

Homo sapiens siRNAs



RNAi experiments in mammalian systems require the application of either *in vitro* diced long dsRNAs (esiRNAs) or synthetic siRNAs. NEXT-RNAi was used to design a genome-wide siRNA library targeting all human genes annotated by the [NCBI RefSeq](#) database (release 40). To this end regions common to all RefSeq transcripts belonging to the same gene were computed for the complete human genome (37,627 regions). Further, sequences of low-complexity were filtered using mdust and remaining sequences longer 100 nt were splitted into two sequences to obtain a higher number of potential target sites for NEXT-RNAi reagent designs. Overall this resulted in 100,270 input sequences used as input for NEXT-RNAi.

NEXT-RNAi results

NEXT-RNAi HTML outputs are available [here](#)

Overall 100,264 designs were obtained, covering 99.9% of the genome. 83.4% of all genes are covered by at least one design that does not show homology of 19 nt to any other gene. 97% of all genes are additionally covered by at least one second, independent design.

Input files and settings used

Input FASTA file

human.rnaMOD.COMMON_mdust_split_crsplit.zip (20MB) containing target sequences as input file (-i input).

```
>LOC729774_cr1:1
CCAAGCCTGCAGCAGGGAGAGCAACAAGCCCTGGCCCTCAGAGCTCAGCCGGATGAGCGCAGCCCAGAG
ACAGCAGCTTCTCGAGGAAGGAAGGACCCGGTTTCAGGAGCTGCTGTCCAGTCCGGCCTACAGAGCCAG
CACCTTGGTGGCCATCGGGCAGACGCTGGCCCGGCAGATGCAGCTGGAAGATGGCGGCCAGCTCTGA
>LOC100128610_cr3:1
ATGAGGCTGAGTCTTATCCCTCGGAACACGGGCACCCACAGAGGGTCCCTGCCTCCTGTGGTCTGGAGC
TCCCCCTCAAGGAAGAAACCCCTTGCTGTCTGCTTGCAACTCCATGATGTTTGGACACCTCAGCCCCGTG
AGGATCCCTTATCTCAGAGGCAAGTTTAAC
>RNF185_cr2:1
AAGTCCCTCCGAGAGGGGCGGCTCCGCGTCATGTGACTGGAGTCCGCGTAGGAGGGGTCGGAGGTCTTA
CCCAACAGATTGACGCGCGTTAGTATTGGCCGTGTACCCGAAAAAAGTATTGACTGGGCTGGCGTTAA
CTGTGCGGAGG
```

Targetgroup file (tab-delimited)

TargetGroups_GeneID.tab (1.3MB) defining which RefSeq transcripts belong to the same gene (headers Target and TargetGroup) ([TARGETGR OUPS](#) option)

Target	TargetGroup	TargetGroup2
NM_001012993.2	C9orf152	401546
NM_001015.3	RPS11	6205
XM_002348062.1	LOC100291269	100291269
NM_182764.1	ELMO2	63916
NM_133171.3	ELMO2	63916
NM_052854.2	CREB3L1	90993
NM_006029.4	PNMA1	9240
NM_004530.4	MMP2	4313
NM_001127891.1	MMP2	4313

Bowtie database/index for off-target evaluation

Bowtie database/index containing annotated RefSeq transcripts (release 40) for specificity calculations (-d input):
[human.rnaMOD.tar.gz \(114MB\)](#)

Feature file with UTR and SNP locations

Tab-delimited feature file containing mappings of UTRs and SNPs (from NCBI [dbSNP](#)) to chromosomes that is used to calculate UTR and SNP 'contents' ([FEATURE](#) option) of designed reagents: [Hs_UTR_SNP.tar.gz \(484MB\)](#)

ID	FeatureName	FeatureLoc	FeatureStart	FeatureEnd
UTR_1	UTR	11	11643172	11643172
UTR_2	UTR	11	6741799	6741799
UTR_3	UTR	11	127219	127219
UTR_4	UTR	11	2662385	2662385
UTR_5	UTR	11	45952962	45952962
rs80303196	SNP	6	29985811	29985811
rs80303196	SNP	6	32995534	32995534
rs80303196	SNP	6	29080247	29080247
rs80303196	SNP	6	30038873	30038873
rs80303196	SNP	6	32181031	32181031

FASTA file for homology evaluation

Transcriptome FASTA file to evaluate the homology of the designs using Blast ([HOMOLOGY](#) and [TXNFASTA](#) options):
[human.rnaMOD.fna.zip \(32MB\)](#)

Bowtie database/index for seed match evaluation

To compute the number of siRNA seed matches (seed complement frequency) a Bowtie database/index containing all annotated 3'-UTR sequences (RefSeq release 40) was generated to be used with the [SEEDMATCH](#) option ([Hs_3UTR.tar.gz \(26MB\)](#)).

Design criteria

Start of program

```
perl nextrnai.pl -i human.rnaMOD.COMMON_mdust_split_crsplit.fa -s 7500 -r s
-d human.rnaMOD -e NO -o options.txt -n Hs_RefSeq40
```

Descriptions for start parameters used are available [here](#).

Options file

```
SIRNALENGTH=19
EFFICIENCY=SIR,63
BOWTIE=/usr/bin/
TARGETGROUPS=TargetGroups_GeneID.tab
SOURCE=CDS
BLAT=/usr/bin/
BLATPROGRAM=gfClient
BLATHOST=b110-cellarray3
BLATPORT=3500
TXNFASTA=human.rnaMOD.fna
GFF=GFF3
GBROWSEBASE=http://www.dkfz.de/signaling/cgi-bin/gbrowse_img/hsrefseq/
GBROWSETRACK=GENE+TXN+ENSEMBLGENESPAN+ENSEMBLGENE
AFF=YES
CANEVAL=6
HOMOLOGY=/usr/bin/,human.rnaMOD.fna,0.1
FEATURE=Hs_UTR_SNP.tab
SEEDMATCH=6,2500,Hs_3UTR
```

Descriptions for all options used are available [here](#).