Designing Custom GoldenGate® Methylation Profiling Panels

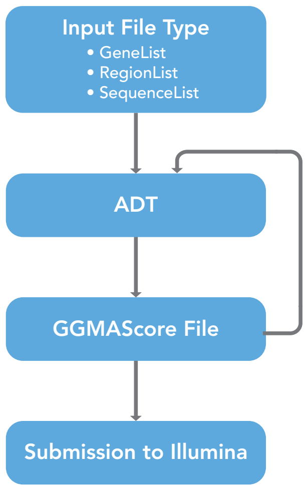
Guidelines for efficiently creating and ordering high-quality custom panels for methylation profiling using the Illumina Assay Design Tool.

Introduction

Illumina GoldenGate Assay for Methylation is a powerful and customizable method to simultaneously analyze the methylation status of 96–384 CpG sites. Illumina has created an Assay Design Tool (ADT) for efficiently creating custom panels for methylation profiling. This technical note provides guidelines for researchers to design successful custom content panels.

Customers create an assay panel by selecting specific assays to target CpG loci within genes or regions of interest. To ensure successful assay development, Illumina provides ADT, which customers use to evaluate a list of CpG loci of interest and ultimately prepare a final set of assays to order from Illumina. ADT creates CpG sequence lists from either human chromosome regions or human gene identifiers. In addition, a customer may submit their proprietary sequences via ADT for evaluation on the Illumina platform. ADT produces and returns a GGMAScore file from a submitted file. The GGMAScore file can then be edited to either remove CpGs predicted to be poor performers, or to add CpGs and further refine the oligonucleotide set. ADT can then accept the edited GGMA-Score file as input for evaluation in an iterative manner. A GGMAScore file is also the format for the final set of CpG loci that are submitted to Illumina when the product is ordered. The set of loci in this file defines the Oligo Pool for Methylation Assay (OMA) that is manufactured and delivered to the customer.

Figure 1: Custom methylation Panel Design Workflow



Preliminary Input Files

There are three different file types used by ADT for the three different methods of evaluating CpG loci: GeneList, RegionList, and SequenceList. Additionally, a GGMAScore output file can be used as an input file in subsequent rounds of evaluation. At this time, ADT returns only human sequences from GeneList and RegionList input files. It is important to note that ADT only supports one build of the human genome at a time. Illumina keeps the supported version of the human genome current and gives users at least two weeks notice when switching to a new version. Technical Support Scientists[1] can confirm which version of the human genome is currently supported.

Input files may be created or edited in any text editor or spreadsheet program. However, before submitting them to ADT, files must be saved in a comma-separated values (*.csv) format. The examples provided in this document show files created in Microsoft Excel and Notepad. Blank lines are generally not permitted in the data fields or between lines in the heading. The following formatting requirements must be followed precisely so ADT can properly evaluate your requests.

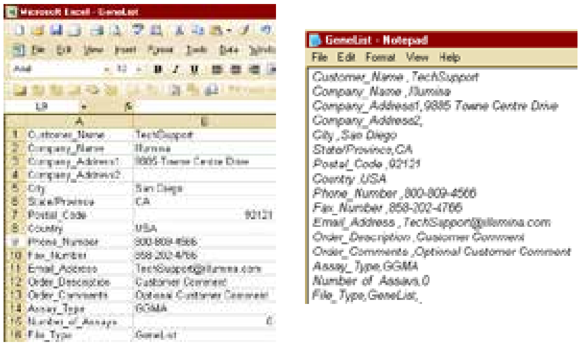
- **Format is comma-separated values with a *.csv file extension.**
- **File includes a file header section. File header format is the same for all file types, and declares the file type.**
- **File includes specific column headings for the data. As described below, each file type has different requirements for column headings.**
- **File contains fewer than 65,000 CpGs. If the number of CpGs exceeds this limit, the file must be split into smaller segments for serial ADT evaluation.**
- File Header

The file header is common to all preliminary input files. Since the input file format is comma delimited, no commas may be used within the heading entries. Table 1 lists and describes the required headings. Figure 2 provides examples of properly formatted file headers.

GeneList

The GeneList file type provides a method for querying all CpGs in the regions upstream and downstream from the transcription start site (TSS) of the indicated gene. A GeneList allows for the interrogation of the currently supported build of the human genome using HUGO or RefSeq gene ID and gene number identifiers. The column description information shown in Table 2 must be provided in the GeneList input file. Figure 3 provides examples of proper GeneList entries in Notepad and Excel.

Figure 2: Examples of File Header Section


Microsoft Excel - GeneList
File Edit View Insert Format Tools Data Window
Arial 12 B I U
L9
A
B
GeneList - Notepad
File Edit Format View Help

	A	B
1	Customer_Name	TechSupport
2	Company_Name	Illumina
3	Company_Address1	9885 Towne Centre Drive
4	Company_Address2	
5	City	San Diego
6	State/Province	CA
7	Postal_Code	92121
8	Country	USA
9	Phone_Number	800-809-4566
10	Fax_Number	858-202-4766
11	Email_Address	TechSupport@illumina.com
12	Order_Description	Customer Comment
13	Order_Comments	Optional Customer Comment
14	Assay_Type	GGMA
15	Number_of_Assays	0
16	File_Type	GeneList

```
Customer_Name ,TechSupport
Company_Name ,Illumina
Company_Address1,9885 Towne Centre Drive
Company_Address2,
City ,San Diego
State/Province,CA
Postal_Code ,92121
Country ,USA
Phone_Number ,800-809-4566
Fax_Number ,858-202-4766
Email_Address ,TechSupport@illumina.com
Order_Description ,Customer Comment
Order_Comments ,Optional Customer Comment
Assay_Type,GGMA
Number_of_Assays,0
File_Type,GeneList,
```

File header section displayed as a screen shot from Excel (left) and Notepad (right).

RegionList

The RegionList file type provides a method for selecting CpGs located between specified locations of a human chromosome. A RegionList file contains a list of regions in the human genome identified by chromosome and coordinate range that ADT will search and evaluate from among CpGs cataloged in the currently supported human genome build. The headings and information shown in Table 3 must be provided in the RegionList input file. Figure 4 provides examples of properly formed RegionList entries.

SequenceList

The SequenceList file provides a method for evaluating CpGs from private databases or other sources. The Sequence_ID field is used to name sequences for easy identification. Sequence_ID entries contained in this file must not begin with "cg," because that prefix designates cg names in the Illumina database and will trigger a database search. To include loci from a previous OMA that were designed using a previous build of the human genome, it is best to use the sequence record from the original score file as input for ADT. Any such sequences should have a prefix other than "cg" manually added to the cg ID to prevent the ADT from using an updated (and perhaps different) sequence for the locus.

Any number of CpG dinucleotides can be specified within one sequence submission for design. To specify a CpG, simply put brackets around one or more CG dinucleotides in the sequence (e.g., ATGCTG**[**CG**]**GCATGCTAAT). Sequences should include at least 50 base pairs on either side of a marked CpG site and the total length of the sequence should be limited to 2,000 base pairs. There is no maximum for the number of CpG sites that can be indicated in a sequence. If more than one site is indicated, ADT provides a separate line of output in the GGMAScore file for each bracketed. If no dinucleotides are bracketed in the sequence, ADT identifies all CpGs in the submission, then evaluates and reports on all CpG sites in the sequence. The headings and information shown in Table 4 must be

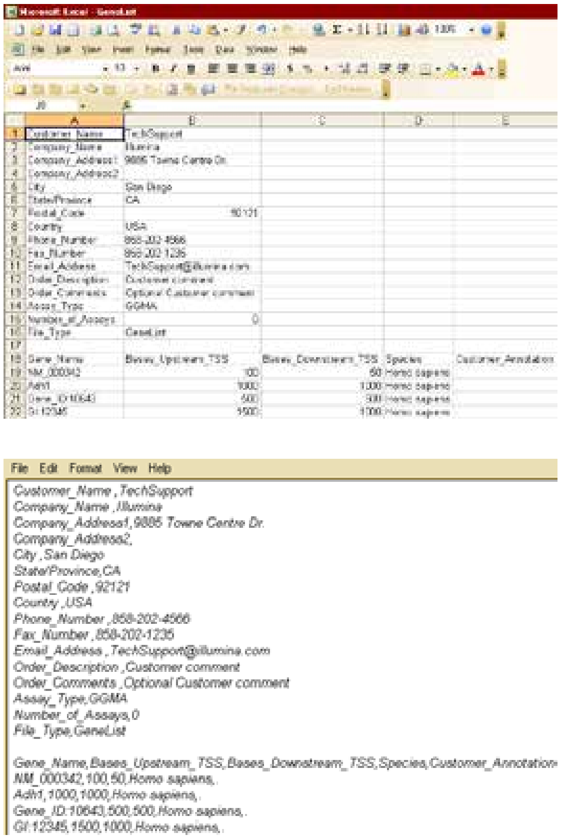
Table 1: Required File Headings For ADT Input File

Heading	Description	Entry 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
Number_of_Assays	Number of CpG loci in file (may be **0** for GeneList, RegionList, and SequenceList file if the number of CpG loci is unknown)	Yes
File_Type	**GeneList, RegionList, SequenceList, or GGMA_Score**	Yes

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


Microsoft Excel - GeneList
File Edit View Insert Format Tools Data Window Help
File Edit Format View Help

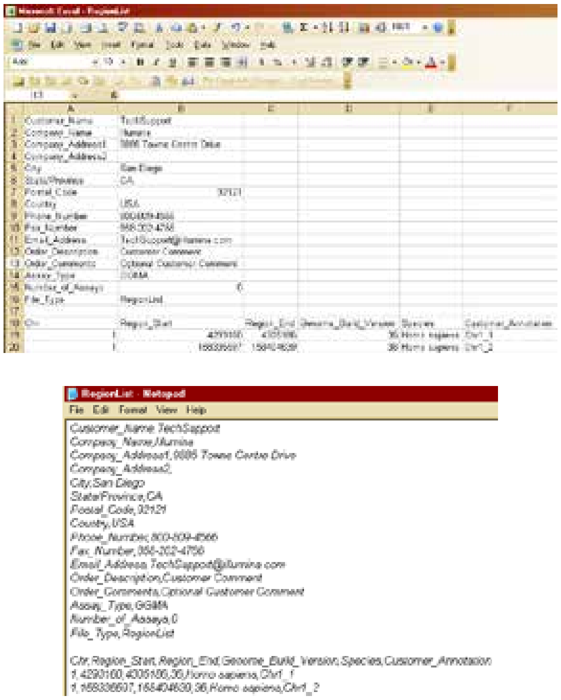
	A	B	C	D	E
1	Customer_Name	TechSupport			
2	Company_Name	Illumina			
3	Company_Address1	9885 Towne Centre Dr.			
4	Company_Address2				
5	City	San Diego			
6	State/Province	CA			
7	Postal_Code	92121			
8	Country	USA			
9	Phone_Number	858-202-4566			
10	Fax_Number	858-202-1235			
11	Email_Address	TechSupport@illumina.com			
12	Order_Description	Customer comment			
13	Order_Comments	Optional Customer comment			
14	Assay_Type	GGMA			
15	Number_of_Assays	0			
16	File_Type	GeneList			
17					
18	Gene_Name	Bases_Upstream_TSS	Bases_Downstream_TSS	Species	Customer_Annotation
19	NM_000342	100	50	Homo sapiens	
20	Adh1	1000	1000	Homo sapiens	
21	Gene_ID:10643	500	500	Homo sapiens	
22	GI:12345	1500	1000	Homo sapiens	

```
Customer_Name ,TechSupport
Company_Name ,Illumina
Company_Address1,9885 Towne Centre Dr.
Company_Address2,
City ,San Diego
State/Province,CA
Postal_Code ,92121
Country ,USA
Phone_Number ,858-202-4566
Fax_Number ,858-202-1235
Email_Address ,TechSupport@illumina.com
Order_Description ,Customer comment
Order_Comments ,Optional Customer comment
Assay_Type,GGMA
Number_of_Assays,0
File_Type,GeneList

Gene_Name,Bases_Upstream_TSS,Bases_Downstream_TSS,Species,Customer_Annotation
NM_000342,100,50,Homo sapiens,
Adh1,1000,1000,Homo sapiens,
Gene_ID:10643,500,500,Homo sapiens,
GI:12345,1500,1000,Homo sapiens,
```

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

Figure 4: Regionlist File Example


Microsoft Excel - RegionList
File Edit View Insert Format Tools Data Window Help
RegionList - Notepad
File Edit Format View Help

	A	B	C	D	E	F
1	Customer_Name	TechSupport				
2	Company_Name	Illumina				
3	Company_Address1	9885 Towne Centre Drive				
4	Company_Address2					
5	City	San Diego				
6	State/Province	CA				
7	Postal_Code	92121				
8	Country	USA				
9	Phone_Number	800-809-4566				
10	Fax_Number	858-202-4756				
11	Email_Address	TechSupport@illumina.com				
12	Order_Description	Customer Comment				
13	Order_Comments	Optional Customer Comment				
14	Assay_Type	GGMA				
15	Number_of_Assays	0				
16	File_Type	RegionList				
17						
18	Chr	Region_Start	Region_End	Genome_Build_Version	Species	Customer_Annotation
19	1	4290100	4305106	36	Homo sapiens	Chr1_1
20	1	158336607	158404630	36	Homo sapiens	Chr1_2

```
Customer_Name,TechSupport
Company_Name,Illumina
Company_Address1,9885 Towne Centre Drive
Company_Address2,
City,San Diego
State/Province,CA
Postal_Code,92121
Country,USA
Phone_Number,800-809-4566
Fax_Number,858-202-4756
Email_Address,TechSupport@illumina.com
Order_Description,Customer Comment
Order_Comments,Optional Customer Comment
Assay_Type,GGMA
Number_of_Assays,0
File_Type,RegionList

Chr,Region_Start,Region_End,Genome_Build_Version,Species,Customer_Annotation
1,4290100,4305106,36,Homo sapiens,Chr1_1
1,158336607,158404630,36,Homo sapiens,Chr1_2
```

Example of properly formed entries in a RegionList 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

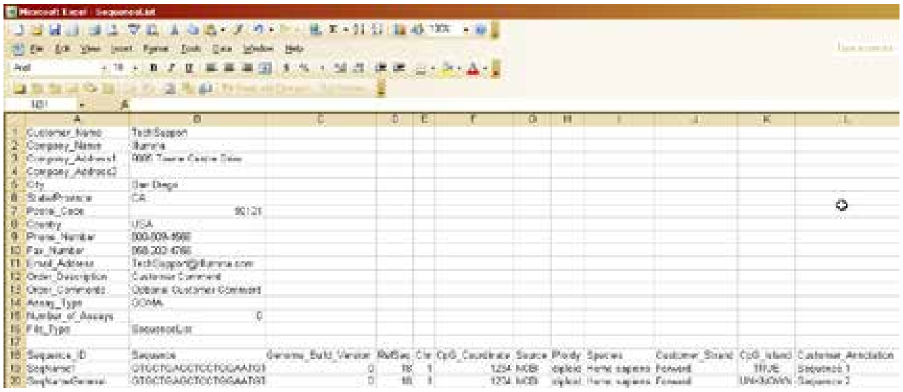
After an input file is submitted via email to technical support, a Technical Support Scientist will submit the file to ADT for processing. ADT generates and returns an output file called the GGMAScore file, which is returned to the customer by email or secure FTP. The GGMAScore file lists the CpGs, sequences, 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

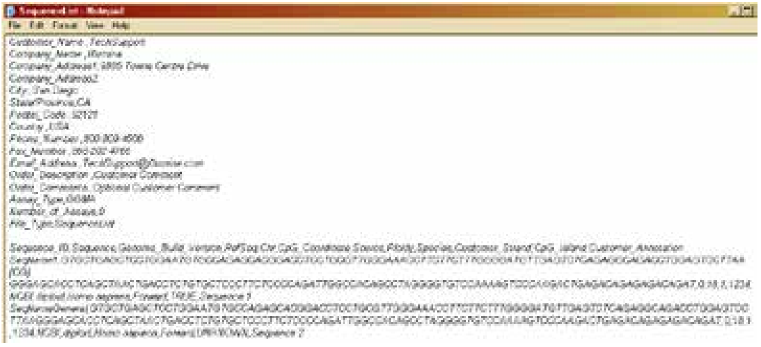
Heading	Description
Genome_Build_Version	Genome build that will be queried. Check with a Technical Support Scientist1 for the currently supported build.
Chr	Chromosome on which the CpG site is located.
Species	Customer-defined species (currently human only). Valid entries are **human**, **man**, or Homo 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



Microsoft Excel - SequenceList

	A	B	C	D	E	F	G	H	I	J	K	L
1	Customer_Name	TechSupport										
2	Company_Name	Illumina										
3	Company_Address1	9885 Towne Centre Drive										
4	Company_Address2											
5	City	San Diego										
6	State/Province	CA										
7	Postal_Code	92121										
8	Country	USA										
9	Phone_Number	800-809-4566										
10	Fax_Number	858-202-4766										
11	Email_Address	TechSupport@illumina.com										
12	Order_Description	Customer Comment										
13	Order_Comments	Optional Customer Comment										
14	Assay_Type	GGMA										
15	Number_of_Assays	0										
16	File_Type	SequenceList										
17												
18	Sequence_ID	Sequence	Genome_Build_Version	RefSeq	Chr	CpG_Coordinate	Source	Ploidy	Species	Customer_Strand	CpG_Island	Customer_Annotation
19	SeqName1	GTGCTGAGCTCCTGGAATGT	0	18	1	1234	NCBI	diploid	Homo sapiens	Forward	TRUE	Sequence 1
20	SeqNameGeneral	GTGCTGAGCTCCTGGAATGT	0	18	1	1234	NCBI	diploid	Homo sapiens	Forward	UNKNOWN	Sequence 2



SequenceList - Notepad

```
Customer_Name,TechSupport
Company_Name,Illumina
Company_Address1,9885 Towne Centre Drive
Company_Address2
City,San Diego
State/Province,CA
Postal_Code,92121
Country,USA
Phone_Number,800-809-4566
Fax_Number,858-202-4766
Email_Address,TechSupport@illumina.com
Order_Description,Customer Comment
Order_Comments,Optional Customer Comment
Assay_Type,GGMA
Number_of_Assays,0
File_Type,SequenceList

Sequence_ID,Sequence,Genome_Build_Version,RefSeq,Chr,CpG_Coordinate,Source,Ploidy,Species,Customer_Strand,CpG_Island,Customer_Annotation
SeqName1,[illegible]
[CG]
[illegible],0,18,1,1234
,NCBI,diploid,Homo sapiens,Forward,TRUE,Sequence 1
SeqNameGeneral,[illegible]
[illegible],0,18,1
,1234,NCBI,diploid,Homo sapiens,Forward,UNKNOWN,Sequence 2
```

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.

Table 4: Sequencelist Column Headers

Heading	Description
Sequence _ID	Customer-supplied sequence identifier.
Sequence	Limited to 10,000 bases. May contain zero or more bracketed CpG sites. Output will be ≤ 122 bases per line.
Genome_Build_Version	Version number supplied by customer. Otherwise, enter **unknown**.
Chr	Chromosome on which the CpG is located. Must be an integer, **X**, or **Y**. Enter **0** if unknown.
CpG_Coordinate	Chromosome coordinate of CpG. Enter **0** if unknown.
CpG_Island	The value indicates whether the CpG site falls in a CpG Island. Use Takai and Jones relaxed criteria2. Must be **true**, **false**, or **unknown**.
Source	Identify the source of the sequence and annotation data. Must be completed. Enter **unknown** if no information is available.
RefSeq	A RefSeq accession field does not have to be supplied. If your list contains some regions which have RefSeq IDs and some which do not, enter **0** for those which do not rather than leaving the cells empty.
Species	Customer defined species (only human is supported currently). Valid entries are **human**, **man**, Homo sapiens.
Customer_Strand	Must contain one of the following 3 values: **forward**, **reverse**, or **unknown**. Information is customer-supplied and is not validated.
Symbol	A gene symbol for the sequence or gene. May be left empty.
Accession	Identify an accession number for the sequence or gene. May be left empty.
Ploidy	Since human is currently the only supported species, this entry must be **diploid**.
Customer_Annotation	Required annotation. This column heading is required and the field must contain a value.
CustomerDefinedPassThrus	These column headings are optional and not limited in number. Entries must not contain commas and should be concise.

Final Input File

After the selection of CpGs is finalized, the GGMAScore file needs to be converted to a final design request by the addition of three lines of information to the header section of the file. The additional heading lines (Design_Iteration, Scale, and Purchase_Order_Number) are indicated in bold in Table 7. This final order file is used to order the oligo pool from Illumina.

Immediately following the header section is the data section, which is generated from previous ADT output. The set of CpGs desired to be in the final pool should be selected and copied from any GGMAScore files and included in the data section of the final input file used for final analysis and ordering. When removing CpGs from a GGMAScore file, it is essential to remove the entire row of text corresponding to that CpG. This will ensure data integrity when the file is resubmitted to ADT.

When compiling a GGMAScore file, sequences for CpGs identified by a “cg” number will always be looked up in Illumina’s internal database, whereas sequences with CpG_Names starting with characters other than “cg” will be scored as submitted. Non-cg entries must have one and only one CpG site identified in square brackets.

Considerations for final selection of oligo pool

When selecting CpGs for the final list, it is important to use the information in the GGMAScore file to select assays that achieve the scientific aims of the experiment and have the highest chances for generating meaningful results. Below are some recommendations for filtering GGMAScore files.

Final_Score

Illumina recommends preferentially choosing assays with scores ≥ 0.8. Internal testing shows little difference in performance between assays with scores from 0.8–1.0.

Loci with scores below 0.4 are set to 0 by default and are given critical failure code 108. Based on internal testing, assays with scores < 0.4 consistently perform poorly in the GoldenGate Assay and negatively impact other assays in the same OMA. These loci are considered undesignable.

Loci with scores from 0.4–0.79 should be chosen by weighing their score and scientific importance. The probability for assay success decreases along with the score. Additionally, as the score decreases, the chance that other assays in the same OMA will be negatively impacted (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 dbSNP[2]. 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 criteria[3].

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.

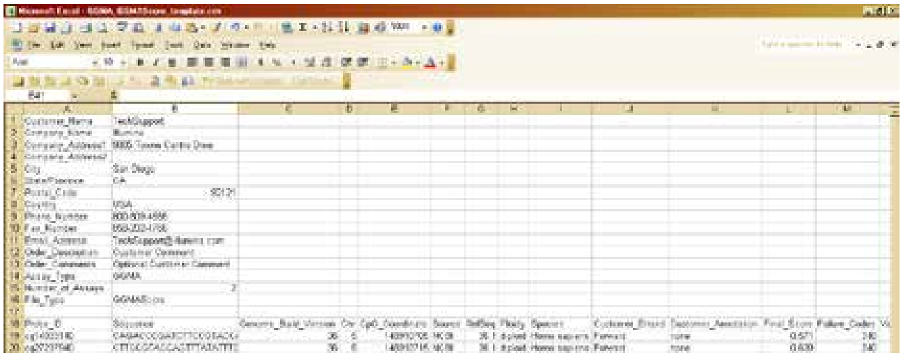
CpG_Offset

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

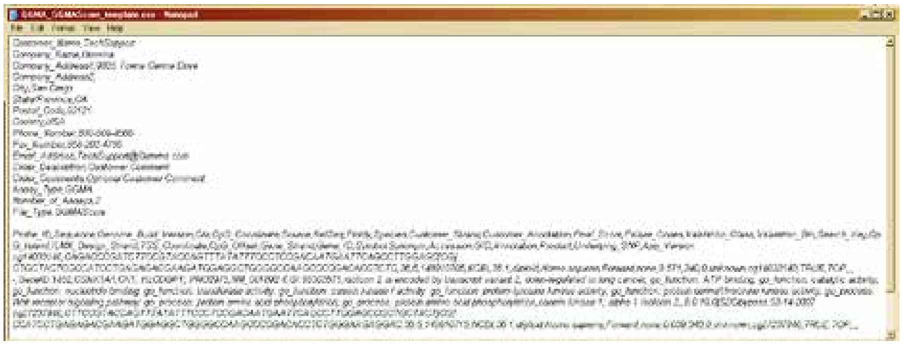
Illumina custom Methylation Assay Panels allow scientists to perform experiments tailored directly to specific hypotheses. By following the guidelines in this technical note, researchers can ensure that their orders are designed and placed quickly and easily. Evaluating potential loci with ADT ensures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

Figure 6: GGMAscore File Examples



Microsoft Excel - GGMA_GGMAScore_template.csv
File Edit View Insert Format Tools Data Window Help
Arial
B41

	A	B	C	D	E	F	G	H	I	J	K	L	M	Va
1	Customer_Name	TechSupport												
2	Company_Name	Illumina												
3	Company_Address1	9885 Towne Centre Drive												
4	Company_Address2													
5	City	San Diego												
6	State/Province	CA												
7	Postal_Code	92121												
8	Country	USA												
9	Phone_Number	800-809-4566												
10	Fax_Number	858-202-4766												
11	Email_Address	TechSupport@illumina.com												
12	Order_Description	Customer Comment												
13	Order_Comments	Optional Customer Comment												
14	Assay_Type	GGMA												
15	Number_of_Assays	2												
16	File_Type	GGMAScore												
17														
18	Probe_ID	Sequence	Genome_Build_Version	Chr	CpG_Coordinate	Source	RefSeq	Ploidy	Species	Customer_Strand	Customer_Annotation	Final_Score	Failure_Codes	Va
19	cg14003140	CAGAACCCGATCTTCCGTACG	36	5	140910705	NCBI	36.1	diploid	Homo sapiens	Forward	none	0.571	340	
20	cg27237940	CTTCCGTACCAGTTTATATT	36	5	140910715	NCBI	36.1	diploid	Homo sapiens	Forward	none	0.609	340	



GGMA_GGMAScore_template.csv - Notepad
File Edit Format View Help

```
Customer_Name,TechSupport
Company_Name,Illumina
Company_Address1,9885 Towne Centre Drive
Company_Address2,
City,San Diego
State/Province,CA
Postal_Code,92121
Country,USA
Phone_Number,800-809-4566
Fax_Number,858-202-4766
Email_Address,TechSupport@illumina.com
Order_Description,Customer Comment
Order_Comments,Optional Customer Comment
Assay_Type,GGMA
Number_of_Assays,2
File_Type,GGMAScore

Probe_ID,Sequence,Genome_Build_Version,Chr,CpG_Coordinate,Source,RefSeq,Ploidy,Species,Customer_Strand,Customer_Annotation,Final_Score,Failure_Codes,Validation_Class,Validation_Bin,Search_Key,CpG_Island,ILMN_Design_Strand,TSS_Coordinate,CpG_Offset,Gene_Strand,Gene_ID,Symbol,Synonym,Accession,GID,Annotation,Product,Underlying_SNP,Ade_Version
cg14003140,CAGAACCCGATCTTCCGTACG[illegible]
[illegible],36,5,140910705,NCBI,36.1,diploid,Homo sapiens,Forward,none,0.571,340,0,unknown,cg14003140,TRUE,TOP,[illegible]
[illegible]
[illegible]
[illegible]
cg27237940,CTTCCGTACCAGTTTATATT[illegible]
[illegible],36,5,140910715,NCBI,36.1,diploid,Homo sapiens,Forward,none,0.609,340,0,unknown,cg27237940,TRUE,TOP,[illegible]
```

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

Table 5: Column Headers For GGMAscore File

Heading	Description
Probe_ID	Illumina “cg” identifier.
Sequence	The bracketed CpG identified by the Probe_ID and 60 nucleotides of flanking sequence.
Genome_Build_Version	Genome build that will be queried. Contact Customer Solutions1 for currently supported build.*
Chr	Chromosome on which the CpG site is located.*
CpG_Coordinate	Chromosome coordinate of CpG.*
Source	Database source.*
RefSeq	RefSeq version used from Source.
Ploidy	Only **diploid** is allowed.*
Species	Only **human**, **man**, or Homo sapiens are valid entries.*
Customer_Strand	**Forward** for all GeneList and RegionList results.*
Customer_Annotation	Annotation passed through from input files.
Final_Score	Quality score assigned by Illumina. Ranges from 0–1.
Failure_Codes	If applicable, reasons why a successful assay at this CpG is unlikely. For a complete list of failure codes, see Table 6.
Validation_Class	Must be **0**. This is a place holder for future applications.
Validation_Bin	Must be **unknown**. This is a place holder for future applications.
App_Version	A concatenated field containing the software versions used to analyze submitted data and the date on which analysis was completed.
Search_Key	Same as Probe_ID for cg IDs. Same as Sequence_ID for SequenceList submissions.
CpG_Island	Reports whether the CpG is within a predicted CpG island (**True**, **False**, or **Unknown**).**
ILMN_Designed_Strand	The strand on which the assay has been designed (**TOP** or **BOT**).
TSS_Coordinate	Chromosomal coordinate of the TSS.
CpG_Offset	Offset of CpGs identified in submitted sequence. Position 1 is the first nucleotide in the sequence.
Gene_Strand	Genomic orientation of the coding strand for the gene (**+** or **-**).
Gene_ID	GeneID for the submitted gene.
Symbol	HUGO gene symbol for the submitted gene.
Synonym	Known synonyms for the submitted gene.
Accession	RefSeq accession ID for the submitted gene.
GID	GI number for the submitted gene.
Annotation	NCBI gene ontology information for the submitted gene.
Product	Gene product definition.
Underlying_SNP	rsIDs known to fall within the assay design region. **-99** if none.

*Information is customer defined and passed through without modification from SequenceList input files.

**The Illumina ADT uses relaxed criteria to define CpG islands, based on the method described by Takai and Jones2.

AAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAA
AATCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACC
AACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTA
TTAACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGAT
AAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAA
AAGATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGA
AACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACA

Figure 6: List Of Failure Codes For The ADT

Code #	Definition
Critical Failures (undesignable)	
101	Flanking sequence is too short.
102	CpG or sequence formatting error. CpG must match the format [CG]. Possible causes: Spaces found in submitted sequence Missing brackets around CpG Locus to design (bracketed sequence) is not “CG”
103	TOP/BOT strand cannot be determined.
106	Degenerate nucleotide(s) in assay design region (e.g., W, R, S, N)
108	Final score falls below assay limit of 0.4.
Warnings (designable)	
304	There are known SNPs within the probe region. Reference Underlying_SNP column for details.
340	Another CpG site in the list is too close. Including more than one assay within 61 base pairs may reduce the chance of success for both assays.
399	Multiple contributing issues.

References

1. To contact Technical Support, send email to techsupport@illumina.com or call 1.800.809.4566.
2. http://www.ncbi.nlm.nih.gov/projects/SNP/
3. Takai D, Jones PA (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc Natl Acad Sci U S A 99: 3740-3745.

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

FOR RESEARCH USE ONLY

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