

Engineering the Human Genome: Reflections on the Beginning

Frederic D. Bushman

Introduction

THIS ARTICLE PRESENTS a history of the gene therapy field, seen from my perspective as a participant, and a few speculations on where it is all going. Stepping back, it has been remarkable to see the current wave of successes in gene and cell therapy. A thousand years from now, if there are still humans around, I bet they will look back on our era as pivotal—the period when humans first successfully engineered their own genetic makeup.

Proteins That Bind and Rearrange DNAs

My road to gene therapy started with studies of DNA-binding proteins. I had the good fortune to carry out my doctoral work with Mark Ptashne, a colorful figure if ever there was one. Mark isolated phage lambda repressor in the 1960s (Ptashne, 1967a,b), leading to the first biochemical demonstration of gene control by a sequence-specific DNA-binding protein. His race with Walter Gilbert to isolate a repressor protein (Mark working on lambda and Wally on lac repressor) was the subject of a television documentary narrated by George C. Scott. When I joined the lab in the 1980s, Mark's group was driving the field of transcriptional regulation. Work from this period is summarized in Mark's book *A Genetic Switch*, which describes with exceptional clarity the operation of the bistable switch regulating induction of the lambda prophage (Ptashne, 1986). My graduate studies of gene control were published in fancy journals (Bushman *et al.*, 1985, 1989; Bushman and Ptashne, 1986, 1988), but merit only a few paragraphs in the book, a testament to the amount of research summarized.

In Mark's lab in the late 1980s, one of the leading questions concerned how proteins bound at one DNA site could influence events far away along the DNA chain. Transcriptional enhancers provided one example. Some of the best work on this question was from the field of DNA recombination, notably work from Kiyoshi Mizuuchi and Robert Craigie. Mizuuchi had established an *in vitro* system that carried out transposition of phage Mu DNA efficiently (Mizuuchi, 1983), a reaction related to retroviral integration, which provided the basis for a striking series of mechanistic experiments, many together with Craigie. They showed that the Mu recombination proteins bound to distant DNA rec-

ognition sites and folded up into an ordered complex—only after correct assembly were the DNA breaking and joining steps initiated (Craigie *et al.*, 1984, 1985; Craigie and Mizuuchi, 1985, 1987). Eventually, many integrating elements, including retroviral preintegration complexes, would be shown to assemble into related structures.

Around this time, the newly discovered HIV virus was much in the news, and I grew interested in the mechanism of viral DNA integration. Retroviral particles contain RNA genomes, which after infection are copied to DNA by reverse transcription. The DNA copy is then integrated into a host-cell chromosome. This reaction involved DNA-binding proteins (i.e., HIV integrase) and also captured problems in DNA enzymology. Today retroviral vectors of course are widely used for gene transfer in human gene therapy.

I joined Craigie and Mizuuchi at the National Institutes of Health as a postdoc in 1989 to study retroviral integration, and in short order we had purified integrase proteins working *in vitro* (Bushman *et al.*, 1990; Craigie *et al.*, 1990). HIV integrase alone, it turned out, could form a covalent bond between model viral DNA and model target DNA under the right conditions. This provided the tool for a series of mechanistic studies, a line of work we are still in to this day (Bushman and Wang, 1994; Pruss *et al.*, 1994b; Farnet *et al.*, 1996; Carteau *et al.*, 1999; Gao *et al.*, 2004; Diamond and Bushman, 2005, 2006; Gupta *et al.*, 2010).

Later, the *in vitro* HIV integration system was picked up by the pharmaceutical industry. After an immense effort at Merck, the first HIV integrase inhibitors were approved by the Food and Drug Administration in 2007 (Summa *et al.*, 2008). Several of my trainees were recruited by Merck and helped with this project.

Both Kiyoshi and Mark are prizewinning researchers, but their styles could not have been more different. Kiyoshi has the reserve of a samurai warrior and usually wears a black headband to hold back his long hair. Mark is mercurial and drives a Porsche. Mark was interested in sweeping simplifying ideas and tended to ignore data that failed to match his latest theory. Kiyoshi was intensely focused on technology and seemed to love the engineering for its own sake. At one point, projects in Kiyoshi's lab required oligonucleotide substrates with phosphorothioates of defined stereochemistries—Kiyoshi disappeared for several weeks

and then returned with the needed DNAs, having synthesized them himself (Engelman *et al.*, 1991; Mizuuchi and Adzuma, 1991). Seeing both Kiyoshi and Mark up close was remarkably instructive—it requires allegiance to your idea to drive it ahead through inevitable setbacks, but you sure better pay close attention to methods and data quality as well.

HIV Integration in Chromosomes

In 1993 I started my own laboratory at the Salk Institute, focusing on molecular mechanisms of HIV replication. I trained on phage, and HIV was the successor, providing a rich window into the molecular biology of vertebrate cells. Our early studies explored the viral and cellular components of viral integration complexes (Miller *et al.*, 1995) and interactions with cellular proteins such as nucleosomes (Pruss *et al.*, 1994a; Bor *et al.*, 1995).

From the start, we were interested in gene therapy, influenced by neighbors such as Ted Friedman at University California–San Diego (UCSD) and Inder Verma at Salk. In one early experiment, I fused the DNA-binding domain of lambda repressor onto HIV integrase and showed that integration *in vitro* was favored near repressor-binding sites in target DNA (Bushman, 1994). In 1994 this made a splash, though the technology has so far not been translated effectively into vector systems useful in patients. Nevertheless, the repressor–integrase fusions did provide a conceptual prototype for the fusions of DNA-binding domains and nucleases that are popular today.

In the late 1990s, I was really excited by the first articles on complete genome sequences. The data established among other things the tremendous force that mobile DNA represented in biological evolution. Thanks to training with Kiyoshi and Mark, I had a good overview of mobile DNA mechanisms, and retrovirology provided another perspective. In 2001 I wrote about it in a book on how DNA moves between cells and what happens as a result (*Lateral DNA Transfer—Mechanisms and Consequences*) (Bushman, 2001). Happily the book was favorably reviewed in several major journals. One of the more memorable callouts was from the science fiction writer Greg Bear, who called the book “one of my favorite science books of the decade” and wrote that it helped him in writing his novel *Darwin’s Radio*, in which mobile DNA runs amok in the human genome!

The draft genome sequence of humans, newly available in 2001, suggested that about 8% of the human genome is composed of retroviral sequences. Most of these sequences were found outside genes, probably selected to minimize disruption of the human transcription units (Brady *et al.*, 2009). So where in the human genome did HIV target integration?

In 2002, we set out to sequence a large number of HIV integration sites, read out the flanking human DNA bases, align the reads on the newly available draft human genome sequence, and assess associations with chromosomal features mapped by others (Schroder *et al.*, 2002). Happily, Joe Ecker at Salk had a high-throughput sequencing line set up, which he had used to help sequence the *Arabidopsis* genome. We sequenced about 500 sites of HIV integration, aligned them to the human genome, and began the job of looking for patterns.

For this we worked closely with statistician Charles Berry at UCSD, a fruitful collaboration that continues to this day (Berry *et al.*, 2006, 2012, 2014). Chuck introduced me to what was possible when sequence information is analyzed using advanced methods and to the value of the statistical mindset in general. My early mentors from the phage school almost never used statistics—experiments involved forming a hypothesis; carrying out a simple experiment, often involving a single pair of experimental and control samples; and assessing whether or not the idea was supported. The general feeling was that if statistics were needed, then the battle was already over and you had lost. However, given many thousands of bases of integration site sequence, and billions of bases of human DNA that you want to compare against, which is itself annotated in dozens of ways, you have no hope of eyeballing the answer. Chuck calmly taught us to arrange the data in a manageable form and articulate answerable questions. He then devised clever computational routes to answers.

By comparing integration site positions to their genomic annotations, we found that HIV favors integration quite strongly in cellular transcription units (Schroder *et al.*, 2002). We then analyzed transcription using Affymetrix microarrays and found that active transcription units were particularly favored. We also found strong clustering of integration sites. For this analysis, Chuck made randomly distributed control sequences, and then compiled a list of distances between integration sites as you progressed along the chromosomes. He made a similar list for the experimental HIV integration sites and then compared the two. After normalizing the numbers of sequences, there were far more short distances between authentic HIV sites, allowing a statistically rigorous demonstration of clustering. The findings made biological sense—HIV infection kills off cells quickly, but integration in transcription units can help assure efficient viral gene expression during the short life of infected cells.

Later we collaborated with several groups to show that binding of HIV integrase to the cellular transcription protein LEDGF/p75 mediates targeting to transcription units via a simple tethering mechanism (Cherepanov *et al.*, 2003; Ciuffi *et al.*, 2005; Ciuffi and Bushman, 2006; Marshall *et al.*, 2007; Shun *et al.*, 2007). One experiment that made this point was the demonstration that hybrid tethers, containing the LEDGF integrase-binding domain fused to a novel chromatin-binding domain, could retarget HIV integration to new locations in chromosomes (Gijbsers *et al.*, 2010).

Gammaretroviruses favor integration in another location—transcription start sites (Wu *et al.*, 2003; Mitchell *et al.*, 2004)—and just this year a tethering factor was found for this reaction as well (cellular bromodomain and extra-terminal [BET] proteins) (Ciuffi and Bushman, 2006; De Rijck *et al.*, 2013; Gupta *et al.*, 2013; Sharma *et al.*, 2013). These findings may allow more control over integration site selection in gene therapy applications—small molecules are currently available that interfere with binding of integrases to both LEDGF/p75 and BET proteins, resulting in more random integration targeting in treated cells.

DNA Integration in Human Gene Therapy

The methods developed in mechanistic studies of retroviral integration were perfect for monitoring outcome in

human gene therapy, and happily we were invited to help analyze several influential clinical trials. Around this time I was fortunate to receive an offer from the University of Pennsylvania, and so moved there, lured in part by the chance to work with the distinguished community of gene therapy researchers, which includes Jean Bennett, Kathy High, Carl June, James Riley, Pablo Tebas, and Jim Wilson.

Shortly after arriving at Penn, we undertook the exciting job of applying the new deep sequencing methods to analysis of integration site distributions. The newly available 454/Roche pyrosequencing method was able to produce close to a million sequence reads in a single instrument run—technology so amazing one wondered if space aliens were secretly behind it. After several false starts, in 2006 we managed to sequence about 40,000 sites of HIV integration in a cell line, by far the largest data set at the time. This allowed much more detailed analysis of integration site positions than was previously possible and for the first time allowed association of integration site data with information on epigenetic modification of chromatin in target cells (Wang *et al.*, 2007).

Our first work with samples from gene-corrected subjects—started at the invitation of Marina Cavazzana-Calvo, Alain Fischer, and Salima Hacein-Bey-Abina at the Hôpital Necker-Enfants Malades in Paris—involved analyzing outcome in the first trial to treat severe combined immunodeficiency (SCID)-X1 (Cavazzana-Calvo *et al.*, 2000; Hacein-Bey-Abina *et al.*, 2003b, 2008; Thrasher *et al.*, 2006; Bushman, 2007; Howe *et al.*, 2008). Bone marrow cells were removed from the affected children, transduced using a gammaretroviral vector harboring an intact copy of the *IL2RG* gene, and then reinfused. Of 20 subjects treated in London and Paris, most achieved dramatic clinical benefit, but 5 of the children suffered adverse events in the form of leukemia associated with integration of the gammaretroviral vector near a cellular cancer gene. Unfortunately, it seemed quite likely that the transcriptional enhancer sequences in the therapeutic vector caused the activation of dangerous cellular genes such as *LMO2* (Hacein-Bey-Abina *et al.*, 2003a, 2008). One child died, the other four responded to chemotherapy and have continued to benefit from gene therapy. Despite the setbacks, the results were remarkable—for the first time in history, humans had stably engineered their genetic makeup to cure a disease.

We sequenced integration sites from longitudinal blood samples from the treated subjects and began a long-term analysis of the population biology of the gene-corrected cells (Wang *et al.*, 2008, 2010; Hacein-Bey-Abina *et al.*, 2010). The clinical setting is unique—the children initially had no T cells at all, but after gene correction, T cells grew out robustly and every cell was marked with a unique integration site. Initially, we helped characterize some of the adverse events, identifying *CCND2* and *BM11* as genes likely involved in insertional activation in addition to *LMO2*. We tracked cell clones longitudinally through gene correction, revealing that in some cases cells with integration sites near cancer-related genes were particularly abundant, potentially because of boosted transcription from the vector enhancer increasing expression of cellular genes favoring proliferation or survival. Another application was monitoring the outcome of chemotherapy for leukemia—we could determine which integration site near a cancer gene

shot up in abundance associated with adverse events and then confirm that the clone was erased from the population by successful chemotherapy.

After the first SCID-X1 trial, many groups turned to newer types of retroviral vectors that lacked the strong enhancers characteristic of gammaretroviruses, and we used integration site information to track cell clones in several of them. One type of vector maintained most of the gammaretroviral backbone that had worked in the SCID-X1 trial but simply deleted the enhancers. The SCID2 trial with the revised vector is now under way and publications should be appearing shortly.

Another direction was to adapt HIV as a vector for gene delivery. Despite the millions of people infected with HIV worldwide, each harboring enormous numbers of infected cells, there are no reported cases of cancer by insertional activation of proto-oncogenes. Proponents of HIV-based vectors (rebranded “lentiviral vectors” to make them sound less scary) suggest that this indicates that vectors also should be safer. However, HIV encodes a protein that arrests the cell cycle (Vpr), and another that is lethal to cells (Env), both of which are removed in vectors. Thus, the potential genotoxicity may be less clear than suggested by the HIV patient data.

The first introduction of lentiviral vectors into humans was carried out by a large team led by Carl June at the University of Pennsylvania (Levine *et al.*, 2006; Wang *et al.*, 2009). T cells were harvested from HIV-positive patients, a vector containing an HIV inhibitory gene was introduced on a lentiviral vector, and then the cells were reinfused. The procedure was found to be safe, the goal of the trial, and we showed that the distribution of lentiviral vector integration sites in the T cells was indistinguishable from the distribution in test infections of tissue culture cells in the lab—that is, there was no indication that vector integration was driving particular cells to proliferate differentially, good news for safety (Levine *et al.*, 2006; Wang *et al.*, 2009).

This first HIV trial showed some possible clinical benefit, though no strong effects on the amount of circulating HIV. However, two other groups shortly thereafter reported dramatic clinical successes with lentiviral vectors. Natalie Cartier, Patrick Auborg, and their coworkers reported successful use of lentiviral gene delivery to treat adrenoleukodystrophy (Cartier *et al.*, 2009), helping patients who had few other therapeutic options. In a second study, Phillipe Leboulch, Marina Cavazzana-Calvo, and a large team succeeded in using lentiviral vector-based gene delivery to treat one patient with beta-thalassemia, rendering the subject independent of the blood transfusions required before gene correction (Cavazzana-Calvo *et al.*, 2010).

We carried out a longitudinal integration site analysis in the beta-thalassemia subject, with surprising results (Cavazzana-Calvo *et al.*, 2010). A large fraction of all transduced cells contained an integration site in *HMGA2*, a known cancer gene. It turned out that integration within the gene resulted in formation of a truncated mRNA, which removed a negatively acting binding site for the let7 microRNA, thereby upregulating expression of *HMGA2*. The transcriptional apparatus of the vector also appeared to be acting on the *HMGA2* promoter, further boosting *HMGA2*

expression and apparently promoting cell proliferation. This provided the first indication of possible “vector driving” by lentiviral vectors in humans—today, the picture seems to be that lentiviral vectors are less likely to affect cellular proliferation than early version gammaretroviral vectors, but there is still some potential for doing so. Happily, at the time of this writing, the treated subject is doing well.

While these studies have yielded exciting new data, they have also left us quite familiar with the limitations of the analytical methods. Recovery methods are limited in the efficiency and reliability, and communicating effectively with clinical colleagues is an ongoing challenge. Going forward, we and others in the field need to optimize methods (Berry *et al.*, 2006, 2012; Hoffmann *et al.*, 2007; Bushman *et al.*, 2008; Wang *et al.*, 2008; Ciuffi *et al.*, 2009; Hacein-Bey-Abina *et al.*, 2010; Brady *et al.*, 2011) and then make them as high throughput and simple as possible for expanded implementation as human gene therapy comes into widespread use.

Looking Forward

Today, numerous successful gene therapy trials are underway. Lentiviral delivery is being applied to a wide array of disease indications (Levine *et al.*, 2006; Cartier *et al.*, 2009; Wang *et al.*, 2009; Cavazzana-Calvo *et al.*, 2010; Grupp *et al.*, 2013), and AAV-based vectors are also showing dramatic success (Maguire *et al.*, 2008; Li *et al.*, 2011a,b; Nathwani *et al.*, 2011; Kaeppl *et al.*, 2013). Cells modified by designer nucleases are now in humans—again Carl June and coworkers have pioneered a new technology in the search for cures for HIV infection (Tebas *et al.*, 2014). Genome editing is taking off as a correction method for many indications, though my guess is we still have a lot to learn about the unwanted side effects of engineered nucleases on genomes.

Whole new worlds are still opening up thanks to the new deep-sequencing methods. For example, gene correction in immunodeficient states can be explored by sequencing products of VDJ recombination (Robins *et al.*, 2009, 2010), providing an exceptionally rich picture of newly formed populations of antigen-recognizing cells. Deep sequencing can also report the response of the human microbiome as immunodeficiencies are corrected (McKenna *et al.*, 2008; Hildebrandt *et al.*, 2009; Hoffmann *et al.*, 2009; Wu *et al.*, 2011; Charlson *et al.*, 2012; Minot *et al.*, 2012, 2013; Peterfreund *et al.*, 2012; Sonnenberg *et al.*, 2012; Alenghat *et al.*, 2013), providing a unique opportunity to investigate how the human immune system shapes our microbial populations.

Without much fanfare, human gene therapy has gone from controversial to almost standard for some indications. It has been a privilege to be able to help along the way. Way down the road, I’m betting that this transition will be seen as a turning point in history.

Acknowledgments

I particularly thank the people who worked in my laboratory and the collaborators who added so much to our projects. Thanks very much to the National Institutes of Health, which has funded most of my research.

Author Disclosure Statement

The author has no conflict of interest.

References

- Alenghat, T., Osborne, L.C., Saenz, S.A., *et al.* (2013). Histone deacetylase 3 coordinates commensal-bacteria-dependent intestinal homeostasis. *Nature* 504, 153–157.
- Berry, C., Hannenhalli, S., Leipzig, J., and Bushman, F.D. (2006). Selection of target sites for mobile DNA integration in the human genome. *PLoS Comput. Biol.* 2, e157.
- Berry, C.C., Gillet, N.A., Melamed, A., *et al.* (2012). Estimating abundances of retroviral insertion sites from DNA fragment length data. *Bioinformatics* 28, 755–762.
- Berry, C.C., Ocwieja, K.E., Malani, N., and Bushman, F.D. (2014). Comparing DNA integration site clusters with scan statistics. *Bioinformatics* [Epub ahead of print]; doi: 10.1093/bioinformatics/btu035.
- Bor, Y.C., Bushman, F., and Orgel, L. (1995). In Vitro integration of human immunodeficiency virus type 1 cDNA into targets containing protein-induced bends. *Proc. Natl. Acad. Sci. USA* 92, 10334–10338.
- Brady, T., Lee, Y.N., Ronen, K., *et al.* (2009). Integration target site selection by a resurrected human endogenous retrovirus. *Genes Dev.* 23, 633–642.
- Brady, T., Roth, S.L., Malani, N., *et al.* (2011). A method to sequence and quantify DNA integration for monitoring outcome in gene therapy. *Nucleic Acids Res.* 39, e72.
- Bushman, F.D. (1994). Tethering human immunodeficiency virus 1 integrase to a DNA site directs integration to nearby sequences. *Proc. Natl. Acad. Sci. USA* 91, 9233–9237.
- Bushman, F.D. (2001). *Lateral DNA Transfer: Mechanisms and Consequences* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).
- Bushman, F.D. (2007). Retroviral integration and human gene therapy. *J. Clin. Invest.* 117, 2083–2086.
- Bushman, F.D., and Ptashne, M. (1986). Activation of transcription by the bacteriophage 434 repressor. *Proc. Natl. Acad. Sci. USA* 83, 9353–9357.
- Bushman, F.D., and Ptashne, M. (1988). Turning lambda Cro into a transcriptional activator. *Cell* 54, 191–197.
- Bushman, F.D., and Wang, B. (1994). Rous sarcoma virus integrase protein: mapping functions for catalysis and substrate binding. *J. Virol.* 68, 2215–2223.
- Bushman, F.D., Anderson, J.E., Harrison, S.C., and Ptashne, M. (1985). Ethylation interference and X-ray crystallography identify similar interactions between 434 repressor and operator. *Nature* 316, 651–653.
- Bushman, F.D., Shang, C., and Ptashne, M. (1989). A single glutamic acid residue plays a key role in the transcriptional activation function of lambda repressor. *Cell* 58, 1163–1171.
- Bushman, F.D., Fujiwara, T., and Craigie, R. (1990). Retroviral DNA integration directed by HIV integration protein *in vitro*. *Science* 249, 1555–1558.
- Bushman, F.D., Hoffmann, C., Ronen, K., *et al.* (2008). Massively parallel pyrosequencing in HIV research. *AIDS* 22, 1411–1415.
- Carteau, S., Gorelick, R., and Bushman, F.D. (1999). Coupled integration of human immunodeficiency virus cDNA ends by purified integrase *in vitro*: stimulation by the viral nucleocapsid protein. *J. Virol.* 73, 6670–6679.
- Cartier, N., Hacein-Bey-Abina, S., Bartholomae, C.C., *et al.* (2009). Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. *Science* 326, 818–823.

- Cavazzana-Calvo, M., Hacein-Bey, S., De Saint Basile, G., *et al.* (2000). Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science* 288, 669–672.
- Cavazzana-Calvo, M., Payen, E., Negre, O., *et al.* (2010). Transfusion independence and HMG2 activation after gene therapy of human beta-thalassaemia. *Nature* 467, 318–322.
- Charlson, E.S., Diamond, J.M., Bittinger, K., *et al.* (2012). Lung-enriched organisms and aberrant bacterial and fungal respiratory microbiota after lung transplant. *Am. J. Respir. Crit. Care Med.* 186, 536–545.
- Cherapanov, P., Maertens, G., Proost, P., *et al.* (2003). HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J. Biol. Chem.* 278, 372–381.
- Ciuffi, A., and Bushman, F.D. (2006). Retroviral DNA integration: HIV and the role of LEDGF/p75. *Trends Genet.* 22, 388–395.
- Ciuffi, A., Llano, M., Poeschla, E., *et al.* (2005). A role for LEDGF/p75 in targeting HIV DNA integration. *Nat. Med.* 11, 1287–1289.
- Ciuffi, A., Ronen, K., Brady, T., *et al.* (2009). Methods for integration site distribution analyses in animal cell genomes. *Methods* 47, 261–268.
- Craigie, R., and Mizuuchi, K. (1985). Mechanism of transposition of bacteriophage Mu: structure of a transposition intermediate. *Cell* 41, 867–876.
- Craigie, R., and Mizuuchi, K. (1987). Joining of Mu to target DNA can be uncoupled from cleavage at the end of Mu. *Cell* 51, 493–501.
- Craigie, R., Mizuuchi, M., and Mizuuchi, K. (1984). Site-specific recognition of the bacteriophage Mu ends by the Mu A protein. *Cell* 39, 387–394.
- Craigie, R., Arndt-Jovin, D.J., and Mizuuchi, K. (1985). A defined system for the DNA strand-transfer reaction at the initiation of bacteriophage Mu transposition: protein and DNA substrate requirements. *Proc. Natl. Acad. Sci. USA* 82, 7570–7574.
- Craigie, R., Fujiwara, T., and Bushman, F. (1990). The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration *in vitro*. *Cell* 62, 829–837.
- De Rijck, J., De Kogel, C., Demeulemeester, J., *et al.* (2013). The BET family of proteins targets moloney murine leukemia virus integration near transcription start sites. *Cell Rep.* 5, 886–894.
- Diamond, T.L., and Bushman, F.D. (2005). Division of labor within human immunodeficiency virus integrase complexes: determinants of catalysis and target DNA capture. *J. Virol.* 79, 15376–15387.
- Diamond, T.L., and Bushman, F.D. (2006). Role of metal ions in catalysis by HIV integrase analyzed using a quantitative PCR disintegration assay. *Nucleic Acids Res.* 34, 6116–6125.
- Engelman, A., Mizuuchi, K., and Craigie, R. (1991). HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell* 67, 1211–1221.
- Farnet, C., Lipford, R., Wang, B., and Bushman, F.D. (1996). Differential inhibition of HIV-1 preintegration complexes and purified integrase protein by small molecules. *Proc. Natl. Acad. Sci. USA* 93, 9742–9747.
- Gao, K., Wong, S., and Bushman, F. (2004). Metal binding by the D₁DX35E motif of human immunodeficiency virus type 1 integrase: selective rescue of Cys substitutions by Mn²⁺ *in vitro*. *J. Virol.* 78, 6715–6722.
- Gijsbers, R., Ronen, K., Vets, S., *et al.* (2010). LEDGF hybrids efficiently retarget lentiviral integration into heterochromatin. *Mol. Ther.* 18, 552–560.
- Grupp, S.A., Kalos, M., Barrett, D., *et al.* (2013). Chimeric antigen receptor-modified T cells for acute lymphoid leukemia. *N. Engl. J. Med.* 368, 1509–1518.
- Gupta, K., Diamond, T., Hwang, Y., *et al.* (2010). Structural properties of HIV integrase. Lens epithelium-derived growth factor oligomers. *J. Biol. Chem.* 285, 20303–20315.
- Gupta, S.S., Maetzig, T., Maertens, G.N., *et al.* (2013). Bromo- and extraterminal domain chromatin regulators serve as cofactors for murine leukemia virus integration. *J. Virol.* 87, 12721–12736.
- Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M., *et al.* (2003a). A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N. Engl. J. Med.* 348, 255–256.
- Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M., *et al.* (2003b). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* 302, 415–419.
- Hacein-Bey-Abina, S., Garrigue, A., Wang, G.P., *et al.* (2008). Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J. Clin. Invest.* 118, 3132–3142.
- Hacein-Bey-Abina, S., Hauer, J., Lim, A., *et al.* (2010). Efficacy of gene therapy for X-linked severe combined immunodeficiency. *N. Engl. J. Med.* 363, 355–364.
- Hildebrandt, M.A., Hoffmann, C., Sherrill-Mix, S.A., *et al.* (2009). High-fat diet determines the composition of the murine gut microbiome independently of obesity. *Gastroenterology* 137, 1716–1724.e1–2.
- Hoffmann, C., Minkah, N., Leipzig, J., *et al.* (2007). DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. *Nucleic Acids Res.* 35, e91.
- Hoffmann, C., Hill, D.A., Minkah, N., *et al.* (2009). Community-wide response of the gut microbiota to enteropathogenic *Citrobacter rodentium* infection revealed by deep sequencing. *Infect. Immun.* 77, 4668–4678.
- Howe, S.J., Mansour, M.R., Schwarzwaelder, K., *et al.* (2008). Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *J. Clin. Invest.* 118, 3143–3150.
- Kaeppl, C., Beattie, S.G., Fronza, R., *et al.* (2013). A largely random AAV integration profile after LPLD gene therapy. *Nat. Med.* 19, 889–891.
- Levine, B.L., Humeau, L.M., Boyer, J., *et al.* (2006). Gene transfer in humans using a conditionally replicating lentiviral vector. *Proc. Natl. Acad. Sci. USA* 103, 17372–17377.
- Li, H., Haurigot, V., Doyon, Y., *et al.* (2011a). *In vivo* genome editing restores haemostasis in a mouse model of haemophilia. *Nature* 475, 217–221.
- Li, H., Malani, N., Hamilton, S.R., *et al.* (2011b). Assessing the potential for AAV vector genotoxicity in a murine model. *Blood* 117, 3311–3319.
- Maguire, A.M., Simonelli, F., Pierce, E.A., *et al.* (2008). Safety and efficacy of gene transfer for Leber’s congenital amaurosis. *N. Engl. J. Med.* 358, 2240–2248.
- Marshall, H.M., Ronen, K., Berry, C., *et al.* (2007). Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. *Plos One* 2, e1340.
- McKenna, P., Hoffmann, C., Minkah, N., *et al.* (2008). The macaque gut microbiome in health, lentiviral infection, and chronic enterocolitis. *PLoS Pathogens* 4, e20.
- Miller, M.D., Wang, B., and Bushman, F.D. (1995). HIV-1 preintegration complexes containing discontinuous plus strands are competent to integrate *in vitro*. *J. Virol.* 69, 3938–3944.

- Minot, S., Grunberg, S., Wu, G.D., *et al.* (2012). Hypervariable loci in the human gut virome. *Proc. Natl. Acad. Sci. USA* 109, 3962–3966.
- Minot, S., Bryson, A., Chehoud, C., *et al.* (2013). Rapid evolution of the human gut virome. *Proc. Natl. Acad. Sci. USA* 110, 12450–12455.
- Mitchell, R.S., Beitzel, B.F., Schroder, A.R.W., *et al.* (2004). Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol.* 2, 1127–1137.
- Mizuuchi, K. (1983). *In vitro* transposition of bacteriophage Mu: a biochemical approach to a novel replication reaction. *Cell* 35, 785–794.
- Mizuuchi, K., and Adzuma, K. (1991). Inversion of the phosphate chirality at the target site of Mu DNA strand transfer: evidence for a one-step transesterification mechanism. *Cell* 66, 129–140.
- Nathwani, A.C., Tuddenham, E.G., Rangarajan, S., *et al.* (2011). Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. *N. Engl. J. Med.* 365, 2357–2365.
- Peterfreund, G.L., Vandivier, L.E., Sinha, R., *et al.* (2012). Succession in the gut microbiome following antibiotic and antibody therapies for clostridium difficile. *PLoS One* 7, e46966.
- Pruss, D., Bushman, F.D., and Wolffe, A.P. (1994a). Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. *Proc. Natl. Acad. Sci. USA* 91, 5913–5917.
- Pruss, D., Reeves, R., Bushman, F.D., and Wolffe, A.P. (1994b). The influence of DNA and nucleosome structure on integration events directed by HIV integrase. *J. Biol. Chem.* 269, 25031–25041.
- Ptashne, M. (1967a). Isolation of the lambda phage repressor. *Proc. Natl. Acad. Sci. USA* 57, 306–313.
- Ptashne, M. (1967b). Specific binding of the lambda phage repressor to lambda DNA. *Nature* 214, 232–234.
- Ptashne, M. (1986). *A Genetic Switch* (Cell and Blackwell Scientific Press, Cambridge, Massachusetts).
- Robins, H.S., Campregher, P.V., Srivastava, S.K., *et al.* (2009). Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. *Blood* 114, 4099–4107.
- Robins, H.S., Srivastava, S.K., Campregher, P.V., *et al.* (2010). Overlap and effective size of the human CD8+ T cell receptor repertoire. *Sci. Transl. Med.* 2, 47ra64.
- Schroder, A.R.W., Shinn, P., Chen, H.M., *et al.* (2002). HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* 110, 521–529.
- Sharma, A., Larue, R.C., Plumb, M.R., *et al.* (2013). BET proteins promote efficient murine leukemia virus integration at transcription start sites. *Proc. Natl. Acad. Sci. USA* 110, 12036–12041.
- Shun, M.C., Raghavendra, N.K., Vandegraaff, N., *et al.* (2007). LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes Dev.* 21, 1767–1778.
- Sonnenberg, G.F., Monticelli, L.A., Alenghat, T., *et al.* (2012). Innate lymphoid cells promote anatomical containment of lymphoid-resident commensal bacteria. *Science* 336, 1321–1325.
- Summa, V., Petrocchi, A., Bonelli, F., *et al.* (2008). Discovery of raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection. *J. Med. Chem.* 51, 5843–5855.
- Tebas, P., Stein, D., Tang, W.W., *et al.* (2014). Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N. Engl. J. Med.* 370, 901–910.
- Thrasher, A.J., Gaspar, H.B., Baum, C., *et al.* (2006). Gene therapy: X-SCID transgene leukaemogenicity. *Nature* 443, E5–E6; discussion E6–E7.
- Wang, G.P., Ciuffi, A., Leipzig, J., *et al.* (2007). HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res.* 17, 1186–1194.
- Wang, G.P., Garrigue, A., Ciuffi, A., *et al.* (2008). DNA bar coding and pyrosequencing to analyze adverse events in therapeutic gene transfer. *Nucleic Acids Res.* 36, e49.
- Wang, G.P., Levine, B.L., Binder, G.K., *et al.* (2009). Analysis of lentiviral vector integration in HIV+ study subjects receiving autologous infusions of gene modified CD4+ T cells. *Mol. Ther.* 17, 844–850.
- Wang, G.P., Berry, C.C., Malani, N., *et al.* (2010). Dynamics of gene-modified progenitor cells analyzed by tracking retroviral integration sites in a human SCID-X1 gene therapy trial. *Blood* 115, 4356–4366.
- Wu, X., Li, Y., Crise, B., and Burgess, S.M. (2003). Transcription start regions in the human genome are favored targets for MLV integration. *Science* 300, 1749–1751.
- Wu, G.D., Chen, J., Hoffmann, C., *et al.* (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science* 334, 105–108.

Address correspondence to:

Dr. Frederic D. Bushman

Department of Microbiology

University of Pennsylvania School of Medicine

3610 Hamilton Walk

Philadelphia, PA 19104-6076

E-mail: bushman@mail.med.upenn.edu