

Quantifying Multi-Mapped Reads in Our Ribo-Seq Data V2

AMIR RANJBAR

2025-06-01

Deschamps-Francoeur et al. (2020) point out that rRNA genes, their pseudogenes, and other short noncoding RNAs have a lot of sequence similarity across the genome, which makes them especially likely to produce multi-mapped reads in RNA-seq. Many of these sequences are either nearly identical to each other or embedded within longer transcripts, making it hard for aligners to place reads uniquely. The paper explains that while an aligner tool (like STAR) often ignores multi-mapped reads to avoid uncertainty, this also means important RNA types—especially rRNA and lncRNA—end up undercounted (as you know Mark, this is not important for our Ribo-seq analysis). These RNAs are not only repetitive but also highly expressed, so filtering them out can skew results (again not important for us). This fits well with our data, where most of the multi-mapped reads came from rRNA and lncRNA.

Extracting (via `--outReadsUnmapped Fastx`) Multi-Mapped Reads from STAR Output

```
STAR --runThreadN 4 --genomeDir  
/global/home/sa104119/ormistonLab/riboDiffV8/reference --readFilesIn  
/global/home/sa104119/ormistonLab/riboDiffV9/3bowtie2V1/2RW9VV_trimmed_filtered.fastq --outFileNamePrefix  
/global/home/sa104119/ormistonLab/riboDiffV9/4starV2/2RW9VV_trimmed_filtered_--outSAMtype BAM SortedByCoordinate --alignEndsType EndToEnd --  
outFilterMultimapNmax 1 --outSAMattributes All --outReadsUnmapped Fastx
```

```
# Output:  
Number of input reads | 2279000  
Uniquely mapped reads number | 670531  
Number of reads mapped to too many loci | 1253246  
Number of reads unmapped: too short | 355220  
Number of reads unmapped: other | 3
```

Now we have a fastq `file` (`2RW9VV_trimmed_filtered_Unmapped.out.mate1`), containing all multi-mapped `reads` (plus too short and other).

Quality Control of Unmapped Reads

```
fastqc  
/global/home/sa104119/ormistonLab/riboDiffV9/4starV2/2RW9VV_trimmed_filtered_  
Unmapped.out.mate1 -o  
/global/home/sa104119/ormistonLab/riboDiffV9/4starV2FastqcV2  
  
# Output - As expected, the basic statistics for our fastq file is:  
Total Sequences 1608469  
Sequence length 17-34
```

Randomly Selecting 20 Reads Using seqkit

```
seqkit sample -n 20 -s 100 2RW9VV_trimmed_filtered_Unmapped.out.mate1 -o  
20_Random_Reads_2RW9VV_trimmed_filtered_Unmapped.fastq  
  
# Output:  
[INFO] sample by number  
[INFO] loading all sequences into memory...  
[INFO] 20 sequences outputted
```

Using STAR for Aligning Those 20 Reads

```
STAR --runThreadN 4 --genomeDir /reference --readFilesIn  
/20_Random_Reads_2RW9VV_trimmed_filtered_Unmapped.fastq --outFileNamePrefix  
/20_Random_Reads_2RW9VV_trimmed_filtered_Unmapped_ --outSAMtype BAM  
SortedByCoordinate --outFilterMultimapNmax 500 --winAnchorMultimapNmax 500 --  
outSAMmultNmax 1000  
  
# Output:  
- 0 uniquely mapped → STAR couldn't confidently place any reads in one clear  
location.  
- All 20 (100%) multi-mapped → Every read aligned to multiple locations (I  
allowed up to 500).  
- 0 reads unmapped → All reads aligned somewhere.  
- All 20 reads did map, just not uniquely.
```

This gave us around ~130 "reads" (alignments).

SAM Records Extraction from the BAM File

I then extracted some useful info from our STAR alignment output
(20_Random_Reads_2RW9VV_trimmed_filtered_Unmapped_Aligned.sortedByCoord.out.b
am):

```

samtools view
/20_Random_Reads_2RW9VV_trimmed_filtered_Unmapped_Aligned.sortedByCoord.out.b
am | awk '{nh="NA"; hi="NA"; for(i=12;i<=NF;i++) {if($i ~ /^NH:i:/) nh=$i;
if($i ~ /^HI:i:/) hi=$i;} print $1 "\t" $3 "\t" $4 "\t" $6 "\t" $10 "\t" nh
"\t" hi}' > readID_RNAME_POS_CIGAR_seq_NH_HI.txt

```

Our Excel Workbook with SAM Records

And add them to Excel workbook (readID_RNAME_POS_CIGAR_seq_NH_HI_split.xlsx):

```

from google.colab import drive
drive.mount('/content/drive')

import pandas as pd

# Load the updated file
file_path = '/content/drive/My Drive/BMIF
898/multiMappedRpfReads/readID_RNAME_POS_CIGAR_seq_NH_HI.txt'
df = pd.read_csv(file_path, sep='\t', header=None, names=['ReadID', 'RNAME',
'POS', 'CIGAR', 'Sequence', 'NH', 'HI'])

# Show the first few rows
df.head()

# Count unique ReadIDs
unique_read_ids = df['ReadID'].nunique()
print(f'Number of unique ReadIDs: {unique_read_ids}')

# Create an Excel file with a sheet per unique ReadID
output_excel_path = '/content/drive/My Drive/BMIF
898/multiMappedRpfReads/readID_RNAME_POS_CIGAR_seq_NH_HI_split.xlsx'
with pd.ExcelWriter(output_excel_path, engine='openpyxl') as writer:
    for read_id, group_df in df.groupby('ReadID'):
        sheet_name = read_id[:31]
        group_df.to_excel(writer, sheet_name=sheet_name, index=False)

```

This Excel workbook contains SAM records extracted from the BAM file (20_Random_Reads_2RW9VV_trimmed_filtered_Unmapped_Aligned.sortedByCoord.out.bam). Each sheet corresponds to one of 20 unique read IDs. For each read ID, all its alignment instances are included, with associated information such as the reference sequence name (RNAME), alignment position (POS), CIGAR string (CIGAR), read sequence (SEQ), and the optional alignment tags NH (number of reported alignments for the read) and HI (hit index indicating the rank of the alignment).

Please refer to readID_RNAME_POS_CIGAR_seq_NH_HI_split.xlsx.

My next and last step was to use IGV (used UCSC instead), manually input the relevant positional info of each of ~130 alignments (RNAME, POS, CIGAR, ...), and check whether the mappings cluster in certain regions or not.

Please refer to UCSC.pdf.

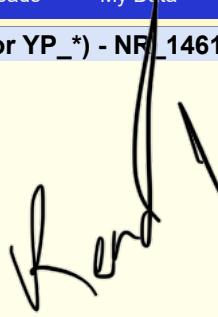
Reference

Deschamps-Francoeur, G., Simoneau, J., & Scott, M. S. (2020). Handling multi-mapped reads in RNA-seq. *Computational and structural biotechnology journal*, 18, 1569–1576.
<https://doi.org/10.1016/j.csbj.2020.06.014>



hg38 NCBI RefSeq genes, curated subset (NM_*, NR_*, NP_* or YP_*) - NR_146148.1

RefSeq Gene RNA28SN2



RefSeq: [NR_146148.1](#) **Status:** Validated
Description: RNA, 28S ribosomal RNA N2
Molecule type: rRNA
Source: BestRefSeq
Biotype: rRNA
HGNC: [53520](#)
Entrez Gene: [109864282](#)
GeneCards: [RNA28SN2](#)
AceView: [RNA28SN2](#)

Summary of RNA28SN2

45S ribosomal DNA (rDNA) arrays, or clusters, are present on human chromosomes 13, 14, 15, 21 and 22, designated RNR1 through RNR5, respectively. Each cluster consists of multiple 45S rDNA repeat units that vary in number among individuals and chromosomes, with total diploid copy number estimates ranging from 60 to >800 repeat units in a human genome. The 45S rDNA repeat unit encodes a 45S rRNA precursor, transcribed by RNA polymerase I, which is processed to form the 18S, 5.8S and 28S rRNAs. This gene represents a copy of the 28S ribosomal RNA on chromosome 21. [provided by RefSeq, Mar 2017]. Sequence Note: The RefSeq transcript was derived from the reference genome assembly. The genomic coordinates were determined from alignments. Publication Note: This RefSeq record includes a subset of the publications that are available for this gene. Please see the Gene record to access additional publications.

mRNA/Genomic Alignments (NR_146148.1)

BROWSER	SIZE	IDENTITY	CHROMOSOME	STRAND	START	END	QUERY	START	END	TOTAL
browser	5053	100.0%	21	+	8213888	8218941	NR_146148.1	1	5054	5054

[View details of parts of alignment within browser window.](#)

Position: [chr21:8213888-8218941](#)

Band: 21p11.2

Genomic Size: 5054

Strand: +

Gene Symbol: RNA28SN2

Links to sequence:

- [Predicted mRNA](#) may be different from the genomic sequence.
- [Genomic Sequence](#) from assembly

[Data schema/format description and download](#)

[Go to NCBI RefSeq track controls](#)

Source data version: NCBI RefSeq GCF_000001405.40-RS_2024_08 (2024-08-27)

Data last updated at UCSC: 2024-09-11

Description

The NCBI RefSeq Genes composite track shows human protein-coding and non-protein-coding genes taken from the NCBI RNA reference sequences collection (RefSeq). All subtracks use coordinates provided by RefSeq, except for the UCSC RefSeq track, which UCSC produces by realigning the RefSeq RNAs to the genome. This realignment may result in occasional differences between the annotation coordinates provided by UCSC and NCBI. For RNA-seq analysis, we advise using NCBI aligned tables like RefSeq All or RefSeq Curated. See the [Methods](#) section for more details about how the different tracks were created.

Please visit NCBI's [Feedback for Gene and Reference Sequences \(RefSeq\)](#) page to make suggestions, submit additions and corrections, or ask for help concerning RefSeq records.

For more information on the different gene tracks, see our [Genes FAQ](#).



hg38 Repeat

RepeatMasker Information

Name: [LSU-rRNA_Hsa](#) (link requires [registration](#))

Family: rRNA

Class: rRNA

SW Score: 40632

Divergence: 0.2%

Deletions: 1.2%

Insertions: 0.8%

Begin in repeat: 878

End in repeat: 5034

Left in repeat: 1

Position: [chr21:8214773-8218941](#)

Band: 21p11.2

Genomic Size: 4169

Strand: +

[View DNA for this feature](#) (hg38/Human)

[Data schema/format description and download](#)

[Go to RepeatMasker track controls](#)

Data last updated at UCSC: 2022-10-18



Description

This track was created by using Arian Smit's [RepeatMasker](#) program, which screens DNA sequences for interspersed repeats and low complexity DNA sequences. The program outputs a detailed annotation of the repeats that are present in the query sequence (represented by this track), as well as a modified version of the query sequence in which all the annotated repeats have been masked (generally available on the [Downloads](#) page). RepeatMasker uses the [Repbase Update](#) library of repeats from the [Genetic Information Research Institute](#) (GIRI). Repbase Update is described in Jurka (2000) in the References section below.

This track and the masking information in our [hg38 genome download FASTA files](#) was created in 2010 with the original RepBase library from 2010-03-02 and RepeatMasker 3.0.1. Since April 2019, RepBase is under a commercial license, we cannot distribute it or update the track using the RepBase library without a license. Therefore, and for compatibility with past results, given how central the masking is for many other annotations, we decided to not update the repeatmasking of hg38. However, you can show the small differences between the RepeatMasker 3/RepBase from 2010 and RepeatMasker 4/DFAM from 2020 using the track "RepeatMasker Viz" in the same track group. It contains two subtracks, one with the old and one with the new data. Also, these tracks have many more visualisation options than the original RepeatMasker track.

However, the last track update time of this track at UCSC is not 2010, because we had to add repeatmasking annotations to the rarely used _alt and _fix "patch" sequences of the hg38 genome. The repeatmasking annotations of the main chromosomes were unaffected and have not changed since 2010. For more information on genome patches, see our [blog post](#).

Display Conventions and Configuration

In full display mode, this track displays up to ten different classes of repeats:

- Short interspersed nuclear elements (SINE), which include ALUs
- Long interspersed nuclear elements (LINE)
- Long terminal repeat elements (LTR), which include retrotransposons
- DNA repeat elements (DNA)
- Simple repeats (micro-satellites)
- Low complexity repeats
- Satellite repeats
- RNA repeats (including RNA, tRNA, rRNA, snRNA, scRNA, srpRNA)
- Other repeats, which includes class RC (Rolling Circle)
- Unknown



hg38 Repeat

RepeatMasker Information

Name: [LSU-rRNA_Hsa](#) (link requires [registration](#))

Family: rRNA

Class: rRNA

SW Score: 21300

Divergence: 0.4%

Deletions: 1.7%

Insertions: 0.6%

Begin in repeat: 878

End in repeat: 2852

Left in repeat: 2183

Position: [chr21:8259002-8260971](#)

Band: 21p11.2

Genomic Size: 1970

Strand: +

[View DNA for this feature](#) (hg38/Human)

[Data schema/format description and download](#)

[Go to RepeatMasker track controls](#)

Data last updated at UCSC: 2022-10-18



Description

This track was created by using Arian Smit's [RepeatMasker](#) program, which screens DNA sequences for interspersed repeats and low complexity DNA sequences. The program outputs a detailed annotation of the repeats that are present in the query sequence (represented by this track), as well as a modified version of the query sequence in which all the annotated repeats have been masked (generally available on the [Downloads](#) page). RepeatMasker uses the [Repbase Update](#) library of repeats from the [Genetic Information Research Institute](#) (GIRI). Repbase Update is described in Jurka (2000) in the References section below.

This track and the masking information in our [hg38 genome download FASTA files](#) was created in 2010 with the original RepBase library from 2010-03-02 and RepeatMasker 3.0.1. Since April 2019, RepBase is under a commercial license, we cannot distribute it or update the track using the RepBase library without a license. Therefore, and for compatibility with past results, given how central the masking is for many other annotations, we decided to not update the repeatmasking of hg38. However, you can show the small differences between the RepeatMasker 3/RepBase from 2010 and RepeatMasker 4/DFAM from 2020 using the track "RepeatMasker Viz" in the same track group. It contains two subtracks, one with the old and one with the new data. Also, these tracks have many more visualisation options than the original RepeatMasker track.

However, the last track update time of this track at UCSC is not 2010, because we had to add repeatmasking annotations to the rarely used _alt and _fix "patch" sequences of the hg38 genome. The repeatmasking annotations of the main chromosomes were unaffected and have not changed since 2010. For more information on genome patches, see our [blog post](#).

Display Conventions and Configuration

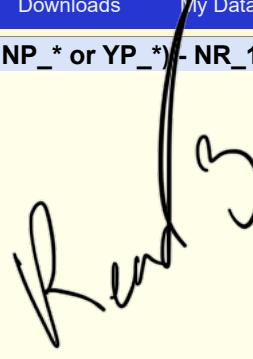
In full display mode, this track displays up to ten different classes of repeats:

- Short interspersed nuclear elements (SINE), which include ALUs
- Long interspersed nuclear elements (LINE)
- Long terminal repeat elements (LTR), which include retrotransposons
- DNA repeat elements (DNA)
- Simple repeats (micro-satellites)
- Low complexity repeats
- Satellite repeats
- RNA repeats (including RNA, tRNA, rRNA, snRNA, scRNA, srpRNA)
- Other repeats, which includes class RC (Rolling Circle)
- Unknown



hg38 NCBI RefSeq genes, curated subset (NM_*, NR_*, NP_* or YP_*) - NR_146151.1

RefSeq Gene RNA45SN3

RefSeq: [NR_146151.1](#) Status: Reviewed

Description: RNA, 45S pre-ribosomal N3

Molecule type: rRNA

Source: BestRefSeq

Biotype: rRNA

HGNC: [53522](#)Entrez Gene: [109910379](#)GeneCards: [RNA45SN3](#)AceView: [RNA45SN3](#)

Summary of RNA45SN3

45S ribosomal DNA (rDNA) arrays, or clusters, are present on human chromosomes 13, 14, 15, 21 and 22, designated RNR1 through RNR5, respectively. Each cluster consists of multiple 45S rDNA repeat units that vary in number among individuals and chromosomes, with total diploid copy number estimates ranging from 60 to >800 repeat units in a human genome. The 45S rDNA repeat unit encodes a 45S rRNA precursor, transcribed by RNA polymerase I, which is processed to form the 18S, 5.8S and 28S rRNAs. This gene represents a copy of the 45S ribosomal RNA on chromosome 21. [provided by RefSeq, Mar 2017]. Sequence Note: The RefSeq transcript was derived from the reference genome assembly. The genomic coordinates were determined from alignments. Publication Note: This RefSeq record includes a subset of the publications that are available for this gene. Please see the Gene record to access additional publications.

mRNA/Genomic Alignments (NR_146151.1)

The alignment you clicked on is first in the table below.

BROWSER	SIZE	IDENTITY	CHROMOSOME	STRAND	START	END	QUERY	START	END	TOTAL
browser	13309	100.0%	21	+	8389035	8402343	NR_146151.1	1	13309	13309
browser	13309	100.0%	21_ML143377v1_fix	+	436160	449468	NR_146151.1	1	13309	13309

[View details of parts of alignment within browser window.](#)

Position: [chr21:8389035-8402343](#)

Band: 21p11.2

Genomic Size: 13309

Strand: +

Gene Symbol: RNA45SN3

Position: [chr21_ML143377v1_fix:436160-449468](#)

Band: 21_ML143377v1_fix

Genomic Size: 13309

Strand: +

Gene Symbol: RNA45SN3

Links to sequence:

- [Predicted mRNA](#) may be different from the genomic sequence.
- [Genomic Sequence](#) from assembly

[Data schema/format description and download](#)

[Go to NCBI RefSeq track controls](#)

Source data version: NCBI RefSeq GCF_000001405.40-RS_2024_08 (2024-08-27)

Data last updated at UCSC: 2024-09-11

Description

The NCBI RefSeq Genes composite track shows human protein-coding and non-protein-coding genes taken from the NCBI RNA reference sequences collection (RefSeq). All subtracks use coordinates provided by RefSeq, except for the UCSC RefSeq track, which UCSC produces by realigning the RefSeq RNAs to the genome. This realignment may result in occasional differences between the annotation coordinates provided by UCSC and



hg38 Repeat

RepeatMasker Information

Name: [LSU-rRNA_Hsa](#) (link requires [registration](#))

Family: rRNA

Class: rRNA

SW Score: 37669

Divergence: 0.4%

Deletions: 1.2%

Insertions: 0.8%

Begin in repeat: 878

End in repeat: 5034

Left in repeat: 1

Position: [chr21:8397807-8401980](#)

Band: 21p11.2

Genomic Size: 4174

Strand: +

[View DNA for this feature](#) (hg38/Human)

[Data schema/format description and download](#)

[Go to RepeatMasker track controls](#)

Data last updated at UCSC: 2022-10-18



Description

This track was created by using Arian Smit's [RepeatMasker](#) program, which screens DNA sequences for interspersed repeats and low complexity DNA sequences. The program outputs a detailed annotation of the repeats that are present in the query sequence (represented by this track), as well as a modified version of the query sequence in which all the annotated repeats have been masked (generally available on the [Downloads](#) page). RepeatMasker uses the [Repbase Update](#) library of repeats from the [Genetic Information Research Institute](#) (GIRI). Repbase Update is described in Jurka (2000) in the References section below.

This track and the masking information in our [hg38 genome download FASTA files](#) was created in 2010 with the original RepBase library from 2010-03-02 and RepeatMasker 3.0.1. Since April 2019, RepBase is under a commercial license, we cannot distribute it or update the track using the RepBase library without a license. Therefore, and for compatibility with past results, given how central the masking is for many other annotations, we decided to not update the repeatmasking of hg38. However, you can show the small differences between the RepeatMasker 3/RepBase from 2010 and RepeatMasker 4/DFAM from 2020 using the track "RepeatMasker Viz" in the same track group. It contains two subtracks, one with the old and one with the new data. Also, these tracks have many more visualisation options than the original RepeatMasker track.

However, the last track update time of this track at UCSC is not 2010, because we had to add repeatmasking annotations to the rarely used _alt and _fix "patch" sequences of the hg38 genome. The repeatmasking annotations of the main chromosomes were unaffected and have not changed since 2010. For more information on genome patches, see our [blog post](#).

Display Conventions and Configuration

In full display mode, this track displays up to ten different classes of repeats:

- Short interspersed nuclear elements (SINE), which include ALUs
- Long interspersed nuclear elements (LINE)
- Long terminal repeat elements (LTR), which include retrotransposons
- DNA repeat elements (DNA)
- Simple repeats (micro-satellites)
- Low complexity repeats
- Satellite repeats
- RNA repeats (including RNA, tRNA, rRNA, snRNA, scRNA, srpRNA)
- Other repeats, which includes class RC (Rolling Circle)
- Unknown



hg38 NCBI RefSeq genes, curated subset (NM_*, NR_*, NP_* or YP_*) - NR_145819.1

RefSeq Gene RNA45SN1

*RefSeq
fix*

RefSeq: [NR_145819.1](#) Status: Reviewed

Description: RNA, 45S pre-ribosomal N1

Molecule type: rRNA

Source: BestRefSeq

Biotype: rRNA

Synonyms: RNA45S4

HGNC: [53514](#)Entrez Gene: [106631777](#)GeneCards: [RNA45SN1](#)AceView: [RNA45SN1](#)

Summary of RNA45SN1

45S ribosomal DNA (rDNA) arrays, or clusters, are present on human chromosomes 13, 14, 15, 21 and 22, designated RNR1 through RNR5, respectively. Each cluster consists of multiple 45S rDNA repeat units that vary in number among individuals and chromosomes, with total diploid copy number estimates ranging from 60 to >800 repeat units in a human genome. The 45S rDNA repeat unit encodes a 45S rRNA precursor, transcribed by RNA polymerase I, which is processed to form the 18S, 5.8S and 28S rRNAs. Gene and RefSeq, in collaboration with HGNC, currently describe one 45S rDNA cluster, and one set of 45S precursor and product rRNAs, for each of the five human chromosomes to which these loci are localized. This gene is a representative copy of the 45S pre-ribosomal RNA on chromosome 21. [provided by RefSeq, Feb 2017]. Sequence Note: The RefSeq transcript was derived from the reference genome assembly. The genomic coordinates were determined from alignments. Publication Note: This RefSeq record includes a subset of the publications that are available for this gene. Please see the Gene record to access additional publications.

mRNA/Genomic Alignments (NR_145819.1)

The alignment you clicked on is first in the table below.

BROWSER	SIZE	IDENTITY	CHROMOSOME	STRAND	START	END	QUERY	START	END	TOTAL	
browser	13351	100.0%	21	+	8433222	8446572	NR_145819.1	1	13351	13351	
browser	13351	100.0%	21	ML143377v1_fix	+	480347	493697	NR_145819.1	1	13351	13351

[View details of parts of alignment within browser window.](#)

Position: [chr21:8433222-8446572](#)

Band: 21p11.2

Genomic Size: 13351

Strand: +

Gene Symbol: RNA45SN1

Position: [chr21_ML143377v1_fix:480347-493697](#)

Band: 21_ML143377v1_fix

Genomic Size: 13351

Strand: +

Gene Symbol: RNA45SN1

Links to sequence:

- [Predicted mRNA](#) may be different from the genomic sequence.
- [Genomic Sequence](#) from assembly

[Data schema/format description and download](#)

[Go to NCBI RefSeq track controls](#)

Source data version: NCBI RefSeq GCF_00001405.40-RS_2024_08 (2024-08-27)

Data last updated at UCSC: 2024-09-11

Description

The NCBI RefSeq Genes composite track shows human protein-coding and non-protein-coding genes taken from the NCBI RNA reference sequences collection (RefSeq). All subtracks use coordinates provided by RefSeq.



hg38 Repeat

RepeatMasker Information

Name: [LSU-rRNA_Hsa](#) (link requires [registration](#))

Family: rRNA

Class: rRNA

SW Score: 27358

Divergence: 0.2%

Deletions: 1.1%

Insertions: 0.8%

Begin in repeat: 878

End in repeat: 5035

Left in repeat: 0

Position: [chr21:8442037-8446211](#)

Band: 21p11.2

Genomic Size: 4175

Strand: +

[View DNA for this feature](#) (hg38/Human)

[Data schema/format description and download](#)

[Go to RepeatMasker track controls](#)

Data last updated at UCSC: 2022-10-18



Description

This track was created by using Arian Smit's [RepeatMasker](#) program, which screens DNA sequences for interspersed repeats and low complexity DNA sequences. The program outputs a detailed annotation of the repeats that are present in the query sequence (represented by this track), as well as a modified version of the query sequence in which all the annotated repeats have been masked (generally available on the [Downloads](#) page). RepeatMasker uses the [Repbase Update](#) library of repeats from the [Genetic Information Research Institute](#) (GIRI). Repbase Update is described in Jurka (2000) in the References section below.

This track and the masking information in our [hg38 genome download FASTA files](#) was created in 2010 with the original RepBase library from 2010-03-02 and RepeatMasker 3.0.1. Since April 2019, RepBase is under a commercial license, we cannot distribute it or update the track using the RepBase library without a license. Therefore, and for compatibility with past results, given how central the masking is for many other annotations, we decided to not update the repeatmasking of hg38. However, you can show the small differences between the RepeatMasker 3/RepBase from 2010 and RepeatMasker 4/DFAM from 2020 using the track "RepeatMasker Viz" in the same track group. It contains two subtracks, one with the old and one with the new data. Also, these tracks have many more visualisation options than the original RepeatMasker track.

However, the last track update time of this track at UCSC is not 2010, because we had to add repeatmasking annotations to the rarely used _alt and _fix "patch" sequences of the hg38 genome. The repeatmasking annotations of the main chromosomes were unaffected and have not changed since 2010. For more information on genome patches, see our [blog post](#).

Display Conventions and Configuration

In full display mode, this track displays up to ten different classes of repeats:

- Short interspersed nuclear elements (SINE), which include ALUs
- Long interspersed nuclear elements (LINE)
- Long terminal repeat elements (LTR), which include retrotransposons
- DNA repeat elements (DNA)
- Simple repeats (micro-satellites)
- Low complexity repeats
- Satellite repeats
- RNA repeats (including RNA, tRNA, rRNA, snRNA, scRNA, srpRNA)
- Other repeats, which includes class RC (Rolling Circle)
- Unknown



hg38 NCBI RefSeq genes, curated subset (NM_*, NR_*, NP_* or YP_*) - NR_146117.1

RefSeq Gene RNA45SN4



RefSeq: [NR_146117.1](#) **Status:** Reviewed

Description: RNA, 45S pre-ribosomal N4

Molecule type: rRNA

Source: BestRefSeq

Biotype: rRNA

Synonyms: RNA45S5

HGNC: [53526](#)

Entrez Gene: [109864271](#)

GeneCards: [RNA45SN4](#)

AceView: [RNA45SN4](#)

Summary of RNA45SN4

45S ribosomal DNA (rDNA) arrays, or clusters, are present on human chromosomes 13, 14, 15, 21 and 22, designated RNR1 through RNR5, respectively. Each cluster consists of multiple 45S rDNA repeat units that vary in number among individuals and chromosomes, with total diploid copy number estimates ranging from 60 to >800 repeat units in a human genome. The 45S rDNA repeat unit encodes a 45S rRNA precursor, transcribed by RNA polymerase I, which is processed to form the 18S, 5.8S and 28S rRNAs. This gene is a representative copy of the 45S ribosomal RNA on chromosome 22. [provided by RefSeq, Mar 2017]. Sequence Note: The RefSeq transcript was derived from the reference genome assembly. The genomic coordinates were determined from alignments. Publication Note: This RefSeq record includes a subset of the publications that are available for this gene. Please see the Gene record to access additional publications.

mRNA/Genomic Alignments (NR_146117.1)

BROWSER	SIZE	IDENTITY	CHROMOSOME	STRAND	START	END	QUERY	START	END	TOTAL
browser	13373	100.0%	22_KI270733v1_random	+	122273	135645	NR_146117.1	1	13373	13373

[View details of parts of alignment within browser window.](#)

Position: [chr22_KI270733v1_random:122273-135645](#)

Band: 22_KI270733v1_random

Genomic Size: 13373

Strand: +

Gene Symbol: RNA45SN4

Links to sequence:

- [Predicted mRNA](#) may be different from the genomic sequence.
- [Genomic Sequence](#) from assembly

[Data schema/format description and download](#)

[Go to NCBI RefSeq track controls](#)

Source data version: NCBI RefSeq GCF_000001405.40-RS_2024_08 (2024-08-27)

Data last updated at UCSC: 2024-09-11

Description

The NCBI RefSeq Genes composite track shows human protein-coding and non-protein-coding genes taken from the NCBI RNA reference sequences collection (RefSeq). All subtracks use coordinates provided by RefSeq, except for the UCSC RefSeq track, which UCSC produces by realigning the RefSeq RNAs to the genome. This realignment may result in occasional differences between the annotation coordinates provided by UCSC and NCBI. For RNA-seq analysis, we advise using NCBI aligned tables like RefSeq All or RefSeq Curated. See the [Methods](#) section for more details about how the different tracks were created.

Please visit NCBI's [Feedback for Gene and Reference Sequences \(RefSeq\)](#) page to make suggestions, submit additions and corrections, or ask for help concerning RefSeq records.

For more information on the different gene tracks, see our [Genes FAQ](#).



hg38 Repeat

RepeatMasker Information

Name: [LSU-rRNA_Hsa](#) (link requires [registration](#))

Family: rRNA

Class: rRNA

SW Score: 40361

Divergence: 0.2%

Deletions: 1.6%

Insertions: 0.9%

Begin in repeat: 878

End in repeat: 3492

Left in repeat: 1543

Position: [chr22_KI270733v1_random:131095-133689](#)

Band: 22_KI270733v1_random

Genomic Size: 2595

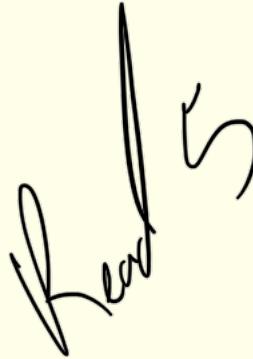
Strand: +

[View DNA for this feature](#) (hg38/Human)

[Data schema/format description and download](#)

[Go to RepeatMasker track controls](#)

Data last updated at UCSC: 2022-10-18



Description

This track was created by using Arian Smit's [RepeatMasker](#) program, which screens DNA sequences for interspersed repeats and low complexity DNA sequences. The program outputs a detailed annotation of the repeats that are present in the query sequence (represented by this track), as well as a modified version of the query sequence in which all the annotated repeats have been masked (generally available on the [Downloads](#) page). RepeatMasker uses the [Repbase Update](#) library of repeats from the [Genetic Information Research Institute](#) (GIRI). Repbase Update is described in Jurka (2000) in the References section below.

This track and the masking information in our [hg38 genome download FASTA files](#) was created in 2010 with the original RepBase library from 2010-03-02 and RepeatMasker 3.0.1. Since April 2019, RepBase is under a commercial license, we cannot distribute it or update the track using the RepBase library without a license. Therefore, and for compatibility with past results, given how central the masking is for many other annotations, we decided to not update the repeatmasking of hg38. However, you can show the small differences between the RepeatMasker 3/RepBase from 2010 and RepeatMasker 4/DFAM from 2020 using the track "RepeatMasker Viz" in the same track group. It contains two subtracks, one with the old and one with the new data. Also, these tracks have many more visualisation options than the original RepeatMasker track.

However, the last track update time of this track at UCSC is not 2010, because we had to add repeatmasking annotations to the rarely used _alt and _fix "patch" sequences of the hg38 genome. The repeatmasking annotations of the main chromosomes were unaffected and have not changed since 2010. For more information on genome patches, see our [blog post](#).

Display Conventions and Configuration

In full display mode, this track displays up to ten different classes of repeats:

- Short interspersed nuclear elements (SINE), which include ALUs
- Long interspersed nuclear elements (LINE)
- Long terminal repeat elements (LTR), which include retrotransposons
- DNA repeat elements (DNA)
- Simple repeats (micro-satellites)
- Low complexity repeats
- Satellite repeats
- RNA repeats (including RNA, tRNA, rRNA, snRNA, scRNA, srpRNA)
- Other repeats, which includes class RC (Rolling Circle)
- Unknown



hg38 Repeat

RepeatMasker Information

Name: [LSU-rRNA_Hsa](#) (link requires [registration](#))

Family: rRNA

Class: rRNA

SW Score: 32955

Divergence: 0.2%

Deletions: 1.8%

Insertions: 0.9%

Begin in repeat: 878

End in repeat: 3492

Left in repeat: 1543

Position: [chr22_KI270733v1_random:176174-178768](#)

Band: 22_KI270733v1_random

Genomic Size: 2595

Strand: +

[View DNA for this feature](#) (hg38/Human)

[Data schema/format description and download](#)

[Go to RepeatMasker track controls](#)

Data last updated at UCSC: 2022-10-18



Description

This track was created by using Arian Smit's [RepeatMasker](#) program, which screens DNA sequences for interspersed repeats and low complexity DNA sequences. The program outputs a detailed annotation of the repeats that are present in the query sequence (represented by this track), as well as a modified version of the query sequence in which all the annotated repeats have been masked (generally available on the [Downloads](#) page). RepeatMasker uses the [Repbase Update](#) library of repeats from the [Genetic Information Research Institute](#) (GIRI). Repbase Update is described in Jurka (2000) in the References section below.

This track and the masking information in our [hg38 genome download FASTA files](#) was created in 2010 with the original RepBase library from 2010-03-02 and RepeatMasker 3.0.1. Since April 2019, RepBase is under a commercial license, we cannot distribute it or update the track using the RepBase library without a license. Therefore, and for compatibility with past results, given how central the masking is for many other annotations, we decided to not update the repeatmasking of hg38. However, you can show the small differences between the RepeatMasker 3/RepBase from 2010 and RepeatMasker 4/DFAM from 2020 using the track "RepeatMasker Viz" in the same track group. It contains two subtracks, one with the old and one with the new data. Also, these tracks have many more visualisation options than the original RepeatMasker track.

However, the last track update time of this track at UCSC is not 2010, because we had to add repeatmasking annotations to the rarely used _alt and _fix "patch" sequences of the hg38 genome. The repeatmasking annotations of the main chromosomes were unaffected and have not changed since 2010. For more information on genome patches, see our [blog post](#).

Display Conventions and Configuration

In full display mode, this track displays up to ten different classes of repeats:

- Short interspersed nuclear elements (SINE), which include ALUs
- Long interspersed nuclear elements (LINE)
- Long terminal repeat elements (LTR), which include retrotransposons
- DNA repeat elements (DNA)
- Simple repeats (micro-satellites)
- Low complexity repeats
- Satellite repeats
- RNA repeats (including RNA, tRNA, rRNA, snRNA, scRNA, srpRNA)
- Other repeats, which includes class RC (Rolling Circle)
- Unknown



hg38 NCBI RefSeq genes, curated subset (NM_*, NR_*, NP_* or YP_*) - NR_046235.3

RefSeq Gene RNA45SN5

RefSeq: [NR_046235.3](#) **Status:** Reviewed

Description: RNA, 45S pre-ribosomal N5

Molecule type: rRNA

Source: BestRefSeq

Biotype: rRNA

Synonyms: RNA45S5

HGNC: [53530](#)

Entrez Gene: [100861532](#)

GeneCards: [RNA45SN5](#)

AceView: [RNA45SN5](#)

Summary of RNA45SN5

45S ribosomal DNA (rDNA) arrays, or clusters, are present on human chromosomes 13, 14, 15, 21 and 22, designated RNR1 through RNR5, respectively. Each cluster consists of multiple 45S rDNA repeat units that vary in number among individuals and chromosomes, with total diploid copy number estimates ranging from 60 to >800 repeat units in a human genome. The 45S rDNA repeat unit encodes a 45S rRNA precursor, transcribed by RNA polymerase I, which is processed to form the 18S, 5.8S and 28S rRNAs. This gene is a representative copy of the 45S pre-ribosomal RNA whose chromosomal location is unknown. [provided by RefSeq, Mar 2017].

Sequence Note: The RefSeq transcript was derived from the reference genome assembly. The genomic coordinates were determined from alignments. **Publication Note:** This RefSeq record includes a subset of the publications that are available for this gene. Please see the Gene record to access additional publications.

mRNA/Genomic Alignments (NR_046235.3)

BROWSER	SIZE	IDENTITY	CHROMOSOME	STRAND	START	END	QUERY	START	END	TOTAL
browser	13357	100.0%	Un_GL000220v1	+	105424	118780	NR_046235.3	1	13357	13357

[View details of parts of alignment within browser window.](#)

Position: [chrUn_GL000220v1:105424-118780](#)

Band: Un_GL000220v1

Genomic Size: 13357

Strand: +

Gene Symbol: RNA45SN5

Links to sequence:

- [Predicted mRNA](#) may be different from the genomic sequence.
- [Genomic Sequence](#) from assembly

[Data schema/format description and download](#)

[Go to NCBI RefSeq track controls](#)

Source data version: NCBI RefSeq GCF_00001405.40-RS_2024_08 (2024-08-27)

Data last updated at UCSC: 2024-09-11

Description

The NCBI RefSeq Genes composite track shows human protein-coding and non-protein-coding genes taken from the NCBI RNA reference sequences collection (RefSeq). All subtracks use coordinates provided by RefSeq, except for the UCSC RefSeq track, which UCSC produces by realigning the RefSeq RNAs to the genome. This realignment may result in occasional differences between the annotation coordinates provided by UCSC and NCBI. For RNA-seq analysis, we advise using NCBI aligned tables like RefSeq All or RefSeq Curated. See the [Methods](#) section for more details about how the different tracks were created.

Please visit NCBI's [Feedback for Gene and Reference Sequences \(RefSeq\)](#) page to make suggestions, submit additions and corrections, or ask for help concerning RefSeq records.

For more information on the different gene tracks, see our [Genes FAQ](#).



hg38 Repeat

RepeatMasker Information

Name: [LSU-rRNA_Hsa](#) (link requires [registration](#))

Family: rRNA

Class: rRNA

SW Score: 38060

Divergence: 0.2%

Deletions: 1.0%

Insertions: 0.8%

Begin in repeat: 878

End in repeat: 5034

Left in repeat: 1

Position: [chrUn_GL000220v1:114239-118417](#)

Band: Un_GL000220v1

Genomic Size: 4179

Strand: +

[View DNA for this feature](#) (hg38/Human)

[Data schema/format description and download](#)

[Go to RepeatMasker track controls](#)

Data last updated at UCSC: 2022-10-18



Description

This track was created by using Arian Smit's [RepeatMasker](#) program, which screens DNA sequences for interspersed repeats and low complexity DNA sequences. The program outputs a detailed annotation of the repeats that are present in the query sequence (represented by this track), as well as a modified version of the query sequence in which all the annotated repeats have been masked (generally available on the [Downloads](#) page). RepeatMasker uses the [Repbase Update](#) library of repeats from the [Genetic Information Research Institute](#) (GIRI). Repbase Update is described in Jurka (2000) in the References section below.

This track and the masking information in our [hg38 genome download FASTA files](#) was created in 2010 with the original RepBase library from 2010-03-02 and RepeatMasker 3.0.1. Since April 2019, RepBase is under a commercial license, we cannot distribute it or update the track using the RepBase library without a license. Therefore, and for compatibility with past results, given how central the masking is for many other annotations, we decided to not update the repeatmasking of hg38. However, you can show the small differences between the RepeatMasker 3/RepBase from 2010 and RepeatMasker 4/DFAM from 2020 using the track "RepeatMasker Viz" in the same track group. It contains two subtracks, one with the old and one with the new data. Also, these tracks have many more visualisation options than the original RepeatMasker track.

However, the last track update time of this track at UCSC is not 2010, because we had to add repeatmasking annotations to the rarely used _alt and _fix "patch" sequences of the hg38 genome. The repeatmasking annotations of the main chromosomes were unaffected and have not changed since 2010. For more information on genome patches, see our [blog post](#).

Display Conventions and Configuration

In full display mode, this track displays up to ten different classes of repeats:

- Short interspersed nuclear elements (SINE), which include ALUs
- Long interspersed nuclear elements (LINE)
- Long terminal repeat elements (LTR), which include retrotransposons
- DNA repeat elements (DNA)
- Simple repeats (micro-satellites)
- Low complexity repeats
- Satellite repeats
- RNA repeats (including RNA, tRNA, rRNA, snRNA, scRNA, srpRNA)
- Other repeats, which includes class RC (Rolling Circle)
- Unknown



hg38 Repeat

RepeatMasker Information

Name: [LSU-rRNA_Hsa](#) (link requires [registration](#))

Family: rRNA

Class: rRNA

SW Score: 33388

Divergence: 0.2%

Deletions: 1.1%

Insertions: 0.8%

Begin in repeat: 878

End in repeat: 4452

Left in repeat: 583

Position: [chrUn_GL000220v1:158211-161802](#)

Band: Un_GL000220v1

Genomic Size: 3592

Strand: +

[View DNA for this feature](#) (hg38/Human)

[Data schema/format description and download](#)

[Go to RepeatMasker track controls](#)

Data last updated at UCSC: 2022-10-18



Description

This track was created by using Arian Smit's [RepeatMasker](#) program, which screens DNA sequences for interspersed repeats and low complexity DNA sequences. The program outputs a detailed annotation of the repeats that are present in the query sequence (represented by this track), as well as a modified version of the query sequence in which all the annotated repeats have been masked (generally available on the [Downloads](#) page). RepeatMasker uses the [Repbase Update](#) library of repeats from the [Genetic Information Research Institute](#) (GIRI). Repbase Update is described in Jurka (2000) in the References section below.

This track and the masking information in our [hg38 genome download FASTA files](#) was created in 2010 with the original RepBase library from 2010-03-02 and RepeatMasker 3.0.1. Since April 2019, RepBase is under a commercial license, we cannot distribute it or update the track using the RepBase library without a license. Therefore, and for compatibility with past results, given how central the masking is for many other annotations, we decided to not update the repeatmasking of hg38. However, you can show the small differences between the RepeatMasker 3/RepBase from 2010 and RepeatMasker 4/DFAM from 2020 using the track "RepeatMasker Viz" in the same track group. It contains two subtracks, one with the old and one with the new data. Also, these tracks have many more visualisation options than the original RepeatMasker track.

However, the last track update time of this track at UCSC is not 2010, because we had to add repeatmasking annotations to the rarely used _alt and _fix "patch" sequences of the hg38 genome. The repeatmasking annotations of the main chromosomes were unaffected and have not changed since 2010. For more information on genome patches, see our [blog post](#).

Display Conventions and Configuration

In full display mode, this track displays up to ten different classes of repeats:

- Short interspersed nuclear elements (SINE), which include ALUs
- Long interspersed nuclear elements (LINE)
- Long terminal repeat elements (LTR), which include retrotransposons
- DNA repeat elements (DNA)
- Simple repeats (micro-satellites)
- Low complexity repeats
- Satellite repeats
- RNA repeats (including RNA, tRNA, rRNA, snRNA, scRNA, srpRNA)
- Other repeats, which includes class RC (Rolling Circle)
- Unknown



hg38 NCBI RefSeq genes, curated subset (NM_*, NR_*, NP_* or YP_*) - NR_146144.1

RefSeq Gene RNA45SN2

Review ✓

RefSeq: [NR_146144.1](#) **Status:** Reviewed

Description: RNA, 45S pre-ribosomal N2

Molecule type: rRNA

Source: BestRefSeq

Biotype: rRNA

HGNC: [53518](#)

Entrez Gene: [109864279](#)

GeneCards: [RNA45SN2](#)

AceView: [RNA45SN2](#)

Summary of RNA45SN2

45S ribosomal DNA (rDNA) arrays, or clusters, are present on human chromosomes 13, 14, 15, 21 and 22, designated RNR1 through RNR5, respectively. Each cluster consists of multiple 45S rDNA repeat units that vary in number among individuals and chromosomes, with total diploid copy number estimates ranging from 60 to >800 repeat units in a human genome. The 45S rDNA repeat unit encodes a 45S rRNA precursor, transcribed by RNA polymerase I, which is processed to form the 18S, 5.8S and 28S rRNAs. This gene represents a copy of the 45S ribosomal RNA on chromosome 21. [provided by RefSeq, Mar 2017]. Sequence Note: The RefSeq transcript was derived from the reference genome assembly. The genomic coordinates were determined from alignments. Publication Note: This RefSeq record includes a subset of the publications that are available for this gene. Please see the Gene record to access additional publications.

mRNA/Genomic Alignments (NR_146144.1)

BROWSER	SIZE	IDENTITY	CHROMOSOME	STRAND	START	END	QUERY	START	END	TOTAL
browser	13314	100.0%	21	+	8205988	8219302	NR_146144.1	1	13315	13315

[View details of parts of alignment within browser window.](#)

Position: [chr21:8205988-8219302](#)

Band: 21p11.2

Genomic Size: 13315

Strand: +

Gene Symbol: RNA45SN2

Links to sequence:

- [Predicted mRNA](#) may be different from the genomic sequence.
- [Genomic Sequence](#) from assembly

[Data schema/format description and download](#)

[Go to NCBI RefSeq track controls](#)

Source data version: NCBI RefSeq GCF_000001405.40-RS_2024_08 (2024-08-27)

Data last updated at UCSC: 2024-09-11

Description

The NCBI RefSeq Genes composite track shows human protein-coding and non-protein-coding genes taken from the NCBI RNA reference sequences collection (RefSeq). All subtracks use coordinates provided by RefSeq, except for the UCSC RefSeq track, which UCSC produces by realigning the RefSeq RNAs to the genome. This realignment may result in occasional differences between the annotation coordinates provided by UCSC and NCBI. For RNA-seq analysis, we advise using NCBI aligned tables like RefSeq All or RefSeq Curated. See the [Methods](#) section for more details about how the different tracks were created.

Please visit NCBI's [Feedback for Gene and Reference Sequences \(RefSeq\)](#) page to make suggestions, submit additions and corrections, or ask for help concerning RefSeq records.

For more information on the different gene tracks, see our [Genes FAQ](#).



hg38 NCBI RefSeq genes, curated subset (NM_*, NR_*, NP_* or YP_*) - NR_146153.1

RefSeq Gene RNA5-8SN3



RefSeq: [NR_146153.1](#) **Status:** Validated
Description: RNA, 5.8S ribosomal RNA N3
Molecule type: rRNA
Source: BestRefSeq
Biotype: rRNA
HGNC: [53525](#)
Entrez Gene: [109910381](#)
GeneCards: [RNA5-8SN3](#)
AceView: [RNA5-8SN3](#)

Summary of RNA5-8SN3

45S ribosomal DNA (rDNA) arrays, or clusters, are present on human chromosomes 13, 14, 15, 21 and 22, designated RNR1 through RNR5, respectively. Each cluster consists of multiple 45S rDNA repeat units that vary in number among individuals and chromosomes, with total diploid copy number estimates ranging from 60 to >800 repeat units in a human genome. The 45S rDNA repeat unit encodes a 45S rRNA precursor, transcribed by RNA polymerase I, which is processed to form the 18S, 5.8S and 28S rRNAs. This gene represents a copy of the 5.8S ribosomal RNA on chromosome 21. [provided by RefSeq, Mar 2017]. Sequence Note: The RefSeq transcript was derived from the reference genome assembly. The genomic coordinates were determined from alignments. Publication Note: This RefSeq record includes a subset of the publications that are available for this gene. Please see the Gene record to access additional publications.

mRNA/Genomic Alignments (NR_146153.1)

The alignment you clicked on is first in the table below.

BROWSER	SIZE	IDENTITY	CHROMOSOME	STRAND	START	END	QUERY	START	END	TOTAL
browser	157	100.0%	21	+	8395606	8395762	NR_146153.1	1	157	157
browser	157	100.0%	21_ML143377v1_fix	+	442731	442887	NR_146153.1	1	157	157

[View details of parts of alignment within browser window.](#)

Position: [chr21:8395606-8395762](#)

Band: 21p11.2

Genomic Size: 157

Strand: +

Gene Symbol: RNA5-8SN3

Position: [chr21_ML143377v1_fix:442731-442887](#)

Band: 21_ML143377v1_fix

Genomic Size: 157

Strand: +

Gene Symbol: RNA5-8SN3

Links to sequence:

- [Predicted mRNA](#) may be different from the genomic sequence.
- [Genomic Sequence](#) from assembly

[Data schema/format description and download](#)

[Go to NCBI RefSeq track controls](#)

Source data version: NCBI RefSeq GCF_000001405.40-RS_2024_08 (2024-08-27)

Data last updated at UCSC: 2024-09-11

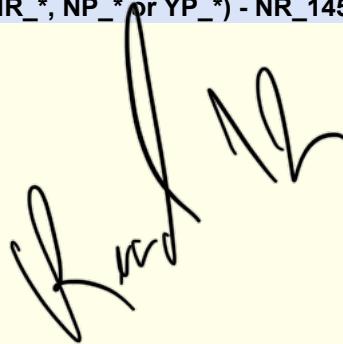
Description

The NCBI RefSeq Genes composite track shows human protein-coding and non-protein-coding genes taken from the NCBI RNA reference sequences collection (RefSeq). All subtracks use coordinates provided by RefSeq, except for the UCSC RefSeq track, which UCSC produces by realigning the RefSeq RNAs to the genome. This realignment may result in occasional differences between the annotation coordinates provided by UCSC and



hg38 NCBI RefSeq genes, curated subset (NM_*, NR_*, NP_* or YP_*) - NR_145821.1

RefSeq Gene RNA5-8SN1



RefSeq: [NR_145821.1](#) **Status:** Validated

Description: RNA, 5.8S ribosomal RNA N1

Molecule type: rRNA

Source: BestRefSeq

Biotype: rRNA

Synonyms: RNA5-8S4

HGNC: [53517](#)

Entrez Gene: [106632260](#)

GeneCards: [RNA5-8SN1](#)

AceView: [RNA5-8SN1](#)

Summary of RNA5-8SN1

45S ribosomal DNA (rDNA) arrays, or clusters, are present on human chromosomes 13, 14, 15, 21 and 22, designated RNR1 through RNR5, respectively. Each cluster consists of multiple 45S rDNA repeat units, that vary in number among individuals and chromosomes, with total diploid copy number estimates ranging from 60 to >800 repeat units in a human genome. The 45S rDNA repeat unit encodes a 45S rRNA precursor, transcribed by RNA polymerase I, which is processed to form the 18S, 5.8S and 28S rRNAs. Gene and RefSeq, in collaboration with HGNC, currently describe one 45S rDNA cluster, and one set of 45S precursor and product rRNAs, for each of the five human chromosomes to which these loci are localized. This gene is a representative copy of the 5.8S ribosomal RNA on chromosome 21. [provided by RefSeq, Feb 2017]. Sequence Note: The RefSeq transcript was derived from the reference genome assembly. The genomic coordinates were determined from alignments. Publication Note: This RefSeq record includes a subset of the publications that are available for this gene. Please see the Gene record to access additional publications.

mRNA/Genomic Alignments (NR_145821.1)

The alignment you clicked on is first in the table below.

BROWSER	SIZE	IDENTITY	CHROMOSOME	STRAND	START	END	QUERY	START	END	TOTAL	
browser	157	100.0%	21	+	8439822	8439978	NR_145821.1	1	157	157	
browser	157	100.0%	21	ML143377v1_fix	+	486947	487103	NR_145821.1	1	157	157

[View details of parts of alignment within browser window.](#)

Position: [chr21:8439822-8439978](#)

Band: 21p11.2

Genomic Size: 157

Strand: +

Gene Symbol: RNA5-8SN1

Position: [chr21_ML143377v1_fix:486947-487103](#)

Band: 21_ML143377v1_fix

Genomic Size: 157

Strand: +

Gene Symbol: RNA5-8SN1

Links to sequence:

- [Predicted mRNA](#) may be different from the genomic sequence.
- [Genomic Sequence](#) from assembly

[Data schema/format description and download](#)

[Go to NCBI RefSeq track controls](#)

Source data version: NCBI RefSeq GCF_00001405.40-RS_2024_08 (2024-08-27)

Data last updated at UCSC: 2024-09-11

Description

The NCBI RefSeq Genes composite track shows human protein-coding and non-protein-coding genes taken from the NCBI RNA reference sequences collection (RefSeq). All subtracks use coordinates provided by RefSeq.



hg38 Repeat

RepeatMasker Information

Name: [SSU-rRNA_Hsa](#) (link requires [registration](#))

Family: rRNA

Class: rRNA

SW Score: 14403

Divergence: 8.3%

Deletions: 1.3%

Insertions: 0.5%

Begin in repeat: 1

End in repeat: 1869

Left in repeat: 0

Position: [chr20:30486122-30487974](#)

Band: 20q11.21

Genomic Size: 1853

Strand: -

[View DNA for this feature](#) (hg38/Human)

[Data schema/format description and download](#)

[Go to RepeatMasker track controls](#)

Data last updated at UCSC: 2022-10-18



Description

This track was created by using Arian Smit's [RepeatMasker](#) program, which screens DNA sequences for interspersed repeats and low complexity DNA sequences. The program outputs a detailed annotation of the repeats that are present in the query sequence (represented by this track), as well as a modified version of the query sequence in which all the annotated repeats have been masked (generally available on the [Downloads](#) page). RepeatMasker uses the [Repbase Update](#) library of repeats from the [Genetic Information Research Institute](#) (GIRI). Repbase Update is described in Jurka (2000) in the References section below.

This track and the masking information in our [hg38 genome download FASTA files](#) was created in 2010 with the original RepBase library from 2010-03-02 and RepeatMasker 3.0.1. Since April 2019, RepBase is under a commercial license, we cannot distribute it or update the track using the RepBase library without a license. Therefore, and for compatibility with past results, given how central the masking is for many other annotations, we decided to not update the repeatmasking of hg38. However, you can show the small differences between the RepeatMasker 3/RepBase from 2010 and RepeatMasker 4/DFAM from 2020 using the track "RepeatMasker Viz" in the same track group. It contains two subtracks, one with the old and one with the new data. Also, these tracks have many more visualisation options than the original RepeatMasker track.

However, the last track update time of this track at UCSC is not 2010, because we had to add repeatmasking annotations to the rarely used _alt and _fix "patch" sequences of the hg38 genome. The repeatmasking annotations of the main chromosomes were unaffected and have not changed since 2010. For more information on genome patches, see our [blog post](#).

Display Conventions and Configuration

In full display mode, this track displays up to ten different classes of repeats:

- Short interspersed nuclear elements (SINE), which include ALUs
- Long interspersed nuclear elements (LINE)
- Long terminal repeat elements (LTR), which include retrotransposons
- DNA repeat elements (DNA)
- Simple repeats (micro-satellites)
- Low complexity repeats
- Satellite repeats
- RNA repeats (including RNA, tRNA, rRNA, snRNA, scRNA, srpRNA)
- Other repeats, which includes class RC (Rolling Circle)
- Unknown



hg38 Repeat

RepeatMasker Information

Name: [SSU-rRNA_Hsa](#) (link requires [registration](#))

Family: rRNA

Class: rRNA

SW Score: 14100

Divergence: 9.0%

Deletions: 1.6%

Insertions: 0.5%

Begin in repeat: 1

End in repeat: 1869

Left in repeat: 0

Position: [chrY:10197256-10199103](#)

Band: Yp11.2

Genomic Size: 1848

Strand: +

[View DNA for this feature](#) (hg38/Human)

[Data schema/format description and download](#)

[Go to RepeatMasker track controls](#)

Data last updated at UCSC: 2022-10-18



Description

This track was created by using Arian Smit's [RepeatMasker](#) program, which screens DNA sequences for interspersed repeats and low complexity DNA sequences. The program outputs a detailed annotation of the repeats that are present in the query sequence (represented by this track), as well as a modified version of the query sequence in which all the annotated repeats have been masked (generally available on the [Downloads](#) page). RepeatMasker uses the [Repbase Update](#) library of repeats from the [Genetic Information Research Institute](#) (GIRI). Repbase Update is described in Jurka (2000) in the References section below.

This track and the masking information in our [hg38 genome download FASTA files](#) was created in 2010 with the original RepBase library from 2010-03-02 and RepeatMasker 3.0.1. Since April 2019, RepBase is under a commercial license, we cannot distribute it or update the track using the RepBase library without a license. Therefore, and for compatibility with past results, given how central the masking is for many other annotations, we decided to not update the repeatmasking of hg38. However, you can show the small differences between the RepeatMasker 3/RepBase from 2010 and RepeatMasker 4/DFAM from 2020 using the track "RepeatMasker Viz" in the same track group. It contains two subtracks, one with the old and one with the new data. Also, these tracks have many more visualisation options than the original RepeatMasker track.

However, the last track update time of this track at UCSC is not 2010, because we had to add repeatmasking annotations to the rarely used _alt and _fix "patch" sequences of the hg38 genome. The repeatmasking annotations of the main chromosomes were unaffected and have not changed since 2010. For more information on genome patches, see our [blog post](#).

Display Conventions and Configuration

In full display mode, this track displays up to ten different classes of repeats:

- Short interspersed nuclear elements (SINE), which include ALUs
- Long interspersed nuclear elements (LINE)
- Long terminal repeat elements (LTR), which include retrotransposons
- DNA repeat elements (DNA)
- Simple repeats (micro-satellites)
- Low complexity repeats
- Satellite repeats
- RNA repeats (including RNA, tRNA, rRNA, snRNA, scRNA, srpRNA)
- Other repeats, which includes class RC (Rolling Circle)
- Unknown



hg38 Repeat

RepeatMasker Information

Name: tRNA-Glu-GAG (link requires [registration](#))

Family: tRNA

Class: tRNA

SW Score: 575

Divergence: 9.3%

Deletions: 0.0%

Insertions: 0.0%

Begin in repeat: 1

End in repeat: 75

Left in repeat: 0

Position: [chr8:12712565-12712639](#)

Band: 8p23.1

Genomic Size: 75

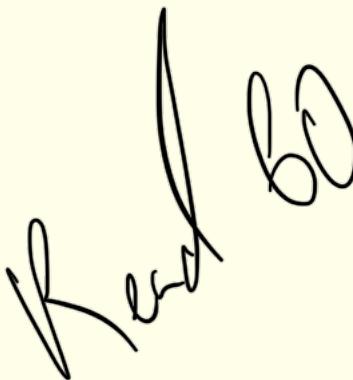
Strand: +

[View DNA for this feature](#) (hg38/Human)

[Data schema/format description and download](#)

[Go to RepeatMasker track controls](#)

Data last updated at UCSC: 2022-10-18



Description

This track was created by using Arian Smit's [RepeatMasker](#) program, which screens DNA sequences for interspersed repeats and low complexity DNA sequences. The program outputs a detailed annotation of the repeats that are present in the query sequence (represented by this track), as well as a modified version of the query sequence in which all the annotated repeats have been masked (generally available on the [Downloads](#) page). RepeatMasker uses the [Repbase Update](#) library of repeats from the [Genetic Information Research Institute](#) (GIRI). Repbase Update is described in Jurka (2000) in the References section below.

This track and the masking information in our [hg38 genome download FASTA files](#) was created in 2010 with the original RepBase library from 2010-03-02 and RepeatMasker 3.0.1. Since April 2019, RepBase is under a commercial license, we cannot distribute it or update the track using the RepBase library without a license. Therefore, and for compatibility with past results, given how central the masking is for many other annotations, we decided to not update the repeatmasking of hg38. However, you can show the small differences between the RepeatMasker 3/RepBase from 2010 and RepeatMasker 4/DFAM from 2020 using the track "RepeatMasker Viz" in the same track group. It contains two subtracks, one with the old and one with the new data. Also, these tracks have many more visualisation options than the original RepeatMasker track.

However, the last track update time of this track at UCSC is not 2010, because we had to add repeatmasking annotations to the rarely used _alt and _fix "patch" sequences of the hg38 genome. The repeatmasking annotations of the main chromosomes were unaffected and have not changed since 2010. For more information on genome patches, see our [blog post](#).

Display Conventions and Configuration

In full display mode, this track displays up to ten different classes of repeats:

- Short interspersed nuclear elements (SINE), which include ALUs
- Long interspersed nuclear elements (LINE)
- Long terminal repeat elements (LTR), which include retrotransposons
- DNA repeat elements (DNA)
- Simple repeats (micro-satellites)
- Low complexity repeats
- Satellite repeats
- RNA repeats (including RNA, tRNA, rRNA, snRNA, scRNA, srpRNA)
- Other repeats, which includes class RC (Rolling Circle)
- Unknown



Genomes

Genome Browser

Tools

Mirrors

Downloads

My Data

Projects

Help

About Us

hg38 NCBI RefSeq genes, curated subset (NM_*, NR_*, NP_* or YP_*) - NM_012309.5

RefSeq Gene SHANK2

RefSeq: [NM_012309.5](#) **Status:** Reviewed

Description: SH3 and multiple ankyrin repeat domains 2, transcript variant 1

Molecule type: mRNA

Source: BestRefSeq

Biotype: protein_coding

Synonyms: AUTS17,CORTBP1,CTTNBP1,ProSAP1,SHANK,SPANK-3

Other notes: isoform 1 is encoded by transcript variant 1

OMIM: [603290](#)

Protein: [NP_036441.2](#)

HGNC: [14295](#)

Entrez Gene: [22941](#)

GeneCards: [SHANK2](#)

AceView: [SHANK2](#)

Summary of SHANK2

This gene encodes a protein that is a member of the Shank family of synaptic proteins that may function as molecular scaffolds in the postsynaptic density of excitatory synapses. Shank proteins contain multiple domains for protein-protein interaction, including ankyrin repeats, and an SH3 domain. This particular family member contains a PDZ domain, a consensus sequence for cortactin SH3 domain-binding peptides and a sterile alpha motif. The alternative splicing demonstrated in Shank genes has been suggested as a mechanism for regulating the molecular structure of Shank and the spectrum of Shank-interacting proteins in the postsynaptic densities of the adult and developing brain. Alterations in the encoded protein may be associated with susceptibility to autism spectrum disorder. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Feb 2014].

mRNA/Genomic Alignments (NM_012309.5)

BROWSER	SIZE	IDENTITY	CHROMOSOME	STRAND	START	END	QUERY	START	END	TOTAL
browser	10830	100.0%	11	-	70467854	71252577	NM_012309.5	1	10830	10830

Position: [chr11:70467854-71252577](#)

Bands: 11q13.3 - 11q13.4

Genomic Size: 784724

Strand: -

Gene Symbol: SHANK2

CDS Start: complete

CDS End: complete

Links to sequence:

- [Predicted Protein](#)
- [Predicted mRNA](#) may be different from the genomic sequence.
- [Genomic Sequence](#) from assembly

[Data schema/format description and download](#)

[Go to NCBI RefSeq track controls](#)

Source data version: NCBI RefSeq GCF_000001405.40-RS_2024_08 (2024-08-27)

Data last updated at UCSC: 2024-09-11

Description

The NCBI RefSeq Genes composite track shows human protein-coding and non-protein-coding genes taken from the NCBI RNA reference sequences collection (RefSeq). All subtracks use coordinates provided by RefSeq, except for the UCSC RefSeq track, which UCSC produces by realigning the RefSeq RNAs to the genome. This realignment may result in occasional differences between the annotation coordinates provided by UCSC and NCBI. For RNA-seq analysis, we advise using NCBI aligned tables like RefSeq All or RefSeq Curated. See the [Methods](#) section for more details about how the different tracks were created.



hg38 NCBI RefSeq genes, curated subset (NM_*, NR_*, NP_* or YP_*) - NM_002204.4

RefSeq Gene ITGA3

RefSeq: [NM_002204.4](#) **Status:** Reviewed

Description: integrin subunit alpha 3

Molecule type: mRNA

Source: BestRefSeq

Biotype: protein_coding

Synonyms: CD49C,FRP-2,GAP-B3,GAPB3,ILNEB,JEB7,MSK18,VCA-2,VL3A,VLA3a

OMIM: [605025](#)

Protein: [NP_002195.1](#)

HGNC: [6139](#)

Entrez Gene: [3675](#)

GeneCards: [ITGA3](#)

AceView: [ITGA3](#)

Summary of ITGA3

The gene encodes a member of the integrin alpha chain family of proteins. Integrins are heterodimeric integral membrane proteins composed of an alpha chain and a beta chain that function as cell surface adhesion molecules. The encoded preproprotein is proteolytically processed to generate light and heavy chains that comprise the alpha 3 subunit. This subunit joins with a beta 1 subunit to form an integrin that interacts with extracellular matrix proteins including members of the laminin family. Expression of this gene may be correlated with breast cancer metastasis. [provided by RefSeq, Oct 2015]. Publication Note: This RefSeq record includes a subset of the publications that are available for this gene. Please see the Gene record to access additional publications. Evidence Data: Transcript exon combination: BC150190.1, M59911.1 [ECO:0000332]; RNAseq introns: single sample supports all introns SAMEA1965299, SAMEA1966682 [ECO:0000348]. RefSeq Attributes: MANE Ensembl match: ENST00000320031.13/ ENSP00000315190.8; RefSeq Select criteria: based on conservation, expression, longest protein.

mRNA/Genomic Alignments (NM_002204.4)

BROWSER	SIZE	IDENTITY	CHROMOSOME	STRAND	START	END	QUERY	START	END	TOTAL
browser	4889	100.0%	17	+	50056110	50090481	NM_002204.4	1	4889	4889

Position: [chr17:50056110-50090481](#)

Band: 17q21.33

Genomic Size: 34372

Strand: +

Gene Symbol: ITGA3

CDS Start: complete

CDS End: complete

Links to sequence:

- [Predicted Protein](#)
- [Predicted mRNA](#) may be different from the genomic sequence.
- [Genomic Sequence](#) from assembly

[Data schema/format description and download](#)

[Go to NCBI RefSeq track controls](#)

Source data version: NCBI RefSeq GCF_000001405.40-RS_2024_08 (2024-08-27)

Data last updated at UCSC: 2024-09-11

Description

The NCBI RefSeq Genes composite track shows human protein-coding and non-protein-coding genes taken from the NCBI RNA reference sequences collection (RefSeq). All subtracks use coordinates provided by RefSeq, except for the UCSC RefSeq track, which UCSC produces by realigning the RefSeq RNAs to the genome. This realignment may result in occasional differences between the annotation coordinates provided by UCSC and NCBI. For RNA-seq analysis, we advise using NCBI aligned tables like RefSeq All or RefSeq Curated. See the [Methods](#) section for more details about how the different tracks were created.



hg38 Human: lincRNA and TUCP transcripts (TCONS_I2_00023649)

Position: [chr5:4451930-4866221](#)

Band: 5p15.32

Genomic Size: 414292

Strand: -

Links to sequence:

- Non-protein coding gene or gene fragment, no protein prediction available.
- [Predicted mRNA](#) from genomic sequences
- [Genomic Sequence](#) from assembly

[Data schema/format description and download](#)

[Go to lincRNA TUCP track controls](#)

Data coordinates converted via liftOver from: Feb. 2009 (GRCh37/hg19)

Data last updated at UCSC: 2015-04-08

Note: lifted from hg19

Description

This track displays the Human Body Map lincRNAs (large intergenic non coding RNAs) and TUCPs (transcripts of uncertain coding potential), as well as their expression levels across 22 human tissues and cell lines. The Human Body Map catalog was generated by integrating previously existing annotation sources with transcripts that were de-novo assembled from RNA-Seq data. These transcripts were collected from ~4 billion RNA-Seq reads across 24 tissues and cell types.

Expression abundance was estimated by Cufflinks (Trapnell et al., 2010) based on RNA-Seq. Expression abundances were estimated on the gene locus level, rather than for each transcript separately and are given as raw FPKM. The prefixes tcons_ and tcons_I2_ are used to describe lincRNAs and TUCP transcripts, respectively. Specific details about the catalog generation and data sets used for this study can be found in Cabili et al (2011). Extended characterization of each transcript in the human body map catalog can be found at the [Human lincRNA Catalog](#) website.

Expression abundance scores range from 0 to 1000, and are displayed from light blue to dark blue respectively:



Credits

The body map RNA-Seq data was kindly provided by the Gene Expression Applications research group at [Illumina](#).

References

Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. [Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses](#). *Genes Dev.* 2011 Sep 15;25(18):1915-27. PMID: [21890647](#); PMC: [PMC3185964](#)

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. [Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation](#). *Nat Biotechnol.* 2010 May;28(5):511-5. PMID: [20436464](#); PMC: [PMC3146043](#)



hg38 Repeat

RepeatMasker Information

Name: [\(TG\)n](#) (link requires [registration](#))

Family: Simple_repeat

Class: Simple_repeat

SW Score: 34

Divergence: 5.6%

Deletions: 0.0%

Insertions: 0.0%

Begin in repeat: 1

End in repeat: 38

Left in repeat: 0

Position: [chr9:29288340-29288377](#)

Band: 9p21.1

Genomic Size: 38

Strand: +

[View DNA for this feature](#) (hg38/Human)

[Data schema/format description and download](#)

[Go to RepeatMasker track controls](#)

Data last updated at UCSC: 2022-10-18



Description

This track was created by using Arian Smit's [RepeatMasker](#) program, which screens DNA sequences for interspersed repeats and low complexity DNA sequences. The program outputs a detailed annotation of the repeats that are present in the query sequence (represented by this track), as well as a modified version of the query sequence in which all the annotated repeats have been masked (generally available on the [Downloads](#) page). RepeatMasker uses the [Repbase Update](#) library of repeats from the [Genetic Information Research Institute](#) (GIRI). Repbase Update is described in Jurka (2000) in the References section below.

This track and the masking information in our [hg38 genome download FASTA files](#) was created in 2010 with the original RepBase library from 2010-03-02 and RepeatMasker 3.0.1. Since April 2019, RepBase is under a commercial license, we cannot distribute it or update the track using the RepBase library without a license. Therefore, and for compatibility with past results, given how central the masking is for many other annotations, we decided to not update the repeatmasking of hg38. However, you can show the small differences between the RepeatMasker 3/RepBase from 2010 and RepeatMasker 4/DFAM from 2020 using the track "RepeatMasker Viz" in the same track group. It contains two subtracks, one with the old and one with the new data. Also, these tracks have many more visualisation options than the original RepeatMasker track.

However, the last track update time of this track at UCSC is not 2010, because we had to add repeatmasking annotations to the rarely used _alt and _fix "patch" sequences of the hg38 genome. The repeatmasking annotations of the main chromosomes were unaffected and have not changed since 2010. For more information on genome patches, see our [blog post](#).

Display Conventions and Configuration

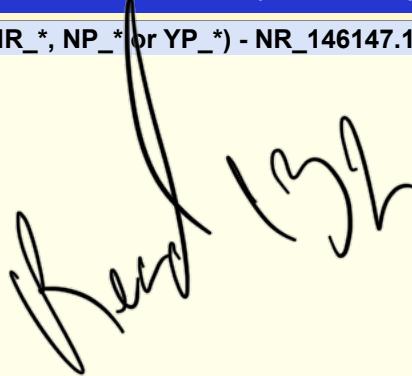
In full display mode, this track displays up to ten different classes of repeats:

- Short interspersed nuclear elements (SINE), which include ALUs
- Long interspersed nuclear elements (LINE)
- Long terminal repeat elements (LTR), which include retrotransposons
- DNA repeat elements (DNA)
- Simple repeats (micro-satellites)
- Low complexity repeats
- Satellite repeats
- RNA repeats (including RNA, tRNA, rRNA, snRNA, scRNA, srpRNA)
- Other repeats, which includes class RC (Rolling Circle)
- Unknown



hg38 NCBI RefSeq genes, curated subset (NM_*, NR_*, NP_* or YP_*) - NR_146147.1

RefSeq Gene RNA5-8SN2



RefSeq: [NR_146147.1](#) **Status:** Validated

Description: RNA, 5.8S ribosomal RNA N2

Molecule type: rRNA

Source: BestRefSeq

Biotype: rRNA

Synonyms: RNA5-8N2

HGNC: [53521](#)

Entrez Gene: [109864281](#)

GeneCards: [RNA5-8SN2](#)

AceView: [RNA5-8SN2](#)

Summary of RNA5-8SN2

45S ribosomal DNA (rDNA) arrays, or clusters, are present on human chromosomes 13, 14, 15, 21 and 22, designated RNR1 through RNR5, respectively. Each cluster consists of multiple 45S rDNA repeat units that vary in number among individuals and chromosomes, with total diploid copy number estimates ranging from 60 to >800 repeat units in a human genome. The 45S rDNA repeat unit encodes a 45S rRNA precursor, transcribed by RNA polymerase I, which is processed to form the 18S, 5.8S and 28S rRNAs. This gene represents a copy of the 5.8S ribosomal RNA on chromosome 21. [provided by RefSeq, Mar 2017]. Sequence Note: The RefSeq transcript was derived from the reference genome assembly. The genomic coordinates were determined from alignments. Publication Note: This RefSeq record includes a subset of the publications that are available for this gene. Please see the Gene record to access additional publications.

mRNA/Genomic Alignments (NR_146147.1)

BROWSER	SIZE	IDENTITY	CHROMOSOME	STRAND	START	END	QUERY	START	END	TOTAL
browser	157	100.0%	21	+	8212571	8212727	NR_146147.1	1	157	157

[View details of parts of alignment within browser window.](#)

Position: [chr21:8212571-8212727](#)

Band: 21p11.2

Genomic Size: 157

Strand: +

Gene Symbol: RNA5-8SN2

Links to sequence:

- [Predicted mRNA](#) may be different from the genomic sequence.
- [Genomic Sequence](#) from assembly

[Data schema/format description and download](#)

[Go to NCBI RefSeq track controls](#)

Source data version: NCBI RefSeq GCF_000001405.40-RS_2024_08 (2024-08-27)

Data last updated at UCSC: 2024-09-11

Description

The NCBI RefSeq Genes composite track shows human protein-coding and non-protein-coding genes taken from the NCBI RNA reference sequences collection (RefSeq). All subtracks use coordinates provided by RefSeq, except for the UCSC RefSeq track, which UCSC produces by realigning the RefSeq RNAs to the genome. This realignment may result in occasional differences between the annotation coordinates provided by UCSC and NCBI. For RNA-seq analysis, we advise using NCBI aligned tables like RefSeq All or RefSeq Curated. See the [Methods](#) section for more details about how the different tracks were created.

Please visit NCBI's [Feedback for Gene and Reference Sequences \(RefSeq\)](#) page to make suggestions, submit additions and corrections, or ask for help concerning RefSeq records.

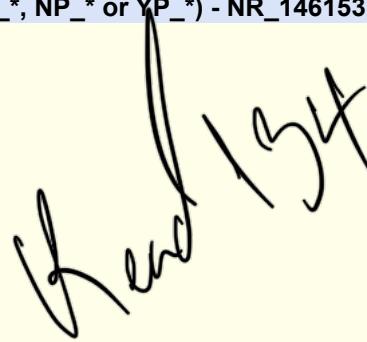
For more information on the different gene tracks, see our [Genes FAQ](#).



hg38 NCBI RefSeq genes, curated subset (NM_*, NR_*, NP_* or YP_*) - NR_146153.1

RefSeq Gene RNA5-8SN3

RefSeq: [NR_146153.1](#) **Status:** Validated
Description: RNA, 5.8S ribosomal RNA N3
Molecule type: rRNA
Source: BestRefSeq
Biotype: rRNA
HGNC: [53525](#)
Entrez Gene: [109910381](#)
GeneCards: [RNA5-8SN3](#)
AceView: [RNA5-8SN3](#)



Summary of RNA5-8SN3

45S ribosomal DNA (rDNA) arrays, or clusters, are present on human chromosomes 13, 14, 15, 21 and 22, designated RNR1 through RNR5, respectively. Each cluster consists of multiple 45S rDNA repeat units that vary in number among individuals and chromosomes, with total diploid copy number estimates ranging from 60 to >800 repeat units in a human genome. The 45S rDNA repeat unit encodes a 45S rRNA precursor, transcribed by RNA polymerase I, which is processed to form the 18S, 5.8S and 28S rRNAs. This gene represents a copy of the 5.8S ribosomal RNA on chromosome 21. [provided by RefSeq, Mar 2017]. Sequence Note: The RefSeq transcript was derived from the reference genome assembly. The genomic coordinates were determined from alignments. Publication Note: This RefSeq record includes a subset of the publications that are available for this gene. Please see the Gene record to access additional publications.

mRNA/Genomic Alignments (NR_146153.1)

The alignment you clicked on is first in the table below.

BROWSER	SIZE	IDENTITY	CHROMOSOME	STRAND	START	END	QUERY	START	END	TOTAL
browser	157	100.0%	21	+	8395606	8395762	NR_146153.1	1	157	157
browser	157	100.0%	21_ML143377v1_fix	+	442731	442887	NR_146153.1	1	157	157

[View details of parts of alignment within browser window.](#)

Position: [chr21:8395606-8395762](#)

Band: 21p11.2

Genomic Size: 157

Strand: +

Gene Symbol: RNA5-8SN3

Position: [chr21_ML143377v1_fix:442731-442887](#)

Band: 21_ML143377v1_fix

Genomic Size: 157

Strand: +

Gene Symbol: RNA5-8SN3

Links to sequence:

- [Predicted mRNA](#) may be different from the genomic sequence.
- [Genomic Sequence](#) from assembly

[Data schema/format description and download](#)

[Go to NCBI RefSeq track controls](#)

Source data version: NCBI RefSeq GCF_000001405.40-RS_2024_08 (2024-08-27)

Data last updated at UCSC: 2024-09-11

Description

The NCBI RefSeq Genes composite track shows human protein-coding and non-protein-coding genes taken from the NCBI RNA reference sequences collection (RefSeq). All subtracks use coordinates provided by RefSeq, except for the UCSC RefSeq track, which UCSC produces by realigning the RefSeq RNAs to the genome. This realignment may result in occasional differences between the annotation coordinates provided by UCSC and



hg38 NCBI RefSeq genes, curated subset (NM_*, NR_*, NP_* or YP_*) - NR_145819.1

RefSeq: [NR_145819.1](#) Status: Reviewed

Description: RNA, 45S pre-ribosomal N1

Molecule type: rRNA

Source: BestRefSeq

Biotype: rRNA

Synonyms: RNA45S4

HGNC: [53514](#)Entrez Gene: [106631777](#)GeneCards: [RNA45SN1](#)AceView: [RNA45SN1](#)

Summary of RNA45SN1

45S ribosomal DNA (rDNA) arrays, or clusters, are present on human chromosomes 13, 14, 15, 21 and 22, designated RNR1 through RNR5, respectively. Each cluster consists of multiple 45S rDNA repeat units that vary in number among individuals and chromosomes, with total diploid copy number estimates ranging from 60 to >800 repeat units in a human genome. The 45S rDNA repeat unit encodes a 45S rRNA precursor, transcribed by RNA polymerase I, which is processed to form the 18S, 5.8S and 28S rRNAs. Gene and RefSeq, in collaboration with HGNC, currently describe one 45S rDNA cluster, and one set of 45S precursor and product rRNAs, for each of the five human chromosomes to which these loci are localized. This gene is a representative copy of the 45S pre-ribosomal RNA on chromosome 21. [provided by RefSeq, Feb 2017]. Sequence Note: The RefSeq transcript was derived from the reference genome assembly. The genomic coordinates were determined from alignments. Publication Note: This RefSeq record includes a subset of the publications that are available for this gene. Please see the Gene record to access additional publications.

mRNA/Genomic Alignments (NR_145819.1)

The alignment you clicked on is first in the table below.

BROWSER	SIZE	IDENTITY	CHROMOSOME	STRAND	START	END	QUERY	START	END	TOTAL	
browser	13351	100.0%	21	+	8433222	8446572	NR_145819.1	1	13351	13351	
browser	13351	100.0%	21	ML143377v1_fix	+	480347	493697	NR_145819.1	1	13351	13351

[View details of parts of alignment within browser window.](#)

Position: [chr21:8433222-8446572](#)

Band: 21p11.2

Genomic Size: 13351

Strand: +

Gene Symbol: RNA45SN1

Position: [chr21_ML143377v1_fix:480347-493697](#)

Band: 21_ML143377v1_fix

Genomic Size: 13351

Strand: +

Gene Symbol: RNA45SN1

Links to sequence:

- [Predicted mRNA](#) may be different from the genomic sequence.
- [Genomic Sequence](#) from assembly

[Data schema/format description and download](#)

[Go to NCBI RefSeq track controls](#)

Source data version: NCBI RefSeq GCF_00001405.40-RS_2024_08 (2024-08-27)

Data last updated at UCSC: 2024-09-11

Description

The NCBI RefSeq Genes composite track shows human protein-coding and non-protein-coding genes taken from the NCBI RNA reference sequences collection (RefSeq). All subtracks use coordinates provided by RefSeq.