

Traineeship Part 1 NCBI

April 20, 2020

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

Author: Iris Raes

The University of Antwerp, Medical Biochemistry, Campus Drie Eiken

Loading required packages

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

Personal API-key

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

1) Entrez Nucleotide Search - mRNA Transcript Variants

```
[3]: ### Creating query
transcriptmRNA_eseach = eclient.esearch(db='nucleotide',
    term='DPP9[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_eseach.ids)
print("\nSearch results: {} \n".format(transcriptmRNA_eseach.count))
```

Loading currently available ids from Entrez nucleotide...

=====

Transcript variant ids:

[1370476185, 1034610004, 1034610002, 768004626, 768004622, 768004618, 768004616, 578833714, 1677498370, 1677499978]

Search results: 10

2) dbVar Search - Pathogenic Copy Number Variation in Human

```
[4]: ### Creating query
CNV_eseach = eclient.esearch(db='dbVar',
                             term='DPP9[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_eseach.ids)
print("\nSearch results: {} \n".format(CNV_eseach.count))
```

Loading currently available ids from dbVar...

=====

dbVar ids:

[49623411, 49353191, 49353005, 49350830, 49349701, 49349293, 49345450, 49344315, 48468240, 48466558, 48466447, 48453939, 45807136, 17813982, 17813734, 3740775, 3739972, 3738955, 3738954, 3738649, 1212838, 1137112]

Search results: 22

3) dbVar Search - Insertions in Human

```
[5]: ### Creating query
insertion_eseach = eclient.esearch(db='dbVar',
                                    term='DPP9[All Fields] AND ("Homo sapiens"[Organism] AND_
→"insertion"[Variant Type])')
print("\nLoading currently available ids from dbVar...")
print("="*50)
print("dbVar ids: ")
print(insertion_eseach.ids)
print("\nSearch results: {} \n".format(insertion_eseach.count))
```

Loading currently available ids from dbVar...

=====

dbVar ids:

[49597698, 49580472, 48530760, 48377645, 48377627, 47753859, 47564069, 47178696, 46791711, 45897195, 45896455, 45807279, 36885535, 24618684, 24516168, 24501143, 17814018, 17813982, 14212055, 14211117, 14209696, 13414404, 11399938, 8023314, 7738722, 7694891, 7590450, 7474153, 6477950, 6451851, 6354196, 5661470, 5431842, 5195919, 1297001, 1028299, 286824, 285317, 284926, 40396]

Search results: 40

4) dbVar Search - Inversions in Human

```
[6]: ### Creating query
inversion_esearch = eclient.esearch(db='dbVar',
                                     term='DPP9[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))
```

Loading currently available ids from dbVar...

=====

dbVar ids:

[48377627, 47178696, 46791711, 45807289, 45807279, 36885535, 24618684, 24516168, 24501143, 17814018, 17813982, 5195919, 1297001, 1028299]

Search results: 14

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

```
[7]: ### Creating query
STR_esearch = eclient.esearch(db='dbVar',
                               term='DPP9[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))
```

Loading currently available ids from dbVar...

=====

dbVar ids:

[35728959, 35728956, 35728945, 35728942, 35728939, 35728922, 35728913, 35728902, 35728888, 35728883, 35728872, 35728679, 35728652, 35728650, 35728640, 35728610, 35728601, 35728076, 35727391, 35727380, 35727364, 35727355, 35727352, 35727332, 35727324, 35726686, 35726677, 35726669, 35726663, 35726639, 30349921]

Search results: 31

6) ClinVar Search - Genetic Variations in Human

```
[8]: ### Creating query
ClinVar_eseach = eclient.eseach(db='ClinVar',
                                term='DPP9[gene] AND "Single gene"')
print("\nLoading currently available ids from ClinVar...")
print("="*50)
print("\nClinVar ids: ")
print(ClinVar_eseach.ids)
print("\nSearch results: {}".format(ClinVar_eseach.count))
```

Loading currently available ids from ClinVar...

=====

ClinVar ids:

[788833, 779179, 778595, 769947, 717743, 713315, 615908]

Search results: 7