Data sets:

**Worm:**

~ 200 TFs (90 from modENCODE, 100+ from modERN and extra year) at various stages of development –

* one from stage of strong expression
* sometimes others.

~ 15 deletion RNA-seq embryonic time series

~ 5 RNA-seq embryonic time series of FAC sorted cells

~ 200 TFs with 3D movies of early embryo expression

**Fly:**

~ 200 TFs (~20from modENCODE, ~180 from modERN and extra year) at various stages of development –

~ 3 RNAi/deletion RNA-seq embryonic time series

~9000 RNA embryonic in situ patterns

**Analysis:**

**ChIPseq peaks**

Peak calls all with the same peak caller.

HOT regions – What is the definition? Overlapping peaks? By density? Mid-point of the peak within x bases of one another?

* Across all samples
* Stage specific
* Tissue specific?

Peaks assigned to genes distinguish 5’5’ vs 5’3’ vs 3’3’ orientation

* Promoter proximal
* Distal (how to handle 5’5’ orientation)
* Genic
* 3’

GO analysis etc of genes associated with non-HOT peaks.

Temporal relationship of expression between peaks and targets over the life cycle.

Spatial relationship between peaks and targets as above.

Clustering of non-HOT peaks -- what TFs bind near each other?

Motifs in not-HOT peaks, conservation

**Expression data**

What genes are under/overexpressed in deletion/RNAi strains?

What is the overlap with ChIP-seq peaks?

What genes are differentially expressed in TF-labeled cells?

What is overlap with ChIP-seq peaks?

How much of the differential expression can be explained by ChIP-seq peaks?

Can combinations of transcription factors do a better job?

Are these combinations actually expressed in the same cells (based on RNA-seq data and 3D movies/in situ data)?

Are there domain/chromatin contributions?

What differences/similarities are there among members of the same family (DNA binding domain) TFs?

What kind of a network does this produce? Can temporal/spatial data be used to prune it?

**Comparative analysis:**

What are the orthologous relationships? This is a question that never seems to have a fully satisfactory answer.

What features are conserved between orthologs of the two species?

* Proximity to promoter
* Targets, especially itself and other TFs
* Tissue
* Co-associations of TFs, modules