

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 upstream of the gene starting coordinate.
Bases_Downstream	Number of bases downstream of the gene ending coordinate.

to identify gene regions. The sizes of upstream and downstream regions queried by ADT are specified by the user. Loci 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 input file. Figure 2 provides examples of proper Gene entries in Excel.

Region File

The Region file type provides a method for selecting loci between specified locations of a human chromosome. A Region file contains a list of regions in the human genome identified by chromosome and coordinate range which ADT will search and evaluate from among cataloged markers in the current Illumina-internal version of dbSNP. This internal database does not contain MNPs, SSRs, or SNPs with ambiguous or multiple localizations. SNPs with a source molecule type of cDNA in dbSNP are also not included, as these may result in primers being inadvertently designed across intron-exon boundaries, resulting in a non-functional assay design. 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. 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 File

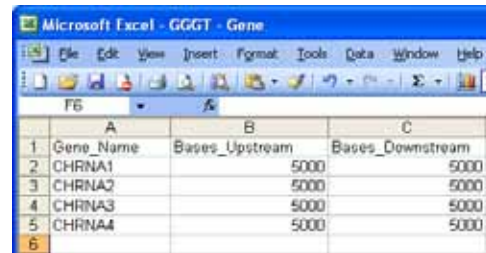
Known loci described in the current version of dbSNP can be requested specifically using the Identity file type. A current internal version of dbSNP is the source for rs loci and flanking sequence data. The column headings and description information shown in Table 3 must be provided in the Identity input file. Figure 4 provides an example of properly formed Identity entries.

Sequence File

The Sequence file format provides a method for evaluating loci 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.

To specify a locus, 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 a single minus sign indicating the deletion from the bases representing the insertion (TGC[-/AT]CCG). A minimum of 50 bp of sequence on either side of the variant is required; however, 60 bp flanking sequence is preferred. ADT will also accept IUPAC codes for degenerate bases

Figure 2: Gene File Format Examples

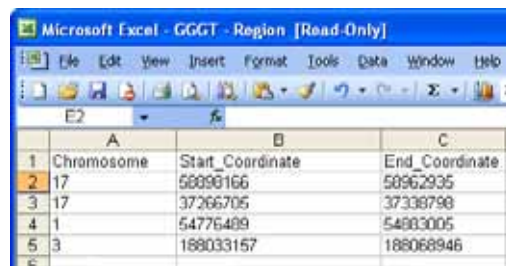


Example of properly formed entries in a Gene file shown from Excel.

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; 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 an example of properly formed Sequence entries.

Figure 3: Region File Format Examples



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

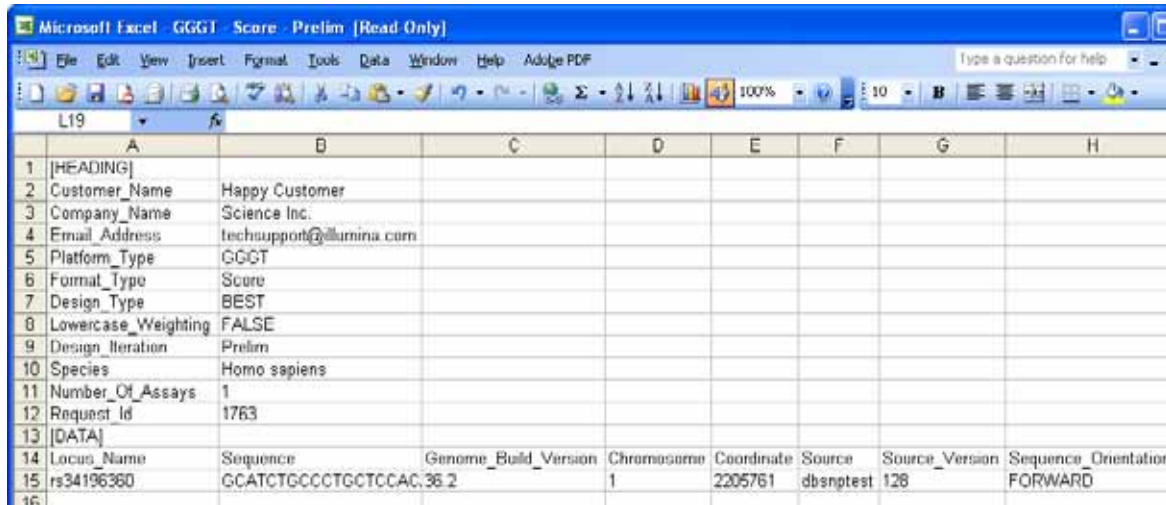
Table 5: Column Headings for Existing Design File

Heading	Description
llmn_id	llmn_id from original design manifest.

Table 6: 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
Email_Address	Customer's email address	Yes
Platform_Type	GGGT	Yes
Format_Type	Gene, Region, Sequence, Identity, ExistingDesign, or Score	Yes
Design_Type	BEST, OTHER, TOP, BOT, PLUS or MINUS	No
Lowercase_Weighting	FALSE	No
Design_iteration	prelim	Yes
Species	Species Name	Yes
Number_of_Assays	Number of loci in file (may be 0 for Gene, Region, Identity, and Sequence files if the number of loci is unknown)	No
VeraCode_Pool	FALSE	No
Request_Id	Unique ADT identification number for file submission	No
Is_MSI	FALSE	No

Figure 7: 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. This header is only required for email submissions, and is formatted the same way for any preliminary file format.

