Regions of Active Transcription Favored by Retroviral Integration

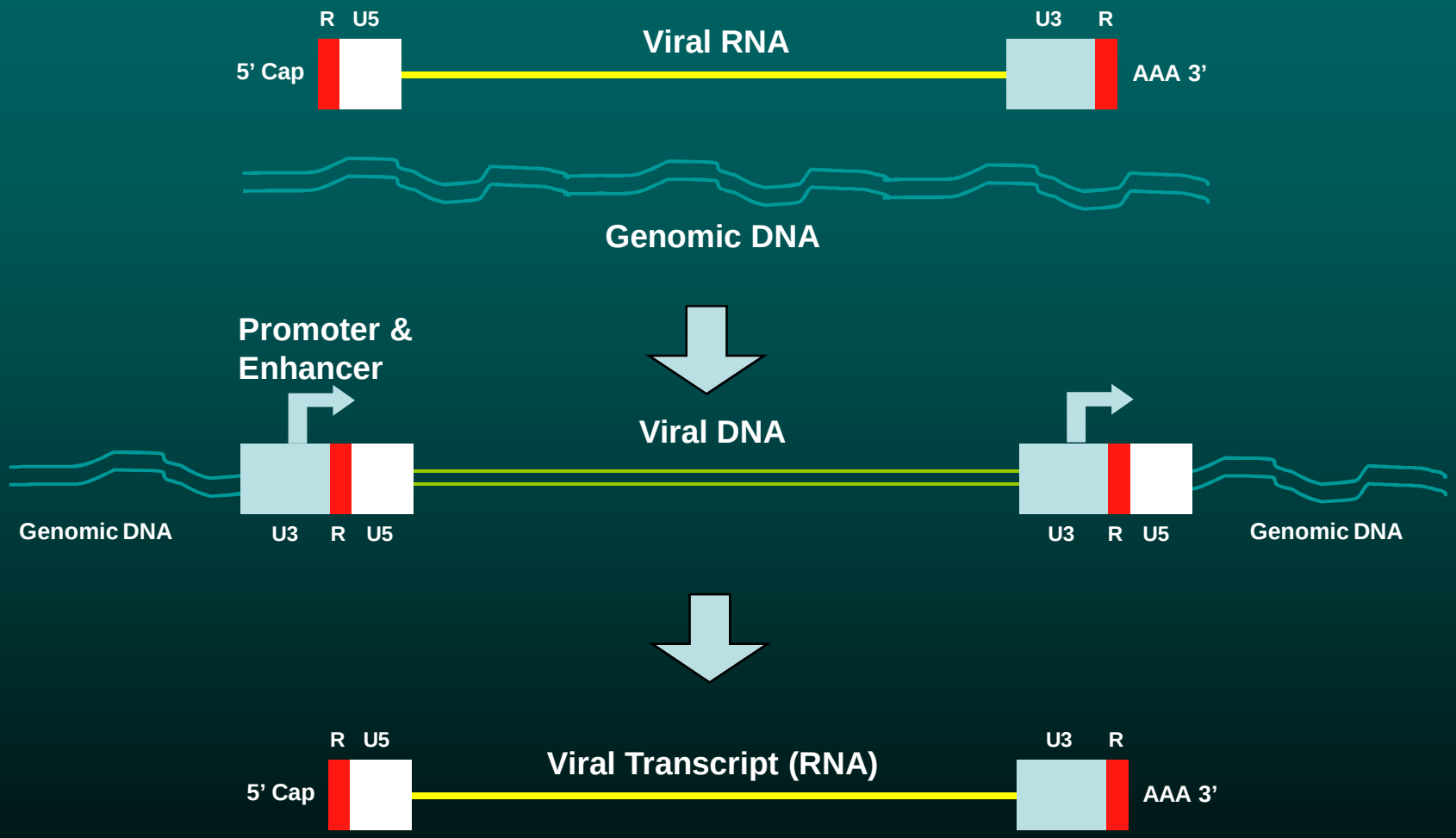
Transcription Start Regions in the Human Genome Are Favored Targets for MLV Integration

Xiaolin Wu,[1] Yuan Li,[2] Bruce Crise,[2] Shawn M. Burgess[1]*

Factors contributing to retroviral integration have been intractable because past studies have not precisely located genomic sites of proviruses in sufficient numbers for significant analysis. In this study, 903 murine leukemia virus (MLV) and 379 human immunodeficiency virus–1 (HIV-1) integrations in the human genome were mapped. The data showed that MLV preferred integration near the start of transcriptional units (either upstream or downstream) whereas HIV-1 preferred integration anywhere in the transcriptional unit but not upstream of the transcriptional start. Defining different integration site preferences for retroviruses will have important ramifications for gene therapy and may aid in our understanding of the factors directing the integration process.

SCIENCE VOL 300 13 JUNE 2003

Integration Is an Essential Part of the Retroviral Life Cycle



Retroviruses and Disease

- Oncoviruses
  - Alpharetroviruses: ALV
  - Betaretroviruses: MMTV
  - Gammaretroviruses: MuLV
  - Deltaretroviruses: HTLV
    - Mechanism of HTLV disease induction not well defined
- Lentiviruses (HIV, SIV)
  - Limited evidence of gene deregulation by HIV infections

Interaction of Retroviruses with Genomes

Mobilization of proto

RSV and identificat

- Deleterious integrations

  Peyton Rous, early 1

  Varmus and Bishop,

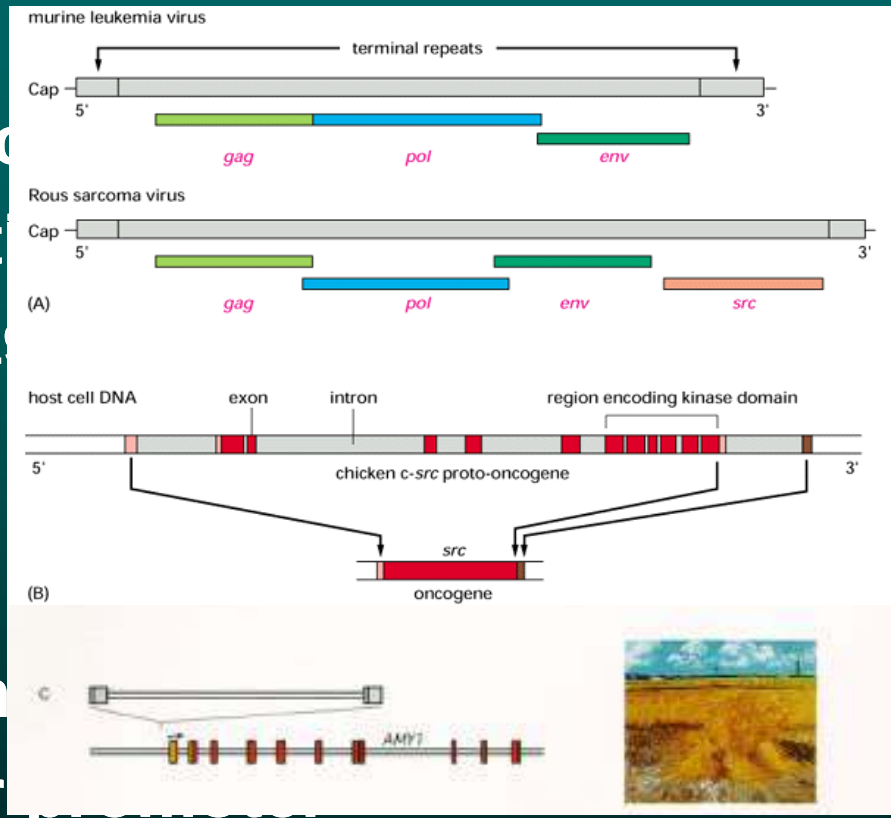
- Phenotypic changes

- Approximately 10% of the human genome consists of retroviral elements

Insertional mutagen

Activation of tumor

Inactivation of tumor suppressor



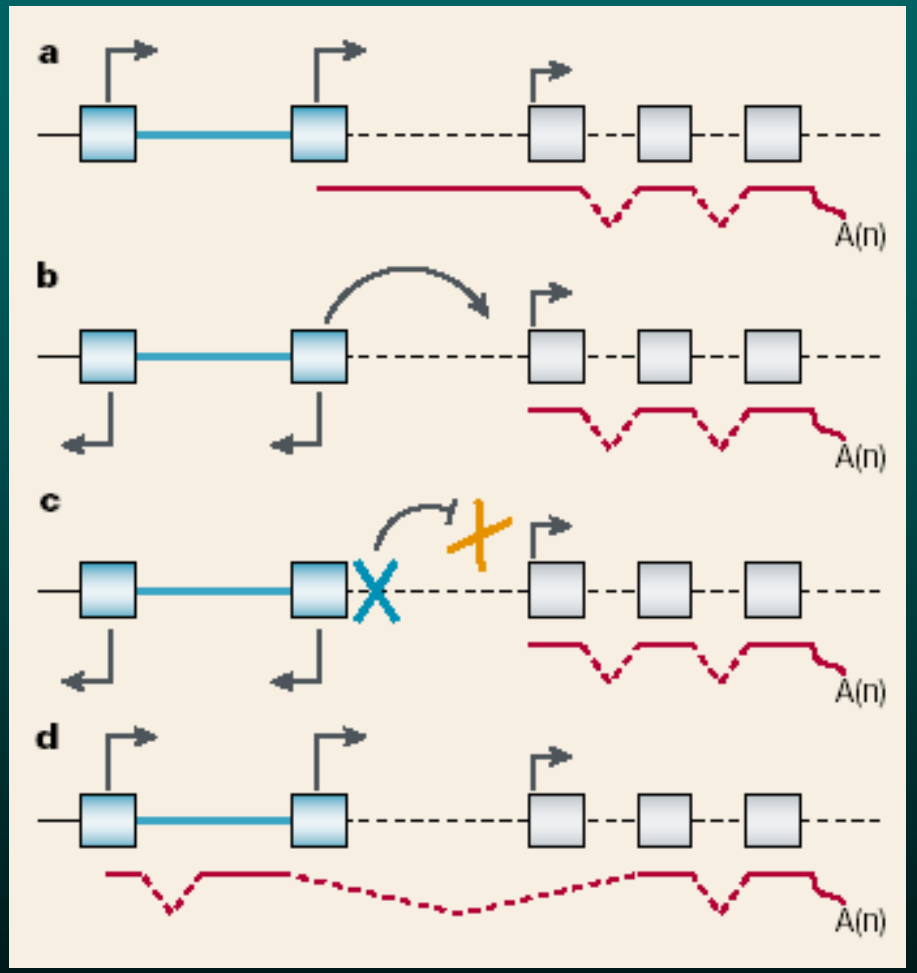
Examples of Retroviral Insertional Mutagenesis

a.) 3’ LTR drives expression of a downstream gene

b.) 5’ LTR enhancer augments expression of a downstream gene

c.) Disruption of transcriptional or post-transcriptional control elements

d.) Fusion of retroviral and host cell trasncripts



Gene Therapy Leads to Human Disease

LMO2-Associated Clonal T Cell Proliferation in Two Patients after Gene Therapy for SCID-X1

S. Hacein-Bey-Abina,[1,2*] C. Von Kalle,[6,7,8] M. Schmidt,[6,7] M. P. McCormack,[9] N. Wulffraat,[10] P. Leboulch,[11] A. Lim,[12] C. S. Osborne,[13] R. Pawliuk,[11] E. Morillon,[2] R. Sorensen,[19] A. Forster,[9] P. Fraser,[13] J. I. Cohen,[15] G. de Saint Basile,[1] I. Alexander,[16] U. Wintergerst,[17] T. Frebourg,[18] A. Aurias,[19] D. Stoppa-Lyonnet,[20] S. Romana,[3] I. Radford-Weiss,[3] F. Gross,[2] F. Valensi,[4] E. Delabesse,[4] E. Macintyre,[4] F. Sigaux,[20] J. Soulier,[21] L. E. Leiva,[14] M. Wissler,[6,7] C. Prinz,[6,7] T. H. Rabbitts,[9] F. Le Deist,[1] A. Fischer,[1,5†‡] M. Cavazzana-Calvo[1,2†]

We have previously shown correction of X-linked severe combined immunodeficiency [SCID-X1, also known as γ chain (γc) deficiency] in 9 out of 10 patients by retrovirus-mediated γc gene transfer into autologous CD34 bone marrow cells.

Determine the Location of HIV Integrations Relative to MuLV Integrations

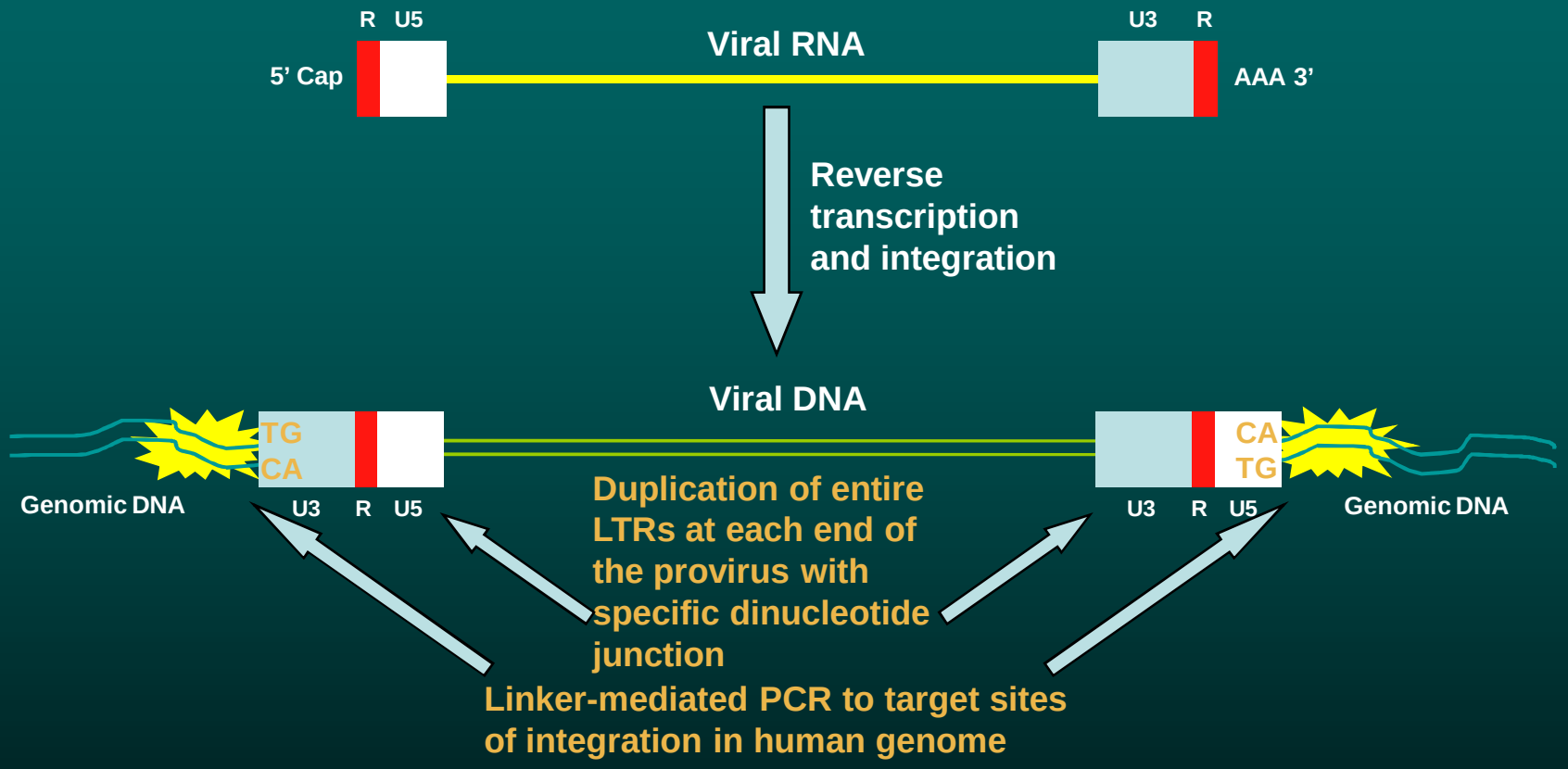
Recent concerns from X-SCID gene therapy trials:

- 2 of 11 recipients developed leukemia and
- MuLV vector carrying γC was found upstream of LMO2 gene

Schroeder et al. indicate that HIV integrates in transcriptional hot spots, but precise location not identified

Are there differences between HIV-1 and MuLV integrations?

Retroviral Genome and Integrated Provirus

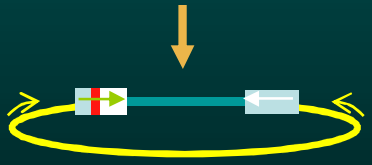


Method for Mapping Retroviral Integrations

Linker-mediated PCR (LM-PCR)

Restriction enzyme digestion (MseI: TTAA)
Linker ligation
PCR

Nested PCR to increase the number of specific clones



Clone

Sequence

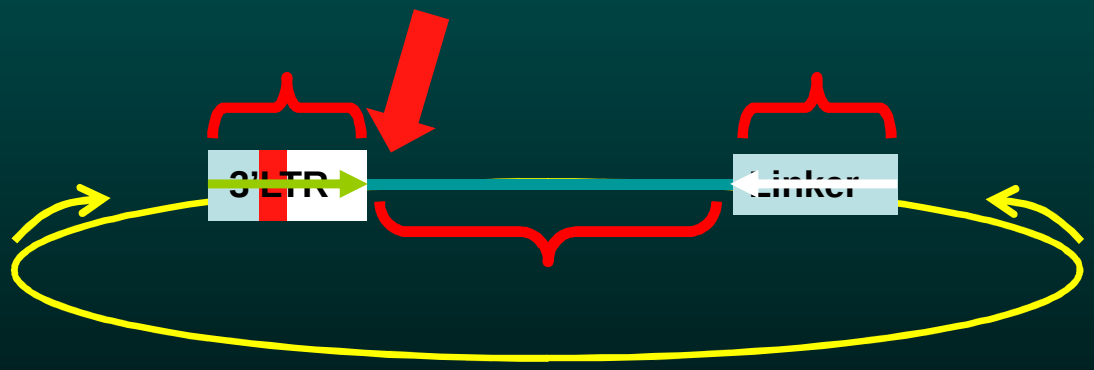
Sort and eliminate artifact sequences

BLAST/BLAT genomic sequence

Definition of an Integration Site

Contains the following:

1) 3’ LTR sequence from nested primer and linker seq.
2) Genomic sequence within 3bp of end of 3’ LTR
3) 95% or greater sequence identity
4) Matched to only one genomic locus



Mapping Retroviral Integrations in the Human Genome

Comparison of MuLV and HIV-1 integration sites

1) MuLV:

    2304 total sequence reads

    903 unambiguously mapped

2) HIV-1:

    379 unambiguously mapped

    Lower quality library:

    Schroeder et al. integrations (500)

Integrations Mapping to Genes

	MuLV	HIV-1	Random
Integrations in RefSeq Genes	34.2% (309/904)	57.8% (219/379)	22.4%

Integrations Found in Introns

	MuLV	HIV-1
Integrations in Introns	95% (292/309)	94% (205/219)

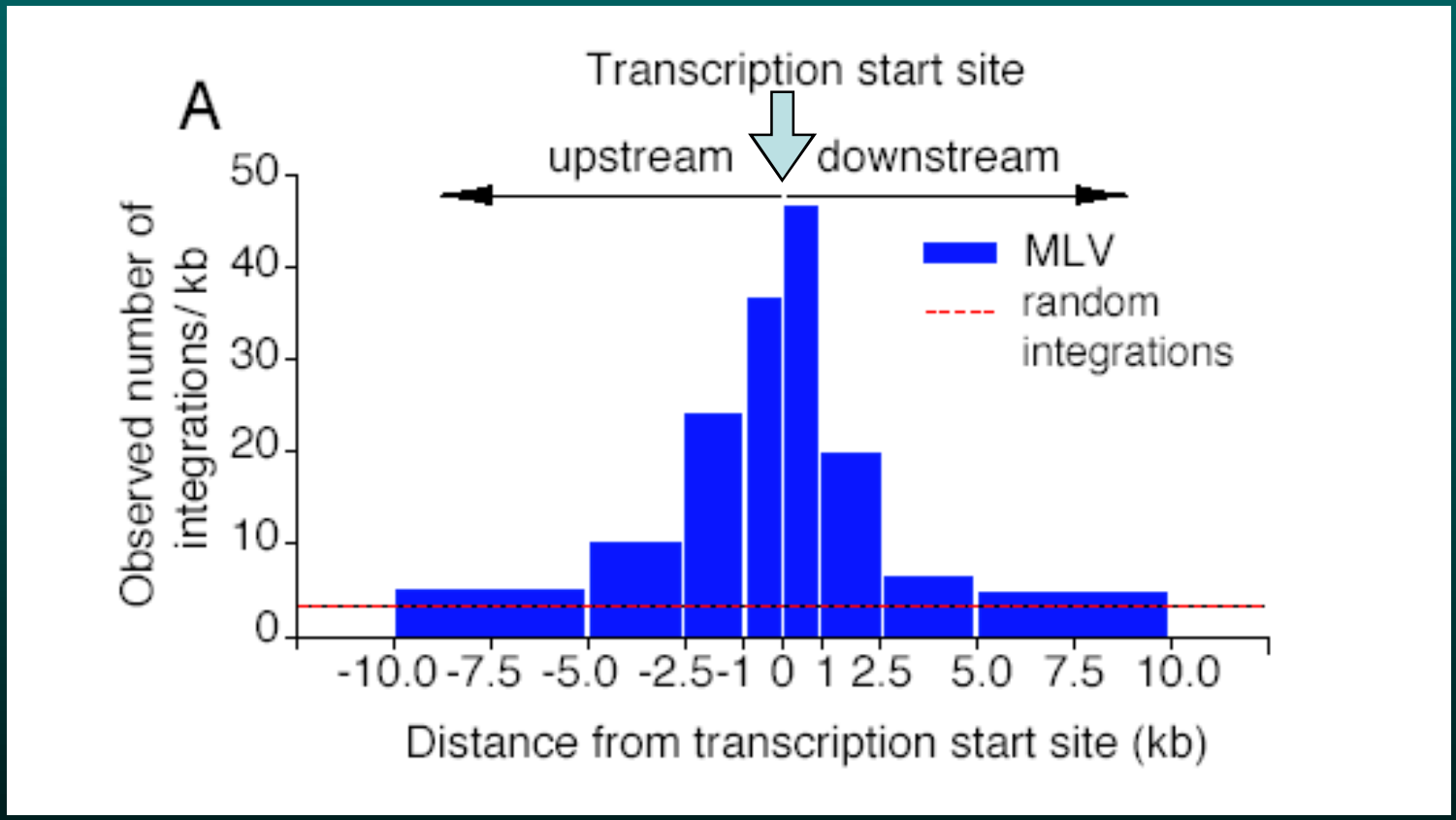
Did not differ significantly from random integrations.

This reflects that the intron regions are much larger than exons.

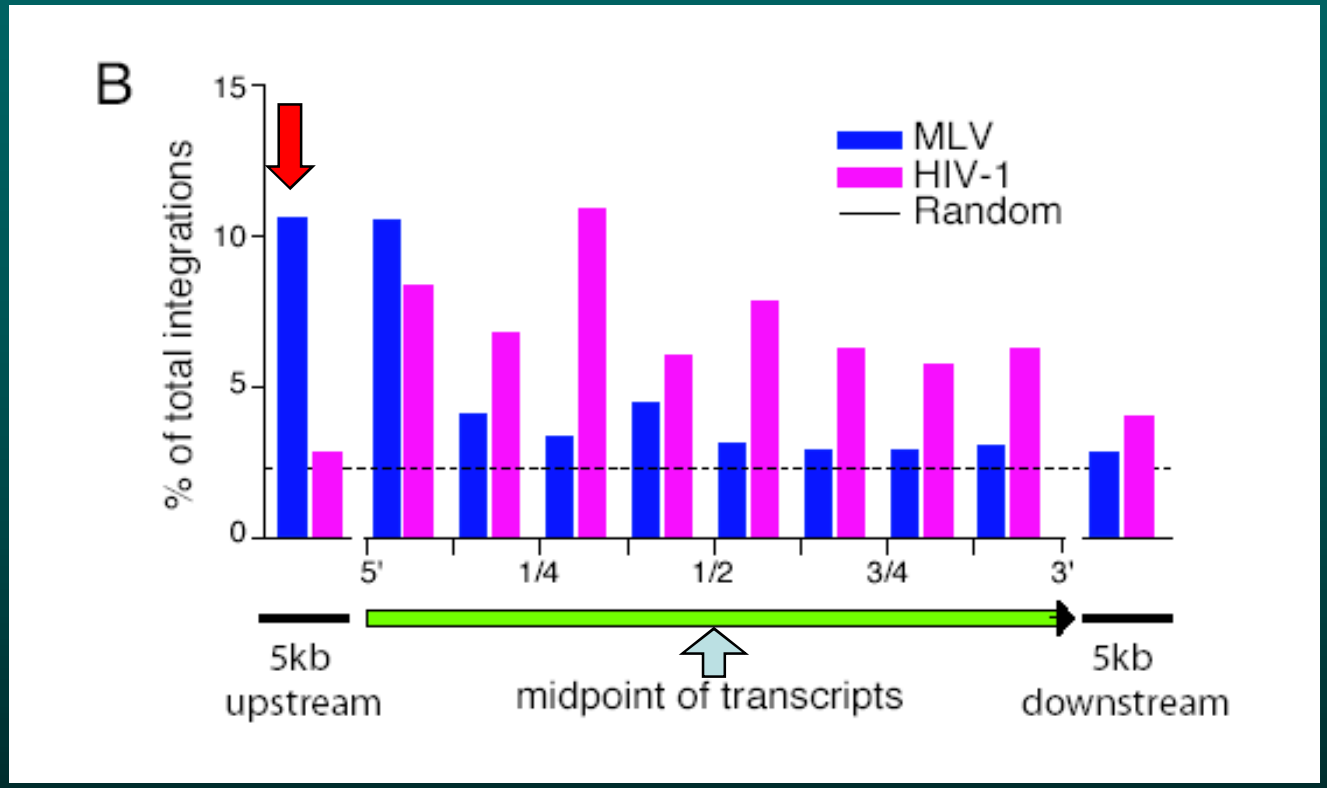
Integrations at 5’ Ends of Genes

Integrations at	MuLV	HIV-1	Expected
CpG islands	16.8%	2.1%	2.1%

Position of MuLV Integration Relative to Transcriptional Start



Comparison of MuLV and HIV-1 Integrations in Transcriptional Unit



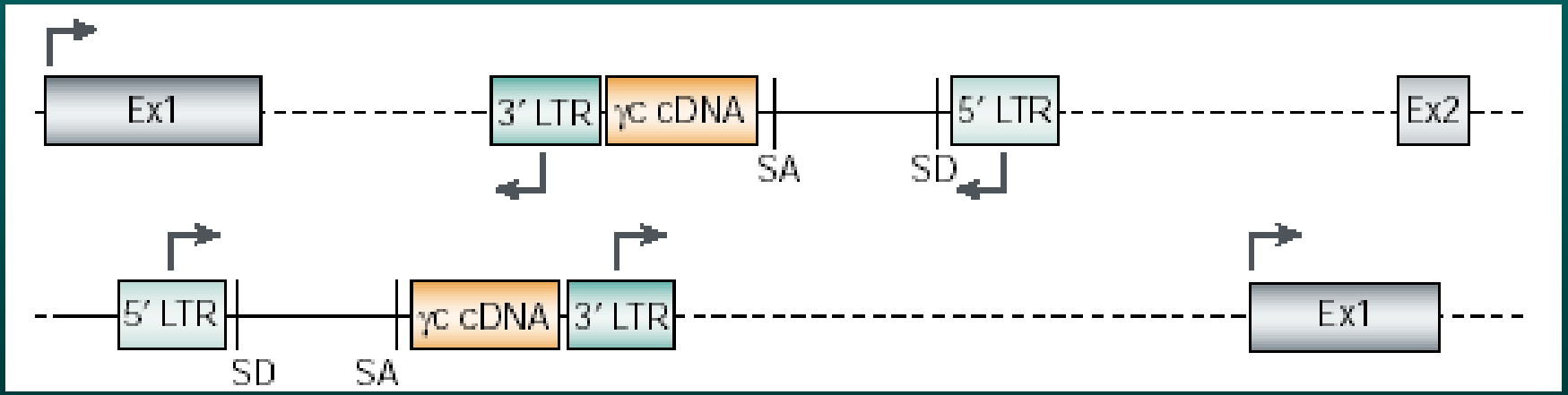
MuLV and HIV-1 integrate into actively transcribed regions

MuLV integration favors 5' ends of gene

HIV-1 shows no strong preference for 5' integration

X-SCID Integration Sites

- γC gene gene therapy vector from two of the X-SCID patients. Note the location at the 5’ end of the gene.



- >5x10^6 cells with MuLV integrations per patient
- 20% of integrations in 18,214 RefSeq genes or
  55 integrations into 5’ region of LM02 per treatment

Summary

- MuLV integration favors transcriptional start sites and 5’ ends of genes and these regions are critical for gene regulation.
- The closer to the transcriptional start site, the greater chance of integration observed, more than 10x higher than that expected for random integration.
- HIV-1 integration favors transcribed gene-coding regions in the human genome. No preference for upstream or downstream regions of genes. More than half (57%) of HIV integrations landed in genes.

Conclusions

- Both MuLV and HIV-1 target integration into important regions of the human genome at frequencies higher than previously thought.
- The risk of insertional mutagenesis by retroviral gene therapy vectors may need to be re-assessed.
- Specific animal model systems will be useful to avoid the problems encountered with the X-SCID gene therapy trials.
- Hope for the future? With over 100 gene therapy trials since 1989, only the X-SCID trial has shown a negative interaction of the vector with the host genome. Is this specific to X-SCID and γC, or has too little time passed for negative outcomes to become apparent?

Supplemental Slides

- Seven supplemental slides with:
  - More integration data
  - Retroviral life cycle
  - IL-2 receptor funtion

Genes and Promoters within the Human Genome

Integration within a gene:

- Location between the transcriptional start and stop of the 18214 RefSeq genes mapped to the human genome

Integration distribution relative to gene location:

- Transcriptional start sites
- Transcriptional end sites
- CpG islands: commonly associated with the 5' end of genes. There are 27,704 CpG islands documented

MuLV Integrations Target Actively Transcribed Genes

	Data set 1	Data set 2	Data set 3
Median expression level of targeted genes	2055	1209	734
Median expression level of all genes	1228	487	378

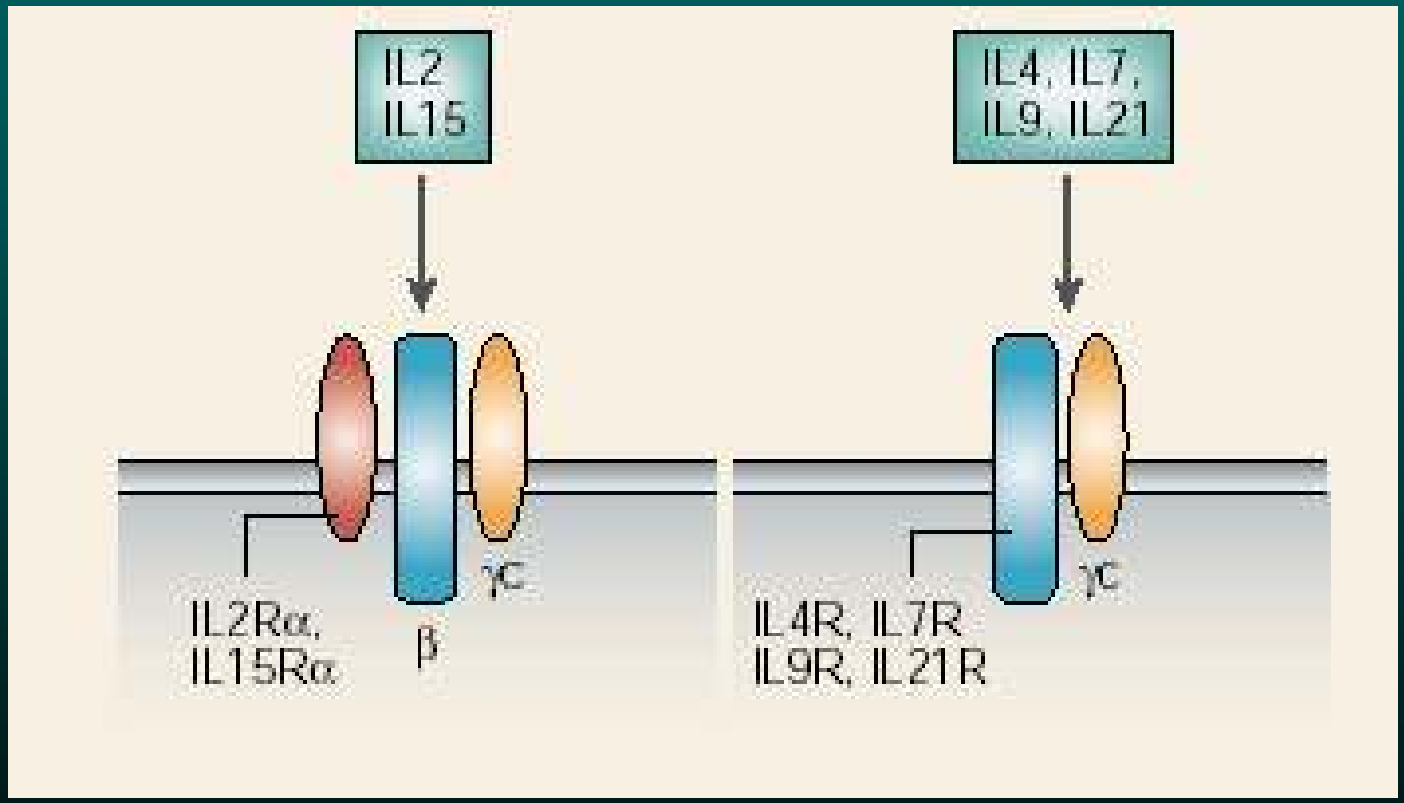
Avoid LM-PCR Bias

- Mse I digestion [TTAA]: 4 bp cutter
  - Expected size 4x4x4x4=256
  - Generates smaller than expected fragments with median size of 70 bp
  - 95% of fragments <500 bp in human genome

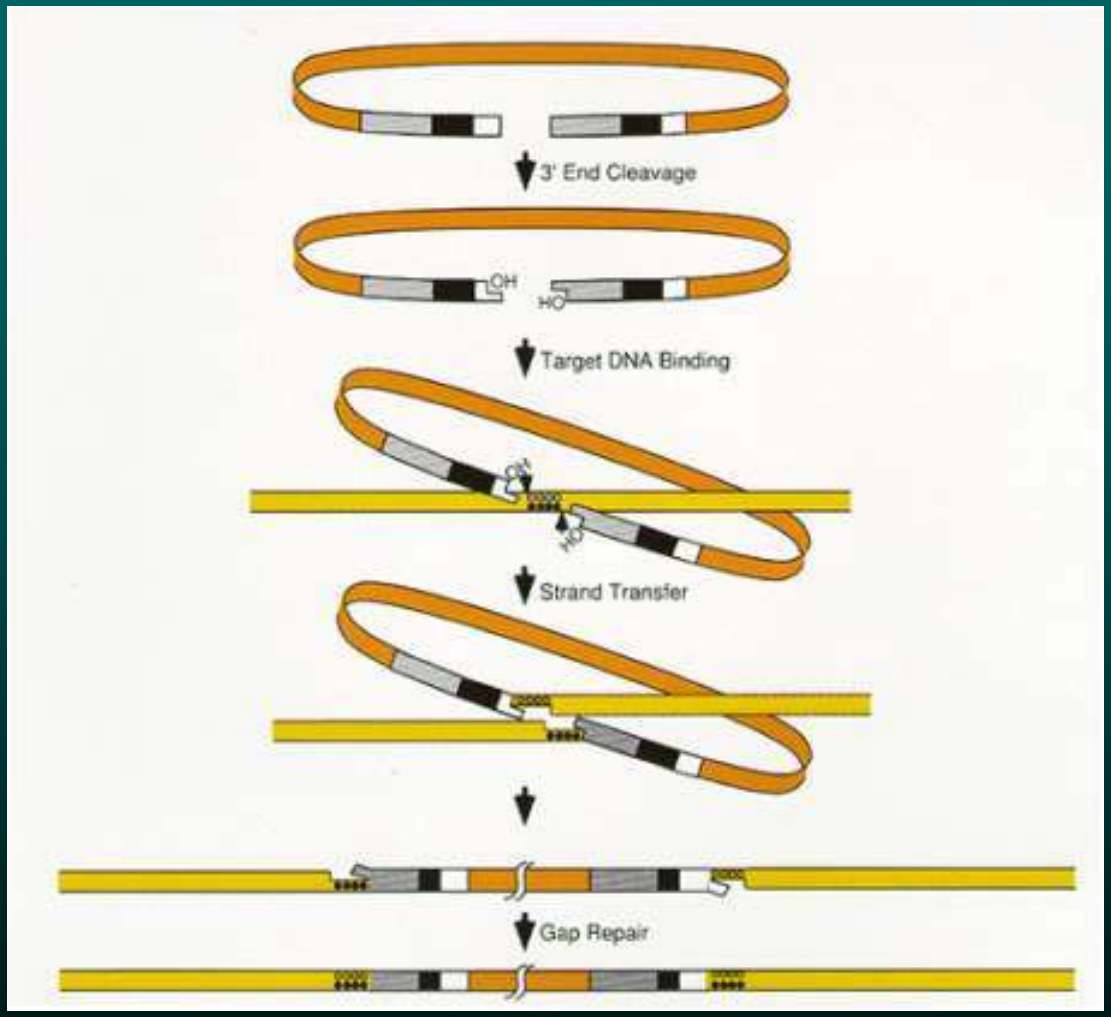
- Small fragments are easy to amplify and clone

Multiple Functions of γC Subunit of IL-Receptors

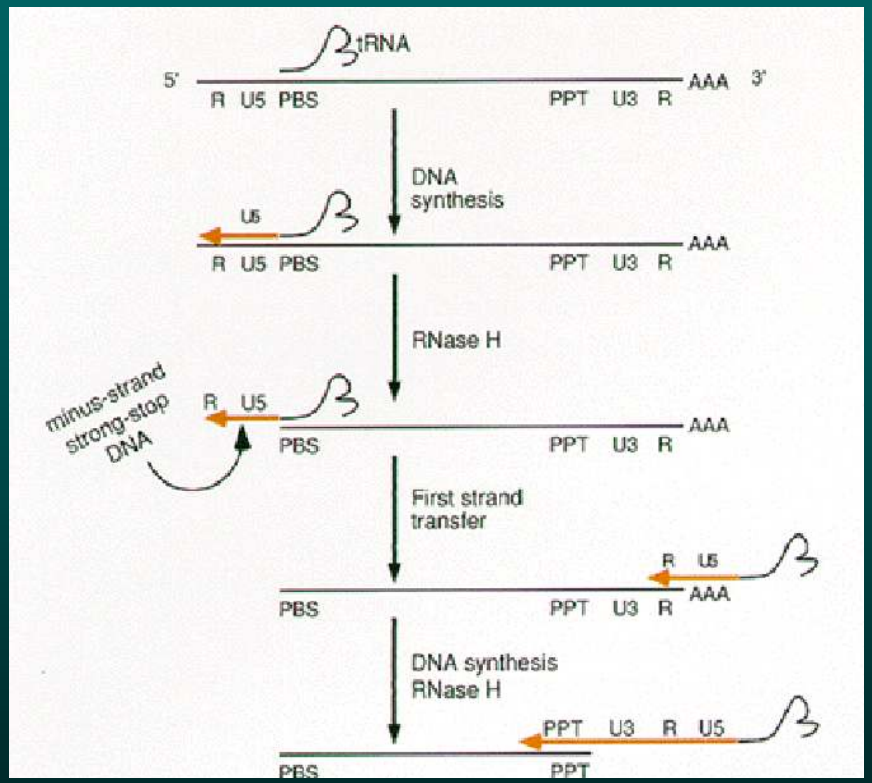
- γC gene product and it function



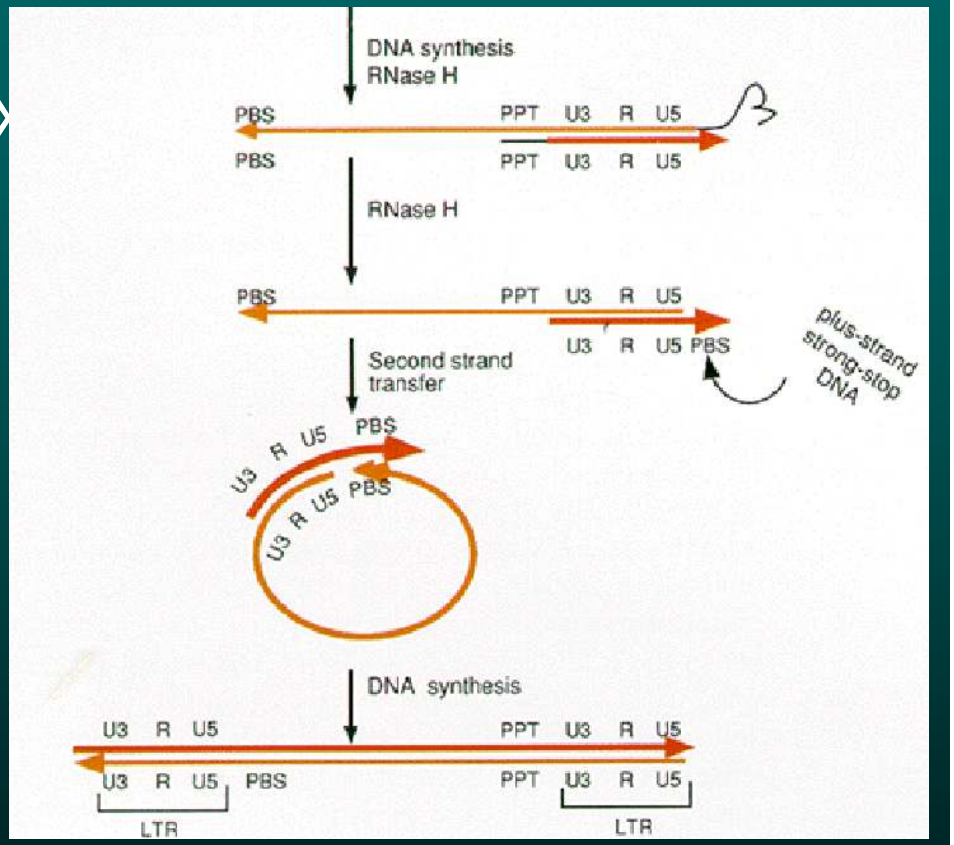
Schematic of Integration Reaction



Retroviral Replication: Preintegration



tex



Discoveries in Retrovirology: Setting the Stage for Paradigm Shifts in Biology

- Isolation of first retroviruses (1904, 1908, 1911)
- Development of focus assay for RSV (1938, 1957)
- Discovery of reverse transcription (1970)
- Retroviral oncogenes shown to be of cellular origin (1976)
- Isolation of human retroviruses (1980)