Genome-wide identification of conserved non-coding elements associated with subterranean mammal phenotypes using deep learning

Elysia Saputra[1], Nathan L. Clark[1,2], Maria Chikina[1]

[1] Department of Computational and Systems Biology, University of Pittsburgh
[2] Department of Human Genetics, University of Utah
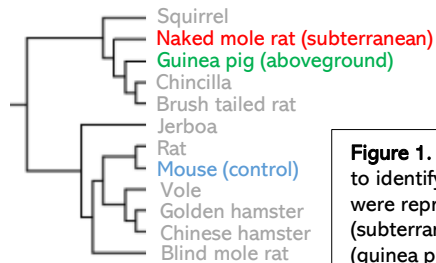
Introduction

Subterranean environment imposes selection pressures that drive the phenotypic adaptation of fossorial animals, which includes the convergent loss of ocular structures. Previous work in our lab have identified significant correlations between ocular functions and protein-coding elements with accelerated relative evolutionary rates in subterranean mammals (Partha, 2017). The same study also demonstrated convergent acceleration in evolutionary rates of ocular-specific transcriptional enhancers. In this work, we extended the investigation genome-wide to identify potential regulatory elements that are associated with subterranean adaptation.

Methods

Genome-wide Prediction of FANTOM and ENCODE signal tracks with *Basenji* (Kelley, 2019)

1-hot-encoded DNA sequence → Basenji (Deep Learning model) → Signal tracks (FANTOM, ENCODE)
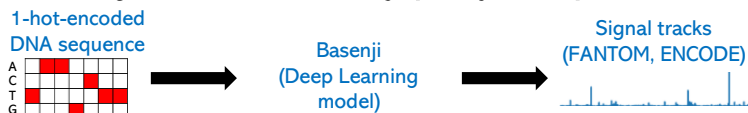
Prediction of the Activity of Conserved Non-coding Elements for 3 Species in the Rodent Clade

Squirrel
Naked mole rat (subterranean)
Guinea pig (aboveground)
Chincilla
Brush tailed rat
Jerboa
Rat
Mouse (control)
Vole
Golden hamster
Chinese hamster
Blind mole rat

Make predictions for:
320,299 conserved non-coding region (phastCons calls)

Figure 1. Use predicted signal tracks from Basenji to identify conserved non-coding regions that were repressed in subterranean mammal (subterranean) and not in aboveground mammal (guinea pig), relative to control (mouse).

Results

Example Prediction of Signal Track of the FANTOM Embryonic Mouse Eyeball (Day 12) Annotation at the Local Region of PAX6 Gene
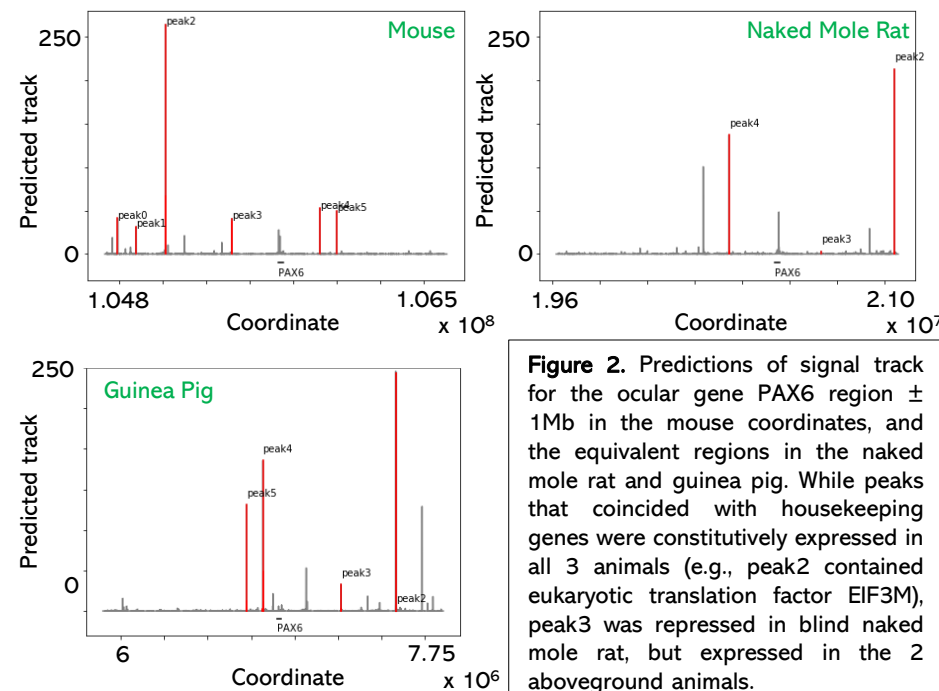

Figure 2. Predictions of signal track for the ocular gene PAX6 region ± 1Mb in the mouse coordinates, and the equivalent regions in the naked mole rat and guinea pig. While peaks that coincided with housekeeping genes were constitutively expressed in all 3 animals (e.g., peak2 contained eukaryotic translation factor EIF3M), peak3 was repressed in blind naked mole rat, but expressed in the 2 aboveground animals.

Identification of Conserved Non-coding Elements that are Significantly Repressed in Naked Mole Rat

16,787 Conserved non-coding regions suppressed in subterranean mammals

Functional enrichment (tool: *GREAT*)
- skin thickening
- eye degeneration
- abnormal kidney functions,
- and heart enlargement

Top Repressed Conserved Non-coding Elements Matched Known DNA Motifs of Regulatory Elements Consistent with the Subterranean Phenotype

DNA motif comparison tool: *STAMP* (Mahony & Benos, 2007)

Element	Alignment	Consensus sequence
TEAD1	TATCACATTCTTGCGCGACCA ---YACATTCCWSNG------	
Functional relevance: mutations causes chorioretinal atrophy		
ROX1	:CGGAGAACAAAGTCCC' --GAGAACAATNGR--	
Functional relevance: repressor of hypoxic gene		
Optix	TATCACAT' TATCA---	
Functional relevance: eye development		

Future Work

- Retrain deep learning model to optimize for cross-species predictions of enhancer regions
- Predict signal tracks of multiple functional annotations for all species in the phylogenetic tree
- Identify conserved non-coding elements that are convergently repressed in independent lineages of subterranean mammals

References

- R. Partha *et al.*, Subterranean mammals show convergent regression in ocular genes and enhancers, along with adaptation to tunneling. *Elife*. 6 (2017), doi:10.7554/eLife.25884
- D.Kelley, Cross-species regulatory sequence activity prediction. *bioRxiv* (2019), doi: 10.1101/660563
- S Mahony, PV Benos, "STAMP: a web tool for exploring DNA-binding motif similarities", *Nucleic Acids Research* (2007) 35(Web Server issue):W253-W258