

Exploring the Hemibrain

Chelsea, Alec, Matt, Jing, Jamie

Introduction



What is our Dataset?

- Hemibrain of an adult female virgin *drosophila melanogaster* (fruit fly); excludes majority of left side brain and visual regions
 - Includes Electron Microscopy data and a segmentation layer for 3D neuron reconstruction

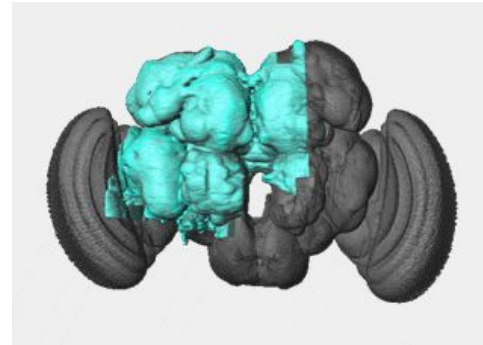
Why did we chose this Data Set?

- Well documented, API access and a wealth of data to explore
- No central research question that we came up with since no one is a fly scientist, more exploratory analysis
- Our team is made up of non biology and biology members!

Largest traced brain data set to date

What can this data set provide us as Data Scientists?

- Great for new DS
- Updated data will continue to be added as data set grows
 - VNC- Ventral Nerve Cord currently being worked on (larger than the hemibrain)



Hemibrain highlighted in cyan



Background - Basic neuron knowledge

Relationship between two neurons: $A \rightarrow B$

- A sends information to B

Invertebrate Neurons differ from vertebrates: Soma migrates to outside of brain and does not receive information since it has no dendrites on it

Terminology

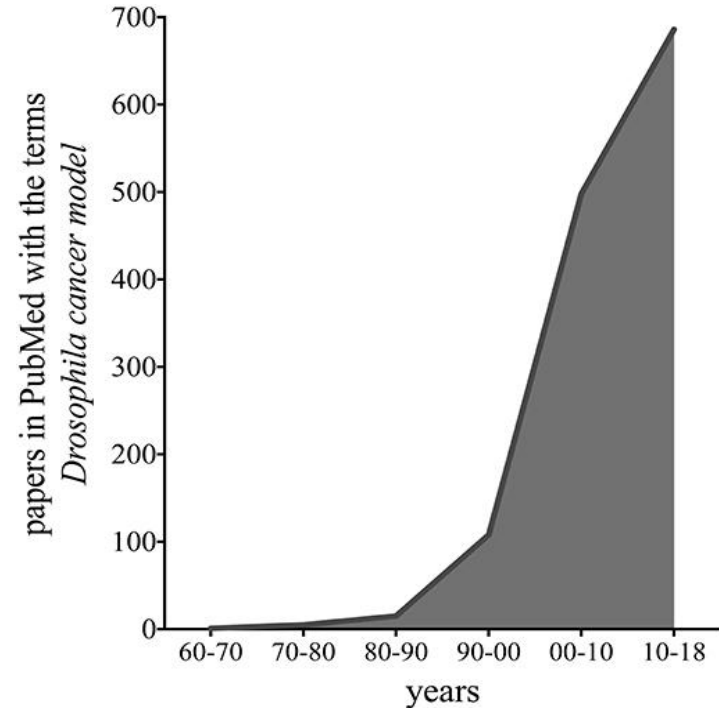
- Please refer to handout!

Important Terms

- **Connectomics** - the production and study of comprehensive maps of connections within an organism's nervous system.
- **ROI/primary ROI** - Region of Interest. Areas within the brain that are grouped together due to the shared function/which types of call pass through. Primary ROI will hold smaller sub ROIs
- **T-Bar/Pre** - Location on neuron that send out neurotransmitters into synaptic cleft
- **PSD/Post** - Location on neuron where receptors have absorbed neurotransmitters.
Postsynaptic Density
- **Morphology** - General physical structure of a neurons body, related to function of neuron
- **Antennal lobe** - Region of Interest that receives olfactory inputs from olfactory sensory neurons
- **MBONs** - Mushroom Body Output Neurons

Some Sanity Checks before we get started:

- Do our queries match the literature?
- Growing field and unfamiliar dataset
- Cancer researchers have better marketing
- ❖ AL -- MBONS- -> Mushroom body
- ❖ APL regulation of memory
- ❖ Ring neuron innervation is Mutually exclusive



Part 1: NeuPrint Data Summary

Summary Statistics - comparing two versions of the data set

Oldest Dataset = V.1.0.1

Latest Dataset = V.1

General Notes:

- Number of ROIs increased from oldest to latest dataset
- Roughly 6 months of work conducted between updates

Approach:

- Compare descriptive statistics across the two versions
- Focus on completion statistics

Summary Statistics - comparing two versions of the data set

Dataset Version	ROI Min	Percent Completeness Min	ROI Max	Percent Completeness Max
V1.1- Latest	FLA(R)	18.794567%	EBr3am	93.917381%
V1.0.1 - Oldest	EPA(L)	20.558988%	EBr3am	94.22657%

- No change in most complete ROI between data sets but change in lowest completed ROI
- General decrease in mean counts of post/pre synaptic counts between data sets which can be attributed to the addition of more ROIs in the latest update
- The ratio of Post: Pre synaptics sites between datasets remained almost the same at ~6.65
 - Almost always have more postsynaptic sites than presynaptic sites.

Difference in mean percent postsynaptic sites completed between data sets: 3.22%

Difference in mean percent presynaptic sites completed between data sets: 0.028%

Table comparing descriptive stats on each data set

Latest Dataset				
	p_presyn	t_pre	p_postsyn	t_post
count	229.000000	2.290000e+02	229.000000	2.290000e+02
mean	93.442736	9.165484e+04	60.196853	6.044246e+05
std	3.332198	2.164123e+05	20.172745	1.555558e+06
min	81.192053	5.100000e+01	20.558988	1.950000e+02
25%	91.983696	5.226000e+03	44.004432	2.437500e+04
50%	93.942688	1.252800e+04	63.723538	5.465600e+04
75%	95.710166	6.143800e+04	74.094809	3.371190e+05
max	99.854581	1.861182e+06	94.226570	1.356352e+07
Oldest Dataset				
	p_presyn	t_pre	p_postsyn	t_post
count	150.000000	1.500000e+02	150.000000	1.500000e+02
mean	93.415063	1.343923e+05	56.973805	8.942202e+05
std	3.603109	2.573065e+05	24.110162	1.859341e+06
min	81.575311	4.040000e+02	18.794567	2.403000e+03
25%	91.769063	5.898000e+03	32.603004	2.993750e+04
50%	93.964872	3.668650e+04	58.965348	1.664785e+05
75%	96.042271	1.325665e+05	81.910875	8.148918e+05
max	99.854581	1.861218e+06	93.917381	1.357282e+07

Dataset and Clean-up

Neuron dataset was acquired by querying NeuPrint database.

```
QUERY = " MATCH (n :Neuron) RETURN n "
```

Returned results: a dataset of Dictionary structure.

```
len(results) = 186,649
```

Convert Dictionary to Dataframe

The initial dataframe dimension: 186,649 X 245 columns, ROI matrix(230 columns)

Remove Left-side brain neurons. Based on ROIs matrix

Right Brain dataset: 123,494 rows

Remove un-traced neurons (including orphans, leaves, segments, etc)

Final dataset: 20,026 rows:

Summary Statistics

- There are 20,026 completed neurons on the right side of the hemibrain
- The average number of ROIs a neuron passes through is: 10
 - * Body ID 1418618235 passes through 117 ROIs
 - * Body ID 356131764 passes through 1 ROI

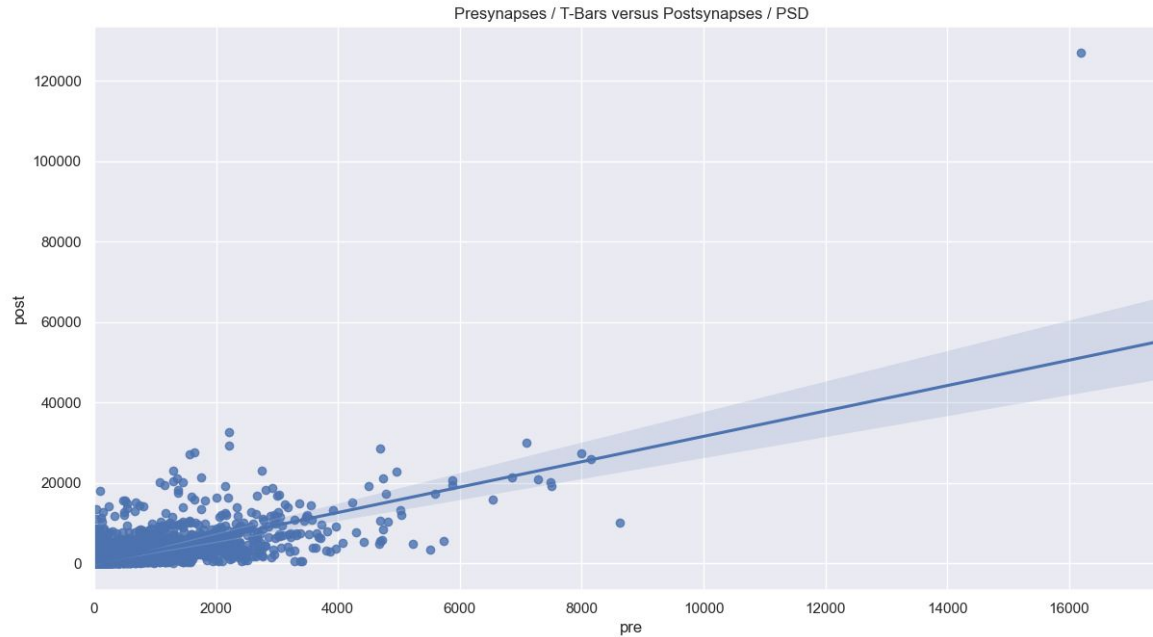
Summary Statistics

- There are 226 unique ROIs in the right side of the hemibrain
- The average number of ROIs passed through per neuron is: 918
 - *VLNP(R) has 11,314 neurons pass through

Summary Statistics

- The average number of presynapses/T-Bars is 281
 - min: 0
 - max: 17,628
- The average number of postsynapse/PSD is 857
 - min: 6
 - max: 127,148

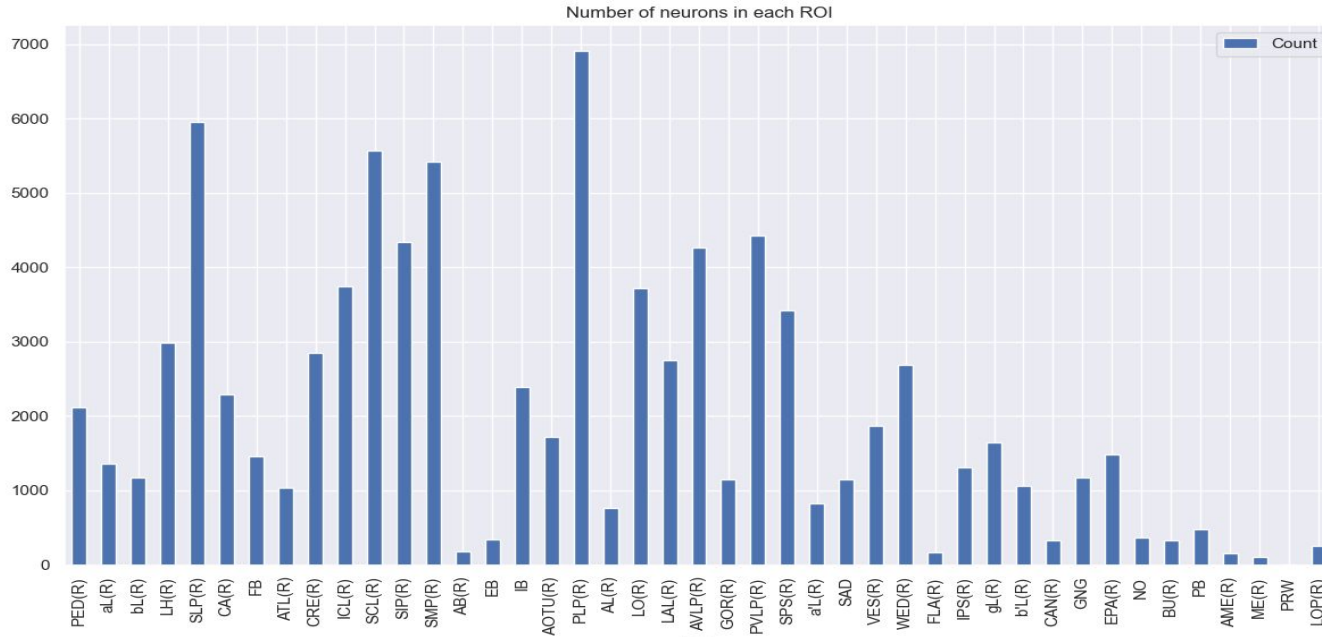
Presynapses/T-Bars VS. Postsynapses/PSD



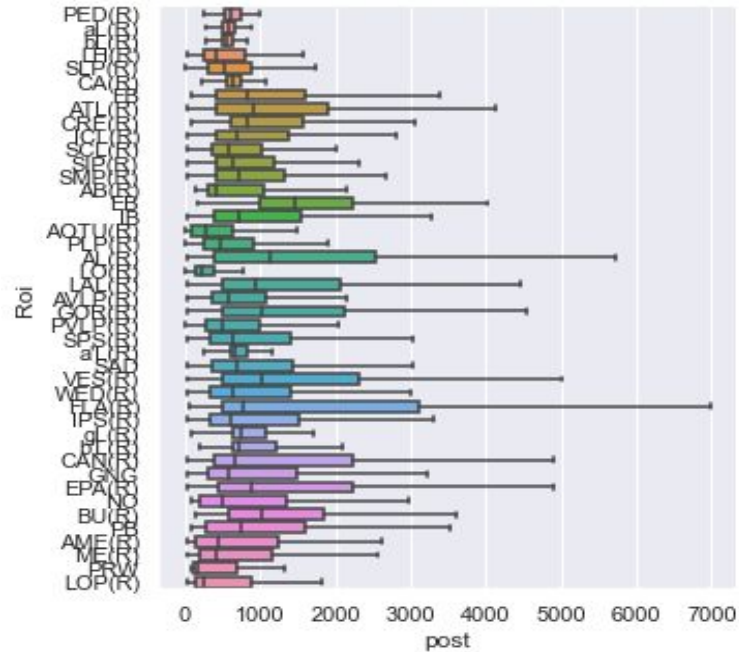
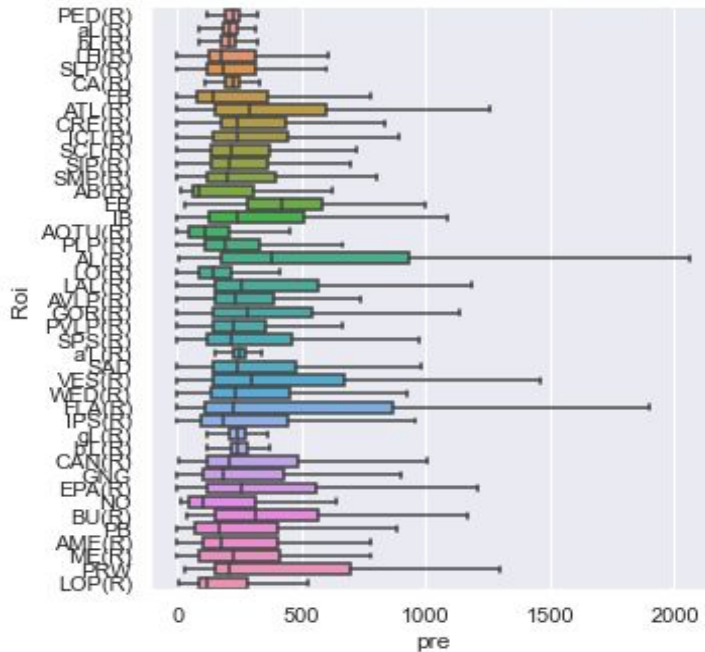
Summary Statistics - MBONs

- There are 64 MBONs on the right side of the hemibrain
- The average number of ROIs for MBONs is 23, while the average number of ROIs for all right side neurons is 10
- The average number of presynapses is 890 (>281)
- The average number of postsynapses is 10,307(>857)

Neuron counts in primary ROIs (right-side)



Pre and post synapses in primary ROIs



Connectivity of Primary ROIs

Relationship between neurons:

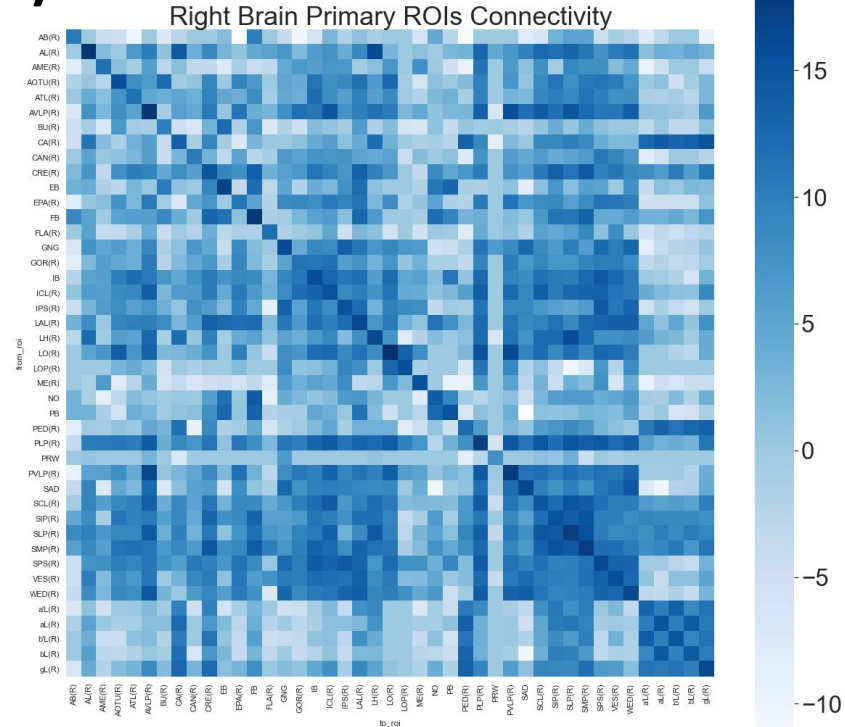
```
fetch_roi_connectivity(format='pandas')
```

pre-computed connectivity statistics between primary ROIs in the dataset

```
In [3]: RightBrain.head(20)
```

```
Out [3]:
```

	from_roi	to_roi	count	weight
25	AB(R)	AB(R)	112	10.591788
26	AB(R)	AL(R)	1	0.059706
27	AB(R)	AME(R)	1	-4.564785
28	AB(R)	AOTU(R)	1	-5.142958
30	AB(R)	ATL(R)	1	-10.221587
32	AB(R)	BU(R)	3	-7.226553
33	AB(R)	CA(R)	1	-2.564785
35	AB(R)	CRE(R)	11	-0.014807
36	AB(R)	EB	18	1.807405
37	AB(R)	EPA(R)	1	-11.647908



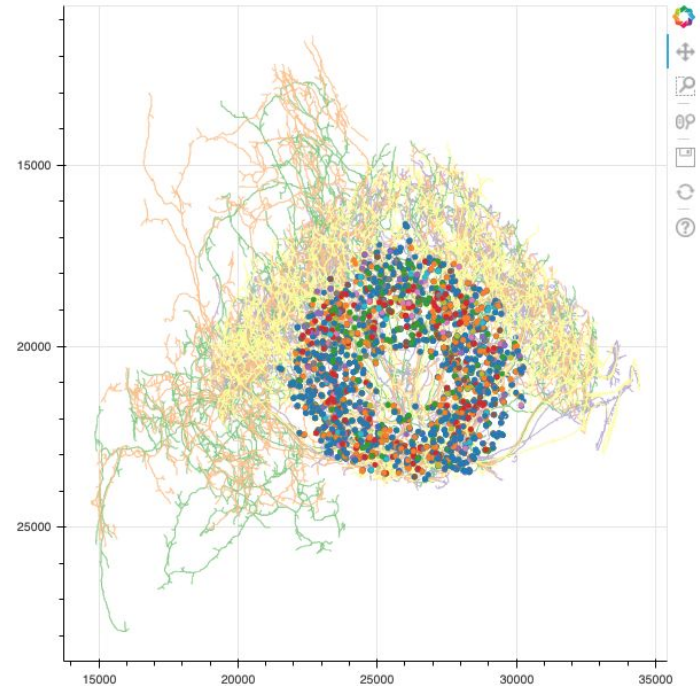
Skeleton Graph

- Visualize the 3D physical location of neurons and synapses in the brain in a 2D plot

Neuron/Synapse Criteria:

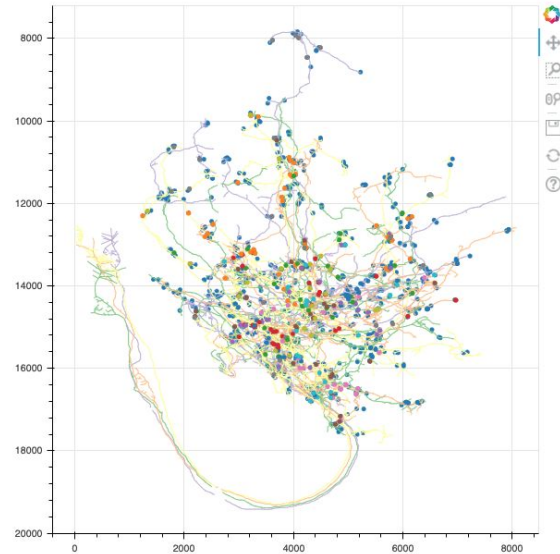
ROIs: EB

Cell Type: FB4Y

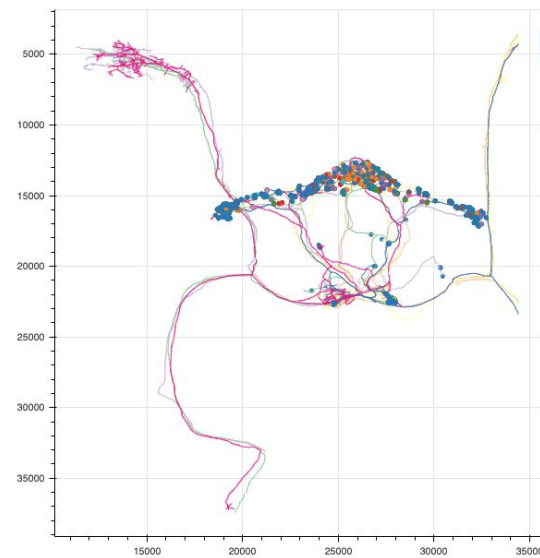


Visually trace connections in the brain

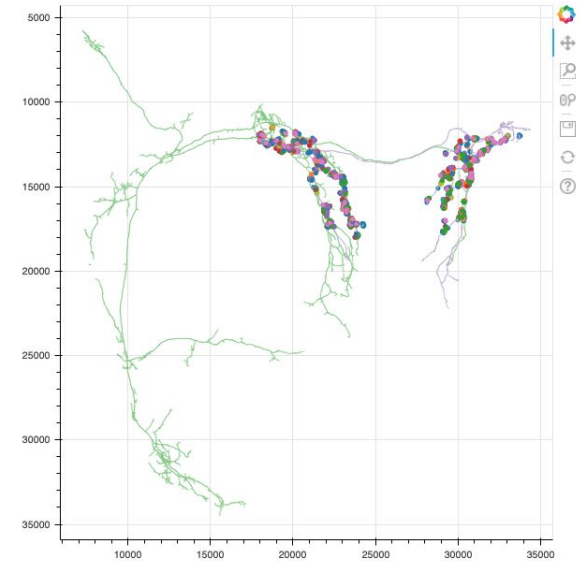
Neuron/Synapse Criteria:
ROIs: LH(R)
Cell Type: LHAV4g2



Neuron/Synapse Criteria:
ROIs: FB
Cell Type: SAF



Neuron/Synapse Criteria:
ROIs: ATL(R), ATL(L)
Cell Type: ATL014



Our Approach

Created a SkeletonGraph class that accepts a “cell type” string and a list of ROIs as parameters

Automatically generates and returns a scatter plot of the tbar synapse connections, and the associated neurons overlaid on a skeleton graph plot of the neurons.

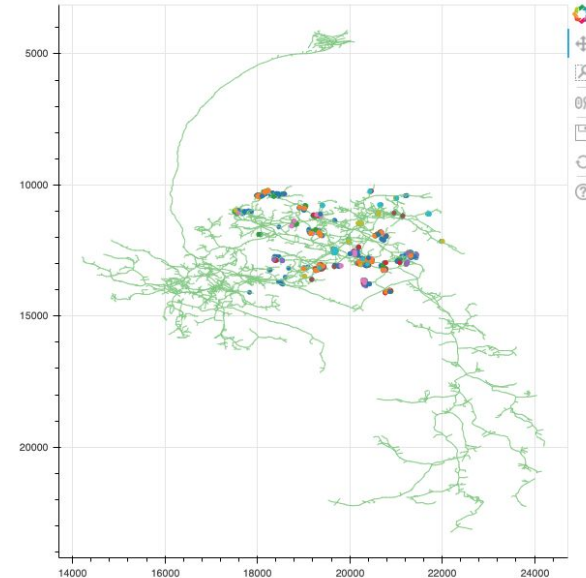
Heavily Utilized the neuprint-Python API

- **fetch_synapses** - returns the set of synapses that meets the input criteria
- **fetch_synapse_connections** - returns the related synapse connection points
- **fetch_neurons** - returns neuron information for associated neurons
- **merge_neuron_properties** - merges the neuron coordinate information with the connections data frame to group connection points by color
- **fetch_skeleton** - returns neuron skeleton coordinate information as a DataFrame

Extending Our Class: Cell Type of the Day

- Pulls the current day's 'Cell Type of the Day' information via the `fetch_daily_type` API call
- Creates an instance of the Skeleton Visualization class that plots the skeleton graph for all neurons of that cell type.

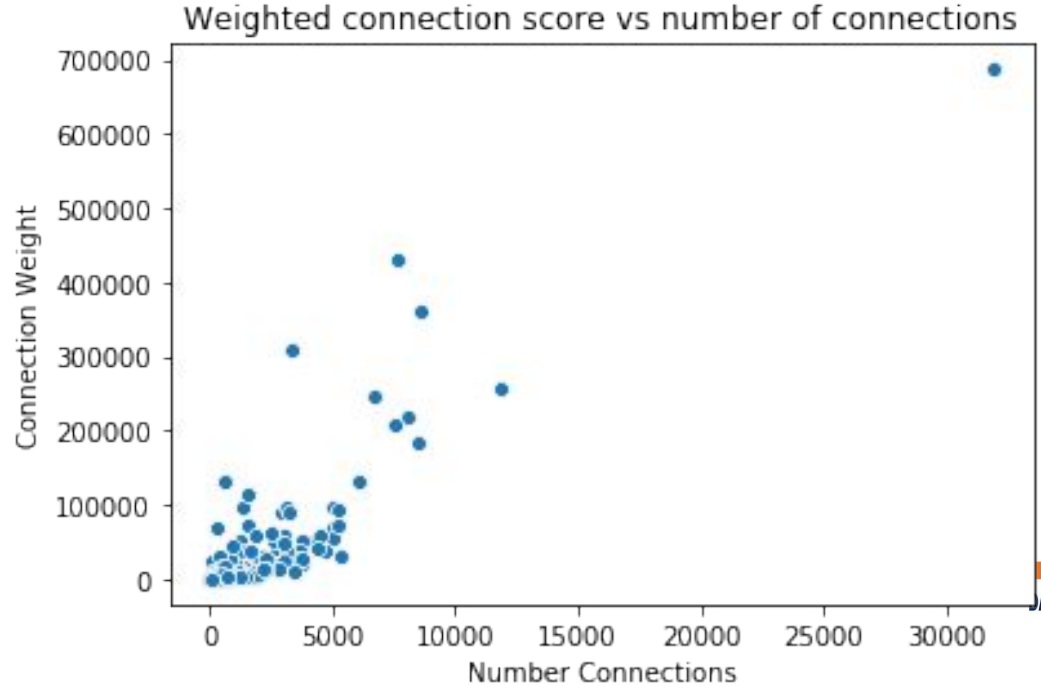
Neuron/Synapse Criteria:
ROIs: SMP(R), SMP(L)
Cell Type: SMP441



Part 2: Dataset Analysis

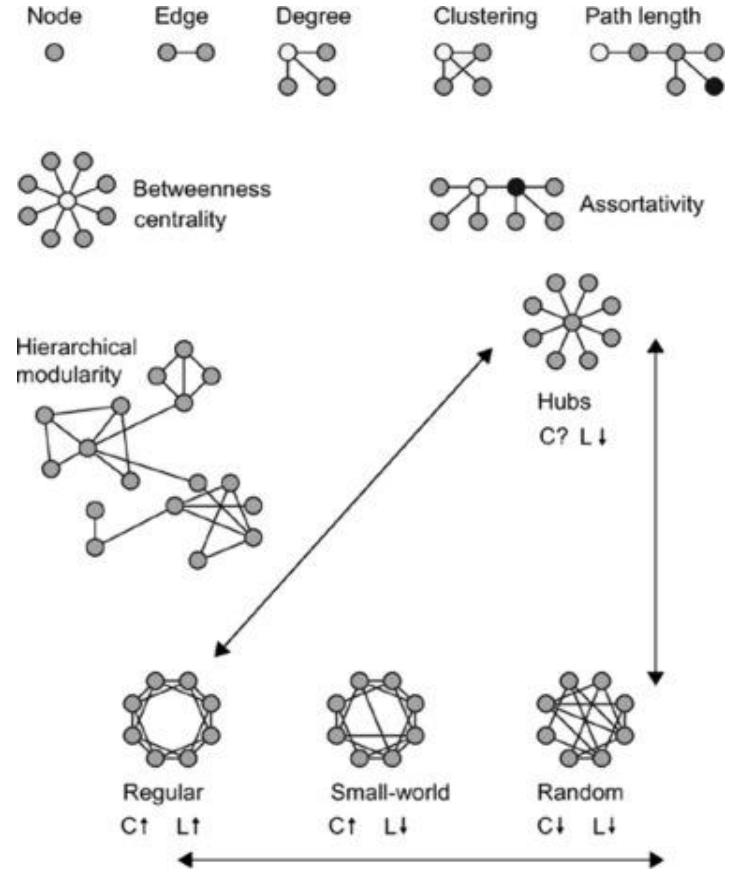
Is Weight per connection constant over ROIS?

- No!
- Some ROIs have higher weight per connection
- Unknown how much of this is structural vs sample bias



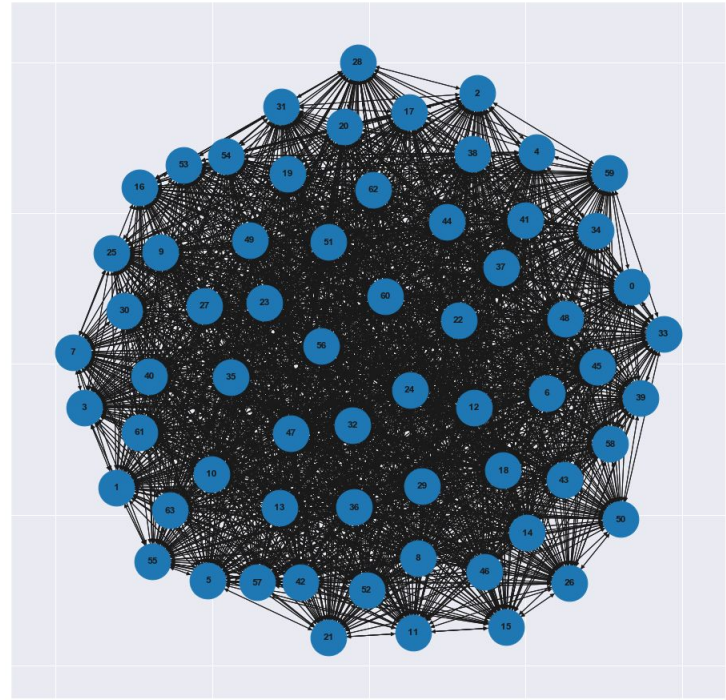
Introducing Graphs

- For us:
- Node = brain region
- Edge = Neurons connecting two Brain regions
- Want to see evidence of Hierarchy
- What is a spring representation?



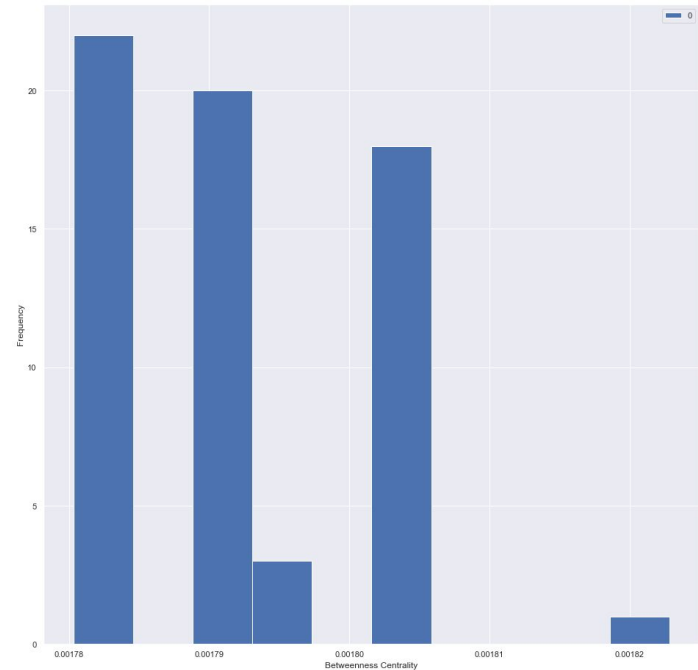
An introduction to graphs (Continued):

- This is an erdos-enyi random graph
- The null with which we compare our Brain
- This visualization is a spring plot (closely connected nodes are attracted)
- Each dot is a node and each line is an Edge (connection)



Betweenness Centrality in the random case

- Narrow distribution across a range of Very small values
- Do we see something different in the Brain?

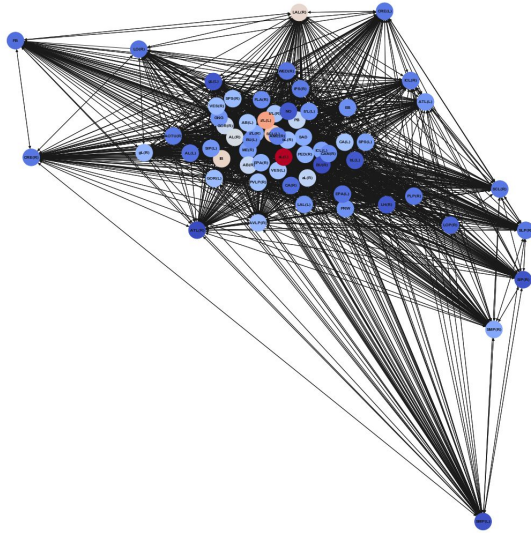


Factors considered in graph exploration

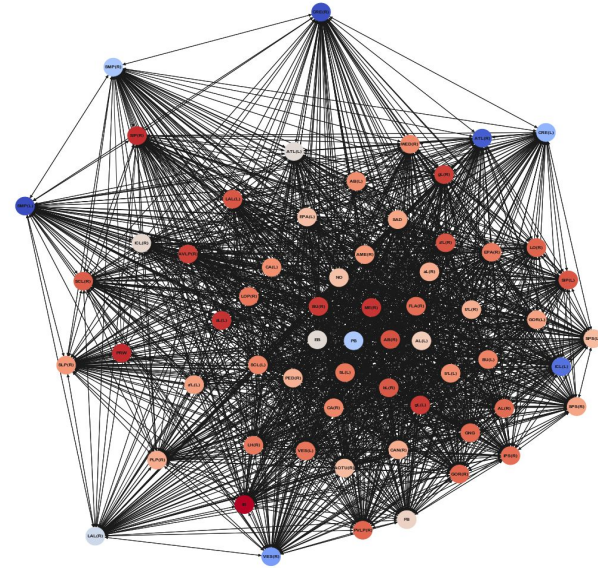
- Do weight and number of connections result in different conclusions
- Is there a single 'most influential' brain region as measured by eigen centrality
- Is there a defined information flow path as measured by betweenness centrality
- Which brain regions have the biggest reach?
- Can we define putative areas for study?

Number of Connections vs Weight

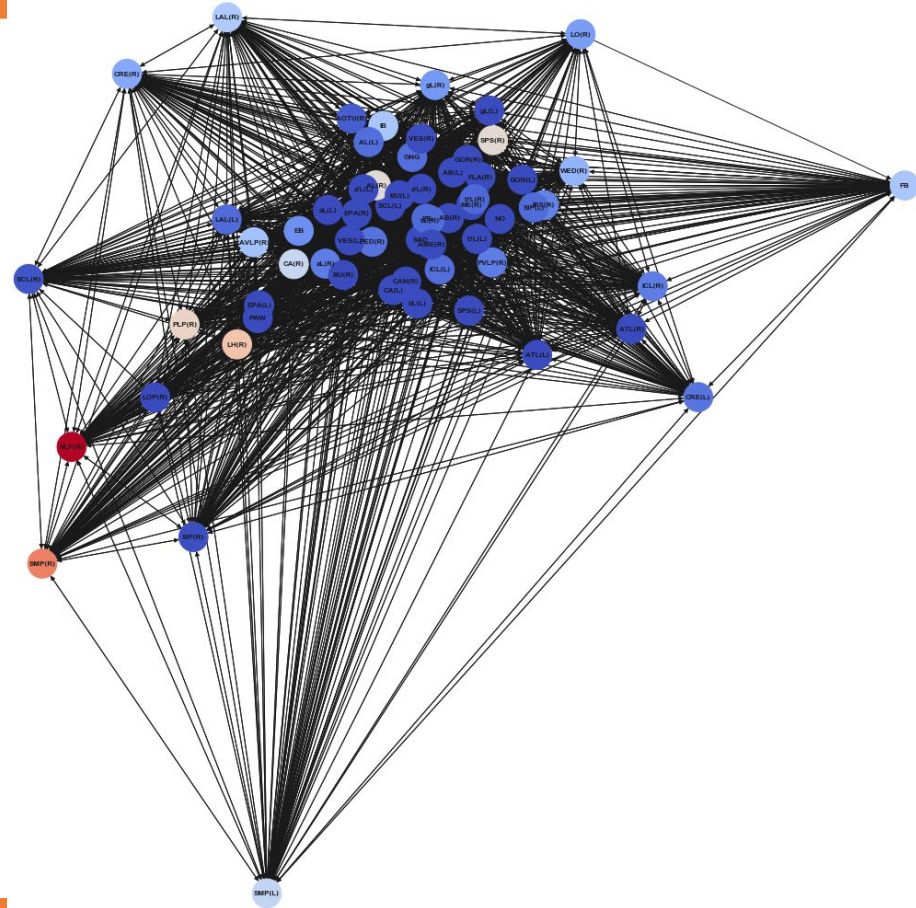
Local Reach by weight



Local Reach by connectivity



- We define a pathway of interest that passes mostly through the protocerebrum
- This circuit is over-represented in terms of involvement in shortest paths between brain regions



Betweenness for Weight

-Protocerebrum



SLP(R) 0.395293

-Sensory input /
learning



SMP(R) 0.318350



LH(R) 0.246959

-Motor Control



PLP(R) 0.222105



SPS(R) 0.207033



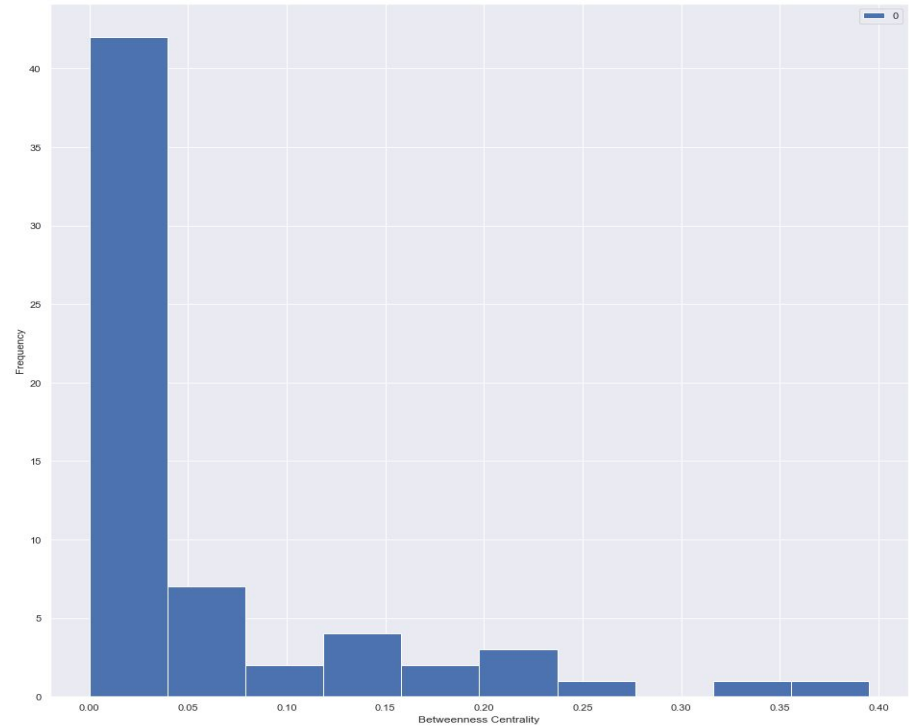
AL(R) 0.205711

CA(R) 0.166843

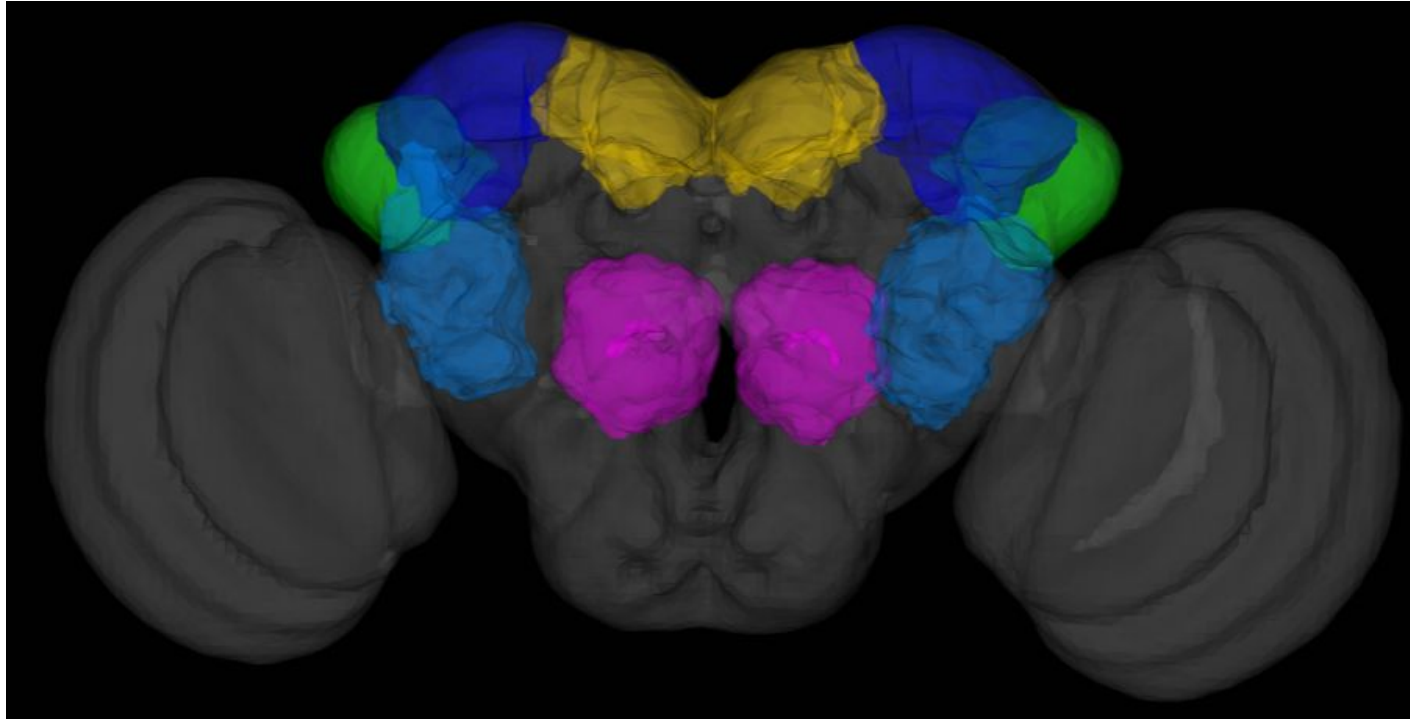
SMP(L) 0.161555

LAL(R) 0.137493

FB 0.132470

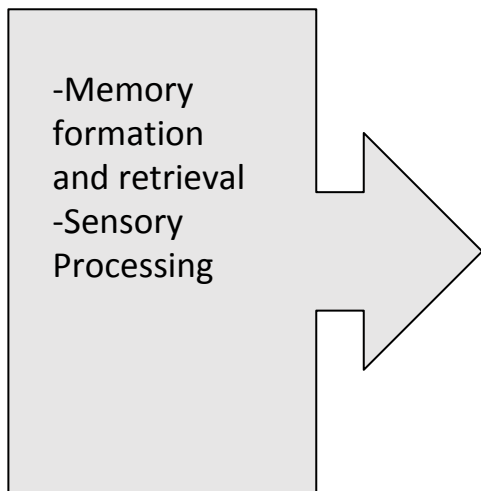


What does this look like in the brain?

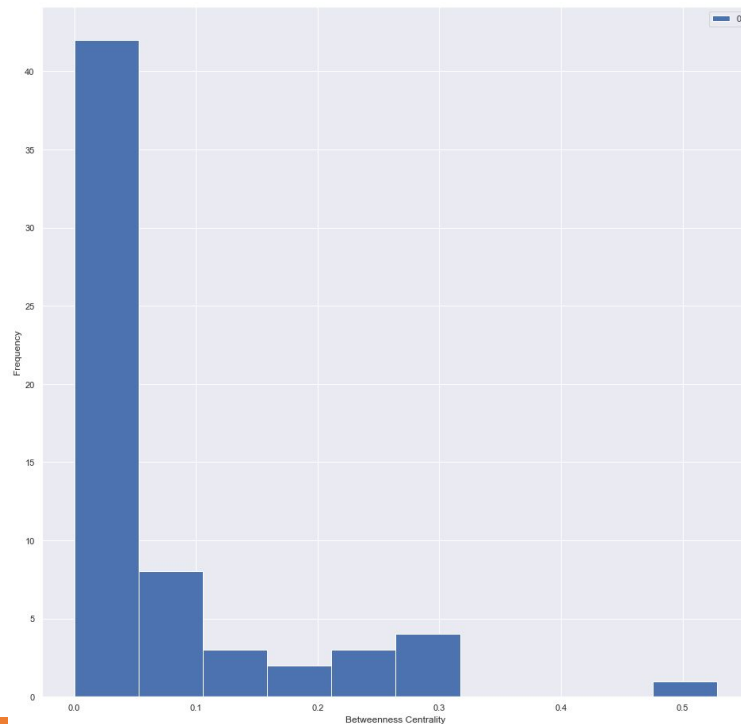


Eigencentrality for Weight

Poorly Defined



SAD	0.528454
a'L(R)	0.315810
gL(R)	0.300056
AB(L)	0.296509
NO	0.295431
EPA(L)	0.259042
aL(R)	0.236455
b'L(L)	0.220793
GOR(L)	0.204093
PB	0.162846

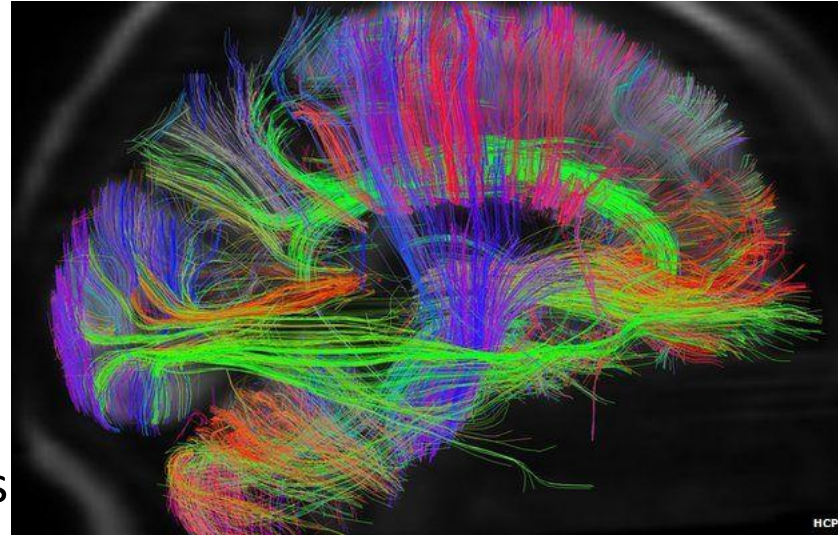


So what are these things we found?

- local reaching: memory
- eigencentality and betweenness centrality:
 - “Terra incognita” -Janelia
 - Possible centers for the control of learning and complex social behavior
 - Poorly characterized and under increasing recent scrutiny

How in the heck is this useful?

- Our dataset: ~20,000 neurons
- The drosophila brain: ~10X more
- The human brain: ~100 billion neurons
- Eventual goal: the human connectome
- Utility of an anatomical approach
- Need to develop single neuron methods



Takeaway

- Drosophila are behaviorally complex creatures with brains the size of sand
- This structural data can be used by researchers to identify experimental targets
- Deeper understanding of the fruit fly highlights the challenges we face in obtaining deeper understanding of ourselves

End of presentation material:

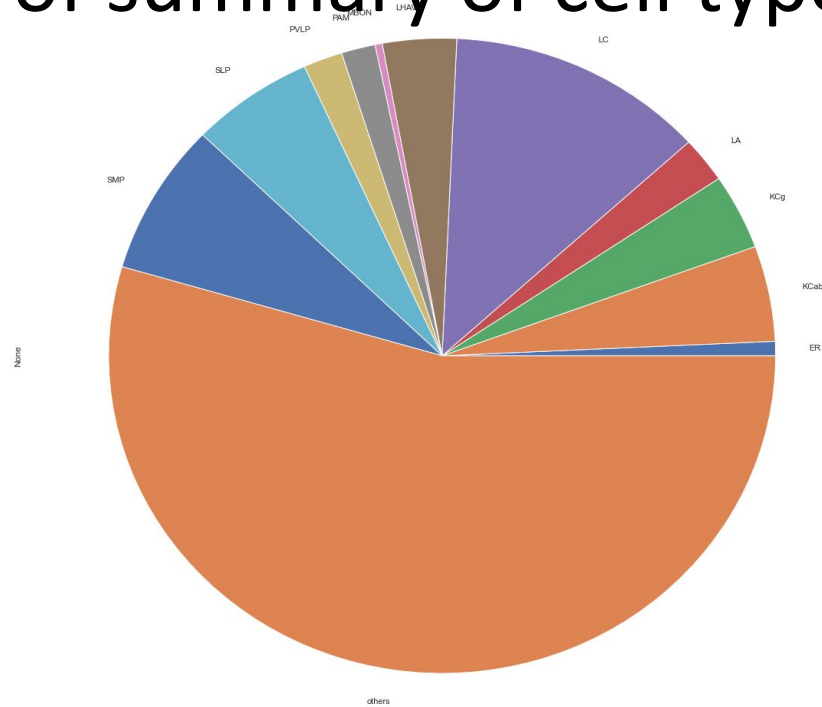
Hi peer reviewers and instructors! Thank you for making it through the presentation.

All following slides did not make the cut for the final presentation and are simply here for your interest if you so desire.

Limitations of this dataset

- Incomplete: about 5-10% of the drosophila brain
- Not representative: female virgin drosophila
- massive (requires substantial cloud based resources for some queries)
(but Janelia provides them for free!)
- Longest completed query time: ~17 hours
- Graph calculations

Pie chart of summary of cell types



Neuron size statistics in primary ROIs

