

# Investigating Trends in Transposable Element Insertion within Regulatory Regions

**Carri-Lyn Mead**  
**cmead@bcgsc.ca**

**Thursday, January 13, 2005**

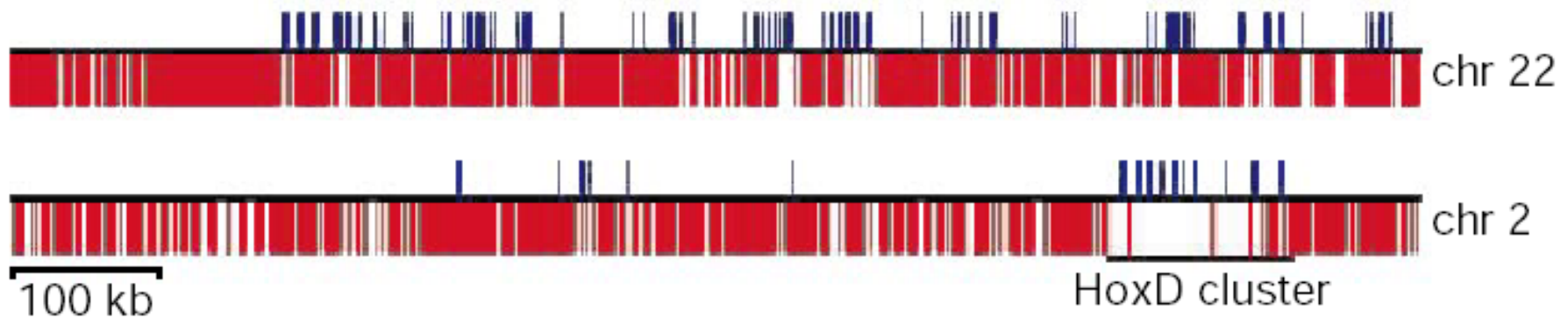
# Outline

- Transposable Element (TE) Overview
- Background
- Human Genome Results
- Mouse Genome Results

# Transposable Element (TE) Overview

- Definition: A defined length of DNA that translocates to other sites in the genome, essentially independent of sequence homology.
- Occupy ~45% of the human genome
- Most have been inserted since the divergence of Mouse and Human
- Types of TEs:
  - Long Interspersed Elements (LINEs)
  - Short Interspersed Elements (SINEs)
  - Long Terminal Repeat (LTR) retrovirus-like elements
  - DNA Transposons

# Human Genome TE Distribution



Two regions of about 1 Mb on chromosomes 2 and 22. Red bars, interspersed repeats; blue bars, exons of known genes. Note the deficit of repeats in the HoxD cluster, which contains a collection of genes with complex, interrelated regulation.

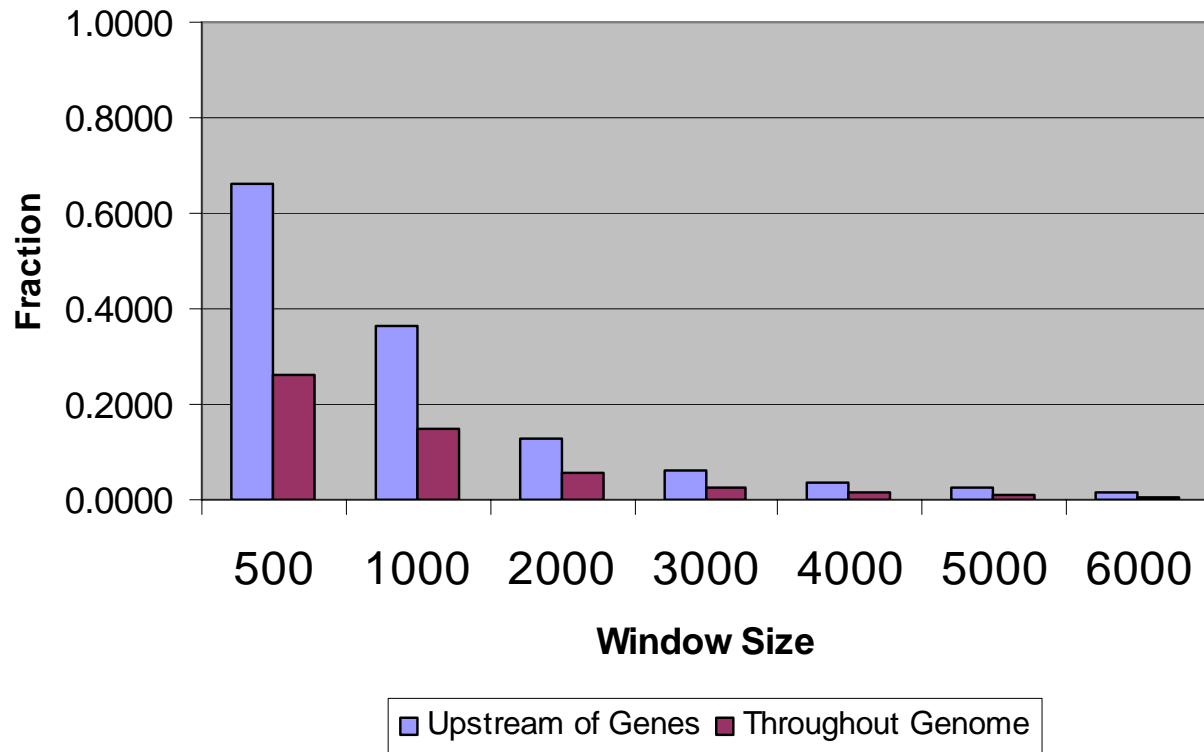
# Hypothesis

Selective pressures are acting to exclude transposable elements from regions around genes whose functions are critical to development.

We will focus on the upstream regulatory region



# Fraction of Windows with No TE Insertion - Human



Window Size	500	1000	2000	3000	4000	5000	6000	n
Upstream of Genes	0.6620	0.3624	0.1288	0.0607	0.0361	0.0243	0.0173	17030
Throughout Genome	0.2636	0.1482	0.0571	0.0274	0.0156	0.0101	0.0072	2851331617

# Gene Ontology Analysis

Window Size 6000

- # Genes with TEs absent from Upstream Window: 306
- # with no other Gene overlapping Upstream Window: 111
- Use GoMiner for Gene Ontology Analysis

# GO Results: Cellular Component

GO Terms	Fold Enrichment	P Values	# Candidate Genes	% Candidate Genes (n=111)
<b>Cellular Component</b>	1.05	0.1079	79	71.17%
<b>Organelle</b>	<b>1.68</b>	<b>0.0000</b>	<b>64</b>	<b>57.66%</b>
<b>Protein Complex</b>	<b>2.55</b>	<b>0.0000</b>	<b>34</b>	<b>30.63%</b>
<b>Cell</b>	<b>1.14</b>	<b>0.0072</b>	<b>75</b>	<b>67.57%</b>
Extracellular Matrix	1.42	0.3087	4	3.60%
Cellular Component Unknown	1.03	0.5141	9	8.11%
Extracellular	0.64	0.9638	10	9.01%



# GO Results: Molecular Function

GO Terms	Fold Enrichment	P Values	# Candidate Genes	% Candidate Genes (n=111)
<b>Molecular Function</b>	1.06	0.0147	84	75.68%
<b>Binding</b>	<b>1.40</b>	<b>0.0000</b>	<b>72</b>	<b>64.86%</b>
<b>Transcription Regulator Activity</b>	<b>5.39</b>	<b>0.0000</b>	<b>50</b>	<b>45.05%</b>
Molecular Function Unknown	1.40	0.1472	12	10.81%
Enzyme Regulator Activity	0.89	0.6678	4	3.60%
Structural Molecule Activity	0.76	0.7827	4	3.60%
Signal Transducer Activity	0.60	0.9864	11	9.91%
Transporter Activity	0.45	0.9897	5	4.50%
Catalytic Activity	0.50	0.9999	15	13.51%

# GO Results: Biological Process

GO Terms	Fold Enrichment	P Values	# Candidate Genes	% Candidate Genes (n=111)
<b>Biological Process</b>	1.10	0.0015	84	75.68%
Regulation of Biological Process	2.83	0.0000	62	55.86%
Development	2.63	0.0000	48	43.24%
Physiological Process	1.13	0.0142	75	67.57%
Viral Life Cycle	4.04	0.2203	1	0.90%
Cellular Process	0.97	0.6485	46	41.44%
Biological Process Unknown	0.89	0.6874	7	6.31%
Behaviour	0.66	0.7855	1	0.90%

# Mouse

Window Size 6000

- # Genes with TEs absent from Upstream Window: 329
- # with no other Gene overlapping Upstream Window: 111

# Human – Mouse Orthologs

- 16% orthology between Human and Mouse genes with no TE insertion or gene overlap in upstream 6000 bp

BAI3	LHX5	PCDH10
C6orf82	LHX9	PITX1
CDK6	MEIS1	PITX2
HOXC12	NFIX	SALL1
HOXD11	NKX2-2	SIX1
IMP-3	NTNG2	SOX1

- GoMiner Overrepresentation:

Biological Process – Development

Biological Process – Regulation of Biological Process

Molecular Function - Transcription Regulator Activity

# Iwama and Gojobori Oct 2004

- a study of upstream sequences of 3,750 human-mouse orthologue pairs, high upstream conservation are predominantly transcription factor genes, particularly those developmentally associated

Iwama, H, and Gojobori, T. Highly Conserved Upstream Sequences for Transcription Factor Genes and Implications for the Regulatory Network. PNAS (2004) 101:17156-17161.

# Potential Contribution

- Results may indicate genes for further investigation
- Results may contribute to determining limits of regulatory region

# References

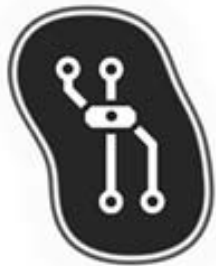
1. The Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* (2001) 409:860-921.
2. Smit, AFA. The origin of interspersed repeats in the human genome. *Current Opinion in Genetics & Development* (1996) 6:743-748.
3. Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, Sunshine M, Narasimhan S, Kane DW, Reinhold WC, Lababidi S, Bussey KJ, Riss J, Barrett JC, Weinstein JN. GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol.* 2003;4:R28.
4. Iwama, H, and Gojobori, T. Highly Conserved Upstream Sequences for Transcription Factor Genes and Implications for the Regulatory Network. *PNAS* (2004) 101:17156-17161.

# Acknowledgements

Terry Fox Laboratory

Dr. Dixie Mager

Louie van de Lagemaat



*CIHR/MSFHR Strategic Training Program in*

**BIOINFORMATICS**