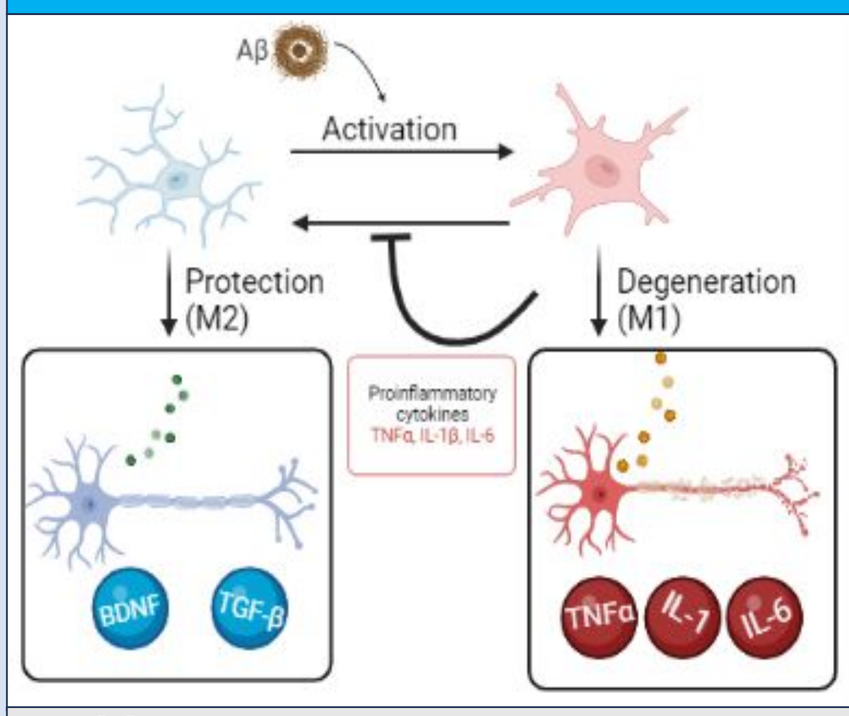
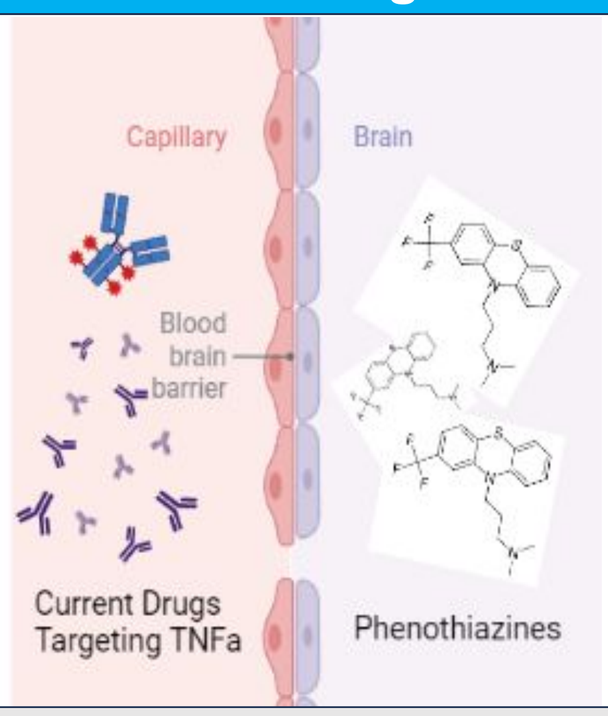
Novel Drug Discovery Methodology using Machine Learning:

Gene Expression-Based Virtual Screening Predicts Novel Compounds to Reverse Alzheimer's Disease with Applications to Cancer and Longevity by Inhibiting CtBP2 Expression

Background





TNF-α and Microglia in Alzheimer's Disease

- **Microglia**: brain immune cells activated into the following two types:
  - **M1 phenotype:** promotes **neuroinflammation**
    - **Pro-inflammatory molecules** and cytokines, such as **TNF-α**
    - Production of **neurotoxic** reactive oxygen species (ROS) and nitric oxide (NO) from conversion of arginine
    - **Glycolytic** state
  - **M2 phenotype:** promotes neuroprotection, **neuronal repair**
    - Anti-inflammatory molecules such as TGF-beta, BDNF
    - Production of neuronal repair factors such as collagen from conversion of arginine
    - **OxPhos** state
- **Healthy immune response:** M1 microglia kill pathogens using the pro-inflammatory cytokines and phagocytosis, then transition to M2
- **Neuroinflammation Hypothesis:** Aβ activates microglia to M1, which induces high levels of **TNF-α** that in turn form a positive feedback loop of sustained M1 state. This sustained M1 activation leads to neurodegeneration and Alzheimer's Disease.

Problems: **Problem 1**: how to find candidate drugs?

- Target- or phenotype-based assays
  → High throughput phenotypic assays show **phenothiazines** were highly effective in reducing TNF-a.

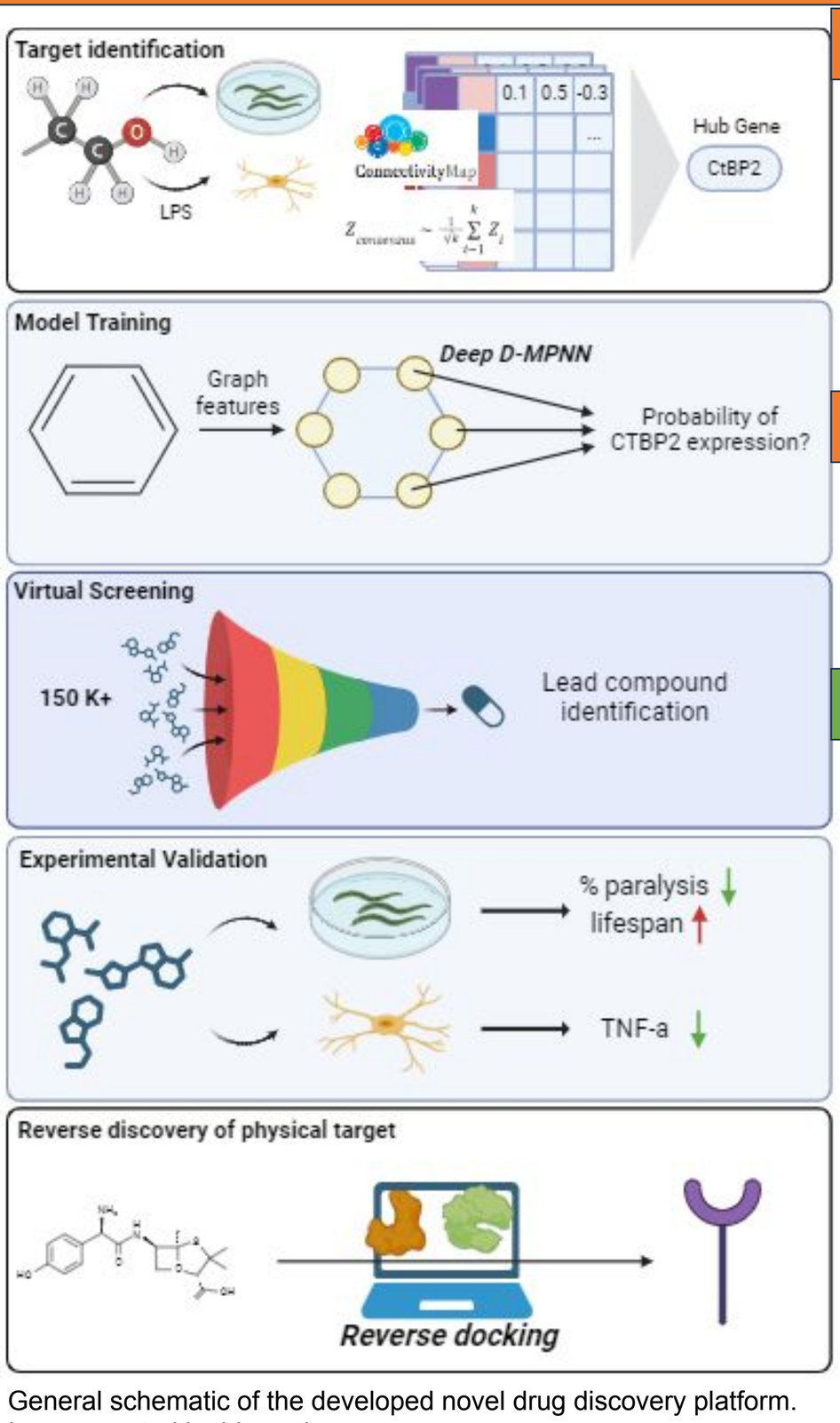
Problem 2: how do phenothiazines work?

- Pharmacogenomic analysis
- Discovery of CtBP2 as a common transcriptional target

Problem 3: CtBP2 is too complex, traditional methods of structure binding is impossible. (~98% "undruggable" targets)

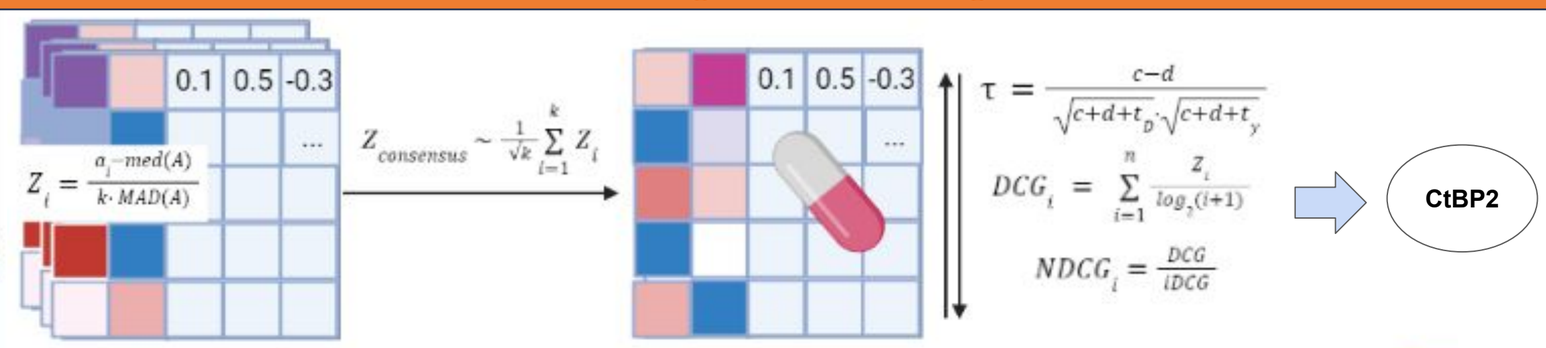
Objective: Using machine learning, screen (predict) for compounds that inhibit CtBP2 gene **expression**.

Methods



General schematic of the developed novel drug discovery platform. Image created by biorender.com

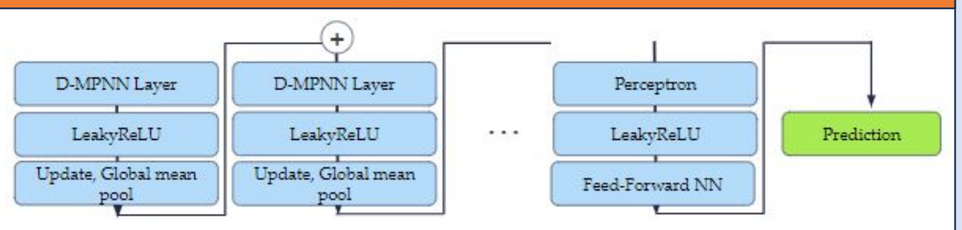
Pharmacogenomic algorithm



D-MPNN-BorderSMOTE algorithm

General MPNN Layer

$$H = \phi\left(x_u, \bigoplus_{v \in N_u} (\psi(x_u, x_v))\right)$$

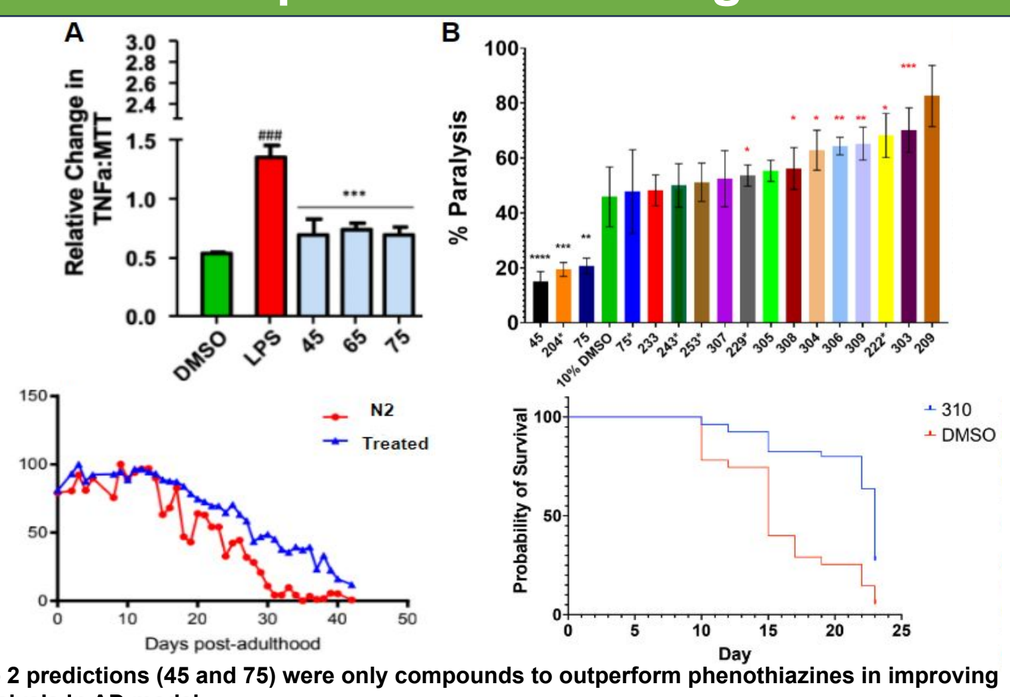


Model Training Results

	Model Type	R2
Regression	RF	Undefined
	XGBoost	0.00
	Deep NN (ECFP)	0.01
	GAT	0.01
	D-MPNN-BorderSMOTE	0.02

	Model Type	AUC-ROC	MCC, CV	MCC, Test
Classification	RF	0.59	0.00	0.00
	XGBoost	0.56	0.22	0.21
	Deep NN (ECFP)	0.63	0.01	0.00
	GAT	0.56	0.01	0.00
	D-MPNN-BorderSMOTE	**0.97**	**0.912544**	**0.914848**

Experimental Testing



- **Top 2 predictions (45 and 75) were only compounds to outperform phenothiazines in improving paralysis in AD model**
- **Lead compound 310 normalizes TNF-a and increases lifespan**

Implications

1. **Therapeutic potential of undruggable targets (which constitutes 98% of human proteins) are now unlocked**
2. **Developed a suite of ML algorithms which allow discovery of novel small molecules to treat any diseases**
   - Addresses problem where expression data is not translated to therapeutics well
3. **Equitable access of drug discovery for all labs to create novel small molecules to treat previously incurable diseases**
4. **Breadth of application can readily expand to any other incurable diseases such as cancer.**
5. **Future work will focus on diabetic complications, Parkinson's disease, and certain cancer types where transcriptional abnormalities are involved**