Article

SOURCE DATA | TRANSPARENT PROCESS | OPEN ACCESS

THE EMBO JOURNAL

Unbiased assessment of genome integrity and purging of adverse outcomes at the target locus upon editing of CD4$^+$ T-cells for the treatment of Hyper IgM1

Daniele Canarutto[1,2,3], Claudia Asperti[1], Valentina Vavassori[1], Simona Porcellini[1], Elisabetta Rovelli[1], Marianna Paulis[4,5], Samuele Ferrari[1], Angelica Varesi[1], Martina Fiumara[1], Aurelien Jacob[1], Lucia Sergi Sergi[1], Ilaria Visigalli[1], Francesca Ferrua[1,3], Luis Ignacio González-Granado[6], Vassilios Lougaris[7], Andrea Finocchi[8], Anna Villa[1,5], Marina Radrizzani[1] & Luigi Naldini[1,2,*]

Abstract

Hyper IgM1 is an X-linked combined immunodeficiency caused by *CD40LG* mutations, potentially treatable with CD4$^+$ T-cell gene editing with Cas9 and a "one-size-fits-most" corrective template. Contrary to established gene therapies, there is limited data on the genomic alterations following long-range gene editing, and no consensus on the relevant assays. We developed drop-off digital PCR assays for unbiased detection of large on-target deletions and found them at high frequency upon editing. Large deletions were also common upon editing different loci and cell types and using alternative Cas9 and template delivery methods. In *CD40LG* edited T cells, on-target deletions were counter-selected in culture and further purged by enrichment for edited cells using a selector coupled to gene correction. We then validated the sensitivity of optical genome mapping for unbiased detection of genome wide rearrangements and uncovered on-target trapping of one or more vector copies, which do not compromise functionality, upon editing using an integrase defective lentiviral donor template. No other recurring events were detected. Edited patient cells showed faithful reconstitution of *CD40LG* regulated expression and function with a satisfactory safety profile. Large deletions and donor template integrations should be anticipated and accounted for when designing and testing similar gene editing strategies.

Keywords AAV6; CRISPR; IDLV; large deletions; large integrations
Subject Categories Biotechnology & Synthetic Biology; Genetics, Gene Therapy & Genetic Disease; Immunology
DOI 10.15252/embj.2023114188 | Received 4 April 2023 | Revised 15 September 2023 | Accepted 25 September 2023 | Published online 2 November 2023
The EMBO Journal (2023) 42: e114188

Introduction

Hyper IgM1 (HIGM1, OMIM #3082) is a rare X-linked combined immunodeficiency caused by mutations in the CD40 ligand (*CD40LG*) gene, which encodes for CD40 ligand (CD40L), also known as CD154 (Tangye *et al*, 2020). CD154 is promptly displayed on the surface of CD4$^+$ T-cells upon activation, in a tightly regulated manner. It is required to induce class-switch recombination and somatic hypermutation of B cells. *CD40LG* mutations result in a characteristic immunoglobulin phenotype of normal to high IgM levels and low levels of the other immunoglobulin classes (Notarangelo *et al*, 2014). HIGM1 patients are vulnerable to infections, liver disease, and cancer (Yazdani *et al*, 2019), with a median survival of 25 years (de la Morena *et al*, 2017). Hematopoietic stem cell transplantation (HSCT) can be curative but can entail high risks (Ferrua *et al*, 2019). HIGM1 cannot be treated with gene addition (Brown *et al*, 1998; Sacco *et al*, 2000), and we and others (Hubbard *et al*, 2016; Kuo *et al*, 2018;

1 San Raffaele Telethon Institute for Gene Therapy, IRCCS San Raffaele Scientific Institute, Milan, Italy
2 Università Vita-Salute San Raffaele, Milan, Italy
3 Pediatric Immunohematology Unit and BMT Program, IRCCS San Raffaele Scientific Institute, Milan, Italy
4 Humanitas Clinical and Research Center IRCCS, Milan, Italy
5 UOS Milan Unit, Istituto di Ricerca Genetica e Biomedica (IRGB), CNR, Milan, Italy
6 Unidad de Immunodeficiencias Primarias y la Unidad de Hematología y Oncología Pediátrica, Instituto de Investigacíon Hospital 12 de Octubre, Madrid, Spain
7 University of Brescia, Brescia, Italy
8 Academic Department of Pediatrics (DPUO), Research Unit of Clinical Immunology and Vaccinology, Bambino Gesú Children's Hospital, Istituto di Ricovero e Cura a Carattere Scientifico, Rome, Italy
*Corresponding author. Tel: +39 02 26434 681; Fax: +39 02 26434 668; E-mail: naldini.luigi@hsr.it

Vavassori *et al*, 2021) previously showed the restoration of regulated *CD40LG* expression by *in situ* gene editing with Cas9 and an adenoviral associated virus serotype 6 (AAV6) donor template. We developed a strategy whereby a corrective cassette encoding for a splice acceptor sequence and the cDNA of *CD40LG* exons 2–5 is integrated into the first intron of the gene, thus bypassing the endogenous exon 2 splicing and reconstituting functional regulation of the gene (Fig 1A). Here we further pursued clinical translation of autologous *in situ* corrected $CD4^+$ T-cell therapy, addressing genome integrity upon editing. While other knock-out based gene editing products have already entered clinical testing (Frangoul *et al*, 2021; Fu *et al*, 2022), little is known about the genotoxicity of the procedure and, despite concerns (Boutin *et al*, 2021; Leibowitz *et al*, 2021; Ferrari *et al*, 2022; Nahmad *et al*, 2022; Sasu *et al*, 2022), there is no consensus on the relevant characterizations that should be performed. Here we report the development of relevant and portable unbiased genome integrity assessments, both on-target and genome-wide, validating high coverage optical genome mapping (OGM) for the latter purpose. We then completed our assessment of genome integrity by OGM in cells edited with Cas9 and an integrase defective lentiviral vector (IDLV) and expanded in optimized culture conditions potentially compatible with clinical testing.

Results

Large deletions occur on-target but are counter selected in culture

The experimental setup is summarized in Fig 1B. Briefly, $CD4^+$ T-cells were isolated from buffy coats or unstimulated leukapheresis, activated, edited with Cas9 ribonucleoprotein (RNP) and a donor adeno-associated viral vector (AAV6), optionally enriched for surface low affinity nerve growth factor receptor (LNGFR) expression, and expanded. Having previously shown any off-target activity of HiFi Cas9 with our gRNA to be below the limit of detection (Vavassori *et al*, 2021), we initially focused on events occurring on-target. As shown in other contexts (Adikusuma *et al*, 2018; Boutin *et al*, 2021; Nahmad *et al*, 2022; Park *et al*, 2022), we hypothesized that the DNA DSB could result in some long-range deletions, and sought to quantify them in a routine manner not relying on the presence of primer binding sites in edited cells (Fig 1C). In Hyper IgM1 patients, the hemizygous CD40LG is deficient, and indels or deletions further disrupting the gene are expected to be neutral. Therefore, we decided to investigate the occurrence of deletions reaching the closest genes to *CD40LG*, i.e. *LNC00892* and *ARHGEF6*, using droplet digital PCR (ddPCR). *MECP2*, located near the Xq telomere was additionally probed to assess loss of the Xq arm. In this drop-off assay deletions are called when fewer targets can be amplified from the treatment sample as compared to the reference sample. We detected a median loss of 14% of amplicons (range 3–71%, Fig 1D) mapping on *LNC00892* early after editing, indicating deletions spanning at least 7 kb towards the centromere, with a trend towards increased frequency of amplicon loss in female cells, likely reflecting greater target abundance and tolerance to genetic inactivation in the biallelic condition. Deletions could be detected 2 h after editing and were most abundant after 6–48 h (Fig EV1A and E). Cells harboring deletions were counter-selected in culture, as the copy number variation (CNV) difference between edited cells and untreated ones progressively decreased, and 6 days after editing little difference could be seen (Fig 1E). Small CNV differences in *ARHGEF6* and *MECP2* (Fig EV1B–D) were also seen, although not reaching statistical significance, possibly due to the greater distance from the DNA DSB.

As could be expected from the coupling of LNGFR expression with *CD40LG* correction (Figs 1A and F, and EV1E), selecting edited $CD4^+$ T-cells reduced the proportion of deletions involving *LNC00892* (Fig 1G). Interestingly, deletions were more frequent in the presence

Figure 1. Experimental setup and on-target deletions in edited $CD4^+$ T-cells.

A Overview of gene correction strategy. The corrective cassette consists of homology arms (dashed lines), a splice acceptor sequence (SA), the corrective cDNA of exons 2–5, and is followed in its final design by an internal ribosome entry sequence (IRES) cassette that couples reconstitution of *CD40LG* with surface expression of the low affinity nerve growth factor receptor (LNGFR). It is integrated in intron 1 of the *CD40LG* gene, targeted by Cas9 ribonucleoprotein.

B Experimental design. $CD4^+$ lymphocytes were isolated Day 0 from buffy coats or leukapheresis, cultured for 2–3 days in presence of CD3/CD28 stimulation, IL7 and IL15, and then electroporated and transduced with an adenoviral associated vector serotype 6 (AAV6) or an integrase defective vector (IDLV) on day 2 or 3. Optional enrichment for surface LNGFR took place on day 6; cells were then expanded until day 13–14.

C Schematic representation of ddPCR CNV assays' positions on chromosome X with respect to the *CD40LG* locus (colored boxes).

D Loss of LNC00892 amplicons 1–2 days after editing and HDR efficiency 6 days after editing in males (M, $n = 11$) and female (F, $n = 15$) from seven independent experiments. Two tailed Mann Whitney test.

E Variation over time of *LNC00892* copies/diploid genome in males (blue) and females (pink) with respect to untreated samples. Median + IQR. Kruskal–Wallis test for area under the curve, $n = 6$ males + 6 females.

F Expression of LNGFR in patients' cells expressing CD40LG after gene editing and stimulation ($n = 4$).

G Variation in *LNC00892* copy number after enrichment (immunomagnetic or sorting), on the day of selection. Pool of five male donors (blue) and six female (pink) donors from 3 independent experiments. Wilcoxon test.

H–J Comparison of LNC00892 copies/diploid genome in $CD4^+$ T-cells edited at the CD40LG locus in absence or presence of AAV6 donor template 1–2 days after editing ($n = 4$ female and 8 male donors, in pink and blue respectively; (I) RNP cutting efficiency and (J) HDR efficiency, both 6 days after editing). Two tailed Wilcoxon test.

K, L Frequency of LNC00892 deletions 1 day after editing with increasing MOI and corresponding HDR efficiency 6 days after editing (K; $n = 3$ biological replicates).

M, N Variation in HDR edited alleles early after editing (M) and corresponding reduction in proportion of deletions in the same samples (N, $n = 9$ female donors, in pink, and 5 male donors, in blue). Two tailed Wilcoxon test.

O–Q Frequency of deletions in IDLV-edited cells 6 days after editing before and after LNGFR selection (IDLV POS) ($n = 6$ female donors, in pink and 3 male donors, in blue; two-tailed Wilcoxon matched-pairs signed rank test) and corresponding cutting efficiency (P) and HDR efficiency (Q).

R, S Copy number variation of target amplicon (AAVS1, B2M, or IL2RG) in colonies derived from cord blood (CB) or mobilized peripheral blood (MPB), placed at 126–800 bp from the DNA DSB. Cas9 was delivered as RNP or RNA. Two sided Fisher's exact test comparing proportion of cells below the cutoff (dashed line).

Source data are available online for this figure.

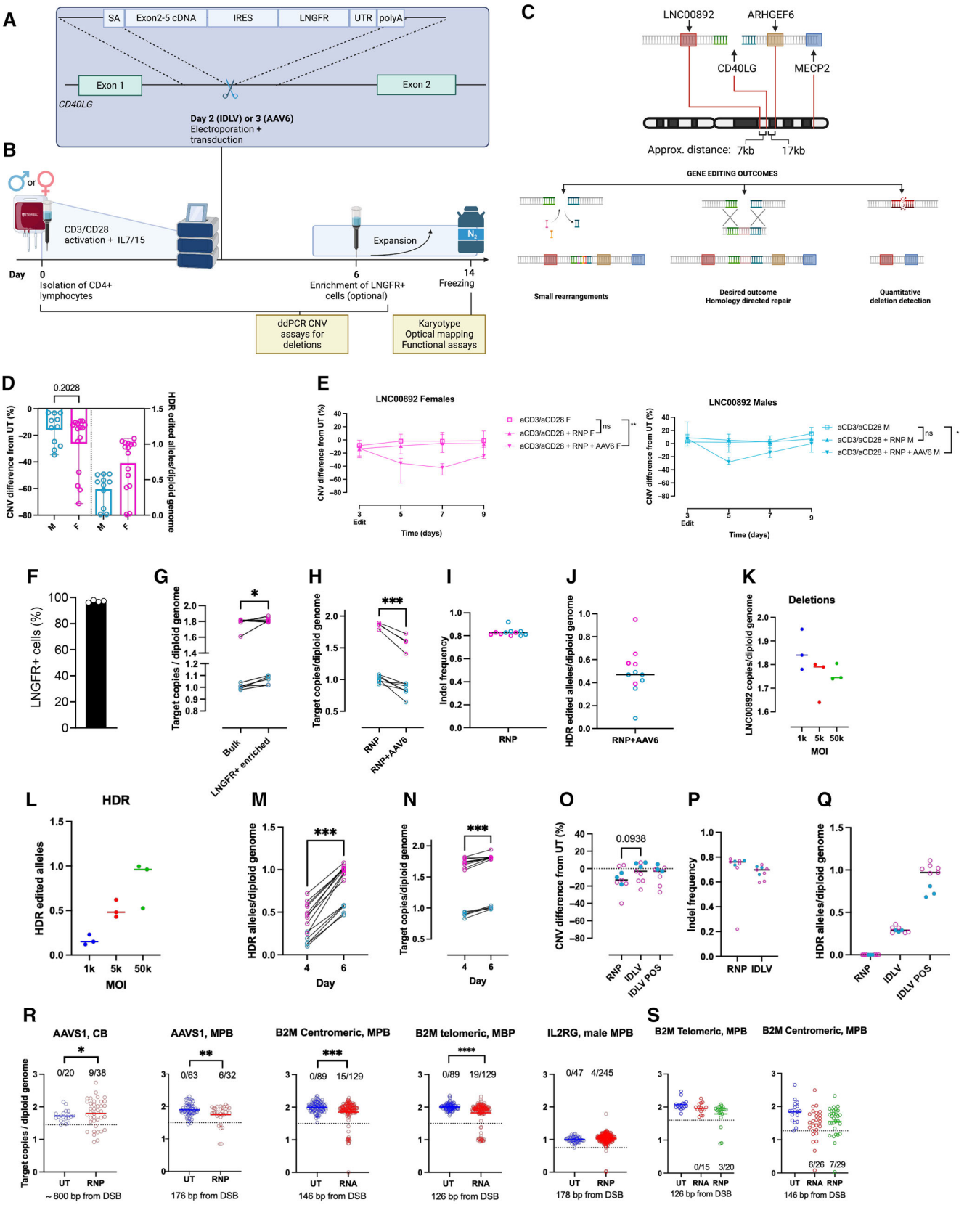
A
SA
Exon2-5 cDNA
IRES
LNGFR
UTR
polyA
Exon 1
Exon 2
CD40LG
Day 2 (IDLV) or 3 (AAV6)
Electroporation +
transduction
B
or
CD3/CD28
activation + IL7/15
Expansion
Day
0
Isolation of CD4+
lymphocytes
6
Enrichment of LNGFR+
cells (optional)
14
Freezing
ddPCR CNV
assays for
deletions
Karyotype
Optical mapping
Functional assays
C
LNC00892
ARHGEF6
CD40LG
MECP2
Approx. distance: 7kb 17kb
GENE EDITING OUTCOMES
Small rearrangements
Desired outcome
Homology directed repair
Quantitative
deletion detection
D
0.2028
CNV difference from UT (%)
HDR edited alleles/diploid genome
M
F
M
F
E
LNC00892 Females
aCD3/aCD28 F
aCD3/aCD28 + RNP F
aCD3/aCD28 + RNP + AAV6 F
ns
**
CNV difference from UT (%)
Edit
Time (days)
LNC00892 Males
aCD3/aCD28 M
aCD3/aCD28 + RNP M
aCD3/aCD28 + RNP + AAV6 M
ns
*
F
LNGFR+ cells (%)
G
Target copies / diploid genome
*
Bulk
LNGFR+ enriched
H
Target copies/diploid genome
***
RNP
RNP+AAV6
I
Indel frequency
RNP
J
HDR edited alleles/diploid genome
RNP+AAV6
K
Deletions
LNC00892 copies/diploid genome
1k
5k
50k
MOI
L
HDR
HDR edited alleles
MOI
M
HDR alleles/diploid genome
***
Day
N
Target copies/diploid genome
***
Day
O
0.0938
CNV difference from UT (%)
RNP
IDLV
IDLV POS
P
Indel frequency
RNP IDLV
Q
HDR alleles/diploid genome
RNP
IDLV
IDLV POS
R
AAVS1, CB
*
0/20
9/38
Target copies / diploid genome
UT
RNP
~ 800 bp from DSB
AAVS1, MPB
**
0/63
6/32
UT
RNP
176 bp from DSB
B2M Centromeric, MPB
***
0/89
15/129
UT
RNA
146 bp from DSB
B2M telomeric, MBP
****
0/89
19/129
UT
RNA
126 bp from DSB
IL2RG, male MPB
0/47
4/245
UT
RNP
178 bp from DSB
S
B2M Telomeric, MPB
0/15
3/20
UT
RNA
RNP
126 bp from DSB
B2M Centromeric, MPB
6/26
7/29
UT
RNA
RNP
146 bp from DSB

Figure 1.

of AAV6 (Fig 1H–J), with a dose-dependent trend (Fig 1K and L). We also noticed that the proportion of HDR edited cells increased over time in culture (Fig 1M), paralleling the decrease in deletions (Fig 1N) and reflecting dynamic changes in culture composition.

Given the lower geno-toxicity concerns observed for IDLV over AAV6 as template delivery vehicle in the context of HSPC gene editing (Ferrari *et al*, 2022), we favored the former for the development of scalable GMP compliant T-cells manufacturing process (Asperti *et al*, 2023), whose editing reagents could be portable to HSPCs. We chose to include the low affinity nerve growth factor receptor (LNGFR) selector (Fig 1A) in the *CD40LG* corrective cassette, functionally coupled to gene correction by an Internal Ribosome Entry Sites (IRES) sequence to allow for *in vitro* enrichment of edited cells and better reconstitution of CD40L expression (Vavassori *et al*, 2021) and function, based on the following data. First, infusion of non-functional cells together with the functional ones resulted in lower total and antigen-specific IgG upon vaccination in the murine model of the disease (Fig EV2A–D). Second, considering that *CD40LG* is hemizygous in the male affected subjects, the selection was expected to benefit genome integrity by purging adverse outcomes at the target site (Fig 1F and G). While the burden of deletions did not appear to be worsened by the presence of IDLV (Fig 1O–Q), deletions were still detected in female cells.

We validated the rationale and portability of our unbiased drop-off ddPCR based tiling assay in colony forming units (CFUs) derived from hematopoietic stem/progenitor cells (HSPCs) sourced from mobilized peripheral blood (MPB) or cord blood (CB) and edited with Cas9, either delivered as RNA or RNP, at different clinically relevant loci (Fig 1R), using CNV assays placed at variable distance from the cut site. Here, the clonal nature of CFUs allowed for unequivocal detection also of events occurring at low frequency in the bulk population, overcoming the technical sensitivity limit stemming from stochastic error. In four target regions we reproducibly found around 15–20% colonies bearing deletions spanning over hundreds of base pairs, indicating viability and conservation of the genetic lesions over the clonogenic outgrowth. A lower but still detectable proportion of deletions was observed in male cells edited at *IL2RG*, possibly due to counterselection of X-linked deletions as observed for *CD40LG* edited cells; little difference was observed between delivering Cas9 as RNA or RNP, at comparable levels of knock-out efficiency (Figs 1S and EV1F).

Validation of high coverage optical genome mapping for detection of rare structural variations

One hundred-metaphase karyotype is the gold standard for clinical assessment of structural rearrangement, but intrinsically limited in sensitivity and resolution (Silva *et al*, 2019). Therefore, we sought to validate OGM, a novel genomic technology not relying on PCR amplification but rather on fluorescent labelling of restriction sites across long stretches of DNA molecules for the assessment of $CD4^+$ T-cells upon editing (Smith *et al*, 2022). To this end, we generated a male lymphoblastoid cell line simultaneously harboring two structural variations (SV) on the X chromosome, close to either the lower (5 kb) or upper reported limit of detection (250–300 kb), as well as a bystander chromosome 14 aneuploidy (Fig 2A–C). The modified clone was then precisely admixed with the unmodified one in different proportions, each corresponding to a specific variant allele frequency (VAF), for *in silico* comparison against the pure unmodified one in order to accurately assess technology performance at the desired VAF (Fig 2D). At 10% VAF the 6.9 kb integration was correctly detected reporting a 0.16 VAF (confidence 0.99, Fig 2E). The position of the new optical label mapping in the 6.9 kb insertion was consistent with the position of the DLE-1 restriction sequence CTTAAG already present in the donor plasmid. The rearrangement occurring around *CD40LG* locus was called by a 149 kb duplicated inversion (confidence −1, VAF 0.12) partially overlapping a 238 kb inversion (confidence 0.83, VAF 0.12) which were manually aligned to a total length of approximately 296 kb (Fig EV3). Among other events (Table EV1), the chromosome 14 aneuploidy was correctly detected (CNV 2.11). An isolated hemizygous centromere-less 3q chromosome gain already present in the parental cell line was interpreted in light of a translocation with chromosome 21 [t(3;21)(q22.1;p11.2)] that had been filtered out due to masking and low confidence (0.02), which matched the derivative chromosome 21 observed by karyotype (Table EV2 and Fig EV4). We then assessed performance at VAF of 2.5%: the 6.9 kb insertion was correctly detected (VAF 0.04, confidence 0.99), but the large rearrangement

Figure 2. Validation of optical mapping.

A Schematic workflow: (1) A subclone (in red) was generated from bulk SKW 6.4 cell line. (2) A linearized plasmid constitutively expressing GFP was integrated into the *IL2RG* locus by trapping in a Cas9 induced DSB. A single clone containing the correct insert was characterized and expanded. (3) The same GFP^+ clone was cut simultaneously at two 300-kb spaced loci (*GPR112* and *ARHFEF6*), respectively upstream and downstream the *CD40LG* locus on chromosome X. A single cell clone was derived again after this modification, characterized, and expanded. (4) The clone of interest (purple) was admixed in known proportions with the unmodified clone (step 1, in red) by single cell sorting. (5) Samples admixed at different proportions (pink-purple) were compared against their unmodified counterpart derived in step 1 (in red).

B Reconstruction of the 6.9 kb insertion in the *IL2RG* locus. Sequenced junctions are shown by dark red arrows.

C Large genomic rearrangements detected in the selected clone. Bars indicate ratios between ddPCR CNV assays located on chromosome X, near the centromere (*ARHGEF9*), on chromosome X, near the telomere (*MECP2*), within the rearranged region comprised within the Cas9 target sites (*ARHGEF6* and *LNC00892*), chromosome 14 (*TTC5*), chromosome 19 (*EPS8L1*), chromosome 1 (*EIF2C1*). Bars indicate 95% confidence interval of the copy number.

D Comparison of Sample #1, containing 10% of cells with genomic rearrangements of the X chromosome, against Sample #0 consisting of unmodified cells. Circos plot: circles from the most external to the most internal represent: (1) Chromosome number. (2) Conventional banding. (3) Annotated genes. (4) Structural variants, color coded. (5) Copy number variations. (6) Translocations (none present in current plot). The green dot on chromosome X indicates an insertion in the *IL2RG* locus, while the blue and purple dots indicate a rearrangement around the *CD40LG* locus.

E Integration of the reporter plasmid in the *IL2RG* locus. Tracks from top to bottom: (1) Banding of the X chromosome. (2) Copy number variations (light green track). (3) Structural variants; the horizontal dark green line indicates an insertion. (4) Annotated genes. (5) Reference assembly (green, labels in blue). (6) Sample assembly (light blue), with labels matching the reference (dark blue) or not (yellow).

Source data are available online for this figure.

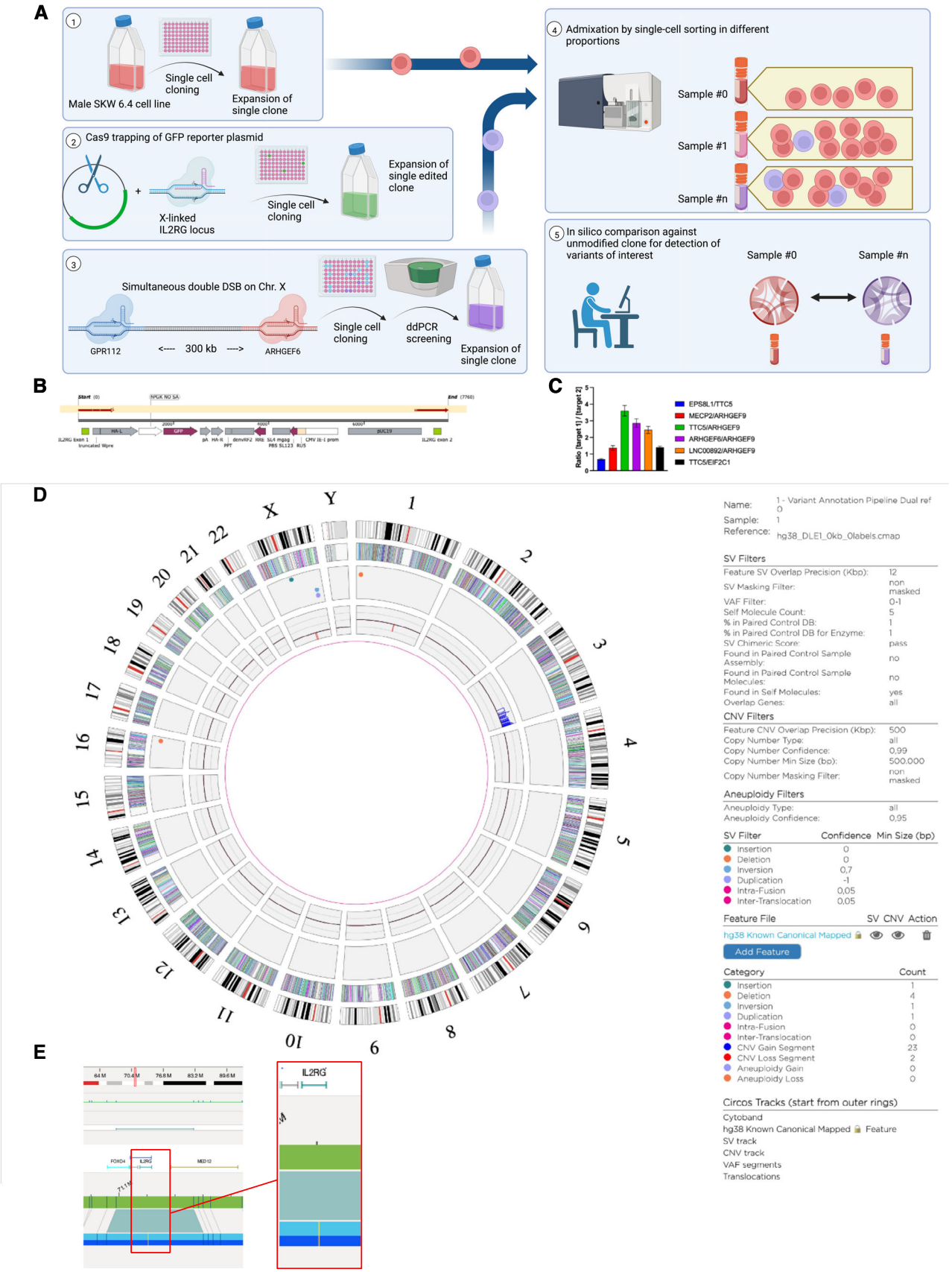
A
1
Single cell cloning
Male SKW 6.4 cell line
Expansion of single clone
2
Cas9 trapping of GFP reporter plasmid
X-linked IL2RG locus
Single cell cloning
Expansion of single edited clone
3
Simultaneous double DSB on Chr. X
GPR112
<---- 300 kb ---->
ARHGEF6
Single cell cloning
ddPCR screening
Expansion of single clone
4
Admixation by single-cell sorting in different proportions
Sample #0
Sample #1
Sample #n
5
In silico comparison against unmodified clone for detection of variants of interest
Sample #0
Sample #n
B
C
Ratio [target 1] / [target 2]
EPS8L1/TTC5
MECP2/ARHGEF9
TTC5/ARHGEF9
ARHGEF6/ARHGEF9
LNC00892/ARHGEF9
TTC5/EIF2C1
D
Name: 1 - Variant Annotation Pipeline Dual ref 0
Sample: 1
Reference: hg38_DLE1_0kb_0labels.cmap
SV Filters
Feature SV Overlap Precision (Kbp): 12
SV Masking Filter: non masked
VAF Filter: 0-1
Self Molecule Count: 5
% in Paired Control DB: 1
% in Paired Control DB for Enzyme: 1
SV Chimeric Score: pass
Found in Paired Control Sample Assembly: no
Found in Paired Control Sample Molecules: no
Found in Self Molecules: yes
Overlap Genes: all
CNV Filters
Feature CNV Overlap Precision (Kbp): 500
Copy Number Type: all
Copy Number Confidence: 0,99
Copy Number Min Size (bp): 500,000
Copy Number Masking Filter: non masked
Aneuploidy Filters
Aneuploidy Type: all
Aneuploidy Confidence: 0,95
SV Filter Confidence Min Size (bp)
Insertion 0
Deletion 0
Inversion 0,7
Duplication -1
Intra-Fusion 0,05
Inter-Translocation 0,05
Feature File SV CNV Action
hg38 Known Canonical Mapped
Add Feature
Category Count
Insertion 1
Deletion 4
Inversion 1
Duplication 1
Intra-Fusion 0
Inter-Translocation 0
CNV Gain Segment 23
CNV Loss Segment 2
Aneuploidy Gain 0
Aneuploidy Loss 0
Circos Tracks (start from outer rings)
Cytoband
hg38 Known Canonical Mapped Feature
SV track
CNV track
VAF segments
Translocations
E
IL2RG

Figure 2.

was filtered out due to failed chimeric score, i.e., the software could not exclude them to be artifacts. Other events are reported in Table EV3. Having validated the platform performance in detecting a small SV at 2.5% VAF and a larger one at 10% VAF, outperforming the resolution and sensitivity of karyotyping, we were thus confident that we could capture unexpected gene editing events escaping conventional approaches.

High coverage optical genome mapping of edited CD4$^+$ T-cells

We then edited CD4 T-cells from three male donors in our GMP-compliant manufacturing process and analyzed them at the end of the process (day 14). Editing efficiency, as measured by LNGFR expression, was 10–12% in the bulk edited population and 78–84% in the positive selected fraction (Fig 3A). We first performed 100-metaphase karyotype on bulk edited cells, which albeit intrinsically limited in sensitivity and resolution, was considered within limits as non-recurrent anomalies were attributed to sample preparation (Fig 3B and C). Having validated the OGM technology for the intended use, we performed dual OGM analyses of LNGFR-enriched fractions against their respective untreated counterpart are reported in Fig 4A. The length of the intended HDR outcome (3.1 kb) was expected to be below the technique's detection limit of 5 kb. However, an insertion in the *CD40LG* locus was found to be recurring in donors #2 and #3. The 16.9–17 kb insert was called at a VAF of 0.13–0.21 with high confidence (0.99). The most abundant SV (Fig 4B) contained one additional DLE-1 label. Other SVs contained either an additional label in a different position, or additional regularly spaced labels. To ascertain whether the LNGFR-enriched fraction of donor #1 was truly devoid of insertions at the *CD40LG* locus, we checked if the SV was filtered by the software. In fact, five insertions were found at the same locus after filters removal: three had been filtered out due to low molecule counts and two for being also apparently present in the untreated sample. However, no insertion event at the same locus was found upon manual revision of the untreated sample #1. We thus concluded that insertions occurred also with donor #1, but somehow had been incorrectly filtered out by the software.

We used VAF and relative molecule abundance from each sample annotated rare variant analysis to estimate the frequency of such insertion events across every donor and condition (Fig 4C). The 8–10 kb events were the most abundant, and the overall frequency of insertions increased with the percentage of LNGFR$^+$ cells in the sample. We speculated that the new labels present in Fig 4B correspond to recognition of the long terminal repeats (LTRs) present in the donor IDLV, which contain a DLE-1 restriction site. Of note, the resolution of the technique was not sufficient to discriminate between one label and two closely (< 1–2 kb) spaced ones, as could be expected in case of IDLV concatemer formation. We sought to confirm the presence of concatemers with a ddPCR linkage assay. We reasoned that concatenated IDLVs do not segregate independently but rather fall in the same oil droplet, leading to underestimation of IDLV copies in the sample. Enzymatic DNA digestion within the vector genome, outside the target amplicon, before droplet generation indeed separated the tandem repeats of target amplicons, uncovering a median VCN increase of 0.1, indirectly confirming the presence of concatemers Fig 4D. Overall, this analysis revealed that around 20–50% of the edited cells expressing LNGFR did not undergo precise HDR at both sides of the DNA DSB but rather trapped the full donor vector DNA, or its concatemer, at either or both sides (bilateral HDR integration of a concatemer being another theoretical possibility). Supporting data, albeit at very low coverage, was generated with targeted nanopore sequencing (Figs 4E and F, and EV5). Beyond a majority of intended edits with precise bilateral HDR, we sequenced a 9.8 kb insert consisting of a concatemer of two IDLVs joined at an intervening single LTR and integrated by HDR, which reproduced the most frequent size of insert recovered by OGM. Importantly, both healthy donors' and patients' edited cells recapitulated the physiological expression of CD40L and downstream signaling (Fig 4G and H), indicating preserved functionality of also these imprecise editing outcomes.

Other events detected by OGM in the bulk and LNGFR enriched fraction are reported in Table 1. We did not detect recurrent events, translocations, nor events occurring at the previously validated non-HiFi Cas9 off-targets (Vavassori *et al*, 2021). Notably, we also did not detect on-target deletions involving *LNC00892* by optical mapping at the end of the manufacturing process in any edited sample, selected or not, even without event filtering, consistently with their aforementioned low abundance in males, counter-selection in culture, and purging. In conclusion and within the confidence and limits of the adopted technology, our unbiased genome-wide analysis of a putative therapeutic cell product did not reveal any recurrent or concerning SV or CNV outside of the target locus.

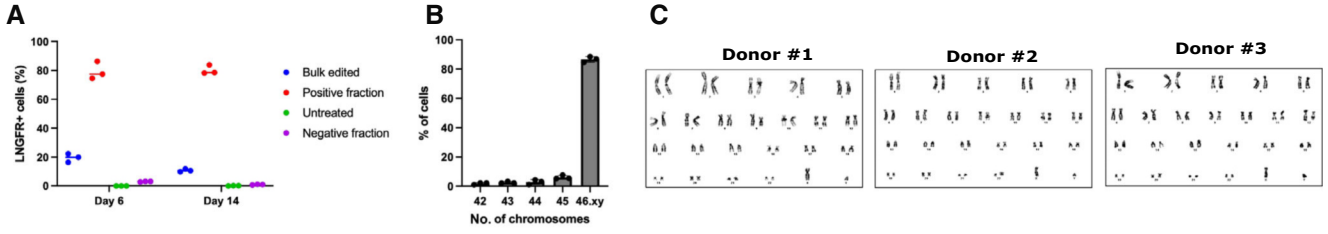

Figure 3. Karyotype of edited CD4$^+$ T-cells.

A LNGFR expression by flow cytometry in CD4$^+$ T-cells at day 6 (day of immunomagnetic enrichment of LNGFR$^+$ cells) and day 14 (end of process).

B, C Results of karyotype analysis performed on three bulk edited biological replicates 14 days after editing and (C) representative images (n = 92, 95 and 91 metaphases respectively for donor 1, 2, and 3).

Source data are available online for this figure.

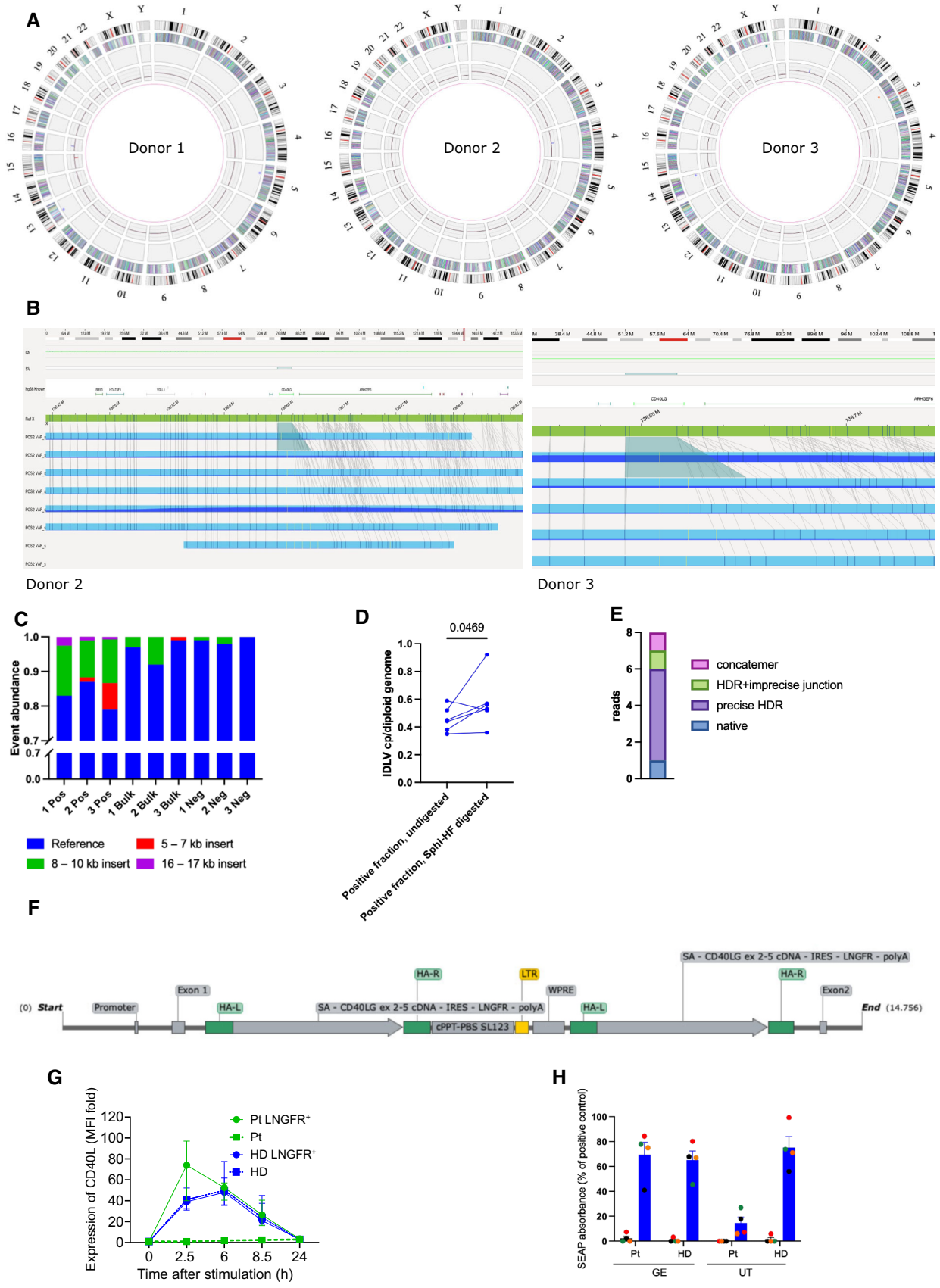
A
Donor 1
Donor 2
Donor 3
B
Donor 2
Donor 3
C
Event abundance
1 Pos
2 Pos
3 Pos
1 Bulk
2 Bulk
3 Bulk
1 Neg
2 Neg
3 Neg
Reference
5 – 7 kb insert
8 – 10 kb insert
16 – 17 kb insert
D
0.0469
IDLV cp/diploid genome
Positive fraction, undigested
Positive fraction, SphI-HF digested
E
reads
concatemer
HDR+imprecise junction
precise HDR
native
F
(0) Start
Promoter
Exon 1
HA-L
SA - CD40LG ex 2-5 cDNA - IRES - LNGFR - polyA
HA-R
LTR
cPPT-PBS SL123
WPRE
HA-L
SA - CD40LG ex 2-5 cDNA - IRES - LNGFR - polyA
HA-R
Exon2
End (14.756)
G
Expression of CD40L (MFI fold)
Pt LNGFR+
Pt
HD LNGFR+
HD
Time after stimulation (h)
H
SEAP absorbance (% of positive control)
Pt
HD
Pt
HD
GE
UT

Figure 4.

Figure 4. Optical mapping and functional characterization of edited CD4$^+$ T-cells.

A Circos plots of LNGFR$^+$ CD4$^+$ T-cells compared with each donor cultured unedited counterpart, annotated as in Fig 2D. The green dot on chromosome X indicates an insertion in the *CD40LG* locus.

B Representation of insertion events around *CD40LG* locus in donor #2 (left) and donor #3 (right). Tracks from top to bottom: (1) Banding of the X chromosome. (2) Copy number variations (green track). (3) Structural variants; green line indicates insertion (also represented by shaded trapezium). (4) Annotated genes. (5) Reference assembly (green, labels in blue). (6) Sample assemblies (light blue, shaded in dark blue according to relative abundance), with labels matching the reference (dark blue) or not (yellow).

C Proportion of insertions at *CD40LG* locus of different color-coded size in bulk, LNGFR-depleted (neg) and enriched (pos) fractions of donors 1, 2 and 3.

D IDLV VCN in the same LNGFR enriched samples of Figs 3A and 4A and three additional LNGFR enriched samples from similar experiments, before and after digestion with SphI. One-tailed Wilcoxon matched-pairs signed rank test.

E, F Distribution and absolute number of on target reads from two samples of gene edited CD4$^+$ T-cells after LNGFR selection and expansion (F) Detailed map of concatemer.

G, H Median fluorescent intensity (MFI) if CD40L in LNGFR+ cells (median-+IQR). H. CD40 downstream signal transduction by SEAP colorimetric assay (mean ± SEM). Each color represents a different experiment. Pt, Patient; HD, healthy donor; UT, untreated and GE, gene edited.

Source data are available online for this figure.

Table 1. Dual analysis by variant annotation pipeline of edited and unedited CD4 T-cells.

#	Sample	chr:position in hg38	Event	Size (kb)	VAF	Notes
Donor 1						
1	Bulk	9:41545212–41552664	Deletion	3	0	Present in 0.6% of control samples database*
2	Bulk/pos	15:75759464–77912434 (bulk)/ 15:75759464–77912434 (pos)	CNV Loss	2,152 (bulk)/2,153 (pos)	0.5 (bulk)/ 0.46 (pos)	Present in UT*
3	Bulk/pos	16:34944870–35898364 (bulk)/ 16:34905599–35614048 (pos)	CNV Gain	953 (bulk)/ 708 (pos)	0.48 (bulk)/0.43 (pos)	Partial CNV mask overlap, near centromere*
4	Bulk	Y:23230658–23264381	Duplication	33	0.01	Masked region, overlapping DAZ2*
5	Pos	5:46874164–46890530	Inverted duplication	16	0.5	Present in 0.6% of control samples database*
6	Pos	13:57143300–57261030	Inverted duplication	117	0.5	Overlapping PRR20A-E, PRRFP. 7 molecules
Donor 2						
7	Bulk/pos	4:91501905–92528369 (bulk)/ 4:91481313–92478681 (pos)	CNV Gain	1,026 (bulk)/997 (pos)	0.42 (bulk)/0.41 (pos)	Present in UT*
Donor 3						
8	Bulk/pos	1:144453902–145230161 (bulk)/ 1:144416847–145264785 (pos)	CNV Gain	776 (bulk)/ 847 (pos)	0.60 (bulk)/0.06 (pos)	Present in UT*
9	Bulk	2:57670638–57700364	Deletion	29	0.02	No annotated genes
10	Pos	3:58006840–58123739	Deletion	104	0.04	Overlapping FLNB, 9 molecules
11	Pos	14:98239205–98427434	Duplication	188	1	Overlapping RN7SL714P. Detected also in bulk with VAF 1 and failed chimeric score, not in negative fraction or UT*

Chr., chromosome; VAF, variant allele frequency; Pos LNGFR enriched cell fraction. Asterisk indicates likely false positives.

Discussion

It is increasingly recognized that editing-induced DNA DSB are not only repaired with small indels originating from NHEJ and microhomology mediated end joining (MM-EJ) or by HDR but may also result in long-range deletions stemming from the DNA DSB and propagating towards one or both ends of the chromosome. These adverse on-target events, possibly aggravated by the recurring action of the nuclease on the target site, have likely been underestimated due to the loss of the primer binding sites used to evaluate the frequency of NHEJ/MM-EJ and HDR, and are in our setting exponentially more frequent than off-target indels. Although in gene correction settings like this one, where the target gene is already non-functional, deletions limited to the gene itself are not expected *a priori* to be problematic, larger ones reaching surrounding, functional genes are more concerning. By tiling the regions of interest near the nuclease target site we found deletion events of at least 7 kb reaching the neighboring genes that were spontaneously counter-selected over time, providing another mean by which such events may escape detection. Although the accuracy and precision

of ddPCR CNV assays are somewhat limited when analyzing heterogeneous populations, they can be suitable for routine product characterization to exclude unexpected clonal expansion stemming for example from the deletion of an oncosuppressor gene or truncation of a proto-oncogene. We showed reproducible occurrence of large deletions at different target sites and robustness of the assay design across different hematopoietic cell types; single-CFU analysis, when possible, increases the granularity and sensitivity of the analysis, overcoming the limits imposed by stochastic error in analyzing bulk populations. In principle, both on-target and off-target genome integrity – as long as the off-targets are known – can be quantitatively assessed with this kind of design, tiling these loci as much as desired. Conventional assays such as the T7 endonuclease 1 mismatch detection assay, Tracking of Indels by Decomposition (TIDE) and targeted NGS assess the cutting efficiency of the nuclease as the proportion of modified amplicons with respect to unmodified ones (Sentmanat *et al*, 2018); if primer binding sites are deleted, the deletion itself is missed. Linkage-based ddPCR assays are more robust (Nahmad *et al*, 2022), but still cannot capture large deletions spanning both targets or neighboring genes. Instead, because our assays are designed to report the lack rather than the presence of PCR primer binding sites, they do not risk underestimating the frequency of deletions. Their relative simplicity has the advantage of being routinely applicable; the drawback is that as DNA "mileposts" they detect deletions of a defined length and not the exact extent of each deletion event. Placing ddPCR assays mileposts on relevant regions, like annotated genes or regulatory elements, is in our opinion a reasonable strategy for clinical translation. Of note, the presence of large deletions raises the issue of quantifying translocation and inversion events with heterogeneous junctions, arising from the joining of DNA DSBs that have undergone long range deletions, which cannot be easily done by conventional ddPCR assays.

Precautionary principles suggest that unanticipated outcomes are to be investigated with the least possible degree of bias. While 100-metaphase karyotyping remains a routine and necessary first step, its low sensitivity and resolution might not be considered sufficient when applied to a novel genetic engineering technology for clinical use. High coverage optical mapping promises unbiased, amplification-free detection of CNVs and rare genomic SVs greater than 5 kb in a very large number of cells, with the obvious downside of not providing sequence information. Indeed, we confirmed sensitivity of the platform in detecting a hemizygous 6.9 kb insertion at 2.5% VAF and a large hemizygous genomic rearrangement of about 300 kb at 10% VAF. As for other technologies, masked regions and filter settings are intrinsic limits and potential caveats. We noticed that integration of karyotype analysis and optical mapping results can be helpful in detection of events that cannot be solved separately by each platform.

We analyzed $CD4^+$ T-cells edited with Cas9 and IDLV in optimized conditions modeling a clinical applicable process and both karyotype and OGM yielded reassuring results, with no detection of on-target deletions nor translocations involving the off-target sites previously detected with non-HiFi Cas9 after expansion in culture (Vavassori *et al*, 2021). As we were unable to generate stable lymphoblastoid clones carrying deletions encompassing *CD40LG*, and the editing induced deletion events were purged over time, we speculate that cells harboring large deletions in this region are not viable. In principle, deletions of heterogeneous length could still be present at low frequency and not be detected by optical mapping; still, even if present, cells harboring them would have not undergone clonal expansion reaching the limit of detection of the technology. Of note, we validated the accuracy of *ARHGEF6* and *LNC00892* CNV assays in detecting deletions present with VAF of 5 and 10%, respectively; while this detection limit is higher than for OGM, nevertheless it may be more robust in detecting events of heterogeneous length that individually would not be captured by the latter.

Interestingly, we were able to further detail the heterogeneity of repair outcomes of the DNA DSB, while also estimating their relative frequency. Imprecise HDR, whereby a longer than expected template is incorporated into the genome, appears to be a relatively common event. Our data indicate that these integrations represent IDLV templates containing LTRs, often arranged in concatemers, complementing observations in other settings (Ferrari *et al*, 2022). A significant fraction of these editing events are functional, as indicated by enrichment through LNGFR, which by design is coupled to faithful reconstitution of CD40L expression. Characterization of these large SVs is a clear advantage of OGM, whose interpretation is facilitated by the presence of the DLE-1 target sequence within the vector LTR, and can guide targeted long read sequencing. Given the currently prohibitive costs of routine application of high coverage OGM on each cellular product, we believe that this technology is currently fit for non-routine characterization and may guide in the choice of custom, simpler ddPCR and functional assays.

We cannot draw definitive conclusions concerning the few non-recurring genomic events found in edited $CD4^+$ T-cells. Most appear to be false positives, either because they were also found in the untreated controls (events 2, 7, 8), in healthy subjects (events 1, 5), falling into masked regions (events 3, 4) or at unreasonably high VAF (event 11). A low-scoring CRISPOR-predicted off-target could only be found in the region involved in event 11, and it is hard to reconcile with the VAF of 1 observed in both bulk and LNGFR enriched fraction. For events 6, 9, and 10, it is currently impossible to determine whether they might be truly consequences of editing, but the likelihood appears to be low, given the absence of predicted/validated off-targets in the region and the very lax similarity with the gRNA. Of note, we found an even higher number of false positive events with high VAF, not compatible with the sample admixing rate, when comparing the *ad hoc* edited validation clone with its unmanipulated counterpart.

In summary, for the purpose of clinical translation of our HIGM1 gene therapy, we believe that the optical mapping analysis has provided a reasonable body of data to support the inclusion of only few and selected *ad hoc* ddPCR assays in the final product quality control panel. In the future, novel single cell DNA assays and sequencing studies may possibly elucidate more granularly the genomic outcomes following gene editing.

From a more general perspective, the frequent on-target occurrence of large deletions and template concatemers reported here are findings that must be accounted for when designing integration strategies. Combining orthogonal technologies may be required to reliably assess them, overcoming the limitations of each technology. In principle, the protocol outlined here is applicable to other cell types and tissue samples, as long as sufficient high molecular weight DNA can be successfully extracted for OGM. Given the relatively high occurrence of these events, we suggest evaluation of their nature and proportion in every prospective clinical application of cell therapy that exploits nucleases, and possibly nickases.

Indeed, while counterselection of large deletions is likely the rule due to impaired cellular fitness, deletions and translocations are signatures of cancer, including T lymphoblastic leukemia, and could cause transformation (Girardi *et al*, 2017). Generally speaking, the functional consequences of deletions are expected to depend on the genes that are involved, and thus specific of the target locus, and secondarily on the individual genetic landscape. While *ex vivo* gene therapy allows for quality control of the drug product before infusion, in the context of *in vivo* gene editing events can only be detected after treatment. Clinical implementation of *in vivo* gene editing must thus be balanced against this potential risk.

Wen *et al* (2021) previously reported a reduced proportion of large deletions in $CD4^+$ T-cells 3 days after editing with AAV6 at MOI 1 k. In our setting with MOI 50 k, deletions were most abundant 6–48 h after editing, and the proportion of deletions diminished over time in culture, mirroring the increase in HDR, as correctly edited cells likely outcompeted those with deletions. Differences in experimental setup (preprint: Tsuchida *et al*, 2023), competition dynamics between HDR-edited cells and those bearing large deletions and timing of analysis could explain these opposite findings; irrespectively of the cause they reiterate the need for implementation of this kind of analytical approach as presence of a donor template does not guarantee absence of large deletions.

Cell cycle is a major driver of adverse genomic outcomes following nuclease-induced DSB (Cullot *et al*, 2023); unfortunately for HDR applications cell stimulation and cell cycle progression remain a requirement. In our setting, targeting a hemizygous locus and including a selection step coupled to successful editing outcome allows generation of an highly enriched population of cells with rescued physiological regulation and function of the edited gene. The reported genome-wide unbiased validation of preserved genomic integrity supports clinical translation with a favorable risk–benefit profile of one of the first applications of long-range HDR mediated gene editing.

Materials and Methods

Culture and gene editing of $CD4^+$ T-cells with AAV6

Culture and editing was performed as previously described (Vavassori *et al*, 2021) with the following modifications. White blood cells were isolated from buffy coats using Human Lympholyte® Cell Separation Media (Euroclone) and SepMate™-50 (Stemcell Technologies), according to the latter's instructions. $CD4^+$ lymphocytes cells were isolated by negative selection with $CD4^+$ T Cell Isolation Kit, human (Miltenyi Biotec). Cells were cultured X-Vivo + 100 IU/ml penicillin + 100 μg/ml streptomycin (Lonza, cat. no. DE17-602E), IL-15 5 ng/ml (Peprotech), IL-7 5 ng/ml (Peprotech), and activated with Dynabeads™ Human T-Activator CD3/CD28 for T Cell Expansion and Activation (ThermoFisher Scientific). A bicistronic AAV6 containing exon 2–5 *CD40LG* cDNA and LNGFR under EIF1alfa short constitutive promoter was used for initial experiments.

Culture and gene editing of $CD4^+$ T-cells with IDLV

Briefly, $CD4^+$ T-cells were isolated from buffy coats of healthy donors or, for patients' samples, from whole blood. $CD4^+$ T-cells were cultured in X-Vivo 15 (Lonza) supplemented with 0.5% Human Serum Albumin (HSA, Baxalta), 1% Pen/Strep (Lonza), 100 IU/ml IL7 and 200 IU/ml IL15 (both from Miltenyi Biotec) and stimulated with T Cell TransAct (Miltenyi Biotec) at 2×10^6 cells/ml. On day 2 cells were transduced at 1×10^6 cells/ml with CD40LG-IRES_LNGFR-IDLV at MOI 40, and after 8 ± 1 h electroporated with SpyFi™ high fidelity (HiFi) Cas9 complexed with a single guide RNA (gRNA) targeting the first intron of *CD40LG*. On day 6 cells were optionally enriched by immunomagnetic CD271 beads (Mitenyi Biotec) (Asperti *et al*, 2023).

HSPCs cell culture

CB $CD34^+$ HSPCs were thawed and cultured in StemSpan medium (serum free, StemCell Technologies) with 100 IU/ml penicillin, 100 μg/ml streptomycin, 2% glutamine, 100 ng/ml hSCF (PeproTech), 100 ng/ml hFlt3-L (PeproTech), 20 ng/ml hTPO (PeproTech) and 20 ng/ml hIL-6 (PeproTech), 10 μM PGE2 (at the beginning of the culture, Cayman), 1 μM SR1 (Biovision) and 50 nM UM171 (STEMCell Technologies) (5% CO_2, 37°C).

Moblized peripheral blood $CD34^+$ HSPCs were thawed and cultured in StemSpan medium (serum free, StemCell Technologies) with 100 IU/ml penicillin, 100 μg/ml streptomycin, 2% glutamine, 300 ng/ml hSCF, 300 ng/ml hFlt3-L, 100 ng/ml hTPO and 10 μM PGE2 (at the beginning of the culture), 1 μM SR1 and 35 nM UM171 (5% CO_2, 37°C).

Editing of human HSPCs

The detailed gene editing protocol can be found in Ferrari *et al* (2021). Briefly, CB and mPB HSPCs were stimulated for 3 days in culture and then gene edited by electroporation (P3 Primary Cell 4D-Nucleofector X Kit Lonza, program EO-100) with 1.25–2.5 μM RNPs or 75 pmol exon 1 *B2M* sgRNA and 7.5 μg Cas9 mRNA. Were indicated, 15 min after electroporation cells were transduced with AAV (2×10^4 vg/cell).

All gRNAs were designed with CRISPOR (Concordet & Haeussler, 2018). AAVS1 gRNA: GTCACCAATCCTGTCCCTAG, B2M exon 1 gRNA: AGTGGAGGCGTCGCGCTGGC, IL2RG gRNA: ACTGGCCATT ACAATCATGT.

Clonogenic assay

Colonies were plated 24 h after electroporation at the density of 400–800 cells in 1.5 ml MethoCult H4434 (StemCell Technologies), three technical replicates/condition. After 14 days single colonies were manually picked and freezed at −80°C.

mRNA *in vitro* transcription

All mRNA constructs have been *in vitro* transcribed following previously published protocol (Schiroli *et al*, 2019; Ferrari *et al*, 2020).

Culture of SKW 6.4 cells

SKW6.4 (ATCC-TIB-215TM lot 5082527) human B lymphoblasts were cultured in RPMI 1640 medium (Corning), 2% glutamine, 100 IU/ml penicillin, 100 μg/ml streptomycin, 10–20% heat-inactivated fetal bovine serum (FBS, Euroclone).

Enzymatic digestion and purification

40 μg of plasmid DNA were digested overnight at 37°C with 100 U of Kpn1-HF (New England Biolabs) and 100 U of Sca1-HF (New England Biolabs) in rCutsmart buffer 1×. The desired band was excised from 1% agarose gel and purified with NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel) according to manufacturer's instructions.

Editing of SKW 6.4 cell line

Briefly, SKW6.4 were edited at IL2RG intron 1 as previously described (Schiroli *et al*, 2019) using electroporation code for K562 and SF Cell Line 4D-Nucleofector™ X Kit S (Lonza).

gRNAs targeting ARHGEF6 and GPR112 were designed with CRISPOR (Concordet & Haeussler, 2018). The following gRNAs were used to target ARHGEF6: #1 CTCTCCCCAAAAGCCGTCAA(AGG), #2 GTATAATAATTCTTGGTAAG(TGG), #3 TACCAAGAATTATTATA CTG(TGG). The following gRNAs were used to target GPR112: #1 AAATAGCGTTGCCCCATACA(CGG), #2 TGTATACTTGCTCGGGTG AT(GGG), #3 CCCATACACGGTCCTGAAAC(TGG). Each gRNA targeting ARHGEF6 was electroporated with each gRNA targeting GPR112 and vice versa. Single cell clones were assessed for CNV of the genes contained within the targeted region (*LNC00892* or *ARHGEF6*).

Sorting and freezing of SKW 6.4 clones

Admixing of clone #2E04 and clone #2 was performed by counting single cells with BD FACSAria Fusion (BD Biosciences), as TC20 Automated Cell Counter (Bio-Rad) coefficient of variation was experimentally determined to be too high (> 15%). Cells, eluted in 1.5 ml of sheath fluid, were sorted in cryovials containing 1.5 ml of heat inactivated fetal bovine serum. Dimethyl sulfoxide (DMSO, D8418, Sigma) was added to a final concentration of 10%. Cells were frozen with KryoPlaner 560–16 Software version v5.26 (Planer Limited, Sunbury-On-Thames, UK) and stored in nitrogen vapors.

DNA extraction

Genomic DNA was extracted with QIAamp DNA Micro Kit (QIAGEN) or KingFisher Flex (Thermo Scientific) according to manufacturers' instructions. Single colonies were lysed with QuickExtract (Epicentre), as indicated.

Sequencing

Primers for sequencing of IL2RG intron 1 have been previously reported (Schiroli *et al*, 2017). The following primers were used for amplification and sequencing:

Application	Primer	Sequence
SKW 6.4 clone #23 5′ junction	Forward	CACCCTCTGTAAAGCCCTGG
	Reverse	AAGACGGCATGGGGTTGGGT
SKW 6.4 clone #23 3′ junction	Forward	TTCATTCAGCTCCGGTTCCC
	Reverse	TCATCTCCCCTCAACCGACT

ddPCR assays

For digital droplet PCR (ddPCR) analyses we analyzed 5–50 ng of gDNA per reaction with the QX200 Droplet Digital PCR System (Bio-Rad), with the following assays: ARHGEF9 dHsaCP2506327/dHsaCP2506706; ARHGEF6 dHsaCNS170742226; MECP2 dHsaCP1000579; EIF2C1 dHsaCP000002; EPS8L1 dHsaCNS737827302 (Bio-Rad), LNC00892 Hs04111584_cn, ARHGEF6 Hs00889052_cn (ThermoFisher). TTC5 dHsaCP2506733 or dHsaCP2506310 (Bio-Rad) was used for normalization. The following custom assays were also used:

ddPCR assays for SKW 6.4

Application	Primer	Sequence (5′–3′)
GFP	Forward	CAGCTCGCCGACCACTA
	Reverse	GGGCCGTCGCCGAT
	Probe	CCAGCAGAACACCCCC (MGBNFQ/FAM)
TELO	Forward	GGCACACGTGGCTTTTCG
	Reverse	GGTGAACCTCGTAAGTTTATGCAA
	Probe	TCAGGACGTCGAGTGGACACGGTG (TAMRA/VIC)
IDLV	Forward	TCACTCCCAACGAAGACAAGATC
	Reverse	GAGTCCTGCGTCGAGAGAG

ddPCR assays for HSPCs CFUs

Application	Primer	Sequence (5′–3′)	Normalizer
Human *AAVS1* ddPCR > Chr 19	(HEX)	PrimePCR ddPCR Copy Number Assay: AAVS1, Human (Biorad)	Human *TTC5* (ddPCR Assay, Biorad, dHsaCP2506310, FAM)
IL2RG deletions TALE ddPCR	FW	AGGTGGTTGAGAATGGTGCT	Human *TTC5* (HEX) or Human *GAPDH* (q Assay, Applied Biosystems, HS00483111_CN VIC)
	RV	CTGTGTGACACGGGCTAAGT	
	PROBE (FAM)	CCTCAGCCCACCTAGATCCT GGG	
AAVS1 deletions TALE ddPCR	FW	ATCCTGGGAGGGAGAGCTT	Human *TTC5* (HEX)
	RV	GGATCAGTGAAACGCACCAG	
	PROBE (FAM)	AGCTGCTCTGACGCGGCCGT	
B2M deletions exon1, 5′	FW	GCCGATGTACAGACAGCAAA	Human *TTC5* (HEX)
	RV	GCAGTGCCAGGTTAGAGAGA	
	PROBE (FAM)	CTCACCCAGTCTAGTGCATGCC TTCT	
B2M deletions exon1, 3′	FW	CCGTGACTTCCCTTCTCCAA	Human *TTC5* (HEX)
	RV	GATCCAGCCCTGGACTAGC	
	PROBE (HEX)	TCTCCTTGGTGGCCCGCCGT	

Thermal conditions for commercial ddPCR assays were 95°C × 10′, (94°C × 30″, 60°C × 1′) × 40 cycles, 98°C × 10′. The assay reported in Fig 1R was performed using Nanoplate 26 K 24-well plates and the QIAcuity Four Platform System (Qiagen) using the following thermal protocol: priming, 95°C × 2′, (95°C × 30″, 60°C × 1′) × 40 cycles.

To estimate the frequency of concatemers, samples were digested or not with SphI-HF in rCutsmart buffer (R3182S, New England Biolabs) for 1 h at 37°C and then heat inactivated for 20 min at

80°C. ddPCR assay was performed in parallel on digested and undigested samples using IDLV primers with EvaGreen reagents (mastermix and oil). The concentration of TELO amplicons was used as reference. Thermal conditions were 95°C × 5′, (95°C × 30″, 63°C × 1′) × 40 cycles, 4°C × 5′, 90°C × 5′. Copy numbers and confidence intervals were calculated with QXManager version 1.2 (Bio-Rad).

Quantification of indels and HDR at *CD40LG* locus

Previously described in Vavassori *et al* (2021).

High coverage optical mapping

After cell thawing (2 min at 37°C), DNA was extracted according to Bionano Prep SP Frozen Cell Pellet DNA Isolation Protocol v2-30398, Rev B, from the SP Blood & Cell Culture DNA Isolation Kit. DNA was labeled according to Bionano Prep Direct Label and Stain (DLS) Protocol-30206, Rev G and loaded on the chip the day after.

Target coverage was 1,600× (effective coverage 1,411.77×–1,491.58×) for all samples except untreated SKW 6.4 human B lymphoblastoid cells (effective coverage 397×), which was run with lower target coverage. The Rare Variant Pipeline (RVP) and Variant Calling were executed on Bionano Solve software (v3.7). Reporting and direct visualization of structural variants was done on Bionano Access (v1.7), using Hg38 as reference genome. Each sample was compared against its unedited, cultured counterpart in paired analysis. The following filters were used: (i) SV confidence: recommended. (ii) Non masked SV only (iii) SV in less or equal to 1% of the paired control db samples. (iv) SV found in self molecules (v) SV in less than or equal to 1% of paired control db samples (vi) SV not found in paired control assemblies (vii). SV nor found in paired control molecules (viii). CNV filters: all variants, recommended confidence, min size 500,000 bp, non masked only (ix). Aneuploidy: all, recommended confidence scores. Events were annotated after filtering. The frequency of insertions of different lengths at the *CD40LG* locus was estimated from the unpaired rare variant analysis by multiplying the VAF for the relative molecule abundance of each event.

Nanopore sequencing

DNA from 10 mln NGFR enriched gene edited $CD4^+$ T-cells (5–10 μg) was extracted with Monarch® HMW DNA Extraction Kit for Cells & Blood (T3050S, New England Biolabs) and processed with Ligation sequencing gDNA – Cas9 enrichment kit (SQK-CS9109, Oxford Nanopore Technologies) according to manufacturing instructions using the following gRNAs: GGTCATGACCGGCCATAGAG-TGG ACCTCGGTGACAGTACAGGA-GGG TGAATCACCCAATGAAACCA-TGG CAACAATTGAGGTACAAGCA-GGG. DNA was sequenced on MinION Mk1B with a FLO-MIN106 flow cell. Reads were aligned using EPI2M wf-alignment pipeline v0.5.0234983b.

Murine experiments

All the mice were maintained in specific pathogen-free (SPF) conditions, and all animal procedures were designed and performed with the approval of the Animal Care and Use Committee of the San Raffaele Hospital (IACUC #818) and communicated to the Ministry of Health and local authorities according to Italian law.

Functional assays

After thawing and overnight resting, $CD4^+$ T-cells were activated using phorbol myristate acetate (PMA, 10 ng/ml Sigma) plus ionomycin (500 ng/ml, Sigma) for 5 h and surface expression of CD40L was followed over 2.5–24 h. T-cells were stained with a panel of antibodies including monoclonal mouse anti-human CD271, CD3, CD4, CD154. CD40L fold expression or CD40 binding were calculated as follows: median fluorescence intensity (MFI) of the stimulated sample/MFI of the unstimulated sample.

The ability of CD40L expressing cells to trigger CD40 downstream signaling was evaluated by co-culturing cells with HEK-BlueTM CD40 cells (InvivoGen). $CD4^+$ T-cells from both HDs and patients were stimulated with PMA/Ionomycin or left unstimulated. After 2.5 h $CD4^+$ T-cells were harvested, washed and co-cultured in 1:1 ratio with HEK-BlueTM CD40 cells for 24 h in presence of anti hIL-1beta (Invivogen) and anti-hTNF-alfa (InvivoGen). Recombinant human CD40L (InvivoGen) and unstimulated T-cells were used respectively as positive and negative control. After 24 h, secreted SEAP was measured in the supernatant by QUANTI-Blue solution (InvivoGen). The absorbance was read by Omega reader (BMG Labtech) at 650 nm.

Flow cytometry

50–500 k cells or 50 μl of PB were incubated with antibodies for 10 min at 4°C and then washed with PBS (Corning) + 2% heat inactivated FBS (Euroclone). At least 20,000 events were acquired with BD FACSCanto II (Becton Dickinson) or Cytoflex (Beckman Coulter). Data was analyzed with FCS Express 7 Research (*De Novo* Software), FlowJo software (Tree Star).

Antibodies

Target	Fluorochrome	Vendor	Clone	Cat. No.	Dilution
CD3	PeCy7	Biolegend	HIT3a	300316	5:100
CD3	FITC	BD Pharmingen	SK7	345763	1:33
CD4	PB	BD Pharmingen	RPA-T4	558116	1:100
CD4	Viogreen	Miltenyi biotec	REA 623	130-113-230	1:100
NGFR	APC	Miltenyi biotec	ME20.4-1.H4	130-113-418	2:100
NGFR	AF647	BD Pharmingen	C40-1457	560326	1:25
CD154	PE	Invitrogen	24-31	12-548-42	5:100
CD154	PE	Biolegend	24-31	310806	1:50
7-AAD Viability Staining Solution	7-AAD	Biolegend	-	420404	1:100

Statistical analysis

Non-parametric statistical analyses were performed using Prism 9 (GraphPad Software), as reported in figure legends.

Human subjects research

Informed consent for biological samples' collection and anonymized biological sample/data sharing for HIGM1 patients were obtained by the referring physician, according to local research protocols, reviewed and approved by local ethics committees or institutional review board (IRB). Buffy coats were obtained as anonymized residues of blood donations, used upon signature of specific institutional informed consent for blood product donation by healthy blood donors.

Data availability

For original data please contact canarutto.daniele@hsr.it. Additional methods may be found in a data supplement available with the online version of this article. OGM data (PRJEB66171) is available at https://www.ebi.ac.uk/ena/browser/home.

Expanded View for this article is available online.

Acknowledgements
We wish to thank all patients; the following key opinion leaders, internal and external collaborators: Pietro Genovese (Harvard Medical School), Harry Malech (National Institutes of Health), Suk See DeRavin (National Institutes of Health), Isabelle Meyts (University Hospitals Leuven), Arjan C. Lankester (Leiden University Medical Center), Andrew R. Gennery (Newcastle University), Sandro Plebani (Università di Brescia), Pere Soler Palacín (Vall d'Hebron Hospital), Luigi D. Notarangelo (National Institutes of Health), Alessandro Aiuti (Università Vita-Salute San Raffaele), for guidance; Akiva Zablocki and the Hyper IgM foundation for patient referral; present and past SR-TIGET members including Tiziana Plati, Elisabetta Mercuri, Luisa Albano, Chiara Gaddoni, Mariasole Vespasiano, Maria Carmina Castiello, Martina Fiumara, Paola Albertini, Francesca Cecere, and the SR-TIGET GLP team, and San Raffaele Flow Cytometry facility (FRACTAL) for technical support; Francesca Sanvito for collaboration with the project; Ivan Merelli and Stefano Beretta (SR-TIGET) for informatic support; Cristina Montagna (Albert Einstein College of Medicine), Marzia Rossato e Massimo Delledonne (Università di Verona) for brainstorming sessions on genome integrity; Francesca Giannese of San Raffaele Center of OMICS Sciences for technical support with Oxford Nanopore, Alessandro Nonis (Centro Universitario di Statistica per le Scienze Biomediche) for statistical support; the Genespire team for support. Some schematics were created using BioRender.com. This work was supported by Fondazione Telethon ETS (TIGET Core Grant support to LN), E-Rare-3-JTC 2017 to LN, and Genespire, a startup company aiming to develop *ex vivo* gene editing in genetic diseases. The present work was performed by D.C. in partial fulfillment of the requirements for obtaining the PhD degree at Vita-Salute San Raffaele University, Milano, Italy. Open access funding provided by BIBLIOSAN.

Author contributions
Daniele Canarutto: Conceptualization; data curation; formal analysis; investigation; methodology; writing – original draft; writing – review and editing. **Claudia Asperti:** Conceptualization; investigation; visualization; writing – original draft. **Valentina Vavassori:** Investigation. **Simona Porcellini:** Investigation; visualization. **Elisabetta Rovelli:** Investigation. **Marianna Paulis:** Investigation; visualization. **Samuele Ferrari:** Investigation. **Angelica Varesi:** Investigation. **Martina Fiumara:** Investigation. **Aurelien Jacob:** Investigation. **Lucia Sergi Sergi:** Resources. **Ilaria Visigalli:** Investigation. **Francesca Ferrua:** Investigation; resources. **Luis Ignacio González-Granado:** Investigation; resources. **Vassilios Lougaris:** Investigation; resources. **Andrea Finocchi:** Investigation; resources. **Anna Villa:** Supervision. **Marina Radrizzani:** Supervision. **Luigi Naldini:** Supervision; writing – review and editing.

Disclosure and competing interests statement
LN, AV, SF, SP, DC, MF and AJ are inventors of patent applications owned by Ospedale San Raffaele and Fondazione Telethon, including one patent application on *CD40LG* gene editing. LN is founder, quota holder and consultant of Genespire. LN holds grants and funding to develop a therapy against HIGM1.

References

Adikusuma F, Piltz S, Corbett MA, Turvey M, McColl SR, Helbig KJ, Beard MR, Hughes J, Pomerantz RT, Thomas PQ (2018) Large deletions induced by Cas9 cleavage. *Nature* 560: E8 – E9

Asperti C, Canarutto D, Porcellini S, Sanvito F, Cecere F, Vavassori V, Ferrari S, Rovelli E, Albano L, Jacob A *et al* (2023) Scalable GMP-compliant gene correction of CD4+ T cells with IDLV template functionally validated in vitro and in vivo. *Mol Ther Methods Clin Dev* 30: 546 – 557

Boutin J, Rosier J, Cappellen D, Prat F, Toutain J, Pennamen P, Bouron J, Rooryck C, Merlio JP, Lamrissi-Garcia I *et al* (2021) CRISPR-Cas9 globin editing can induce megabase-scale copy-neutral losses of heterozygosity in hematopoietic cells. *Nat Commun* 12: 4922

Brown MP, Topham DJ, Sangster MY, Zhao J, Flynn KJ, Surman SL, Woodland DL, Doherty PC, Farr AG, Pattengale PK *et al* (1998) Thymic lymphoproliferative disease after successful correction of CD40 ligand deficiency by gene transfer in mice. *Nat Med* 4: 1253 – 1260

Concordet J-P, Haeussler M (2018) CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res* 46: W242 – W245

Cullot G, Boutin J, Fayet S, Prat F, Rosier J, Cappellen D, Lamrissi I, Pennamen P, Bouron J, Amintas S *et al* (2023) Cell cycle arrest and p53 prevent ON-target megabase-scale rearrangements induced by CRISPR-Cas9. *Nat Commun* 14: 4072

de la Morena MT, Leonard D, Torgerson TR, Cabral-Marques O, Slatter M, Aghamohammadi A, Chandra S, Murguia-Favela L, Bonilla FA, Kanariou M *et al* (2017) Long-term outcomes of 176 patients with X-linked hyper-IgM syndrome treated with or without hematopoietic cell transplantation. *J Allergy Clin Immunol* 139: 1282 – 1292

Ferrari S, Jacob A, Beretta S, Unali G, Albano L, Vavassori V, Cittaro D, Lazarevic D, Brombin C, Cugnata F *et al* (2020) Efficient gene editing of human long-term hematopoietic stem cells validated by clonal tracking. *Nat Biotechnol* 38: 1298 – 1308

Ferrari S, Beretta S, Jacob A, Cittaro D, Albano L, Merelli I, Naldini L, Genovese P (2021) BAR-Seq clonal tracking of gene-edited cells. *Nat Protoc* 16: 2991 – 3025

Ferrari S, Jacob A, Cesana D, Laugel M, Beretta S, Varesi A, Unali G, Conti A, Canarutto D, Albano L *et al* (2022) Choice of template delivery mitigates the genotoxic risk and adverse impact of editing in human hematopoietic stem cells. *Cell Stem Cell* 29: 1428 – 1444

Ferrua F, Galimberti S, Courteille V, Slatter MA, Booth C, Moshous D, Neven B, Blanche S, Cavazzana M, Laberko A *et al* (2019) Hematopoietic stem cell transplantation for CD40 ligand deficiency: Results from an EBMT/ESID-IEWP-SCETIDE-PIDTC study. *J Allergy Clin Immunol* 143: 2238 – 2253

Frangoul H, Altshuler D, Cappellini MD, Chen Y-S, Domm J, Eustace BK, Foell J, de la Fuente J, Grupp S, Handgretinger R *et al* (2021) CRISPR-Cas9 gene

editing for sickle cell disease and β-thalassemia. *N Engl J Med* 384: 252 – 260

Fu B, Liao J, Chen S, Li W, Wang Q, Hu J, Yang F, Hsiao S, Jiang Y, Wang L *et al* (2022) CRISPR–Cas9-mediated gene editing of the BCL11A enhancer for pediatric β0/β0 transfusion-dependent β-thalassemia. *Nat Med* 28: 1573 – 1580

Girardi T, Vicente C, Cools J, De Keersmaecker K (2017) The genetics and molecular biology of T-ALL. *Blood* 129: 1113 – 1123

Hubbard N, Hagin D, Sommer K, Song Y, Khan I, Clough C, Ochs HD, Rawlings DJ, Scharenberg AM, Torgerson TR (2016) Targeted gene editing restores regulated CD40L function in X-linked hyper-IgM syndrome. *Blood* 127: 2513 – 2522

Kuo CY, Long JD, Campo-Fernandez B, de Oliveira S, Cooper AR, Romero Z, Hoban MD, Joglekar AV, Lill GR, Kaufman ML *et al* (2018) Site-specific gene editing of human hematopoietic stem cells for X-linked hyper-IgM syndrome. *Cell Rep* 23: 2606 – 2616

Leibowitz ML, Papathanasiou S, Doerfler PA, Blaine LJ, Sun L, Yao Y, Zhang C-Z, Weiss MJ, Pellman D (2021) Chromothripsis as an on-target consequence of CRISPR–Cas9 genome editing. *Nat Genet* 53: 895 – 905

Nahmad AD, Reuveni E, Goldschmidt E, Tenne T, Liberman M, Horovitz-Fried M, Khosravi R, Kobo H, Reinstein E, Madi A *et al* (2022) Frequent aneuploidy in primary human T cells after CRISPR–Cas9 cleavage. *Nat Biotechnol* 40: 1807 – 1813

Notarangelo LD, Giliani S, Plebani A (2014) CD40 and CD40 ligand deficiencies. In *Primary Immunodeficiency Diseases*, Ochs HD, Smith CIE, Puck JM (eds), pp 324 – 342. Oxford: Oxford University Press

Park SH, Cao M, Pan Y, Davis TH, Saxena L, Deshmukh H, Fu Y, Treangen T, Sheehan VA, Bao G (2022) Comprehensive analysis and accurate quantification of unintended large gene modifications induced by CRISPR-Cas9 gene editing. *Sci Adv* 8: eabo7676

Sacco MG, Ungari M, Catò EM, Villa A, Strina D, Notarangelo LD, Jonkers J, Zecca L, Facchetti F, Vezzoni P (2000) Lymphoid abnormalities in CD40 ligand transgenic mice suggest the need for tight regulation in gene therapy approaches to hyper immunoglobulin M (IgM) syndrome. *Cancer Gene Ther* 7: 1299 – 1306

Sasu BJ, Opiteck GJ, Gopalakrishnan S, Kaimal V, Furmanak T, Huang D, Goswami A, He Y, Chen J, Nguyen A *et al* (2022) Detection of chromosomal alteration after infusion of gene-edited allogeneic CAR T cells. *Mol Ther* 31: 676 – 685

Schiroli G, Ferrari S, Conway A, Jacob A, Capo V, Albano L, Plati T, Castiello MC, Sanvito F, Gennery AR *et al* (2017) Preclinical modeling highlights the therapeutic potential of hematopoietic stem cell gene editing for correction of SCID-X1. *Sci Transl Med* 9: eaan0820

Schiroli G, Conti A, Ferrari S, Della Volpe L, Jacob A, Albano L, Beretta S, Calabria A, Vavassori V, Gasparini P *et al* (2019) Precise gene editing preserves hematopoietic stem cell function following transient p53-mediated DNA damage response. *Cell Stem Cell* 24: 551 – 565

Sentmanat MF, Peters ST, Florian CP, Connelly JP, Pruett-Miller SM (2018) A survey of validation strategies for CRISPR-Cas9 editing. *Sci Rep* 8: 888

Silva M, de Leeuw N, Mann K, Schuring-Blom H, Morgan S, Giardino D, Rack K, Hastings R (2019) European guidelines for constitutional cytogenomic analysis. *Eur J Hum Genet* 27: 1 – 16

Smith AC, Neveling K, Kanagal-Shamanna R (2022) Optical genome mapping for structural variation analysis in hematologic malignancies. *Am J Hematol* 97: 975 – 982

Tangye SG, Al-Herz W, Bousfiha A, Chatila T, Cunningham-Rundles C, Etzioni A, Franco JL, Holland SM, 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

Tsuchida CA, Brandes N, Bueno R, Trinidad M, Mazumder T, Yu B, Hwang B, Chang C, Liu J, Sun Y *et al* (2023) Mitigation of chromosome loss in clinical CRISPR-Cas9-engineered T cells. *bioRxiv* https://doi.org/10.1101/2023.03.22.533709 [PREPRINT]

Vavassori V, Mercuri E, Marcovecchio GE, Castiello MC, Schiroli G, Albano L, Margulies C, Buquicchio F, Fontana E, Beretta S *et al* (2021) Modeling, optimization, and comparable efficacy of T cell and hematopoietic stem cell gene editing for treating hyper-IgM syndrome. *EMBO Mol Med* 13: e13545

Wen W, Quan Z-J, Li S-A, Yang Z-X, Fu Y-W, Zhang F, Li G-H, Zhao M, Yin M-D, Xu J *et al* (2021) Effective control of large deletions after double-strand breaks by homology-directed repair and dsODN insertion. *Genome Biol* 22: 236

Yazdani R, Fekrvand S, Shahkarami S, Azizi G, Moazzami B, Abolhassani H, Aghamohammadi A (2019) The hyper IgM syndromes: Epidemiology, pathogenesis, clinical manifestations, diagnosis and management. *Clin Immunol* 198: 19 – 30