Gene Therapy (2008) 15, 143–153

www.nature.com/gt

REVIEW

Insertional mutagenesis and clonal dominance: biological and statistical considerations

B Fehse[1,2] and I Roeder[3]

[1]*Clinic for Stem Cell Transplantation, University Medical Centre, Hamburg, Germany;* [2]*Experimental Paediatric Haematology and Oncology, Johann Wolfgang Goethe University, Frankfurt, Germany and* [3]*Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany*

Improvements of (retroviral) gene transfer vectors, stem cell isolation and culture techniques as well as transduction protocols eventually resulted not only in the successful genetic modification of cells capable of reconstituting the haematopoietic system in various animal models, but also human beings. This was a conditio sine qua non for the successful application of gene therapy for inherited diseases as meanwhile achieved for severe combined immune deficiencies (SCID-X1, ADA-SCID) and chronic granulomatous disease (CGD). Unexpectedly, in long-term animal experiments as well as in the follow up of patients from the CGD trial, haematopoietic clones bearing insertions in certain gene loci became dominant, which was most apparent in the myeloid blood compartment. Accumulating data strongly suggest that this clonal dominance was due to some growth and/or survival advantage conferred by gene-activating or -suppressing effects of the integrated retroviral vector (insertional mutagenesis). Importantly, such induced clonal dominance seems not to lead to malignant transformation of affected cell clones inadvertently. The latter finding has become the basis for the concept of 'induced haematopoietic stem cells', a potentially powerful tool to investigate genes involved in the regulation of mechanisms underlying competitive advantages of stem cells, but also in the multistep nature of malignant transformation. Here we discuss promises and open issues of this concept as well as the important question of common insertion sites statistics and its pitfalls.

Gene Therapy (2008) **15,** 143–153; doi:10.1038/sj.gt.3303052; published online 1 November 2007

Keywords: *stem cells; retroviral vectors; insertional mutagenesis; clonal dominance; SCID*

Introduction: genetic modification of haematopoietic stem cells to treat inherited blood diseases

Inherited monogenetic diseases may be viewed as the 'natural target' of gene therapy, since in principle correction of the one defective gene underlying the disease should be sufficient to cure a given patient.[1,2] This is particularly true if the respective disease is pathophysiologically linked to one defined organ system. In the latter case, replacement therapy in the given organ should allow to induce disease remission or, if possible even full recovery.

Due to its easy accessibility, blood is one of the organs in the human body that is of special interest for gene therapeutic interventions. Moreover, the blood system clearly reveals a hierarchic structure with haematopoietic stem cells (HSC) being the origin of any mature blood cell.[3] It has been calculated that human beings as well as other mammals do have a relatively limited number (in the range of about 10 000–30 000) of long-term repopulating HSC;[4,5] thus, correction of a small number of cells might be sufficient to eradicate a genetic disease in the blood system. In line with these theoretical considerations, it has been shown almost 40 years ago that transplantation of allogeneic (from an major histocompatibility complex-matched, sibling donor), healthy HSC contained in the bone marrow is sufficient to cure inherited immune deficiencies.[6]

An even more striking proof for the power of genetic therapy has been presented by Mother Nature herself. In some rare cases, inherited mutations are corrected in pathophysiologically relevant cells by chance. This leads to the phenomenon of somatic mosaicism that has been observed for various genetic diseases including tyrosinaemia, severe combined immune deficiency (SCID) syndromes and Fanconi anaemia (reviewed in Hirschhorn[7]). If such natural reversion results in a sequence polymorphism as compared to the normal gene, gene-corrected cells are readily identifiable. This provides estimates for the number of corrected (stem) cells that are required to obtain various levels of disease correction.

An interesting example of natural gene therapy has recently been reported for Fanconi anaemia.[8] As convincingly proven by the group of Grompe and colleagues, intrauterine correction of the mutated FANCA gene in a single (embryonic) HSC was sufficient to result in fully normal blood systems of two identical twins. Since

Correspondence: Professor Dr B Fehse, Experimental Paediatric Haematology and Oncology, Paediatric Clinic III, University Hospital, Johann Wolfgang Goethe University, Theodor-Stern-Kai 7, Frankfurt 60590, Germany.
E-mail: b.fehse@kinderkrebsstiftung-frankfurt.de
Received 11 July 2007; revised 30 August 2007; accepted 14 September 2007; published online 1 November 2007

correction of the HSC could only have appeared in one of the two affected twins, this case report impressively demonstrates the power of transplanting gene-corrected cells.[8]

Considering the above findings, one could easily envision the ideal gene therapy–directed gene repair in the pathophysiologically relevant cells to induce the mentioned somatic mosaicism. Again, in a hierarchic system like haematopoiesis gene correction would be sufficient in a limited number of stem cells. If such repair system worked error free, there would be no side effects to be expected—a future vision of gene therapy recently described as 'precision genome surgery'.[9]

Unfortunately, 'genomic surgery' is still science fiction, but it is moving closer. Non-integrating vectors that are needed for getting the correct gene version into the target cell bearing the defect have become efficient gene transfer vehicles. To replace the mutant by the 'healthy' gene, homologous recombination between the two DNA copies is required. Natural recombination is not very efficient in most cells which has so far precluded its clinical use.[9,10] However, homologous recombination could be strongly facilitated by inducing double-strand breaks in (the vicinity of) the target sequence. With the advent of designer nucleases (reviewed in Paques and Duchateau[10]), those double-strand breaks could be induced right in the place where they are needed. This resulted in much better recombination efficiencies (10–20%)[10] reaching the level of clinical relevance. On the other hand, the still quite high levels of off-target cleavage with those designer nucleases were not acceptable for clinical purposes. Two recent publications have presented a novel generation of Zinc finger nucleases with a much better safety profile.[11,12] Thus, it might soon be possible that gene therapists eventually really start curing diseases by correcting genes.

So far, however, all one can do in a clinical setting is to add a normal, 'healthy' copy of the missing (or mutated) gene to the target cell. Since the correct gene copy is required to be inherited with every cell division, it should be incorporated into the cell's genome. Adding something to the genome per definition means introducing a mutation. That is the reason why we have to discuss the obstacle of insertional mutagenesis (see below).

Besides the problem of correcting a genetic misinformation, there is the task of defining and identifying the right target cells. As discussed above, in the blood system one could go for HSC. Although we do know where to find the HSC (see above), we are not yet able to non-ambiguously identify them based on phenotypical markers. Instead, using magnetic cell sorting for cells expressing the CD34 surface molecule, an enrichment of HSC by a factor of approximately 100 is being achieved in clinical HSC transplantation settings. Since those CD34 cells are reckoned as enriched HSC they are also being used for genetic modification. However, it should be taken into account that long-term repopulating, multipotent HSC have been estimated to constitute less than two-tenths of a per cent of the CD34 fraction.[5,13] Complementary to theoretical considerations, these estimates are also supported by surrogate assays for HSC, such as the numbers of SCID mouse-repopulating units.[13,14] This has several practical implications. Importantly, if *ex vivo* transduction of CD34 cells is used to gene-modify HSC, most corrected cells will be progenitor, precursor or even almost mature blood cells rather than HSC. Therefore, CD34 'bulk' transduction is not only associated with high, unnecessary cell production (that is GMP-) costs (for retroviral vector supernatant as well as cell culture reagents, including cytokines), but also leads to a high degree of uncertainty with regard to the eventual transduction efficiency for long-term repopulating HSC. In fact, the bitter experience of the first years in gene therapy clearly showed that high-level gene transfer even into $CD34^+$ cells not necessarily implies high transduction of long-term reconstituting HSC.[14,15]

Although transduction of cells capable of reconstituting the entire blood system is possible on the basis of current protocols (Schmidt *et al*.[16]), transduction of stem cells still seems to be less efficient than transduction of more mature cells. Therefore, it might be necessary to allow very high gene transfer rates (that is multiple vector copies)[17,18] in progenitor and mature cells to ensure efficient genetic modification of the less susceptible HSC. This, of course, brings us to the next problem, namely the possible consequences of high transduction rates in cells capable of (potentially long-term) self-renewal. Genetic modification of large (instead of the necessary low) cell numbers is associated with the accumulation of huge numbers of vector integrations thus leading to a high probability of unintentional insertion events in the vicinity of growth-regulatory genes and their possible consequences, namely insertional mutagenesis.[19–21]

Altogether, HSC are a very good model to illustrate the chances but also the current limitations of gene therapy. Convincing successes in the treatment of inherited severe immune deficiencies[22–25] have for the first time proven the enormous potential of gene therapy in patient cohorts with a very bad quality of life, poor prognoses and absent or very limited alternative treatment options. At the same time, insertional mutagenesis has not remained a theoretical risk of gene therapy, but has already been observed in pre-clinical models and even clinical studies.[25–28] This review will focus on one particular side effect observed in pre-clinical models as well as in one clinical study—the phenomenon of 'clonal dominance' induced by retroviral insertional mutagenesis.

Clonal dominance and 'clonality'

The *terminus technicus* clonal dominance has been introduced in the mid 70s of the past century.[29] At that time it was used to illustrate a dominant immune response, for example by a certain B-cell clone (thus one could probably link this type of clonal dominance to the existence of dominant antigen epitopes). Up to recently, the term has been used relatively seldom and mostly in the context of prevailing immune answers. In fact, as of 30 April 2007, a *PubMed* search (www.pubmed.gov) using clonal dominance resulted in just 145 entries (for comparison, there were almost 3000 entries when using clonality). However, based on past years' observations the term clonal dominance has more recently become quite popular in the field of gene transfer.

Although the terms clonality and clonal dominance are now frequently used, there is a considerable potential for confusion and it is necessary to define these terms more precisely. Clonality refers to the origin of a particular cell population. If all members of a cell population are derived from one particular cell, the population is clonally derived and is sometimes also called a clone. That means, to unambiguously define a cell clone it is necessary to show that all clone members have been derived from the same founder cell. Practically, the identification of clonally derived cell populations requires the presence or introduction of unique, inheritable (in the sense that the marker is transmitted from the mother to the daughter cells) clonal markers that can be detected experimentally. In the context discussed in this review, the insertion site of a particular gene vector can serve as a clonal marker because the probability for identical insertions of multiple vectors is sufficiently small.

There are two important consequences of such a perspective. First, it is possible that the same cell belongs to two (or more) clones, if different founder cells (for example characterized by subsequent clonal marking events) are considered (Figure 1). In this sense, an individual cell population or colony might be simultaneously considered as monoclonal (if referring to the colony-inducing cell) and as oligo- or polyclonal (if considering multiple marking events at later daughter cell generations). Furthermore, it is not necessary that all cells of a clonally derived population exhibit identical properties as long as these are not strictly linked to the clonal marking event. If, in contrast, the clone-marking event itself induces a particular phenotype, this phenotype might serve as a defining criterion for the clonality.

In contrast to the term clonality, which refers to the origin of a cell population, clonal dominance quantifies the relative contribution of particular clones. The most extreme case of clonal dominance is monoclonality. In this situation, an entire tissue consists of the progeny of one particular cell. However, clonal dominance might also occur in oligo- or polyclonal tissues, if some clones dominate the others in terms of their contributions to cell production. Because a precise threshold that defines dominance is hard to be specified in general, a clear and unmistakable definition of what is meant by clonal dominance is required in any particular study.

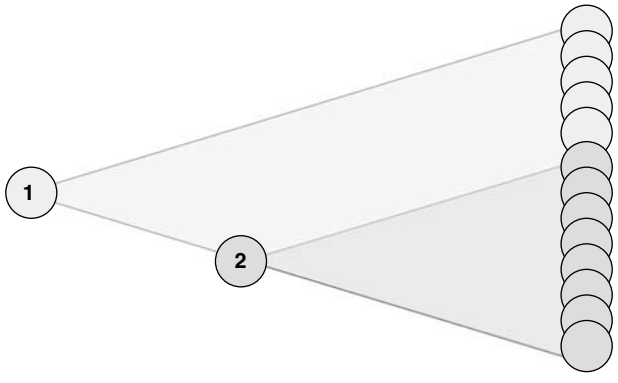


Figure 1 Clonality and nested clones. All cells shown (bright and dark grey) are descendants of cell 1, wherefore all these cells belong to the clone defined by this founder cell (clone 1). However, the cells marked by the dark grey background at the same time belong to another (sub)clone (clone 2) defined by founder cell 2. If subclones are characterized by different properties (for example by a mutation of cell 2) it is possible that cells of the same clone (here clone 1) contain cells with different properties and/or functionalities.

Last but not least, it should be pointed out that clonal dominance refers to the actual state of a system. It is in general not possible to conclude the displacement of other clones from the existence of dominant clones, that is it is possible that a particular dominant clone coexists with other (smaller) clones in a polyclonal situation. To describe the dynamics of clonal development (for example the conversion to monoclonality due to clonal competition) it is necessary to monitor clonal contributions over time.

Clonal dominance and retroviral gene marking

The probably first link between the terms clonal dominance and retroviral insertions can be found in a paper of Fanning *et al*.,[30] where the authors conclude that 'Restriction enzyme analysis suggested that pre-malignant outgrowths and pregnancy-dependent tumours both consisted largely of heterogeneous cell populations, although some evidence of clonal dominance was detected.' Thus, Fanning *et al*.[30] have used the retrovirus integration sites as markers to assess the dominance of some clones over others, but they have not found (searched for?) indications that the observed dominance was causally linked to the insertions.

The term clonal dominance has since also been used in the context of clonal disorders of multi-potent HSC, to some extent as a synonym of mono-clonality.[31,32] In an interesting review, Kerbel *et al*.[33] underline the usefulness of genetic marking to investigate the clonal evolutionary dynamics of tumour growth and the lineage relationship of primary tumours to their metastases. In fact, the reviewed data on genetic marking clearly indicate that some malignant cells do have a strong growth advantage over others. Subsequently, these advantages may lead to different contributions of various clones to tumour progression and to the eventual phenomenon of clonal dominance within the primary tumour and its metastases. In 1994, using a similar model based on retroviral gene marking Cornetta *et al*.[34] were able to confirm clonal dominance within metastases of breast cancer cells, but not in the primary tumours. Notably, in these early studies a possible influence of insertion sites themselves was not considered relevant.

The promising data with retroviral gene marking in different animal models were the basis for the use of this technique in human gene therapy. In fact, the very first official gene therapy study initiated in 1989 by Anderson and colleagues[35] at the NIH had no therapeutic implications but was a gene-marking study instead (aimed at following the *in vivo* fate of tumour-infiltrating lymphocytes). Subsequently, numerous marking studies were carried out in the setting of autologous haematopoietic stem cell transplantation (HSCT) mostly addressing the cause of malignant relapses after transplantation (reviewed in Tey and Brenner[36]). In principle, two different sources of recurring tumour/leukaemia cells are possible after autologous HSCT: the malignant cells may either have survived high-dose therapy in the body of the

patient or they may have been contained in the autologous stem cell graft. Determining the origin of the recurring malignant clone could therefore help to improve treatment regimens (for example better purging in case of graft contamination vs intensified therapy in case of *in vivo* survival). In fact, valuable insights were obtained by retroviral gene marking and, for many years, those marking studies were considered the most successful applications in human gene therapy.[36] Only recently, interpretation of some of the marking data has been questioned due to a possible influence of the marking vector itself (see below).

Clonal evolution in normal, stress and malignant haematopoiesis

Several groups used natural genetic markers (polymorphism) or limiting dilution approaches to investigate clonal evolution in steady-state and stress haematopoiesis. For example, based on data from multiple animal experiments in various species together with a mathematical model analysis Abkowitz and colleagues[37–40] concluded that haematopoiesis was maintained through clonal succession, that is a subsequent initiation of differentiating clones from a (self-renewing) stem cell pool. The application of their stochastic model of stem cell organization explains a possible appearance of clonal dominance of individual HSC clones after HSCT by chance.[37,38] The same is true for the establishment of macroscopic leukaemic clones after mutation of a single cell. Considering a competition of (possibly heterogeneous) HSC clones for common resources (for example stem cell-supporting microenvironments) that includes stochastic decisions, different theoretical models predict clonal dominance and the ultimate displacement of normal haematopoiesis by the leukaemic clone to be a stochastic event.[41,42] Although the probability of clonal dominance/leukaemic overgrowth depends on the growth advantage induced by the malignant transformation, it is still possible that the leukaemic clone might randomly disappear due to the process of stochastic fluctuation of clone sizes. This probability is decreasing with the size of the malignant clone. Furthermore, different theoretical models[41,42] predict that the malignant clone (if persistent) will completely displace normal haematopoiesis in the long run.

But not only malignant clones might have the potential to outcompete other stem cell clones. There is experimental[43–45] as well as theoretical evidence[44] that small genetically or epigenetically determined differences in cellular properties might lead to predictable patterns of clonal evolution including the dominance of particular stem cell clones.

On the basis of various mathematical models, conversion to monoclonality (as a particular type of clonal dominance) has been discussed also for normal (that is non-leukaemic) stem cell systems.[46–50] Whereas there is strong experimental evidence for (stochastic) clonal selection with the ultimate result of monoclonal systems in the crypts of the small intestine,[51–54] such a process has not been demonstrated for the haematopoietic system yet. It is currently unclear whether such phenomena do not exist in the haematopoietic system or whether they could not been detected due to the lack of appropriate experimental monitoring protocols. However, novel transplantation studies utilizing 'neutral' marking strategies and elaborated monitoring techniques (including non-invasive techniques) will allow experimental assessment of this hypothesis in the near future.[36]

It should be noted that the theoretically predicted time scales for the processes of clonal expansion, the generation of dominant clones and finally the conversion to monoclonality can be quite different and that they are strongly dependant on the status of the system. For example, after a unique but neutral marking of 300 stem cells in a mathematical model of murine haematopoiesis the predicted time to monoclonality has been calculated to be more than 65 years,[50] which is about 30 times the lifespan of a mouse. However, this time can be dramatically reduced if the system is in a highly activated, regenerative situation (for example after transplantation) or if the clonal markers induce competitive growth advantages. The latter scenario might be of particular importance for the interpretation of experimental results on the interrelation between retroviral gene marking and insertional mutagenesis.

In summary, clonal dominance previously described in malignant haematopoiesis is most likely the natural consequence of a clonal competition process that is biased by the malignant progression. However, clonal dominance may also occur (although on a much longer time scale) in normal haematopoiesis, even in the case of truly neutral clonal markers. Such a phenomenology can be explained consistently by stochastic models of stem cell organization.[37,48,50] As an important consequence of such a stochastic perspective, the fate of particular individual (stem cell) clones can only be predicted in a probabilistic sense. The latter perception should be taken into account when the phenomenon of induced clonal dominance is being investigated (see below).

Retroviral gene transfer and insertional mutagenesis

Based not only on the above theoretical considerations, but also in the light of experimental findings, clonal dominance was, until recently, considered a relevant result of malignant progression and a rare but possible event in normal, particularly post-transplantation haematopoiesis. However, many of the data used to establish the respective models were based on gene-marking experiments. The most commonly used marking tools were γ-retroviral vectors derived from murine leukaemia viruses (MLV). At the same time it was well established that these parental retroviruses, at least when replication competent, are highly mutagenic *in vivo*.[55,56] In spite of this, the risk of insertional mutagenesis in clinical gene therapy settings was thought to be negligible since the probability of hitting a relevant gene (for example a proto-oncogene, POG) in a cell, potentially susceptible to transformation events (a putative stem cell), seemed to be very small. Moreover, from numerous animal studies there was no evidence for strong mutagenicity of retroviral vectors.[19,57,58] Retrospectively, those previous animal data probably reflect limitations of early gene transfer protocols (low gene transfer rates into relevant target cells) in conjunction

with the use of models directed at studying efficacy rather than (non-expected) toxicity.[58]

Unfortunately, standard γ-retroviral vectors eventually turned out to be not as neat as we would like them to be. A first observation of leukaemia in a mouse model[27] was causally linked to the upregulation of a well-known POG—ecotropic viral integration 1 (*Evi1*). Only a few months later, a first leukaemia case[26] was diagnosed in an initially very successful gene therapy trial aiming at the treatment of SCID caused by the lack of a functional common γ-chain (SCID-X1).[22] Meanwhile, already four out of nine successfully treated kids developed leukaemia, one of them unfortunately died.[59] Although the more recent cases are still under investigation, molecular analysis of the first two clonal lymphoproliferative diseases provided strong evidence for insertional upregulation of a POG, the LIM domain-only protein LMO2 (Hacein-Bey-Abina *et al*.;[26] for a more detailed discussion on possible contributing factors please refer to references [57–62] and references therein).

Malignant transformation has been shown to require a multi-step process.[63] In fact, there are early stages of malignant diseases that might be considered as pre-malignant (for example hyperproliferative). In those stages, respective (pre-malignant) cell clones may already show signs of an upcoming malignancy (for example unlimited growth potential), but still respond to regulatory signals. Accumulation of further mutations may eventually lead to a fully transformed phenotype—a process that may take many years.

Based on these considerations it is tempting to speculate that insertional mutagenesis may also have different outcomes. It may, for instance, represent the initial or just one event during malignant transformation. Thus, a possible outcome of insertional mutagenesis could be some kind of intermediate stage between normal and transformed cells (not searched for or overseen in former experimental models), for example non-malignant clones bearing some growth advantage leading to the dominance over non-mutated cells. In a best-case scenario, such a growth advantage would never be followed by further transformation. This would be one possible basis for a phenotype of clonal dominance. An overview of different outcomes of insertional mutagenesis/retroviral gene marking is provided in Figure 2.

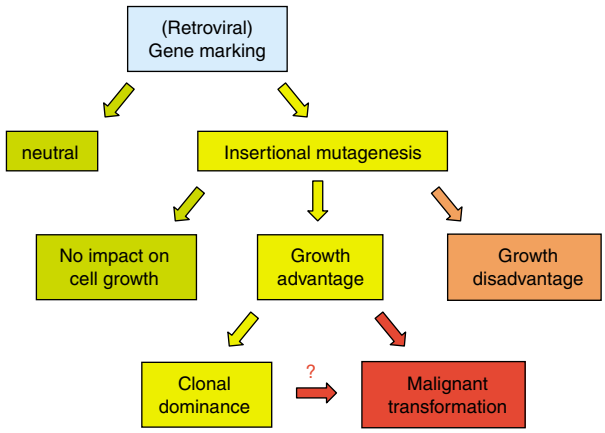


Figure 2 Possible outcomes after gene marking with mutagenic vectors. Gene marking with mutagenic vectors may have different consequences—from a non-measurable impact up to the malignant transformation of a given cell (clone).

Induced haematopoietic stem cells

In fact, molecular analysis of haematopoiesis in mice after serial transplantation of retrovirally marked HSC revealed a surprising accumulation of vector insertions in genes known to be involved in the growth/survival regulation of HSC.[28] A relatively large database of those insertions confirmed that integrations into growth-regulatory genes become particularly dominant upon serial transplantation.[64] Moreover, distribution of insertions in long-term reconstituting cells significantly differs from that in freshly transduced cells, suggesting an *in vivo* selection towards 'supportive insertions'. Strikingly, *Evi1*, the gene identified in the first retroviral vector-mediated leukaemia, also represented the most frequent hit in dominant, non-malignant clones (see below). Similarly, a non-random distribution of retroviral vector insertion sites with the MDS1/EVI1 locus representing the most hit locus was established in HSC clones reconstituting long-term haematopoiesis in non-human primates.[65] However, the frequent observation of clones bearing insertions in the MDS1/EVI1 locus has not been associated with clonal dominance in that particular study. In line, dominating clones induced by retroviral insertions were also not detected in a recent non-human primate study, but the animal number was limited to two.[66]

Taken together a lot of experimental data strongly suggest insertional mutagenesis being a potential driving force of clonal survival and/or eventual dominance in the used animal models of (serial) HSCT. In line with the *in vivo* data, Du *et al*.[67] and Modlich *et al*.[68] confirmed the possibility of generating dominant clones of primary murine bone marrow–derived haematopoietic cells by insertional mutagenesis *in vitro*. In this context, an interesting observation was reported back in 1981 by Greenberger *et al*.[69]. They found increased survival of murine bone marrow cultures *in vitro* after infection with MLV. It might well be that those early data were also related to insertional mutagenesis rather than non-genetic, virus-mediated effects. However, no method for sensitive and high-throughput cloning of insertion sites was available at that time.

As in the murine transplantation models, the *in vitro* studies identified the POG *Evi1* (beside others) as one prevalent hit in dominant clones. In conjunction, these data seem to support the above speculation that insertional mutagenesis may result in some pre-malignant growth advantage. However, further serial transplantation of dominant *in vivo* clones as well as infusion of clones obtained *in vitro* did not result in malignant outgrowth, arguing against a pre-transformed status of the dominant clones.[28,67]

In contrast, it points to the possibility of 'creating dominant stem cell clones' by (retroviral) insertion-mediated gene activation. One could imagine two mechanisms for that—the first would be based on the induction of a growth programme in susceptible stem cells present in the target cell population. Given the frequent activation of growth-promoting genes such as

Evi1 in dominant clones this seems to be an obvious hypothesis. However, there is one drawback. If activation of the growth-promoting gene by retroviral vector insertion has to appear in long-term repopulating stem cells, the probability of such an event should be very low, considering the very limited number of these cells present in the transduction process and accessible for retroviral vectors. However, one particular locus, Mds1/Evi1 or MDS1/EVI1, was found to be the most frequent insertion site in serially transplantable dominant HSC clones in various mouse experiments (approximately 5% of all identified dominant integrations)[64] and almost exclusively hit in long-term reconstituting HSC in a clinical gene therapy trial[25] (see below) while obviously not representing an integration hotspot in (human) HSC.[70] The second possible mechanism would include the activation of the (haematopoietic) 'stem cell programme' in cells which already underwent some commitment. In this case, the vector-mediated gene activation should affect one or several so-called 'stemness genes'. Such induction of 'stemness' may in principle occur without the initiation of any process eventually leading to malignant transformation (see below). The possibility of creating even pluripotent embryonic stem cell-like cells, so-called induced pluripotent stem cells (iPS), from differentiated skin fibroblasts by retroviral transfer of just four genes was first demonstrated by Takahashi and Yamanaka.[71] This work has been confirmed recently in a number of studies.[72–74] At the same time, those studies also pointed to the possibility of malignant transformation in some of the generated iPS cells. In analogy, one might hypothesize that activation of certain genes in already committed cells transforms those cells into 'induced HSC' (iHSC).

Whereas the generation of iPS cells has not yet been achieved in the human system, there is clear evidence for the possibility of inducing clonal dominance after vector-mediated gene activation from a clinical gene therapy study. In 2006, Ott *et al*.[25] reported the first successful correction of an inherited immune deficiency (chronic granulomatous disease, CGD) in adult patients. That study was associated with a number of unexpected turns: the initial, relatively high gene transfer levels (approximately 40% in both patients) reflected the progress in transduction conditions. However, the unmatched *in vivo* gene-marking rates (15–20%) in the myeloid compartment of both patients (treated according to the novel protocol developed by Aiuti *et al*.[23]) came as a most welcome surprise, in particular since gene-corrected HSC seem not to have any selection advantage *in vivo*. The optimistic perception changed when both patients showed an unexpected increase in the percentage of gene-corrected cells approximately 6 months after the infusion of transduced CD34-selected HSC by a factor of about 3. Surprisingly, this increase stopped without any clinical intervention and the levels of gene marking remained constant for about 12 months. In fact, molecular analysis of insertion sites confirmed that the raise was not due to the clonal outgrowths of some malignant clone. However, the obtained data brought another set of quite worrying news: an over time increasing percentage of insertions was detected in just three different genes—*MDS1-EVI1*, *PRDM16* or *SETBP1*. Eventually, the vast majority of surviving clones had insertions exclusively in the *MDS1-EVI1* locus.[25] Still, clones dominating a large part of the haematopoiesis for several months were regularly replaced by others indicating at least partially conserved mechanisms of clonal succession and/or a limited lifespan of the respective progenitor cells.

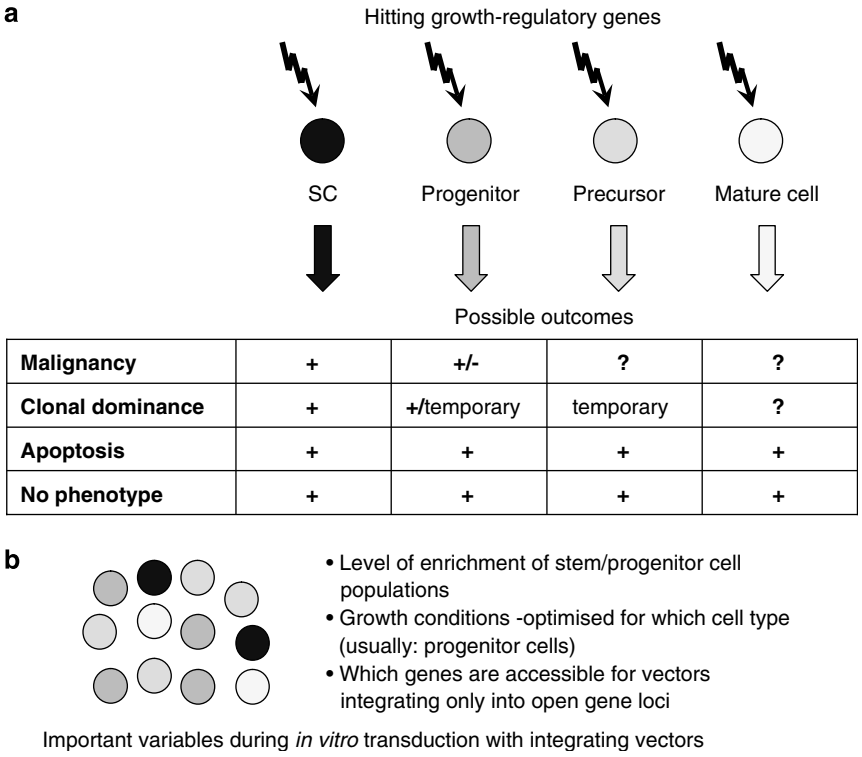
Together, the data from that clinical study strongly support the hypothesis of gene vector-induced clonal dominance. The limited lifespan of many of the genetically modified clones indicates that activation of genes such as *EVI1* in more committed target cells may result in a temporarily, possible self-limiting (for example based on telomere length) growth advantage. It remains to be investigated whether some of the induced clones represent progenies of 'true', long-living HSC.

It is still under investigation whether induced clonal dominance may only be observed with γ-retroviral vectors and when progenitor cells are transduced. A recent report by Evans-Galea *et al*.[75] provides evidence that inclusion of insulator elements into a lentiviral vector results in the suppression of clonal dominance in cultured human lymphoid cells. This data suggest that lentiviral vectors not only do have the potential to induce clonal dominance, but this might be even true, at least *in vitro*, in other than haematopoietic progenitor, namely lymphoid cells. The near future will obviously bring further insights with regard to this potentially very relevant observation.

Induced clonal dominance and malignant transformation

As already mentioned before, there are two possible ways to get closer to the phenomenon of induced clonal dominance: one might either view induced clonal dominance as one step in the multi-step process towards malignant transformation or postulate that clonal dominance may be achieved by artificially generating some kind of 'induced stem cells'. Although seemingly somewhat antagonistic at a first glance, both theories have several common features.

Only stem cells are capable of permanently sustaining regenerative tissues (including their own population), such as the haematopoietic system. Furthermore, for many malignant diseases the presence of tumour stem cells has been demonstrated.[76–78] Because obviously not every cell may act as a stem cell, it is also suggestive that not every cell can be transformed into a tumour stem cell. Based on these considerations one might argue that hitting a POG such as *Evi1* in one cell (for example a putative stem cell) may represent the first step of establishing a malignant clone, whereas hitting the same gene in a more mature cell may not be sufficient to induce a process of malignant transformation, but still result in a growth advantage (Figure 3). In most cases, the self-renewal probability of the latter, initially more differentiated cells will probably be limited (that is below 0.5) by intrinsic properties. This would result in a specific type of asymmetric cell fate, that is a given clone would lose its stem cell character while expanding. In those induced clones the obtained growth advantage would *per se* be of transient nature, that is they are doomed to eventually extinct. A temporary growth advantage that is limited by the absence of telomerase function might serve as a typical example for such an intrinsic limitation.



Malignancy	+	+/-	?	?
Clonal dominance	+	+/temporary	temporary	?
Apoptosis	+	+	+	+
No phenotype	+	+	+	+

Figure 3 Variables influencing the impact of insertional mutagenesis. (**a**) Long-term effects of insertional mutagenesis largely depend on the target cell—activation of a given gene in a stem cell may result in a growth advantage, whereas overexpression of the same gene in differentiated cells may lead to their apoptosis. (**b**) Consequently, the cell culture/transduction conditions do represent important variables influencing the effects mediated by gene marking.

It remains to be investigated whether the induction of growth-promoting genes (such as *Evi1*) itself may potentially be associated with activation of growth programmes with decreased self-renewal (see below).

The hypothesis of limited self renewal capacity is in good agreement with *in vivo* data from human as well as animal studies.[20,25,28] In fact, besides the common presence of insertions within the *MDS1/EVI1* locus in almost all haematopoietic cells from 1 year after transplantation on the observed clonal dominance in the CGD trial has been characterized by a permanent exchange between different clones. Even clones almost dominating the whole haematopoiesis for some period of time eventually disappeared.[25] To definitely decide whether this indicates that insertional mutagenesis indeed did not create true stem cells but rather conferred a temporary growth advantage still needs to be a analysed in detail. It should be mentioned that disappearance of dominating clones with Evi1 insertions has also been observed after serial transplantation in the murine system.[28]

On the other hand, clonal fluctuations and transient clonal contributions have also been suggested for long-term repopulating HSC on the basis of cell population data in chimaeric animal models as well as simulation studies (see above).[37–42,50] Therefore, one could argue that appearance and disappearance of those Evi-1-positive clones just reflect physiological behaviour of stem cells. In the same line of argument, experimental data from the murine models indicate that the dominant clones were able to reconstitute mice even after serial transplantation, that is they fulfilled a major criterion for stem cells. Still one might argue that due to the relatively short observation periods in mice this stem cell definition might be too fuzzy. Thus, the distinct findings in various settings may either reflect limitations of the resolution strength in a given system or actual differences of human and murine haematopoiesis.

Altogether, current data do not support the idea of 'induced clonal dominance' representing the first step of inadvertent malignant transformation. At the same time it remains to be verified in both humans and mice whether normal HSC (iHSC) can be created by insertional mutagenesis. Independently on the status of the dominant clones (HSC or just temporary contributing), their existence provokes a number of questions which may help to better understand the regulation of hierarchic cell systems, but also the process of malignant progression. Those questions include: (1) Under which conditions does the permanent, non-regulated activation of a POG such as Evi1 lead to 'benign' clonal dominance instead of malignant transformation? (2) Could the expression of single genes turn on stem cell programmes and if yes, in which target cells? (3) Why do stem cell clones bearing gene transfer vector insertions dominate the non-transduced stem cells, even in the apparent absence of a sustained selective advantage (as in the CGD trial), which would be expected to result in an ultimate displacement of non-transduced clones? Some of the above questions can be (and currently are) addressed in suitable experimental systems. Thus, in near future we hopefully will get much information about the mechanisms behind induced clonal dominance.

The search for stemness genes or the power and the jeopardy of statistics

One obvious objection against the proposed induced clonal dominance is reflected by the question whether the accumulation of gene hits in common loci represents an integration bias rather than the result of *in vivo* selection. In this case, the insertions identified in dominant clones would just mark those genes that are hit in 'true', long-term reconstituting HSC, but not being related to the clonal dominance. In line, a recent study examining >1000 γ-retroviral insertion sites found a number of insertion 'hotspots' in human $CD34^+$ cells with an overrepresentation of POGs among them.[70]

Taken together, the data from the above *in vivo* as well as *in vitro* studies argue against the idea of an integration bias being sufficient to determine later clonal dominance.[25,28,64,65,67,68] In fact, analysis of insertion sites in $CD34^+$ cells immediately after transduction, but also in non-obese diabetes (NOD)-SCID repopulating cells did not reveal any comparable partiality as in the dominant clones.[70,79] Here one could argue that the significance of integration data obtained right after transduction is limited since the proportion of HSC might be rather low. However, even if one admits that only a limited number of gene loci is accessible in long-term repopulating HSC, this would not explain the frequent insertions into just one gene locus (Evi1) in many mouse experiments and the almost exclusive appearance of clones with hits in this locus after *in vitro* immortalization and in the CGD trial. Moreover, one would have expected to observe the same integration bias towards Evi1 at least when analysing insertion sites in NOD-SCID repopulating cells, a population suggested to be highly enriched for HSC.[79]

Since statistics do play an important role in those evaluations, it seems to be important to get a closer look into those. Today, the identification of common insertion sites (CIS) is one of the most often used criteria to assess the importance of a given gene hit. This approach indeed needs some reconsideration.[80] In fact, taken as a single parameter, the CIS criterion may be misleading, that is the sole finding of CIS is far from being sufficient to postulate any biological significance.[80] This is nicely illustrated by the so called 'Birthday Paradoxon',[81] a statistical problem that points to the difference between intuitive and mathematical probability: What is the probability that two persons on a soccer field (22 players+1 referee) have their birthday on the same day? Guided by the probability (under the simplifying assumptions that a year always has 365 days and that birthdays are uniformly distributed throughout the year) of about 0.003 that the birthday of two randomly chosen persons coincide, intuition might suggest a quite low probability for a common birthday among the 23 persons as well. However, because the probability depends on the number of possible pairings (22 for the first player, 21 for the second and so on) it increases very rapidly if the number of persons rises (Figure 4). For the 23 sportsmen it is already greater than 0.5! If you add the linesmen and three substitutes per team, the probability of having at least one paired birthday will already be more than 0.7 and it is almost sure (that is >0.99) to find at least two persons with common birthday if you consider more than 57 persons.

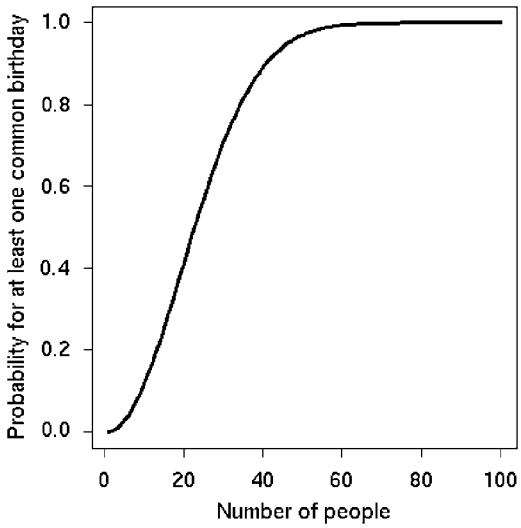


Figure 4 The Birthday Paradoxon—illustrating the difference between intuitive and mathematical probability. Shown is the probability of the occurrence of at least two common birthdays in a group of persons depending on the number of persons. As depicted, this probability is greater than 0.5 for just 23 people (for example the 22 players and the referee on a soccer ground) and exceeds 0.99 already for 57 persons.

This short side trip gives us two important clues—the probability of coinciding events can be much higher than intuitively anticipated and it increases in a non-linear fashion with higher sample size. Statistics could, of course, also be applied to predict insertion site distribution. One could, for instance, divide the whole genome (3×10^9 bp) into 30 000 possible insertion loci of 100 000 kb size (this quite nicely fits with the estimated gene number in the human genome, so in a quite simplified manner, one could also consider this as 30 000 gene loci with a total size of 100 kb including 5′ and the 3′ susceptible regions. Windows of such size (100 kb) are frequently used to analyse gene-surrounding regions potentially susceptible to insertional mutagenesis[70,80]). The above theoretical segmentation of the genome allows estimating the probability of CIS with increasing numbers of obtained integrations. Such an approach is not only very easy; it also allows taking into account other parameters such as the limited accessibility of genes in different target populations. For example, with 30 000 accessible gene loci (for simplicity, all of them assumed to have equal insertion probability) the probability of any CIS (that is at least two identical integration sites in the sample) between just 100 identified insertions would be approximately 0.15; with (a more realistic estimate of) 3000 accessible gene loci already about 0.81. Even more strikingly, if one analyses 1000 insertion sites, the probability of observing at least one CIS would amount to almost 1 (that is it is almost sure) even if considering the whole genome (all 30 000 loci) equally accessible (Figure 5).

These theoretical considerations have some significant implications. First, even when analysing relatively small numbers of integration sites, the probability of finding a CIS may (depending on the cell system) be quite high. In fact, given the obviously limited number of genes accessible (for example for γ-retroviral vectors) in a given

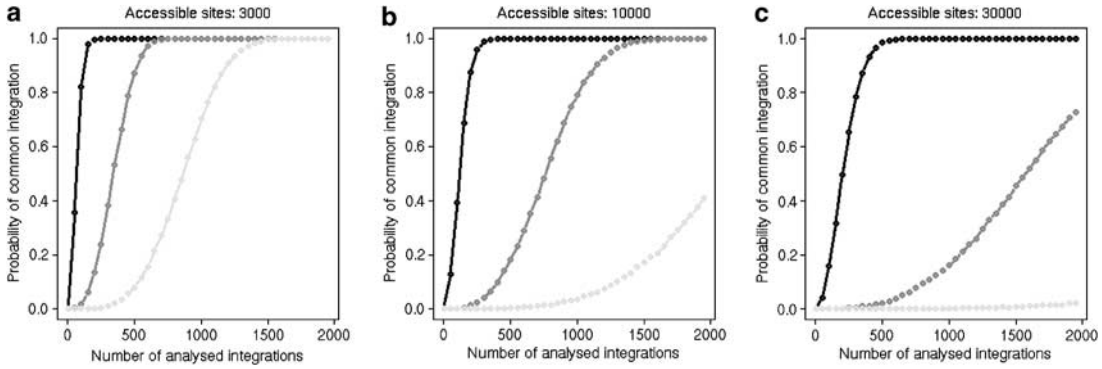


Figure 5 Probabilities for observing CIS by chance. Shown are the probabilities for observing CIS assuming that integration into all sites occurs equally likely with a probability of 1/(number of totally accessible integration sites). The probabilities, depending on the number of analysed insertions, are given for different numbers of totally accessible integration sites: (**a**) 3000, (**b**) 10 000 and (**c**) 30 000. Gray scale values for the probability of observing common integrations: black, at least two (CIS pair); dark grey, at least three (CIS triple); light grey, at least four (CIS quadruple). Even when postulating a non-selective integration over the whole genome (30 000 accessible sites), the probability of finding paired CIS increases very fast. Moreover, also multiple CIS may occur by chance, particularly for limited numbers of accessible gene loci and increasing numbers of analysed insertions.

cell population (for example $CD34^+$ cells), such probability could be expected to be considerably larger than 0.5 for almost any target cell after analysing just 100 insertion sites. Second, the probability of observing CIS, very quickly approaches 1 with increasing numbers of analysed insertion site, even if considering a huge number of potentially accessible loci. If the number of analysed insertions gets in the same range as the number of possible different insertions (=accessible sites), CIS will become the rule rather than an exception. Third, given the above numbers, there is a certain probability of finding potentially growth-regulating or even established POGs among the CIS. In fact, if one would assume the number of POGs to be in the range of 300 (Baum *et al*.[58]), approximately 1% of all CIS would fall into a POG by chance. Considering the likely case of an overrepresentation of growth-regulatory genes among the active gene loci in stem cells, common integrations into POG loci may be even more frequent in stem cells compared to mature, non-dividing cells.

Importantly, the above mathematical considerations did not take into account the integration behaviour of commonly used integrating (for example retroviral) vectors. Since it has been shown that retro- (including lenti-)viral vectors show a strong integration bias,[79,82,83] CIS by chance may happen even much more frequently than anticipated by the outlined statistical considerations (please also refer to Wu *et al*.[80] for further details).

One important conclusion of these mathematical reflections is the necessity to consider the possibility that with increasing numbers of analysed integration sites there is a high probability that a given CIS may have occurred by chance. Therefore, the sole identification of two insertions into one gene locus (even an 'interesting one') might not be sufficient to regard this insertion site biologically significant. In addition, since the probability of randomly detectable CIS increases in a non-linear fashion with an increasing number of analysed integration sites, the insertion pattern of different vectors should not be compared based on different numbers of analysed insertions. Otherwise, the vector with less insertion sites analysed might appear to be 'safer' since it seemingly generates less CIS.

Beyond these simplified theoretical illustrations, an algorithm to calculate the expected number of CIS in a given study has been described during revision of this manuscript.[84] Based on a Poisson approximation of the CIS numbers, the proposed algorithm allows the statistical testing of whether the observed CIS numbers can be expected to be non-random.

Conclusions

Insertional mutagenesis by γ-retroviral vectors may result in malignant transformation as well as in induced clonal dominance. This has important implications for their future use in clinical gene therapy where less genotoxic vectors are a must. The concept of induced clonal dominance may be of interest for future cell therapeutic strategies. It will, however, require better insights into the underlying mechanisms. Investigation of those mechanisms will also contribute to our understanding of the biology of normal and malignant stem cells. Herein, a careful statistical analysis is required to avoid misleading interpretations of the experimental and clinical results.

Acknowledgements

This work has been supported by the Deutsche Forschungsgemeinschaft DFG within the Priority Program SPP1230 by grants to BF (FE568/9-1) and IR (RO 3500/1-1). BF's professorship has been funded by the Deutsche Krebshilfe.

References

1 Friedmann T, Roblin R. Gene therapy for human genetic disease? *Science* 1972; **175**: 949–955.
2 Blaese RM, Culver KW, Miller AD, Carter CS, Fleisher T, Clerici M *et al*. T lymphocyte-directed gene therapy for ADA-SCID: initial trial results after 4 years. *Science* 1995; **270**: 475–480.

3 Bryder D, Rossi DJ, Weissman IL. Hematopoietic stem cells: the paradigmatic tissue-specific stem cell. *Am J Pathol* 2006; **169**: 338–346.
4 Sieburg HB, Cho RH, Muller-Sieburg CE. Limiting dilution analysis for estimating the frequency of hematopoietic stem cells: uncertainty and significance. *Exp Hematol* 2002; **30**: 1436–1443.
5 Abkowitz JL, Catlin SN, McCallie MT, Guttorp P. Evidence that the number of hematopoietic stem cells per animal is conserved in mammals. *Blood* 2002; **100**: 2665–2667.
6 Gatti RA, Meuwissen HJ, Allen HD, Hong R, Good RA. Immunological reconstitution of sex-linked lymphopenic immunological deficiency. *Lancet* 1968; **2**: 1366–1369.
7 Hirschhorn R. *In vivo* reversion to normal of inherited mutations in humans. *J Med Genet* 2003; **40**: 721–728.
8 Mankad A, Taniguchi T, Cox B, Akkari Y, Rathbun RK, Lucas L *et al*. Natural gene therapy in monozygotic twins with Fanconi anemia. *Blood* 2006; **107**: 3084–3090.
9 Pingoud A, Silva GH. Precision genome surgery. *Nat Biotechnol* 2007; **25**: 743–744.
10 Paques F, Duchateau P. Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. *Curr Gene Ther* 2007; **7**: 49–66.
11 Miller JC, Holmes MC, Wang J, Guschin DY, Lee YL, Rupniewski I *et al*. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol* 2007; **25**: 778–785.
12 Szczepek M, Brondani V, Buchel J, Serrano L, Segal DJ, Cathomen T. Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat Biotechnol* 2007; **25**: 786–793.
13 Bhatia M, Wang JC, Kapp U, Bonnet D, Dick JE. Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice. *Proc Natl Acad Sci USA* 1997; **94**: 5320–5325.
14 Larochelle A, Vormoor J, Hanenberg H, Wang JC, Bhatia M, Lapidot T *et al*. Identification of primitive human hematopoietic cells capable of repopulating NOD/SCID mouse bone marrow: implications for gene therapy. *Nat Med* 1996; **2**: 1329–1337.
15 Anderson WF. Gene therapy. The best of times, the worst of times. *Science* 2000; **288**: 627–629.
16 Schmidt M, Hacein-Bey-Abina S, Wissler M, Carlier F, Lim A, Prinz C *et al*. Clonal evidence for the transduction of CD34+ cells with lymphomyeloid differentiation potential and self-renewal capacity in the SCID-X1 gene therapy trial. *Blood* 2005; **105**: 2699–2706.
17 Kustikova OS, Wahlers A, Kühlcke K, Stähle B, Zander AR, Baum C *et al*. Dose finding with retroviral vectors: correlation of retroviral vector copy numbers in single cells with gene transfer efficiency in a cell population. *Blood* 2003; **102**: 3934–3937.
18 Fehse B, Kustikova OS, Bubenheim M, Baum C. Pois(s)on—it's a question of dose…. *Gene Therapy* 2004; **11**: 879–881.
19 Baum C, Düllmann J, Li Z, Fehse B, Meyer J, Williams DA *et al*. Side effects of retroviral gene transfer into hematopoietic stem cells. *Blood* 2003; **101**: 2099–2114.
20 Modlich U, Kustikova OS, Schmidt M, Rudolph C, Meyer J, Li Z *et al*. Leukemias following retroviral transfer of multidrug resistance 1 (MDR1) are driven by combinatorial insertional mutagenesis. *Blood* 2005; **105**: 4235–4246.
21 Baum C, Kustikova O, Modlich U, Li Z, Fehse B. Mutagenesis and oncogenesis by chromosomal insertion of gene transfer vectors. *Hum Gene Ther* 2006; **17**: 253–263.
22 Hacein-Bey-Abina S, Le Deist F, Carlier F, Bouneaud C, Hue C, De Villartay JP *et al*. Sustained correction of X-linked severe combined immunodeficiency by *ex vivo* gene therapy. *N Engl J Med* 2002; **346**: 1185–1193.
23 Aiuti A, Slavin S, Aker M, Ficara F, Deola S, Mortellaro A *et al*. Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning. *Science* 2002; **296**: 2410–2413.
24 Gaspar HB, Bjorkegren E, Parsley K, Gilmour KC, King D, Sinclair J *et al*. Successful reconstitution of immunity in ADA-SCID by stem cell gene therapy following cessation of PEG-ADA and use of mild preconditioning. *Mol Ther* 2006; **14**: 505–513.
25 Ott MG, Schmidt M, Schwarzwaelder K, Stein S, Siler U, Koehl U *et al*. Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. *Nat Med* 2006; **12**: 401–409.
26 Hacein-Bey-Abina S, von Kalle C, Schmidt M, McCormack MP, Wulffraat N, Leboulch P *et al*. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* 2003; **302**: 415–419.
27 Li Z, Düllmann J, Schiedlmeier B, Schmidt M, von Kalle C, Meyer J *et al*. Murine leukemia induced by retroviral gene marking. *Science* 2002; **296**: 497.
28 Kustikova O, Fehse B, Modlich U, Yang M, Düllmann J, Kamino K *et al*. Clonal dominance of hematopoietic stem cells triggered by retroviral gene marking. *Science* 2005; **308**: 1171–1174.
29 McMichael AJ, Willcox N. Radioactive antigen suicide of an anti-DNP (2,4-dinitrophenyl) clone. I. Recovery and escape from clonal dominance by suicide resistant precursors. *Eur J Immunol* 1975; **5**: 58–64.
30 Fanning TG, Vassos AB, Cardiff RD. Methylation and amplification of mouse mammary tumor virus DNA in normal, premalignant, and malignant cells of GR/A mice. *J Virol* 1982; **41**: 1007–1013.
31 Abkowitz JL, Fialkow PJ, Niebrugge DJ, Raskind WH, Adamson JW. Pancytopenia as a clonal disorder of a multipotent hematopoietic stem cell. *J Clin Invest* 1984; **73**: 258–261.
32 Green PL, Kaehler DA, Risser R. Clonal dominance and progression in Abelson murine leukemia virus lymphomagenesis. *J Virol* 1987; **61**: 2192–2197.
33 Kerbel RS, Waghorne C, Korczak B, Lagarde A, Breitman ML. Clonal dominance of primary tumours by metastatic cells: genetic analysis and biological implications. *Cancer Surv* 1988; **7**: 597–629.
34 Cornetta K, Moore A, Johannessohn M, Sledge GW. Clonal dominance detected in metastases but not primary tumors of retrovirally marked human breast carcinoma injected into nude mice. *Clin Exp Metastasis* 1994; **12**: 3–12.
35 Rosenberg SA, Aebersold P, Cornetta K, Kasid A, Morgan RA, Moen R *et al*. Gene transfer into humans—immunotherapy of patients with advanced melanoma, using tumor-infiltrating lymphocytes modified by retroviral gene transduction. *N Engl J Med* 1990; **323**: 570–578.
36 Tey SK, Brenner MK. The continuing contribution of gene marking to cell and gene therapy. *Mol Ther* 2007; **15**: 666–676.
37 Abkowitz JL, Catlin SN, Guttorp P. Evidence that hematopoiesis may be a stochastic process *in vivo*. *Nat Med* 1996; **2**: 190–197.
38 Abkowitz JL, Catlin SN, Guttorp P. Strategies for hematopoietic stem cell gene therapy: insights from computer simulation studies. *Blood* 1997; **89**: 3192–3198.
39 Abkowitz JL, Golinelli D, Harrison DE, Guttorp P. *In vivo* kinetics of murine hematopoietic stem cells. *Blood* 2000; **96**: 3399–3405.
40 Abkowitz JL, Golinelli D, Guttorp P. Strategies to expand transduced hematopoietic stem cells *in vivo*. *Mol Ther* 2004; **9**: 566–576.
41 Catlin SN, Guttorp P, Abkowitz JL. The kinetics of clonal dominance in myeloproliferative disorders. *Blood* 2005; **106**: 2688–2692.
42 Roeder I, Horn M, Glauche I, Hochhaus A, Mueller MC, Loeffler M. Dynamic modeling of imatinib-treated chronic myeloid leukemia: functional insights and clinical implications. *Nat Med* 2006; **12**: 1181–1184.
43 Muller-Sieburg CE, Cho RH, Thoman M, Adkins B, Sieburg HB. Deterministic regulation of hematopoietic stem cell self-renewal and differentiation. *Blood* 2002; **100**: 1302–1309.

44 Roeder I, Kamminga LM, Braesel K, Dontje B, de Haan G, Loeffler M. Competitive clonal hematopoiesis in mouse chimeras explained by a stochastic model of stem cell organization. *Blood* 2005; **105**: 609–616.
45 Sieburg HB, Cho RH, Dykstra B, Uchida N, Eaves CJ, Muller-Sieburg CE. The hematopoietic stem compartment consists of a limited number of discrete stem cell subsets. *Blood* 2006; **107**: 2311–2316.
46 Loeffler M, Birke A, Winton D, Potten CS. Somatic mutation, monoclonality and models of stem cell organization in the intestinal crypt. *J Theor Biol* 1993; **160**: 471–491.
47 Loeffler M, Bratke T, Paulus U, Li YQ, Potten CS. Clonality and life cycles of intestinal crypts explained by a state dependent stochastic model of epithelial stem cell organization. *J Theor Biol* 1997; **186**: 41–54.
48 Loeffler M, Roeder I. Tissue stem cells: definition, plasticity, heterogeneity, self organization and models—a conceptual approach. *Cells Tissues Organs* 2002; **171**: 8–26.
49 Meineke FA, Potten CS, Loeffler M. Cell migration and organization in the intestinal crypt using a lattice-free model. *Cell Prolif* 2001; **34**: 253–266.
50 Roeder I, Braesel K, Lorenz R, Loeffler M. Stem cell fate analysis revisited: interpretation of individual clone dynamics in the light of a new paradigm of stem cell organization. *J Biomed Biotechnol* 2007; **2007**: 84656.
51 Winton DJ, Blount MA, Ponder BA. A clonal marker induced by mutation in mouse intestinal epithelium. *Nature* 1988; **333**: 463–466.
52 Park HS, Goodlad RA, Wright NA. Crypt fission in the small intestine and colon. A mechanism for the emergence of G6PD locus-mutated crypts after treatment with mutagens. *Am J Pathol* 1995; **147**: 1416–1427.
53 Bjerknes M, Cheng H. Modulation of specific intestinal epithelial progenitors by enteric neurons. *Proc Natl Acad Sci USA* 2001; **98**: 12497–12502.
54 Potten CS, Booth C, Pritchard DM. The intestinal epithelial stem cell: the mucosal governor. *Int J Exp Pathol* 1997; **78**: 219–243.
55 van der Putten H, Quint W, van Raaij J, Maandag ER, Verma IM, Berns A. M-MuLV-induced leukemogenesis: integration and structure of recombinant proviruses in tumors. *Cell* 1981; **24**: 729–739.
56 Mikkers H, Berns A. Retroviral insertional mutagenesis: tagging cancer pathways. *Adv Cancer Res* 2003; **88**: 53–99.
57 Kohn DB, Sadelain M, Dunbar C, Bodine D, Kiem HP, Candotti F *et al*. American Society of Gene Therapy (ASGT) *ad hoc* subcommittee on retroviral-mediated gene transfer to hematopoietic stem cells. *Mol Ther* 2003; **8**: 180–187.
58 Baum C, von Kalle C, Staal FJ, Li Z, Fehse B, Schmidt M *et al*. Chance or necessity? Insertional mutagenesis in gene therapy and its consequences. *Mol Ther* 2004; **9**: 5–13.
59 ##www.esgct.org/upload/4th_CaseofLeukemial.pdf.
60 Davé U, Jenkins NA, Copeland NG. Gene therapy insertional mutagenesis insights. *Science* 2004; **303**: 333.
61 Kohn DB, Sadelain M, Glorioso JC. Occurrence of leukaemia following gene therapy of X-linked SCID. *Nat Rev Cancer* 2003; **3**: 477–488.
62 Nienhuis AW, Dunbar CE, Sorrentino BP. Genotoxicity of retroviral integration in hematopoietic cells. *Mol Ther* 2006; **13**: 1031–1049.
63 Hahn WC, Weinberg RA. Modelling the molecular circuitry of cancer. *Nature Rev Cancer* 2002; **2**: 331–341.
64 Kustikova O, Geiger H, Li Z, Brugman MH, Chambers SM, Shaw CA *et al*. Retroviral vector insertion sites associated with dominant hematopoietic clones mark 'stemness' pathways. *Blood* 2007; **109**: 1897–1907.
65 Calmels B, Ferguson C, Laukkanen MO, Adler R, Faulhaber M, Kim HJ *et al*. Recurrent retroviral vector integration at the Mds1/Evi1 locus in nonhuman primate hematopoietic cells. *Blood* 2005; **106**: 2530–2533.
66 Bozorgmehr F, Laufs S, Sellers SE, Roeder I, Zeller WJ, Dunbar CE *et al*. No evidence of clonal dominance in primates up to four years following transplantation of multidrug resistance 1 retrovirally-transduced long-term repopulating cells. *Stem Cells* 2007; **25**: 2610–2618.
67 Du Y, Jenkins NA, Copeland NG. Insertional mutagenesis identifies genes that promote the immortalization of primary bone marrow progenitor cells. *Blood* 2005; **106**: 3932–3939.
68 Modlich U, Bohne J, Schmidt M, von Kalle C, Knöss S, Schambach A *et al*. Cell culture assays reveal the importance of retroviral vector design for insertional genotoxicity. *Blood* 2006; **108**: 2545–2553.
69 Greenberger JS, Shadduck RK, Jaenisch R, Waheed A, Sakakeeny MA. Effects of murine leukemia virus infection on long-term hematopoiesis *in vitro* emphasized by increased survival of bone marrow cultures derived from BALB/Mo mice. *Cancer Res* 1981; **41**: 3556–3565.
70 Cattoglio C, Facchini G, Sartori D, Antonelli A, Miccio A, Cassani B *et al*. Hot spots of retroviral integration in human CD34+ hematopoietic cells. *Blood* 2007; **110**: 1770–1778.
71 Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 2006; **126**: 663–676.
72 Okita K, Ichisaka T, Yamanaka S. Generation of germline-competent induced pluripotent stem cells. *Nature* 2007; **448**: 313–317.
73 Wernig M, Meissner A, Foreman R, Brambrink T, Ku M, Hochedlinger K *et al*. *In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* 2007; **448**: 318–324.
74 Maherali N, Sridharan R, Xie W, Utikal J, Eminli S, Arnold K *et al*. Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* 2007; **1**: 55–70.
75 Evans-Galea MV, Wielgosz MM, Hanawa H, Srivastava DK, Nienhuis AW. Suppression of clonal dominance in cultured human lymphoid cells by addition of the cHS4 insulator to a lentiviral vector. *Mol Ther* 2007; **15**: 801–809.
76 Lapidot T, Sirard C, Vormoor J, Murdoch B, Hoang T, Caceres-Cortes J *et al*. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* 1994; **367**: 645–648.
77 Wang JC, Dick JE. Cancer stem cells: lessons from leukemia. *Trends Cell Biol* 2005; **15**: 494–501.
78 Wicha MS, Liu S, Dontu G. Cancer stem cells: an old idea—a paradigm shift. *Cancer Res* 2006; **66**: 1883–1890.
79 Laufs S, Nagy KZ, Giordano FA, Hotz-Wagenblatt A, Zeller WJ, Fruehauf S. Insertion of retroviral vectors in NOD/SCID repopulating human peripheral blood progenitor cells occurs preferentially in the vicinity of transcription start regions and in introns. *Mol Ther* 2004; **10**: 874–881.
80 Wu X, Luke BT, Burgess SM. Redefining the common insertion site. *Virology* 2006; **344**: 292–295.
81 Singh S. *Fermats letzter Satz*. Deutscher Taschenbuch Verlag GmbH & Co. KG: München, 2000, pp 66–67. English original edition: Singh S. *Fermat's Last Theorem*. Fourth Estate: London, 1997.
82 Wu X, Li Y, Crise B, Burgess SM. Transcription start regions in the human genome are favored targets for MLV integration. *Science* 2003; **300**: 1749–1751.
83 Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, Bushman F. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* 2002; **110**: 521–529.
84 Abel U, Deichmann A, Bartholomae C, Schwarzwaelder K, Glimm H, Howe S *et al*. Real-time definition of non-randomness in the distribution of genomic events. *PLoS ONE* 2007; **2**: e570.