UCSC Table Browser

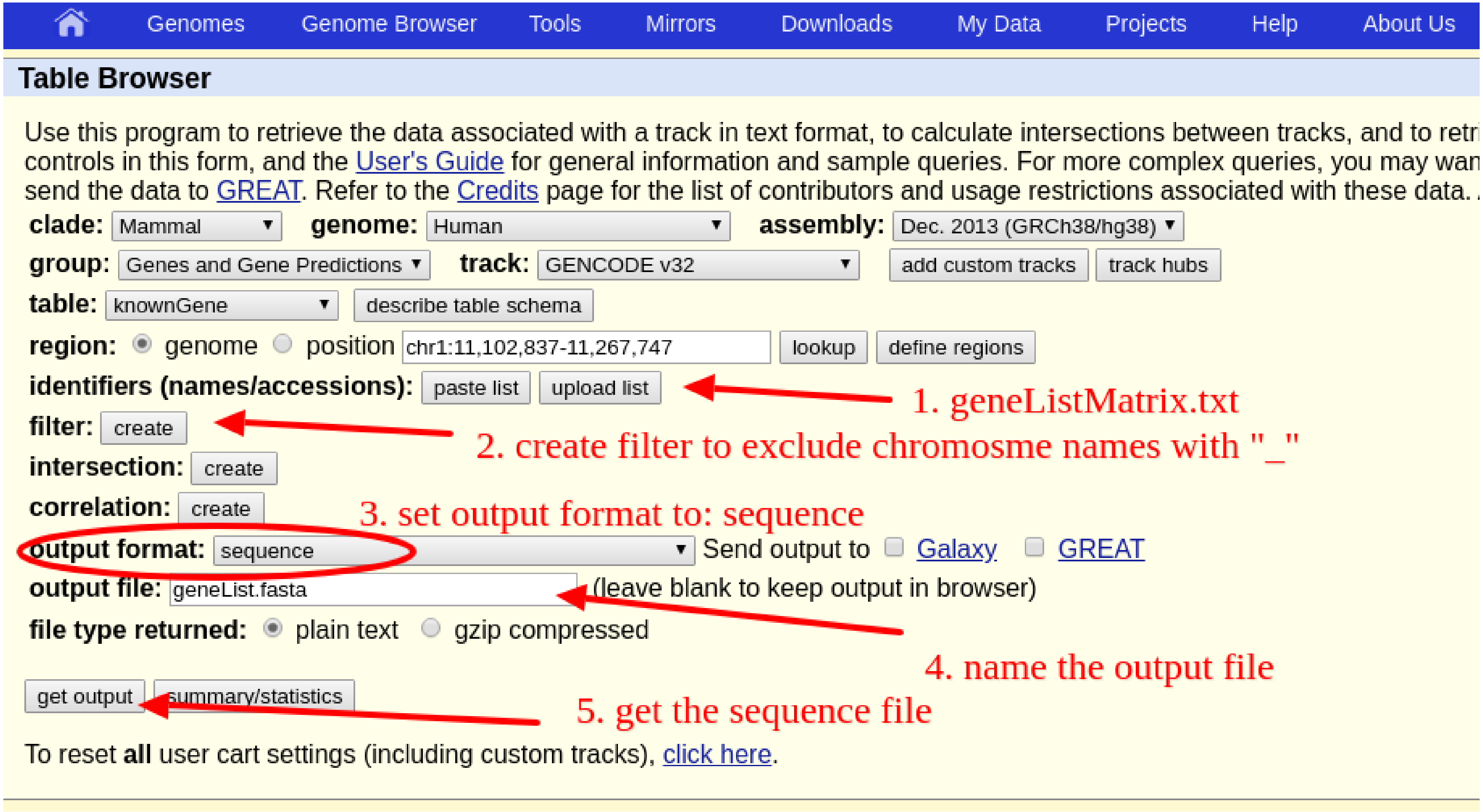


Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retri
controls in this form, and the User's Guide for general information and sample queries. For more complex queries, you may wan
send the data to GREAT. Refer to the Credits page for the list of contributors and usage restrictions associated with these data. .

clade: Mammal ▾ **genome:** Human ▾ **assembly:** Dec. 2013 (GRCh38/hg38) ▾

group: Genes and Gene Predictions ▾ **track:** GENCODE v32 ▾ add custom tracks | track hubs

table: knownGene ▾ describe table schema

region: ● genome ○ position chr1:11,102,837-11,267,747 lookup | define regions

identifiers (names/accessions): paste list | upload list

filter: create

intersection: create

correlation: create

output format: sequence ▾ Send output to ☐ Galaxy ☐ GREAT

output file: geneList.fasta (leave blank to keep output in browser)

file type returned: ● plain text ○ gzip compressed

get output | summary/statistics

To reset **all** user cart settings (including custom tracks), click here.

1. geneListMatrix.txt

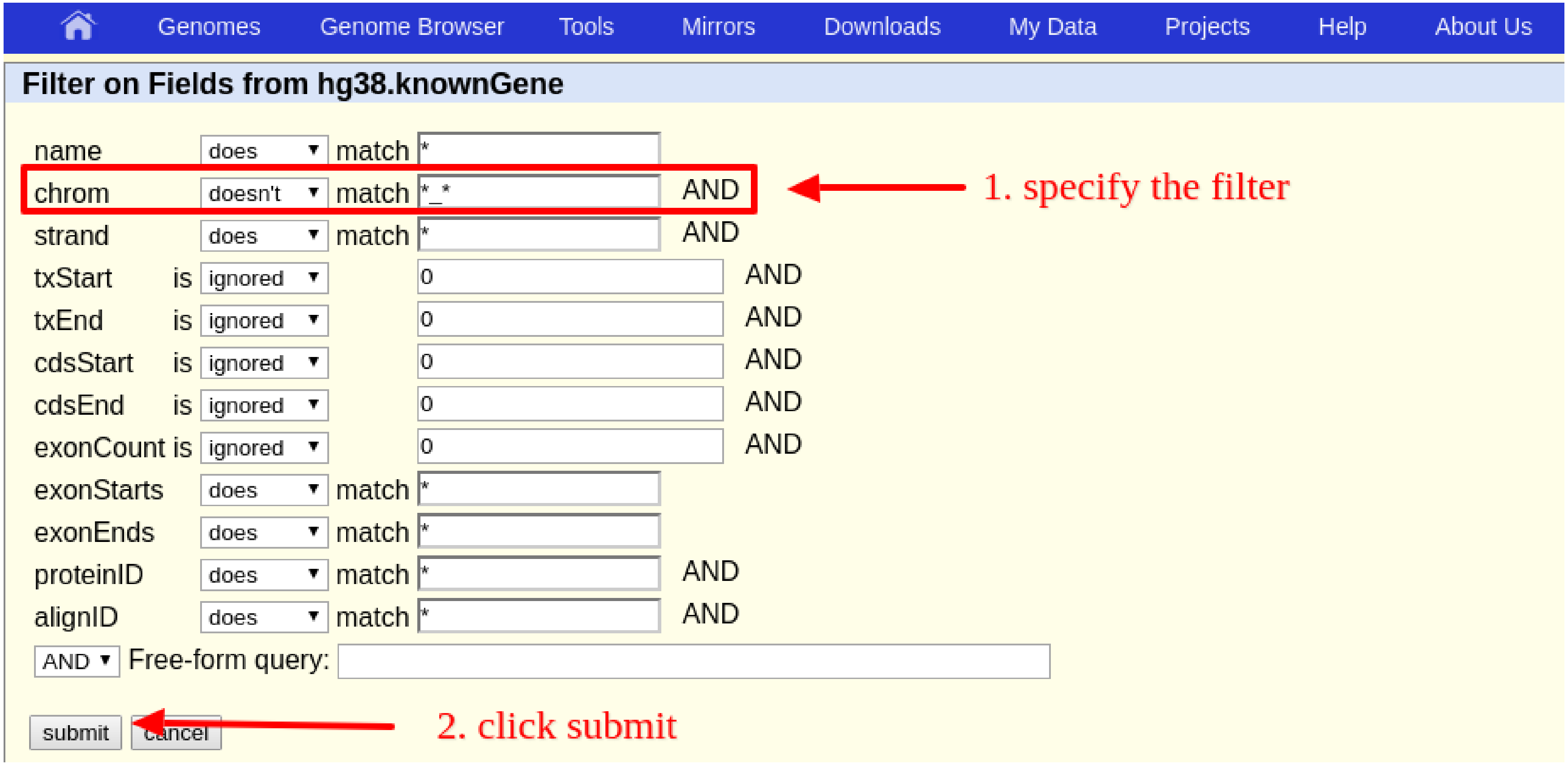
2. create filter to exclude chromosme names with "_"

3. set output format to: sequence

4. name the output file

5. get the sequence file

UCSC Table Browser: set filter



Filter on Fields from hg38.knownGene

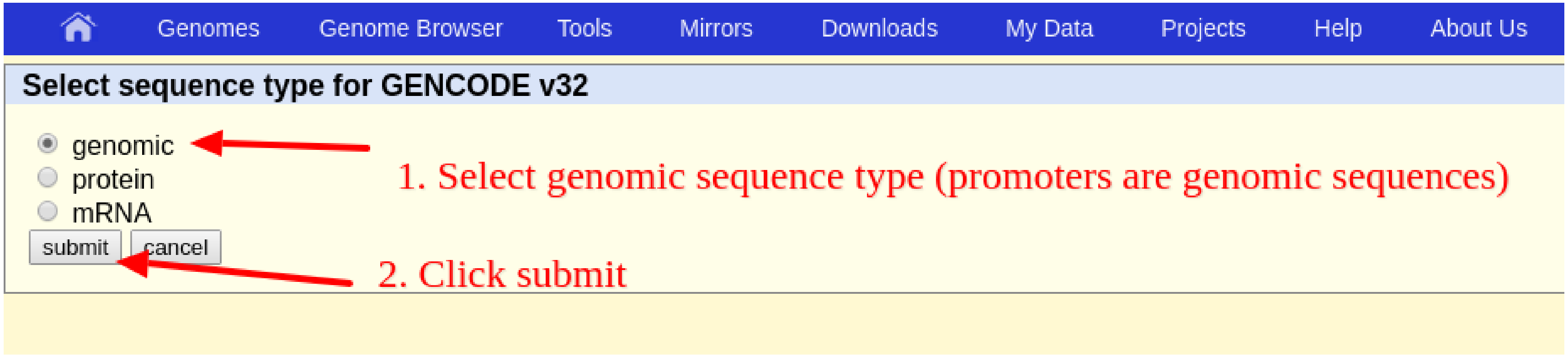
name	does	match	*	
chrom	doesn't	match	*_*	AND
strand	does	match	*	AND
txStart is	ignored		0	AND
txEnd is	ignored		0	AND
cdsStart is	ignored		0	AND
cdsEnd is	ignored		0	AND
exonCount is	ignored		0	AND
exonStarts	does	match	*	
exonEnds	does	match	*	
proteinID	does	match	*	AND
alignID	does	match	*	AND

AND Free-form query:

1. specify the filter

2. click submit

UCSC Table Browser: select sequence type



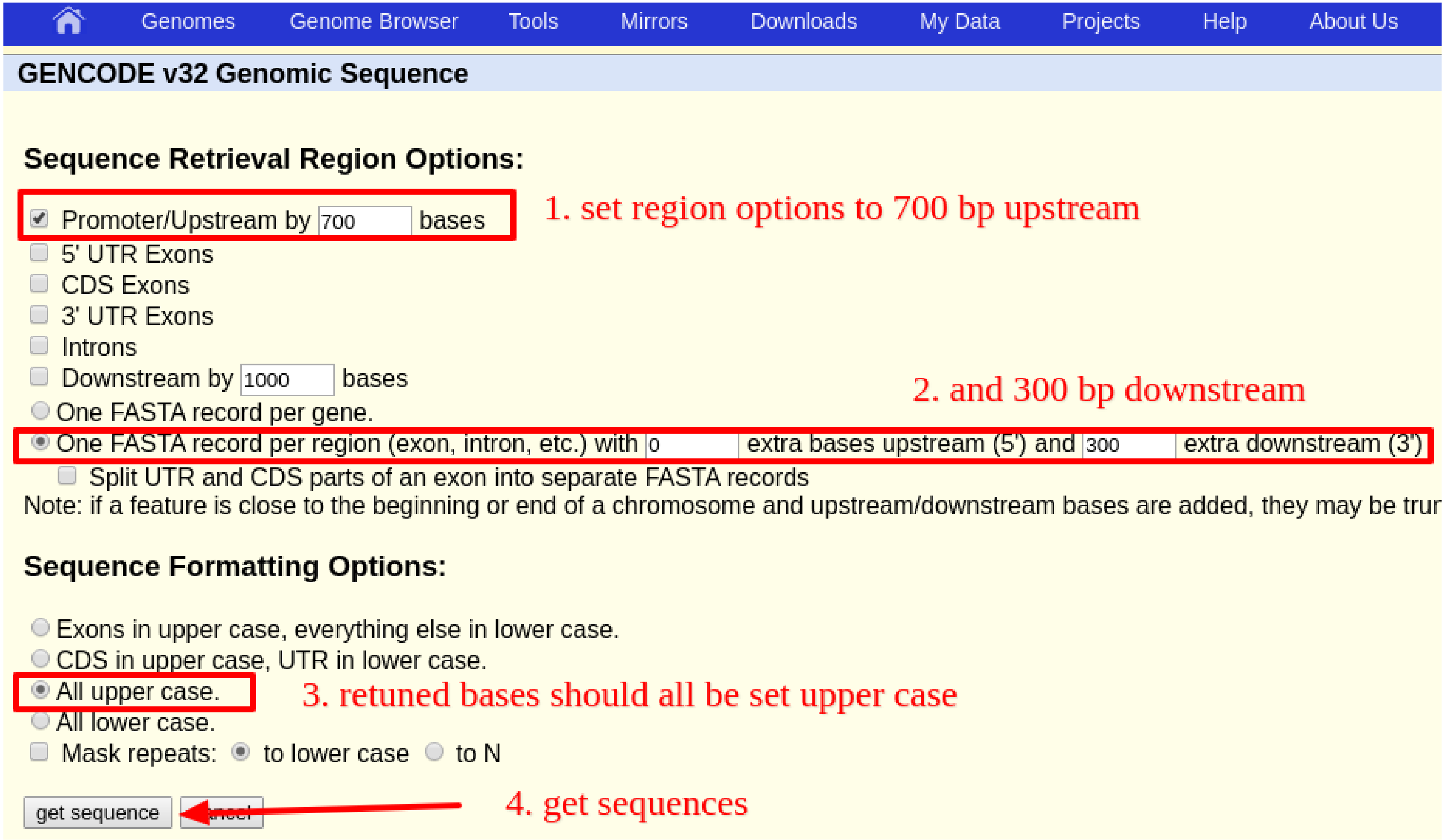
Select sequence type for GENCODE v32

● genomic
○ protein
○ mRNA

1. Select genomic sequence type (promoters are genomic sequences)

2. Click submit

UCSC Table Browser: set region options



GENCODE v32 Genomic Sequence

Sequence Retrieval Region Options:

☑ Promoter/Upstream by 700 bases
☐ 5' UTR Exons
☐ CDS Exons
☐ 3' UTR Exons
☐ Introns
☐ Downstream by 1000 bases
○ One FASTA record per gene.
● One FASTA record per region (exon, intron, etc.) with 0 extra bases upstream (5') and 300 extra downstream (3')
☐ Split UTR and CDS parts of an exon into separate FASTA records

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be trun

Sequence Formatting Options:

○ Exons in upper case, everything else in lower case.
○ CDS in upper case, UTR in lower case.
● All upper case.
○ All lower case.
☐ Mask repeats: ● to lower case ○ to N

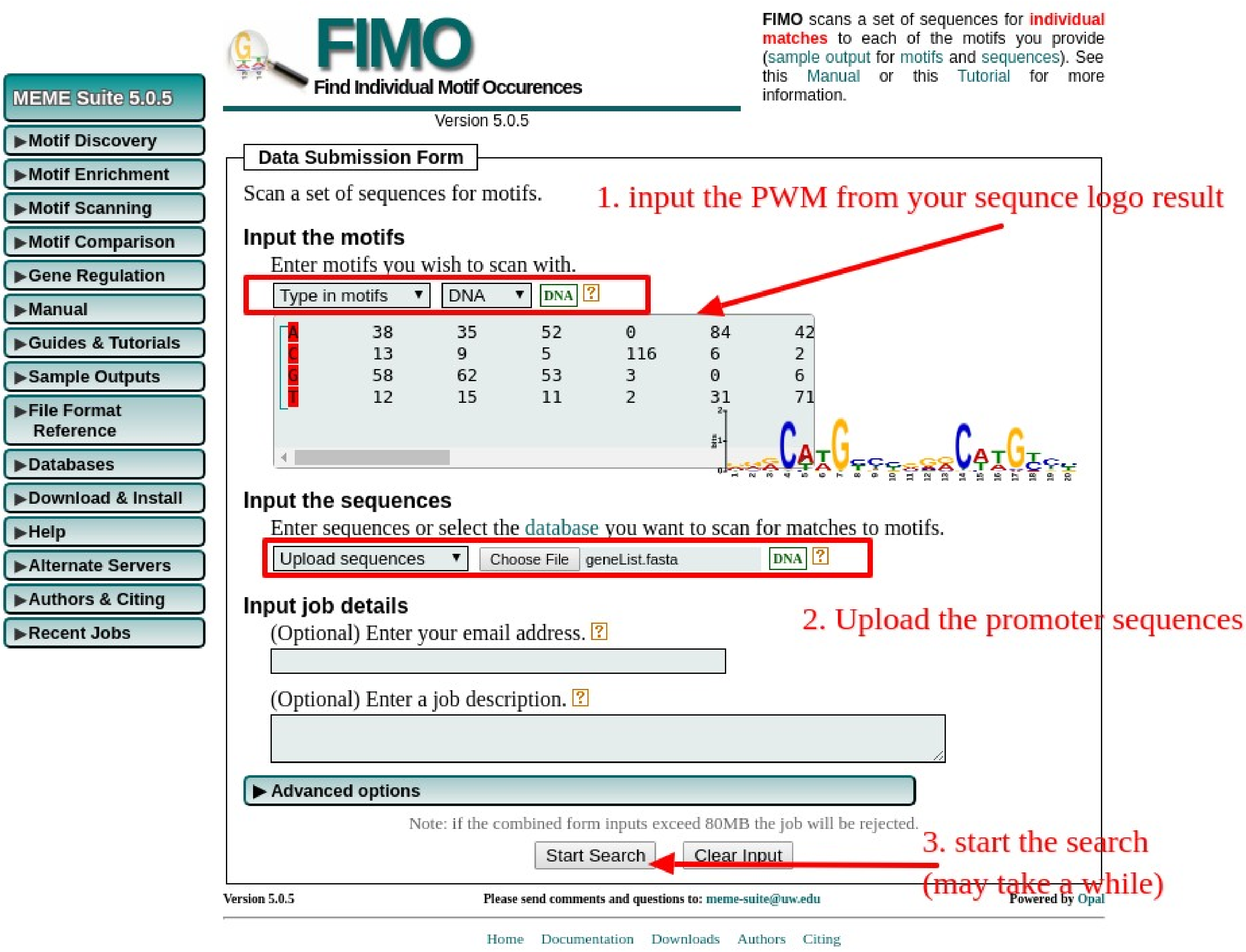
1. set region options to 700 bp upstream

2. and 300 bp downstream

3. retuned bases should all be set upper case

4. get sequences

FIMO: scan promoter sequences for TFBS



Version 5.0.5

FIMO scans a set of sequences for **individual matches** to each of the motifs you provide (sample output for motifs and sequences). See this Manual or this Tutorial for more information.

Data Submission Form

Scan a set of sequences for motifs.

Input the motifs

Enter motifs you wish to scan with.

Type in motifs ▼ DNA ▼ DNA ?

```
A     38    35    52    0     84    42
C     13    9     5     116   6     2
G     58    62    53    3     0     6
T     12    15    11    2     31    71
```

Input the sequences

Enter sequences or select the database you want to scan for matches to motifs.

Upload sequences ▼ Choose File geneList.fasta DNA ?

Input job details

(Optional) Enter your email address. ?

(Optional) Enter a job description. ?

▶ Advanced options

Note: if the combined form inputs exceed 80MB the job will be rejected.

Start Search Clear Input

Version 5.0.5 Please send comments and questions to: meme-suite@uw.edu Powered by Opal

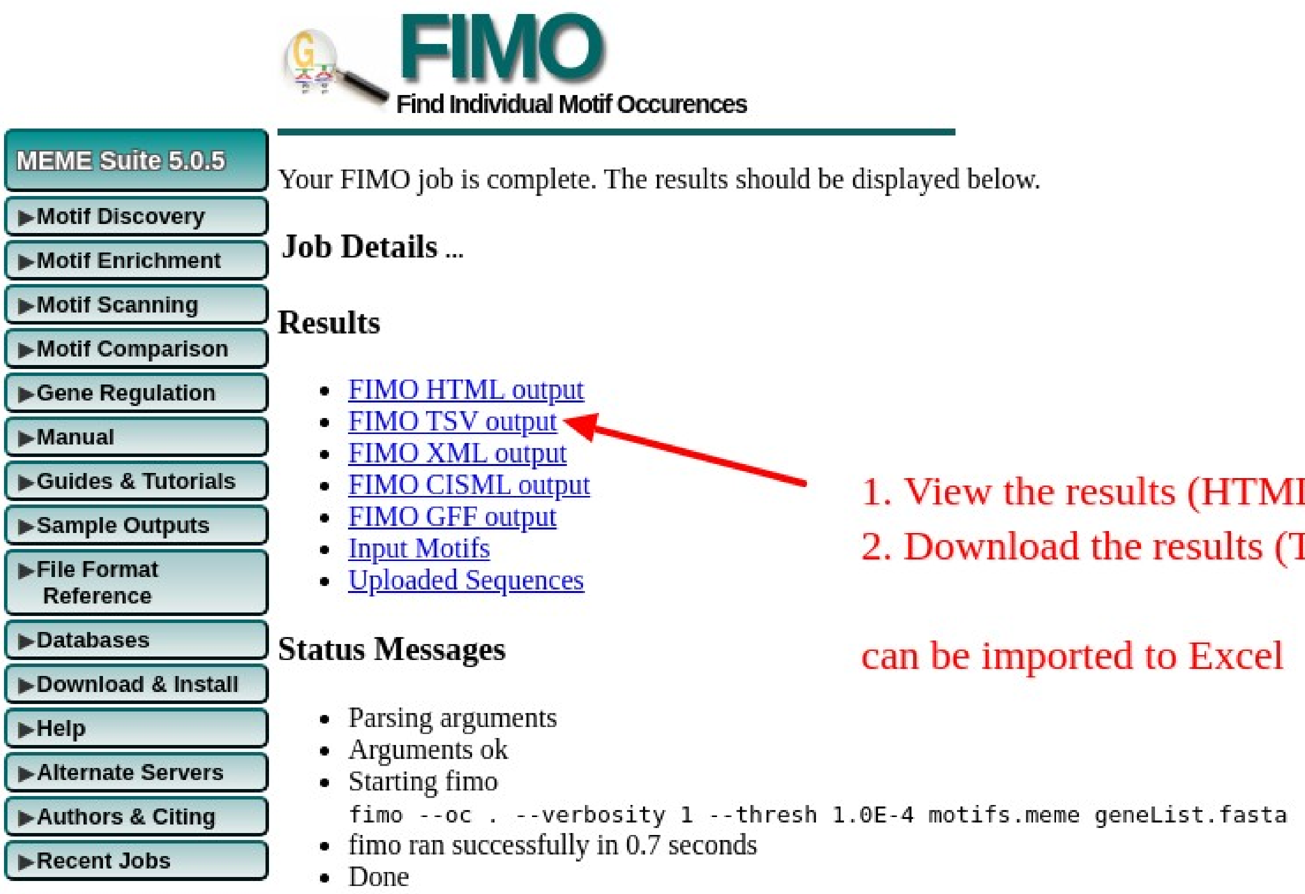
Home Documentation Downloads Authors Citing

1. input the PWM from your sequnce logo result

2. Upload the promoter sequences

3. start the search (may take a while)

FIMO: scan promoter sequences for TFBS results



Your FIMO job is complete. The results should be displayed below.

Job Details ...

Results

- FIMO HTML output
- FIMO TSV output
- FIMO XML output
- FIMO CISML output
- FIMO GFF output
- Input Motifs
- Uploaded Sequences

Status Messages

- Parsing arguments
- Arguments ok
- Starting fimo

```
fimo --oc . --verbosity 1 --thresh 1.0E-4 motifs.meme geneList.fasta
```

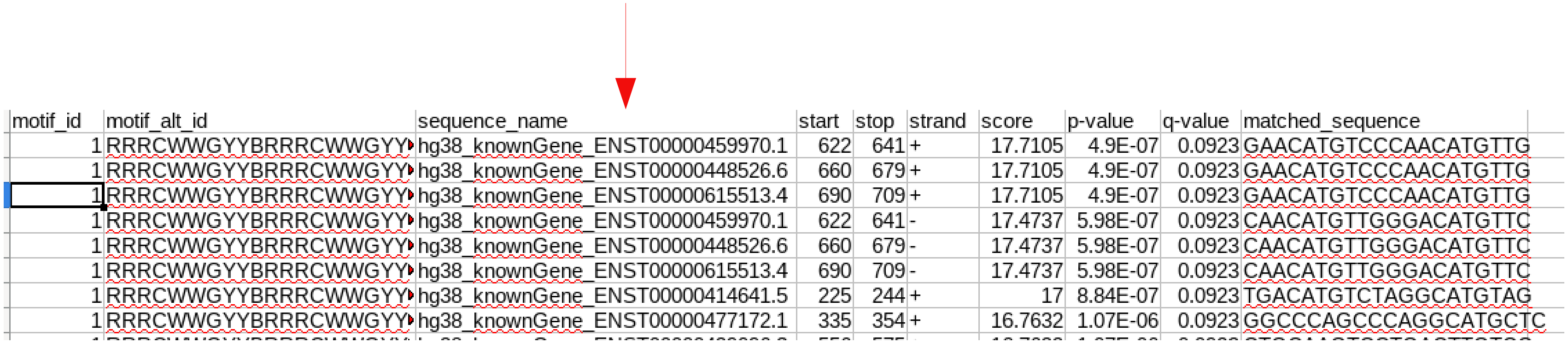
- fimo ran successfully in 0.7 seconds
- Done

1. View the results (HTML)
2. Download the results (TSV)

can be imported to Excel

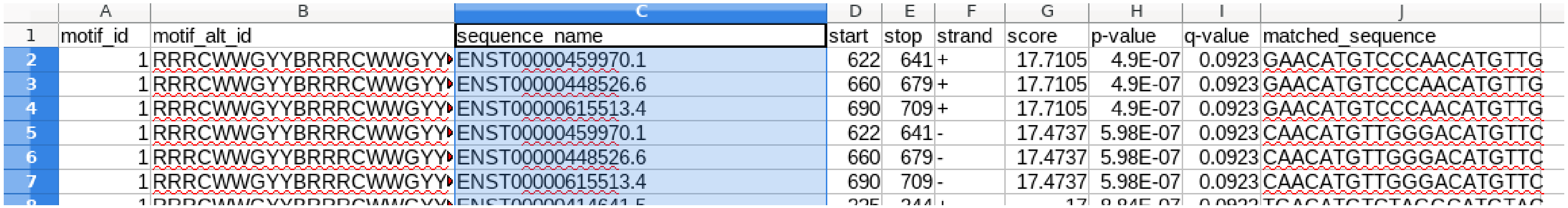
Ensembl transcript ID (ENST...) to HGNC symbol (gene symbol) transformation

The FIMO result table shows the Ensembl transcript start for which a P53 TFBS was found, we need to translate these IDs into gene names in order to know the number of genes with a TFBS (one gene can have multiple transcripts)



motif_id	motif_alt_id	sequence_name	start	stop	strand	score	p-value	q-value	matched_sequence
1	RRRCWWGYYBRRRCWWGYY	hg38_knownGene_ENST00000459970.1	622	641	+	17.7105	4.9E-07	0.0923	GAACATGTCCCAACATGTTG
1	RRRCWWGYYBRRRCWWGYY	hg38_knownGene_ENST00000448526.6	660	679	+	17.7105	4.9E-07	0.0923	GAACATGTCCCAACATGTTG
1	RRRCWWGYYBRRRCWWGYY	hg38_knownGene_ENST00000615513.4	690	709	+	17.7105	4.9E-07	0.0923	GAACATGTCCCAACATGTTG
1	RRRCWWGYYBRRRCWWGYY	hg38_knownGene_ENST00000459970.1	622	641	-	17.4737	5.98E-07	0.0923	CAACATGTTGGGACATGTTC
1	RRRCWWGYYBRRRCWWGYY	hg38_knownGene_ENST00000448526.6	660	679	-	17.4737	5.98E-07	0.0923	CAACATGTTGGGACATGTTC
1	RRRCWWGYYBRRRCWWGYY	hg38_knownGene_ENST00000615513.4	690	709	-	17.4737	5.98E-07	0.0923	CAACATGTTGGGACATGTTC
1	RRRCWWGYYBRRRCWWGYY	hg38_knownGene_ENST00000414641.5	225	244	+	17	8.84E-07	0.0923	TGACATGTCTAGGCATGTAG
1	RRRCWWGYYBRRRCWWGYY	hg38_knownGene_ENST00000477172.1	335	354	+	16.7632	1.07E-06	0.0923	GGCCCAGCCCAGGCATGCTC

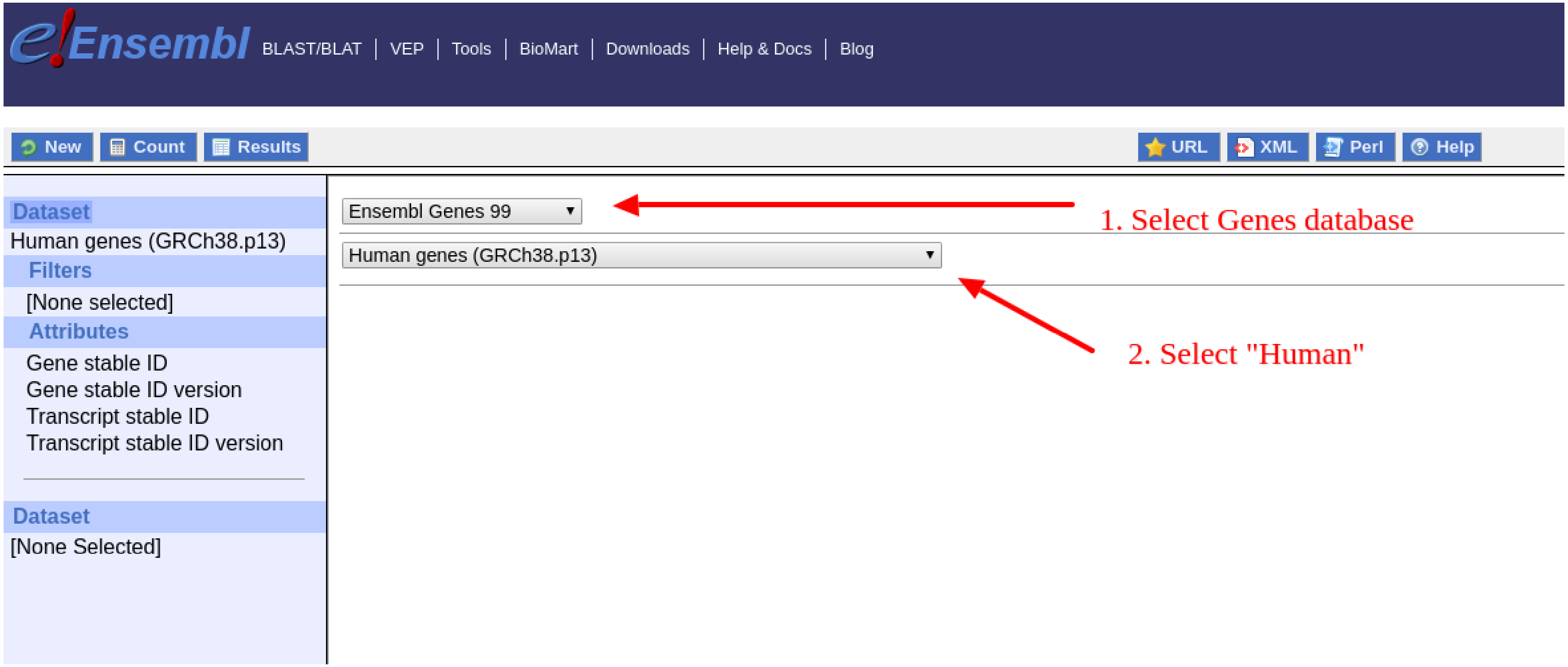
1. remove (seach/replace) the “hg38_knownGene_” part from the sequence names
2. copy all Ensembl transcript IDs (ENST…) into the clipboard (mark column, Strg+c)



	A	B	C	D	E	F	G	H	I	J
1	motif_id	motif_alt_id	sequence_name	start	stop	strand	score	p-value	q-value	matched_sequence
2	1	RRRCWWGYYBRRRCWWGYY	ENST00000459970.1	622	641	+	17.7105	4.9E-07	0.0923	GAACATGTCCCAACATGTTG
3	1	RRRCWWGYYBRRRCWWGYY	ENST00000448526.6	660	679	+	17.7105	4.9E-07	0.0923	GAACATGTCCCAACATGTTG
4	1	RRRCWWGYYBRRRCWWGYY	ENST00000615513.4	690	709	+	17.7105	4.9E-07	0.0923	GAACATGTCCCAACATGTTG
5	1	RRRCWWGYYBRRRCWWGYY	ENST00000459970.1	622	641	-	17.4737	5.98E-07	0.0923	CAACATGTTGGGACATGTTC
6	1	RRRCWWGYYBRRRCWWGYY	ENST00000448526.6	660	679	-	17.4737	5.98E-07	0.0923	CAACATGTTGGGACATGTTC
7	1	RRRCWWGYYBRRRCWWGYY	ENST00000615513.4	690	709	-	17.4737	5.98E-07	0.0923	CAACATGTTGGGACATGTTC

3. navigate to the Ensembl Biomart website

Ensembl transcript ID (ENST...) to HGNC symbol (gene symbol) transformation



Human genes (GRCh38.p13)

[None selected]

Gene stable ID
Gene stable ID version
Transcript stable ID
Transcript stable ID version

[None Selected]

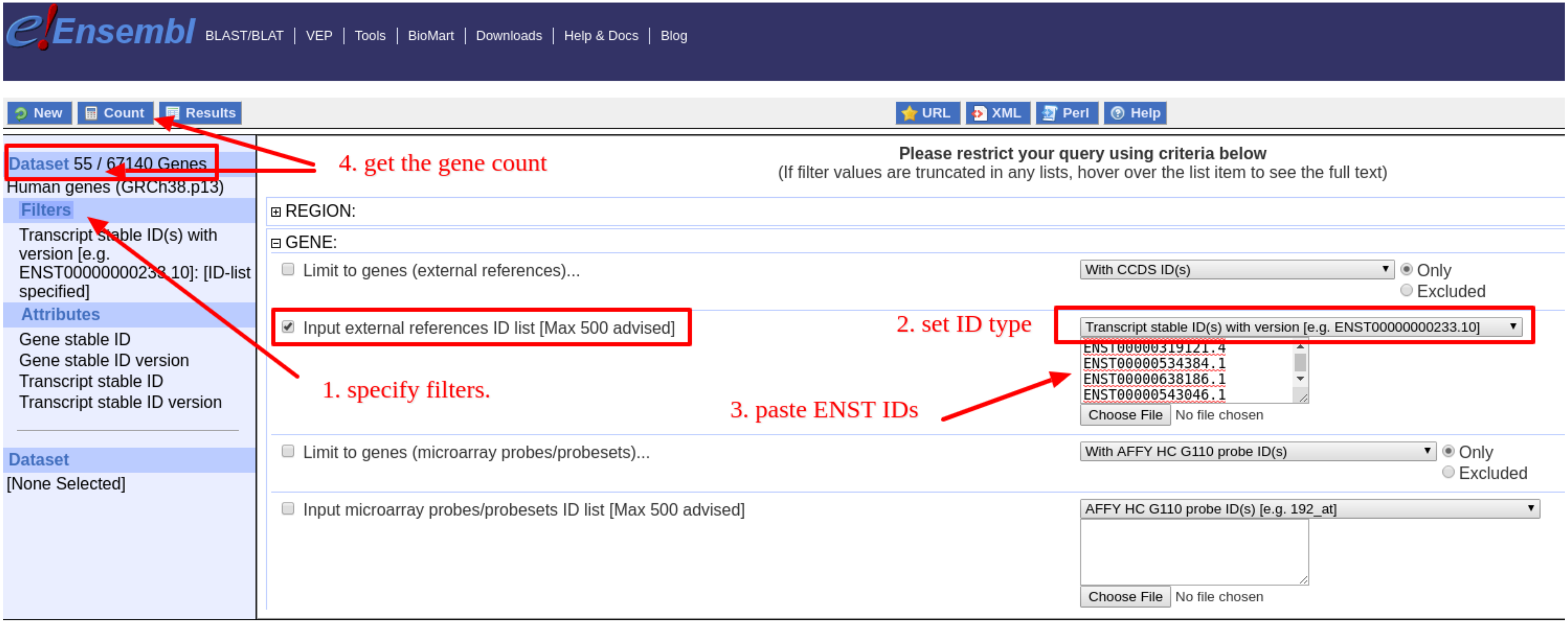
Ensembl Genes 99 ▾

Human genes (GRCh38.p13) ▾

1. Select Genes database

2. Select "Human"

Ensembl transcript ID (ENST...) to HGNC symbol (gene symbol) transformation



4. get the gene count

1. specify filters.

2. set ID type

3. paste ENST IDs

Dataset 55 / 67140 Genes
Human genes (GRCh38.p13)

Filters
Transcript stable ID(s) with version [e.g. ENST00000000233.10]: [ID-list specified]

Attributes
Gene stable ID
Gene stable ID version
Transcript stable ID
Transcript stable ID version

Dataset
[None Selected]

Please restrict your query using criteria below
(If filter values are truncated in any lists, hover over the list item to see the full text)

⊞ REGION:

⊟ GENE:

☐ Limit to genes (external references)... — With CCDS ID(s) — ● Only ○ Excluded

☑ Input external references ID list [Max 500 advised] — Transcript stable ID(s) with version [e.g. ENST00000000233.10]

```
ENST00000319121.4
ENST00000534384.1
ENST00000638186.1
ENST00000543046.1
```

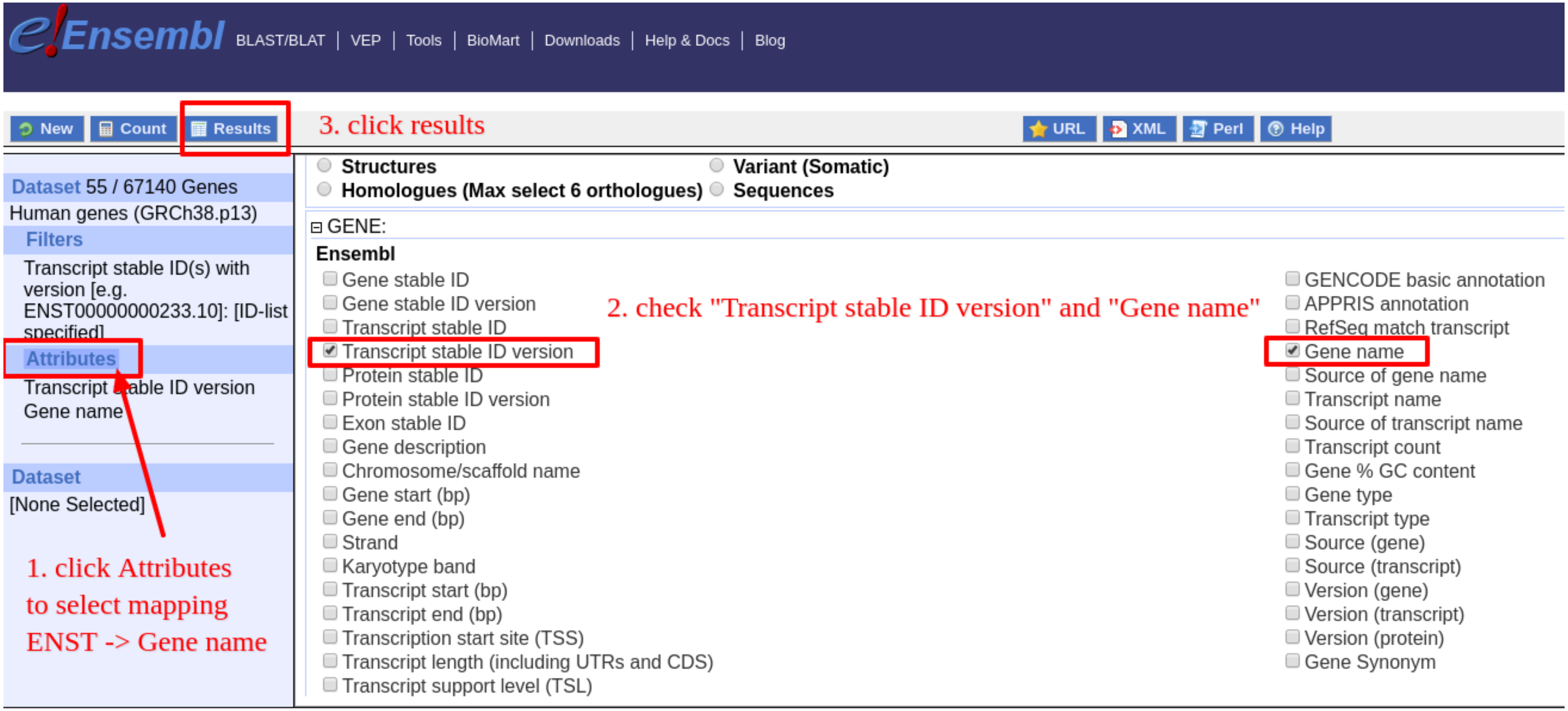
Choose File No file chosen

☐ Limit to genes (microarray probes/probesets)... — With AFFY HC G110 probe ID(s) — ● Only ○ Excluded

☐ Input microarray probes/probesets ID list [Max 500 advised] — AFFY HC G110 probe ID(s) [e.g. 192_at]

Choose File No file chosen

Ensembl transcript ID (ENST...) to HGNC symbol (gene symbol) transformation



3. click results

Dataset 55 / 67140 Genes
Human genes (GRCh38.p13)

Filters
Transcript stable ID(s) with version [e.g. ENST00000000233.10]: [ID-list specified]

Attributes
Transcript stable ID version
Gene name

Dataset
[None Selected]

1. click Attributes to select mapping ENST -> Gene name

○ **Structures**
○ **Homologues (Max select 6 orthologues)**
○ **Variant (Somatic)**
○ **Sequences**

⊟ GENE:

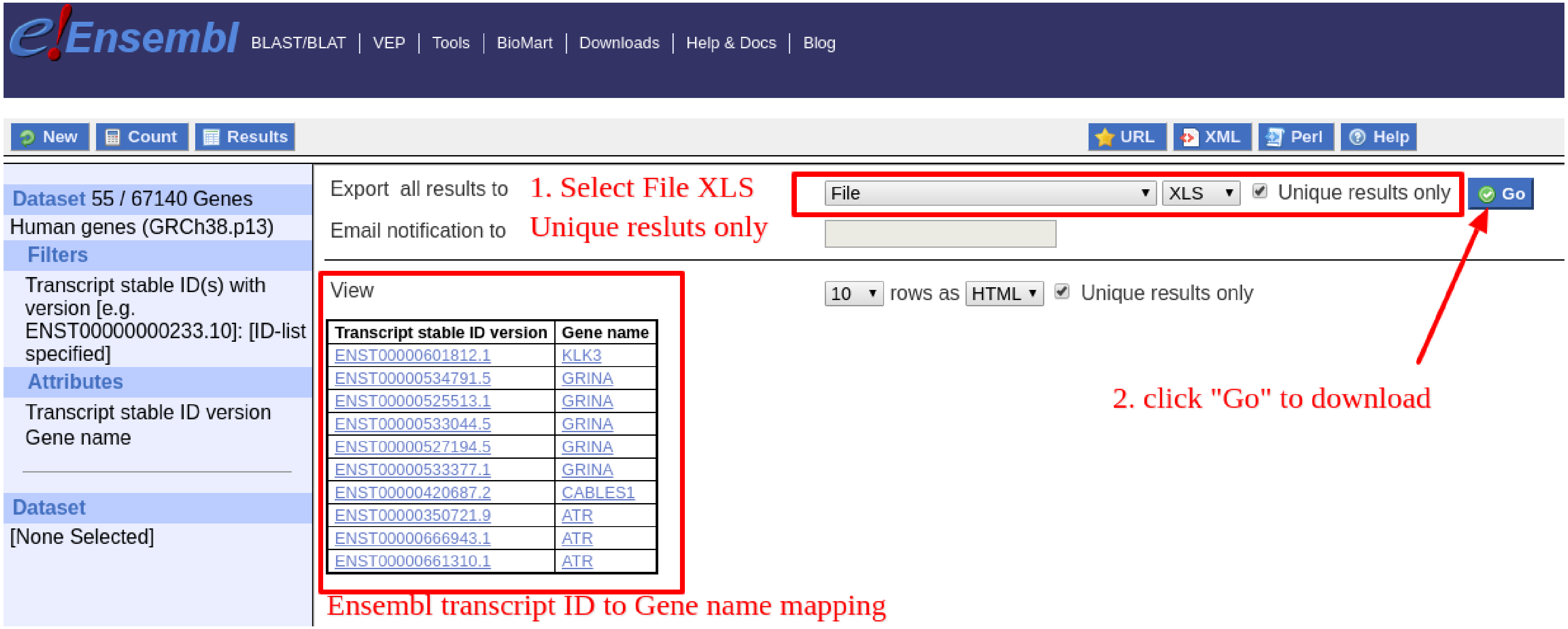
Ensembl

2. check "Transcript stable ID version" and "Gene name"

☐ Gene stable ID
☐ Gene stable ID version
☐ Transcript stable ID
☑ Transcript stable ID version
☐ Protein stable ID
☐ Protein stable ID version
☐ Exon stable ID
☐ Gene description
☐ Chromosome/scaffold name
☐ Gene start (bp)
☐ Gene end (bp)
☐ Strand
☐ Karyotype band
☐ Transcript start (bp)
☐ Transcript end (bp)
☐ Transcription start site (TSS)
☐ Transcript length (including UTRs and CDS)
☐ Transcript support level (TSL)

☐ GENCODE basic annotation
☐ APPRIS annotation
☐ RefSeq match transcript
☑ Gene name
☐ Source of gene name
☐ Transcript name
☐ Source of transcript name
☐ Transcript count
☐ Gene % GC content
☐ Gene type
☐ Transcript type
☐ Source (gene)
☐ Source (transcript)
☐ Version (gene)
☐ Version (transcript)
☐ Version (protein)
☐ Gene Synonym

Ensembl transcript ID (ENST...) to HGNC symbol (gene symbol) transformation



Dataset 55 / 67140 Genes
Human genes (GRCh38.p13)

Filters
Transcript stable ID(s) with version [e.g. ENST00000000233.10]: [ID-list specified]

Attributes
Transcript stable ID version
Gene name

Dataset
[None Selected]

Export all results to File ▾ XLS ▾ ☑ Unique results only

Email notification to

1. Select File XLS
Unique resluts only

View 10 ▾ rows as HTML ▾ ☑ Unique results only

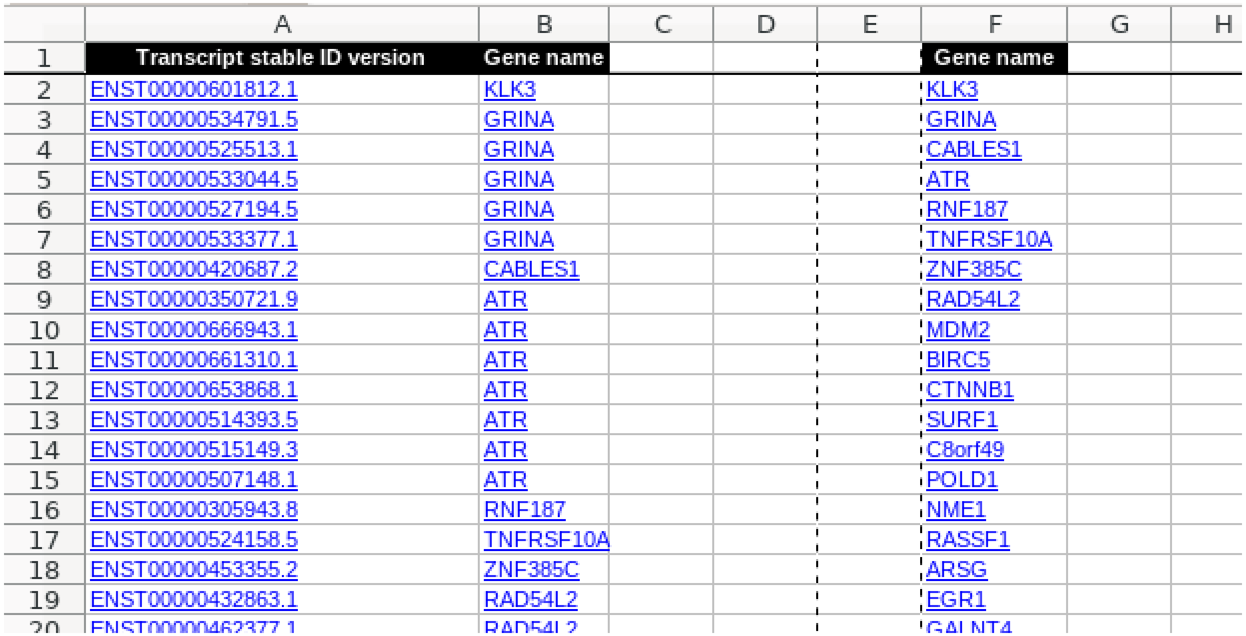
Transcript stable ID version	Gene name
ENST00000601812.1	KLK3
ENST00000534791.5	GRINA
ENST00000525513.1	GRINA
ENST00000533044.5	GRINA
ENST00000527194.5	GRINA
ENST00000533377.1	GRINA
ENST00000420687.2	CABLES1
ENST00000350721.9	ATR
ENST00000666943.1	ATR
ENST00000661310.1	ATR

2. click "Go" to download

Ensembl transcript ID to Gene name mapping

Ensembl transcript ID (ENST...) to HGNC symbol (gene symbol) transformation

Open in Excel and remove duplicate “Gene name” in separate column



	A	B	C	D	E	F	G	H
1	Transcript stable ID version	Gene name				Gene name		
2	ENST00000601812.1	KLK3				KLK3		
3	ENST00000534791.5	GRINA				GRINA		
4	ENST00000525513.1	GRINA				CABLES1		
5	ENST00000533044.5	GRINA				ATR		
6	ENST00000527194.5	GRINA				RNF187		
7	ENST00000533377.1	GRINA				TNFRSF10A		
8	ENST00000420687.2	CABLES1				ZNF385C		
9	ENST00000350721.9	ATR				RAD54L2		
10	ENST00000666943.1	ATR				MDM2		
11	ENST00000661310.1	ATR				BIRC5		
12	ENST00000653868.1	ATR				CTNNB1		
13	ENST00000514393.5	ATR				SURF1		
14	ENST00000515149.3	ATR				C8orf49		
15	ENST00000507148.1	ATR				POLD1		
16	ENST00000305943.8	RNF187				NME1		
17	ENST00000524158.5	TNFRSF10A				RASSF1		
18	ENST00000453355.2	ZNF385C				ARSG		
19	ENST00000432863.1	RAD54L2				EGR1		
20	ENST00000462377.1	RAD54L2				GALNT4		