

Lentivector cryptic splicing mediates increase in CD34+ clones expressing truncated *HMGA2* in human SCID-X1

Suk See De Ravin (✉ sderavin@niaid.nih.gov)

National Institutes of Health <https://orcid.org/0000-0002-9800-774X>

Siyuan Liu

C RTP, Leidos Biomedical Research Inc

Colin Sweeney

National Institute of Allergy and Infectious Diseases, National Institutes of Health

<https://orcid.org/0000-0002-9442-1134>

Julie Brault

NIAID/NIH

Narda Whiting-Theobald

NIAID/NIH

Michelle Ma

National Institute of Allergy and Infectious Diseases

Uimook Choi

National Institutes of Health

Janet Lee

NIAID/NIH

Sandra Anaya-O'Brien

NIAID/NIH

Priscilla Quackenbush

NIAID/NIH

Tyra Estwick

NIAID/NIH

Anita Karra

NIAID/NIH

Nana Kwatema

NIAID/NIH

Shuang Guo

Leidos, Frederick

Ling Su

National Institutes of Health

David sun

Sheng Zhou

St Jude Children's Research

Harry Malech

National Institute of Allergy and Infectious Diseases, NIH <https://orcid.org/0000-0001-5874-5775>

Xiaolin Wu

Frederick National Laboratory for Cancer Research <https://orcid.org/0000-0002-6432-1300>

Taylor Liu

NIAID/NIH

Article

Keywords: vector integration site (VIS), lentivectors, cryptic splicing

Posted Date: October 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-858393/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Communications on June 28th, 2022. See the published version at <https://doi.org/10.1038/s41467-022-31344-x>.

Abstract

A 4–8 year follow-up of vector integration site (VIS) analysis in blood lineages from patients with X-linked severe combined immune deficiency receiving lentivector gene therapy found a > 60-fold increase in frequency of forward-orientated VIS within intron 3 of *HMG2*. Some patients demonstrated emergence of dominant *HMG2* VIS clones in progenitor and myeloid lineages, but with no disturbance of hematopoiesis. Molecular analysis demonstrated a cryptic splice site within the cHS4 insulator generating truncated mRNA transcripts from any transcriptionally active gene containing forward-oriented intronic insert, but with detectable effect on VIS frequency only for intron 3 *HMG2* inserts. A two base-pair change at the splice site eliminated splicing activity while retaining vector functional capability. Functional analysis of lentivectors for cryptic splicing should become routine for pre-clinical safety assessment.

Introduction

Ex vivo lentivector transduction of autologous CD34 + hematopoietic stem cells (HSC) as gene therapy has achieved significant clinical benefit for many inherited diseases (1–5). Some clinical trials using murine gamma retrovirus vectors for autologous HSC gene therapy have resulted in malignancy and myelodysplasia from insertional mutagenesis activation of nearby oncogenes from the action of enhancer elements within the vector long terminal repeat (LTR). However, vector integration mediated oncogenesis has not been observed in any of the many hundreds of patients treated with self-inactivating (SIN) lentivector transduced autologous HSC over the past decade. The observed lack of vector integration mediated oncogenesis by SIN lentivectors is presumed to be due to the introduction of the U3 deletion of enhancer elements in the LTR that confers the SIN characteristic, plus a genome insertional pattern for lentivectors that is less likely to affect nearby gene promoter activation compared to murine gamma retrovirus vectors. None the less, long term genetic analysis of lineage specific vector integration site (VIS) patterns remain an essential element of safety assessment of patients treated with lentivector gene therapy.

Surveillance for potential genotoxicities rely on characterization of the VIS landscape across blood lineages. Observing a diverse polyclonal pattern, lack of emergence of increasing dominance of a blood lineage by a single clone and lack of significant progressive disturbance of hematopoietic lineage number or pattern are generally considered indications of safety. Since 2012 we have been conducting a clinical trial of SIN lentivector transduced autologous HSC for older children and young adults with X-linked severe combined immune deficiency who entered this clinical trial because haploidentical transplant with none or minimal conditioning in infancy had resulted in incomplete immune reconstitution (ClinicalTrials.gov Identifier: NCT01306019). The patients at entry to our study had donor engraftment in only the T lymphocyte lineage, required life-long supplemental IgG, and had increasing medical problems including in most cases chronic norovirus infection. In 2016 we reported clinical outcomes for the first 5 patients demonstrating significant restoration of T cell, B cell and NK cell immunity in all patients. The first two patients (both young adults) at > 2 years follow up had achieved

independence from their need for life-long IgG supplementation, had demonstrated protective responses to immunization and eradication of their long-standing chronic norovirus infection(3). In that report we observed in those first two patients that there was an increasing over-representation of VIS within the *HMGA2* gene relative to the otherwise abundant diverse VIS in other genes and intergenomic regions. We noted no disturbance in hematopoiesis pattern, no expansion of any lineage significantly beyond the normal range and no emergence of leukemic clones. By 2016 a total of 8 patients had been treated with pattern of clinical benefit from the gene therapy for all but Subject P8 reproducing the improvement in immune function together with the clinical benefit reported for the first 5 patients. However, more recently we noted in Subject 6 the appearance of a single clonal population containing a forward-oriented insert in intron 3 of *HMGA2* comprising 15 to 22% of the CD34 + hematopoietic progenitor and myeloid lineages without any evidence for disturbance in hematopoiesis or any cytogenetic abnormalities. The effort to delineate the molecular mechanism for this clonal dominance in Subject P6 and the relative increase in VIS within *HMGA2* in the other patients is the basis of this report.

Transcriptional profiling of peripheral blood immune cells and single cell analysis of CD34 + HSCs and inducible pluripotent stem cells (iPSCs) unraveled the vector-induced alternate splicing events to create aberrant fusion transcripts that led to the clonal dominance in P6 and clonal expansions in other patients. We confirm the potent gene trapping effect of the vector can be mitigated by the removal of its cryptic splice acceptor (SA).

Results

Long term multi-lineage gene marking with abundant and diverse clones in patients treated by lentiviral vector gene therapy for SCID-X1.

Lentivector gene therapy was used to treat older patients with SCID-X1 who continue to experience significant medical problems due to persistent immune deficiency despite prior HSC transplant without myeloid conditioning that achieves only limited engraftment of T cell progenitors. We have treated 15 subjects in this study to date (the first 8 subjects between 2012 and 2016; the second 7 subjects treated since 2019), but will only present detailed analysis of first 8 subjects that allowed an extensive evaluation of the long term kinetics of VIS patterns (Table 1).

Stable multi-lineage gene marking is observed following gene therapy in all of the first 8 patients treated in this study (Fig. 1A). As expected there was substantial increase in VCN in T, B and NK cell lineages over time because of the physiologic growth advantage provided by *IL2RG* transgene production of common gamma chain protein essential to the restoration of function of the receptors for Interleukins 2, 4, 7, 9, 15 and 21. However, there was also a slight to modest increase in VCN in myeloid lineages and in the analysis of peripheral blood circulating CD34+ progenitors. In general, the greatest increase in VCN was observed in NK cells, often exceeding 1 copy/cell (e.g. P1, 4, 5, and 6), suggesting a selection for higher copy number of the *IL2RG* transgene during the reconstitution of NK cells in these patients. Subject P2 had longstanding very severe bronchiectasis pre-gene therapy and died from a fatal pulmonary bleed at

2.5 years after treatment; and Subject P8 did not achieve levels of gene marking sufficient to provide clinical benefit and was successfully retreated with gene therapy at 30 months. For this reason, data in Fig S1 for P2 and P8 are truncated at 24 months and 30 months, respectively.

Extensive VIS analysis was performed over time and across lineages for all 8 patients allowing assessment of over 2 million unique vector integration sites from the 8 patients during the 4-8 year follow-up period tracked (Table 1). This formed the basis for assessing the relative frequency of inserts within individual genes that allowed detection of an increased frequency of VIS in *HMGA2*.

HMGA2 VIS are enriched in all patients.

We compare the cumulative incidence of VIS in *HMGA2* from peripheral blood lineages with the baseline incidence of *HMGA2* VIS in the infusion product *in vitro* (Fig. 2A). For this analysis we distinguish between *HMGA2* VIS that are in the same (forward) orientation versus the opposite (reverse) orientation relative to the *HMGA2* gene. The frequency of forward and reverse orientation *HMGA2* VIS are approximately equal in the baseline transduced CD34+ product and are at a frequency similar to that expected for a gene of this size. However, there is a very significant increase of the *in vivo* frequency of *HMGA2* VIS observed in all patients specifically for VIS in the same orientation as the *HMGA2* gene (Fig 2A), suggesting a cis-element (s) in the vector contributing to the selective clonal expansion. Most of the VIS landed in intron 2, 3, 4 of *HMGA2*, with the largest intron 3 harboring the great majority of these forward-oriented VIS (Fig.2B). Next, our VIS analysis in individual peripheral blood (PB) bead-selected myeloid (CD14 and polymorphonuclear (PMN)) and lymphoid (CD3+T, CD19+B, and NK) cells showed the highest frequency of *HMGA2* VIS clones in myeloid cells (especially in P2, 3, 4, 5, 6), reaching ~20% of all VIS (Fig. 2C). Of note, the frequency declined in all patients (except P4) after a peak by 1-2 years following treatment with no changes in their immune cell numbers, indicating that while *HMGA2* VIS clones have a growth advantage over other clones, overall lineage cell production remained under homeostatic regulation.

The growth advantage of *HMGA2* VIS clones is associated with generation of a truncated *HMGA2* Exon 1-3 / lentivector fusion transcript in CD34 HPSCs.

There is a previous report of lentivector insertion activation of *HMGA2* causing a large expansion of a single clone restricted to the erythroid lineage in human lentiviral HSPC β -thalassemia gene therapy (10). The activation was caused by alternative splicing of *HMGA2* transcript to the integrated lentivector LTR which increase the stability of the transcription. To evaluate for potential alternative splicing as a mechanism for the *HMGA2* clonal expansion in the myeloid cells in our patients, we performed reverse transcription of RNA from patient PMN, CD14 cells which failed to detect any endogenous *HMGA2* transcripts in those samples. *HMGA2* is highly expressed in early embryonic development is lost with differentiation in adult tissues except in some stem cells including CD34+ HSC. We reasoned that the high frequency of *HMGA2* VIS clones in the short-lived myeloid (PMN, CD14+) cells may simply reflect an increased pool of *HMGA2* VIS clones in their upstream progenitors from which they are derived. To evaluate the upstream progenitors, we isolated CD34+ cells normally found at low (<0.1%) levels

circulating in peripheral blood from P6 who had a dominant HMGA2 VIS clone at 18 month following treatment, and by RT-PCR, identified a fusion HMGA2-LV transcript (Fig.3A). Sequencing of the fusion transcript confirmed the alternate splicing mechanism generated by the splicing donor (SD) of HMGA2 exon 3 and a cryptic splicing acceptor (SA) in the cHS4-400 element in the vector LTR (Fig. S1).

We then quantified the expression of the fusion transcript (contains only the first 3 exons) relative to the full length HMGA2 transcript (includes exons 4, 5) using the ratio of exon 1-2/exon 4-5. The ratio in HD CD34⁺ cells was ~1, while in P6 CD34⁺ cells the ratio increased to 5.91 (Fig. 3B), consistent with an accumulation of truncated fusion transcripts of *HMGA2* that lack exon 4-5 and the 3' UTR as a consequence of the LV integration. HMGA2 transcript can also be detected in NK cells from P6 but with exon 1-2/exon 4-5 ratio of 7.55 (Fig. 3B). Since both the full and the truncated transcript retains the same endogenous promoter, it is interesting that HMGA2 expression is detected in CD34⁺ as well as NK cells, suggesting a physiological role for HMGA2 in these cell types and not in differentiated PMN, CD14, T and B cells. This expression pattern is also confirmed by RNAseq. The fusion transcript of HMGA2-cHS4 has been detected by RNAseq in P6 CD34 cells and can be detected in CD34 cells from other patients. These data confirm that the truncated HMGA2 fusion transcripts initiated in CD34⁺ cells led to their clonal expansion that is reflected in the short-lived PMN and CD14 cells that serve as useful surrogate biomarkers for the genetic composition in the rarer CD34⁺ progenitors from they are derived.

Expanded HMGA2 clones are commonly associated with multicopy transgenes.

In P6, the largest HMGA2 clone remained stable from 24m to 48m in CD34 cells, which constituted about 4% of VIS in PMN, CD14 or CD34 cells. Strikingly, there are 10 other VIS tracked at the high frequency throughout this time course, also observed in multiple cell lineages (CD19, NK, PMN, and CD14 cells) (Fig. 1B, Fig. 3C). This synchronized kinetics of multiple VIS is also observed in other patients (Fig. 1B), raising the possibility that HMGA2 insertion and multiple other insertions reside in the same cell clones. To evaluate this possibility, we established single cell CD34⁺ colonies from P6 (at 36 months) for VISA or ddPCR analyses and found 20/88 (%) harbored the same (10 inserts) or closely similar (11 inserts) set of VIS with one additional integration in TALDO1 gene, possibly acquired during in vitro transduction (Fig. 3D). The presence of multiple copies of transgene per cell or clone leads to an automatic increase in the denominator with each increase in the numerator insert, causing an under-estimation of the frequency of the clones and invalidates the use of VIS frequency alone to determine the actual frequency of the clones. To address this, we used ddPCR to quantify all integration site junctions in this set relative to the human genome reference gene and confirmed from multiple assays that the true cell clone frequency to be 23% of CD34 and CD14 cells from P6, much higher than ~4% estimated by VIS (Fig. 3C). We developed a simple statistical method to predict potential multicopy cell clones based on VIS frequency correlation coefficient. Based on this approach, we identified numerous other multi-insert clones in many of the patients (Suppl. Excel Table S1). that generally also harbor an insert in HMGA2, suggesting that while insertion in the HMGA2 may drive the clonal expansion, the association with multiple copies of transgene also raise the possibility that activated HMGA2 works in synergy with high dose of IL2 receptor common gamma chain (IL2R γ c) for the observed growth advantage.

Further enrichment of P6 HMGA2 10+ copy clone during induced pluripotent stem (iPS) cell generation.

Single cell CD34+ colonies provide limited DNA or RNA too challenging for detailed molecular characterization. To further characterize the P6 HMGA2 10+ copy cell clone and its impact on host gene expression, we established (induce pluripotent stem) iPS single cell derived clones from peripheral blood CD34+ HSC from P6. Karyotyping showed that these iPS cell clones have normal karyotypes (Fig. 3E).

Over half (56%) of the iPSCs (15 of 27 total) harbored the HMGA2 10+ clone which is significantly higher compared to CD34+ cells (23%) ($p=0.0002$, Fisher's Exact test) (Fig. 3F). This may be attributable to higher expression of HMGA2 in the more primitive iPSCs.

cHS4-400 cause widespread alternative splicing of host genes that it integrates into in the sense (forward) orientation.

To determine if the HMGA2-Exon3-cHS4 splicing is sequence context dependent specifically for the HMGA2 gene, or a more general vector-specific phenomenon, we performed RNA sequencing of nine iPSC clones harboring different sets of integrations, each with multiple copies of VIS in different genes (Suppl Excel Table 2). The RNAseq analysis supports the latter addition of TALDO1 insertion to the 10+clone, instead of a later loss of TALDO insert from the 11+ clone, ascertained by the TALDO1 expression in #21 #26 (no TALDO1) being twice that of #10#11(with TALDO1 insertion), consistent with no loss of an allele in #21#26, but that insertion in TALDO1 exon 3 in clone #10#11knocks out the expression of one allele. Details of additional iPSC clones are provided in Supplementary, but of the 3 different VIS in HMGA2, and 26 in other genes, 14/29 were in introns and in the same orientation of the gene. Chimeric transcripts of target genes fused to the cHS4-400 SA in 93% (13/14) of the same orientation integration events, with the exception of only one gene CTSS, which is not expressed or at extremely low level in iPS cells (Fig. 4, Suppl. Table S2). In a few cases such as MARK3 gene, UBE2G1 gene, and GOSR1 gene, multiple fusion transcripts were detected in addition to the expected upstream exons. One VIS is outside the nearest known gene ZFN586 and in the opposite orientation as the gene nevertheless resulted in fusion transcripts between cHS4-400 SA of this insertion and another nearby gene ZFN552, which is upstream in the same orientation. The splicing occurred with ZFN552 exon2 SD, skipping ZFN552 exon3, and fused to cHS4-400 SA that is more than 32 kb 3' to the exon2 SD. Figure 4 summarizes all insertions in each target gene in these clones and illustrates fusion transcripts discovered by RNAseq. Our data show that cHS4 in the vector can cause widespread fusion transcripts with host genes when inserted in the sense (forward) orientation.

cHS4 is a strong gene trap that causes complete transcription termination of genes that it integrates into in the sense orientation

We next asked if the cHS4 is a strong gene trap, the extent of its impact on the host gene transcription. Low level alternative splicing of host gene is known to happen in integrated lentivectors (11, 12). We quantified the impact of transcription termination of cHS4 on host genes By comparing the sequence read numbers of the upstream exon-cHS4 gene-trap event and the normal exon-exon splicing event (Fig. 5A), where insertion causing loss of one allele should result in a 50:50 ratio in RNAseq. Our data

demonstrated that for many genes, the fusion transcripts of exon-cHS4 are detected at similar or even higher level as normally spliced transcript. For the HMGA2 gene, the data here seem to contradict what was observed in patient CD34+ HSC or NK cells. The ratio of the fusion transcript vs the normal transcript is roughly equal here in iPS cells, whereas the fusion transcripts was much more abundant than the normal transcript in CD34+ HSC or NK cells. This analysis relied on the homogeneity of single cell iPS clones. Interestingly, the increased stability of truncated HMGA2 fusion transcript due to loss of let7b regulation provides a growth advantage to HMGA2 VIS clones in CD34 or NK cells where let7b is normally expressed. However, during the iPS reprogramming, lin28 is turned on (as confirmed by RNAseq) which in turn blocks let7b expression. It appears that the HMGA2 fusion transcript is equally stable as normal transcript in iPS cells. The HMGA2 normal transcript became less stable again than the fusion transcript when we induced the iPS cell to differentiate into macrophages, during which Lin28 expression is turned off. These data confirm that normal full length HMGA2 transcript is subjected to miRNA regulated degradation when let7b is expressed. Increased transcripts were also observed from other genes (DIP2A, STAG3, GOSR). In total, these data show that the cryptic splice acceptor in cHS4-400 is a very strong gene trap that can cause orientation specific transcription termination of any genes it integrates into via its splicing acceptor and polyadenylation signal in the LTR. It can even cause dysregulation of target genes even if it is outside the gene. This could lead to either gain of function if the truncated protein has new functions or loss of function of the target genes that is needed for the cellular growth.

Clonal expansion of cells with vector insertion in other genes.

Next, we evaluated for potential gene trapping phenomenon in other genes to account for their expansion and their potential implications. A notable clone was one identified in P3 at 20 months following transplant that accounted for ~ 50% of his NK cells only harbors a VIS in the sense orientation in *NF1*, a tumor suppressor gene that negatively regulates Ras activity. Heterozygous *NF1* mutations lead to neurofibromatosis type 1 (13, 14).. Anxieties regarding the clone contributing to ~50% of his NK cells were balanced by the lower-than-normal numbers of peripheral blood NK cells, and his stable clinical status. This *NF1*+ clone is stably maintained at high frequency till the last assayed timepoint at 54 months, with his absolute NK cell count still below normal range. RNAseq analysis on three NK cell RNA samples from P3 revealed a fusion transcript of *NF1* upstream exon to cHS4 in two out of three samples that was subsequently also identified in other cell types though to a lesser degree. Also of note is a VIS in *ETS1* oncogene in CD3+ cells in P6 (10% of VIS) that has also remained stable from 36-48months. These examples highlight the expanse of genes that may be impacted by the cHS4 gene trap activity and lead to clonal expansion. However, it is not easy to establish the causal relationship in each case.

Site-directed mutation of the SA in cHS4 eliminates gene trap activity.

Next, we removed the cryptic SA in the cHS4 region of the vector by site-directed mutagenesis, located outside the CTCF binding site and should have minimal impact on the insulator function. The conserved AG dinucleotide sequence for SA (Fig. 5B, C)) was changed to TG dinucleotide in the modified vector. We

transduced SCID-X1 patient CD34+ HSPCs with either the clinical CI20-i4-hyc OPT vector or the modified version. We performed *in vitro* IL2RG gc-dependent differentiation of transduced HSCs to T cells using an artificial thymic organoid system (15) (Fig. 6A). XSCID patient CD34+ HSCs transduced with modified LV maintained T-cell differentiation properties (Fig. 6B), gc expression in *in vitro*-differentiated T (Fig. 6C) and NK cells (Fig. 6D) as well as restored IL2RGgc-STAT5 signaling (Fig. 6E) comparable to the original clinical LV, confirming that the minor change to the cryptic SA has not disrupted vector function or transducing capability (Fig. 6F). *In vitro* assessment confirmed γ c expression in transduced patient cells that indicates the modification did not disrupt transgene function. To investigate whether the gene trapping activity is removed from the new vector, we performed RNAseq on CD34+ colonies as either single colonies (to increased sensitivity) or pooled colonies (to increase the chance of detecting various trapped genes). No fusion transcripts indicative of gene trapping activity were identified in cells transduced with the modified vector, while many fusion transcripts were identified in cells transduced with the clinical vector, indicating successful functional removal of the described SA.

Clinical benefits and stability following treatment

Of the eight patients treated in this cohort, 7 are alive and well with gradual recovery of immune cell numbers over years (Suppl). P2 had a fatal pulmonary hemorrhage due to pre-existing chronic lung disease at ~ two years following treatment. The most significant clinical benefit is observed in clearance of chronic noro-or other gastrointestinal pathogen, improvement of gastrointestinal function and in the improvement in B cell compartment.

Discussion

Although no cancers have been reported with over a hundred patients treated with LV gene therapy over the past 12–15 years, insertional mutagenesis remains a safety concern for gene therapy using integrating vectors. Long term surveillance remains critical as events may present late, as demonstrated in the most recent leukemia in γ RV-treated SCID-X1 patient 16 years after treatment (16). Clonal expansion of HMGA2 VIS clones was noted in our first two patients of SCID-X1 gene therapy following a relative short time period. Long term surveillance of these clones reveal the kinetics of these clones that were also identified in the entire cohort of treated 8 patients, indicating a system-wide problem that is related to the cryptic SA in the cHS4 region of the vector. Single cell derived RNA sequencing data characterized fusion transcripts of multiple genes, including tumor suppressor genes.

Integrating vectors may trigger oncogenesis by a) interactions between viral enhancer sequences and cellular promoters to cause transcriptional activation and dysregulation of oncogenes (7, 9, 17), b) insertion of vector splicing and polyadenylation signals that may lead to dysregulation of gene expression through aberrant splicing, premature transcript termination, and the generation of chimeric transcripts. The preferential integration of LV within transcription bodies likely increases risks of aberrant transcripts. Although aberrant transcripts have been demonstrated *in vitro* (11, 18–20) and in tumor model (21), the long-term biological impact in human following lentivector gene therapy remains unclear.

Analysis of the gene-trapping events reveal interesting biology *in vivo* not previously appreciated. The High Mobility Group AT-hook 2 (*HMGA2*) gene encodes a small protein with three DNA-binding AT-hook domains and modulates chromatin structure, transcription, and epigenetic regulation of multiple genes (22, 23). It plays crucial roles in the proliferation, cell-cycle progression, apoptosis, and senescence of cells, and given its high expression level selectively in primitive embryonic tissues, it is not surprising that *HMGA2* has been recently shown to be involved in self renewal of hematopoietic stem cells (24, 25) and neural stem cells (26, 27). *HMGA2* clonal expansion in our patients was first noticed in peripheral blood, especially myeloid cells. However, *HMGA2* transcripts were not identified from circulating myeloid cells, but only from CD34 cells, highlighting the importance of surveilling expressing cells. While PMNs with short half-life are a good surrogate for stem and progenitor cells with regards to genomic VIS, the transcriptional profile is highly dependent on the specific cell stage. Our data demonstrated that LV-insertional activation of *HMGA2* can provide a growth advantage for CD34 + cells and supports its role in hematopoietic stem cell renewal and promoting survival and proliferation after stem cell transplantation.

The activation of *HMGA2* in our patients caused by the cHS4 insulator element in the lentivector was also observed in β -thalassemia gene therapy that used a vector also including the cHS4 element, indicating that the gene trapping activity is vector specific, especially since no *HMGA2* clonal expansion has been reported in lentivector gene therapies for other primary immunodeficiency diseases such as WAS (4), MLD (5) or CGD (2) using vectors without the cHS4 insulator. Interestingly, although the expanded *HMGA2* VIS clone was restricted to the myeloid lineage in the β -thalassemia study, we identified *HMGA2* VIS clones in both myeloid and lymphoid lineages in our study, raising the possibility that there may be additional factors affecting the clonal expansion of *HMGA2* clones. A key difference between our studies is the transgene itself, where the multiple *IL2RG* transgenes in the present study may super-impose an additional growth advantage. Interestingly, Wang *et al.* reported enrichment of *HMGA2* clones in two SCID-X1 patients treated with γ RV gene therapy (28) even though the vector does not involve cHS4, supporting the concept that *IL2RG* transgene may play a synergistic role in *HMGA2* clonal expansion.

The chromatin insulator was introduced to act as a barrier to prevent silencing of the transgene or interference with the host cell or transgene expression as an enhancer blocker. Two copies of insulators flanking the transgene are necessary to achieve protection from gene silencing and enhancer blocking. Since the large (1200bp full length cHS4) size and repetitive nature can generate problems with instability in the lentivector LTR and reduce the titer of the virus, a shorter 400 bp fragment with full protection function was used instead. Our data suggest that cHS4 should not be used in its current form due to its strong transcription termination activity. There are many other cryptic splicing sites in lentivectors, although most of them are known to cause only low-level alternative splicing of host genes (17). We did not identify any other aberrant transcripts in this study, although it is possible that less active cryptic splice sites may manifest in the absence of cHS4 SA. We demonstrated that introduction of a simple point mutation in the cHS4 SA eliminated its gene trap activity and recommend such modification for clinical LV to reduce genotoxicity risks.

Interestingly, although the expanded *HMGA2* VIS clone was restricted to the myeloid lineage in the β -thalassemia study, we identified *HMGA2* VIS clones in both myeloid and lymphoid lineages in our study, raising the possibility that there may be additional factors affecting the clonal expansion of *HMGA2* clones. A key difference between our studies is the transgene itself, where the multiple *IL2RG* transgenes may super-impose an additional growth advantage. Interestingly, Wang *et al.* reported enrichment of *HMGA2* clones in two SCID-X1 patients treated with γ RV gene therapy [23] even though the vector does not involve *cHS4*, supporting the concept that *IL2RG* transgene may play a synergistic role in *HMGA2* clonal expansion.

While the proliferative advantage from increased *HMGA2* expression in CD34 + HSCs may appear advantageous for a CD34 + HSC product, its potential causative involvement in multiple types of cancers (22, 29) pose a potential safety concern that requires long-term follow up. The concerns are somewhat offset by our observations that to date, 1) it has not caused un-controlled expansion, and 2) there has been spontaneous attrition of the *HMGA2* VIS clones after a peak at 1–2 years in most patients. We attribute the limitation of *HMGA2* effects from increased stability to regulation by its endogenous promoter restricting its expression to specific phases and lineages of cell development. Increased levels of the truncated *HMGA2* protein may provide an advantage to CD34 + HSCs and animal models to evaluate whether the potential benefits and safety issue with *HMGA2* dysregulation for hematopoiesis would be of value. In a study with a transgenic mouse model that express 3'UTR truncated *HMGA2*, a growth advantage was seen for *HMGA2* expressing bone marrow cells after serial transplantation but no malignancy was observed (30). In a recent rhesus macaque model (Bonner *et al.*) transplanted with CD34 + HSCs transduced with a lentivector to constitutively over-express 3' UTR truncated *HMGA2* under control of the MSCV promoter in transplanted CD34⁺ hematopoietic stem cells, these cells demonstrated increased self-renewal potential that was self-limiting and did not result in any hematological malignancies even after 3 generations of transplant.

An alternative mechanism likely accounts for the expansion of the *NF1* VIS clone in P3 NK cells if it is not a stochastic event. The strong *cHS4* gene trap activity can have two different consequences, the commonest being a loss of function for one allele of an autosomal gene, that can be compensated by the other functional allele. However, loss of one allele may cause haplo-insufficiency that may be problematic for tumor suppressor genes such as *NF1*, a potential cause of dysregulated hematopoiesis (31). Alternatively, the *cHS4* gene trap may result in truncated transcripts/proteins with a gain of function as demonstrated by the increased stability of *HMGA2* fusion transcript. Most of the truncated *HMGA2* proteins retain the first three exons, which encode all three AT-hook DNA binding domains of the protein. The truncated protein is (or is partially) functional in providing a growth advantage to cells in which it is expressed. Our data showing enrichment for multi-insert clones suggest that the multiple copies of *IL2RG* may augment *HMGA2* activation, acting in synergy to provide a growth advantage.

In summary, we have reported potent gene trapping due to a cryptic splice acceptor in the lentivector *cHS4* region of the vector causing substantial increase in *HMGA2* clones in CD34 + HSCs. Identification of the aberrant fusion transcripts is dependent on the use of appropriate cells in which the genes are

expressed. A simple mutagenesis of the strong SA removed the gene trapping activity to improve safety of the vector.

Methods

Human subjects. This is a phase 1/2 nonrandomized clinical trial of *ex vivo* HSPC gene transfer treatment for SCID-X1 using a SIN insulated LV. The study is approved by the National Institute of Allergy and Infectious Diseases (NIAID) Institutional Regulatory Board (clinical protocol #11-I-0007, ClinicalTrials.gov ID NCT01306019) and Institutional Bio-safety Committee, FDA (investigational new drug #15041), sponsored by the NIAID Regulatory Compliance and Human Subjects Protection Branch.

Lentiviral vector. The CI20-i4-EF1 α -hycOPT used in this clinical trial is a vesicular stomatitis virus G (VSV-G)-pseudotyped, third-generation SIN vector that uses a promoter fragment from the eukaryotic EF1 α gene to express a codon-optimized human γ c cDNA and contains a 400-bp cHS4 insulator fragment from the chicken β -globin locus within the SIN LTR (1, 3).

Autologous CD34⁺ HSPCs collection and isolation. Patients received G-CSF at 16 mg/kg per day by subcutaneous (sc) injection for five consecutive days, supplemented by plerixafor (0.24 mg/kg, sc, 11 hours before collection) [National Institutes of Health (NIH) protocol 94-I-0073] before apheresis at the NIH Clinical Center. The products were processed using FDA-approved Isolex immune anti-CD34 magnetic bead system to isolate and enrich CD34⁺ cells by the NIH Department of Transfusion Medicine Cell Processing Facility. Purified CD34⁺ HSPCs were cryopreserved until gene therapy.

Transduction of CD34⁺ HSPCs. *Ex vivo* culture and transduction of the patient's autologous CD34⁺ HSPCs with VSV-G-pseudotyped CI20-i4-EF1 α -hycOPT LV were performed and certified by the NIH Department of Transfusion Medicine Cell Processing Facility. Transduction involves thawing and suspension of patient CD34⁺ HSPCs in X-VIVO 10 serum-free growth medium [containing 1% human serum albumin plus cytokines (stem cell factor, 50 to 100 ng/ml; Flt-3 ligand, 50 to 100 ng/ml; thrombopoietin, 50 to 100 ng/ml; IL-3, 5 ng/ml)]. The cells were cultured in T-175 tissue culture flasks coated with the recombinant fibronectin fragment known as RetroNectin and exposed to LV for 6 to 8 hours each day for two consecutive days after an overnight pre-stimulation. Transduced CD34⁺ HSPCs were harvested and infused fresh after required safety testing and quality control testing.

Conditioning regimen. A 6 mg/kg busulfan dose given over 2 days was used prior to infusion of transduced cells. No T cell- or B cell-depleting agents were given.

Clinical course. No adverse events were noted with the infusion of the transduced CD34⁺ HSPC product. The nadir for the expected busulfan-related effects of neutropenia and thrombocytopenia occurred at 2 to 3 weeks after treatment and were recovered without requiring any cellular support or intervention. Subjects were discharged home by 3 to 4 weeks after gene therapy. Three of the eight patients developed febrile neutropenia that responded to empiric antimicrobial therapy.

Cell lineage separation for gene marking and integration analysis. After PMN separation using histopaque, the mononuclear leukocyte layer was fractionated by magnetic beads per manufacturer's instructions (Dyna Beads, Invitrogen).

Peripheral blood CD34⁺ cell purification and expansion for vector copy number (VCN) and integration site analyses. PBMCs were purified from non-mobilized peripheral blood by Ficoll separation (Lymphocyte Separation Medium, MP Biomedicals). CD34⁺ cells were isolated from PBMCs using magnetic cell sorter magnetic beads (Miltenyi Biotec) according to the manufacturer's protocol. CD34⁺ cells were expanded for 8 to 12 days in StemSpan II media (STEMCELL Technologies) supplemented with 100 ng/ml each of human stem cell factor, FLT-3L, and thrombopoietin (PeproTech). DNA was isolated using the DNeasy Blood and Tissue Kit (QIAGEN).

Quantitative determination of VCN by ddPCR. To measure the vector-carrying cells, we used a ddPCR (Bio-Rad) assay. The ddPCR assay allows the measurement of absolute copy number without using standard curve. The vector-specific primers and probes are as follows: HIV forward, 5'-CTGTTGTGTGACTCTGGTAACT-3'; HIV reverse, 5'-TTCGCTTTCAAGTCCCTGTT-3'; and HIV probe, 5'-/56-FAM/AAATCTCTA/ZEN/GCAGTGGCGCCCG/3IABkFQ/-3'. We multiplexed a reference gene assay for cell counts in the same reaction [myocardin-like protein 2 (MKL2): forward, 5'-AGATCAGAAGGGTGAGAAGAATG-3'; reverse, 5'-GGATGGTCTGGTAGTTGTAGTG-3'; and probe, 5'-/56-HEX/TGTTCTGC/ZEN/AACTGCAGATCCTGA/3IABkFQ/-3']. Cell number was calculated as half of the *MKL2* counts because each cell is diploid. VCN was calculated as vectors per cell.

Quantification of clonal expansion of specific integration site. Top expanded clones identified by the VISA sequencing assay were followed up and monitored by specific ddPCR assay. All of our ddPCR assays consist of a common LTR primer, an LTR probe, and a specific primer for genomic DNA junction (5LTR reverse, 5'-CTGCAGGGATCTTGTCTTCTT-3'; 5LTR junction probe, 5'-/56-FAM/TGGAAGGGC/ZEN/TAATCACTCCCA/3IABkFQ/-3'; 3LTR forward, 5'-CCCACTGCTTAAGCCTCAATA-3'; 3LTR junction probe, 5'-/56-FAM/AAGTAGTGT/ZEN/GTGCCCGTCTGTTGT/3IABkFQ/-3'). We multiplexed the integration site-specific assay together with the *MKL2* reference gene assay, which measures cell counts in the same reaction.

Vector Integration Site Analysis (VISA). Patient DNA was subjected to VISA as described previously (3).

CD34⁺ single cell clone VISA assay. To verify multi-copy transgene integrated in the same cell, we isolated CD34⁺ cells from patient and performed colony forming cell (CFC) assay in methylcellulose media. Colonies formed from single cells were plucked and transferred into 50 µL of water and then Proteinase K was added for digestion. Lysate (9 µL) containing cellular DNA was used for whole genome amplification using GenomePlex Single Cell Whole Genome Amplification kit WGA4 (MilliporeSigma). The amplified DNA was subjected to VISA and ddPCR analysis. Integration sites identified from single colonies are considered from single cell.

iPSC single cell clones. We established iPSC single cell-derived clones from patient CD34⁺ cells by Sendai virus reprogramming as previously described (32). The established single cell clones were subjected to VISA and ddPCR analysis. G-band karyotype analysis was performed commercially (WiCell).

Detection of HMGA2 fusion transcripts. Total RNA was extracted from patient or healthy donor blood cells. Reverse transcription was carried out to generate cDNA using Invitrogen kit. PCR was performed using nested primers, with forward primers matching *HMGA2* exon 2 and exon 3, and reverse primers matching the R region of lentivector LTR (HMGA2-Ex2-F CGGTGAGCCCTCTCCTAA; Vector-HIV-R AGCAGTGGGTTCCCTAGTTA; HMGA2-E3-M13Fnest GTAAAACGACGGCCAGTGCAGAAGCCACTGGAGAAA; Vector-HIV-R-M13Rnest GGAAACAGCTATGACCATGCCAGGCTCAGATCTGGTCTAAC). Two amplicons were observed with slightly different sizes on the gel. The amplicons of the fusion transcripts were sequenced by Illumina MiSeq. We were able to deduce two different alternatively spliced variants of *HMGA2* exon 3 into the *cHS4* insulator core element of the vector.

Quantification of HMGA2 transcripts. Full length HMGA2 and truncated HMGA2 transcripts were quantified by ddPCR after converting RNA to cDNA by reverse transcription. Two ddPCR assays were designed for HMGA2, one 5' to the integration site cluster at exon 2/exon 3 junction, the other 3' to the integration site cluster at exon 4/exon 5 junction. These two assays were used to examine the gene expression of *HMGA2* and the 5'/3' ratio for the 3'-truncated form that might be caused by vector integration. CD34⁺ cells from healthy donors were used as control. Primer and probe sequences for 5' assay are forward CCC TCT CCT AAG AGA CCC A, Probe /56-FAM/TT TGC TGC C/ZEN/T TTG GGT CTT CCC /3IABkFQ/, Reverse CTG CCT CTT GGC CGT TT. Primer and probe sequences for 3' assay are: forward GTT CAG AAG AAG CCT GCT CA, Probe /56-FAM/AG AGT CTG C/ZEN/C GAA GAG GAC TAG GG/3IABkFQ/, Reverse TGC TGA GGT AGA AAT CGA ACG.

RNAseq. Total RNA was isolated from iPS cell clones. We performed both mRNAseq and total RNAseq using NEB Next Ultra RNA library Prep Kit for Illumina (E7530S) following to manufacturer's recommended protocols (New England Biolabs, Ipswich, MA). The libraries were sequenced with 2x75 bp paired end reads on Illumina NextSeq2000 with P2 reagents (Illumina, San Diego, CA). Data analysis was performed with NextSeq2000 onboard Dragon RNAseq pipeline. For fusion transcript detection with *cHS4*, fastq files were used to search all reads that contain the *cHS4* SA site. The junctions were then split and sorted and mapped to human genome hg38 to find the fusion partners.

T cell differentiation: Artificial Thymic Organoid (ATO) system CD34⁺ HSPCs were differentiated into T cells *in vitro* using a 3D artificial thymic organoid system (15, 33). Briefly, CD34⁺ cells were co-cultured with MS5 murine stromal cell line modified to express DLL4 ligand in a volume of 5 μ L on 0.4 μ m transwell inserts (EMD Millipore). Culture medium containing 5 ng/ml rhFlt3-ligand and 5 ng/ml rhIL-7 (PeproTech) was changed every 3–4 days. After 6 weeks, ATOs were disaggregated and filtered through a 70 μ m cell strainer before staining for flow cytometry analysis.

NK differentiation

HSPCs were differentiated into NK cells *in vitro* using a protocol adapted CD34⁺ cells were resuspended in Iscove's Modified Dulbecco's Medium supplemented with 20% Fetal Bovine Serum (FBS) and 50 μ M β -mercaptoethanol containing 5 ng/mL rhIL-3 (first week only), 10 ng/mL rhIL-15, 20 ng/mL rhIL-7, 20 ng/mL rhSCF, 10 ng/mL rhFlt3-ligand and 100 U/mL rhIL-2 for 35 days. 40Gy-irradiated K562-mb15-41BBL cells (kindly provided by St Jude Children's Research Hospital, Memphis, TN) at days 7 and 21 of differentiation. Cells were harvested at day 35 for molecular and phenotypic analysis, and cytotoxicity (34).

Statistics.

Chi square test or Fisher's exact test were used as described in the text. For prediction of multicopy integration sites in the same cell clone, we calculated the frequency of all integration sites at each timepoint and in each cell lineages. We then calculated correlation coefficient of VIS frequencies. The correlation coefficient and total counts of each VIS is considered together to make calls for multicopy clones.

Data availability

Patient integration site data are available upon request. RNAseq data for iPSC clones are available at NCBI Sequence Read Archive (PRJNAxxxxx: SRRnnnnnnn - which can all be found on the National Center for Biotechnology Information database)

Declarations

Study approval

The study is approved by the National Institute of Allergy and Infectious Diseases (NIAID) Institutional Regulatory Board (clinical protocol #11-I-0007, ClinicalTrials.gov ID [NCT01306019](#)) and Institutional Biosafety Committee, FDA (investigational new drug #15041), sponsored by the NIAID Regulatory Compliance and Human Subjects Protection Branch. Written informed consent was received from participants prior to inclusion in the study.

Author contributions

XW, SDDR, HLM designed the experiments, performed data analysis, and wrote the manuscript. SL, LS, DS, SG carried out integration site analysis, RT-PCR, RNAseq. CS, MM performed iPSC clones derived from CD34 cells. UC performed site-directed mutagenesis. NTW, TL performed patient sample collection and lineage sorting.

Acknowledgements

This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN261200800001E. The content of this publication

does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

References

1. E. Mamcarz *et al.*, Lentiviral Gene Therapy Combined with Low-Dose Busulfan in Infants with SCID-X1. *N Engl J Med* **380**, 1525-1534 (2019).
2. D. B. Kohn *et al.*, Lentiviral gene therapy for X-linked chronic granulomatous disease. *Nat Med* **26**, 200-206 (2020).
3. S. S. De Ravin *et al.*, Lentiviral hematopoietic stem cell gene therapy for X-linked severe combined immunodeficiency. *Sci Transl Med* **8**, 335ra357 (2016).
4. A. Aiuti *et al.*, Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott-Aldrich syndrome. *Science* **341**, 1233151 (2013).
5. A. Biffi *et al.*, Lentiviral hematopoietic stem cell gene therapy benefits metachromatic leukodystrophy. *Science* **341**, 1233158 (2013).
6. C. J. Braun *et al.*, Gene therapy for Wiskott-Aldrich Syndrome-Long-term reconstitution and clinical benefits, but increased risk for leukemogenesis. *Rare Dis* **2**, e947749 (2014).
7. S. J. Howe *et al.*, Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *J Clin Invest* **118**, 3143-3150 (2008).
8. S. Hacein-Bey-Abina *et al.*, A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N Engl J Med* **348**, 255-256 (2003).
9. S. Hacein-Bey-Abina *et al.*, LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* **302**, 415-419 (2003).
10. M. Cavazzana-Calvo *et al.*, Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia. *Nature* **467**, 318-322 (2010).
11. A. Moiani *et al.*, Lentiviral vector integration in the human genome induces alternative splicing and generates aberrant transcripts. *J Clin Invest* **122**, 1653-1666 (2012).
12. R. K. Merling *et al.*, Transgene-free iPSCs generated from small volume peripheral blood nonmobilized CD34+ cells. *Blood* **121**, e98-107 (2013).
13. D. Viskochil *et al.*, Deletions and a translocation interrupt a cloned gene at the neurofibromatosis type 1 locus. *Cell* **62**, 187-192 (1990).

14. M. R. Wallace *et al.*, Type 1 neurofibromatosis gene: identification of a large transcript disrupted in three NF1 patients. *Science* **249**, 181-186 (1990).
15. M. Bosticardo *et al.*, Artificial thymic organoids represent a reliable tool to study T-cell differentiation in patients with severe T-cell lymphopenia. *Blood Adv* **4**, 2611-2616 (2020).
16. M. Cavazzana, F. D. Bushman, A. Miccio, I. Andre-Schmutz, E. Six, Gene therapy targeting haematopoietic stem cells for inherited diseases: progress and challenges. *Nat Rev Drug Discov* **18**, 447-462 (2019).
17. S. Hacein-Bey-Abina *et al.*, Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J Clin Invest* **118**, 3132-3142 (2008).
18. M. Bokhoven *et al.*, Insertional gene activation by lentiviral and gammaretroviral vectors. *J Virol* **83**, 283-294 (2009).
19. S. Knight, M. Bokhoven, M. Collins, Y. Takeuchi, Effect of the internal promoter on insertional gene activation by lentiviral vectors with an intact HIV long terminal repeat. *J Virol* **84**, 4856-4859 (2010).
20. D. Cesana *et al.*, Whole transcriptome characterization of aberrant splicing events induced by lentiviral vector integrations. *J Clin Invest* **122**, 1667-1676 (2012).
21. E. Montini *et al.*, The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy. *J Clin Invest* **119**, 964-975 (2009).
22. A. Fusco, M. Fedele, Roles of HMGA proteins in cancer. *Nat Rev Cancer* **7**, 899-910 (2007).
23. R. Sgarra *et al.*, Nuclear phosphoproteins HMGA and their relationship with chromatin structure and cancer. *FEBS Lett* **574**, 1-8 (2004).
24. M. R. Copley *et al.*, The Lin28b-let-7-Hmga2 axis determines the higher self-renewal potential of fetal haematopoietic stem cells. *Nat Cell Biol* **15**, 916-925 (2013).
25. P. Kumar *et al.*, HMGA2 promotes long-term engraftment and myeloerythroid differentiation of human hematopoietic stem and progenitor cells. *Blood Adv* **3**, 681-691 (2019).
26. J. Nishino, I. Kim, K. Chada, S. J. Morrison, Hmga2 promotes neural stem cell self-renewal in young but not old mice by reducing p16Ink4a and p19Arf Expression. *Cell* **135**, 227-239 (2008).
27. J. Nishino, S. Kim, Y. Zhu, H. Zhu, S. J. Morrison, A network of heterochronic genes including Imp1 regulates temporal changes in stem cell properties. *Elife* **2**, e00924 (2013).
28. G. P. Wang, A. Ciuffi, J. Leipzig, C. C. Berry, F. D. Bushman, HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res* **17**,

1186-1194 (2007).

29. A. R. Young, M. Narita, Oncogenic HMGA2: short or small? *Genes Dev* **21**, 1005-1009 (2007).
30. K. Ikeda, P. J. Mason, M. Bessler, 3'UTR-truncated Hmga2 cDNA causes MPN-like hematopoiesis by conferring a clonal growth advantage at the level of HSC in mice. *Blood* **117**, 5860-5869 (2011).
31. F. C. Yang *et al.*, Nf1-dependent tumors require a microenvironment containing Nf1+/- and c-kit-dependent bone marrow. *Cell* **135**, 437-448 (2008).
32. R. K. Merling *et al.*, Transgene-free iPSCs generated from small volume peripheral blood nonmobilized CD34+ cells. *Blood* **121**, e98-107 (2013).
33. C. S. Seet *et al.*, Generation of mature T cells from human hematopoietic stem and progenitor cells in artificial thymic organoids. *Nat Methods* **14**, 521-530 (2017).
34. H. Fujisaki *et al.*, Expansion of highly cytotoxic human natural killer cells for cancer cell therapy. *Cancer Res* **69**, 4010-4017 (2009).

Table 1

Table 1

	P1	P2	P3	P4	P5	P6	P7	P8	Total unique IS
IL2RG mutation	c. 823T>G	c447 del A	c.923C>A	c.341G>A	c. 31T> A	c.938+1G>A	c694G>A	c.31T> A	
Age	23	22	7	16	10	23	3	12	
Norovirus	NA	NA	Cleared		Cleared	Cleared	NA	Transient	
IgG production	Restored	Restored	Restored	Increase	Increased	Increased	Increased	Increased	
VISA followup (month)	84	24	54	48	54	48	42	38	
Unique IS in B cell	73,816	26,041	51,620	256,327	248,447	185,693	21,462	36,193	899,599
Unique IS in T cell	14,628	3,863	15,208	88,605	96,946	31,681	12,390	13,951	277,272
Unique IS in NK cell	19,142	8,779	9,590	115,598	98,103	79,869	10,290	47,208	388,579
Unique IS in granulocytes	29,947	10,762	36,275	122,735	211,781	157,350	11,266	82,961	663,077
Unique IS in monocytes	25,289	10,343	32,174	110,785	181,202	141,994	10,590	76,722	589,099
Total unique IS by patient	116,660	55,354	97,543	495,223	589,147	479,035	44,213	198,214	2,075,389

Footnotes

1. Cumulative unique integration sites (IS) for each lineage is shown. The unique IS are often shared among different lineages within a patient so the sum of all lineage in each patient is not equal to number of total unique IS in each patient. Unique IS are not shared among different patients.
2. P2 is deceased at 2.5 years after treatment, so IS accumulation in P2 is truncated.
3. P8 was retreated at 30 months.

NA Not applicable

N normal

Figures

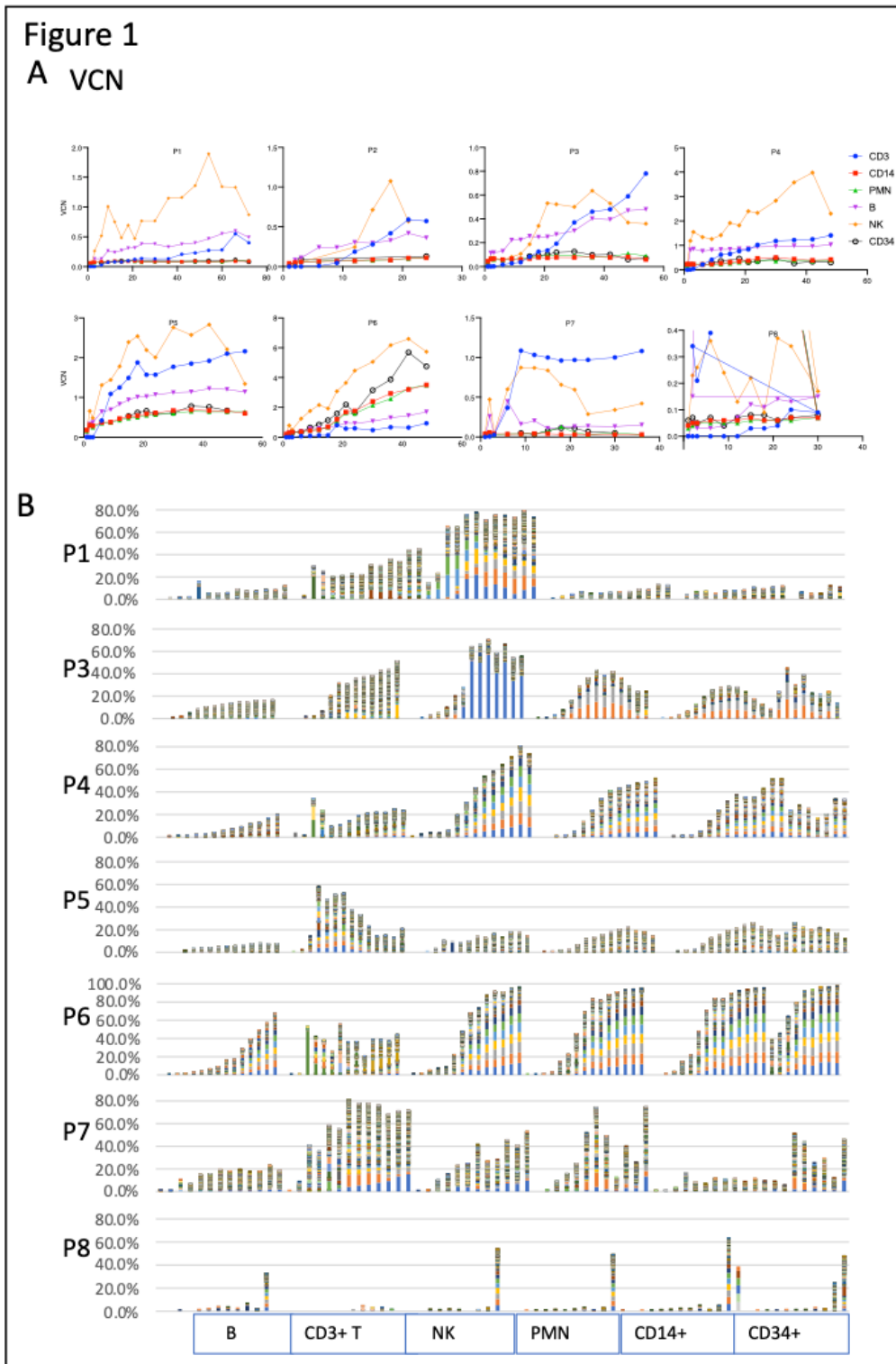


Figure 1

Chronological surveillance of gene marking and polyclonal vector insertion sites (VIS) in sorted immune cells following lentivector (LV) gene therapy in X-SCID patients (n=8). A. Surveillance of multi-lineage gene marking in sorted immune (CD3, CD14, PMN, B, NK and CD34+) cell lineages over time following gene therapy. B. Tracking of vector integration sites at different timepoint in sorted immune cell lineages for each patient treated with LV-gene therapy. Bar-chart shows top 256 VIS in each patient. Low frequency clones not shown. C. Tracking of vector integration sites at different timepoint in sorted immune cell lineages for each patient treated with LV-gene therapy. Bar-chart shows top 256 VIS in each patient. Low frequency clones not shown.

Figure 2

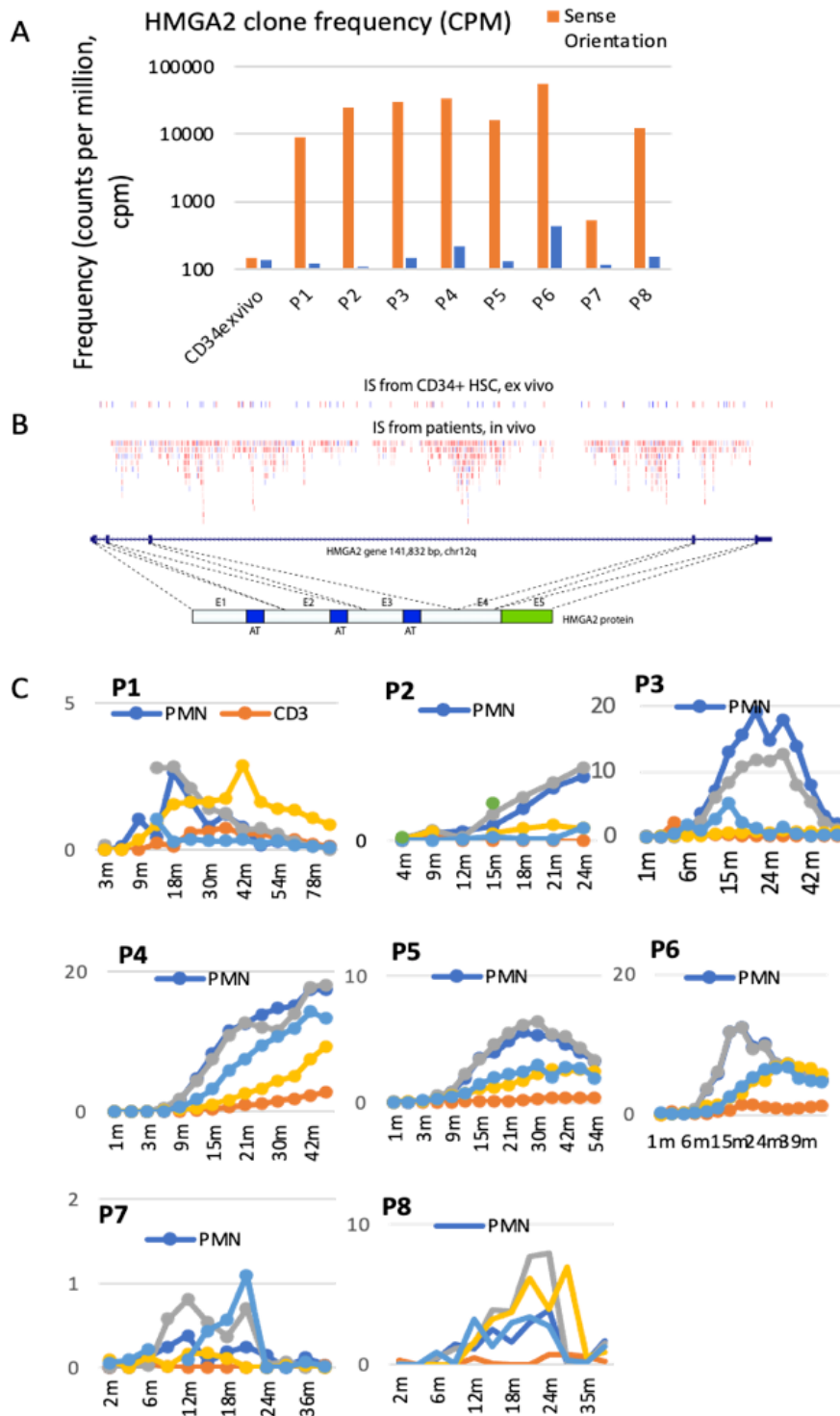
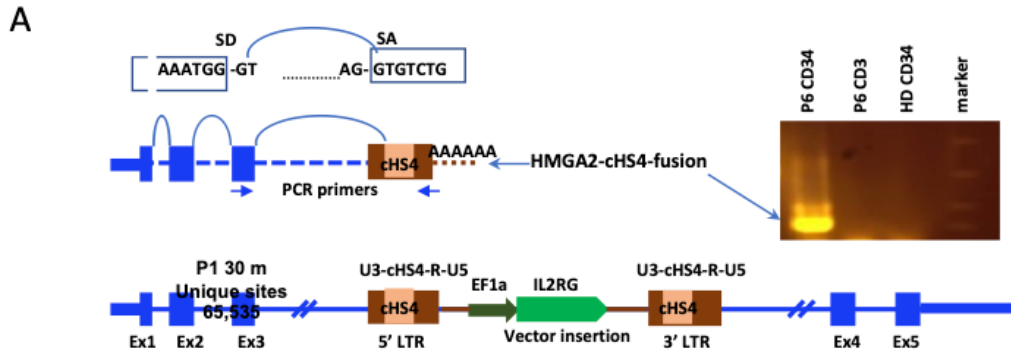


Figure 2

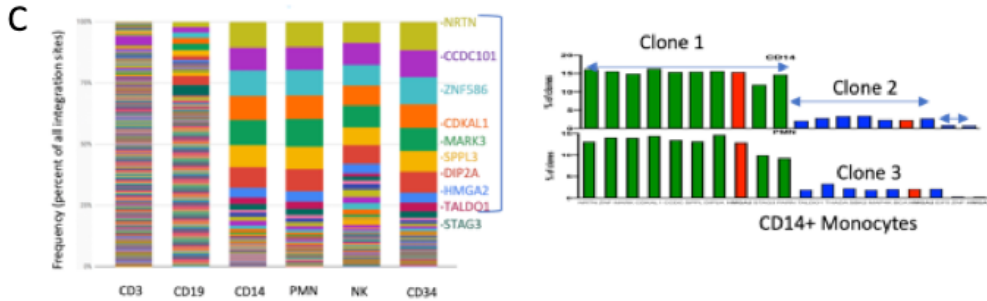
In vivo enrichment of HMGA2 clones with vector insertion in sense orientation and tracking of the highly enriched HMGA2 clones in all patients following LV gene therapy. A. Unique HMGA2 VIS containing clones are detected at higher frequency and HMGA2 VIS represent a greater proportion of all inserts detected from all patient samples compared to baseline ex vivo transduced CD34+ HSC product data sets. Shown is the frequency of detection of HMGA2 VIS calculated from cumulative VIS data for each patient including all timepoints and cell lineages, and then normalized as counts per million total VIS detected. Frequencies were calculated separately for both orientations of HMGA2 VIS provirus (Forward/Sense versus Reverse/Anti-Sense orientation as the HMGA2 gene). The in vivo frequency is much higher than that observed from the ex vivo transduced CD34+ HSC product for sense orientation, but not for antisense orientation. Note the Y axis is log10 scale. B. Shown is the distribution and frequency of unique Vector IS across the HMGA2 gene as detected in the ex vivo transduced CD34+ HSC product data sets (upper row) compared to the in vivo patient blood lineage derived data sets (second row). The blue or red colors, respectively, denote each unique HMGA2 VIS containing clone where provirus is in the sense orientation (blue) or anti-sense orientation (red) of the gene. The darker the color shading the larger the unique clone. Most insertion sites are in intron 3 of HMGA2 gene, which separates the Hmga2 protein to the N-terminal AT-hook DNA binding domain and C-terminal acidic domain. C. HMGA2 VIS clone dynamics in vivo by peripheral blood lineage in each Subject. HMGA2 VIS clone frequency was calculated as percent of all VIS at each timepoint in different lineages. In general, the highest frequency of HMGA2 VIS clones is seen in PMN, and CD14 cells, but also can be seen in other lineages like NK and CD19 cells. Note the different scale for the y-axis (percentage of total VIS).

Figure 3



B

	HD		P6				
	CD34	CD34	NK	PMN	CD14	CD3	CD19
Expression (Copies/ng)	90.5	153.7	25.9	ND	ND	ND	ND
Ex1-2/Ex4-5 ratio	1.01	5.91	7.55	ND	ND	ND	ND



D

10+copy clone 1A	10+copy clone 1B	10+copy clone 2	6+copy clone 3	3+copy clone 4
HMGA2(chr12+66327484)	HMGA2(chr12+66327484)	HMGA2(chr12+66249148)	HMGA2(chr12+66238698)	HMGA2(chr12+66308146)
CCDC101(chr16-28588950)	CCDC101(chr16-28588950)	ACACA(chr17-35624498)	BCAS3(chr17-58950329)	HCG4(chr6-29754777)
CDKAL1(chr6+21048330)	CDKAL1(chr6+21048330)	CHSY1(chr15-101690556)	EIF5(chr14-103808065)	BAIAP2L2(chr22-38487560)
DIP2A(chr21+47931106)	DIP2A(chr21+47931106)	CTSS(chr1-150731112)	MAP4K2(chr11+64553688)	
MARK3(chr14+103866754)	MARK3(chr14+103866754)	DEDO2(chr19-42711507)	SBK2(chr19+56031761)	
NRTN(chr19+5812281)	NRTN(chr19+5812281)	INPP5B(chr1-38352834)	THADA(chr2-43647506)	
PARN(chr16+14652221)	PARN(chr16+14652221)	SLC7A5(chr16-87899252)		
SPPL3(chr12+121246774)	SPPL3(chr12+121246774)	TEP1(chr14+20872943)		
STAG3(chr7+99783923)	STAG3(chr7+99783923)	TMEM187(chrX-153254921)		
ZNF586(chr19-58292195)	ZNF586(chr19-58292195)	TUBA1A(chr12-49598267)		
TALDO1(chr11-758954)		ZNF638(chr2-71613861)		

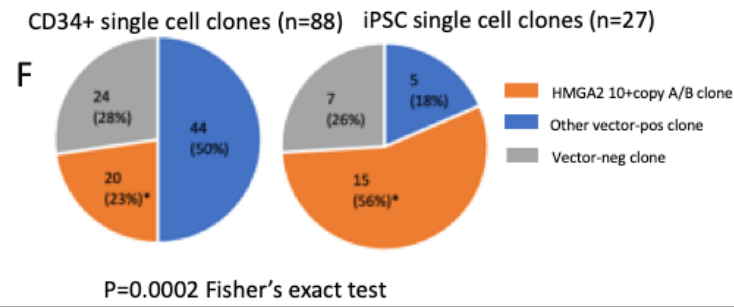


Figure 3

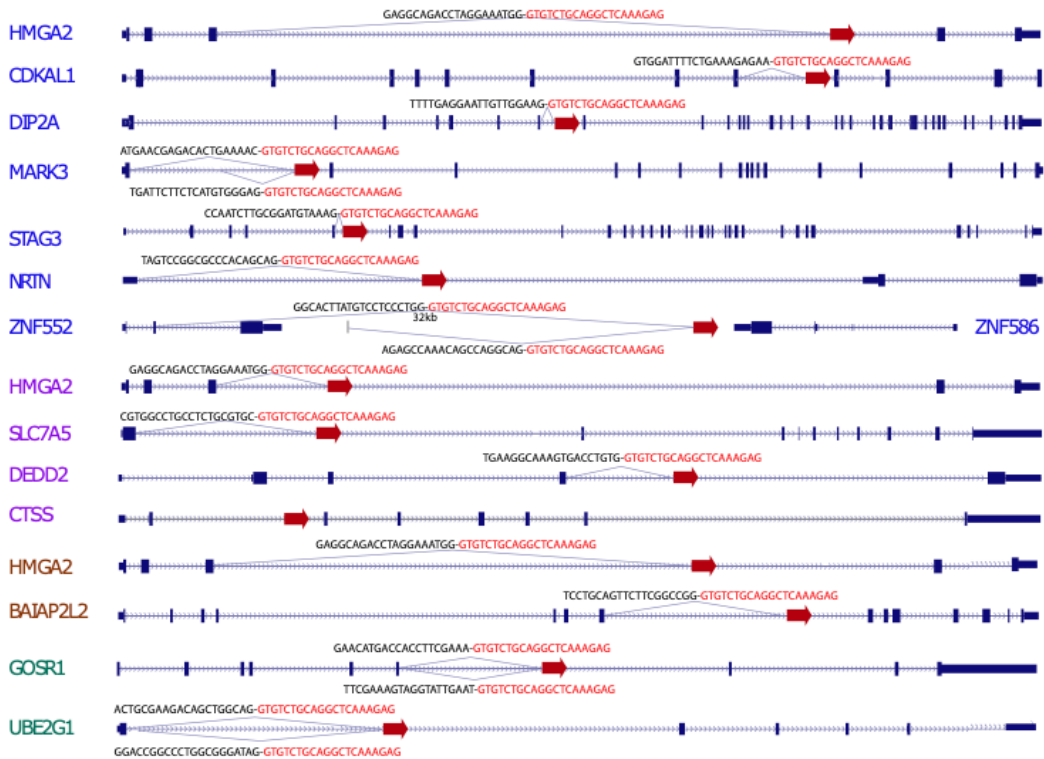
Characterization of HMGA2-cHS4 fusion transcripts in single cell CD34+ and derived-iPSCs with potential synergy in multi-insert clones. A. Illustration of lentivector insertion in the 3rd intron of HMGA2 gene in the bottom of the panel. Fusion transcript was generated by HMGA2 exon1-exon2-exon3 and spliced into the cHS4 SA in the 5LTR of the vector. Splicing Donor (SD) of HMGA2 exon3 and splicing acceptor (SA) of cHS4 sequences are shown on the top. PCR primers used to amplify the fusion transcript from P6 CD34

cells are located on HMGA2 exon3 and vector LTR. Amplified fusion transcript from RTPCR is shown in the right. B. Quantification of HMGA2 transcript in different blood lineages in P6 and the 5'/3' ratio, both determined by ddPCR as described in methods. HMGA2 is expressed in healthy donor (HD) CD34 cells, P6 CD34 cells, as well as P6 NK cells. It is not detected (ND) in P6 PMN, CD14, CD3, CD19 cells. The elevated 5'/3' ratio suggest accumulation of truncated transcript relative to full length transcript. C. P6 Clonality at 30m based on VISA. D. Single cell colony assay identified clones with multicopy transgenes. E. Karyotype in iPSCs derived from P6 peripheral blood CD34+cells. F. HMGA2-10+copy clones are the most abundant clones in peripheral CD34+ cells in PT6 and this clone is significantly enriched in iPSC cells derived from CD34+ cells (*p=0.0002, Fisher's exact test).

Figure 4

Gene Trap Events by lentivector insertions in 4 iPSC clones derived from patient CD34 cells

A. Lentivector insertion in the same orientation in the target gene and sequence junction of fusion transcripts of exons and cHS4



B. Lentivector insertion in opposite orientation of target gene or outside gene which no fusion transcripts observed

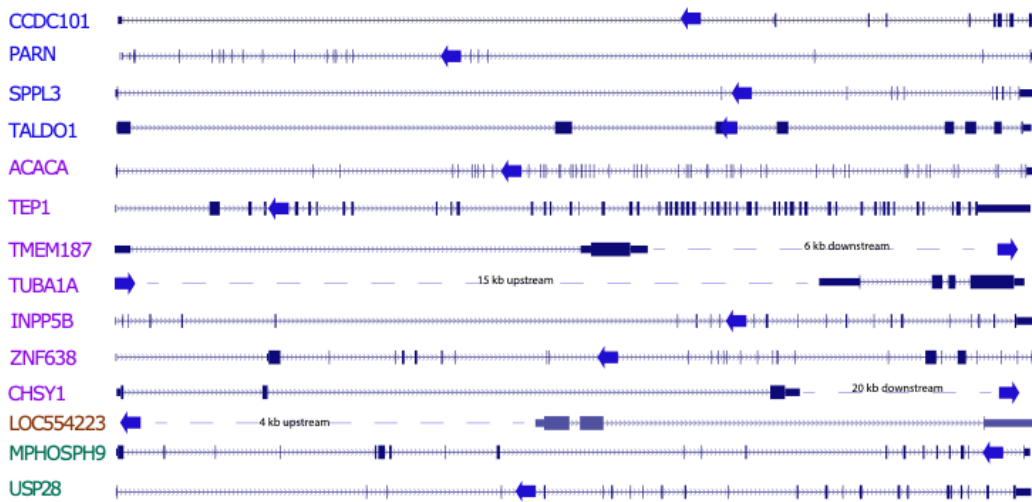
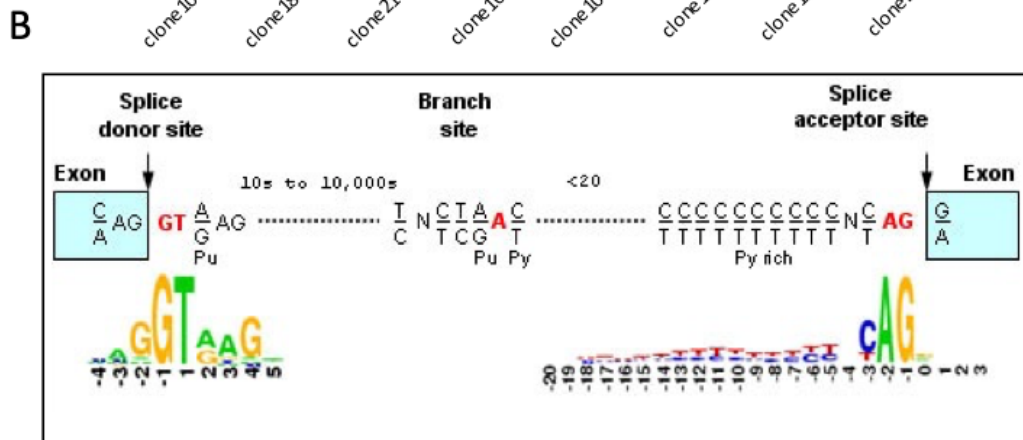
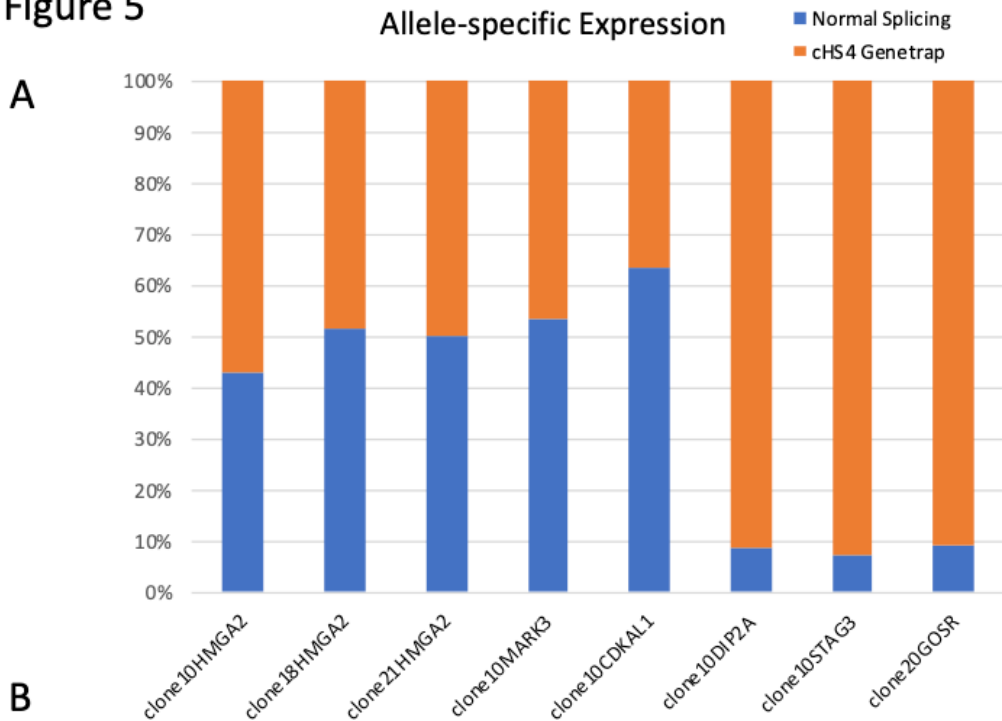


Figure 4

Cryptic splice acceptor in cHS4 causes transcription termination and aberrant fusion transcripts when inserted in same orientation of almost all target genes confirmed in iPS clones. No fusion transcript is observed for proviral integration with reverse orientation as the gene. Four iPSC clones were analyzed by RNAseq. Gene labels with the same color denote LV-insertion in the same iPSC clone, A. Almost all LV insertions that are in the same orientation as the insert gene generated fusion transcripts of the upstream

exons or even upstream introns into cHS4 in the LV, except for CTSS (not expressed in iPSCs, and therefore not possible to detect transcripts). Red arrows denote LV insertion site in the gene in the same orientation. Fusion transcript sequences were identified by RNAseq and labeled above the junction. Black text indicates the sequences from upstream exons/introns and red text indicates the sequences derived from cHS4. B. LV insertions in the reverse orientation of target gene or outside the gene do not produce fusion transcripts. Blue arrows indicate LV integration site and orientation relative to the gene. Insertion in TALDO1 is the only exonic insertion.

Figure 5



C

```

+
CCTGCAGGCA TCAAGGCCA GGCTGGATGT GGCTCTGGC AGCCTGGCT GCTGGTTGAT GACCCTGCAC 70
ATAGCAGGGG GTTGGATCTG GATGAGCACT GTGCTCCTTT GCAACCCAGG CCGTTCTATG ATTCTGTCAT 140
TCTAAATCTC TCTTTCAGCC TAAAGCTTTT TCCCCGATC CCC#1CAGG TG TCT#2CAGGCT CAAAGAGCAG 210
CGAGAAGCGT TCAGAGGAAA GCGATCCCGT GCCACCTCC CCGTGCCCGG GCTGTCCCG CACGCTGCCG 280
GCTCGGGGAT GCGGGGGGAG CGCCGACCG GAGCGGAGCC CCGGGCGGCT CGCTGCTGCC CCCTAGCGGG 350
GGAGGGACGT AATTACATCC CTGGGGGCTT TGGGGGGGG CTGTCCCCGT GAGCTCCCCA GATC
  
```

Figure 5

Dominant termination of insert alleles due to a cryptic splice acceptor in the cHS4 region of the vector. A. Quantification of exon-cHS4 gene trap splicing and normal exon-exon splicing of target genes in single cell iPS clones. RNAseq was performed for the iPS clones. Read counts across the exon-cHS4 splicing junction and normal exon-exon junction are reported as fraction of the total reads to measure the frequency of each splicing event. B. Figure depicting the consensus sequences for splice donor, branch and acceptor sites that corresponds to the lentivector cHS4 sequence. C. Vector sequence indicating branch site and cryptic splice sites. Single base pair changes from A to T were introduced at #1 and #2 to modify the vector.

Figure 6

In vitro CD34-derived T cells (ATO system)

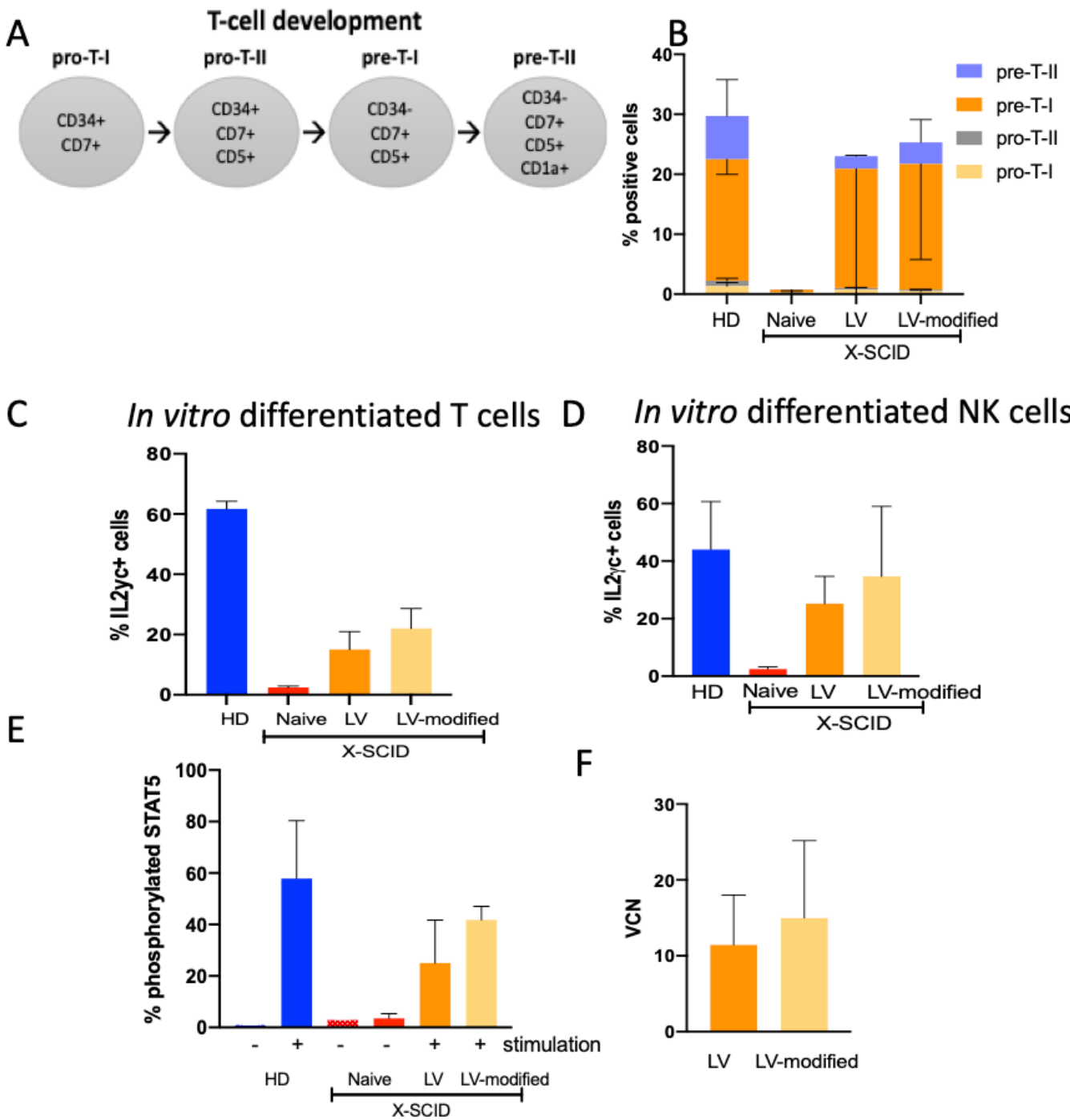


Figure 6

Modified vector with removal of cryptic SA corrects X-SCID CD34+ HSCs. Functional restoration of X-SCID CD34+ HSCs transduced by modified lentivector A. Stages of T-cell differentiation using the Artificial thymic organoid. B. Percentages of in vitro CD34-derived T-cell progenitors. C. Percentages of γ^+ cells (by flow cytometry) following in vitro CD34+ differentiation into T cells. D. Percentages of γ^+ NK cells (by flow cytometry) following in vitro differentiation. E. Phospho-STAT5 signaling in in vitro differentiated T

cells derived from transduced X-SCID CD34+ HSCs. F. Percentage of γ c+ cells analyzed by flow cytometry (CD132) following in vitro NK differentiation. G. Vector copy number determined by ddPCR in HSCs transduced by clinical lentivector (LV) or the modified vector (LV-modified).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [LVXSCIDHMGA2Suppl82621NCOMM.docx](#)
- [TableS2multiinsertclones.pdf](#)
- [TableS3SingleCelliPSCclonesFusionTranscripts.pdf](#)