

**UCLA**

**UCLA Previously Published Works**

**Title**

Creating New  $\beta$ -Globin-Expressing Lentiviral Vectors by High-Resolution Mapping of Locus Control Region Enhancer Sequences

**Permalink**

<https://escholarship.org/uc/item/4sh7x4pq>

**Authors**

Morgan, Richard A  
Ma, Feiyang  
Unti, Mildred J  
et al.

**Publication Date**

2020-06-01

**DOI**

10.1016/j.omtm.2020.04.006

Peer reviewed

# Creating New $\beta$ -Globin-Expressing Lentiviral Vectors by High-Resolution Mapping of Locus Control Region Enhancer Sequences

Richard A. Morgan,<sup>1,2</sup> Feiyang Ma,<sup>3</sup> Mildred J. Unti,<sup>4</sup> Devin Brown,<sup>4</sup> Paul George Ayoub,<sup>4</sup> Curtis Tam,<sup>4</sup> Lindsay Lathrop,<sup>4</sup> Bamidele Aleshe,<sup>4</sup> Ryo Kurita,<sup>5</sup> Yukio Nakamura,<sup>5</sup> Shantha Senadheera,<sup>4</sup> Ryan L. Wong,<sup>2</sup> Roger P. Hollis,<sup>4</sup> Matteo Pellegrini,<sup>3</sup> and Donald B. Kohn<sup>2,4,6,7</sup>

<sup>1</sup>Charles R. Drew University of Medicine and Science, Los Angeles, CA 90059, USA; <sup>2</sup>Department of Molecular and Medical Pharmacology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; <sup>3</sup>Molecular Biology Institute Interdepartmental Doctoral Program, University of California, Los Angeles, Los Angeles, CA 90095, USA; <sup>4</sup>Department of Microbiology, Immunology & Molecular Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; <sup>5</sup>Cell Engineering Division, RIKEN BioResource Center, Tsukuba, Ibaraki, Japan; <sup>6</sup>Department of Pediatrics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; <sup>7</sup>The Eli & Edythe Broad Center of Regenerative Medicine & Stem Cell Research, University of California, Los Angeles, Los Angeles, CA, USA

**Hematopoietic stem cell gene therapy is a promising approach for treating disorders of the hematopoietic system. Identifying combinations of *cis*-regulatory elements that do not impede packaging or transduction efficiency when included in lentiviral vectors has proven challenging. In this study, we deploy LV-MPRA (lentiviral vector-based, massively parallel reporter assay), an approach that simultaneously analyzes thousands of synthetic DNA fragments in parallel to identify sequence-intrinsic and lineage-specific enhancer function at near-base-pair resolution. We demonstrate the power of LV-MPRA in elucidating the boundaries of previously unknown intrinsic enhancer sequences of the human  $\beta$ -globin locus control region. Our approach facilitated the rapid assembly of novel therapeutic  $\beta^{\text{AS3}}$ -globin lentiviral vectors harboring strong lineage-specific recombinant control elements capable of correcting a mouse model of sickle cell disease. LV-MPRA can be used to map any genomic locus for enhancer activity and facilitates the rapid development of therapeutic vectors for treating disorders of the hematopoietic system or other specific tissues and cell types.**

## INTRODUCTION

Hematopoietic stem cell gene therapy is a promising approach for treating many monogenic disorders of the hematopoietic system, having demonstrated remarkable success in a number of phase I/II clinical trials.<sup>1–12</sup> Therapeutic lentiviral vectors typically contain human *cis*-regulatory elements that provide high-level, lineage-specific expression of a therapeutic transgene. Identification of appropriate *cis*-regulatory elements that do not impede packaging or transduction efficiency when included in lentiviral vectors (LVs) has been a major challenge in developing novel gene therapy vectors.<sup>13–19</sup>

The elements used in LVs to confer tightly regulated expression of therapeutic transgenes are often identified by contrasting mouse

knockout studies<sup>20–28</sup> to histone marking and accessibility data<sup>29</sup> to determine putative boundaries of *cis*-regulatory elements. Although aforementioned technologies have enabled the identification of candidate *cis*-regulatory elements for incorporation into LVs, these technologies fail to provide insight on the exact base pair (bp) boundaries of bona fide “sequence-intrinsic” enhancers (i.e., the actual sequences providing enhancer activity). Moreover, when chromatin marking and accessibility data are used to identify putative regulatory elements, detection of appropriate cell type-specific control elements can be confounded by extrinsic factors such as chromatin state, cell cycle, cell state, cell type, antibody sensitivity/specificity, and protein evanescence. The proposed control sequences must still be functionally tested for their ability to enhance transgene expression, which is a low-throughput process when attempting to identify the best combination of elements to include in LVs.

To overcome the limitations of current methods used to detect cell-specific enhancer elements for use in therapeutic vector design, we developed a massively parallel reporter assay (MPRA) to pinpoint and identify boundaries of human genomic sequences capable of driving cell type-specific LV reporter gene expression. Similar assays have been used to map promoter activity<sup>30</sup> and measure enhancer activity,<sup>31–34</sup> and they have provided novel insights into the complex relationships between nucleotide composition and regulatory activity.<sup>35–40</sup>

Received 8 April 2020; accepted 13 April 2020;  
<https://doi.org/10.1016/j.omtm.2020.04.006>.

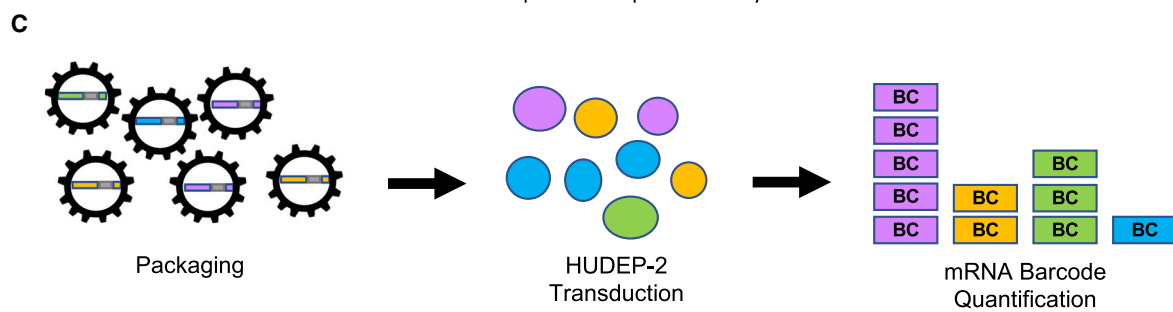
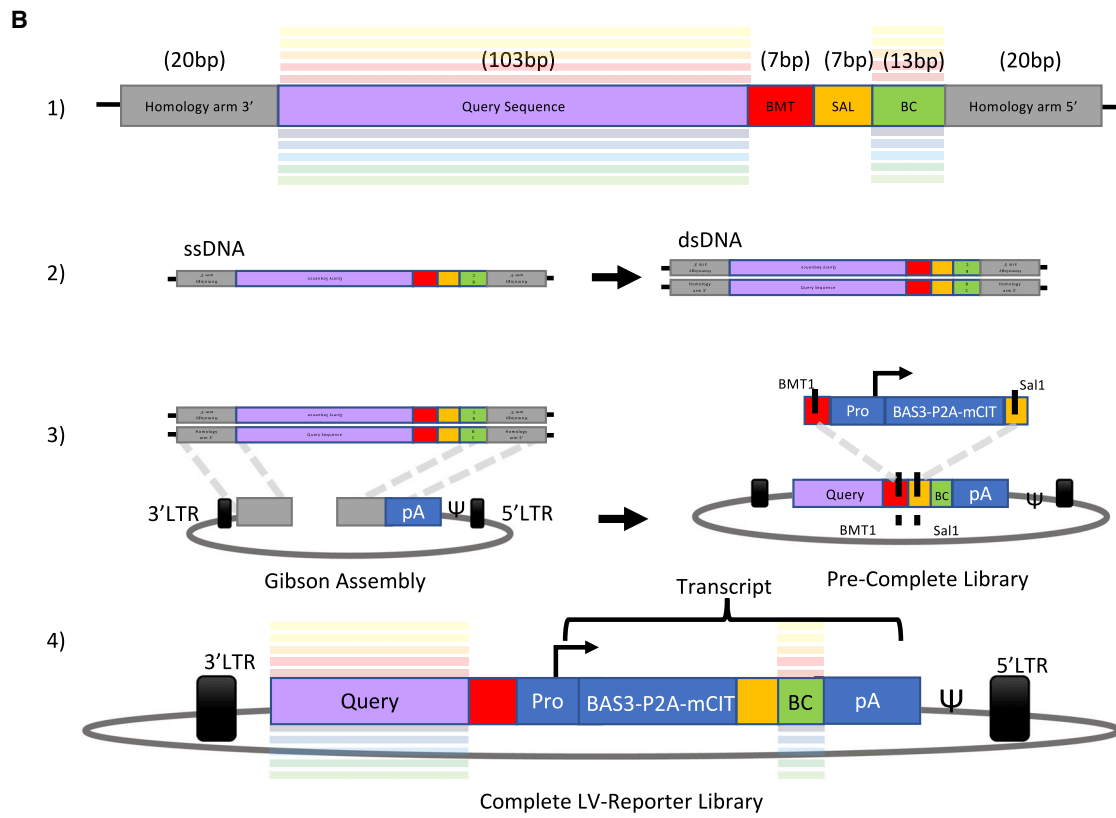
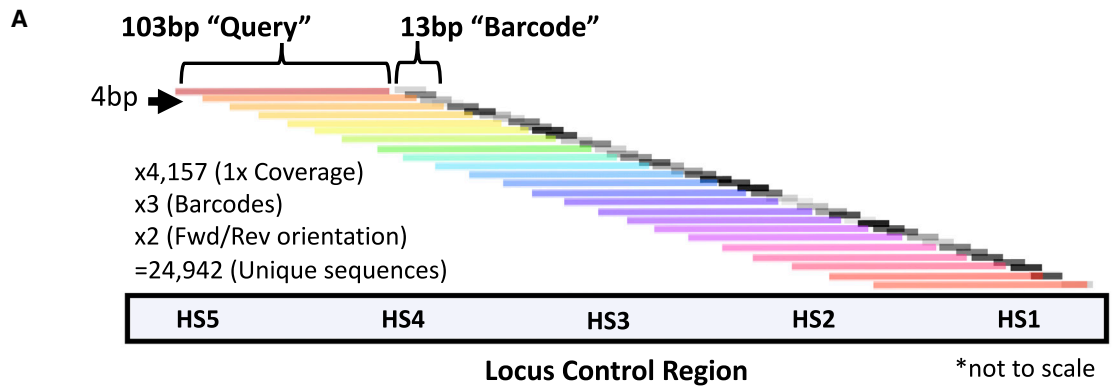
**Correspondence:** Richard A. Morgan, Charles R. Drew University of Medicine and Science, Los Angeles, CA 90059, USA.

**E-mail:** [rmorgan@mednet.ucla.edu](mailto:rmorgan@mednet.ucla.edu)

**Correspondence:** Donald B. Kohn, MD, Department of Molecular and Medical Pharmacology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA.

**E-mail:** [dkohn@mednet.ucla.edu](mailto:dkohn@mednet.ucla.edu)





(legend on next page)

We chose the well-studied human  $\beta$ -globin locus control region (LCR) as our target, a region known to contain potent lineage-specific enhancers that drive the temporal and developmentally specific expression of the  $\beta$ -globin gene locus.<sup>41</sup> Previous studies examining the influence of each LCR DNase hypersensitive site on transgene expression incorporated minimal hypersensitive sites (HS) sequences (typically 200–400 bp [referred to as HS core regions]) with flanking sequences into LVs to achieve position-independent expression.<sup>18,42</sup> Full enhancer activity of any given HS sequence was only realized once larger sequences containing both core and associated flanking sequences were incorporated into LVs, work that was largely enabled by the significant contributions of the Engel, Grosveld, and Stamatoyannopoulos laboratories, among others.<sup>20,43,44</sup> While DNA sequences within the cores are very well conserved, as are blocks of sequences found outside the core regions,<sup>45</sup> an exhaustive survey examining the functional significance of LCR core and flanking sequences had yet to be implemented at high-resolution before our effort.

In this study, we show that LV-based MPRA (LV-MPRA) was able to map the functional *cis*-regulatory elements possessing enhancer activity at near-base-pair resolution. The map generated by LV-MPRA guided the construction of an array of novel recombinant regulatory elements that conferred a range of transgene expression in human erythrocytes that correlated with the length of concatenated enhancer combinations, allowing us to generate, in mere weeks, a vector containing enhancers of novel composition that rivaled the performance of those in a clinical vector that have undergone nearly two decades of refinement.<sup>13,46</sup>

Finally, we demonstrated that novel therapeutic  $\beta$ -globin LV constructs generated using LV-MPRA data, selected for optimal size and expression, faithfully corrected the disease phenotype in a mouse model of sickle cell disease (SCD), demonstrating the promise of this approach in accelerating development of clinically relevant and novel vectors.

## RESULTS

### LV-MPRA Identifies Sequence-Intrinsic Enhancers of the $\beta$ -Globin Locus Control Region

To map enhancers throughout the LCR, we designed a reporter vector that contained only the minimal human  $\beta$ -globin promoter as the

major *cis*-regulatory element driving expression of the  $\beta^{\text{AS3}}$ -P2A-mCitrine (mCit) transgene.<sup>47,48</sup> We then incorporated a 103-bp sequence derived from the LCR's HS2 core region into the multiple cloning site of LV-reporter (Figure S1A), as HS2 has been shown to possess robust erythroid-specific enhancer activity throughout all stages of human development.<sup>28</sup> The 103-bp HS2 enhancer sequence provided increased expression of the reporter gene in human umbilical cord blood-derived erythroid progenitor clone 2 (HUDEP-2)<sup>49</sup> cells when compared to control (Figures S1B and S1C). This finding suggested that  $\sim$ 100-bp regions of the LCR could enhance expression from the minimal human  $\beta$ -globin promoter. Thus, LV-reporter became the basis of what was used to analyze thousands of overlapping  $\sim$ 100-bp LCR sequences in parallel for enhancer activity.

The LV-MPRA experimental strategy consisted of three steps: (1) *in silico* bar-coded LCR segment library design and construction, (2) library packaging, cell line transduction, and culture, and (3) barcode acquisition, sequencing, and data analysis.

First, a library of overlapping 103-bp LCR “query” sequences was designed by *in silico* tiling. Overlapping sequences were generated with the start of subsequent sequences beginning 4 bp after the start of a preceding sequence. The collection of sequences was duplicated three times by associating three unique 13-bp barcodes with each sequence. The resultant sequence collection was duplicated a second time by replacing all sequences with reverse complement sequences and assigning new barcodes. In total, a collection of  $\sim$ 25,000 unique LCR/barcode sequence combinations were generated with a maximum of 150 $\times$  coverage per bp (Figure 1A).

The library of LCR sequence and barcode pairs were synthesized via DNA microarray as 170-bp oligonucleotides. A schematic of a single 170-mer is shown in Figure 1B. The length of any given query sequence was fixed at 103 bp, as the DNA microarray was limited to synthesizing 170 mer, and 67 bp of sequence was needed for backend library construction/sequencing. A cloning strategy was deployed that allowed for placement, en masse, of query sequences upstream of the minimal  $\beta$ -globin promoter, as well as placement of corresponding barcodes between the  $\beta^{\text{AS3}}$ -P2A-mCit sequence and 3' UTR of the transcriptional cassette. Thus, the strength of a

### Figure 1. Overview of LV-MPRA Library Design and Experimental Workflow

(A) Overlapping 103-bp  $\beta$ -globin locus control region (LCR) sequences were generated with 4-bp tiling. The start of subsequent sequences began 4 bp after the start of a preceding sequence. Three unique barcodes (BCs) were assigned to each sequence, and the entire sequence collection was duplicated in reverse orientation and assigned new BCs. A total of  $\sim 4.2 \times 10^3$  unique oligonucleotides were needed to achieve 1 $\times$  coverage of the larger 16-kb LCR sequence. Each “query” sequence was assigned three unique 13-bp barcodes, tripling the diversity of sequences to  $\sim 1.2 \times 10^4$ . Antisense versions of the query sequences were also included, doubling the total number of unique sequences to  $\sim 2.5 \times 10^4$ . (B) A schematic of a single 170-mer is provided. (1) The 170-mer is flanked by 20-bp arms at each end, with each possessing homology to the plasmid backbone and required for downstream cloning. The 103-bp LCR (query) sequence and 13-bp barcode are separated by *BmtI* and *SalI* restriction sites to facilitate downstream cloning. (2) A pool of 170 mer generated by DNA microarray were converted into dsDNA by primer extension and then PCR amplified. (3) The pool of 170-bp dsDNA fragments were joined to lentiviral vector plasmid backbones by Gibson assembly to ensure 1:1 fusions of 170-bp dsDNA fragments to plasmid. The library was then digested with restriction enzymes and an expression cassette ( $\beta$ -globin promoter/ $\beta^{\text{AS3}}$ -globin gene/p2A/mCitrine) introduced by ligation. (4) The complete library provides placement of the query sequence upstream of the promoter and placement of the BC upstream of a polyadenylation signal to allow for determination of query sequence strength by BC abundance. (C) The complete LV reporter library was packaged into lentiviral particles and virus used to transduce the erythroid progenitor-like cell line, HUDEP-2. The abundance of cDNA BCs were quantified and normalized to the abundance of plasmid DNA BCs to detect and measure the strength of sequence-intrinsic enhancers.

query sequence could be quantified by the abundance of barcodes expressed in mRNA. The transcriptional cassette was kept in reverse orientation in respect to viral RNA production to retain introns during viral genome packaging. Completeness of the starting plasmid library was confirmed by sequencing (Figure S2A).

Second, the plasmid library was packaged into viral particles and used to transduce the HUDEP-2 cell line at a multiplicity of infection (MOI) of 4. To verify that packaging and cell line transduction did not negatively influence library complexity, we quantified barcode abundance in the starting plasmid pool and in genomic DNA (gDNA) of transduced HUDEP-2 cells. The correlation between barcodes in the plasmid pool and gDNA of transduced HUDEP-2 cells was strong ( $r = 0.854$ ), demonstrating that the diversity of barcodes in the starting plasmid pool was efficiently transferred to the integrated proviral barcode pool of transduced HUDEP-2 cells (Figure S2B; each dot in the dot plot represents the  $\log_{10}$  value of an individual barcode found in the plasmid barcode pool, integrated proviral barcode pool, or both). Vector copy number (VCN) in the bulk transduced HUDEP-2 cell population was found to be 9.7 by digital droplet (DD) PCR.

Lastly, barcodes were obtained from mRNA of transduced HUDEP-2 cells and sequenced after cells were differentiated down the erythroid lineage for 4 days. The correlation between mRNA and gDNA barcodes was low ( $r = 0.304$ ), demonstrating that only a subset of query sequences possessed enhancer activity (Figure S2C). To generate a map of enhancer activity across the LCR, barcode reads were normalized by sequencing depth and enhancer activity calculated by dividing RNA barcode counts by plasmid DNA barcode counts. Plasmid barcode counts were chosen as the normalization factor, as we were only interested in identifying enhancer sequences that were successfully packaged to enhance expression once viral genomes were stably integrated into HUDEP-2 cells.

Expression values were generated for each starting position by averaging normalized counts of three corresponding barcodes per position. Averaged normalized counts were then plotted to their corresponding positions across a map of the LCR. Statistical bootstrapping was then used to determine an “enhancing score” for a given query sequence by resampling the neighboring 50 values 100,000 times and then replacing that value with the estimated value in an iterative fashion. The bootstrapping process allowed for reduction of noise, as clusters of neighboring sequences offered similar levels of expression activity due to the narrow sequence length assayed.

The resulting map of enhancer activity across the LCR is provided in Figure 2A. Highlighted regions represent boundaries of LCR regulatory elements characterized in transgenic mice as erythroid cell-specific DNase I hypersensitive sites required for high-level, position-independent expression.<sup>20–28</sup> As shown, most peaks within the LCR that represent sequences with strong intrinsic enhancer activity fall within the boundaries of classically defined LCR control elements. The same map is shown in Figure S2D, this time aligned

to ENCODE track sets denoting regions of open chromatin and other markings in K562 cells <http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&position=chr11:5296481-5313207> (GRCh37/hg19; University of California Santa Cruz [UCSC] Genome Browser), markings that when present in specific combinations suggest the presence of enhancers.<sup>50</sup> Furthermore, when components of composite enhancers were aligned with the endogenous human  $\beta$ -globin LCR, those components fell well within the boundaries of Lenti/ $\beta^{\text{AS3}}$ -FB's full-length HS sequences (Figure S3).

To determine whether the presence of specific transcription factor binding sites (TFBSs) were predictive of intrinsic enhancer strength, the presence of key putative erythroid-specific TFBSs (GATA1, KLF1, and TAL1) were identified and aligned to the map of normalized barcode counts (Figure 2B). The presence of putative erythroid-specific TFBSs were seen to align closely with intrinsic enhancer peaks above the 95th percentile of highest expression. These findings are consistent with previous reports that highlight the role of GATA1, KLF1, and TAL1 in modulating expression of erythroid-specific genes.<sup>51,52</sup>

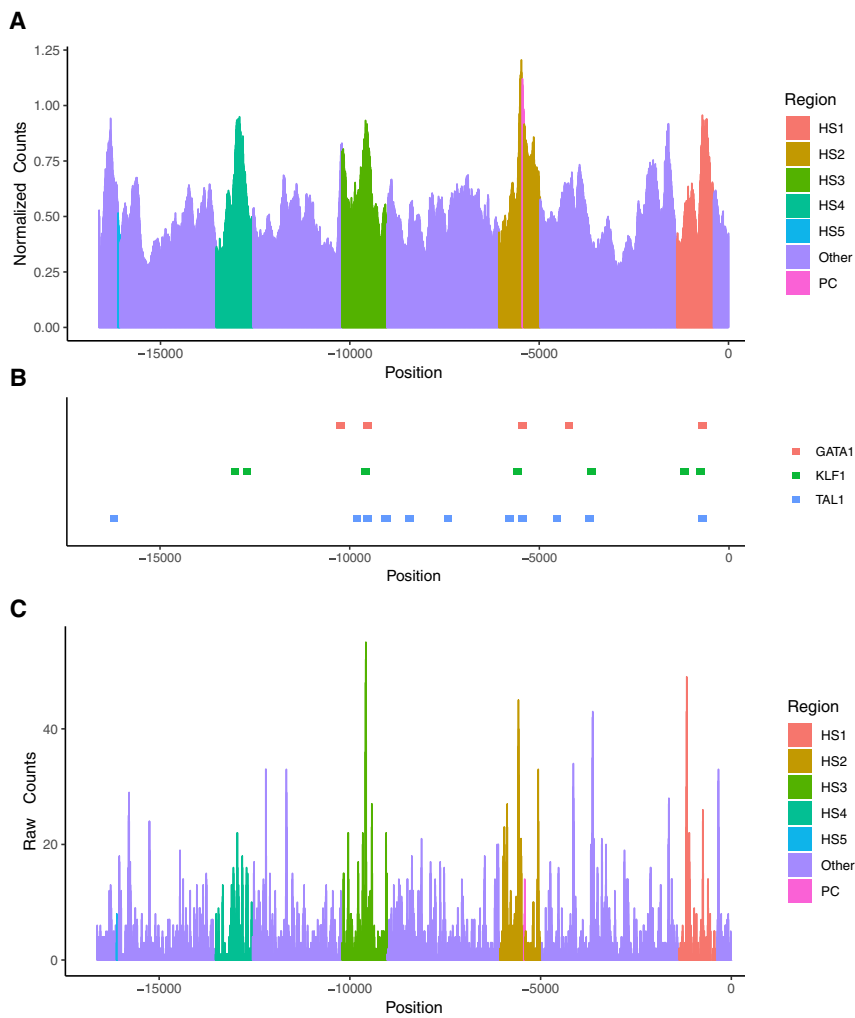
Densities of putative TFBSs were also found to be predictive of enhancer strength. Putative TFBSs were identified along the LCR using FIMO (Find Individual Motif Occurrences),<sup>53</sup> an algorithm that scans a given DNA region for individual TFBS consensus sequence matches. Figure 2C provides a density map of FIMO TFBS matches plotted by location across the LCR. Locations with quantities of putative TFBSs greater than 40 were seen to align closely with sequence-intrinsic enhancer peaks that were above the 95th percentile of highest expression.

#### LV-MPRA-Guided Therapeutic Vector Design and Characterization

A series of  $\beta$ -globin expression constructs were designed by imposing thresholds to cluster LCR sequences that were within the 80th, 90th, 95th, 97.5th, or 98.75th percentiles of highest enhancer activity (Figure 3A) (see Figure S3 for alignment of novel LCR enhancers to the entire  $\beta$ -globin LCR and to the LCR element contained within the Lenti/ $\beta^{\text{AS3}}$ -FB vector). Concatenated enhancer combinations were each cloned in sense orientation into the identical  $\beta$ -globin expression vector backbone and compared head-to-head against Lenti/ $\beta^{\text{AS3}}$ -FB, which was deemed to be the strongest expressing globin vector by way of its full-length HS2, HS3, and HS4 elements (totaling  $\sim 3.6$  kb in sequence length).

The LV-MPRA-guided constructs, termed 80, 90, 95, 97.5, and 98.75 (based on the percentile cutoff thresholds used to identify sequences for concatenation within a given percentile of expression), were packaged in parallel against Lenti/ $\beta^{\text{AS3}}$ -FB using HEK293T cells and titered head-to-head on HT-29 cells. An inverse correlation was observed when titer was plotted as a function of enhancer length, demonstrating that overall enhancer length strongly influenced infectious particle production (Figure 3B).

To measure the infectivity of the different vectors, human CD34<sup>+</sup> hematopoietic stem and progenitor cells (HSPCs) were transduced



**Figure 2. Sequence-Intrinsic Enhancer Map Spanning the  $\beta$ -Globin Locus Control Region**

(A) A map of enhancer activity across the LCR is provided. Highlighted regions represent boundaries of LCR-hypersensitive sites (HS1–HS5) as defined in previous literature. PC, positive control with known enhancer sequence from HS2. Track sets available through ENCODE denoting regions of open chromatin and other markings are provided as a reference. (B) A map of important erythroid-specific transcription factor binding sites. (C) A map of putative transcription factor binding sites across the LCR. Highlighted regions represent boundaries of LCR-hypersensitive sites as defined in previous literature.

We then sought to examine the relationship between enhancer length and expression per vector genome. Thus, primary human CD34<sup>+</sup> HSPCs were transduced at various MOIs and cultured under erythroid differentiation conditions for 14 days. Expression by each vector was measured as percentages of  $\beta^{\text{AS3}}$ -globin transcripts to total  $\beta$ -globin-like transcripts, normalized to VCN, and plotted by function of enhancer length. Normalized expression levels were seen to correlate strongly with enhancer length (Figure 3E).

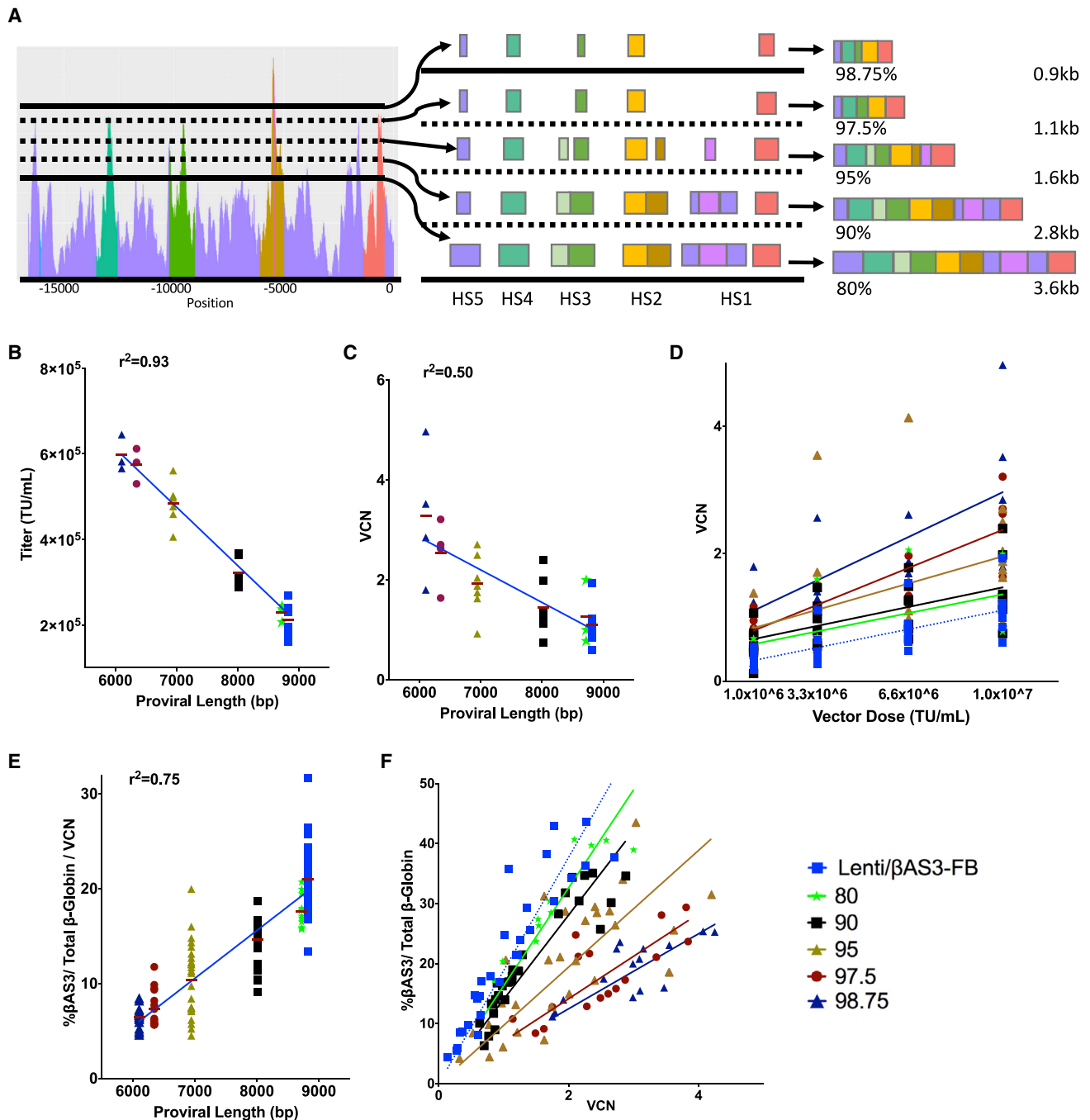
Comparing % $\beta^{\text{AS3}}$ -globin transcript levels and VCN, each vector had a characteristic linear relationship (Figure 3F). The larger vectors achieved higher expression per VCN, but the smaller vectors were able to reach a higher VCN and compensate for lower expression per VCN to yield total amounts of expression that were similar. For example, Lenti/ $\beta^{\text{AS3}}$ -FB was seen to provide slightly higher levels of expression than

at a fixed MOI of 10 ( $1.0 \times 10^7$  TU/mL; based on titers measured on HT-29 cells). Transduced CD34<sup>+</sup> cells were cultured for 14 days followed by extraction of gDNA for measurement of VCN. An inverse relationship was observed between enhancer length and VCN, which reflects relative infectivity of the different vectors when used at the same MOI (Figure 3C; Table 1).

Transduction efficiency was then assessed across a range of MOIs (1, 3.3, 6.6, or 10). Linear regression analysis revealed a positive correlation between infectivity and vector dose while demonstrating an inverse correlation between infectivity and enhancer length (Figure 3D). Indeed, the observation that overall vector length is the strongest factor influencing both packaging and transduction efficiency, rather than the presence of discrete sequences, is well supported.<sup>54–56</sup> However, the role sequence plays in packaging efficiency cannot be overlooked. For example, enhancer elements containing cryptic polyadenylation sequences can prematurely terminate viral genome synthesis, leading to significant decreases in titers.

did the similarly sized 80 vector (as shown by the increased slope of the fitted line), but this difference failed to reach significance. All LV-MPRA vectors displayed erythroid-specific expression patterns, as expression was non-detectable in transduced HSPCs cultured under myeloid differentiation conditions (Table 1).

We then sought to determine optimal enhancer orientation and arrangement. Sequences within the 95th percentile of highest expression were arranged in sense orientation with those associated with HS1 placed closest to the promoter (95-Sense). Enhancers were also arranged in an anti-sense orientation, with those elements associated with HS1 placed farthest from the promoter (95-Antisense). Enhancers were also arranged in anti-sense orientation with those associated with HS1 now placed closest to the promoter (95-Alt Antisense). A schematic outlining enhancer orientation and arrangement is provided in Figure S4A. While sequences containing cryptic polyadenylation sites should have been depleted during viral genome synthesis and thus eliminated from the initial screen, any cryptic polyadenylation sites present within incorporated enhancer



**Figure 3. LV-MPRA-Guided Therapeutic Vector Design and Characterization**

(A) Percentile cutoff thresholds were established to identify sequences within a given percentile of expression. Those sequences were then concatenated to produce composite enhancer elements. (B) Composite enhancers were cloned into the plasmid backbone of a therapeutic lentiviral vector (Lenti/ $\beta$ <sup>AS3</sup>-FB with LCR DNase-hyper-sensitive sites (HSs) 2, 3, and 4 removed, packaged, and titered head-to-head, and the quantity of infectious particles were plotted as a function of proviral length (bp). Each point in the plot represents an individual 10-cm plate of virus titered on HT-29 cells. Proviral length is defined as sequence length from the beginning of the 5' long terminal repeat (LTR) U3 through the end of the 3' LTR U5.  $n = 3$ –9 per arm. (C) Human CD34<sup>+</sup> hematopoietic stem and progenitor cells (HSPCs) were transduced with constructs at a multiplicity of infection (MOI) of 10 ( $1 \times 10^7$  TU/mL) and cultured under myeloid culture conditions to assess infectivity. Vector copy number (VCN) was determined by droplet digital polymerase chain reaction (ddPCR) 14 days after transduction. Each point in the plot represents an individual transduction, and the VCN of each transduction is plotted by function of proviral length.  $n = 4$  per arm. (D) Human CD34<sup>+</sup> HSPCs were transduced at MOIs of 1, 3.3, 6.6, or 10 and cultured under myeloid culture conditions to assess

(legend continued on next page)

sequences would be eliminated when enhancers were placed in anti-sense or alt-antisense orientation.

The collections of enhancer configurations were cloned into the identical  $\beta$ -globin expression vector backbone and packaged and titered in parallel. No significant differences in titer were observed (Figure S4B). Human CD34<sup>+</sup> HSPCs were transduced with constructs at  $1 \times 10^7$  TU/mL and cultured for 14 days. No differences in infectivity (Figure S4C), expression (Figure S4D), or normalized expression levels per VCN were observed, demonstrating that element directionality or element proximity to the promoter did not diminish vector production, transduction, or expression (Figure S4E).

Lastly, we sought to confirm that sequences within the 95th percentile of lowest expression were incapable of enhancing expression above basal levels offered by the minimal promoter. Sequences within the 95th percentile of lowest expression were identified, concatenated, and cloned into a  $\beta$ -globin expression vector backbone to generate 95-Negative. This new construct was then compared against a  $\beta$ -globin expression vector containing only a minimal  $\beta$ -globin promoter as the major regulatory element driving expression (Pro.Only) and the previously defined (above) 95-Sense construct. Human CD34<sup>+</sup> HSPCs were transduced at an MOI of 10 and cultured under erythroid conditions for 14 days. The previously defined 95-Sense construct offered significantly higher levels of normalized expression when compared head-to-head against other constructs. As expected, no significant differences in normalized expression between 95-Negative and Pro.Only were observed, demonstrating that sequences within the 95th percentile of lowest expression were incapable of enhancing expression while sequence within the 95th percentile of highest expression could clearly enhance expression (Figure S5D).

Taken together, these data demonstrate that *cis*-regulatory elements identified using LV-MPRA fulfill the major tenet that defines enhancers, which is the ability to increase transcription from a promoter independently of direction or juxtaposition.<sup>57,58</sup> These data also demonstrate that LV-MPRA can be used to concatenate regulatory elements of various lengths that (in this case) offer predictable patterns of performance across multiple categories (additional data are provided in Figure S6).

#### **In Vivo Characterization of LV-MPRA-Based Therapeutic Vectors in the SCD Mouse Model**

The best performing LV-MPRA constructs, 95 and 97.5, which were intermediate for length, titer, and expression, were then compared to Lenti/ $\beta^{\text{AS3}}$ -FB in the “Townes” mouse model of SCD to evaluate their ability to induce hematologic correction. Lineage-depleted bone marrow (BM) cells were obtained from homozygous  $\beta^{\text{S}}/\beta^{\text{S}}$

donor mice and pre-stimulated for 1 day. Cells were transduced at equal MOIs by the different vectors and after 1 day delivered by retro-orbital injection into lethally irradiated GFP-transgenic mouse recipients (B6-GFP; Jackson Laboratory). Three independent experiments were conducted, and *in vitro* VCN was determined from the transduced cell product 14 days after transduction (Table 2). Peripheral blood (PB) samples, acquired at 4 and 16 weeks post-transplantation, were assessed for engraftment by flow cytometry for GFP-negative donor cells, gene marking in circulating cells (VCN by droplet digital PCR [ddPCR]), and blood hemoglobin (Hb) concentration and composition by high-performance liquid chromatography (HPLC).

Mice with BM donor engraftment <97% at week 16 were excluded from analyses, as  $\geq 4\%$  residual wild-type (WT) recipient RBCs could mask adverse pathophysiology induced by  $\beta^{\text{S}}/\beta^{\text{S}}$  donor cells.<sup>59</sup> Week 16 engraftment efficiency in PB or BM (by fluorescence-activated cell sorting [FACS] or HPLC) were not different among experimental arms (Figures S7A–S7C). Average gene transfer efficiency seen in circulating PB (Figure 4A) and BM cells (Figure S7D) differed significantly among experimental arms, and it demonstrated that constructs with decreased total enhancer lengths offered superior transduction efficiency.

Quantification of Hb $\beta^{\text{AS3}}$  tetramers in PB lysates was accomplished using HPLC. To compare differences in normalized expression between each experimental arm, the %Hb $\beta^{\text{AS3}}$ /total Hb tetramers was normalized to PB VCN for each mouse and plotted. Constructs with larger enhancers offered superior expression per vector genome, with overall trends reflecting those seen in *in vitro* experiments, where constructs with larger enhancers offered superior expression per vector genome (Figure 4C).

Surprisingly, average total levels of Hb $\beta^{\text{AS3}}$ /total Hb tetramers were not different among experimental arms, demonstrating that lower expression levels from smaller vectors were compensated for by gains in transduction efficiency (Figure 4B). Expression findings were also confirmed by measuring % $\beta^{\text{AS3}}$ -globin mRNA to  $\beta$ -globin mRNA in BM obtained 16 weeks post-transplantation and then normalizing those values to BM VCN (Figures S6E and S6F).

Finally, hematologic indices were measured using PB obtained at 4 and 16 weeks post transplantation. When compared to mice transplanted with mock-transduced BM cells, Hb levels, red blood cell counts, and hematocrits were significantly higher in mice that were transplanted with BM transduced with vectors designed using LV-MPRA.

---

vector infectivity. The VCN of each transduction is plotted by function of vector dose. Slopes represent linear regressions.  $n = 12$ – $28$  per arm. (E) Human CD34<sup>+</sup> HSPCs were transduced at MOIs of 1, 3.3, 6.6, or 10 and differentiated under erythroid culture conditions. Percentages of  $\beta^{\text{AS3}}$ -globin RNA to total  $\beta$ -globin-like RNA were determined by reverse transcription (RT) ddPCR and normalized to VCN. Normalized expression values are plotted as a function of proviral length.  $n = 12$ – $28$  per arm. (F) Human CD34<sup>+</sup> HSPCs were transduced at MOIs of 1, 3.3, 6.6, or 10 and differentiated under erythroid culture conditions. Percentages of  $\beta^{\text{AS3}}$ -globin RNA to total  $\beta$ -globin RNA are plotted by function of their corresponding VCNs.  $n = 12$ – $28$  per arm.



**Table 1. Comparison of Normalized Expression between Culture Conditions**

Vector	Erythroid VCN	Erythroid Expression (% $\beta^{AS3}$ /VCN)	Myeloid VCN	Myeloid Expression (% $\beta^{AS3}$ /VCN)
98.75	3.4 ( $\pm 0.5$ )	6.8 ( $\pm 1.3$ )	3.3 ( $\pm 1.3$ )	not detectable
97.5	3.1 ( $\pm 0.8$ )	7.9 ( $\pm 2.7$ )	2.6 ( $\pm 0.7$ )	not detectable
95	2.1 ( $\pm 0.4$ )	10.5 ( $\pm 3.3$ )	1.9 ( $\pm 0.7$ )	not detectable
90	1.1 ( $\pm 0.1$ )	16.1 ( $\pm 0.7$ )	1.5 (0.7)	not detectable
Lenti/ $\beta^{AS3}$ -FB	1.1 ( $\pm 0.2$ )	19.9 ( $\pm 2.4$ )	1.0 ( $\pm 0.3$ )	not detectable

At week 16, Hb levels of mice that received mock-transduced cells were 7.9 g/dL on average while the Hb levels of mice that received BM transduced with Lenti/ $\beta^{AS3}$ -FB, 95, or 97.5 were 11.0, 11.2, and 11.6 g/dL on average, respectively (Figure 4D). RBC counts were also significantly higher for recipients of Lenti/ $\beta^{AS3}$ -FB-, 95-, or 97.5-transduced BM cells ( $8.5 \times 10^6$ ,  $9.0 \times 10^6$ , and  $8.5 \times 10^6$  cells/ $\mu$ L on average, respectively) compared to recipients of mock-transduced BM cells ( $6.2 \times 10^6$  cells/ $\mu$ L on average). Similar improvements were seen for hematocrits, where mice that received mock-transduced cells had hematocrits of 25.2 on average, while mice that received Lenti/ $\beta^{AS3}$ -FB-, 95-, or 97.5-transduced BM had hematocrits of 30.9, 32.2, and 30.6 on average, respectively. The *in vivo* data demonstrate that when LV-MPRA was deployed to aid therapeutic vector development, identification of correct combinations of enhancers capable of providing sufficient levels of transgene expression was achieved.

## DISCUSSION

Control over LV transgene expression is typically achieved by including *cis*-regulatory elements that enhance transcription of a LV's internal promoter in a cell type-specific pattern.<sup>60</sup> While genome-scale cell line-based genetic and epigenetic studies have generated an impressive collection of candidate enhancers, these studies utilize indirect measurements (histone modification, chromatin accessibility, and bound transcriptional co-activators) to predict enhancer locations.<sup>61</sup> Moreover, the limitations of these studies fail to provide exact boundaries of sequence-intrinsic enhancers (i.e., the actual sequences that provide enhancer activity). To date, the vast majority of putative *cis*-regulatory elements have yet to be characterized in LV systems.

Therapeutic expression vectors are traditionally designed by testing individual candidate enhancers for their ability to potently drive cell type-specific transgene expression, in addition to determining how incorporation influences titer and infectivity. While criteria can be imposed to reduce the number of candidate enhancer elements for evaluation, only a handful of elements can be tested at a time, as current throughput limitations restrict the number of LVs that can be evaluated in parallel. Confounding the issue of low throughput,

**Table 2. *In Vitro* VCN of Gene-Modified  $\beta$ S/ $\beta$ S Lin<sup>-</sup> BM Cells before Transplant**

	Lenti/ $\beta^{AS3}$ -FB	95	97.5
Transduction condition (TU/mL)	$6.6 \times 10^6$	$6.6 \times 10^6$	$6.6 \times 10^6$
<i>In vitro</i> VCN experiment 1	1.6	ND	7.2
<i>In vitro</i> VCN experiment 2	1.5	2.7	4.9
<i>In vitro</i> VCN experiment 3	1.2	3.0	5.7

TU, transduction units; VCN, vector copy number; ND, not done.

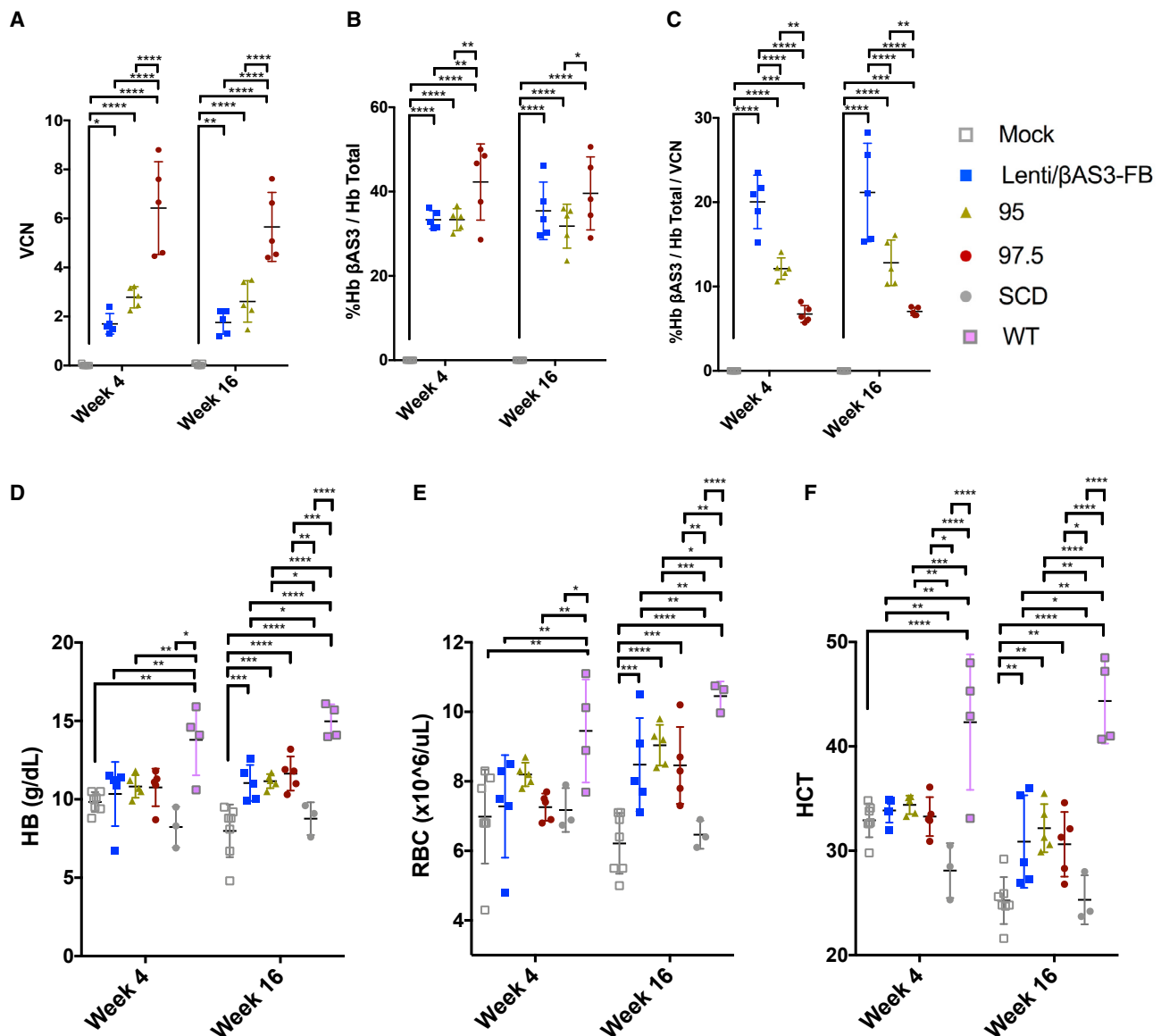
specific combinations of elements must also be uncovered to achieve temporal and spatial control over transgene output. Thus, the challenges in developing expression vectors produced at high titers, while offering both robust infectivity and appropriate levels of cell type-specific protein expression, are elevated.

Multiple groups have used high-throughput approaches toward dissecting the functional activities of *cis*-regulatory elements in various contexts,<sup>30–39</sup> however, no group has applied their findings toward therapeutic vector design. While each group developed an assay that possessed aspects of an ideal functional assay for therapeutic vector development, each approach had its own unique advantages and limitations. The ideal functional assay needed was one that can unbiasedly screen thousands of DNA sequences derived from large genomic regions for intrinsic enhancer activity and do so in the context of a therapeutic vector integrated into a relevant cell type.

To this end, we developed an LV-MPRA to generate a continuous and quantitative map on which to dissect the ability of putative *cis*-regulatory sequences to enhance expression in an erythroid progenitor cell line. By using DNA oligonucleotide synthesis to create the starting material required to create thousands of barcoded LVs with each harboring a known LCR fragment, we were able to transcend the traditional limitations of therapeutic vector development by using barcode sequencing to digitally measure enhancer activity across thousands of sequences.

While results of our initial pilot study (provided in Supplemental Information) demonstrated that a construct harboring the minimal  $\beta$ -globin promoter alone generated detectable levels of basal activity, the construct harboring an HS2 fragment provided ~5-fold higher levels of activity, suggesting that other sequences possessing comparable intrinsic enhancer activity could be detected. The use of 103-bp insert size influenced the dynamic range in the assay by decoupling the influence of flanking DNA on expression offered by core sequences, which is consistent with previous findings demonstrating the weak role of even larger core fragments (200–400 bp) on enhancing expression when compared to appropriate controls.<sup>42</sup>

Averaging expression between three barcodes was sufficient to generate a map of expression activity once statistical bootstrapping was implemented. The resultant map possessed peaks that aligned with previously published chromatin



immunoprecipitation sequencing (ChIP-seq) data denoting putative positions of enhancer elements. Notably, the same HS2 sequence tested in the pilot experiment exhibited the highest levels of activity, while HS5, a sequence demonstrated in previous studies to be a potent insulator element,<sup>23</sup> was shown to lack any enhancing activity in our analysis. Moreover, the consistent level of noise across the map (ranging from  $\sim 0.3$  to  $\sim 0.5$ ) was consistent with levels of activity offered by the minimal promoter alone, as demonstrated in the pilot study.

When the percentile cutoff of 80% was imposed to identify sequences within the 80th percentile of highest expression, we were able to develop a novel regulatory sequence that when incorporated into a therapeutic expression vector backbone produced a construct, termed 80, with no significant differences in titer, infectivity, or expression when compared to Lenti/ $\beta^{AS3}$ -FB. Given that boundaries defining the enhancer elements of Lenti/ $\beta^{AS3}$ -FB have been undergoing continuous refinement for the past 20 years,<sup>14</sup> we found it noteworthy that our approach was

able to create a comparable regulatory sequence of novel composition in a matter of weeks.

Since the 80 vector was so analogous to Lenti/ $\beta^{\text{AS3}}$ -FB (in relationship to robust expression but diminished titer and gene transfer), we decided to test the smaller 95 and 97.5 in the Townes mouse model of SCD. Although 95 and 97.5 were found to possess 50% and 30% of the expression activity of Lenti/ $\beta^{\text{AS3}}$ -FB per vector genome, respectively, these losses in expression were overcome by gains in infectivity, resulting in production of comparable total amounts of  $\beta^{\text{AS3}}$ -mRNA and protein in circulating murine RBCs, thus reversing the disease phenotype. While an increase in VCN to compensate for expression is typically undesirable, gene transfer is typically less than one copy per cell in the case of traditional  $\beta$ -globin-expressing LVs. Thus, an increase to a VCN of 2–4 is not excessive, as seen with the 95 vector. More importantly, 95 and 97.5 were found to be produced at significantly higher titers when compared to Lenti/ $\beta^{\text{AS3}}$ -FB, which may hold implications for improving access to gene therapy by significantly reducing the cost of vector production through providing higher levels of gene transfer using the smallest amounts of vector. When combined with transduction enhancers,<sup>62</sup> the use of 95 and 97.5 would lead to even further advantages in drug product quality and cost reduction per patient.

This proof-of-principle study demonstrates that in mere weeks, lineage-specific therapeutic expression vectors were generated that rivalled a vector that has undergone years of refinement. Designing LVs using sequence-intrinsic enhancers of the LCR was done not just because the locus is such a hot gene therapy target, but also because fine mapping represents a high bar to clear since the locus has been studied in such detail for decades. The real power of this LV-MPRA approach should come into play when far less studied loci are targeted for gene therapy vector design and coupled with state-of-the-art machine learning algorithms.<sup>63</sup> The approach outlined herein facilitates the rapid development of any therapeutic vector requiring cell type-specific expression while advancing the prospect of realizing gene therapy's promise.

As an alternative to gene addition using LVs, a variety of approaches are being developed for genome editing using CRISPR/Cas9 to treat SCD. Disruption of the key repressor of fetal globin expression, *BCL11A*, or its binding site near the  $\gamma$ -globin gene or even site-specific correction of the A→T transversion underlying SCD are in early clinical or pre-clinical stages. It will remain to be determined which gene therapy approaches ultimately provides the best efficacy, safety, and cost-effectiveness. However, LVs have an established track record of clinical efficacy and safety and thus continue to be worthy of efforts for optimization.

## MATERIALS AND METHODS

### LV-MPRA Oligonucleotide Design and Preparation

Pools of synthetic 170-mer DNA sequences were ordered from CustomArray (Bothell, WA, USA). Each oligonucleotide was designed according to the following scheme: 5' primer amplification sequence (ATGTTTTTCTAGGTCTCGAG)/103-bp LCR query

sequence/Bmtf site/4-bp spacer/SalI site/barcode/3' primer amplification sequence (CTTTGTTCCCTAAGTCCAAC).

Each oligonucleotide possessed a 103-bp query sequences derived from a larger ~16-kb LCR sequence (<http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&position=chr11:5296481-5313207> GRCh37/hg19; Chr11:5,296,481-5,313,207; UCSC Genome Browser), with each subsequent oligonucleotide possessing a similar query sequence offset by 4 bp until complete coverage of the larger sequence was achieved. A total of  $\sim 4.2 \times 10^3$  unique oligonucleotides were needed to achieve  $1 \times$  coverage of the larger 16-kb LCR sequence. Each query sequence was assigned three unique 13-bp barcodes, tripling the diversity of sequences to  $\sim 1.2 \times 10^4$ . Antisense versions of the query sequences were also included, doubling the total number of unique query/barcode pairs to  $\sim 2.5 \times 10^4$ .

Primer “Custom Array Rev” (Table S1) was used to convert full-length DNA microarray synthesized oligonucleotides into double-stranded DNAs (dsDNAs) by primer extension. Twelve reactions using MyTaqRed (Bioline Meridian Life Science, Memphis, TN, USA) were established in parallel, with each reaction containing 200 ng of template and 4  $\mu$ L of reverse primer. A single cycle of PCR amplification was performed at an annealing temperature of 45°C. Reaction products were purified using a PureLink PCR purification kit (Thermo Fisher Scientific, Waltham, MA, USA) to deplete residual primers and single-stranded DNA (ssDNA) products and pooled.

Primers “CustomArray FWD” and “CustomArray REV” were then used to amplify 170-bp dsDNA fragments. Twelve PCR reactions were performed using MyTaqRed (Bioline Meridian Life Science, Memphis, TN, USA) with 200 ng of dsDNA template and amplified for six cycles with an annealing temperature of 45°C. Reaction products were purified by column and pooled. The purified reaction products were then used in LV library construction.

### LV-MPRA Library Construction

We created a streamlined version of a  $\beta$ -globin expression vector for use in LV-MPRA library construction. The modified construct was made by replacing the reverse-oriented human genomic elements (full-length LCR HS2 and HS3 enhancer elements, minimal  $\beta$ -globin gene promoter,  $\beta^{\text{AS3}}$  transgene with introns, exons, and 3'UTR) of Globe1-AS3-FB with a minimal  $\beta$ -globin gene 3' UTR sequence and an EcoRV site. The plasmid backbone was linearized using EcoRV enzyme in buffer 3.1 (New England Biolabs, Ipswich, MA, USA) and 5' DNA ends dephosphorylated using recombinant shrimp alkaline phosphatase (New England Biolabs, Ipswich, MA, USA). Reaction products were pooled and run on 1% agarose gel, and DNA of correct length was gel extracted and purified using a PureLink quick gel extraction kit (Invitrogen, Carlsbad, CA, USA). Pools of 170-bp dsDNA fragments were cloned into linearized plasmid backbones using an NEBuilder HiFi DNA assembly (New England Biolabs, Ipswich, MA, USA) mix to create the “pre-complete library” (the 5' and 3' primer amplification sequences of the 170-bp dsDNA fragments had homology with the DNA ends of the linearized plasmid

backbones). A total of six 20- $\mu$ L NEBuilder reactions were established, each containing 400 ng of purified plasmid backbone and 50 ng of purified 170-bp dsDNA fragments. Reactions were incubated at 50°C for 60 min. The NEBuilder reactions were pooled, purified using a 2.5 $\times$  Agencourt AMPure XP beads kit (Beckman Coulter, Brea, CA, USA), and eluted into a final volume of 24  $\mu$ L of DNase-free water. A total of six transformation reactions were performed with NEB Stable Competent *E. coli* (New England Biolabs, Ipswich, MA, USA) and 4  $\mu$ L of purified plasmid product. Each transformation was recovered in 500  $\mu$ L of SOC medium for 30 min, pooled, and expanded in 500 mL of Luria-Bertani (LB) medium overnight. Large-scale plasmid DNA isolation was then performed using a PureLink HiPure plasmid maxiprep kit (Invitrogen, Carlsbad, CA, USA) to create a pre-complete library.

The pre-complete library was then linearized with Sall and BmtI enzymes in buffer 3.1 (New England Biolabs, Ipswich, MA, USA). A total of 12 digestion reactions were established with each containing 2  $\mu$ g of plasmid. Digests were incubated overnight at 37°C with recombinant shrimp alkaline phosphatase. Reaction products were pooled and run on 1% agarose gel, and DNA of correct length was gel extracted and purified.

The  $\beta^{\text{AS3}}$ -P2A-mCit expression cassette was liberated from its pCR-BluntII-TOPO vector backbone using Sall and BmtI restriction enzymes in buffer 3.1 (New England Biolabs, Ipswich, MA, USA). A total of 12 digestion reactions were established, each containing 2  $\mu$ g of plasmid. Digests were incubated overnight at 37°C. Reaction products were pooled and run on 1% agarose gel, and DNA of correct length was gel extracted and purified.

The expression cassette was then ligated into the linearized pre-complete library. A total of six ligation reactions were established, with each containing 200 ng of linearized plasmid and 430 ng of insert. The ligation reactions were pooled, purified using a 2.5 $\times$  Agencourt AMPure XP beads kit (Beckman Coulter, Brea, CA, USA), and eluted into a final volume of 24  $\mu$ L. A total of six transformation reactions were performed with NEB Stable Competent *E. coli* (New England Biolabs, Ipswich, MA, USA) and 4  $\mu$ L of purified plasmid product. Each transformation was recovered in 500  $\mu$ L of SOC medium for 30 min, pooled, and expanded in 500 mL of LB medium overnight. Large-scale plasmid DNA isolation was then performed using a PureLink HiPure plasmid maxiprep kit (Invitrogen, Carlsbad, CA, USA) to create a “complete library.”

#### Vector Production and Titration

Transient transfection of 293T cells using the third-generation LV packaging system<sup>64</sup> provided packaged virus particles. Viral supernatants were then directly used for titer determination or concentrated by tangential flow filtration, as described by Cooper et al.<sup>65</sup> Briefly, the HT-29 human colorectal carcinoma cell line was transduced with different dilutions of both raw and concentrated vectors. To calculate titers, cells were harvested and VCNs were determined by ddPCR approximately 60 h post-transduction.

#### HUDEP-2 Cell Culture and Transduction

HUDEP-2 cells (provided by Dr. Y. Nakamura, RIKEN BioResource Center, Tsukuba, Ibaraki, Japan) were maintained in Iscove's modified Dulbecco's medium (IMDM; Gibco, Grand Island, NY, USA) supplemented with 1  $\mu$ M dexamethasone (Sigma-Aldrich, St. Louis, MO, USA), 1  $\mu$ g/mL doxycycline (Sigma-Aldrich, St. Louis, MO, USA), 50 ng/mL human stem cell factor (SCF), 3 U/mL erythropoietin (EPO) (all cytokines were acquired from PeproTech, Rocky Hill, NJ, USA), and 1 $\times$  glutamine, penicillin, and streptomycin (Gemini Bio-Products, Sacramento, CA, USA). A total of  $2.5 \times 10^6$  HUDEP-2 cells were transduced with 5 mL of raw virus in a total volume of 10 mL in a T75 flask. Approximately 24 h later, cells were pelleted, medium was changed, and cells were expanded for 7 days. Approximately  $1.0 \times 10^7$  transduced HUDEP-2 cells were then harvested for gDNA isolation, and the remaining cells were plated on an MS5 stromal cell layer in IMDM supplemented with 1 $\times$  glutamine, penicillin, and streptomycin, holo-human transferrin (330  $\mu$ g/mL, Sigma-Aldrich, St. Louis, MO, USA), heparin (2 IU/mL, Sigma-Aldrich, St. Louis, MO, USA), recombinant human insulin (10  $\mu$ g/mL, Sigma-Aldrich, St. Louis, MO, USA), 3 U/mL EPO, and 5% inactivated human plasma (Grifols USA, Los Angeles, CA, USA). Cells were co-cultured for 4 days, after which  $\sim 1.0 \times 10^7$  cells were harvested for RNA extraction.

#### Barcode Generation

gDNA was extracted from HUDEP-2 cells 8 days after transduction using a PureLink gDNA mini kit (Thermo Fisher Scientific, Carlsbad, CA, USA). Total RNA was extracted from HUDEP-2 cells 4 days after the start of differentiation using the RNeasy mini kit (QIAGEN, Valencia, CA, USA). Total RNA was separated into eight 10- $\mu$ L aliquots for storage. Reverse transcription of total RNA was carried out in parallel across six reactions each using 10  $\mu$ L of total RNA, 1  $\mu$ L of first-strand primer, 2  $\mu$ L of 0.1 mM DTT, 4  $\mu$ L of first-strand synthesis buffer, 1  $\mu$ L of 10 mM 2'-deoxynucleoside 5'-triphosphate (dNTP), 1  $\mu$ L of RNaseOUT, and 1  $\mu$ L of 200 U/ $\mu$ L Moloney murine leukemia virus (M-MLV) reverse transcriptase (all from Invitrogen, Carlsbad, CA, USA). The first-strand primer was designed to enrich for only cDNAs that contained barcodes and thus possessed homology to mCit to allow for creation of cDNAs comprised of a sequence spanning mCit and the 3' UTR of the barcoded  $\beta^{\text{AS3}}$ -P2A-mCit mRNAs. To amplify barcodes from both gDNA, RNA, and complete library plasmid, first-strand primer (nested) and second-strand primer were used, as they flank the barcoded region. Following PCR amplification, both the gDNA and cDNA barcodes were purified and submitted to the University of California Los Angeles (UCLA) Technology Center for Genomics & Bioinformatics Sequencing Core, Department of Pathology and Laboratory Medicine, for library construction and PEX150 Illumina sequencing. The KAPA LTP library preparation kit (catalog no. KK8232, Roche) was used for library preparations, and libraries were sequenced using a single lane of the Illumina HiSeq 3000. Reads that perfectly matched the first 14 nt of the amplicon were included in subsequent analysis. We generated 124 million reads from the complete library plasmid, 49 million reads from the gDNA, and 117 million reads from the

cDNA. Plasmid barcode reads were highly correlated with gDNA reads ( $r = 0.854$ ), allowing the plasmid barcode reads to be used for normalization.

### Barcode Quantification and Sequencing

The constant sequences before (5'-ACTTAGGGAACAAAG-3') and after (5'-GTCGACATGCTAGC-3') the barcode were used to locate barcodes from sequencing data, and the frequencies of barcodes were determined in plasmid, gDNA, and cDNA. To quantify the enhancing ability of each LCR fragment, the read counts were first normalized by sequencing depth, and cDNA barcode counts were divided by plasmid barcode counts after the sums of matched barcode triplets were calculated (each LCR fragment was associated three unique barcodes). The  $\log(x + 1)$  values were calculated and used to represent enhancing ability.

The LCR sequences tested were 103 bp in length, with each neighboring sequence beginning 4 bp away from the next. Thus, a single nucleotide should be covered by 50 different sliding fragments. Statistical bootstrapping was used to calculate the mean of the non-zero counts in the 50 covering fragments to represent the enhancing score. The bootstrap is a widely applicable and extremely powerful statistical tool that is applied in machine learning models to make predictions on data not included in training data. Bootstrapping allowed for determination of an enhancing score for a given query sequence by resampling the neighboring 50 values 100,000 times and then replacing that value with the estimated value in an iterative fashion. This process allows for reduction of noise, as a cluster of neighboring sequences should offer similar averaged levels of expression activity given the narrow sequence length being assayed. The enhancing score was then used to make the map of enhancer activity across the LCR in [Figure 2A](#).

To quantify putative TFBSs present within the LCR in the enhancer, FIMO 5.0.1<sup>53</sup> was used to infer locations of TFBSs across the query sequence. A total of 771 motifs for human transcription factors from HOCOMOCOv11 were input to FIMO for searching, and the p value threshold was set to 0.0001. The number of motifs binding to each nucleotide was calculated and plotted in [Figure 2C](#).

### BM CD34<sup>+</sup> Cell Culture and Transduction

All BM aspirates were obtained from voluntary healthy donors supplied by AllCells (Alameda, CA, USA). BM mononuclear cells were isolated by Ficoll-Hypaque density gradient centrifugation. CD34<sup>+</sup> HSPCs were enriched using a CD34<sup>+</sup> MicroBead kit (Miltenyi Biotec, Bergisch Gladbach, Germany). Enriched CD34<sup>+</sup> HSPCs were cryopreserved in fetal bovine serum supplemented with 10% dimethyl sulfoxide (Sigma-Aldrich, St. Louis, MO, USA) in liquid nitrogen. Cells were thawed and plated on non-tissue culture-treated six-well plates pre-coated with RetroNectin (20  $\mu$ g/mL, Takara Shuzo, Otsu, Japan) at  $1 \times 10^6$  cells/mL. Cells were pre-stimulated for 16–24 hours in X-VIVO 15 medium (Lonza, Basel, Switzerland) supplemented with  $1 \times$  glutamine, penicillin, and streptomycin (Gemini Bio-Products, Sacramento, CA, USA), human

SCF (50 ng/mL), human Flt-3 ligand (50 ng/mL), human thrombopoietin (50 ng/mL), and human interleukin-3 (20 ng/mL; all cytokines were acquired from PeproTech, Rocky Hill, NJ, USA). Concentrated viral supernatants were used at various MOIs to transduce CD34<sup>+</sup> HSPCs for 24 h. These cells were washed, re-plated, and cultured under myeloid or erythroid culture conditions, as described by Romero et al.<sup>66</sup> On day 14 of culture, gDNA and/or mRNA was extracted from transduced cells.

### ddPCR for VCN and % $\beta^{AS3}$ mRNA Quantification

gDNA was extracted using a PureLink gDNA mini kit (Invitrogen, Carlsbad, CA, USA). VCN was calculated by using probes SCD4 (human Syndecan 4) as a reference and HIV-1 Psi as a target. ddPCR was carried out as described in Urbinati et al.<sup>67</sup> An RNeasy Plus mini kit (QIAGEN, Valencia, CA, USA) was used for RNA extraction followed by reverse transcription as described by Urbinati et al.<sup>67</sup> Probes HBB<sup>Total</sup> as a reference and Hb $\beta^{AS3}$  as a target were used to generate droplets for ddPCR, as described by Hindson et al.<sup>68</sup> Droplets were analyzed for absolute quantification of the  $\beta^{AS3}$  gene expression normalized to the total  $\beta$ -globin gene expression.

### In Vivo Experiment in SCD Mouse Model

BM from 8- to 12-week-old homozygous  $\beta$ S/ $\beta$ S Townes mice (Jackson Laboratory stock #013071) were lineage-depleted using the lineage cell depletion kit from Miltenyi Biotec. Lineage-negative ( $\text{Lin}^-$ ) cells were pre-stimulated for 24 h in StemSpan (STEMCELL Technologies, Vancouver, BC, Canada) supplemented with murine SCF (100 ng/mL), human interleukin 11 (100 ng/mL), murine interleukin 3 (20 ng/mL), and human FLT-3 ligand (100 ng/mL). Pre-stimulated  $\text{Lin}^-$  cells were plated at  $2 \times 10^6$  cells/mL in a total volume of 2 mL in non-tissue culture-treated six-well plates that were pre-coated with recombinant retronectin. Cells were then transduced at  $1 \times 10^7$  TU/mL, with volumes used of each vector adjusted based on the titers of the vector preparations. Thus, the MOI was 5 TU/cell for all vectors, based on vector titers determined with HT29 cells.<sup>65</sup> Twenty-four hours later, 1–2 million transduced cells were delivered by retro-orbital injection after recipient mice (B6-GFP transgenic; Jackson Laboratory) were lethally irradiated (1,075 cGy, split in two fractions).

PB samples were collected at weeks 4 and 16 to measure VCN of engrafted cells by ddPCR, expression of Hb $\beta^{AS3}$  Hb by HPLC, and to determine red blood cell (RBC) indices. At week 16, mice were euthanized and BM cells were used to measure engraftment by flow cytometry (GFP<sup>+/-</sup>), VCN, and expression.

### HPLC

To characterize and quantify Hb tetramers, including human HbS and Hb $\beta^{AS3}$ , and murine HbA and HbF, 1  $\mu$ L of murine PB was lysed in 25  $\mu$ L of hemolysate and incubated at room temperature. Hemolysates were then centrifuged at  $500 \times g$  for 10 min at 4°C to remove RBC ghosts. The lysates were then stored frozen at  $-80^\circ\text{C}$  and later thawed and processed as described by Urbinati et al.<sup>67</sup>

### Statistical Analysis

All data are reported as mean  $\pm$  standard deviation of the mean unless otherwise stated. Statistical analyses were performed using GraphPad Prism version 7.0 (GraphPad, San Diego, CA, USA). The statistical significance between two averages was established using unpaired *t* tests. When the statistical significance between three or more averages was evaluated, a one-way ANOVA was applied, followed by multiple paired comparisons for normally distributed data (Tukey's test). When the normality assumption was violated, a Mann-Whitney *U* test was performed for group-wise comparison instead. Linear regression analyses were used to determine the correlation between VCN and  $\beta^{\text{AS3}}$ -globin RNA transcript quantities. All statistical tests were two-tailed, and a *p* value of  $< 0.05$  was deemed significant.

### Human and Animal Subject Oversight

The use of anonymous, commercially-purchased human CD34<sup>+</sup> cells is considered to not be human subjects research and is exempt from IRB review. The Townes sickle cell disease model mice were used under UCLA Animal Care Committee protocol ARC# 2014-025.

### Data Availability

The data and code that support the findings of this study are available from the corresponding author upon reasonable request.

### SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.omtm.2020.04.006>.

### AUTHOR CONTRIBUTIONS

R.A.M., R.P.H., and D.B.K. conceived and designed all experiments. R.A.M. and M.J.U. executed all experiments. R.A.M. and F.M. performed all bioinformatics analyses. P.G.A., D.B., C.T., L.L., B.A., R.L.W., and S.S. helped execute portions of experiments. R.A.M. and M.J.U. analyzed all data. R.A.M., R.P.H., P.G.A., R.K., and Y.N. provided research materials. R.P.H. and D.B.K. advised experiments. R.A.M. and D.B.K. provided financial and administrative support. R.A.M., R.P.H., and D.B.K. wrote the manuscript. R.A.M. and D.B.K. approved the final manuscript.

### CONFLICTS OF INTEREST

The novel locus control region elements and resulting LVs described herein are covered under a pending patent application from the University of California Board of Regents, with R.A.M., R.P.H., and D.B.K. as inventors.

### ACKNOWLEDGMENTS

This work was supported by training grants provided to R.A.M. (NIH/NHLBI F31 HL134313; PhRMA Foundation Paul Calabresi Predoctoral Fellowship 20174717; and UCLA MSTP T32 GM00804). Additional support was offered by NIH National Center for Advancing Translational Sciences (NCATS) UCLA CTSI grant number UL1TR001881. We would also like to thank the Technology Center for Genomics & Bioinformatics Sequencing Core, Department of Pathology & Laboratory Medicine, in the David Geffen School of

Medicine at UCLA for offering technical support and Dr. Aaron Cooper for his generous consultation on the project. The Flow Cytometry Core of the UCLA Broad Stem Cell Research Center was an essential resource for these studies. Please contact Dr. Donald B. Kohn ([dkohn1@mednet.ucla](mailto:dkohn1@mednet.ucla)) to request access to any reagents developed under this project by a material transfer agreement.

### REFERENCES

- Hacein-Bey-Abina, S., Hauer, J., Lim, A., Picard, C., Wang, G.P., Berry, C.C., Martinache, C., Rieux-Laucat, F., Latour, S., Belohradsky, B.H., et al. (2010). Efficacy of gene therapy for X-linked severe combined immunodeficiency. *N. Engl. J. Med.* 363, 355–364.
- Aiuti, A., Biasco, L., Scaramuzza, S., Ferrua, F., Cicalese, M.P., Baricordi, C., Dionisio, F., Calabria, A., Giannelli, S., Castiello, M.C., et al. (2013). Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott-Aldrich syndrome. *Science* 341, 1233151.
- Hacein-Bey Abina, S., Gaspar, H.B., Blondeau, J., Caccavelli, L., Charrier, S., Buckland, K., Picard, C., Six, E., Himoudi, N., Gilmour, K., et al. (2015). Outcomes following gene therapy in patients with severe Wiskott-Aldrich syndrome. *JAMA* 313, 1550–1563.
- Thompson, A.A., Walters, M.C., Kwiatkowski, J., Rasko, J.E.J., Ribeil, J.-A., Hongeng, S., Magrin, E., Schiller, G.J., Payen, E., Semeraro, M., et al. (2018). Gene therapy in patients with transfusion-dependent  $\beta$ -thalassemia. *N. Engl. J. Med.* 378, 1479–1493.
- Ribeil, J.A., Hacein-Bey-Abina, S., Payen, E., Magnani, A., Semeraro, M., Magrin, E., Caccavelli, L., Neven, B., Bourget, P., El Nemer, W., et al. (2017). Gene therapy in a patient with sickle cell disease. *N. Engl. J. Med.* 376, 848–855.
- Biffi, A., Montini, E., Lorioli, L., Cesani, M., Fumagalli, F., Plati, T., Baldoli, C., Martino, S., Calabria, A., Canale, S., et al. (2013). Lentiviral hematopoietic stem cell gene therapy benefits metachromatic leukodystrophy. *Science* 341, 1233158.
- Sessa, M., Lorioli, L., Fumagalli, F., Acquati, S., Redaelli, D., Baldoli, C., Canale, S., Lopez, I.D., Morena, F., Calabria, A., et al. (2016). Lentiviral haemopoietic stem-cell gene therapy in early-onset metachromatic leukodystrophy: an ad-hoc analysis of a non-randomised, open-label, phase 1/2 trial. *Lancet* 388, 476–487.
- Aiuti, A., Cattaneo, F., Galimberti, S., Benninghoff, U., Cassani, B., Callegaro, L., Scaramuzza, S., Andolfi, G., Mirolo, M., Brigida, I., et al. (2009). Gene therapy for immunodeficiency due to adenosine deaminase deficiency. *N. Engl. J. Med.* 360, 447–458.
- Hacein-Bey-Abina, S., von Kalle, C., Schmidt, M., Le Deist, F., Wulffraat, N., McIntyre, E., Radford, I., Villeval, J.L., Fraser, C.C., Cavazzana-Calvo, M., et al. (2003). A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N. Engl. J. Med.* 348, 255–256.
- Pawliuk, R., Westerman, K.A., Fabry, M.E., Payen, E., Tighe, R., Bouhassira, E.E., Acharya, S.A., Ellis, J., London, I.M., Eaves, C.J., et al. (2001). Correction of sickle cell disease in transgenic mouse models by gene therapy. *Science* 294, 2368–2371.
- Ott, M.G., Schmidt, M., Schwarzwaelder, K., Stein, S., Siler, U., Koehl, U., Glimm, H., Kuhlcke, K., Schilz, A., Kunkel, H., et al. (2006). Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EV11, PRDM16 or SETBP1. *Nat. Med.* 12, 401–409.
- Boztug, K., Schmidt, M., Schwarzer, A., Banerjee, P.P., Diez, I.A., Dewey, R.A., Böhm, M., Nowrouzi, A., Ball, C.R., Glimm, H., et al. (2010). Stem-cell gene therapy for the Wiskott-Aldrich syndrome. *N. Engl. J. Med.* 363, 1918–1927.
- May, C., Rivella, S., Callegari, J., Heller, G., Gaensler, K.M., Luzzatto, L., and Sadelain, M. (2000). Therapeutic haemoglobin synthesis in  $\beta$ -thalassaemic mice expressing lentivirus-encoded human  $\beta$ -globin. *Nature* 406, 82–86.
- Weber, L., Poletti, V., Magrin, E., Antoniani, C., Martin, S., Bayard, C., Sadek, H., Felix, T., Meneghini, V., Antoniou, M.N., et al. (2018). An optimized lentiviral vector efficiently corrects the human sickle cell disease phenotype. *Mol. Ther. Methods Clin. Dev.* 10, 268–280.
- Molet, J.M., Petrykowska, H., Bouhassira, E.E., Feng, Y.Q., Miller, W., and Hardison, R.C. (2001). Sequences flanking hypersensitive sites of the beta-globin locus control region are required for synergistic enhancement. *Mol. Cell. Biol.* 21, 2969–2980.

16. Zhan, J., Johnson, I.M., Wielgosz, M., and Nienhuis, A.W. (2016). The identification of hematopoietic-specific regulatory elements for WASp gene expression. *Mol. Ther. Methods Clin. Dev.* *14*, 16077.
17. Lisowski, L., and Sadelain, M. (2007). Locus control region elements HS1 and HS4 enhance the therapeutic efficacy of globin gene transfer in  $\beta$ -thalassemic mice. *Blood* *110*, 4175–4178.
18. Sadelain, M., Wang, C.H., Antoniou, M., Grosveld, F., and Mulligan, R.C. (1995). Generation of a high-titer retroviral vector capable of expressing high levels of the human beta-globin gene. *Proc. Natl. Acad. Sci. USA* *92*, 6728–6732.
19. Delzor, A., Dufour, N., Petit, F., Guillemier, M., Houitte, D., Auregan, G., Brouillet, E., Hantraye, P., and Dégion, N. (2012). Restricted transgene expression in the brain with cell-type specific neuronal promoters. *Hum. Gene Ther. Methods* *23*, 242–254.
20. Grosveld, F., van Assendelft, G.B., Greaves, D.R., and Kollias, G. (1987). Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* *51*, 975–985.
21. Ryan, T.M., Behringer, R.R., Martin, N.C., Townes, T.M., Palmiter, R.D., and Brinster, R.L. (1989). A single erythroid-specific DNase I super-hypersensitive site activates high levels of human  $\beta$ -globin gene expression in transgenic mice. *Genes Dev.* *3*, 314–323.
22. Peterson, K.R., Clegg, C.H., Navas, P.A., Norton, E.J., Kimbrough, T.G., and Stamatoyannopoulos, G. (1996). Effect of deletion of 5'HS3 or 5'HS2 of the human beta-globin locus control region on the developmental regulation of globin gene expression in beta-globin locus yeast artificial chromosome transgenic mice. *Proc. Natl. Acad. Sci. USA* *93*, 6605–6609.
23. Tanimoto, K., Sugiura, A., Omori, A., Felsenfeld, G., Engel, J.D., and Fukamizu, A. (2003). Human  $\beta$ -globin locus control region HS5 contains CTCF- and developmental stage-dependent enhancer-blocking activity in erythroid cells. *Mol. Cell Biol.* *23*, 8946–8952.
24. Fedosyuk, H., and Peterson, K.R. (2007). Deletion of the human  $\beta$ -globin LCR 5'HS4 or 5'HS1 differentially affects  $\beta$ -like globin gene expression in  $\beta$ -YAC transgenic mice. *Blood Cells Mol. Dis.* *39*, 44–55.
25. Peterson, K.R., Fedosyuk, H., and Harju-Baker, S. (2012). LCR 5' hypersensitive site specificity for globin gene activation within the active chromatin hub. *Nucleic Acids Res.* *40*, 11256–11269.
26. Collis, P., Antoniou, M., and Grosveld, F. (1990). Definition of the minimal requirements within the human beta-globin gene and the dominant control region for high level expression. *EMBO J.* *9*, 233–240.
27. Fraser, P., Hurst, J., Collis, P., and Grosveld, F. (1990). DNaseI hypersensitive sites 1, 2 and 3 of the human  $\beta$ -globin dominant control region direct position-independent expression. *Nucleic Acids Res.* *18*, 3503–3508.
28. Fraser, P., Pruzina, S., Antoniou, M., and Grosveld, F. (1993). Each hypersensitive site of the human  $\beta$ -globin locus control region confers a different developmental pattern of expression on the globin genes. *Genes Dev.* *7*, 106–113.
29. The ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* *489*, 57–74.
30. van Arensbergen, J., FitzPatrick, V.D., de Haas, M., Pagie, L., Sluimer, J., Bussemaker, H.J., and van Steensel, B. (2017). Genome-wide mapping of autonomous promoter activity in human cells. *Nat. Biotechnol.* *35*, 145–153.
31. Maricque, B.B., Dougherty, J.D., and Cohen, B.A. (2017). A genome-integrated massively parallel reporter assay reveals DNA sequence determinants of cis-regulatory activity in neural cells. *Nucleic Acids Res.* *45*, e16.
32. Inoue, F., Kircher, M., Martin, B., Cooper, G.M., Witten, D.M., McManus, M.T., Ahituv, N., and Shendure, J. (2017). A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Res.* *27*, 38–52.
33. Kwasnieski, J.C., Fiore, C., Chaudhari, H.G., and Cohen, B.A. (2014). High-throughput functional testing of ENCODE segmentation predictions. *Genome Res.* *24*, 1595–1602.
34. Inoue, F., Kreimer, A., Ashuach, T., Ahituv, N., and Yosef, N. (2018). Massively parallel characterization of regulatory dynamics during neural induction. *bioRxiv*. <https://doi.org/10.1101/370452>.
35. Yáñez-Cuna, J.O., Arnold, C.D., Stampfel, G., Boryń, L.M., Gerlach, D., Rath, M., and Stark, A. (2014). Dissection of thousands of cell type-specific enhancers identifies dinucleotide repeat motifs as general enhancer features. *Genome Res.* *24*, 1147–1156.
36. Smith, R.P., Taher, L., Patwardhan, R.P., Kim, M.J., Inoue, F., Shendure, J., Ovcharenko, I., and Ahituv, N. (2013). Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. *Nat. Genet.* *45*, 1021–1028.
37. White, M.A., Myers, C.A., Corbo, J.C., and Cohen, B.A. (2013). Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks. *Proc. Natl. Acad. Sci. USA* *110*, 11952–11957.
38. Grossman, S.R., Zhang, X., Wang, L., Engreitz, J., Melnikov, A., Rogov, P., Tewhey, R., Isakova, A., Deplancke, B., Bernstein, B.E., et al. (2017). Systematic dissection of genomic features determining transcription factor binding and enhancer function. *Proc. Natl. Acad. Sci. USA* *114*, E1291–E1300.
39. Kreimer, A., Yan, Z., Ahituv, N., and Yosef, N. (2019). Meta-analysis of massive parallel reporter assay enables functional regulatory elements prediction. *bioRxiv*. <https://doi.org/10.1101/202002>.
40. Patwardhan, R.P., Hiatt, J.B., Witten, D.M., Kim, M.J., Smith, R.P., May, D., Lee, C., Andrie, J.M., Lee, S.L., Cooper, G.M., et al. (2012). Massively parallel functional dissection of mammalian enhancers in vivo. *Nat. Biotechnol.* *30*, 265–270.
41. Philipsen, S., and Hardison, R.C. (2018). Evolution of hemoglobin loci and their regulatory elements. *Blood Cells Mol. Dis.* *70*, 2–12.
42. Jackson, J.D., Petrykowska, H., Philipsen, S., Miller, W., and Hardison, R. (1996). Role of DNA sequences outside the cores of DNase hypersensitive sites (HSs) in functions of the  $\beta$ -globin locus control region. Domain opening and synergism between HS2 and HS3. *J. Biol. Chem.* *271*, 11871–11878.
43. Engel, J.D., and Tanimoto, K. (2000). Looping, linking, and chromatin activity: new insights into beta-globin locus regulation. *Cell* *100*, 499–502.
44. Stamatoyannopoulos, G., Nienhuis, A.W., Majerus, P., and Varmus, H. (1994). *The Molecular Basis of Blood Diseases* (W.B. Saunders).
45. King, D.C., Taylor, J., Elnitski, L., Chiaromonte, F., Miller, W., and Hardison, R.C. (2005). Evaluation of regulatory potential and conservation scores for detecting cis-regulatory modules in aligned mammalian genome sequences. *Genome Res.* *15*, 1051–1060.
46. Urbinati, F., Campo Fernandez, B., Masiuk, K.E., Poletti, V., Hollis, R.P., Koziol, C., Kaufman, M.L., Brown, D., Mavilio, F., and Kohn, D.B. (2018). Gene therapy for sickle cell disease: a lentiviral vector comparison study. *Hum. Gene Ther.* *29*, 1153–1166.
47. Levasseur, D.N., Ryan, T.M., Reilly, M.P., McCune, S.L., Asakura, T., and Townes, T.M. (2004). A recombinant human hemoglobin with anti-sickling properties greater than fetal hemoglobin. *J. Biol. Chem.* *279*, 27518–27524.
48. Donnelly, M.L.L., Luke, G., Mehrotra, A., Li, X., Hughes, L.E., Gani, D., and Ryan, M.D. (2001). Analysis of the aphthovirus 2A/2B polyprotein “cleavage” mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal “skip”. *J. Gen. Virol.* *82*, 1013–1025.
49. Kurita, R., Suda, N., Sudo, K., Miharada, K., Hiroyama, T., Miyoshi, H., Tani, K., and Nakamura, Y. (2013). Establishment of immortalized human erythroid progenitor cell lines able to produce enucleated red blood cells. *PLoS ONE* *8*, e59890.
50. King, A.D., Huang, K., Rubbi, L., Liu, S., Wang, C.Y., Wang, Y., Pellegrini, M., and Fan, G. (2016). Reversible regulation of promoter and enhancer histone landscape by DNA methylation in mouse embryonic stem cells. *Cell Rep.* *17*, 289–302.
51. Tallack, M.R., Magor, G.W., Dartigues, B., Sun, L., Huang, S., Fitzcock, J.M., Fry, S.V., Glazov, E.A., Bailey, T.L., and Perkins, A.C. (2012). Novel roles for KLF1 in erythropoiesis revealed by mRNA-seq. *Genome Res.* *22*, 2385–2398.
52. Kang, Y., Kim, Y.W., Yun, J., Shin, J., and Kim, A. (2015). KLF1 stabilizes GATA-1 and TAL1 occupancy in the human  $\beta$ -globin locus. *Biochim. Biophys. Acta* *1849*, 282–289.
53. Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* *27*, 1017–1018.
54. Kumar, M., Keller, B., Makalou, N., and Sutton, R.E. (2001). Systematic determination of the packaging limit of lentiviral vectors. *Hum. Gene Ther.* *12*, 1893–1905.

55. Morgan, R.A., Unti, M.J., Aleshe, B., Brown, D., Osborne, K.S., Koziol, C., Ayoub, P.G., Smith, O.B., O'Brien, R., Tam, C., et al. (2020). Improved titer and gene transfer by lentiviral vectors using novel, small  $\beta$ -globin locus control region elements. *Mol. Ther.* 28, 328–340.
56. Canté-Barrett, K., Mendes, R.D., Smits, W.K., van Helsing-van Wijk, Y.M., Pieters, R., and Meijerink, J.P. (2016). Lentiviral gene transfer into human and murine hematopoietic stem cells: size matters. *BMC Res. Notes* 9, 312.
57. Pott, S., and Lieb, J.D. (2015). What are super-enhancers? *Nat. Genet.* 47, 8–12.
58. Pennacchio, L.A., Bickmore, W., Dean, A., Nobrega, M.A., and Bejerano, G. (2013). Enhancers: five essential questions. *Nat. Rev. Genet.* 14, 288–295.
59. Levasseur, D.N., Ryan, T.M., Pawlik, K.M., and Townes, T.M. (2003). Correction of a mouse model of sickle cell disease: lentiviral/antisickling beta-globin gene transduction of unmobilized, purified hematopoietic stem cells. *Blood* 102, 4312–4319.
60. Morgan, R.A., Gray, D., Lomova, A., and Kohn, D.B. (2017). Hematopoietic stem cell gene therapy: progress and lessons learned. *Cell Stem Cell* 21, 574–590.
61. Zhang, Y., An, L., Yue, F., and Hardison, R.C. (2016). Jointly characterizing epigenetic dynamics across multiple human cell types. *Nucleic Acids Res.* 44, 6721–6731.
62. Masiuk, K.E., Zhang, R., Osborne, K., Hollis, R.P., Campo-Fernandez, B., and Kohn, D.B. (2019). PGE2 and poloxamer syneronic F108 enhance transduction of human HSPCs with a  $\beta$ -globin lentiviral vector. *Mol. Ther. Methods Clin. Dev.* 13, 390–398.
63. Shigaki, D., Adato, O., Adhikari, A.N., Dong, S., Hawkins-Hooker, A., Inoue, F., Juven-Gershon, T., Kenlay, H., Martin, B., Patra, A., et al. (2019). Integration of multiple epigenomic marks improves prediction of variant impact in saturation mutagenesis reporter assay. *Hum. Mutat.* 40, 1280–1291.
64. Dull, T., Zufferey, R., Kelly, M., Mandel, R.J., Nguyen, M., Trono, D., and Naldini, L. (1998). A third-generation lentivirus vector with a conditional packaging system. *J. Virol.* 72, 8463–8471.
65. Cooper, A.R., Patel, S., Senadheera, S., Plath, K., Kohn, D.B., and Hollis, R.P. (2011). Highly efficient large-scale lentiviral vector concentration by tandem tangential flow filtration. *J. Virol. Methods* 177, 1–9.
66. Romero, Z., Urbinati, F., Geiger, S., Cooper, A.R., Wherley, J., Kaufman, M.L., Hollis, R.P., de Assin, R.R., Senadheera, S., Sahagian, A., et al. (2013).  $\beta$ -globin gene transfer to human bone marrow for sickle cell disease. *J. Clin. Invest.* 123, 3317–3330.
67. Urbinati, F., Wherley, J., Geiger, S., Fernandez, B.C., Kaufman, M.L., Cooper, A., Romero, Z., Marchioni, F., Reeves, L., Read, E., et al. (2017). Preclinical studies for a phase 1 clinical trial of autologous hematopoietic stem cell gene therapy for sickle cell disease. *Cytotherapy* 19, 1096–1112.
68. Hindson, B.J., Ness, K.D., Masquelier, D.A., Belgrader, P., Heredia, N.J., Makarewicz, A.J., Bright, I.J., Lucero, M.Y., Hiddessen, A.L., Legler, T.C., et al. (2011). High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal. Chem.* 83, 8604–8610.