We are IntechOpen,
the world's leading publisher of Open Access books
Built by scientists, for scientists

4,000
Open access books available

116,000
International authors and editors

120M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Systemic Approach to the Genome Integration Process of Human Lentivirus

Felipe García-Vallejo and Martha Cecilia Domínguez

Additional information is available at the end of the chapter

1. Introduction

The human genome is one of the most complex molecular structures ever seen in nature. Its extraordinary information content has revealed a surprising mosaicims between coding and non-coding sequences [1-4]. This highly regionalized structure introduces complex patterns for understanding the gene structure and repetitive DNA sequence composition providing a new scenario to study biological process such as Lentivirus cDNA integration into host genome. In the field of genome analysis, bioinformatics provides the key connection between all different forms of data gathered by new high-throughput techniques such as systematic sequencing, expression arrays, and high throughput screenings among others. Although the success of bioinformatics in the genome analysis is undeniable, in some cases has complicated the relationship of computation with experimental biology. There is a need to attend to our pressing needs of bioinformatics applications without forgetting other, perhaps less evident but equally important, aspects of computation in biology.

The study of particular systems is the source of inspiration that guides the formation of general ideas from specific cases to general principles. Therefore the systemic approach extends towards the study of fundamental biological questions, such as gene assembly, protein folding and the nature of functional specificity. Such issues extend beyond the current perception of bioinformatics as a support discipline and address aspects of biological complexity, including the simulation of molecular interaction networks.

2. An overview to human genome

The genome coding regions are defined, in part, by an alternative series of motifs responsible for a variety of functions that take place on the DNA and RNA sequences, such as, gene regulation, RNA transcription, RNA splicing, and DNA methylation. For example,

sequencing of the human genome revealed a controversial number of interrupted genes (25,000-32,000) with their regulatory sequences [1, 2] representing about 2% of the genome. These genes are immersed in a giant sea of different types of non-coding sequences which make up around 98% of the genome. The non-coding regions are characterized by many kinds of repetitive DNA sequences, where almost 10.6% of the human genome consists of Alu sequences, a type of SINE (short interspersed elements) sequence [3]. [Alu] elements are not randomly distributed throughout the genome but rather are biased toward gene-rich regions [5]. They can act as insertional mutagens and the vast majority appears to be genetically inert (6). LINES, MIR, MER, LTRs, DNA transposons, and introns are other kinds of noncoding sequences, which together conform about 86% of the genome. In addition, some of these sequences are overlapped one to another, for example, the CpG islands (CGI), which complicates analysis of the genomic landscape. In turn, each chromosome is characterized by some particular properties of structure and function.

3. Human lentiviral integration

The two closely related human lentiviruses HIV-1 and HIV-2 are responsible for the 21th century AIDS pandemic [7-9]. Most current therapeutic approaches use combinations of antiviral drugs that inhibit activities of viral enzymes such as reverse transcriptase, protease and integrase; nevertheless none of those have succeeded in controlling infection [10-12]. One option to overcome the problem is to explore new therapies that include the study of the integration dynamics of human Lentiviruses because it would permit to understand the underpinnings behinds of alterations of cellular homeostasis when a cell is infected [13]. Additionally, analysis of integration process is important in HIV-induced disease and in Lentivirus-based gene therapy [14].

Integration is a crucial step in the life cycle of retrovirus permitting the incorporation of viral cDNA into the host genome [15-17]. cDNA integration is mediated by the virally encoded integrase enzyme and other viral and cellular proteins in a molecular complex called the pre-integration complex (PIC) [18]. One cellular factor involved in HIV targeting is the lens epithelium-derived growth factor (LEDGF) [19, 20], which binds to both HIV-1 integrase and chromatin, tethering the viral integration machinery to chromatin [21]. HIV-1 integration has been extensively studied using a wide array of molecular biology, biochemistry and structural biology approaches [22]. However, is critical to directly identify the viral distribution inside human genome in order to understand at genomic level the relationship between the composition and topology of chromatin and the target site selection.

As shown by previous studies, the preferences in target site selection for integration are not entirely random [23-26]; being pronounced favored and disfavored chromosomal regions which differ among retroviruses [27]. These preferential regions of host genomes are characterized by having a high frequency of integrational events, as known as "hotspot" and are distributed along the genome of host cell [28, 29]. In HIV-1, most of proviruses are localized into transcriptionally active regions not only in exons and introns, but also in sequences around start transcription sites [30, 31].

An additional related study performed by Felice et al, 2009 [32] compared and contrasted the chromosomal integration patterns between gamma retrovirus (Moloney Leukemia virus, MLV) and Lentivirus (human immunodeficiency virus type 1, HIV-1), finding that gamma-retroviral, but not lentiviral vectors, integrate in genomic regions enriched in cell-type specific subsets of transcription factors binding sites (TFBSs), independently from their relative position with respect to genes and transcription start sites. Therefore, is proposed that TFBSs could be differential genomic determinants of retroviral target site selection in the human genome.

Several *in vitro* and *in vivo* studies have shown that HIV-1 integrate predominantly in active transcription units and in genome zones with high gene density, high frequency of Alu elements, low content of CpG islands and open chromatin regions [33]. Notwithstanding this evidence, the identification of particular characteristics of local chromatin that facilitate integration in a wider genomic manner still remains to be elucidated.

The objective of the this chapter is to show the main results that our group of investigation have obtained of statistically testing those genomic variables that define a preferred genomic environment for human lentiviral integration and localize them in specific chromosome loci; moreover in the construction of gene/protein interaction networks among those cellular genes located around several Lentivirus integration sites in naturally infected humans as a systemic approach to better understand the lentiviral integration process.

To test our hypothesis we conducted *in silico* studies of the integration profile in the genomic DNA of peripheral blood mononuclear cells (PBMCs) and macrophages for both human Lentiviruses (HIV-1 and HIV-2) in a window size analysis of 100K. The statistical analyses included several genomic variables such as the chromosomal loci, the numbers of CpG Island, protein coding genes, transcripts and also the distribution of SINEs, LINEs, LTRs and others; moreover the exploration genomic regions in which epigenetics mechanisms would be associated with the integration process. Together, the results allow us to propose common genomic environments that favor the target chromatin zones for both human Lentiviruses.

4. Data mining and statistical analyses

A total of 352 human genome sequences flanking the 5'LTR of human Lentiviruses (176 sequences of HIV-1 [27] and 176 of HIV-2 [33] were obtained from GenBank (NCBI) under accession numbers: CL529260 to CL529766 (HIV-1) and DQ632388 to DQ632563 (HIV-2). Using the BLAST algorithm (NCBI; *http://blast.ncbi.nlm.nih.gov/Blast.cgi*), the sequences were aligned to the draft human genome (hg18) and those that met the following criteria were considered authentic integration sites: (i) contained the terminal 3' end of the HIV-1 or HIV-2 LTR; (ii) had matching genomic DNA within five bp of the end of the viral LTR; (iii) had at least 95% homology to human genomic sequence across the entire sequenced region; (iv) matched a single human genetic locus with at least 95% homology across the entire sequenced region (v) had minimum size of 50 bp.

BLAST of NCBI and the BLAT algorithm of the Genome Browser (University of California, Santa Cruz, Human Genome Project) (*http://www.genome.ucsc.edu/)* were used to obtain information about coding protein genes (RefSeq), transcripts, CpG islands and repetitive elements. Additional genomic information included molecular process and molecular function, was obtained from Gene Ontology (GO) (*http://www.geneontology.org/index.shtml*), GenCard (*http://www.genecards.org/cgi-bin/carddisp.pl*) and Gene Entrez (*http://www.ncbi.nlm.nih.gov/ncbi/geneentrez*). The chromosomal localization of the HIV-1 and HIV-2 proviruses was identified using the G pattern banding of each chromosome, as proposed by the Paris Conference (1971) [35], with updating of 850 times resolution. As the highest number of HIV-1 and HIV-2 proviruses was recorded on chromosome 17, an extensive characterization of its chromatin structure was performed including the genomic information available in several platforms of the Genome Browser: shows the CpG islands and distribution of its methylation; of histone H3 in the Lysine 4 and 27 methylation data obtained from ENCODE Histone modification by University of Washington CHIP-seq; Nucleosoma occupancy probabilities from A375 by Washington University and DNase1 hypersensitivity (ENCODE University of Washington) in GM12878 cells. All statistical analyses were performed using STATISTICA 7 [35]. The Mann-Whitney test (Wilcoxon rank) was used to establish differences between HIV-1 and HIV-2 chromosomal integration. Differences in function, molecular process and cell localization were analyzed using the t-test for independent samples. The Kolgomorov-Smirnov test was used for determining normality of data. In order to avoid an erroneous significance level for multiple comparisons a Bonferroni correction test was applied. To calculate the significant association among CpG numbers, genes numbers and integrations multiple regression analyses were performed. CpG numbers and genes per Mpb per chromosomes were determined from the NCBI and Ensemble databases (update 2010).

5. Patterns of provirus distribution

No significant differences were observed in the integration lymphocytic profiles between HIV-1 and HIV-2 ($p > 0.05$, Mann-Whitney test). The integrational events for both human Lentiviruses were recorded in all chromosomes except the Y (figure 1). However, significant differences between the number of HIV-1 and HIV-2 provirus were observed for chromosomes 4, 8, 9, 11 and 16 ($p < 0.05$, X^2 test). Most of the total integrations (39/352) occurred in chromosome 17 (figure 1). A tendency to a differential distribution of provirus towards telomeric and subtelomeric regions of the most of human chromosomes was observed. In this sense, other authors showed that centromeric alphoid repeat regions are disfavored as integration sites [36]. Although proviruses were observed in all chromosomes, we identified some chromatin regions with only HIV-1 integrations in chromosomes 4, 6 and 9 and only HIV-2 in chromosome 21.

5.1. Functional characterization of genes flanking integration sites

The ontology of genes hosting HIV integrations events were analyzed using G.O (Gene Ontology from NCBI). 83% (146/176) of HIV-1 and 77% (135/176) of HIV-2 integrations

occurred close to chromatin regions containing protein coding genes (p>0.05, t-student test). In a 100Kb extension of chromatin that harbored both HIV-1 and HIV-2 proviruses no differences were observed for the gene functional categories (p>0.05, Bonferoni´s correction). According to molecular function, 46% of HIV-1 integrations and 57% of HIV-2 were associated with molecular binding, while 19% and 18% respectively occurred in regions that code for genes associated with enzymatic function (figure 2a). Otherwise an exploring about the biological process revealed a preferential integration in a collection of genes involved in metabolism and gene expression for HIV-1 (36%) and HIV-2 (37%) (p>0.05, Bonferoni´s correction) (figure 2b).
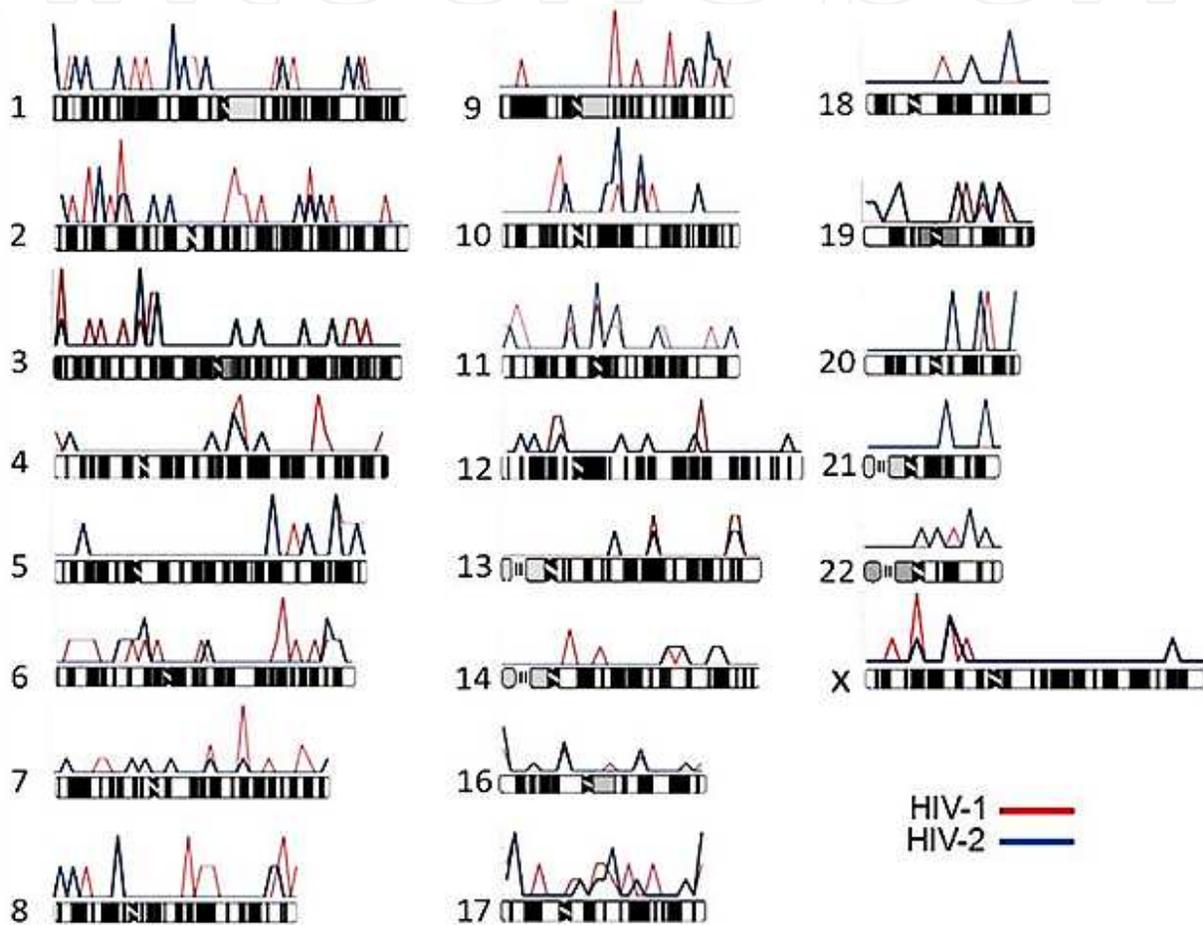

Figure 1. Chromosomal loci where 352 HIV-1 and HIV-2 cDNA have integrated into the human genome. Localization of chromosomal sequences matching both lentivirus are indicated in the graphics. Upper for each chromosome. Blue lines identify HIV-2 integrations and red lines identify HIV-1 integrations.

5.2. Distribution of the repetitive elements flanking integration sites

A low number of repetitive elements including SINEs, LINEs and LTRs were identified associated with provirus in an extension of 100Kb of flanking host chromatin. In general, there were no differences in the distribution of repetitive elements categories (SINEs, LINEs and LTRs) between HIV-1 and HIV-2 integrations (p>0.05, X^2 test). Our results showed that

both lentiviruses had a preferential integration close to Alu elements which correspond to SINEs. Within LINEs, differences among L1, L2 and L3 were recorded. The other class of repetitive elements like LTR, simple repeats and low complexity represented a minor proportion of the integration associated chromatin (figure 3).
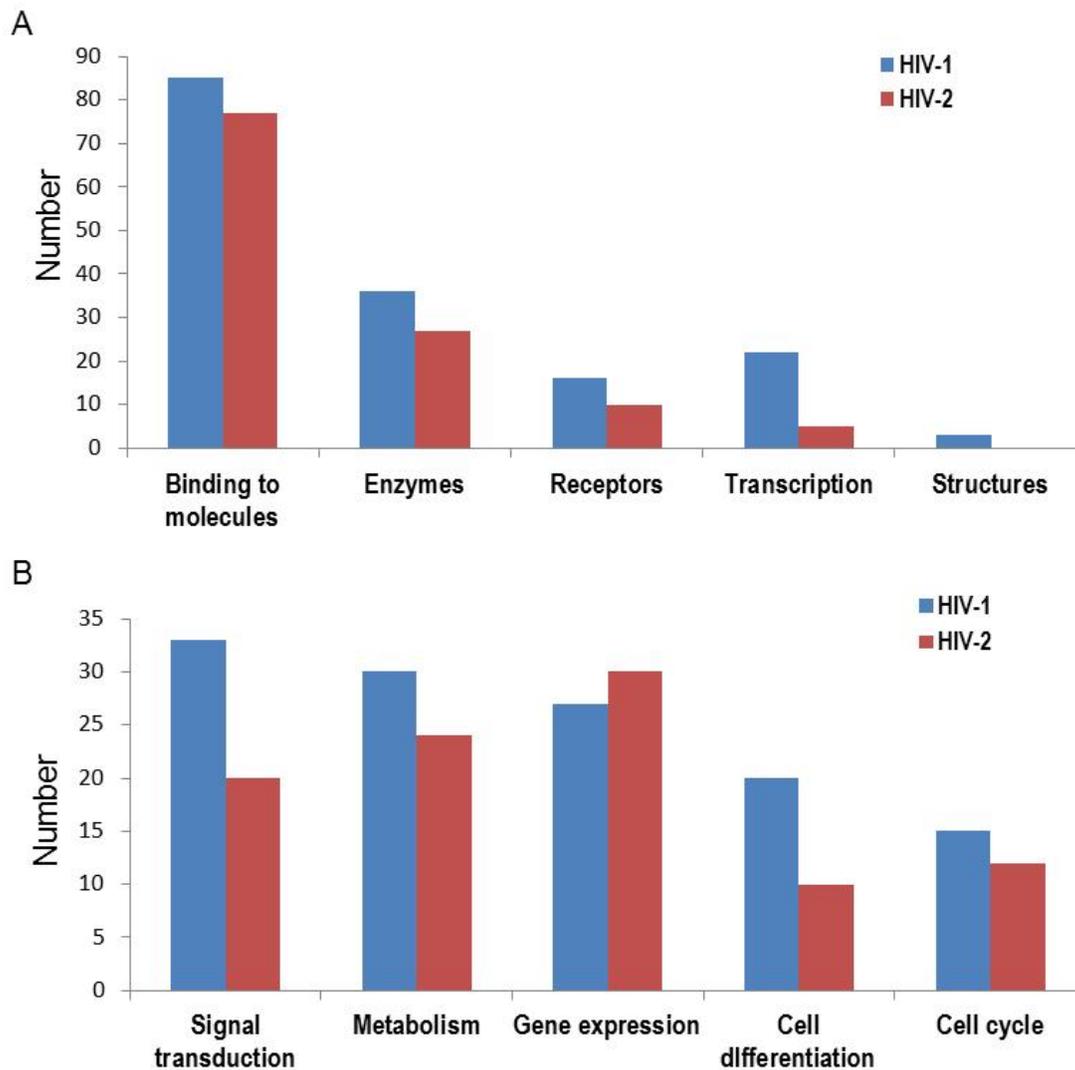

Figure 2. Functional characterization of the coding protein genes located in genome regions around 100 kb of human lentivirus. (a) Molecular function by GO of genes associated with HIV-1 and HIV-2 integration. (b) Biological process by GO of genes associated with HIV-1 and HIV-2 integration. Blue blocks correspond to HIV-1. Red blocks to HIV-2

5.3. Definition of the common genomic environment of integrations

As the integration do not follow a random model [23-25], some characteristics of the chromatin associated with regions with high level of provirus integration, support the hypothesis that a preferential integration is conditioned by structural and functional states of local chromatin; these states are defined by several genomic variables which were studied in this work, and together would define genomic environments.
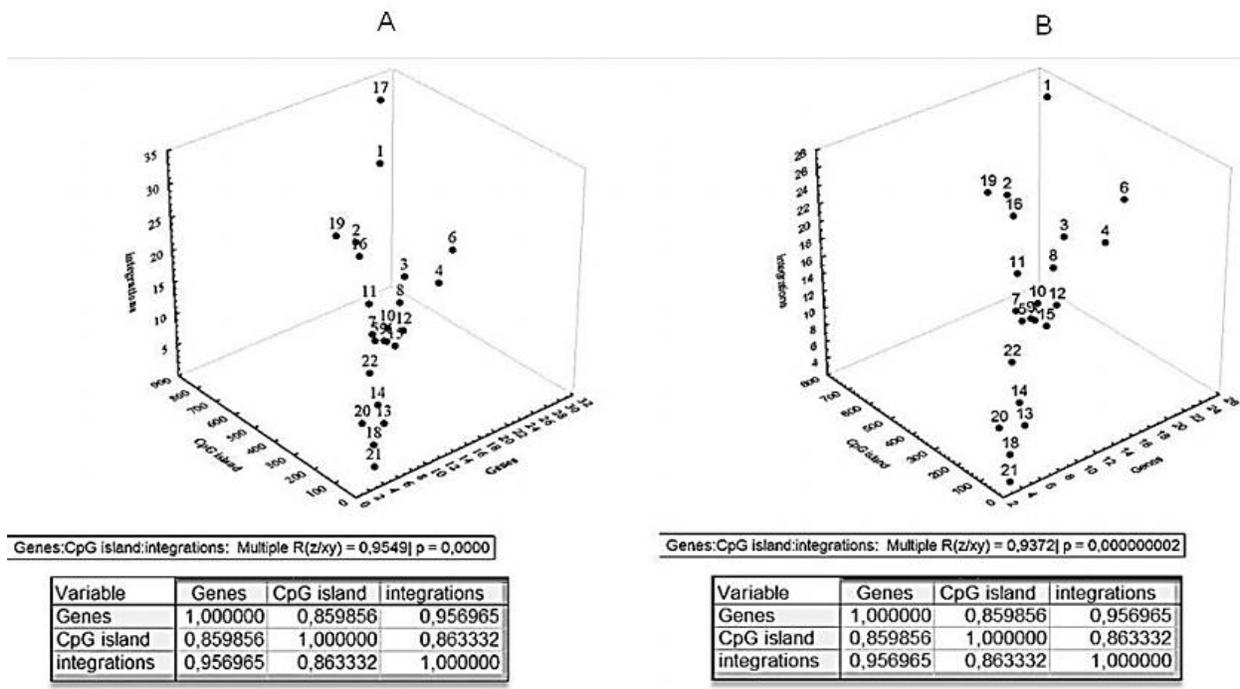
The results of multiple-regression analysis conducted on the HIV-1 and HIV-2 data sets showed that there were differential distributions of CpG island, genes, and Alu elements that together conditioned a specific genomic environment per chromosome (R^2=0.91, p<0.05). Gene density was the independent variable contributed most in the prediction of the dependent variable (integrations) due to the highest regression coefficients (B= 0.83; p<0.05). The highest relative likelihood of hosting a lentiviral integration event in the human genome was registered in chromosome 17 (figure 4a). To test that integration events are favored by gene-rich regions in all chromosomes, a comparison between those variables was done indicating that a high gene density in chromatin regions determine a favorable environment for integration, even when the chromosome 17 is excluded (Figure 4b). Because chromosome 17 registered the highest percentage of Lentiviral integration events, a detailed analysis of chromatin structure correlating several variables that give data about the cellular chromatin status was performed. In general the distal chromatin regions of p and q arms showed similarities in the distribution of methylation in CpG islands, methylation in several lysine residues of histone H3 (K4, K27 and K36) and variable levels of open chromatin and nucleosome occupancy (figure 5a and b).
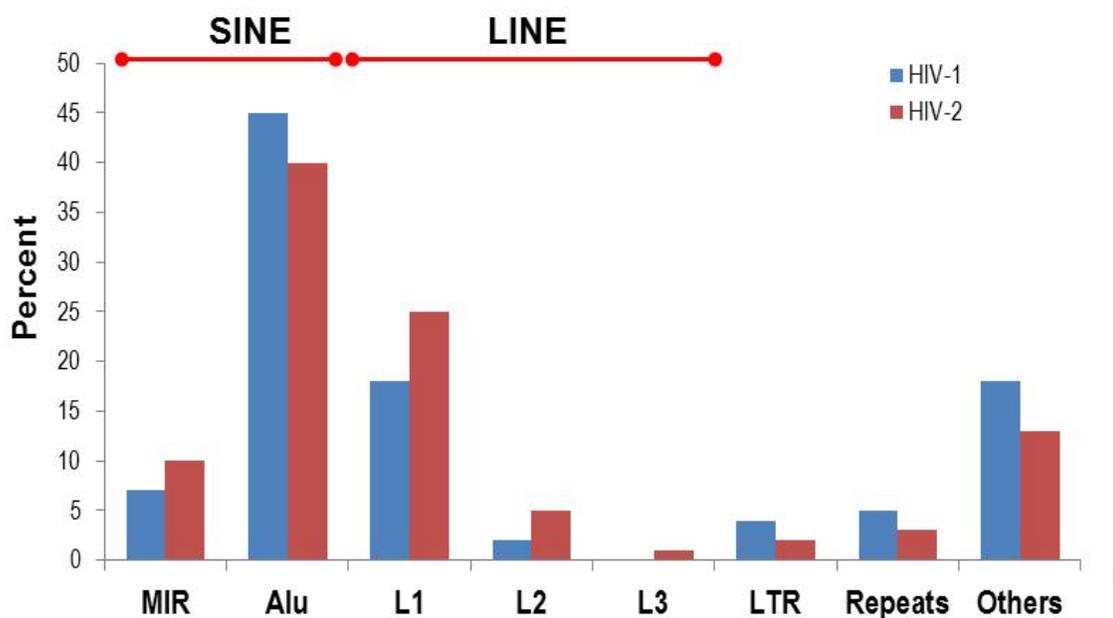
Figure 3. Frequencies of several repetitive elements associated with regions of 100 kb around the HIV-1 and HIV-2 proviruses. SINEs, short interspersed nuclear element; LINEs, long interspersed nuclear element.

Experimental studies have demonstrated that regulatory regions in general and promoters in particular, tend to be DNase sensitive and are target for integration of the majority of retroviruses [37, 38]. In 2006, the complete nucleotide sequence of chromosome 17 was published [39]. This chromosome is rich in protein coding genes, having the second highest gene density in the genome, (16.2 genes per Mb), with a relative excess of short interspersed elements (SINEs, 22.3%) and a deficit of long interspersed elements (LINEs, 14.4%). Likewise, this chromosome has high average CpG content (45.5%) and high euchromatin

density [39] (figure 6). Our statistical analysis determined that chromosome 17 had the highest number of integrations, mainly concentrate towards the telomeres of both arms.


Figure 4. Multiple-regression analysis among gene density, CpG island number, and frequency of HIV-1 and HIV-2 proviruses including every human chromosome. A high statistical correlation is observed mainly for chromosome 17. (a) Analysis including all human chromosomes. (b) The same analysis but excluding chromosome 17.

The most relevant relationship was related to the conformational state of chromatin including the nucleosomes occupancy, methylation of CpG Islands, DNase hypersensitive regions and transcriptionally active genes that are found in open-decondensed chromatin regions. These regions provide the environment for DNA regulatory processes such as DNA replication, repairs and transcription. Albanese et al. (2008) [40], found histone and IN acetylation may favor integration by tethering the virus to acetylated/decondensed regions of the chromatin. We concluded that the structural characteristics and the epigenetic modifications observed in those regions with high frequency of cDNA viral integrations would synergistically configure a local "genomic environment" that facilitates the target site selection during the retroviral integration.

5.4. Construction of HIV-1 gene/protein networks

Host-virus interactions is a complex level of systems information that permits a thorough understanding of how the virus exploits the host cell and uses the cellular machinery to integrate into host genome. Recently, the HIV-1 Human Protein Interaction Database (HHPID) registered 3959 interactions among 1452 human proteins and nineteen HIV proteins (fifteen of them structural and four intermediate proteins) [41].
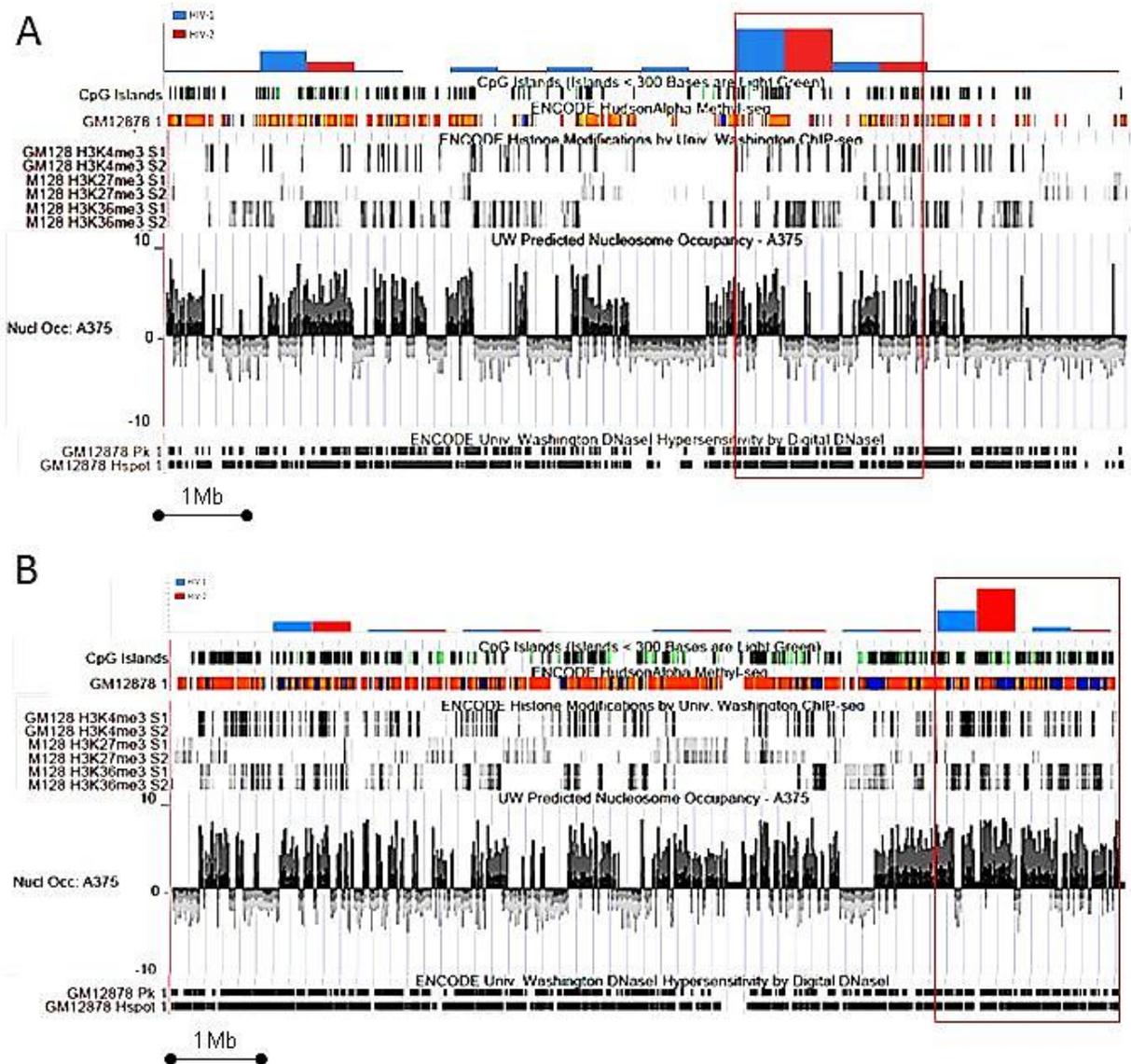
Figure 5. Flash image of the Genome Browser showing the distribution of several characteristics of the chromatin along 9.5 Mb of the p and q arm representing 25% of chromosome 17. (a) p arm, (b) q arm. The figure shows the GC percentage for each: 5pb (black), refseq Genes (black), CpG islands (several tones of gray), levels of open chromatin (ENCODE, Duke) in GM12878 cells with DNaseI and FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements) (black), DNase1 hypersensitivity (ENCODE, University of Washington) in GM12878 cells (gray), pk (sites identified as signal peaks within FDR 0.5% hypersensitive zones), Hspots (zones identified using the HotSpot algorithm), and predicted nucleosome occupancy in A375 cells (black peaks).

Previous studies have identified most of human cell pathways been disturbed by at least one interaction with an HIV-1 protein during the virus life cycle [42-44]. Those interactions are of two types: either direct, via host cell protein-viral proteins or indirect, such as regulatory interactions that alter expression of human genes [45, 46]; the signaling network cc-cytokine is both disrupted and exploited by HIV at various stages of infection. 22 candidates human class E proteins were connected into coherent network by 43 different protein-protein interactions, in which AIP1 play a key role in linking complexes that act

early (TSG101/ESCRT-I) and late (CHMP4/ESCRT-III) in the HIV infection pathways [47, 49]. Monocyte/macrophage infection is characterized by a viral dynamic substantially different from that of T lymphocytes. In fact, *in vivo* HIV infection of activated CD4-T lymphocytes accounts for the majority of the daily production of virus particles. However, a large number of lymphocytes are in a resting state, thus unable to sustain a complete and productive virus life cycle, and contribute only minimally to the daily virus production [50-52]. Because of the limited HIV-induced cytopathic effect and of their ability to accumulate high levels of HIV particles in intracellular compartments, HIV-infected macrophages serve as a potentially important reservoir, and as "Trojan horses" exploited by the virus to favor its dissemination in different tissues. [53, 54].

Cytoscape v.2.63 [55] was also used to construct a gene expression network from two kinds of files: The first one from gene expression profiles as a text file (.pvals) that were imported of expression data microarray experiments (GEO profiles, NCBI). The second, as data annotation in text files (.sif) that corresponds to each one gene-gene interactions (online databases). In the first one, gene expression values were collected from the microarray data series GSE19236 composed by two Agilent platforms (**GPL6480** and **GPL6848**) with 48 samples of monocytes to macrophages, macrophages and dendritic cells. These are available from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) repository (accession number GEO: GSE19236) and for our analysis, we selected all macrophages expression samples (GSM476720, GSM476721, GSM476722, GSM476723, GSM476724, GSM476725). To identify which genes were significant among samples in microarrays; considering a p-value< 0.001 as significant, an ANOVA test was calculated. Additionally, a Hierarchical clustering analysis of the samples using Euclidean Distance Method and mean linking were performed. MultiExperiment Viewer v4.1 [56] was applied to make the corresponding statistical analyses. Using data from BOND (Biomolecular Network Data Bank, http://bond.unleashedinformatics.com/Action), BioGird (Biological General Repository for Interaction Datasets, http://thebiogrid.org/), KEEG (Kyoto Encyclopedia of Genes and Genomes, http://www.genome.jp/kegg/), available online, a new file with the interaction data of 28 genes located close to integration sites was constructed.

Cytoscape v2.6 was used for visualizing and analyzing the genetic interaction networks among 28 human macrophages genes and their interactions. BiNGO v2.6 plugin (Biological Networks Gene Ontology tool) was used to determine which Gene Ontology (GO) terms are significantly overrepresented in a set of genes. A hypergeometric test was applied to determine which categories were significantly represented (p-value< 0.01); significant value was adjusted for multiple hypotheses testing using the Bonferroni Family-wise error rate correction [57]. The network topology parameters were calculated using Network Analyzer plug-in, which includes network diameter, the number of connected pairs of nodes and average number of neighbors; it also analyses node degrees, shortest paths, clustering coefficients, and topological coefficients (Max Planck Institute Informatik).

To identify active sub-networks as highly connected regions of the main network we used j ActiveModules plug-in that grouped genes according with significant p-values of gene expression over particular subsets of samples. The result shows active modules, listed

according to the number of nodes, and an associated Z-score. An active module with Z-scores greater than 3.0 indicated significant response upon the conditions of the experiment. We kept the standard default values, as being the most effective for initial analyses (58).
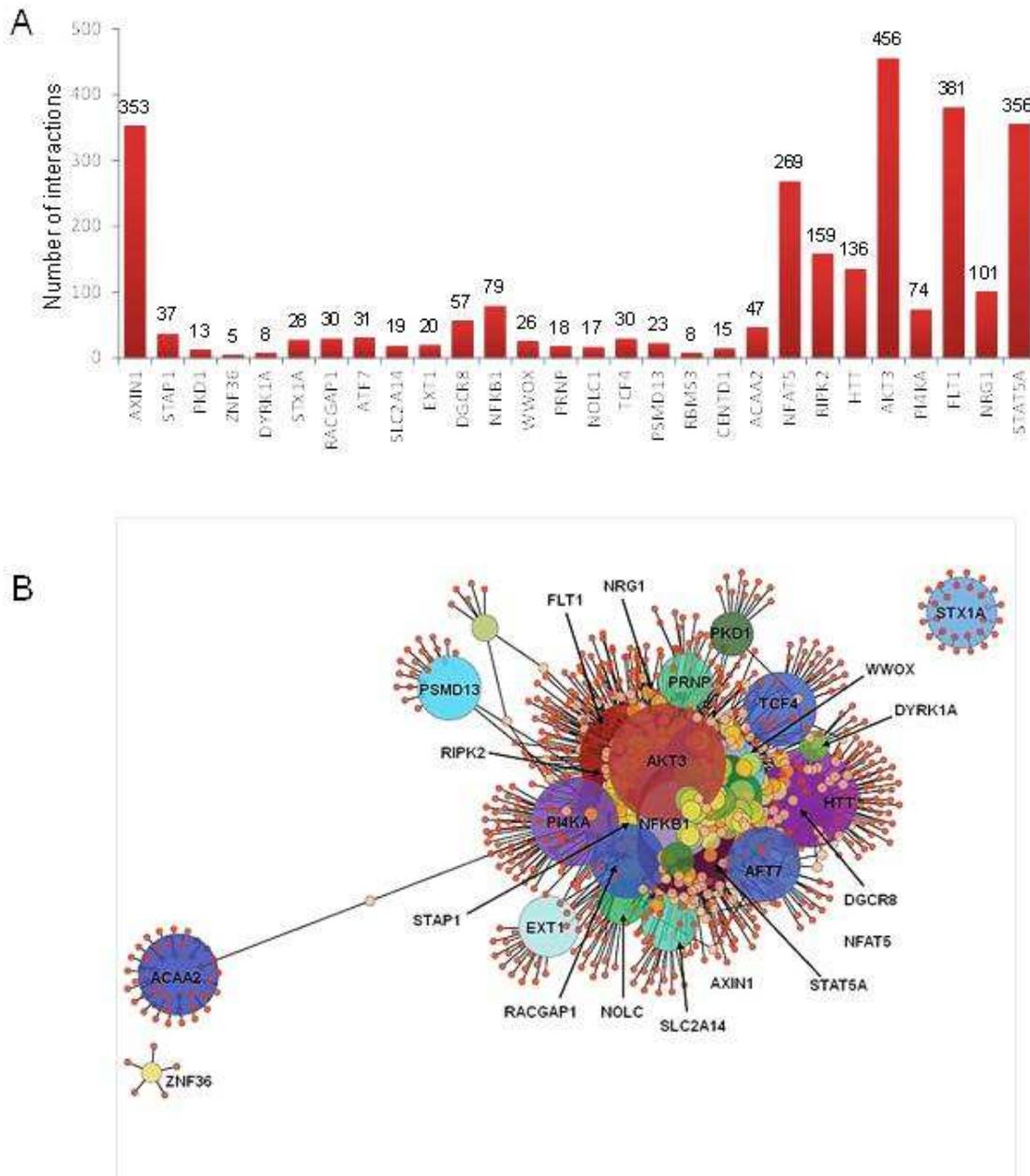
Figure 6. (A). Interaction values of 28 genes of human macrophages interrupted by HIV-1 cDNA integration. (B). Gene expression network in non HIV-1 infected macrophage. Visualization of gene network composed by 28 genes located close to regions with high frequency of HIV-1 provirus in human macrophages. These genes interact with 1202 genes through 2770 interactions. The network was constructed using Cytoscape. Each node corresponds to a gene and edges represent interactions among genes. The color gradient represents the expression values

Eleven thousand and seven hundred and thirteen (11,713) significant genes of 41,000 probes were clustered in two significant different groups of cells; one of them included only

dendritic cells, meanwhile the second grouped monocyte to macrophages and macrophages which are sharing similar gene expression patterns. A total of 2,770 interactions among 28 genes which were located closed to HIV-1 proviruses in human macrophages were recorded. AKT3 was gene with highest number of interactions (456), followed by FLT1 (381), STAT5A (356) and AXIN1 (328) (figure 6a). In contrast ZNF36, DYRK1A and RBMS3 genes had the lowest number of gene interactions. The normal macrophage gene network showed tree components: the main cluster composed by 26 macrophages genes and its interactions and two minor clusters in which ZNF36 gene was the central node with five interactions; and STX1A as central node with twelve interactions (Figure 6a).

To further identify active sub-networks inside the main gene network, we performed an expression clustering analysis using p-values calculated by comparison of gene level expression among five macrophage samples. We found 5 subnetworks, in which the most significant active module was integrated by 222 genes with a score of 3.15 (p<0.01). Within them 12 genes related with provirus integration sites were found: AXIN1, NFAT5, STAT5A, FLT1, AKT3, HTT, RIPK2, DGCR8, WWOX, NRG1, DYRK1A and SLC2A14 (figure 6b).

The GO functional significant categories in this active module showed enrichment for positive regulation of biological process and cell proliferation. Most of the genes identified in this sub-network were associated with cellular pathways that play significant role by modulating cell signaling networks including Wnt signaling, MAPK signaling and ErbB signaling.

5.5. Effects on normal macrophage gene networks by HIV-1 integration

In order to better understand the alteration of macrophages homeostasis by the HIV-1 integration, our analyses were focused to simulate what are the effects of viral cDNA integration in the alteration of several gene expression networks in human macrophage. In general the topology of non-infected macrophage network gene was dramatically changed by the HIV-1 integration events that lead to turned off the expression of five genes by the integration of proviral cDNA (Figure 7).

The evaluation of the several topological parameters such as clustering coefficient, shortest paths, network heterogeneity, the centralization, average number of neighbors and characteristic path length, showed a changed in the values of HIV-1 macrophage infected gene network, compared with normal macrophage network. The non-altered network was more condensed, had more number of interactions, was wide open rich in shortest paths and also was composed by one major component and two minor clusters being more heterogeneous and multi-functional (table 1).

Statistical differences between the topology states of two networks were registered for topological coefficients, closeness centrality and neighborhood connectivity distribution (Kolgomorov-Smirnov test p<0.05), but not in average clustering coefficient distribution. These results indicate that normal network was significantly more central and densely connected in comparison with that of HIV-1 macrophage infected network.
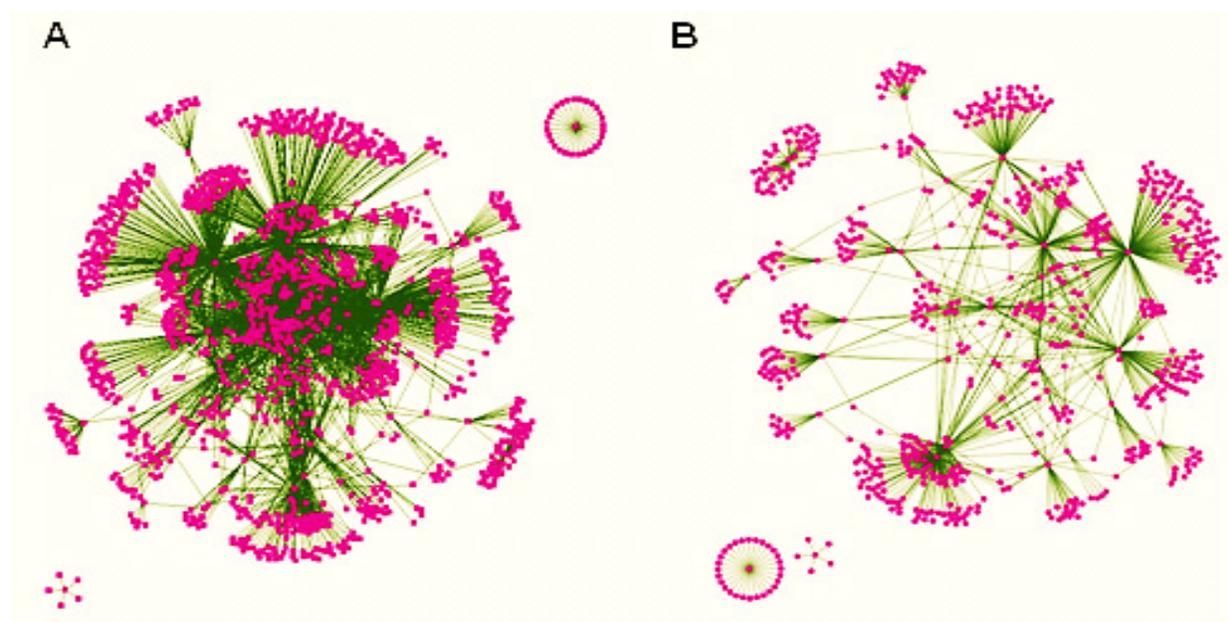
Figure 7. Effects in the topology gene expression network in macrophage by HIV-1 integration. (A) Normal macrophage genes expression network. (B) HIV-1 integration network when five macrophage genes were turned off.

Parameter	Normal macrophage	HIV-1 infected macrophage
Clustering coefficient	0.30	0.04
Connected components	3	3
Shortest paths	94%	90%
Network heterogeneity	5.63	3.75
Centralization	0.34	0.21
Avg. number of neighbors	4.2	2.70
Characteristic path length	3.30	4.13

Table 1. Comparison of network parameter values in normal and HIV-1 infected macrophages.

Using Random network plugin by Cytoscape we found the Clustering Coefficient of the non-infected Network and simulated infected Network in comparison with those generated at random showed not statistical differences (Kruskal-Wallis test, P= 0.317). The data confirmed that the topology of both reported networks have a strong support that the simulation of our gene network is valid.

We test our hypothesis that integration HIV-1 generate disturbs in the gene expression having a global effect in cellular networks and essential biological pathways. The enriched GO terms were categorized for normal and infected macrophages networks to identify the functional cellular change by HIV-1 integration. From all the GO categories covered by the 28 macrophages genes and its interactions, we have listed the ten most significant categories of the enriched GO terms in table 2.

The Gene Ontology (GO) enrichment analysis that normal network was composed by 423 significant functional categories of a total of 1190. These individual significant categories

could be further classified into two major groups; cell function regulation and signaling of biological process. In contrast HIV-1 infected macrophage gene network was enriched with 10 significant functional categories of a total of 40. The significantly overrepresented categories indicated that this emergent new gene network was composed by genes involved in metabolic process and DNA repair process.

In this study we simulated at systemic level, the alterations of cellular pathways when HIV provirus integrates into genes by turning them off and produce dysregulation of several local signaling pathways. One of the target gene associated with HIV-1 integration was AKT3, also called PKB, which is a serine/threonine protein kinase family member. It is involved in a wide range of biological processes including cell proliferation, differentiation, apoptosis, stimulating cell growth, and regulating other biological responses (59, 60). Also, it have been identified playing important roles of regulation in the G2/M transition of the cell cycle.

Normal Network	p-value [b]	HIV-1 infected Network	p-value [b]
Signal transduction [GO-ID: 7165]	3,90E-123	Biopolymer biosynthetic process [GO-ID: 43284]	6,27E-05
Cell communication [GO-ID: 7154]	4,63E-120	Metabolic process [GO-ID: 8152]	8,00E-05
Positive regulation of cellular process [GO-ID: 51242]	1,39E-95	Macromolecule biosynthetic process [GO-ID: 9059]	1,40E-03
Positive regulation of biological process [GO-ID: 48518]	1,93E-92	Biosynthetic process [GO-ID: 9058]	2,43E-03
Biological regulation [GO-ID: 65007]	1,45E-88	alcohol metabolic process [GO-ID: 6066]	3,19E-03
Intracellular signaling cascade [GO-ID: 7242]	1,40E-86	Maintenance of fidelity during DNA-dependent DNA replication [GO-ID: 45005]	6,24E-03
Regulation of cellular process [GO-ID: 51244]	2,29E-85	Mismatch repair [GO-ID: 6298]	6,24E-03
Regulation of biological process [GO-ID: 50791]	2,65E-84	Furaldehyde metabolic process [GO-ID: 33859]	6,68E-03
Phosphate metabolic process [GO-ID: 6796]	3,01E-80	Age-dependent response to reactive oxygen species during chronological cell aging [GO-ID: 1320]	6,68E-03
phosphorus metabolic process [GO-ID: 6793]	3,01E-80	oxidation reduction [GO-ID: 55114]	6,90E-03

a. The description of the gene ontology biological processes and the corresponding gene ontology identifiers are given.
b. p-Value calculated as an exponential function.

Table 2. The top 10 of significant biological process of normal and HIV-1 infected macrophages networks [a].

AKT3 via JNK interacts with NFTA and Jun that are targets for the HIV-1 macrophage integration network and are included in the mitogen-activated protein kinase (MAPK) cascade which perform essential functions such as proliferation, survival and inflammation, apoptosis in all cell types. This pathway is associated with others that include the phosphatidylinositol signaling system, Wnt signaling pathway, ERK5 pathway, P53 signaling pathway. (61-63). According with these previous data, we propose that, when AKT3 is turned off by HIV-1 integration, the cross talk with others is disrupted leading to a signaling dysfunction of metabolic associated pathways. When AKT3 was inactive the direct interaction with MKK7 produce a disruption of JNK and after with JUN that would result in a non activation by phosphorilation of apoptotic and cell cycle process. On the other hand inactivation of the MAPK pathway in both macrophages and dendritic cells leads to inhibition of proinflammatory cytokine secretion, downregulation of co-stimulatory molecules such as CD80 and CD86, and ineffective T cell priming. The net result is an impaired innate and adaptive immune response (64, 65).

Recently it have been reported that HIV-1 infection triggers the activation of the PI3K/Akt cell survival pathway in primary human macrophages as reflected by decreased PTEN protein expression and increased Akt kinase activity and renders these cells resistant to cytotoxic insults (54, 61, 64, 65). As result of HIV-1 integration close to AKT3, PTEN, AKT1 and 2, FOXO 1 and MDM2 that are included into the macrophage gene network, would expected a disruption of the apoptotic process.

6. Conclusions

We can conclude that a general effect of HIV-1 integrations in macrophages DNA is to disrupt several signaling pathways that control the normal cell homeostasis. Comparison between normal and infected macrophages of top 10 GO function categories showed the dramatic change of one non-infected macrophage whose main cellular functions are devoted to maintain a cell signaling crucial functions, to one infected in which the most important function are macromolecular biosynthetic process, maintenance of fidelity during DNA-dependent DNA replication, mismatch repair, age-dependent response to reactive oxygen species during chronological cell aging and oxidation reduction. As HIV infected macrophage is an abnormal reservoir in which the metabolic cascades are altered, it is possible to propose that the metabolism of macrophage adapt to perform survival functions where the apoptotic process is interrupted and a SOS metabolism make that the macrophage change of its life style

In silico studies are based upon statistical calculations which permit the drawing of generalizations about a biological process; however since some variables could affecting the *in toto* process, in order to get a real history of Lentivirus integration it would be important to consider that there is another factors, including physiological process and cellular compartments that would be influencing the *in vivo* integration site selection. Some of these are cell-cycle phase, the transcriptional state of the cell, the topology of chromosomal DNA, cell type infected, and presence of co-helper molecules during the PIC complex conformation

By providing these new testable hypotheses we hope that our results will accelerate experimental efforts to define a reliable disturbing in the gene complex relationship by lentivirus integration in PBMC and macrophages which are critical immune cells responsible for a wide range of immune functions and play multifaceted roles in HIV pathogenesis

Author details

Felipe García-Vallejo
Physiological Sciences Department, Scientific Director of the Laboratory of Molecular Biology and Pathogenesis, School of Basic Sciences, Health Sciences Faculty, Universidad del Valle, Cali, Colombia

Martha Cecilia Domínguez
Laboratory of Molecular Biology and Pathogenesis, School of Basic Sciences, Health Sciences Faculty, Universidad del Valle, Cali, Colombia

7. References

[1] Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K,Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, et al. (2001) Initial sequencing and analysis of the human genome. *Nature.* 409:860–921.

[2] Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Miklos G.L. Gabor, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, et al. (2001) The sequence of the human genome. *Science.* 291:1304–1351.

[3] International human genome sequencing consortium. *(*2004) Finishing the euchromatic sequence of the human genome. *Nature.* 431: 931-945.

[4] Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC, Remington KA, Abril JF, Gill J, Borman J, Rogers YH, Frazier ME, Scherer SW, Strausberg RL, et al 2007 The diploid genome sequence of an individual human. *PLoS Biol*. 5:e254.

[5] Versteeg R, van Schaik BDC, van Batenburg MF, Roos M, Monajemi R, Caron H, Bussemaker HJ, van Kampen AHC (2003) The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. *Genome Research* 13:1998-2004.

[6] DeCerbo J, Carmichael GG (2005) SINEs point to abundant editing in the human genome. *Genome Biology.* 216:1-4.

[7]   United Nations AIDS Program on HIV/AIDS (2008). Report of the Global AIDS. UNAIDS/UN.

[8]   Simon V, Ho DD.and Karim QA (2007) HIV/AIDS Epidemiology, Pathogenesis, Prevention and Treatment. Lancet. 368: 489-504

[9]   Inciardi J.A. and Williams M.L Editor's introduction: (2005) The global epidemiology of HIV and AIDS. AIDS Care. 17 (suppl 1): S1–8

[10] Balakrishnan S (2009) Alternative paths in HIV-1 targeted human signal transduction pathways. BMC Genomics. 10 (Suppl 3), S30

[11] Wang YJ, McKenna PM, Hrin R, Felock P, Lu M, Jones KG, Coburn CA and Grobler JA (2010) Assessment of the susceptibility of mutant HIV-1 to antiviral agents. J. virol methods. 165: 230-237.

[12] Hanson K and Hicks C (2006) New antiretroviral drugs. Current HIV/AIDS Rep. 3: 93-101.

[13] Moore J.P. and Stevenson M : New Targets for Inhibitors of HIV-1 Replication. Nature Rev. 2000; 1: 40-49.

[14] Cereseto A, and Giacca M (2004) Integration site selection by retroviruses. AIDS Rev 6: 13-21.

[15] Coffin J.M: Retroviridae and their replication In Virology, ed. B.N. Fields et al., Raven Press, New York, 1996, pp. 1767–1848.

[16] Sierra S, Kupfer B and Kaiser R (2005) Basics of the virology of HIV-1 and its replication. J. Clin. Virol. 34: 233-244

[17] Hindmarsh P, Leis J (1999) Retroviral DNA integration. Microbiol mol. biol. rev. 63: 836-84

[18] Van Maele B, Busschots K, Vandekerckhove L, Christ F, and Debyser Z (2006) Cellular co-factors of VIH-1 integration. Trends biochem. sci. 31: 98–105.

[19] Bushman F, Lewinski M, Ciuffi A, Barr S, Leipzig J, Hannenhalli S. and Hoffmann C (2005) Genome-wide analysis of retroviral DNA integration. Na.t rev. microbiol. 3: 848-58.

[20] Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, Shinn P, Ecker JR and Bushman F (2005) A role for LEDGF/p75 in targeting HIV DNA integration. Nat. med. 11: 1287-9.

[21] Ciuffi A (2008) Mechanisms governing lentivirus integration site selection. Curr. gene. ther. 8: 419-29.

[22] Lewinski M, Yamashita M, Emerman M, Ciuffi A, Marshall H, and Crawford G (2006) Retroviral DNA Integration: Viral and Cellular Determinants of Target-Site Selection. PLoS pathog. 2: 0611-0622

[23] Ferris AL, Wu X, Hughes CM, Stewart C, Smith SJ, Milne TA, Wang GG, Shun MC, Allis CD, Engelman A and Hughes SH (2010) Lens epithelium-derived growth factor fusion proteins redirect HIV-1 DNA integration. Proc. natl. acad. sci. USA. 107: 3135-40

[24] Jordan A, Defechereux P, Verdin E (2001) The site of HIV-1 integration in the human genome determines basal transcriptional activity and response to Tat transactivation. The EMBO J. 20: 1726-1738

[25] Wu X, Li Y, Crise B, Burgess S.M (2003) Transcription start regions in the human genome are favored targets for MLV integration. Science. 300: 1749-1751

[26] Hematti P, Hong BK, Ferguson C, Adler R, Hanawa H, Sellers S, Ingeborg E (2004) Distinct Genomic Integration of MLV and SIV Vectors in Primate Hematopoietic Stem and Progenitor Cells. PLoS Biol. 2: E423

[27] Rick SM, 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: 1127-1137

[28] Maxfield L, Fraize C, Coffin JM (2005) Relationship between retroviral DNA-integration-site selection and host cell transcription. Proc natl. acad. Sci USA. 102: 1436-1441

[29] Soto J, Peña A, Salcedo M, Domínguez MC, Sánchez A, García-Vallejo F (2010) Caracterización Genómica de la Integración In vitro del VIH-1 en células mononucleares de sangre periférica, macrófagos, y células T de Jurkat. Infectio. 14: 20-30

[30] Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, Bushman F (2002) HIV-1 integration in the human genome favors active genes and local hotspots. Cell. 110: 521–529

[31] Derse D, Crise B, Li Y, Princler G, Stewart C, Connor F, Hughes H, Munroe D, Wu X (2007) HTLV-1 integration target sites in the human genome: comparison with other retroviruses. J. virol. 81: 6731-6741

[32] Felice B, Cattoglio C, Cittaro D, Testa A, Miccio A, Ferrari G, Luzi L, Recchia A, Malivio, F (2009) Transcription Factors Binding Sites Are Genetic Determinants of Retroviral Integration in the Human Genome. PLoS One. 4: e4571

[33] MacNeil A, Sankale JL, Meloni S, Sarr A, Mboup S, Kanki P (2006) Genomic Sites of Human Immunodeficiency Virus Type 2 (HIV-2) Integration: Similarities to HIV-1 In Vitro and Possible Differences In Vivo. J. virol. 80: 7316–7321

[34] PARIS CONFERENCE. Supplement (1975) Standardization in human cytogenetic. Cytogenet. Cell Genet. 1971; 15: 203-238

[35] STATSOFT, INC: STATISTICA (data analysis software system), 2004, version 7. Available on: www.statsoft.com

[36] Carteau S, Hoffmann C, Bushman F (1998) Chromosome structure and human immunodeficiency virus type 1 cDNA integration: centromeric alphoid repeats are a disfavored target. J. virol. 72: 4005–4014.

[37] Taganov KD, Cuesta I, Daniel R, Cirillo LA, Katz RA (2004) Integrase specific enhancement and suppression of retroviral DNA integration by compacted chromatin structure in vitro. J. virol. 78: 5848–5855.

[38] Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A (2004) International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature. 431: 931-945

[39] Zodyl MC, Garber M, Adams D, Sharpe T, Harrow J, Ames R, Nicholson C (2006) DNA sequence of human chromosome 17 and analysis of rearrangement in the human lineage. Nature. 440: 1045-1049

[40] Albanese A, Arosio D, Terreni M, Cereseto A (2008) HIV-1 Pre-Integration Complexes Selectively Target Decondensed Chromatin in the Nuclear Periphery. PLoS one. 3: e2413. doi:10.1371/journal.pone.0002413

[41] Fu W, Sanders-Beer BE, Katz KS, Maglott DR, Pruitt KD, Ptak RG (2009) Human immunodeficiency virus type 1, human protein interaction database at NCBI. Nucleic acid. res. 37(Database issue):D417-22

[42] Song, G, Ouyang G, Bao S (2005) The activation of Akt/PKB signaling pathway and cell survival. J. cell. mol. med. 9 (1): 59-71.

[43] Balakrishnan S, Tastan O, Carbonell J, Klein-Seetharaman J (2009) Alternative paths in HIV-1 targeted human signal transduction pathways. BMC Genomics. 10 (Suppl. 3):S30.

[44] Fu W, Sanders-Beer BE, Katz KS, Maglott DR, Pruitt KD, Ptak RG (2009) Human immunodeficiency virus type 1, human protein interaction database at NCBI. Nucleic acids res. 37: D417–D422.

[45] Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD (1996) HIV-1 dynamics in vivo: virion clearance rate, infected cell lifespan, and viral generation time. Science. 271 (5255):1582–1586.

[46] Sirskyj D, Thèze J, Kumar A, Kryworuchko M (2008) Disruption of the cc cytokine network in T cells during HIV infection. Cytokine. 43 (1): 1–14.

[47] von Schwedler UK, Stuchell M, Müller B, Ward DM, Chung HY, Morita E, Wang HE, Davis T, He GP, Cimbora DM, Scott A, Kräusslich HG, Kaplan J, Morham SG, Sundquist WI (2003) The protein network of HIV budding. Cell. 114 (6): 701–713.

[48] Bandyopadhyay S, Kelley R, Ideker T (2006) Discovering regulated networks during HIV-1 latency and reactivation. Pac. symp. biocomput. 354–366.

[49] Chun T, Carruth LM, Finzi D (1997) Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. Nature. 387:83–188.

[50] Bagnarelli P, Valenza S, Menzo R, Sampaolesi PE, Varaldo L, Butini M, Montoni CF, Perno S, Aquaro A, Mathez D (1996) Dynamics and modulation of human immunodeficiency virus type 1 transcripts in vitro and in vivo. J. virol. 70 (11):7603–7613.

[51] Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD (1996) HIV-1 dynamics in vivo: virion clearance rate, infected cell lifespan, and viral generation time. Science. 271(5255):1582–1586.

[52] Gendelman HE, Orenstein JM, Baca LM, Weiser B, Burger H, Kalter DC, Meltzer MS (1989) The macrophage in the persistence and pathogenesis of HIV infection. AIDS. 3:475–495.

[53] Herbein G, Gras G, Khan KA, Abbas W (2010) Macrophage signaling in HIV-1 infection. Retrovirology. 9:7–34.

[54] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome res. 13 (11):2498–2504.

[55] Saeed A, Bhagabati NK, Braisted JC, Liang W, Sharov W, Howe V, Li J, Thiagarajan M, White JA, Quackenbush J, (2006) TM4 microarray software suite. Methods enzymol. 411:134–193.

[56] Maere S, Heymans K, Kuiper M (2005) BiNGO: a cytoscape plug-in to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics. 21 (16):3448–3449.

[57] Ideker T, Ozier O, Schwikowski B, Siegel AF (2002) Discovering regulatory and signalling circuits in molecular interaction networks. Bioinformatics. 18:S233–S240.

[58] Maxfield L, Fraize C, Coffin JM (2005) Relationship between retroviral DNAintegration-site selection and host cell transcription. Proc natl acad.sci. 102 (5):1436–1441.

[59] Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, Bushman F (2002) HIV-1 integration in the human genome favors active genes and local hotspots. Cell. 110 (4):521–529.

[60] Dérijard B, Hibi M, Wu IH, Barrett T, Su B, Deng T, Karin M, Davis RJ (1994) JNK1: a protein kinase stimulated by UV light and Ha-Ras that binds and phosphorylates the c-Jun activation domain. Cell. 76:1025–1037.

[61] Mordret G, (1993) MAP kinase: a node connecting multiple pathways. Biol. cell. 79:193–207.

[62] Rao KM, (2001) MAP kinase activation in macrophages. J. Leukoc. Biol. 69 (1):3–10.

[63] Osaki M, Oshimura M, Ito M (2004) PI3K-Akt pathway: its functions and alterations in human cancer. Apoptosis. 9 (6):667–676.

[64] Chugh P, Bradel-Tretheway B, Monteiro-Filho CM, Planelles V, Maggirwar SB, Dewhurst S, Kim B (2008) Akt inhibitors as an HIV-1 infected macrophagespecific anti-viral therapy. Retrovirology. 5:11.