

Call for proposals for your final projects

- Write a simple proposal up to one page of text (≥ 10 point font) and up to one page of figure(s).
 - Think about designing a project that will bring together a few different concepts we have talked about in the course. See a few examples of good project presentations from last year on the next slides.
 - Remember you only have three days to work on the project, so don't try to do too much.
 - Great projects use a subset of the data and get something simple working as a proof of principle.
 - We recommend you talk to other students to see if they have similar ideas for a project. Several of you are very welcome to write a proposal together or you can write a proposal on your own.
 - Include author name(s) on the proposal
 - In your proposal include a brief background description of the problem you want to work on without any unnecessary details, input data you will use and the design of any pipeline, database or website you will be building.
- Email a PDF of your proposal to Simon (prochnik at gmail dot com) or Sofia (sofia2341 at gmail dot com) by Wednesday 9am.

What happens next??

- The Instructors and TAs will read all the proposals and discuss which will make for successful projects.
- We will pick five projects and assign each project a dedicated TA who will help with the project.
- The projects will be posted in Bush and students can sign up to any project.
- We hope to get roughly equal numbers of students per project and may need to shuffle projects and/or students a little to achieve this.

Example project ideas follow...

Goal: Pipeline for quantifying allele-specific expression

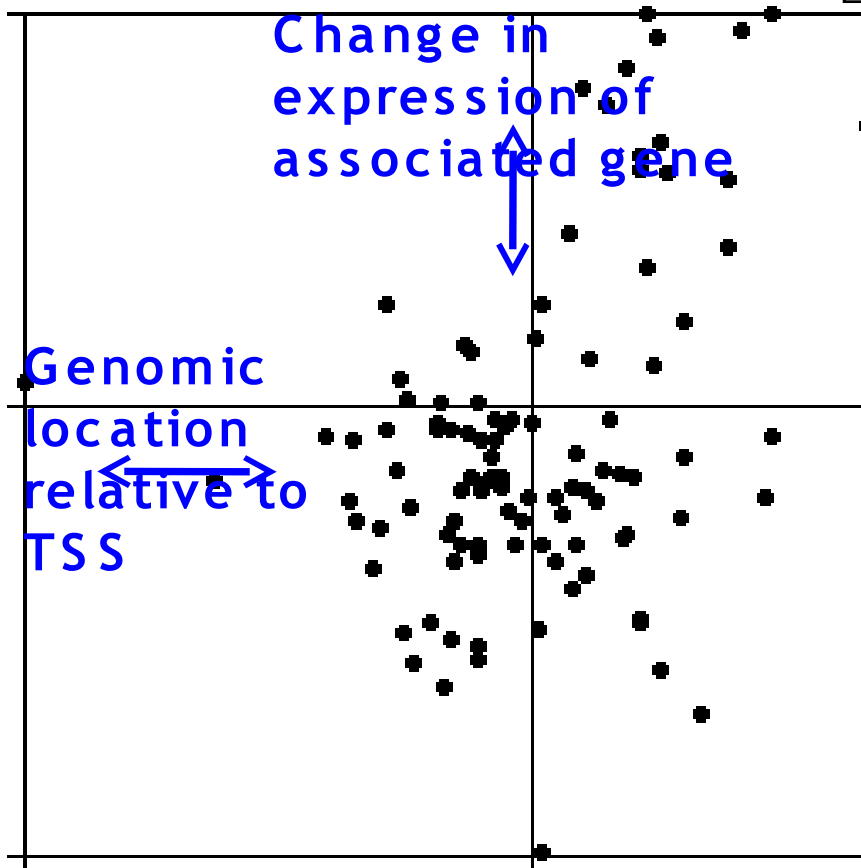
Methodology

- Curate Reads
 - Remove/trim low quality sequence
 - Remove poly A sequences
- Align to Reference and call SNPs
 - choose alignment software (BWA, Bowtie)
 - choose reference (genome, annotations)
 - call SNPs and confirm with known SNP set if available
 - format output for analysis package or individual scripts
- Analyze
 - quantify expression and test for allele-specific expression
 - output for GBrowse or other display in genomic context
 - develop display to extract condition specificity

ChIP-seq/expression Relational Databases

- Query relational databases using MySQL to identify groups of peaks/genes of interest.
- Send data to R to graphically represent phenomena/calculate statistical significance

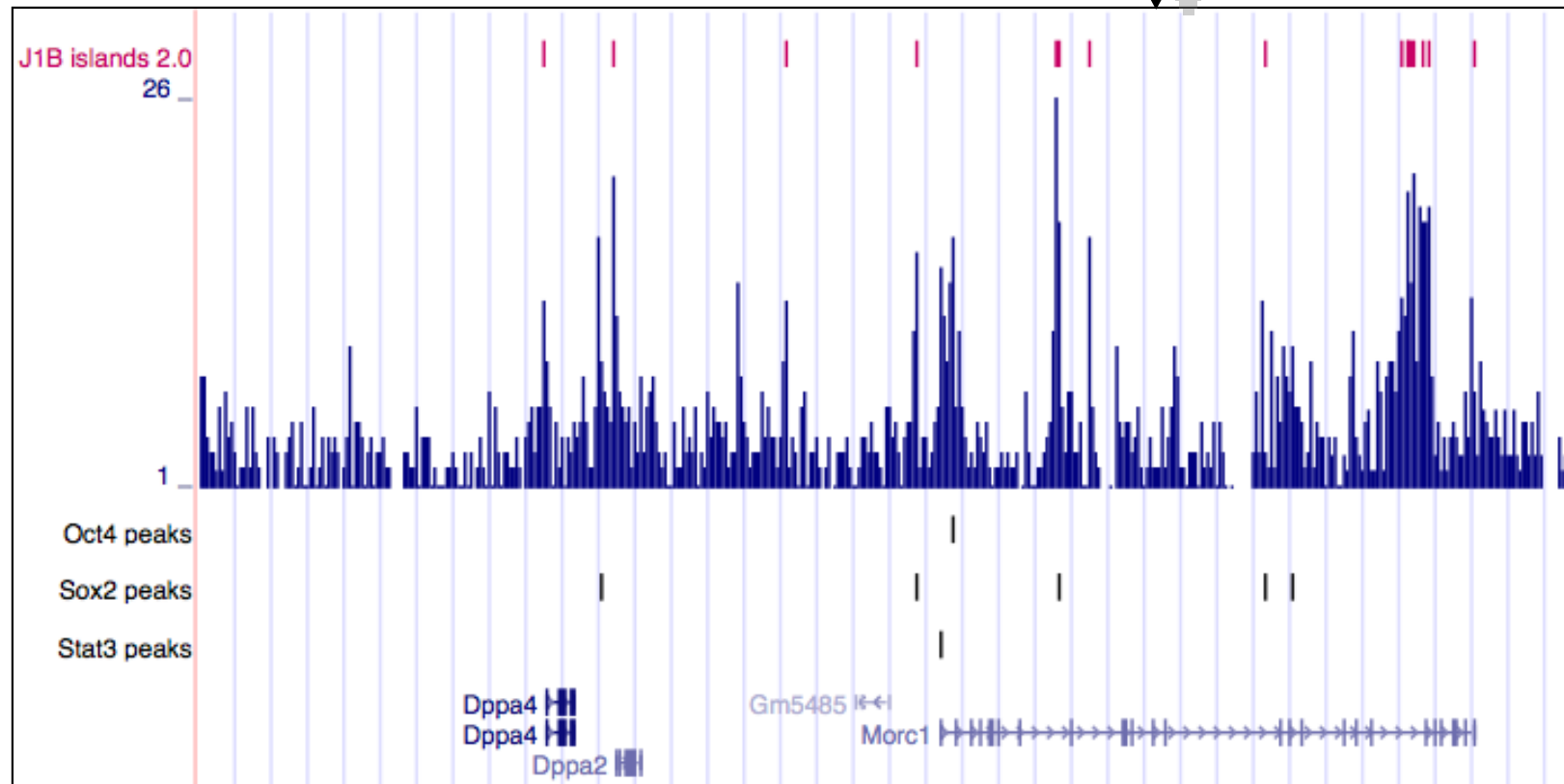
Ex: Correlate gene regulation with genomic location



- Query relational databases using MySQL to identify groups of peaks/genes of interest.
- Generate primer sets for ChIP-seq peaks for experimental validation by sending sequences to primer3 website

Forward: actgcatgctatcgatcgatgc | Stats...
Reverse: tacgatgcatgcgaggctagca | Stats...

- Set up a genome browser with ChIP-seq and/or expression data
- Allow access to browser for individual entries through SQL queries.



Pipeline to discover and estimate abundance of novel noncoding RNAs .

