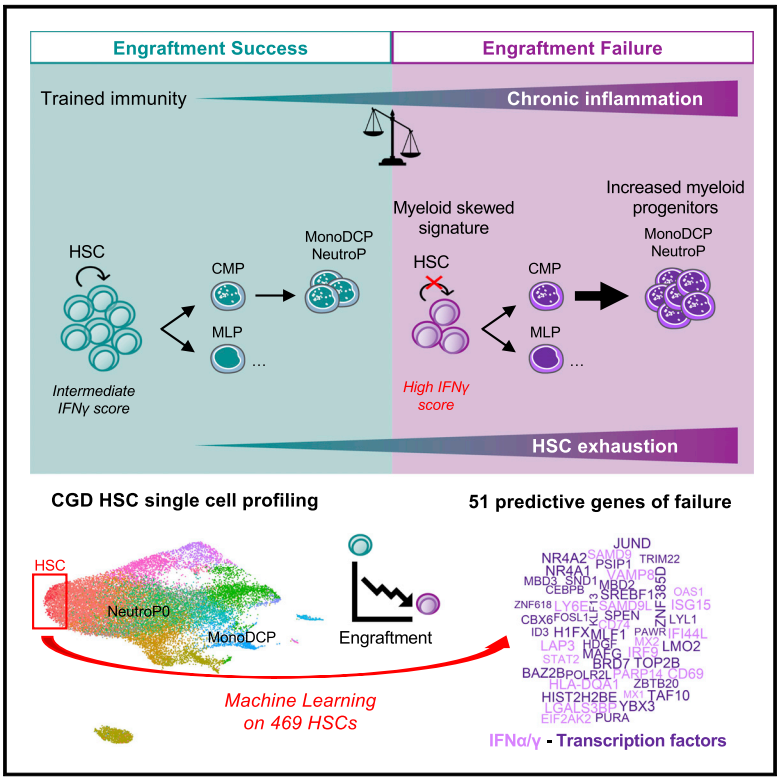
Cell Reports
Medicine

Article

Severe hematopoietic stem cell inflammation compromises chronic granulomatous disease gene therapy

Graphical abstract



Highlights

- CGD gene therapy efficiency depends on the level of chronic inflammation
- Loss of gene-corrected HSCs correlates with the level of interferon score in HSCs
- Chronic inflammation drives HSCs exhaustion and increased myeloid progenitors
- Machine learning identified 51 predictive markers of engraftment failure

Authors

Steicy Sobrino, Alessandra Magnani, Michaela Semeraro, ..., Stephane Blanche, Marina Cavazzana, Emmanuelle Six

Correspondence

m.cavazzana@aphp.fr

In brief

Sobrino et al. report heterogeneous efficacy of a gene therapy approach for CGD patients. Transcriptomic exploration at the single-cell level shows lower HSC frequency and elevated myeloid progenitor frequency in the most severe patients. This study unravels 51 markers in HSCs predictive of the engraftment failure of gene-corrected cells.

Sobrino et al., 2023, Cell Reports Medicine *4*, 100919
February 21, 2023
https://doi.org/10.1016/j.xcrm.2023.100919

CellPress

Cell Reports Medicine

CellPress
OPEN ACCESS

Article

Severe hematopoietic stem cell inflammation compromises chronic granulomatous disease gene therapy

Steicy Sobrino,[1,27] Alessandra Magnani,[2,3,27] Michaela Semeraro,[4] Loredana Martignetti,[5] Akira Cortal,[5] Adeline Denis,[1] Chloé Couzin,[2,3] Capucine Picard,[6,7,8] Jacinta Bustamante,[6,9,10] Elisa Magrin,[2,3] Laure Joseph,[2] Cécile Roudaut,[2,3] Aurélie Gabrion,[2,3] Tayebeh Soheili,[1] Corinne Cordier,[11] Olivier Lortholary,[12] François Lefrere,[2,13] Frédéric Rieux-Laucat,[14] Jean-Laurent Casanova,[9,10] Sylvain Bodard,[15,16] Nathalie Boddaert,[17]

(Author list continued on next page)

[1]Human Lymphohematopoiesis Laboratory, Université Paris Cité, Imagine Institute, INSERM UMR 1163, Paris, France
[2]Biotherapy Department, Necker-Enfants Malades Hospital, AP-HP, Paris, France
[3]Biotherapy Clinical Investigation Center, Groupe Hospitalier Universitaire Ouest, AP-HP, INSERM, Paris, France
[4]Clinical Investigation Center CIC 1419, Necker-Enfants Malades Hospital, GH Paris Centre, Université Paris Cité, AP-HP, Paris, France
[5]Clinical Bioinformatics Laboratory, Université Paris Cité, Imagine Institute, INSERM UMR 1163, Paris, France
[6]Study Center for Primary Immunodeficiencies, Necker-Enfants Malades Hospital, AP-HP, Université Paris Cité, Paris, France
[7]Lymphocyte Activation and Susceptibility to EBV Infection Laboratory, INSERM UMR 1163, Imagine Institute, Paris, France
[8]Centre de Références des Déficits Immunitaires Héréditaires (CEREDIH), Necker-Enfants Malades Hospital, AP-HP, Paris, France
[9]Human Genetics of Infectious Diseases Laboratory, Université Paris Cité, Imagine Institute, INSERM UMR 1163, Paris, France
[10]St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, NY, USA

(Affiliations continued on next page)

SUMMARY

X-linked chronic granulomatous disease (CGD) is associated with defective phagocytosis, life-threatening infections, and inflammatory complications. We performed a clinical trial of lentivirus-based gene therapy in four patients (NCT02757911). Two patients show stable engraftment and clinical benefits, whereas the other two have progressively lost gene-corrected cells. Single-cell transcriptomic analysis reveals a significantly lower frequency of hematopoietic stem cells (HSCs) in CGD patients, especially in the two patients with defective engraftment. These two present a profound change in HSC status, a high interferon score, and elevated myeloid progenitor frequency. We use elastic-net logistic regression to identify a set of 51 interferon genes and transcription factors that predict the failure of HSC engraftment. In one patient, an aberrant HSC state with elevated CEBPβ expression drives HSC exhaustion, as demonstrated by low repopulation in a xenotransplantation model. Targeted treatments to protect HSCs, coupled to targeted gene expression screening, might improve clinical outcomes in CGD.

INTRODUCTION

Chronic granulomatous disease (CGD) is a recessive inborn error of immunity[1,2] caused by loss-of-function (LOF) mutations in the X-linked or autosomal genes that encode the five components of the NADPH oxidase complex.[3]

The complex's membrane-bound catalytic core is a heterodimer of gp91^{phox} and p22^{phox}, encoded respectively by the X-linked *CYBB* gene and the autosomal *CYBA* gene. The regulatory part is a cytosolic heterotrimer composed of p40^{phox}, p47^{phox}, and p67^{phox}, encoded respectively by *NCF4*, *NCF1*, and *NCF2*.[4] Following phagocyte activation, the NADPH oxidase complex assembles on the phagosomal membrane and produces reactive oxygen species (ROS).

Patients with CGD suffer from specific, recurrent, invasive, life-threatening bacterial and fungal infections.[5,6] Prominent inflammatory manifestations (particularly affecting the respiratory and gastrointestinal tracts) are also common, especially in patients with the X-linked form of the disease.[7,8] In some patients, CGD is revealed by these inflammatory manifestations. Others present initially with unexplained granulomatosis, which is associated with a poor prognosis.[8,9]

Patients routinely receive antimicrobial prophylaxis and, eventually, anti-inflammatory treatments according to the clinical

Adrian J. Thrasher,[18,19] Fabien Touzot,[2,3] Sophie Taque,[20] Felipe Suarez,[12,21] Ambroise Marcais,[12] Agathe Guilloux,[22] Chantal Lagresle-Peyrou,[1,3] Anne Galy,[23,24,28] Antonio Rausell,[5,25,28] Stephane Blanche,[26,28] Marina Cavazzana,[2,3,21,28,29,*] and Emmanuelle Six[1,28]

[1]Plateforme de Cytométrie en Flux, Structure Fédérative de Recherche Necker, INSERM US24-CNRS UAR3633, Paris, France
[12]Necker-Pasteur Center for Infectious Diseases and Tropical Medicine, Necker-Enfants Malades Hospital, AP-HP, Université Paris Cité, Imagine Institute, Paris, France
[13]Department of Adult Hematology, Necker-Enfants Malades Hospital, AP-HP, Paris, France
[14]Immunogenetics of Pediatric Autoimmune Diseases Laboratory, Université Paris Cité, Imagine Institute, INSERM UMR 1163, Paris, France
[15]Department of Adult Radiology, Necker Enfants-Malades Hospital, AP-HP, Université Paris Cité, Paris, France
[16]Laboratoire d'Imagerie Biomédicale, LIB, Sorbonne Université, CNRS, INSERM, Paris, France
[17]Département de Radiologie Pédiatrique, INSERM UMR 1163 and UMR 1299, Imagine Institute, AP-HP, Necker-Enfants Malades Hospital, Paris, France
[18]UCL Great Ormond Street Institute of Child Health, London, UK
[19]Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK
[20]CHU de Rennes, Département de Pédiatrie, Rennes, France
[21]Imagine Institute, Université Paris Cité, Paris, France
[22]Mathematics and Modelization Laboratory, CNRS, Université Paris-Saclay, Université d'Evry, Evry, France
[23]Genethon, Evry-Courcouronnes, France
[24]Université Paris-Saclay, University Evry, Inserm, Genethon (UMR_S951), Evry-Courcouronnes, France
[25]Service de Médecine Génomique des Maladies Rares, AP-HP, Necker-Enfants Malades Hospital, Paris, France
[26]Department of Pediatric Immunology, Hematology, and Rheumatology, Necker-Enfants Malades Hospital, AP-HP, Paris, France
[27]These authors contributed equally
[28]These authors contributed equally
[29]Lead contact
*Correspondence: m.cavazzana@aphp.fr
https://doi.org/10.1016/j.xcrm.2023.100919

manifestations. The only widely available, curative treatment is allogeneic hematopoietic stem cell transplantation (HSCT).[10–13] After conditioning, $CD34^+$ hematopoietic stem and progenitor cells (HSPCs) are transplanted; engraftment of the most immature hematopoietic stem cells (HSCs) in the bone marrow (BM) then enables full immune reconstitution. In the absence of a compatible donor for HSCT, gene therapy (GT) is a treatment option. Although several research groups have developed GT protocols for CGD, the previous clinical trials have been compromised by the absence of stable engraftment of the gene-corrected cells.[14] In the first trials (without a conditioning regimen), the lack of engraftment was probably due to the absence of a selective advantage for the transduced cells.[14,15] The use of a low-intensity conditioning regimen in the subsequent trials with gammaretroviral vectors resulted in temporary engraftment, although insertional mutagenesis favored the development of myelodysplastic syndromes in a few patients.[16,17] More recently, significant improvements were achieved after the gammaretroviral vectors were replaced by self-inactivating lentiviral vectors in which a chimeric internal promoter drove $gp91^{phox}$ expression specifically in myeloid cells.[18] The combination of this new-generation vector with a full busulfan-based myeloablative conditioning regimen resulted in significantly better clinical and biological outcomes and a better safety profile (i.e., the absence of GT-related adverse events) in two trials (a trial in London sponsored by Généthon [ClinicalTrials.gov identifier: NCT01855685] and Kohn et al.'s investigator-led trial in the United States [NCT02234934][19]). In 9 of the 13 treated patients, the stably engrafted cells cured the underlying X-linked CGD (X-CGD).

Several recent studies have reported that chronic inflammation harms HSPCs in patients with CGD and patients with other conditions. Mice and humans with X-CGD have low HSC counts in the BM.[20] Furthermore, human HSCs from patients with CGD showed rapid exhaustion after *in vitro* culture. In the presence of high levels of pro-inflammatory cytokines (such as IL-1β), mouse HSCs showed increased cycling and a lower long-term engraftment potential.[20] Elevated levels of IL-18 and interferon γ (IFN-γ) have been observed in inflamed tissue from patients with CGD.[21]

Here, we report the results of a phase I/II clinical trial of GT (based on a G1XCGD lentiviral vector and gene-modified HSPCs) in four patients with X-CGD lacking a human leukocyte antigen (HLA)-compatible donor for HSCT. A fifth patient was included in the clinical trial but was not treated because the investigational medicinal product (IMP) did not meet the release criteria. The degrees of cell engraftment and clinical efficacy varied markedly from one patient to another. To fully understand the molecular alterations in HSCs associated with the success or failure of GT for CGD, we profiled the transcriptome of HSPCs at the single-cell level. We found that the gene-corrected cell engraftment defect observed in two patients was correlated with the upregulation of the IFN pathway. Last, we identified a set of biomarkers (including IFI44L and CEBPB) that were predictive of GT failure.

RESULTS

Clinical presentation of patients

We performed a nonrandomized, open-label, phase I/II clinical study (ClinicalTrials.gov identifier: NCT02757911) including five patients with X-CGD (referred to hereafter as P1 to P5). The patients had a severe deficiency in $gp91^{phox}$ protein, due to a mutation in the *CYBB* gene and the absence of NADPH oxidase activity. Four patients received autologous $CD34^+$ cells transduced with a lentiviral GT vector after cryopreservation of the IMP. One patient (P3) was not treated because the level of

$CD34^+$ cell transduction did not meet the release specification. The patients' age at the time of GT ranged from 8 to 28 years.

All four treated patients had severe X-CGD-related infections, some of which were active at the time of GT. Three of the patients (P1, P2, and P5) also had severe inflammatory manifestations (Table 1). Prior to GT, all four patients received standard antimicrobial/antifungal prophylaxis and (for those with inflammatory features) long-term anti-inflammatory treatments.

In particular, P1 (8 years of age at the time of GT) presented with many deep abscesses, gut inflammation, and severe lung disease with infectious and inflammatory components; hence, P1 was receiving oxygen therapy and enteral nutrition in addition to steroids and antimicrobial prophylaxis.

P2 and P5 (respectively 19 and 28 years of age at the time of GT) had similar clinical profiles, with very severe, long-lasting, corticoresistant inflammation and typical CGD-associated infections. Since infancy, P2 had presented with treatment-resistant granulomatous cystitis. He also had a history of tibial osteomyelitis and actinomycotic abscesses of the liver with portal hypertension, requiring surgery. P5 presented with long-lasting severe colitis that was refractory to various anti-inflammatory treatments, together with pulmonary aspergillosis, osteitis, and *Campylobacter* and *Salmonella* infections.

In contrast to the other patients, P4 (23 years of age at the time of GT) did not have a history of inflammation but presented with life-threatening, invasive, treatment-resistant pulmonary aspergillosis, *Salmonella* infections, cervical adenitis, and folliculitis. In view of P4's critical condition and the absence of other treatment options, compassionate-use GT was authorized; at that time, the gene correction process was under optimization by the addition of prostaglandin E2 (PGE2) (see below).

Manufacturing and characteristics of the IMP

IMPs were manufactured from HSPCs harvested from the BM (P1) or via leukapheresis (P1 to 5) after granulocyte colony-stimulating factor (G-CSF)/plerixafor cell mobilization. The gene-corrected cells were infused after targeted myeloablative conditioning (median [range] area under the curve for total exposure to busulfan: 75,610 [71,973–85,478] ng × h/mL) (Table 1 and STAR Methods). The infused $CD34^+$ cell doses ranged from 3.0 to 15.67 × 10^6/kg. The infused IMPs are described in detail in Table 1, and the transduction procedure is summarized in Figure 1A (see also STAR Methods).

P1 received a final IMP containing genetically modified $CD34^+$ HSPCs sourced from BM and mobilized peripheral blood (MPB) (G-CSF + plerixafor-mobilized leukapheresis), as specified in the initial protocol design. For P2, a low yield of $CD34^+$ cells after BM harvest prevented gene correction, and so the unmodified cell product was cryopreserved. He underwent two subsequent aphereses leading to below-specification levels of gene transduction. As a consequence, the transduction protocol was modified by the addition of PGE2 (which reportedly favors HSC transduction and repopulation).[22] The addition of this transduction adjuvant considerably improved the level of lentiviral transduction (see STAR Methods and Figure S1). The following three procedures (for P4, P2, and P5) were therefore performed with the optimized protocol, starting from G-CSF + plerixafor (P4, P5)- or plerixafor-only (P2)-mobilized leukapheresis. Using classical HSPC phenotyping, we did not detect differences in the HSC frequency (defined as $CD34^+Lin^-CD38^-CD133^+CD90^+CD45RA^-$ cells), neither in the IMP (Figures S2A and S2C) nor in the apheresis before engineering (Figures S2A and S2B), and we showed that the patients had received similar doses of HSCs per kilogram (Figure S2D).

Indeed, the addition of PGE2 during the transduction step was associated with a significantly greater vector copy number (VCN) both in pre-clinical tests and in the IMPs ($p = 0.0022$ and $p = 0.0357$, respectively) (Figures S1A and S1B). Moreover, transcriptomic analysis of PGE2-treated and control HSPCs from P2 and P4 showed that the addition of this adjuvant was associated with a less inflammatory expression profile (Figure S1C). For the four treated patients, the median (range) VCN in the IMP was 1.42 (0.99–1.73).

Clinical outcomes

After a myeloablative conditioning regimen, patients were infused with the IMP containing genetically modified HSPCs. The infusion was well tolerated in all cases. The only adverse events were related to the conditioning (e.g., mucositis), rather than the IMP. P1 presented with *Staphylococcus epidermidis* sepsis 6 days after infusion of the IMP, and P2 presented with cytolysis and cholestasis 14 days after infusion. These adverse events resolved after engraftment, and hematopoietic reconstitution was satisfactory for all patients; the median (range) time to neutrophil and platelet engraftment was 18.5 days (15–21 days) (Figure S3A).

As of January 2022, the median (range) follow-up period was 42 months (24–60). For all treated patients, the VCN in neutrophils ranged from 0.17 to 0.96 in the first month post-GT. P1 showed an initial decrease in the level of gene marking, which stabilized at around 10%–15% after a few months (Figure 1B). Although this level was not optimal, it provided P1 with clinical benefit, particularly with regard to the regression of infectious manifestations and as shown by the post-GT lung scan results (Figure 1E); this enabled P1 to discontinue nocturnal oxygen therapy, enteral nutrition, steroids, and antimicrobial prophylaxis. However, the inflammatory manifestations continue to progress (particularly in the gut and lung), requiring the recent introduction of Janus kinase (JAK) 1 and 2 inhibitors.

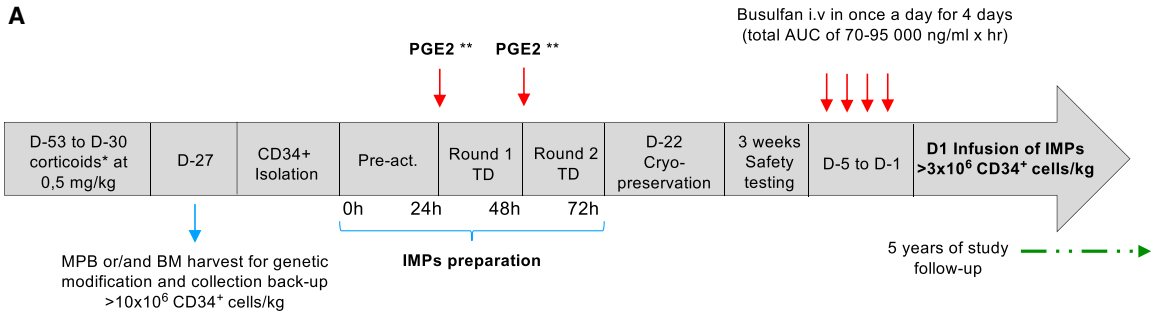
In P2 and P5, a progressive decrease in the engraftment of gene-corrected cells was observed 2 to 3 months after GT, and the patients regressed to their pre-GT condition (Figure 1B). Similar results were observed for monocytes, B cells, natural killer (NK) cells, and T cells. The level of gene marking was lower in T cells, given the absence of T cell depletion during the conditioning (Figure S3B). Due to the recurrence of inflammation and infections, P2 underwent HSCT with an unrelated, partially matched donor (1 of 10 HLA alleles was mismatched) 3.5 years after GT. Twenty-seven days after the HSCT, P2 developed ultimately fatal septic shock. Seven months after infusion of the IMP, P5 presented with submandibular lymphadenopathy that resolved progressively with oral antibiotic treatment. The patient continued to receive antimicrobial prophylaxis, and his clinical condition is currently stable.

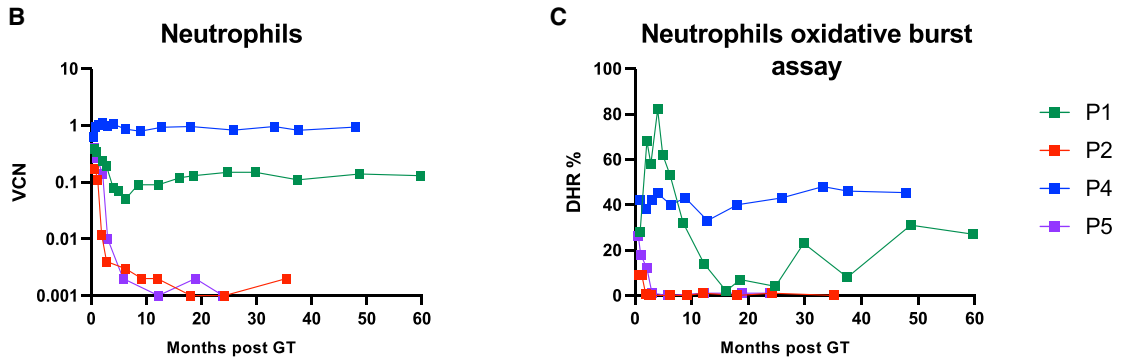
Shortly after GT, P4's life-threatening lung aspergillosis resolved completely, with *ad integrum* healing of the lytic costal

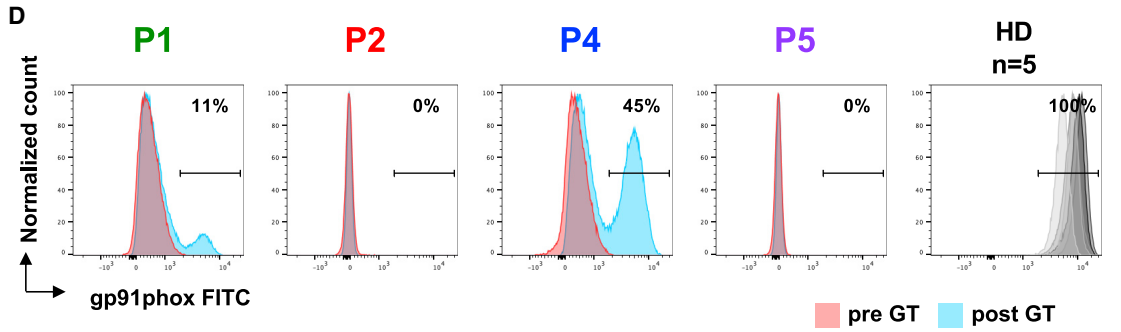
Table 1. Clinical features of the patients with X-linked CGD before and after GT, and characteristics of the infused, gene-corrected, autologous cell product

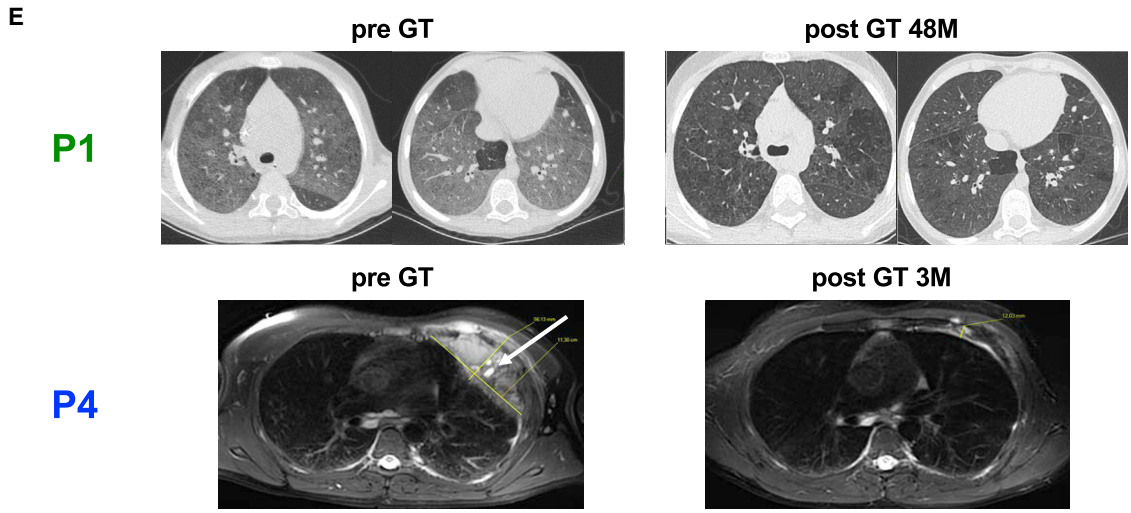
		P1	P2	P3	P4	P5
Age at GT (years)		8	19	6	23	28
X-CGD (*CYBB*) mutation		c.779C>G; (p.Pro260Arg); exon 7	c.1083G>A; (p.Trp361X); exon 9	c.736C>T; (p.Gln246X); exon 7	c.469C>T; (p.Arg157X); exon 5	c.736C>T; (p.Gln246X); exon 7
Follow-up post-GT (months)		60	36	not infused	48	24
HSPC source		BM + MPB	MPB	MPB	MPB	MPB
HSPCs infused (10^6/kg)		BM 3; MPB 5.86	3.24	not infused	15.67	14.27
VCN drug product		BM 0.6; MPB 1.2	1.26	0.35	1.73	1.57
Transduction adjuvant: PGE2		–	+	–	+	+
Busulfan conditioning, AUC (ng × h/mL)		85,478	77,330	not infused	71,973	73,890
DHR^+ neutrophils (%)	At 12 months	14	1	not infused	33	<1
	At last follow-up	27	<1	not infused	47	1
Treatments ongoing at GT		antimicrobial prophylaxis, steroids, enteral nutrition, sleeping O_2 treatment	antimicrobial prophylaxis, steroids	antimicrobial prophylaxis, steroids	antimicrobial prophylaxis, antifungal treatment	antimicrobial prophylaxis, hydroxychloroquine
Infectious history (before GT)		multiple deep abscesses (*S. marcescens*), recurrent pneumonitis	tibial osteomyelitis, liver abscess, *Actinomyces* hepatic abscess with portal hypertension	mucormycosis, *Aspergillosis*, *Clostridium*	severe invasive pulmonary aspergillosis, pneumonitis, salmonellosis, folliculitis, cervical adenitis	aspergillosis, pneumonitis, salmonellosis, genitourinary *Campylobacter*, osteitis (*S. marcescens*)
Inflammatory history (before GT)		corticodependent inflammatory colitis, pulmonary granuloma	corticodependent, long-lasting granulomatous cystitis resistant to multiple treatments	early colitis, granulomatous gastritis	folliculitis	long-lasting severe colitis resistant to multiple treatments
Clinical follow up after GT		lung scan improvement, progression/relapse of gut inflammation requiring anti-inflammatory treatment	decreased corticosteroids; at month 11, pulmonary aspergillosis infection, gastric hemorrhage; at month 24, lymphadenitis; at 3.5 years, MMUD HSCT, septic shock, deceased	not infused	marked improvement in thoracic fungal lesions; disappearance of folliculitis; off treatment	clinically stable; antimicrobial prophylaxis; at month 7, submandibular lymphadenopathy

AUC, area under the curve; BM, bone marrow; MPB, mobilized peripheral blood; GT, gene therapy; PGE2, prostaglandin E2; HSCT, hematopoietic stem cell transplantation; VCN, vector copy number; MMUD, mismatched unrelated donor; MRI, magnetic resonance imaging.









(legend on next page)

erosions as early as 4 months post-GT (Figure 1E). This stable, clinical benefit was associated with the presence of functional circulating neutrophils (50% of the normal proportion) (Figure 1C). P4 resumed his education and is now working full time. He discontinued all treatments 2 months post-GT.

The gene marking results were correlated with the results of the functional oxidative burst (dihydrorhodamine 123 [DHR]) assay; the proportion of positive neutrophils stabilized at 27% and 47% of neutrophils, for P1 and P4, respectively (Figure 1C). At last follow-up, the expression of gp91^{phox} protein in CD15^{+} neutrophils were significantly enhanced for P1 and P4 (Figure 1D).

The analysis of vector integration over time in the patients highlighted the polyclonal reconstitution of both peripheral blood mononuclear cells (PBMCs) and neutrophils (Figure S4). The mean (range) number of unique integration sites at last follow-up was 3,252 (119–10,650) in PBMCs and 4,865 (158–15,344) in neutrophils. Lower values were observed for P2 and P5, due to the progressive loss of gene-corrected cells. Integrations close to oncogenes (such as MECOM [MDS/EVI1]) previously targeted by gammaretroviral vectors were present at a low frequency (below 2%) in all patients and did not increase over time.

A low frequency of HSCs and a high frequency of myeloid progenitors

To understand the interindividual differences in engraftment, we analyzed transcriptomic differences in PBMCs and HSPCs from patients versus healthy donors (HDs). This analysis highlighted the upregulation of the type 1 and 2 IFN response pathway in PBMCs and in HSPCs (Figures S5A and S5B, respectively). We also used the ROMA method[23] to quantify the activity of sets of genes in individual samples. This analysis did not reveal any interindividual differences among the patients' PBMCs. However, P2's and P5's HSPCs had a higher IFNα score, and a more intense IFNγ score, relative to the other patients. In contrast, P4 (the patient with the best engraftment of gene-corrected cells) displayed an only slightly higher IFNα score (Figures S5C and S5D). To further explore these interindividual differences in HSPC subpopulations, we performed single-cell transcriptomic analyses to determine the transcriptional profiles of 53,412 MPB HSPCs from the four treated patients with CGD and four HDs (Figure 2A). Due to the continuum between the various stem and progenitor cells, the use of classical clustering analysis to distinguish the different cell types remains challenging.[24] For this reason, and to preserve the strength of the single-cell resolution, we used an automated cell annotation method, Cell-ID,[25] and 16 different BM HSPC reference signatures (Table S1).[26] We developed a single-cell RNA-sequencing (RNA-seq) pipeline for HSPC data processing and analysis as depicted in Figure S6.

To define the most immature HSC subpopulation, we used a diffusion mapping approach to determine the origin in the trajectory map (Figures S7A–S7C). We then compared the frequency of each subpopulation in the patients with that in the HDs (Figures 2B and 2C). We showed that CGD patients were likely to have around half the number of HSCs found in HDs (odds ratio [OR] = 0.53; $p = 2.2 \times 10^{-16}$, using logistic regression). Importantly, patients with engraftment failure (P2 and P5) had an even lower proportion of HSCs than patients with successful engraftment (P1 and P4) (OR = 0.58, $p = 3.78 \times 10^{-8}$, using logistic regression). Moreover, P2 presented a high frequency of B cell progenitors and monocyte/dendritic cell progenitors (Figure 2C). P5 had a high proportion of neutrophil progenitors (NeutroP0; Figure 2C) and a low proportion of immature progenitors (ImP1; similar to common myeloid progenitors, Figure S7D). No differences in frequencies of the various HSPC cell types were observed for the other two patients (P1 and P4), relative to HDs (Figure S7D).

An aberrant HSC profile, with a mixture of HSC/neutrophil signatures

To further explore the abnormally large NeutroP0 subpopulation in P5, we used the Cell-ID method to identify cells that simultaneously matched multiple cell-type signatures. Thus, by testing each individual cell against the 16 reference signatures, each individual cell could be identified as simultaneously displaying signatures from several cell types (Figure S8A). In fact, the majority of cells (ranging from 61% to 90%) are matching two or more signatures, while cells matching a single signature range between 6% and 26% of the total population (Figure S8B). For example, the *NeutroP0^{Match}* population (Figure S8C) encompassed not only the NeutroP0^{ID} population (Figure S8D) but also cells displaying other top signatures, yet showing significant enrichments for the NeutroP0 gene signature. An UpSet plot of the various mixed signatures showed that there were 19 distinct combinations of the NeutroP0 signature with other cell types in P5 but only three distinct combinations in HDs (black arrow, Figure S8E). We therefore looked further at the most frequent combinations in P5 that comprised NeutroP0 signatures (Figure 3). This analysis revealed that 438 cells matched the NeutroP0, MPP, and All HSC signatures. This mixed signature (depicted in black in the uniform manifold approximation and projection [UMAP] plot, Figure 4A) was found in P5 but not in the other

Figure 1. Patient follow-up after gene therapy, and clinical improvements

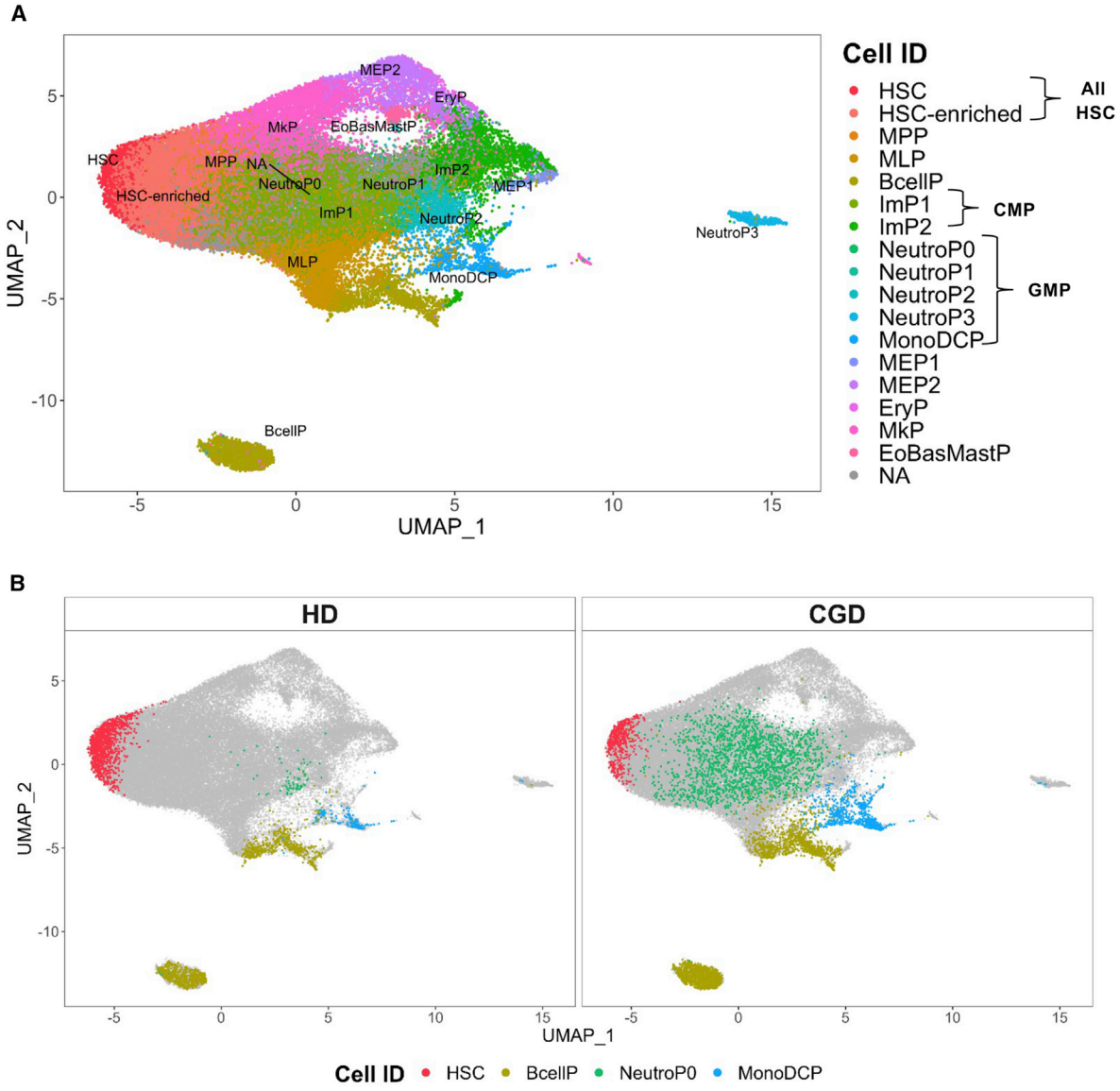
(A) Clinical trial scheme. MPB, mobilized peripheral blood; D, day; IMPs, investigational medicinal products; TD, transduction. *Corticoid treatment only for P1; **PGE2, prostaglandin E2 only for P2, P4, and P5.

(B) The VCN per cell, measured as a guide to changes over time in the level of gene marking in the treated patients' CD15^{+} neutrophils.

(C) Percentage of neutrophils positive for dihydrorhodamine (DHR) 123 after phorbol myristate acetate stimulation, for each patient and at different time points post-GT.

(D) gp91^{phox} membrane expression on neutrophils pre-GT and post-GT (last follow-up) for each patient, as measured with flow cytometry.

(E) An axial chest CT scan (lung window) of P1 pre-GT and 48 months afterward (post-GT). Before GT, P1 presented with large ground-glass opacities, corresponding to pulmonary granulomas. Thoracic MRI (with a T2 fat-saturation sequence) of P4 before GT (repetition time [RT] 14,118 ms; echo time [ET] 95.48 ms) and 3 months after GT (RT 10,610 ms; ET 86.72 ms). For P4, the arrow indicates lesions caused by *Aspergillus* infection.



(legend on next page)

patients or the HDs. In this pathological subpopulation, the top genes (defined using Cell-ID) included the CEBPβ transcription factor (Figure 4B), which is typically expressed more in committed myeloid progenitors. In HD cells, CEBPβ was expressed from the NeutroP3 stage onward. In P5's cells, CEBPβ was expressed in the most immature HSC subpopulation and in early HSC progenitors and the MPP population (Figure 4C). These results suggest strongly that P5 not only presents a large NeutroP0 population but also has a strong alteration in the most immature HSC state, with aberrant expression of the CEBPβ myeloid factor.

In P2, the Cell-ID analysis highlighted a large number of cells with mixed BcellP-MonoDCP signatures (Figure 3). This mixed signature was also detected (albeit to a lesser extent) in HDs and in the other patients (Figure S9A). The gene signature detected with Cell-ID in this mixed BcellP-MonoDCP population evidenced the expression of genes known to have a role in B cell and dendritic cell lineages (e.g., *IRF8*, *SPIB*, *TLR7*, and *TNFRSF17*) and that were not detected in the equivalent population of HDs (in red in Figures S9B and S9C). These results emphasized the closeness of the relationship between the BcellP and the MonoDCP lineages, both of which were unusually prominent in P2.

The interferon pathway score highlights HSC alterations correlated with poor engraftment

To better understand the molecular changes in the most immature HSCs, we used a model-based analysis of single-cell transcriptomics (MAST) to define differentially expressed genes (DEGs) in patients versus HDs.[27] We identified 369 DEGs in the most immature HSC subpopulation and then tested for functional enrichment using the Molecular Signature Database.[28] The IFNα, IFNγ, and TNFα pathways were more prominent in patients with CGD than in HDs (Figure 5A).

Cell-ID allowed us to compare individual cell signatures with well-defined gene sets (such as the Hallmarks collection) associated with particular biological states or processes.[28] The highest IFNγ response scores were found for P2 and P5, especially in most immature HSCs at the apex of the UMAP (red arrow, Figure 5B). In contrast, HSCs in P1 and P4 (the patients with the best correction and engraftment) did not present with significant IFNγ response scores, as for the HDs (Figure 5C). Patient P4 showed a significant IFNγ score but in more committed cells.

To further understand these interindividual differences and DEGs, we also performed a MAST analysis for each individual patient's HSCs. Sixty-one of the DEGs were part of the IFNγ pathway (Figure 5D). The degree of deregulation was higher in P2 and P5 than in P1 and P4, in accordance with the IFN pathway enrichment score, which did not reach significance level in P1 and P4 in the HSC population (Figure 5C). We hypothesize that a higher level of IFNγ activation could lead to HSC exhaustion. To test this hypothesis and to determine which IFNγ pathway together with transcription factor genes might contribute significantly to engraftment failure, we took advantage of the large number of samples provided by the single-cell RNA-seq transcriptomic profiling with 469 individual patient HSCs. Furthermore, we made use of a machine learning approach (elastic-net logistic regression) to predict graft ability for each individual cell[29] (STAR Methods and Figure S10). We used Monte Carlo cross-validation with 50 iterations that demonstrate good predictive power, as shown by the scores for each patient (Figures 5E and S10A–S10D, median accuracy 0.97, median area under the curve 0.99, for 50 cross-validation models). This approach identified a set of 51 IFN genes and transcription factors as being predictive of the engraftment defect in P2 and P5 (Figures 5F, S10E, and S10F). The IFN-stimulated genes (ISGs) included *IFI44L*, *MX1*, *STAT2*, *IRF9*, and SAMD9L, all of which were significantly upregulated in P2's and P5's HSC subpopulation (Figure 5G). In P1 and P4 (patients with successful engraftment), these genes were expressed to the same extent as in HDs or only slightly more. The model also selected predictive transcription factors, which interacted in a functional protein association network (Figure 5F) linking *CEBPB* (already identified in P5) with other factors, such as *JUND*, *SREBF1*, and *MAFG*.

Taken as a whole, these transcriptomic data identified specific biomarkers in CGD HSCs. Elevated inflammatory pathway activity was predictive of poor engraftment.

HSC exhaustion revealed by impaired xenotransplantation of HSPCs from patients with severe CGD

To further understand the changes in HSCs associated with defective engraftment in patients with severe CGD, we evaluated xenotransplantation in a humanized NOD-SCID-$\gamma c^{-/-}$ (NSG) mouse model. The transplanted HSPCs came from P4 and P5, whose IMPs were similar.

Using an aliquot of the patient's IMP, we infused engineered HSPCs into NSG mice (P4, 4.3×10^5 cells; P5, 3.5×10^5 cells; n = 4 mice per patient). As controls, we infused nontransduced cord blood (CB) (2.7×10^5 cells, n = 3) and a sample of MPB

Figure 2. Low HSC frequency and elevated myeloid progenitor frequency, revealed by single-cell HSPC transcriptional mapping

(A) Unsupervised analysis of 53,412 cells from merged $CD34^+$ MPB from four patients with CGD and four HDs, represented as two-dimensional UMAP plots. Each individual cell in our dataset was annotated using the Cell-ID method and reference BM HSPC signatures (Table S1). HSC, hematopoietic stem cell; HSC-enriched, hematopoietic stem cell enriched; MPP, multipotent progenitors; MLP, multipotent lymphoid progenitors; ImP1 and ImP2, immature myeloid progenitors; NeutroP0, NeutroP1, NeutroP2, and NeutroP3, neutrophil progenitors; MonoDCP, monocyte and dendritic cell progenitors; BcellP, B cell progenitors; MEP1 and MEP2, megakaryocyte and erythrocyte progenitors; EryP, erythroid progenitors; MkP, megakaryocyte progenitors; EoBasMastP, eosinophil, basophil, and mast cell progenitors; NA, not annotated.

(B) UMAP plots of 30,225 HSPCs from four HDs (left) and of 23,187 HSPCs from four patients with CGD (right), showing the HSC, BcellP, NeutroP0, and MonoDCP subpopulations.

(C) Bar plots showing the percentages of HSC, BcellP, NeutroP0, and MonoDCP subpopulations in each patient and each HD. The HSC frequency was significantly lower in CGD patients than in the HDs (odds ratio [OR] = 0.53, 95% confidence interval [CI] = 0.33–0.46, $p = 2 \times 10^{-1}$, using logistic regression). We also observed lower HSC frequencies in P2 and P5 than in P1 and P4 (OR = 0.58, 95% CI = 0.48–0.70, $p = 3.78 \times 10^{-8}$, using logistic regression). The frequencies of the other cell types are shown in Figure S7D.

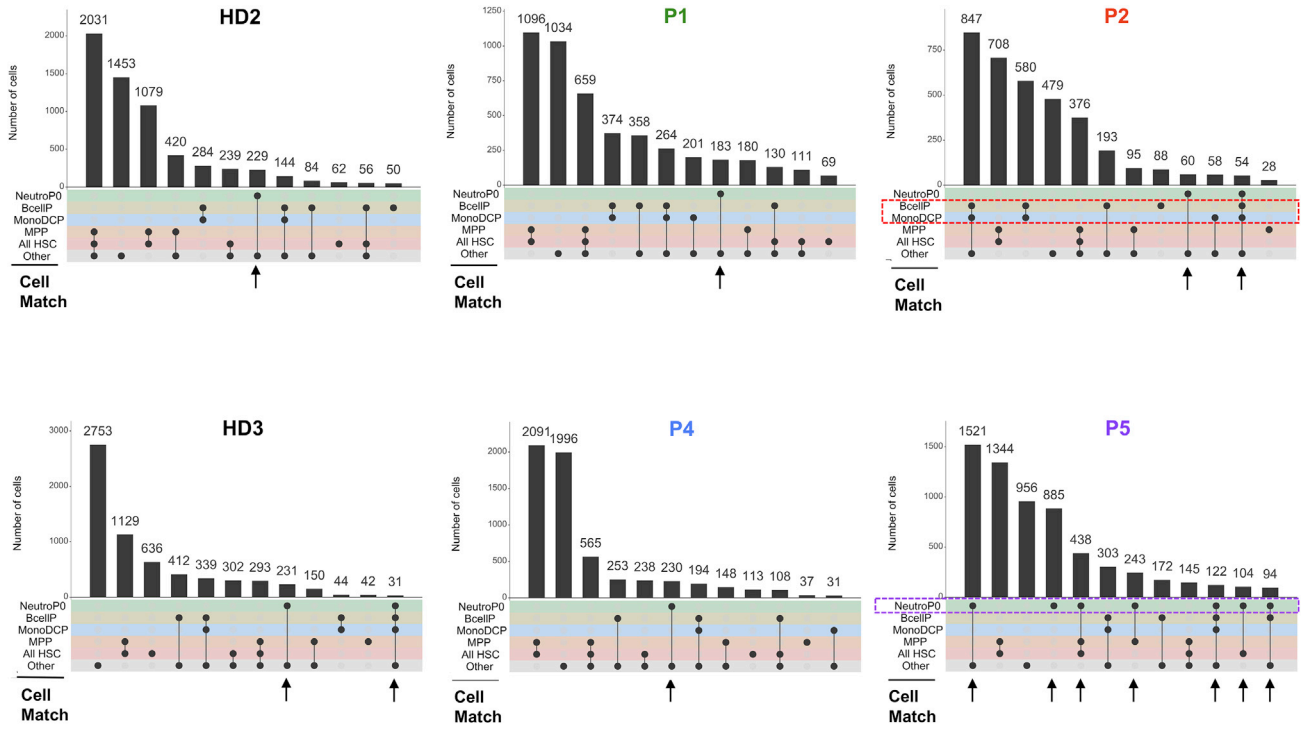


Figure 3. Identification of HSPC subpopulations displaying mixed signatures

UpSet plots showing the number of cells significantly matching one or more cell types (see Figure S8A): All HSC, MPP, MonoDCP, BcellP, and NeutroP0 signatures and Others (i.e., the 11 other HSPC cell-type signatures) are shown. The top 12 lineage combinations are shown for each patient and for two HDs. A full UpSet plot covering the 16 cell types is shown in Figure S8D. The arrows indicate cells identified as simultaneously displaying signatures for NeutroP0 and other cell type(s). P5 presented a larger number of NeutroP0 cells with mixed signatures compared with other patients and HDs. Dotted lines highlight the predominant cell types in P5 (NeutroP0) and P2 (MonoDCP and BcellP).

transduced with the same protocol as for the IMP (3.7×10^5 cells, n = 5). We then analyzed engraftment in the BM and spleen after 16 weeks. The mean level of BM chimerism was 35% in P5 recipients and 60% in P4 recipients; this difference was statistically significant (p = 0.0286; Figure 6A). P5 recipients had a lower absolute human $CD45^+$ (hCD45) cell count than P4 recipients, although the difference did not reach statistical significance. Human BM HSPCs (defined as $CD45^+CD34^+$ cells) were also significantly less frequent in P5 recipients than in P4 recipients (Figure 6B). These results were confirmed by a chimerism analysis of the spleen (Figure 6C). Although P4 and P5 recipients had similar levels of gene correction in the IMP (1.73 and 1.58, respectively), the VCN in hCD45 cells from BM and spleen was significantly lower after transplantation, especially for P5 recipients (a mean value of 0.25 versus 1.05 for P4; p = 0.0286) (Figures 6D and 6E). Similarly, the level of correction (estimated from the $gp91^{phox}$ protein expression by the hCD45 cells) was significantly lower in P5 recipients than in P4 recipients (p = 0.0286, Figure 6F).

These *in vivo* experiments demonstrated that P5's HSPCs had a lower engraftment ability and $gp91^{phox}$ expression than their counterparts from P4; this finding was in line with the corresponding clinical outcomes in the GT trial.

Taken as a whole, our results showed that P5's HSPCs presented with a chronic inflammatory profile and molecular alterations that strongly impaired their functional capacity.

DISCUSSION

Our present results revealed that a severe inflammation score can profoundly alter the HSCs and compromise the effectiveness of GT in patients with CGD. GT remains a potentially curative treatment option for patients with CGD who lack an HLA-compatible donor for HSCT and do not present exacerbated inflammatory markers. We observed the engraftment of gene-corrected cells in two patients (P1 and P4), leading to a complete remission in P4 and to an intermediate level in P1 that was sufficient for clinical benefit. The significant, stable correction of HSPCs has been maintained for more than 4 years now and is correlated with neutrophilic NADPH oxidase activity. In contrast, P2 and P5 progressively lost the corrected cells. This was also observed for four child patients in two other GT clinical trials.[19] The results of an in-depth single-cell transcriptomic analysis in our patients suggested that this defect might be linked to (1) a high inflammation score in the most immature HSCs and (2) the upregulation of specific biomarkers not currently detectable by classical immunophenotyping.

The fact that HLA-identical HSCT gives excellent outcomes in patients with CGD (i.e., low graft failure and mortality rates) suggests that the HSC niche in the BM microenvironment is not significantly altered. In a multicenter study of allo-HSCT in 712 patients with CGD, the estimated overall survival

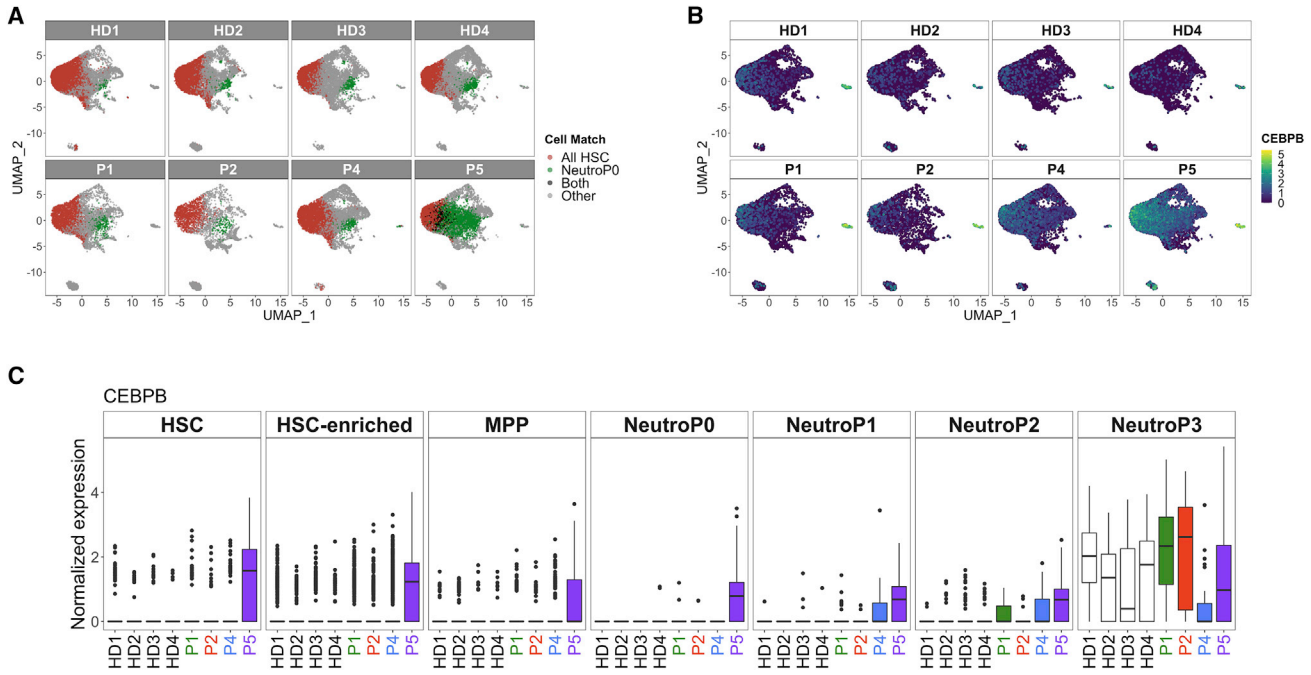


Figure 4. HSCs with an altered state and aberrant *CEBPB* expression

(A) UMAP plots of cells significantly matching the All HSC and NeutroP0 signatures for each HD (n = 4) and CGD patient (n = 4). Cells matching All HSC and NeutroP0 cell types are shown in black, and cells matching with only one cell type are shown in red (All HSC) and green (NeutroP0).

(B) UMAP visualization of *CEBPB* mRNA expression in each HD (n = 4) and CGD patient (n = 4). Normalized expression is represented by a color-coded gradient.

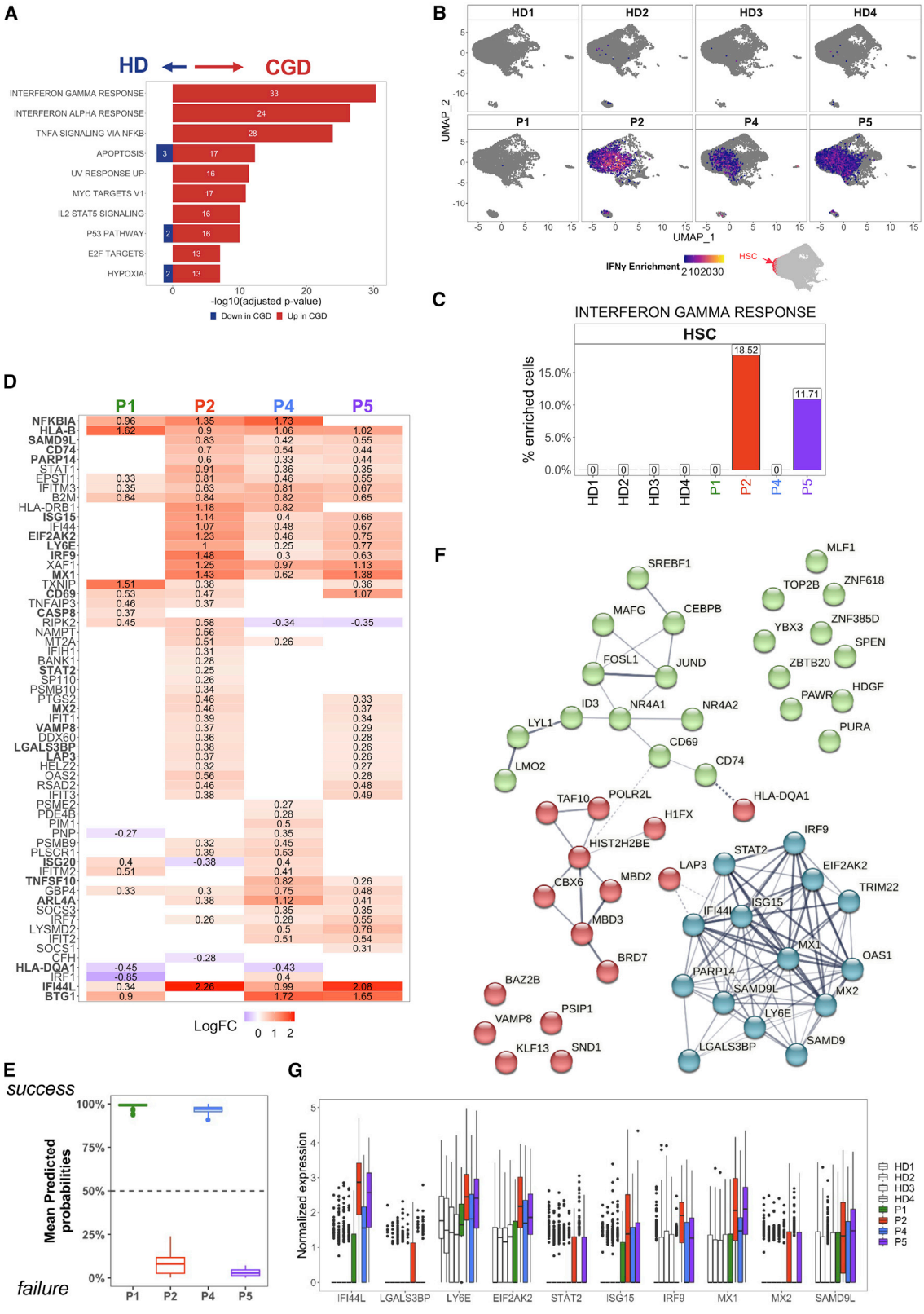
(C) Boxplots of *CEBPB* mRNA expression in the HSC, HSC-enriched, MPP, NeutroP0, NeutroP1, NeutroP2, and NeutroP3 populations, in each HD (n = 4) and CGD patient (n = 4).

and event-free survival rates at 3 years were 85.7% and 75.8%, respectively.[10] These results suggest that the HSC niche is mostly normal and that exposure to chronic inflammation caused an intrinsic HSC alteration. Using single-cell transcriptome profiling, we identified specific inflammatory signatures (including IFNγ and IFNα responses) in CGD HSCs and myeloid progenitors (monocyte, dendritic cell, and neutrophil progenitors). Moreover, the two patients with the highest inflammation scores presented a high frequency of myeloid progenitors and a low frequency of immature HSCs.

Our results are in line with the increased myeloid differentiation observed in response to various inflammatory emergency signals, such as IL-1,[30] IFN, and lipopolysaccharide.[31–33] Indeed, microbial infections and other stimuli (e.g., metabolic stress) can drive HSCs out of dormancy and favor proliferation and myeloid differentiation (facilitating the host's defense). This emergency granulopoiesis is initiated by the key transcription factor CEBPβ.[34] Signaling through G-CSF and STAT3 can induce a switch from CEBPα-dependent steady-state granulopoiesis to CEBPβ-dependent emergency granulopoiesis.[35] The impact of persistent inflammation has also been reported in a mouse model of X-CGD, with greater HSC proliferation and differentiation toward myeloid lineages.[20] P2 displayed an expansion of both MonoDC progenitors and B cell progenitors, which share a number of markers. These findings are reminiscent of the pro-B cell progenitor expansion that occurs after IFN stimulation[36] and the reprogramming of myeloid lineages in a context of inflammation.[37,38]

Even though acute inflammation can be beneficial, it is known that long-lasting, chronic inflammation strongly impairs HSC function. Whereas acute treatment with IFNα can promote the proliferation of murine HSCs, chronic IFNα activation compromises the HSCs' repopulating activity.[39] Several studies have shown that elevated IFN signaling in chronic infection is the primary cause of HSC exhaustion and depletion.[39–43]

A pathway analysis of the most immature HSCs indicated that the higher IFN score in P2 and P5 might be responsible for the loss of gene-corrected cells and for HSC exhaustion. In contrast, the intermediate level of IFN pathway activity in P1 and P4 might have helped to maintain a beneficial response and the HSCs' repopulating ability. By taking advantage of regularized logistic regression and the large number of cells provided by single-cell analyses, we identified a set of 51 IFN genes and transcription factors that were upregulated specifically in P2 and P5 (including *IFI44L*, *STAT2*, *IRF9*, *MX1*, *SAMD9L*, and *CEBPB*) and that appeared to be predictive of defective gene-modified HSC engraftment. These transcriptomic alterations and biomarkers appear to be specific to the HSC compartment, as no significant differences in the inflammation score were observed in PBMCs, while they were detected in the global HSPC population. In view of clinical applicability, the targeted gene expression analysis of the predictive genes on the HSPC population should provide a reliable test to assess HSC fitness before patient enrollment.



(legend on next page)

It has been shown that IFN pathway activation in HSCs involves STAT1 and IRF9 signaling pathways[44] by forming the DNA-binding STAT1-STAT2-IRF9 ternary complex ISGF3, which then activates ISGs.[45] The strong activation of the IFN pathway observed in patients with CGD resulted in marked overexpression of *STAT1*, *STAT2*, and *IRF9* genes, especially in P2. This patient displayed a high frequency of monocyte/dendritic cell progenitors with strong inflammatory profile but also the upregulation of several stress-induced factors (such as *JUND* or *SREBF1*) in HSCs, which might have been responsible for the functional defects.[46,47] This situation was reminiscent of HSC exhaustion through chronic IFN pathway activation.[39] Based on the strong activation of the Jun/Fos pathway, we cannot exclude the synergic contribution of additional inflammatory pathways participating in HSC exhaustion, such as the TLR4/TRIF pathway.[33]

P5 had a large neutrophil progenitor population and aberrant expression of *CEBPB* very early in the HSC differentiation process. G-CSF mobilization is known to induce CEBPβ in myeloid progenitors[34,35] and therefore could contribute to the exacerbated myeloid skewing in patient P5. However, G-CSF did not trigger similar skewing in P1 and P4 and the used of plerixafor alone as the mobilizing agent in P2 was not sufficient to avoid HSC exhaustion, suggesting that the main cause resides in the HSC intrinsic alteration driven by chronic inflammation. The epigenetically inscribed infection history is known to make HSCs more responsive to secondary stimulation.[48] However, chronic lipopolysaccharide stimulation drives HSC exhaustion and dysfunction.[49]

We hypothesize that in P2 and P5, chronic IFN stimulation epigenetically blocked HSCs in an aberrant state and thus drove exhaustion. One of the downstream markers observed in both patients was sterile α motif domain-containing protein 9-like encoded by *SAMD9L*, an ISG-induced gene in which mutations are known to predispose to pancytopenia and myeloid malignancies. Indeed, gain-of-function (GOF) mutations in *SAMD9L* have been reported in people with ataxia pancytopenia syndrome.[50] The antiproliferative effect of these GOF mutations led to greater DNA damage and apoptosis (responsible for BM hypocellularity). A secondary mutation (monosomy 7) would favor the development of myelodysplastic syndromes.[51] Enhanced expression of *SAMD9L* (correlating with the higher IFN scores in P2 and P5) might therefore contribute to HSC exhaustion in a context of chronic IFN activation.

Kohn et al. reported a higher frequency of stable correction and engraftment, for which there are no obvious clues to explain the difference with the present study.[19] Engraftment failure does not seem to correlate with patient age, since in our case it was mainly observed in adult patients. To be noted is their higher frequency of the use of fresh cells compared with cryopreserved cell products that could constitute an additional stress factor for HSCs, already compromised by the chronic inflammation.

The sometimes poor transduction ability in CGD HSPCs also prompted us to optimize the transduction procedure by adding PGE2; this adjuvant is known to favor HSC homing, survival, proliferation, and repopulation ability.[22,52,53] PGE2's pro-inflammatory role during vasodilatation, vascular leakiness, and pain has long been known,[54] but this compound can also mediate anti-inflammatory effects,[55] as also shown in our transcriptomic analysis. Despite its indubitable benefit, this short course of PGE2 was not enough to counter the HSC alteration induced by chronic inflammation in P2 and P5. Moreover, PGE2 does not completely restore transduction efficiency, which is lower than in HDs, probably due to the upregulation of genes encoding restriction factors like *MX1*, *MX2*, and *IFITM3*.[56,57] The expression of these factors by HSCs during an innate immune response inhibited lentiviral entry but could be overcome by exposure to cyclosporine H[58] or other transduction enhancers that are currently being investigated. This aspect might be important in the further development of GT in the context of inflammatory diseases.

Chronic inflammation in CGD might eventually favor the emergence of mutated clones with a proliferative advantage; in turn, this might lead to tumor events[59] and so further highlights the need to control hyperinflammation.

The impaired repopulating ability of CGD HSCs has been previously reported in a mouse model of X-CGD exposed to a high IL-1 concentration. Pre-treatment of X-CGD mice with anakinra (an IL-1R antagonist) improves HSC engraftment.[20] More recently, p38MAPK (a downstream target of IL-1β) was identified in a

Figure 5. An elastic-net model identified the inflammatory genes in HSCs that are predictive of engraftment failure after GT

(A) MAST identified 369 DEGs in the HSCs in CGD patients (n = 4) versus HDs (n = 4). The top 10 pathways (in terms of p value, identified using a hypergeometric test and MSigDB and Hallmark gene sets) among the 369 DEGs in the HSCs are shown. In each pathway, genes that are upregulated in CGD (relative to HDs) are shown in red, and those that are downregulated in CGD are shown in blue. The false discovery rate (−log10 adjusted p value) is shown for each pathway. The numbers of upregulated and downregulated genes in each pathway are also shown.

(B) UMAP plot of the interferon γ response pathway enrichment score for each HD (n = 4) and CGD patient (n = 4), determined with Cell-ID for each cell (see STAR Methods). Red arrow indicates the HSC at the apex of the UMAP.

(C) Percentage of HSCs presenting a significant interferon γ enrichment score ($p < 0.01$).

(D) MAST identified 1,136 DEGs in the HSCs in each individual CGD patient versus HDs. Heatmap showing the 61 DEGs in the Hallmark interferon γ response signature. Genes identified as predictors in the following model are shown in bold (see Figure S10). The color code shows the logFC for each gene in a patient's HSCs (versus HDs).

(E) Boxplot showing predicted engraftment scores per patient (elastic-net model, see Figure S10), corresponding to the distribution of the 50 mean probabilities per patient for HSCs (test datasets). A probability <50% corresponds to a prediction of engraftment failure, and a probability >50% corresponds to a prediction of engraftment success.

(F) The network of predicted interactions between the 51 ISGs and transcription factors in HSCs selected by the elastic-net model as being predictive of the engraftment failure observed in P2 and P5 (see also Figures S10E and S10F). The network was generated using StringDB and clustered with k-means (k = 3); dotted lines show the edges between clusters.

(G) Boxplot of the expression by HSCs of 10 representative ISGs (*IFI44L*, *LGALS3BP*, *LY6E*, *EIF2AK2*, *STAT2*, *ISG15*, *IRF9*, *MX1*, *MX2*, and *SAMD9L*) identified by the elastic-net model as being predictive of engraftment failure in P2 and P5 and shown for each HD (n = 4) and CGD patient (n = 4).

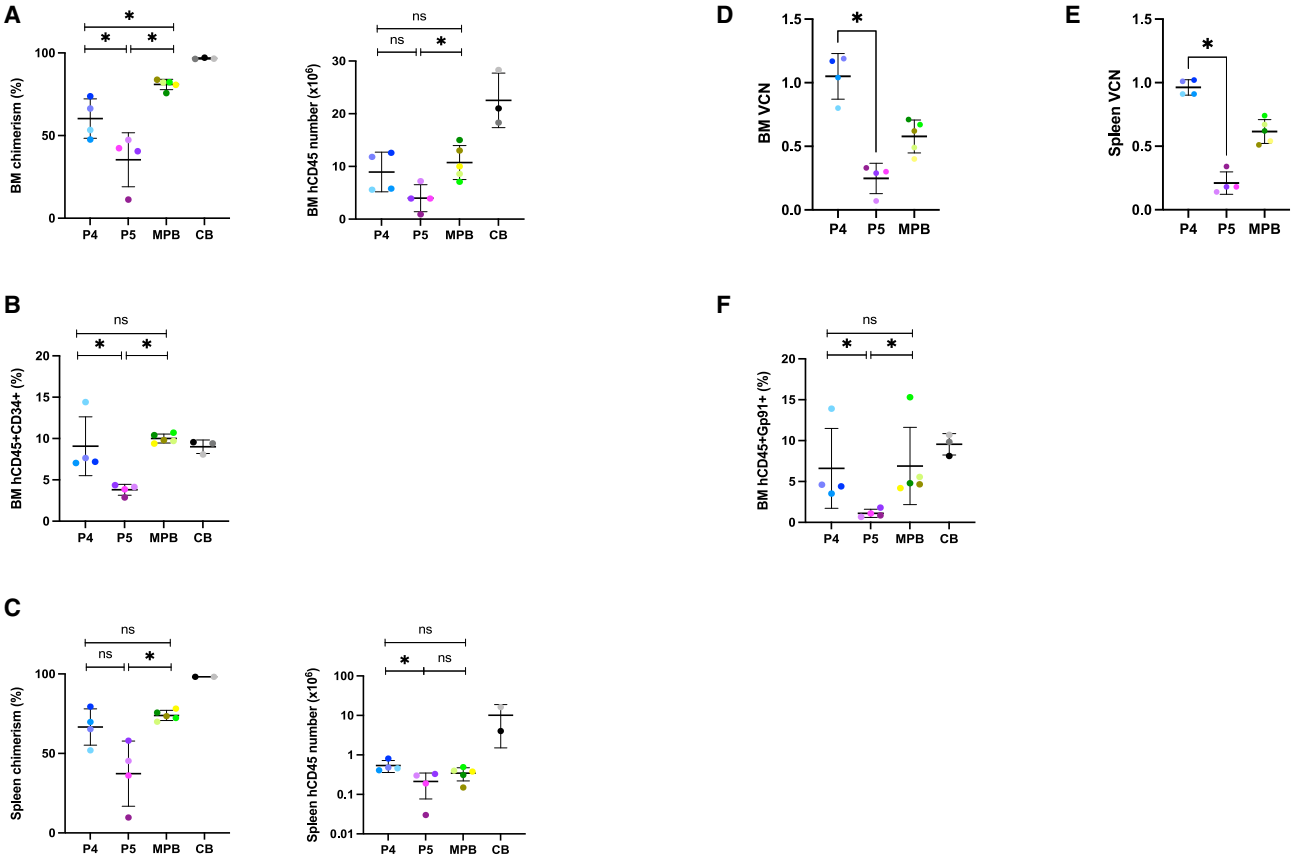


Figure 6. Xenotransplantation of the patients' corrected HSPCs in a humanized mouse model

HSPCs from P4's and P5's IMPs were infused into NOD-SCID-$\gamma c^{-/-}$ (NSG) mice (n = 4 per group, VCN_{P4} = 1.7, VCN_{P5} = 1.6). Nontransduced CB samples (n = 3) and transduced MPB samples (n = 5, VCN_{MPB} = 3.47) were used as controls, with the same (clinical) *ex vivo* cell engineering protocol. Mean ± SD. Engraftments in the BM and spleen were analyzed 16 weeks after transplantation. Mann-Whitney test was performed for the statistical analysis; ns, not significant; *p < 0.05.

(A) Left, human chimerism (% $hCD45^+$/[$hCD45^+$ + $mCD45^+$]) in the BM 16 weeks after transplantation (P4, n = 4; P5, n = 4; MPB, n = 5; CB, n = 3). The level of chimerism was significantly lower in P5 than in P4 (p = 0.0286), also in comparison with P5/MPB (p = 0.0159) and P4/MPB (p = 0.0159). Right, the number of human $CD45^+$ cells in the BM was significantly lower in P5 than in MPB (p = 0.0317).

(B) Frequency of human $CD45^+CD34^+$ cells in the BM. The frequency was significantly lower in P5 than in P4 (p = 0.0286) and also in comparison with P5/MPB (p = 0.0159).

(C) Left, human chimerism (% $hCD45^+$/[$hCD45^+$ + $mCD45^+$]) in the spleen 16 weeks after transplantation, in the same mice. The level of chimerism was significantly lower in P5 than in MPB (p = 0.0159). Right, the number of human $CD45^+$ cells in the spleen. The level of chimerism was significantly lower in P5 than in P4 (p = 0.0286).

(D) The VCN per cell was measured to assess the level of gene marking in total BM, using a droplet digital PCR (ddPCR) technique. The VCN was significantly lower in P5 than in P4 (p = 0.0286).

(E) The VCN per cell was measured to assess the level of gene marking in the total spleen, using a ddPCR technique. The VCN was significantly lower in P5 than in P4 (p = 0.0286).

(F) The frequency of $gp91^{phox+}$ cells in human $CD45^+$ BM. The frequency was significantly lower in P5 than in P4 (p = 0.0286) and also in comparison with P5/MPB (p = 0.0159).

CRISPR-Cas9 screening step as a druggable target for increasing HSC engraftment. *Ex vivo* culture of CGD HSPCs in the presence of a p38MAPK inhibitor increased chimerism significantly (1.5-fold).[60]

Inhibition of the JAK/STAT pathway would be another way to target the hyperactivated IFN pathway. Given that several studies have described encouraging results for JAK1 inhibition in type I interferonopathies,[61–63] this approach could also be considered for controlling inflammation before HSPC harvesting in patients with CGD and thus avoiding HSC exhaustion. If *ex vivo* treatments are enough to improve HSC engraftment rate or if we need to control *in vivo* the inflammation before HSPC harvesting is now under deep investigation.

Together with the results of a recently published study on GT for CGD, our present findings show that GT is a potentially curative treatment option in patients with CGD lacking an HLA-compatible donor. However, the specific clinical and cellular characteristics of good candidates for GT (notably with regard to the inflammatory background) need to be taken into account. Our present study identified an IFN-pathway-related transcriptional signature that was specific to HSCs from patients with poor engraftment. The

present results might open the way to (1) specific anti-inflammatory treatments for patients prior to HSPC harvesting, (2) the optimization of *ex vivo* HSPC engineering, and (3) identification of predictive biomarkers for validating the GT product prior to infusion.

Limitations of the study

Statistical power to understand the loss of gene-corrected cells has been obtained thanks to single-cell RNA-seq and machine learning approaches on hundreds of individual cells. Still, exploring the efficiency of new therapeutics in rare immune deficiencies remains highly challenging due to low patient sample size.

The 51 identified predictive markers will be useful for tracking patient HSPC status. Further HSPC transcriptomic profiling on additional CGD patients along with immunosuppressive treatments (such as JAK inhibitors) will provide confirmation of the most relevant biomarkers.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Study design and investigational therapy
  - Patients
  - Healthy donors
  - Mouse experiments and xenotransplantation assays
- METHOD DETAILS
  - G1XCGD lentiviral vector production
  - G1XCGD lentiviral vector-modified $CD34^+$ cell product manufacture
  - Determination of the VCN
  - The DHR assay
  - Isolation of mononuclear cells
  - Flow cytometry
  - Analysis of vector integration sites
  - Bulk RNA-seq
  - Single-HSPC RNA-seq
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Single-HSPC RNA-seq
  - Other statistical analysis
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrm.2023.100919.

ACKNOWLEDGMENTS

The study was sponsored by Généthon (Evry-Courcouronnes, France) and funded by grants from the European Union's FP7 (NET4CGD, project 305011), the Dior Chair for Tailored Medicine (to M.C. and A.R.), the DIM-Thérapie Génique/Région Ile-de-France (INFLAGT) program (to E.S.), the Fondation pour la Recherche Médicale (FDT202106012898 to S. Sobrino), and the French state (the Agence Nationale de la Recherche's "Investissements d'Avenir" program, ANR-10-IAHU-01). We thank the patients and their families for their cooperation during the study. We are grateful to M. Granier and colleagues from the Clinical Development Department of Généthon, as well as V. Jolaine and J. Marouene (respectively project manager and study coordinator at Assistance Publique-Hôpitaux de Paris/URC Paris Center) for study implementation, monitoring, and data management. We thank the Center of Clinical Investigation (CIC-NEM) for their support in the follow-up of the patients, as well as M. Roelens from the Centre Etudes des Déficits Immunitaires (CEDI). We are grateful to members of the I. André laboratory, especially H. Vinçon, T. Blein, and Y. Aslan, for help with the animal experiments and I. André for her advice and support. We thank P. Revy for critical discussion of the results, M. Guiot for help with the oxidative burst assay, M. Lucas and M. Menager at the Imagine Institute's single-cell analysis facility, C. Bole and colleagues at the Imagine Institute's genomics facility, N. Cagnard and C. Masson at the Imagine Institute's bioinformatics facility, E. Panafieu and colleagues at the Laboratoire d'Expérimentation Animale et Transgénèse at the Structure Fédérative de Recherche Necker, and J. Mégret from the SFR Necker cytometry facility.

AUTHOR CONTRIBUTIONS

S.S. designed and performed the experiments and analyzed the results. A. Magnani, A. Galy, A.R., S. Blanche, M.C., and E.S. supervised and funded the research, designed the study, and collected and analyzed the data. A. Magnani, M.S., C. Couzin, C.P., J.B., E.M., L.J., O.L., F.L., F.R.L., J.L.C., S. Bodard, N.B., F.T., S.T., F.S., A. Marcais, C.L.P., S. Blanche, and M.C. performed the transplantations, provided patient and family care, and/or monitored the patients and performed analysis on patient samples. L.M., A.C., A.D., T.S., A. Guilloux, A.R., and E.S. performed the transcriptomic and statistical analysis. C. Cordier performed cytometry analysis on patient samples. S.S., A. Magnani, M.C., and E.S. primarily wrote the manuscript and all authors critically revised the manuscript.

DECLARATION OF INTERESTS

S.S., L.M., A.C., A.D., A. Guilloux, A.R., M.C., and E.S. have submitted a patent related to the use of the predictive markers for assessing HSC exhaustion induced by chronic inflammation (EP22305722.5 filed on May 16, 2022).

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: July 6, 2022
Revised: October 20, 2022
Accepted: January 6, 2023
Published: January 26, 2023

REFERENCES

1. Bousfiha, A., Jeddane, L., Picard, C., Al-Herz, W., Ailal, F., Chatila, T., Cunningham-Rundles, C., Etzioni, A., Franco, J.L., Holland, S.M., et al. (2020). Human inborn errors of immunity: 2019 update of the IUIS phenotypical classification. J. Clin. Immunol. *40*, 66–81. https://doi.org/10.1007/S10875-020-00758-X.
2. Tangye, S.G., Al-Herz, W., Bousfiha, A., Chatila, T., Cunningham-Rundles, C., Etzioni, A., Franco, J.L., Holland, S.M., Klein, C., Morio, T., et al. (2020). Human inborn errors of immunity: 2019 update on the classification from the international union of immunological societies expert committee. J. Clin. Immunol. *40*, 24–64. https://doi.org/10.1007/S10875-019-00737-X.
3. Roos, D., Tool, A.T.J., van Leeuwen, K., and de Boer, M. (2017). Biochemical and genetic diagnosis of chronic granulomatous disease. Immunology and immune system disorders. In Chronic granulomatous disease

genetics, biology and clinical management, R.A. Seger, D. Roos, H. Segal, and T.W. Kuijpers, eds. (Nova Science Publishers), pp. 231–300.

4. Roos, D. (2016). Chronic granulomatous disease. Br. Med. Bull. *118*, 50–63. https://doi.org/10.1093/BMB/LDW009.
5. van den Berg, J.M., van Koppen, E., Ahlin, A., Belohradsky, B.H., Bernatowska, E., Corbeel, L., Español, T., Fischer, A., Kurenko-Deptuch, M., Mouy, R., et al. (2009). Chronic granulomatous disease: the European experience. PLoS One *4*, e5234. https://doi.org/10.1371/JOURNAL.PONE.0005234.
6. Marciano, B.E., Spalding, C., Fitzgerald, A., Mann, D., Brown, T., Osgood, S., Yockey, L., Darnell, D.N., Barnhart, L., Daub, J., et al. (2015). Common severe infections in chronic granulomatous disease. Clin. Infect. Dis. *60*, 1176–1183. https://doi.org/10.1093/CID/CIU1154.
7. Magnani, A., Brosselin, P., Beauté, J., de Vergnes, N., Mouy, R., Debré, M., Suarez, F., Hermine, O., Lortholary, O., Blanche, S., et al. (2014). Inflammatory manifestations in a single-center cohort of patients with chronic granulomatous disease. J. Allergy Clin. Immunol. *134*, 655–662.e8. https://doi.org/10.1016/J.JACI.2014.04.014.
8. Marciano, B.E., Segal, B.H., and Holland, S.M. (2017). Inflammatory manifestations in chronic granulomatous disease. In Chronic Granulomatous Disease. Genetics, Biology And Clinical Management, R.A. Seger, D. Roos, H. Segal, and T.W. Kuijpers, eds. (Nova Biomedical Press), pp. 163–201.
9. van de Veerdonk, F.L., and Dinauer, M.C. (2017). Dysfunctional processes regulating IL-1B in chronic granulomatous disease. In Chronic Granulomatous Disease. Geneticcs, Biology And Clinical Management, R.A. Seger, D. Roos, H. Segal, and T.W. Kuijpers, eds. (Nova Biomedical Press), pp. 69–87.
10. Chiesa, R., Wang, J., Blok, H.J., Hazelaar, S., Neven, B., Moshous, D., Schulz, A., Hoenig, M., Hauck, F., Al Seraihy, A., et al. (2020). Hematopoietic cell transplantation in chronic granulomatous disease: a study of 712 children and adults. Blood *136*, 1201–1211. https://doi.org/10.1182/BLOOD.2020005590.
11. Cole, T., Pearce, M.S., Cant, A.J., Cale, C.M., Goldblatt, D., and Gennery, A.R. (2013). Clinical outcome in children with chronic granulomatous disease managed conservatively or with hematopoietic stem cell transplantation. J. Allergy Clin. Immunol. *132*, 1150–1155. https://doi.org/10.1016/J.JACI.2013.05.031.
12. Gennery, A.R., Slatter, M.A., Grandin, L., Taupin, P., Cant, A.J., Veys, P., Amrolia, P.J., Gaspar, H.B., Davies, E.G., Friedrich, W., et al. (2010). Transplantation of hematopoietic stem cells and long-term survival for primary immunodeficiencies in Europe: entering a new century, do we do better? J. Allergy Clin. Immunol. *126*, 602–610.e1-11. https://doi.org/10.1016/J.JACI.2010.06.015.
13. Martinez, C.A., Shah, S., Shearer, W.T., Rosenblatt, H.M., Paul, M.E., Chinen, J., Leung, K.S., Kennedy-Nasser, A., Brenner, M.K., Heslop, H.E., et al. (2012). Excellent survival after sibling or unrelated donor stem cell transplantation for chronic granulomatous disease. J. Allergy Clin. Immunol. *129*, 176–183. https://doi.org/10.1016/J.JACI.2011.10.005.
14. Grez, M., Reichenbach, J., Schwäble, J., Seger, R., Dinauer, M.C., and Thrasher, A.J. (2011). Gene therapy of chronic granulomatous disease: the engraftment dilemma. Mol. Ther. *19*, 28–35. https://doi.org/10.1038/mt.2010.232.
15. Malech, H.L., Maples, P.B., Whiting-Theobald, N., Linton, G.F., Sekhsaria, S., Vowells, S.J., Li, F., Miller, J.A., Decarlo, E., Holland, S.M., et al. (1997). Prolonged production of NADPH oxidase-corrected granulocytes after gene therapy of chronic granulomatous disease. Proc. Natl. Acad. Sci. USA *94*, 12133–12138. https://doi.org/10.1073/PNAS.94.22.12133.
16. Ott, M.G., Schmidt, M., Schwarzwaelder, K., Stein, S., Siler, U., Koehl, U., Glimm, H., Kühlcke, K., Schilz, A., Kunkel, H., et al. (2006). Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. Nat. Med. *12*, 401–409. https://doi.org/10.1038/NM1393.
17. Stein, S., Ott, M.G., Schultze-Strasser, S., Jauch, A., Burwinkel, B., Kinner, A., Schmidt, M., Krämer, A., Schwäble, J., Glimm, H., et al. (2010). Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease. Nat. Med. *16*, 198–204. https://doi.org/10.1038/NM.2088.
18. Brendel, C., Rothe, M., Santilli, G., Charrier, S., Stein, S., Kunkel, H., Abriss, D., Müller-Kuller, U., Gaspar, B., Modlich, U., et al. (2018). Non-clinical efficacy and safety studies on G1XCGD, a lentiviral vector for ex vivo gene therapy of X-linked chronic granulomatous disease. Hum. Gene Ther. Clin. Dev. *29*, 69–79. https://doi.org/10.1089/humc.2017.245.
19. Kohn, D.B., Booth, C., Kang, E.M., Pai, S.Y., Shaw, K.L., Santilli, G., Armant, M., Buckland, K.F., Choi, U., de Ravin, S.S., et al. (2020). Lentiviral gene therapy for X-linked chronic granulomatous disease. Nat. Med. *26*, 200–206. https://doi.org/10.1038/S41591-019-0735-5.
20. Weisser, M., Demel, U.M., Stein, S., Chen-Wichmann, L., Touzot, F., Santilli, G., Sujer, S., Brendel, C., Siler, U., Cavazzana, M., et al. (2016). Hyperinflammation in patients with chronic granulomatous disease leads to impairment of hematopoietic stem cell functions. J. Allergy Clin. Immunol. *138*, 219–228.e9. https://doi.org/10.1016/j.jaci.2015.11.028.
21. Meda Spaccamela, V., Valencia, R.G., Pastukhov, O., Duppenthaler, A., Dettmer, M.S., Erb, J., Steiner, U.C., Hillinger, S., Speckmann, C., Ehl, S., et al. (2019). High levels of IL-18 and IFN-γ in chronically inflamed tissue in chronic granulomatous disease. Front. Immunol. *10*, 2236. https://doi.org/10.3389/FIMMU.2019.02236.
22. Zonari, E., Desantis, G., Petrillo, C., Boccalatte, F.E., Lidonnici, M.R., Kajaste-Rudnitski, A., Aiuti, A., Ferrari, G., Naldini, L., and Gentner, B. (2017). Efficient ex vivo engineering and expansion of highly purified human hematopoietic stem and progenitor cell populations for gene therapy. Stem Cell Rep. *8*, 977–990. https://doi.org/10.1016/J.STEMCR.2017.02.010.
23. Martignetti, L., Calzone, L., Bonnet, E., Barillot, E., and Zinovyev, A. (2016). ROMA: representation and quantification of module activity from target expression data. Front. Genet. *7*, 18. https://doi.org/10.3389/FGENE.2016.00018.
24. Kiselev, V.Y., Andrews, T.S., and Hemberg, M. (2019). Challenges in unsupervised clustering of single-cell RNA-seq data. Nat. Rev. Genet. *20*, 273–282. https://doi.org/10.1038/S41576-018-0088-9.
25. Cortal, A., Martignetti, L., Six, E., and Rausell, A. (2021). Gene signature extraction and cell identity recognition at the single-cell level with Cell-ID. Nat. Biotechnol. *39*, 1095–1102. https://doi.org/10.1038/S41587-021-00896-6.
26. Velten, L., Haas, S.F., Raffel, S., Blaszkiewicz, S., Islam, S., Hennig, B.P., Hirche, C., Lutz, C., Buss, E.C., Nowak, D., et al. (2017). Human haematopoietic stem cell lineage commitment is a continuous process. Nat. Cell Biol. *19*, 271–281. https://doi.org/10.1038/NCB3493.
27. Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K., Miller, H.W., McElrath, M.J., Prlic, M., et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol. *16*, 278. https://doi.org/10.1186/S13059-015-0844-5.
28. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425. https://doi.org/10.1016/J.CELS.2015.12.004.
29. Torang, A., Gupta, P., and Klinke, D.J. (2019). An elastic-net logistic regression approach to generate classifiers and gene signatures for types of immune cells and T helper cell subsets. BMC Bioinf. *20*, 433. https://doi.org/10.1186/S12859-019-2994-Z.
30. Pietras, E.M., Mirantes-Barbeito, C., Fong, S., Loeffler, D., Kovtonyuk, L.v., Zhang, S., Lakshminarasimhan, R., Chin, C.P., Techner, J.M., Will, B., et al. (2016). Chronic interleukin-1 exposure drives haematopoietic stem cells towards precocious myeloid differentiation at the expense of self-renewal. Nat. Cell Biol. *18*, 607–618. https://doi.org/10.1038/NCB3346.

31. Chen, C., Liu, Y., Liu, Y., and Zheng, P. (2010). Mammalian target of rapamycin activation underlies HSC defects in autoimmune disease and inflammation in mice. J. Clin. Invest. *120*, 4091–4101. https://doi.org/10.1172/JCI43873.

32. Takizawa, H., Fritsch, K., Kovtonyuk, L.v., Saito, Y., Yakkala, C., Jacobs, K., Ahuja, A.K., Lopes, M., Hausmann, A., Hardt, W.D., et al. (2017). Pathogen-induced TLR4-TRIF innate immune signaling in hematopoietic stem cells promotes proliferation but reduces competitive fitness. Cell Stem Cell *21*, 225–240.e5. https://doi.org/10.1016/J.STEM.2017.06.013.

33. Zhang, H., Rodriguez, S., Wang, L., Wang, S., Serezani, H., Kapur, R., Cardoso, A.A., and Carlesso, N. (2016). Sepsis induces hematopoietic stem cell exhaustion and myelosuppression through distinct contributions of TRIF and MYD88. Stem Cell Rep. *6*, 940–956. https://doi.org/10.1016/J.STEMCR.2016.05.002.

34. Hirai, H., Zhang, P., Dayaram, T., Hetherington, C.J., Mizuno, S.I., Imanishi, J., Akashi, K., and Tenen, D.G. (2006). C/EBPbeta is required for "emergency" granulopoiesis. Nat. Immunol. *7*, 732–739. https://doi.org/10.1038/NI1354.

35. Manz, M.G., and Boettcher, S. (2014). Emergency granulopoiesis. Nat. Rev. Immunol. *14*, 302–314. https://doi.org/10.1038/NRI3660.

36. Montandon, R., Korniotis, S., Layseca-Espinosa, E., Gras, C., Mégret, J., Ezine, S., Dy, M., and Zavala, F. (2013). Innate pro-B-cell progenitors protect against type 1 diabetes by regulating autoimmune effector T cells. Proc. Natl. Acad. Sci. USA *110*, E2199–E2208. https://doi.org/10.1073/PNAS.1222446110.

37. Xie, H., Ye, M., Feng, R., and Graf, T. (2004). Stepwise reprogramming of B cells into macrophages. Cell *117*, 663–676. https://doi.org/10.1016/S0092-8674(04)00419-2.

38. Audzevich, T., Bashford-Rogers, R., Mabbott, N.A., Frampton, D., Freeman, T.C., Potocnik, A., Kellam, P., and Gilroy, D.W. (2017). Pre/pro-B cells generate macrophage populations during homeostasis and inflammation. Proc. Natl. Acad. Sci. USA *114*, E3954–E3963. https://doi.org/10.1073/PNAS.1616417114.

39. Essers, M.A.G., Offner, S., Blanco-Bose, W.E., Waibler, Z., Kalinke, U., Duchosal, M.A., and Trumpp, A. (2009). IFNalpha activates dormant haematopoietic stem cells in vivo. Nature *458*, 904–908. https://doi.org/10.1038/NATURE07815.

40. Matatall, K.A., Jeong, M., Chen, S., Sun, D., Chen, F., Mo, Q., Kimmel, M., and King, K.Y. (2016). Chronic infection depletes hematopoietic stem cells through stress-induced terminal differentiation. Cell Rep. *17*, 2584–2595. https://doi.org/10.1016/J.CELREP.2016.11.031.

41. King, K.Y., Baldridge, M.T., Weksberg, D.C., Chambers, S.M., Lukov, G.L., Wu, S., Boles, N.C., Jung, S.Y., Qin, J., Liu, D., et al. (2011). Irgm1 protects hematopoietic stem cells by negative regulation of IFN signaling. Blood *118*, 1525–1533. https://doi.org/10.1182/BLOOD-2011-01-328682.

42. Sato, T., Onai, N., Yoshihara, H., Arai, F., Suda, T., and Ohteki, T. (2009). Interferon regulatory factor-2 protects quiescent hematopoietic stem cells from type I interferon-dependent exhaustion. Nat. Med. *15*, 696–700. https://doi.org/10.1038/NM.1973.

43. Lin, F.C., Karwan, M., Saleh, B., Hodge, D.L., Chan, T., Boelte, K.C., Keller, J.R., and Young, H.A. (2014). IFN-γ causes aplastic anemia by altering hematopoietic stem/progenitor cell composition and disrupting lineage differentiation. Blood *124*, 3699–3708. https://doi.org/10.1182/BLOOD-2014-01-549527.

44. Baldridge, M.T., King, K.Y., Boles, N.C., Weksberg, D.C., and Goodell, M.A. (2010). Quiescent haematopoietic stem cells are activated by IFN-gamma in response to chronic infection. Nature *465*, 793–797. https://doi.org/10.1038/NATURE09135.

45. Crow, Y.J., and Stetson, D.B. (2022). The type I interferonopathies: 10 years on. Nat. Rev. Immunol. *22*, 471–483. https://doi.org/10.1038/S41577-021-00633-9.

46. Roy, A., Wang, G., Iskander, D., O'Byrne, S., Elliott, N., O'Sullivan, J., Buck, G., Heuston, E.F., Wen, W.X., Meira, A.R., et al. (2021). Transitions in lineage specification and gene regulatory networks in hematopoietic stem/progenitor cells over human development. Cell Rep. *36*, 109698. https://doi.org/10.1016/J.CELREP.2021.109698.

47. Lu, Y., Zhang, Z., Wang, S., Qi, Y., Chen, F., Xu, Y., Shen, M., Chen, M., Chen, N., Yang, L., et al. (2022). Srebf1c preserves hematopoietic stem cell function and survival as a switch of mitochondrial metabolism. Stem Cell Rep. *17*, 599–615. https://doi.org/10.1016/J.STEMCR.2022.01.011.

48. de Laval, B., Maurizio, J., Kandalla, P.K., Brisou, G., Simonnet, L., Huber, C., Gimenez, G., Matcovitch-Natan, O., Reinhardt, S., David, E., et al. (2020). C/EBPβ-Dependent epigenetic memory induces trained immunity in hematopoietic stem cells. Cell Stem Cell *26*, 793. https://doi.org/10.1016/j.stem.2020.03.014.

49. Esplin, B.L., Shimazu, T., Welner, R.S., Garrett, K.P., Nie, L., Zhang, Q., Humphrey, M.B., Yang, Q., Borghesi, L.A., and Kincade, P.W. (2011). Chronic exposure to a TLR ligand injures hematopoietic stem cells. J. Immunol. *186*, 5367–5375. https://doi.org/10.4049/JIMMUNOL.1003438.

50. Davidsson, J., Puschmann, A., Tedgård, U., Bryder, D., Nilsson, L., and Cammenga, J. (2018). SAMD9 and SAMD9L in inherited predisposition to ataxia, pancytopenia, and myeloid malignancies. Leukemia *32*, 1106–1115. https://doi.org/10.1038/S41375-018-0074-4.

51. Thomas, M.E., Abdelhamed, S., Hiltenbrand, R., Schwartz, J.R., Sakurada, S.M., Walsh, M., Song, G., Ma, J., Pruett-Miller, S.M., and Klco, J.M. (2021). Pediatric MDS and bone marrow failure-associated germline mutations in SAMD9 and SAMD9L impair multiple pathways in primary hematopoietic cells. Leukemia *35*, 3232–3244. https://doi.org/10.1038/S41375-021-01212-6.

52. Hoggatt, J., Singh, P., Sampath, J., and Pelus, L.M. (2009). Prostaglandin E2 enhances hematopoietic stem cell homing, survival, and proliferation. Blood *113*, 5444–5455. https://doi.org/10.1182/BLOOD-2009-01-201335.

53. North, T.E., Goessling, W., Walkley, C.R., Lengerke, C., Kopani, K.R., Lord, A.M., Weber, G.J., Bowman, T.v., Jang, I.H., Grosser, T., et al. (2007). Prostaglandin E2 regulates vertebrate haematopoietic stem cell homeostasis. Nature *447*, 1007–1011. https://doi.org/10.1038/nature05883.

54. Chace, J.H., Fleming, A.L., Gordon, J.A., Perandones, C.E., and Cowdery, J.S. (1995). Regulation of differentiation of peritoneal B-1a (CD5+) B cells. Activated peritoneal macrophages release prostaglandin E2, which inhibits IgM secretion by peritoneal B-1a cells. J. Immunol. *154*, 5630–5636.

55. Weissmann, G. (1993). Prostaglandins as modulators rather than mediators of inflammation. J. Lipid Mediat. *6*, 275–286.

56. Goujon, C., Moncorgé, O., Bauby, H., Doyle, T., Ward, C.C., Schaller, T., Hué, S., Barclay, W.S., Schulz, R., and Malim, M.H. (2013). Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. Nature *502*, 559–562. https://doi.org/10.1038/NATURE12542.

57. Colomer-Lluch, M., Ruiz, A., Moris, A., and Prado, J.G. (2018). Restriction factors: from intrinsic viral restriction to shaping cellular immunity against HIV-1. Front. Immunol. *9*, 2876. https://doi.org/10.3389/FIMMU.2018.02876.

58. Petrillo, C., Thorne, L.G., Unali, G., Schiroli, G., Giordano, A.M.S., Piras, F., Cuccovillo, I., Petit, S.J., Ahsan, F., Noursadeghi, M., et al. (2018). Cyclosporine H overcomes innate immune restrictions to improve lentiviral transduction and gene editing in human hematopoietic stem cells. Cell Stem Cell *23*, 820–832.e9. https://doi.org/10.1016/J.STEM.2018.10.008.

59. Jofra Hernández, R., Calabria, A., Sanvito, F., de Mattia, F., Farinelli, G., Scala, S., Visigalli, I., Carriglio, N., de Simone, M., Vezzoli, M., et al. (2021). Hematopoietic tumors in a mouse model of X-linked chronic granulomatous disease after lentiviral vector-mediated gene therapy. Mol. Ther. *29*, 86–102. https://doi.org/10.1016/J.YMTHE.2020.09.030.

60. Klatt, D., Ha, T.C., Schinke, M., Selich, A., Lieske, A., Dahlke, J., Morgan, M., Maetzig, T., and Schambach, A. (2020). Competitive sgRNA screen identifies p38 MAPK as a druggable target to improve HSPC engraftment. Cells *9*, 2194. https://doi.org/10.3390/CELLS9102194.

61. Vanderver, A., Adang, L., Gavazzi, F., McDonald, K., Helman, G., Frank, D.B., Jaffe, N., Yum, S.W., Collins, A., Keller, S.R., et al. (2020). Janus

kinase inhibition in the aicardi-goutières syndrome. N. Engl. J. Med. *383*, 986–989. https://doi.org/10.1056/NEJMC2001362.

62. Sanchez, G.A.M., Reinhardt, A., Ramsey, S., Wittkowski, H., Hashkes, P.J., Berkun, Y., Schalm, S., Murias, S., Dare, J.A., Brown, D., et al. (2018). JAK1/2 inhibition with baricitinib in the treatment of autoinflammatory interferonopathies. J. Clin. Invest. *128*, 3041–3052. https://doi.org/10.1172/JCI98814.
63. Forbes, L.R., Vogel, T.P., Cooper, M.A., Castro-Wagner, J., Schussler, E., Weinacht, K.G., Plant, A.S., Su, H.C., Allenspach, E.J., Slatter, M., et al. (2018). Jakinibs for the treatment of immune dysregulation in patients with gain-of-function signal transducer and activator of transcription 1 (STAT1) or STAT3 mutations. J. Allergy Clin. Immunol. *142*, 1665–1669. https://doi.org/10.1016/J.JACI.2018.07.020.
64. Schuettpelz, L.G., Borgerding, J.N., Christopher, M.J., Gopalan, P.K., Romine, M.P., Herman, A.C., Woloszynek, J.R., Greenbaum, A.M., and Link, D.C. (2014). G-CSF regulates hematopoietic stem cell activity, in part, through activation of Toll-like receptor signaling. Leukemia *28*, 1851–1860. https://doi.org/10.1038/LEU.2014.68.
65. Walter, D., Lier, A., Geiselhart, A., Thalheimer, F.B., Huntscha, S., Sobotta, M.C., Moehrle, B., Brocks, D., Bayindir, I., Kaschutnig, P., et al. (2015). Exit from dormancy provokes DNA-damage-induced attrition in haematopoietic stem cells. Nature *520*, 549–552. https://doi.org/10.1038/NATURE14131.
66. Schmidt, M., Hoffmann, G., Wissler, M., Lemke, N., Müssig, A., Glimm, H., Williams, D.A., Ragg, S., Hesemann, C.U., and von Kalle, C. (2001). Detection and direct genomic sequencing of multiple rare unknown flanking DNA in highly complex samples. Hum. Gene Ther. *12*, 743–749. https://doi.org/10.1089/104303401750148649.
67. Afzal, S., Wilkening, S., von Kalle, C., Schmidt, M., and Fronza, R. (2017). GENE-IS: time-efficient and accurate analysis of viral integration events in large-scale gene therapy data. Mol. Ther. Nucleic Acids *6*, 133–139. https://doi.org/10.1016/J.OMTN.2016.12.001.
68. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/S13059-014-0550-8.
69. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550. https://doi.org/10.1073/pnas.0506580102.
70. Lun, A.T.L., Bach, K., and Marioni, J.C. (2016). Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. Genome Biol. *17*, 75. https://doi.org/10.1186/S13059-016-0947-7.
71. Couckuyt, A., Seurinck, R., Emmaneel, A., Quintelier, K., Novak, D., van Gassen, S., and Saeys, Y. (2022). Challenges in translational machine learning. Hum. Genet. *141*, 1451–1466. https://doi.org/10.1007/S00439-022-02439-8.
72. Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. J. Roy. Stat. Soc. B *67*, 301–320.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
CD3 antibody (clone BW264/56) FITC	Miltenyi Biotec	Cat#130-080-401; RRID: AB_244231
CD56 antibody (clone B159) APC	BD Biosciences	Cat# 555518; RRID: AB_398601
CD19 antibody (clone LT19) PE	Miltenyi Biotec	Cat# 130-091-247; RRID: AB_244223
StraightFrom® Whole Blood CD15 MicroBeads, human	Miltenyi Biotec	Cat# 130-091-058
CD14 antibody (clone M5E2) PercPCy5.5	BD Biosciences	Cat# 550787; RRID: AB_393884
CD2 antibody FITC	BD Biosciences	Cat# 347404; RRID: AB_2868849
CD3 antibody FITC	BD Biosciences	Cat# 345763; RRID: AB_2811220
CD4 antibody FITC	BD Biosciences	Cat# 345768; RRID: AB_2868797
CD8 antibody FITC	BD Biosciences	Cat# 345772; RRID: AB_2868800
CD14 antibody FITC	BD Biosciences	Cat# 345784; RRID: AB_2868810
CD15 antibody FITC	BD Biosciences	Cat# 332778; RRID: AB_2868627
CD16 antibody FITC	BD Biosciences	Cat# 335035; RRID: AB_2868680
CD19 antibody FITC	BD Biosciences	Cat# 345776; RRID: AB_2868804
CD20 antibody FITC	BD Biosciences	Cat# 345792; RRID: AB_2868818
CD33 antibody FITC	BD Biosciences	Cat# 345798; RRID: AB_2868822
CD56 antibody FITC	BD Biosciences	Cat# 345811; RRID: AB_2868832
CD235a antibody FITC	BD Biosciences	Cat# 559943; RRID: AB_397386
CD133 antibody (clone 293C3) PE	Miltenyi Biotec	Cat# 130-090-853; RRID: AB_244346
CD34 antibody (clone 581) PECY7	Beckman	Cat# A21691
CD38 antibody (clone HIT2) APC	BD Biosciences	Cat# 555462; RRID: AB_398599
CD45RA antibody (clone T6D11) APCVIO770	Miltenyi Biotec	Cat# 130-096-604; RRID: AB_2660986
CD90 antibody (clone 5E10) BV421	BD Biosciences	Cat# 562556; RRID: AB_2737651
CD45 human antibody (clone HI30) BV421	Sony Biotechnology	Cat# 2120160
CD34 antibody (clone 581) APC-Cy7	Sony Biotechnology	Cat# 2317570
gp91^{phox} antibody (clone 7D5) FITC	MBL Bio	Cat# D162-4; RRID: AB_591390
CD45 murin antibody (clone 30-F11) APC	BD Biosciences	Cat# 559864; RRID: AB_398672
Biological samples		
Gene Therapy patient peripheral blood sample	Necker's Hospital Biotherapy Clinical Investigation Center	
Mobilized peripheral blood	HemaCare	
Cord blood	Saint Louis's Hospital	
Chemicals, peptides, and recombinant proteins		
SCF	CellGenix	1018-050
FLT3-L	CellGenix	1015-050
TPO	CellGenix	1017-050
IL-3	CellGenix	1002-050
PGE2 Prostin E2 10mg/ml	Pfizer	
X-Vivo 20 medium	Lonza	BESP1058F
Sulfate de protamine, Protamine Choay 10 000UAH- 10ml	Sanofi	
Busulfan	Sigma	B1170000
Critical commercial assays		
DNeasy kit	Qiagen	
RNeasy micro kit	Qiagen	

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chromium Single Cell 3' GEM kit v3	10X Genomics	
Experimental models: Organisms/strains		
Mouse: NSG NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ	The Jackson Laboratories	Strain code 614
Oligonucleotides		
Primer Alb Forward GCTGTCATCTCTTGTGGGCTGT		
Primer Alb Reverse ACTCATGGGAGCTGCTGGTTC		
Probe Alb VIC- CCTGTCATGCCCACACAAATCTCTCC - TAMRA		
Primer HIV Forward CAGGACTCGGCTTGCTGAAG		
Primer HIV Reverse TCCCCCGCTTAATACTGACG		
Probe HIV FAM- CGCACGGCAAGAGGCGAGG -TAMRA		
Primer Alb Forward GCTGTCATCTCTTGTGGGCTGT		
Primer Alb Reverse ACTCATGGGAGCTGCTGGTTC		
Probe Alb VIC-CCTGTCATGCCCACACAAATCTCTCC-QSY		
Primer HIV Forward TCCCCCGCTTAATACTGACG		
Primer HIV Reverse CAGGACTCGGCTTGCTGAAG		
Probe HIV FAM-CGCACGGCAAGAGGCGAGG-IowaBlackFQ		
Deposited data		
Bulk RNAseq data	Biostudies, EMBL-EBI	https://www.ebi.ac.uk/biostudies/studies/S-BSST958
Single cell RNAseq data	Biostudies, EMBL-EBI	https://www.ebi.ac.uk/biostudies/studies/S-BSST959
Full code and post-processed single cell data	Zenodo	https://doi.org/10.5281/zenodo.6580036
Software and algorithms		
Flowjo software (version 10.8)	FlowJo	https://www.flowjo.com
Prism software (version 9)	GraphPad	https://www.graphpad.com
R Studio Software (version 4.0.4)	R Core Team	https://www.R-project.org/
Cell Ranger (version 3.0.2)	10X Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/installation
Cell-ID method	Github	https://github.com/RausellLab/CelliD
Others		
CliniMACS system	Miltenyi Biotech	
FACSCanto	BD Biosciences	
Viia 7	Applied Biosystems	

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Marina Cavazzana (m.cavazzana@aphp.fr).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Bulk RNAseq and Single Cell RNAseq data are available at Biostudies EMBL-EBI :https://www.ebi.ac.uk/biostudies/studies/S-BSST958 and https://www.ebi.ac.uk/biostudies/studies/S-BSST959 respectively.
- The full code listing and the set of post-processed data used for the analysis and the figures are available online in the following repository, Zenodo: https://doi.org/10.5281/zenodo.6580036
- Lentiviral integration sites sequence data used in the present study are available upon request.
- Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Study design and investigational therapy

The study, sponsored by Genethon and registered under NCT02757911, is a phase I/II non-randomized monocentric open-label study conducted at Hôpital Necker-Enfants Malades in Paris under the responsibility of Pr. S. Blanche. Inclusion and exclusion criteria are detailed in the protocol synopsis (see Data S1, supplemental information). In brief, eligible participants were male patients with X-CGD, aged 23 months and older, with molecular diagnosis including DNA sequencing and absent or substantially reduced (>70%) biochemical activity of NADPH-oxidase, and absence of 10/10 human leukocyte antigen-matched donor (sibling or unrelated). The primary objectives included the evaluation of safety and of efficacy by the biochemical and functional reconstitution of the progeny of engrafted cells and the stability of these effects at 12 months. The secondary objectives included clinical efficacy and longitudinal evaluation of augmented immunity against bacterial and fungal infections; assessment of hematopoietic stem cell transduction by the G1XCGD lentiviral vector and engraftment of the gene-modified cells. The study aimed to include up to 5 patients and the data analysis was planned to be mostly descriptive (see Consort diagram, Data S2, supplemental information). During the study, a Data Safety Monitoring Board (DSMB) reviewed the data on a regular basis. The study, still ongoing, includes a follow-up of 2 years to evaluate safety and immune reconstitution parameters, followed by a 3 year period to assess the long-term safety and efficacy.

The protocol and informed consent documents were reviewed and approved by the french institutional ethical review board « Ile-de-France V » and by the french National Agency for the Safety of Medicines and Health Products (ANSM) following an agreement from the French Ministry of Research and Innovation for the use of genetically-modified organisms.

The drug product consists of autologous $CD34^+$ cells transduced with the G1XCGD lentiviral vector. Following informed consent and eligibility confirmation, $CD34^+$ cells were recovered from bone marrow or by apheresis, were transduced *ex vivo* by the G1XCGD lentiviral vector and were frozen until administration to the patient. Patients received busulfan myeloablative conditioning with pharmacokinetic monitoring. After a wash-out period of 24-48 hours the drug products were infused intravenously (i.v.) through a central venous line. Patient vital signs and clinical condition were monitored closely during and after the infusion for adverse reactions.

Patients

All the patients were treated in the Pediatric Immunohematology Department or the Adult Hematology Department at Necker Children's Hospital (Paris, France). The IMP was manufactured in the hospital's Cell and Gene Therapy Laboratory at the Biotherapy Department.

The follow-up included regular patient visits in the Center of Clinical Investigations at Necker Enfants Malades Hospital and laboratory tests. This included clinical status assessment, adverse event recording, immune cell hematological reconstitution, gene marking in cell subpopulations (VCN analyses), $gp91^{phox}$ expression in specific cell subsets, and the DHR oxidative burst assay used to assess the activity of NADPH oxidase. Additional cell characterization assays were performed on an *ad hoc* basis.

Healthy donors

Mobilized peripheral blood (MPB) samples from healthy donors were provided by HemaCare (Northridge, CA, USA). $CD34^+$ cells were mobilized with G-CSF and plerixafor (for HD1-2) or with plerixafor only (for HD3-4). HD5-7 were mobilized with G-CSF, and the $CD34^+$ cells were harvested and immunoselected in the Department of Biotherapy at Necker Children's Hospital. The HDs provided written, informed consent to the use of their samples for research purposes, and their data were anonymized. No nominative data concerning the donor were sent to the investigators.

Cord blood was obtained from a biological resource center (Centre Ressources Biologiques (CRB)) – Banque de Sang de Cordon) at Saint-Louis Hospital (Paris, France). HSPCs were isolated using standard Ficoll density gradient centrifugation and then magnetic selection with anti-$CD34^+$ antibody.

Blood samples from HD8-12 were obtained from the French Blood Establishment (Etablissement Français du Sang, Paris, France; reference: C CPSL UNT-N°18/EFS/032). Again, the HDs provided written, informed consent to the anonymous use of their samples for research purposes. PBMCs were isolated using standard Ficoll density gradient centrifugation.

Mouse experiments and xenotransplantation assays

All animal procedures were approved by the animal care and use committee at the University of Paris (Paris, France; February 16th, 2021) and the French Ministry of Agriculture (APAFIS#29592-2020120216106476). The procedures were performed in accordance with European Union (EU) Directive 2010/63/EU. NOD-SCID-$\gamma c^{-/-}$ strain (NSG) mice were obtained from Charles River Laboratories. 3.5 to $4.2x10^5$ engineered HSPCs from a patient's IMP were injected into 16 NSG mice previously conditioned with one dose per day of busulfan at 15 mg/kg (45 mg/kg in total). Engraftment in BM, spleen and thymus were analyzed after 16 weeks, using flow cytometry. The antibodies used are described in the key resources table.

Culture conditions for the MPB control were the same as in the clinical trial: 18 h of pre-activation in a cytokine cocktail (SCF: 300ng/ml, FLT3L: 300ng/ml, TPO: 100ng/ml, IL3: 20ng/ml), addition of 10 μM PGE2 2h before transduction, and then 2 rounds of transduction with the $gp91^{phox}$ clinical vector and. The VCN in the MPB control sample (3.47) was measured in a ddPCR assay (see below).

The cells from the CB control were not transduced or cultured.

The VCN in the mice was assessed by ddPCR assay on a Bio-rad QX200 ddPCR System. Quantitative PCR on droplets was performed using TaqMan PCR Master Mix for probes in an Applied Biosystems SimpliAmp thermocycler, using a standard protocol. 80 ng of total gDNA, 900 nM primers and 250 nM probes were used in a total volume of 17 μL for absolute quantification with the droplet reader. The primers and probes sequences used are related in the Key Ressources Tables. Fluorescence in PCR-positive droplets was quantified according to a Poisson distribution, and VCNs were calculated according as (PSI*2)/(ALB).

METHOD DETAILS

G1XCGD lentiviral vector production

The lot of clinical-grade G1XCGD lentiviral vector was manufactured at Yposkesi (Evry, France) under good manufacturing practices using identical processes as lots produced for other studies.[19] Briefly, the G1XCGD vector was produced by transient transfection of 293T cells with plasmids encoding the G1XCGD transfer cassette, HIV-1 gag/pol, HIV-1 rev and the VSV-G glycoprotein. The viral particles were purified from culture supernatants following clarification, ion exchange chromatography, tangential flow filtration and gel filtration steps, formulated in X-Vivo 20 medium (Lonza), aliquoted and cryopreserved at < -70°C. The lot of vector used to treat patients titered $3.7x10^9$ infectious genomes (IG) ml^{-1} measured at Genethon by qPCR using HCT116 colon carcinoma cells, and $2.1x10^4$ HIV1 P24 core antigen (P24) ml^{-1} as measured by ELISA assay. The lot of vector tested negative for replication-competent lentivirus and was conform to all release specifications.

G1XCGD lentiviral vector-modified $CD34^+$ cell product manufacture

Cells were manufactured onsite at the Cell and Gene Therapy Laboratory in the Biotherapy department of Necker-Enfants Malades Hospital in Paris. Patient's cells were obtained from apheresis procedures after a mobilization regimen based on G-CSF and plerixafor, and when appropriate from bone marrow (BM) harvest under general anaesthesia. A back-up harvest of at least $3x10^6$ unmanipulated $CD34^+$ cells per kg was retained in case of failure of hematopoietic reconstitution following gene therapy. As repetitive G-CSF mobilizations could have a detrimental effect on HSC fitness,[64,65] after two unsuccessful collections and engineering of P2 HSPC, a Plerixafor-only mobilization was performed for the third apheresis to avoid further detrimental effects of G-CSF treatment. The transduction procedure is summarized in Figure 1A. Briefly, $CD34^+$ cells were immunoselected using the CliniMACS system (Miltenyi Biotec). The selected $CD34^+$ cells were pre-activated into cell culture in serum-free medium (X-vivo20, Lonza) with recombinant human cytokines (stem cell factor at 300 ng/ml, flt-3 ligand at 300 ng/ml, thrombopoietin at 100 ng/ml and interleukin-3 at 20 ng/ml, CellGenix). On the next two successive days, the G1XCGD lentiviral vector was added to the cell culture at a final concentration of $1x10^8$ $IG.ml^{-1}$. PGE2 was added as transduction adjuvant for the manufacturing of P2 (third manipulation), P4 and P5 IMP. The following day the cells were removed from culture, washed and cryopreserved, to allow the release quality controls. The IMP was infused after targeted myeloablative conditioning with a minimum $CD34^+$ cell dose to be administered of $3x10^6$ cells/kg body weight.

Determination of the VCN

Genomic DNA was extracted from HSPCs in the IMP (14 days after transduction) and during the follow-up from sorted neutrophils, monocytes, T cells, B cells and NK cells and on total PBMCs using a DNeasy Kit (Qiagen).

The VCN was determined in a quantitative PCR assay (Viia 7, Applied Biosystems) and the PSI and ALB human probes (see key resources table).

The DHR assay

Neutrophils were stimulated with phorbol myristate acetate to induce superoxide anion production, according to the manufacturer's protocol. The non-fluorescent dye DHR is reduced by H_2O_2 and then converted into fluorescent rhodamine, which is quantified using flow cytometry.

Isolation of mononuclear cells

In line with the trial protocol, BM and/or MPB was collected for IMP manufacturing and peripheral blood was sampled regularly during the follow-up period. Mononuclear cells were isolated from PB or MPB using standard Ficoll density gradient separation. The absolute lymphocyte count was determined using Trucount Tubes (BD Bioscience).

Flow cytometry

The neutrophil subpopulation was purified from PB on a column with magnetic beads and fluorochrome-coupled anti-CD15 antibodies. Monocytes, T cells, B cells and NK cells were sorted on a cell sorter (FACSAria II, BD Biosciences), using fluorochrome-coupled antibodies against CD14, CD3, CD19, and CD56.

Total PBMCs were surface-stained for $gp91^{phox}$, using an anti-flavocytochrome b558 7D5 clone (human) mAb-FITC (MBL Bio) and gating for neutrophils.

The patients' HSPCs were characterized using a multilabeled panel with the antibodies listed in the key resources table: lineage cocktail, CD34, CD133, CD38, CD90, and CD45RA. Staining was analyzed with a FACSCanto II cell analyzer.

Analysis of vector integration sites

Integration site were identified on PBMC and neutrophil samples, using the S-EPTS/LM-PCR protocol, an advanced version of EPTS/LM-PCR,[66] and thereafter analyzed using the GENE-IS tool suite.[67]

Bulk RNA-seq

RNA was isolated using RNeasy Micro Kit (Qiagen) with a DNase step. RNA integrity and concentration were assessed using capillary electrophoresis and the Fragment Analyzer (Agilent). RNA-seq libraries were prepared from 100 ng of total RNA, using the Universal Plus mRNA (Nugen-Tecan). The amplified cDNA produced was sequenced on a NovaSeq6000 system (Illumina). There were ~50 million reads per library.

The raw read counts were normalized with DESeq2 package, based on the library size and testing for differential expression between conditions.[68] Coding genes were extracted from gencodeV30. Next, the noise filter was used to retain only genes that had at least one sample with an expression value greater than 20 before the pathway enrichment analysis. Normalized enrichment scores were calculated for all deregulated coding genes, using GSEA software.[69]

Gene set enrichment was investigated with MSigDB, using a hypergeometric test on a pre-filter dataset ($p<0.05$ and fold-change (FC) >1.2 or <1/-1.2). The output false discovery rate had to be below 0.05.

Representation and quantification of module activity (ROMA) was applied to DEGs in PBMCs and HSPCs. ROMA calculate a module score for a set of samples and is based on the simplest single-factor linear model of gene regulation whose first principal component approximates the expression data.[23]

Single-HSPC RNA-seq

Library preparation

Frozen HSPCs from each individual were thawed and resuspended in PBS + 1% BSA. The cell preparation was loaded onto a Chromium Single-Cell Chip (10x Genomics) for co-encapsulation with barcoded Gel Beads at a target capture rate of ~7000 individual cells per sample. Captured mRNAs were barcoded during cDNA synthesis, using the Chromium Single-Cell 3' reagents v3 (10x Genomics) according to the manufacturer's instructions. All samples were processed simultaneously with the Chromium Controller (10x Genomics), and the resulting libraries were prepared in parallel in a single batch. We pooled all the libraries for sequencing in a single SP Illumina flow cell. Libraries were sequenced with 28 read 1 cycle containing cell-identifying barcodes and unique molecular identifiers (UMIs), 8 i7 index cycles, and 91 read 2 cycles containing transcript sequences on an Illumina NovaSeq 6000 (Illumina). Sequencing reads were demultiplexed and aligned with the human reference genome (GRCh38), using the CellRanger pipeline v3.1.

The pipeline for data processing and analysis of scRNAseq data is shown in Figure S6.

Integration and data pre-processing

Empty droplets were excluded with DropletUtils package, with an FDR threshold of 0.01. Cells with more than 15% of mitochondrial genes and less than 3000 UMI were removed. As HSPCs differ in their maturity (translating into difference in expression abundance from one cell to another), the expression matrix for each sample was normalized using deconvolution rather than standard library size methods.[70] The gene expression was then restricted to protein-encoding genes. The 6000 highly variable genes were found with the Seurat FindVariableFeatures function and its default parameters.

QUANTIFICATION AND STATISTICAL ANALYSIS

Single-HSPC RNA-seq

Cell-ID annotation of individual cells, using HSPC reference signatures

Cell-ID is a robust statistical method for gene signature extraction and cell identity recognition on the basis of single-cell RNA-seq data.[25] It incorporates a multiple correspondence analysis and simultaneously represents cells and genes in low-dimension space. The genes are then ranked by their Euclidean distance from each individual cell, which provides unbiased per-cell gene signatures.

Using published data,[26] Cell-ID, and the 200 most specific genes, we extracted 16 reference signatures: HSC, MPP, MLP, ImP1, ImP2 (corresponding to common myeloid progenitors), NeutroP0, NeutroP1, NeutroP2, NeutroP3, MonoDCP (corresponding to granulocyte-monocyte progenitors), BcellP, MEP1, MEP2, EryP, MkP, and EoBasMastP. The Cell-ID method defines the gene ranking in each cell in the dataset (53,412 cells in total), evaluates whether a cell accurately matches a particular reference signature, and determines the cell's identity (Cell ID) on the basis of the top p-value ($p<0.01$) (Figure S8A). The enrichment score is based on the $-\log_{10}$(p-value).

HSC identification and the diffusion map

Since the annotated population was enriched in HSCs (corresponding to 30% of all HSPCs), we combined diffusion map analysis (for determining the differentiation trajectory (Figure S7A)) with an analysis of the enrichment strength for Velten et al.'s HSC signature (to determine the origin of the diffusion map (Figure S7B) and isolate the most immature HSC subpopulation, corresponding to 3% of the total HSPC (Figure S7C)).

The other annotated HSCs are referred to as "HSC-enriched". The "All HSC" subpopulation includes the most immature HSCs and the HSC-enriched subpopulations.

Following cell-type annotation, samples were integrated by applying the *Harmony* package using the first 30 principal component axis and the default parameter as input.

The Cell-ID score for signaling pathway enrichment

The Cell-ID method was used to assess the statistical enrichment of individual-cell gene signatures vs. signaling pathway gene sets (such as Hallmark gene sets, MSigDB collections, v7.5.1) based on hypergeometric test p-values with Benjamini–Hochberg correction for the number of tested gene signatures. Enrichment scores were calculated as the $-\log_{10}$(p-value) in the test. A cell was considered to be enriched in a given pathway when the score was >2 ($p<0.01$).

Cell-ID identification of mixed signatures, and UpSet plots representation

To further understand the heterogeneity and diversity of cell state among the cells, we took advantage of Cell-ID enrichment system to identify cells that were significantly enriched ($p<0.01$) for several reference signatures. These cells were then represented on an UpSet plot with the *UpSetR* package for all labels (Figure S8D) or on selected labels such as NeutroP0, BcellP, MonoDCP, MPP, All HSC and others (Figure 3).

Identification of deregulated genes with MAST

In order to identify DEGs in HSC subpopulations, we made use of the MAST approach[27] (https://github.com/RGLab/MAST) based on statistical models tailored to single-cell data, allowing inference for genes with sparse expression. These models can handle a more complex variance structure, such as expected correlations between cells derived from the same individual.

DEGs were identified with a Hurdle model (implemented with *MAST* v1.16.0) testing the 6000 highly variable genes in the dataset, and adjusting with a cellular detection rate parameter that correspond to the number of genes detected in a cell.

To analyze the enrichment pathway, we applied a hypergeometric test for pathway enrichment using Hallmark geneset, MSigDB database.

Elastic-net logistic regression for the identification of predictive markers

We used an elastic-net logistic regression model[29,71,72] with the *glmnet* package in order to predict the HSCs' ability to engraft or not.

We constructed the model at the cell level and defined the capacity of engraftment based on the patient it belongs to (success for the 304 HSCs of P1 and P4, failure for the 165 cells of P2 and P5). The genes of interest used as variables were the pre-selected 239 IFNa, IFNg genes and transcription factors differentially expressed in at least one patient (Figure S10). We performed cross-validation using the *caret* package to determine the optimal lambda (6.87e-05) and alpha (0.7) parameters on a training dataset composed of 75% of patients' HSCs. Then we performed a Monte Carlo cross validation by randomly splitting the HSCs in a training set (75% of the cells) and a test set (25% remaining HSCs) 50 times with the tuned parameters to check the stability of the results. We obtained AUC values between 0.98 and 1, and accuracy between 0.95 and 1 (Figures S10A–S10C). Using those models, we determined the selection's frequency of each gene and put a threshold at 70% to select the most informative ones (Figure S10D). 78 significantly contributing factors were selected as engraftment predictors with this method. We then concentrated on the 51 factors with a negative estimate (i.e. corresponding to detrimental factors for engraftment that were up-regulated in P2 and P5) (Figures S10E and S10F).

The network of genes selected by the elastic-net model as being detrimental for engraftment was visualized using *StringDB* (Figure 5F).

Other statistical analysis

Statistical analyses in Figures 6 and S1 were performed using GraphPad Prism9 software (GraphPad, La Jolla, CA, USA) as indicated in the figure legend.

ADDITIONAL RESOURCES

Here, we report the results of a Phase I/II clinical trial of GT (NCT02757911) in four patients with X-CGD lacking a human leukocyte antigen (HLA)-compatible donor for HSCT, sponsored by Genethon. A fifth patient was included in the clinical trial but was not treated because the investigational medicinal product (IMP) did not meet the release criteria.