Comparing RNA-seq vs DNA-seq methods for identifying T cell repertoires

Michelle Miron

12/2/2017

The goal of this project

T cell clones can be identified by the sequence of their T cell receptor gene. Unlike most cells in the body, T cells have rearragned segnemtn of DNA that is esentially a barcode to identify that T cells and all subsequent daughter cells originating from that cell.

There is newly developed software to extract T cell receptor (TCR) sequences from bulk RNA-seq data.

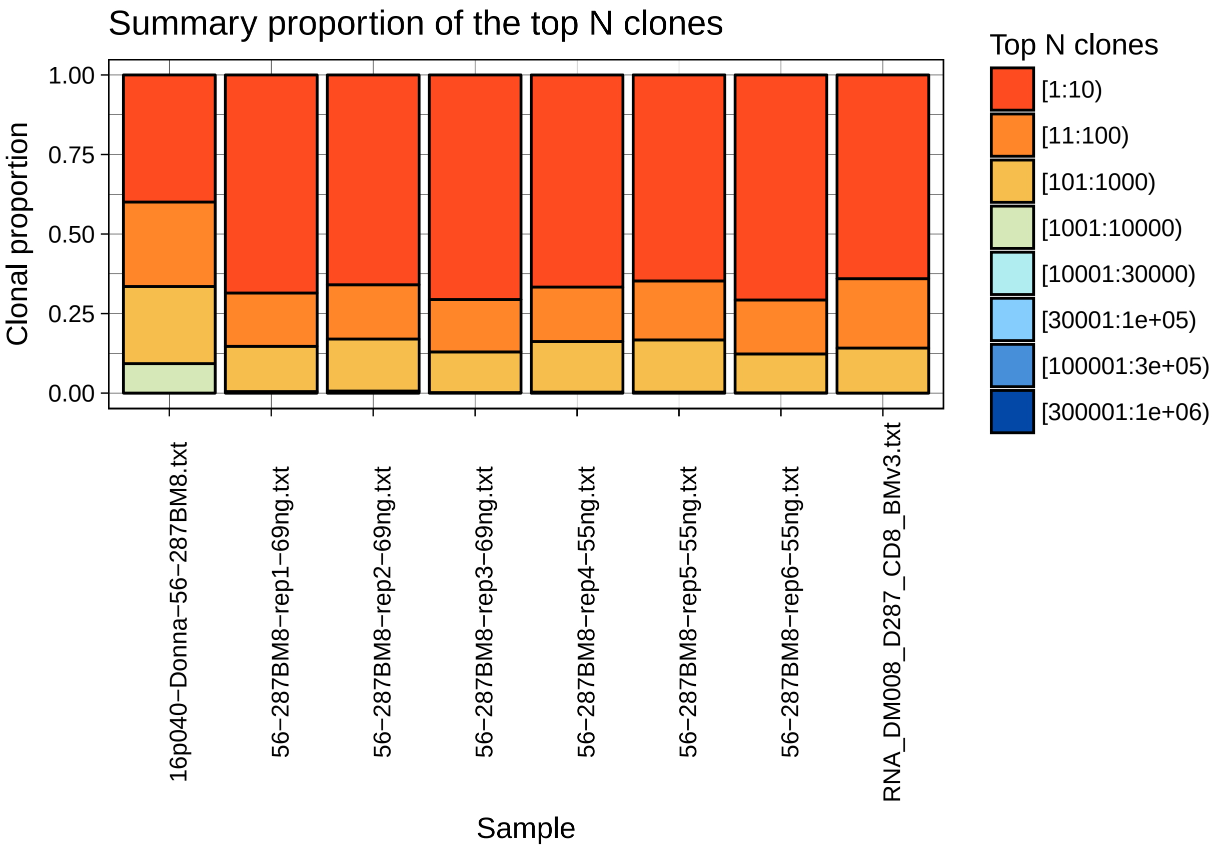
The novelty here, is that usually in order to detect T cell receptor sequences, you design specific primers to sequence the T cell receptor locus within the DNA. However, recently people have designed algorithms to extract TCR sequences from bulk-RNA seq data.

A major limitation in analyzing TCR sequences that were extracted from bulk RNA-seq data, is that you will have very limited sampling. How limited is this sampling? What portion of the reperotire is this capturing? Equivalent to about about many cells? That is the question I will be addressing here.

I am uniquely positioned to answer this questions, with experimenta data from both TCR seq of DNA and bulk RNA-seq of sorted T cells.

Methods

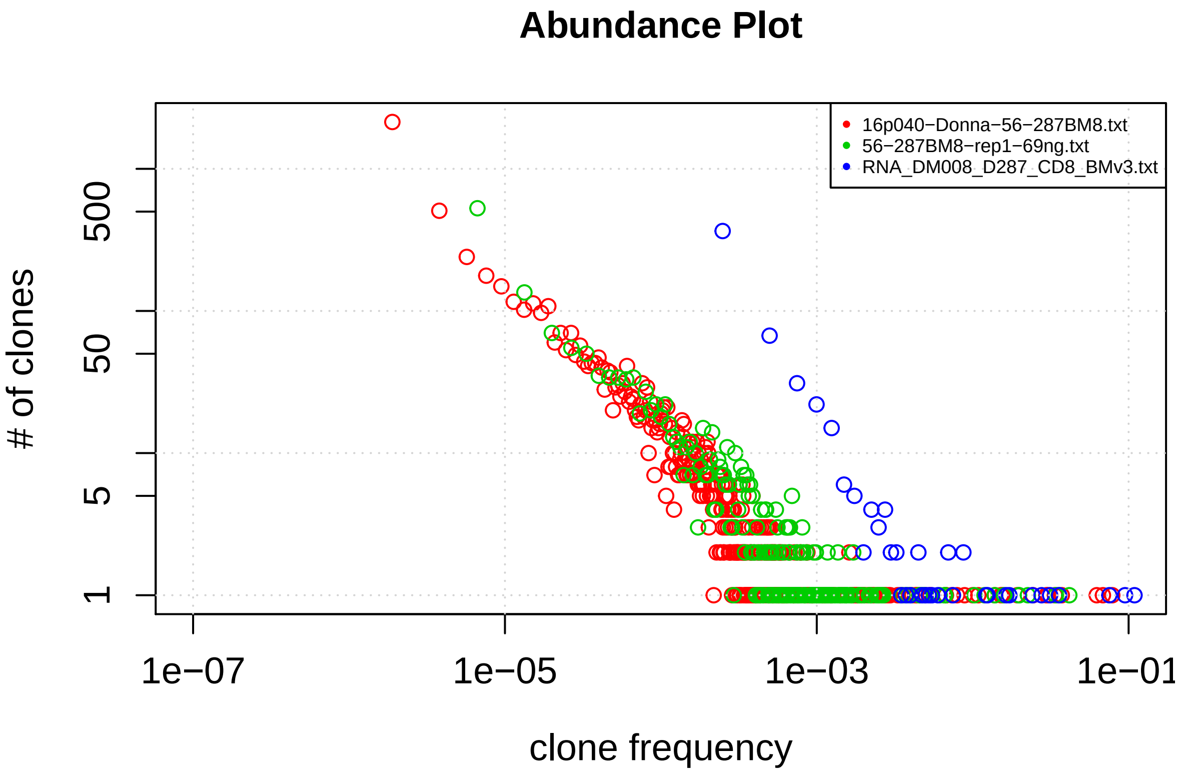
I will be using MiXCR https://milaboratory.com/software/mixcr/.



```
##                         CDR3.nucleotide.sequence Umi.count.x Umi.count.y
## 1     AGCAGTGCTAGGCGAGGGGGCGCAGGACAGATACGCAGTATTTT           1           0
## 2       AGTGCCACCAGCACGGAGACGGAACAAAAGACCCAGTACTTC           1           1
## 3             AGTGCCAGCAGCGTAGAGGGGGGCGAGCAGTACTTC           1           0
## 4 AGTGCCAGCAGCTATATGGCAGGCCCCCGAGTGGGCGGCTACACCTTC           1           0
## 5             AGTGCCAGCAGCTTAGTAAACGATGAGCAGTTCTTC           8           8
## 6             AGTGCCAGCAGCTTAGTAAGCGATGAGCAGTTCTTC           1           1
##   Umi.count
## 1         0
## 2         0
## 3         0
## 4         0
## 5         0
## 6         0
```

```
##                           CDR3.nucleotide.sequence Umi.count.x
## 6695         TGTTCCTGGGACTCCGGGACAGGGAGCGGCTACACCTTC           3
## 6696         TGTTCGAGTTATACCAGTGGCTGGTTGACTGACTCGTGG           0
## 6697    TTCTCTTTATTCTCCCAGGGGGCACTAATGAAAAACTGTTTTTT           9
## 6698    TTCTCTTTATTCTCCCAGGGGGCTCTAATGAAAAACTGTTTTTT           1
## 6699 TTTGCCACCAGCAGAGATCTTTGGGGGGGGGCGATGAGCAGTTCTTC           2
## 6700         TTTGCCAGCGCCACTCAGTCGGATCCTGGCTACACCTTC           1
##      Umi.count.y Umi.count
## 6695           3         0
## 6696           0         1
## 6697           9         0
## 6698           1         0
## 6699              0         0
## 6700              0         0
```

```
## [1] 4
```



```
##                               CDR3.nucleotide.sequence Umi.count
## 1           TGCAGTGCCGAAAGGACAGGGGTGACTGAAAAACTGTTTTTT     41398
## 2              TGTGCCAGTAGTACCGATTTCGACTACGAGCAGTACTTC     36074
## 3 TGTGCCAGCAGTCCCCCCCGTGTGCCGCGGCTCCTACAATGAGCAGTTCTTC     32925
## 4           TGTGCCAGCAACATGGGGGGGGGAAATCAGCCCCAGCATTTT     19645
## 5  TGTGCCACCAGCAGAGATCTCACCCAGGGTAGAAGAAACACCATATATTTT     17757
## 6                 TGTGCCAGCAGTTCAGCTAACTATGGCTACACCTTC     15986
```

```
##                               CDR3.nucleotide.sequence Umi.count
## 1           TGTGCCAGCAGTTTACCCCAGGTACCCACTGAAGCTTTCTTT     70567
## 2           TGCAGTGCCGAAAGGACAGGGGTGACTGAAAAACTGTTTTTT      6248
## 3 TGTGCCAGCAGTCCCCCCCGTGTGCCGCGGCTCCTACAATGAGCAGTTCTTC      5345
## 4              TGTGCCAGTAGTACCGATTTCGACTACGAGCAGTACTTC      5121
## 5           TGTGCCAGCAACATGGGGGGGGGAAATCAGCCCCAGCATTTT      3390
## 6                 TGTGCCAGCAGTTCAGCTAACTATGGCTACACCTTC      2973
```

```
##                           CDR3.nucleotide.sequence Umi.count
## 1       TGTGCCAGCAGTTTACCCCAGGTACCCACTGAAGCTTTCTTT       837
## 2 TGTGCTGTGGAGTTAATGAATAATAATGCAGGCAACATGCTCACCTTT       438
## 3                      TGTGCCACCTGGGATAAGAAACTCTTT       381
## 4                TGTGCTCTGAACACCGGTAACCAGTTCTATTTT       302
## 5   TGTGCTGCGTGGGATCCCGCCGGCCACTGGTTGGTTCAAGATATTT       144
## 6 TGTGCCAGCAGTACACCGGCGGGACAGGGGATCAATGAGCAGTTCTTC       126
```