Published in final edited form as:

Blood Cells Mol Dis. 2010 December 15; 45(4): 343–350. doi:10.1016/j.bcmd.2010.08.005.

A Functional Screen for Regulatory Elements that Improve Retroviral Vector Gene Expression

Amy C. Groth¹ and David W. Emery^{1,2}

¹Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA

²Institute for Stem Cell and Regenerative Medicine, University of Washington, Seattle, WA

Abstract

Recombinant retroviruses constitute the most common class of gene delivery vectors used in hematopoietic cell-based gene therapy. However, the use of these vectors can be limited by inadequate levels of transgene expression, often mediated by expression variegation and vector silencing due to chromosomal position effects. Toward the goal of addressing this problem, we sought to identify cis-regulatory elements from the human genome that can improve the level and stability of retroviral vector gene expression. Libraries of size-selected fragments from the human genome were cloned into the “double-copy” position of the gammaretroviral reporter vector MGP2, and the resulting vectors underwent several rounds of transduction and selection for high-level vector GFP expression. From this screen we identified both enhancer-like elements and vector mutations associated with increased vector expression. One element, H-11, exhibited enhancer activity in a mouse bone marrow progenitor colony assay, a human promoter trap assay, and a long-term mouse bone marrow transplant assay. This element seems to be an orientation-dependent, tissue-independent enhancer.

Keywords

enhancer; insulator; retrovirus; regulatory element; screen

Introduction

Recombinant vectors based on retroviruses, such as gammaretroviruses and, more recently, lentiviruses, constitute the second most common means of gene delivery used in the setting of clinical gene therapy, with over 330 trials reported to date [1]. In the setting of hematopoietic gene therapy, this vector class is used almost exclusively due to its ability to efficiently integrate therapeutic genes into the target cell genome, resulting in long-term and stable gene transfer, even after many rounds of cell division [2,3]. However, vector integration occurs at sites throughout the target cell genome [4], rendering vector gene

© 2010 Elsevier Inc. All rights reserved.

Corresponding Author: David W. Emery, Research Associate Professor, Institute for Stem Cell and Regenerative Medicine, University of Washington, 815 Mercer St, Box 358056, Seattle, WA 98195, Phone 206-221-0609, demery@uw.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Contributions:

AG designed and performed experiments, assembled data, helped prepare the manuscript, and approved the final submission. DWE helped design and perform experiments, assembled data, helped prepare the manuscript, provided funding, and approved the final submission.

expression subject to the influence of the surrounding chromatin. This in turn can result in variations in vector gene expression and even overt vector silencing, a phenomenon known as chromosomal position effects (as reviewed in [5]).

Progress has been made toward reducing the impact of negative chromosomal position effects on retroviral vector expression. Several lines of evidence indicate that vectors based on recombinant gammaretroviruses are inherently more sensitive to these effects than vectors based on recombinant lentiviruses, presumably due to differences in integration site preferences [4]; however, even lentiviral vectors are still prone to position-effect silencing [6]. In addition, silencing position effects can be reduced through the use of potent enhancers and chromatin-opening elements such as those found at the β -globin locus control region [7]. The choice of promoter sequences can likewise affect the rate of vector silencing [8], with some promoter elements such as that from the HNRPA2B1-CBX3 locus proving highly resistant to vector silencing [9]. We and others have found that silencing chromosomal position effects can also be mitigated, at least in part, by flanking retroviral vectors with a class of cis-regulatory elements known as barrier chromatin insulators [8,10–15]. This approach has proven especially attractive for two main reasons. First, these elements do not exhibit classical enhancer activity on their own [16], and as such they do not override other potentially important tissue-specific regulatory elements that may be included in a vector. Second, at least some of these elements also include another class of chromatin insulator, known as enhancer-blockers [17,18], that can further prevent the inappropriate activation of vector sequences by cellular enhancers located near some sites of vector integration.

However, even with these advances, vector silencing remains problematic: barrier chromatin insulators identified to date are not fully effective [8,12,13,14]; many applications require tissue-specific expression, precluding the use of ubiquitously active promoters or enhancers; and many enhancers are often tissue-specific [19], precluding their use in settings where ubiquitous expression is required. In addition, even ideal regulatory elements may be incompatible with some vectors due to the fortuitous presence of short stretches of homologous sequences that can lead to vector deletions and limit vector titers [20,21]. For these reasons, we sought to expand the repertoire of available genetic elements that could be used to increase the level and stability of retroviral vector expression without inversely impacting vector titer or stability. We used a candidate-independent approach based on a functional genetic screen involving a well-characterized gammaretroviral reporter vector, MGPN2. We describe here the products that came out of this screen, including the characterization of a potentially useful element that appears to be an orientation-dependent, tissue-independent enhancer.

Materials

Functional screen

Genomic DNA was prepared from the human fibrosarcoma cell line HT1080 [22], and digested individually, doubly and triply with the methylation-independent restriction enzymes *Dra*I, *Sca*I, and *Stu*I. Fragments ranging from 0.5 to 2 kb were size-fractionated on an agarose gel for each digestion, and then combined and cloned into the *Not*I site located in the "double-copy" position of the 3' long-terminal repeat (LTR) of the gammaretroviral reporter vector MGPN2 [8,23]. As described [8,24], elements inserted at this site are copied into the 5' LTR during generation of vector provirus, resulting in the flanking arrangement diagrammed in Fig. 1A. Greater than 180,000 bacterial colonies from three separate ligations and transformations were pooled into 1 large-scale bacterial culture, and DNA was prepared using the Qiafilter DNA Maxiprep kit (Qiagen Valencia, CA). As diagrammed in Fig. 1B, the functional screen was initiated by transfecting the Phoenix-Eco packaging cell

line [25] with 6 μ g of library plasmid using FuGENE6 (Roche Applied Science, Indianapolis, IN). Media was changed at 24 hrs, and viral supernatant was harvested at 48 hrs post-transfection, filtered, and stored at -80°C . This viral supernatant was subsequently used to transduce the amphotropic packaging cell lines PA317 [26] at a sub-saturating multiplicity of infection to maintain an average of less than one viral genome per cell. As diagrammed in Fig. 1C, at about 1 week post transduction, between 1,000 to 10,000 of the brightest 2–10% GFP(+) cells were collected using a FACS Aria cell sorter (BD Biosciences, San Jose, California). The sorted populations were expanded and virus was again harvested and used to transduce the ecotropic packaging cell line PE501 [27]. The sort was repeated, and virus was used to transduce PA317 cells, etc. After the second and third PA317 sorts (3rd and 5th rounds of transduction), genomic DNA was prepared using the QIAamp Blood Mini kit (Qiagen). Candidate elements were isolated by PCR from this genomic DNA using primers flanking the original site of insertion in the 3' LTR, and subcloned before sequencing. Candidate sequences were compared to the May 2004 build of the human genome using the BLAT program of the UCSC genome browser [28]. GC content was determined using the GEECEE program [29]. All cell lines were maintained at 37°C , 7.5% CO_2 in Dulbecco's Modified Eagle's Medium with 8% heat-inactivated fetal bovine serum. This media was further supplemented with 4 $\mu\text{g}/\text{ml}$ polybrene for transduction cultures.

Transient enhancer assay

A GATEWAY® (Invitrogen, Carlsbad, CA) attP4-*ccdB*-attP1 cloning cassette [30] was cloned into the *Sma*I site of the commercially available firefly luciferase reporter plasmid pGL3-P (Promega, Madison, WI), creating the pGL3-P4-1+ luciferase vector. To test for enhancer activity, candidates were amplified using candidate specific attB-containing primers, and cloned into the attP cassette in one orientation via a Clonase reaction, according to the manufacturer's protocol (Invitrogen). The cHS4 insulator, SV40 enhancer and Zeocin resistance gene (as a neutral spacer element) were also cloned into this vector as controls. 100 ng of each pGL3-based plasmid was co-transfected with 5 ng of the control renilla luciferase reporter plasmid pRL via FuGENE6 transfection into NIH3T3 or HT1080 cells. At 48 hours post-transfection, cells were lysed and analyzed for luciferase expression using a Dual-Luciferase Reporter (DLR) assay (Promega).

Promoter trap enhancer assay

The plasmid reporter construct for the promoter trap enhancer assay is diagrammed in Fig. 5A. In short, this involved first generating a conventional promoter trap cassette consisting of an internal ribosomal entry site (IRES) sequence, followed by a neomycin phosphotransferase (Neo) drug resistance gene and two polyadenylation elements (one from the SV40 and one from the BGH gene locus). The candidate regulatory element H-11 was then inserted between the two polyadenylation elements. Constructs containing the following control elements were also generated: the promoter from SV40; the enhancer from SV40; the cHS4 chromatin insulator; and a neutral spacer we have used previously (from the G6PD gene [13]). 10 μg of each construct was transfected into HT1080 cells using FuGENE6, and the cells were subsequently selected with 500 $\mu\text{g}/\text{mL}$ G418. Approximately 10 days later, G418-resistant colonies were stained with methylene blue in methanol and counted. Colony numbers were adjusted for molar amounts of plasmid transfected.

Mouse bone marrow transduction assays

Candidate regulatory elements from the functional screen were amplified by PCR using *Not*I-containing candidate-specific primers and cloned into the *Not*I site in the double copy position of the gammaretroviral vector MGP2. We also included the following control vectors: MGP2 with no insert; MGP2 containing the cHS4 insulator (vector INS4(+)[8]); and MGP2 containing a neutral spacer (from the G6PD gene [13]). All vectors were

prepared in the ecotropic packaging cell line GP+E86 [31]. Mouse bone marrow cells were transduced as previously described [8,13]. Briefly, bone marrow cells were harvested from 5-fluorouracil-treated, 6–16 week old female B6xD2 F1 mice, and pre-induced by 48 hr culture with the cytokines IL-3, IL-6, and SCF. The cells were then transduced by a further 48 hr co-culture on irradiated producer cells in the same media supplemented with 8 μ g/mL polybrene. For bone marrow progenitor assays, cells were plated in complete methylcellulose (Stemcell Technologies, Vancouver, BC), without and with 1.0 mg/ml G418. For bone marrow transplantation assays, cells were transplanted into irradiated (1050 cGy) syngeneic B6xD2 F1 recipients. Peripheral blood cells were collected by retro-orbital bleed, red blood cells were lysed by hypotonic shock, and the remaining cells were analyzed by flow cytometry (FACScan, BD Biosciences).

Results

Functional Screen

In order to identify cis-regulatory elements from the human genome that can consistently improve the level and stability of retroviral vector gene expression, we carried out the functional screen outlined in Fig. 1. In short, this entailed inserting size-fractionated fragments from a human DNA library into the "double-copy" position of the GFP gammaretroviral reporter vector MGP2. As diagrammed in Fig. 1A, from this position the candidate element is copied into the 5' LTR during proviral integration, such that the integrated provirus is flanked on both sides by the candidate element. This offers two main advantages. First, the flanking arrangement serves to insulate the integrated provirus from the target cell genome on both sides, rather than on just one side. Second, some cis-regulatory elements, such as chromatin insulators, are thought to function best when used in pairs [32]. As outlined in Fig. 2B, this vector library was then subjected to multiple rounds of vector transduction and isolation of those cells containing vector provirus that consistently expressed the highest level of vector GFP using complementary vector packaging cell lines (see Materials for details). As diagrammed in Fig. 1C, this entailed the use of cell sorting to isolate the brightest 2–10% of transduced cells at each round of selection.

After three and five rounds of transduction and selection, we isolated individual cell clones and sequenced the candidate inserts. As summarized in Table 1, in three independent experiments we saw an apparent convergence on a small number of inserts, from 18 independent inserts out of 46 clones analyzed after three rounds, to only 8 independent inserts out of 62 clones analyzed at 5 rounds. This represents a 3-fold reduction in the complexity of the pools ($P=0.002$) between these two stages in the selection. After both three and five rounds of selection we noted at least one dominant insert in each pool. These inserts, listed as the "most frequent" in Table 1, typically accounted for about 58% of all clones analyzed after 3 rounds of selection, and about 82% of all clones analyzed after 5 rounds of selection, again providing evidence of convergence in the pools. In one case, the same dominant insert observed after 3 rounds of selection was also the dominant insert after 5 rounds of selection (experiment 3, clone I-3). In the other two cases, the dominant clones changed with additional selection.

As summarized in Table 2, we identified a total of 24 independent inserts after three or five rounds of functional enrichment. These inserts ranged in size from 355 bp to 1109 bp, with an average length of 764 bp. We mapped these inserts onto build hg17 of the human genome. One was in repetitive DNA and could not be localized; fourteen of the remaining 23 were intergenic. Of the nine that were in genes, five were in the first intron of at least one predicted transcript, while two spanned at least one exon. Distance from the nearest

transcription start ranged from 2.2 kb to 323 kb, with an average distance of 62 kb. Finally, the GC content of these inserts ranged from 38% to 64%, with an average of 49%.

Vector Mutations

In addition to the three independent experiments summarized in Table 1, we also had two library pools that were essentially taken over by a vector population that expressed GFP at strikingly high levels. In these cases, we were not able to isolate candidate inserts from the resulting provirus. We also observed some instances where candidates from the three experiments summarized in Table 1 showed increased GFP expression levels when virus was isolated from the final sorted clones, but not when the candidate inserts were recloned into the original reporter vector MGP2. We concluded that, in both of these cases, the elevated levels of vector GFP expression were most likely due to mutations in the vector backbone, rather than effects of the candidate inserts. We conducted a limited amount of sequencing to identify these mutations. As diagrammed in Fig. 2A, we found several small mutations throughout the vector sequence, including: i) A point mutation 11 bp downstream of the 5' LTR; ii) A polymorphism 208 bp upstream of the GFP gene; iii) Two non-synonymous point mutations near the end of the coding region for the 254 amino acid GFP protein (E252K and K245G); iv) A 12 bp deletion in the P_{gk} promoter; and v) A point mutation in the coding region of the 267 amino acid Neo protein (D230N). The mutation closest to the 5' LTR was found in three different clones, and therefore may represent a polymorphism found in the original vector plasmid preparation. In addition, we found many examples that appeared to have arisen either before, or concurrent with, the formation of the first provirus, and involved apparent crossovers between the vector and plasmid sequences (Fig. 2A). We identified three separate crossovers in three separate pools. The crossover points occur in a 31 bp region near the end of the GFP gene, and a 95 bp region of the vector backbone, ~200 bp upstream of the 5' LTR. Each of the three identified crossovers alters and extends the GFP open reading frame (ORF) (Fig. 2B). Each also creates an ORF in the antisense direction with a distinct 5' end (Fig. 2C).

Assessment of Candidates in a Classical Transient Enhancer Assay

As a first approach for functionally characterizing the candidate inserts isolated after three and five rounds of selection, we turned to a well established luciferase enhancer assay based on transient transfection in mouse NIH3T3 cells, the same parental cell line used to generate the viral packaging cell lines used in the screen. As diagrammed in Fig. 3, 11 of 26 candidate elements (42%) showed at least some enhancer activity in this assay, with only 8 of 26 candidate elements (31%) reaching a P value of ≤ 0.05 after Bonferroni correction for multiple testing. These eight candidates increased expression from 1.4-fold to 4.0-fold compared to the neutral spacer control. However, none of the candidates reached the 5.4-fold enhancement achieved with the SV40 enhancer that was used as the positive control. Indeed, the top four candidates (I-9, I-15, 7-5-17L and H-51L) only increased expression an average 2.5-fold. This compares to no statistically significant increase compared to the no-insert control and a statistically significant but very modest 1.2-fold increase for the cHS4 chromatin insulator. It is unclear why the cHS4 insulator increased expression at all in this assay, since it is reportedly devoid of enhancer activity [16]. We hypothesize that this activity reflects the ability of the cHS4 element to reduce epigenetic silencing through the constitutive maintenance of an "open" histone code [33,34]. When we carried out similar studies in human HT1080 cells, none of the candidates were found to enhance expression beyond a statistically significant threshold, even though the SV40 positive control was twice as effective in this setting (data not shown). From these studies we concluded that many of the candidate elements that were isolated in the screen appeared to have some enhancer activity, but that this activity appeared to be species- or tissue-specific.

Assessment of Candidates in a Mouse Bone Marrow Culture Assay

We next sought to extend the analysis of the candidate elements to primary cells. For this purpose, we turned to a well-characterized mouse bone marrow progenitor colony assay [13]. In short, the candidate elements were again cloned into the gammaretroviral reporter vector MGP2 used in the initial screen, and these vectors were used to transduce mouse bone marrow. Transduced progenitor cell colonies were then selected in semi-solid methylcellulose cultures, and analyzed for vector GFP expression. As seen in Fig. 4, almost all of the elements had little to no effect on vector GFP expression, and one, F-15, actually appeared to decrease expression. This data suggests that the activity of these elements seen in the original screen may reflect a high degree of tissue specificity. However, one element, H-11, did increase the amount of GFP expression in the mouse bone marrow transduction assay 3.3-fold, from an average of 64 mean fluorescent units (MFU) for the uninsulated vector to an average of 213 mfu. This was almost as much as the 3.7-fold increase observed with the cHS4 chromatin insulator control, and far greater than the statistically indistinguishable effect seen with the SV40 enhancer control in this setting.

Assessment of Candidate H-11 in a Promoter-Trap Enhancer Assay

The candidate element H-11 produced a small and marginally significant effect in the classical transient enhancer assay in mouse NIH3T3 cells (Fig. 3), but no discernable effect using the same assay in human HT1080 cells (data not shown). However, this element did present a fairly strong effect in the mouse bone marrow transduction and culture assay. These two models differ in several key aspects. First, in the bone marrow transduction model, the effects of the candidate element are assessed in multiple cell lineages and states of differentiation. Second, in the bone marrow transduction model, vector expression can be influenced both by intact chromatin and by the unique context of the surrounding genetic loci, including possible sources of silencing heterochromatin, enhancers, promoters, etc. Third, the retroviral vector platform used in the bone marrow transduction model allows for the candidate element to be copied into the 5' LTR, essentially doubling the number of candidate copies and thereby providing an opportunity to increase or even alter the resulting activity. And finally, the retroviral vector platform also includes additional enhancers and multiple promoters, providing many opportunities for enhancer-promoter cross-talk and competition.

In order to start discriminating between these variables, we turned to a modified version of the stable promoter trap assay. As diagrammed in Fig. 5A, this entailed starting with a conventional promoter trap construct in which a neomycin drug-resistance gene (Neo) is placed downstream of an internal ribosomal entry site (IRES) sequence. On the occasions that this promoterless gene integrates into an actively transcribed gene (in the sense orientation), the Neo gene is expressed and thereby confers drug resistance to the corresponding clone. In our modification, we added the H-11 candidate, or other control sequences, downstream of this cassette. If the added sequence contains an enhancer, it should increase the frequency of drug-resistant colonies by activating the promoters of otherwise silent genes into which the reporter integrates. This assay differs from the transient enhancer assay both by assessing activity in the setting of intact chromatin, and by surveying activity at many sites of the target cell genome. However, this assay also differs from the mouse bone marrow transduction assay and the attendant retroviral vector platform by using only one copy of the candidate element and by eliminating potential competition with other enhancers or promoters. Finally, we carried out these studies in human HT1080 cells in order to further test the lineage- and species-specificity of this element.

As seen in Fig. 5B, the construct containing no insert formed colonies at an average rate of $119/10^{12}$ molecules of transfected DNA. Inclusion of the SV40 promoter showed no

increase in colonies, indicating that the Neo gene was not being expressed through a tandem integration or transcription read-through mechanism. The SV40 promoter also did not decrease colony numbers when cloned in the reverse orientation, indicating there were no problems with antisense transcription. The neutral spacer, G6PD, likewise had no significant effect on the colony number. When the candidate element was tested in the (+) orientation, the number of colonies increased 2-fold, to an average of 247 ($P=0.002$ vs no insert control). However, when this same element was tested in the (−) orientation, there was no significant increase in colony formation (avg. 141, $P=0.14$ vs no insert control and $P=0.006$ vs the (+) orientation). In comparison, the positive control SV40 enhancer increased the colony formation nearly 3-fold in both orientations, to an average 321 for the (+) orientation and 301 for the (−) orientation. Interestingly, the cHS4 chromatin insulator also increased the frequency of colonies in this assay, about 3-fold in the (+) orientation and 2-fold in the (−) orientation. This was again surprising since the cHS4 element reportedly does not exhibit enhancer activity on its own [16], and our own data from the classic transient enhancer assay showed only a very modest effect, averaging less than 10% of the activity seen with the SV40 enhancer. However, this element does exhibit a potent chromatin-opening activity and the ability to block the encroachment of silencing heterochromatin [33,34], and both of these activities could also conceivably lead to the activation of otherwise silent promoters depending on the context.

Given these positive results, we conducted additional studies with H-11 using the transient luciferase assay in HT1080 cells. This included eliminating the GATEWAY cloning sequences in order to rule out any potential effect on transgene expression, and testing H-11 in both orientations. As before, no enhancer activity was detected (data not shown). In addition, H-11 was cloned into the promoter position of the transient luciferase reporter plasmid pGL3-C in both orientations, and tested in a transient assay in HT1080 cells. Neither construct showed any promoter activity (data not shown).

Assessment of Candidate H-11 in a Mouse Bone Marrow Transplantation Assay

The studies described so far demonstrate that H-11 can enhance or otherwise activate gene expression in a manner that is (mostly) dependent on intact chromatin, and that this activity extends to multiple cell lineages from mouse and humans. However, a question remained as to whether this element simply enhanced the expression of linked genes (e.g. classical enhancer activity), or if it also exhibited the capacity to actively resist silencing (e.g. chromatin opening activity). In order to address this issue, we turned to our well-described mouse bone marrow transduction and transplantation assay [8]. In this system, mice are transplanted with vector-transduced bone marrow, and the level of vector expression is assessed in peripheral white blood cells (WBC) over time. In this system, vector expression is typically silenced after a few months, so that only a small fraction of WBC that contain vector provirus actually express this provirus. This silencing presumably reflects the genome-wide changes in heterochromatin and gene expression patterns that occur during the engraftment and differentiation of primitive hematopoietic stem cells. In this system, strong enhancers such as those in the viral LTR or the alpha-globin locus, can enhance vector expression for a month or two, but eventually they succumb to silencing [8,11]. In contrast, this silencing can be reduced through the use of either barrier insulators such as cHS4, or regulatory elements such as those found in the beta-globin locus control region that have autonomous chromatin opening activity [7].

As seen in Fig. 6A, we found that the candidate element H-11 initially increased the frequency of white blood cells expressing vector GFP compared to the no-insert control, although not to the degree seen with the positive control cHS4 insulator. However, this advantage was eventually lost, so that by 24 weeks the fraction of GFP-positive white blood cells had dropped by ~70% for the cohort of mice treated with the H-11 vector, to a level

that was even lower than that seen for the no-insert controls. In contrast, when we looked at the level of expression for those cells that continued to express vector GFP, we saw a different story. As seen in Fig. 6B, the candidate element H-11 appeared to increase the mean fluorescence of vector GFP (in those cells that continued to express GFP) at both early and late time points post-transplant compared to the no-insert control. Taken together, we believe these results indicate that H-11 is acting as a classical enhancer, but does not exhibit additional chromatin opening or barrier insulator activity.

Discussion

It was our initial intention to carry out a functional screen to identify cis-regulatory elements from the human genome that could consistently improve the level and stability of retroviral vector gene expression. Two lines of evidence indicated that this goal was nominally met: the library pools appeared to converge on a limited number of clones, especially after 5 rounds of selection; and 42% of isolated elements demonstrated at least some statistically significant activity in a transient enhancer assay. In particular, the top 4 candidates demonstrated nearly half the activity of the highly potent SV40 enhancer. However, this screen also had two key drawbacks. First, the selection schema allowed for the serial passage and selection of vectors with backbone mutations, rather than candidate inserts, that resulted in greatly elevated levels of vector GFP expression. Although we did not seek to capitalize on these mutations in the studies presented here, they nevertheless point to possible areas for further improvements in vector engineering. We presume that this drawback could be overcome by recloning the insert library into the original reporter vector after each round of selection. Second, the vast majority of cis-regulatory elements we did identify appeared to only be active in the cell line used for the screen, namely mouse fibroblast NIH3T3 cells. Recent efforts to catalog the regulatory elements of the human genome suggest that many enhancers are likely to be tissue-specific [19], although the relative frequency of ubiquitously active enhancers is presumably greater than the rate of 1 in 23 observed in our studies. We presume that this drawback could be overcome by carrying out the selection in multi-potential cells such as hematopoietic progenitor cells or embryonic stem cells. As an added bonus, the high degree of silencing chromosomal position effects in these cell types would likely increase the chances of isolating barrier insulators or enhancers with chromatin-opening activity as well.

In spite of these drawbacks, we did identify one candidate element with noteworthy properties - H-11. This element is located on chromosome 12, p13.32, near the end of the long third intron of the TSPAN9 (tetraspanin 9) gene. The human fragment that we isolated is 548 bp, and maps to one of the putative regulatory elements computationally identified by Prabhakar, et al. [35]. Although not one of the most conserved elements, the 159 bp VISTA conserved noncoding sequence (chr12:3,256,889-3,257,048) almost completely overlaps with H-11 (chr12:3,256,913- 3,257,460). Our data indicates that H-11 can enhance or otherwise activate gene expression in both transient and stable assays, although it appears to function best in the context of intact chromatin. It works in both mouse and human cells, and appears to function in multiple hematopoietic and non-hematopoietic lineages. In total, the evidence presented here suggests that this element functions as a classical enhancer, elevating the expression of linked genes. However, H-11 exhibited two properties that are not consistent with some definitions of enhancers. First, data from the promoter-trap enhancer assay suggest that this element may only function in one orientation. This attribute was confirmed in a pilot mouse bone marrow culture assay, where elevated GFP expression was only observed when the reporter vector was flanked with the H-11 element in the same orientation used for the studies shown in Fig. 4, but not in the opposite orientation. Second, data from the long-term bone marrow transplantation study suggest that H-11 can elevate expression of a linked transgene while that gene is active, but on its own cannot prevent that

gene from being silenced over time. The candidate element is similar to many other enhancers in this regard, including among others the viral LTR of the reporter vector MGP2 [8,12] and the HS-40 enhancer from the α -globin locus [11]. However, in both these examples the enhancers could remain active if the vector was also flanked with a barrier insulator (the cHS4 element). Future studies will be needed to determine whether a similar approach would allow the H-11 enhancer to remain functional long-term as well.

Acknowledgments

The authors wish to thank D.W. Russell and A. Lieber for starting materials for the promoter trap enhancer assay construct, and G.P. Nolan for the Phoenix-Eco packaging cell line. This work was funded by P01 HL53750 (GS) and HL75713 (DWE).

References

1. www.wiley.co.uk/genmed/clinical
2. Emery DW, Nishino T, Murata K, et al. Hematopoietic Stem Cell Gene Therapy. *Int. J. Hemat.* 2002; 75:228–236.
3. Kohn DB, Sadelain M, Dunbar C, et al. American Society of Gene Therapy (ASGT) ad hoc subcommittee on retroviral-mediated gene transfer to hematopoietic stem cells. *Molec. Ther.* 2003; 8:180–187. [PubMed: 12907140]
4. Berry C, Hannenhalli S, Leipzig J, Bushman FD. Selection of target sites for mobile DNA integration in the human genome. *PLoS Comput. Biol.* 2006; 2:e157. [PubMed: 17166054]
5. Emery, DW.; Aker, M.; Stamatoyannopoulos, G. Chromatin Insulators and Position Effects. In: Makrides, SC., editor. *Gene Transfer and Expression in Mammalian Cells*. Norwood, MA: EIC Laboratories, Inc.; 2003. p. 381–395.
6. Hong S, Hwang DY, Yoon S, et al. Functional analysis of various promoters in lentiviral vectors at different stages of *in vitro* differentiation of mouse embryonic stem cells. *Mol. Ther.* 2007; 15:1630–1639. [PubMed: 17609656]
7. May C, Rivella S, Callegari J, et al. Therapeutic haemoglobin synthesis in beta-thalassaemic mice expressing lentivirus-encoded human beta-globin. *Nature*. 2000; 406:82–86. [PubMed: 10894546]
8. Emery DW, Yannaki E, Tubb J, Stamatoyannopoulos G. A chromatin insulator protects retrovirus vectors from position effects. *Proc. Natl. Acad. Sci. U. S. A.* 2000; 97:9150–9155. [PubMed: 10908661]
9. Zhang F, Thornhill SI, Howe SJ, et al. Lentiviral vectors containing an enhancer-less ubiquitously acting chromatin opening element (UCOE) provide highly reproducible and stable transgene expression in hematopoietic cells. *Blood*. 2007; 110:1448–1457. [PubMed: 17456723]
10. Rivella S, Callegari JA, May C, et al. The cHS4 insulator increases the probability of retroviral expression at random chromosomal integration sites. *J. Virol.* 2000; 74:4679–4687. [PubMed: 10775605]
11. Emery DW, Yannaki E, Tubb J, et al. Development of virus vectors for gene therapy of β chain hemoglobinopathies: flanking with a chromatin insulator reduces γ -globin gene silencing *in vivo*. *Blood*. 2002; 100:2012–2021. [PubMed: 12200360]
12. Yannaki E, Tubb J, Aker M, et al. Topological constraints governing the use of the chicken HS4 chromatin insulator in oncoretrovirus vectors. *Mol. Ther.* 2002; 5:589–598. [PubMed: 11991750]
13. Aker M, Tubb J, Groth AC, et al. Extended core sequences from the cHS4 insulator are necessary for protecting retroviral vectors from silencing position effects. *Hum. Gene. Ther.* 2007; 18:333–343. [PubMed: 17411365]
14. D'Apolito D, Baiamonte E, Bagliesi M, et al. The sea urchin *sns5* insulator protects retroviral vectors from chromosomal position effects by maintaining active chromatin structure. *Mol. Ther.* 2009; 17:1434–1441. [PubMed: 19352322]
15. Arumugam PI, Scholes J, Perelman N, et al. Improved human beta-globin expression from self-inactivating lentiviral vectors carrying the chicken hypersensitive site-4 (cHS4) insulator element. *Mol. Ther.* 2007; 15:1863–1871. [PubMed: 17622240]

16. Chung JH, Bell AC, Felsenfeld G. Characterization of the chicken beta-globin insulator. *Proc. Natl. Acad. Sci. U. S. A.* 1997; 94:575–580. [PubMed: 9012826]
17. Bell AC, West AG, Felsenfeld G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell.* 1999; 98:387–396. [PubMed: 10458613]
18. Recillas-Targa F, Pikaart MJ, Burgess-Beusse B, et al. Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proc. Natl. Acad. Sci. U. S. A.* 2002; 99:6883–6888. [PubMed: 12011446]
19. The ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature.* 2007; 447:799–816. [PubMed: 17571346]
20. Emery DW, Chen H, Li Q, Stamatoyannopoulos G. Development of a condensed Locus Control Region cassette and testing in retrovirus vectors for A-gamma-globin. *Blood Cells Mol. Dis.* 1998; 24:322–339. [PubMed: 10087991]
21. Pathak VK, Temin HM. Broad spectrum of in vivo forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: deletions and deletions with insertions. *Proc. Natl. Acad. Sci. U. S. A.* 1990; 87:6024–6028. [PubMed: 2166940]
22. Rasheed S, Nelson-Rees WA, Toth EM, et al. Characterization of a newly derived human sarcoma cell line (HT-1080). *Cancer.* 1974; 33:1027–1033. [PubMed: 4132053]
23. Cheng L, Du C, Murray D, et al. A GFP reporter system to assess gene transfer and expression in human hematopoietic progenitor cells. *Gene Ther.* 1997; 4:1013–1022. [PubMed: 9415306]
24. Hantzopoulos PA, Sullenger BA, Ungers G, Gilboa E. Improved gene expression upon transfer of the adenosine deaminase minigene outside the transcriptional unit of a retroviral vector. *Proc. Natl. Acad. Sci. U. S. A.* 1989; 86:3519–3523. [PubMed: 2542934]
25. Kinsella TM, Nolan GP. Episomal vectors rapidly and stably produce high-titer recombinant retrovirus. *Hum. Gene Ther.* 1996; 7:1405–1413. [PubMed: 8844199]
26. Miller AD, Buttimore C. Redesign of retrovirus packaging cell lines to avoid recombination leading to helper virus production. *Mol. Cell. Biol.* 1986; 6:2895–2902. [PubMed: 3785217]
27. Miller AD, Rosman GJ. Improved retroviral vectors for gene transfer and expression. *BioTechniques.* 1989; 7:980–990. [PubMed: 2631796]
28. <http://genome.ucsc.edu/>
29. <http://bioweb.pasteur.fr/seqanal/interfaces/geecee.html>
30. Tubb J, Groth AC, Leong L, Emery DW. Simultaneous sequence transfer into two independent locations of a reporter vector using multisite gateway technology. *Biotechniques.* 2005; 39:553–557. [PubMed: 16235567]
31. Markowitz D, Goff S, Bank A. A safe packaging line for gene transfer: separating viral genes on two different plasmids. *J. Virol.* 1988; 62:1120–1124. [PubMed: 2831375]
32. Wallace J, Felsenfeld G. We gather together: insulators and genome organization. *Curr. Opin. Genet. Dev.* 2007; 17:400–407. [PubMed: 17913488]
33. Litt MD, Simpson M, Gaszner M, et al. Correlation between histone lysine methylation and developmental changes at the chicken beta-globin locus. *Science.* 2001; 293:2453–2455. [PubMed: 11498546]
34. Li CL, Emery DW. The cHS4 chromatin insulator reduces gammaretrovirus vector silencing by epigenetic modifications of integrated provirus. *Gene. Ther.* 2008; 15:49–53. [PubMed: 17989708]
35. Prabhakar S, Poulin F, Shoukry M, et al. Close sequence comparisons are sufficient to identify human cis-regulatory elements. *Genome Res.* 2006; 16:855–863. [PubMed: 16769978]

Fig. 1.

Schema for functional screen. (A) Screening vector. A library of DNA fragments from the human genome were cloned into the "double-copy" position of the 3' LTR of the gammaretroviral reporter vector MGP2 [23]. This vector expresses GFP from the 5' LTR promoter, and Neo from an internal Pkg promoter. During reverse transcription and integration of the vector genome, the insert is copied into the 5' LTR, resulting in the flanking arrangement diagrammed. LTR, long terminal repeat; GFP, Green fluorescent protein; PGKNeo, phosphoglycerate kinase promoter, Neomycin resistance gene; Insert, library DNA or control sequences. (B) Selection strategy workflow. (C) Example of flow

cytometric data demonstrating the gating used to isolate the brightest transduced cells by fluorescence-activated cell sorting.

Fig. 2.

Vector mutations. (A) Map of the plasmid for the reporter vector MGP2 indicating the location of various mutations. * = unique point mutation or small deletion; x = unique crossover. See the legend for Fig. 1 for vector details. (B) The ends of the GFP open reading frame (ORF) and the ORFs generated by the 3 unique crossovers, as aligned by ClustalW2 (the previous 240 amino acids are identical). (C) The beginning of the 3 unique ORFs running antisense to the GFP gene that were created by the three unique crossovers, as aligned by ClustalW2 (the remaining 217 aa are identical).

Fig. 3.

Candidate activity in transient enhancer assay. A transient luciferase assay was performed with the indicated candidates in NIH3T3 cells (see Materials for details). Data represent the ratio of firefly luciferase to the control luciferase, Renilla, from 3 to 6 independent transfections. Error bars represent the standard deviation. None, no insert control; SV40, positive SV40 enhancer control; Neutral, neutral spacer control consisting of a fragment of Zeocin resistance gene; cHS4, the 1.2 kb cHS4 chromatin insulator. P values based on T-test versus the neutral spacer control, adjusted for multiple testing by Bonferroni correction.

Fig. 4.

Candidate activity in mouse bone marrow transduction culture assay. Mouse bone marrow cells were transduced with the MGP2 retroviral reporter vector containing the indicated candidates, and the resulting progenitor colonies were analyzed for expression of vector GFP. Data represent the average mean fluorescence of GFP positive cells across individual G418-selected clones (8 clones were analyzed per candidate in each of 1 to 4 experiments). Error bars represent the standard deviation. Controls are the same as in Fig. 3, except that sequence from G6PD was used as the neutral control. P values based on T-test versus the neutral spacer control, adjusted for multiple testing by Bonferroni correction.

Fig. 5. Candidate H-11 activity in a Promoter-Trap Enhancer Assay. (A) Structure of reporter vector. The base vector for the promoter trap is shown. IRES, Internal Ribosome Entry Site, Neo, Neomycin resistance gene; pA, polyadenylation sequence; ori/amp, bacterial sequences; insert, inserts to be tested. (B) Colony Assay. Linearized plasmids containing the indicated insertions were transfected into HT1080 cells, and drug-resistant colonies were selected with G418. Colony numbers were normalized for the size of the plasmid transfected (in molecules). Data represent the average $\pm$ standard deviation from 3 independent experiments. P values based on T-test versus the empty vector. None, no insert control; Prom, SV40 promoter; SV40, positive SV40 enhancer control; Neutral, neutral spacer

consisting of G6PD sequence; cHS4, the 1.2 kb cHS4 chromatin insulator; H-11, candidate element H-11; (+) and (−), indicated element in one of two orientations.

Fig. 6.

Candidate H-11 activity in a mouse bone marrow transduction and transplantation assay. The GFP reporter vector MGP2 containing either no insert, the 1.2 kb cHS4 element, or candidate H-11 were used to transduce mouse bone marrow cells, which were then transplanted into myeloablated syngeneic recipients. The fraction of peripheral white blood cells expressing vector GFP (A), as well as the mean fluorescence of these GFP-positive cells (B), was determined at the indicated times post-transplant by flow cytometry. Data represent the average $\pm$ standard deviation for 4–5 mice per group.

Table 1

Clonal complexity.

Experiment	3 Rounds of Selection			5 Rounds of Selection		
	Clones analyzed (no.)	Independent clones (no.)	Most frequent	Clones analyzed (no.)	Independent clones (no.)	Most frequent
1	18	9	6 of 18 (cl. 7-4)	20	3	17 of 20 (cl. 7-5-3)
2	14	5	10 of 14 (cl. H-2)	24	3	17 of 24 (cl. H-5-1L)
3	14	4	11 of 14 (cl. I-3)	18	2	17 of 18 (cl. I-3)

Table 2

Candidate Sequences.

Candidate	Size (bp)	Location (5' border) (May 2004 freeze)	Gene location	G/C content (%)	Nearest TSS (bp)*
7-2	589	chr9:42,366,656	intergenic	38	153,621
7-4	727	chr8:142,803,891	intergenic	60	217,379
7-5	714	chr3:52,125,091	intronic	51	29,124
7-7	648	chr8:72,067,297	intergenic	46	323,351
7-8	1,030	chr9:113,432,701	intronic/exonic	56	2,233
7-10	779	chr4:54,258,318	intronic	46	6,199
7-11	810	chr9:44,365,489	intergenic	54	11,777
7-15	1,048	chr16:22,748,338	intronic	46	15,339
7-16	777	chr10:99,274,619	intronic	51	25,861
F-15	710	chr19:37,690,526	intergenic	40	41,752
F-16	1,050	chr1:68,734,282	intergenic	38	59,496
H-1	929	chr8:10,446,239	intergenic	51	21,907
H-2	987	chr4:613,687	intronic	62	4,292
H-3	355	chr11:2,182,378	intergenic	64	40,642
H-11	548	chr12:3,256,913	intronic	56	6,543
H-13	1,109	chr11:118,428,348	intronic/exonic	53	3,578
I-1	649	chr4:126,167,771	intergenic	42	221,579
I-3	571	chr11:122,289,235	intronic	43	30,552
I-9	655	chr7:73,276,739	intergenic	62	10,233
I-15	757	chr7:24,153,425	intergenic	43	56,378
7-5-3	893	chr6:83,152,017	intergenic	38	21,759
7-5-17L	676	couldn't localize	repetitive	41	
H-5-1L	761	chr2:15,815,033	intergenic	52	132,665
H-5-5	556	chr6:150,777,842	intergenic	54	3,745
Average	764		14/23 intergenic	49	62,609

* TSS - Transcription Start Site