

High-Resolution Genome-Wide Mapping of integration sites of the DNA transposon *Hermes* in *S. cerevisiae*

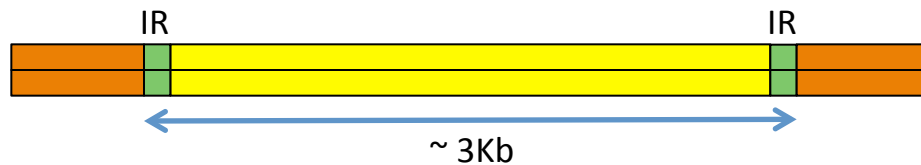
Loris Mularoni, Sunil Gangadharan, Sarah Wheelan, Nancy Craig

Johns Hopkins University School of Medicine, Baltimore
Howard Hughes Medical Institute



Introduction

HERMES



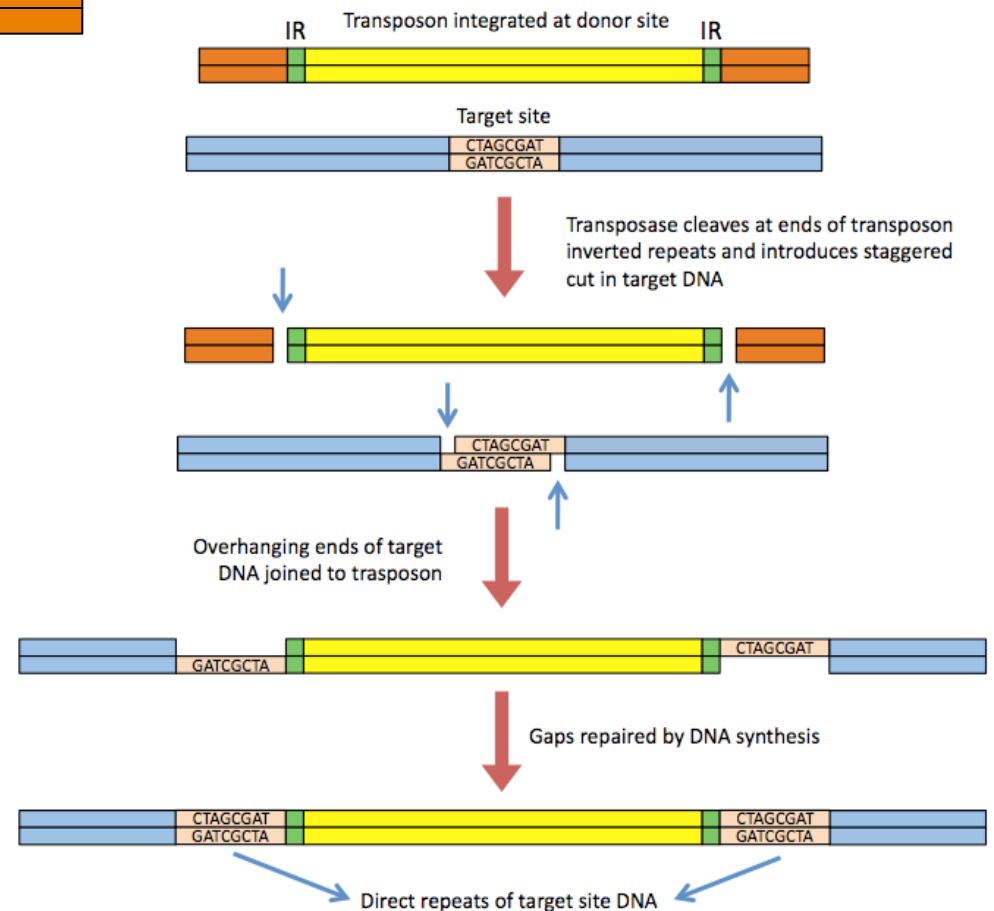
Features of *Hermes*

- DNA transposon (cut and paste mechanism)
- initially isolated from *Musca domestica* (Warren et al., 1994)

Most transposon insertions are non-random

- can target specific sequences
- can target genomic features

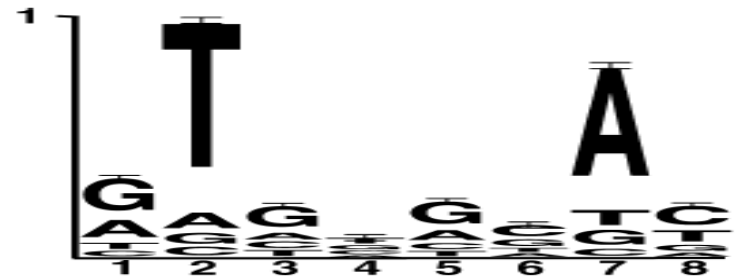
TRANSPOSITION



Introduction

PREVIOUS STUDIES OF *HERMES* INTEGRATION SITE PREFERENCE

- in *D. melanogaster* (Guimond et al. 2003)
Genome size= 122.6 Mb
Number of unique insertions=252



- in *A. aegypti* (Sarkar A. et al. 1997)
Genome size=3.9Kb
Number of unique insertions=24

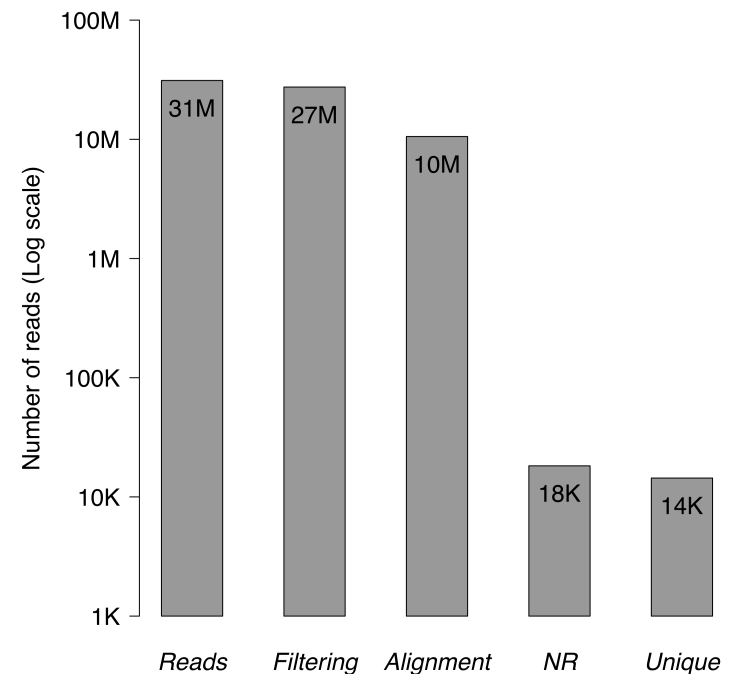
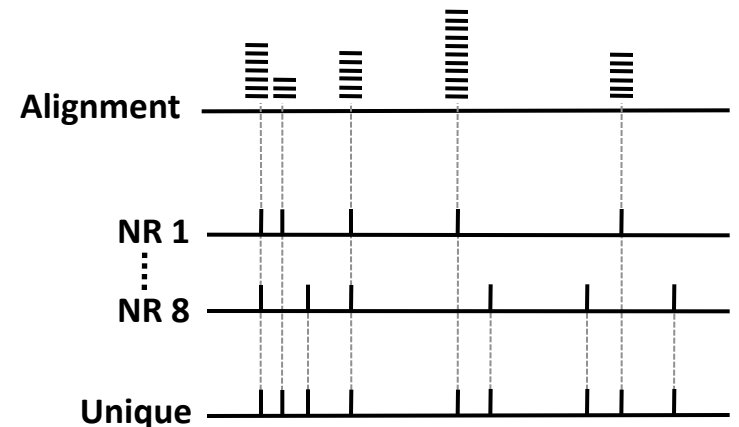


- in *S. pombe* (Evertts A. G. et al. 2007)
Genome size=14.1 Mb
Number of unique insertions=26



EXPERIMENTAL DESIGN

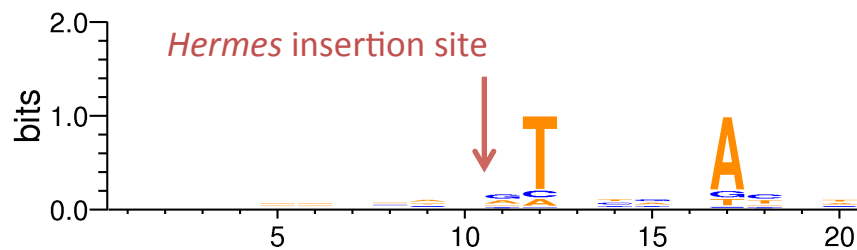
- 1) Transformation of yeast with *Hermes* transposon + transposase
- 2) Digestion of the DNA with MseI and EAH (EcoRI, HindIII, AflII) enzymes
- 3) Amplification by transposon specific linker mediated PCR
- 4) Massively parallel sequencing (454 and Solexa, **31M reads**)
- 5) Filtering of the reads (**27M reads**)
- 6) Alignment of the reads to the yeast genome with Bowtie (**10M reads**)
- 7) Identification of the non-redundant (NR) and unique insertions (**14394 insertions**)



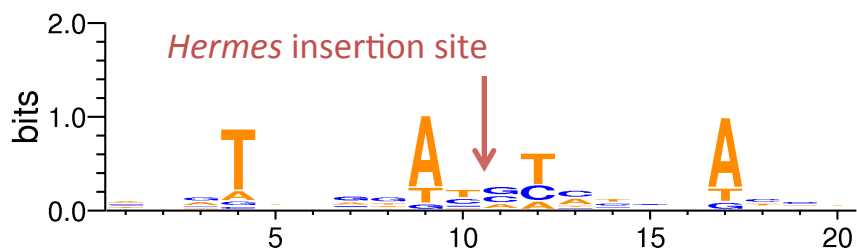
Results

SEQUENCE ANALYSIS OF *HERMES* INTEGRATION SITE

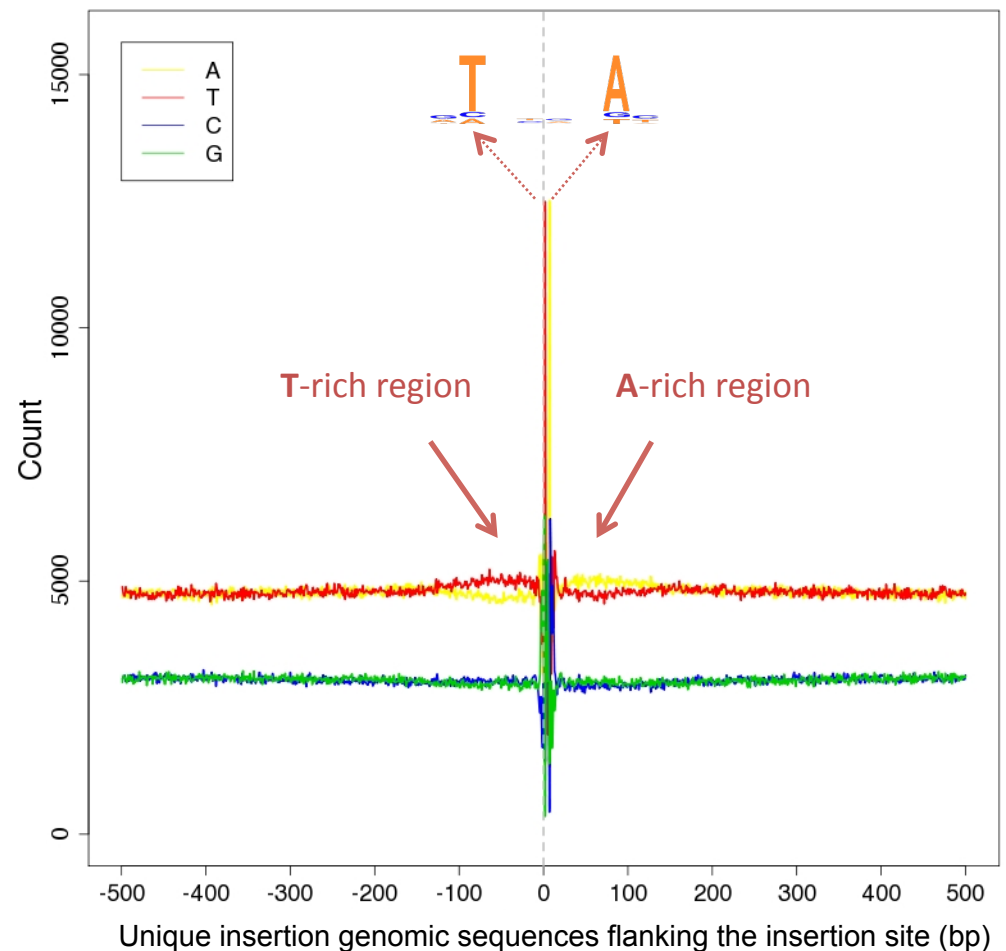
All 14394 unique insertions



46 insertion sites are bidirectional



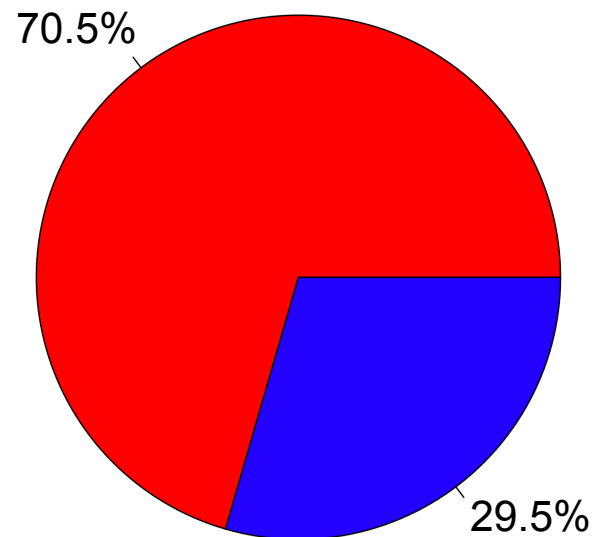
Nucleotide composition of target sequences



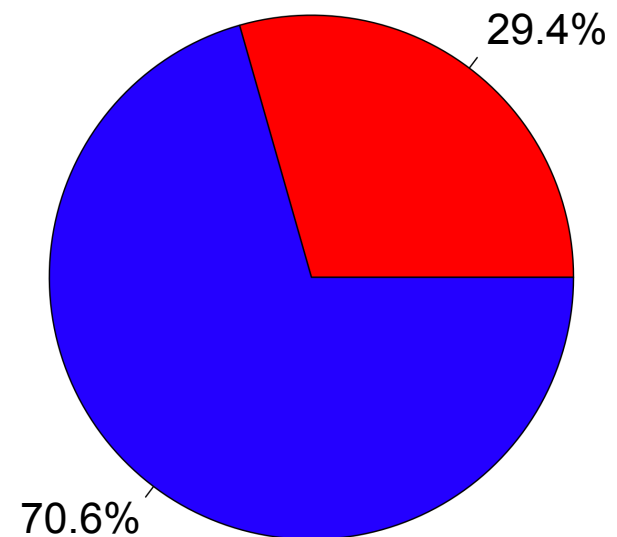
Results

HERMES INTEGRATES OUTSIDE OF ORFS

Yeast genome



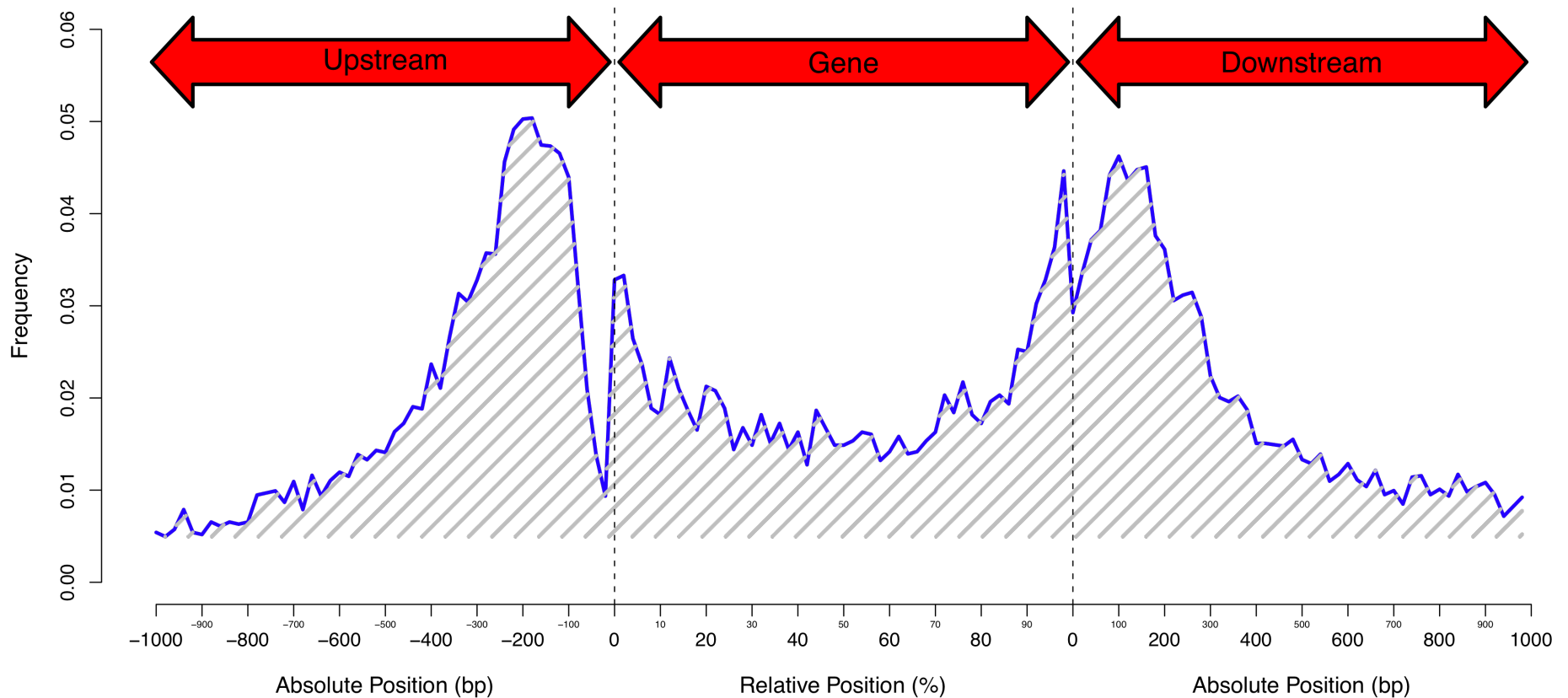
Hermes insertions



Results

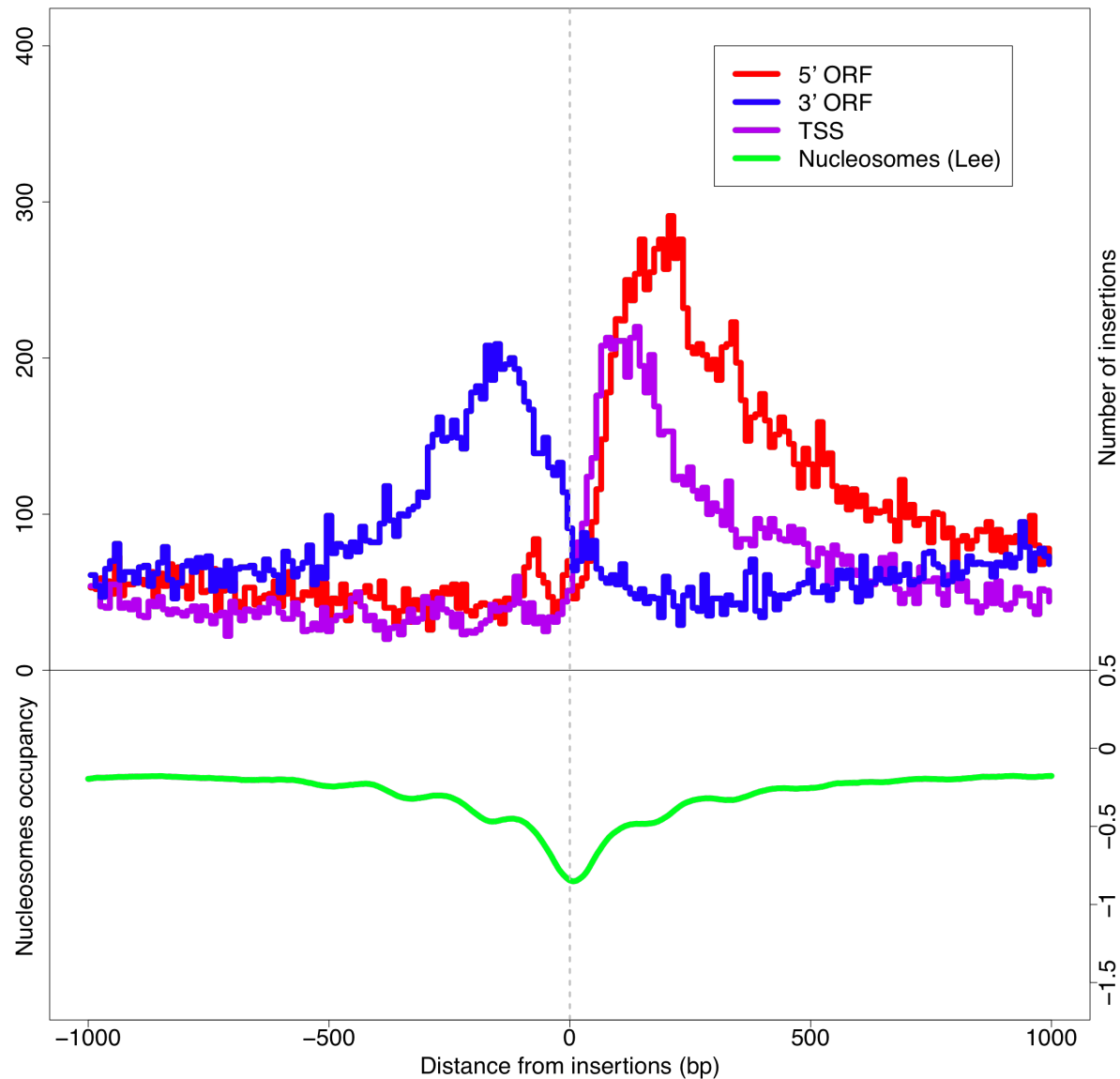
INSERTION SITE DISTRIBUTION IN THE GENOME

Distribution of insertions inside and around genes



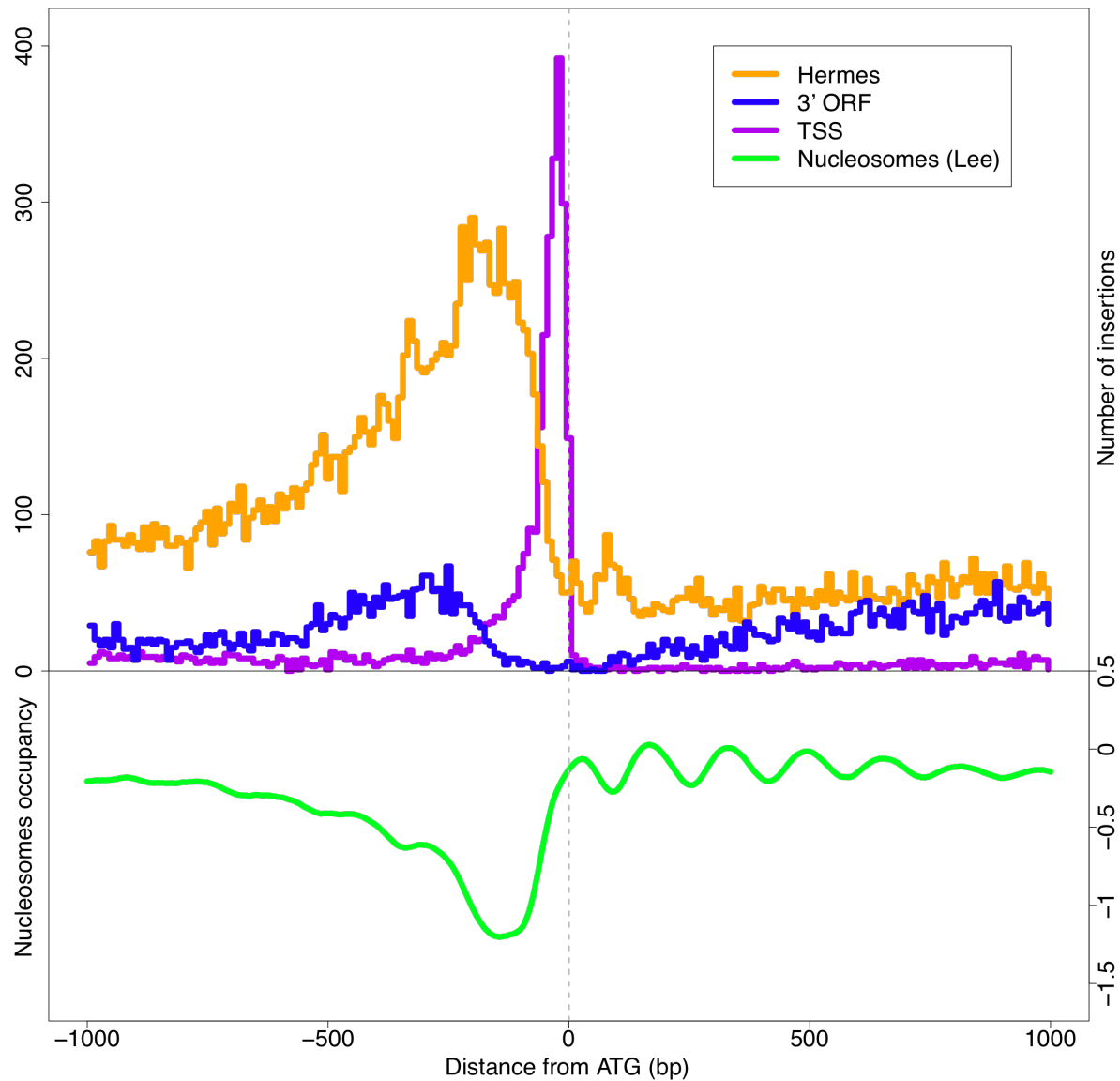
Results

INSERTION SITE FEATURES



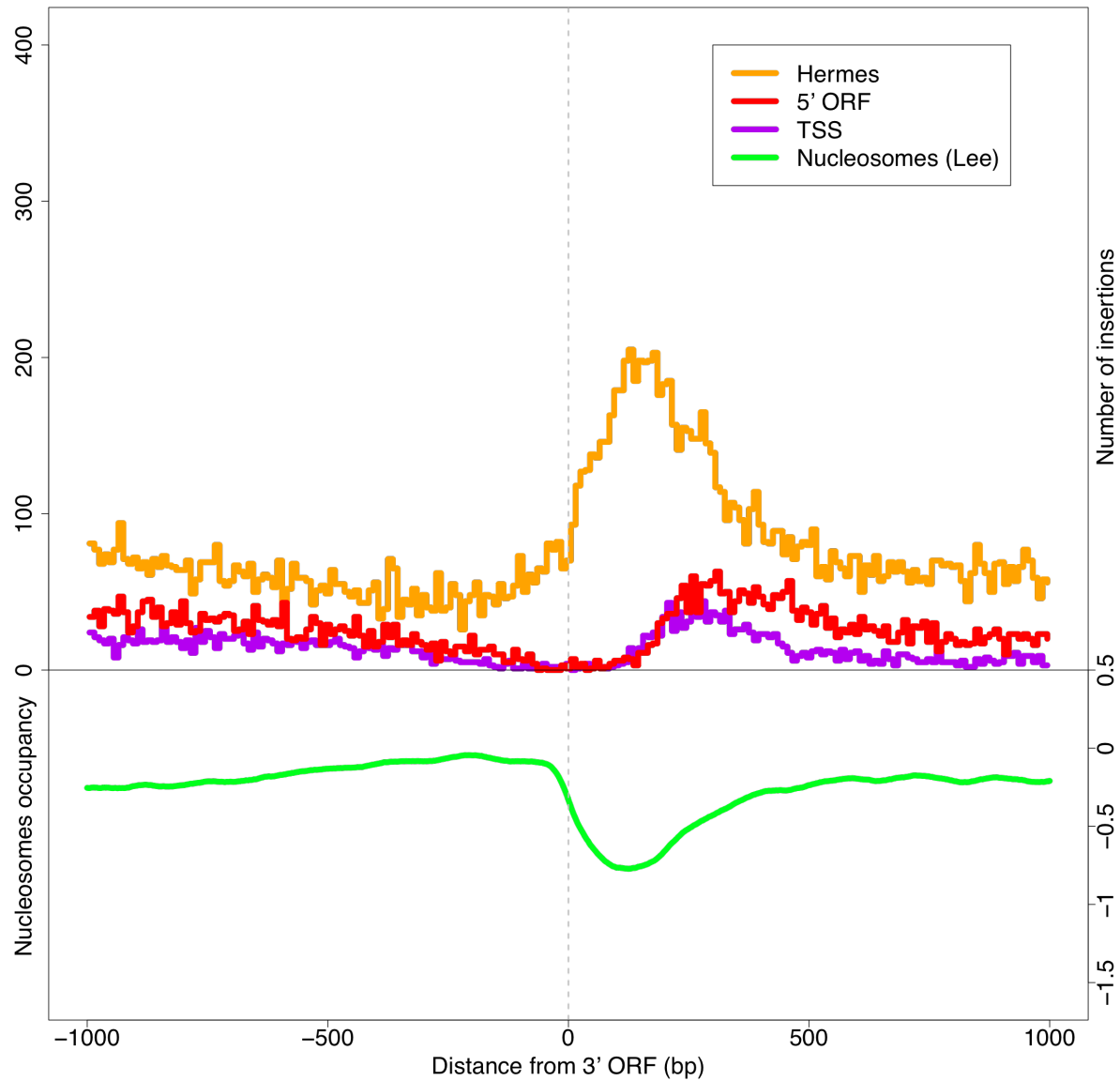
Results

ATG-CENTERED VIEW



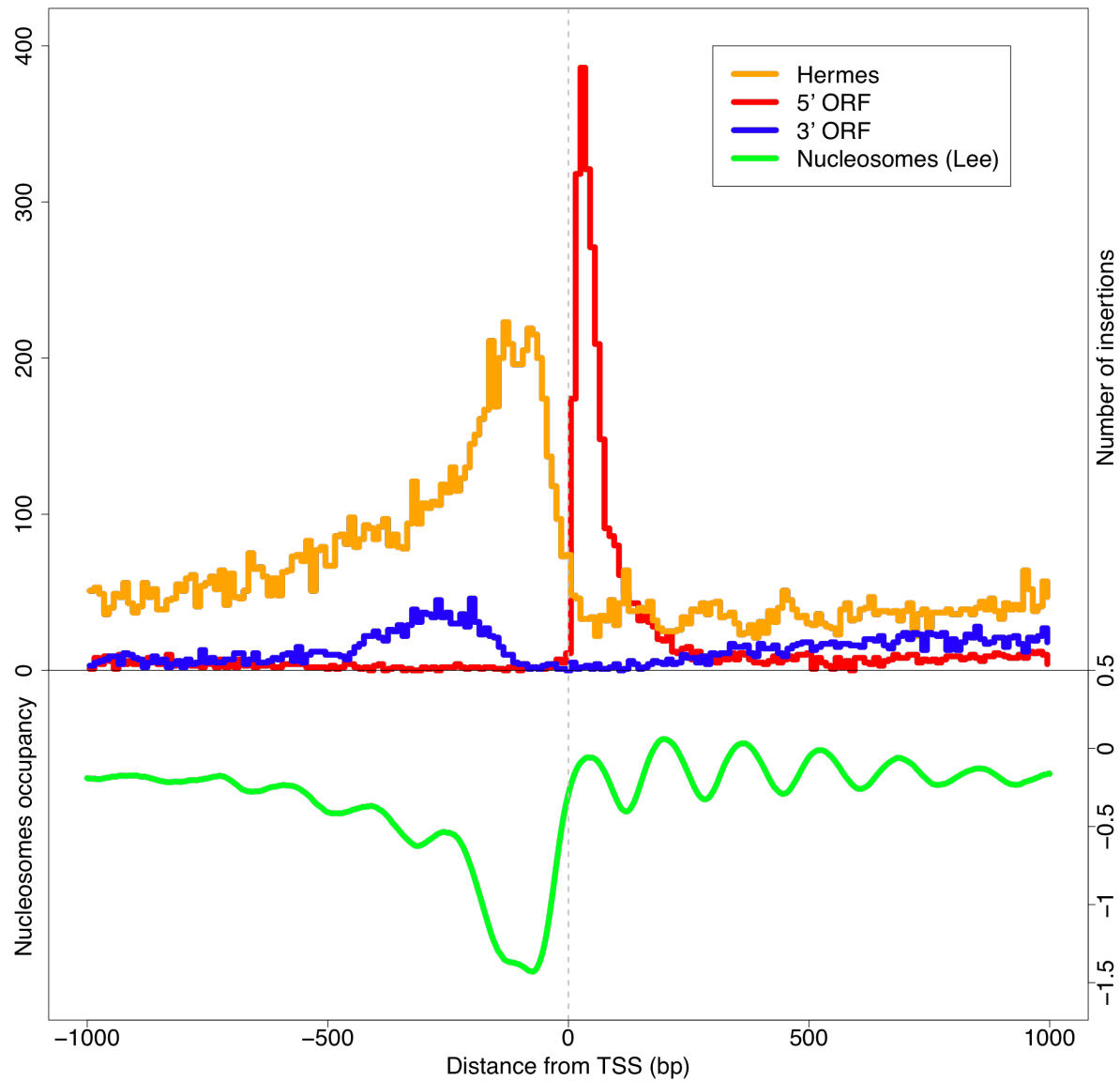
Results

ORF 3p END-CENTERED VIEW



Results

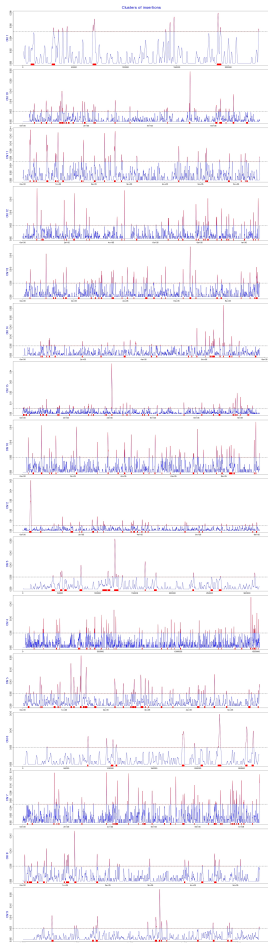
TSS-CENTERED VIEW



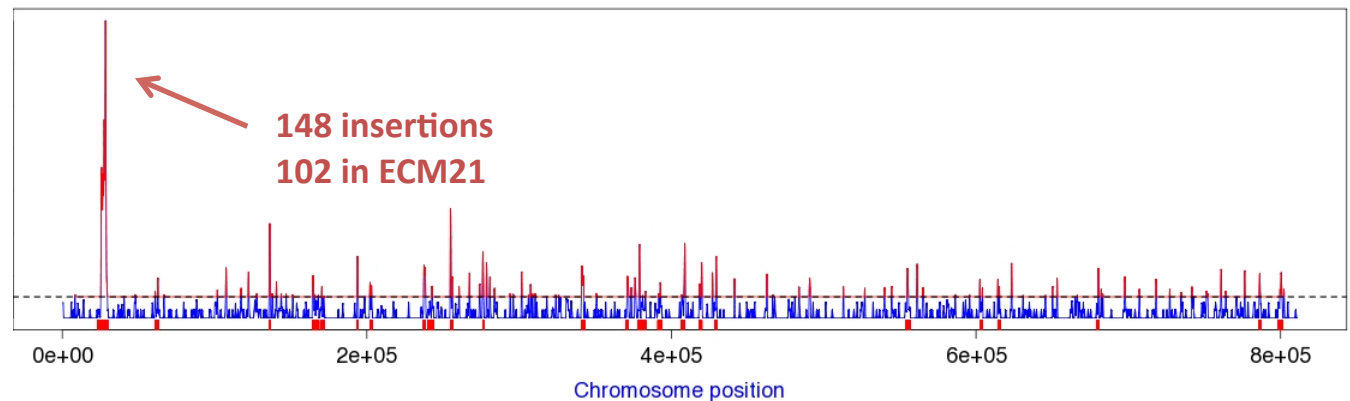
Results

HOT REGIONS OF THE YEAST GENOME

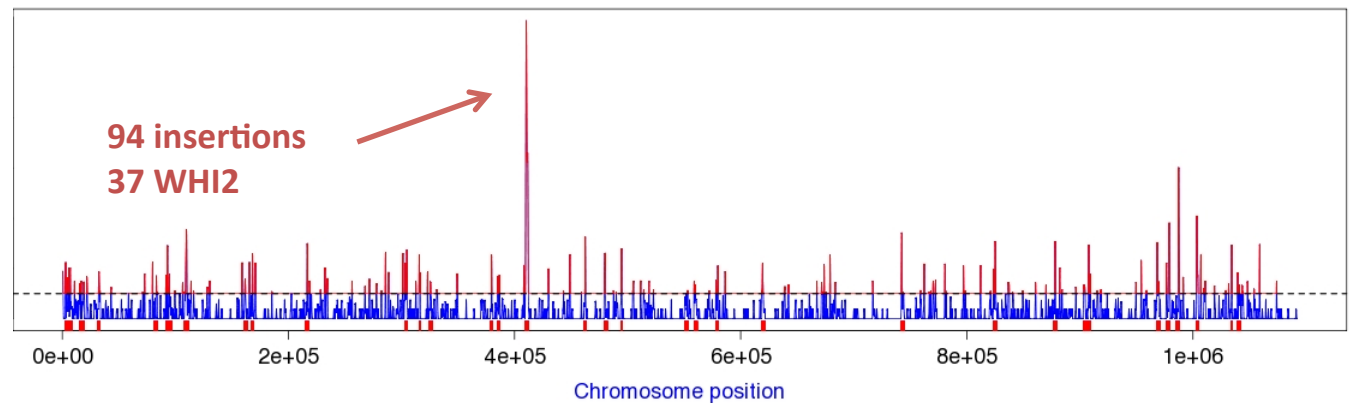
1. Clusters: **365** regions where at least 10 insertions are no more than 1Kb apart



Clusters of insertions in Chromosome 2



Clusters of insertions in Chromosome 15



Results

HOT REGIONS OF THE YEAST GENOME

2. Number of times (frequency) the same event occurs in independent experiments

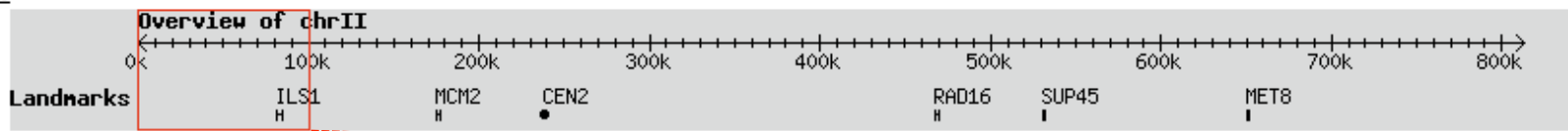
# Repeated sites	1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8
insertions	14394	2563	826	354	71	12	6	5

# Repeated sites	1	2	3	4	5	6	7	8+
Simulated data	17151	501	23	1	1	0	0	0

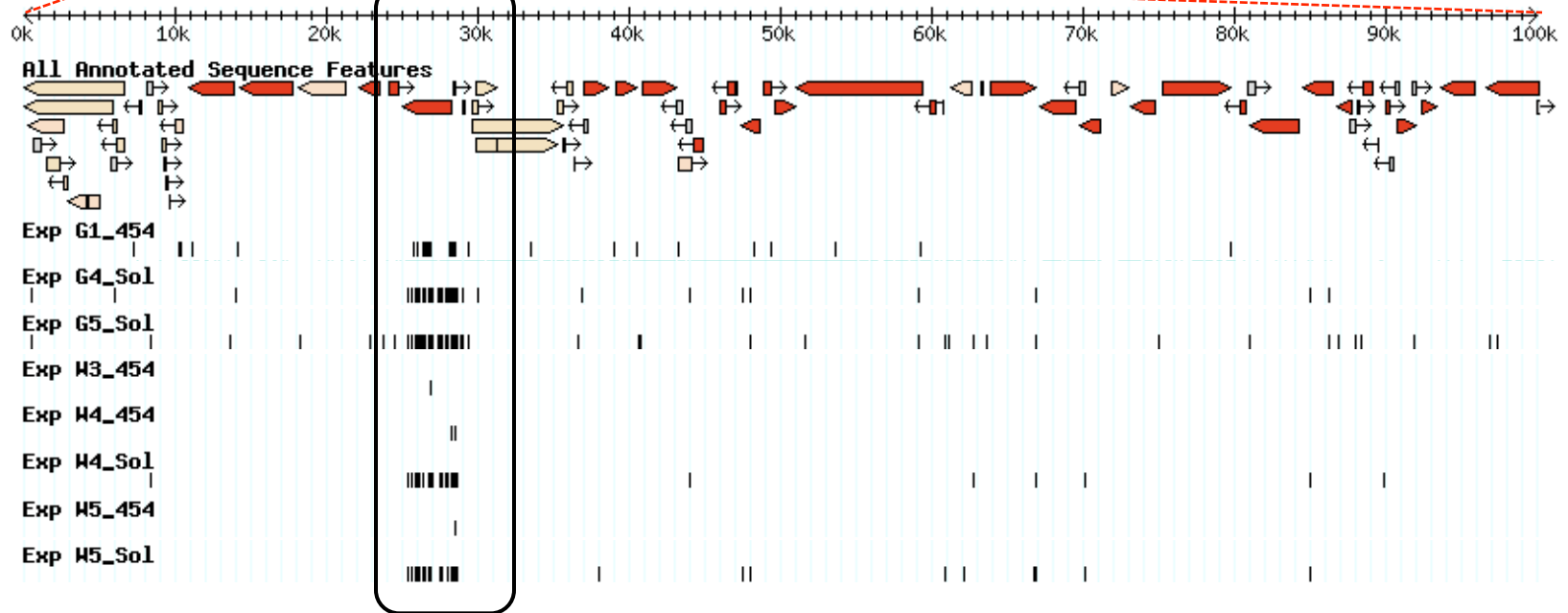
Results

ECM21 LOCUS IN CHROMOSOME II

Overview



Details

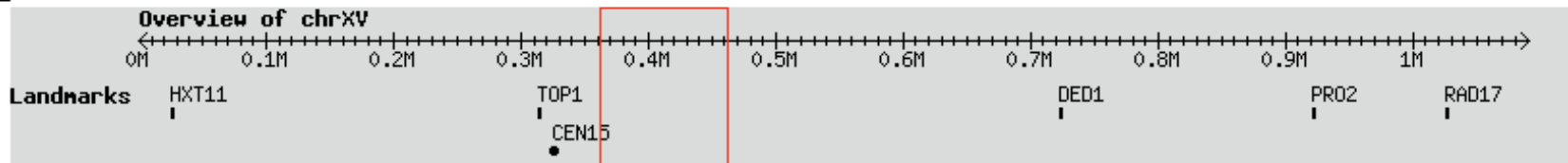


ECM21

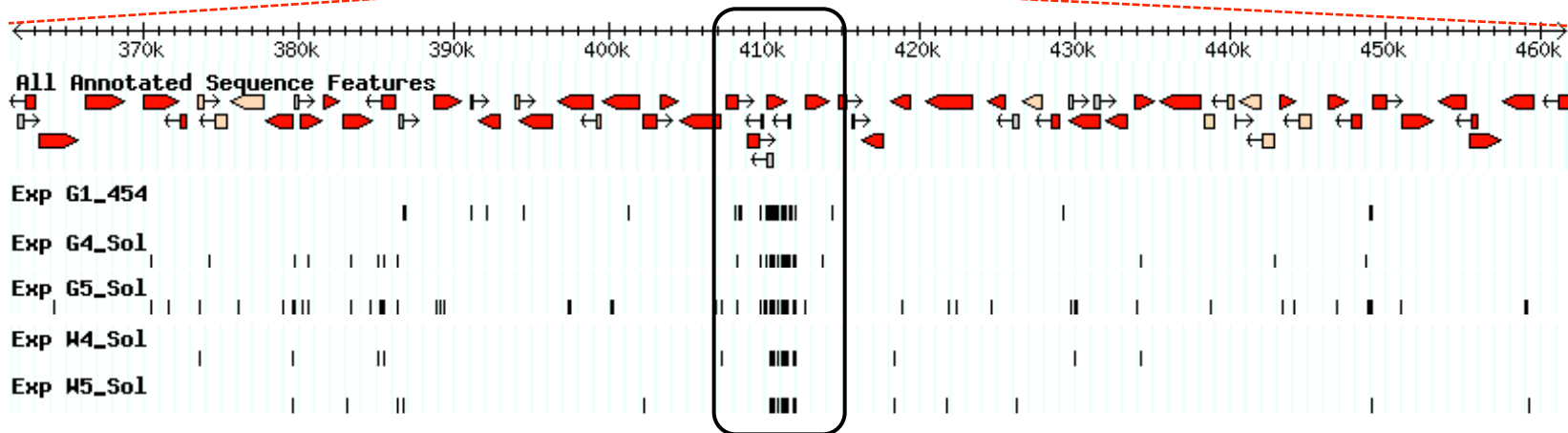
Results

WHI2 LOCUS IN CHROMOSOME XV

Overview



Details



WHI2

Conclusion

- *Hermes* insertion sites are determined in part by the DNA sequence and in part by the genome features
- *Hermes* tends to insert upstream of ATG and TSS, and downstream of ORF 3p-end, but the strongest correlation is with nucleosome free regions (NFRs)
- *Hermes* preferentially inserts in intergenic regions (IGRs) over ORFs
- Some regions in the yeast genome are more prone to accumulate insertions (for examples ECM21 and WHI2)

Acknowledgements

Oncology Biostatistics and Bioinformatics



Sarah Wheelan

Giovanni Parmigiani

Rafael Irizarry

Luigi Marchionni

Jeff Leek

Hector Corrada Bravo

Martin Aryee

Department of Molecular Biology & Genetics

Nancy Craig

Sunil Gangadharan

www.genomics.jhu.edu