NimbleGen SeqCap EZ Exome Library v3.0 – Design and Annotation Files

Release Notes

November 2011

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 v3.0, 4 Reactions (Catalog No. 06465684001)
- SeqCap EZ Human Exome Library v3.0, 48 Reactions (Catalog No. 06465692001)

Introduction

The SeqCap EZ Exome Library v3.0 product covers more than 20,000 genes in the human genome. The following sources provided information about the genes:

- NCBI Reference Sequence (RefSeq) RefGene from UCSC (GRCh37_CDS_06092011)
- CCDS.2 from NCBI GRCh37_20110422
- Vega (GRCh37_CDS_42)
- Gencode(GRCh37_CDS_v3C)
- Ensemble (GRCh37_CDS_v63)
- miRNAs from miRBase (version 16,)
- miRNAs from snoRNABase (version 3)
- 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.

More than two million long oligonucleotide DNA probes were designed to capture the target regions. Because the flanking regions of some coding exons and miRNAs are also covered by probes, the total size of the regions covered by probes is 64 Mb.

File Descriptions

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

The folder contains these three files:

■ **SeqCap_EZ_Exome_v3_primary.bed:** This file contains primary target intervals along with associated annotation IDs in BED format.

■ **SeqCap_EZ_Exome_v3_capture.bed:** This file contains capture target intervals along with associated annotation IDs in BED format. The coordinates listed here correspond to locations where capture probes were actually designed and placed. If an exon was originally targeted for capture, but probes could not be placed in that region (for example, due to highly repetitive sequences), then the coordinates would be included in the SeqCap_EZ_Exome_v3_primary.bed file but *not* included in the SeqCap_EZ_Exome_v3_capture.bed file. Annotations provided in this file assume a 100 bp theoretical padding surrounding the capture targets.

Annotation provided in the BED file is derived from Ensemble Genes (version 64) and includes IDs for RefSeq, CCDS, Ensembl, Vega, miRBase along with the associated gene_name for each interval.

The BED file can be used to create a custom annotation track for display in the UCSC Genome browser (*http://genome.ucsc.edu/*).

■ **SeqCap_EZ_Exome_v3.gff:** This file contains both primary target and capture target intervals in GFF format.

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

Technical Support

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