neurodata.io

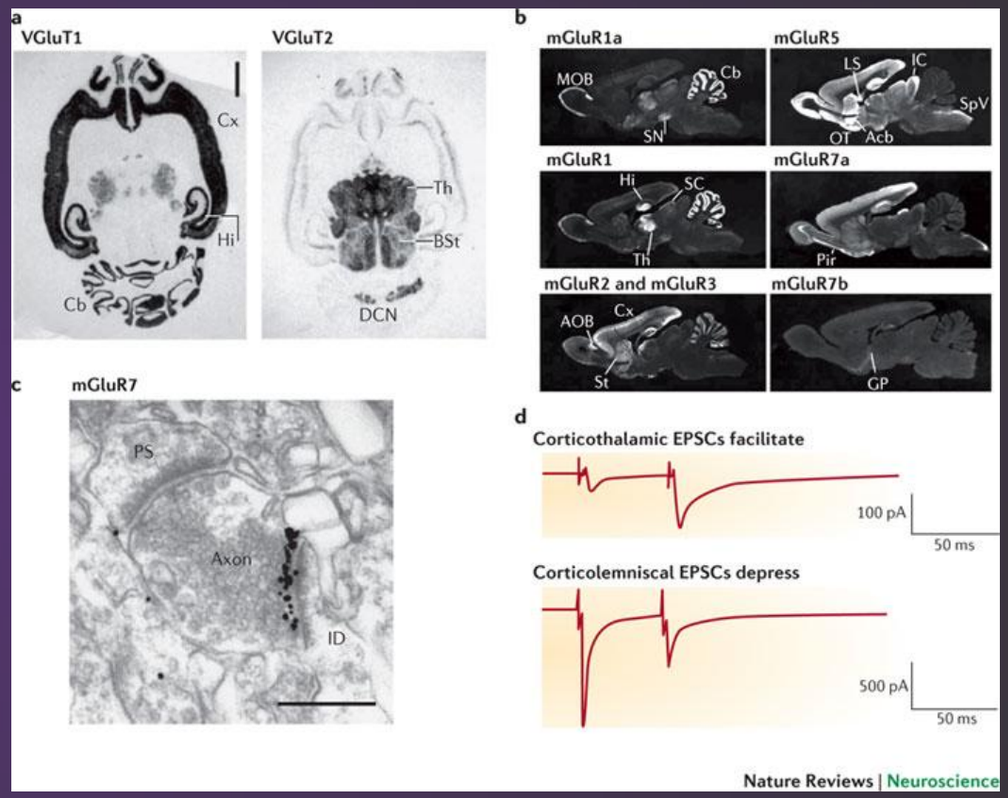
Synapse Diversity: Characterizing How Our Synapse Data Are Clustered

Adam Li, Tyler Tomita

March 10th, 2016

Significance

	Total PSD	G2C PSD	cPSD
Channels and receptors	80	60	39
MAGUKs/adaptors/scaffolders	54	31	33
Serine/threonine kinases	46	23	20
Tyrosine kinase	3	3	2
Protein phosphatases	18	13	11
G proteins and modulators	77	40	34
Signalling molecules and enzymes	278	167	98
Transcription and translation	119	36	41
Cytoskeletal and cell adhesion molecules	153	99	97
Synaptic vesicles and protein transport	159	119	76
Novel	107	29	11
Other	30	0	4
Summary	1124	620	466



synapse classification -> spatial distribution of synapse type -> relationship to disease

Gap

- Currently, very little is understood about the synaptic connections with our brain. It was originally believed that there were two types of synapses: 1) excitatory and 2) inhibitory
- It is now known that synapse diversity is actually much more vast

Challenge

- Computational challenge: >1M observations and 144 dimensions
- Observations are noisy
- Synaptic Connections Are Inherently Dependent and Not Much Information Is Known
- Validation of putative synapse types is hard

Formal Statement of Problem

Ho: Di ∈ U ∀ i∈x,y,z

Ha: ∃i ∈ x,y,z s.t. Di ∉ U

Hypothesis of Spatial Distribution of Synapses

$\vec{X} f_{\vec{X}} \in F_{\vec{X}}(\cdot;\theta) : \theta \in \Theta$

We assume f is a GMM and $\theta = [\mu, \Sigma, \vec{\pi}, \mathbf{k}]$, where k is the number of mixture components and $\vec{\pi}$ are the mixing weights of each mixture component.

Hypothesis of Number of Clusters There Are

$Fn(x)=1/n*\sum(I(Xi))$

Kolmogorov-Smirnov Test Statistic

$\Lambda=L(\theta1;X)L(\theta0;X)$

Likelihood Ratio Test Statistic

Data

Data Acquisition:

- Data acquired using serial section tramsission electron microscopy (ssTEM) imaging.

Data Features:

- 6 total features (each with channels n=24), but we threw out f4, f5 as suggested by Jovo.

- f0 = integrated brightness

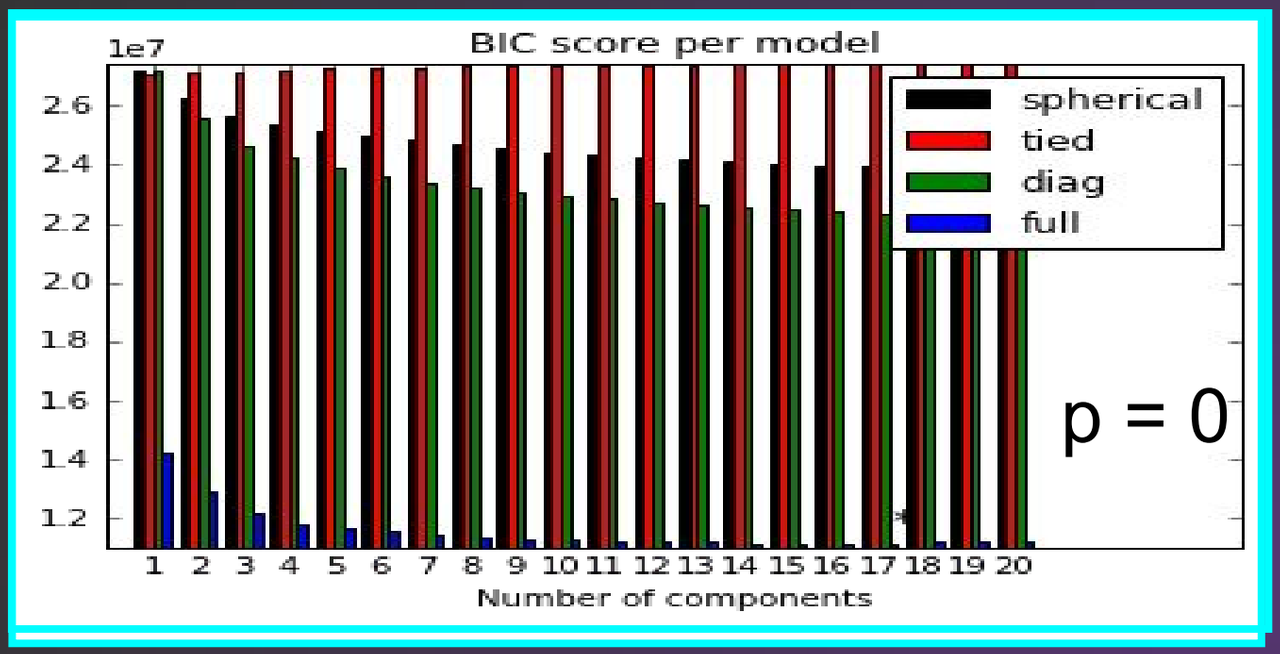
- f1 = local brightness

- f2 = distance to Center of Mass

- f3 = moment of inertia around synapsin maxima

- Each 24 channels for each feature represent some protein expression

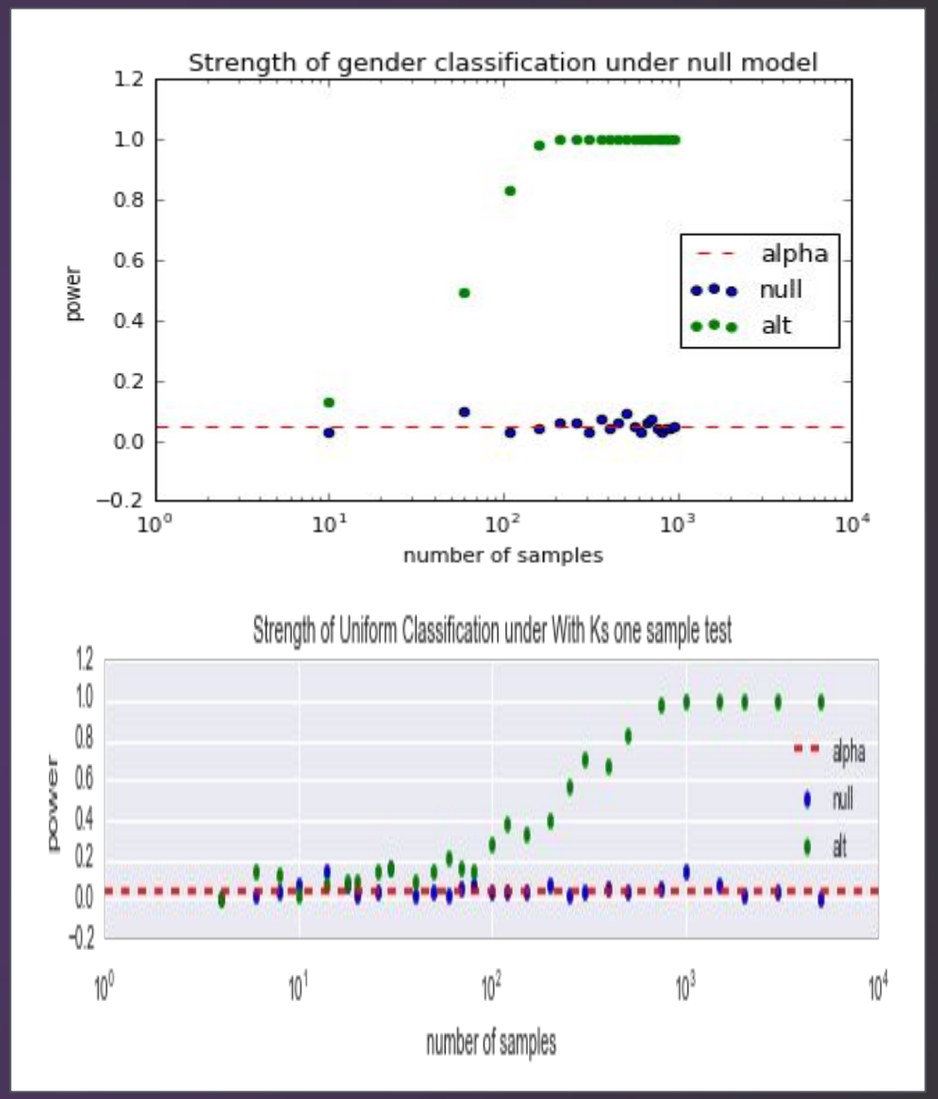
Results



KstestResult(statistic=0.081, pvalue=0.0)

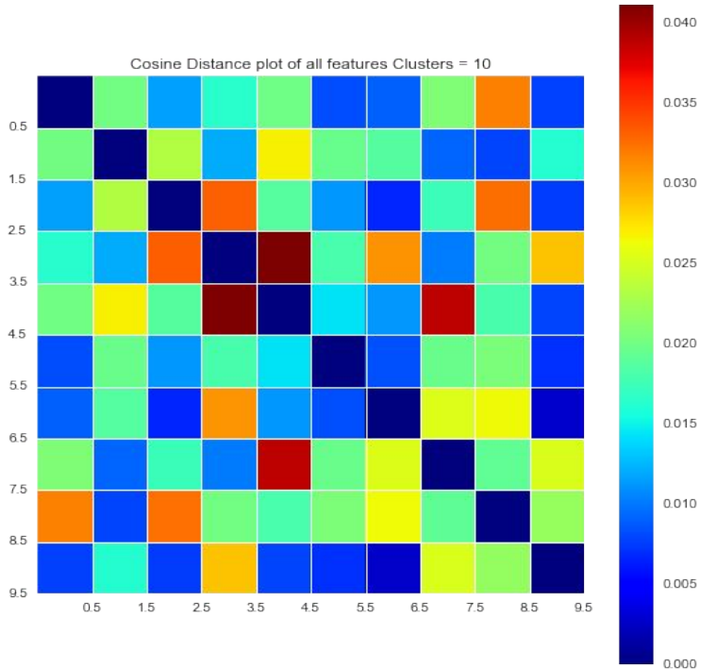
KstestResult(statistic=0.061, pvalue=0.0)

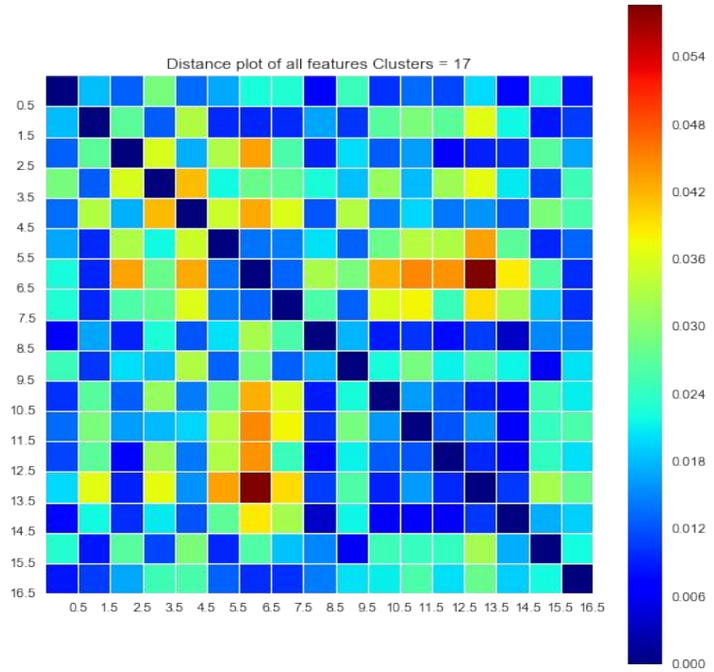
KstestResult(statistic=0.070, pvalue=0.0)

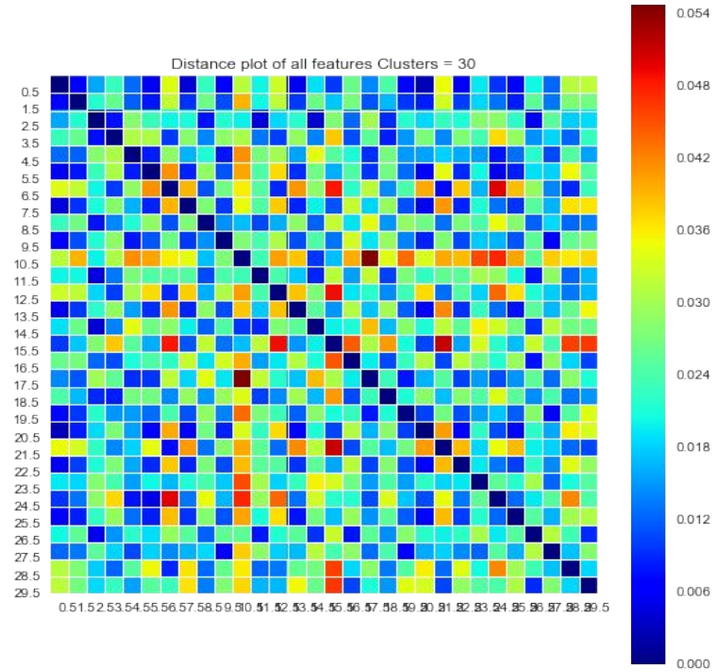


- 17 Optimal Clusters
- <x,y,z> are not uniformly distributed

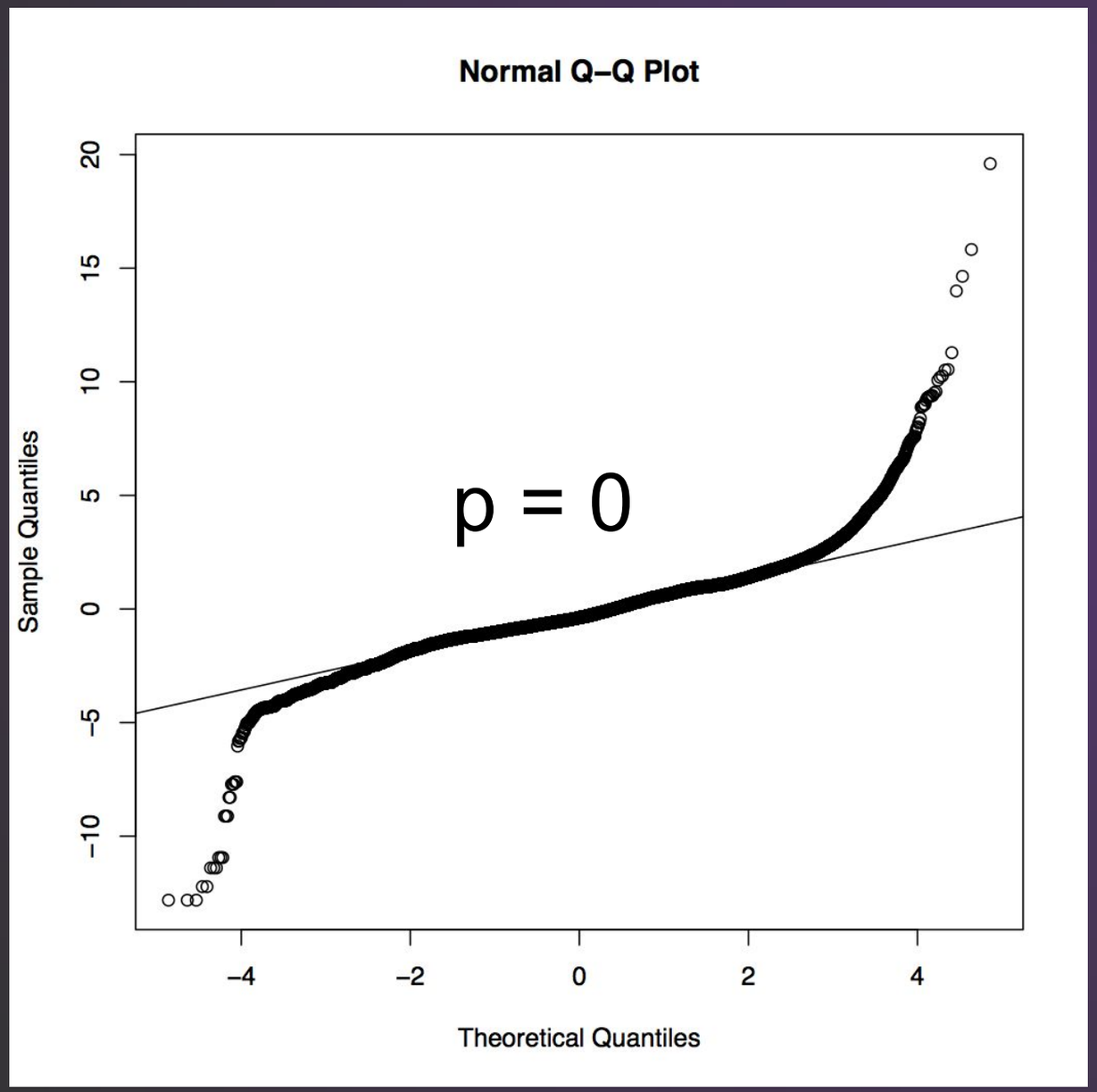
Model Checking







Model Checking



Henze-Zirkler test for multivariate normality

H0: Cluster is normally distributed
H1: Cluster is not

Conclusion/Future Steps

- Our results give us reason to suspect that there are more than two types of clusters, which supports the idea that synapses are much more diverse than previously believed.

- Due to computational restraints, we had to downsample by ~10 fold. Using all of the data will give us better estimates of clusters.

- Development of a way to validate putative synapse types identified in this work.

- Testing independence among features f0, f1, f2, f3