

## Array-Ready Oligo Set™ for the Rat Genome Version 1.1

We are pleased to announce Version 1.1 of our Rat Genome Oligo Set containing 5705 70mer probes representing 5705 *Rattus norvegicus* genes. Version 1.1 contains all 4273 probes from Version 1.0, plus 1432 probes representing 1432 additional genes (Upgrade). This set of 5705 probes (Version 1.1) represents most cloned rat genes and some ESTs. For our probe design we use state-of-the-art methods and proprietary software. An amino linker is attached to the 5' end of each oligo.

### Gene sequence source and selection

All probes are designed from the UniGene Database Build Rn 90 (Version 1.0), UniGene Database Build Rn 108 (Upgrade), and the Rat RefSeq database, developed and maintained at the National Center of Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov>).

### Advantages of using gene sequences from UniGene and RefSeq

UniGene is an open source database freely available to everyone. UniGene automatically clusters GenBank sequences into a nonredundant set of gene-oriented clusters. It has become one of the most widely used *de facto* standards in the public domain for cataloging Rat genes. Each UniGene cluster contains cloned genes and EST sequences that represent a unique gene. UniGene sequences are filtered for contaminant sequences, genomic repetitive regions, and low-complexity sequences using NCBI's Dust program. All oligos are designed from the representative sequence. This chosen representative sequence is the sequence with the longest region of high-quality sequence in each cluster.

The NCBI Reference Sequence project (RefSeq) is an effort to standardize gene sequence references by providing a NCBI-staff curated gene sequence and avoiding gene sequence redundancy. Each RefSeq is linked through a LocusID number to NCBI's LocusLink database. RefSeqs genes are quickly becoming the *de facto* standard and are used by a large research community.

The following are advantages of designing a probe directly from a gene sequence from the RefSeq collection. Each of these advantages are provided by the NCBI LocusLink interface.

Features provided for each RefSeq through LocusLink	Example
Alternate gene symbols	Gnas1, Gnpas, Gnasx
Full-length coding regions known	Present
Gene Ontology™ (GO) terms	Molecular function: GTP binding Biological process: Protein amino acid ADP-ribosylation, G-protein coupled receptor protein signaling pathway
Molecular function: GTP binding Biological process: Protein amino acid ADP-ribosylation, G-protein coupled receptor protein signaling pathway	3
Official gene symbol and name (RATMAP)	Gnas: Guanine nucleotide-binding protein G-s, alpha subunit

Number of oligos designed from sequences:	Number of oligos (Version 1.0)	Number of oligos (Upgrade)
Simultaneously represented in UniGene and RefSeq	2431	379
Represented in UniGene only	1842	1053
Total number oligos in Rat Genome Oligo Set	4273	1432

### Probe design and selection rules

Once a gene has been selected to be included in the set, a probe is selected with an optimal set of parameters. Large numbers of 70mer candidate probes for each gene are selected using the following criteria for the Rat Genome Oligo Set.

1) All oligos are within  $78^{\circ}\text{C} \pm 5^{\circ}\text{C}$ . using the following formula:

$$T_m = 81.5 + 16.6 \times \log[\text{Na}^+] + 41 \times (\#G + \#C)/\text{length} - 500/\text{length} \text{ where } [\text{Na}^+] = 0.1 \text{ M and length} = \#A + \#C + \#G + \#T$$

2) Each oligo is within 750 bases from the 3' end of the available gene sequence.

3) An oligo cannot have a contiguous single nucleotide base repeat or poly (N) tract longer than 10 bases.

4) An oligo cannot have a potential hairpin structure with a stem length longer than 9 bases.

5) A normalized score is assigned to each oligo based on the number of repeats. Oligos with more repeats having a normalized score greater than a certain threshold are filtered out.

6) Each oligo has less than or equal to 70% identity to all other genes. For all oligos in the Rat Genome Oligo Set, using BLAST, each oligo is aligned against all 59,018 representative sequences in Rat UniGene Build Rn 90 and 62,758 representative sequences in Rat UniGene Build Rn 108. Using the alignment with the candidate oligo versus the highest scoring non-self gene, a BLAST percent identity score is computed. The highest scoring non-self gene is defined as the sequence that yields the most matched bases in an alignment. This BLAST percent identity is also referred to as cross-hybridization identity of the oligo.

This calculated cross-hybridization identity is dependent on the size of the sequence database used to BLAST against, oligo sequence, and use of either gapped or no-gap alignment method.

7) Each oligo of any length cannot have greater than 20 contiguous bases common to any other gene.

Once oligo candidates have been selected satisfying all the selection rules mentioned above, each oligo is ranked based on BLAST percent identity as computed in Step 6. One final oligo for each gene is selected with the minimum percent identity or crosshybridization similarity.

Note that for Version 1.0 a small number of genes (313 genes or 7.3%) that did not yield oligos satisfying all the above criteria, certain rules were relaxed. For those genes, the oligo is selected greater than 750 bases from the 3' end, greater than 70% crosshybridization identity, or a longer stem length in a potential hairpin.

SUMMARY

Oligo selection criteria	Criteria values	Number of oligos in genome set satisfying these criteria (Version 1.0)	Number of oligos in genome set satisfying these criteria (Upgrade)
Length Melting temperature Location from 3' end Poly(N)tract length Stem length in potential hairpin Cross-hybridization identity to all other genes Contiguous base match to any other gene	70mer 78°C ± 5°C ≤ 750 ≤ 10 ≤ 9 ≤ 70% ≤ 20	3960	1266
Length Melting temperature Location from 3' end Poly(N)tract length Stem length in potential hairpin Cross-hybridization identity to all other genes Contiguous base match to any other gene	70mer 78°C ± 5°C ≤ 750 ≤ 10 ≤ 9 > 70% ≤ 20	230	0
Length Melting temperature Location from 3' end Poly(N)tract length Stem length in potential hairpin Cross-hybridization identity to all other genes Contiguous base match to any other gene	70mer 78°C ± 5°C ≤ 750 ≤ 10 ≤ 9 ≤ 70% > 20	23	5
Length Melting temperature Location from 3' end Poly(N)tract length Stem length in potential hairpin Cross-hybridization to all other genes Contiguous base match to any other gene	70mer 78°C ± 5°C ≤ 750 ≤ 10 ≤ 9 > 70% > 20	21	0
Stem length in potential hairpin	9 < x ≤ 12	8	6

Location from 3' end	750 < x < 850	31	54
Location from 3' end	850 < x < 1500	0	112
Contiguous base match to any other gene	> 20	44	8
Cross-hybridization to all other genes	> 70%	255	0
Total		4273	1432

The following illustrations show the distribution of all 5705 oligos representing Version1.0 and the upgrade for melting temperature, GC content, location from 3' end of gene sequence, length of maximum stem length, and BLAST percent identity or crosshybridization similarity.

Figure 1. Melting Temperature

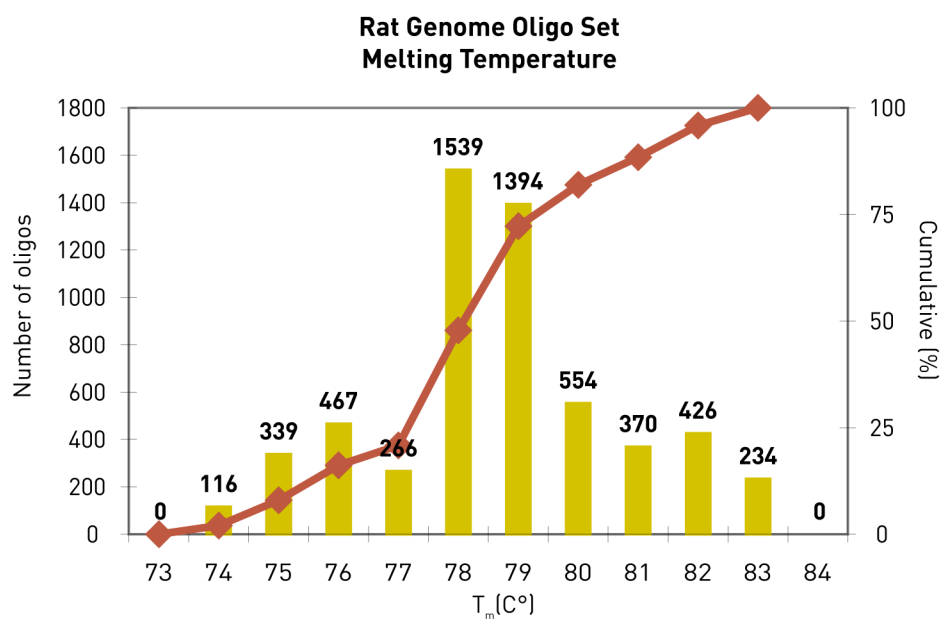


Figure 2. GC Content

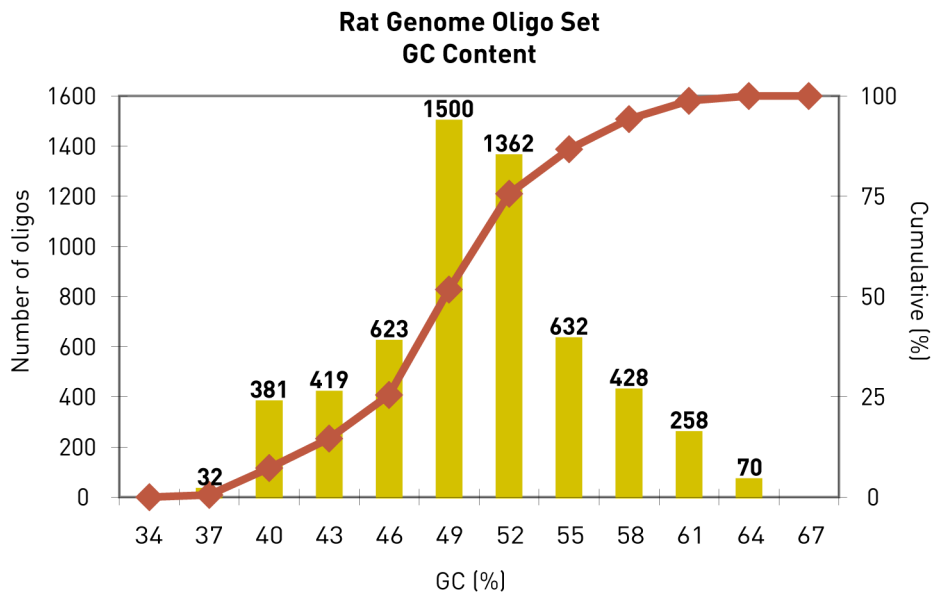


Figure 3. Location from 3' End

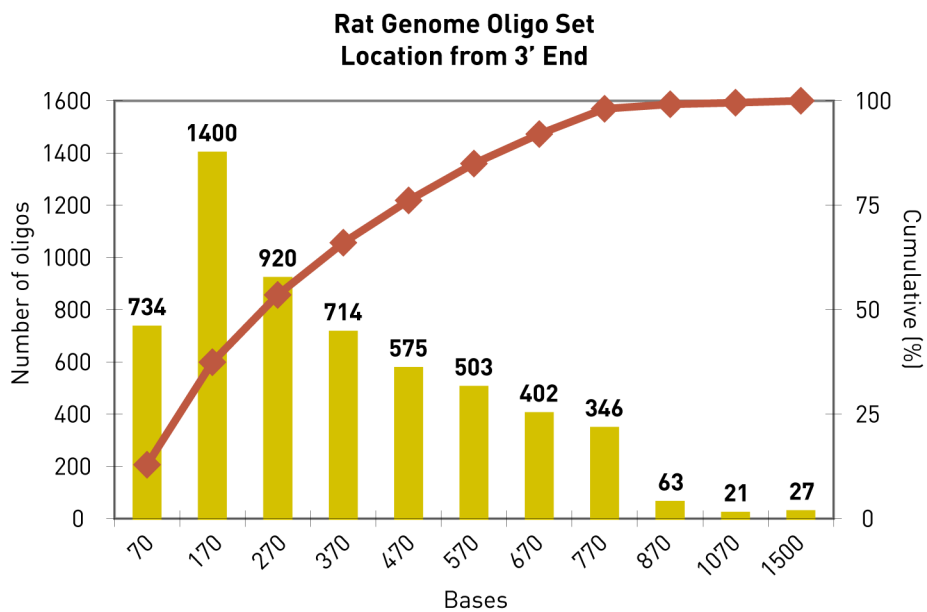


Figure 4. Length of the Longest Hairpin Stem

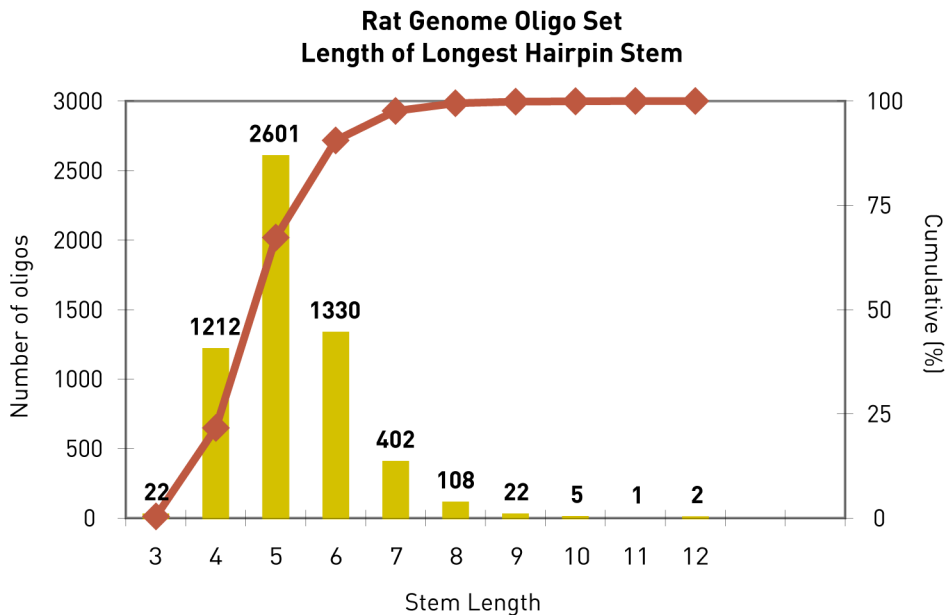
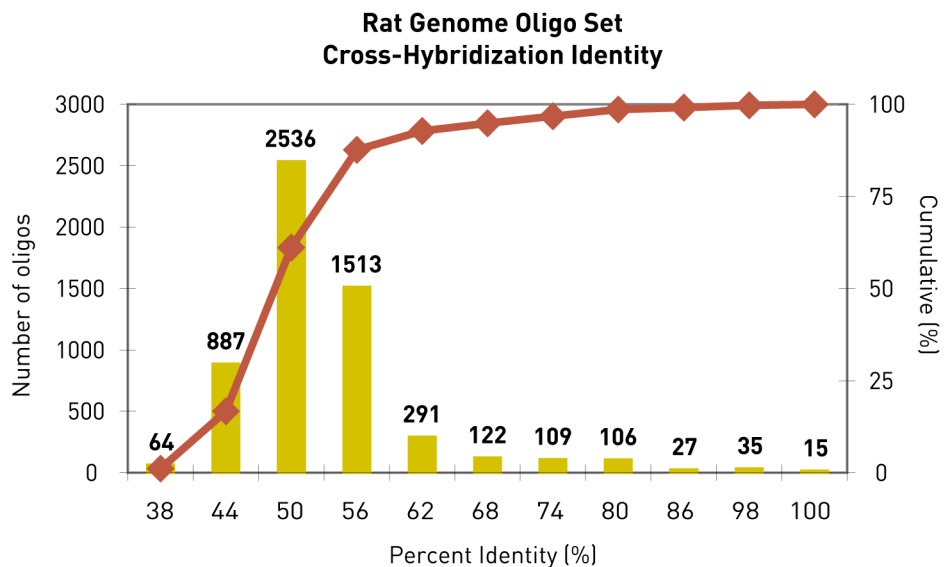


Figure 5. Cross-Hybridization Identity



### Quality check of probe design specifications

Once the final oligo set has been selected to represent a gene, each oligo undergoes design specifications quality control where we use an independent method to confirm that all oligos have met the specified design specifications. The table below summarizes data from our quality check for probe design specifications for all 4273 probes in the set Version 1.0 and 1432 oligos in the Upgrade set (Version 1.1).

Probe design specification	Expected value	Verified range	Number of oligos (Version 1.0)	Number of oligos (Upgrade)
Melting temperature [C°]	78°C ± 5°C	73.6–82.9	4273	1432
GC content (%)	35%–62%	38.6–61.4	4273	1432
Cross-hybridization similarity (%)	≤ 70	38–70	4018	1432
Cross-hybridization similarity (%)	> 70	71–100	255	0