**CpGPNP MANUAL**

The CpGPNP is a program that predicts CpG islands and designs primer sequences based on predicted CpG island location.

The CpGPNP searches the CpG island using the sliding-window method. and Primer sequence design programs are divided into standard PCR and bisulfite PCR programs according to bisulfite treatment. A methylation specific PCR (MSP)program is also performed to identify the methylation of particular CpG sites.

CpGPNP is developed using perl script and consists of the following programs.

A. CpG Island predition

1. CpGPredictor-2.0.pl

2. graphMake-1.0.pl

B. Primer design

1. standardPCR-2.0.pl

2. standardPCRmulti-2.0.pl

3. bisulfitePCR-1.0.pl

4. bisulfitePCRmulti-2.0.pl

5. msp-2.0.pl

6. positionFinder.pl

- Install perl and R package on your computer

Perl : https://www.perl.org/get.html

R package : https://www.r-project.org/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

A. CpG Island predition

**1. CpGPredictor-2.0.pl**

[Usage]

> perl CpGPredictor-2.0.pl -f INPUTFILE[FASTA] (-l INT -w INT -m STRING -g INT -o INT) -s OUTPUTFILE

[Parameter]

-f[--file] : INPUT file (FASTA format)

-l[--searchlength] : Length of CpG island, [default 500]

-w[--slide] : moving window, [default 1]

-m[--motif] : motif which user searches

-g[--GC] : Threshold of GC contents percent, [default 55]

-o[--OBEX] : Threshold of Observed /Expected CpG Ratio, [default 0.65]

-fa : make FASTA file, [y or yes]

-s[--save] : OUTPUT file (.txt)

[Example]

For Gardiner-Garden and Frommer algorithm (1987),

> perl CpGPredictor-2.0.pl -f test.fa -l 200 -g 50 -o 0.6 -s test\_result

\* If you want to get a fasta file, type 'y' or 'yes' with the [-fa] option.

[Result]

.rawdata : all detected CpG islands results (1st result)

.groups : the result grouped from redundant CpG islands of .rawdata (2nd result)

.contig : CpG island results merged between the overlapped groups from .groups (Final result)

.summary : result and parameter summary file

.graph : Files to make graphs using R

.contig result example:

No. START LAST Length GC\_Contents obs/exp\_CpG\_ratio SEQEUNCE

1 10327 11348 1022 70.1565557729941 0.954322620807719 taacccctaaccctaaccctaaccctaccctaaccctaaccctaaccctaacc

2 16469 16668 200 51.5 0.60790273556231 ACCAAGTAGAACAAGATATTTGAAATGGAAACTATTCAAAAAATTGAGAA

3 28625 29807 1183 70.0760777683855 0.841022056539298 ATTCTTTTTAAGTGACAAAACTTGTACATGTGTATCGCTC

4 51434 51995 562 54.8042704626335 0.866281310211946 GAATCACACGTATTGGAAAACCAGCGGAAGAGTAAGTCT

5 121169 121614 446 51.3452914798206 0.646969002901206 GTGCAAGTAAAGAGCCTTACTGCTGATGAGGTTTGAGG

6 134071 134376 306 52.6143790849673 0.666874221668742 ttttttttttttttttttttttttgagaccgagtcttgctctgtcgcccaggctggagtgcagtg

**2. graphMake-1.0.pl**

[Usage]

>perl graphMake-1.0.pl test.graph

[Result]

.pdf : graph visualization pdf file

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

B. Primer design

**1. standardPCR-2.0.pl**

[Usage]

>perl standardPCR-2.0.pl

[Parameter]

- FASTA file : fasta sequence file

- START position : start of target region

- END position : end of target region

- PRIMER LENGTH (minimum, maximum) : set minimum/maximum primer length

- Tm threshold

[Result]

.stdPrimer

.stdPrimer result example:

Category Forward/Reverse FirstPosition LastPosition primer Tm GC(%) PrimerLength 5'Position 3'Position hair-pin\_structure product\_size

primer 1 Forward 150 500 5'-CCTCTCAGCCTTTGAAAGAAA-3' 50.2496052226556 42.8571428571429 21 129 150 ...<.............>...

primer 1 Reverse 150 500 5'-AACAGGAGGAGCAGAGAGCGAAGCG-3' 62.2166134462012 60 25 609 584 ..<........<.>........>.. 434

primer 2 Forward 150 500 5'-AGCCTTTGAAAGAAAGAAAGGG-3' 52.0199749134786 40.9090909090909 22 135 157 ..<................>..

primer 2 Reverse 150 500 5'-AACAGGAGGAGCAGAGAGCGAAGCG-3' 62.2166134462012 60 25 609 584 ..<........<.>........>.. 427

primer 3 Forward 150 500 5'-AGCCTTTGAAAGAAAGAAAGGG-3' 52.0199749134786 40.9090909090909 22 135 157 ..<................>..

primer 3 Reverse 150 500 5'-AACAGGAGGAGCAGAGAGCGAAGCG-3' 62.2166134462012 60 25 609 584 ..<........<.>........>.. 427

**2. standardPCRmulti-2.0.pl**

[Usage]

>perl standardPCRmulti-2.0.pl -p [POSITIONFILE] (-minlen [int] -maxlen [int] -mintm[int] -maxtm [int] -opt [int]) -r [FASTAFILE] -o [OUTPUTFILE]

[Parameter]

-p[--position] : the POSITION file(.groups, .contigs) resulted from CpGPredictor-2.0.pl

-minlen[--minlength] : minimum primer length, [default 20]

-maxlen[--maxlength] : maximum primer length, [default 30]

-mintm[--mintm] : the minimum Tm, [default 50]

-maxtm[--maxtm] : the maximum Tm, [default 65]

-opt[--optTm] : the optimum Tm, [default 55]

-r[--reference] : the REFERECE FASTA file

-o[--output] : the OUTPUT file

\* The POSITION file can use the result file(.groups or .contig) obtained from CpGPrediction.pl, or user can manually make the POSITION file as follows

start end

54 523

39 540

46 1707

**3. bisulfitePCR-1.0**

[Usage]

>perl bisulfitePCR-1.0.pl

[Parameter]

- FASTA file : fasta sequence file

- START position : start of target region

- END position : end of target region

- PRODUCT SIZE : target size. It determines the direction of the primer design

- PRIMER LENGTH (minimum, maximum) : set minimum/maximum primer length

- Tm threshold

[Result]

.bisPrimer

.bisPrimer result example:

Category Forward/Reverse Original\_Sequence Primer Tm GC(%) PrimerLength 5'Position 3'Position Enthalpy Energy Salt\_concent(M). hair-pin\_structure product\_size

primer 1 Forward CGAGAAAGAAGGTGAGAAAGACAGAGC 5'-YGAGAAAGAAGGTGAGAAAGATAGAGT-3' 55.9281150879021 40.7407407407407 26 68 95 205.6 37.15 0.05 .....<......<.>......>.....

primer 1 Reverse CCCCACGGGAACCGCCCGTGC 5'-ACACRAACRATTCCCRTAAAA-3' 54.4994538670536 42.8571428571429 20 360 339 178.5 32.7 0.05 .........<.>......... 244

primer 2 Forward CGAGAAAGAAGGTGAGAAAGACAGAGC 5'-YGAGAAAGAAGGTGAGAAAGATAGAGT-3' 55.9281150879021 40.7407407407407 26 68 95 205.6 37.15 0.05 .....<......<.>......>.....

primer 2 Reverse CCCCACGGGAACCGCCCGTGC 5'-ACACRAACRATTCCCRTAAAA-3' 54.4994538670536 42.8571428571429 20 360 339 178.5 32.7 0.05 .........<.>......... 244

primer 3 Forward CGAGAAAGAAGGTGAGAAAGACAGAGC 5'-YGAGAAAGAAGGTGAGAAAGATAGAGT-3' 55.9281150879021 40.7407407407407 26 68 95 205.6 37.15 0.05 .....<......<.>......>.....

primer 3 Reverse CGCCGTAAGGTGCCGCAGTCCCGAATG 5'-CATTCRAAACTACRACACCTTACRACR-3' 59.2409703587063 48.1481481481481 26 252 225 215.2 40.25 0.05 ..<...<<..<.....>..>>...>.. 130

**4. bisulfitePCRmulti-2.0**

[Usage]

>perl bisulfitePCRmulti-2.0 -p [POSITIONFILE] (-minlen [int] -maxlen [int] -mintm[int] -maxtm [int] -t [int]) -r [FASTAFILE] -o [OUTPUTFILE]

[Parameter]

-p[--position] : the POSITION file(.groups, .contigs) resulted from CpGPredictor-2.0.pl

-minlen[--minlength] : minimum primer length, [default 20]

-maxlen[--maxlength] : maximum primer length, [default 30]

-mintm[--mintm] : the minimum Tm, [default 50]

-maxtm[--maxtm] : the maximum Tm, [default 65]

-t[--threshold] : the threshold of product size, [default 300]

-r[--reference] : the REFERECE FASTA file

-o[--output] : the OUTPUT file

\* The POSITION file is the same as that used in standardPCR-2.0.pl

**5. msp-2.0.pl**

[Usage]

>perl msp-2.0.pl -f [FASTAFILE] -s [int] -e [int] -a [int] -o OUTPUTFILE

[Parameter]

-f[--fasta] : Input your fasta file

-s[--start] : Start Position where you want to design MSP in your FASTA file

-e[--end] : End Position where you want to design MSP in your FASTA file

-a[--amplicon] : amplicon size (default 100)

-o[--output] : the output file

\* If you do not specify start and end position, map to the entire fasta file.

[Result]

The MSP program produces two results as follows.

1) .CpGMAP

2) .msp.result

1) .CpGMAP

- CpGMAP is a file that maps CG dimer positions

Modified DNA:

GTAGGCGAGTCGGACGTCGTTCGTAGTATCGGAGAGGGCGTATTGTAAAGGCGGGTAGTAGATCGTGGAGAGTTCGGGAGCGGAGTTGGATATCGTTTCGGAGGGAAGAAATGAGGT

Map of CpG :

.....CG...CG..CG.CG..CG......CG.......CG...........CG..........CG.........CG....CG...........CG...CG..................

Position of CpG :

.....6....11...15..18...22.......30........39............52...........64..........75.....81............94....99...................

2) .msp.result

- The msp.result file shows the M and U primer set.

CpG\_position sense/antisense M/U Primer primer\_sequence Tm 5'-end position 3'-end position

30 forward M\_Primer 5-TCGGACGTCGTTCGTAGTATC-3 59.2123686846054 8 29

forward U\_Primer 5-GGTGAGTTGGATGTTGTTTGTAGTATT-3 54.1874323739344 2 29

reverse Universal(M&U) 3-RCCTAARCCAATACAARCAA-5 53.0214800943587 150 130

39 forward M\_Primer 5-GTTCGTAGTATCGGAGAGGGC-3 57.3465929748468 17 38

forward U\_Primer 5-GTTGTTTGTAGTATTGGAGAGGGT-3 53.2584006661494 14 38

reverse Universal(M&U) 3-AATACAARCAAAAAACAAAARCC-5 53.0914563859459 162 139

52 forward M\_Primer 5-TCGGAGAGGGCGTATTGTAAAGGC-3 60.9128262082993 27 51

forward U\_Primer 5-ATTGGAGAGGGTGTATTGTAAAGGT-3 54.0504502968375 26 51

reverse Universal(M&U) 3-AACAAAARCCTCAAATCRCA-5 53.9172787032532 172 152

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**positionFinder.pl**

[Usage]

>perl positionFinder.pl

[Parameter]

- FASTA file : fasta sequence file

- starting motif : Position of Starting motif

- end motif : Position of End motif