

Control peak pipeline for eCLIP-seq

Yue Hu

June 10, 2025

1 Introduction

A pipeline to generate control peaks based on eCLIP-seq peaks. Each control peak would match the transcript region and gene with the original eCLIP-seq peak.

2 Installation

2.1 Quick Install

```
git clone https://github.com/jechia/control_peaks.git
cd control_peaks
```

2.2 Requirements

- Bash 4.0+
- Standard Unix tools (awk, sed, grep)
- bedtools
- python-intervals

3 Basic Usage

3.1 Prepare for the annotation files

```
# Download annotation from GENCODE database (GENCODE_V46)
sh downAnno.sh

# Generate customized annotation based on GENCODE
sh genAnno.sh --gtf annotation/gencode.v46.primary_assembly.annotation
.gtf --rna-list RNA.list
```

3.2 Generate control peaks

```
# Annotate the peaks
bedtools intersect -a <sample>_peaks.bed -b annotation/
    gencode_v46_transcripts.bed -f 1 -wa -wb -s > anno/<sample>.bed

# Generate control peaks
python eCLIP_control_v4.py -i anno/<sample>.bed -a annotation/
    gencode_v46_anno.bed -g annotation/genes.bed -p 10

# Get fasta for the peak and control peaks
sh getFasta.sh <sample>
```