



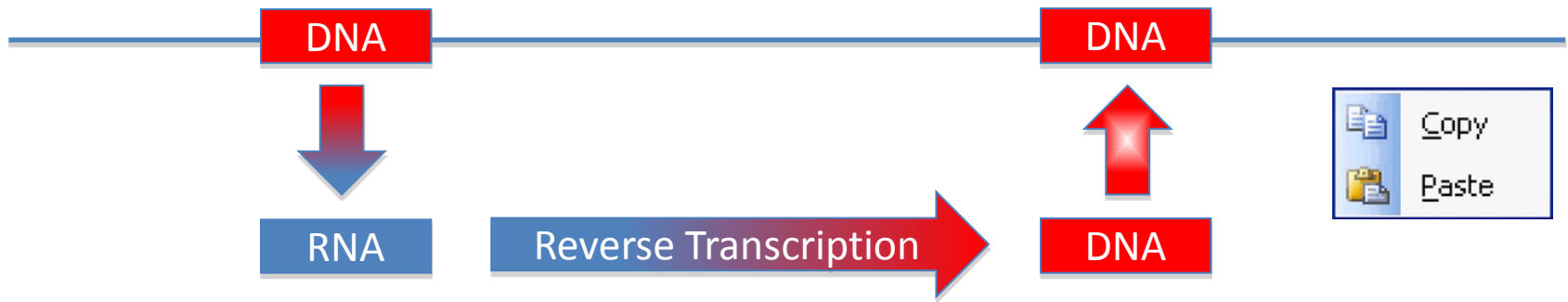
RetroSeq: A Tool to Discover Somatic Insertions of Retrotransposons

Elena Helman

The Cancer Genome Atlas Symposium
November 18, 2011



Retrotransposons

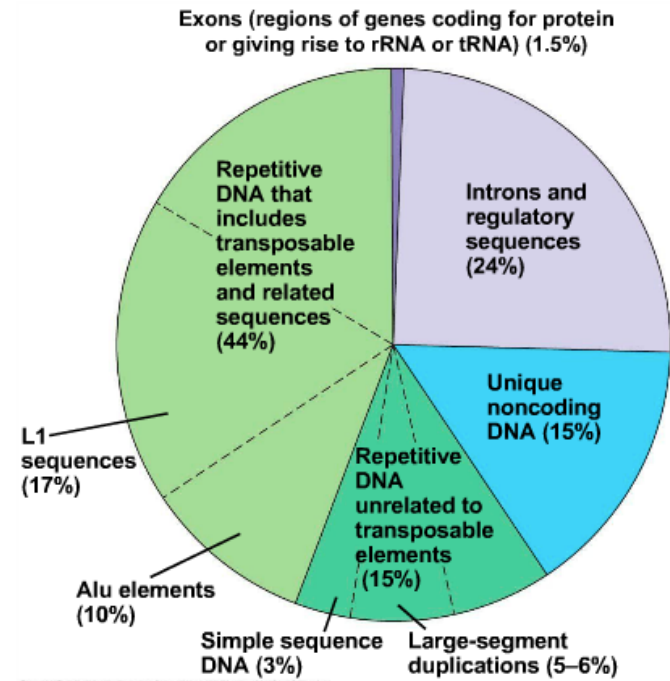


- Retrotransposons
 - Mobile genomic elements that copy and paste themselves across the genome via an RNA intermediate

Drivers of genome evolution



- Comprise >40% of the human genome
- Most are no longer active...but some remain “hot”
- Major source of genetic variation
 - ~10,000 polymorphic sites
 - Estimated 600-1000 retrotransposon differences between two European individuals

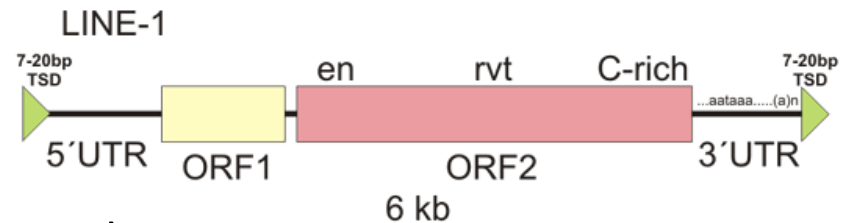


Abundant retrotransposon elements



- L1 (LINE-1)

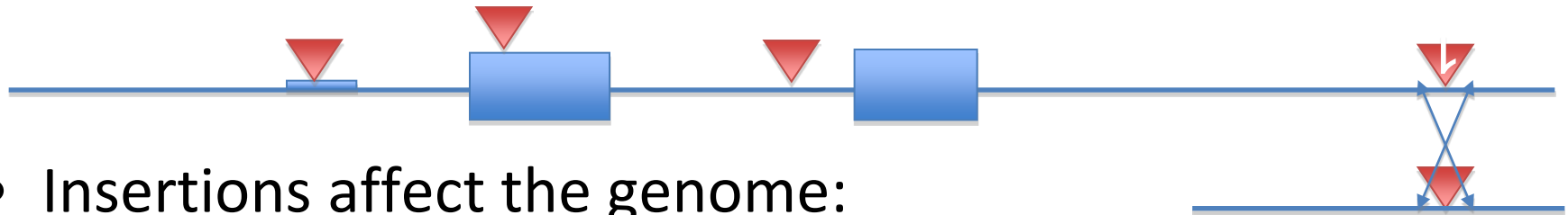
- 6,000 bp long
- 500,000 elements (17% of genome)
- 80-100 still active
- Autonomous
- ORF1: RNA-binding protein
- ORF2: endonuclease and reverse transcriptase



- ALU

- 300 bp long
- >1 million elements (11% of the genome)
- Relies on L1 retrotransposition machinery

Effect of retrotransposon insertions



- Insertions affect the genome:
 - Disrupt protein function
 - Affect promoters
 - Create or disrupt sites for RNA splicing
 - Lead to further genomic rearrangement
- Aberrant retrotransposons insertions in cancer:
 - L1 in APC exon in colorectal cancer (Miki et al., 1992)
 - L1 in MYC intron in breast cancer (Morse et al., 1988)
 - 9 L1 insertions in 6 out of 20 lung tumors (Iskow et al., 2010)

Overall goal

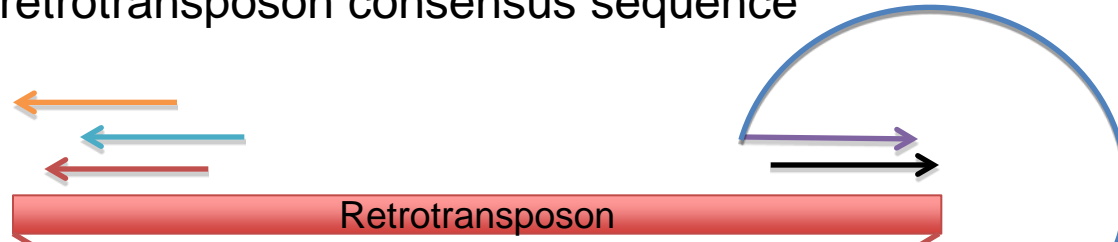


Identify the extent of somatic retrotransposon insertions throughout the cancer genome, using paired-end sequencing data

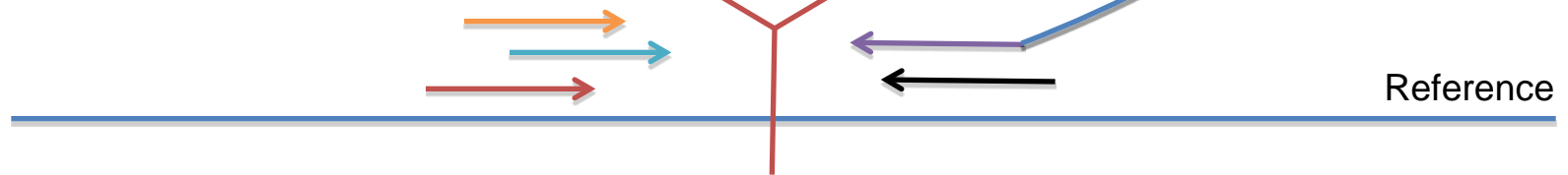
RetroSeq



1. Align reads to retrotransposon consensus sequence



2. Locate clusters of pair-mates



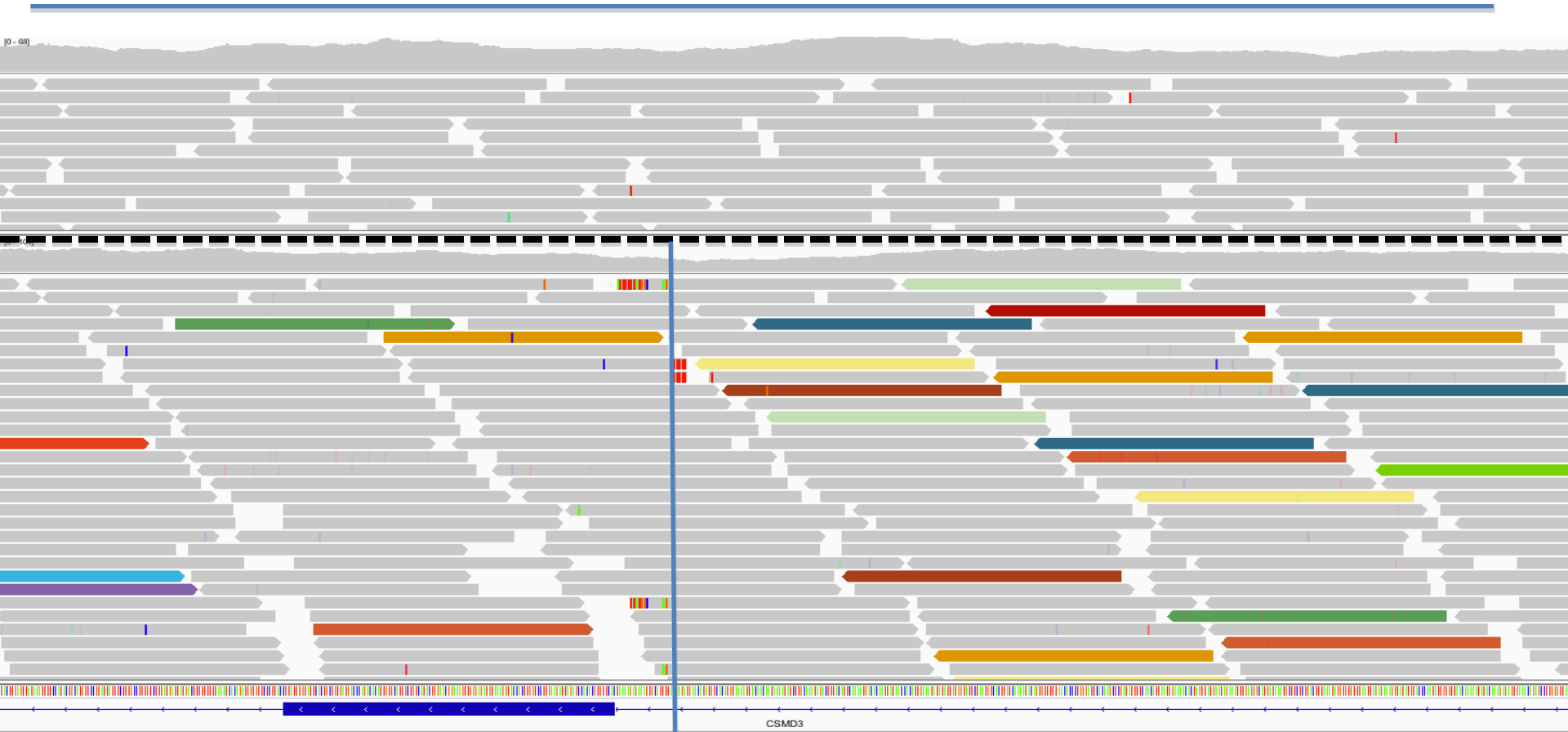
3. Identify putative retrotransposon insertion position



Somatic retrotransposon insertion



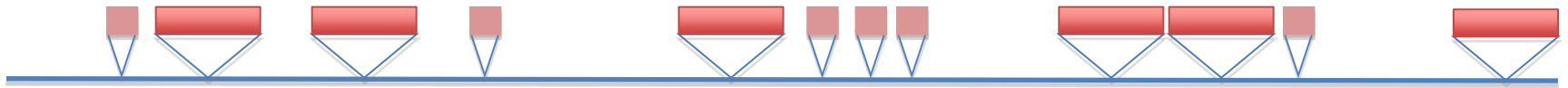
Normal genome



Tumor genome

Retrotransposon

Simulation Performance



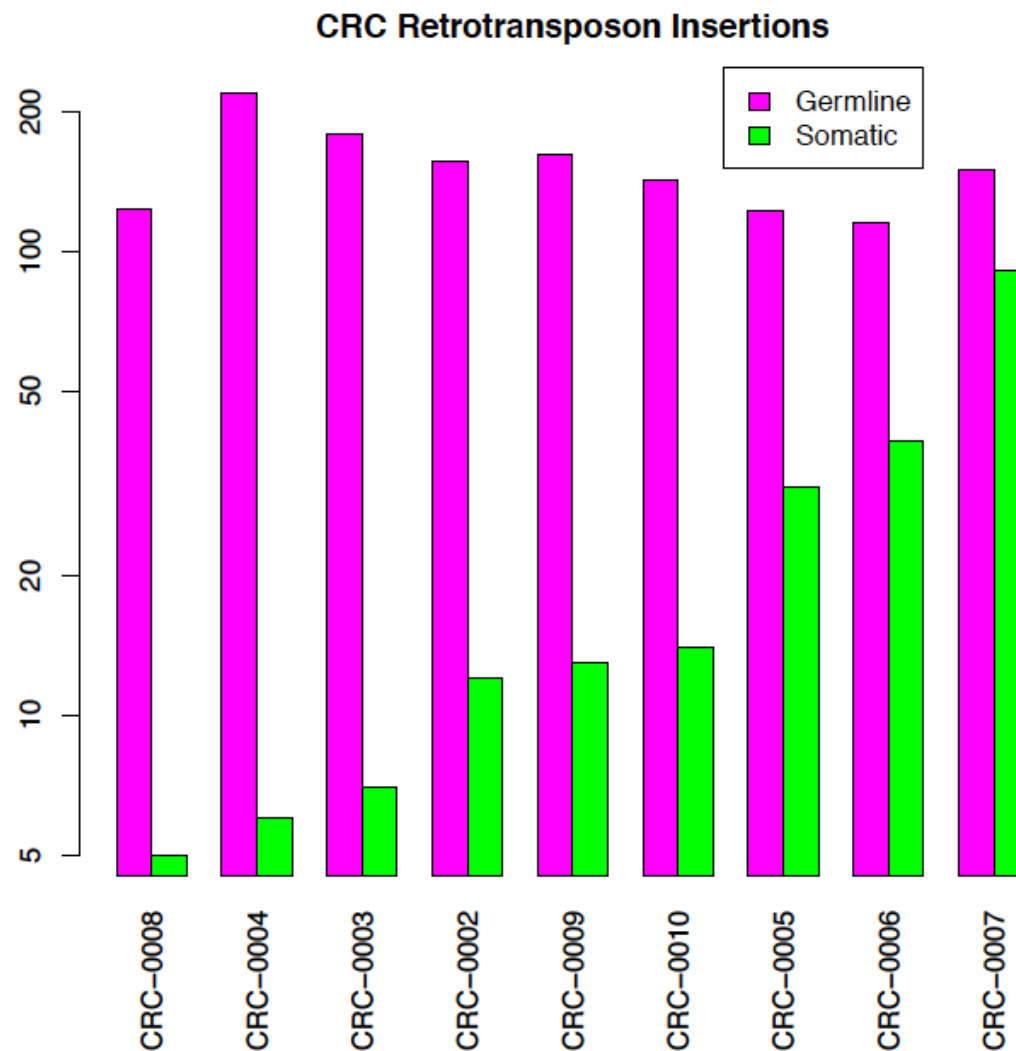
- Inserted 226 L1s and 732 ALUs into BAM file

	Inserted	Sensitivity	Specificity
L1	226	100%	98.3%
ALU	732	99.9%	99.8%

LINE-1 insertions in CRC



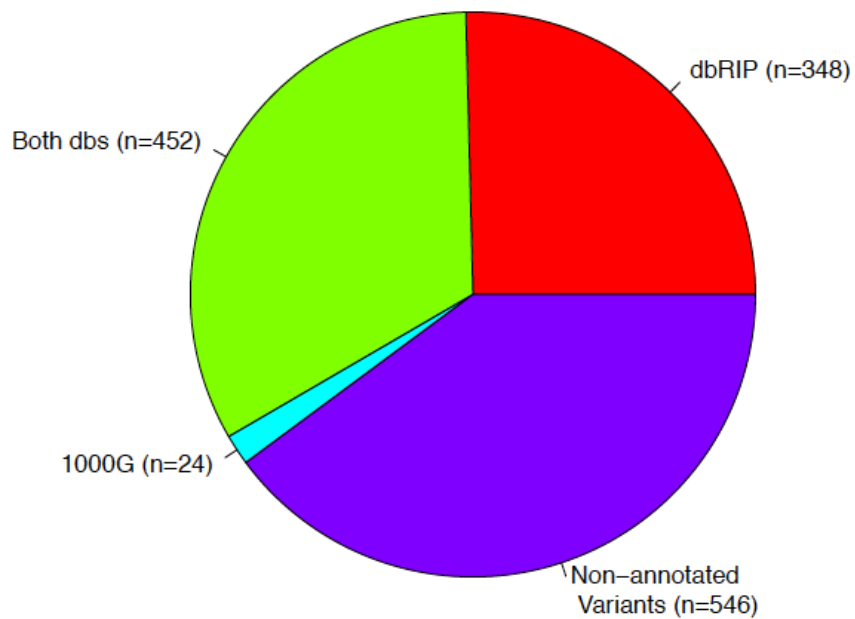
- 9 WGS colorectal tumor/normal pairs
- Retrotransposon consensus sequence database
 - L1 family
 - GIRI Repbase



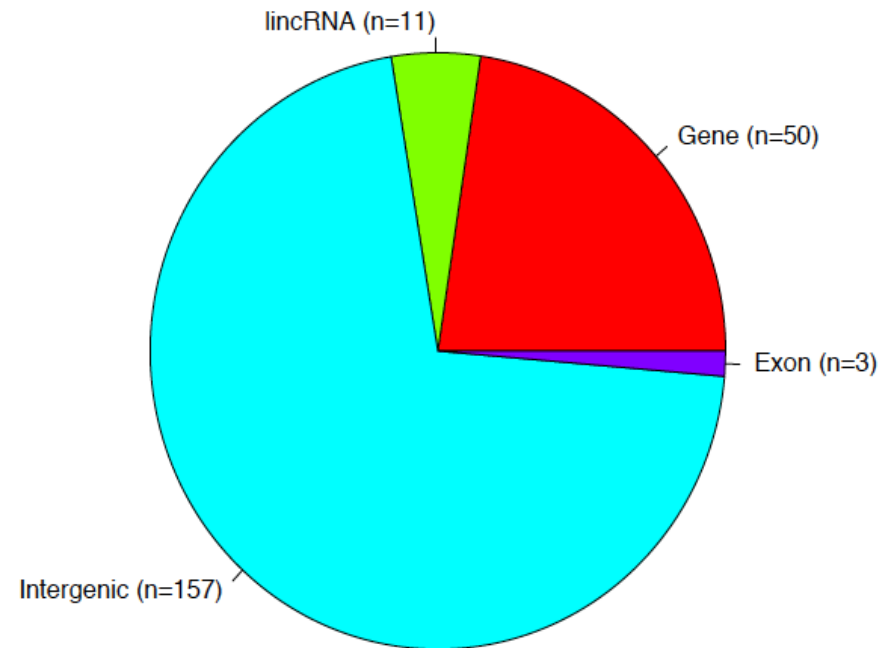
Composition of LINE-1 insertions



CRC L1 Germline Events (n=1470)

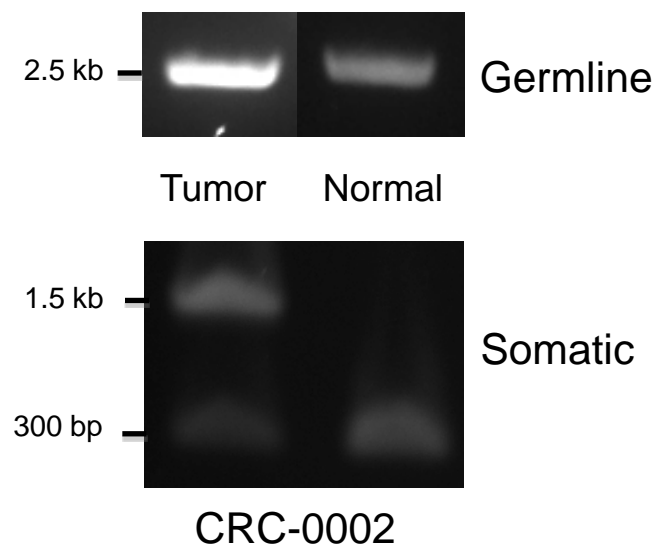


CRC L1 Somatic Events (n=221)



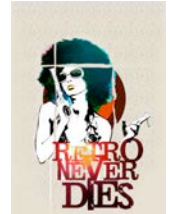
Future studies

- Experimental validation in progress
- Extension to other tumor types
- Orthogonal data integration
 - Expression
 - Methylation



Conclusions

- RetroSeq leverages paired-end sequencing data to computationally localize somatic retrotransposon insertions
- Discovered novel retrotransposon insertions present in tumor, but not matched normal tissue
 - Insertions in genes and regulatory regions
- Evidence for reactivation of retrotransposon mobilization in cancer



Acknowledgements



The Cancer Genome Atlas  *Understanding genomics
to improve cancer care*

- Mike Lawrence
- Chip Stewart
- Gad Getz
- Matthew Meyerson

- Broad Institute Cancer Genome Analysis Group

