LncRNA Transcript Collection

-Arraystar Human LncRNA Microarray V4.0

The set of LncRNAs covered by the Arraystar Human LncRNA Microarray V4.0 is carefully constructed using the most highly-respected public transcriptome databases (Refseq, UCSC knowngenes, Gencode, etc), as well as landmark publications [1-17]. Our LncRNA database is continually being updated to ensure that all the latest annotated LncRNAs are included on the array.

1. RefSeq (Updated Aug 2015)
The Reference Sequence (RefSeq) database maintained by NCBI (http://www.ncbi.nlm.nih.gov/projects/RefSeq/) is a comprehensive and well-annotated collection of genome, transcript, and protein sequences [1]. There are 4,927 human LncRNAs in Refseq as of August 2015, all of which are included on the Arraystar Human LncRNA Microarray V4.0.

2. UCSC Known genes dataset (Known Genes 7)
The UCSC Known genes dataset (http://genome.ucsc.edu/cgi-bin/hgTables) contains predicted genes based on data from RefSeq, Genbank, CCDS and UniProt [2]. After removing small RNAs and other unrelated transcripts, the Arraystar Human LncRNA Microarray V4.0 covers 3,521 LncRNAs from UCSC Known genes.

3. GENCODE version 19
The GENCODE project is a database of annotations of all human protein-coding and noncoding genes using evidence-based gene features [3]. After analyzing the noncoding sequences and removing those unrelated to LncRNAs, we have designed probes to detect 13,332 LncRNAs from GENCODE.

4. LncRNAdb
LncRNAdb (http://lncrnadb.com/) is a database of functional LncRNAs that are connected in one way or another with eukaryotic biological function, including expression patterns, subcellular localization, etc. [6]. 122 human LncRNAs in this database are represented on the Arraystar Human LncRNA Microarray V4.0.

5. NRED
The Noncoding RNA Expression Database (NRED) (http://jsm-research.imb.uq.edu.au/nred/cgi-bin/ncrnadb.pl) includes human and mouse LncRNAs with experimental expression and ancillary information [5]. 645 human LncRNAs from NRED are covered by the Arraystar Human LncRNA Microarray V4.0.

6. RNAdb 2.0
The RNAdb (http://research.imb.uq.edu.au/rnadb/) database is archived by the Mattick group at the Institute for Molecular Bioscience (IMB) [4]. This database contains the legacy sequences and annotations for thousands of non-coding RNAs. 1,318 human LncRNAs from RNAdb are represented on the Arraystar Human LncRNA Microarray V4.0.
7. LncRNAs identified by Khalil et al.
Khalil et al. identified and characterized 3,289 large intergenic noncoding RNAs (lincRNAs) by searching for regions of chromatin methylation (H3K4me3 and H3K36me3) outside of known protein-coding loci [7]. By mapping these chromatin state data to transcriptome databases, eliminating all annotated non-lincRNA transcripts (e.g., annotated protein-coding genes, rRNAs and tRNAs), and evaluating their coding potential, 2,193 of the lincRNAs described by Khalil et al. were included on the Arraystar Human LncRNA Microarray V4.0.

8. LncRNAs identified by Cabili et al.
Cabili et al. defined a reference catalog of more than 8,000 human lincRNA genes using their RNA sequencing results and public database information [8]. 14,353 transcripts expressed from 4,662 stringently-defined human lincRNA genes were identified. 6,969 out of these lincRNA transcripts are covered by the Arraystar Human LncRNA Microarray V4.0.

9. LincRNAs identified by Iyer et al. & Clark et al.
Clark et al. used CaptureSeq to greatly improve RNA-seq coverage and support the identification of 16,453 lncRNA transcripts in 78 tissue samples. Iyer et al. integrated 7,256 RNA-seq data from 25 independent studies, including TCGA, ENCODE and others, to derive 58,648 LncRNAs [17]. 20,142 of these LncRNAs are covered by the Arraystar Human LncRNA Microarray V4.0.

10. Ultraconserved regions encoding LncRNAs (T-UCRs)
Ultraconserved regions (UCRs) are intra- and intergenic sequences greater than 200 nt in length that are 100% identical among humans, mice, and rats. 481 human UCRs were identified by Bejerano et al. [9]. A large fraction of UCRs transcribe a subset of LncRNAs, known as T-UCRs, that are aberrantly expressed in several human cancers. All T-UCRs are represented on the Arraystar human LncRNA Microarray V4.0. To help discover potential non-coding transcripts from these regions, we also designed 962 probes to target both strands of these UCRs (http://users.soe.ucsc.edu/~jill/ultra.html).

11. HOX loci LncRNAs (HOX LncRNAs)
HOX cluster genes are fundamental regulators of pattern and axis formation during animal development. Rinn et al. identified 407 transcribed regions within the four HOX loci in humans (101 HOX gene exons, 75 introns and 231 intergenic ncRNA transcripts) [10]. All of these distinct transcribed regions are targeted by probes on the Arraystar Human LncRNA Microarray V4.0. Furthermore, 68 potential LncRNAs, whose transcript units (TUs) overlap HOX cluster genes on the same or antisense genomic strand, are covered by the Arraystar Human LncRNA Microarray V4.0.

12. LncRNAs with Enhancer-like Function (LncRNA-a)
Using the human GENCODE annotation, Orom et al. identified 3,019 human LncRNAs with enhancer-like function expressed from 2,286 unique genes [11]. LncRNAs with enhancer-like function are included on the Arraystar Human LncRNA Microarray V4.0.
References: