An ML-Enhanced Platform Leveraging Stem Cell Secretomes to Identify Therapeutic Proteins

Rohit R. Jadhav[1], Edward J. Moler[1], Tejaswini Ganapathi[1], Munir Yousef[1], Roie Levy[1], Thach Mai[1], Jeremy D. O'Connell[1], Hanadie Yousef[1]

[1]Juvena Therapeutics, Redwood City, CA. juvenatherapeutics.com

Introduction

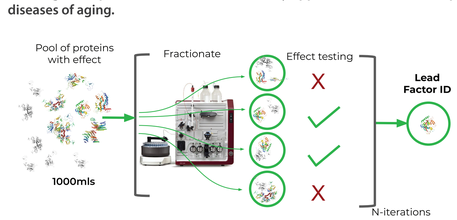
Stem cell secreted proteins are increasingly appreciated to drive regenerative effects across a multitude of tissues and cell types. An important application of this effect is leveraging it to identify and develop biologics that elicit rejuvenation in specific biological processes and systems that undergo degeneration in aging, disease, or developmental conditions.

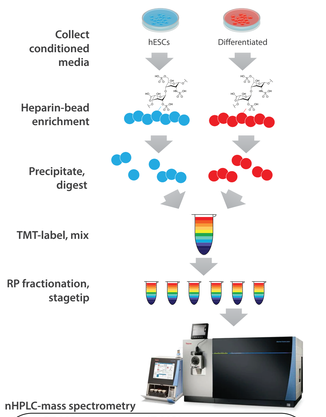
To accelerate the identification and verification of proteins with targeted tissue and disease modifying effects, we a) created a stem cell secretome protein library by combining LC-MS3 experiments with phenotype screening, b) generated and assessed a feature database combining sequence-derived metrics, tissue and cell-type specific expression patterns, biochemical properties, and evolutionary conservation scores for each protein in the library, and c) applied these feature sets as inputs to predictive algorithms designed to enrich the early success of library screening experiments.

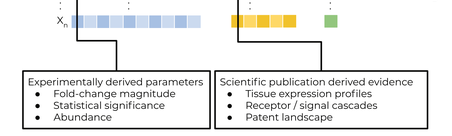
This strategy led to the identification of multiple candidate therapeutic factors. Juvena's first lead candidate, JUV-161, is currently in pre-IND studies for Myotonic Dystrophy type 1. We have built on that strategy and developed a platform, consisting of an integrated wet-lab with compounding knowledgebase and predictive modeling framework, that supports a growing number of disease/tissue areas. Multiple candidates in several rare and degenerative disease areas are under investigation. Our unique platform, built around developing insights into stem-cell secreted proteins through a combination of predictive modeling, empirical testing for phenotypic effects, and biological knowledge, drives insights into the pathways and candidate therapeutics for the treatment of diverse diseases.

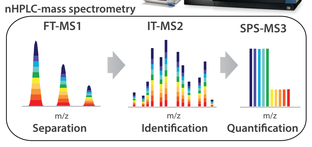
Methods

Deploying recent advances in quantitative proteomics methodologies we overcome limitations inherent in the hESC secretome, allowing a comprehensive, unbiased discovery approach to rank ordered proteins for high-through put screening as therapies for diseases of aging.

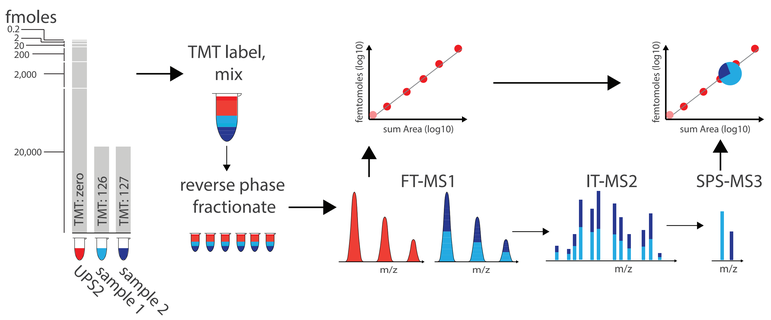


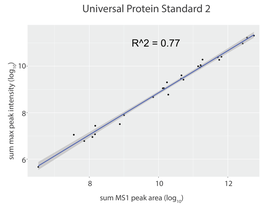


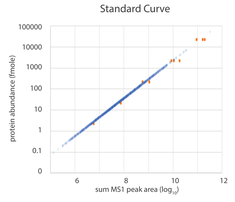


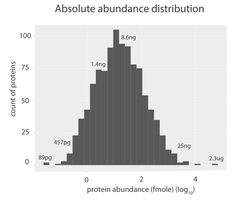


We developed and deployed a new, multiplexed method for sample agnostic absolute abundance measurement (SAAM) with good linearity over 5 orders of magnitude in abundance, spanning from femtomolar to low micrograms simultaneously across multiple samples (below).



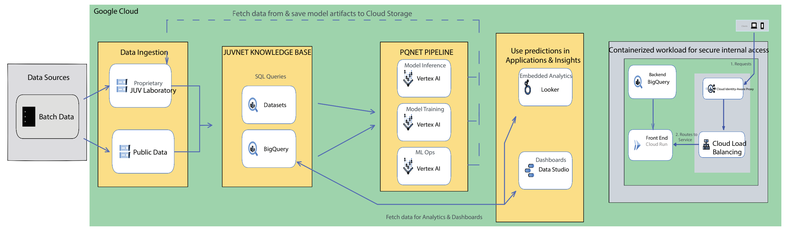




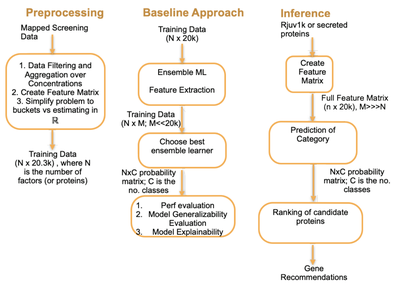


Methods

JuvNET Knowledgebase & ML Architecture enables prediction enriched target identification



PQNET predictive modelling flow



Modelling Strategy

- Model as Classification problem - multi-model outcome distributions with long tails.
- Ensemble Learning Algorithms to facilitate "soft predictions" to aid our Bayesian ranking model.
- Ensembles based on sampling technique: Adaboost+trees, XGBoost, Random Forests, ExtraTrees.

Model generated screening recommendations accelerate hit identification.

Results

orphaNet rare disease classification	Library Proteins
genetic diseases	329
neurological diseases	155
developmental anomalies during embryogenesis	146
ophthalmic disorders	93
transplant-related disorders	91
inborn errors of metabolism	65
bone diseases	61
skin diseases	48
systemic and rheumatological diseases	41
endocrine diseases	40
renal diseases	40
hematological diseases	35
neoplastic diseases	31
immunological diseases	29
otorhinolaryngological diseases	28
odontological diseases	24
abdominal surgical diseases	19
cardiac diseases	19
gynecological and obstetric diseases	19
infertility disorders	18
surgical maxillo-facial diseases	16
hepatic diseases	15
surgical thoracic diseases	15
gastroenterological diseases	12
circulatory system diseases	11
cardiac malformations	10
respiratory diseases	9
systemic or rheumatologic diseases of childhood	9
urogenital diseases	9
allergic disease	4

Selection of initial screening candidates

The intersection of the secretome library with gene expression GWAS and other genetic disease and pathway association data yields potential disease targeted screening candidates for model training.

This intersection is quantified for identifying screening candidates as shown in the example barplot.

Secretome Library Disease Associations

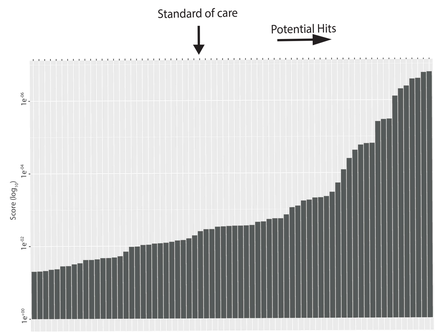
344 rJuv1k secretome proteins are directly associated with 578 rare diseases based on literature mining and curation

Of top three disease areas:

Enrichment of developmental anomalies in embryogenesis validates the enrichment of the secretome for factors involved in tissue generation and orchestration

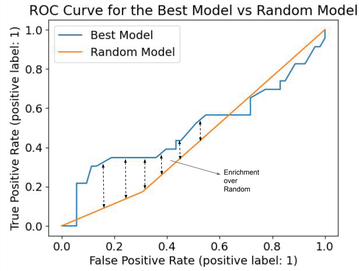
A notable number of rare degenerative diseases are classified as neurological, including Duchenne Muscular Dystrophy

Genetic Diseases are heavily represented in Orphanet database due to extensive research in that area.



Results

ROC Curve for the Best Model vs Random Model



Metric		Random Model	Best Model (model M1)
Number of targets correctly identified		4/23	10/23
Overall Accuracy		0.34	0.37
Performance Metrics corresponding to class corresponding to readout value > 1.2 (class of interest corresponding to "hits")	Precision	0.20	0.28
	Recall	0.17	0.43
	F-Score	0.19	0.34

Distance map of human proteome based on Juvena's curated feature library shows the clustering of protein families in human stem-cell secretomes.

Our model is able to predict hits across a range of diverse families of the secretome library.

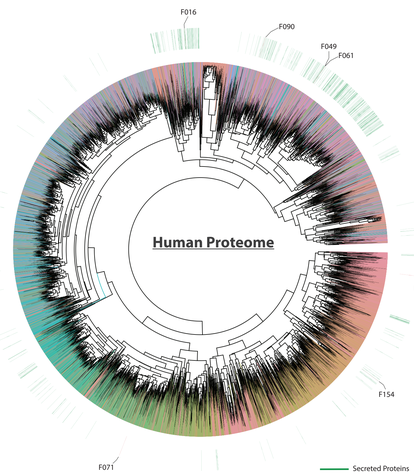
Predicted screening hits

The models are trained on the results of initial in-house screening runs.

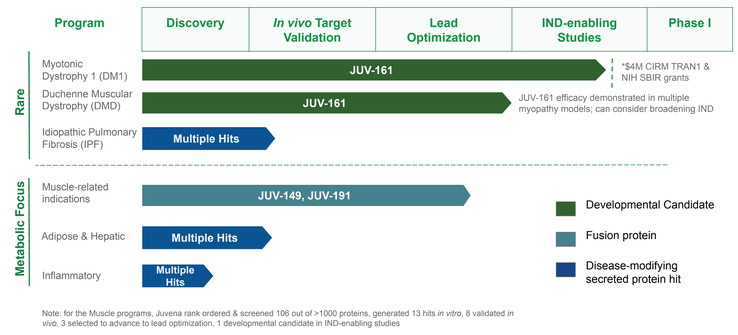
The trained models are then used to rank screening candidates for predicted disease modifying activity.

This accelerates the process for in-vitro and in-vivo efficacy testing in the disease models.

The landscape of Juvena's Library



Juvena's Pipeline



Note: for the Muscle programs, Juvena rank ordered & screened 106 out of >1000 proteins, generated 13 hits *in vitro*, 8 validated *in vivo*, 3 selected to advance to lead optimization, 1 developmental candidate in IND-enabling studies

Conclusions

Juvena's platform, knowledgebase and ML has enabled the identification of multiple candidates with significant disease modifying effects.

Juvena's unique discovery platform has generated a pipeline of Lead Candidates across multiple programs with an initial focus on rare and metabolic disease.

The therapeutic potential of these candidates are in various stages of pre-clinical assessment.