Integration of retroviral vectors induces minor changes in the transcriptional activity of T cells from ADA-SCID patients treated with gene therapy

Barbara Cassani1*, Eugenio Montini1*, Giulietta Maruggi2, Alessandro Ambrosi2, Massimiliano Mirolo1, Silvia Selleri1, Erika Biral1, Ilaria Frugnoli1, Vivian Hernandez-Trujillo4, Clelia Di Serio3,5, Maria Grazia Roncarolo1,5, Luigi Naldini1,5, Fulvio Mavilio2,6 and Alessandro Aiuti1,7.

(1) San Raffaele Telethon Institute for Gene Therapy (HSR-TIGET), Milan, Italy; (2) Dep. of Biomedical Sciences, Univ. of Modena and Reggio Emilia, Modena, Italy; (3) CUSSB, Istituto Scientifico H. San Raffaele, Milan, Italy; (4) Division of Allergy and Immunology Miami Children’s Hospital, USA; (5) Vita-Salute San Raffaele University, Milan, Italy; (6) IIT Unit of Molecular Neuroscience, Istituto Scientifico H. San Raffaele, Milan, Italy; (7) University of Rome “Tor Vergata”, Italy.

* BC and EM equally contributed to the manuscript.

Running title: Influence of Retroviral Integration on Gene Expression

Correspondence should be addressed to: Alessandro Aiuti, HSR-TIGET, via Olgettina 58, 20132 Milan, Italy.

E-mail: a.aiuti@hsr.it
Abstract

Gene transfer into HSCs by gammaretroviral vectors (RV) is an effective treatment for inherited blood disorders, although potentially limited by the risk of insertional mutagenesis. We evaluated the genomic impact of RV integration in T lymphocytes from ADA-SCID patients 10 to 30 months after infusion of autologous, genetically-corrected CD34+ cells. Expression profiling on ex vivo T-cell bulk population revealed no difference with respect to healthy controls. To assess the effect of vector integration on gene expression at single cell level, primary T-cell clones were isolated from two patients. T-cell clones harboured either one (89.8%) or two (10.2%) vector copies per cell and displayed partial to full correction of ADA expression, purine metabolism and TCR-driven functions. Analysis of retroviral integration sites (RIS) indicated a high diversity in T-cell origin, consistently with the polyclonal TCR-Vbeta repertoire. Quantitative transcript analysis of 120 genes within a 200kb-window around RIS showed modest (2.8- to 5.2-fold) disregulation of 5.8% genes in 18.6% of the T-cell clones compared to controls. Nonetheless, affected clones maintained a stable phenotype and normal in vitro functions. These results confirm that RV-mediated gene transfer for ADA-SCID is safe, and provide crucial information for the development of future gene therapy protocols. The trials described herein have been registered at http://www.clinicaltrials.gov under identifiers NCT00598481 and NCT00599781.
Introduction

Gene transfer into hematopoietic stem/progenitor cells (HSC) by gammaretroviral vectors has been demonstrated an effective treatment for inherited blood disorders, including X-linked severe combined immunodeficiency (X1-SCID)\(^1,2\), adenosine deaminase-deficient SCID (ADA-SCID)\(^3\) and X-linked Chronic Granulomatous Disease (X-CGD)\(^4\). However, following the occurrence of leukemic proliferation in 5 out of 19 SCID-X1 patients\(^5,6\) and of clonal expansion in 2 CGD patients\(^7\), the safety of the use of retroviral vectors turned out to be a primary concern. Since the viral RNA genome is reverse transcribed into a DNA copy that becomes integrated into the host cell genome, RV may induce cell transformation by insertional mutagenesis leading to proto-oncogene trans-activation, in combination to transgene- and disease-specific co-factors.

The genomic features driving the preferential integration of retroviral vectors have been described in several large-scale survey of mapping studies in mice, non-human primates and humans enrolled in clinical trials\(^8-14\). These studies have shown that RV integrate non-randomly in the host genome favoring transcription start sites and expressed genes, which makes productive interactions between the vector and host transcriptional machinery more likely. As a result of this bias, integrations which alter the expression of flanking genes involved in cell cycle, transcriptional activity, and signal transduction, might influence the biological fate of the affected cell clone, thereby conferring a growth or survival advantage\(^4,15,16\). However, to gain relevant information on the influence of vector integrations on transcriptional activity of nearby genes, studies should be carried out in relevant primary cells at the level of a clonal population containing defined insertion sites. Previous work conducted in the context of gene therapy with retrovirally transduced peripheral blood T cells revealed that RV insertion induced over-expression of a proximal gene at a frequency of almost 20%, although without measurable consequences on the biology and function of infused cells\(^17\).

We showed previously that promoters and transcriptionally active genes are preferential sites of RV integration also in BM-derived, transduced CD34\(^+\) progenitor cells, and in their progeny isolated ex vivo from ADA-SCID patients. We also found that, in an ADA-deficient context, selective growth advantage of ADA\(^+\) lymphocytes is associated with a significant bias towards integrations in genomic regions allowing a sustained transgene expression. Nevertheless, infusion of ADA-transduced HSCs is safe long-term and does not result in \textit{in vivo} clonal expansion or malignant transformation up to 8 years after treatment\(^11,18\).
Here, we report a detailed analysis aimed at fully assessing the potential transcriptional interference caused by vector integration, in both bulk populations and single clones of transduced T cells. Our data show that RV integration induce minor changes in the transcriptional activity of nearby endogenous genes at clonal level, which, however, does not impact on the biology of the T-cell clones. In addition, they provide crucial information for the development of future gene therapy protocols.
Methods

Patients

The ADA-SCID GT clinical trial was approved by the San Raffaele Scientific Institute and Hadassah University Ethical Committees and the respective National Regulatory authorities. The clinical trials are registered at ClinicalTrials.gov (numbers NCT00598481 and NCT00599781). The patients’ characteristics and treatment have been described previously. Patients received a single infusion of autologous CD34+ cells transduced with GIADAl vector. Blood samples from ADA-SCID patients and healthy donors were obtained after informed consent, in accordance with the Declaration of Helsinki, following standard ethical procedure and with approval of the Ethical Committee.

T-cell cloning and manipulation

Mononuclear cells from peripheral blood of Pt2 and Pt3 were isolated by Ficoll-Hypaque gradient separation and enriched for T lymphocytes by immunomagnetic-based depletion of monocytes (CD14+), B (CD19+) and NK (CD56+) cells. Individual T-cell clones were obtained by plating cells in 96-well plates at a concentration of 0.3-1-3 cells/well. Cells were stimulated with 1 μg/ml PHA (Roche Diagnostics Corp.), 100IU/ml rhIL-2 (Chiron Corp., USA) in the presence of an allogeneic feeder mixture of 1x10^5/well PBMC (X-ray irradiated at 60 Gy) and 10⁴/well JY cells (X-ray irradiated at 100 Gy). The cultures were performed in IMDM (Cambrex Bio Science, Belgium) supplemented with 10% YSSEL medium (Dyaclone, France), 5% Hyclone (Cambrex Bio Science) and 100U/ml penicillin/streptomycin (Bristol-Myers Squibb, Italy).

Under these conditions, cloning efficiencies were 1.2% for Pt2, 1.9% for Pt3, and 5.1% for healthy donor. T-cell clones were expanded by adding fresh medium with 600IU/ml rhIL-2 and restimulated as described every 14 days. The generation of T-cell clones at sufficient numbers for performing experiments required 4 cycles of restimulation. Phenotypic analysis, qRT-PCR, biochemical and functional assays were performed in resting T-cell clones sampled following 4-6 cycles at day 14 after restimulation.

Analysis of T-cell phenotype and function

Anti-CD4, -CD8, -TCRαβ, -TCRγδ mAbs directly coupled with FITC or PE were from BD Bioscience. Expression of ADA protein was assessed by intracellular staining using an anti-human ADA mAb, as described. To assess TCR b-chain usage, T cells (2x10⁵) were stained with a mix of directly conjugated
TCR Vbeta Abs corresponding to 24 different specificities, according to manufactuary instructions (IOTest Beta Mark kit; Immunotech, France). Samples were acquired using FACSCalibur flow cytometer (Becton Dickinson) and analyzed with FlowJo software (TreeStar Inc.). Intracellular ADA and SAHH enzyme activities were analyzed in cell lysate (0.5x10^6 cells) as described 19,20.

Resting T-cell clones were seeded at 1x10^5 cells/well in a 96-well plate round-bottom pre-coated with the indicated dose of anti-CD3 mAb (OKT3, Janssen Cilag, Ltd.) with or without soluble anti-CD28 mAb (10 μg/ml; BD Biosciences), in a final volume of 200 μl. Proliferation was assessed after 48h by [3H]thymidine incorporation. All experiments were performed in duplicate or triplicate wells. The secretion of cytokines was measured in the culture supernatants by luminex technique, according to the manufacturer’s instructions (Bio-Rad Laboratories, USA).

Analysis of RV integration sites

Genomic DNA was extracted from T-cell clones using QIAamp DNA Blood Mini kit (Qiagen Inc., USA). Inverse-PCR was performed as previously described 11. The resultant nested PCR products were either separated on 2% agarose excised and directly sequenced. Integration were mapped onto human genome by BLAST analysis using the ENSEMBL search engine (November 2004 freeze).

Quantitative tracking of specific integrants

Quantitative PCR analysis was performed on 50 ng genomic DNA, directly isolated from sorted CD3+ T cells, using the specific genomic flanking primers in combination with common LTR primer and probe (Primm, Italy). The primer sequence were as follows: 3’LTR primer, 5’-GTTTGCATCCGAATCGTGGT-3’; 3’LTR probe, 5’-6-FAM-TCTCCTCTGAGTGATTGACTACCCACGG-TAMRA-3’; 5’ LTR primer, 5’-CCTGACCTTGATCTGAACCTTCTCTCT-3’; ID-04 primer, 5’-TTGCCAGTCTAATGGGTGTG-3’; 5’ LTR probe, 5’-6-FAM-CACCTGTAGGTTTGGCAAGCTAGCTT-TAMRA-3’; ID-13 primer, 5’-CAGAGGTAGACAGAGAAGGCAT-3’; ID-16 primer, 5’-TGCACAGACTGATCAAAAGTCTCTAA-3’; ID-20 primer, 5’-CGGCTCGGGTGTCCTG-3’. The frequency of each individual clone, expressed as the frequency of CD3+ T cells containing the specific vector-genome junction sequence, was calculated on the basis of a standard curve of the relative integrants diluted in untransduced cells.
Gene Expression analysis

Transcriptional profiling was carried out in CD4+ and CD8+ T cells purified with immunomagnetic beads (Miltenyi Biotec, Germany) from the peripheral blood lymphocytes of ADA-SCID patients (Pt1, 3 and 4) 10-30 months after GT and age-matched healthy controls (n=4). The study was not performed in Pt2 due to limited availability of cells. RNA was isolated from 1x10⁶ cells using Eurozol reagent (Euroclone S.p.A., Italy). The transcribed biotinylated cRNA was fragmented and hybridized to Affymetrix HG-U133A GeneChip arrays for 16 hours (Affymetrix Inc., USA). Scanned images were processed by the Affymetrix MAS 5.0 suite. Data analysis was performed with the GeneSpring® GX 7.3.1 version (Agilent Technologies). Custom 384 wells TaqMan Low Density Arrays (LDA; Applied Biosystems) were used to assess potential transcriptional interference at the site of RV integration. LDA format was customize on-line with four replicates per target gene. A total of 135 probes (assays) corresponding to 120 single genes were included in this study. Samples were analyzed using the 7900HT system with a Taqman LDA Upgrade (Applied Biosystems) and SDS2.2 software, according to manufacturer’s instructions. Data were normalized to the housekeeping gene GAPDH. In case of genes analyzed with two probes the average of corresponding ΔCTs was considered. Relative quantification was calculated by using the ΔΔCt methods as relative to the average of ΔCTs in all samples as calibrator. The amount of target relative to the calibrator was 2^{-ΔΔCt}, and is indicated in Table 2 as fold changing.

Statistical analysis

In the gene expression study on ex vivo bulk populations, transcript levels have been compared by means t-type statistic (Welch statistic for not equal variances assumptions). p-values have been computed based on distribution free permutation procedures and corrected for multiplicity by means of Benjamini and Hochberg False Discovery Rate. In details the permutation procedure was considered for each single RIS to detect significant difference between the expression of gene(s) hit by RV in the target cell clone(s) and that of control clones. The computation of p-values by means of this re-sampling non parametric procedure is not affected from the sparness of observations and the different size of the two groups of interest, what, in standard parametric methods (like t-test) could lead to strong biases in estimating the false positives due to large variances that would artificially spread up the t-values. In functional assays, a two-tailed Mann
Whitney U test was used for statistical comparison of patients vs healthy controls. Responses within the same group were compared by the (two-tailed) Wilcoxon signed rank test for paired data.
Results

Normal gene expression profile in T lymphocytes purified ex vivo from GT treated patients

Global gene expression profiling analysis of ex vivo PB T lymphocytes from ADA-SCID patients was performed to assess potential alterations in the transcriptional activity after GT. CD4⁺ and CD8⁺ T cells were isolated from three patients (Pt1, Pt3, and Pt4) 10 to 30 months after autologous transplantation with genetically corrected CD34⁺ cells. At the indicated time points, the percentage of vector-positive T cells by qPCR was >75%. Transcriptional profiles were determined on Affymetrix HG-U133A microarray, and compared to those of age-matched healthy controls. The microarray data has been deposited into the GEO public database under accession number GSE17354. Heat maps originated from CD4⁺ and CD8⁺ data sets are depicted in Figure 1. An unsupervised cluster analysis identified two main branches corresponding to the CD4⁺ and CD8⁺ cell type (not shown). Within each T-cell subset, samples from GT-treated patients did not form distinct clusters compared to healthy controls, indicating substantial overlapping between the gene expression profiles of the two groups (Fig. 1A-B).

Using a supervised hierarchical ordering approach, one-way analysis of variance was performed to search for genes which varied significantly between patients’ and control samples. We identified as low as 235 (2.5%) genes, among the CD4⁺ dataset, and 479 (5.1%) genes, among the CD8⁺ dataset, differentially expressed in the cluster of patients as compared to that of controls (Fig. 1C-D). However, when we applied a multiple testing corrections with the least stringent Benjamini and Hochberg False Discovery Rate procedure, no significant differences were observed in the gene expression patterns between the two groups (not shown). Such findings are in keeping with the lack of sample segregation according to the treatment, as observed by the clustering analysis.

These results indicate that our GT protocol causes no major alterations in the gene expression program of T lymphocytes developed from genetically-corrected CD34⁺ cells.

Analysis of retroviral vector integrations and diversity of T-cell clones.
To study the effect of vector integration on gene expression and cell function at clonal level, primary T-cell clones were isolated ex vivo by limiting dilution from PB lymphocytes of two ADA-SCID patients 18 months after GT 11 (Pt2 (X), 26 clones; Pt3 (Y), 27 clones). Among the clones generated from Pt2, four (15.4%) did not contain an integrated provirus, in agreement with the lower percentage (70%) of gene marking in CD3+ cells at the time of cell cloning. In all the vector-marked clones, vector-genome junctions were retrieved by inverse-PCR, sequenced and mapped onto the human genome by BLAST analysis. Overall, 41 RIS were unambiguously assigned to a chromosomal position, 26 of which from Pt3 and 15 from Pt2 (Suppl. Table1). The large majority of clones contained a single provirus per cell (n=44, 89.8%), with few clones carrying two proviruses (n=5, 10.2%), as confirmed by quantitative PCR.

The distribution of RIS with respect to the closest RefSeq genes is shown in Table 1, in comparison with the insertion profile described for pre-transplant CD34+ cells and for cells obtained ex vivo during the patients’ follow-up 11. Intergenic and promoter-proximal (± 5 kb from TSSs) RIS were over-represented in T-cell clones with respect to pre-GT and post-GT bulk samples (70.7% vs 49.1% and 58.7%, and 36.6% vs 23.6% and 28.8%, respectively). The characteristic cluster of RIS in gene-dense chromosomal regions was confirmed in T-cell clones (83% RIS in >10 genes/Mb regions; Fig. 2).

Phenotypically, all clones were CD4+ T cells and displayed a high diversity in the T-cell receptor (TCR) β chain usage (Fig. 3A), mirroring the complexity of the immune reconstitution occurred in patients. To track the in vivo origin of T-cell clones, we combined the RIS mapping with TCR Vβ analysis. In 5 cases we could identify T-cell clones harboring distinct Vβ TCR that shared the same vector insertion, indicating that they derived from the same pre-thymic progenitor (Fig. 3B and Suppl. Table 1). Interestingly, RIS identified in multiple cell clones (i.e. ID-04 and ID-16 found in five and three clones, respectively) are present at relatively higher frequency within the ex vivo isolated CD3+ population, thus demonstrating that RIS sampled in the T-cell clones are directly representative of the in vivo clonal composition at the time of cloning. Moreover, some of these T-cell clones shared the same RIS previously identified in CD15+ granulocytes and CD19+ B lymphocytes in the same patients 11, confirming their common origin from transduced multipotent HSCs (Fig. 3D).
Quantitative transcript analysis of genes targeted by retroviral vectors.

We next investigated any potential transcriptional perturbation of the genes surrounding RIS in prospectively isolated ex vivo T-cell clones. Adequate RNA material was available for 43 clones, corresponding to 35 different RV insertion sites. We studied the expression level of all genes contained in an interval of 200 kb centered on the site of RV integration. This broad interval was chosen because previous studies have suggested that insertional gene activation is not merely a function of the distance of the RV from the TSSs. Overall, we analyzed a total of 120 different genes (T-cell clone gene-set). The transcriptional profile of the T-cell clone gene-set was preliminary evaluated within the Affymetrix datasets of purified T cells from HD, to assess their relative expression with respect to the global average transcription levels. The proportion of expressed genes in the array did not increase significantly when the analysis was restricted to genes landed within 100 kb from RIS (43.3% vs 44.9%; Suppl. Figure 1A), indicating that no bias towards expressed genes was introduced with the chosen cut-off. Functional classification indicated that RV hit genes belonged to major networks involved in cellular process and metabolism groups, and without any particular skewing with respect to the expected distribution (Suppl. Fig 1B). Three genes (2.5%) are known proto-oncogenes, of which one (BLM) was identified as common insertion site (CIS) in our previously published collection.

Using custom Taqman Low Density Arrays (LDA), the expression of the genes hit by the vector was evaluated in the clone(s) in which the integration was identified and compared to the level of expression found in the other clones carrying integrations in different loci. Untransduced (UT) clones from Pt2 (n=4) as well as clones from healthy donors (HD, n=3) were used as additional controls. Normalized expression levels of genes neighbouring RV insertion sites are represented in Figure 4A-B. Since normal distribution of data was doubtful and normal approximations to obtain standard errors and confidence intervals cannot be correctly used, we employed a non-parametric permutation test based on re-sampling, which shifts the assignment of sample within each single RIS and re-computes the t statistic each time. This analysis revealed that 7 out of the 120 genes assayed (5.8%) had a significantly perturbed expression in the clones carrying a RV integration at a distance of <100 kb (Table 2 and Suppl. Table 2). ACSL4, MADD, IQGAP1, STAM and LOC57228 genes were upregulated (frequency 11.4%; 2.8-4.5 fold increase), whereas
C14orf159 and RUNDC2A appeared to be downregulated (frequency 5.7%; 4.7-5.2 fold decrease; a description of deregulating RIS’ genomic position is available in the Suppl. Table 1).

All the upregulated genes are functionally involved in metabolic and signaling pathways, and are expressed at low (ACSL4) or intermediate level (MADD, IQGAP1, STAM) in T cells, as assayed by Affymetrix profiling (Suppl. Fig 1 and data not shown).

To compare the genotoxic potential of RV vector insertion in T-cells in the absence of in vivo selection, we used the same quantitative assay and analysis criteria to re-analyzed perturbation of gene expression around RV insertion site in T-cell clones from one of our previous studies. By this analysis, 6/29 (20.7%) randomly selected clones showed significantly higher expression in 6/68 (8.8%) genes close to an integrated provirus. In non transplanted T-cells, activating RIS increased mRNA transcript level at a similar frequency but in a more prominent fashion, ranging from 3.2-fold (ZNF217) to more than 1,900-fold (LY64) the average value in the control clones (Fig. 4C and Suppl. Table 3). Deregulation of gene expression occurred preferentially when a provirus was integrated in close proximity of a TSS (≤50 kb; Suppl. Fig. 2) and in opposite transcriptional orientation, in both ex vivo and in vitro clones.

RIS-associated altered gene expression has no impact on phenotype and functional behavior of T-cell clones.

To evaluate the consequences of vector integration on the biology of T-cell clones, we first asked whether perturbation in neighboring gene transcription could influence the ADA transgene expression. ADA transcript analysis was measured with LDA in (i) clones carrying deregulating RIS (n=6), (ii) other transduced clones (n=37), (iii) untransduced clones from ADA-SCID patients (n=4) and (iv) control clones from healthy subjects (n=3) (Fig.5A). Relative ADA mRNA level was significantly lower in untransduced clones when compared to transduced and normal clones (p=0.023), whereas no difference in ADA expression was observed between the last two groups. Interestingly, clones with two proviruses generally showed higher ADA mRNA levels with respect to clones carrying one provirus, and closer to normal levels. Remarkably, deregulating RIS had no discernible effect on transgene expression in all the hit clones with the exception of clone Y7, which harbored two integration events (Fig.5A). Similar results were achieved when
ADA expression was evaluated at the protein level by intracellular FACS staining (Fig. 5B). Nonetheless, a more robust correlation was observed between the protein expression and the vector copy number, since only clones carrying two proviruses restored ADA to normal levels (p=0.036, Fig. 5B). A significant tendency to select for integrations in genomic regions potentially more favorable to vector-derived ADA expression has been previously reported for *ex vivo* purified bulk T cells. This tendency was further enhanced in *in vitro* expanded clones (Table 1), and confirmed by the higher protein expression detected in T-cell cultures (bulk T-cell line MFI:20.6; average MFI in T-cell clones: 24.2) compared to freshly isolated T cells from Pt3 (MFI: 12.2; Suppl. Fig.3A). Possible position effects were additionally evaluated in T-cell clones, after grouping according to the type of integration and genomic location harboring RIS. No difference in ADA expression was observed among clones carrying intragenic versus intergenic RIS, or RIS landing in high dense gene regions (not shown). However, when the analysis was restricted to clones with TSS-proximal (<5 kb) intergenic RIS, ADA protein levels showed a tendency to increase, although not statistically significant (Suppl. Fig.3B). Interestingly, hit genes were transcriptionally active, and enriched for the intermediate-high level of expression (not shown). Overall, these data confirm that productive interaction between the vector and host transcriptional machinery is one of the crucial parameters determining the selective advantage of T-cell clones expressing adequate ADA levels.

We then asked whether insertional dysregulation of genes flanking the retroviral insertion could confer preferential proliferative advantage to the targeted clones. None of the cell clones here analyzed showed a cytokine-independent *in vitro* growth during relatively long-term culture period. The functional behavior of cell clones was further analyzed upon TCR stimulation with plate-bound anti-CD3 mAb (OKT3). Proliferative responses of gene-corrected T clones were significantly increased with respect to those observed for untransduced ADA-deficient T-cell clones (p=0.0012), and reached the range of normality (Fig. 5C). Interestingly, three out of four clones carrying 2 VCN displayed the highest range of proliferation, further confirming the strong selective advantage conferred by ADA expression to T-cell function. The ability to produce both Th1 and Th2-like cytokines was also restored to normal level (Fig. 5D). Finally, it is worth noting that cell clones displaying significantly dysregulated gene expression did not show increased basal proliferation or aberrant functional responses upon mitogen stimulation.
ADA transgene expression correlates with normalization of purine metabolism.

We took advantage of the availability of untransduced, gene-corrected and normal T-cell clones to study at single cell level the effect of restored ADA expression on other enzymes involved in purine metabolism (Fig. 6A). Indeed, in the absence of ADA, increased amounts of adenosine (Ado) and deoxy-adenosine (dAdo) may “spill over” into additional intracellular metabolic pathways normally only minimally utilized \(^{22-24}\), thus contributing to the pathogenetic mechanisms of the disease.

First, we focused on S-Adenosylhomocysteine Hydrolase (SAHH) enzyme, as reduction of its activity has been extensively documented in cells from ADA-SCID patients and considered an hallmark of the disease \(^{19,25}\). Whereas no difference in the relative mRNA expression was detected between untransduced and gene-corrected or normal clones (Fig. 6B), a marked block in SAHH activity was observed in untransduced clones (\(p=0.029\) vs gene-corrected clones, \(p=0.012\) vs normal clones; Fig.6C). In contrast, normalization of SAHH activity was detected in gene-corrected cells (Fig. 6C). Of note, T-cell clones carrying two proviruses showed the best recovery of enzyme function, in keeping with the highest expression of ADA protein (Fig. 6C).

Although dAdo is a weak substrate for Adenosine Kinase (AK) and Deoxycytidine Kinase (DCK), in the absence of ADA these enzymes can phosphorylate dAdo. In turn, the resulting dATP pool expansion may interfere with a number of metabolic pathways critically required for cell viability \(^{26}\). Similarly, Ado phosphorylation-induced pyrimidine depletion has been one of the first mechanism proposed as the cause of lymphopenia in ADA deficiency \(^{27}\).

Remarkably, expression analysis in our cohort of T-cell clones revealed that mRNA transcript levels of both kinases were significantly upregulated (2 and 2.6 fold increase for DCK and AK, respectively; Fig. 6D-E) in untransduced as compared to gene-corrected and normal clones (\(p=0.0293\) for DCK; \(p=0.014\) for AK), strongly supporting the cooperative role of DCK and AK-mediated nucleoside phosphorylation in the intracellular pathogenetic mechanisms of the disease.
Discussion

The great therapeutic potential of retroviral-mediated gene transfer is accompanied by the inherent risk of insertional mutagenesis through deregulation of host gene transcription. Indeed, gene therapy has been shown to be a beneficial treatment for patients affected by SCID-X1 and CGD but seriously limited by the emergence of leukemic proliferation (5 out of 19 treated SCID-X1 patients) and clonal expansion of myeloid cells (2 CGD patients) associated with RV insertions near cellular proto-oncogenes. In the context of ADA-SCID gene therapy, the cumulative experience of our study of 10 patients as well as of other clinical trials did not unveil such complications, indicating that retroviral-mediated gene transfer for ADA-SCID retains a favorable safety profile. Consistently, our previous extensive sequencing analysis of RIS in multiple hematopoietic lineages revealed the typical genomic preferences of gammaretroviral vectors, but the lack of in vivo skewing for sensitive genes. Here, we have performed a comprehensive study of the influence of vector integration on gene expression in primary bulk populations and in single T-cell clones obtained from ADA-SCID patients transplanted with autologous, transduced CD34+ HSCs. A global gene expression profiling was performed on CD4+ and CD8+ T-cell subsets purified ex vivo from three ADA-SCID patients at different times after GT. The microarray analysis showed a substantial overlap with the expression patterns of T cells from controls, indicating the absence of gross abnormalities in the development and function of T cells deriving from genetically corrected HSCs. However, gene expression studies in bulk cell populations may underscore vector-induced gene deregulation occurring only in a small fraction of transduced cells, unless the altered cell clones contribute to a relevant amount of the analyzed cell population. To overcome the limited resolution power of the microarray technique, and provide a direct estimate of the transcriptional interference caused by vector integration at single-cell level, we carried out a quantitative transcript analysis of genes close to integrated proviruses in mature T-lymphocyte clones isolated from two of the patients. T-cell clones expanded ex vivo from single-cells represent a unique and powerful experimental tool to directly establish a relationship between integration sites and host gene expression profile. A potential bias of this approach is that circulating T cells represent only a limited fraction of the whole body T-cell pool. Additionally, we cannot exclude that certain clones may be less prone to ex vivo stimulation and long-term culture and thus potentially negatively selected. The large majority of the clones contained a unique integration, demonstrating the heterogeneity of
T-cell population. Indeed, the few cell clones carrying the same vector integrant showed distinct patterns of TCR rearrangement, indicating the major contribution of polyclonal thymopoiesis in immune reconstitution rather than peripheral expansion. Moreover, the occurrence of shared integrants between lymphoid and myeloid cells further confirmed the engraftment of transduced HSC with multilineage potential.

The analysis of the RIS genomic distribution in T-cell clones reflected the known preferences of gammaretroviral vectors. Although not significant, an additional tendency to favor integrations in intergenic regions, and particularly in the proximity of TSSs, was observed in T-cell clones with respect to RIS retrieved from uncultured T-cell populations collected during the follow-up. The increased expression of ADA protein detected, particularly, in clones with RIS closer to TSS of active genes, suggests that such vector insertions profile could be biased by a selection in vitro for clones carrying RIS in genome sites more favorable for high level of vector-derived expression, without altering the transcription of the target gene(s) in the majority of cases. Beyond the position effects, it appears that a complete normalization of ADA activity in peripheral T cells was mainly dependent on two copies of ADA transgene inserted into HSC, leading to full correction of TCR-driven effector function.

In the absence of ADA, adenosine and deoxy-adenosine may be metabolized differently, and give rise to alternative pathogenetic pathways, ultimately causing the intracellular poisoning. Our data indicate that untransduced T-cell clones had significantly higher expression of the kinases in charge of the purine nucleoside phosphorylation, as compared to gene-corrected and healthy clones. Therefore, the increased conversion of ADA metabolites to nucleotide derives is likely responsible for the intracellular toxicity associated with SAHH inhibition, which persisted in vitro in ADA-deficient but not in corrected clones.

The quantitative transcript analysis in single cells revealed an incidence of transcriptional perturbation in 18.6% of the interrogated ex vivo clones, with 5.8% of genes assayed showing altered expression as consequence of a nearby vector insertion (7/35, frequency 20%). Enhanced gene expression, due to the transcriptional read-through from the nearby LTR or the activity of viral enhancer present in the LTR, was observed in 4 out of the 120 genes analyzed (3.3%), whereas 1 gene (0.83%) showed actual insertional activation in two independently isolated clones sharing the same RIS. Interestingly, in two cases (1.6%) the integrated vector caused down-regulation of gene expression. Reduced transcript accumulation might be the
consequence of aberrant splicing and/or premature transcript termination due to viral polyadenilation signals in intronic integrations, or by transcriptional interference in the case of intergenic insertions.

The extent of dysregulation in the expressed genes ranged from 2.8 to 5.2-fold the average value detected in control clones, whereas in case of putative RIS-mediated insertional activation the expression level of target gene was at the limit of detection in quantitative PCR for both the cellular clones. The Affymetrix analysis performed on polyclonal population of Pt3, purified 8 months after the cell cloning, revealed a remarkable transcriptional identity with the expression profile detected in other GT-treated patients and healthy controls, with none of the genes deregulated by RIS at the clonal level showing significant difference. Hence, minor alterations of gene expression at single cell level have no discernible effect on the bulk population. Interestingly, the extent of altered gene expression detected in our cohort of *ex vivo* derived clones was on average 100-fold lower compared to that induced by RIS in T-cells cloned transduced *in vitro* and not subjected to *in vivo* selection. It is also worth noting that an upregulation of more than 10 fold was detected for *LMO2* and *CCND2* in leukemic cells (with respect to control T cells or thymocytes) from patients of the SCID-X1 trial 5,6 and for MDS-EVI1 (36-74 fold increase) in CGD patients 4.

Overall, our results suggest that in the ADA-SCID context dangerous insertional activations, if any, result in a competitive disadvantage, leading to extinction rather than selection of the affected clones. An alternative and non-mutually exclusive hypothesis for the lack of dominant/malignant clones in ADA-SCID disease settings is that the selective advantage conferred by the restoration of ADA expression is strong and acts on a wide number of cell types. Thus, any possible additional proliferative advantage conferred by vector integrations near oncogenes will be not relevant in an environment in which virtually all transduced cells compete equally in repopulating the host 32. Previous studies on mice transplanted with RV-transduced bone marrow have proposed a causal link between insertional deregulation and enhancement of cell “fitness”. Accordingly, dominant clones were consistently found to have increased expression of a gene, within the region of the integrated RV, that could be implicated in potentially causing survival or proliferative advantage, even in the lack of malignant transformation 15,16. The murine system, however, might be biased by the strong selection process determined by serial transplantation and repopulation experiments, which could cause *per se* the emergence of clones displaying altered expression of sensitive genes. In the case of the X-SCID and CGD cells the nature of the therapeutic transgene or of the particularly strong LTR enhancer
could have played a role in the selection of malignant or pre-malignant cell clones carrying integrations into proto-oncogenes \cite{33,35}. In our study, the majority of the genes showing vector-induced transcriptional perturbation belonged to "housekeeping" pathways in T cells and, remarkably, none of them is a known proto-oncogene. Based thereon, overexpression of such genes induced by the nearby vector insertion may theoretically confer a growth advantage to the affected clones. Although we were unable to directly assess the effect of increased gene transcription at the protein level, due to the reduced availability and limited life span of T-cell clones, the stable phenotype and the normal functional behavior observed in the target clones ruled out the possibility that any transformation event has occurred. Accordingly, the polyclonal pattern of vector integrations and T-cell repertoire, and the inability to retrieve these insertions by random cloning in the \textit{ex vivo} and \textit{in vitro}-expanded bulk populations at later time points (11 and unpublished observations) suggest that these clones do not represent a major fraction of the cell populations tested.

In summary, beside the actual risk for RV-mediated perturbation in the host cellular transcriptional activity, these results indicate that GT for ADA-SCID retains a low risk profile and that disease-specific co-factors need to be accounted in future designing of clinical trial \cite{33,36,38}.

In this regard, the construction of so-called self-inactivating (SIN) vectors should allow to further reduce the risk of insertional gene activation \cite{39,41}. Different approaches would be to incorporate suicide genes in therapeutic vector \cite{42}, or insulators \cite{43}. However, the impact of such modifications on vector safety remains speculative and it has to be demonstrated in appropriate pre-clinical models.
Acknowledgments

We gratefully acknowledge Dr. Filippo Carlucci and Dr. Antonella Tabucchi from University of Siena for biochemical assays. Dr. Samantha Scaramuzza for technical assistance and helpful discussions.

Funding

Supported by grants from the Italian Telethon Foundation (HSR-TIGET and GGP06101), from Association Française contre les Myopathies–Telethon (GAT0205), the independent drug research program of the Italian Medicines Agency (AIFA) (FARM5JRXRM), and the European Commission (Concerted Safety and Efficiency Evaluation of Retroviral Transgenesis in Gene Therapy of Inherited Diseases [CONSERT] LSHB-CT-2004-005242 and Clinigene LSHB-CT2006-018933).

Authorship

Contribution: BC designed and performed research experiments, analyzed data, and wrote the paper; EM contributed to research design, acquisition and data analysis, and revised the paper; GM contributed to research experiment and data analysis; AAmbsori performed statistical analysis; MM and SS performed research experiments; EB, IF and VHT provided clinical care of the patients; CDS performed statistical analysis; MGR and LN supervised the research and revised the paper; FM contributed to study design and revised the paper; AAiuti designed and supervised the research and wrote the paper; all authors checked the final version of the manuscript.

Competing Interests

The authors declare that no competing interests exist.

Abbreviations: ADA, adenosine deaminase; SCID-X1, X-linked SCID; CGD, chronic granulomatous disease; GT, gene therapy; PB, peripheral blood; RIS, retroviral insertion site; Pt, patient; TSS, transcription start site; VCN, vector copy number, TCR, T-cell receptor.
References


Figure Legend.

Figure 1. Gene expression profile in purified T-cell subset from gene therapy treated patients. Heat maps representing the unsupervised Affymetrix gene expression analysis of peripheral blood CD4+ (A) and CD8+ (B) T cells isolated from three gene therapy-treated patients (Pt1, +30 mo; Pt3, +26 mo; Pt4, +10 mo) and four age-matched healthy controls (c). Expression levels are represented by a colour key in which bright red represents the highest levels (upregulated genes) and bright blue represents the lowest levels (downregulated genes), less saturated shades represent intermediated levels of expression. Heat maps resulting from supervised hierarchical clustering of CD4+ (C) and CD8+ (D) T-cell samples according to the type of treatment (i.e. GT versus normal controls). p-value cut-off 0.05, multiple testing correction: none. This restriction tested 9.364 genes. About 468,2 genes would be expected to pass the restriction by chance.

Figure 2. Distribution of vector integration site in ex vivo T-cell clones. Distribution of integration sites, retrieved from T-cell clones isolated ex vivo from Pt2 (X, n=22) and Pt3 (Y, n=27), relative to gene density within a 1Mbp window, as compared to insertion sites retrieved from ex vivo and in vitro cell samples, as reported in Aiuti et al.,2007 (*). Each bar corresponds to the percentage of integration sites within the corresponding gene density region.

Figure 3. Analysis of TCR diversity and haematopoietic origin of T-cell clones. (A) Immune repertoire of T-cell clones isolated from the peripheral blood of Pt2 and Pt3. Each bar corresponds to the percentage of clones displaying the specific TCR Vβ chain rearrangement, as analyzed by flow cytometry. (B) Analysis of Vβ repertoire in three clones from Pt3 sharing a common RIS. (C) Quantification of specific RIS-containing clones (ID-04 and ID-13 from Pt2 identified in five and two independent T-cell clones; ID-16 and ID-30 from Pt3 identified in three and one T-cell clone, respectively) within the total T-cell population. The frequency of insertions was measured by sequence-specific real-time PCR in CD3+ T cells purified from PB at different time points (months) during the follow up after GT. Rectangle highlights the time point at which the T-cell cloning was performed. (D) T-cell clones from Pt3 share RIS with distinct hematopoietic cell subsets. Integration analysis was performed, by random cloning and sequence-specific
PCR, on cells purified ex vivo from peripheral blood (PB) or bone marrow (BM) at different time points (months, mo) during the follow up.

Figure 4. Expression analysis of gene targeted by RIS in T-cell clones. Quantitative transcript analysis of genes landed in a 200kb window around RV integration site in individual T-cell clones isolated ex vivo from Pt2 (A) and Pt3 (B), as performed by real-time RT-PCR on low density arrays. Only genes found expressed by the majority of clones (≥ 95%) are depicted. (C) Quantitative PCR analysis of the expression of genes hit by RV integration in randomly selected T-cell clones transduced in vitro. Expression level (y axis) of each gene (x axis) hit by the nearby RIS, measured as relative mRNA quantity after normalization for the level of the housekeeping gene GAPDH (ΔCt), is presented as a dot for the target clone(s) and by means of box-and-whisker plots for all control clones. Box-and-whisker plots illustrate the median, the interquartile range, and the range that contains the central 95% of the observation and the maximum and the minimum values. Red arrows and dots indicate the expression of the dysregulated gene in the T-cell clone containing the RIS and its relative p value.

Figure 5. Phenotypic and functional characterization of T-cell clones. (A) Relative mRNA expression of ADA in T-cell clones. Expression was measured with microfluidic card in UT clones from Pt2 (gray dots), clones from healthy donors (black dots), clones carrying upregulating (red dots) or downregulating (green dots) RIS, clones with 2 vector copy/cell (white dots) and 1 vector copy/cell (box and whiskers). (B) ADA protein expression. Expression was determined by intracytoplasmic FACS staining as described in methods. M.F.I.: Mean Fluorescence Intensity. (C) Proliferative responses and cytokine production. T-cell clones generated from patients and age-matched controls were stimulated with immobilized anti-CD3 monoclonal antibody (OKT3, 1 μg/ml) and 3H-Thymidine incorporation was assessed after 48h. In parallel, cytokine secretion was evaluated in the culture supernatants collected at different time points (18 hr for IL-2 and 48 hr for the other cytokines) (D). *p<0.05 and ** p<0.01 vs transduced clones.

Figure 6. Expression analysis of purine metabolic enzymes. (A) Illustration of purine pathway. HPRT, hypoxanthine phosphorybosil transferase; PNP, purine nucleoside phosphorylase; SAHH,
S-adenosyl homocysteine hydrolase; MeT, methyl transferase, AK, adenosine kinase; DCK, deoxycytidine kinase; RR, rebonucleotide reductase. (B) Relative mRNA expression of SAHH, as measured with q-RT-PCR, in UT clones from Pt2 (n=4; gray dots) and controls (transduced and healthy clones n= 46; box-and-whiskers). (C) SAHH enzymatic activity. Relative mRNA expression of DCK (D) and AK (E) in UT clones and controls. p values refer to gene expression of UT clones versus controls.
Table 1. RIS distribution in T-cell clones from ADA-SCID patients. Distribution of RV integration sites mapped in T-cell clones isolated ex vivo from the peripheral blood of Pt2 and Pt3 at 18 months after gene therapy, compared to RV integrations in pre-GT and post-GT bulk samples from a previously published collection (Aiuti et al., 2007). Integrations were distributed as inside (intragenic) or outside (intergenic) RefSeq genes, and at a distance of <10 kb upstream or ± 5 kb from TSSs.
Table 2. Genes with deregulated expression induced by the nearby RIS. List of genes displaying altered mRNA level in the ex vivo isolated T-cell clone(s) harbouring RIS. T-cell clones retrieved from Pt2 are indicated with X, clones from Pt3 with Y. Vector orientation is relative to targeted gene transcription. Relative gene expression level is indicated as normalized for the housekeeping gene (GAPDH) in each hit clone, and as fold change relative to the calibrator, as described in Methods. n.a., not applicable.
Figure 1

A

B

C

D

CD4⁺

CD8⁺

C1 C2 Pt4 C4 Pt1 C3 Pt3

C1 Pt3 C2 C4 C3 Pt4

C1 C2 C4 C3 Pt1 Pt3 Pt4

C1 C2 C3 C4 C3 Pt4
**Figure 3**

A

![Bar graph showing percentage of T-cell clones by TCR Vbeta Family.](image)

B

![Flow cytometry plots for Y23, Y27, and Y9 showing Vbeta expression.](image)

C

![Graph showing percentage of specific RIS in CD3 T cells over time after GT.](image)

D

![Follow-up timeline for ID_16 with CD3 PB, CD15 PB, CD15 BM, CD19 PB, and CD19 BM samples.](image)
Integration of retroviral vectors induces minor changes in the transcriptional activity of T cells from ADA-SCID patients treated with gene therapy

Barbara Cassani, Eugenio Montini, Giulietta Maruggi, Alessandro Ambrosi, Massimiliano Mirolo, Silvia Selleri, Erika Biral, Ilaria Frugnoli, Vivian Hernandez-Trujillo, Clelia Di Serio, Maria Grazia Roncarolo, Luigi Naldini, Fulvio Mavilio and Alessandro Aiuti