

# How T2T Human Genome Assembly Enables Better Understanding of the Human Mobilome

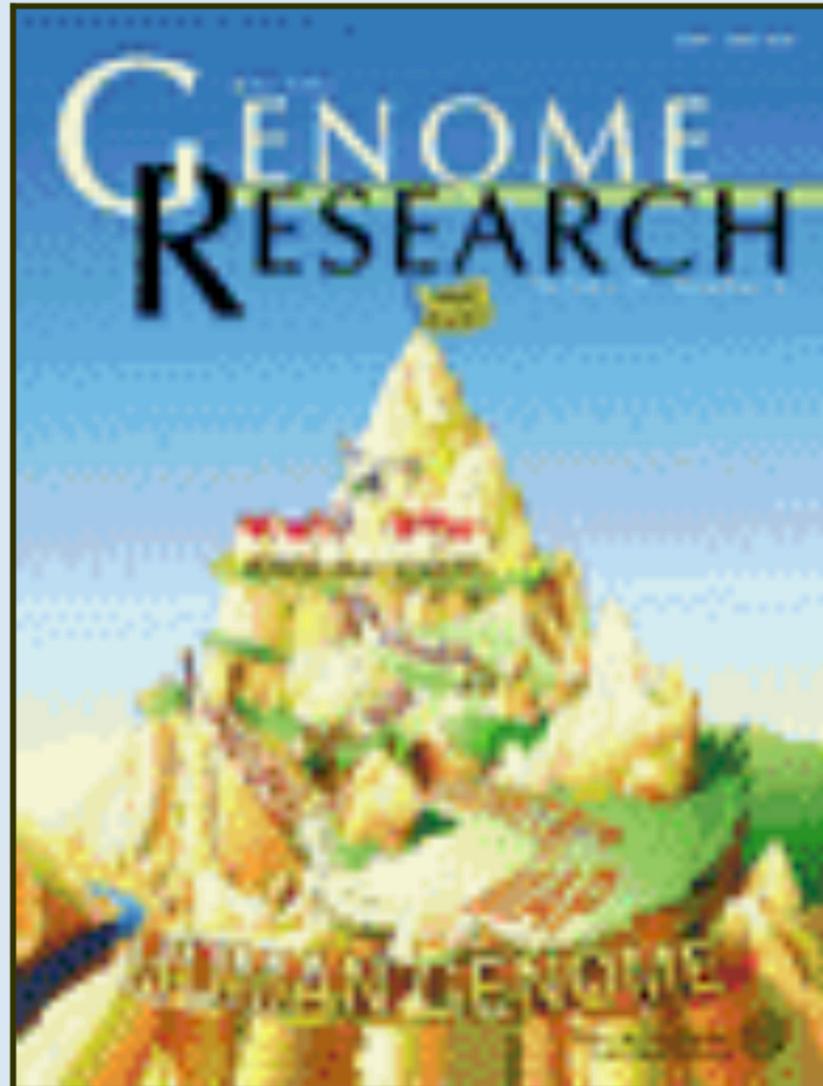
Wojciech Makałowski

Institute of Bioinformatics

University of Münster, Münster, Germany



May 2001; 11 (5)



## ▲ INSIGHT/OUTLOOK

**Are We Polyploids? A Brief History of One Hypothesis**

Wojciech Makalowski

Genome Res. May 1, 2001 11: 667-670; doi:10.1101/gr.188801

» [Extract](#) » [Full Text](#) » [Full Text \(PDF\)](#)

**Complete DNA Sequence of *Lactococcus lactis* Adds Flavor to Genomics**

Oscar P. Kuipers

Genome Res. May 1, 2001 11: 673-674; doi:10.1101/gr.188501

» [Extract](#) » [Full Text](#) » [Full Text \(PDF\)](#)

[Clear](#)

[Get All Checked Abstracts](#)

## ▲ REPORT

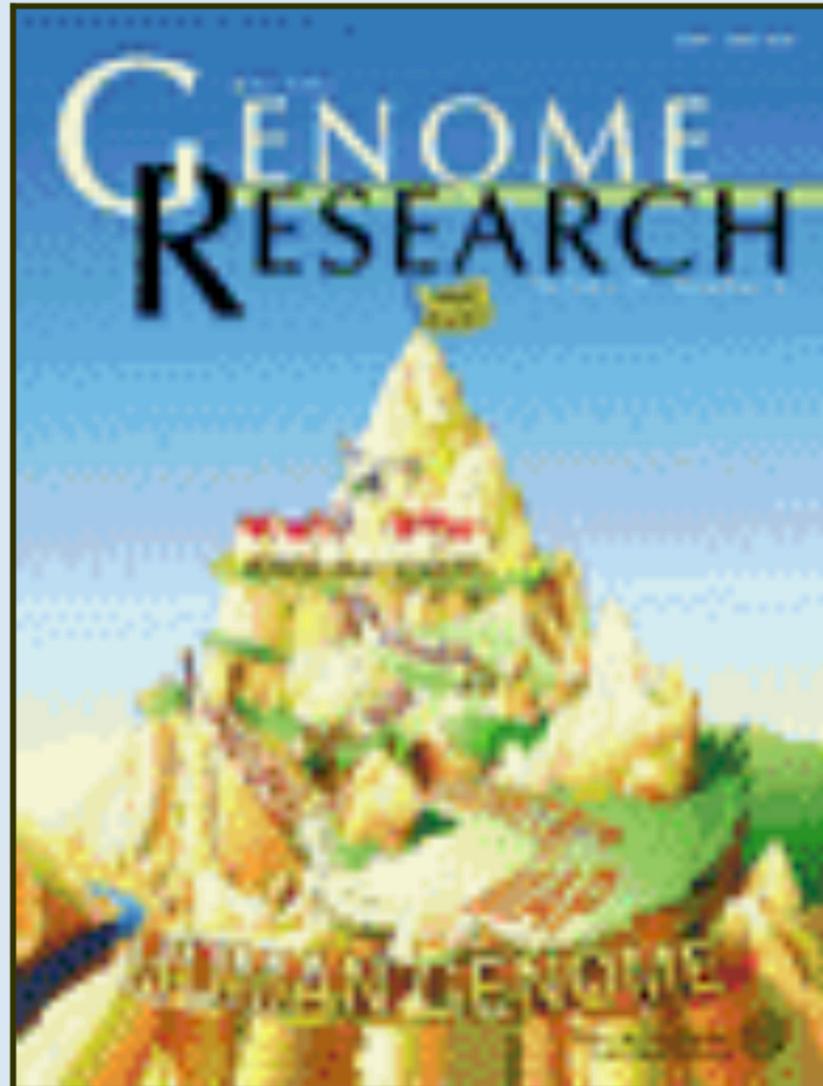
**Identification and Characterization of the Potential Promoter Regions of 1031 Kinds of Human Genes**

Yutaka Suzuki, Tatsuhiro Tsunoda, Jun Sese, Hirotohi Taira, Junko Mizushima-Sugano, Hiroko Hata, Toshio Ota, Takao Isogai, Toshihiro Tanaka, Yusuke Nakamura, Akira Suyama, Yoshiyuki Sakaki, Shinichi Morishita, Kousaku Okubo, and Sumio Sugano

Genome Res. May 1, 2001 11: 677-684; Published in Advance April 11, 2001, doi:10.1101/gr.164001

» [Abstract](#) » [Full Text](#) » [Full Text \(PDF\)](#)

May 2001; 11 (5)



- The Human Genome Sequence Expedition: Views from the "Base Camp"**  
Eric D. Green and Aravinda Chakravarti  
Genome Res. May 1, 2001 11: 645-651; doi:10.1101/gr.188701  
» [Extract](#) » [Full Text](#) » [Full Text \(PDF\)](#)
  
  - Segmental Duplications: What's Missing, Misassigned, and Misassembled—and Should We Care?**  
Evan E. Eichler  
Genome Res. May 1, 2001 11: 653-656; doi:10.1101/gr.188901  
» [Extract](#) » [Full Text](#) » [Full Text \(PDF\)](#)
  
  - Through the Lens of the Sequence**  
Ellen Wright Clayton  
Genome Res. May 1, 2001 11: 659-664; doi:10.1101/gr.187801  
» [Extract](#) » [Full Text](#) » [Full Text \(PDF\)](#)
- [Clear](#) [Get All Checked Abstracts](#)
- 
- ▲ INSIGHT/OUTLOOK**
- Are We Polyploids? A Brief History of One Hypothesis**  
Wojciech Makalowski  
Genome Res. May 1, 2001 11: 667-670; doi:10.1101/gr.188801  
» [Extract](#) » [Full Text](#) » [Full Text \(PDF\)](#)
  
  - Complete DNA Sequence of *Lactococcus lactis* Adds Flavor to Genomics**  
Oscar P. Kuipers  
Genome Res. May 1, 2001 11: 673-674; doi:10.1101/gr.188501  
» [Extract](#) » [Full Text](#) » [Full Text \(PDF\)](#)
- [Clear](#) [Get All Checked Abstracts](#)
- 
- ▲ REPORT**
- Identification and Characterization of the Potential Promoter Regions of 1031 Kinds of Human Genes**  
Yutaka Suzuki, Tatsuhiko Tsunoda, Jun Sese, Hirotohi Taira, Junko Mizushima-Sugano, Hiroko Hata, Toshio Ota, Takao Isogai, Toshihiro Tanaka, Yusuke Nakamura, Akira Suyama, Yoshiyuki Sakaki, Shinichi Morishita, Kousaku Okubo, and Sumio Sugano  
Genome Res. May 1, 2001 11: 677-684; Published in Advance April 11, 2001, doi:10.1101/gr.164001  
» [Abstract](#) » [Full Text](#) » [Full Text \(PDF\)](#)

# Telomere-to-Telomere

- The human genome is finally finished from T to T!
- Technology development: first 92% took 10 years, last 8% took 20 years
- Solved with combination of PacBio HiFi + ONT ultra-long

**nature**

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [articles](#) > article

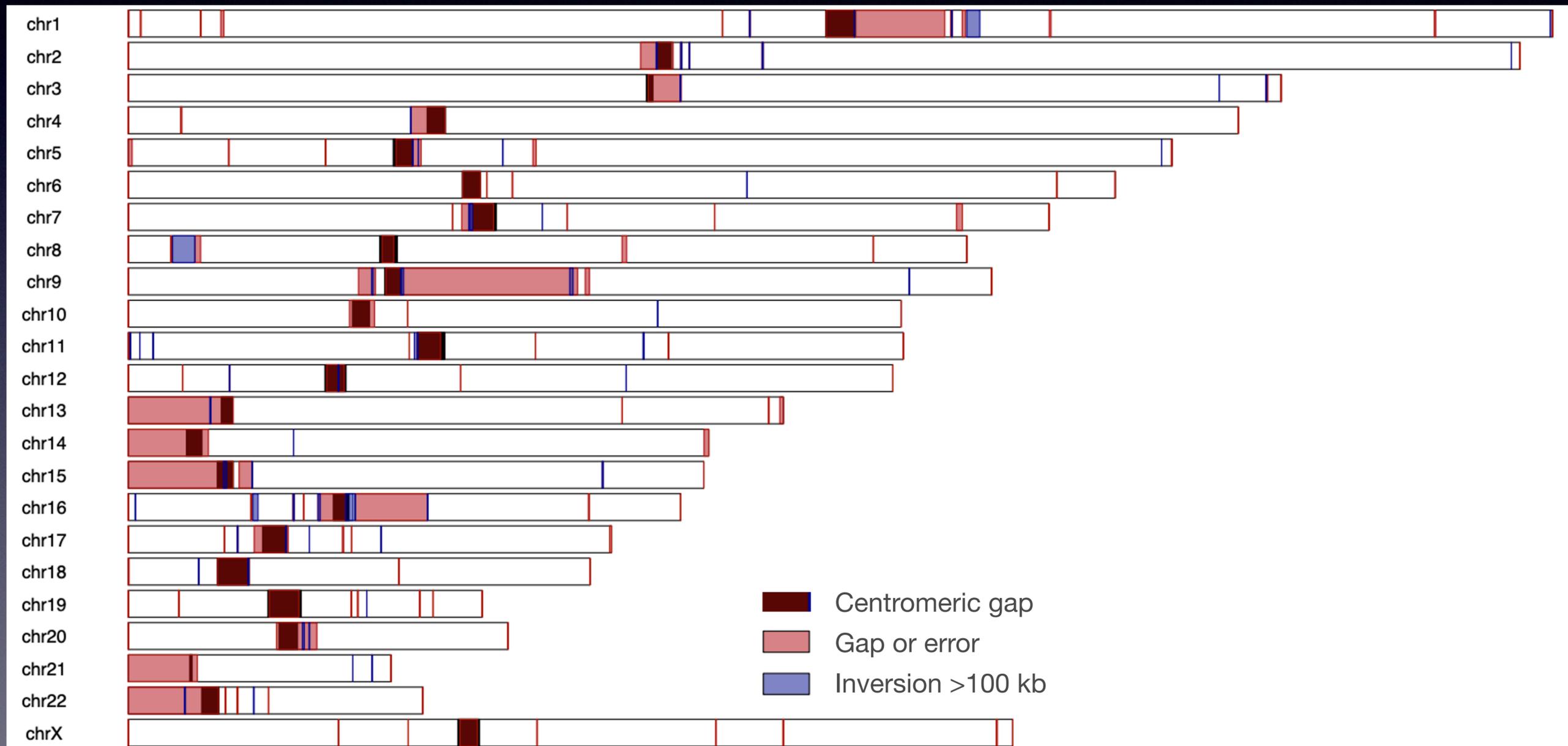
Article | [Published: 23 August 2023](#)

## The complete sequence of a human Y chromosome



The slide courtesy of Adam Phillippy, NHGRI, NIH

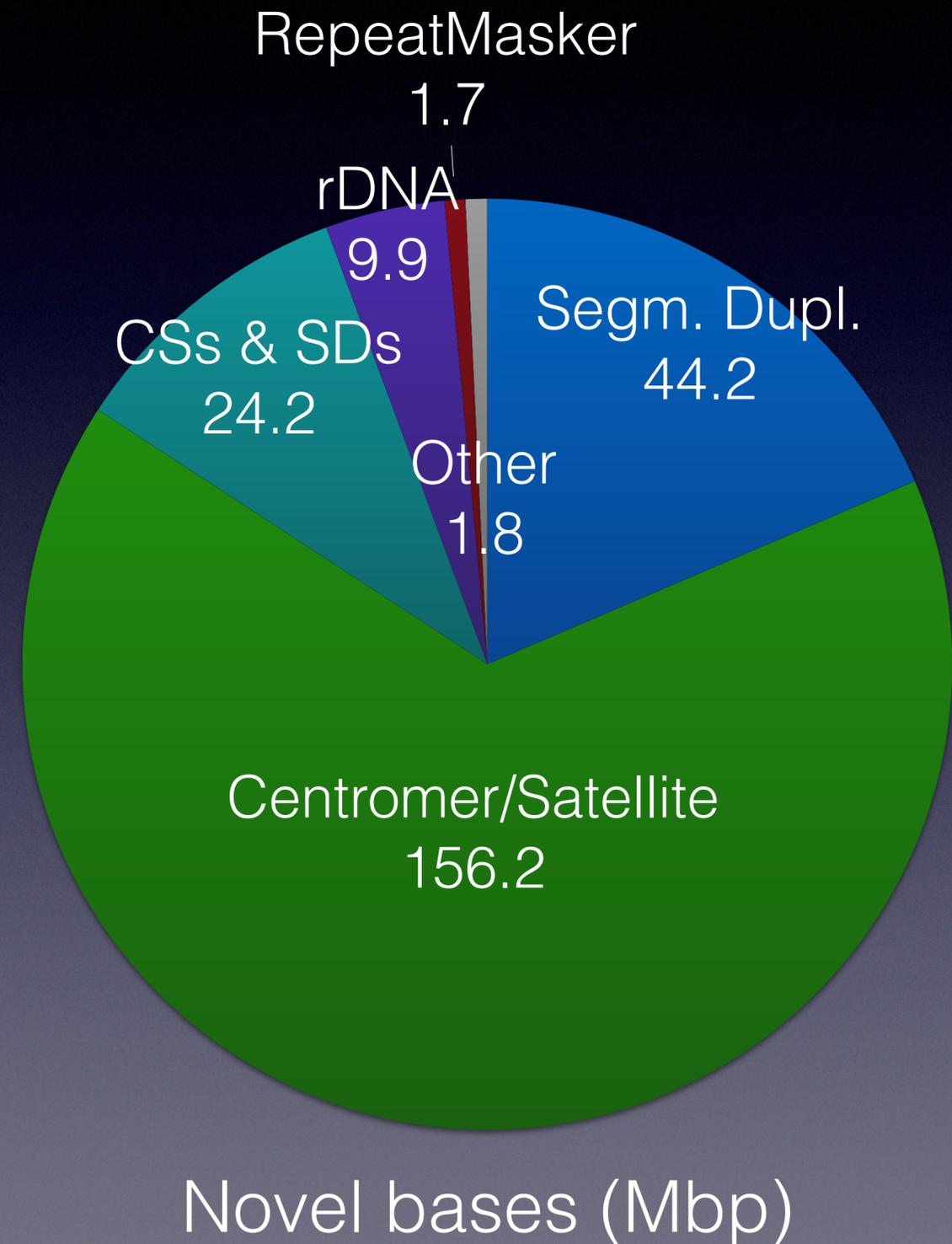
# The last 8% of the human genome



# The last 8% of the human genome

## What's there? Repeats!

- Centromeres
- Telomeres
- Segmental duplications
- Tandem gene arrays
- Acrocentric p-arms
- >200 Mbp of new sequence
- ~2,000 new genes predicted



# From telomere to telomere: The transcriptional and epigenetic state of human repeat elements

Savannah J. Hoyt

Jessica M. Storer

Gabrielle A. Hartley

Patrick G. S. Grady

Ariel Gershman

Leonardo G. de  
Lima

Charles Limouse

**Reza Halabian**

Luke Wojenski

**Matias Rodriguez**

Nicolas Altemose

Arang Rhie

Leighton J. Core

Jennifer L. Gerton

**Wojciech  
Makalowski**

Daniel Olson

Jeb Rosen

Arian F. A. Smit

Aaron F. Straight

Mitchell R. Vollger

Travis J. Wheeler

Michael C. Schatz

Evan E. Eichler

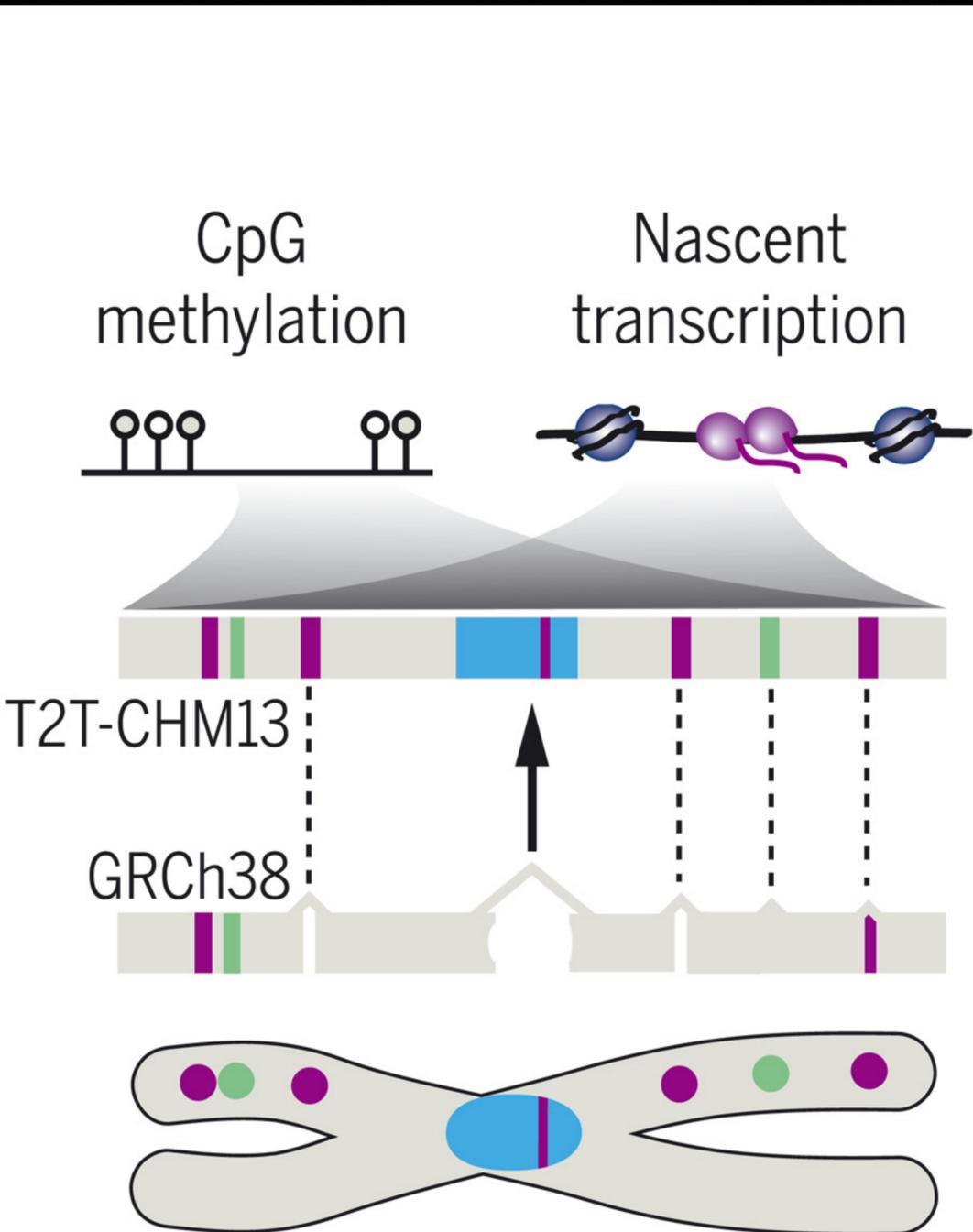
Adam M. Phillippy

Winston Timp

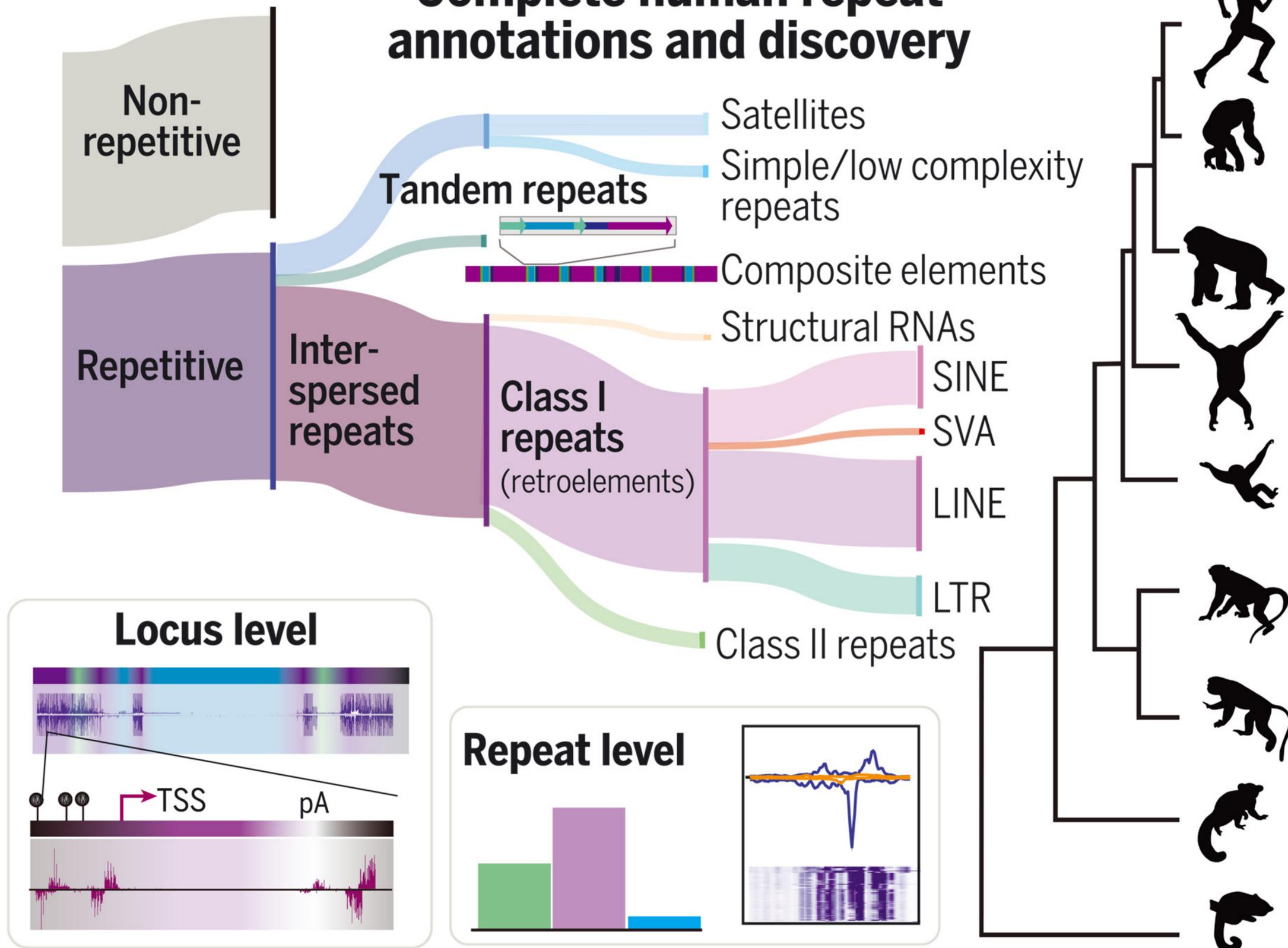
Karen H. Miga

Rachel J. O'Neill





# Complete human repeat annotations and discovery



# TE annotation is more refined for CHM13v1.1

	CHM13v1.1		GRCh38 (excluding Y)	
<b>Repeat class</b>	<b>Mbp</b>	<b>% of assembly</b>	<b>Mbp</b>	<b>% of assembly</b>
LTR	269.91	8.83	267.52	9.15
LINE	631.64	20.68	626.33	21.43
SINE	394.72	12.93	390.78	13.38
DNA	109.34	3.58	108.53	3.71
<b>Total</b>	<b>1405.61</b>	<b>46.02</b>	<b>1393.16</b>	<b>47.67</b>

# TE annotation on chromosome Y

Repeat class	T2T-Y	GRCh38-Y	Difference (%)
LTR	4,613,537	4,604,368	0.2
LINE	6,456,888	6,378,323	1.2
SINE	4,404,417	2,643,856	67.1
DNA	4,387,030	2,626,425	67.0
<b>Total</b>	<b>19,861,872</b>	<b>16,252,972</b>	<b>22.2</b>



L1-mediated DNA transductions

First Glimpses/Reports

# Frequent Human Genomic DNA Transduction Driven by LINE-1 Retrotransposition

Oxana K. Pickeral,<sup>1,2</sup> Wojciech Makalowski,<sup>2</sup> Mark S. Boguski,<sup>1,2</sup> and Jef D. Boeke<sup>1</sup>

<sup>1</sup>Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA; <sup>2</sup>National Center for Biotechnology Information, National Institutes of Health, Bethesda, Maryland 20894 USA

Human L1 retrotransposons can produce DNA transduction events in which unique DNA segments downstream of L1 elements are mobilized as part of such a reaction in tissue culture cells. Structures of L1 element target sites that ~15% of full-length L1 elements findings to the 600,000 copies represents ~1% of the genome

transduction events in which unique DNA segments downstream of L1 elements are mobilized as part of such a reaction in tissue culture cells. Structures of L1 element target sites that ~15% of full-length L1 elements findings to the 600,000 copies represents ~1% of the genome

## Exon Shuffling by L1 Retrotransposition

John V. Moran,\*† Ralph J. DeBerardinis, Haig H. Kazazian Jr.†

Long interspersed nuclear elements (LINE-1s or L1s) are the most abundant retrotransposons in the human genome, and they serve as major sources of reverse transcriptase activity. Engineered L1s retrotranspose at high frequency in cultured human cells. Here it is shown that L1s insert into transcribed genes and retrotranspose sequences derived from their 3' flanks to new genomic locations. Thus, retrotransposition-competent L1s provide a vehicle to mobilize non-L1 sequences, such as exons or promoters, into existing genes and may represent a general mechanism for the evolution of new genes.

Published: 19 September 2002  
The electronic version of this article is the complete one and found online at <http://genomebiology.com/2002/3/10/research>  
© 2002 Szak et al., licensee BioMed Central Ltd  
(Print ISSN 1465-6906; Online ISSN 1465-6914)

Correspondence: Jef D Boeke. E-mail: [jboeke@jhmi.edu](mailto:jboeke@jhmi.edu)

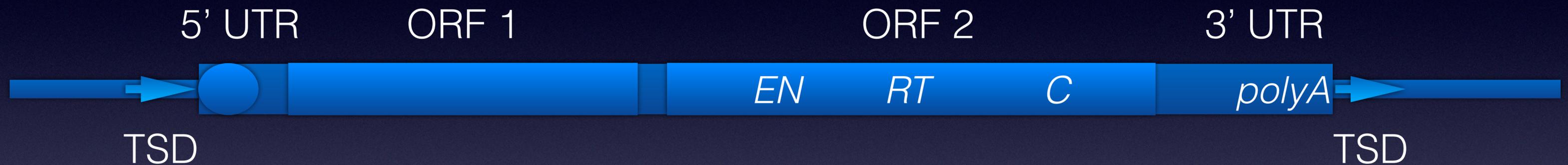
Addresses: \*National Center for Biotechnology Information (NCBI), National Library of Medicine, Bethesda, MD 20894, USA. †Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. Current addresses: ‡Biogen, Inc., Cambridge, MA 02142, USA. §Human Genome Sciences, Inc., Rockville, MD 20850, USA. ¶Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue, North Seattle, WA 98109, USA. \*These authors contributed equally to this work.

Research  
**Molecular archeology of L1 insertions in the human genome**  
Suzanne T Szak\*†§, Oxana K Pickeral\*†§#, Wojciech Makalowski\*†, Mark S Boguski\*†¶, David Landsman\* and Jef D Boeke†

Received: 12 February 2002  
Revised: 2 July 2002  
Accepted: 13 August 2002

<http://genomebiology.com/>

# L1 transposons



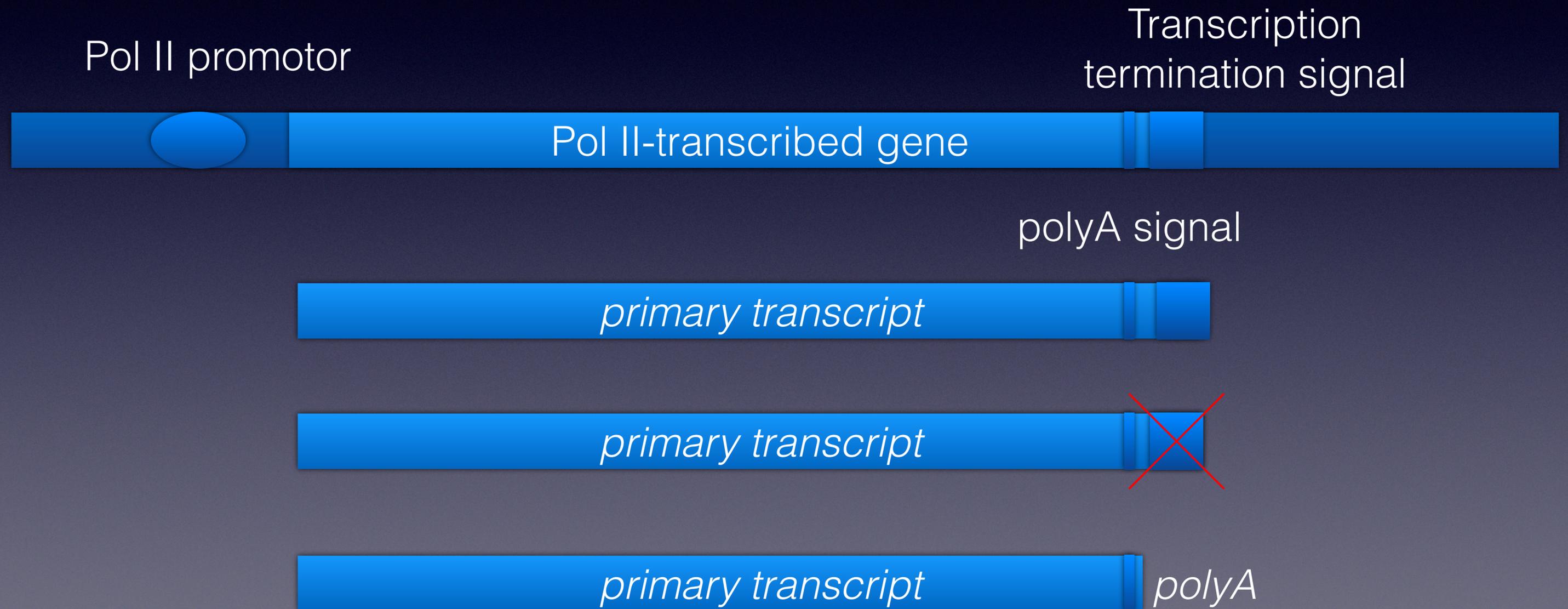
Full length human element is about 6 kb long while it is 7 kb long in mice

Two open reading frames

982,161 copies in the human genome (CHM13 assembly)

80,000 copies in the mouse genome (10% of the mass)

# Pol II transcription



# L1 transcription

Pol II promotor

Transcription  
termination signal

L1 element

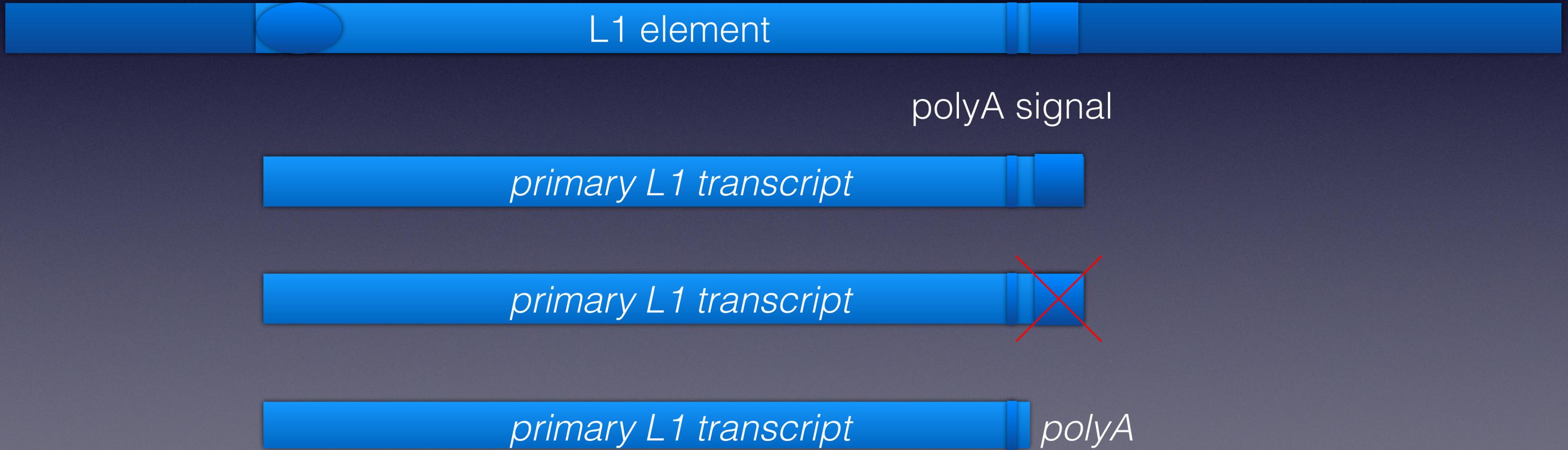
polyA signal

*primary L1 transcript*

*primary L1 transcript*

*primary L1 transcript*

*polyA*



# L1 transcription

Pol II promotor

Weak transcription termination signal

Strong transcription termination signal

L1 element

polyA signal

*primary L1 transcript*

*primary L1 transcript*

*primary L1 transcript*

*extra RNA*

*polyA*



# L1-driven DNA transduction

Pol II promotor

Weak transcription  
termination signal

Strong transcription  
termination signal

L1 element

transcription

*L1 transcript*

*extra RNA*

*polyA*

reverse transcription

*L1 element*

*extra DNA*

*polyA*

genomic reintegration into a new location

*L1 element*

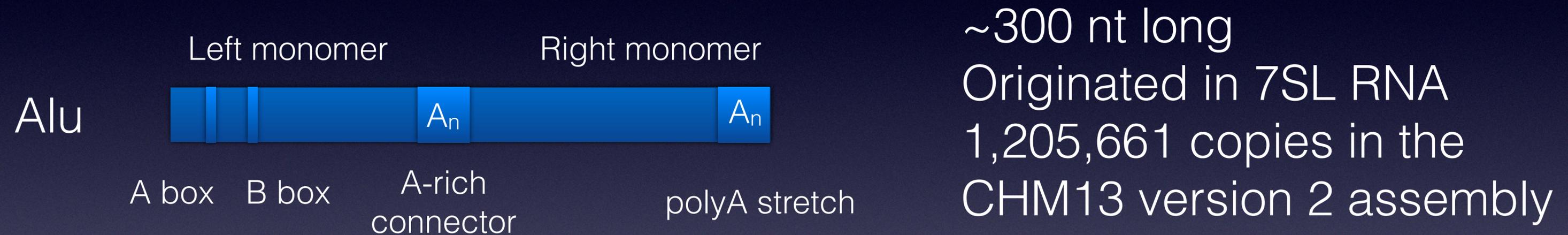
*transduced DNA*

*polyA*

TSD

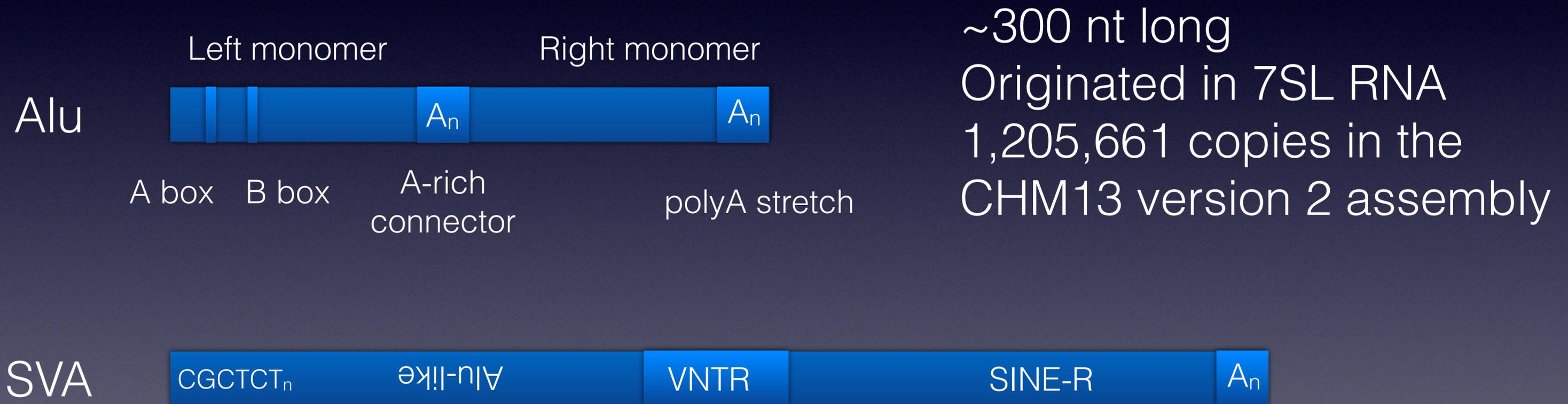
TSD

# There are two other active transposons in humans

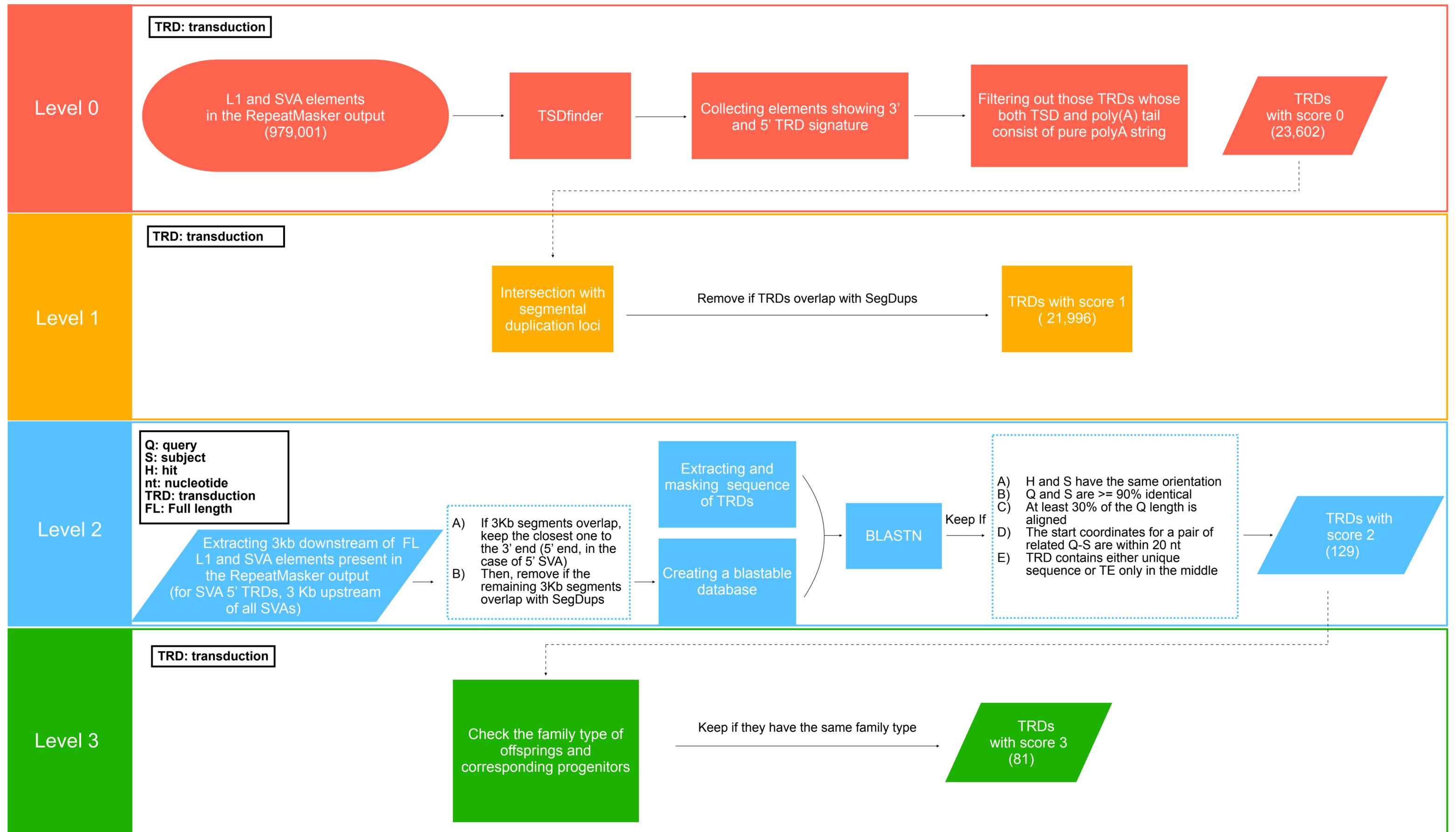


Composite element over 1 kb long  
7,380 copies in the CHM13 version 2 assembly

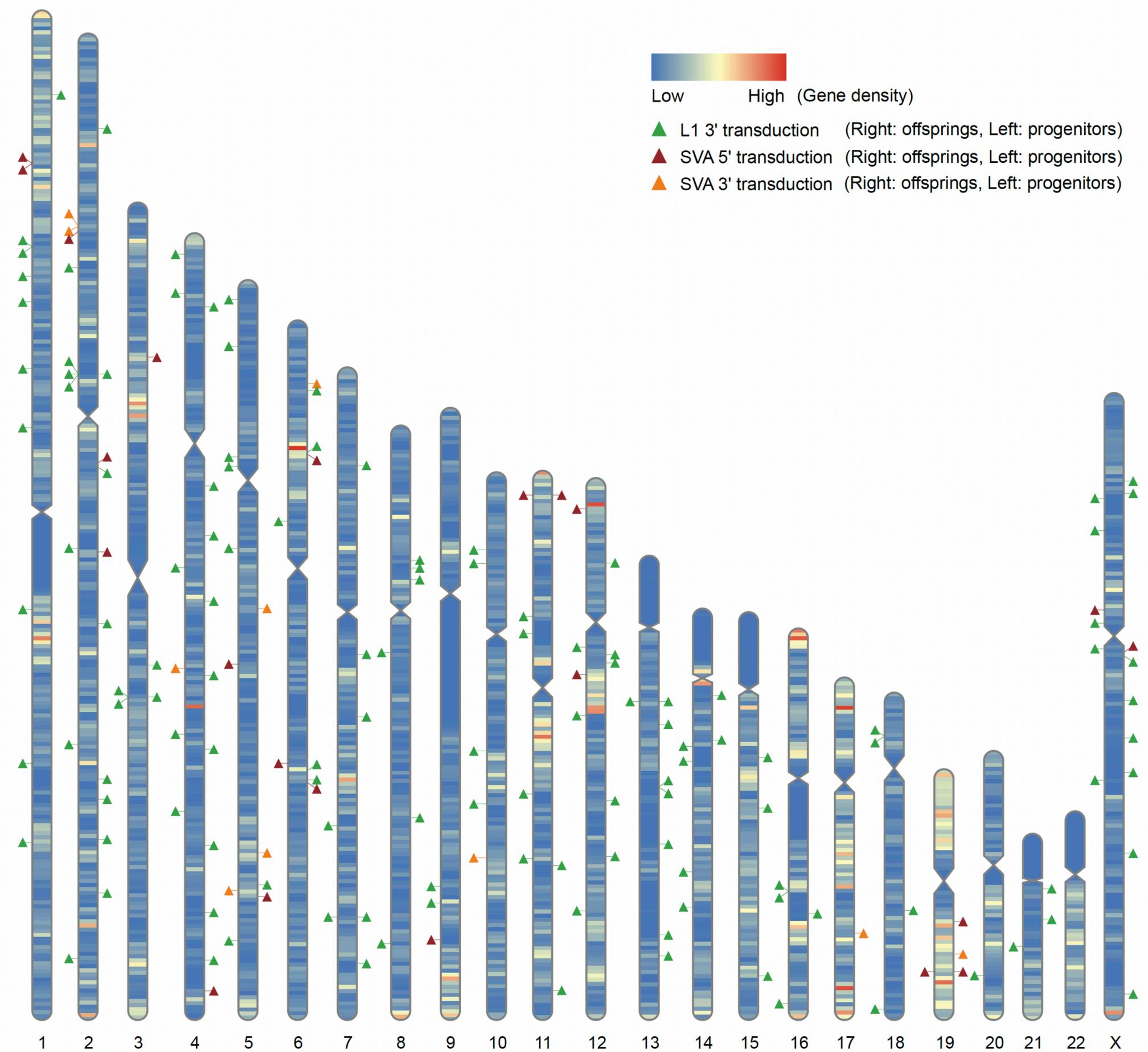
# There are two other active transposons in humans



Do they drive transductions as well?

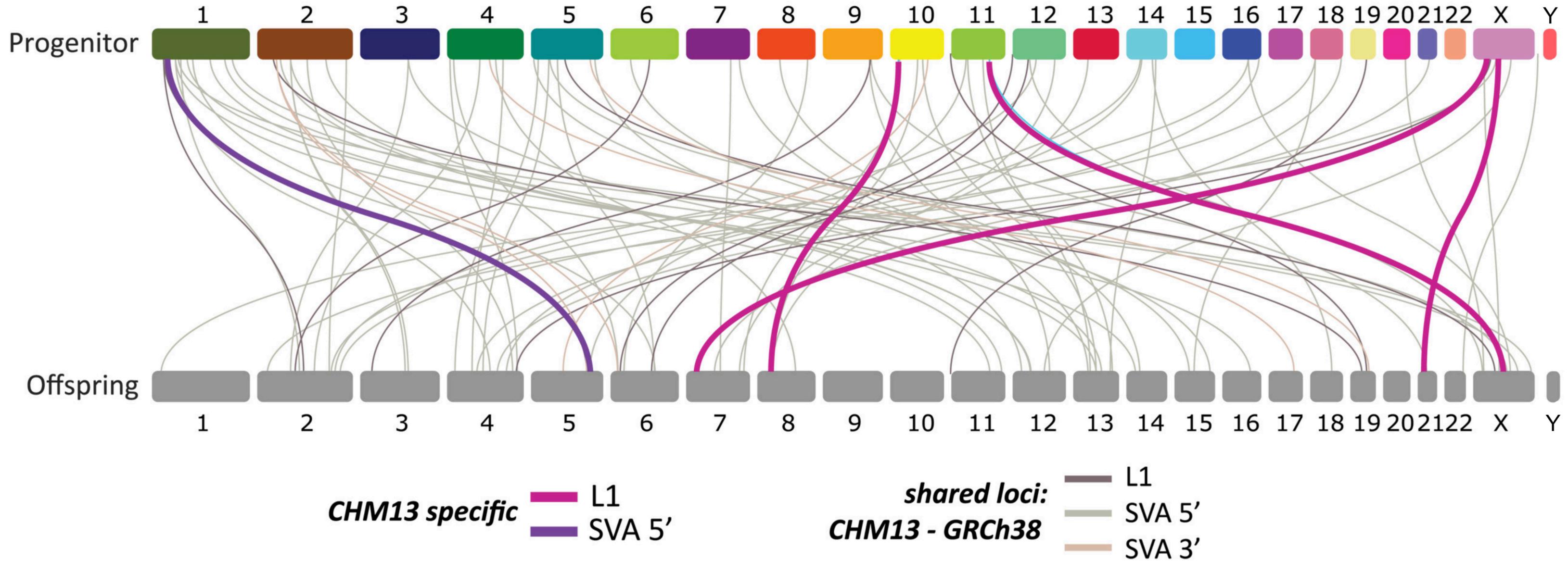


Transduction events are found genome wide in T2T-CHM13





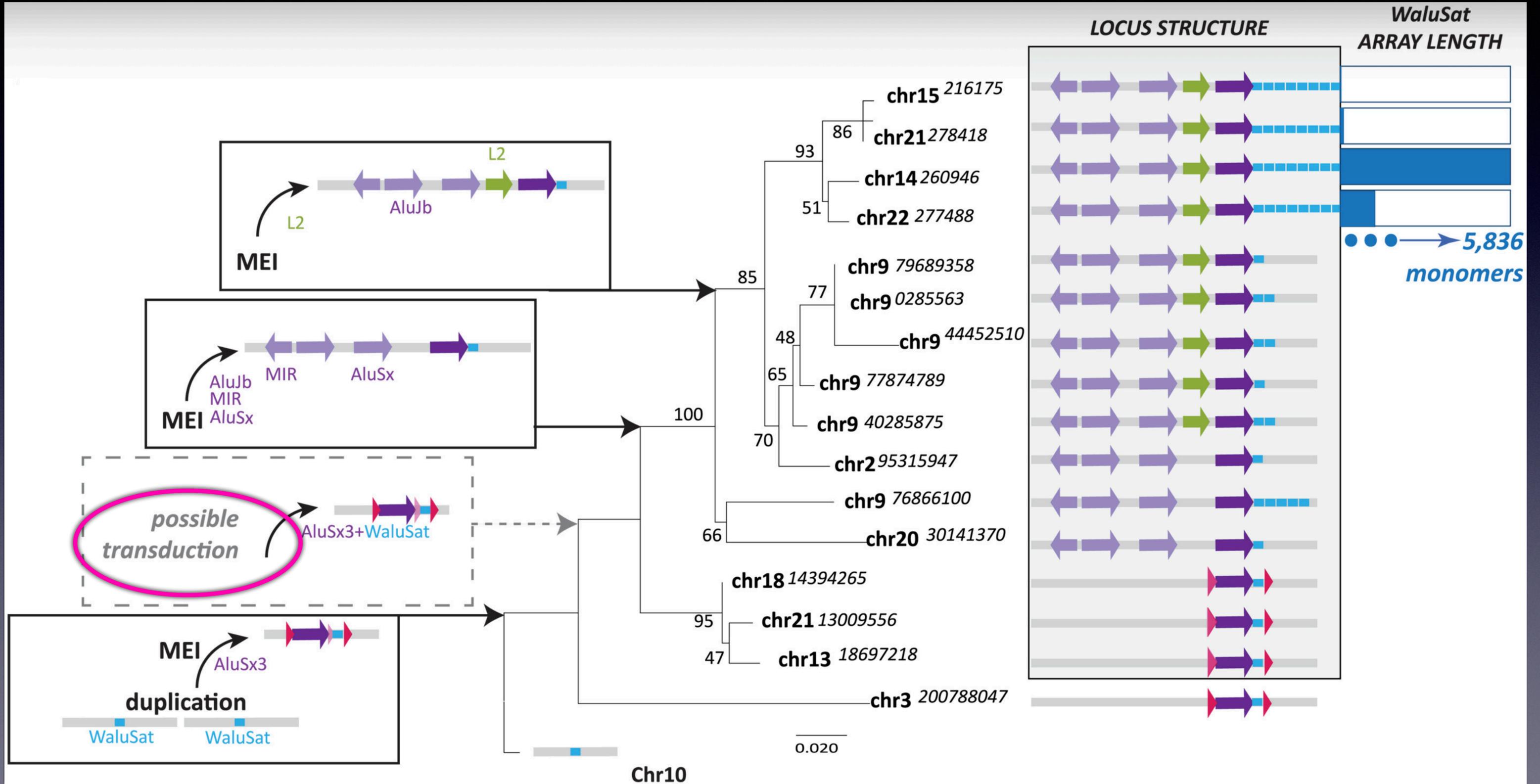
# TE-driven DNA transductions



What about Alu elements?



# WaluSat Repeat



> [Mol Biol Evol.](#) 2011 Jan;28(1):13-5. doi: 10.1093/molbev/msq218. Epub 2010 Aug 16.

# Alu monomer revisited: recent generation of Alu monomers

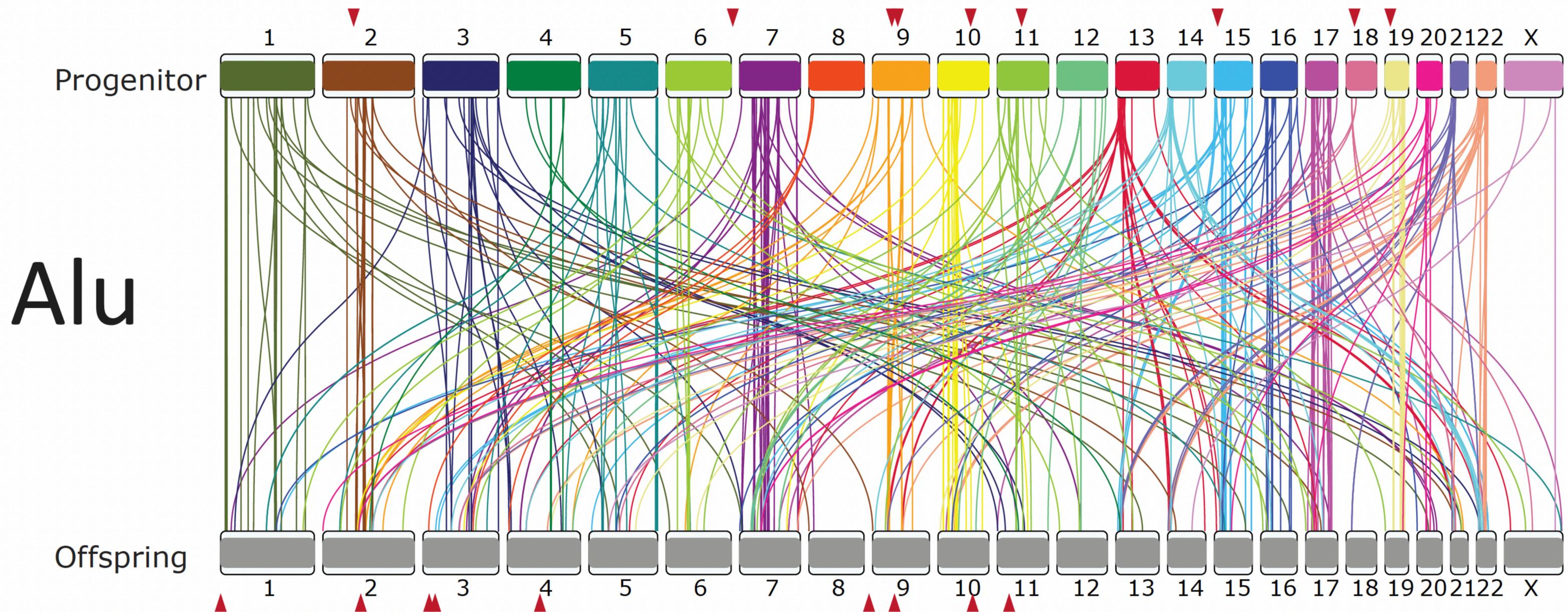
[Kenji K Kojima](#)

PMID: 20713470 DOI: [10.1093/molbev/msq218](#)

## Abstract

Alu is a predominant short interspersed element (SINE) family in the human genome and consists of two monomer units connected by an A-rich linker. At present, dimeric Alu elements are active in humans, but Alu monomers are present as fossilized sequences. A comparative genome analysis of human and chimpanzee genomes revealed eight recent insertions of Alu monomers. One of them was a retroposed product of another Alu monomer with 3' transduction. Further analysis of 1,404 loci of the Alu monomer in the human genome revealed that some Alu monomers were recently generated by recombination between the internal and 3' A-rich tracts inside of dimeric Alu elements. The data show that Alu monomers were generated by 1) retroposition of other Alu monomers and 2) recombination between two A-rich tracts.

# Alu-driven DNA transductions



# BIOINFORMATICS CREED

Remember about biology

Do not trust the data

Use comparative approach

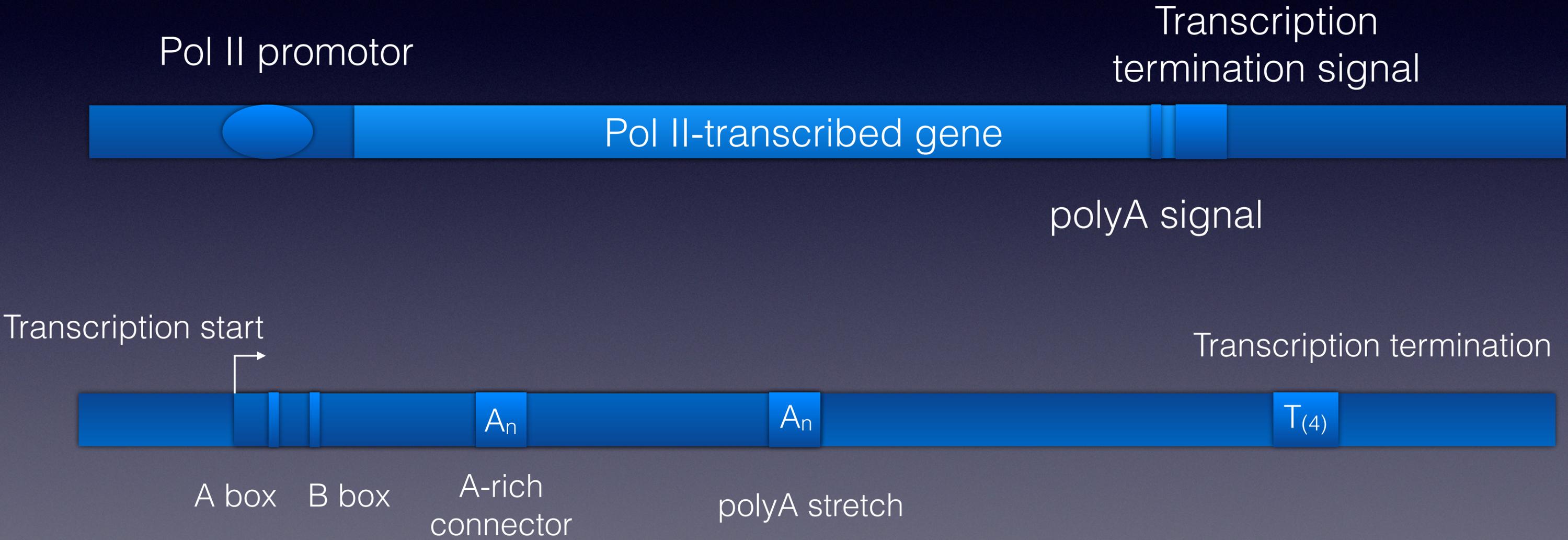
Use statistics

Know the limits

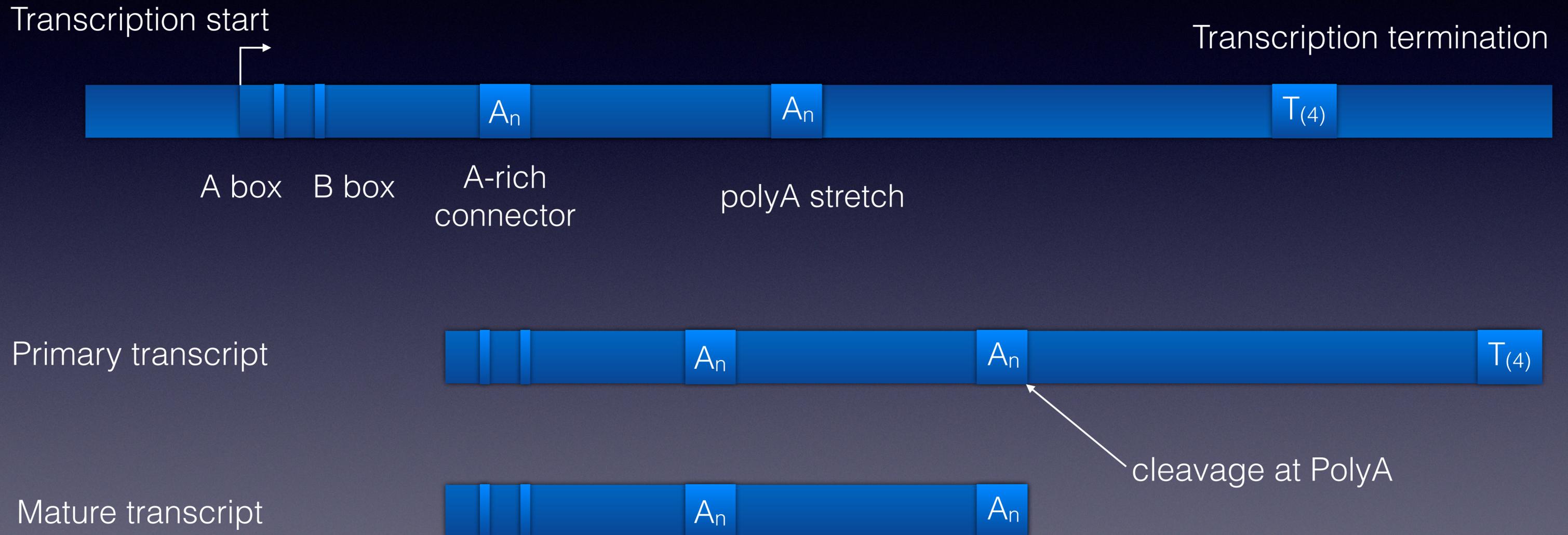
Remember about biology!!!



# Alu transcription by polymerase III



# Alu transcription by polymerase III



A close-up photograph of a monkey with a prominent red face and white, spiky fur. The monkey is looking directly at the camera with a serious expression. It is sitting on a large, textured rock. The background is a soft, out-of-focus blue-grey color.

New software  
required!

# Alu-mediated DNA transductions

	Detected items
Full-length AluYs	118,489
AluYs with TSD	118,489
Potential AluYs with transduction signatures including polyA upstream of 3' TSD	4978
Potential AluYs with transduction signatures not overlapping with other TEs	742
Verified transductions*	24





# Team T2T (...and many more)

