Intrinsically Disordered Transcription Factors:

Correlating Sub-Nuclear Localization & Evolutionary Conservation

Amanda C. Ross[1], Colleen Hannon, Ph.D.[2], Michael Eisen, Ph.D.[2]

[1] San Jose City College, [2] University of California, Berkeley

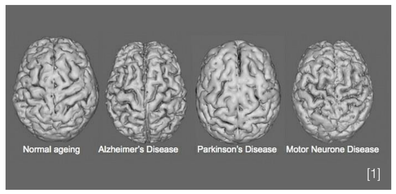
2018 Transfer-to-Excellence Research Experiences for Undergraduates Program (TTE REU Program)

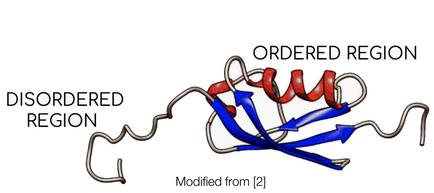
Abstract

Transcription factors (TFs) are proteins recognized for having a crucial role in gene expression. There is a great deal known about these proteins and their structured DNA binding domains. However, many TFs also contain intrinsically disordered regions (IDRs), the characterization of which has largely been neglected because they are flexible and unstructured, making them difficult to research. The IDRs of many TFs have been shown to play a role in regulating transcriptional activation, but how it affects this activation is relatively unknown. The goal of this project is to characterize these IDRs by comparing their specific amino acid sequences and sub-nuclear distributions in the *Drosophila melanogaster* embryo and determine if these features are correlated. *Drosophila* (the common fruit fly) is investigated because its uncommon embryonic development makes it easier to study nuclei, specifically the intrinsically disordered TFs inside the nuclei. IDR analysis may determine a correlation between amino acid sequences and TF localization within the nucleus that can be used for characterization of IDRs. The classification of different IDRs within *Drosophila* TFs will shed light on the intrinsically disordered proteomics as a whole.

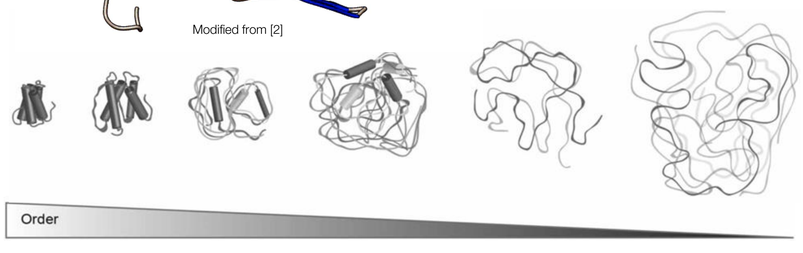
Overview

- Intrinsically disordered proteins are a contributor of numerous neurodegenerative diseases.
- By characterizing these proteins, through correlation of amino acid composition & localization of protein within nuclei and conservation of disordered regioned amino acid sequences between species, we may better understand the function and evolution of disordered regions.
- Proteins are on a spectrum based on their regions of order and disorder having varying levels of both parts.
- This research focuses on seven proteins of study that are primarily disordered: Da, Zelda, CG1620, Lilli, D1, Su(var)2-HP2, and Bap111.





Modified from [2]



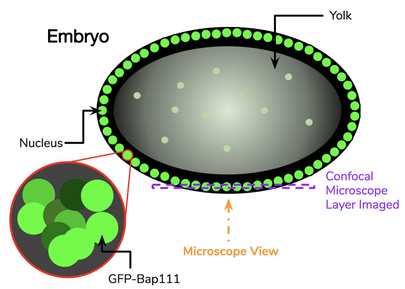
Modified from [3]

Contact Information
Amanda Ross
(408) 340-3722
amanda.cm.ross@gmail.com

Support Information
This work was funded by Centers for Excellence in Genomic Science of the National Institutes of Health Award RM1HG009490

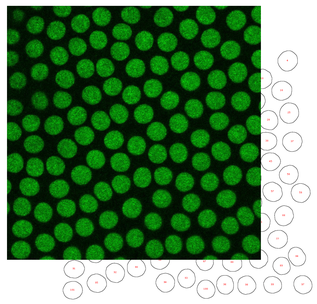
National Institutes of Health

Methods: Sub-Nuclear Localization



Images of protein taken with Confocal Fluorescence Microscope. Used Fiji, image analytics tool, to overlay masks over each nuclei to determine protein skewness.

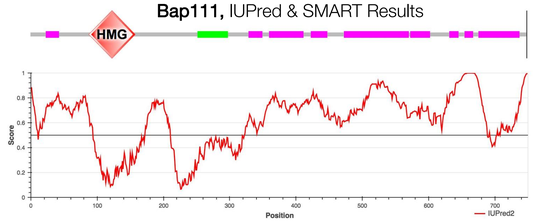
Bap111, Nuclear Cycle 13



	Skew
1	0.564
2	0.551
3	0.356
4	0.354
5	0.444
6	0.347
7	0.397
8	0.400
9	0.248
10	0.407
11	0.364
12	0.317
13	0.311
14	0.283
15	0.446
16	0.338
17	0.318
18	0.278
...	

Methods: Evolutionary Conservation

Bap111, IUPred & SMART Results



Amino acids make up proteins (like letters make up words). FruitiFly[4], an open-source Python program developed by Joseph Heady of the SJCC Robotics and Automation Club, automates the process of comparing disordered regioned amino acid sequences between species.

SMART

IUPred2A

Protein BLAST NIH protein ▸ protein

SJCC ROBOTICS & AUTOMATION

Jalview
A workbench for multiple sequence alignment and analysis

Results: Sub-Nuclear Localization & Evolutionary Conservation

Figure A

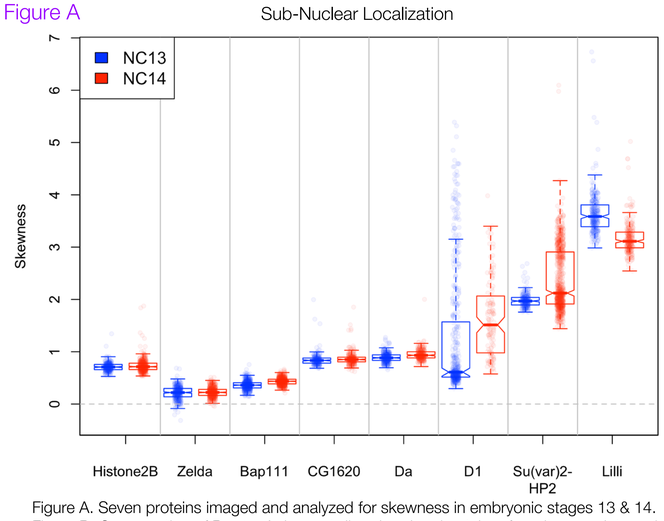


Figure B

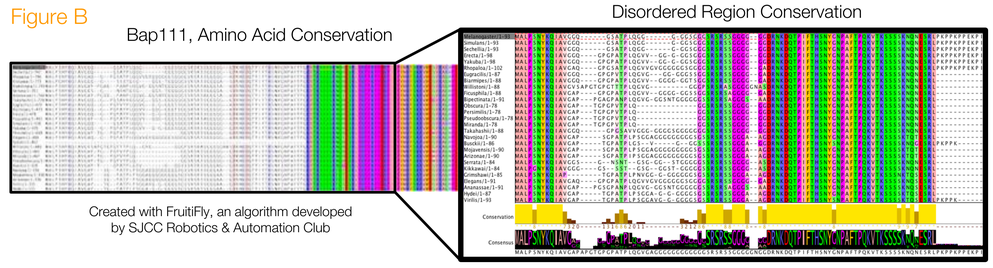


Figure C

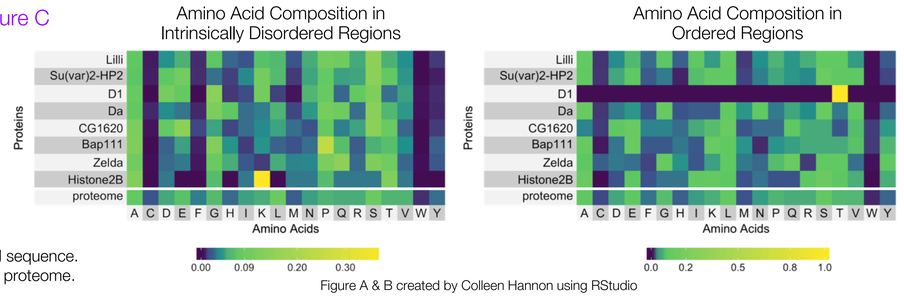


Figure A. Seven proteins imaged and analyzed for skewness in embryonic stages 13 & 14.
Figure B. Conservation of Bap111's longest disordered region taken from its complete amino acid sequence.
Figure C. Amino acid compositions of each protein in comparison to the Drosophila Melanogaster proteome.

Figure A & B created by Colleen Hannon using RStudio

Conclusions

The characterization of intrinsically disordered proteins is still being developed and the methods for proper study is widely unknown. We have determined through our methods that:

- For more skewed proteins (D1, Su(var)2-HP2, Lilli), there are more Lysine (K) in disordered regions in comparison to the proteome. For all disordered proteins studied, there are generally more disorder-promoting amino acids: Glutamine, Serine, and Proline (Q, S, & P).
- The disordered regions for our proteins of study have various strict areas of evolutionary conservation that we speculate are essential to their function.

References

[1] Nuffield Department of Clinical Neurosciences at University of Oxford. (2018). *FMRIB Neurodegeneration Group* [Online]. Available: https://www.ndcn.ox.ac.uk/research/neurodegeneration-imaging

[2] L. Kozlowski (2014) *SUMO-1 morphing based on NMR structure 1a5r* [Online]. Available FTP: https://en.wikipedia.org/wiki/Intrinsically_disordered_proteins#/media/File:1a5r_SUMO-1_protein.gif

[3] L. Staby *et al.* "Eukaryotic Transcription Factors: Paradigms of Protein Intrinsic Disorder" in Biochemical Journal, 2017 © Portland Press. doi: 10.1042/BCJ20160631

[4] J. Heady (2018) *FruitiFly* [Online]. Available: https://github.com/SJCCRAC/fruitifly

Future Work

- Continue to explore properties of and apply characteristics to disordered regions for the entire proteome of Drosophila Melanogaster.
- Remove disordered regions from proteins to see if they retain function.
- Continue to build out FruitiFly[4] functions and develop methods of study and comparison for disordered regions.