Exploring Distal Enhancers Through Histone SNPs

Chris Seward, Stubbs Lab
Cell and Developmental Biology
Institute for Genomic Biology
Molecular Roots of the Social Brain

- Project to identify gene networks that respond to social stimuli in mice, honeybees, and stickleback fish.
- Utilized RNASeq and ChIP-seq after social stimulus to identify differentially expressed genes and differentially accessible regulatory elements.
- In Honeybee, studies revealed stronger differences in histone modifications between colonies, than due to social stimuli.
- Additionally, it is always difficult to assign distal enhancers to the genes they regulate.
Differential Histone SNPs

- Can SNPs explain differences we see between colonies in histone peaks?
- Recent study in humans examined SNP preference in pooled histone and TF ChIP peaks.

- Can we apply a similar approach to data we have already collected from honey bee and mouse?
- Can we also use this data to help link distal peaks to genes?
When we do ChIP we collect input DNA, basically genomic DNA.

Pooling of all input DNA gives ~30x coverage → call SNPs in population.

Also pooled control H3k27ac peaks from each colony → call peaks.

Then checked SNP frequency in gDNA vs. inside Histone peaks.

---

**Differentially Enriched Histone SNPs Pipeline**

- **ChIP Inputs**
  - Pool CK samples by colony
  - Call SNPs inside peak regions
  - Compare SNP frequency of input+IP
  - Differential Histone SNPs

- **ChIP H3k27ac IPs**
  - Call Peaks
  - HOMER
  - /SAMtools
  - /FreeBayes

**Gene Expression**
Two Possible Histone SNP Effects

I. Genomic DNA has a particular SNP frequency in the pool, e.g., 50%, but Histone peak has a significantly different SNP frequency, e.g., 90%.

II. One of the two colonies has a SNP, but the other colony does not. Colony with the SNP has a peak, but the other colony does not.

Variation could be from heterozygous DNA OR from multiple individuals in pool.
Example: Unnamed Zinc Finger

Histone peak is only present in colony that has SNPs
One SNP allele is preferentially found in 83% of histone reads vs. 48% of genomic reads
Second Approach: BL6/CD1 Hybrid Histone SNPs

- Generating a Hybrid Mouse line allows for testing with less coverage

Expected 50/50 non-Influencing SNP

Histone Influencing SNP

Most SNPs are 50/50

Histone IP

HET gDNA

No preference

Strong preference for one genotype
Example: Med13l (100/0 SNP)
Future Directions

- Re-run improved Histone SNP calling pipeline on honeybee dataset, taking ploidy into account
- Bring in gene expression data to see if Histone SNPs also change gene expression pattern of nearby genes, allowing us to link distal peaks to the genes they regulate
- Identify co-incidence of Histone SNPs with regulatory motifs
- Use CRISPR to test hypotheses and validate methods in vivo
  - Eg. Introduce a SNP predicted to be influential to a primary cell line and see if that alters nearby chromatin and expression
Acknowledgements

MRSB Group

IGB GNDP Theme

- Lisa Stubbs
- David Zhao
- Gene Robinson
- Alison Bell
- Michael Saul
- Arian Avalos

Stubbs Lab

- Lisa Stubbs
- Huimin Zhang
- Joe Troy
- Soumya Negi
- Jae Eun Yoo
- Wei Chun
- Alex Pohlman

Keck Biotech Center

- Alvaro Hernandez