Gammaretroviral Vector Integration Occurs Overwhelmingly Within and Near DNase Hypersensitive Sites

Mingdong Liu,¹ Chang Long Li,^{1,*} George Stamatoyannopoulos,¹ Michael O. Dorschner,^{2,†}
Richard Humbert,² John A. Stamatoyannopoulos,² and David W. Emery^{1,3}

Abstract

Concerns surrounding the oncogenic potential of recombinant gammaretroviral vectors has spurred a great deal of interest in vector integration site (VIS) preferences. Although gammaretroviral vectors exhibit a modest preference for integration near transcription start sites (TSS) of active genes, such associations only account for about a third of all VIS. Previous studies suggested a correlation between gammaretroviral VIS and DNase hypersensitive sites (DHS), which mark chromatin regions associated with *cis*-regulatory elements. In order to study this issue directly, we assessed the correlation between 167 validated gammaretroviral VIS and a deep genome-wide map of DHS, both determined in the same cell line (the human fibrosarcoma HT1080). The DHS map was developed by sequencing individual DNase I cleavage sites using massively parallel sequencing technologies. These studies revealed an overwhelming preference for integrations associated with DHS, with a median distance of only 238 bp between individual VIS and the nearest DHS for the experimental dataset, compared to 3 kb for a random dataset and 577 to 1457 bp for two unrelated cell lines ($p < 0.001$). Indeed, nearly 84% of all VIS were found to be located within 1 kb of a DHS ($p = 10^{-43}$). Further, this correlation was statistically independent from the association with TSS. The preference for DHS far exceeds that seen for other hallmarks of gammaretroviral VIS, including TSS, and may help explain several aspects of gammaretroviral vector biology, including the mechanism of VIS selection, as well as the relative frequency and underlying biology of gammaretroviral vector-mediated genotoxicity.

Introduction

GENE DELIVERY VECTORS BASED ON replication-defective gammaretroviruses have been used widely in biomedical research, and comprise the most widely used class of integrating viral vectors in clinical gene therapy applications (Kohn *et al.*, 2003; www.wiley.co.uk/genmed/clinical/). Although such vectors have proven generally safe, and in some cases effective, several lines of evidence indicate that interactions between these vectors and the genomic environment surrounding sites of vector integration can lead to vector-mediated genotoxicity (Baum *et al.*, 2006). These observations in turn have spurred investigations into the genomic properties of vector integration sites (VIS), as well as the mechanism(s) underlying VIS preferences. Large-scale studies of gammaretroviral VIS published to date essentially all point to a modest preference for integrations near the

following: transcription start sites (TSS), especially for transcriptionally active genes; CpG islands; and DNase hypersensitive sites (DHS) (Wu, *et al.*, 2003; Berry *et al.*, 2006; Lewinski *et al.*, 2006). For example, our own studies found that 36% of all gammaretroviral VIS in murine bone marrow progenitor cells resided within a ± 5 -kb window of a TSS, compared to only 14% for a simulated random control set (Aker *et al.*, 2006). In a more recent study with the human fibrosarcoma cell line HT1080, we found similar results, with 27% of all VIS within a ± 5 -kb window of a TSS, compared to 6% for a simulated random control set (Li *et al.*, 2009). However, it is important to note that this preference for TSS is at best only modest, typically accounting for about a third or less of all VIS. Of the parameters studied to date, the least well studied is the association with DHS. Such sites reflect focal remodeling of the chromatin in association with active *cis*-regulatory foci within the living genome and are detectable experimentally on the basis of

¹Department of Medicine, Division of Medical Genetics, ²Department of Genome Sciences, and ³Institute for Stem Cell and Regenerative Medicine, University of Washington, Seattle, Washington.

*Current address: Department of Biochemistry and Molecular Biology, School of Preclinical and Forensic Medicine, State Key Laboratory of Biotherapy of Human Diseases, Sichuan University, Chengdu, China.

†Current address: The Permanente Medical Group, Department of Genetics, San Jose, California.

pronounced sensitivity to cleavage when intact nuclei are exposed to DNA modifying agents, canonically the nonspecific endonuclease DNase I (Wu, 1980; Elgin, 1984; Gross and Garrard, 1988). Early studies of a very limited number of integration sites ($n=12$) associated with wild-type Moloney murine leukemia virus found that essentially all sites were located within a few hundred base pairs of DHS (Vijaya *et al.*, 1986; Rohdewohld *et al.*, 1987). In contrast, more recent studies with recombinant gammaretroviral vectors found that only about 11% of VIS were associated with DHS (Berry *et al.*, 2006; Lewinski *et al.*, 2006). Although this preference was statistically significant compared to a simulated random control dataset, it nevertheless represents a striking difference from the earlier studies with wild-type virus. One difference between the earlier and later studies involve the criteria for defining DHS. In the earlier studies, DHS were defined by a conventional method of DNase treatment and Southern blotting. In the latter study, DHS were defined by the presence of two or more sites of DNase cleavage identified by high-throughput sequencing of DNA ends (Crawford *et al.*, 2006). However, deep sequencing data generated under the National Human Genome Research Institute ENCODE Project (ENCODE Project Consortium, 2007; <http://www.encodeproject.org>) suggest that this DHS mapping study may have significantly underestimated the total number of DHS in the target cell population. In addition, these previous studies either relied on integration events that had been derived under some form of selection, such as tumor formation (Vijaya *et al.*, 1986) or drug selection (Lewinski *et al.*, 2006), or on DHS mapping preformed in sources of cells other than those used for the original infection/transduction (Berry *et al.*, 2006; Lewinski *et al.*, 2006). The former has been shown to skew the integration sites repertoire (Aker *et al.*, 2006; Berry *et al.*, 2006), while the latter can lead to misidentification of DHS, which can often be tissue and lineage specific (Crawford *et al.*, 2006; Sekimata *et al.*, 2009; John *et al.*, 2011).

In order to resolve this apparent discrepancy in published studies, we carried out a deep genome-wide mapping of DHS in the human fibrosarcoma cell line HT1080 and then compared this map to a map of 167 gammaretroviral VIS that were derived without selection and also determined in HT1080 cells. Overall, we found an extremely high preference for gammaretroviral VIS within and near DHS. This preference was independent of the preference for TSS and accounted for nearly 84% of all gammaretroviral VIS based on a 1-kb window. The implications of these findings for understanding several aspects of vector-mediated genotoxicity associated with gammaretroviral vectors and for the mechanisms underlying vector integration are discussed.

Materials and Methods

Tissue culture

The human fibrosarcoma HT1080 (Rasheed *et al.*, 1974) was cultured in Dulbecco's modified Eagle's medium with 10% heat-inactivated fetal bovine serum at 37°C and 7.5% CO₂. All studies were carried out using aliquots from the same master cell bank to minimize sample variation.

Vector integration site datasets

The gammaretroviral VIS used for this analysis have been previously published (Li *et al.*, 2009). In short, HT1080 cells

were transduced with the gammaretroviral vectors MGP2 and INS4(+) described previously (Emery *et al.*, 2000) using viral supernatant from amphotropic packaging lines. Individual cell clones were subsequently isolated in the absence of selection, and transduced clones were identified by Southern blotting. VIS were subsequently determined by a variety of methods, including inverse PCR, linear amplification-mediated polymerase chain reaction (PCR), and in some cases, with a directed DNA library screening approach (see Li *et al.*, 2009 for details). Sequences were BLAST searched against the human genome using the UCSC Genome Browser (<http://genome.ucsc.edu/>) as previously described (Aker *et al.*, 2006). The simulated random integration dataset was generated essentially as described (Aker *et al.*, 2006).

Genome-wide mapping of DHS

Genome-wide mapping of DHS in HT1080 cells was carried out by combining portions of two previously described methods (Sabo *et al.*, 2004, 2006) and essentially are recently described (John *et al.*, 2011). In short, nuclei were isolated and treated with DNase I as for standard DHS mapping (Reitman *et al.*, 1993; Sabo *et al.*, 2006). Genomic DNA was then prepared and size fractionated so as to isolate fragments between 100 and 300 bp in length. Such "DNase double hit fragments" (DDHF) are generated by two cleavage "hits" occurring in close proximity, and their use imparts a much greater degree of specificity compared with analysis of individual free DNA ends, since the latter are contaminated with both nonspecific cleavage and random shear sites (Sabo *et al.*, 2006). The DDHF were then ligated directly to sequencing adapters and sequenced with an Illumina Genome Analyzer II. Twenty-seven base reads were mapped to the human reference genome using Illumina's ELAND extended alignment tool. From two independent runs we mapped 22,469,362 sequence tags. Hotspots of DHS were subsequently identified using the following algorithm.

First, enrichment was gauged in a small window (250 bp) relative to a local background model based on the binomial distribution, using the observed tags in a 50-kb surrounding window. Each mapped tag was assigned a z-score for the 250- and 50-kb windows centered on the tag. A hotspot was defined as a succession of neighboring tags within a 250-bp window, each with a z-score ≥ 2 . Once a hotspot was identified, the hotspot itself was assigned a z-score relative to the 250- and 50-kb windows centered on the average position of the tags forming the hotspot. A z-score is calculated as follows. Suppose n observed tags fall in the 250-bp window, and N total tags fall in the 50 kb surrounding the background window ($N \geq n$). Each tag in the background window is considered an "experiment," with favorable outcome if it falls in the smaller window. Assuming each base in the 50-kb window is equally likely, the probability of success for each tag is therefore $p = (250/50,000)$. Under these assumptions, the binomial distribution applies, and the expected number of tags falling in the smaller window is $\mu = Np$. The standard deviation of this expected value is $\sigma = [Np(1-p)]^{1/2}$. Finally, the z-score for the observed number of tags in the smaller window is $z = (n - \mu) / \sigma$. As a null model, we computationally generate tags uniformly over the uniquely mappable bases of the genome, using the same number of tags for the observed and random data. The random data also coalesce into

hotspots, which are identified, scored, and resolved into peaks using the same technique as for the observed data. For a given z -score threshold T , the false discovery rate (FDR) for the observed hotspot peaks with z -scores greater than T is estimated as $\text{FDR}(T) = (\text{no. of random peaks with } z \geq T) / (\text{no. of observed peaks with } z \geq T)$. Since the numerator, which is calculated on a dataset that is entirely null, likely overestimates the number of false positives in the observed data, this is likely a conservative estimate of the false discovery rate. For studies presented here, we used hotspots identified with an FDR of 0.05. Datasets for the human erythroid cell line K562 (<http://genome.ucsc.edu/>) and from primary human T helper 1 (Th1) cells (Sekimata *et al.*, 2009) derived in a similar manner were used as controls.

Results and Discussion

Correlation between VIS and DHS

Several lines of evidence suggest that many DHS are lineage and tissue specific. For this reason, we chose to compare DHS and VIS in the same cultured cell line. In particular, we chose to focus on the human fibrosarcoma HT1080 because it has a mostly diploid karyotype that has proven highly stable in our hands (Li *et al.*, 2009). As discussed in the Introduction and outlined in Table 1, we have previously carried out an extensive analysis of gammaretroviral VIS in this cell line, including the mapping of 167 independent sites of integration for conventional gammaretroviral reporter vectors (Li *et al.*, 2009). These VIS were derived in the absence of selection and thus represent an unbiased repertoire. As described at length elsewhere (Li *et al.*, 2009) and summarized in Table 1, these sites exhibit the conventional modest preference for gene-rich regions and TSS.

In order to map DHS in this cell line, we used a novel approach based on massively parallel sequencing. As described at length in the Methods section, this involved treating isolated nuclei with a low level of DNase I, isolating small DDHFs, and then sequencing the digested ends using the Illumina high-throughput sequencing platform. We mapped over 22 million independent sequence tags, and identified 193,104 DHS hotspots (based on a 5% false-discovery rate). These results are similar to those seen in several different cell lines with this technology and reflect the general frequency of DHS seen by hybridization-based technologies for the ENCODE region (Sabo *et al.*, 2006). As controls, we used DHS maps previously determined from two unrelated sources of human cells: the erythroid cell line K562 (<http://genome.ucsc.edu/>) and primary Th1 cells (Sekimata *et al.*, 2009). We also used a control VIS dataset of simulated random integration sites as previously described (Li *et al.*, 2009).

We subsequently determined the distance between each VIS and the nearest DHS using these datasets. As diagrammed in Fig. 1A, comparisons of the VIS and DHS datasets in HT1080 cells indicated that nearly 20% of all VIS resided within DHS, while the remaining VIS ranged from just a few bases to over 69 kb from the nearest DHS, with a median distance of 238 bp. The simulated VIS dataset ranged over a similar distance, but with a significantly longer median distance of 3010 bp and the occurrence of VIS within DHS of only 2% ($p < 0.001$). Comparison of the cumulative distribution of these two datasets using the non-parametric and distribution-free two-sample Kolmogorov-Smirnov (KS) test (Horn, 1977) indicates that these distributions are significantly different ($D = 0.67$, $p < 0.001$). Comparison between the VIS dataset determined in HT1080 cells and DHS datasets determined in the unrelated cell populations gave intermediate results. In both cases, approximately 20% of VIS were located within DHS. However, the remaining VIS ranged to well over 100,000 kb from the nearest DHS, with a commensurate skewing in the median distance to 577 bp for K562 cells and 1457 bp for primary Th1 cells. Analysis by the KS test indicated that both of these distributions are significantly different than the HT1080 dataset ($D = 0.28$, $p < 0.001$ for K562; $D = 0.38$, $p < 0.001$ for Th1), but similar to each other ($D = 0.12$, $p = 0.17$). This skewing was most evident by segregating the datasets into discrete bins. In particular, as seen in Fig. 1B, there was a statistically significant overrepresentation of VIS located within 250 bp of the nearest DHS in HT1080 cells compared to all three controls ($p < 0.05$), while there was a commensurate underrepresentation of VIS located over 10,000 kb from the nearest DHS in HT1080, again compared to all three controls ($p < 0.05$). In fact, over half (52%) of all authentic VIS were located within 250 bp of the nearest DHS, expanding to 84% for the 1-kb window.

Direct comparisons of the overlap between VIS and the regions around DHS serve to emphasize the overwhelming nature of this preference. For this purpose, p values were obtained using the binomial distribution, with the expectation (probability, p) equal to the total DHS space (approximately 1% of the HT1080 genome), divided by the total 27-mer uniquely mapable genomic space (2.334 gb for hg17). As summarized in Table 2, the probability of overlap between VIS and the regions around DHS ranged from $p = 10^{-16}$ for integrations directly in DHS to a high of $p = 10^{-43}$ for integrations located within 1 kb of a DHS.

Deeper analysis of the various datasets indicates that many of the HT1080 DHS associated with an internal VIS were either conserved in the other two DHS datasets (33%) or located within 250 bp of a DHS in the other two DHS

TABLE 1. PROPERTIES OF VECTOR INTEGRATION SITE DATASET

Dataset	No. of cell clones	No. of VIS ^a	Median distance to nearest TSS ^b	No. VIS ± 2 kb of TSS (%)	No. VIS ± 5 kb of TSS (%)
Vectors	87	167	23,852 bp	34 (20.4)	47 (28.1)
Simulated	n.a.	100	85,082 bp	3 (3.0)	6 (6.0)

^aVIS, vector integration sites.

^bTSS, transcription start sites.

FIG. 1. Distance between vector integration sites (VIS) and nearest DNase hypersensitive sites (DHS) for gammaretroviral vectors. **(A)** Cumulative distributions of the distances between individual VIS and the nearest DHS hotspot. VIS datasets include 167 gammaretroviral VIS determined experimentally in HT1080 cells, as well as 100 simulated integration sites as a random control. DHS datasets include genome-wide analysis in HT1080 cells, the human erythroid cell line K562, and human primary Th1 T cells. *p* values were determined by Kolmogorov–Smirnov (KS) test. **(B)** Discrete distribution of data presented in panel (A), for the indicated bins. *p* value comparing experimental and simulated random datasets within each bin was determined using the Z test for two proportions, and were Bonferroni corrected for multiple testing.

datasets (another 33%). This roughly reflects the degree of DHS conservation seen between the multiple cell types in smaller scale studies (ENCODE Project Consortium, 2007) and emphasizes the importance of comparing VIS and DHS in the same cell type. Deeper analysis of the matched DHS-VIS datasets in HT1080 cells further demonstrated there was no statistical difference in the preferences seen with the un-insulated vector MGP2 and this vector flanked with the cHS4 chromatin insulator ($D=0.18$, $p=0.13$), indicating that the presence of the cHS4 insulator has no significant effect on

TABLE 2. PROBABILITY OF OVERLAP BETWEEN VECTOR INTEGRATION SITE AND REGIONS AROUND DNASE HYPERSENSITIVE SITES

Distance to Nearest DHS	Cumulative no. of VIS	Probability*
Within DHS	33	10^{-16}
±250 bp	87	10^{-35}
±500 bp	108	10^{-36}
±1 kb	140	10^{-43}
±2 kb	149	10^{-25}
±10 kb	163	Insignificant
>10 kb	167	Insignificant

*Binomial distribution.

the integration site preference, consistent with previous studies (Li *et al.*, 2009).

Preference for DHS is independent of TSS

As discussed in the Introduction, several studies indicate that DHS are often found in association with TSS, presumably reflecting promoters and other promoter-proximal regulatory elements. Three lines of evidence indicate that the high incidence of VIS near DHS is not simply due to the prevalence of DHS near promoters. First, the frequency of VIS near DHS is far greater than the frequency of VIS near TSS (84% vs. 13% in a ±1-kb window, respectively). Second, when we parsed the VIS-DHS datasets into two groups, those with VIS within 5 kb of a TSS and those with VIS greater than 5 kb from a TSS, the resulting cumulative distributions were statistically indistinguishable (Fig. 2A, $D=0.21$, $p=0.09$). Third, linear regression analysis comparing the distance of each VIS to the nearest TSS vs. to the nearest DHS indicated no direct correlation (Fig. 2B, $R^2=0.004$, $p=0.44$). Indeed, nearly three quarters of all VIS located greater than 100 kb from the nearest TSS were nevertheless also located within 1 kb of a DHS (32 of 43 cases). Nevertheless, the fact that transcriptionally active genes are essentially always associated with promoter-proximal DHS (ENCODE Project Consortium, 2007), likely helps explain the modest preference for VIS near the TSS of transcriptionally active genes (for example, see Aker *et al.*, 2006).

Implications regarding vector-mediated genotoxicity

The findings reported here may help explain two important attributes of vector-mediated genotoxicity associated with gammaretroviral vectors. First, it may help explain why vectors based on gammaretroviruses appear to exhibit a higher rate of genotoxicity than vectors based on lentiviruses (Montini *et al.*, 2006), which generally do not target DHS (Lewinski *et al.*, 2006; Berry *et al.*, 2006; data not shown). The co-localization of DHS with *cis*-active elements within the mammalian genome spans the spectrum of known transcriptional and chromosomal regulatory activities, including transcriptional enhancers, promoters, silencers, insulators, locus control regions, and domain boundary elements (Gross and Garrard, 1988; Felsenfeld, 1996; Burgess-Beusse *et al.*, 2002). DHS have also been observed to coincide with sequences governing fundamental genomic

FIG. 2. VIS preference for DHS is independent of preference for TSS. **(A)** Cumulative distributions of the distances between DHS and gammaretroviral VIS that are located proximal (<5 kb) and distal (>5 kb) from TSS. *p* value determined by KS test. **(B)** Scatter plot showing distance between individual gammaretroviral VIS and nearest DHS versus nearest TSS. The coefficient of determination R^2 was determined by linear regression, with the resulting *p* value determined using the *F* statistic.

processes including attachment to the nuclear matrix (Jarman and Higgs, 1988; Kieffer *et al.*, 2002), and recombination (Zhang *et al.*, 2002). The observation that gammaretroviral vectors target DHS more frequently than lentiviral vectors means they have greater opportunity to disrupt the activity of these sites, leading in turn to a greater opportunity to disrupt the expression of cellular genes at many levels.

Second, the results described here could help explain two of the unexpected findings from a recent report from our laboratory involving an unbiased, genome-wide assessment of gammaretroviral vector-mediated dysregulation of cellular genes (Li *et al.*, 2009). Previous to this report, it was assumed that gammaretroviral vectors could only influence the expression of cellular genes that were located near the integrated provirus (generally within 100 kb of a TSS), and that most, if not all, of such effects would involve either an increase in the level of expression or some form of alter-

native splicing (Baum *et al.*, 2006). However, we found that about one third of all cases of vector-mediated cellular gene dysregulation involved a decrease in the level of expression of the affected gene (Li *et al.*, 2009). In addition, we found many examples where dysregulated genes were located at megabase distances from the nearest vector provirus. If we assume that the majority of *cis*-regulatory elements identified as DHS are involved in the positive regulation of gene expression, then their disruption by vector integration would likely lead to a decrease in the expression of genes dependent on those regulatory elements. Also, there is growing evidence that some classes of *cis*-regulatory elements marked by DHS, such as locus control regions and chromatin insulators, may function over very large distances and can even be involved in interchromosomal interactions (Zhao *et al.*, 2006; Dekker, 2008). As such, the disruption of such sites by vector integration could likewise affect cellular gene expression over very large distances or even on other chromosomes.

Implications regarding mechanisms underlying integration site preferences

Little is known about the mechanisms underlying the targeting of gammaretroviral vectors to specific regions of the genome. Our studies point to two possible mechanisms, diagrammed in Fig. 3. First, it is possible that the vector pre-integration complex is simply targeting the regions within or near DHS because of some physical property of these regions, such as ease of access or unique structures or tension in the DNA helix. Experimentally, the physiological resolution of DHS averages ~200–300 bp. Within this region, a core domain can be identified of approximately 80–120 bp over which critical DNA–protein interactions take place

Physical nature of DHS and border region:

FIG. 3. Models for VIS preferences. Two models are presented for why gammaretroviral vectors target DHS. Top: the physical nature of the boarder region, such as ease of access or helix tension, may facilitate vector integration. Bottom: the vector pre-integration complex may be tethered to proteins at the DHS or histone modifications near the DHS.

(Lowrey *et al.*, 1992). Cooperative binding of transcription factors to such core regions is sufficient to either exclude nucleosomes (Adams and Workman, 1995) or to replace canonical nucleosomes with less stable nucleosome variants containing histones H3.3 and H2A.Z (Jin *et al.*, 2009). However, less is known about the physical structure of the chromatin immediately surrounding DHS. The VIS patterns described in Fig. 1 suggest that gammaretroviral vectors preferentially integrate within DHS or within the distance of approximately two nucleosomes of the border of a DHS. More studies will be needed to determine whether these adjoining regions share unique structural properties. As diagrammed in the bottom of Fig. 3, it is also possible that the gammaretroviral vector pre-integration complex is targeting DHS through the tethering of this complex to proteins commonly found at these sites, such as common transcription or chromatin remodeling factors. This tethering could also conceivably involve specific histone modifications or even the variant histones H3.3/H2A.Z often found in association with DHS (Jin *et al.*, 2009). Future studies will also be needed to distinguish between these various possibilities. Given the large variation in DHS patterns observed between different cell types (ENCODE Project Consortium, 2007), our data in particular indicate that any such studies into the association between VIS and specific epigenetic properties need to be carried out within the same cell type and optimally within a homogeneous cell line.

Acknowledgments

This work was supported by grants from the National Heart, Lung, and Blood Institute, National Institutes of Health (RO1 HL75713 and PO1 HL53750).

Author Disclosure Statement

No competing financial interests exist.

References

- Adams, C.C., and Workman, J.L. (1995). Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. *Mol. Cell. Biol.* 15, 1405–1421.
- Aker, M., Tubb, J., Miller, D.G., *et al.* (2006). Integration bias of gammaretrovirus vectors following transduction and growth of primary mouse hematopoietic progenitor cells with and without selection. *Mol. Ther.* 14, 226–235.
- Baum, C., Kustikova, O., Modlich, U., *et al.* (2006). Mutagenesis and oncogenesis by chromosomal insertion of gene transfer vectors. *Hum. Gene Ther.* 17, 252–263.
- Berry, C., Hannonhalli, S., Leipzig, J., and Bushman, F.D. (2006). Selection of target sites for mobile DNA elements in the human genome. *PLoS Comput. Biol.* 2, e157.
- Burgess-Beusse, B., Farrell, C., Gaszner, M., *et al.* (2002). The insulation of genes from external enhancers and silencing chromatin. *Proc. Natl. Acad. Sci. U.S.A.* 99(Suppl. 4), 16433–16437.
- Crawford, G.E., Holt, I.E., Whittle, J., *et al.* (2006). Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). *Genome Res.* 16, 123–131.
- Dekker, J. (2008). Gene regulation in the third dimension. *Science* 319, 1793–1794.
- Elgin, S.C. (1984). Anatomy of hypersensitive sites. *Nature* 309, 213–214.
- Emery, D.W., Yannaki, E., Tubb, J., and Stamatoyannopoulos, G. (2000). A chromatin insulator protects retrovirus vectors from position effects. *Proc. Natl. Acad. Sci. U. S. A.* 97, 9150–9155.
- ENCODE Project Consortium. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799–816.
- Felsenfeld, G. (1996). Chromatin unfolds. *Cell* 86, 13–19.
- Gross, D.S., and Garrard, W.T. (1988). Nuclease hypersensitive sites in chromatin. *Annu. Rev. Biochem.* 57, 159–197.
- Horn, S.D. (1977). Goodness-of-fit tests for discrete data: a review and an application to a health impairment scale. *Biometrics* 33, 237–247.
- Jarman, A.P., and Higgs, D.R. (1988). Nuclear scaffold attachment sites in the human globin gene complexes. *EMBO J.* 7, 3337–3344.
- Jin, C., Zhang, C., Wei, G., *et al.* (2009). H3.3/H2A.Z double variant-containing nucleosomes mark ‘nucleosome-free regions’ of active promoters and other regulatory regions. *Nat. Genet.* 41, 941–945.
- John, S., Sabo, P.J., Thurman, R.E., *et al.* (2011). Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat. Genet.* 43, 264–268.
- Kieffer, L.J., Grealley, J.M., Landres, I., *et al.* (2002). Identification of a candidate regulatory region in the human CD8 gene complex by colocalization of DNase I hypersensitive sites and matrix attachment regions which bind SATB1 and GATA-3. *J. Immunol.* 168, 3915–3922.
- Kohn, D.B., Sadelain, M., Dunbar, C., *et al.* (2003). American Society of Gene Therapy (ASGT) ad hoc subcommittee on retroviral-mediated gene transfer to hematopoietic stem cells. *Mol. Ther.* 8, 180–187.
- Lewinski, M.K., Yamashita, M., Emerman, M., *et al.* (2006). Retroviral DNA integration: viral and cellular determinants of target-site selection. *PLoS Pathog.* 2, e60.
- Li, C.L., Xiong, D., Stamatoyannopoulos, G., and Emery, D.W. (2009). Genomic and functional assays demonstrate reduced gammaretroviral vector genotoxicity associated with use of the chS4 chromatin insulator. *Mol. Ther.* 17, 716–724.
- Lowrey, C.H., Bodine, D.M., and Nienhuis, A.W. (1992). Mechanism of DNase I hypersensitive site formation within the human globin locus control region. *Proc. Natl. Acad. Sci. U. S. A.* 89, 1143–1147.
- Montini, E., Cesana, D., Schmidt, M., *et al.* (2006). Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration. *Nat. Biotechnol.* 24, 687–696.
- Rasheed, S., Nelson-Rees, W.A., Toth, E.M., *et al.* (1974). Characterization of a newly derived human sarcoma cell line (HT-1080). *Cancer* 33, 1027–1033.
- Reitman, M., Lee, E., Westphal, H., and Felsenfeld, G. (1993). An enhancer/locus control region is not sufficient to open chromatin. *Mol. Cell. Biol.* 13, 3990–3998.
- Rohdewohld, H., Weiher, H., Reik, W., *et al.* (1987). Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. *J. Virol.* 61, 336–343.
- Sabo, P.J., Hawrylycz, M., Wallace, J.C., *et al.* (2004). Discovery of functional noncoding elements by digital analysis of

- chromatin structure. *Proc. Natl. Acad. Sci. U. S. A.* 101, 16837–16842.
- Sabo, P.J., Kuehn, M.S., Thurman, R., *et al.* (2006). Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat. Methods* 3, 511–518.
- Sekimata, M., Perez-Melgosa, M., Miller, S.A., *et al.* (2009). CCCTC-binding factor and the transcription factor T-bet orchestrate T helper 1 cell-specific structure and function at the interferon-gamma locus. *Immunity* 31, 551–564.
- Vijaya, S., Steffen, D.L., and Robinson, H.L. (1986). Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites in chromatin. *J. Virol.* 60, 683–692.
- Wu, C. (1980). The 5' ends of *Drosophila* heat shock genes in chromatin are hypersensitive to DNase I. *Nature* 286, 854–860.
- Wu, X., Li, Y., Crise, B., and Burgess, S.M. (2003). Transcription start sites in the human genome are favored sites for MLV integration. *Science* 300, 1749–1751.
- Zhang, Y., Strissel, P., Strick, R., *et al.* (2002). Genomic DNA breakpoints in AML1/RUNX1 and ETO cluster with topoisomerase II DNA cleavage and DNase I hypersensitive sites in t(8;21) leukemia. *Proc. Natl. Acad. Sci. U. S. A.* 99, 3070–3075.
- Zhao, Z., Tavoosidana, G., Sjolinder, M., *et al.* (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* 38, 1341–1347.

Address correspondence to:

Dr. David W. Emery
Division of Medical Genetics & Institute for Stem Cell
and Regenerative Medicine
University of Washington
815 Mercer St, Room N536
Box 358056
Seattle, WA 98109

E-mail: demery@uw.edu

Dr. John A. Stamatoyannopoulos
Department of Genome Sciences
University of Washington
Foege S310A
1705 NE Pacific Street
Box 355065
Seattle, WA 98195

E-mail: jstam@uw.edu

Received for publication August 30, 2010;
accepted after revision August 7, 2011.

Published online: October 7, 2011.