

Integration profile of retroviral vector in gene therapy treated patients is cell-specific according to gene expression and chromatin conformation of target cell

Luca Biasco^{1,2}, Alessandro Ambrosi³, Danilo Pellin³, Cynthia Bartholomae⁴, Immacolata Brigida¹, Maria Grazia Roncarolo^{1,2}, Clelia Di Serio^{2,3}, Christof von Kalle⁴, Manfred Schmidt⁴, Alessandro Aiuti^{1,5*}

Keywords: chromatin modifications; gene therapy; haematopoietic stem cells; lymphocytes; retroviral integrations

DOI 10.1002/emmm.201000108

Received May 25, 2010
Revised November 19, 2010
Accepted November 22, 2010

→ See accompanying
Closeup by Baum
DOI: <http://dx.doi.org/10.1002/emmm.201000110>.

The analysis of genomic distribution of retroviral vectors is a powerful tool to monitor 'vector-on-host' effects in gene therapy (GT) trials but also provides crucial information about 'host-on-vector' influences based on the target cell genetic and epigenetic state. We had the unique occasion to compare the insertional profile of the same therapeutic moloney murine leukemia virus (MLV) vector in the context of the adenosine deaminase-severe combined immunodeficiency (ADA-SCID) genetic background in two GT trials based on infusions of transduced mature lymphocytes (peripheral blood lymphocytes, PBL) or a single infusion of haematopoietic stem/progenitor cells (HSC). We found that vector insertions are cell-specific according to the differential expression profile of target cells, favouring, in PBL-GT, genes involved in immune system and T-cell functions/pathways as well as T-cell DNase hypersensitive sites, differently from HSC-GT. Chromatin conformations and histone modifications influenced integration preferences but we discovered that only H3K27me3 was cell-specifically disfavoured, thus representing a key epigenetic determinant of cell-type dependent insertion distribution. Our study shows that MLV vector insertional profile is cell-specific according to the genetic/chromatin state of the target cell both *in vitro* and *in vivo* in patients several years after GT.

INTRODUCTION

Retroviral vectors have been extensively used in gene therapy (GT) applications as an effective tool to permanently integrate a therapeutic gene of interest into a target cell, conferring in most cases stable and efficient transgene expression over time (Kohn

& Candotti, 2009; Naldini, 2009). Both wild-type retroviruses and retroviral vectors were initially thought to insert randomly into the host genome, but it is now well established that the majority of them display a semi-random or a non-random integration profile possibly influencing the fate of transduced cells (Brady et al, 2009; Bushman, 2007; Cattoglio et al, 2007; Ciuffi, 2008; Mitchell et al, 2004; Wu et al, 2003). Initial reports of leukaemia cases from the X-linked severe combined immunodeficiency (SCID-X1) GT clinical trial suggested a potential link between leukaemogenesis and insertional activation of the *LMO2* proto-oncogene (Hacein-Bey-Abina et al, 2003a,b). Indeed, it is now recognized that vector bearing enhancer sequences can alter the expression of neighbouring genes (Maruggi et al, 2009) and several studies associated events of clonal dominance *in vivo* to vector insertion sites

(1) The San Raffaele Telethon Institute for Gene Therapy (HSR-TIGET), Milano, Italy.

(2) Università Vita-Salute San Raffaele, Milano, Italy.

(3) CUSSB, Università Vita-Salute, Milan, Italy.

(4) National Center for Tumor Diseases (NCT), German Cancer Research Center (DKFZ), Heidelberg, Germany.

(5) University of Rome 'Tor Vergata', Rome, Italy.

*Corresponding author: Tel: +39 02 2643 4671; Fax: +39 02 2643 4668; E-mail: aiuti.alessandro@hsr.it

(Kustikova et al, 2007; Montini et al, 2006, 2009). On the other hand, *in vitro* analyses of transduced clones purified from patients several years after GT failed to show clear signs of perturbation of neighbouring genes (Cassani et al, 2009; Recchia et al, 2006). Additionally, although the analyses of integration sites from adenosine deaminase deficient SCID (ADA-SCID), SCID-X1, and chronic granulomatous disease (CGD) trials (Aiuti et al, 2007; Deichmann et al, 2007; Ott et al, 2006) show the presence of specific regions with recurrent integrations (common integration sites, CIS), it remains undefined to what extent the presence of CIS is the result of positive clonal selection after cell infusion or instead derives from preferential targeting for integration at the time of transduction (Cattoglio et al, 2007). Indeed, insertion site selection during *in vitro* transduction could be driven by tethering of transcription factors (TF) to specific regions according to TF binding sites location (Felice et al, 2009) and seems dependent on cellular determinants as well as on vector design (Lewinski et al, 2006).

Along this line, we studied the impact of vector integrations on clonal expansion and the frequency of CIS after GT in five patients from a clinical trial of haematopoietic stem cell gene therapy (HSC-GT) for ADA-SCID that has been shown to achieve immune and metabolic correction in the absence of adverse events related to gene transfer (Aiuti et al, 2002a, 2009). Our data did not reveal any sign of clonal dominance or aberrant expansions even in the presence of CIS in the *LMO2* or *CCND2* proto-oncogene loci (Aiuti et al, 2007). It is now believed that other factors including the disease background, the nature of the transgene, and the acquisition of other genetic abnormalities unrelated to vector insertions are also needed for aberrant expansion of transduced clones (Hacein-Bey-Abina et al, 2008; Howe et al, 2008). While most of these studies have been focusing on vector-on-host effects, there is still limited information on the role of the target cell status at the time of transduction and host-on-vector influences upon *in vivo* engraftment. Toward this aim, two recent publications addressed the possible influence of target cell type on integration profile of retroviral vectors in murine LSK subpopulations (Kustikova et al, 2009) and T-cells (Newrzela et al, 2008), the latter suggesting that the cell-dependent insertional pattern of transduced, mature murine lymphocytes play only a secondary role in inducing resistance to transformation. However, these findings remain to be defined in GT clinical trials on patient samples and a characterization of genomic features influencing integration preferences in different human target cells is currently missing. New information now provide a detailed genome-wide map of the epigenetic and chromatin state of different human cell types, like HSC and peripheral blood T-cells, allowing to compare retroviral vector distribution with several high-throughput mapped genomic features such as DNase I hypersensitive sites (HSS; Boyle et al, 2008; Xi et al, 2007) and histone methylation distribution (Barski et al, 2007; Cui et al, 2009). In this regard, recent reports have shown an interesting correlation between retroviral insertions and histone methylations in human cells (Brady et al, 2009; Wang et al, 2007, 2010).

In order to study the influences of host cell condition on vector insertion sites *in vitro* and after clonal selection in patients, we identified integration sites from patients affected by ADA-SCID and treated with infusions of moloney murine leukemia virus (MLV) transduced HSC (HSC-GT) or peripheral blood lymphocytes (PBL-GT; Aiuti et al, 2002b). This represents a unique model that allows the comparison of insertions derived from the same ADA-encoding vector in the context of the same genetic background but in two different target cells, before and several years after infusion in patients. We found that the MLV-derived vector displayed a different integration distribution in the two groups of patients under PBL-GT or HSC-GT. These cell-specific insertion preferences were directly related to the epigenetic state and expression profile of the cell type at the time of transduction and overall did not show any particular vector-driven bias in patients even long-term after GT.

RESULTS

Distribution of vector integrations from PBL-GT and HSC-GT

To compare the insertional profile of MLV vector in PBL-GT and HSC-GT, we identified a total of 4157 unique insertions from ADA-SCID patients (2198 from PBL-GT and 1959 from HSC-GT) by the combination of shotgun cloning-generated insertion libraries of ligation mediated PCR (LM-PCR) products and pyrosequencing of linear amplification mediated PCR (LAM-PCR) products (Table 1). The integrations were divided in two categories according to the samples from which they were collected for each group of patients, *in vitro* (PRE-GT) and *in vivo* (POST-GT). From HSC-GT patients, we retrieved retroviral integration sites (RIS) from both CD3+ T-cells and CD15+ granulocytes *in vivo*, whereas only T-cells were analysed in the PBL-GT group since in these patients the vector is undetectable in other lineages. The overall distribution of the vector integrants confirmed the classical preferences of MLV vector to integrate inside or in the proximity of coding regions

Table 1. Amounts of unique integration sites retrieved through LM-PCR/LAM-PCR from PBL-GT and HSC-GT *in vitro* pre-transplant (PRE-GT) and *in vivo* after transplant (POST-GT)

Patient samples	Pt1	Pt2	Pt3	Pt4
PBL-GT (2198 unique integrations)				
T cells (PRE-GT) pre-infusion	669	121	451	124
T cells (POST-GT) 6–9 years after last infusion	40	22	669	102
% of vector positive T cells	1.8	1.4	100	0.7
Patient samples				
HSC-GT (1959 unique integrations)				
CD34+ cells (PRE-GT) pre-infusion	NA	NA	558	329
T cells (POST-GT) 3–6 years after GT	85	204	300	382
% of vector positive T cells	52.6	100	100	69.8
Granulocytes (POST-GT) 3–6 years after GT	6	24	71	NA
% of vector positive granulocytes	1.1	5.8	3.4	NA

Percentages of vector positive cells *in vivo* at last follow up, measured through Q-PCR as copies of vector per cell genome, are also reported for each patient (NA = not available).

Table 2. Genomic distribution of MLV vector integration sites in PBL-GT and HSC-GT

Patient samples	Inside gene (%)	Outside gene (%)	10 kb window on TSS (%)	Clustered in CIS (%)
PBL-GT pre-GT CD3+	55.9	44.1	27.6	43.4
PBL-GT post-GT CD3+	51.3	48.7	23.6	34.0
PBL-GT total	54.1	45.9	26.1	39.8
HSC-GT pre-GT CD34+	52.6	47.4	20.0	36.1
HSC-GT post-GT CD3+	40.7	59.3	23.0	44.8
HSC-GT post-GT CD15+	40.6	59.4	23.8	15.8
HSC-GT total	46.1	53.9	21.8	39.4
Randomly generated RIS	40.7	59.3	2.1	NA
Human genome	35.5	64.5	ND	NA

Percentages of integrations *in vitro* and *in vivo* from PBL-GT (blue) and HSC-GT (red) landing inside genes, outside genes, in a 10 kb window centred on TSS of genes and involved in CIS *in vitro* and *in vivo* in PBL-GT or HSC-GT. Percentages relative to 100.000 random insertions and human genome reference are also shown for comparison (ND, not done; NA, not available).

(Table 2). In PBL-GT, 54.1% of RIS are inside genes and 26.1% in a window of 10 kb surrounding the transcription start site (TSS) of genes, while in HSC-GT their frequency was 46.1 and 21.8%, respectively.

Comparing the data obtained *in vitro* before infusion and *in vivo*, we found that the percentage of insertions inside genes was reduced after *in vivo* engraftment. PRE-GT *in vitro* samples showed an enriched integration frequency inside genes both in PBL-GT and HSC-GT (55.9 and 52.6%, respectively) while POST-GT *in vivo* purified cell populations showed a lower frequency of RIS into genes as compared to *in vitro* data in PBL-GT T cells (51.3%) and in CD15+ cells (40.6%) or CD3+ cells (40.7%, $p < 0.001$. Test on proportions with Holm-Bonferroni correction) from HSC-GT. *In vivo* T-cells and granulocytes from HSC-GT displayed a slightly higher percentage of insertions surrounding the TSS compared to the *in vitro* dataset, while there is no indication of *in vivo* skewing in the PBL-GT samples.

Specific CIS were identified from PBL-GT and HSC-GT

We then analysed the presence of hotspots of integrations and CIS in all our datasets *in vitro* and *in vivo* normalizing the size of the clustering window for the number of integrations retrieved in each group. CIS, defined as clusters of two or more insertions, were present already *in vitro* at the time of transduction both in PBL-GT (43.4%) and HSC-GT (36.1%) groups (Table 2). The proportion of insertions involved in clusters increased upon *in vivo* selection only in CD3+ cells differentiated from transduced HSC (44.8%, $p < 0.05$. Test on proportions with Holm-Bonferroni correction) while it decreased in those derived from gene-corrected mature PBLs (34.0%, $p < 0.001$). Integrations from granulocytes did not display any particular clustering of insertions around specific regions to the level present *in vitro* in CD34+ cells (15.8%, $p < 0.05$).

By comparing the two clinical trials, only a small number of CIS genes (12 out of 474 total CIS) were in common between *in vitro* transduced PBL or HSC while the majority of hotspots were specific for the two target cells (Fig S1 of Supporting Information). This low similarity is only slightly increased *in vivo* where 14 out of 402 total hotspots were shared between T-cell from HSC-GT and T cells from PBL-GT patients. Among CIS genes retrieved *in vivo* only from HSC-GT, we found *LMO2*

and *MDS1-EVII*. These genes were targeted both in CD3+ and CD15+ cells from HSC-GT treated patients, while *in vitro* we found only one insertion near *LMO2* from transduced CD34+ cells. The *TCRA* locus was instead a clear CIS only in PBL-GT while no integration was retrieved in this region from HSC-GT samples.

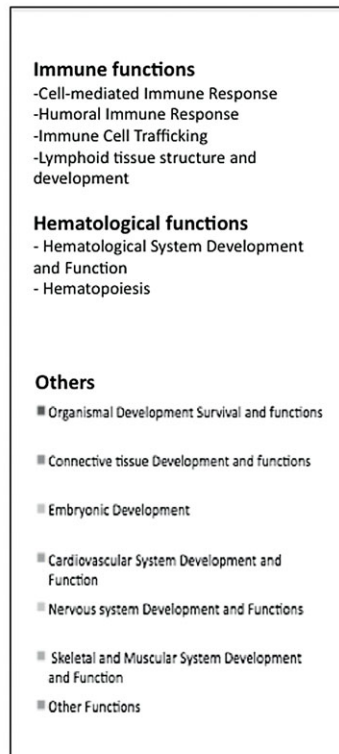
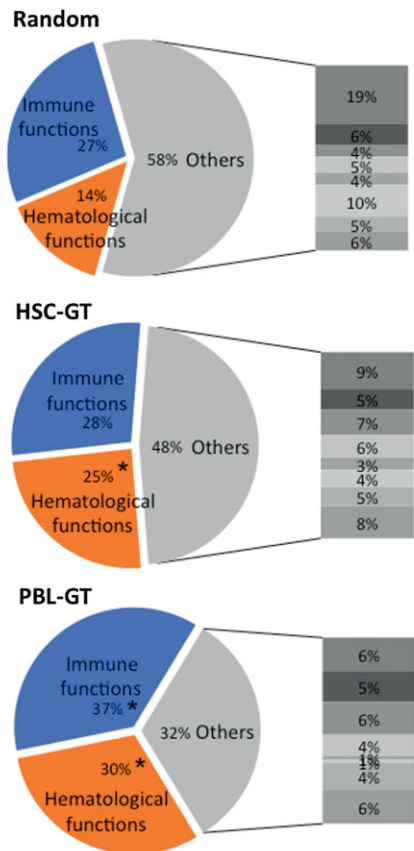
T-cell specific genes are favoured insertion sites in PBL-GT

To analyse the functions of genomic loci involved by RIS, we collected data regarding the gene directly hit or the single gene closest to each insertion site in order to assign one gene to one integration. Hit gene lists for all insertion sites from PBL-GT, HSC-GT and a random *in silico* dataset ($n = 100,000$) were uploaded on the ingenuity pathways analysis (IPA) to look for related biological functions.

We calculated the contribution of hit genes with respect to the functional categories listed in the IPA database. We analysed the frequency of hit genes listed in the two main categories 'Haematological functions' and 'Immune functions' with respect to the remaining functions (Others; Fig 1A). As compared to random reference both HSC-GT and PBL-GT displayed a higher frequency of hit genes belonging to haematological functions. PBL-GT displayed an even higher contribution of hit genes from this category together with a significant over-representation of immune functions genes as compared to HSC-GT (37% vs. 28%, $p < 0.001$. Test on proportions with Holm-Bonferroni correction) and random (27%, $p < 0.001$). The PBL-GT dataset showed also a lower frequency of genes involved in other functions, particularly in cardiovascular and nervous system. These general preferences were confirmed by restricting the analysis to CIS genes *in vitro* and *in vivo* from PBL-GT and HSC-GT (Fig S2 of Supporting Information).

Considering the nature of the cell type transduced in PBL-GT, we extrapolated T-cell specific functions from the IPA library and cross-compared these categories with our list of hit genes. We found that genes involved in proliferation ($p < 0.001$. Test on proportions with Holm-Bonferroni correction), activation ($p < 0.001$), differentiation and development ($p < 0.05$) of T lymphocytes were clearly a preferential target for integrations in PBL-GT when compared to the HSC-GT insertions and random dataset (Fig 1B). We then looked closer at the canonical

A Physiological functions



B T cell functions

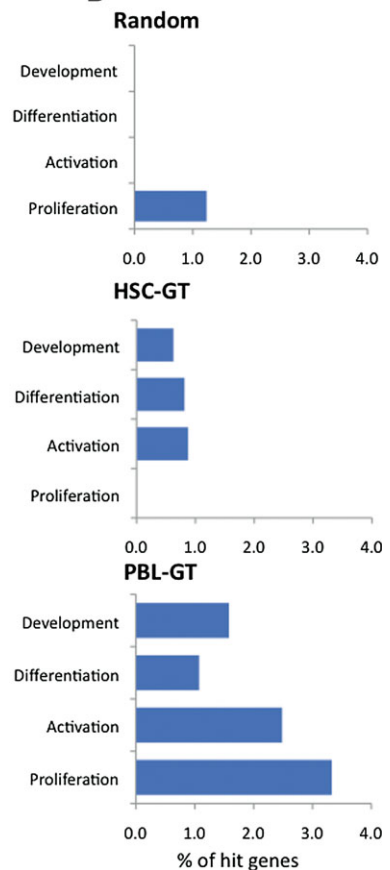


Figure 1. Functions of genes hit by the MLV vector in PBL-GT and HSC-GT.

- A.** Physiological functions of genes hit by random *in silico* insertions as well as by all HSC-GT and PBL-GT integrations *in vitro* and *in vivo*. Haematological functions (orange), immune functions (light blue) and other functions (scale of grey) derive from a pool of categories listed in the Ingenuity pathway analysis software (IPA) database. Percentages are calculated over the total number of hit genes involved in physiological functions. Asterisks indicate over-represented categories with respect to random dataset ($p < 0.001$. Test on proportions with Holm–Bonferroni correction).
- B.** T-cell functions genes hit by random *in silico*, HSC-GT and PBL-GT integrations. Percentages are calculated over the total number of hit genes. Categories extrapolated from haematological and immune functions through the tool ‘effect on function’ from IPA software.

T-cell receptor (TCR) signalling pathway from the IPA library. As shown in Fig S3 of Supporting Information, the TCR pathway was strongly and preferentially hit by integrations derived from mature lymphocytes as compared to the ones from HSC and their differentiated progeny. We also analysed the contribution of hit genes to cytokine and interleukin signalling pathways (Fig 2). T-cell specific pathways like signalling of TCR, CD28 and IL-2 were significantly targeted only in PBL-GT as compared to the expected random frequency calculated by IPA ($p < 0.05$, Benjamini–Hochberg correction for multiple testing). On the contrary, IL-6 signalling genes were significantly over-represented in the HSC-GT group of patients and not in the PBL-GT one ($p < 0.05$). We also divided the list of hit genes according to their different representation in the two GT trials (Fig S4 of Supporting Information). Again, the preference for T-cell specific pathways was found only in the PBL-GT specific hit genes and PBL-GT/HSC-GT commonly hit genes.

No *in vivo* enrichment for ‘Cancer’ genes in PBL-GT and HSC-GT

We then studied the contribution of hit genes to the disease category ‘Cancer’ (from IPA software) for all the integration subsets (Fig S5 and Supporting Information ‘List of Cancer Genes hit in PBL-GT and HSC-GT’). Proto-oncogenes were over-represented in both trials when compared to randomly generated integrations ($n = 12,323$ genes, $p < 0.05$. Test on proportions, Holm–Bonferroni correction). However, by comparing *in vitro* and *in vivo* data, we did not find any significant enrichment in terms of hit proto-oncogenes among the different subsets in both groups of patients. The in depth analysis of genes related to myeloid leukaemia, lymphocytic leukaemia and lymphomas *in vitro* in both the trials showed no significant *in vivo* skewing for any of these categories (Fig S5 and Supporting Information ‘List of Leukaemia–Lymphoma Genes hit in PBL-GT and HSC-GT’). Although CD15+ cells from HSC-GT seem to display a slightly higher contribution of myeloid leukaemia and lymphocytic leukaemia genes, these differences were not

Cytokine signaling pathways

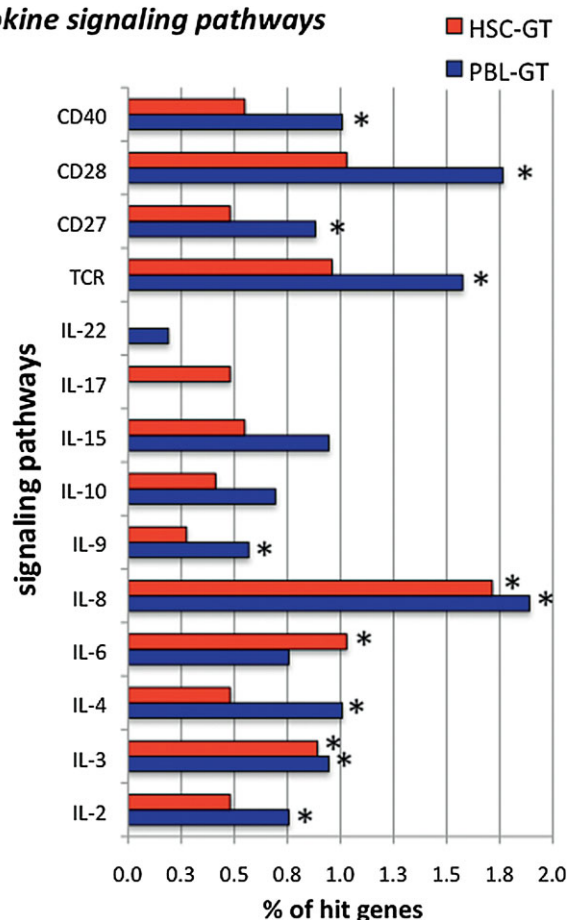


Figure 2. Signalling pathways hit by retroviral integrations in PBL-GT and HSC-GT. Percentages of genes belonging to different signalling pathways calculated on the total number of hit genes (* p -value < 0.05, Fisher's exact test with B-H correction, random reference from IPA software). Frequency of hit genes is shown by blue bars (PBL-GT) and red bars (HSC-GT).

significant by the comparison with *in vitro* CD34+ cells integration dataset (Test on proportions, Holm–Bonferroni correction).

Differential gene expression profile influences vector integrations in PBL and HSC

To get more insight into the differential distribution of the vector in the two transduced cell types, we analysed the expression profile of target cells at the time of transduction through HG-U133A Affymetrix microarray chips. We then restricted the analysis to the list of genes found to be hit in PBL-GT or HSC-GT and compared the expression levels of hit genes to the overall expression profile in T cells and CD34+ cells (Fig 3A). We found that the expression of hit genes was skewed towards the highest expression categories in both trials as compared to the random reference ($p < 0.0001$ for all subgroups except for HSC CD15+ with $p = 0.052$, Mann–Whitney test). In contrast, no significant skewing was found between the expression profile of hit genes retrieved from *in vivo* samples as compared to *in vitro*, in both PBL-GT and HSC-GT groups.

By the comparison between the two Affymetrix chip datasets, we were also able to extrapolate differentially expressed genes in T cells and CD34+ cells at the time of transduction from the distribution of differences (deltas) based on expression values (Fig 3B). We identified a small fraction of probesets (225 out of 22,284; 1.01%) that were differentially expressed above an arbitrary level of ± 3 delta robust multi-array averaging analysis (RMA) values. A total of 105 probesets (corresponding to 89 genes) were found highly expressed in T cells at the time of transduction, while 120 probesets (relative to 113 genes) were specifically highly expressed in HSCs (Fig 3 heatmap and Supporting Information ‘Differentially expressed genes in T cells vs. HSC’).

We then tried to assess how many integrations hit these two groups of genes in PBL-GT or in HSC-GT (Fig 3B column graph), considering that the calculated random frequency of insertions is 0.5% in each group. Integrations from PBL-GT involved 19 out of 89 T-cell specifically expressed genes (35 RIS, 1.6% of PBL-GT insertions) and only 3 out of 113 CD34+ cells specific genes (4 RIS, 0.2% of PBL-GT insertions). On the contrary, the HSC-GT integration dataset contained 11 out of 113 of HSC specifically expressed genes (23 RIS, 1.2% of HSC-GT insertions), whereas T-cell expressed genes were significantly less represented (3 out of 89; 5 RIS, 0.3% of HSC-GT insertions; $p < 0.005$. Test on proportions, Holm–Bonferroni correction).

We then analysed the distribution of hit genes *in vitro* and *in vivo* in both RIS datasets on the basis of their differential expression levels between the two chips (Fig 3C). We found that the insertions in HSC-GT versus PBL-GT were distributed among the profile of expression of hit genes in a significantly different fashion (p -value < 0.001 / < 0.0001 Mann–Whitney test). Moreover, PBL-GT RIS were more skewed towards the T-cell specifically expressed genes already *in vitro* at the time of transduction, with no significant difference as compared to the *in vivo* insertions. HSC-GT insertions were instead more related to CD34+ expressed genes both *in vitro* and *in vivo*, with CD3+ cells displaying a distribution more similar to the random reference as compared to *in vitro* dataset. These observations were confirmed and strengthened by limiting the analysis to CIS genes in all the subgroups *in vitro* and *in vivo* (Fig S6 of Supporting Information).

Cell-specific epigenetic marks correlate with integration preferences

We then assessed whether the differential vector bias for some loci was related to the overall chromatin accessibility of these regions in the two different target cell types. In a first analysis, we compared the distribution of our integrations to DNaseI HSSs mapped in human CD4+ T cells (Boyle et al, 2008). Overall, in both trials the MLV vector displayed a strong preference to integrate directly inside these regions as compared to the random reference (2% randomly simulated RIS $n = 100,000$; Fig 4A). Nonetheless, these features were hit in a significantly higher proportion in PBL-GT (34% of RIS from PBL-GT vs. 28% from HSC-GT; $p < 0.001$. Test on proportions with Holm–Bonferroni correction) and integrations from PBL-GT were on average more than two times closer to HSS mapped on T cells as

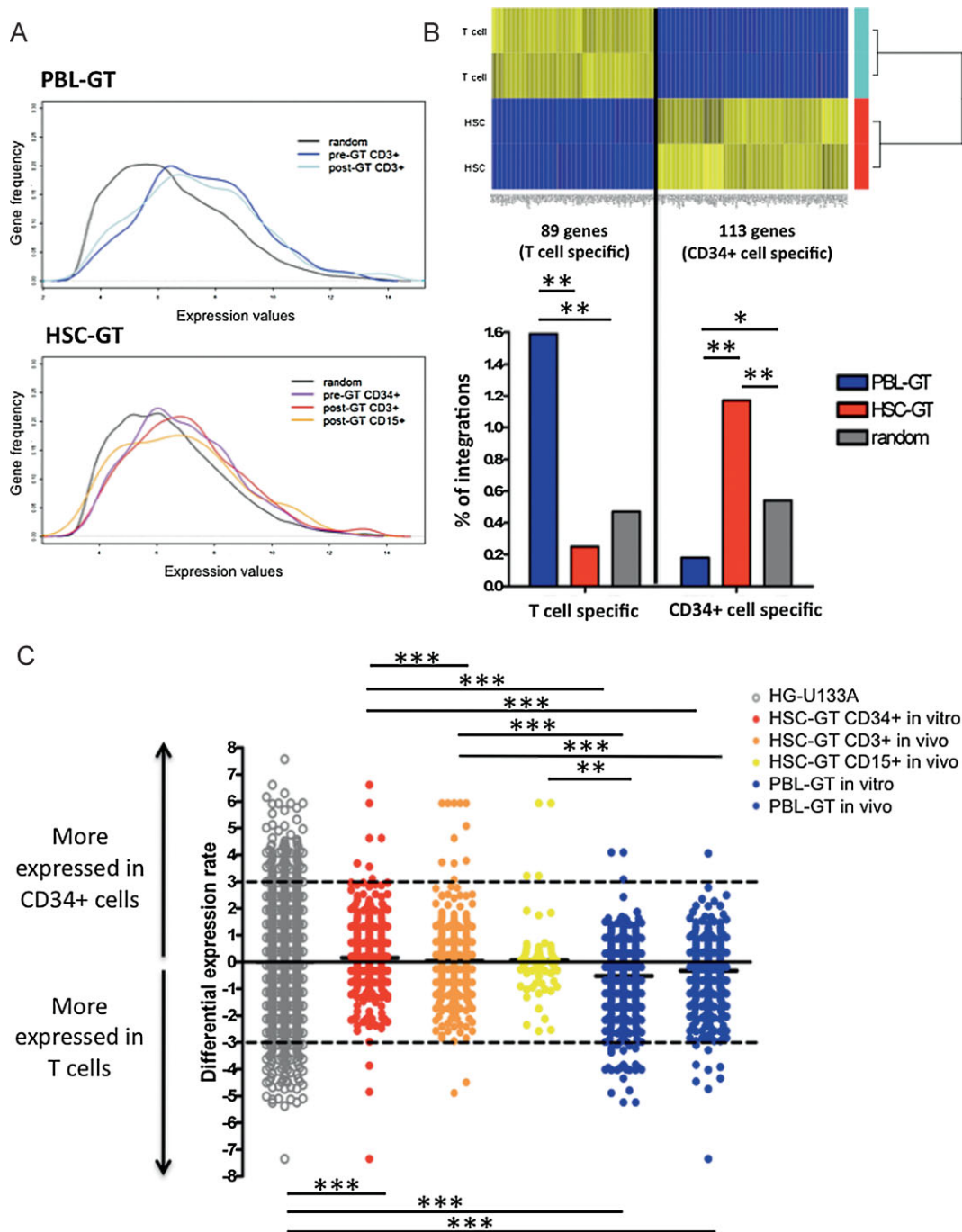


Figure 3. Correlation between expression profile of target cells and RIS in PBL-GT and HSC-GT.

- A.** Distribution of gene expression from HG-U133A affymetrix chips data (RMA values) in T-cells (upper panel) and in CD34+ cells (lower panel; random references, black lines). The analysis is also restricted to the genes found to be hit *in vitro* and *in vivo* from PBL-GT as well as *in vitro* and in CD3+ and CD15+ cells *in vivo* from HSC-GT.
- B.** Heatmap of differentially expressed genes in T cells and CD34+ cells. The column graph below shows the percentage of insertions from PBL-GT (blue column) and HSC-GT (red column) landing in these two group of genes as compared to random expected frequencies from HG-U133A probesets ($p < 0.05$, $**p < 0.005$. Test on proportions, Holm–Bonferroni correction).
- C.** Differential gene expression of hit genes in PBL-GT and HSC-GT. The dotted line in correspondence with +3 and –3 value refers to the threshold used for the comparative analysis of panel (B). Gene expression in target cells is represented in a scale of differential expression rate (RMA values) on y-axis for all genes in HG-U133A affymetrix chip (grey dots), hit genes *in vitro* and *in vivo* in CD3+ and CD15+ cells from HSC-GT (red, orange and yellow dots, respectively) as well as *in vitro* and *in vivo* in PBL-GT (dark blue and light blue dots; $**p < 0.001$, $***p < 0.0001$, Mann–Whitney test).

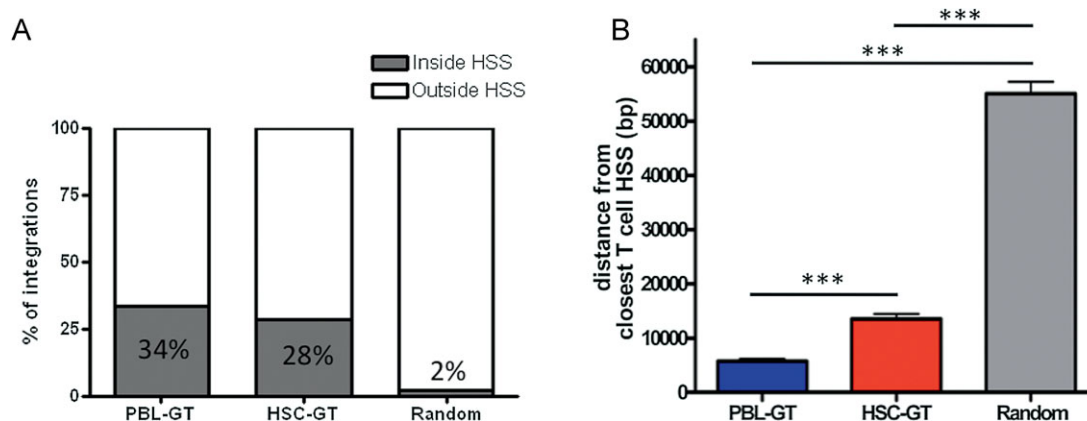


Figure 4. Comparison between insertion sites and T-cell DNase HSSs distribution.

A. Percentages of integrations landing inside a T-cell HSS in PBL-GT, HSC-GT and random *in silico* insertions.

B. Average distance in bp from the closest T-cell HSS of RIS from PBL-GT (blue bar), HSC-GT (red bar) and random *in silico* insertions (grey bar; Mean and SEM, *** $p < 0.0001$ Mann-Whitney test).

compared to HSC-GT (5.7 kb *vs.* 13.5 kb of average distance, respectively; $p < 0.0001$ Mann-Whitney test; Fig 4B).

To further dissect the relationship between integration preferences and chromatin status, we took advantage of the data available on histone methylation profile in CD4+ T cells and CD133+/CD34+ haematopoietic stem/progenitor cells (HSC/HPC; Barski et al, 2007; Cui et al, 2009). We analysed the distances of each of these features from every integration site considering a ± 45 kb window centred on RISs both from PBL-GT and HSC-GT *in vitro* and *in vivo*. This analysis was performed on histone methylations mapped both on T cells and HSC/HPC (Fig 5A). Remarkably, PBL-GT and HSC-GT datasets showed a higher contribution of integrations close to histone modifications associated with open chromatin state and gene activation, like H3K4me1 and H3K27me1 and with active TSS like H3K4me3. These preferences were present at the time of transduction and were maintained in POST-GT samples. On the other hand, histone methylations associated with heterochromatin regions, like H3K9me3, were disfavoured by the MLV vector equally in both cell types. Strikingly, the disfavoured for H3K27me3 was cell-specific, since it was evident only when the RIS of one trial were compared to this feature mapped on their relative target cell. On the contrary, when compared to the H3K27me3 mapped on the unrelated target cell, the distribution of RIS related to this histone modification was more similar to the random reference. A complete analysis on other epigenetic features is shown in Figs S7 and S8 of Supporting Information. We also subdivided the insertional datasets *in vitro* and *in vivo* in two subgroups containing CIS insertions and NOT CIS insertions, respectively, and we extended the analysis of histone methylation density distribution on each of these loci (Fig S9 of Supporting Information). Integration sites from HSC-GT *in vivo* CD15+ cells were excluded in this analysis due to the small number of CIS retrieved. The distribution of histone methylations around CIS was overall similar to the regions hosting single integration events (NOT CIS). We found additional confirmation of cell-specific disfavoured for H3K27me3 modification,

although CIS from HSC-GT CD3+ cells *in vivo* showed a different probability density distribution of this feature, more similar to the random reference, as compared to NOT CIS integrations and insertions from CD34+ cells *in vitro*.

Finally, we selected two cell-specific hotspots, *LMO2* and *TCRA*, to study the local epigenetic landscape in relation to the specific insertion sites (Fig 5B). The *LMO2* region, that was found to be hit only in HSC-GT, displayed a clear heterochromatic configuration (high H3K27me3 density) in T cells being instead open in HSC/HPC. On the other hand, the *TCRA* locus, which was an hotspot in PBL-GT but did not host any HSC-GT insertions, showed in correspondence to insertion site cluster a strong signal of open chromatin conformation (high H3K4me3 density) in T cells while it is mostly in closed configuration in HSC/HPC.

DISCUSSION

Integration site analysis has now become a critical tool to monitor the activity of retroviral vectors on *in vivo* selection of patient clones (Aiuti et al, 2007; Deichmann et al, 2007; Ott et al, 2006). In view of the adverse events that occurred in the SCID-X1 GT clinical trials, a number of studies have focused their attention on the influences of the vector on neighbouring genes (Hacein-Bey-Abina et al, 2003b, 2008; Howe et al, 2008; Maruggi et al, 2009; Ott et al, 2006) as well as how these could lead to clonal dominance and possibly to malignant transformation (Deichmann et al, 2007; Kustikova et al, 2007; Montini et al, 2006, 2009). In contrast, the impact of host cell genetic/epigenetic conditions on vector target site selection *in vitro* and *in vivo* is often overlooked in GT clinical studies.

Here, we had the unique opportunity to compare integration sites in two GT trials, in the absence of adverse events and major clonal expansions, sharing the same MLV vector gene transfer approach and the same ADA-SCID genetic background and only differing in the transduced cell type infused in patient,

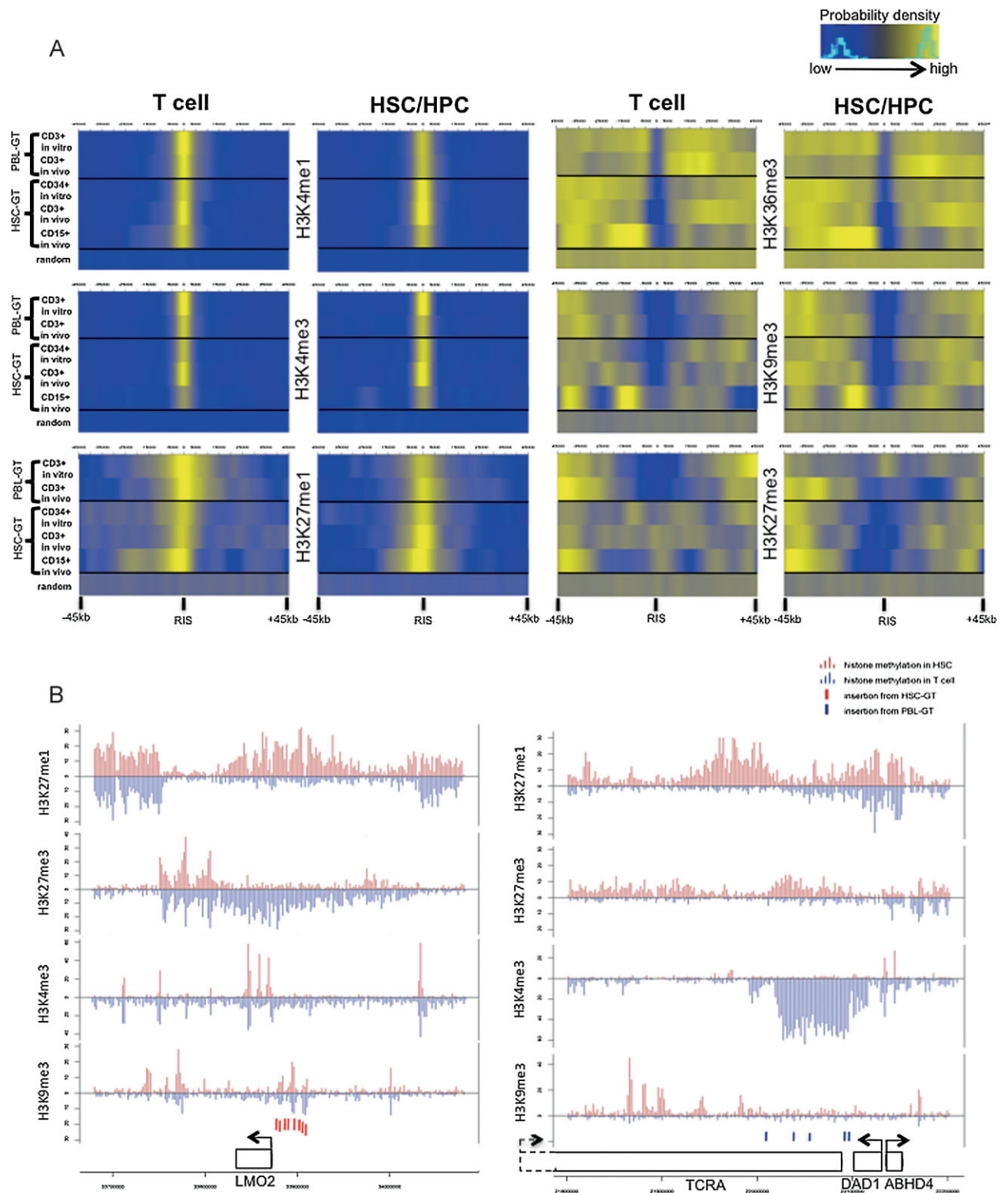


Figure 5. Comparison between histone tail modifications distribution in T-cell and HSC versus RIS in PBL-GT and HSC-GT.

- A.** Distribution of distances from six different histone tail modifications (H3K4me1, H3K4me3, H3K9me3, H3K27me1, H3K27me3 and H3K36me3) in a ± 45 kb window centred on RIS from PBL-GT, HSC-GT and random database. Probability density of distances is represented by a gradient from blue (low) to yellow (high). All insertion datasets are compared to the six histone tail modifications mapped in T cells or in HSC/HPC.
- B.** Distribution of four histone modifications in a 350 kb window centred on *LMO2* (left) and *TCRA* (longest transcript; right). Signals of four histone modifications mapped on the selected windows are represented for T cells as blue columns below the axis and for HSC/HPC as red columns above the axis of the mirror graph. Insertion sites are shown as blue lines (RIS from PBL-GT) and as red lines (RIS from HSC-GT).

mature T lymphocytes or haematopoietic progenitors in PBL-GT and HSC-GT, respectively. Our high-throughput analyses revealed a cell-specific vector preference that is related to the host cell status in terms of chromatin state and transcriptional activity at the time of transduction. Long-term follow up of treated patients showed that this cell-specific pattern is overall maintained *in vivo*.

The number of unique insertions retrieved from the different samples analysed correlate well with the percentage of vector positive cells calculated by q-PCR. As expected, the contribution of insertions from granulocytes in HSC-GT patients was lower with respect to other lineages, having proportions of transduced cells ranging between 1.1 and 5.8%. Nevertheless, the integration dataset from granulocyte cells is more complex and abundant than in other studies on GT clinical trials (Ott et al, 2006; Wang et al, 2010). In addition, in the present work, we collected a relevant number of insertions from PRE-GT samples (1365 RIS and 887 RIS from PBL-GT and HSC-GT, respectively) allowing a coherent comparison of *in vitro* data with integration sites detected long-term in patients. We were thus able to better discriminate between intrinsic insertional preferences in target cells before infusion and selective pressures or vector-driven bias after *in vivo* engraftment and selection.

Integrations inside genes were less represented in POST-GT samples, suggesting that clones carrying insertions into coding regions affecting cell survival were negatively selected *in vivo*. In addition, when focusing on CIS we found that clusters of insertions were present at a relevant frequency already at the time of transduction both in PBL-GT and HSC-GT, but the contribution of integrations involved in CIS increased in CD3+ cells differentiated from transduced progenitors of HSC-GT patients. This *in vivo* skewing could be explained by a positive role of vector position on growth-promoting genes, such as *LMO2* and *MDS1-EV11*. Nonetheless, the polyclonality of the RIS and the TCR repertoire as well as the lack of expansion of *LMO2* insertions by clonal tracking over time (Aiuti et al, 2007) would suggest that the finding of a CIS in the proximity of a proto-oncogene is not necessarily linked to aberrant proliferation *in vivo*. Although we found only one RIS in the *LMO2* locus *in vitro* in HSC-GT in the present dataset, it should be noted that this region has been previously shown to be a hotspot of integrations in CD34+ cells (Aiuti et al, 2007; Cattoglio et al, 2007). One could also argue that clones carrying insertions in HSC active loci would have a survival advantage due to a better detoxification from toxic metabolites, particularly in the early phases after GT during which purine metabolites are elevated due to enzyme replacement therapy discontinuation in combination with the effects of chemotherapy (Aiuti et al, 2009). Monitoring these clones in longer follow up will provide clearer information about a possible effect of vector on clonal dynamics.

The highest overlap of hotspots (6.6% of CIS genes) was found comparing the integrations from *in vitro* and *in vivo* lymphocytes in PBL-GT. On the other hand, few CIS genes were in common between lymphocytes and their transduced progenitors in HSC-GT, likely as a result of the complex developmental path from CD34+ cells to T cells.

Having observed differential genomic distribution and clustering of insertions in the two GT trials, we hypothesized that the different status of the two target cell types (lymphocytes and HSC/HPC) could have influenced in a cell-specific fashion the integration profile of MLV vector *in vitro* at the time of transduction and *in vivo* after selection. Our finding that MLV insertions from PBL-GT showed a strong preference for 'immune functions' genes and a lower contribution of categories unrelated to haematological/immune system was a first confirmation of a cell-dependent integration profile. This observation was further strengthened when we looked deeper at T-cell specific functions and pathways, which were significantly hit by integrations only in PBL-GT and not in HSC-GT. Analysis of the 'Cancer' category and leukaemia/lymphoma related genes from IPA software showed a higher contribution of hit genes with respect to random, with no significant *in vivo* selection for these categories, suggesting that integrations in these loci have not provided any growth advantage to transduced cells. It should also be noted that the definition of 'Cancer-related' genes is somehow arbitrary and a relevant number of genes belonging to this category are also involved in physiological functions of haemopoietic system.

We then wondered whether cell-specific preferences were linked to the expression profile of the two different target cells at the time of transduction. In line with previous reports on integration studies in haematopoietic cells (Aiuti et al, 2007; Cattoglio et al, 2007; Mitchell et al, 2004; Recchia et al, 2006), MLV vector favoured genes belonging to higher expression categories in both trials *in vitro*, without significant *in vivo* skewing. By the comparison of the expression profile in the two target cell types we found that MLV vector favoured genes specifically expressed in T cells in PBL-GT and genes specifically expressed in CD34+ cells in HSC-GT. Interestingly, the comparison between integrations in CD3+ cells *in vivo* from HSC-GT *versus* PBL-GT revealed that the same kind of lymphoid cells displayed in the two trials a completely different integration profile that was related to the type of transduced cells from which they derived *in vivo*.

However, gene expression is not the only mechanism driving integration preferences of MLV vector. Indeed, we found a substantial number of genes hit by the vector but not significantly expressed in the two target cells and a relevant overlap in the expression pattern of hit genes from the different sample subsets, suggesting that the differential expression profile could not account *per se* to the clustering of insertions in specific regions of the genome. This observation extends the information from previous studies on MLV insertional hotspots (Cattoglio et al, 2007; Recchia et al, 2006) and is also supported by our analysis of expression profiles of CIS genes (Fig S6 of Supporting Information). We therefore hypothesized that the overall accessibility of the genome at the time of transduction, of which 'gene expression' is an indicator, could play a more general role in inducing the cell-specific insertional preferences of MLV vector. For this reason, we correlated insertion sites with other epigenetic features accounting for different chromatin conformations.

DNaseI HSSs are commonly related to open chromatin state as well as with the presence of active DNA binding sites for regulatory proteins and have been associated with MLV vectors insertions in HeLa cells and CD34+ cells (Lewinski et al, 2006; Wang et al, 2010). Our results show that on average PBL-GT insertions were two times closer to HSS mapped on T cells (Boyle et al, 2008) as compared to HSC-GT RIS, thus accounting again for an additional mechanism linked to cell-specific preferences of MLV vector.

Histone modifications are other important epigenetic markers of open/closed chromatin state and their distribution was recently mapped through ChIP-Seq technique both in human T cells and HSC/HPC (Barski et al, 2007; Cui et al, 2009). Correlation of insertion sites and histone methylations was previously shown in recent studies (Brady et al, 2009; Wang et al, 2010). Our work provides novel information on how the general chromatin state of two different haematopoietic cells influences vector integrations by the cross-comparison of vector insertion profiles in PBL-GT and HSC-GT with several histone marks mapped in both target cell types. As compared to random distribution, insertion sites were preferentially located in proximity of histone modifications associated with an open chromatin state (Fig 5A). The strongest preference was noted for H3K4me3, a modification that is mainly present in correspondence of TSS of genes (Barski et al, 2007), similarly to recent observations in the X-SCID GT trial (Wang et al, 2010) thus indicating that this is a common feature of MLV vectors. Importantly, no cell-specific behaviour was found when comparing H3K4me3 mapped in the two target cells. One could envisage that, at a genome wide level, many of these H3K4me3 modifications share the same position in both T cells and CD34+ cells. Indeed, in a recent work on chromatin remodelling upon lymphoid differentiation, histone modifications involved in chromatin decondensation related to lymphoid-affiliated genes were already detected in HSCs (Maes et al, 2008). Histone methylations associated with heterochromatic conformations are, on the other hand, disfavoured by MLV vector integrations. Strikingly, among all histone markers, we discovered that H3K27me3 alone represents a key epigenetic determinant of cell-dependent integration profile of MLV vector, since it was the only modification that was disfavoured in a cell-specific fashion. These results are in agreement with the findings of *in vitro* studies showing that H3K27me3 distribution was significantly changed upon differentiation of haematopoietic progenitor cells (Cui et al, 2009) and from the analysis of epigenetic marks in distinct lymphoid lineages were the genomic association with H3K27me3 signals was both gene- and cell-specific (Wei et al, 2009). Our results support the concept that the different distribution of these cell-specific 'closed windows' in the genome of our target cells is indirectly driving the vector to 'open windows' associated with general chromatin accessibility and gene expression. In this view, the bias of MLV for some specific active loci could also be the consequence of chromatin inaccessibility to integration events of other regions. It should be also noted that the majority of the histone marks analysed, even when mapped in correspondence of genes, are not necessarily

associated to TSS, a preferential target for MLV vectors, thus suggesting a general active role of epigenetic signals also on integrations retrieved inside coding regions or collected from intergenic spaces. The different influence of some of these signals in T cells and HSC/HPC on integration events is further evident when focusing on *LMO2* and *TCRA* regions that were specific insertional hotspots in HSC-GT and PBL-GT, respectively (Fig 5B). By the analysis of CIS and NOT CIS loci we found that the cell-specific disfavoured for H3K27me3 signal is a general rule of integration site selections both in hotspots or regions hosting single integration events (Fig S9 of Supporting Information). Nevertheless, it is interesting to notice that CIS derived from HSC-GT *in vivo* CD3+ cells displayed a different distribution of this histone modification as compared to NOT CIS subset and to the CD34+ cells *in vitro* from which they derived. This is in agreement with the results regarding the transcriptional activity of CIS genes (significantly different between HSC *in vitro* and T cells *in vivo* from HSC-GT patients, Fig S6 of Supporting Information) and again points to a more complex mechanism of physiological selection of insertions upon *in vivo* differentiation.

In conclusion, our high-throughput analysis shows that MLV vector displays integration preferences that are cell-specific and closely related to the genomic and chromatin state of target cells at the time of transduction. Most of these insertional features are mirrored in *in vivo* patient samples without showing any particular deviation from *in vitro* vector distribution even several years after GT. These results better define the 'physiological' MLV vector behaviour in GT patients treated with different gene-corrected cells in absence of adverse events, thus providing information of reference for the follow up of ongoing clinical trials based on the use of gammaretroviral vectors and for the future design of novel gene transfer approaches for genetic correction of haematopoietic cells.

MATERIALS AND METHODS

Patients and clinical trials

ADA-SCID lacking HLA-identical sibling donors were enrolled in the clinical trials. PBL-GT patients were enrolled between 1992 and 1998 in a phase I/II clinical trial with repeated infusions of transduced autologous PBL (#NCT00599781) PBL-GT Pt1, 2, and 3 have been previously described (Aiuti et al, 2002b; Bordignon et al, 1995) while Pt4 received a similar treatment (Aiuti et al, in preparation). HSC-GT patients were enrolled between 2002 and 2005 in a phase I/II clinical trial for ADA-SCID GT with transduced autologous CD34+ cells (#NCT00598481) and HSC-GT Pt4-7 were previously described (Aiuti et al, 2009). For both groups the GIADAL retroviral vector encoding ADA cDNA under the MLV long terminal repeat (LTR) promoter was used (Aiuti et al, 2002b). Vector production and transduction protocol have been previously described (Aiuti et al, 2002a,b, 2009). The clinical trials described in this study were approved by San Raffaele Scientific Institute Ethical Committee and Italian national regulatory authorities. All patients signed the informed consent to the experimental treatment and follow up analyses.

The paper explained

PROBLEM:

Retroviral vectors have been used as effective tools to transfer therapeutic genes for the treatment of haematological inherited disorders. Studies on how the vectors integrate into the host genome of different cell types and how gene-corrected cells engraft and survive long-term in patients are crucial to provide information on the safety and efficacy of GT approaches.

RESULTS:

We studied the properties of genomic integration sites of a gammaretroviral vector in haematopoietic cells from GT-treated patients affected by adenosine deaminase deficient-severe combined immunodeficiency (ADA-SCID), treated either with mature lymphocytes or haematopoietic stem cells, in clinical

contexts free from adverse events. We analysed the influence of target cell type on vector integrations both at the time of gene transfer and years after infusion in GT treated patients. Our study uncovered a cell-specific insertion profile of retroviral vector dependent on functional and transcriptional activity as well as on epigenetic and chromatin conformations of host genome.

IMPACT:

This work unveiled the genomic features influencing the 'physiological' integration site selection of a gammaretroviral vectors in two different target cell types, providing crucial information for the follow up of current and future GT trials based on genetic modifications of haematopoietic cells.

Cell purification and measurement of vector positive cells

In vitro samples were derived from a fraction of transduced T cells or CD34+ cells that were infused in patients, kept in culture for additional 3 days after transduction.

In vivo cell subsets were purified from peripheral blood of ADA-SCID patients as previously reported (Aiuti et al, 2007). To detect the presence of the vector in patient cells we extracted genomic DNA by QIAamp DNA Blood Mini kit or Micro kit (Qiagen) and we performed qPCR for vector positivity with primers specific for Neo^R reporter gene and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) as shown previously (Aiuti et al, 2002a; Cassani et al, 2008). Percentage of transduced cells were calculated on the basis of a standard curve and expressed as proportion of cells carrying the Neo^R gene.

Analysis of vector integration

To collect integration sites from patient samples we performed LM-PCR with the use of *MseI* enzyme as elsewhere reported (Aiuti et al, 2007; Wu et al, 2003). Fasta sequences from insertions previously retrieved with the same technique in CD3+ cells of HSC-GT Pt4 and Pt5 were also added to the analysis (Aiuti et al, 2007). We also set up another protocol on the basis of LAM-PCR technique previously described (Schmidt et al, 2007) which consisted in two steps of linear amplifications using two 5'-biotinylated primers designed in forward direction on MLV LTR (MLV1: 5'-GACTGAGTCGCCCGGTACCCGTGT-3' and MLV2: 5'-CCAATAAACCTCTTGCACTTGCA-3') under the following conditions: 95°C, for 5 min, 50 cycles at 95°C for 45 s, 60°C for 45 s and 72°C for 1.5 min, and a final step at 72°C for 10 min. After ligation o/n with streptavidine magnetic beads (Invitrogen Dynabeads) linear amplified products went through a Klenow-mediated second strand reconstitution, *MseI* digestion and linker ligation. Fragments were then detached from beads with 15 min incubation with NaOH 0.1 M at 25°C, exponentially amplified with primers for LTR and linker as in LM-PCR protocol and sanger sequenced. In addition, we retrieved integration sites through the combination of LAM-PCR with the use of up to four enzymes (TSP509I, Hinp1I, HpyCH4IV and *MseI*) and 454-pyrosequencing (Roche) technique as described in a previous work

(Bushman et al, 2008; Howe et al, 2008; see also Supporting Information 'Integration sequences from PBL-GT and HSC-GT', 'List of CIS from HSC-GT', 'List of CIS from PBL-GT'). In total, 6% of the integrations (137 out of 2198) from PBL-GT and 2% of the integrations (47 out of 1959) from HSC-GT were retrieved with *MseI* enzyme only, while the majority derived from the use of four restriction enzymes in both the trials.

Bioinformatics and statistical analysis

Details of bioinformatics and statistical methods used for the mapping of vector integrants and the analysis of hit gene functions as well as correlations of vector insertions with expression profile and epigenetic statuses of target cells are available in Supporting Information 'Bioinformatics and Statistical analysis'. Expression data from microarrays experiments are deposited on ArrayExpress public database (<http://www.ebi.ac.uk/arrayexpress/>).

Author contributions

LB designed the project, performed experiments, analysed data and wrote the manuscript. AAm, DP and CDS performed bioinformatics and statistical analyses. IB performed patients' samples purifications. CB, CK and MS revised the manuscript and performed integration sites retrieval. MGR provided scientific support and revised the manuscript. AAi supervised the project, wrote and revised the manuscript.

Acknowledgements

We are grateful to the Italian Telethon Foundation and the European Commission [Concerted Safety and Efficiency Evaluation of Retroviral Transgenesis in Gene Therapy of Inherited Diseases (CONSERT) and CliniGene] for financial support. We also thank Dr. Kevin Goudy for his kind revision of the manuscript.

Supporting information is available at EMBO Molecular Medicine online.

The authors declare that they have no conflict of interest.

For more information

European Society of Gene and Cell Therapy:

<http://www.esgct.eu/>

American Society of Gene and Cell Therapy:

<http://www.asgct.org/>

Telethon Foundation:

<http://www.telethon.it/english/default.aspx>

Clinigene-NoE European Network:

<http://www.clinigene.eu/>

The San Raffaele Telethon Institute for Gene Therapy (HSR-TIGET):

http://www.fondazione.sanraffaele.it:80/ENG-home/research/divisions_centers_institutes_research_programs/TIGET/index.html

ADA-SCID on Orphanet:

[http://orphanet.orpha.net/consor/cgi-bin/Disease_Search.php?lng=EN&data_id=8023&disease\(s\)/group%20of%20diseases=Severe-combined-immunodeficiency-due-to-adenosine-deaminase-deficiency&title=Severe-combined-immunodeficiency-due-to-adenosine-deaminase-deficiency&search=Disease_Search_Simple](http://orphanet.orpha.net/consor/cgi-bin/Disease_Search.php?lng=EN&data_id=8023&disease(s)/group%20of%20diseases=Severe-combined-immunodeficiency-due-to-adenosine-deaminase-deficiency&title=Severe-combined-immunodeficiency-due-to-adenosine-deaminase-deficiency&search=Disease_Search_Simple)

Accompanying Closeup:

<http://dx.doi.org/10.1002/emmm.201000110>

References

- Aiuti A, Slavina S, Aker M, Ficara F, Deola S, Mortellaro A, Morecki S, Andolfi G, Tabucchi A, Carlucci F *et al* (2002a) Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning. *Science* 296: 2410-2413
- Aiuti A, Vai S, Mortellaro A, Casorati G, Ficara F, Andolfi G, Ferrari G, Tabucchi A, Carlucci F, Ochs HD *et al* (2002b) Immune reconstitution in ADA-SCID after PBL gene therapy and discontinuation of enzyme replacement. *Nat Med* 8: 423-425
- Aiuti A, Cassani B, Andolfi G, Mirole M, Biasco L, Recchia A, Urbinati F, Valacca C, Scaramuzza S, Aker M *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, Cattaneo F, Galimberti S, Benninghoff U, Cassani B, Callegaro L, Scaramuzza S, Andolfi G, Mirole M, Brigida I *et al* (2009) Gene therapy for immunodeficiency due to adenosine deaminase deficiency. *N Engl J Med* 360: 447-458
- Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129: 823-837
- Bordignon C, Notarangelo LD, Nobili N, Ferrari G, Casorati G, Panina P, Mazzolari E, Maggioni D, Rossi C, Servida P *et al* (1995) Gene therapy in peripheral blood lymphocytes and bone marrow for ADA-immunodeficient patients. *Science* 270: 470-475
- Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE (2008) High-resolution mapping and characterization of open chromatin across the genome. *Cell* 132: 311-322
- Brady T, Agosto LM, Malani N, Berry CC, O'Doherty U, Bushman F (2009) HIV integration site distributions in resting and activated CD4+ T cells infected in culture. *AIDS* 23: 1461-1471
- Bushman FD (2007) Retroviral integration and human gene therapy. *J Clin Invest* 117: 2083-2086
- Bushman FD, Hoffmann C, Ronen K, Malani N, Minkah N, Rose HM, Tebas P, Wang GP (2008) Massively parallel pyrosequencing in HIV research. *AIDS* 22: 1411-1415
- Cassani B, Mirole M, Cattaneo F, Benninghoff U, Hershfield M, Carlucci F, Tabucchi A, Bordignon C, Roncarolo MG, Aiuti A (2008) Altered intracellular and extracellular signaling leads to impaired T-cell functions in ADA-SCID patients. *Blood* 111: 4209-4219
- Cassani B, Montini E, Maruggi G, Ambrosi A, Mirole M, Selleri S, Biral E, Frugnoli I, Hernandez-Trujillo V, Di Serio C *et al* (2009) Integration of retroviral vectors induces minor changes in the transcriptional activity of T cells from ADA-SCID patients treated with gene therapy. *Blood* 114: 3546-3556
- Cattoglio C, Facchini G, Sartori D, Antonelli A, Miccio A, Cassani B, Schmidt M, von Kalle C, Howe S, Thrasher AJ *et al* (2007) Hot spots of retroviral integration in human CD34+ hematopoietic cells. *Blood* 110: 1770-1778
- Ciuffi A (2008) Mechanisms governing lentivirus integration site selection. *Curr Gene Ther* 8: 419-429
- Cui K, Zang C, Roh TY, Schones DE, Childs RW, Peng W, Zhao K (2009) Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell Stem Cell* 4: 80-93
- Deichmann A, Hacein-Bey-Abina S, Schmidt M, Garrigue A, Brugman MH, Hu J, Climm H, Gyapay G, Prum B, Fraser CC *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
- Felice B, Cattoglio C, Cittaro D, Testa A, Miccio A, Ferrari G, Luzi L, Recchia A, Mavilio F (2009) Transcription factor binding sites are genetic determinants of retroviral integration in the human genome. *PLoS ONE* 4: e4571
- Hacein-Bey-Abina S, von Kalle C, Schmidt M, Le Deist F, Wulffraat N, McIntyre E, Radford I, Villeval JL, Fraser CC, Cavazzana-Calvo 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, McCormack MP, Wulffraat N, Leboulch P, Lim A, Osborne CS, Pawliuk R, Morillon E *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 GP, Soulier J, Lim A, Morillon E, Clappier E, Caccavelli L, Delabesse E, Beldjord K *et al* (2008) Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J Clin Invest* 118: 3132-3142
- Howe SJ, Mansour MR, Schwarzwaelder K, Bartholomae C, Hubank M, Kempinski H, Brugman MH, Pike-Overzet K, Chatters SJ, de Ridder D *et al* (2008) Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *J Clin Invest* 118: 3143-3150
- Kohn DB, Candotti F (2009) Gene therapy fulfilling its promise. *N Engl J Med* 360: 518-521
- Kustikova OS, Geiger H, Li Z, Brugman MH, Chambers SM, Shaw CA, Pike-Overzet K, de Ridder D, Staal FJ, von Keudell G *et al* (2007) Retroviral vector insertion sites associated with dominant hematopoietic clones mark "stemness" pathways. *Blood* 109: 1897-1907
- Kustikova OS, Schiedmeier B, Brugman MH, Stahlhut M, Bartels S, Li Z, Baum C (2009) Cell-intrinsic and vector-related properties cooperate to determine the incidence and consequences of insertional mutagenesis. *Mol Ther* 17: 1537-1547
- Lewinski MK, Yamashita M, Emerman M, Ciuffi A, Marshall H, Crawford G, Collins F, Shinn P, Leipzig J, Hannehalli S *et al* (2006) Retroviral DNA integration: viral and cellular determinants of target-site selection. *PLoS Pathog* 2: e60
- Maes J, Maleszewska M, Guillemin C, Pflumio F, Six E, Andre-Schmutz I, Cavazzana-Calvo M, Charron D, Francastel C, Goodhardt M (2008) Lymphoid-affiliated genes are associated with active histone modifications in human hematopoietic stem cells. *Blood* 112: 2722-2729
- Maruggi G, Porcellini S, Facchini G, Perna SK, Cattoglio C, Sartori D, Ambrosi A, Schambach A, Baum C, Bonini C *et al* (2009) Transcriptional enhancers induce insertional gene deregulation independently from the vector type and design. *Mol Ther* 17: 851-856
- Mitchell RS, Beitzel BF, Schroder AR, Shinn P, Chen H, Berry CC, Ecker JR, Bushman FD (2004) Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol* 2: E234

- Montini E, Cesana D, Schmidt M, Sanvito F, Ponzone M, Bartholomae C, Sergi L, Benedicenti F, Ambrosi A, Di Serio C *et al* (2006) Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration. *Nat Biotechnol* 24: 687-696
- Montini E, Cesana D, Schmidt M, Sanvito F, Bartholomae CC, Ranzani M, Benedicenti F, Sergi LS, Ambrosi A, Ponzone M *et al* (2009) The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy. *J Clin Invest* 119: 964-975
- Naldini L (2009) Medicine. A comeback for gene therapy. *Science* 326: 805-806
- Newrzela S, Cornils K, Li Z, Baum C, Brugman MH, Hartmann M, Meyer J, Hartmann S, Hansmann ML, Fehse B *et al* (2008) Resistance of mature T cells to oncogene transformation. *Blood* 112: 2278-2286
- Ott MG, Schmidt M, Schwarzwaelder K, Stein S, Siler U, Koehl U, Glimm H, Kuhlcke K, Schilz A, Kunkel H *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
- Recchia A, Bonini C, Magnani Z, Urbinati F, Sartori D, Muraro S, Tagliafico E, Bondanza A, Stanghellini MT, Bernardi M *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
- Schmidt M, Schwarzwaelder K, Bartholomae C, Zaoui K, Ball C, Pilz I, Braun S, Glimm H, von Kalle C (2007) High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). *Nat Methods* 4: 1051-1057
- Wang GP, Ciuffi A, Leipzig J, Berry CC, Bushman FD (2007) HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res* 17: 1186-1194
- Wang GP, Berry CC, Malani N, Leboulch P, Fischer A, Hacein-Bey-Abina S, Cavazzana-Calvo M, Bushman FD (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
- Wei G, Wei L, Zhu J, Zang C, Hu-Li J, Yao Z, Cui K, Kanno Y, Roh TY, Watford WT *et al* (2009) Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating CD4+ T cells. *Immunity* 30: 155-167
- Wu X, Li Y, Crise B, Burgess SM (2003) Transcription start regions in the human genome are favored targets for MLV integration. *Science* 300: 1749-1751
- Xi H, Shulha HP, Lin JM, Vales TR, Fu Y, Bodine DM, McKay RD, Chenoweth JG, Tesar PJ, Furey TS *et al* (2007) Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS Genet* 3: e136