**29 July 2014**

**In silico modeling of performance of RNA mixes**

Refer to excel file “140730\_Final SequIn Mixes\_improvedv2.xlsx”.

**Mix A**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **REF** | **VAR** | **Grp** | **REF length** | **VAR length** | **AIM** | **RATIO** | **RATIO** | **CON** | **CON** | **CON READ** | **CON READS** | **PER KB** | **PER KB** | **ROUND** | **ROUND** |
| R\_1\_1\_R | R\_1\_1\_V | A | 703 | 785 | 10000000 | 1 | 1 | 0.5 | 0.5 | 5000000 | 5000000 | 3515000 | 3925000 | 3515000 | 3925000 |
| R\_1\_2\_R | R\_1\_2\_V | A | 1310 | 1582 | 10000000 | 3 | 1 | 0.75 | 0.25 | 7500000 | 2500000 | 9825000 | 3955000 | 9825000 | 3955000 |
| R\_1\_3\_R | R\_1\_3\_V | A | 1940 | 698 | 10000000 | 1 | 3 | 0.25 | 0.75 | 2500000 | 7500000 | 4850000 | 5235000 | 4850000 | 5235000 |
| R\_1\_4\_R |  | D | 664 |  | 152 | 1 | 0 | 1 | 0 | 152 |  | 100.928 |  | 100 |  |
| R\_10\_1\_R | R\_10\_1\_V | A | 719 | 430 | 2500000 | 1 | 1 | 0.5 | 0.5 | 1250000 | 1250000 | 898750 | 537500 | 898750 | 537500 |
| R\_10\_2\_R | R\_10\_2\_V | A | 1490 | 1362 | 2500000 | 1 | 3 | 0.25 | 0.75 | 625000 | 1875000 | 931250 | 2553750 | 931250 | 2553750 |
| R\_10\_3\_R | R\_10\_3\_V | A | 1754 | 1856 | 2500000 | 3 | 1 | 0.75 | 0.25 | 1875000 | 625000 | 3288750 | 1160000 | 3288750 | 1160000 |
| R\_2\_1\_R | R\_2\_1\_V | B | 867 | 794 | 625000 | 1 | 1 | 0.5 | 0.5 | 312500 | 312500 | 270937.5 | 248125 | 270937 | 248125 |
| R\_2\_2\_R | R\_2\_2\_V | B | 1169 | 1041 | 625000 | 1 | 3 | 0.25 | 0.75 | 156250 | 468750 | 182656.25 | 487968.75 | 182656 | 487968 |
| R\_2\_3\_R | R\_2\_3\_V | B | 1689 | 1182 | 625000 | 3 | 1 | 0.75 | 0.25 | 468750 | 156250 | 791718.75 | 184687.5 | 791718 | 184687 |
| R\_2\_4\_R |  |  | 4577 |  | 625000 | 1 | 0 | 1 | 0 | 625000 |  | 2860625 |  | 2860625 |  |
| R\_3\_1\_R | R\_3\_1\_V | B | 661 | 1466 | 156250 | 1 | 1 | 0.5 | 0.5 | 78125 | 78125 | 51640.625 | 114531.25 | 51640 | 114531 |
| R\_3\_2\_R | R\_3\_2\_V | B | 944 | 783 | 156249 | 1 | 3 | 0.25 | 0.75 | 39062.25 | 117186.75 | 36874.764 | 91757.22525 | 36874 | 91757 |
| R\_3\_3\_R | R\_3\_3\_V | B | 2361 | 1022 | 156249 | 3 | 1 | 0.75 | 0.25 | 117186.75 | 39062.25 | 276677.9168 | 39921.6195 | 276677 | 39921 |
| R\_4\_1\_R | R\_4\_1\_V | B | 439 | 1001 | 39061 | 3 | 1 | 0.75 | 0.25 | 29295.75 | 9765.25 | 12860.83425 | 9775.01525 | 12860 | 9775 |
| R\_4\_2\_R | R\_4\_2\_V | B | 975 | 604 | 39062 | 1 | 1 | 0.5 | 0.5 | 19531 | 19531 | 19042.725 | 11796.724 | 19042 | 11796 |
| R\_4\_3\_R | R\_4\_3\_V | B | 2617 | 2907 | 39061 | 1 | 3 | 0.25 | 0.75 | 9765.25 | 29295.75 | 25555.65925 | 85162.74525 | 25555 | 85162 |
| R\_5\_1\_R | R\_5\_1\_V | C | 884 | 890 | 9765 | 1 | 3 | 0.25 | 0.75 | 2441.25 | 7323.75 | 2158.065 | 6518.1375 | 2158 | 6518 |
| R\_5\_2\_R | R\_5\_2\_V | C | 1118 | 1504 | 9765 | 3 | 1 | 0.75 | 0.25 | 7323.75 | 2441.25 | 8187.9525 | 3671.64 | 8187 | 3671 |
| R\_5\_3\_R |  | C | 1986 | 1969 | 9765 | 1 | 0 | 1 | 0 | 9765 |  | 19393.29 |  | 19393 |  |
| R\_6\_1\_R | R\_6\_1\_V | C | 663 | 535 | 2440 | 1 | 1 | 0.5 | 0.5 | 1220 | 1220 | 808.86 | 652.7 | 808 | 652 |
| R\_6\_2\_R | R\_6\_2\_V | C | 955 | 627 | 2441 | 1 | 3 | 0.25 | 0.75 | 610.25 | 1830.75 | 582.78875 | 1147.88025 | 582 | 1147 |
| R\_6\_3\_R | R\_6\_3\_V | C | 2336 | 1506 | 2441 | 3 | 1 | 0.75 | 0.25 | 1830.75 | 610.25 | 4276.632 | 919.0365 | 4276 | 919 |
| R\_7\_1\_R | R\_7\_1\_V | C | 825 | 476 | 609 | 3 | 1 | 0.75 | 0.25 | 456.75 | 152.25 | 376.81875 | 72.471 | 376 | 72 |
| R\_7\_2\_R | R\_7\_2\_V | C | 988 | 1227 | 610 | 1 | 1 | 0.5 | 0.5 | 305 | 305 | 301.34 | 374.235 | 301 | 374 |
| R\_7\_3\_R | R\_7\_3\_V | C | 1914 | 2865 | 609 | 1 | 3 | 0.25 | 0.75 | 152.25 | 456.75 | 291.4065 | 1308.58875 | 291 | 1308 |
| R\_8\_1\_R | R\_8\_1\_V | D | 889 | 1651 | 152 | 3 | 1 | 0.75 | 0.25 | 114 | 38 | 101.346 | 62.738 | 101 | 62 |
| R\_8\_2\_R | R\_8\_2\_V | D | 1400 | 982 | 152 | 1 | 1 | 0.5 | 0.5 | 76 | 76 | 106.4 | 74.632 | 106 | 74 |
| R\_8\_3\_R | R\_8\_3\_V | D | 1648 | 945 | 152 | 1 | 3 | 0.25 | 0.75 | 38 | 114 | 62.624 | 107.73 | 62 | 107 |
| R\_9\_1\_R | R\_9\_1\_V | D | 605 | 954 | 38 | 1 | 1 | 0.5 | 0.5 | 19 | 19 | 11.495 | 18.126 | 11 | 18 |
| R\_9\_2\_R | R\_9\_2\_V | D | 1637 | 727 | 37 | 1 | 3 | 0.25 | 0.75 | 9.25 | 27.75 | 15.14225 | 20.17425 | 15 | 20 |
| R\_9\_3\_R | R\_9\_3\_V | D | 1725 | 1177 | 37 | 3 | 1 | 0.75 | 0.25 | 27.75 | 9.25 | 47.86875 | 10.88725 | 47 | 10 |

**Mix B**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **REF** | **VAR** | **Grp** | **REF length** | **VAR length** | **AIM** | **RATIO** | **RATIO** | **CON** | **CON** | **CON READ** | **CON READS** | **PER KB** | **PER KB** | **ROUND** | **ROUND** |
| R\_1\_1\_R | R\_1\_1\_V |  | 703 | 785 | 2500000 | 1 | 1 |  | 0.5 | 0.5 | 1250000 | 1250000 | 878750 | 981250 | 878750 |
| R\_1\_2\_R | R\_1\_2\_V |  | 1310 | 1582 | 2500000 | 1 | 3 |  | 0.25 | 0.75 | 625000 | 1875000 | 818750 | 2966250 | 818750 |
| R\_1\_3\_R | R\_1\_3\_V |  | 1940 | 698 | 2500000 | 3 | 1 |  | 0.75 | 0.25 | 1875000 | 625000 | 3637500 | 436250 | 3637500 |
| R\_1\_4\_R |  |  | 664 |  | 304 | 1 | 0 |  | 1 |  | 304 |  | 201.856 |  | 201 |
| R\_10\_1\_R | R\_10\_1\_V |  | 719 | 430 | 625000 | 1 | 1 |  | 0.5 | 0.5 | 312500 | 312500 | 224687.5 | 134375 | 224687 |
| R\_10\_2\_R | R\_10\_2\_V |  | 1490 | 1362 | 625000 | 3 | 1 |  | 0.75 | 0.25 | 468750 | 156250 | 698437.5 | 212812.5 | 698437 |
| R\_10\_3\_R | R\_10\_3\_V |  | 1754 | 1856 | 625000 | 1 | 3 |  | 0.25 | 0.75 | 156250 | 468750 | 274062.5 | 870000 | 274062 |
| R\_2\_1\_R | R\_2\_1\_V |  | 867 | 794 | 312500 | 1 | 1 |  | 0.5 | 0.5 | 156250 | 156250 | 135468.75 | 124062.5 | 135468 |
| R\_2\_2\_R | R\_2\_2\_V |  | 1169 | 1041 | 312500 | 3 | 1 |  | 0.75 | 0.25 | 234375 | 78125 | 273984.375 | 81328.125 | 273984 |
| R\_2\_3\_R | R\_2\_3\_V |  | 1689 | 1182 | 312500 | 1 | 3 |  | 0.25 | 0.75 | 78125 | 234375 | 131953.125 | 277031.25 | 131953 |
| R\_2\_4\_R |  |  | 4577 |  | 312500 | 1 | 0 |  | 1 |  | 312500 |  | 1430312.5 |  | 1430312 |
| R\_3\_1\_R | R\_3\_1\_V |  | 661 | 1466 | 78124 | 1 | 1 |  | 0.5 | 0.5 | 39062 | 39062 | 25819.982 | 57264.892 | 25819 |
| R\_3\_2\_R | R\_3\_2\_V |  | 944 | 783 | 78124 | 3 | 1 |  | 0.75 | 0.25 | 58593 | 19531 | 55311.792 | 15292.773 | 55311 |
| R\_3\_3\_R | R\_3\_3\_V |  | 2361 | 1022 | 78124 | 1 | 3 |  | 0.25 | 0.75 | 19531 | 58593 | 46112.691 | 59882.046 | 46112 |
| R\_4\_1\_R | R\_4\_1\_V |  | 439 | 1001 | 19530 | 1 | 3 |  | 0.25 | 0.75 | 4882.5 | 14647.5 | 2143.4175 | 14662.1475 | 2143 |
| R\_4\_2\_R | R\_4\_2\_V |  | 975 | 604 | 19530 | 1 | 1 |  | 0.5 | 0.5 | 9765 | 9765 | 9520.875 | 5898.06 | 9520 |
| R\_4\_3\_R | R\_4\_3\_V |  | 2617 | 2907 | 19530 | 3 | 1 |  | 0.75 | 0.25 | 14647.5 | 4882.5 | 38332.5075 | 14193.4275 | 38332 |
| R\_5\_1\_R | R\_5\_1\_V |  | 884 | 890 | 9765 | 3 | 1 |  | 0.75 | 0.25 | 7323.75 | 2441.25 | 6474.195 | 2172.7125 | 6474 |
| R\_5\_2\_R | R\_5\_2\_V |  | 1118 | 1504 | 9765 | 1 | 3 |  | 0.25 | 0.75 | 2441.25 | 7323.75 | 2729.3175 | 11014.92 | 2729 |
| R\_5\_3\_R |  |  | 1986 | 1969 | 9765 | 1 | 0 |  | 1 |  | 9765 |  | 19393.29 |  | 19393 |
| R\_6\_1\_R | R\_6\_1\_V |  | 663 | 535 | 2440 | 1 | 1 |  | 0.5 | 0.5 | 1220 | 1220 | 808.86 | 652.7 | 808 |
| R\_6\_2\_R | R\_6\_2\_V |  | 955 | 627 | 2441 | 3 | 1 |  | 0.75 | 0.25 | 1830.75 | 610.25 | 1748.36625 | 382.62675 | 1748 |
| R\_6\_3\_R | R\_6\_3\_V |  | 2336 | 1506 | 2441 | 1 | 3 |  | 0.25 | 0.75 | 610.25 | 1830.75 | 1425.544 | 2757.1095 | 1425 |
| R\_7\_1\_R | R\_7\_1\_V |  | 825 | 476 | 609 | 1 | 3 |  | 0.25 | 0.75 | 152.25 | 456.75 | 125.60625 | 217.413 | 125 |
| R\_7\_2\_R | R\_7\_2\_V |  | 988 | 1227 | 610 | 1 | 1 |  | 0.5 | 0.5 | 305 | 305 | 301.34 | 374.235 | 301 |
| R\_7\_3\_R | R\_7\_3\_V |  | 1914 | 2865 | 609 | 3 | 1 |  | 0.75 | 0.25 | 456.75 | 152.25 | 874.2195 | 436.19625 | 874 |
| R\_8\_1\_R | R\_8\_1\_V |  | 889 | 1651 | 304 | 1 | 3 |  | 0.25 | 0.75 | 76 | 228 | 67.564 | 376.428 | 67 |
| R\_8\_2\_R | R\_8\_2\_V |  | 1400 | 982 | 304 | 1 | 1 |  | 0.5 | 0.5 | 152 | 152 | 212.8 | 149.264 | 212 |
| R\_8\_3\_R | R\_8\_3\_V |  | 1648 | 945 | 304 | 3 | 1 |  | 0.75 | 0.25 | 228 | 76 | 375.744 | 71.82 | 375 |
| R\_9\_1\_R | R\_9\_1\_V |  | 605 | 954 | 76 | 1 | 1 |  | 0.5 | 0.5 | 38 | 38 | 22.99 | 36.252 | 22 |
| R\_9\_2\_R | R\_9\_2\_V |  | 1637 | 727 | 76 | 3 | 1 |  | 0.75 | 0.25 | 57 | 19 | 93.309 | 13.813 | 93 |
| R\_9\_3\_R | R\_9\_3\_V |  | 1725 | 1177 | 76 | 1 | 3 |  | 0.25 | 0.75 | 19 | 57 | 32.775 | 67.089 | 32 |

Create simulated reads with Sherman (samtools)

$ module load samtools/0.1.18 cufflinks/2.1.1

$ while read X

$ do

$ wgsim -d 400 -N 10000 -1 101 -2 101 ${X} ${X}.R1.fq ${X}.R2.fq

$ done < /home/wenche/Simulations/fastas/standardsfasta/RNA.standards.fa.list.txt

#where RNA.standards.fa.list.txt is a list of all the RNA standards in .fa, i.e.,

R\_10\_1\_R.fa

R\_10\_1\_V.fa

R\_10\_2\_R.fa

R\_10\_2\_V.fa

R\_10\_3l\_R.fa

R\_10\_3\_V.fa

…

#Create bowtie2 index files

bowtie2-build -f /hox/u/uqtmerce/Wendy/ChrT.5.10.fa ChrT.5.10

# Aliqn your simulated reads to the reference genome

$ while read X

$ do

$ bowtie2 -x ChrT.5.10 -1 ${X}.fa.R1.fq -2 ${X}.fa.R2.fq -S ${X}.sam

$ done < RNA.standards.list.txt

# where RNA.standards.list.txt is a list of all the RNA standards, i.e.,

R\_10\_1\_R

R\_10\_1\_V

R\_10\_2\_R

R\_10\_2\_V

R\_10\_3l\_R

R\_10\_3\_V

…

$ head -3515000 sams/R\_1\_1\_R.sam > ./tempA2/R\_1\_1\_R.mixA.sam

# do for all the .sam files

$ cat \*.sam > combined.sam

#grep any line that DOES NOT contain “@” at the beginning of the line

$ grep -v ^@ combined.sam >combined.noheader.sam

$ cat ../header.txt combined.sam > combined.head.sam

$ samtools view -bt ~/SimData/RNAsim/Align/hg19\_ercc\_chrT.fa combined.head.sam >combined.head.bam

$ samtools sort combined.head.bam combined.head.sort

$ cufflinks -p 8 -G ../RNAstandards.ref.iso.gtf combined.head.sort.bam

#to get the genes and isoforms FPKM values

$ less genes.fpkm\_tracking

$ less isoforms.fpkm\_tracking

In the ideal situation, the mixes will appear as…

 

 

 