Synaptic database quering examples

Oksana Sorokina

08.04.2021

1 Connect DB

```
con <- DBI::dbConnect(RSQLite::SQLite(), dbname = dbname)
```

2 Query

2.1 Get specific gene information

```
gns<-con %>% tbl("FullGeneFullPaperFullRegion") %>% filter(HumanEntrez == "1742") %>%
  select(Localisation, HumanEntrez, HumanName, PaperPMID, Paper, Year, BrainRegion)

gns %>% show_query()
```

```
## <SQL>
## SELECT `Localisation`,`HumanEntrez`,`HumanName`,`PaperPMID`,`Paper`,`Year`,`BrainRegion`
## FROM `FullGeneFullPaperFullRegion`
## WHERE (`HumanEntrez` = '1742')
```

```
gns %>% collect() %>% pander()
```

Localisation	HumanEntrez	HumanName	PaperPMID	Paper	Year	BrainRegion
Postsynaptic	1742	DLG4	10818142	WALIKONIS_2000	2000	Forebrain
Postsynaptic	1742	DLG4	10862698	HUSI_2000	2000	Forebrain
Postsynaptic	1742	DLG4	11895482	SATON_2002	2002	Forebrain
Postsynaptic	1742	DLG4	14532281	LI_2004	2004	Forebrain
Postsynaptic	1742	DLG4	14720225	YOSHIMURA_2004	2004	Forebrain
Postsynaptic	1742	DLG4	1.5e+07	PENG_2002	2004	Forebrain
Postsynaptic	1742	DLG4	15169875	JORDAN_2004	2004	Brain
Postsynaptic	1742	DLG4	15748150	TRINIDAD_2005	2005	Brain
Postsynaptic	1742	DLG4	16332460	DOSEMESI_2006	2006	Hippocampus
Postsynaptic	1742	DLG4	16507876	CHENG_2006	2006	Forebrain
Postsynaptic	1742	DLG4	16507876	CHENG_2006	2006	Cerebellum
Postsynaptic	1742	DLG4	16635246	COLLINS_2006	2006	Forebrain
Postsynaptic	1742	DLG4	17623647	DOSEMESI_2007	2007	Cerebral cortex
Postsynaptic	1742	DLG4	18056256	TRINIDAD_2008	2008	Midbrain
Postsynaptic	1742	DLG4	18056256	TRINIDAD_2008	2008	Cerebellum
Postsynaptic	1742	DLG4	18056256	TRINIDAD_2008	2008	Hippocampus
Postsynaptic	1742	DLG4	19402746	SELIMI_2009	2009	Cerebellum
Postsynaptic	1742	DLG4	19455133	FERNANDEZ_2009	2009	Forebrain
Synaptosome	1742	DLG4	20309889	FILIOU_2010	2010	Brain
Postsynaptic	1742	DLG4	21170055	BAYES_2011	2010	Cerebral cortex

Localisation	HumanEntrez	HumanName	PaperPMID	Paper	Year	BrainRegion
Synaptosome	1742	DLG4	21398567	DAHlHAUS_2011	2011	Visual cortex
Postsynaptic	1742	DLG4	22632720	SCHWENK_2013	2013	Brain
Postsynaptic	1742	DLG4	23071613	BAYES_2012	2012	Cerebral cortex
Presynaptic	1742	DLG4	23622064	BOYKEN_2013	2013	Cerebral cortex
Synaptosome	1742	DLG4	24413018	BIESEMANN_2014	2014	Forebrain
Presynaptic	1742	DLG4	24534009	WEINGARTEN_2014	2014	Brain
Presynaptic	1742	DLG4	24876496	WILHELM_2014	2014	Cerebellum
Postsynaptic	1742	DLG4	25211037	DISTLER_2014	2014	Hippocampus
Synaptosome	1742	DLG4	25211037	DISTLER_2014	2014	Hippocampus
Synaptosome	1742	DLG4	25352669	LIU_2014	2014	Prefrontal cortex
Postsynaptic	1742	DLG4	25429717	BAYES_2014	2014	Frontal lobe
Synaptosome	1742	DLG4	2.6e+07	CHANG_2015	2015	Hippocampus
Synaptosome	1742	DLG4	2.6e+07	CHANG_2015	2015	Motor cortex
Synaptosome	1742	DLG4	27115346	KOHANSAL_NODEHI_2016	2016	Cerebellum
Synaptosome	1742	DLG4	27115346	KOHANSAL_NODEHI_2016	2016	Cerebral cortex
Postsynaptic	1742	DLG4	27507650	LI_2016	2016	Hippocampus
Postsynaptic	1742	DLG4	27609886	UEZU_2016	2016	Cerebral cortex
Synaptosome	1742	DLG4	27748445	GONZALEZ_LOZANO_2016	2016	Cerebral cortex
Postsynaptic	1742	DLG4	27898073	FOCKING_2016	2016	Brain
Postsynaptic	1742	DLG4	28671696	LI_2017	2017	Hippocampus
Synaptosome	1742	DLG4	28713243	ALFIERI_2017	2017	Telencephalon
Postsynaptic	1742	DLG4	2.9e+07	FERNANDEZ_2017	2017	Forebrain
Postsynaptic	1742	DLG4	29203896	ROY_2017	2017	Frontal lobe
Postsynaptic	1742	DLG4	29203896	ROY_2017	2017	Occipital lobe
Postsynaptic	1742	DLG4	29203896	ROY_2017	2017	Temporal lobe
Postsynaptic	1742	DLG4	29203896	ROY_2017	2017	Parietal lobe
Synaptosome	1742	DLG4	29610302	HEO_2018	2018	Hippocampus
Synaptosome	1742	DLG4	29610302	HEO_2018	2018	Cerebral cortex
Postsynaptic	1742	DLG4	30071621	ROY_2018	2018	Cerebellum
Postsynaptic	1742	DLG4	30071621	ROY_2018	2018	Hypothalamus
Postsynaptic	1742	DLG4	30071621	ROY_2018	2018	Hippocampus
Postsynaptic	1742	DLG4	30071621	ROY_2018	2018	Striatum
Postsynaptic	1742	DLG4	30071621	ROY_2018	2018	Frontal lobe
Presynaptic	1742	DLG4	30301801	KOKOTOS_2018	2018	Cerebellum
Postsynaptic	1742	DLG4	3.1e+07	WILSON_2019	2019	Cerebral cortex

2.2 Get paper information for gene list

```
gns<-con %>% tbl("FullGeneFullPaperFullRegion") %>% filter(HumanEntrez %in% c("4130","10458", "57554")) %>%
  select(Localisation, HumanEntrez, HumanName, PaperPMID, Paper, Year, BrainRegion)
gns %>% collect() %>% pander()
```

Localisation	HumanEntrez	HumanName	PaperPMID	Paper	Year	BrainRegion
Postsynaptic	4130	MAP1A	14532281	LI_2004	2004	Forebrain
Postsynaptic	4130	MAP1A	14720225	YOSHIMURA_2004	2004	Forebrain
Postsynaptic	4130	MAP1A	1.5e+07	PENG_2002	2004	Forebrain
Postsynaptic	4130	MAP1A	15169875	JORDAN_2004	2004	Brain
Postsynaptic	4130	MAP1A	15447677	FARR_2004	2004	Brain
Postsynaptic	4130	MAP1A	15748150	TRINIDAD_2005	2005	Brain
Postsynaptic	4130	MAP1A	16332460	DOSEMESI_2006	2006	Hippocampus
Postsynaptic	4130	MAP1A	16507876	CHENG_2006	2006	Forebrain

Localisation	HumanEntrez	HumanName	PaperPMID	Paper	Year	BrainRegion
Postsynaptic	4130	MAP1A	16507876	CHENG_2006	2006	Cerebellum
Postsynaptic	4130	MAP1A	16635246	COLLINS_2006	2006	Forebrain
Presynaptic	4130	MAP1A	17110340	TAKAMORI_2006	2006	Cerebral cortex
Postsynaptic	4130	MAP1A	17623647	DOSEMESI_2007	2007	Cerebral cortex
Postsynaptic	4130	MAP1A	18056256	TRINIDAD_2008	2008	Midbrain
Postsynaptic	4130	MAP1A	18056256	TRINIDAD_2008	2008	Cerebellum
Postsynaptic	4130	MAP1A	18056256	TRINIDAD_2008	2008	Hippocampus
Postsynaptic	4130	MAP1A	19455133	FERNANDEZ_2009	2009	Forebrain
Presynaptic	4130	MAP1A	19562802	ABUL_HUSN_2009	2009	Hippocampus
Presynaptic	4130	MAP1A	19562802	ABUL_HUSN_2009	2009	Striatum
Presynaptic	4130	MAP1A	20053882	GRONBORG_2010	2010	Cerebral cortex
Synaptosome	4130	MAP1A	20309889	FILIOU_2010	2010	Brain
Postsynaptic	4130	MAP1A	21170055	BAYES_2011	2010	Cerebral cortex
Synaptosome	4130	MAP1A	21398567	DAHlHAUS_2011	2011	Visual cortex
Postsynaptic	4130	MAP1A	23071613	BAYES_2012	2012	Cerebral cortex
Presynaptic	4130	MAP1A	24876496	WILHELM_2014	2014	Cerebellum
Postsynaptic	4130	MAP1A	25211037	DISTLER_2014	2014	Hippocampus
Synaptosome	4130	MAP1A	25211037	DISTLER_2014	2014	Hippocampus
Synaptosome	4130	MAP1A	25352669	LIU_2014	2014	Prefrontal cortex
Postsynaptic	4130	MAP1A	25429717	BAYES_2014	2014	Frontal lobe
Synaptosome	4130	MAP1A	2.6e+07	CHANG_2015	2015	Hippocampus
Synaptosome	4130	MAP1A	2.6e+07	CHANG_2015	2015	Motor cortex
Synaptosome	4130	MAP1A	27115346	KOHANSAL_NODEHI_2016	2016	Cerebellum
Synaptosome	4130	MAP1A	27115346	KOHANSAL_NODEHI_2016	2016	Cerebral cortex
Postsynaptic	4130	MAP1A	27609886	UEZU_2016	2016	Cerebral cortex
Synaptosome	4130	MAP1A	27748445	GONZALEZ_LOZANO_2016	2016	Cerebral cortex
Postsynaptic	4130	MAP1A	27898073	FOCKING_2016	2016	Brain
Postsynaptic	4130	MAP1A	29203896	ROY_2017	2017	Frontal lobe
Postsynaptic	4130	MAP1A	29203896	ROY_2017	2017	Occipital lobe
Postsynaptic	4130	MAP1A	29203896	ROY_2017	2017	Temporal lobe
Postsynaptic	4130	MAP1A	29203896	ROY_2017	2017	Parietal lobe
Synaptosome	4130	MAP1A	29610302	HEO_2018	2018	Hippocampus
Synaptosome	4130	MAP1A	29610302	HEO_2018	2018	Cerebral cortex
Postsynaptic	4130	MAP1A	30071621	ROY_2018	2018	Cerebellum
Postsynaptic	4130	MAP1A	30071621	ROY_2018	2018	Hypothalamus
Postsynaptic	4130	MAP1A	30071621	ROY_2018	2018	Hippocampus
Postsynaptic	4130	MAP1A	30071621	ROY_2018	2018	Striatum
Postsynaptic	4130	MAP1A	30071621	ROY_2018	2018	Frontal lobe
Presynaptic	4130	MAP1A	30301801	KOKOTOS_2018	2018	Cerebellum
Postsynaptic	4130	MAP1A	3.1e+07	WILSON_2019	2019	Cerebral cortex
Postsynaptic	10458	BAIAP2	10818142	WALIKONIS_2000	2000	Forebrain
Postsynaptic	10458	BAIAP2	11895482	SATON_2002	2002	Forebrain
Postsynaptic	10458	BAIAP2	14532281	LI_2004	2004	Forebrain
Postsynaptic	10458	BAIAP2	14720225	YOSHIMURA_2004	2004	Forebrain
Postsynaptic	10458	BAIAP2	1.5e+07	PENG_2002	2004	Forebrain
Postsynaptic	10458	BAIAP2	15169875	JORDAN_2004	2004	Brain
Postsynaptic	10458	BAIAP2	15748150	TRINIDAD_2005	2005	Brain
Postsynaptic	10458	BAIAP2	16635246	COLLINS_2006	2006	Forebrain
Postsynaptic	10458	BAIAP2	17623647	DOSEMESI_2007	2007	Cerebral cortex
Postsynaptic	10458	BAIAP2	18056256	TRINIDAD_2008	2008	Midbrain
Postsynaptic	10458	BAIAP2	18056256	TRINIDAD_2008	2008	Cerebellum
Postsynaptic	10458	BAIAP2	18056256	TRINIDAD_2008	2008	Hippocampus

3

Localisation	HumanEntrez	HumanName	PaperPMID	Paper	Year	BrainRegion
Postsynaptic	10458	BAIAP2	19402746	SELIMI_2009	2009	Cerebellum
Postsynaptic	10458	BAIAP2	19455133	FERNANDEZ_2009	2009	Forebrain
Synaptosome	10458	BAIAP2	20309889	FILIOU_2010	2010	Brain
Postsynaptic	10458	BAIAP2	21170055	BAYES_2011	2010	Cerebral cortex
Synaptosome	10458	BAIAP2	21398567	DAHlHAUS_2011	2011	Visual cortex
Postsynaptic	10458	BAIAP2	23071613	BAYES_2012	2012	Cerebral cortex
Presynaptic	10458	BAIAP2	24534009	WEINGARTEN_2014	2014	Brain
Presynaptic	10458	BAIAP2	24876496	WILHELM_2014	2014	Cerebellum
Postsynaptic	10458	BAIAP2	25211037	DISTLER_2014	2014	Hippocampus
Synaptosome	10458	BAIAP2	25211037	DISTLER_2014	2014	Hippocampus
Synaptosome	10458	BAIAP2	25352669	LIU_2014	2014	Prefrontal cortex
Postsynaptic	10458	BAIAP2	25429717	BAYES_2014	2014	Frontal lobe
Synaptosome	10458	BAIAP2	2.6e+07	CHANG_2015	2015	Hippocampus
Synaptosome	10458	BAIAP2	2.6e+07	CHANG_2015	2015	Motor cortex
Synaptosome	10458	BAIAP2	27115346	KOHANSAL_NODEHI_2016	2016	Cerebellum
Synaptosome	10458	BAIAP2	27115346	KOHANSAL_NODEHI_2016	2016	Cerebral cortex
Postsynaptic	10458	BAIAP2	27507650	LI_2016	2016	Hippocampus
Postsynaptic	10458	BAIAP2	27609886	UEZU_2016	2016	Cerebral cortex
Synaptosome	10458	BAIAP2	27748445	GONZALEZ_LOZANO_2016	2016	Cerebral cortex
Postsynaptic	10458	BAIAP2	27898073	FOCKING_2016	2016	Brain
Postsynaptic	10458	BAIAP2	28671696	LI_2017	2017	Hippocampus
Synaptosome	10458	BAIAP2	28713243	ALFIERI_2017	2017	Telencephalon
Postsynaptic	10458	BAIAP2	2.9e+07	FERNANDEZ_2017	2017	Forebrain
Postsynaptic	10458	BAIAP2	29203896	ROY_2017	2017	Frontal lobe
Postsynaptic	10458	BAIAP2	29203896	ROY_2017	2017	Occipital lobe
Postsynaptic	10458	BAIAP2	29203896	ROY_2017	2017	Temporal lobe
Postsynaptic	10458	BAIAP2	29203896	ROY_2017	2017	Parietal lobe
Synaptosome	10458	BAIAP2	29610302	HEO_2018	2018	Hippocampus
Synaptosome	10458	BAIAP2	29610302	HEO_2018	2018	Cerebral cortex
Postsynaptic	10458	BAIAP2	30071621	ROY_2018	2018	Cerebellum
Postsynaptic	10458	BAIAP2	30071621	ROY_2018	2018	Hypothalamus
Postsynaptic	10458	BAIAP2	30071621	ROY_2018	2018	Hippocampus
Postsynaptic	10458	BAIAP2	30071621	ROY_2018	2018	Striatum
Postsynaptic	10458	BAIAP2	30071621	ROY_2018	2018	Frontal lobe
Postsynaptic	10458	BAIAP2	3.1e+07	WILSON_2019	2019	Cerebral cortex
Postsynaptic	57554	LRRC7	10818142	WALIKONIS_2000	2000	Forebrain
Postsynaptic	57554	LRRC7	14532281	LI_2004	2004	Forebrain
Postsynaptic	57554	LRRC7	14720225	YOSHIMURA_2004	2004	Forebrain
Postsynaptic	57554	LRRC7	1.5e+07	PENG_2002	2004	Forebrain
Postsynaptic	57554	LRRC7	15169875	JORDAN_2004	2004	Brain
Postsynaptic	57554	LRRC7	15748150	TRINIDAD_2005	2005	Brain
Postsynaptic	57554	LRRC7	16507876	CHENG_2006	2006	Forebrain
Postsynaptic	57554	LRRC7	16507876	CHENG_2006	2006	Cerebellum
Postsynaptic	57554	LRRC7	16635246	COLLINS_2006	2006	Forebrain
Postsynaptic	57554	LRRC7	17623647	DOSEMESI_2007	2007	Cerebral cortex
Postsynaptic	57554	LRRC7	18056256	TRINIDAD_2008	2008	Midbrain
Postsynaptic	57554	LRRC7	18056256	TRINIDAD_2008	2008	Cerebellum
Postsynaptic	57554	LRRC7	18056256	TRINIDAD_2008	2008	Hippocampus
Synaptosome	57554	LRRC7	20309889	FILIOU_2010	2010	Brain
Postsynaptic	57554	LRRC7	21170055	BAYES_2011	2010	Cerebral cortex
Synaptosome	57554	LRRC7	21398567	DAHlHAUS_2011	2011	Visual cortex
Postsynaptic	57554	LRRC7	23071613	BAYES_2012	2012	Cerebral cortex

Localisation	HumanEntrez	HumanName	PaperPMID	Paper	Year	BrainRegion
Presynaptic	57554	LRRC7	24876496	WILHELM_2014	2014	Cerebellum
Postsynaptic	57554	LRRC7	25211037	DISTLER_2014	2014	Hippocampus
Synaptosome	57554	LRRC7	25211037	DISTLER_2014	2014	Hippocampus
Synaptosome	57554	LRRC7	25352669	LIU_2014	2014	Prefrontal cortex
Postsynaptic	57554	LRRC7	25429717	BAYES_2014	2014	Frontal lobe
Synaptosome	57554	LRRC7	27115346	KOHANSAL_NODEHI_2016	2016	Cerebellum
Synaptosome	57554	LRRC7	27115346	KOHANSAL_NODEHI_2016	2016	Cerebral cortex
Postsynaptic	57554	LRRC7	27507650	LI_2016	2016	Hippocampus
Postsynaptic	57554	LRRC7	27609886	UEZU_2016	2016	Cerebral cortex
Synaptosome	57554	LRRC7	27748445	GONZALEZ_LOZANO_2016	2016	Cerebral cortex
Postsynaptic	57554	LRRC7	28671696	LI_2017	2017	Hippocampus
Synaptosome	57554	LRRC7	28713243	ALFIERI_2017	2017	Telencephalon
Postsynaptic	57554	LRRC7	29203896	ROY_2017	2017	Frontal lobe
Postsynaptic	57554	LRRC7	29203896	ROY_2017	2017	Occipital lobe
Postsynaptic	57554	LRRC7	29203896	ROY_2017	2017	Temporal lobe
Postsynaptic	57554	LRRC7	29203896	ROY_2017	2017	Parietal lobe
Synaptosome	57554	LRRC7	29610302	HEO_2018	2018	Hippocampus
Synaptosome	57554	LRRC7	29610302	HEO_2018	2018	Cerebral cortex
Postsynaptic	57554	LRRC7	30071621	ROY_2018	2018	Cerebellum
Postsynaptic	57554	LRRC7	30071621	ROY_2018	2018	Hypothalamus
Postsynaptic	57554	LRRC7	30071621	ROY_2018	2018	Hippocampus
Postsynaptic	57554	LRRC7	30071621	ROY_2018	2018	Striatum
Postsynaptic	57554	LRRC7	30071621	ROY_2018	2018	Frontal lobe
Postsynaptic	57554	LRRC7	3.1e+07	WILSON_2019	2019	Cerebral cortex

gns %>% show_query()

```
## <SQL>
## SELECT `Localisation`, `HumanEntrez`, `HumanName`, `PaperPMID`, `Paper`, `Year`, `BrainRegion`
## FROM `FullGeneFullPaperFullRegion`
## WHERE (`HumanEntrez` IN ('4130', '10458', '57554'))
```

2.3 Get disease information for gene list

```
gns<-con %>% tbl("FullGeneFullDisease") %>% filter(HumanEntrez %in% c("4130","10458", "57554")) %>%
  select(HumanName, HDOID, Description)
gns %>% collect() %>% pander()
```

HumanName	HDOID	Description
MAP1A	DOID:936	brain_disease
MAP1A	DOID:331	central_nervous_system_disease
MAP1A	DOID:863	nervous_system_disease
MAP1A	DOID:5419	schizophrenia
MAP1A	DOID:2468	psychotic_disorder
MAP1A	DOID:1443	cerebral_degeneration
MAP1A	DOID:1561	cognitive_disorder
MAP1A	DOID:150	disease_of_mental_health
BAIAP2	DOID:150	disease_of_mental_health
LRRC7	DOID:150	disease_of_mental_health
MAP1A	DOID:7	disease_of_anatomical_entity
LRRC7	DOID:7	disease_of_anatomical_entity

HumanName	HDOID	Description
BAIAP2	DOID:0060037	developmental_disorder_of_mental_health
LRRC7	DOID:0060037	developmental_disorder_of_mental_health
LRRC7	DOID:0014667	disease_of_metabolism
BAIAP2	DOID:12849	autistic_disorder
BAIAP2	DOID:0060041	autism_spectrum_disorder
BAIAP2	DOID:0060040	pervasive_developmental_disorder
BAIAP2	DOID:0060038	specific_developmental_disorder
LRRC7	DOID:0060038	specific_developmental_disorder
LRRC7	DOID:1287	cardiovascular_system_disease
MAP1A	DOID:10907	microcephaly
LRRC7	DOID:2914	immune_system_disease
LRRC7	DOID:74	hematopoietic_system_disease
MAP1A	DOID:2490	congenital_nervous_system_abnormality
LRRC7	DOID:0060158	acquired_metabolic_disease
BAIAP2	DOID:1094	attention_deficit_hyperactivity_disorder
LRRC7	DOID:1094	attention_deficit_hyperactivity_disorder
MAP1A	DOID:0080015	physical_disorder
LRRC7	DOID:0080015	physical_disorder
LRRC7	DOID:178	vascular_disease
LRRC7	DOID:10763	hypertension
LRRC7	DOID:0050828	artery_disease
LRRC7	DOID:9500	leukocyte_disease
LRRC7	DOID:12987	agranulocytosis
LRRC7	DOID:374	nutrition_disease
LRRC7	DOID:654	overnutrition
LRRC7	DOID:615	leukopenia
LRRC7	DOID:9970	obesity
LRRC7	DOID:0050567	orofacial_cleft
MAP1A	all	all
BAIAP2	all	all
LRRC7	all	all
MAP1A	DOID:4	disease
BAIAP2	DOID:4	disease
LRRC7	DOID:4	disease

```r
gns %>% show_query()
```

```
## <SQL>
## SELECT `HumanName`, `HDOID`, `Description`
## FROM `FullGeneFullDisease`
## WHERE (`HumanEntrez` IN ('4130', '10458', '57554'))
```

```r
write.table(gns, file = "SomeDisease.txt", sep = "\t", row.names = F)
```

2.4 Get SynGO annotation for gene list

test

```r
gns<-con %>% tbl("FullGeneFullPaper") %>% filter(HumanEntrez %in% c("1742","10458", "57554")) %>%
  select(Localisation, HumanEntrez,HumanName, SynGO)
gns %>% collect() %>% unique() %>%
  mutate(SynGO = paste0("[",SynGO,"](https://syngoportal.org/gene.html?",SynGO,")")) %>%
```

```
pander()
```

Localisation	HumanEntrez	HumanName	SynGO
Postsynaptic	1742	DLG4	HGNC:2903
Synaptosome	1742	DLG4	HGNC:2903
Presynaptic	1742	DLG4	HGNC:2903
Postsynaptic	10458	BAIAP2	HGNC:947
Synaptosome	10458	BAIAP2	HGNC:947
Presynaptic	10458	BAIAP2	HGNC:947
Postsynaptic	57554	LRRC7	HGNC:18531
Synaptosome	57554	LRRC7	HGNC:18531
Presynaptic	57554	LRRC7	HGNC:18531

```
gns %>% show_query()
```

```
## <SQL>
## SELECT `Localisation`, `HumanEntrez`, `HumanName`, `SynGO`
## FROM `FullGeneFullPaper`
## WHERE (`HumanEntrez` IN ('1742', '10458', '57554'))
```

```
write.table(gns, file = "SomeSynGO.txt", sep = "\t", row.names = F)
```

2.5 Get number of papers for each gene

```
gns<-con %>% tbl("Gene") %>% filter( !is.na(MGI))
paps<- con %>% tbl("Paper")
gps<- con %>% tbl("PaperGene")
t<- gns %>%
  inner_join(gps,by=c("ID"="GeneID")) %>%
  inner_join(paps,by=c("PaperPMID"="PMID"))
```

```
t %>% show_query()
```

```
## <SQL>
## SELECT `ID`,`MGI`,`HumanEntrez`,`MouseEntrez`,`HumanName`,`MouseName`,`RatEntrez`,`RatName`,`S
## FROM (SELECT `ID`,`MGI`,`HumanEntrez`,`MouseEntrez`,`HumanName`,`MouseName`,`RatEntrez`,`RatNa
## FROM (SELECT *
## FROM `Gene`
## WHERE (NOT(((`MGI`) IS NULL)))) AS `LHS`
## INNER JOIN `PaperGene` AS `RHS`
## ON (`LHS`.`ID` = `RHS`.`GeneID`)
## ) AS `LHS`
## INNER JOIN `Paper` AS `RHS`
## ON (`LHS`.`PaperPMID` = `RHS`.`PMID`)
```

```
t %>% group_by(MGI,MouseName) %>%
  summarise(numP=n()) %>%
  arrange(desc(numP)) %>%
  head(25) %>%
  collect() %>%
  pander()
```

```
## `summarise()` has grouped output by 'MGI'. You can override using the `.groups` argument.
```

MGI	MouseName	numP
MGI:99441	Ckmt1	76
MGI:104560	Nsf	61
MGI:88256	Camk2a	61
MGI:96568	Ina	58
MGI:103020	Syn2	57
MGI:105384	Hspa8	57
MGI:107363	Stxbp1	57
MGI:95739	Glul	57
MGI:99667	Syt1	57
MGI:95781	Gnb1	56
MGI:107384	Dnm1	55
MGI:109618	Atp6v1b2	55
MGI:1277959	Dlg4	55
MGI:88107	Atp1a3	55
MGI:88437	Cnp	55
MGI:101921	Ap2a1	54
MGI:1919020	Ap2b1	54
MGI:2388633	Cltc	54
MGI:87994	Aldoa	54
MGI:98460	Syn1	54
MGI:1335094	07-Sep	53
MGI:98388	Sptbn1	53
MGI:1277955	Bsn	52
MGI:1353495	Slc25a4	52
MGI:1298405	Ap2m1	51

2.6 Make the full graph

```
gns<-con %>% tbl("Gene") %>% filter( !is.na(HumanEntrez)) %>% collect()
pps <- con %>% tbl("PPI") %>% select(A,B) %>%collect()
g <- graph_from_data_frame(pps, directed = FALSE, vertices = gns)
summary(g)
```

```
## IGRAPH 9492e18 UN-- 7949 97600 --
## + attr: name (v/c), MGI (v/c), HumanEntrez (v/n), MouseEntrez (v/n),
## | HumanName (v/c), MouseName (v/c), RatEntrez (v/n), RatName (v/c),
## | SynGO (v/c)
```

```
gr <- simplify(g)
summary(gr)
```

```
## IGRAPH 146a37d UN-- 7949 44517 --
## + attr: name (v/c), MGI (v/c), HumanEntrez (v/n), MouseEntrez (v/n),
## | HumanName (v/c), MouseName (v/c), RatEntrez (v/n), RatName (v/c),
## | SynGO (v/c)
```

2.7 Make graph from specific paper or localisation

```
gnp<-con %>% tbl("FullGeneFullPaper") %>% filter(PaperPMID == "10818142") %>% select(GeneID, HumanEntrez, H
#gnp<-con %>% tbl("FullGeneFullPaper") %>% filter(Localisation == "Presynaptic") %>% select(GeneID, HumanEntre
ppi <- con %>% tbl("PPI")
edges<- gnp %>%
```
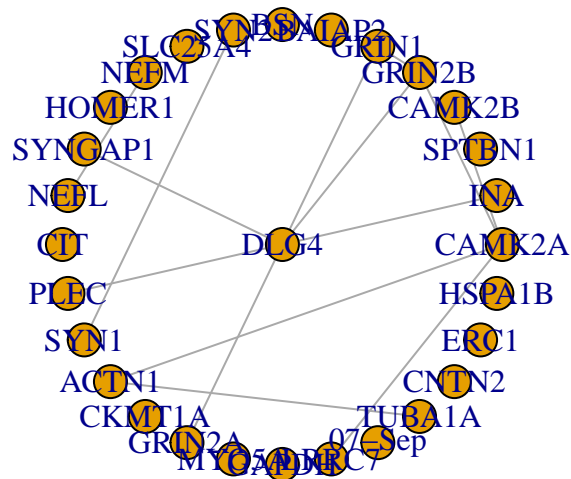
```r
  inner_join(ppi,by=c("GeneID"="A")) %>%
  inner_join(gnp,by=c("B"="GeneID")) %>% select(GeneID,B) %>% collect()
vert <- gnp %>% collect()

gp <- graph_from_data_frame(edges, directed = FALSE, vertices = vert)
gp <- simplify(gp)
summary(gp)
```

```
## IGRAPH 8014a37 UN-- 29 13 --
## + attr: name (v/c), HumanEntrez (v/n), HumanName (v/c)
```

```r
coords <- layout_(gp, as_star())
plot(gp, layout = coords, vertex.label = V(gp)$HumanName)
```


3 Appendix

3.1 Functions

```r
## Custom functions used in the analysis should go into this chunk.
## They will be listed in their own section of the appendix.
#' Function to create UpSetR compatible table from taxonomy table
#'
#' @param data taxonomy data
#' @param cname name of the column with taxon of interest
#' @param ename name of the column with entity names
#'
#' @return UpSetR compatible data.frame
prepareUpSet<-function(data,cname,ename='GeneID'){
data<-as.data.frame(data)
myFromList<-function (input)
{
  elements <- unique(unlist(input))
  data <- unlist(lapply(input, function(x) {
    x <- as.vector(match(elements, x))
  }))
  data[is.na(data)] <- as.integer(0)
  data[data != 0] <- as.integer(1)
```

9

```r
  data <- data.frame(matrix(data, ncol = length(input), byrow = F))
  rownames(data)<-elements
  data <- data[which(rowSums(data) != 0), ]
  names(data) <- names(input)
  return(data)
}
  if(any(is.na(data[,cname]))){
    data[is.na(data[,cname]),cname]<-'Unspecified'
  }
  sets<-unique(data[,cname])
  l<-lapply(sets, function(.x){
   as.character(data[data[,cname]==.x,ename])
  })
  names(l)<-sets
  pt<-myFromList(l)
}
```

3.2 Setup R

```r
## This chunk should contain global configuration commands.
## Use this to set knitr options and related things. Everything
## in this chunk will be included in an appendix to document the
## configuration used.
#output <- opts_knit$get("rmarkdown.pandoc.to")
opts_knit$set(stop_on_error = 2L)


## Cache options
opts_chunk$set(cache=FALSE)


## Set 'hide.fig.code' to FALSE to include code chunks that
## produce Figures in the output. Note that this affects all chunks
## that provide a figure caption.
opts_chunk$set(hold=TRUE, hide.fig.code=FALSE)

## Pander options
panderOptions("digits", 3)
panderOptions("table.split.table", 160)
```

3.3 Versions

3.3.1 Session Info

Error in get(genname, envir = envir) : object 'testthat_print' not found

version	R version 3.6.1 (2019-07-05)
os	macOS Catalina 10.15.7
system	x86_64, darwin15.6.0
ui	X11
language	(EN)
collate	en_GB.UTF-8
ctype	en_GB.UTF-8

	tz	Europe/Moscow
	date	2021-04-08

	package	ondiskversion	loadedversion	attached	is_base	date	source
assertthat	assertthat	0.2.1	0.2.1	FALSE	FALSE	2019-03-21	CRAN (R 3.6.0)
backports	backports	1.2.1	1.2.1	FALSE	FALSE	2020-12-09	CRAN (R 3.6.2)
bit	bit	4.0.4	4.0.4	FALSE	FALSE	2020-08-04	CRAN (R 3.6.2)
bit64	bit64	4.0.2	4.0.2	FALSE	FALSE	2020-07-30	CRAN (R 3.6.2)
blob	blob	1.2.1	1.2.1	FALSE	FALSE	2020-01-20	CRAN (R 3.6.0)
callr	callr	3.4.3	3.4.3	FALSE	FALSE	2020-03-28	CRAN (R 3.6.2)
cli	cli	2.4.0	2.4.0	FALSE	FALSE	2021-04-05	CRAN (R 3.6.1)
codetools	codetools	0.2.16	0.2-16	FALSE	FALSE	2018-12-24	CRAN (R 3.6.1)
colorspace	colorspace	1.4.1	1.4-1	FALSE	FALSE	2019-03-18	CRAN (R 3.6.0)
crayon	crayon	1.4.1	1.4.1	FALSE	FALSE	2021-02-08	CRAN (R 3.6.2)
DBI	DBI	1.1.1	1.1.1	FALSE	FALSE	2021-01-15	CRAN (R 3.6.2)
dbplyr	dbplyr	2.1.0	2.1.0	FALSE	FALSE	2021-02-03	CRAN (R 3.6.2)
debugme	debugme	1.1.0	1.1.0	FALSE	FALSE	2017-10-22	CRAN (R 3.6.0)
desc	desc	1.2.0	1.2.0	FALSE	FALSE	2018-05-01	CRAN (R 3.6.0)
devtools	devtools	2.3.1	2.3.1	FALSE	FALSE	2020-07-21	CRAN (R 3.6.2)
digest	digest	0.6.27	0.6.27	FALSE	FALSE	2020-10-24	CRAN (R 3.6.2)
dplyr	dplyr	1.0.5	1.0.5	TRUE	FALSE	2021-03-05	CRAN (R 3.6.2)
ellipsis	ellipsis	0.3.1	0.3.1	FALSE	FALSE	2020-05-15	CRAN (R 3.6.2)
evaluate	evaluate	0.14	0.14	FALSE	FALSE	2019-05-28	CRAN (R 3.6.0)
fansi	fansi	0.4.2	0.4.2	FALSE	FALSE	2021-01-15	CRAN (R 3.6.2)
fs	fs	1.5.0	1.5.0	FALSE	FALSE	2020-07-31	CRAN (R 3.6.2)
generics	generics	0.1.0	0.1.0	FALSE	FALSE	2020-10-31	CRAN (R 3.6.2)
ggplot2	ggplot2	3.3.2	3.3.2	TRUE	FALSE	2020-06-19	CRAN (R 3.6.2)
glue	glue	1.4.2	1.4.2	FALSE	FALSE	2020-08-27	CRAN (R 3.6.2)

	package	ondiskversion	loadedversion	attached	is_base	date	source
gridExtra	gridExtra	2.3	2.3	FALSE	FALSE	2017-09-09	CRAN (R 3.6.0)
gtable	gtable	0.3.0	0.3.0	FALSE	FALSE	2019-03-25	CRAN (R 3.6.0)
htmltools	htmltools	0.5.0	0.5.0	FALSE	FALSE	2020-06-16	CRAN (R 3.6.2)
igraph	igraph	1.2.5	1.2.5	TRUE	FALSE	2020-03-19	CRAN (R 3.6.0)
knitr	knitr	1.29	1.29	TRUE	FALSE	2020-06-23	CRAN (R 3.6.2)
lifecycle	lifecycle	1.0.0	1.0.0	FALSE	FALSE	2021-02-15	CRAN (R 3.6.2)
magrittr	magrittr	2.0.1	2.0.1	FALSE	FALSE	2020-11-17	CRAN (R 3.6.2)
memoise	memoise	1.1.0	1.1.0	FALSE	FALSE	2017-04-21	CRAN (R 3.6.0)
munsell	munsell	0.5.0	0.5.0	FALSE	FALSE	2018-06-12	CRAN (R 3.6.0)
pander	pander	0.6.3	0.6.3	TRUE	FALSE	2018-11-06	CRAN (R 3.6.0)
pillar	pillar	1.5.1	1.5.1	FALSE	FALSE	2021-03-05	CRAN (R 3.6.2)
pkgbuild	pkgbuild	1.1.0	1.1.0	FALSE	FALSE	2020-07-13	CRAN (R 3.6.2)
pkgconfig	pkgconfig	2.0.3	2.0.3	FALSE	FALSE	2019-09-22	CRAN (R 3.6.0)
pkgload	pkgload	1.1.0	1.1.0	FALSE	FALSE	2020-05-29	CRAN (R 3.6.2)
plyr	plyr	1.8.6	1.8.6	FALSE	FALSE	2020-03-03	CRAN (R 3.6.0)
prettyunits	prettyunits	1.1.1	1.1.1	FALSE	FALSE	2020-01-24	CRAN (R 3.6.0)
processx	processx	3.4.3	3.4.3	FALSE	FALSE	2020-07-05	CRAN (R 3.6.2)
ps	ps	1.3.4	1.3.4	FALSE	FALSE	2020-08-11	CRAN (R 3.6.2)
purrr	purrr	0.3.4	0.3.4	FALSE	FALSE	2020-04-17	CRAN (R 3.6.2)
R6	R6	2.5.0	2.5.0	FALSE	FALSE	2020-10-28	CRAN (R 3.6.2)
Rcpp	Rcpp	1.0.5	1.0.5	FALSE	FALSE	2020-07-06	CRAN (R 3.6.2)
remotes	remotes	2.2.0	2.2.0	FALSE	FALSE	2020-07-21	CRAN (R 3.6.2)
rlang	rlang	0.4.10	0.4.10	FALSE	FALSE	2020-12-30	CRAN (R 3.6.2)
rmarkdown	rmarkdown	2.3	2.3	FALSE	FALSE	2020-06-18	CRAN (R 3.6.2)
rprojroot	rprojroot	1.3.2	1.3-2	FALSE	FALSE	2018-01-03	CRAN (R 3.6.0)
RSQLite	RSQLite	2.2.0	2.2.0	FALSE	FALSE	2020-01-07	CRAN (R 3.6.0)

	package	ondiskversion	loadedversion	attached	is_base	date	source
scales	scales	1.1.1	1.1.1	FALSE	FALSE	2020-05-11	CRAN (R 3.6.2)
sessioninfo	sessioninfo	1.1.1	1.1.1	FALSE	FALSE	2018-11-05	CRAN (R 3.6.0)
stringi	stringi	1.4.6	1.4.6	FALSE	FALSE	2020-02-17	CRAN (R 3.6.0)
stringr	stringr	1.4.0	1.4.0	FALSE	FALSE	2019-02-10	CRAN (R 3.6.0)
testthat	testthat	2.3.2	2.3.2	FALSE	FALSE	2020-03-02	CRAN (R 3.6.0)
tibble	tibble	3.1.0	3.1.0	FALSE	FALSE	2021-02-25	CRAN (R 3.6.2)
tidyselect	tidyselect	1.1.0	1.1.0	FALSE	FALSE	2020-05-11	CRAN (R 3.6.2)
UpSetR	UpSetR	1.4.0	1.4.0	TRUE	FALSE	2019-05-22	CRAN (R 3.6.0)
usethis	usethis	1.6.1	1.6.1	FALSE	FALSE	2020-04-29	CRAN (R 3.6.2)
utf8	utf8	1.2.1	1.2.1	FALSE	FALSE	2021-03-12	CRAN (R 3.6.2)
vctrs	vctrs	0.3.7	0.3.7	FALSE	FALSE	2021-03-29	CRAN (R 3.6.2)
withr	withr	2.4.1	2.4.1	FALSE	FALSE	2021-01-26	CRAN (R 3.6.2)
xfun	xfun	0.16	0.16	FALSE	FALSE	2020-07-24	CRAN (R 3.6.2)
yaml	yaml	2.2.1	2.2.1	FALSE	FALSE	2020-02-01	CRAN (R 3.6.0)