

## Count sequences frequencies using x- and y- coordinates

Source code: count\_CGR.cpp

Example input file: Eya1\_1\_CGR.txt

Example output file: Eya1\_1\_CGR\_out.txt

*Instructions:*

1) Compile source code to app called CGR using the line command:

```
g++ count_CGR.cpp -o count_CGR
```

2) Run the program by invoking the following line command:

```
./count_CGR 4 1 Eya1_1_CGR.txt Eya1_1_CGR_out.txt
```

Note you can change the resolution of the box size by changing the number in the command line. The number '4' means you divide the CGR plot into 16 boxes (e.g. AA, AT, AC... TT). The number 8 means you have divided the CGR plot into 64 boxes (e.g., AAA, AAC... TTT).

The second number 1 is a toggle switch for the header. If you want a header to be produced set the toggle to 1. If you do not want a header, set the toggle switch to 0.

Here is the output you should get:

```
GG  CG  GC  CC  AG  TG  AC  TC  GA  CA  GT  CT  AA  TA  AT  TT
Eya1_1_CGR.txt 4 237 113 231 267 304 325 279 252 262 356 249
                293 348 235 271 324 4346
```

Transferring the output to MS Excel produces a table that looks like this:

Transcript	Size	GG	CG	GC	CC	AG	TG	AC	TC	GA	CA	GT	CT	AA	TA	AT	TT	Len
Eya1_1_CGR.txt	4	237	113	231	267	304	325	279	252	262	356	249	293	348	235	271	324	4346

Eya1\_1\_CGR.txt is the file name. the resolution of 4 means there are 16 boxes. The boxes (di-nucleotides in this case) are shown in the top row and the corresponding counts are shown in the second row. The last column is the total number of nucleotides in the gene, 4,296.