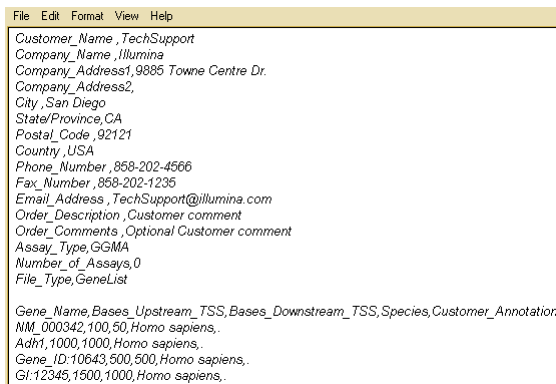
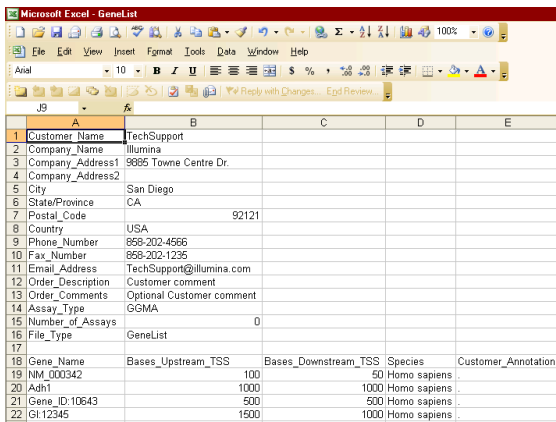


Table 2: Genelist File Column Descriptions

Heading	Description
Species	Entered by customer. Valid entries are human, man, or Homo sapiens.
Gene_Name	Customer-supplied gene name. Can be HUGO gene symbol, RefSeq accession ID, GI number, or Gene ID.
Bases_Upstream_TSS	Number of bases upstream of the TSS to search. Must be between 1 and 2000.
Bases_Downstream_TSS	Number of bases downstream of the TSS to search. Must be between 1 and 1000.
Customer_Annotation	Customer comments. Limited to 30 characters.
CustomerDefinedPassThrus	These columns are optional and not limited in number. Entries must not contain commas and should be concise.

Figure 3: Genelist File Examples



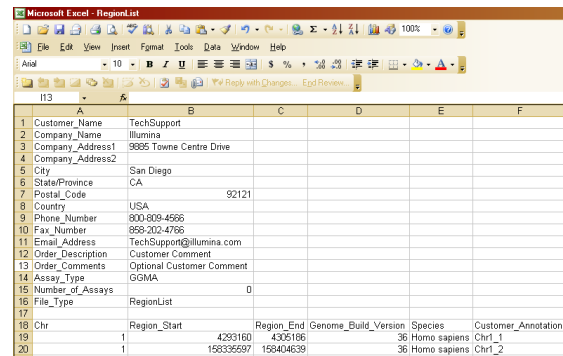
Example of properly formed entries in a GeneList file shown from Excel (top) and Notepad (bottom).

provided in the SequenceList input file. Figure 5 provides examples of properly formed SequenceList entries.

GGMAScore Output File

After an input file is submitted via email to technical support, a Technical Support Scientist will submit the file to ADT for process-

Figure 4: Regionlist File Example



Example of properly formed entries in a RegionList file shown from Excel (top) and Notepad (bottom).

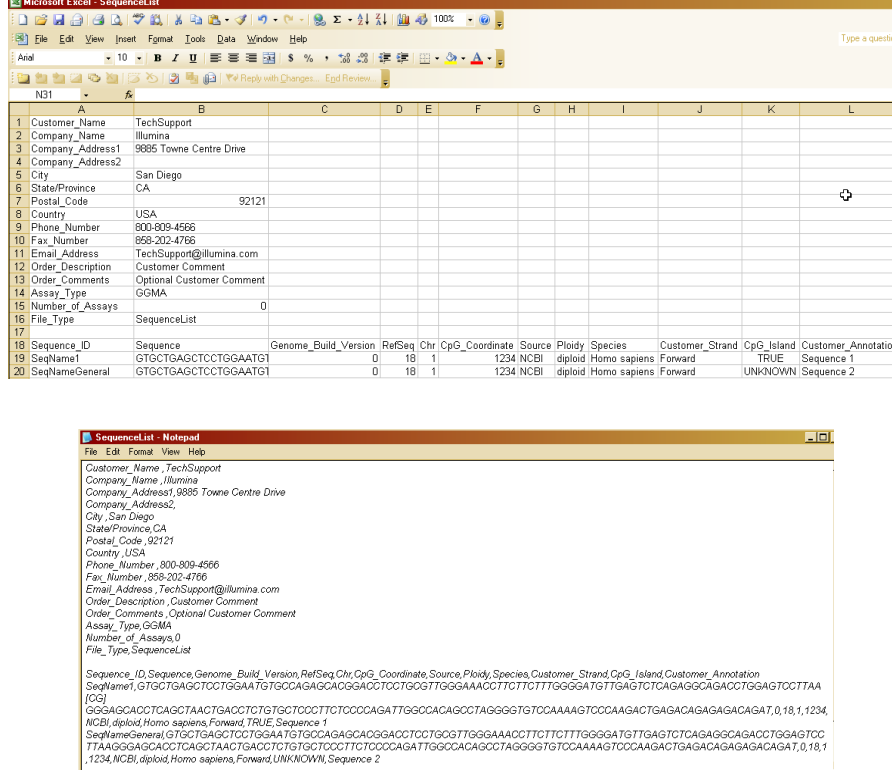
ing. ADT generates and returns an output file called the GGMAScore file, which is returned to the customer by email or secure FTP. The GGMAScore file lists the CpGs, sequences, coordinates, score, and gene annotation information of the CpG list defined in the input file. The GGMAScore file contains the same columns as the SequenceList file, along with results of the oligonucleotide evaluation and gene annotation information (if available).

The GGMAScore file is the output file of ADT for custom methylation design. This file can be edited on a per row basis to remove CpGs predicted to perform poorly or those spaced too closely together. CpGs identified using more than one input search method (e.g., GeneList,

Table 3: Regionlist File Column Descriptions

Heading	Description
Genome_Build_Version	Genome build that will be queried. Check with a Technical Support Scientist1 for the currently supported build.
Chr	Chromosome on which the CpG site is located.
Species	Customer-defined species (currently human only). Valid entries are human , man , or Homo sapiens.
Region_Start	First chromosome coordinate of region to search.
Region_End	End chromosome coordinate of region to search. Total region size must be at least 1 and less than 10,001.
Customer_Annotation	This column heading is required but the entry field may be left empty.
CustomerDefinedPassThrus	These column headings are optional and not limited in number. Entries must not contain commas and should be concise.

Figure 5: Sequencelist File Example



Example of properly formed entries in a SequenceList file shown from Excel (top) and Notepad (bottom).

RegionList, or SequenceList) may be combined as one GGMAScore file and resubmitted to ADT as a single input file for evaluation as a single oligo pool. The column heading information listed in Table 5 must be provided in the GGMAScore input file. Figure 6 provides examples of properly formed GGMAScore file entries.

Output File Failure Codes

Failure codes indicate the reasons why a CpG might be inappropriate for the Illumina platform or incompatible with other CpGs in the same pool. Table 6 defines the various failure codes that might be returned by ADT in the GGMAScore file.

Technical Note: Epigenetics

(poisoned) increases. Therefore, high scoring loci should be substituted for low scoring loci whenever possible.

Validation_Class/Validation_Bin

Because the GoldenGate Assay for Methylation is relatively new, Illumina does not have validation information available yet. This column is a place holder for information that may be added at a later date.

Underlying_SNP

This column lists the rs IDs for any SNPs found within the assay design region for a CpG. Such a SNP will also trigger failure code 304. Information about these SNPs can be obtained from dbSNP². These assays are designable, but the failure code is issued as a warning to customers that performance may be affected by a polymorphism in the assay design region. In some instances allele frequency may be such that a SNP will not pose a risk to assay success in a given sample population.

CpG_Island

This field indicates whether a CpG falls within a CpG island. CpG islands are defined by the Illumina ADT according to Takai and Jones relaxed criteria³.

ILMN_Design_Strand

This field describes which strand (TOP/BOT) has been used for designing the assay. When two loci in the assay pool fall within 60nt of each other, both loci receive failure code 340. If these assays

target different strands, the proximity of the assays will not affect performance and they can be included in the same oligo pool without risk. An important caveat is that the TOP strand for one sequence may be on the same genomic strand as the BOT strand for another sequence. Therefore, the TOP/BOT designation must be correlated to the genomic orientation before determining whether two CpG sites in close proximity are suitable for use in the same pool.

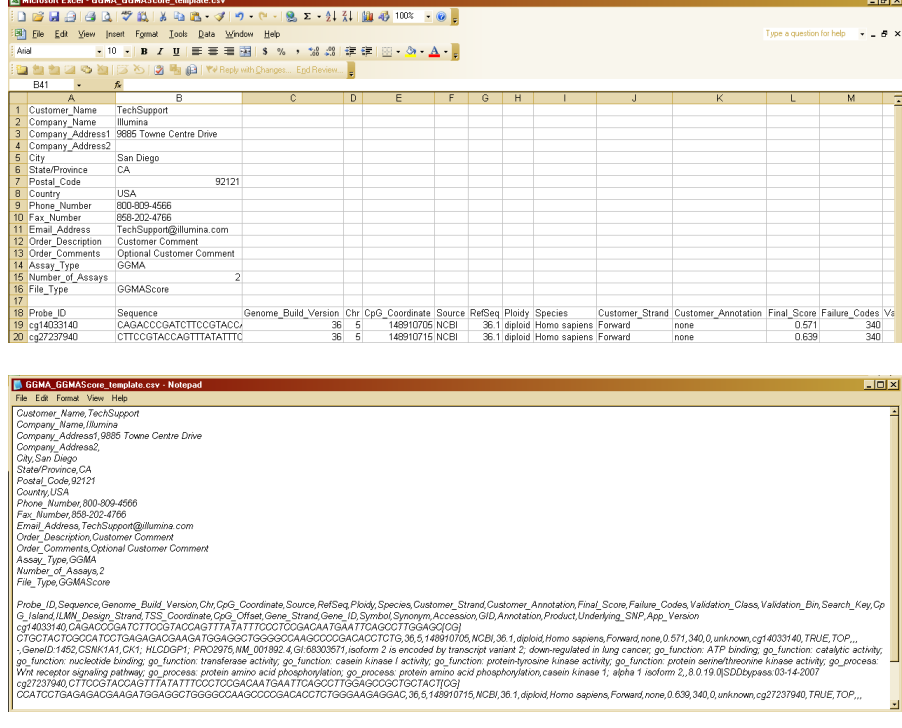
CpG_Offset

When designing an OMA from SequenceList files, unannotated sequences may be submitted. ADT identifies all CG dinucleotides in this sequence and assigns an offset position starting with first base as position 1. ADT uses the CpG_offset and the Search_Key to identify subsequences that came from the same original sequence, and to determine whether they are too close to include in the same pool. Loci that are positioned too close to each other will receive failure code 340. Multiple unannotated sequences submitted in the same SequenceList input file that cover overlapping genomic regions will not be compared to each other. Thus, in such cases, care should be taken not to include overlapping CpG sequences.

Summary

Illumina custom Methylation Assay Panels allow scientists to perform experiments tailored directly to specific hypotheses. By following the guidelines in this technical note, researchers can ensure that their orders are designed and placed quickly and easily. Evaluating potential loci with ADT ensures the high-quality assays that scientists expect and Illumina delivers.

Figure 6: GGMAsscore File Examples



Examples of properly formed entries in a GGMAsscore file shown from Excel (top) and Notepad (bottom).

Figure 7: Header Section For Final Order File

Heading	Description	Value required
Customer_Name	Name of person submitting the ADT File	Yes
Company_Name	Company name (no commas)	Yes
Company_Address1	Line 1 of customer's address	Yes
Company_Address2	Line 2 of customer's address (optional)	No
City	Customer's city	Yes
State/Province	Customer's state or province	Yes
Postal_Code	Customer's postal code	Yes
Country	Customer's country	Yes
Phone_Number	Customer's phone number	Yes
Fax_Number	Customer's fax number	Yes
Email_Address	Customer's email address	Yes
Order_Description	Description of work	Yes
Order_Comments	Additional comments (optional)	No
Assay_Type	GGMA	Yes
Design_Iteration	Final	Yes
Scale (Number_of_Tubes)	Must be 1 or greater	Yes
Purchase_Order_Number	Customer purchase order number	Yes
Number_of_Assays	Number of CpG loci in file	Yes
File_Type	GGMAScore	Yes

Illumina, Inc. • 9885 Towne Centre Drive, San Diego, CA 92121 USA • 1.800.809.4566 toll-free • 1.858.202.4566 tel • techsupport@illumina.com • illumina.com

FOR RESEARCH USE ONLY

© 2011 Illumina, Inc. All rights reserved.

Illumina, illuminaDx, BeadArray, BeadXpress, cBot, CSPRO, DASL, DesignStudio, Eco, GAllx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iSelect, MiSeq, Nextera, Sentrix, Solexa, TruSeq, VeraCode, the pumpkin orange color, and the Genetic Energy streaming bases design are trademarks or registered trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners.

Pub. No. 270-2010-001 Current as of 12 September 2011



TTAAGATGATAACAGTAAACACACTTCTGTTAAACCTTAAGATTACTTTGATCCACTGATTCAACGTACCGTAAACGAAAGTATCAATTGAGACTAAATATAACGTACCCATTAAGAGCTACCGTCTTCTGTTAAACCTTAAGATTACTTTGATCCACTGATTCAACGTAT
 WKTCAACGTACCGTAAACGAAAGTATCAATTGAGACTAAATATAACGTACCCATTAAGAGCTACCGTCTTCTGTTAAACCTTAAGATTACTTTGATCCACTGATTCAACGTATCAACGTACCGTAAACGAAAGTATCAATTGAGACTAAATATAACGTACCCATTAAGAGCTACCGTCTTCTGTTAAACCTTAAGATTACTTTGATCCACTGATTCAACGTAT
 CGACGAAAGTATCAATTGAGACTAAATATAACGTACCCATTAAGAGCTACCGTCTTCTGTTAAACCTTAAGATTACTTTGATCCACTGATTCAACGTATCAACGTACCGTAAACGAAAGTATCAATTGAGACTAAATATAACGTACCCATTAAGAGCTACCGTCTTCTGTTAAACCTTAAGATTACTTTGATCCACTGATTCAACGTAT
 CTAAACGTACCGTAAACGAAAGTATCAATTGAGACTAAATATAACGTACCCATTAAGAGCTACCGTCTTCTGTTAAACCTTAAGATTACTTTGATCCACTGATTCAACGTATCAACGTACCGTAAACGAAAGTATCAATTGAGACTAAATATAACGTACCCATTAAGAGCTACCGTCTTCTGTTAAACCTTAAGATTACTTTGATCCACTGATTCAACGTAT
 TAAAGATGATAACAGTAAACACACTTCTGTTAAACCTTAAGATTACTTTGATCCACTGATTCAACGTACCGTAAACGAAAGTATCAATTGAGACTAAATATAACGTACCCATTAAGAGCTACCGTCTTCTGTTAAACCTTAAGATTACTTTGATCCACTGATTCAACGTAT
 AAGATTACTTTGATCCACTGATTCAACGTATCAACGTACCGTAAACGAAAGTATCAATTGAGACTAAATATAACGTACCCATTAAGAGCTACCGTCTTCTGTTAAACCTTAAGATTACTTTGATCCACTGATTCAACGTATCAACGTACCGTAAACGAAAGTATCAATTGAGACTAAATATAACGTACCCATTAAGAGCTACCGTCTTCTGTTAAACCTTAAGATTACTTTGATCCACTGATTCAACGTAT
 AACGTATCAACGTACCGTAAACGAAAGTATCAATTGAGACTAAATATAACGTACCCATTAAGAGCTACCGTCTTCTGTTAAACCTTAAGATTACTTTGATCCACTGATTCAACGTATCAACGTACCGTAAACGAAAGTATCAATTGAGACTAAATATAACGTACCCATTAAGAGCTACCGTCTTCTGTTAAACCTTAAGATTACTTTGATCCACTGATTCAACGTAT
 AAGATTACTTTGATCCACTGATTCAACGTATCAACGTACCGTAAACGAAAGTATCAATTGAGACTAAATATAACGTACCCATTAAGAGCTACCGTCTTCTGTTAAACCTTAAGATTACTTTGATCCACTGATTCAACGTATCAACGTACCGTAAACGAAAGTATCAATTGAGACTAAATATAACGTACCCATTAAGAGCTACCGTCTTCTGTTAAACCTTAAGATTACTTTGATCCACTGATTCAACGTAT
 AAGATTACTTTGATCCACTGATTCAACGTATCAACGTACCGTAAACGAAAGTATCAATTGAGACTAAATATAACGTACCCATTAAGAGCTACCGTCTTCTGTTAAACCTTAAGATTACTTTGATCCACTGATTCAACGTATCAACGTACCGTAAACGAAAGTATCAATTGAGACTAAATATAACGTACCCATTAAGAGCTACCGTCTTCTGTTAAACCTTAAGATTACTTTGATCCACTGATTCAACGTAT