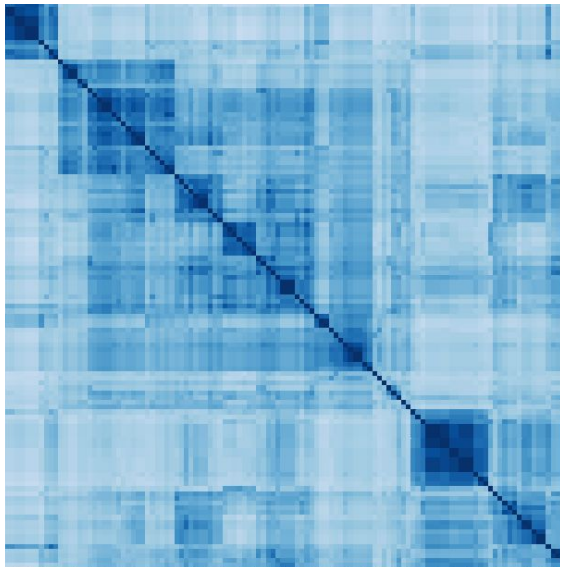
BENG 212 Final Project

Revealing location-specific variation and drug transport specificity in the Allen Brain Atlas



Kevin Rychel and Margot Wagner

March 19, 2019

Introduction to the Allen Brain Atlas

Allen Brain Atlas Overview

- Project by the Allen Institute for Brain Science since 2003
- Goal: insights into whole brain function
  - Emphasize disease treatment: Parkinson's, Alzheimer's, autism, etc.
- Contents
  - 'All genes - All structures' microarray
    - Used to obtain transcriptomic differences between structures
  - Human and Mouse
  - Development, aging, and disease
  - Imaging: histology, MRI
  - Tools for visualization
  - **More recently: RNA-seq of two human brains**
  - Single cell data

Chosen dataset

RNAseq

- 121 samples from 82 unique areas
- 22,318 genes

Preprocessing

- log(TPM + 1)
- Remove genes → 7,530 remain
  - Low/constant expression
  - Sequence < 100 nt

Donor H0351.2002 – Microarray Survey	
Tissue Receipt Date	8/25/2009
Sex	Male
Age	39 years
Race/Ethnicity	African American
Handedness	Left
Postmortem Interval	10 hours (estimated time of death to time that tissue is frozen)
Serology	Pass
Toxicology	Positive for atropine, caffeine, lidocaine and monoethylglycinexylidide (MEGX) at levels usually not toxicologically significant
Tissue pH	6.86
RNA Quality	Pass
Neuropathology	MRI-based Radiology Report: Normal; possible small pituitary adenoma Microneuropathology: Normal; single neurofibrillary tangle in entorhinal cortex
Tissue Received	25 cerebral slabs in coronal orientation; 5 mm thickness 17 cerebellar slabs in sagittal orientation; 5 mm thickness; 1 broken and irreparable 1 brainstem, whole
Additional Medical Information	None known

Region Tested	RIN value (Mean ± SD)
Frontal pole (left & right)	7.5 ± 0.2
Occipital pole (left & right)	7.1 ± 1.0
Cerebellum (left & right)	8.6 ± 0.6
Brainstem	7.3 ± 0.0

Project Summary

Part 1: Understanding the dataset
- Brain Structures
- PCA
- Agglomerative and K-Means clustering

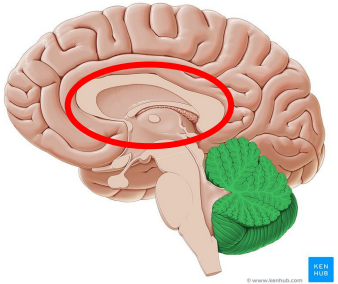
Part 2: Supervised learning
- Different model performances
- Different brain resolutions

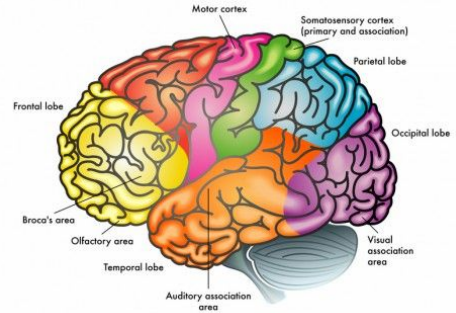
Part 3: Drug Transport
- Workflow for estimating structure-specific drug susceptibility
- Prediction of drug uptake

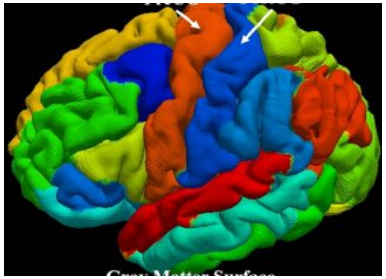
Does RNA expression predict region?

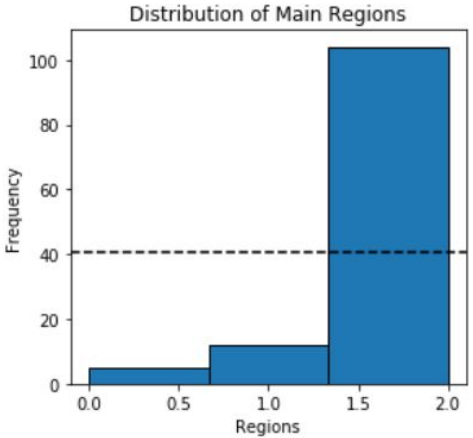
Can we predict where drugs end up in brain?

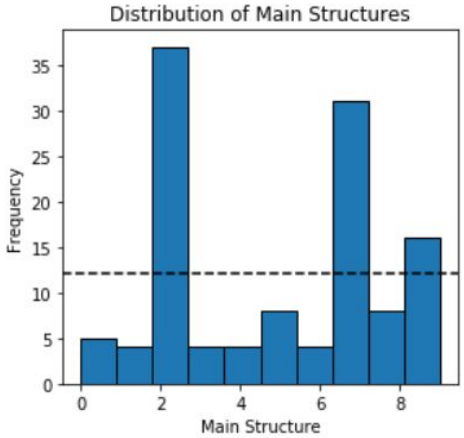
Data Visualization

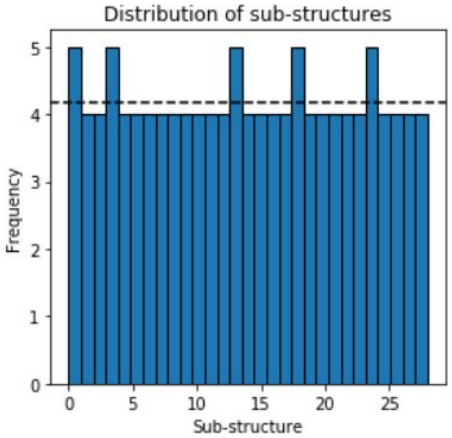


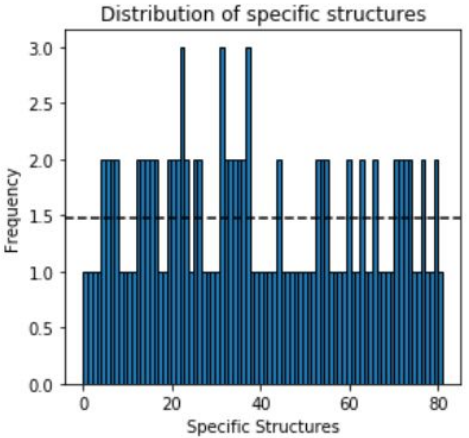






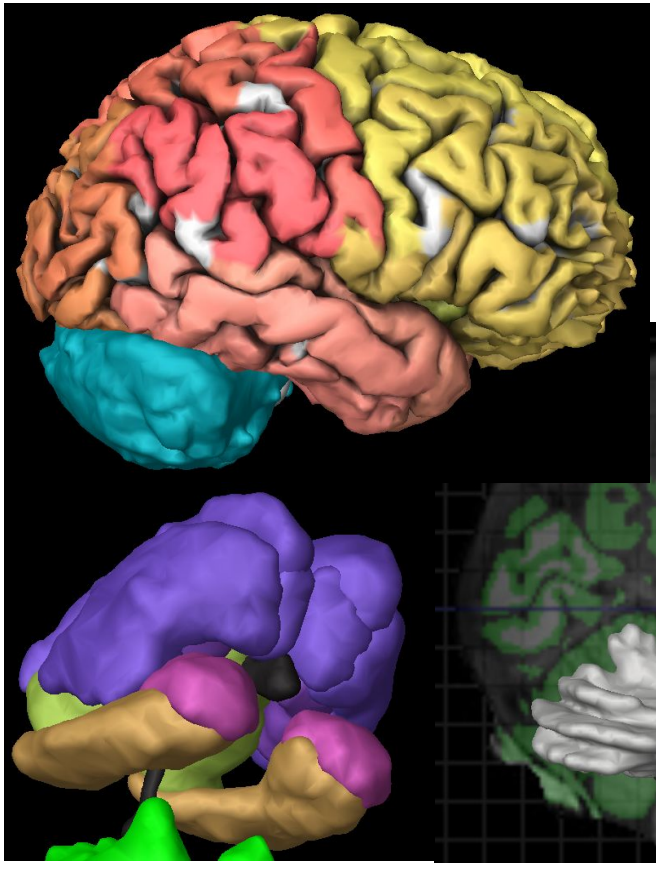


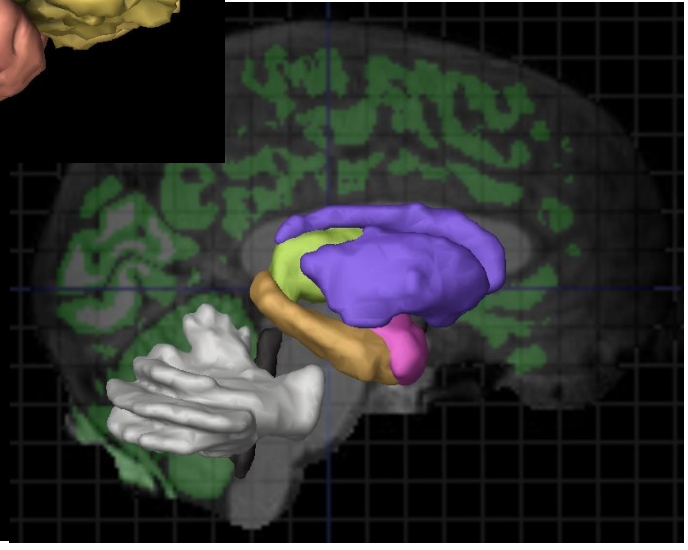


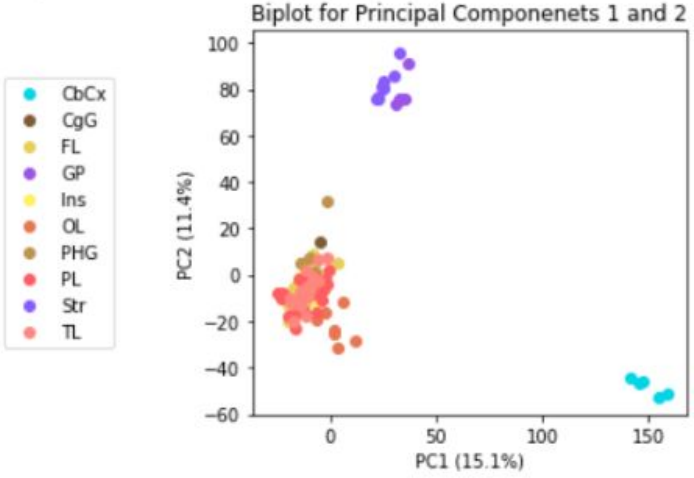


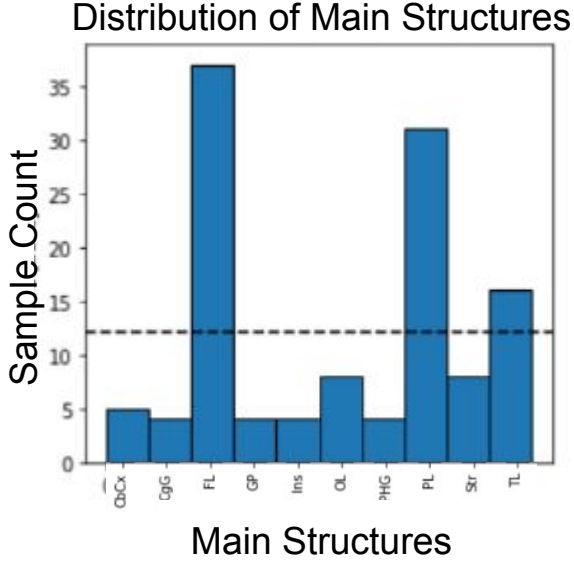
Increasing Resolution

Structures

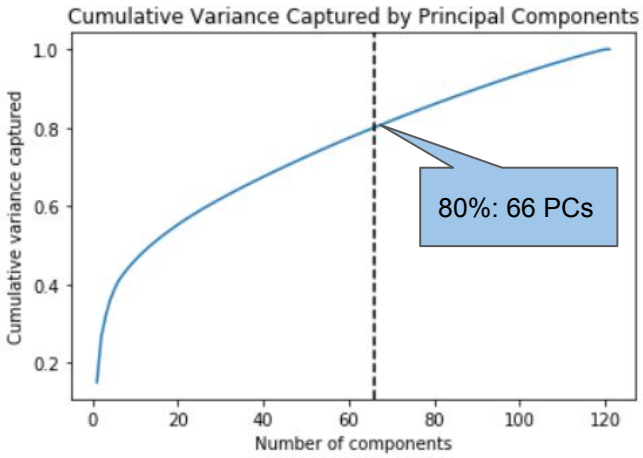





Biplot for Principal Componenets 1 and 2
CbCx
CgG
FL
GP
Ins
OL
PHG
PL
Str
TL
PC2 (11.4%)
PC1 (15.1%)
100
80
60
40
20
0
−20
−40
−60
0
50
100
150

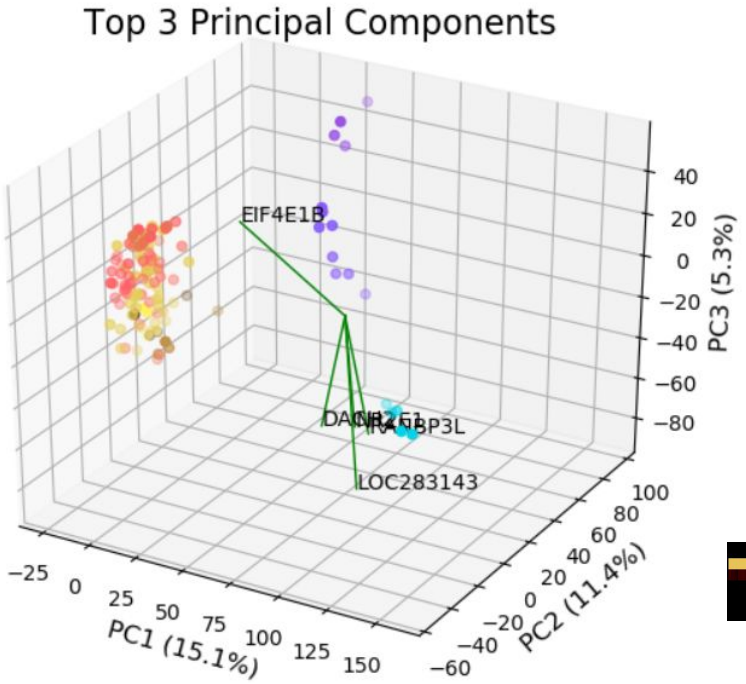

Distribution of Main Structures
Sample Count
35
30
25
20
15
10
5
0
CbCx
CgG
FL
GP
Ins
OL
PHG
PL
Str
TL
Main Structures

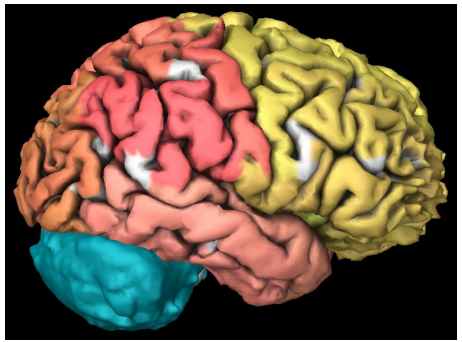
PCA

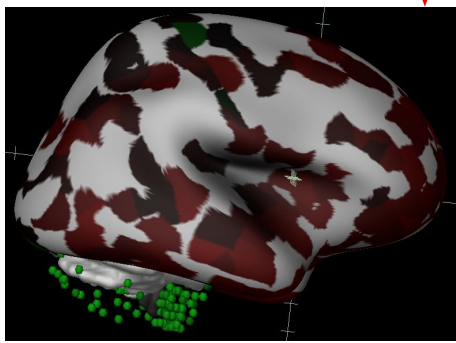


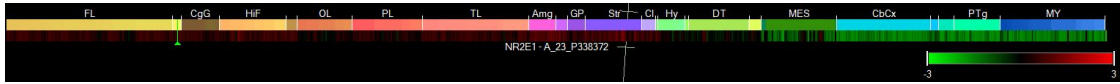
Top 5 Genes

- NR2E1
- DACH2
- RANBP3L
- EIF4E1B
- LOC283143



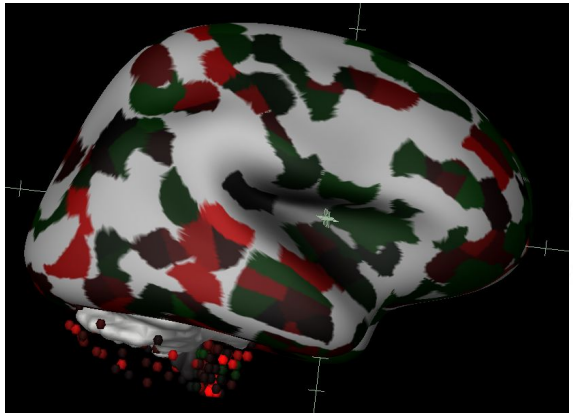




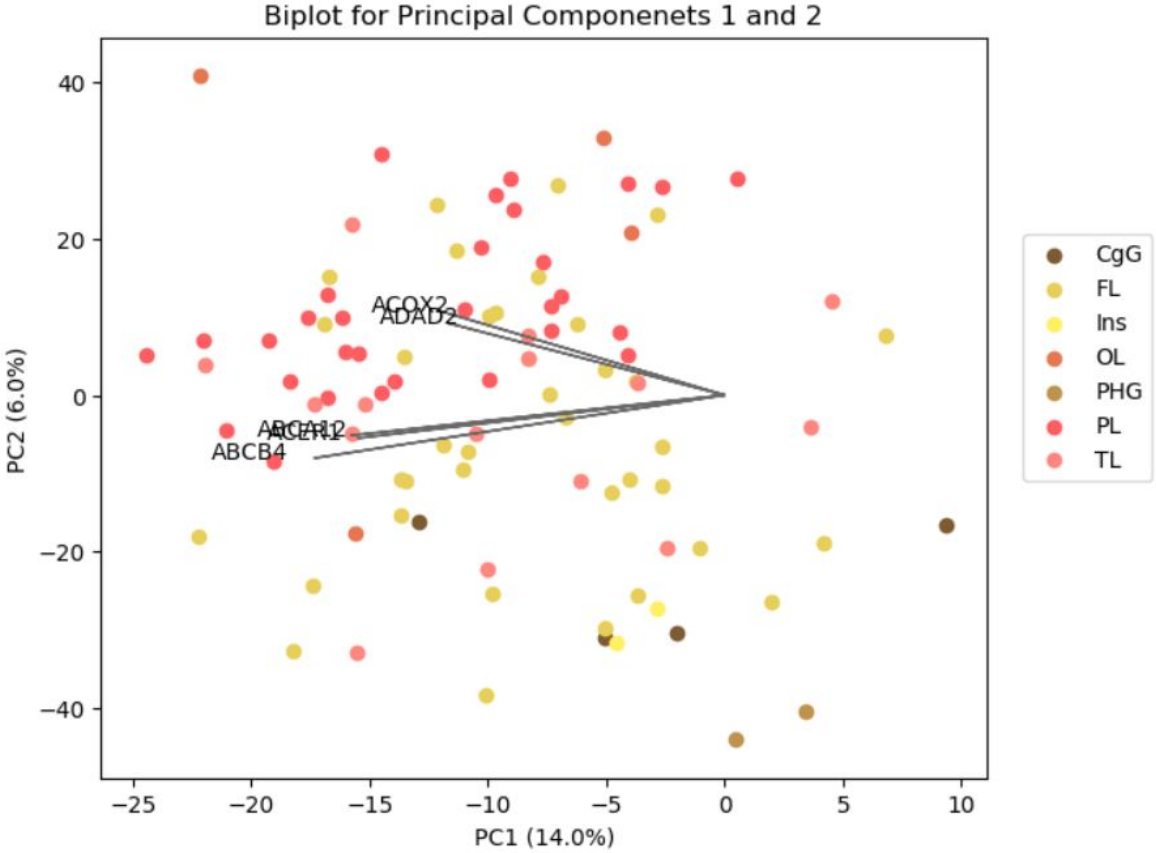


Z-score of log(TPM+1)

Cortex PCA



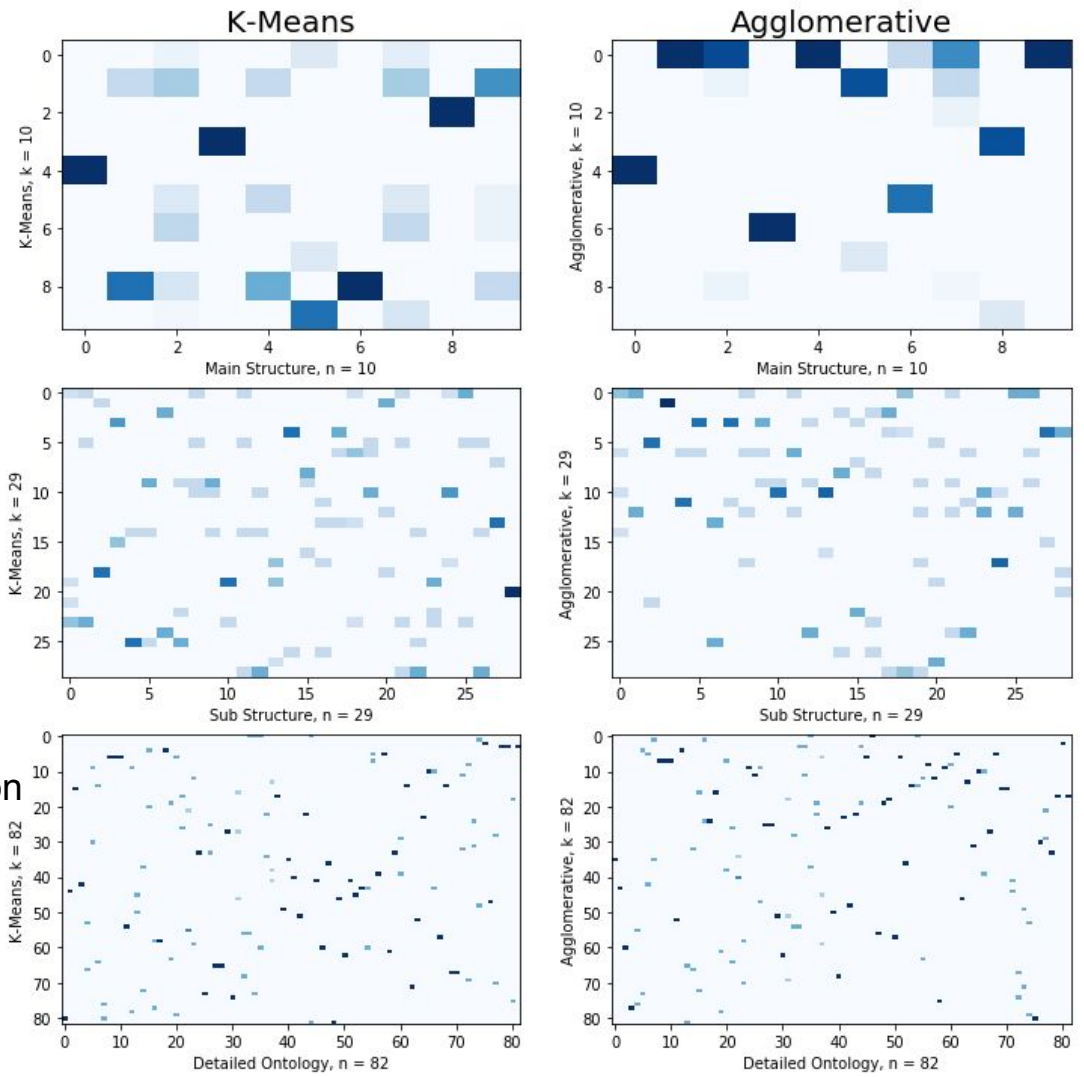
Top 5 Genes	
1	ABCB4
2	ACOX2
3	ACER1
4	ABCA12
5	ADAD2



Cluster members

No overlap Complete overlap

- Cluster data with k = [number of structures]
- Agglomerative → better performance
- Clustering does not totally recapitulate region



Supervised Learning:

How well can a model differentiate between brain regions from gene expression data?

Overview

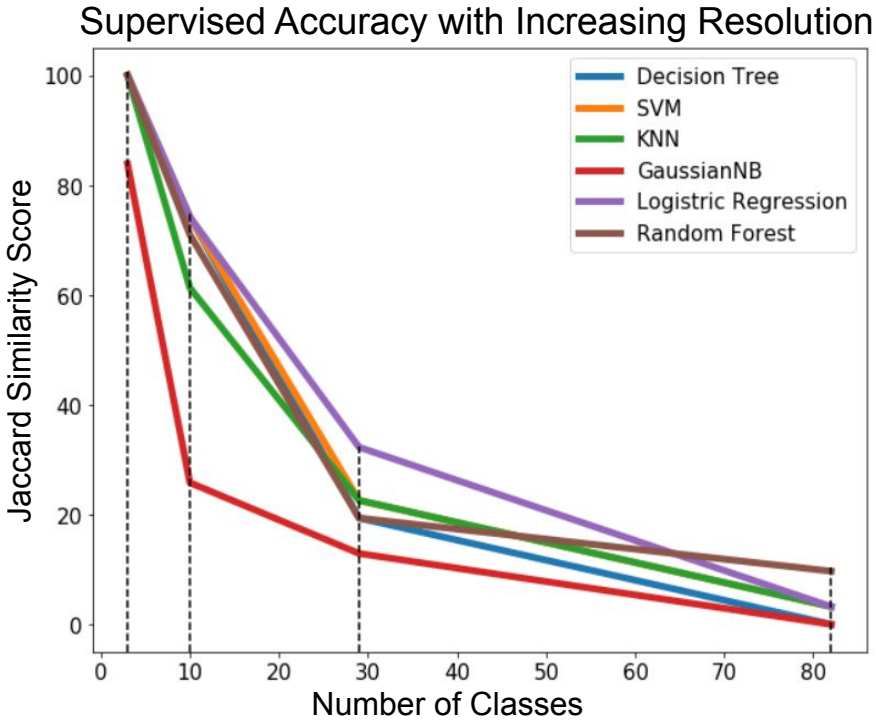
Shotgun classifier testing

- Decision Tree
- Support Vector Machine (SVM)
- K-Nearest Neighbors
- Logistic Regression
- Gaussian Naive Bayes
- Random Forest

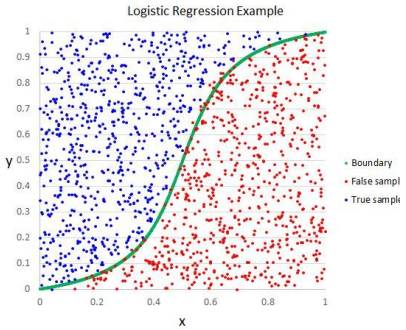
Top model refinement

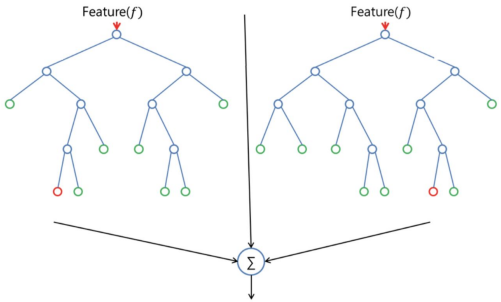
- Bootstrapping
- Cross-validation
- Regularization

Coarse Grain Training



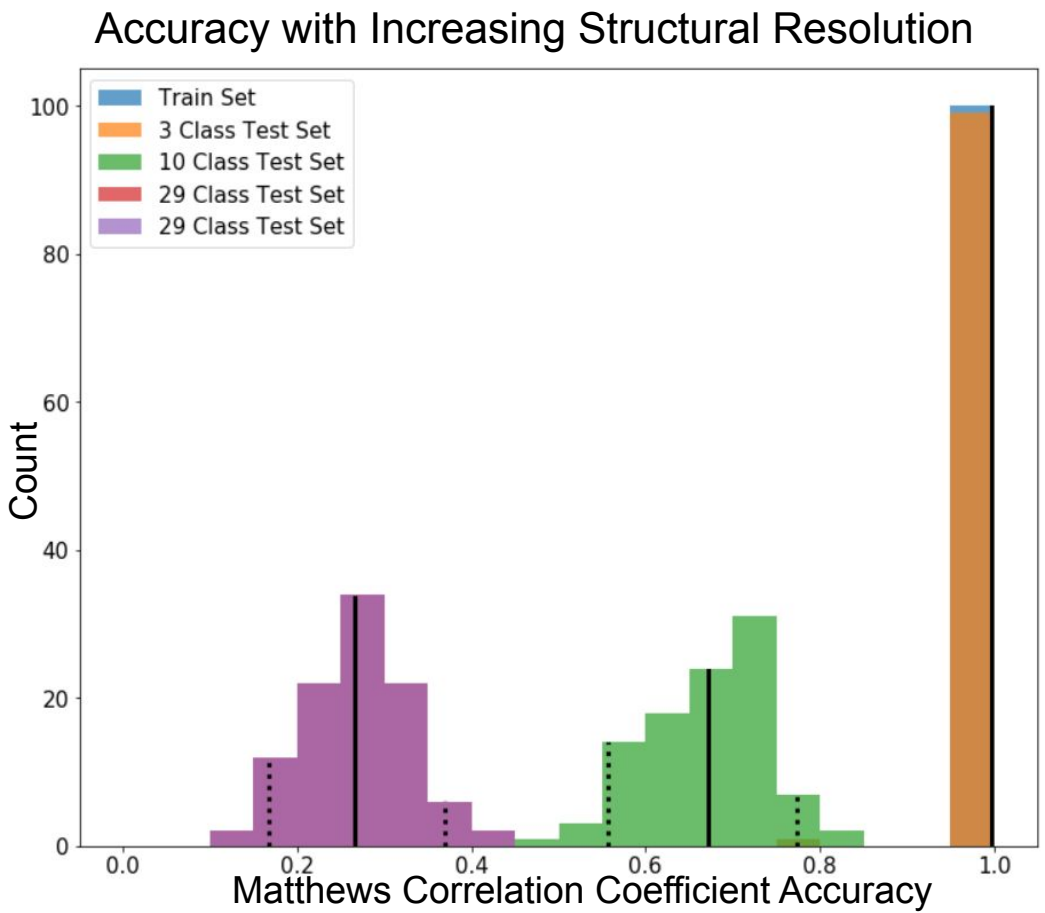
- Trained and tested 5 multiclass classifier for each resolution
- **Multinomial Logistic Regression** and **Random Forest** performed the best across resolutions





Multinomial Regression

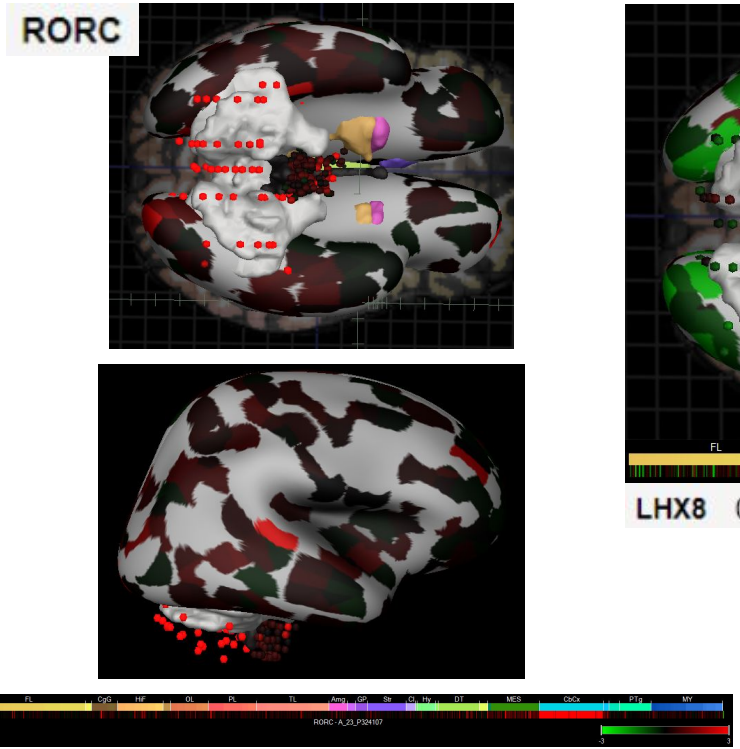
- Performance decreases as number of classes increases
- Cross-validation proves no overfitting for 3 class
  - Others not enough samples
- L2 regularization
  - L1 could not converge
- One-vs-all
- 3 class: 50% overlap in genes between brainstem and cortex
  - No overlap with cerebellum

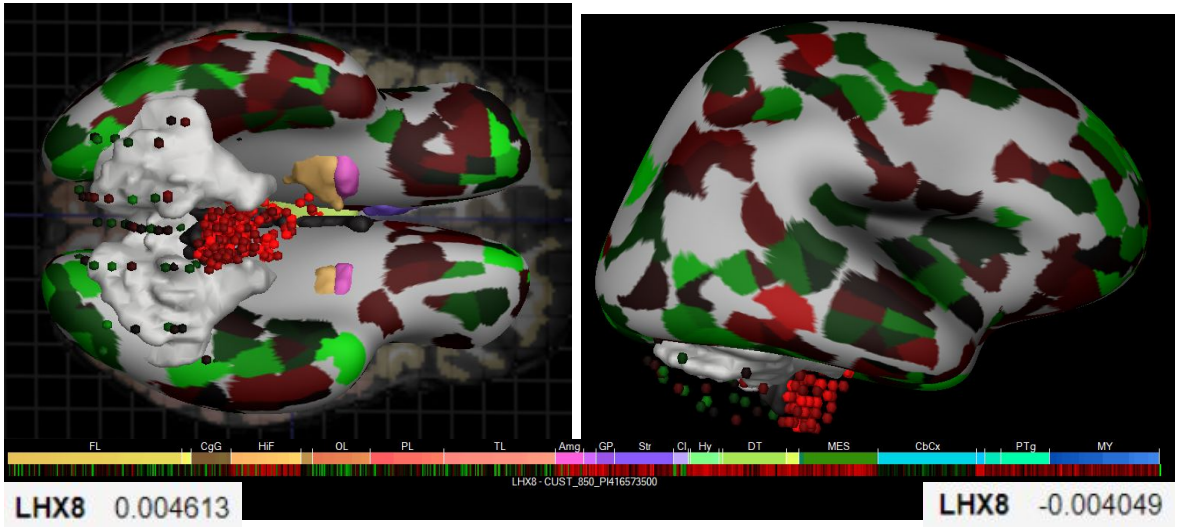


Cerebellum

Brainstem

Cortex

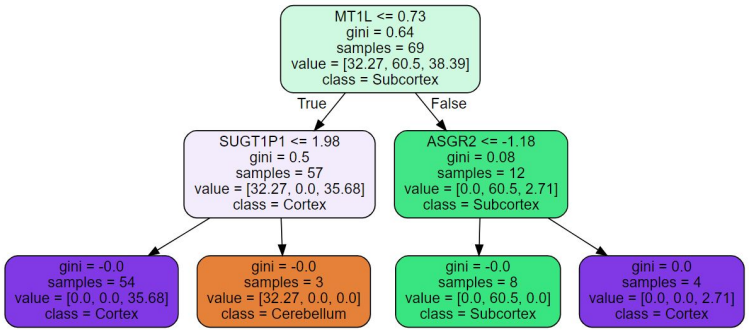




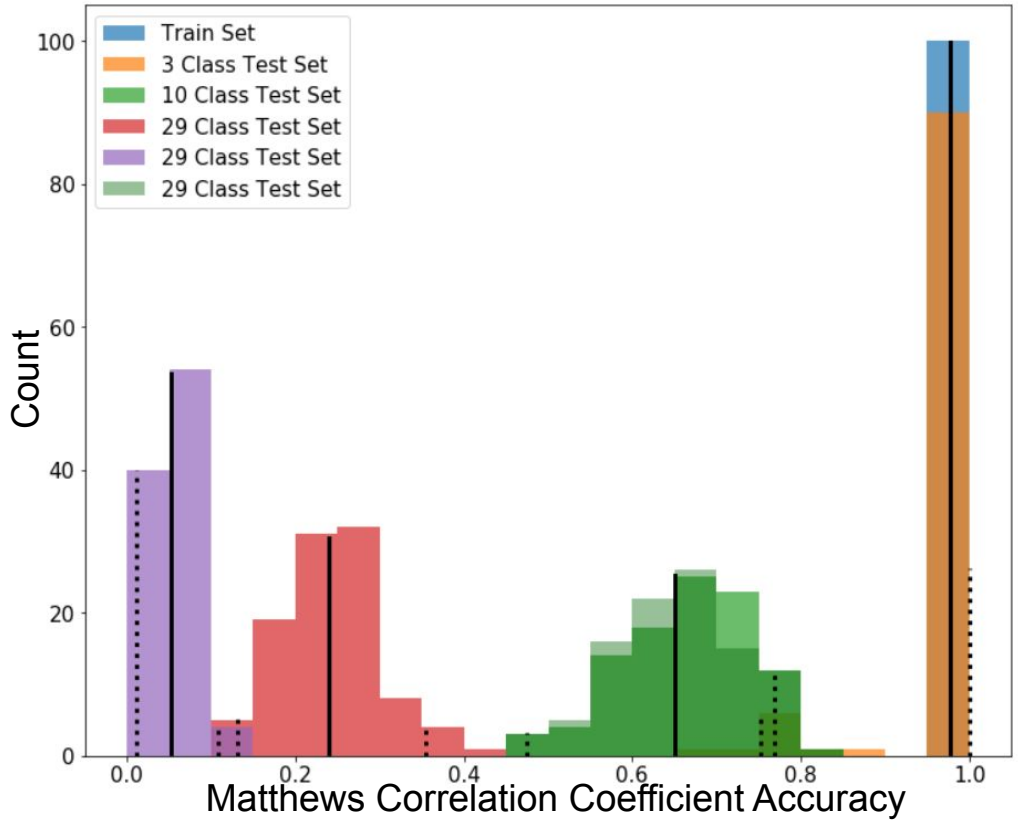
Z-score of log(TPM+1)

Random Forest

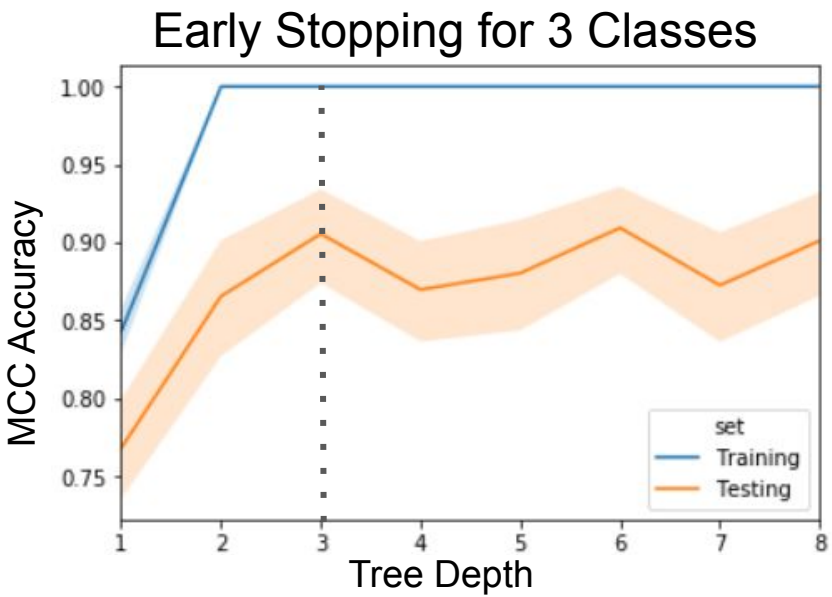
- Performance decreases as number of classes increases
- Initially built until fully expanded
- Inherently multiclass

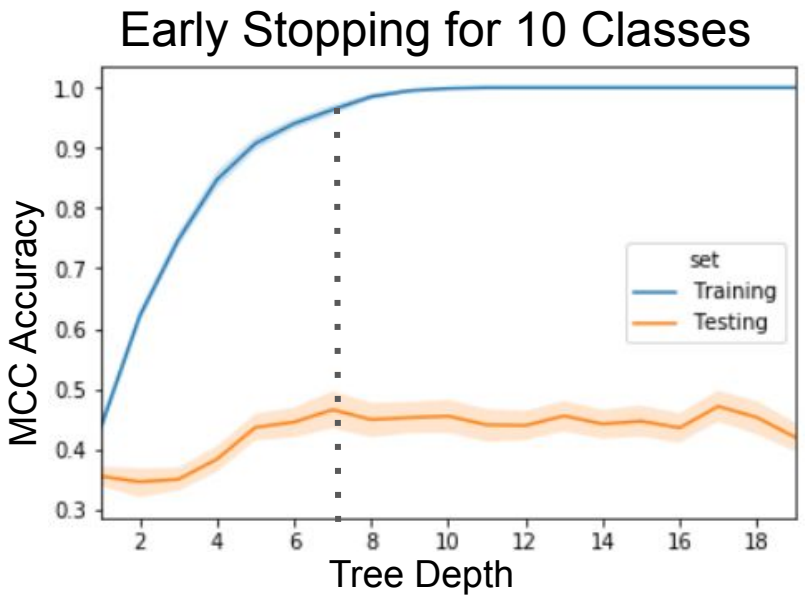


Accuracy with Increasing Structural Resolution



Early Stopping





- Both trees can do early stopping while maintaining performance

Outcomes

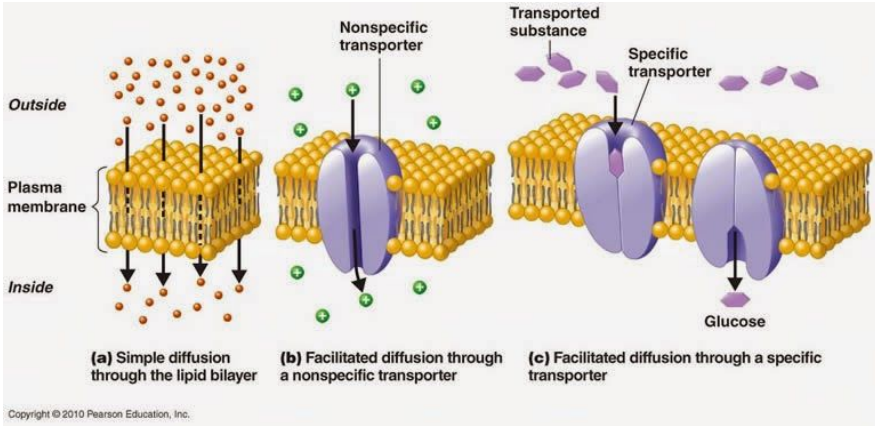
- Very good performance for 3 class
  - Significant drop off after that
- Can obtain useful information from 3 and 10 class models
- Multinomial regression can more easily show biological information
- Transcription factor expression useful for 3 class differentiation

Supervised learning possible at low resolution from this dataset

Drug Transport:

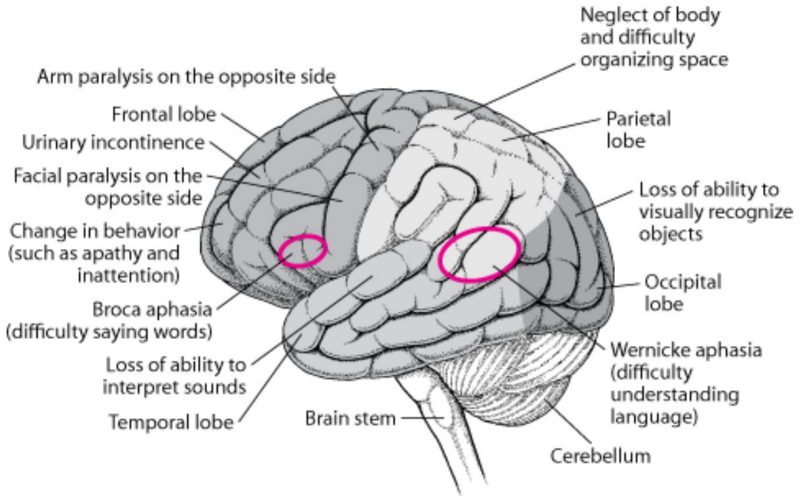
How much does a given brain region take up a given drug?

Motivation



Apply a new scientific paradigm:
carrier-mediated drug uptake

(Dobson & Kell, Nature Reviews Drug Discovery, 2008)

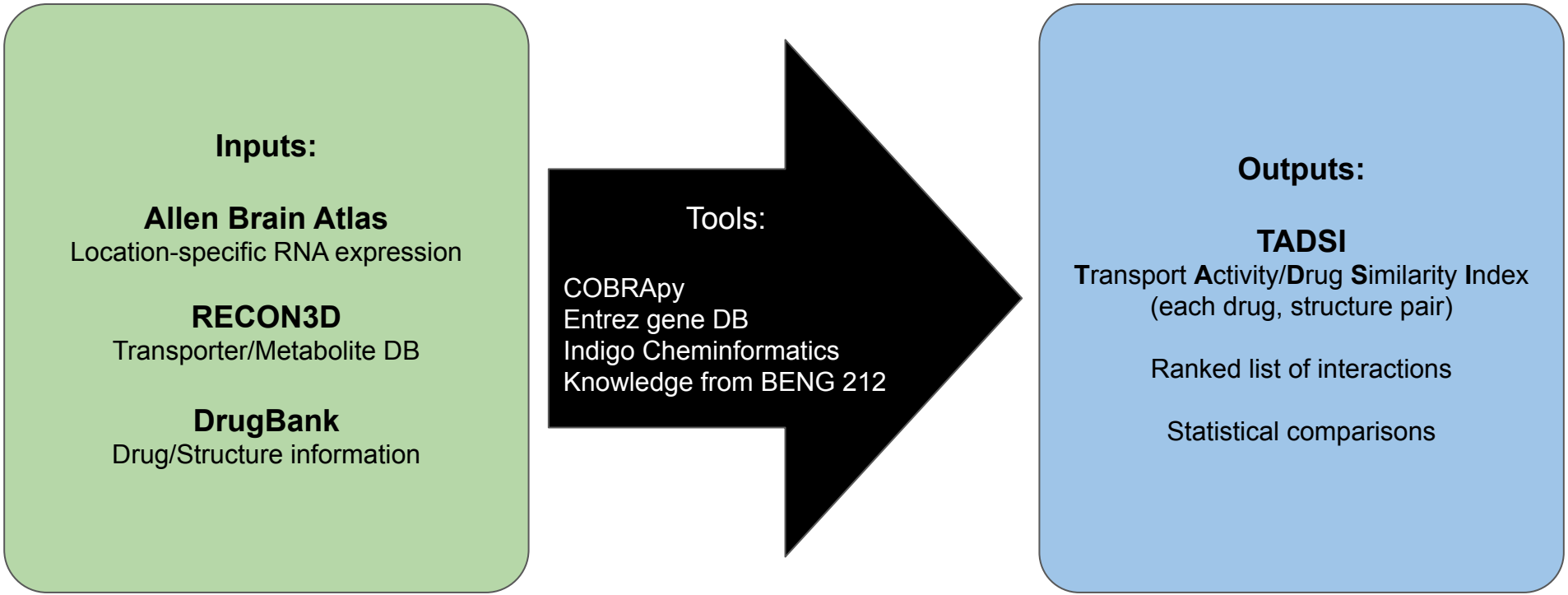


Inform targeted drug discovery

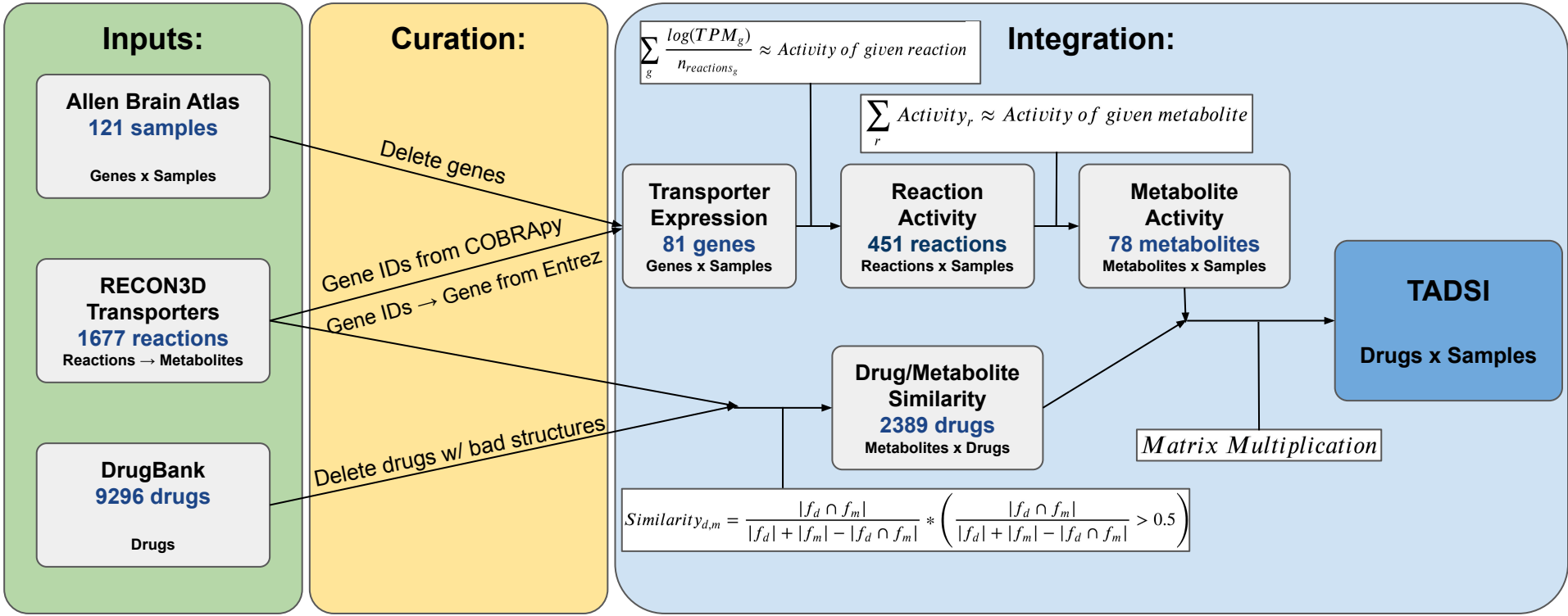
Understand off-target effects

Images:
https://www.merckmanuals.com/home/brain,-spinal-cord,-and-nerve-disorders/brain-dysfunction/brain-dysfunction-by-location
https://www.pearson.com/us/higher-education/program/Mathews-Biochemistry-4th-Edition/PGM39253.html

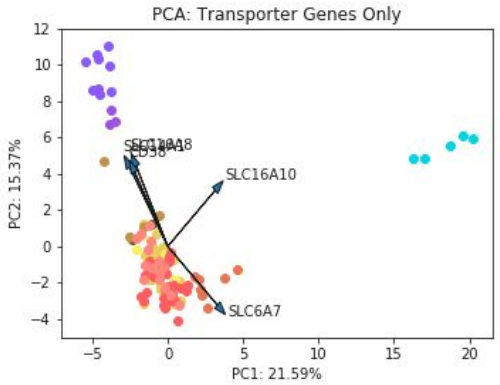
Overview of workflow

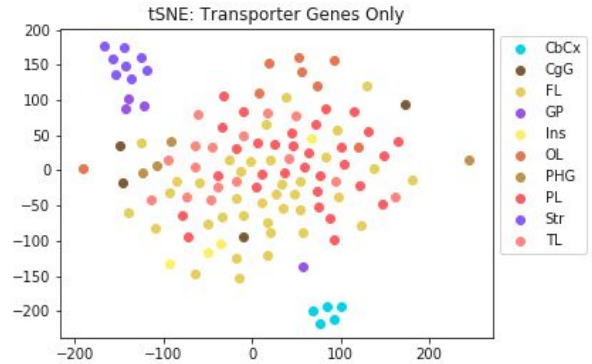


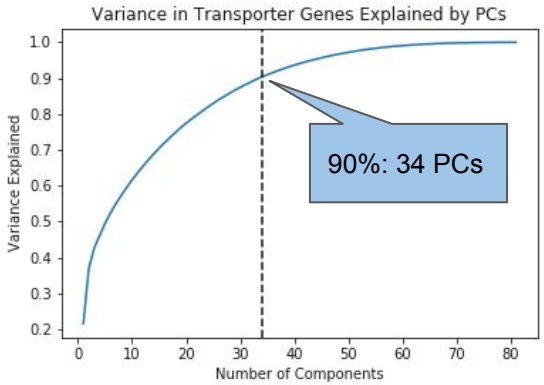
Detailed workflow



Reduced dimensions of transporter geneset





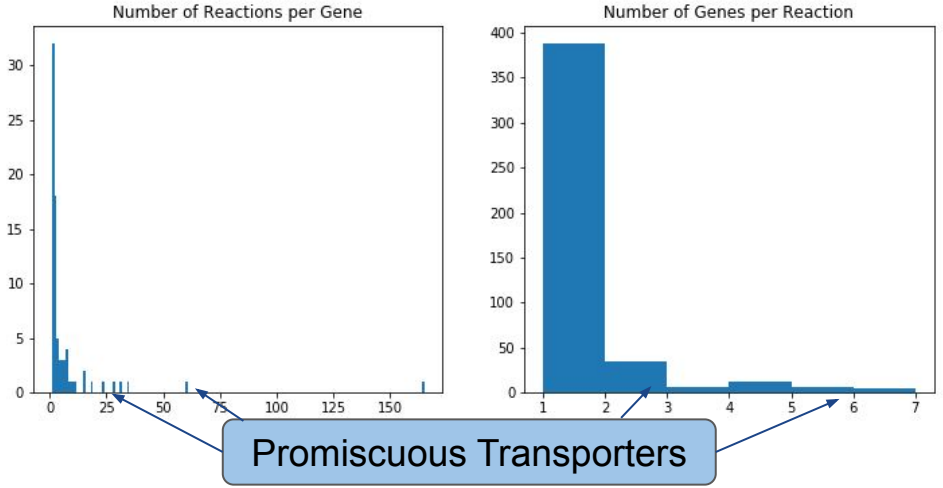


Gene	Metabolite	Expression location/Details
SLC14A1	Urea	Expressed in erythrocytes and the kidney
SLC16A8	Monocarboxylates	Cerebellar choroid plexus: basal epithelia
SLC6A7	L-proline, Na+	Expressed in brain. Proline acts as neurotransmitter

Gene-Reaction Mapping

Ex: SLC7A7 (cationic AA transporter):

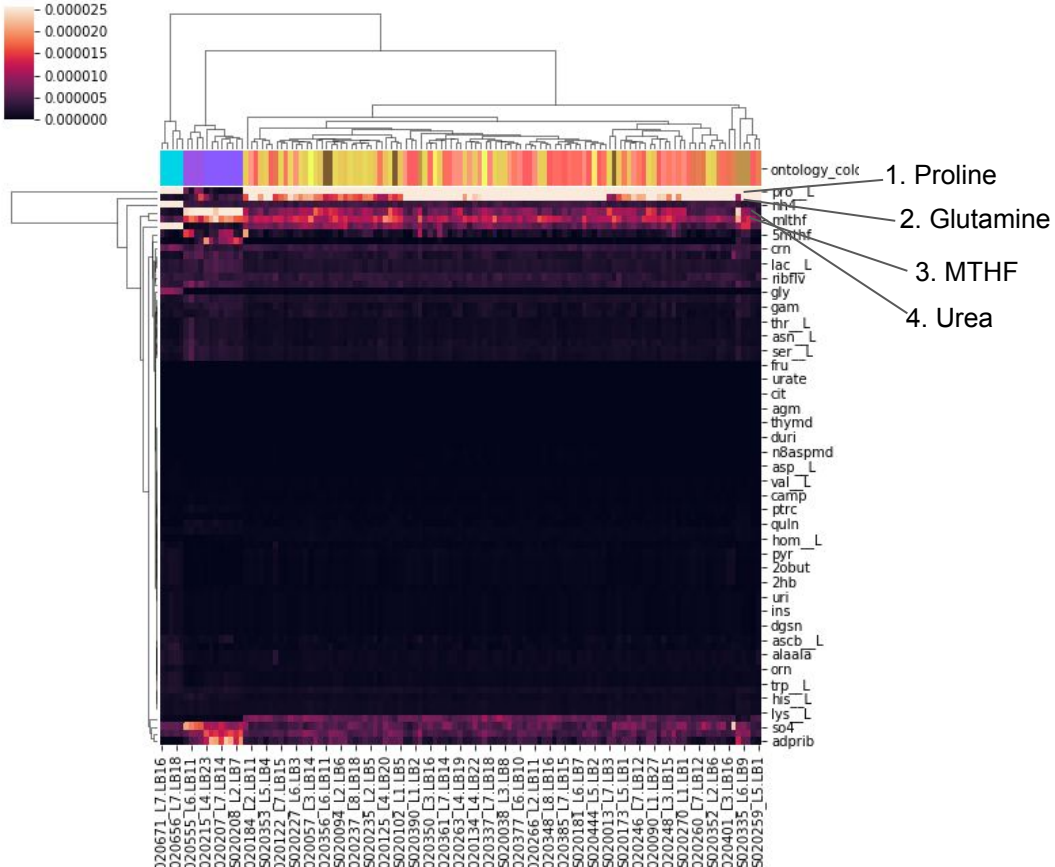
	Reaction	Stoichiometry	Transported Metabolite
405	SERLYSNaex	na1_e + ser__L_e + lys__L_c -> na1_c + ser__L_...	lys__L
406	SERLYSNaex	na1_e + ser__L_e + lys__L_c -> na1_c + ser__L_...	ser__L
2157	ALAyLATthc	h_e + ala__L_e + arg__L_c -> h_c + ala__L_c + ...	ala__L
2158	ALAyLATthc	h_e + ala__L_e + arg__L_c -> h_c + ala__L_c + ...	arg__L
2181	GLNyLATthc	h_e + gln__L_e + arg__L_c -> h_c + gln__L_c + ...	arg__L
2182	GLNyLATthc	h_e + gln__L_e + arg__L_c -> h_c + gln__L_c + ...	gln__L
2189	HISyLATtc	na1_e + arg__L_c + his__L_e -> na1_c + arg__L_...	arg__L
2190	HISyLATtc	na1_e + arg__L_c + his__L_e -> na1_c + arg__L_...	his__L
2191	HISyLATthc	h_e + arg__L_c + his__L_e -> h_c + arg__L_e + ...	arg__L
2192	HISyLATthc	h_e + arg__L_c + his__L_e -> h_c + arg__L_e + ...	his__L
2199	LEUyLATthtc	h_e + arg__L_c + leu__L_e -> h_c + arg__L_e + ...	arg__L



- Not proteomics → ignore protein-level regulation
  - Assume no transport complexes
- Assume each gene has equal activity for each reaction it performs
- Assume the contributions of each gene are additive

$$\sum_{g} \frac{log(TPM_g)}{n_{reactions_g}} \approx Activity\ of\ given\ reaction$$

Metabolite activity in each brain region



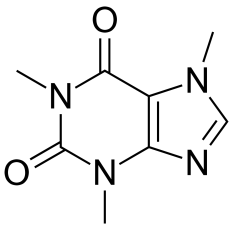
$$\sum_r Activity_r \approx Activity\ of\ given\ metabolite$$

Potential Improvements:

- Expand database
- Single-cell omics
- Network information
  - Flux direction
  - Transporter affinity
  - Metabolite concentrations

Cheminformatics Workflow

Known Structure



“SMILES” Structure

CN1C=NC2=C1C(=O)N(C)C(=O)N2C

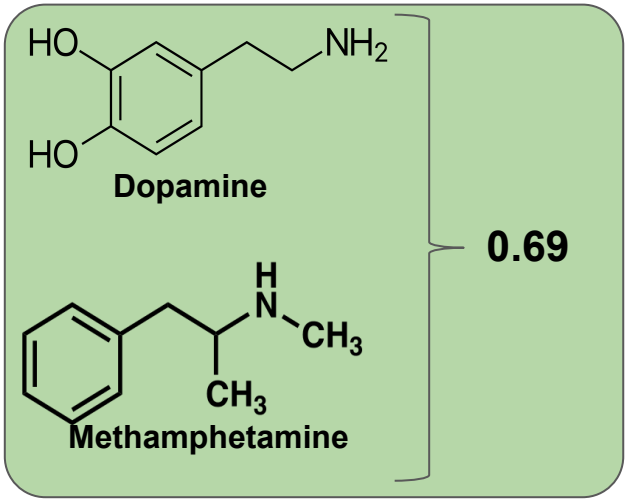
Fingerprint

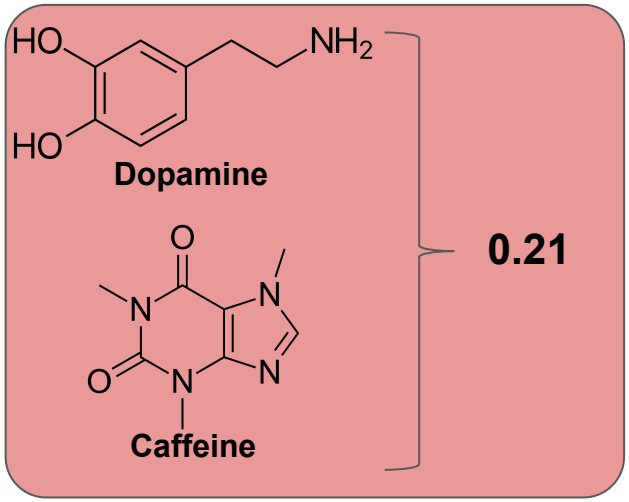
Long list of attribute presence/ absence

$$Similarity_{d,m} = \frac{|f_d \cap f_m|}{|f_d| + |f_m| - |f_d \cap f_m|}$$

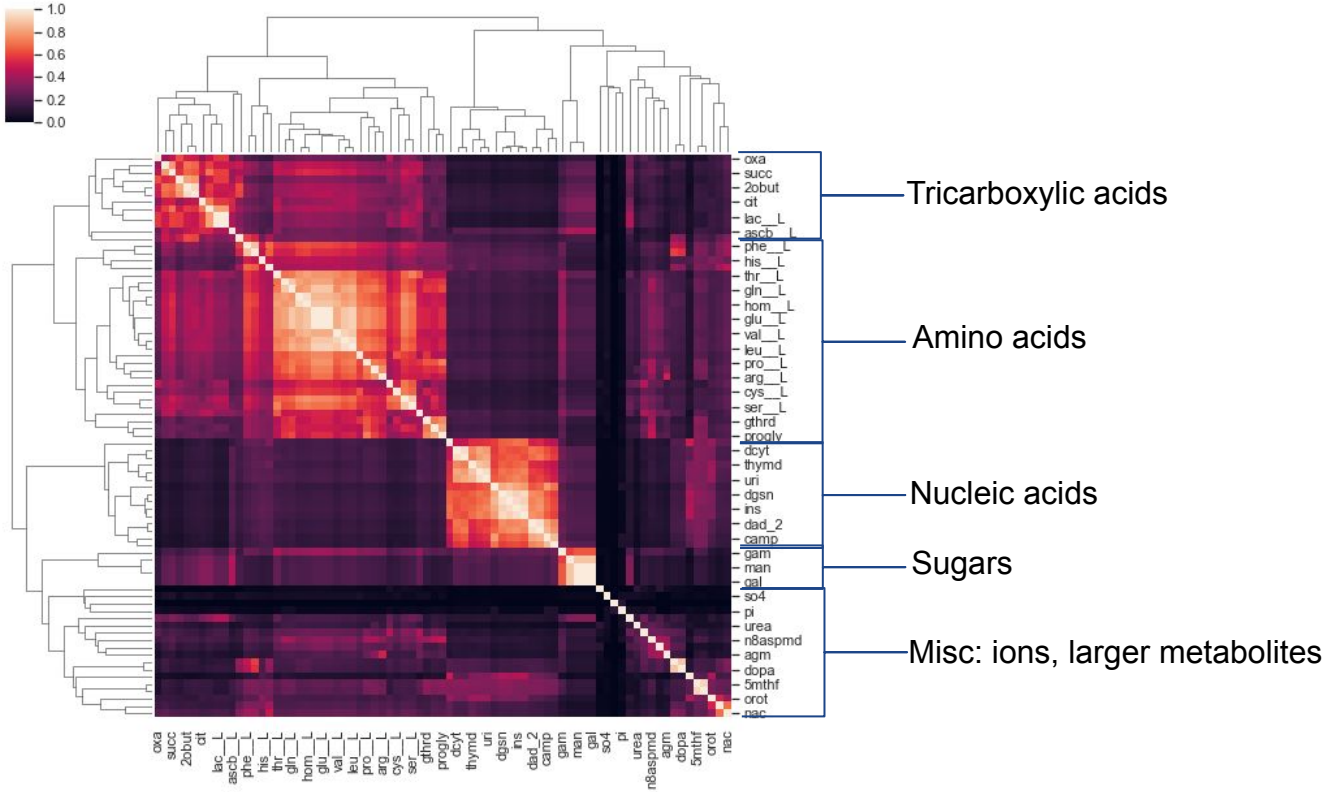
Also called:
Tanimoto Index
Jaccard Similarity

Indigo python package: PMC3083596

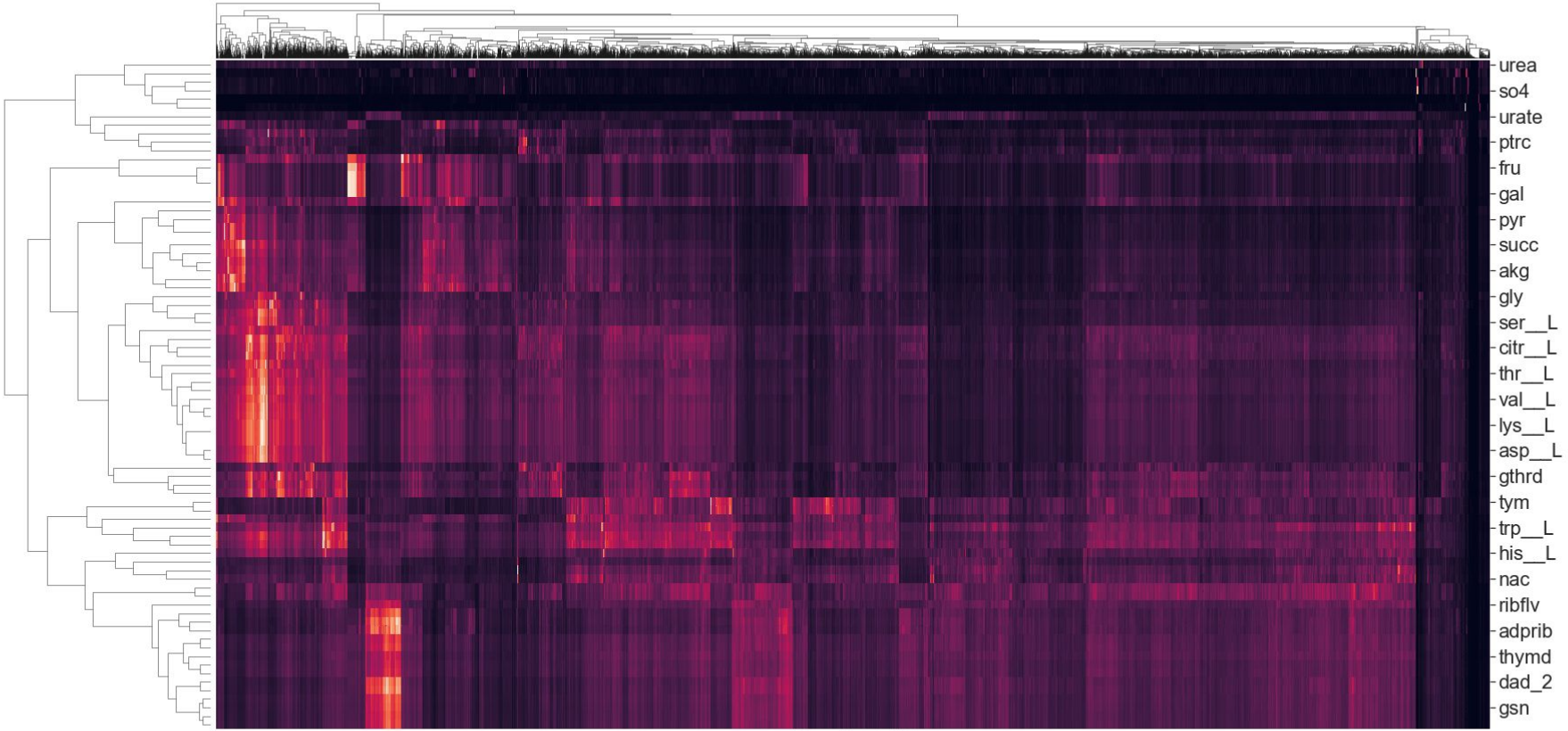




Metabolite:Metabolite Similarity



Metabolite:Drug Similarity

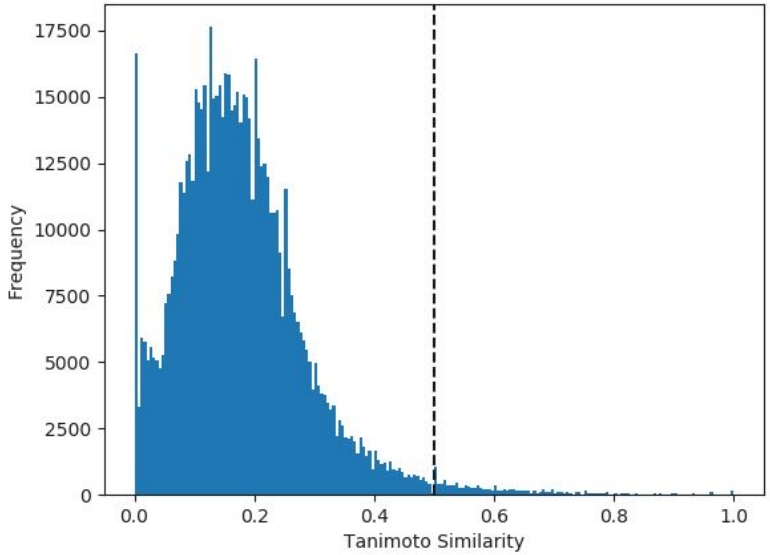


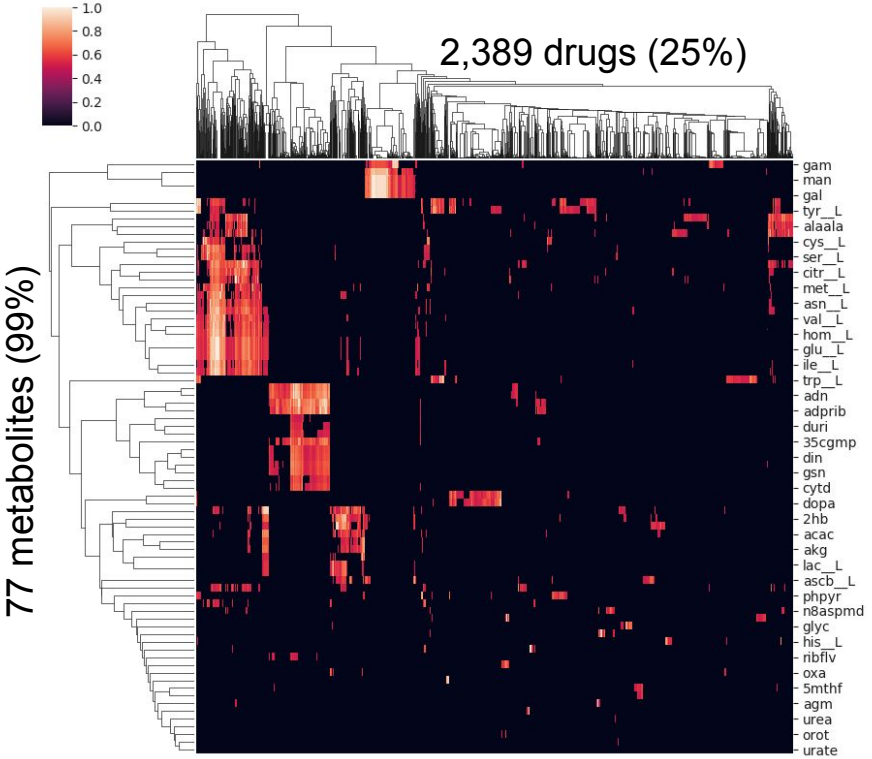
Rule of 0.5

$$Similarity_{d,m} = \frac{|f_d \cap f_m|}{|f_d| + |f_m| - |f_d \cap f_m|} * \left(\frac{|f_d \cap f_m|}{|f_d| + |f_m| - |f_d \cap f_m|} > 0.5 \right)$$

Drugs with similarity < 0.5 cannot use a metabolite's transporter

S. O'Hagan et al, Metabolomics, 2015





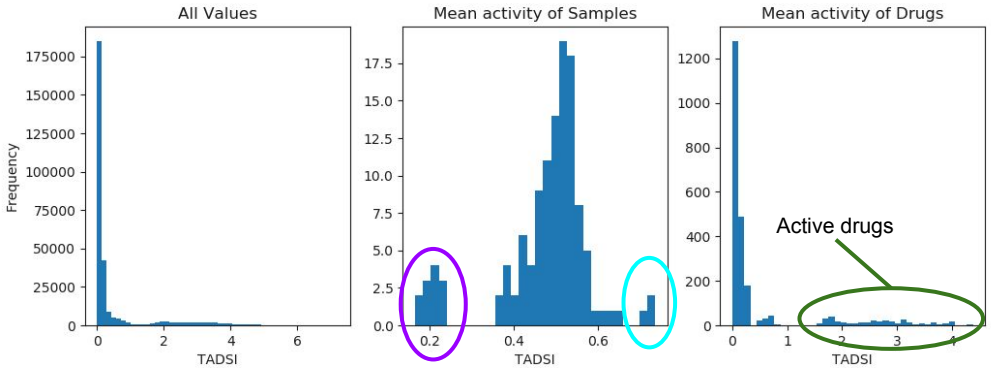
TADSI Matrix

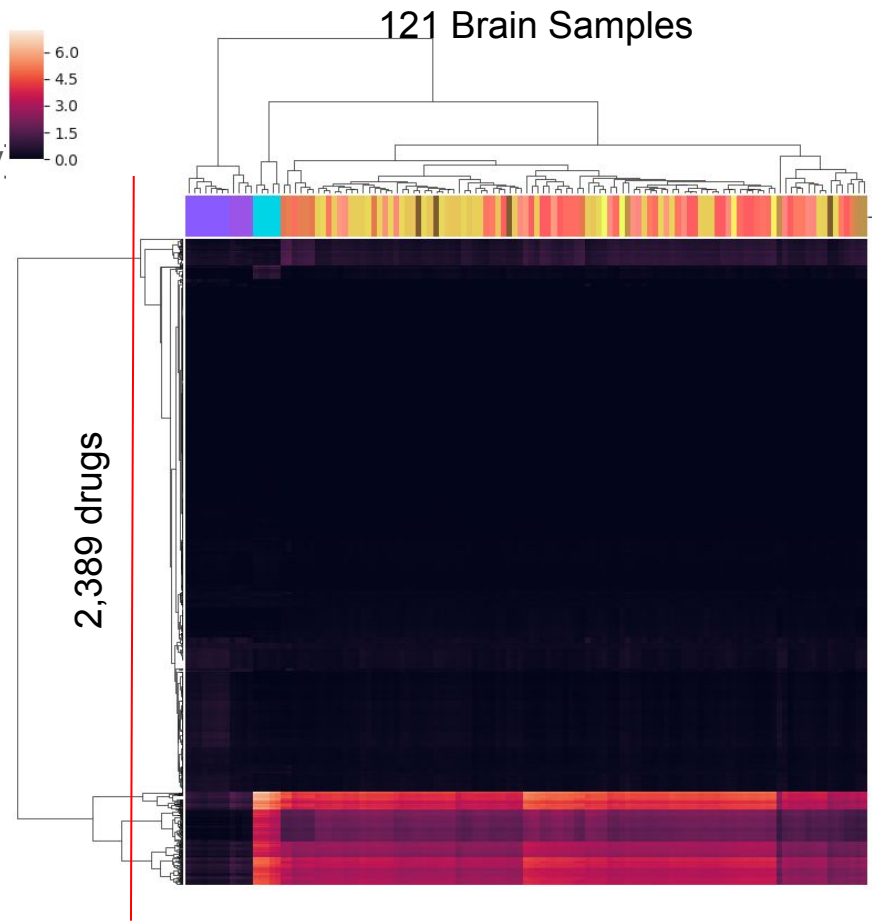
Transport **A**ctivity **D**rug **S**imilarity **I**ndex

$[TADSI] = [Met:Drug\ Similarity]^T \times [Met:Sample\ Activity]$

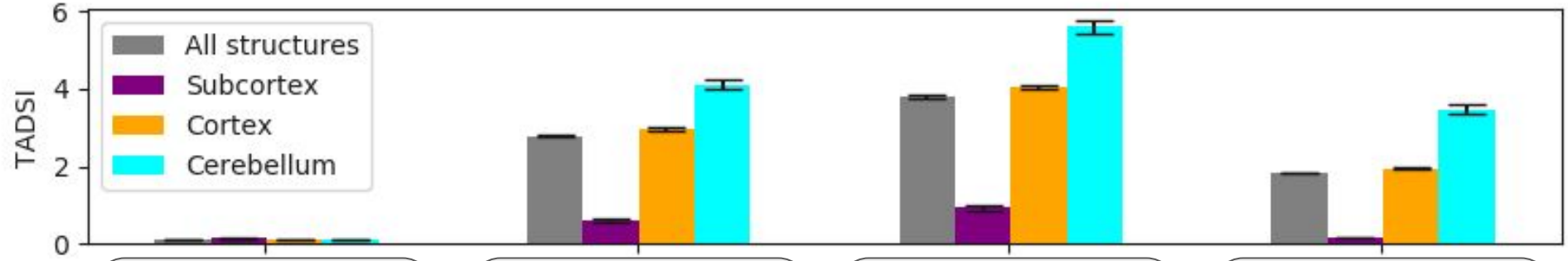
Scaled by standard deviation of full matrix

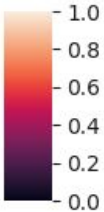
- No highly specific drugs
- 4 clusters, 3 of which are active
  - 1 somewhat specific to cerebellum





TADSI Drug Clusters

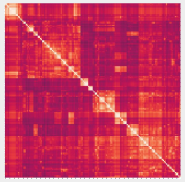


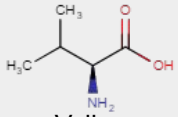


Cluster 0:
2,046 drugs

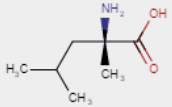
Very low TADSI

Cluster 1:
160 drugs
Least difference cortex/cerebellum

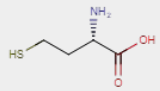




Valine



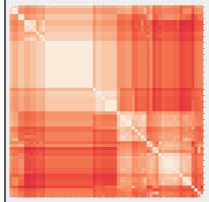
2-Methylleucine

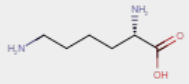


2-Amino-4-mercapto-butyric acid

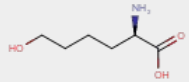
Amino Acids

Cluster 2:
65 drugs
Highest uptake

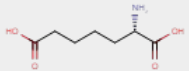




Lysine



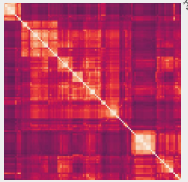
6-Hydroxy-norleucine

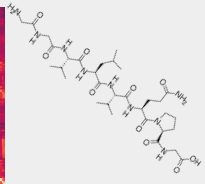


2-Aminopimelic acid

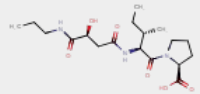
Amino Acids

Cluster 3:
118 drugs
Lower uptake

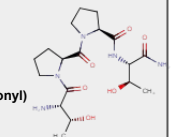




Larazotide



N-(3-Propylcarbamoyloxirane-2-Carbonyl) -Isoleucyl-Proline



Rapastinel

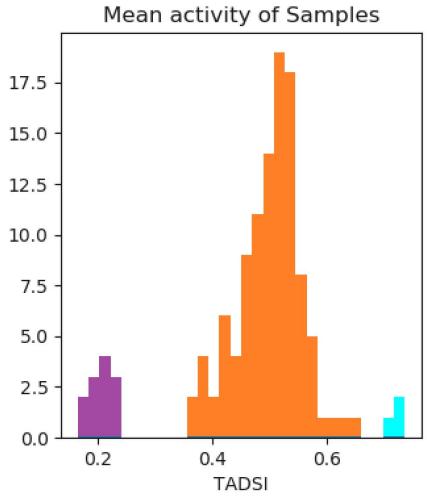
Peptides

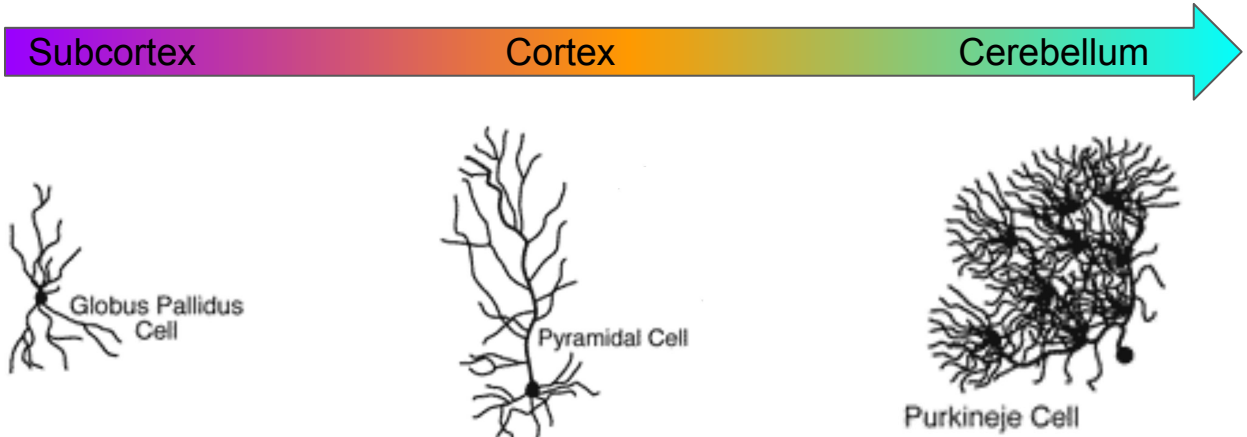
Cluster 3: Representative Drugs

1	Larazotide	Cell permeability suppression for celiac disease
2	N-(3-Propylcarbamoyloxirane-2-carbonyl)-isoleucyl-proline	Experimental cathepsin B inhibitor (proteolysis)
3	Rapastinel	Clinical trials for **depression, OCD**
4	Perindopril	ACE inhibitor (hypertension)
5	Lisinopril	ACE inhibitor (hypertension)
6	Enalaprilat	ACE inhibitor (hypertension)
7	N-[1-Hydroxycarboxyethyl-Carbonyl]Leucylamino-2-Methyl-Butane	Experimental cathepsin B inhibitor (proteolysis)
8	Ethylaminobenzylmethylcarbonyl Group	Experimental candidapepsin-2 inhibitor (proteolysis)
9	Methyl-n-({(2s,3s)-3-[(Propylamino)Carbonyl]Oxiran-2-yl}Carbonyl)-l-isoleucyl-l-prolinate	Experimental cathepsin B inhibitor
10	Ciclosporin	Immunosuppression

SA:V ratio may contribute to differential uptake

Hypothesis: More surface area → more transporters





Images:
https://canvas.brown.edu/courses/851434/pages/looking-at-neurons?module_item_id=6586407

Outcomes

- Predicted uptake of 343 drugs in cortex and cerebellum
  - Mainly amino acids and peptides
  - At least one experimental antidepressant
- Demonstrated differences between three brain parts
  - Highest uptake: cerebellum
  - Lowest uptake: subcortex
- Identified areas for improvement
  - Single cell resolution
  - Thorough annotation
  - Integration with other omics data/networks
  - Drug localization experiments for validation

Limitations

- Very incomplete transporter list
  - Master’s project: complete annotation
  - 81 genes, 451 reactions, 78 metabolites
- Tanimoto similarity
  - May not predict affinity
- Coarse granularity
  - RNAseq run on sections of brain instead of single cells
  - Blood-Brain-Barrier permeability ignored
- Disease state ignored
- Long list of assumptions

Assumptions

1. Transporter activity is only determined by its RNA concentration
    a. Ignores protein level regulation
    b. Ignores kinetics, affinities, and metabolite concentrations
2. Transporters carrying out the same transport event behave independently
    a. No complexes or preferential transport affinities
3. Each unique transport reaction occupies an equal fraction of a promiscuous transporter’s activity
4. Flux direction is ignored
5. For drug/metabolite similarities above a threshold, the activity of the drug scales with its similarity
6. Drug and metabolite activities through each transporter in a region are additive

Conclusion

Conclusion

Part 1:

- **Supervised learning possible at low resolution from this dataset**
- 3 and 10 class analysis works well
- Biological relevance vs. minimizing overfitting

Part 2:

- **Predicted uptake of 343 drugs in cortex and cerebellum**
- Demonstrated differences between three brain parts
- Identified areas for improvement
  - Single cell resolution
  - Thorough annotation
  - Integration with other omics data/networks
  - Drug localization experiments for validation

Thanks for listening!

References

1. 2010 Allen Institute for Brain Science. Allen Human Brain Atlas. Available from:human.brain-map.org
2. Shen, E.H., Overly, C.C., Jones, A.R. The Allen Human Brain Atlas: comprehensive gene expression mapping of the human brain. *Trends Neurosci* vol. 35, 12 (2012): 711-4.
3. Kukurba, K.R., Montgomery, S.B. RNA Sequencing and Analysis. *Cold Spring Harb Protoc* vol. 11 (2015): 951-69.
4. Hawrylycz, M.J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* vol. 489 (2012): 391-99.
5. Hawrylycz, M. *et al.* Canonical genetic signatures of the adult human brain. *Nature Neuroscience* vol 18 (2015): 1832-44.
6. Mendes, P., Oliver, S.G., Kell, D.B. Fitting Transporter Activities to Cellular Drug Concentrations and Fluxes: Why the Bumblebee Can Fly. *Trend Pharmacol Sci* vol. 36, 11 (2015): 710-23.
7. Bajusz, D., Racz, A., Heberge, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *Journal of Cheminformatics* vol. 7, 20 (2015)
8. P.D. Dobson, D.B. Kell, Carrier-mediated cellular uptake of pharmaceutical drugs: an exception or the rule?, *Nature Reviews Drug Discovery*. 7 (2008) 205–220. doi:10.1038/nrd2438.
9. S. O'Hagan, N. Swainston, J. Handl, D.B. Kell, A 'rule of 0.5' for the metabolite-likeness of approved pharmaceutical drugs, *Metabolomics*. 11 (2015) 323–339. doi:10.1007/s11306-014-0733-z.
10. Wishart DS, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 2017 Nov 8. doi: 10.1093/nar/gkx1037.
11. E. Brunk, et al., Recon3D enables a three-dimensional view of gene variation in human metabolism, Nat. Biotechnol. 36 (2018) 272–281. doi:10.1038/nbt.4072.
12. D. Pavlov, M. Rybalkin, B. Karulin, M. Kozhevnikov, A. Savelyev, A. Churinov, Indigo: universal cheminformatics API, J Cheminform. 3 (2011) P4. doi:10.1186/1758-2946-3-S1-P4.
13. Entrez Help [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2005-. Entrez Help. 2006 Jan 20 [Updated 2016 May 31].

Questions?

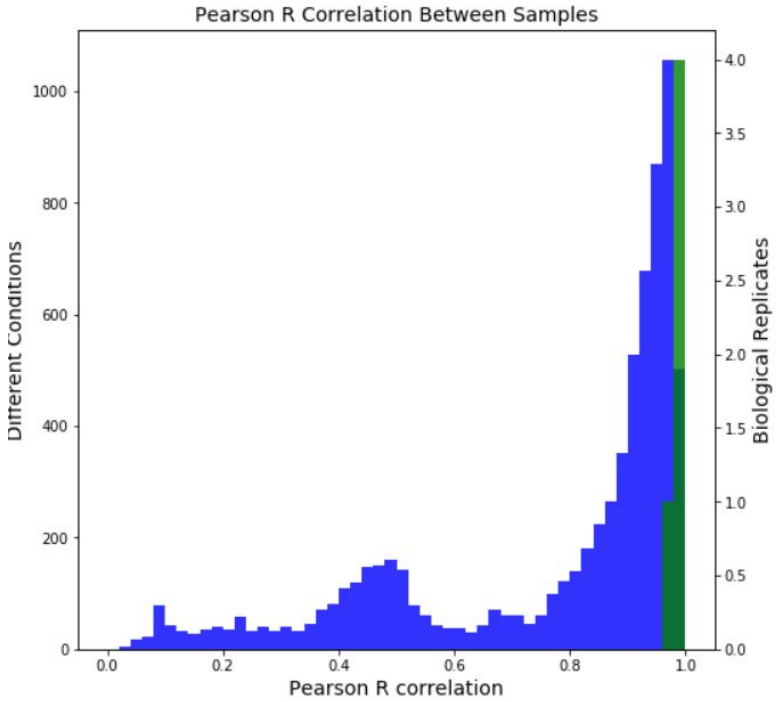
Correlations

Non-replicate pairs with highest correlations:

	Sample 1	Structure 1	Substructure 1	Ontology 1	Hemisphere 1	Sample 2	Structure 2	Substructure 2	Ontology 2	Hemisphere 2	Correlation
0	S020134_L4.LB22	FL	MFG	MFG-s	L	S020291_L8.LB15	FL	MFG	MFG-i	L	0.996198
1	S020190_L6.LB5	PL	SMG-i	SMG-i	L	S020348_L8.LB16	PL	SMG-i	SMG-i	R	0.995369
2	S020198_L2.LB10	TL	MTG	MTG-i	L	S020262_L8.LB20	PL	AnG-i	AnG-i	L	0.994884
3	S020024_L8.LB22	FL	orlFG	orlFG	L	S020094_L2.LB6	FL	OrbGyri	MOrG	L	0.994715
4	S020038_L3.LB8	FL	MFG	MFG-s	R	S020235_L2.LB5	FL	PCLa-i	PCLa-i	L	0.994557

Pairs with the lowest correlations:

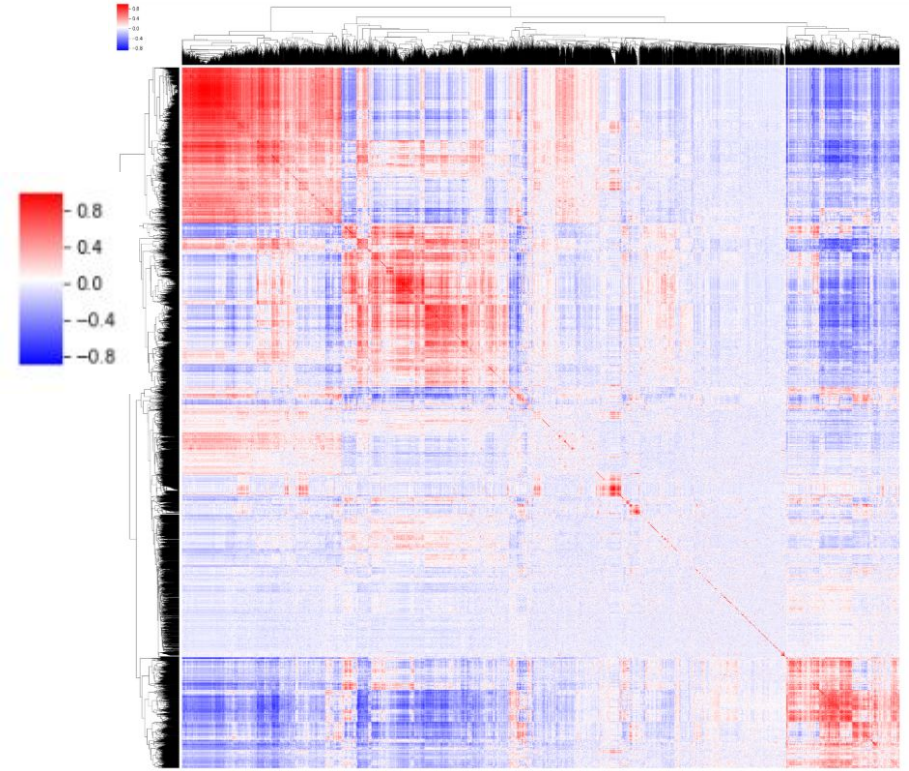
	Sample 1	Structure 1	Substructure 1	Ontology 1	Hemisphere 1	Sample 2	Structure 2	Substructure 2	Ontology 2	Hemisphere 2	Correlation
0	S020215_L4.LB23	Str	Putamen	Pu	R	S020722_L4.LB25	CbCx	CbCx	He-Crus II	L	0.022171
1	S020215_L4.LB23	Str	Putamen	Pu	R	S020656_L7.LB18	CbCx	CbCx	He-VIIIA	R	0.022912
2	S020215_L4.LB23	Str	Putamen	Pu	R	S020671_L7.LB16	CbCx	CbCx	PV-IV	R	0.023257
3	S020215_L4.LB23	Str	Putamen	Pu	R	S020697_L1.LB3	CbCx	CbCx	PV-VIIB	L	0.024174
4	S020215_L4.LB23	Str	Putamen	Pu	R	S020671_L7.LB16b	CbCx	CbCx	PV-IV	R	0.025753
5	S020206_L6.LB7	Str	Putamen	Pu	L	S020722_L4.LB25	CbCx	CbCx	He-Crus II	L	0.041097
6	S020206_L6.LB7	Str	Putamen	Pu	L	S020697_L1.LB3	CbCx	CbCx	PV-VIIB	L	0.042085
7	S020206_L6.LB7	Str	Putamen	Pu	L	S020656_L7.LB18	CbCx	CbCx	He-VIIIA	R	0.044123
8	S020055_L3.LB12	Str	Caudate	HCd	R	S020722_L4.LB25	CbCx	CbCx	He-Crus II	L	0.047550
9	S020109_L3.LB13	FL	SFG-m	SFG-m	L	S020671_L7.LB16	CbCx	CbCx	PV-IV	R	0.047973

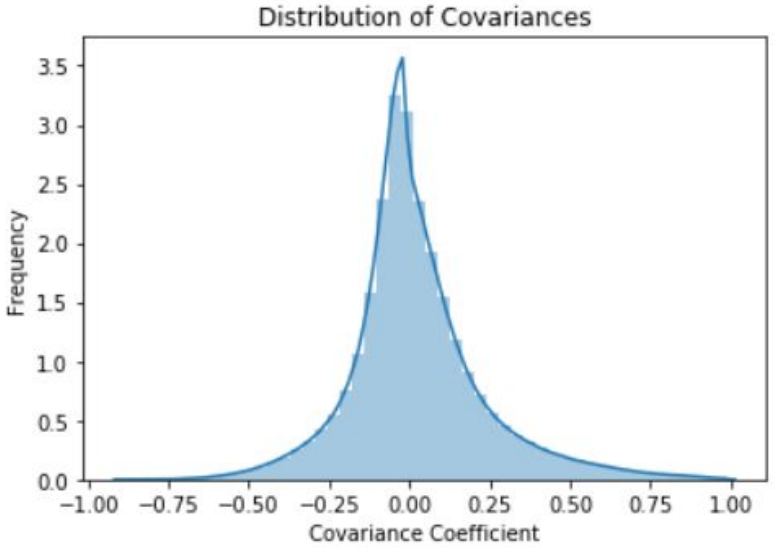


Replicate Correlations:

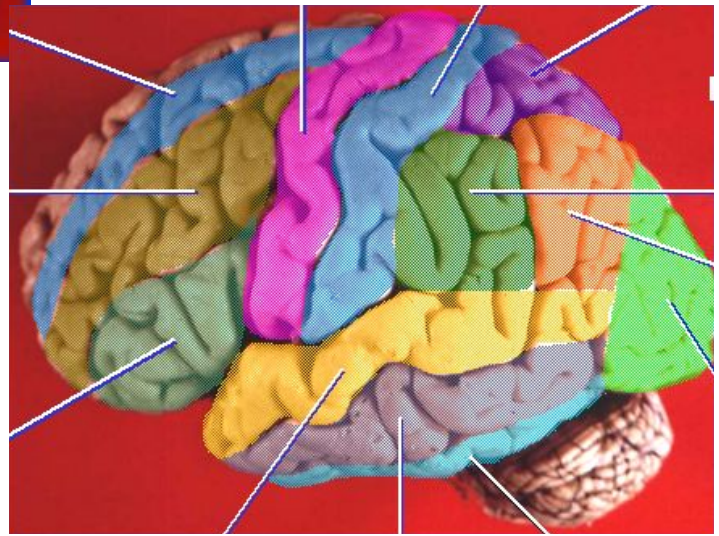
```
{'S020173': 0.9609132285955708,
 'S020181': 0.9918103371114042,
 'S020183': 0.9908425213183575,
 'S020237': 0.9945104162802172,
 'S020671': 0.9927718829425117}
```

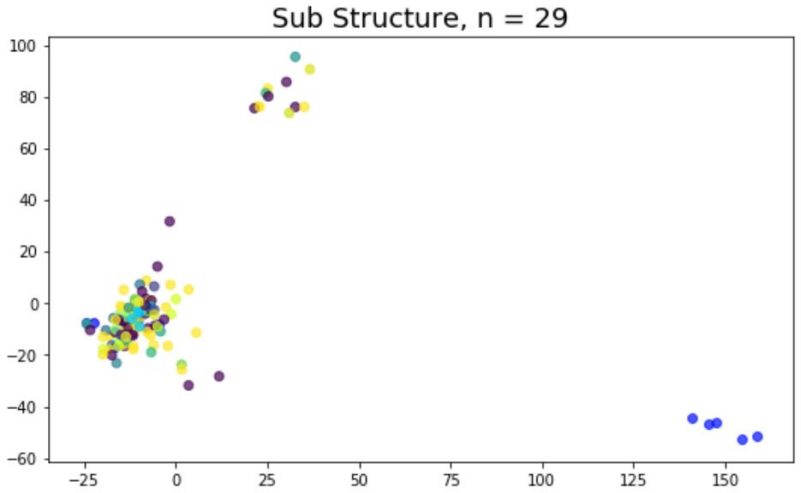
Covariance Matrix

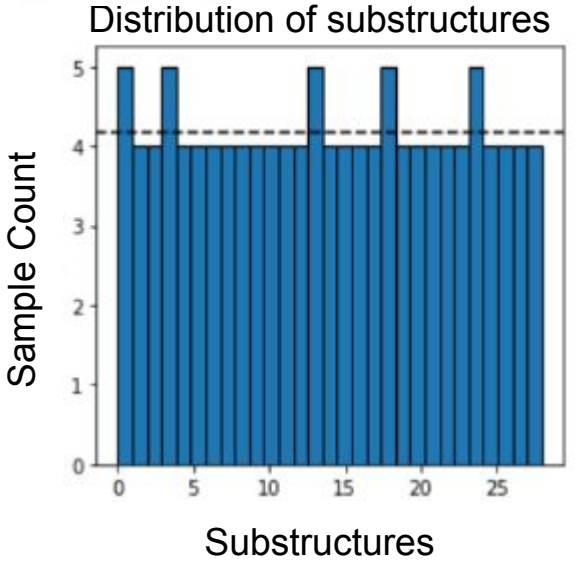




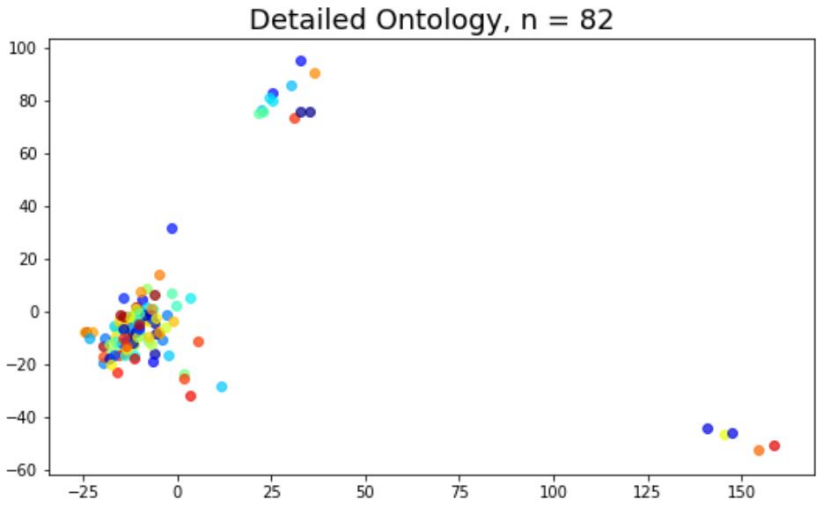
Substructures

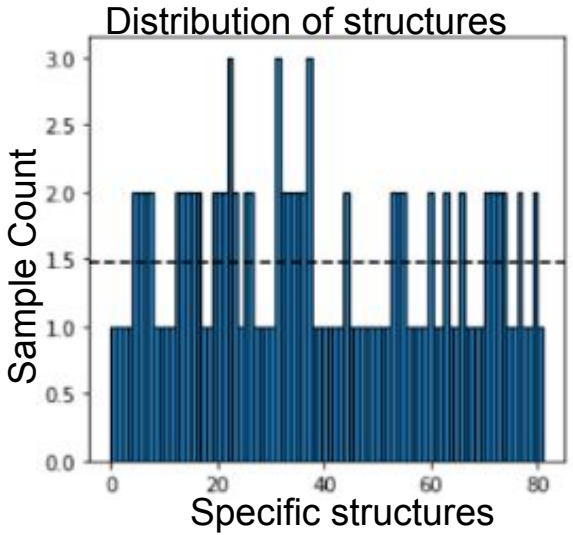




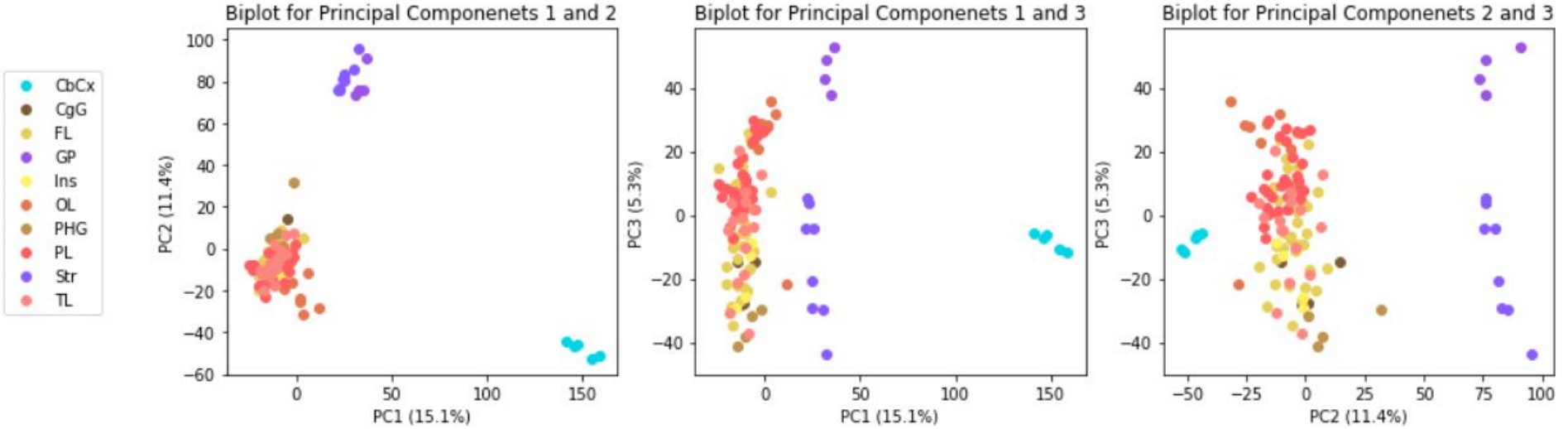


Specific Structures





PCA



Gene	Description	Gene Ontology Annotations
NR2E1	Nuclear Receptor Subfamily 2 Group E Member 1	DNA-binding transcription factor activity, enzyme binding
DACH2	Dachshund Family Transcription Factor 2	DNA-binding transcription factor activity, transcription factor activity, RNA polymerase II core promoter sequence-specific binding involved in preinitiation complex assembly.
RANBP3L	RAN Binding Protein 3 Like	nuclear export factor
EIF4E1B	Eukaryotic Translation Initiation Factor 4E Family Member 1B	RNA binding and translation initiation factor activity
LOC283143	Long Non-Protein Coding RNA	non-coding protein region

Cortex PCA Genes

Gene	Description	Gene Ontology Annotations
ABCB4	ATP Binding Cassette Subfamily Member 4	ATPase activity, ATPase activity coupled to transmembrane movement of substances
ACOX2	Acyl-CoA Oxidase 2	Signaling receptor binding, oxidoreductase activity acting on the CH-CH group of donors
ACER1	Alkaline ceramidase 1	Hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amides, dihydroceramidase activity
ABCA12	ATP Binding Cassette Subfamily A Member 12	Signaling receptor binding, ATPase activity coupled to transmembrane movement of substances
ADAD2	Adenosine Deaminase Domain Containing 2	RNA binding, adenosine deaminase activity

Multinomial LR Genes (3 Classes)

	coef	|coef|
RORC	0.002476	0.002476
TFAP2B	0.002468	0.002468
DEFB1	0.002468	0.002468
GCOM1	0.002467	0.002467
C7orf16	0.002466	0.002466
KRT31	0.002464	0.002464
PAX2	0.002463	0.002463
PCP2	0.002454	0.002454
SCNN1G	0.002454	0.002454
BARHL1	0.002452	0.002452

	coef	|coef|
LHX8	0.004613	0.004613
SFTA3	0.004611	0.004611
ECEL1	0.004548	0.004548
SDS	0.004443	0.004443
HPSE2	0.004361	0.004361
NKX2-1	0.004295	0.004295
APOC1	0.004129	0.004129
GBX2	0.004045	0.004045
LOC150381	0.004016	0.004016
FABP6	0.003981	0.003981

	coef	|coef|
LHX8	-0.004049	0.004049
SFTA3	-0.003985	0.003985
ECEL1	-0.003976	0.003976
NCAPG	-0.003906	0.003906
GBX2	-0.003868	0.003868
KCNE1L	-0.003866	0.003866
FAM180B	-0.003853	0.003853
MPPED1	0.003839	0.003839
SDS	-0.003824	0.003824
INSRR	-0.003817	0.003817

Multinomial LR Genes (3 Classes)

Gene	Description	Gene Ontology Annotations
RORC	RAR Related Orphan Receptor C	DNA-binding transcription factor activity, steroid hormone receptor activity
TFAP2B	Transcription Factor AP-2 Beta	DNA-binding transcription factor activity, sequence-specific DNA binding
DEFB1	Defensin Beta 1	Hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amides, dihydroceramidase activity
GCOM1	FRINL1A Complex Locus 1	Readthrough transcription variation
C7orf16	Protein Phosphate 1 Regulatory Subunit 17	Microbicidal and cytotoxic peptide activity

Gene	Description	Gene Ontology Annotations
LHX8	LIM Homeobox 8	Sequence-specific DNA binding
SFTA3	Surfactant Associated 3	Metabolism activity
ECEL1	Endothelin Converting Enzyme Like 1	metalloendopeptidase activity, metallopeptidase activity
SDS	Serine Dehydratase	protein homodimerization activity, L-serine ammonia-lyase activity
HPSE2	Heparanase 2	heparan sulfate proteoglycan binding, heparanase activity

Gene	Description	Gene Ontology Annotations
LHX8	LIM Homeobox 8	Sequence-specific DNA binding
SFTA3	Surfactant Associated 3	Metabolism activity
ECEL1	Endothelin Converting Enzyme Like 1	metalloendopeptidase activity, metallopeptidase activity
NCAPG	Non-SMC Condensin I Complex Subunit G	binding
GBX2	Gastrulation Brain Homeobox 2	DNA-binding transcription factor activity, sequence-specific DNA binding

Random Forest Top Genes (3, 10, 29, 82)

	coef	|coef|
MYOZ1	0.011939	0.011939
TFCP2L1	0.011648	0.011648
HRK	0.011129	0.011129
BUB1	0.010904	0.010904
DNAJC5G	0.010798	0.010798
TRIB3	0.010715	0.010715
NHLH2	0.010085	0.010085
C21orf128	0.009653	0.009653
NCRNA00246B	0.009576	0.009576
TRIM54	0.009400	0.009400

	coef	|coef|
ATP2C2	0.005976	0.005976
RSPH10B2	0.005390	0.005390
CTXN3	0.005369	0.005369
BTK	0.005274	0.005274
MAB21L1	0.005003	0.005003
DUSP4	0.004884	0.004884
KRT31	0.004673	0.004673
SLC5A7	0.004138	0.004138
ONECUT2	0.004100	0.004100
BCL11B	0.003971	0.003971

	coef	|coef|
CTXN3	0.002611	0.002611
LXN	0.002586	0.002586
METTL7B	0.002558	0.002558
CELSR1	0.002386	0.002386
CHRM2	0.002080	0.002080
LRRC38	0.001942	0.001942
SLN	0.001911	0.001911
ZSCAN5B	0.001897	0.001897
ST8SIA2	0.001861	0.001861
FAM46C	0.001857	0.001857

	coef	|coef|
HAPLN3	0.001519	0.001519
LCT	0.001421	0.001421
ADAMTSL5	0.001421	0.001421
FLJ42351	0.001384	0.001384
RGPD3	0.001365	0.001365
C2orf54	0.001311	0.001311
AOX1	0.001282	0.001282
REM1	0.001187	0.001187
FMOD	0.001174	0.001174
ICAM5	0.001168	0.001168