High Throughput Integration Site Detection

Frederic Bushman

University of Pennsylvania School of Medicine

Possible questions to be addressed with integration site analysis

What is the distribution of integration target sites relative to genomic land marks?

Has an adverse event taken place?

What molecular events were associated with an adverse event?

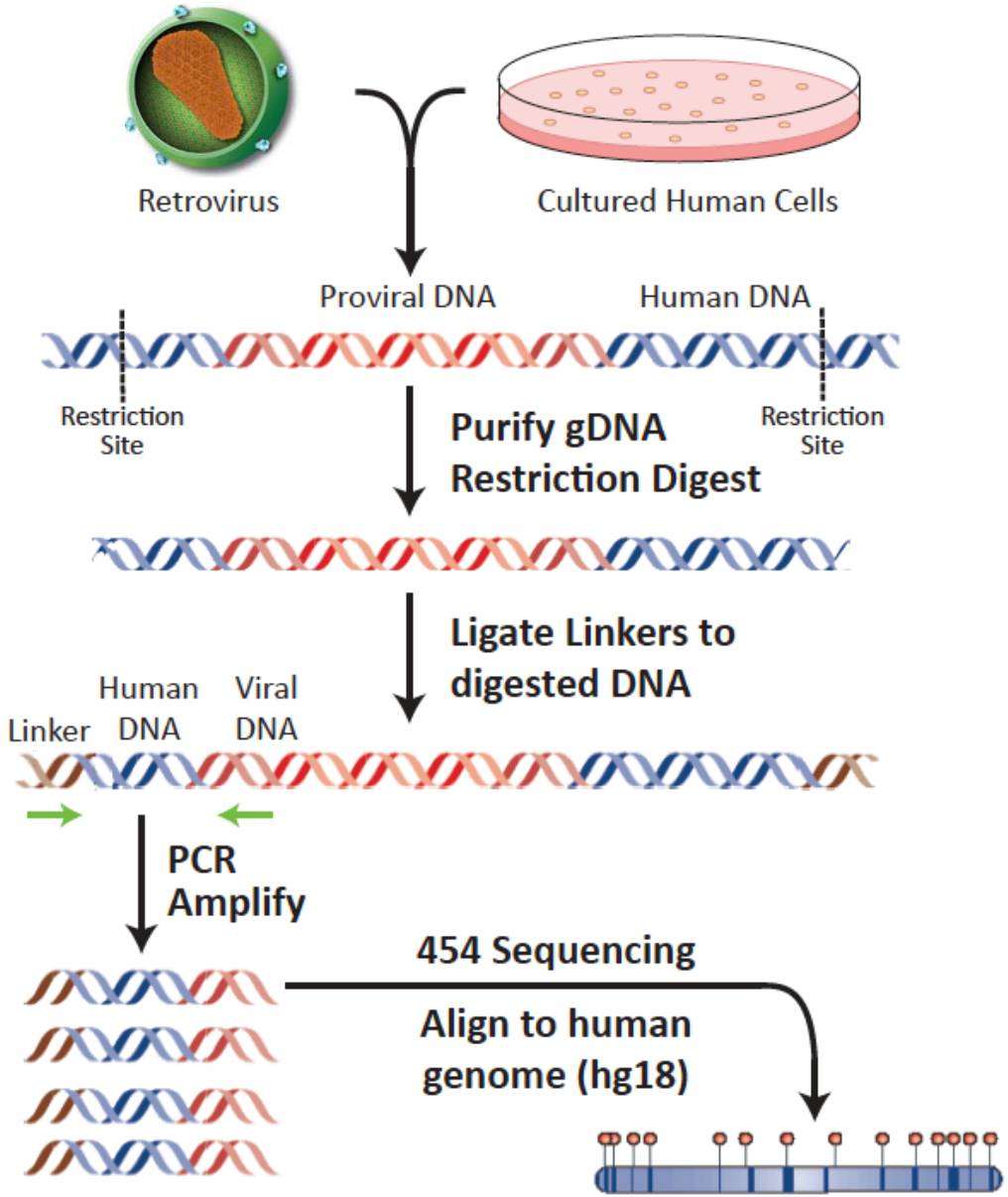
How many transduced cells are present in a subject?

What is the clonal structure of a transduced cell population?

Challenges in tracking integration sites during gene therapy

- Widely used isolation methods are severely biased
- We have no method for fully sequencing any integration site population
- Some literature wildly overstates completeness and certainty, leading to unrealistic expectations
- Statistical methods for quantifying uncertainty at an early stage
- PCR cross over

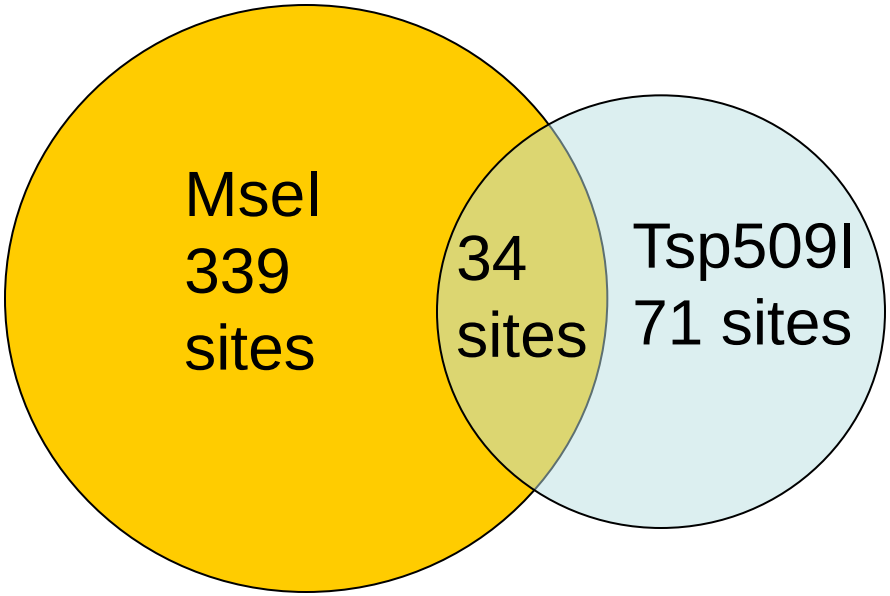
Technology for analyzing integration sites

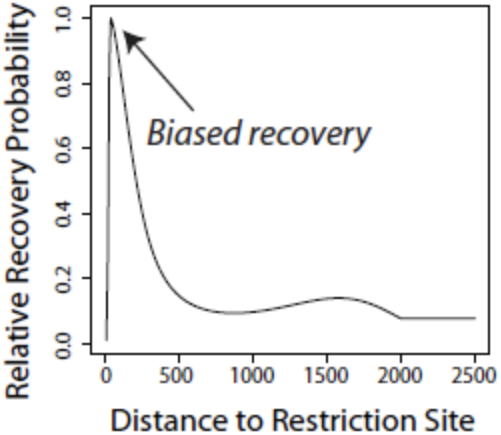


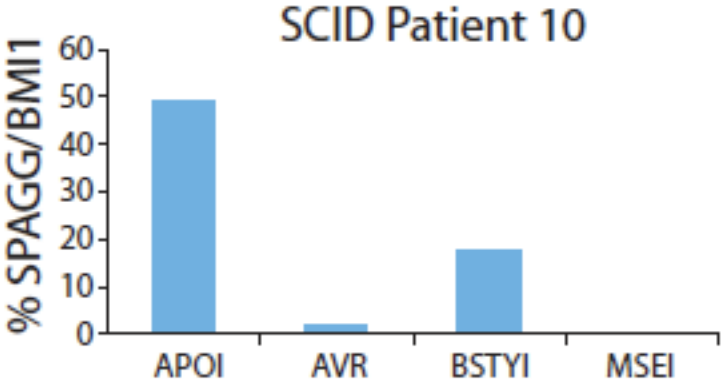
Sequencing:
454/Roch
Illumina
…

Use of restriction enzymes to cleave genomic DNA leads to severe biases

G. Wang, A. Garrigue, A. Ciuffi, K. Ronen, J. Leipzig, C. C. Berry, C. Legrasle-Peyrou, F. Benjelloun, S. Hacein-Bey-Abina, A. Fischer, M. Cavazzana-Calvo, and F. D. Bushman, 2008



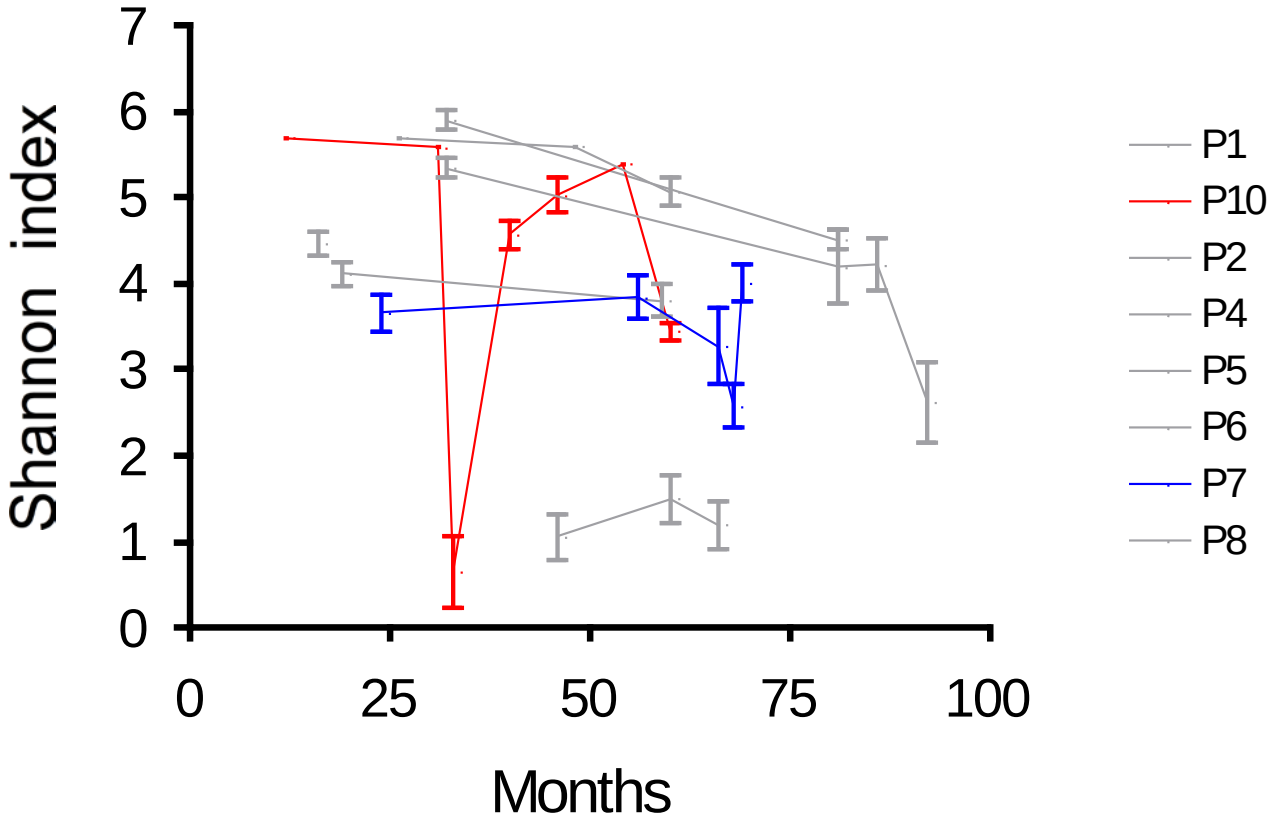




Brady et al, submitted

Longitudinal analysis of 8 SCID-X1 subjects

Longitudinal analysis of 8 subjects
>200,000 sequence reads
Up to six restriction enzymes used per sample



Wang, Cavazzana-Calvo, Fisher, Hacein-Bey-Abina, Bushman et al., 2009, 2010, 2010

Methods for recovering integration sites in proportion to their abundance

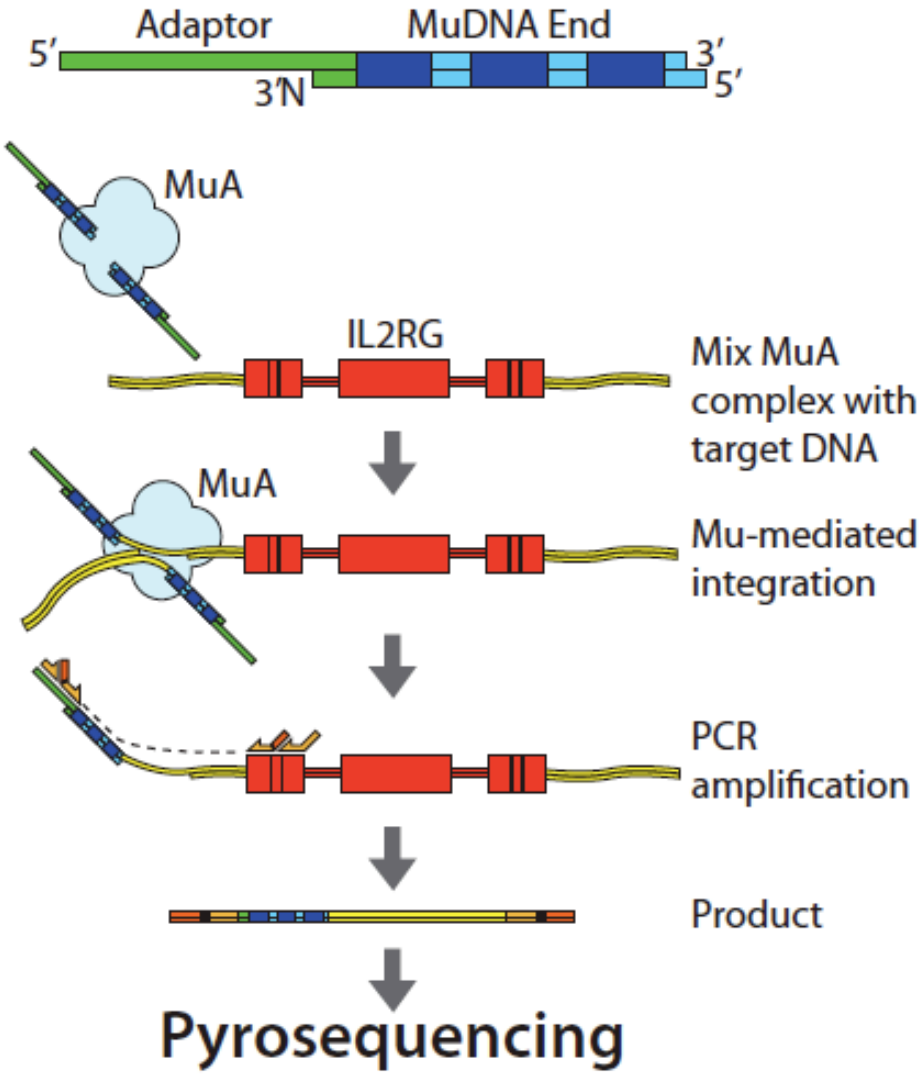
Name	Method	Sequencing method
nrPCR *Gabriel et al., 2010*	Limited polymerization, RNA ligase to attach linkers	454/Roche
Mu-mediated capture *Brady et al., submitted*	Mu transposition in vitro to install linkers	454/Roche
DNA shearing, ligation *Williams-Carrier et al., 2010* *Gillet et al., submitted*	Shear DNA, blunt end ligation to attach linkers	Solexa/Illumina

Considerations: lack of bias, efficiency, amount of starting DNA needed, ease of use, throughput

Isolation of newly integrated DNA using Mu transposition in vitro

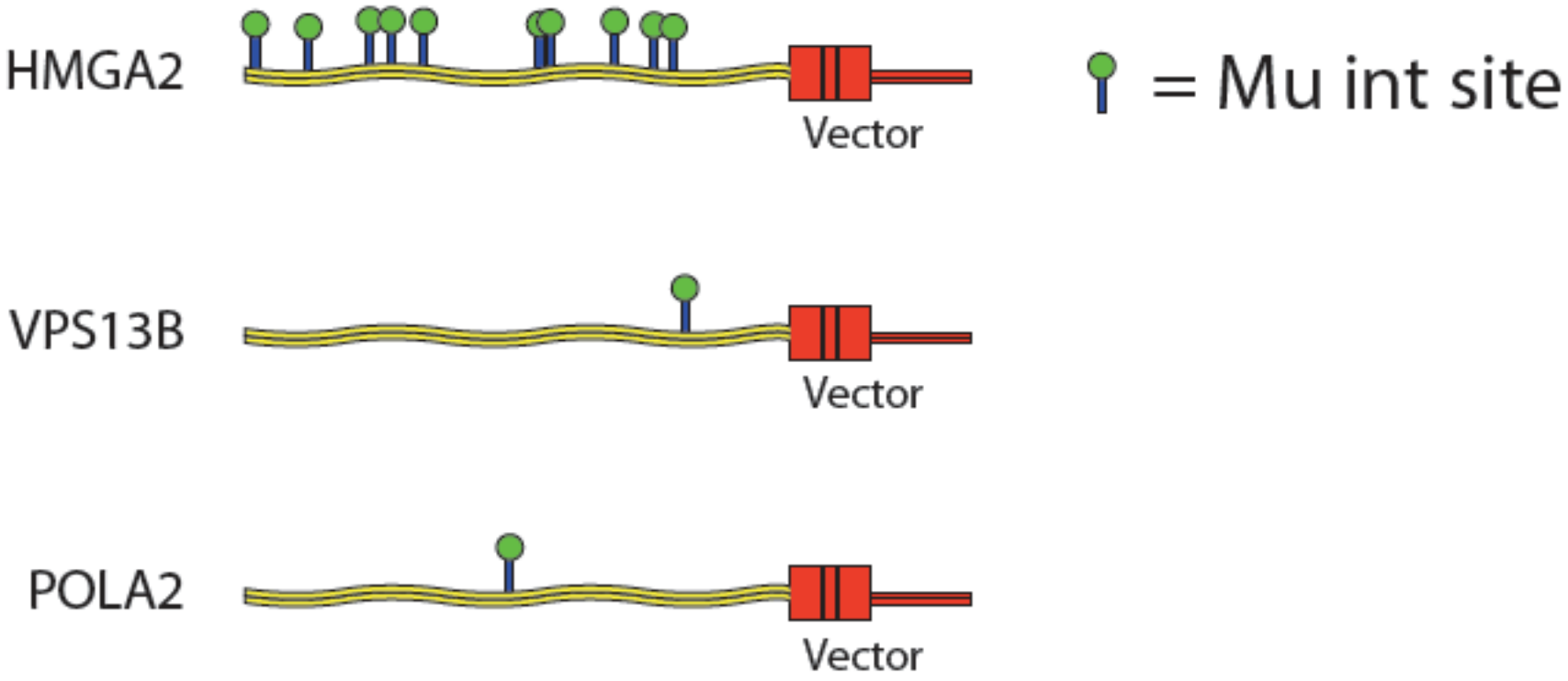
Troy Brady, Nirav Malani, Gary P. Wang, Charles C. Berry, Philippe Leboulch, Salima Hacein-Bey-Abina, Marina Cavazzana-Calvo, Harri Savilahti and Frederic D. Bushman

Integration site capture using DNA transposition in vitro



Much less recovery bias than with restriction enzymes

Abundance estimation in the beta thalassemia trial by counting Mu hops



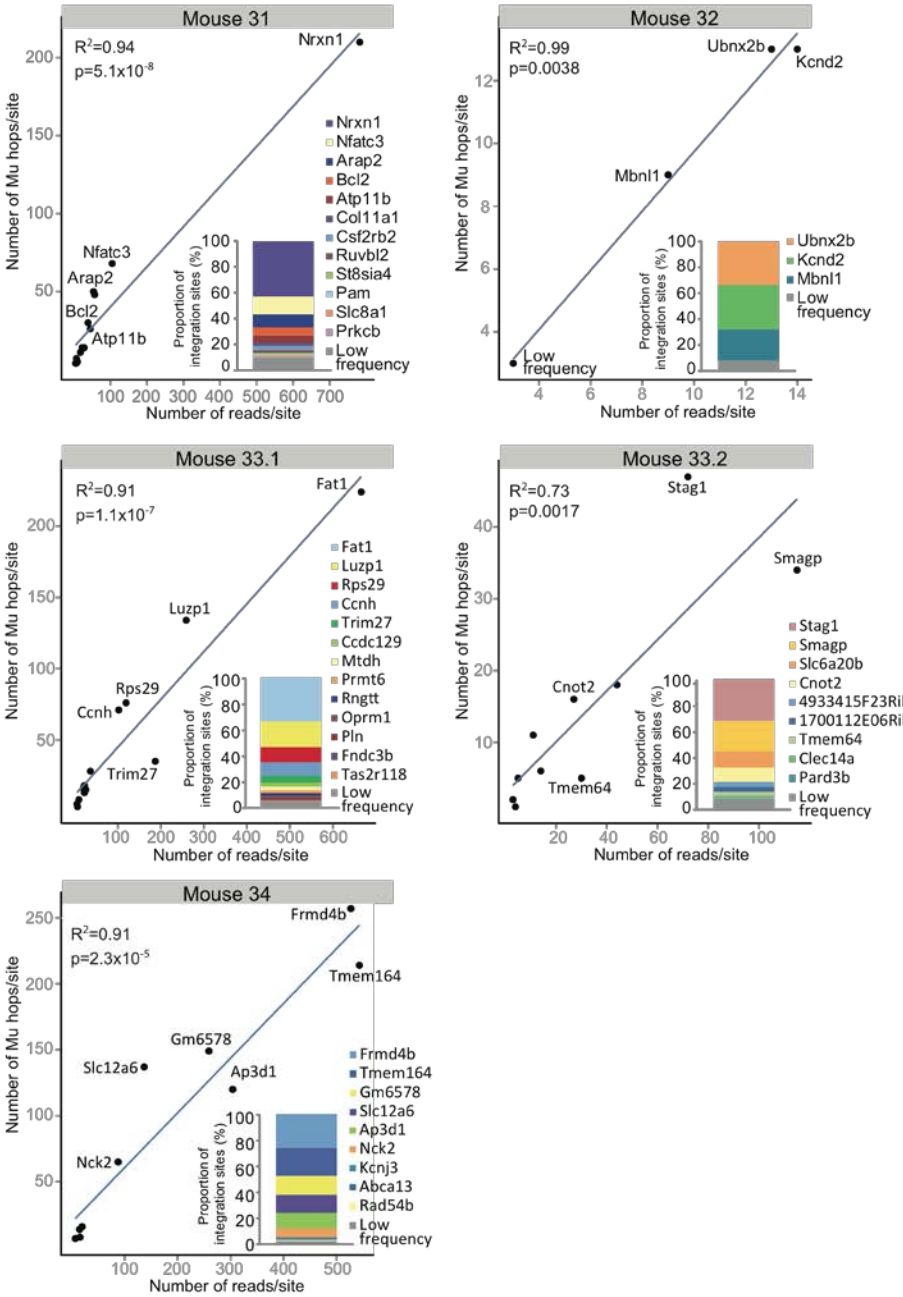
The HMGA2 site is the *only* site recovered at elevated frequency

Correlating numbers of sequence reads and numbers of Mu hops per site

Data from murine model of beta-thal gene correction

Good agreement on rank order of abundance

Ronen, Negre, Roth, Leboulch, Payen, Bushman, submitted



Summary of Mu study

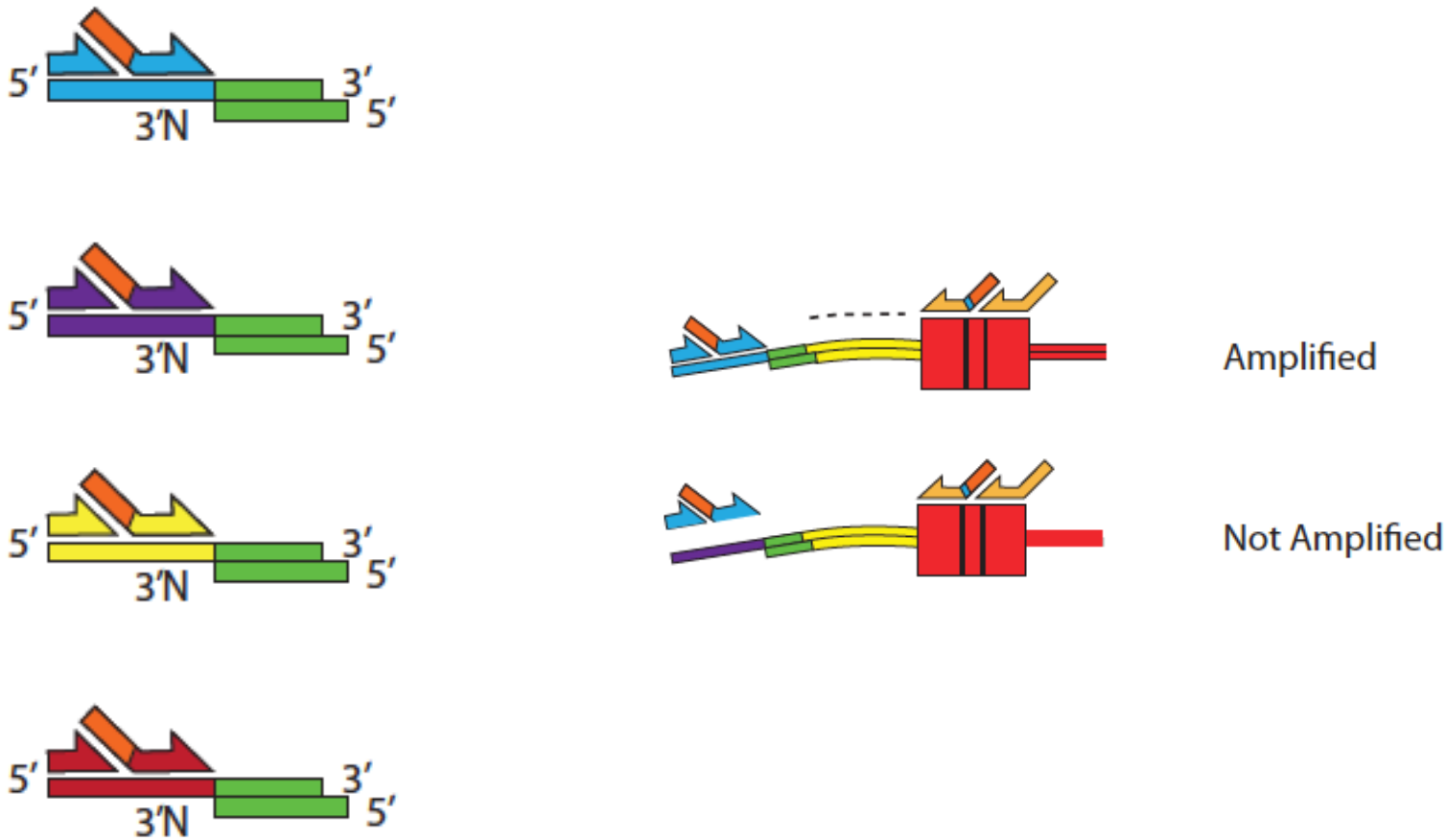
•Capture of integration sites using restriction enzyme cleavage of genomic DNA is highly biased

•Use of MuA-catalyzed transposition in vitro is far less biased

•Counting Mu transposition events leading to site recovery provides a simple abundance estimate

•The HMGA2 site is the only high abundance site harbored by the expanded clone from the beta thalassemia trial

Suppressing PCR cross over by cycling adaptors



Possible questions to be addressed with integration site analysis

What is the distribution of integration target sites relative to genomic land marks?

Has an adverse event taken place?

What molecular events were associated with an adverse event?

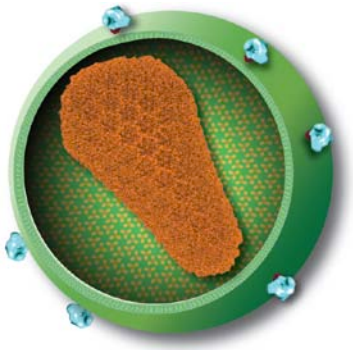
What is the clonal structure of a transduced cell population?

Credits

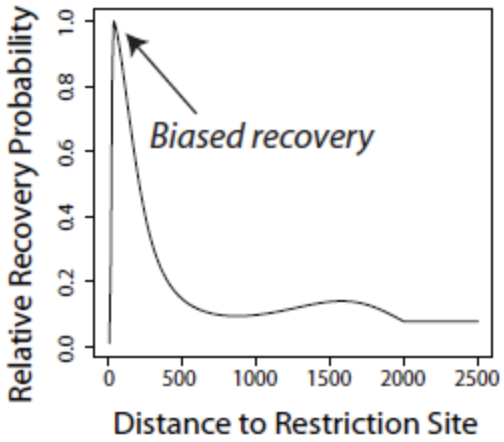
Bushman laboratory	Former Bushman Lab	Collaborators
Troy Brady	Rick Mitchell	Charles Berry
Kyle Bittinger	Brett Beitzel	Sumit Chanda
Rohini Sinha	Mary Lewinski	John Young
Keshet Ronen	Astrid Schroder	Renate Koenig
Karen Ocwieja	Angela Ciuffi	Joe Ecker
Christian Hoffmann	Heather Marshall Rose	Craig Hyde
Nirav Malani	Jeremy Leipzig	Mark Yeager
Rithun Mukherjee	Matt Culyba	Kushol Gupta
Young Hwang	Gary Wang	Greg Van Duyne
Rebecca Custers-Allen	Tracy Diamond	Masahiro Yamashita
Serena Dollive	Scott Sherrill-Mix	Mike Emerman
Jennifer Hwang	Nana Minkah	Francis Collins
Jimmy Hu		Nam Lee
Sam Minot		Paul Bieniasz
Greg Peterfreund		Philippe Leboulch
Shannah Roth		Alain Fischer
Frances Male		Marina Cavazzana-Calvo
Emily Charlson		Zeger Debyser
		Rik Gijsbers

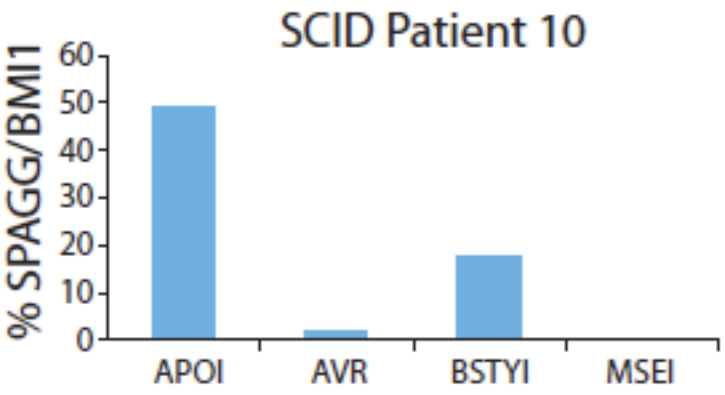
Lentiviral vectors for gene therapy

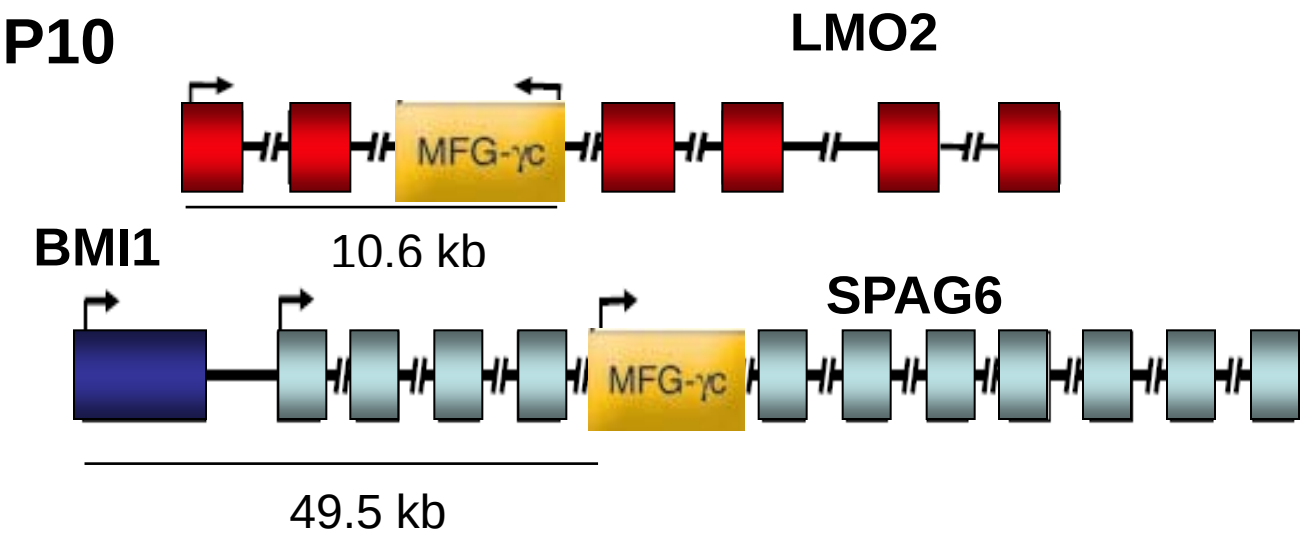
- No insertional mutagenesis in HIV-positive individuals
- Can infect non-dividing cells
- Other advantages



Use of restriction enzymes to cleave genomic DNA leads to severe biases







Beta thalassemia trial: Summary and perspective

•Lentiviral vector based stem cell gene therapy for beta-thalassemia rendered patient transfusion independent, healthy at present.

•Much of therapeutic benefit derived from a single clone with vector integration in the HMGA2 third intron.

•Cells overexpress HMGA2, associated with altered 3' end and increased rate of initiation.

•Cause and effect?

•Expanded clones with HMGA2 integration sites detected in SCID-X1 gene therapy. Overexpression and altered 3' end structure in patient cells confirmed in SCID.

•Examples of 3' end truncation and miRNA binding site removal associated with proto-oncogene activation in mice: Pim1, Gfi1, c-Myc, and Int-2

•No evidence for insertional mutagenesis leading to clonal dominance in mouse models of beta-thalassemia.

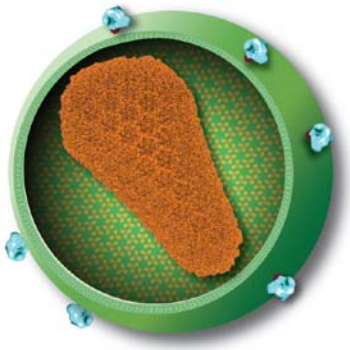
•No evidence for expansion of cells with HMGA2 integration sites during HIV infection.

- **Introduction**

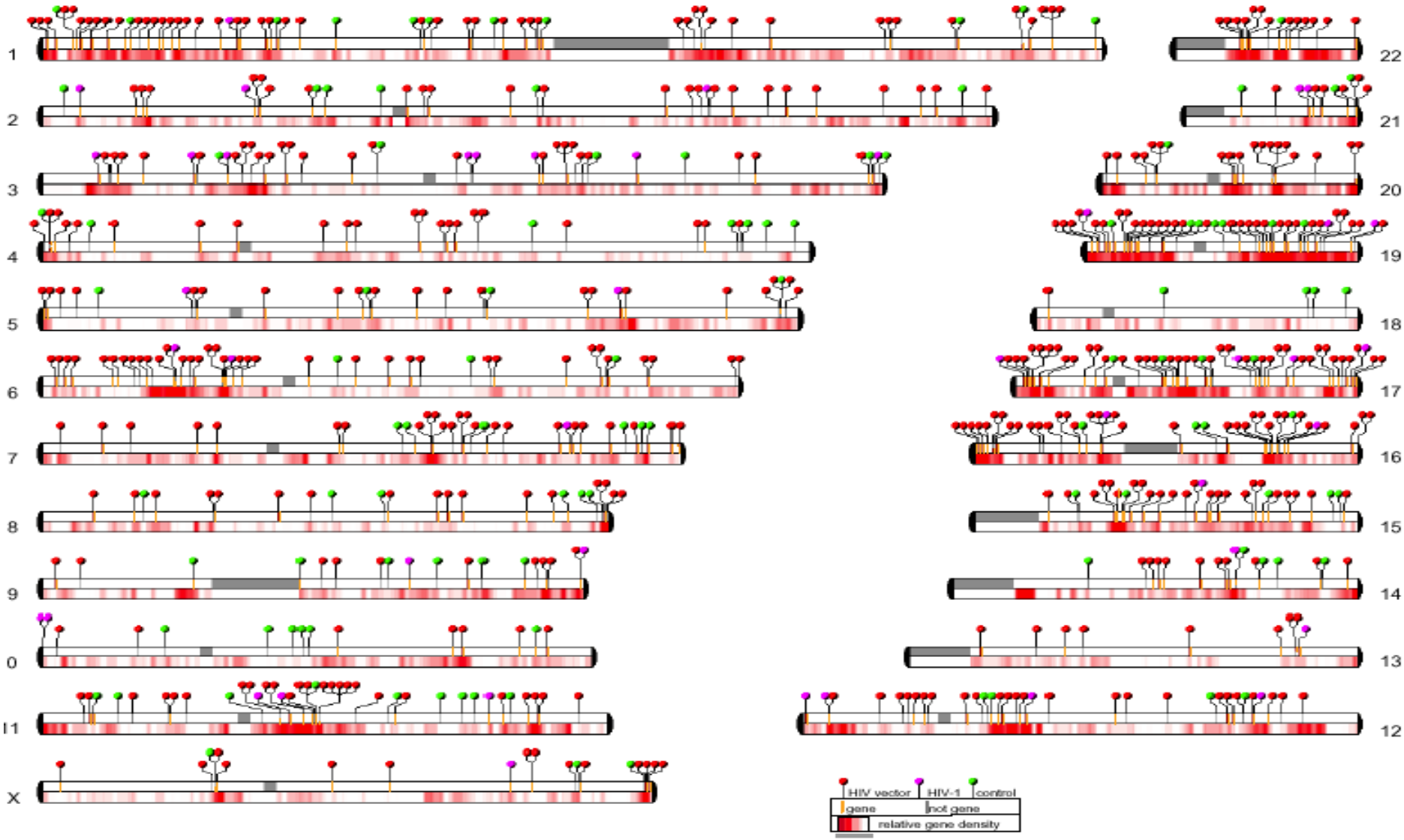
- **Retargeting lentiviral integration with LEDGF/p75 fusions**

- **Retargeting lentiviral integration by altering nuclear entry pathways**

- **Clonal expansion during lentiviral gene therapy associated with vector integration in a proto-oncogene**



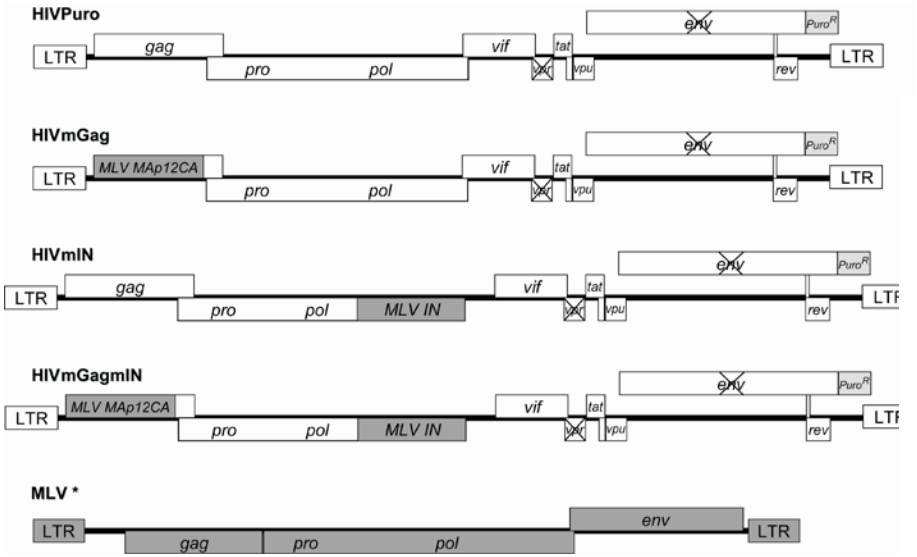
Sites of HIV-1 cDNA Integration in SupT1 Cells



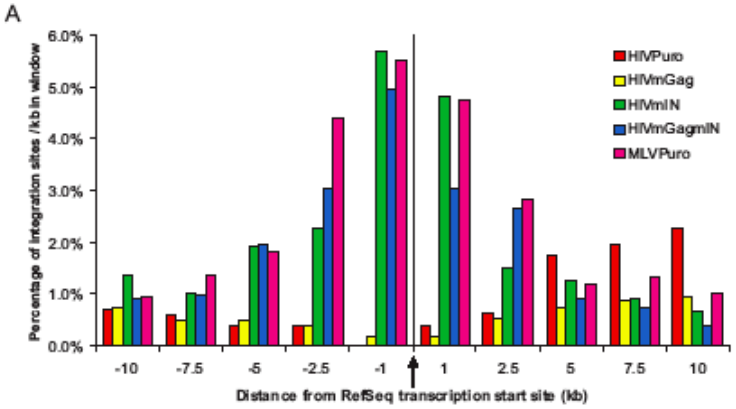
Schroder, Shinn, Chen, Berry, Ecker and Bushman, Cell, 2002

Integrase is a viral determinant of integration targeting

HIV/MLV chimeras studied



Chimeras with MLV IN show favored integration near transcription start sites



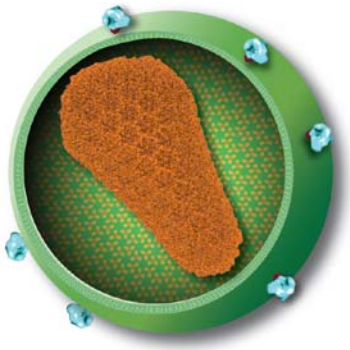
MK Lewinski, M Yamashita, M Emerman, A Ciuffi, H Marshall,G Crawford, F Collins, P Shinn, J Leipzig, JR Ecker, FD Bushman, PLOS Pathogens, 2006

- **Introduction**

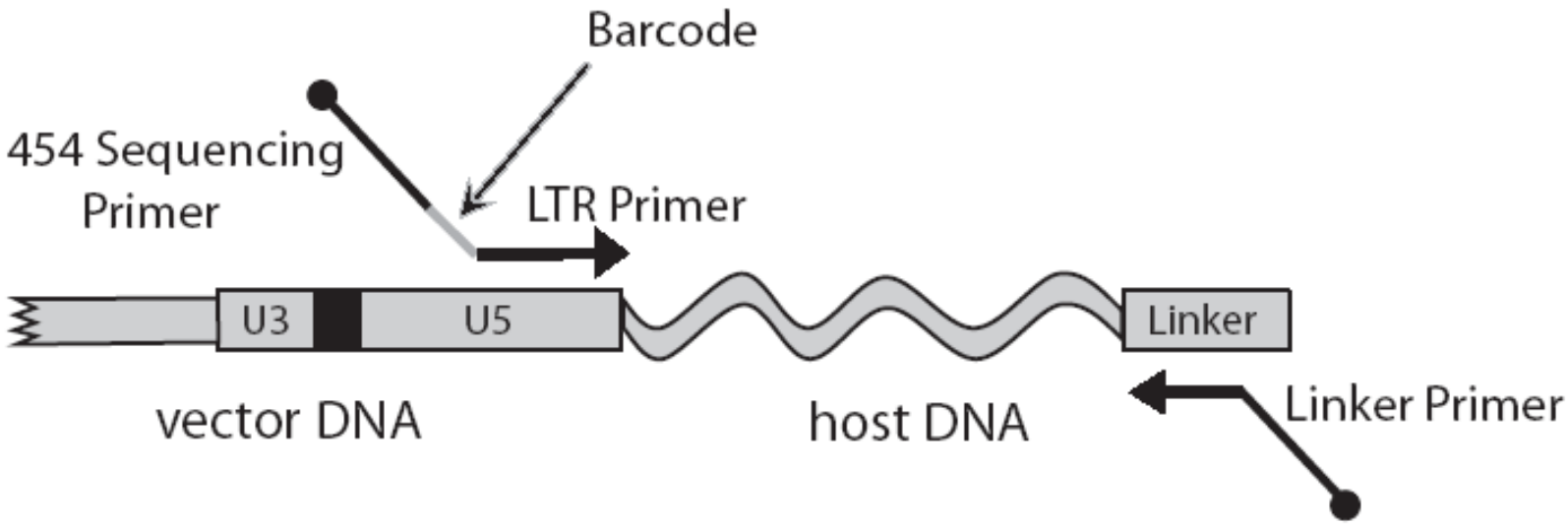
- **Retargeting lentiviral integration with LEDGF/p75 fusions**

- **Retargeting lentiviral integration by altering nuclear entry pathways**

- **Clonal expansion during lentviral gene therapy associated with vector integration in a proto-oncogene**



DNA bar coding



Sample primer sequence

GCCTTGCCAGCCCGCTCAG - CACG - AGACCCTTTTAGTCAGTGTGGAAAATC

454 Sequencing Primer Barcode LTR Primer

454/Roche sequencing of model populations

40,000 HIV integration site sequences from Jurkat cells

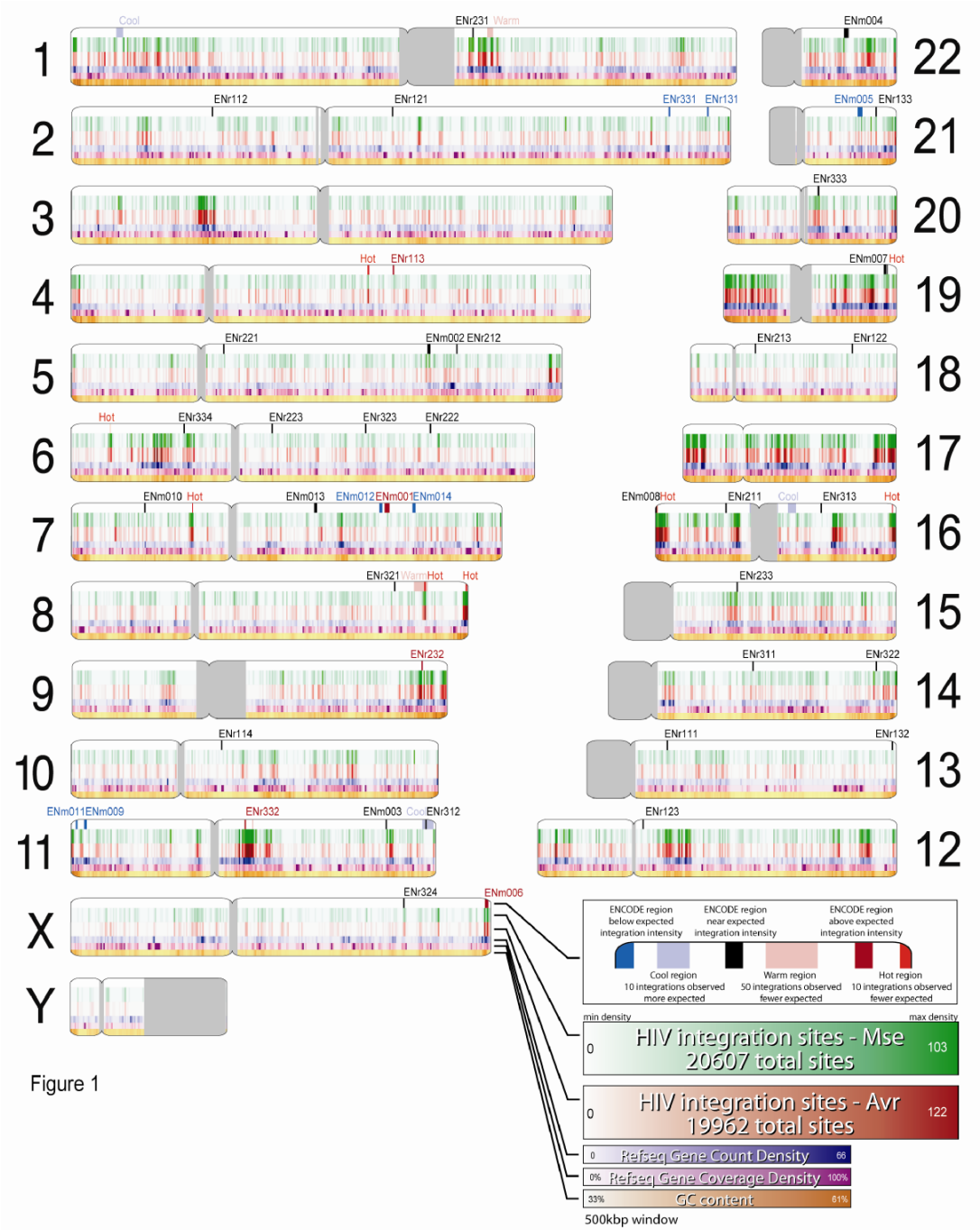


Figure 1

Wang, Leipzig, Berry, and Bushman, Genome Res., 2007

PSIP1/LEDGF/p75: a host cell factor involved in HIV integration targeting

Transcriptional coactivator

Ge et al. 1988

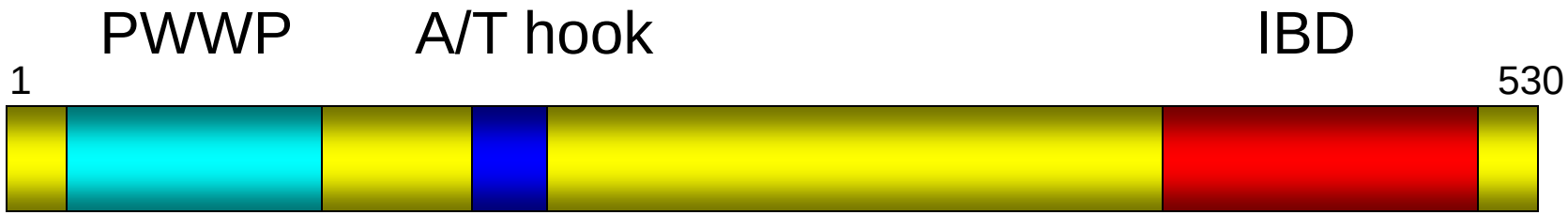
Binds tightly to HIV IN

Cherepanov et al. 2003, Maertens et al. 2003, Llano et al., 2004, Turlure et al., 2004

Promotes integration

Emiliani et al., 2005, *Llano et al., 2006, Shun et al. 2007, Marshall et al., 2007*

LEDGF/p75 protein

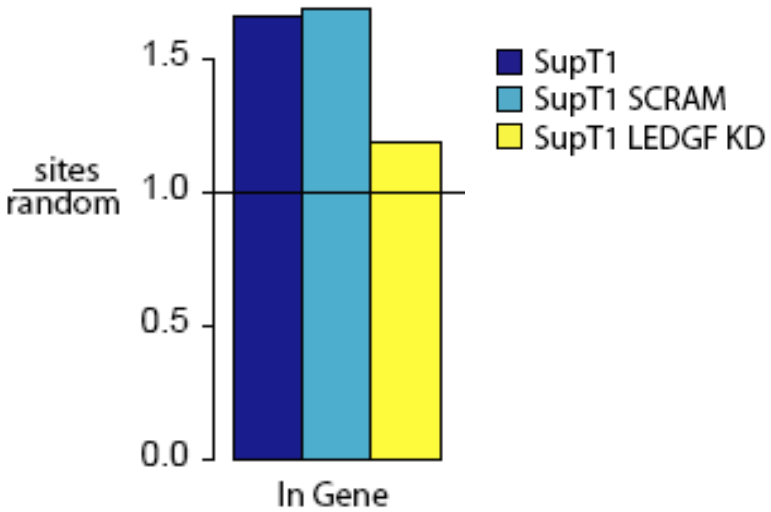


Data from labs of Poeschla, Engelman, Debyser, Cherepanov, Benharous, Bushman…

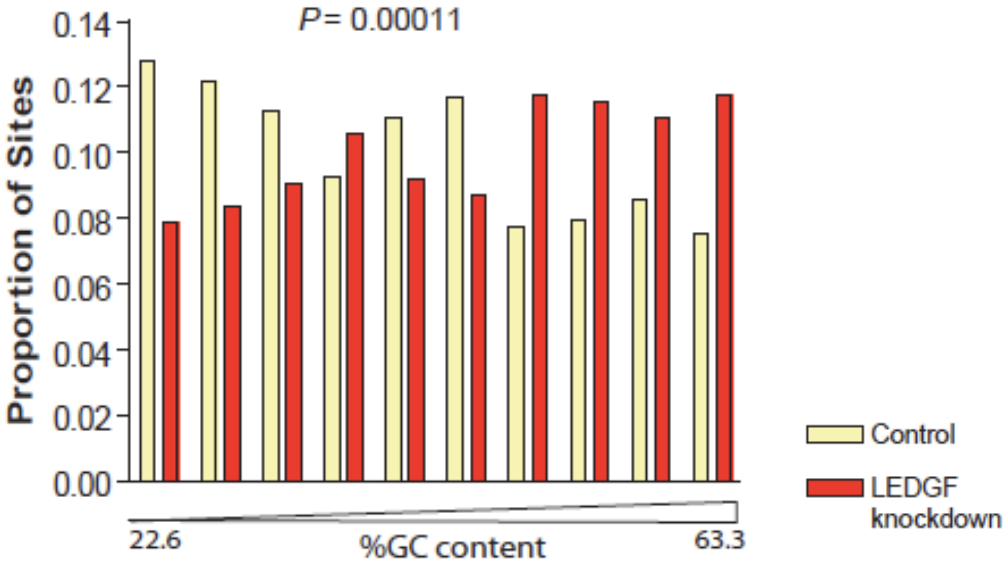
Depletion of PSIP1/LEDGF/p75 Redirects Integration

Data from SupT1 knockdown (human), MEF gene trap (murine)
Studied HIV and equine infectious anemia virus (another lentivirus)

Integration in transcription units

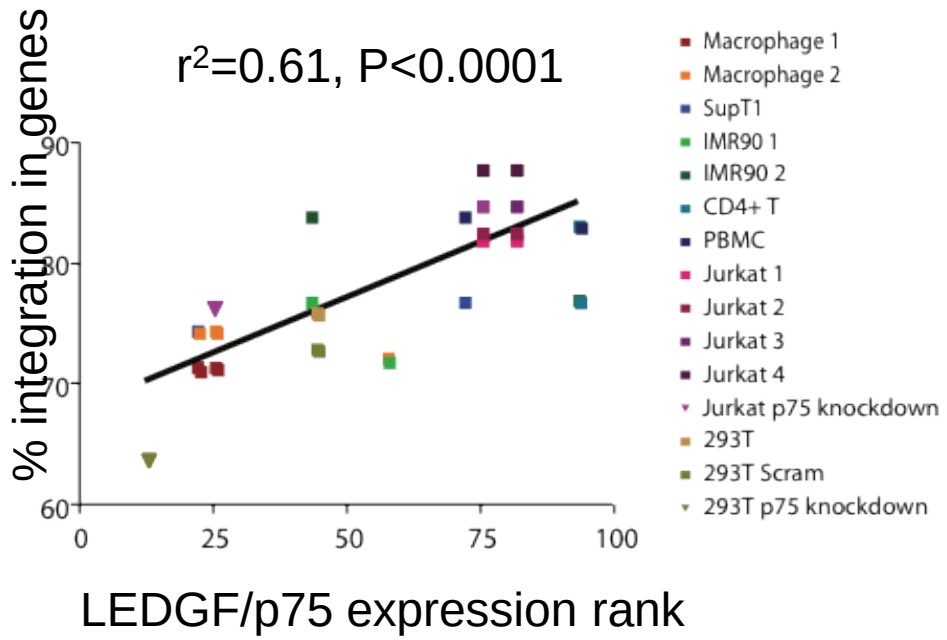


Integration in GC-rich regions



Ciuffi, M. Llano, E. Poeschla, H. Marshall, C. Hoffmann, P. Shinn, S. Hannenhalli, J. Ecker, F. Bushman. Nature Medicine, 2005.
Shun, Cherepanov, Engelman et al. Genes and Dev. 2007
Marshall, Ronen, Bickmore, Poeschla, Bushman et al., PLoS One, 2007

LEDGF/p75 expression and integration frequency in transcription units



- Different cell types show reproducible differences in the frequency of HIV integration in transcription units
- These differences correlate with the relative expression level of LEDGF/p75

The more LEDGF/p75, the greater fraction of integration events in transcription units

Retargeting lentiviral integration away from genes

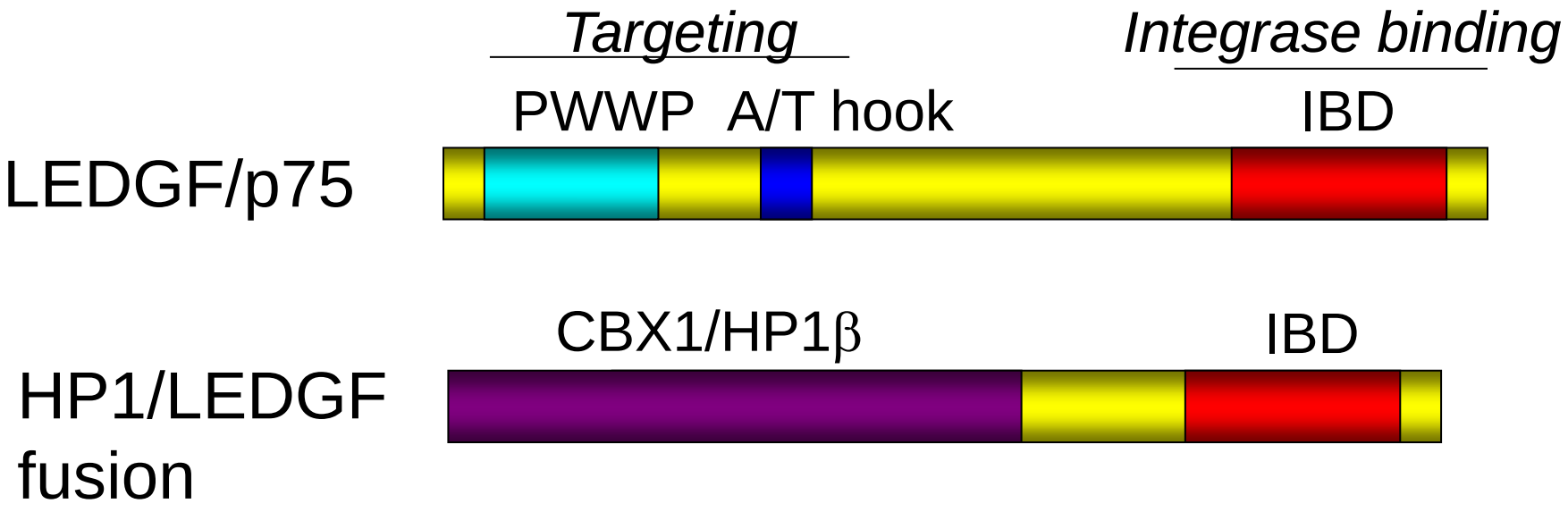
Rik Gijsbers, Keshet Ronen, Sofie Vets, Nirav Malani, Jan De Rijck, Melissa McNeely, Frederic D. Bushman, and Zeger Debyser

See also papers from Daniel, Engelman, Hughes and colleagues

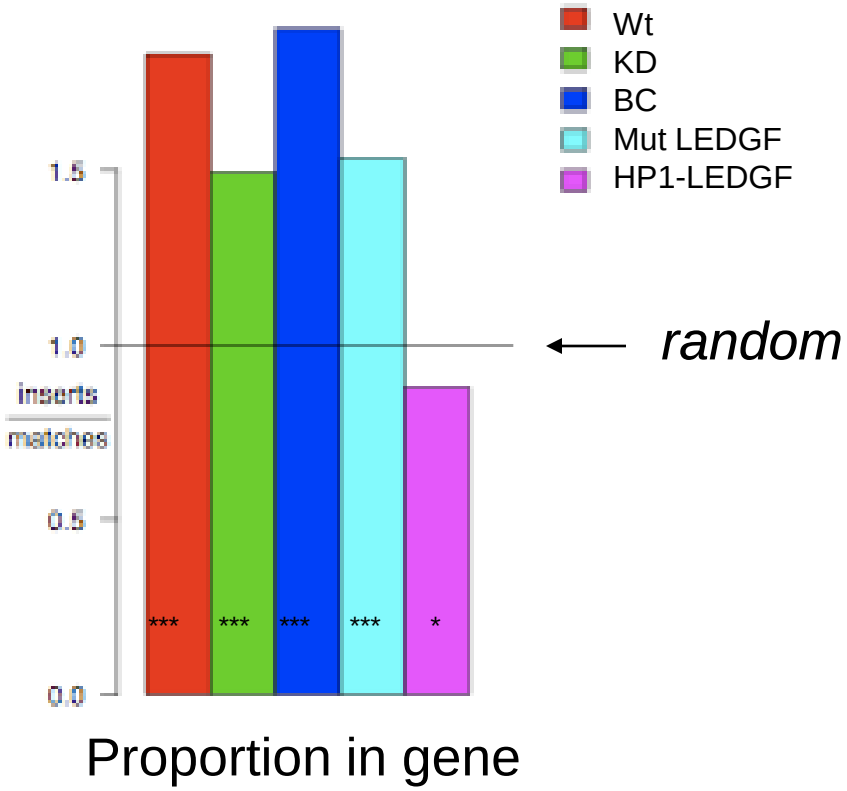
Retargeting lentiviral integration using fusions of HP1β(CBX) to LEDGF/p75

HP1 enriched in heterochromatic regions particularly around centromeres

Binds to H3K9me2&3

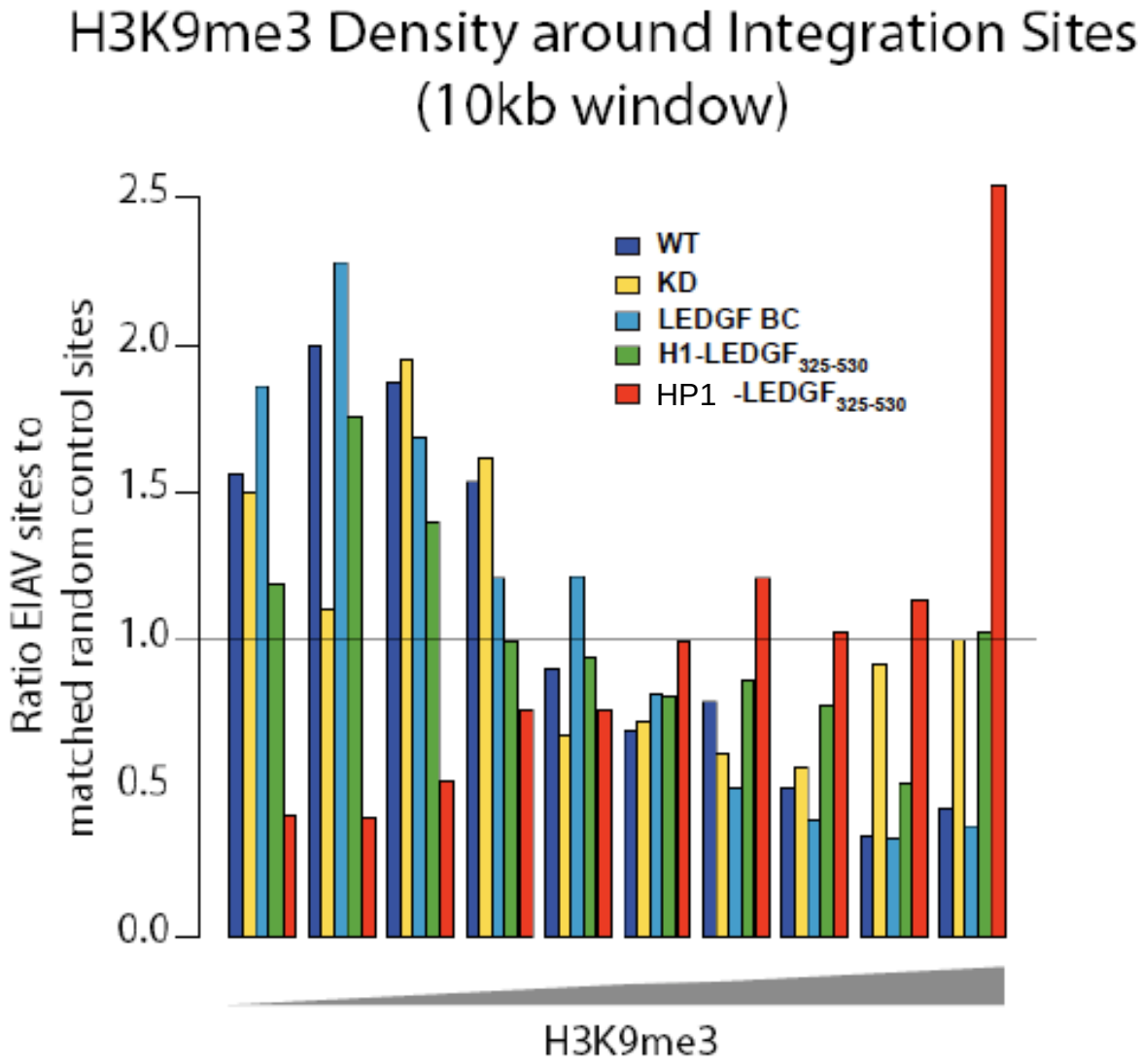


Integration in Transcription Units



Expression of HP1-LEDGF fusion directs integration away from genes

Integration in presence of HP1-LEDGF fusion favored near sites of H3K9me3



Summary

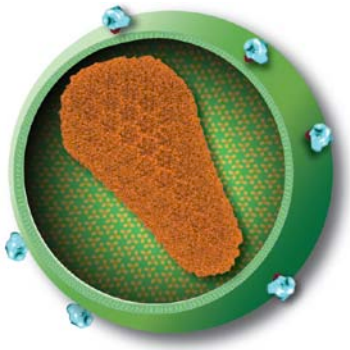
- HIV favors integration in active transcription units
- Other retroviruses have different favored targets
- PSIP1/LEDGF/p75 tethers HIV integration complexes to DNA
- Applications: controlling integration inside cells? New drug target?

- **Introduction**

- **Retargeting lentiviral integration with LEDGF/p75 fusions**

- **Retargeting lentiviral integration by altering nuclear entry pathways**

- **Clonal expansion during lentviral gene therapy associated with vector integration in a proto-oncogene**

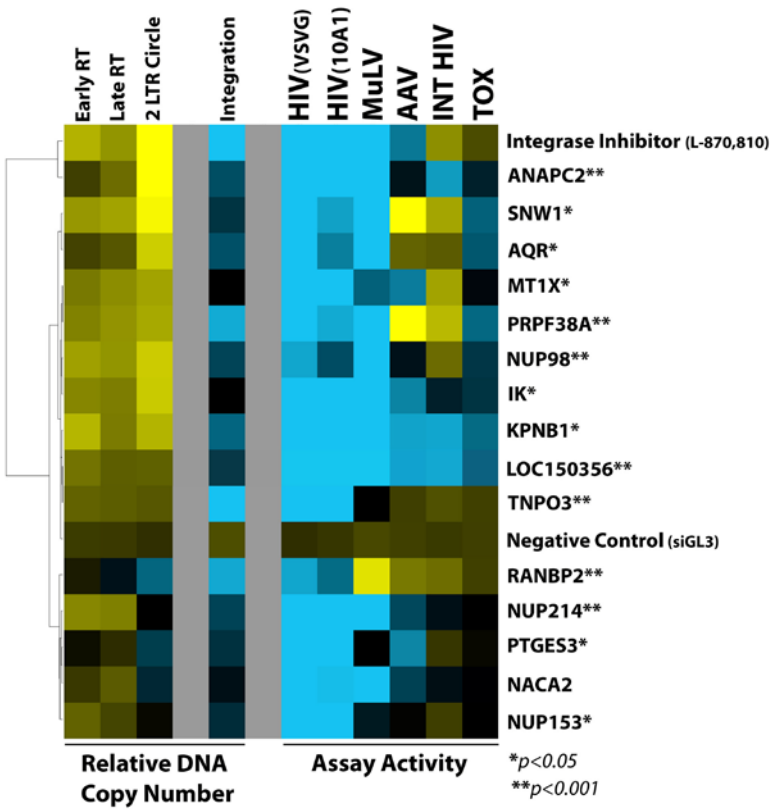


Host Factors in HIV Integration Targeting: TNPO3 and RANBP2 (NUP358)

Karen E. Ocwieja, Troy Brady, Keshet Ronen, Nirav Malani, Alyssa Huegel, Torsten Schaller, Greg Towers, Renate Konig, Sumit Chanda, John Young, Charles C. Berry, Frederic D. Bushman

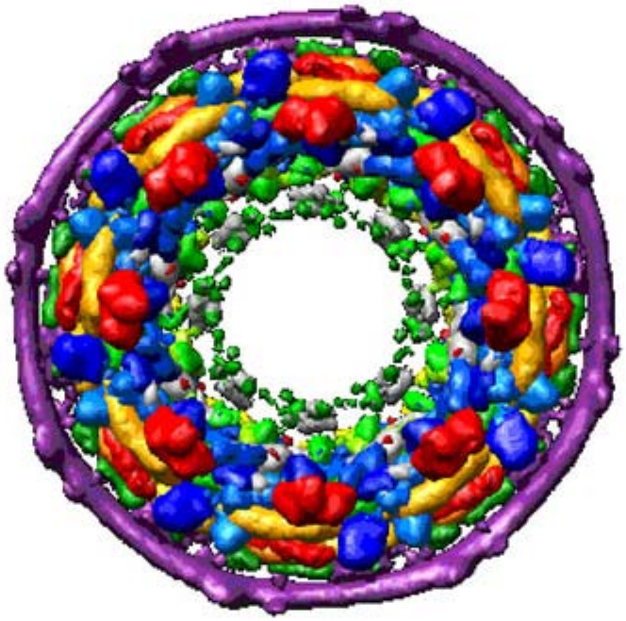
Submitted

Genes affecting integration



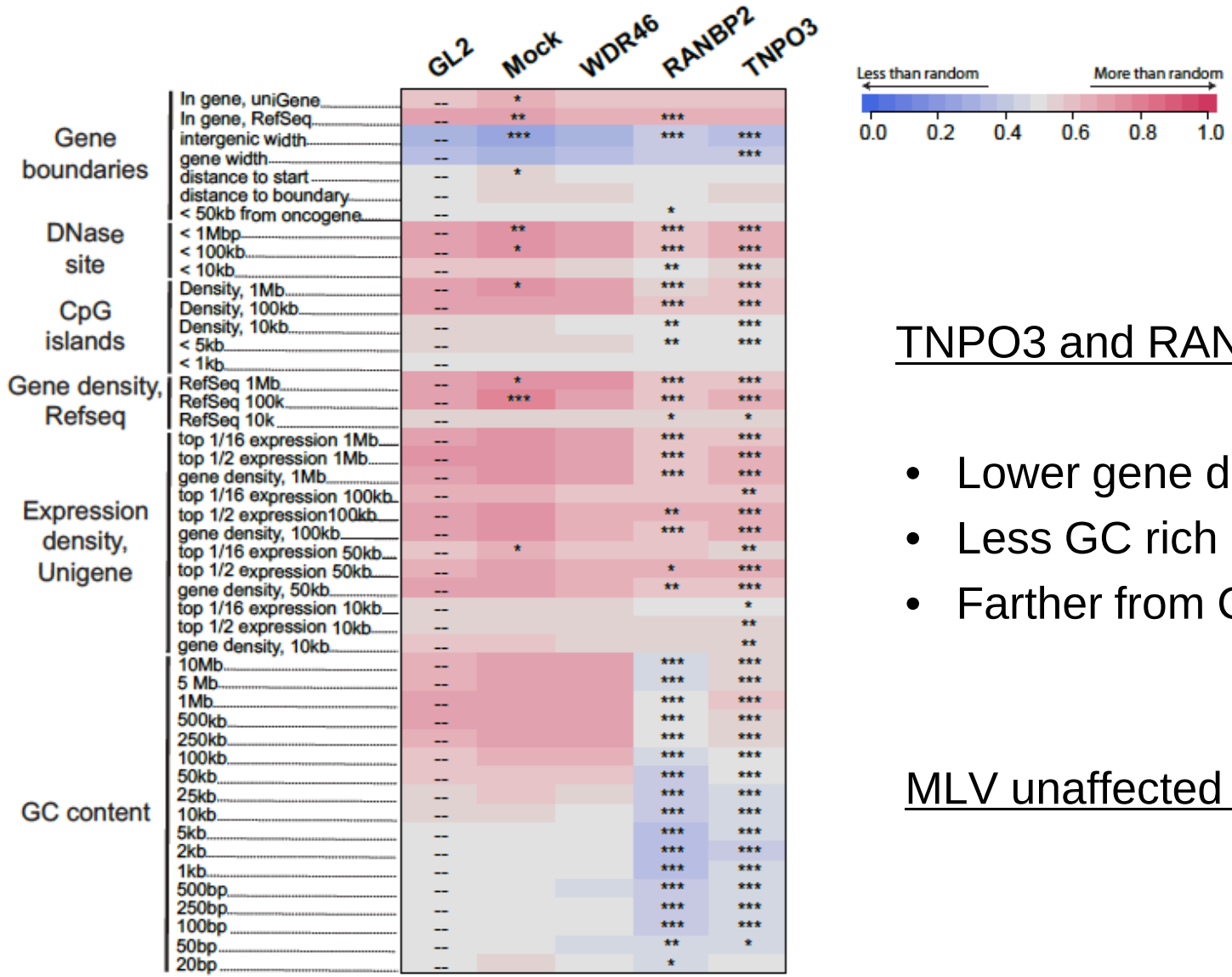
Konig et al., Cell 2007

Role for nuclear pore/import machinery even in dividing cells; both before and after entry



Coupling of subcellular sorting, nuclear import, integration?

Genomic features at integration sites



TNPO3 and RANBP2 knockdowns

- Lower gene density
- Less GC rich
- Farther from CpG islands

MLV unaffected by knockdowns

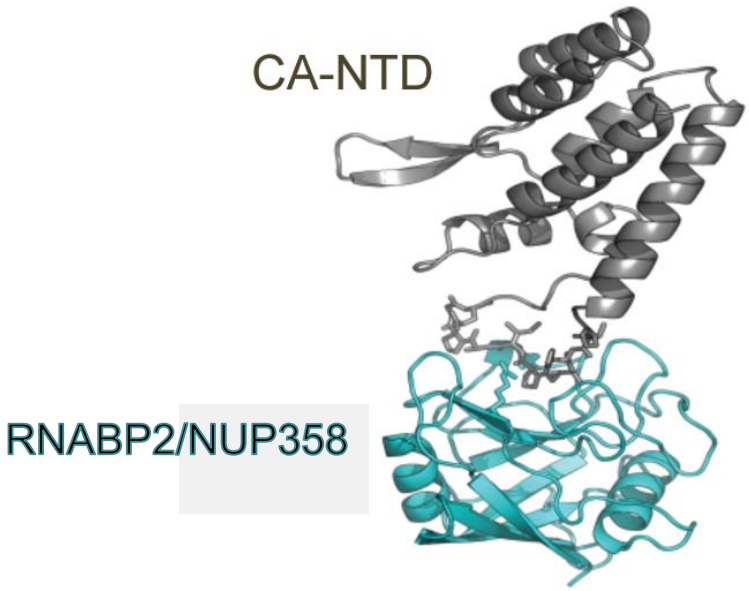
TNPO3/RANBP2 knockdowns influence integration targeting

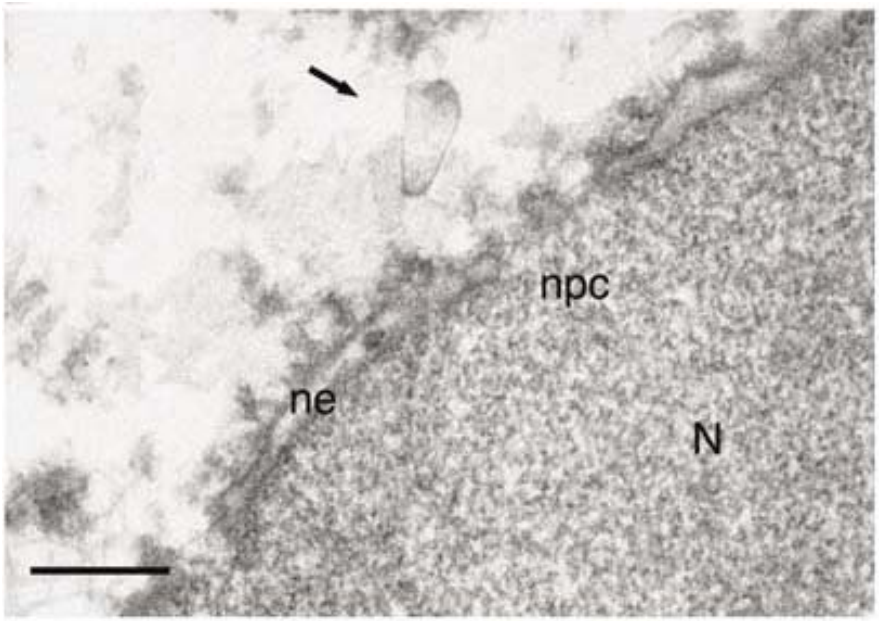
HIV-1 Capsid-Cyclophilin interactions determine nuclear import pathway and integration targeting

Torsten Schaller, Amanda J. Price, Karen E. Ocwieja, Jane K. Rasaiyaah, Troy Brady, Shannah Roth, Adam J. Fletcher, KyeongEun Lee, Vineet N. KewalRamani, Mahdad Noursadeghi, Richard G. Jenner, Frederic D. Bushman, Leo C. James, Greg J. Towers

Submitted

Capsid and nuclear pores

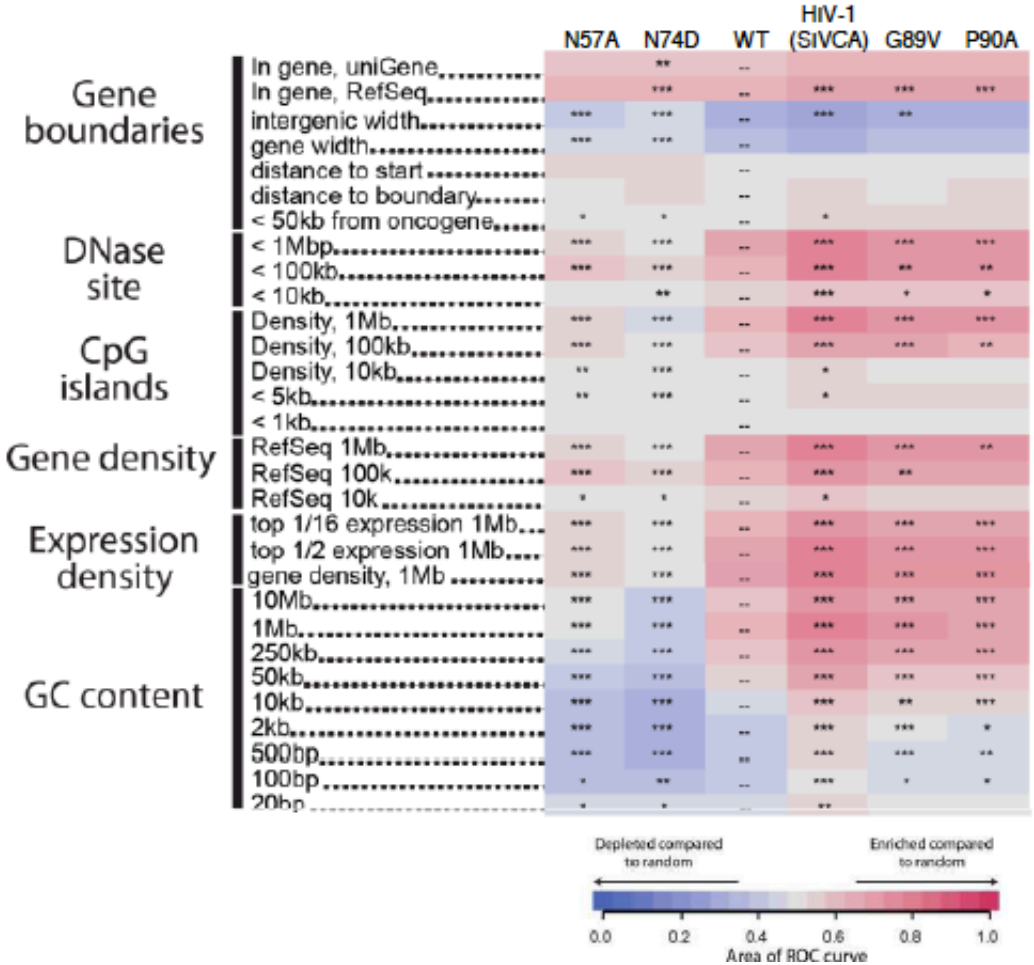


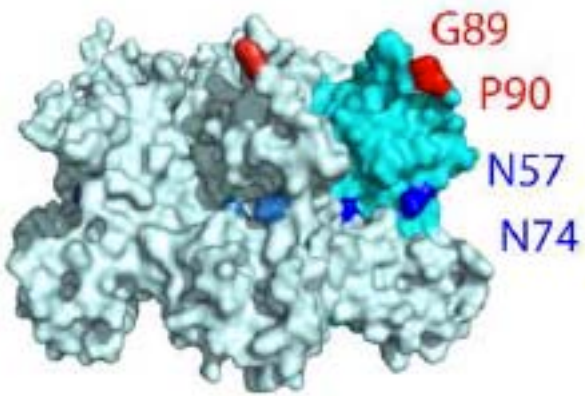


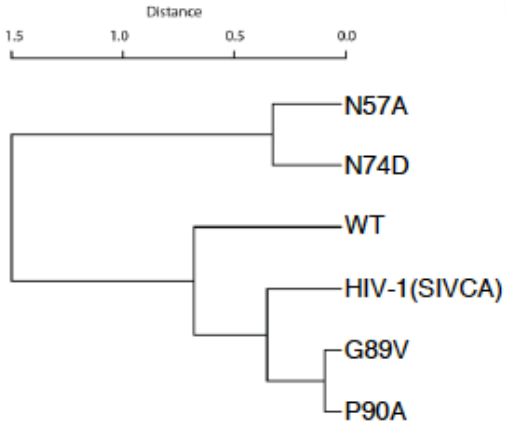
Schaller et al., submitted

Arhel et al., EMBO 2007

Capsid mutants alter integration targeting







- Determined ~20,000 unique integration site sequences
- Targeting redirected toward more gene dense regions (cyclophillin non-binders) and less gene dense regions (non-TNPO3 users)
- Cyclosporin treatment phenocopies loop mutants

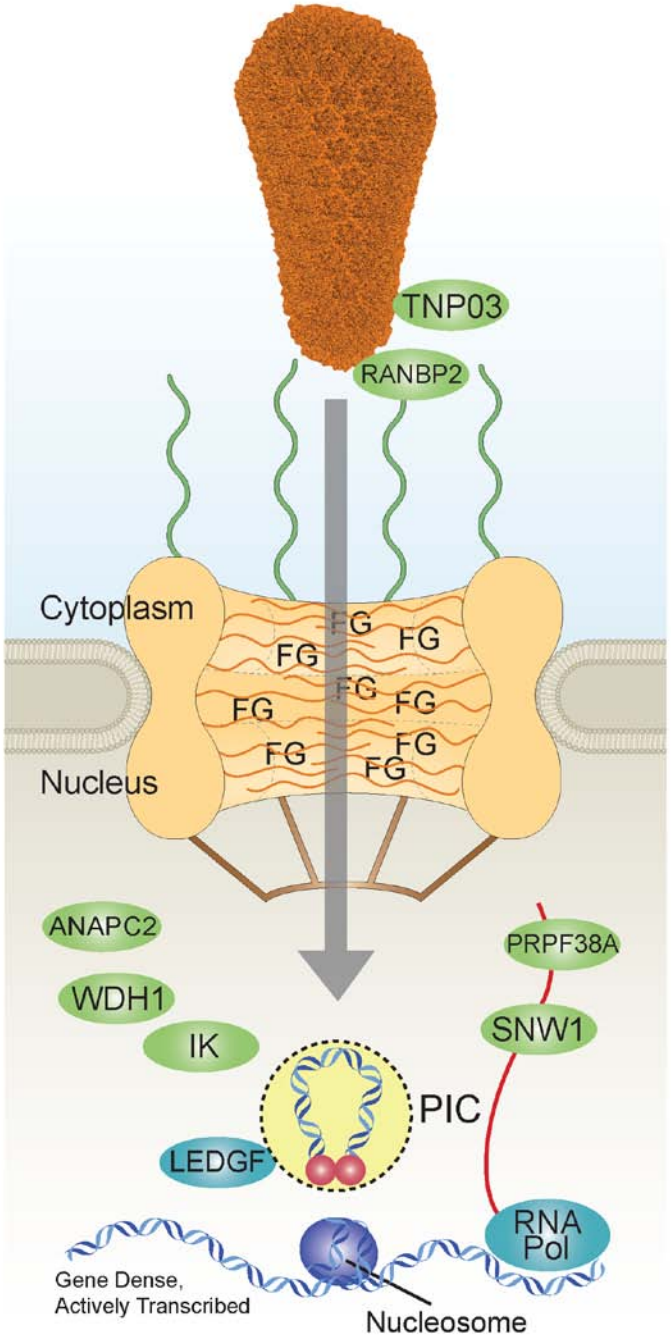
Summary

Knockdown of nuclear targeting/pore factors TNPO3 and RANBP2 alters integration site selection

HIV CA binds tightly to RANBP2

Amino acid substitutions in HIV CA alter integration targeting

Supports model involving a railroad track from the nuclear pore to favored integration sites



Use of iPS to engineer cells with vectors in “Safe Harbors”

•Isolate bone marrow cells from beta thalassemia patients
•Create iPS
•Transduce with lentiviral vectors expressing beta globin
•Single cell clone transduced iPS cells
•Analyze integration site locations one by one
•Select cell lines that have a vectors integrated in safe harbors

In 20 cell clones, 2 were identified with integration sites outside of genes, 50 kb from gene 5’ end, 300 kb from proto-oncogenes, not within ultraconserved regions.

Eirini P. Papapetrou, Gabsang Lee, Nirav Malani, Isabelle Riviere, Frederic D. Bushman, Lorenz Studer and Michel Sadelain,

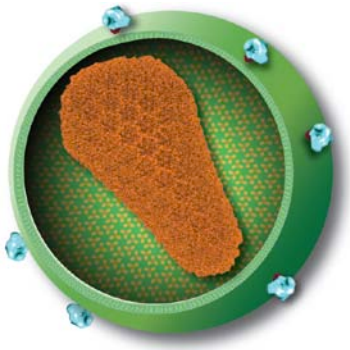
Nature Biotech., in press

- **Introduction**

- **Retargeting lentiviral integration with LEDGF/p75 fusions**

- **Retargeting lentiviral integration by altering nuclear entry pathways**

- **Clonal expansion during lentviral gene therapy associated with vector integration in a proto-oncogene**



Clinical benefit with partial clonal dominance after lentiviral gene therapy for severe human thalassemia

M. Cavazzana-Calvo, E. Payen, O. Negre, G. Wang, K. Hehir, F. Fusil, J. Down M. Denaro, R. Pawliuk, K.A. Westerman, R. Cavallesco, B. Gillet-Legrand, L. Cacavelli, S. Sunyaev, F. Bernaudin, R. Girot, R. Dorazio, G.-J. Mulder, A. Polak, A. Bank, J. Soulier, J. Larghero, N. Kabbara, G. Socie, S. Chretien, N. Cartier, P. Aubourg, A. Fischer, K. Cornetta, Y. Beuzard, E. Gluckman, F. Bushman, S. Hacein-Bey-Abina, P. Leboulch.

Nature, 2010

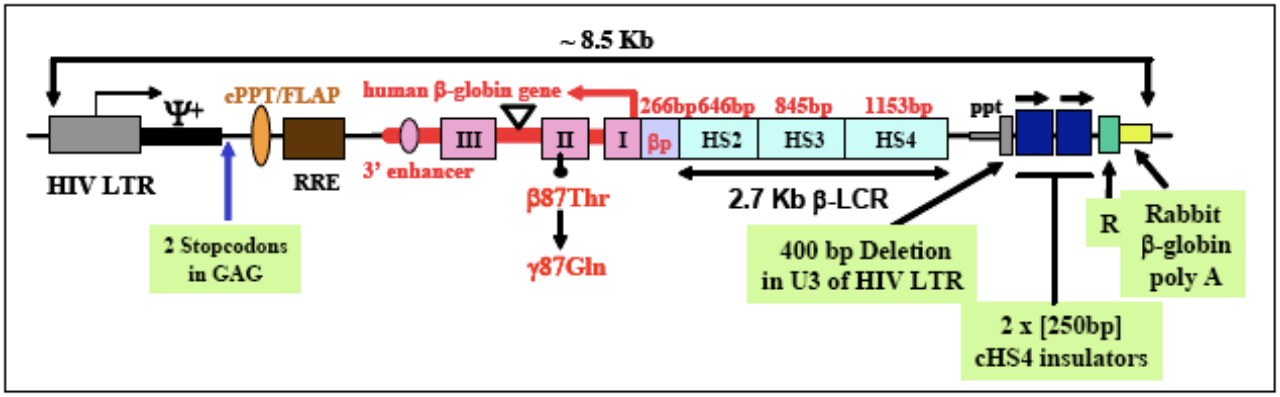
Lentivirus-based gene therapy for beta-thalassemia

Transduced hematopoietic stem cells with lentiviral vector
Successful engraftment in one out of two patients

Beta-thalassemia patient Px converted to transfusion independence

Healthy to date

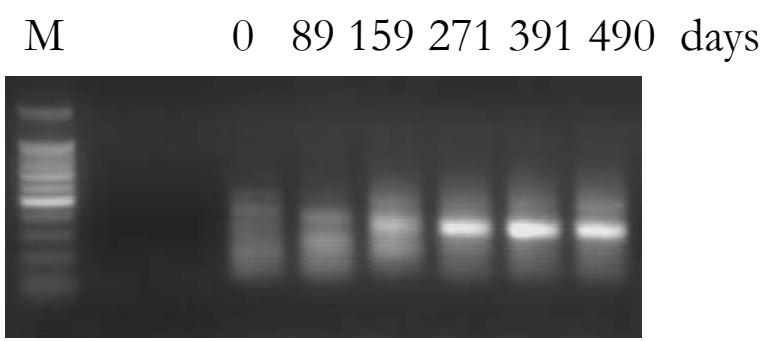
Integration site analysis disclosed that much of the therapeutic benefit from a single dominant clone with integration site in HMGA2



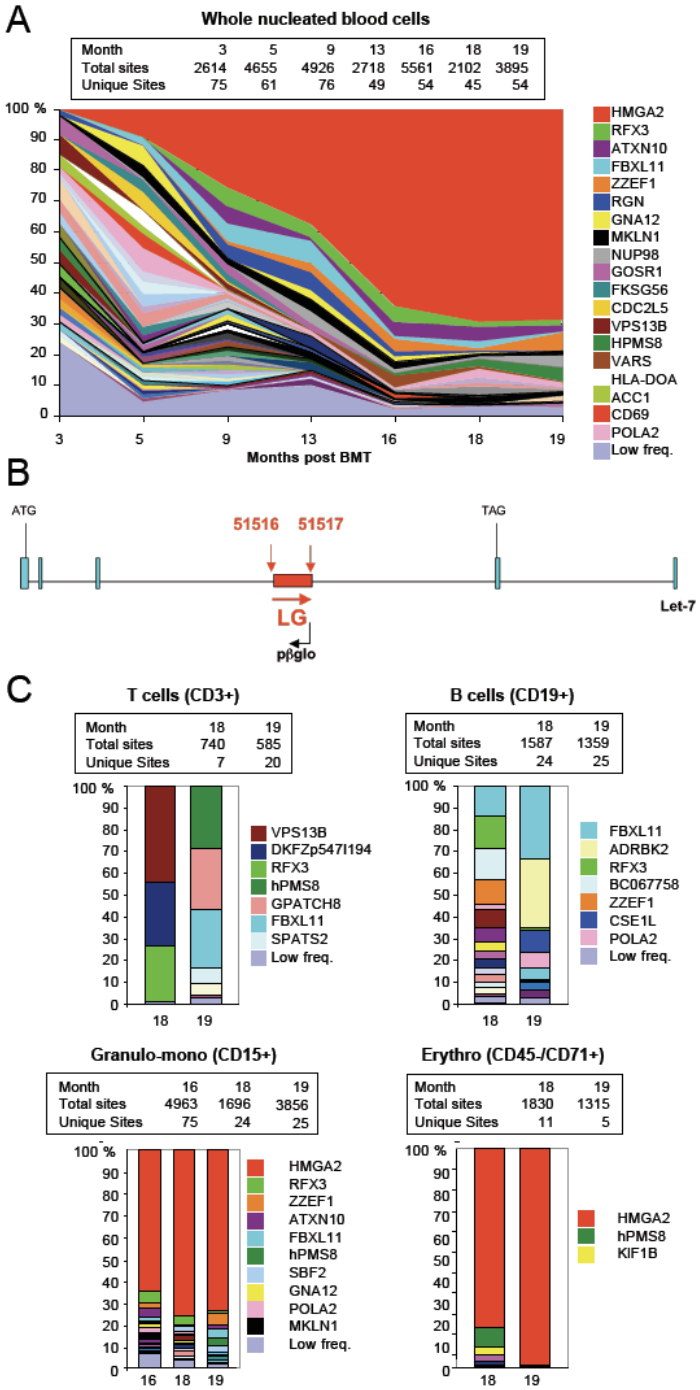
Integration site analysis

Ligation-mediated PCR to recover sites

DNA bar coding and 454 sequencing



Massive expansion of a cell clone with the lentivector integrated in the 3rd intron of the proto-oncogene HMGA2



HMGA2

Architectural DNA binding protein

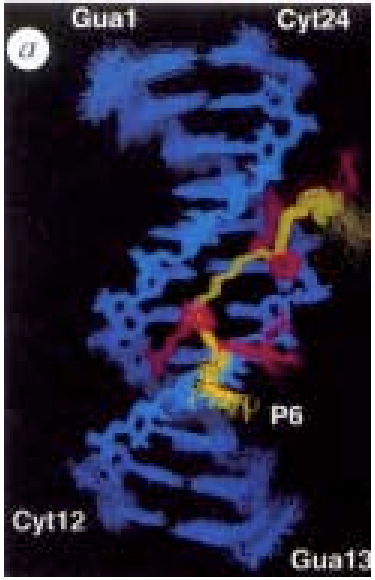
Three A/T hooks, bring together DNA duplexes

Expression up-regulated in many cancers

Gene rearranged in benign lipomas, others

miRNA regulation via 3' end, important in transformation

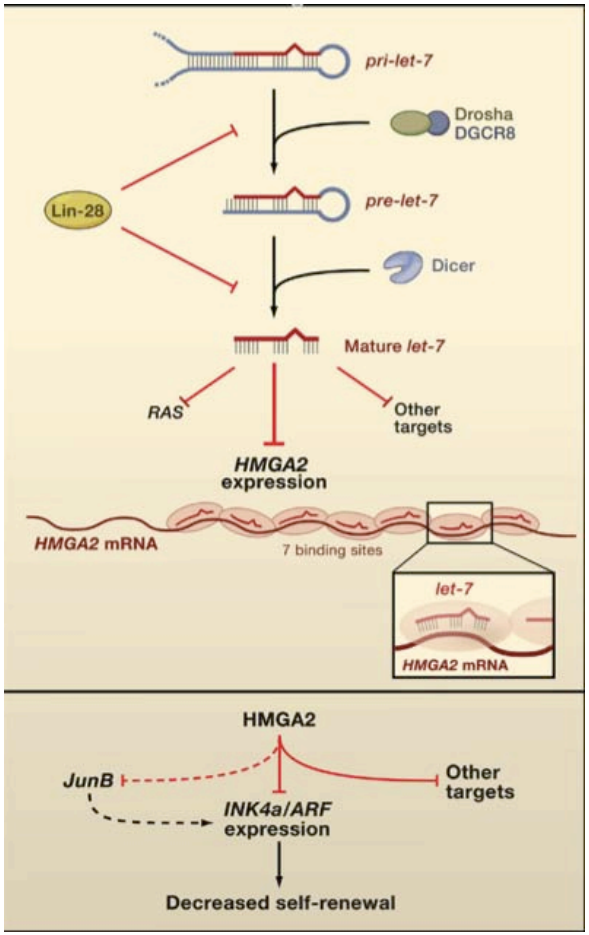
Implicated in reversing stem cell senescence



NMR structure of AT-hook/DNA complex

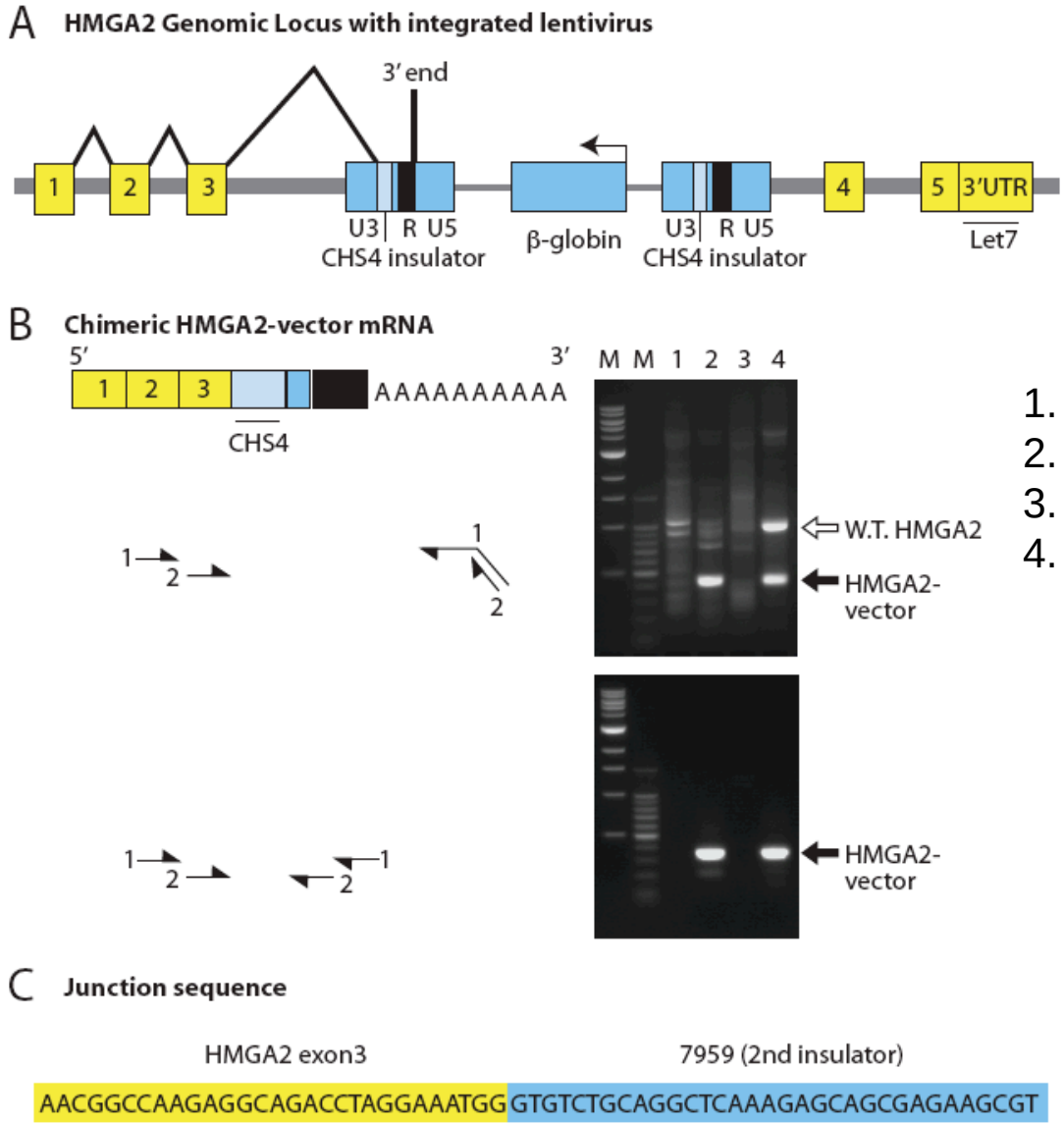
Huth et al., 1997
Nat. Struct. Bio.

HMGA2 regulation of cell proliferation



Hammond and Sharpless
Cell 135, 1013-1016 2008

HMGA2 overexpression and 3’ end substitution



1. Pre-transplant
2. Whole blood day 587
3. CD3+ cells day 587
4. Erythroblasts day 587

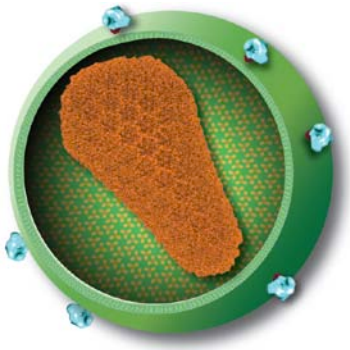
Conclusion: HMGA2 mRNA overexpressed with substituted 3’ end

- **Introduction**

- **Retargeting lentiviral integration with LEDGF/p75 fusions**

- **Retargeting lentiviral integration by altering nuclear entry pathways**

- **Clonal expansion during lentviral gene therapy associated with vector integration in a proto-oncogene**



Summary and speculations

Knockdowns of several host factors encoding nuclear pore (RANBP2) and transport (TNPO3) proteins affect integration targeting.

X-ray structure for RNABP2(NUP358) bound to CA. TNPO3 reported to bind CA, IN.

Capsid mutants can alter integration targeting.

Cyclophillin and cyclosporin also players.

Speculation: multiple “railroad tracks” into the nucleus; which track you take influences where you integrate; at least some tracks are influenced by CA interactions with RANBP2(NUP358), TNPO3 and cyclophillin.

Milk Duds
Milk Duds
CHARLESTON CHEWS
PACK-A SNACK!
1.79
1.69
1.29
RAISINETS
RAISINETS
SOUR PATCH
Sweet
Sour then Sweet
Kids
1.79
1.02
Mike Ike
Original Fruits
FAT FREE
FLAVORS
FLAVORS
1.79
Mike Ike Berry Blast
Mike Ike Berry Blast
Mike Ike Berry Blast
Mike Ike Berry Blast
1.02

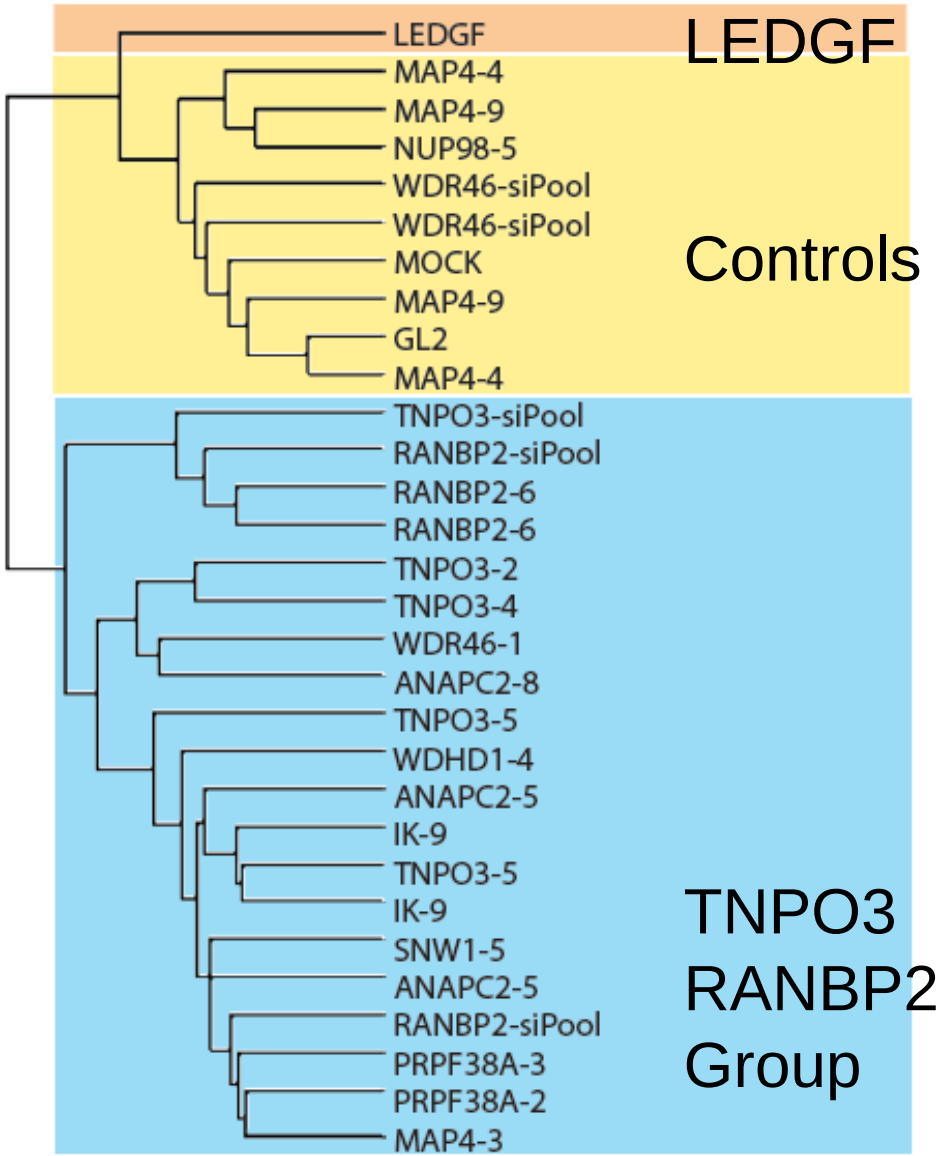
Clustering over many forms of genomic annotation

(i. e. in genes, local sequence, GC content, gene density, CpG island density, etc.)

<u>Conditional logit model</u>

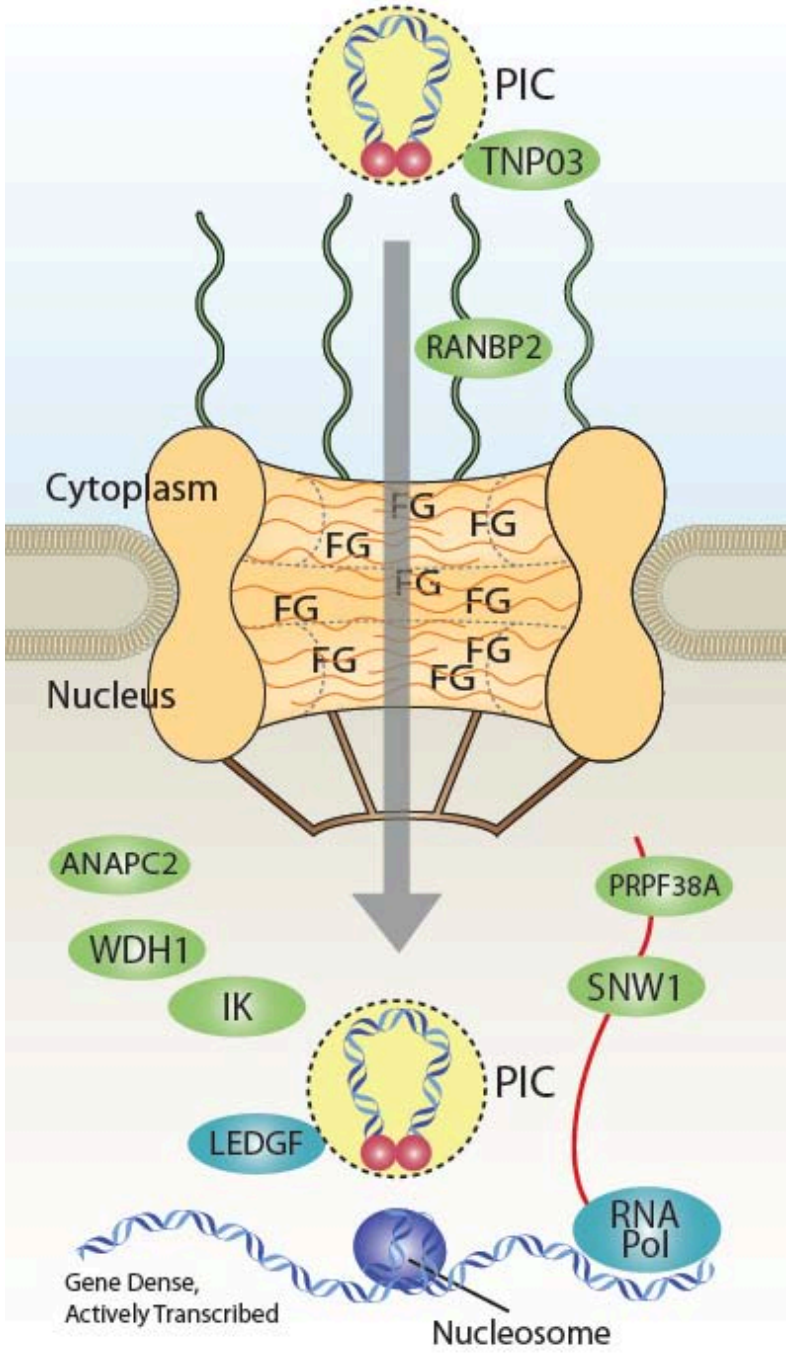
- Cluster based on relative resemblance over collection of genomic features

Charles Berry



Separation of LEDGF, Controls, and TNPO3/RANBP2 group

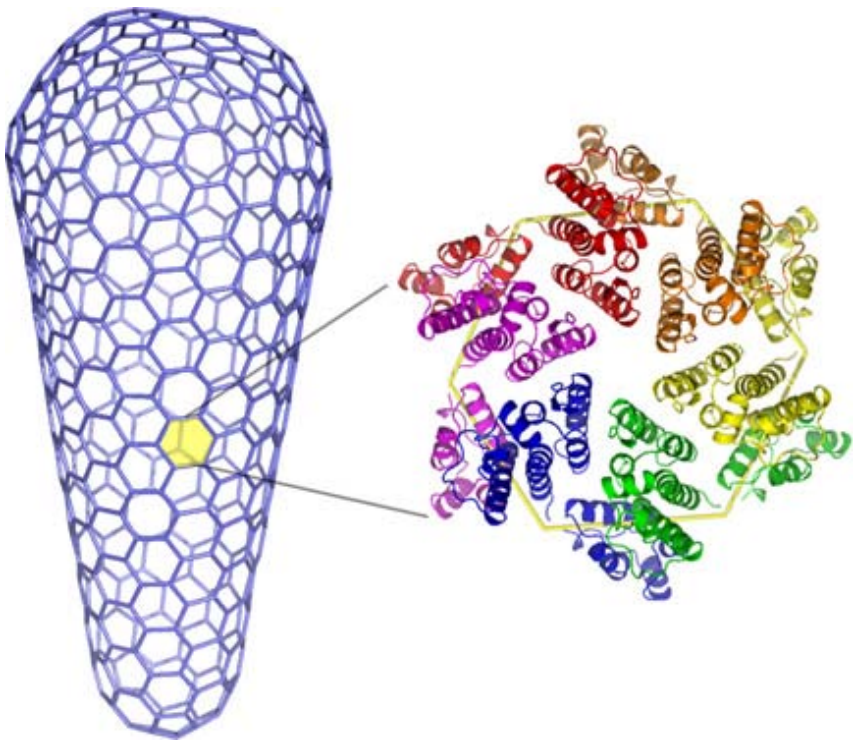
Model for “railroad track”



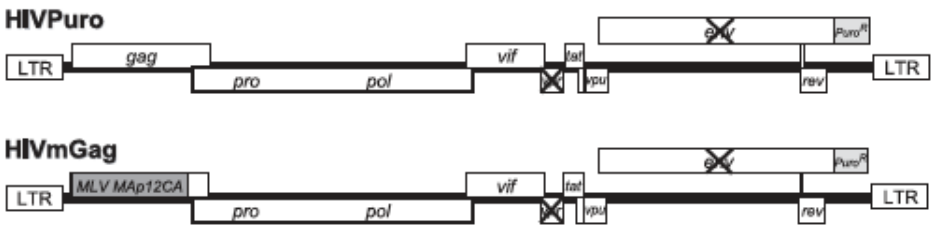
HIV Capsid and the Nuclear Pore

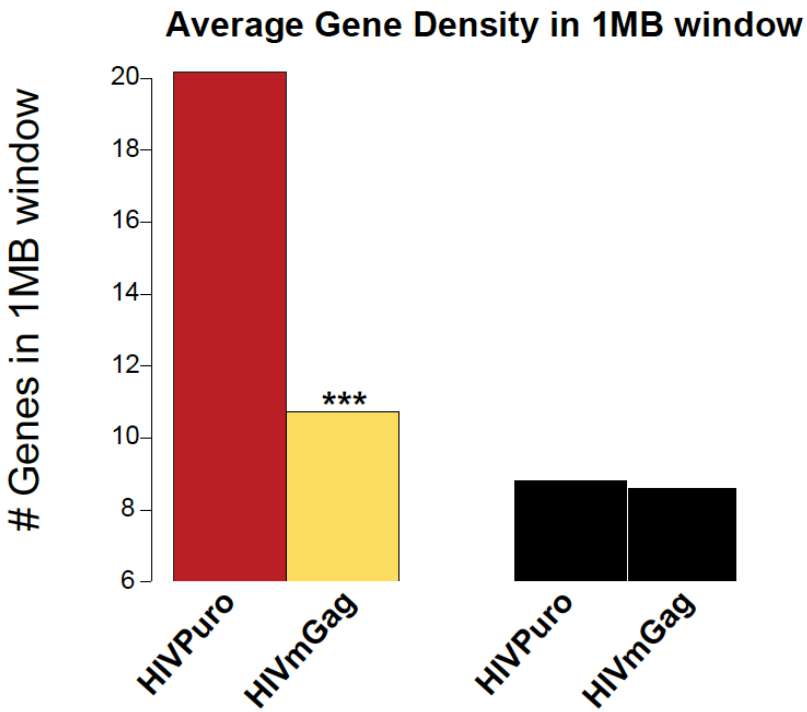
- HIV with MLV capsid does not require TNPO3 for infection (Krishnan et al., 2009)
- Capsid mutant N74D does not require TNPO3 for infection (Lee et al., 2010)
- Capsid mutant T54A/N57A enters nucleus in non dividing cells but does not integrate (Yamashita et al., 2007)

HIV Capsid



Integration targeting by HIV substituted with MLV Gag





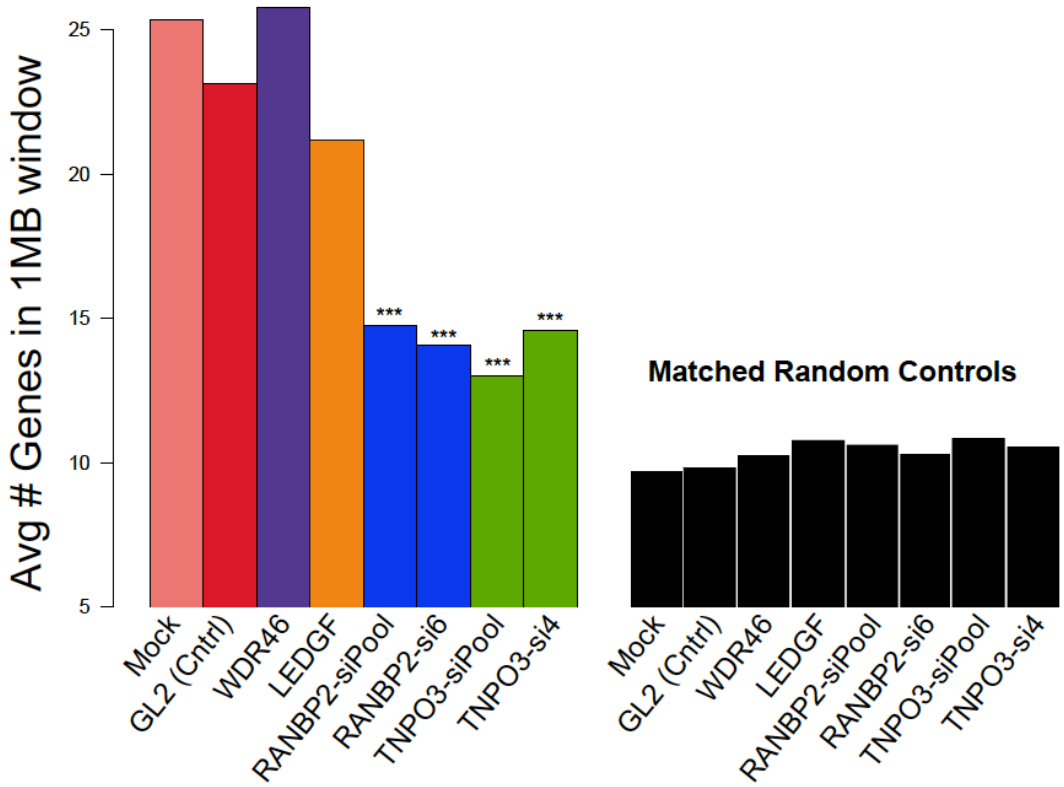
HIVmGag compared to HIV

- Less Gene dense
- Less GC rich
- Farther from CpG islands

HIV Gag is a viral determinant of HIV integration in gene dense regions

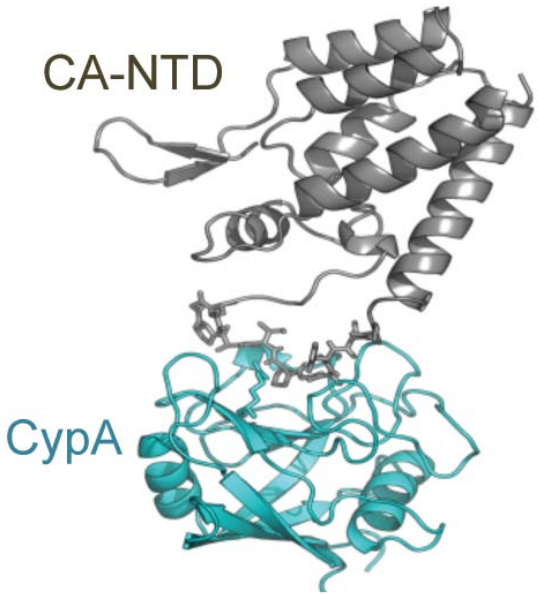
Gene density at integration sites

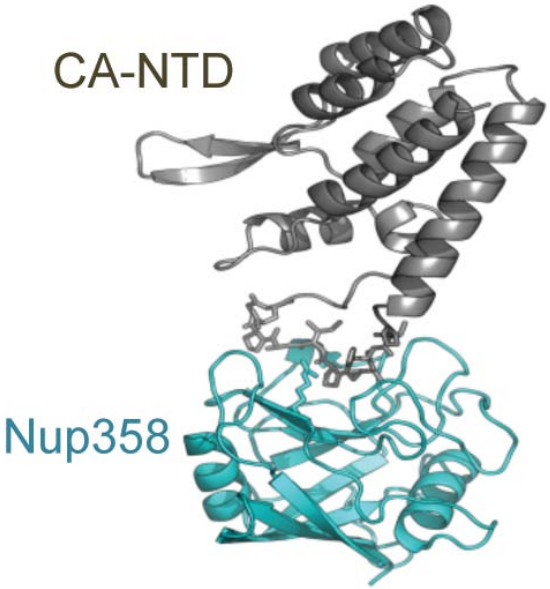
RANBP2 and TNPO3 knockdowns reduce gene density at integration sites

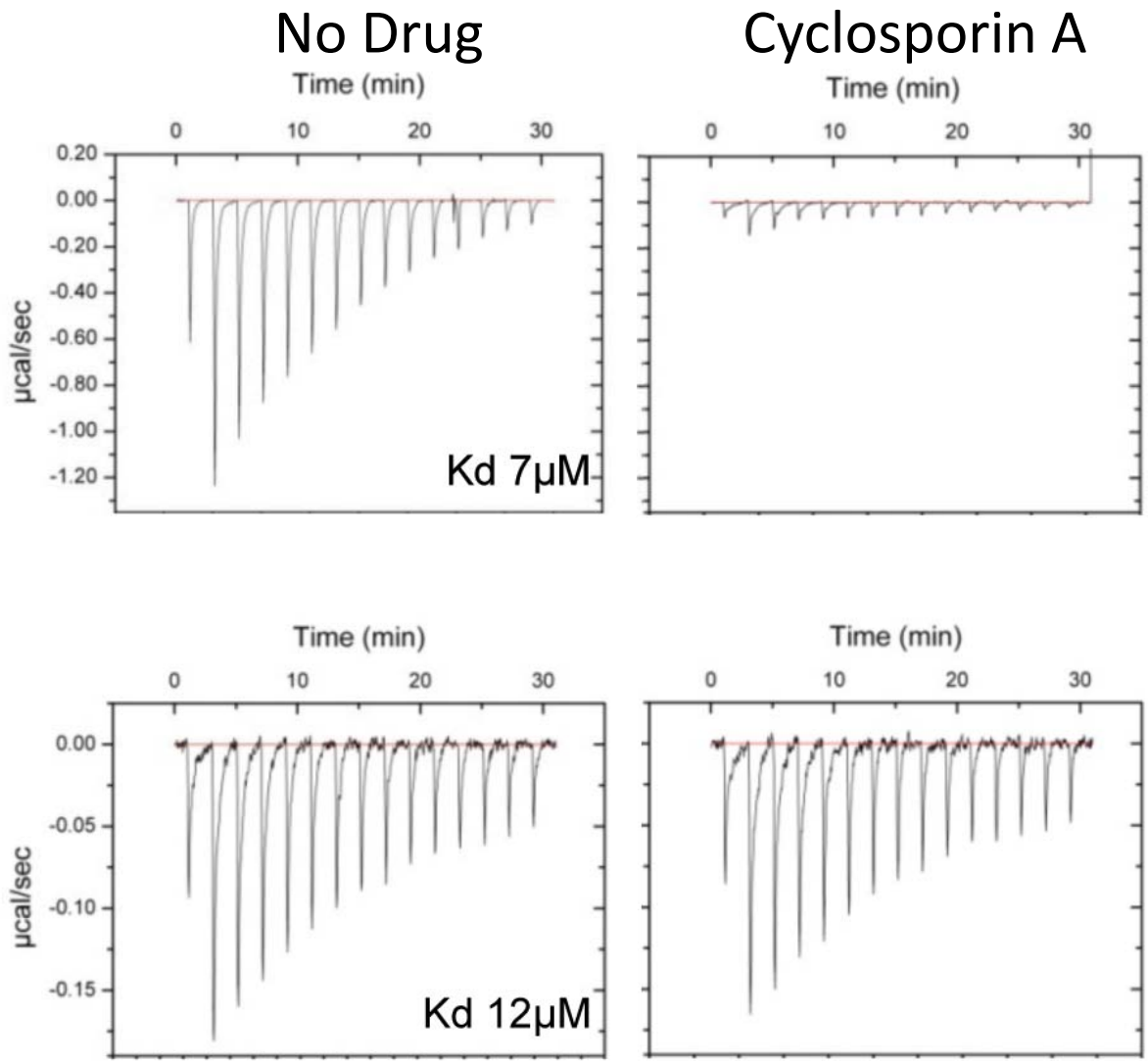


Integration in less gene dense regions in TNPO3 and RANBP2 knockdows

CA binds RanBP2 (NUP358)







Data from Leo James, Greg Towers and coworkers

- **Introduction**

- **Host factors in early steps: role of LEDGF/p75 in integration targeting**

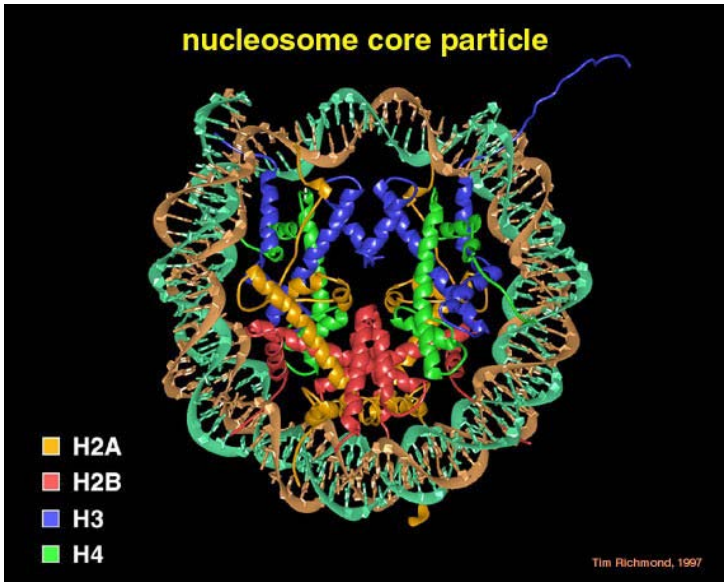
- **Host factors in early steps: trafficking to the nucleus**

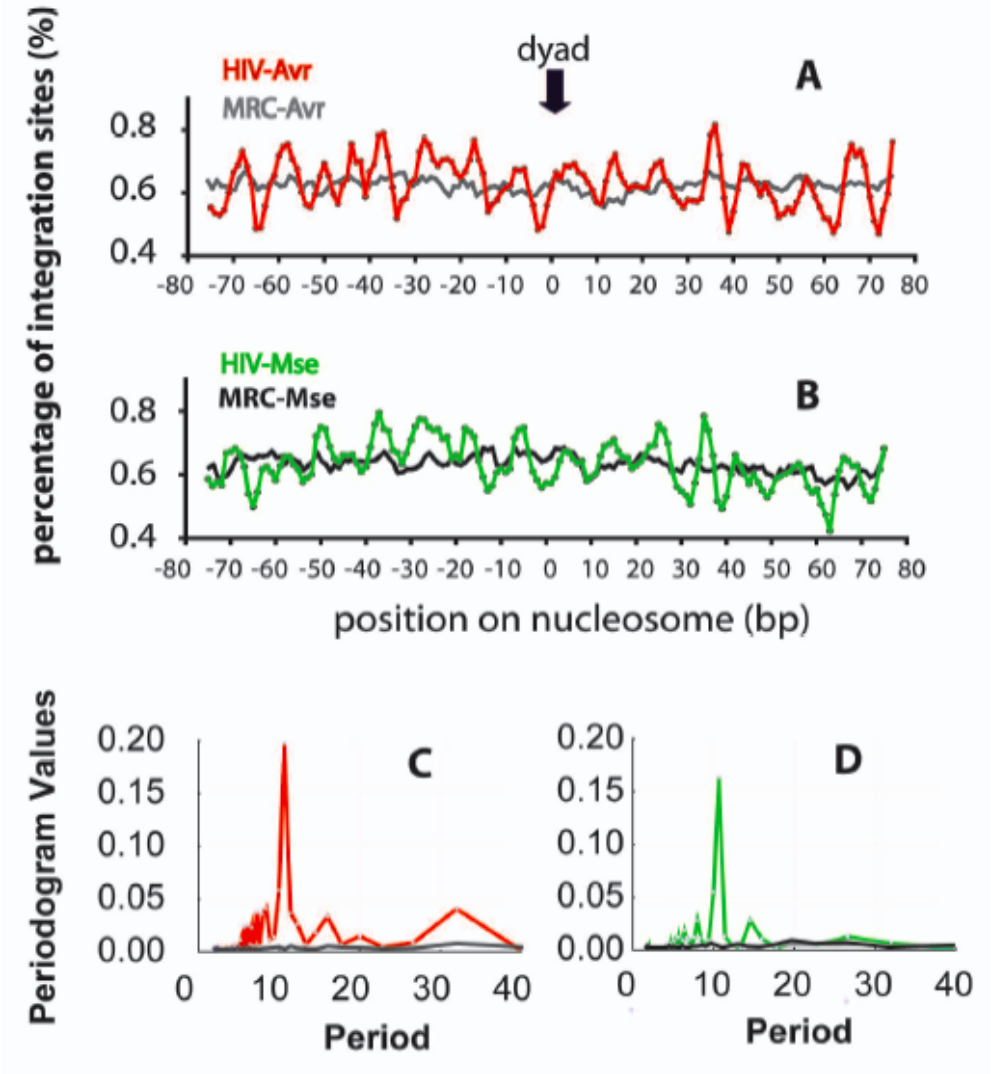
- **Host factors in early steps: nucleosomes**

- **Clonal expansion during lentiviral gene therapy for beta-thalassemia**

Periodic integration pattern relative to nucleosome dyad

"A second genetic code" for nucleosome positioning. Segal et al. Nature 2006, 442, 772-778.





40,000 HIV integration sites *(Wang et al., Genome Res. 2007)*
Periodicity matches that of the DNA helix
Distribution indicative of integration in nucleosomal targets in chromosomes

- **Introduction**

- **Host factors in early steps: role of LEDGF/p75 in integration targeting**

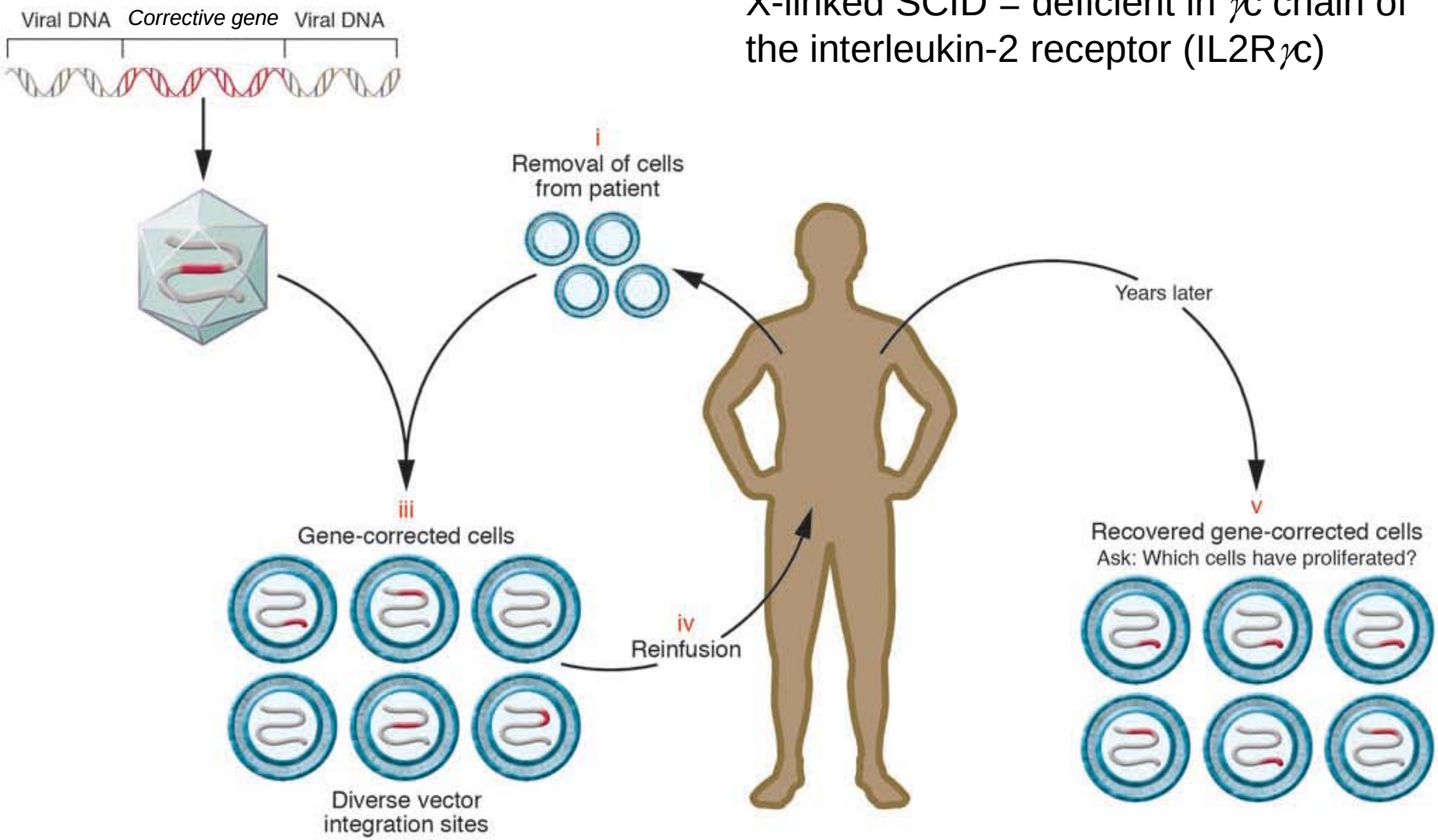
- **Host factors in early steps: trafficking to the nucleus**

- **Host factors in early steps: nucleosomes**

- **Clonal expansion during lentiviral gene therapy for beta-thalassemia**

Gene therapy to treat SCID-X1

X-linked SCID = deficient in γc chain of the interleukin-2 receptor (IL2Rγc)



Cavazzana-Calvo et al., Science 2000

- **Introduction**

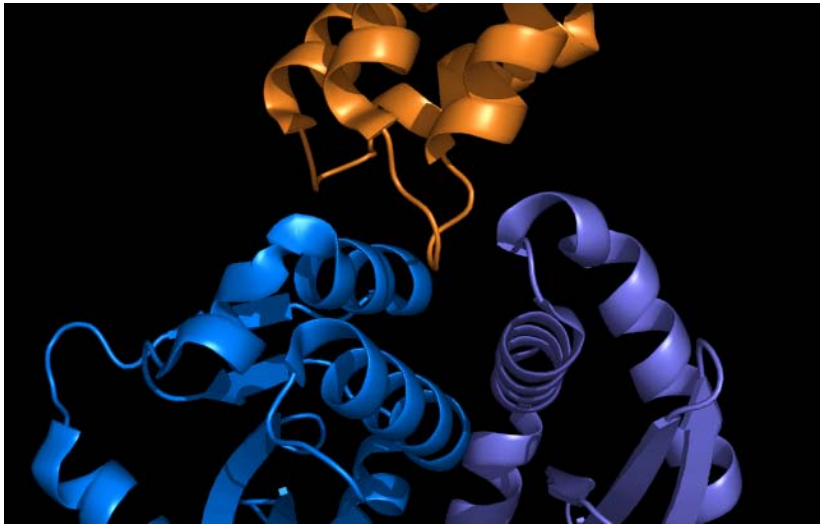
- **Host factors in early steps: role of LEDGF/p75 in integration targeting**

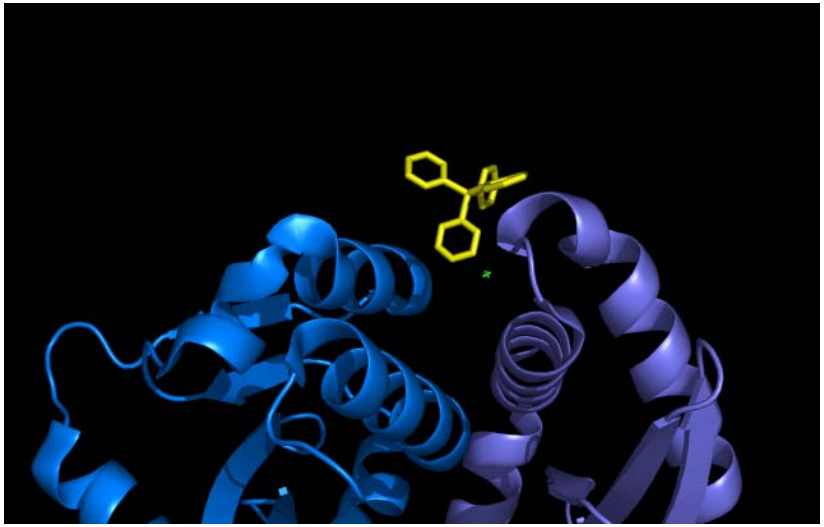
- **Host factors in early steps: trafficking to the nucleus**

- **Host factors in early steps: nucleosomes**

- **Clonal expansion during lentiviral gene therapy for beta-thalassemia**

Small molecule binding at the LEDGF/p75 interaction site





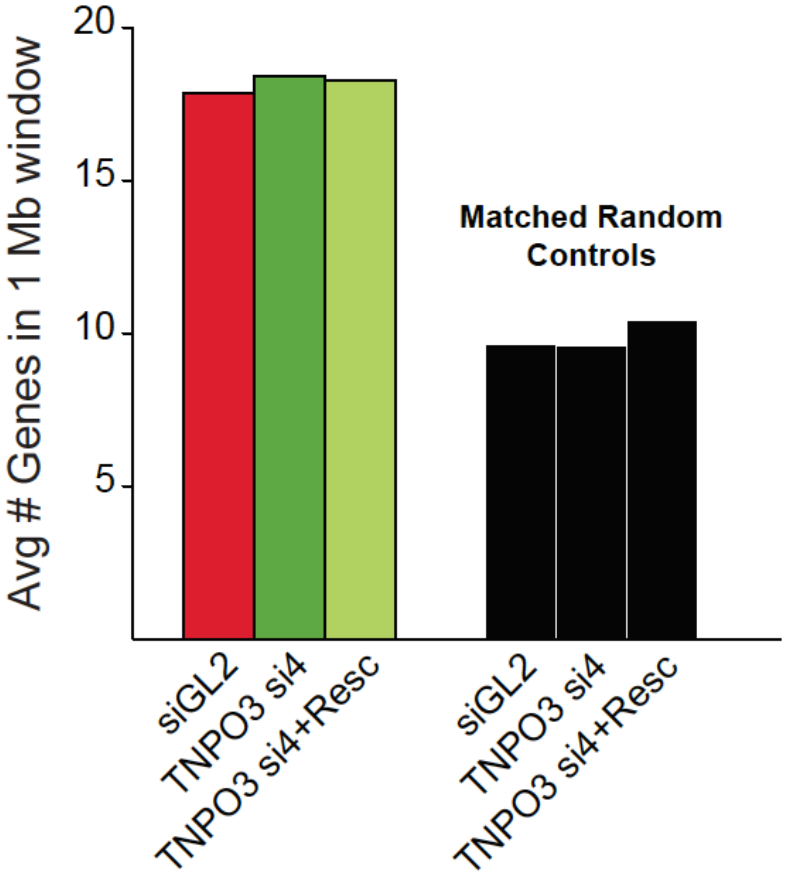
PSIP1/LEDGF/p75 binds to the dimer interface of the IN catalytic domain.
Cherepanov et al., 2005

A small molecule (TPA) binds to this site--possible “drugable” target in addition to active site?
Greenwald et al., 2001

MLV Integration in knockdowns

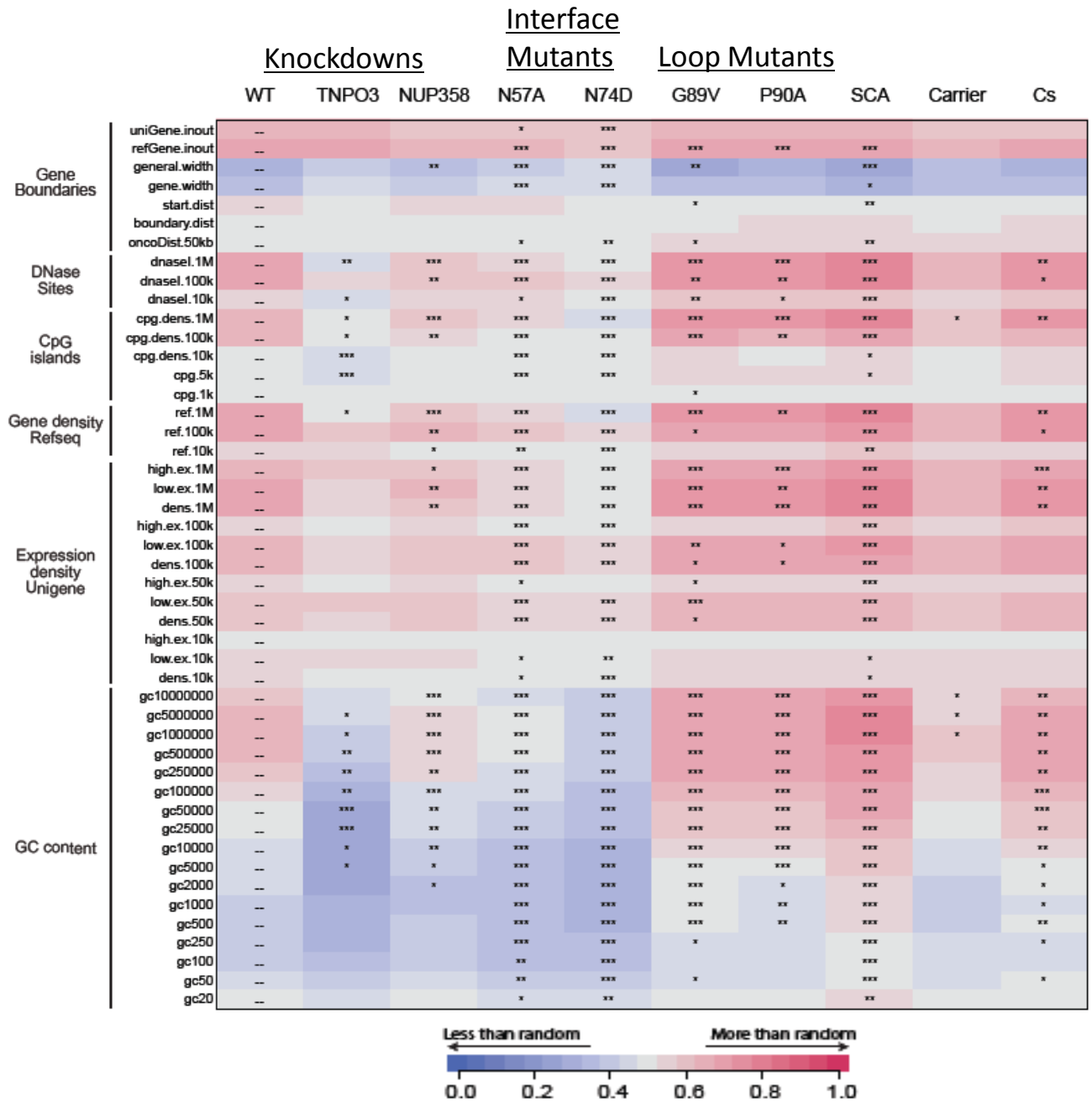
- MLV does not access nucleus in arrested cells
  Roe et al. 1993, Lewis et al. 1994

- MLV infection does not require TNPO3 or RANBP2
  Konig et al 08, Krishnan et al. 2010



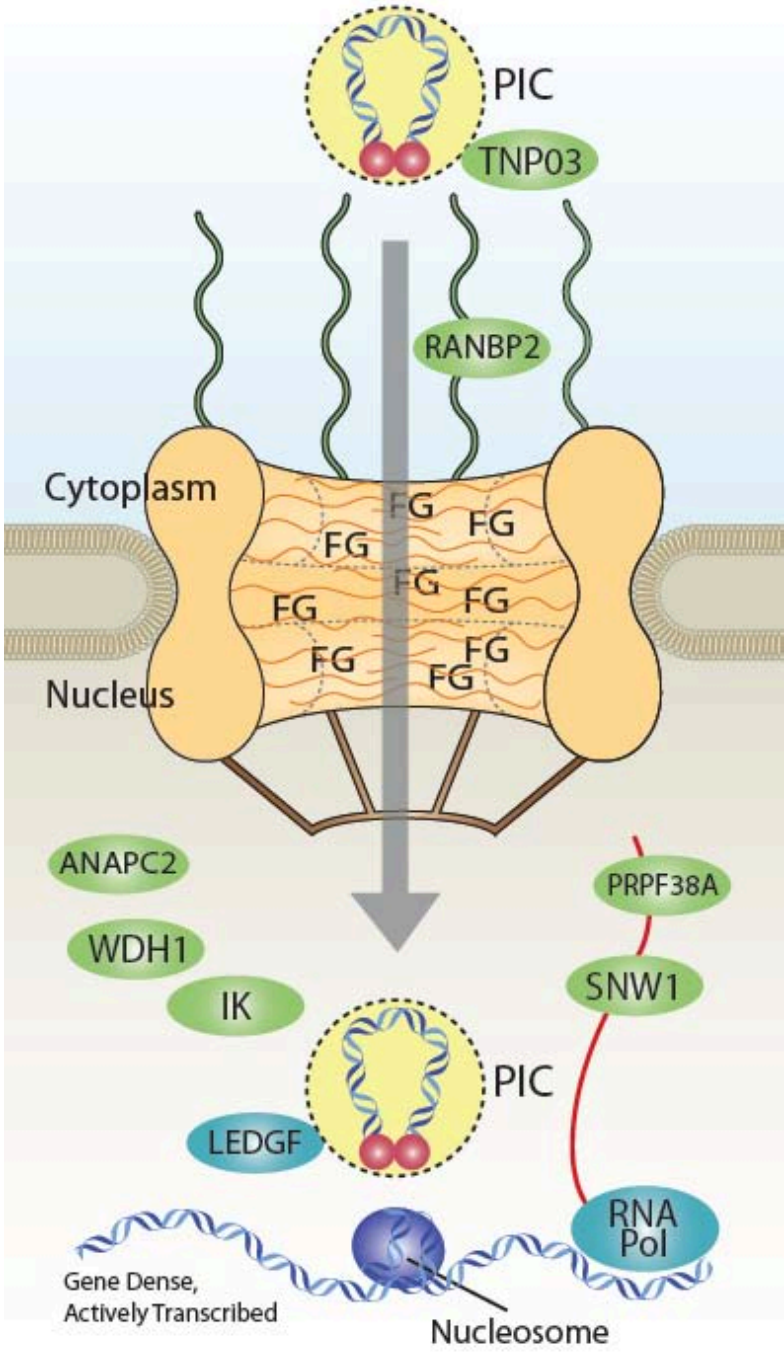
Gene density at MLV integration sites is unchanged in TNPO3 knockdowns

NUP358-Capsid Interactions in HIV Integration site selection

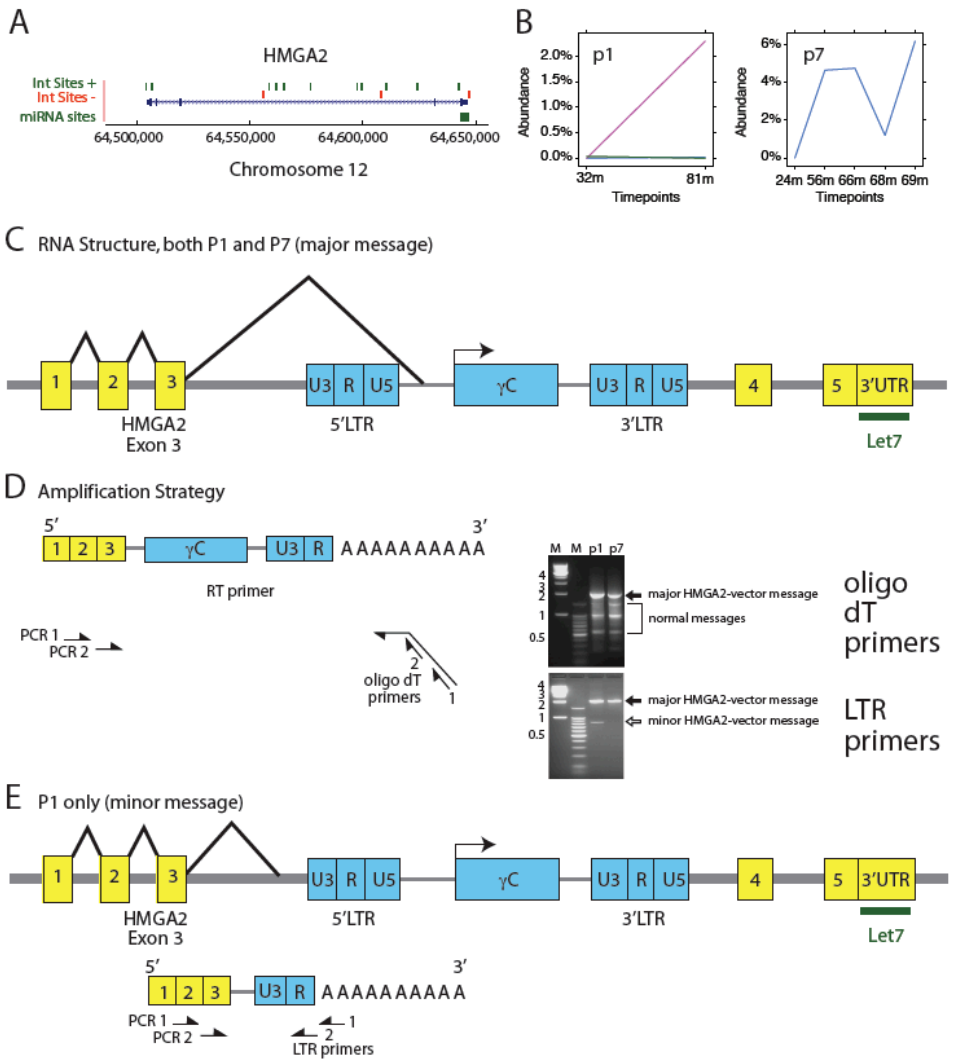


Summary

- Depletion of TNPO3 and RANBP2 by siRNA results in HIV integration in less gene dense regions
- HIV Gag is a viral determinant of integration in gene dense regions



SCID-X1 gene therapy and HMGA2



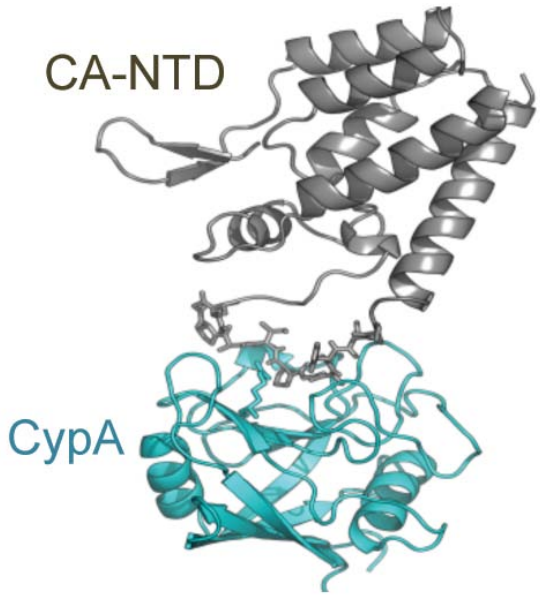
Loose cluster of integration sites in 3rd intron

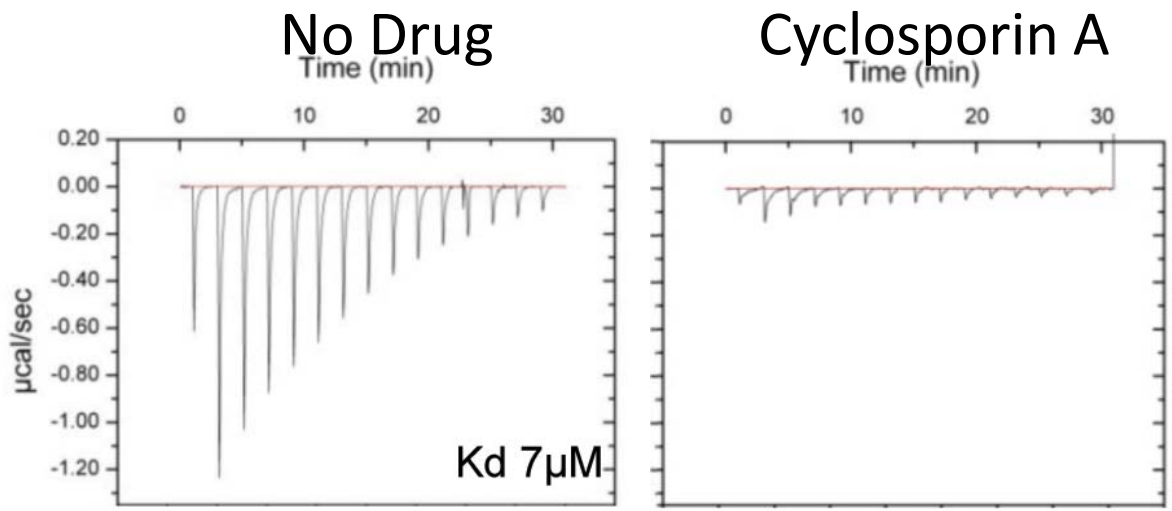
Two patients show clonal expansion

HMGA2 mRNA terminates prematurely in vector sequences

Wang et al., Blood 2010

Background: CA interacts with CypA

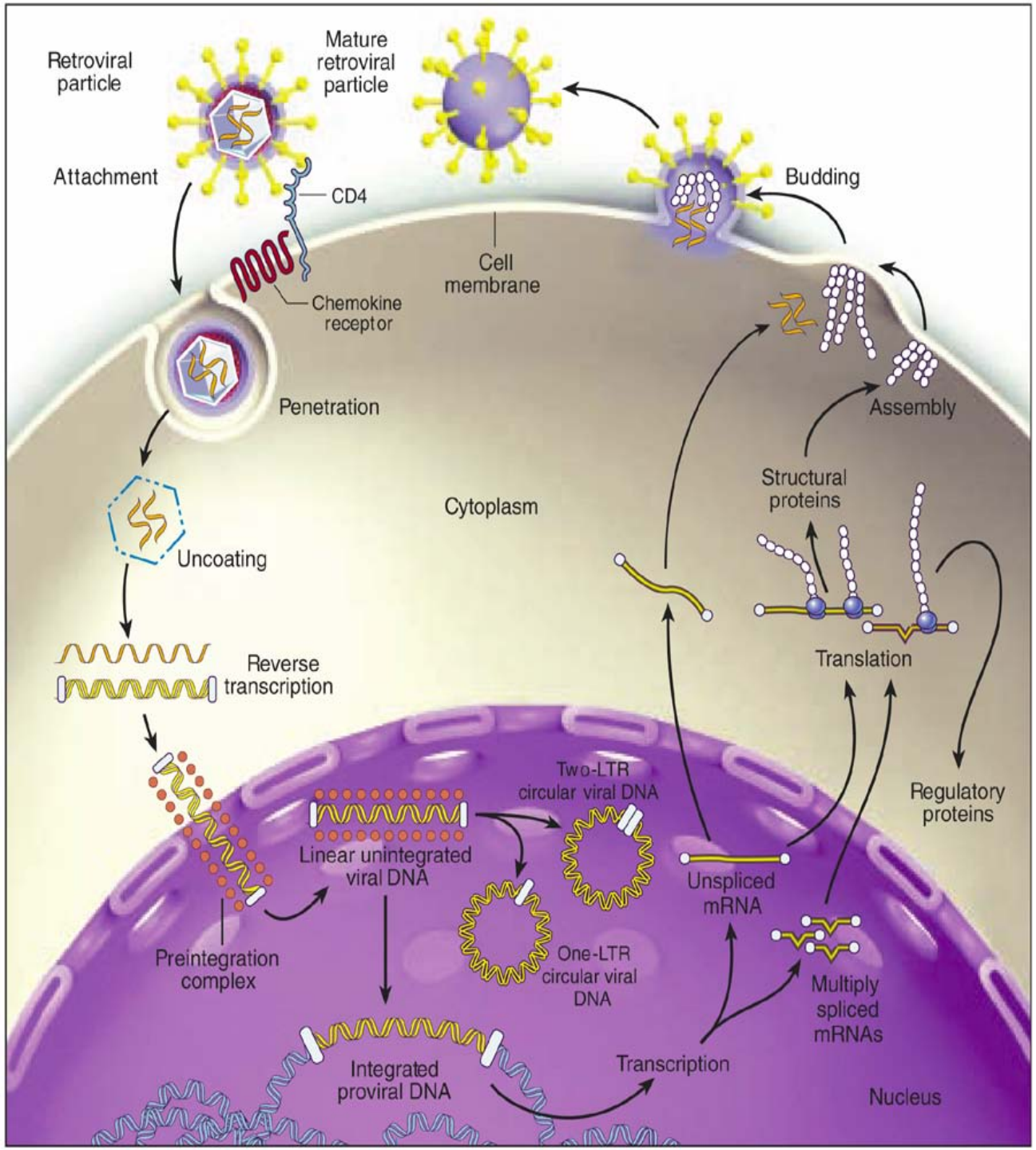




Data from Greg Towers

<u>Cyclophilin A (CypA)</u>

- Binds CA (packaged in virions)
- Binding inhibited by cyclosporine (Cs, CSA)
  - Antiviral in some cell types
- Cores depleted of CypA show signs of instability *Briones et al 2010*
- Mutants dependent on Cs are blocked in arrested cells between nuclear import and integration *Qi et al., 2008*


Retroviral particle
Mature retroviral particle
Attachment
CD4
Chemokine receptor
Cell membrane
Budding
Penetration
Assembly
Structural proteins
Cytoplasm
Uncoating
Translation
Reverse transcription
Two-LTR circular viral DNA
Regulatory proteins
Linear unintegrated viral DNA
Unspliced mRNA
Preintegration complex
One-LTR circular viral DNA
Multiply spliced mRNAs
Integrated proviral DNA
Transcription
Nucleus

- HIV DNA integration introduction
- Viral and host cell factors directing HIV integration targeting
- Integration targeting in human gene therapy

First-in-humans use of lentiviral vectors

Sustained Gene Transfer in Humans Using Lentiviral Vectors.

B. Levine, L. M. Humeau, J. Boyer, R.-R. MacGregor, T. Rebello, X. Lu, G. K. Binder, V. Slepushkin, J. R. Mascola, F. D. Bushman, B. Dropulic, and C. H. June. PNAS 2006

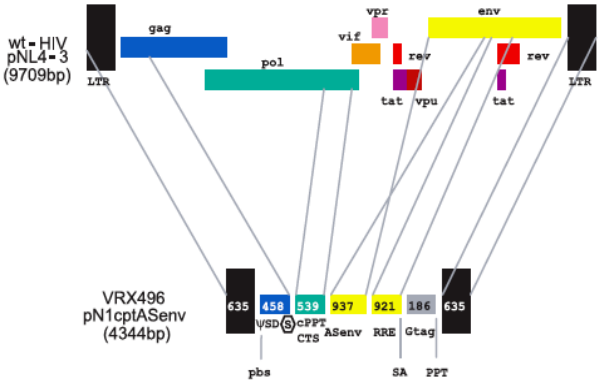
Analysis of Lentiviral Vector Integration in HIV+ Study Subjects Receiving Autologous Infusions of Gene Modified CD4+ T Cells

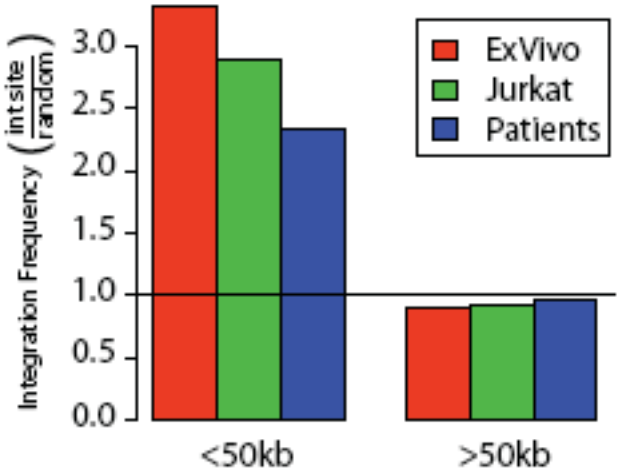
Gary P. Wang, Bruce L. Levine, Gwendolyn K. Binder, Charles C. Berry , Nirav Malani, Gary McGarrity, Pablo Tebas, Carl H. June, Frederic D. Bushman
Mol Ther 2009

Lentiviral vectors to treat HIV: first in humans

Env antisense introduced into mature T-cells

Vector





Proto-oncogene 5' ends

Recovered sites from patients at several times after infection; no sign of increased frequency of sites near oncogenes after transplantation

Levine et al., 2006
Wang et al., Mol. Ther. 2009

Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy.

Cartier N, Hacein-Bey-Abina S, Bartholomae CC, Veres G, Schmidt M, Kutschera I, Vidaud M, Abel U, Dal-Cortivo L, Caccavelli L, Mahlaoui N, Kiermer V, Mittelstaedt D, Bellesme C, Lahlou N, Lefrère F, Blanche S, Audit M, Payen E, Leboulch P, l'Homme B, Bougnères P, Von Kalle C, Fischer A, Cavazzana-Calvo M, Aubourg P.

Science. 2009 Nov 6;326(5954):818-23.

Credits

Bushman laboratory	Former Bushman Lab	Collaborators
Troy Brady	**Rick Mitchell**	Charles Berry
Kyle Bittinger	Brett Beitzel	Sumit Chanda
Rohini Sinha	Mary Lewinski	John Young
Keshet Ronen	Astrid Schroder	Renate Koenig
Karen Ocwieja	Angela Ciuffi	Joe Ecker
Christian Hoffmann	Heather Marshall Rose	Craig Hyde
Nirav Malani	Jeremy Leipzig	Mark Yeager
Rithun Mukherjee	Matt Culyba	Kushol Gupta
Young Hwang	Gary Wang	Greg Van Duyne
Rebecca Custers-Allen	Tracy Diamond	Masahiro Yamashita
Serena Dollive	Scott Sherrill-Mix	Mike Emerman
Jennifer Hwang	Nana Minkah	Francis Collins
Jimmy Hu		Nam Lee
Sam Minot		Paul Bieniasz
Greg Peterfreund		Philippe Leboulch
Shannah Roth		Alain Fischer
Frances Male		Marina Cavazzana-Calvo
Emily Charlson		Zeger Debyser
	Rik Gijsbers	
		Pablo Tebas
		Mike Miller

Retroviral DNA Integration

- *Target for therapy*
- *Model protein-DNA "machine"*
- *Basis of genomic evolution*
- *Central to gene therapy*

TABLE 10.2. Number of Copies and Fraction of Human Genome for Classes of Interspersed Repeat

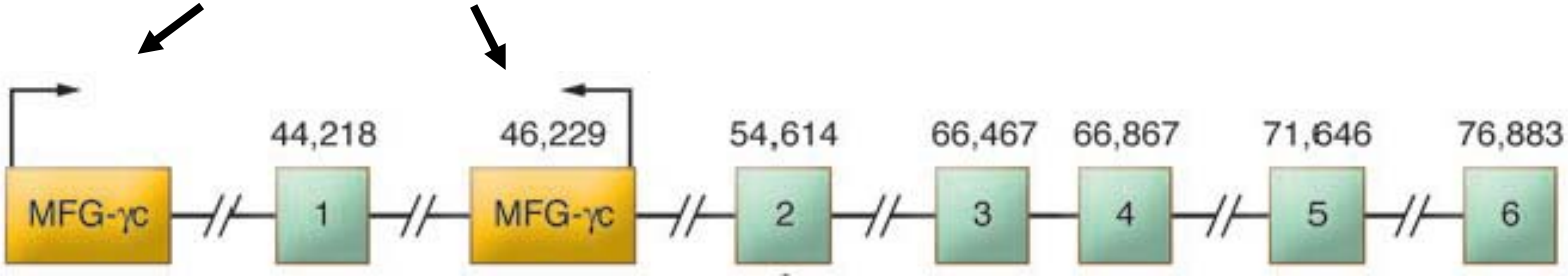
	Number of copies (x1,000)	Total number of bases in the draft genome sequence (Mb)	Fraction of the draft genome sequence (%)	Number of families (subfamilies)
SINEs	1,558	359.6	13.14	
Alu	1,090	290.1	10.60	1(~20)
Mir	393	60.1	2.20	1(1)
Mir3	75	9.3	0.34	1(1)
LINEs	868	558.8	20.42	
LINE1	516	462.1	16.89	1(~55)
LINE2	315	88.2	3.22	1(2)
LINE3	37	8.4	0.31	1(2)
LTR elements (proviruses)	443	227.0	8.29	
HERV-class I	112	79.2	2.89	72(132)
HERV (K)-class II	8	8.5	0.31	10(20)
HERV (L)-class III	83	39.5	1.44	21(42)
MaLR	240	99.8	3.65	1(31)
DNA elements	294	77.6	2.84	
hAT group				
MER1-Charlie	182	38.1	1.39	25(50)
Zaphod	13	4.3	0.16	4(10)
Tc-1 group				
MER2-Tigger	57	28.0	1.02	12(28)
Tc2	4	0.9	0.03	1(5)
Mariner	14	2.6	0.10	4(5)
PiggyBac-like	2	0.5	0.02	10(20)
Unclassified	22	3.2	0.12	7(7)
Unclassified	3	3.8	0.14	3(4)
Total interspersed repeats		1,226.8	44.83	

(Reprinted, with permission, from Lander et al. 2001 [© Macmillan Magazines Ltd.]. [29])

Insertional activation in adverse events

20 patients treated in Paris and London
5 adverse events

vector integration sites



P5 P4

Also chromomsomal abberations present in blast cells

*Hacein-Bey-Abina (2003) Science **302**, 415-419.*
*Cavazzana-Calvo and Fischer (2007) J. Clin. Invest. **117** ,1456-65*

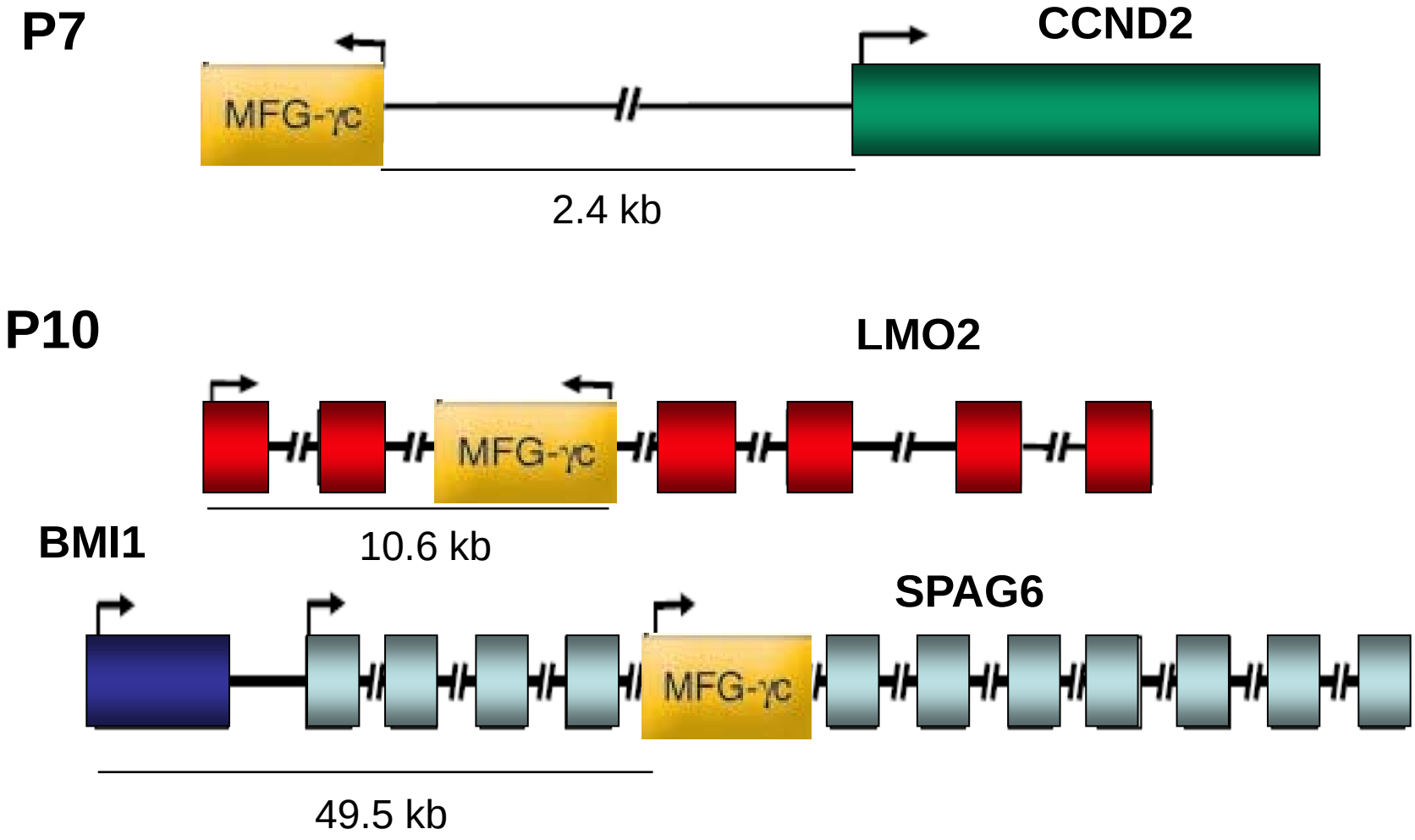
Pyrosequencing analysis of integration sites in SCID-X1 gene therapy

Longitudinal samples from 8 patients
Isolate sites by cleaving genome with restriction enzymes, ligate linkers, PCR
203,805 integration site sequence reads

Questions:

- What proto-oncogenes were involved in most recent two adverse events?
- What long term trends in population structure are detectable using these methods?
- What is the influence of chemotherapy on the population of transduced clones?

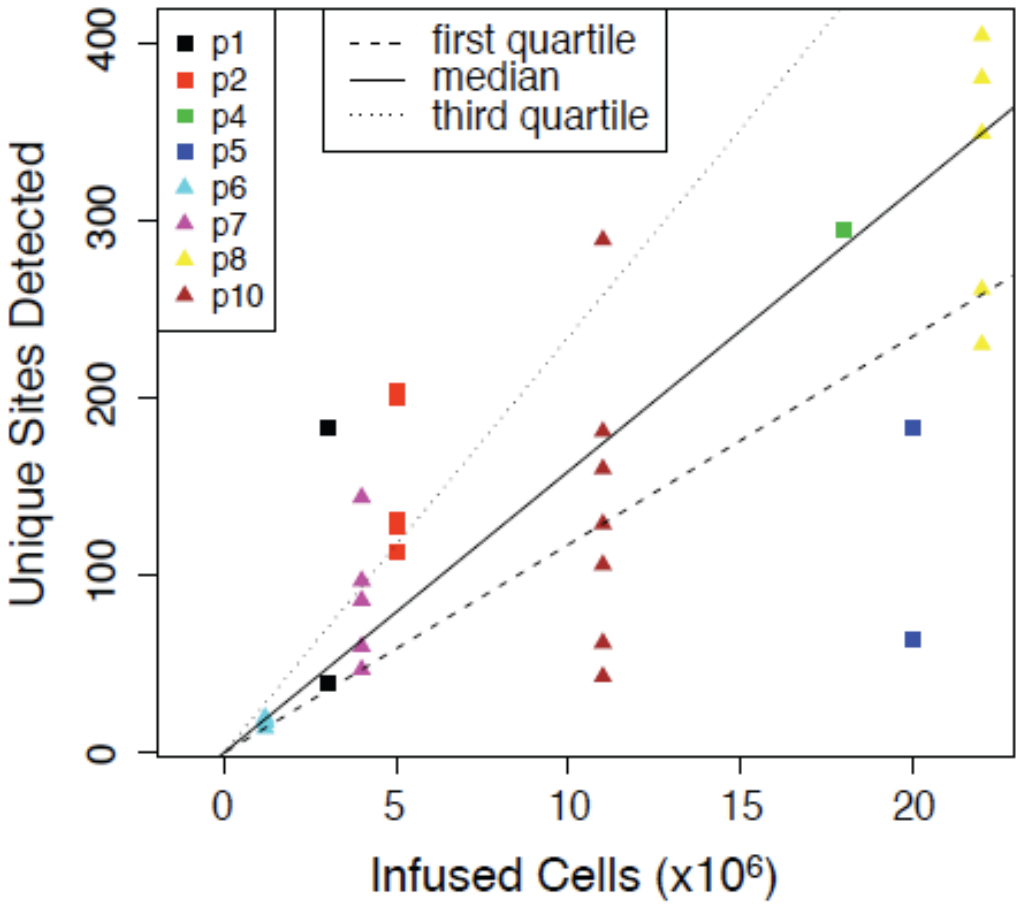
Genes at integration sites in blast cells



Transcription increased at CCND2, LMO2, and BMI1

Hacein-Bey et al., 2008 JCI

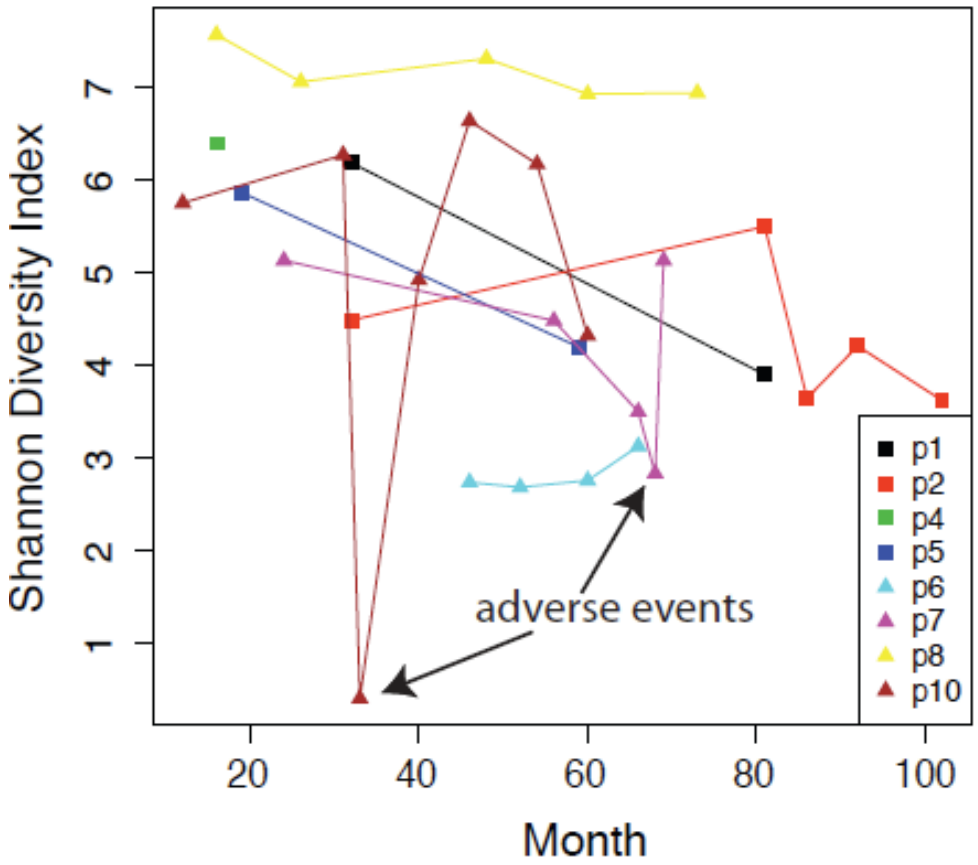
Relationship between cells infused and numbers of unique integration sites recovered



The more cells infused, the more unique sites

Wang et al., Blood 2010

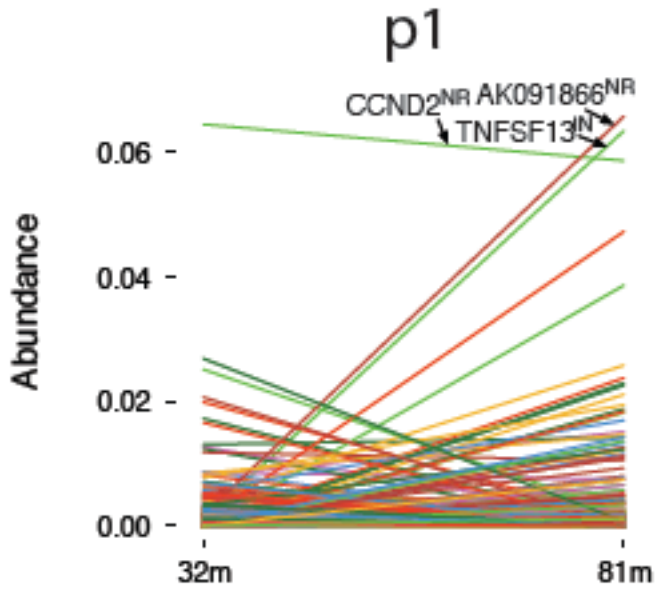
Longitudinal analysis of diversity

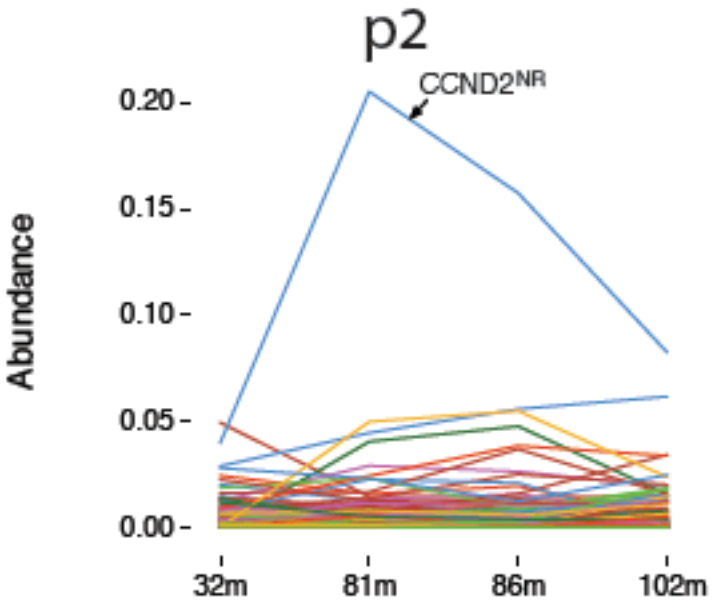


Measurable slow decline in diversity in some patients

Wang et al., Blood 2010

Clonal expansion without adverse events

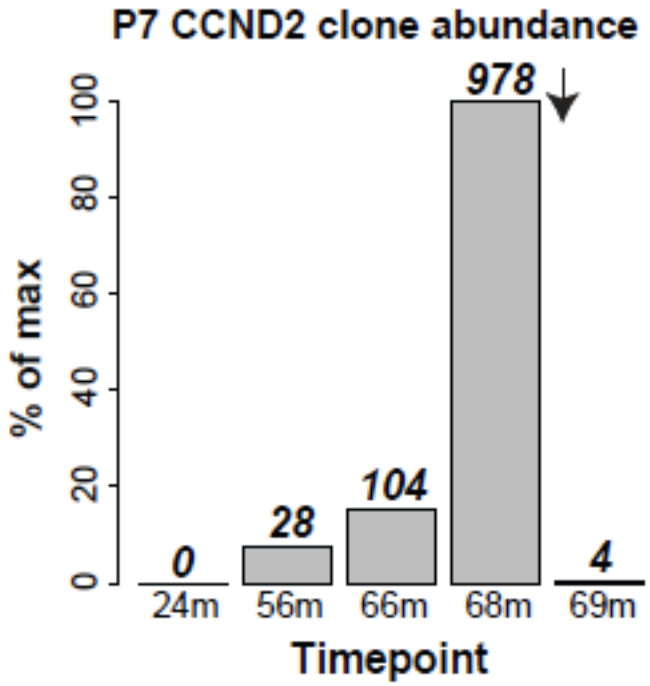


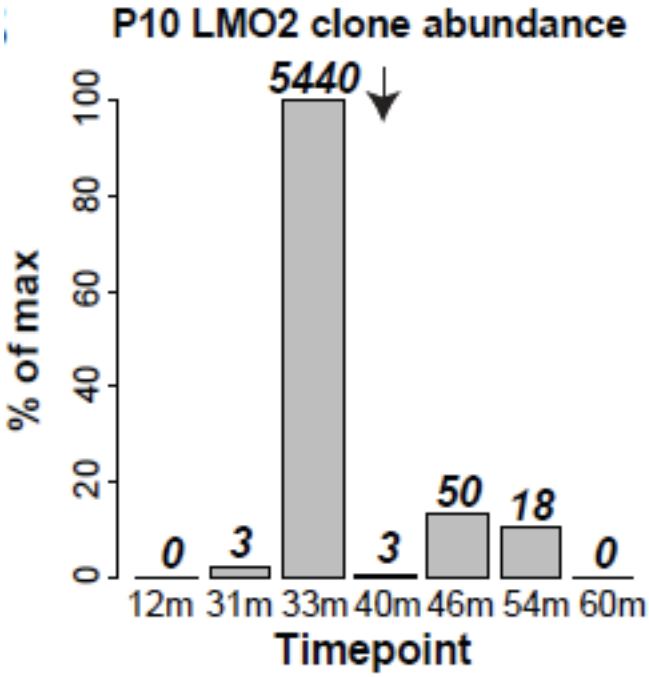


Cells harboring CCND2 reach high levels of abundance, but patients are healthy to date

Wang et al., Blood 2010

Monitoring the outcome of chemotherapy by tracking integration sites

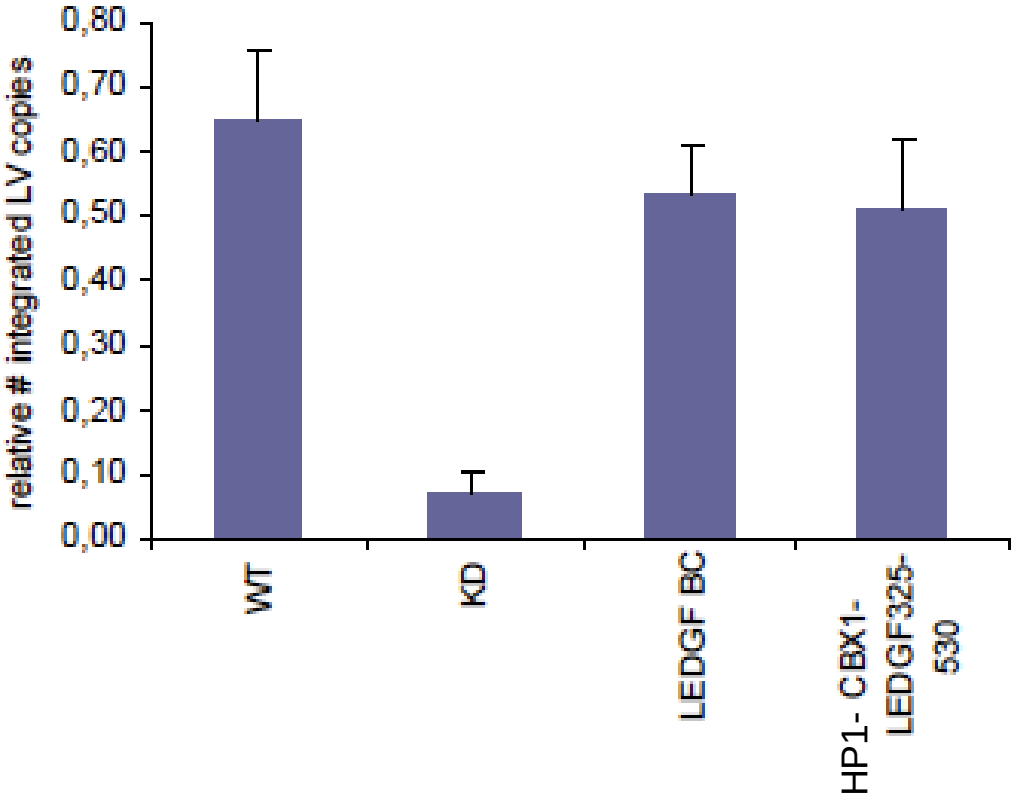




Chemotherapy so far appears to have eradicated malignant clone
Patients continue to benefit from gene therapy

Wang et al., Blood 2010

HP1(CBX)-LEDGF rescues lentivirus integration



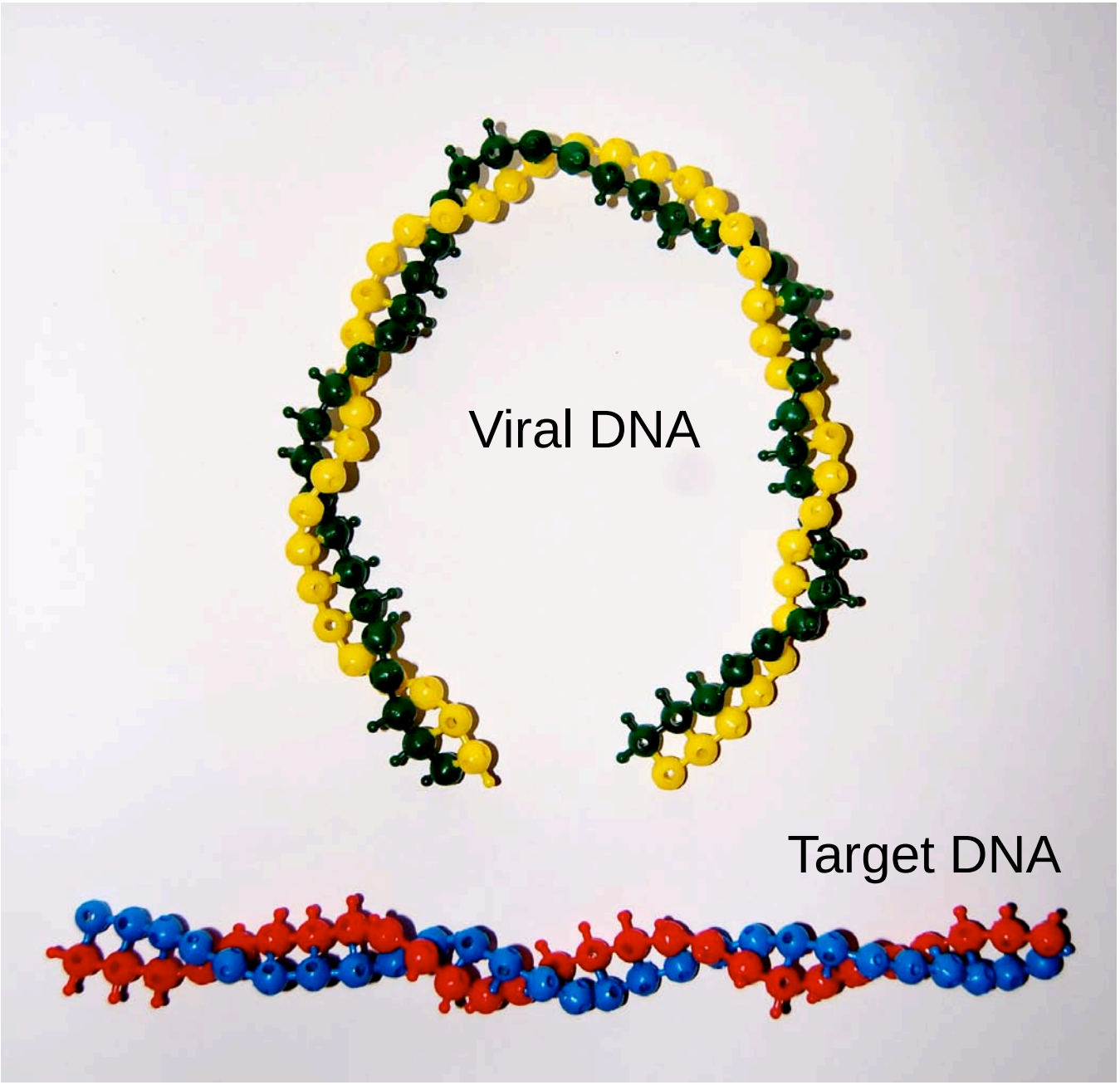
Summary

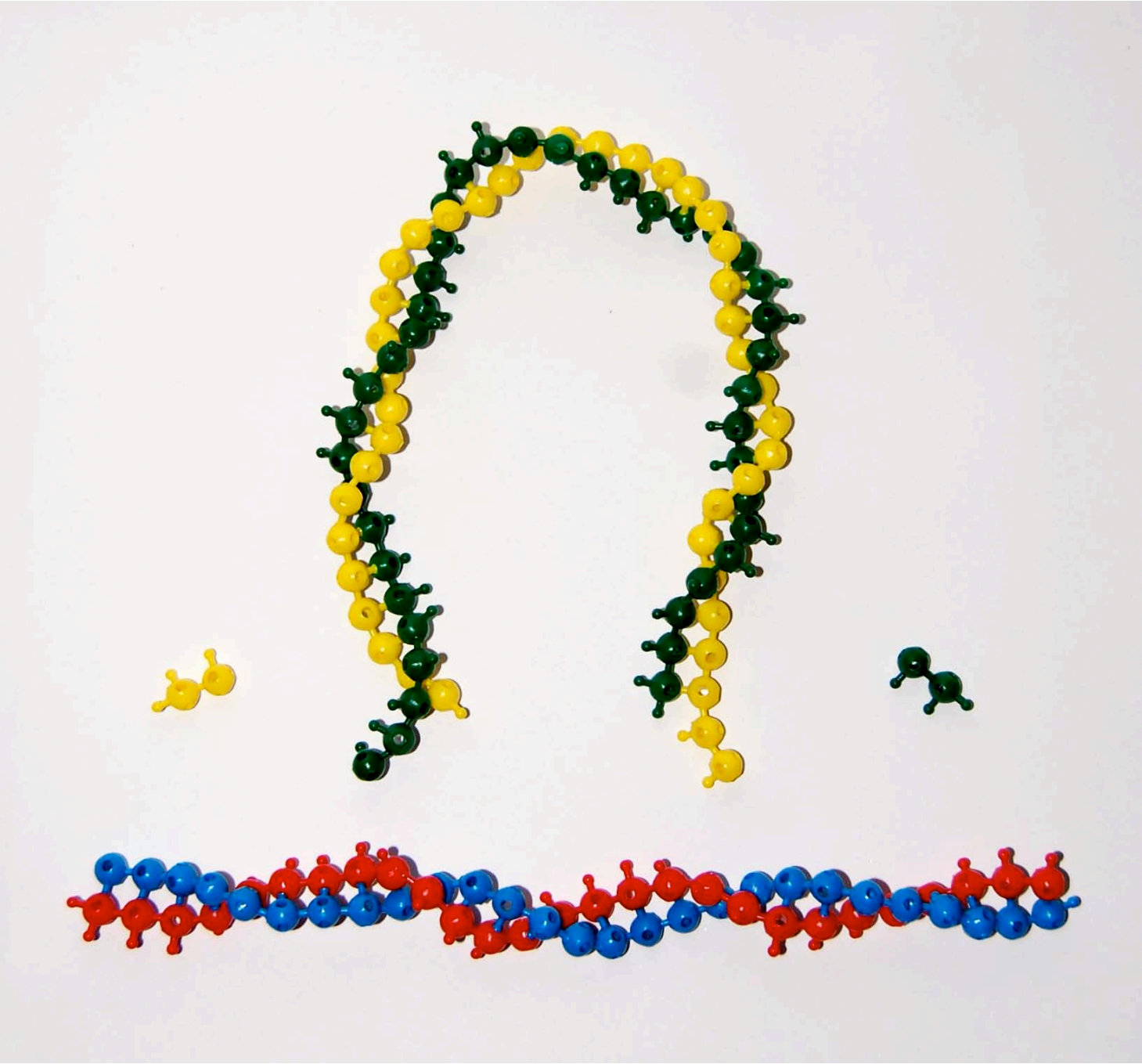
•SCID-X1 gene therapy (transduction of stem cells using gamma-retroviral vector) reversed immunodeficiency in most patients treated.

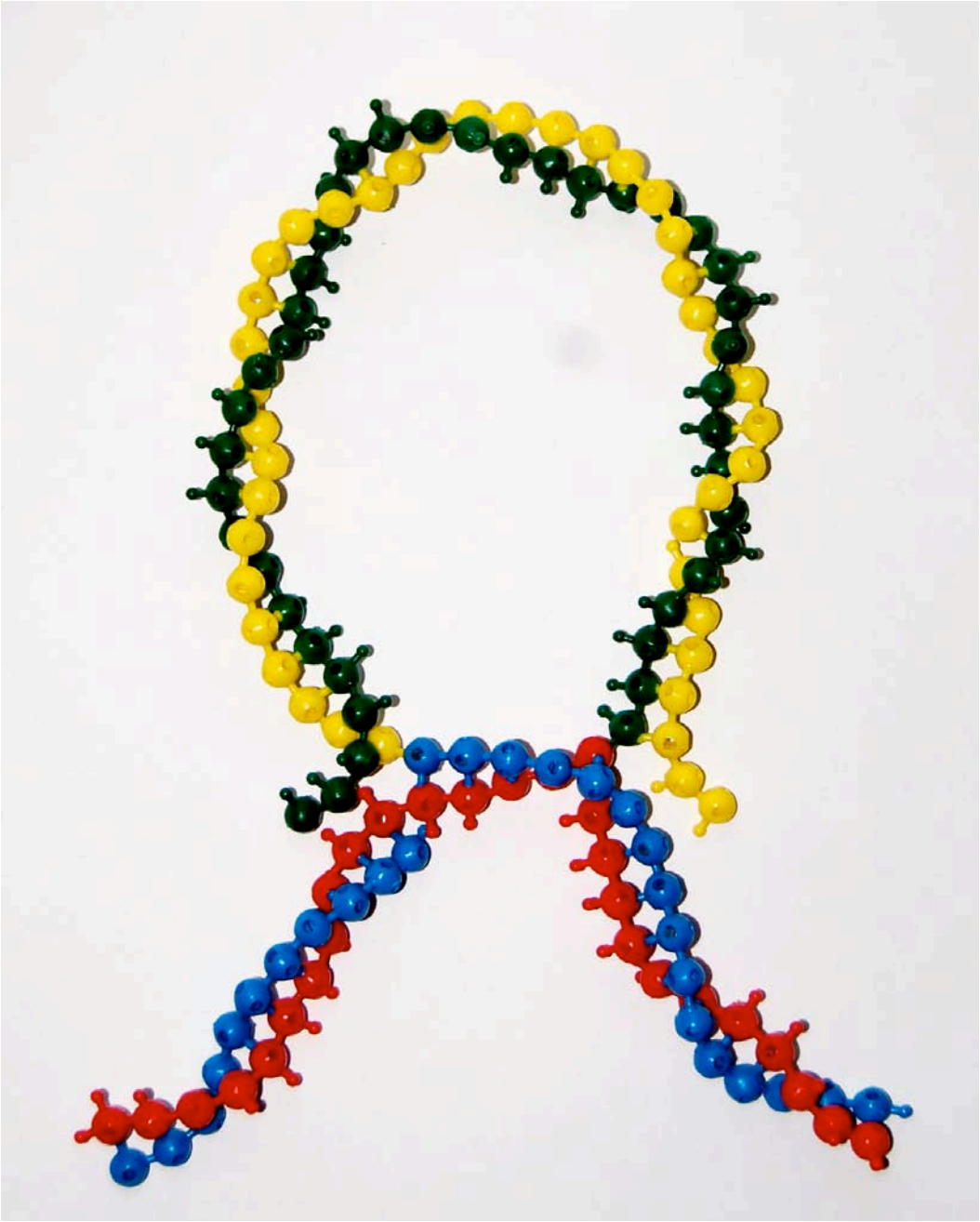
•Five adverse events out of twenty patients treated; insertional activation of LMO2, CCND2, and BMI1, also other molecular events.

•Evidence for longitudinal clonal expansion in some cells without adverse events associated with integration near proto-oncogenes.

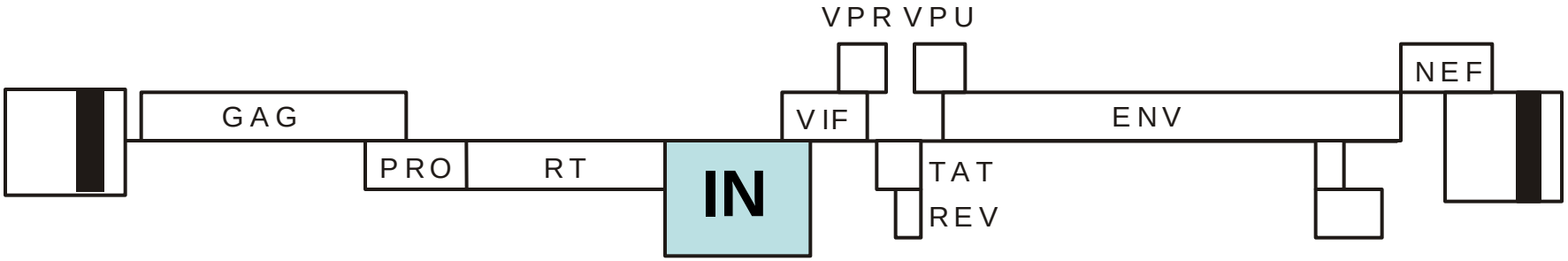
•Chemotherapy appears to have eradicated malignant cell clones.


Viral DNA
Target DNA

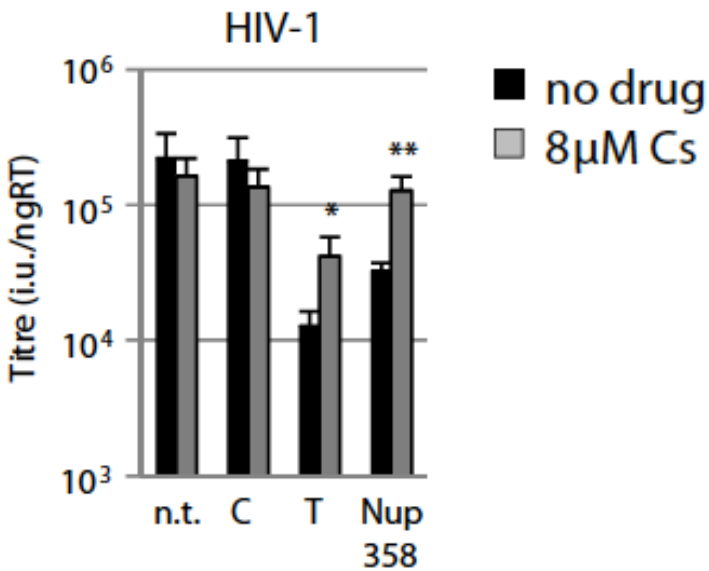


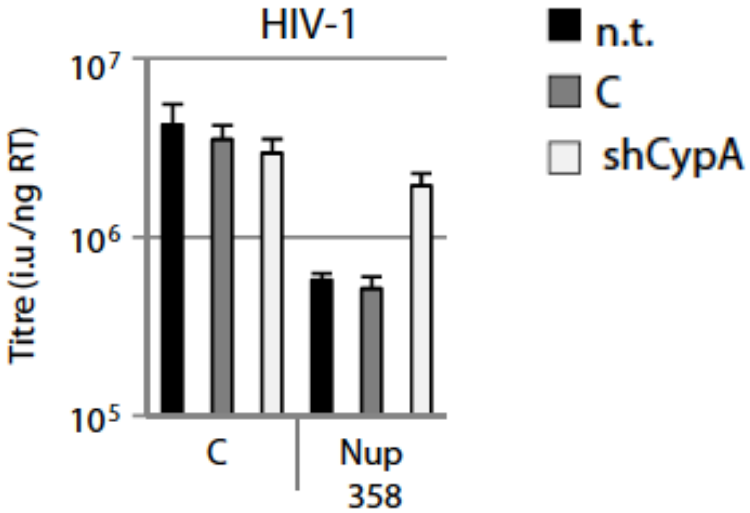


The integrase coding region



HIV infection in knockdowns + Cs

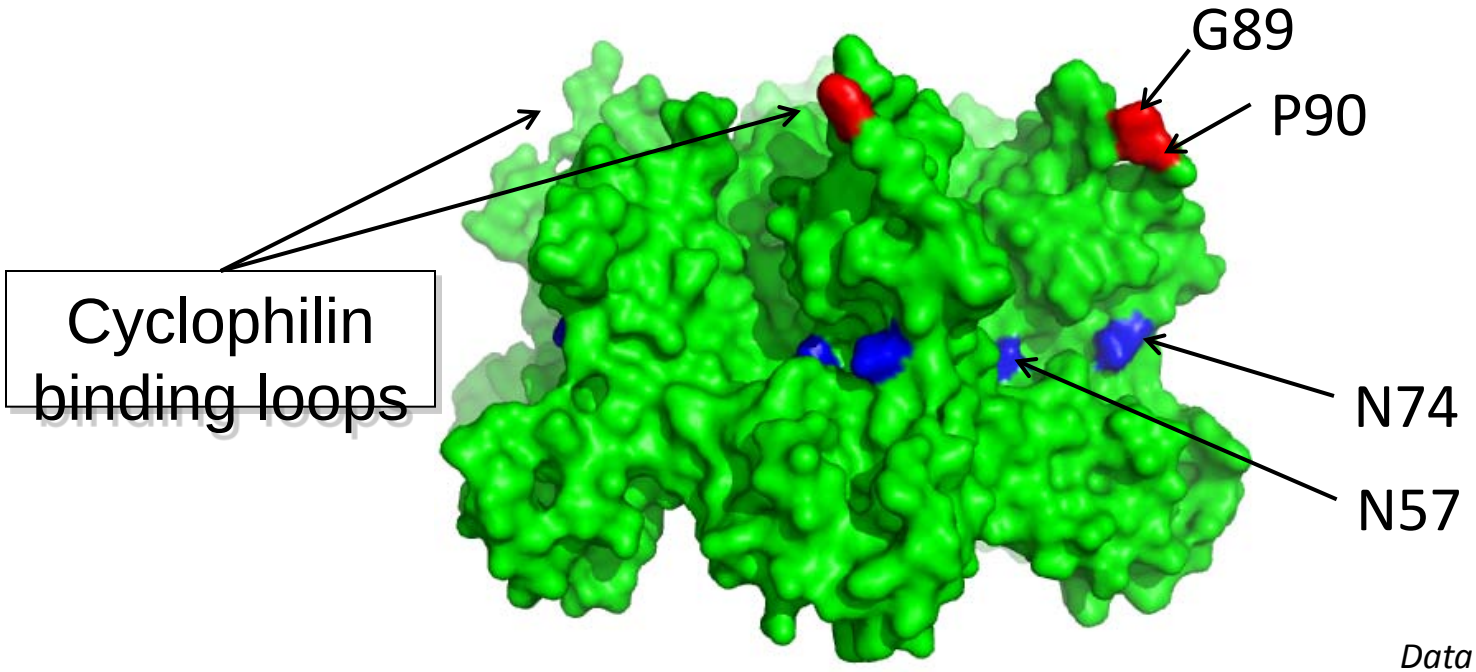




- Cs inhibits Cyclophilin binding but not RanBP2 binding
- Cs partially rescues infection in TNPO3 knockdowns
- Cs and shCypA rescue infection in RanBP2 knockdowns:
  - **RanBP2 is unnecessary when Cyclophilin can't bind HIV capsid**

Cyclophilin may stabilize capsid and RanBP2 may help remove it at the pore

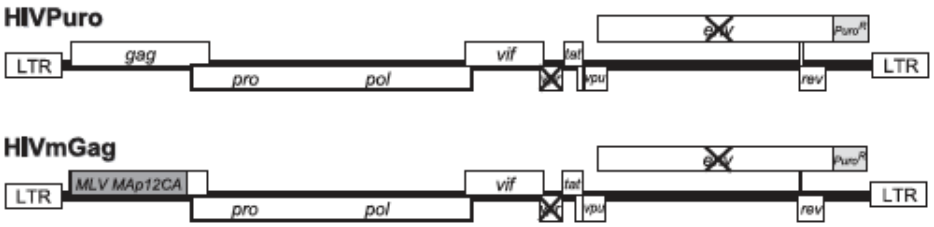
HIV CA Mutants



Data from Greg Towers

CA Mutant	Description	CypA Kd (uM)	Nup358 Kd (uM)	Sensitive to NUP358 KD?
WT		7	12	Y
N57A	monomer interface	7	55	N
N74D	hexamer interface	7	95	N
G89V	CypA binding loop	Nonbinding	Nonbinding	N
P90A	CypA binding loop	46	Nonbinding	N
SCA	SIV CA	Nonbinding	Nonbinding	N

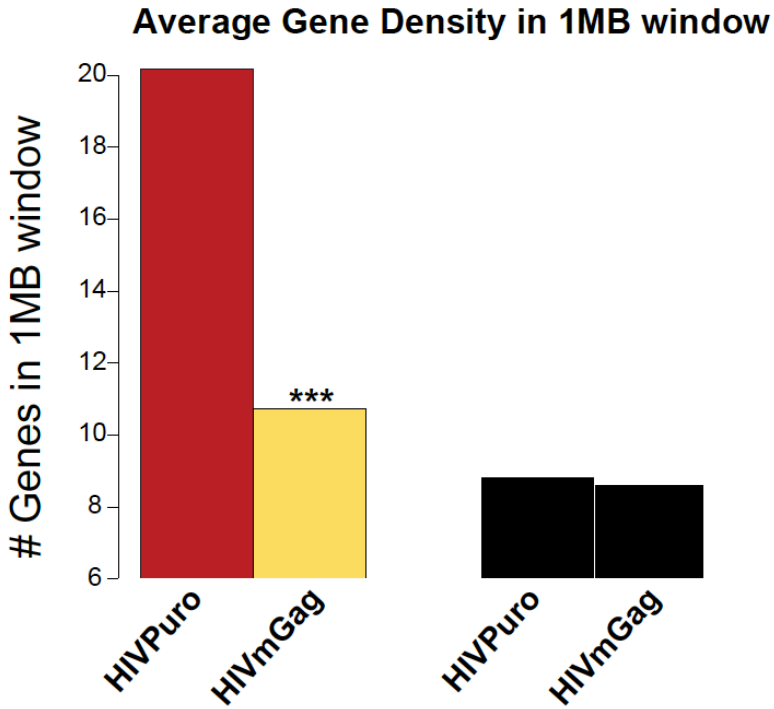
Integration of HIV with MLV Gag



Re-analysis of data from Lewinsky et al. 2006

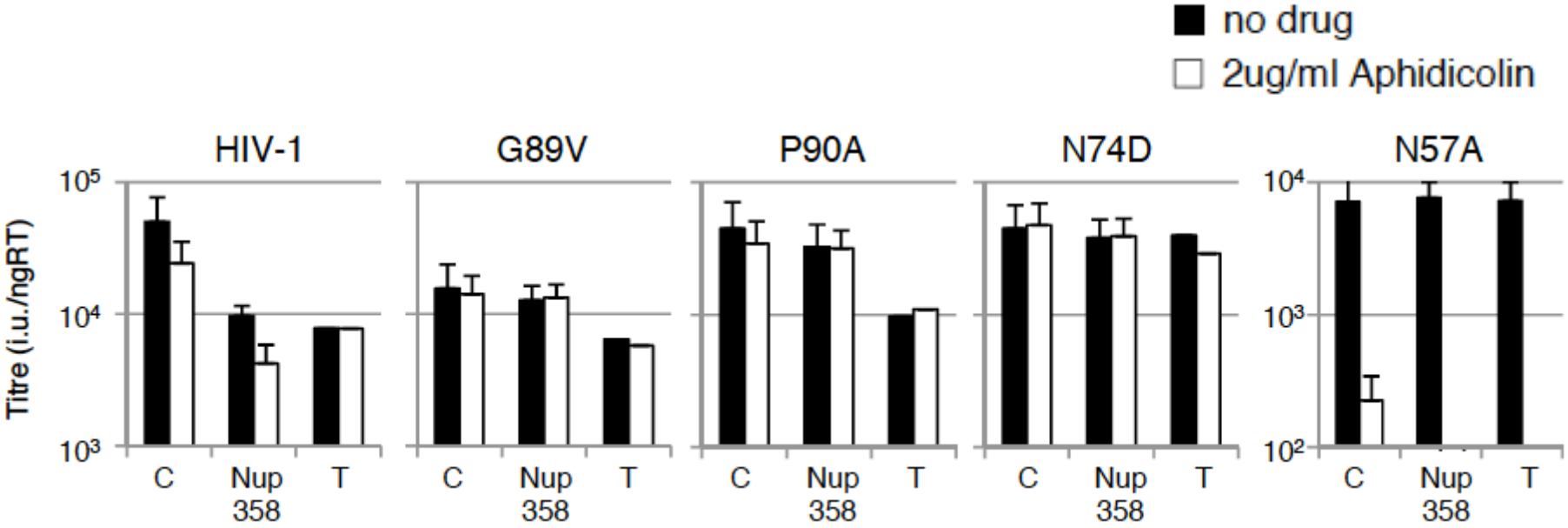
HIVmGag compared to HIV

- Less Gene dense
- Less GC rich
- Farther from CpG islanc



HIV Gag is a viral determinant of HIV integration in gene dense regions

Infection eficiency of Mutants



	G89V and P90A	N57A and N74D
Mutant Type	Loop Mutants	Interface mutants
Bind CypA?	No	Yes
Bind RanBP2?	No	No
RanBP2 kd	Independent	Independent
TNPO3 kd	Dependent	Independent

NUP358-Capsid Interactions in HIV Integration site selection

	G89V and P90A	N57A and N74D
Mutant Type	Loop Mutants	Interface mutants
Bind CypA?	No	Yes
Bind RanBP2?	No	No
RanBP2 kd	Independent	Independent
TNPO3 kd	Dependent	Independent
Gene Density at Proviruses	Increased	Decreased

Targeting mutants in Capsid

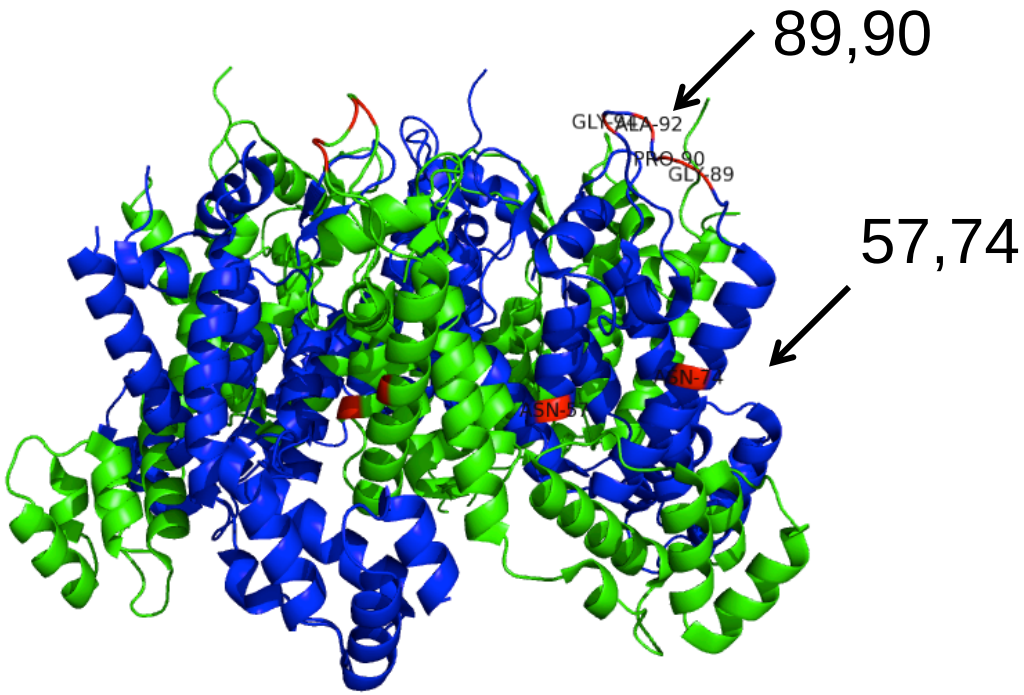
CA binds tightly to the cyclophillin domain of RanBP2

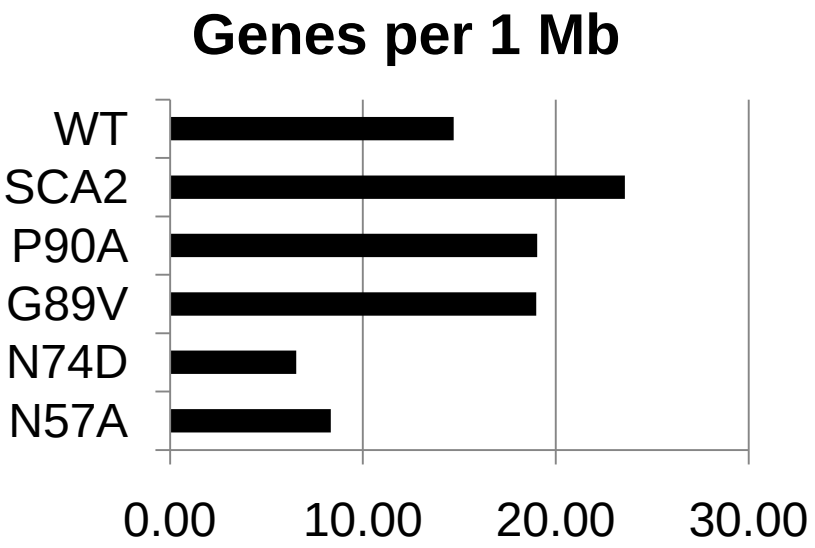
Mutants in CA affect targeting:

Mutants in cyclophilin/RanBP2 binding loop increase integration in gene-dense regions

Mutant N74D decreases binding in gene-dense regions

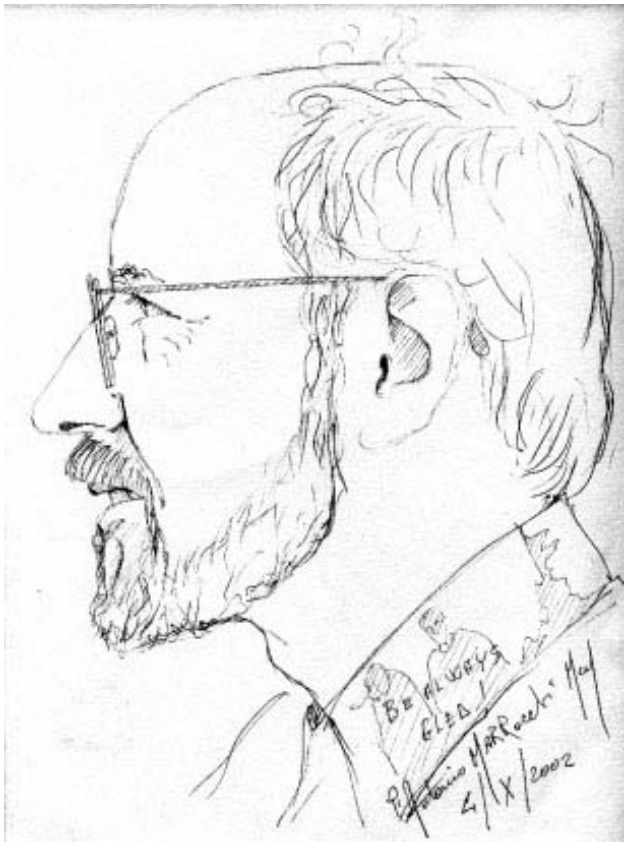
Thorsten Shaller, Karen Ocwieja, Troy Brady, Shannah Roth, Frederic Bushman, Leo James, Greg Towers





- HIV DNA integration introduction
- Viral and host cell factors directing HIV integration targeting
- Integration targeting in human gene therapy

Special Thanks to Charles Berry



3 What does $\hat{z}_{jl}$ estimate?

The expectation of $\hat{z}_{jl}$ is

$$\mathrm{E}(\hat{z}_{jl}) = \sum_{m=1}^{M} \left(\tilde{\pi}_{m|l} \int_{yx} \Phi(y,x) \partial F_{j|lm}(y) \partial G_{j|lm}(x) \right)$$

where $\tilde{\pi}_{m|l}$ is the expected relative frequency of stratum m for sites of type l, $F_{j|lm}$ is the cumulative distribution function of feature j given that the integration site is of type l and in stratum m, and $G_{j|lm}$ is the corresponding CDF for the matched random controls in collection l and stratum m.

Each of the integrals in the summation is the expected ROC curve area for type l and stratum m. These are weighted according according to $\tilde{\pi}_{m|l}, m = 1, \ldots, M$ to determine $\hat{z}_{j|l}$

Note that

$$\tilde{\pi}_{m|l} = \frac{\Pr(\text{recovery of site}|\text{site is type } l \text{ in stratum } m)\pi_{m|l}}{\sum_m \Pr(\text{recovery of site}|\text{site is type } l \text{ in stratum } m)\pi_{m|l}}$$

is not the relative frequency of genomic locations in stratum m, which we denote by $\pi_{m|l}$. If integration sites in stratum m are hard to recover then $\tilde{\pi}_{m|l} < \pi_{m|l}$ and if they are easy to recover then $\tilde{\pi}_{m|l} > \pi_{m|l}$. Truly, we would prefer our estimand to be the area under the ROC curve comparing all integration sites (not just those that were recovered) of the type represented in collection l to the full genome.

An unbiased estimate of $z_{j|l}$ could be based on $\frac{\pi_{m|l}}{\tilde{\pi}_{m|l}}$ if it were known exactly. Taking $m(i,l)$ as the index of the stratum for integration site i in collection l, this estimator would given by

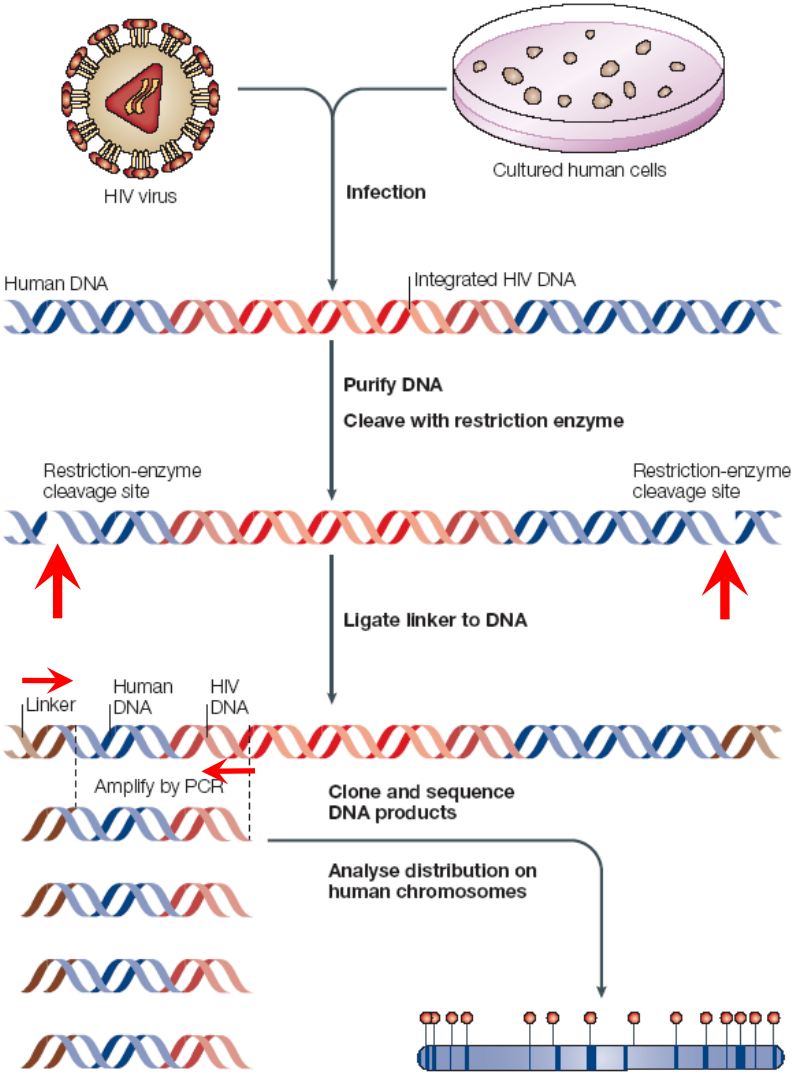
$$\bar{z}_{jl} = \frac{\sum_{ik} \Phi(y_{ijl}, x_{ijkl}) \frac{\pi_{m(i,l)|l}}{\tilde{\pi}_{m(i,l)|l}}}{K \sum_i \frac{\pi_{m(i,l)|l}}{\tilde{\pi}_{m(i,l)|l}}}$$

The expected difference between $\hat{z}_{j|l}$ and $\bar{z}_{j|l}$ is

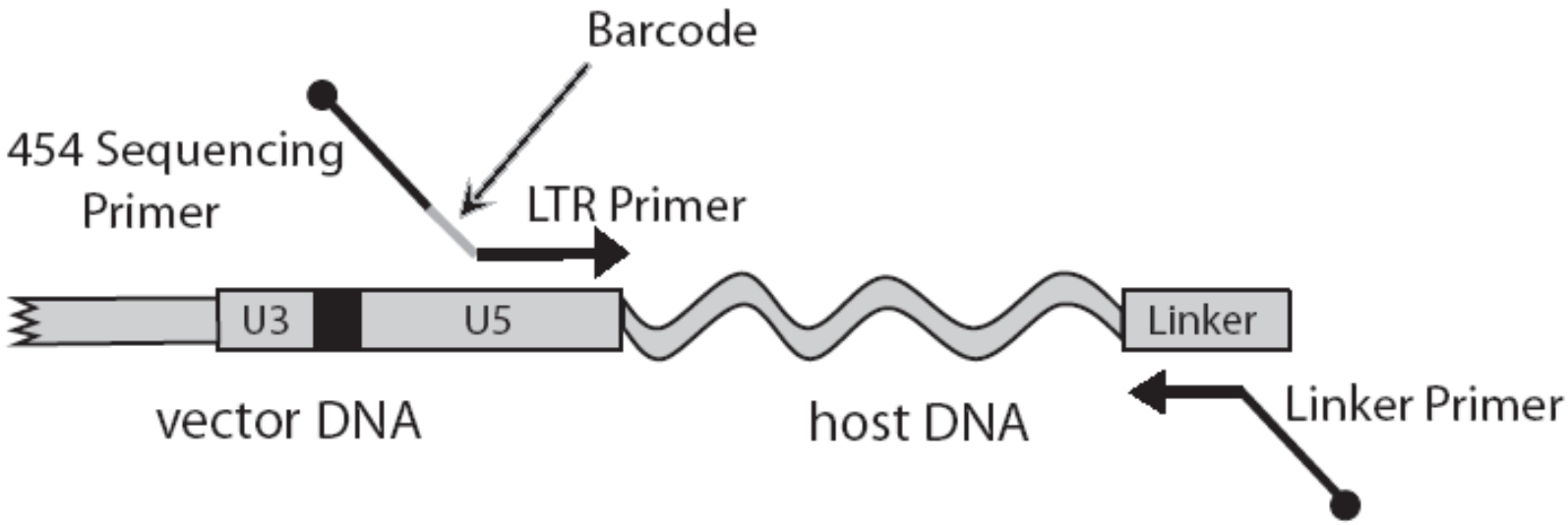
$$\mathrm{E}(\hat{z}_{j|l} - \bar{z}_{j|l}) = \sum_{m=1}^{M} (\tilde{\pi}_{m|l} - \pi_{m|l}) \int_{yx} \Phi(y,x) \partial F_{j|lm}(y) \partial G_{j|lm}(x)$$

From Supplementary Material 3, Brady et al., Genes and Dev. 2009

Technology for analyzing integration site distributions



DNA bar coding



Sample primer sequence

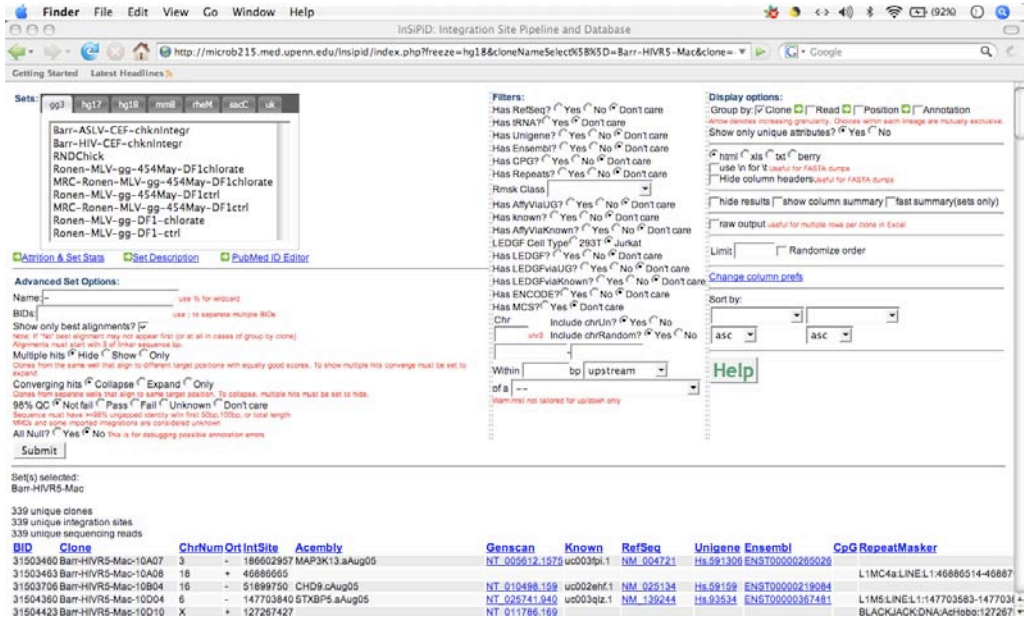
GCCTTGCCAGCCCGCTCAG - CACG - AGACCCTTTTAGTCAGTGTGGAAAATC

454 Sequencing Primer Barcode LTR Primer

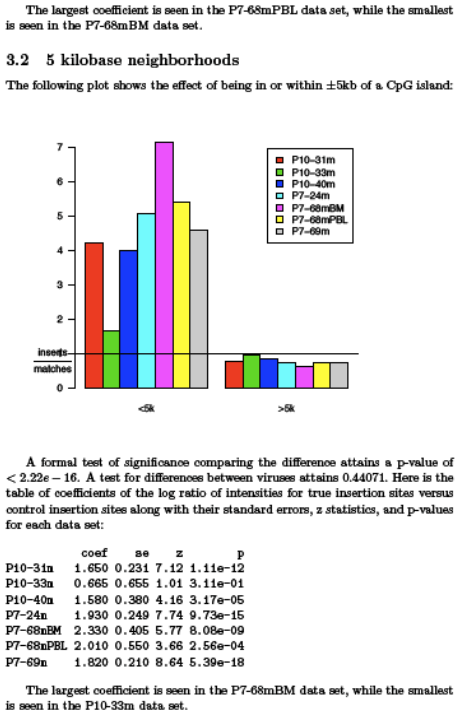
Analytical pipeline

Trim sequence reads
Align to human genome using Blat
Generate matched random controls
(~4,000,000 sequence reads in database)

Manual analysis



Automated reports



The largest coefficient is seen in the P7-68mPBL data set, while the smallest is seen in the P7-68mBM data set.

3.2 5 kilobase neighborhoods

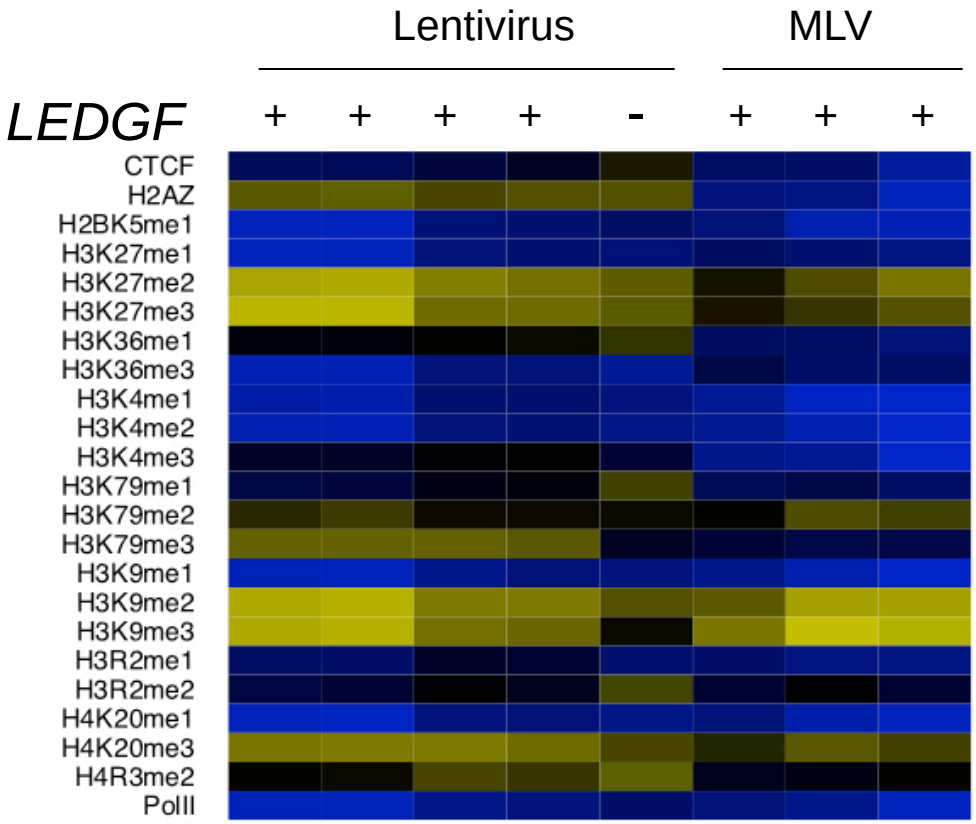
The following plot shows the effect of being in or within ±5kb of a CpG island:

A formal test of significance comparing the difference attains a p-value of $< 2.22e-16$. A test for differences between viruses attains 0.44071. Here is the table of coefficients of the log ratio of intensities for true insertion sites versus control insertion sites along with their standard errors, z statistics, and p-values for each data set:

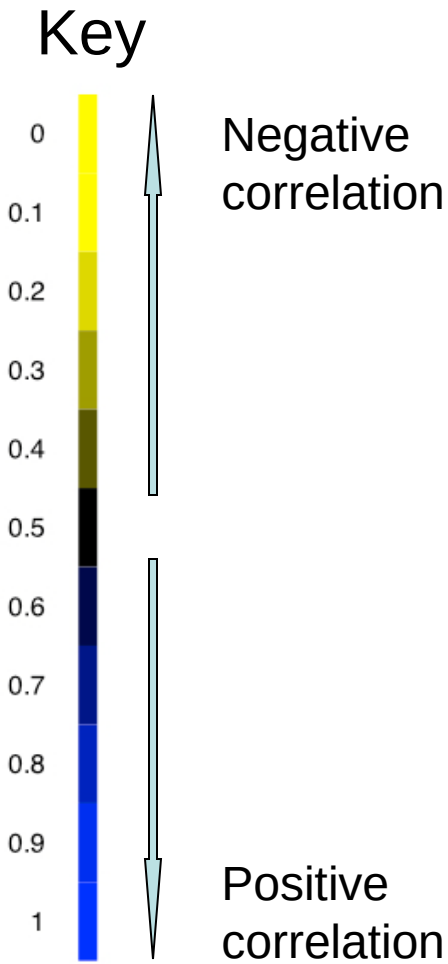
	coef	se	z	p
P10-31m	1.650	0.231	7.12	1.11e-12
P10-33m	0.665	0.655	1.01	3.11e-01
P10-40m	1.580	0.380	4.16	3.17e-05
P7-24m	1.930	0.249	7.74	9.73e-15
P7-68mBM	2.330	0.405	5.77	8.08e-09
P7-68mPBL	2.010	0.550	3.66	2.56e-04
P7-69m	1.820	0.210	8.64	5.39e-18

The largest coefficient is seen in the P7-68mBM data set, while the smallest is seen in the P10-33m data set.

Histone modification and integration frequency

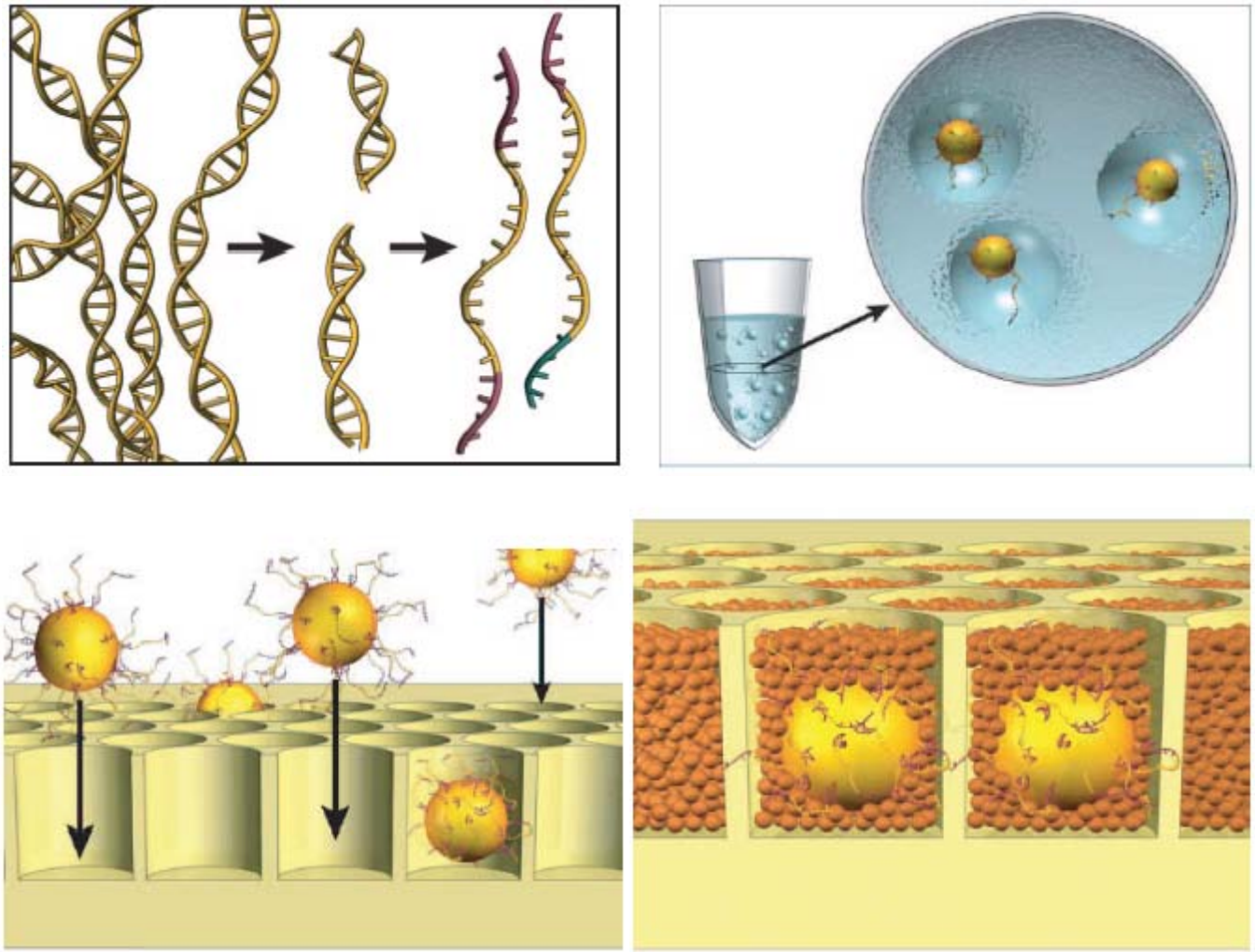


Heat map method described in Berry et al., 2006, PLoS Comp Bio



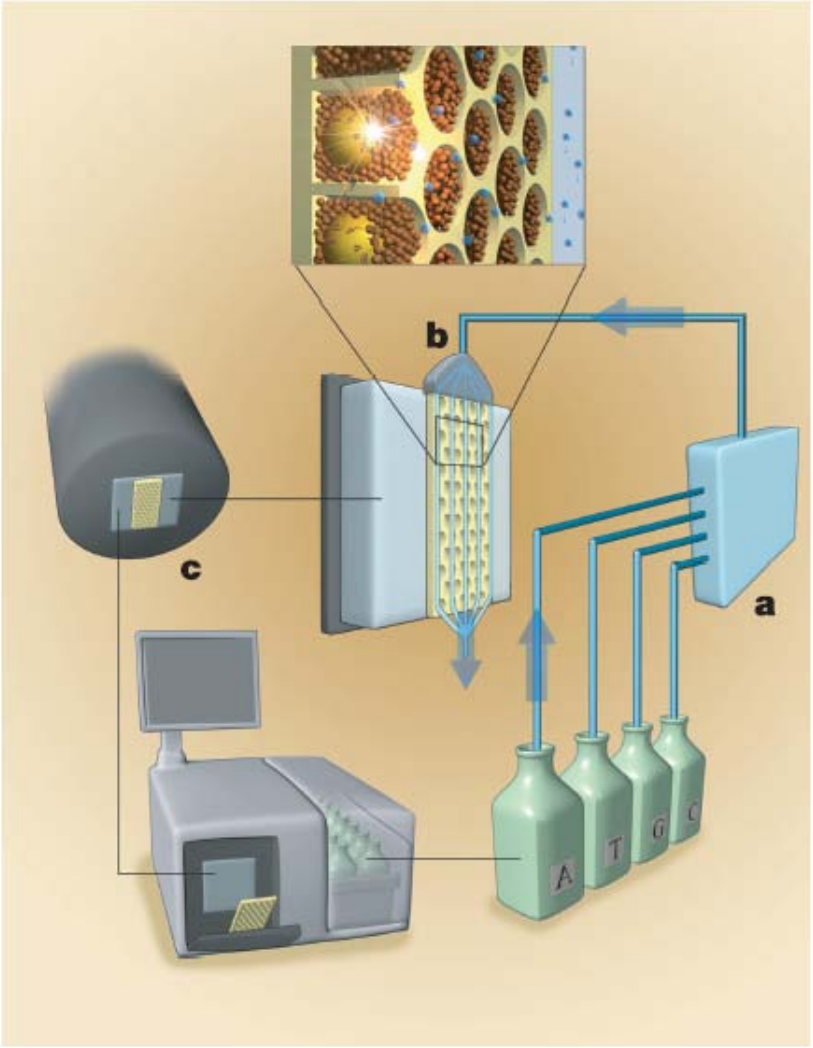
Strong correlations between histone modification and integration frequency

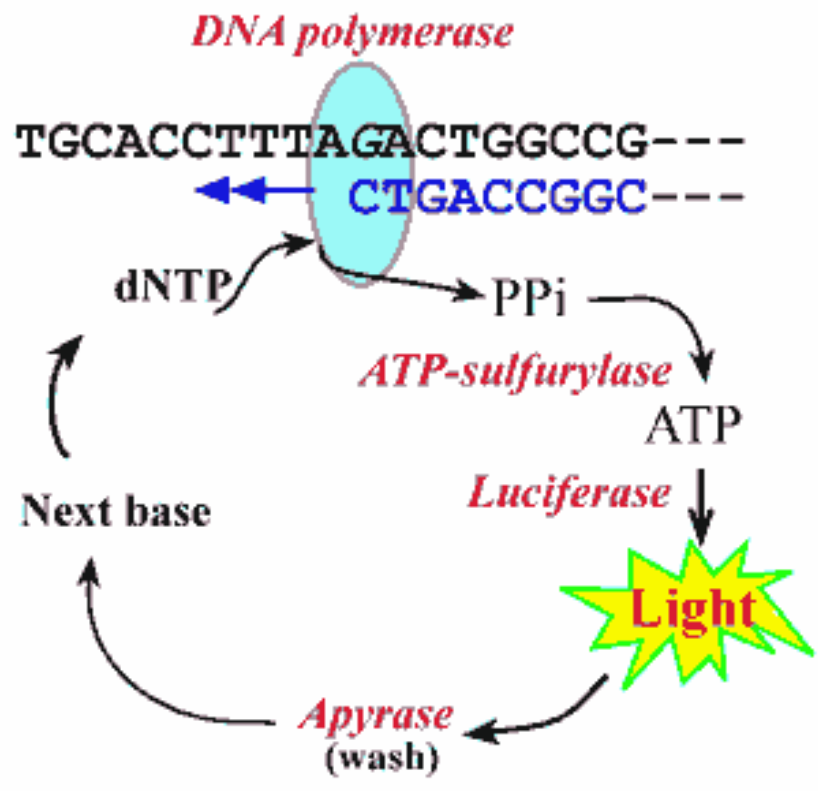
454 sequencing



Margulies et al. (2005) Nature 437, 376-380

454 sequencing





Using this method, a single person sequenced a bacterial genome in a day!!

Margulies et al. (2005) Nature 437, 376-380

- Introduction and methods
- A new host factor in HIV DNA integration
- Gene therapy with integrating vectors: the SCID-X1 trial
- Gene therapy with integrating vectors: trials with lentiviral vectors

Pyrosequencing to analyze 40,000 sites of HIV DNA Integration

New findings: nucleosome bound DNA usually target, correlations with histone post-translational modification

Wang et al., Genome Res. 2007

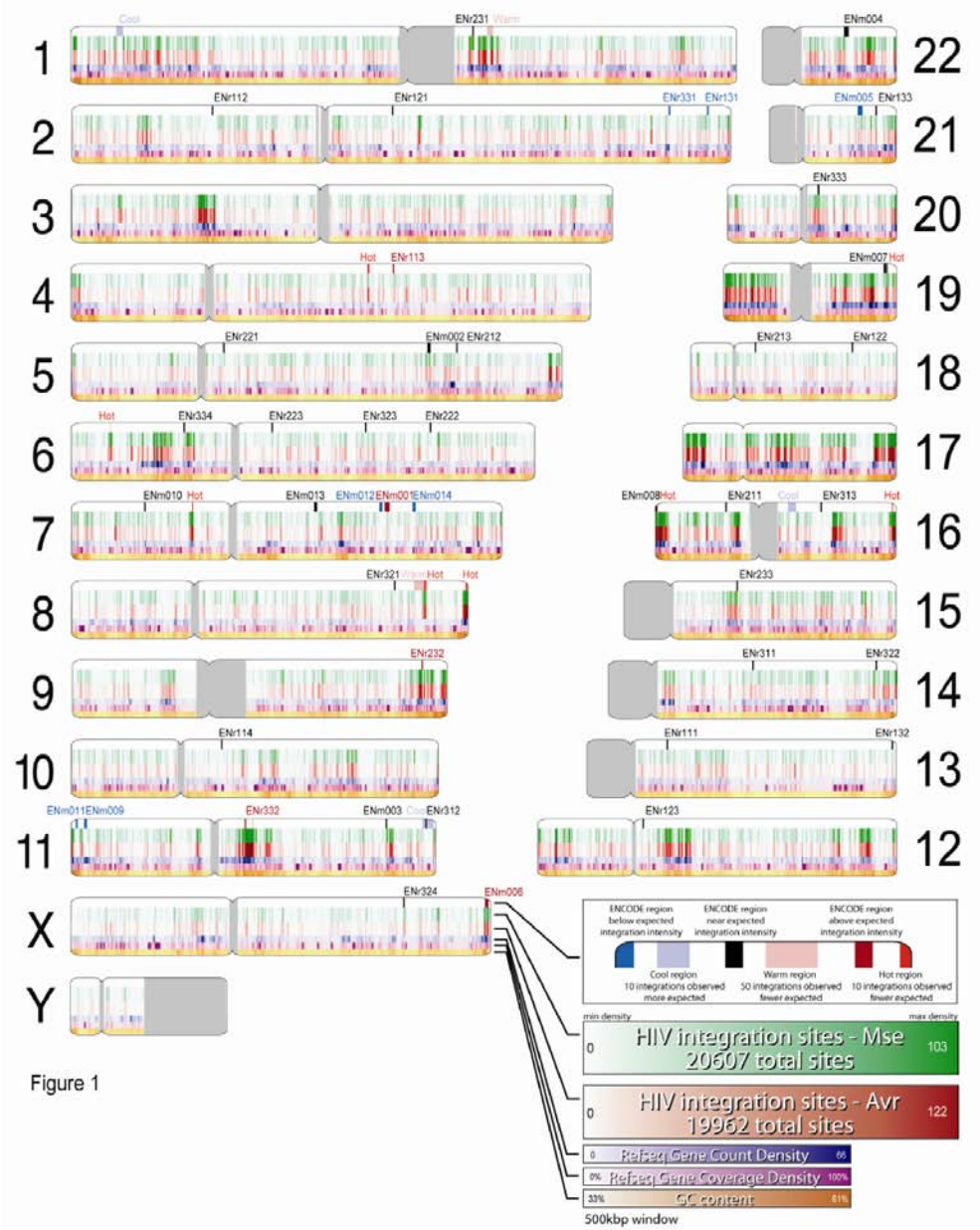
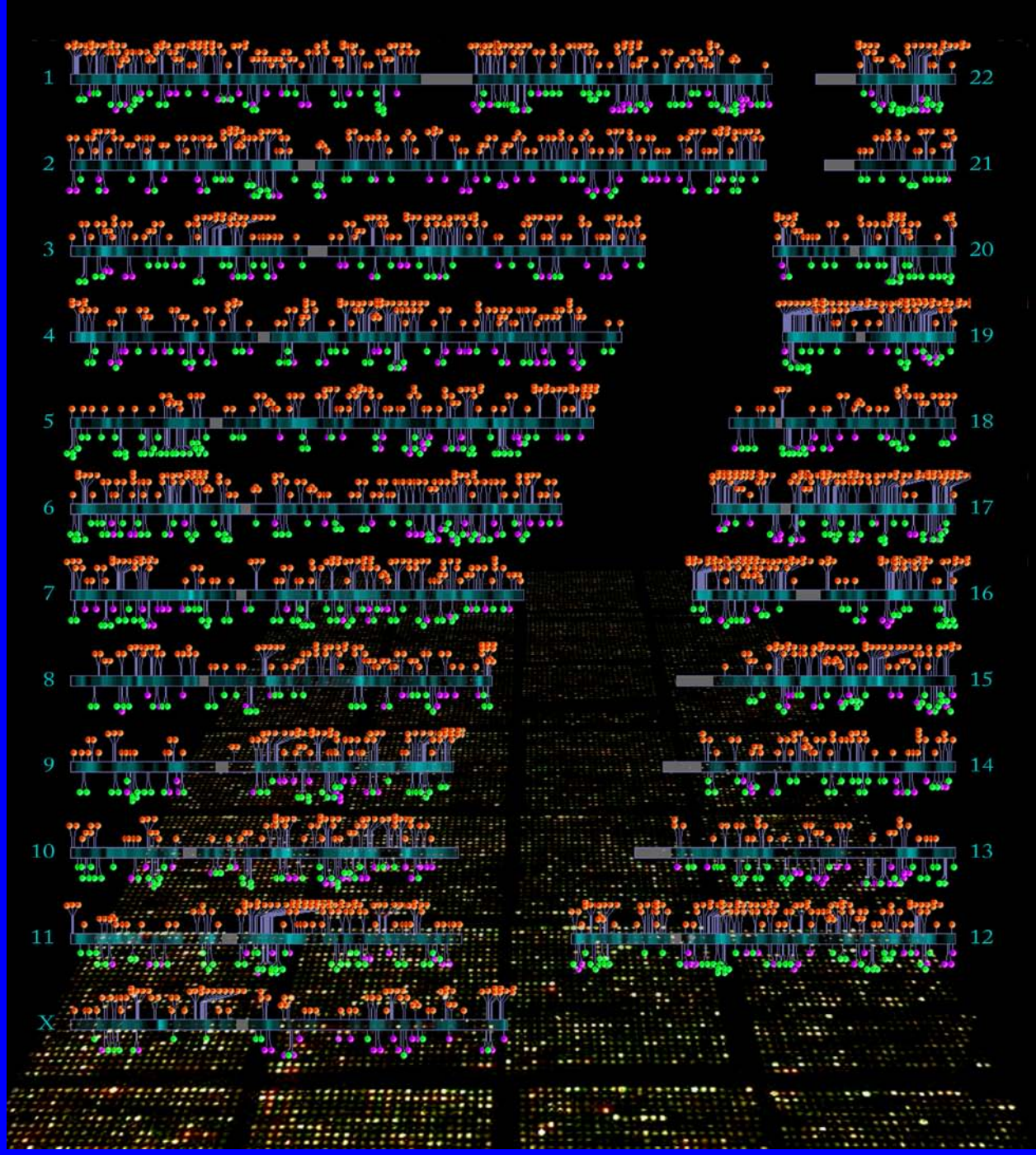


Figure 1

CREDITS

R. S. Mitchell
B. F. Beitzel
A. R. W. Schroder
P. Shinn
H. Chen
C. C. Berry
J. R. Ecker
F. D. Bushman



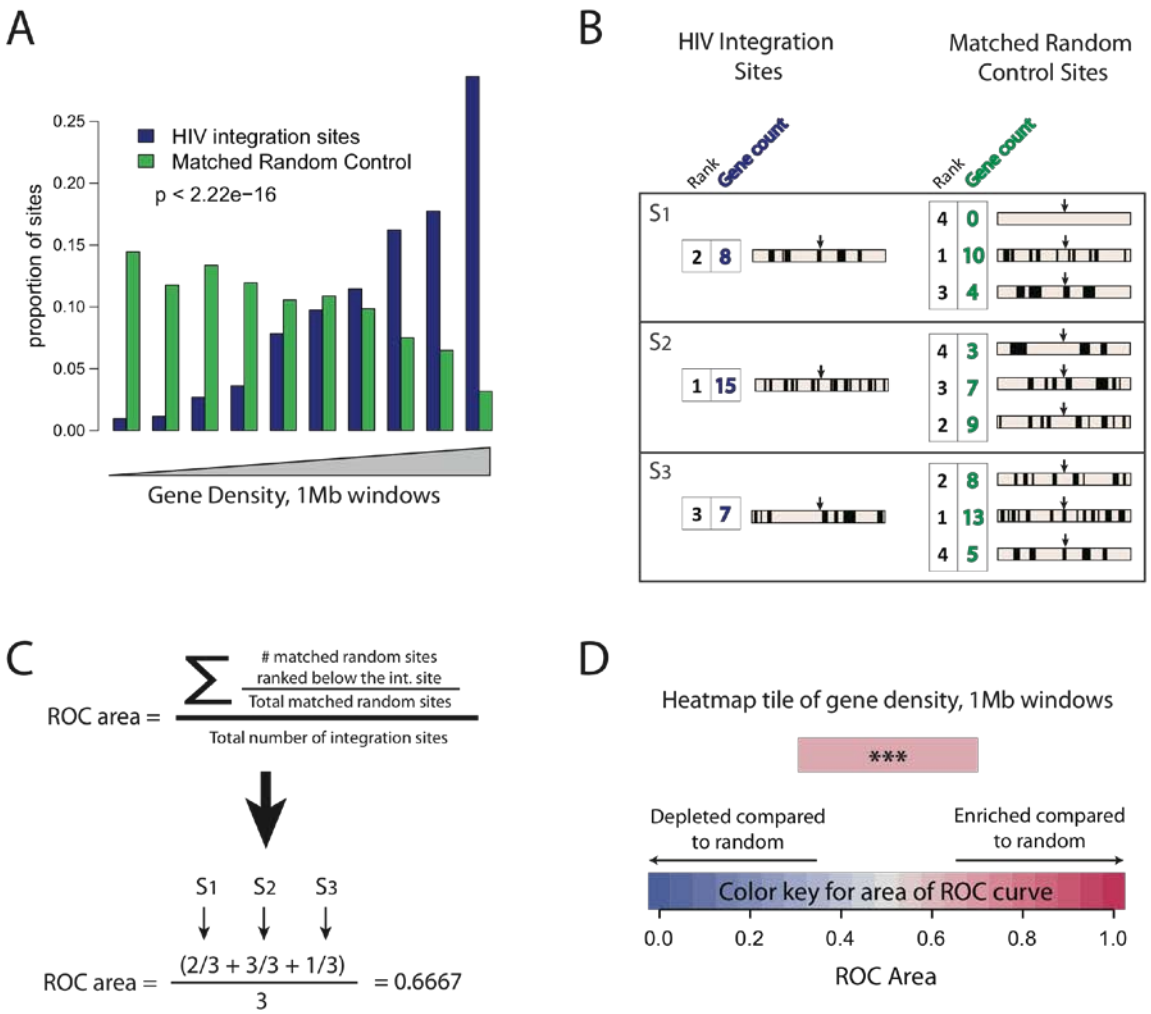
Different target site preferences for different viruses

HIV favors integration in active transcription units
Schroder, Shinn, Chen, Berry, Ecker and Bushman, Cell, 2002

MLV favors integration near transcription start sites
Wu, Crise, and Burgess, Science 2003

ASLV integration close to random
Mitchell, Beitzel, Schroder, Shinn, Chen, Berry, Ecker and Bushman, PLoS Bio. 2004
Narezkina, Taganov, Litwin, Stoyanova, Hayashi, Seeger, Skalka, Katz, J. Virol 2004

Summarizing departures from random integration as ROC areas



Berry et al., PLOS Comp. Bio., 2006

- Introduction and methods
- A new host factor in HIV DNA integration
- Gene therapy with integrating vectors: the SCID-X1 trial
- Gene therapy with integrating vectors: trials with lentiviral vectors

Retroviral DNA Integration: ASLV, HIV, and MLV Show Distinct Target Site Preferences

R. S. Mitchell, B. F. Beitzel, A. R. W. Schroder, P. Shin, H. Chen, C. C. Berry, J. R. Ecker, F. D. Bushman

CSH presentation, 2004

Integration is Favored in Genes

Gene defined by Acembly program

vector	cell	% in genes	# of sites
HIV	PBMC	82	528
HIV	IMR90	75	465
HIV	SupT1	80	436
MLV	HeLa	61	822
ASLV	293T-TVA	57	469
human genome		45	

- HIV strongly favors integration into genes independent of cell type
- MLV and ASLV weakly favor integration into genes

PLoS Biology, in press

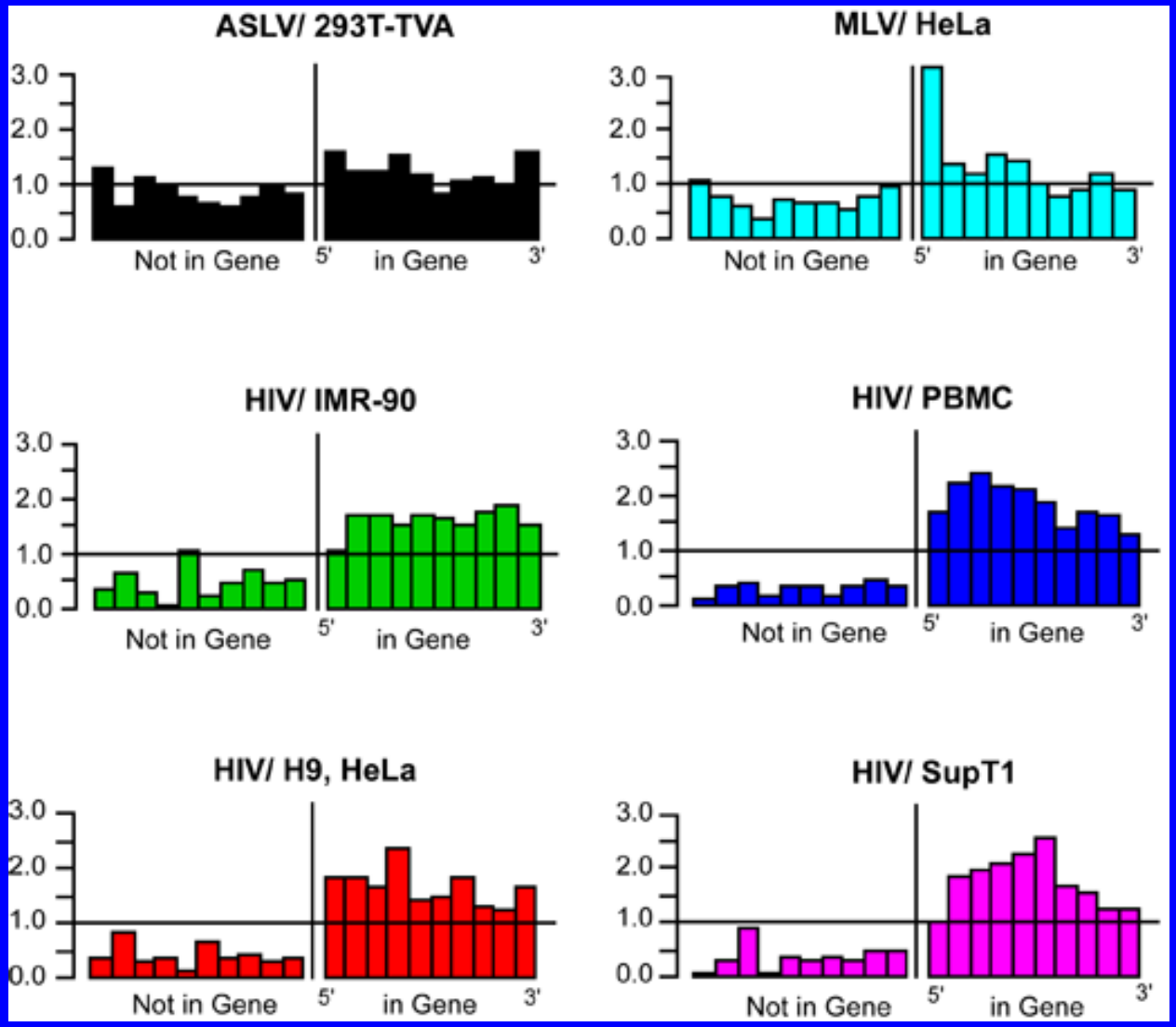
Vector integration sites in human genome

Blue lollipops = HIV

Red lollipops = MLV

Green lollipops = ASLV

Integration site selection by HIV, MLV, and ASLV



MLV and HIV/H9,Hela data from Burgess and coworkers (Science 2003)

Statistical study by Chuck Berry:

Incorporates matched random control.

ASLV, MLV, and HIV all significantly different from one another

HIV favors transcription units, no preference for location

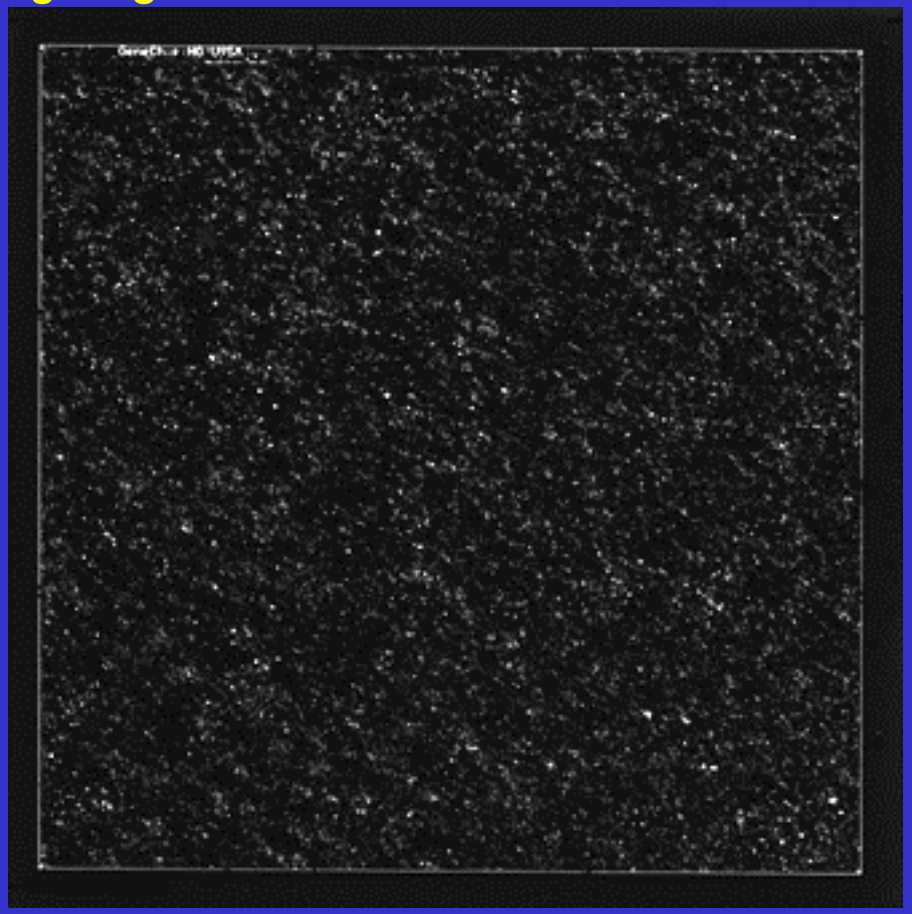
MLV strongly favors 5' ends of genes (Burgess)

ASLV shows only very weak preference for genes, no preference for 5' ends

Does integration targeting correlate with transcriptional activity?

Assess by comparing transcriptional profiling data(Affymetrix chips)

to integration targeting data



Rank all genes on chip (12,500) by expression.

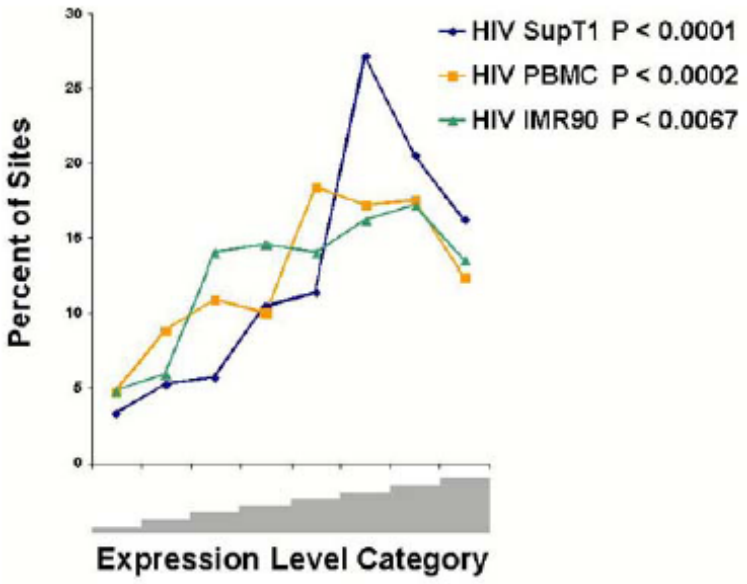
Place them into 8 bins:
bin#1= 1/8 lowest expressing genes

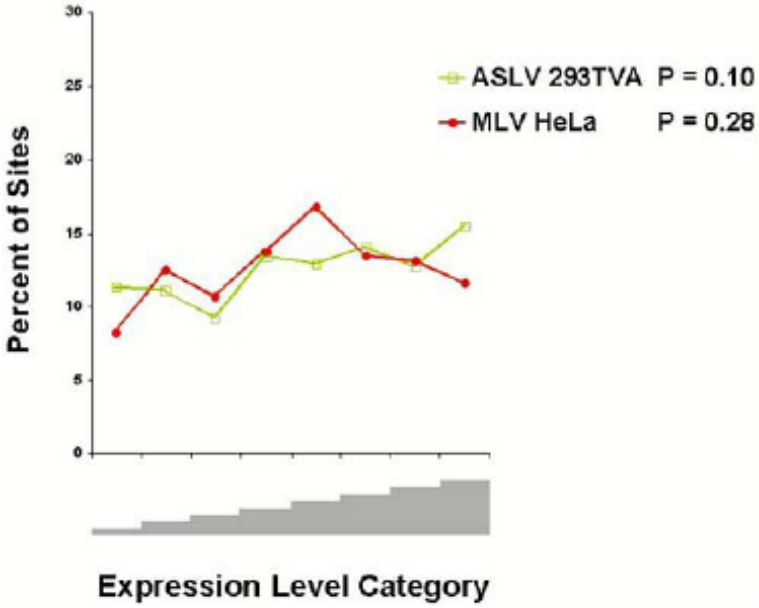
On up to bin #8= 1/8 highest expressing genes

Influence of gene activity on integration frequency

HIV favors integration into active genes in all cell types

MLV and ASLV do not show significant trend for integration into active genes



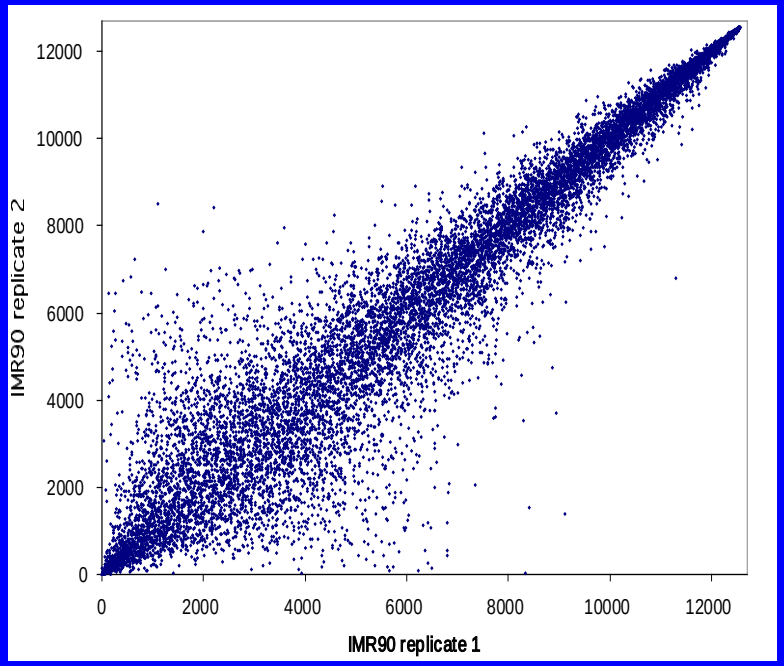


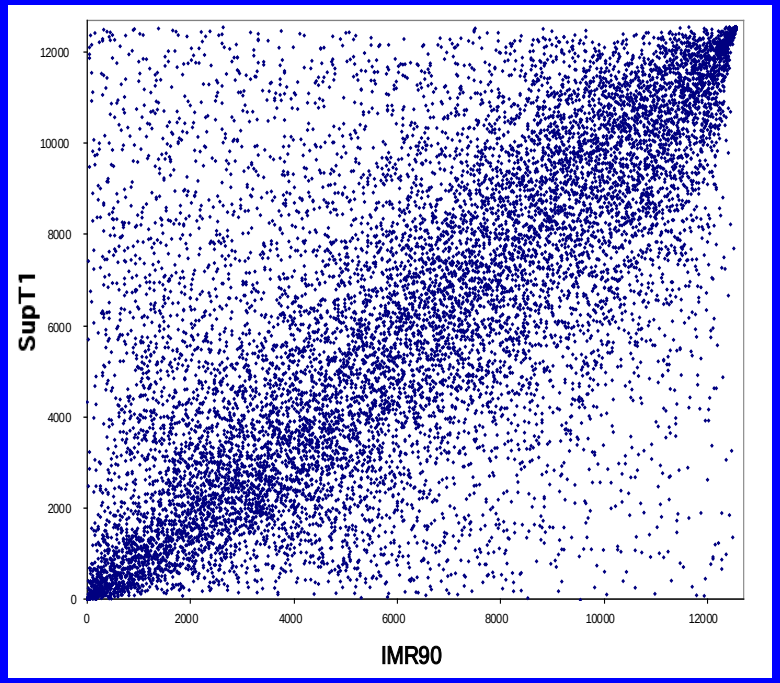
- How different will integration targeting be in different cell types due to different transcriptional programs?

- How different are the transcriptional programs between different cell types?

Tissue-Specific Transcription and HIV Integration Targeti

Documenting tissue-specific gene expression in SupT1, PBMC, IMR90





Cell Types	R
SupT1 and PBMC	0.7936
SupT1 and IMR90	0.6842
PBMC and IMR90	0.7350

Relatively low correlation coefficients indicate tissue-specific differences in transcriptional patterns

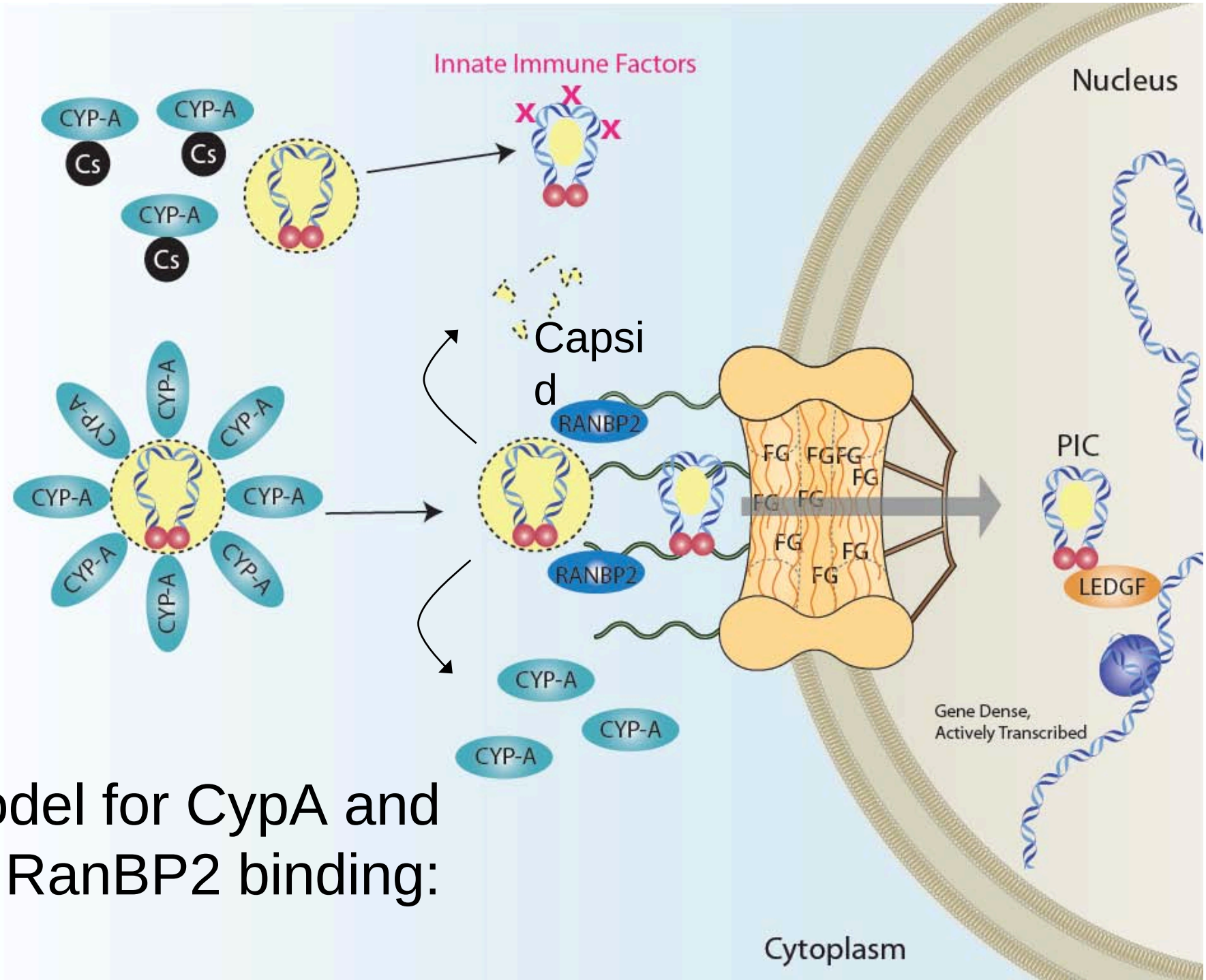
Tissue-Specific Transcription and HIV Integration Targeting

Question:
How different will integration targeting be in different cell types due to different transcriptional programs?

Method:
Compare HIV integration targeting in PBMC, SupT1, and IMR-90. Ask whether targeted genes are more highly expressed in the authentic host cell compared to the other two.

Result:
Each set of integration sites showed higher expression in the authentic host cell type compared to the other two. (IMR-90 P=0.00045; PBMC P=0.027; SupT1 P=0.046)

Conclusion: cell-type-specific transcription does bias integration site selection, but the effect is quantitatively modest.

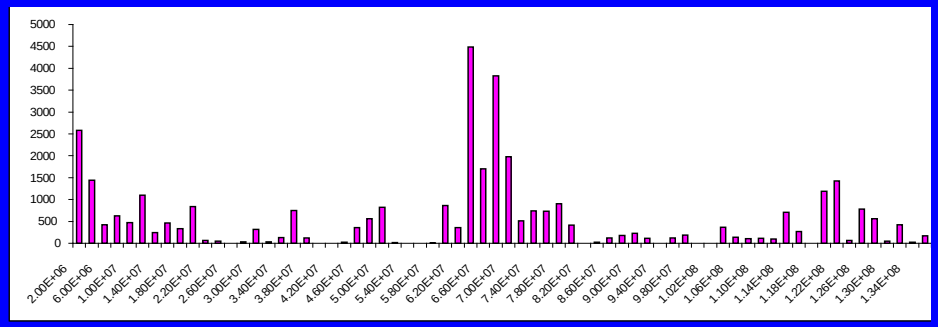


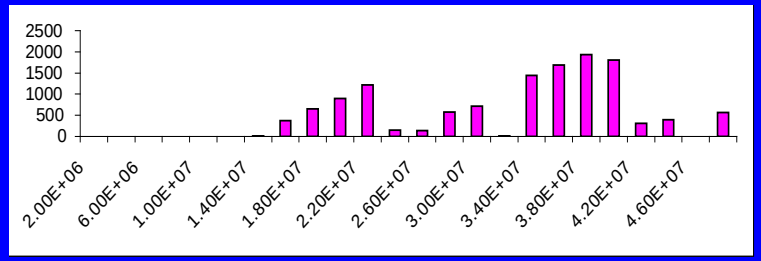
Model for CypA and RanBP2 binding:

Transcriptional Intensity Versus Integration Intensity

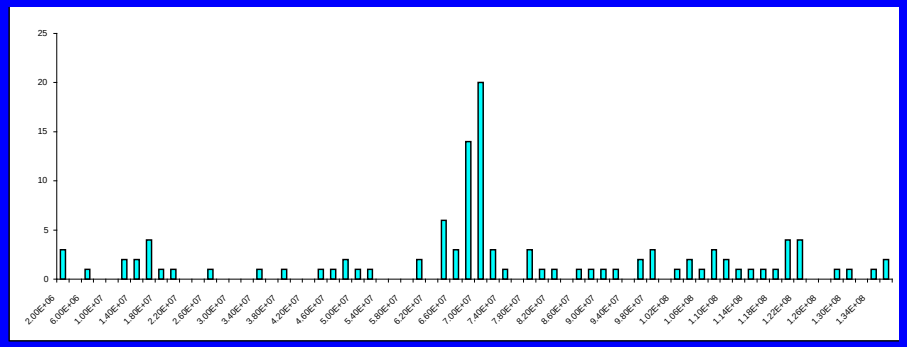
Chromosome 11 Chromosome 22

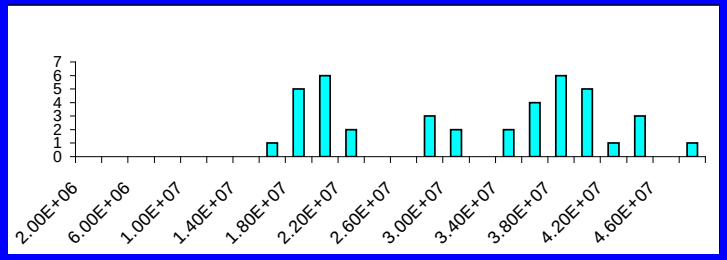
Transcriptional Intensity (EST data from Sanger Center)





HIV Integration Intensity





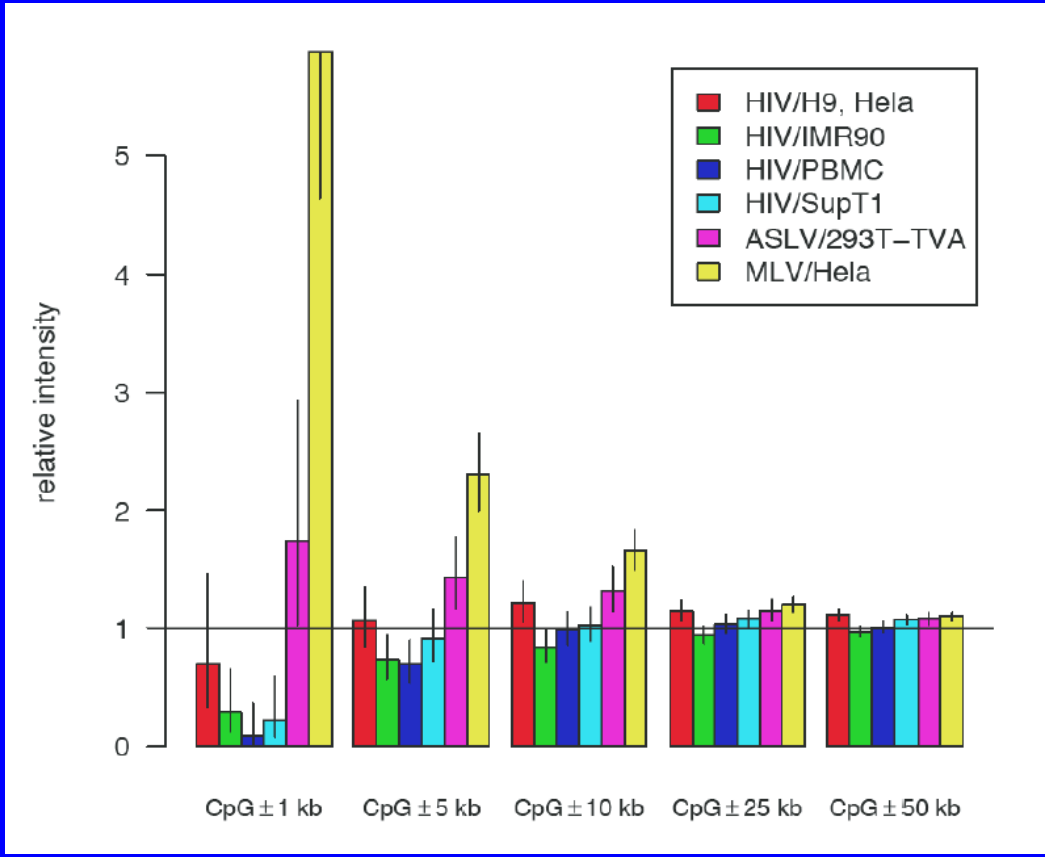
Regions of high transcriptional intensity are favored for HIV integration.
Weaker but detectable effect for MLV and ASLV

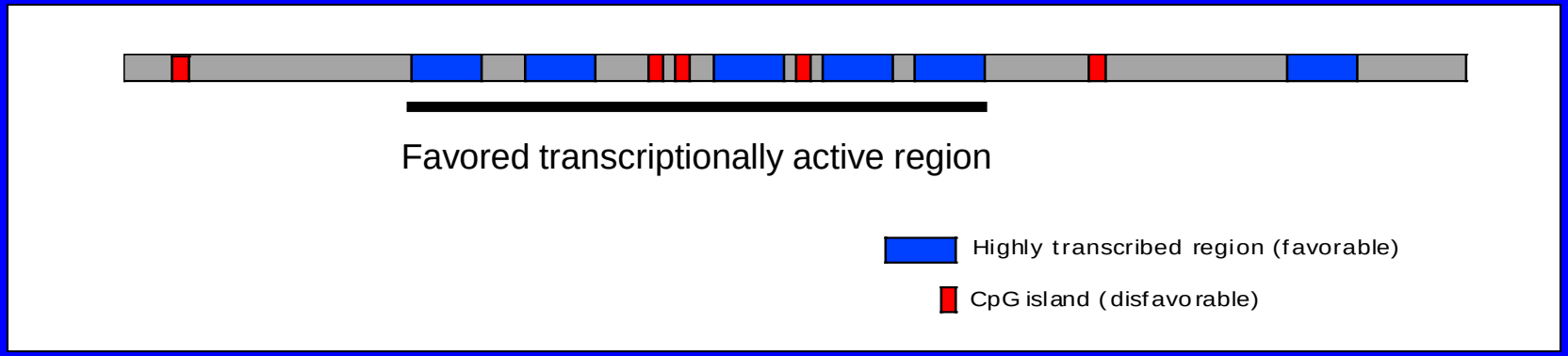
Sub-structure within favored chromosomal regions

Two lines of study allow us to subdivide favored chromosomal regions

Statistical study, assessing the relationship between integration intensity and transcriptional intensity, indicates the size of favored features is much smaller than highly expressed chromosomal domains (100-250 kb).

CpG islands, regions upstream of genes that are enriched in transcription factor binding sites, are enriched in highly expressed chromosomal regions but disfavored for HIV integration.



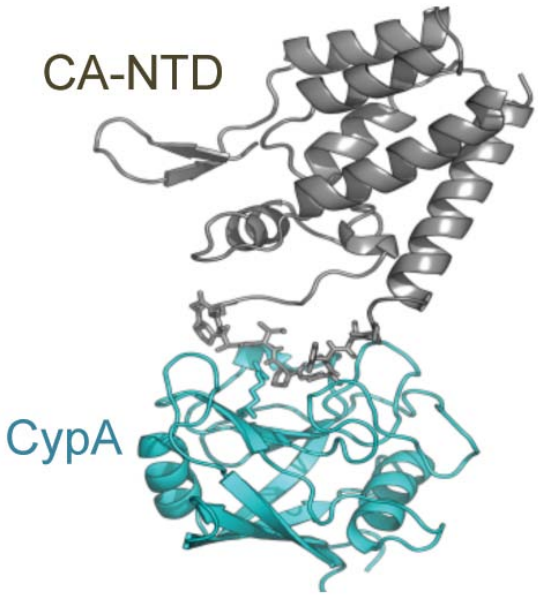


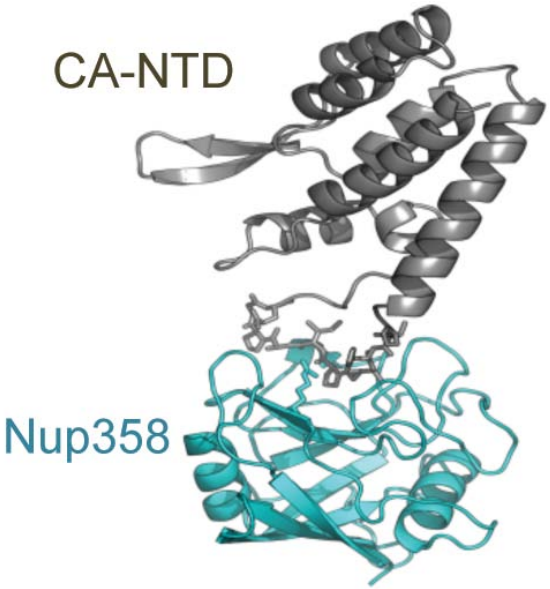
Conclusions: Target site selection by integration complexes in human cells

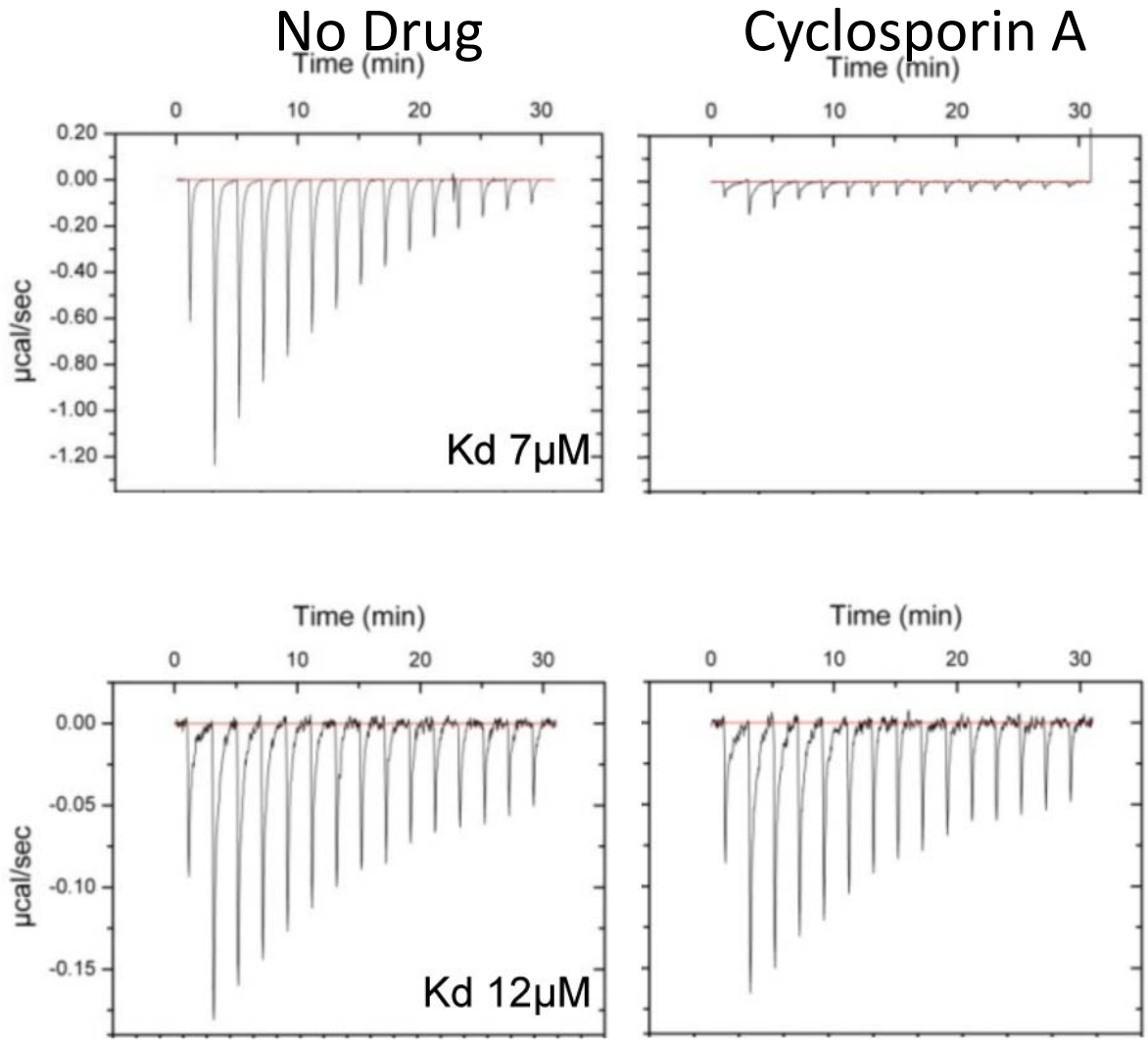
- HIV favors integration in active genes (in primary cells as well as transformed cell lines)
- MLV favors gene transcription start regions (Burgess and coworkers)
- ASLV displays no strong biases
- For HIV, clear though modest influence of tissue specific transcription on integration targeting
- Favored regions small in size (100-250 kb)--length of one or a few genes. For HIV, favorable regions interspersed with unfavorable CpG islands. (MLV islands favorable, weakly so for ASLV)

- Introduction and methods
- A new host factor in HIV DNA integration
- Gene therapy with integrating vectors: the SCID-X1 trial
- Gene therapy with integrating vectors: trials with lentiviral vectors

CA interacts with RanBP2 (NUP358)







Data from Greg Towers

Genes Studied

Gene	Function	
TNPO3	Transportin, nuclear import	Christ et al. 2008, Lee et al. 2010
RANBP2 (Nup358)	Nuclear pore	Lee et al. 2010, Torsten Schaller (Towers)
NUP98	Nuclear pore	Ebina et al. 2004
MAP4	Microtubule-associated protein	
ANAPC2	Anaphase promoting complex	
SNW1	Splicing/chromatin protein	
PRPF38A	Splicing protein	
IK	Nuclear protein	
WDHD1	Binds LEDGF, DNA binding?	
WDR46	Binds LEDGF	

Knockdowns analyzed

Treatment	# Integration Sites	
Mock	317	Controls
siGL2	4152	
TNPO3	2590	Candidate Factors
RANBP2	1656	
NUP98	1226	
MAP4	8709	
IK	1028	
ANAPC2	1061	
PRPF38A	2124	
SNWI	696	
WDHD1	243	
WDR46	2622	