



Operon Biotechnologies, Inc.  
2211 Seminole Drive  
Huntsville, AL 35805  
www.operon.com

## ***Homo sapiens* (human) Promoter CoRe™ Version 1.0 Probe Set**

---

Researchers interested in probing promoter areas of the human genome may find great utility using the *Homo sapiens* (human) Promoter Cooperative Research (CoRe) Version 1.0 Probe Set. Developed in a collaboration among Carsten Müller-Tidow and Hubert Serve at University of Münster, Münster, Germany and Michael McClelland, Fred Long, and Yipeng Wang at the Sidney Kimmel Cancer Center in San Diego, CA, USA.

*Homo sapiens* (human) Promoter Version 1.0 CoRe Probe Set consists of 30,178 non-redundant oligonucleotide probes representing both areas on both 'plus' and 'minus' strands of the human genome. Probe location is detailed in the Human Promoter CoRe Version 1.0 Probe Set Genelist. Researchers may find that supplementing this set with the *Homo sapiens* (human) AROS Version 4.0 Control plate is useful for their experiments. The control plate consists of positive controls, randomly-generated negative controls, stringency controls, SpotReport™ Alien™ spike controls and a tracking control for experimental validation and quality control.

Release: February, 2007

### **I. Characteristics of *Homo sapiens* (human) Promoter CoRe™ Version 1.0 Probe**

The *Homo sapiens* (human) Promoter CoRe Probe Set encompasses 30,178 oligo probes. General characteristics of the probe set are included here. Specific information on the individual probes can be accessed through Operon's Online Microarray (OMAD™) Database at [www.operon.com](http://www.operon.com).

1. Human putative promoter sequences of all known genes and sca/miRNA genes were downloaded from UCSC Golden Path Database based on NCBI Build 35.
2. The putative promoter sequences were defined as 2,000 bases upstream and 1,000 bases downstream from transcription start site.
3. Sequences were repeat-masked.
4. Probes were designed from the sense strand of the non-repeat-masked portions of the promoter regions, with these probe restrictions:
  - a. Length: 46 – 55 bases
  - b. Tm: 69 – 79 °C
  - c. Creatable by NimbleGen in 148-rounds of A, C, G, T.
  - d. Probe must not contain simple repeats (AAAAAAA, TTTTTTTT, CCCCCCCC, GGGGGGGG, GTGTGTGT, or CACACACA).
  - e. Probe must not have tendency to form secondary structure. (For every 10-base subsequence *s* in the probe, its reverse-complement *rc(s)* must not also appear.)

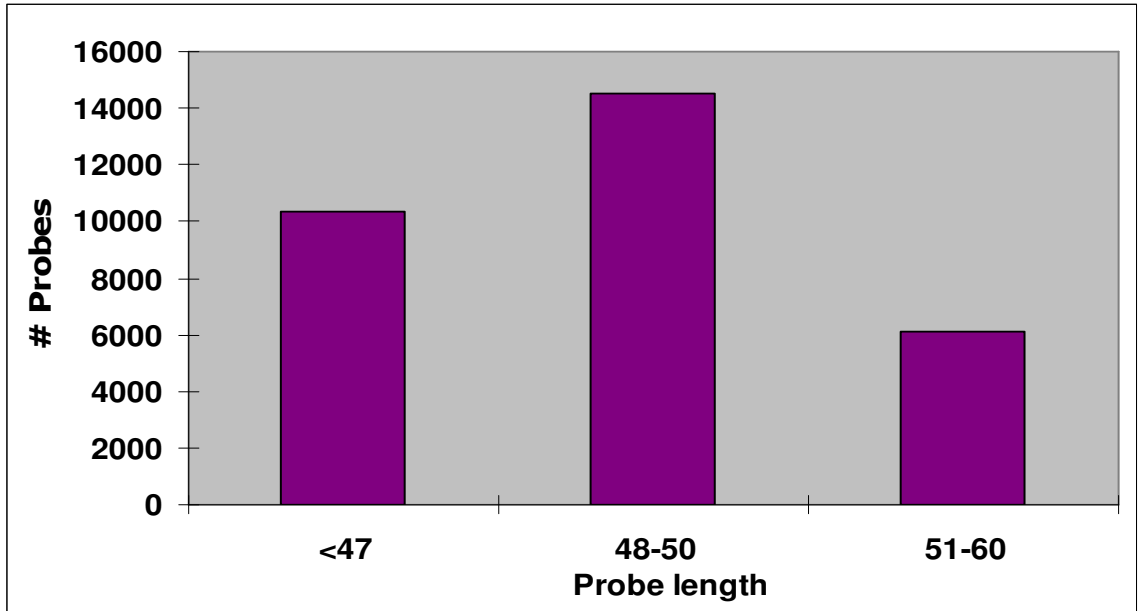
5. A second set of miRNA probes was created with a slightly looser length restriction of 40 – 60 bases, to be used for hard-to-cover miRNAs.
6. The number of probes was reduced by choosing the closest probe to every position in the promoter that was a multiple of 25, thus spacing the probes roughly 25 bases apart.
7. Probe specificity was screened by aligning them to the human genome using Blat (Kent, W.J. 2002. *BLAT—The BLAST-Like Alignment Tool*. Genome Research 4: 656-664). Probes that had imperfect matches in the genome of 30 bases or more were thrown out, as were probes that had more than 2 exact matches.
8. Up to 4 upstream probes and 1 downstream probe were selected for each promoter region, with an attempt made at sharing probes among overlapping promoter regions. Priority was given to probes that contained the degenerate sequence RCG or CGY, and that were found within 1000 bases upstream and 500 bases downstream of the start of transcription.
9. Genes with known functions were prioritized for oligo synthesis. Majority of genes have three representative oligos. See the following table for details.

	n. of probes	n. of genes	n. of genes with three probes	n. of genes with two probes	n. of genes with one probes
CDBox	187	66	57	7	2
HAcabox	100	34	32	2	0
miRna	811	311	230	40	41
scaRna	48	17	14	3	0
refSeq Genes	23799	8399	7476	448	475
Other Known Genes	5997	2169	1839	150	180
<b>TOTAL</b>	<b>30942</b>	<b>10996</b>	<b>9648</b>	<b>650</b>	<b>698</b>

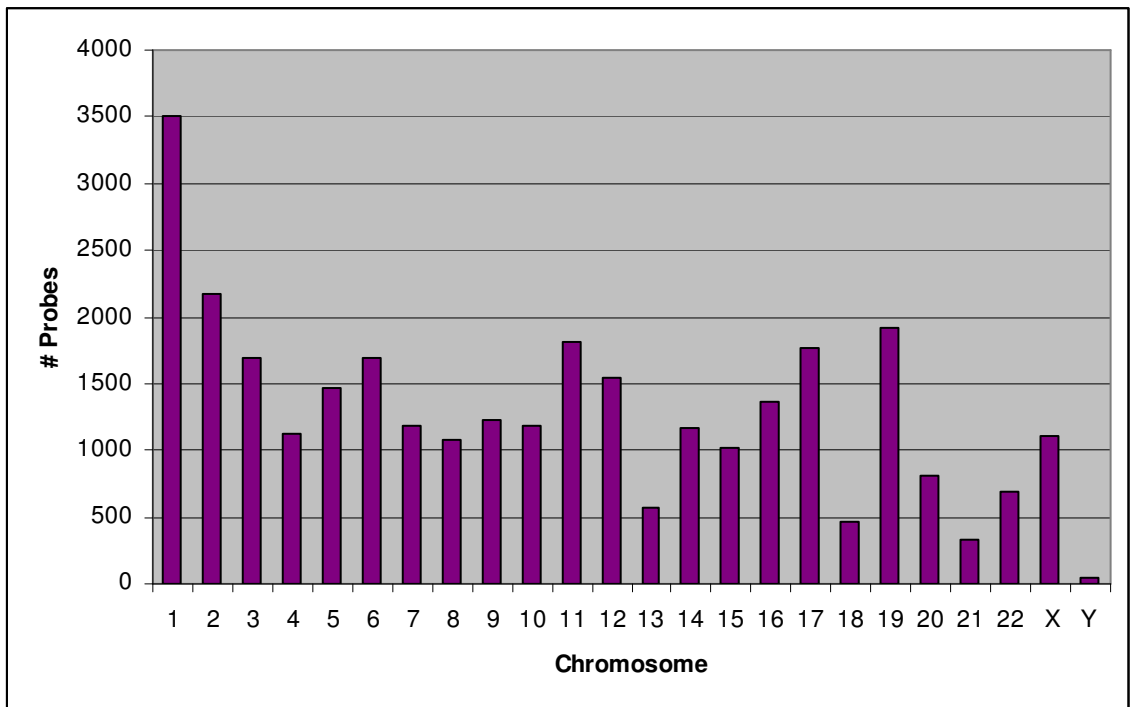
10. Within each genes, the distance between two oligos ranges from 87 to 1126 bases, with a median of 197 bases.
11. For the genes with three oligos designed, 90.2% (8703 out of 9648) of the genes have one downstream oligo and two upstream oligos.
12. 95.2% (29459 out of 30942) of probes are within 1000 bases upstream and 500 downstream of gene transcription start site.

**Figure 1. Distribution of oligo probe length**

**Average probe length = 49.4 nt**



**Figure 2. Distribution of oligo probes among the chromosomes**



**Figure 3. Distribution of Probes on Plus(+) and Minus(-) Strands**

