Transgenics, bombardments, and chips

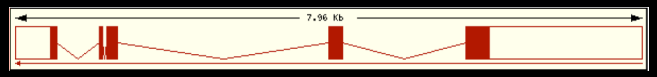
Steven Kuntz

August 10, 2005

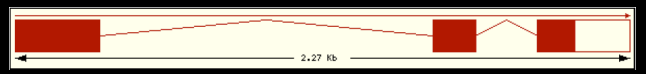
- Hox project status
- Ballistic Injections
- ChIP-chip algorithm and results

C. elegans vs. C. briggsae vs. CB5161: lin-39 and ceh-13

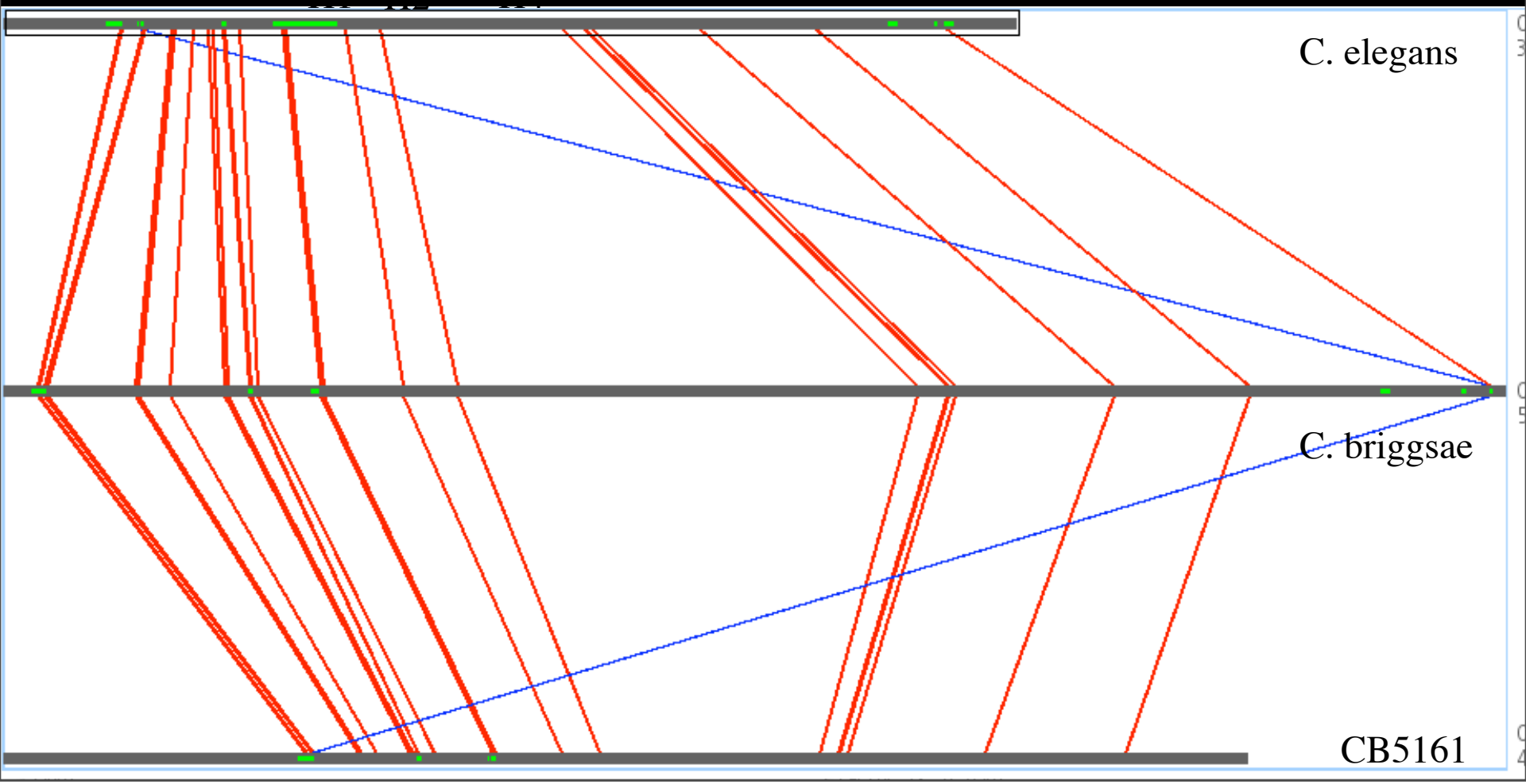
- Window: 30 Threshold: 25



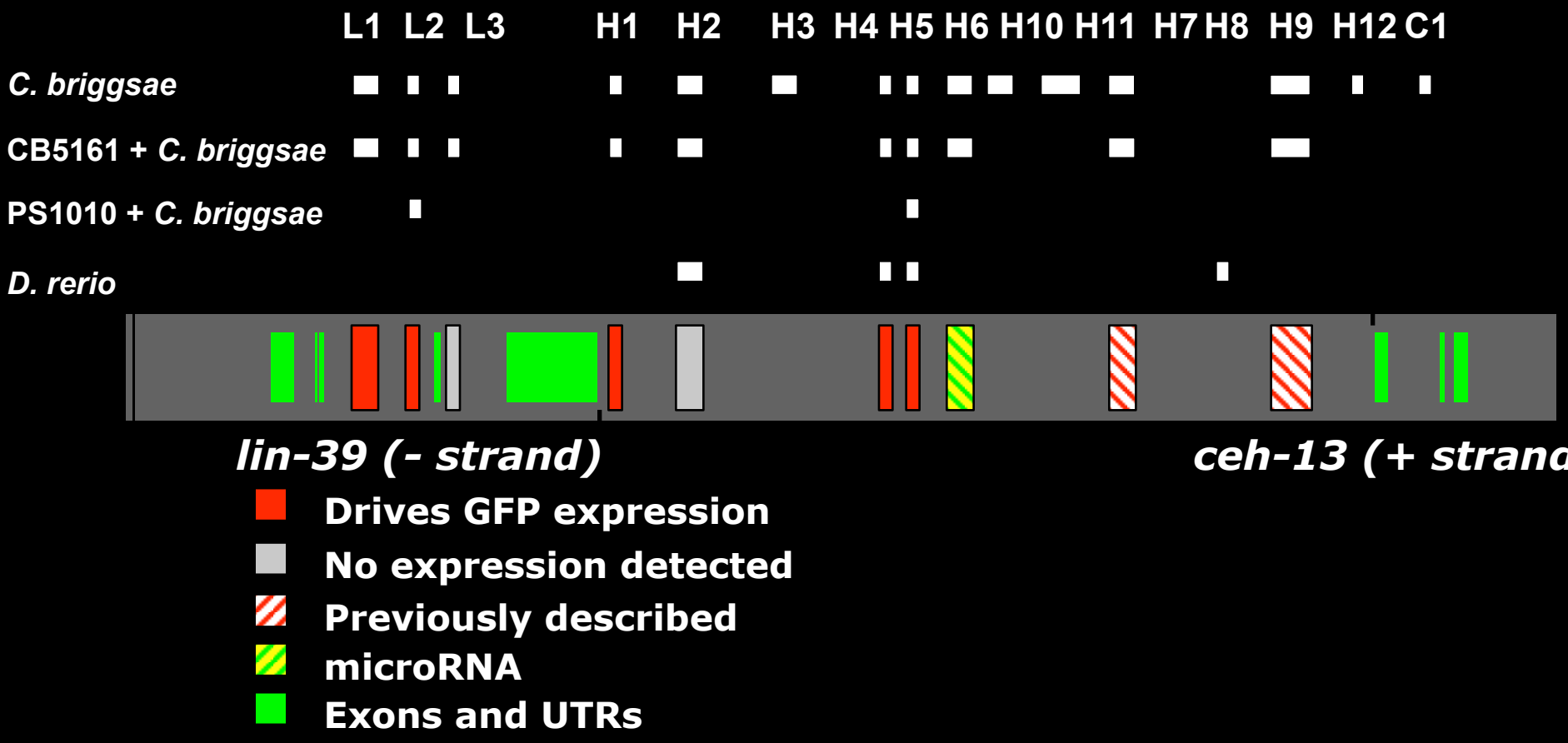
Lin-39 (- strand)



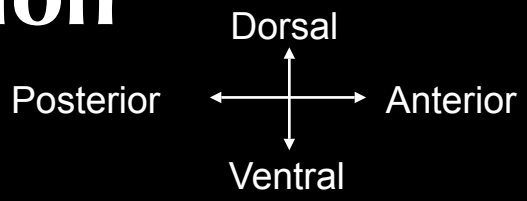
Ceh-13 (+ strand)



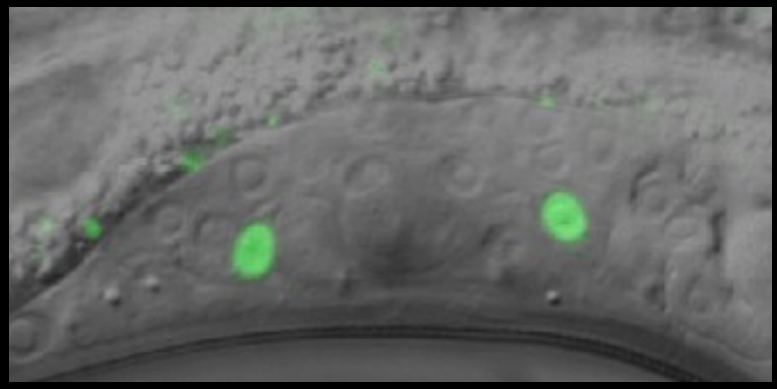
Summary of Computationally Identified Elements in Common with *C. elegans*



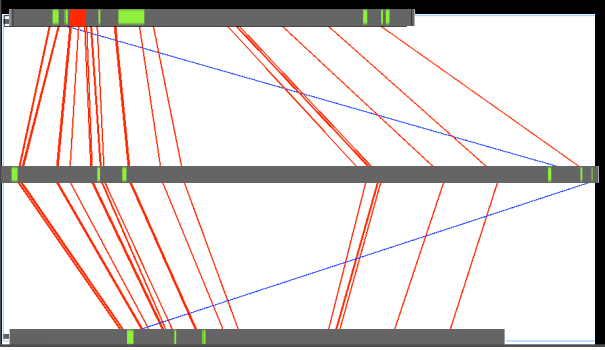
Enhancer L1 expression



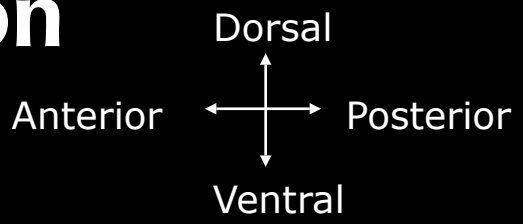
Expression in L2 posterior intestine

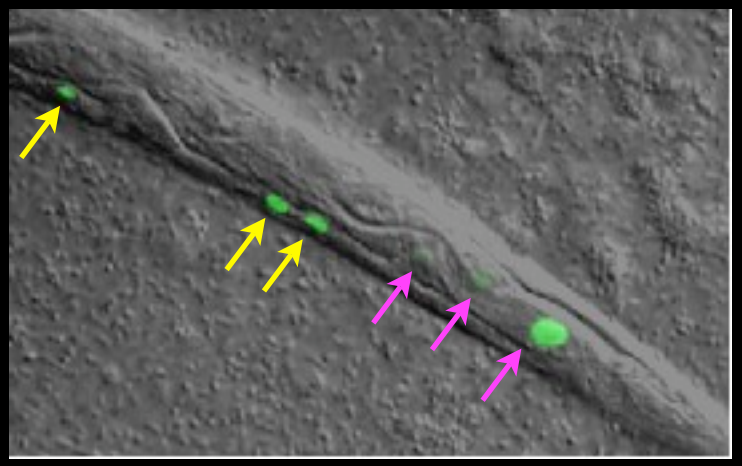


Expression in adult vulval muscles



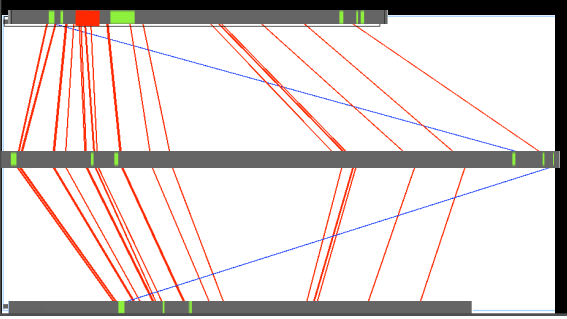
Enhancer L2 expression



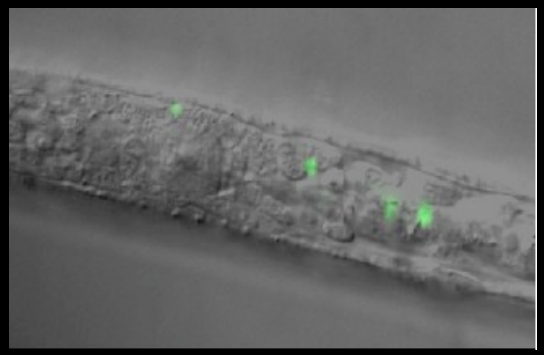


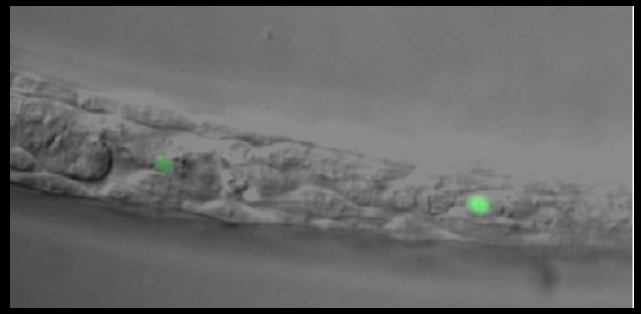
Expression in L1 ventral chord neurons

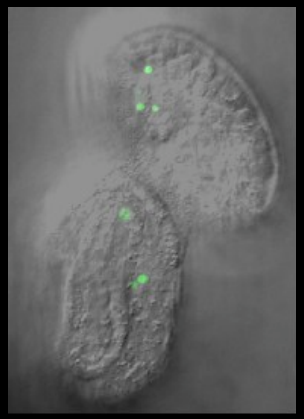
Expression in L1 posterior ventral chord neurons and Q cells

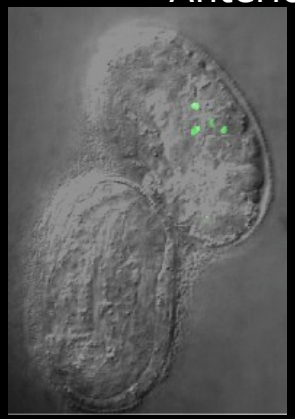


Enhancer H1 expression

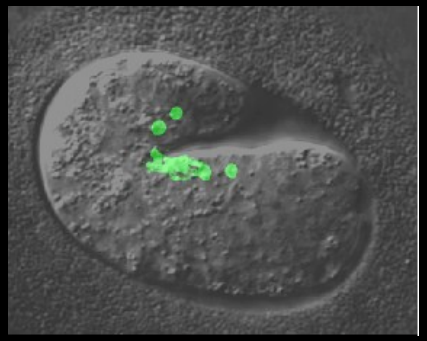




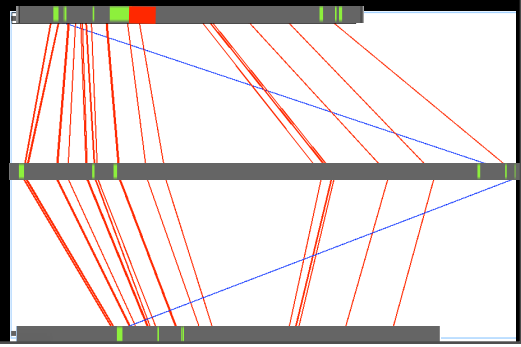




Dorsal
Anterio
Posterior
Ventral

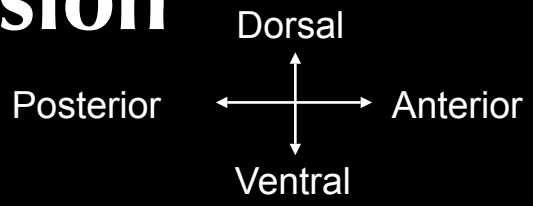


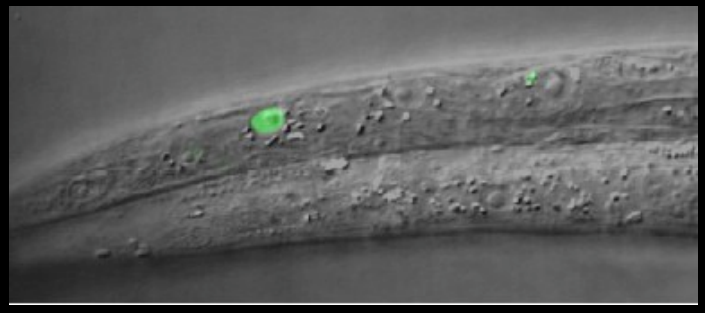
Expression in embryos



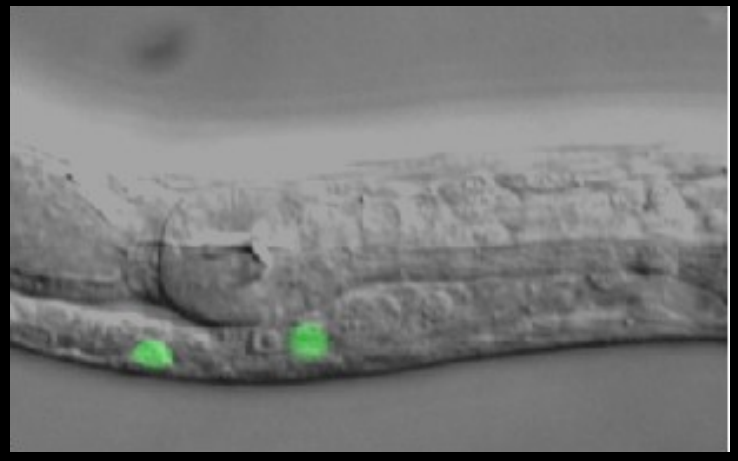
Expression in L1 in intestine and V cells?

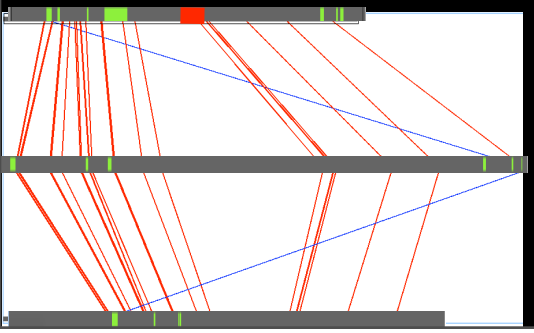
Enhancer H4 expression





Expression in L2 posterior

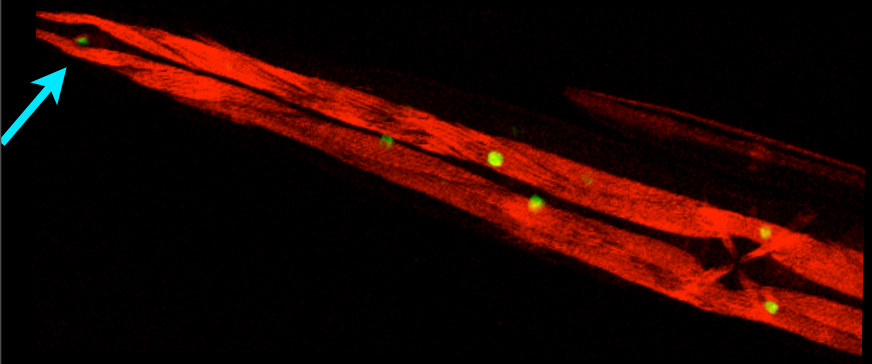


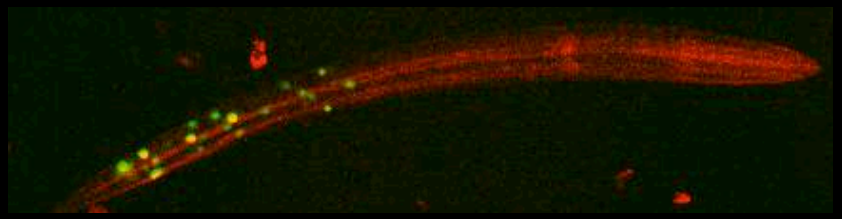


Expression in L2 anterior in excretory and G cells?

Enhancer H5 expression

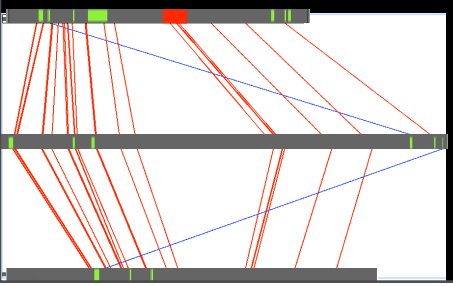
Posterior ⟷ Anterior





Expression in larval posterior bodywall muscle and anal depressor cell

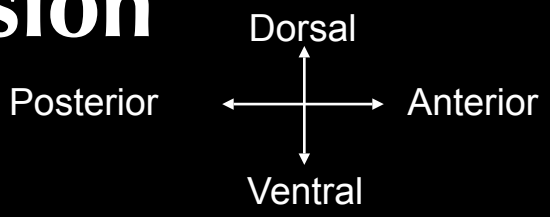
Expression in adult posterior bodywall muscle



Not pictured: Expression in posterior enteric muscle and posterior intestine

Enhancer H7 expression

- Described by Streit, et al.

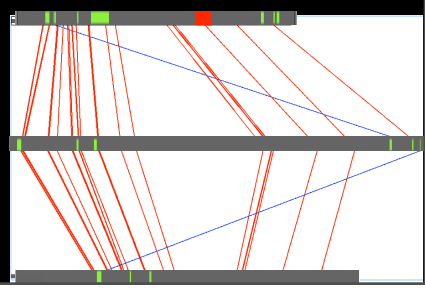


Expression in adult mid-body in gonad?

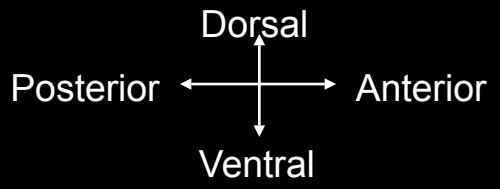
Expression in adult anterior bodywall muscle

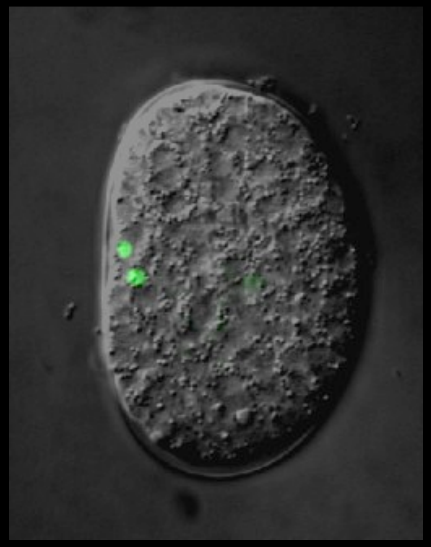
Ventral view of adult mid-body expression

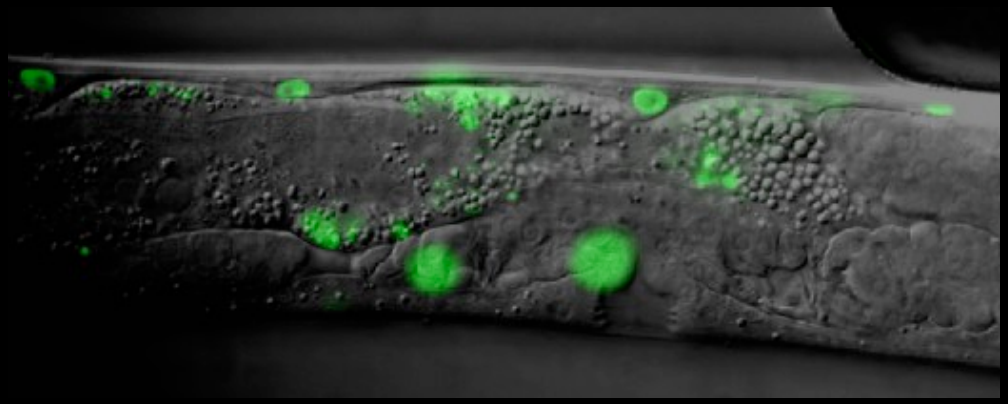
Not pictured: rectal dorsal neuron



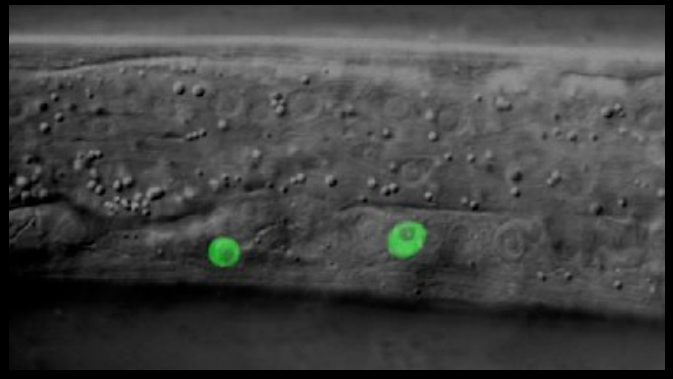
Upstream Region W2

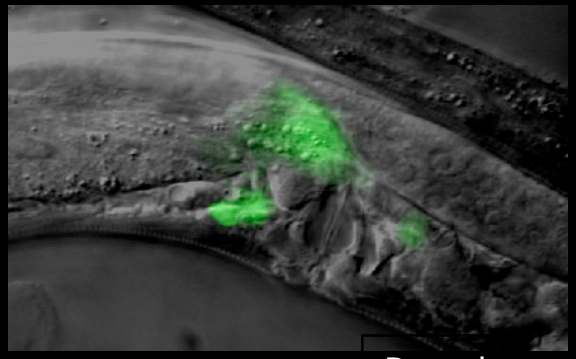


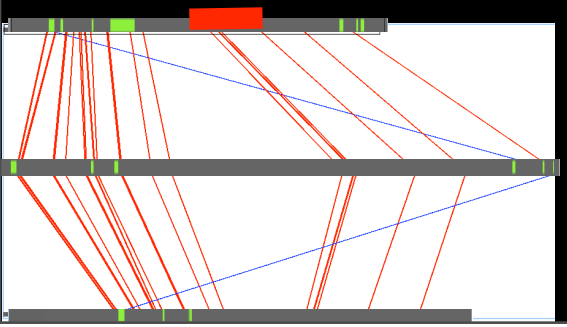




Embyronic expression







Adult mid-body expression in bodywall muscle, dermis, and intestine?

Future directions

- Male expression
  - him-14 RNAi
- Intervening expression
  - elements cloned, awaiting injection
- Finish summation expression
  - elements cloned, awaiting injection

Status of Ballistic Injections (Biolistics)

- Procedure
  - Growth conditions
  - Nematode purification
  - Gold bead preparation
  - DNA-coating
- Co-injection markers
- DNA
  - Preparation
  - Topology

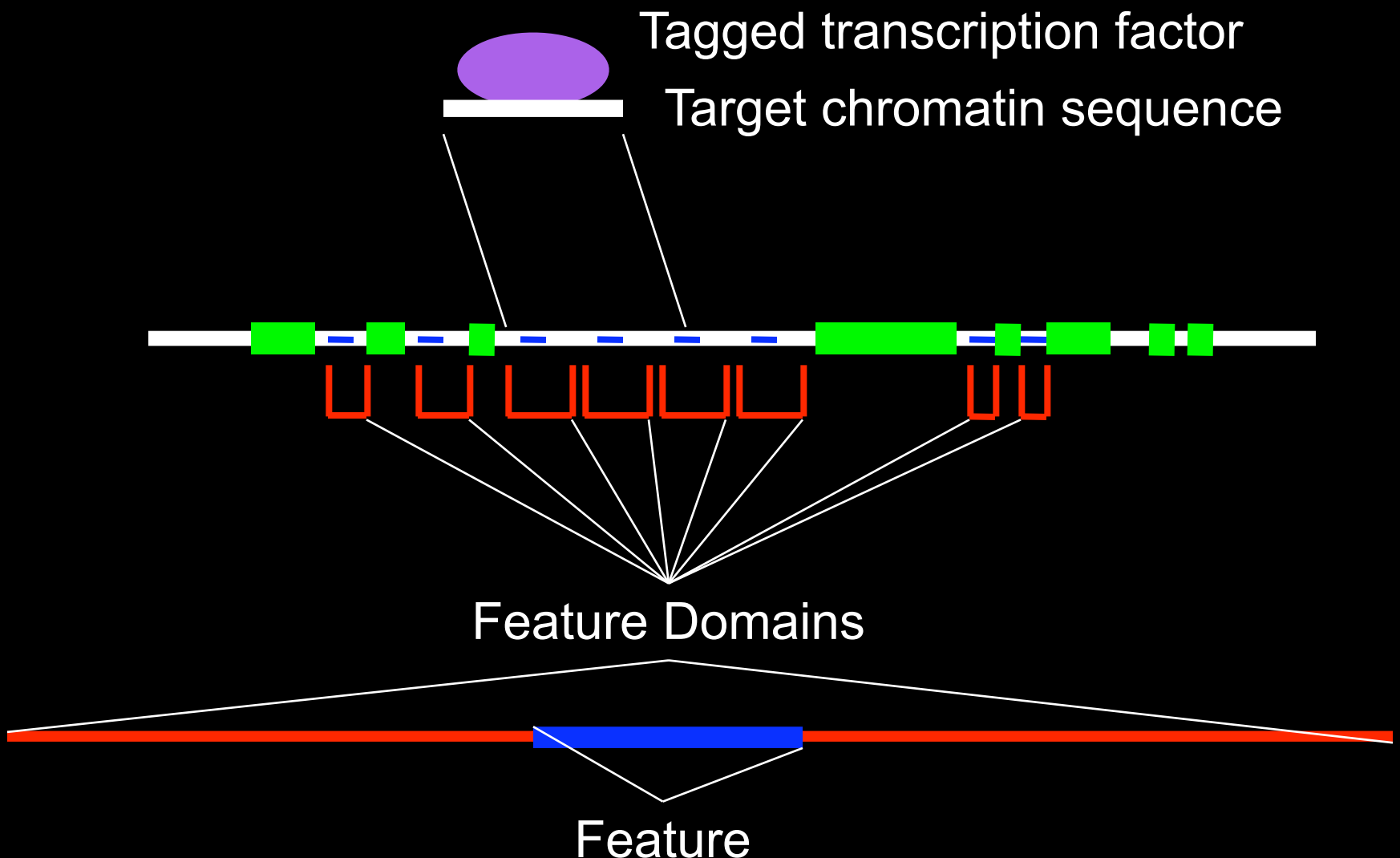
Cotransformation

- Myo-2::gfp and unc-54::gfp
  - 100% cotransformation in 41 F1's
- Myo-2::gfp and pPD107.94
- Myo-2::gfp and pha-1 rescue
- Unc-119 rescue and experimental constructs

Time and efficiency for liquid culture-based ballistics

- 10-40 shots in 2 hours
  - 1-6 constructs per hour if in triplicate
  - Minimum size determination *in process*
  - Not including time to make solutions/set up bacteria
- ~1 stable line per shot
  - Published rate of 30% integration
- Several week turn-around time

ChIP-chip feature determination analysis



Feature Domains and Features

- Feature Domain of variable size
  - Minimum size set at 200 bp
- Features always at 50 bp
- Variations
  - Interexonic regions
    - Intergenic plus introns
    - Exons
    - Small introns
  - Indiscriminant placement
- Based on Li and Stormo. *Bioinformatics* 17(11):1067-1076, 2001.

Control features

- Random sequence
  - Generated 1000 unique sequences not found in *C. elegans*
  - *De novo* generation of IDT oligos to clone these fragments
    - ~$30 per oligo
- Complex repeats
  - ~682 independently defined complex repeats
- *C. briggsae* and *C. remanei*
  - 90 genes
  - 10 kb upstream and downstream included

Fatal errors

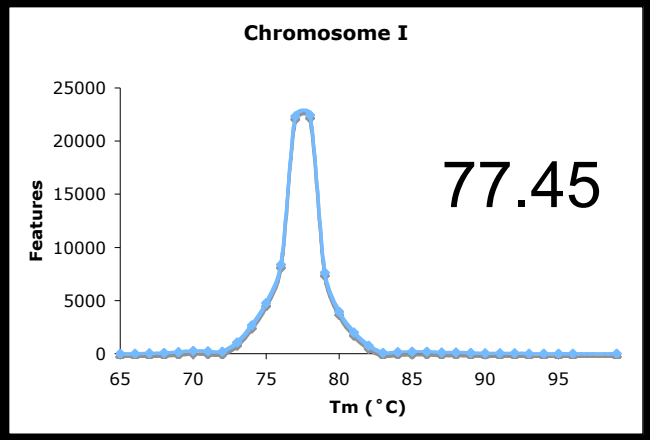
- Masked sequence
  - Threshold set at 10 bp
  - Set at 0 for *C. remanei* and *C. briggsae*
- Base content
  - Threshold for content set at 50% (25 bp)
- Strings of bases
  - Threshold set at 30% (15 bp)

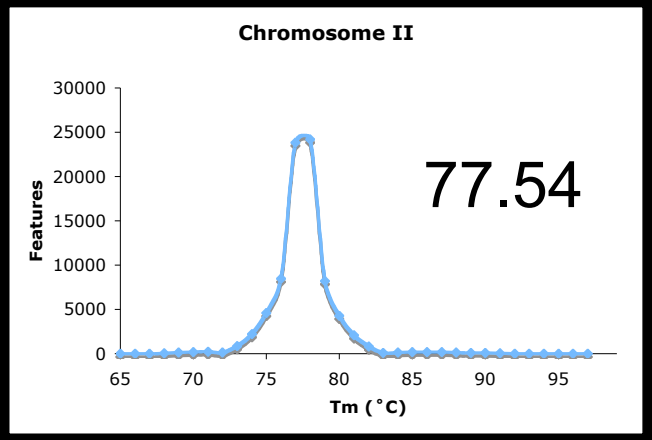
Self-hybridization

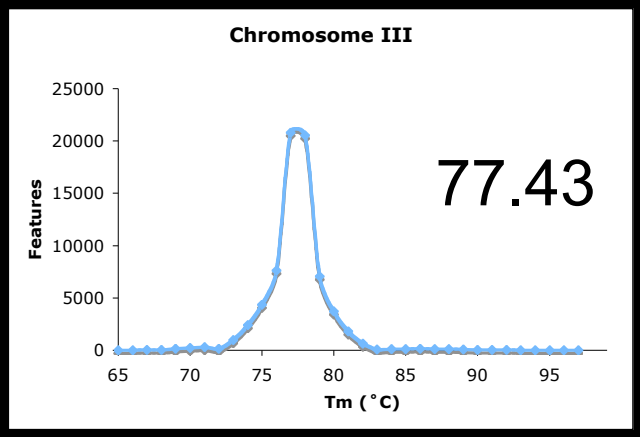
- Sliding window analysis
- Optimized for fewer than 15 bp self-hybridization

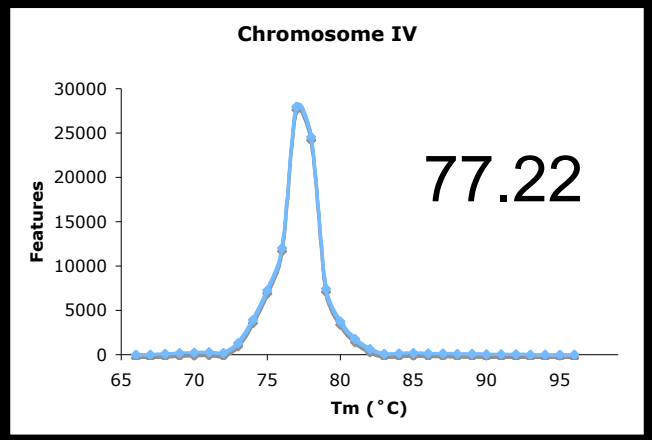
Tm

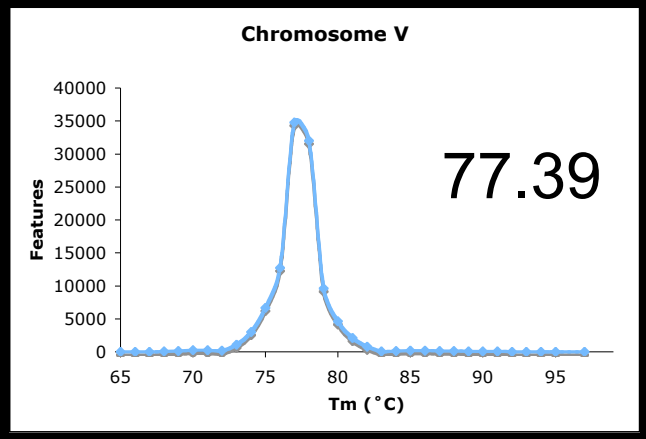
- Nearest-neighbor and initialization approach
- $\Delta H = \sum H_{partial} + \sum H_{init}$ and $\Delta S = \sum S_{partial} + \sum S_{init}$
- Salt correction
- $Tm = \Delta H/(\Delta S+R\ln(c/4)) - 273.15 + 16.6\log[K+]$
- Based on
  - SantaLucia. *PNAS* 95:1460-1465 1998.
  - Allawi and SantaLucia. *Biochemistry* 36(34):10581-94, 1997.
  - Rychlik, Spencer, and Rhoads. *Nucleic Acids Research* 18(21):6409-6412 1990.

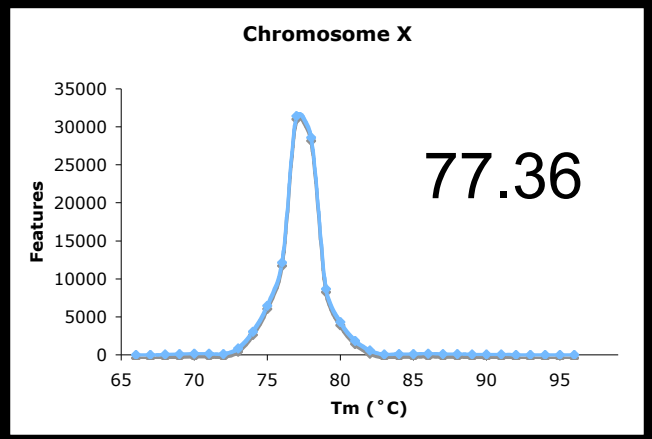

Chromosome I
Features
25000
20000
15000
10000
5000
0
65
70
75
80
85
90
95
Tm (˚C)
77.45

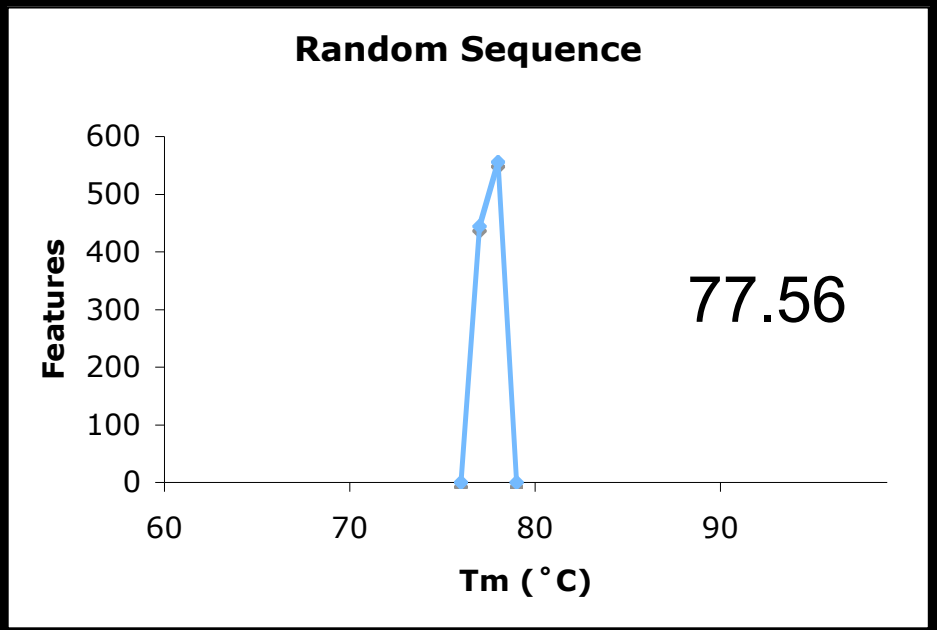

Chromosome II
Features
30000
25000
20000
15000
10000
5000
0
65
70
75
80
85
90
95
Tm (˚C)
77.54

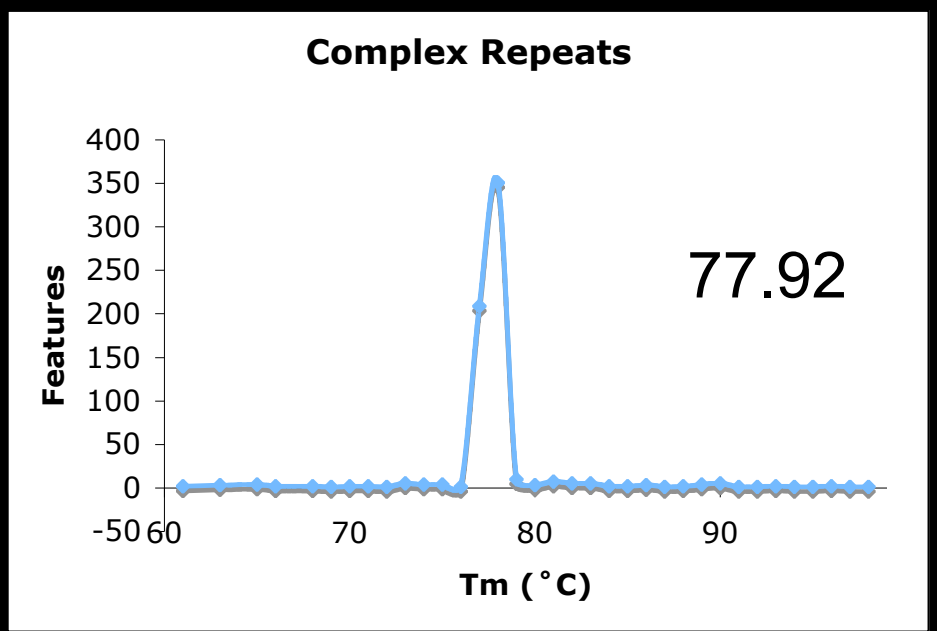

Chromosome III
Features
25000
20000
15000
10000
5000
0
65
70
75
80
85
90
95
Tm (˚C)
77.43


Chromosome IV
Features
30000
25000
20000
15000
10000
5000
0
65
70
75
80
85
90
95
Tm (˚C)
77.22


Chromosome V
Features
40000
35000
30000
25000
20000
15000
10000
5000
0
65
70
75
80
85
90
95
Tm (˚C)
77.39

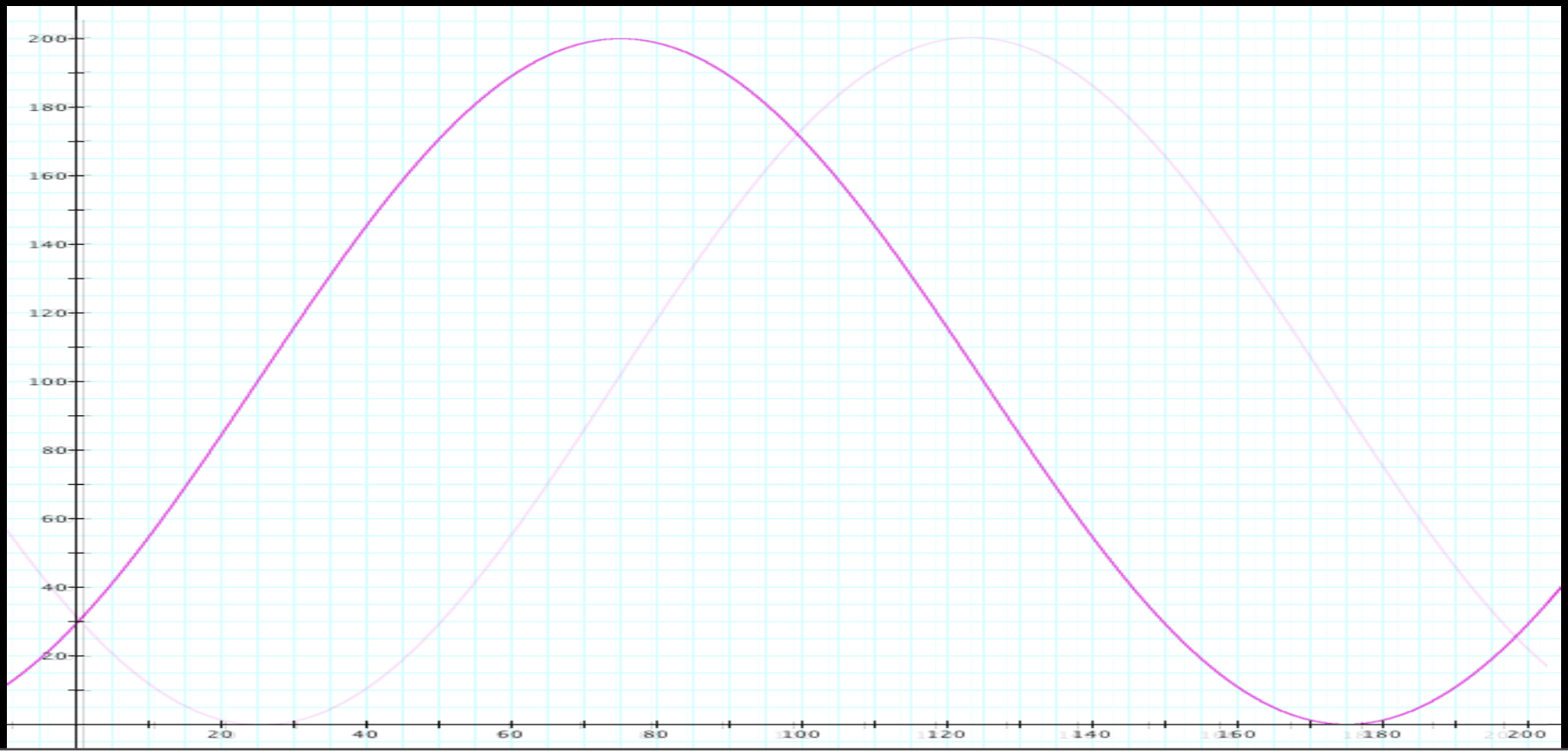

Chromosome X
Features
35000
30000
25000
20000
15000
10000
5000
0
65
70
75
80
85
90
95
Tm (˚C)
77.36


Random Sequence
600
500
400
300
200
100
0
Features
60
70
80
90
Tm (˚C)
77.56


Complex Repeats
400
350
300
250
200
150
100
50
0
-50
Features
60
70
80
90
Tm (˚C)
77.92

Spacing

- Position weight array created
- 1-cos((2*π/featureDomain)*(k+25))



Blat

- Order sequences according to Tm, position, self-binding properties
- Similar to blast
  - ~150-fold faster
  - Different mismatch handling from Blast
- Permit
  - Single blat match
  - Blat match with double the identity of the remaining matches

Speed

- ~2 Mbp/hour for indiscriminant analysis
  - Longer for excluded domains
  - Shorter for included domains
  - Processor speed limited
- ~1 Mbp/hour for feature selective analysis
  - Speed a function of region length
  - Processor speed limited

Results and Limitations

- Chromosome III at minimal domain size of 200:
  - Even, indiscriminant tiling:
    - 85% of 68,916 potential domains had acceptable features
    - Yield of 58,722 features
  - Selective feature tiling:
    - 85% of 73,392 potential domains had acceptable features
    - Yield of 62,347 features
    - Breakdown:
      - Interexonic: 40,470
      - Exons: 18,749
      - Small introns: 3125

Results and Limitations

- Genome scan at minimal domain size of 200:
  - Even, indiscriminant tiling:
    - 86.7% of potential domains had acceptable features
    - Yield of 434,608 features
- Genome scan at minimal domain size of 150:
  - Even, indiscriminant tiling:
    - ??% of potential domains had acceptable features
    - Yield of ~555,700 features

Final spacing

Final spacing of ~160 base pairs for a feature domain

532,183 features

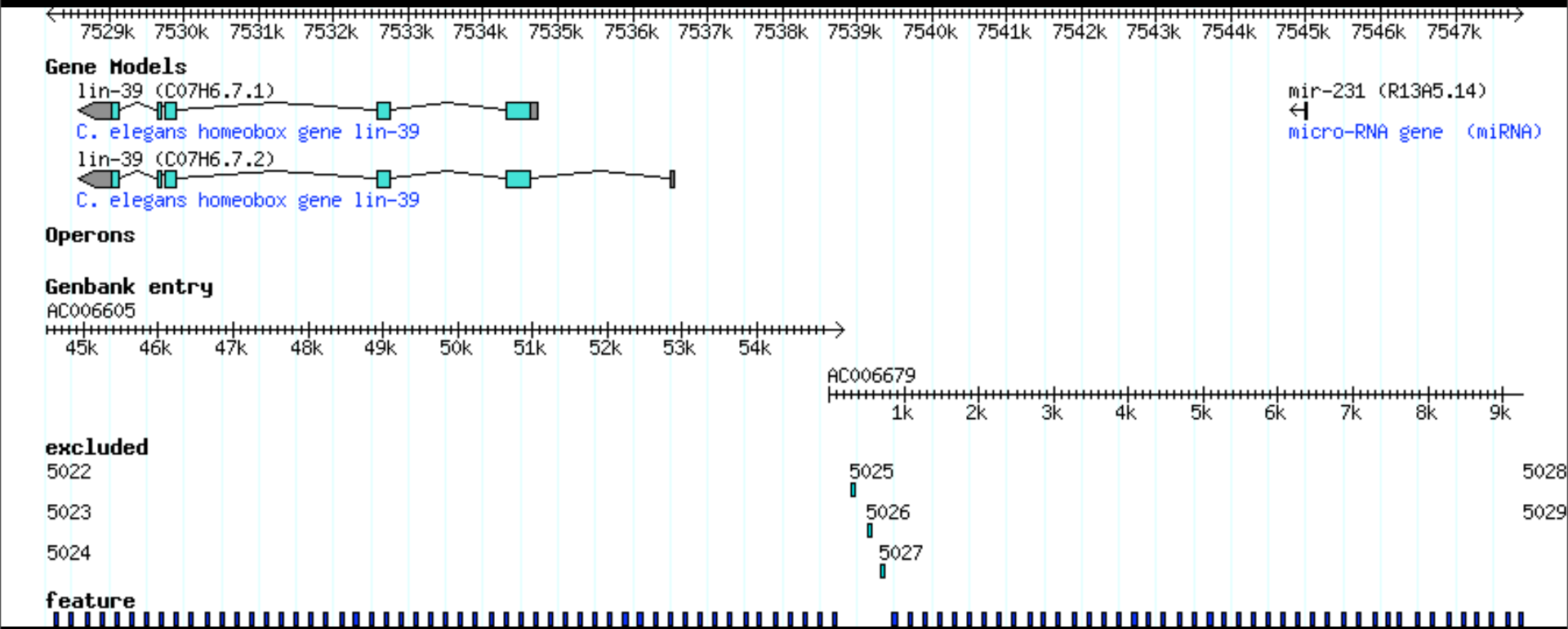
Chrom.	Total	Included	%
I	94202	77903	82.70
II	95495	81336	85.17
III	86145	71642	83.16
IV	109336	92530	84.62
V	130763	109380	83.64
X	110742	99392	89.75
total	626683	532183	84.92

Final Tm Distribution

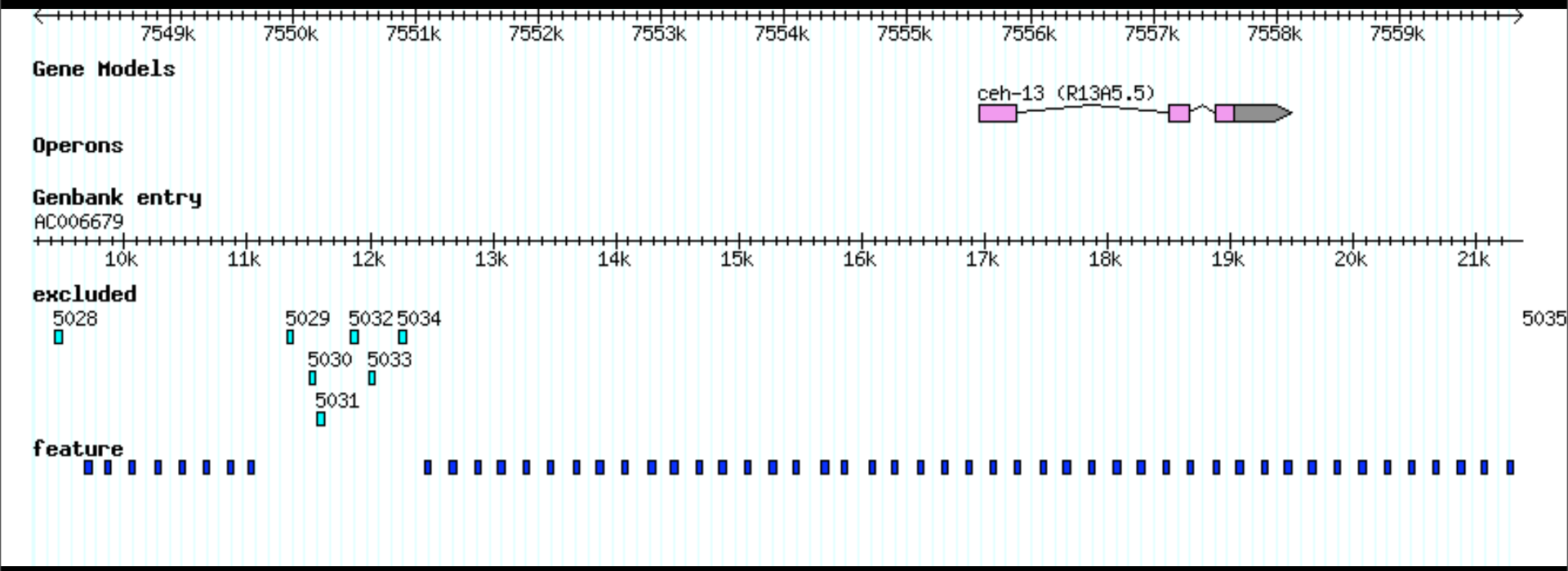
Chrom	% in +/- 5˚C	% in +/- 3˚C
I	97.67	89.18
II	97.87	90.45
III	97.78	89.40
IV	98.12	89.69
V	98.37	91.82
X	98.68	92.15
Total	98.13	90.59

Result visualization

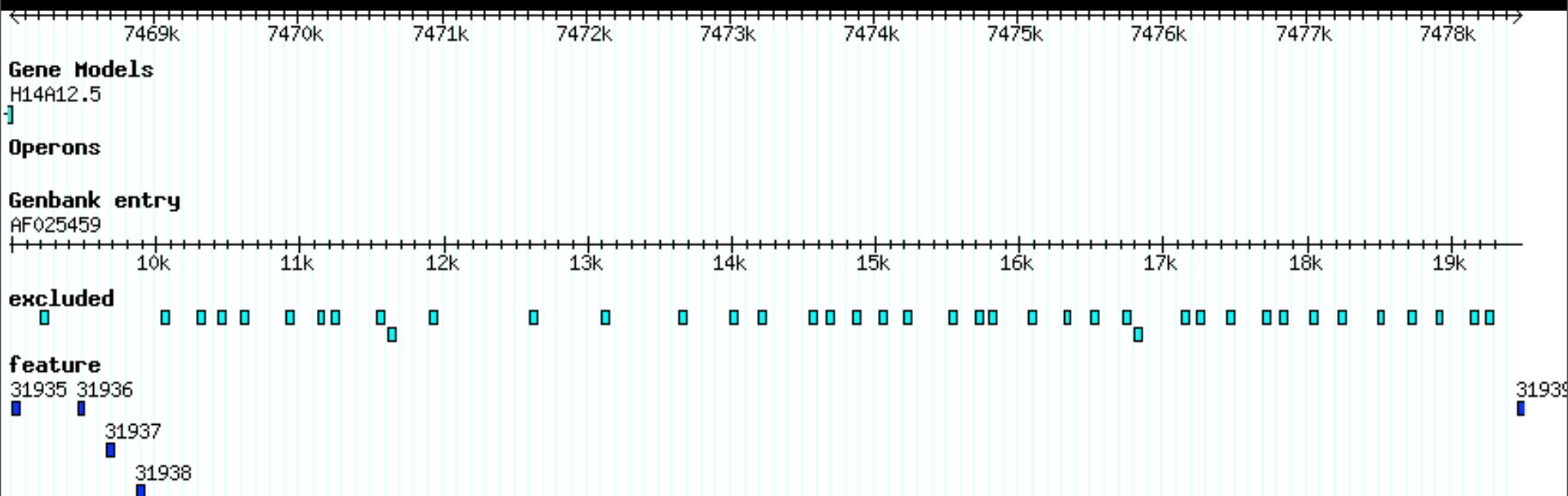
lin-39 and hox cluster



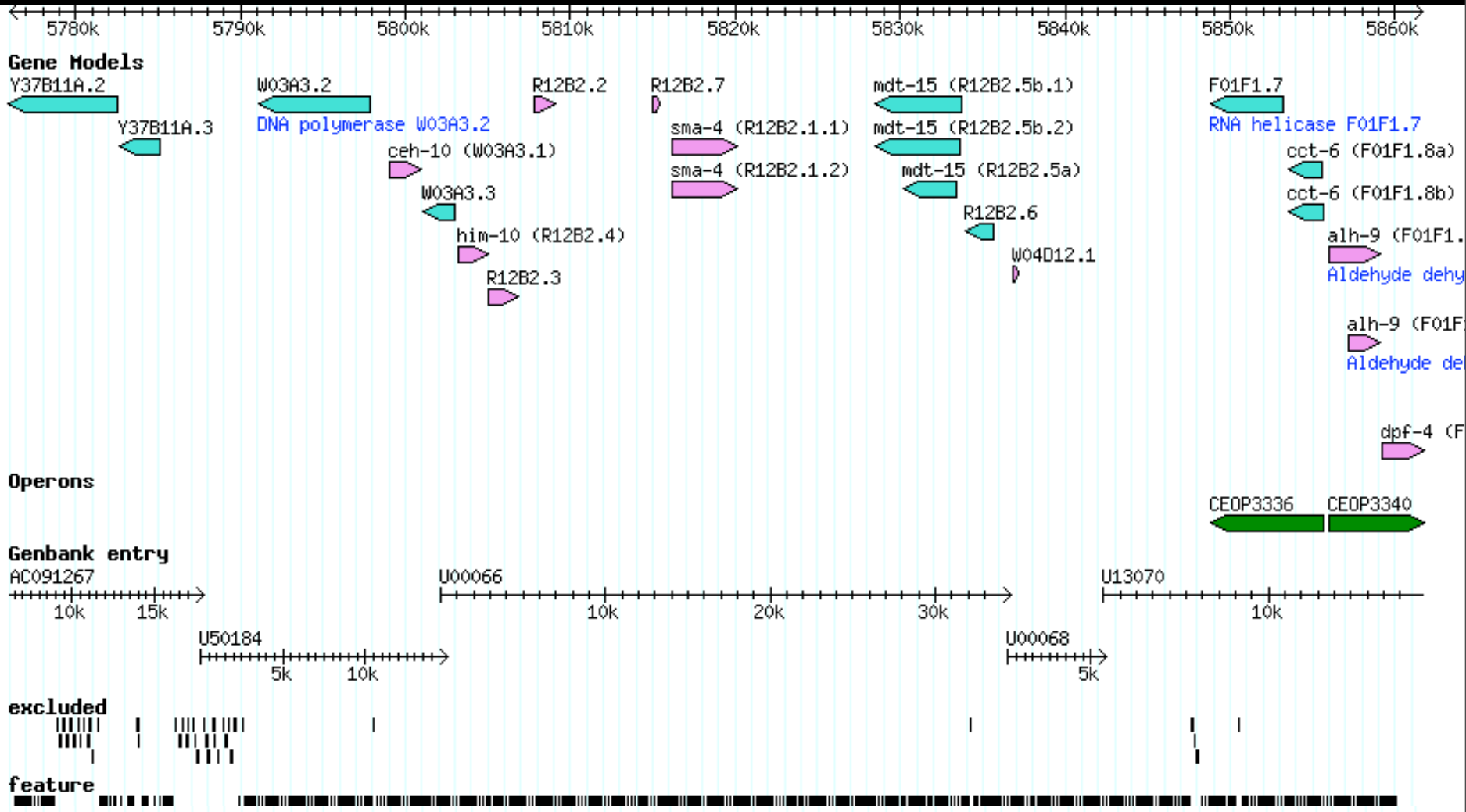
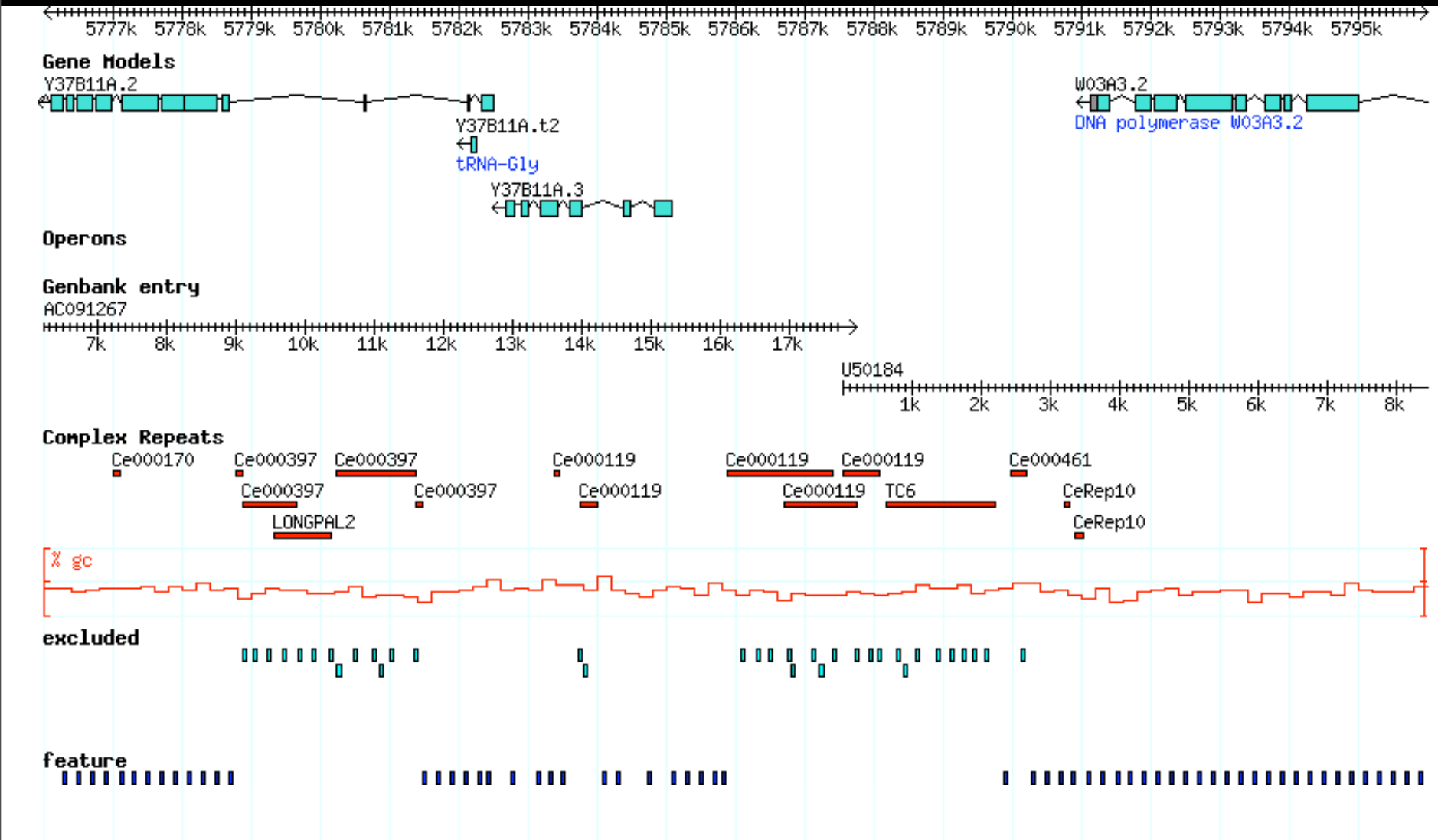
ceh-13 and hox cluster



Gene depleted region



Gene rich region



Zoomed in portion of gene-rich