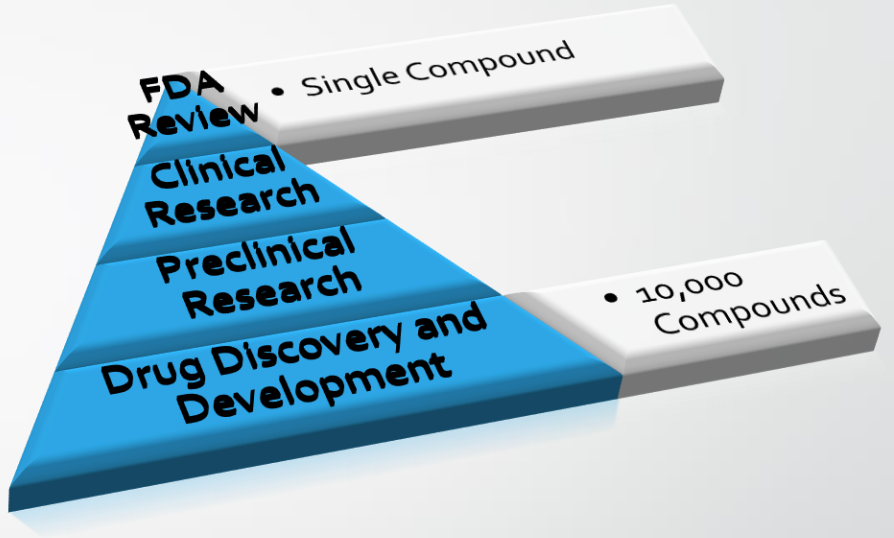
Computational Visualization Method for Drug Discovery

Student: Shy Alon

Supervisor: Dr. Abraham Yosipof

Motivation



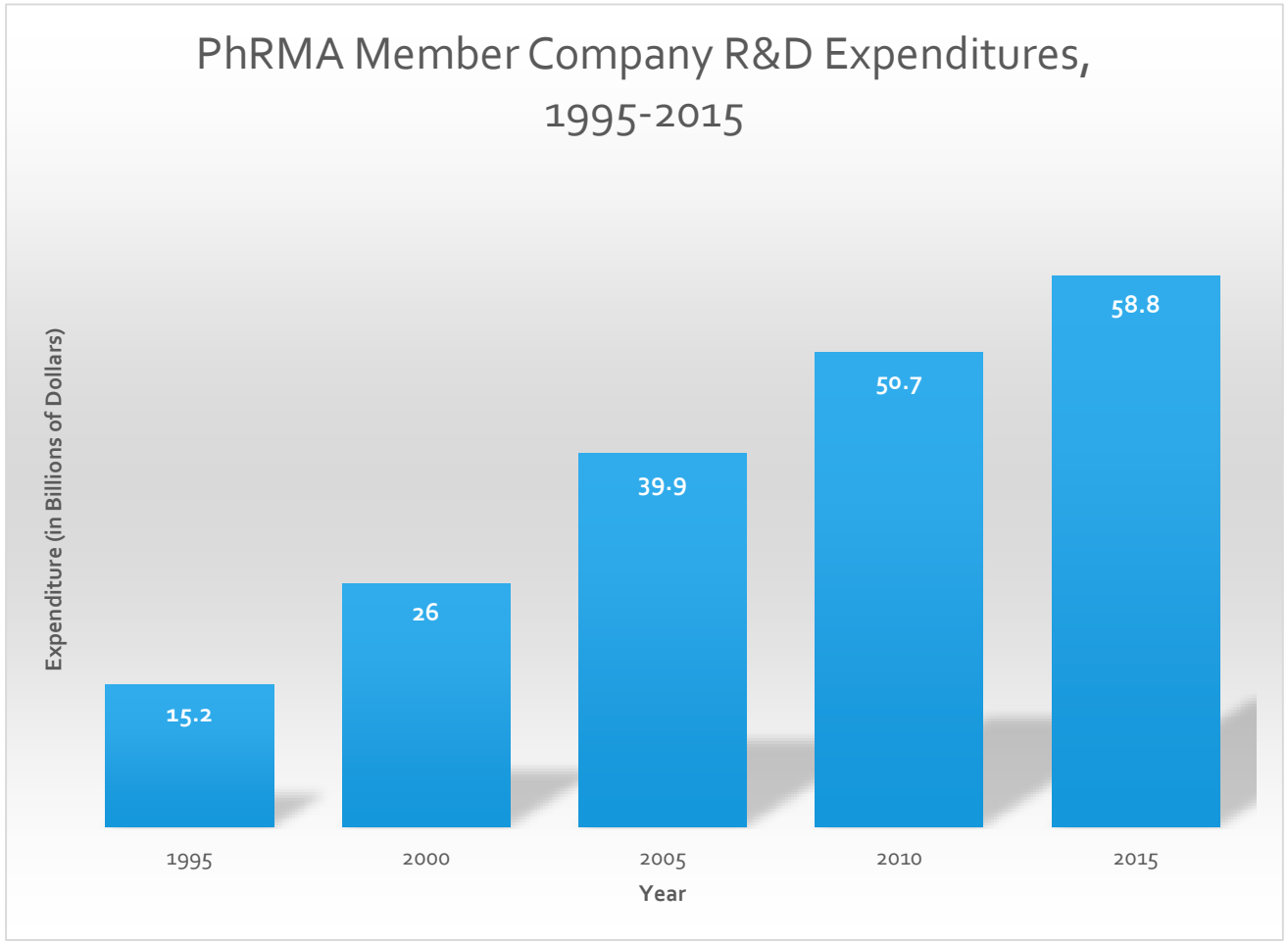
- The cost of a new drug is estimated at 5.5 billion USD
- Out of which the discovery costs 1.25 billion USD

Research Hypothesis

- There exists a computational cost effective method for the drug discovery stage.
- Such a method will be able to predict the biological activity of a compounds prior to their synthesis

Application

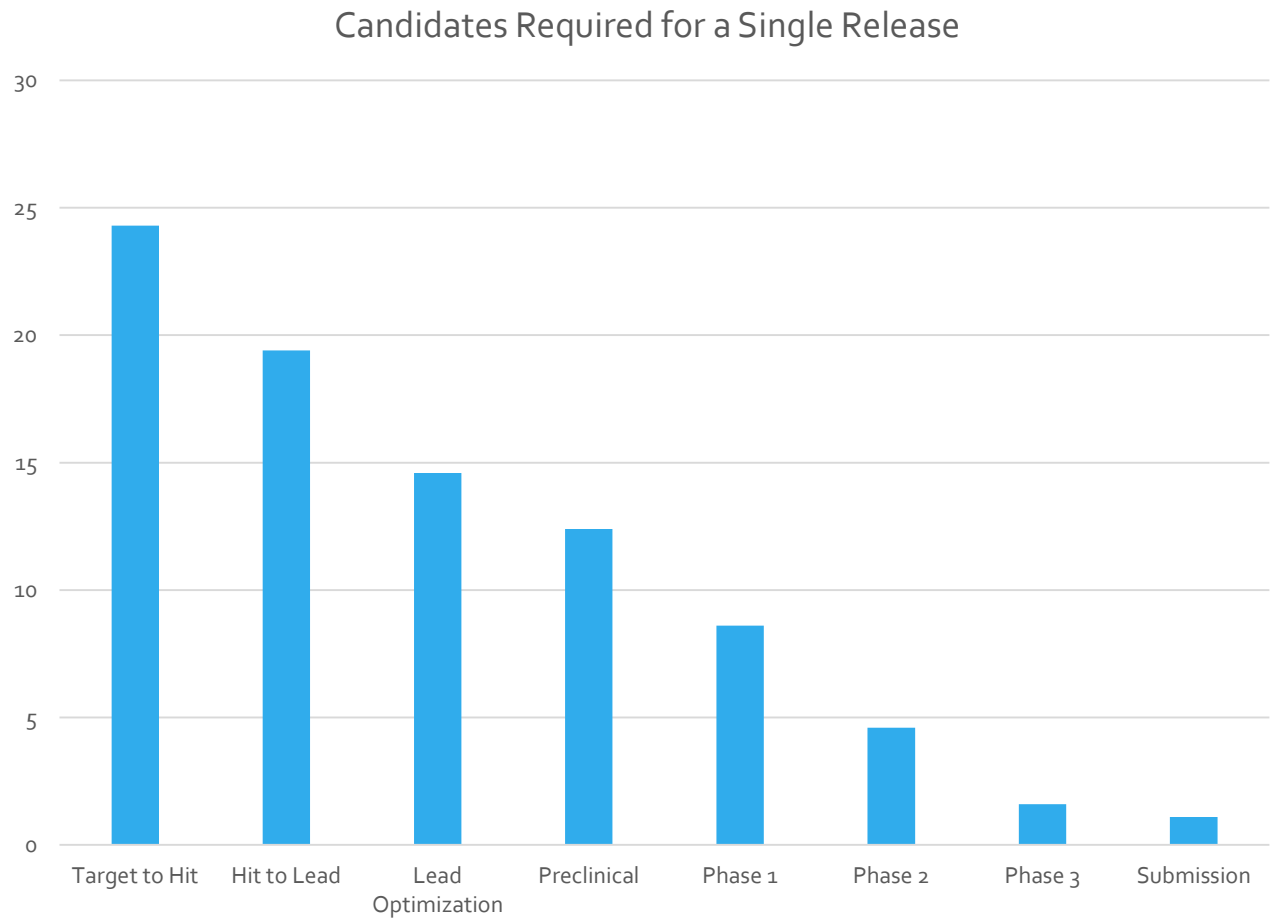
- G-protein-coupled receptors (GPCRs) are the largest single superfamily of proteins
- GPCRs represent ~45% of current drug targets and thus have excellent potential for drug discovery



Pharma Research

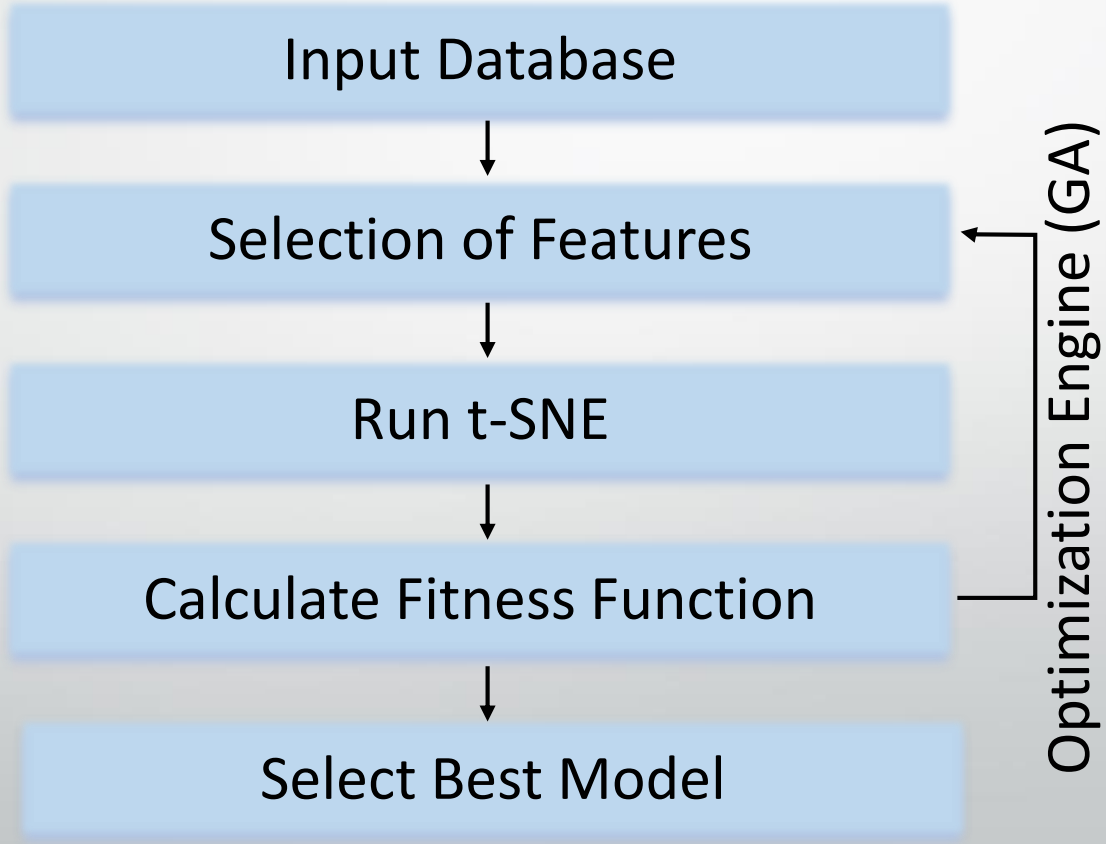
Pharma Research

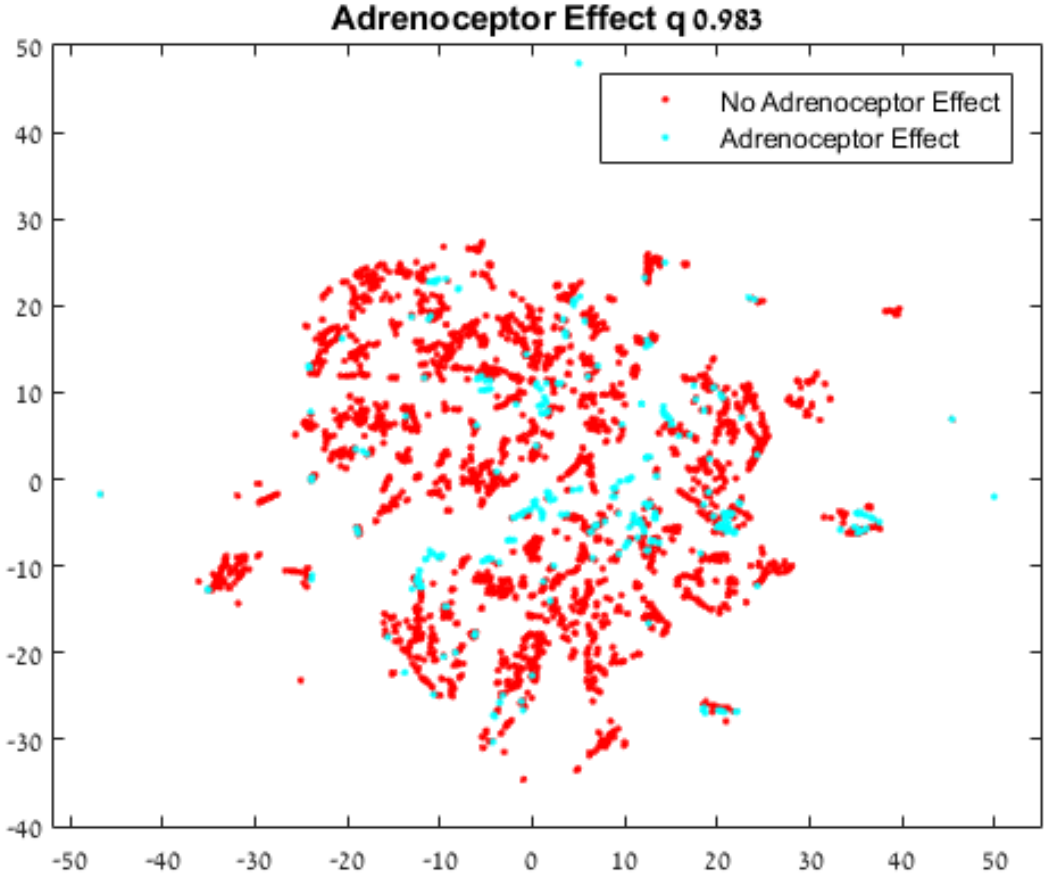
- 30% of total cost of an approved new compound is due to the drug discovery phase and the pre-clinical phase.
- It takes 24.3 targets to create a hit and 19.4 hits to create a lead - one out of 471 targets become a lead.
- These numbers can be significantly improved.



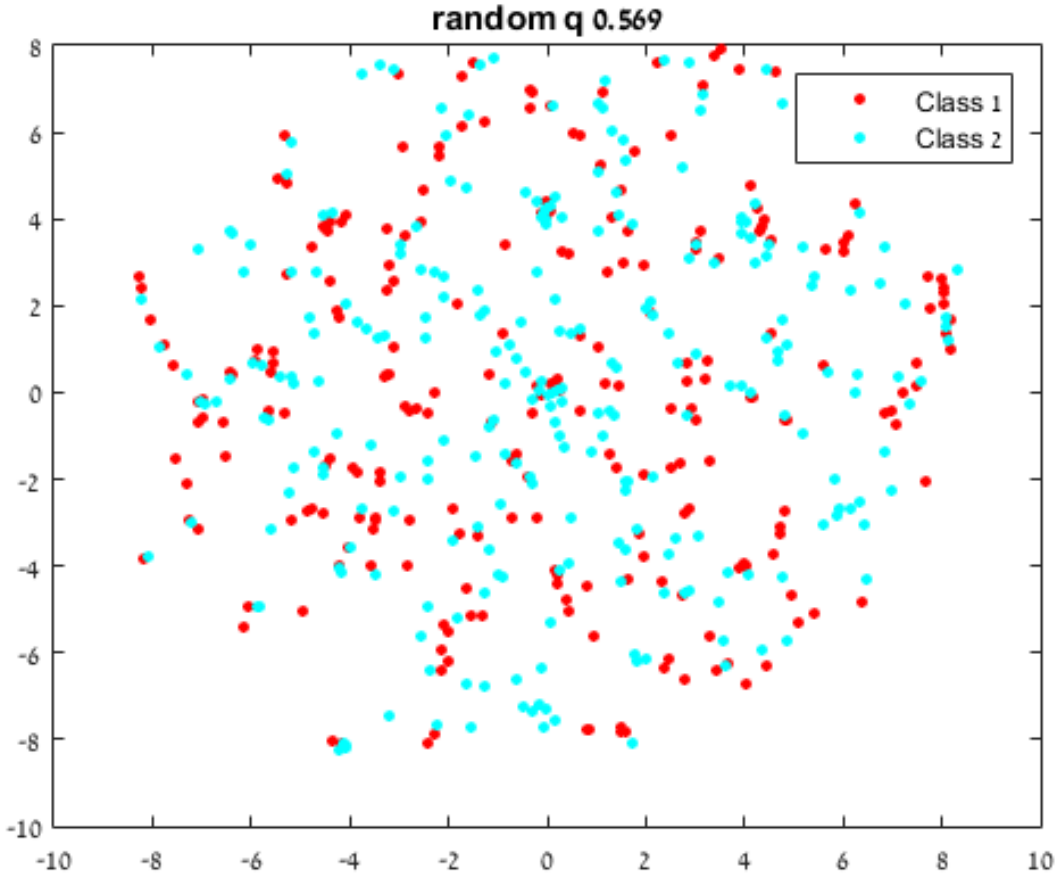
Pharma Research

The Model


Input Database
Selection of Features
Run t-SNE
Calculate Fitness Function
Select Best Model
Optimization Engine (GA)



Result Example



Verification – Random Result

Result Summary

	Q-1	*Q-3*	*Q-5*	*Trust*
Dopamine	0.965	0.965	0.961	0.087
Adrenoceptors	0.983	0.985	0.984	0.077
Histamine	0.967	0.963	0.964	0.078
Muscarinic	0.993	0.993	0.993	0.076
Serotonin	0.938	0.934	0.921	0.090
Average	0.9692	0.9680	0.9646	0.0816

Validation

- A compound with unknown effect is very likely to have the same effect as their nearest neighbor.
- Choosing by a nearest neighbor has a very high probability of choosing a compound with the same effect.
- Compounds are classified by nearest neighbor.

Conclusions

- An algorithm for drastically reducing costs for pre-clinical phase has been demonstrated as effective.
- Limitations exist because of algorithm complexity (time and space).

Future Work

- Distributed implementation of the algorithm
- Use of lower complexity alternatives to t-SNE