

GeneRIF Analysis

Genetics Research
and
Natural Language Processing

Dylan Craine

GeneRIFs

(RIF: Reference Into Function)

Examples for gene SIR2 (Silent Information Regulator 2):

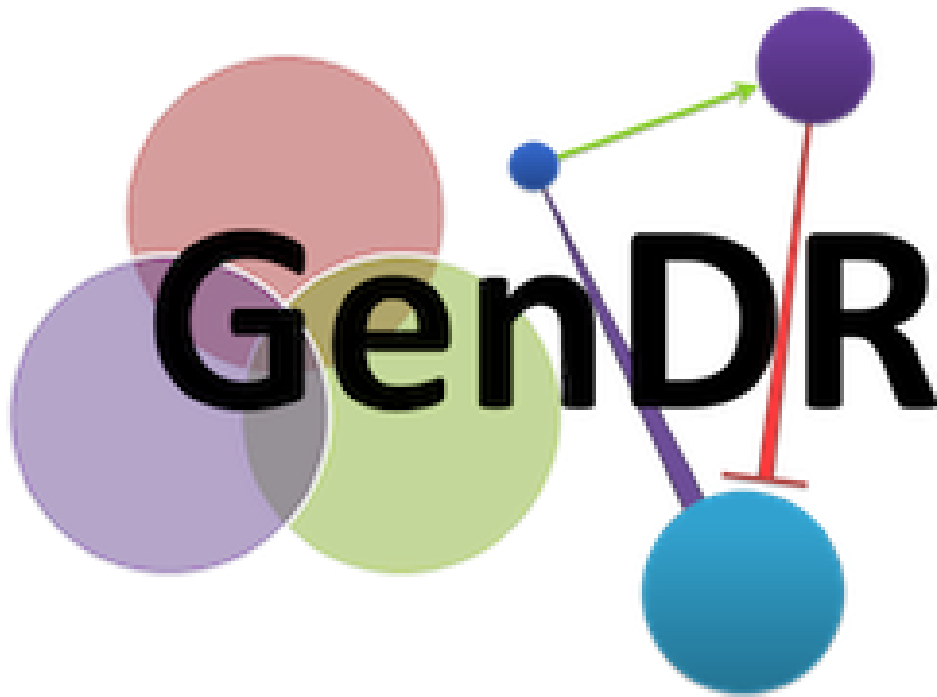
"sir2 and sir4, but not sir3, have roles in silencing of DNA breakage and recombination"

"Sir2p does not have a role in caloric restriction-augmented reactive oxygen species defense"

"Extends longevity by overexpression."

"We propose that lifespan extension by caloric restriction is independent of sirtuins and that nicotinamide has sirtuin-independent effects on lifespan extension by CR."

GenDR



Database of genes related to the longevity-enhancing effects of caloric reduction diets

Goal

- Use GeneRIF text to predict genes that might be relevant to the longevity effects of dietary restriction but that aren't in the GenDR database
- Find underlying patterns in the perceived significance of the GenDR genes that might produce insights into the shape of the research that has been conducted

Phase 1: Prediction with Naive Bayes

2018 GeneRIFs for genes in GenDR

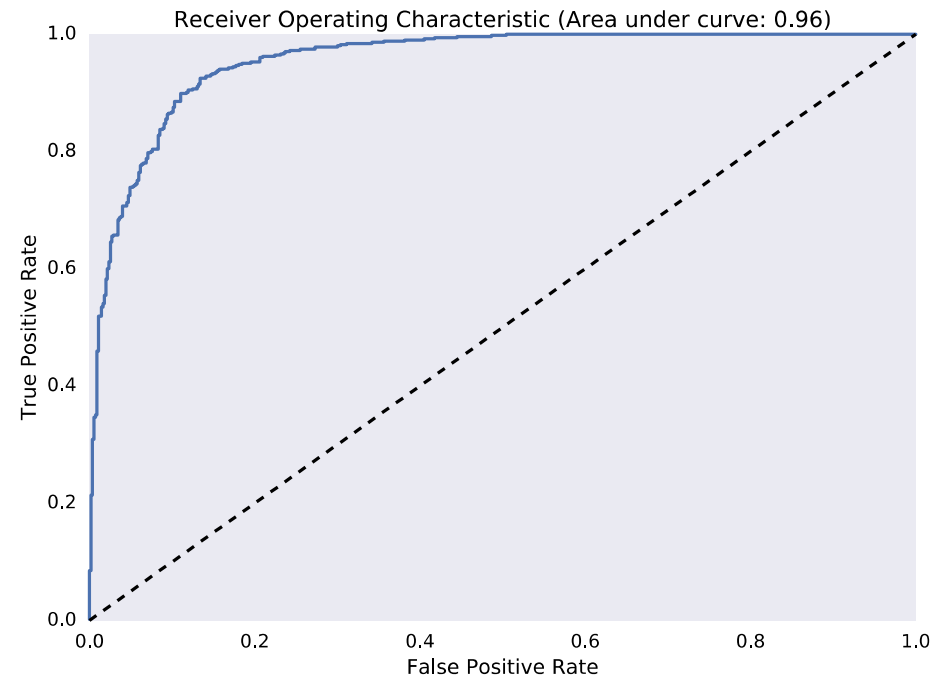
1106122 GeneRIFs overall

~2000 RIFs in non-GenDR training subset

Phase 1: Prediction with Naive Bayes

Accuracy: ~90%

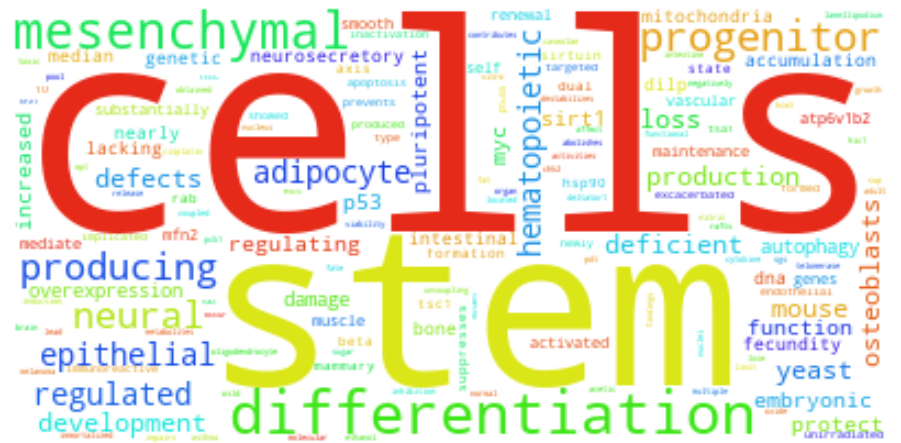
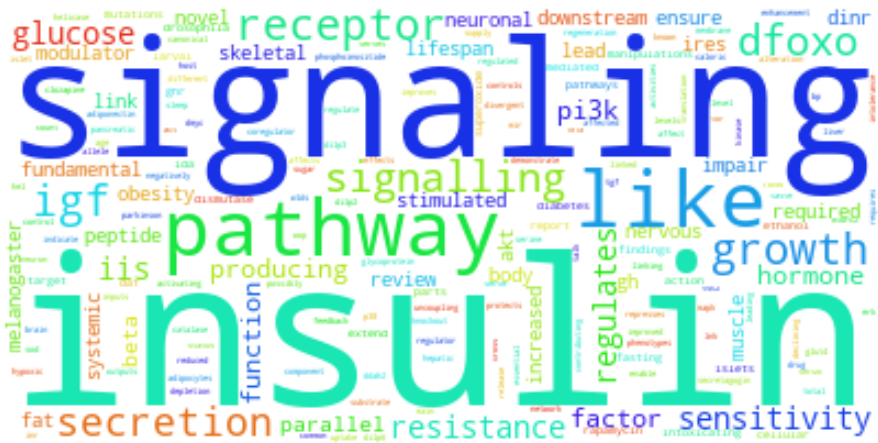
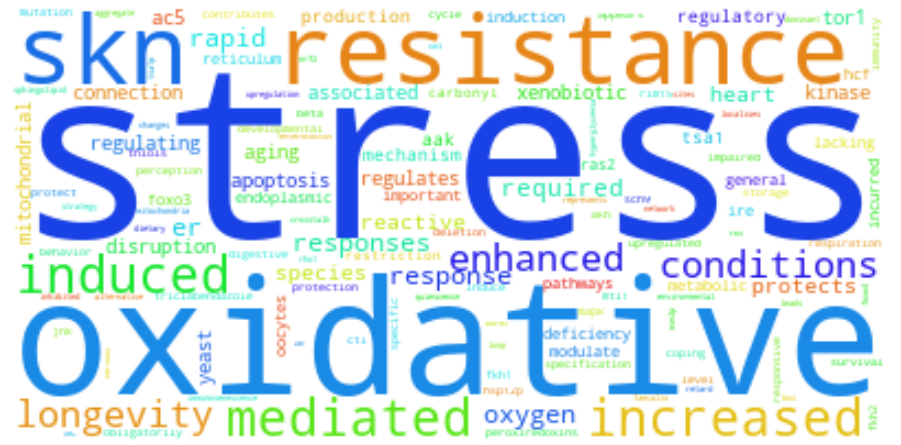
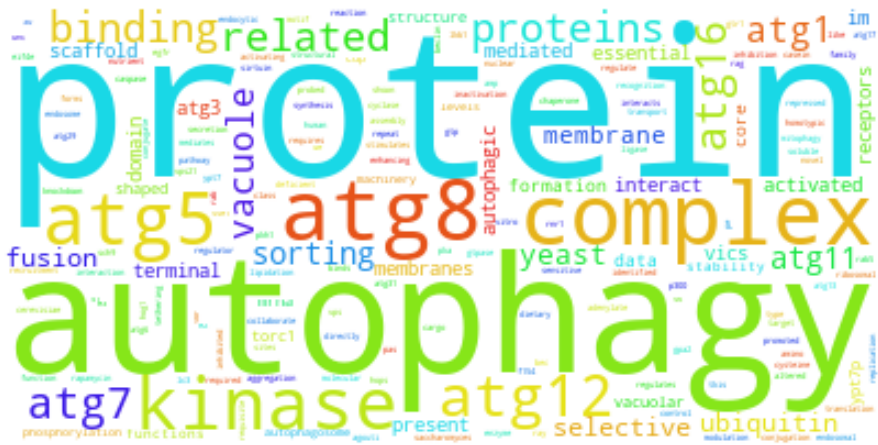
Found 40314 potential
DR-relevant genes out
of 88409 total



Phase 1: Prediction with Naive Bayes



Phase 2: Categorization with NMF



Data Sources:

GeneRIFs:

<https://www.ncbi.nlm.nih.gov/>

GenDR:

<http://genomics.senescence.info/diet/>

Contact Me:

Dylan Craine

dylan.craine@deepwatercreations.com

 github.com/deepwatercreations

 linkedin.com/in/dylancraine