HUMAN GENE THERAPY 25:475–481 (June 2014)

DOI: 10.1089/hum.2014.2525

Pioneer Perspective

Vector Integration and Tumorigenesis

Christof von Kalle,[1,2] Annette Deichmann,[1,2] and Manfred Schmidt[1,2]

The Beginning

With some early molecular work in oncology and virology in Harald zur Hausen's lab in Heidelberg, Germany, a background in internal medicine and oncology (courtesy of learning to treat Hodgkin's disease from Volker Diehl, Cologne, Germany), I had developed a fascination with stem cell transplantation as a way to treat otherwise intractable diseases in humans. So I jumped at the opportunity to work on a gene therapy project in Rainer Storb's division at the Hutch in Seattle, together with Hans-Peter Kiem and the late Friedrich Schuening.

Gene therapy was all the rage of the day; many people convinced—or trying to convince others—that gene transfer would solve many of the major medical problems of humankind. My own view, under the influence of Rainer's sharp scientific wit and guidance, was that perhaps the stably integrating retrovirus vectors, biologically inert as we thought they were, could at least help decipher the structure of human hematopoiesis. At the very least, they could help to understand its regeneration after hematopoietic stem cell transplantation—that meant as soon as we would succeed making them enter human blood-forming stem cells in significant numbers.

I Did Not Do It

All of what will be discussed in the next few pages is the result of team science, working together, throwing ideas back and forth, and several years of frustrating but important tinkering interrupted by innovative spurts. Single-molecule precision is hard work, punishing work, as well as wet lab and very high-intensity dry lab work on the computer 24/7. It is their work and pride, not mine: Manfred, Cynthia, Annette, Ali, Anna, Anne, Christine, Eliana, Kerstin, Raffaele, Richard, Simone, Uwe, and many, many others.

The Behavior of Retroviral Vectors

Integration of retroviral vectors into the host genome results in a unique fusion sequence of the provirus with the genome sequence that will be passed on to all progeny cells. This vector–genome fusion sequence serves as a molecular marker of a single cell, allowing the determination of the number and clonal contribution of individual gene-modified cells (Scherdin *et al*., 1990; Collas *et al*., 1996). Such clonality analyses can be applied in preclinical and clinical gene transfer studies to investigate the physiology of hematopoiesis and the biology and pathogenesis of cancer (Dilloo *et al*., 1996; Kohn, 1997; Rosenberg *et al*., 1997). They allow conclusions about self-renewal capacity, differentiation capacity, and long-term activity of initially transduced cells (e.g., hematopoietic stem and progenitor cells) and the detection of vector-induced side effects (Guenechea *et al*., 2001; Lemischka and Jordan, 2001; Hacein-Bey-Abina *et al*., 2003a, 2008; Schmidt *et al*., 2003, 2005; Glimm *et al*., 2005; Ott *et al*., 2006; Howe *et al*., 2008a).

Admiring the purity and simplicity of the polymerase chain reaction (PCR) (Mullis *et al*., 1986) encouraged me to try and solve our little gene vector insertion problem at hand by hard thinking before venturing into broad experimentation. Some say PCR is just the right amount of molecular biology that medical doctors can still understand. More importantly, the considerations on the efficiency of polymerase, which is close to one (Saiki *et al*., 1988), reproducing a copy from each template at every cycle, made it instantly clear to me (an MD by training) that a process could be generated for sequencing unknown DNA segments adjacent to known sequences also with an efficiency of close to one, if the polymerase would be the only or the first enzyme in the order of steps of a strategy. None of the protocols like inverse PCR (Silver and Keerikatte, 1989) or ligation-mediated PCR (Pfeifer *et al*., 1989) or any other protocols to characterize unknown flanking DNA we knew had succeeded to address this problem, making them ineligible for the analysis of highly complex mixtures, low copy numbers, or large clonal diversity.

How Hypermutation of B Cells Can Influence Retrovirus Sequencing

Ralf Küppers (Essen, Germany) is a brilliant geneticist and immunologist from the school of Klaus Rajewski (Harvard, Boston), as straight as they come, and discussions with him are always enlightening and beautiful—all contents, no politics. Jan and his team were trying to figure out the role of immunoglobulin hypermutation in B cell immune responses, reasoning that capturing and sequencing open DNA strands from B cells should tell them when, where, and why B cell immunity readjusts to its proper targets.

[1]Nationales Centrum für Tumorerkrankungen, 69120 Heidelberg, Germany.
[2]Abteilung für Translationale Onkologie, Deutsches Krebsforschungszentrum, 69120 Heidelberg, Germany.

Trying to write a convincing grant proposal on insertion site sequencing, I discussed with him our desire to more directly sequence such events than by cloning. He made me aware of the work by Mueller and Wold (Mueller and Wold, 1989) using ligation of abundant oligonucleotides for cloning and direct sequencing of any unknown DNA strand.

My Little Eureka Moment

From the available literature and our own set of experiments, we had figured out that the efficiency of ligating strands of DNA in a reproducible order and fashion was indeed much worse than in the ideal world of polymerases, around 1% of available molecules or less. The efficiencies of combinations of DNA restriction digests, purifications, and then ligations were even much more catastrophic. One morning in 1997, while in the shower getting ready for my day job as a physician on the intensive care unit, I went over the different steps of obtaining an amplifiable stretch of DNA around an insertion site once more in my mind. It must be in there somewhere, it must be in there somewhere.... Formerly preoccupied with the idea of obtaining known DNA sequences and both ends of the site to be discovered, I realized, all of a sudden, that we could harness the superior efficiency of polymerases for writing not an exponential, but at least a linear, multiple number of copies for each insertion event prior to engaging less efficient steps in the procedure, offsetting the later losses early on. Let us worry about the second strand later, I thought. My skin was a bit wrinkled, and I had no time for breakfast, but the idea proved kind of useful.

Manfred Schmidt immediately figured out that, all things combined, it would take 50–200 cycles of linear amplification to generate a number of copies that would bring the likelihood of discovering each individual event into the vicinity of one. The advantage of that scenario was that we could now prepare a second strand, restriction digest, and ligation for a much more robust double-stranded technology: the idea of linear amplification-mediated (LAM) PCR was born. It allowed the precise characterization of proviral insertion flanks directly from patient samples down to the single-cell level (Schmidt *et al*., 2002, 2007). This method has also allowed for the first time the comprehensive determination of the clonality in hematopoietic repopulation after transplantation in humans, helping, as we hoped, to render gene therapy clinically feasible, safe, and efficient.

Trashed, Scooped, and Still Happy...

It still took a couple of man-years of experimentation to figure out all the parameters to really make it work as of then. To make things worse, the department we were working for had fallen on hard times; our funding threatened to go away completely. In addition, we of course had not invented the principle of insertion site analysis in gene transfer as such. Jan Nolta's extremely elegant article in *PNAS* showed the feasibility of clonality analysis in a xenotransplantation experiment (Nolta *et al*., 1996).

We believed in the LAM-PCR idea, with the potential of the precise characterization of proviral insertion flanks directly from patient samples down to the single-cell level. In our first experiments, we observed promising results: the competitiveness of the amplification reaction in polyclonal samples seemed to allow a fairly quantitative analysis of clone copy numbers as long as the fragment size was in the same ballpark. More information must be in there somewhere, we figured, and started collaborating with Cynthia Dunbar at the NIH and Hans-Peter Kiem at the Hutch, who had the most advanced large animal models for gene transfer. With these analyses it could be shown that hematopoiesis in nonhuman primate gene-marking models (rhesus monkey and baboon monkey) is polyclonal and driven by long-term active clones. It has further been demonstrated that the initial transduction occurred also in primitive hematopoietic progenitor and stem cells (Schmidt *et al*., 2002). This method allowed the comprehensive determination of the clonality in hematopoietic repopulation after transplantation in nonhuman primates. The long-term follow-up suggested that clonal activity *in vivo* is influenced to some extent by the retroviral insertion (Schmidt *et al*., 2002; Hematti *et al*., 2004; Calmels *et al*., 2005).

Furthermore, we examined the clonality of gene-corrected cells in Don Kohn's first successful admirable clinical gene therapy trial treating adenosine deaminase deficiency (ADA-SCID). The treatment had worked on a percentage of the patients' cells long term but did not result in complete correction of the genetic defect (Kohn *et al*., 1998). With LAM-PCR it could be shown that the lymphopoietic regeneration in both patients remained stable but mono- to oligoclonal. In contrast to later trials in X-SCID, no malignant monoclonal cell proliferation could be observed. T cell receptor analyses could demonstrate that most of the gene-modified cells originate from a prethymic stem or progenitor cell, and that the LAM-PCR works also on human peripheral blood leukocytes: It is in there (Schmidt *et al*., 2003)!

In 2000 the first therapeutically successful clinical gene therapy was carried out on SCID-X1 patients—a great success for Marina Cavazzana-Calvo and Alain Fischer from the Hôpital Necker in Paris (Cavazzana-Calvo *et al*., 2000) and for the whole field of gene therapy. Shortly thereafter, Adrian Thrasher and Bobby Gaspar from the University College of London reached the same success in their SCID-X1 trial (Gaspar *et al*., 2004). Insertion site data provided the first molecular evidence that gene-corrected CD34+ cells with both lymphomyeloid potential and self-renewal capacity are involved and can maintain blood formation (Schmidt *et al*., 2005).

Sailing the Integrome with a Genomic Map

Based on the advent and public availability of the almost complete murine and human genome sequence in 2002 and 2003, the characteristics of insertion of viral vectors have been extensively reevaluated (Schroder *et al*., 2002; Waterston *et al*., 2002; Wu *et al*., 2003). Frederic Bushman and his group have put a lot of effort into this, especially regarding the exploration of the wild-type virus insertion patterns. These studies have not only refuted the theory that retroviral integration is a more or less random event but also identified significant integration preferences for different retroviruses. For the human immunodeficiency virus type 1 (HIV-1) and derived vectors, it has been shown that gene-coding regions

are preferred integration sites (IS): about two-thirds of all integrates were uniformly distributed within gene regions (Schroder *et al*., 2002; Mitchell *et al*., 2004). In contrast, oncoretroviral vectors derived from murine leukemia virus (MLV) or spleen focus-forming virus showed a significant accumulation of IS around the transcription start site of a gene and in gene regulatory CpG islands (Wu *et al*., 2003; Mitchell *et al*., 2004). This integration preference of MLV is found also in human hematopoietic cells in the xenograft model and in clinical trials (Laufs *et al*., 2003; Glimm *et al*., 2005; Ott *et al*., 2006; Recchia *et al*., 2006; Aiuti *et al*., 2007; Cattoglio *et al*., 2007; Deichmann *et al*., 2007; Schwarzwaelder *et al*., 2007). However, the oncoretroviral vector ASLV, which derives from avian sarcoma leukosis virus, showed a random, uniform distribution over the entire genome (Mitchell *et al*., 2004).

One thing that dawned on us in the early days of LAM-PCR, however, was not to buy too deeply into the details on biological inertness of retroviral vectors. I had initially sought to save a copy of every article on gene therapy that was published but had rapidly given up on that task in favor of electronic searches that had recently become available. Early on, I had stumbled on the wide body of literature published by expert retrovirologists describing more or less each brand of insertional virus as an ideal tool for inducing and studying mutagenesis in living cells. At the time, the clinical gene therapy field still was in a bit of denial on this particular subject.

Be that as it may, we reasoned, to obtain the exact location where insertional vectors integrated should be done in a large-scale analysis also in patient samples, being highly informative both for the purpose of barcoding individual stem cell progeny themselves and to understand the mutational side effects of genetic transfer, should any develop.

At that point, in June 2002, one of our wonderful, spirited discussions and collaborations with Christopher Baum and his team brought about the startling discovery that a single retrovirus insertion can initiate a sequence of development leading to malignant clonal outgrowth.

The End of the Beginning

Unfortunately, and completely unfairly for Alain and his team, the beautiful X-SCID trial brought a rude awakening, demonstrating the influence of insertional mutagenesis caused by the vector insertion. Four of the nine patients developed T-cell leukemia because of vector insertion (Hacein-Bey-Abina *et al*., 2003a). The vector used in this trial was based on a gamma-retroviral MLV vector containing complete long terminal repeats (LTRs). These LTR regions contained a powerful enhancer that was able to activate the neighboring proto-oncogene *LMO2* (Hacein-Bey-Abina *et al*., 2003b).

The pluripotency and long-term activity of individual gene-modified clones can be monitored by a clone-specific tracking PCR. In conjunction with provirus-specific primers from the LTR region, the detected genomic insertion sequence serves as a template for the production of primers. The quantitative assessment of the contribution of individual clones within the examined material can be determined by simultaneous coamplification (qPCR) of an artificial internal standard that has shortened by a few base pairs.

From 2002 on, depth and precision of insertion site studies increased steadily, showing that a significant correlation between transcriptional activity and HIV-1 and MLV vector insertion exists (Mitchell *et al*., 2004; Deichmann *et al*., 2007; Schwarzwaelder *et al*., 2007). It has further been demonstrated that insertion of MLV vectors occurs significantly more frequently in common fragile sites (Bester *et al*., 2006) that are correlated with chromosomal breakpoints in tumors (Popescu, 2003). The insertion process of retroviruses and their derived vectors is obviously not only influenced by the accessibility of the euchromatin. Rather, there must be complex mechanisms that include active binding of the preinsertion complex to cellular DNA motifs or DNA-binding proteins (Bushman, 2003). For the insertion of lentiviral vectors, it has been shown that lens epithelium-derived growth factor (LEDGF/p75) is responsible for the insertion in actively transcribed gene regions (Ciuffi *et al*., 2005). Recently, bromodomain and extraterminal (BET) proteins have been validated as MLV insertion targeting factors (De Rijck *et al*., 2013). Wang *et al.* could demonstrate a clustering of gamma-retroviral IS in enhancer and promoter areas associated with posttranslational histone modifications and RNA polymerase II binding (Wang *et al*., 2010).

Fine? Refine, Find, Define!

LAM-PCR development still had and has further life in it. The recognition motifs for the enzymes used are not uniformly distributed throughout the genome; a proportion of insertion sites in transduced samples could not be recognized by insertion site analysis using LAM-PCR. Too large of a distance (>1 kb) to the nearest restriction site leads to inefficient amplification of the PCR products; very short distances (<20–30 bp) generate, on the other hand, fusion sequences that are too short to be aligned correctly to the reference genome. To maximize the accessibility of genomic insertion sites, we therefore went on to model LAM on genome-wide restriction sites as well as doing away with restriction enzymes altogether (Gabriel *et al*., 2009).

The establishment and the use of high-throughput next-generation sequencing technologies combined with the individual marking of nonrestrictive (nr) LAM-PCR products by barcoding enabled the simultaneous sequencing of a large number of different PCR products (Paruzynski *et al*., 2010). The number of obtained sequence reads of an individual cell clone can be used to estimate its individual clonal contribution.

Throughout these first ''big data'' years, David Williams has been our wonderful mentor and staunch supporter, helping us to accomplish scientifically and clinically meaningful output and getting things organized as well.

One Log More Every Season

The advent of the information from the human genome project on the web was an incredible boon to our field. Even though in the beginning the information was not very precise (e.g., we had to readjust the coordinates of the *LMO2* gene in the available release by realigning Thomas Boehms original exome sequences), the human and murine genome projects had thrown on the light switch for gene therapy's

little needle in a haystack project in a very big way. Even more, the availability of high-throughput devices for sequencing had allowed us to process larger and larger amounts of materials and information, arriving from a small random sampling to a more complete repertoire analysis of many, eventually all insertion events present in a system.

This will provide new possibilities to detect the clonal repertoire of a sample. Because of the currently decreasing sequencing costs, the enrichment of the proviral vector sequences combined with subsequent direct sequencing of the DNA is manageable (Okou *et al*., 2007; Porreca *et al*., 2007). The first promising experiments have been carried out also in our laboratory. We enriched the integrated vector sequences with custom-designed baits and were able to detect the insertion sites by direct genome sequencing on a third-generation sequencing platform (unpublished results). In the future, whole-genome sequencing for the detection of clonality will play a crucial role.

With the exponential increase of sequencing data, the need for fast and reliable analysis pipelines was high and depended upon strong bioinformatics tools. We and others have therefore developed analysis pipelines such as HISAP (Arens *et al*., 2012), QuickMap (Appelt *et al*., 2009), or SeqMap (Peters *et al*., 2008) for the characterization of insertion sites. These pipelines allow for trimming of sequences, clustering to reduce redundancy, readout of read counts, and alignment to a reference genome to obtain the precise genomic location.

Really brilliant people (not all MDs then, probably) by now had devised wonderful strategies that allow the statistical comparison of the experimental insertion site profile with the expected profile, considering the typical gamma- and lentiviral distribution of insertion sites (De Ridder *et al*., 2006; Abel *et al*., 2007, 2011; Berry *et al*., 2012). Bioinformatics is very important for a comprehensive analysis of such a large number of sequencing data and often the bottleneck for a fast analysis of the data.

The monitoring of vector insertions has been expanded also to other vector types such as adeno-associated vectors (AAV). We performed such analysis on a clinical trial for the treatment of lipoprotein lipase deficiency (LPLD). This clinical trial led to the first approval and marketing authorization of a gene therapy vector (Glybera) in the Western world by the European Medicines Agency. As expected, the majority of the vector-fusion sequences represented episomal concatemeric AAV rearrangements and showed large deletions of the vector-inverted terminal repeat sequences (Nakai *et al*., 2003; Inagaki *et al*., 2007; Li *et al*., 2011; Nowrouzi *et al*., 2012). For the analysis of the AAV concatemers, we designed a specific analysis pipeline that allows the identification of unique genomic insertion sites. In contrast to other insertion studies, the insertions were distributed throughout the genome with no preference for gene regions or other regulatory regions. But we could identify AAV insertion hotspots within the mitochondrial genome in samples with intramuscular injection but not in samples with intravenous injection of the vector (Kaeppel *et al*., 2013). This finding might be useful for the treatment of mitochondria-based disorders by AAV-mediated therapies (Yu *et al*., 2012).

On the wet lab side, we have generated amplicon sequencing protocols for all types of high-throughput machinery we could lay our hands on. The reproducibility of these systems has allowed us and the field to use random sizing of DNA samples, getting away from the artifacts of restriction enzyme digestion and quantification by electrophoresis. These advances together with one-touch pipelines for the analysis of fragments have led to consistent growth in our capacity and precision to analyze and understand data and insertional repertoires. Rapidly, these repertoire analyses will encompass the complete vector and the larger vicinity of each event. Literally, over the last decade the ability to study insertion has increased by more than a log for each season, much to the dismay of some of our younger lab members. ''Considering that we are analyzing more reads in one afternoon now than in the first five years of doing this,'' calculated one of our team members recently, ''we could have spent the first six years of this project on the beach, with very little difference . . . . '' Or so it seems.

Translation into the Clinic

All these activities have allowed a comprehensive analysis of the clonal inventory in human gene therapy trials starting from our early days in 2003 up to now. Severe side effects played a major role in this history. This analysis has matured gene therapy with all types of vectors, making it better. In 2002, the first proof of vector-mediated malignant transformation was observed in a mouse model, with LAM-PCR results confirming this finding (Li *et al*., 2002). In 2003, we could dissect the association between LMO2 integration and associated leukemias in the two first patients of the successful SCID-X1 trial (Hacein-Bey-Abina *et al*., 2003a,b). Two other patients were followed in this trial, one patient in the London trial. Fortunately, in four of the five patients the leukemias could be treated successfully. To know where and why a vector goes is, in essence, the equivalent of classical pharmacodynamics in gene therapy studies. This is now being performed in many preclinical and clinical gene therapy trials to monitor *in vivo* the safety of existing and new vectors as recommended by the national regulatory authorities worldwide.

Because of the severe side effects in the first successful SCID-X1 study, intensive studies of vector insertion sites were carried out for other trials. The overall aim was to improve the safety of gene therapy and to monitor the clonality of gene-corrected cells in the treated patients to make the therapy as safe as possible.

The activation of proto-oncogenes by gamma-retroviral vectors containing a full LTR region was also observed in a second clinical trial treating SCID-X1 patients (Howe *et al*., 2008b) in a gene therapy trial for the treatment of chronic granulomatous disease (Ott *et al*., 2006; Stein *et al*., 2010) and for the treatment of Wiskott–Aldrich syndrome (Boztug *et al*., 2010; Avedillo Díez *et al*., 2011; Braun *et al*., 2014).

Several other clinical trials used lentiviral vectors for gene transfer. The first clinical trial using HIV-1-based vector systems was applied for the treatment of X-linked adrenoleukodystrophy, conducted by Nathalie Cartier and Patrick Aubourg in Paris. The identification of identical IS in the myeloid and lymphoid cell fraction could prove the initial transduction of early progenitor cells. Our functional insertion site analysis, the comparison of pre- and post-transplant samples, and the analysis of preferred insertion loci did not show any signs of a potential vector-induced

expansion of gene-corrected cell clones. This and the missing *in vivo* expansion of gene-corrected cells make lentiviral vectors promising vehicles for various transgenes (Cartier *et al*., 2009).

Current clinical protocols for the treatment of Wiskott-Aldrich syndrome or metachromatic leukodystrophy use a self-inactivating (SIN)-lentiviral HIV-derived vector to overcome the problem of insertional mutagenesis as has been successfully translated in San Raffaele, Milan. This vector lacks a strong enhancer in the LTR region; instead, it contains a weak internal promoter that allows the transcription of the transgene (Aiuti *et al*., 2013; Biffi *et al*., 2013). These groundbreaking vector safety improvements were developed in the 1990s in Inder Verma's lab at the Salk Institute by Didier Trono and Luigi Naldini. The following years have shown that lentiviral and retroviral SIN vectors are really much more inert to insertional mutagenesis effects than the first generation of gamma-retroviral vectors.

A fascinating possibility is site-directed gene therapy mediated by designer nucleases such as zinc finger nucleases (ZFNs), transcription activator-like effector nucleases, or the Crisp/Cas9 system. Our method for insertion site analysis is able to identify the activity of such designer nucleases not only for on-target but also for off-target insertion sites. Together with Luigi's group, we have performed a comprehensive analysis of the off-target activity of ZFNs for the human genes *CCR5* and *IL2RG* for the first time (Gabriel *et al*., 2011). The precision of this determination of off-target insertion events is very important for the further development of this technology.

Outlook

With the implementation of more and more efficient next-generation sequencing technologies, hundreds of thousands of events are now visualized in preclinical and clinical clonal repertoire studies. A sophisticated analysis of this vast amount of accessible sequence data requires excellent practical and bioinformatical tools and skills. As in former days, the retrieval of the relevant biological (and clinical) information about vector-mediated influences on proliferation and selection of gene-corrected cells has remained the same.

I may hypothesize that we can learn from the ongoing cancer omics studies that currently revolutionize our knowledge of the molecular mechanisms driving primary cells into malignant transformation. The steadily growing increase of sequence output by minimizing costs renders it indispensable that insertion site sequencing will be accomplished in the near future by direct-sequencing approaches, avoiding excessive PCR steps prior to sequencing. Targeted vector genome and insertion sequencing as known from current all-exome sequencing studies should allow comprehensive and reliable quantitative measurements on the diversity of the analyzed insertion site pool, individual clonal contributions, and even vector genome stability.

At the beginning of the gene therapy era, this new form of treatment suffered a lot of setbacks, starting with ineffective outcomes, and some serious adverse events caused by insertional mutagenesis. All the more amazing and enlightening it is how much progress has been made over the last decade—starting with the exponential improvement of sequencing technologies accompanied by decreasing costs, the strong development of bioinformatical tools, and, of course, the fascinosum of targeted editing. And it is great to see what potential all these efforts carry for patients with otherwise incurable diseases. I would never have dreamed of such an acceleration of development.

Acknowledgments

We do apologize that we could not mention all individual collaborators and investigators with whom and for whom we have worked for the benefit of patients. This work would not have been possible without the sustained contribution of all former and present group members, the exciting inspirations given by internal and external collaborators, and the invaluable support of the gene therapy community as well as the national and international funding agencies.

Author Disclosure Statement

No competing financial interests exist.

References

Abel, U., Deichmann, A., Bartholomae, C., *et al.* (2007). Real-time definition of non-randomness in the distribution of genomic events. PLoS One 2, e570.

Abel, U., Deichmann, A., Nowrouzi, A., *et al.* (2011). Analyzing the number of common integration sites of viral vectors—new methods and computer programs. PLoS One 6, e24247.

Aiuti, A., Cassani, B., Andolfi, G., *et al.* (2007). Multilineage hematopoietic reconstitution without clonal selection in ADA-SCID patients treated with stem cell gene therapy. J. Clin. Invest. 117, 2233–2240.

Aiuti, A., Biasco, L., Scaramuzza, S., *et al.* (2013). Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott-Aldrich syndrome. Science 341, 1233151.

Appelt, J.U., Giordano, F.A., Ecker, M., *et al.* (2009). QuickMap: a public tool for large-scale gene therapy vector insertion site mapping and analysis. Gene Ther. 16, 885–893.

Arens, A., Appelt, J.U., Bartholomae, C., *et al.* (2012). Bioinformatic clonality analysis of next-generation sequencing-derived viral vector integration sites. Hum. Gene Ther. Methods 23, 111–118.

Avedillo Díez, I., Zychlinski, D., Coci, E.G., *et al.* (2011). Development of novel efficient SIN vectors with improved safety features for Wiskott–Aldrich syndrome stem cell based gene therapy. Mol. Pharm. 8, 1525–1537.

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.

Bester, A.C., Schwartz, M., Schmidt, M., *et al.* (2006). Fragile sites are preferential targets for integrations of MLV vectors in gene therapy. Gene Ther. 13, 1057–1059.

Biffi, A., Montini, E., Lorioli, L., *et al.* (2013). Lentiviral hematopoietic stem cell gene therapy benefits metachromatic leukodystrophy. Science 341, 1233158.

Boztug, K., Schmidt, M., Schwarzer, A., *et al.* (2010). Stem-cell gene therapy for the Wiskott–Aldrich syndrome. N. Engl. J. Med. 363, 1918–1927.

Braun, C.J., Boztug, K., Paruzynski, A., *et al.* (2014). Gene therapy for Wiskott-Aldrich syndrome—long-term efficacy and genotoxicity. Sci. Transl. Med. 6, 227ra33.

Bushman, F.D. (2003). Targeting survival: integration site selection by retroviruses and LTR-retrotransposons. Cell 115, 135–138.

Calmels, B., Ferguson, C., Laukkanen, M.O., *et al.* (2005). Recurrent retroviral vector integration at the Mds1/Evi1 locus in nonhuman primate hematopoietic cells. Blood 106, 2530–2533.

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.

Cattoglio, C., Facchini, G., Sartori, D., *et al.* (2007). Hot spots of retroviral integration in human CD34+ hematopoietic cells. Blood 110, 1770–1778.

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.

Ciuffi, A., Llano, M., Poeschla, E., *et al.* (2005). A role for LEDGF/p75 in targeting HIV DNA integration. Nat. Med. 11, 1287–1289.

Collas, P., Husebye, H., and Alestrom, P. (1996). The nuclear localization sequence of the SV40 T antigen promotes transgene uptake and expression in zebrafish embryo nuclei. Transgenic Res. 5, 451–458.

Deichmann, A., Hacein-Bey-Abina, S., Schmidt, M., *et al.* (2007). Vector integration is nonrandom and clustered and influences the fate of lymphopoiesis in SCID-X1 gene therapy. J. Clin. Invest. 117, 2225–2232.

De Ridder, J., Uren, A., Kool, J., *et al.* (2006). Detecting statistically significant common insertion sites in retroviral insertional mutagenesis screens. PLoS Comput. Biol. 2, e166.

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.

Dilloo, D., Rill, D.R., Grossmann, M.E., *et al.* (1996). Gene marking and gene therapy for transplantation medicine. J. Hematother. 5, 553–555.

Gabriel, R., Eckenberg, R., Paruzynski, A., *et al.* (2009). Comprehensive genomic access to vector integration in clinical gene therapy. Nat. Med. 15, 1431–1436.

Gabriel, R., Lombardo, A., Arens, A., *et al.* (2011). An unbiased genome-wide analysis of zinc-finger nuclease specificity. Nat. Biotechnol. 29, 816–823.

Gaspar, H.B., Parsley, K.L., Howe, S., *et al.* (2004). Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector. Lancet 364, 2181–2187.

Glimm, H., Schmidt, M., Fischer, M., *et al.* (2005). Efficient marking of human cells with rapid but transient repopulating activity in autografted recipients. Blood 106, 893–898.

Guenechea, G., Gan, O.I., Dorrell, C., and Dick, J.E. (2001). Distinct classes of human stem cells that differ in proliferative and self-renewal potential. Nat. Immunol. 2, 75–82.

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.

Hematti, P., Hong, B.K., Ferguson, C., *et al.* (2004). Distinct genomic integration of MLV and SIV vectors in primate hematopoietic stem and progenitor cells. PLoS Biol. 2, e423.

Howe, S.J., Mansour, M.R., Schwarzwaelder, K., *et al.* (2008a). Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. J. Clin. Invest. 118, 3143–3150.

Inagaki, K., Lewis, S.M., Wu, X., *et al.* (2007). DNA palindromes with a modest arm length of greater, similar 20 base pairs are a significant target for recombinant adeno-associated virus vector integration in the liver, muscles, and heart in mice. J. Virol. 81, 11290–11303.

Kaeppel, C., Beattie, S.G., Fronza, R., *et al.* (2013). A largely random AAV integration profile after LPLD gene therapy. Nat. Med. 19, 889–891.

Kohn, D.B. (1997). Gene therapy for haematopoietic and lymphoid disorders. Clin. Exp. Immunol. 107 Suppl 1, 54–57.

Kohn, D.B., Hershfield, M.S., Carbonaro, D., *et al.* (1998). T lymphocytes with a normal ADA gene accumulate after transplantation of transduced autologous umbilical cord blood CD34+ cells in ADA-deficient SCID neonates. Nat. Med. 4, 775–780.

Laufs, S., Gentner, B., Nagy, K.Z., *et al.* (2003). Retroviral vector integration occurs in preferred genomic targets of human bone marrow-repopulating cells. Blood 101, 2191–2198.

Lemischka, I.R., and Jordan, C.T. (2001). The return of clonal marking sheds new light on human hematopoietic stem cells. Nat. Immunol. 2, 11–12.

Li, Z., Duellmann, F., Schiedlmeier, B., *et al.* (2002). Murine leukemia induced by retroviral gene marking. Science 296, 497.

Li, H., Malani, N., Hamilton, S.R., *et al.* (2011). Assessing the potential for AAV vector genotoxicity in a murine model. Blood 117, 3311–3319.

Mitchell, R.S., Beitzel, B.F., Schroder, A.R., *et al.* (2004). Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol. 2, E234.

Mueller, P.R., and Wold, B. (1989). *In vivo* footprinting of a muscle specific enhancer by ligation mediated PCR. Science 246, 780–786.

Mullis, K., Faloona, F., Scharf, S., *et al.* (1986). Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. Cold Spring Harb. Symp. Quant. Biol. 51 Pt 1, 263–273.

Nakai, H., Montini, E., Fuess, S., *et al.* (2003). AAV serotype 2 vectors preferentially integrate into active genes in mice. Nat. Genet. 34, 297–302.

Nolta, J.A., Dao, M.A., Wells, S., *et al.* (1996). Transduction of pluripotent human hematopoietic stem cells demonstrated by clonal analysis after engraftment in immune-deficient mice. Proc. Natl. Acad. Sci. USA 93, 2414–2419.

Nowrouzi, A., Penaud-Budloo, M., Kaeppel, C., *et al.* (2012). Integration frequency and intermolecular recombination of rAAV vectors in non-human primate skeletal muscle and liver. Mol. Ther. 20, 1177–1186.

Okou, D.T., Steinberg, K.M., Middle, C., *et al.* (2007). Microarray-based genomic selection for high-throughput resequencing. Nat. Methods 4, 907–909.

Ott, M.G., Schmidt, M., Schwarzwaelder, K., *et al.* (2006). Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. Nat. Med. 12, 401–409.

Paruzynski, A., Arens, A., Gabriel, R., *et al.* (2010). Genome-wide high-throughput integrome analyses by nrLAM-PCR and next-generation sequencing. Nat. Protoc. 5, 1379–1395.

Peters, B., Dirscherl, S., Dantzer, J., *et al.* (2008). Automated analysis of viral integration sites in gene therapy research using the SeqMap web resource. Gene Ther. 15, 1294–1298.

Pfeifer, G.P., Steigerwald, S.D., Mueller, P.R., *et al.* (1989). Genomic sequencing and methylation analysis by ligation mediated PCR. Science 246, 810–813.

Popescu, N.C. (2003). Genetic alterations in cancer as a result of breakage at fragile sites. Cancer Lett. 192, 1–17.

Porreca, G.J., Zhang, K., Li, J.B., *et al.* (2007). Multiplex amplification of large sets of human exons. Nat. Methods 4, 931–936.

Recchia, A., Bonini, C., Magnani, Z., *et al.* (2006). Retroviral vector integration deregulates gene expression but has no consequence on the biology and function of transplanted T cells. Proc. Natl. Acad. Sci. USA 103, 1457–1462.

Rosenberg, S.A., Blaese, R.M., Brenner, M.K., *et al.* (1997). Human gene marker/therapy clinical protocols. Hum. Gene Ther. 8, 2301–2338.

Saiki, R.K., Gelfand, D.H., Stoffel, S., *et al.* (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239, 487–491.

Scherdin, U., Rhodes, K., and Breindl, M. (1990). Transcriptionally active genome regions are preferred targets for retrovirus integration. J. Virol. 64, 907–912.

Schmidt, M., Zickler, P., Hoffmann, G., *et al.* (2002). Polyclonal long-term repopulating stem cell clones in a primate model. Blood 100, 2737–2743.

Schmidt, M., Carbonaro, D.A., Speckmann, C., *et al.* (2003). Clonality analysis after retroviral-mediated gene transfer to CD34+ cells from the cord blood of ADA-deficient SCID neonates. Nat. Med. 9, 463–468.

Schmidt, M., Hacein-Bey-Abina, S., Wissler, M., *et al.* (2005). Clonal evidence for the transduction of CD34+ cells with lymphomyeloid differentiation potential and self-renewal capacity in the SCID-X1 gene therapy trial. Blood 105, 2699–2706.

Schmidt, M., Schwarzwaelder, K., Bartholomae, C., *et al.* (2007). High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). Nat. Methods 4, 1051–1057.

Schroder, A.R., Shinn, P., Chen, H., *et al.* (2002). HIV-1 integration in the human genome favors active genes and local hotspots. Cell 110, 521–529.

Schwarzwaelder, K., Howe, S.J., Schmidt, M., *et al.* (2007). Gammaretrovirus-mediated correction of SCID-X1 is associated with skewed vector integration site distribution *in vivo*. J. Clin. Invest. 117, 2241–2249.

Silver, J., and Keerikatte, V. (1989). Novel use of polymerase chain reaction to amplify cellular DNA adjacent to an integrated provirus. J. Virol. 63, 1924–1928.

Stein, S., Ott, M.G., Schultze-Strasser, S., *et al.* (2010). Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease. Nat. Med. 16, 198–204.

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.

Waterston, R.H., Lander, E.S., and Sulston, J.E. (2002). On the sequencing of the human genome. Proc. Natl. Acad. Sci. USA 99, 3712–3716.

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.

Yu, H., Koilkonda, R.D., Chou, T.H., *et al.* (2012). Gene delivery to mitochondria by targeting modified adenoassociated virus suppresses Leber's hereditary optic neuropathy in a mouse model. Proc. Natl. Acad. Sci. USA 109, E1238–E1247.

Address correspondence to:
Prof. Dr. Christof von Kalle
National Center for Tumor Diseases (NCT)
and German Research Cancer Center (DKFZ)
Im Neuenheimer Feld 460
69120 Heidelberg
Germany

E-mail: christof.kalle@nct-heidelberg.de