

Designing and Ordering iSelect® HD Custom Genotyping Assays

Guidelines for efficiently creating and ordering high quality custom Infinium® Genotyping Assays using the Illumina Assay Design Tool.

INTRODUCTION

The Illumina Infinium HD Assay is a powerful and widely used assay for highly multiplexed genotyping. Along with a broad portfolio of fixed content BeadArray™ products, Illumina offers custom and semi-custom genotyping panels deployed on various BeadChips. Researchers can design their own content panels for custom and semi-custom BeadChips with the assistance of the online Assay Design Tool (ADT) and Illumina scientists. The iSelect HD Custom Genotyping BeadChip can be completely custom designed to assay between 3,000 and 200,000 genotyping markers per sample in any species. The BeadChips can be deployed on either the 24-sample format (3,000 to 60,800 attempted bead types), or the 12-sample format (60,801 to 200,000 attempted bead types). The Human1M-Duo+ and HumanHap550-Quad+ are semi-custom genotyping BeadChips that combine standard human genome-wide content and custom content.

Researchers create custom panels by selecting and

submitting a list of requested loci to Illumina. To ensure successful assay development, this list is evaluated with the Assay Design Tool to select an initial assay panel that includes desired assays that are predicted to have a high likelihood of success. Metrics returned by ADT provide predicted success information, validation status, and minor allele frequencies from published studies.

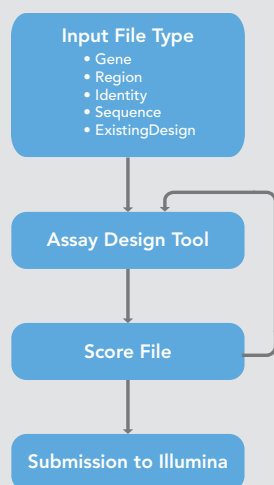
This technical note describes the process of designing, analyzing, and ordering a custom panel of markers. Researchers must submit certain files describing the custom panel. This information is evaluated by ADT, and results are used to refine the panel definition. Each of the file type format options for ADT input and output are described with examples. Template files can be downloaded from iCom¹, or by contacting Illumina Technical Support².

CUSTOM AND SEMI-CUSTOM GENOTYPING PRODUCTS

For array customization in human studies, Illumina offers the opportunity to add up to 60,800 markers to the proven whole-genome SNP genotyping and CNV identification content available on the standard Infinium HD BeadChips. The Human1M-Duo+ and HumanHap550-Quad+ are semi-custom HD genotyping BeadChips that combine standard genome-wide content (more than 1,100,000 or 560,000 tag SNPs, respectively) plus 3,000 to 60,800 custom bead types to assay additional markers.

Custom Genotyping BeadChips can also include copy number variation (CNV) assays. For these loci, a long genomic region is assayed with multiple markers to determine if the region is present, absent, or present in multiple copies. Since the customer can specify the number of probes for a particular CNV assay, the number of bead types for a CNV region can vary significantly.

FIGURE 1: ISELECT CUSTOM GENOTYPING ASSAY DESIGN WORKFLOW



Bead Type Definition

Depending on the type of SNP or marker being assayed, the Infinium HD Assay uses one of two probe (or bead type) designs, Infinium I or Infinium II. The Infinium II probe design, which stops at the base before the SNP of interest, uses only one probe per loci (i.e., one probe for both alleles). This probe design is suitable for the majority of loci in most organisms. Infinium I probe design is required for relatively less common A/T and C/G SNPs and requires two probes (or bead types) per SNP because the probe stops at the base representing the SNP of interest (i.e., one probe for each of both alleles).

For ordering and pricing, custom BeadChip products are defined by their number of attempted bead types, not number of loci assayed. Only if a project design is exclusively limited to loci using Infinium II design (A/G, A/C, T/G, T/C SNPs; and all indels) are the number of markers equal to the number of bead types. The more markers ordered that require Infinium I design, the fewer total markers that can be attempted on a custom BeadChip.

For example, suppose that a researcher wants to create an iSelect BeadChip to assay 10,000 markers. 9,500 are Infinium II design and 500 are Infinium I design. 9,500 bead types are required for the first set of markers (one probe per marker), and 1,000 bead types are required for the second set of markers (two probes per markers). Thus, the total number of attempted bead types is 10,500 (not 10,000).

Bead Type Success Rate

Illumina guarantees a bead type success rate of at least 80%. The bead type success rate is the proportion of the total number of attempted assays present on the final product. Manufacturing a custom BeadChip encompasses many steps, including synthesizing all required oligonucleotide probes, attaching probes to silica beads, combining the various bead types into the final bead pool, loading the bead pools onto the BeadChip, and decoding the bead identities on the BeadChip. The manufacturing steps required to build custom BeadChips have been optimized through years of experience, yet no step is 100% efficient. The manifest file supplied with BeadChips will contain all successful assays (i.e., those present on the final product), not all of the ordered assays. Illumina does not guarantee analytical success rates due to the fully custom nature of these projects.

Bead Pool and Product Shelf Life

Illumina guarantees the life of iSelect HD bead pools for 12 months from the date of manufacture. Illumina keeps custom bead pools in inventory for the entire 12 months to enable reorders of the same project. Once the 12 month shelf life has expired, any remaining bead pools are discarded.

BeadChips manufactured from iSelect HD bead pools have a six month shelf life from the date of manufacture. BeadChips are typically made to order and are not kept in inventory. Therefore, the overall life of an iSelect project is up to 18 months, assuming the last order of BeadChips is manufactured during the last month of bead pool shelf life.

Illumina notifies customers of an impending bead pool expiration in order to give them a last opportunity to order products using a custom bead pool. In some cases, a bead pool may be kept past the 12 month expiration (by customer request), but the quality of that bead pool cannot be guaranteed, and Illumina reserves the right to not manufacture if the yield does not meet quality requirements.

PRELIMINARY INPUT FILES

To submit preliminary input files, researchers can use five different file types corresponding to the five different methods for preliminary evaluation of custom SNP loci: Gene, Region, Identity, Sequence, and CNVRegion. Additionally, requests for probe designs from a previously ordered iSelect product or current Illumina standard product use the ExistingDesign file type. After preliminary evaluation, ADT generates a Score file that can be used as an input file in subsequent rounds of evaluation or for ordering (Figure 1).

At this time, ADT returns only human sequences from Gene, Region, Identity, and CNVRegion 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 before switching to a new version. Technical Support Scientists² can confirm which version of the human genome is currently supported.

ADT supports full duplicate and repeat region checking of files for human assays. When designs are created for other species supported by ADT, varying levels of support for duplicate and repeat region checking will be given, depending on the availability of public sequence data. If ADT does not support the species, no duplicate and repeat region checking will be performed and customers should ensure their assays target unique sequences. For

TABLE 1: GENE FILE COLUMN HEADING DESCRIPTIONS

HEADING	DESCRIPTION
Gene_Name	Customer-supplied gene name. Can be a RefSeq accession ID or HUGO gene symbol.
Bases_Upstream	Number of bases to search upstream of the gene's starting coordinate.
Bases_Downstream	Number of bases to search downstream of the gene starting coordinate.

non-human supported species, Illumina Technical Support Scientists can provide information about the level of duplicate and repeat region support provided by ADT.

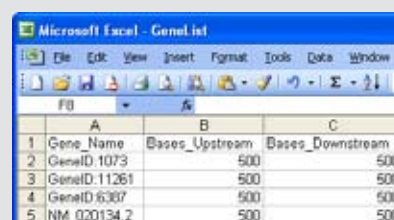
Input files may be created or edited with any text editor or spreadsheet program. However, before submitting them to ADT, files must be saved in a comma-separated values (*.csv) format. The examples provided in this document show files created in Microsoft Excel. Blank lines are generally not permitted 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. Since the input file format is comma-delimited, no commas may be used within the values.
- Each file type includes specific column headings for the data, as described below.
- File contains fewer than 1,000,000 markers or indels. If the number of markers exceeds this limit, the file must be split into smaller segments for serial ADT evaluation.
- If the file is submitted by email rather than on iCom, it must include a file header section. File header format is the same for all file types (Figure 7 and Table 7).

GENE

The Gene file type provides a method for querying all markers within a gene and in the regions upstream and downstream from the indicated gene. A Gene file allows for the interrogation of the currently supported build of the human genome using RefSeq NM accession ID (preferred) or HUGO identifiers. ADT maps these

FIGURE 2: GENE FILE FORMAT EXAMPLES



	A	B	C
1	Gene Name	Bases_Upstream	Bases_Downstream
2	GeneID:1073	500	500
3	GeneID:11261	500	500
4	GeneID:6307	500	500
5	NM_020134.2	500	500

Example of properly formed entries in a Gene file shown from Excel.

accession numbers to the human genome to identify gene regions. The sizes of upstream and downstream regions queried by ADT are specified by the user. Markers in overlapping gene regions will be listed in the Score output file only once, but will be annotated as being present in both regions in the **Region_Description** field. The column headings and description information shown in Table 1 must be provided in the Gene file. Figure 2 provides examples of proper Gene entries in Notepad and Excel. A general guideline is that a Gene file with up to 600 genes and a range of 10 kb upstream and downstream will be within the marker limit.

REGION

The Region file type enables researchers to select markers present between specified locations of a human chromosome. A Region file contains a list of regions in the human genome identified by chromosome and coordinate

TABLE 2: REGION FILE COLUMN HEADING DESCRIPTIONS

HEADING	DESCRIPTION
Chromosome	Chromosome on which the marker is located. Must be a valid chromosome for the species specified in the metadata.
Start_Coordinate	Non-negative integer indicating the start of the region of interest (inclusive).
End_Coordinate	Non-negative integer indicating the end of the region of interest (inclusive).

FIGURE 3: REGION FILE FORMAT EXAMPLES

	A	B	C
1	Chromosome	Start_Coordinate	End_Coordinate
2	17	58898166	58862935
3	17	37266705	37338798

Example of properly formed entries in a Region file shown from Excel.

FIGURE 4: IDENTITY FILE FORMAT EXAMPLES

	A
1	Locus_Name
2	rs10403552
3	rs13343438
4	rs11671249

Example of properly formed entries in a Identity file shown from Excel.

range that ADT will search and evaluate from among cataloged markers in a current Illumina-internal version of dbSNP. This internal database does not contain MNPs, SSRs, or markers with ambiguous or multiple localizations. markers in overlapping regions will be listed in the Score output file only once, but will be annotated as being present in both regions in the **Region_Description** field. Since ADT limits output to 1,000,000 markers, submitting less than 60 Mb of regions per file is recommended. The

column headings and description information shown in Table 2 must be provided in the Region input file. Figure 3 provides examples of properly formed Region entries.

IDENTITY

Known markers described in the current version of dbSNP can be requested specifically using the Identity file type. A current internal version of dbSNP is the source for rs marker and flanking sequence data. The column headings

TABLE 3: IDENTITY FILE COLUMN HEADING DESCRIPTIONS

HEADING	DESCRIPTION
Locus_Name	rs number taken from dbSNP. Value must begin with "rs" (case insensitive) followed by an integer.

TABLE 4: SEQUENCE FILE COLUMN HEADING DESCRIPTIONS

HEADING	DESCRIPTION
Locus_Name	Customer-supplied unique locus name (cannot begin with rs or cg). Name must not contain any of the following characters: " % / = \ ? ` , @ ; ' , or space.
Target_Type	Must be "SNP" or "INDEL" (case insensitive).
Sequence	Limited to 10,000 bases. May only contain one bracketed SNP or indel. Output will be ≤ 122 bases per line.
Chromosome	Chromosome on which the marker is located. Must be a valid chromosome for the species being analyzed. Enter 0 if unknown.
Coordinate	Chromosome coordinate of marker. Enter 0 if unknown.
Genome_Build_Version	Version number supplied by customer. Otherwise, enter 0.
Source	Identify the source of the sequence and annotation data. Must be completed. Enter unknown if no information is available.
Source_Version	Source version number. Enter 0 if unknown.
Sequence_Orientation	Must be either "Forward" or "Reverse" (case insensitive).
Plus_Minus	Must be either "PLUS" or "MINUS" (case insensitive).

FIGURE 5: SEQUENCE FILE FORMAT EXAMPLES

	A	B	C	D	E	F	G	H	I	J
	Locus_Name	Target_Type	Sequence	Chromosome	Coordinate	Genome_Build_Version	Source	Source_Version	Sequence_Orientation	Plus_Minus
1	seqrs10403652	SNP	ATTCTCTCCAGGGCT	19	59900243	36	dbSNP	128	Forward	Plus
2	seqrs13343438	SNP	cagggcccaatctcagctca	19	59751810	36	dbSNP	128	Forward	Plus
3	seqrs11671249	SNP	tagaggcattagcctaagfgr	19	59814896	36	dbSNP	128	Forward	Plus
4	seqrs34785303	SNP	GAACTCAGGCTGCTAA	19	59976309	36	dbSNP	128	Forward	Plus
5	seqrs36836276	SNP	CCCAGCCTGGGTTTTTC	19	59937831	36	dbSNP	128	Forward	Plus

Example of properly formed entries in a Sequence file shown from Excel.

and description information shown in Table 3 must be provided in the Identity input file. Figure 4 provides examples of properly formed Identity entries.

SEQUENCE

The Sequence file type allows researchers to evaluate markers from private databases or other sources, as well as from non-human species. The **Locus_Name** field is used to name sequences for easy identification. **Locus_Name** entries contained in this file must not begin with “rs,” because that prefix designates rs ID names in the Illumina database and will trigger a database search.

To specify a SNP, put brackets around a polymorphic locus in the submitted sequence. Separate the two alleles with a forward slash (TGC[A/C]CCG). Similarly, to specify an indel, use a slash to separate the allele from a plus or minus (insertion or deletion) sign (TGC[-/AT]CCG). A minimum of 50 bp of sequence on either side of the SNP is required, but 60 bp is preferred. ADT will also accept IUPAC codes for degenerate bases in the flanking sequence and take these into consideration during design. If the **Lowercase_Weighting** checkbox on the iCom submission form (or file header value) is unchecked, lowercase nucleotides will be considered for oligo design; or if it is checked, then lowercase nucleotides are masked. In either case, an Illumina algorithm will identify repetitive or duplicated regions in the unmasked sequence. Since lowercasing in public databases is not a standard way to indicate masking, we recommend clearing the **Lowercase_Weighting** checkbox by default.

The column headings and description information shown in Table 4 must be provided in the Sequence input file. Figure 5 provides examples of properly formed Sequence entries.

CNVREGION

The CNVRegion file type provides a method for researchers to select CNV markers present between specified locations of a human chromosome. CNV regions are typically analyzed by combining data from a set of polymorphic and nonpolymorphic probes. A CNVRegion file contains a list of regions in the human genome identified by a chromosome and coordinate range that are to be assayed for copy number variation (Table 5). The number of required probes in each region must also be specified.

CNVRegion files may optionally include a list of required polymorphic probes (Figure 6). When these markers are contained in a CNV region, they will be included in the set of markers for that region. The design module will intelligently select additional markers by balancing good probe scores with even spacing, resulting in a design consisting of high scoring and evenly spaced markers.

FIGURE 6: CNVREGION FILE FORMAT EXAMPLE

	A	B	C	D
	Chromosome	Start_Coordinate	End_Coordinate	Probes_Per_Region
1	17	50980166	50962905	5
2	17	37266705	37338798	4
3	1	54776489	54883005	5
4	3	188033157	188066946	3
5	1	201166576	201204034	5
6	[MISC]			
7	Required_Locus_Name			
8	rs10403652			
9	rs13343438			
10	rs11671249			

Example of properly formed entries in a CNVRegion file shown from Excel.

USING PREVIOUS ASSAY DESIGNS

Illumina has created a separate method for conveniently ordering the exact same assays that were designed and used on a previous iSelect or current standard Illumina product. As described in Table 6 and shown in Figure 7, an ExistingDesign file only contains a list of the Ilmn_Ids from the original design manifest.

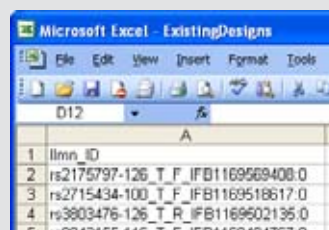
WEB SUBMITTING VIA ICOM

Preliminary input files can be submitted to ADT for evaluation either directly via a web-based interface, or by emailing the file to a technical support scientist who will submit the file to ADT. Submitting via the web is preferred since it provides rapid turnaround and 24-hour access. To submit a preliminary design file, log in to <https://iCom.illumina.com> and select **Prelim Assay Design Tool (ADT)**. The ADT interface allows users to enter necessary file type information and attach a *.csv formatted input file. After the file has been scored, an email notification is sent to the user.

SUBMITTING VIA EMAIL

Preliminary ADT input files submitted by email must include an additional file header section before the data entry section (Table 7). The format of the file header is common to all preliminary input file types. The entire *.csv file should then be emailed to techsupport@illumina.com.

FIGURE 6: EXISTINGDESIGN FILE FORMAT EXAMPLES



Example of properly formed entries in an ExistingDesign file shown from Excel.

SCORE OUTPUT FILE

If preliminary input files are submitted to ADT via iCom, an email notification is sent when scoring is complete, often within an hour. The results are returned as a Score file that can be downloaded from iCom on the Prelim Assay Design files page. If an input file is submitted via email to Technical Support, an Illumina scientist will submit the file to ADT for processing. ADT generates the Score output file, which is returned to the customer by email or secure FTP in 1–2 days. The Score file provides researchers information to help them filter and select their marker lists, as described in the following section.

The Score file header section includes additional summary information, such as the total number of markers in the file. This is further broken down into numbers of

TABLE 5: CNVREGION FILE COLUMN HEADING DESCRIPTIONS

HEADING	DESCRIPTION
Chromosome	Chromosome on which the marker is located. Must be a valid chromosome for the species specified in the metadata.
Start_Coordinate	Non-negative integer indicating the start of the region of interest (inclusive).
End_Coordinate	Non-negative integer indicating the end of the region of interest (inclusive).
Probes_Per_Region	The number of probe designs that should be returned for the region of interest.
Required_Locus_Name*	RS number taken from dbSNP; value must begin with "rs" (case insensitive) followed by an integer.

* Optionally included to specify required polymorphic probes

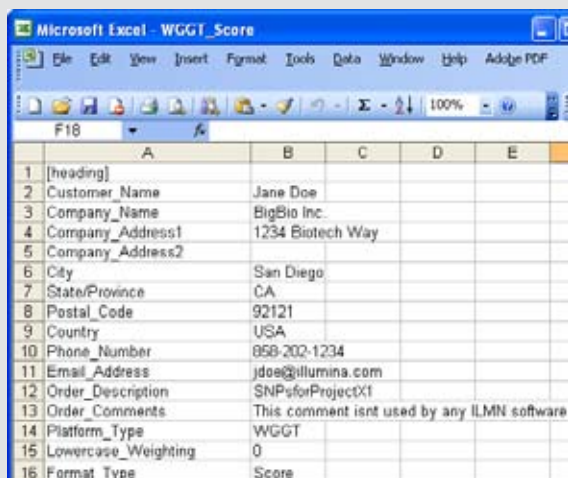
TABLE 6: COLUMN HEADINGS FOR EXISTINGDESIGN FILE

HEADING	DESCRIPTION
ILMN_ID	ID value of an ExistingDesign file, which must map to a design with the same species and assay type as specified in the metadata.

TABLE 7: FILE HEADINGS FOR ADT INPUT FILE REQUIRED ONLY FOR EMAIL SUBMISSION

HEADING	DESCRIPTION	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	No
Company_Address2	Line 2 of customer's address	No
City	Customer's city	No
State/Province	Customer's state or province	No
Postal_Code	Customer's postal code	No
Country	Customer's country	No
Phone_Number	Customer's phone number	No
Fax_Number	Customer's fax number	No
Email_Address	Customer's email address	Yes
Order_Description	Description of work (30 character limit)	No
Order_Comments	Additional comments	No
Platform_Type	WGGT	Yes
Number_of_Assays	Number of loci in file (enter 0 for GeneList and RegionList)	No
Lowercase_Weighting	0 for no masking or 1 to mask lower cased characters	No
Format_Type	Gene, Region, Sequence, Identity, ExistingDesign, or Score	Yes
Species	Species name	Yes

FIGURE 8: EXAMPLES OF FILE HEADER SECTION REQUIRED ONLY FOR EMAIL SUBMISSION



Examples of properly formed entries in the header of a Score file shown from Excel (left) and Notepad (right). This header is only required for email submissions, and is formatted the same for any preliminary file format.

TABLE 8: SCORE FILE COLUMN HEADERS

HEADING	DESCRIPTION
Locus_Name	rs number or customer's unique name.
Sequence	The bracketed site identified by the Locus_Name with > 50 bases of flanking sequence.
Genome_Build_Version	Genome build that will be queried. Contact Technical Support ² for the currently supported build.
Chromosome	Chromosome on which the marker is located. Must be a valid chromosome for the species being analyzed. Enter 0 if unknown.
Coordinate	Chromosome coordinate of marker. Enter 0 if unknown.
Source	Identify the source of the sequence and annotation data. Enter unknown if no information is available.
Source_Version	Source version number or 0 if unknown.
Sequence_Orientation	Must contain one of the following three values: forward , reverse , or unknown . Information is customer-supplied and is not validated.
Region_Description	Description of the region of interest.
Final_Score	Ranges from 0–1.1, with higher values reflecting greater ability to design a successful assay.
Failure_Codes	If applicable, reasons why a successful assay at this marker locus is unlikely (Table 8).
Validation_Class	Numerical representation of validation_bin (Table 9).
Validation_Bin	Manner in which designed assays have been validated (Table 9).
MAF_Caucasian	
Chr_Count_Caucasian	
Study_Name_Caucasian	
MAF_African	
Chr_Count_African	
Study_Name_African	
MAF_African_American	
Chr_Count_African_American	
Study_Name_African_American	
MAF_Japanese	
ChrCount_Japanese	
Study_Japanese	
MAF_Chinese	
Chr_Count_Chinese	
Study_Name_Chinese	
MAF_Other	
Chr_Count_Other	
Study_Name_Other	

Minor allele frequency from the largest peer-reviewed study conducted, the study size in terms of number of chromosomes, and the study type. Data are retrieved from dbSNP for each population:

- Caucasian
- African
- African-American
- Han Chinese
- Japanese
- Unknown

TABLE 8: SCORE FILE COLUMN HEADERS CONTINUED

HEADING	DESCRIPTION
Normalization_Bin	A, B, or C.
Bead_Types/Assay	1 for Infinium II or 2 for Infinium I.
Assay_Type	Infinium I or Infinium II.
ILMN_ID	Unique identifier assigned by ADT for the designed assay.
Gene_ID*	Gene ID number from NCBI.
Gene_symbol	HUGO identifier.
Accession*	RefSeq Accession number.
Location*	Structural location of the marker: intron, coding, flanking_5UTR, flanking_3UTR, 5UTR, 3UTR, UTR.
Location_relative_to_gene*	If the marker does not fall within an exon, the value is the actual base pair distance from gene start. The absolute value of this number is the distance to the closest transcript. The negative sign is a formatting symbol and is not meant to imply strand or direction. If the SNP is within an exon, two values separated by a '/' are given, which represent distances to the exon-intron boundaries.
Coding_status*	NONSYN or SYNON. If the marker falls within an exon, this field notes a synonymous or nonsynonymous amino acid change.
Amino_acid_change*	If the marker falls within an exon, this field notes the actual change to the amino acid, followed by the GenBank protein sequence used in numbering the change.
Id_with_mouse*	Ratio of identical bases within 60 bp of flanking sequence compared to mouse sequence that has been aligned with the homologous human sequence and covers the marker in question.
Phast_conservation*	Metric used by the UCSC Genome Browser to identify highly conserved markers among species.
Design_Date	Date of design.

*Additional gene annotation only in Score output file from submitted GeneList, RSList, and RegionList files.

markers in each of three **Normalization_Bins**: A, B, and C. Bin C assays are Infinium II designs. Bin A and B assays are Infinium I designs. Infinium I assays are classified into one of these two bins based on the sequence context (the base pair after the locus) and strand that the assay is designed against. For a final order, if there are any loci in a given normalization bin, there must be at least 190 loci in that bin to ensure normalization functions as expected. If Bin A or Bin B needs to be supplemented to reach 190, submit additional A/T or C/G SNPs to ADT for scoring. ADT will report whether each of these SNPs are in Bin A or Bin B. The appropriate assays can then be added to the original design as needed.

Of particular importance for ordering is the Number of Bead Types value, since iSelect BeadChip orders are priced based on the number of attempted bead types. The number of bead types may be different than the number of assays because Infinium I assays require two bead types per markers and Infinium II assays only require one bead type per marker.

Following the Score file header section, detailed information for each marker is listed in the data section. All Score file data section column headers are described in Table 8. Important performance values are also presented for each SNP. The **Final_Score** indicates the expected success for designing a given assay, and may be supported with **Failure_Codes** for further information. Validation status is also indicated to provide even greater confidence in design success. To assist researchers in ordering the most applicable markers for their studies, minor allele frequencies (MAFs) in several populations are provided for SNPs when available from dbSNP. MAF from the largest study is reported, and is qualified based on peer-reviewed publication, study design and size, and verified results.

FILTERING AND SELECTING CUSTOM LIST

In addition to being an output file format, Score files can be used as input files to ADT. Thus, users can easily cre-

TABLE 9: LIST OF FAILURE CODES FOR THE ADT

CRITICAL FAILURES (UNDESIGNABLE)	
101	Flanking sequence is too short.
102	Formatting error. Variant must match the format [A/B]. Possible causes: <ul style="list-style-type: none"> • Spaces in submitted sequence. • More than one set of brackets in sequence. • Missing brackets around variant. • Variant alleles not separated by "/".
103	TOP/BOT strand cannot be determined due to low sequence complexity.
104	Variant is not appropriate for Illumina platform. Possible causes: <ul style="list-style-type: none"> • Tri- or quad-allelic variant. • Variant contains characters other than A, G, C, and T.
106	Degenerate nucleotide(s) in assay design region (e.g., W, R, S, N).
WARNINGS (DESIGNABLE)	
301	Variant in duplicated/repetitive region.
399	Multiple contributing issues.

TABLE 10: VALIDATION STATUS DESCRIPTIONS

VALIDATION_BIN	VALIDATION_CLASS	DESCRIPTION
Infinium II validated	5	Variant has been previously designed and successfully generated polymorphic results using the Infinium II Assay.
Infinium I validated	4	Variant has successfully generated polymorphic results with the Infinium I Assay.
Two-hit validated	2	Both alleles have been seen in two independent methods and populations.
Non-validated	1	Variant has been seen in only one method or population. Even if it has a high design score, there is an increased chance that it is monomorphic.

ate a filtered or edited output file (with markers removed or added) for iterative ADT analysis while determining the optimal set of loci to order. Markers identified using more than one input search method (e.g., Gene, Region, Identity, Sequence, CNVRegion, or ExistingDesign) can be combined as one Score file and resubmitted to ADT as an input file for evaluation as a single product.

Illumina recommends applying the following criteria for discriminating marker lists to create a successful product that achieves the scientific aims of the experiment and has the highest chances of generating meaningful results.

- 1) Remove markers that cannot be ordered (error codes in the 101–199 range).
- 2) Consider research requirements (e.g., tags, spacing, or MAF).
- 3) When appropriate, favor Infinium-validated markers, since they have the highest chance of converting into functional assays.
- 4) Use two-hit validated markers based on **Final_Score**. Higher Final_Scores are preferred.
- 5) Do not be hindered by proximity limits for Infinium assays. Markers can be close to each other without potential interference.
- 6) Submit a number of bead types that is within the required range. Human1M-Duo+ or Human550-Quad+ BeadChips require 3,000 to 60,800 attempted bead types. iSelect HD BeadChips can accommodate 3,000 to 200,000 attempted bead types.

TABLE 11: HEADER SECTION FOR FINAL ORDER FILE

HEADING	DESCRIPTION	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	No
Company_Address2	Line 2 of customer's address (optional)	No
City	Customer's city	No
State/Province	Customer's state or province	No
Postal_Code	Customer's postal code	No
Country	Customer's country	No
Phone_Number	Customer's phone number	No
Fax_Number	Customer's fax number	No
Email_Address	Customer's email address	Yes
Order_Description	Description of work (30 character limit)	No
Order_Comments	Additional comments (optional)	No
Platform_Type	WGGT	Yes
Number_of_Assays	Number of loci in file	No
Lowercase_weighting	0 or 1	No
Design_Iteration	Final	Yes - for final file
Scale(Number_of_Tubes)	Must be 1 or greater	Yes - for final file
Purchase_Order_Number	Customer purchase order number	Yes - for final file
Product_Name	Text to appear on the product packaging	Yes - for final file
Format_Type	Score	Yes
Normalization_Bin_A	Either 0 or ≥ 190 loci	Yes
Normalization_Bin_B	Either 0 or ≥ 190 loci	Yes
Normalization_Bin_C	Either 0 or ≥ 190 loci	Yes
Number_of_Bead_Types	Must be within limits of quotation or contract (not number of loci)	Yes
Species	Species name.	Yes

FINAL SCORE FILE

After selecting SNPs that meet your research criteria and the ADT criteria, a final Score file must be created to place an order. A preliminary Score file is converted to a final Score file by the completion of four header rows (white in Table 10): **Design_Iteration**, **Scale(Number_of_Tubes)**, **Purchase_Order_Number**, and **Product_Name**. It is important to ensure that the Number of Bead Types (not

SNPs) value in the final file matches the number on the corresponding quotation or contract.

ORDERING

To place an order, submit the final Score file on the iCom website (<https://iCom.illumina.com>) (Figure 9). Researchers can also reorder an existing iSelect panel or obtain customized project pricing information before

FIGURE 9: UPLOAD ISELECT DESIGN FILES WEB PAGE



Use the Upload iSelect File web page to input design files.

placing an order. All of these options are available after logging in to iCom. To get an iCom account, contact Illumina Technical Support².

Upload iSelect HD Design Files

To upload files to iCom, follow these steps:

- 1) On the iCom welcome page, select **Orders | Upload iSelect HD Files**.
- 2) Click **Upload Final Score File**.
- 3) Select the File Type (Score or ExistingDesign).
- 4) Select the Assay Type (WGGT).
- 5) Select the Species.
- 6) Enter the Number of Assays, which is the number of different markers or variants (e.g., SNPs or CNVs) being interrogated in the custom project, not the number of bead types.
- 7) Enter the Description (user-defined and optional).
- 8) Enter the location and file name of the target file or click **Browse** to navigate to the file.
- 9) Click **Upload and Continue with Validation**. The uploaded file will then be listed in the iSelect Files table, with an initial status of "In Scoring." Within two days, ADT scoring will be completed, and an email notification will be sent. At that point, the file status will be updated to "Ready to Order" or "Warning." See Table 8 for a description of possible warning codes.
- 10) After a file is uploaded and successfully validated, it can be used for product order placement by selecting the checkbox next to the file and clicking **Continue to Project Pricing**.

iSelect HD Project Pricing Calculator

The Project Pricing Calculator can be used at any time. It opens automatically when you select **Continue to Project Pricing** for a design file as described above. To get project pricing information at any other time, go to the iCom web page and select **Tools | iSelect Project Pricing Calculator** or **Order Now**. On the next screen, click **Infinium Genotyping | Project Wizard**.

- 1) Select the product type (Figure 10).
- 2) If a design file is already selected, skip to step 4.
- 3) Do one of the following:
 - To determine pricing based on a design file, click **Select Files** and either select an existing file, or upload a new file by clicking **Upload File** and following the instructions in the previous section.
 - After a file is selected, click **Continue to Project Pricing**.
- 4) To determine pricing based on an assumed number of bead types rather than a design file, enter the number of bead types to assume for the pricing calculation.
- 5) Enter the number of samples. The number of samples must be between 1,152 and 116,352, in multiples of 48.
- 6) Enter the product name. A product name is required when using a design file, but is optional if pricing is to be based only on the number of bead types. The product name must have fewer than 30 characters and contain only alphanumeric, underscore, or dash characters.
- 7) Click **Calculate Price** to display pricing information for the selected design specifications.
- 8) To place an order based on the selected design file and number of samples entered in the pricing calculator,

FIGURE 10: ISELECT PROJECT PRICING CALCULATOR WEBPAGE



Use the iSelect Project Pricing Calculator to get pricing information based on an uploaded design file or for a hypothetical product based on a specified number of bead types.

click **Preview**.

- 9) Click **Add to Cart** to add the product to the cart and complete the order.

Reordering an Existing Bead Pool or Product

To place an order based on an existing product or custom bead pool, click **Order Now**, and then click **Reorder iSelect**. Reorders have a minimum order requirement of 288 samples, and must be ordered in multiples of 48. Only non-expired BeadChips and bead pools are available for reordering. Bead pools expire after 12 months. Unused final design files are removed 90 days after being uploaded. If a design file has been used to place an BeadChip order, that design file is never removed.

Two types of reorders can be placed for custom products.

- To order a custom product identical to one that has already been ordered:
 - 1) Select the desired custom product in the list.
 - 2) Click **Reorder** to go to the iSelect Project Calculator.
 - 3) In the iSelect HD Project Calculator, select the sample number desired.
 - 4) Click **Preview** to view the order summary.
 - 5) Click **Add to Cart** to add the reordered product to the cart, and complete the order as described in the next section.
- To order a new product using an existing pool (e.g., to order a semi-custom Human1M-Duo+ BeadChip using the custom pool from a prior iSelect BeadChip):
 - 1) Select **Reorder iSelect**.
 - 2) Click on **Order from existing bead pool**.
 - 3) Select the file that is associated with the bead pool.
 - 4) Select the bead pool.
 - 5) Select the product type.
 - 6) Enter the number of samples.
 - 7) Click **Preview** to view the order summary.
 - 8) Click **Add to Cart** to add the product to the cart, and complete the order as described in the next section.

Completing an Order

Once all desired orders are added to the shopping cart, click **Continue to Checkout**. The shopping cart is automatically displayed after clicking **Add to Cart**; otherwise click **View Cart**. During the checkout process, shipping and payment information is required. An initial shipping date and partial shipment request can be entered before final submission. After clicking **Submit**, the order will be sent to Illumina Customer Solutions for review. If needed, customers can make modifications to an order by emailing orders@illumina.com during the 48 hour review period.

Shipping Schedule

Shipping schedules can be defined after an order is placed. To specify a shipping schedule, fill out the Ship Schedule Form (available from a link in the order confirmation or directly from <https://icom.illumina.com/EAI/ShipScheduleForm.xls>) and email it within two weeks of the order confirmation date to orders@illumina.com. Otherwise, orders may ship complete or according to a default schedule. The first shipment of custom BeadChips should arrive within eight to ten weeks after order confirmation.

Projects with fewer than 12,000 samples must be shipped complete within 6 months. Projects of 1,152 to 2,304 samples have a maximum of 2 shipments, 2,305 to 5,760 samples have a maximum of 3 shipments, and 5,761 to 11,999 have a maximum of 4 shipments. Projects of over 12,000 samples must be shipped within 12 months and a weekly shipping schedule can be arranged.

Ordering By Email

To place an iSelect HD BeadChip order by email rather than through iCom, submit the final Score file and an official purchase order to orders@illumina.com. Illumina will send a confirmation and contact you if the final list lacks essential information or contains unorderable designs.

SUMMARY

Custom Infinium DNA Analysis products by Illumina allow researchers to create assays tailored directly to their specific needs for targeted region genotyping or fine-mapping of candidate disease association regions. The ADT provides a simple and powerful method for evaluating individual loci and creating the most successful custom genotyping BeadChips. By following the guidelines in this technical note, researchers can ensure that their orders are designed and placed quickly and easily.

REFERENCES

- (1) <https://icom.illumina.com/>
- (2) To contact Technical Support, send email to techsupport@illumina.com or call 1.800.809.4566.

ADDITIONAL INFORMATION

Visit our website or contact us to learn more about iSelect Custom Infinium products from Illumina.

Illumina, Inc.
Customer Solutions
9885 Towne Centre Drive
San Diego, CA 92121-1975
1.800.809.4566 (toll free)
1.858.202.4566 (outside the U.S.)
techsupport@illumina.com
www.illumina.com

FOR RESEARCH USE ONLY

© 2009 Illumina, Inc. All rights reserved.
Illumina, IlluminaDx, Solexa, Making Sense Out of Life, Oligator, Sentrix, GoldenGate, GoldenGate indexing, DASL, BeadArray, Array of Arrays, Infinium, BeadXpress, VeraCode, IntelliHyb, iSelect, CSPRO, and GenomeStudio are registered trademarks or trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners.
Pub. No. 370-2007-021 Current as of 13 August 2009

