



# Feature Map

<u>Name</u>	<u>Start</u>	<u>End</u>	
Histone H3.3 K27M ORF*	2488	2949	
Amp Resistance	181	4745	Reverse
Bacterial Rep. Origin	360	1163	
pTF3	1640	1665	
Txn Terminator	1760	1336	Reverse
CMV Promoter	1813	2413	
T7 Promoter	2446	2462	
BGH polyA Site	2959	3081	
SV40 Rep. Origin	3193	3536	
Neo Resistance	3600	4394	
SV40 polyA Site	4566	4696	

\* includes DDDDK tag and spacer (see below)

## Expressed Protein Sequence (Histone H3.3 K27M)

The amino acid sequence of DDDDK epitope tag is underlined and in bold. The amino acids in blue are not part of the H3.3 sequence but added to provide spacing for the epitope tag, as described in:

Lewis *et al* (2013) Science **340**: 857-861.

MARTKQTARKSTGGKAPRKQLATKAAR **M**SAPSTGGVKK  
PHRYRPGTVLREIRRYQKSTELLIRKLPFQRLVREIAQDF  
KTDLRFQSAAIGALQEASEAYLVGLFEDTNLCAIHAKRVT  
IMPKDIQLARRIRGERA **AAAGG****DYKDDDDKS****AAAG**

# pEPI-H3.3K27M Vector DNA Sequence

```
1  CGGGGCGAAA  ACTCTCAAGG  ATCTTACCGC  TGTTGAGATC  CAGTTCGATG  TAACCCACTC
61  GTGCACCCAA  CTGATCTTCA  GCATCTTTTA  CTTTCACCAG  CGTTTCTGGG  TGAGCAAAAA
121  CAGGAAGGCA  AAATGCCGCA  AAAAAGGGAA  TAAGGGCGAC  ACGGAAATGT  TGAATACTCA
181  TATTCTTCCT  TTTTCAATAT  TATTGAAGCA  TTTATCAGGG  TTATTGTCTC  ATGAGCGGAT
241  ACATATTTGA  ATGTATTTAG  AAAAATAAAC  AAATAGGGGT  CAGTGTTACA  ACCAATTAAC
301  CAATTCTGAA  CATTATCGCG  AGCCCATTTA  TACCTGAATA  TGGCTCATAA  CACCCCTTGC
361  TCATGACCAA  AATCCCTTAA  CGTGAGTTAC  GCGCGCGTCG  TTCCACTGAG  CGTCAGACCC
421  CGTAGAAAAG  ATCAAAGGAT  CTTCTTGAGA  TCCTTTTTTT  CTGCGCGTAA  TCTGCTGCTT
481  GCAAACAAAA  AAACCACCGC  TACCAGCGGT  GGTTTGTGTT  CCGGATCAAG  AGCTACCAAC
541  TCTTTTTTCC  AAGGTAAGT  GCTTCAGCAG  AGCGCAGATA  CCAAATACTG  TTCTTCTAGT
601  GTAGCCGTAG  TTAGCCACCC  ACTTCAAGAA  CTCTGTAGCA  CCGCCTACAT  ACCTCGCTCT
661  GCTAATCCTG  TTACCAGTGG  CTGCTGCCAG  TGGCGATAAG  TCGTGTCTTA  CCGGGTTGGA
721  CTCAAGACGA  TAGTTACCGG  ATAAGGCGCA  GCGGTCGGGC  TGAACGGGGG  GTTCGTGCAC
781  ACAGCCCAGC  TTGGAGCGAA  CGACCTACAC  CGAACTGAGA  TACCTACAGC  GTGAGCTATG
841  AGAAAGCGCC  ACGCTTCCCG  AAGGGAGAAA  GCGGACAGG  TATCCGGTAA  GCGGCAGGGT
901  CGGAACAGGA  GAGCGCACGA  GGGAGCTTCC  AGGGGGAAAC  GCCTGGTATC  TTTATAGTCC
961  TGTCGGGTTT  CGCCACCTCT  GACTTGAGCG  TCGATTTTTG  TGATGCTCGT  CAGGGGGGCG
1021  GAGCCTATGG  AAAAACGCCA  GCAACGCGGC  CTTTTTACGG  TTCTGGCCT  TTTGCTGGCC
1081  TTTTGCTCAC  ATGTTCTTTC  CTGCGTTATC  CCCTGATTCT  GTGGATAACC  GTATTACCGC
1141  CTTTGAGTGA  GCTGATACCG  CTCGCCGAG  CCGAACGACC  GAGCGCAGCG  AGTCAGTGAG
1201  CGAGGAAGCG  GAAGGCGAGA  GTAGGGAACT  GCCAGGCATC  AAATAAGCA  GAAGGCCCT
1261  GACGGATGGC  CTTTTTGCGT  TTCTACAAAC  TCTTTCTGTG  TTGTAAAACG  ACGGCCAGTC
1321  TTAAGCTCGG  GCCCCTGGG  CGGTTCTGAT  AACGAGTAAT  CGTTAATCCG  CAAATAACGT
1381  AAAAACCCGC  TTCGGCGGGT  TTTTTTATGG  GGGGAGTTTA  GGGAAAGAGC  ATTTGTCAGA
1441  ATATTTAAGG  GCGCCTGTCA  CTTTGCTTGA  TATATGAGAA  TTATTTAACC  TTATAAATGA
1501  GAAAAAAGCA  ACGCACTTTA  AATAAGATAC  GTTGCTTTTT  CGATTGATGA  ACACCTATAA
1561  TTAAACTATT  CATCTATTAT  TTATGATTTT  TTGTATATAC  AATATTTCTA  GTTTGTTAAA
1621  GAGAATTAAG  AAAATAAATC  TCGAAAATAA  TAAAGGGAAA  ATCAGTTTTT  GATATCAAAA
1681  TTATACATGT  CAACGATAAT  ACAAATATA  ATACAAACTA  TAAGATGTTA  TCAGTATTTA
1741  TTATCATTTA  GAATAAATTT  TGTGTCGCC  TTAATTGTGA  GCGGATAACA  ATTACGAGCT
1801  TCATGCACAG  TGGCGTTGAC  ATTGATTATT  GACTAGTTAT  TAATAGTAAT  CAATTACGGG
1861  GTCATTAGTT  CATAGCCCAT  ATATGGAGTT  CCGCGTTACA  TAACTTACGG  TAAATGGCCC
1921  GCCTGGCTGA  CCGCCCAACG  ACCCCCGCCC  ATTGACGTCA  ATAATGACGT  ATGTTCCCAT
1981  AGTAACGCCA  ATAGGGACTT  TCCATTGACG  TCAATGGGTG  GAGTATTTAC  GGTAAACTGC
2041  CCACTTGGCA  GTACATCAAG  TGTATCATAT  GCCAAGTACG  CCCCTATTG  ACGTCAATGA
2101  CGGTAAATGG  CCCGCTGGC  ATTATGCCCA  GTACATGACC  TTATGGGACT  TTCCTACTTG
2161  GCAGTACATC  TACGTATTAG  TCATCGCTAT  TACCATGGTG  ATGCGGTTTT  GGCAGTACAT
2221  CAATGGGCGT  GGATAGCGGT  TTGACTCACG  GGGATTTCCA  AGTCTCCACC  CCATTGACGT
2281  CAATGGGAGT  TTGTTTTGGC  ACCAAAATCA  ACGGGACTTT  CAAAATGTC  GTAACAACTC
2341  CGCCCATTTG  ACGCAAATGG  GCGGTAGGCG  TGTACGGTGG  GAGGTCTATA  TAAGCAGAGC
2401  TCTCTGGCTA  ACTAGAGAAC  CCACTGCTTA  CTGGCTTATC  GAAATTAATA  CGACTCACTA
2461  TAGGGGAAGC  TGGCTAGCGC  CGCCACCATG  GCCAGAACTA  AGCAGACCGC  CCGCAAATCA
2521  ACTGGAGGAA  AAGCGCCACG  AAAGCAACTC  GCTACTAAGG  CGGCTAGGAT  GTCGGCACCC
2581  AGCACCAGGG  GTGTGAAGAA  ACCTCATCGG  TACCGCCCGG  GCACGGTGGC  GCTGCGGGAG
2641  ATTAGACGCT  ACCAGAAGTC  CACTGAACTC  CTTATCAGAA  AGCTGCCATT  CCAGCGCCTG
2701  GTCAGAGAGA  TCGCCCAAGA  TTTCAAGACC  GATCTCCGGT  TTCAGAGCGC  AGCTATCGGA
2761  GCACTGCAAG  AAGCCTCGGA  AGCCTACCTG  GTGGGACTGT  TCGAGGACAC  CAACCTCTGC
```

2821 GCCATTCACG CGAAGCGCGT CACCATCATG CCGAAGGATA TCCAGTTGGC CCGGCGGATC  
2881 AGGGGAGAAC GCGCCGCAGC CGCAGGCGGG GACTACAAAG ACGATGACGA CAAATCCGCG  
2941 GCTGGCTAGC TCGAGAAAAT CAGCCTCGAC TGTGCCTTCT AGTTGCCAGC CATCTGTTGT  
3001 TTGCCCTCC CCCGTGCCTT CCTTGACCCT GGAAGGTGCC ACTCCCCTG TCCTTTCCTA  
3061 ATAAAATGAG GAAATTGCAT CACAACACTC AACCCATATCT CGGTCTATTC TTTTGATTTA  
3121 TAAGGGATTT TGCCGATTTT GGCCTATTGG TTAAAAAATG AGCTGATTTA ACAAAAATTT  
3181 AACGCGAATT AATTCTGTGG AATGTGTGTC AGTTAGGGTG TGGAAAGTCC CCAGGCTCCC  
3241 CAGCAGGCAG AAGTATGCAA AGCATGCATC TCAATTAGTC AGCAACCAGG TGTGGAAAGT  
3301 CCCCAGGCTC CCCAGCAGGC AGAAGTATGC AAAGCATGCA TCTCAATTAG TCAGCAACCA  
3361 TAGTCCC GCC CTAACCTCCG CCCATCCC GC CCCTA ACTCC GCCCAGTTCC GCCCATTCTC  
3421 CGCCCCATGG CTGACTAATT TTTTTTATTT ATGCAGAGGC CGAGGCCGCC TCTGCCTCTG  
3481 AGCTATTCCA GAAGTAGTGA GGAGGCTTTT TTGGAGGCCT AGGCTTTTGC AAAAAGCTCC  
3541 CGGGAGCTTG TATATCCATT TTCGGATCTG ATCAAGAGAC AGGATGAGGA TCGTTTCGCA  
3601 TGATTGAACA AGATGGATTG CACGCAGGTT CTCCGCCCGC TTGGGTGGAG AGGCTATTCTG  
3661 GCTATGACTG GGCACAACAG ACAATCGGCT GCTCTGATGC CGCCGTGTTC CGGCTGTCAG  
3721 CGCAGGGGCG CCCGGTTCTT TTTGTCAAGA CCGACCTGTC CGGTGCCCTG AATGA ACTGC  
3781 AGGACGAGGC AGCGCGGCTA TCGTGGCTGG CCACGACGGG CGTTCCTTGC GCAGCTGTGC  
3841 TCGACGTTGT CACTGAAGCG GGAAGGGACT GGCTGCTATT GGGCGAAGTG CCGGGGCAGG  
3901 ATCTCCTGTC ATCTCACCTT GCTCCTGCCG AGAAAGTATC CATCATGGCT GATGCAATGC  
3961 GCGGGCTGCA TACGCTTGAT CCGGCTACCT GCCCATTCGA CCACCAAGCG AAACATCGCA  
4021 TCGAGCGAGC ACGTACTCGG ATGGAAGCCG GTCTTGTGCA TCAGGATGAT CTGGACGAAG  
4081 AGCATCAGGG GCTCGCGCCA GCCGA ACTGT TCGCCAGGCT CAAGGCGCGC ATGCCCGACG  
4141 GCGAGGATCT CGTCGTGACC CATGGCGATG CCTGCTTGCC GAATATCATG GTGGAAAATG  
4201 GCCGCTTTTC TGGATTCATC GACTGTGGCC GGCTGGGTGT GCGGACCGC TATCAGGACA  
4261 TAGCGTTGGC TACCCGTGAT ATTGCTGAAG AGCTTGGCGG CGAATGGGCT GACCGCTTCC  
4321 TCGTGCTTTA CGGTATCGCC GCTCCCGATT CGCAGCGCAT CGCCTTCTAT CGCCTTCTTG  
4381 ACGAGTTCTT CTGAGCGGGA CTCTGGGGTT CGAAATGACC GACCAAGCGA CGCCCAACCT  
4441 GCCATCACGA GATTTTCGATT CCACCGCCGC CTTCTATGAA AGGTTGGGCT TCGGAATCGT  
4501 TTTCCGGGAC GCCGGCTGGA TGATCCTCCA GCGCGGGGAT CTCATGCTGG AGTTCTTCGC  
4561 CCACCCCAAC TTGTTTATTG CAGCTTATAA TGTTACAAA TAAAGCAATA GCATCACAAA  
4621 TTTCAAAAT AAAGCATTTT TTTACTGCA TTCTAGTTGT GGTTTGTCCA AACTCATCAA  
4681 TGTATCTTAT CATGTCTGTA TACCGTCGAC CTCTAGCTAG AGCTTGGCGT AATCATGGTC  
4741 ATTACCAATG CTTAATCAGT GAGGCACCTA TCTCAGCGAT CTGTCTATTT CGTTCATCCA  
4801 TAGTTGCCTG ACTCCCCGTC GTGTAGATAA CTACGATACG GGAGGGCTTA CCATCTGGCC  
4861 CCAGCGCTGC GATGATACCG CGAGAACCAC GCTCACC GGC TCCGGATTTA TCAGCAATAA  
4921 ACCAGCCAGC CGGAAGGGCC GAGCGCAGAA GTGGTCCTGC AACTTTATCC GCCTCCATCC  
4981 AGTCTATTAA TTGTTGCCGG GAAGCTAGAG TAAGTAGTTC GCCAGTTAAT AGTTTGC GCA  
5041 ACGTTGTTGC CATCGCTACA GGCATCGTGG TGTCACGCTC GTCGTTTGGT ATGGCTTCAT  
5101 TCAGCTCCGG TTCCAACGA TCAAGGCGAG TTACATGATC CCCCATGTTG TGCAAAAAG  
5161 CGGTTAGCTC CTTCCGGTCT CCGATCGTTG TCAGAAGTAA GTTGGCCGCA GTGTTATCAC  
5221 TCATGGTTAT GGCAGCACTG CATAATTCTC TTAGTGT CAT GCCATCCGTA AGATGCTTTT  
5281 CTGTGACTGG TGAGTACTCA ACCAAGTCAT TCTGAGAATA GTGTATGCGG CGACCGAGTT  
5341 GCTCTTGCCC GCGTCAATA CGGGATAATA CCGCGCCACA TAGCAGA ACT TTAAAAGTGC  
5401 TCATCATTGG AAAACGTTCT T

# RESTRICTION ENZYME SITES

<u>Enzyme</u>	<u>Restriction Site</u>	<u>Position of 5' Cut</u>
Acc65I	GGTACC	2609
Accl	GTMKAC	4699, 4706
AclI	AACGTT	5041, 5414
AfeI	AGCGCT	4865
AlwNI	CAGNNNCTG	680, 2697
Apal	GGGCCC	1333
ApaLI	GTGCAC	61, 775
Asel	ATTAAT	1840, 2445, 3189, 4987
AvaI	CYCGRG	1326, 2616, 2950, 3539
AvrII	CCTAGG	3518
BclI	TGATCA	3569
BglI	GCCNNNNNGGC	1337, 1923, 2045, 2116, 2488, 4935
BspEI	TCCGGA	4901
BspHI	TCATGA	229, 361
BsrBI	CCGCTC	235(C), 1160, 1781(C), 4341, 4395(C)
BsrDI	GCAATG	3961
BssHII	GCGCGC	391, 4125
BstBI	TTCGAA	4410
BtsI	GCAGTG	2421(C), 2761(C), 4643(C), 5215, 5235(C)
EagI	CGGCCG	3634
EcoRV	GATATC	1673, 2859
HincII	GTYRAC	1691, 1817, 4707
KasI	GGCGCC	1450, 3727
KpnI	GGTACC	2613
MscI	TGGCCA	2491, 3810
NarI	GGCGCC	1451, 3728
NcoI	CCATGG	2193, 2486, 3425, 4160
NdeI	CATATG	2067
NheI	GCTAGC	2473, 2945
NruI	TCGCGA	318
NsiI	ATGCAT	3268, 3340
PciI	ACATGT	1089, 1685
PspOMI	GGGCCC	1329
PstI	CTGCAG	3781
PvuI	CGATCG	5185
PvuII	CAGCTG	3834
RsrII	CGGWCCG	4244
SacI	GAGCTC	2401
SacII	CCGCGG	2939

Sall	GTCGAC	4705
Sapl	GCTCTTC	4072(C), 4282(C)
SmaI	CCCGGG	2618, 3541
SnaBI	TACGTA	2173
SpeI	ACTAGT	1832
SphI	GCATGC	3266, 3338, 4133
SspI	AATATT	198, 1442, 1603
XhoI	CTCGAG	2950
XmaI	CCCGGG	2616, 3539
XmnI	GAANNNTTC	3190, 5414
AgeI	ACCGGT	no cuts
AscI	GGCGGCC	no cuts
AsiSI	GCGATCGC	no cuts
BamHI	GGATCC	no cuts
BbsI	GAAGAC	no cuts
BglII	AGATCT	no cuts
BlnI	GCTNAGC	no cuts
BsaI	GGTCTC	no cuts
BsiWI	CGTACG	no cuts
BsmBI	CGTCTC	no cuts
BsrGI	TGTACA	no cuts
BstEII	GGTNACC	no cuts
BstXI	CCANNNTTGG	no cuts
Clal	ATCGAT	no cuts
EcoRI	GAATTC	no cuts
FseI	GGCCGGCC	no cuts
HindIII	AAGCTT	no cuts
HpaI	GTTAAC	no cuts
MfeI	CAATTG	no cuts
MluI	ACGCGT	no cuts
NotI	GCGGCCGC	no cuts
PacI	TTAATTA	no cuts
PmeI	GTTTAAAC	no cuts
PmlI	CACGTG	no cuts
PpuMI	RGGWCCY	no cuts
PshAI	GACNNNGTC	no cuts
PspXI	VCTCGAGB	no cuts
SanDI	GGGWCCC	no cuts
SbfI	CCTGCAGG	no cuts
SfiI	GGCCNNNNGGCC	no cuts
Swal	ATTTAAAT	no cuts
XbaI	TCTAGA	no cuts