

Article

Fernando Freire

December 23, 2018

Contents

1	Initializations	2
2	Lookup	2
3	Get sequences	2
4	Get triplets	3

1 Initializations

Script 1.0.1 (python)

```
1 import requests, sys
2 import json
3 import numpy as np
4 import matplotlib.pyplot as plt
5 try:
6     import pandas as pd
7     if pd.__version__ > "0.22.0":
8         has_pandas = True
9     else:
10         has_pandas = False
11 except ImportError:
12     has_pandas = False
13
14 %matplotlib inline
15
16 server = "http://rest.ensembl.org"
17 headers={ "Content-Type" : "application/json", "Accept" : "application/json"}
```

2 Lookup

Script 2.0.1 (python)

```
1 ext = "/lookup/id"
2 # Get lookup data
3 r = requests.post(server+ext, headers=headers, data='{ "expand" : 1, "ids" :
  ↳ ["ENSG00000157764"] }')
4
5 if not r.ok:
6     r.raise_for_status()
7     sys.exit()
8
9 found = json.loads(r.text)
```

3 Get sequences

Script 3.0.1 (python)

```
1 # Create list of ID
2 IDs = []
3 for transcripts in found.values():
4     for transcript in transcripts["Transcript"]:
5         IDs.append(transcript["id"])
6
```

```

7 dquotes_ids=json.dumps({"ids" : IDs})
8 print(dquotes_ids)
9
10 # Get sequence data
11 ext = "/sequence/id"
12
13 # Make request and check status
14 r = requests.post(server+ext, headers = headers, data = dquotes_ids)
15 if not r.ok:
16     r.raise_for_status()
17     sys.exit()
18
19 json_data = json.loads(r.text)

```

Output

```

{"ids": ["ENST00000496384", "ENST00000644120", "ENST00000642875", "ENST00000644969",
↪ "ENST00000646891", "ENST00000644905", "ENST00000642228", "ENST00000288602",
↪ "ENST00000645443", "ENST00000646730", "ENST00000479537", "ENST00000647434",
↪ "ENST00000644650", "ENST00000497784", "ENST00000646334", "ENST00000642272",
↪ "ENST00000643356", "ENST00000642808", "ENST00000643790", "ENST00000646427",
↪ "ENST00000469930"]}

```

4 Get triplets

From position 102 of each sequence. I am taking into account the position 102 of the string in python, which starts counting from position 0, as most programming languages do. If we wanted position 102 to start counting from 1 as the position of first base, line 4 would be:

```
triplet = record["seq"][101:104]
```

Script 4.0.1 (python)

```

1 count_triplets = {}
2 count_all = 0
3 for record in json_data:
4     triplet = record["seq"][102:105]
5     count_all += 1
6     if triplet in count_triplets:
7         count_triplets[triplet] += 1
8     else:
9         count_triplets[triplet] = 1
10
11 if has_pandas:
12     df = pd.DataFrame.from_dict(count_triplets, orient='index', columns=['Count'])
13     display(df)
14 else:
15     display(count_triplets)
16 plt.bar(count_triplets.keys(), count_triplets.values())

```

```
17 plt.xlabel('Triplets')
18 _ = plt.ylabel('Count')
```

	Count
ATA	1
GAC	3
GTA	2
GCC	3
CCG	4
GGC	2
TAG	1
TCC	1
CAT	1
CTC	1
GCG	1
TCT	1

