PhenoGen Informatics

The site for quantitative genetics of the transcriptome

http://phenogen.ucdenver.edu

The PhenoGen Group
Initiated and directed by Dr. Boris Tabakoff
Presented by Dr. Laura Saba
Anschutz Medical Campus
University of Colorado Denver

**Please use Firefox for optimal results and to follow along in the workshop

Acknowledgements

The PhenoGen Group:

Boris Tabakoff, PhD; Paula L. Hoffman, PhD; Spencer Mahaffey, MA; Laura Saba, PhD; Stephen Flink, PhD; Lauren Vanderlinden, MS; Yinni Yu, MS

Collaborators:

Morton Printz, PhD; Michal Pravenec, PhD

Technical Support:

Adam Chapman; James Huntley; Laura Breen; Donna Moye

Grant Support:

National Institute on Alcohol Abuse and Alcoholism; the Banbury Fund (Robertson Family)

Outline

1. Introduction to PhenoGen Informatics
2. Interactive demo of candidate gene approach
3. Summary of phenotype approach
4. Downloadable data
5. Future directions

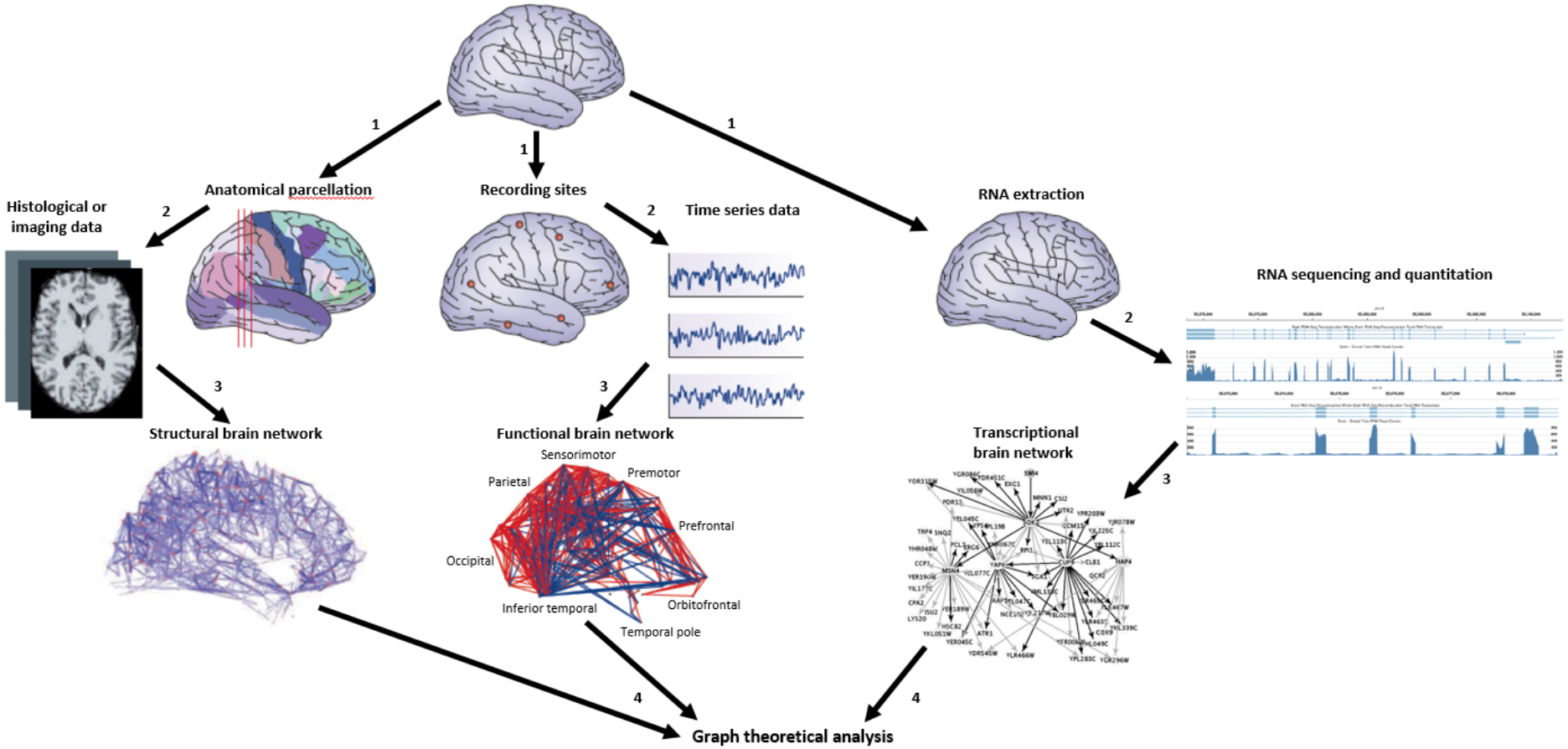
Goal for the PhenoGen Project

Transcriptional Connectome

To generate a new image of organs as networks of interacting elements (transcripts)

- Collect genome sequence and full transcriptome information for organs (brain, liver, heart).
  - We have completed exon array analysis for these three organs
  - RNA-Seq for brain in 30 strains of the HXB/BXH panel completed
  - RNA-Seq for heart and liver of the RI progenitor strains completed

Brain Transcriptional Connectome



Why Networks?

1. The brain is a complex hierarchical network spatio-temporally linked through structure and function.
2. Complex pathologic traits can be conceptualized as systems disorders of failed network regulation.
3. The brains structural and functional systems have features of complex networks that can be described through application of “graph theory” (small world topology).
4. The generation of a “Transcriptional Connectome” representing the Resting State transcriptional networks provides power for understanding predisposition to disease, etiology of organ or behavioral pathology and response to medications or toxins.

Why Study the RNA Dimension

Transcriptome links DNA and complex traits/diseases

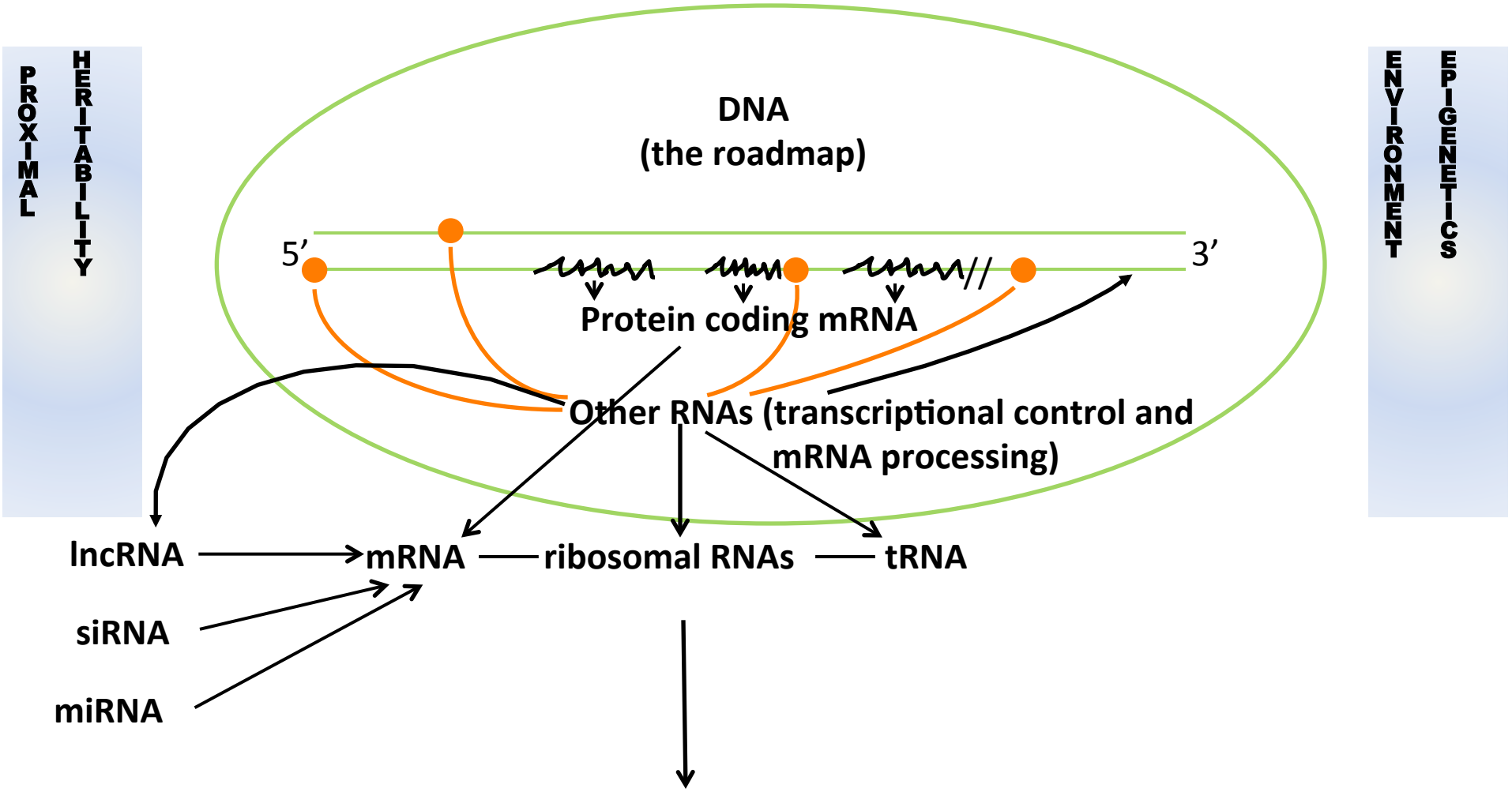
A. RNA is one of the first quantitative links between DNA sequence and phenotype (an endophenotype).

B. Transcriptome information addresses part of the GWAS Gap: how does an identified DNA polymorphic locus contribute to disease?

C. First step where DNA sequence and environment interact.

D. Implementation of graph theory at the transcript level provides insight into genetic/environmental interactions that are the basis for susceptibility to complex diseases.

The RNA Dimension (the true intermediate phenotype)



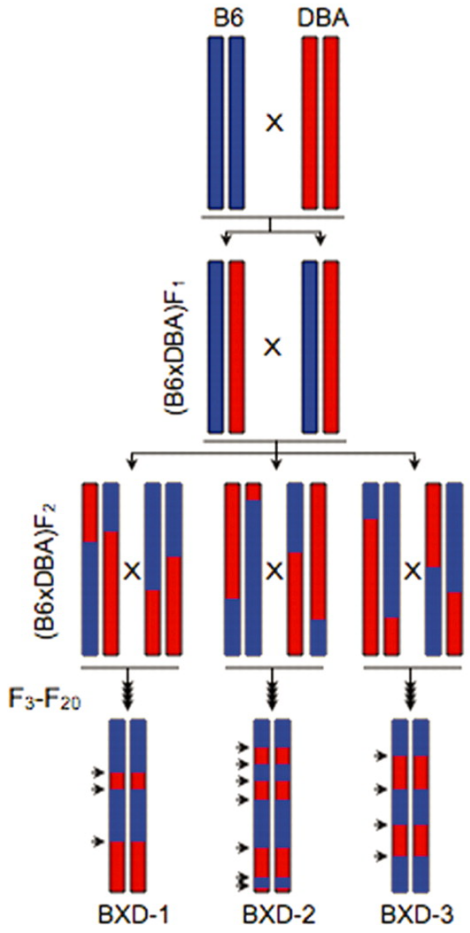
Co-expression as a measure of the "connectome"

- **Theory** – if the magnitude of RNA expression of two transcripts correlates over multiple "environments" (genomes), then the two transcripts are involved in similar biological processes

- Caveats when multiple environments are multiple genetic backgrounds (false positive correlations)

  - Linkage Disequilbrium
  - Cell-type mixing proportions

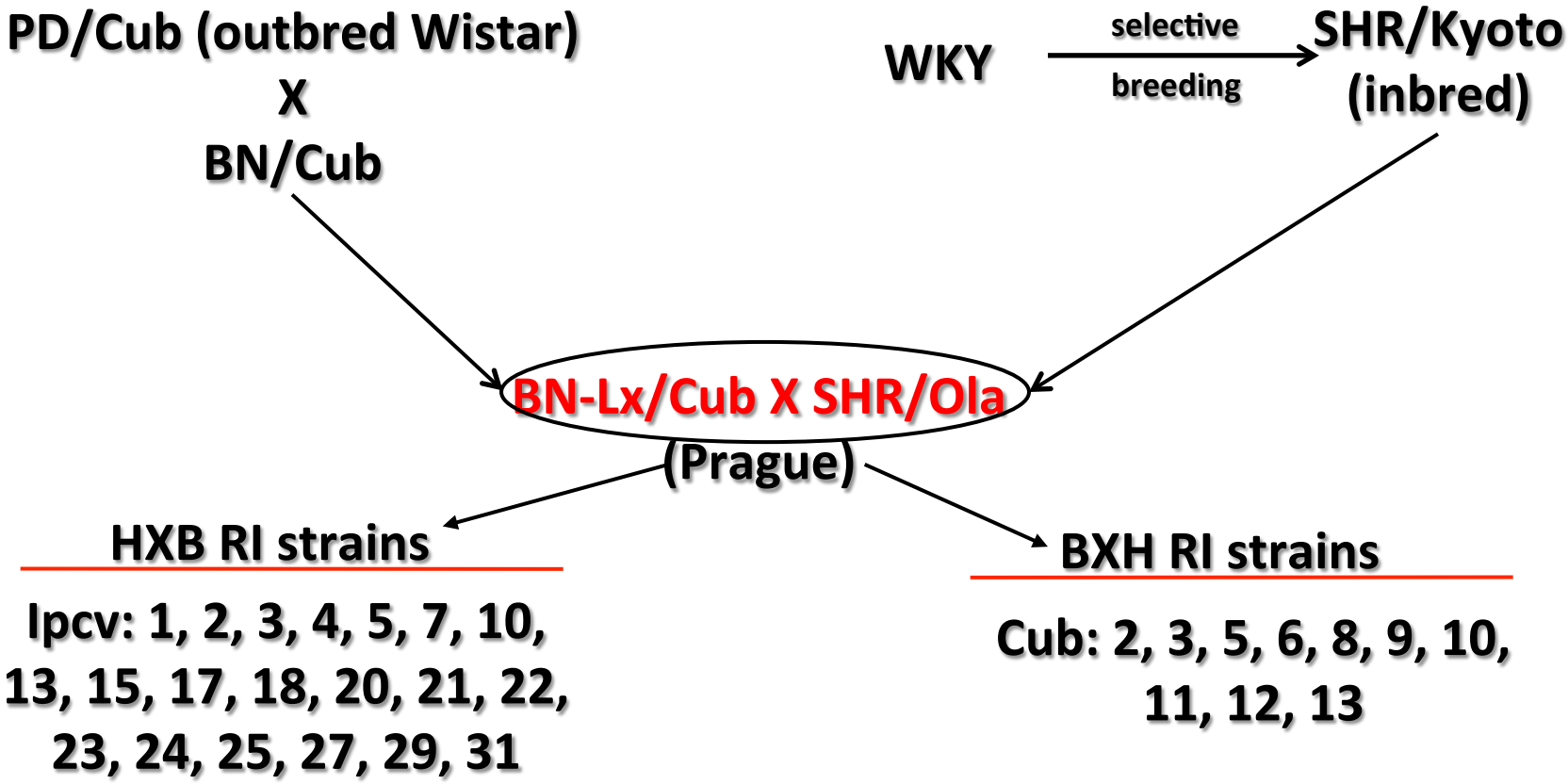
Recombinant Inbred Rodent Panel



- Genetic identity is retained over generations
- Cumulative genetic and phenotype data across labs
- Ideal genetic controls for studying interventions/ environmental effects

Wu Z K et al. Nucl. Acids Res. 2010;nar.gkp1251

Origin of HXB/BXH RI Panel



- > 4,500,000 SNPs/indels between progenitors
- Haplotype map generated

Current Data – RNA-Seq

Strains	Tissue	Sex	Number of Biological Replicates Per Strain	Number of Paired-End Reads (rRNA-depleted total RNA)	Number of Paired-End Reads (polyA+-selected RNA)	Number of Single- End Reads (small RNA)
SHR and BNLx	brain	male	3	645 million	192 million	96 million
SHR and BNLx	brain	female	4	982 million	---	297 million
SHR and BNLx	liver	male	3	583 million	---	342 million
SHR and BNLx	heart	male	4	790 million	---	300 million
30 RI Strains	brain	male	1 to 2	7.3 billion	----	1.9 billion

Current Data - Microarrays

Panel	Platform	Tissue	Number of Biological Replicates
21 HXB/BXH RI Strains	Exon Array	Brain	2 to 4
21 HXB/BXH RI Strains	Exon Array	Liver	2 to 4
21 HXB/BXH RI Strains	Exon Array	Heart	2 to 4
21 HXB/BXH RI Strains	Exon Array	Brown Adipose	2 to 4
60 LXS Mouse RI Strains	Exon Array	Brain	6
30 BXD Mouse RI Strains	3' Array	Brain	4 to 7
26 Inbred Mouse Strains	3' Array	Brain	4 to 6

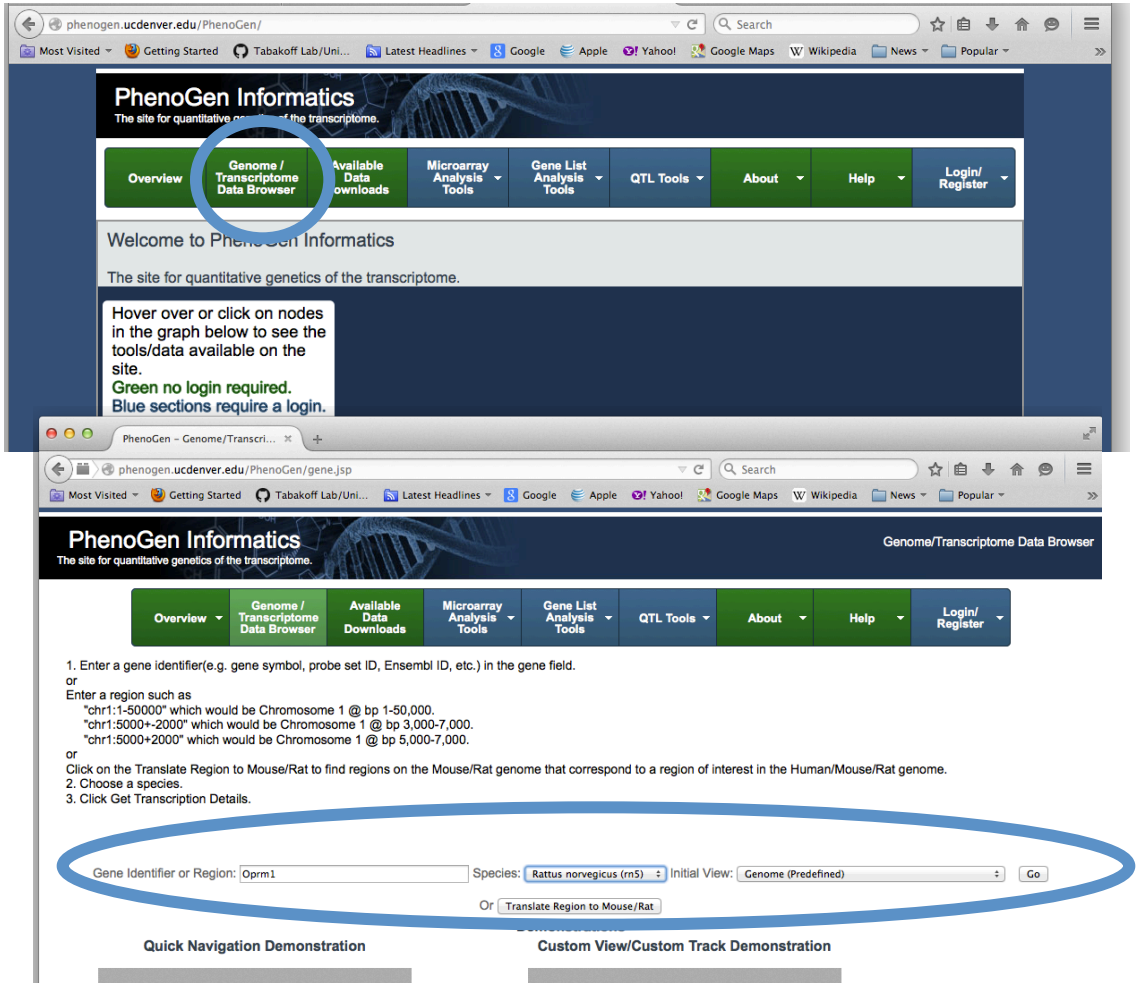
CANDIDATE GENE EXAMPLE

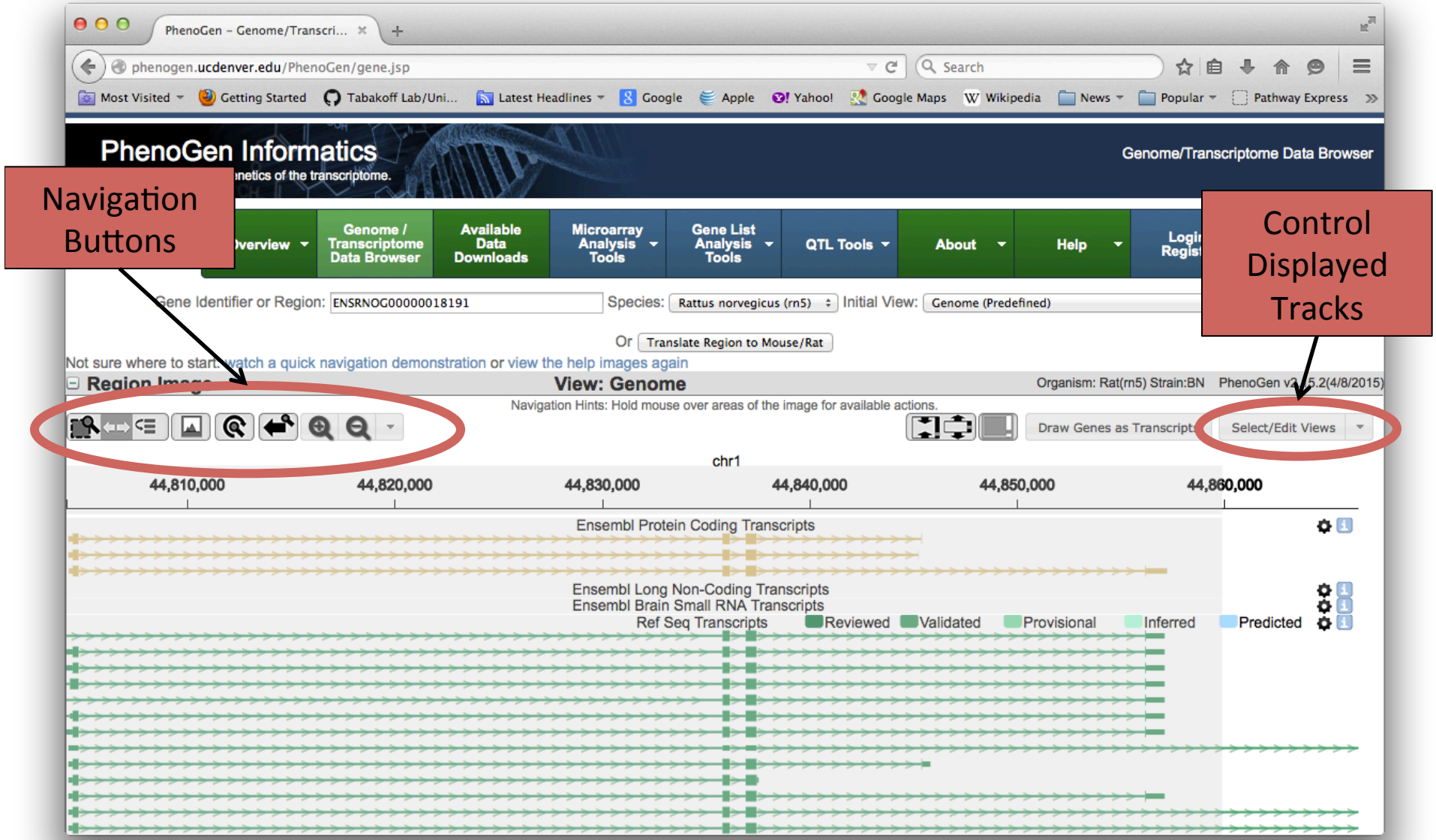
Transcriptome Information on a Gene of Interest

- Genome View
  - Annotated isoforms
  - DNA variants in parental strains
  - Published QTL
- Transcriptome View
  - Exon array information
  - RNA-Seq results
- Selected Feature
  - Gene details and external links
  - Gene eQTL
  - Weighted gene co-expression network analysis
  - Probe set level data

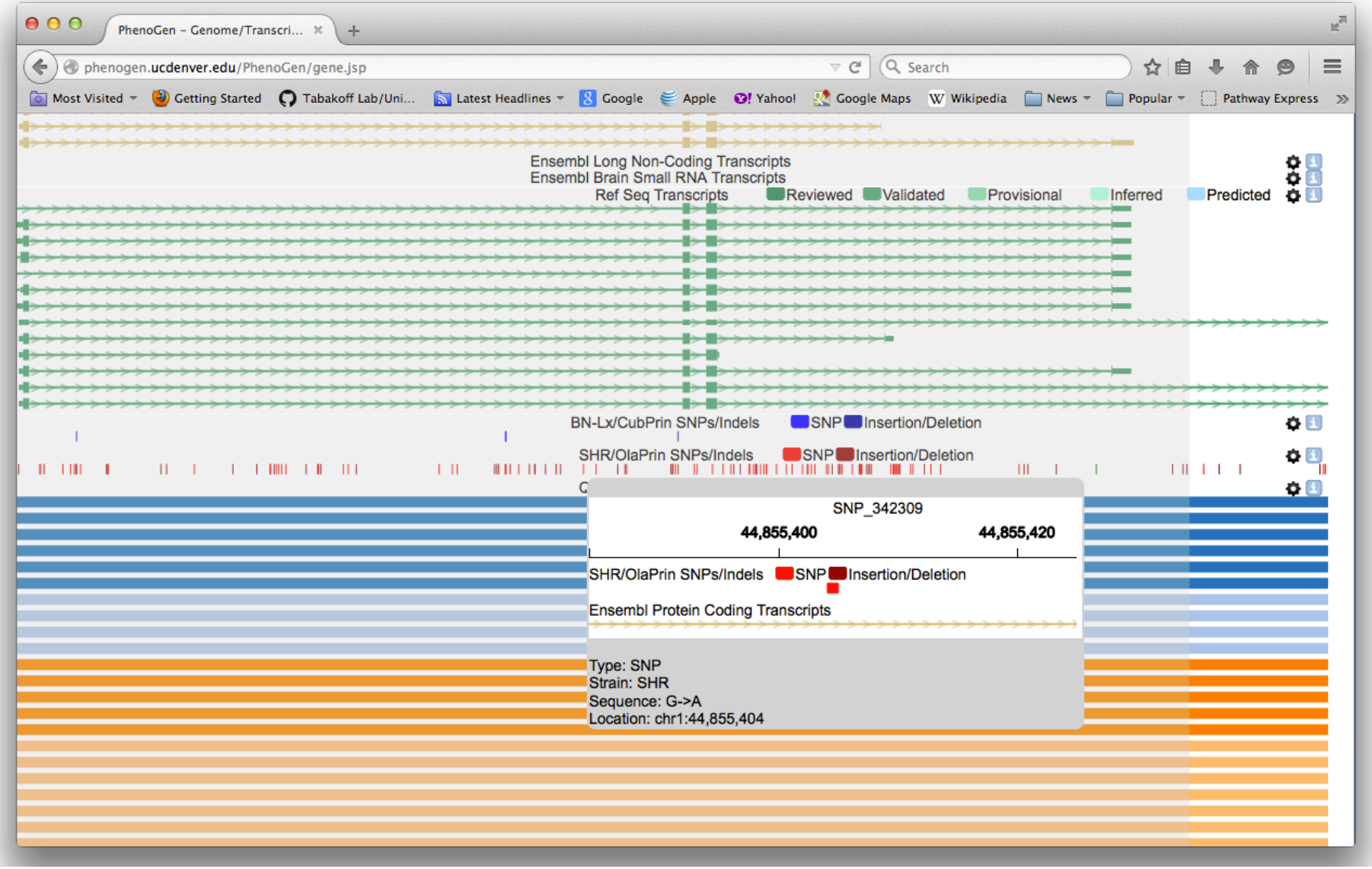
Genome View

1. Click Genome/ Transcriptome Data Browser
2. Type in Rat/Mouse Official Gene Symbol into “Gene Identifier or Region” box
3. Select **Rattus Norvegicus (rn5)** from “Species” box
4. Select **Genome (Predefined)** from “Initial View” box
5. Click “Go”

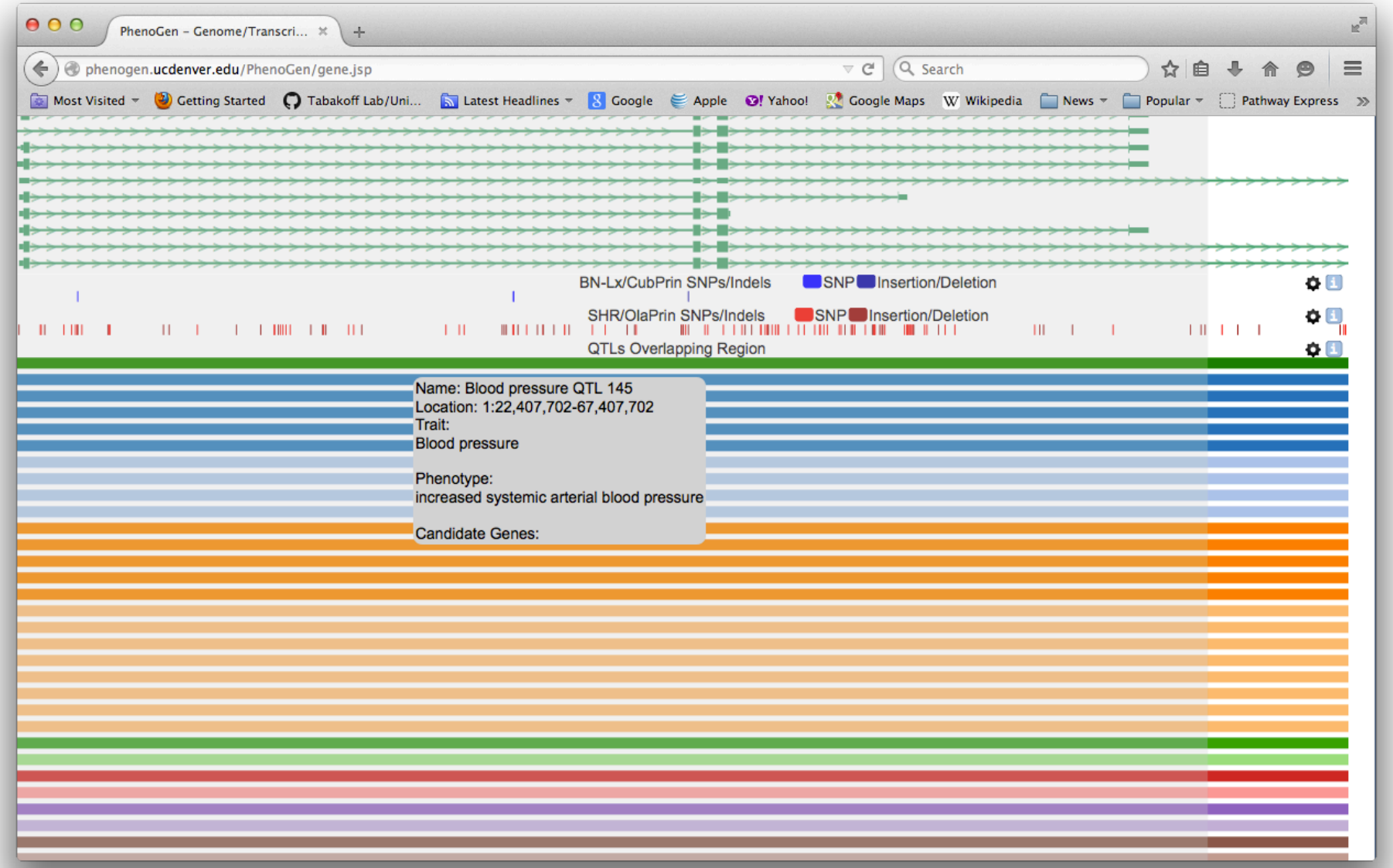



Navigation Buttons
Control Displayed Tracks
PhenoGen - Genome/Transcri...
phenogen.ucdenver.edu/PhenoGen/gene.jsp
Search
Most Visited
Getting Started
Tabakoff Lab/Uni...
Latest Headlines
Google
Apple
Yahoo!
Google Maps
Wikipedia
News
Popular
Pathway Express
PhenoGen Informatics
Genome/Transcriptome Data Browser
Genome / Transcriptome Data Browser
Available Data Downloads
Microarray Analysis Tools
Gene List Analysis Tools
QTL Tools
About
Help
Gene Identifier or Region:
ENSRNOG00000018191
Species:
Rattus norvegicus (rn5)
Initial View:
Genome (Predefined)
Or
Translate Region to Mouse/Rat
Not sure where to start? watch a quick navigation demonstration or view the help images again
View: Genome
Organism: Rat(rn5) Strain:BN
Navigation Hints: Hold mouse over areas of the image for available actions.
Draw Genes as Transcripts
Select/Edit Views
chr1
44,810,000
44,820,000
44,830,000
44,840,000
44,850,000
44,860,000
Ensembl Protein Coding Transcripts
Ensembl Long Non-Coding Transcripts
Ensembl Brain Small RNA Transcripts
Ref Seq Transcripts
Reviewed
Validated
Provisional
Inferred
Predicted

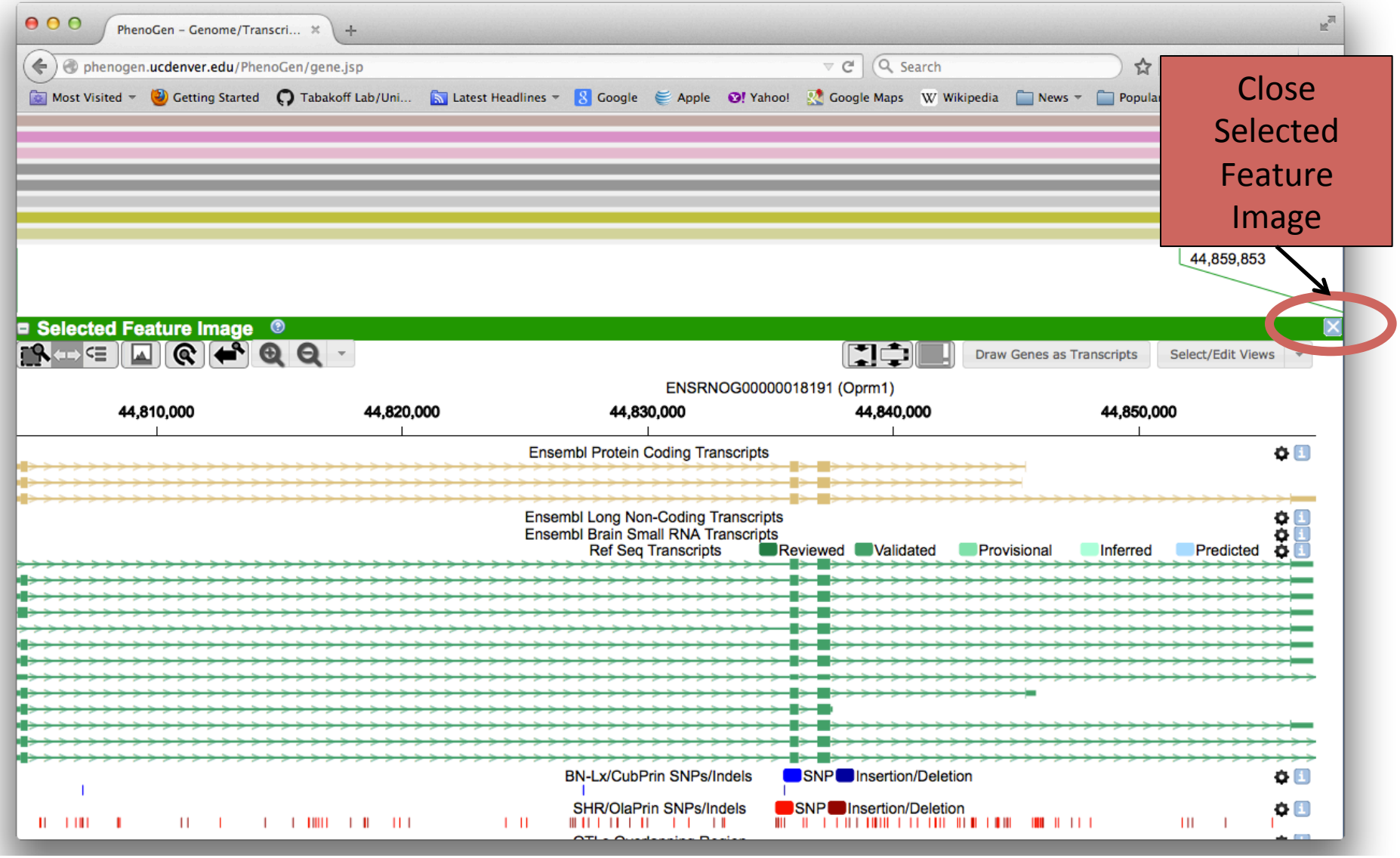
Genetic Variants Between Parental Strains of RI Panel



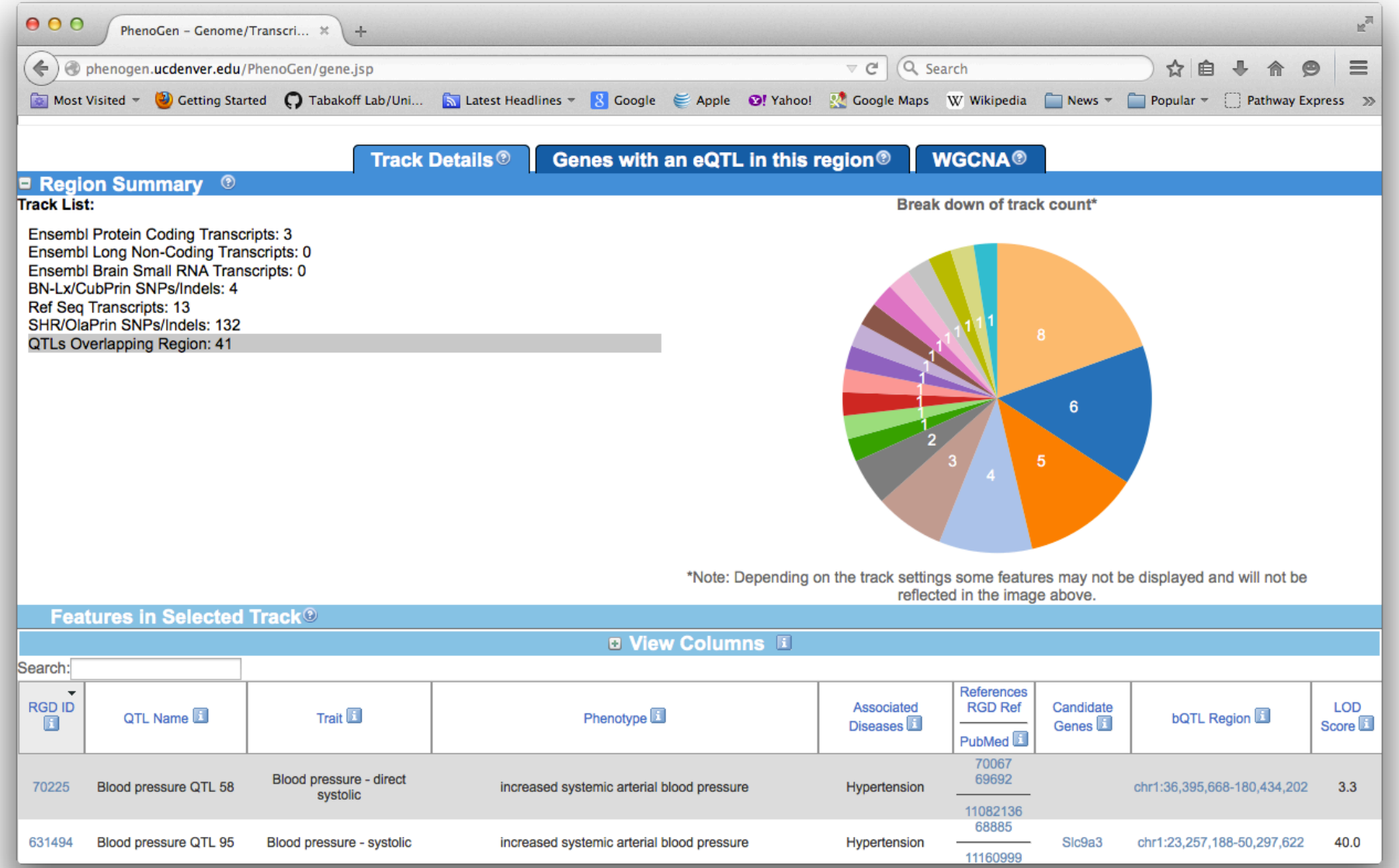
Published Behavioral/Physiologic QTL



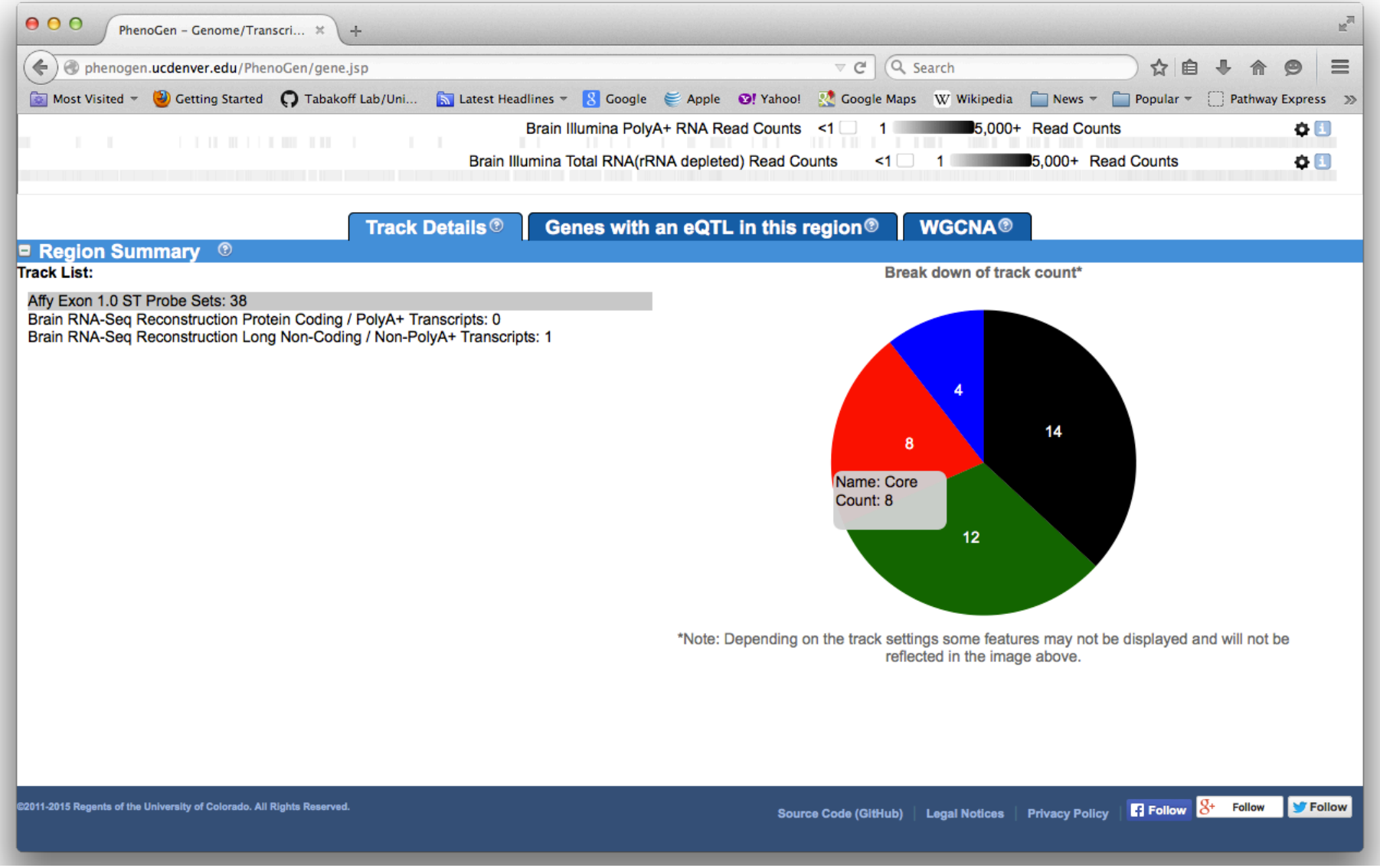
Get Region Summary



Region Summary

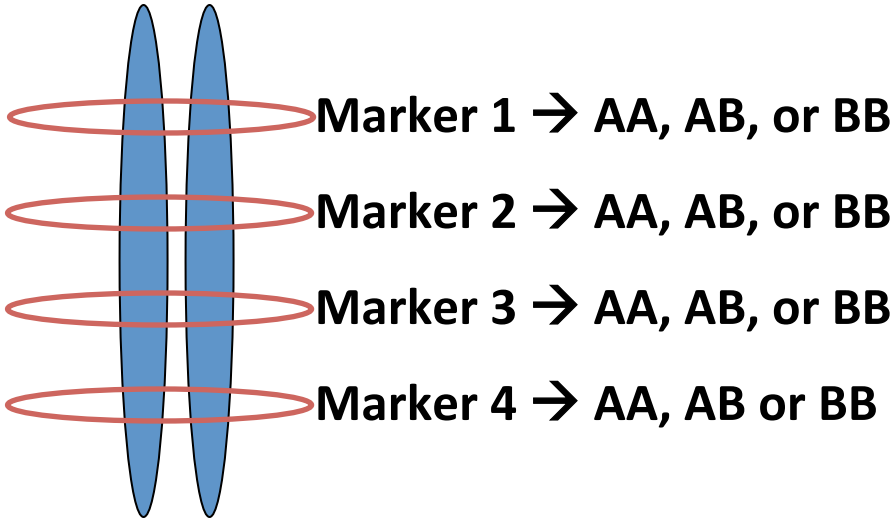


Alternate Region Summary Under Different View



Quantitative Trait Loci

- Definition – area of the genome where polymorphisms are associated with a quantitative trait

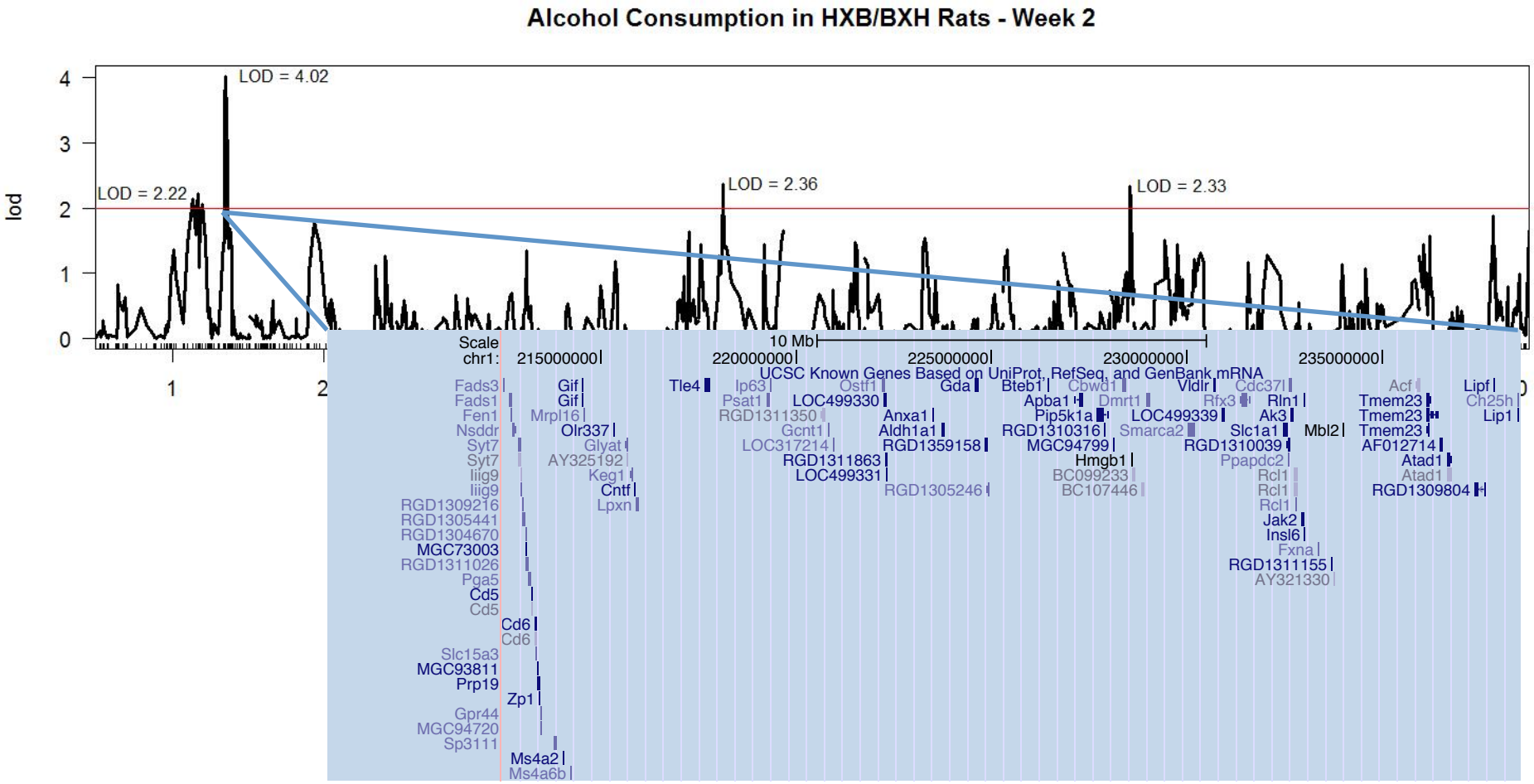


Chromosome G

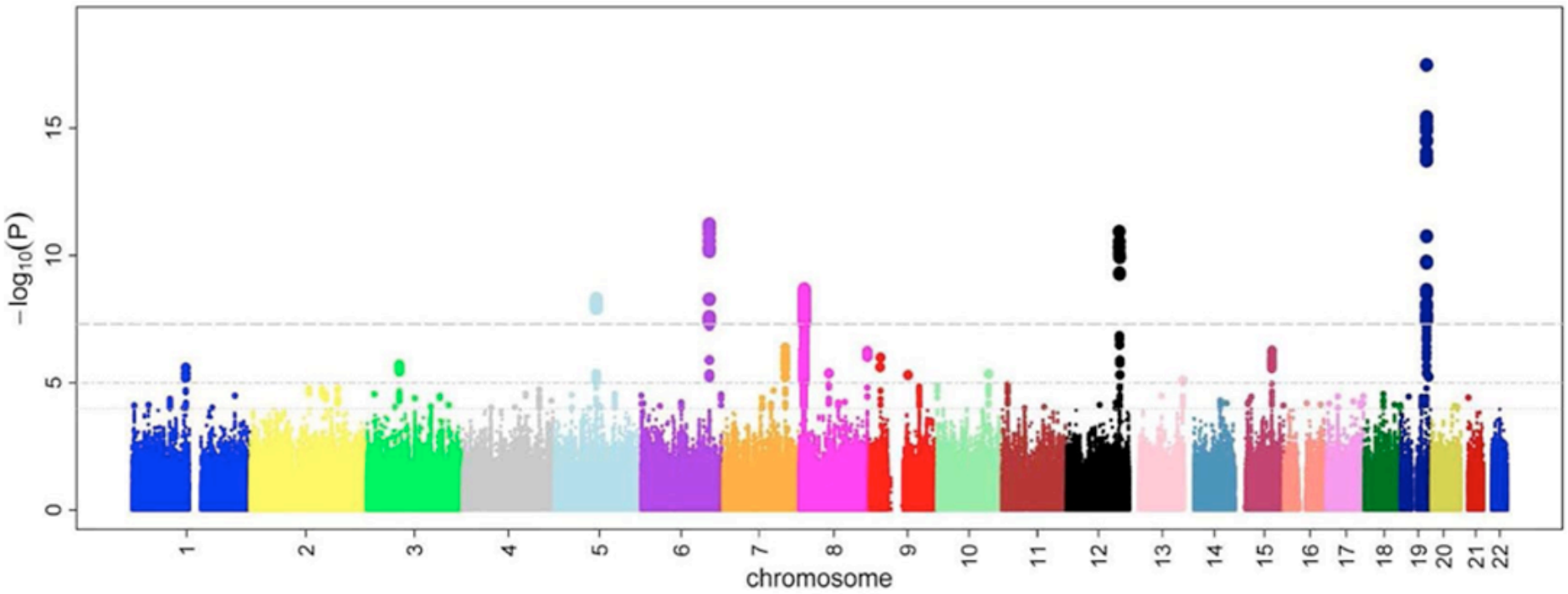
At each marker:

1. **Split population into two/three groups based on genotype**
2. **Compare mean values for the quantitative trait (phenotype) between the groups**
3. **LARGE Statistically Significant Differences → QTL**

Example LOD Plot

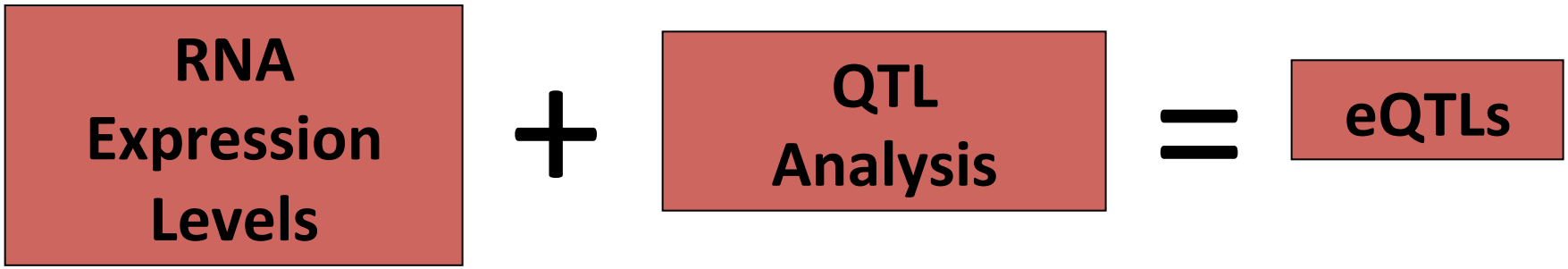


Example Manhattan Plot

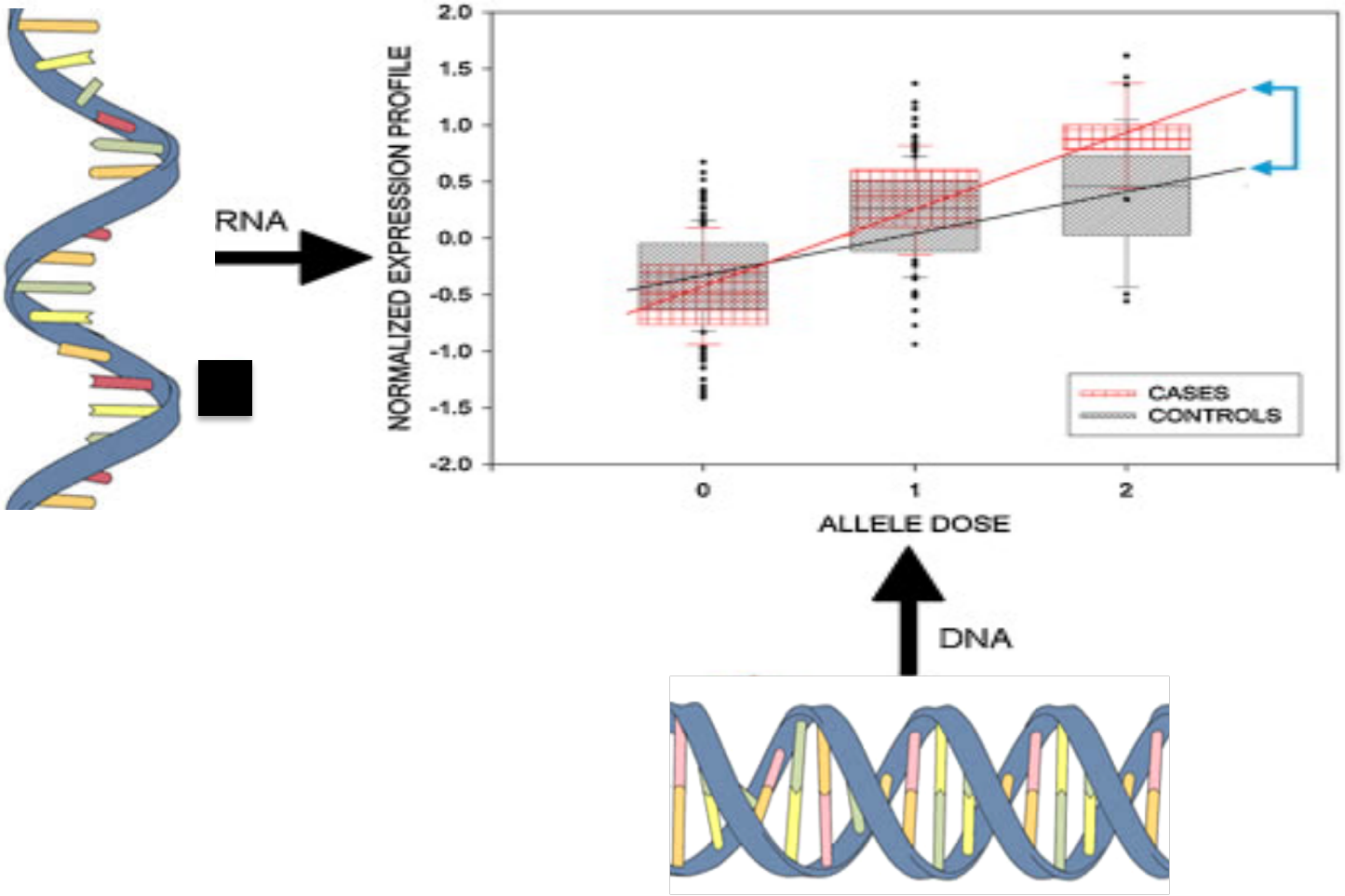


Definition of Genetical Genomics

The study of the genomic location where transcription is controlled by determining expression Quantitative Trait Loci (eQTL).



eQTL Definition



Myers, AJ. The age of the "ome": Genome, transcriptome and proteome data set collection and analysis. Brain Research Bulletin Volume 88, Issue 4 2012 294 - 301

Types of eQTL

- *cis*-eQTL (or local eQTL) – the locus controlling transcription is "near" the physical location of the gene in the genome
- *trans*-eQTL (or distal eQTL) – the locus controlling the transcription is NOT "near" the physical location of the gene in the genome

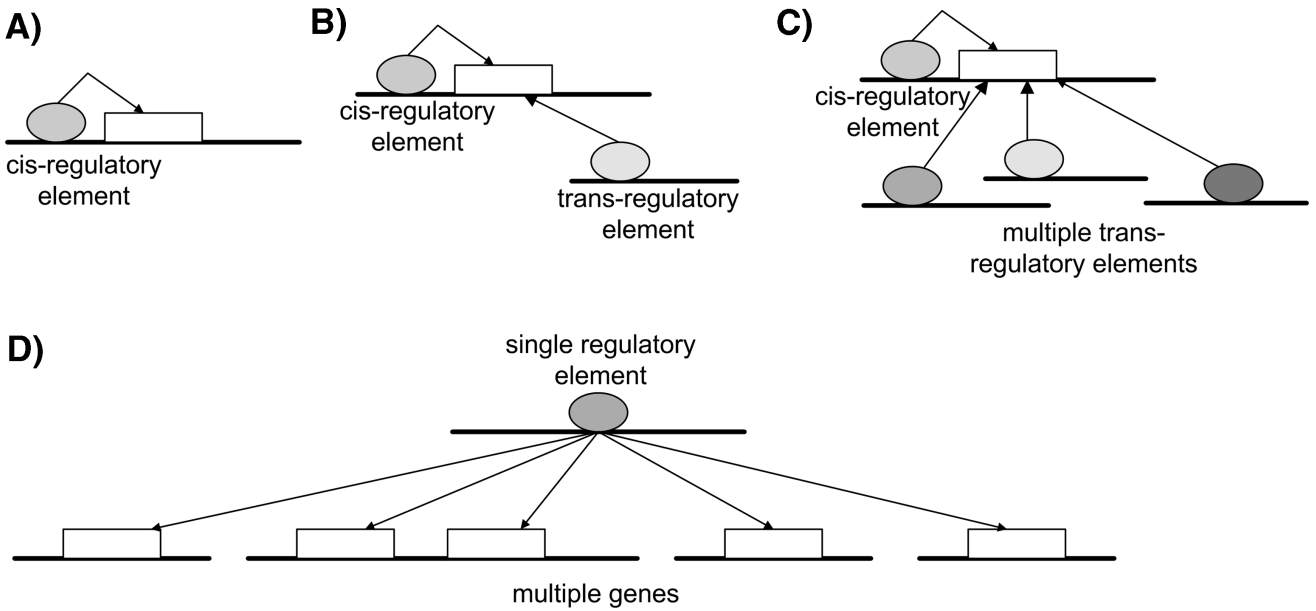
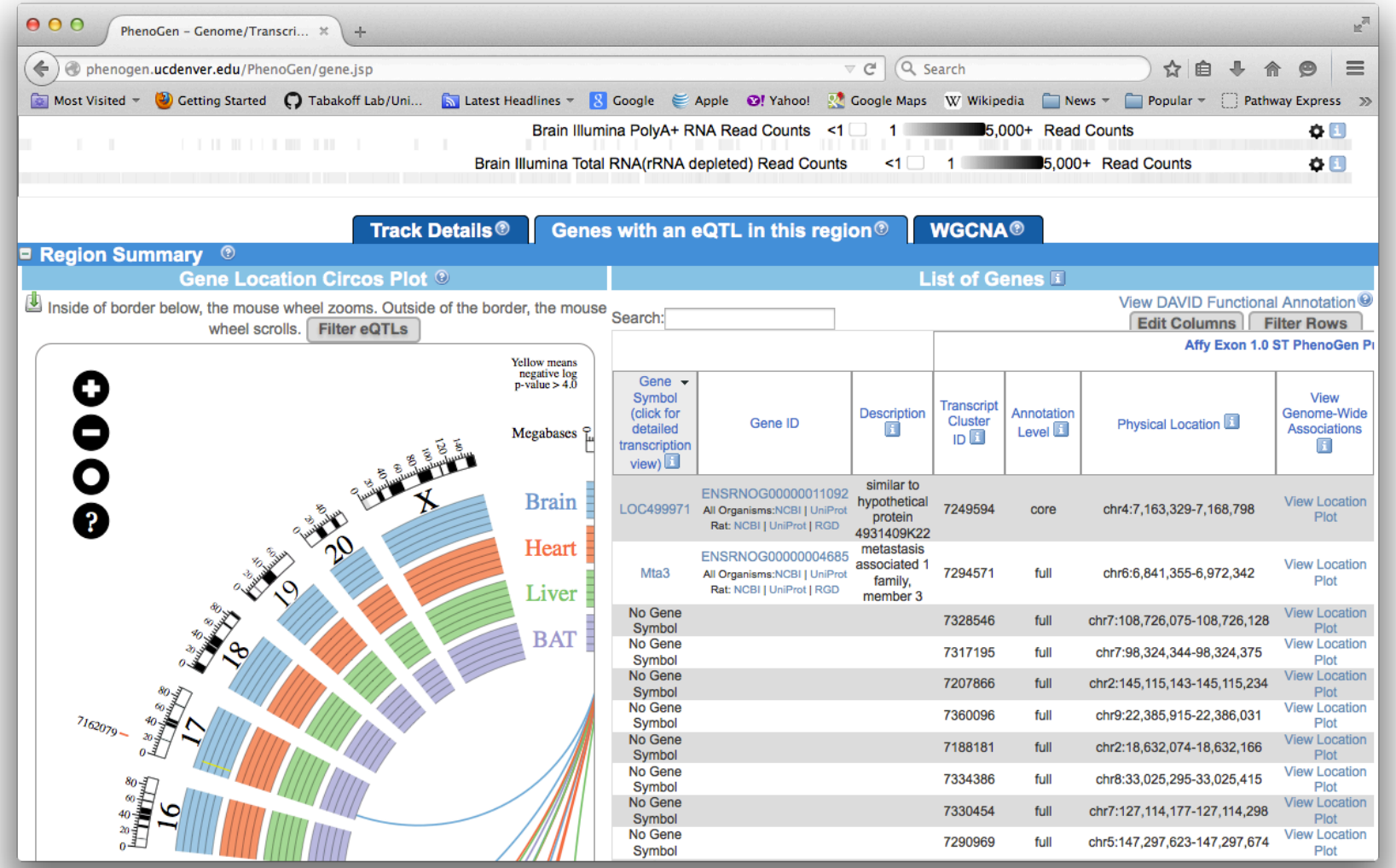


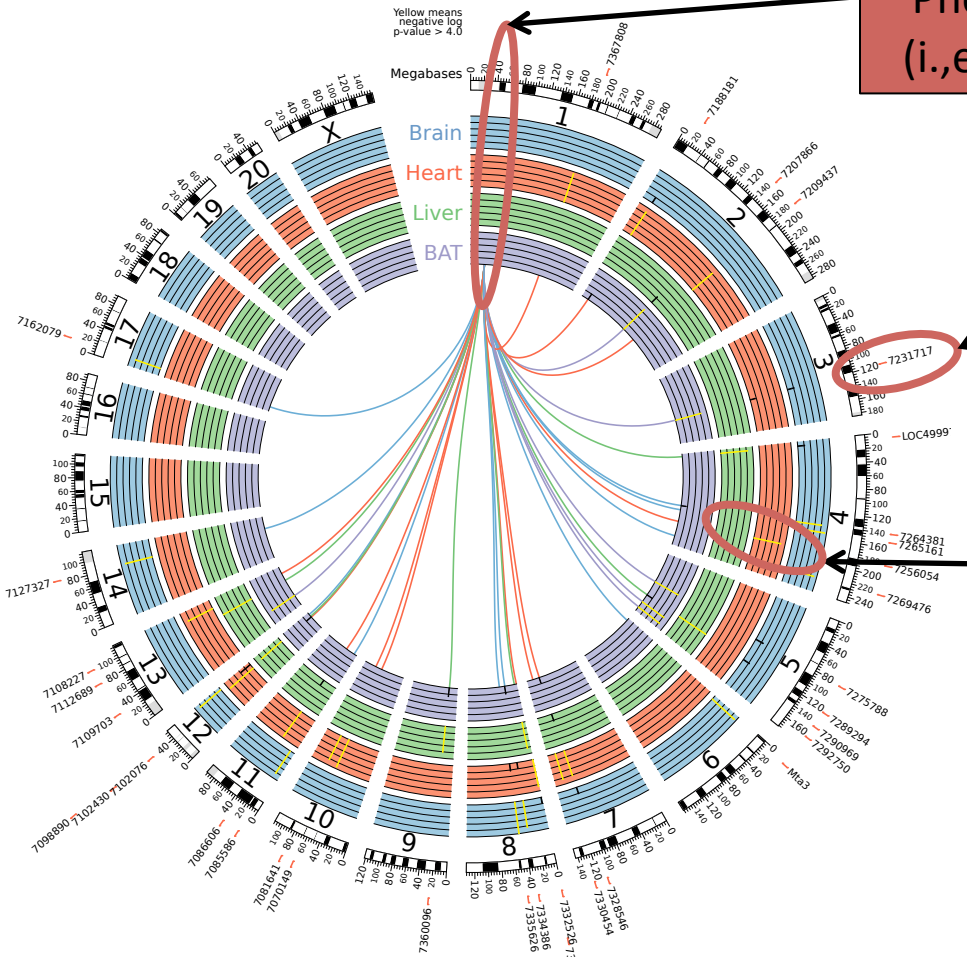
Image copied from Sieberts SK and Schadt EE (2007). Moving toward a system genetics view of disease. Mammalian Genome 18(6): 389-401.

Genes with an eQTL in this region



Circos Plot

Genes within an eQTL in this region

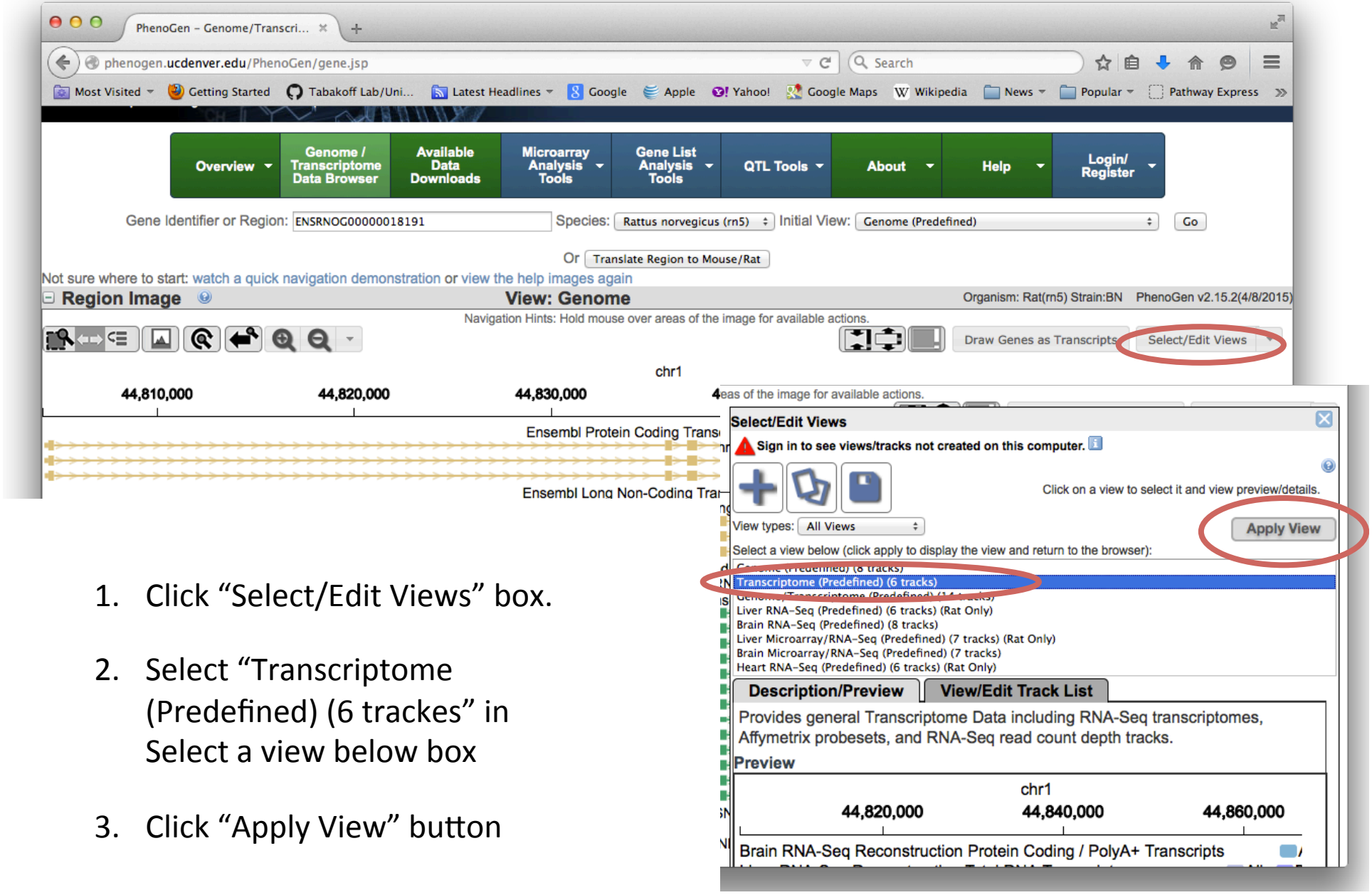


Region from PhenoGen (i.,e., eQTL)

Physical Location of Transcript Cluster

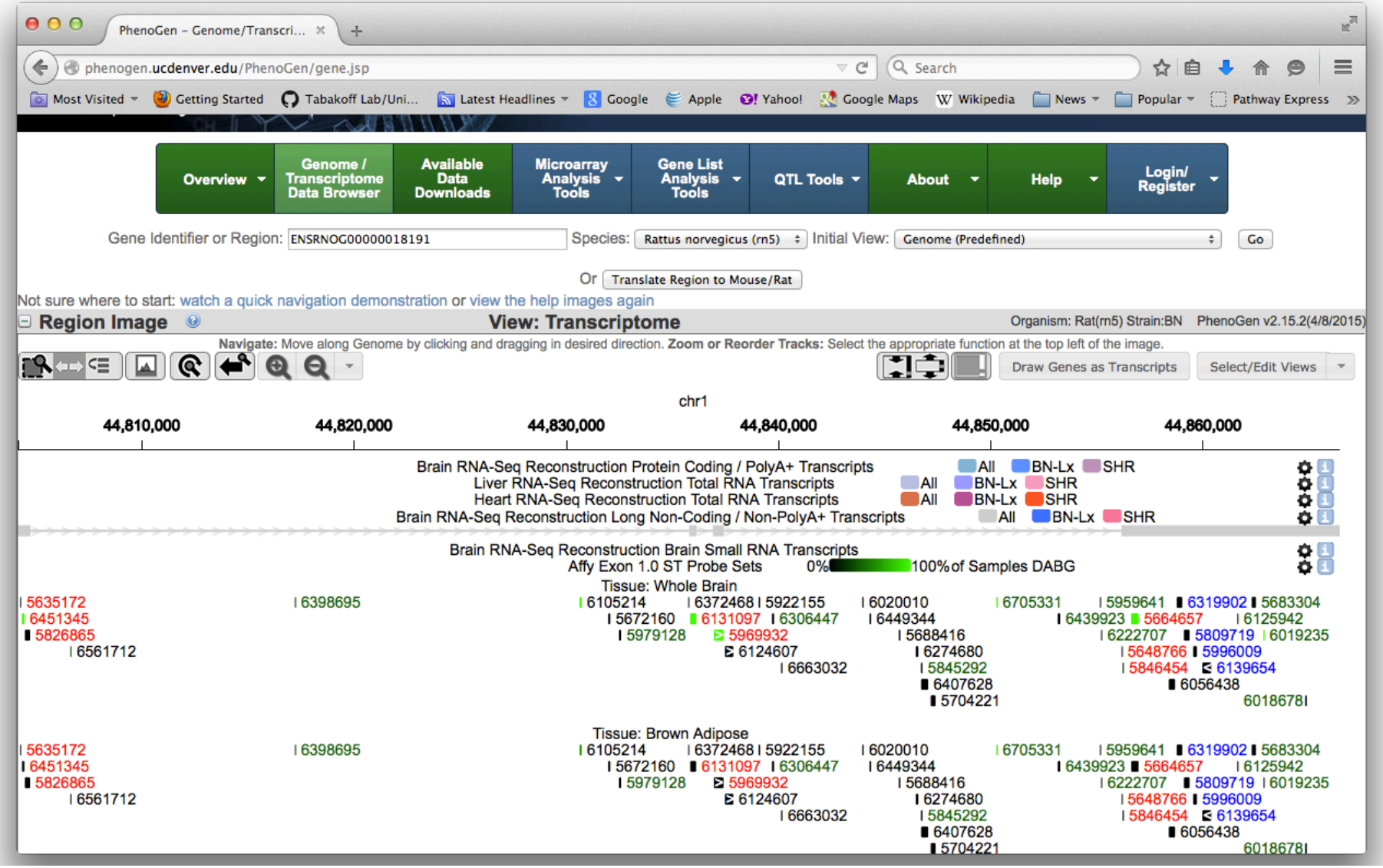
P-value of eQTL in Heart

Transcriptome View

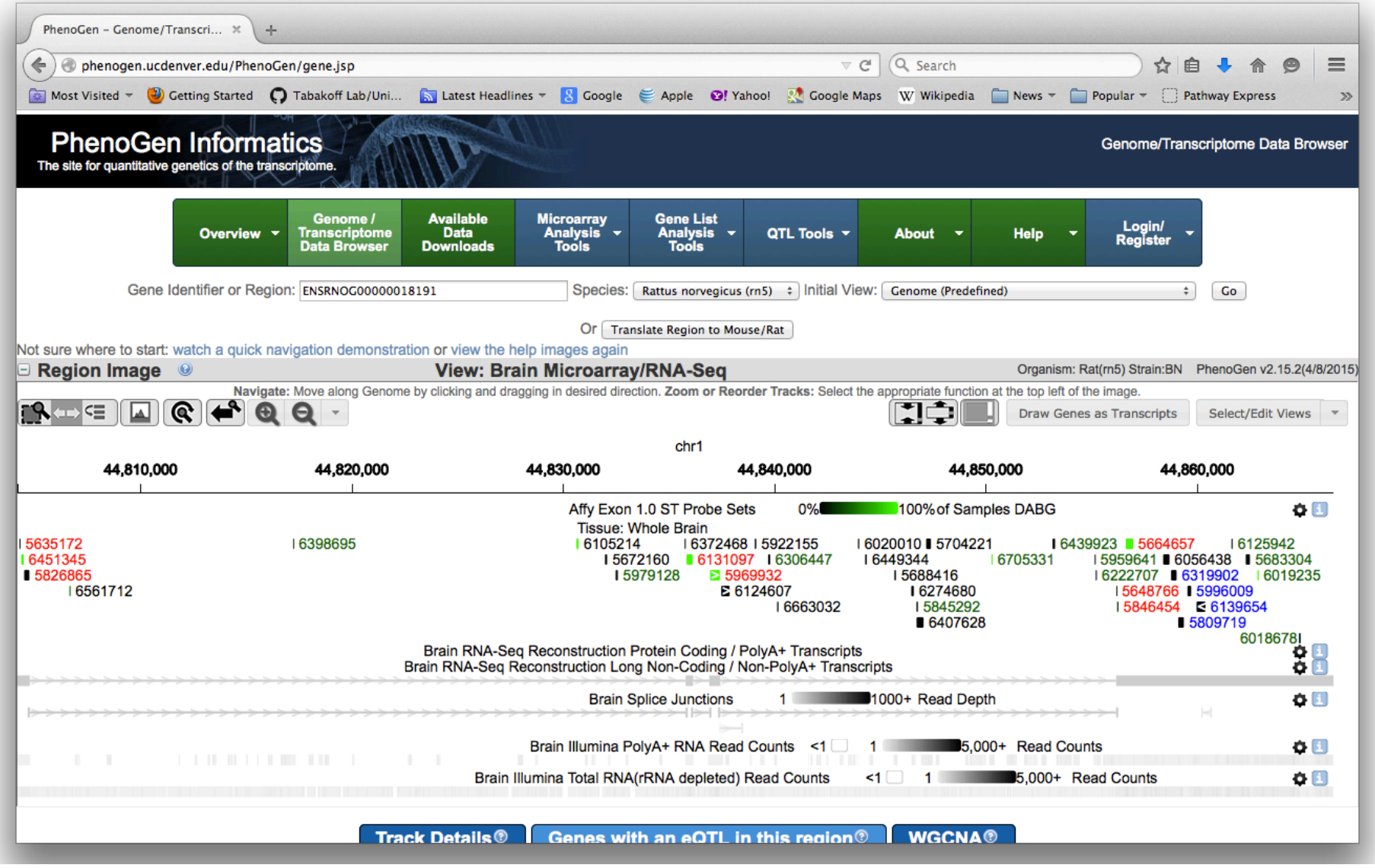


1. Click “Select/Edit Views” box.
2. Select “Transcriptome (Predefined) (6 trackes” in Select a view below box
3. Click “Apply View” button

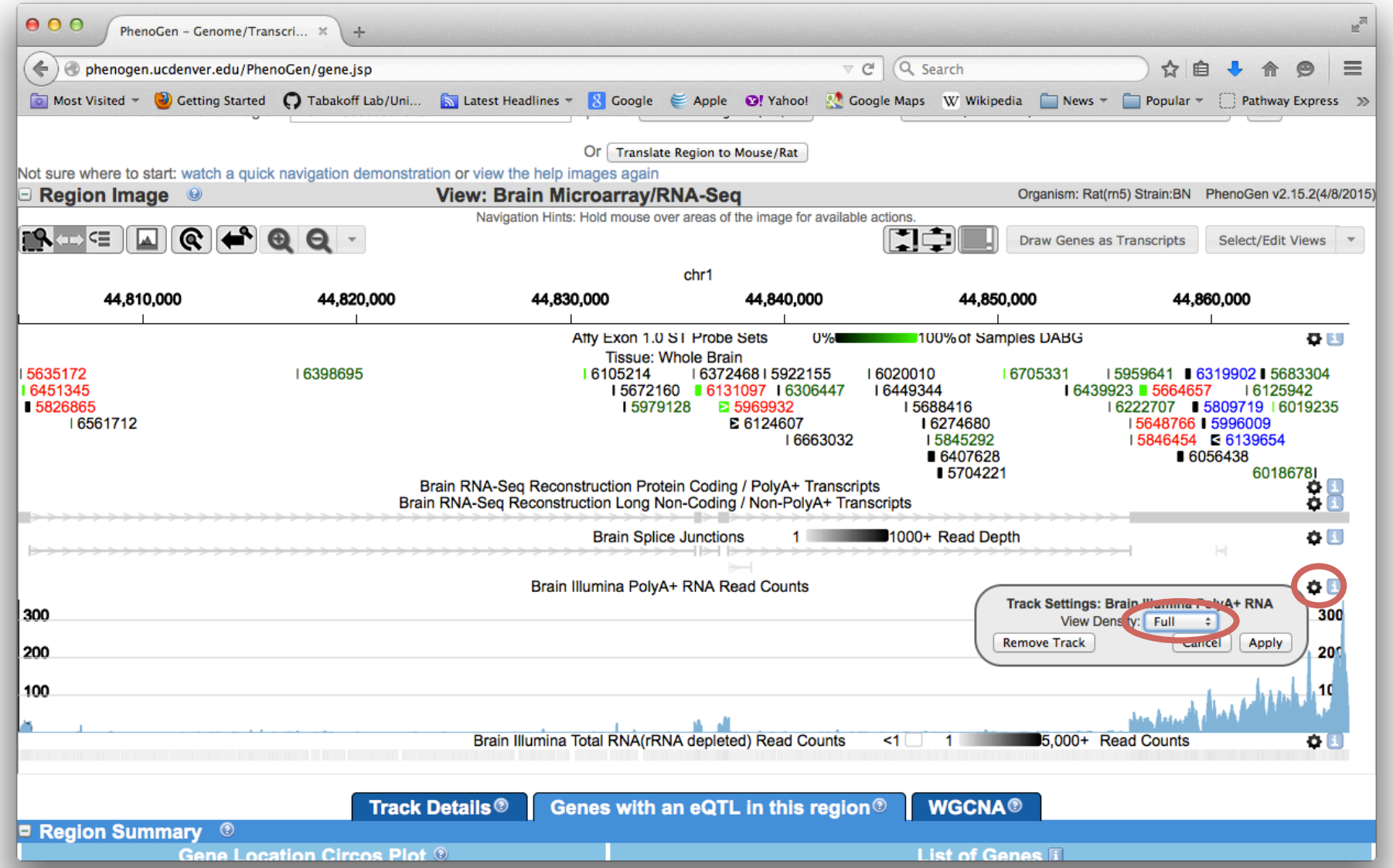
Oprm1 – Only in Brain Reconstruction



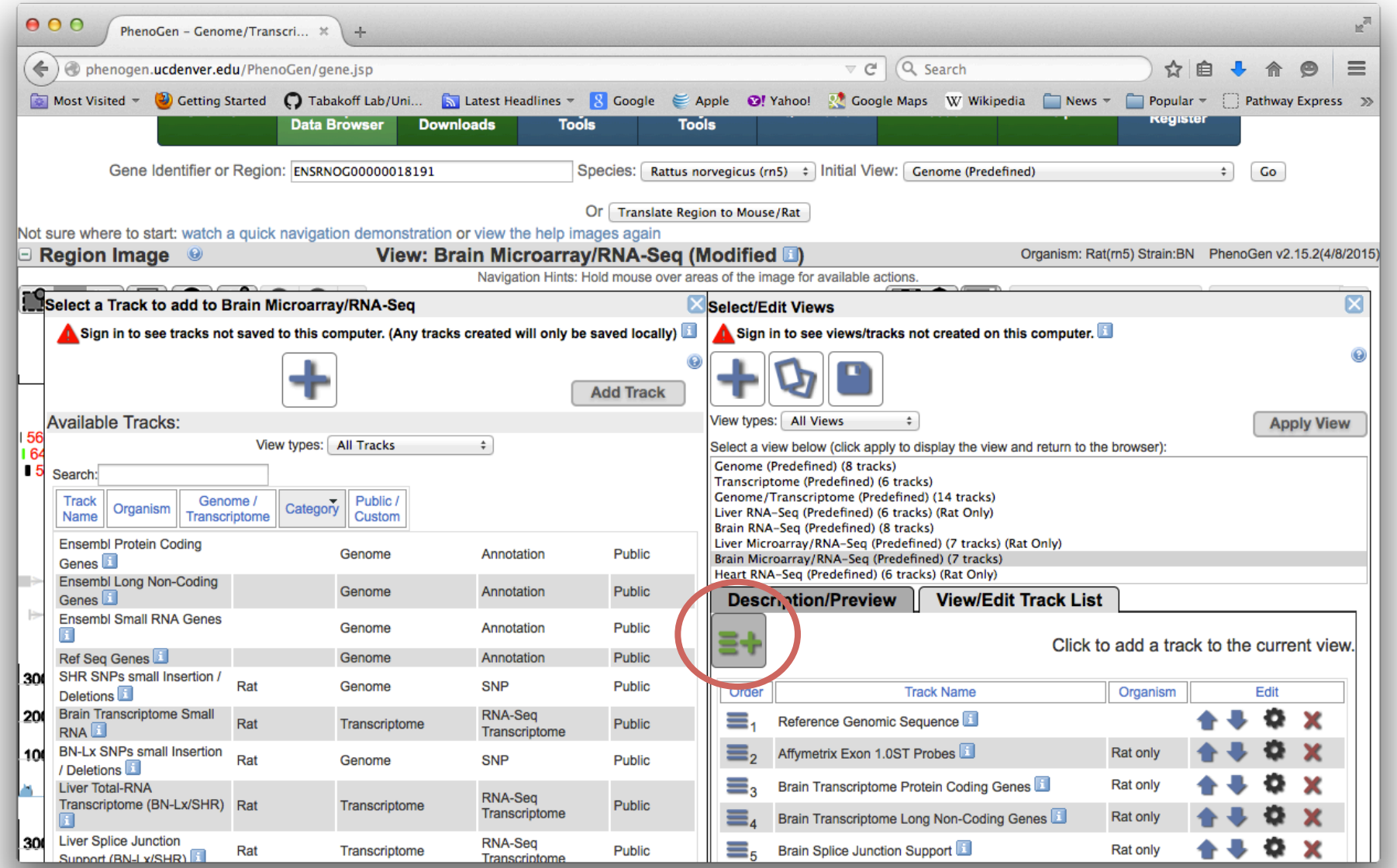
Brain Microarray/RNA-Seq View



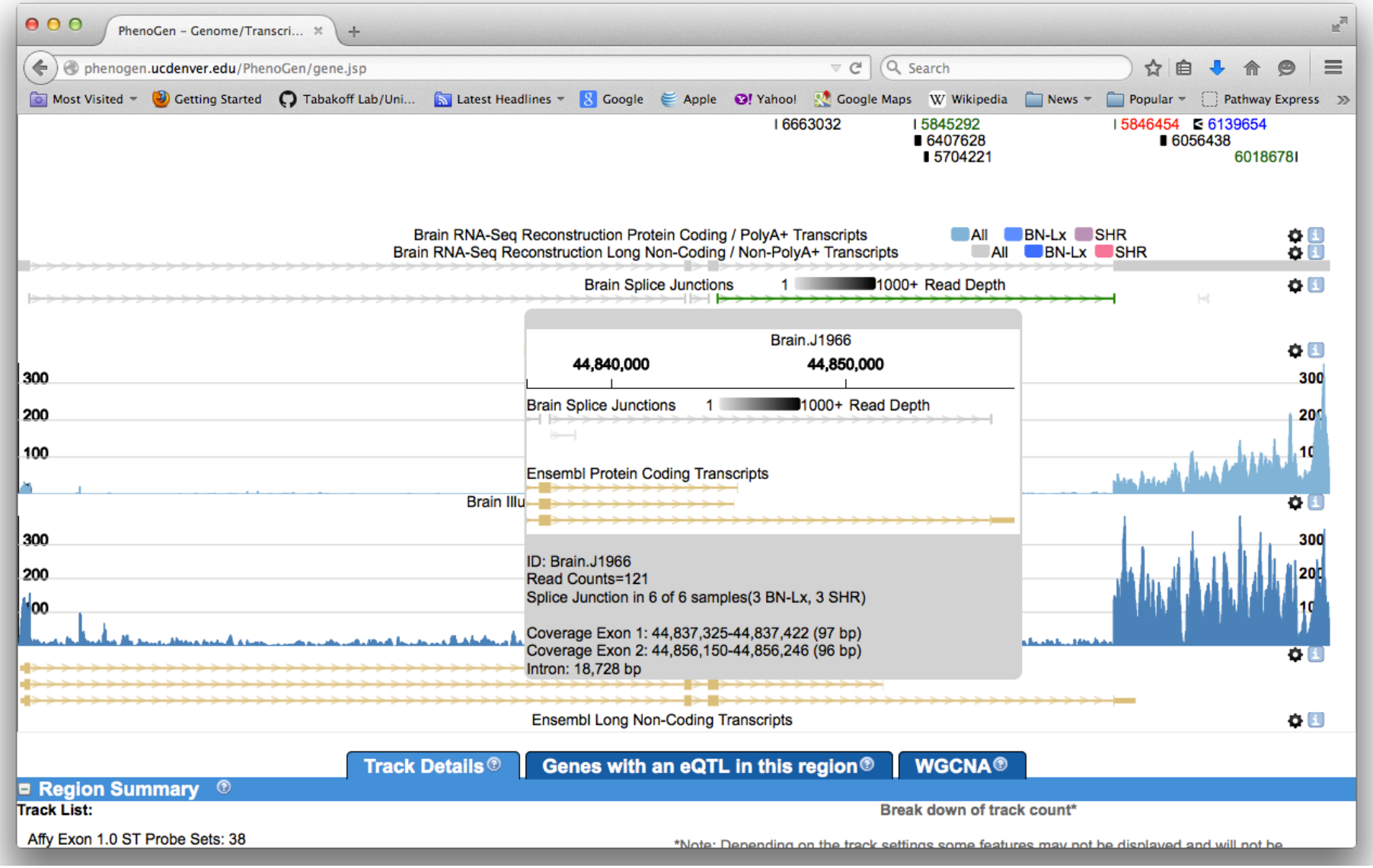
Expand RNA-Seq Read Counts



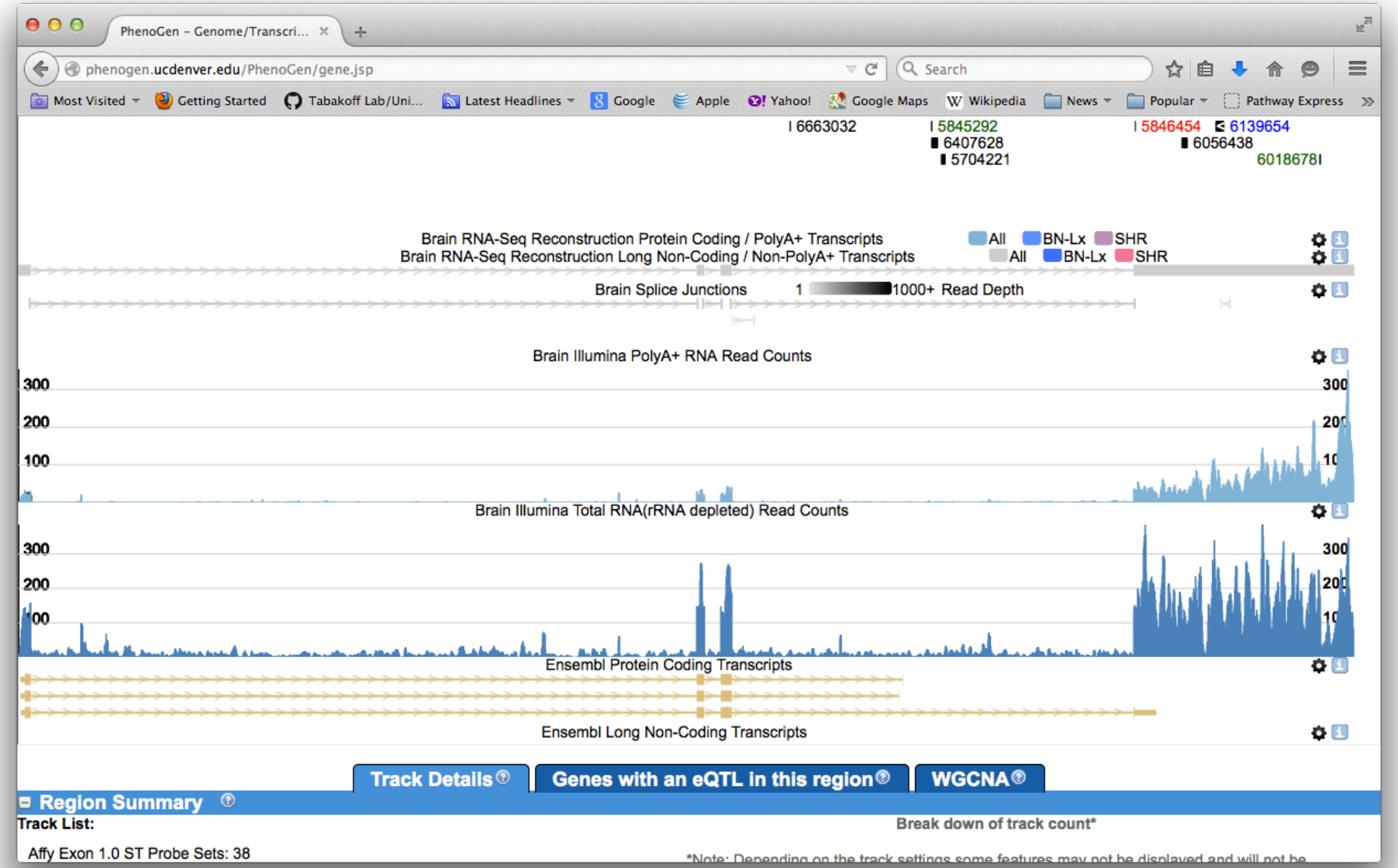
Add Ensembl Track



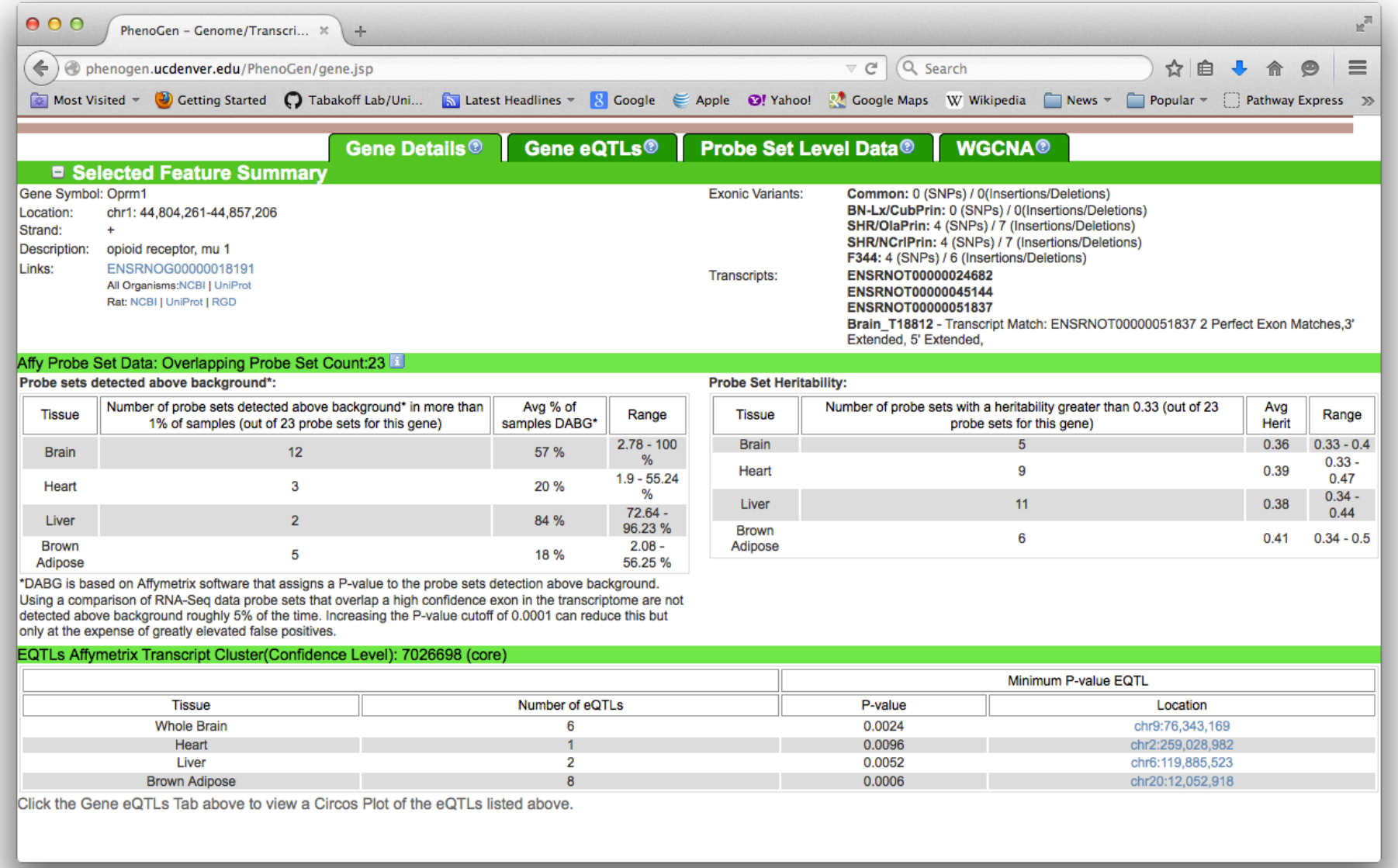
Junction Reads



Alternative 3’ UTR



Selected Feature Summary



PhenoGen - Genome/Transcri...

phenogen.ucdenver.edu/PhenoGen/gene.jsp

Search

Most Visited | Getting Started | Tabakoff Lab/Uni... | Latest Headlines | Google | Apple | Yahoo! | Google Maps | Wikipedia | News | Popular | Pathway Express

Gene Details | Gene eQTLs | Probe Set Level Data | WGCNA

Selected Feature Summary

Gene Symbol: Oprm1
Location: chr1: 44,804,261-44,857,206
Strand: +
Description: opioid receptor, mu 1
Links: ENSRNOG00000018191
All Organisms:NCBI | UniProt
Rat: NCBI | UniProt | RGD

Exonic Variants:
Common: 0 (SNPs) / 0(Insertions/Deletions)
BN-Lx/CubPrin: 0 (SNPs) / 0(Insertions/Deletions)
SHR/OlaPrin: 4 (SNPs) / 7 (Insertions/Deletions)
SHR/NCrlPrin: 4 (SNPs) / 7 (Insertions/Deletions)
F344: 4 (SNPs) / 6 (Insertions/Deletions)

Transcripts:
ENSRNOT00000024682
ENSRNOT00000045144
ENSRNOT00000051837
Brain_T18812 - Transcript Match: ENSRNOT00000051837 2 Perfect Exon Matches,3' Extended, 5' Extended,

Affy Probe Set Data: Overlapping Probe Set Count:23

Probe sets detected above background*:

Tissue	Number of probe sets detected above background* in more than 1% of samples (out of 23 probe sets for this gene)	Avg % of samples DABG*	Range
Brain	12	57 %	2.78 - 100 %
Heart	3	20 %	1.9 - 55.24 %
Liver	2	84 %	72.64 - 96.23 %
Brown Adipose	5	18 %	2.08 - 56.25 %

*DABG is based on Affymetrix software that assigns a P-value to the probe sets detection above background. Using a comparison of RNA-Seq data probe sets that overlap a high confidence exon in the transcriptome are not detected above background roughly 5% of the time. Increasing the P-value cutoff of 0.0001 can reduce this but only at the expense of greatly elevated false positives.

Probe Set Heritability:

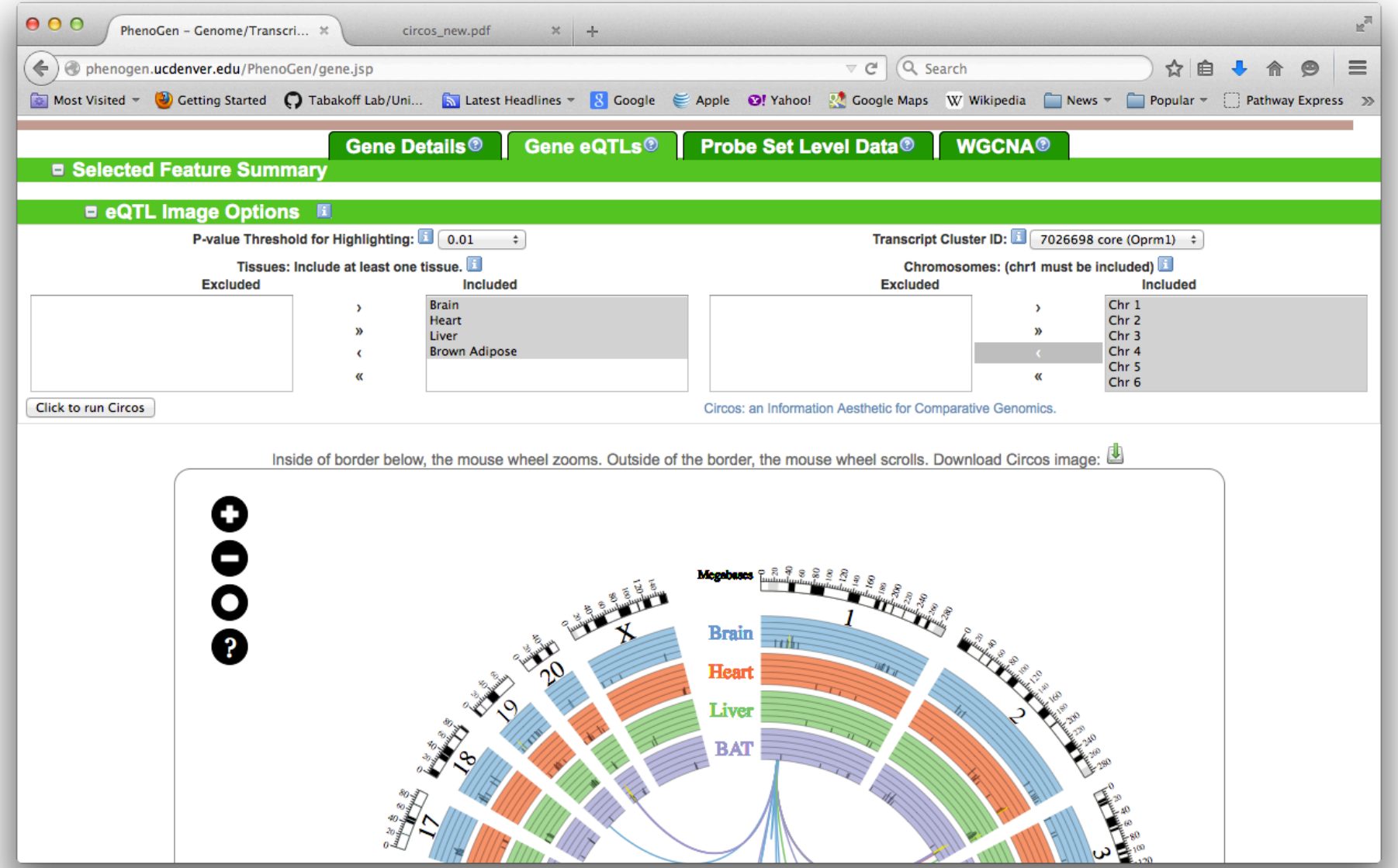
Tissue	Number of probe sets with a heritability greater than 0.33 (out of 23 probe sets for this gene)	Avg Herit	Range
Brain	5	0.36	0.33 - 0.4
Heart	9	0.39	0.33 - 0.47
Liver	11	0.38	0.34 - 0.44
Brown Adipose	6	0.41	0.34 - 0.5

EQTLs Affymetrix Transcript Cluster(Confidence Level): 7026698 (core)

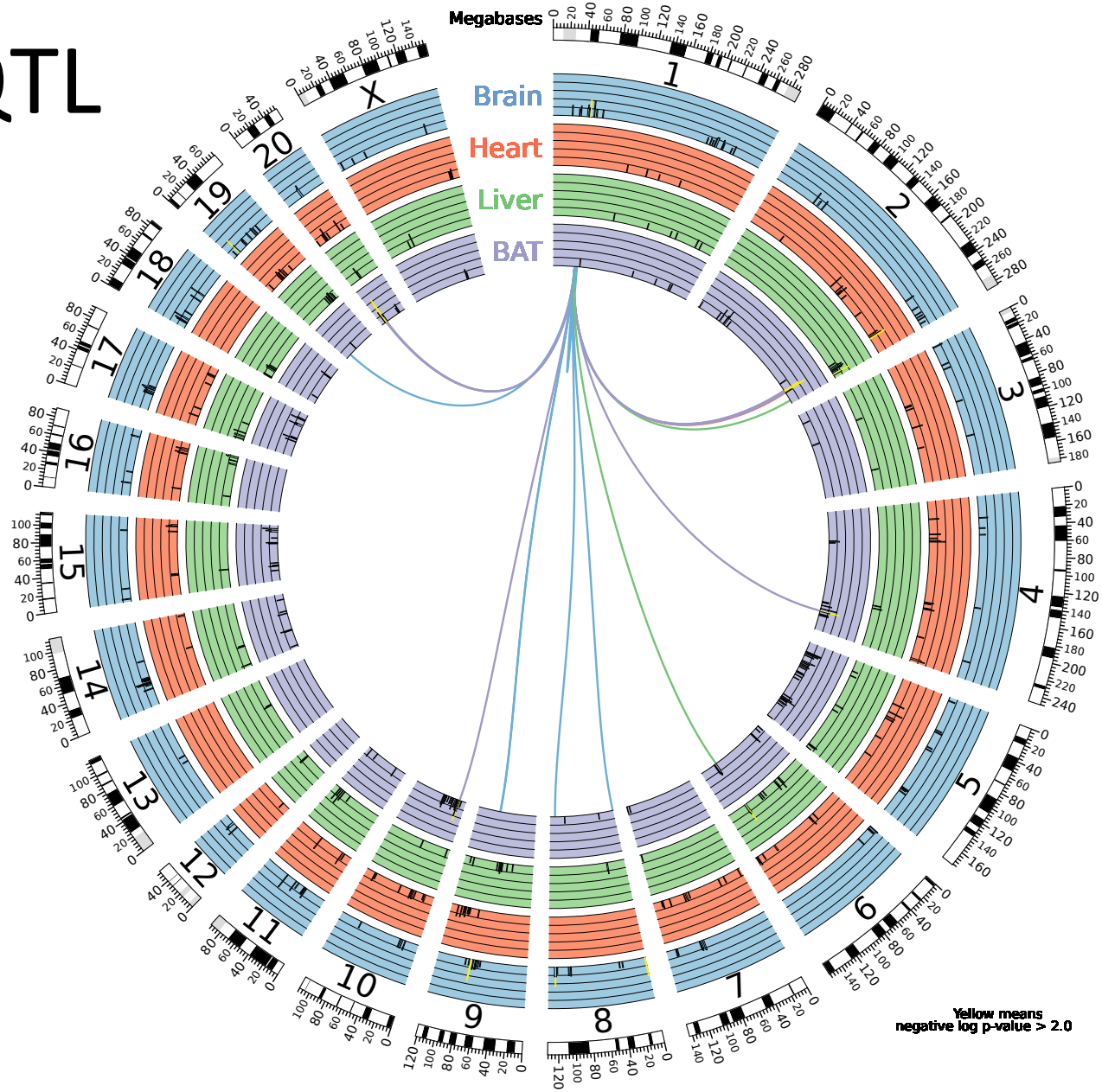
		Minimum P-value EQTL	
Tissue	Number of eQTLs	P-value	Location
Whole Brain	6	0.0024	chr9:76,343,169
Heart	1	0.0096	chr2:259,028,982
Liver	2	0.0052	chr6:119,885,523
Brown Adipose	8	0.0006	chr20:12,052,918

Click the Gene eQTLs Tab above to view a Circos Plot of the eQTLs listed above.

Gene eQTL


PhenoGen – Genome/Transcri...
circos_new.pdf
phenogen.ucdenver.edu/PhenoGen/gene.jsp
Search
Most Visited
Getting Started
Tabakoff Lab/Uni...
Latest Headlines
Google
Apple
Yahoo!
Google Maps
Wikipedia
News
Popular
Pathway Express
Gene Details
Gene eQTLs
Probe Set Level Data
WGCNA
Selected Feature Summary
eQTL Image Options
P-value Threshold for Highlighting:
0.01
Transcript Cluster ID:
7026698 core (Oprm1)
Tissues: Include at least one tissue.
Excluded
Included
Brain
Heart
Liver
Brown Adipose
Chromosomes: (chr1 must be included)
Excluded
Included
Chr 1
Chr 2
Chr 3
Chr 4
Chr 5
Chr 6
Click to run Circos
Circos: an Information Aesthetic for Comparative Genomics.
Inside of border below, the mouse wheel zooms. Outside of the border, the mouse wheel scrolls. Download Circos image:
Megabases
Brain
Heart
Liver
BAT

Gene eQTL


Megabases
Brain
Heart
Liver
BAT
Yellow means
negative log p-value > 2.0

WEIGHTED GENE CO-EXPRESSION NETWORK ANALYSIS (WGCNA)

What are we missing by considering each candidate gene individually?

- Define relationships among genes
- Infer biological function from other co-expressed genes
- Define the context in which the gene exerts its effect
- Find multiple therapeutic targets within the same pathway

What do we gain by building networks and identifying modules?

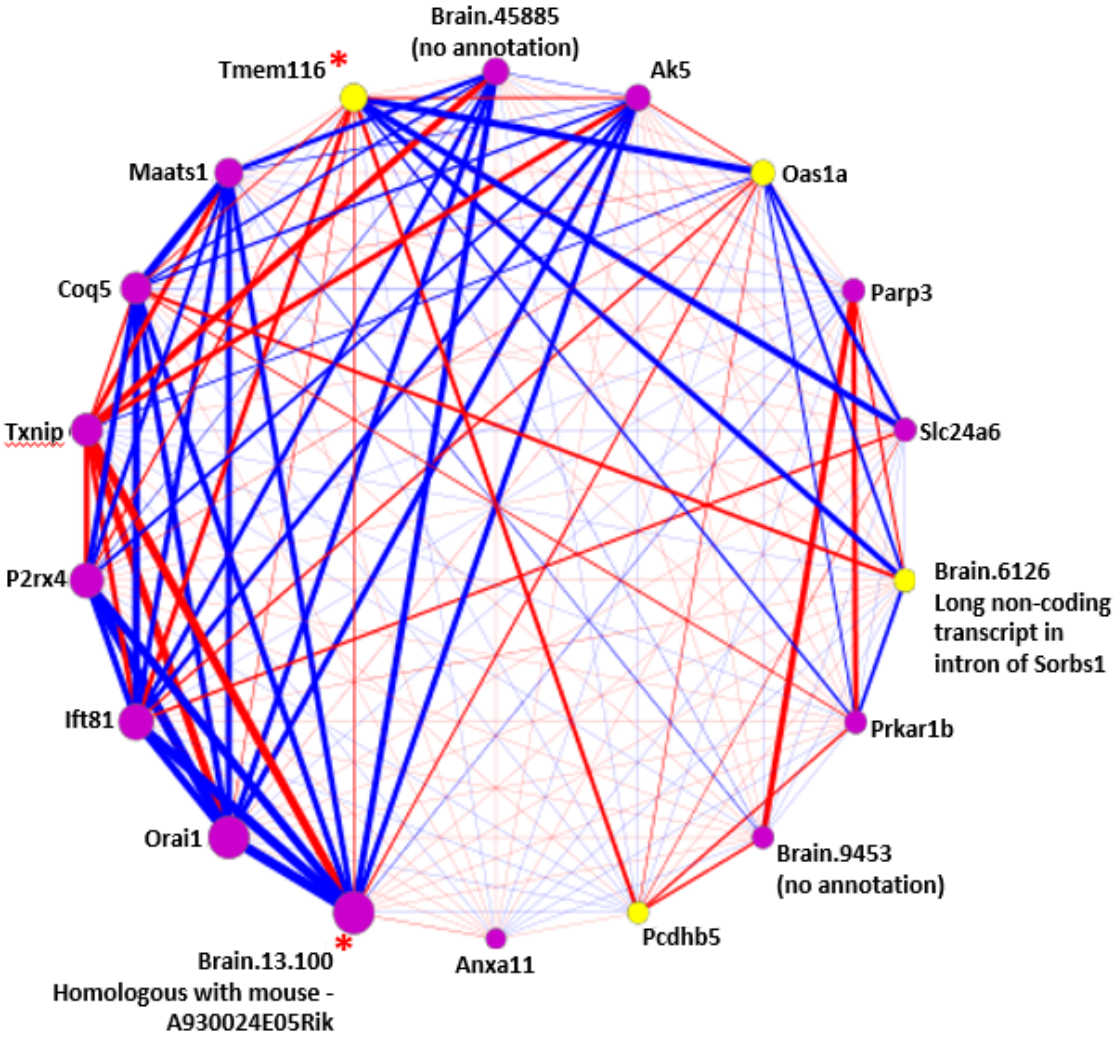
- No gene product acts **independently** in the cell
- Information about biological function in cell
- Information on causes/ consequences of differential expression

Weighted Gene Co-Expression Network Analysis
Why Not Just Use Correlation?

1. Simple correlation does not give connectivity.
2. How are we measuring co-expression?
   - Scale-Free Network
     - Network has few highly connected genes rather then each gene have similar connectivity
     - **Biologically motivated**, fewer highly connected genes means that a system is more robust to failure of any one gene
3. How do we get a **robust** measure of connectivity for identifying modules?
   - Topological Overlap Measure
     - Includes a measure of how many "friends" two genes have in common
     - Protects against spurious correlations among genes

Example Co-Expression Module

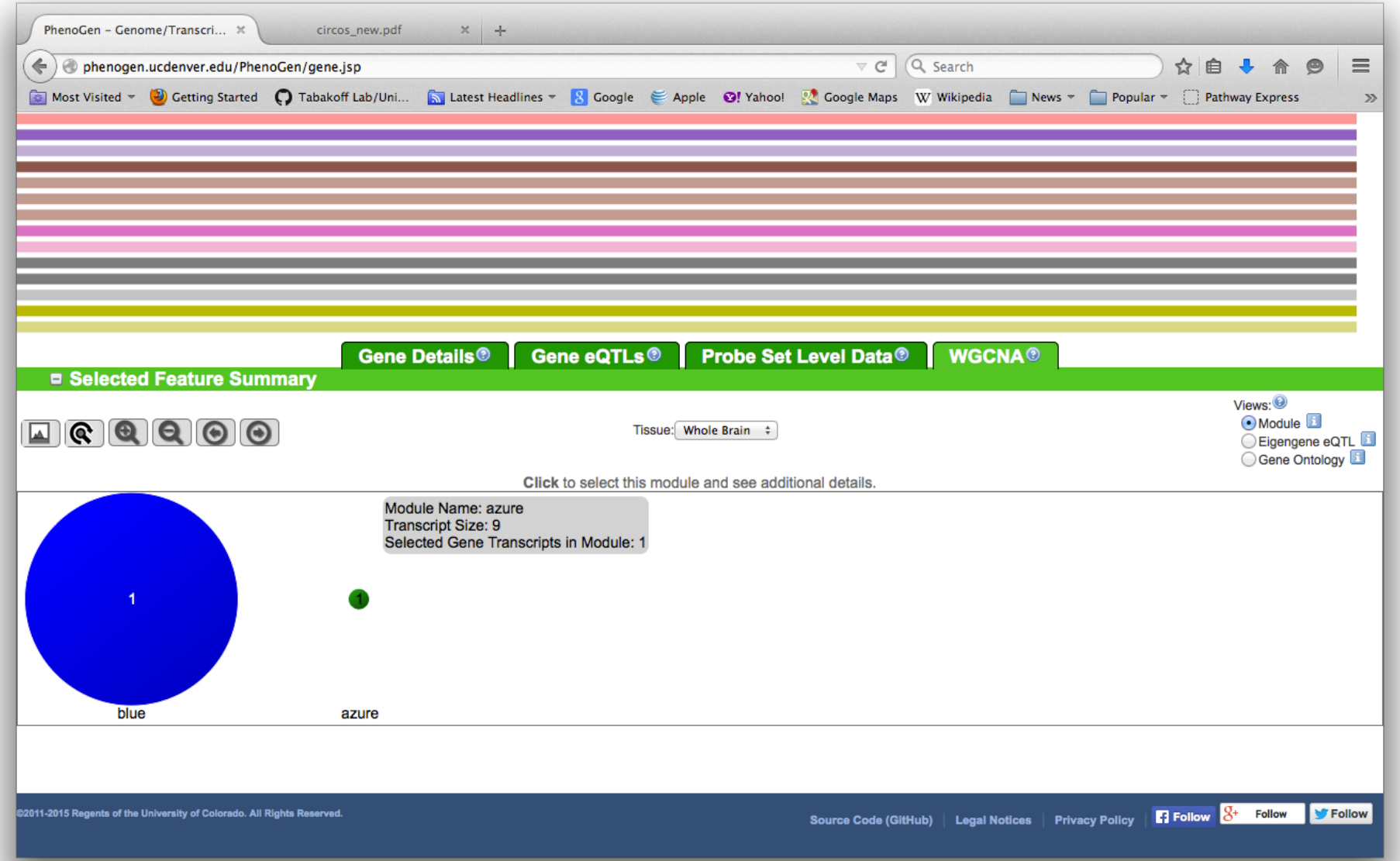
- **Edge thickness is weighted based on magnitude of correlation between nodes.**
- **Blue edges represent a positive correlation between nodes.**
- **Red edges represent a negative correlation between nodes.**
- **Node size is weighted based on connectivity within module.**



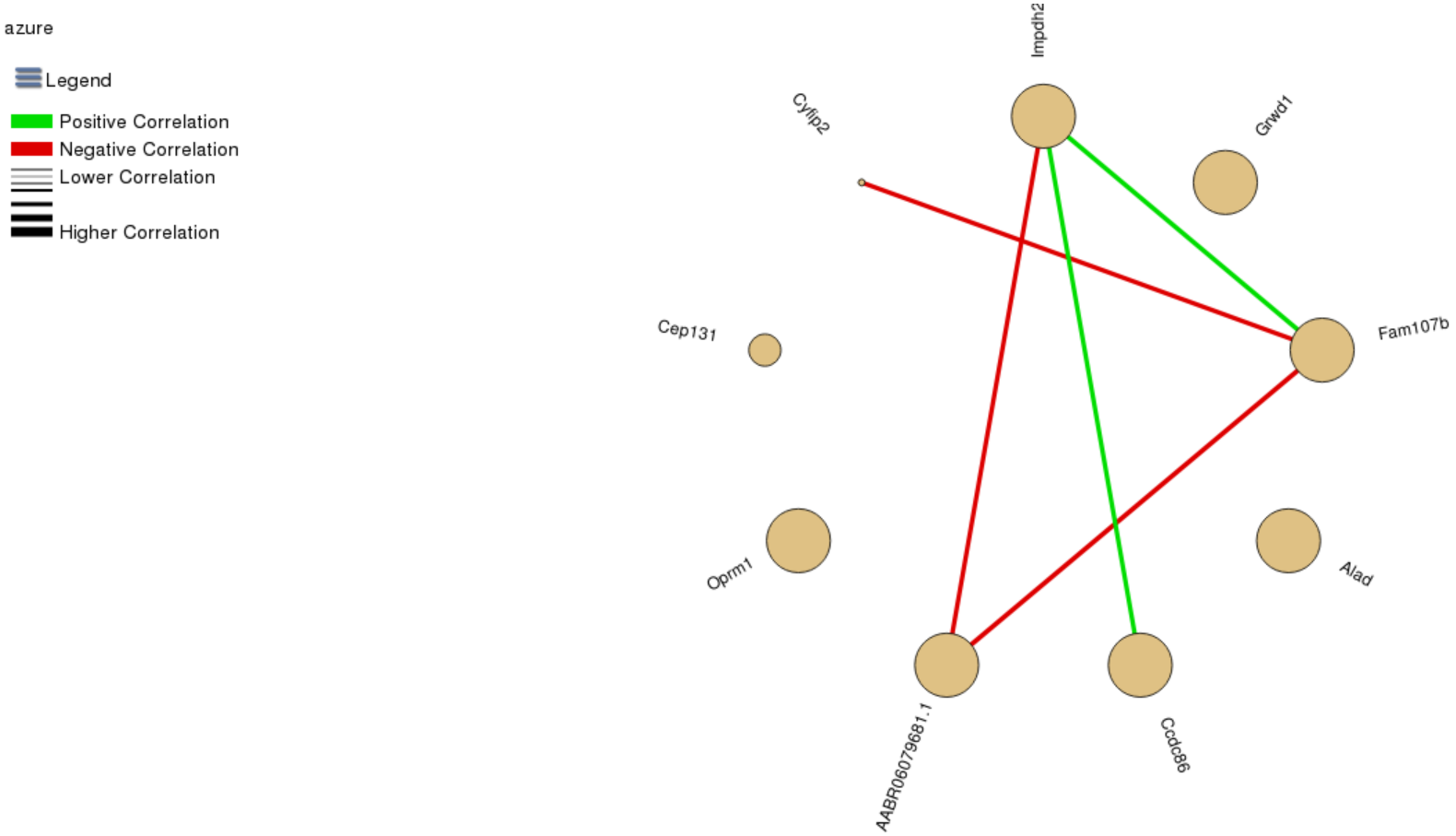
Summary of Expression Pattern Within Co-Expression Module

- Eigengene
  - First Principal Component
    - Maximize the amount of variance in expression captured by a single value per strain/sample
- Hub Gene
  - Most ‘connected’ gene within a module
  - May have biological implications

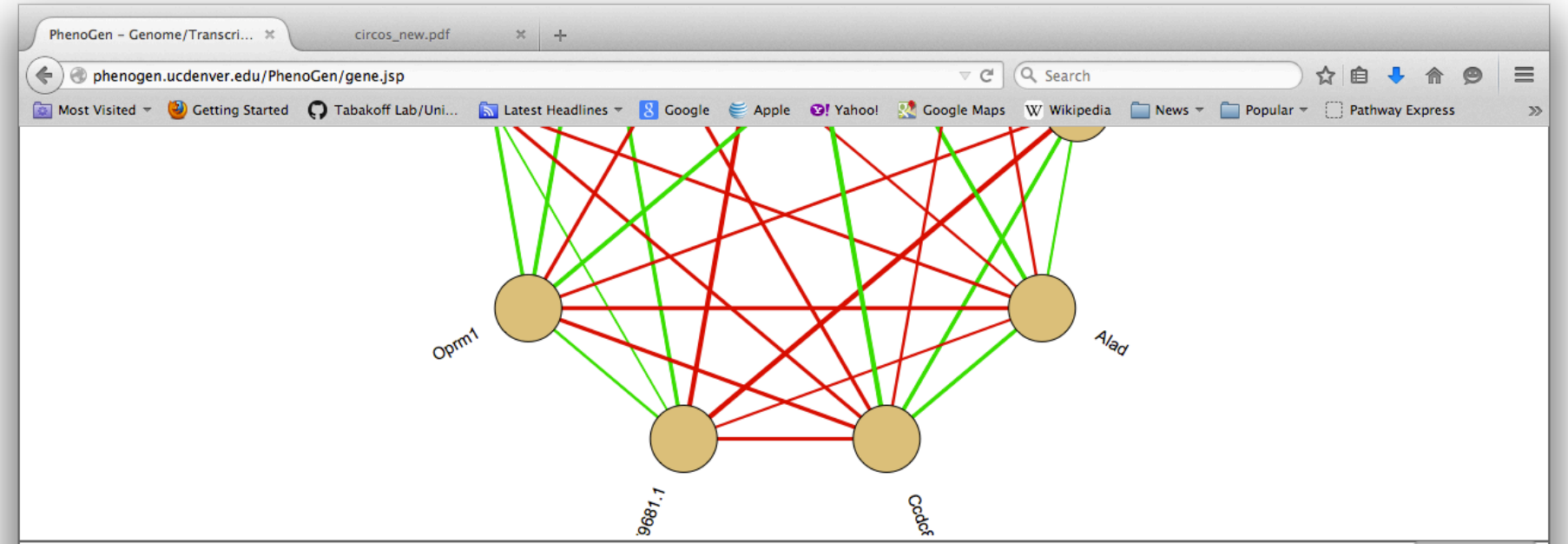
WGCNA


PhenoGen – Genome/Transcri...
circos_new.pdf
phenogen.ucdenver.edu/PhenoGen/gene.jsp
Search
Most Visited
Getting Started
Tabakoff Lab/Uni...
Latest Headlines
Google
Apple
Yahoo!
Google Maps
Wikipedia
News
Popular
Pathway Express
Gene Details
Gene eQTLs
Probe Set Level Data
WGCNA
Selected Feature Summary
Tissue: Whole Brain
Views:
Module
Eigengene eQTL
Gene Ontology
Click to select this module and see additional details.
Module Name: azure
Transcript Size: 9
Selected Gene Transcripts in Module: 1
1
1
blue
azure
©2011-2015 Regents of the University of Colorado. All Rights Reserved.
Source Code (GitHub)
Legal Notices
Privacy Policy
Follow
Follow
Follow

Azure Module



Azure Module Table



Transcripts in azure Module

Export CSV

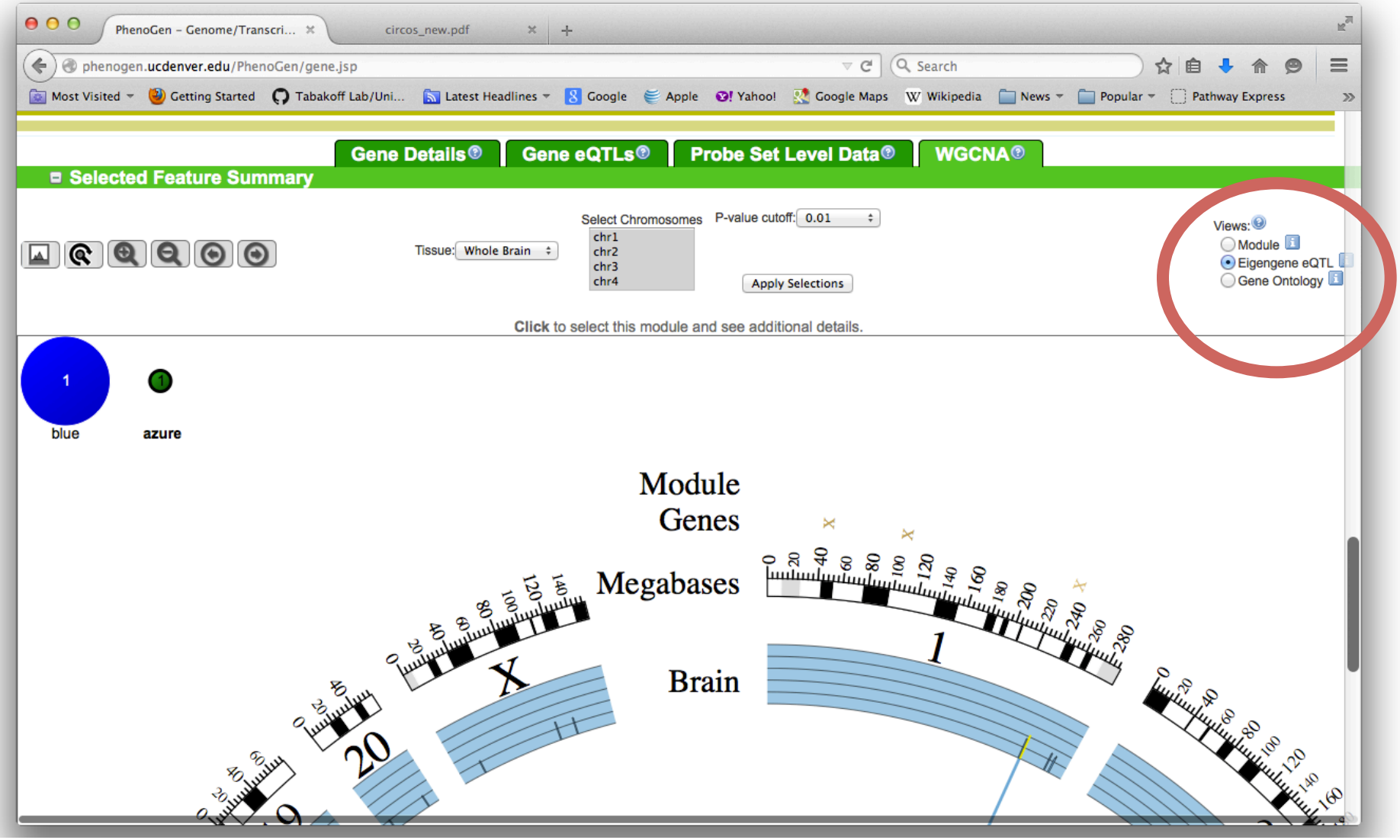
Search:

Showing 1 to 9 of 9 entries

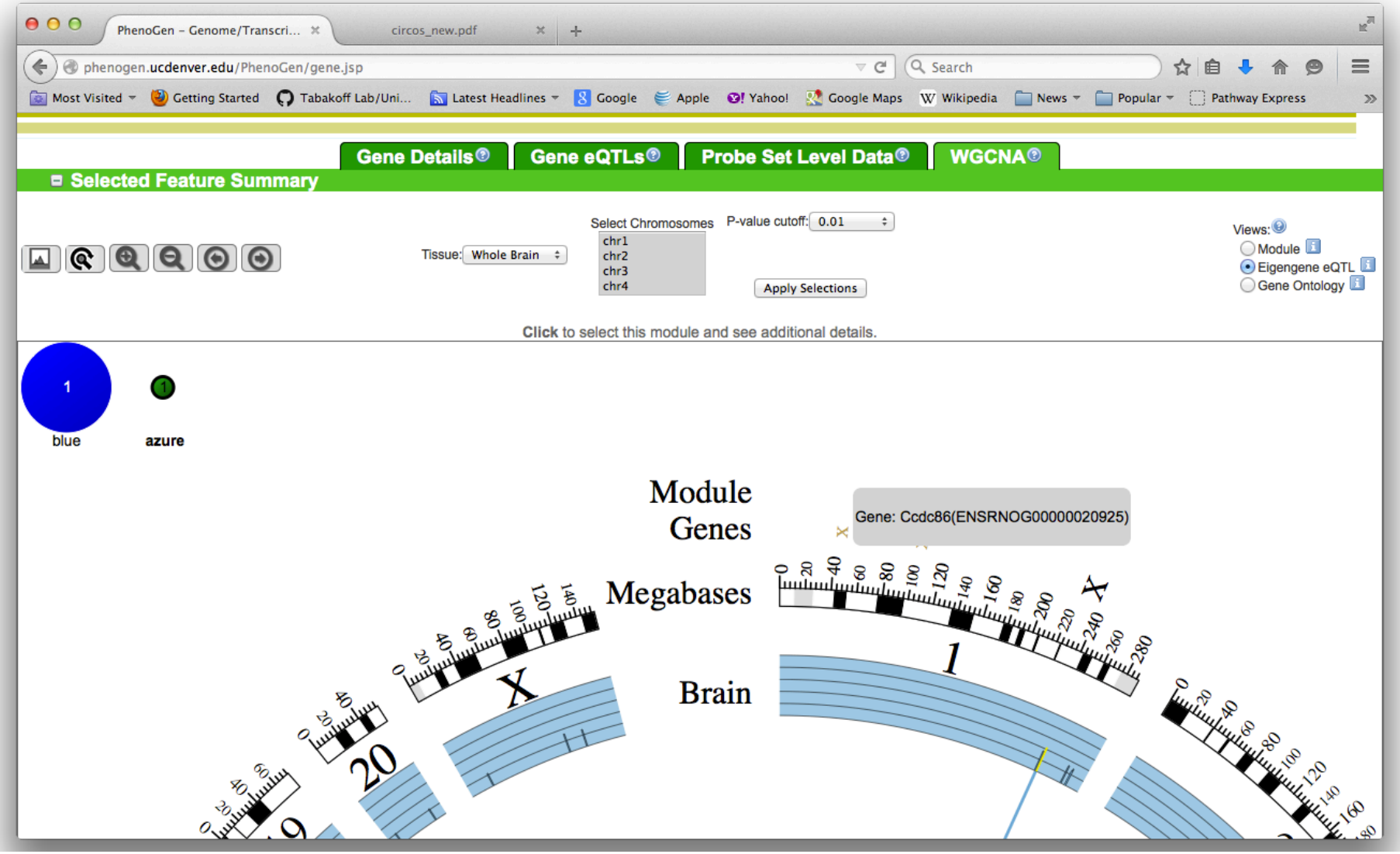
Gene Symbol	Gene ID	Transcript	Probe Sets	Link Total	Link Rank
Impdh2	ENSRNOG00000031965	Brain_C8804.3	1	4.90	1
Grwd1	ENSRNOG00000021058	Brain_C9091.1	1	3.48	2
Fam107b	ENSRNOG00000014886	Brain_C6674.1	5	3.21	3
Alad	ENSRNOG00000015206	Brain_C6729.2	1	3.14	4
Ccdc86	ENSRNOG00000020925	Brain_C4185.1	1	2.48	5
AABR06079681.1	ENSRNOG00000036375	Brain_C4097.4	1	1.41	6
Oprm1	ENSRNOG00000018191	Brain_C29606.1	1	1.29	7
Cep131	ENSRNOG00000004430	Brain_C1276.2	1	0.55	8
Cyfip2	ENSRNOG00000006557	Brain_C1260.2	2	0.00	9

Source Code (GitHub) | Legal Notices | Privacy Policy | Follow | Follow | Follow

Eigengene eQTL

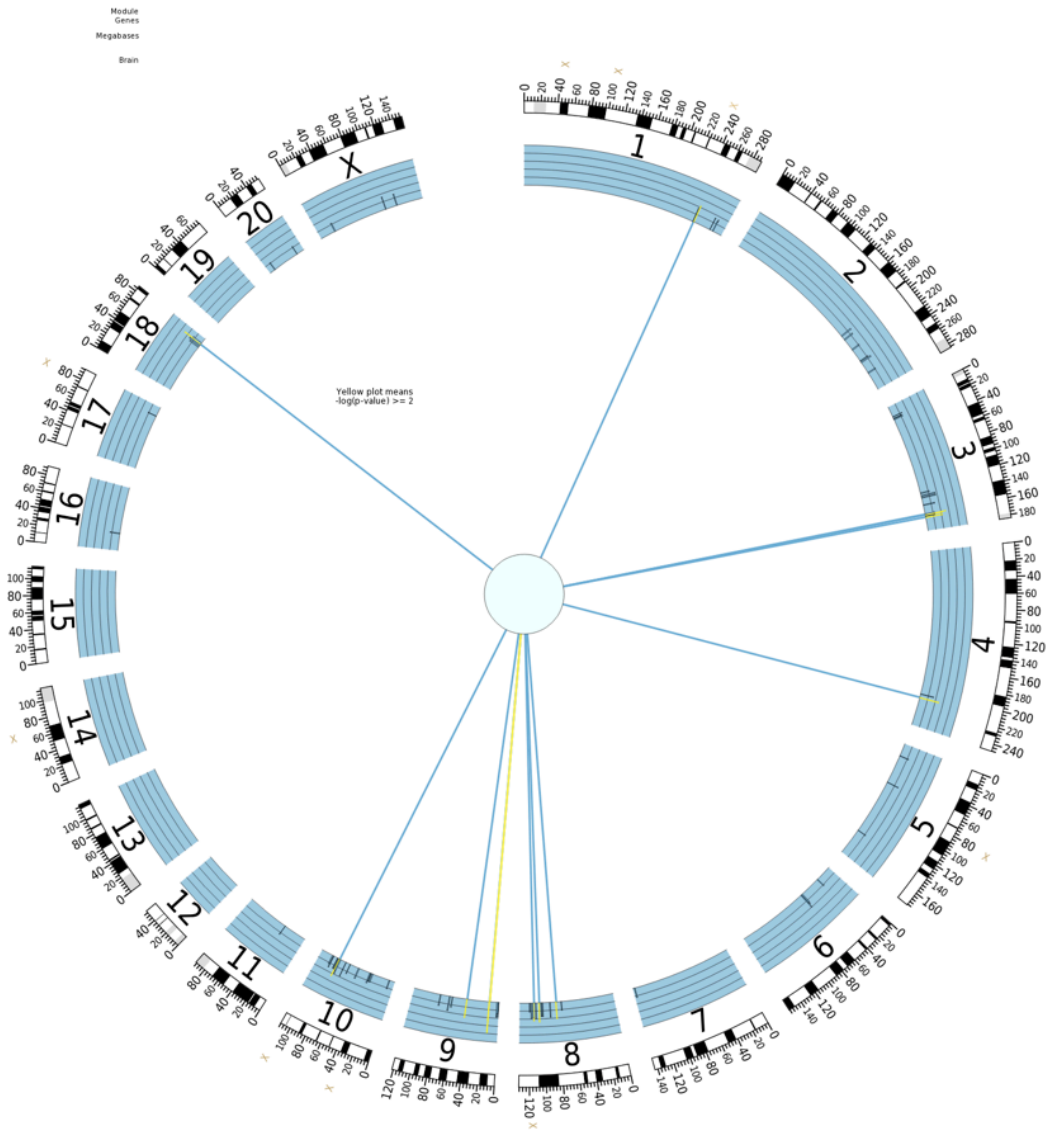


Physical Location of Genes with Module

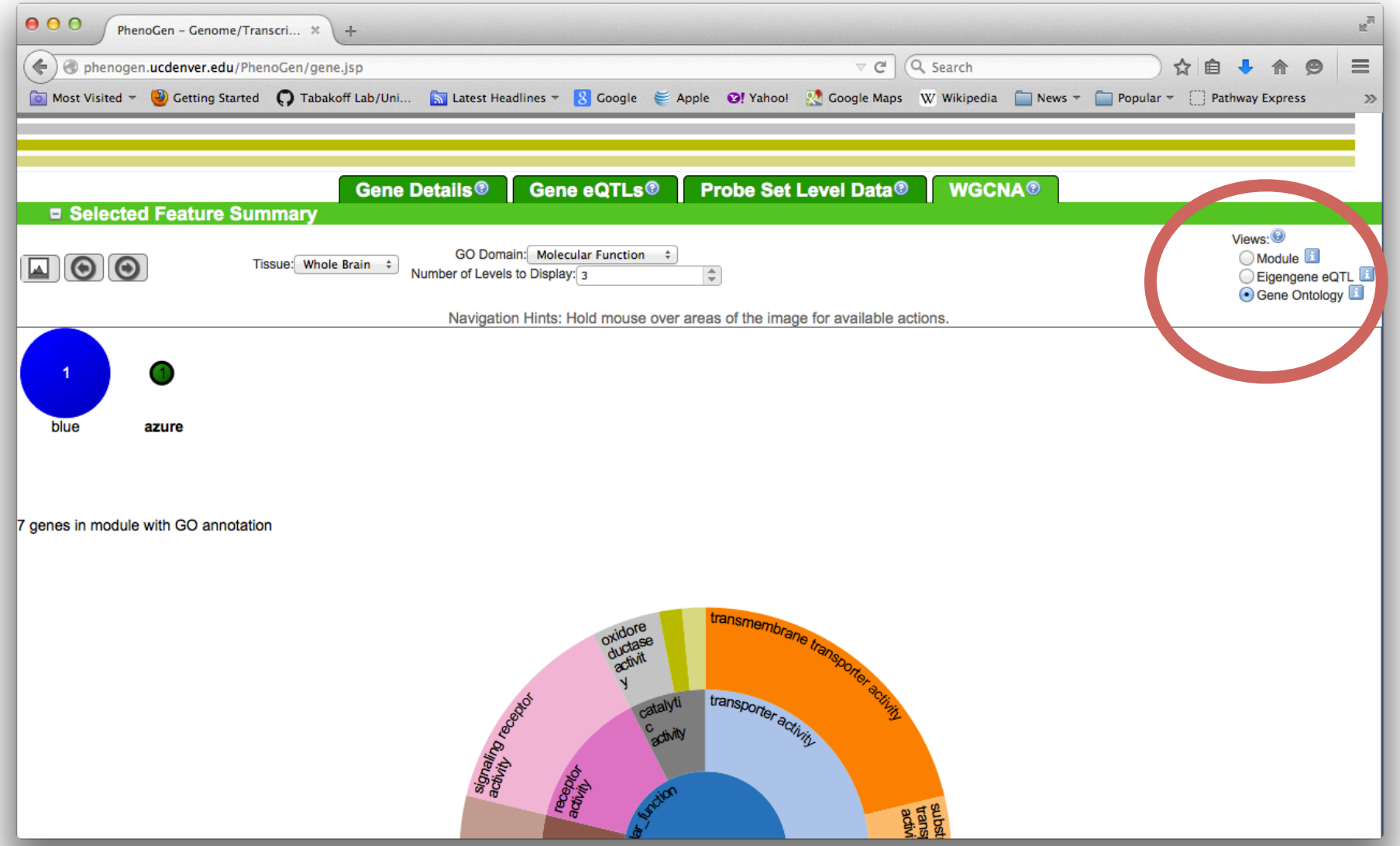


Circos

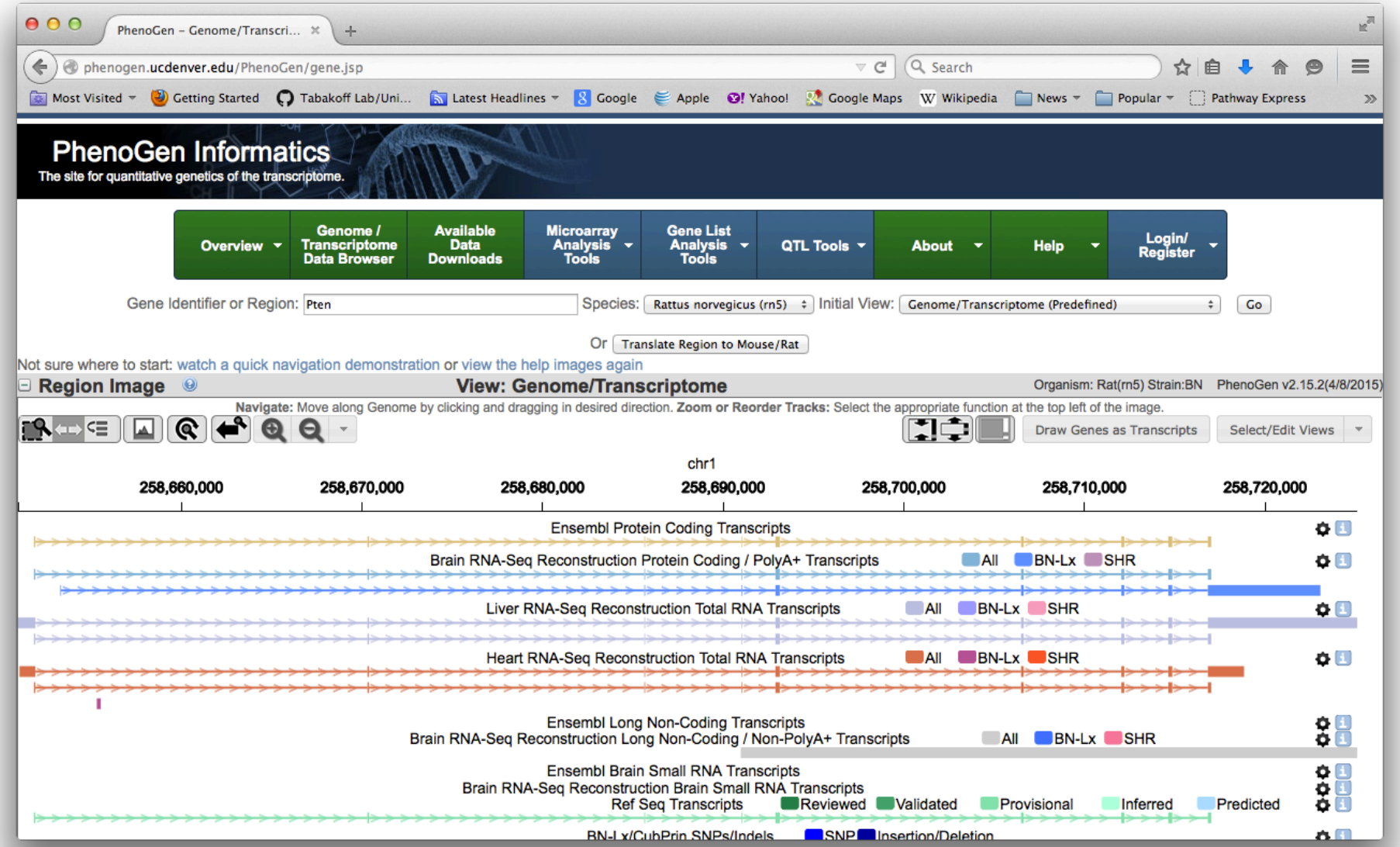
Eigengene eQTL



Gene Ontology

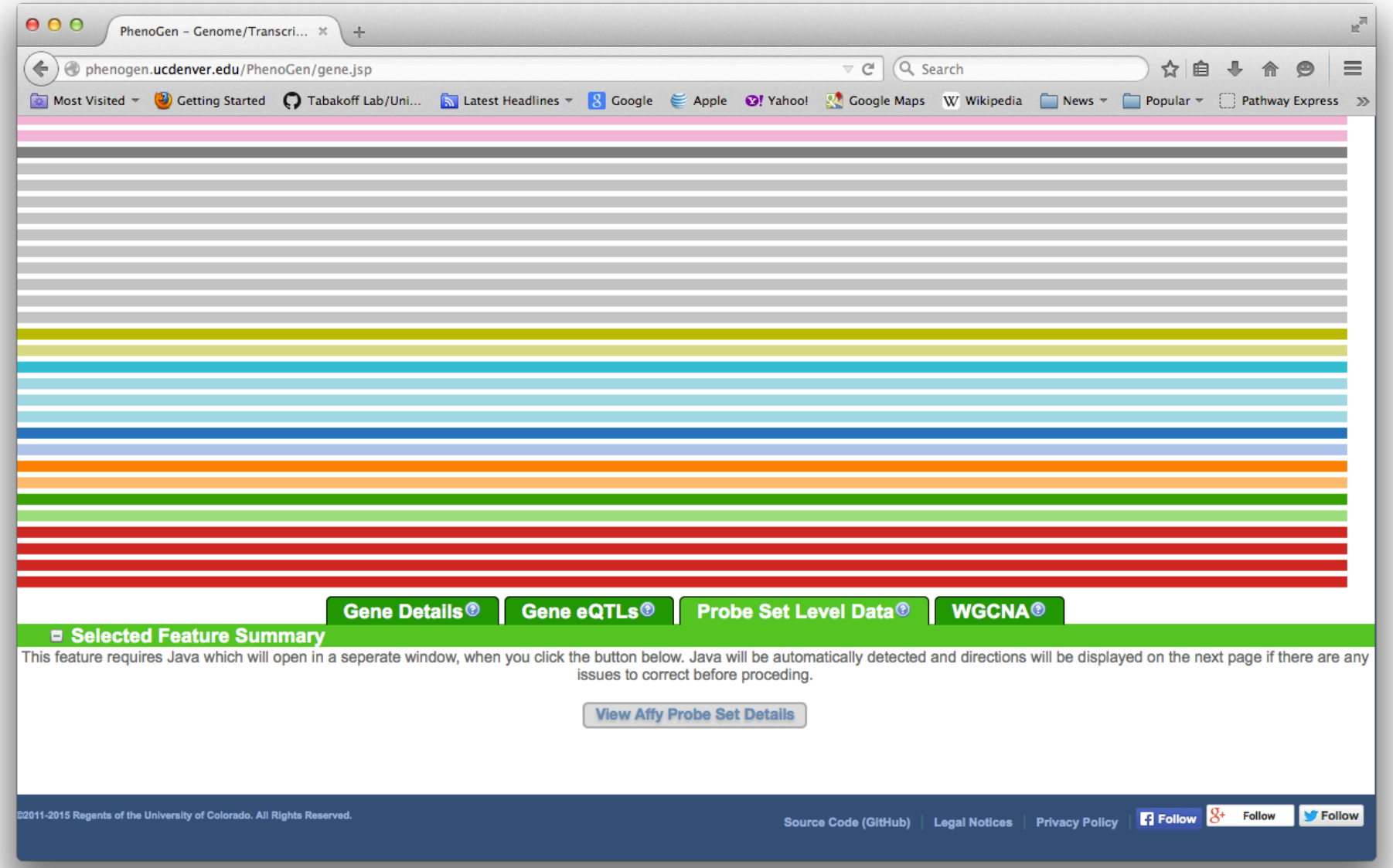


RNA Expression Across Tissues

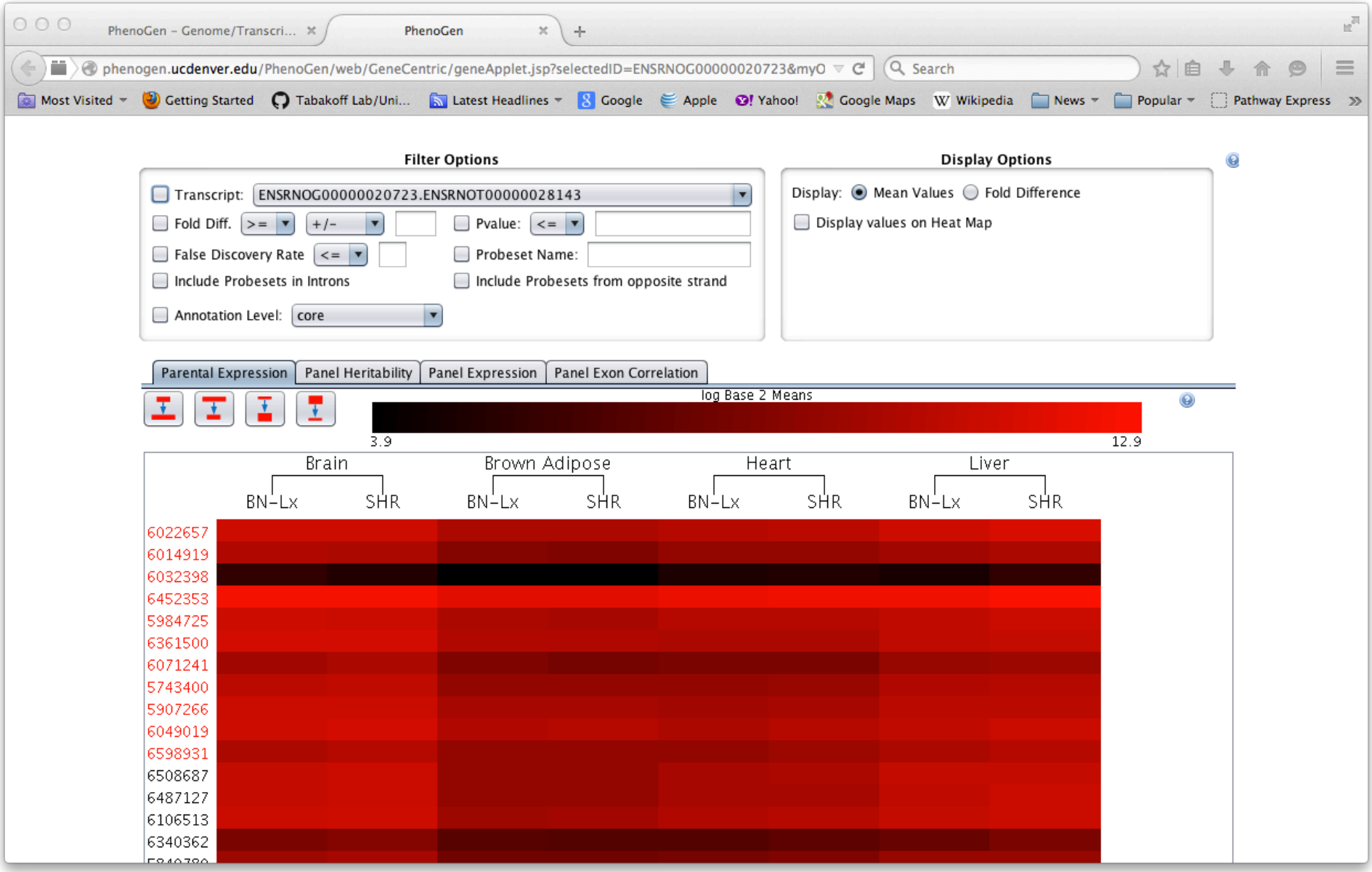


Probe Set Level Data

Probe Set Level Data needs to collect and parse lots of data.
Click on the tab, and wait a minute before pushing the View Affy Probe Set Details button

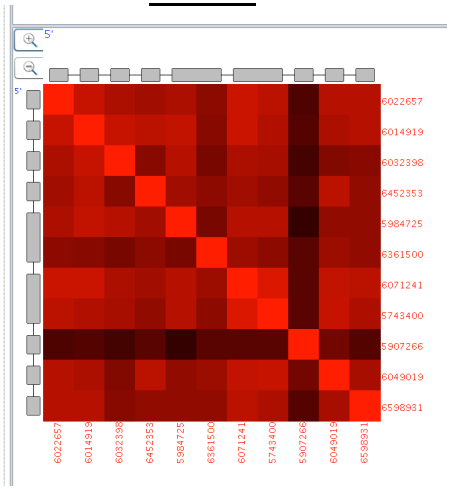


Differences Between Parental Strains Across Tissues

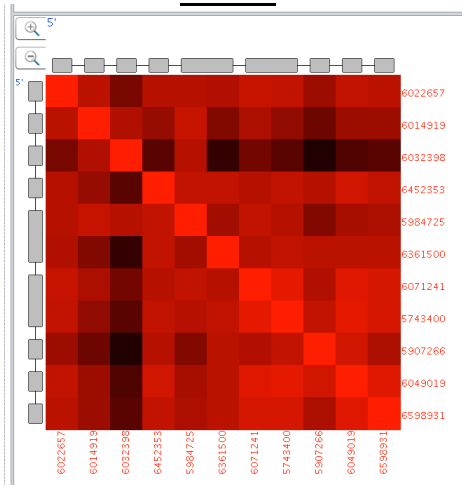


Panel Exon Correlation In Different Tissues

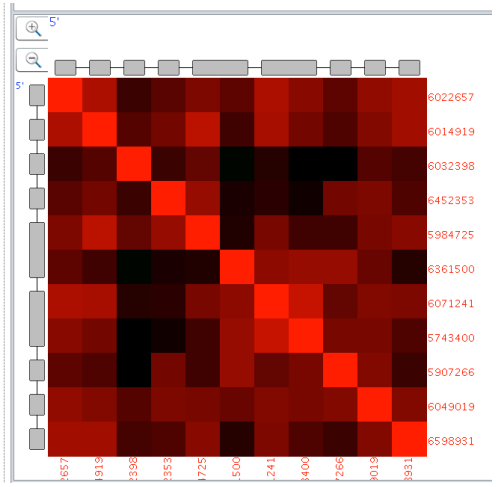
Brain



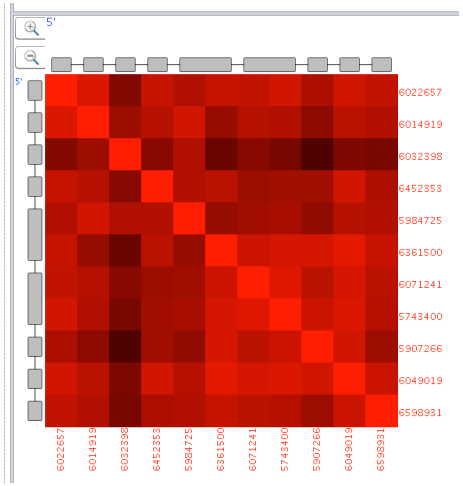
Liver



Heart



Brown Adipose



FROM PHENOTYPE TO CANDIDATE GENES

Genetical Genomics/Phenomics Approach

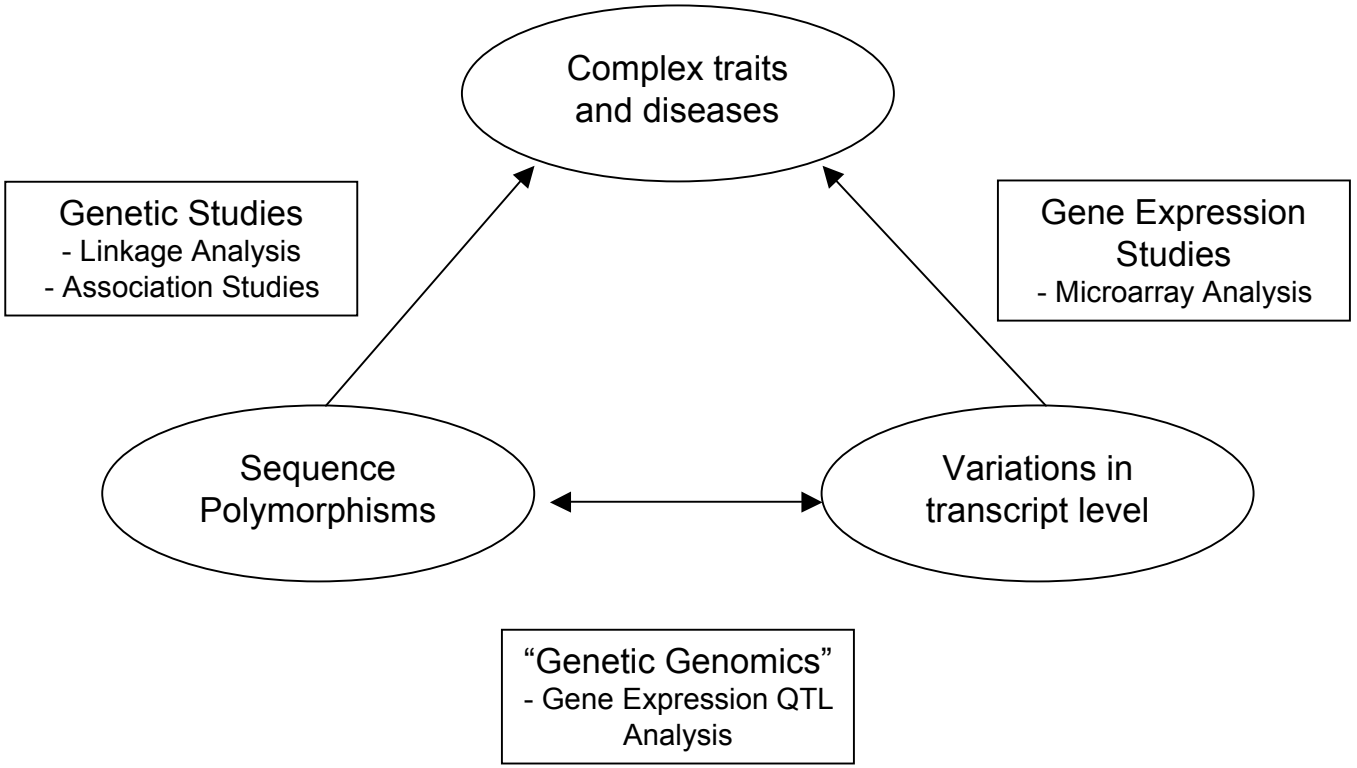


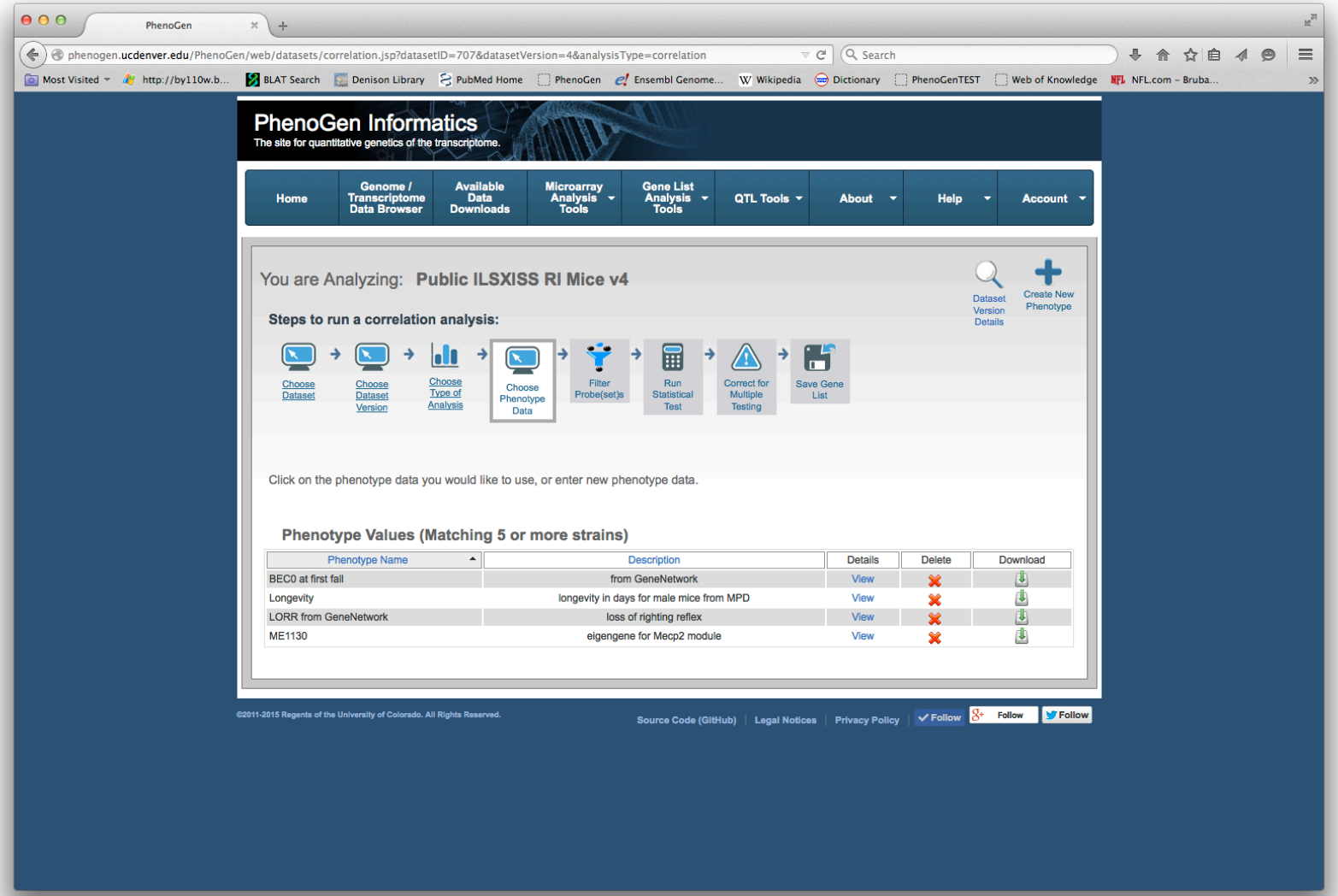
Image copied from "The Marriage of Phenomics and Genetical Genomics: A Systems Approach to Complex Trait Analysis" in Systems Biology in Psychiatric Research: From High-Throughput Data to Mathematical Modeling, edited by Tetter F, Winterer G, Gebicke-Haerter PG, and Mendoza E. Wiley-VCH 2010.

Data

- What you provide
  - Quantitative phenotype data from one of the following panels:
    - BXD RI mice
    - LXS RI mice
    - HXB/BXH RI rats
- What we provide
  - RNA expression levels
  - SNP information

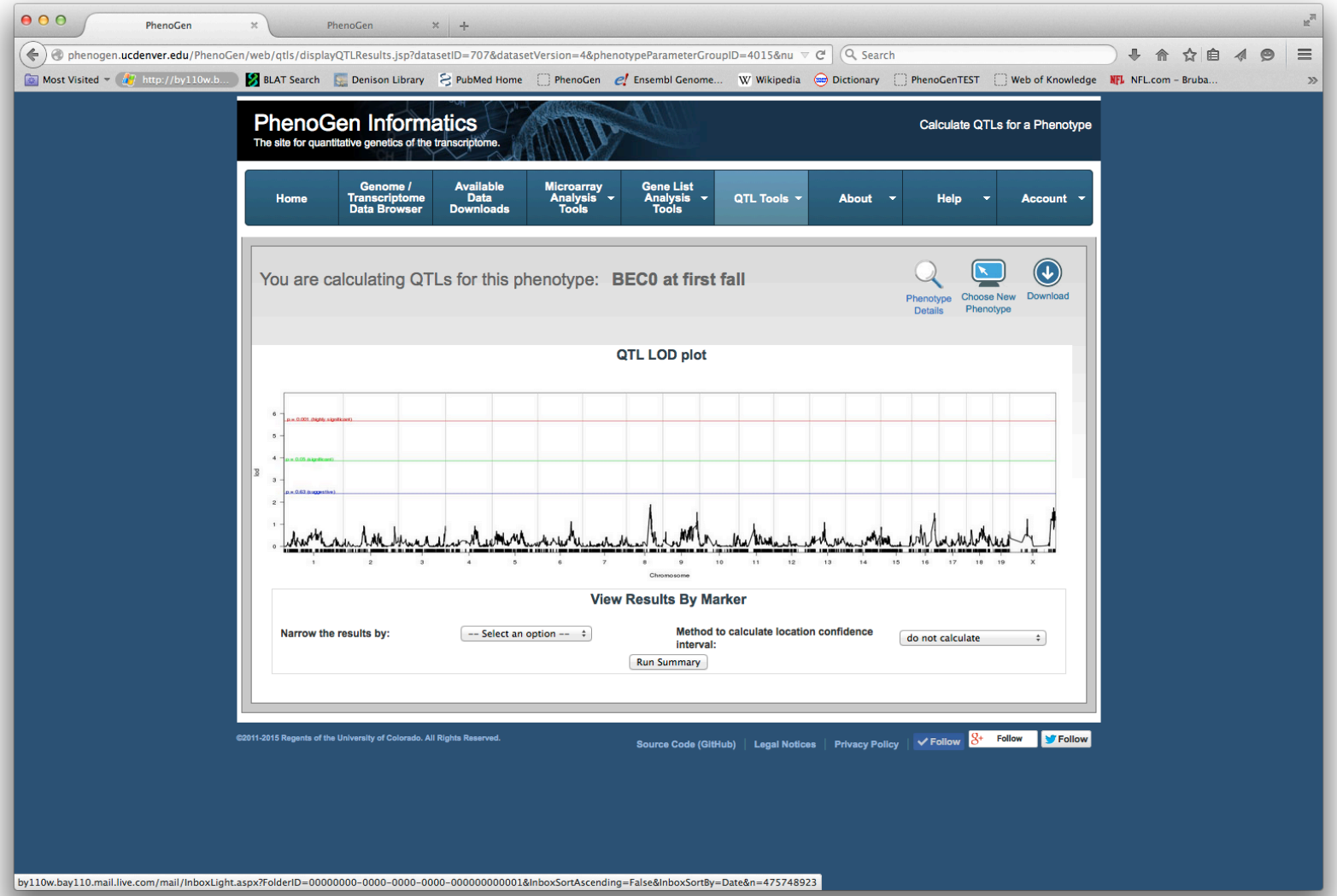
Genetic Correlation

RNA Expression Levels Correlated with Phenotype



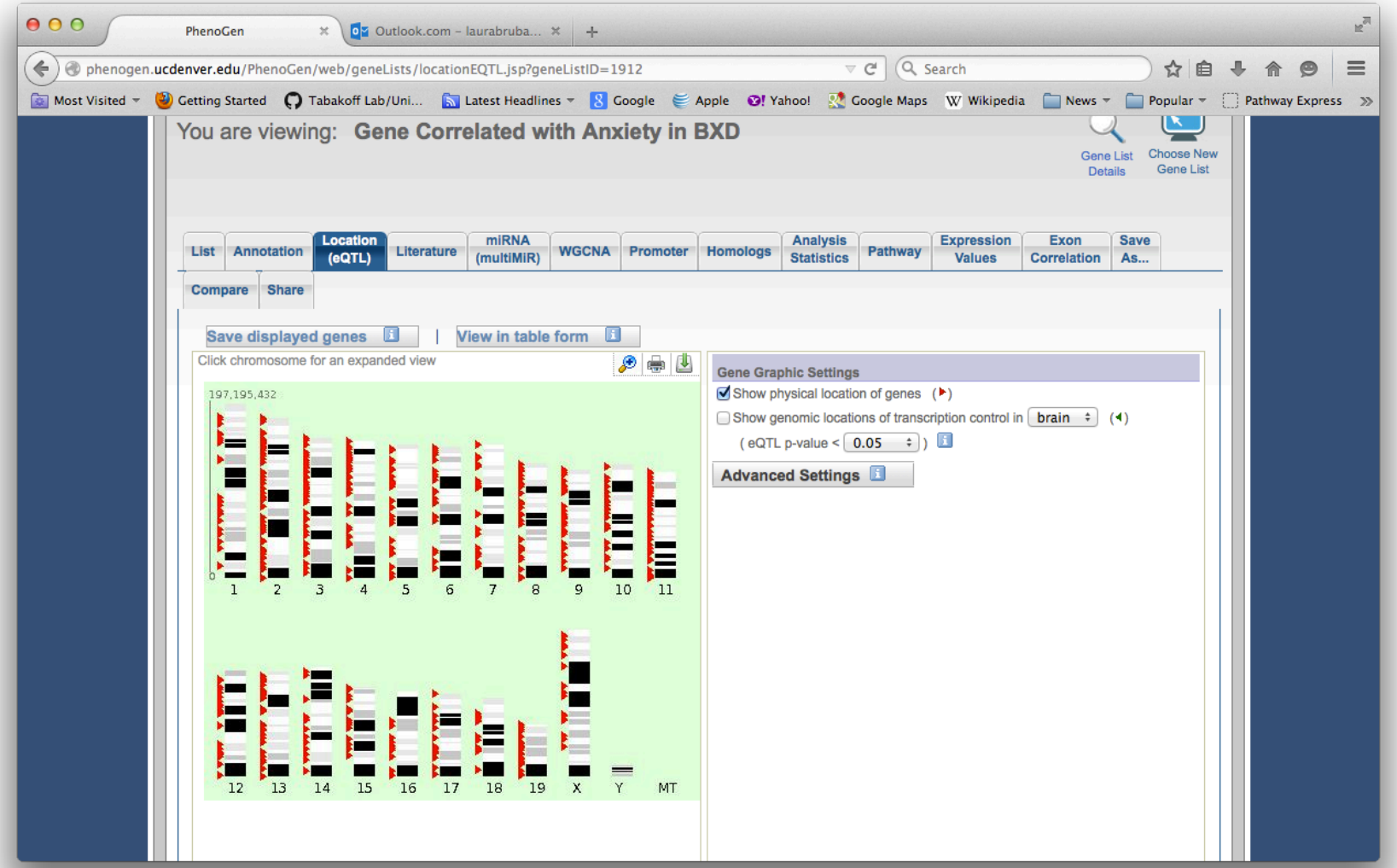
Phenotypic QTL

DNA Variants Associated with Phenotype



Expression QTL

DNA Variants Associated with RNA Expression Levels



DOWNLOADS

Downloadable Files

- Microarray Expression Data
  - Raw
  - Processed
  - eQTL/heritability
- Genotype Data
- RNA-Seq Data
  - Raw
  - Processed
- Code
  - Github

FUTURE DIRECTIONS

Future Data Plans

Strains	Tissue	Sex	Number of Biological Replicates Per Strain	Number of Paired-End Reads (rRNA-depleted total RNA)	Number of Single- End Reads (small RNA)
30 Classic Inbred Rat Strains	brain	male	4	18 trillion	5.5 billion
30 RI Rat Strains	brain	male	4	18 trillion	5.5 billion
30 Classic Inbred Rat Strains	liver	male	4	18 trillion	5.5 billion
30 RI Rat Strains	liver	male	4	18 trillion	5.5 billion

Functions to Add in Future

- multiMiR (miRNA/mRNA pairs) for rat
  - Currently implemented for mouse
- WGCNA for uploaded data sets
- Causal inference in co-expression modules
- Google drive integration