Discovery and assessment of new target sites for anti-HIV therapies: an approach to utilize genome wide gene expression changes and computational models.

Sanjive Qazi, Gustavus Adolphus College, St. Peter, MN.

Introduction

- HIV infection of human immune cells is a good model system to study gene activation pathways.

- HIV-1 contains six additional regulatory and accessory proteins (tat,rev, vif,vpu,vpr and nef) that are required for the integration of the virus into the human cell and its replication.

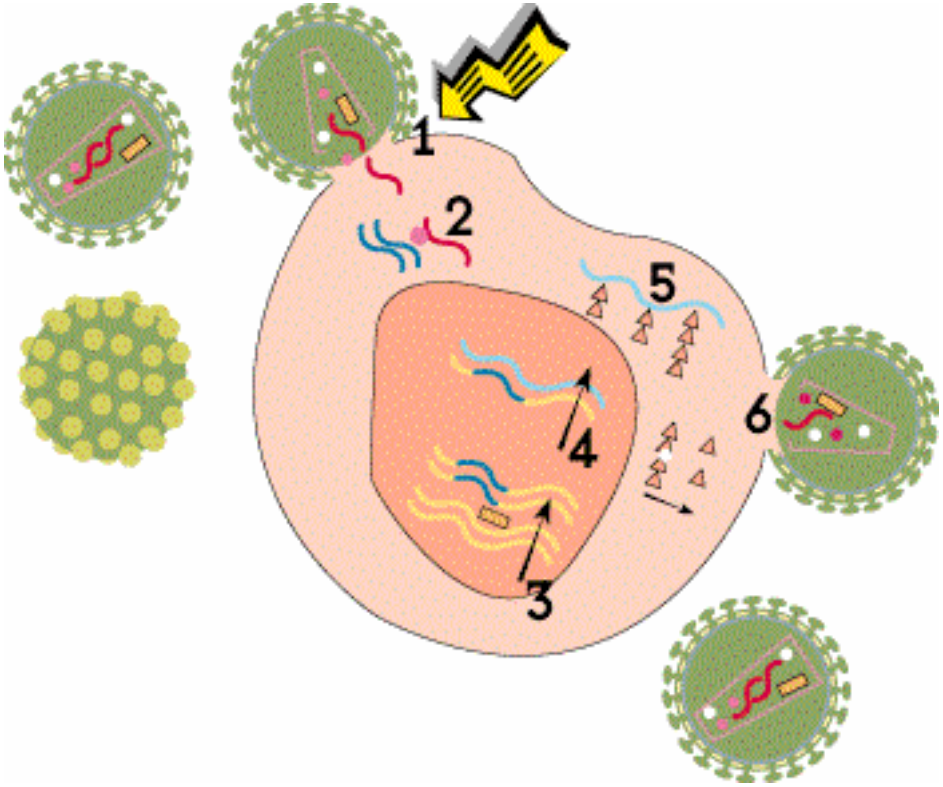
- Two types of studies can begin to decipher the complex interplay between virus proteins and thousands of host cell proteins to escape cell defense mechanisms and replicate more virus particles: RNAi knockdown and Gene Expression profiling.

- Presently the viral enzymes are important target sites for anti-HIV therapeutics. Since viral replication also requires host cell proteins, identification of these targets would greatly expand the number of strategies for drug development: reduce the development of drug resistant strains as host cell proteins will also be targeted; design of complementary therapies to act in synergy with common anti-viral drugs; enhance specific immune responses of cells infected with virus; inhibit key steps in viral replication such as integration of viral DNA or assembly of viral particles requiring host proteins.

3D Animation of HIV infection.

http://www.scienceinreview.com/2007/3d-animation-of-hiv-infection.html

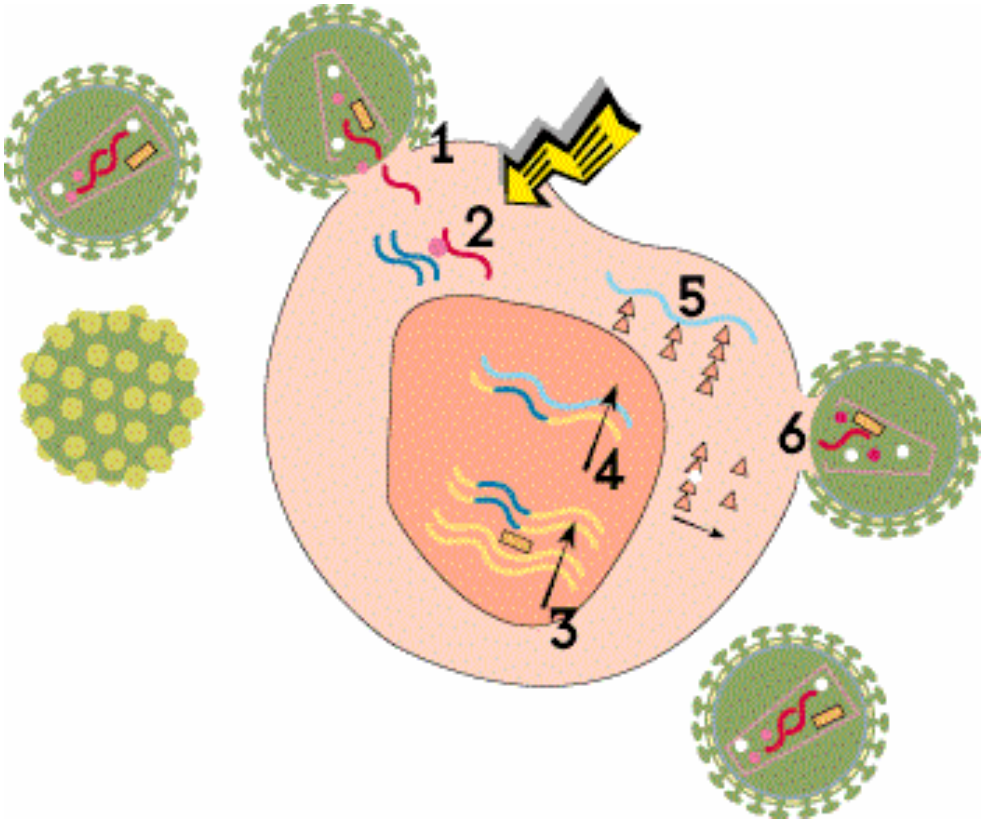
The HIV Life Cycle



Step 1: Binding

A virus consists of an outer envelope of protein, fat and sugar wrapped around a set of genes (in the case of HIV, genetic information is carried as RNA instead of DNA) and special enzymes.

HIV has proteins on its envelope that are strongly attracted to the CD4+ surface receptor on the outside of the T4-cell. When HIV binds to a CD4+ surface receptor, it activates other proteins on the cell's surface, allowing the HIV envelope to fuse to the outside of the cell.

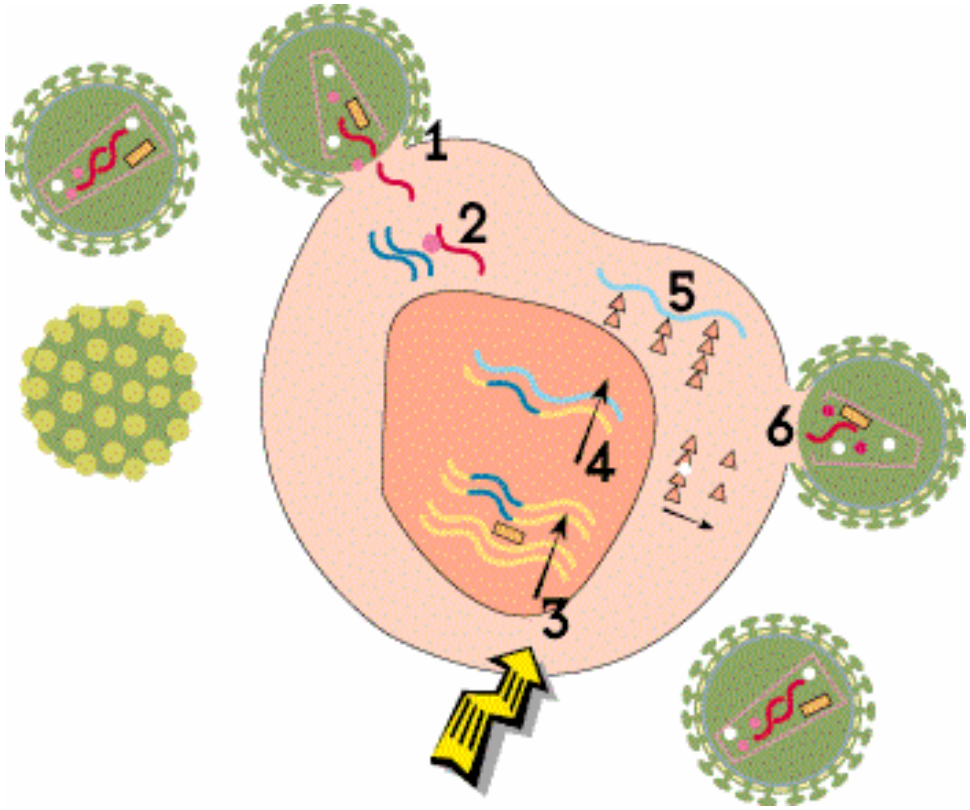


Step 2: Reverse Transcription

HIV's genes are carried in two strands of RNA, while the genetic material of human cells is found in DNA. In order for the virus to infect the cell, a process called "reverse transcription" makes a DNA copy of the virus's RNA.

After the binding process, the viral capsid (the inside of the virus which contains the RNA and important enzymes) is released into the host cell. A viral enzyme called reverse transcriptase makes a DNA copy of the RNA. This new DNA is called "proviral DNA."

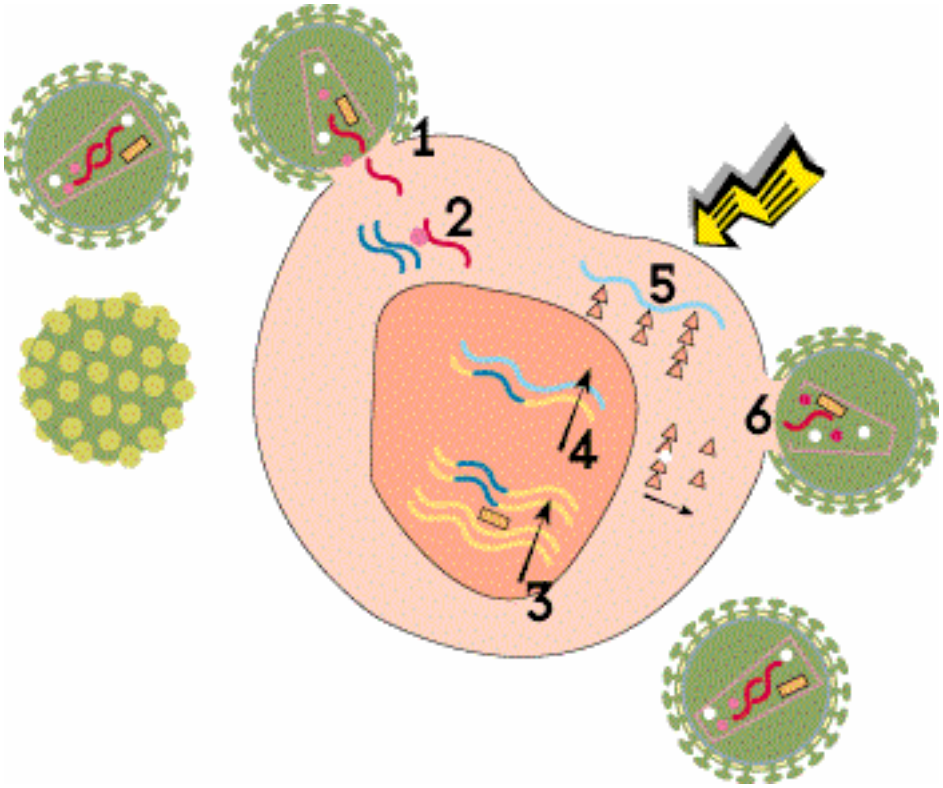
Reverse transcription can be blocked by: Nucleoside Reverse Transcriptase Inhibitors (NRTIs), Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs), and Nucleotide Reverse Transcriptase Inhibitors.



Step 3: Integration

The HIV DNA is then carried to the cell's nucleus (center), where the cell's DNA is kept. Then, another viral enzyme called integrase hides the proviral DNA into the cell's DNA. Then, when the cell tries to make new proteins, it can accidentally make new HIVs.

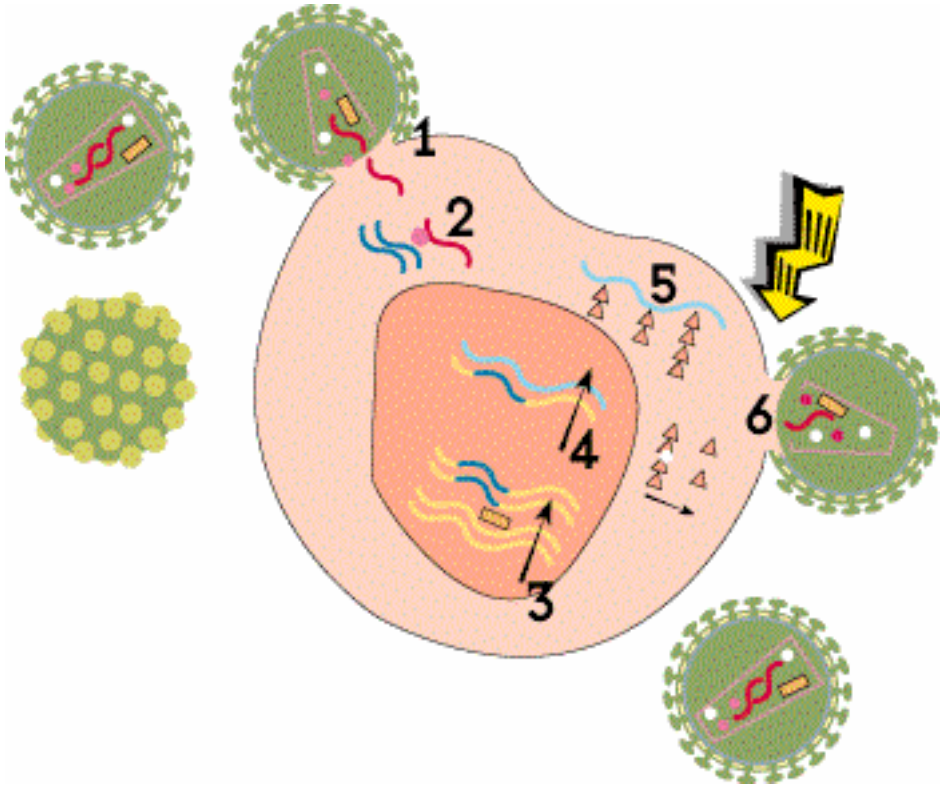
Integration can be blocked by integrase inhibitors, a new class of drugs that are in the earliest stage of research.



Step 5: Translation

The mRNA carries instructions for making new viral proteins from the nucleus to a kind of workshop in the cell. Each section of the mRNA corresponds to a protein building block for making a part of HIV.

As each mRNA strand is processed, a corresponding string of proteins is made. This process continues until the mRNA strand has been transformed or "translated" into new viral proteins needed to make a new virus.



Step 6: Viral Assembly

Finally, a new virus is assembled. Long strings of proteins are cut up by a viral enzyme called protease into smaller proteins. These proteins serve a variety of functions; some become structural elements of new HIV, while others become enzymes, such as reverse transcriptase.

Once the new viral particles are assembled, they bud off the host cell, and create a new virus. This virus is then able to infect new cells. Each infected cell can produce a lot of new viruses.

Viral assembly can be blocked by Protease Inhibitors (PIs).

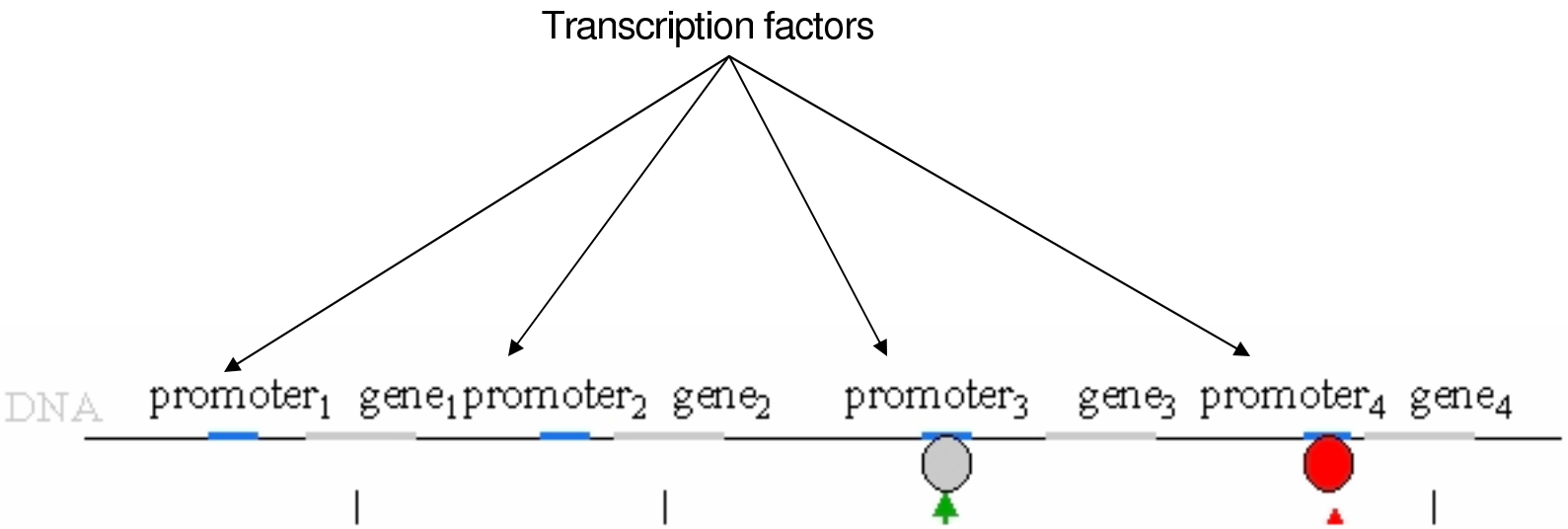
New strategies to target host cell proteins

· Viral proteins interact with host proteins to modify the transcription program of the cell; process of switching genes on and off.

· Identify these potential new target sites for future therapeutics.

·Two types of whole genome screens can now be used (genechips and RNAi) to discover these potential target sites.

DNA Transcription

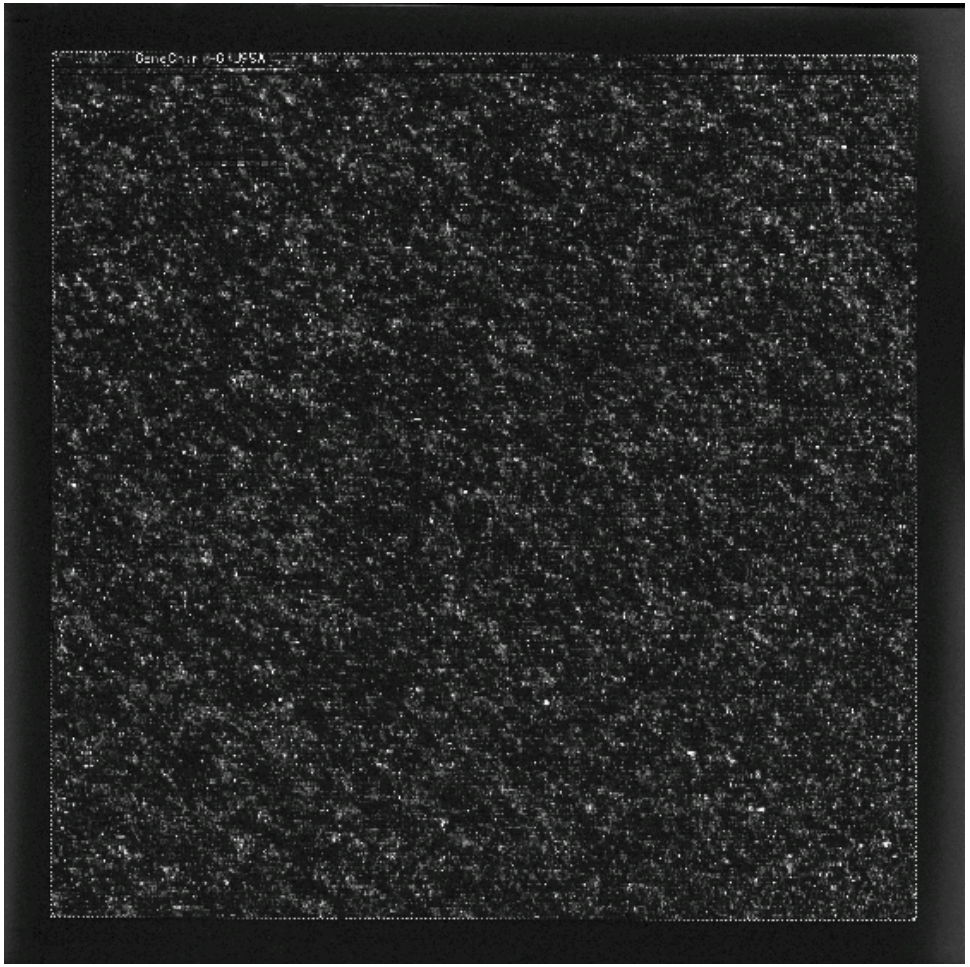


mRNA levels ↑ ↓

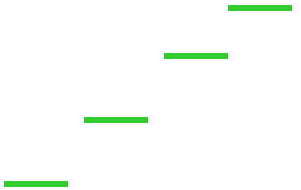
Gene Expression Microarray

- A 1x1 cm chip that contains probes used to detect the presence of genes
- Determine if genes are upregulated or downregulated post-HIV infection
- Compares treatment samples with controls

The Affymetrix Oligonucleotide Chip.



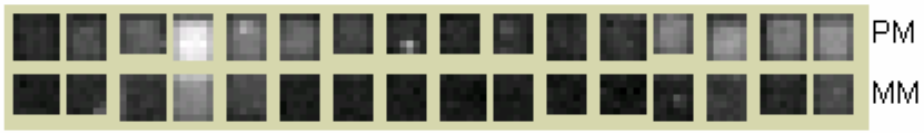
Segment of a piece of DNA



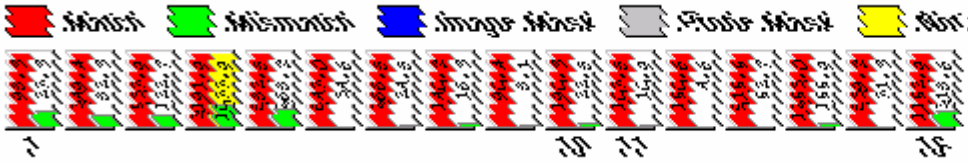
Oligonucleotides (25mer)

- 12,625 genes per chip.
- 20 oligonucleotides per gene.
- 64 pixels per probe cell (3µm/pixel)
- The chip contains numerous spiked controls, housekeeping genes, background cells, noise determination and normalization methods.

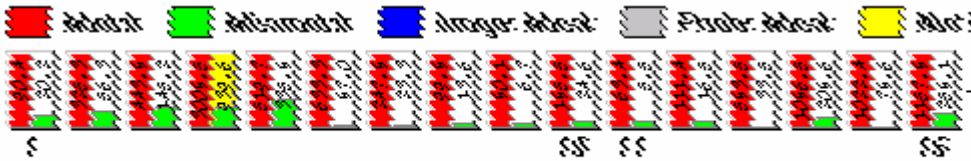
Data from a probe set on the chip.



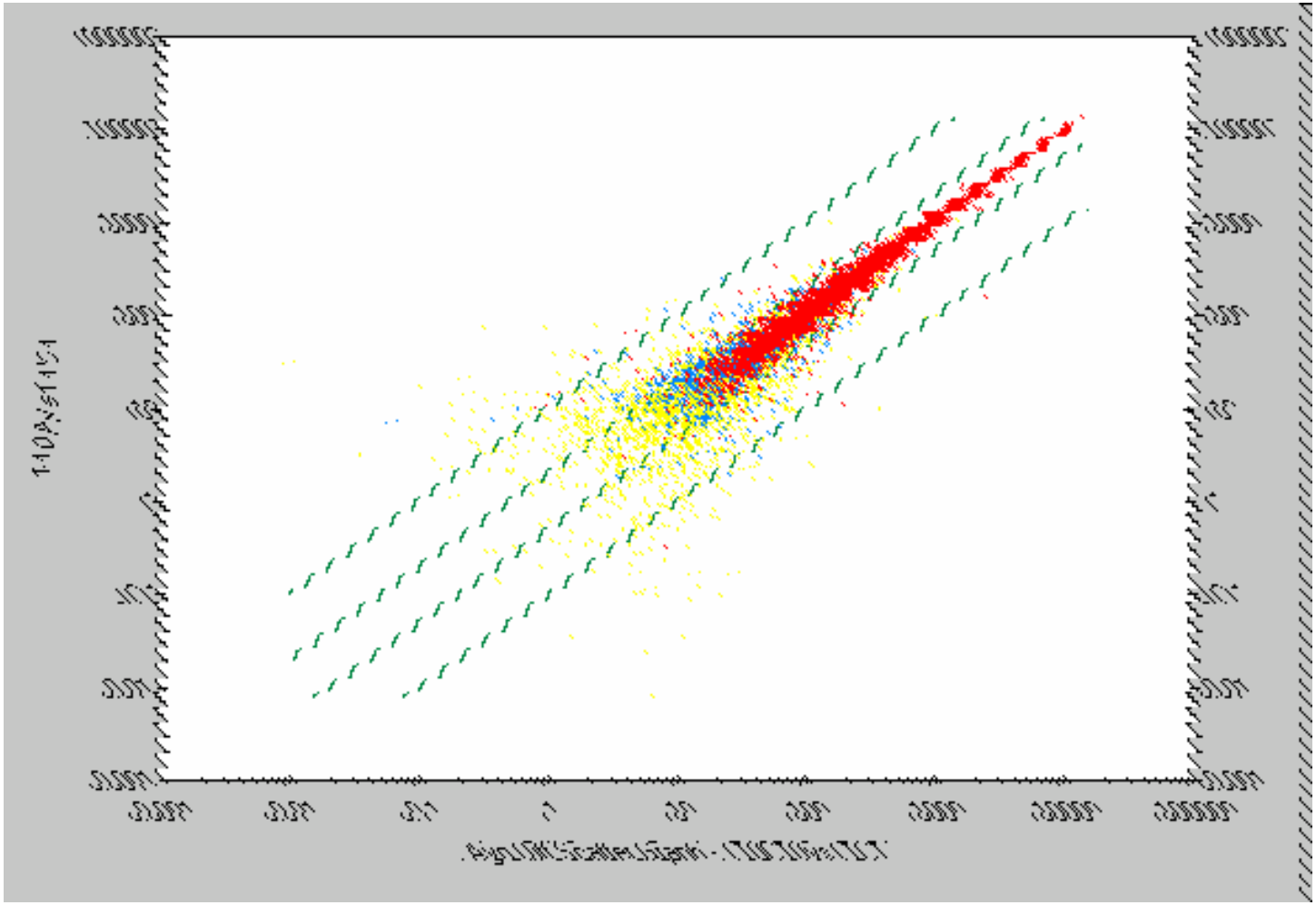
36647_at (091801MT2ctl24h001)
Intensity: 32 - 2730



36647_at (091801MT2BR9201924h001)
Intensity: 39 - 2278



Comparing data from two experiments.



HIV Gene Expression Microarray

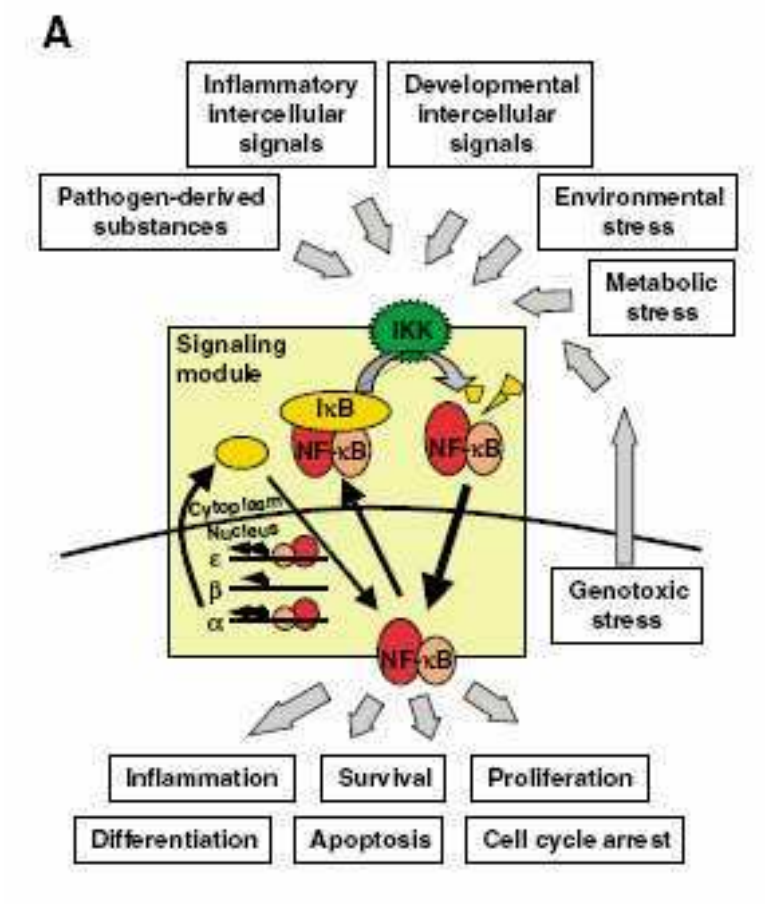
- Current study: 12,558 probes that correspond to 7,531 genes
- Obtained microarray gene expression profiles at 24 h, 48 h, and 7 d HIV post-infection.
- Virus treated samples compared with control at each time point and then performed ANOVA across the three time points.

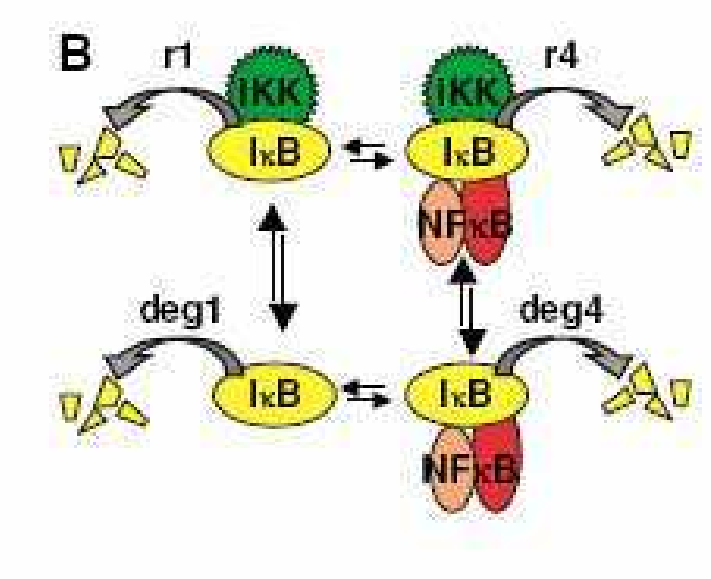
Components of NF-Kappa B pathway affected by HIV treatment

Gene ID	mean24n	SEM	mean48n	SEM	mean7dn	SEM
100_g_at	-8.625	18.3852	35.525	18.74546	1.833333	6.653654
1000_at	12.5	12.55833	-5.175	12.55855	29.4	6.305817
1001_at	15.05	6.410473	8.95	5.95266	8.7	3.7072
1002_f_at	6.45	5.777326	-2.45	7.830017	12.3	3.164385
1003_s_at	20.95	8.569763	-12.3	36.50961	-16.66667	9.617057
1004_at	9.45	4.146987	-5.1	9.52348	3.733333	11.60062
1005_at	248.95	48.042	175.3	55.60298	145.0333	108.5049
1006_at	3.675	3.675907	-2.25	0.429146	4.3	4.895236
1007_s_at	15.35	8.942548	-21.175	7.36471	-12.83333	15.38845
1008_f_at	-52.475	20.48176	232.475	74.84287	106.5333	45.21085
1009_at	-196.35	52.52975	123	41.15777	-199.4667	46.04781
101_at	-7.5	6.49859	5.5	7.407429	-3.333333	17.67129
1010_at	8.125	14.29431	-30.125	7.11552	-21.66667	16.77342

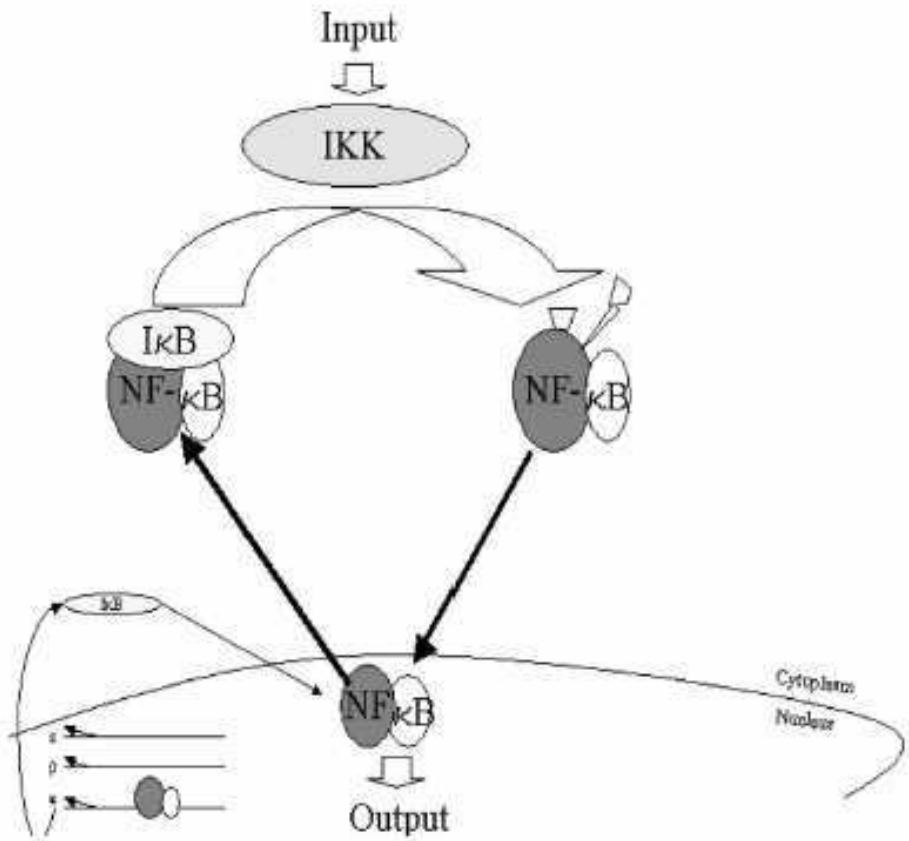
GeneInfo	Genebank	ANOVA p-val	24hr	48hr	7day
IkB kinase alpha subunit (IKK alpha) mRNA	AF009225	0.006	UP	DOWN	UP
Human NF-kappa-B transcription factor p65 subunit	L19067	0.002	UP	DOWN	DOWN
p65 protein	AJ002425	0.024	UP	DOWN	UP
IkB kinase gamma subunit (IKK-gamma)	AF074382	0.020	UP	DOWN	UP
HIV-1 transcriptional elongation factor TAT cofactor TAT-SF1	Z97632	0.003	UP	DOWN	UP

NF-Kappa B pathway

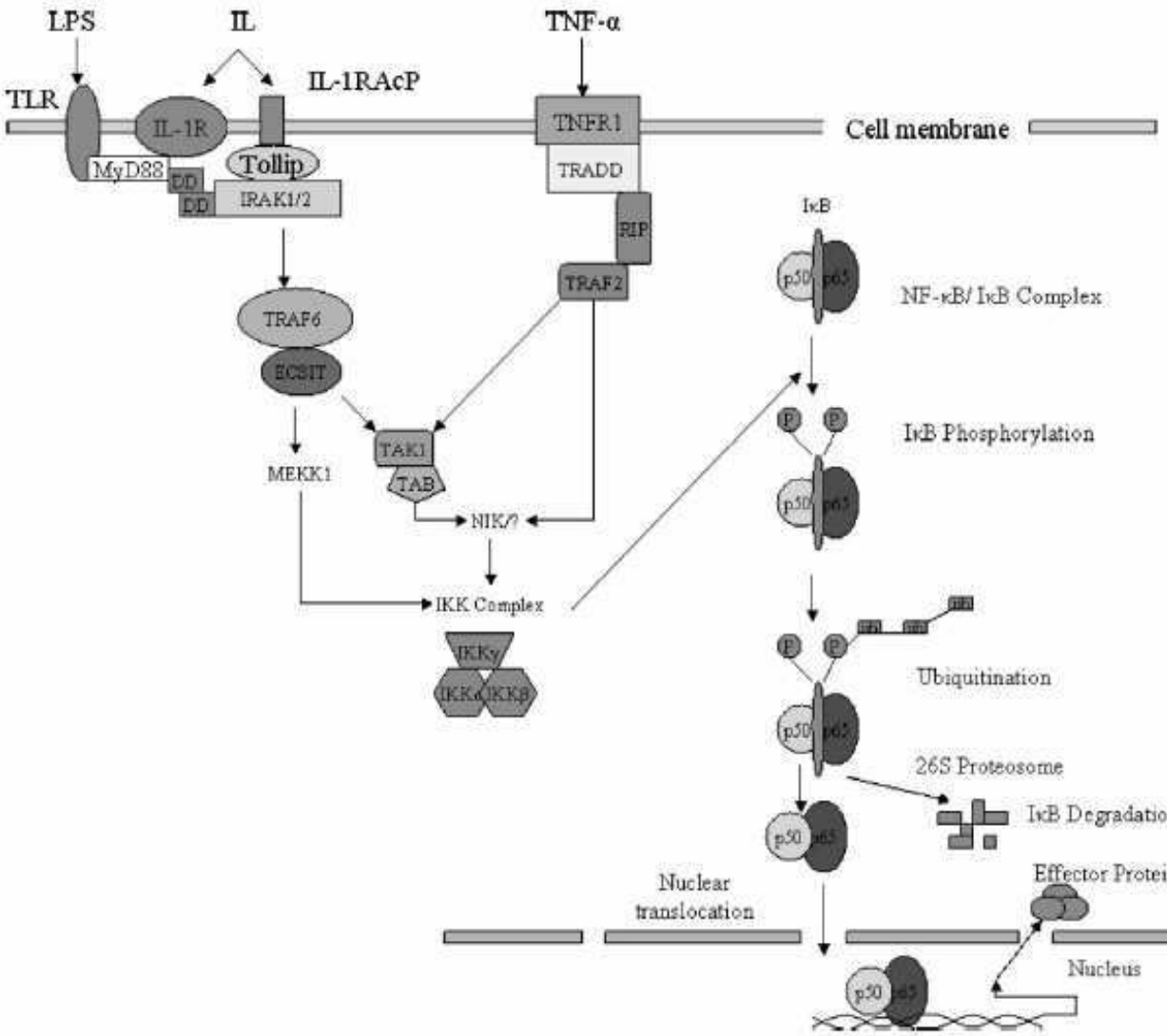




(O'Dea et al 2007, Molecular Systems Biology, 3, pp1)

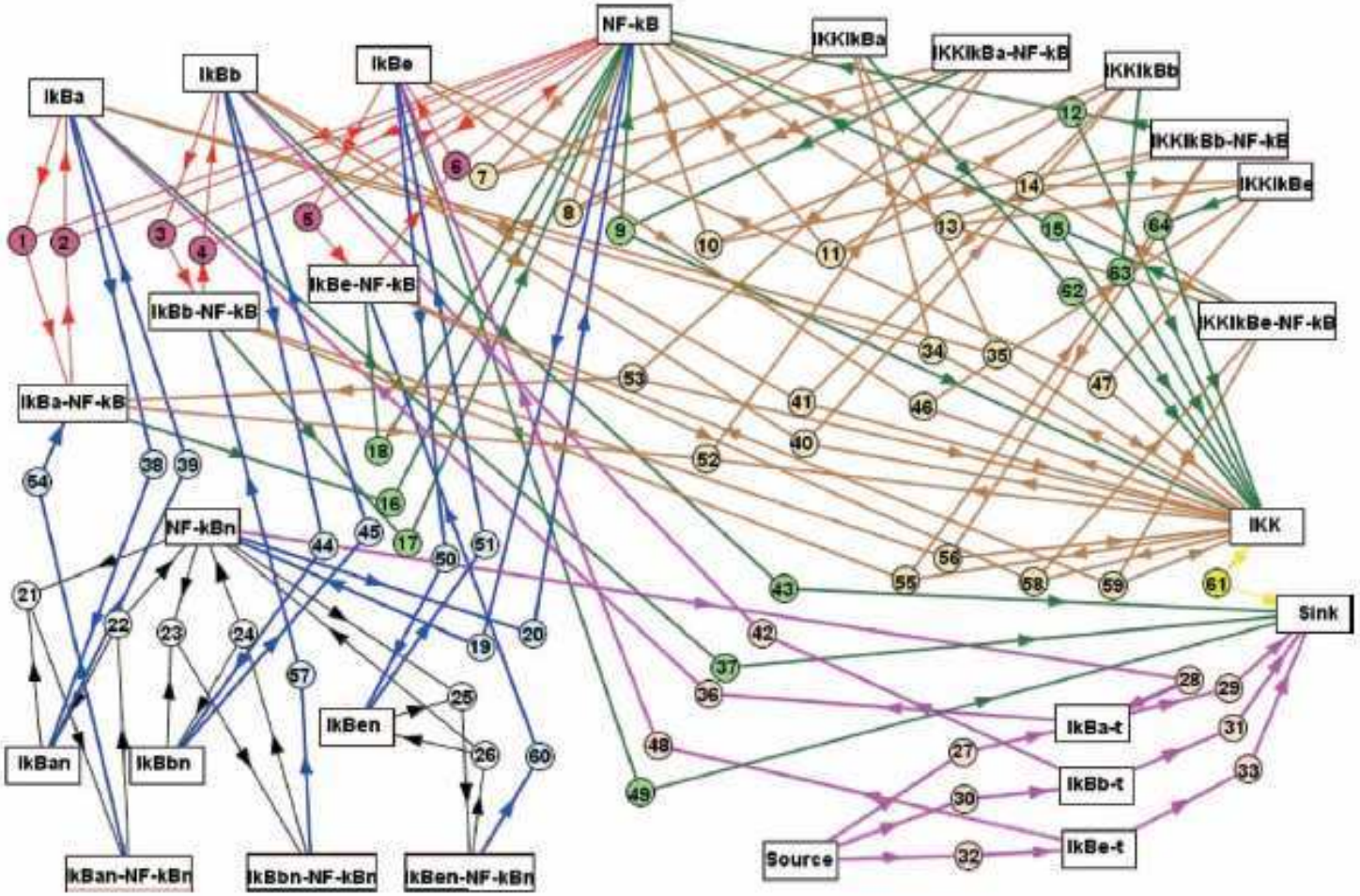


Basic IκB-NF-κB signalling model. NF-κB is held inactive in the cytoplasm of non-stimulated cell by three IκB isoforms. During cell stimulation, IKK complex is activated, leading to phosphorylation and ubiquitination of the IκB proteins. Free NF-κB translocates to the nucleus, activating genes including IκBα. IκBβ& -ε are synthesised at steady rate, allowing for complex temporal control of NF-κB activation involving negative feedback



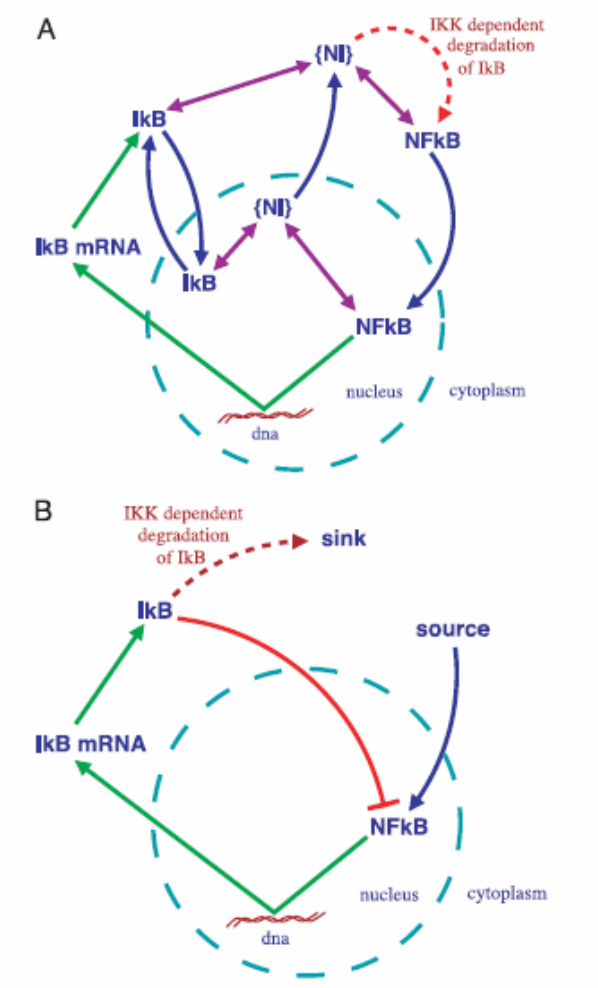
A schematic representation of signalling cascades for LPS, IL and TNF-α stimulation and activation of NF-κB (p50/p65)

Ihekwaba et al 2004, Sys. Biol. Vol 1, pp93)



Connection of the reactions of the NF-κB model analysed in the present work. Red arrows and violet red circles = IκB-NF-κB cytoplasmic reactions; blue arrows and circles = nuclear transport; magenta arrows and pink circles = IκB mRNA synthesis (including transcription, translation and degradation); black arrows and white circles = IκB-NF-κB nuclear reactions; light green arrows and circles = IκB phosphorylation and degradation reactions; brown arrows and brown circles = Bimolecular IKK-IκB and tri-molecular IKK-IκB-NF-κB; yellow arrows and circles = IKK slow adaptation coefficient

Minimal NF-Kappa B model



Three-Variable Model of NF-κB Oscillations. The core feedback loop, we find, consists of only three constituents (Fig. 1*B*): nuclear NF-κB (N_n), cytoplasmic IκB (I), and IκB mRNA (I_m). NF-κB dimers activate production of IκB mRNA, which translated to IκB inhibits nuclear NF-κB production, completing the feedback loop.

The dynamics of the system in Fig. 1*B* is captured by three coupled ordinary differential equations

$$\frac{dN_n}{dt} = A\frac{(1-N_n)}{\varepsilon + I} - B\frac{IN_n}{\delta + N_n}, \qquad [1a]$$

$$\frac{dI_m}{dt} = N_n^2 - I_m, \qquad [1b]$$

$$\frac{dI}{dt} = I_m - C\frac{(1-N_n)I}{\varepsilon + I}. \qquad [1c]$$

Fig. 1. Schematic diagram of key interactions in the NF-κB signaling system. Green arrows indicate transcription and translation; blue arrows indicate transport in and out of the nucleus; purple arrows indicate complex formation; and {NI} denotes the NF-κB–IκB complex. The red barred arrow in *B* indicates the effective inhibition of nuclear NF-κB by IκB. (*A*) Seven-variable model. The variables are the concentrations of NF-κB inside and outside the nucleus, IκB inside and outside, the complex {NI} inside and outside, and finally the concentration of the mRNA of IκB. (*B*) Three-variable model. The variables are the concentrations of nuclear NF-κB, the mRNA of IκB, and cytoplasmic IκB.

(Krishna et al, PNAS, 103, pp 10840

Proposal Objectives.

- Write and run the published models for NF-KB activation and compare with the range of dynamics already published for this pathway.

- Gene expression changes suggest that HIV infection affects at least two components of this pathway (IKKB and RelA/p65). Determine how the dynamics are affected by changing the levels of NF-KB and Rel A.

- Screen for low-dimensional manifolds in this system of ODEs.

Small Interfering RNA (siRNA) Knockdown Study

- Elledge et al. Science. 2008 Feb 15;319
- Study found 273 (237 new) HIV dependant factors (HDFs)
- Our study showed that 100 of the 212 genes on the Affymetrix gene chip represented were affected with viral treatment.
- We now know the activation status of these 100 primary targets.

1: Science. 2008 Feb 15;319(5865):921-6. Epub 2008 Jan 10.

Identification of host proteins required for HIV infection through a functional genomic screen.

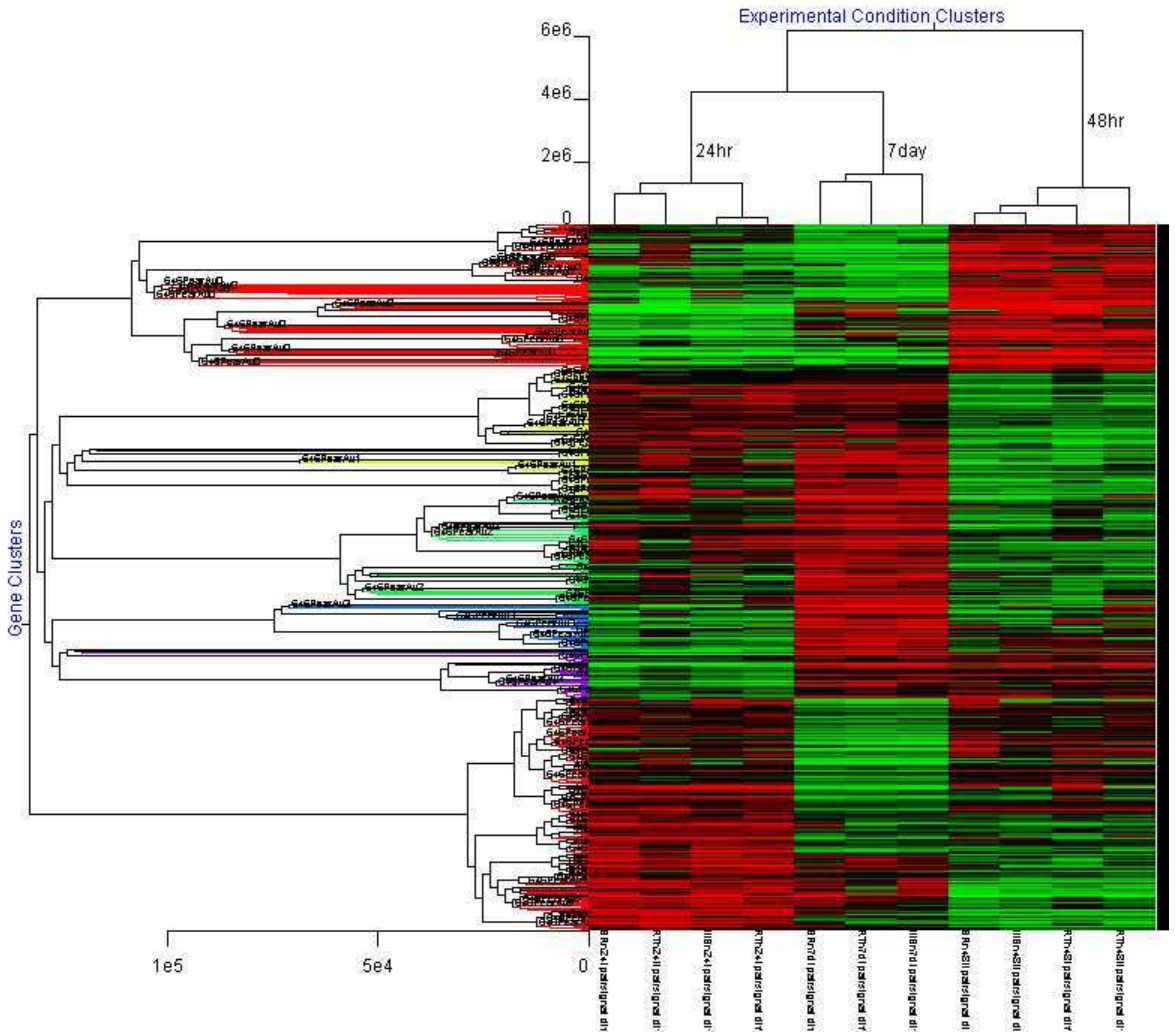
Brass AL, Dykxhoorn DM, Benita Y, Yan N, Engelman A, Xavier RJ, Lieberman J, Elledge SJ.

Department of Genetics, Center for Genetics and Genomics, Brigham and Women's Hospital, Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA.

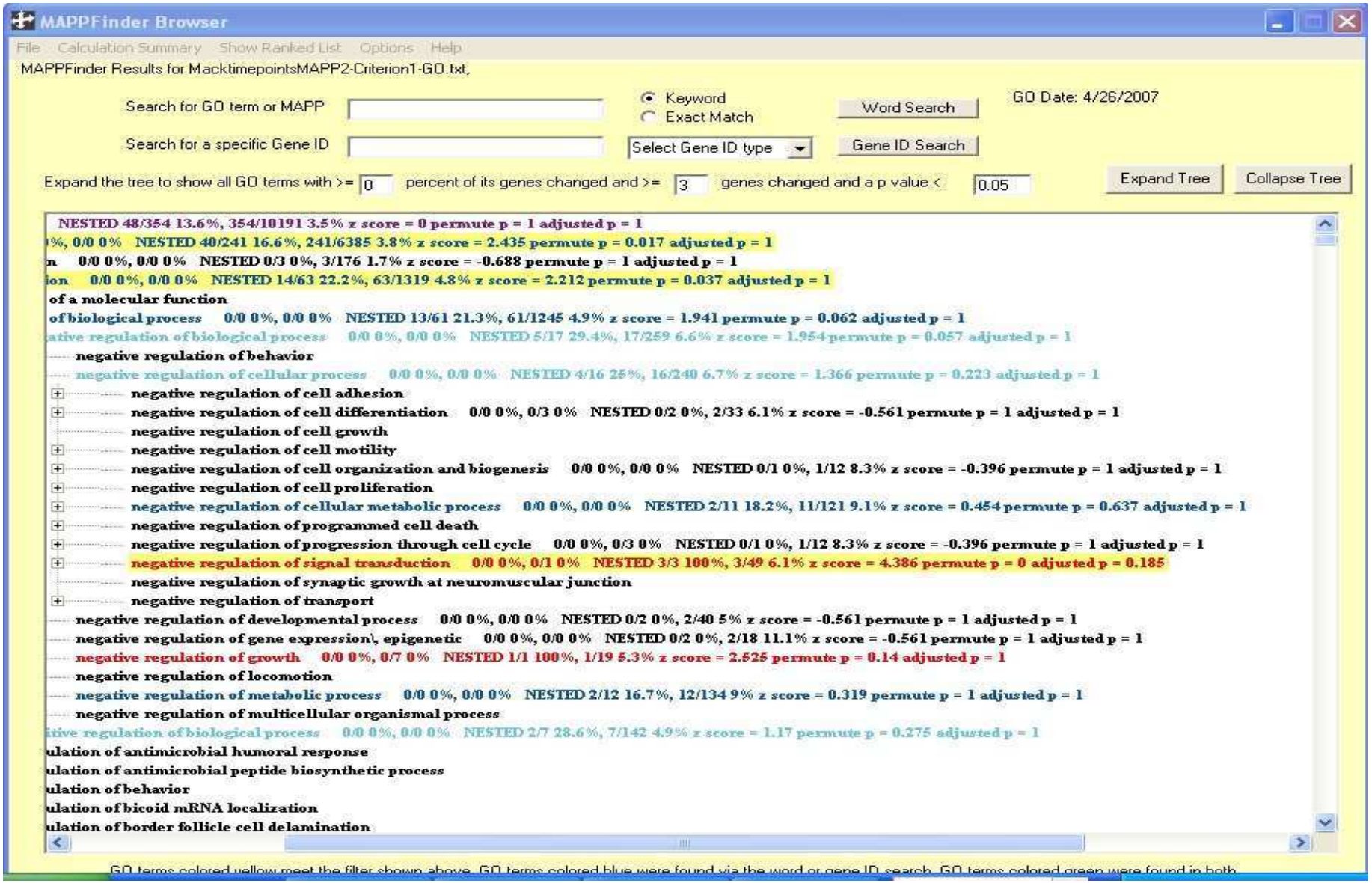
HIV-1 exploits multiple host proteins during infection. We performed a large-scale small interfering RNA screen to identify host factors required by HIV-1 and identified more than 250 HIV-dependency factors (HDFs). These proteins participate in a broad array of cellular functions and implicate new pathways in the viral life cycle. Further analysis revealed previously unknown roles for retrograde Golgi transport proteins (Rab6 and Vps53) in viral entry, a karyopherin (TNPO3) in viral integration, and the Mediator complex (Med28) in viral transcription. Transcriptional analysis revealed that HDF genes were enriched for high expression in immune cells, suggesting that viruses evolve in host cells that optimally perform the functions required for their life cycle. This effort illustrates the power with which RNA interference and forward genetics can be used to expose the dependencies of human pathogens such as HIV, and in so doing identify potential targets for therapy.

Genes that change across time with virus treatment

At each time point the virus effect was assessed using two measures of signal log ratios(SLR) and absolute difference between the control (no virus) and virus treated samples (signal difference values were used). An analysis of variance was performed to screen for genes that were most affected across 24hr, 48hr and 7 days. Genes that were affected by a p-value of less than 0.05 for both changes in signal log ratios and absolute differences were analyzed using a two way clustering algorithm (639 genes were affected). The figure shows the dendrograms that group both the time of treatment and the genes using average linkage and Pearson Correlation as the distance measure. The clustering is represented graphically as a colored image in which red represents up-regulation and green represents down-regulation of genes. The values were obtained from absolute difference between control and virus treated. Clustering of the genes and the time points shows rows of gene with similar expression patterns. The bottom left corner represents the genes that are switched on at 24 hours and the top right corner shows the genes that are switched on at 7 days.



Gene Ontology Classifications.

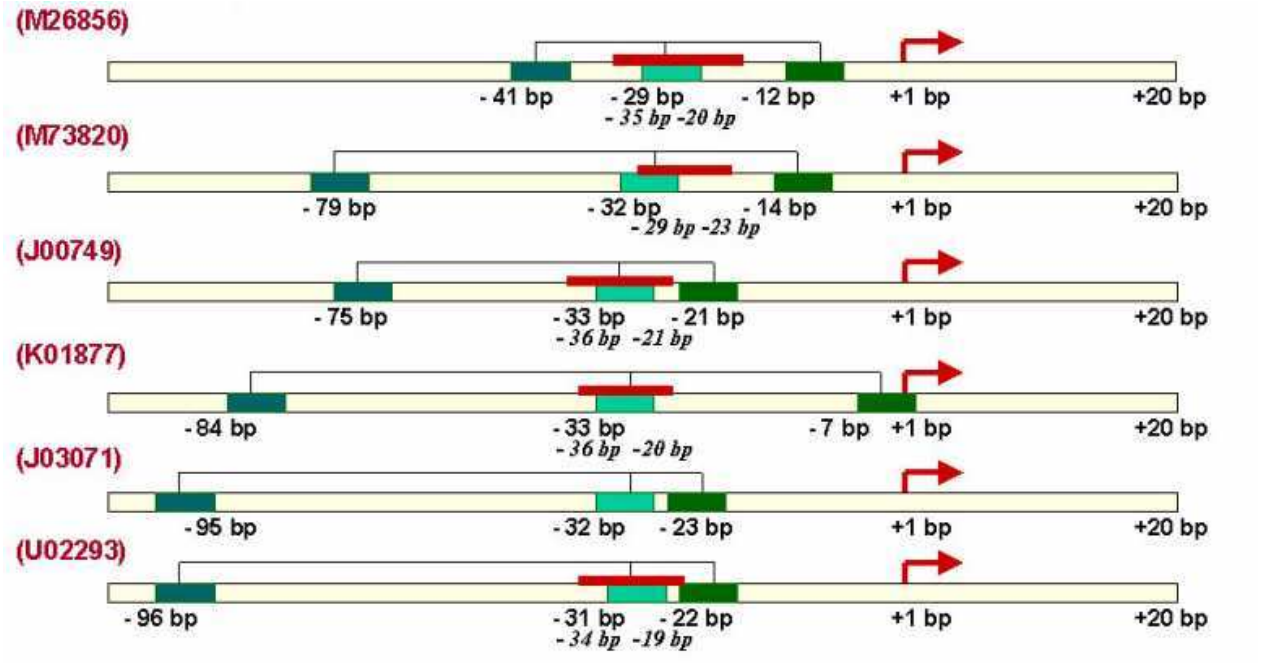


GOID	GO Name	GO Type	Number Changed	Number Measured	Number in GO	Percent Changed	Percent Present	Z Score	Permute P
16408	C-acyltransferase activity	F	8	8	11	100	72.72727	3.123	0
226	microtubule cytoskeleton organization and biogenesis	P	24	34	52	70.58823	65.38461	2.995	0.002
5667	transcription factor complex	C	41	66	96	62.12121	68.75	2.794	0.004
31109	microtubule polymerization or depolymerization	P	6	6	9	100	66.66666	2.704	0.004
7019	microtubule depolymerization	P	6	6	8	100	75	2.704	0.004
5159	insulin-like growth factor receptor binding	F	6	6	7	100	85.71429	2.704	0.006
51051	negative regulation of transport	P	10	12	16	83.33334	75	2.665	0.007
45185	maintenance of protein localization	P	10	12	14	83.33334	85.71429	2.665	0.007

Gene Ontology terms containing 5-100 genes and Permuted P < 0.05 for overall change in expression value.

Genes affected by HIV treatment were organized using the ontology terms. Significance of the distribution of these genes across all the ontology terms was tested using a Chi-square test with permuted p-values. Transcription factor complex showed 41 genes out of 66 measured resulted in viral affected changes in gene expression. Gene ontologies can be used to describe in molecular detail how the host cell responds to HIV infection.

Promoter Analysis.



The promoter sequences are aligned relative to the transcription start (position +1 bp), indicated by arrows. The EMBL identifiers of promoters studied are given in parentheses on the left. The eight-bp oligonucleotide motifs composing the complex signal are shown as shaded green rectangles; positions of the first nucleotides are indicated relative to the transcription start. Red rectangles mark positions of the TATA-boxes, indicated in the TRRD database; positions of its first and last nucleotides are italicised. Interestingly, only a single oligonucleotide in the complex signal corresponds to the real annotated site, whereas the others could correspond to potential transcription factor binding sites or to the double-stranded DNA regions with specific physicochemical properties.(Vityaev et al, In Silico Biology 2, 0024 (2002))

Genebank	ANOVA p-val	Transcription factor mRNA	24hrprof	48hrprof	7dayprof
AC004084	0.0409		2	2	2
AJ011915	0.0078		2	2	2
D13969	0.0489	polycomb group ring finger 2	2	2	2
AF052105	0.0048		2	2	2
X86809	0.0263		2	2	2
M36035	0.0390		2	2	2
N58130	0.0451		2	2	2
AB011177	0.0094		2	2	2
AL046961	0.0430		2	2	2
AI338355	0.0271		2	2	2
U08191	0.0349	Nuclear factor related to kappaB binding protein	2	2	2
D88208	0.0024		2	2	2
J02973	0.0303		2	2	2
AF072242	0.0036		2	2	2
AA142942	0.00381615		2	1	2
X97548	0.005787532	tripartite motif-containing 28	2	1	2
Z46629	0.03709349	SRY (sex determining region Y)-box 9 (campomelic dysplasia, autosomal sex-reversal)	2	1	2
D15050	0.005192797	SNF1-like kinase	2	1	2
D30756	0.007516497		2	1	2
W29105	0.018412744		2	1	2
M63962	0.001953308		2	1	2
X90761	0.026201326		2	1	2
AF055634	0.008455801		2	1	2
AB017788	0.045981557		2	1	2
AJ001454	2.55799E-05		2	1	2
U52111	0.017352296		2	1	2
AI435898	0.016716926	zinc finger and BTB domain containing 25	2	1	2
L29385	0.025570548		2	1	2
AF030335	0.003133566		2	1	2
Y14768	0.013998956		2	1	2
D87011	0.019885187		2	1	2
AA812696	0.03598914		2	1	2
W26627	0.044268715		2	1	2

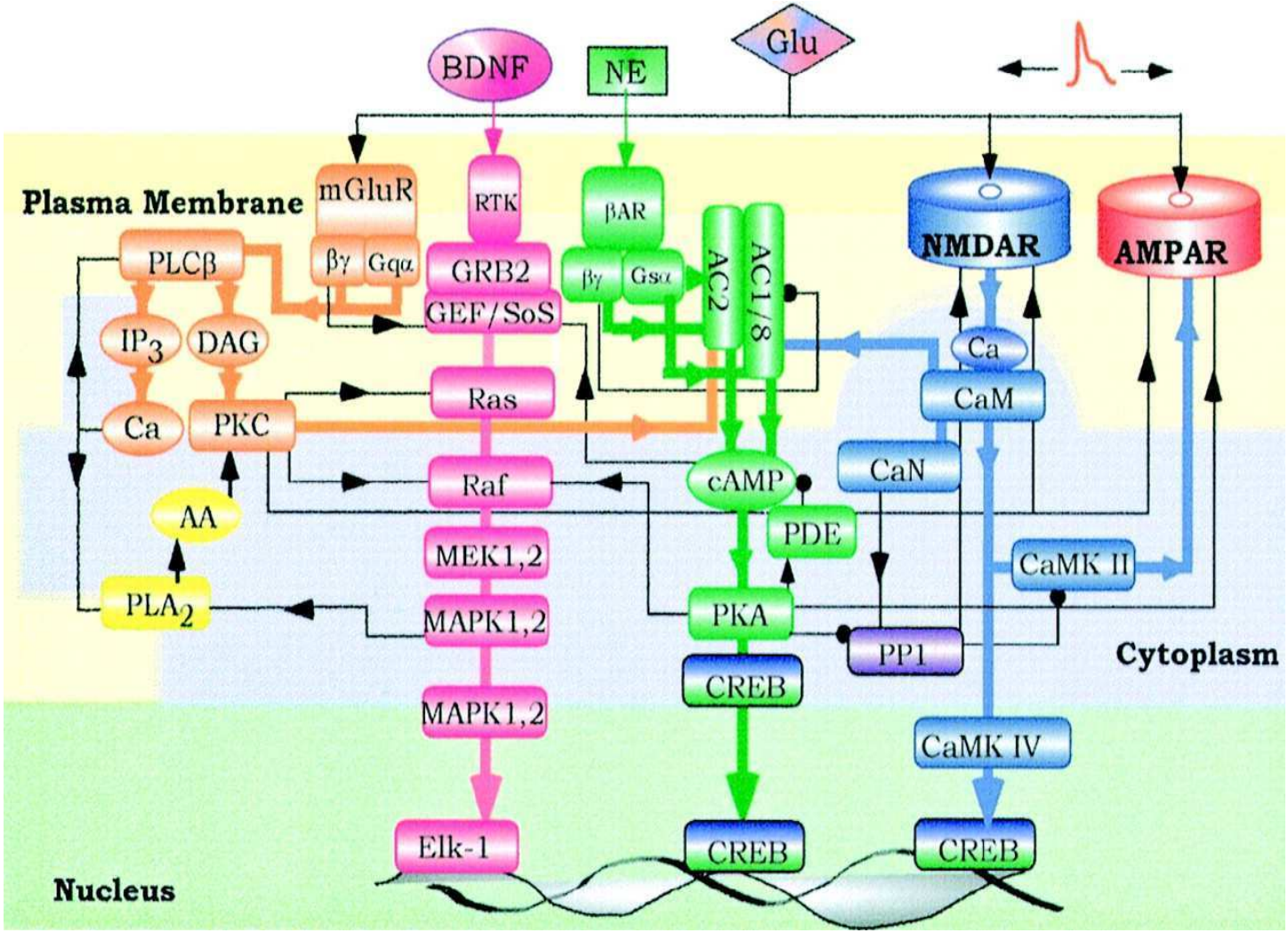
Discovery of potential new target sites.

- Use discovery-driven approaches for potential new pathways leading to gene activation through novel transcription factors.

- Write computational models for these pathways to extends other inputs into the NF-Kappa B pathways.

- Perform sensitivity analysis to screen for potential new drug target sites in these pathways for gene activation.

Control of transcription.



(Weng et al., 1999, Science, pp 92)

Reverse Engineering