

NimbleScan v2.3 User's Guide Addendum

DNA Methylation Analysis

This addendum assists you with installing and using the DNA methylation analysis tools within NimbleScan™ v2.3 software:

- Find-Peaks analysis – Identifies regions of significant positive enrichment for DNA methylation data as peaks.
- Map-Peaks analysis – Associates identified peaks to genomic features, such as transcription start sites or CpG islands.

Analysis	
Auto Align	Ctrl+A
Manually Align	Ctrl+M
Histogram/Profile	Ctrl+B
Reports	▶
CGH	▶
ChIP	▶
Expression RMA...	Ctrl+Alt+E
Methylation	▶
Sequencing PBC...	Ctrl+Alt+P

Installing the DNA Methylation Analysis Tools

If you do not see a Methylation selection within NimbleScan's Analysis menu, install the DNA methylation analysis tools:

1. Go to <http://www.nimblegen.com/software/methylation/install.html>.
2. Follow the on-screen instructions to download and install the DNA methylation analysis tools.

Analyzing Your DNA Methylation Data

Running an Analysis

To analyze your DNA methylation data in NimbleScan v2.3:

1. Open your microarray image file (.tiff), design file (.pos), and optionally the gene description file (.ngd). (See Chapter 2 in the *NimbleScan v2.3 User's Guide*.)
2. Adjust the image's appearance to better view features. (See Chapter 2 in the *NimbleScan v2.3 User's Guide*.)
3. Align the grid to the array image. (See Chapter 3 in the *NimbleScan v2.3 User's Guide*.)
4. Create pair files (.txt) for your data. (See Chapter 5 in the *NimbleScan v2.3 User's Guide*.)
5. Create scaled log₂-ratio data. (**Analysis -> ChIP -> Chip**; see Chapter 6 in the *NimbleScan v2.3 User's Guide*.)
6. Run the Find-Peaks analysis (**Analysis -> Methylation -> Find Peaks**). See "Find-Peaks Analysis" on the next page for background information on this analysis tool.
7. Run the Map-Peaks analysis (**Analysis -> Methylation -> Map Peaks**). See "Map-Peaks Analysis" on page 4 for background information on this analysis tool.

Find-Peaks Analysis

Find-Peaks analysis identifies regions of significant positive enrichment in ChIP-based methylation microarray data, using a modified ACME algorithm for peak identification as reported by [Scacheri, et al. Methods Enzymol. 2006; 411:270-82](#). The analysis accepts normalized, scaled \log_2 -ratio data in GFF format.

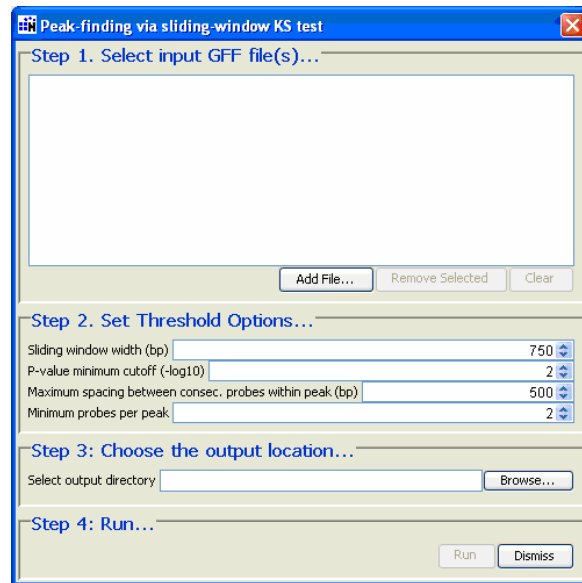
Find-Peaks analysis does the following:

- 1) Positions a fixed-length window around each consecutive probe.
- 2) Produces per-probe scores reflecting the probability of positive enrichment at or near each probe, using a non-parametric, one-sided Kolmogorov-Smirnov (KS) test. The KS test essentially determines whether the probes in the window are drawn from a significantly more positive distribution of intensity log-ratios than those in the rest of the array.
- 3) To each peak, assigns a p-value score ($-\log_{10}$), which is the average of p-values from all probes that comprise the peak.
Note: Because of several confounding issues including multiple-hypothesis correction as well as significant correlation between nearby probes, these scores should not be interpreted as p-values, but rather as relative scores.
- 4) Applies a threshold to the p-value scores to identify peaks, or blocks of enrichment, above a certain level.
- 5) Creates a GFF file, whose data can be displayed in the NimbleGen's SignalMap™ Data Browser.

Analysis Parameters

You can fine tune your Find-Peaks analysis by choosing the desired settings in the “Peak-finding via sliding window KS test” dialog box (**Analysis -> Methylation -> Find Peaks**).

Note: If you are uncertain about the quality of your experiment, use the default settings.



- **Sliding window width (bp):** Adjusts the width of the sliding window surrounding each probe. Increasing this value tests more points around each probe against the rest of the array. This in turn gives additional statistical power for resolving positive enrichment regions at the expense of lower resolution. The default setting is 750bp.
- **P-value minimum cutoff (-log₁₀):** Probes scoring above this cutoff (default set at 2) comprise peaks. Higher -log₁₀ values increase stringency and result in fewer peaks being identified.
- **Maximum spacing between nearby probes within peak (bp):** Nearby probes equal to or less than the specified spacing are consolidated into a single peak. This parameter specifies the maximum distance over which high-scoring probes are joined to form peaks. The default setting is 500bp. Increasing the distance between nearby probes (e.g. 1000bp) within peak may merge peaks that may otherwise be identified as separate peaks.
- **Minimum probes per peak:** Specifies the minimum number (default set at 2) of probes that must be above the cutoff before the region is identified as a peak. Increasing the minimum probes per peak (e.g. 4 probes) increases the stringency of peak calling because a peak requires more probes to comprise that peak.

Analysis Results

Find-Peaks analysis can generate the following files:

- Peaks GFF (.gff) file. Contains all of the identified peaks, along with their p-values. View this GFF file and the genome annotation GFF file provided with your microarray using NimbleGen's SignalMap Data Browser or spreadsheet software, such as Microsoft Excel.

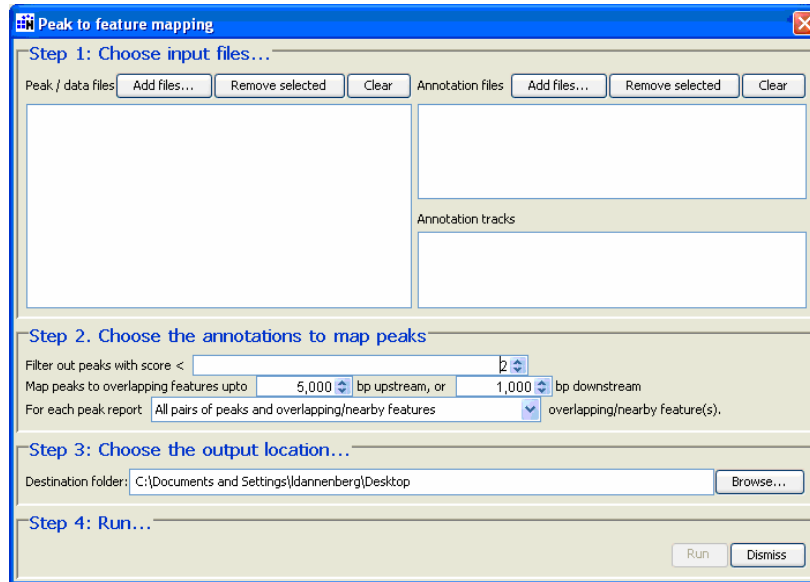
Map-Peaks Analysis

Map-Peaks analysis associates peaks to genomic features, such as transcription start sites or CpG islands.

Analysis Parameters

You can fine tune your Map-Peaks analysis by choosing the desired settings in the "Peak to feature mapping" dialog box (**Analysis -> Methylation -> Map Peaks**).

Note: *If you are uncertain about the quality of your experiment, use the default settings.*



- **Peak /data files.** Add the appropriate peak GFF files (.gff) to use when mapping peaks to genomic features.
- **Annotation files.** Add annotation GFF files (.gff) to use when mapping peaks to genomic features.
- **Annotation tracks.** Choose the track from which to map the peak. For promoter mapping, choose “transcription_start_site” to map each peak relative to the +1 site of each gene.
- **Filter out peaks with score <.** Use this parameter to eliminate peaks of low confidence from the summary report.
- **Map peaks to overlapping features up to.** This parameter sets the limit on how far upstream or downstream from a given feature the peak must be located in order to consider it a mapped peak.
- **For each peak report.** Select between these report types:
 - All pairs of peaks: Generates a report containing every combination of peak and annotation record that map together.
 - Nearest feature (if any) to each peak: Generates a report listing only the nearest feature to a peak.
 - Nearest peak (if any) to each feature: Generates a report listing only the nearest peak to a feature.

Analysis Results

Map-Peaks analysis can generate the following file:

- Summary report (tab-delimited, .xls) file. Associates peaks with genomic features, such as transcription start sites or CpG islands. View using spreadsheet software.