



Guide for Reviewing and Approving Custom Designs

NimbleGen Tiling Designs

This document describes how to review and approve the proposed custom design for a NimbleGen CGH, Chip-ChIP or Methylation microarray based on the genomic regions or sequence provided by you.

Before You Begin

The Roche NimbleGen designer will send an email with the information to access our sFTP server, including the web address, user name and password.

The zip file on the sFTP server will have the following tab-delimited design file(s):

- ▢ **<base name>_probes.txt:** This file is the probe file and includes probe sequence, probe_id and the seq_id. It is viewable in a text editor such as WordPad and Notepad or in any program capable of opening spreadsheets, such as Microsoft Excel or OpenOffice
- ▢ **<base name>_probe_locations.gff:** This file displays the genomic probe locations of each probe sequence and is viewable using Roche NimbleGen SignalMap software.
- ▢ **<base name>_stats.txt:** This file details a statistical analysis for each provided (and merged) region. It is viewable in a text editor such as WordPad and Notepad or in any program capable of opening spreadsheets, such as Microsoft Excel or OpenOffice.
- ▢ **<base name>_annotation.gff:** This file contains three tracks labeled “Tiled Regions”, “primary transcript” and “transcription start site”. The “Tiled Regions” track will display the regions you requested that are covered by probe sequences in this design. The other two tracks are based on publically available information, such as from UCSC, for your model organism. Our ability to provide this annotation file depends on public resources and information you provide. This file is viewable in SignalMap or DEVA.

Further Notes to Consider

- ▢ Regions not covered by the microarray design are usually repetitive regions, or known gaps in the reference sequence. There are pluses and minus to including repetitive regions and those decisions should be made based on your experiments. Ask your designer for advice if you need assistance in making decisions about the trade-offs of coverage versus repetitiveness/uniqueness.
- ▢ The stringency filter Roche NimbleGen uses by default will not include low copy repeats (appearing two or more times in the genome) in the microarray design. If these regions are necessary to your research, indicate this requirement in the email to Roche NimbleGen (refer to [Step 4. Approving or Requesting Changes to the Microarray Design](#) on page 6). Roche NimbleGen will use less stringent criteria in generating the microarray design on request. Be aware that the less stringent design may provide more genomic coverage but at the cost of less precise data.

- n Prior to layout, designs must be reviewed and approved by you. In order to confirm that the design will meet the needs of your experiment, you will need to review the details of the design and the files provided. Review the files downloaded from sFTP server following Steps 1-4 outlined below and confirm the accuracy of the design information provided to you via email:
 - Organism/build
 - Number of replicates
 - Probe tiling density
- n If you have questions, contact Roche Microarray Technical Support (refer to page 7 for contact details).



Mac and Linux Users: SignalMap can only be used on a Windows computer. If you do not have access to a Windows computer and your custom design uses a genome supported by UCSC Genome Browser (<http://genome.ucsc.edu/>) Roche NimbleGen can provide you a BED file which will contain the same information as the probe locations gff file. If you need a BED file, indicate this requirement in the email to Roche NimbleGen (refer to [Step 4. Approve or Request Changes to the Microarray Design](#) on page 6). See [Appendix A](#) for details on how to view the BED file.

Step 1. Visually Reviewing the Microarray Design

The two gff files mentioned above (<base_name>_probe_locations.gff, <base_name>_annotation.gff) are viewable using SignalMap. You can find instructions on how to acquire this software in the email from your designer. The probe locations files are used to visually review the probe selection and compare it to the provided annotation gff. Review the “Probe Locations” track for specific genomic locations of the probe sequences selected for your design and take note of any gaps in coverage. When viewed in SignalMap, the X-axis represents the genomic location while the Y-axis refers to how unique each probe sequence is. A Y-axis value of 1, means the probe was unique in the genome, while higher numbers indicate more matches to other closely related sequences in the genome.

Using SignalMap Software

1. Save the two GFF files from your sftp account to your computer or a network drive.
2. Open SignalMap software.
3. Select **File** -> **New** and import the two GFF files. Figure 1 is an example microarray design file displayed in the SignalMap window.

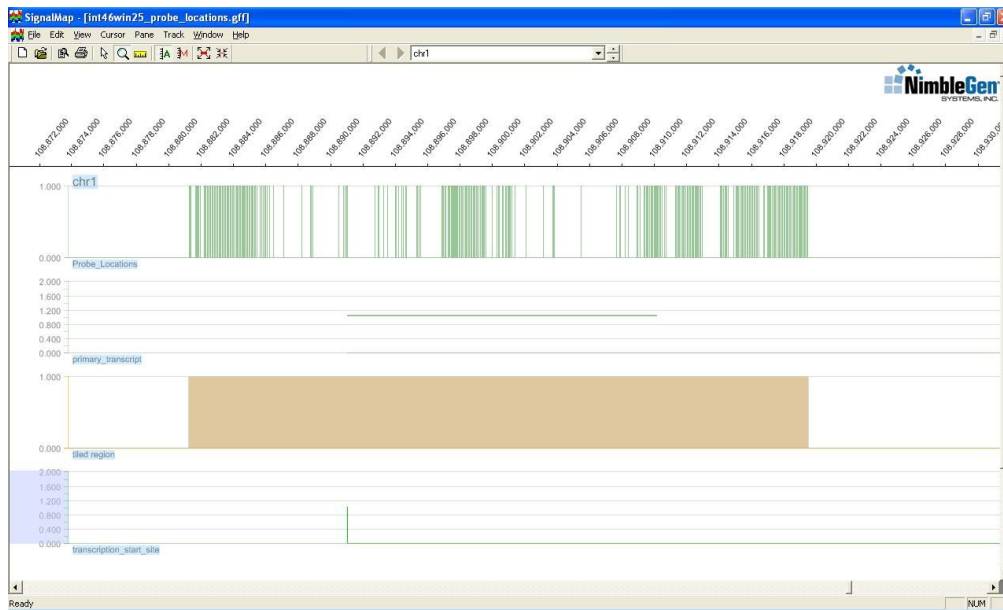


Figure 1: Example Microarray Design File Displayed in the SignalMap Window

4. Review the microarray design using SignalMap software's functions, such as the following:
 - **Zoom:** Click the magnifier button on the toolbar, and then click and drag to draw a bounding box around the region to magnify. Alternatively, hold down Ctrl and press + to zoom in or - to zoom out.
 - **Arrange chromosomes:** View data tracks in many different ways by chromosome. Use the *pane selector* field below the toolbar to change the view to either all chromosomes (All Tracks) or a selected chromosome (e.g. chr2).
 - **Go to specific genomic location:** Select **View -> Go to position**, select the chromosome, and enter the start and end positions.
5. Refer to the *SignalMap User's Guide* for additional details about the software's capabilities.

Step 2. Reviewing the Statistics File: <base name>_stats.txt

The statistics file contains information for each region provided (and merged) for your custom design. Each line lists the number of probes selected, the average probe coverage, and additional coverage statistics including the largest gap found within that region (see below for field definitions).



If there were regions without any probe coverage, they will be listed at the end of this file. See instructions at beginning of document for possible ways to improve coverage.

The regions in this file should be carefully reviewed to ensure that any gaps in probe coverage and/or missing regions do not coincide with a region important to your experiment. In addition, review the number of probe sequences selected for each region and review the final line of the report for information about the selection as a whole. A description of each column is listed below. If the results are not what you were expecting or differ from the instructions you specified in the *Design Specification Form*, contact your designer via email.

1. Open the statistics file using any program capable of opening spreadsheets, such as Microsoft Excel or OpenOffice.
2. Review each field to ensure the design meets the specifications for your design.

- Refer to the “Definitions of Summary Design File Fields” below to become familiar with what each field means.

	A	B	C	D	E	F	G	H	I	J	K
1	SEQ_ID	LENGTH	PROBES	DENSITY	MEAN_IN	MEDIAN_I	1ST_QUAF	3RD_QUAI	MIN_INTE	MAX_INTE	COVERAGE
2	chr1:12885	21453	3	7151	6161	6161	6125	6125	6125	6198	154
3	chr1:1317	25418	1	25418	-	-	-	-	-	-	-
4	chr1:1387	82901	16	5181	3632	3666	3577	4917	3511	16468	828
5	chr1:1662	47239	13	3633	3622	3623	3589	3680	3522	4056	688
6	chr1:1833	25973	8	3246	3601	3604	3592	3608	3521	3696	416
7	chr1:2498	24654	7	3522	3631	3626	3524	3630	3519	3642	357
8	chr1:6236	141552	39	3629	3609	3610	3582	3665	3529	5066	2041
9	chr1:6606	87643	24	3651	3605	3609	3595	3662	3537	4438	1325
10	chr1:7934	13679	17	3635	3610	3609	3579	3710	3561	3907	613

Figure 2: Example Summary Design File Displayed in Microsoft Excel

Definitions of Summary Design File Fields

SEQ_ID: The genomic coordinate range requested.

LENGTH: The genomic length of the region.

PROBES: The number of probe sequences the design has for a particular SEQ_ID

DENSITY: A calculation based on the length of the SEQ_ID divided by the number of probe sequences selected for that SEQ_ID.

MEAN_INTERVAL: The calculated Tukey's biweight mean base pair spacing between the start position of one probe sequence and the start position of the next probe sequence. Tukey biweight mean, which is a median weighted mean, is used to accommodate the occasional large gaps (large blocks of N's, repeat stretches, etc.) found in tiling designs.

MEDIAN_INTERVAL: The calculated median base pair spacing between the start position of one probe sequence and the start position of the next probe sequence.

1st_QUARTILE: The 25th percentile of the set of all gap intervals for a particular SEQ_ID.

3rd_QUARTILE: The 75th percentile of the set of all gap intervals for a particular SEQ_ID.

MIN_INTERVAL: The smallest gap interval for a particular SEQ_ID.

MAX_INTERVAL: The largest gap interval for a particular SEQ_ID.

COVERAGE: The number of base pairs covered by selected probes for a particular SEQ_ID.

Step 3. Reviewing the Probe File: <base name>_probes.txt

The probe file contains detailed information for each probe sequence including the actual probe sequence. Review this file if you have specific questions about the probe sequences and attributes. A description of each column is listed below.

Definitions of Probe Design File Fields

PROBE_SEQUENCE: The specific probe sequence for each PROBE_ID.

PROBE_ID: A unique alphanumeric identifier assigned to each probe sequence. It consists of three pieces of information - typically the chromosome, the strand information and the genomic start position. For example, if the PROBE_ID was CHR01FS034894307, the identifier notes that this probe sequence is on chromosome 1, is on the forward strand and begins at genomic position 34894307.

SEQ_ID: The requested genomic coordinate range. These values may be slightly different than the ones initially provided because Roche NimbleGen merges overlapping regions before selecting probes.

POSITION: The genomic start position of the probe sequence.

DESIGN_NOTE: A Roche NimbleGen internal note that will always be set to "interval rank"

SELECTION_CRITERIA: A listing of all parameters weighed in the selection of the probe sequence. Parameters are listed in this column delineated by “;” and as separate columns.

PROBE_CLASS: A Roche NimbleGen internal note that is always set to experimental for any customer requested probe sequences. This column is not used to indicate which probe sequences are customer controls.

PROBE_TM: Roche NimbleGen’s computed melting temperature for each probe sequence. This is a computational analysis and may differ from actual lab results.

FREQUENCY: Roche NimbleGen’s measure of repetitiveness. We use the average 15-mer frequency within a probe to judge whether or not a probe falls in a repetitive region. The normal cut-off for human designs is a value of 100; therefore greater than 100 is repetitive and less than 100 is okay. The scoring algorithm, however, will always try to select probes with smaller values.

RULES: A score indicating the presence of how often runs of bases appear in the probe sequence (i.e. TTTTTTT). Runs of 6 or more of either A or T result in a score of $-50 * \text{the length of the homopolymer run}$ for each instance. Runs of 4 or more of G or C result in a score of $-100 * \text{the length of the homopolymer run}$ for each instance. A score of 0, the highest score possible, means there are no homopolymer runs in the sequence.

SCORE: A combination of different bioinformatic metrics (including PROBE_TM, FREQUENCY and RULES) that are used internally by Roche NimbleGen to evaluate a particular probe sequence. This is not to be taken as an indication to how well a particular probe will perform on the array, as only an actual experiment can evaluate that. It is a unitless number that allows Roche NimbleGen to rank probes for a particular SEQ_ID.

RANK: This column is not applicable to most tiling designs. The value is set to 1 unless multiple probes are being selected for a defined region. For example, a design targeting exons where 8 probes are requested for each exon, regardless of size, rather than a fixed tiling density. In those cases, the best ranked probe will be numbered 1, and the successive probes will be numbered up to the number of probes requested for the region.

Step 4. Approving or Requesting Changes to the Microarray Design



Roche NimbleGen cannot finalize the microarray design and manufacture the Tiling microarray until approval is received.

To approve your design reply to the email that included the data deliverables files with the proposed microarray design. In the reply, indicate you approve the design.

To request changes to your design or annotation file, reply to the same email and provide information for your requested changes. You will then receive a Design Change Form which will list the requested changes and their impact on your project. This Design Change Form will need to be approved by you prior to any modifications being made to your design or annotation file.

What Happens After You Approve the Microarray Design?

Full-Service Orders:

You receive the following once your microarray design is completed:

- ▢ An email from the array designer containing the Design Name for your custom design as well as any notes that we entered into our database regarding your custom design.
- ▢ **Sample barcodes and a shipping label:** Sample barcodes will be used to label your tubes of sample prior to shipment to the Roche NimbleGen Service Lab in Iceland. The shipping label should be used to label the box that your samples are sent in to the Service Lab.
- ▢ **Passkey to the Roche NimbleGen online sample submission form (eForm):** Once your order is set up in our system, you will receive an automated email from madison.customerservice@roche.com. This email contains a link to the online sample submission form (eForm) and mailing instructions. This form will need to be completed prior to sending your samples into our Service Lab.

Array Delivery Orders:

You will receive an email from the array designer containing the Design Name for your custom design as well as any notes that we entered into our database regarding your custom design. The order will then be sent by our Customer Service team to the Manufacturing group so that it can be added to the synthesis queue. Any questions regarding the synthesis or delivery of your order can be sent to: madison.customerservice@roche.com

Design Delivery Orders:

You will receive an email from the array designer containing the Design Name for your custom design as well as any notes that we entered into our database regarding your custom design. The Design Name can then be used for placing future orders with Roche NimbleGen. We will also send a data CD to you that contains all the deliverables associated with this order (this is generally the main probe file (.ndf), positions file (.pos), probe locations file (.gff) annotation file (.gff), and layout file (.ncd)).

Customer Service

Any questions regarding the synthesis or delivery of your order can be sent to:
madison.customerservice@roche.com

Technical Support

If you have technical questions, contact your local Roche Microarray Technical Support. Go to:
www.nimblegen.com/arrayssupport for contact information

Appendix A. Using UCSC Genome Browser to Review the Microarray Design with BED files

1. Save the BED file provided via email to your computer or a network drive.
2. Open the UCSC Genome browser homepage: <http://genome.ucsc.edu>.
3. In the left pane of the UCSC Genome browser home page, click **Genome Browser**.
4. In the Genome Browser Gateway page, choose the appropriate options in the *genome* and *assembly* list boxes for your microarray design (e.g., *Human* and *Feb. 2009 (GRCh37.hg19)*).
5. Click the **add custom tracks** button.

*Tip: If the Gateway page displays the **manage custom tracks** button, select the **Click here to reset** link to display the **add custom tracks** button.*

6. In the Add Custom Tracks page:
 - a. Click the **Browse** button to select and upload the BED file.
 - b. Click the **Submit** button.
7. In the Manage Custom Tracks page, click the **go to genome browser** button. To view your regions see the 'User Track' section. Figure 3 shows an example microarray design displayed in the UCSC Genome browser.

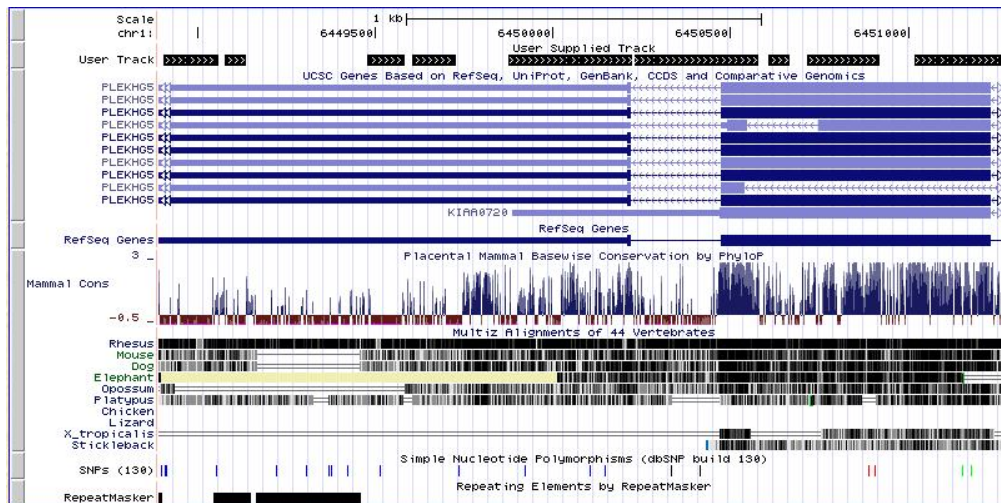


Figure 3: Example Microarray Design File Displayed in the UCSC Genome Browser

8. Review the microarray design using the UCSC Genome browser's functions, such as the following:
 - **Zoom:** Click the **zoom in** and **zoom out** buttons to zoom in or out on the center of the annotation tracks window by 1.5-, 3-, or 10-fold.
 - **Scroll:** Click the **move** buttons to scroll to the left or right.
 - **Display a different position in the genome:** Enter the coordinates in the *position/search* text box and click the **jump** button.
 - **View base composition:** Click the **base** button to view the base composition of the sequence underlying the current annotation track display.
9. Click the **Help** link for additional details about the browser's capabilities.



For life science research only.
Not for use in diagnostic procedures.

NIMBLEGEN and DEVA are trademarks of Roche.
All other product names and trademarks are the property of their
respective owners.

Published by:
Roche NimbleGen, Inc
504 S. Rosa Road
Madison, WI 53719
USA
www.nimblegen.com

© 2012 Roche NimbleGen, Inc. All rights
reserved.

06499503001 | 03/12