

## ML in DMEL

### Classify REs in DMEL (HMM)

EpiRegio

REGGAE

Create database for all REs identified in DMEL

follow one of the two routes

Split REs based on type and make them type agnostic

Retain RE type

Train algorithm on all REs in DMEL and predict in DMEL

Train another algorithm to predict RE type

Train algorithm on all REs in DMEL and predict in DMEL (with classification in same step)

Identify (HMM-Classified  $\cup$  Predicted)' and classify into TP and FP

## Encapsulating TP and FP

### Identify Distribution of FP REs

Run UCLUST to cluster TP REs to identify "known" types

\* "known" types based on EpiRegio and REGGAE predictions

Run UCLUST to cluster FP REs to identify unknown types

Classify FP clusters based on TP cluster annotations

Based on cluster annotation and genome data, check if each FP RE is a putative true RE

FALSE

TRUE

Omit and skip to next FP RE

Add to database of known RE

