19th Annual Meeting of the Complex Trait Community (CTC) and the Rat Genomics & Models Community
September 29th and 30th

University of Colorado Anschutz Medical Campus
Seminar Room (V20-1000), Pharmacy and Pharmaceutical Sciences Building
12850 E. Montview Blvd, Aurora, CO 80045

Thursday, September 29th

8:00 AM	***Registration open/Light Breakfast***	
9:00 AM	***Plenary Session #1***	
	Arjun Krishnan, PhD Associate Professor Department of Biomedical Informatics School of Medicine University of Colorado Anschutz Medical Campus	*Leveraging public data with machine learning to study complex diseases*
10:00 AM	***Discussion***	
10:30 AM	***Break***	
10:45 AM	***Session 1 - Submitted Talks***	
	10:45 AM (virtual)	**Michal Pravenec** Institute of Physiology, Czech Academy of Science *Spontaneous nonsense mutation in Tuft1 (tuftelin 1) gene is associated with amelioration of glucose and lipid metabolism in the BXH6 subline of rat recombinant inbred strains*
	11:00 AM (virtual)	**Aman Kumar** University of Tennessee Health Science Center *Identifying modulators of optic nerve numbers and health in glaucoma using systems genetics*
	11:15 AM (virtual)	**David Ashbrook** University of Tennessee Health Science Center *A novel pre-clinical model identifies genetic modifiers of triple negative breast cancer risk and progression*
	11:30 AM (virtual)	**Suheeta Roy** University of Tennessee Health Science Center *Replicated loci that modulate lifespan and longevity in females of the BXD family on Chr 1 and Chr 3*
11:45 AM	***Lunch - box lunches provided***	
12:45 PM	***Session 2 - Submitted Talks***	
	12:45 PM (in-person)	**Cody Cousineau** University of Michigan School of Public Health *Distinct genetic mechanism associated with phenotype response in chow-fed versus high fat diet-fed mice*
	1:00 PM (in-person)	**Jessica Zhou** University of California San Diego *Cocaine addiction is associated with long-term changes in gene regulation, metabolic pathways, and GABAergic inhibition within the amygdala*
	1:15 PM (in-person)	**Christopher King** University at Buffalo *Attribution of incentive salience to reward cues is genetically associated with other addiction-related traits in outbred heterogeneous stock rats*
	1:30 PM (in-person)	**Apurva Chitre** University of California San Diego *Genome-wide association study in a rat model of temperament identifies multiple loci for exploratory locomotion and anxiety-like traits*
	1:45 PM (in-person)	**Jennifer Jacob** Wayne State University *Identification of actionable target genes that regulate breast cancer progression by genetic linkage analysis in HER2/Neu Transgenic Diversity Outbred (DO) F1 mice*
2:00 PM	***Break***	

2:15 PM	***Session 3 - Submitted Talks***	
	2:15 PM (virtual)	**Michelle Perry** The Jackson Laboratory *Multiple genome viewer (MGV): a new tool for visualization and comparison of multiple annotated genomes*
	2:30 PM (virtual)	**Flavia Villani** University of Tennessee Health Science Center *Initial effort in generating a rat pangenome*
	2:45 PM (virtual)	**Gregory Keele** The Jackson Laboratory *Which mouse multiparental population is right for your study? The Collaborative Cross inbred strains, their F1 hybrids, or the Diversity Outbred population*
	3:00 PM (in-person)	**Denghui Chen** University of California San Diego *Hybrid genotyping pipeline on Lc-WGS and ddGBS sequences for heterogeneous stock rats*
	3:15 PM (in-person)	**Jennifer Smith** Medical College of Wisconsin *The Rat Genome Database resources for de novo rat strain assemblies*
3:30 PM	***Break***	
3:45 PM	***Session 4 - Submitted Talks***	
	3:45 PM (virtual)	**Gary Churchill** The Jackson Laboratory *Genetic analysis of multi-omic data identifies drivers of protein phosphorylation*
	4:00 PM (virtual)	**William Valdar** University of North Carolina at Chapel Hill *Increasing power in inbred strain association mapping by recognizing variance heterogeneity*
	4:15 PM (virtual)	**Dan Munro** University of California San Diego & Scripps Research *The regulatory landscape of multiple brain regions in outbred heterogeneous stock rats*
	4:30 PM (in-person)	**Cynthia Wu** University of California San Diego *A novel quantitative trait locus implicates Msh3 in the propensity for genome-wide short tandem repeat expansions in mice*
	4:45 PM (in-person)	**Abraham Palmer** University of California San Diego *Polygenic transcriptome risk scores can translate genetic results between species*
5:00 PM	***Poster Session***	
		Douglas Adams University of Colorado Anschutz Medical Campus *Genetic loci associated with skeletal response to PTH therapy*
		Mary Kaldunski Medical College of Wisconsin *PhenoMiner: RGD's quantitative phenotype data repository has data update and data mining tool improvements*
		Spencer Mahaffey University of Colorado Anschutz Medical Campus *PhenoGen: HRDP Transcriptome Data and API*
		Ryan Eveloff University of California San Diego *Characterization of polymorphic short tandem repeats and their association with phenotypes in HS rats*
		Panjun Kim University of Tennessee Health Science Center *Improving de novo genome assemblies of the HXB/BXH family of inbred strains using Hi-C data*
		Monika Tutaj Medical College of Wisconsin *Completion of genomic rat variants analysis of the Hybrid Rat Diversity Panel*
6:00 PM	***Optional Group Dinner***	TBD

Friday, September 30th

8:00 AM	***Registration open/Light Breakfast***	
9:00 AM	***Plenary Session #2***	
	L. Cinnamon Bidwell, PhD Assistant Professor Director of CU REACH Co-Director of CUChange Institute of Cognitive Science Department of Psychology and Neuroscience University of Colorado Boulder	*Developing a Model of Cannabis Psychopharmacology and Harm Reduction: Empirical Investigations Using Legal Market Products*
10:00 AM	***Break***	
10:15 AM	***Session 5 - Submitted Talks***	
	10:15 AM (virtual)	**Helene Tonnele** Centre for Genomic Regulation of Barcelona *Dissecting the mechanisms of indirect genetic effects in outbred mouse population*
	10:30 AM (virtual)	**Cara Trivett** University of Glasgow *Alternative splicing of osteopontin in CRISPR/Cas9 knock out rats*
	10:45 AM (virtual)	**Teresa McGee** University of North Carolina Chapel Hill *Incorporating strain-specific variance into genome-wide inbred strain mapping for zero-inflated phenotypes*
	11:00 AM (in-person)	**Luke Dillard** University of Virginia *Scalable in vitro model using scRNA-seq to generate cell-type specific transcriptome profiles of osteogenic cells*
	11:15 AM (in-person)	**Jack Pattee** University of Colorado Anschutz Medical Campus *Adjusting for cis-regulatory variation in gene expression data can improve biological network analysis in model organisms*
11:30 AM	***Lunch - box lunches provided***	
12:30 PM	***Session 6 - Submitted Talks***	
	12:30 PM (virtual)	**Emily Holt** University of Vermont *Genetic analysis of CNS autoimmunity using the diversity of the Collaborative Cross reveals unique phenotypes and mechanisms*
	12:45 PM (virtual)	OPEN
	1:00 PM (virtual)	**Karthickeyan Chella Krishnan** University of Cincinnati *Genetic regulation of heart mitochondrial proteome influencing heart function*
	1:15 PM (virtual)	**Ellen Risemberg** University of North Carolina Chapel Hill *Gene-by-treatment QTL mapping to dissect genetic susceptibility to severe coronavirus disease*
1:30 PM	***Break***	
1:45 PM	***Session 7 - Submitted Talks***	
	1:45 PM (virtual)	**Montana Lara** University of Vermont *Network-based analysis predicts interacting genetic modifiers from a mapping study of multiple absence epilepsy mouse models*
	2:00 PM (virtual)	**Mary Seramur** Wake Forest University *GRK5 is required for adipogenesis through activation of ERK pathway*
	2:15 PM (virtual)	**Mackenzie Roberts** Wake Forest University *Adcy3 knock-out rats show increased adiposity without changes in behavioral phenotypes*
	2:30 PM (in-person)	**Carl Litif** University of Wyoming *Classifying the discrete epigenetic environment involved in drug-seeking behavior*

2:45 PM	***Break***	
3:00 PM	***Session 8 - Submitted Talks***	
	3:00 PM (virtual)	**Callan O'Connor** The Jackson Laboratory *A genetic basis for oxidative stress response variation*
	3:15 PM (virtual)	**Valerie Wagner** Medical College of Wisconsin *Genetic background in the rat impacts metabolic outcomes of post-wean BPF exposure*
	3:30 PM (virtual)	**Joel Leal-Gutierrez** University of California San Diego *Structural equation modeling uncovers candidate genes for cocaine's conditioned effects in heterogeneous stock rats*
	3:45 PM (in-person)	**Brittany Leger** University of California San Diego *Genome-wide association studies of human and rat body mass index converge on a conserved molecular network*
	4:00 PM (in-person)	**Oksana Polesskaya** University of California San Diego *Genome-wide association study finds multiple loci associated with intraocular pressure in HS rats*
4:15 PM	***Closing Remarks***	

Spontaneous nonsense mutation in *Tuft1* (tuftelin 1) gene is associated with amelioration of glucose and lipid metabolism in the BXH6 subline of rat recombinant inbred strains

[1]Šilhavý, J., [1]Mlejnek, P., [1,2]Liška, F., [1]Šimáková, M., [3]Marková, I., [3]Hüttl, M., [3]Malínská, H., [4]Stránecký, V., [4]Kmoch, S. & [1,2]*Pravenec, M.

*Michal.Pravenec@fgu.cas.cz

[1]Institute of Physiology, Czech Academy of Sciences, Prague, Czech Republic
[2]Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University, General University Hospital, Prague, Czech Republic
[3]Institute for Clinical and Experimental Medicine, Prague, Czech Republic
[4]Department of Pediatrics and Inherited Metabolic Disorders, First Faculty of Medicine, Charles University, Prague, Czech Republic

Recently, we detected a recessive mutation, an abnormal coat appearance, in the BXH6 strain, a member of the BXH/HXB set of recombinant inbred strains derived from SHR (spontaneously hypertensive rat) and BN-*Lx* (Brown Norway) progenitors. Whole genome sequencing of mutant rats identified 195875980 G/A mutation in *Tuft1* (tuftelin 1) gene on chromosome 2 resulting in premature stop codon. Mutant rats exhibited significantly increased *Tuft1* mRNA expression in all tissues examined which is likely secondary to defective message and absence of the TUFT1 protein. Mutant BXH6^{Tuft1} versus BXH6 wild type rats showed lower body weight due to reduced visceral fat and ectopic fat accumulation in liver and heart, decreased serum glucose and insulin, and increased insulin stimulated glycogenesis in skeletal muscle. Thus absence of TUFT1 protein is associated with reduced adiposity and increased sensitivity of muscle tissue to insulin action.

TUFT1, a phosphorylated glycoprotein, was originally reported to play a role in dental enamel mineralization. Recently, TUFT1 expression was found to be elevated in several types of cancers. In addition, C57BL/6NTac-*Tuft1*$^{tm1a(KOMP)Wtsi}$ mice with targeted *Tuft1* gene showed improved glucose tolerance, lower circulating HDL cholesterol, and decreased circulating alkaline phophatase (https://www.mousephenotype.org/data/genes/MGI:109572). The latter results are congruent with our findings and together provide evidence for important role of *Tuft1* gene in regulation of glucose and lipid metabolism.

Identifying Modulators of Optic Nerve Numbers and Health in Glaucoma Using Systems Genetics

[1]Edwards, W.S., [1]*Kumar, A., [2]Ashbrook, D.G., [2]Stiemke, A.B., [2]Lu, L., [2]Williams, R.W., [1]Jablonski, M.M.

*akumar28@uthsc.edu

[1]Hamilton Eye Institute, Department of Ophthalmology, University of Tennessee Health Science Center, Memphis, TN 38163

[2]Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, Memphis, TN 38163

Glaucoma is a multifactorial, neurodegenerative disease characterized by the progressive loss of retinal ganglion cells (RGCs), optic nerve (ON) damage, and subsequent vision loss. Current treatment regimens focus on lowering intraocular pressure (IOP); however, it is known that some patients continue to experience RGC death despite adequate IOP reduction. It is crucial to identify alternative therapeutics to halt the progression of RGC death for more effective glaucoma prevention and treatment. In this study, we identify a quantitative trait locus (QTL) that modulates the number of intact and seemingly healthy axons in PPD-stained optic nerves from a large family of recombinant inbred strains of mice. We also define a set of high priority positional candidate genes, a subset of which are likely to modulate ON resilience and health.

A large cohort of the BXD family was aged to greater than 13 months-of-age. ONs from 75 strains and the DBA/2J (D2) parent were harvested, sectioned, and stained with p-phenylenediamine. Numbers of intact axons per ON cross-section were counted from 1–10 cases per strain (GeneNetwork trait BXD_18613).

Numbers of intact axons per nerve ranged from 30,000 to nearly 62,000. Linear mixed model mapping defines a locus on chromosome 9 between 94.6 and 105.7 Mb with a –logP linkage of 4.27. Of 424 positional candidates, three genes— *nephronophthisis 3* (*Nphp3*); *armadillo repeat containing 8* (*Armc8*); and *zinc finger and BTB domain containing 38* (*Zbtb38*) passed stringent criteria and are highlighted here as high priority candidates for follow-up analyses. Overall, these genes were significantly correlated with a higher number of intact axons per ON and a protective *D* allele effect.

A novel pre-clinical model identifies genetic modifiers of triple negative breast cancer risk and progression

[1]Laura M. Sipe, [1]Jeremiah R. Holt, Robert W. Read, D. [1,2]Neil Hayes, [3]Lu Lu, [2,3]Robert W. Williams, [2,3]*David G. Ashbrook, [1,2]Liza Makowski

*dashbroo@UTHSC.edu
[1]Division of Hematology and Oncology, Department of Medicine
[2]UTHSC Center for Cancer Research
[3]Department of Genetics, Genomics, Informatics, College of Medicine, The University of Tennessee Health Science Center; Memphis, TN 38163, USA

Triple negative breast cancer (TNBC) is an aggressive breast cancer subtype with poor outcomes. There is a grave clinical challenge to successfully treat the ~30,000 patients diagnosed with this disease every year. It is clear that risk and disease progression vary both between individuals and between populations. Discovering genetic modifiers of differential TNBC vulnerability and disease progression is critical to improving predictive and personalized treatments. In humans, it is often difficult to collect data on disease progression on a large scale, whereas in mice progression is carefully monitored. Disease models are necessary that reflect the genetic heterogeneity of the population, and therefore the heterogeneity of disease and of treatment response.

We hypothesized that different variants modulate the risk, aggression, and progression of TNBC. The C3(1)-T antigen (C3Tag; 013591) genetically engineered mouse model in the FVB/N strain recapitulates human basal-like TNBC. However, this model is highly constrained by its inbred genetic background, where all aspects of the disease (e.g. initiation, progression, and response to therapy) are confounded. By adding genetic diversity to the model, we are able to deconvolute these different aspects of disease, and identify quantitative trait loci (QTL) that modulate them.

To add genetic diversity to the model, we crossed the C3Tag to the largest and best characterized mammalian genetic reference population, the BXD family, to create a BXD breast cancer model (BXD-BC). In this initial study of 28 strains, we demonstrate high heritability of phenotypes including latency (onset when a tumor was first identified), progression (how long between latency and endpoint), and multiplicity (number of tumors developed). Whereas 100% of hemizygous C3Tag females on the FVB/N background develop mammary adenocarcinoma within 6 months, we identified two BXD-BC strains that did not develop tumors even after 12 months of age.

Quantitative trait loci (QTL) analysis identified a genomic region modulating onset, and three, potentially interacting, regions modulating multiplicity. RNA-seq of the tumors, combined with systems genetics analysis, identified high-priority candidates within all of these regions. Cross-species comparison of our findings with publicly available human GWAS and genomic databases is an effective approach to validate conserved biologically relevant and targetable pathways.

To our knowledge, this is the first study to explore modifier genes for TNBC phenotypes using a systems genetics approach in a mouse model of TNBC. Our results will contribute to understanding the genetics of risk and disease progression, and provide a novel model in which to test treatments for triple negative breast cancer.

Replicated loci that modulate lifespan and longevity in females of the BXD family on Chr 1 and Chr 3

Suheeta Roy[1*], David G. Ashbrook[1], Danny Arends[5], Jesse F. Ingels[1], Casey J. Chapman[1], Melinda S. McCarty[1], Arthur G. Centeno[1], Pjotr Prins[1], Saunak Sen[2], Collin Ewald[3], Johan Auwerx[4], Lu Lu[1], Robert W. Williams[1], Evan G. Williams[6]

* sroy12@uthsc.edu (presenting)
[1]Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, Memphis, TN, USA
[2]Department of Preventive Medicine, University of Tennessee Health Science Center; Memphis, TN, USA
[3]Department of Health Sciences and Technology, Institute of Translational Medicine Eidgenössische Technische Hochschule (ETH) Zürich, CH-8603 Zürich, Switzerland
[4]Laboratory of Integrative Systems Physiology, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne CH-1015, Switzerland
[5]Department of Applied Sciences, Northumbria University, Newcastle upon Tyne, UK
[6]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

Background: We measured lifespan as a function of strain and diet across cohorts of 82 BXD strains (n = 1665 females) in a well-controlled environment. Details on the design are presented in Roy et al. 2021 and Williams et al. 2022 with a focus on understanding predictors of lifespan and body mass as a function of diet. Here we focus on loci that modulate lifespan as a function of two diets using new mapping methods and genotypes.
Methods: We measured lifespan of ~10 replicates per strains on a standard chow diet (CD, 6% calories from fat) or a high fat diet (HFD, 60% calories from fat). Means were mapped using ~20,000 sequence-based markers and linear mixed model methods (GEMMA) that account for family substructure and cofactors (FAIR data @GeneNetwork, BXD_18441 and BXD_18435). We mapped ascending thresholds at ~200, 600, 700, and 750 days for CD; and ~200, 515, 603 days, and 660 days for HFD.
Results: The high fat diet shortens mean lifespan by ~80 days (HFD 612 ± 13 d, n = 809; CD 692 ± 13, *n* = 856), equivalent to ~7-year loss in humans. We consistently detect a highly significant locus on Chr 1 (–logP 5.1 at ~22–25 Mb, +50 d per *D* allele) on the HFD. This locus is also detected on CD (–logP 2.0). Likewise, mapping individual BXD data controlling for diet yielded a single joint locus here (–logP 3.85 at ~28 Mb). This locus—*Vita1a*—has been replicated in the NIA ITP cohort of 6200 UM-HET3 mice (Arends D et al., in progress) although the polarity of effect is reversed. Finally, we also replicate the *Vita3a* locus detected initially in UM-HET3 females (Bou Sleiman M, Roy S, et al., *Science*, in press). In BXD females this locus has a –logP of 2.5 on Chr 3 at 98 Mb, and an additive effect of +34 d per *D* allele.

Roy, Suheeta et al. 2021. "Gene-by-Environment Modulation of Lifespan and Weight Gain in the Murine BXD Family." *Nature Metabolism* 3 (9): 1217–27.
Williams, Evan G et al. 2022. "Multiomic Profiling of the Liver across Diets and Age in a Diverse Mouse Population." *Cell Systems* 13 (1): 43-57. e6.

Distinct Genetic Mechanisms Associated With Phenotype Responses in Chow-Fed Versus High Fat Diet-Fed Mice

[1]*Cousineau, C.M. & [1]Bridges, D.

*ccousz@umich.edu

[1]Department of Nutritional Sciences, University of Michigan School of Public Health

Interactions between genetics and diet greatly impact complex traits such as cholesterol in mice and humans. Using public data from diversity outbred (DO) mice, we performed a Genome Wide Association Study and identified three regions associated with cholesterol levels in chow-fed mice. Closer evaluation of the individual loci indicated inconsistent associations with cholesterol on a high fat, high sucrose diet (HFHS). Using these data we generated a polygenic risk model. This showed a strong predictive relationship between the score and cholesterol levels in chow-fed mice ($r^2 = 0.429$). Using the same score, the relationship was much weaker in predicting cholesterol levels in HFHS-fed mice ($r^2 = 0.061$). To nominate causal pathways, liver transcriptomic data from these mice was used, identifying 213 genes with significant associations with cholesterol on chow, but none of these genes showed significant associations in the same direction on HFHS diets. Independently, in BXD mice, we confirm a lack of concordance between cholesterol levels on chow and HFHS diets when comparing genetically similar inbred mice. To identify additional *a priori* predictors of cholesterol, we utilized a machine learning approach to test how 162 phenotypes predicted cholesterol levels in DO mice. A pruned regression tree was generated with four of the phenotypes as classifiers. Aside from HFHS diet, calcium was the strongest predictor for cholesterol in both chow ($r^2 = 0.325$, $p = <0.001$) and HFHS diet fed-mice ($r^2 = 0.278$, $p = <0.001$). A combined model comprising of sex, diet, and calcium levels predicted cholesterol levels with 45% accuracy. Together, these data indicate the predominance of distinct diet-dependent genetic mechanisms controlling cholesterol levels. This work also highlights the importance of evaluating gene-diet interactions when considering the genetic risk of hypercholesterolemia.

Cocaine addiction is associated with long-term changes in gene regulation, metabolic pathways, and GABAergic inhibition within the amygdala

[1,2*]Zhou, J.L., [3]Guglielmo, G.d., [1]Ho, A.J., [3]Kallupi, M., [4]Li, H., 3Chitre, A.S., [3]Carrette, L.L., [5]George, O., [3,6]Palmer, A.A., [1]McVicker, G., & [4]Telese, F.

*jlz014@eng.ucsd.edu

[1]Integrative Biology Laboratory, Salk Institute for Biological Studies
[2]Bioinformatics and Systems Biology Program, University of California San Diego
[3]Department of Psychiatry, University of California San Diego
[4]Department of Medicine, University of California San Diego
[5]The Scripps Research Institute
[6]Institute for Genomic Medicine, University of California San Diego

The amygdala contributes to negative emotional states associated with relapse to drug seeking, but the cell type-specific gene regulatory programs that are involved in addiction are unknown. Here we generate an atlas of single nucleus gene expression and chromatin accessibility in the amygdala of rats with low and high cocaine addiction-like behaviors following a prolonged period of abstinence. Between rats with different addiction indexes, there are thousands of cell type-specific differentially expressed genes which are enriched for molecular pathways including GABAergic synapses in excitatory and somatostatin neurons. Here we show that higher addiction severity is linked to excessive GABAergic inhibition in the amygdala, and electrophysiological and behavioral addiction-related phenotypes are reversed by the pharmacological inhibition of the metabolic enzyme glyoxalase 1. By analyzing chromatin accessibility, we identify thousands of cell type-specific chromatin sites enriched in transcription factor (TF) motifs associated with addiction-like behavior, most notably motifs for pioneer TFs in the Fox, Sox, and helix-loop-helix families. Overall, we provide a comprehensive characterization of cell type-specific regulatory mechanisms associated with addictive behaviors.

Attribution of Incentive Salience to Reward Cues is Genetically Associated with Other Addiction-Related Traits in Outbred Heterogeneous Stock Rats.

[1,2,*]King, C.P., [3]Chitre, A.S., [3]Polesskaya, O., [4,5]Flagel, S.B., [6]Robinson, T.E., [2]Ishiwari, K., [7]Dietz, D.M., [8]Solberg Woods, L.C., [9]Chen, H., [10,11]Palmer, A.A., [1]Meyer, P.J.

[*]cpking@buffalo.edu

[1]Dept. of Psychology, [2]Clinical and Research Inst. on Addictions, [7]Dept. of Pharmacology and Toxicology, Univ. at Buffalo
[3]Dept. of Psychiatry, Univ. of California San Diego
[4]Dept. of Psychiatry, [5]Michigan Neuroscience Inst., [6]Dept. of Psychology, Univ. of Michigan
[8]Dept. of Internal Medicine, Molecular Medicine, Center on Diabetes, Obesity and Metabolism, Wake Forest School of Medicine, Winston-Salem
[9]Dept. of Pharmacology, Addiction Science and Toxicology, Univ. of Tennessee Health Science Center
[10]Dept. of Psychiatry, [11]Inst. for Genomic Medicine, Univ. of California San Diego

Individuals show variability across many addiction-related traits, such as tendency to attribute incentive salience to reward cues. These individual differences are frequently associated with features of drug-taking and drug-response. For example, we have shown that heightened incentive salience to reward cues is associated with other traits including action impulsivity (King et al. 2016) and reinstatement to nicotine seeking (Versaggi et al. 2016).

To examine the genetic basis of these relationships, we first conducted a genome-wide association study (GWAS) for the attribution of incentive salience using the Pavlovian Conditioned Approach task in a large cohort of Heterogeneous Stock (HS) rats (n = 1,645). Rats displayed two types of conditioned responses: cue-directed "sign-tracking" as a measure of incentive salience attribution, and food-cup directed "goal-tracking". We found that various measures of sign-tracking were moderately heritable (h^2 = .189-.215) and were associated with 18 unique quantitative trait loci (QTL). Interval sizes of these QTLs varied widely, many were located and chromosome 1 and contained few genes (*e.g. Tenm4*, *Mir708*). To help identify behaviorally relevant candidate genes, we further identified genes with moderately impacted coding variants (*e.g. Usp35, Alg8*) or contained genes from expression-QTL in mesocorticolimbic regions of the central nervous system (*e.g. Wnt11, Capn5).*

To compare sign-tracking with the genetic basis of other addiction related traits, such as reinstatement to nicotine seeking and impulsivity, we performed a phenome-wide association study (PheWAS) using results from other GWAS conducted using separate cohorts of HS rats across three centers. We have tentatively identified genes associated with reinstatement to nicotine self-administration (*Tenm4, Usp35, Alg8)* that overlap with tendency to sign-track, reflecting overlapping regions of chromosome 1. Ongoing work includes examination of other traits that are genetically correlated with measures of sign-tracking, such as delayed discounting (Rg = 0.689) and social reinforcement (Rg = 0.637). We show the utility for HS rats in identifying genetic variants across different complex behaviors.

Genome-Wide Association Study in a Rat Model of Temperament Identifies Multiple Loci for Exploratory Locomotion and Anxiety-Like Traits

Apurva S. Chitre[1]*, Elaine K. Hebda-Bauer[2]*, Peter Blandino[2], Hannah Bimschleger[1], Khai-Minh Nguyen[1], Pamela Maras[2], Fei Li[2], A. Bilge Ozel[3], Oksana Polysskaya[1], Riyan Cheng[1], Shelly B. Flagel[2], Stanley J. Watson, Jr.[2], Jun Li[3], Huda Akil[2]**,Abraham A Palmer[1,4]**#

*Equal first authors, **Equal last authors, # Corresponding Author

achitre@ucsd.edu

1 Department of Psychiatry, University of California San Diego

2 Michigan Neuroscience Institute, University of Michigan

3 Department of Human Genetics, University of Michigan

4 Institute for Genomic Medicine, University of California San Diego

Common genetic factors likely contribute to multiple psychiatric diseases including mood and substance use disorders. Certain stable, heritable traits reflecting temperament, termed externalizing or internalizing, play a large role in modulating vulnerability to these disorders. To model these heritable tendencies, we selectively bred rats for high and low exploration in a novel environment (bred High Responders (bHR) vs. Low Responders (bLR)). To identify genes underlying the response to selection, we phenotyped and genotyped 558 rats from an F_2 cross between bHR and bLR. Several behavioral traits show high heritability, including the selection trait: exploratory locomotion (EL) in a novel environment.

We identified multiple significant loci for six behavioral traits. Five of the six traits reflect different facets of EL that were captured by three behavioral tests. Distance traveled measures from the open field and the elevated plus maze map onto different loci, thus may represent different aspects of novelty-induced locomotor activity. The sixth behavioral trait, number of fecal boli, is the only anxiety-related trait mapping to a significant locus on chromosome 18. The identification of a locomotor activity-related QTL on chromosome 7 encompassing the *Pkhd1l1* and *Trhr* genes is consistent with our previous finding of these genes being differentially expressed in the hippocampus of bHR vs. bLR rats.The strong heritability coupled with identification of several loci associated with EL and emotionality provide compelling support for this selectively bred rat model in discovering relatively large effect causal variants tied to elements of internalizing and externalizing behaviors inherent to psychiatric and substance use disorders.

Identification of Actionable Target Genes that Regulate Breast Cancer Progression by Genetic Linkage Analysis in HER2/Neu Transgenic Diversity Outbred (DO) F1 Mice

[1*]Jacob, J.B., [1]Wei, K-C., [1]Bepler, G., [1]Reyes, J.D., [1]Polin, L., [1]White, K., [2]Yin, C., [2]Mi, Q-S., [1]Wei, W-Z.

*jacobj@wayne.edu
[1]Department of Oncology, Karmanos Cancer Institute, Wayne State University, Detroit, MI
[2]Department of Immunology, Henry Ford Health System, Detroit, MI

Our ultimate goal is to prevent breast cancer through the identification and targeting of cancer-associated molecules. We have thus established a functional genomics platform utilizing Diversity Outbred (DO) F1 mice that express an oncogene in order to map candidate genes of breast cancer. In the current study, BALB NeuT females expressing the HER2/Neu oncogene were crossed with DO mice to generate (BALBxDO)F1 NeuT mice. By associating the age of spontaneous mammary tumor onset and tumor growth rates with the GigaMUGA genetic marker profile, we identified 3 QTL in mouse Chr 1, X and 10. In addition, PWK/PhJ (PWK) and CAST/EiJ (CAST) were identified as the driving haplotypes for more aggressive tumors.

Candidate gene(s) need to be relevant to humans as well as targetable. We first determined which genes were found within the QTL confidence intervals using the UCSC Genome Browser, then identified genes which harbor SNP unique to the driver strain(s) with the Sanger Mouse Database (*no longer available to the public*). From this query, 173 genes were located within the 3 QTL and 34 harbored consequential SNP within coding region(s) unique to PWK and CAST haplotypes. Twenty-six (26) of those have human homologues and were further pursued. Relevance to breast cancer was determined by querying human databases and 21 of the 26 genes were significantly associated with survival outcomes. To prioritize the 21 genes, we determined whether candidate genes impacted shared pathways using Ingenuity Pathway Analysis. A set of 8 genes emerged, mainly affecting cytokine and chemokine production. Of particular interest was the transcription factor TSC22D3 in Chr X, which regulates LILRB4 in mouse Chr 10 (human Chr 19). This pair of genes were elevated for further analysis.

Cell type expression of candidate genes was assessed using single cell RNA sequencing and flow cytometry in normal and tumorous mouse mammary tissues. While TSC22D3 is expressed ubiquitously, LILRB4 is expressed primarily on the membrane of tissue infiltrating macrophages. LILRB4 has been described in both humans and mice as a myeloid cell checkpoint molecule that negatively regulates tumor immunity and is a prime clinical candidate. To initiate analysis of LILRB4 activity, (BALBxPWK)F1 NeuT mice were generated and electrovaccinated with DNA encoding a truncated Neu protein. Significantly enhanced HER2/neu Ab response was induced in the F1 mice compared with BALB NeuT parental mice (304±38 vs. 147±19, $p < 0.01$) and tumor growth inhibition was observed in both strains despite earlier tumor onset in (BALBxPWK)F1 mice. These findings indicate a more vigorous immune response in mice of PWK background, neutralizing the aggressive tumor growth. Further studies are ongoing to determine the role of LILRB4 in tumor immunity and the efficacy of modulating this molecule.

This work demonstrates the power of the diversity background coupled with a functional genomics approach as a means of identifying humanly-relevant, targetable candidate genes. Supported by CA76340 (WZW), Herrick Foundation (WZW), and CCSG P30 CA022453 (GB).

Multiple genome viewer (MGV): a new tool for visualization and comparison of multiple annotated genomes

[1]*Perry, M.N., [1]Smith, C.L., [1]Richardson, J.E., [1]Baldarelli, R.M., & [1]Bult, C.J.

*michelle.perry@jax.org

[1]The Jackson Laboratory

The multiple genome viewer (MGV) is a powerful tool to aid in the exploration of genome features between mouse strains and different species, evaluation of missing annotations between mouse strains, and detection of genome issues with genome assemblies. MGV enables visualization, exploration, and comparison of annotated genomes from sixteen inbred and two wild-derived mouse strains. Horizontal genome tracts allow users to explored annotation similarities and differences across the genomes of multiple mouse strains and compare the organization of homologous genome features from different organisms (human, fruit fly, zebrafish, nematode, rat, and yeast. Basic navigation features make it easy to view rearrangements and inversions, and to zoom out or in, or down to the level of individual base pairs. Users can search with genome coordinates or generate custom sets of genes annotated to phenotype, disease, or functional terms. For example, restricting the comparison to genome features associated with atherosclerosis identifies 25 loci that can be compared between mouse strains. The MGV offers the ability to download a FASTA formatted file for an entire genomic region, or for select genomic, transcript or coding sequences. For the latter, amino acid translation can be provided.

Future improvements will allow users to upload their own annotated genomes and utilize the paralog assertions and gene list functionality in line with those available through the Alliance of Genome Resources.

Initial effort in generating a rat pangenome

*Flavia Villani[1], Andrea Guarracino[2], Jun Huang[3], David Ashbrook[1], Robert W. Williams[1], Vincenza Colonna[1,4], Hao Chen[3], and Erik Garrison[1]

*fvillan1@uthsc.edu

[1] *Department of Genetics, Genomics, and Informatics, University of Tennessee Health Science Center, Memphis, TN, United States*
[2] *Genomics Research Centre, Human Technopole, Viale Rita Levi-Montalcini 1, Milan, 20157, Italy*
[3] *Department of Pharmacology, Addiction Science and Toxicology, University of Tennessee Health Science Center, Memphis, TN, United States.*
[4] *Institute of Genetics and Biophysics "Adriano Buzzati-Traverso", National Research Council, Naples, Italy.*

A pangenome can contain the full genomic information of a species. Pangenome graphs provide a compact representation of the mutual alignment of collections of genomes. In these graphs, nodes represent sequences in the pangenome, and paths describe genomes as walks through the graph. The members of the HXB/BXH family of recombinant inbred strains of rats have been used in many genetic mapping studies of physiological and behavior traits. It is part of the hybrid rat diversity panel, which is used in several ongoing large-scale genetic mapping studies on substance abuse related traits and phenome-wide association analysis. Current genomic studies using this family assume a single linear reference genome, making it difficult to observe sequences diverging from the reference, therefore limiting the accuracy and completeness of analyses. We sequenced all 30 members of the HXB family using linked-read libraries and built a pangenome graph with the PanGenome Graph Builder (PGGB) (*Garrison et al., 2021*) to study genetic variation. The pangenome enhanced the discovery of complex variants not seen by traditional genomics methods and provided calls with good precision and sensitivity. The ratio of the number of transitions to the number of transversions (Ts/Tv) in the pangenomic calls is 2.2, which is slightly higher than figures from LongRanger (2.1), this potentially reflects a slight enrichment of complex variants in the pangenomic set. In masked regions *(Gonzalez et al., 2021)* (i.e. regions that do not contain SINE, ALUs, LINE, LTR, and other simplex and complex DNA repeats), precision and sensitivity of vg (Hickey et al., 2020) calls are 90% and 84%, respectively, and for SNPs these figures rise to 94% and 80%. Overall we were able to reproduce data on known genetic variants and capture novel variations, which are being validated using Sanger sequencing. In summary, we demonstrate that pangenomes can be accurately built from linked-reads and that the pangenome produced by short linked reads can be informative.

Which mouse multiparental population is right for your study? The Collaborative Cross inbred strains, their F1 hybrids, or the Diversity Outbred population

[1*]Keele, Gregory R.

*greg.keele@jax.org

Multiparental populations (MPPs) encompass more genetic diversity than traditional experimental crosses, enabling deeper interrogations of genetic variation regulating complex traits. For mouse models, two related MPPs have emerged: the Collaborative Cross (CC) inbred panel and the Diversity Outbred (DO) population. Additionally, the F1 intercrosses of the CC (CC-RIX) allow for studies of replicable outbred mice. Researchers often seek to characterize and genetically dissect traits through heritability estimation and mapping regulatory genetic loci. Here we evaluate the relative merits of these populations for these tasks through simulation, as well as provide recommendations for performing the quantitative analyses. We find that sample populations that include replicate animals, as possible with the CC and CC-RIX, provide more efficient and precise estimates of heritability. CC-RIX populations with replicates enable heritability to be decomposed into additive and non-additive components. All populations of approximately 200 animals were well-powered to detect large effect loci that explain ≥ 40% of the phenotypic variation, but only large sample populations of 500 DO mice were well-powered to detect smaller effect loci (≤10%) for highly polygenic traits. All results were produced with our R package musppr, which we developed to simulate data and evaluate genetic analyses from user-provided genotypes from these MPPs.

Hybrid Genotyping Pipeline on Lc-WGS and ddGBS Sequences for Heterogeneous Stock Rats

[1*]Chen, D., [1]Chitre, A.S., [1]Nguyen, K., [1]Cohen, K., [1]Cheng, R., [1]Polesskaya, O., [1,2]Palmer, A.A.

* dec037@health.ucsd.edu

[1]Department of Psychiatry, University of California San Diego
[2]Institute for Genomic Medicine, University of California San Diego

Mammalian model organisms, such as rats, offer the chance of direct experimental manipulations for biologists to study relationships between genes and behavioral traits. Our laboratory has been focusing on the use of outbred heterogeneous stock (HS) rats, which were derived in 1984 by intercrossing 8 inbred rat strains, to study drug abuse-related phenotypes. HS rats have been maintained as an outbred population for more than 90 generations, and provide both genetic diversity and excellent mapping resolution. In order to perform Genome-wide association studies (GWAS) for drug-abuse related traits, we have developed a cost-effective and high-throughput hybrid genotyping pipeline based on double digest genotyping by sequencing (ddGBS) and low-coverage whole-genome sequencing (Lc-WGS). The ddGBS sequence data (~5x coverage on captured sites) were sequenced on an Illumina HiSeq 2500 with double digest GBS library preparation, and the Lc-WGS sequence data (~0.25x coverage) were sequenced on an Illumina NovaSeq 6000 with Riptide (Twist) library preparation. Our genotyping pipeline uses BWA for alignment, and follows STITCH for genotype imputation through constructing haplotypes. With this hybrid genotyping pipeline, we are able to call 6.5 million SNPs in more than 13,000 HS rats with 99.7% concordance rate compared to the SNPs called with 30x coverage sequence data in 88 samples that were also processed using both ddGBS and Lc-WGS. In conclusion, we developed a genotyping pipeline that allows combining sequencing data obtained by different library preparation methods to produce dense genotype data, which provides a foundation to reveal novel genetic mechanisms underlying drug abuse and other biomedically important traits.

The Rat Genome Database resources for de novo rat strain assemblies

[1]*Smith, J.R., [1]Tutaj, M.A., [1]Laulederkind, S.J.F., [1]Hayman, G.T., [1]Wang, S.J., [1]Tutaj, M., [1]Kaldunski, M.L., [1]Vedi, M., [1]Demos, W.M., [1]Thota, J., [1]Nalabolu, H.S., [1]Lamers, L., [1]Gibson, A.C., [1]Thorat, K., [1]Brodie, K.C., [1]Zacher, S.M., [1]De Pons, J.L., [2]Doris, P.A., [1, 3]Dwinell, M.R., [1, 3]Kwitek, A.E.

*jrsmith@mcw.edu

[1]The Rat Genome Database, Medical College of Wisconsin, Milwaukee, WI, USA.
[2]Center for Human Genetics, McGovern Medical School, University of Texas Health Science Center, Houston, TX, USA.
[3]Department of Physiology, Medical College of Wisconsin, Milwaukee, WI, USA.

As the premiere online location for consolidated rat genomic, genetic, phenotypic and disease data, the Rat Genome Database (RGD, https://rgd.mcw.edu) is committed to the integration of the full spectrum of genomic data to support research into the underlying causes of human disease. One exciting new development in rat genomics is the whole genome sequencing and de novo assembly of rat strains that are established models of disease. The recent release of reference quality assemblies and preliminary genome annotations for three rat strains, SHRSP/BbbUtx (RGD:8142383, GCA_021556685), SHR/Utx (RGD:8142385, GCA_023515785) and WKY/Bbb (RGD:1581635, GCA_023515805) has prompted RGD to incorporate these data into our comparative genomics platform. Currently, genomic positions on all three of these assemblies are available on RGD rat gene pages. RGD has also created JBrowse instances for these assemblies to facilitate exploration of genes and regions of interest in their genomic context. In addition, problematic genomic regions in the mRatBN7.2 reference assembly that were reported based on analysis of these assemblies have been added to the queue for review by RGD curators as part of our ongoing collaboration with the Genome Reference Consortium (GRC). Finally, RGD has begun work to upgrade our genome browsers to the more feature-rich JBrowse 2.0. JBrowse 2.0's breakpoint split viewer, whole genome alignment dotplot viewer and linear synteny viewer will facilitate comparisons between these three de novo assemblies, the mRatBN7.2 reference assembly, and in the future, any additional de novo assemblies being produced.

Genetic analysis of multi-omics data identifies drivers of protein phosphorylation

Tian Zhang[1], Gregory R. Keele[2], Steven C. Munger[2], Fernando Pardo-Manuel de Villena[3], Martin Ferris[3], Joao A. Paulo[1], Steven P. Gygi[1], Gary A. Churchill[2*]

1. Harvard University, Boston, MA
2. The Jackson Laboratory, Bar Harbor, ME
3. University of North Carolina, Chapel Hill, NC

Abstract

Site-specific phosphorylation of proteins is a key step in the regulation of many cellular processes including activation of enzymes and signaling cascades. Phosphorylated peptides (phosphopeptides) can be detected and quantified by mass spectrometry. The abundance of a phosphopeptide is determined by the abundance of its parent protein and the proportion of target sites that are phosphorylated. We quantified phosphopeptides, proteins, and transcripts in heart, liver, and kidney tissue samples from 116 genetically diverse female and male mice of the Collaborative Cross strain panel. We mapped ~700 phosphorylation quantitative trait loci (phQTL) across the three tissues and applied genetic mediation analysis to identify causal drivers of phosphorylation. We identified kinases, phosphatases, cytokines, and other factors, including both known and potentially novel interactions between target proteins and genes that regulate site-specific phosphorylation. We find evidence of coordination of phosphorylation across multiple sites within a protein and across proteins that form complexes. Our analysis highlights multiple targets of pyruvate dehydrogenase kinase 1 (PDK1), a regulator of metabolic function that shows reduced activity in the NZO/HILtJ mouse, a polygenic model of obesity and type 2 diabetes.

Increasing power in inbred strain association mapping by recognizing variance heterogeneity

[1]Ashner, M.C., [1]Corty, R.W. & [1]*Valdar, W.

*william.valdar@unc.edu

[1]University of North Carolina at Chapel Hill

Modern quantitative trait locus (QTL) mapping in panels of inbred strains uses a linear mixed model (LMM) to test for SNP-phenotype association while accounting for a random effect of population structure. A decade of mathematical tricks have mitigated the computational expense of repeatedly fitting this complex model for genome-wide applications. Existing procedures, however, assume that the phenotype of each strain (or individual) is known with equal precision. In reality, this assumption does not always hold. We propose a method, weighted Inbred Strain Association Mapping (wISAM), which accounts for heteroscedastic residual variance in the study population using a weighted regression technique and makes use of variance shrinkage methods to stably estimate these weights. Simulation studies comparing wISAM to existing methods demonstrate that it can provide additional statistical power for GWAS. The method is then illustrated using data from studies on the Hybrid Mouse Diversity Panel (HMDP).

The regulatory landscape of multiple brain regions in outbred heterogeneous stock rats

[1,2]*Munro, D., [3]Wang, T., [1]Chitre, A.S., [1]Polesskaya, O., [2]Ehsan, N., [1]Gao, J., [4]Gusev, A., [5]Solberg Woods, L.C., [6]Saba, L.M., [3]Chen, H., [1,7]Palmer, A.A. & [2,8]Mohammadi, P.

*dmunro@scripps.edu

[1]Department of Psychiatry, University of California San Diego, La Jolla, CA
[2]Department of Integrative Structural and Computational Biology, Scripps Research, La Jolla, CA
[3]Department of Pharmacology, Addiction Science and Toxicology, University of Tennessee Health Science Center, Memphis, TN
[4]Division of Population sciences, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA
[5]Wake Forest University School of Medicine, Department of Internal Medicine, Section on Molecular Medicine, Winston Salem, NC
[6]Department of Pharmaceutical Sciences, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of Colorado Anschutz Medical Campus, Aurora, CO
[7]Institute for Genomic Medicine, University of California San Diego, La Jolla, CA
[8]Scripps Research Translational Institute, Scripps Research, La Jolla, CA

Heterogeneous Stock (HS) rats are a genetically diverse outbred rat population that is widely used for studying genetics of behavioral and physiological traits. Mapping Quantitative Trait Loci (QTL) associated with transcriptional changes would help to identify mechanisms underlying these traits. We generated genotype and transcriptome data for five brain regions from 88 HS rats. We identified 21,392 cis-QTLs associated with expression and splicing changes across all five brain regions and validated their effects using allele specific expression data. We identified 80 cases where eQTLs were colocalized with GWAS results from nine physiological traits. Comparing our dataset to human data from the Genotype-Tissue Expression (GTEx) project, we found that the HS rat data yields twice as many significant eQTLs as a similarly sized human dataset. We also identified a modest but highly significant correlation between genetic regulatory variation among orthologous genes. Surprisingly, we found less genetic variation in gene regulation in HS rats relative to humans, though we still found eQTLs for the orthologs of many human genes for which eQTLs had not been found. Finally, we compared these eQTL results to preliminary analysis of a larger RNA-Seq dataset from 339 HS rat brain hemispheres. These data are available from the RatGTEx data portal (RatGTEx.org) and will enable new discoveries of the genetic influences of complex traits.

A novel quantitative trait locus implicates *Msh3* in the propensity for genome-wide short tandem repeat expansions in mice

Mikhail Maksimov[1,2], Cynthia Wu[3*], David G. Ashbrook[4], Flavia Villani[4], Vincenza Colonna[4,5], Nima Mousavi[6], Nichole Ma[1], Lu Lu[4], Jonathan K. Pritchard[7,8], Alon Goren[1,10], Robert W. William[4], Abraham A. Palmer[9,10], Melissa Gymrek[1,2,10]

*cew003@ucsd.edu

[1]Department of Medicine, University of California San Diego, La Jolla, CA
[2]Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA
[3]Bioinformatics and Systems Biology Program, University of California San Diego, La Jolla, CA
[4]Department of Genetics, Genomics, and Informatics, University of Tennessee Health Science Center, Memphis, TN
[5]Institute of Genetics and Biophysics, National Research Council, Naples, 80111, Italy
[6]Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA
[7]Department of Genetics, Stanford University, Stanford, CA
[8]Department of Biology, Stanford University, Stanford, CA
[9]Department of Psychiatry, University of California San Diego, La Jolla, CA
[10]Institute for Genomic Medicine, University of California San Diego, La Jolla, CA

Short tandem repeats (STRs) are a class of rapidly mutating genetic elements typically characterized by repeated units of 1-6 nucleotides. We leveraged whole genome sequencing data for 152 recombinant inbred (RI) strains from the BXD family derived from C57BL/6J and DBA/2J mice to study the effects of genetic background on genome-wide patterns of new mutations arising during parent-to-offspring transmission at STRs. We defined quantitative phenotypes describing the numbers and types of germline STR mutations in each strain and identified a locus on chromosome 13 associated with the propensity of STRs to expand. Several dozen genes lie in the QTL region, including *Msh3,* a known modifier of STR stability at pathogenic repeat expansions in mice and humans and current drug target for Huntington's Disease. Detailed analysis of the locus revealed a cluster of variants near the 5' end of *Msh3*, including multiple protein-coding variants within the DNA mismatch recognition domain of MSH3, and a retrotransposon insertion overlapping an annotated exon. Additionally, gene expression analysis demonstrates co-localization of this QTL with expression QTLs for multiple nearby genes, including *Msh3*. Our results suggest a novel role for *Msh3* in regulating genome-wide patterns of germline STR mutations and demonstrate that inherited genetic variation can contribute to variability in accumulation of new mutations across individuals.

Polygenic Transcriptome Risk Scores Can Translate Genetic Results Between Species

[1]Santhanam, N., [2,3]Sanchez-Roige, S., [1]Liang, Y., [2]Chitre, A.S., [2,4]Munro, D., Chen, D., [2]Cheng, R., [1]Nyasimi, F., [1]Perry, M., [2]Gao, J., [5]George, A.M., [2]Gileta, A., [6]Holl, K., [7]Hughson, A., [8]King, C.P., [9]Lamparelli, A.C., [5]Martin, C.D., [10]Garcia-Martinez, A., [1]Mi, S., [2]St. Pierre, C.L., [8]Tripi, J., [10]Wang, T., [10]Chen, H., [11]Flagel, S., [5,8]Ishiwari, K., [8]Meyer, P., [12]Saba, L., [13]Solberg-Woods, L.C., [2]Polesskaya, O., *[2]Palmer, A.A., [1]Im H.K.

* aap@ucsd.edu

[1]Department of Medicine, Section of Genetic Medicine, The University of Chicago;
[2]Department of Psychiatry, University of California San Diego
[3]Department of Medicine, Division of Genetic Medicine, Vanderbilt University
[4]Department of Integrative Structural and Computational Biology, Scripps Research
[5]University at Buffalo, Clinical and Research Institute on Addictions University at Buffalo
[6]Medical College of Wisconsin, Department of Pediatrics
[7]University of Michigan, Department of Psychiatry,
[8]University at Buffalo, Department of Psychology
[9]Department of Psychology, University of California Los Angeles
[10]Department of Pharmacology, University of Tennessee Health Science Center
[11]Department of Psychiatry, University of Michigan, Ann Arbor, Michigan, USA
[12]University of Colorado, Department of Pharmaceutical Sciences
[13]Wake Forest University School of Medicine, Department of Internal Medicine

Genome-wide association studies have demonstrated that most traits are highly polygenic; however, translating these polygenic signals into biological insights remains difficult. A lack of satisfactory methods for translating polygenic results across species has precluded the use of model organisms to address this problem. Here we explore the use of polygenic transcriptomic risk scores (PTRS) for translating polygenic results across species. Unlike polygenic risk scores (PRS), which rely on SNPs for predicting traits, PTRS use imputed gene expression for prediction, which allows cross-species translation to orthologous genes. We first developed RatXcan, which is a framework for transcriptome-wide association studies (TWAS) in outbred rats. Leveraging predicted transcriptome and genotype data from UK Biobank, and the genetically trained gene expression models from RatXcan, we scored more than 3,000 rats using a human-derived PTRS for height. Strikingly, we found that human-derived height PTRS significantly predicted body length in rats ($P<0.013$). The genes included in the PTRS were enriched for biological pathways including skeletal growth and metabolism and were over-represented in tissues including pancreas and brain. This approach provides an empirical metric by which to evaluate the suitability of specific animal models and identify their shared biological underpinnings.

Dissecting the mechanisms of Indirect Genetic Effects in outbred mouse population

[1*]Tonnelé, H. & [1,2]Baud, A

[*]helene.tonnele@crg.eu

[1] Centre for Genomic Regulation, Barcelona Institute of Science and Technology, Barcelona, Spain
[2] Universitat Pompeu Fabra, Barcelona, Spain

Increasing evidence shows that the phenotype of an individual can be affected by the genotypes of other, interacting individuals. These effects are called "indirect genetic effects" (IGE), in contrast with "direct genetic effects" (DGE) arising from the individual's own genotypes. IGE have in particular been detected between familiar and unfamiliar cage mates in populations of outbred laboratory mice[1,2]. These associations were detected using a variance decomposition approach that models the effect of the genotypes of the cage mates on the phenotype of interest *without any reference to the cage mates' phenotypes*, thus, the traits of cage mates mediating IGE are unknown. The genome-wide association study of IGE (igeGWAS) was previously developed to address this question and identify specific loci giving rise to IGE, which can point to the traits of cage mates involved[2]. However, very large samples sizes are required for igeGWAS.

Here we present a new, complementary approach applicable to smaller sample sizes. This new approach leverages the phenome data available in several outbred rodent datasets, including multiple behavioural, physiological and morphological traits. Considering each of those phenotypes in turn as "phenotype of interest", our approach consists in scanning all the traits in the dataset and identifying those whose genetic basis (DGE) is correlated with IGE on the phenotype of interest. From this analysis we obtain a correlation matrix for all pairs of phenotypes that helps understand whether IGE arise from behavioural, immune or metabolic interactions, for example.

To calculate the correlation between DGE on a trait of the dataset and IGE on the phenotype of interest, we developed bivariate linear mixed models including both DGE and IGE. We sought to validate this new model using simulations based on the real genotypes and cages of one outbred mouse dataset. So far we have simulated phenotypes affected by DGE and IGE but no correlation between DGE and IGE. Our preliminary results show the bivariate model yields unbiased estimates of DGE and IGE in that case. In the future we will simulate and analyse phenotypes with cross-phenotypes DGE-IGE correlations to check that we get accurate and precise estimates for this parameter. We will then analyse several mouse and rat datasets to gain insights into the mechanisms of social influence.

1. Baud A, Mulligan MK, Casale FP, Ingels JF, Bohl CJ, Callebert J, et al. *Genetic variation in the social environment contributes to health and disease*. PLoS genetics. 2017;13(1):e1006498.
2. Baud A, Casale FP, Barkley-Levenson AM, Farhadi N, Montillot C, Yalcin B, et al. *Dissecting indirect genetic effects from peers in laboratory mice*. Genome Biology. 2021;22(1):216. doi: 10.1186/s13059-021-02415-x.

Alternative Splicing of Osteopontin in CRISPR/Cas9 Knock Out Rats

Trivett, C., [1*], Graham, D. [1], McBride, M. [1]

* c.trivett.1@research.gla.ac.uk

[1] School of Cardiovascular and Metabolic Health, University of Glasgow, Glasgow, UK

Introduction:
In hypertensive populations, patients non-uniformly develop an increased left ventricle mass (LVM), without in an increase in LV function, associated with increased morbidity and mortality. Following identification of a quantitative trait locus (QTL) for LVM on rat chromosome 14, congenic strains on the normotensive (WKY) and hypertensive (SHRSP) backgrounds were generated (WKY.SPGla14a & SP.WKYGla14a respectively). Studies using these animals positionally and functionally implicated the osteopontin gene (*Spp1*) in development of left ventricular hypertrophy, independent of blood pressure. To investigate the molecular actions of *Spp1*, in the context of a genetic model of hypertension, the CRISPR/Cas9 system was used to introduce a mutation into the *Spp1* gene of SHRSP/A3NCrl embryos (SHRSP-*Spp1*^{em1Mcwi}).

Methods:
WKY, SHRSP and SHRSP-*Spp1*^{em1Mcwi} strains were bred and selectively phenotyped at neonatal (days 1-3) and 5-week timepoints. Liver tissue was collected for DNA extraction and genotype confirmation of SHRSP-*Spp1*^{em1Mcwi} as heterozygous, wild type or knock-out. Heart and kidney tissues were removed, and snap frozen in liquid nitrogen. Neonatal hearts were split to extract both protein and RNA from individual hearts. Protein was analysed by SDS-PAGE gel electrophoresis and immunoblotting. Gene expression was assessed by Taqman qRT-PCR. End-point PCR primers were designed to flanking exon containing deletion mutation. PCR fragments capturing *Spp1* open reading frame (ORF) were cloned using pTarget Mammalian Expression Vector into competent cells.

Results:
There were no differences in heart, body or heart to body weight ratio of neonatal (KO = 5.33 mg/kg, WT= 6.34mg/kg p= 0.669) or 5-week-old Spp1+/+ (WT) and Spp1-/- (KO) animals. (KO = 4.44mg/kg, WT=4.47mg/kg, p= 0.759). Cardiac assessment by echocardiography revealed no significant difference in LVM (p = 0.961) between WT and KO littermates. *Spp1* mRNA was significantly increased in neonate hearts of SHRSP animals ($p<.001$) versus both WKY and KO neonates. *Spp1* mRNA expression was also significantly reduced in KO animals versus WKY ($p<.001$). Western blotting analyses revealed *Spp1* protein expression was maintained in neonate hearts from KO animals. End point PCR suggested Spp1-/- animals produce an alternative RNA form of *Spp1* producing a truncated osteopontin protein. Sanger sequencing of spliced *Spp1* transcripts will confirm end-point PCR findings.

Conclusions:
Spp1-/- animals produced using CRISPR/Cas9 technology produce a truncated form of *Spp1* protein, potentially through alternative splicing to remove exons containing INDEL mutation. Residual protein expression in CRISPR/Cas9 generated knock-outs should be fully characterised to determine whether preservation of *Spp1* function has occurred.

Incorporating strain-specific variance into Genome-wide inbred strain mapping for zero-inflated phenotypes

[1,3]*McGee, T.K., [2]Ashner, M., [1,3]Corty, R. & [3,4]Valdar, W.
*tmcgee@unc.edu
[1]Bioinformatics and Computational Biology Curriculum, UNC Chapel Hill
[2] Biostatistics Curriculum, UNC Chapel Hill
[3] Department of Genetics, UNC Chapel Hill
[4] Lineberger Comprehensive Cancer Center, UNC Chapel Hill

Modern quantitative trait locus (QTL) mapping in panels of inbred strains uses a linear mixed model (LMM) to test for SNP-phenotype association while accounting for a random effect of population structure. Existing procedures, however, assume strains are equally variable. In an accompanying presentation, we propose a method, weighted Inbred Strain Association Mapping (wISAM), that accounts for heterogeneous strain variances by down-weighting more variable strains. Here we show in a Collaborative Cross study of drug metabolism-related peptide abundances how accounting for strain variance--although usually beneficial--can introduce problems for the analysis, specifically when phenotypes include observations with near zero, intra-strain variation. We propose a system of analysis that expands on identifying drivers of zero-inflation and zero-variance, whether those correspond to truly zero outcomes or whether they simply represent measurements below the limit of detection. We explain why such zero-inflated phenotypes cause problems for wISAM, and for any variance-aware analysis, and describe practical solutions.

Scalable *in vitro* model using scRNA-seq to generate cell-type specific transcriptomic profiles of osteogenic cells

[1*]Dillard, L., [1]Rosenow, W., [1]Calabrese, G., [1,3]Mesner, L., [1,2]Al-Barghouthi, B., [1,2]Abood, A., [1]Farber, E.A., [1,3]Onengut-Gumuscu, S., [4]Tommasini, S., [4]Horowitz, M., [5]Rosen, C., [6]Yao, L., [6]Qin, L., [1,2,3]Farber, C.R.

* ljd9az@virginia.edu

[1]Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908
[2]Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA 22908
[3]Department of Public Health Sciences, University of Virginia, Charlottesville, VA 22908
[4]Department of Orthopaedics and Rehabilitation, Yale School of Medicine, New Haven, CT 06520
[5]Maine Medical Center Research Institute, Scarborough, ME 04074
[6]Department of Orthopaedic Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104

Skeletal development, bone remodeling, and fracture healing are maintained by numerous cell-types. Historically, genomic studies of bone cells have been challenging due to difficulties in isolating homogenous cell-types from marrow or bone. Recently, this has begun to change with the emergence of single-cell technologies. Here, we profile the transcriptomes of cultured bone marrow-derived stromal cells (BMSCs), a popular model of osteoblast differentiation and activity, from five Diversity Outbred (DO) mice using single-cell RNA-seq (scRNA-seq). The goals of the study were to explore technical challenges, evaluate cellular heterogeneity, and determine if BMSCs could serve as a model for the generation of cell-type specific transcriptomic profiles of osteogenic cells derived from hundreds of mice in order to inform genetic studies. We demonstrate that dissociation of BMSCs from a heavily mineralized matrix has little effect on viability or their transcriptomic signatures. Furthermore, we show that BMSCs cultured under osteogenic conditions are diverse and consist of cells with characteristics of mesenchymal progenitors, marrow adipogenic lineage precursors (MALPs), osteoblasts, osteocyte-like cells, and non-osteogenic immune cells. Importantly, all cells were representative and similar from a transcriptomic perspective to their *in vivo* counterparts. We also demonstrate the ability to multiplex single-cells and subsequently assign cells to their "sample-of-origin" using demultiplexing approaches based on genotypes inferred from coding SNPs. We employed other current scRNA-seq analytical tools to rigorously investigate the biological integrity of the identified cell-types. SCENIC was used to reconstruct gene regulatory networks (GRNs) and we show that identified cell-types retain GRNs expected of osteogenic and adipogenic lineage cells. Further, CELLECT analysis showed that genes with higher expression in osteoblasts, osteocyte-like cells, and MALPs contribute more of the genetic signal identified by bone mineral density GWAS than the other cell-types profiled. Together, these data suggest cultured BMSCs derived from DO mice coupled with scRNA-seq can be used as a scalable and biologically-informative model to generate cell-type specific transcriptomic profiles of osteogenic lineage cells. Ultimately, the BMSC model can be leveraged in genetic studies requiring cellular phenotypes representative of large and diverse mouse or human populations.

Adjusting for cis-regulatory variation in gene expression data can improve biological network analysis in model organisms

[1*]Pattee, J., [2,3]Munro, D., [2,4]Mohammadi, P., [5]Palmer, A., & [6]Saba, L.M.,

*jack.pattee@cuanschutz.edu

[1] Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO, USA
[2] Department of Integrative Structural and Computational Biology, Scripps Research, La Jolla, CA
[3] Department of Psychiatry, University of California San Diego, La Jolla, CA
[4] Scripps Research Translational Institute, Scripps Research, La Jolla, CA
[5] Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA
[6] Department of Pharmaceutical Sciences, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

Often, studying gene co-expression patterns as networks provides insight into the etiology of biological processes. Weighted gene correlation network analysis (WGCNA) is a popular method for modeling RNA co-expression. WGCNA transforms gene-gene correlations into a scale-free network from which cluster assignments can be estimated. Model organisms are a rich data source for network analyses as the reduced genetic and environmental variability in these populations can facilitate co-expression modeling. However, in some model organism data, cis-SNPs may substantially affect the gene expression levels of many nearby genes located in large blocks of high linkage disequilibrium. As a result, close physical proximity between two genes may induce correlation in their expression levels even if they are not functionally related. Thus, cluster assignments estimated via WGCNA may represent co-localization rather than a common biological process in the analysis of model organism data. We investigated this problem by comparing gene clustering with and without adjustment for cis-SNP effects. We used WGCNA to analyze gene expression data from the nucleus accumbens of a panel of 75 heterogeneous stock rats. To determine the biological relevance of the estimated clusters, we conducted overrepresentation tests for gene ontology terms and genes associated with small molecule perturbations in the Signature Commons (SigCom) LINCS database. Adjusting for cis-SNPs generated more clusters in which multiple genes associated with the same small molecule perturbation were included, likely indicating a higher functional relatedness and thus more biological relevance. However, it is of note that for many clusters, adjusting for cis-SNPs did not dramatically influence module membership. In summary, adjusting for cis-SNPs in co-expression network analyses can reduce the number of clusters driven by co-localization while preserving clusters driven by shared biological processes. Supported by a Skaggs Scholars Grant Award from the ALSAM Foundation, by NIDA (P50 DA037844; P30 DA044223), and by NIH R01GM140287.

Genetic analysis of CNS autoimmunity using the diversity of the Collaborative Cross reveals unique phenotypes and mechanisms

[1*]Holt, E.A., [1]Lahue, K.G., [2]Mahoney, J.M., [3]Teuscher, C. & [1]Krementsov, D.N.

*Emily.A.Holt@uvm.edu

[1]Department of Biomedical and Health Sciences, University of Vermont, Burlington, VT 05401, USA
[2]The Jackson Laboratory, Bar Harbor, ME 04609, USA
[3]Department of Medicine, University of Vermont, Burlington, VT 05401, USA

Multiple Sclerosis (MS) is a complex disease with remarkable heterogeneity in disease course and progression, the genetic basis of which remains obscure. Here we leveraged the Collaborative Cross (CC) - a highly genetically diverse mouse strain panel - and myelin oligodendrocyte glycoprotein peptide 35-55 (MOG_{35-55}) induced experimental autoimmune encephalomyelitis (EAE), to model genetics of MS disease course. 32 CC strains were selected based on compatible MHC haplotypes ($H2^b$ and $H2^{g7}$), which captured a wide spectrum of distinct EAE phenotypes, compared with typical chronic EAE in C57BL/6 mice. CC028 mice exhibited severe and rapidly progressing disease. In contrast, several strains, including CC011 and CC040, were highly resistant to EAE. Sex differences in EAE course were observed in 4 strains, including CC042. Remitting-relapsing EAE was observed in 4 strains, including CC002. In addition to classical EAE clinical signs (ascending paralysis), we identified two strains, CC004 and CC083 that exhibited high incidence of axial rotary (AR)-EAE, including profound ataxia, head tilt, and axial rotation. Preliminary quantitative trait locus (QTL) analysis revealed a distinct linkage pattern in each sex for classical EAE severity, with several emerging peaks on chromosome (Chr) 2, 8, 12, 18, and X in females, Chr10 in males, and Chr4 in both sexes. QTL analysis of AR-EAE severity revealed a narrow interval on Chr18 (39.2-41.0Mb) passing suggestive linkage significance at the 90% confidence threshold. Experiments are ongoing to determine the immunopathologic basis of distinct genetically controlled EAE phenotypes in CC strains of interest.

Genetic regulation of heart mitochondrial proteome influencing heart function

[1*]Karthickeyan Chella Krishnan, [2]Elie-Julien El Hachem, [3]Luke Carroll, [3]Alexis Diaz Vegas, [4]Christine Light, [5]Yang Cao, [5]Calvin Pan, [5]Karolina Elżbieta Kaczor-Urbanowicz, [5]Varun Shravah, [5]Diana Anum, [5]Matteo Pellegrini, [4,6]Chi Fung Lee, [7]Marcus M. Seldin, [8]Benjamin L. Parker, [3]David E. James and [5]Aldons J. Lusis

[*]chellakn@ucmail.uc.edu
[1]University of Cincinnati College of Medicine, Cincinnati, OH, USA
[2]Sorbonne Université, Paris, France
[3]University of Sydney, Sydney, New South Wales, Australia
[4]Oklahoma Medical Research Foundation, Oklahoma City, OK, USA
[5]University of California, Los Angeles, CA, USA
[6]University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA
[7]University of California Irvine, CA, USA
[8]University of Melbourne, Melbourne, Victoria, Australia

Mitochondria play a key role in the normal function of the heart as well as in the pathogenesis of diseases. We report analysis of common genetic variations contributing to mitochondrial and heart functions using an integrative proteomics approach in a panel of inbred mouse strains called the Hybrid Mouse Diversity Panel (HMDP). We performed a whole heart proteomic analysis in the HMDP (72 strains, n=2-3 mice) and retrieved 840 mitochondrial proteins (quantified in ≥50 strains). High-resolution association mapping on their respective abundance levels identified three *trans*-acting genetic loci, located on chromosome (chr) 7, chr13 and chr17, that control distinct classes of mitochondrial proteins as well as heart hypertrophy. Follow-up high resolution regional mapping identified NDUFS4, LRPPRC and COQ7 as the candidate genes for chr13, chr17 and chr7 loci, respectively, and both experimental and statistical analyses supported their causal roles. Variations of all three were associated with heart mass in two independent heart stress models, namely, isoproterenol (ISO)-induced heart failure and diet-induced obesity (DIO) models. To identify the aspects of mitochondrial metabolism regulated by these loci, we constructed co-expression protein networks using weighted gene co-expression network analysis (WGCNA). DAVID enrichment analyses of genes regulated by each of the loci revealed that the chr13 locus was highly enriched for complex-I proteins (24 proteins, P = 2.2E-61), the chr17 locus for mitochondrial ribonucleoprotein complex (17 proteins, P = 3.1E-25) and the chr7 locus for ubiquinone biosynthesis (3 proteins, P = 6.9E-05). These results indicate that common variations of certain mitochondrial proteins can act in *trans* to influence mitochondrial functions and contribute to heart hypertrophy, elucidating mechanisms that may underlie genetic susceptibility to heart failure in human populations. Overall, our unbiased systems genetics analyses identified three loci regulating mitochondrial function in the heart. It also provides strong support for a role of the mitochondrial proteome in heart pathophysiology.

Gene-by-treatment QTL mapping to dissect genetic susceptibility to severe coronavirus disease

[1,2*]Riseberg, E.R., [3]Leist, S.R., [3]Schäfer, A., [4,5]Anderson, E., [2]Taft-Benz, S., [3]Gully, K.L., [2]Shaw, G.D., [2]Bell, T.A., [2]Hock, P., [2]Miller, D.R., [4,5]Baxter, V.K., [2,6]Pardo-Manuel de Villena, F., [2,6]Valdar, W., [3,6,7]Baric, R., [2,6,7]Heise, M.T., [2]Ferris, M.T.

*erisembe@email.unc.edu

[1]Bioinformatics and Computational Biology Curriculum, UNC Chapel Hill
[2]Department of Genetics, School of Medicine, UNC Chapel Hill
[3]Department of Epidemiology, School of Public Health, UNC Chapel Hill
[4]Department of Pathology, School of Medicine, UNC Chapel Hill
[5]Department of Comparative Medicine, UNC Chapel Hill
[6]Lineberger Comprehensive Cancer Center, UNC Chapel Hill
[7]Department of Microbiology & Immunology, School of Medicine, UNC Chapel Hill

Gene-by-treatment (GxT) effects are an important component of the genetic architecture of complex traits, especially those involving a large environmental component such as diet or infection response. GxT effects are most easily modeled in inbred organisms where genetic replicates allow for calculation of a response phenotype. GxT in outbred organisms requires more sophisticated modeling. Here we describe a statistical model for identifying GxT quantitative trait loci (QTL), and describe results from applying this model to data from a treatment-control experiment in an outbred population. In this experiment, an F2 cross was performed between coronavirus susceptible (CC006/TauUnc) and coronavirus resistant (CC044/UncJ) Collaborative Cross strains to identify genetic regulators of coronavirus disease severity. F2 mice were exposed to one of three coronaviruses or saline to establish a control group. We report results from GxT QTL mapping in these mice, including several loci associated with disease severity and immune response.

Network-based analysis predicts interacting genetic modifiers from a mapping study of multiple absence epilepsy mouse models

[1]*Lara, M.K., [1]Brabec, J.L., [2]Tyler, A.L., [1,2]Mahoney, J.M.

*mlara@uvm.edu

[1] Department of Neurological Sciences, University of Vermont, Burlington, VT, United States
[2] The Jackson Laboratory, Bar Harbor, ME, United States

Absence epilepsy (AE) is characterized by brief lapses in awareness accompanied by the hallmark spike-and-wave discharge (SWD) EEG pattern, commonly attributed to abnormalities in the corticothalamic neural circuit. In humans, most AE is idiopathic and has a polygenic etiology. While many genes have been associated with AE, including Mendellian forms with a single causal allele, there are many unknown genetic modifiers of AE influencing variation in risk and severity. Similar to this human genetic complexity, multiple transgenic mice separately carrying SWD-causing alleles, including *Gabrg2*$^{tm1Spet(R43Q)}$, *Scn8a*^{8j}, and *Gria4*^{spkw1}, have demonstrated strain-specific variation in SWD phenotypes. Using a series of experimental crosses between transgenic seizure-resistant C57BL/6J (B6) and seizure-prone C3HeB/FeJ (C3H) strains carrying one of these mutations, Tyler *et al.* (2014) identified epistatically interacting loci on mouse chromosomes 2 and 7 influencing SWD. The C3H allele at either locus increases SWD severity, but C3H alleles at both loci have a sub-additive effect, independent of the SWD-causing mutations. These results implicate universal modifiers in the B6 background that mitigate SWD severity across all causes through a common pathway. However, because of the low mapping resolution of experimental crosses, these loci are each around 10 MB and contain hundreds of positional candidate genes. Here, we sought to prioritize candidate interacting modifiers within these loci responsible for the observed epistasis. By integrating evidence from human AE GWAS, differential gene expression in B6 and C3H mouse brains, and cortex- and thalamus-specific gene interaction networks, we ranked all gene-pairs spanning the chromosome 2 and 7 loci for association with AE risk gene networks. Our analysis yielded highly plausible candidate gene-pairs—including *Tubgpc4-Ppme1*, *Chp1-Rab6a*, and *Mertk-Myo7a*—which have strong functional associations to genes involved in human AE risk and plausible mechanistic pathways supporting a possible epistatic interaction at the gene level. Our top-ranking gene pairs represent novel hypotheses about the sources of SWD variation across strain backgrounds, which could clarify universal mechanisms driving differences in AE severity.

GRK5 is required for adipogenesis through activation of ERK pathway

[1]*Mary Seramur, [1]Bailey McDonald, [2]Swapan Das, [1]Leah Solberg Woods, [1]Chia-Chi Chuang Key

* mseramur@wakehealth.edu

1. Department of Internal Medicine, Section on Molecular Medicine, Wake Forest University School of Medicine, Winston Salem, NC, USA.
2. Department of Internal Medicine, Section on Endocrinology and Metabolism, Wake Forest University School of Medicine, Winston Salem, NC, USA.

As the prevalence of obesity continues to grow in the U.S., adult obesity and its comorbidities are now a major health threat to our population. Obesity is caused by the chronic imbalance between energy intake and energy expenditure resulting in adipose tissue expansion which can occur through adipocyte hypertrophy and/or hyperplasia (i.e., adult adipogenesis). Although both genetics and environment contribute to obesity, the genetic basis remains poorly understood. Data from our human cohort, an outbred rat model, and diet-induced obese mice demonstrated a positive correlation between the gene expression of G protein-coupled receptor (GPCR) kinase 5 (GRK5) in white adipose tissue and adiposity, and mediation analysis in the outbred rat suggests GRK5 is playing a causal role in obesity. This is further supported by Wang et al. who reported that whole body GRK5 knockout (KO) mice exhibited decreased adipogenesis, and protection from diet-induced obesity. However, the mechanism by which GRK5 regulates adipogenesis is unknown.

To fill this knowledge gap, we created a GRK5 KO 3T3-L1 pre-adipocyte cell line. We found that, during adipogenic stimulation, GRK5 KO pre-adipocytes had decreased white adipogenesis and lipid accumulation. Similar results were observed in brown-like adipocyte differentiation. In pre-adipocytes, insulin of the adipogenic cocktail acts through insulin-like growth factor-1 receptor (IGF-1R), a receptor tyrosine kinase (RTK), which can activate two downstream pathways: Extracellular signal-regulated kinase (ERK) and AKT. We found that deletion of GRK5 suppressed insulin stimulated ERK, but not AKT, phosphorylation and activation. These findings indicate that adipogenesis is regulated by GRK5 potentially through insulin/IGF-1R/ERK pathways, suggesting that GRK5 is not only a serine/threonine GPCR kinase, but might also regulate RTK signaling in pre-adipocytes.

Adcy3 knock-out rats show increased adiposity without changes in behavioral phenotypes

[1]Roberts, M.K., [1]Szalanczy, A.M., [1]Vora, A., [2]Grzybowski, M., [2]Geurts, A., & [1]Solberg Woods, L.C.

mackrobe@wakehealth.edu

[1]Wake Forest, Department of Molecular Medicine
[2]Medical College of Wisconsin

Both environmental and genetic factors contribute to a person's risk of developing obesity. In a genome-wide association study (GWAS) using heterogeneous stock (HS) rats, our lab previously identified a protein-coding variant in the transmembrane domain of the gene *adenylate cyclase 3* (*Adcy3*) that is associated with adiposity. Both expression-altering variants and protein-coding variants in *Adcy3* have been associated with adiposity and also with depression in humans. Existing rodent studies investigating this gene have used *Adcy3* knockout (KO) models in mice, showing increases in both adiposity and emotional behaviors. We developed two *in vivo* models, both on the Wistar Kyoto (WKY) rat background: a full *Adcy3* KO strain and a strain with a protein-coding mutation ("Adcy3$^{mut/mut}$") in the ADCY3 transmembrane domain. We have used the *Adcy3* KO model to investigate the role of *Adcy3* in adiposity, metabolic health, and behavior, hypothesizing that *Adcy3* KO rats would show increased adiposity, hyperphagia, poor metabolic health, and increased emotional behaviors relative to wild-type (WT) rats.

We found that *Adcy3* KO rats weigh more than WT rats due to an increase in fat mass without an increase in lean mass. Although *Adcy3* KO rats show a slight increase in food intake, further investigation is needed to fully explain the large differences in adiposity. Surprisingly, we found that *Adcy3* KO rats do not display worsened emotional behaviors relative to WT rats. These results indicate that the adiposity and emotional behaviors observed in previous *Adcy3* KO models may be caused by distinct underlying mechanisms, or that *Adcy3* KO phenotypes may differ between mouse and rat models. Further, as the WKY rat is a depressive rat strain, it is also possible that the WT rats are demonstrating a ceiling effect in the behavioral tests, masking any differences by genotype. Future studies will investigate thermogenesis and energy expenditure in *Adcy3* KO rats, as well as assess adiposity, metabolic health, and behavior in Adcy3$^{mut/mut}$ rats that have a protein-coding *Adcy3* mutation instead of a KO.

Classifying the Discrete Epigenetic Environment Involved in Drug-Seeking Behavior

[1]Litif, C., [1]Sandum, K., [1]Flom, L., [1]Vaccaro, L., [1]Gigley, J., [1]Blouin, N., & [1]Bobadilla, A.C.

*clitif@uwyo.edu

[1]University of Wyoming

Discrete neuronal networks, or neuronal ensembles, within the the nucleus accumbens core (NAcore) are conserved amongst species to motivate the acquisition of survival needs through goal-directed behaviors. Cocaine use disorder (CUD) is linked to the formation and maintenance of specific ensembles, within the NAcore. We know ensembles localized in the NAcore underpin goal-directed seeking of cocaine and non-drug sucrose. This is in part due to discrete epigenetic signatures associated with cocaine- and sucrose-seeking that create reward-specific ensembles. Within the NAcore, we know that 70% of cocaine- and sucrose-seeking ensembles are exclusive to each reward while 30% of neurons overlap. Although we know that epigenetic mechanisms in the NAcore are involved in cocaine-seeking, it is not well understood which epigenetic components are distinguished from non-drug sucrose-seeking. Isolating the epigenetic environment of reward-specific ensembles requires conditioning of both cocaine- and sucrose-seeking within the same brain, followed by separation of reward-specific neurons. This approach has made determining discrete cocaine-seeking related genes technically difficult. To address this gap in knowledge, we aimed to characterize shared or exclusive NAcore-localized epigenetic components involved in cocaine- and sucrose-seeking ensembles. TRAP2 mice underwent innovative polyreward self-administration conditioning to identify and isolate reward-specific ensembles. Using fluorescently-activated cell sorting (FACS) and low-input RNA-sequencing, we were able to define cocaine- and sucrose-seeking neuronal ensembles and their underlying epigentic components. Findings give insight into the genetics mediating cocaine-seeking separate from non-drug reward-driven behaviors. Ultimately this study advances the understanding and prevention of relapse that is associated wtih CUD.

A Genetic Basis for Oxidative Stress Response Variation

[1,2*]O'Connor, C., [1]Keele, G., Martin, W., [1]Gatti, D., [1]Korstanje, R., [1,2]Churchill, G., [1,2]Reinholdt, L.G.

*callan.oconnor@jax.org

1. The Jackson Laboratory
2. Tufts University

Cell morphology can be used to summarize complex molecular response traits in the cell culture environment using high content imaging platforms. Taking advantage of these technologies, we sought to use an *in vitro* systems genetics approach to discover the genetic drivers of chemical exposure response variation. We derived a panel of primary fibroblast cell lines derived from Diversity Outbred (DO) mouse population, exposed them to a range of monomethylarsonous acid (MMAIII) concentrations, and we used high content imaging to quantify oxidative stress response. We performed Quantitative Trait Loci (QTL) mapping of dose-response model parameters fit to >300 cellular morphometric features extracted from more than 190,000 images of fluorescently labeled fibroblasts. Of 1329 dose-response model parameters (DRPs; i.e., min, max, slope, and EC50s), 800 had non-zero heritability (h^2) ranging up to 0.71. QTL mapping identified 57 DRPs with maximum LOD scores above the per-feature permutations threshold set at $\alpha = .05$. Additionally, the confidence intervals for the significant LOD peaks contain 392 genes previously shown to interact with arsenic. Taken together, these results indicate that high content screening of DO cell lines can be used to reliably study gene by environment interactions. Among the most interesting results was for the slope DRP of mitochondrial area (MitoTracker Deep Red) with a significant peak on Chr 10: 82.9 Mbp (LOD score = 9.60), where variant association mapping identified private NZO/HlLtJ (NZO) SNPs in *Txnrd1*, a selenoprotein previously shown to be directly inhibited by MMAIII and modulate redox homeostasis. Several of the highest associated SNPs were in the 3'-UTR or downstream regions of *Txnrd1,* a region critical to the recoding of the selenocysteine amino acid into the protein which serves as critical component of selenoprotein activity. To determine whether *Txnrd1* expression varied across outlying strains, we performed differential expression analysis on a subset of DO fibroblast lines exposed to 0 μM and 0.75 μM MMAIII. *Txnrd1* expression increased significantly (*p adj.* = .0116) in exposed cells, however we did not detect an allele-specific response. Currently, our efforts are focused on determining whether TXNRD1 is affected at the protein level, either through a reduction in protein abundance or reduced enzymatic activity, which both may indicate inefficient selenocysteine recoding during translation. In summary, this example shows genetic variation in the oxidative stress pathway, which *Txnrd1* plays an essential role in, influences the mitochondrial response to MMAIII. Limitations to our study include common pitfalls in dose response modeling, a small (n = 204) DO population for mapping, and an acute exposure of MMAIII beyond levels commonly observed in human populations.

Genetic Background in the Rat Impacts Metabolic Outcomes of Post-wean BPF Exposure

[1]*Wagner, V.A., [1]Clark, K.C., [1]Holl, K.L., [1]Reho, J.J., [2]Solberg Woods, L.C., [3]Lehmler, H., [1,4]Raff, H., [1]Grobe, J.L., [1]Kwitek, A.E.

* vawagner@mcw.edu

[1]Medical College of Wisconsin, Milwaukee, WI
[2]Wake Forest School of Medicine, Winston-Salem, NC
[3]University of Iowa, Iowa City, IA
[4]Advocate Aurora Research Institute, Milwaukee, WI

Bisphenol F (BPF) is a common substitute for bisphenol A (BPA), a known obesogen, and is found in polycarbonates and consumer products. Interindividual variation in human bisphenol levels suggests that gene x environment (GxE) interactions influence cardiometabolic disease risk from bisphenol exposure. BPF is a potent endocrine disruptor affecting the thyroid, reproductive health, and neuroendocrine functions. Traditional *in vivo* toxicity studies are performed in isogenic or genetically undefined outbred rodents, leading to conflicting results possibly due to GxE interactions. This limitation can be avoided by studying the N/NIH Heterogeneous Stock (HS) rats, an outbred population derived from eight inbred strains that are amenable to genetic study.

We hypothesize that the development of BPF-induced cardiometabolic disease has underlying genetic risk identifiable using the HS rat and its founding inbred strains. We previously demonstrated that five weeks post-weaning BPF exposure significantly impacts body growth and adiposity in male HS rats. This project aimed to evaluate the metabolic impact of post-weaning BPF exposure in a subset of HS founding inbred strains. Weanling littermate pairs of male ACI/EurMcwi (ACI), BN/NHsdMcwi (BN), BUF/Mna (BUF), F344/Stm (F344), M520/NMcwi (M520), and WKY/NCrl (WKY) rats randomly received either vehicle (0.1% EtOH) or 1.125 mg BPF/L in 0.1% EtOH for 10 weeks in drinking water. Cardiometabolic measures, tissues, urine, and feces were taken.

BPF affected male endocrine glands. Briefly, ACI males had increased thyroid gland mass, BN males showed a trend in increased pituitary gland and testes masses, and WKY males had increased adrenal gland mass. Preliminary experiments indicated trends in increased expression of thyroid hormone synthetic genes in BPF ACI males. BPF BN males showed a trend in increased circulating total testosterone that may be explained by increased expression of testosterone synthetic genes responding to a change in prolactin secretion. BPF WKY males did not show evidence of altered adrenal cortex hormones or their synthetic genes despite significantly decreased gene expression of all *Nr4a* subfamily members. BUF and M520 males did not show BPF effects.

These preliminary data suggest that post-weaning BPF exposure has varied effects on the endocrine system depending on genetic background, demonstrating that the HS rat founding inbred strains contributed diverse bisphenol-exposure risk alleles to the population. This supports the hypothesis that BPF exposure is a cardiometabolic disease risk factor and indicates that the HS rat will be a useful model for dissecting GxBPF interactions on metabolic health.

Structural equation modeling uncovers candidate genes for cocaine's conditioned effects in heterogeneous stock rats

[1]*Leal-Gutiérrez J.D., [2]King C.P., [1]Chen D., [1]Li B., [1]Chitre A., [1]Polesskaya O., [3]Holl K., [1]Gao J., [1]Cheng R., [1]Bimschleger H., [4]Garcia A., [5]George T., [1,6]Gileta A., [4]Han W., [7]Horvath A., [7]Hughson A., [5]Ishiwari K., [2]Lamparelli A., [2]Versaggi C., [5]Martin C., [8]St. Pierre C., [2]Tripi J., [4]Wang T., [4]Chen H., [9]Flagel S., [5]Richards J., [10]Robinson T., [11]Solberg Woods L, [2]Meyer P. and [1,12]Palmer A.

*jdleal@health.ucsd.edu

1 Department of Psychiatry, University of California, San Diego, La Jolla, California, USA
2 Department of Psychology, University at Buffalo, Buffalo, New York, USA
3 Human and Molecular Genetic Center, Medical College of Wisconsin, Milwaukee, Wisconsin, USA
4 Department of Pharmacology, University of Tennessee Health Science Center, Memphis, Tennessee, USA
5 Clinical and Research Institute on Addictions, University at Buffalo, Buffalo, New York, USA
6 Department of Human Genetics, University of Chicago, Chicago, Illinois, USA
7 Department of Psychiatry, University of Michigan, Ann Arbor, Michigan, USA
8 Department of Genetics, Washington University, St. Louis, Missouri, USA
9 Molecular and Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, Michigan, USA
10 Department of Psychology, University of Michigan, Ann Arbor, Michigan, USA
11 Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA.
12 Institute for Genomic Medicine, University of California San Diego, La Jolla, California, USA

Abstract

Drug addiction is a complex phenotype determined by a collection of environmental and genetic factors. We performed a Genome Wide Association study based on genomic Structural Equation Modeling (genomic SEM) for cocaine conditioning phenotypes in heterogeneous stock rats. We measured cocaine contextual conditioning (CCC) and cocaine conditioned cue preference (CCP) phenotypes (n=1,376 for CCC and n=1,674 for CCP). A total of 3,513,321 autosomal markers were available for analysis. Each observed phenotype was quantile normalized and adjusted by batch number, coat color, and sex using a linear model estimation. Residuals were fitted in a univariate GWAS using the leave-one-chromosome-out (LOCO) approach; the association threshold was determined using permutation. Univariate GWAS summary statistics were used in conjunction with the R package GenomicSEM to perform a gene mapping based on multivariate analysis. Activity-Distance-Velocity defined the final CCC model. Several significant loci were identified (and candidate genes), including for Activity: Chr7:15Mb (*ORLs*) and Chr10:46 Mb (*ORLs* and *TRPV2*). For Distance, Chr5:17-19 Mb (*PENK*) was found. No associated locus was identified for Velocity. Time-Distance-Activity defined the CFA model for CCP. The associated loci identified for Time included chr19:12 Mb (*LARGE1*). Distance and Activity included chr2:170 Mb (*SLITRK3*) and chr7:144 Mb (*HNRNPA1*), respectively. The genomic SEM analysis outperformed the standard univariate approach based on identified associated loci for CCC and CCP. Six associated loci were identified for CCC and eight for CCP using the multivariate genomic SEM approach, while only one locus was identified using the univariate association analysis. An evident gain in power is shown when using a multivariate approach.

Genome-wide association studies of human and rat body mass index converge on a conserved molecular network

*[1]Brittany S. Leger, *[2]Sarah N. Wright, [3]Sara Brin Rosenthal, [2]Sophie Liu, [2]Tongqiu Jia, [1]Apurva S. Chitre, [1]Oksana Polesskaya, [4]Katie Holl, [1]Jianjun Gao, [1]Riyan Cheng[1], [5]Angel Garcia Martinez, [6]Tony George, [1,7]Alexander F. Gileta, [5]Wenyan Han, [8]Aidan Horvath, [8]Alesa Hughson, [9]Christopher P. King, [9]Alexander Lamparelli, [9]Cassandra L. Versaggi, [4]Connor Martin, [10]Celine L. St. Pierre, [5]Tengfei Wang[5], [1]Hannah Bimschleger, [6]Jerry Richards, [4]Keita Ishiwari, [5]Hao Chen, [10]Shelly B. Flagel, [9]Paul Meyer, [8]Terry E. Robinson, [14]Leah C. Solberg Woods, [2]Jason F. Kreisberg, [2,13]Trey Ideker, [1,13]Abraham A. Palmer

*bsleger@health.ucsd.edu

Affiliations

1. Department of Psychiatry, University of California San Diego
2. Department of Medicine, University of California San Diego
3. Center for Computational Biology & Bioinformatics, Department of Medicine, University of California, San Diego
4. Human and Molecular Genetic Center, Medical College of Wisconsin
5. Department of Pharmacology, University of Tennessee Health Science Center
6. Clinical and Research Institute on Addictions, University at Buffalo
7. Department of Human Genetics, University of Chicago
8. Department of Psychology, University of Michigan
9. Department of Psychology, University at Buffalo
10. Department of Genetics, Washington University
11. Molecular and Behavioral Neuroscience Institute, University of Michigan
12. Department of Internal Medicine, Wake Forest School of Medicine
13. Institute for Genomic Medicine, University of California San Diego

* These authors contributed equally to this work

A vexing issue in genome-wide association studies of a trait (GWAS) is that the genetic variants identified in one species are often differ greatly from those identified in another. Here, we demonstrate that cross-species translation of GWAS findings can be greatly improved by an analysis of co-localization in molecular networks. Using body mass index (BMI) as a complex phenotype, we show that the genes associated with BMI in humans versus rats lack significant agreement. However, network analysis synthesizes prior knowledge of molecular interactions with the findings from GWAS, amplifying biological signal. The networks interconnecting these BMI GWAS-implicated genes have highly significant overlap, highlighting common mechanisms of BMI mediation, including hormonal regulation, epigenetic modification, and synaptic signaling transduction. Genetic perturbations in mice of genes within this network cause abnormal BMI phenotypes in mice, supporting their broad conservation across mammals. We further produced species-specific networks which highlight differential pathways for BMI mediation, including carbohydrate biosynthesis (humans) and glycerolipid metabolism (rodents). Finally, we perform network colocalization for body height/length, yielding similar cross-species convergence. This study advances a general paradigm for determining whether and how phenotypes measured in a model species recapitulate human biology.

Genome-Wide Association Study Finds Multiple Loci Associated with Intraocular Pressure in HS Rats

[1]Kumar, A., [1]Fowler, S. [2,3]Munro, D., [4]Wang, T., [2]Chitre, A.S., [1]Hollingsworth, T.J., [3]Martinez, A.G., [2]St. Pierre, C.L., [2]Bimschleger, H., [2]Gao, J.J., [2]Cheng, R., [3,5]Mohammadi, P., [4]Chen, H., [2,6]Palmer, A.A., [2]*Polesskaya, O., [1]Jablonski, M.M.

*opolesskaya@ucsd.edu

[1]Hamilton Eye Institute Department of Ophthalmology, University of Tennessee Health Science Center, Memphis, TN, USA

[2]Department of Psychiatry, University of California San Diego, La Jolla, CA, USA

[3]Department of Integrative Structural and Computational Biology, Scripps Research, La Jolla, CA, USA

[4]Department of Pharmacology, Addiction Science, and Toxicology, University of Tennessee Health Science Center, Memphis, TN, USA

[5]Department of Integrative Structural and Computational Biology, Scripps Research, La Jolla, CA, USA

[6]Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA

Elevated intraocular pressure (IOP) is influenced by environmental and genetic factors. Increased IOP is a major risk factor in most types of glaucoma. Investigating the genetic basis of IOP may lead to a better understanding of the molecular mechanisms of this disease.

The goal of this study was to identify genetic loci involved in regulating IOP using outbred heterogeneous stock (HS) rats. HS rats are a multigenerational outbred population derived from eight inbred strains that have been fully sequenced. This population is ideal for genome-wide association studies (GWAS) owing to the accumulated recombinations among well-defined haplotypes, the relatively high allele frequencies, the accessibility to a large collection of tissue samples, and the large allelic effect size compared to human studies. Both male and female HS rats (N=1,812) were used in the study. Genotyping-by-sequencing was used to obtain ~3.5 million single nucleotide polymorphisms (SNP) for each individual.

SNP heritability for IOP in HS rats was 0.32, which agrees with other studies. We performed a GWAS for the IOP phenotype using a linear mixed model and used permutation to determine a genome-wide significance threshold. We identified three genome-wide significant loci for IOP on chromosomes 1, 5, and 16. Next, we sequenced the mRNA of 51 whole eye samples to find cis-eQTLs to aid in identification of candidate genes. We report 5 candidate genes within those loci: *Tyr*, *Ctsc*, *Plekhf2, Ndufaf6 and Angpt2*. *Tyr, Ndufaf6* and *Angpt2* genes have been previously implicated by human GWAS of IOP-related conditions. *Ctsc* and *Plekhf2* genes represent novel findings that may provide new insight into the molecular basis of IOP.

This study highlights the efficacy of HS rats for investigating the genetics of elevated IOP and identifying potential candidate genes for future functional testing.

Genetic Loci Associated with Skeletal Response to PTH Therapy

Douglas J. Adams, Dana A. Godfrey, Nicole S. Szeto, Raphael R. Hatami, Veronica M. Butler, Robert D. Maynard, Michael A. David, Gregory K. Ottenberg, Cheryl L. Ackert-Bicknell

Douglas.Adams@CUAnschutz.edu

Department of Orthopedics, University of Colorado Anschutz Medical Campus

Introduction: Teriparatide (Human PTH:1-34) is an anabolic therapeutic used in the treatment of osteoporosis. Variable response to PTH has been reported, with up to 44% of patients showing no change in hip bone mineral density after 24 months of treatment. We have shown that the anabolic response of bone to intermittent PTH (i.e., daily injection) varies across mouse inbred strains, suggesting that the degree of the response is genetic. In this study we conducted a genome-wide association mapping study to identify genetic loci that may interact with PTH treatment.

Methods: The Diversity Outbred (DO) mouse population provides genetic diversity that is ideal for high-resolution mapping studies. Beginning at 12 weeks of age, 782 DO mice were treated with intermittent PTH (40mg/kg/day) or saline for 4 weeks (female: 195 PTH, 184 saline; male: 206 PTH, 197 saline). Mice were genotyped using the Giga Mouse Universal Genotyping Array. The amount of trabecular bone in the distal femur was measured as volume fraction (BV/TV) using 3D microCT imaging, and femur diaphyseal strength was measured via 3-point bending. Loci were mapped by fitting two mixed-effects linear models at each genetic marker and regressing each bone phenotype on the haplotype probabilities. PTH treatment was used as an additive covariate in model 1, and as an additive covariate and interactive covariate with genotype in model 2. The peak logarithm of the odds (LOD) for each locus from model 2 was subtracted from the LOD for model 1 to detect genomic loci that interacted with PTH treatment.

Results: We detected 5 loci for femur strength in model 2 that exceeded a statistical LOD threshold of 8 and were shared with model 1, and 4 such loci for trabecular BV/TV. By subtraction of the LODs, we identified 1 peak for strength and 2 peaks for BV/TV that were suggestive of interacting with PTH treatment to modify these phenotypes. The largest LOD difference occurred at the Chromosome 12 locus for BV/TV. Applying a 1.5 LOD-drop statistical confidence interval for this locus yielded 37 candidate protein-coding genes, including *Alkbh1*, a histone dehydrogenase.

Discussion: Loss of *Alkbh1* in mice results in bone ossification defects, and among the DO founder strains both missense and splice mutations have been noted which could impact gene function, suggesting that *Alkbh1* is a plausible candidate gene associated with bone anabolic response to PTH. This ongoing study demonstrates that genetic mapping using the DO mouse population has the potential to reveal genes that interact with teriparatide treatment in the regulation of bone physiology and treatment of osteoporosis, toward opening new avenues of investigation for identifying safe therapies to treat bone disease.

Characterization of polymorphic short tandem repeats and their association with phenotypes in HS rats

[1]*Eveloff, R.J.,[1] Wu, C.,[2] Chen, D., [3,4]Munro, D., [3,5]Palmer, A.A., [1,5,6]Gymrek, M.

*reveloff@ucsd.edu

[1]Department of Computer Science and Engineering, University of California, San Diego
[2]Bioinformatics and Systems Biology Program, University of California, San Diego
[3]Department of Psychiatry, University of California, San Diego
[4]Department of Integrative Structural and Computational Biology, Scripps Research, La Jolla, CA
[5]Institute for Genomic Medicine, University of California, San Diego
[6]Department of Medicine, University of California, San Diego

Short tandem repeats (STRs), consisting of repeating units of 1-6bp, represent one of the largest sources of genetic variation in mammalian genomes. In humans, variation in repeat copy number at STRs has been implicated in both Mendelian and complex traits including gene expression. Compared to single nucleotide polymorphisms (SNPs), which are primarily bi-allelic, STRs are often highly polymorphic and may exhibit a range of distinct alleles at a single locus, offering unique opportunities for performing genetic mapping. Despite their high polymorphism rates and known functional roles, STR variation has not been well-studied in model rat populations. Here, we applied HipSTR, a computational method for STR genotyping, to analyze approximately 1.5 million STRs based on the Rn7.2 reference genome using available whole genome sequencing for the eight inbred "founders" of the heterogeneous stock (HS) and 88 outbred HS rats derived from these founders. We identified more than 480,000 polymorphic STR loci segregating in the HS population. We used this callset to identify and characterize new mutations that have arisen at STRs during generations of outbreeding from the original founders by comparing observed repeat lengths to those expected based on the inferred founder haplotypes at each locus. Finally, we integrated our STR genotypes with RNA-sequencing for 5 brain regions to identify STRs whose repeat lengths are associated with the expression of nearby (cis-eSTRs) or distal (trans-eSTRs) genes. Overall, we envision incorporating STRs into quantitative trait loci (QTL) mapping will add a rich new layer of polymorphic loci that have not been typically considered in large-scale genomic analyses in model organisms.

PhenoMiner: RGD's quantitative phenotype data repository has data update and data mining tool improvements

[1]*Kaldunski, M.L., [1]Smith, J.R., [1]Thota, J., [1]Tutaj, M.A., [1]Laulederkind, S.J.F., [1]Hayman, G.T., [1]Wang, S-J., [1]Tutaj, M., [1]Vedi, M., [1]Demos, W.M., [1]Nalabolu, H.S., [1]Lamers, L., [1]Gibson, A.C., [1]Thorat, K., [1]Brodie K.C., [1]Zacher S.M., [1]De Pons, J.L., [2]Dwinell, M.R., [1,2]Kwitek, A.E.

* mlkal@mcw.edu

[1]Rat Genome Database, Dept. of Biomedical Engineering, Medical College of Wisconsin, Milwaukee, WI
[2]Department of Physiology, Medical College of Wisconsin, Milwaukee, WI

RGD (https://rgd.mcw.edu) is the primary online resource for laboratory rat genetic, genomic, disease and phenotypic data. The PhenoMiner database and mining tool components were developed for rat quantitative phenotype measurements, from both manual curation of scientific literature as well as uploaded data provided by investigators. Data includes detailed information about what (CMO, clinical measurement ontology), how (MMO, measurement method ontology), and under what conditions (XCO, experimental conditions ontology) phenotypes were measured in what animals (species, strain) for each measurement value. Recently, a concerted effort of manual data curation was undertaken to increase data from scientific literature for HRDP founder strains, especially those underrepresented in the data repository. In addition, based on user feedback, the data mining tool was reworked to improve the user interface for data interactivity. In particular, improvements in the search functionality help make navigating the complex datasets more straightforward. The query results page was redesigned to simplify filtering of the data returned, to facilitate tailoring specific query results. New functionality makes it possible to view results for related terms with the same unit of measurement in the same graph. The tool provides the ability to download either filtered results or all results matching the user's original query, providing users with documentation at multiple levels. Upcoming planned improvements will include the availability of imported high throughput phenotyping data from individual inbred and outbred (i.e., heterogeneous stock) rats that will utilize the new graphing functions. The improvements in the PhenoMiner user interface will enable better data access, filtering, and visualization, and continued ability to download data for further analyses.

Improving *de novo* genome assemblies of the HXB/BXH family of inbred strains using Hi-C data

*Panjun Kim[1], Rachel Ward[1], Robert W. Williams[1], Burt M. Sharp[1], Hao Chen[2]

*pkim11@uthsc.edu

[1]Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, Memphis, TN, USA
[2]Department of Pharmacology, Addiction Science, and Toxicology, University of Tennessee Health Science Center, Memphis, TN, USA

Background: Hi-C is an assay that examines chromatin contact frequencies across a genome to identify three-dimensional genomic interactions. These contact frequencies are highly valuable for *de novo* genome assembly because it provides long-range linking information. This information can be used to join contigs or identify potential misassemblies and improve the quality of genome assemblies at chromosome scale. Various genome assembly techniques, including the latest telomere-to-telomere efforts that produce gapless assemblies, all benefit from Hi-C data. The HXB/BXH recombinant inbred (**RI**) family of strains have been used in many genetic mapping studies on physiological and behavior traits. Having high quality *de novo* genome assemblies of these strains will improve the discovery of causal genetic variants for phenotypes of interest.

Results: We built 30 genome assemblies for the HXB/BXH family of inbred strains, including their parent strains, BN-Lx/Cub and SHR/OlaIpcv, using the Supernova pipeline using 60x coverage of linked-read data. The genome assemblies achieved an average of contig length of 34,088 kb at N50. We improved the scaffold N50 for the genomes of BN-Lx/Cub and SHR/OlaIpcv from 6.8 Mb to 36 Mb and from 2.2 Mb to 26 Mb, respectively using Hi-C data (55x–66x coverage) generated from the prefrontal cortex tissue of the strains.

Discussion: These higher quality genome assemblies will be used for the generation of a rat pangenome assembly and to enable the pangenome to encompass many complete rat genomes as well as DNA variants, including large structural variants. We expect this contribution to lead to efficiently finding causal variants for phenotypes of the HXB family and to enable better exploitation of rat genome assemblies. We are generating additional Hi-C data for other strains in the family and anticipate these data will improve all existing assemblies and be useful for future efforts on telomere-to-telomere genomes of these strains.

PhenoGen: HRDP Transcriptome Data and API

[1]*Mahaffey, S., [2]Wood, C., [2]Pattee, J., [1,3]Hoffman, P., [1]Tabakoff, B., [1]Saba, L.

*Spencer.Mahaffey@cuanschutz.edu

[1]Department of Pharmaceutical Sciences, Skaggs School of Pharmacy & Pharmaceutical Sciences, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA

[2]Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

[3]Department of Pharmacology, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

PhenoGen(https://phenogen.org) is a website that provides the Hybrid Rat Diversity Panel transcriptome data across multiple tissues. The HRDP is composed of 96 strains which include strains from 2 recombinant inbred panels (BXH/HXB and FXLE/LEXF) and genetically diverse inbred strains. The primary method available to browse data is through a genome browser where you can view many detailed tracks and reports related to the ribosome depleted totalRNA sequencing and smallRNA sequencing that has been collected on a growing number of strains in the HRDP. Data available for each tissue includes a reconstructed transcriptome, splice junctions, read depth, variants, and standard annotation tracks. Recent releases of PhenoGen included several updates to the data available and extended the Application Programming Interface to make data more easily accessible by any application using well established web standards.

The HRDPv6 dataset has been partially released with whole brain, heart, liver, and kidney data. HRDPv6 is the first dataset to be based on rn7 and which used strain specific genomes containing SNP variant calls from DNA sequencing of most strains. The brain and liver datasets include 58 strains up from 45 strains with HRDPv5. The transcriptome reconstruction and normalized expression are now available. Additionally, eQTLs for each gene/transcript/tissue and WGCNA modules will be available later this year. HRDPv7 is already in progress which will include 75 strains in whole brain and liver.

The API (https://rest.phenogen.org) has been extended to directly access expression data and the transcriptome reconstruction. The API provides functions to browse the datasets and tissues available and drill down to downloadable files. Extensive metadata is provided for each dataset including preprocessing steps performed for the analysis. Alternatively, this can be viewed through functions provided for R and the results can be downloaded directly into a table in R. These functions are available now as a simple file (see Downloads->REST API) while work is ongoing to provide it as a PhenoGen R package in the future. Extensive documentation is provided for each API function on the documentation page at https://rest-doc.phenogen.org. Documentation includes code snippets for many popular languages as well. Help is also available directly through the API by appending. ?help=Y at the end of any of the function URLs. Current development focuses on extending the data API and providing other useful tools through the site. Please reach out by email or through the site if you have additional requests or suggestions.

Completion of genomic rat variants analysis of the Hybrid Rat Diversity Panel

[1]*Monika Tutaj, [1]Kent C Brodie, [1]Jeffrey L De Pons, [1]Wendy M Demos, [1]Adam Gibson, [1]G Thomas Hayman, [1]Mary L Kaldunski, [1]Logan Lamers, [2]Lynn Lazcares, [1]Stanley JF Laulederkind, [1]Harika S Nalabolu, [2]Rebecca Shilling, [1]Jennifer R Smith, [2]Akiko Takizawa, [1]Ketaki Thorat, [1]Jyothi Thota, [1]Marek A Tutaj, [1]Mahima Vedi, [1]Shur-Jen Wang, [1]Stacy M Zacher, [1,2]Anne E Kwitek, [1,2]Melinda R Dwinell

*motutaj@mcw.edu

[1]The Rat Genome Database, Medical College of Wisconsin, Milwaukee, WI, USA.
[2]Department of Physiology, Medical College of Wisconsin, Milwaukee, WI, USA

The Hybrid Rat Diversity Panel (HRDP) is a group of 96 rat inbred strains selected to study mechanisms of complex traits similar in their pathology to common human diseases. We have analyzed the HRDP whole genomic sequencing data (Illumina short reads) using the high-quality rat reference mRatBN7.2 and variant discovery GATK4 Best Practices recommendations (Broad Institute 2019). The average sequence coverage ranges from 15x to 69x per rat strain sample. We have found that more than 18 mln of germline variants, ~5 mln of short indels and ~13 mln of SNVs, characterize the rat cohort. In addition, we observed a remarkable drop in the number of indels discovered with the new genome reference compared with the old rat assembly from 2014, Rnor6. Our results provide high confidence variants that represent the sequence cohort heterogeneity but also variants that require additional quality testing. Low coverage variants, a fraction of private variants of Brown Norway reference rat strain and genomic regions with accumulation of heterozygous variants will require further analysis. Currently we are incorporating the data into Rat Genome Database (https://rgd.mcw.edu) to provide functional annotations to the rat community, as well as assist in prioritization of potential disease-causing mutations. Thus, researchers can access information about strain specific variants on gene report pages, in Variant Visualizer tool and in genomic browser. The data in vcf format are available for retrieval in the 'Download' section of the RGD site.