

Figure S3C 재현

홍무범

1. 데이터 전처리

- **pileup**: CLIP-35L33G.bam → CLIP-35L33G.pileup
- **filtering**
 - read depth > 50
 - autosome only (chr1 ~ 19)
 - remove sequencing tags and INDELs
 - split (+) & (-) stranded basereads

Figure S3C 재현

홍무범

2. 분석: 높은 오류율 위치 찾기

- 각 포지션 별 baseread의 Shannon's Entropy를 계산
(strand 별로 각자 계산)
- binding site: entropy > 0.8 & read depth > 50
→ (+): 17,144 & (-): 16,022개의 위치 발견
- UCSC reference genome (mm39)에서 해당 위치 및 hexamer 추출

Figure S3C 재현

홍무민

3. 결과

C LIN28A binding motif and position-specific probabilities detected in three biological replicates

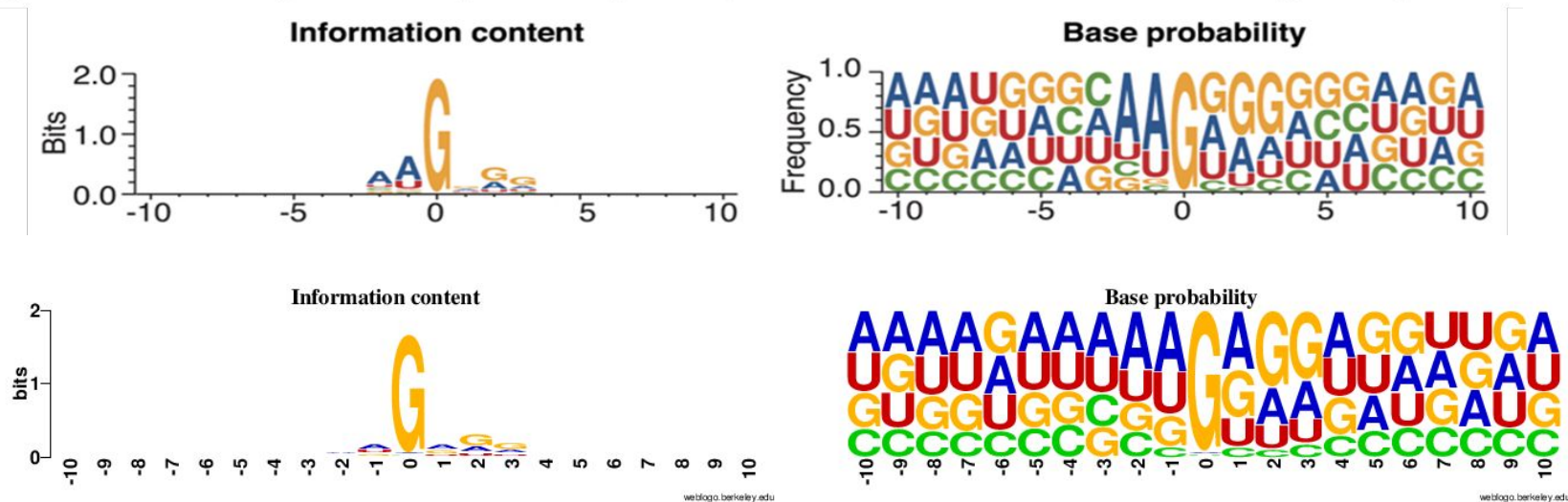


Figure S3C 재현

홍무범

4. 논의점

- Lin28a binding site의 consensus sequence가 “AAGNGG”임을 검증
- filtering threshold 조정 및 base loss를 최소화해서 재현성 개선
- 추가분석으로 hexamer 및 flanking sequence의 빈도도 같이 계산 가능
(Figure S3B & S3D)