Committee Meeting Report **Sahar Mozaffari**
Professor: Carole Ober
GGSB Matriculated 2013 Date: 09/08/16

Progress since last Committee Meeting - June 16, 2015

Awards

- Awarded F31 Ruth L. Kirschstein NRSA
- FASEB MARC Travel Award to ASHG 2015 in Baltimore, MD
- FASEB MARC Travel Award to ASHG 2014 in San Diego, CA
- Genetics and Regulation Training Grant: 2013-2016

Publications

- Gamazon, E. R., Wheeler, H. E., Shah, K. P., Mozaffari, S. V., Aquino-Michaels, K., Carroll, R. J., et al. (2015). *A gene-based association method for mapping traits using reference transcriptome data.* Nature Genetics, 47(9), 10911098. http://doi.org/10.1038/ng.3367[**?**]

Presentations

- Genetics of Model Organisms Club April 21, 2016
  Mozaffari, SV. *Parent of Origin Effects in the Hutterites*
- Complex Trait Mapping Journal Club March 25, 2016
  Paper: Amin V, Harris RA, Onuchic V, et al. *Epigenomic footprints across 111 reference epigenomes reveal tissue-specific epigenetic regulation of lincRNAs.* Nat Commun. 2015;6:6370.
- Human Genetics Work in Progress January 20, 2016
  Mozaffari, SV. *Parent of Origin Effects*
- Molecular Biosciences Retreat November 5, 2015
  Mozaffari SV, DeCara J, Shah S, Herman C, Lang R, Nicolae D, Ober C., *Parent of Origin GWAS with Cardiovascular Disease Associated Traits in the Hutterites.* 2015: Nov 5; Galena, IL.
- ASHG October 9, 2015
  Mozaffari SV, DeCara J, Shah S, Herman C, Lang R, Nicolae D, Ober C., *Parent of Origin GWAS of CVD-Associated Phenotypes in the Hutterites* (Abstract Program #310). Presented at the Annual Meeting of The American Society of Human Genetics; 2015: Oct 9; Baltimore, MD.

Posters

- Mozaffari SV, Gamazon E, Aquino-Michaels K, Cox NJ, Im HK. *Quantifying Context Specificity of Gene Regulation using Predicted Gene Expression Levels.* Poster presented at the Annual Meeting of The American Society of Human Genetics Conference; 2014: Oct 18-22; San Diego, CA

Teaching Assistantship Requirements Completed

- MGCB 31400 (BIOS 21236) *Genetic Analysis of Model Organisms* Fall 2014
  Graduate & Undergraduate Course: Introduction to genetic tools, experiments, and model organisms
- HGEN 47000 *Human Genetics* Fall 2015
  Graduate Course: Classic and modern approaches to studying cytogenetic, Mendelian, and complex human diseases. Grant proposal writing course.

Additional Courses

- STAT 24500 *Statistical Theory & Methods II* Winter 2015
- HGEN 46900 *Human Variation & Disease* Spring 2015
- STAT 35500 *Statistical Genetics* Spring 2015
- myChoice Mini-Course: *Effective Writing in the Biological Sciences* Fall 2015
- HGEN 48600 *Fundamentals of Computational Biology: Models & Inference* Winter 2016
- PBHS 31831 *Genetic & Molecular Epidemiology* Spring 2016

Additional Workshops

- Master R Developer Workshop taught by Hadley Wickham May 2015
- Summer Institute in Statistical Genetics at the University of Washington July 2016

Extracurricular

- Museum of Science & Industry: Science Connections volunteer Fall 2014-current
- myCHOICE Internship: Institute of Translational Medicine Summer 2015
  Translate complex research into dynamic science stories. Share translational research stories in weekly newsletter, ITM website, and social media platforms

Proposal Updates:

I would like to revise my aims to focus on Parent of Origin Effects. There are many different aspects of the project to explore and I would like to take advantage of it. These were my previous aims.

AIM 1 To identify and characterize parent of origin effects and allele specific effects on gene expression in 430 Hutterites.
AIM 2 To investigate association of gene age and eQTLs in endometrial-expressed genes in humans.
AIM 3 To characterize the contributions of genetic variants, IBD segments, parent of origin, functional variants, and relatedness among individuals to predict clinical phenotypes.

The following are my proposed revised aims:

AIM 1a To identify and characterize parent of origin effects on quantitative traits in the Hutterites.
AIM 1b To estimate maternal and paternal heritability measures on quantitative traits in the Hutterites.
AIM 2a To identify and characterize parent of origin effects on gene expression in 430 Hutterites.
AIM 2b To identify and characterize parent of origin and allele specific effects on gene expression in 430 Hutterites.

Previously, I shared results on parent of effects on gene expression where I tested maternally- and paternally- inherited alleles separately with the sum of gene expression.[**?**] I received feedback to separate the gene expression into reads that map maternally and those that map paternally, and to use these measures for POeQTL.

Since my Qualifying Exam

AIM 1
I have ran Parent of Origin GWAS, testing maternally and paternally- inherited alleles separately with 22 different traits. The manhattan plots for the Maternal and Paternal GWAS for LDL are in Figure 1. Table 1 has the phenotypes with significant maternal and paternal allele associations.

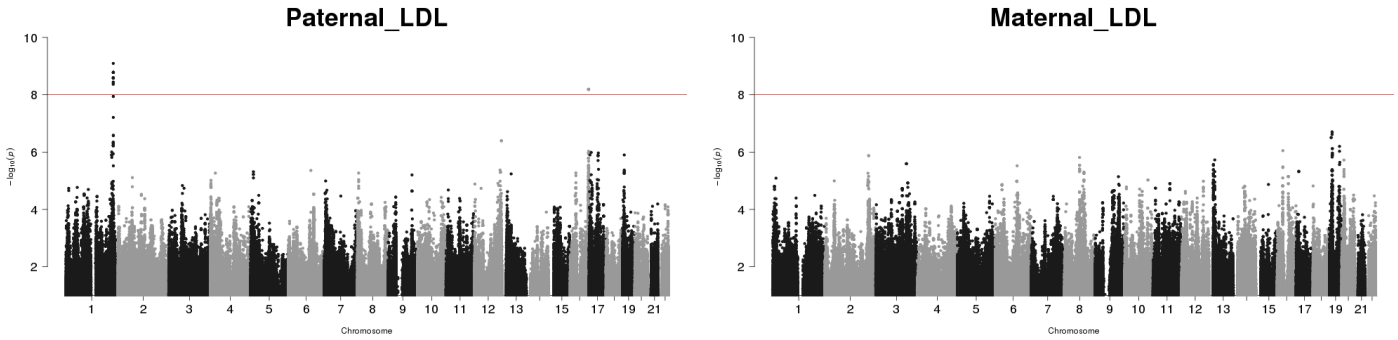

Figure 1: Paternal and Maternal allele PO- GWAS Manhattan plots for LDL.

	Phenotypes
Significant Maternal Association	Carotid Intima Media Thickness, FEV_1, Trigylcerides
Significant Paternal Associations	Systolic Blood Pressure, Blood eosinophil count LDL-C Total cholesterol
No Significant Parental Associations	Left Atrial Volume Index, Left Ventricular Mass Index, FEV_1/FVC, Bronchial Responsiveness Index, Fraction exhaled nitric oxide Diastolic Blood Pressure, Lymphocyte Count, Monocyte Count, Neutrophil Count, IgE, Chitin, YKL40, BMI, Height, HDL-C

Table 1: Phenotypes with Significant Maternal or Paternal Allele Associations.

I have also ran a new model (equation 1) that tests for difference of parental effects in two of these phenotypes (LDL & BMI), and I am working on running them on the remaining 20 phenotypes. I am working on replicating the interesting findings with BMI in the Framingham cohort.

$$Y = \mu + (\beta_p - \beta_m)\frac{(X_p - X_m)}{2} + \frac{(\beta_p + \beta_m)}{2}X_{mp} + \epsilon \quad (1)$$

AIM 1b

With the help of Mark Abney, I have models to test for parent of origin heritability. I have estimated the average maternal and average paternal heritability for each of the 22 traits and I am working on getting a more accurate measure (as opposed to average).

AIM 2a

I have remapped the LCL RNA-seq data using STAR[**?**] and corrected sample swaps using verifyBamID (fixed 2 samples that were already included, and gained 4 samples).[**?**] I used WASP[**?**] to remove mapping bias and mapped reads to the maternal and paternal haplotypes.

For this aim I have methods to detect patterns of parent of origin effects (i.e. imprinting) using maternal and paternal gene expression but not any SNPs (no POeQTL). First I test for asymmetry in maternal and paternal gene expression (normalized total gene expression) with permutations of the data. The second test uses a binomial test to get a Z-score of maternal and paternal expression within each sample (not normalized gene expression). The distribution of the Z-score can be tested against a normal distribution with the Shapiro-Wilk test. The most significant genes from this test have a lot of overlap with known imprinted genes as shown in Table 2 & 3.

Gene	Imprinted	Asymmetry	Shapiro
CACNA2	no	0	0.58
UNC93B1	no	0	0.72
HCAR2	no	0	NA
CTAGE10	no	0	5e^{-08}
MTRNR2	no	0	0.46
CRYBB2P	no	0	0.18
H19	yes	0	1.57e^{-08}
SUPT4H1	no	0	0.75
EIF4A3	no	0	0.14
ZNF597	yes	0	5.63e^{-13}
DIDO1	no	0	1.05e^{-03}
EIF2AK1	no	0	0.12
LOC388897	no	0	1.65e^{-08}
KCNQ1	yes	0	0
SEC61A1	no	0	2.43e^{-06}
CPNE1	no	0	0
SEC22B	no	0	0

Table 2: Paternally Imprinted (maternally expressed) genes significant from the Asymmetry test (p-values in column 3); top 17 genes with p-value 0. Known imprinting status in column 2, and Shapiro test p-value in column 4. Colored in pink are known paternally imprinted genes from geneimprint.com.

Gene	Imprinted	Asymmetry	Shapiro
MEST	yes	0	2.3e^{-07}
MRPL28	no	0	0.13
BMP8A	yes	0	3.67e^{-06}
EIF5AL1	no	0	0.08
FAM50B	yes	0	6.05e^{-12}
PRIM2	conflicting	0	0.72
TCEA1	no	0	0.02
LPAR6	no	0	3.65e^{-05}
PCGF5	no	0	2.18e^{-17}
DUSP22	no	0	0.24
ZNF331	no	0	0.24
NHP2L1	yes	0	3.13e^{-05}
ZDBF2	yes	0	1.31e^{-09}
PEG10	yes	0	2.29e^{-10}
TMEM30A	no	0	0.98
BCLAF1	no	0	1.5e^{-04}
MAP2K3	no	0	7.2e^{-13}

Table 3: Maternally Imprinted (paternally expressed) genes significant from the Asymmetry test (p-values in column 3); top 17 genes with p-value 0. Known imprinting status in column 2, and Shapiro test p-value in column 4. Colored in blue are known maternally imprinted genes from geneimprint.com.

AIM 2b

I will test for POeQTLs in this aim testing maternally inherited SNPs with the maternal gene expression and paternally inherited SNPs with paternal expression. I will combine this with POeQTL results from before using the sum of gene expression, especially for genes which we don't have maternal or paternal expression.

References

[1] Gamazon, E. R., Wheeler, H. E., Shah, K. P., Mozaffari, S. V., Aquino-Michaels, K., Carroll, R. J., et al. (2015). A gene-based association method for mapping traits using reference transcriptome data. Nature Genetics, 47(9), 10911098. http://doi.org/10.1038/ng.3367

[2] Cusanovich, D. A., Caliskan, M., Billstrand, C., Michelini, K., Chavarria, C., De Leon, S., et al. (2016). Integrated analyses of gene expression and genetic association studies in a founder population. Human Molecular Genetics, ddw061. http://doi.org/10.1093/hmg/ddw061

[3] Dobin, A., & Gingeras, T. R. (2015). Mapping RNA-seq Reads with STAR. Current Protocols in Bioinformatics. 51, 11.14.119. http://doi.org/10.1002/0471250953.bi1114s51

[4] G. Jun, M. Flickinger, K. N. Hetrick, Kurt, J. M. Romm, K. F. Doheny, G. Abecasis, M. Boehnke,and H. M. Kang, (2012) Detecting and Estimating Contamination of Human DNA Samples in Sequencing and Array-Based Genotype Data, AJHG doi:10.1016/j.ajhg.2012.09.004 (volume 91 issue 5 pp.839 - 848)

[5] van de Geijn B, McVicker G, Gilad Y, Pritchard JK. (2015) WASP: allele-specific software for robust molecular quantitative trait locus discovery. Nat Meth. 12:1061-1063. doi:10.1038/nmeth.3582.