The sceUpstr support in package harbChIP

HJB/VJC

November 1, 2026

Contents

1 Introduction **1**

2 Building the sceUpstr object **1**

3 Checking a finding of Harbison et al. **2**

1 Introduction

The intent of this package is to allow code like the following:

```
> library(harbChIP)
> data(sceUpstr)
> sceUpstr

upstreamSeqs instance, organism  sce
There are  6674  entries
first keys:
[1] "YAL001C" "YAL002W" "YAL003W" "YAL004W" "YAL005C"

> getUpstream("YAL001C", sceUpstr)

$YAL001C
500-letter DNAString object:
CTGTACCACTATAATAATTTATCTTGATCGTATTATGTG...AGAGGACGTTTGGTTGAAGCCAACTAGCCACAAGAAAA
```

2 Building the sceUpstr object

Upstream sequences of length 500bp were obtained from the SGD website:
`www.yeastgenome.org -> Download Data -> FTP`
`sequence/genomic_sequence/orf_dna/archive/utr5_sc_500.20040206.fasta.gz`

```
> fname = system.file("extdata/utr5_sc_500_20040206.fasta", package="sceUpstr")
> utr5 = readFASTA(fname)
> sceUpstr = buildUpstreamSeqs2(utr5)
> save(sceUpstr, file="sceUpstr.rda")
```

3 Checking a finding of Harbison et al.

It is asserted in Fig 1 B of the paper that GGCGCTA is specifically bound by Snt2. We will examine the frequency of this heptamer in upstream regions and relate to the binding ratio distribution.