PAR-17-482 Comparative Genomics Research Program (R01-Clinical Trial Not Allowed)

**Hypothesis and specific aims:**

**Synopsis (direct response to the funding opportunity purpose paragraph)** We have developed a comparative, functional genomic approach to investigate promoter structure and function across the eukaryotic domain. We will use this approach to examine the putative modular nature of effectors downstream of promoters thereby linking common *cis*-regulatory elements to conserved cell types and life cycle transitions. Our approach utilizes a novel technique to generate transcription start site (TSS) profiles at base-pair resolution, identifying promoter locations genome-wide and facilitating *de novo* motif discovery within core and proximal promoter regions. We will thus provide a new framework for understanding eukaryotic promoter evolution, as this approach can be extended to vertebrate and plant systems, as well as parasites.

**Introductory Statements:** Complex life cycles, which include amoeboid, flagellate, and cyst stages are present across major taxonomic groups across the eukaryotic tree of life. Precisely how complex life cycles are regulated is largely unknown due to initial focus on genome content over regulatory features. Gene regulation has been studied in plants, animals, and yeasts (REFs), but *cis*-regulatory features of diverse eukaryotes remain largely uninvestigated. This means that an understanding of the diversity and unity of regulatory features across eukaryotes is lacking. Co-PI Raborn has codeveloped a technique, STRIPE-seq that enables high throughput TSS mapping and promoter identification in any eukaryotic species. PI-Wideman has obtained a dataset of conserved orthologues across the major eukaryotic lineages. Combining our expertise will enable (1) the generation of promoter atlases of diverse, representative eukaryotes (2) a comparative analysis tracing the evolutionary history of promoters and their diversity across eukaryotes (3) the discovery of conserved gene regulatory networks. These findings will bring unprecedented insight into eukaryotic *cis*-regulatory elements, associated downstream effectors, and their evolution. The knowledge gained from this study will provide context for investigations into multicellular organisms, and how regulatory complexity evolved numerous times from simpler systems.

**Aim 1.** Generate promoter atlases for ~ 20 representative species across eukaryote diversity by identifying TSSs in a compendium of species using the novel TSS profiling method STRIPE-seq.

**Aim 2.** Use the results from Aim 1 to infer ancestral promoter states across eukaryotes. Incorporate ancestral and extant states to trace promoter evolution across eukaryotes.

**Aim 3.** Identify homologous developmental processes by inferring conserved gene regulatory networks across eukaryotic diversity.

Upon conclusion we will have promoter atlases for 20 eukaryotic lineages for which little or nothing is known about their regulatory sequences. We will have a basic understanding of how core and proximal promoters have evolved from the Last Eukaryotic Common Ancestor (LECA) to extant lineages. Finally, we will be able to trace the homology of gene regulatory networks across eukaryotic lineages with complex unicellular life cycles. These findings will have direct impacts on human health research by providing general insight into the evolution of eukaryotic *cis*-regulatory sequences. This insight can be applied to the biology of gene regulation in vertebrates, plants, and their parasites.

FIGURE 1: Synopsis of eukaryote diversity (new burki figure with asterisks of lineages remaining uninvestigated). Figure 2. Stripe-seq figure – very simple??