



Gene Content of the GoldenGate[®] Methylation Cancer Panel I

I. Background

The main purpose of this document is to provide a broad view of the content of Illumina's GoldenGate Methylation Cancer Panel I product. The annotation information, extracted from NCBI's RefSeq mRNA records (updated as of December 11, 2006), is provided in the accompanying gene annotation file.

For detailed product information, such as information about targeted CpG sites in these genes, consult the product white paper or other marketing materials.

II. Functional Categories of Genes

The GoldenGate Methylation Cancer Panel I product contains a panel of cancer-related genes collected from Illumina's scientific collaborators, literature, and public databases.

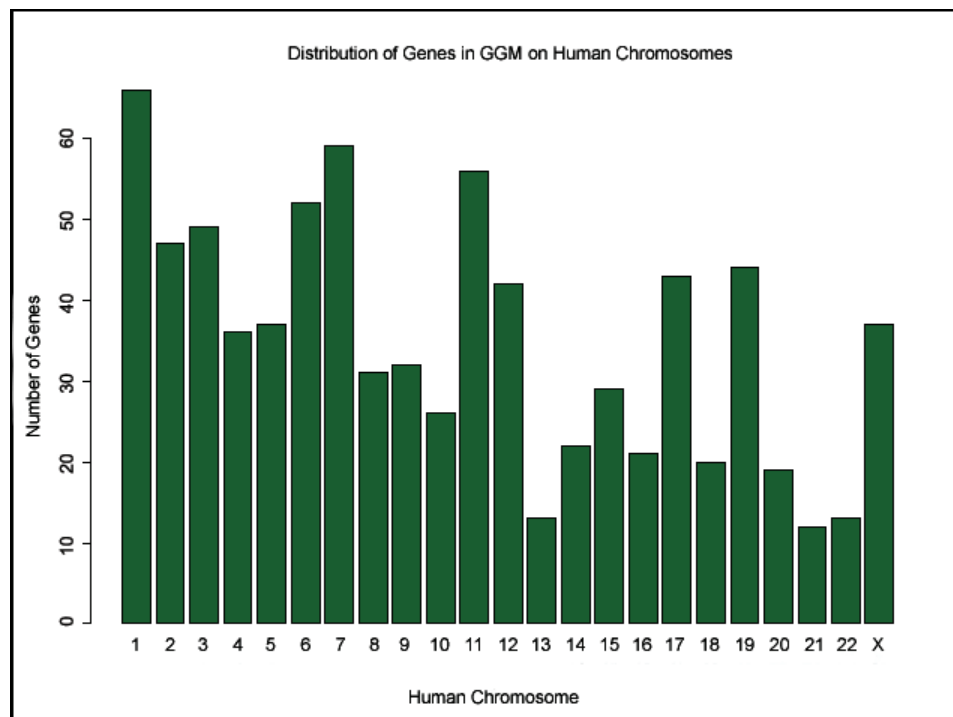
For the distribution of these genes in the Gene Ontology (<http://www.geneontology.org>) categories, please see Appendix A of this document and the accompanying data file.

III. Number of Genes

There are 807 genes in the cancer panel according to human RefSeq 36.2 annotation.

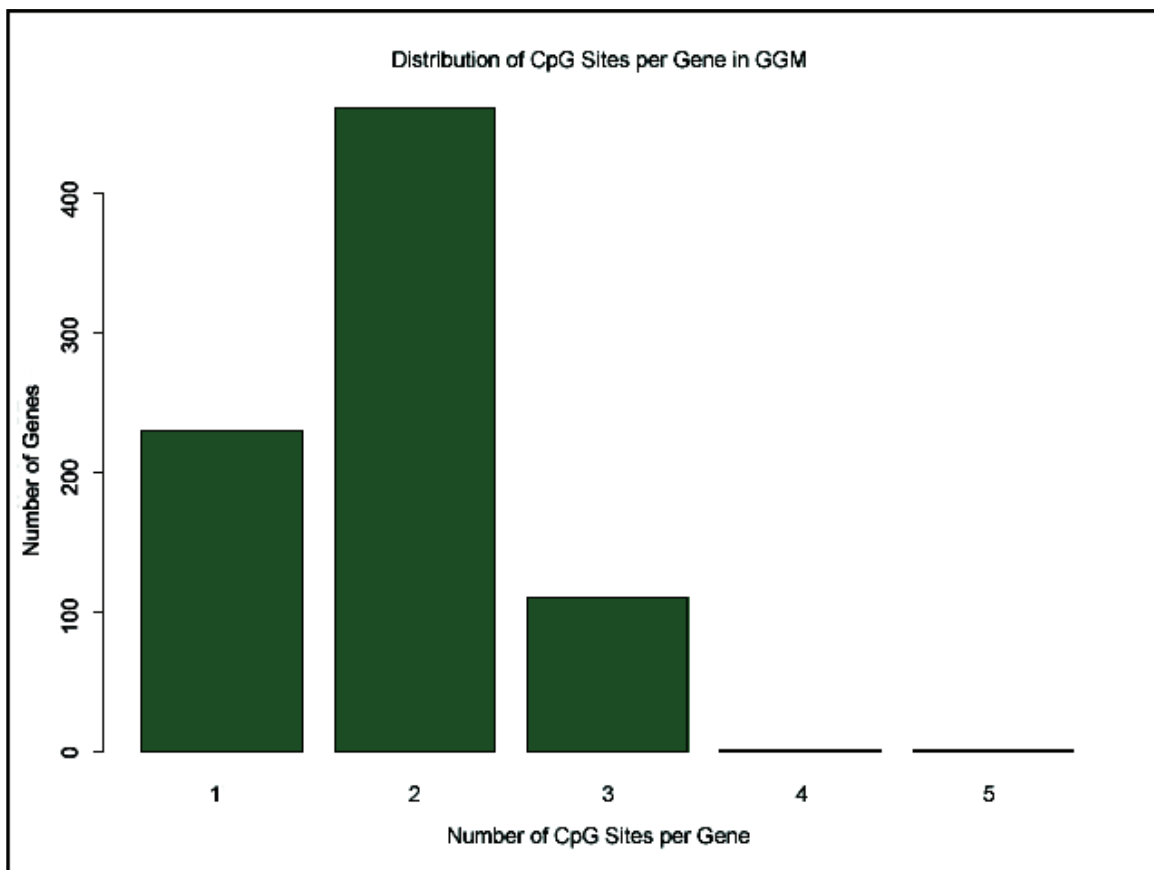
IV. Distribution of Genes by Chromosome

Figure 1 shows the distribution of genes by chromosome.



V. Distribution of CpG Sites by Gene

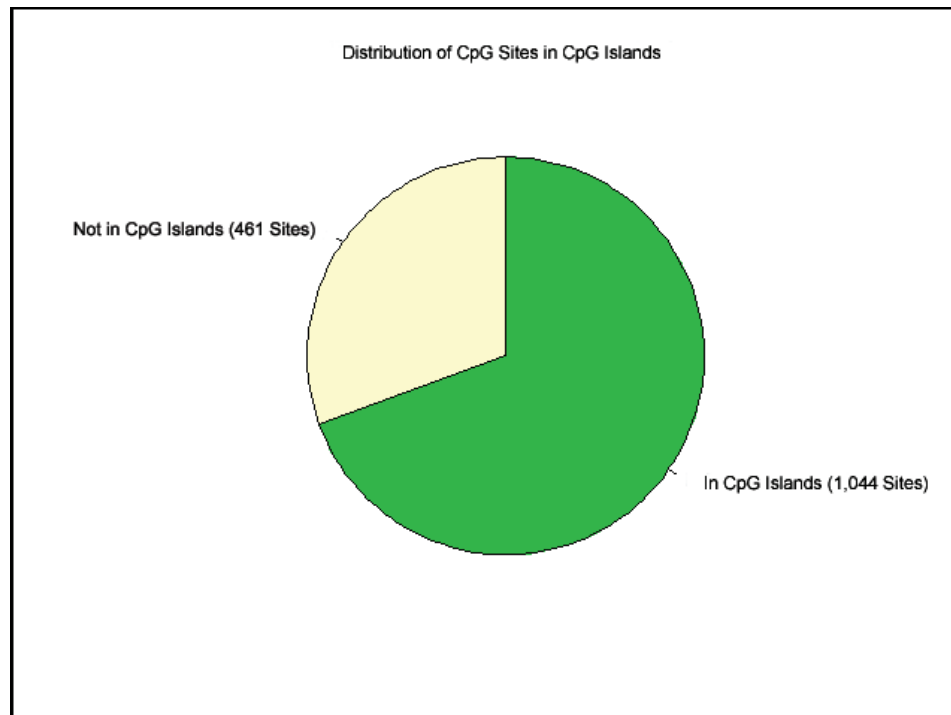
Figure 2 shows the distribution of number of CpG sites per gene.





VI. Distribution of CpG Sites in CpG Islands

Figure 3 shows the 1,505 CpG sites represented in the GoldenGate Methylation Cancer Panel I. 1,044 CpG sites are located within CpG islands, and 461 are located outside of CpG islands.



CpG islands are defined using the relaxed criteria found in "Takai D and Joines PA. 2002. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. PNAS USA. 99(6):3740-5." Specifically, a CpG Island is defined as a nucleotide sequence of (1) 200-bp or greater in length, (2) 50% or greater in GC-percent, and (3) 0.60 or greater in the ratio of observed CpG sites over expected CpG sites.

Appendix A: Gene Annotation Data File - Fields and Definitions

The file `GoldenGate_Methylation_Cancer_Panel_I_gene_info.csv` does not contain RefSeq annotation entries for the following six genes, which are present in the panel. NCBI's RefSeq permanently suppressed or temporarily removed these entries for a variety of reasons. Annotation for these genes can be obtained from NCBI's GENE and RefSeq databases through NCBI's online Entrez tool. The GeneID, Accession number, or both are provided below.

GeneID	RefSeq Accession Number	Notes
7682	N/A	Gene: MKRN4
4611	N/A	Gene: MYCL2
54551	NM_019066.2	Temporarily removed as of December 22, 2006
55025	XM_940022.1	Gene model; permanently removed
79136	NM_024123.1	Permanently suppressed; a pseudogene
9834	NM_014792.2	Permanently suppressed



Field Number	Field Name	Definition
1	Acc.Ver	Composed of RefSeq's nucleotide sequence accession number, followed by a dot ("."), followed by the annotation version number of the record.
2	GI	NCBI's sequence identification number, known as GenInfo Identifier, which uniquely identifies a nucleotide sequence.
3	Transcript-Length	Length in nt of the annotated RNA sequence.
4	CDS-Length	Length in nt of the annotated CDS feature, if the sequence is mRNA.
5	Chr_No	Human chromosome number, 1-22, X and Y.
6	Cyto_Map	Cytochromosomal location of the annotated gene.
7	State	RefSeq annotation state.
8	Gene_Symbol	Gene symbol extracted from NCBI's RefSeq records.
9	Synonyms	Synonym(s) of the gene.
10	Protein_Id	Protein accession version of the encoded peptide, if the entry is an mRNA sequence.
11	Product	Protein product name, if the sequence is mRNA.
12	gene-Note	Value of the "note" qualifier for the feature key "gene," extracted from the feature table of NCBI's RefSeq record.
13	CDS-Note	Value of the "note" qualifier for the feature key "CDS," extracted from the feature table of NCBI's RefSeq record. This is applicable for the mRNA sequence entry.
14	GO_Component	Gene Ontology Component annotation, extracted from the RefSeq record.
15	GO_Function	Gene Ontology Function annotation, extracted from the RefSeq record.
16	GO_Process	Gene Ontology Process annotation, extracted from the RefSeq record.
17	PubMed_Id	List of all PubMed database IDs extracted from the "Reference" field of the RefSeq record.
18	N_PubMed	Number of PubMed_Ids for the record.
19	MIM_Id	ID or IDs of MIM (Mendelian Inheritance in Man), extracted from the RefSeq record.
20	GENE_Id	NCBI's unique Gene ID to identify the gene record.

FOR RESEARCH ONLY

© 2006 Illumina, Inc.

Illumina, Making Sense Out of Life, Sentrix, GoldenGate, DASL, BeadArray, Array of Arrays, Infinium, BeadXpress, VeraCode, IntelliHyb, iSelect, and CSPro are registered trademarks or trademarks of Illumina. All other brands and names contained herein are the property of their respective owners.

Pub. No. 270-2006-004 20Dec06