

C.1 Eponine transcription termination parameters

The parameters used to create Eponine transcription termination model –

```
<? xml version="1.0" ?>

<app xmlns=http://www.sanger.ac.uk/Users/td2/specs/epoapps/0/2
jclass="eponine.TrainingCore">

    <bean name="dataSource" jclass="eponine.datasources.XMLDataSource">
        <string name="fileName" value="Datasets/trainingdata.xml" />
    </bean>

    <bean name="basisSource" jclass="eponine.model.MultiplexedBasisSource">
        <int name="reweightFrequency" value="15" />
        <double name="reweightPseudocounts" value="10.0" />

        <child jclass="eponine.model.NewBasisSource">
            <boolean name="maximize" value="false" />
            <double name="stringency" value="0.55" />
            <double name="stringencyVariance" value="0.03" />
            <int name="minLength" value="4" />
            <int name="maxLength" value="8" />
            <double name="minDistWidth" value="2.5" />
            <double name="maxDistWidth" value="200.0" />
            <boolean name="reversible" value="false" />
            <string name="name" value="nbs1_narrow" />
            <int name="minPos" value="-190" />
            <int name="maxPos" value="1990" />
        </child>

        <child jclass="eponine.model.SampleWMBasisSource">
            <double name="nullModelWeighting" value="7.0" />
            <double name="nullModelPerMarginalColumn" value="1.0" />
            <int name="sampleCounts" value="203" />
            <double name="nullModelWeightingN" value="9.0" />
            <int name="sampleCountsN" value="120" />
            <string name="name" value="samplewm2" />
        </child>

        <child jclass="eponine.model.DropColumnBasisSource" />

        <child jclass="eponine.model.DistributionBasisSource">
            <double name="distChangeWidth" value="3.0" />
            <double name="distChangeGamma" value="3.0" />
            <double name="distChangeScale" value="25.0" />
            <double name="distChangeBias" value="0.06" />
            <!-- double name="shapeChangeProbability" value="0.05" / -->
            <double name="flipEnvelopeProbability" value="0.00" />
        </child>
    </bean>
</app>
```

```

        <string name="name" value="distwidth" />
    </child>

    <child jclass="eponine.model.PositionBasisSource">
        <double name="shiftWidth" value="4" />
    </child>

    <child jclass="eponine.model.CrossWMBasisSource" />

    <child jclass="eponine.model.AppendColumnBasisSource" />

    <child jclass="eponine.model.FlipMaxBasisSource" />
</bean>

<bean name="trainer" jclass="stats.glm.VRVMTrainer">
    <int name="numThreads" value="4" />
    <int name="maxCycles" value="11000" />
    <!--int name="cleaningCycles" value="0" /-->
    <int name="maxWorkingSet" value="28" />
    <int name="minWorkingSet" value="25" />
    <int name="initialWorkingSet" value="50" />
    <double name="initialAlpha" value="1.0" />
    <boolean name="unityHack" value="true" />
    <double name="unityHackThreshold" value="1.0" />
    <boolean name="resetAlphaHack" value="true" />
    <boolean name="insertUnity" value="true" />
</bean>

<bean name="retrainer" jclass="stats.glm.VRVMTrainer">
    <int name="maxCycles" value="100" />
    <!--int name="cleaningCycles" value="0" /-->
    <double name="initialAlpha" value="1.0" />
    <boolean name="unityHack" value="true" />
    <double name="unityHackThreshold" value="1.0" />
</bean>

<string name="fileName" value="Models/terminationmodel.xml" />
<int name="checkpointFrequency" value="500" />
</app>

```

C.2 GAZE gene structure models

The configuration file explaining the gene model with translation features for predicting genes using GenePred –

```

<? xml version="1.0" encoding="US-ASCII" ?>

<gaze>
    <declarations>
        <feature id="tss" st_off="0" en_off="1" />
        <feature id="tis" st_off="0" en_off="3" />
        <feature id="5ss" st_off="1" en_off="1" />
        <feature id="3ss" st_off="1" en_off="1" />
    
```

```

<feature id="tts" st_off="3" en_off="0"/>
<feature id="polyA" st_off="1" en_off="1"/>

<feature id="tss_rev" st_off="1" en_off="0" />
<feature id="tis_rev" st_off="3" en_off="0" />
<feature id="5ss_rev" st_off="1" en_off="1" />
<feature id="3ss_rev" st_off="1" en_off="1" />
<feature id="tts_rev" st_off="0" en_off="3" />
<feature id="polyA_rev" st_off="1" en_off="1"/>

<!--lengthfunction id="intron_pen" />
<lengthfunction id="intergene_pen" />
<lengthfunction id="inital_exon_pen" />
<lengthfunction id="internal_exon_pen" />
<lengthfunction id="terminal_exon_pen" />
<lengthfunction id="single_exon_gene_pen" /-->
</declarations>

<gff2gaze>
  <!-- Features -->
  <gfffeat feature="TSS" strand="+" source="Eponine">
    <feat id="tss"/>
  </gfffeat>

  <gfffeat feature="TSS" strand="-" source="Eponine">
    <feat id="tss_rev"/>
  </gfffeat>

  <gfffeat feature="TIS" strand="+" source="Eponine">
    <feat id="tis"/>
  </gfffeat>

  <gfffeat feature="TIS" strand="-" source="Eponine">
    <feat id="tis_rev"/>
  </gfffeat>

  <gfffeat feature="5SS" strand="+" source="Eponine">
    <feat id="5ss"/>
  </gfffeat>

  <gfffeat feature="5SS" strand="-" source="Eponine">
    <feat id="5ss_rev"/>
  </gfffeat>

  <gfffeat feature="3SS" strand="+" source="Eponine">
    <feat id="3ss"/>
  </gfffeat>

  <gfffeat feature="3SS" strand="-" source="Eponine">
    <feat id="3ss_rev"/>
  </gfffeat>

  <gfffeat feature="TTS" strand="+" source="Eponine">
    <feat id="tts"/>
  </gfffeat>

```

```
<gfffeat feature="TTS" strand="-" source="Eponine">
  <feat id="tts_rev"/>
</gfffeat>

<gfffeat feature="POLYA" strand="+" source="Eponine">
  <feat id="polyA"/>
</gfffeat>

<gfffeat feature="POLYA" strand="-" source="Eponine">
  <feat id="polyA_rev"/>
</gfffeat>
</gff2gaze>

<dna2gaze>
  <!--dnafeat pattern="tataaa">
    <feat id="tss" />
  </dnafeat>

  <dnafeat pattern="atg" score="0.001">
    <feat id="tis" />
  </dnafeat>

  <dnafeat pattern="taa" score="0.001">
    <feat id="tts" />
  </dnafeat>

  <dnafeat pattern="tag" score="0.001">
    <feat id="tts" />
  </dnafeat>

  <dnafeat pattern="tga" score="0.001">
    <feat id="tts" />
  </dnafeat>

  <dnafeat pattern="aataaa" score="0.001">
    <feat id="polyA" />
  </dnafeat>

  <dnafeat pattern="tttata">
    <feat id="tss_rev" />
  </dnafeat>

  <dnafeat pattern="cat" score="0.001">
    <feat id="tis_rev" />
  </dnafeat>

  <dnafeat pattern="tta" score="0.001">
    <feat id="tts_rev" />
  </dnafeat>

  <dnafeat pattern="cta" score="0.001">
    <feat id="tts_rev" />
  </dnafeat>
```

```

<dnafeat pattern="tca" score="0.001">
  <feat id="tts_rev" />
</dnafeat>

<dnafeat pattern="tttatt" score="0.001">
  <feat id="polyA_rev" />
</dnafeat-->

<!--takedna id="5ss_1" st_off="0" en_off="1"/>
<takedna id="3ss_1" st_off="1" en_off="-1"/>
<takedna id="5ss_2" st_off="-1" en_off="1"/>
<takedna id="3ss_2" st_off="1" en_off="0"/>
<takedna id="5ss_1_rev" st_off="1" en_off="0"/>
<takedna id="3ss_1_rev" st_off="-1" en_off="1"/>
<takedna id="5ss_2_rev" st_off="1" en_off="-1"/>
<takedna id="3ss_2_rev" st_off="0" en_off="1"/-->
</dna2gaze>

<model>
  <target id="END">
    <source id="BEGIN" out_feat="No_genes"/>
    <source id="polyA" out_feat="GEN_DNA" />
    <source id="tss_rev" out_feat="GEN_DNA"/>
  </target>

  <!--Forward strand gene-->

  <target id="tss">
    <source id="BEGIN" out_feat="GEN_DNA"/>
    <source id="polyA" mindis="1" out_feat="intergenic"/>
    <source id="tss_rev" mindis="1" out_feat="intergenic"/>
  </target>

  <target id="tis">
    <source id="tss" mindis="1" out_feat="5UTR" out_str="+"/>
  </target>

  <target id="5ss">
    <!--killfeat id="tts" /-->
    <source id="tis" out_feat="inital_exon" mindis="3" maxdis="10000" out_str="+"/>
    <source id="3ss" out_feat="internal_exon" mindis="6" maxdis="10000" out_str="+"/>
  />
</target>

  <target id="3ss">
    <source id="5ss" out_feat="intron" mindis="6" out_str="+"/>
  </target>

  <target id="tts">
    <!--killfeat id="tts" /-->
    <!--source id="tis" out_feat="single_exon_gene" mindis="60" out_str="+"/-->
    <source id="3ss" out_feat="terminal_exon" mindis="3" out_str="+"/>
  </target>

  <target id="polyA">

```

```

        <source id="tts" out_feat="3UTR" mindis="1" out_str="+"/>
    </target>

    <!--Reverse strand gene-->

    <target id="polyA_rev">
        <source id="BEGIN" out_feat="GEN_DNA"/>
        <source id="polyA" out_feat="intergenic" mindis="1"/>
        <source id="tss_rev" out_feat="intergenic" mindis="1"/>
    </target>

    <target id="tts_rev">
        <source id="polyA_rev" out_feat="3UTR" mindis="1" out_str="-"/>
    </target>

    <target id="3ss_rev">
        <!--killfeat id="tts_rev"/-->
        <source id="tts_rev" out_feat="terminal_exon" mindis="3" maxdis="10000"
out_str="-"/>
        <source id="5ss_rev" out_feat="internal_exon" mindis="6" maxdis="10000"
out_str="-"/>
    </target>

    <target id="5ss_rev">
        <source id="3ss_rev" out_feat="intron" mindis="6" out_str="-"/>
    </target>

    <target id="tis_rev">
        <!--killfeat id="tts_rev" phase="0"/-->
        <!--source id="tts_rev" out_feat="single_exon_gene" mindis="60" out_str="-"/-->
        <source id="5ss_rev" out_feat="initial_exon" mindis="3" out_str="-"/>
    </target>

    <target id="tss_rev">
        <source id="tis_rev" out_feat="5UTR" mindis="1" out_str="-"/>
    </target>
</model>

<lengthfunctions>
    <!-- lengthfunc id="intron_pen" file="./tables/intron_penalty"/>
    <lengthfunc id="initial_exon_pen" file="./tables/exon_penalty.initial"/>
    <lengthfunc id="terminal_exon_pen" file="./tables/exon_penalty.terminal"/>
    <lengthfunc id="internal_exon_pen" file="./tables/exon_penalty.internal"/-->

    <!--lengthfunc id="single_exon_gene_pen">
        <point x="500" y="0.001"/>
        <point x="20000" y="0.2"/>
    </lengthfunc-->

    <!--lengthfunc id="intergene_pen">
        <point x="200000" y="0.01"/>
        <point x="200001" y="0.01"/>
    </lengthfunc -->
</lengthfunctions>
</gaze>

```

The configuration file explaining the gene model without translation features for predicting genes using GenePred –

```
<?xml version="1.0" encoding="US-ASCII"?>

  <gaze>
    <declarations>
      <feature id="tss" st_off="0" en_off="1" />
      <!--feature id="tis" st_off="0" en_off="3" /-->
      <feature id="5ss" st_off="1" en_off="1" />
      <feature id="3ss" st_off="1" en_off="1" />
      <!--feature id="tts" st_off="3" en_off="0" /-->
      <feature id="polyA" st_off="1" en_off="1"/>

      <feature id="tss_rev" st_off="1" en_off="0" />
      <!--feature id="tis_rev" st_off="3" en_off="0" /-->
      <feature id="5ss_rev" st_off="1" en_off="1" />
      <feature id="3ss_rev" st_off="1" en_off="1" />
      <!--feature id="tts_rev" st_off="0" en_off="3" /-->
      <feature id="polyA_rev" st_off="1" en_off="1"/>

      <!--lengthfunction id="intron_pen" />
      <lengthfunction id="intergene_pen" />
      <lengthfunction id="inital_exon_pen" />
      <lengthfunction id="internal_exon_pen" />
      <lengthfunction id="terminal_exon_pen" />
      <lengthfunction id="single_exon_gene_pen" /-->
    </declarations>

    <gff2gaze>
      <!-- Features -->
      <gfffeat feature="TSS" strand="+" source="Eponine">
        <feat id="tss" />
      </gfffeat>

      <gfffeat feature="TSS" strand="-" source="Eponine">
        <feat id="tss_rev" />
      </gfffeat>

      <!--gfffeat feature="TIS" strand="+" source="Eponine">
        <feat id="tis" />
      </gfffeat>

      <gfffeat feature="TIS" strand="-" source="Eponine">
        <feat id="tis_rev" /-->
      </gfffeat>

      <gfffeat feature="5SS" strand="+" source="Eponine">
        <feat id="5ss" />
      </gfffeat>

      <gfffeat feature="5SS" strand="-" source="Eponine">
        <feat id="5ss_rev" />
      </gfffeat>
    </gff2gaze>
  </gaze>
```

```

</gfffeat>

<gfffeat feature="3SS" strand="+" source="Eponine">
  <feat id="3ss"/>
</gfffeat>

<gfffeat feature="3SS" strand="-" source="Eponine">
  <feat id="3ss_rev"/>
</gfffeat>

<!--gfffeat feature="TTS" strand="+" source="Eponine">
  <feat id="tts"/>
</gfffeat>

<gfffeat feature="TTS" strand="-" source="Eponine">
  <feat id="tts_rev"/-->
</gfffeat>

<gfffeat feature="POLYA" strand="+" source="Eponine">
  <feat id="polyA"/>
</gfffeat>

<gfffeat feature="POLYA" strand="-" source="Eponine">
  <feat id="polyA_rev"/>
</gfffeat>
</gff2gaze>

<dna2gaze>
  <!--dnafeat pattern="tataaa">
    <feat id="tss" />
  </dnafeat>

  <dnafeat pattern="atg" score="0.001">
    <feat id="tis" />
  </dnafeat>

  <dnafeat pattern="taa" score="0.001">
    <feat id="tts" />
  </dnafeat>

  <dnafeat pattern="tag" score="0.001">
    <feat id="tts" />
  </dnafeat>

  <dnafeat pattern="tga" score="0.001">
    <feat id="tts" />
  </dnafeat>

  <dnafeat pattern="aataaa" score="0.001">
    <feat id="polyA" />
  </dnafeat>

  <dnafeat pattern="tttata">
    <feat id="tss_rev" />
  </dnafeat>

```



```

<dnafeat pattern="cat" score="0.001">
  <feat id="tis_rev" />
</dnafeat>

<dnafeat pattern="tta" score="0.001">
  <feat id="tts_rev" />
</dnafeat>

<dnafeat pattern="cta" score="0.001">
  <feat id="tts_rev" />
</dnafeat>

<dnafeat pattern="tca" score="0.001">
  <feat id="tts_rev" />
</dnafeat>

<dnafeat pattern="tttatt" score="0.001">
  <feat id="polyA_rev" />
</dnafeat-->

<!--takedna id="5ss_1" st_off="0" en_off="1"/>
<takedna id="3ss_1" st_off="1" en_off="-1"/>
<takedna id="5ss_2" st_off="-1" en_off="1"/>
<takedna id="3ss_2" st_off="1" en_off="0"/>
<takedna id="5ss_1_rev" st_off="1" en_off="0"/>
<takedna id="3ss_1_rev" st_off="-1" en_off="1"/>
<takedna id="5ss_2_rev" st_off="1" en_off="-1"/>
<takedna id="3ss_2_rev" st_off="0" en_off="1"/-->
</dna2gaze>

<model>
  <target id="END">
    <source id="BEGIN" out_feat="No_genes"/>
    <source id="polyA" out_feat="GEN_DNA" />
    <source id="tss_rev" out_feat="GEN_DNA"/>
  </target>

  <!--Forward strand gene-->
  <target id="tss">
    <source id="BEGIN" out_feat="GEN_DNA"/>
    <source id="polyA" mindis="1" out_feat="intergenic"/>
    <source id="tss_rev" mindis="1" out_feat="intergenic"/>
  </target>

  <!--target id="tis">
    <source id="tss" mindis="1" out_feat="5UTR" out_str="+"/>
  </target-->

  <target id="5ss">
    <!--killfeat id="tts"/-->
    <!--source id="tis" out_feat="inital_exon" mindis="3" maxdis="10000" out_str="+"
  /-->
    <source id="tss" mindis="1" out_feat="initial_exon" out_str="+"/>

```

```

        <source id="3ss" out_feat="internal_exon" mindis="6" maxdis="10000"
out_str="+" />
    </target>

    <target id="3ss">
        <source id="5ss" out_feat="intron" mindis="6" out_str="+" />
    </target>

    <!--target id="tts"-->
        <!--killfeat id="tts" /-->
        <!--source id="tis" out_feat="single_exon_gene" mindis="60" out_str="+" /-->
        <!--source id="3ss" out_feat="terminal_exon" mindis="3" out_str="+" />
    </target-->

    <target id="polyA">
        <!--source id="tts" out_feat="3UTR" mindis="1" out_str="+" /-->
        <source id="3ss" out_feat="terminal_exon" mindis="3" out_str="+" />
    </target>

    <!--Reverse strand gene-->

    <target id="polyA_rev">
        <source id="BEGIN" out_feat="GEN_DNA" />
        <source id="polyA" out_feat="intergenic" mindis="1" />
        <source id="tss_rev" out_feat="intergenic" mindis="1" />
    </target>

    <!--target id="tts_rev">
        <source id="polyA_rev" out_feat="3UTR" mindis="1" out_str="-" />
    </target-->

    <target id="3ss_rev">
        <!--killfeat id="tts_rev" phase="0" /-->
        <source id="polyA_rev" out_feat="terminal_exon" mindis="3" maxdis="10000"
out_str="-" />
        <source id="5ss_rev" out_feat="internal_exon" mindis="6" maxdis="10000"
out_str="-" />
    </target>

    <target id="5ss_rev">
        <source id="3ss_rev" out_feat="intron" mindis="6" out_str="-" />
    </target>

    <!--target id="tis_rev"-->
        <!--killfeat id="tts_rev" phase="0" /-->
        <!--source id="tts_rev" out_feat="single_exon_gene" mindis="60" out_str="-" /-->
        <!--source id="5ss_rev" out_feat="initial_exon" mindis="3" out_str="-" />
    </target-->

    <target id="tss_rev">
        <source id="5ss_rev" out_feat="initial_exon" mindis="1" out_str="-" />
    </target>
</model>
<lengthfunctions>
    <!-- lengthfunc id="intron_pen" file="./tables/intron_penalty" />

```

```
<lengthfunc id="initial_exon_pen" file="./tables/exon_penalty.initial"/>
<lengthfunc id="terminal_exon_pen" file="./tables/exon_penalty.terminal"/>
<lengthfunc id="internal_exon_pen" file="./tables/exon_penalty.internal"/-->

<!--lengthfunc id="single_exon_gene_pen">
  <point x="500" y="0.001"/>
  <point x="20000" y="0.2"/>
</lengthfunc-->

<!--lengthfunc id="intergene_pen">
  <point x="200000" y="0.01"/>
  <point x="200001" y="0.01"/>
</lengthfunc -->
</lengthfunctions>
</gaze>
```