

*Supplementary information*

## **Tissue Specific Distribution and Dynamic Changes of 5-hydroxymethylcytosine in Mammalian Genomes**

Shannon Morey Kinney\*, Hang Gyeong Chin\*, Romualdas Vaisvila, Jurate Bitinaite, Yu Zheng, Pierre-Olivier Estève, Suhua Feng<sup>2</sup>, Hume Stroud, Steven E. Jacobsen<sup>2</sup> and Sriharsa Pradhan

New England Biolabs, Inc; 240 County Road, Ipswich, Massachusetts 01938. <sup>2</sup>HHMI; Department of MCD Biology, UCLA, Los Angeles, CA 90095-1606

\*Contributed equally to this study

## **Supplementary information.**

**Supplementary Figure 1:** Optimal enzyme to DNA ratio for DNA cleavage specificity (A) Titration of MspI with 5-hydroxymethylcytosine containing synthetic DNA as shown in Figure 2A. Units of MspI enzyme used for digestion are shown on the top. (B) Quantitative PCR on MspI cleaved 5-hmC containing synthetic DNA for determination of complete cleavage. Units of enzyme used are shown at the bottom. Uncut DNA was used for normalization. (C) Quantitative PCR on MspI cleaved 5-hmC containing genomic DNA from mouse brain for determination of complete cleavage. Units of enzyme used are shown at the bottom. Uncut DNA was used for normalization. The PCR primer used are obtained from NEB (EpiMark 5-hmC and 5-mC analysis kit).

**Supplementary Figure 2:** Validation of Tet1 specific antibody. HEK293 lysate was probed with bleed specific for Tet1 enzyme. The left panel is without competitor peptide and the right panel is with. Please note the Tet1 specific band is missing in the presence of the competitor peptide.

**Supplementary Figure 3:** Quantitative PCR analysis of nine *VANGLI* loci for 5-mC and 5-hmC in human normal brain (HNB), spleen, liver, heart and HeLa DNA. All the experiments were performed in triplicate. The enzymatic modification and treatments are shown on the right. Numbers on the panels correlate with numbers in diagram from Fig 4A upper panel.

**Supplementary Figure 4:** Quantitative PCR analysis of nine *EGFR* loci for 5-mC and 5-hmC in human normal brain (HNB), spleen, liver, heart and HeLa DNA. All the experiments were performed in triplicate. The enzymatic modification and treatments are shown on the right. Numbers on the panels correlate with numbers in diagram from Fig 4B upper panel.

**Supplementary Figure 5:** Quantitative PCR analysis of EGFR primer sets 1 and 2 and two primer sets containing no CCGG site for normalization. Upper panel: Diagram of

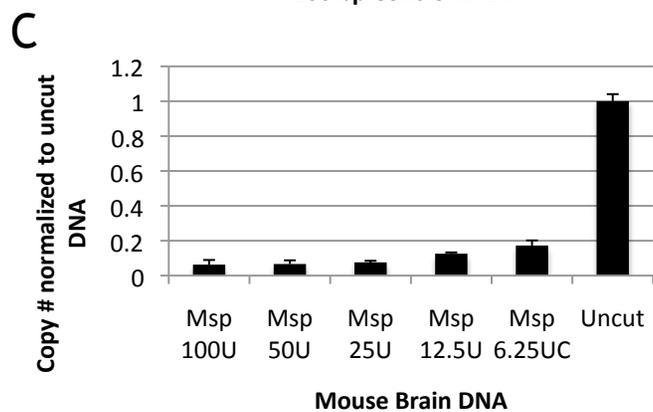
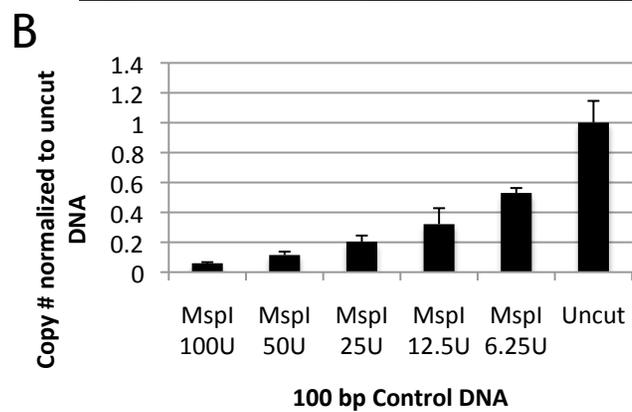
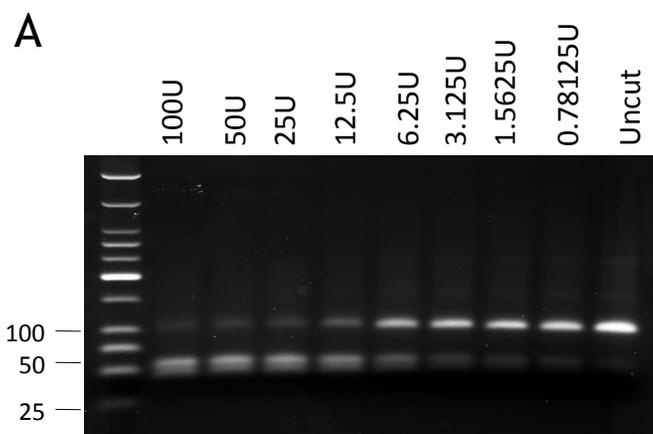
the EGFR locus with Exons as black rectangles. \* indicates the approximate location of No CCGG primer sets 1 and 2 in intron 4 of the EGFR gene. (A) Results for EGFR primer sets 1 and 2 with HNB2 sample with no normalization to No CCGG primer set values. (B) Results for EGFR primer sets 1 and 2 with HNB2 sample with normalization to No CCGG primer set 1 value for each respective sample. (C) Results for EGFR primer sets 1 and 2 with HNB2 sample with normalization to No CCGG primer set 2 values for each respective sample.

**Supplementary Table 1:** Discovering 5-hmC loci in mouse embryonic stem cell DNA. Sanger sequencing was performed using one primer only.

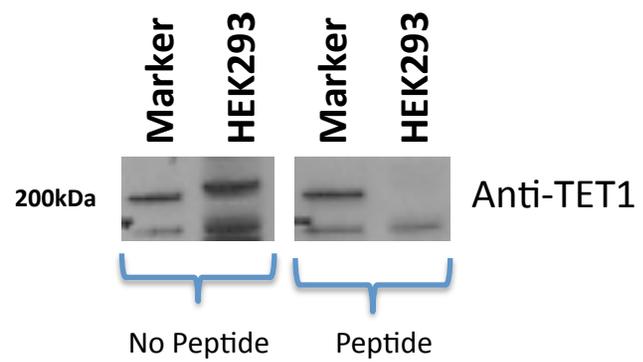
**Supplementary Table 2:** Discovering 5-hmC loci in human normal brain DNA. Sanger sequencing was performed using one primer only. Concatameric ligated products displayed more than two CCGG sites.

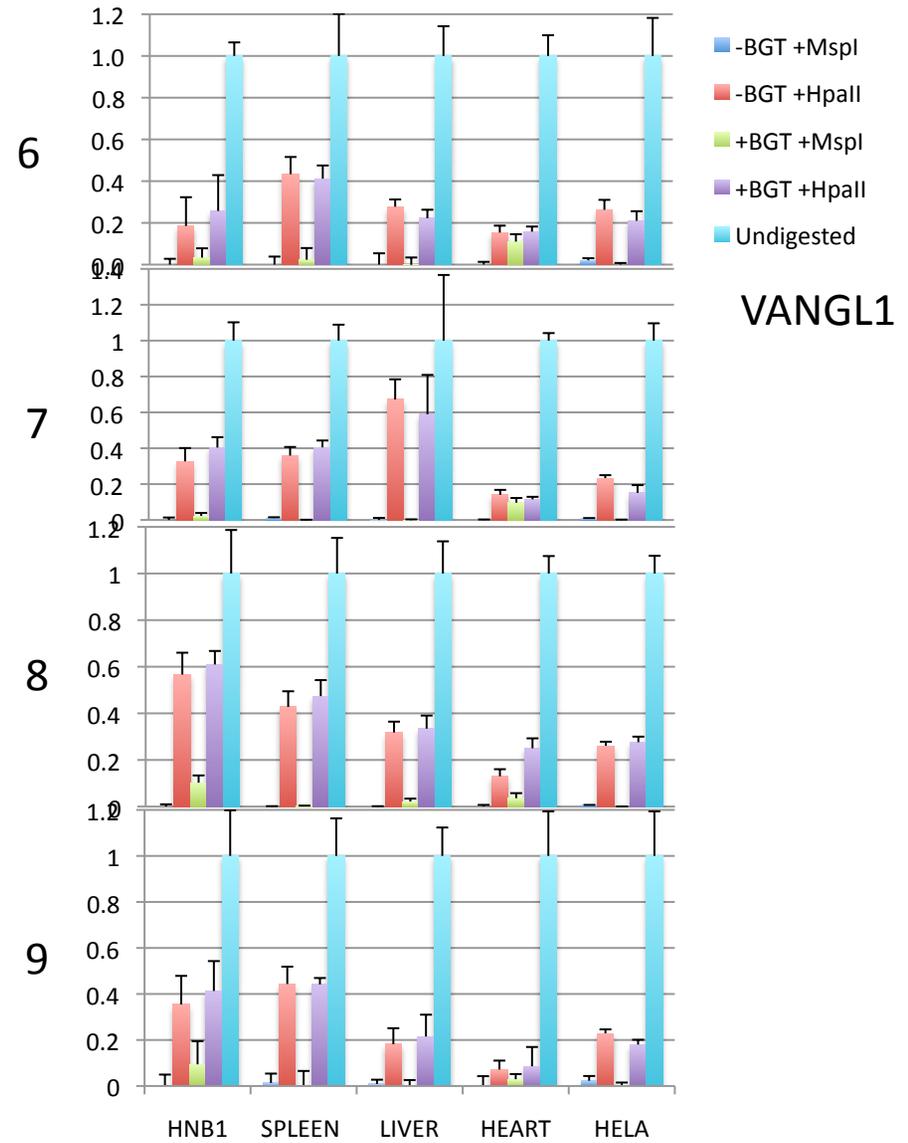
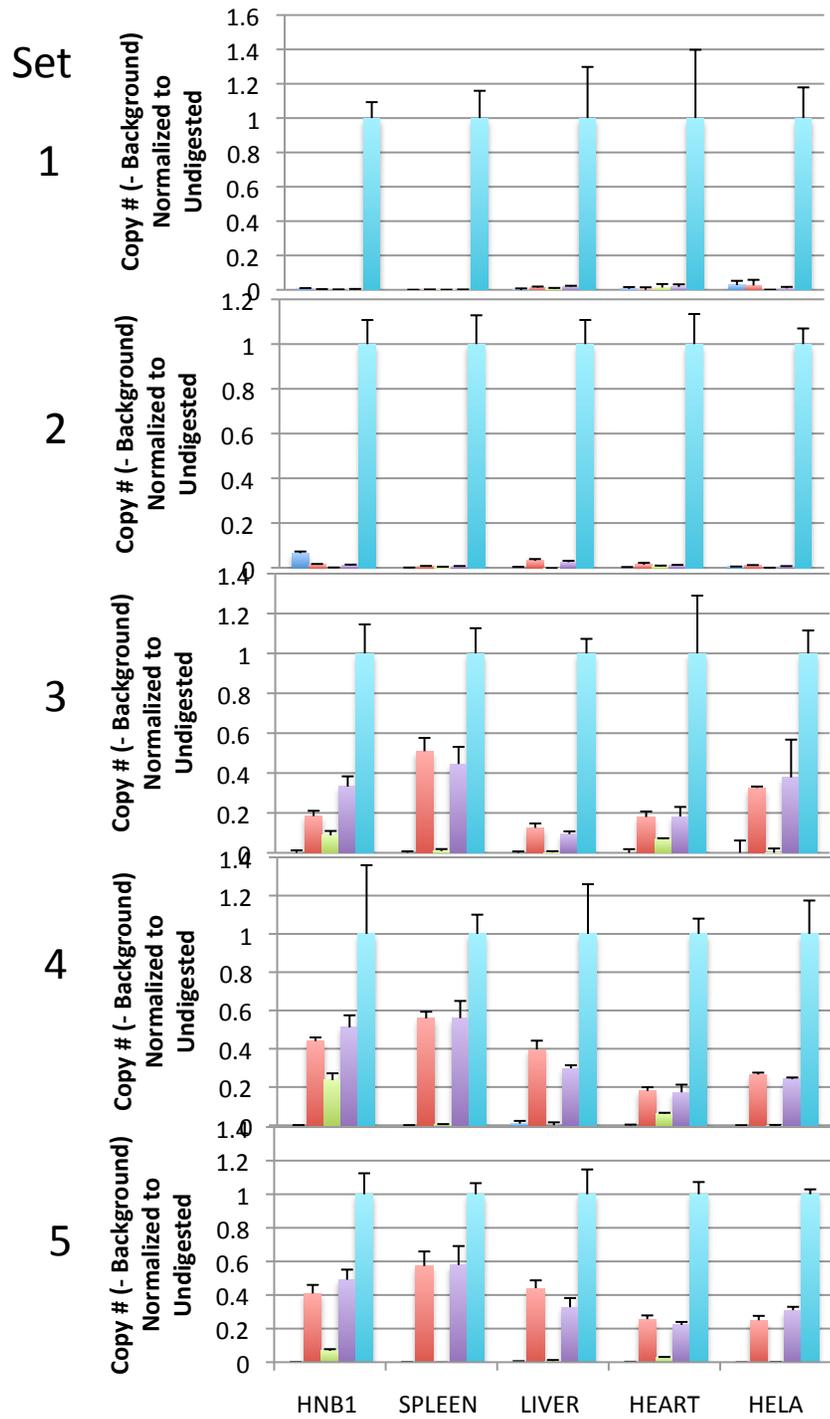
**Supplementary Table 3:** PCR primer sets for endpoint PCR and q-PCR of mouse and human CCGG loci.

# Supplementary figure 1

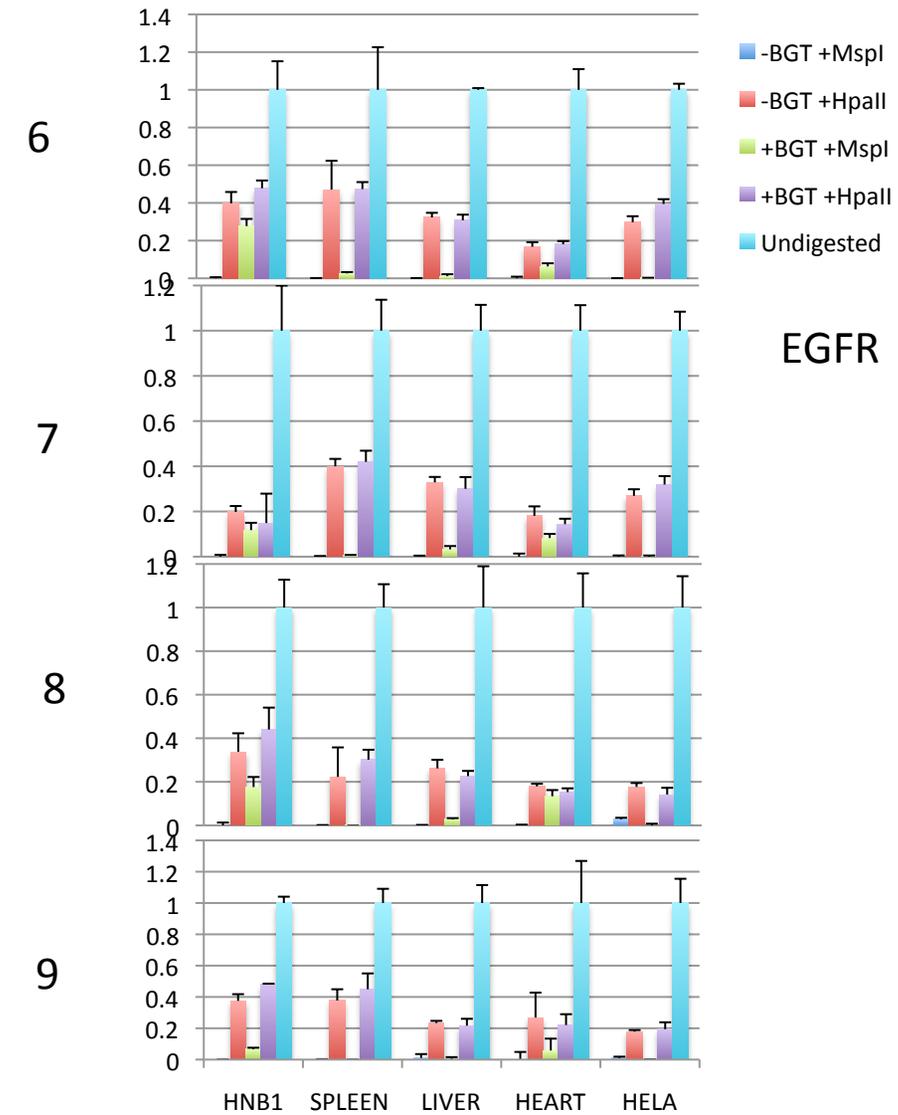
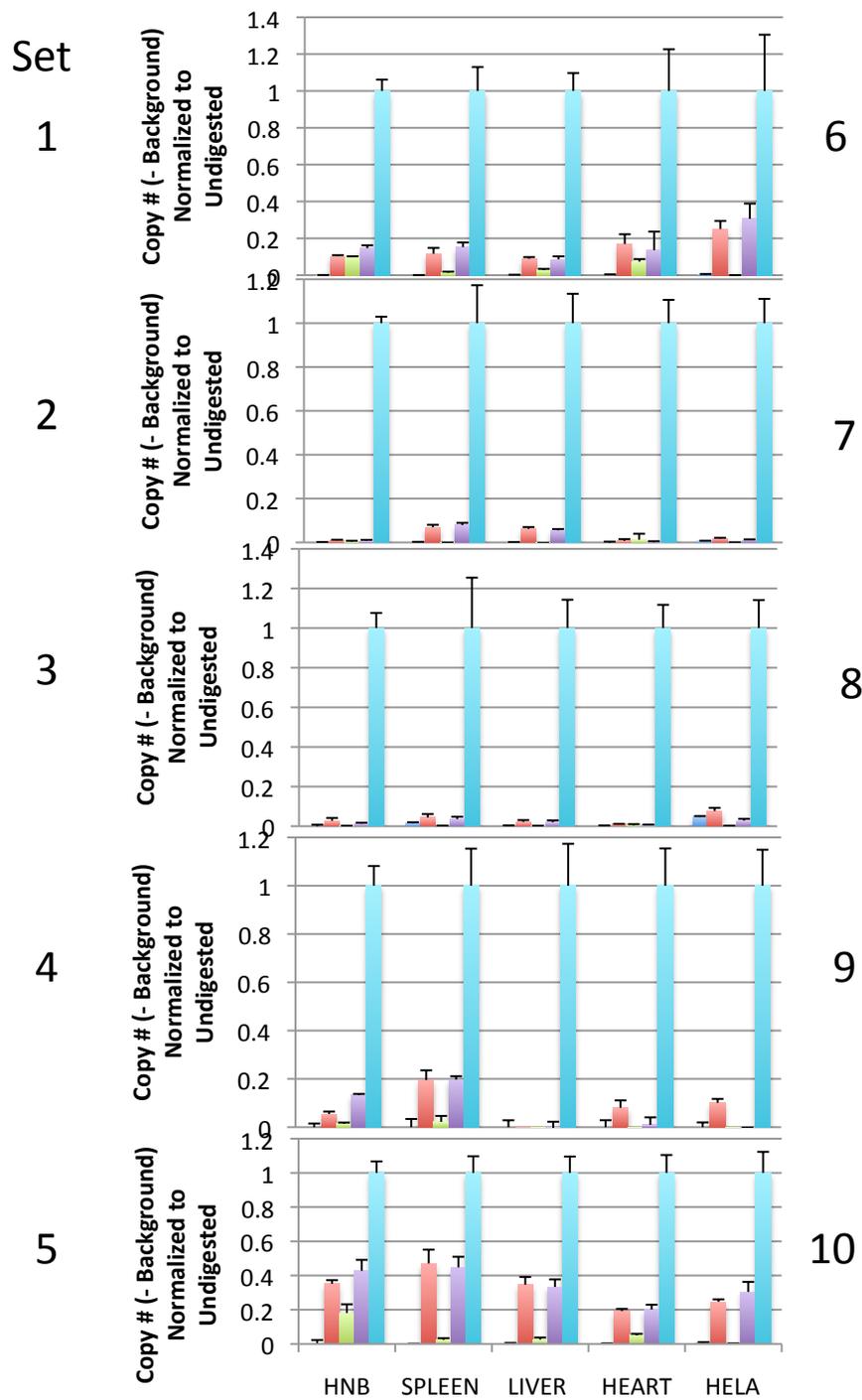


## Supplementary figure 2



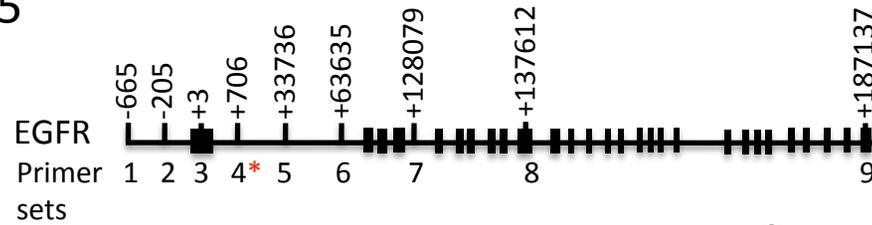


Supplementary figure 3

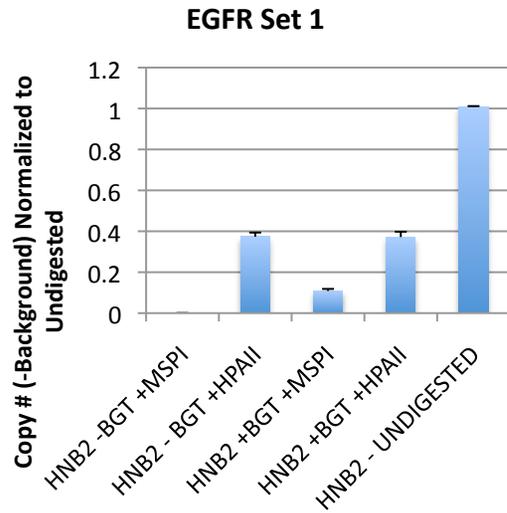


Supplementary figure 4

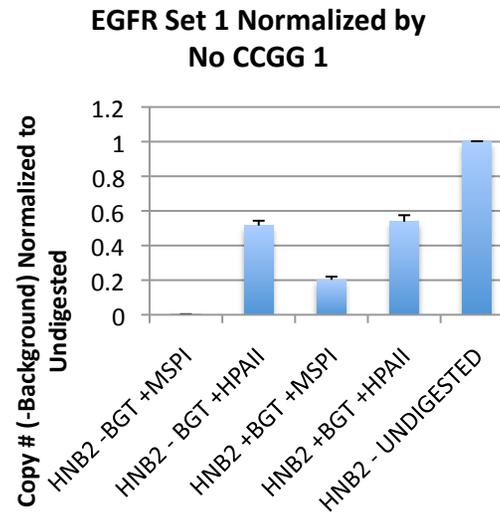
# Supplementary figure 5



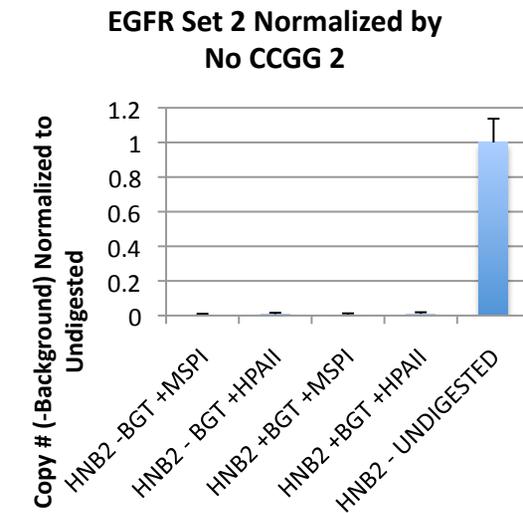
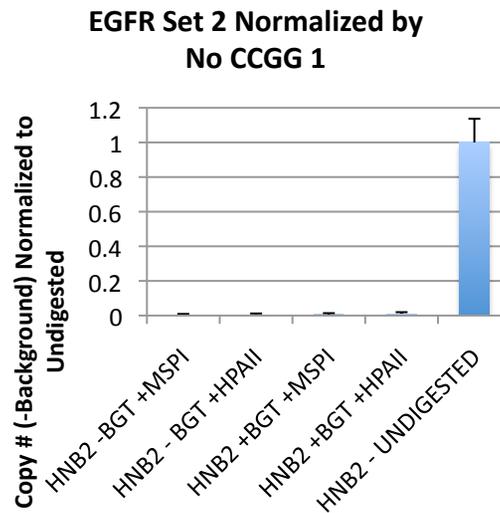
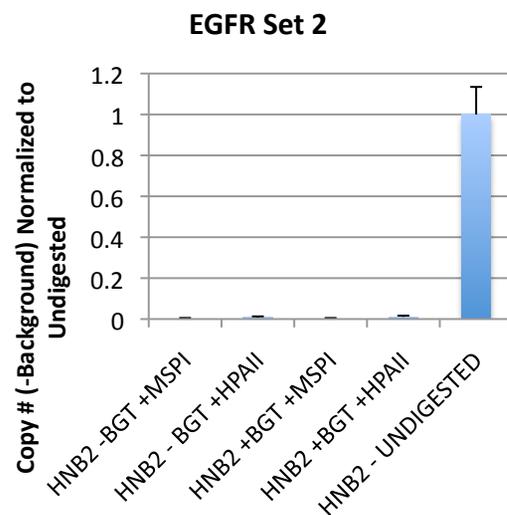
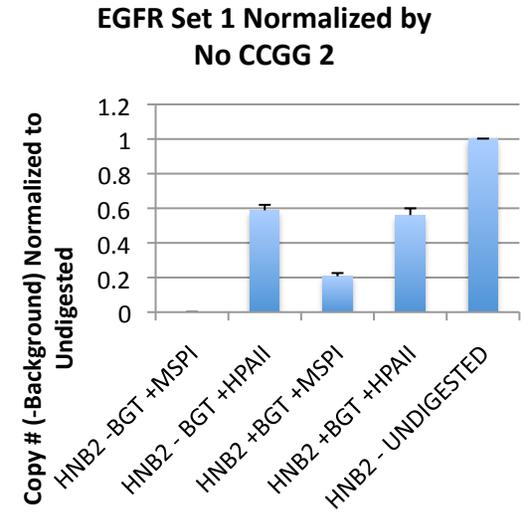
A



B



C



# of clones	# MSP1 sites	Gene	Information
2	2	Intergenic	Mouse Chromosome 10, bp 34574152
2	2	Intergenic	Mouse Chromosome 11, bp 33898156
1	2	Intergenic	Mouse Chromosome 12, bp 17432255
1	2	Cyclin F	Intronic region, bp 1073249
1	2	Intergenic	Mouse Chromosome 2, bp 12557113
1	2	Mxra7	Intronic region, bp 28228129
1	2	Lrp1	Intronic region, bp 2372508
1	2	Intergenic	Mouse Chromosome 8, bp 48334398
1	2	Intergenic	Mouse Chromosome 11, bp 25210706
1	2	Intergenic	Mouse Chromosome 5, bp 10658147
1	2	Intergenic	Mouse Chromosome 13, bp 3661176
3	1	Krt18	Intronic region, bp 63148003
2	1	Intergenic	Mouse Chromosome 13, bp 4870651
1	1	Kirrel3	Intronic region, bp 36276208
1	1	Intergenic	Mouse Chromosome 15, bp 54727587

Table 1

Table 2

# of clones	# MSPI sites	Gene	Information
3	1	Unknown	Aligns to several chromosomes
3	2	ATXN2	Ataxin 2
3	2	INPP4A	inositol polyphosphate-4-phosphatase, type I
3	2	PRKAA2	protein kinase, AMP-activated, alpha 2 catalytic subunit
3	3	RPL11P5	ribosomal protein L11 pseudogene 5
3	2	THRAP3	thyroid hormone receptor associated protein 3
3	2	TTL10	tubulin tyrosine ligase-like family, member 10
2	2	BCR	Breakpoint cluster region
2	1	DDHD1	DDHD domain containing 1
2	1	Intergenic	Human Chromosome 16, bp 574535550
2	1	Intergenic	Human Chromosome 16, bp 2037850
2	3	Intergenic	Human Chromosome 1, bp 229203550
2	2	Intergenic	Human Chromosome 5, bp 68477050
2	2	KPNA6	karyopherin alpha 6 (importin alpha 7)
2	1	SND1	staphylococcal nuclease and tudor domain containing 1
1	2	Unknown	Aligns to several chromosomes
1	2	Unknown	Homo sapiens Chromosome 1
1	2	Unknown	Aligns to several chromosomes
1	2	Unknown	Homo sapiens chromosome 1
1	2	Unknown	Aligns to several chromosomes
1	2	Unknown	Aligns to several chromosomes
1	2	Unknown	Aligns to several chromosomes
1	3	Unknown	Aligns to several chromosomes
1	2	Unknown	Aligns to several chromosomes
1	1	ATP8A2	ATPase, aminophospholipid transporter, class I, type 8A, member 2
1	1	AUTS2	AUTS2 autism susceptibility candidate 2
1	3	BANP	BTG3 associated nuclear protein
1	5	CACNA1H	calcium channel, voltage-dependent, T type, alpha 1H subunit
1	2	CHID1	chitinase domain containing 1
1	2	DPM3	dolichyl-phosphate mannosyltransferase polypeptide 3
1	2	EGFR	epidermal growth factor receptor
1	2	Intergenic	Human Chromosome 3, bp 196319050
1	2	Intergenic	Human Chromosome 9, bp 102235809
1	2	Intergenic	3kb upstream of COP9 constitutive photomorphogenic homolog subunit 6 (Arabidopsis)
1	4	Intergenic	Human Chromosome 5, bp 13600000
1	2	Intergenic	Human Chromosome 18, bp 44237300
1	1	Intergenic	Human Chromosome 3, bp 8482970
1	1	Intergenic	Human Chromosome 2, bp 95649950
1	1	LMF1	lipase maturation factor 1
1	3	NASP	nuclear autoantigenic sperm protein (histone-binding)
1	2	NT5DC2	5'-nucleotidase domain containing 2
1	1	RALGAPB	Ral GTPase activating protein, beta subunit (non-catalytic)
1	2	Repetitive element	Alu
1	2	Repetitive element	Alu
1	2	Repetitive element	Alu
1	2	Unkown	Human Chromosome 15
1	3	RTN4RL1	reticulon 4 receptor-like 1
1	2	TCF7L2	transcription factor 7-like 2 (T-cell specific, HMG-box)
1	3	VANGL1	vang-like 1 (van gogh, Drosophila)

Table 3

Genomic position/Gene	Species	Analysis Type	MSP1 site	Forward Primer sequence	Reverse Primer sequence
Chr. 10, bp 34574152 (Set#2)	M. musculus	PCR	5'	GAACAGCAGAGGGAGATAG	CAAGCCTGGAGTTAAGAGAG
Chr. 10, bp 34574152 (Set#3)	M. musculus	PCR	3'	GTGTGTTCTCCACCAAGTGT	CTTTCTTCTCCAGCATCAG
Chr. 11, bp 33898156	M. musculus	PCR	5'	GGGAAAAGACACATCATGGC	CGGCAGGATGAAGGGGAATC
Chr. 11, bp 33898156	M. musculus	PCR	3'	CTTAAGCTAGACTGCAAGACC	TCCCATCTTTTCAATGATG
Chr. 12, bp 17432255 (Set#4)	M. musculus	PCR	5'	ATACAGTGGCTTGGGAGAGG	GTGACATAGACTGAGAGGAGAC
Chr. 12, bp 17432255	M. musculus	PCR	3'	GTCTCCTCTCAGTCTATGTCAC	CCTCTCTTACCATCTTGTGCC
Cyclin F, Intron, bp 1073249	M. musculus	PCR	5'	GGTCACAGGAAGAAGAACCA	AAAGTCAGGGTAGGAGCTTG
Cyclin F, Intron, bp 1073249	M. musculus	PCR	3'	CTAGATTGCTATGGGAGGAC	TCTAAACTTGTATGGCCACAG
Chr. 2, bp 12557113	M. musculus	PCR	5'	CAGTCCCACCTATAATCGAA	GTGGATGAGCTATCTAGTGG
Chr. 2, bp 12557113	M. musculus	PCR	3'	CCCCTAGATAGCTCATCCA	CCAAGAACCCTAGAACTCA
Mxra7, Intron, bp 28228129	M. musculus	PCR	5'	GTAGGGTGGCTTGTATCTT	AGTGGGTTTCAGTGTAGAAG
Mxra7, Intron, bp 28228129	M. musculus	PCR	3'	GACAATTTCTAAGCAGCGAC	TCAACTCCCTACTCCAGT
Lrp1, Intron, bp 2372508 (Set#12)	M. musculus	PCR	5'	TCCTTACCCTGAATGACTCC	CAACCCACACTATTCCCTTG
Lrp1, Intron, bp 2372508	M. musculus	PCR	3'	CAAGGGAAATAGTGTGGGTTG	CAGCTGGATGTTTATGGTCC
Chr. 8, bp 48334398	M. musculus	PCR	5'	CCCCAAAGTTTCAGTTTATCC	GCCATCTTTCTGTTGTGGTC
Chr. 8, bp 48334398	M. musculus	PCR	5' 3'	CTCCCAAAGTTTCAGTTTATC	TTCTCCTGCTCCTGGCCAC
Chr. 11, bp 25210706	M. musculus	PCR	3'	CTAACGAGAGAGGCCAAAG	CTGGGATGATGAGGAAGTGA
Chr. 5, bp 10658147	M. musculus	PCR	5'	GCTACTGGGTTGCTTTGTC	CCTCGAGTGTTCATCAAG
Chr. 5, bp 10658147	M. musculus	PCR	3'	GAGTCACACCCCAAAGAATC	CAATTCTCCCGTACTCCCA
Chr. 13, bp 3661176	M. musculus	PCR	5'	CTTTTTCTGTGTAGCATCT	CCTTATTCTGGGCAACT
Chr. 13, bp 3661176	M. musculus	PCR	3'	GTTGCCAGAGAATAAGGAAT	ATGACCTTCTACGCAGTC
Krt18, Intron, bp 63148003	M. musculus	PCR	5'	GCTGGAGGATGGAGAAGATT	GGGAGACAGAGTATGTGGAA
Chr. 13, bp 4870651	M. musculus	PCR	5'	CATGGAACACAGCCTCTTCA	ATTTCTCTCCAGTCACTCCC
Kirrel3, Intron, bp 36276208	M. musculus	PCR	3'	GTTTCTCCCTTFCCTCTGT	TCCATGCATTCCCTAGTC
Chr. 15, bp 54727587	M. musculus	PCR	5'	TGACGCTCATTCCAACGATG	CTCTCGCTCTCTGATATTG
Oct4	M. musculus	PCR	NA	ATTGGGGAGGGAGAGGTGAAA	TGGGGGTGAGAAGGCCAAGT
Nanog	M. musculus	PCR	NA	GGCCTTCTAGTCTGAAATAG	AACCTCACAAAAAATGGG
VANGL1, Intron 4, bp 43955	H. sapiens	PCR/qPCR	5'	CCAGGGCTTAGCAGACACAGCC	GGCTGGGAACTATCAGGGGA
VANGL1, Intron 4, bp 44454	H. sapiens	PCR/qPCR	3'	ACGCTGTTACACAGATCCGCA	ACAGCAAGGTTTGCACACTCTGA
miR17A	H. sapiens	PCR	NA	TTTAAACAGGATATTTACGTTCTGC	GAGGAAATCTTACATCCACG
VANGL1	H. sapiens	qPCR	1	GGCGGTTGTCGGAGATCAATCC	TCCCCCTCACCTCGTACC
VANGL1	H. sapiens	qPCR	2	CCACTCTTCTCCCATGCGTTGC	GGGACGTCCGAGGTGTGAGC
VANGL1	H. sapiens	qPCR	3	GCAGCGGACTGAGATGTGTGCT	GCACTCTTTGCGGGTCCGA
VANGL1	H. sapiens	qPCR	4	TGCTAGCTCATTGTTGTTTTTCAGGGA	TGGCCAAGTGGCAGTGTGCTT
VANGL1	H. sapiens	qPCR	5	CCGCAAGCGGAGAGCTGACA	GAAGCGGACTCGCCATCGG
VANGL1	H. sapiens	qPCR	6	GCGGCTACAGTGGGATGGC	CCAGGGAAACCAACAGCAGCA
VANGL1	H. sapiens	qPCR	7	GGCGGAGGAATTGGCTGCTC	TGTCCTCAAGCTGCAGCAAGTGTG
VANGL1	H. sapiens	qPCR	8	GCCTCCAAATCCGAGCAGCCA	GCAGGAAACACACACCCTGC
VANGL1	H. sapiens	qPCR	9	AGCAGAATACCACAGCATGGAGA	ACACGCTATAATCAACTGGCCAAGC
VANGL1	H. sapiens	qPCR	10	CTTGGCCCTTGGGTGCAGGG	CAAGGAGGCCAGGCCAGGGA
EGFR	H. sapiens	qPCR	1	GCACCTCCAGCGGTGTTGTGT	GGCGTCCGAGGTGCTCT
EGFR	H. sapiens	qPCR	2	CAGAGGGGCAAGTGTGGGAACG	GGCCCCGCGGGACCTAGT
EGFR	H. sapiens	qPCR	3	TGGGTCCCCGCTGCTGGTT	GTCGGGCGCTCACACCGT
EGFR	H. sapiens	qPCR	4	GTGCTGGTGGGCGCTGG	CAGTGCAGTCTCGGCGGGTG
EGFR	H. sapiens	qPCR	5	GGAGCAAGGCCCCGTAGGA	GCTGTGCGAGCTCTGCCTTC
EGFR	H. sapiens	qPCR	6	TCATGGTTTCAATTTTCGCTCCAGTGT	GCCATGGCAAGGATGGTATCTGT
EGFR	H. sapiens	qPCR	7	GCTGACTGCGCCAGACTCTACA	GGACAGGGACCAGGCCACCA
EGFR	H. sapiens	qPCR	8	CCTGCTTCCCTCTGCCTGTGG	ACTGAACCTGTGACTCACCCCT
EGFR	H. sapiens	qPCR	9	CAGGCCCTCAGAGCACTTGCC	AAGAACGGCTTGGTCAGCCCC
EGFR	H. sapiens	qPCR	10	CCAGGCCCAACTGTGAGCAAGG	TGCCAGGAAGGACAGATGG
VANGL1	H. sapiens	Bisulfite Sequencing	5' 3'	GATATAGGTATTTAAATATGGAATGA	TTAATAAAACAAAAAATAAACTCTAATA
EGFR	H. sapiens	Bisulfite Sequencing	5' 3'	TTTTGTTTTTTGTTTTTATTTTTA	ATAACTAAAACTACCTAACTACC
EGFR	H. sapiens	qPCR	No CCGG 1	TGTCCCTGGTTCAAGTGTGCCA	CCTTGCCCTCCCCAACGA
EGFR	H. sapiens	qPCR	No CCGG 2	TGAGCGAGTCTGGCTTCGTGACT	TCTATGCCTTGCCCTCCCC