Data Distillery Demo: LINCS use case

NIH Demo July 13th

For a *selected* drug transporter or drug processing enzyme, find the tissue where these are highly expressed (GTEx), and the drugs that may induce or suppress their gene expression (LINCs).

Use Case

- For a specific drug transporter or drug processing enzyme, find the tissue where these transporters and enzymes are highly expressed (GTEx), and the drugs that may induce or suppress the expression of these genes (LINCS).

Motivation for use case:

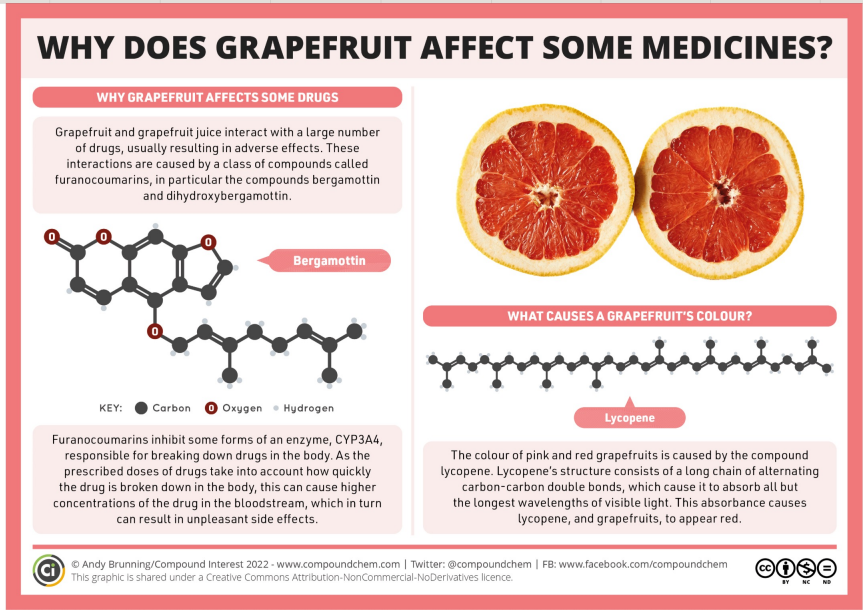
- Cytochrome P450 (CYP450) proteins are a superfamily of enzymes involved in the synthesis and metabolism of a range of internal and external cellular components
- An overly active CYP enzyme will render the drug ineffective. However, if these enzymes are not active enough, the drug can stay in the body for a prolonged duration leading to toxicity.
- Of all the different CYP proteins that are present in the human body, six of them are involved in the metabolism of 90% of drugs. These proteins are CYP1A2, CYP2C9, CYP2D6, **CYP3A4**, and CYP3A5. The most important are CYP3A4 and CYP2D6.

Motivation for use case (cont.)

Grapefruit juice was often used as a vehicle for drugs and placebos in double-blind clinical trials, since the strong bitter taste masked the taste of drugs. This preserved the double-blind nature of the trial.

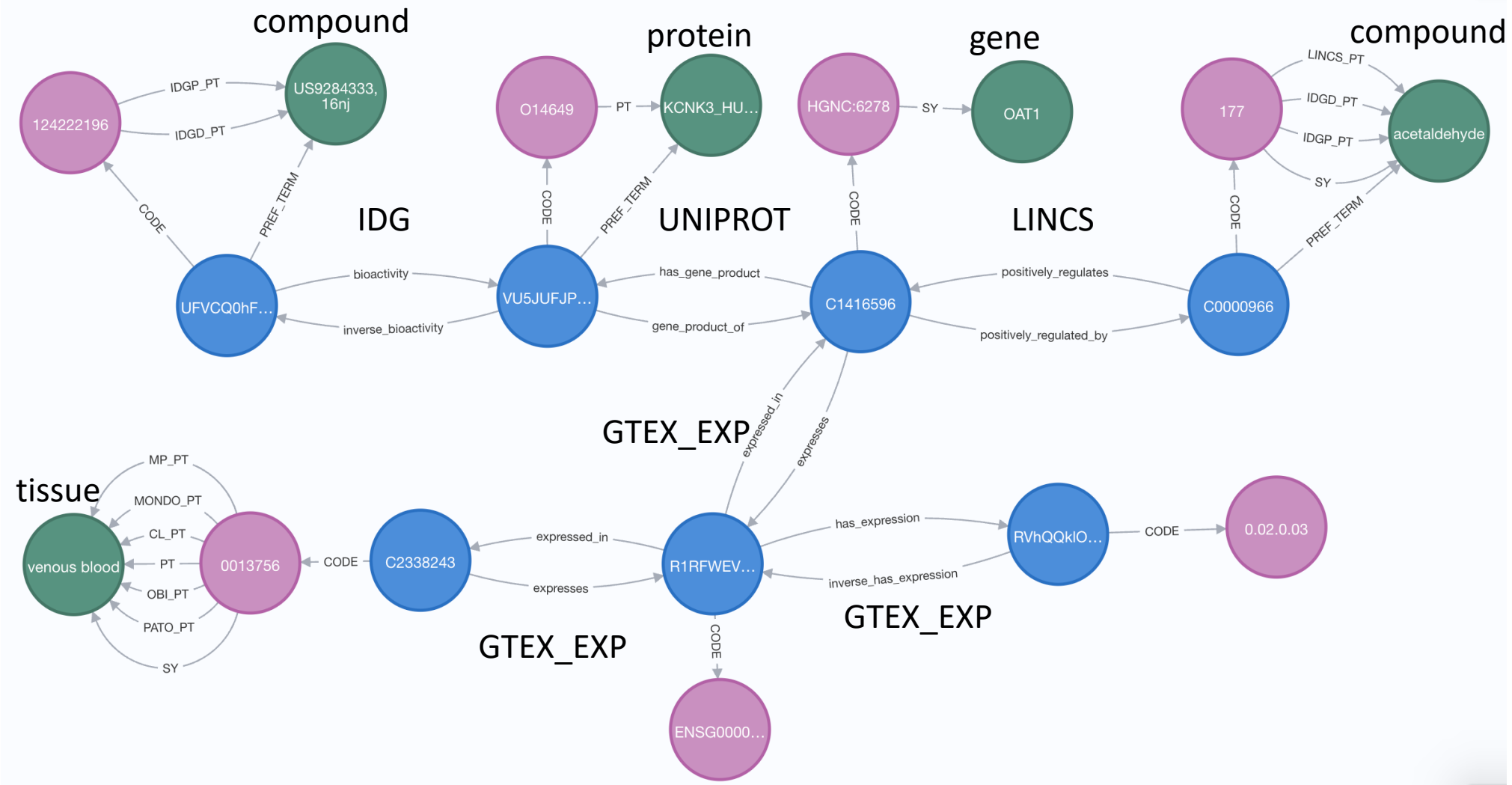
However, grapefruit juice contains psoralens which produce "suicide inhibition" of CYP3A4 (a reactive intermediate forms a covalent bond to irreversibly inactivate the enzyme). This causes large increases in blood levels of drugs that are metabolized by CYP3A4, notably lovastatin (see below data from [15]). Untoward reactions from this effect have been noted for midazolam, triazolam and buspirone (impaired CNS function), and felodipine (hypotension).

The Psoralens (linear furanocoumarins) , **bergamottin** and **dihydroxybergamottin**.



Use case iv) For a specific drug transporter or drug processing enzyme, find the tissue where these transporters and enzymes are highly expressed (GTEx), and the drugs that may induce or suppress the expression of these genes (LINCS).

Specific drug transporter: OAT1



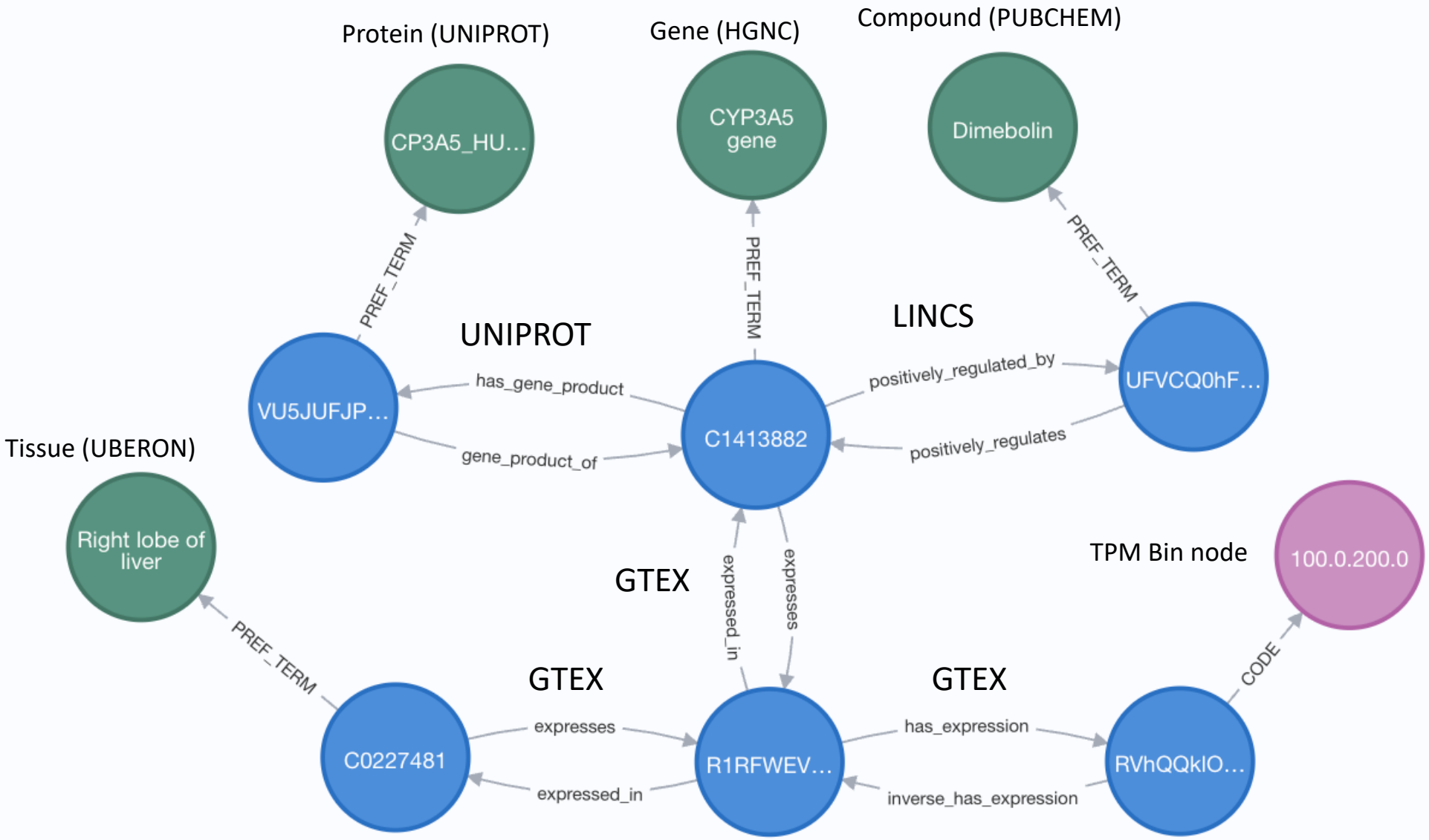
*Need to:

1. Remove Code nodes where theyre not needed
2. load newer version of graph to get rid of duplicated edges
3. Add cypher query
4. Add (SAB) after protein/gene/tissue text to show where they are from

Graphical Representation of Query

Sources of edges:
UNIPROT
LINCS
GTEX

Sources of nodes:
UNIPROT
HGNC
PUBCHEM
GTEX
UBERON



Query the GTEX data and find the top ~100 expressed genes (TPM) in Kidney and Liver

FIND TOP ~100 EXPRESSED GENES IN KIDNEY.

```
WITH 'cortex of kidney' AS TISSUE
MATCH (gtexexp_code:Code {SAB:'GTEXEXP'})<-[:CODE]-(gtexexp_cui:Concept)<-[:expresses]-(hgnc_cui:Concept)-[:CODE]-
>(hgnc_code:Code {SAB:'HGNC'})-[:PT]-(hgnc_term:Term)
MATCH (expbins_code:Code {SAB:'EXPBINS'})<-[:CODE]-(expbins_cui:Concept)<-[:has_expression]-(gtexexp_cui)-[]-
(ub_cui:Concept)-[:CODE]-(ub_code:Code {SAB:'UBERON'} )-[:PT]-(ub_term:Term {name:TISSUE})
WHERE expbins_code.lowerbound >= 300
RETURN COUNT(DISTINCT hgnc_term.name)
```

--> Returns 114 genes

FIND TOP ~100 EXPRESSED GENES IN LIVER

```
WITH 'right lobe of liver' AS TISSUE
MATCH (gtexexp_code:Code {SAB:'GTEXEXP'})<-[:CODE]-(gtexexp_cui:Concept)<-[:expresses]-(hgnc_cui:Concept)-[:CODE]->
(hgnc_code:Code {SAB:'HGNC'})-[:PT]-(hgnc_term:Term)
MATCH (expbins_code:Code {SAB:'EXPBINS'})<-[:CODE]-(expbins_cui:Concept)<-[:has_expression]-(gtexexp_cui)-[]-(ub_cui:Concept)
-[:CODE]-(ub_code:Code {SAB:'UBERON'} )-[:PT]-(ub_term:Term {name:TISSUE})
WHERE expbins_code.lowerbound >= 500
RETURN COUNT(DISTINCT hgnc_term.name)
```

--> Returns 109 genes

Define Gene List

```
WITH ['CYP1A2', 'CYP2C9', 'CYP2D6', 'CYP3A4', 'CYP3A5'] AS GENE_LIST
```

Match GTEX to HGNC

```
MATCH (a:Code {SAB:'GTEXEXP'})<-[:CODE]-(b:Concept)<-[:expresses]-(c:Concept)-[:CODE]->
(d:Code {SAB:'HGNC'})-[]->(e:Term) WHERE e.name IN GENE_LIST
```

Match TPM Expression to UBERON through GTEX

```
MATCH (f:Code {SAB:'EXPBINS'})<-[:CODE]-(g:Concept)-[:has_expression]-(b)-[:expressed_in]->
(h:Concept)-[:CODE]-(i:Code {SAB:'UBERON'})
```

Define variables to return

```
WITH REDUCE(m='',w in split(a.CODE,'-')[2..] |m+w+' ') AS tissueStr, f.lowerbound AS TPM_lb,
f.upperbound AS TPM_ub, e.name AS Gene
```

Return variables

```
RETURN DISTINCT Gene, tissueStr, TPM_lb,TPM_ub ORDER BY TPM_lb DESC LIMIT 5
```

	Gene	tissueStr	TPM_lb	TPM_ub
1	"CYP2C9"	"Liver "	400.0	500.0
2	"CYP3A4"	"Liver "	300.0	400.0
3	"CYP2D6"	"Liver "	200.0	300.0
4	"CYP3A5"	"Liver "	100.0	200.0
5	"CYP1A2"	"Liver "	71.0	72.0

Same query but with PUBCHEM IDs returned for each gene

Find compounds that affect expression of genes

```
WITH ['CYP1A2', 'CYP2C9', 'CYP2D6', 'CYP3A4', 'CYP3A5'] AS GENE_LIST
MATCH (hgnc_cui:Concept)-[:CODE]->(hgnc_code:Code {SAB:'HGNC'})-[]->(hgnc_term:Term)
WHERE hgnc_term.name IN GENE_LIST
MATCH (hgnc_cui)-[:expresses]->(gtexexp_cui:Concept)-[:CODE]->(gtexexp_code:Code {SAB:'GTEXEXP'})
MATCH (expbins_code:Code {SAB:'EXPBINS'})<-[:CODE]-(expbins_cui:Concept)-[:has_expression]-(gtexexp_cui)
-[:expressed_in]->(ub_cui:Concept)-[:CODE]-(ub_code:Code {SAB:'UBERON'})
WITH expbins_code.lowerbound AS TPM_lb, expbins_code.upperbound AS TPM_ub,
hgnc_term.name AS Gene, hgnc_cui
MATCH (hgnc_cui)-[ {SAB:'LINCS'}]-(pubchem_cui:Concept)-[:CODE]-
(pubchem_code:Code {SAB:'PUBCHEM'})-[:LINCS_PT]-(pubchem_term:Term)
RETURN DISTINCT Gene, COLLECT(DISTINCT pubchem_term.name) AS PUBCHEM_TERMs
```

Gene	PUBCHEM_TERMs
"CYP3A4"	["MW-SHH-97"]
"CYP3A5"	["latrepirdine", "KU-C104226", "ST-056792"]

```
MATCH (hgnc_cui:Concept)-[:CODE]->(hgnc_code:Code {SAB:'HGNC'})-[:PT]->(hgnc_term:Term)
WHERE hgnc_term.name contains 'transporter'
MATCH (hgnc_cui)-[:expresses]->(gtexexp_cui:Concept)-[:CODE]->(gtexexp_code:Code {SAB:'GTEXEXP'})
MATCH (expbins_code:Code {SAB:'EXPBINS'})<-[:CODE]-(expbins_cui:Concept)-[:has_expression]-(gtexexp_cui)
WITH expbins_code.lowerbound AS TPM_lb, expbins_code.upperbound AS TPM_ub,
hgnc_term.name AS Gene, hgnc_cui,hgnc_code
MATCH (hgnc_cui)-[ {SAB:'LINCS'}]-(pubchem_cui:Concept)-[:CODE]-(pubchem_code:Code {SAB:'PUBCHEM'})-[:LINCS_PT]-(pubchem_term:Term)
RETURN DISTINCT Gene,hgnc_code.CODE AS hgnc_code ,COLLECT(DISTINCT pubchem_term.name) AS PUBCHEM_IDs LIMIT 50
```

Gene	hgnc_code	tpm_lb	tpm_ub	tissue	PUBCHEM_IDs
"Fc gamma receptor and transporter"	"HGNC:3621"	100.0	200.0	"right lobe of liver"	["Clonazepam", "Carisoprodol", "Clioquinol", "Aceproma
"solute carrier organic anion transporter family member 1B3"	"HGNC:10961"	41.0	42.0	"right lobe of liver"	["KU-C103875", "immethridine", "Mepacrine", "Norfloxa
"magnesium transporter 1"	"HGNC:28880"	31.0	32.0	"right lobe of liver"	["PIK-75", "Methylphenidate", "ML-9"]
"NPC intracellular cholesterol transporter 2"	"HGNC:14537"	18.0	19.0	"right lobe of liver"	["Nortriptyline", "Protriptyline", "Amiodarone", "Zaleplon
"transporter 1, ATP binding cassette subfamily B member"	"HGNC:43"	13.0	14.0	"right lobe of liver"	["idarubicin", "Valrubicin", "Epirubicin", "SIB-1893", "do
"NPC intracellular cholesterol transporter 1"	"HGNC:7897"	3.0	4.0	"right lobe of liver"	["BGT-226", "HC-toxin", "Metronidazole", "NTNCB", "Im
"receptor transporter protein 4"	"HGNC:23992"	2.0	3.0	"right lobe of liver"	["Dacarbazine"]
"solute carrier organic anion transporter family member 4A1"	"HGNC:10953"	1.0	2.0	"right lobe of liver"	["clonixin-lysinate", "EMF-BCA1-64", "Calcifediol", "Sulf
"solute carrier organic anion transporter family member 3A1"	"HGNC:10952"	1.0	2.0	"right lobe of liver"	["N-methylquipazine", "Bindarit", "mecamylamine", "Dar
"solute carrier organic anion transporter family member 2A1"	"HGNC:10955"	1.0	2.0	"right lobe of liver"	["L-theanine"]

Return compounds affecting genes containing the word ‘transporter’ , as well as their TPM in a specified tissue