Traineeship Part 1b NCBI

April 21, 2020

Traineeship Part 1: Data collection (ids) using NCBI eUtils (DPP8)

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_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)
print("\nSearch results: {}\n".format(transcriptmRNA_esearch.count))
```

```
Loading currently available ids from Entrez nucleotide…
==================================================

Transcript variant ids:
[1370466850, 1370466849, 1370466848, 1370466847, 1370466846, 1370466845,
1370466844, 1370466843, 1370466842, 1370466841, 1370466840, 1370466839,
1370466838, 1034591191, 1034591189, 530406104, 1676355481, 1675159331,
1675115520, 1675107575, 1674995210, 1519241926]
```

```
Search results: 22
```

2) dbVar Search - Pathogenic Copy Number Variation in Human

[4]:
```
### Creating query
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)
print("\nSearch results: {}\n".format(CNV_esearch.count))
```

```
Loading currently available ids from dbVar…
==================================================
dbVar ids:
[49355208, 49345988, 48482823, 48479604, 48476936, 48468493, 48467441, 48463636,
48462914, 48458970, 48456310, 48440267, 45807182, 45806585, 45805231, 45804309,
45803873, 45802836, 17813982, 17813734, 11417959, 3738417, 1137112]

Search results: 23
```

3) dbVar Search - Insertions in Human

[5]:
```
### 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))
```

```
Loading currently available ids from dbVar…
==================================================
dbVar ids:
[50024441, 48549155, 48377627, 36885535, 24618684, 24576021, 24558392, 24516168,
24501143, 24501142, 17814018, 17813982, 17336238, 14119771, 14115029, 14114189,
14113866, 14112658, 14112430, 14112429, 14111508, 14111245, 14107369, 14105300,
14104726, 14104193, 14103616, 13414404, 11399938, 8197100, 8159847, 8122040,
8057196, 8044108, 8007639, 7768471, 7752235, 7738302, 7722208, 7705457, 7688648,
7664161, 7609031, 7591442, 7570618, 7474009, 6648623, 6628573, 6602416, 6598575,
```

```
6568283, 6558551, 6491732, 6435184, 6381100, 6339301, 6327518, 6309426, 6307845,
6273396, 6271310, 6266973, 6208346, 6201794, 5661470, 5637858, 5637856, 5637855,
5637848, 5637847, 5637844, 5637843, 5469512, 5469511, 5195919, 1297001, 1028299,
200347, 200332, 198303, 198286, 197576, 196707, 194999, 193349, 193218, 40396]

Search results: 87
```

4) dbVar Search - Inversions in Human

[6]:
```
### 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))
```

```
Loading currently available ids from dbVar…
==================================================
dbVar ids:
[48549155, 48377627, 36885535, 25050625, 25020883, 24618684, 24618666, 24516168,
24501143, 24501142, 17814018, 17813982, 17336238, 5195919, 1297001, 1028299]

Search results: 16
```

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

[7]:
```
### 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))
```

```
Loading currently available ids from dbVar…
==================================================
dbVar ids:
[35556668, 35556667, 35556666, 35556665, 35556663, 35556662, 35556661, 35556660,
35556659, 35556658, 35556657, 35556656, 35556654, 35556653, 35556652, 35556651,
```

```
35556650, 35556649, 35556648, 35556647, 35556646, 35554677, 35554676, 35554675,
35554674, 35554672, 35554671, 35554670, 35554669, 35554667, 35554666, 35554665,
35554664, 35554663, 35554662, 35554661, 35554659, 35554658, 35553038, 35553037,
35553036, 35553035, 35553034, 35553033, 35553032, 35553031, 35553029, 35553028,
35553027, 35553026, 35553024, 35553023, 35553022, 35553021, 35553019, 35553018,
35553017, 35552871, 35552869, 35552868, 35552867, 35552866, 35552865, 35552864,
35552863, 35552862, 35552861, 35552860, 35552859, 35552858, 35552857, 35552856,
35552855, 35552854, 35552853, 35552852, 35552851, 35552850, 35552849, 30349921]

Search results: 80
```

6) ClinVar Search - Genetic Variations in Human

[8]:
```
### 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))
```

```
Loading currently available ids from ClinVar…
==================================================

ClinVar ids:
[614697, 614696]

Search results: 2
```