Traineeship Part 1 NCBI-esummary

April 22, 2020

Traineeship Part 1: Data collection (ids) using NCBI eUtils and esummary

Author: Iris Raes

The University of Antwerp, Medical Biochemistry, Campus Drie Eiken

Loading required packages

[ ]:
```
# pip3 install --user eutils
from eutils import Client
from Bio import Entrez
```

Personal API-key

[ ]:
```
eclient = Client(api_key="8ecce891e7fa036ff84bccc7c74e5138dc09")
```

---

1) Entrez Nucleotide Search - mRNA Transcript Variants

[ ]:
```
### Creating query
mRNAtranscripts = []
transcriptmRNA_esearch = eclient.esearch(db='nucleotide',
            term='DPP8[gene] AND "Homo sapiens"[Primary Organism] AND␣
 ↪(biomol_mrna[PROP] AND refseq[filter])')
print("\nLoading currently available ids from Entrez nucleotide...")
print("="*50)
print("\nTranscript variant ids: ")
print(transcriptmRNA_esearch.ids)
for item in transcriptmRNA_esearch.ids:
    mRNAtranscripts.append(item)
print("\nSearch results: {}\n".format(transcriptmRNA_esearch.count))
```

[ ]:
```
### Esummary for retrieving information
Entrez.email = "iris.raes@hotmail.com"
### For each id in mRNAtranscripts
counter = 1
for ids in mRNAtranscripts:
    handle = Entrez.esummary(db="nucleotide", id=ids)
```

```
    record = Entrez.read(handle)
    handle.close()
    print("[{}] Esummary for id ---> {}".format(counter,record[0]["Id"]))
    print(record[0]["Title"])
    print(record[0]["AccessionVersion"])
    print("Length: {} bp".format(record[0]["Length"]))
    counter += 1
    print("\n")
```

2) dbVar Search - Pathogenic Copy Number Variation in Human

[ ]:
```
### Creating query
CNV = []
CNV_esearch = eclient.esearch(db='dbVar',
            term='DPP8[All Fields] AND ("Homo sapiens"[Organism] AND "copy␣
 ↪number variation"[Variant Type] AND "Pathogenic"[clinical_interpretation])')
print("\nLoading currently available ids from dbVar...")
print("="*50)
print("dbVar ids: ")
print(CNV_esearch.ids)
for item in CNV_esearch.ids:
    CNV.append(item)
print("\nSearch results: {}\n".format(CNV_esearch.count))
```

[ ]:
```
### Esummary for retrieving information
Entrez.email = "iris.raes@hotmail.com"
### For each id in CNV
counter = 1
for ids in CNV:
    handle = Entrez.esummary(db="dbVar", id=ids)
    record = Entrez.read(handle)
    handle.close()
    #print(record)
    print("[{}] Esummary for id ---> {}".format(counter,ids))
    print("Variant Region ID: {}".
 ↪format(record['DocumentSummarySet']['DocumentSummary'][0]['SV']))
    print("Type: {}".
 ↪format(record['DocumentSummarySet']['DocumentSummary'][0]['dbVarVariantTypeList'][0]))
    print("Study ID: {}".
 ↪format(record['DocumentSummarySet']['DocumentSummary'][0]['ST']))
    print("Clinical Assertion: {}".
 ↪format(record['DocumentSummarySet']['DocumentSummary'][0]['dbVarClinicalSignificanceList'][
    print("-"*30)
    print("Position on chromosome assembly: {}".
 ↪format(record['DocumentSummarySet']['DocumentSummary'][0]['dbVarPlacementList'][0]['Assembl
```

```
    print("--> Start: {}".
↪format(record['DocumentSummarySet']['DocumentSummary'][0]['dbVarPlacementList'][0]['Chr_sta
    print("--> End: {}".
↪format(record['DocumentSummarySet']['DocumentSummary'][0]['dbVarPlacementList'][0]['Chr_end
    print("- -"*9)
    print("Position on chromosome assembly: {}".
↪format(record['DocumentSummarySet']['DocumentSummary'][0]['dbVarPlacementList'][1]['Assembl
    print("--> Start: {}".
↪format(record['DocumentSummarySet']['DocumentSummary'][0]['dbVarPlacementList'][1]['Chr_sta
    print("--> End: {}".
↪format(record['DocumentSummarySet']['DocumentSummary'][0]['dbVarPlacementList'][1]['Chr_end
    counter += 1
    print("\n")
```

3) dbVar Search - Insertions in Human

[ ]:
```
### Creating query
insertion_esearch = eclient.esearch(db='dbVar',
            term='DPP8[All Fields] AND ("Homo sapiens"[Organism] AND␣
 ↪"insertion"[Variant Type])')
print("\nLoading currently available ids from dbVar...")
print("="*50)
print("dbVar ids: ")
print(insertion_esearch.ids)
print("\nSearch results: {}\n".format(insertion_esearch.count))
```

4) dbVar Search - Inversions in Human

[ ]:
```
### Creating query
inversion_esearch = eclient.esearch(db='dbVar',
            term='DPP8[All Fields] AND ("Homo sapiens"[Organism] AND␣
 ↪"inversion"[Variant Type])')
print("\nLoading currently available ids from dbVar...")
print("="*50)
print("dbVar ids: ")
print(inversion_esearch.ids)
print("\nSearch results: {}\n".format(inversion_esearch.count))
```

5) dbVar Search - Short Tandem Repeats in Human (seems to be less important)

[ ]:
```
### Creating query
STR_esearch = eclient.esearch(db='dbVar',
            term='DPP8[All Fields] AND ("Homo sapiens"[Organism] AND "short␣
 ↪tandem repeat"[Variant Type])')
print("\nLoading currently available ids from dbVar...")
```

```
print("="*50)
print("dbVar ids: ")
print(STR_esearch.ids)
print("\nSearch results: {}\n".format(STR_esearch.count))
```

6) ClinVar Search - Genetic Variations in Human

[ ]:
```
### Creating query
ClinVar_esearch = eclient.esearch(db='ClinVar',
            term='DPP8[gene] AND "Single gene"')
print("\nLoading currently available ids from ClinVar...")
print("="*50)
print("\nClinVar ids: ")
print(ClinVar_esearch.ids)
print("\nSearch results: {}\n".format(ClinVar_esearch.count))
```