

# TF Kandidaten für ein Gen finden und verifizieren

## Daten

- Multiple Sequence Alignment, Maus als Referenz, Spalax unter den Queries
  - Speicherort Samrock, /moreAddSpace/abrzoska/mm10/multi.anno.maf
  - ! Bei allen Annotationen verifizieren, dass sie auf der selben Assembly (hg38/mm10) beruhen.
- Annotation (cCRE/Encode, CNEs, etc. in .bed-Format)

### 1. Annotation filtern auf upstream-Region des zu untersuchenden Gens

```
reg_ex = rex.RegionExtractor()  
reg_ex.find_bed_candidates(bfile, start, end, scaffold,  
output_dir)
```

bfile: Annotation in .bed format

start/end/scaffold: Genomausschnitt, e.g. *start=86700000, end=86750000, scaffold='chr17'*

Output: .bed mit Regionen, die Annotationen in der angegebenen Region repräsentieren

### 2. TFmotifView Kandidaten generieren

<http://bardet.u-strasbg.fr/tfmotifview/>

Assembly selektieren, Genomregionen aus 1. hochladen/einfügen, starten.

Ergebnisse kommen per Mail/Link

### 3. Erstellen der ENCODE TF Liste

<https://screen.wenglab.org/>

Gen suchen, cCRE eins nach dem anderen auswählen, unter TF and His-mod

Intersection die TF rauskopieren und in eine Liste einfügen (Jedes auf neuer Zeile,

Duplikate sind kein Problem)

factor	# of experiments that support TF binding	# experiments in total
POLR2A	14	16
CTCF	5	31
HCFC1	3	3
MAZ	2	2
TBP	2	2

### 4. Vergleich