Regulatory Mechanisms of Gender Associated Differentially Expressed Genes

MAHEETHA BHARADWAJ
Biomedical Informatics, Stanford University School of Medicine
Stanford, CA 94305, USA
Email: maheetha@stanford.edu

WARREN CHAN
Biomedical Informatics, Stanford University School of Medicine
Stanford, CA 94305
Email: wchan28@stanford.edu

SHIRBI ISH-SHALOM
Biomedical Computation, Stanford University
Stanford, CA 94305
Email: shirbi@stanford.edu

TIMOTHY LEE
Biomedical Informatics, Stanford University School of Medicine
Stanford, CA 94305
Email:timothyl@stanford.edu

JOHN LOUIE
Computer Science, Stanford University
Stanford, CA 94305
Email: jwlouie@stanford.edu

Cardiovascular disease related mortality is a leading cause of death and costs $444 billion dollars annually. Since it is a multifactorial disorder, the disease stems from a complex web of causes and it is difficult to pinpoint the perpetrators. To add an additional layer of complexity, the causes seem to manifest themselves in men and women differently since there is a 30% mortality rate in men, versus only 21% in women. One such cause is the underlying gender-specific genetic risk factors; however, the regulatory mechanisms of these genetic factors are unknown. Determining the underlying regulatory mechanisms of the gender-specific genetic factors could help diagnose and treat, and even potentially prevent cardiovascular disease in the future. We thus utilize the Genotype-Tissue Expression (GTEx) database and the Ensembl Genome database to obtain genomic data in order to be able to produce cohorts for significantly differentially expressed genes and their respectively regulatory regions for males and females. Then, we identified significant SNPs within the male and female gene and regulatory regions to be able to compare the allele frequencies between the male and female cohorts. The discovery of these differentially expressed SNPs in males and females can offer new targets for drug therapy, and can be used to generate a framework for predicting new treatments, making more accurate diagnoses, and initiating preventive care based on gender-specific expression.

1. Introduction

Cardiovascular disease related mortality is the leading cause of death and costs $444 billion dollars annually [9]. Since it is a multifactorial disorder, the disease stems from a complex web of causes and it is difficult to pinpoint the perpetrators. To add an additional layer of complexity, the causes seem to manifest themselves in men and women differently since there is a 30% mortality rate in men, versus only 21% in women even after taking lifestyle choices into account [19].

According to a review by Regitz-Zagrosek in 2006, there is an abundance of research that shows that the treatment and research of cardiovascular diseases are significantly influenced by a subject's sex and that sex differences should be considered by clinical and research professionals when developing treatments and evaluating their efficacy [13]. Hypertrophic cardiomyopathy, an ailment that results in thickened myocardium of the heart, has been found to manifest differently in men and women. In a paper by Olivotto et al., researchers found that "[women] with HCM[hypertrophic cardiomyopathy] were under-represented, older, and more symptomatic than men, and showed higher risk of progression to advanced heart failure or death," all of which are indicators and conditions that should be taken into consideration during HCM diagnosis and treatment [12].

Even basic heart functions like cardiac rhythm are sex differentiated [13]. Ventricular repolarization in females, specifically the rate-corrected QT (QTc) interval, is prolonged relative to males, a phenomenon that predisposes women to a life-threatening condition known as “torsade des pontes” or TdP [1]. TdP, a form of ventricular tachycardia initiated by a prolonged QT interval, is often drug-induced in women and has resulted in the removal of several cardiac drugs from the market [1]. Additionally, different mutations in cardiac potassium ion channels indicate lifetime cardiac events for individuals of different sexes. In a genetic study by Zareba et. al. in 2003, researchers found that male children with a KCNQ1 potassium channel gene mutation (LQT1) exhibited a significantly higher risk of cardiac events, whereas female adults with LQT1 and HERG potassium channel gene mutations (LQT2) exhibited a significantly higher risk of cardiac events [10].

Considering the abundance of evidence that shows that sex is a significant factor in multiple heart disease risk factors, we believe that an eQTL study of male and female SNPs can help elucidate genetic factors that contribute to sex-specific cardiac diseases. We hope to utilize our findings to help guide potential drug development for future research.

2. Methods

2.1 *Identify Significantly Differentially Expressed Genes for Men and Women*

First, we identified genes connected to heart tissue that were differentially expressed in men and women. The expression data was procured from the Genotype-Tissue Expression (GTEx) database, and was already fully processed, normalized, and filtered. The GTEx project is a project started by the Broad Institute of MIT and Harvard, and aims to provide to the scientific community a resource with which to study human gene expression and regulation and its relationship to genetic variation. This portal has a total of approximately 8000 samples taken from 450 donors.

The genes analyzed are eQTL genes, meaning that they have a significant cis-eQTL acting upon the gene. The gene set analyzed in this section corresponded to the atrial appendage of the heart with 159 samples. We downloaded the Heart_Atrial_Expression_Data file and the Heart_Atrial_Covariate_Data file from the GTEx website. Both of these are readily available on the GTex Portal [8]. This is also provided in the files we have uploaded.

The gender of each de-identified patient was known from a separate covariates file. Several methods were used to determine the significance of the differences in gene expression between males and females. First, a student’s t-test was used to assess the significance for each gene. To control for false discovery rate, corresponding p-values for each set of genes were converted into q-values. A q-value with a significance below 0.05 was considered statistically significant. In addition, non-parametric tests such as the Mann-Whitney test were used to assess the significance for each gene, avoiding the assumption made in the student’s t-test that expression levels for a gene follow a normal distribution. Again, corresponding p-values were corrected to q-values before determination of statistical significance. The intersection of the sets of genes determined as significantly differentially expressed by these separate techniques was deemed as truly differentially expressed.

The data used within this section differed from that used in the subsequent analysis because the gene expression of some tissue samples was quantified but not the specific genotypes of their eQTLs.

2.2 ***Map Significantly Differentially Expressed Genes to Regulatory Regions***

We mapped the differentially expressed eQTL genes attained in the previous section for each gene set to specific regulatory regions upstream of these genes.

We looked to the Ensembl genome browser for information on which regulatory regions mapped to each of our differentially expressed genes. Ensembl contains a page with a table of regulatory region information for each gene in the human genome. We made a Python script to iteratively download the HTML document of this table for each of of our differentially expressed genes. We then found a Python library called Beautiful Soup that pulls data out of HTML files using an HTML parser, and we used this tool to find all elements in the table necessary for producing an effective mapping of differentially expressed gene to its upstream regulatory regions, namely: Ensembl stable id's for all regulatory regions associated with the queried gene, the chromosome on which each regulatory region resides, and each regulatory region's start and end location on that chromosome.

In this way, we procured every regulatory region for all differentially expressed eQTL genes from the previous section that had this specific regulatory region table element in Ensembl. We arranged the data into the bed file format that could easily be used in conjunction with SNP location information in the next section to map SNPs in the regulatory regions to each differentially expressed eQTL.

2.3 ***Use eQTL Data to Find Significant SNPs correlated with Heart Tissue Genes***

With information about the regulatory regions of the differentially expressed genes, we needed to find SNPs within these regulatory regions that are correlated with heart tissue genes. Specifically, we wanted to find SNP-eQTL gene relationships which significantly showed that a change in genotype of the SNP would cause a change in eQTL gene expression.

First, we used the same gene expression data as section 2.1 from the GTEx website. Again, the samples we used in this project are atrial heart samples (78 samples). In addition, we required the chromosome locations of the eQTL genes. Gene location data was taken from GENCODE, [7]. We downloaded the "Comprehensive Annotation" in GTF format. The gene location data was in a bed file format, with gene (ENSGXXXXXXXXXXX) and start and end positions.

The last file we needed was not publicly available on the GTex Portal due to sensitive information. It is the variant calling file (VCF), which gives the phased haplotype information for all 450 donors. We gained access to this file, through a lab (that Warren and Maheetha work in). Overall, the VCF file for GTex data had approximately 11,000,000 heterozygous variants across all donors.

We used the R package, "MatrixeQTL," to compare the VCFs of SNPs to our gene expression data, accounting for confounding variables listed in the covariates file [17]. For each SNP and gene, MatrixeQTL creates a linear regression model with gene expression as the response variable and multiple covariates and the genotype as the explanatory variables. The genotype is defined quantitatively as the number of recessive alleles (as further explicated in step 1 of the code flow portion.).

$$expression = \alpha + \sum_k \beta_k \cdot covariate_k + \gamma \cdot genotype\_additive \quad [16]$$

Expression = level of gene expression α = intercept-term K = number of covariate variables

β_k = linear coefficient of kth covariate $covariate_k$ = value of kth covariate variable

γ = linear coefficient of genotype_additive Genotype_additive = number of recessive alleles

After fitting each SNP to the linear model, a t-test with a significance threshold of 0.01 for the q-value was used to determine whether the genotype of a SNP is significantly correlated to a difference in gene expression. We kept SNPs that had at least one significant relationship with a eQTL gene, and removed the other SNPs from downstream

analysis. In addition, we separated the analyses for males and females based on the assumption that the genotype of these SNPS may regulate gene expression in different ways.

The code flow of this portion of the methods has been uploaded. It comes in four parts:

1. Separating the VCF file into parsable portions that can then be modified to represent haplotype information. At first, the giant VCF file was separated into smaller 50000 lined VCF files. Then a directory was made for each of the smaller VCF files, which were then copied into their respective directories and separated into 1000 line VCF files. For each VCF file, "0|0" which indicated the homozygous reference was converted to 0, "0|1" or "1|0" which indicated heterozygous was converted to 1, and "1|1" which indicated homozygous recessive was converted to 2. Any SNP that was marked ".|." suggested that you couldn't confirm the genotype. These were converted to "NA". SNPs that didn't code "PASS" were not included in the analysis.
2. The VCF files were also copied and manipulated to show the frequency of the major allele for each individual. Individuals for ".|." had negative numbers, so they could be easily filtered out. Each line from the VCF was taken, modified, and output with the SNP ID name. This was later used for the chisq test. Thus each directory ended up having the same number of individual SNP_vcf files as in the smaller VCF file.
3. Running MatrixeQTL on each of the individual smaller 1000-line VCF files. The data was first separated into male and female samples and MatrixeQTL was run separately for each gender. The output was a significant SNP-Gene correlation file for females and males separately. Due to the large number of SNPs, correction of the p-value by conversion to q-value was used to control for false discovery rate.

This portion of the program involved incredible parallelization of parts due to the large size of the VCF files and the computational complexity of the data cleaning.

2.4 ***Mapping SNPs to Regulatory Regions from Ensembl***

We then mapped each of our significant SNPs from section 2.3 to their corresponding regulatory regions. Any significant SNP that was not in one of the regulatory regions from section 2.2 was removed from downstream analysis. This was done by intersecting the gene regulatory bed file from section 2.2 with the eQTL-SNP files from section 2.3 using the bedtools intersectBed method.

2.5 ***Constructing the ChiSq Analysis for SNPs to Detect Difference Between Genders***

We took the intersection of significant SNPs found in Section 2.4 and queried for their corresponding VCF files within our file system. The VCF files, created by splitting up the larger VCF as detailed in part 2.3, provides the copy number information for each sample for each SNP. These SNP VCF files are separated by gender, meaning that we had to query independent files for men and women.

We used the copy numbers from each SNP VCF query to calculate the genotypic and allelic frequencies for both males and females. We calculated genotypic frequency by pulling out copy number counts for one of three different ranges: 0 to 0.5 (aa), 0.5 to 1.5 (aA or Aa) and 1.5 to 2.0 (AA). We calculated the allelic frequency based on copy number counts that were either over (dominant) or under (recessive). Given the different frequency values, we conducted a chi-squared test for each significant SNP using both genotype and allele as the first categorical variable and gender as the second.

3. Results

3.1 ***Specific Aim 1: Build Gene Cohorts for Significantly Expressed Male and Female Genes***

After conducting the student's t test and the non-parametric Mann-Whitney test on the eQTL genes for patient atrial appendage samples, correcting for false discovery rate and filtering for significant p- and q-values, and taking the intersection of the parametric and non-parametric test, we ended up with 214 significantly differentially expressed genes for males and females for this tissue type.

3.2 *Specific Aim 2: Map Significant Genes to their Respective Regulatory Region*

	A	B	C	D	E
1	gene	regulatory region	chr	start	end
2	ENSG00000130940.10	ENSR00000163274	1	10633784	10636183
3	ENSG00000130940.10	ENSR00000530507	1	10639047	10639739
4	ENSG00000130940.10	ENSR00000530508	1	10640004	10640389
5	ENSG00000130940.10	ENSR00000163277	1	10642350	10642695
6	ENSG00000130940.10	ENSR00000163280	1	10643984	10644783
7	ENSG00000130940.10	ENSR00000530509	1	10646372	10646736
8	ENSG00000130940.10	ENSR00000530512	1	10648555	10650211
9	ENSG00000130940.10	ENSR00000163290	1	10651563	10652281
10	ENSG00000130940.10	ENSR00000530517	1	10662984	10663983
11	ENSG00000130940.10	ENSR00000530520	1	10665874	10671150
12	ENSG00000130940.10	ENSR00001576889	1	10677150	10677471
13	ENSG00000130940.10	ENSR00000530527	1	10677576	10677862
14	ENSG00000130940.10	ENSR00001576890	1	10678384	10678783
15	ENSG00000130940.10	ENSR00000530529	1	10678984	10679983
16	ENSG00000130940.10	ENSR00001576891	1	10679584	10679983
17	ENSG00000130940.10	ENSR00000530530	1	10680384	10681183
18	ENSG00000130940.10	ENSR00001576892	1	10681384	10681783
19	ENSG00000130940.10	ENSR00001517360	1	10681984	10683383
20	ENSG00000130940.10	ENSR00001576893	1	10684984	10685183
21	ENSG00000130940.10	ENSR00000530531	1	10685984	10687583

We then mapped all of our differentially expressed genes to each of their upstream regulatory regions as defined in the Ensembl genome browser.

In this table we show the bed format of the file that we generated mapping each differentially expressed gene to all of its associated regulatory regions, as well as each regulatory region's location in the human genome. For the regulatory region's location itself, we provide the chromosome number as well as the start and end location numbers so that we can easily intersect SNP location numbers in the next section to get the mapping from our regulatory region SNPs to our differentially expressed genes.

Figure 2. Gene-to-Regulatory Region Mapping

3.3 *Specific Aim 3: Determine Significant Gene-SNPs Relationships in Gender Cohorts*

We fit a linear model for each eQTL gene- regulatory region SNP to test whether the SNP genotype significantly affects gene expression. Because of the relatively small sample sizes (78 samples), only 10 of the 36 covariate variables, in addition to genotype, were included in the model. If we were to use all 36 covariate variables, the risk of overfitting to the experimental data would have been higher, given the small sample size.

The Variant Calling File had a total of 11,000,000 SNPs across all sample-types. After running Matrix eQTL to fit each gene- SNP relationship, and finding those relationships that were significant at a p-value of 0.01, there were 3 million SNPs that had at least one significant relationship with a cardiac atrial appendage gene in females. On the other hand, there were approximately 6 millions SNPs that had at least one significant relationship with a cardiac atrial appendage gene in females.

After finding only the SNPs that were within the regulatory regions of part 2.2, and adding information about which regulatory region each SNP resides in, we had approximately 3000 SNPs each for males and females each that were significant at a p-value of 0.01. Stated previously, we had separate files for males and females in order to compare how many SNPs significantly affected gene expression and thus a cardiac event between genders. This was based on the assumption that the genotype of these SNPS may regulate gene expression in different ways between genders.

3.2 *Specific Aim 4: Identify Significant SNPs in Regulatory Regions for Gender Cohorts*

After conducting the chi-square test, we came up with 30 significant SNPs in males and 18 significant SNPs in females. We had fewer cardiac related SNPs in the female cohort than we did in the male cohort.

Gene Name	SNP(s)	Genotypic p-value	Allelic p-value	Associated Phenotype(s)
KCTD18	2_201353083	0.002588617	0.002161047	Blood Pressure
GDF15	19_17968392	0.008377695	0.00518551	Coronary Artery Disease
KDM6B	17_7882428	0.029116715	0.013003546	Height

ASB7	15_101459177	0.021311258	0.016031315	Coronary Heart Disease
ANO6	12_45235884	0.019381276	0.020107537	Blood Coagulation, Carotid, Artery Diseases
**BCAN	1_156752857	0.03032453	0.029325072	Arteries, Stroke
*ADCYAP1	18_492998	0.03075287	0.01237411	Cholesterol
COBLL1	2_164535875	0.048683799	0.034192055	Stroke

Figure 3. Significant SNPs in Males and Females

Figure 3 shows the major SNPs that came up in males versus females. Only 1 SNP appeared in both. No star indicates that the SNP was only found to be significant in the male cohort. One star indicates that the SNP was found to be significant only in the female cohort. Two stars indicate that it was found significantly associated in both female and males.

We plotted the distribution of chisq p-values with respect to location on the genome for the genotypic and allelic differences in male and female SNPs based on the SNP VCF files. These two plots represent the distribution of different significant SNPs. Overall, for males, more SNPs come up as significant in the 5-11 chromosome region, compared to a dip in the same location for females. Here we see a clear difference in gene regulation between males and females. The male and female chi-square tests for alleles are shown in Figures 4A and 5A, and chisq tests for genotypes are shown in Figures 4B and 5B.

4. Discussion/Future Directions/Conclusion

The results above indicate that there are certain SNPs that have significant differences in genotype between genders and can significantly affect eQTL gene expression levels. In Specific Aim 1, we acknowledge that there are eQTL heart genes that are differentially expressed between men and women. In Specific Aim 2 and 3, we pinpoint those SNPs located in regulatory regions that significantly change the gene expression levels of the differentially expressed eQTL heart genes. In Specific Aim 4, we pinpoint from the resulting set of SNPs, the SNPs that had significantly different allelic frequencies or genotypic frequencies between males and females.

At the end of the analysis, there were more significant SNPs associated with eQTL heart genes in males than in females. This may be important in discovering why males may still be more susceptible to heart disease, even after accounting for lifestyle differences [19]. eQTL heart gene underexpression or overexpression due to these varying genotypes in these SNPs may lead to an increased risk of cardiac events. That males inherently have more of these significant SNPs may signify that heart eQTL gene expression levels are more likely to be imbalanced, leading to this increased risk.

However, we had some causes of concerns in our results derived from our methods. In specific aim 3, we found an extreme number of SNPs that had at least one significant relationship with a eQTL heart gene (3 million SNPs for females, 6 million for males). Given that the number of total SNPs in the human genome is 11 million, these numbers seem highly unlikely. We may not have properly accounted for multiple hypothesis testing, and many of SNP-gene relationships may have been falsely discovered, allowing these SNPs to be considered as significant. However, the reduction from 3 million to 3000 SNPs after intersection with the true regulatory regions of the eQTL genes have helped remove these false positives.

In addition, the chi-square analysis on the different genotype and allelic frequencies between males and females may not hold up. There does not seem to be a known biological mechanism that accounts for genotype frequencies of SNPs on autosomal chromosomes to significantly differ between males and females. The allelic and genotypic gender imbalances as shown in the results of Specific Aim 4 may be a result of false discovery from multiple

hypothesis testing. Further analyses on the obtained p-values and scrutinization of our p-value correction methods may be necessary. However, it may still be possible that even with no actual significant difference in the genotypic frequencies of SNPs between genders, different epigenetic patterns such as methylation between genders coupled with differences in SNP genotypes within a population as a whole may account for differential gene expression between genders.

In the end, we have provided a preliminary pipeline for the analysis of differential gene expression between genders and its relationship with SNPs on regulatory regions. The pipeline we have used for eQTL heart genes may be used for genes of other tissue types. It may also be interesting to see whether this differential gene expression may necessitate different medication protocols between men and women. It would be very exciting to see to what new conclusions we can find with this new form of analysis.

5. Acknowledgments

We would sincerely like to thank the Michael Snyder lab for giving us access to the GTex eQTL Variant Calling File data. We would like to thank Michael Snyder for his support and guidance as to the direction of the project. We would like to thank SCG3 cluster for giving us the necessary memory to support our computational work. We would like to extend our thanks to the Broad Institute, MIT, and Harvard for allowing us to use the publicly available GTex data. We would like to thank the members of the team for supporting each other throughout the process.

6. Supplementary Data:

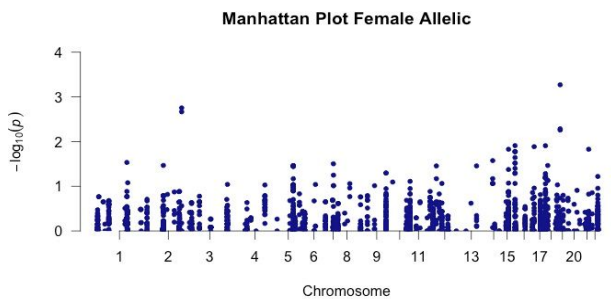

Figure 4A. Distribution of Chi-Square Test p-Values for Female Alleles

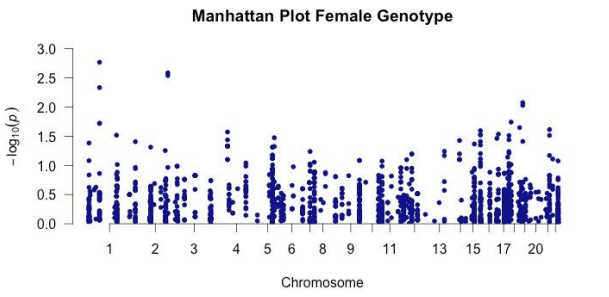

Figure 4B. Distribution of Chi-Square Test p-Values for Female Genotypes

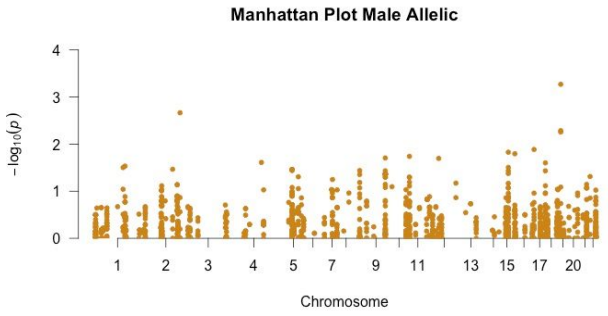

Figure 5A. Distribution of Chi-Square Test p-Values for Male Alleles

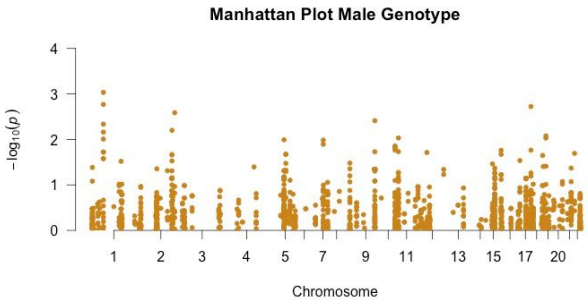

Figure 5B. Distribution of Chi-Square Test p-Values for Male Genotypes

7. References

1. Abi-Gerges, Najah, Karen Philp, Chris Pollard, Ian Wakefield, Tim G. Hammond, and Jean-Pierre Valentin. "Sex Differences in Ventricular Repolarization: From Cardiac Electrophysiology to Torsades De Pointes." Fundam Clin Pharmacol Fundamental and Clinical Pharmacology 18.2 (2004): 139-51. Web.
2. Aguila, Alian, Wassim Jawad, Khyati Baxi, and Joel A. Strom. "Valvular Heart Disease in Women." Management of Cardiovascular Disease in Women (2013): 175-216. Web.
3. Alomar, Muaed Jamal. "Factors Affecting the Development of Adverse Drug Reactions (Review Article)." Saudi Pharmaceutical Journal 22.2 (2014): 83-94. Web.
4. El-Eraky, Hala, and Simon H. L. Thomas. "Effects of Sex on the Pharmacokinetic and Pharmacodynamic Properties of Quinidine." British Journal of Clinical Pharmacology 56.2 (2003): 198-204. Web.
5. Fairweather, Delisa, Leslie T. Cooper, and Lori A. Blauwet. "Sex and Gender Differences in Myocarditis and Dilated Cardiomyopathy." Current Problems in Cardiology 38.1 (2013): 7-46. Web.
6. Franconi, F., S. Brunelleschi, L. Steardo, and V. Cuomo. "Gender Differences in Drug Responses." Pharmacological Research 55.2 (2007): 81-95. Web.
7. "Release 24 (GRCh38.p5)." GENCODE. Wellcome Trust Sanger Institute. Web.
8. "GTEx Portal." GTEx Portal. Broad Institute, 2015. Web.
9. Heart disease and stroke - an overview of our nation's leading killers (2012, February). *National Conference of State Legislatures*, Retrieved from http://www.ncsl.org/issues-research/health/heart-disease-and-stroke-an-overview.aspx
10. Locati, E. H., W. Zareba, A. J. Moss, P. J. Schwartz, G. M. Vincent, M. H. Lehmann, J. A. Towbin, S. G. Priori, C. Napolitano, J. L. Robinson, M. Andrews, K. Timothy, and W. J. Hall. "Age- and Sex-Related Differences in Clinical Manifestations in Patients With Congenital Long-QT Syndrome : Findings From the International LQTS Registry." Circulation 97.22 (1998): 2237-244. Web.
11. Makkar, R. R. "Female Gender as a Risk Factor for Torsades De Pointes Associated with Cardiovascular Drugs." JAMA: The Journal of the American Medical Association 270.21 (1993): 2590-597. Web.
12. Olivotto, Iacopo, Martin S. Maron, A. Selcuk Adabag, Susan A. Casey, Daniela Vargiu, Mark S. Link, James E. Udelson, Franco Cecchi, and Barry J. Maron. "Gender-Related Differences in the Clinical Presentation and Outcome of Hypertrophic Cardiomyopathy." Journal of the American College of Cardiology 46.3 (2005): 480-87. Web.
13. Regitz-Zagrosek, Vera. "Therapeutic Implications of the Gender-specific Aspects of Cardiovascular Disease." Nature Reviews Drug Discovery Nat Rev Drug Discov 5.5 (2006): 425-239. Web.
14. Rinn, John L., Joel S. Rozowsky, Ian J. Laurenzi, Petur H. Petersen, Kaiyong Zou, Weimin Zhong, Mark Gerstein, and Michael Snyder. "Major Molecular Differences between Mammalian Sexes Are Involved in Drug Metabolism and Renal Function." Developmental Cell 6.6 (2004): 791-800. Web.
15. Roberts, W. C., and R. Virmani. "Aschoff Bodies at Necropsy in Valvular Heart Disease. Evidence from an Analysis of 543 Patients over 14 Years of Age That Rheumatic Heart Disease, at Least Anatomically, Is a Disease of the Mitral Valve." Circulation 57.4 (1978): 803-07. Web.
16. Sauer, A. J., and C. Newton-Cheh. "Clinical and Genetic Determinants of Torsade De Pointes Risk." Circulation 125.13 (2012): 1684-694. Web.
17. Shabalin, A.A. Matrix eQTL: Ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 28, no. 10 (2012): 1353-1358.
18. Si, Han, Ramandeep S. Banga, Pinelopi Kapitsinou, Manjunath Ramaiah, Janis Lawrence, Ganesh Kambhampati, Antje Gruenwald, Erwin Bottinger, Daniel Glicklich, Vivian Tellis, Stuart Greenstein, David B. Thomas, James Pullman, Melissa Fazzari, and Katalin Susztak. "Human and Murine Kidneys Show Gender- and Species-Specific Gene Expression Differences in Response to Injury." PLoS ONE 4.3 (2009). Web.
19. Silander, Kaisa, et al. "Gender differences in genetic risk profiles for cardiovascular disease." PloS one 3.10 (2008): e3615.

20. Warnes, C. A. "Sex Differences in Congenital Heart Disease: Should a Woman Be More Like a Man?" Circulation 118.1 (2008): 3-5. Web.
21. Waxman, D. J., and M. G. Holloway. "Sex Differences in the Expression of Hepatic Drug Metabolizing Enzymes." Molecular Pharmacology 76.2 (2009): 215-28. Web.
22. Xia, K., A. A. Shabalin, S. Huang, V. Madar, Y.-H. Zhou, W. Wang, F. Zou, W. Sun, P. F. Sullivan, and F. A. Wright. "SeeQTL: A Searchable Database for Human EQTLs." Bioinformatics 28.3 (2011): 451-52. Web.
23. Zuppiroli, Alfredo, Michael Rinaldi, Randi Kramer-Fox, Silvia Favilli, Mary J. Roman, and Richard B. Devereux. "Natural History of Mitral Valve Prolapse." Journal of Diagnostic Medical Sonography 12.1 (1996): 50. Web.

8. Contributions

Timothy
tim_diff_genes_code_final.R
Worked on finding differentially expressed genes between males and females, final paper, presentation

Maheetha
matrixEqtl_maheetha.R
wrap_matrixEQTL_maheetha.sh
Worked on general data analysis, collection, final paper, presentation

Shirbi
createSNPBedFiles_shirbi.R
chisqMaleFemale_shirbi.R
Worked on generating chi Squared pvalues from SNP regulatory region data, final paper, presentation

John
Worked on background research, code review/debugging, managed workflow, final paper, presentation.

Warren
warren_geneRegMap.R
warren_geneRegMap.py
Worked on mapping genes to regulatory regions, final paper, presentation