

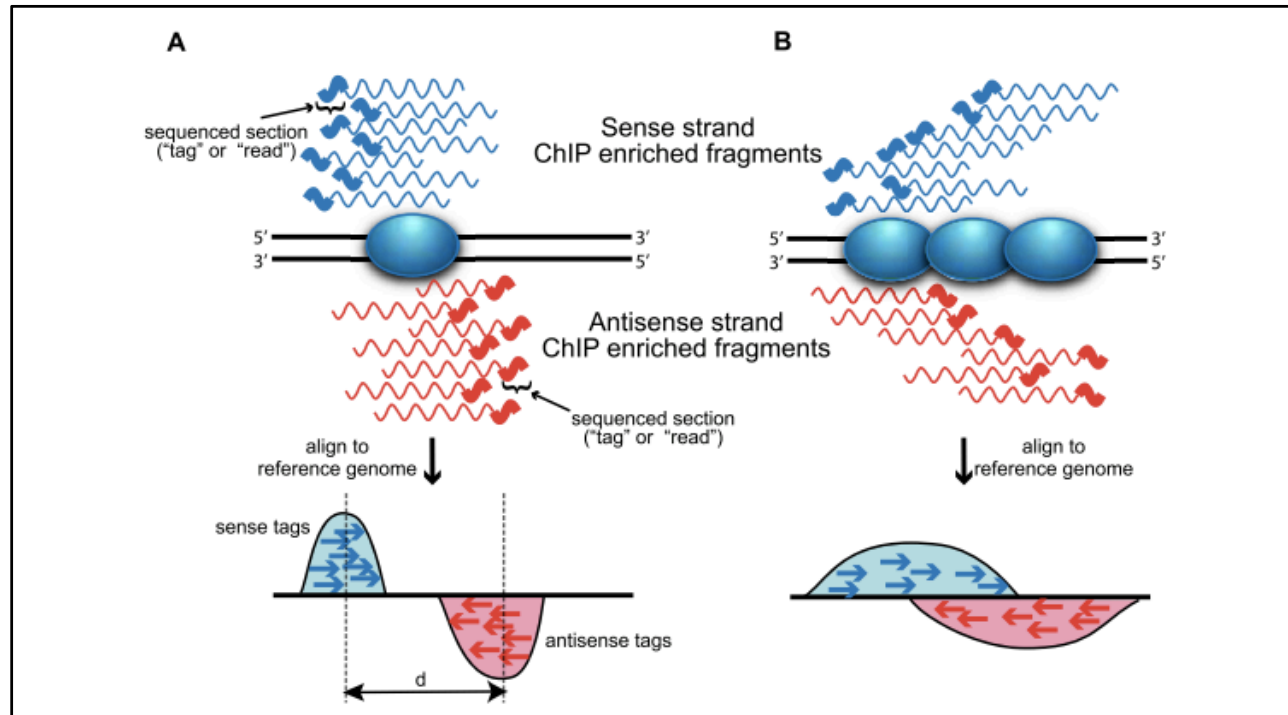
---

# **bPeaks: a R package to perform ChIP-seq peak calling**

How to detect transcription factor binding sites from  
ChIP-seq data in small eukaryotic genomes ?

# What is “peak calling” ?

↳ Location of DNA binding sites of proteins (transcription factors, histones, etc.)



*Wilbanks et al., Plos One (2010)*

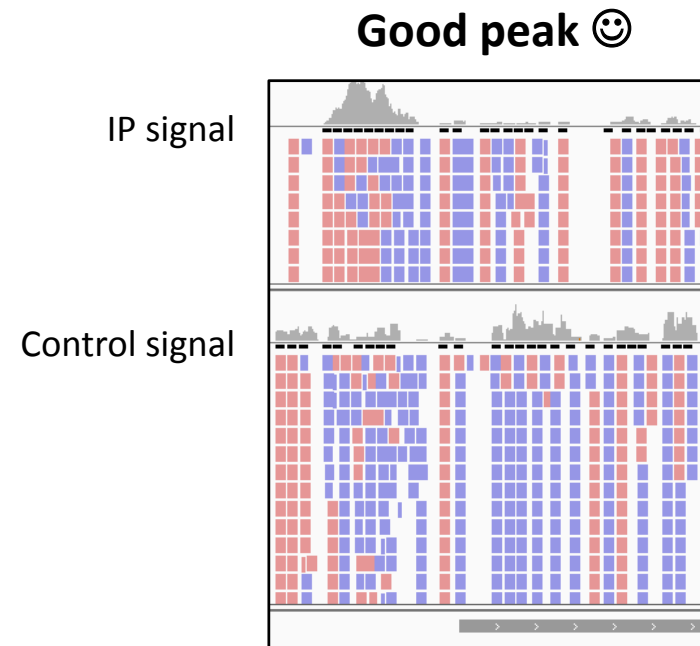
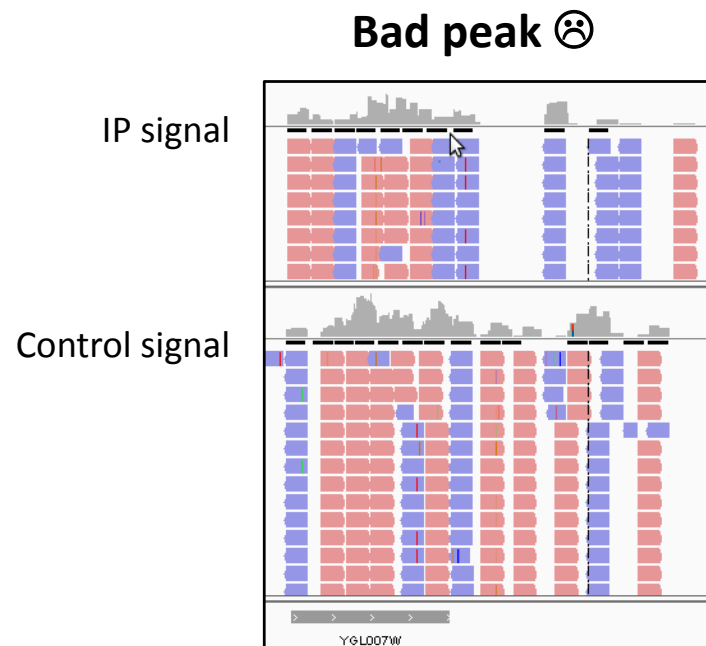
↳ “Peak calling” (ChIP-seq data) = identification of genomic regions with a high density of sequences (reads)

# What is bPeaks?

➔ Simple approach for detection of basic peaks (bPeaks) from ChIP-seq data

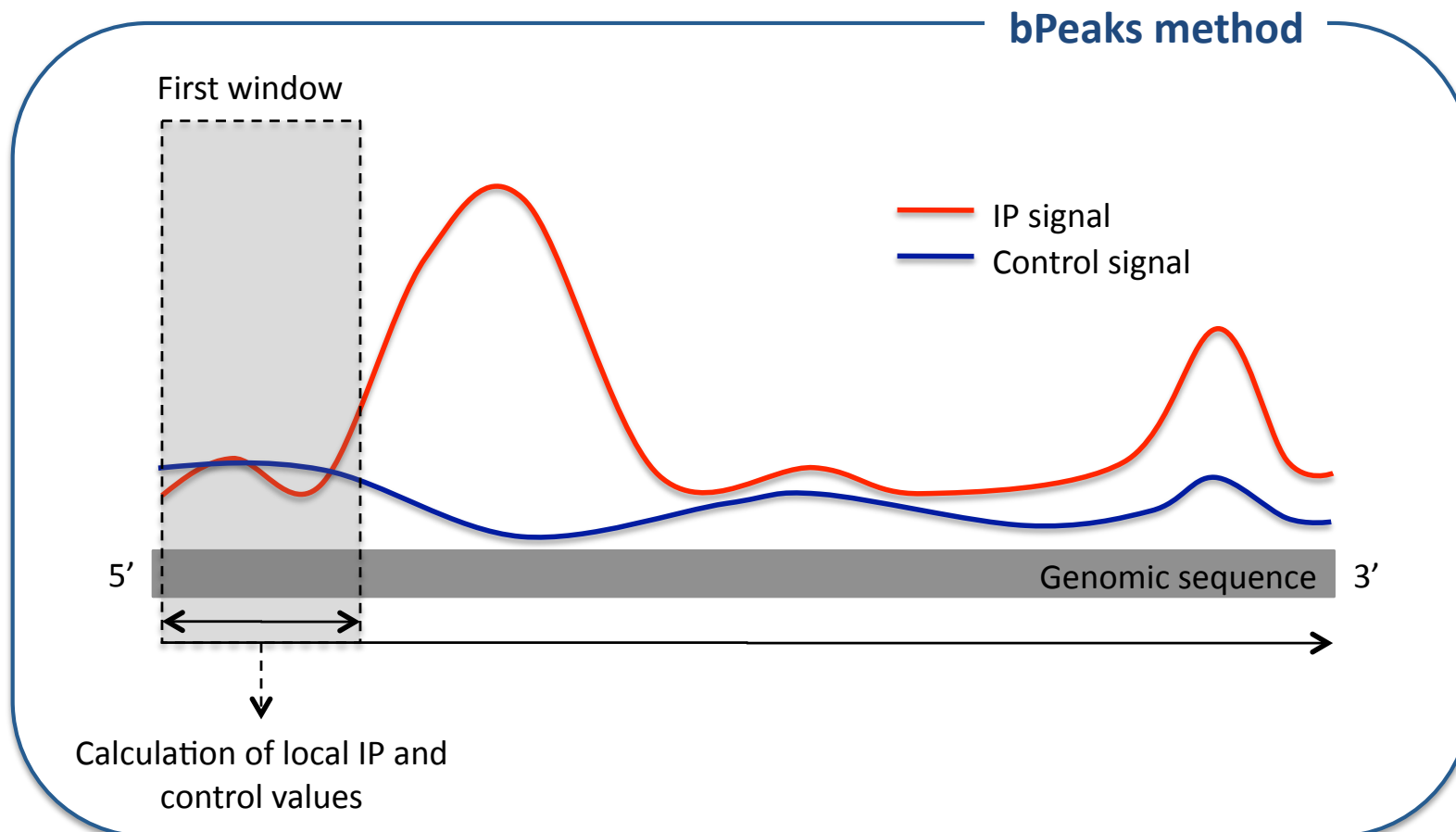
## General philosophy

Easy-to-use tool based on an intuitive definition of peaks by a biologist who visually inspects the ChIP-seq data on a genome browser



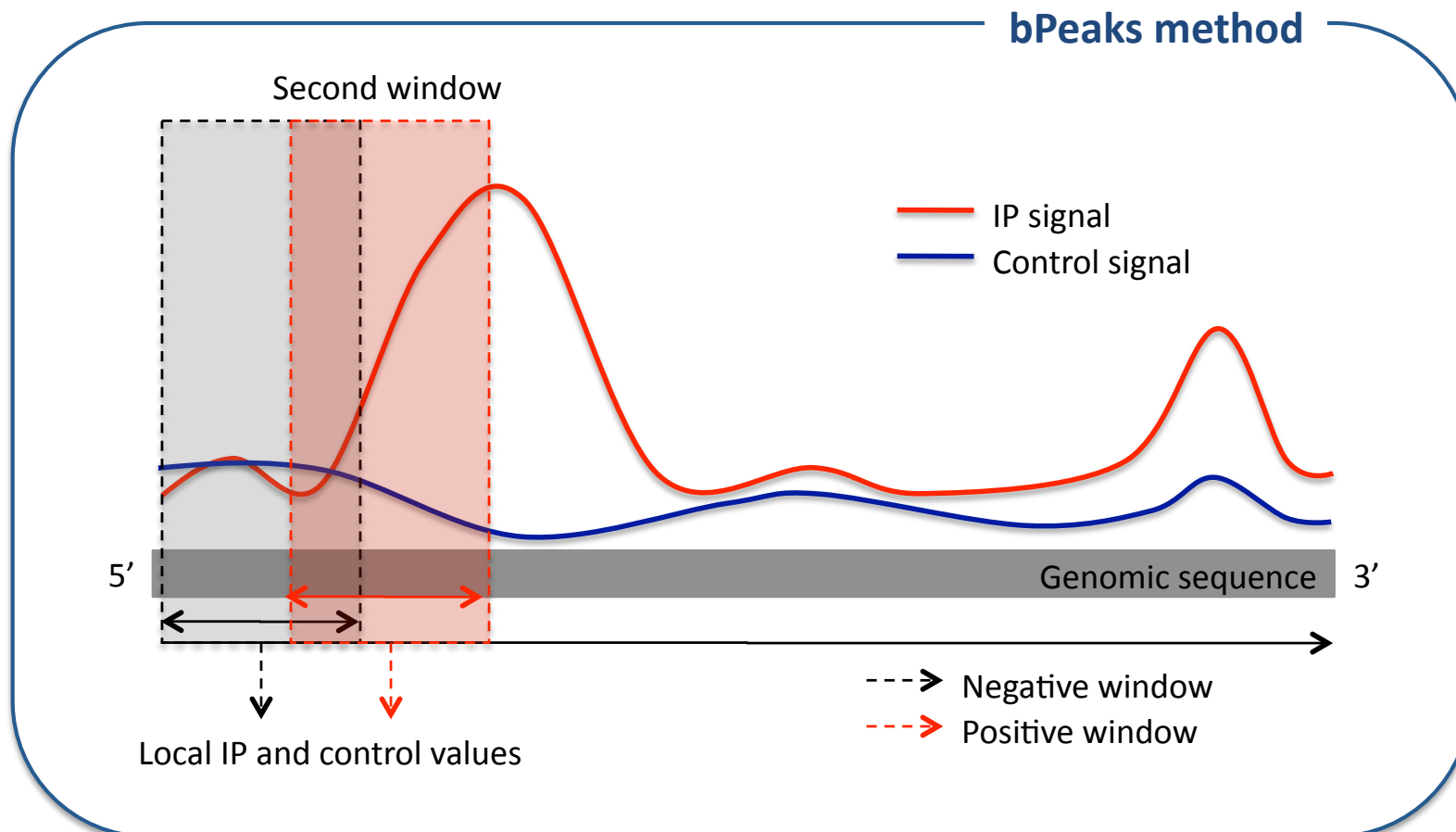
# General principle (1/3)

↳ Sliding window to scan the genomic sequence and compare IP and control signals



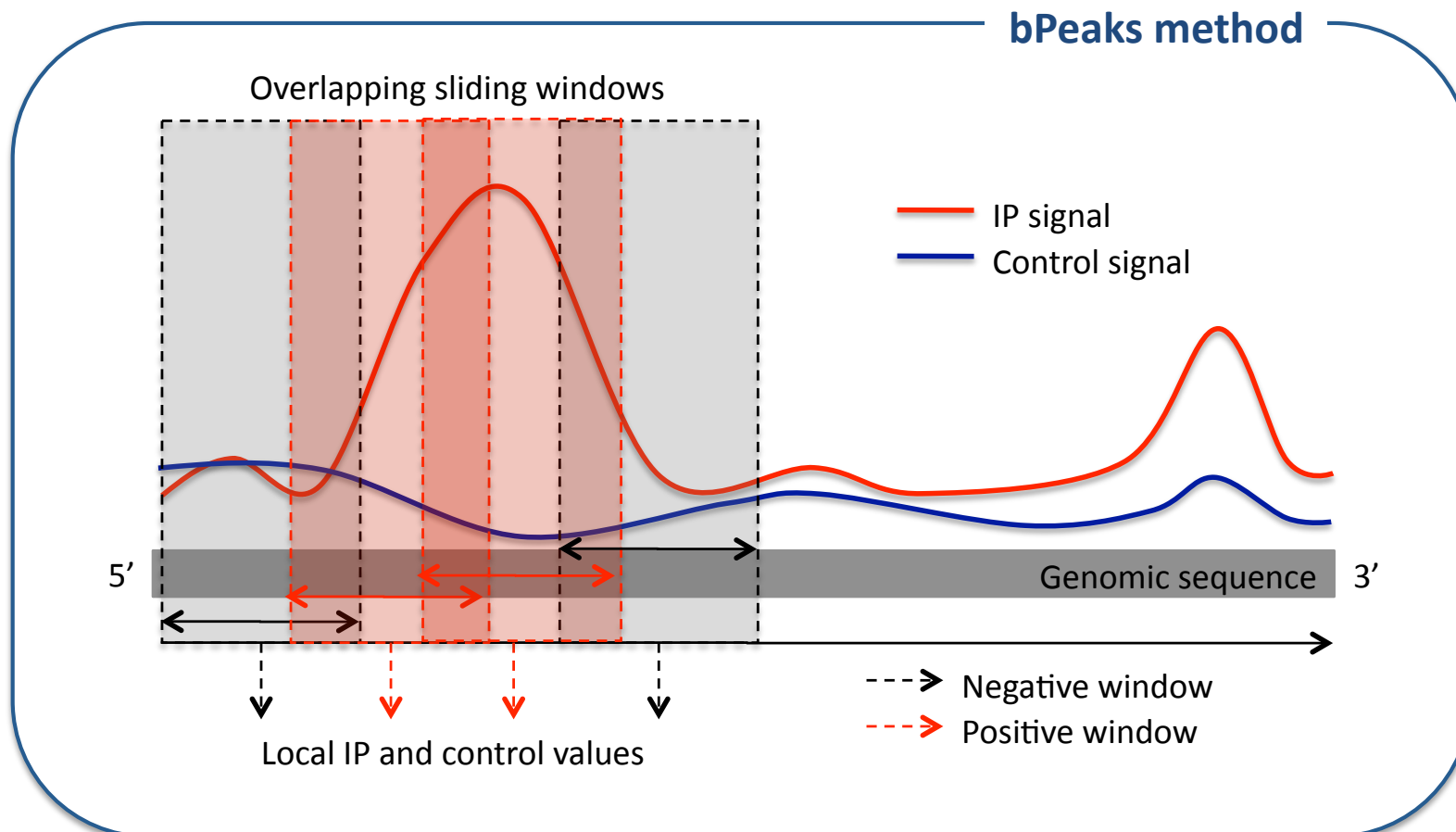
# General principle (1/3)

↳ Sliding window to scan the genomic sequence and compare IP and control signals



# General principle (1/3)

↳ Sliding window to scan the genomic sequence and compare IP and control signals



# General principle (2/3)

➔ Four criterion define interesting genomic regions

## Criteria #1

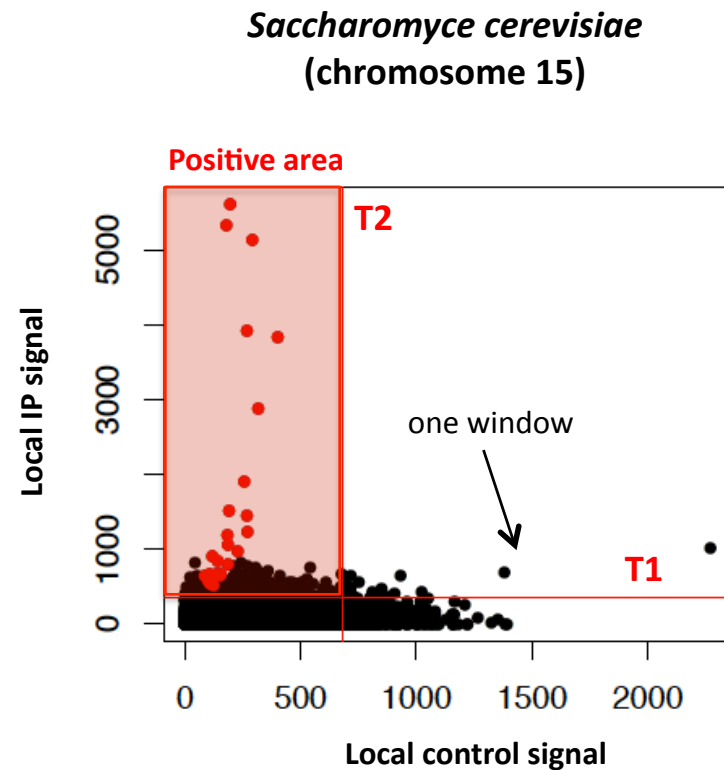
High number of sequences  
(reads) in IP sample

= T1

## Criteria #2

Low number of sequences  
(reads) in control sample

= T2



# General principle (2/3)

➔ Four criterion define interesting genomic regions

## Criteria #1

High number of sequences  
(reads) in IP sample

= T1

## Criteria #2

Low number of sequences  
(reads) in control sample

= T2

## Criteria #3

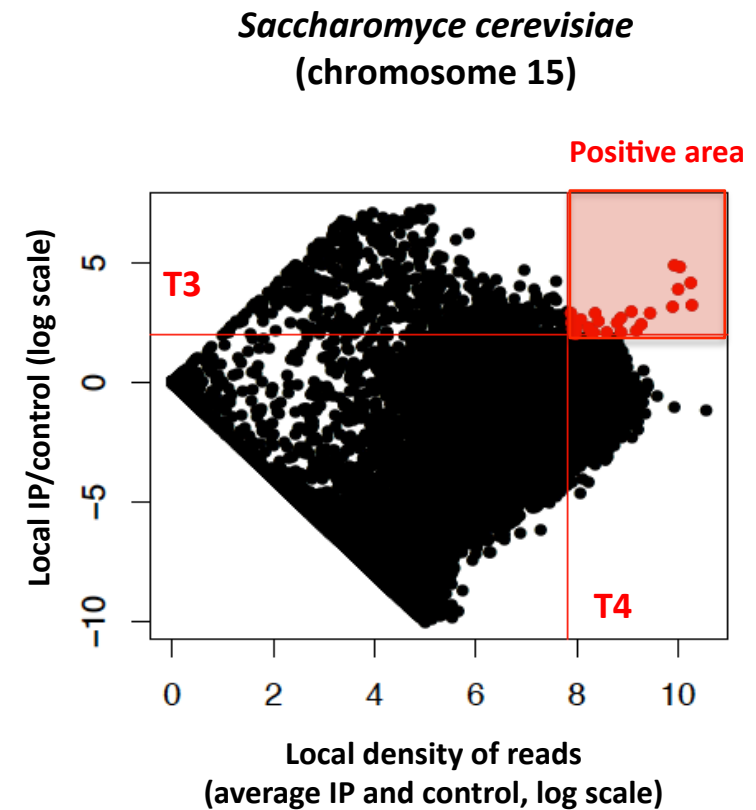
High value of  $\log(\text{IP}/\text{control})$

= T3

## Criteria #4

High density of reads (IP and  
control samples)

= T4





# General principle (2/3)

➔ Four criterion define interesting genomic regions

## Criteria #1

High number of sequences  
(reads) in IP sample

## Criteria #2

Low number of sequences  
(reads) in control sample

## Criteria #3

High value of  $\log(\text{IP}/\text{control})$

## Criteria #4

High density of reads (IP and  
control samples)

T1

+

T2

+

T3

+

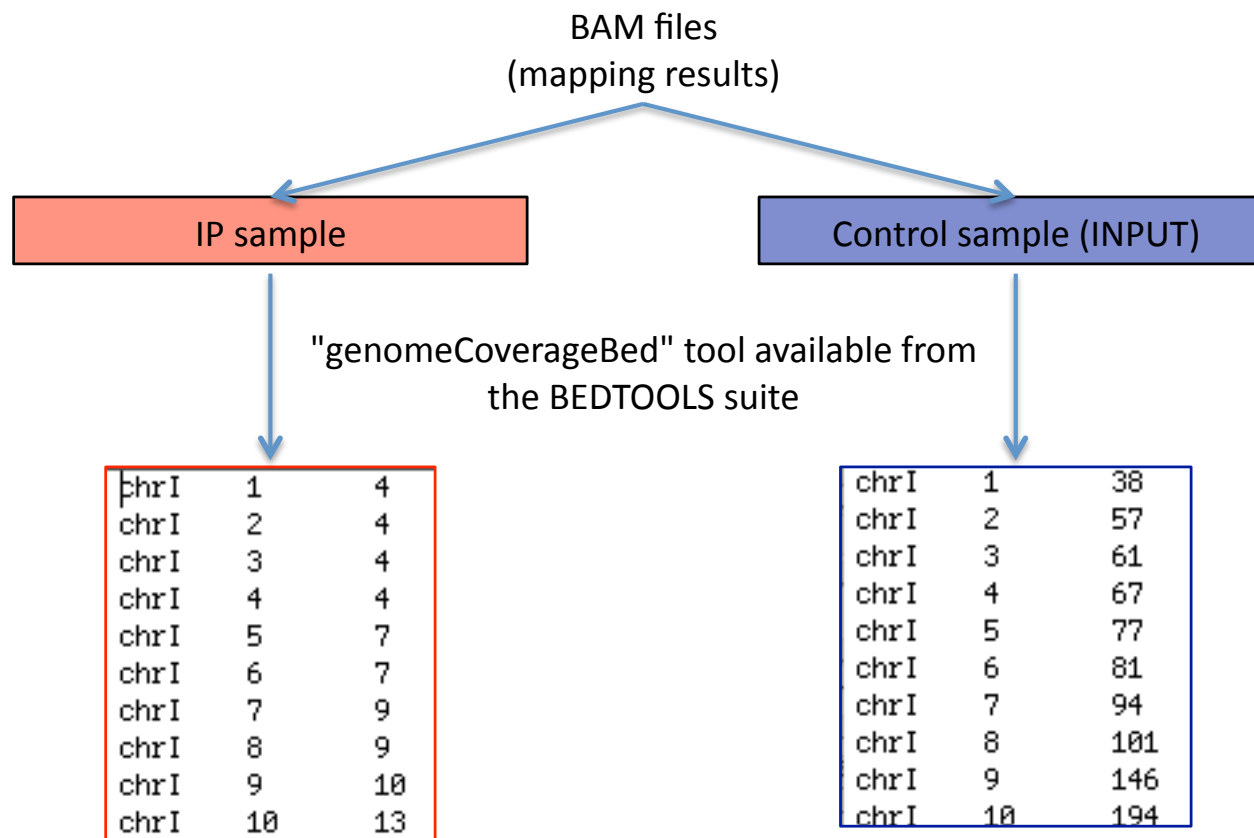
T4

## Interesting region

Successive positive windows  
=  
Basic peaks (bPeaks)

# bPeaks input datasets

- ➔ Sequencing results should be datafiles with numbers of sequences mapped on each nucleotide in the reference genome



# Getting started with the R package

## Package 'bPeaks'

July 30, 2013

**Type** Package

**Title** bPeaks: an intuitive peak-calling strategy to detect transcription factor binding sites from CHIP-seq data in small eukaryotic genomes

**Version** 1.2

**Date** 2013-07-30

**Author** Jawad MERHEJ and Gaelle LELANDAIS

**Maintainer** Gaelle LELANDAIS <gaelle.lelandais@univ-paris-diderot.fr>

**Description** bPeaks is a simple approach to identify transcription factor binding sites from CHIP-seq data. Our general philosophy is to provide an easy-to-use tool, well-adapted for small eukaryotic genomes (< 20 Mb). bPeaks uses a combination of 4 cut-offs (T1, T2, T3 and T4) to mimic "good peak" properties as described by biologists who visually inspect the CHIP-seq data on a genome browser. For yeast genomes, bPeaks calculates the proportion of peaks that fall in promoter sequences. These peaks are good candidates as transcription factor binding sites.

**License** GPL

**Depends** R (>= 2.10)

### R topics documented:

bPeaks-package	2
baseLineCalc	4
bPeaksAnalysis	5
dataPDR1	8
dataReading	11
dataSmoothing	13
peakDetection	14
peakDrawing	17
peakLocation	19
yeastCDS	20

**Index** 22

### Examples

```
# get library
library(bPeaks)

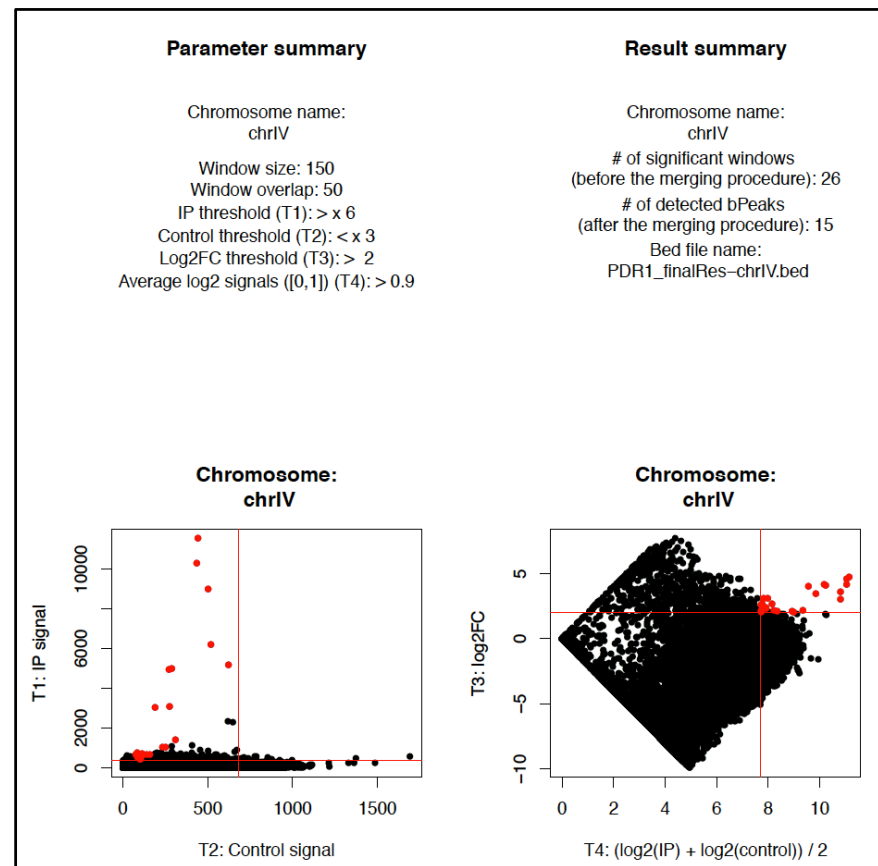
# STEP 1: get PDR1 data
data(dataPDR1)

# STEP 2 : bPeaks analysis
bPeaksAnalysis(IPdata = dataPDR1$IPdata, controlData = dataPDR1$controlData,
               chromosomalFeatures = dataPDR1$chromosomalFeatures,
               smoothingValue = c(20),
               windowSize = c(150), windowOverlap = 50,
               IPcoeff = c(6), controlCoeff = c(4), log2FC = c(2),
               averageQuantiles = c(0.9),
               resultName = "bPeaks_PDR1",
               peakDrawing = TRUE, promSize = 800)
```



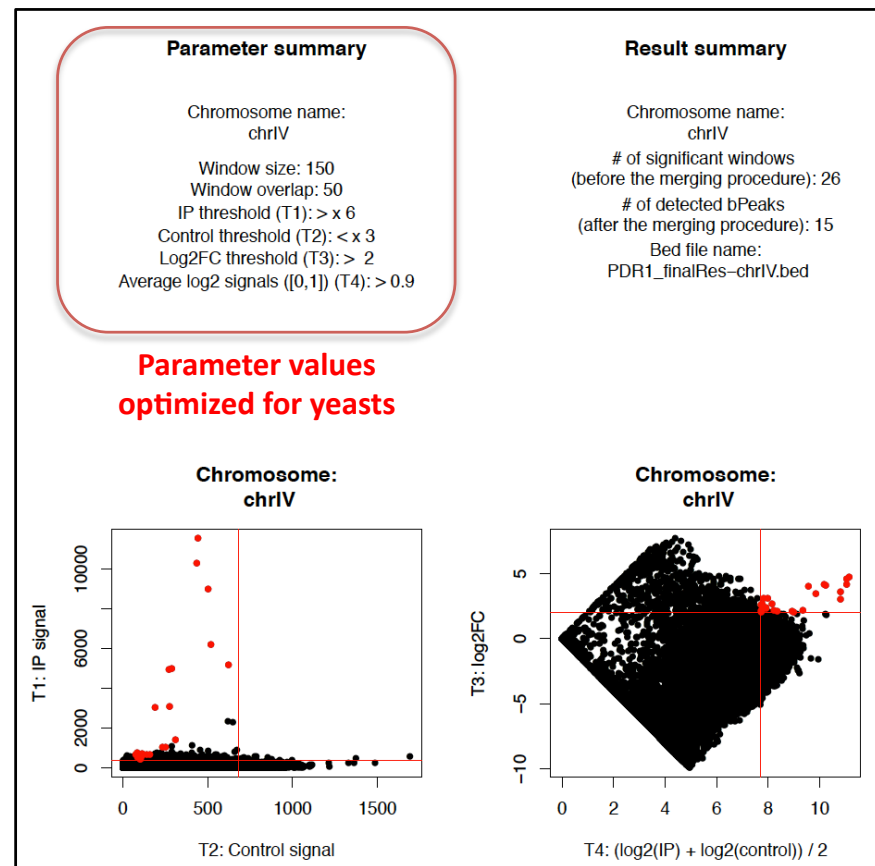
# bPeaks output files (1/4)

➔ PDF file with IP and control signal information for each chromosome



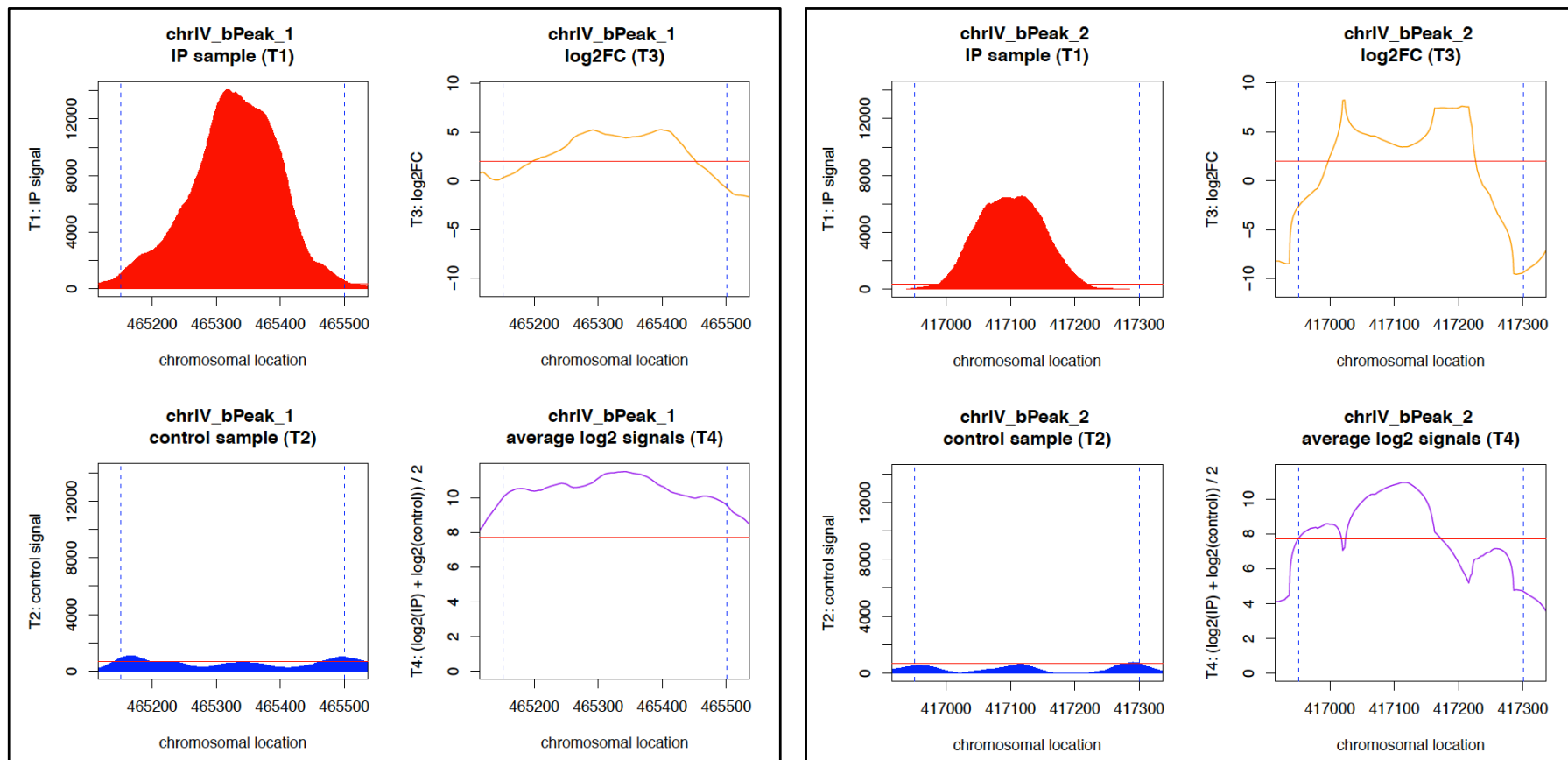
# bPeaks output files (1/4)

➔ PDF file with IP and control signal information for each chromosome



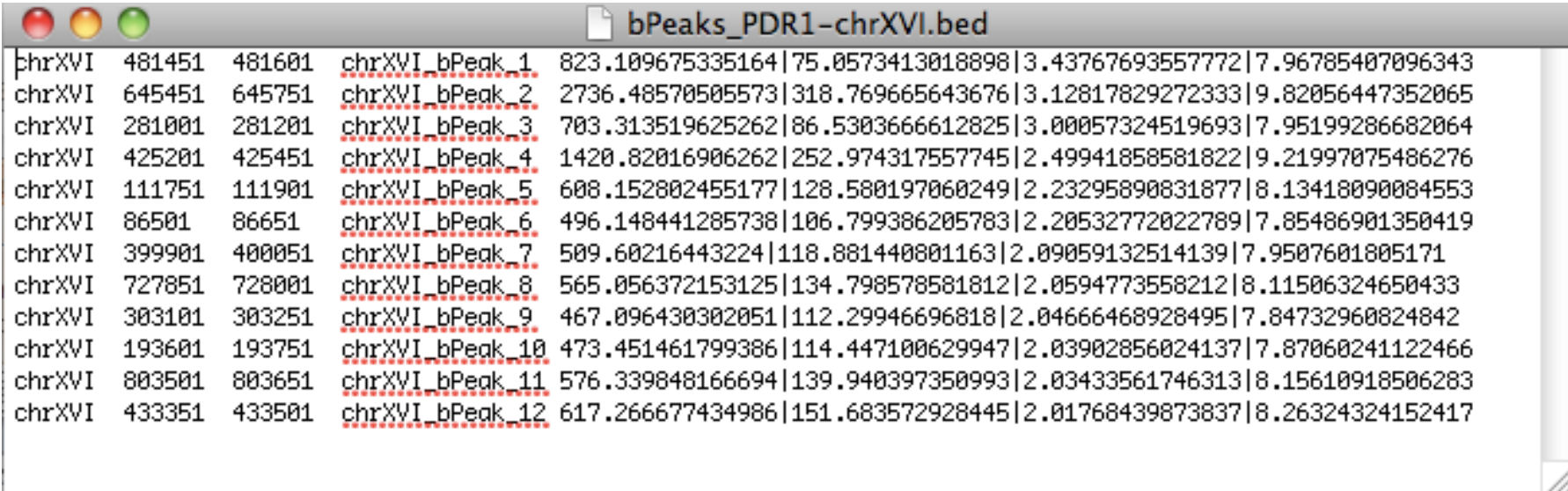
# bPeaks output files (2/4)

➡ PDF files with threshold information for all detected regions



# bPeaks output files (3/4)

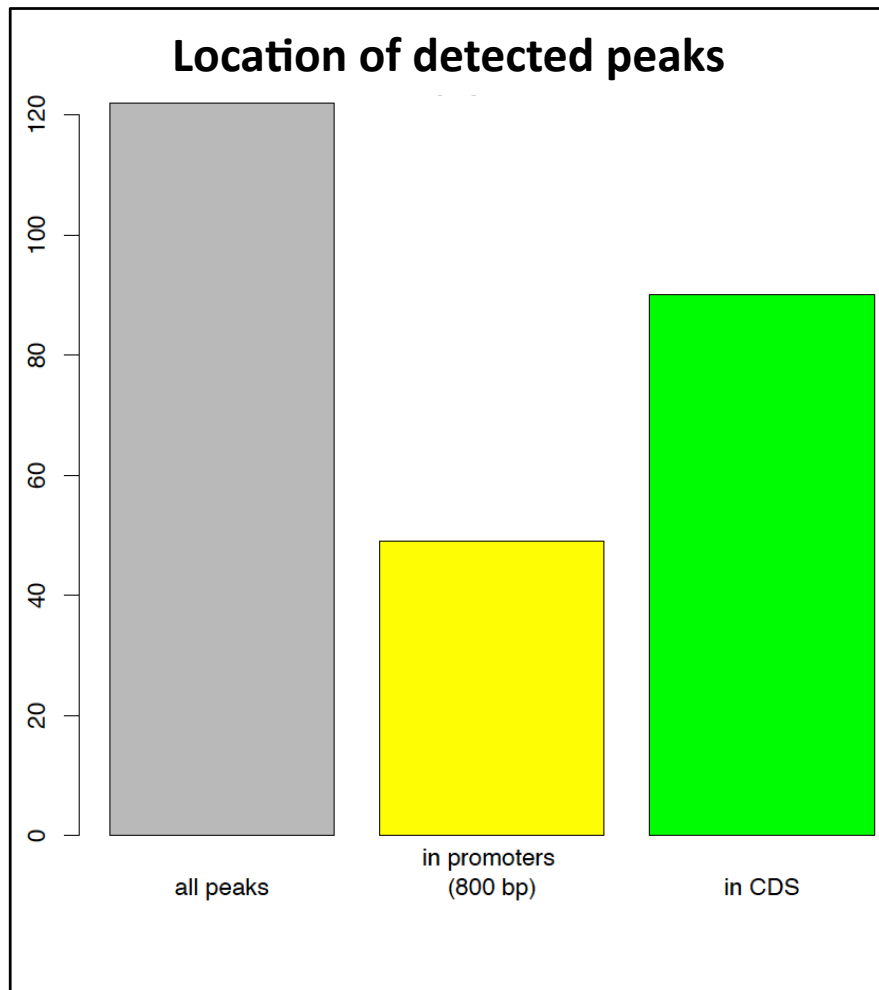
↳ BED files for result visualization with genome browsers



chrXVI	start	end	chrXVI_bPeak_1	823.109675335164	75.0573413018898	3.43767693557772	7.96785407096343
chrXVI	645451	645751	chrXVI_bPeak_2	2736.48570505573	318.769665643676	3.12817829272333	9.82056447352065
chrXVI	281001	281201	chrXVI_bPeak_3	703.313519625262	86.5303666612825	3.00057324519693	7.95199286682064
chrXVI	425201	425451	chrXVI_bPeak_4	1420.82016906262	252.974317557745	2.49941858581822	9.21997075486276
chrXVI	111751	111901	chrXVI_bPeak_5	608.152802455177	128.580197060249	2.23295890831877	8.13418090084553
chrXVI	86501	86651	chrXVI_bPeak_6	496.148441285738	106.799386205783	2.20532772022789	7.85486901350419
chrXVI	399901	400051	chrXVI_bPeak_7	509.60216443224	118.881440801163	2.09059132514139	7.9507601805171
chrXVI	727851	728001	chrXVI_bPeak_8	565.056372153125	134.798578581812	2.0594773558212	8.11506324650433
chrXVI	303101	303251	chrXVI_bPeak_9	467.096430302051	112.29946696818	2.04666468928495	7.84732960824842
chrXVI	193601	193751	chrXVI_bPeak_10	473.451461799386	114.447100629947	2.03902856024137	7.87060241122466
chrXVI	803501	803651	chrXVI_bPeak_11	576.339848166694	139.940397350993	2.03433561746313	8.15610918506283
chrXVI	433351	433501	chrXVI_bPeak_12	617.266677434986	151.683572928445	2.01768439873837	8.26324324152417



# bPeaks output files (4/4)



- For yeast genomes, bPeaks calculates the proportion of peaks falling in promoter regions.
- These peaks are good candidates as potential transcription factor binding sites.
- Annotations of genes positions for ten different yeasts species are available:
  - *Saccharomyces cerevisiae*,
  - *Candida albicans*,
  - *Candida glabrata*,
  - *Debaryomyces hansenii*,
  - *Eremothecium gossypii*,
  - *Kluyveromyces lactis*,
  - *Pichia sorbitophila*,
  - *Saccharomyces kluyveri*,
  - *Yarrowia lipolytica*
  - *Zygosaccharomyces rouxii*.



# bPeaks summary

## R package

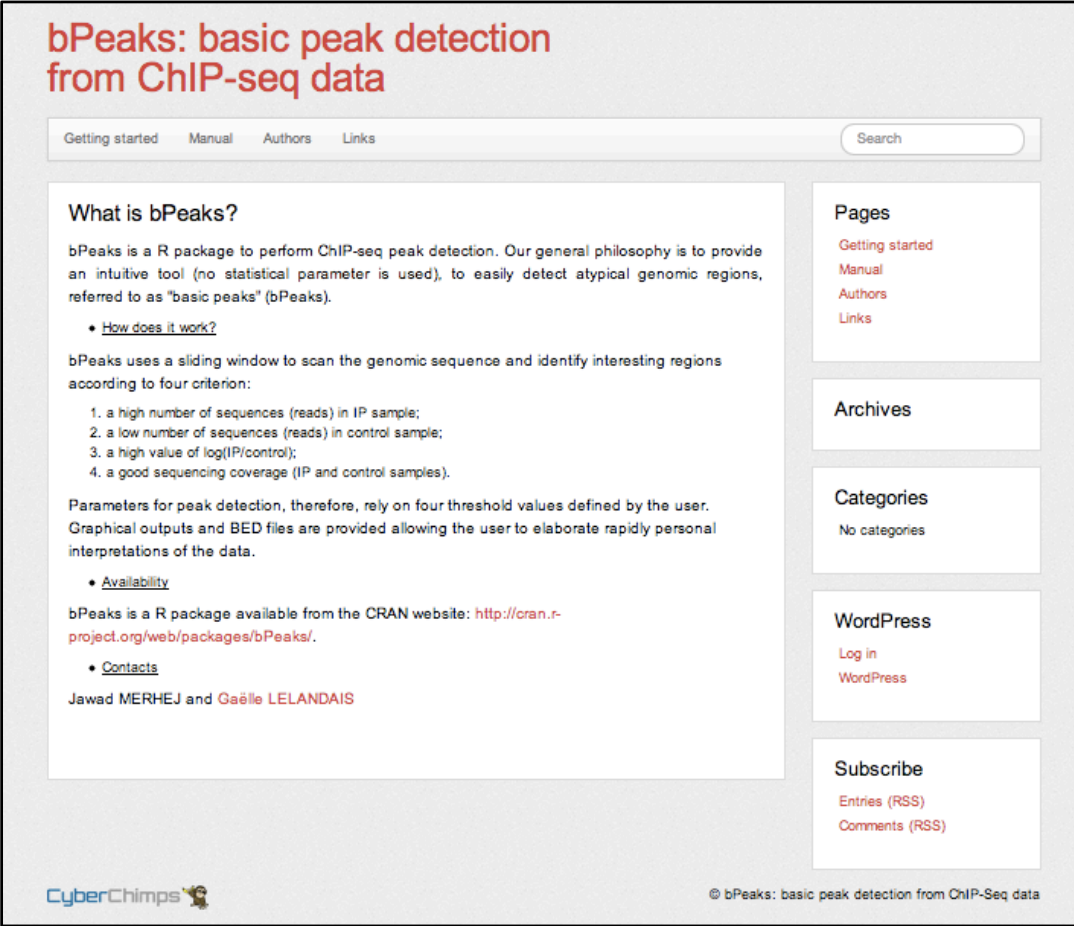
- Easy-to-use tool based on an intuitive definition of peaks by a biologist who visually inspects the ChIP-seq data on a genome browser
- Graphical outputs and BED files are provided to rapidly assess the relevance of the chosen parameters (T1, T2, T3 and T4)

## Method performances

- Average size of peaks detected with bPeaks is remarkably small
- Whereas bPeaks does not use elaborate statistical model, the detected peaks are relevant according to others biological information
- Visual inspection of the peaks and motif discovery show that bPeaks are well focused and centered around the protein DNA binding sites

# More information?

- ➔ Supplementary data together with detailed tutorials (R programming and bPeaks use) are available online: <http://bpeaks.gene-networks.net>



The screenshot shows the homepage of the bPeaks website. The main heading is "bPeaks: basic peak detection from CHIP-seq data" in red. Below the heading is a navigation bar with links for "Getting started", "Manual", "Authors", and "Links", and a search box. The main content area is titled "What is bPeaks?" and contains a paragraph describing the tool, a list of four criteria for peak detection, and information about availability and contacts. The right sidebar contains sections for "Pages" (with links to Getting started, Manual, Authors, and Links), "Archives", "Categories" (No categories), "WordPress" (with links to Log in and WordPress), and "Subscribe" (with links to Entries (RSS) and Comments (RSS)). The footer includes the CyberChimps logo and the copyright notice "© bPeaks: basic peak detection from CHIP-Seq data".

**bPeaks: basic peak detection from CHIP-seq data**

Getting started Manual Authors Links

### What is bPeaks?

bPeaks is a R package to perform CHIP-seq peak detection. Our general philosophy is to provide an intuitive tool (no statistical parameter is used), to easily detect atypical genomic regions, referred to as "basic peaks" (bPeaks).

- [How does it work?](#)

bPeaks uses a sliding window to scan the genomic sequence and identify interesting regions according to four criterion:

1. a high number of sequences (reads) in IP sample;
2. a low number of sequences (reads) in control sample;
3. a high value of  $\log(\text{IP}/\text{control})$ ;
4. a good sequencing coverage (IP and control samples).

Parameters for peak detection, therefore, rely on four threshold values defined by the user. Graphical outputs and BED files are provided allowing the user to elaborate rapidly personal interpretations of the data.

- [Availability](#)

bPeaks is a R package available from the CRAN website: <http://cran.r-project.org/web/packages/bPeaks/>.

- [Contacts](#)

Jawad MERHEJ and Gaëlle LELANDAIS

### Pages

- [Getting started](#)
- [Manual](#)
- [Authors](#)
- [Links](#)

### Archives

### Categories


No categories

### WordPress

- [Log in](#)
- [WordPress](#)

### Subscribe

- [Entries \(RSS\)](#)
- [Comments \(RSS\)](#)

CyberChimps  © bPeaks: basic peak detection from CHIP-Seq data