Using Unsupervised Machine Learning Techniques and 3D Visualization Tools to Better Understand Cardiovascular Disease

Kevin McCoy, David Kartchner, Stephen Allegri, Michael Davis, Cassie Mitchell

Motivation

Can machine learning techniques be leveraged to solve long-standing biomedical problems?

Background

30 million

PubMed indexed articles

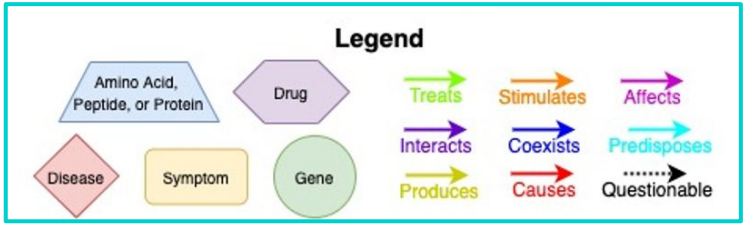
3,000

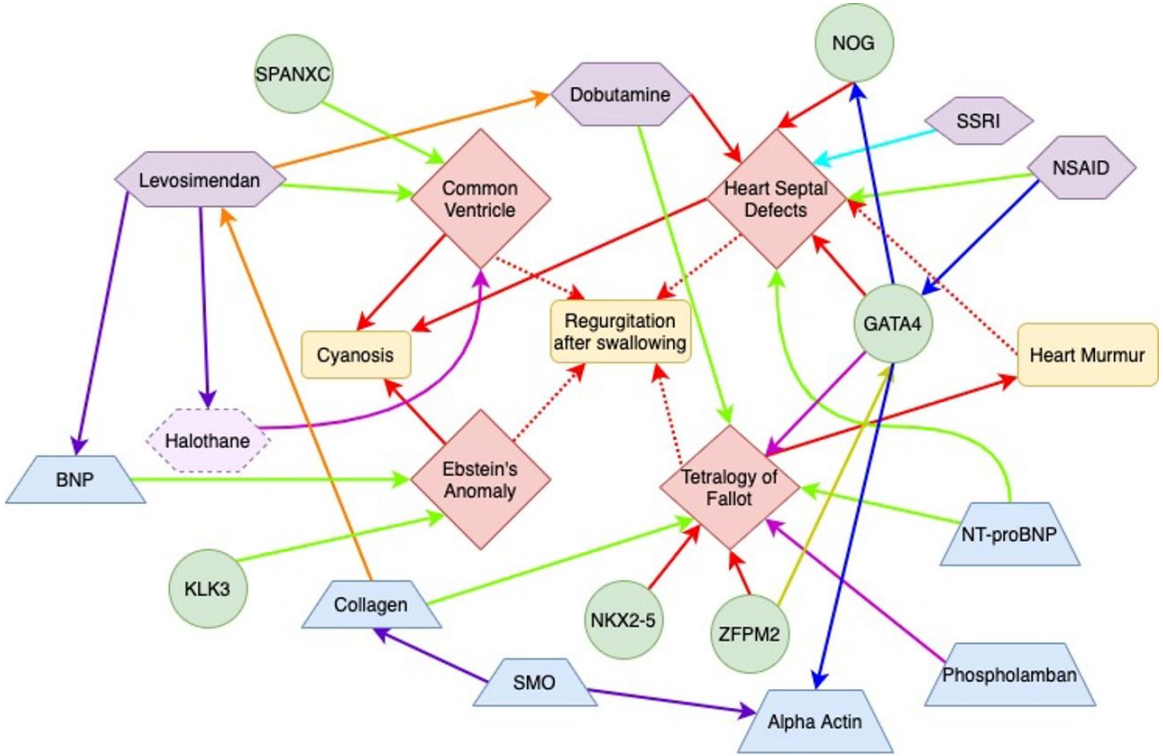
New articles published to PubMed every day

2 million

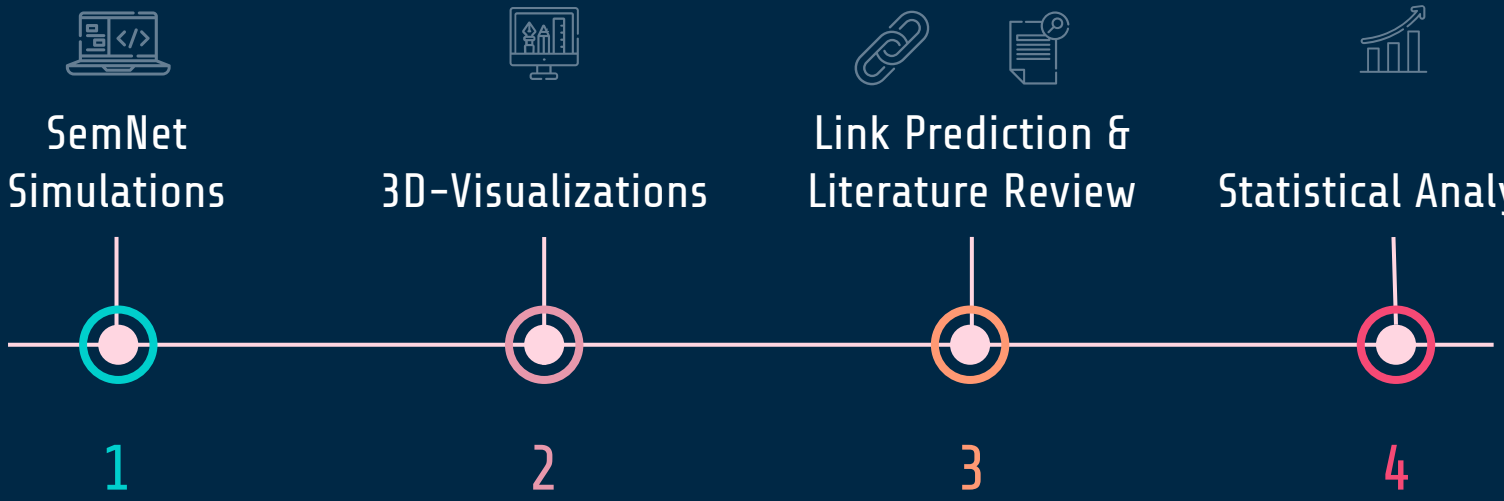
PubMed articles relating to cardiovascular disease

SemNet Knowledge Graph





Project Timeline



SemNet Simulations

Amino Acids, Peptides, and Proteins

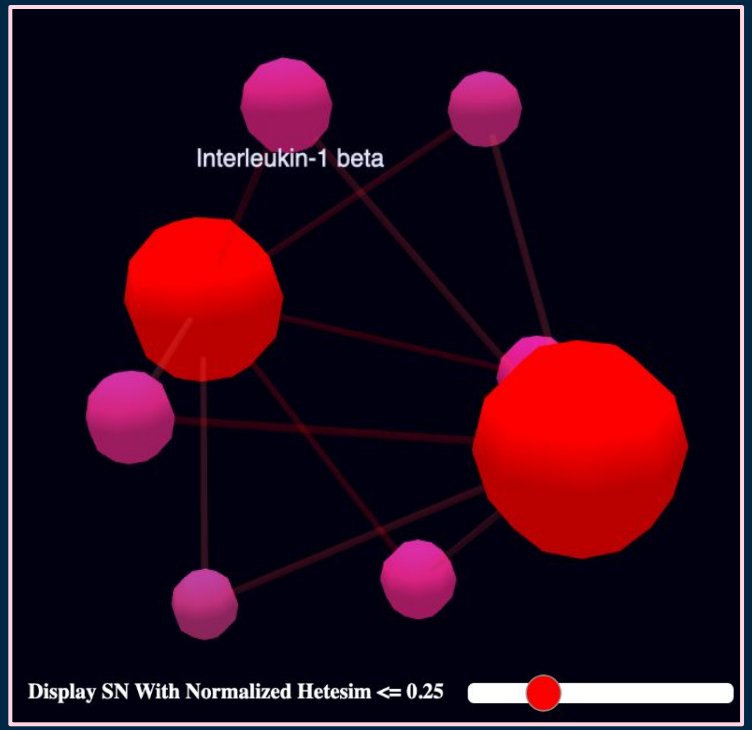
- Inflammation + Angiogenesis
- Inflammation + Fibrosis
- Inflammation + Ejection Fraction

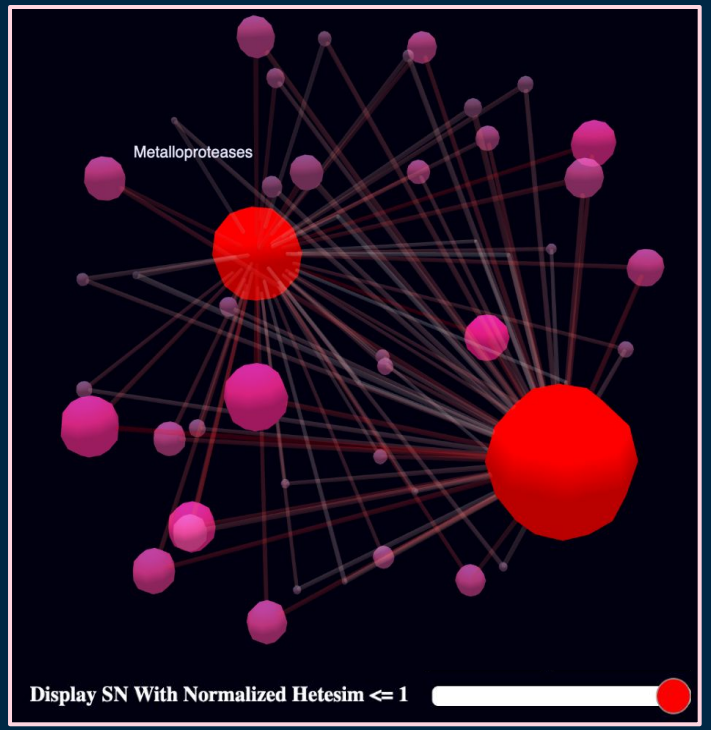
Genes

- Inflammation + Angiogenesis
- Inflammation + Fibrosis
- Inflammation + Ejection Fraction

Example Results

	hetesim	count	dwpc	novelty
VEGF protein, human|VEGFA	1.878639	1.754198	2.036228	0.494895
Transforming Growth Factor beta	2.165431	1.868706	2.238096	0.470761
cytokine	2.197022	1.266344	1.804237	0.374406
TGFB1 protein, human|TGFB1	2.215703	2.054812	2.365146	0.491828
Interleukin-1 beta	2.233374	1.792378	2.244316	0.449281
NOS3 protein, human|NOS3	2.234343	2.141264	2.422347	0.502297
Interleukin-6	2.26039	1.428502	1.944651	0.389925
Collagen	2.316587	1.862808	2.272523	0.447948
Tumor Necrosis Factor-alpha	2.323183	1.967266	2.304573	0.462906
NF-kappa B	2.331665	1.953982	2.318746	0.459653
Matrix Metalloproteinases	2.338937	2.242177	2.477714	0.502507
Fibroblast Growth Factor 2	2.347991	2.428073	2.470615	0.529517
cyclooxygenase 2	2.358848	2.35965	2.494467	0.517518
Gelatinase B	2.367822	2.362102	2.563111	0.516591
NOS2A protein, human|NOS2	2.414956	2.345532	2.574755	0.507233
receptor	2.466775	1.482516	2.074689	0.368232
chemokine	2.539888	2.048699	2.418544	0.443895
Angiotensin II	2.560165	2.215863	2.539162	0.466424
Interleukin-8	2.562655	2.11479	2.434365	0.450663
Gelatinase A	2.571594	2.599605	2.695811	0.523233
Integrins	2.588631	2.421104	2.695769	0.493567
Tumor Necrosis Factor-alpha|TNF	2.642429	2.518037	2.690563	0.500535
Cell Adhesion Molecules	2.666065	2.489679	2.685474	0.492788
Adiponectin	2.721662	2.641276	2.76277	0.507825
TRANSCRIPTION FACTOR	2.752101	2.075896	2.498444	0.417271





3D Visualizations

Relation prediction

Given two entities predict the relation

Query Params: | HEAD ENTITY: Metalloproteases | TAIL ENTITY: Congestive heart failure | SIZE: 4 | METHOD | D

Filter: Search...

RELATION	NOVEL	RANK	SCORE
ASSOCIATED_WITH_AAPPaswtDSYN	yes	0	1.0
CAUSES_AAPPcausDSYN	yes	1	1.42
PREDISPOSES_AAPPpredDSYN	yes	2	1.19
DISRUPTS_AAPPdrptDSYN	yes	3	1.16

Link Prediction

National Library of Medicine
National Center for Biotechnology Information

Log in

PubMed.gov

metalloproteases angiogenesis inflammation

Search

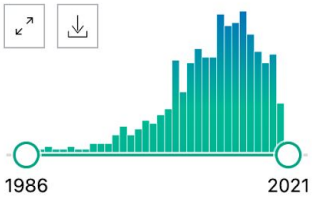
Advanced Create alert Create RSS User Guide

Save Email Send to Sorted by: Best match Display options

MY NCBI FILTERS

526 results

RESULTS BY YEAR



TEXT AVAILABILITY

- Abstract
- Free full text

1. Matrix metalloproteinases: regulators of the tumor microenvironment.
Kessenbrock K, Plaks V, Werb Z.
Cell. 2010 Apr 2;141(1):52-67. doi: 10.1016/j.cell.2010.03.015.
PMID: 20371345 **Free PMC article.** Review.
In cancer, altered proteolysis leads to unregulated tumor growth, tissue remodeling, **inflammation**, tissue invasion, and metastasis. The matrix metalloproteinases (MMPs) represent the most prominent family of proteinases associated with tumorigenesis. ...In addition to thei ...

2. Biochemical and Biological Attributes of Matrix Metalloproteinases.
Cui N, Hu M, Khalil RA.
Prog Mol Biol Transl Sci. 2017;147:1-73. doi: 10.1016/bs.pmbts.2017.02.005. Epub 2017 Mar 22.
PMID: 28413025 **Free PMC article.** Review.
MMPs play a role in tissue remodeling during various physiological processes such as **angiogenesis**, embryogenesis, morphogenesis, and wound repair, as well as in pathological conditions such as myocardial infarction, fibrotic disorders, osteoarthritis, and cancer. ...MMPs a ...

3. A disintegrin and **metalloproteases**: molecular scissors in **angiogenesis**, **inflammation** and atherosclerosis.

Cite Share

Literature Review

Statistical Analysis

- Compare the HeteSim scores between link prediction labels and literature review labels to assess the performance of the SemNet model.

- Highlight particular nodes of interest that have significant yet unexplored relevance to cardiovascular disease.

Future Work

Q&A

Thank you!