# **Supporting Information, Text S1**

# 24 hours in the life of HIV-1

Pejman Mohammadi, Sébastien Desfarges, Istvan Bartha, Beda Joos, Nadine Zangger, Miguel Muñoz, Huldrych F. Günthard, Niko Beerenwinkel, Amalio Telenti, Angela Ciuffi

Table of contents:

- 1. Modeling viral progression
- 2. Clustering of gene expression time courses
- **3.** Supplementary Figures
- 4. Supplementary Tables

#### 1. Modeling viral progression

**The model.** The raw measurements of viral intermediates consist of nine independent time series, each quantifying the progression of a specific step in the life cycle of the virus. Each representative marker (e.g., extracellular p24) accumulates in the sample from the start of infection until the time of the measurement. Ideally, each marker level,  $x_t$ , at time point t starts from zero and monotonically increases up to its maximum as more and more cells go through the corresponding phase of the viral life cycle. The net viral activity, v, during time span  $\Delta t$  can be estimated as

(Eq. 5) 
$$\overline{V}_{t,t+\Delta t} = \frac{x_t - x_{t+\Delta t}}{\Delta t}$$

In practice, however, the measurements are not monotonic and linear estimation of **Eq. 5** can lead to negative viral activity estimates, which are biologically implausible given that the measured viral processes are not reversible. We address this problem by incorporating two further components into the model.

First, we account for a constant rate of marker loss during the experiment, which can be due to the experimental procedure or due to natural degradation of the marker in the sample. Marker loss is modeled by an exponential decay term, such that

(Eq. 6) 
$$\frac{x_t - x_{t+\Delta t}}{\Delta t} = \overline{v}_{t,t+\Delta t} - \lambda x_t, \quad \lambda \in [0,\infty)$$

where  $\lambda$  is the decay rate, and for  $\lambda = 0$ , **Eq. 5** is recovered. Second, we constrain the viral activity,  $v_t$ , to be a non-negative parametric function over time. Specifically, assuming that one

unit of marker (e.g., one 2LTR circle) is produced only after a sequence of sub-events has taken place, the distribution of the waiting time until production of one unit of the marker can be described by a gamma distribution function,

(Eq. 7) 
$$\Gamma(t;k,\theta) = \frac{1}{\theta^k \int_0^\infty u^{k-1} e^{-u} du} t^{k-1} e^{-t/\theta}, \quad k,\theta \in (0,\infty)$$

with shape and scale parameters k and  $\theta$ , respectively. The overall expected production rate of the marker in the sample,  $v_b$  is then proportional to the density function at time t. Using an additional normalization constant  $\alpha$  this relationship can be written as

(Eq. 8) 
$$v_t^{(\theta,k)} = \alpha \Gamma(t;k,\theta) = \alpha \hat{v}_t^{(\theta,k)}, \quad \alpha \in [1,\infty)$$

where we have defined  $\hat{v}_t = \hat{v}_t^{(\theta,k)} = \Gamma(t;k,\theta)$ , and  $\alpha$  corresponds to the number of identical subprocesses. For  $\Delta t \to 0$ , **Eq. 6** becomes the following ordinary differential equation (ODE)

(Eq. 9) 
$$\frac{dx}{dt} = v_t^{(\theta,k)} - \lambda x_t$$

Assuming  $x_0 \ge 0$ , this deterministic equation can be solved as

(Eq. 10) 
$$x_t = e^{-\lambda t} \left[ \int_0^t v_{t'}^{(\theta,k)} e^{\lambda t'} dt' + x_0 \right] \quad x_0 \in [0,\infty)$$

The normalization constant,  $\alpha$ , can be factored out of the equation,

(Eq. 11) 
$$x_t = \alpha e^{-\lambda t} \left[ \int_0^t \hat{v}_{t'}^{(\theta,k)} e^{\lambda t'} dt' + \hat{x}_0 \right] \quad \hat{x}_0 = \frac{x_0}{\alpha}$$

We denote the right hand side of this equation by  $f(\hat{v}_t^{(\theta,k)}, \lambda, \hat{x}_0, \alpha)$ . Eq. 11 allows for deriving the pattern of viral activity in the absence of an absolute measure of the marker, which is often the case, for instance, when the data is produced by qPCR relative to an internal reference substrate. The viral activity model consists of four main components, namely activity shape  $(\hat{v}_t^{(\theta,k)})$ , marker decay  $(\lambda)$ , initial marker  $(\hat{x}_0)$ , and scaling constant  $(\alpha)$ . In the absence of marker decay, the model reduces to a linear transformation of the cumulative gamma distribution function.

**Parameter estimation and confidence intervals.** For each of the viral activity markers, model parameters were estimated using a least squares fit to the experimental data vector *y* with

(Eq. 12) 
$$\tau(y_t) = \tau(f(\hat{v}_t^{(\theta,k)}, \lambda, \hat{x}_0, \alpha)) + \varepsilon_t$$

where the function  $\tau(.)$  is variance stabilizing transformation (described below), and  $\varepsilon$  is a vector of i.i.d. Gaussian noise. Confidence intervals of the peak viral activity, which is given by  $k\theta$ , were specifically computed by using parametric bootstrapping [1]. For this, we generated bootstrap samples from the model using the estimated Gaussian noise, and estimated the distribution of the peak viral activity directly from the bootstrap samples.

**Variance stabilization.** Viral data was produced on three main platforms, namely qPCR, FACS, and ELISA (**Figure S2**). The data are heteroscedastic, which can significantly degrade the quality of the model fit. In order to address this problem, the logarithm was applied as a variance stabilizing transformation prior to least squares fitting. This choice is well-justified in case of qPCR data, which comprises seven out of the nine data sets. As **Figure S3** illustrates, the variance in the qPCR values can be almost perfectly described by the mean using a set of parallel regression lines on a logarithmic scale with data set-specific intercepts,  $b_j$ , and a common slope, a, corresponding to the equation

(Eq. 13) 
$$\operatorname{var}(y_{j,m}) = e^{b_i} \overline{y}_{j,m}^a$$

where  $\overline{y}_{j,m}$  is the mean value of the technical replicates of the  $m^{\text{th}}$  measurement in the  $j^{\text{th}}$  set of qPCR measurements. Hence, the variance stabilizing transformation,  $\tau$ , can be found as

(Eq. 14) 
$$\tau(\mathbf{y}_j) = e^{-\frac{b_j}{2}} \int^{\mathbf{y}} \omega^{-\frac{a}{2}} d\omega$$

The constant exponential term can be disregarded in practice, and the integral in **Eq. 14** becomes the log transformation for a slope parameter value of a = 2. Log transformation of qPCR data is identical to using the raw  $-\Delta\Delta$ CT values before the exponentiation step that is usually employed prior to reporting experimental values. For the FACS and ELISA datasets, there was not enough data available for *de novo* derivation of a suitable transformation. Hence, the log transformation was chosen in accordance with common practice [2,3].

**Model fitting and further remarks.** In order to fit the model to the experimental data, we use the trust region algorithm for nonlinear least-squares optimization implemented in Matlab [4]. At each iteration of the algorithm, the value of  $x_t$  was calculated numerically from the ODE in **Eq.9** using the Dormand-Prince method for solving non-stiff ODEs [5] implemented in the Matlab *ode45* function. All parameters were searched over the domain  $[0, +\infty)$ , which spans the valid domain of the function parameters. We used a difference of  $10^{-10}$  in subsequently updated parameters as a threshold to detect convergence of the algorithm. Each set of measurements was once used with the full model, and once with a reduced model in which the decay constant was fixed to zero. The likelihood was then calculated for the full and the reduced model and the preferred model was chosen based on the Bayesian Information Criterion (BIC) score [6]. **Figure S4** illustrates the model fits to data. **Figure S5** depicts the extracted pattern of activity and progression for each of the viral life steps. Bootstrapping was performed over the whole procedure using 500 samples. The resulting bootstrap distribution is shown by the white inner violins for the peak activity of each viral step in **Figure 1B**. We repeated the modeling

using three alternative activity shapes namely, Weibull, lognormal, and truncated normal distribution functions, all of which yielded very similar results (results not shown here). In addition to the conventional confidence intervals based on the Jacobian matrix, and the parametric bootstrapping, we further assessed predictive power and stability of the fits qualitatively by fitting the model to data after removing the last time points. **Figure S6** illustrates the resulting fits for the case of the last, and two last time points removed.

#### 2. Clustering of gene expression time courses

Gene expression profiles over the 24h observation time were clustered to identify co-regulated sets. For this purpose, we analyzed all 7,991 genes that were significantly described by the regression model, i.e., for which at least one regression coefficient in  $w_i$  was significantly different from zero, defined by a q-value below 0.05. Each gene *i* was represented by a normalized vector of regression weights,  $\hat{w}_i$ , such that

(Eq. 16) 
$$\hat{w}_i = \frac{w_i}{\|w_i\|}$$

We used the squared Euclidian distance for measuring the similarity of gene expression profiles. This choice is equivalent to the cosine distance between two un-normalized data vectors. This distance measure is scale-independent and sensitive to sign changes (additive constants). By contrast, the correlation distance would be insensitive to downregulation and upregulation as long as the relative expression pattern is unchanged. Clustering was carried out using the k-means algorithm [7] repeated 10,000 times using different sets of randomly chosen cluster seeds, and the best clustering was selected based on the lowest sum of point-to-cluster distances.

The optimal number of clusters was chosen based on the BIC score [6] calculated as

(Eq. 17) 
$$BIC(k) = -2\ln(L_k) + p_k \ln(N)$$

for *N* observations, the data likelihood,  $L_k$ , evaluated at the maximum likelihood parameters,  $p_k$  degrees of freedom, and *k* clusters. The likelihood was derived under the assumption of identical spherical Gaussian distributions of the length-normalized weight vectors [8]. The number of free parameters,  $p_k$ , was counted as the sum of (k - 1) cluster probabilities, 3k centroid coordinates, and one intra-cluster variance parameter [8]. Calculating the BIC score over a range of *k*, one can select an optimal number of clusters,  $k^*$ , as

(Eq. 18) 
$$k^* = \underset{k}{\operatorname{argmin}} BIC(k)$$

However, in order to account for the uncertainty in the data and the non-deterministic nature of the k-means algorithm, we add a bootstrapping scheme for choosing the optimal number of clusters. For each tested parameter value, k, bootstrap datasets were produced using re-sampling. Clustering was performed, and the corresponding BIC score was calculated for each bootstrap dataset separately. Sample average,  $B_k$ , and standard deviation,  $\delta_k$ , of the bootstrap BIC scores

were calculated for each tested value of *k*. Accounting for additional uncertainty in the mean estimator, we defined the following quantity:

(Eq. 19) 
$$s_k = \delta_k \sqrt{1 + \frac{1}{\beta}}$$

with  $\beta$  denoting the number of bootstrap samples. The number of clusters,  $k^+$ , was chosen as the most parsimonious solution within one standard deviation from the global minimum average BIC,

(Eq. 20) 
$$k^{+} = \min \left\{ k \left| B_{k} \leq B_{k^{*}} + s_{k^{*}} \right|, k^{*} = \arg \min_{k'} B_{k'} \right\}$$

This 1-standard-error distance is a commonly used threshold [9]. We tested the quality of clustering for a range of k values between 1 and 50 and the number of clusters was chosen as  $k^+$  = 18 (**Figure S7**). The resulting clusters are shown in **Figure S8**.

#### References

- 1. Efron B, Tibshirani RJ (1994) An Introduction to the Bootstrap.
- 2. Herzenberg LA, Tung J, Moore WA, Herzenberg LA, Parks DR (2006) Interpreting flow cytometry data: a guide for the perplexed. Nat Immunol 7: 681-685.
- 3. Peterman JH, Butler JE (1989) Application of theoretical considerations to the analysis of ELISA data. Biotechniques 7: 608-615.
- 4. Coleman TF, Li Y (1996) An Interior, Trust Region Approach for Nonlinear Minimization Subject to Bounds. SIAM Journal on Optimization 6: 418-445.
- 5. Dormand JR, Prince PJ (1980) A family of embedded Runge-Kutta formulae. J Comp Appl Math 6: 19-26.
- 6. Schwarz G (1978) Estimating the dimension of a model. Ann Satist 6: 461-464.
- 7. Jain AK, Dubes RC (1981) Algorithms for Clustering Data: Prentice-Hall.
- 8. Pelleg D, Moore A (2000) X-means: Extending K-means with Efficient Estimation of the Number of Clusters. International Conference on Machine Learning. Stanford, CA.
- 9. Tibshirani R, Walther G, Hastie T (2001) Estimating the number of clusters in a data set via the gap statistic. J R Stat Soc Ser B Stat Meth 63: 411-423.

### 3. Supporting Figure Legends

### Figure S1. Experimental design.

**Figure S2. Primary measurement data on nine viral intermediates.** Progression was measured by qPCR (Early RT, Late RT, 2LTR, and Integrated viral DNAs), RT–qPCR (Multiply spliced, Singly spliced, and Unspliced viral transcripts), FACS (GFP mean fluorescent intensity), and ELISA (viral p24 release) every two hours and normalized with respect to the last time point (24hr).

**Figure S3. Technical variance as a function of mean for qPCR data.** The  $R^2$  statistic represents the described variance by the seven parallel fitted lines sharing the same slope (a = 1.99), and having independent intercepts for each of the seven qPCR measurement sets.

**Figure S4. Viral progression model fitted curves:** Panel A illustrates the model fits in the variance stabilized coordinates (logarithmic scale). Panel B illustrated the fits when transformed back to the original coordinates of measurements (See figure S3). Blue dots indicate the experimental measurement values, and solid red lines are the fitted curves with the dashed lines representing the 95% confidence intervals of the fit derived by linear approximation around the fitted parameters using the Jacobian matrix.

**Figure S5. Estimated activity and progression of viral replication steps.** Accounting for the initial viral input, experimental noise, and decay of the measured species yields a refined estimation of activity rates (panel A) and the progression (panel B) of the viral cycle steps. The vertical dashed lines indicate the peak of activity. The values are normalized such that asymptotically progression reaches 100%.

**Figure S6. Qualitative assessment of the predictive power and stability of the fits.** Shown are the fits using full data (solid dark blue), data missing 24hr. (dashed blue), and data missing 22hr. and 24hr. (dashed light blue) time points extrapolated 10 hours beyond the original experiment. This mimics the case that the experiment was finished earlier than 24 hours. Fits derived after removing the last time point are all very close to the full model. This still remain the case for all viral steps after removing the two last time points, except for the case of viral release.

Figure S7. Determining the number clusters. BIC score (Mean  $\pm$  one standard deviation calculated from 500 bootstrapped samples) of the clustering for different numbers of clusters (k). The global minimum of the BIC was reached at  $k^* = 23$  (red asterisk). The number of clusters,  $k^+ = 18$  (green asterisk), was chosen as the most parsimonious solution within one standard deviation from the global minimum (dashed red line).

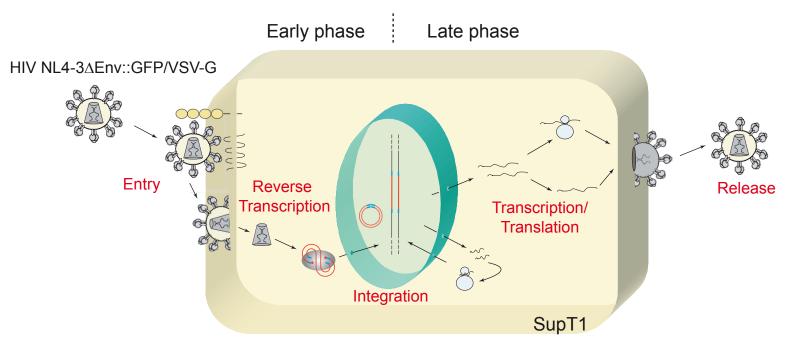
Figure S8. Clusters of host genes correlated with viral progression. Temporal expression patterns of 7991 genes were grouped into 18 clusters with differential expression profiles at three phases of the viral life cycle, namely reverse transcription, integration, and late phase. The boxplots on the left show the distribution of the normalized regression weights for each viral phase used in clustering. Cluster codes were defined for the three phases using characters '+' and '-' marking significant ( $p < 10^{-2}$ ) upregulation and downregulation, respectively, and 'o'

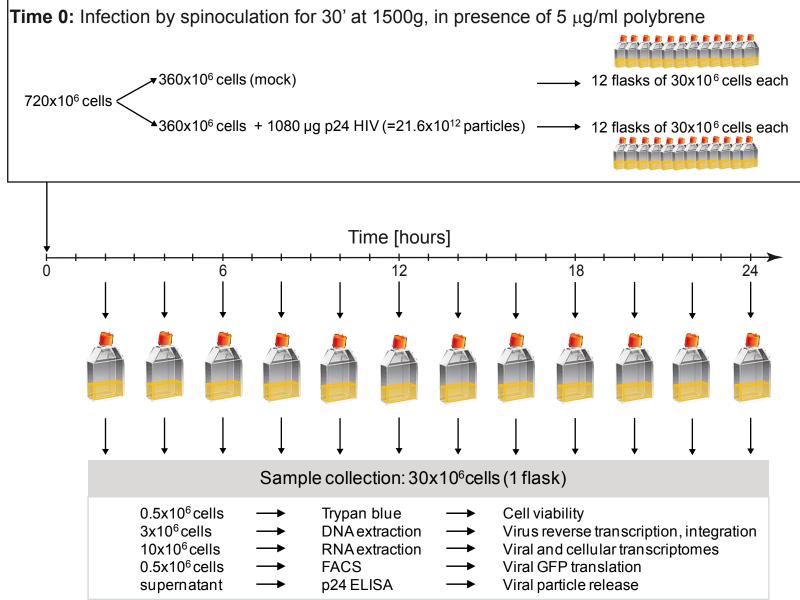
indicating no significant deviation. In total, six upregulated clusters (A), four clusters with mixed patterns of regulation (B), and eight downregulated clusters (C) were found. Expression change pattern of each gene over time is plotted for mock and HIV-1 samples (dotted orange lines) along with the cluster median (bold red line). Details of clusters are available at the dedicated web resource.

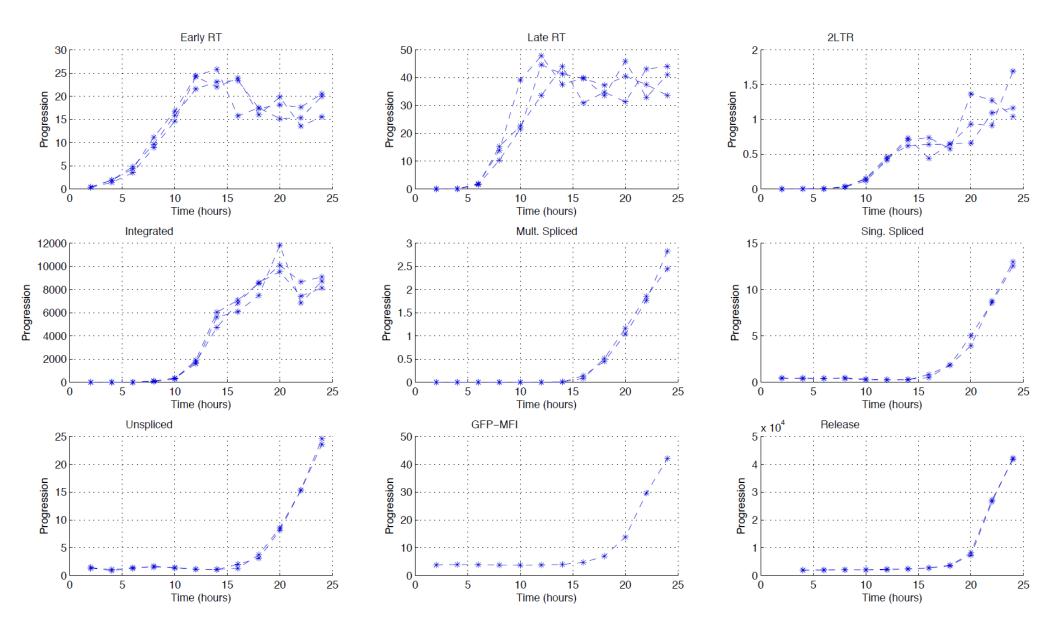
**Figure S9. Expected population-level changes due to viral integration.** Shown is the empirical null distribution of expression in mock samples and the corresponding empirical p-value boundaries. The blue line represents the expected population effect of viral integration assuming expression knock-out in the host gene (first scenario). The red line represents the expected population effect of viral integration assuming an increase in expression in the host gene by a factor of 105 (second scenario). Given the expected frequency of viral integrations, neither of these two opposite extreme scenarios implies significant perturbation in the population-level expression as compared to the variation in expression of the mock samples.

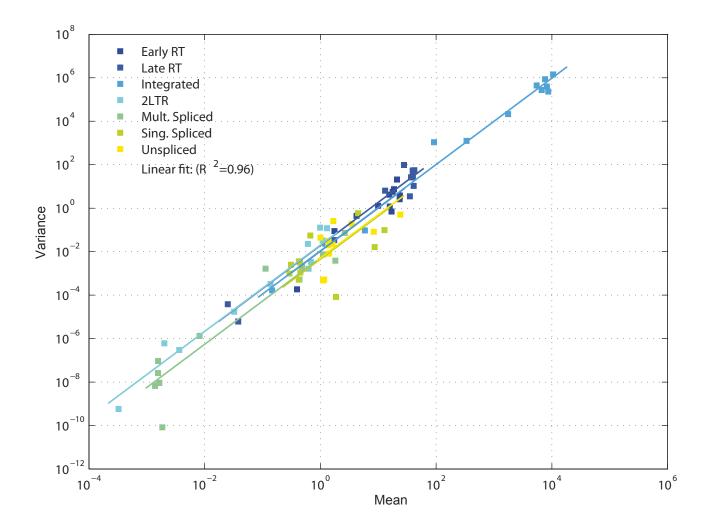
Figure S10. Infection efficiency assessed by FACS analysis of GFP expression. SupT1 cells (red and orange lines) or activated primary CD4+ T cells (dark and light blue lines) were transduced by VSV-G pseudotyped HIV vector carrying a *GFP* reporter gene. SupT1 cell infection was carried out with 3  $\mu$ g p24 equivalent of HIV-based vector, either competent (red line), or heat-inactivated (hi; red dash line) or a 1:10 mix of competent and heat-inactivated virus (orange line). Expression of GFP was assessed by FACS over time. Data from original infection of SupT1 cells were also plotted (black line).

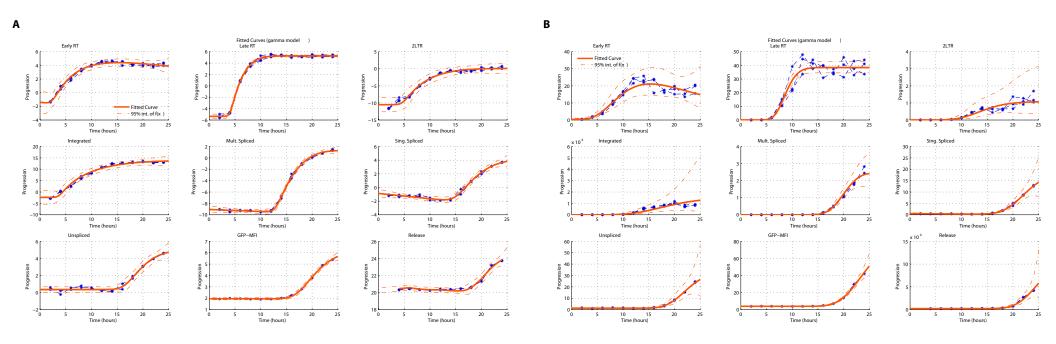
Figure S11. Correlation analysis between SAGE-Seq and RT-qPCR. Fourteen genes representative of diverse clusters were compared between the two techniques used, SAGE-Seq and RT-qPCR, by plotting the log<sub>2</sub> fold change of HIV-1 over mock. Each dot corresponds to one selected gene at one time point. Correlation analysis of the log<sub>2</sub> fold change of HIV-1 over mock was calculated by linear regression as  $r^2 = 0.5183$ ,  $p < 10^{-4}$ . (Spearman r = 0.7869,  $p < 10^{-4}$ ). Correlation improves towards late time points from  $r^2 = 0.23$  at 2 hours (p = 0.01) to  $r^2 = 0.77$  at 24 hours ( $p < 10^{-4}$ ).

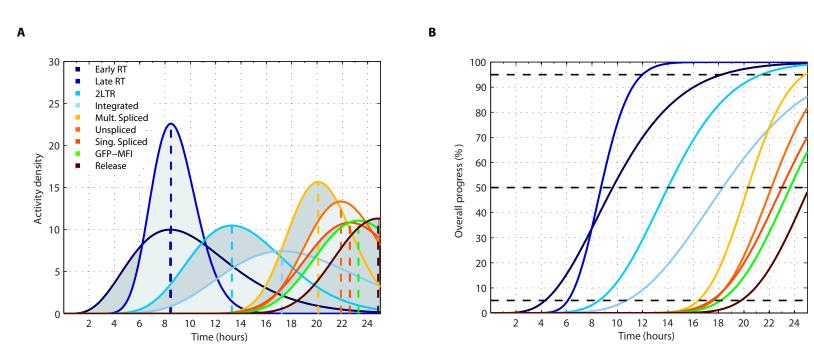


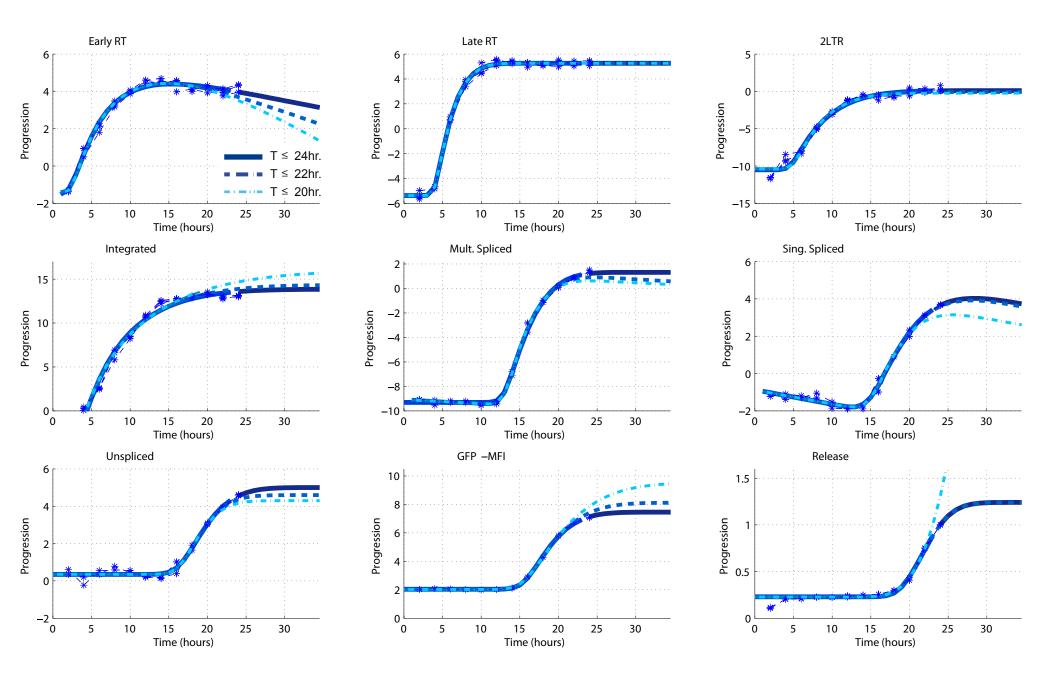


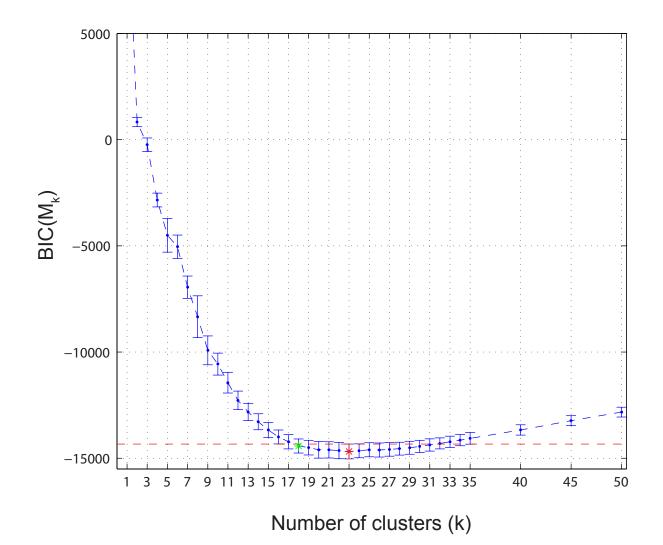


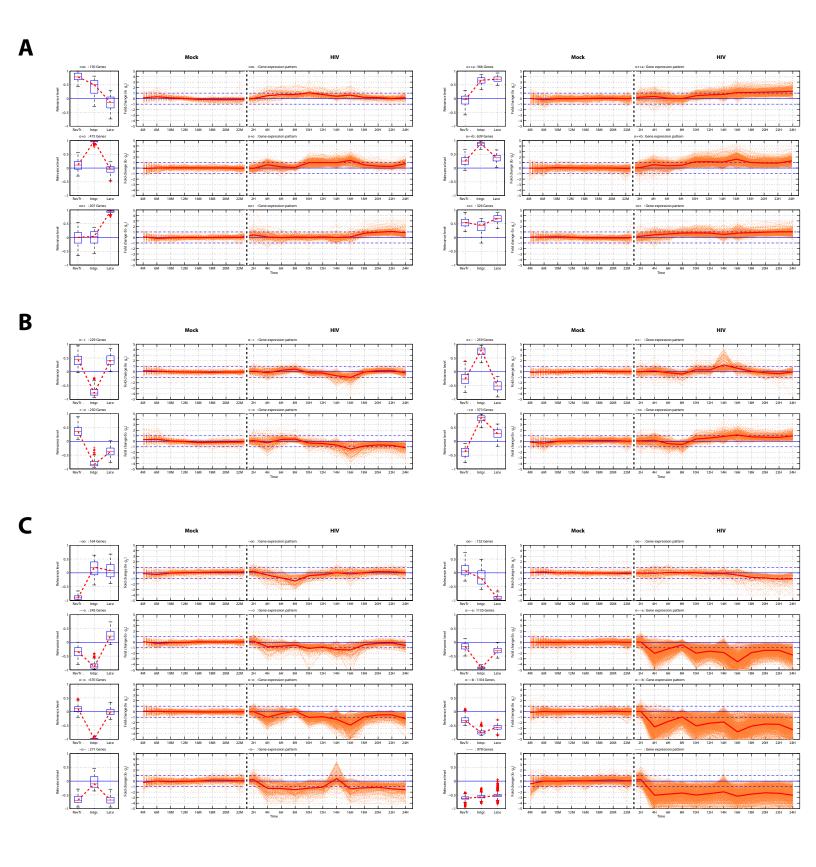


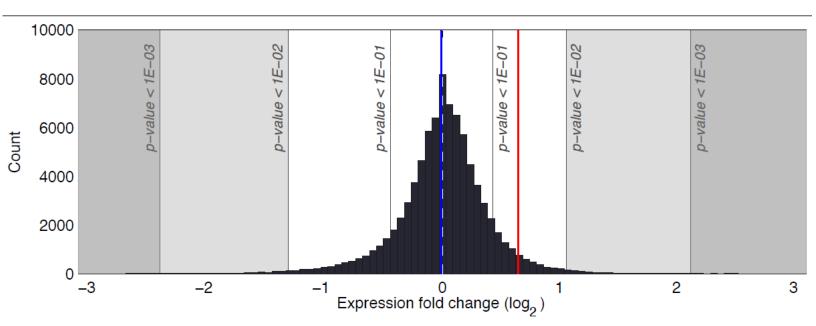


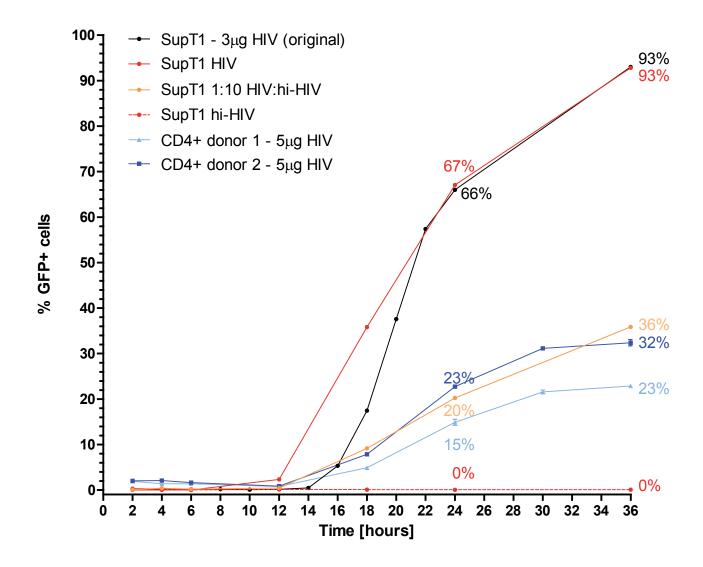




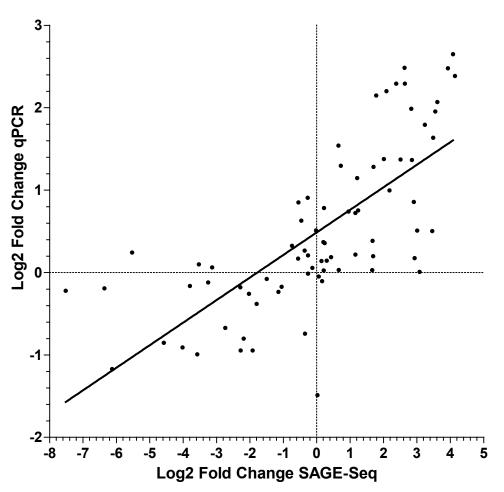












### 4. Supplementary Tables

### Supplementary Tables S1 and S2

Analysis of transcriptional regulators. Of 1391 transcription factors from the curated set defined by Vaguerizas et al. [1], 612 were found expressed, of which 421 (69%) were modulated in concordance with viral progression features, with 38%, 21%, and 10% in down-, up-, and mixed regulated clusters, respectively; a pattern shared with the general distribution of cellular transcripts. Genes sharing transcription factor binding motifs were inspected for evidence of co-regulation. Nine gene clusters were found to be enriched in targets of at least one transcription factor. Out of the total 25 cognate transcription factors identified with coregulated target genes, 17 were expressed in our system and 10 of them showed the same direction of modulation in expression as expected from their target set (Table S1). We detected 399 cellular miRNAs from the miRbase database [2], out of which 176 (44%) showed modulation through viral progression. The distribution across clusters of expressed miRNAs was 17%, 17%, and 10% in down-, up-, and mixed-regulated clusters, respectively, indicating a lower enrichment of miRNAs in down-regulated clusters compared to the distribution of cellular mRNAs ( $p < 10^{-7}$ ). Genes sharing 3'UTR miRNA binding sites were inspected for evidence of co-regulation. Seven gene clusters were found to be enriched for a number of miRNA targets. Out of the total 17 miRNA motifs identified with co-regulated target genes, 14 were expressed in our system. Four of them (let7b, miR101, miR124, miR142) showed the expected modulation in expression that could be expected from the predicted target set (Table S2).

#### References

- Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM (2009) A census of human transcription factors: function, expression and evolution. Nat Rev Genet 10: 252-263.
- 2. Kozomara A, Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res 39: D152-157.

TF target set	Enriched cluster pattern	Expressed TF	TF patter
V\$PPAR_DR1_Q2		PPARA	0++a
V\$E2F_03	0a	ND	
V\$TEL2_Q6	0a	ETV7	
V\$NRF2_01	0a	GABPB1	+-0
RCGCANGCGY_V\$NRF1_Q6	0a	NRF1	NA
GGGCGGR_V\$SP1_Q6	0a	SP1	oa
V\$E2F_Q6_01	0a	TFDP1/E2F	- 0-
V\$NFMUE1_Q6	ob	ND	
V\$E2F1_Q6_01	ob	E2F1	ob
V\$E2F1_Q4	ob	E2F1	ob
TGTTTGY_V\$HNF3_Q6	ob	FOXA1	
GGGAGGRR_V\$MAZ_Q6	ob	MAZ	oa
V\$MYCMAX_01	ob	MYC	ob
CGTSACG_V\$PAX3_B	ob	PAX3	
V\$E2F_Q3_01	ob	TFDP1/E2F	-0-
V\$E2F1_Q4_01	ob	TFDP1/E2F	-0-
V\$ATF6_01	00-	ATF6	oa
V\$CDX2_Q5	00-	CDX2	
V\$CREB_Q4	00-	CREB1	0+-
V\$CREB_Q2_01	00-	CREB1	0+-
TGGAAA_V\$NFAT_Q4_01	00-	NFAT	oa, ob
V\$OCT1_06	00-	POU2F1	NA
V\$YY1_01	00-	YY1	ob
CGTSACG_V\$PAX3_B	-00	PAX3	
V\$SRF_Q6	-00	SRF	0+-
V\$RORA1_01	O+-	RORA	NA
V\$YY1_01	0+-	YY1	ob
V\$USF_C	+-0	ND	
V\$CEBP_01	+-0	CEBPA	
V\$CEBPB_01	+-0	CEBPB	+-0
V\$GATA6_01	+-0	GATA6	
V\$OCT1_07	- +0	POU2F1	NA
V\$CREBP1_01	+00	ATF2	NA
V\$HLF_01	+00	HLF	
V\$E4BP4_01	+00	NFIL3	- +0
TTAYRTAA_V\$E4BP4_01	+00	NFIL3	- +0
not detected			

Table S2 : Enriched sets of	bi mikiwa target genes		
Enriched sets in clusters	Enriched Cluster pattern	Expressed cognate miRNAs	miRNA pattern
Targets of let-7b-5p	0a	hsa-let-7b	+0+
Targets of MIR 124-3p	oa	hsa-miR-124	+00
Targets of MIR 7-5p	oa	hsa-miR-7	-0-
Targets of MIR 133a	ob	ND	
Targets of MIR 16-5p	ob	hsa-miR-16	NA
Targets of MIR 34a-5p	ob	ND	
Targets of MIR 98	ob	hsa-miR-98	NA
Targets of MIR 99a-5p	ob	hsa-miR-99a	NA
Targets of MIR 30	0	hsa-miR-30e, hsa-miR-30b	00-
Targets of MIR 130b-3p	00-	hsa-miR-130b	00-
Targets of MIR 19b-3p	00-	hsa-miR-19b	00-
Targets of MIR 335-5p	-+0	ND	
Targets of MIR 101-3p	0+-	hsa-miR-101	oa
Targets of MIR 125a-5p	+00	hsa-miR-125a-5p	+0+
Targets of MIR 192-5p	+0+	hsa-miR-192	+00
Targets of MIR 21-5p	+0+	hsa-miR-21	NA
Targets of MIR 101-3p	o++b	hsa-miR-101	oa
Targets of MIR 142-3p	o++b	hsa-miR-142-3p	00-
Targets of MIR 192-5p	o++b	hsa-miR-192	+00
ND: not detected			
NA: not associated with viral progression	n features		

Map-244GIGCCCCTICITICITICITICITICITICITICITICIT	Primer name	5'-3' primer sequence	Orientation	Target (HIV position)	Purpose
MA.p.273     CHAGAGATCCCTCAGACCCTTTAGTCAGTGTGG     FAM-TANRA puebe     US (57-57)     ack RT       MA.p.236     TOTTOCCCGCTGAGAGAGTC     rev     psi (69/48)     late RT       MA.p.237     CAGTGGCGCCCGAACAGGGA     FAM-TANRA puebe     PSi (69/48)     late RT       MA.p.237     ACATGGGGCCCCCAACAGGGA     FAM-TANRA puebe     PSi (69/48)     late RT       MA.p.237     ACATGGGGACCCCGGACTGGTTAGA     Rvd     RVD38-M400     late RT       MA.p.238     TOCACAGATCATGGGGACTGCAGGGCAGGCTTT     Rvd     NdD (51/28)     late Registed, fini FCR       MA.p.238     GCCTCCCAAAGTGCGTGGAGAGCTTT     Rvd     ANd AND (51/28)     attegrated, nested PCR       MA.p.239     GCCTCCACAAGTGCATGAGGGTGGA     rev     late (51/25-59)     mtegrated, nested PCR       MA.p.231     GCCGCACATGAAGGGGTGGA     rev     con 5 (452-450)     mtegrated, nested PCR       MA.p.232     AGGGATTGACTGAGGGCTGGA     rev     con 5 (452-450)     mtegrated, nested PCR       MA.p.232     AGGGATTGACTGAGGGCTGGA     rev     con 5 (452-450)     mtegrated, nested PCR       MA.p.233     AGGGATTGACATGAGGGGTCTGGA     rev     con 5 (453-	MA.pr-243	GTGCCCGTCTGTTGTGTGAC	fwd	U5 (560-579)	early RT
Mp.948 IFGTGGCCQGCGGTGTGTGTGT evd pis(09460) lac RT   Mp.9266 AGGTGGGGCGGAGAGAGATC evc pis(09460) lac RT   Mp.9278 AGGTGGGGCGGAGAGAGT fxd RQ49480) 2.11R circles   Mp.9278 AGGTGGGGGGGGGGATAGG fxd RQ49480) 2.11R circles   Mp.928 TGCAGAGATGAGGATATCTTGTGC rxd U3 (51-26) 2.11R circles   Mp.928 GGCAGCGAATGGGGGATTAGA Md Nd (bost) mcgrand, fint PCR   Mp.928 GGCAGCGAATGGGGGGGGGATGA Md Nd (bost) mcgrand, fint PCR   Mp.928 GGCGTGGGATGTGACTGAGGGGGATTGA Nd Nd (bost) mcgrand, fint PCR   Mp.928 GGCATGTGAGCGACTGAGGGGGGGATGA Nd Nd (bost) mcgrand, fint PCR   Mp.928 GGCATGTGAGTGAGATGAAGCGCTGTGA rev U5 (563-90) mcgrand, instel PCR   Mp.929 GGCAGTGAGTCAGCGCGCTGAGA Nd mton 5 (980-340) HMBS (PBGD) (GancD 3165)   Mp.929 GGCAGGTGAGTCAGCAGCTGTGT NCMGB_NPQ pub mtorn 5 can 5 (482-4850) HMBS (PBGD) (GancD 3165)   Mp.929 GGCAGGGAGCTCAGCAGCGCC rev enc 7 (482-4857) millips splecit (182 b class)   Map 200 GGGATGTGGTCTGTGAGGAGCGCA rev enc 7 (482-4857) millips splecit (182 b class) </td <td>1A.pr-244</td> <td>GGCGCCACTGCTAGAGATTT</td> <td>rev</td> <td>U5-PBS (642-623)</td> <td>early RT</td>	1A.pr-244	GGCGCCACTGCTAGAGATTT	rev	U5-PBS (642-623)	early RT
Ap p236 GAGTCCTGOCGAGAGAGTC ev pH096000 bac RT   Ap p237 AGTGGGGGCCCGAGAGGAG FAM-TAMR probe PH86(83462) bac RT   Ap p238 TCGAGGAGTCACTGCTTAAG fxd R(9128-465) bal R et RT   Ap p238 TCGAGGAGTCACTGGAGGAGGTTT FAM-TAMR probe R(9128-465) bal R ecteds   Ap p238 GCCTCCAAAGTCAGAGGCAGGTTT FAM-TAMR probe R(9128-900) megrated, first PCR   Ap p238 GCCTCCAAAGTGGTGGGATACCA Fvd Alk obs1 megrated, first PCR   Ap p238 GCCTCCAAAGGCATGGAGGTA Fvd R(9128-900) megrated, netred PCR   Ap p234 CCCGCTCGTTGTGGCTGAAGGTAG Fvd R(9128-900) megrated, netred PCR   Ap p234 CCCGCTCGTTGTGGCTGAAGCTAG Fvd R(9128-900) megrated, netred PCR   Ap p234 CCGGCGATTGGAGCTCTGG ev cons (4184-430) HMIS (FKD) (netred) 146)   Ap p234 CCGGCGATTGGAGCTCATGGGCTT fvd mtrov fvd megrated, netred PCR   Ap p234 CCGGCGATTGGAGCTCATGGGCTT fvd mtrov fvd megrated, netred PCR   Ap p234 CCGGCGATTGGAGCTGAGGAGCAAGCCTGT fvd mtrov fvd mtrov fvd   Ap p234 CCGGGCAGTTGGAGCTGAGCGCGTG ev cons (4184-430) HMIS (FKD) (netre) 136	1A.pr-275	CTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGG	FAM-TAMRA probe	U5 (588-621)	early RT
MA.pr.27 CAGTGGGCGCGAACAGGGA FAMT-ANRA probe PK (9428-9490) LLTR cicks   MA.pr.247 ACATGGGGAACAGGACTGTTAAG fed R (9428-9490) LLTR cicks   MA.pr.247 TCCACAGATCAAGGGATACTGTGTC rev U3 (5): 280 LLTR cicks   MA.pr.247 ACACTACTTTGAGGACTCAAGGCAAGCTT FAMT-ANRA probe R-15 (5): 480-9487) LLTR cicks   MA.pr.249 GCCTCCCAAATGCTGGGGATTACA Kod Mat (5): 480 mitsgrated, fist PCR   MA.pr.253 GCCTCGCAACTGCTTGCTGGTAATAGGGCTGAC Rv US (622-599) mitsgrated, nested PCR   MA.pr.253 GCCTGTGGTGGTGGGGATTACA Rv US (622-599) mitsgrated, nested PCR   MA.pr.254 GCCATGTGTGTGGGATCAGGCTGCTG Rv Rv Rv Rv   MA.pr.253 GCGCATGTGTGGGAGATAAGGCTGT Rv Rv Rv Rv   MA.pr.254 GCGCATGTGCAGGGAGAAAGGCTGT Rv Rv Rv Rv   MA.pr.274 GCGGGAGCTATCAAGGCTCTTGGGAAGAAGGCA Rv Rv mittans (4)(4)(4)(1)(1)(1)(1)(1)(1)(1)(1)(1)(1)(1)(1)(1)	MA.pr-245	TGTGTGCCCGTCTGTTGTGT	fwd	U5 (557-576)	late RT
MA, pr.24 AACTAGGGAACCACTGCTTAAG ev U3(4):248   MA, pr.248 TOCACAGATACTIGGC ev U3(5):28) 24.1R cicks   MA, pr.248 TOCACAGATACTIGGCAAGCTTACAGGCAAGCTTT FAM-TANRA probe R-U5 (458-947) 24.1R cicks   MA, pr.249 GCCTTOCGACCCAAGGTCGTGAAGCTACA fwd M(100) micgrated, finit PCR   MA, pr.249 GCCTTOCGACCCCATCTCTTCTCC ev ga (80-372) micgrated, nist PCR   MA, pr.231 GCCTCATAAAGCGTGCTTGA fwd Md (252) micgrated, nist PCR   MA, pr.232 TOCACACTGACTACAGGTCTGATACTAG fwd micgrated, nist PCR   MA, pr.233 TOCACACTGACATCAGCTCTGTTACTAG fwd micgrated, nist PCR   MA, pr.234 CCCGTCTTGTGTGGCACTCGAGGTCTTTC fwd micgrated, nist PCR   MA, pr.235 AGAGGATTGCACCAGGCTCCTTTG fwd micro fx884-430) HMIS (MSD) (GenelD 145)   MA, pr.237 CCGGGCAGATTGAGAGGAAAAGCCTGT fwd micro fx884-430) HMIS (MSD) (GenelD 145)   MA, pr.238 GGAGTCTGCTCTCTCTCTCTCTCTCCCCC fwd micro fx884-430) HMIS (MSD) (GenelD 145)   MA pr.239 CCGGCGAGATTGAGAGGAAAAGCCTGT fwd micro fx884-430) mitry spliced (18/2 16/2 16/2 16/2 16/2 16/2 16/2 16/2 16	MA.pr-246	GAGTCCTGCGTCGAGAGATC	rev	psi (699-680)	late RT
MA.pr.28 TCCACAGATCAAGGATATCHGTC ev U (91-28) 24.TR circles   MA.pr.276 ACACTACTTTGACGACTCAAGGCAACCTTT FAM-TAMRA probe R45 (9488-9467) 24.TR circles   MA.pr.249 GCCTCCCAAAGTCGTGGGATTACA fxd Ahu (host) integrated, first PCR   MA.pr.251 GCCTCCCAACACTCTCTOTCC ev gg (905-72) integrated, nested PCR   MA.pr.252 GCCCTCAATAAGGGTGGCTTGA fxd R(522-54) integrated, nested PCR   MA.pr.252 CCCCCTGACTGAAGGAGCAGCATGGCTGGA ev U (5(62-59)) integrated, nested PCR   MA.pr.253 AAGGGATTCACTGAGGCATGTG fxd inton 5 (4308-4330) IMMIS (PBGD) (GeneID:3145)   MA.pr.254 GCGCGGACATGAGGAGGAGAAAGCGTGT fxd inton 5 (4308-4330) IMMIS (PBGD) (GeneID:3145)   MA.pr.274 GCGGAGGTTGAGGCAGCAGCAGCAGCGT rv env/rv env/rv2 (8308-276) maliply spliced (18:21b class)   mtR4 ACGTCAGACCACATCACAGCGCCTCTTACAAGCCA fxd tat/rvv1 (011-6044) maliply spliced (18:21b class)   mtR2 GGACGGGAACTTGAGGCGCA rv env/rvv2 (8308-276) maliply spliced (18:21b class)   mtR3 GGACTCATCACAGCACTATCAACGATTCATCTACAGCA fxd env/rvv2 (8308-276) maliply spliced (18:21b class)   mtR2 GGGAGGGACCCACACATCAGCACGACCACC rv	MA.pr-276	CAGTGGCGCCCGAACAGGGA	FAM-TAMRA probe	PBS (633-652)	late RT
MA.pr.26     CACCTACTTCAAGCCAAGGCAAGCTTT     FAM-TAMRA probe     RJ/S (94S8-9487)     2LTR cricks       MA.pr.29     GCCTTCCCAAAGTGCTGGGAGTTACA     find     Alu (host)     integrated, fini PCR       MA.pr.291     GCCTTCCCAACGTCCTCTCTCC     rev     gg (807-782)     integrated, fini PCR       MA.pr.292     GCCTCCCAACTAAAGCGTTGCA     find     integrated, nested PCR       MA.pr.292     CCCAACGTGCCTTGGTGAACTAG     find     introm 5 (4384-4330)     HMIS (PBCD) (GeneID345)       MA.pr.293     AGGGATTCACAGGCTCTTGG     find     introm 5 (4384-4330)     HMIS (PBCD) (GeneID345)       MA.pr.294     GCGCTGTGTGAGGACTCATGGGACCTAG     rev     end (482-1463)     HMIS (PBCD) (GeneID345)       MA.pr.294     GGCATGTCACAGACTCCTTGG     rev     end (482-1463)     HMIS (PBCD) (GeneID345)       MA.pr.294     CGGCAGTCGCAAGGCCCCTATCAAAGCCTGT     rev     end (482-1463)     multiply spleed (18.2 bb class)       mtB4     GAGTTGTCACAGACGCCCAC     rev     end (482-1463)     multiply spleed (18.2 bb class)       mtB4     GGGGAGCCCAAGGCCCC     rev     end (493-422)     singly spleed (14.2 bb class)       mtB4     GGGGAGCCCGAAGGGCCC	MA.pr-247	AACTAGGGAACCCACTGCTTAAG	fwd	R (9428-9450)	2-LTR circles
MA.pr-249 GCCTCCCAAAGTGCTGGGGATTACA fed Alu (host) megrated, first PCR   MA.pr-230 GCTCTCGCACCAACGTGCTTCCC rev gg (80-782) megrated, first PCR   MA.pr-231 GCCTCGCACCACGTTGCTCA fed R(52-243) megrated, nested PCR   MA.pr-234 CCCGTCGTGTGACTGGCTGA rev U5 (62-599) megrated, nested PCR   MA.pr-234 CCCGTCGTGTGTGGGACCCATGGCTGAC fed intron 5 (4388-433) HMRS (PBGD) (GeneID 345)   MA.pr-234 GGCAGTGATTGAAGGAGAAAAGCCTGT fred intron 5 (4382-436) HMRS (PBGD) (GeneID 345)   MA.pr-234 GGCAGTGATTGAAGGAGAAAAGCCTGT WCAGB_NPQ probe intron 5 (4382-436) HMRS (PBGD) (GeneID 345)   MA.pr-237 CCGGCAGATTGAAGGCTCATCGAAGGCTCTCTCTATCAAAGCA fred intron 1 (6011-6041) matiply spliced (1.82 2b class)   mt245 AGGGAACCCAACGCCCC rev env/rat/ruv2 (830-827) mtiply spliced (1.82 2b class)   mt222 GGCAGGGATATTGACGATTGAGA fred env/ruv2 (830-827) mtiply spliced (1.82 2b class)   mt224 AGGGAACCCAACGCCCCC rev env/ruv2 (830-827) singly spliced (1.82 2b class)   mt224 GGCAGGGATATTGACGATTGAGA fred env/ruv2 (830-827) singly spliced (1.82 2b class)   mt225 GGCAGGGATATTGACGATTGATGGTTTCACAC rev	MA.pr-248	TCCACAGATCAAGGATATCTTGTC	rev	U3 (51-28)	2-LTR circles
MA.pr-29     GCCTCCCAAAAGCTGCGGGGATTACAA     Nod     Nuthons)     megrated, first PCR       MA.pr-230     GCCTCGCAAAAAGCTGCTCTCC     rev     gg (803-702)     integrated, first PCR       MA.pr-231     GCCTCGCAATAAAGCTGCTCTCGA     rev     U5 (622-599)     megrated, nested PCR       MA.pr-232     CCCACGTGCGTTGACATCAGCATGACTAGG     FAM-TAMRA probe     U5 (632-591)     megrated, nested PCR       MA.pr-234     CCGCGCGGTTGTGACATCTGGCTTTGC     find     mitron 5 (4308-4330)     HMSR (PBCD) (GencID 345)       MA.pr-234     GGCAGTGTCAAGCAGAGAAAGCCTGT     rev     exon 5 (4324-436)     HMSR (PBCD) (GencID 345)       MA.pr-247     CCGGCGCAATTGCAAGGCTCATTGCAAAGCCTGT     rev     exon 5 (4324-436)     HMSR (PBCD) (GencID 345)       MA.pr-278     CGGGAGCTCTCTCTGTCTCTCTACAAAGCTGTTCCTATCAAAGCA     fird     ttt/rvv1 (6011-604)     multiply spliced (1.82 bb class)       mt884     GGAGTCTGTCTCTGTCTCTCTCTCTCCACC     rev     env/ra2/rvv2 (8304-827)     multiply spliced (1.82 bb class)       mt224     GGCAGGGATTGACGCATTGAGAGCAC     FAM-TAMRA probe     env/ra2/rv2 (8404-827)     singly spliced (1.82 bb class)       mt225     GGAGGGATTGACGCATTGAGGGCC     rev     env/ra2/rv2	MA.pr-276	ACACTACTTTGAGCACTCAAGGCAAGCTTT	FAM-TAMRA probe	R-U5 (9458-9487)	2-LTR circles
MA.pr.250 GCTCT0GACCCATCTCTCCC rv gg (80-782) integrated, inst PCR   MA.pr.251 GCCTCAATAAAGCTTGCTCGC Nvd R (22-54) integrated, insted PCR   MA.pr.252 TCCACACTGACTTAAAGGTCTGA rvv U5 (63-59) integrated, insted PCR   MA.pr.254 GCCGTCTGTTGTGTGACTCTGGTAACTGA FAM-TAMRA probe U5 (63-59) integrated, insted PCR   MA.pr.254 AGGGATTGCAGGGCTCTTGT field integrated, insted PCR   MA.pr.254 GGCATGGAGAGAAAGCCCGTGG vv exo field Sta2-463) IMBS (PBG) (GnelD 3145)   MA.pr.254 GGCATGGAGCACATCAAGCCTCTTAGA field tat/irvv1 (6011-6044) maliply spliced (18.2 kb class)   mf84.AK145 ACAGTCAGACTCATCAAAGCTTTCTACAAGCA field tat/irvv1 (6011-6044) maliply spliced (18.2 kb class)   mf84.AK145 GGACTGGAGCAACAGGCCC field env (809-8222) singly spliced (18.2 kb class)   mf84. GGACTGGACTGACTTATAGCTTTCAGA field env (809-8227) singly spliced (4 kb class)   mf85 GGACTGGACTGACTAGGCCC field pol (256-2562) unspliced (9 kb class)   mf84 AGGGATAGTGATGTAGTGCTT field pol (266-2561) unspliced (9 kb class)   mf84 AGGGAGACCGACAGGACCC field pol (266-2562) unspliced (9 kb class) </td <td></td> <td>GCCTCCCAAAGTGCTGGGATTACA</td> <td></td> <td>Alu (host)</td> <td>integrated, first PCR</td>		GCCTCCCAAAGTGCTGGGATTACA		Alu (host)	integrated, first PCR
MA.pr-21     GCCTCAATAAAGCTTGCCTTGA     Nul     R (32,54)     integrated, nested PCR       MA.pr-23     TCCACACTGACTAAAAGCGTTGA     rev     U5 (62,599)     integrated, nested PCR       MA.pr-230     CCCGCTGTGTTGGGTACTGGGTAACTAG     FAM-TAMRA probe     U5 (66,541)     integrated, nested PCR       MA.pr-231     AGGCAAGTCAACTGAGCGCTCTTTG     Null     inton 5 (4382-4453)     HMBS (PRG2) (cnelD:3145)       MA.pr-232     CCGGCAAGTTGGAAGCAGCAGGCGT     VCAGB NPQ probe     inton 5 -exon 5 (4384-4453)     HMBS (PRG2) (cnelD:3145)       MA.pr-23     CCGGCAAGTTGGAAGCTCATCAAAGCCTGT     VCAGB NPQ probe     inton 5 -exon 5 (4384-453)     HMBS (PRG2) (cnelD:3145)       MA.pr-23     CCGGCAAGTTGGAAGCTCATCAAAGCCTGT     VCAGB NPQ probe     inton 5 -exon 5 (4384-453)     mBtply spliced (18.2 b class)       MB24     AGGGCACCGACCCACAGCCCCC     rev     env/rev2 (820-827)     maltply spliced (18.2 b class)       M226     AGGGGACCGACAGGCCCC     rev     env/rev2 (820-827)     singly spliced (4 b class) + unspliced (9 b class)       M28     GGACGTGTGTGTGTGTGTGTCTCTCACCAC     rev     env/rev2 (820-827)     singly spliced (4 b class) + unspliced (9 b class)       M26     AGGGGACCCGACAGGCCCC	-	GCTCTCGCACCCATCTCTCCC	rev		
MA.pr-252   TCCACACTGACTAAAAGGGTCTGA   rev   U5 (62-599)   integrated, nested PCR     MA.pr-234   CCCGTCTGTGTGGTAACTCAGGTACTGG   FAM-TAMRA prob   U5 (663-591)   integrated, nested PCR     MA.pr-234   AGGGATTGCAGGCTCTTGG   fvd   inton 5 (4388-430)   HMIS (PRGD) (GeneID-3145)     MA.pr-234   GGCATGTCAAGCTCCTTGG   rev   exon 5 (4382-4363)   HMIS (PRGD) (GeneID-3145)     MA.pr-234   GGCAGGTCATCAAGACAGCTGT   VCAGB NPQ probe   inton 5-oxon 5 (4382-4363)   HMIS (PRGD) (GeneID-3145)     MA.pr-234   GGCAGGTCTCTGTGCTCTCTGTCACAGCTGTCTATCAAAGCA   fvd   tatl/rev1 (6011-6044)   maltiply spliced (1.8.2 bk class)     m84   ACAGGCAGCACGCCCCC   FAM-TAMRA probe   env/rev2 (830-8276)   maltiply spliced (1.8.2 bk class)     m825   GGCAGGGCACGACGCCCC   FAM-TAMRA probe   env/rev2 (830-8276)   singly splied (4 bk class) + unsplied (9 bk class)     m826   GGCAGGGACCGCACAGGCCC   FAM-TAMRA prob   env(rat2rvev2 (820-8276)   singly splied (4 bk class) + unsplied (9 bk class)     m826   GGCAGGGACCGCGCACAGGCCCC   FAM-TAMRA prob   pol (236-263)   unsplied (9 bk class)     m827   GGCAGGGACGCGACAGGCCC   FAM-TAMRA prob   pol (236-263)   unsplied (9 bk class) </td <td>-</td> <td>GCCTCAATAAAGCTTGCCTTGA</td> <td>fwd</td> <td></td> <td></td>	-	GCCTCAATAAAGCTTGCCTTGA	fwd		
MA.pr.234   CCCGTCTIGTIGTGTGACTCTGGTAACTAG   FAM-TAMRA probe   US (563-591)   integrated, nested PCR     MA.pr-235   AAGGGATTCACTCAGCGCGTTTTC   fvd   intron 5 (438-4330)   HMBS (PRGD) (cenelD:3145)     MA.pr-237   CCGGCAGATTGGAGGAGAAAAGCCTGG   rev   exm 5 (4382-436).   HMBS (PRGD) (cenelD:3145)     MA.pr-237   CCGGCAGATTGGAGGAGAAAAGCCTGT   VICMGB_NFQ probe   intron 5-exon 5 (4334-4359)   HMBS (PRGD) (cenelD:3145)     mtB4   ACAGTCAGACTCATCAAGCTTCTCTACAAGGCA   fvd   tat1/rev1 (6011-6044)   multiply spliced (1.82 kb class)     mtB3   GGACTGTTACAGACTCATCACGCCC   rev   env/tat2/rev2 (820-827)   multiply spliced (1.82 kb class)     mt22   GGCAGGGATATTCACCATTATCGTTTCAGA   fvd   env/tat2/rev2 (820-827)   singly spliced (4 kb class) + unspliced (9 kb class)     mt23   GGCATGTTAAATTTGCACTATATCGTTTCAGA   fvd   env/tat2/rev2 (820-827)   singly spliced (4 kb class) + unspliced (9 kb class)     mt24   AGGGGACCGACAGGGCCC   rev   env/tat2/rev2 (820-827)   singly spliced (4 kb class) + unspliced (9 kb class)     mt25   AGGGGACCGACAGGGCCC   FAM-TAMRA probe   env/tat2/rev2 (820-827)   singly spliced (4 kb class) + unspliced (9 kb class)     mt26   GGCAGTTAAAGTTTTACCATTATGGTTTATTC </td <td>-</td> <td></td> <td>rev</td> <td></td> <td></td>	-		rev		
MA.pr-233   AAGGGATTCACTCAGGCTCTTTC   fwd   intron 5 (4384-430)   HMBS (PBGD) (GeneID:3145)     MA.pr-234   GGCATTGTCAAGCTCCTTGG   rev   ewn 6 (4382-436)   HMBS (PBGD) (GeneID:3145)     MA.pr-279   CCGCGCAGATTGGAGAGAAAGCCTGT   VIC/MGB_NFQ probe   intron 5-exon 5 (4334-4359)   HMBS (PBGD) (GeneID:3145)     mtB4-AK145   ACAGTCAACCTCCTCTGTCTCTCTCACAGCA   fwd   tu1/rev1 (6011-6044)   mulphy spliced (1.82 kb class)     mtB3   GGACATTCTCTGTCTCTCTCTCACACC   rev   env/rev2 (820-8277)   mulphy spliced (1.82 kb class)     mt224   GGCAGGGATTTCACCATTATCGTTTCAGA   fwd   env(rev2 (820-8277)   singly spliced (1.4b class) + unspliced (9 kb class)     mt224   GGCAGGGATCTCATCTGCTCTCTCACACC   rev   env/rev2 (820-8277)   singly spliced (1.4b class) + unspliced (9 kb class)     mt224   GGCAGGGATCTTATCATTCGTTTCACACC   rev   env/rev2 (820-8277)   singly spliced (1.4b class) + unspliced (9 kb class)     mt225   AGGGGACCCGACAGGCCC   FAM-TAMRA probe   env(rev2 (820-8277)   singly spliced (1.4b class) + unspliced (9 kb class)     mt26   AGGGGACCGAATGATGATGGCC   rev   rev = nv1/co <sup>2</sup> (280-8264)   unspliced (9 kb class)     mt27   GGCAATTTATATTTTTTTTTCC   rev   rev =	•				
MA.pr-254GGCATGITICAAGCTCCTIGGrevexon 5 (4382-4363)HMBS (PBGD) (GeneID:3145)MA.pr-279CGCAGATTIGGAGGAGAAAGCCTGATMCAB_NPQ probeintron 5-exon 5 (434-4359)HMISS (PBGD) (GeneID:3145)ntR4-AK145ACAGTCAAGCTCATCAAGCTCATCAAAGCAArevmatipv spiced (1.82 kb class)matiply spiced (1.82 kb class)ntR54AGGGGACCGACAGGCCCrevenv(rev2 (8302-827)matiply spiced (1.82 kb class)nt220GGCAGGGATATTCACCATTATCGTTTCAGArevenv(rev2 (8302-827)matiply spiced (1.82 kb class)nt231GGCAGGGATATTCACCATTATCGTTTCAGArevenv(rev2 (8302-827)singly spiced (1 kb class) + unspiced (9 kb class)nt232GGCAGGGACAGGCCrevenv(rev2 (8302-827)singly spiced (1 kb class) + unspiced (9 kb class)nt234AGGGGACCGACAGGCCrevenv(rev2 (8302-827)singly spiced (1 kb class) + unspiced (9 kb class)nt235GGACTTGTATATTTTCCTATTAGTCTAfvdpol (235-263)unspiced (9 kb class)nt236CAAATTTCTACTAATGCTTTATGTTTCACAfvdexon (145-64)unspiced (9 kb class)nt304CAAATTTCTACAATTGCACTTAGTGCTAfvdexon (145-64)GAPDH (GeneID: 297)nt305CGACAGGAATGGCCGCGATTTGCTGTCTfvdexon (145-64)GAPDH (GeneID: 297)nt406GGCAACATATCCACTTTAGCGGATGGTGTframeexon (145-64)GAPDH (GeneID: 297)nt407AGGCGGGGGGAGGAAGCGCGCTTAAGGGACCATGantisenseintegration site determination, linker Msent408GGCAACATATCCACTTAAGCGGAGAGACCGTantisenseintegration site determination, linker Ms	-				
MA.pr-279     CCGGCAGATTGGAGAGAAAAGCCTGT     VIC.MGB_NFQ probe     intron 5-exon 5 (4334-4359)     HMBS (PBGD) (GeneID:3145)       MA.pr-279     CCGGCAGACTCATCAAGGCTCTCTATCAAAGCAA     fwd     tat1/rev1 (6011-6044)     multiply spliced (1.8/2 kb class)       m83     GGATCTGTCTGTCTCTGTCTCCACC     rev     env/rev2 (8302-827)     multiply spliced (1.8/2 kb class)       m222     GGCAGGGATTGTCCTGTCTCTGTCCACAT     fwd     env/rev2 (8302-827)     singly spliced (4 kb class) + unspliced 9 kb class)       m223     GGCAGGGATTGTCTGTGTCTGTGTCTGTCTCACAC     rev     env/rev2 (8302-827)     singly spliced (4 kb class) + unspliced 9 kb class)       m224     AGGGGACCCGACAGGGCC     rev     env/rat2/rev2 (8302-827)     singly spliced (4 kb class) + unspliced 9 kb class)       m225     GGCAGTTTAAATTTGCCATTAGTCTTA     fwd     pol (256-250)     unspliced 9 kb class)       m248     AAGCGCAGGATGGACGCAC     FAM-TAMRA probe     pol (256-2634)     unspliced 9 kb class)       m454     AGGGGAGTCAGATGGCC     FAM-TAMRA probe     pol (256-204)     unspliced 9 kb class)       m454     AGGGCGAGTATTACCAGTTTACCAG     rev     exon 1 (45-64)     GAPDH (GeneID: 297)       m454     AGGGTCGGAGTCAACTGACTTAAGGGAGAC	-	GGCATGTTCAAGCTCCTTGG	rev	,	
Imital Instruction Instruction   milling ACAGTCAGACTCATCAAGCTTCTCTATCAAAGCA fived tat1/rev1 (6011-6044) multiply spliced (1.8/2 kb class)   milling GGATCTGTCTCTGTCTCTCTCCCACC rev env/rat2/rev2 (8302-8276) multiply spliced (1.8/2 kb class)   milling GGACCGGACAGCGCC FAM-TAMRA probe env/rat2/rev2 (840-8257) multiply spliced (1.8/2 kb class)   milling GGACCGGACAGCCCC FAM-TAMRA probe env/rat2/rev2 (840-8257) singly spliced (1 kb class) + unspliced (9 kb class)   milling GGACTGTCTCTGTCTCTCTCTCCCACC rev env/rat2/rev2 (840-8257) singly spliced (4 kb class) + unspliced (9 kb class)   milling GGACCCGACAGCGCC FAM-TAMRA probe env/rat2/rev2 (840-8257) singly spliced (4 kb class) + unspliced (9 kb class)   milling GGCACCCGACAGCGCC FAM-TAMRA probe env/rat2/rev2 (840-8257) singly spliced (4 kb class) + unspliced (9 kb class)   milling GGCACCGCAGCAGCCC FAM-TAMRA probe pol (2536-2562) unspliced (9 kb class)   milling CAAATTGCTTTATTATTTTTTTTTTTTTTTTTTTTTTTT	•				
ntll33   GGATCTGTCTCTGTCTCTCTCCACC   rev   env/rev2 (8302-8276)   miltiply spliced (1.8/2 kb class)     ntl226   AGGGGACCCGACAGGCCC   FAM-TAMRA probe   env/rev2 (8240-8257)   multiply spliced (1.8/2 kb class)     ntl22   GGCAGGGATATTCACCATTATCGTTTCAGA   fivd   env (8193-8222)   singly spliced (1.4/2 kb class) + unspliced (9 kb class)     ntl236   AGGGGACCCGACAGGCCC   rev   env/rev2 (8302-8276)   singly spliced (1.4/c class) + unspliced (9 kb class)     ntl236   AGGGGACCCGACAGGCCC   rev   env/rev2 (8240-8257)   singly spliced (1.4/c class) + unspliced (9 kb class)     ntl236   AGGGGACCCGACAGGCCC   FAM-TAMRA probe   env/rev2 (8240-8257)   singly spliced (1.4/c class) + unspliced (9 kb class)     ntl236   AGGGGACCCGACAGGCCC   FAM-TAMRA probe   env/rev2 (8240-8257)   unspliced (9 kb class)     ntl236   AGGGGACCCGACAGGACTGCT   rev   pol (266-264)   unspliced (9 kb class)     ntl348   AAGCCAGGAATGGATGGCC   FAM-TAMRA probe   exon 1 (45-64)   GAPDH (GeneID: 2597)     ntl446   GGCAACAATATCCACTTTACCAG   rev   exon 3 (2070-2048)   GAPDH (GeneID: 2597)     ntl446   GGCGGGGGGTCAGCGTAAGGGCG   rev   exon 3 (2070-2048)   GAPDH (GeneID: 2597)	1		_ (1		
AGGGACCCGACAGGCCC   FAM-TAMRA probe   env/tat2/rev2 (8248-8257)   multiply spliced (1.8/2 kb class)     m222   GGCAGGGATATTCACCATTATCGTTTCAGA   fwd   env (8193-8222)   singly spliced (1.8/2 kb class)     m183   GGATCGTCTCTGTCTCTCTCTCCCACC   rev   env/rev2 (830-8227)   singly spliced (1.8/2 kb class)     m1226   AGGGGACCCGACAGGCCC   rev   env/rev2 (830-827)   singly spliced (1.8/2 kb class)     m1299   GCACTTTAAATTTCCCCATTAGTCCTA   fwd   env/rat2/rev2 (8240-8257)   singly spliced (1.8/2 kb class)     m1299   GCACTTTAAATTTCCCATTAGTCCTA   fwd   pol (256-2562)   unspliced (9 kb class)     m302   CAAATTTCTACTAATGCTTTATTTTTTTTC   rev   pol (256-2634)   unspliced (9 kb class)     m434   AAGCCAGGAATGGATGGACG   FAM-TAMRA probe   pol (2586-2604)   unspliced (9 kb class)     m434   AGGCAGGCAATTTCACCTTT   fwd   exon 1 (45-64)   GAPDH (cnerD: 2597)     m445   TCGACAGTCAACGGACTTT   fwd   exon 1 (45-64)   GAPDH (cnerD: 2597)     m470q   AAGGTCGGAGTCAACGGATTGGCCG   rev   exon 3 (355-371/204-2013)   GAPDH (cnerD: 2597)     m470q   AAGGTCGGAGTCAACGAGATAGGGGACCGATTAGGGGACCA   integration site determina	nf84-AK145	ACAGTCAGACTCATCAAGCTTCTCTATCAAAGCA	fwd	tat1/rev1 (6011-6044)	multiply spliced (1.8/2 kb class)
H222GGCAGGGATATTCACCATTATCGTTTCAGAfvdenv (8193-8222)singly spliced (4 kb class) + unspliced (9 kb class)mR3GGATCTGTCTCTGTCTCTCTCCCACCrevenv/rev2 (8302-8276)singly spliced (4 kb class) + unspliced (9 kb class)mR26AGGGGACCCGACAGGGCCFAM-TAMRA probeenv/rat2/rev2 (8240-8257)singly spliced (4 kb class) + unspliced (9 kb class)mR30GCAACTTTAAATTTTCCCATTAGTCCTAfivdpol (2536-2562)unspliced (9 kb class)mR302CAAATTGCTACTAATGCTTTATTTTTTCrevpol (256-2634)unspliced (9 kb class)mR45TCGACAGTCAGCCGCACTCTTfivdexon 1 (45-64)GAPDH (GeneID: 2597)mR46GGCAACAATATCCACTTTACCAGrevexon 3 (2070-2048)GAPDH (GeneID: 2597)mR46GGCAACAATATCCACTTTACCAGrevexon 3 (355-371/2004-2013)GAPDH (GeneID: 2597)mR47Phosp]TAGTCCCTTAAGGGGAG-[AmC7-Q]senseintegration site determination, linker MsemA.pr-526[Phosp]TAGTCCCTTAAGGGGAG-[AmC7-Q]senseintegration site determination, linker MseMA.pr-528CTTAAGCACTCACTATAGGGGCCCCCCCTTAAGGGACCATGantisenseintegration site determination, linker NlaMA.pr-529GTAATACGACTCACTATAGGGCCCCCCCTTAAGGGACfivdintegration site determination, PCR1, HIV primerMA.pr-532GTAATACGACTCACTATAGGGCCCGCTTAAGGGACfivdintegration site determination, PCR2, A-HIV primerMA.pr-532GTAATACGACTCACTATAGGGCCGCGCTTAAGGGACfivdintegration site determination, PCR2, A-HIV primerMA.pr-532GTAATACGACTCACTATAGGGCCGCGTTAAGGGACfivdintegration	mf83	GGATCTGTCTCTCTCTCCACC	rev	env/rev2 (8302-8276)	multiply spliced (1.8/2 kb class)
ntl83GGATCTGTCTGTCTCTCTCCCACCrevenv/rev2 (8302-8276)singly spliced (4 kb class) + unspliced (9 kb class)ntl226AGGGGACCCGACAGGCCCFAM-TAMRA probeenv/tat2/rev2 (8240-8257)singly spliced (4 kb class) + unspliced (9 kb class)ntl299GCACTTTAAATTTTCCCATTAGTCCTAfvdpol (255-262)unspliced (9 kb class)ntl302CAAATTTCTACTAATGCTTTTATTTTTTCrevpol (262-2634)unspliced (9 kb class)ntl348AAGCCAGGAATGGATGGCCFAM-TAMRA probepol (268-2604)unspliced (9 kb class)ntl45TCGACAGTCAGCCGCACTTTfvdexon 1 (45-64)GAPDH (CnerD): 2597)ntl46GGCAACAATATCCACTTTACCAGrevexon 3 (2070-2048)GAPDH (CnerD): 2597)ntl70tqAAGGTCGGAGTCAACGGATTTGGTCGTFAM-TAMRA probeexon 2 exon 3 (355-371/2004-2013)GAPDH (GneID): 2597)ntl70tqAGGTCCGGAGTCAACGGAGTTTGGCGTsenseintegration site determination, linker Msentl70tqAGGTCGGAGTCAACGGAGTAAGGGGACCATGantisenseintegration site determination, linker Msentl70tqGTAATACGACTCACTATAGGGCTCGCTTAAGGGACCantisenseintegration site determination, linker Msentl70tqGTAATACGACTCACTATAGGGCTCGCTTAAGGGACCATGantisenseintegration site determination, linker Msentl70tqGTAATACGACTCACATATAGGGCTCGCCTTAAGGGACCATGantisenseintegration site determination, linker Nlant_pr-526GTAATACGACTCACATATAGGGCTCGCCTTAAGGGACCATGantisenseintegration site determination, PCR, I, HIVprimernt_pr-528GTAATACGACTCACATATAGGGCTCGCCTGAGGGAGAfvdintegratio	mf226	AGGGGACCCGACAGGCCC	FAM-TAMRA probe	env/tat2/rev2 (8240-8257)	multiply spliced (1.8/2 kb class)
htt226AGGGGACCCGACAGGCCCFAM-TAMRA probeenv/tat2/rev2 (8240-8257)singly spliced (4 kb class) + unspliced (9 kb class)ht1299GCACTTTAAATTTTCCCATTAGTCCTAfwdpol (2536-2562)unspliced (9 kb class)ht1302CAAATTTCTACTAATGCTTTATTTTTTCrevpol (2662-2634)unspliced (9 kb class)ht1348AAGCCAGGAATGGATGGCCFAM-TAMRA probepol (2586-2604)unspliced (9 kb class)ht1345TCGACAGTCAGCCGCATCTTfwdexon 1 (45-64)GAPDH (GenelD: 2597)ht1346GGCAACAATATCCACTTTACCAGrevexon 3 (2070-2048)GAPDH (GenelD: 2597)ht1347AAGGTCGGAGTCAACGGATTIGGTCGTFAM-TAMRA probeexon 2 exon 3 (355-371/2004-2013)GAPDH (GenelD: 2597)ht1348GGCAACAATATCCACTTAAGGGAG-[AmC7-Q]senseintegration site determination, linker Mseht1349MAGGTCGGAGTCACCTATAGGGGACCGCTTAAGGGACantisenseintegration site determination, linker Mseht1349MA.pr-526[Phosp]GTCCTTAAGCGGAG-[AmC7-Q]senseintegration site determination, linker Mseht1352GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACCATGantisenseintegration site determination, linker Nlaht1354GTAATACGACTCACTATAGGGCTCAGCTTAAGGGACCATGfwdintegration site determination, PCR1, linker primerht1355GTAATACGACTCACTATAGGGCTCAGCTTAAGGGACCfwdintegration site determination, PCR2, A-HIV primerht1355GTAATACGACTCACTATAGGGCTCAGCTTAAGGGACCfwdintegration site determination, PCR2, A-HIV primerht1355GTAATACGACTCACTATAGGGCTCAGCTTAAGGGACCfwdintegration site det	mf222	GGCAGGGATATTCACCATTATCGTTTCAGA	fwd	env (8193-8222)	singly spliced (4 kb class) + unspliced (9 kb class)
mt299   GCACTTTAAATTTTCCCATAGGTCCTA   fwd   pol (2536-2562)   unspliced (9 kb class)     mt302   CAAATTTCTACTAATGCTTTTATTTTTC   rev   pol (2662-2634)   unspliced (9 kb class)     mt348   AAGCCAGGCAATGGATGGCC   FAM-TAMRA probe   pol (2586-2604)   unspliced (9 kb class)     mt45   TCGACAGTCAGCCGCATCTT   fwd   exon 1 (45-64)   GAPDH (CeneID: 2597)     mt46   GGCAACAATATCCACTTTACCAG   rev   exon 3 (2070-2048)   GAPDH (CeneID: 2597)     mt70tq   AAGGTCGGAGTCAACGGATTGGTCGT   FAM-TAMRA probe   exon 2-exon 3 (355-371/2004-2013)   GAPDH (GeneID: 2597)     MA.pr-524   [Phosp]TAGTCCCTTAAGCGGAG-[AmC7-Q]   sense   integration site determination, linker Mse     MA.pr-525   GTAATACGACTCACTATAGGGGTCCGCTTAAGGGAC   antisense   integration site determination, linker Mse     MA.pr-526   [Phosp]GTCCCTTAAGCGCAGAC_IAmC7-Q]   sense   integration site determination, linker Na     MA.pr-527   GTAATACGACTCACTATAGGGCTCGCCTTAAGGGACCATG   antisense   integration site determination, PCR1, HIV primer     MA.pr-528   CTTAAGCGAGACAATAAGGGTGCCGCTTGAGG   fwd   integration site determination, PCR1, HIV primer     MA.pr-529   GTAATACGACTCACTATAGGGCCCCGCTTAAGGGAC   fwd <td>mf83</td> <td>GGATCTGTCTCTCTCTCCACC</td> <td>rev</td> <td>env/rev2 (8302-8276)</td> <td>singly spliced (4 kb class) + unspliced (9 kb class)</td>	mf83	GGATCTGTCTCTCTCTCCACC	rev	env/rev2 (8302-8276)	singly spliced (4 kb class) + unspliced (9 kb class)
m302CAAATTICTACTAATGCTTTATTITTTCrevpol (262-2634)unspliced (9 kb class)m348AAGCCAGGAATGGATGGCCFAM-TAMRA probepol (2586-2604)unspliced (9 kb class)m45TCGACAGTCAGCCGCGCATCTTfwdexon 1 (45-64)GAPDH (GeneID: 2597)m46GGCAACAATATCCACTTTACCAGrevexon 3 (2070-2048)GAPDH (GeneID: 2597)m470tqAAGGTCGGAGTCAACGGATTGGTCGTFAM-TAMRA probeexon 2 -exon 3 (355-371/2004-2013)GAPDH (GeneID: 2597)m470tqAAGGTCGGAGTCAACGGAGTAGCGGAG-[AmC7-Q]senseintegration site determination, linker MseMA.pr-524[Phosp]TAGTCCCTTAAGGGGTCCGCTTAAGGGACantisenseintegration site determination, linker MseMA.pr-525GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACsenseintegration site determination, linker MseMA.pr-526[Phosp]GTCCCTTAAGCGAGAG-[AmC7-Q]senseintegration site determination, linker NlaMA.pr-527GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACCATGantisenseintegration site determination, linker NlaMA.pr-528CTTAAGCCTCACTATAAGGCTCGCCTTGAGfwdintegration site determination, PCR1, HIV primerMA.pr-529GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACCfwdintegration site determination, PCR2, A-HIV primerMA.pr-530gcctectcogcgcctatgGCACTATAGGGCTCCGCTTTAAGGGACAATCfwdintegration site determination, PCR2, A-HIV primerMA.pr-532gccttgccagcccgtcatgCACTATAGGGCTCCGCTTTAAGGGACCACTfwdintegration site determination, PCR2, A-HIV primer	mf226	AGGGGACCCGACAGGCCC	FAM-TAMRA probe	env/tat2/rev2 (8240-8257)	singly spliced (4 kb class) + unspliced (9 kb class)
M348AAGCCAGGAATGGATGGCCFAM-TAMRA probepol (2586-2604)unspliced (9 kb class)m45TCGACAGTCAGCCGCATCTTfwdexon 1 (45-64)GAPDH (GeneID: 2597)m46GGCAACAATATCCACTTTACCAGrevexon 3 (2070-2048)GAPDH (GeneID: 2597)m770tqAAGGTCGGAGTCAACGGATTTGGTCGTFAM-TAMRA probeexon 2 exon 3 (355-371/2004-2013)GAPDH (GeneID: 2597)m770tqAAGGTCCGCAGTCAACGGAGTCAACGGAGTGTGGTFAM-TAMRA probeexon 2 exon 3 (355-371/2004-2013)GAPDH (GeneID: 2597)m770tqAAGGTCCGCAGTAACGGGAG-[AmC7-Q]senseintegration site determination, linker MsemA.pr-525GTAATACGACTCACTATAGGGGCTCCGCTTAAGGGACantisenseintegration site determination, linker MseMA.pr-526[Phosp]GTCCCTTAAGCGGGAG-[AmC7-Q]senseintegration site determination, linker MseMA.pr-527GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACCATGantisenseintegration site determination, linker NlaMA.pr-528CTTAAGCCTCAATAAAGCTTGCCTTGAGfwdintegration site determination, PCR1, HIV primerMA.pr-530gcctccctgcgccatcagCACTATAGGGCCCCGCTTAAGGGACfwdintegration site determination, PCR2, A-HIV primerMA.pr-532gccttgccagcccgtcag <u>TCATGAGC</u> AGACCCTTTTAGTCAGTGTGGGAAAATCfwdintegration site determination, PCR2, B-barcode Mse-linker primer	mf299	GCACTITAAATTITCCCATTAGTCCTA	fwd	pol (2536-2562)	unspliced (9 kb class)
m445TCGACAGTCAGCCGCATCTTfwdexon 1 (45-64)GAPDH (GeneID: 2597)m446GGCAACAATATCCACTTTACCAGrevexon 3 (2070-2048)GAPDH (GeneID: 2597)m70tqAAGGTCGGAGTCAACGGATTTGGTCGTFAM-TAMRA probeexon 2-exon 3 (355-371/2004-2013)GAPDH (GeneID: 2597)m70tqAAGGTCCGCAGCCAACGGAGTCAACGGAGTGAGCAGFAM-TAMRA probeexon 2-exon 3 (355-371/2004-2013)GAPDH (GeneID: 2597)mA.pr-524[Phosp]TAGTCCCTTAAGCGGAG-[AmC7-Q]senseintegration site determination, linker MseMA.pr-525GTAATACGACTCACTATAGGGGACCGGCTTAAGGGGACantisenseintegration site determination, linker MseMA.pr-526[Phosp]GTCCCTTAAGCGGAG-[AmC7-Q]senseintegration site determination, linker MseMA.pr-527GTAATACGACTCACTATAGGGGCTCCGCTTAAGGGGACCATGantisenseintegration site determination, linker NlaMA.pr-528CTTAAGCCTCAATAAAGCTTGCCTTGAGfwdintegration site determination, pCR1, HIV primerMA.pr-530gcctcctcggcgccatcagCACTATAGGGCTCCGCTTAAGGGACfwdintegration site determination, PCR2, A-HIV primerMA.pr-532gccttgccagcccgctcagTCAGGCAGAGCCCTTTTAGGTGTGGAAAATCrevintegration site determination, PCR2, B-barcode Mse-linker primer	mf302	CAAATTTCTACTAATGCTTTTATTTTTTC	rev	pol (2662-2634)	unspliced (9 kb class)
m46GGCAACAATATCCACTTTACCAGrevexon 3 (2070-2048)GAPDH (GeneID: 2597)m170tqAAGGTCGGAGTCAACGGATTTGGTCGTFAM-TAMRA probeexon 2-exon 3 (355-371/2004-2013)GAPDH (GeneID: 2597)m170tqIPhosp]TAGTCCCTTAAGCGGAG-[AmC7-Q]senseintegration site determination, linker MseMA.pr-524[Phosp]GTCCCTTAAGGGGAG-[AmC7-Q]senseintegration site determination, linker MseMA.pr-525GTAATACGACTCACTATAGGGGCTCCGCTTAAGGGACantisenseintegration site determination, linker MseMA.pr-526[Phosp]GTCCCTTAAGCGGAG-[AmC7-Q]senseintegration site determination, linker MseMA.pr-527GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACCATGantisenseintegration site determination, linker NlaMA.pr-528CTTAAGCCTCAATAAAGCTTGCCTTGAGfwdIntegration site determination, pCR1, HIV primerMA.pr-529GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACfwdintegration site determination, pCR2, A-HIV primerMA.pr-530gcctcctcggccatcagCACTATAGGGCACCCTTTAGTGGGAAAATCfwdintegration site determination, pCR2, A-HIV primerMA.pr-532gccttgccagccgtcagTCATGAGCAGCCCTTTAGTGGGAAAATCfwdintegration site determination, pCR2, A-HIV primer	mf348	AAGCCAGGAATGGATGGCC	FAM-TAMRA probe	pol (2586-2604)	unspliced (9 kb class)
MAGGTCGGAGTCAACGGATTTGGTCGTFAM-TAMRA probeexon 2-exon 3 (355-371/2004-2013)GAPDH (GeneID: 2597)MA.pr-524[Phosp]TAGTCCCTTAAGCGGAG-[AmC7-Q]senseintegration site determination, linker MseMA.pr-525GTAATACGACTCACTATAGGGGTCCGGCTTAAGGGACantisenseintegration site determination, linker MseMA.pr-526[Phosp]GTCCCTTAAGCGGAG-[AmC7-Q]senseintegration site determination, linker MseMA.pr-527GTAATACGACTCACTATAGGGGCTCCGCTTAAGGGACCATGantisenseintegration site determination, linker NlaMA.pr-528CTTAAGCCTCAACTATAGGGCTCCGCTTGAGfwdintegration site determination, pCR1, HIV primerMA.pr-529GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACCfwdintegration site determination, pCR2, A-HIV primerMA.pr-530gcctcctcggccatcagCACTATAGGGCAGAGCCCTTTAGGTGTGGAAAATCrevintegration site determination, PCR2, A-HIV primerMA.pr-532gccttgccagcccgtcagTCATGAGGAGACCCTTTTAGTCAGTGTGGGAAAATCrevintegration site determination, PCR2, B-barcode Mse-linker primer	mf45	TCGACAGTCAGCCGCATCTT	fwd	exon 1 (45-64)	GAPDH (GeneID: 2597)
MA.pr-524[Phosp]TAGTCCCTTAAGCGGAG-[AmC7-Q]senseintegration site determination, linker MseMA.pr-525GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACantisenseintegration site determination, linker MseMA.pr-526[Phosp]GTCCCTTAAGCGGAG-[AmC7-Q]senseintegration site determination, linker NaMA.pr-527GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACCATGantisenseintegration site determination, linker NlaMA.pr-528CTTAAGCCTCACTATAGGGCTCCGCTTGAGfwdintegration site determination, PCR1, HIV primerMA.pr-529GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACCrevintegration site determination, PCR1, linker primerMA.pr-530gcctcctcggccatcagCACTATAGGGCTCCGCTTAAGGGAACATCrevintegration site determination, PCR2, A-HIV primerMA.pr-532gccttgccagcccgctcagTCAGGCAGACCCTTTTAGTCAGTGTGGAAAATCrevintegration site determination, PCR2, B-barcode Mse-linker primer	mf46	GGCAACAATATCCACTTTACCAG	rev	exon 3 (2070-2048)	GAPDH (GeneID: 2597)
MA.pr-525GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACantisenseintegration site determination, linker MseMA.pr-526[Phosp]GTCCCTTAAGCGGAG-[AmC7-Q]senseintegration site determination, linker NlaMA.pr-527GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACCATGantisenseintegration site determination, linker NlaMA.pr-528CTTAAGCCTCACTATAAGGGCTCCGCTTGAGfwdintegration site determination, PCR1, HIV primerMA.pr-529GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACCrevintegration site determination, PCR1, linker primerMA.pr-530gcctcctcggccatcagCACTATAGGGCTCCGCTTAAGGGAACATCrevintegration site determination, PCR2, A-HIV primerMA.pr-532gccttgccagcccgtctag <u>TCATGAGC</u> AGACCCTTTTAGTCAGTGTGGAAAATCrevintegration site determination, PCR2, B-barcode Mse-linker primer	mf70tq	AAGGTCGGAGTCAACGGATTTGGTCGT	FAM-TAMRA probe	exon 2-exon 3 (355-371/2004-2013)	GAPDH (GeneID: 2597)
MA.pr-525GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACantisenseintegration site determination, linker MseMA.pr-526[Phosp]GTCCCTTAAGCGGAG-[AmC7-Q]senseintegration site determination, linker NlaMA.pr-527GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACCATGantisenseintegration site determination, linker NlaMA.pr-528CTTAAGCCTCACTATAGGGCTCCGCTTGAGfwdintegration site determination, PCR1, HIV primerMA.pr-529GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACCrevintegration site determination, PCR1, linker primerMA.pr-530gcctcctcggccatcagCACTATAGGGCTCCGCTTAAGGGAACATCrevintegration site determination, PCR2, A-HIV primerMA.pr-532gccttgccagcccgtcagTCATGAGGCAGACCCTTTAGTCAGTGTGGAAAATCrevintegration site determination, PCR2, B-barcode Mse-linker pri	MA				
A. pr-526[Phosp]GTCCCTTAAGCGGAG-[AmC7-Q]senseintegration site determination, linker NlaMA.pr-527GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACCATGantisenseintegration site determination, linker NlaMA.pr-528CTTAAGCCTCAATAAAGCTTGCCTTGAGfwdintegration site determination, PCR1, HIV primerMA.pr-529GTAATACGACTCACTATAGGGCrevintegration site determination, PCR1, linker primerMA.pr-530gcctcctcggccatcagCACTATAGGGCTCCGCTTAAGGGAACATCfwdintegration site determination, PCR2, A-HIV primerMA.pr-532gccttgccagcccgtctagTCATGAGCAGACCCTTTAGGGAACATCrevintegration site determination, PCR2, A-HIV primerMA.pr-532gccttgccagcccgtctagTCATGAGCAGACCCTTTAGTCAGTGTGGAAAATCrevintegration site determination, PCR2, B-barcode Mse-linker primer	•				
MA.pr-527   GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACCATG   antisense   integration site determination, linker Nla     MA.pr-528   CTTAAGCCTCAATAAAGCTTGCCTTGAG   fwd   integration site determination, PCR1, HIV primer     MA.pr-529   GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACCATG   rev   integration site determination, PCR1, linker primer     MA.pr-530   gcctccctcggcccatcagCACTATAGGGCTCCGCTTAAGGGACC   fwd   integration site determination, PCR2, A-HIV primer     MA.pr-532   gccttgccagcccgctcagTCATGAGGCAGACCCTTTAGGGAACATC   rev   integration site determination, PCR2, A-HIV primer					
MA.pr-528   CTTAAGCCTCAATAAAGCTTGCCTTGAG   fwd   integration site determination, PCR1, HIV primer     MA.pr-529   GTAATACGACTCACTATAGGGC   rev   integration site determination, PCR1, linker primer     MA.pr-530   gcctccctcgcgccatcagCACTATAGGGCTCCGCTTAAGGGAC   fwd   integration site determination, PCR2, A-HIV primer     MA.pr-532   gccttgccagcccgctcagTCATGAGGCAGACCCTTTAGGGACAAAATC   rev   integration site determination, PCR2, B-barcode Mse-linker primer					
MA.pr-529   GTAATACGACTCACTATAGGGC   rev   integration site determination, PCR1, linker primer     MA.pr-530   gcctccctcgcgccatcagCACTATAGGGCTCCGCTTAAGGGAC   fwd   integration site determination, PCR2, A-HIV primer     MA.pr-532   gccttgccagcccgctcagTCATGAGGCAGACCCTTTAGTGAGGAAAATC   rev   integration site determination, PCR2, B-barcode Mse-linker primer	-				
MA.pr-530   gcctccccgcgccatcagCACTATAGGGCTCCGCTTAAGGGAC   fwd   integration site determination, PCR2, A-HIV primer     MA.pr-532   gccttgccagccgctcagTCATGAGCAGACCCTTTTAGTCAGTGGGAAAATC   rev   integration site determination, PCR2, B-barcode Mse-linker primer					
MA.pr-532 gccttgccagcccgctcagTCATGAGCAGACCCTTTTAGTCAGTGTGGAAAATC rev integration site determination, PCR2, B-barcode Mse-linker pr	-				
	•				
MA.pr-534 gccugccagcccgcucag <u>cCaCGATG</u> AGACCCTTTAGTCAGTGTGGGAAAATC rev integration site determination, PCR2, B-barcode Nla-linker pri	•				
	MA.pr-534	gccttgccagcccgctcag <u>CTACGATG</u> AGACCCTTTTAGTCAGTGTGGAAAATC	rev		integration site determination, PCR2, B-barcode Nla-linker primer

Table S4 : Gene expression assays			
Gene Symbol	ENSG_ID	Gene Name	AB Assay ID
DZIP3	ENSG00000198919	DAZ interacting protein 3, zinc finger	Hs00978125_m1
GIMAP6	ENSG00000133561	GTPase, IMAP family member 6	Hs00226776_m1
GNG2	ENSG00000186469	guanine nucleotide binding protein (Gprotein), gamma 2	Hs00828232_m1
GRIP1	ENSG00000155974	glutamate receptor interacting protein 1	Hs00402711_m1
HDGF	ENSG00000143321	hepatoma-derived growth factor	Hs00610314_m1
HSP90AB1	ENSG0000096384	heat shock protein 90kDa alpha (cytosolic), class B member 1	Hs01546474_g1
IFI16	ENSG00000163565	interferon, gamma-inducible protein 16	Hs00194261_m1
IGSF3	ENSG00000143061	immunoglobulin superfamily, member 3	Hs00155437_m1
LY96	ENSG00000154589	lymphocyte antigen 96	Hs01026734_m1
PIGS	ENSG0000087111	phosphatidylinositol glycan anchor biosynthesis, class S	Hs00264209_m1
RCBTB2	ENSG00000136161	regulator of chromosome condensation (RCC1) and BTB (POZ) domain containing protein 2	Hs01048815_m1
RPL31	ENSG0000071082	ribosomal protein L31	Hs01015497_g1
RPS9	ENSG00000170889	ribosomal protein S9	Hs02339424_g1
SLC7A5	ENSG0000103257	solute carrier family 7 (amino acid transporter light chain, L system), member 5	Hs00185826_m1
SPON2	ENSG00000159674	spondin 2, extracellular matrix protein	Hs00202813_m1