**Bioinformatics for NGS – Exercise Sheet 5**

**Task 5:**

5 miRNA precursors from miRBase followed by the RNAfold links (randomly chosen):

1.

<http://www.mirbase.org/cgi-bin/mirna_entry.pl?acc=MI0000266> hsa-mir-10a

<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi?PAGE=3&ID=q7v0XUZuh5>

2.

<http://www.mirbase.org/cgi-bin/mirna_entry.pl?acc=MI0000069> hsa-mir-15a

<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi?PAGE=3&ID=YWc0JC8vxm>

3.

<http://www.mirbase.org/cgi-bin/mirna_entry.pl?acc=MI0000089> hsa-mir-31

<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi?PAGE=3&ID=wQLSp6OEpr>

4.

<http://www.mirbase.org/cgi-bin/mirna_entry.pl?acc=MI0000483> hsa-mir-186

<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi?PAGE=3&ID=PTpFY_IxnB>

5.

<http://www.mirbase.org/cgi-bin/mirna_entry.pl?acc=MI0000102> hsa-mir-100

<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi?PAGE=3&ID=1KgMABmmv2>

5 stretches from the human genome (hg38, randomly chosen):

1.

Chr. 1 unplaced bases(no pos. available)

DNA:

GAATTCAGCTGAGAAGAACAGGCAAGGACTTAGGAAATATTCCTTATTTGAAGGGGCCTGAAAGTGTGGTCTGGGGTACAGCAGTGACCTGTCATACTTGAGAGGATTAA

RNA-complement:

CUUAAGUCGACUCUUCUUGUCCGUUCCUGAAUCCUUUAUAAGGAAUAAACUUCCCCGGACUUUCACACCAGACCCCAUGUCGUCACUGGACAGUAUGAACUCUCCUAAUU

<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi?PAGE=3&ID=mrJOd0vFgi>

2.

Chr. 2 pos. 755-838

DNA:

GTTTACTGCCCACGCTTATCCGCCTAGTGAAACAAGCTAGAACAAAGAGAAAATCACTCTCAAAAGGAAACTTGCCGTTTTGA

RNA-complement:

CAAAUGACGGGUGCGAAUAGGCGGAUCACUUUGUUCGAUCUUGUUUCUCUUUUAGUGAGAGUUUUCCUUUGAACGGCAAAACU

<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi?PAGE=3&ID=IoeJW75ex7>

3.

Chr. 3 pos. 100-171

DNA:

GGCCAGGGAGACCAGAGACCATGTTGCCTTTGGAAGTGAGAGAAACTCTG

AGGAAATCCATGGTAAATCCT

RNA-complement:

CCGGUCCCUCUGGUCUCUGGUACAACGGAAACCUUCACUCUCUUUGAGACUCCUUUAGGUACCAUUUAGGA

<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi?PAGE=3&ID=a5jeuFNF0X>

4.

Chr. 4 pos. 13672-13755

DNA:

TCCTGAATTTTTTTCACTTTTAAAACTGCAGGATTGTGAGGGTTGTTT

TTAATTTTTTAATTTTTTTCCAATAATTTGAAATCAAA

RNA-complement:

AGGACUUAAAAAAAGUGAAAAUUUUGACGUCCUAACACUCCCAACAAAAAUUAAAAAAUUAAAAAAAGGUUAUUAAACUUUAGUUU

<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi?PAGE=3&ID=0Aifw7uPBO>

5.

Chr. 17 pos. 11400-11480

DNA:

GGAGCCAATTGCCCATCAGTTCTCTAGGACAGTGCTGAAGAGTGAGTAAA

TCAGTGAAGGAATGCTATCCTTGGATCATT

RNA-complement:

CCUCGGUUAACGGGUAGUCAAGAGAUCCUGUCACGACUUCUCACUCAUUUAGUCACUUCCUUACGAUAGGAACCUAGUAA

<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi?PAGE=3&ID=4MQbOQn0Kz>

**Task 7:**

Filled text:

For GSEA a **running sum** statistic has to be computed for each **Category C**, representing whether the genes of C are accumulated: on top of the **sorted** list, on **bottom** of the sorted list or if they are **randomly** distributed. Given a set of m genes of which l belong to C, the sorted list is processed from **top to bottom**. Whenever a gene of C is found, the running sum is **increased** by a certain amount, otherwise it is **decreased**.