Ancient elements in genomes

what is a transposon?

- Contiguous piece of DNA; different types vary in length (300 bp to 6.5kb or so)
- Repeated with minor variations throughout the host genome
- Can replicate itself by cut and paste or copy and paste mechanisms (can move around!)
- No known function — synthetic genome projects aim to remove them
- Structural and functional analogies to viruses (much of the terminology reflects this)
- Possibly associated with large-scale genomic structural changes
types of transposons

- Cut and paste (DNA transposons)
- Copy and paste (retrotransposons)
  - Autonomous retrotransposons
    - ERVs *possibly active in human genome
    - L1 & relatives *active in human genome
  - Nonautonomous retrotransposons
    - SINEs (Alu) *active in human genome
    - SVA *active in human genome
      - Composite element (SINE, VNTR, Alu)
    - Processed pseudogenes
transposons comprise nearly 45% of the human genome

- DNA transposons 3%
- Autonomous retrotransposons
  - ERVs
    - L1 18% (500,000 copies)
    - L2 3%
    - L3 & relatives 1%
  \[\text{LTR retrotransposons}\]
- Nonautonomous retrotransposons
  - SINEs (Alu) 15% (1 million+ copies)
  - SVA (3000 copies)
  - Processed pseudogenes (>8000)
- (Simple repeats occupy almost another 10%)

“Junk DNA”?

- What do transposons do?
  - Make more of themselves
  - Move genes around
  - Serve as reservoirs of new sequence
  - Cause genetic instability (repeats stimulate translocation; L1 causes chromosome breakage)
  - Can contribute to genes and gene expression
    - 5% of alternatively spliced internal human exons come from Alus
    - 80% of genes have some L1 sequence in noncoding portion
    - 1-4% of coding sequence is L1-derived
    - Act as methylation centers
unusual evolutionary mechanism

this leads to accumulation of older and degenerate (nonfunctional) elements in the genome

importance in genomics

- Transposons are a source of human variability
  - At least 5% of people have a transposon not found in either parent (not due to nonpaternity!)
  - Overall polymorphism variable; between 2 unrelated individuals there are typically thousands of transposons found in only one of the people
- Transposons can be useful in medicine
  - Occasionally cause disease (de novo insertion in factor VIII clotting gene led to L1 discovery in 1980s)
  - May often be linked to disease loci
importance in genomics

- Transposons in introns may disrupt gene expression
  - Mechanism depends on whether they are on the sense or antisense strand
- (+) strand orientation — transcription stalling
- (-) strand orientation — premature polyadenylation, gene splitting

**L1 elements can reshuffle the genome**

Introduction of foreign sequence

- 5' transduction
- 3' transduction

Non-allelic homologous recombination
transposon domestication

- Host cells use many mechanisms to control transposons
  - Methylation (original role?)
  - miRNA defense
  - Sequestered in stress granules
  - Nucleic acid editing
    - APOBEC family of proteins edits cytosines to uracils
    - ADARs edit dsRNA adenosine to inosine
- Transposons can spread by horizontal transfer!
  - Evidence in opossum, little brown bat, grey mouse lemur
- Alu elements are dying out in orangutans

importance in genomics

- Can have large-scale effects, through chromosomal translocation, inversion, breakage
- Generation of diversity (identical twins can have different genome-wide transposon insertion sites)
- May cause or mark gross abnormalities in genome structure
- Potential for gene therapy vector
what to do with transposons?

- Difficult to manage in high throughput sequencing experiments (microarrays simply leave them off)
- cannot determine variation in repetitive regions
- cannot confidently call copy number changes in repetitive regions
- cannot unambiguously place transposon insertions without specialized techniques

• Study them
• Try to ignore them
  • RepeatMasker (Smit & Jurka)
  • Problem: many elements are in part unique, and RepeatMasker will mask that too
  • Can’t find variation in transposons or due to transposons (very useful in linkage studies)
short read sequencing & repeats

unique  single repeat  unique

unique  two copies of the repeat  unique

the problem is not limited to research on transposons and simple repeats -

-pseudogenes
-gene families
-telomeres
-cancer: breakpoints are often at repeats
-CNVP boundaries difficult to determine if repetitive elements are nearby
what to do with transposons?
what to do with transposons?
Studying transposons

- specialized experimental techniques allow profiling of most transposon insertion sites in a genome

Studying transposons

- compare sites between two genomes (tumor/normal, family members etc)

Mobile Interspersed Repeats Are Major Structural Variants in the Human Genome

Studying transposons
Studying transposons

- looking at transposon expression is becoming popular though specificity and quantitation are problems

Recent work with, on, or around mobile elements

- mutagenesis screens

---

A Transposon-Based Genetic Screen in Mice Identifies Genes Altered in Colorectal Cancer

Timothy K. Starr,1,2 Raha Allani,1 Kevin A. T. Silverstein,1 Rodney A. Staggs,1
Aaron L. Savar,1 Tracy L. Bergemann,1 Rithir Gupta,1,2 M. Gerard O'Sullivan,1
Ruz Mattei,1 Adam J. Dupuy,1 Lara S. Geller,1 Scott Powers,1 Art L. Ong,1
Yan W. Amara,1,2 Stephen N. Filalobre,1 Uluo Tersaard,1,2 Noel G. Copeland,1,3
Nancy A. Jenkins,1,4 Robert T. Counter,1,4 David A. Largaespada4

Human colorectal cancer (CRC) display a large number of genetic and epigenetic alterations, some of which are causally involved in tumorigenesis (driver) and others that have little functional impact (passenger). To help distinguish between these two classes of alterations, we used a transposon-based genetic screen in mice to identify candidate genes for CRC. Mice harboring mutagenic sterility-breaking DSB transposons were crossed with mice expressing SB transposase in gastrointestinal tract epithelia. Most of the offspring developed intestinal lesions, including intraperitoneal neoplasms, adenomas, and adenocarcinomas. Analysis of over 16,000 transposon insertions identified 77 candidate CRC genes, 60 of which are mutated and/or deregulated in human CRC and thus are most likely to drive tumorigenesis. These genes include APC, PKN, and SMAD. The screen also identified 17 candidate genes that had not previously been implicated in CRC, including POU, PTP, and JSPQ.
Recent work with, on, or around mobile elements

• link to natural variation

Natural Genetic Variation Caused by Transposable Elements in Humans
E. Andrew Bennett,⁎,1,2 Laura E. Coleman,⁎,1 Circe Tsui,⁎,†,2 W. Stephen Pittard⁎ and Scott E. Devine⁎,†,2,‡,3

Recent work with, on, or around mobile elements

Natural Mutagenesis of Human Genomes by Endogenous Retrotansposons
Rebecca C. Iskow,⁎,1,2 Michael T. McCabe,⁎,1,2,3,4 Ryan E. Mills,⁎,1,2,3,4 Spencer Tomene,⁎,1,2,3,4 W. Stephen Pittard,⁷ Andrew F. Neuwald,⁎,1,2,3,4 Erwin G. van Meijl,⁎,1,2,3,4 Paula M. Vertino,⁎,1,2 and Scott E. Devine⁎,1,2,3,4,‡,3,4,†,2,3,4
Recent work with, on, or around mobile elements

A Novel Correlation between LINE-1 Hypomethylation and the Malignancy of Gastrointestinal Stromal Tumors
Shinohi Igashira, Hiromu Suzuki, Takeaki Nituma, Haruo Shinbuku, Masanori Nojima, Hirono Ikeda, Takayuki Nishida, Yasuaki Miyazaki, Hiroyuki Takamura, Hiroyuki Yamanoboshi, Takashi Tokino, Tadashi Hasegawa, Koichi Hirata, Kohroh Imai, Minoru Toyota, and Yasuhisa Shinomura

Recent work with, on, or around mobile elements

• look for activity in adult cells

Cocaine dynamically regulates heterochromatin and repetitive element unsilencing in nucleus accumbens
Ian Maze, Jian Feng, Matthew R. Wilkinson, HaoSheng Sun, Li Shen, and Eric J. Nestler
Recent work with, on, or around mobile elements

- probes of chromatin structure

DNA transposon *Hermes* inserts into DNA in nucleosome-free regions in vivo

Sunil Gangadharan\textsuperscript{a,b}, Loris Meleron\textsuperscript{a,b}, Jennifer Fahn-Thornton\textsuperscript{a,c}, Sarah J. Wheelan\textsuperscript{a,c}, and Nancy L. Craig\textsuperscript{a,b}
**PLANTS**

other option: forget about where they are and just study the transposons

Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data

Petr Novák, Pavel Neumann and Jiří Macas*
graph. We have found that high graph density and maximal degree values are indicative of short tandem repeats or satellites, and that graph diameter is proportional to the repetitive element length (see examples provided below).

Figure 4 Examples of graph layouts derived from clusters of repetitive sequences. Graph layouts were calculated using the 3D version of Fruchterman and Reingold algorithm from which a 2D projection is shown. Individual reads are represented by vertices and similar reads are connected by edges. Individual clusters are described further in Table 1.

Figure 5 Visualization of contigs in the F. solium cluster PsCL7 using SeqGraphite. (A) The graph layout was calculated using a 3D version of the Fruchterman and Reingold algorithm. The colors of the nodes are based on the results of the sequence assembly into contigs using the CAP3 program, which are then schematically represented in panel B. Numbers label the loops discussed in the text. The contig alignment is shown below the graph together with a diagram of corresponding regions of the Acgel LTR-retrotransposon (C).