Opening points to make:

 Major goals of current project: 1) ChIPSeq on 60 rising to100 ChIPseq data sets per year for each worm and fly , 2) attempt to validate with RNAi/mutational analysis by RNA-seq. Successful 4 years, meeting or nearly meeting goals for each year for each species for strain generation and ChIP-seq, highly useful to the community for analysis of individual TF and larger-scale functions. Early returns on RNA-seq results intriguing. We now seek to amplify and expand on this progress, incorporating new technologies to complete the catalog of the remaining factors and further define the functionality of binding events in the genome.

1) Complete binding site discovery for remaining factors

 a. This goal requires two steps – generation of animal strains expressing GFP-tagged TFs, and performing ChIP-seq on these strains. At the end of the existing project, we estimate that, for worm, XX factors will still require strain generation and XX factors ChIP-seq. For fly, XX factors will still require strain generation and XX factors ChIP-seq. We will use the existing strain construction pipeline in each organism for the subset of TFs with available DNA constructs (~100 small in fly, ~100 large), and we will use CRISPR technology to engineer GFP into endogenous loci for the remaining 41 fly genes with no BAC. ChIP experiments will continue to be performed in whole animals to complement our previous experiments and facilitate comparisons.

b. In addition for factors expressed in only a few cells, which makes peak detection in whole animals problematic, we will explore methods for cell sorting or dissection,to isolate tissue- or cells for organ-specific ChIP-seq. \*\*\*How many of these are there?\*\*\* These methods would also be useful for determining whether TFs that are broadly expressed bind to distinct sites in different tissues/cell types.. \*\*\*This is more exploratory. How many should we try?\*\*\*

 c. Something new might be Reciprocal IPs to get chromatin- and RNA-associated complexes. (I had suggested GFP pull downs followed by MS to look at TF complexes but realized NHGRI would not want to fund such experiments – too far from the genome)

2) Test the biological relevance of the transcription factor binding sites for gene expression.

The assignment of TF binding sites to the appropriate regulated genes is not always reliable, and TF binding sites detected by ChIP-seq may not always influence expression of nearby genes.

a) In the current project, we have undertaken an analysis of how loss of TF activity affects candidate target gene expression. We propose to continue these efforts and expand to a greater number of factors to be able to better decipher the contribution of binding to gene regulation. Our pipelines are up and running and extremely efficient. We can perform 70 RNAi experiments (in replicate with three times and matched controls 840 samples) per year (with one full-time FTE). \*\*\*This is a big commitment to something we haven’t shown works.\*\*\*

 b) Many factors are co-localized more frequently than expected by chance, suggesting that they may interact in regulating gene expression. Further, loss-of-function mutations in about 80% of TF genes (in worm) are viable. To determine if these co-localizations are functionally relevant, we will generate double mutants of likely interacting genes (identified by their co-localization and overlapping expression patterns in time and space) and examine their phenotypic effects, including their impact on gene expression.

3) Construct a temporally and spatially resolved TF regulatory network.

We will continue to use our current data analysis pipeline to assess data quality and identify binding sites, and to place all experimental data in the ENCODE DCC to ensure rapid public accessibility. We will perform additional bioinformatics analysis to integrate the experimental data from Aims 1 (binding sites) and 2 (expression responses and gene interactions) with computational analysis of sequence motifs, spatial regulation, RNA expression, and other data sources, to build models of TF interactions that guide gene regulation in vivo.