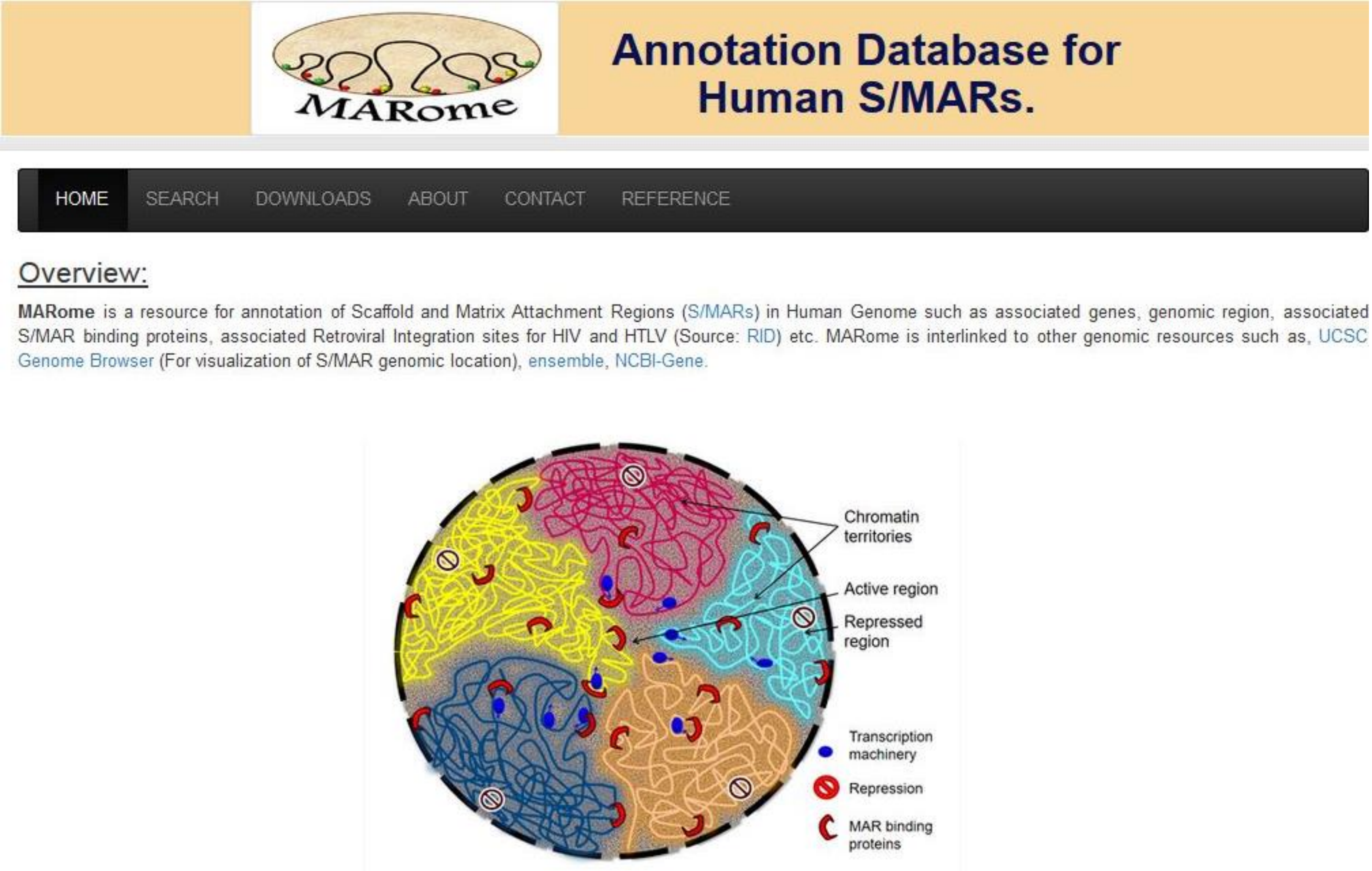


Annotation Database for Human S/MARs.

Overview:

MARome is a resource for annotation of Scaffold and Matrix Attachment Regions (S/MARs) in Human Genome such as associated genes, genomic region, associated S/MAR binding proteins, associated Retroviral Integration sites for HIV and HTLV (Source: RID) etc. MARome is interlinked to other genomic resources such as, UCSC Genome Browser (For visualization of S/MAR genomic location), ensemble, NCBI-Gene.

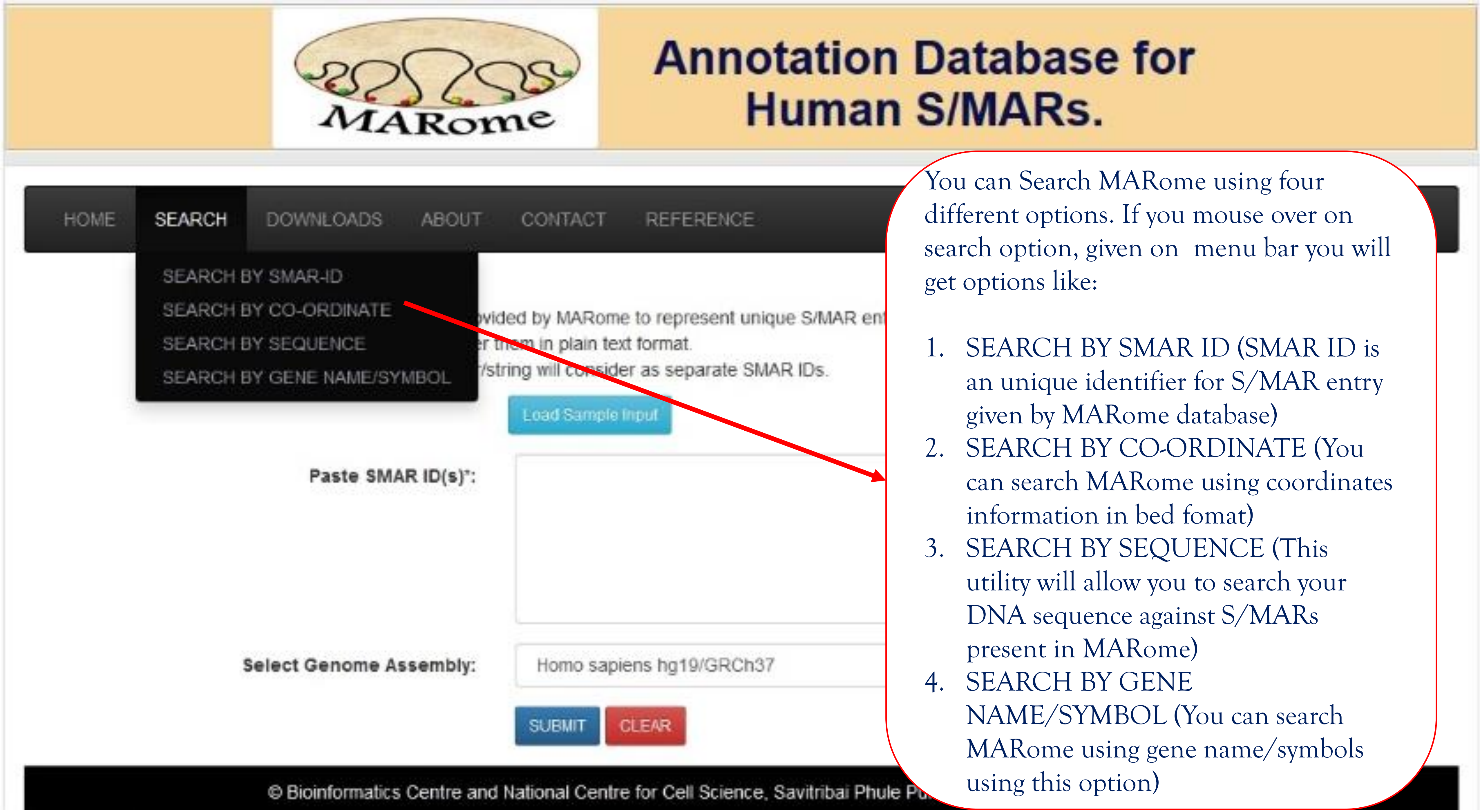
- MARome is a simple web interface designed to access the data obtained in the present study. It is not a prediction tool/server which you could use for S/MAR OR S/MAR like sequence prediction. In case of sequence search utility you can search any DNA sequence against MARome. This utility searches the input sequence using NCBI blastn (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/) program and retain a closet match along with all hits.

- We have integrated few search utilities for easy and complete search, like Coordinate search, Sequence search, text and ID search. You can search your own data against MARome using these utilities.

- You can export your searched results into EXCEL/CSV/PDF/TXT formats.

- Complete dataset including protein wise BED files, complete S/MAR sequences in FASTA format and complete annotation is available for download, you can find the same under download tab available on menu bar.

- MARome holds data for two different assembly versions i.e. hg38 and hg19.



Annotation Database for Human S/MARs.

vided by MARome to represent unique S/MAR ent

r them in plain text format.

/string will consider as separate SMAR IDs.

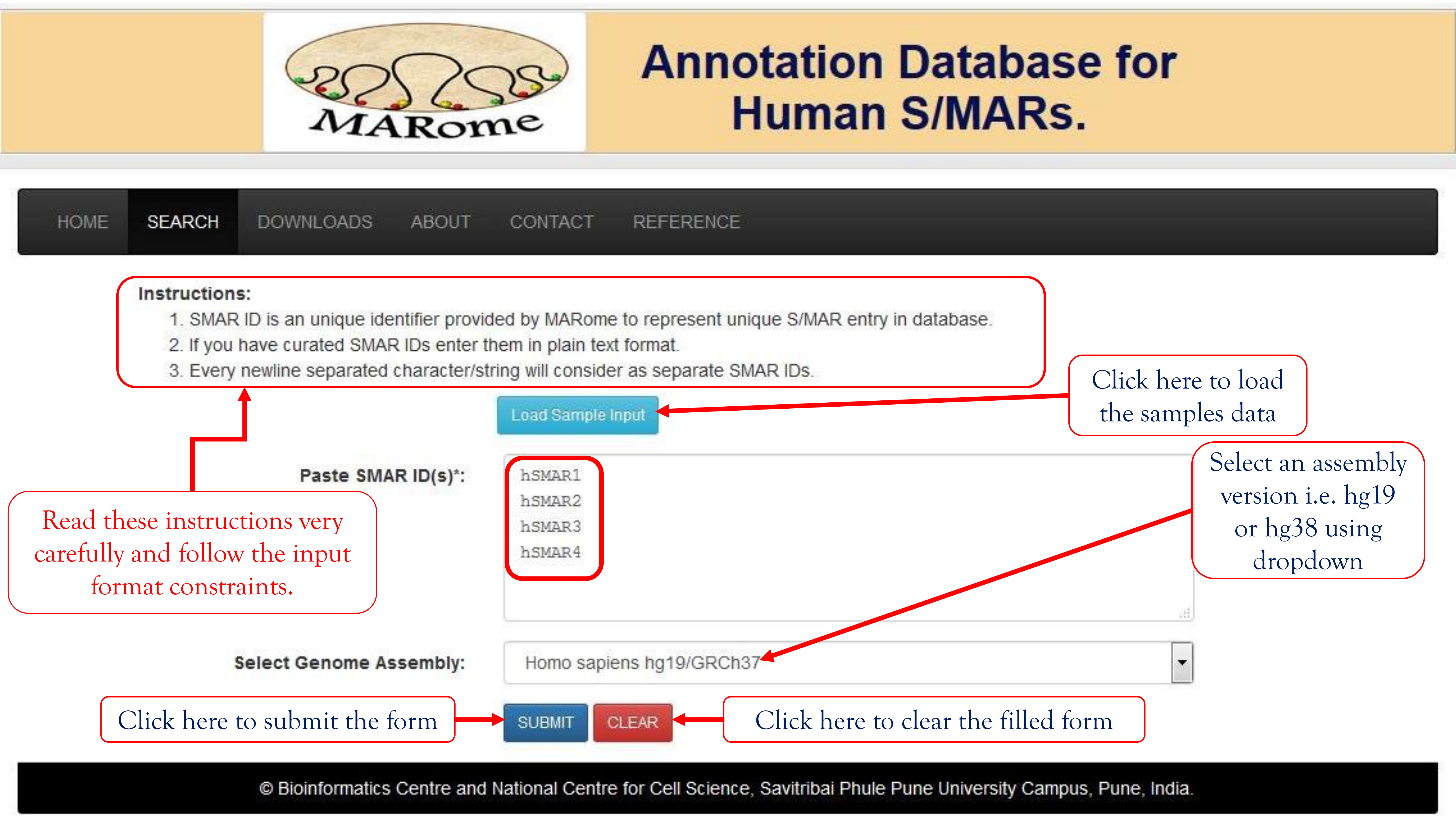
Paste SMAR ID(s)*:

Select Genome Assembly: Homo sapiens hg19/GRCh37

© Bioinformatics Centre and National Centre for Cell Science, Savitribai Phule Pu

You can Search MARome using four different options. If you mouse over on search option, given on menu bar you will get options like:

1. SEARCH BY SMAR ID (SMAR ID is an unique identifier for S/MAR entry given by MARome database)
2. SEARCH BY CO-ORDINATE (You can search MARome using coordinates information in bed fomat)
3. SEARCH BY SEQUENCE (This utility will allow you to search your DNA sequence against S/MARs present in MARome)
4. SEARCH BY GENE NAME/SYMBOL (You can search MARome using gene name/symbols using this option)



Annotation Database for Human S/MARs.

Instructions:

1. SMAR ID is an unique identifier provided by MARome to represent unique S/MAR entry in database.
2. If you have curated SMAR IDs enter them in plain text format.
3. Every newline separated character/string will consider as separate SMAR IDs.

Click here to load the samples data

Paste SMAR ID(s)*:

```
hSMAR1
hSMAR2
hSMAR3
hSMAR4
```

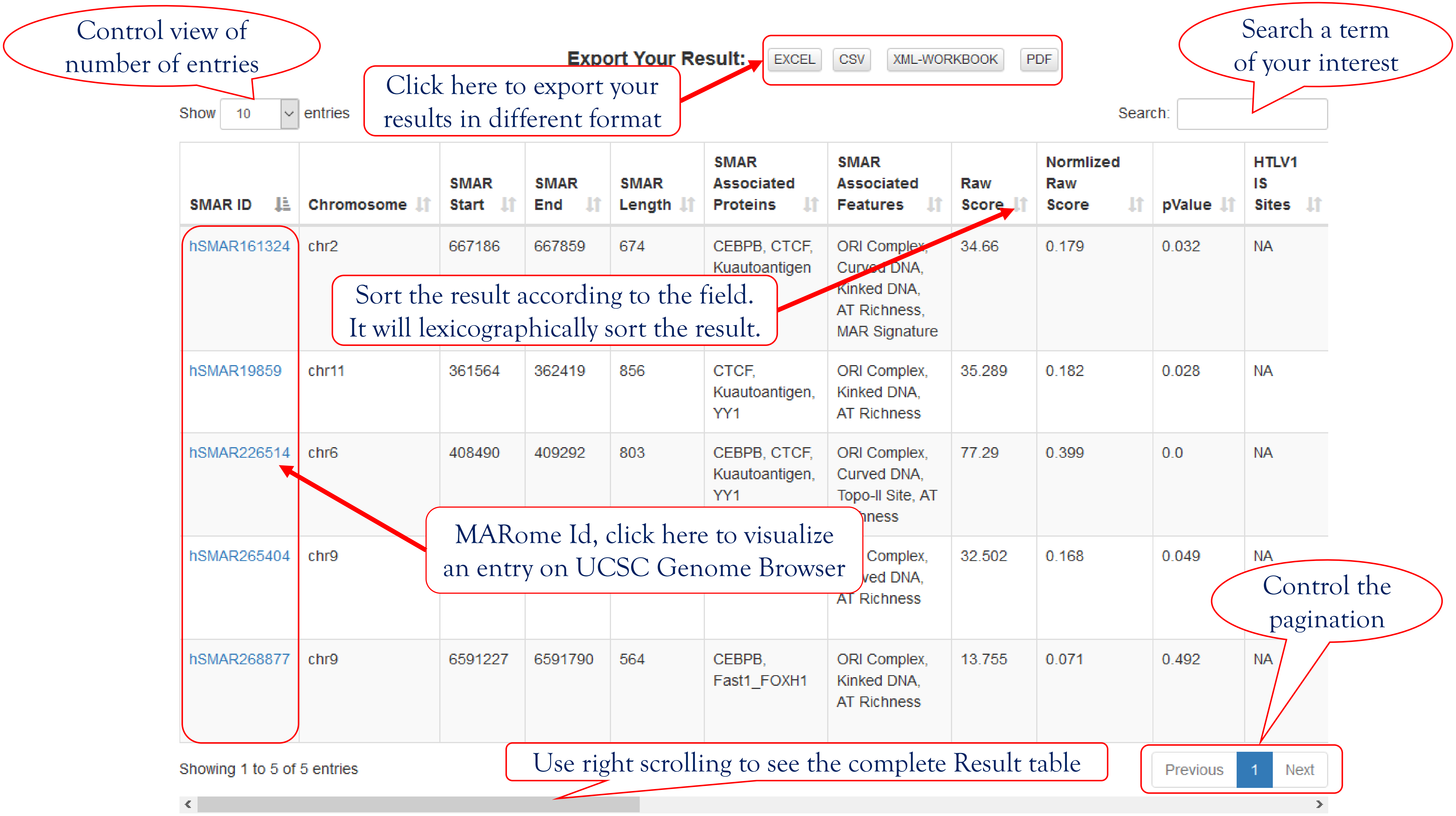
Read these instructions very carefully and follow the input format constraints.

Select an assembly version i.e. hg19 or hg38 using dropdown

Select Genome Assembly: Homo sapiens hg19/GRCh37

Click here to submit the form

Click here to clear the filled form



Control view of number of entries

Click here to export your results in different format

Search a term of your interest

Sort the result according to the field. It will lexicographically sort the result.

MARome Id, click here to visualize an entry on UCSC Genome Browser

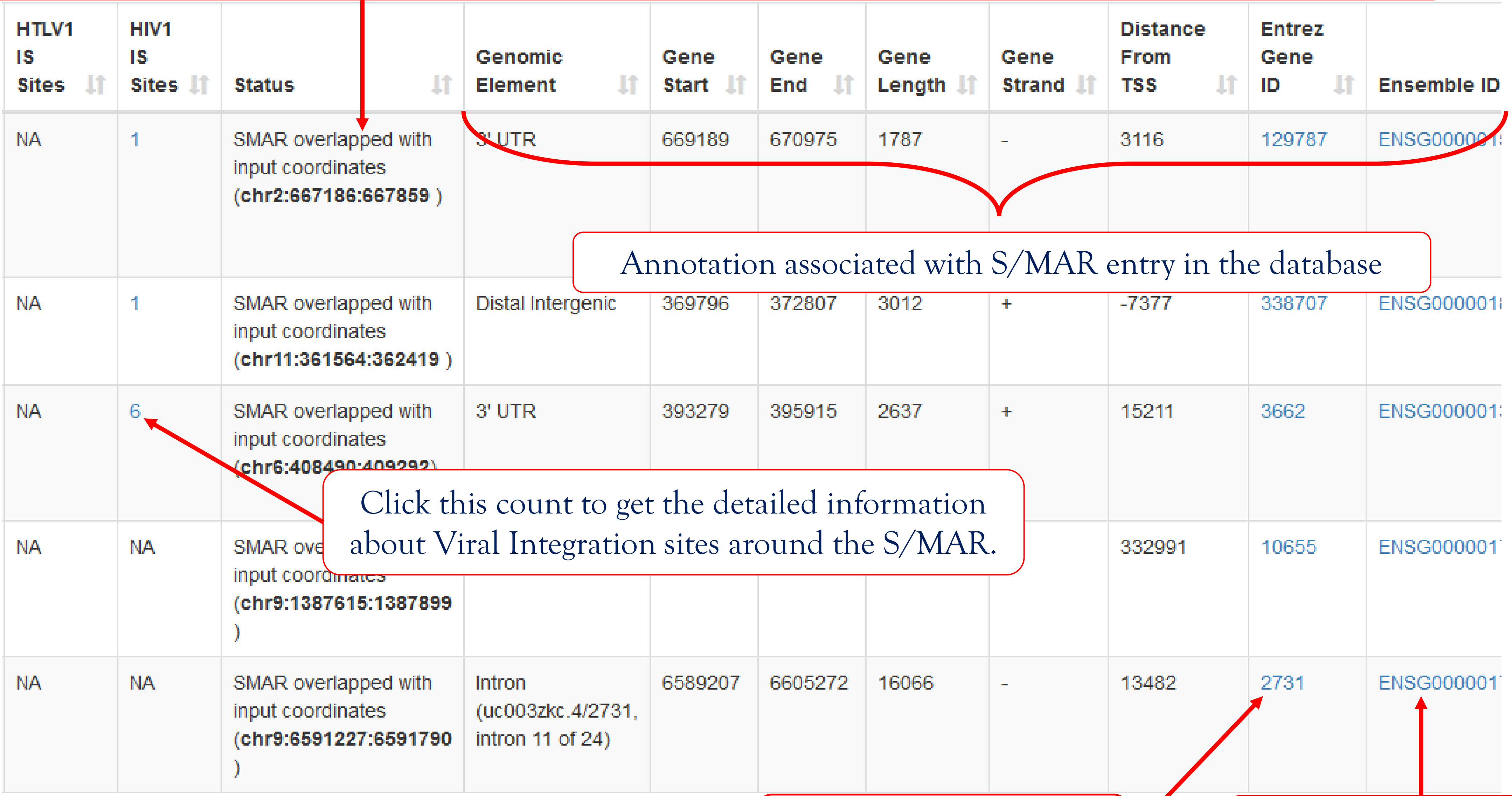
Control the pagination

Use right scrolling to see the complete Result table

SMAR ID	Chromosome	SMAR Start	SMAR End	SMAR Length	SMAR Associated Proteins	SMAR Associated Features	Raw Score	Normlized Raw Score	pValue	HTLV1 IS Sites
hSMAR161324	chr2	667186	667859	674	CEBPB, CTCF, Kuautoantigen	ORI Complex, Curved DNA, Kinked DNA, AT Richness, MAR Signature	34.66	0.179	0.032	NA
hSMAR19859	chr11	361564	362419	856	CTCF, Kuautoantigen, YY1	ORI Complex, Kinked DNA, AT Richness	35.289	0.182	0.028	NA
hSMAR226514	chr6	408490	409292	803	CEBPB, CTCF, Kuautoantigen, YY1	ORI Complex, Curved DNA, Topo-II Site, AT nness	77.29	0.399	0.0	NA
hSMAR265404	chr9					Complex, ved DNA, AT Richness	32.502	0.168	0.049	NA
hSMAR268877	chr9	6591227	6591790	564	CEBPB, Fast1_FOXH1	ORI Complex, Kinked DNA, AT Richness	13.755	0.071	0.492	NA

Showing 1 to 5 of 5 entries

When you search MARome using "search by co-ordinate" utility, this information will display. It will give a clue for the association between provided coordinates and S/MAR present in the database.



HTLV1 IS Sites	HIV1 IS Sites	Status	Genomic Element	Gene Start	Gene End	Gene Length	Gene Strand	Distance From TSS	Entrez Gene ID	Ensemble ID
NA	1	SMAR overlapped with input coordinates (**chr2:667186:667859**)	3' UTR	669189	670975	1787	-	3116	129787	ENSG0000
NA	1	SMAR overlapped with input coordinates (**chr11:361564:362419**)	Distal Intergenic	369796	372807	3012	+	-7377	338707	ENSG000001
NA	6	SMAR overlapped with input coordinates (**chr6:408490:409292**)	3' UTR	393279	395915	2637	+	15211	3662	ENSG000001
NA	NA	SMAR ove input coordinates (**chr9:1387615:1387899**)						332991	10655	ENSG000001
NA	NA	SMAR overlapped with input coordinates (**chr9:6591227:6591790**)	Intron (uc003zkc.4/2731, intron 11 of 24)	6589207	6605272	16066	-	13482	2731	ENSG000001

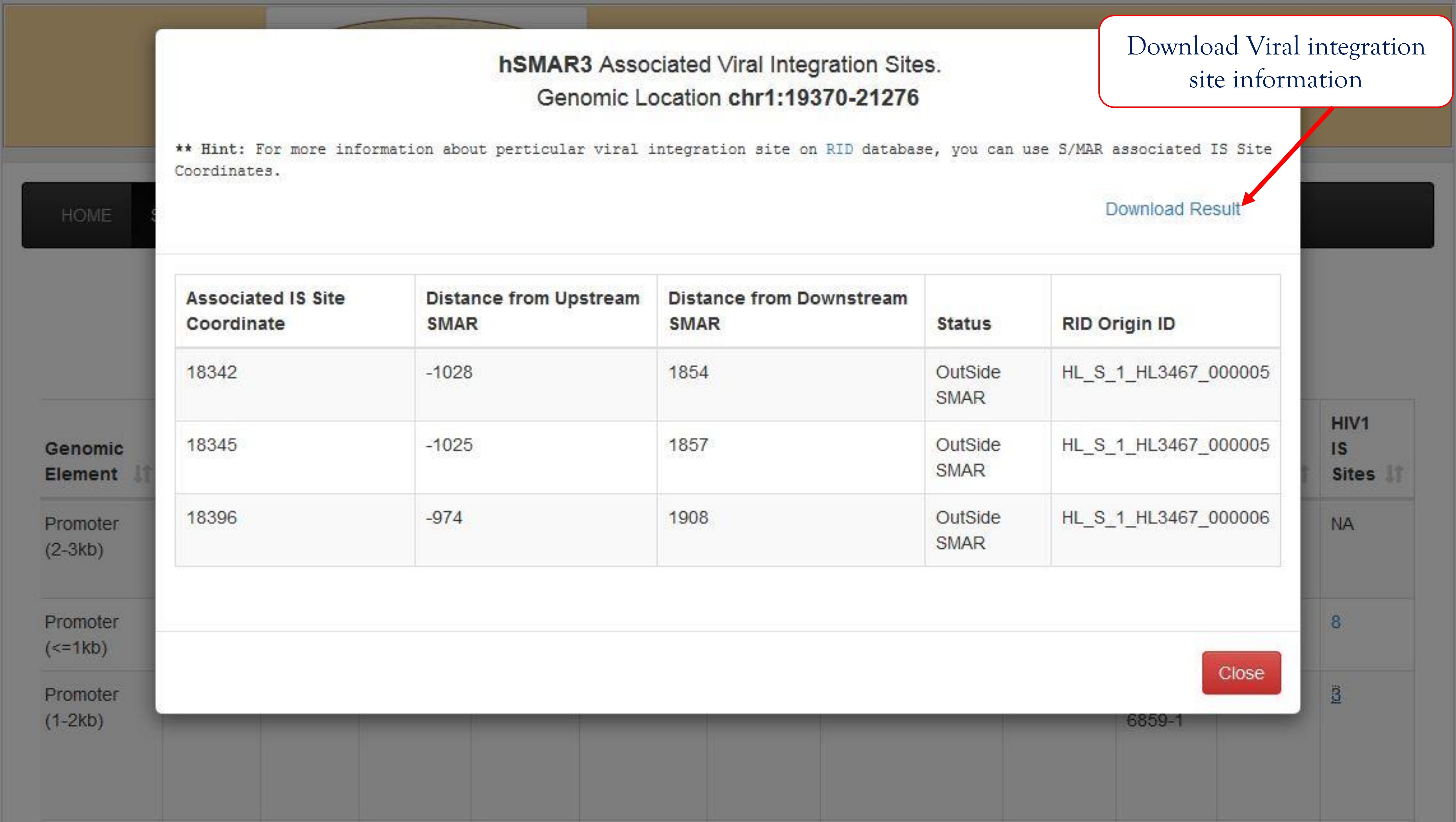
Annotation associated with S/MAR entry in the database

Click this count to get the detailed information about Viral Integration sites around the S/MAR.

Click this id to redirect on NCBI gene database

Click this id to redirect on Ensemble gene database

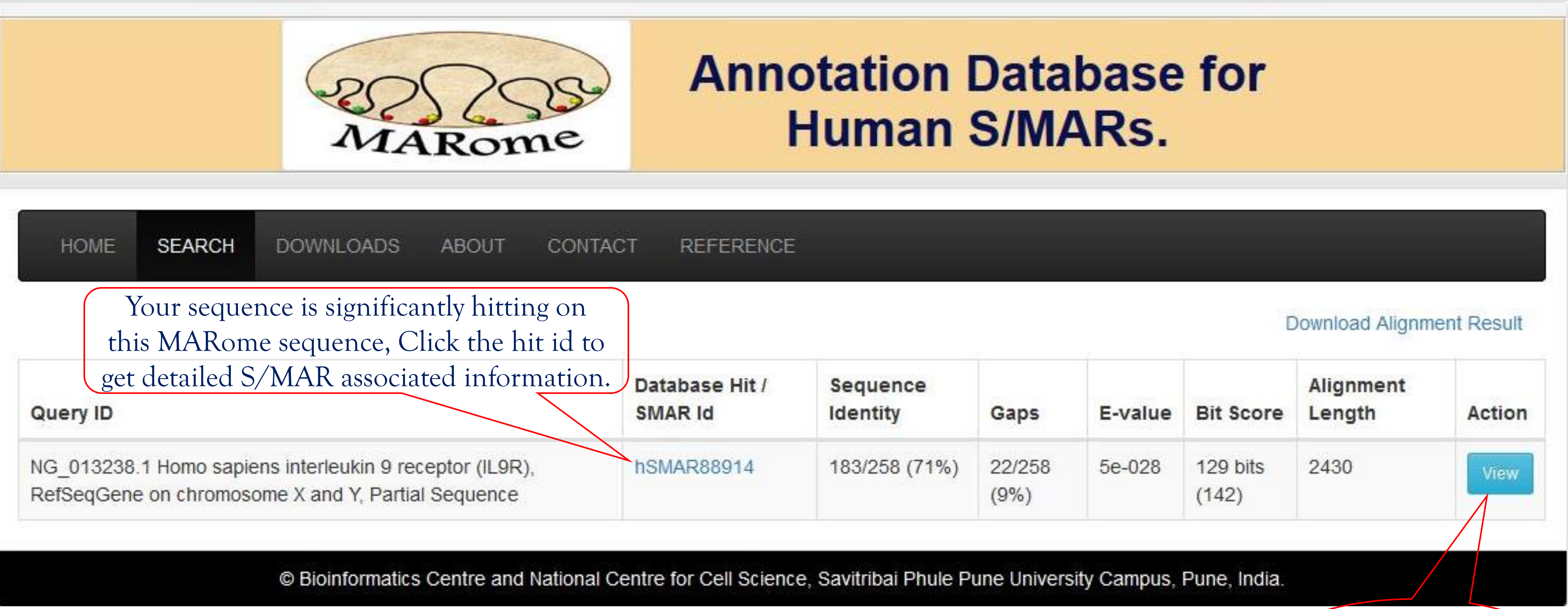
Download Viral integration site information



hSMAR3 Associated Viral Integration Sites.
Genomic Location **chr1:19370-21276**

`** Hint: For more information about perticular viral integration site on RID database, you can use S/MAR associated IS Site Coordinates.`

Associated IS Site Coordinate	Distance from Upstream SMAR	Distance from Downstream SMAR	Status	RID Origin ID
18342	-1028	1854	OutSide SMAR	HL_S_1_HL3467_000005
18345	-1025	1857	OutSide SMAR	HL_S_1_HL3467_000005
18396	-974	1908	OutSide SMAR	HL_S_1_HL3467_000006



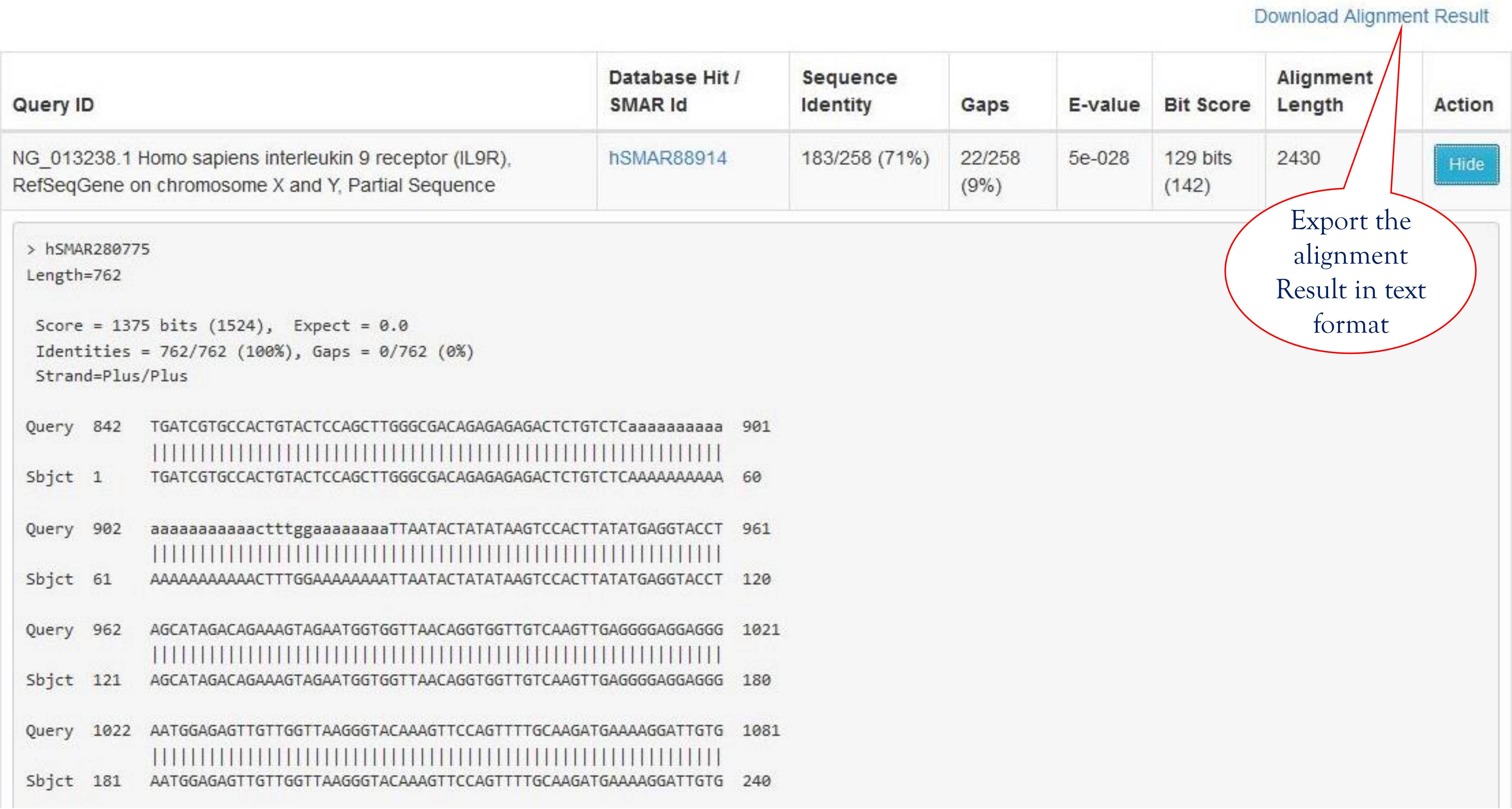
Annotation Database for Human S/MARs.

Query ID	Database Hit / SMAR Id	Sequence Identity	Gaps	E-value	Bit Score	Alignment Length	Action
NG_013238.1 Homo sapiens interleukin 9 receptor (IL9R), RefSeqGene on chromosome X and Y, Partial Sequence	hSMAR88914	183/258 (71%)	22/258 (9%)	5e-028	129 bits (142)	2430	View

© Bioinformatics Centre and National Centre for Cell Science, Savitribai Phule Pune University Campus, Pune, India.

Your sequence is significantly hitting on this MARome sequence, Click the hit id to get detailed S/MAR associated information.

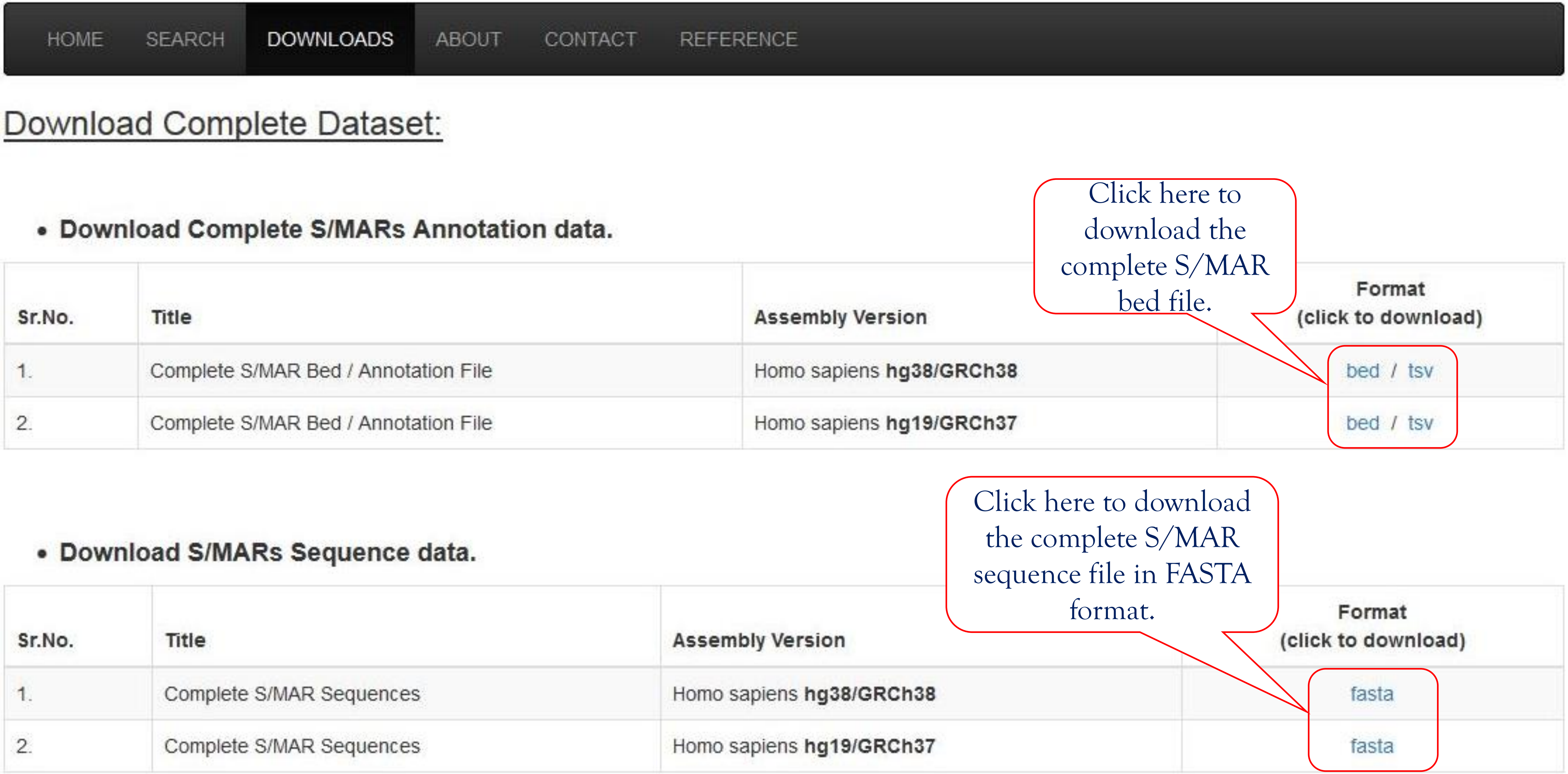
Click here to view detailed blast alignment



Query ID	Database Hit / SMAR Id	Sequence Identity	Gaps	E-value	Bit Score	Alignment Length	Action
NG_013238.1 Homo sapiens interleukin 9 receptor (IL9R), RefSeqGene on chromosome X and Y, Partial Sequence	hSMAR88914	183/258 (71%)	22/258 (9%)	5e-028	129 bits (142)	2430	Hide

```
> hSMAR280775
Length=762

 Score = 1375 bits (1524),  Expect = 0.0
 Identities = 762/762 (100%), Gaps = 0/762 (0%)
 Strand=Plus/Plus

Query  842   TGATCGTGCCACTGTACTCCAGCTTGGGCGACAGAGAGAGACTCTGTCTCaaaaaaaaaa  901
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1     TGATCGTGCCACTGTACTCCAGCTTGGGCGACAGAGAGAGACTCTGTCTCAAAAAAAAAA  60

Query  902   aaaaaaaaaaactttggaaaaaaaaTTAATACTATATAAGTCCACTTATATGAGGTACCT  961
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61    AAAAAAAAAAACTTTGGAAAAAAAATTAATACTATATAAGTCCACTTATATGAGGTACCT  120

Query  962   AGCATAGACAGAAAGTAGAATGGTGGTTAACAGGTGGTTGTCAAGTTGAGGGGAGGAGGG  1021
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121   AGCATAGACAGAAAGTAGAATGGTGGTTAACAGGTGGTTGTCAAGTTGAGGGGAGGAGGG  180

Query  1022  AATGGAGAGTTGTTGGTTAAGGGTACAAAGTTCCAGTTTTGCAAGATGAAAAGGATTGTG  1081
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181   AATGGAGAGTTGTTGGTTAAGGGTACAAAGTTCCAGTTTTGCAAGATGAAAAGGATTGTG  240
```

Export the alignment Result in text format



Download Complete Dataset:

- **Download Complete S/MARs Annotation data.**

Sr.No.	Title	Assembly Version	Format (click to download)
1.	Complete S/MAR Bed / Annotation File	Homo sapiens **hg38/GRCh38**	bed / tsv
2.	Complete S/MAR Bed / Annotation File	Homo sapiens **hg19/GRCh37**	bed / tsv

Click here to download the complete S/MAR bed file.

- **Download S/MARs Sequence data.**

Sr.No.	Title	Assembly Version	Format (click to download)
1.	Complete S/MAR Sequences	Homo sapiens **hg38/GRCh38**	fasta
2.	Complete S/MAR Sequences	Homo sapiens **hg19/GRCh37**	fasta

Click here to download the complete S/MAR sequence file in FASTA format.

Scoring Function

We ascribed the feature weight (*Fw*) to individual S/MAR associated feature as shown equation **1**. For example, OriC feature is present 271541 times in entire dataset (including multiple occurrence in same sequence) then its weight will be 0.959. Similarly, Topoisomerase-II sites that occurs 9953 times in dataset will carry weight of 0.035.

$$Fw = \frac{Individual\ S/MAR\ Feature\ Abundance}{Total\ Number\ Of\ SMARs\ in\ dataset} \qquad \mathbf{1}$$

Then for every SMAR entry and every SMAR associated feature we calculated the feature score (*Fs*) as show in equation **2**. To calculate *Fs*, we considered feature count per kilobase (*Kb*) of the S/MAR length and multiplied it with weight of a feature in consideration (calculated in equation **1**).

$$Fs = \left(\frac{Feature\ Count}{Length\ of\ the\ S/MAR\ (Kb)}\right) \times Fw \qquad \mathbf{2}$$

Such distinct calculated *Fs* were then summed up and multiplied by the number of S/MARBP concomitant with the S/MAR entry. This S/MAR score is called as *Raw Score* and denoted by **‘S’** as shown in equation **3**.

$$S = \sum_{i=1}^{n} Fs_i \times Pc \qquad \mathbf{3}$$

$$(where\ n = S/MAR\ features\ present\ in\ an\ S/MAR\ entry\ ; Pc = Number\ of\ S/MARBPs)$$

Then the obtained raw score was subjected to Min-Max Normalization method using equation **4** to scale it between 0 and 1. Such normalized score is denoted by N.

$$\mathrm{N} = \frac{S - \min(S)}{\max(S) - \min(S)} \qquad \mathbf{4}$$

To get the normal distribution, the individual raw scores were then transformed into *z-score* using equation **5**. The *p-value* was then calculated using the *z-score*. Database users can use the *Min-Max Normalized score* for filtering quality S/MARs over others. they can also take help of *p-value*.

$$z = \frac{S - \mu}{\sigma} \quad (Where\ S = raw\ score\ ;\ \mu = mean;\ \sigma = standard\ deviation) \qquad \mathbf{5}$$