BB - 8/13/19

Importance of Motif Position

First\_Half vs Back\_Half

Front Back

S P Y A P Y D R

- Only worked with sequences of length 20 for now. WDWDWD\*\*\* sequences were removed.

-Motif: (?=[WYF].{0,4}[DE]|[DE].{0,4}[WYF])

Front: 160 live, 6633 die, 2.4% live

Back: 153 live, 6037 die, 2.5% live

-Motif: [RHK]

Front: 235 live, 22342 die, 1.0% live

Back: 249 live, 22449 die, 1.1% live

-Motif: [WYF].{0,4}[DE]

Front: 100 live, 4091 die, 2.4% live

Back: 106 live, 3725 die, 2.8% live

-Motif: [DE].{0,4}[WYF]

Front: 105 live, 4061 die, 2.5% live

Back: 101 live, 3567 die, 2.6% live

Taking a step up to look at frequency of motifs and not just presence

-Motif: [DE] frequency

Front:

|  |  |  |  |
| --- | --- | --- | --- |
| Motif Count | Live\_count | Die\_count | Live\_percent |
| 0 | 111 | 12876 | 0.8 % |
| 1 | 140 | 9793 | 1.4 % |
| 2 | 67 | 3341 | 2.0 % |
| 3 | 22 | 650 | 3.3 % |
| 4 | 4 | 82 | 4.7 % |
| 5 | 0 | 6 | 0 % |
| 6 | 0 | 1 | 0 % |

Back:

|  |  |  |  |
| --- | --- | --- | --- |
| Motif Count | Live\_count | Die\_count | Live\_percent |
| 0 | 122 | 13591 | 0.9 % |
| 1 | 124 | 9559 | 1.3 % |
| 2 | 71 | 2972 | 2.3 % |
| 3 | 24 | 550 | 4.2 % |
| 4 | 3 | 69 | 4.2 % |
| 5 | 0 | 8 | 0 % |

-Motif: [WYF] frequency

Front:

|  |  |  |  |
| --- | --- | --- | --- |
| Motif Count | Live\_count | Die\_count | Live\_percent |
| 0 | 95 | 11459 | 0.8 % |
| 1 | 138 | 10096 | 1.3 % |
| 2 | 78 | 4040 | 1.9 % |
| 3 | 27 | 966 | 2.7 % |
| 4 | 5 | 168 | 2.9 % |
| 5 | 1 | 20 | 4.8 % |

Back:

|  |  |  |  |
| --- | --- | --- | --- |
| Motif Count | Live\_count | Die\_count | Live\_percent |
| 0 | 93 | 12020 | 0.8 % |
| 1 | 131 | 9953 | 1.3 % |
| 2 | 82 | 3812 | 2.1 % |
| 3 | 35 | 842 | 4.0 % |
| 4 | 3 | 111 | 2.6 % |
| 5 | 0 | 11 | 0 % |

Motif matching in sections FOR SEQUENCES of LENGTH 20 aa

ALMNQ WTHEL QQQNR TCGDL

Motif: ‘([WYF].{0,3}[DE]|[DE].{0,3}[WYF])’

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Section 1 | Section 2 | Section 3 | Section 4 | Live\_count | Die\_count | Live\_percent |
| 1 | 1 | 1 | 1 | 2 | 0 | 100% |
| 1 | 0 | 1 | 1 | 3 | 13 | 18.75% |
| 0 | 1 | 1 | 1 | 2 | 9 | 18.18% |
| 1 | 1 | 1 | 0 | 3 | 15 | 16.67% |
| 1 | 1 | 0 | 1 | 3 | 19 | 13.64% |
| 0 | 1 | 1 | 0 | 10 | 148 | 6.33% |
| 1 | 0 | 1 | 0 | 11 | 184 | 5.64% |
| 1 | 0 | 0 | 1 | 9 | 154 | 5.52% |
| 0 | 1 | 0 | 1 | 8 | 157 | 4.85% |
| 0 | 0 | 1 | 1 | 5 | 161 | 3.01% |
| 0 | 1 | 0 | 0 | 40 | 1833 | 2.14% |
| 0 | 0 | 0 | 1 | 33 | 1522 | 2.12% |
| 1 | 1 | 0 | 0 | 4 | 201 | 1.95% |
| 0 | 0 | 1 | 0 | 30 | 1773 | 1.66% |
| 1 | 0 | 0 | 0 | 25 | 1909 | 1.29% |
| 0 | 0 | 0 | 0 | 156 | 18651 | 0.83% |

Motif: ‘[WYF]’

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Section 1 | Section 2 | Section 3 | Section 4 | Live\_count | Die\_count | Live\_percent |
| 1 | 1 | 1 | 1 | 16 | 360 | 4.26% |
| 0 | 1 | 1 | 1 | 22 | 658 | 3.24% |
| 1 | 1 | 1 | 0 | 23 | 735 | 3.03% |
| 1 | 1 | 0 | 1 | 19 | 693 | 2.67% |
| 1 | 0 | 1 | 1 | 15 | 670 | 2.19% |
| 1 | 0 | 0 | 1 | 23 | 1223 | 1.85% |
| 0 | 1 | 0 | 1 | 24 | 1284 | 1.83% |
| 0 | 1 | 1 | 0 | 21 | 1346 | 1.54% |
| 1 | 0 | 1 | 0 | 22 | 1422 | 1.52% |
| 0 | 0 | 1 | 1 | 18 | 1223 | 1.45% |
| 0 | 0 | 0 | 1 | 31 | 2406 | 1.27% |
| 1 | 1 | 0 | 0 | 17 | 1525 | 1.10% |
| 0 | 1 | 0 | 0 | 27 | 2617 | 1.02% |
| 1 | 0 | 0 | 0 | 20 | 2757 | 0.72% |
| 0 | 0 | 1 | 0 | 17 | 2709 | 0.63% |
| 0 | 0 | 0 | 0 | 29 | 5121 | 0.56% |

Motif: ‘[DE]’

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Section 1 | Section 2 | Section 3 | Section 4 | Live\_count | Die\_count | Live\_percent |
| 1 | 1 | 1 | 1 | 11 | 188 | 5.53% |
| 1 | 0 | 1 | 1 | 24 | 480 | 4.76% |
| 1 | 1 | 1 | 0 | 19 | 512 | 3.59% |
| 1 | 1 | 0 | 1 | 18 | 513 | 3.39% |
| 0 | 1 | 1 | 1 | 11 | 482 | 2.23% |
| 0 | 1 | 1 | 0 | 27 | 1189 | 2.22% |
| 1 | 0 | 1 | 0 | 23 | 1209 | 1.87% |
| 0 | 1 | 0 | 1 | 18 | 1068 | 1.66% |
| 1 | 0 | 0 | 1 | 18 | 1137 | 1.56% |
| 0 | 0 | 1 | 1 | 15 | 1008 | 1.47% |
| 1 | 1 | 0 | 0 | 16 | 1325 | 1.19% |
| 1 | 0 | 0 | 0 | 29 | 2921 | 0.98% |
| 0 | 0 | 0 | 0 | 58 | 6496 | 0.88% |
| 0 | 0 | 0 | 1 | 20 | 2601 | 0.76% |
| 0 | 1 | 0 | 0 | 19 | 2849 | 0.66% |
| 0 | 0 | 1 | 0 | 18 | 2771 | 0.65% |

Conclusion: Frequency of a motif occurring is more important than position. Increasing aromatics or acidics by themselves is good but much better to increase together

Motif: ‘[RHK]’

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Section 1 | Section 2 | Section 3 | Section 4 | Live\_count | Die\_count | Live\_percent |
| 0 | 0 | 0 | 0 | 33 | 733 | 4.31% |
| 0 | 0 | 1 | 0 | 30 | 999 | 2.92% |
| 0 | 0 | 0 | 1 | 30 | 1061 | 2.75% |
| 0 | 1 | 0 | 0 | 29 | 1056 | 2.67% |
| 1 | 0 | 0 | 0 | 17 | 968 | 1.73% |
| 1 | 0 | 0 | 1 | 25 | 1524 | 1.61% |
| 0 | 1 | 0 | 1 | 17 | 1533 | 1.10% |
| 0 | 1 | 1 | 0 | 17 | 1571 | 1.07% |
| 1 | 1 | 0 | 0 | 16 | 1543 | 1.03% |
| 1 | 1 | 1 | 0 | 23 | 2273 | 1.00% |
| 0 | 0 | 1 | 1 | 16 | 1614 | 0.98% |
| 1 | 1 | 0 | 1 | 22 | 2284 | 0.95% |
| 1 | 0 | 1 | 0 | 13 | 1499 | 0.86% |
| 1 | 0 | 1 | 1 | 18 | 2306 | 0.77% |
| 1 | 1 | 1 | 1 | 25 | 3468 | 0.72% |
| 0 | 1 | 1 | 1 | 13 | 2317 | 0.56% |