The Construction of a Tool that Predicts the Pathways from the Genome to Phenome

Mahaffey S.[1] , Saba L.[1], Hoffman P.[1], Pravenec M.[2], Vanderlinden L.[1] , Yu Y.[1] , Lusk R.[1], Tabakoff B.[1]
[1]Skaggs School of Pharmacy and Pharmaceutical Sciences and School of Medicine, University of Colorado Anschutz Medical Campus, Aurora CO [2] Czech Academy of Sciences, Prague, Czech Republic

PhenoGen Informatics
The site for quantitative genetics of the transcriptome.

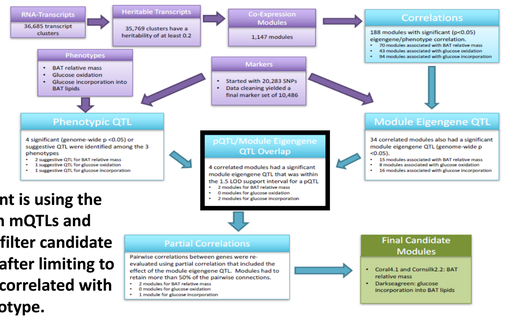
University of Colorado Anschutz Medical Campus

Abstract

The power of systems biology as a concept for decoding the relationship between genome and phenotype is becoming evident. We are continuing development of the Hybrid Rat Diversity Panel, an informative population in which to apply the systems biology approach. This panel will comprise 96 inbred rat strains, all of which will have deep sequencing of their genome, RNASeq, and exon array information on gene expression in four organs, and phenotype information. The analysis of these data allows for selection of proper markers for high resolution mapping of gene expression and additional traits to the genome. The gene expression phenotype is mapped as QTLs for modules derived through weighted gene co-expression network analysis (WGCNA), and represented by their eigengenes. Our approach requires that the eigengene values across strains correlate with the phenotype of interest and that the module eigengene QTL overlaps the QTL for the phenotype, i.e., gene expression within the module is controlled from the same area of the genome as the phenotype. Even though we are approximately half way through the transcriptome data collection, we have applied this approach to several complex traits to produce a solid proof of concept. We de novo uncovered a long non-coding RNA that is a hub gene for a module predisposing various levels of alcohol consumption; we found that liver, as well as brain, contributes to levels of alcohol consumption; we identified brown fat modules that contribute to "metabolic syndrome" phenotypes; we demonstrated that the approach identifies the expected enzymes responsible for alcohol metabolism, but also links the alcohol dehydrogenases to immune system function. Furthermore, we deleted (CRISPR/Cas9) the hub gene associated with alcohol drinking level, resulting in a change in phenotype and in expression of transcripts within and outside the module. Supported by NIAAA/NIH (R24AA013162) and the Banbury Fund.

Methods

Phenotype, Genotype, Transcriptome Analysis Flow Chart

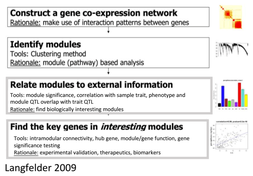


A key point is using the overlap in mQTLs and pQTLs to filter candidate modules after limiting to modules correlated with the phenotype.

Weighted Gene Coexpression Network Analysis (WGCNA)

Weighted gene co-expression network analysis (WGCNA) is a systems biology method for describing the correlation patterns among genes based on their expression levels.

Premise = Genes with similar expression patterns may form complexes, pathways, or participate in identifiable regulatory and signaling circuits (i.e., the gene products are functionally related).



Langfelder 2009

Rat Genotype-Tissue Expression (RGTEx)

With the HRDP we will have 96 sequenced strains with RNA-Seq expression of annotated and novel genes/isoforms in multiple tissues of males and females. With the sequencing / genotyping highly accurate eQTLs can be calculated.

Current RNA-Seq Data
Total RNA / Small RNA(<200bp)
- 40 strains in Brain and Liver
- BN-Lx/SHR in Heart, Female Brain
- BN-Lx/SHR Transcriptome Reconstruction Brain, Heart, and Liver
- 22 Billion paired-end reads (totalRNA)
- 5.7 billion reads(smallRNA)
- 4.2 TB of compressed raw RNA-Seq reads.

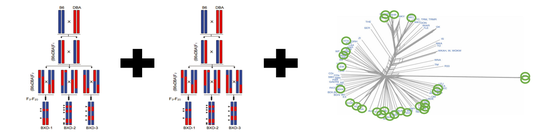
Future RNA-Seq Data
Total RNA / Small RNA(<200bp)
- 24 inbred strains in Whole Brain and Liver
- 32 LEXF/FXLE Recombinant Inbred Strains in Whole Brain and Liver
- RI Panel Transcriptome Reconstruction for Whole Brain and Liver

HRDP

Hybrid Rat Diversity Panel (HRDP)

A Renewable Genetically Defined Population Consisting of Recombinant Inbred and Classical Inbred Rats for Cumulative Biology ●96 Strains ●Genetic and Omic-Analysis ●Ascertainment of Heritability ●Susceptibility/predisposition studies ●Mechanistic Genome Biology Studies ●Systems/Network Biology Studies ●Toxicologic Analysis ●Pharmaco/toxicokinetics ●Proof of Concept Studies for Medication Development

Panel Built for Power



30 HXB/BXH Recombinant Inbred Strains | 32 LEXF/FXLE Recombinant Inbred Strains | 34 Classic Inbred Strains

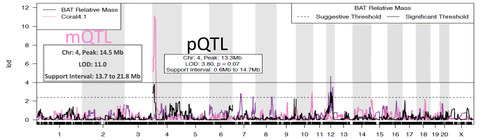
Power to detect loci that contribute 10-20% to genetic variance
Power to detect significant genetic correlations as small as 0.28
Power to perform high resolution mapping (median haplotype block size = 225 Kb)

Metabolic Syndrome Example

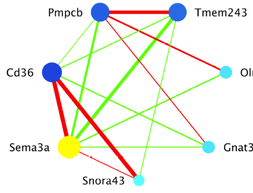
Pravenec M , et. al.,**Systems Genetic Analysis of Brown Adipose Tissue Function.** Nov 10, 2017, Physiol Genomics

Metabolic syndrome is a cluster of insulin resistance associated conditions – increased blood pressure, high blood sugar, excess body fat around the waist, and abnormal cholesterol or triglyceride levels – that occur together, increasing the risk of heart disease, stroke and diabetes (T2D). The phenotype used for this was the weight of interscapular brown fat per 100g body weight in rats (relative mass).

LOD plot showing the phenotype QTL and module(coral4.1) QTL overlap on chr 4

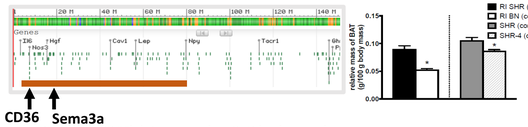


Transcripts Co-Expressed and Forming the Coral4.1 Module Associated with Relative Mass of Brown Adipose Tissue (BAT)



SEMA3a (semaphorin) – decrease noradrenergic innervation of BAT and reduce hypertrophy.
CD36 – transporter for long chain fatty acids and lipoproteins (fat sensor?).
PMPCb (mitochondrial peptidase) – activates SIRT3 which is a deacetylase-regulating mitochondrial function (increase ox-phos).
SNORA43 (nucleolar RNA) – function in expression of CD36
OLR371 (olfactory receptor) – first demonstration of olfactory receptor expression in BAT, but expressed in other organs, e.g., gut.
GNAT3 (α subunit gustducin) – G protein in taste transduction, coupled to PDE6 to keep basal [cGMP] low.
TMEM243 – mitochondrial protein of unknown function.

Using Animal Models for Confirmation: BN, SHR and SHR-4 Congenic Rats



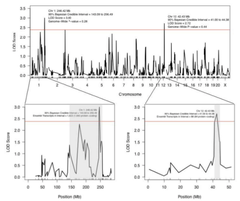
CD36 Sema3a
The BN Ch 4:5-80 Mb intergressed on SHR background

Alcohol Consumption Example

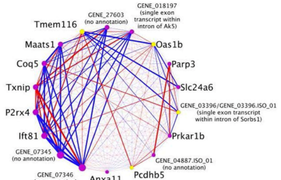
Saba L, et. al., **The sequenced rat brain transcriptome, its use in identifying networks predisposing alcohol consumption.** July 2015, FEBS

The HXB/BXH panel was used to evaluate genetic predisposition to alcohol consumption. The phenotype is voluntary alcohol consumption measured by providing 10% alcohol and water as described in Tabakoff et. al. 2009. Whole Brain WGCNA modules correlated with alcohol consumption where filtered by overlap with the alcohol consumption QTL.

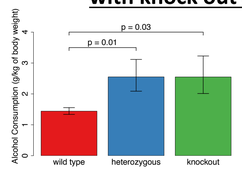
Phenotype QTL



Brain Module correlated with phenotype and with a module QTL overlapping the phenotype QTL



Animal Confirmation: Alcohol Consumption of male rats with knock out of the module hub gene

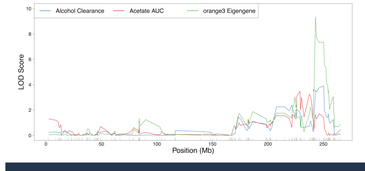


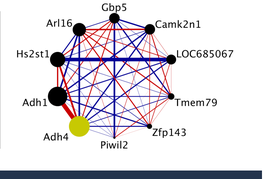
The possible ORF in the unannotated hub gene was deleted using CRISPR/Cas9. Alcohol consumption was measured in the knock outs. Males showed a significant difference(shown at left) in consumption with the knock out, while females showed no difference(not shown). P2rx4, which is one of the genes positively correlated with the hub gene, had decreased expression in the knock out. In addition many other genes also showed differential expression in the knock out versus the wild type.

Alcohol Metabolism Example

Lusk R, et. al., **Unsupervised, statistically-based systems biology approach for unraveling the genetics of complex traits: A demonstration with ethanol metabolism** Submitted, Genome Biology

The same approach was used with alcohol clearance and acetate under the curve phenotypes. The LOD plot below is just chr 2 and displays the overlap of the QTLs for both phenotypes and the candidate module in liver. The module, shown below right, contains 2 alcohol dehydrogenases one of which is the hub gene. These alcohol dehydrogenases are associated with alcohol clearance and are expected to be expressed in the liver.



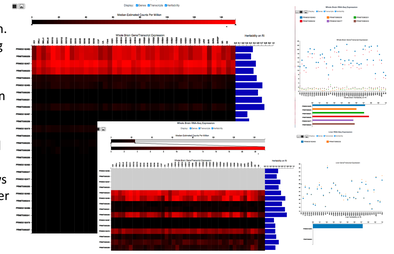


PhenoGen

https://phenogen.ucdenver.edu

PhenoGen's Genome/Transcriptome Data Browser provides an intuitive interface based on a genome browser but with access to various types of interactive data summaries from circos plots of eQTLs, to scatter plot/heat maps of expression to summaries of WGCNA modules to miRNA targeting summaries(multiMiR, Ru 2014).

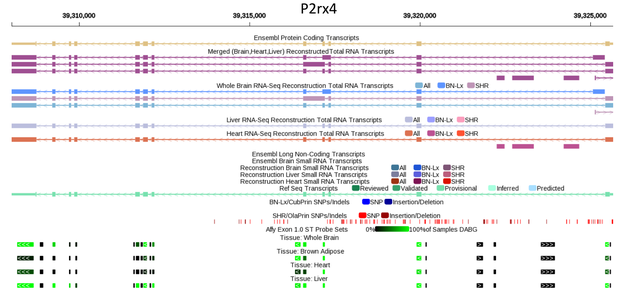
Expression / Heritability for genes and isoforms in the region. A heat map is displayed showing median estimated counts per million across all the HRDP strains. In a small region this can also be viewed as a scatter plot as in the gene view. Controls at the top allow viewing gene level or isoform level expression. A sliding scale on the legend allows viewing expression over a smaller range by graying out cells with expression outside the selected range.



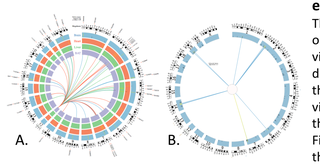
PhenoGen

https://phenogen.ucdenver.edu

Genome/Transcriptome Data Browser

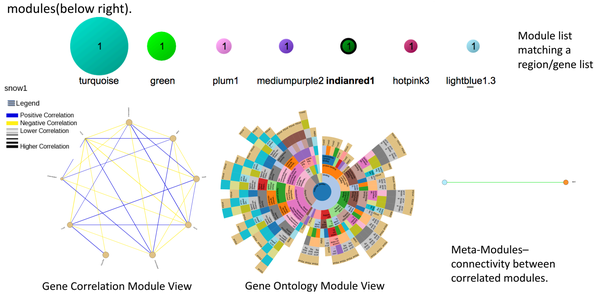


Example Transcriptome view of P2rx4 shows different isoforms in each tissue with strain specific isoforms if present. It also provides information on SNPs and small insertions or deletions. Other tracks can show splice junction read counts, read depth across the region, DNA Sequence, AA Sequence, bQTLs, UCSC Repeat Masker or you can import your own in bed, bedGraph, bigBed, or bigWig files. Hovering over a gene/transcript provides further detail including isoform quantitation from RNA-Seq data.



eQTLs
This tab/view summarizes all of the array based or RNA-Seq based eQTLs. Based on what is being viewed the eQTL ciros plot adjusts. For a region it displays genes at their location in the genome that have a significant eQTL in the region being viewed(Fig A). For a gene it displays the all of the locations of eQTLs above a set threshold. Finally for a module it displays eQTL locations for the module's eigengene(Fig B).

Weighted Gene Co-expression Network Analysis
WGCNA modules from microarrays and RNA-Seq. The tool will display all the modules relevant to the current context (region, gene list, or gene). Each module provides a view of the connectivity in the module(below left), Eigen gene eQTLs(above Fig B), gene ontology summaries(below middle), and miRNAs targeting of genes in the module(not shown). Modules are tissue specific and can be selected from the three tissues available whole brain, heart, and liver. For RNA-Seq Whole Brain and Liver are available and include a meta-module view which shows correlation between modules so you can browse related modules(below right).



Conclusions

The information provided on PhenoGen can be used without additional data, or with a phenotype assessment of a trait using the HRDP rats. The PhenoGen data reflects a basal state of the animal and should be considered to reflect a predisposition to a complex trait. The analysis of quantitative phenotypes by the researcher and the use of data provided by PhenoGen can be used 1) to explore the factors predisposing to disease or the normal variance in measured phenotypes, 2) predict differences in response to therapy, 3) Identify novel therapeutic targets. If you can model an addiction or any other phenotype in a rat you can deduce the predisposing genetic factors for that phenotype.

Support: R24AA013162 NIAAA.

References

Langfelder P et al. BMC Bioinformatics 2008
Lusk R et al. Genome Biol. Submitted
Pravenec M et al. Physiol Genomics 2017
Ru Y et al. NAR 2014
Saba L, et al. FEBS 2015
Tabakoff B et al. BMC Biol. 2009