

Icahn School of Medicine at Mount Sinai LINCS Center for Drug
Toxicity Signatures

Standard Operating Procedure:
High-Throughput mRNASeq Library Construction for
3' Digital Gene Expression (DGE)

DToxS SOP Index: A-6.0

Last Revision: 12/13/2015

Written By: Magali Soumillon (Broad Institute), Tarjei Mikkelsen (Broad Institute),
Jaehee Shim (Sinai)

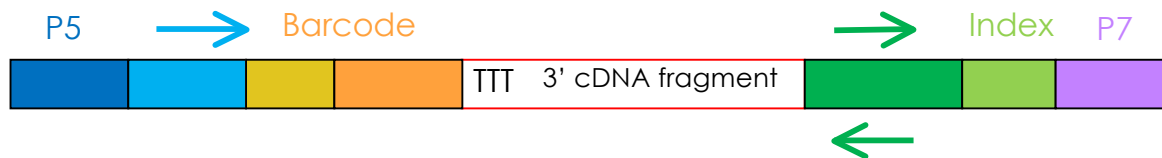
Approvals (Date): Joseph Goldfarb (12/13/2015)
Marc Birtwistle (12/13/2015)
Eric Sobie (12/13/2015)
Ravi Iyengar (12/13/2015)

Quality Assurance/Control (QA/QC) steps are indicated with **green highlight**.

Metadata recording is highlighted with **yellow highlight and superscript indices**.

This protocol was developed at the Broad Institute by Magali Soumillon and Tarjei Mikkelsen and is performed there.

0. Final library construct for RNA-DGE, for reference purposes
 - a. These libraries should be run as paired-end with a 50 bp cycle kit (see SOP A-7.0).



Barcode: Well/Cell index (N_6) + Unique Molecular Identifier (N_{10})
Index: Plate indexing
 P5: Flow Cell Sequence preceding I5 (5' end Index)
 P7: Flow Cell Sequence following I7 (3' end Index)

1) Reverse Transcription (RT) / Template Switching

- a) Primers for this step:
i) RT Primer: E3V6NEXT

5'-/5Biosg/ACACTCTTTCCCTACACGACGCTCTTCCGATCT[BC6]N
NNNNNNNNNTTTVN-3'

5Biosg: 5' biotin
V: (AGC)
N: (AGCT)

[BC6]: 6bp sample-specific barcode**

The barcodes were designed such that each barcode differs from the others by at least two nucleotides, so that a single sequencing error cannot lead to the misidentification of the barcode.

***See supplementary pages for the complete list of BC6 sequences*

(N)10: Unique Molecular Identifier (UMI)—molecule-specific sequence tag

- ii) Template Switching Primer: E5V6NEXT

5'-iCiGiCACACTCTTTCCCTACACGACGCrGrGrG-3'

iC : iso-dC
iG: iso-dG
rG: RNA G

- b) Set up the following RT/template switching mix, for each sample (e.g., one sample per well of a 96 well plate):

- i) 10 ng of total RNA
ii) 2 µL of 5x RT buffer (250 mM Tris-HCl, pH 8.3 at 25°C)
iii) 1 µL of dNTPs¹
iv) 1 µL of 10µM E5V6NEXT adapter²
v) 1 µL of barcoded, 10µM, E3V6NEXT adapter³
vi) 0.125 µL of Maxima H Minus Reverse Transcriptase⁴
vii) complete to 10 µL with Nuclease-Free Water⁵ (not DEPC-Treated)

***If total RNA is too dilute, either re-precipitate the sample and resuspend in the appropriate volume or scale up the reaction mix proportional to the volume given above.*

- c) Incubate the plate at 42°C for 1 hour 30 minutes, then at 80°C for 10 minutes

2) cDNA pooling and purification

- a) Pool all wells (samples) together
b) Add 7x volume of DNA Binding Buffer⁶ to the pooled cDNAs.
i. For example, 96 samples from Step 1 each 10 µL in volume yields ~960 µL total volume; to this one adds 0.960 mL*7=6.72 mL.

- c) Purify all pooled cDNAs through one single DNA Clean & Concentrator-5 column⁷ according to the manufacturer's protocol.
- d) Elute cDNAs in 18 µL of Nuclease-Free Water

3) Exonuclease I treatment

- a) Add 2 µL of 10X Exonuclease I reaction buffer (from NEB) and 1µL of Exonuclease I⁸ to the cDNAs
- b) Incubate the reaction at 37°C for 30 minutes, then at 80°C for 20 minutes

4) Full length cDNA amplification

- a) Full Length cDNA Amplification Primer-SINGV6

5'-/5Biosg/ACACTCTTTCCCTACACGACGC-3'

5Biosg: 5' biotin

- b) Full length cDNA is amplified by single primer PCR using the Advantage 2 PCR Enzyme System⁹

- c) Set up the PCR reaction

- i) 20 µL of cDNA from previous step (essentially the whole sample)
- ii) 5 µL of 10X Advantage 2 PCR buffer (from kit)
- iii) 1 µL of dNTPs (from kit)
- iv) 1 µL of the 10µM, SINGV6 primer¹⁰
- v) 1 µL of the Advantage 2 Polymerase Mix (from kit)
- vi) 22 µL of Nuclease-Free Water (PCR grade water, preferably not DEPC treated water)

- d) Perform PCR using the following program:

- i) 95°C 1 min
- ii) x cycles of (see below for cycle #):
 - 95°C 15 sec
 - 65°C 30 sec
 - 68°C 6 min
- iii) 72°C 10 min
- iv) 4°C storage

- e) PCR cycle numbers: x=10 cycles for 48-96 samples pooled together
x=15 cycles for < 48 samples

5) Full length cDNA purification and quantification

- a) Purify the full length cDNAs with 30 µL of Agencourt AMPure XP magnetic beads¹¹ according to the manufacturer's protocol.
- b) Elute the full length cDNAs in 12µL of Nuclease-Free Water (preferably not DEPC-treated)
- c) **QA/QC1**: quantify eluted cDNA on the Qubit 2.0 Fluorometer¹² using the dsDNA HS Assay¹³ according to the manufacturer's protocol. Optionally (**QA/QC2**), the eluted cDNA can also be checked using Bioanalyzer²⁰ DNA High Sensitivity Assay²¹ according to the manufacturer's protocol.

6) Sequencing Library Preparation

- a) Replacement of i5 Primer for 2nd PCR : P5NEXTPT5

5'-
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC
C*G*A*T*C*Ts-3'

*: phosphorothioate bond

General Structure of i7** Primer for 2nd PCR:

5' CAAGCAGAAGACGGCATACGAGAT[i7]GTCTCGTGGGCTCGG

**See supplementary pages for the complete list of i7 sequences

- b) Proceed to the tagmentation using the Nextera DNA kit¹⁴, suitable for 50 ng of input, using the manufacturer's protocol. For less than 50 ng of pooled, purified and pre-amplified cDNA, scale down all reagents and reaction volume by the input concentration, according to the manufacturer's protocol.
- c) Purify the tagmented cDNA on a single DNA Clean & Concentrator-5 column⁷ according to the manufacturer's protocol.
- d) Set up the PCR reaction
- 25 µL of cDNA from previous step c
 - 15 µL of NPM reaction mix (from the Nextera kit)
 - 5 µL of the 5 µM, Illumina i7 index primer (N70*)¹⁶
 - 5 µL of the 5 µM, P5NEXTPT5¹⁵
- e) Perform PCR:
- 72°C 3 min
 - 98°C 30 sec
 - x cycles of (see below for cycle #):
 - 98°C 10sec
 - 63°C 30 sec
 - 72°C 3min
 - 4°C storage
- f) PCR cycle numbers: x=10 cycles if the input cDNA was less than 20ng
x=5 cycles if the input cDNA was more than 20ng
***Depending on the input cDNA measured in QA/QC1 step (prior to tagmentation), PCR cycle number is determined.*

7) Sequencing Library Purification and Size Selection

- a) Purify the sequencing library with 30 µL of Agencourt AMPure XP¹¹ magnetic beads according to the manufacturer's protocol and elute it in 20 µL of nuclease-free water (preferably not DEPC-treated).
- b) The entire library is run on an E-Gel EX Gel, 2%¹⁷ and the band corresponding to a size range of 350 to 800 bp (QA/QC3) is excised, purified using the QIA quick Gel Extraction Kit¹⁹ according to the manufacturer's protocol and eluted in 15 µL of nuclease-free water. Use E-gel 50bp Ladder¹⁸

8) Sequencing Library Quality Assessment

- a) **QA/QC4**: The library is quantified on the Qubit 2.0 Fluorometer¹² using the dsDNA HS Assay¹³.
- b) **QA/QC5** (optional): The quality and average size of the library are assessed by BioAnalyzer²⁰ with the DNAHigh Sensitivity kit²¹.

9) Metadata

1. dNTPs (New England Biolabs, #N0447L)
2. E5V6NEXT adapter (Eurogentec/Integrated DNA Technologies, #custom, refer to the protocol for sequences/modification)
3. E3V6NEXT adapter(Integrated DNA Technologies, #custom, refer to the protocol sequences/modification)
4. Maxima H Minus Reverse Transcriptase(Thermo Scientific, #EP0753)
5. Nuclease-Free Water(LifeTechnologies, #AM9937)
6. DNA Binding Buffer(Zymo Research, #D4004-1-L)
7. DNA Clean & Concentrator-5 column (Zymo Research, #D4013)
8. Exonuclease I (New England Biolabs, #M0293L)
9. Advantage 2 PCR Enzyme System (Clontech, #639206)
10. SINGV6 primer (Integrated DNA Technologies, #custom, refer to the protocol for sequences/modification)
11. AgencourtAMPure XP magnetic beads (Beckman Coulter, #A63880)
12. Qubit 2.0 Fluorometer (Life Technologies, Q32866)
13. dsDNA HS Assay (Life Technologies. #Q32851)
14. Nextera DNA kit (FC-121-1031)
15. P5NEXTPT5 (Integrated DNA Technologies, #custom, refer to the protocol for sequences/modification)
16. i7Primers(Integrated DNA Technologies,TruGrade, #custom, refer to the protocol for sequences/modification)
17. E-Gel EX Gel, 2% (Life Technologies, #G4010-02)
18. E-gel 50bp Ladder (Life Technologies, #10488-099)
19. QIAquick Gel Extraction Kit (Qiagen, #28704)
20. BioAnalyzer (Agilent, G2939AA)
21. High Sensitivitykit (Agilent, #5067- 4626)

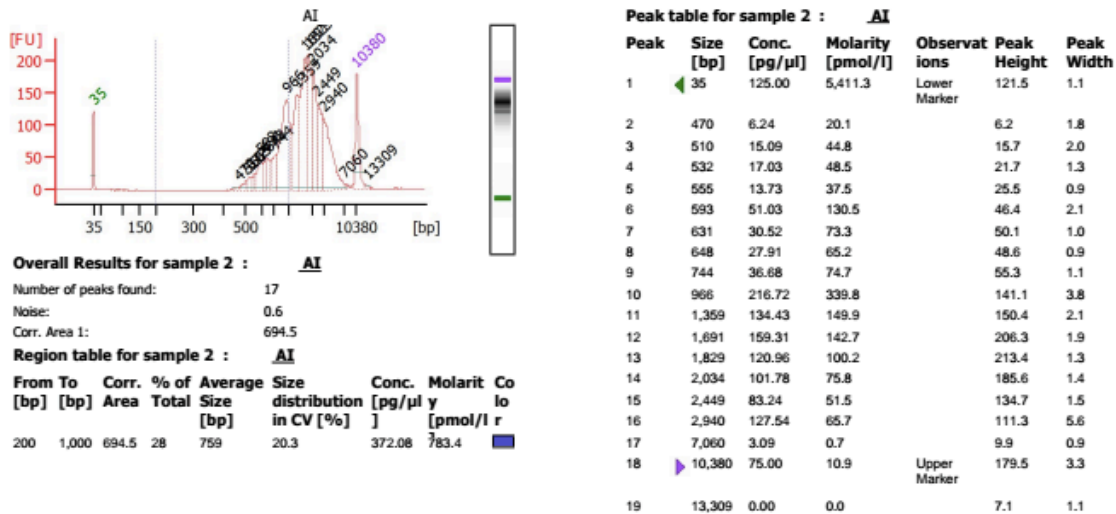
Quality Assurance/Control Steps (QA/QC)

QA/QC1: Quantify eluted cDNA from the first PCR on the Qubit 2.0 Fluorometer¹² using the dsDNA HS Assay¹³ according to the manufacturer's protocol. Passing QC1 is defined by the total amount of cDNA > 30ng.

QA/QC2 (Optional): this QC step can be supplemented with Bioanalyzer High Sensitivity Assay^{20,21} to check for the full length cDNA integrity.

A sample trace of the bioanalyzer high sensitivity assay of the cDNA for a successful prep is shown below:

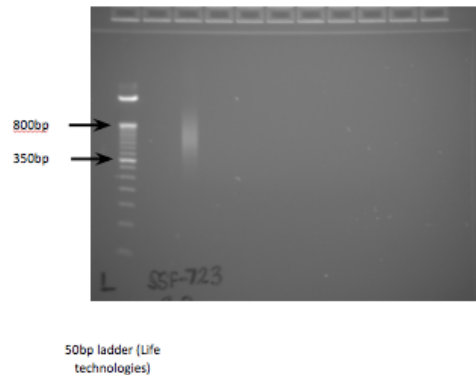
If the peak sizes corresponds to the range of mRNA transcript length, as shown below, it can be assumed that the cDNA step was successful.



QA/QC3: During the gel purification step after the 2nd PCR, the size of the library is visualized using the E-Gel EX Gel, 2%¹⁶ (manufacturer's protocol). This step can be used to check if the library cDNA length distribution is primarily from ~350bp to ~800bp.

A sample gel picture of a passing test is shown below. If the library size distribution spans across 350-800bp as it is shown, QC result is pass. If not, QC result is fail.

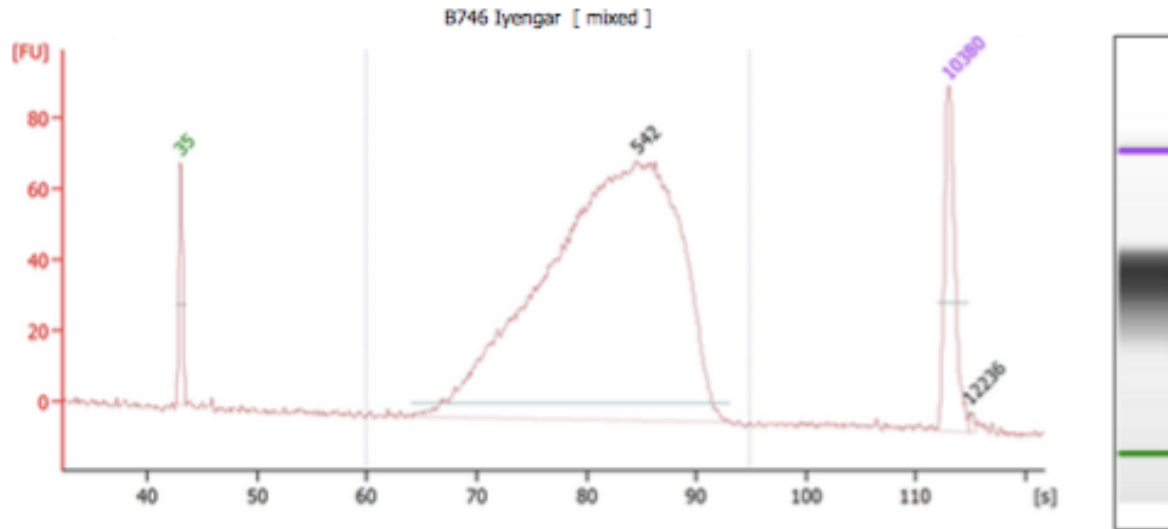
**** Do not go lower than 300bp as it will risk enriching small insert libraries and concatemers of adapters (i.e. no insert). Final library size may appear smaller than what is shown below if fragmentation step was too long or an incorrect amount of input cDNA (too little) was used.**



QA/QC4: The library is quantified on the Qubit 2.0 Fluorometer¹² using the dsDNA HS Assay¹³, as in **QA/QC1**

QA/QC5 (optional): The quality and library length distribution are assessed by BioAnalyzer²⁰ with the DNA High Sensitivity kit²¹.

A sample bioanalyzer trace of the final library (after 2nd PCR) is shown below. The left peak labeled 35 and the right peak labeled 10380 are lower and upper markers used as length standards, and the broad distribution between is the library, whose mode is 542 bp. If the library size (middle bump in the trace below) corresponds to the size selection of 350pb-800bp, QC result is pass. Lower than 350bp peaks mean that the size selection might have not done properly and cut into the adapter dimers. If that happened, we recommend repeating gel purification and excise the gel at the proper library size again. QC fails if the library size is wrong (peaks < 350 bp) and the sample concentration is too low to repeat the gel purification step.



Overall Results for sample 3 : B746 Iyengar

Number of peaks found: 2 Corr. Area 1: 1,387.6
 Noise: 0.7

Peak table for sample 3 : B746 Iyengar

Peak	Size [bp]	Conc. [$\mu\text{g}/\mu\text{l}$]	Molarity [pmol/l]	Observations
1	35	125.00	5,411.3	Lower Marker
2	542	1,346.46	3,763.4	
3	10,380	75.00	10.9	Upper Marker
4	12,236	0.00	0.0	

Supplementary Information for Oligo Sequences:

i7 Index Sequences

This index sequence can be used for high throughput setting (i.e. prepping library for more than 384 samples at a time)

i7 bases in adapter	i7 index name	i7 bases for entry on sample sheet
TCGCCTTA	N701	TAAGGCGA
CTAGTACG	N702	CGTACTAG
TTCTGCCT	N703	AGGCAGAA
GCTCAGGA	N704	TCCTGAGC
AGGAGTCC	N705	GGACTCCT
CATGCCTA	N706	TAGGCATG
GTAGAGAG	N707	CTCTCTAC
CCTCTCTG	N708	CAGAGAGG
AGCGTAGC	N709	GCTACGCT
CAGCCTCG	N710	CGAGGCTG
TGCCTCTT	N711	AAGAGGCA
TCCTCTAC	N712	GTAGAGGA

http://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/experiment-design/illumina-customer-sequence-letter.pdf

Barcode Sequences:

BC1	AAAAC
BC2	AAAATC
BC3	AAAGTT
BC4	AAATAC
BC5	AAATTG
BC6	AACAAT
BC7	AAGATT
BC8	AAGTAT
BC9	AATACA
BC10	AATAGT
BC11	AATCTT
BC12	AATGAT
BC13	AATTCT
BC14	AATTGA
BC15	ACAATA
BC16	ACATAA

BC17	ACTTAT
BC18	ACTTTA
BC19	AGATTA
BC20	AGTAAT
BC21	ATAAAC
BC22	ATAACA
BC23	ATACAA
BC24	ATACTT
BC25	AAACAT
BC26	AAACTA
BC27	AAATCA
BC28	AAATGT
BC29	AACATA
BC30	AACTAA
BC31	AAGTTA
BC32	AATAAC
BC33	AATATG
BC34	AATCAA
BC35	AATGTA
BC36	AATTAG
BC37	AATTTT
BC38	ACAAAT
BC39	ACTAAA
BC40	ACTATT
BC41	AGAATT
BC42	AGATAT
BC43	AGTATA
BC44	AGTTAA
BC45	ATAAGT
BC46	ATAATG
BC47	ATAGAT
BC48	ATAGTA
BC49	ATATAG
BC50	ATATCT
BC51	ATCAAA
BC52	ATCATT
BC53	ATGAAT
BC54	ATGATA
BC55	ATTACT

BC56	ATTAGA
BC57	ATTCTA
BC58	ATTGAA
BC59	ATTTCA
BC60	ATTTGT
BC61	CAAATA
BC62	CAATAA
BC63	CATTAT
BC64	CATTTA
BC65	CTATAT
BC66	CTATTA
BC67	CTTTAA
BC68	GAAATT
BC69	GATAAT
BC70	GATATA
BC71	GTAATA
BC72	GTATAA
BC73	ATATGA
BC74	ATATTC
BC75	ATCTAT
BC76	ATCTTA
BC77	ATGTAA
BC78	ATTAAG
BC79	ATTATC
BC80	ATTCAT
BC81	ATTGTT
BC82	ATTTAC
BC83	ATTTTG
BC84	CAAAAT
BC85	CATAAA
BC86	CATATT
BC87	CTAAAA
BC88	CTAATT
BC89	CTTAAT
BC90	CTTATA
BC91	GAATAT
BC92	GAATTA
BC93	GATTAA
BC94	GTAAAT

BC95	GTAAAA
BC96	GTTATT
BC97	GTTTAT
BC98	GTTTTA
BC99	TAAAGT
BC100	TAAATG
BC101	TAAGAT
BC102	TAAGTA
BC103	TAATGA
BC104	TAATTC
BC105	TACTAT
BC106	TACTTA
BC107	TAGTAA
BC108	TAGTTT
BC109	TATAGA
BC110	TATATC
BC111	TATGAA
BC112	TATGTT
BC113	TATTGT
BC114	TATTTG
BC115	TCATAT
BC116	TCATTA
BC117	TCTTAA
BC118	TGAAAT
BC119	TGATTT
BC120	TGTAAA
BC121	TAAAAC
BC122	TAAACA
BC123	TAACAA
BC124	TAACTT
BC125	TAATAG
BC126	TAATCT
BC127	TACAAA
BC128	TACATT
BC129	TAGAAT
BC130	TAGATA
BC131	TATAAG
BC132	TATACT
BC133	TATCAT

BC134	TATCTA
BC135	TATTAC
BC136	TATTCA
BC137	TCAAAA
BC138	TCAATT
BC139	TCTAAT
BC140	TCTATA
BC141	TGAATA
BC142	TGATAA
BC143	TGTATT
BC144	TGTTAT
BC145	TGTTTA
BC146	TTAAAG
BC147	TTAATC
BC148	TTACAT
BC149	TTAGTT
BC150	TTATAC
BC151	TTATTG
BC152	TTCAAT
BC153	TTGAAA
BC154	TTGATT
BC155	TTTACA
BC156	TTTAGT
BC157	TTTCTT
BC158	TTTGTA
BC159	TTTTGA
BC160	TCTTTC
BC161	AGACCT
BC162	AGGGAT
BC163	CACCAA
BC164	CAGTCA
BC165	CCACAT
BC166	CCGATT
BC167	CTAGTG
BC168	CTTCTG
BC169	TTAACT
BC170	TTAAGA
BC171	TTACTA
BC172	TTAGAA

BC173	TTATCA
BC174	TTATGT
BC175	TTCATA
BC176	TTCTAA
BC177	TTGTTA
BC178	TTTAAC
BC179	TTTATG
BC180	TTTCAA
BC181	TTTTAG
BC182	TTTTCT
BC183	TTGGAT
BC184	ACCGTA
BC185	ATCGAG
BC186	CAAGCT
BC187	CATCAG
BC188	CATGGT
BC189	CGACTT
BC190	CGATTG
BC191	GAAGAC
BC192	GATCGT
BC193	GCTAGA
BC194	GCTTAC
BC195	GGGATT
BC196	GTACAC
BC197	GTTCGA
BC198	TAGTGG
BC199	TCTGCA
BC200	TTCCTC
BC201	CCAACC
BC202	CCTTCC
BC203	GTACCG
BC204	ACCCCC
BC205	ACCGGC
BC206	ACGCCG
BC207	ACGGGG
BC208	AGCCCG
BC209	AGCGGG
BC210	AGGCCC
BC211	AGGGGC

BC212	CACCCC
BC213	CACGGC
BC214	CAGCCG
BC215	CAGGGG
BC216	CCACCG
BC217	GGACAT
BC218	GGCAAT
BC219	GTCAAG
BC220	GTGACT
BC221	TCCAAC
BC222	TCGAAG
BC223	TTGTCC
BC224	TTTGGC
BC225	CTCTCC
BC226	GGACCA
BC227	ACCCGG
BC228	ACCGCG
BC229	ACGCGC
BC230	ACGGCC
BC231	AGCCGC
BC232	AGCGCC
BC233	AGGCGG
BC234	AGGGCG
BC235	CACCGG
BC236	CACGCG
BC237	CAGCGC
BC238	CAGGCC
BC239	CCACGC
BC240	CCAGGG
BC241	CCCACG
BC242	CCCAGC
BC243	CCCCGT
BC244	CCCCTG
BC245	CCCTGG
BC246	CCGAGG
BC247	CCGGAC
BC248	CCGGCA
BC249	CCGTGC
BC250	CCGTGC

BC251	CCTGGC
BC252	CGACCC
BC253	CGAGGC
BC254	CGCACC
BC255	CGCCCT
BC256	CGCCGA
BC257	CGCGCA
BC258	CGCGGT
BC259	CGCTGC
BC260	CGGACG
BC261	CGGCCA
BC262	CGGCGT
BC263	CGGGCT
BC264	CGGGGA
BC265	CCCCAC
BC266	CCCCCA
BC267	CCCGAG
BC268	CCCGGA
BC269	CCGCAG
BC270	CCGCGA
BC271	CCGGGT
BC272	CCGGTG
BC273	CCTCGG
BC274	CCTGCG
BC275	CGACGG
BC276	CGAGCG
BC277	CGCAGG
BC278	CGCCAG
BC279	CGCCTC
BC280	CGCGAC
BC281	CGCGTG
BC282	CGCTCG
BC283	CGGAGC
BC284	CGGCAC
BC285	CGGCTG
BC286	CGGGAG
BC287	CGGGTC
BC288	CGGTCC
BC289	CGGTGG

BC290	CGTCCG
BC291	CGTGGG
BC292	CTCCCG
BC293	CTGCGG
BC294	CTGGCG
BC295	GACCGC
BC296	GACGCC
BC297	GAGCGG
BC298	GAGGCG
BC299	GCACGG
BC300	GCAGCG
BC301	GCCAGG
BC302	GCCCAG
BC303	GCCCTC
BC304	GCCGAC
BC305	GCCGTG
BC306	GCCTCG
BC307	GCGAGC
BC308	GCGCAC
BC309	GCGCTG
BC310	GCGGAG
BC311	GCGGTC
BC312	GCGTCC
BC313	CGTCGC
BC314	CGTGCC
BC315	CTCCGC
BC316	CTCGGG
BC317	CTGGGC
BC318	GACCCG
BC319	GACGGG
BC320	GAGCCC
BC321	GAGGGC
BC322	GCACCC
BC323	GCAGGC
BC324	GCCACC
BC325	GCCCCT
BC326	GCCCGA
BC327	GCCGCA
BC328	GCCGGT

BC329	GCCTGC
BC330	GCGACG
BC331	GCGCCA
BC332	GCGCGT
BC333	GCGGCT
BC334	GCGGGA
BC335	GCGTGG
BC336	GCTCCG
BC337	GCTCGC
BC338	GCTGCC
BC339	GGAGCC
BC340	GGAGGG
BC341	GGCCAC
BC342	GGCGAG
BC343	GGCGTC
BC344	GGCTCC
BC345	GGGCAG
BC346	GGGCCT
BC347	GGGGAC
BC348	GGGGCA
BC349	GGGTCG
BC350	GGGTGC
BC351	GGTGGC
BC352	GTCCCC
BC353	GTGCGC
BC354	GTGGCC
BC355	TCCCGC
BC356	TCCGGG
BC357	TCGGGC
BC358	TGCCCC
BC359	TGGCCG
BC360	TGGCGC
BC361	GCTGGG
BC362	GGACGC
BC363	GGCACG
BC364	GGCAGC
BC365	GGCGCT
BC366	GGCGGA
BC367	GGGACC

BC368	GGGAGG
BC369	GGGCGA
BC370	GGGCTC
BC371	GGGGGT
BC372	GGGGTG
BC373	GGTCCC
BC374	GGTGCG
BC375	GTCGCG
BC376	GTCGGC
BC377	GTGGGG
BC378	TCCCCG
BC379	TCGCGG
BC380	TCGGCG
BC381	TGCGCG
BC382	TGCGGC
BC383	TGGGCC
BC384	TGGGGG