WP3 28.01.22

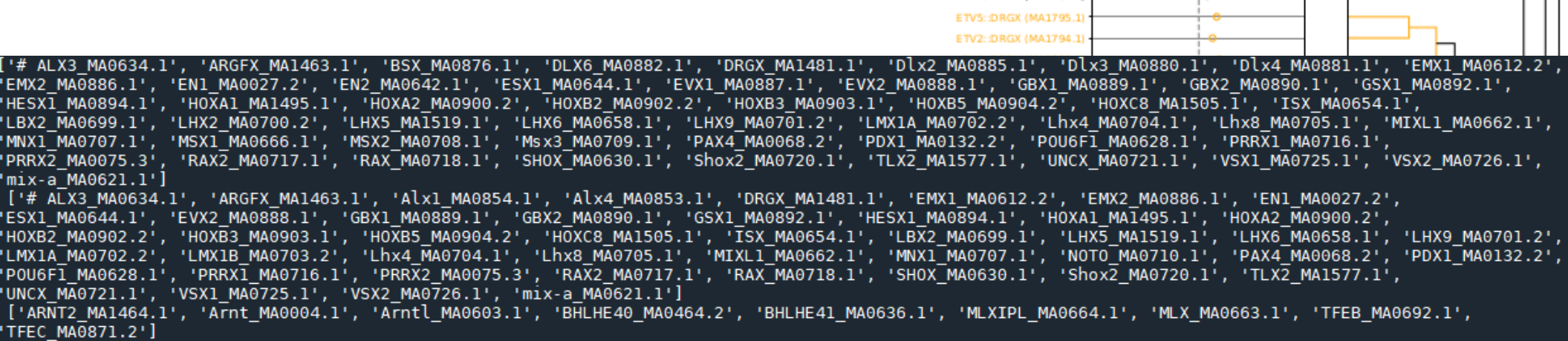# Tasks done since last week

- Sucessfully ran the snakemake pipeline
  - Takes about 3 days
  - Clustering creates a file for easy snakemake import
- Started to attempt Transcription factor grouping by family
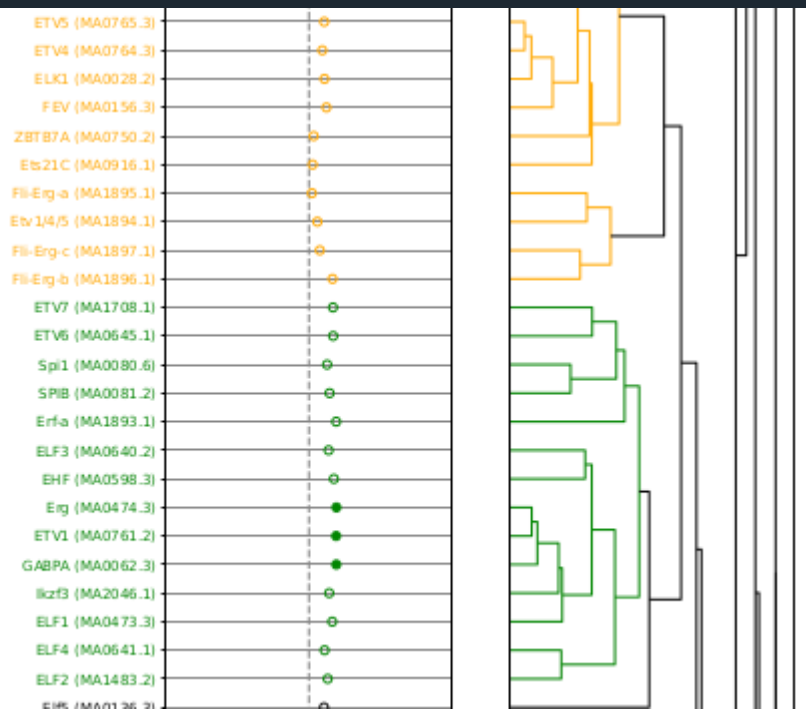- Got bam files from WP2

Problems

- Snakemake pipeline runs some tools we don't need for our analyses
  - Might be able to reduce runtime through that
- Snakemake pipeline output is different; adaptation of scripts required
- Ran into problems with TF family grouping
- Bam files from WP2 cause peak calling error

ETV5::DRGX (MA1795.1)
ETV2::DRGX (MA1794.1)

```
['# ALX3_MA0634.1', 'ARGFX_MA1463.1', 'BSX_MA0876.1', 'DLX6_MA0882.1', 'DRGX_MA1481.1', 'Dlx2_MA0885.1', 'Dlx3_MA0880.1', 'Dlx4_MA0881.1', 'EMX1_MA0612.2',
'EMX2_MA0886.1', 'EN1_MA0027.2', 'EN2_MA0642.1', 'ESX1_MA0644.1', 'EVX1_MA0887.1', 'EVX2_MA0888.1', 'GBX_MA0889.1', 'GBX2_MA0890.1', 'GSX1_MA0892.1',
'HESX1_MA0894.1', 'HOXA1_MA1495.1', 'HOXA2_MA0900.2', 'HOXB2_MA0902.2', 'HOXB3_MA0903.1', 'HOXB5_MA0904.2', 'HOXC8_MA1505.1', 'ISX_MA0654.1',
'LBX2_MA0699.1', 'LHX2_MA0700.2', 'LHX5_MA1519.1', 'LHX6_MA0658.1', 'LHX9_MA0701.2', 'LMX1A_MA0702.2', 'Lhx4_MA0704.1', 'Lhx8_MA0705.1', 'MIXL1_MA0662.1',
'MNX1_MA0707.1', 'MSX1_MA0666.1', 'MSX2_MA0708.1', 'Msx3_MA0709.1', 'PAX4_MA0068.2', 'PDX1_MA0132.2', 'POU6F1_MA0628.1', 'PRRX1_MA0716.1',
'PRRX2_MA0075.3', 'RAX2_MA0717.1', 'RAX_MA0718.1', 'SHOX_MA0630.1', 'Shox2_MA0720.1', 'TLX2_MA1577.1', 'UNCX_MA0721.1', 'VSX1_MA0725.1', 'VSX2_MA0726.1',
'mix-a_MA0621.1']
['# ALX3_MA0634.1', 'ARGFX_MA1463.1', 'Alx1_MA0854.1', 'Alx4_MA0853.1', 'DRGX_MA1481.1', 'EMX1_MA0612.2', 'EMX2_MA0886.1', 'EN1_MA0027.2',
'ESX1_MA0644.1', 'EVX2_MA0888.1', 'GBX1_MA0889.1', 'GBX2_MA0890.1', 'GSX1_MA0892.1', 'HESX1_MA0894.1', 'HOXA1_MA1495.1', 'HOXA2_MA0900.2',
'HOXB2_MA0902.2', 'HOXB3_MA0903.1', 'HOXB5_MA0904.2', 'HOXC8_MA1505.1', 'ISX_MA0654.1', 'LBX2_MA0699.1', 'LHX5_MA1519.1', 'LHX6_MA0658.1', 'LHX9_MA0701.2',
'LMX1A_MA0702.2', 'LMX1B_MA0703.2', 'Lhx4_MA0704.1', 'Lhx8_MA0705.1', 'MIXL1_MA0662.1', 'MNX1_MA0707.1', 'NOTO_MA0710.1', 'PAX4_MA0068.2', 'PDX1_MA0132.2',
'POU6F1_MA0628.1', 'PRRX1_MA0716.1', 'PRRX2_MA0075.3', 'RAX2_MA0717.1', 'RAX_MA0718.1', 'SHOX_MA0630.1', 'Shox2_MA0720.1', 'TLX2_MA1577.1',
'UNCX_MA0721.1', 'VSX1_MA0725.1', 'VSX2_MA0726.1', 'mix-a_MA0621.1']
['ARNT2_MA1464.1', 'Arnt_MA0004.1', 'Arntl_MA0603.1', 'BHLHE40_MA0464.2', 'BHLHE41_MA0636.1', 'MLXIPL_MA0664.1', 'MLX_MA0663.1', 'TFEB_MA0692.1',
'TFEC_MA0871.2']
```

Some TFs have overlap with multiple different families

Fix this issues by working with grouping to shortest Distance?

ETV5 (MA0765.3)
ETV4 (MA0764.3)
ELK1 (MA0028.2)
FEV (MA0156.3)
ZBTB7A (MA0750.2)
Ets21C (MA0916.1)
Fli-Erg-a (MA1895.1)
Etv1/4/5 (MA1894.1)
Fli-Erg-c (MA1897.1)
Fli-Erg-b (MA1896.1)
ETV7 (MA1708.1)
ETV6 (MA0645.1)
Spi1 (MA0080.6)
SPIB (MA0081.2)
Erf-a (MA1893.1)
ELF3 (MA0640.2)
EHF (MA0598.3)
Erg (MA0474.3)
ETV1 (MA0761.2)
GABPA (MA0062.3)
Ikzf3 (MA2046.1)
ELF1 (MA0473.3)
ELF4 (MA0641.1)
ELF2 (MA1483.2)
ELF5 (MA0136.3)

ToDos

- Fix the errors
- Finish TF family grouping
- Run pipeline with Data from WP2