

416   frequencies by next-generation sequencing.

417   **Extended Data Fig. 2 | Efficiency of modification at the *CCR5* locus. a,** Schematic showing the

418   three primer-based genotyping scheme to distinguish mono and bi-allelic integration into the

419   *CCR5* locus on CFA-derived colonies. This strategy did not distinguish WT versus alleles with

420   indels (NHEJ). **b,** Example agarose gels of 40 colonies genotyped in this manner. A single 1.1Kb

421   band was interpreted as WT/NHEJ in both alleles, while a single 0.6 Kb band was read as bi-allelic

422   integration. **c,** Schematic of probe design for ddPCR analysis. Fraction of modified alleles was

423   obtained by using a second reference probe to the *CCRL2* gene also on chromosome 3p. **d,**

424   Originally two probes where each straddled a 5' or 3' homology arm were designed. The accuracy

425   of the assays was verified and compared using genomic DNA from colonies derived from mono-

426 allelic cells (0.5 fraction of alleles modified). Error bars indicate 95% CI. The 3' HA probe was

427 selected. **e,** *CCR5* allele targeting frequencies in CB (red) and PB (blue)-derived IDUA expressing

428 HSPCs as measured by ddPCR. The mean fraction of modified alleles in CB-HSPCs for the SFFV

429 and PGK-driven donors was 23% ± 6, 21% ± 8 respectively, and 41% ± 5 for the cassette lacking

430 YFP.

431 **Extended Data Fig. 3 | a,** Gating scheme for quantification of human CD34+, CD14+, and

432 CD11b+ cells in human HSPCs maintained in standard CD34+ cytokine media (top panel) or

433 media containing M-CSF to induce monocyte/macrophage differentiation (bottom panel). Single

434 and live cell discrimination not shown. **b,** Gating scheme used to analyze human cell engraftment

435 and cell lineages after transplantation. Representative plots for quantification of mouse and human

436 hematopoietic (mCD45+ and hCD45+), all human (CD45+/HLA-ABC+), human B (CD19+),

437 human myeloid (CD33+), human T (CD3+), and YFP+ cells.

438 **Extended Data Fig. 4 | IDUA-HSPCs modified with SFFV containing cassettes are capable**

439 **of long-term repopulation and multi-lineage differentiation. a,** Percent human cell chimerism

440 in BM and PM of mice 16-weeks post-transplant with CB-derived HSPCs. **b,** Percent YFP+ cells

441 in BM and PB of mice in primary transplants. **c,** Percent human cell chimerism in BM of mice in

442 secondary transplants (32 weeks). **d,** Percent YFP+ cells in BM and PB of mice in secondary

443 transplants.

444 **Extended Data Fig. 5 | | IDUA-HSPCs maintain multi-lineage differentiation potential. a,**

445 Representative photos showing morphology and YFP expression in CFU-GM, BFU-E, and CFU-

446 E colonies. **b,** Colony formation unit frequency in mock, YFP- and YFP+ cells (only cells modified

447 with PGK driven cassettes are shown). **c,** Box plot with whiskers (min to max) showing percent

448 human CD33+ (myeloid), CD19+ (B) cell in the BM of mice transplanted with mock, and FAC-

449 sorted YFP+ and YFP- cells in primary transplant mice. Each point represents data from a single

450 mouse. **d,** Scatter plot from mice with human cell chimerism <1% against the percent myeloid

451 cells. **e,** Scatter plot of human cell chimerism >20% against the percent myeloid cells. **f,** Percent

452 human CD33+, CD19+, and CD3+ (T) cells in the PB of mice transplanted with mock, and FAC-

453 sorted YFP+ and YFP- cells. **g,** Box plot with whiskers (min to max) showing percent human

454   CD33+ (myeloid), CD19+ (B) cell in the BM secondary transplant mice. ****: p < .0001 in one-

455   way ANOVA test. Post hoc comparisons were made with the Tukey's multiple comparisons test.

456   **Extended Data Fig. 6 | Biochemical characterization of NSG-IDUA$^{X/X}$ mice. a,** Tissue IDUA

457   enzyme activity in tissues. **b,** Tissue GAGs measure by dimethylmethylene blue reactivity. **c,**

458   Histological sections of paraffin-embedded tissues stained with bromophenol blue (brain) or alcian

459   blue (liver, spleen, and heart). Brain sections on X/X mice showed distended and vacuolated

460   Purkinje cells. In liver, spleen, and heart blue deposit-laden cells can be seen throughout. **d**, Age-

461   related progression of urinary GAGs excretion. **e,** Survival analysis comparing NSG, W/X, and

462   X/X during one year of observation (n=10). **f**, Physical dysmorphisms and visceral enlargement.

463   **g,** Total body, liver, spleen and heart weight in W/X and X/X mice.

464   **Extended Data Fig. 7 | Phenotypic characterization of NSG-IDUA$^{X/X}$ mice. a,** Reconstructed

465   micro-CT images of skull and zygomatic and parietal bone thickness. **b**, CT longitudinal sections

466   of femurs, and cortical thickness, width, and length measurements. **c**, Spontaneous locomotion in

467   open field testing. **d,** Vertical rearing counts for W/X and X/X mice during 10-minute observation

468   in the open field chamber. **e**, Long-term memory in passive inhibitory avoidance test. **f**, Defensive

469   digging in the marble burying task. Total 5 female mice per genotype. Data are presented as mean

470   ± SD for **a-e**, mean ± SEM for **f-g**. Comparisons between groups were performed using unpaired

471   t-test. *: p < .05, **: p < .01, ***: p < .001, and ****: p < .0001. Open field testing was analyzed

472   using within-subject modeling for the entire time course by calculating the are under the curve for

473   each mouse and comparing between genotypes with a t-test.

474   **Extended Data Fig. 8 | Body and organ size in transplantation experiments using bulk IDUA-**

475   **HSPCs. a,** Total body weight in heterozygous transplanted (W/X Tx- dark gray), heterozygous

476   sham-treated (W/X sham- clear), homozygous transplanted (X/X Tx- red), and homozygous sham-

477   treated (X/X Tx- blue) mice. **b,** Normalized liver weight. **c,** Normalized spleen weight. **d,**

478   Normalized heart weight. **e**, Percent human, B (CD19+), and myeloid (CD33+) cells in the spleen

479   of X/X Tx mice measured 18 weeks post-transplant.

480   **Extended Data Fig. 9 | Neurobehavioral studies in mice transplanted with bulk IDUA-**

481   **HSPCs. a,** 10-minute time course of spontaneous locomotor behavior in heterozygous transplanted

482   (W/X Tx-black), heterozygous sham-treated (W/X sham- celar or gray), homozygous transplanted

483   (X/X Tx- red), and homozygous sham-treated (X/X Tx- blue) mice. No comparison was found to

484   be significant. **b,** Total ambulatory distance in 10 minutes. **c,** Memory for inhibitory avoidance

485   training (24h). No comparison was found to be significant. Comparisons between groups were

486   performed using one-way ANOVA test and post-hoc comparisons were made with the Tukey's

487   multiple comparisons test. Open field testing and vertical rearings were analyzed using within-

488   subject modeling by calculating the are under the curve for each mouse for the entire time course

489   and comparing between groups with one-way ANOVA.

490   **Extended Data Fig. 10 | Comprehensive off-target site analysis of the *CCR5* sgRNA.** Table

491   lists all 67 COSMID predicted sites, the sequence, genomic location, percent Indels in two

492   experiments and the number of reads for each site and in each experiment. Samples with percent

493   Indels > 0.1% are highlighted in pink. Samples with low coverage are highlighted in light yellow.

494   CCR5 _OT63 through 67 were located within a repetitive element. Three of these had primers that

495   should have been unique per locus but the Indel analysis showed that this was not the case.

496   Therefore, the true off-target rate at these sites were not ascertainable but should all be < 0.5%.

497   **Methods**

498

499   **AAV donor plasmid construction**

500   The CCR5 donor vectors have been constructed by PCR amplification of ~ 500 bp left and right

501   homology arms for the CCR5 locus from human genomic DNA. SFFV, PGK, IDUA sequences

502   were amplified from plasmids. Primers were designed using an online assembly tool (NEBuilder,

503   New England Biolabs, Ipswich, MA, USA) and were ordered from Integrated DNA Technologies

504   (IDT, San Jose, CA, USA). Fragments were Gibson-assembled into a the pAAV-MCS plasmid

505   (Agilent Technologies, Santa Clara, CA, USA).

506

507   **rAAV production**

508   We followed a protocol that has been previously reported with slight modifications[1]. Briefly, HEK

509   293 cells are transfected with a dual-plasmid transfection system: a single helper plasmid (which

510   contains the AAV rep and cap genes and specific adenovirus helper genes) and the AAV donor

511   vector plasmid containing the ITRs. After 2 days the cells are lysed by three rounds of freeze/thaw,

512   and cell debris is removed by centrifugation. AAV viral particles are purified by

513   ultracentrifugation in iodixanol gradient. Vectors are formulated by dialysis and filter sterilized.

514   Titers are performed using droplet digital PCR. Alternatively, viruses were amplified and purified

515   by Vigene Biosciences (Rockville, MD, USA).

516

517   **Electroporation and transduction of cells**

518   CCR5 sgRNA was purchased from TriLink BioTechnologies (San Diego, CA, USA) and was

519   previously reported[2]. The sgRNA was chemically modified with three terminal nucleotides at both

520   the 5′ and 3′ ends containing 2′ O-Methyl 3′ phosphorothioate and HPLC-purified. The genomic

521   sgRNA   target   sequence   (with   PAM   in   bold)   was: CCR5:   5'-

522   GCAGCATAGTGAGCCCAGAA**GGG**-3'. Cas9 protein was purchased from Integrated DNA

523   Technologies. RNP was complexed by mixing Cas9 with sgRNA at a molar ratio of 1:2.5 at room

524   temperature. CD34$^+$ HSPCs were electroporated 2 days after thawing and expansion by using the

525   Lonza Nucleofector 4D (program DZ-100) in P3 primary cell solution as follows: 10×10^6cells/ml,

526   300 µg/ml Cas9 protein complexed with 150 µg/ml of sgRNA, in 100 µl. Following

527   electroporation, cells were rescued with media at 37°C after which rAAV6 was added (MOI

20

528   15,000 of 15,000 titrated to maximize modification efficiency and cell recovery). A mock-

529   electroporated control was included in most experiments where cells underwent electroporation

530   without Cas9 RNP.

531

532   **Quantification of putative CCR5 gRNA off-target activity by deep sequencing**

533   Potential off-target sites in the human genome (hg19) were identified and ranked using the recently

534   developed bioinformatics program COSMID, allowing up to three base mismatches without

535   insertions or deletions and two base mismatches with either an inserted or deleted base (bulge).

536   The top ranked sites were further investigated. Off-target activity at a total of 67 predicted loci was

537   measured by deep sequencing in two biological replicates of CB-derived HSPCs.

538   Bioinformatically predicted off-target loci were amplified by two rounds of PCR to introduce

539   adaptor and index sequences for the Illumina MiSeq platform. All amplicons were normalized,

540   pooled and quantified using the PerfeCTa NGS quantification kit per manufacturer's instructions

541   (Quantabio, Beverly, MA, USA). Samples were sequenced using a MiSeq Illmina using 2 x 250bp

542   paired end reads. INDELs were quantified as previously described[3].

543

544   **Measuring insertions at the CCR5 locus with ddPCR**

545   Genomic DNA was extracted from either bulk or sorted populations using QuickExtract DNA

546   Extraction Solution. For droplet-digital PCR (ddPCR), droplets were generated on a QX200

547   Droplet Generator (Bio-Rad) per manufacturer's protocol. A HEX reference assay detecting copy

548   number input of the *CCRL2* gene was used to quantify the chromosome 3 input. The assay

549   designed to detect insertions at *CCR5* consisted of: F:5'-GGG AGG ATT GGG AAG ACA -3',

550   R:5'- AGG TGT TCA GGA GAA GGA CA-3', and labeled probe: 5'− FAM/AGC AGG CAT

551   /ZEN/GCT GGG GAT GCG GTG G/3IABkFQ-3'. The reference assay designed to detect the

552   *CCRL2* genomic sequence: F:5'-CCT CCT GGC TGA GAA AAA G -3', R:5'- GCT GTA TGA

553   ATC CAG GTC C -3', and labeled probe: 5'− HEX/TGT TTC CTC /ZEN/CAG GAT AAG GCA

554   GCT GT/3IABkFQ-3'. The accuracy of this assay was established with genomic DNA from a

555   mono-allelic colony (50% allele fraction) as template. Final concentration of primer and probes

556   was 900 nM and 250 nM respectively. 20 µL of the PCR reaction was used for droplet generation,

557   and 40 µL of the droplets was used in the following PCR conditions: 95° - 10 min, 45 cycles of

558   94° - 30 s, 57°C – 30 s, and 72° - 2 min, finalize with 98° - 10 min and 4°C until droplet analysis.

21

559   Droplets were analyzed on a QX200 Droplet Reader (Bio-Rad) detecting FAM and HEX positive

560   droplets. Control samples with non-template control, genomic DNA, and mock-treated samples,

561   and 50% modification control were included. Data was analyzed using QuantaSoft (Bio-Rad).

562

563   **HSPC Selection and Culturing**

564   Human CD34+ HSPCs mobilized peripheral blood purchased from AllCells (Alameda, CA, USA)

565   and thawed per manufacturer's instructions. CD34+ HSPCs were purified from umbilical cord

566   blood collected donated under informed consent via the Binns Program for Cord Blood Research

567   at Stanford University and used without freezing. In brief, mononuclear cells were isolated by

568   density gradient centrifugation using Ficoll Paque Plus. Following two platelet washes, HSPCs

569   were labeled and positively selected using the CD34+ Microbead Kit Ultrapure (Miltenyi Biotec,

570   San Diego, CA, USA) per manufacturer's protocol. Enriched cells were stained with APC anti-

571   human CD34 (Clone 561; Biolegend, San Jose, CA, USA) and sample purity was assessed on an

572   Accuri C6 flow cytometer (BD Biosciences, San Jose, CA, USA). Cells were cultured at 37°C,

573   5% CO_2, and 5% O_2 for 48 hours prior to gene editing. Culture media consisted of StemSpan

574   SFEM II (Stemcell Technologies, Vancouver, Canada) supplemented with SCF (100 ng/ml), TPO

575   (100 ng/ml), Flt3-Ligand (100 ng/ml), IL-6 (100 ng/ml), UM171 (35nM), and StemRegenin1 (0.75

576   mM).

577

578   **Colony-forming unit assay and clonal genotyping**

579   Cells were single-cell sorted into 96-well plates (Corning) pre-filled with 100 μl of methylcellulose

580   (Methocult, StemCell Technologies).

581   Single YFP+, YFP-, and mock-treated cells were sorted into methylcellulose media containing

582   SCF, IL3, erythropoietin and GM-CSF, conditions that support the growth of blood progenitor

583   cells: erythroid progenitors (burst forming unit-erythroid or BFU-E, and colony-forming unit-

584   erythroid or CFU-E), granulocyte-macrophage progenitors (CFU-GM), and multi-potential

585   granulocyte, erythroid, macrophage, megakaryocyte progenitor cells (CFU-GEMM).

586   After 14 days, colonies were counted and scored as BFU-E, CFU-M, CFU-GM and CFU-GEMM

587   per the manual for 'Human Colony-forming Unit (CFU) Assays Using MethoCult' from StemCell

588   Technologies. For DNA extraction from 96-well plates, PBS was added to wells with colonies,

589   and the contents were mixed and transferred to a U-bottomed 96-well plate. Cells were pelleted by

590   centrifugation at 300xg for 5 min followed by a wash with PBS. Finally, cells were resuspended

591   in 25 µl QuickExtract DNA Extraction Solution (Epicentre, Madison, WI, USA) and transferred

592   to PCR plates, which were incubated at 65°C for 10 min followed by 100°C for 2 min. For *CCR5*,

593   a 3-primer PCR was set up with a forward primer outside the left homology arm (5'-

594   CACCATGCTTGACCCAGTTT-3'), a forward primer binding the poly-adenylation signal in all

595   inserts (5'-CGCATTGTCTGAGTAGGTGT-3'), and a reverse primer binding inside the right

596   homology arm (5'-AGGTGTTCAGGAGAAGGACA-3'). Accupower premix was used for PCR

597   reaction and cycled at the parameters: 95° - 5 min, and 35 cycles of 95° - 20 s, 72°C – 60 s. DNA

598   fragments were detected by agarose gel electrophoresis.

599

600   **Macrophage differentiation and flow cytometry**

601   CD34+ HSPCs were seeded at a density of 2x105 cells/mL in untreated 6-well polystyrene plates

602   in differentiation medium (SFEM II supplemented with SCF (200 ng/ml), Il-3 (10 ng/mL), IL-6

603   (10 ng/mL), FLT3-L (50 ng/mL), M-CSF (10 ng/ml), penicillin/streptomycin (10 U/mL), and

604   cultured at 37 °C 5% CO_2, and 5% O_2. After 48 hours, non-adherent cells were removed from

605   plates and reseeded in new non-treated 6-well polystyrene plates at 2x105 cells/mL in

606   differentiation medium. Adherent cells were maintained in the same plates in maintenance medium

607   (RPMI supplemented with FBS (10% v/v), M-CSF (10 ng/ml) and penicillin/streptomycin (10

608   U/mL). After two weeks, adherent cells, comprising terminally differentiated macrophages, were

609   harvested by incubation with 10 mM EDTA and gentle scraping. For phenotypic analysis we

610   harvested 1x105 cells per condition resuspended in 100 µl staining buffer (PBS containing 2%

611   FBS and 0.4% EDTA). Non-specific antibody binding was blocked (5% v/v TruStain FcX,

612   BioLegend, #422302) and cells were stained with 2 µl of each fluorophore-conjugated monoclonal

613   antibody (30 minutes, 4°C, dark). Antibodies used were hCD34-APC (BioLegend #343510),

614   hCD14-BV510 (BioLegend #301842) and hCD11b-PE (BioLegend #101208). Propidium Iodide

615   (1 µg/mL)) was used to detect dead cells and cells were analyzed on a BD FACSAria flow

616   cytometer.

617

618   **Transplantation of CD34[+] HSPCs into NSG mice**

619   Targeted cells (sorted or bulk) were transplanted four to five days after

620   electroporation/transduction. YFP-negative (YFP-), and YFP-positive (YFP+) cells were isolated

621   using FACS and ~400,000 cells were transplanted intra-femorally into sub-lethally irradiated (2.1

622   Gy) 6 to 8-week-old mice. Approximately 1×10^6 cells HPSCs modified with cassettes without

623   YFP and were transplanted in bulk. Mice were randomly assigned to each experimental group and

624   analyzed in a blinded fashion.

625

626   **Assessment of human engraftment**

627   16-18 weeks after transplantation, samples of peripheral blood, bone marrow, and spleen were

628   harvested from recipient mice. Samples were treated with ammonium chloride to eliminate mature

629   erythrocytes. Non-specific antibody binding was blocked (10% vol/vol, TruStain FcX,

630   BioLegend), cells were stained (30 min, 4°C, dark), and analyzed by setting nucleated cell scatter

631   gates using a BD FACSAria II flow cytometer or BD FACSCanto II analyzer (BD Biosciences).

632   Cells were analyzed based on monoclonal anti-human HLA-ABC APC-Cy7 (W6/32, BioLegend),

633   anti-mouse CD45.1 PE-Cy7 (A20, eBioScience, San Diego, CA, USA), CD19 APC (HIB19,

634   BD511 Biosciences), CD33 PE (WM53, BD Biosciences), anti-mouse mTer119 PE-Cy5 (TER-

635   119, BD Biosciences), and CD3 PerCP/Cy5.5 (HiT3A, BioLegend) antibodies, and Propidium

636   Iodide to detect dead cells. Human engraftment was defined as HLA-ABC$^+$/HCD45$^+$ cells.

637

638   **IDUA activity assay**

639   IDUA enzyme activity was measured fluoremetrically using 4-methylumbelliferyl α-L-iduronide

640   (4MU-iduronide) (LC Scientific Inc., Canada) per established assay conditions[4]. Briefly, for

641   IDUA the 4-methylumbelliferyl-iduronide substrate is diluted with sodium formate buffer, 0.4 M,

642   pH 3.5, to 6.6 mM concentration. 25 μL aliquots of substrate are mixed with 25 μL of cell or tissue

643   homogenates and adjusted to a final substrate concentration of 2.5 mM. The mixture is incubated

644   at 37 °C for 60 min, and 200 μL glycine carbonate buffer (pH 10.4) is added to quench the reaction.

645   4-MU (Sigma) is used to make the standard curve. The resulting fluorescence is measured using a

646   SpectraMax M3 plate reader with excitation at 355 nm and emission at 460 nm (Molecular

647   devices).

648

649   **Analysis of glycosaminoglycans**

650   Urine and tissue GAGs were measured with the modified dimethylmethylene blue assay

651   (DMB)[5]. Tissue samples (10 - 30 mg) were incubated for 3 hours at 65 °C in papain digest solution

652   (calcium- and magnesium-free PBS containing 1% papain suspension (Sigma), 5 mM cysteine,

653   and 10 mM EDTA, pH 7.4) to a final concentration of 0.05 mg tissue/mL buffer. 50 µL of extract

654   was incubated with 200 µL DBM reagent (9:1 31 µM DMB stock (in formiate buffer 55 nM): 2

655   M Tris base). The samples were read on a microplate reader at 520 nm.

656

Histology

658   Histology was performed by HistoWiz Inc. (histowiz.com) using standard operating procedures

659   and fully automated workflow. Samples were processed, embedded in paraffin, and sectioned at

660   4µm. Sections were then counterstained with toluidine blue or Alcian blue, dehydrated and film

661   coverslipped using a TissueTek-Prisma and Coverslipper (Sakura). Whole slide scanning (40x)

662   was performed on an Aperio AT2 (Leica Biosystems).

663

Immunocytochemistry

665   MPSI fibroblasts cultures cells were fixed in 4% paraformaldehyde in phosphate-buffered saline

666   (PBS), blocked with 3% bovine serum albumin (BSA) in PBS, and stained with rabbit anti-LAMP1

667   (Abcam) followed by 1:500 dilutions of Alexa 488-conjugated anti-rabbit antibody (Molecular

668   Probes). Mounting and staining of nuclei was done Vectashield with DAPI (Vector labs). Slides

669   were visualized by conventional epifluorescence microcopy using a cooled CCD camera

670   (Hamamatsu) coupled to an inverted Nikon Eclipse Ti microscope. Images were acquired using

671   NIS elements software and analyzed with ImageJ.

672

Computerized Tomography

674   High-resolution Micro-CT scans were acquired at Stanford Center for Innovation in In-Vivo

675   Imaging (SCI[3]) using an eXplore CT 120 scanner (TriFoil imaging). Mice were anesthetized with

676   isoflurane (Baxter Corporation, Mississauga, ON, Canada). The scans were obtained with voxel

677   resolution of 100 μm, an energy level of 80 keV, and 360 degrees of whole mice. Microview

678   software (Parallax innovations) was used for isosurface rendering and measurements. Skull

679   thickness was quantified on Midsagittal images. Femur length was determined by measuring the

680   long axis between the two epiphysis. Zygomatic bone thickness was measures on coronal sections,

681   perpendicular to the axis of the zygoma. Bone lengths were determined using the line measurement

682   tool in MicroView. Femurs were measured from the base of the lateral femoral condyle to the tip

683   of the greater trochanter.

684

Spontaneous locomotor activity

686   All behavioral experimenters were blind to the genotype of the mice throughout testing. All tests

687   were conducted in the light cycle. In all experiments, animals were habituated to the testing room

688   2 h before the tests and were handled by the experimenter for three days before all the behavioral

689   tests. For spontaneous locomotor activity, assessment took place using the open field test in a

690   square arena (76×76 cm^2) with opaque white walls, surrounded with privacy blinds to eliminate

691   external room cues. Mice were placed in the center of the open-field arena and allowed to freely

692   move for 10 min while being tracked by Ethovision (Noldus Information Technology,

693   Wageningen, the Netherlands) automated tracking system. Before each trial, the surface of the

694   arena was cleaned with Virkon disinfectant. For analysis, the arena was divided into a central (53.5

695    $\times$ 53.5 cm^2) and a peripheral zone (11.25-cm wide).

696

Passive Inhibitory Avoidance

698   The passive inhibitory avoidance test was used to assess fear-based learning and memory. We used

699   a dual-compartment system (GEMINI system, San Diego Instruments), where lighted and dark

700   compartments, equipped with grid floor that can deliver electrical shocks, are separated by an

701   automated gate. On day one, each mouse was habituated to the apparatus by placing it into the

702   lighted compartment. After 30 s, the gate opened allowing access to the dark compartment. When

703   the mice entered the dark compartment, the gate closed and the time to cross after the gate opened

704   is recorded (latency time). On day 2 or training day, the mice receive a 0.5 mA shock for 2 s after

705   a 3 s delay after crossing from the lighted to the dark compartment. On day 3, or testing day, after

706   being placed in the lighted compartment for 5 s, the gate opened allowing access to the dark

707   compartment. The latency to enter the dark compartment was recorded. Maximum time to cross

708   was 10 minutes.

709

Marble Burying

711   Repetitive behavior was tested in the marble bury test. Individual mice were introduced into cages

712   containing 20 black glass marbles (1.5 cm diameter, four equidistant rows of five marbles each)

26

713   on top of bedding 5 cm deep. After 30 min under low-light conditions, mice were removed and

714   the number of marbles that were at least half-covered was determined.

715

716   **Mice**

717   NOD.Cg-Prkdc^{scid}IL2rg^{tmlWjl}/Sz (NSG) mice were developed at The Jackson Laboratory[6].

718   Mice were housed in a 12-h dark/light cycle, temperature- and humidity-controlled environment

719   with pressurized individually ventilated caging, sterile bedding, and unlimited access to sterile

720   food and water in the animal barrier facility at Stanford University. All experiments were

721   performed in accordance with National Institutes of Health institutional guidelines and were

722   approved by the University Administrative Panel on Laboratory Animal Care (IACUC 25065).

723

724   **NSG-IDUA$^{X/X}$ mice**

725   We used CRISPR/Cas9 to knock-in the W401X mutation (UniProtKB - Q8BMG0), analogous to

726   the W402X mutation commonly found in patients with severe MPSI, into NSG mouse embryos[7].

727   The guide RNA target sequence was searched using crispr.mit.edu and six shortlisted guides

728   close to the target site were first screened by using an *in vivo* assay in NIH 3T3 cells. Two

729   guides, one each on both sides of the target site, were selected: Guide1 (5'-

730   TTATAGATGGAGAACAACTC-3') cleaves 4 bases upstream and Guide3 (5'-

731   GTTGGACAGCAATCATACAG-3') cleaves 44 bases downstream of the target site. The guides

732   were prepared by in vitro transcription (HiScribe™ T7 High Yield RNA Synthesis Kit, E2040S,

733   New England Biolabs) of a dsDNA template generated by annealing two oligos (with a T7

734   promoter in the sense oligo) followed by a standard PCR reaction. The ssODN donor DNA

735   contained an intended point mutation leading to a STOP codon (TGG to TAG): 5'-

736   ggtgggagctagatattagggtaggaagccagatgctaggtatgagagagccaacagcctcagccctctgcttggcttatagATGGA

737   GAACAA/CTCT**A**GGCAGAGGTCTCAAAGGCTGGGGCTGTGTTGGACAGCAATCATA/

738   CAGTGGGTGTCCTGGCCAGCACCCATCACCCTGAAGGCTCCGCAGCGGCCTGGAGT

739   AC-3' (lower case is intron, upper case is exon, guide cut sites marked by "/" and the mutation in

740   bold).

741   Mouse Zygotes were obtained by mating NSG stud males with super-ovulated NSG females.

742   Female NSG mice 3–4 weeks of age (JAX Laboratories, stock number 005557) were super-

743 ovulated by intraperitoneal injection with 2.5IU pregnant mare serum gonadotropin (National

744 Hormone & Peptide Program, NIDDK), followed 48 hours later by injection of 2.5 IU human

745 chorionic gonadotropin (hCG, National Hormone & Peptide Program, NIDDK). The animals were

746 sacrificed 14 hours following hCG administration and fertilized eggs were collected. CRISPR

747 Injection mixture was prepared by dilution of the components into injection buffer (5 mM Tris,

748 0.1 mM EDTA, pH 7.5) to obtain the following concentrations: 10 ng/μl Cas9 mRNA (Thermo

749 Fisher Scientific, Carlsbad, CA), 10 ng/μl IDUA1F and IDUA3F guide RNA and 10 ng/μl ssODN

750 Donor (Integrated DNA Technologies, Coraville, IA). Zygote injections and embryo transfers

751 were performed using standard protocols described previously[8]. A total of 38 zygotes were

752 injected, the surviving 27 zygotes were transferred, which yielded 7 live offspring. Among these

753 a male homozygous for the mutation was used to establish the NSG-IDUA$^{X/X}$ colony. Mice were

754 genotyped by-PCR based amplification followed by Sanger sequencing using the following

755 primers: GENO F: 5'-CATGGCCCTGTTGGGTGAGTAATGA-3', and GENO R: 5'-

756 TGTGGTACTCCAGGCCGCTG-3'.

757

758 **Statistical analysis**

759 All statistical test including paired and unpaired t-tests, and one-way analysis of variance

760 (ANOVA) followed by Tukey's multiple comparisons test was performed using GraphPad Prism

761 version 7 for Mac OS X, GraphPad Software, La Jolla California USA. Data was reported as means

762 when all conditions passed three normality tests (D'Agostino & Pearson, Shapiro-Wilk, and KS

763 normality test).

764

765 **Acknowledgements**

771 Research Fund, and National Institute of Neurological Disorders and Stroke (NINDS,
772 1K08NS102398-01).

773

Author contributions

775 N.G.-O. conceived the project, collected data, performed experiments, carried out the analyses,
776 and wrote the manuscript; S.G. B. performed macrophage experiments, assisted with mouse
777 studies and figure preparation; N.M. assisted with mouse colony management, transplantation and
778 flow cytometry; R.O.B performed CCR5 guide design and validation; S.M. obtained and purified
779 CD34+-HSPCs from donated cord-blood and assisted with secondary transplants; R.M.Q and
780 C.B.G. designed and generated the NSG-IDUA$^{X/X}$ mice; C.L. and G.B. performed and analyzed
781 off-target studies; L.A. assisted with manuscript preparation and figure design; M.P.H. directed
782 the project, assisted with experimental design, and manuscript preparation.

783

Competing interests

785 M.H.P. is a consultant and has equity interest in CRISPR Tx, but CRISPR Tx had no input or
786 opinions on the subject matter described in this manuscript.

787

Materials and correspondence

789 Materials and correspondence to be addressed to Natalia Gomez-Ospina and Matthew H. Porteus.

790

791

References

1   Wraith, J. E. & Jones, S. Mucopolysaccharidosis type I. *Pediatr Endocrinol Rev* **12 Suppl 1**, 102-106 (2014).

2   Wraith, J. E. *et al.* Enzyme replacement therapy for mucopolysaccharidosis I: a randomized, double-blinded, placebo-controlled, multinational study of recombinant human alpha-L-iduronidase (laronidase). *J Pediatr* **144**, 581-588, doi:10.1016/j.jpeds.2004.01.046 (2004).

3   Mitchell, R. *et al.* Outcomes of haematopoietic stem cell transplantation for inherited metabolic disorders: a report from the Australian and New Zealand Children's Haematology Oncology Group and the Australasian Bone Marrow Transplant Recipient Registry. *Pediatr Transplant* **17**, 582-588, doi:10.1111/petr.12109 (2013).

4   Tanaka, A. *et al.* Long-term efficacy of hematopoietic stem cell transplantation on brain involvement in patients with mucopolysaccharidosis type II: a nationwide survey in Japan. *Mol Genet Metab* **107**, 513-520, doi:10.1016/j.ymgme.2012.09.004 (2012).

5   Wynn, R. F. *et al.* Improved metabolic correction in patients with lysosomal storage disease treated with hematopoietic stem cell transplant compared with enzyme replacement therapy. *J Pediatr* **154**, 609-611, doi:10.1016/j.jpeds.2008.11.005 (2009).

6   Aldenhoven, M. *et al.* Long-term outcome of Hurler syndrome patients after hematopoietic cell transplantation: an international multicenter study. *Blood* **125**, 2164-2172, doi:10.1182/blood-2014-11-608075 (2015).

7   Visigalli, I. *et al.* Gene therapy augments the efficacy of hematopoietic cell transplantation and fully corrects mucopolysaccharidosis type I phenotype in the mouse model. *Blood* **116**, 5130-5139, doi:10.1182/blood-2010-04-278234 (2010).

8   Wang, D. *et al.* Reprogramming erythroid cells for lysosomal enzyme production leads to visceral and CNS cross-correction in mice with Hurler syndrome. *Proc Natl Acad Sci U S A* **106**, 19958-19963, doi:10.1073/pnas.0908528106 (2009).

9   Biffi, A. *et al.* Lentiviral hematopoietic stem cell gene therapy benefits metachromatic leukodystrophy. *Science* **341**, 1233158, doi:10.1126/science.1233158 (2013).

10   McCormack, M. P. & Rabbitts, T. H. Activation of the T-cell oncogene LMO2 after gene therapy for X-linked severe combined immunodeficiency. *N Engl J Med* **350**, 913-922, doi:10.1056/NEJMra032207 (2004).

11   Ranzani, M. *et al.* Lentiviral vector-based insertional mutagenesis identifies genes associated with liver cancer. *Nat Methods* **10**, 155-161, doi:10.1038/nmeth.2331 (2013).

12   Pan, Y. W., Scarlett, J. M., Luoh, T. T. & Kurre, P. Prolonged adherence of human immunodeficiency virus-derived vector particles to hematopoietic target cells leads to secondary transduction in vitro and in vivo. *J Virol* **81**, 639-649, doi:10.1128/JVI.01089-06 (2007).

13   Persons, D. A., Hargrove, P. W., Allay, E. R., Hanawa, H. & Nienhuis, A. W. The degree of phenotypic correction of murine beta -thalassemia intermedia following lentiviral-mediated transfer of a human gamma-globin gene is influenced by chromosomal position effects and vector copy number. *Blood* **101**, 2175-2183, doi:10.1182/blood-2002-07-2211 (2003).

14   Bak, R. O., Gomez-Ospina, N. & Porteus, M. H. Gene Editing on Center Stage. *Trends in Genetics* **34**, 11, doi:10.1016/j.tig.2018.05.004 (2018).

15   Dever, D. P. *et al.* CRISPR/Cas9 beta-globin gene targeting in human haematopoietic stem cells. *Nature* **539**, 384-389, doi:10.1038/nature20134 (2016).

16   Liu, R. *et al.* Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* **86**, 367-377 (1996).

17   Bak, R. O., Dever, D. P. & Porteus, M. H. CRISPR/Cas9 genome editing in human hematopoietic stem cells. *Nat Protoc* **13**, 358-376, doi:10.1038/nprot.2017.143 (2018).

18   Hendel, A. *et al.* Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat Biotechnol* **33**, 985-989, doi:10.1038/nbt.3290 (2015).

19   Clarke, L. A. *et al.* Murine mucopolysaccharidosis type I: targeted disruption of the murine alpha-L-iduronidase gene. *Hum Mol Genet* **6**, 503-511 (1997).

20   Wang, D. *et al.* Characterization of an MPS I-H knock-in mouse that carries a nonsense mutation analogous to the human IDUA-W402X mutation. *Mol Genet Metab* **99**, 62-71, doi:10.1016/j.ymgme.2009.08.002 (2010).

21   Mendez, D. C. *et al.* A novel, long-lived, and highly engraftable immunodeficient mouse model of mucopolysaccharidosis type I. *Mol Ther Methods Clin Dev* **2**, 14068, doi:10.1038/mtm.2014.68 (2015).

22   Lombardo, A. *et al.* Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nat Methods* **8**, 861-869, doi:10.1038/nmeth.1674 (2011).

23   Hu, Z., Van Rooijen, N. & Yang, Y. G. Macrophages prevent human red blood cell reconstitution in immunodeficient mice. *Blood* **118**, 5938-5946, doi:10.1182/blood-2010-11-321414 (2011).

24   Cradick, T. J., Qiu, P., Lee, C. M., Fine, E. J. & Bao, G. COSMID: A Web-based Tool for Identifying and Validating CRISPR/Cas Off-target Sites. *Mol Ther Nucleic Acids* **3**, e214, doi:10.1038/mtna.2014.64 (2014).

25   Vakulskas, C. A. *et al.* A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nat Med* **24**, 1216-1224, doi:10.1038/s41591-018-0137-0 (2018).

26   Aldenhoven, M. *et al.* Hematopoietic cell transplantation for mucopolysaccharidosis patients is safe and effective: results after implementation of international guidelines. *Biol Blood Marrow Transplant* **21**, 1106-1109, doi:10.1016/j.bbmt.2015.02.011 (2015).

27   Sadelain, M., Papapetrou, E. P. & Bushman, F. D. Safe harbours for the integration of new DNA in the human genome. *Nat Rev Cancer* **12**, 51-58, doi:10.1038/nrc3179 (2012).

28   Wang, D. *et al.* Engineering a lysosomal enzyme with a derivative of receptor-binding domain of apoE enables delivery across the blood-brain barrier. *Proc Natl Acad Sci U S A* **110**, 2999-3004, doi:10.1073/pnas.1222742110 (2013).

29   Sharma, R. *et al.* In vivo genome editing of the albumin locus as a platform for protein replacement therapy. *Blood* **126**, 1777-1784, doi:10.1182/blood-2014-12-615492 (2015).

30   Heyer, W. D., Ehmsen, K. T. & Liu, J. Regulation of homologous recombination in eukaryotes. *Annu Rev Genet* **44**, 113-139, doi:10.1146/annurev-genet-051710-150955 (2010).

31   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, doi:10.1038/srep21264 (2016).

32   van Galen, P. *et al.* The unfolded protein response governs integrity of the haematopoietic stem-cell pool during stress. *Nature* **510**, 268-272, doi:10.1038/nature13228 (2014).

33   Fares, I. *et al.* Cord blood expansion. Pyrimidoindole derivatives are agonists of human hematopoietic stem cell self-renewal. *Science* **345**, 1509-1512, doi:10.1126/science.1256337 (2014).

34   Boitano, A. E. *et al.* Aryl hydrocarbon receptor antagonists promote the expansion of human hematopoietic stem cells. *Science* **329**, 1345-1348, doi:10.1126/science.1191536 (2010).

35   Oussoren, E. *et al.* Residual alpha-L-iduronidase activity in fibroblasts of mild to severe Mucopolysaccharidosis type I patients. *Mol Genet Metab* **109**, 377-381, doi:10.1016/j.ymgme.2013.05.016 (2013).

36   Elliott, S. *et al.* Pilot study of newborn screening for six lysosomal storage diseases using Tandem Mass Spectrometry. *Mol Genet Metab* **118**, 304-309, doi:10.1016/j.ymgme.2016.05.015 (2016).
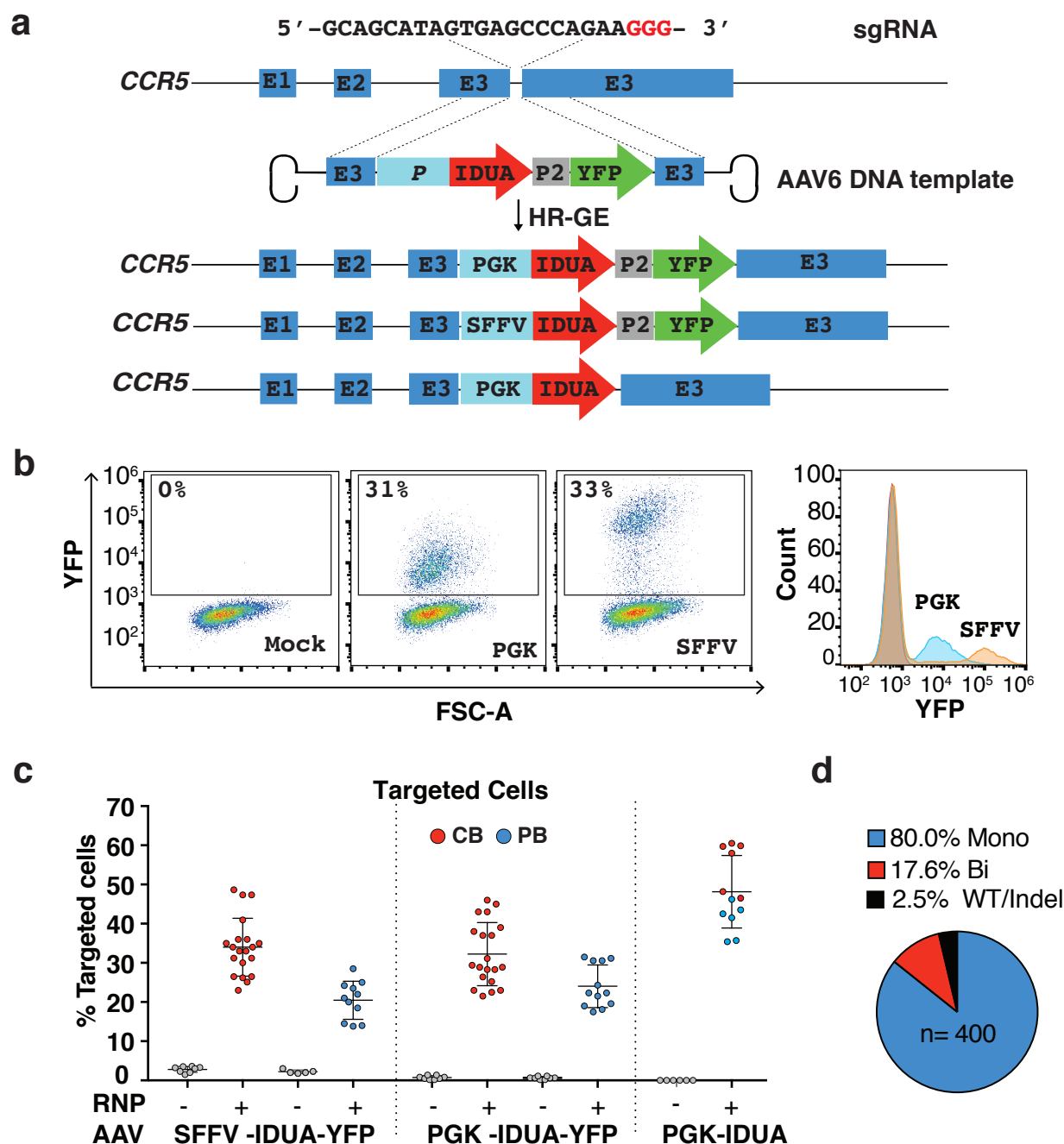
Fig. 1 I Efficient CRIPR/Cas9-mediated integration of IDUA overexpression cassettes into the *CCR5* locus in human CD34+ HSPCs. a, Schematic of targeted integration of IDUA and expression cassettes. The AAV6 genome was constructed to have 500bp arms of homology centered on the cut site, and the IDUA sequence placed under the control of the SFFV or the PGK promoter. In two DNA templates, YFP was expressed downstream of IDUA using the self-cleaving P2A peptide. Analysis was performed 3-days post-modification. **b,** FACs and histogram plots of mock and human HSPCs that underwent RNP and AAV6 exposure with YFP-containing expression cassettes. **c,** Targeting frequencies in CB (red) and PB (blue)-derived HSPCs read by percent fluorescent cells in YFP expressing cassettes and percent colonies with targeted CCR5 alleles by single cell-derived colony genotyping in cassettes without the reporter. Each dot represents the average of duplicates for a single human cell donor. For RNP+AAV6 conditions with YFP templates,CB=20, PB=11. For the template without selection CB=6, PB=6. Lines indicate mean and SD. **d,** Distribution of wildtype (WT), mono and bi-allelically modified cells (n=400) in YFP-positive HSPCs.
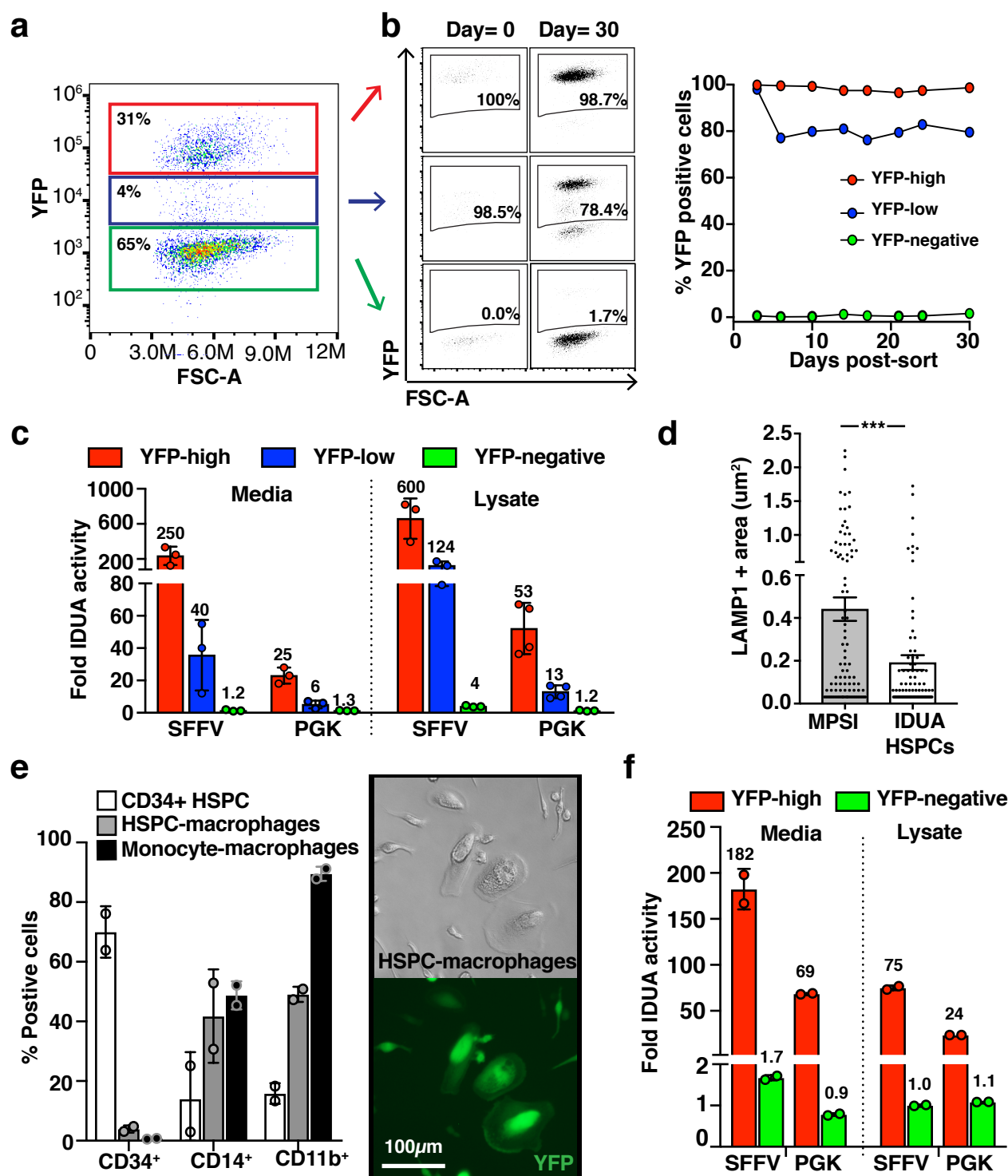
Gomez-Ospina et al, Figure 1

Fig. 2 l Enhanced IDUA expression by IDUA-HSPCs and derived macrophages. a, FACS plot shows distinct populations based on YFP expression 3 days post-modification. **b,** Persistent YFP expression up to 30 days cultures. **c,** Fold increase in IDUA secretion and intracellular expression by YFP-high, YFP-low, and YFP-negative populations compared to mock cells. **d,** Average LAMP-1+ area in MPSI fibroblasts co-cultured with IDUA-HSPCs. Each dot represents a cell. **e,** Human CD34, CD14, and CD11b marker expression in HSPC-derived macrophages and human monocyte-derived macrophages after *in vitro* differentiation compared to undifferentiated cells (CD34+ HSPCs). Macrophage morphology and YFP expression after differentiation. **f,** Fold increase in IDUA secretion and intracellular expression in HSPC-macrophages modified with SFFV and PGK expression cassettes. **c, e, and f,** Each dot represents average of triplicates in a human cell donor. All data expressed as mean ± SD, *** p < .001 in two-sided unpaired t-test.
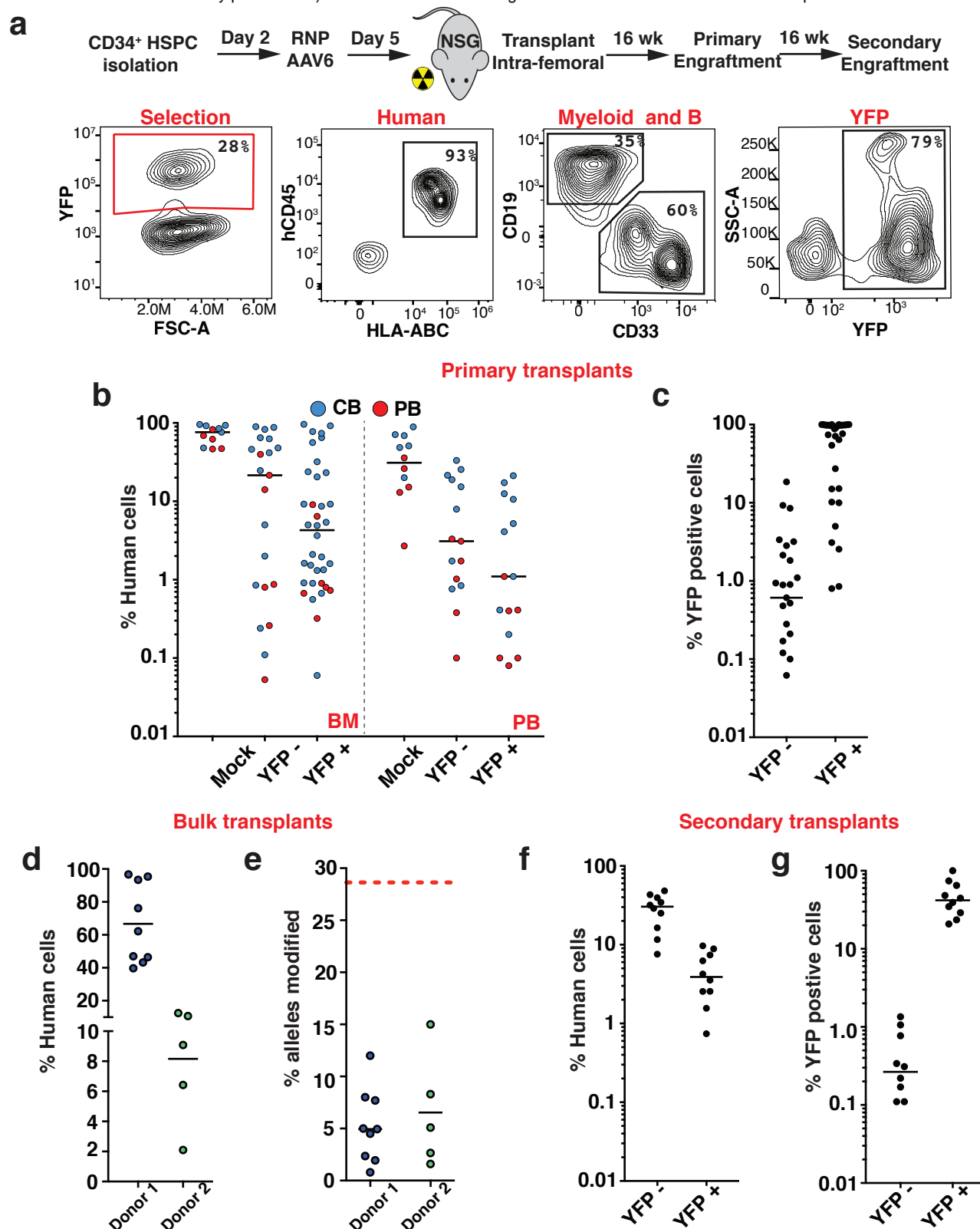
Gomez-Ospina et al, Figure 2

Fig. 3 | IDUA-HSPCs maintain long-term repopulation capacity. a, Schematic and representative FACS plots showing phenotyping by flow of human, myeloid, B-cell, and targeted cells after engraftment. **b,** Percent human cell chimerism in bone marrow (BM) and peripheral blood (PM) in mice 16-weeks post-transplant with CB (blue) and PB (red)-derived HSPCs targeted with PGK cassettes; mock (n=11), YFP- (n=21), and YFP+ (n=36). Each point represents a mouse. **c,** Percent human YFP+ cells in BM of mice in BM 16-weeks post-transplant. **d,** Percent human cell chimerism in BM in mice transplanted with bulk cells without selection from two different human cell donors. **e,** Percent modified alleles in engrafted cells by ddPCR. 28% was the starting allele modification frequency. **f,** Percent human cell chimerism in BM of mice in secondary transplants 32 weeks after genome editing; YFP- (n=10), and YFP+ (n=10). **g,** Percent human, YFP+ cells in BM of mice in secondary transplants.
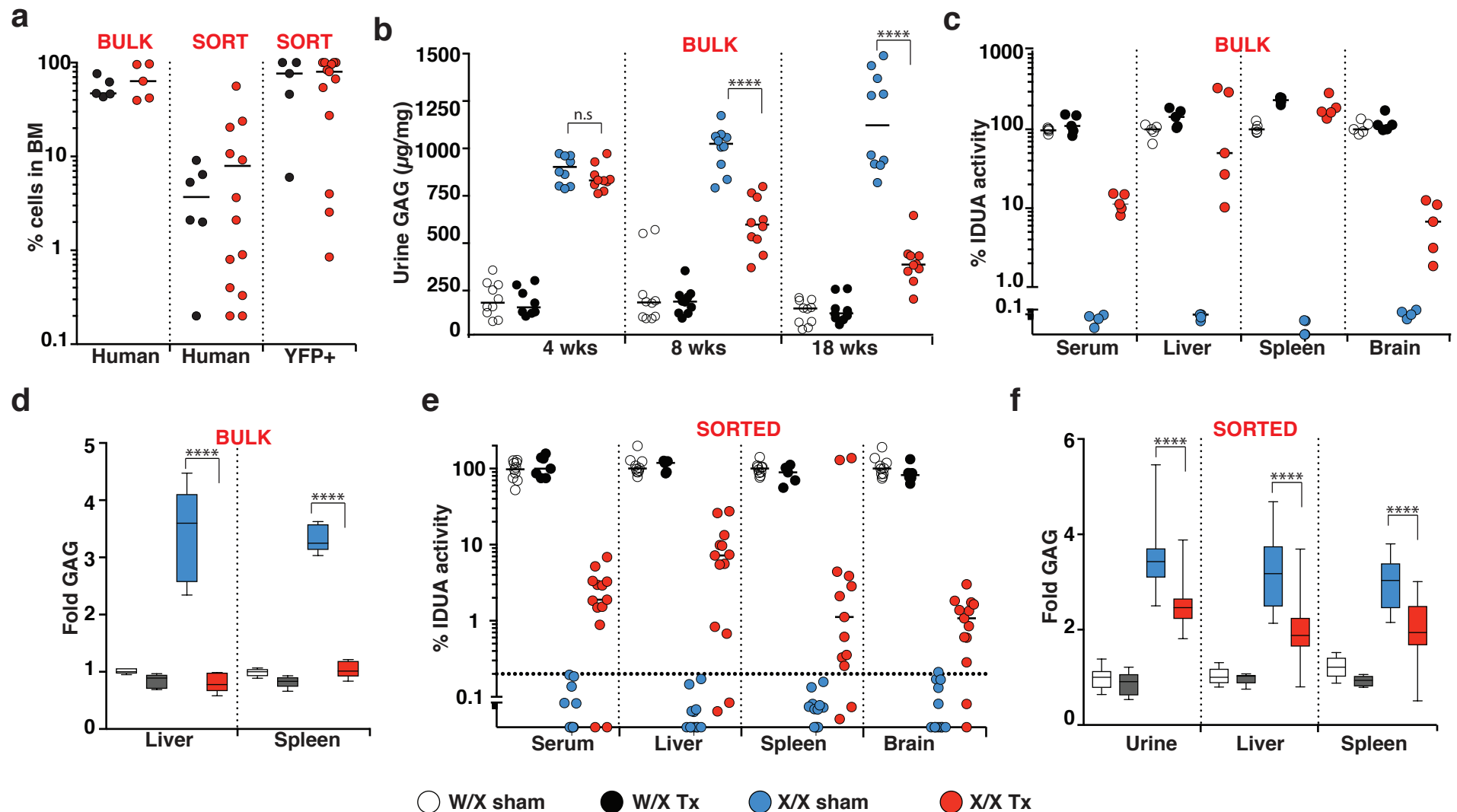
Gomez-Ospina et al, Figure 3

Fig. 4 | Biochemical correction in NSG-IDUAX/X mice by human IDUA-HSPCs. IDUA activity and GAG accumulation in heterozygous sham-treated (W/X sham- clear), heterozygous transplanted (W/X Tx- black), homozygous sham-treated (X/X sham- blue), and homozygous transplanted (X/X Tx- red) mice. a, Percent human and YFP+ cells in BM in experiments using bulk and sorted cells. **b,** Urinary GAGs at 4,8, and 18 weeks in experiments using bulk cells (n=5 mice per cohort, two measurements per mouse). **c,** Serum and tissue IDUA activity in experiments using bulk cells (n=5 per cohort). **d,** Fold GAG storage in liver and spleen (normalized by W/X sham, n=5 per cohort). **e,** Serum and tissue IDUA activity in experiments using sorted cells (n=5 for W/X Tx and sham mice, and n=13 for X/X Tx and sham mice). **f,** Fold GAG urinary excretion and tissue storage in experiments using sorted cells (normalized by W/X sham). Median values in all scatter plots. d and f show box plots with whiskers at the 5-95th percentiles. ****: p < .0001 in one-way ANOVA test. Post hoc comparisons were made with the Tukey's multiple comparisons test.
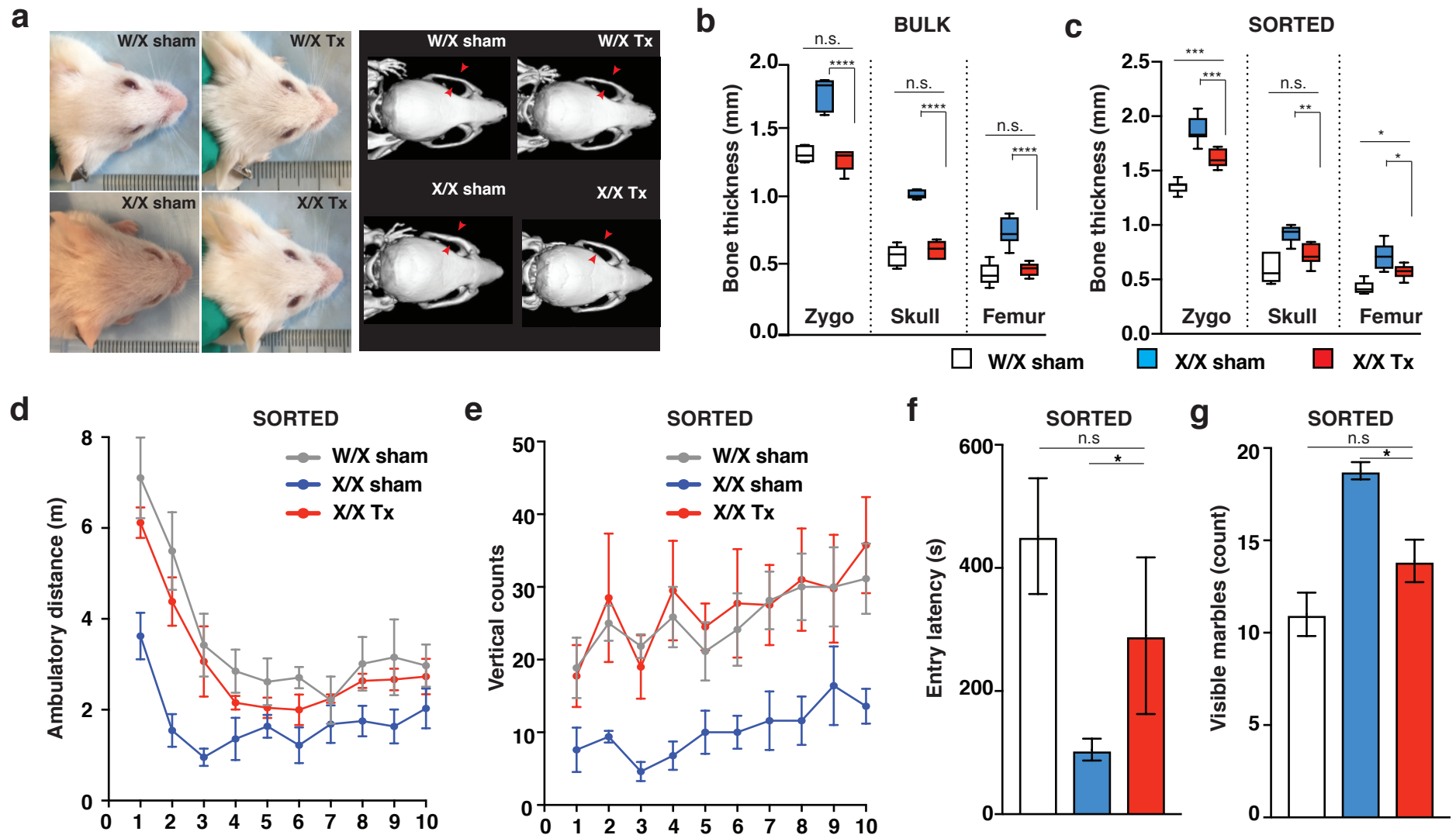
Gomez-Ospina et al, Figure 4

Fig. 5 | Phenotypic restitution in NSG-IDUAX/X mice by human IDUA-HSPCs. Behavioral and skeletal assessment in: W/X sham (clear or gray, n=11), X/X sham (blue, n=10), and X/X Tx (red, n=11). **a,** Representative photos showing facial features in mice transplanted with bulk cells. **b,** Bony features in mice transplanted with bulk and **c,** sorted cells. Box plots with whiskers show min and max. **d,** Ambulatory distance in mice transplanted with sorted cells. W/X sham vs. X/X sham: **; W/X sham vs. X/X Tx: n.s.; X/X sham vs. X/X sham: *. **e,** Vertical rearing in mice transplanted with sorted cells. W/X sham vs. X/X sham: *; W/X sham vs. X/X Tx: n.s.; X/X sham vs. X/X sham: *. **f,** Memory retention in mice transplanted with sorted cells. **g,** Quantification of digging behavior in mice transplanted with sorted cells. Data shown as mean ± SEM. **b-g,** Comparisons between groups were performed using one-way ANOVA test and post-hoc comparisons were made with the Tukey's multiple comparisons test. *: $p < .05$, **: $p < .01$, ***: $p < .001$, and ****: $p < .0001$. Open field testing and vertical rearings were analyzed using within-subject modeling by calculating the are under the curve for each mouse within the first five minutes and comparing between groups with one-way ANOVA.
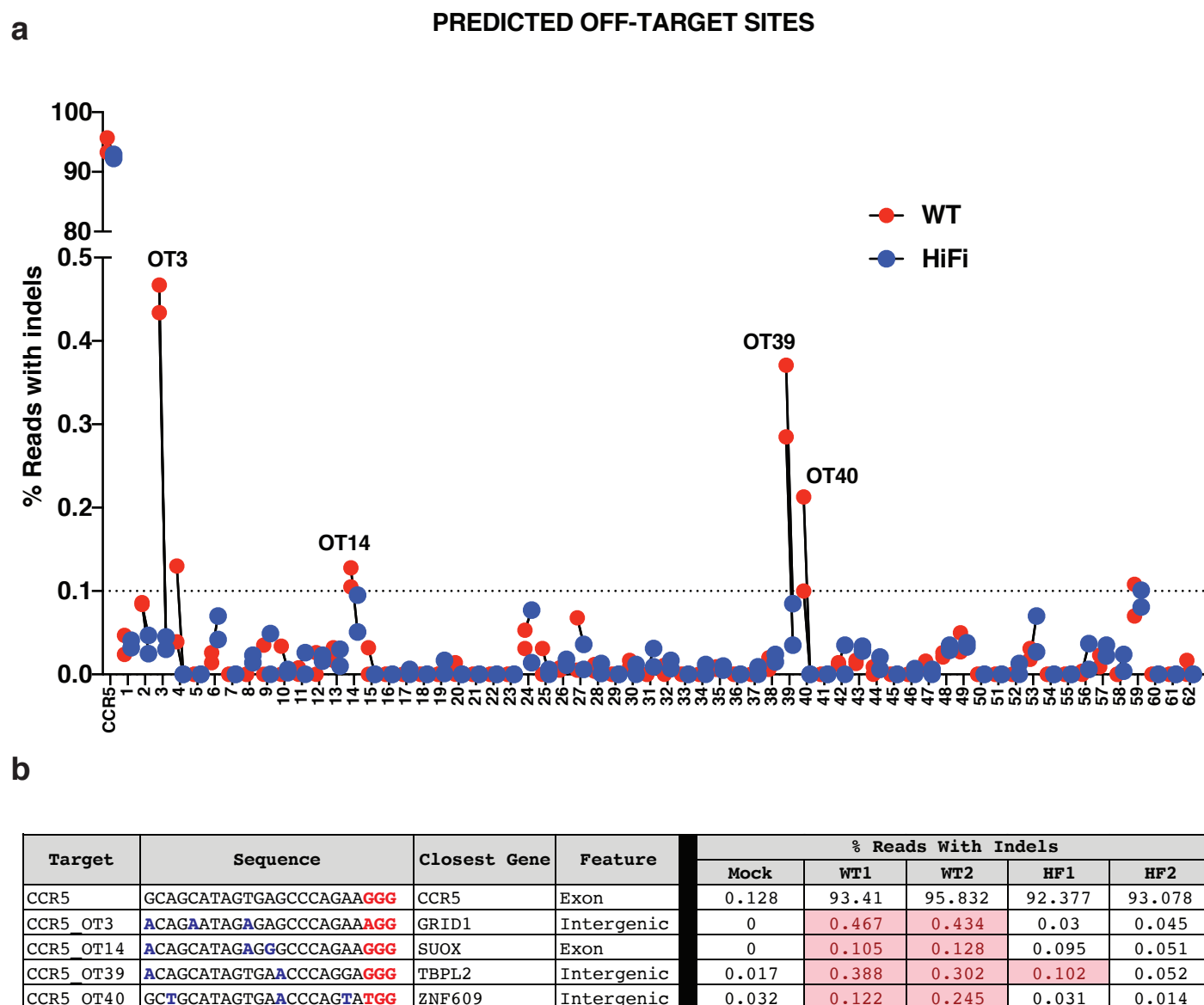
Gomez-Ospina et al, Figure 5

a

PREDICTED OFF-TARGET SITES

b

Target	Sequence	Closest Gene	Feature	% Reads With Indels				
				Mock	WT1	WT2	HF1	HF2
CCR5	GCAGCATAGTGAGCCCAGAAGGG	CCR5	Exon	0.128	93.41	95.832	92.377	93.078
CCR5_OT3	ACAGAATAGAGAGCCCAGAAAGG	GRID1	Intergenic	0	0.467	0.434	0.03	0.045
CCR5_OT14	ACAGCATAGAGGGCCCAGAAGGG	SUOX	Exon	0	0.105	0.128	0.095	0.051
CCR5_OT39	ACAGCATAGTGAACCCAGGAGGG	TBPL2	Intergenic	0.017	0.388	0.302	0.102	0.052
CCR5_OT40	GCTGCATAGTGAACCCAGTATGG	ZNF609	Intergenic	0.032	0.122	0.245	0.031	0.014

Fig. 6 | OFF-target analysis of the CCR5 sgRNA. a, Percent reads with Indels at 62 off-target sites predicted using COSMID. For each site, red dots indicate samples treated with WT Cas9 and blue dots indicate samples treated with HiFi Cas9. The limit of detection for NGS is 0.1% and is indicated on the graph by a dashed line. **b,** Table summarizing four bona fide off-target sites. PAM sequences are shown in red and mismatched bases are shown in blue. For all of these sites the percent of Indels was < 0.5%. For all of these sites, the use of the HiFi Cas9 abolished off-target activity.

Gomez-Ospina et al, Figure 6
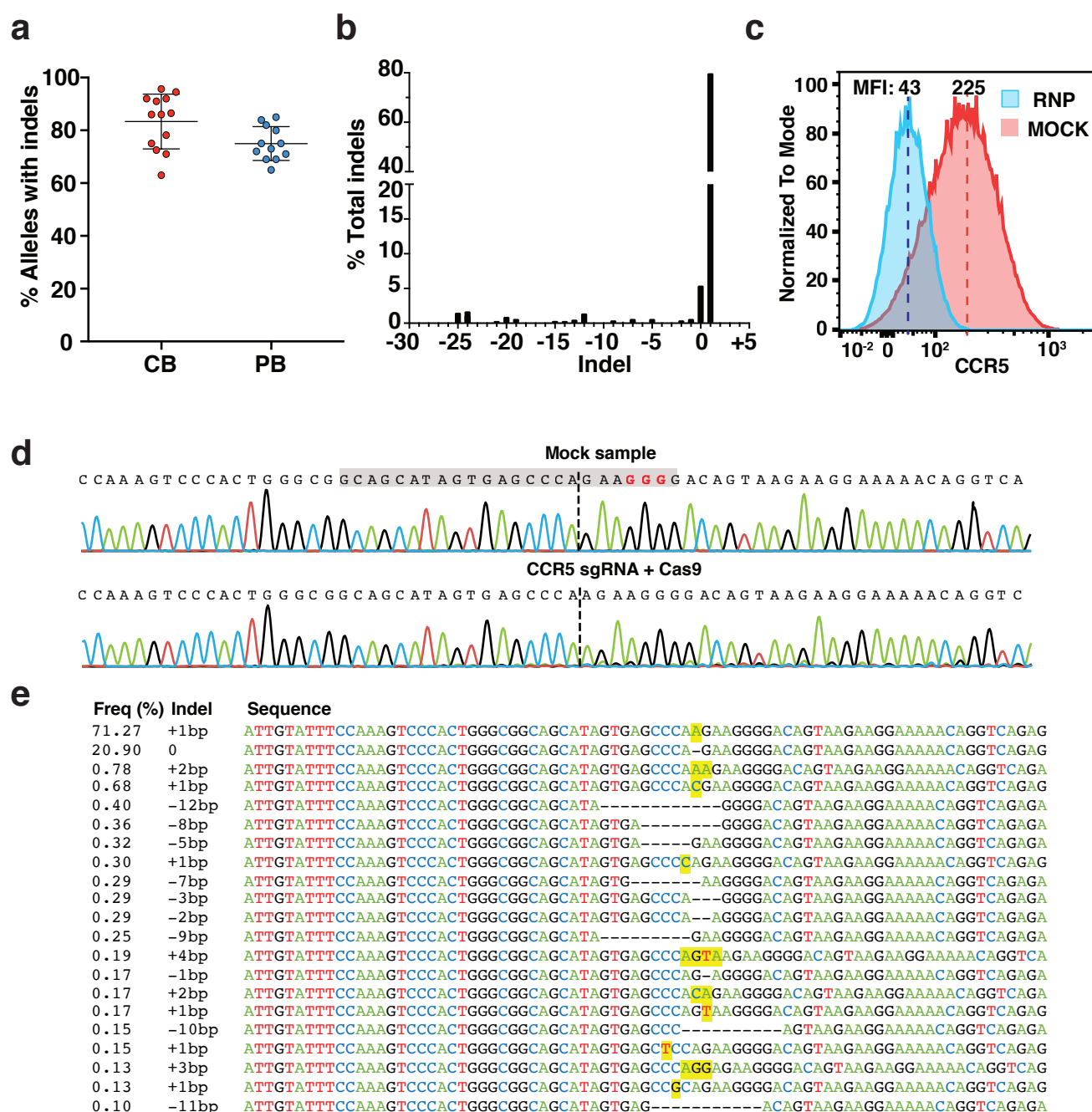
Extended Data Fig. 1 | Characterization of the *CCR5* sgRNA. a, Indel frequency in CB and PB-derived cells by the RNP complex. **b,** Representative indel distribution from next generation sequencing reads. **c,** Histogram of CCR5 protein expression in mock-treated and RNP-treated cells showing an 80% reduction in protein expression after indel induction. **d,** Sample sequence traces around the *CCR5* sgRNA sequence (gray box, PAM in red) in mock samples and RNP-treated CB-derived HSPCs showing predominant single A insertion. **e,** Summary of indels with frequencies greater than 0.1%.

Gomez-Ospina et al, Extended Data Figure 1
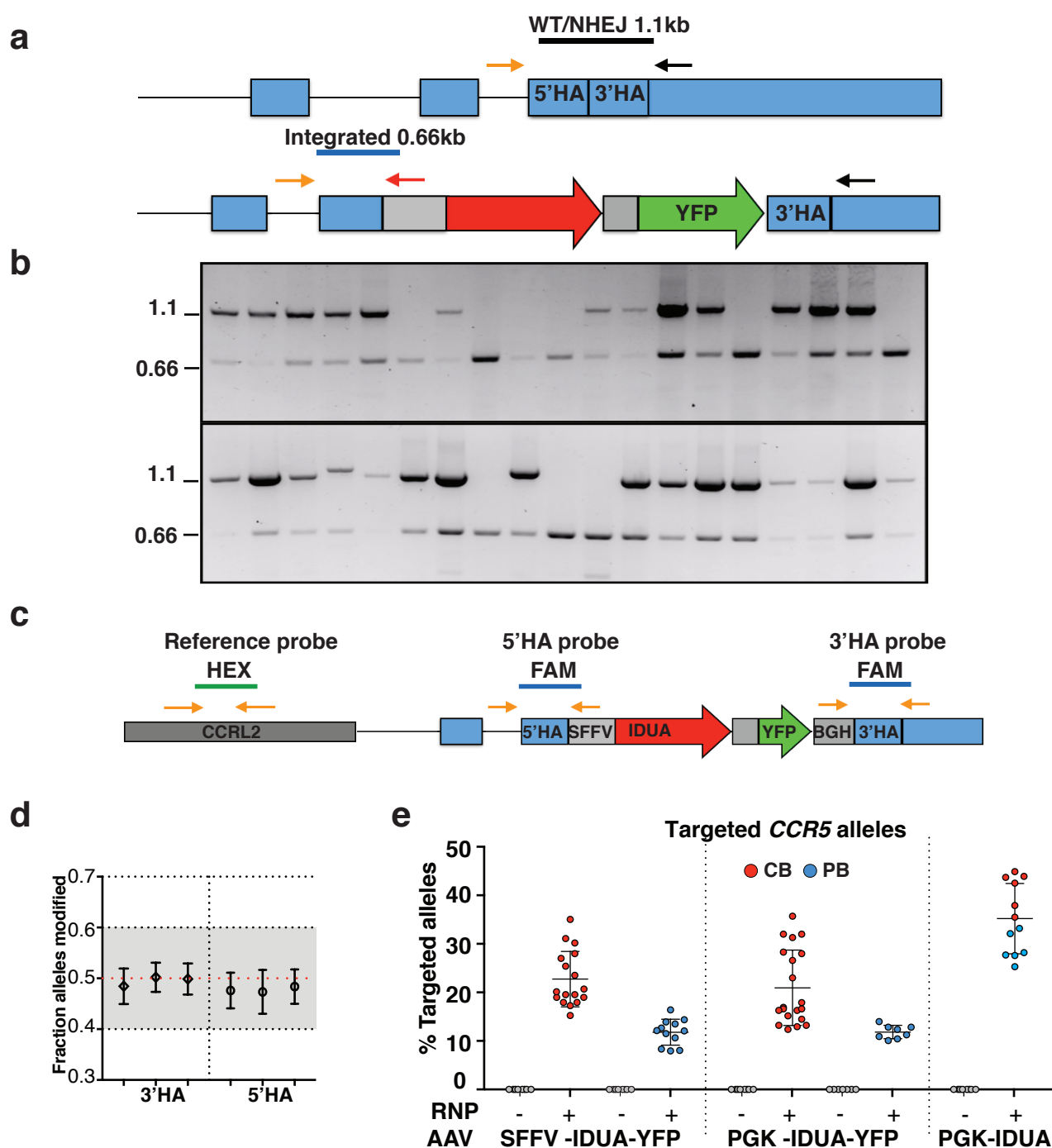
Extended Data Fig. 2 l Efficiency of modification at the *CCR5* locus. a, Schematic showing the three primer-based genotyping scheme to distinguish mono and bi-allelic integration into the CCR5 locus on CFA-derived colonies. This strategy did not distinguish WT versus alleles with indels (NHEJ). **b,** Example agarose gels of 40 colonies genotyped in this manner. A single 1.1Kb band was interpreted as WT/NHEJ in both alleles, while a single 0.6 Kb band was read as bi-allelic integration. **c,** Schematic of probe design for ddPCR analysis. Fraction of modified alleles was obtained by using a second reference probe to the CCRL2 gene also on chromosome 3p. **d,** Originally two probes where each straddled the 5' or 3' homology arm were designed. The accuracy of the assays was verified and compared using genomic DNA from colonies derived from mono-allelic cells (0.5 fraction of alleles modified). Error bars indicate 95% CI. The 3' HA probe was selected. **e,** CCR5 allele targeting frequencies in CB (red) and PB (blue)-derived IDUA expressing HSPCs as measured by ddPCR.

Gomez-Ospina et al, Extended Data Figure 3

a

b

Extended Data Fig. 3 I a, Gating scheme for quantification of human CD34+, CD14+, and CD11b+ cells in human HSPCs maintained for 2 weeks in standard CD34+ cytokine media (top panel) or media with M-CSF to induce macrophage differentiation (bottom panel). Single and live cell discrimination not shown.
b, Gating scheme used to analyze human cell engraftment and cell lineages after transplantation. Representative plots for quantification of mouse and human hematopoietic (mCD45+ and hCD45+), all human (CD45+/HLA-ABC+), human B (CD19+), human myeloid (CD33+), human T (CD3+), and YFP+ cells.

Gomez-Ospina et al, Extended Data Figure 3

Extended Data Fig. 4 I IDUA-HSPCs modified with SFFV containing cassettes are capable of long-term repopulation and multi-lineage differentiation. a, Percent human cell chimerism in BM and PM of mice 16-weeks post-transplant with CB-derived HPSCs. mock (n=11), YFP- (n=19), and YFP+ (n=25). Each point represents a mouse.**b,** Percent YFP+ cells in BM and PB of mice in primary transplants. **c,** Percent human cell chimerism in BM of mice in secondary transplants (32 weeks). **d,** Percent YFP+ cells in BM and PB of mice in secondary transplants. Medians shown.

Gomez-Ospina et al, Extended Data Figure 4
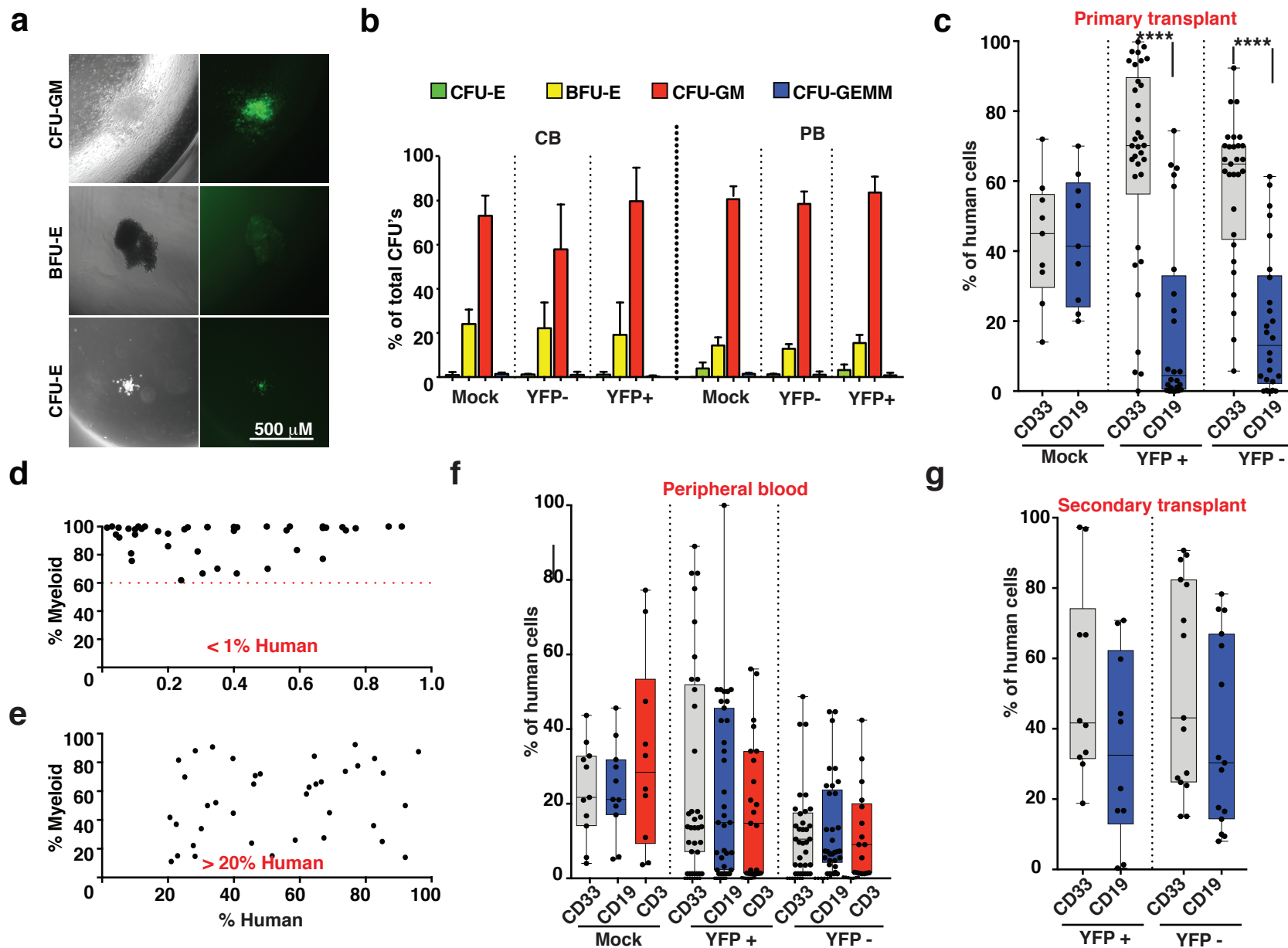
Extended Data Fig. 5 | | IDUA-HSPCs maintain multi-lineage differentiation potential. **a,** Representative photos showing morphology and YFP expression in CFU-GM, BFU-E, and CFU-E colonies in CFAs. **b,** Colony formation unit frequency in mock, YFP- and YFP+ cells. **c,** Box plot with whiskers (min to max) showing percent human CD33+ (myeloid), CD19+ (B) cell in the BM of mice transplanted with mock, and FAC-sorted YFP+ and YFP- cells in primary transplant mice. Each point represents data from a single mouse. **d,** Scatter plot from mice with human cell chimerism <1% against the percent myeloid cells. **e,** Scatter plot of human cell chimerism >20% against the percent myeloid cells. **f,** Percent human CD33+, CD19+, and CD3+ (T) cells in the PB of mice transplanted with mock, and FAC-sorted YFP+ and YFP- cells. **g,** Box plot with whiskers (min to max) showing percent human CD33+ (myeloid), CD19+ (B) cell in the BM secondary transplant mice. ****: $p < .0001$ -in one-way ANOVA test. Post hoc comparisons were made with the Tukey's multiple comparisons test.

Gomez-Ospina et al, Extended data Fgure 5
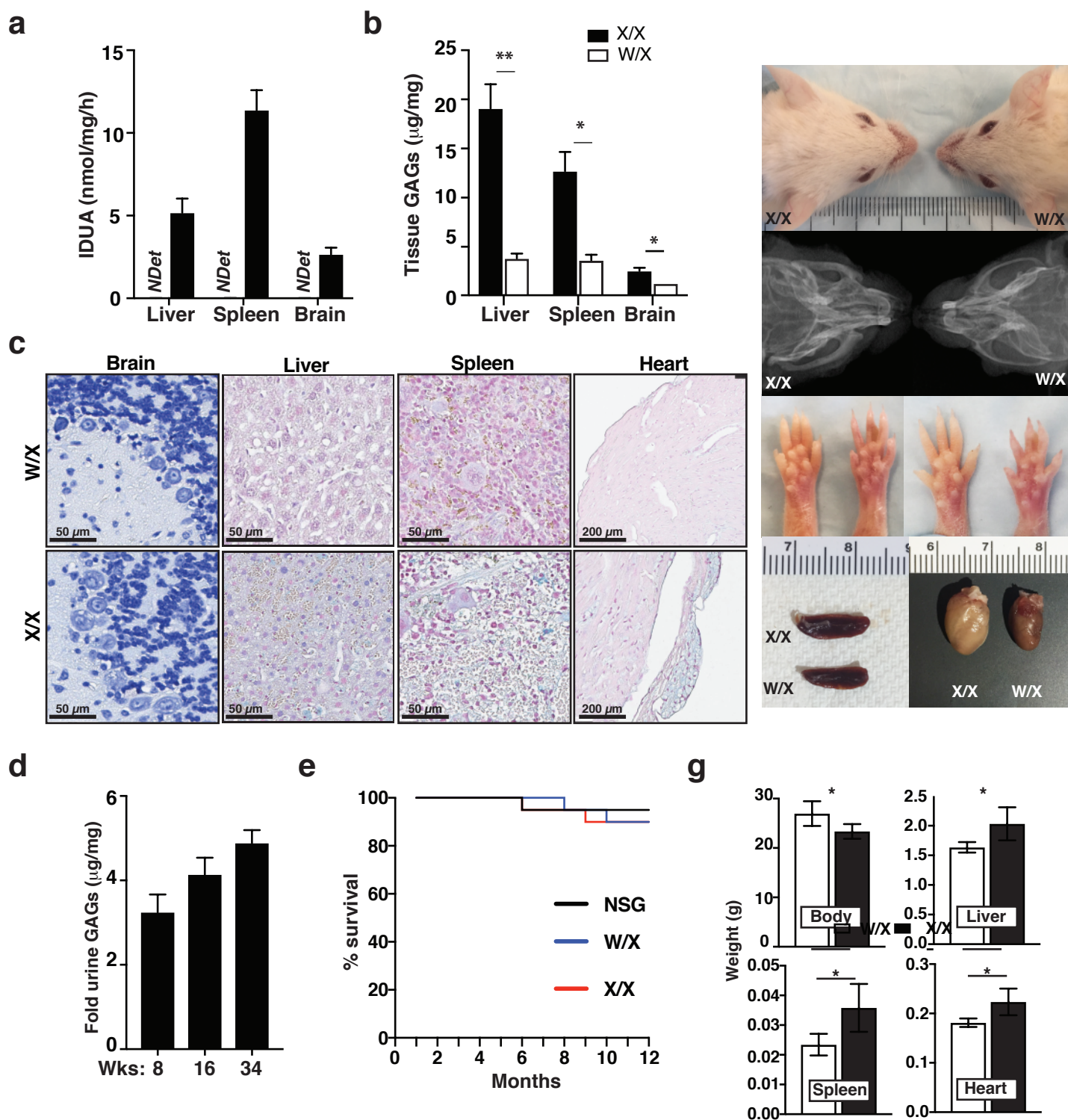
Extended Data Fig. 6 | Biochemical characterization of NSG-IDUAX/X mice. a, Tissue IDUA enzyme activity in tissues. **b,** Tissue GAGs measure by dimethylmethylene blue reactivity. **c,** Histological sections of paraffin-embedded tissues stained with bromophenol blue (brain) or alcian blue (liver, spleen, and heart). Brain sections on X/X mice showed distended and vacuolated Purkinje cells. In liver, spleen, and heart blue deposit-laden cells can be seen throughout. **d,** Age-related progression of urinary GAGs excretion. **e,** Survival analysis comparing NSG, W/X, and X/X during one year of observation (n=10). **f,** Physical dysmorphisms and visceral enlargement. **g,** Total body, liver, spleen and heart weight in W/X and X/X mice

Gomez-Ospina et al, Extended Data Figure 6
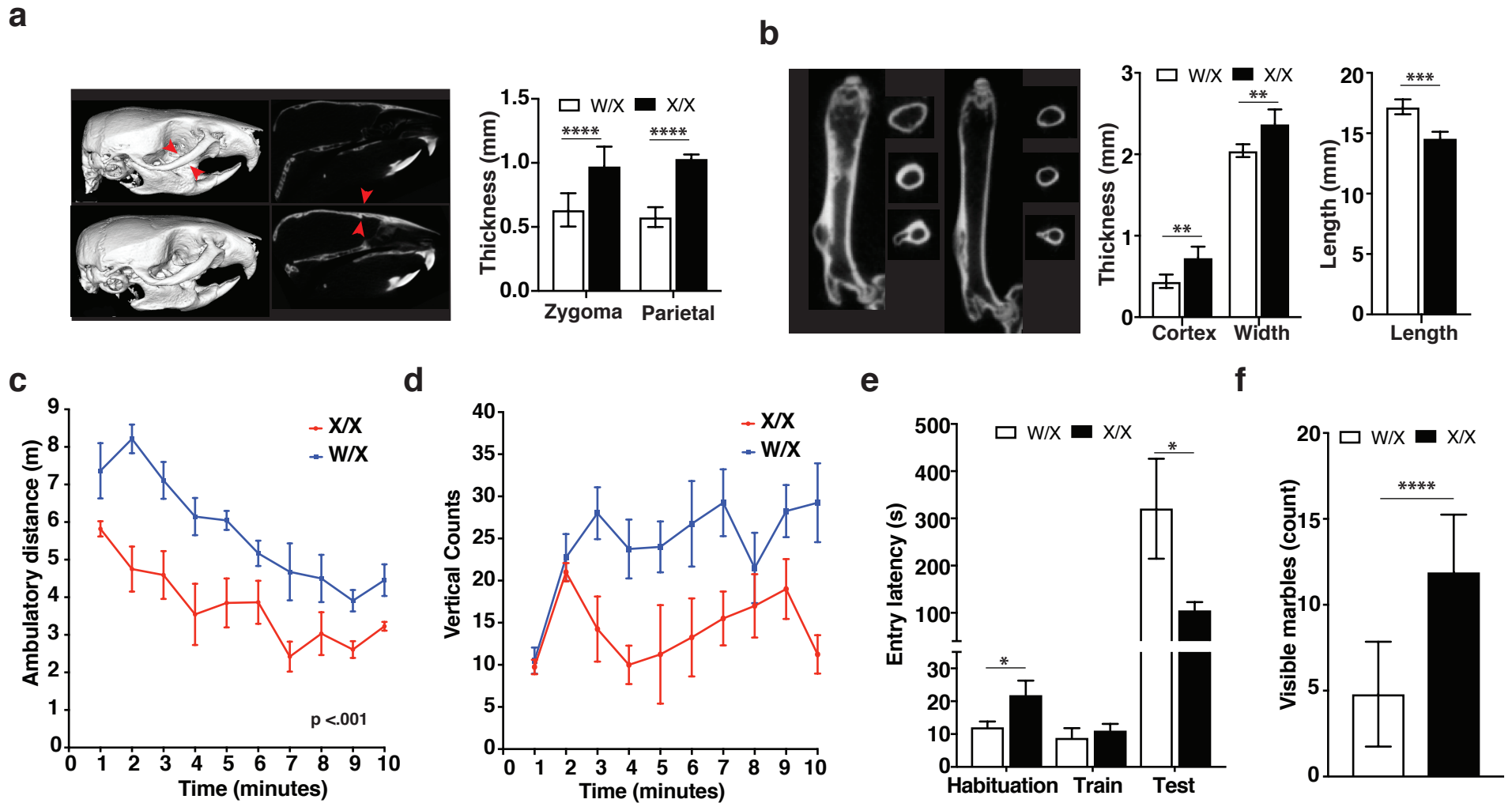
Extended Data Fig. 7 | Phenotypic characterization of NSG-IDUAX/X mice. a, Reconstructed micro-CT images of skull and zygomatic and parietal bone thickness. **b,** CT longitudinal sections of femurs, and cortical thickness, width, and length measurements. **c,** Spontaneous locomotion in open field testing. **d,** Vertical rearing counts for W/X and X/X mice during 10-minute observation in the open field chamber. **e,** Long-term memory in passive inhibitory avoidance test. **f,** Defensive digging in the marble burying task. Total 5 female mice per genotype. Data are presented as mean ± SD for a-e, mean ± SEM for f-g. Comparisons between groups were performed using unpaired t-test. *: $p < .05$, **: $p < .01$, ***: $p < .001$, and ****: $p < .0001$. Open field testing was analyzed using within-subject modeling for the entire time course by calculating the are under the curve for each mouse and comparing between genotypes with a t-test.

Gomez-Ospina et al, Extended Data Figure 7
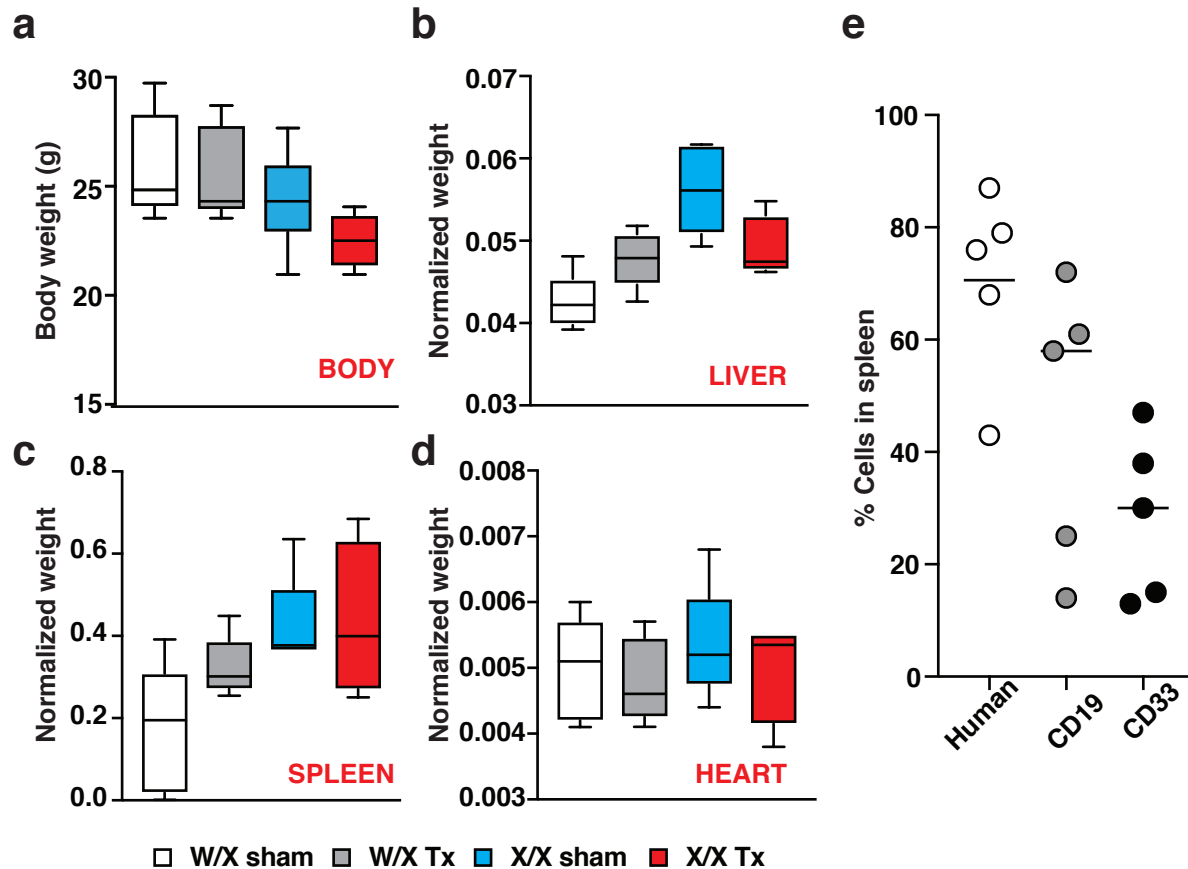
Extended Data Fig. 8 | Body and organ size in transplantation experiments using bulk IDUA-HSPCs. **a,**Total body weight in heterozygous transplanted (W/X Tx- dark gray), heterozygous sham-treated (W/X sham- clear), homozygous transplanted (X/X Tx- red), and homozygous sham-treated (X/X Tx- blue) mice. **b**, Normalized liver weight. **c**, Normalized spleen weight. **d,** Normalized heart weight. **e**, Percent human, B (CD19+), and myeloid (CD33+) cells in the spleen of X/X Tx mice measured 18 weeks post-transplant.

Gomez-Ospina et al, Extended Data Figure 8
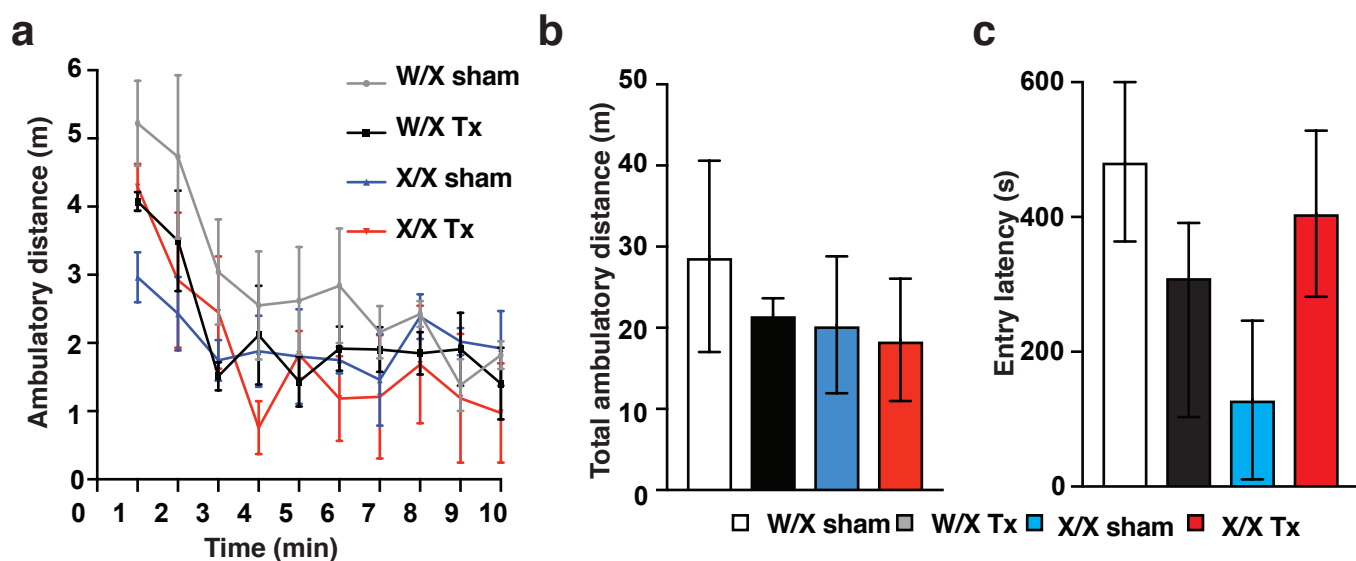
Extended Data Fig. 9 | Neurobehavioral studies in mice transplanted with bulk IDUA-HSPCs. **a,** 10-minute time course of spontaneous locomotor behavior in heterozygous transplanted (W/X Tx-black), heterozygous sham-treated (W/X sham- celar or gray), homozygous transplanted (X/X Tx- red), and homozygous sham-treated (X/X Tx- blue) mice. No comparison was found to be significant. **b,** Total ambulatory distance in 10 minutes. **c,** Memory for inhibitory avoidance training (24h). No comparison was found to be significant. Comparisons between groups were performed using one-way ANOVA test and post-hoc comparisons were made with the Tukey's multiple comparisons test. Open field testing and vertical rearings were analyzed using within-subject modeling by calculating the are under the curve for each mouse for the entire time course and comparing between groups with one-way ANOVA.

Gomez-Ospina et al, Extended Data Figure 9

| Target | Sequence | Closest Gene | Feature | Experiment 1 and 2 % Reads With Indels | | | | | Experiment 1 and 2 # Reads Aligned Total | | | | |
				Mock	WT1	WT2	HF1	HF2	sg11_WT	WT	sg11_HF	HF	Mock
CCR5	GCAGCATAGTGAGCCCAGAAGGG	CCR5	exon	0.128	93.41	95.832	92.377	93.078	8999	8950	11242	7556	3902
CCR5_OT1	AGAGAAGAGTGAGCCCAGAAGGG	FEM1B	intron	0	0.024	0.047	0.041	0.032	33413	17051	31452	18919	2343
CCR5_OT2	ACAGAATAGAGAGCCCAGAAAGG	LOC100129138	Intergenic	0	0.084	0.086	0.047	0.025	25060	19724	25506	15774	1140
CCR5_OT3	ACAGAATAGAGAGCCCAGAAAGG	GRID1	Intergenic	0	0.467	0.434	0.03	0.045	36179	24425	33838	24617	1498
CCR5_OT4	ACAGAATAGAGAGCCCAGAAAGG	LINC00374	intron	0	0.039	0.13	0	0	5083	9198	1145	4421	144
CCR5_OT5	GCAGGAATGTGAGCCCAGAAAGG	SRGAP3	intron	0.058	0.015	0.035	0.036	0.055	19753	16972	22146	9050	3440
CCR5_OT6	TCAGAATAGAGAGCCGAACGG	SERTM2	Intergenic	0	0.014	0.026	0.042	0.07	22218	18997	30971	15789	2255
CCR5_OT7	ACAGAATAGAGAGCCCAGAAAGG	AMER1	Intergenic	0.076	0.05	0.071	0.04	0	23880	15464	22453	59	2641
CCR5_OT8	ACAACAAAATGAGCCCAGAAAGG	MPLKIP	intron	0.045	0	0.025	0.059	0.068	10	8079	3368	2960	4408
CCR5_OT9	GAAGCAAAGAGAGCCCAGAATGG	MYOF	intron	0	0	0.035	0	0.049	1440	5684	2841	4111	4097
CCR5_OT10	ACAGAGTAGAGAGCCCAGAAAGG	MIR4426	Intergenic	0.02	0.017	0.054	0.022	0.026	23972	22021	17890	30204	10103
CCR5_OT11	ACAGCAGTGGGAGCCCAGAAAGG	SPATA5	intron	0.033	0.027	0.041	0.033	0.059	22219	24223	18295	23758	6095
CCR5_OT12	CCATCAAAGTTAGCCCAGAACGG	CISTR	Intergenic	0	0	0.026	0.023	0.016	314	15441	17491	25162	7446
CCR5_OT13	TCAGCAGACTCAGCCCAGAAAGG	QTRT2	intron	0.015	0.032	0.047	0.045	0.025	21616	17111	17849	23533	6757
CCR5_OT14	ACAGCATAGAGGGCCCAGAAGGG	SUOX	exon	0	0.105	0.128	0.095	0.051	27733	4686	7399	5859	1233
CCR5_OT15	TCAGCAGTGTGGGCCCAGAAAGG	LMAN1	intron	0.13	0.058	0.162	0.07	0.095	19004	13541	17090	18998	3847
CCR5_OT16	GCAGCACAGGGGGCCCAGAATGG	PSMD3	intron	0.09	0	0.073	0.066	0.052	3080	16533	6070	13525	4466
CCR5_OT17	CCAGCCCAGTGACCCCAGAAGG	COL12A1	intron	0.073	0.07	0.067	0.049	0.079	35916	25487	14296	22671	6880
CCR5_OT18	GCAGAATTGTGAACCCAGAAGGG	SQSTM1	intron	0.116	0.042	0.053	0.036	0.071	4744	18949	13988	15595	2581
CCR5_OT19	ACAACCTAGTGAGGCCGAAGGG	COA7	intron	0.025	0.008	0.013	0.026	0.042	26311	14882	11557	11775	3987
CCR5_OT20	ACAGAATGGTGAGACCGAAATGG	FAM151A	intron	0	0.014	0	0	0	7013	973	818	197	1163
CCR5_OT21	CCAGCAAAGTGCTCCCAGAATGG	LOC105378853	intron	0.032	0.024	0.01	0.027	0.02	16444	20159	21986	15242	6318
CCR5_OT22	CCAGCACAGTGGACCCAGAAAGG	HIPK2	intron	0.067	0.042	0	0.057	0	30621	5919	8846	1287	1500
CCR5_OT23	CCAGCATGGTGCCCCCAGAAGGG	SCNN1B	intron	0.426	0.201	0.144	0.199	0.266	29778	20829	23107	24454	1876
CCR5_OT24	GCAACAGAGTGAGCGCAGAAGGG	SOX3	Intergenic	0	0.053	0.031	0.077	0.014	9470	22341	22036	7045	3101
CCR5_OT25	GCAGCAGATGGGGCCAGAAAGG	MIR122	intron	0.024	0.024	0.055	0.03	0	41387	21778	23382	16711	4084
CCR5_OT26	CCAGCAAAGGGAGCACAGAAGGG	DRD2	Intergenic	0	0.005	0.008	0.018	0.011	38393	26582	22831	17671	3292
CCR5_OT27	ACAGCAAAGTGAATCCAGAAAGG	CSRNP3	Intergenic	0	0.005	0.068	0.006	0.036	21585	19008	17104	19461	5948
CCR5_OT28	GCAGCATACAGAGCTCAGAAAGG	TMEM132D	intron	0.012	0.016	0.024	0.025	0.013	31658	29533	36085	23368	8472
CCR5_OT29	CCAGCAGGTGAGCCAAGAATGG	RP1	Intergenic	0.121	0.003	0.09	0.021	0.083	33009	20067	18717	19187	7425
CCR5_OT30	CCAGCATAGGAAGCACAGAATGG	PRDM2	Intergenic	0	0.009	0.017	0	0.012	32503	23686	21755	25653	3840
CCR5_OT31	GCAGCAGAGTGAGCCCACAATGG	CCR2	exon	0.039	0.035	0.023	0.048	0.07	28703	8707	12414	4263	5066
CCR5_OT32	TCTCCATAGTGAGCCCATAAAGG	CDH2	Intergenic	0.013	0.012	0.024	0.02	0.03	40255	29035	25314	26394	7689
CCR5_OT33	TCACCACAGTGAGCCCATAAAGG	CCDC88C	intron	0.039	0	0.021	0.024	0.028	225	23811	21270	25438	10253
CCR5_OT34	CCAGCACAGTGAGTCCGGAAGGG	MIR148A	Intergenic	0.117	0.075	0.077	0.102	0.129	19970	16954	21486	20165	9377
CCR5_OT35	ACAGTACAGTGAGCCCAGGATGG	LINC01812	Intergenic	0.023	0.032	0.029	0.028	0.033	37925	27755	24584	21038	13199
CCR5_OT36	CCAGCACGGTGAGCCCAGCATGG	MROH5	intron	0.034	0.014	0.024	0.021	0.018	43525	20787	28047	32663	8896
CCR5_OT37	CCAGCATGCTGAGCCCAGCAGGG	CEL	intron	0.038	0.003	0.026	0.005	0.047	33847	19060	21827	21439	7840
CCR5_OT38	ACAGCAGAGGGAGCCCAGCATGG	LINC00963	Intergenic	0.012	0.018	0.032	0.036	0.027	16951	22074	13984	21861	8178
CCR5_OT39	ACAGCATAGTGAACCCAGGAGGG	TBPL2	Intergenic	0.017	0.388	0.302	0.102	0.052	32448	30133	21500	29042	5732
CCR5_OT40	GCTGCATAGTGAACCCAGTATGG	ZNF609	Intergenic	0.032	0.122	0.245	0.031	0.014	28637	12663	16325	35019	6293
CCR5_OT41	GGAGGATAGTGAGCCCAGAGCGG	MAST2	intron	0.05	0.017	0.032	0.031	0.03	29887	27990	19368	26278	3975
CCR5_OT42	ACAGGAAAGTGAGCCCAGAGGGG	JPH2	intron	0.014	0.026	0.028	0.007	0.049	23064	17835	14052	12207	7124
CCR5_OT43	CCAGCATACAGAGCCCAGAGAGG	EWSAT1	TTS	0.014	0.027	0.031	0.042	0.048	36861	16004	21524	23083	14226
CCR5_OT44	CCAGCATAGTGAGCTCAGACGGG	ADCY10	intron	0	0.009	0	0.006	0.021	22559	9974	17576	19144	7950
CCR5_OT45	CCAGCATAGTAAGCCCCGAGAGG	MED13L	intron	0.125	0.026	0.092	0.045	0	38957	5407	6625	841	4011
CCR5_OT46	ACAGCATGGTGAGCCCACAGTGG	LINC01017	intron	0.045	0.03	0.026	0.028	0.052	36233	26521	17782	21239	6653
CCR5_OT47	GCAGCATAGTGAGACCAGCCTGG	LOC344967	intron	0	0.016	0.007	0.006	0	18870	13807	15733	11530	13
CCR5_OT48	TCAGAAGAGTGAGCCCAGAAAAG	LINC01843	Intergenic	0	0.021	0.026	0.029	0.035	38261	18881	27445	25617	7373
CCR5_OT49	ACAACATTGTGAGCCCGAATAG	LRRC8D	intron	0	0.027	0.05	0.038	0.033	29312	12086	18461	11972	2156
CCR5_OT50	ACAGAATAGAGAGCCCAGAATAG	GRM7-AS3	Intergenic	0.087	0.044	0.036	0.037	0.073	31982	8442	26995	17715	3444
CCR5_OT51	ACAGAATAGAGAGCCCAGAAAGG	ZNF680	intron	0.078	0.047	0.047	0.042	0.024	35796	23271	21580	28583	6449
CCR5_OT52	ACAGCATACTGAACCCAGAAAAG	CSMD3	Intergenic	0.044	0.03	0.021	0.028	0.057	33860	24157	21384	21207	2274
CCR5_OT53	CCAGCTTAGTGAGCCCAGGAAAG	LOC101927974	intron	0	0.031	0.018	0.027	0.07	32037	16375	14992	10030	1335
CCR5_OT54	CCAGCATACTGAGCCCAGACGAG	PTGR1	intron	0.108	0.021	0.052	0.033	0	19350	11638	6130	11	6503
CCR5_OT55	AAAGC^TAGTGAGCCCAGAAAGG	CHD5	intron	0.041	0.014	0.028	0.029	0.02	28555	25087	20953	19641	14519
CCR5_OT56	GAAGCA^AGTGAGCCCAGAAGGG	ABCC4	intron	0.042	0.046	0.039	0.079	0.048	34783	25532	22825	22858	14377
CCR5_OT57	TCAGGA^AGTGAGCCCAGAAAGG	KRT40	intron	0.011	0.02	0.034	0.033	0.046	35535	23675	24208	17298	18371
CCR5_OT58	TCAGAATAGTGA^CCCAGAAGGG	LOC100129034	intron	0.037	0.013	0.034	0.041	0.061	30486	14878	14544	11414	10705
CCR5_OT59	TCAGCAAAGTGAGCCCCAGAAAGG	MTUS2	intron	0	0.07	0.108	0.081	0.101	21472	13948	11093	11919	563
CCR5_OT60	ACAGC^TAGTGAGCCCAGCATGG	CACNA1C	intron	0.039	0.018	0.037	0.038	0.027	43798	21871	7804	7541	5183
CCR5_OT61	CCAGCAGAGTGAGCCCAG^AGGG	ADRA2C	Intergenic	0	0	0	0	0	0	0	0	0	0
CCR5_OT62	TCAGCAT^GTGAGCCCAGCATGG	HDAC8	intron	0.064	0.015	0.081	0.058	0.052	32870	24594	15556	22880	15726

	Sites within repetitive element												
CCR5_OT63	ACAGAATAGAGAACCCAGAAAGG	MAB21L1	intron	0.087	0.386	0.47	0.32	0.432	21759	12544	9683	13436	1147
CCR5_OT64	ACAGAATAGAGAACCCAGAAAGG	MIR924HG	Intergenic	0.055	0.02	0.012	0.012	0.027	24399	25263	17225	18348	5488
CCR5_OT65	ACAGAATAGAGAACCCAGAAAGG	LOC105374820	Intergenic	0.108	0.1696	0.056	0.124	0.061	11967	23168	11334	21152	2789
CCR5_OT66	ACAGAATAGAGAACCCAGAAAGG	BMP3	intron	0	0.025	0.018	0.012	0.01	32291	22189	25417	20336	3015
CCR5_OT67	ACAGAATAGAGAACCCAGAAAGG	CYLC1	Intergenic	0.057	0.039	0.041	0.137	0.101	10282	19301	9518	3949	1768

Extended Data Fig. 10 l Comprehensive off-target site analysis of the CCR5 sgRNA. Table lists all 67 COSMID predicted sites, the sequence, genomic location, percent Indels in two experiments and the number of reads for each site and in each experiment. Samples with percent Indels > 0.1% are highlighted in pink. Samples with low coverage are highlighted in light yellow. CCR5 _OT63 through 67 were located within a repetitive element. Three of these had primers that should have been unique per locus but the Indel analysis showed that this was not the case. Therefore, the true off-target rated at these sites were not ascertainable but should all be < 0.5%.

Gomez et al, Extended Data Figure 10