4

December 13, 2020

1 Combining OmniPath annotations and networks

1.0.1 Marton Olbei

09/12/2020 In this tutorial we show you how to query interactions from one of the many resources included in OmniPath, we map additional annotations to the data, and combine the two to build a small tissue specific network out of them

We'll start by importing libraries, first `omnipath`, and `pandas` for data wrangling.

[14]:
```
import omnipath as op
import pandas as pd
```

With the `omnipath` package loaded, we can download a subset of the interaction data. To take a look at the available databases, run:

[15]:
```
op.interactions.AllInteractions.resources()
```

[15]:
```
('ABS',
 'ACSN',
 'ACSN_SignaLink3',
 'ARACNe-GTEx_DoRothEA',
 'ARN',
 'Adhesome',
 'AlzPathway',
 'BEL-Large-Corpus_ProtMapper',
 'Baccin2019',
 'BioGRID',
 'BioGRID_ICELLNET',
 'CA1',
 'CancerCellMap',
 'CellPhoneDB',
 'CellPhoneDB_ICELLNET',
 'DEPOD',
 'DIP',
 'DOMINO',
 'DeathDomain',
 'Dinarello2013_ICELLNET',
 'DoRothEA',
```

```
'DoRothEA-reviews_DoRothEA',
'ELM',
'EMBRACE',
'ENCODE-distal',
'ENCODE-proximal',
'ENCODE_tf-mirna',
'FANTOM4_DoRothEA',
'Fantom5_LRdb',
'GO-lig-rec_ICELLNET',
'Guide2Pharma',
'Guide2Pharma_CellPhoneDB',
'Guide2Pharma_ICELLNET',
'Guide2Pharma_LRdb',
'HOCOMOCO_DoRothEA',
'HPMR',
'HPMR_ICELLNET',
'HPMR_LRdb',
'HPRD',
'HPRD-phos',
'HPRD_KEA',
'HPRD_LRdb',
'HPRD_MIMP',
'HTRIdb',
'HTRIdb_DoRothEA',
'HuRI',
'I2D_CellPhoneDB',
'ICELLNET',
'IMEx_CellPhoneDB',
'InnateDB',
'InnateDB-All_CellPhoneDB',
'InnateDB_CellPhoneDB',
'InnateDB_ICELLNET',
'InnateDB_SignaLink3',
'IntAct',
'IntAct_CellPhoneDB',
'IntAct_DoRothEA',
'JASPAR_DoRothEA',
'KEA',
'KEGG-MEDICUS',
'Kinexus_KEA',
'Kirouac2010',
'Kirouac2010_ICELLNET',
'LMPID',
'LRdb',
'Li2012',
'Lit-BM-17',
'LncRNADisease',
```

```
'MIMP',
'MINT_CellPhoneDB',
'MPPI',
'Macrophage',
'Macrophage_ICELLNET',
'MatrixDB',
'MatrixDB_CellPhoneDB',
'NCI-PID_ProtMapper',
'NFIRegulomeDB_DoRothEA',
'NRF2ome',
'NetPath',
'NetworKIN_KEA',
'ORegAnno',
'ORegAnno_DoRothEA',
'PAZAR',
'PAZAR_DoRothEA',
'PhosphoNetworks',
'PhosphoPoint',
'PhosphoSite',
'PhosphoSite_KEA',
'PhosphoSite_MIMP',
'PhosphoSite_ProtMapper',
'PhosphoSite_noref',
'ProtMapper',
'REACH_ProtMapper',
'RLIMS-P_ProtMapper',
'Ramilowski2015',
'Ramilowski2015_Baccin2019',
'Ramilowski2015_ICELLNET',
'ReMap_DoRothEA',
'Reactome_ICELLNET',
'Reactome_LRdb',
'Reactome_ProtMapper',
'Reactome_SignaLink3',
'RegNetwork_DoRothEA',
'SIGNOR',
'SIGNOR_ICELLNET',
'SIGNOR_ProtMapper',
'SPIKE',
'SPIKE_ICELLNET',
'STRING_ICELLNET',
'SignaLink3',
'SignaLink3_ICELLNET',
'Sparser_ProtMapper',
'TCRcuration_SignaLink3',
'TFactS_DoRothEA',
'TFe_DoRothEA',
```

```
 'TRED_DoRothEA',
 'TRIP',
 'TRRD_DoRothEA',
 'TRRUST_DoRothEA',
 'TransmiR',
 'UniProt_CellPhoneDB',
 'UniProt_LRdb',
 'Wang',
 'dbPTM',
 'iPTMnet',
 'iTALK',
 'lncrnadb',
 'miR2Disease',
 'miRDeathDB',
 'miRTarBase',
 'miRecords',
 'ncRDeathDB',
 'phosphoELM',
 'phosphoELM_KEA',
 'phosphoELM_MIMP')
```

To query the interactions from one of these sources, we can use the `interactions.AllInteractions` function.

[16]:
```
interactions = op.interactions.AllInteractions.get()
```

Let's get interactions coming from the `SIGNOR` database. It is important to mention here, that the `omnipath` library queries can be replicated in browser too, as they access specific URLs, depending on the parameters we give here. Feel free to give it a go: https://omnipathdb.org/interactions?genesymbols=yes&resources=SIGNOR&datasets=dorothea,kinaseextra,ligr

[17]:
```
interactions_filtered = interactions[interactions.sources.isin(['SIGNOR'])]
interactions_filtered
```

[17]:

	source	target	is_directed	is_stimulation	is_inhibition	\
799	P53355	P46821	True	True	False	
2416	Q8N726	P06748	True	False	True	
2814	Q12933	P61088	True	True	False	
3937	P23458	P38484	True	True	False	
3942	P22681	Q96JA1	True	False	True	
…	…	…	…	…	…	
149118	Q13526	P01574	True	False	True	
149119	Q01860	P31749	True	False	True	
149120	Q01860	Q9Y243	True	False	True	
149121	Q01860	P31751	True	False	True	
149122	Q01860	P26358	True	True	False	

	consensus_direction	consensus_stimulation	consensus_inhibition	\

```
799                  True                  True                 False
2416                False                 False                 False
2814                False                 False                 False
3937                False                 False                 False
3942                False                 False                 False
…                     …                     …                     …
149118               True                 False                  True
149119               True                 False                  True
149120               True                 False                  True
149121               True                 False                  True
149122               True                  True                 False

       dip_url  curation_effort       references sources                type  \
799       None                1  SIGNOR:18806760  SIGNOR  post_translational
2416      None                1  SIGNOR:14636574  SIGNOR  post_translational
2814      None                1  SIGNOR:18635759  SIGNOR  post_translational
3937      None                1  SIGNOR:19041276  SIGNOR  post_translational
3942      None                1  SIGNOR:15282549  SIGNOR  post_translational
…          …                  …               …       …                   …
149118    None                1  SIGNOR:16699525  SIGNOR     transcriptional
149119    None                1  SIGNOR:23041284  SIGNOR     transcriptional
149120    None                1  SIGNOR:23041284  SIGNOR     transcriptional
149121    None                1  SIGNOR:23041284  SIGNOR     transcriptional
149122    None                1  SIGNOR:22795133  SIGNOR     transcriptional

       references_stripped n_references  n_sources  n_primary_sources
799               18806760            1          1                  1
2416              14636574            1          1                  1
2814              18635759            1          1                  1
3937              19041276            1          1                  1
3942              15282549            1          1                  1
…                    …          …         …                 …
149118            16699525            1          1                  1
149119            23041284            1          1                  1
149120            23041284            1          1                  1
149121            23041284            1          1                  1
149122            22795133            1          1                  1

[4793 rows x 17 columns]
```

1.0.2 Importing annotations

To give these interactions a bit more depth, we can map annotation data to the interactions. To take a look at the available annotation resources in OmniPath, call the `requests.Annotations` function by running the line below.

[18]:
```
op.requests.Annotations.resources()
```

```
[18]: ('Adhesome',
       'Almen2009',
       'Baccin2019',
       'CORUM_Funcat',
       'CORUM_GO',
       'CSPA',
       'CSPA_celltype',
       'CancerGeneCensus',
       'CancerSEA',
       'CellCellInteractions',
       'CellPhoneDB',
       'CellPhoneDB_complex',
       'ComPPI',
       'DGIdb',
       'DisGeNet',
       'EMBRACE',
       'Exocarta',
       'GO_Intercell',
       'GPCRdb',
       'Guide2Pharma',
       'HGNC',
       'HPA_secretome',
       'HPA_subcellular',
       'HPA_tissue',
       'HPMR',
       'HPMR_complex',
       'ICELLNET',
       'ICELLNET_complex',
       'IntOGen',
       'Integrins',
       'KEGG-PC',
       'Kirouac2010',
       'LOCATE',
       'LRdb',
       'MCAM',
       'MSigDB',
       'Matrisome',
       'MatrixDB',
       'Membranome',
       'NetPath',
       'OPM',
       'Phobius',
       'Phosphatome',
       'Ramilowski2015',
       'Ramilowski_location',
       'SIGNOR',
       'SignaLink_function',
```

```
 'SignaLink_pathway',
 'Surfaceome',
 'TCDB',
 'TFcensus',
 'TopDB',
 'UniProt_family',
 'UniProt_keyword',
 'UniProt_location',
 'UniProt_tissue',
 'UniProt_topology',
 'Vesiclepedia',
 'Zhong2015',
 'iTALK',
 'kinase.com')
```

Let's import tissue enrichment data from the **Human Protein Atlas**. Calling the necessary function first we select the proteins we'd like to gather information on, followed by the resources we are pulling the data from.

A toy example below:

We assign the annotations of the proteins TP53 and LMNA from the tissue section of the HPA into the HPA_small variable. The "wide" setting pivots the data from a long format to wide, which gives us a nice table from the queried data.

[19]:
```
HPA_small = op.requests.Annotations.get(
    proteins = ['TP53, LMNA'],
    resources = 'HPA_tissue'
)
HPA_small
```

[19]:

	uniprot	genesymbol	entity_type	source	label	value	record_id
0	P04637	TP53	protein	HPA_tissue	organ	lymphoma	632671
1	P04637	TP53	protein	HPA_tissue	tissue	lymphoma	632671
2	P04637	TP53	protein	HPA_tissue	level	Not detected	632671
3	P04637	TP53	protein	HPA_tissue	n_not_detected	9	632671
4	P04637	TP53	protein	HPA_tissue	n_low	3	632671
..	…	…	…	…	…	…	
780	P04637	TP53	protein	HPA_tissue	level	Not detected	
781	P04637	TP53	protein	HPA_tissue	status	Enhanced	
782	P04637	TP53	protein	HPA_tissue	prognostic	False	
783	P04637	TP53	protein	HPA_tissue	favourable	False	
784	P04637	TP53	protein	HPA_tissue	pathology	False	

```
4       632671
..         …
780     632772
781     632772
782     632772
783     632772
784     632772

[785 rows x 7 columns]
```

These queries are also URL accessible. This toy query translates to the following:

URL: https://omnipathdb.org/annotations?resources=HPA_tissue&proteins=TP53,LMNA&license=academic

Let's get the unique proteins from SIGNOR into a list (set)

[20]:
```
SIGNOR_proteins = []
SIGNOR_proteins.extend(interactions_filtered['source'])
SIGNOR_proteins.extend(interactions_filtered['target'])
SIGNOR_proteins = list(set(SIGNOR_proteins)) #keep unique values only
```

We can pass this into `requests.Annotations.get()` just like above. To ensure sensible runtimes for this tutorial here we restrict the queried proteins to the first 500 in the list with the [0:499] slice (starting from 0 as Python uses zero based numbering).

[21]:
```
HPA_signor = op.requests.Annotations.get(
    proteins = SIGNOR_proteins[0:499],
    resources = 'HPA_tissue'
)
HPA_signor
```

[21]:
```
        uniprot genesymbol entity_type      source       label  \
0       P23511       NFYA     protein  HPA_tissue       organ
1       P23511       NFYA     protein  HPA_tissue      tissue
2       P23511       NFYA     protein  HPA_tissue       level
3       P23511       NFYA     protein  HPA_tissue      status
4       P23511       NFYA     protein  HPA_tissue  prognostic
…          …          …           …           …           …
316210  P10147       CCL3     protein  HPA_tissue      tissue
316211  P10147       CCL3     protein  HPA_tissue  prognostic
316212  P10147       CCL3     protein  HPA_tissue  favourable
316213  P10147       CCL3     protein  HPA_tissue       score
316214  P10147       CCL3     protein  HPA_tissue   pathology

                  value  record_id
0              prostate        797
1       glandular cells        797
2                Medium        797
3              Enhanced        797
```

```
4                False        797
…                …          …
316210     liver cancer    1421904
316211            False    1421904
316212            False    1421904
316213           0.3542    1421904
316214             True    1421904

[316215 rows x 7 columns]
```

First we need to pivot this "long" data format to a "wide" one.

[25]:
```
HPA_signor_wide = pd.pivot_table(HPA_signor, index = 'uniprot', columns =␣
↪'label', values = 'value', aggfunc='first')
HPA_signor_wide
```

[25]:
```
label   favourable         level n_high n_low n_medium n_not_detected  \
uniprot
A0PJZ3        True  Not detected      0     0        1              2
A1L390       False           Low      1     6        5              0
A8MT69        True           NaN    NaN   NaN      NaN            NaN
O00170        True        Medium      1     2        8              0
O00206       False  Not detected      0     0        0             10
…              …            …    …   …       …              …
Q9Y3A5        True        Medium      0     3        4              4
Q9Y3M8       False        Medium      0     6        4              1
Q9Y448       False        Medium      0     0        0              4
Q9Y4P8       False           Low      0     0        0              3
Q9Y5X5       False           NaN    NaN   NaN      NaN            NaN

label              organ pathology prognostic      score     status  \
uniprot
A0PJZ3    thyroid cancer      True      False     0.2515        NaN
A1L390     adrenal gland     False      False    0.07214   Enhanced
A8MT69    stomach cancer      True      False    0.02572        NaN
O00170          melanoma      True      False   0.006193   Approved
O00206             ovary     False      False  0.0006289   Approved
…                  …       …         …         …         …
Q9Y3A5   prostate cancer      True      False     0.4591   Approved
Q9Y3M8          prostate     False      False     0.1433   Enhanced
Q9Y448            breast     False      False    0.06306   Enhanced
Q9Y4P8    salivary gland     False      False   0.009274  Uncertain
Q9Y5X5   prostate cancer      True      False      0.151        NaN

label                  tissue
uniprot
A0PJZ3         thyroid cancer
A1L390        glandular cells
```

```
A8MT69         stomach cancer
O00170               melanoma
O00206   ovarian stroma cells
…                        …
Q9Y3A5        prostate cancer
Q9Y3M8        glandular cells
Q9Y448    myoepithelial cells
Q9Y4P8        glandular cells
Q9Y5X5        prostate cancer

[496 rows x 12 columns]
```

Now that we have the data, let's filter it down to a specific case, like breast cancer. We filter out rows where the levels of the proteins are favourable, i.e. we only move forward with the ones that are.

[23]:
```
HPA_breast_cancer = HPA_signor_wide[
    (HPA_signor_wide['tissue'] != 'breast cancer') &
    (HPA_signor_wide['favourable'] == 'True' )]
```

[24]:
```
HPA_breast_cancer
```

[24]:
```
label   favourable         level n_high n_low n_medium n_not_detected  \
uniprot
A0PJZ3        True  Not detected      0     0        1              2
A8MT69        True           NaN    NaN   NaN      NaN            NaN
O00170        True        Medium      1     2        8              0
O14511        True  Not detected      0     3        0              8
O14842        True           NaN    NaN   NaN      NaN            NaN
…              …            …     …    …       …              …
Q9UI33        True           Low      0     6        1              4
Q9UJ55        True           NaN    NaN   NaN      NaN            NaN
Q9ULT6        True        Medium      0     0       11              0
Q9Y261        True        Medium      0     1        6              4
Q9Y3A5        True        Medium      0     3        4              4

label                   organ pathology prognostic      score     status  \
uniprot
A0PJZ3         thyroid cancer      True      False     0.2515        NaN
A8MT69         stomach cancer      True      False    0.02572        NaN
O00170               melanoma      True      False   0.006193   Approved
O14511           renal cancer      True      False    0.00779  Uncertain
O14842      pancreatic cancer      True      False    0.04515        NaN
…                        …       …        …        …        …
Q9UI33      urothelial cancer      True      False  0.0008296        NaN
Q9UJ55   head and neck cancer      True      False     0.3796        NaN
Q9ULT6        prostate cancer      True      False     0.3942        NaN
Q9Y261         ovarian cancer      True       True  0.0006283   Enhanced
```

```
Q9Y3A5        prostate cancer      True      False     0.4591   Approved

label                   tissue
uniprot
A0PJZ3         thyroid cancer
A8MT69         stomach cancer
O00170               melanoma
O14511           renal cancer
O14842      pancreatic cancer
…                        …
Q9UI33      urothelial cancer
Q9UJ55   head and neck cancer
Q9ULT6        prostate cancer
Q9Y261         ovarian cancer
Q9Y3A5        prostate cancer

[74 rows x 12 columns]
```