

**A**

Extract transcription factor protein-coding genes and MinHash the peptide sequences to get a signature

**B**

kNN graph of  
Brawand2011  
TF MinHashes

**C**

kNN graph of  
Brawand2011  
TF expression

**D**

K-mers driving  
similarity in  
brawand2011

**E**

Are the k-mers from  
unmapped reads or  
unannotated genes?