

NimbleGen Design File (NDF)

Version 2005_01_20

A NimbleGen design file contains 17 columns. For a standard 1:4 design, the NDF file will contain approximately 196000 rows. For a standard 1:2 design, the NDF file will contain approximately 393000 rows. The information is tab delimited with a single header line containing the field names. The columns can be in any order.

The columns are:

Field Name	Description	Requirement
PROBE_DESIGN_ID	This a unique, composite key consisting of the DESIGN_ID, ROW_NUM and COL_NUM.	Supplied by database
DESIGN_ID	This is the NimbleGen identifier for the design.	Supplied by NimbleGen
CONTAINER	NimbleGen arrays are divided into containers. One method of using containers is to divide the array into quadrants with reference marks.	Supplied by ArrayScribe. Limited to 50 characters
DESIGN_NOTE	A comment field suitable for placing information necessary for data analysis, for instance if you want to analyze sets of probes and/or genes together that don't separate out using SEQ_ID, PROBE_ID, SELECTION_CRITERIA, CONTAINER, or other information.	Optional – can be used if you want to analyze data using criteria other than CONTAINER (rarely necessary) Limited to 100 characters.
SELECTION_CRITERIA	Generally contains information about how a probe was selected. For rank selection, contains rank, uniqueness, and frequency. For older designs, will contain a criteria category.	Optional – necessary only if you want to evaluate probe sets at different levels of selection criteria. Limited to 100 characters.
SEQ_ID	The NimbleGen sequence identifier. Used to group the probe pairs together to generate average difference values. An NGS SEQ_ID is a 17 character string that looks like: <i>HSAP0001S00001834</i>	Required – must be unique for each sequence/region of interest. Limited to 50 characters

	<p>The first four letters (HSAP) are the species code.</p> <p>The next four characters (0001) are the sequence build number.</p> <p>The 'S' is the designator for sequence (so there's no confusion with the similar looking PROBE_IDs)</p> <p>The last eight digits are the unique sequence number within the sequence build</p>	
POSITION	Position of the PROBE_SEQUENCE in the sequence/region of interest, starting from the left/5' end.	Optional – useful for data analysis. Integer
PROBE_SEQUENCE	The DNA sequence synthesized on the array	Required. Limited to 200 characters
MISMATCH	The mismatch index of the PROBE_SEQUENCE. This will be 0 (for the perfect match probe) , 1 for the first mismatch, 2 for the next, etc... (for the corresponding mismatch probe).	Required for expression arrays. 0 for perfect match, other positive integer for mismatch. Generated by ArrayScribe. May be over-ridden by users if MISMATCH in probe file >= 10,000.
MATCH_INDEX	Integer number that ties probe pairs together. Using the combination of MATCH_INDEX and MISMATCH you can retrieve and distinguish the members of the probe pair. Many controls have a MATCH_INDEX of '0' since they are not paired with a mismatch probe.	Required for expression arrays with mismatches. Must be unique for each probe pair. Integer. Generated by ArrayScribe. May be over-ridden by users if MATCH_INDEX in probe file >= 1,000,000
FEATURE_ID	Unsigned Integer that uniquely identifies a feature. A feature is the set of probes on the array that are to be considered as one entity. A 4:9 array will have 4 probes with the same FEATURE_ID	Integer. Must be unique within a design. Generated by ArrayScribe.

COL_NUM	The X coordinate of the feature on the image	Integer. Supplied by user if specifying coordinate position. Otherwise is the position of the probe in the container.
ROW_NUM	The Y coordinate of the feature on the image	Integer. Supplied by user if specifying coordinate position. Otherwise is the position of the probe in the container.
X	The DLP X coordinate of the feature on the image	Integer. Supplied by ArrayScribe.
Y	The DLP Y coordinate of the feature on the image	Integer. Supplied by ArrayScribe.
PROBE_CLASS	For internal use. Generally 'experimental' or 'control' or 'fiducial'.	'fiducial' is mandatory for those features used for extraction. Limited to 20 characters
PROBE_ID	<p>The NimbleGen probe identifier. A PROBE_ID is a 17 character string that looks like: <i>HSAP00P0001724033</i></p> <p>The first four letters (HSAP) are the species code.</p> <p>The 'P' is the designator for probe (so there's no confusion with the similar looking SEQ_IDs)</p> <p>The last ten digits are the probe number.</p>	Limited to 50 characters.