Reverse transcription and integration

Lecture 9
Biology 4310
Virology
Spring 2020

“One can't believe impossible things,” said Alice.
“I dare say you haven't had much practice,” said the Queen. “Why, sometimes I've believed as many as six impossible things before breakfast.”

—LEWIS CARROLL, Alice in Wonderland
Tumor virus history

- 1908 - Discovery of chicken leukemia virus, Bang & Ellerman
- 1911 - Discovery of Rous sarcoma virus, Peyton Rous (Nobel Prize 55 years later)
- Called tumor viruses
- Found to have RNA genomes
Howard Temin’s insight

- RNA tumor viruses caused permanent changes in cells (transformation)
- Viral DNA was integrated into host genome
- Became permanent part of host DNA
- Provirus hypothesis
David Baltimore’s insight

- (+) RNA viruses: No RdRp in particle
- (-) RNA viruses: RdRp in particle
- An enzyme to copy (+) RNA to DNA must be in virus particle
Baltimore and Temin independently discovered RT in RNA tumor virus particles (Nobel Prize, 1975)

Listen to TWiV #100 (Baltimore) for more insight
Reverse transcriptase

- Retroviruses got their name because of their ability to reverse the flow of genetic information
  
  DNA => RNA => protein

- RT discovery revolutionized molecular biology (e.g. every CoV diagnosis utilizes RT-PCR)
Viruses with RT

- Influenza virus
  - Poliovirus
    - + RNA → − RNA
  - Parovirus
    - + DNA → − DNA
- Adenovirus
- Herpes simplex virus
- Hepatitis B virus
- Reovirus
  - + RNA → − RNA
- Retrovirus
  - + RNA → − DNA

©Principles of Virology, ASM Press
Rous sarcoma virus, a retrovirus
Reverse transcriptase

- Primer can be DNA or RNA
- Template can be RNA or DNA
- Only dNTPs, not rNTPs, are incorporated
RT

- Bacteria, Archaea, eukaryotes have RT activity
- Therefore RT evolved before the separation of Archaea, bacteria, and eukaryotes
- RT might be the bridge between early RNA world and modern DNA world
- RT also in HBV, *Caulimoviridae*
Sequence relationships among polymerases

- Gly-Asp-Asp in (+) strand RNA polymerases
- Asp-Asp in RT, segmented (-) strand polymerases
- Gly-Asp-Asn in nonsegmented (-) strand polymerases
RNAse H: A second activity of RT

- Cleaves RNA only when in duplex form
- RNA can be in RNA:RNA or RNA:DNA duplexes
- Makes endonucleolytic cleavages
- Produces short oligonucleotides with 5′-phosphate, 3′-OH
HIV-1 Reverse transcriptase

DNA synthesis is slow (4 h per 9 kb genome) and error prone (1 misincorporation per $10^4$ to $10^6$ nt)
Reverse transcriptase has revolutionized molecular biology. Which statement about the enzyme is not correct?

A. RT is unique to retroviruses  
B. RT is packaged in the retrovirus particle  
C. The RT protein also has RNAse H activity  
D. The name of the enzyme comes from its ability to reverse the flow of genetic information  
E. Might have bridged the ancient RNA world and the DNA world
RNA coated with NC protein
50-100 molecules RT per virus particle
RNA dimer

- Explains why retroviruses are relatively resistant to UV and ionizing radiation
- Two copies of all genes
- Recombination (copy-choice) during reverse transcription builds one functional genome
Primer tRNA binding to retroviral genome
DNA synthesis: cytoplasmic

Initiation of (−) strand DNA synthesis

The 5' end of the viral RNA genome is degraded by the RNase H activity of RT as the (−) strand DNA is synthesized.
First template exchange

The RNA genome continues to be degraded as (-) strand DNA is synthesized

(+)-stranded DNA synthesis begins, primed by the ppt RNA

ppt = polypurine tract
(+) strand DNA synthesis

The pbs sequence is copied twice:
• once from the RNA genome
• once from the tRNA primer

RNase H endonuclease activity of RT removes both primer RNAs

DNA ends are juxtaposed by annealing at complementary PBS sequences

(+) strong-stop DNA
Second template exchange is facilitated by annealing of PBS sequences

1. PBS

2. PPT U3 R U5 PBS

3. (+) strand synthesis continues from (+) strong-stop DNA

4. Strand displacement synthesis by RT extends the (−) strand DNA to the end of the (+) strand.

5. Sequences bordered by the (−) and (+) strand start sites (pbs and ppt) are copied twice

6. U3 R U5 PBS

7. LTR

(+)-strand DNA

(−)-strand DNA
Animation of reverse transcription: https://youtu.be/RYwVnzYf4V8
Which of the following steps occur during reverse transcription of retroviral genomic RNA?

A. Priming of (-) DNA synthesis by tRNA
B. Two template exchanges
C. Degradation of the viral RNA by RNAse H
D. Generation of two LTRs
E. All of the above
Integration may occur at many sites on both strands, but is not random

<table>
<thead>
<tr>
<th>Site or region</th>
<th>% Integration*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Random</td>
</tr>
<tr>
<td>Within genes</td>
<td>26</td>
</tr>
<tr>
<td>Transcription start sites</td>
<td>5</td>
</tr>
</tbody>
</table>

*The % Integration values are approximate and may vary depending on the specific context and virus type.
BAF binding prevents auto-integration
• One DNA produced from two RNAs by RT

• Strong promoter (the LTR) built during RT

• Proviral DNA directs the host transcription machinery to synthesize many copies of viral mRNA

• Viral mRNA is translated into viral proteins OR encapsidated into virus particles

There is no viral DNA replication and no viral RNA replication
Provirus is a permanent part of host genome

- No mechanism for precise excision of integrated provirus
- Only way out of genome is transcription by host RNA pol II
- Cell genomes are littered with ancient and modern retroelements - via endogenization
Contemporary endogenization in Koalas

~50,000 years ago, cross-species transmission from rodents
Retroelements

- Sequences that move in the genome via RT
- Proviral DNA integrated into the germline = endogenous retroviruses, ERV
- Often replication-defective (all are defective in humans)
- ~42% of human genome comprises mobile genetic elements, including endogenous proviruses and other retroelements
Retroelements in the human genome

Endogenous retrovirus

- LTR
- gag
- pol
- env
- LTR

Retrotransposons

- LTR
- gag
- pol
- LTR

LINEs

- UTR
- orf1
- orf2

Processed pseudogenes

- (A)_n

SINEs

- (A)_n

\{ No RT \}

Retroelements in the Human Genome

- Non-LTR: 33.9%; 2.4 x 10^6
- LTR-Containing: 8.3%; 0.3 x 10^6

- LINEs (L1) 16.9%
- (L2) 3.2%
- Endogenous Retroviral sequences (ERV) 7.7%
- SINEs (Alu) 10.6%
- (MIR) 2.5%
- Others including Retrotransposons 0.6%
- Processed pseudogenes <1.0

<0.05% are active in human genome today

\* Likely progenitors of retroviruses
Syncytins: Exapted retroviral env

- **Syncytin-1 (HERV-W)**
  - Human
  - Chimpanzee
- **Syncytin-2 (HERV-FRD)**
  - Gorilla
  - Orangutan
  - Gibbon
  - Old World Monkeys
  - New World Monkeys
  - Prosimian

![Syncytin-1 and Syncytin-2 evolutionary tree](image)

- **Fetal villi**
  - Fetal vessel
  - Maternal blood
  - Cytotrophoblast (CT)
  - Syncytiotrophoblast (ST)
  - Extravillous trophoblasts (EVT)
  - Uterine artery

![Fetal villi and uterine structures](image)

**References**

- Placenta 33 (2012) 663-671
Retroviral influence on human embryonic development

- HERV-K, infected human ancestors ≈200,000 years ago
- HERVs do not product infectious virus
- HERV-K mRNAs are produced during normal human embryogenesis
- From 8 cell stage to epiblasts
- Virus-like particles observed
- Induces an antiviral protein, IFITM1
A retrovirus makes chicken eggshells blue

Go to:

b.socrative.com/login/student
room number: virus

Which of the following statements about retroelements is not correct?

A. There are many copies in eukaryotic genomes
B. They are currently entering the Koala germline
C. Those in the human genome produce infectious viruses
D. They can be beneficial
E. None of the above
**Hepadnaviridae** - DNA viruses that encode RT
RT commences in the cytoplasm!

No genome integration
Virology Lectures 2020 • Prof. Vincent Racaniello • Columbia University

Yellow = in virus particle

DNA in 5-20% of virus particles
Next time: Assembly