Expanding the genetic code of the human hematopoietic system

Sida Shao[a,b], Minseob Koh[a,b], and Peter G. Schultz[a,b,1]

[a]Department of Chemistry, The Scripps Research Institute, La Jolla, CA 92037; and [b]Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037

Contributed by Peter G. Schultz, February 27, 2020 (sent for review August 20, 2019; reviewed by Abhishek Chatterjee and David R. Liu)

The genetic incorporation of noncanonical amino acids (ncAAs) into proteins has been realized in bacteria, yeast, and mammalian cells, and recently, in multicellular organisms including plants and animals. However, the addition of new building blocks to the genetic code of tissues from human origin has not yet been achieved. To this end, we report a self-replicating Epstein-Barr virus-based episomal vector for the long-term encoding of ncAAs in human hematopoietic stem cells and reconstitution of this genetically engineered hematopoietic system in mice.

genetic code expansion | Epstein-Barr virus | hematopoietic stem cell | hematopoietic reconstruction

The genetic encoding of noncanonical amino acids (ncAAs) with novel physical and chemical properties has provided useful tools to study and manipulate the structures and functions of proteins in vitro and in vivo (1). Examples include genetically encoded photocrosslinkers, photocaged amino acids, biophysical probes, and amino acids with orthogonal chemical reactivity (2). All that is required to encode an ncAA is an orthogonal tRNA/aminoacyl tRNA synthetase (aaRS) pair (i.e., not cross-reactive with the host translational machinery) that encodes the ncAA in response to a unique codon; for example, the amber stop or a frameshift codon. This methodology has been successfully employed in both prokaryotic cells such as *Escherichia coli* and mycobacteria and eukaryotic cells including *Saccharomyces cerevisiae* and mammalian cells on both the laboratory and industrial scales (3–5). Recently, this orthogonal ncAA encoding system has also been adapted from single cells to multicellular organisms, including *Danio rerio*, *Caenorhabditis elegans*, and *Mus musculus* (6–8). However, the methodology has not been extended to human systems to date. Here we demonstrate the genetic encoding of ncAAs in the human hematopoietic system by stable transgenic expression of the pyrrolysyl-tRNA synthetase (PylRS)/$tRNA^{Pyl}_{CUA}$ pair in human hematopoietic stem cells (HSCs) and show that these engineered HSCs can be engrafted into mice.

The hematopoietic system is responsible for the production of the cellular blood components including B cells, T cells, macrophages, neutrophils, and megakaryocytes, all of which play crucial roles in normal and disease processes (9). The entire hematopoietic system undergoes constant turnover throughout life through the self-renewal and multilineage differentiation capacity of HSCs. HSCs are one of the best characterized somatic stem cell populations and can be harvested (10). Moreover, the entire human hematopoietic system can be reconstituted in immunodeficient mice by engrafting human HSCs (11). These properties make the allogenic transplantation of HSCs in mice an attractive system for encoding ncAAs in a population of cells that gives rise to a large number of human cell types at an organism level.

Results and Discussion

We began by developing an efficient gene delivery method for long-term transgene expression and high tRNA expression in mammalian cells. Previous efforts in our laboratory and others involved transient transfection with plasmids, which becomes limiting for long-term maintenance of ncAA incorporation throughout the hematopoietic development process (5). More recent efforts have focused on generating stable cell lines expressing aaRS/tRNA pairs using the PiggyBac transposon system or stable transfection, but such methods are technically challenging with HSCs (12, 13). In addition, inefficient expression of the orthogonal suppressor tRNAs from archaea and bacteria is another major limiting factor for efficient incorporation of ncAAs in eukaryotes (14, 15). To address these challenges, we developed an expression system using an Epstein-Barr Virus (EBV)-based episomal vector for the encoding of ncAAs in human HSCs (16–20). EBV belongs to the human herpesvirus family and is able to maintain its genome extrachromosomally in the nucleus of mitotic primate cells by replicating and partitioning its genome equally into daughter cells (21). This episomal maintenance of its DNA is mediated by two virus-encoded components: a cis-acting latent origin of replication (OriP) and EBV nuclear antigen 1 (EBNA1). Another advantage of this system is that the vector containing OriP and EBNA1 is double-stranded, can be easily manipulated, and is highly amenable to transfection.

The orthogonal, polyspecific PylRS/$tRNA^{Pyl}_{CUA}$ pair from *Methanosarcina barkeri* was selected because of its ability to efficiently incorporate a diverse array of ncAAs, including the ε-tBoc-lysine derivative (**1**) (Fig. 1*B*) (1). To generate an efficient system for ncAA incorporation, we encoded both the PylRS/$tRNA^{Pyl}_{CUA}$ pair and a reporter gene in a single vector. Six copies of $tRNA^{Pyl}_{CUA}$ were included to increase tRNA expression levels, and to minimize the possibility of homologous recombination, two polymerase III promoters, U6 and H1, were used to drive expression of $tRNA^{Pyl}_{CUA}$ (14). EGFP(Y39TAG) under the CMV

Significance

The ability to genetically incorporate noncanonical amino acids (ncAAs) into proteins in bacteria, yeast, mammalian cells, and multicellular organisms has provided useful tools to study and manipulate protein structure and function. Here, we report the development of a method to stably encode ncAAs in human hematopoietic stem cells (HSCs) and their differentiated progenies, using self-replicating plasmids derived from Epstein-Barr virus. We also demonstrate that once transplanted, these manipulated HSCs can give rise to the entire hematopoietic system encoded with ncAAs. The methods described here should be applicable to other progenitor cell populations and provide tools for the study of human proteins in vitro and in vivo.

Author contributions: S.S. and P.G.S. designed research; S.S. and M.K. performed research; S.S., M.K., and P.G.S. contributed new reagents/analytic tools; S.S., M.K., and P.G.S. analyzed data; and S.S. and P.G.S. wrote the paper.

Reviewers: A.C., Boston College; and D.R.L., HHMI, Harvard University, Broad Institute.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: schultz@scripps.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1914408117/-/DCSupplemental.

First published April 6, 2020.

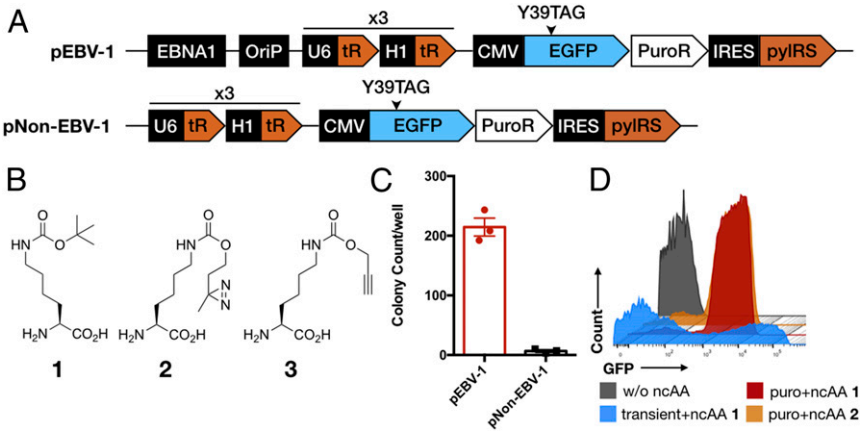

Fig. 1. EBV-based episomal vector for encoding ncAAs in HEK293T cells. (*A*) Structure of pEBV-1 and pNon-EBV-1 encoding *PylRS/tRNA$^{Pyl}_{CUA}$* and *EGFP(Y39TAG)* in one expression cassette. Arrowhead indicates EGFP reporter containing amber stop codon at the permissive site (Y39). IRES, internal ribosome entry site. (*B*) ncAAs used in this study. (*C*) Surviving HEK293T colonies from cells transfected with pEBV-1 and pNon-EBV-1 under puromycin selection. Colonies were counted 5 d after addition of 10 μg/mL puromycin (mean ± SEM). (*D*) Flow cytometry analysis of the HEK293T cells transiently transfected with pEBV-1 and stable HEK293T cells selected with puromycin. Cells were cultured in the presence of either 0.5 mM **1** or 1 mM **2**. Stable HEK293T cells cultured in the absence of ncAA were used as a negative control.

promoter and with an amber mutation at Y39 was used as the reporter, followed by PylRS driven by IRES element to afford the plasmid pNon-EBV-0 (*SI Appendix*, Fig. S1*A*). As a model system, HEK293T cells transfected with this single-vector system gave rise to a significantly higher fraction of cells expressing full-length EGFP in response to 0.5 mM **1** than a double-vector system where the EGFP(Y39TAG) and PylRS were encoded separately (*SI Appendix*, Fig. S1*B*). As expected, no EGFP expressing cells were observed in the absence of ncAA.

Next we investigated whether the *EBNA1-OriP* cassette is able to provide sustained transgene expression of the PylRS/tRNA$^{Pyl}_{CUA}$ pair. Because the *EBNA1-OriP* element has been reported to greatly enhance downstream gene transcription, we inserted the expression cassette after *EBNA1-oriP* in the backbone of an EBV-based episomal vector (22). For selection and enrichment, a puromycin resistance gene (containing a 5′ P2A sequence that directs self-cleavage) was placed between EGFP(Y39TAG) and the IRES to afford pEBV-1. An identical vector without the *OriP-EBNA1* cassette was also constructed as a nonepisomal vector control (pNon-EBV-1; Fig. 1*A*). HEK293T cells were transfected with pEBV-1 or pNon-EBV-1, and subsequently placed under selection with puromycin. Cells transfected with pEBV-1 gave rise to significantly more and larger colonies in the presence of 10 μg/mL puromycin than cells transfected with pNon-EBV-1, suggesting that cells transfected with pEBV-1 have higher plasmid retention and transgene expression levels (Fig. 1*C* and *SI Appendix*, Fig. S2).

To assess the efficiency of ncAA incorporation, colonies were then expanded in the presence of puromycin, and a uniform GFP fluorescence signal was detected in the cells cultured in the presence of 0.5 mM **1** or 1 mM *N6*-[[2-(3-methyl-3H-diazirin-3-yl) ethoxy]carbonyl]-L-lysine (**2**) (a photocrosslinking ncAA); ~90% and ~85% of cells expressed full-length EGFP in the presence of **1** and **2**, respectively, under long-term culture conditions (suppression efficiency 52% with 1mM **1**; Fig. 1 *B* and *D* and *SI Appendix*, Fig. S3*A*). In contrast, heterogenous fluorescence (~45% of cells expressing full-length EGFP with suppression efficiency of 60%) was detected when pEBV-1 was transiently transfected under short-term culture conditions in the presence of **1** (Fig. 1*D* and *SI Appendix*, Fig. S3*B*). Further, to verify that the pEBV-1 replicated inside HEK293T cells, extrachromosomal plasmids were extracted and digested with *DpnI* (which cleaves plasmids of bacterial origin), demonstrating replication of the EBV-based episomal vector in HEK293T cells (*SI Appendix*, Fig. S4 *A* and *B*). Sanger sequencing confirmed no mutations existed at the amber mutation at EGFP(Y39TAG), indicating that the GFP signals were a result of incorporation of **1** or **2** in EGFP(Y39TAG) (*SI Appendix*, Fig. S4*C*). These experiments show that the EBV-based episomal vector is able to provide prolonged transgene expression of the orthogonal tRNA/aaRS pair in the HEK293T cell line.

Given this positive HEK293T result, we asked whether the EBV-based episomal vector is able to maintain long-term encoding of ncAAs in primary human HSCs. We inserted an mCherry-TAG-maxGFP cassette into the EBV-based episomal vector to afford pEBV-2, which served as dual reporter system for transfection efficiency and ncAA incorporation (Fig. 2*A*). Primary human CD34$^+$ cells were then magnetically isolated from umbilical cord blood (CB) with higher than 99% purity as measured by CD34 expression (*SI Appendix*, Fig. S5). Following nucleofection with pEBV-2, CB CD34$^+$ cells were cultured in media containing 0.5 mM **1** and GFP fluorescence was detected within the mCherry-expressing population (mCherry$^+$ cells, 47%; maxGFP$^+$ mCherry$^+$ cells, 17.5%), indicating that **1** was successfully incorporated into the target protein in the primary human HSCs with reasonable efficiency (Fig. 2*B*). To determine whether CB CD34$^+$ cells nucleofected with pEBV-2 were able to maintain the long-term expression of PylRS/tRNA$^{Pyl}_{CUA}$ pair and pass it on to their differentiated progenies, the maxGFP$^+$ cells were FACS sorted and seeded in a colony-forming unit (CFU) assay. mCherry and maxGFP double-positive colonies were observed (percentage of mCherry$^+$ maxGFP$^+$ colonies, 13.1 ± 2.9%, mean ± SEM, $n = 3$) when cultured in the presence of 0.5 mM **1**, while no maxGFP-positive colonies were observed in the absence of ncAAs (Fig. 2*C*). To obtain a more homogenous population of PylRS/tRNA$^{Pyl}_{CUA}$-expressing HSC progenies, CD34$^+$ cells were nucleofected with pEBV-1 (suppression efficiency, 58% in CD34$^+$ cells) followed by puromycin selection and analyzed by CFU assay (*SI Appendix*, Fig. S3*C* and S6). When cultured in the presence of 0.5 mM **1**, 72% of the cells that survived in the presence of puromycin expressed full-length EGFP, whereas in the absence of puromycin, 29% of cells expressed full-length EGFP, likely as a result of the loss of the episomal vector (Fig. 2*D*).

Next, we examined whether multilineage differentiation potential is affected by the amber suppression machinery and ncAA **1**. As a result of stress induced by plasmid nucleofection, the colony-forming efficiency of CD34$^+$ cells nucleofected with pEBV-2 was significantly lower than untreated or mock-nucleofected cells (23).

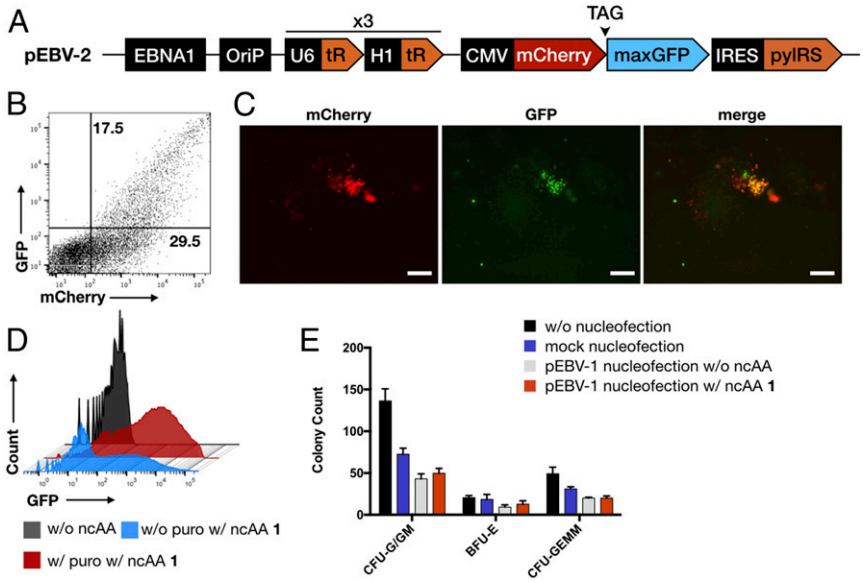

Fig. 2. EBV-based episomal vector mediated stable *PylRS/tRNA*$^{Pyl}_{CUA}$ expression in primary HSCs and the differentiated progenies. (*A*) Structure of pEBV-2, in which mCherry-TAG-maxGFP was used as reporter. (*B*) Flow cytometry analysis of primary CD34$^+$ cells nucleofected with pEBV-2. Cells were cultured for 24 h in the presence of 0.5 mM **1**. (*C*) Representative images of GFP$^+$ colony cultured in the presence of 0.5 mM **1** (scale bars, 100 μm). (*D*) Flow cytometry analysis of colonies cultured in the presence or absence of 7.5 μg/mL puromycin. HSCs were nucleofected with pEBV-1 and cultured in the presence of 0.5 mM **1** for 10 d; cells cultured in the absence of ncAA were the negative control. (*E*) Various types of colonies were counted for pEBV-1 nucleofected, mock nucleofected, or untreated CD34$^+$ in the presence or absence of 0.5 mM **1**. BFU-E, burst-forming unit-erythroid; CFU-G/GM, colony-forming unit-granulocyte/granulocyte, macrophage; CFU-GEMM, colony-forming unit-granulocyte, erythrocyte, macrophage, megakaryocyte.

Thus, 10-fold more pEBV-1 or pEBV-2 nucleofected HSCs were plated for this assay. Interestingly, colonies from each lineage were formed at comparable frequency in the presence or absence of either the amber suppression machinery or 0.5 mM **1**, indicating the differentiation potential of HSCs was not affected by addition of **1** or by the PylRS/tRNA$^{Pyl}_{CUA}$ pair (Fig. 2*E* and *SI Appendix*, Fig. S7). We also determined whether this system can be used to incorporate other ncAAs, which are known substrates of the polyspecific PylRS. Colonies expressing full-length EGFP were observed when cultured in the presence of 1 mM photocrosslinker **2** or 1 mM Nε-([prop-2-yn-1-yloxy]carbonyl)-L-lysine (**3**), a substrate for click reactions (*SI Appendix*, Fig. S8), and again we showed replication of pEBV-1 in HSCs. Taken together, these results demonstrate that the EBV-based vector is able to provide long-term amber suppression in primary HSCs, as well as in in vitro differentiated progenies without disrupting multilineage differentiation potential.

To explore whether this system can be extended to in vivo studies, we next determined whether the CD34$^+$ cells nucleofected with EBV-based vectors are able to engraft in NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ (NSG) mice and maintain their ability to encode ncAAs. Likely as a result of toxicity associated with the exogenous double-strand DNA, the level of engraftment with pEBV-2 nucleofected maxGFP$^+$ CD34$^+$ cells was ∼3-fold lower than that with mock nucleofected cells, as measured by the percentage of human CD45$^+$ population in the bone marrow by the end of week 12 (23) (*SI Appendix*, Fig. S9; pEBV-2 nucleofected, 9.5 ± 2.3%; mock nucleofected, 27.8 ± 3.7%, mean ± SEM, n = 5). Robust repopulation was achieved by engrafting more maxGFP$^+$ CD34$^+$ cells (1 or 2 million/mouse) compared with the mock-nucleofected counterpart (0.25 million/mouse). Bone marrow engraftment analysis revealed that 26.3 ± 5.0% (mean ± SEM, n = 5) of cells are human origin. Moreover, we confirmed the multilineage differentiation potential by detecting the HSC progenies including B cells, myeloid cells, and T cells (*SI Appendix*, Fig. S10). These results suggested that, while having lower viability, CD34$^+$ cells nucleofected with the EBV-based vector are multipotent and competent for hematopoietic system reconstitution.

We next tested the in vivo incorporation of ncAAs using **1**, as it is commercially available, can be incorporated with high efficiency, and has good oral bioavailability (71%) (*SI Appendix*, Fig. S11) and no apparent toxicity at millimolar concentrations in cell culture. Following nucleofection with either pEBV-1 or pEBV-2, GFP$^+$ CD34$^+$ cells were transplanted in NSG mice resulting in consistent chimeric levels in bone marrow (percentage of human CD45$^+$ cells in bone marrow [pEBV-1, 28.1 ± 2.3%; pEBV-2, 23.8 ± 2.6%, mean ± SEM, n = 15; *SI Appendix*, Fig. S12]). Then, 150 mg/kg of **1** was administered daily via oral gavage, starting from the sixth week of repopulation, to afford ∼100 μM plasma levels of the ncAA. GFP$^+$ cells were detected with both pEBV-1 and pEBV-2 nucleofected groups when fed with **1**, while no GFP$^+$ cells were identified in the absence of **1** (Fig. 3 and *SI Appendix*, Fig. S13; percentage of EGFP$^+$ cells: pEBV-1 +**1**, 14.9 ± 2.0%, pEBV-1 −**1**, 0.05 ± 0.01%; pEBV-2 +**1**, 23.1 ± 5.8%, pEBV-2 −**1**, 0.04 ± 0.03%, mean ± SEM, n = 5). As expected, the percentage of the EGFP$^+$ population dropped to basal level when the feeding of **1** was withdrawn for 1 wk, indicating the full-length GFP expression was dependent on the presence of ncAA (Fig. 3 and *SI Appendix*, Fig. S13). Lineage differentiation potential was confirmed by detecting CD19$^+$ cells within the EGFP$^+$ population from engrafted mice. Furthermore, episomal replication of pEBV-1 in isolated CD19$^+$ cells was confirmed by the *DpnI* digestion assay (*SI Appendix*, Fig. S14). Although only ∼20% of the HSCs engrafted in mice encoded the ncAA, development of an appropriate selectable in vivo system could potentially give rise to more homogenous hematopoietic populations. Nonetheless, this level should be sufficient for many cellular studies (e.g., photocrosslinking, bioorthogonal conjugation, imaging).

In conclusion, we have been able to encode ncAAs in primary human HSCs in vitro, and to engraft the engineered hematopoietic system in mice. This was achieved using an EBV-based vector, in which the orthogonal translational suppression machinery was

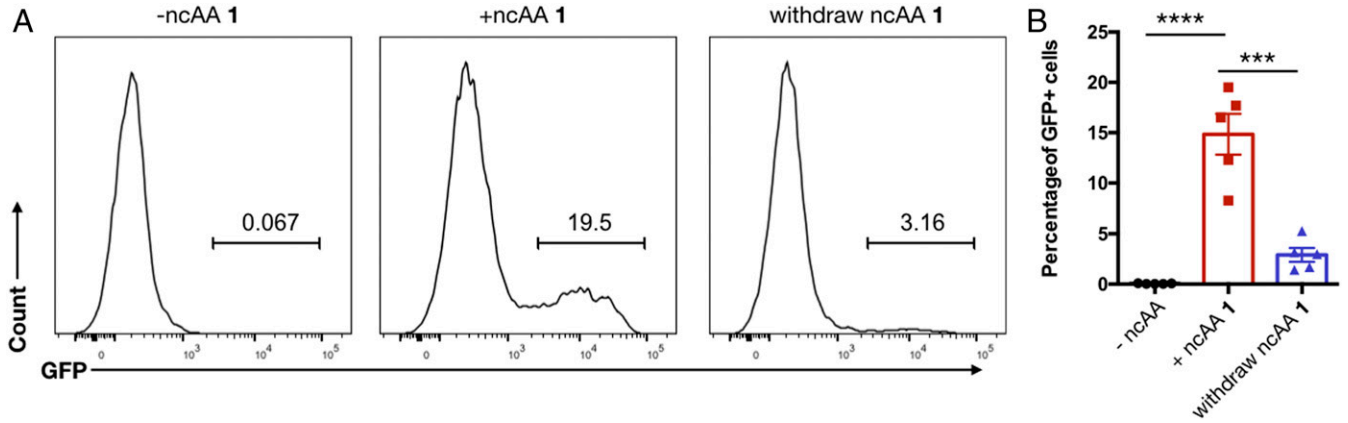

Fig. 3. Incorporation of ncAAs in reconstituted human hematopoietic system. (*A*) Flow cytometry analysis of representative engraftment in bone marrow showing successful incorporation of **1** when administered by mouth at 150 mg/kg daily. The hematopoietic system reconstituted from $CD34^+$ cells nucleofected with pEBV-1 consisted of an $EGFP^+$ population which was dependent on the addition of **1**. Numbers in the plot indicated frequencies of the gated population. (*B*) Percentage of $EGFP^+$ population in the human graft. $n = 5$. $^{***}P < 0.0005$; $^{****}P < 0.0001$, *t* test.

expressed efficiently and stably. This work can likely be expanded to other ncAAs, including photocrosslinkers, biophysical probes, and amino acids with bioorthogonal chemical handles. It will also be of interest to challenge these mice with immunogens and determine whether ncAAs can play a role in the antibody response.

Materials and Methods

HEK293T Cell Culture, Transfection, and Puromycin Selection. HEK293T cells were cultured in high glucose DMEM (Corning) supplemented with 10% heat-inactivated FBS (Gibco) and incubated at 37 °C and 5% CO_2. Plasmid transfection was performed with Fugene HD transfection reagent (Promega) following manufacturer's instruction, and 0.5 mM ncAA **1** (Sigma-Aldrich) or 1 mM ncAA **2** (Tocris) was added at the time of transfection. For transient transfection experiments, cells were monitored 36 h posttransfection, using fluorescent microscopy (Nikon). For clonal selection and generation of stable cell lines, 10 μg/mL puromycin dihydrochloride (Gibco) was included in culture media 36 h posttransfection. Culture media containing puromycin was changed every day for the first 5 d. The puromycin-resistant stable HEK293T cell line was maintained in complete media with 10 μg/mL puromycin.

Plasmid Construction. The backbone of the EBV-based episomal vectors used in this study was derived from pCE-GFP (Addgene, Plasmid #41858), in which the GFP and the terminator element were removed by *BsaAI* and *HindIII* digestion. The resulting linear backbone was used for insertion of gene expression cassettes. A pUC19 backbone was employed as a shuttle vector, in which all expression cassettes were constructed before transfer into the EBV backbone. Elements including CMV promoter, EGFP(Y39TAG), puroR, mCherry-TAG-maxGFP, IRES, and PylRS were constructed in pUC19 stepwise by assembling each fragment with linearized pUC19 vector, using the Gibson Assembly method. Final expression cassettes assembled in pUC19 vector were transferred into the linearized EBV backbone, using the Gibson Assembly method. All DNA oligonucleotides were purchased from IDT. Plasmid maps are provided in *SI Appendix*, Fig. S15.

HSC Isolation and Culture. Human $CD34^+$ cells were purified from fresh umbilical cord blood purchased from the San Diego Blood Bank with patient consensus. Mononuclear cells were separated from cord blood using Ficoll-Paque PLUS (GE Healthcare), and the $CD34^+$ cells were subsequently purified using the EasySep Human Cord Blood CD34 Positive Selection Kit (StemCell Technologies) following manufacturer's instruction. Isolated $CD34^+$ cells were cultured in HSC expansion media (StemSpan SFEM II [StemCell Technologies], supplemented with the following recombinant human cytokines: thrombopoietin, IL6, Flt3 ligand, and stem cell factor [100 ng/mL for all, R&D Systems], and 1 μM SR1 [Selleckchem]) for 5 d before nucleofection.

HSC Nucleofection. For each nucleofection assay, a total of 5×10^6 cells were mixed with 10 μg pEBV-1 or pEBV-2 (plasmid stock at concentration greater than 2 μg/μL) in 100 μL nucleofection buffer (Lonza, Human CD34+ Cell Nucleofector Kit) and transferred into a nucleofection cuvette. Nucleofection was performed on Amaxa Nucleofector II (Lonza) with program U-008 following manufacturer's instruction. The nucleofected cells were resuspended with HSC expansion media and cultured in the presence or absence of ncAAs at the annotated concentrations for 48 ~ 72 h at 37 °C in a 5% CO_2 incubator for further experiments.

CFU Assay. A total of 20,000 untreated human $CD34^+$ cells or 200,000 FACS sorted GFP^+ human $CD34^+$ cells were suspended in 400 μL IMDM (Gibco) supplemented with 2% FBS (Gibco) and mixed with 4 mL semisolid methylcellulose media (StemCell Technologies, MethoCult H4435 Enriched) for triplicate cultures. For ncAA incorporation experiments, ncAAs were added to the culture at the annotated concentration. For puromycin selection experiments, 7.5 μg/mL puromycin was added to the culture. Next, 1.1 mL methylcellulose cell cultures were plated in 3.5-mm cell culture dishes and incubated for 10 d for manual counting and further analysis.

Engraftment Assay. NOD.Cg-*Prkdc*^{scid}*Il2rg*^{tm1Wjl}/SzJ (NSG) mice were purchased from The Jackson Laboratory (Stock #: 005557) and bred in house. The animal study was performed under protocols approved by the Institutional Animal Care and Use Committee (IACUC) of The Scripps Research Institute (IACUC protocol #17-0013). Mice were housed in a restricted-access, pathogen-free vivarium. For $CD34^+$ cell transplantation, 6- to 8-wk-old female NSG mice were sublethally irradiated (250cGy-300cGy, depending on the body weight) with Cesium^{137} (Gammacell 40 Exactor) 24 h ahead of transplantation. A total of 2.5×10^5 untreated $CD34^+$ cells or 10^6 (for *SI Appendix*, Fig. S7) and 2×10^6 (for rest of the studies) sorted GFP^+ $CD34^+$ cells were washed and resuspended in 100 μL PBS and i.v. injected into irradiated mice via tail vain. Mouse activity was monitored on a daily basis during the first week of transplantation. Engraftment was monitored by analyzing mononuclear cells from blood obtained via retro-orbital sinus with flow cytometry. The mice were sacrificed at the end of week 12 (for *SI Appendix*, Fig. S6) or week 8 (for rest of the studies) after transplantation; bone marrow (from femurs and tibiae) was harvested for analysis. For the ncAA feeding experiments, ncAA **1** was dissolved in acidified water at 20 mg/mL, and 200 μL ncAA **1** solution was administered by mouth to mouse through oral gavage daily for 1 wk. For the control group, the same volume of vehicle was administered by mouth for the same period.

Assessment of Human Engraftment. At week 12 after transplantation, mice were euthanized. Bone marrow was harvested from femur and tibia. MNCs were enriched using Ficoll-Paque PLUS (GE Healthcare). B cells, T cells, and myeloid cells were stained with anti-hCD19 (clone HIB19), anti-hCD3 (clone SK7), anti-hCD33 (clone P67.6) antibodies, respectively, and analyzed by flow cytometry with LSR II flow cytometer (Becton Dickinson).

FACS Sorting and Flow Cytometry. Flow cytometric analysis for cultured $CD34^+$ cells and in vivo grafts were performed on a LSR II flow cytometer (Becton Dickinson). Cells were stained in staining media (HBSS supplemented with 2% FBS and 2 mM EDTA) at 4 °C for 1 h with the following antibodies: PE anti-hCD34, clone 581 (StemCell Technologies); Pacific Blue anti-mCD45, clone 30-F11 (BioLegend); APC anti-hCD45, clone HI30 (BioLegend); Pacific

Blue anti-hCD19, clone HIB19 (BioLegend); PE-Cy7 anti-hCD3, clone SK7 (BD Biosciences); and PE-Cy7 anti-hCD33, clone P67.6 (BD Biosciences). FACS sorting experiments were performed on a BD influx cell sorter at the Sanford Consortium for Regenerative Medicine. After 48 ~ 72 h culture of nucleofected $CD34^+$ cells supplemented with ncAA, GFP^+ cells were sorted into complete culture media in the low-pressure mode. Cells were placed on ice until seeding in methylcellulose culture or injecting into NSG mice.

Plasmid Extraction from Mammalian Cells. Episomal plasmid was extracted from HEK293T cells or CFU colonies with a modified Hirt extraction protocol. Briefly, 10^7 HEK293T cells or CFU colonies from one 3.5-mm dish were resuspended in media containing 50 mM Tris·HCl, 10 mM EDTA, and 100 μg/mL RNase A and then lysed by 1.2% SDS. Chromosomal DNA was precipitated by adding an equal volume of 3 M CsCl, 1 M potassium acetate, and 0.67 M acetic acid. After centrifugation, episomal plasmid DNA was isolated from supernatant with a DNA extraction column. Extracted episomal DNA was digested with *DpnI* before transformation into *E. coli* for colony analysis.

Data Availability. Source data for all mouse experiments have been provided in Dataset S1.

ACKNOWLEDGMENTS. This work was supported by NIH R01 GM132071. We thank Jesus Olvera and Cody Fine for help with FACS sorting at stem cell core facility. We acknowledge Dr. Angad P. Mehta and Dr. Han Xiao for helpful discussions and Kristen Williams for her assistance in manuscript preparation.

1. C. C. Liu, P. G. Schultz, Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79**, 413–444 (2010).
2. D. D. Young, P. G. Schultz, Playing with the molecules of life. *ACS Chem. Biol.* **13**, 854–870 (2018).
3. L. Wang, A. Brock, B. Herberich, P. G. Schultz, Expanding the genetic code of Escherichia coli. *Science* **292**, 498–500 (2001).
4. J. W. Chin *et al*., An expanded eukaryotic genetic code. *Science* **301**, 964–967 (2003).
5. W. Liu, A. Brock, S. Chen, S. Chen, P. G. Schultz, Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat. Methods* **4**, 239–244 (2007).
6. S. Greiss, J. W. Chin, Expanding the genetic code of an animal. *J. Am. Chem. Soc.* **133**, 14196–14199 (2011).
7. Y. Chen *et al*., Heritable expansion of the genetic code in mouse and zebrafish. *Cell Res.* **27**, 294–297 (2017).
8. R. J. Ernst *et al*., Genetic code expansion in the mouse brain. *Nat. Chem. Biol.* **12**, 776–778 (2016).
9. J. Seita, I. L. Weissman, Hematopoietic stem cell: Self-renewal versus differentiation. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2**, 640–653 (2010).
10. M. Kondo *et al*., Biology of hematopoietic stem cells and progenitors: Implications for clinical application. *Annu. Rev. Immunol.* **21**, 759–806 (2003).
11. L. D. Shultz, F. Ishikawa, D. L. Greiner, Humanized mice in translational biomedical research. *Nat. Rev. Immunol.*, **7**, 118–130 (2007).
12. S. J. Elsässer, R. J. Ernst, O. S. Walker, J. W. Chin, Genetic code expansion in stable cell lines enables encoded chromatin modification. *Nat. Methods* **13**, 158–164 (2016).
13. L. Si *et al*., Generation of influenza A viruses as live but replication-incompetent virus vaccines. *Science* **354**, 1170–1173 (2016).
14. A. Chatterjee, H. Xiao, M. Bollong, H. W. Ai, P. G. Schultz, Efficient viral delivery system for unnatural amino acid mutagenesis in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11803–11808 (2013).
15. S. Greiss, J. W. Chin, Expanding the genetic code of an animal. *J. Am. Chem. Soc.* **133**, 14196–14199 (2011).
16. F. Grignani *et al*., High-efficiency gene transfer and selection of human hematopoietic progenitor cells with a hybrid EBV/retroviral vector expressing the green fluorescence protein. *Cancer Res.* **58**, 14–19 (1998).
17. E. P. Papapetrou, N. C. Zoumbos, A. Athanassiadou, Genetic modification of hematopoietic stem cells with nonviral systems: Past progress and future prospects. *Gene Ther.* **12** (suppl. 1), S118–S130 (2005).
18. S. Banerjee, E. Livanos, J. M. H. Vos, Therapeutic gene delivery in human B-lymphoblastoid cells by engineered non-transforming infectious Epstein-Barr virus. *Nat. Med.* **1**, 1303–1308 (1995).
19. M. Conese, C. Auriche, F. Ascenzioni, Gene therapy progress and prospects: Episomally maintained self-replicating systems. *Gene ther.* **11**, 1735–1741 (2004).
20. D. H. Crawford, Biology and disease associations of Epstein-Barr virus. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **356**, 461–473 (2001).
21. E. R. Leight, B. Sugden, Establishment of an oriP replicon is dependent upon an infrequent, epigenetic event. *Mol. Cell. Biol.* **21**, 4149–4161 (2001).
22. F. Längle-Rouault *et al*., Up to 100-fold increase of apparent gene expression in the presence of Epstein-Barr virus oriP sequences and EBNA1: Implications of the nuclear import of plasmids. *J. Virol.* **72**, 6181–6185 (1998).
23. P. Genovese *et al*., Targeted genome editing in human repopulating haematopoietic stem cells. *Nature* **510**, 235–240 (2014).