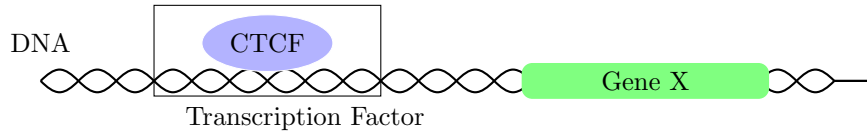


Interval	Extended	Sorted	Index	Diff	Value	P-value	Peak
(0, 10, +)	(0, 24)	(0, 24)	0	1	1	0.93	0
(21, 31, -)	(7, 31)	(2, 26)	2	1	2	0.81	0
(7, 17, +)	(7, 31)	(3, 27)	3	1	3	0.63	0
(2, 12, +)	(2, 26)	(7, 31)	7	2	5	0.27	0
(17, 27, -)	(3, 27)	(7, 31)	13	1	6	0.15	0
(17, 27, +)	(17, 41)	(13, 37)	17	1	7	0.08	0
(33, 43, -)	(19, 43)	(17, 41)	19	1	8	0.03	1
(27, 37, -)	(13, 37)	(19, 43)	24	-1	7	0.08	0
	$O(n)$	$O(n \log n)$	26	-1	6	0.15	0
			27	-1	5	0.27	0
			31	-2	3	0.63	0
			37	-1	2	0.81	0
			41	-1	1	0.93	0
			43	-1	0	0.99	0

yielding the interval $[19, 24)$ as the only peak.

In a real experiment, more intervals would be yielded which would be affected by noise in the experiment. In order to reduce the effect of noise, two postprocessing steps are performed. First, small gaps between intervals are removed, then small peaks are removed:

Interval	Small Gaps Filled	Small Peaks Removed
(121, 125)	(121, 125)	
(142, 147)	(142, 166)	(142, 166)
(151, 166)	(172, 200)	(172, 200)
(172, 200)		



(a) Transcription factor binding to DNA, regulating the expression of Gene X

CTCF

ACGTTTCGTATATCGTAGCTACTCGAGCTGTAGTTTGATAGATAT

CTCF

ACGTTTCGTATATCGTAGCTACTCGAGCTGTAGTTTGATAGATAT
 ACGTTTCGTATATCGTAGCTACTCG.....
TATATCGTAGCTACTCGAGCTGTA.....
TCGTAGCTACTCGAGCTGTAGTTT.....
 ..GTTTCGTATATCGTAGCTACTCGAG.....
 ...TTCGTATATCGTAGCTACTCGAGC.....
CTACTCGAGCTGTAGTTTGATAGA...
ACTCGAGCTGTAGTTTGATAGATA..
GTAGCTACTCGAGCTGTAGTTTGA.....

(b) DNA Fragments obtained from ChIP

CTCF

ACGTTTCGTATATCGTAGCTACTCGAGCTGTAGTTTGATAGATAT
ACGTTTCGTATATCGTAGCTACTCG.....
TATATCGTAGCTACT**TCGAGCTGTA**.....
**TCGTAGCTACT**CGAGCTGTAGTTT.....
 ..**GTTTCGTATAT**CGTAGCTACTCGAG.....
 ...TTCGTATATCGTAG**CTACTCGAGC**.....
**CTACTCGAGCT**GTAGTTTGATAGA...
ACTCGAGCTGTAGT**TTGATAGATA**..
GTAGCTACTCGAGC**TGTAGTTTGA**.....

↓

ACGTTTCGTAT, TACAGCTCGA, TCGTAGCTAC, GTTTCGTATAT
 GCTCGAGTAG, CTACTCGAGC, TATCTATCAA, TCAAACATACA

(c) One end of each fragment is sequenced

ACGTTTCGTATATCGTAGCTACTCGAGCTGTAGTTTGATAGATAT
 [----->.....
<-----].
[----->.....
 ..[----->.....
<-----].
[----->.....
<-----].
<-----].

(0, 10, +), (21, 31, -), (7, 17, +), (2, 12, +)
 (17, 27, -), (17, 27, +), (33, 43, -), (27, 37, -)

(d) Sequenced reads mapped to the reference genome

\$	A	C	G	T	T	A	C	G	G	T	A	C	C	G	T	A	T
A	\$	C	G	T	A	\$	A	C	G	T	A	C	C	G	T	A	T
A	C	C	G	T	A	C	C	G	T	A	\$	A	C	C	G	T	\$
A	C	C	G	T	A	C	C	G	T	A	\$	A	C	C	G	T	A
C	C	C	G	T	A	\$	A	C	G	T	A	C	C	G	T	A	A
C	C	C	G	T	A	\$	A	C	G	T	A	C	C	G	T	A	\$
C	C	C	G	T	A	\$	A	C	G	T	A	C	C	G	T	A	\$
G	G	G	T	A	\$	A	C	G	T	A	\$	A	C	G	T	A	\$
G	G	G	T	A	\$	A	C	G	T	A	\$	A	C	G	T	A	\$
G	G	G	T	A	\$	A	C	G	T	A	\$	A	C	G	T	A	\$
T	T	A	\$	A	C	C	G	T	A	\$	A	C	C	G	T	A	\$
A	T	\$	A	C	C	G	T	A	\$	A	C	C	G	T	A	\$	\$
C	A	C	G	T	A	\$	A	C	G	T	A	\$	A	C	G	T	\$
G	C	C	G	T	A	\$	A	C	G	T	A	\$	A	C	G	T	\$
G	C	C	G	T	A	\$	A	C	G	T	A	\$	A	C	G	T	\$
T	T	A	\$	A	C	C	G	T	A	\$	A	C	C	G	T	A	\$
A	T	\$	A	C	C	G	T	A	\$	A	C	C	G	T	A	\$	\$
C	A	C	G	T	A	\$	A	C	G	T	A	\$	A	C	G	T	\$
G	C	C	G	T	A	\$	A	C	G	T	A	\$	A	C	G	T	\$
G	C	C	G	T	A	\$	A	C	G	T	A	\$	A	C	G	T	\$
T	T	A	\$	A	C	C	G	T	A	\$	A	C	C	G	T	A	\$
A	T	\$	A	C	C	G	T	A	\$	A	C	C	G	T	A	\$	\$
C	A	C	G	T	A	\$	A	C	G	T	A	\$	A	C	G	T	\$
C	C	C	G	T	A	\$	A	C	G	T	A	\$	A	C	G	T	\$
G	C	C	G	T	A	\$	A	C	G	T	A	\$	A	C	G	T	\$
T	G	G	T	A	\$	A	C	G	T	A	\$	A	C	G	T	A	\$
A	T	T	\$	A	C	C	G	T	A	\$	A	C	C	G	T	A	\$

16	\$	A	C	G	T	T	A	C	G	G	T	A	C	C	G	T	A
15	A	\$	C	G	T	A	\$	A	C	G	T	A	C	C	G	T	A
10	A	C	C	G	T	A	C	C	G	T	A	\$	A	C	C	G	T
5	A	C	C	G	T	A	C	C	G	T	A	\$	A	C	C	G	T
0	A	C	C	G	T	A	C	C	G	T	A	\$	A	C	C	G	T
11	C	C	G	T	A	\$	A	C	G	T	A	\$	A	C	C	G	T
6	C	C	G	T	A	\$	A	C	G	T	A	\$	A	C	C	G	T
12	C	C	G	T	A	\$	A	C	G	T	A	\$	A	C	C	G	T
1	C	C	G	T	A	\$	A	C	G	T	A	\$	A	C	C	G	T
7	G	G	T	A	\$	A	C	C	G	T	A	\$	A	C	C	G	T
13	G	G	T	A	\$	A	C	C	G	T	A	\$	A	C	C	G	T
8	G	G	T	A	\$	A	C	C	G	T	A	\$	A	C	C	G	T
2	G	G	T	A	\$	A	C	C	G	T	A	\$	A	C	C	G	T
14	T	A	\$	A	C	C	G	T	A	\$	A	C	C	G	T	A	\$
9	T	A	\$	A	C	C	G	T	A	\$	A	C	C	G	T	A	\$
4	T	A	\$	A	C	C	G	T	A	\$	A	C	C	G	T	A	\$
3	T	A	\$	A	C	C	G	T	A	\$	A	C	C	G	T	A	\$