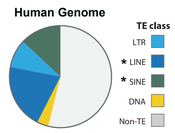
Direct genome-scale mapping of human LINE-1 ORF2p endonuclease activity

Hanna Kodama[1-2], Syndi Koltz[3], John Thompson[3], Gerwald Jogl[2], Martin Taylor[1]
[1]Department of Pathology and Laboratory Medicine, Brown University, Providence, RI; Department of Molecular Biology, Cellular Biology, and Biochemistry, Brown University, Providence, RI; [3]Nabsys 2.0 LLC, Providence, RI

Abstract

Background: The LINE-1 retrotransposon is a virus-like DNA parasite that has written at least a third of the human genome and is our only active protein-coding transposon. It propagates via an RNA intermediate through ORF2p endonuclease (EN) and reverse transcriptase (RT) activities. LINE-1 insertions can cause sporadic genetic disease and contribute to cancer through EN-dependent DNA damage, with γH2AX foci far exceeding completed insertions, suggesting damage outside canonical retrotransposition.
Insertions favor a 5′-TTTTT↓AA motif, but whether this reflects EN cleavage specificity or RT priming requirements is unknown, therefore EN sequence preferences in genomic DNA remain unclear.



Methods: Electronic genome mapping (EGM) uses solid-state nanodetectors to survey long DNA molecules and construct high-density maps to detect structural variants (SVs) and repeat sizing. To address these knowledge gaps, we purified ORF2p EN to homogeneity from *E. coli*, as well as a catalytically inactive mutant as a control and assayed their cleavage patterns on Lambda, *E. coli*, and human DNAs using electronic genome mapping, a technique that directly measures nick positions on long DNAs.

Results: Analytical methods were modified to minimize false positives and enable the detection of low-level nicking. The frequency of nicks increased in an EN activity- and concentration-dependent manner. Mapping of low-frequency nicks to specific sites is underway.

Conclusion: These results reveal a powerful new method to directly measure genome-scale endonuclease nicking. Future applications of this technology may also be used to directly assay other DNA modifications and damage at scale.

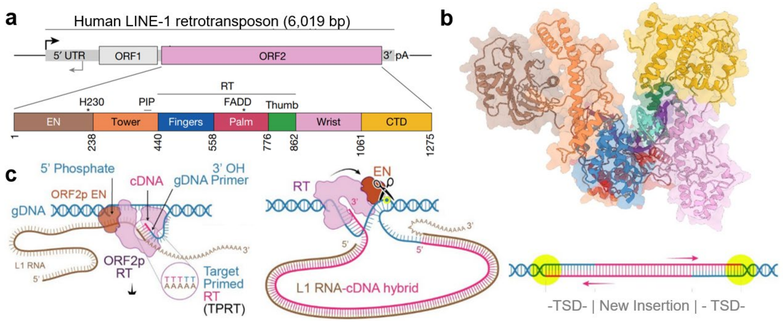


Figure 1. Hur…an L1 encodes retrotransposition-facilitating machinery. (a) L1-ORF2p is a 1275 amino acid, multi-domain protein with EN and RT function and (b) folds into a classic right-handed RT structure, cradling its RNA:DNA hybrid template primer (purple, cyan). (c) Pathogenic DNA cutting by the L1-EN occurs at multiple points during insertion. Final steps include ligation and repair, and potentially rely on host factors.[1]

L1 Insertion Profiling Misrepresents EN activity

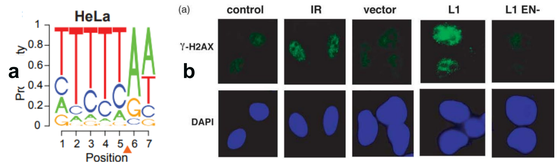


Figure 2. EN Consensus sequence is likely influenced by later steps during L1 insertion. (a) The commonly cited L1-EN consensus sequence (5′-TTTTT/AA-3′) was derived from genomic insertion sites, which inherently select for sequences that support reverse-transcription priming[2] (see Fig. 1c). (b) L1-expressing cells show substantially elevated DNA damage (γH2AX) relative to the number of observed insertion events, indicating that many EN nicks do not progress to full insertions[3].

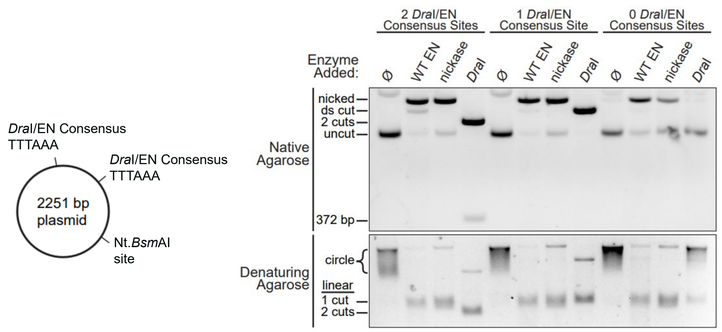


Figure 3. L1-EN nicks plasmids with or without consensus sequence. Purified L1 EN activity on plasmid DNA produces a nicked (open circle) product. Plasmids with two consensus EN/DraI cut sites (one 5′-TTTTAAA and one 5′TTTAAA), one canonical EN/DraI cut site (5′-TTTTAAA), and zero canonical EN/DraI cut sites showed similar rates EN cutting activity on both native and denaturing agarose gel revealing a lack of EN cutting specificity for the consensus site[4].

Method: Electronic genome mapping

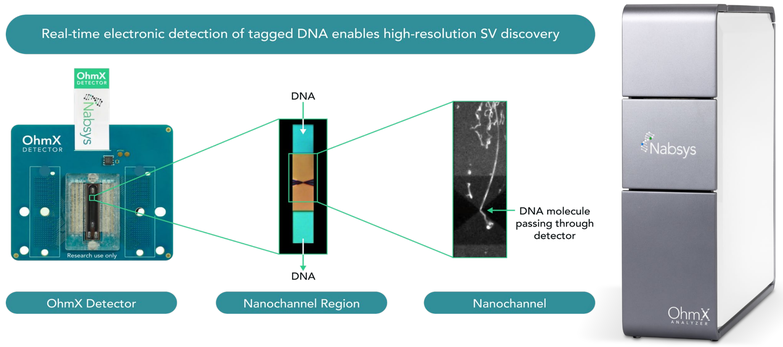


Figure 4. EGM detector and analyzer. EGM is enabled by the novel, solid-state nanodetectors developed by Nabsys. The OhmX Detector houses 256 parallel nanochannels, each with its own electronic sensor. Nabsys has developed DNA tagging chemistry that provides a high signal-to-noise. The high signal and spatial resolution allows tags to be closely spaced on the high molecular weight DNA molecules. This yields optimal tag density such that structural variations as small as 300 bp can be routinely identified and mapped genome-wide.

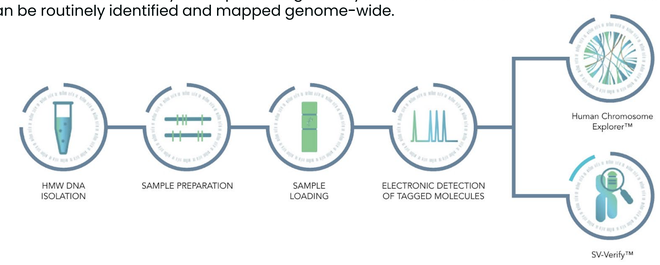


Figure 5. EGM Workflow Overview. The EGM workflow consists of the following steps: (1) high molecular weight (HMW) DNA isolation, (2) sample preparation using the OhmX Simultaneous Nicking and Labeling Sample Preparation Kit, (3) sample loading, (4) electronic detection of tagged molecules using the OhmX Analyzer, and (5) data analysis.

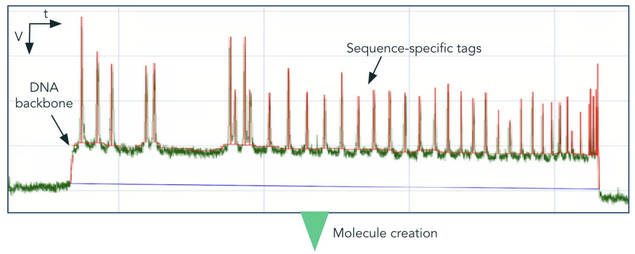


Figure 6. Electronic profile for tagged DNA molecules over 150 kb. The baseline voltage is shown on the left. When the DNA molecules enter the nanochannel, a change in voltage is observed. Whenever a site-specific tag on the DNA is also present, the voltage changes again. The time between tags is converted to distance in base pairs.

Validation of EN nicking detection by EGM

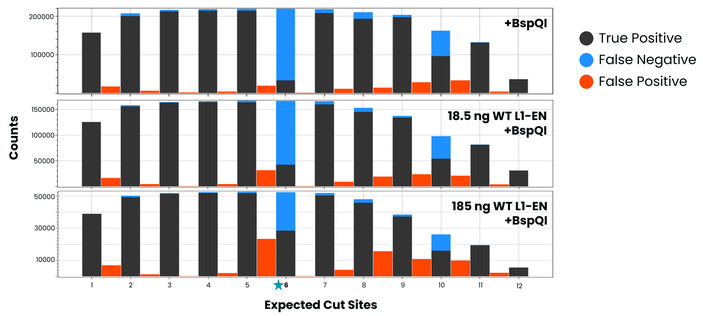
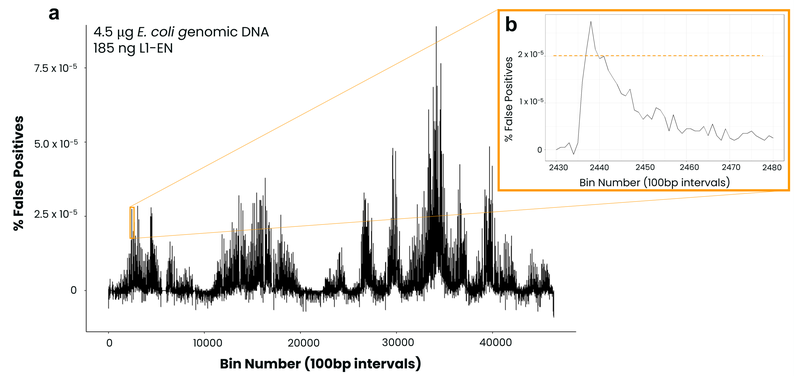


Figure 7. Nicking profile along Lambda DNA shows nicks appearing in an EN-dependent manner. Including sites 6 and 11 (non-BspQI sites) reveals EN-dependent signal at site 6 (starred), where increasing WT EN leads to increased ratio of nick calls (black) to control signal (blue). Red bars denote nicks not present in the reference, comprising both true false positives and bona fide EN nicks occurring outside the TTTTT/AAA consensus.

L1-EN Nicking of Genomic DNA



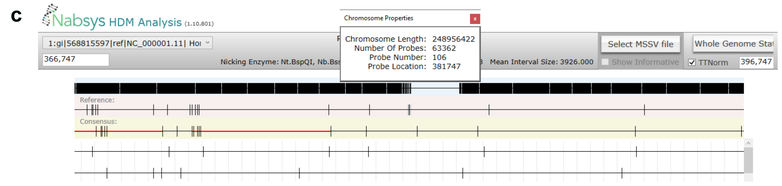


Figure 8. The distribution of nicks across the *E. coli* and human genomes. (a) Across the ~4.5 Mbp genome of *E. coli*, there are hot spots of nicks only appearing in presence of EN. Segments such as that shown in (b) were used for motif analysis. (c) Example view of Nabsys EGM Analysis, showing a 3kb window on Chromosome 12. Individual fragments (bottom), as well as the consensus fragment, show strong alignment to the distribution of nicks on the reference.

a

	EN	Cut Sites Included in Reference	% Filtered Remapped (FR)	%FP	%FN (>500bp)	Δ%FR/site	# of Sites	
E. coli	−	BssSI	63.85	11.6	5.93	-	807	
	+	BssSI	1.2	22.3	8.59	-	807	
	+	BssSI, TTTAAAAA	2.9	19.3	15.25	0.0051	333	
	+	BssSI, TTTAAAAA, **CTTATCAGGCCT**	3.5	19.2	15.48	0.0146	158	*12mer*
	+	BssSI, TTTAAAAA, **GCCGCCGCG**	4.5	19	16.23	0.0244	135	*9mer*
	+	BssSI, TTTAAAAA, **GCCAGC**	39.1	6.28	1.1	0.004	9469	*6mer*

b

	EN	Cut Sites Included in Reference	% Filtered Remapped (FR)	%FP	%FN (>500bp)	Δ%FR/site	# of Sites
human	−	BssSI, BspQI	32.4	8.2	9.29	-	-
	+	BssSI, BspQI	33.1	8.27	8.76	-	-
	+	BssSI, BspQI, **GCCGCCGCG**	32.3	8.2	8.9	-	7388

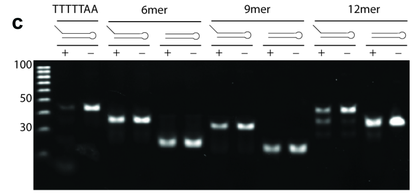


Figure 9. Remapping fragments to reference genomes reveals potential new EN motifs. (a) EN nicking sequence motifs (bold) were identified using MEME Suite from regions of high EN-induced signal in *E. coli* DNA. Remapping fragments using only canonical nickase site(s) yielded poor alignment, whereas including newly identified motifs improved remapping, with the 9-mer performing best after accounting for frequency (underlined). (b) Preliminary EN nicking of human DNA shows strong remapping when the *E. coli*-derived 9mer motif is included in the analysis. (c) Hairpin substrates designed for gel-based nicking assays and containing the new 6-, 9-, and 12-mer motifs are not nicked, suggesting that additional EN–DNA interactions may be required for activity.

Conclusions

- We have validated EGM as a method to detect EN-mediated nicks, confirming activity at previously identified consensus sequences as well as revealing new nicking sites.
- **Workflow advantages:** Compared to large-scale approaches such as ChIP-seq, END-seq, and CUT&RUN, this workflow is faster, lower-cost, and requires less intensive downstream analysis.
- **Impact:** Improving our understanding of LINE-1 endonuclease promiscuity will help link LINE-1 activity to disease-causing mutations and opens new avenues for future experiments, including ongoing studies of EN mutants.

[1] Baldwin, E. T. et al. Structures, functions and adaptations of the human LINE-1 ORF2 protein. Nature 626, 194-206 (2024); **[2]** Flasch, D. A. et al. Genome-wide de novo L1 Retrotransposition Connects Endonuclease Activity with Replication. Cell 177, 837-851 e828 (2019); **[3]** Gasior, S. L., Wakeman, T. P., Xu, B. & Deininger, P. L. The Human LINE-1 Retrotransposon Creates DNA Double-strand Breaks. Journal of Molecular Biology (2006); **[4]** Miller, B.M. et al. Rapid DNA cleavage by the LINE-1 endonuclease proximal to DNA ends and at mismatches. J Biol Chem (2025).