

Intronic Poly-A&T Workflow Manual

Andy Cao

August 2020

1 Finding Poly-A&T Sequences

Find the location of all poly-A&T sequences within the specified genome by calling the 'PolyA&T_Finder.py' script.

```
python PolyA&T_Finder.py directory length mismatch genome
```

Input Parameters	directory: location of directory containing chromosome FASTA files length: minimum length of poly-A&T sequences mismatch: maximum number of mismatches allowed per sequence genome: genome ID
Output Files	genome_output.txt: poly-A&T sequences in BED format genome_freq.txt: sequence length frequency distribution of poly-A&T sequences genome_summary.txt: summary of output

2 Finding Intronic Poly-A&T Sequences

Find the intersection between all poly-A&T sequences and introns within the specified genome by calling the 'Intron_Detector.py' script.

```
python Intron_Detector.py polyat introns genome
```

Input Parameters	<p>polyat: file location containing poly-A&T sequence information (genome.output.txt)</p> <p>introns: file location containing intron sequence information</p> <p>genome: genome ID</p>
Output Files	<p>genome_within.txt: poly-A&T sequences falling completely within introns in BED format</p> <p>genome_partial_within.txt: poly-A&T sequences falling partially within introns in BED format</p> <p>genome_intron_freq.txt: sequence length frequency distribution of intronic poly-A&T sequences</p> <p>genome_intron_summary.txt: summary of output</p>

3 Plotting Frequency Distribution

3.1 Poly-A&T Frequency Distribution

Plot the normalized poly-A&T sequence length frequency distribution across specific chromosomes or the whole genome by calling the ‘FreqDistr_Grapher.py’ script.

```
python FreqDistr_Grapher.py freq chr
```

Input Parameters	freq: file location containing poly-A&T sequence length frequency distribution (genome_freq.txt) chr: chromosome ID (can enter multiple at once separated by spaces) to plot frequency distribution of specified chromosome(s) or ‘whole_genome’ to plot frequency distribution of whole genome
Output Files	chr_A.png: normalized poly-A sequence length distribution plot for specified chromosomes (or whole genome) chr_T.png: normalized poly-T sequence length distribution plot for specified chromosomes (or whole genome)

3.2 Intronic Poly-A&T Frequency Distribution

Plot the normalized intronic poly-A&T sequence length frequency distribution across the whole genome by calling the ‘Intron_FreqDistr_Grapher.py’ script.

```
python Intron_FreqDistr_Grapher.py freq
```

Input Parameters	freq: file location containing intronic poly-A&T sequence length frequency distribution (genome_intron_freq.txt)
Output Files	Intronic_Poly_A.png: normalized intronic poly-A sequence length distribution plot for whole genome Intronic_Poly_T.png: normalized intronic poly-T sequence length distribution plot for whole genome