

i.

UCSC Genome Browser on Human May 2004 (NCBI35/hg17) Assembly

move <<< << < > >> >>> zoom in [1.5x] [2x] [10x] base zoom out [1.5x] [2x] [10x]

chr10:83,728,635-83,728,925 291 bp | enter position, gene symbol or search terms go

Now European server available! Click here for more information.

Scale: chr10: 83,728,466 83,728,456 83,728,506 83,728,556 83,728,606 83,728,656

Human MM9 Human ENSG

Human ETSs That Have Been Selected

Vertebrate Multiple Alignment & Conservation

Conservation

Genomes Genome Browser Tools Mirrors Downloads My Data About Us View Help

UCSC Genome Brow PDF/PS DNA In Other Genomes (Convert)

chr10:83,728,344-83,729,21 New Euro Ensembl Configure Browser Default Tracks Default Track Order Reset All User Settings

Symbol or search terms go here for more information.

Get DNA in Window (hg17/Human)

Get DNA for Position chr10:83,728,344-83,729,216

Note: This page retrieves genomic DNA for a single region. If you would prefer to get DNA for many items in a particular track, or get DNA with formatting options based on gene structure (introns, exons, UTRs, etc.), try using the Table Browser format.

Sequence Retrieval Region Options:

Add 0 extra bases upstream (5') and 0 extra downstream (3')

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome.

Sequence Formatting Options:

All upper case
 All lower case
 Mask repeats: to lower case to N
 Reverse complement (get "L" strand sequence)

Note: The "Mask repeats" option applies only to "get DNA", not to "extended case/color options".

ii.

range=chr10:83728344-83729216 5'pad=0 3'pad=0 strand=+ repeatMasking=none

```

-AATAGGAAGATAAAAGTTAGGTATTGTCTACAGAGCATT
|TCACAGTGATCTTATTTCTGAAACTCTTAATATCATATATATATT
|TGCATCTTCTTTCTTTCTTTCTTTAGAGATAGGGCTTGTCTG
|GTTGCCTAGGTTGCTCTAAACTGCTGGCCTCAAGTGTATCCCTTGGCCTC
|AGTCTCCAAAAAACATGGGATTATAGGTATGAGCCACCATGCCCTGGCCA
|GCATCTTATGTTTGTCAGCATCTAGCCATTAATGTATTTTAAAAATTCC
|TGGCGGGCGTGGTGGCTCATGCCTGTAATCCCAGCACTTGGGAGGCCG
|AGGTGGCGGATCACAAAGGTCAAGAGATCAAGACCCTGGCAACATG
|ATGAAACCCGTCCTATTAAAAGTACAAAAATTAGCTGGCGTAGTGGC
|GCACGCCCTGTAGTCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTT

```

iii.

iv.

```

-AATAGGAAGATAAAAGTTAGGTATTGTCTACAGAGCATT
|TCACAGTGATCTTATTTCTGAAACTCTTAATATCATATATATT
|TGCATCTTCTTTCTTTCTTTCTTTAGAGATAGGGCTTGTCTG
|GTTGCCTAGGTTGCTCTAAACTGCTGGCCTCAAGTGTATCCCTTGGCCTC
|AGTCTCCAAAAAACATGGGATTATAGGTATGAGCCACCATGCCCTGGCCA
|GCATCTTATGTTTGTCAGCATCTAGCCATTAATGTATTTTAAAAATTCC
|TGGCGGGCGTGGTGGCTCATGCCTGTAATCCCAGCACTTGGGAGGCCG
|AGGTGGCGGATCACAAAGGTCAAGAGATCAAGACCCTGGCAACATG
|ATGAAACCCGTCCTATTAAAAGTACAAAAATTAGCTGGCGTAGTGGC
|GCACGCCCTGTAGTCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTT

```

Figure S1. Steps to get DNA sequence via the UCSC genome browser.

Genomes Genome Browser Tools Mirrors Downloads My Data About Us Help

Human BLAT Search

BLAT Search Genome

Genome: Human Assembly: May 2004 (NCBI35/hg17)

Tools

- [Blat](#) (highlighted with a red dotted box)
- [Table Browser](#)
- [Variant Annotation Integrator](#)
- [Gene Sorter](#)
- [Genome Graphs](#)
- [In-Silico PCR](#)
- [LiftOver](#)
- [VisiGene](#)
- [Other Utilities](#)

Output: [Table](#) (highlighted with a red dotted box) Output type: [hyperlink](#)

BLAT Search Results

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
browser details	hg17_dna	533	1	873	873	94.8%	10	+	81315688	81316242	555
browser details	hg17_dna	287	300	604	873	97.1%	5	-	101791646	101791950	305
browser details	hg17_dna	286	282	594	873	96.5%	15	+	83651541	83651862	322
browser details	hg17_dna	281	302	600	873	97.0%	7	-	132943638	132943936	299
browser details	hg17_dna	281	299	602	873	96.4%	4	-	42153003	42153307	305
browser details	hg17_dna	279	289	602	873	95.2%	6	+	49946897	49947210	314
browser details	hg17_dna	277	284	602	873	93.7%	9	+	125798716	125799038	323
browser details	hg17_dna	277	303	598	873	97.0%	5	+	155209819	155210131	313

v.

Genome: Chimp Assembly: Feb. 2011 (CSAC 2.1.4/panTro4) Query type: BLAT's guess Sort output: query,score Output type: hyperlink

```
>hg17_dna range=chr10:83728344-83729216 5' pad=0 3' pad=0 strand=+
repeatMasking=none
AAGAAAACAAATAATAGGAAGATAAAAGTTAGGTATTGTCATCAGAGCATTAA
ACACAGTGATCTTATTTCTGAAACTCTTAATATCATATATATATATATT
TGCATCTTCTTTCTTTCTTTGTTTTAGAGATAAGGTCTTGTGT
GTTGCTAGGTTGTCTAAACTGCTGGCCTCAAGTGTCTCTTGTGT
AGTCTCCAAAAAACATTGGGATTATAGGTATGAGCCACCATGCCCTGGCCCA
GCATCTTATTTGTAGCATCTAGGCCATAATGTATTTTAAAAATTCCC
TGGCCGGCGCTGGGGCTCATGCCTGTAATCCAGCCTTGGGAGGCCG
AGGTGGGCGGATCACAAAGGTCAAGAGATCAGACCATCTGGCCAAACATG
ATGAAAACCCGTCTCTATTAAAAGTACA AAAAATTAGCTGGCGTAGTGC
GCACGCGTGTAGTCCCAGCTACTGGGAGGTAGGGCAGGGAAATCGCTT
GAACCTGGGAGGCAGGGTTGCAGTGAGCCGGATTGTGCCACTGCACTC
CAGCTGGCGACAGGGCGGACTCCGTCAAAAAAAAAAAAAA
```

submit I'm feeling lucky clear

vi.



vii.

Alignment of hg17_dna

[hg17_dna](#)
[Chimp chr10](#)
[block1](#)
[together](#)

Alignment of hg17_dna and chr10:81315688-81316242

Click on links in the frame to the left to navigate through the alignment. Matching bases in cDNA

cDNA hg17_dna

```
AAGAAAACAA TAATAGGAAG ATAAAGTTAG GTATTGTCAT CAGAGCATTAA 50
ACACAGTGAT CTTTATTTCG TAAACTCTTA ATATCAKATA TATATATATT 100
TGCATCTTCTTCTTTCTTTGTTTTAGAGATAAGGTCTTGTGT 150
GTTGCTAGGTTGTCTAAACTGCTGGCCTCAAGTGTCTCTTGTGT 200
AGTCTCCAAAAACATTGGGATTATAGGTATGAGCCACCATGCCCTGGCCCA 250
GCATCTTATTTGTAGCATCTAGGCCATAATGTATTTTAAAAATTCCC 300
TGGCCGGCGCTGGGGCTCATGCCTGTAATCCAGCCTTGGGAGGCCG 350
AGGTGGGCGGATCACAAAGGTCAAGAGATCAGACCATCTGGCCAAACATG 400
ATGAAAACCCGTCTCTATTAAAAGTACA AAAAATTAGCTGGCGTAGTGC 450
GCACGCGTGTAGTCCCAGCTACTGGGAGGTAGGGCAGGGAATCGCTT 500
GAACCTGGGAGGCAGGGTTGCAGTGAGCCGGATTGTGCCACTGCACTC 550
CAGCTGGCGACAGGGCGGACTCCGTCAAAAAAAAAAAAAA 600
GGAGCTGAGGTT GCACTGGAGGTT GGAGTGGAGC GGGATGGGGGCTGGC 650
CAGCTGGCGACAGGGCGGACTCCGTCAAAAAAAAAAAAAA 700
GGATTTATCTC TGATTTCCCTC CACAGAGTCT CTATGGGAAm TATTGTAGAG 750
GTATGGGCTT CCACCCCTAAC AATAATATAA CATAAAAGGA AATTCTAAAA 800
TCTTGCTCAT ATCTAATGGG AGGATTTTG GGAGGGCATTTG GAGGGATAGAGCATACTTTAACACTT 850
GCTAGAGAAC TATGGGGAA CTGTTTCAAA TGAGAGAAAA TGCAAGGCTA 900
GGCAGTATGA AATgtATCTT GGT
```

viii.



ix.

>Human cacagctcgtacggggactgggacccctttccagtcgtccggatggaaacttgcgtttttggcaggaaacttgcgtttttggcaggaaactttttggatctTTACGTAGTGGTATTTCAGTAAGGAAATAGAGCATACTTTATTTAACACTTG
>Chimp cacagctcgtacggggactgggacccctttccagtcgtccggatggaaacttgcgtttttggcaggaaacttgcgtttttggatctTTAtGTAGTGGTATTTCAGTAAGGAAATAGAGCATACTTTATTTAACACTTG
>Gorilla cacagctcgtacggggactgggacccctttccagtcgtccggatggaaacttgcgtttttggcaggaaacttgcgtttttggatctTTAtGTAGTGGTATTTCAGTAAGGAAATAGAGCATACTTTATTTAACACTTG
>Orangutan cacagctcgtacattggggactgggacccctttccagtcgtccggatggaaacttgcgtttttggcaggaaacttgcgtttttggatctTTAtGTAGTGGTATTTCAGTAAGGAAATAGAGCATACTTTATTTAACACTTG
>Gibbon cacagctcgtacggggatgggacccctttccagtcgtccggatggaaacttgcgtttttggcaggaaacttgcgtttttggatctTTAtGTAGTGTATTTCAGTAAGGAAATAGAGCATACTTTATTTAACACTTG
>Rhesus gtctgactctggggactgggacccctttccagtcgtttttggcggaaacttgcgtttttggatctTTAtGTAGTGTATTTCAGTAAGGAAATAGAGCATACTTTATTTAACACTTG

Figure S2. Retrieve similar DNA sequence using BLAT.

Basic Options

or

Sequence:

```
>Human
cacagcttgcacgttggggactgggccccctttccatgtggcggtttggcagga
actctggaaatgttgcacttaaaaactggatTTACGTAGTGGTATTCC
AAGTAAGGAAATAGAGCATACTTTAAACAACCTTGATGTCTGACTTACTTT
TAAACACTAAGATTATGTGGCTTCATTGTTACTCTTACCTCAGGCCACAGACGTCA
TAGAAGAGCTTGGAAAAAGTTAAATTCTGGGAGATAAGATCAAAGAGGGT
```

Search Engine: ablast rmblast hmmer cross_match

Speed/Sensitivity: rush quick default slow

DNA source:

Return Format: html tar file

X. **Return Method:** html email

Results

Right-click and select "Save As" to save results to your computer or click on the link to view the file in the browser.

Annotation File: [RM2ceupload_1372728873.out.html](#) (NEW XHTML Format)

[RM2ceupload_1372728873.out.txt](#) (Text File Format)

Masked File: [RM2ceupload_1372728873.masked](#)

Alignment File: [RM2ceupload_1372728873.align](#)

xii.



SW score	perc div.	perc del.	perc ins.	query sequence	begin	end	(left)	matching repeat	repeat class/family	begin	end	(left)	ID	
551	26.1	8.1	3.2	Chimp	17	226	(844)	+	MER115	DNA/hAT-Tip100	450	669	(24)	1
17	18.5	0.0	0.0	Chimp	802	832	(238)	+	A-rich	Low_complexity	1	31	(0)	2
17	12.1	2.4	7.0	Chimp	959	1000	(70)	+	GA-rich	Low_complexity	1	43	(0)	3
575	26.1	9.7	1.3	Gibbon	32	248	(821)	+	MER115	DNA/hAT-Tip100	459	693	(0)	4
17	18.5	0.0	0.0	Gibbon	801	831	(238)	+	A-rich	Low_complexity	1	31	(0)	5
17	12.1	2.4	7.0	Gibbon	958	999	(70)	+	GA-rich	Low_complexity	1	43	(0)	6
564	27.5	8.2	2.9	Gorilla	17	248	(822)	+	MER115	DNA/hAT-Tip100	450	693	(0)	7
17	18.5	0.0	0.0	Gorilla	802	832	(238)	+	A-rich	Low_complexity	1	31	(0)	8
15	12.8	2.5	7.3	Gorilla	959	998	(72)	+	GA-rich	Low_complexity	1	41	(0)	9
542	28.4	8.2	2.9	Human	17	248	(822)	+	MER115	DNA/hAT-Tip100	450	693	(0)	10
17	18.5	0.0	0.0	Human	802	832	(238)	+	A-rich	Low_complexity	1	31	(0)	11
15	12.8	2.5	7.3	Human	959	998	(72)	+	GA-rich	Low_complexity	1	41	(0)	12
597	27.8	7.8	2.5	Orangutan	17	248	(814)	+	MER115	DNA/hAT-Tip100	450	693	(0)	13
17	14.6	2.3	6.7	Orangutan	951	994	(68)	+	GA-rich	Low_complexity	1	45	(0)	14
609	24.2	10.2	1.4	Rhesus	28	232	(898)	+	MER115	DNA/hAT-Tip100	459	681	(12)	15

Figure S3. TE identification using RepeatMasker.

Sequence source: All

Force translated search:

Search for identity:

Report simple repeats:

Mask pseudogenes:

Enter query file name:
(Up to 2MB; IG-Stanford, FASTA, GENBANK, EMBL formats are supported)

OR

XIII.
Paste query sequences here:
(Up to 2MB; IG-Stanford, FASTA, GENBANK, EMBL formats are supported)

```
cacagctctgacgttggggactggcacccttccagctgccgtttggcaggaactctgag
aatgcttgaactaaaactggcatttggatctTATGTAGTGGTATTCCAAGTAAGAGGAATAGA
GCATACTTTATTAAACAACCTTGATGTCCTGACTACTTTAACACTAAGATTATGTGGGCT
TCCATTGTTACTTACCTCAGGCCACAGAtGTCATAGAAGAGCTTGGAAAAAGTTAAATTCTTGGG
GAGAATAGAATAGAATCTAAAGAGGGTGATTGGACTCCAGCATCTCTCAGGTTCTGTTACCC
CCTGCACCTCTCTGGTTCTGAGaCAGGTGCTTATCTCTTGGAGATGTGCTGAGGCAGAATCA
```



Figure S4. TE identification using Censor