NimbleGen SeqCap EZ Exome Library v2.0 – Design and Annotation Files

Release Notes

August 2010

The design and annotation files provide information about genomic regions covered by the capture probes and the genes included in these regions. These files were designed for use with the following Roche NimbleGen products:

- SeqCap EZ Human Exome Library v2.0, 4 Reactions (Catalog No. 05 860 482 001)
- SeqCap EZ Human Exome Library v2.0, 48 Reactions (Catalog No. 05 860 504 001)

Introduction

The SeqCap EZ Exome Library v2.0 product covers more than 30,000 genes in the human genome. The following sources provided information about the genes:

- NCBI Reference Sequence (RefSeq) RefGene from UCSC (January 2010)
- CCDS from NCBI (September 2009)
- miRNAs from miRBase (version 14, September 2009)
- Customer inputs

All the genome coordinates were based on human genome build GRCh37 (hg19).

For RefSeq genes, only transcripts with an "NM_" prefix were selected, and only protein coding parts of the transcripts were targeted. For exons that are smaller than 100 bp, Roche NimbleGen extended the target region to 100 bp.

The total size of the target regions is 36.5 Mb. Roche NimbleGen selected 2.1 million long oligo probes to cover the target regions. Because some flanking regions are also covered by probes, the total size of regions covered by probes is 44.1 Mb, larger than the initial target regions.

In the file descriptions provided in the next section, "target regions" refer to the 36.5 Mb targets selected from various databases, and "probe-covered regions" refer to the 44.1 Mb regions covered by long oligo capture probes.

File Descriptions

All the genome coordinates were based on human genome build GRCh37 (hg19).

The Target_Regions folder contains these two files:

- SeqCap_EZ_Exome_v2.gff: There are two tracks in this .gff file. The primary_target_region track displays the 36.5 Mb target regions, and the capture_target track displays the 44.1 Mb probe-covered regions.

  The GFF files can be opened using SignalMap software (Roche NimbleGen, *www.nimblegen.com/products/software/index.html*).

- SeqCap_EZ_Exome_v2.bed: There are two tracks in this .bed file. The target_region track displays the 36.5 Mb target regions, and the tiled_region track displays the 44.1 Mb probe-covered regions.

  The BED file can be displayed as a custom annotation track using the UCSC Genome browser (*http://genome.ucsc.edu/*).

The Annotations folder contains these four files:

- SeqCap_EZ_Exome_v2_annotations.xls: This Microsoft Excel file lists the genes and miRNAs that are targeted by the design. There are three worksheets in the file, listing RefSeq genes, miRNA genes, and other genes. The other genes include customer provided genes (using the UCSC Genes database) and CCDS genes that are not in the RefSeq database. Most column headers are self-explanatory. Two of the columns provide information about how well the specific exon target is covered by the capture probes:

  - *ARRAY COVERAGE*. Percentage of target base covered by probes. Be aware that a region not covered by probes can still be captured if its neighboring regions are covered by probes. Refer to the following description of the *ARRAY COVERAGE W 100BP EXTENSION* column for more information.

  - *ARRAY COVERAGE W 100BP EXTENSION*. Percentage of target base covered by probes or located within 100 bp to one or more probes. Because the DNA fragments captured by the SeqCap EZ Exome Library are generally greater than 200 bp, sequencing results typically show sufficient coverage of 100 - 200 bp flanking regions at both sides of a region targeted by probes. Therefore, coverage with 100 bp extension is a better estimate of how much of the target region will receive sequence coverage.

- SeqCap_EZ_Exome_v2_RefSeq.gff, SeqCap_EZ_Exome_v2_miRNA.gff, SeqCap_EZ_Exome_v2_other.gff: There is a single track in each of these .gff files. The track lists the original coordinates of the exon targets as determined from the various databases. The other genes include customer provided genes (using UCSC Genes database) and CCDS genes that are not in the RefSeq database, These files can be loaded into the SignalMap software, and each vertical bar represents one exon target. When using SignalMap software, move the cursor over each exon to display the accession number/sequence identifier.

Technical Support

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

05836611001-RN1008 ➊ 08/10

NimbleGen

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

Published by
Roche NimbleGen, Inc.
504 S. Rosa Rd
Madison, WI 53719 USA

www.nimblegen.com/arraysupport