

```
def (FileName):
```

```
    read genome sequences by using readlines() from txt
```

```
return genome sequences
```

```
def prepareValidData(FileName):
```

```
    create new list to valid lines
```

```
    if( genome sequences >= 2700 and 'N' not in sequences)
```

```
        add genome sequences to valid lines list
```

```
def probes (fileName, probeLen):
```

```
    find valid genome sequences with substract probelen-1 to 27000
```

```
    create dict
```

```
    create set to row check
```

```
    for ( travel i in legth of valid data)
```

```
        clear row_check set in every row to check if you count same probe in same row
```

```
        for ( travel in legth of genomeSequences)
```

```
            if (validData not include in dict)
```

```
                make valid data value equals one
```

```
                add valid data to row_check set to control if you are on same line
```

```
            elif ( validData not in row_check set)
```

```
                increase value of validData in dictionary
```

```
                add valid data to row_check set
```

```
    return dictionary , validData
```

```
__name__ == '__main__':
```

```
Find maximum value in dict
```

```
counter = 0 (counter is the )
```

```
for(travel in dictionary items)
```

```
    if ( value of dictionary equals with max value)
```

```
        counter+=1
```

```
write the legth of the valid data , max value and counter to file
```

The number of Valid sequences	The number of sequences probe is found	Probe length	Time in seconds
5	16	2	0.9784250259399414
10	22378	90	1.363420009613037
100	6195	90	4.944784879684448
250	1289	90	11.269296884536743
500	464	90	21.49469780921936
1000	29	90	42.41370177268982
10	21931	100	1.3194384574890137
50	10584	100	2.8952949047088623
100	5323	100	5.712731838226318
250	888	100	11.303974866867065
500	319	100	22.74420738220215