The genetic link between depression and cardiovascular disease

A Thesis

By Ana Progovac

For completion of the Bachelor of Science degree
In Neuroscience (with Honors)
Submitted April 2010

[Abstract]

Cardiac patients are more likely to suffer from depression, and depression itself predicts increased morbidity among patients with cardiovascular disease. As research uncovers more and more information about the link between depression and cardiovascular disease (CVD), an increasing number of candidate genes which play some role in the association are also being uncovered. Here we chose to study the Serotonin Transporter Length Polymorphism (5-HTTLPR) in the Serotonin Transporter gene (SLC6A4) and Tryptophan Hydroxylase-2 (TPH2). Both genes are in the Serotonin (5-HT) pathway. Subjects for the study are drawn from the Cardiac Rehabilitation Center here at U of M Preventive Cardiology Services under the direction of Dr. Melvyn Rubenfire, and genotyping of the 5-HTTLPR is combined with extensive clinical information collected at the hospital in an attempt to create a model that accurately characterizes the role of 5-HTTLPR and TPH2 in depression cardiovascular disease.

Acknowledgements

This thesis would have been impossible without these people. I thank them warmly.

Dr. Margit Burmeister – Sponsor. You welcomed me into the lab, and have provided guidance and support for this project as well as my future endeavors.

Dr. Sandra Villafuerte – Mentor. Without your interest and expertise in this topic I would not have gotten the opportunity to work on it. Without your patience and understanding I certainly would not have succeeded in finishing it. Thank you very very much!

Ela Śliwerska – Lab manager. Thank you for all your help with both Taqman assays, and, more importantly, for the help, guidance, and sense of humor throughout my time in the lab.

Kathrin Stoßberg - Medical student. Thank you for the hours and hours you spent perfecting the 5-HTTLPR genotyping. It made my job a whole lot easier!

Dr. Richard Hume – Advisor and First Reader of my thesis. For your time and endless patience, even while on sabbatical.

Dr. John Kuwada – Co-sponsor. For your support and mentorship through the independent research program.

Sanguettha Mahadevan – Statistician. A great help!

The Burmeister Lab – Thank you for the help, the friendships, the memories, and the laughter. And the wonderful food!

Dušan and Ljiljana Progovac – My parents. Thank you both for your love and support throughout college, and for always pushing me to do my best.

It is amazing what can be accomplished when nobody cares about who gets the credit.

-**Robert Yates**

***

If I could solve all the problems myself, I would.

-**Thomas Edison**, when asked why he had a team of twenty-one assistants

***

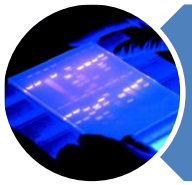
You ought not to attempt to cure the body without the soul. The cure of many diseases is unknown to physicians because they disregard the whole.

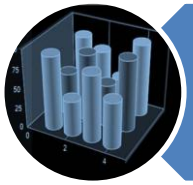
- **Hippocrates**

***

If you want happiness for an hour — take a nap.
If you want happiness for a day — go fishing.
If you want happiness for a year — inherit a fortune.
If you want happiness for a lifetime — help someone else.
Chinese Proverb

Contents





Cover image © 2009 McKibben Landscaping

1. INTRODUCTION

1.1 Background

This project emerges from a long history of research about the relationship between depression and cardiovascular disease (CVD), for which there is a well-established comorbidity, meaning that the two diseases often occur at the same time. Depression is a known risk factor for coronary heart disease (Rugulies 2002), conferring at least two times the risk for developing CVD (Halaris 2009). Likewise, depressive symptoms are present in about 30- 45% of cardiac patients (Schleifer, Macari-Hinson et al. 1989), which is substantial compared to the 7-18% prevalence in the general population. In fact "clinically diagnosed major depression [contributed an increased] risk for the development of CVD… equal [to] the risk of smoking and diabetes" (Van der Kooy, van Hout et al. 2007). But even if depression comes after CVD, it makes this chronic disease worse: patients who have a myocardial infarction (heart attack) and then develop depression have higher mortality rates. Overwhelming data such as these have lead to the exploration of the various ways depression and CVD might lead to and interact with each other.

It is not difficult to imagine how typical symptoms of depression might lead to CVD. Low mood, insomnia, lost of interest in activities, low self-esteem, and feelings of self-worthlessness render an individual less likely to take care of themselves, including through exercise. According to obesity research for example, it is "Consciousness of weight stigma, regardless of objective weight status, [that] may negatively affect individuals' willingness to participate in physical

activity"(Schmalz 2010). Similarly, depressed patients with comorbid CVD are less likely to adhere with their medication or treatment regimen. A meta-analysis of literature from 1968 to 1998 looking at depression and anxiety's impact on non-adherence to treatment looked at 12 articles on depression and 13 on anxiety. Interestingly, anxiety did not have decisive impact on compliance scores. But depression and noncompliance yielded a very significant odds ratio of 3.03 (95% Confidence Interval: 1.96-4.89), meaning that having depression specifically leaves patients three times more likely to have trouble complying to treatment (DiMatteo, Lepper et al. 2000). Surely this loss of interest or ability in self-care can account for a portion of the explanation for why individuals with depression are twice as likely to develop CVD. How CVD might influence depression purely due to the psychological burden is likewise easy to imagine. Chronic illness such as CVD places a great deal of psychological burden on an individual, as we can imagine. Even so, the chronic illness becomes worse when one has depression, as one analysis of 31 published studies including 16,922 patients describes. "Patients with chronic medical illness and comorbid depression or anxiety compared to those with chronic medical illness alone reported significantly higher numbers of medical symptoms when controlling for severity of medical disorder" (Katon, Lin et al. 2007).

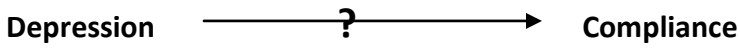
So what is going on here? Researchers have for years now been researching and proposing common underlying mechanisms which may lead to both diseases, as opposed to merely exploring the ways that one might worsen the other. Several biological mechanisms are explored in section 1.3 of this paper, including platelet activation, cytokines, and the body's own stress response. Another portion of the explanation for this relationship could be common genetic risk factors, two of which we explore in this paper.

Even without looking at specific genes and mechanisms, association studies established that there are common genetic risk factors underlying depression and cardiovascular disease. An association study of 2,731 complete pairs of male-male twins which looked at the common genetic and environmental risk factors for depression, heart disease, and hypertension confirmed the co-occurrence of depression and cardiovascular disease, and found a significant genetic correlation between both depressive symptoms and hypertension (r=0.19) and also between depression and heart disease (r=0.42) (Scherrer, Xian et al. 2003).

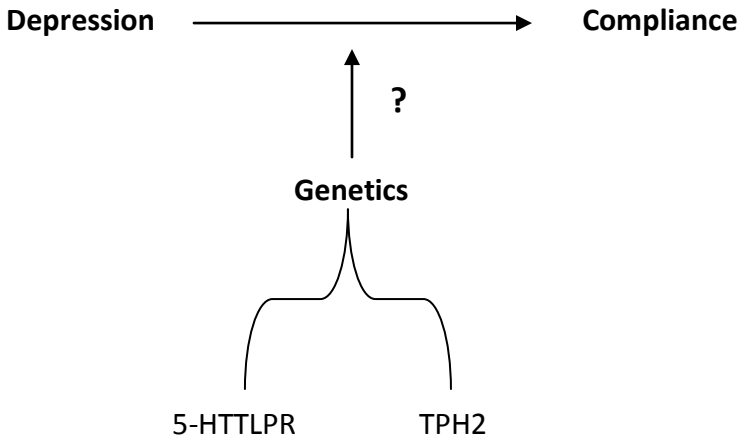
From there, the scramble to identify which specific genetic polymorphisms is part of this shared genetic risk has uncovered numerous candidate genes for study. The two genes we chose to study are the Serotonin Transporter Length Polymorphism (5-*HTTLPR*) and two Single Nucleotide Polymorphisms (SNPs) on Tryptophan Hydroxylase -2 (*TPH2*). The Serotonin Transporter gene has a SNP of interest in addition to the length polymorphism noted above. As will be described in detail, one allele of *5-HTTLPR* has been shown to predispose individuals to depression given environmental exposure to stressful live events (Caspi, Sugden et al. 2003). Similarly, variations in *TPH2* account for a subgroup of depressed patients who are more likely to develop a disorder called Metabolic Syndrome (MetS), which is a cluster of cardiac risk factors (Deen 2004; Kloiber, Kohli et al. 2009).

Our study uses clinical data and DNA from cardiac rehabilitation patients at the University of Michigan to examine the following relationships:

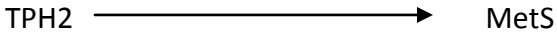
1. Analyzing how *5-HTTLPR* and *TPH2*'s role in mediate standardized mood scores pre and post-cardiac rehabilitation.
2. Verifying the relationship between depression symptoms and compliance in our sample



2.1. Then introducing the *5-HTTLPR* and *TPH2* factors to see whether the genetic risk factors mediate depression's impact on compliance.



3. Verifying *TPH2*'s link to Metabolic Syndrome (MetS) in our sample



1.2 Genetic epidemiology: The genetics of complex traits

In order to understand this study, background knowledge of genetic epidemiology is essential. Complex traits are called so because they are more complicated than phenotypes driven by a single gene or by genes with a larger effect. Because of this, they are much harder to study. In complex traits, there can be many genes involved and the resulting trait can also be influenced or triggered by environmental factors. As a result, studying complex traits in genetics involves careful analysis and extensive knowledge of statistics in order to identify relationships that actually exist and separate out confounding factors that may actually not be strictly associated. There are several ways to study complex traits and their genetic components, but one basic way to start such an inquiry is with association studies.

Given that two individuals have mostly the same genes, they will, of course, differ in the specific versions (alleles) they have for many of those genes. Identifying candidate genes that might contribute to a complex trait can often start with association studies linking specific alleles of a gene to aspects of complex traits and diseases like depression or CVD. Once associations have been found, researchers aim to clarify the contribution that each allele may have to the disease. For complexes diseases or traits, this might mean confirming that a specific allele modifies a normal physical or mental response, making the individual more vulnerable to a disease if they are exposed to enough environmental stimuli. For example, an allele that seems to cause an increased stress response in individuals exposed to stressful life situations may predispose those individuals to develop more complicated diseases like depression later in life. The allele itself does not “cause” depression, but may in conjunction with other factors

increase that individual's risk of developing it. The basis for such inquiries, however, often must start from a statistically sound association study.

Association Studies

Association studies which correlate a specific genotype with a phenotype are performed differently depending on whether the phenotype is categorized as binary, or continuous.

If we start with a binary trait for which individuals can be classified as either having the trait or not having the trait, then a Case-Control design study is a good way to test for association. In a Case-Control study, the frequency of any particular allele is compared between a group of subjects who have the trait of interest (*cases*) and a group selected from the population but matched for age, gender, and race as much as possible (*controls*). To compare the groups, we would use a chi-square test which will determine whether the difference we observe in the data is just a matter of chance, or if it is statistically significant.

Alternatively with a quantitative phenotype, every subject's phenotype is on a continuous scale. There are no cases or controls, just a pool of subjects. The genotypes of interest would be plotted along with a measure of the trait, and correlation established using a regression analysis. It is in this step that one can assess significant covariates for the model in question, and choose the best model to fit the data. Covariates for a trait like height, for example, might include nutritional background or gender of the individual. Height is a classic example of a continuous trait controlled in part by genetics, but which also varies a great deal with the level of nutrition an individual has while growing. Through logistic regression analysis

we sought to develop a model that most accurately describes the relationship between our outcome variable and the genetic predictor variables as well as appropriate covariates.

Regression models identify statistically significant variables and covariates, but there are often common errors in these studies which could lead to erroneous or misleading associations (Cardon and Bell 2001). A few such problems are summarized in **Figure 1**. And even assuming none of these errors are committed, statistical analysis at its best is still a calculation of probabilities. Sometimes chance alone can lead to a false positive.

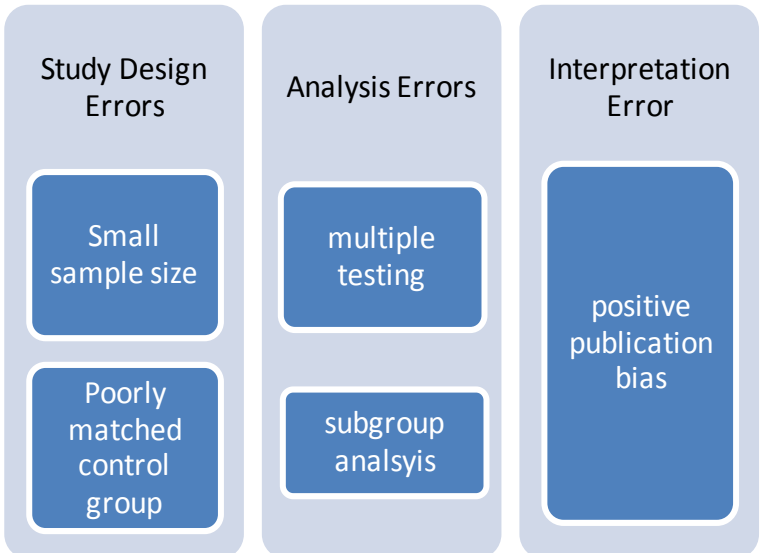


Figure 1. Common Errors in Association Studies. Adapted from *Concepts of Association Studies*, Cardon and Bell, 2001.

In fact, finding positive association can arise from one of at least three possible situations (Lander and Schork 1994). The first possibility, that allele A actually causes the phenotype, is of course the finding that such studies are hypothesizing and are designed for. The second possibility is that allele A is in linkage disequilibrium with allele B which actually

causes the phenotype, meaning that Allele A is found along with Allele B (the phenotype-causing allele) more often than would be expected by chance.

The third possibility is that the association found is actually a product of *population admixture.* In the same paper, they explain: "In a mixed population, any trait present at a higher frequency in an ethnic group will show positive association with any allele that also happens to be more common in that group." (Lander and Schork 1994). The authors provide a theoretical example of a researcher studying a population in San Francisco who finds a certain allele of the HLA complex (*HLA-A1*) is associated with the ability to eat with chopsticks…which is simply because that allele is more common in Asians than Caucasians. The allele itself has absolutely nothing to do with the chopsticks ability trait, but is found to be associated based on population admixture.

The problems of population admixture do present themselves in real-life studies as well, and the authors note an example in Pima Amerindians. Researchers found an association between the *Gm* locus and type II diabetes. The version of the allele they found to be protective happened to be the one more frequent in Caucasians, and it turned out that Pima who had this allele also tended to have more Caucasian ancestry in general. Combine this knowledge with the fact that the Pima are more susceptible to type II diabetes in general, and we can see that the "protective" effects of the allele were much more likely found because of the protective effects of Caucasian ancestry, and not the allele itself (Lander and Schork 1994). Even among Caucasians, stratification based on ancestry could lead to false positive associations. One study demonstrated this for example when they found an association between the Lactase gene *LCT*

and height in a European American sample ($p<10^6$). However in this case again, the association was due to stratification: when they matched individuals based on specific European ancestry the association was greatly reduced, and entirely disappeared in Polish and Scandinavian individuals (Campbell, Ogburn et al. 2005). These real-life examples are much more subtle than the chopsticks example, and a fair warning to avoid population admixture in association studies whenever possible. If this step is not taken, then the results can often be misleading.

Heritability

Heritability (h^2) measures how much phenotypic variance is influenced by genotype. It is based on a simple model where Phenotype (P) is based on two basic variables: Genotype (G) and Environment (E). So while an association can be very useful in establishing which genotypes are related to any given phenotype, h^2 addresses how much of that phenotype's variance is explained by the variance in genes alone, excluding environmental factors. Basically:

Phenotype: P = Genotype (G) + Environment (E)

Heritability: h^2 = Var (G) / Var (P)

For example, speaking Serbian may run in a family but no one would suggest that there is a "speaking Serbian" gene. The phenotype of speaking Serbian would mainly be attributed to environment, and the heritability would be low. The heritability measure is crucial in separating out instances like these, of "traits" that simply run in families, from traits that are actually passed through families genetically.

In genetics, h^2 is calculated by looking at differences in the correlation of phenotype traits between monozygotic (identical) and dizygotic (fraternal) twins. The basic logic is that if

monozygotic twins are more similar than dizygotic twins, then the trait is considered heritable. For example, one calculates the correlation among monozygotic twins for a trait like height, and calculates another correlation among dizygotic twins. The difference between these correlations is doubled to account for the fact that monozygotic twins share 100% of their genes and dizygotic only 50%.

$$h^2 = 2(\text{monozygotic correlation} - \text{dizygotic correlation})$$

The logic is simple: dizygotic twins' phenotypes differ based on both genes AND environment, while two monozygotic twins will differ from each other based on environmental influence alone. So the difference between them (doubled) should represent how much the phenotype is driven by purely genetics.

Complex traits like depression and CVD have been widely studied and are considered heritable in both men (Lyons, Eisen et al. 1998),(Friedlander, Siscovick et al. 1998) and women (Kendler, Neale et al. 1992), (Austin, King et al. 1987). In men, heritability was found to be h^2=0.36 using DSM-III-R criteria for diagnosing major depression (Lyons, Eisen et al. 1998). Another twin study sought to find a genetic explanation for the higher incidence of depression among women than men, but concluded that h^2 is the same for men and women and equal to 0.39 (Kendler and Prescott 1999).

Researchers have also established that many traits which are themselves risk factors for CVD have high heritability (Austin, King et al. 1987). Those traits whose heritability was significant in Austin's 1987 analysis included HDL levels, LDL levels, triglyceride levels, and relative weight. More recent research has accepted that systolic and diastolic blood pressure

also show high rates of h^2 (0.52-0.66 for systolic and 0.44-0.66 for diastolic) (Evans, Van Baal et al. 2003). Such studies strongly suggest evidence for a genetic component to depression, and a genetic component to CVD.

Statistics of complex traits

Studying complex traits in humans through a genetic lens has become an increasingly statistical process. Statistical significance and the type of statistical modeling relevant to our study are described below.

Significance

Every scientific inquiry that seeks concrete results is based on a hypothesis. One might, for example, have a hypothesis that Allele Y is associated with depression. To test this statistically, we frame this hypothesis in terms of a "null" hypothesis (H0) and "alternative" hypothesis:

Null hypothesis: H_0 = Allele Y is not associated with depression (no effect)

Alternative: H_A = Allele Y is associated with depression

The null hypothesis, as noted above, implies that we are not seeing any effect. This is assumed to be true. In hypothesis testing, we test the likelihood of that our data could occur *given* that the null hypothesis is true. If we find that our result is very highly *unlikely* under the H_0 assumption, then we reject the null hypothesis and accept the alternate hypothesis. For example if Allele Y occurs with depression in 97% of our patients, we would be likely to reject the null hypothesis (given the study is large enough and the sample homogeneous enough).

Such a result, one that allows us to reject the null hypothesis, is said to be significant. The significance cut-off is usually at a "p-value" of 0.05. This means that we can expect our data to occur (given the null is true) only 5% of the time, purely by chance. Any p-value of 0.05 or lower is generally considered significant enough to reject the null hypothesis.

Many researchers conduct more than one hypothesis test in a single study. In these cases where multiple testing occurs, more stringent p-values must be used to ensure that each hypothesis is truly accurately accepted or rejected. If, for example, we were to conduct 100 hypothesis tests in a single study, one would *expect* that 5 of them would come up significant based on pure chance. Thus a study that includes multiple testing or multiple comparisons must apply some kind of correction to adjust the p-value that will be considered significant for each individual test. One of the most commonly used adjustments is called the Bonferroni Correction: divide the original p-value (α) by the number of test (n):

Bonferroni Correction: α / n

For example, if we conducted 100 tests, our new p-value should be 0.0005.

Modeling

Human geneticists use statistical modeling in order to describe the variation in a phenotype based on variables like genotype, environment, and interaction between the two, often with regression analysis. Because traits like depression and CVD are influenced by a constellation of genetic *and* environmental factors, we must be especially careful in making sure to include relevant covariates or identify confounding variables.

Confounders occur for example when a previously established association is later discovered to be almost entirely attributable to another variable altogether, previously overlooked, and when that variable is instead used as a covariate all of the significance found in the first association disappears. In this case, the second variable (a confounding variable) is said to have been "driving" the initial association. Besides confounding variables, interactions between variables themselves must also be considered. An example from genetics is the "Gene by Environment" (G X E) interaction. We know now that both genes and environment are important in determining traits and behaviors, and what is more, that they can interact in ways that produce different results than can be explained by the presence of either alone. One of the main candidate genes involved in our study is a perfect example of the G X E interaction. In a longitudinal study of a birth cohort of 1037 children (52% male), Caspi found that one allele (the S version of the well studied 5-HTTLPR) predicted higher rates of depression, but *only* in the presence of stressful life events (**Figure 2**). Without the environmental factor included, there seemed to be no contribution of the allele to higher rates of depression (Caspi, Sugden et al. 2003).

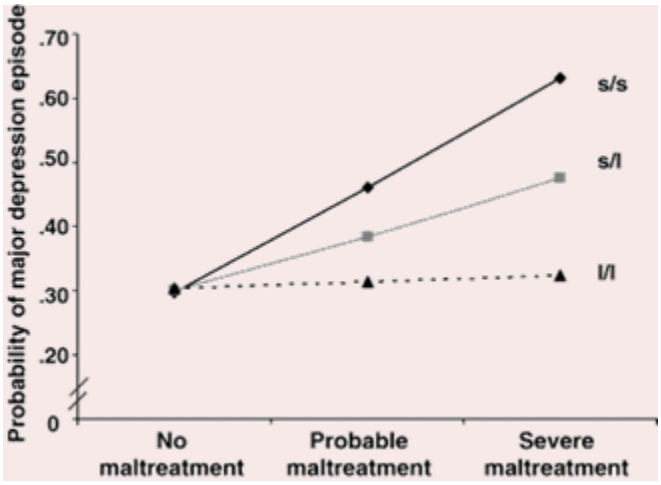


Figure 2. The S allele interacts with increasing instances of stressful life events, in this case childhood maltreatment, to increase the probability of developing major depression in those individuals.

Caspi et. al., *Science* 2003.

As complex traits, studying depression and CVD brings forth similar complications, and as will be clear in the following section, there are many hypotheses linking the two diseases and predictions about the underlying biological mechanisms behind these links.

1.3 Depression and Cardiovascular Disease Interaction

The study of depression and CVD proves challenging especially because the interactions between the two are almost impossible to tease apart. Does the instance of myocardial infarction spurn depressive moods and, in a way, lead to depression? And what about depression's role in developing CVD? A proposed set of mechanisms by which depression might lead to cardiac events is shown in **Figure 3** (adapted from a table in (Whooley 2006). But other studies show cardiac risk factors *leading* to depression as well! Metabolic Syndrome (MetS) is a good example of this type of link, since it is roughly defined as a cluster of risk factors for CVD and yet having the syndrome seems to predispose people for developing depression (Koponen, Jokelainen et al. 2008).

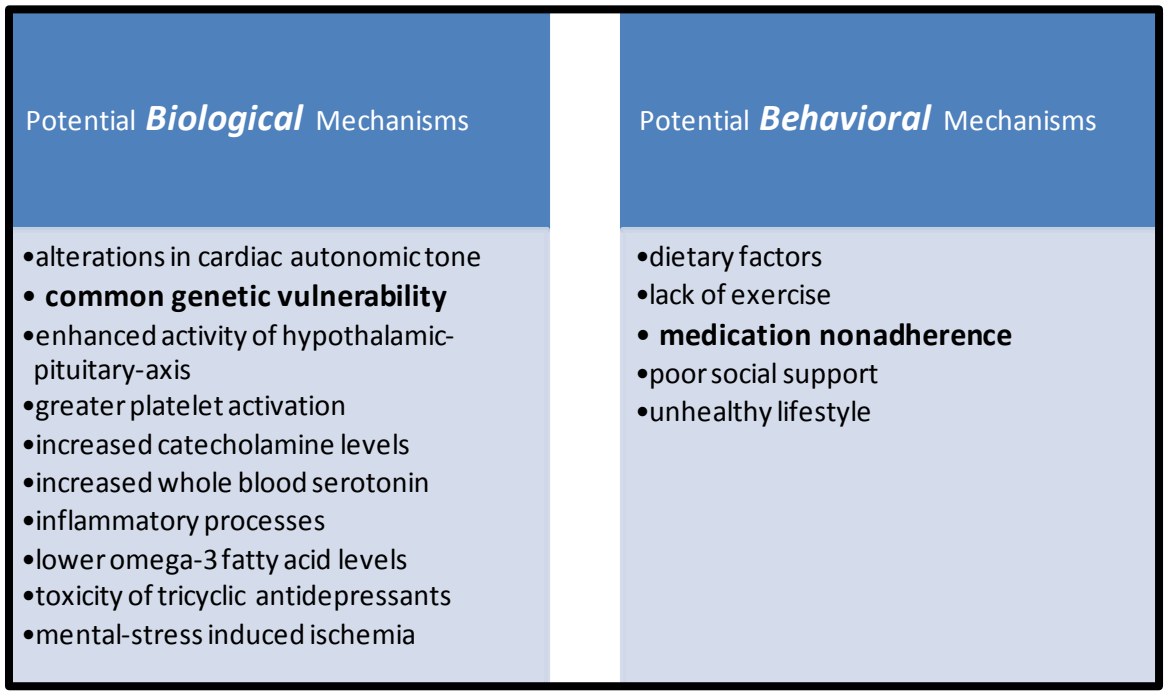


Figure 3. Mechanisms by which Depression may lead to Cardiac Events. Adapted directly from Whooley, 2006.

MetS has been "variously defined, but generally consists of 3 or more of the following components: hyperglycemia, hypertension, hypertriglyceridemia, low HDL, and increased abdominal circumference and/or BMI at >30"(Sutherland, McKinley et al. 2004).

Similar to its contribution to developing depression, there seem to be psychological traits that increase likelihood of MetS, such as depression, hostility, and anger (Goldbacher and Matthews 2007). This is why researchers interested in the biology of depression and CVD are interested in finding common mechanisms that might lead to both diseases. One of these proposed mechanisms involves an immune response mechanism involving platelet activation, cytokines, and the body's stress response.

Platelet Activation, Cytokines, and the Stress Response

Depressed patients are shown to have higher rates of platelet activation than controls (Nemeroff, Musselman et al. 1998) and similarly, depressive symptoms have been shown to predate and to worsen inflammation which is a risk factor for coronary artery disease(Stewart, Rand et al. 2009). Platelets (also called thrombocytes), are the essential factors for functional blood clotting, and while low levels of platelets can lead to excessive bleeding, heightened platelet activation and reactivity can cause blood clots and heart attacks. Platelets in turn can interact with leukocytes (white blood cells) and release factors such as cytokines, which are crucial in regulation of the body's inflammatory response. Both inflammation itself and higher platelet levels, then, can contribute to CVD risk. Research has verified this relationship, including one paper which reviewed the evidence for cytokines in acute heart failure, and noted

that higher pro-inflammatory cytokine levels definitely correlated with more severe symptoms of chronic heart failure and had more negative clinical outcomes (Chen, Assad-Kottner et al. 2008).

The neurotransmitter serotonin (5HT) is also involved in the cardio-depression link. Its link to depression is well known, and it is likely that even most non-scientists have seen advertisements for the popular pharmaceutical depression treatments called Selective Serotonin Reuptake Inhibitors (SSRIs), which directly block the Serotonin Transporter (5-HTT) from reabsorbing 5HT released at the synapse (the gap between two neurons). But 5HT has many roles, and researchers have yet to fully elucidate how they interact. However it is noteworthy to mention that "it has been clearly established that serotonin plays a significant role…in chronotropic and ionotropic effects in the cardiovascular system, and in platelet aggregation in the circulatory system" (Jonnakuty and Gragnoli 2008). Platelets, in fact, provide storage for circulating 5HT. When platelets interact with damaged tissue or several 5HT receptor agonists, they actually release 5HT and contribute part of the platelet aggregation response (Maurer-Spurej, Pittendreigh et al. 2004).

An environmental disruptor of pathways which normally attempt to maintain homeostasis and is involved in these two diseases is chronic stress. The normal stress response functions on a negative feedback system. Under the normal immune response, or during stress, cytokines act on receptors in the brain to cause cortisol release from the adrenal gland, which in turn suppresses cytokine release. In chronic stress situations this balance is tipped, and cortisol no longer acts to suppress cytokine release –leading to excess levels of cortisol and cytokines, and

eventually chronic inflammation: which also has been associated with MetS (Sutherland, McKinley et al. 2004). In this way, environmental stress interacts with the brain and body to increase risk for both depression and CVD (Mosovich, Boone et al. 2008).

Clinical Implications and Research

One reason to study the genetics of the two diseases is precisely because of mixed clinical results about treating depression to improve cardiovascular outcomes (both with medication and therapy). Two studies into this matter are the Sertraline AntiDepressant Heart Attack Trial (SADHART) and the Enhancing Recovery in Coronary Heart Disease (ENRICHD) trial.

SADHART focused on the Sertraline antidepressant (trade names are Zoloft, Lustral), which is a Selective Serotonin Reuptake Inhibitor (SSRI). The study confirmed that Sertraline was in fact a safe and effective antidepressant for cardiac patients with ischemic heart disease, but the study lacked the statistical power to detect any improvement in mortality rates. The ENRICHD trial, which assessed Cognitive-Behavioral Therapy (CBT), confirmed CBT as an effective depression treatment which improved quality of life but again did not find that it improved morbidity or mortality rates (Joynt and O'Connor 2005). Another study compared results of another SSRI called citalopram (trade names Celexa, Cipramil) and Interpersonal Psychotherapy (IPT) to reduce depressive symptoms in cardiac patients with major depression. Their findings suggest that citalopram did significantly improve depressive symptoms but IPT did not add any value to the clinical treatment (Lesperance, Frasure-Smith et al. 2007). It is difficult to draw any definite conclusions from the mixed results, and this has been our motivation to study the genetic predispositions which might potentially influence how patients

respond to therapy (or adhere to it) in order to better understand the cardio-depression interplay.

Perhaps some of the difficulty of studying these outcomes is due to individual differences in alleles for the 5-HT. As we saw in the Caspi paper noted above, there is individual variation for the transporter at which these antidepressants act, 5-HTTLPR. Another study on the same gene found that among patients who had had one acute myocardial infarction (heart attack), the S allele indeed predicted for future cardiac events (but only in the presence of depressive symptoms) (Nakatani, Sato et al. 2005). This may be partially explained by recent findings of "The Heart and Soul Study", which found that "among patients with chronic illness, carriers of the S allele of 5-HTTLPR are more vulnerable to depression, perceived stress, and high norepinephrine secretion, [which] may contribute to worse cardiovascular outcomes" (Otte, McCaffery et al. 2007). The basic difference, then, may very well come down to how individuals with the S allele react to environmental stressors. The following describes both the *5-HTTLPR* and *TPH2* genes in detail.

1.4 Candidate Genes in the Cardio-Depression Link

While many different genes are being studied in the link between Depression and CVD, our focus in this project so far has been on the *Serotonin Transporter Length Polymorphism* (*5-HTTLPR*) and on the *Tryptophan Hydroxylase-2* (*TPH2*). Both of these genes involved in the 5-HT pathway in our brains, which we have already learned is a good pathway to study for our purposes based on its role in depression and platelet activation. In fact 5-HT's role in the body extends to the regulation of other functions as well, such as temperature regulation, memory, sleep, appetite, emotions and wakefulness (Jacobs and Azmitia 1992).

Serotonin Transporter Length Polymorphism (*5-HTTLPR*)

Serotonin transmission occurs when 5-HT is released from a presynaptic neuron to a postsynaptic neuron across a gap called a synapse. Once the 5-HT molecules cross the synapse they bind at the postsynaptic neuron on 5-HT receptors, completing the signal between those two neurons. The 5-HT Transporter (5-HTT) is actually a transporter present on the presynaptic neuron which functions to reuptake 5-HT that has already been released. It serves a necessary function, which is clearing the synapse of 5-HT so that the postsynaptic receptors are not flooded constantly with released 5-HT (and the presynaptic neuron can soon fire another batch of 5-HT), thus allowing the nervous system better control of the neurotransmitter signaling process. It also serves to reuptake 5-HT molecules so they can be recycled or broken down and reused by the presynaptic neuron.

5-HTT-gene-linked-polymorphic region is on the 5-HTT (*SLC6A4*) gene on Chromosome 17 with two major alleles: Short (S) and Long (L). The length variation in these alleles is an example of a Variable Number Tandem Repeat (VNTR), where a sequence of nucleotides is repeated in a single allele of a gene. The L allele has 16 repeats (each 20-23 bp long) while the S allele has 14. The polymorphism in the promoter region of *SLC6A4*. The S allele results in less mRNA transcript and ultimately in fewer 5-HTT molecules present on that individual's neurons compared to the L allele (Canli and Lesch 2007), originally described in (Lesch, Bengel et al. 1996). This image from the Canli paper summarized *5-HTTLPR* and its effects (**Figure 4**).

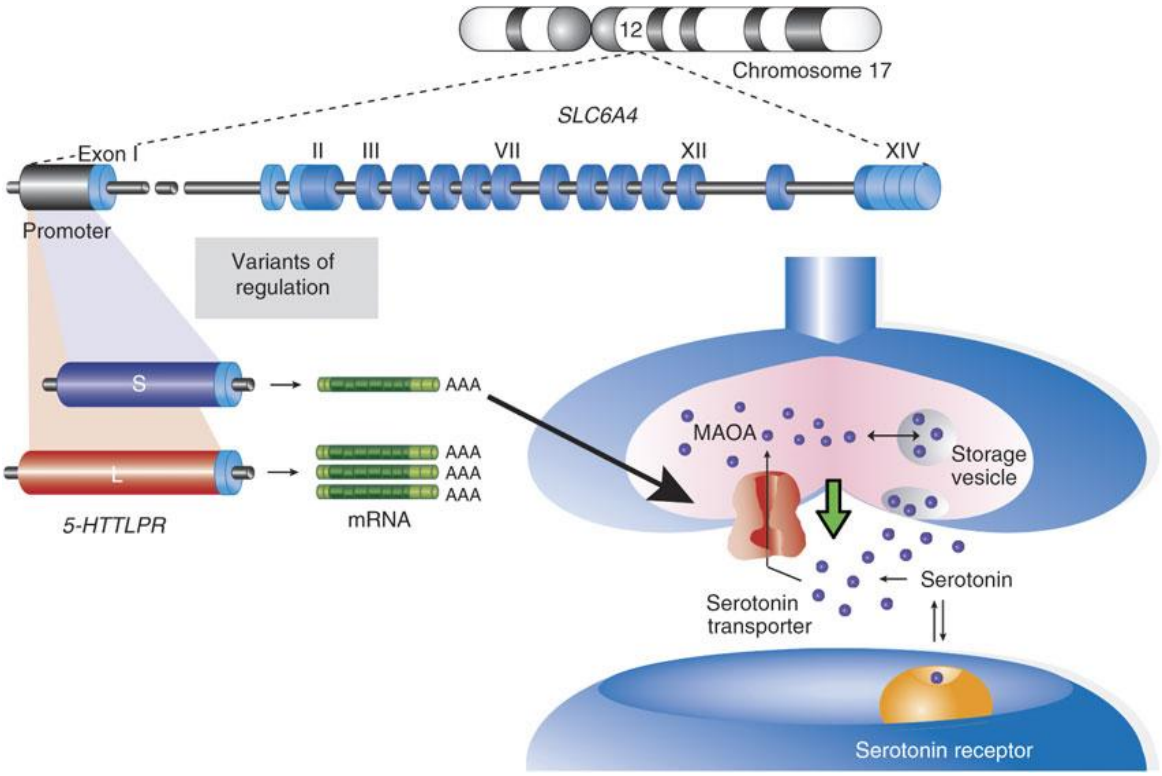


Figure 4. ***5-HTTLPR***.
Shown here: location of the polymorphism and that the S allele ends up leading to lower levels of mRNA than does the L allele. Ultimately, this translates to less 5-HTT protein on the presynaptic neuron in the case of the S allele as well.
From Canli and Lesch 2007.

In addition to the length polymorphism, the *5-HTTLPR* also contains a Single Nucleotide Polymorphism (SNP) on each version of the polymorphism. In this SNP, an A→G substitution creates a binding site for the Activator Protein-2 (AP2) transcription factor which actually suppresses transcription of the *5-HTT.* Thus an L allele with the A→G SNP (denoted L_G from now on, as opposed to L_A) will actually lead to transcription levels comparable to the S allele. Less is known about the transcription levels of the S_G version of the allele, though it is suspected that if anything it leads to even lower transcription levels than the S_A allele. **Figure 5** details the SNP's location on each allele.

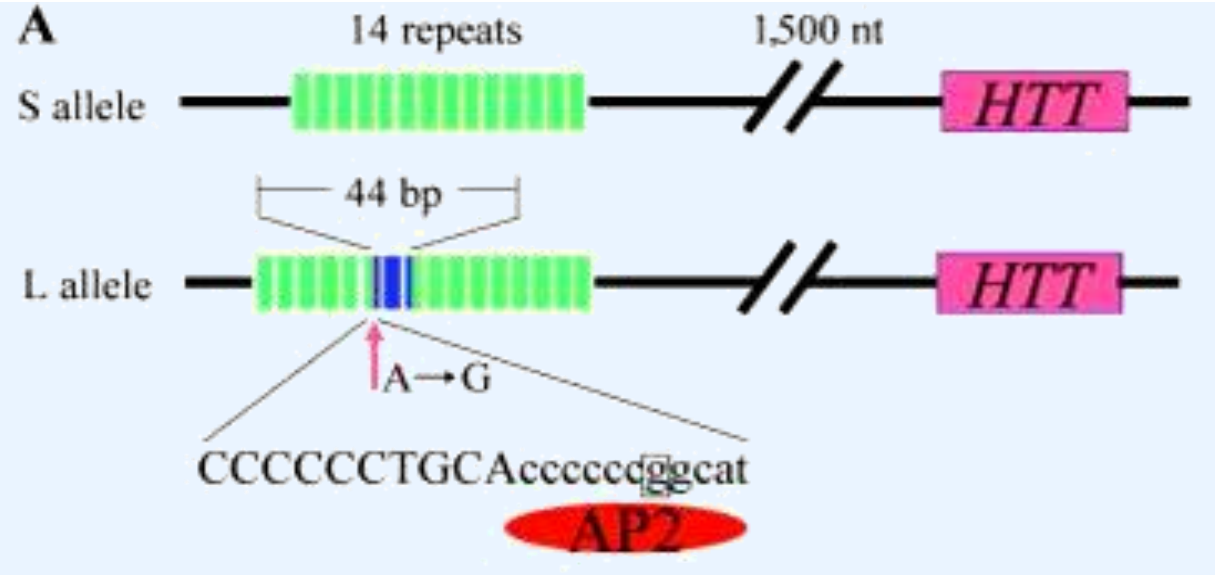


Figure 5. **A→ G SNP on *SCL6A4***
In the presence of the G nucleotide, an AP2 binding site is created, which suppresses transcription of the *SCL6A4* gene downstream of this promoter region.

Tryptophan Hydroxylase-2 [TPH2] Polymorphisms

This gene was chosen in part because of its central role in the pathway of Serotonin synthesis, and also for seeming connections between depression and CVD or CVD-like diseases. *TPH2* is the brain specific version of *TPH* and is the enzyme that facilitates the rate-limiting step in converting Tryptophan to Serotonin in brain. More importantly, "*TPH2* polymorphisms characterize a subgroup of depressed patients who are especially prone to develop metabolic disorders induced by genotype-dependent impairment of serotonergic transmission" (Kloiber, Kohli et al. 2009). The Metabolic Syndrome mentioned above (MetS), is actually a cluster of five risk factors closely tied to CVD: central obesity, high blood pressure, high triglycerides, low HDL-cholesterol, and insulin resistance. An individual who has three or more of the factors is said to have MetS. The syndrome itself is a risk factor for CVD and in fact "Because the U.S. population is aging, and because more than one half of adults are overweight or obese, *it has been estimated that metabolic syndrome soon will overtake cigarette smoking as the primary risk factor for cardiovascular disease*"(Deen 2004).

There are many SNPs on the *TPH2* gene, and they have been linked to various components of mental health such as suicidality, affective disorders, amygdala activation, and executive control. The Kloiber paper from 2009 finds the G allele of the rs17110690 SNP to be associated with MetS among depressed patients. Thus we were interested in adding analysis of this *TPH2* SNP to our study, especially given TPH2's role in 5-HT synthesis. Fortunately we already had data about MetS diagnosis and diabetes diagnosis in our clinical data collected from our subjects.

1.5 Molecular Genetics Tools

While details can be found in *Methods*, I would like to outline the workflow of how subjects are genotyped as it is essential to this thesis. Blood or saliva samples collected from patients were brought to the lab at which point DNA extraction and genotyping took place depending on the gene in question.

5-HTTLPR Genotyping: PCR, Electrophoresis, and Digestion

Genotyping of the Serotonin Transporter polymorphisms was optimized in the Burmeister Lab. After extraction of DNA from the blood or saliva samples and dilution to an appropriate concentration, the DNA is amplified via Polymerase Chain Reaction (PCR).

PCR is a method of amplifying DNA via thermal cycling. As double stranded DNA is heated, the two strands denature (or separate), exposing the "sticky" single nucleotides of each to the reaction mixture. In the reaction mixture specific primers, a DNA polymerase enzyme, and additional deoxynucleoside triphosphates (dNTPs) are added. The primers are short chains of DNA sequence that will anneal to specific segments of the genome as the temperature in the reaction chamber cools. They are designed to flank the 5' and 3' end of the region which we wish to amplify. These primers allow the DNA polymerase enzyme (in our case Taq polymerase) to bind and to start attaching additional dNTPs to the exposed nucleotides from the original DNA strand. In the first cycle, one double stranded DNA segment is replicated. The cycle is repeated over and over again, with the end result being many many copies of our genetic region of interest. A simple diagram of this process can be seen in **Figure 6**.

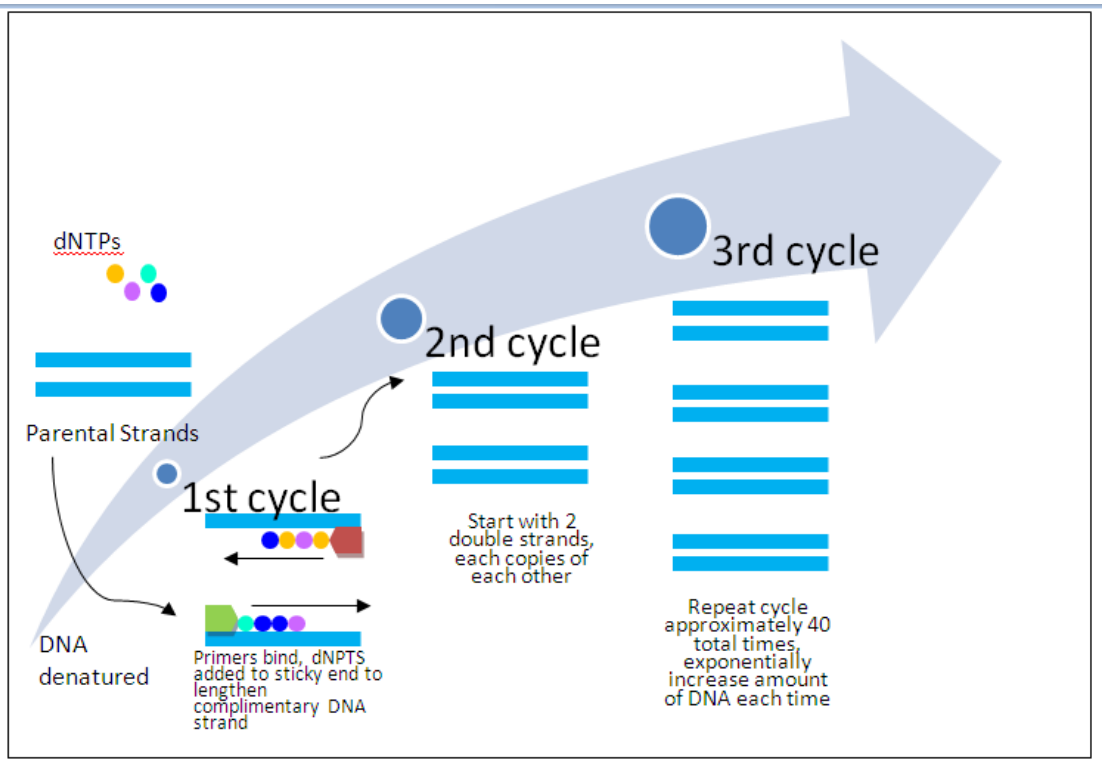
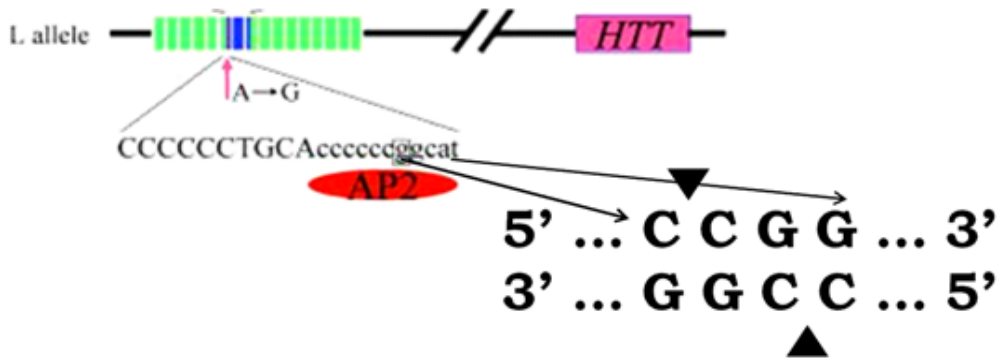


Figure 6. Polymerase Chain Reaction (PCR)
Each step in the process doubles the amount of product.

Once the 5-HTTLPR region has been amplified, we can now exploit the size difference between the L and S allele and the charged nature of DNA molecules to visualize which alleles any given individual has. An electrophoresis gel (solid but porous) is created using agarose, a gelatin-like substance, and DNA mixed with a dye is placed inside pre-formed wells in the gel. The gel is then placed in a gel box filled with buffer and through which electricity runs from one end of the gel to the other. The DNA fragments will flow with the electric charge towards the positive end of the gel box because the sugar-phosphate DNA backbone naturally renders them negatively charged. Since the S allele is smaller, it will migrate down the gel faster than the L allele. After sufficient time for separation of the various DNA strands, the gel can be visualized under UV light due to the ethidium bromide added to the gel. The ethidium bromide is an

intercalating agent (it inserts itself among the bases in DNA) with fluorescent properties, which is why it is useful for visualization. Distinction between L and S alleles is clearly visible and can be compared to a "DNA ladder" which has fragments of set base pair intervals: 100, 200, 300, 400, 500, etc.

After we have characterized whether each individual has the LL, LS, or SS genotype we must further characterize their genotype for the *5-HTTLPR* SNP. This is done using a digestion of the amplified DNA using a restriction enzyme. These enzymes, derived from bacteria where they are useful to recognize and cut foreign DNA, will recognize a specific sequence of nucleotides and perform a "cut" of the double stranded DNA, yielding two smaller pieces. In our study we chose the enzyme *Hpa*II which will specifically cut the L and S allele *but* only in the presence of the G allele of the rs25531 SNP (**Figure 7**). Because the cut produces different lengths of DNA segments, running an electrophoresis gel again after the product has been digested will allow us to easily determine the full genotype of an individual (L_GS_A for example). However, it is necessary to genotype for the length polymorphism only and then for the SNP in two separate steps because the L_G and S_G alleles are the same length.



Allele	Undigested fragment	visible digested fragment	Hidden fragment
L_A	512 bp	---	---
L_G	---	402	110
S_A	469	---	---
S_G	---	402	67

Figure 7. **Digestion with Hpa II (rs25531) Restriction Enzyme.**

The TPH2 Genotyping: rs17110690 with Taqman Assays

Quicker genotyping method developed by Applied Biosystems Inc [Foster City, CA] involves using chemical fluorescence to accurately quantify and characterize the genotype of a PCR product in a real-time PCR reaction. In addition to all the standard PCR reagents, two types of Taqman probe are added which correspond to each allele of the SNP. Each subject's sample, therefore, will have the potential to bind one, the other, or both probes depending on if that person's genotype is AA, GG, or AG for rs17110690. The probes themselves are composed of a

reporter fluorophore on the 5' end, and a quencher fluorophore on the 3' end (see **Figure 8**). The reporter fluorophore cannot fluoresce while it is still attached on the probe to its quencher fluorophore. However, during the PCR reaction when Taq polymerase encounters the probe, it will detach the reporter fluorophore from the probe. At this point, the reporter can fluoresce and be detected. As the reaction proceeds and PCR product is exponentially increasing, a machine such as the ABI PRIZM 7900HT Sequence Analyzer automatically detects the presence and amount of each fluorophore (VIC or FAM, corresponding to each allele) and generally makes a "call" for that individual's genotype: VIC, FAM, or BOTH.

The Taqman method is much faster (and costlier) than the PCR and electrophoresis method. Taqman can genotype hundreds of samples simultaneously in one cycle lasting about 4 hours. However it is currently used for SNPs and not repeats, which is why it was used for *TPH2* genotyping but not *5-HTTLPR*.

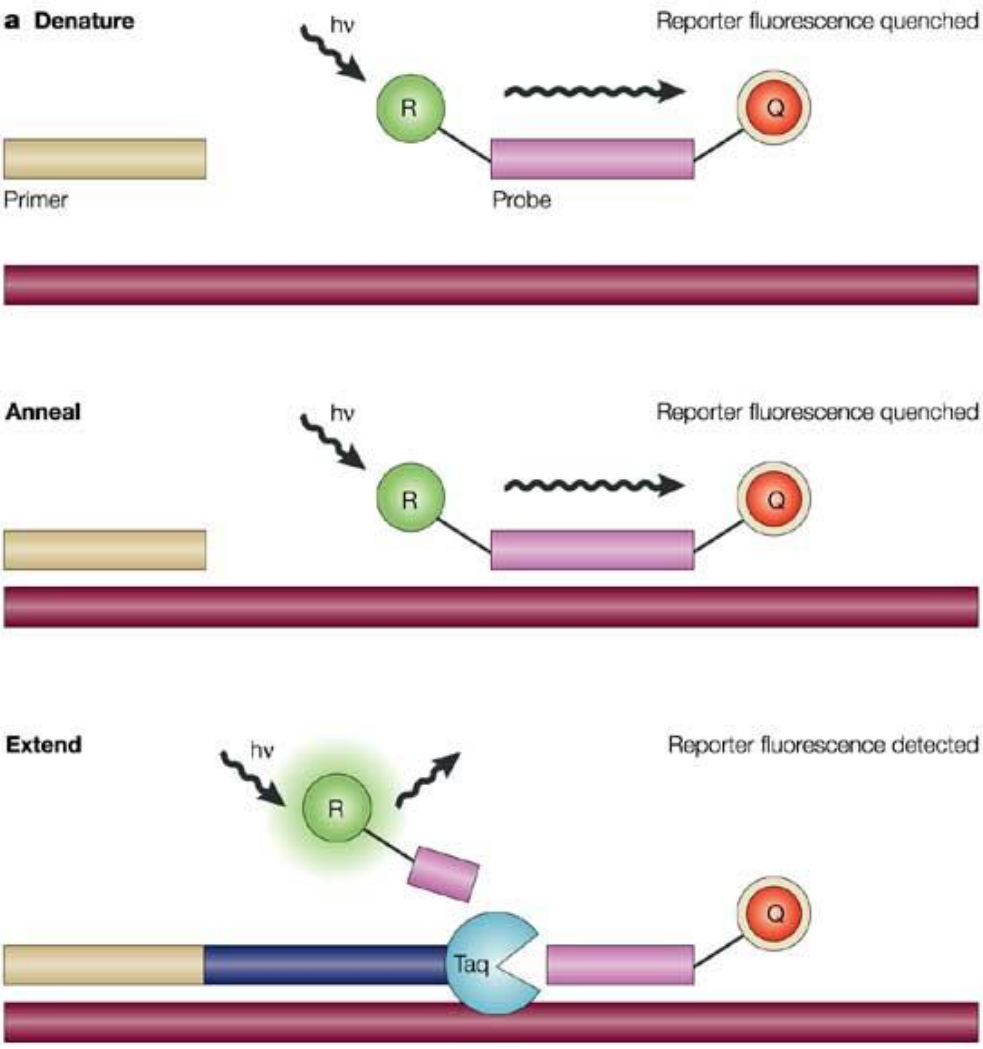


Figure 8. Taqman genotyping method
The Taqman method (named after the game "Pacman") relies on the basic principles of PCR with the addition of fluorescent probes. Each probe will attach to only one type of our alleles, and will only fluoresce when separated from the quencher fluorophore by the Taq polymerase molecule as it extends the complimentary DNA's strand. (Image from Koch, 2004.)

Other Data: Clinical Data and Surveys

In addition to the molecular data, clinical data and surveys are used. Details can be found in *Methods*, but the main surveys used are the Short Form 36 (SF-36) and Brief Symptom Inventory (BSI). Each of these surveys gives us a breadth of data from which we can characterize various phenotypes and endophenotypes. Endophenotypes are usually a more stable, measurable, and have a direct genetic association. For this reason, surveys collected are very detailed and in general do not simply ask "yes or no" questions about our complicated phenotypes like depression or CVD. Rather, we have data about various risk factors and components of each condition. This allows for much greater flexibility in our statistical analysis and generally much more reliable results.

In addition to the primary scales used mentioned above, other clinical questionnaires and health history data were collected as part of the cardiac rehabilitation program, much of which were pertinent to our study as well as potential confounders or co-variants. We had access to extensive information about factors like gender, race, and education level for example. Details about heart-disease related medications were taken, as well as past surgeries and injuries. Significant past and current medical problems were noted in the following categories: General Conditions, Heart/Circulation, Endocrine, Pulmonary, Gastrointestinal, Genitourinary, Bone and Joint, Neuropsychiatric, Hematologic, and Other. Gynecological (if female) and Family History was taken. Tobacco and Alcohol use was assessed (including questioning about types of tobacco, and attempts to quit or intentions to quit). Body weight and nutrition information was taken as well as extensive questioning about physical activity

habits and opinions. Information about “Well-being, Stress, and Emotional Factors” was taken primarily via the SF-36 and BSI questions described above, which were integrated within an overall questionnaire for the subject participants.

The program staff also measured participants’ health status via a number of clinical measures. These included data on resting pulse, resting blood pressure, fasting blood test results (choclesterol, triglycerides, LDL, HDL, glucose, hemoglobin A1c, homocysteine, and ALT), diagnosis status of various disorders such as Diabetes or Metabolic Syndrome, and the results of a Graded Exercise Test (GXT).

2 METHODS

2.1 Subjects

Subjects from our study were drawn from patients who participated in the cardiac rehabilitation program through University of Michigan Preventive Cardiology at the Cardiovascular Center, under the direction of Dr. Melvyn Rubenfire. Patients who were eligible had coronary heart disease (CAD) but *not* heart attack, coronary heart failure, or valvular heart disease. Informed consent was given and our study was approved by the University of Michigan Medical School Institutional Review Board. In addition to the clinical and psychological data collected for the cardiac center, data was collected using the surveys mentioned previously (the SF-36 and the BSI). Blood or saliva collection was also added as part of our study for extraction of DNA. The sample analyzed here includes clinical data from 245 subjects, though the study is ongoing.

2.2 DNA Extraction

DNA was extracted from blood samples using the Puregene protocol for extraction from whole blood (produced by Qiagen). For saliva samples, the Oragene® protocol for extraction from saliva was used (produced by DNA Genotek). At the end of DNA extraction, the concentration of DNA was measured (ng/µl) in each sample using the Nanodrop (ND-1000) Spectrophotometer.

2.3 *5-HTTLPR* Genotyping

We used PCR amplification followed by Electrophoresis and enzymatic digestion with *Hpa*II to characterize the 5-HTTLPR alleles and the rs25531 SNP. After measuring DNA concentration, each DNA sample was diluted to a concentration of 20 ng/ ng/µl. PCR was performed in a PTC 100 thermal cycler in a total volume of 30 µl for every sample: 60 ng (3 µl) of template DNA, 3 µl 10X SCB Buffer, 1.2 µl of 50 mM MgCl, 1.5 µl 1M Betaine, 1.5 µl DMSO, 0.3 µl *Taq* Polymerase, 3 µl of 2.5 mM dNPTs, 1.5 µl each of 5 mM primers HTT-R (5'-TGG GGG TTG CAG GGG AGA TCC TG -3') and HTT-F (5'-TCC TCC GCT TTG GCG CCT CTT CC -3'), and 13.5 µl of H_20.

The PCR reaction started with a 94°C for 5 minutes, then 40 cycles of: 30 seconds at 94°C, 60.2°C for 90 seconds, and 72°C for 1 minute. PCR cycling ended with 72°C step for 10 minutes and remained at 15°C until removal from thermal cycler. Samples were then run on a 3% agarose gel with 5 µl of ethidium bromide added. Visualization on the gel of the bands of L and S alleles was done by mixing 10 µl of the PCR product with a loading dye, under UV light.

Digestion of the rs25531 SNP was performed using *Hpa*II enzyme from New England BioLabs. Only samples that successfully amplified with PCR were digested. *Hpa*II cleaves at the recognition site 5'-C^CGG-3', which only occurs in the presence of the A→G SNP (otherwise the sequence remains 5'-CCAG-3'). As seen in **Figure 7** above, the presence of the G SNP allows for cleavage of the L allele (originally 512 bp long) into a 402 bp visible fragment (and 110 bp byproduct). Only the 402 bp segment is big enough to be characterized in our gel procedure. The S allele (originally 469 bp) is reduced to a 402 bp fragment (and 110 bp byproduct). Samples were digested in a total volume of 20 µl containing: 10 µl of DNA product from PCR, 2

µl of NEB 1 Buffer, 0.4 µl of *Hpa*II enzyme, and 7.6 µl H20. The reaction was performed in the same thermal cycler under the following conditions: 37°C for 4 hours, 65°C (enzyme deactivation) for 20 minutes, and 15°C until removal from thermal cycler.

2.4 *TPH2* Genotyping

Genotyping of TPH2 SNP rs17110690 was performed according to the Taqman method (Applied Biosystems, Inc [Foster City, CA]) described above and shown in Figure 8. The specific Taqman assay for our SNP (C_ _33094005 on ABI website) was used, along with a standard universal PCR master mix (also from ABI). The reactions were run in a 5 µl total volume per sample, which included 2.5 µl of 10X PCR master mix, 0.125 µl of a 40X primers and probes specific to our SNP, and about 40 ng of genomic DNA. PCR cycling and the fluorescence read was conducted in an ABI PRIZM 7900HT Sequence Analyzer under the following conditions: 95°C for 10 minutes, and 40 cycles consisting of 92°C for 15 seconds followed by 60°C for 1 minute. We used the program SDS 2.1 (ABI) for allelic determination, which included manual calls when clustering was hard to determine automatically .

2.5 Clinical Data and Surveys Collected

Subjects are from the University of Michigan Cardiac Rehabilitation Program at Preventive Cardiology Services at the Medical School, and are under the direction of Dr. Melvyn Rubenfire. Patients undergo 6-8 weeks of rehabilitation geared towards increasing their physical health via supervised exercise and helping them integrate exercise at home. Education is also provided about healthy eating, cardiac signs, and stress management. If the patients

have given consent to our part of the study, DNA is collected by blood sample or saliva sample at the end of their rehab. In addition, subjects participate in some additional surveys to give us observable data about mood, comorbid conditions, medications, and other variables that will be essential in our analysis.

Clinical data was taken from subjects upon admission to rehab, completion of rehab, and at 6 months and 12 months after that in the case of part of the surveys. Two questionnaires of clinical data are collected from the patients that are of specific interest in our study: the Short Form 36 (SF-36) and the Brief Symptom Inventory BSI (a subset of questions from the Symptom Checklist 90, SCL-90). The SF-36 was part of the follow-up surveys, but the BSI was only included in the initial baseline collection of clinical data. Our analysis includes SF-36 assessments primarily.

Short Form 36 (SF-36): Quality of Life Assessment, Our Depression Measure

The SF-36 is a quality of life assessment which asks questions about patients' emotions and health states over the past 4 weeks, and contains 36 questions broadly grouped into as pertaining to physical or mental well being, and into four categories within each for 8 total scales. The scales within Physical Health (Dimension A), are: (1) Physical Functioning –PF, (2) Role-Physical –RP, (3) Bodily Pain –BP, (4) General Health. Mental Health (Dimension B) Categories are: (5) Vitality – VT, (6) Social Functioning – SF, (7) Role-Emotional – RE, and (8) Mental-Health –MH. In fact ,scales 4 and 5 (General Health and Vitality respectively) are considered to overlap between the Physical and Mental Health dimensions. We have focused most on the Mental Health (MH) index, a subset of five questions about mental well being.

Patients report their feelings of **Happiness, Nervousness, Downheartedness, Down-in-the-Dumps-ness,** and **Calmness**. Questions are presented in the general format of: "How much of the time during the PAST FOUR WEEKS have you felt….." and a subject would circle from a range of: (1) All of the time (2) Most of the time (3) A good bit of the time (4) Some of the time (5) A little of the time and (6) None of the time. However, scores are inverted between emotions considered "negative"such as Nervousness so that in all cases, the highest score is considered the most desirable state for that particular mood score. Therefore in the case of nervousness for example a score of 6 would actually be *least* nervous. Because of this, we can average the 5 scores into a single variable **Mental Health** which we will use often in the analysis, and higher scores of Mental Health are consistent with averaging in the more desirable ranges for each mood score.

The Brief Symptom Index (BSI): Depression, Hostility, and Anxiety

During the baseline clinical evaluation subjects also answer a number of questions drawn from the BSI, which is a subset of questions the Symptom-Checklist-90 SCL-90 (a 90 question survey). The BSI in total consists of 53 self-reported questions that cover a range of 9 scales (Somatization, Obsessive-Compulsive, Interpersonal Sensitivity, Depression, Anxiety, Hostility, Phobic Anxiety, Paranoid Ideation, and Psychoticism) and ask the subject "How much has that problem distressed you during the PAST SEVEN DAYS including today". Subjects respond by circling a number from 0 to 4 ("Not at all" to "Extremely"). From here we are primarily interested in data about the patients' levels of Hostility, Depression, and Anxiety. However we chose not to use this in our analysis due to not having enough clinical data from these variables yet. It will probably be used in future analysis with this dataset.

Compliance Data

An initial evaluation included questions about difficulty taking prescribed medications. Follow-up health questionnaires included an entire section of questions labeled “Behavior Modification/Program Compliance Problems” which asked “Since your previous evaluation, have you experienced problems with any of the following:”

(1) Taking medications

(2) Exercise

(3) Quitting smoking

(4) Eating healthy

(5) Weight control

(6) Drinking alcohol

(7) Coping with stress

For consistency, compliance values from the first follow-up available for each subject were used due to high variability for the number and frequency of follow-ups at this point of the study. Compliance was converted into 1 binary score for each individual (0 = total compliance, 1 = at least one compliance issue).

2.6 Statistical Analysis

All statistical analysis was performed using SPSS 17.0.0 software package (released August 23, 2008). General trends were observed using plots of mean values and error bars representing +/- 1 standard error of the mean.

Relationship between Mental Health and Compliance were analyzed using Binary Logistic Regression with binary Compliance scores as the outcome variable. Tests assessing whether genetics had any influence on the change in mental health scores between initial baseline scores and follow-up scores were conducting using a general linear model for repeated measures with 2 levels (baseline and follow-up scores of mental health, happiness, or calm) and based on the 5-HTTLPR values or the SNP on *SCL6A4*. Covariates in the analysis were Gender, Ethnicity, and Antiplatelet medication use. The relationship between *TPH2* SNPs and MetS was analyzed using a Chi-Square test.

5-HTTLPR and the SNP

5-HTTLPR genotype was recorded first in the 3 categories mentioned before: LL, LS, or SS, and for analysis purposes coded as 1, 2, or 3 respectively. In analyzing the effect of the SNP, we recoded the full genotypes into three categories based on their functionality as referenced in the literature and earlier in this thesis. For example an L_G allele was treated as an S allele. Therefore Category 1, functionally similar to an LL genotype, included L_AL_A only. Category 2, similar to LS, included L_AL_G as well as L_AS_A and L_AS_G. Category 3, similar to SS, would include any L_GL_G, L_GS_G, L_GS_A, S_AS_A, S_AS_G, or S_GS_G genotypes:

1 - Equivalent to LL	2 – Equivalent to LS	3- Equivalent to SS
L_AL_A	L_AS_A	S_AS_A
	L_AS_G	S_AS_G
	L_AL_G	S_GS_G
		L_GL_G
		L_GS_G
		L_GS_A

3 RESULTS

3.1 Descriptives

Our subject pool contains 184 (75.1%) male and 61 (24.9%) female subjects. The age range of subjects is 34 years old to 94 years old, and age is normally distributed with a mean of 65.23 (s.d. 10.243, n=245). The predominant ethnicity was Caucasian (229 subjects, 94.6%), followed by African American (7 subjects, 2.9%), Asian (4 subjects, 1.7%), and Other (2 subjects, 0.8%). In the sample, 28 subjects had diagnosed MetS, 131 did not have diagnosed MetS, and 117 of the subjects had an unknown MetS status.

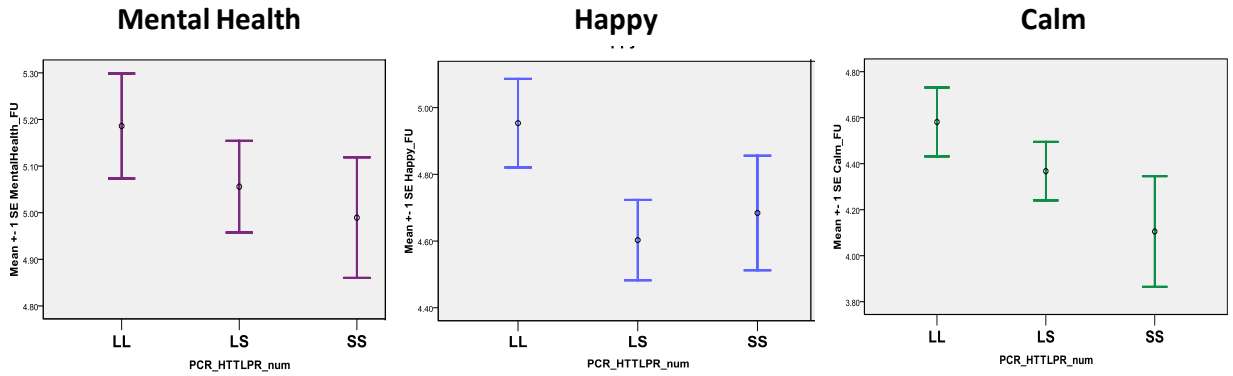
Allele Frequencies

For the *5-HTTLPR*, the allele frequency among Caucasians is 0.4 for the L allele and 0.6 for the S allele according to Hapmap, leading to expected genotype frequencies of: LL 0.36, LS 0.48, SS 0.16. In our sample among the Caucasian subjects the genotypic frequencies were: LL 0.325 , LS 0.528 , SS 0.147 (n=163). For *TPH2* rs2171363, the major allele is G with an expected frequency of 0.64 and the minor allele A has a 0.36 expected frequency (Hapmap). This would lead to expected frequencies of: GG 0.409 , AG 0.461, and AA 0.130. In our sample Caucasians have: GG 0.265, AG 0.565, AA 0.170 (n=147). For the *TPH2* rs17110690 SNP, allele frequencies among Caucasians from Hapmap are 0.8 for the G allele and 0.2 for A. Expected frequencies would be: GG 0.64, AG 0.32, and AA 0.04. Actual frequencies for Caucasians were very close: GG 0.614, AG 0.368, AA 0.018 (n=171).

3.2 Effects of 5-HTTLPR on Mood Scores

We took a preliminary look at how each score seemed to be affected by genotype based on error bar plots and determined that Happy, Calm, and Mental Health were potentially interesting to test for significance. A Bonferroni correction of 0.05/6 = 0.0083 was used for the new p-value acceptable for significance, given we had 6 SF-36 mood scores to start with.

I will start by showing the error bar plots for scores of Mental Health, Happy, and Calm at our first follow-up, stratified only by the *5-HTTLPR* genotype (Figure 9). Though none of the findings are significant, we are seeing some interesting trends. As expected, the LL genotypes tend to have the highest scores of Mental Health, Happy, and Calm. P values are: Mental Health p=0.291, Happy p=0.139, and Calm p=0.087.



p = 0.291 p = 0.139 p = 0.087

Figure 9: ***5-HTTLPR* and Mental Health, Happy, and Calm at Follow-Up**
P-values are listed based on linear regression with only HTTLPR genotype as dependent variable: Mental Health p=0.291, Happy p=0.139, and Calm p=0.087. Error Bars display mean values +/- 1 SE of the Mean

Next we analyzed whether the *5-HTTLPR* genotype affected how individuals' mood scores changed over time. We found no significance but we did find trends in the expected direction. Interestingly, while most subjects improved mood scores regardless of genotype at follow-up versus baseline (due to the positive effect of rehabilitation), those with the LL seemed to generally have higher mood scores and this stayed true after rehabilitation.

A general linear model with repeated measures based on two factors of Mental Health (at baseline and follow-up), with either *5-HTTLPR* genotype (**Figure 10**) or the functionally equivalent SNP genotype as described in *Methods* (**Figure 11**) as the independent variable and covariates of Gender and Ethnicity yields an insignificant result (*5-HTTLPR* p=0.568, SNP p=0.791) but an apparent trend if we diagram only Mental Health versus the genotype over time. Age and Gender were driving more of the effect in the model (Age p-value: 0.123, 0.127; Gender p-value: 0.247, 0.266).

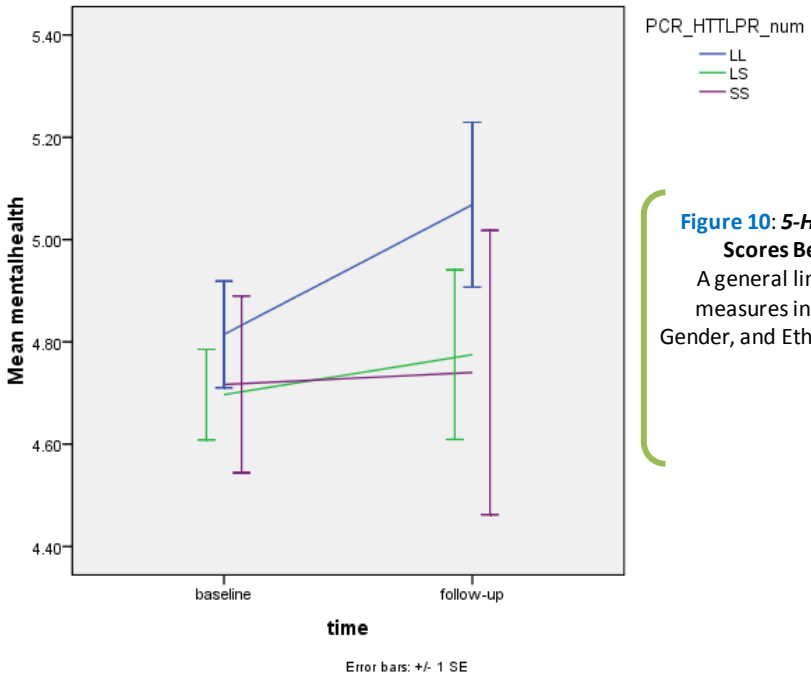
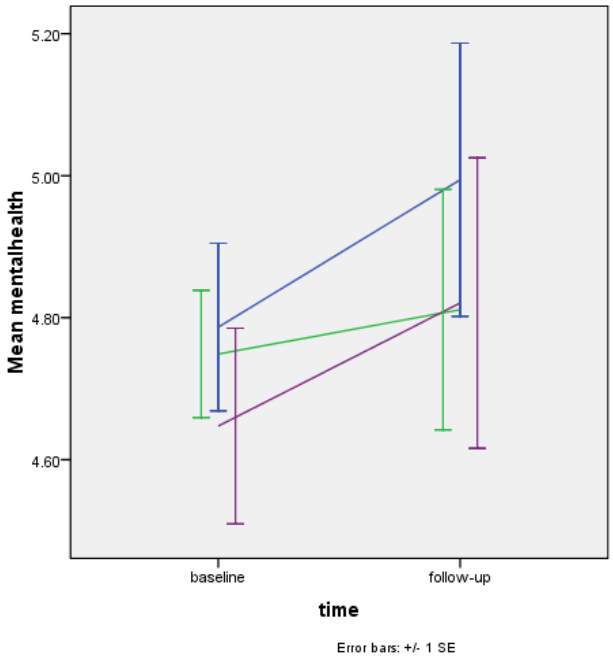


Figure 10: ***5-HTTLPR* and Mental Health Scores Before and After Rehab**
A general linear model with repeated measures including covariates of Age, Gender, and Ethnicity yielded no significance (p=0.568).



1 = equivalent to LL
2 = equivalent to LS
3 = equivalent to SS

Figure 11: *SLC6A4* SNP and Mental Health Scores Before and After Rehab
A general linear model with repeated measures including covariates of Age, Gender, and Ethnicity yielded no significance (p=0.791).

The statistical model verified that in our sample we do not have reason to believe there is an effect of the genotype on the changes in Mental Health Scores seen in rehabilitation. However in looking at the graphs we can see that there is some trend: individuals with the LL genotype, as expected, seem to have slightly higher Mental Health Scores (and perhaps even experience more improvement during rehabilitation). A bigger sample size might verify this.

Also worth noting is that there does not seem to be much of a difference between reporting the genotype without taking the SNP into account, and using the functionally equivalent SNP genotypes. In fact we often get a slightly clearer effect when we use the *5-HTTLPR* genotype rather than the SNP. I will continue to report both kinds of results for Happy and Calm scores.

For the same analysis with Happy scores, *5-HTTLPR* again was not significant though showed a clearer trend (p=0.151), while the functionally equivalent SNP genotype seemed less significant (p= 0.510). For Happy however, Age was a significant covariate (p = 0.031, 0.024). The graph of Happy scores at baseline and follow up are shown in **Figure 12** for *5-HTTLPR* and **Figure 13** for the SNP.

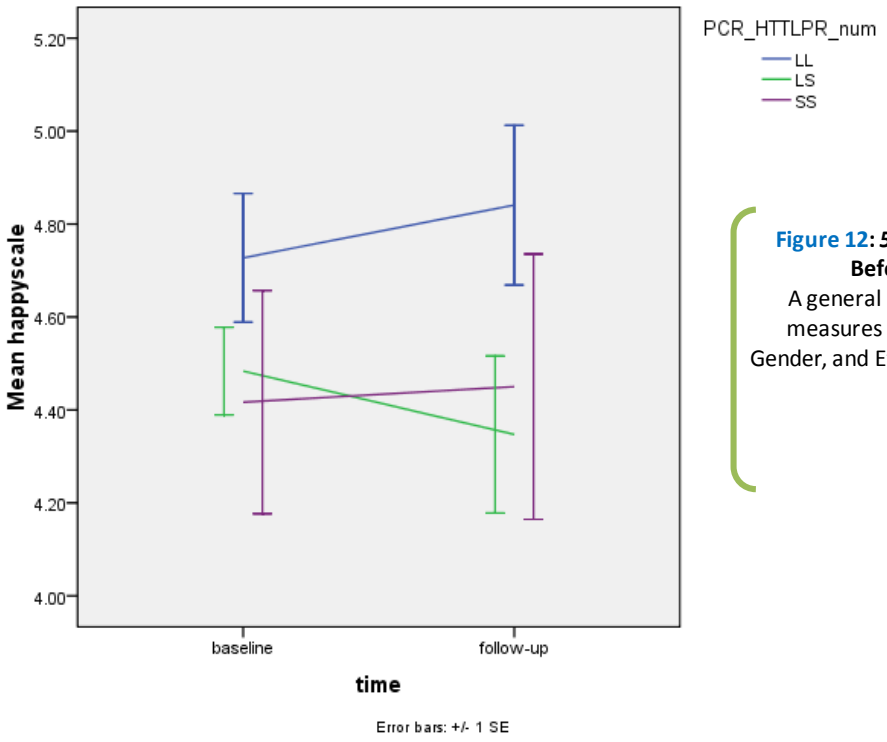


Figure 12: *5-HTTLPR* and Happy Scores Before and After Rehab
A general linear model with repeated measures including covariates of Age, Gender, and Ethnicity yielded no significance (p=0.151).

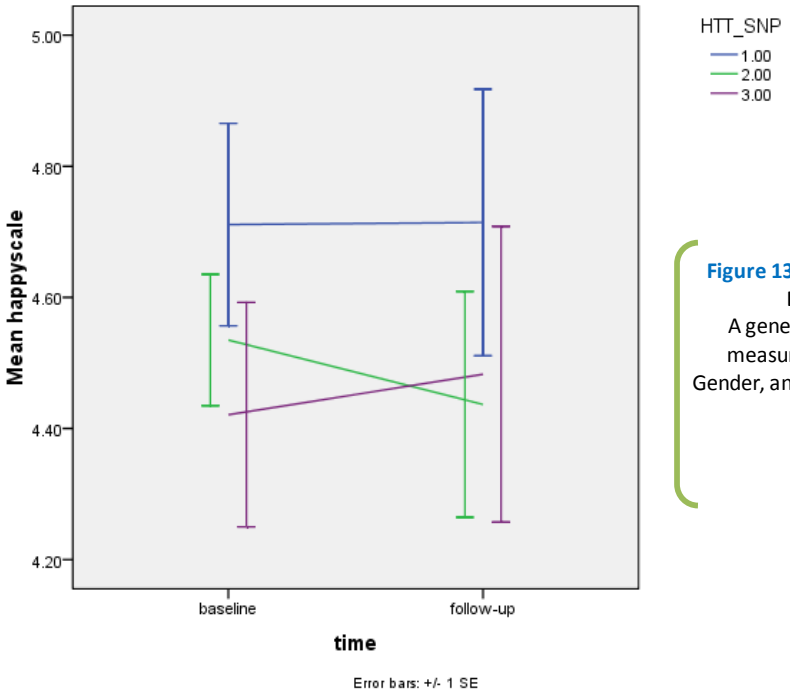


Figure 13: *SLC6A4* SNP and Happy Scores Before and After Rehab
A general linear model with repeated measures including covariates of Age, Gender, and Ethnicity yielded no significance (p=0.510).

For Calm scores, the 5-HTTLPR genotype was again not significant (p= 0.530, **Figure 14**) nor was that of the SNP (p=0.521, **Figure 15**). In this case again, Age was a significant covariate (p=0.007 and 0.008, respectively).

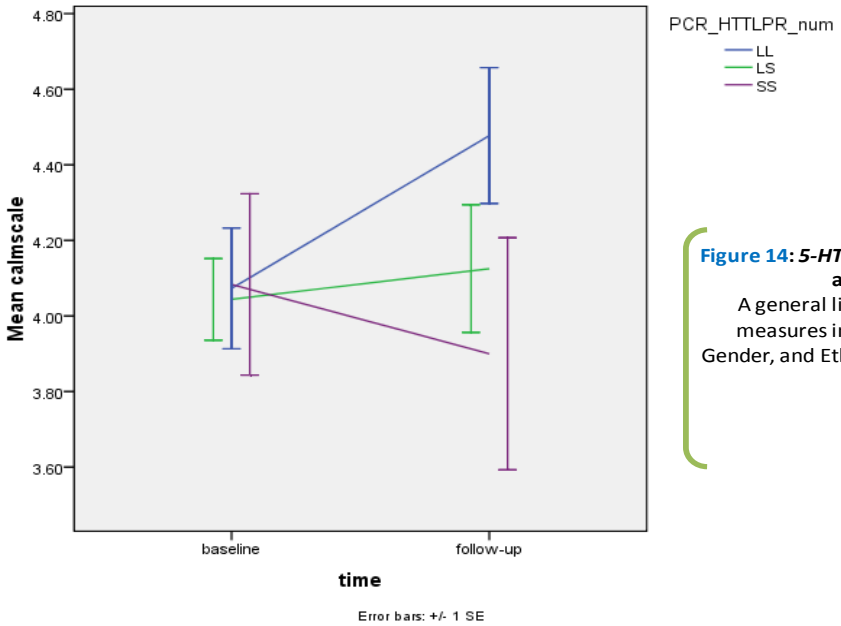


Figure 14: *5-HTTLPR* and Calm Scores Before and After Rehab
A general linear model with repeated measures including covariates of Age, Gender, and Ethnicity yielded no significance (p=0.530).

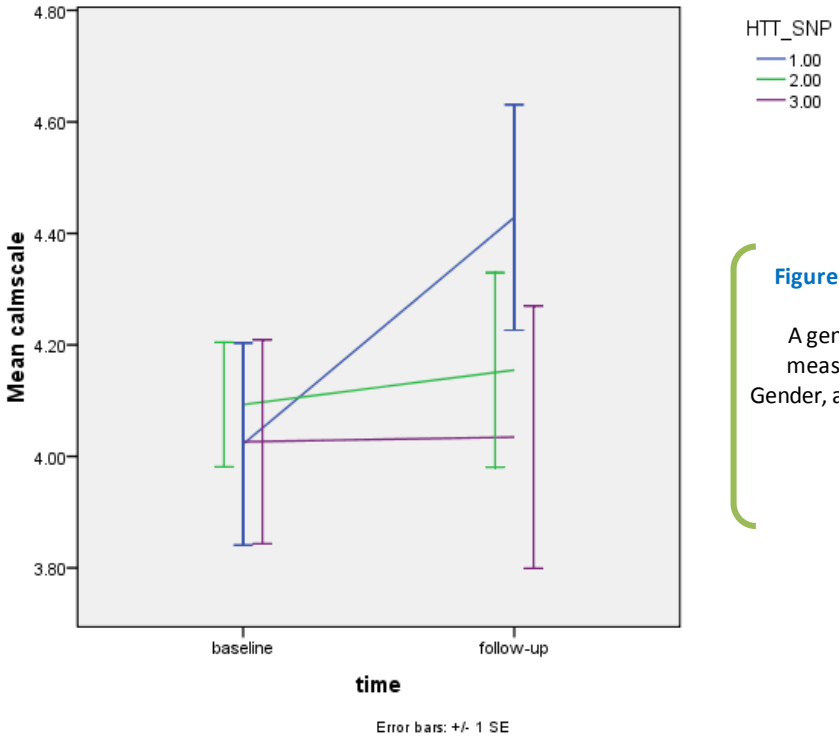


Figure 15: *SLC6A4* SNP and Calm scores Before and After Rehab
A general linear model with repeated measures including covariates of Age, Gender, and Ethnicity yielded no significance (p=0.521).

3.3 Mood Scores, Compliance, and 5-HTTLPR

Logistic Regression was performed with Compliance (1=yes, 2=no) as the binary outcome variable and scores of Mental Health, Happy, or Calm at follow-up as the independent variables, in addition to *5-HTTLPR* genotype. Age and Ethnicity were added as independent variables we well. At a cut-off of p=0.05, two of our mood scores proved significant and one shows a trend: Mental Health p=0.005, Happy p=0.01, and Calm p=0.071. Subjects who were complaint showed higher mental health scores. With our Bonferroni-corrected cut-off 0.0083, Mental Health score still remains significant. The genotype was not a significant factor in any of these. It seems that while Mental Health plays a big role in compliance, the genotype does not.

However, we did observe a similar trend as seen below for the Mental Health scores when stratified by Compliance score and genotype. It seems having the LL genotype and being Compliant is correlated with the Mental Health scores. The LS and SS genotype individuals that comply still seem to rank lower on Mental Health Scores. This is a trend worth exploring as well.

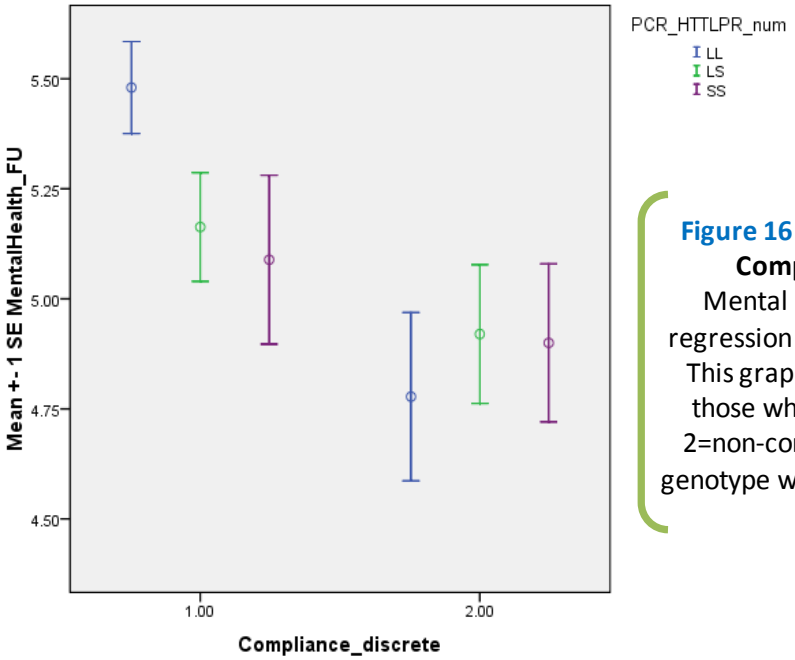


Figure 16 *5-HTTLPR* , Mental Health, and Compliance Scores after Rehab
Mental Health is significant by logistic regression (p=0.005) on Compliance scores. This graph shows a definite trend among those who are Compliant (1=compliant, 2=non-compliant) by genotype. However, genotype was not significant for compliance.

3.4 *TPH2* rs2171363 and rs17110690 and MetS

We cannot report significant association between MetS and either of the *TPH2* SNPs. Admittedly the sample size was small and that only 28 out of 276 subjects had diagnosed MetS. For rs2171363, a Chi-Square test for deviation from expected ratios of MetS per genotype yielded no significance or trend (p=0.803, n=104). The other SNP, rs17110690, was also not significant (p=0.511, n=122). This is why the *TPH2* SNPs have not been added in other parts of the analysis.

4. DISCUSSION

Allelic Frequencies and TPH2

Part of the difficulty of working with this data set has been the small sample size and in some cases, especially for the *TPH2* SNPS, a low minor allele frequency. Additionally, we have had trouble with the Taqman genotyping method failing to characterize all of the TPH2 SNPs. Troubleshooting is difficult in this high throughput method, and therefore a number of samples have been left undetermined. For the rs2171363 SNP, 117 of our samples were not genotyped successfully, which is 42.4 % of those attempted. Similarly though the rs17110690 SNP had close to expected frequencies after genotyping, we were unable to genotype 76 samples of the attempted samples, which was 27.5% of the total number. This may be at least a partial explanation for why our sample genotypic frequencies differed so much from expected values. In the future those samples need to be addressed individually: this would for example involve purifying the DNA samples we have that did not amplify properly in Taqman and retrying the Taqman again, or perhaps individually genotyping those samples with another method.

5-HTTLPR, Mental Health, and Compliance

Although we did not uncover any statistically significant associations between SF-36 scores at follow-up and the LL, LS, or SS genotype, we could see that there are some trends emerging. Without a larger amount of subjects, we cannot yet say we have verified or disproven the expected direction of the effect of genetics, which is that LL individuals would be more likely to have higher Mental Health scores in our sample. We do not have the number of subjects

needed to reduce the range of the confidence intervals quite yet to ensure that trends we are seeing are significant, but it is likely that the trends we are seeing will prove significant when more subjects are present in the analysis.

Our analysis reveals that people did in fact improve their Mental Health scores with rehabilitation (as expected). Unfortunately we could not uncover a true effect by genotype on the slope of that improvement. In other words, they all seemed to improve at the same rate, even though as we saw by the images in **Figure 9** it does seem that those with the LL genotype leave rehab with the highest Mental Health scores. Additionally, if we simply observe the apparent trends between baseline and follow-up as stratified by depression, there are a few cases where it appears the LL genotype has a greater slope of improvement in mood scores compared to LS and SS. This would echo previous literature findings such as those of Caspi mentioned in the introduction. In that case, he found that those with the S allele had greater risk of developing depression after the environmental effect of stressful life events (Caspi, Sugden et al. 2003). Here, we are seeing that a similar relationship to positive outcomes is seen with the LL individuals: with a positive environmental factor like rehabilitation, they seem to increase in mood scores more than the LS or SS individuals. We would hope with more subjects to verify the significance of this trend.

Even assuming the slope of improvement is the same, if LL individuals do in fact start and end with higher Mental Health scores as the trends suggest, this implies that even with the same rehabilitation treatment, LS or SS individuals do not achieve Mental Health scores as high as LL individuals on average. This is important because as noted, Mental Health was significant

at p=0.005 as a determinant of Compliance to the rehabilitation, and Compliance to parameters such as heart medications, diet, and exercise are important for cardiac outcomes. It is also important to remember the basics of what was mentioned in the introduction: depression itself is a risk factor for CVD, and makes the quality of life of those living with CVD worse if present. Interestingly, our studies' finding that Mental Health, Happy, and Calm are some of the most noteworthy scores in this analysis is consistent with a current paper that concluded that "positive affect was protective against 10-year incidence of [Coronary Heart Disease] CHD, suggesting that preventive strategies may be enhanced not only be reducing depressive symptoms but also by increasing positive affect" (Davidson, Mostofsky et al. 2010). This is consistent with the fact that we did not find Dumps, Nervous, or Downhearted to be of any predictive value in our sample.

Finally, in the general linear model with repeated measures we uncovered a significant relationship between Age and Happy, and Age and Calm. In both cases, Age was positively correlated with Happy or Calm scores. In other words, it appeared that being older actually contributed positively to the slope of improvement in mood scores seen from rehabilitation. This was not something we originally expected to have an effect, but is an interesting finding and should be taken into consideration by facilities administering rehabilitation programs.

Future Studies with This Data Set

In a future analysis of this data, it would be interesting to study the sustained improvements in Mental Health over time (or the decay in Mental Health, as the case may be). Genetics might have a role in this. On the other hand, perhaps everyone's Mental Health scores will decline

similarly over time after the end of rehabilitation. More likely, Compliance will play a role in how peoples' Mental Health scores decline over time. In our study, we have a high variability in amount of Compliance measures per individual. In general those who joined the study much earlier have had the chance to provide many more follow-up assessments than those who joined later. We chose to use only the first follow-up for each individual at this point, but when there is more longitudinal data following subjects after rehabilitation, this will be very interesting to explore.

Genetics may have a part to play in that Mental Health decay and perhaps even interacting with Compliance: as we saw there seems to be an interesting trend among individuals who show total Compliance and have the LL allele. Those individuals in our sample seem to have the highest Mental Health scores, even though across all subjects the genotype was not a significant factor.

Yet another possibility to explore given more longitudinal data from this sample is the effect of environmental stressors on mood scores, and possible interaction with genotype. As outlined in the introduction, *5-HTTLPR* seems to predispose people to have higher rates of depression in the presence of stressful live events. There are some data on major health events like heart attacks compiled along with the data in our study, and it would be interesting to see if we could replicate this effect by monitoring stressful life events in our subjects as well. In our case, we would also be able to investigate whether compliance or non-compliance changes the effect of genotype for individuals on average.

Conclusions

Our small sample size did not allow us to conclude any significant associations between *TPH2* SNPs and Mets. Similarly we could not conclude with significance that *5-HTTLPR* genotype had an effect on SF-36 mood scores. However, Mental Health was a significant factor, even in our small study, on scores of Compliance. Age proved to be a significant contributor to the degree of improvement in mood scores due to rehabilitation. Our general findings that only positive affect scores show trends of note are consistent with current research suggesting that positive affect may be important in predicting cardiovascular health, and not only negative affect (Davidson, Mostofsky et al. 2010). We also saw trends in *5-HTTLPR*'s effect on mood scores are consistent with other literature such as the Caspi paper showing environmental interaction with genotype on mood scores. In previous literature, the SS genotype was observed to increased risk for depression in the presence of stressful life events. Similarly in our study, it seems that LL increases the improvement in mood scores in the presence of positive life events such as rehab. We expect that with an increased sample size we could verify this environmental interaction between mood scores and genotype.

Works Cited

Austin, M. A., M. C. King, et al. (1987). "Risk factors for coronary heart disease in adult female twins. Genetic heritability and shared environmental influences." Am J Epidemiol **125**(2): 308-18.

Campbell, C. D., E. L. Ogburn, et al. (2005). "Demonstrating stratification in a European American population." Nat Genet **37**(8): 868-72.

Canli, T. and K. P. Lesch (2007). "Long story short: the serotonin transporter in emotion regulation and social cognition." Nat Neurosci **10**(9): 1103-9.

Cardon, L. R. and J. I. Bell (2001). "Association study designs for complex diseases." Nat Rev Genet **2**(2): 91-9.

Caspi, A., K. Sugden, et al. (2003). "Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene." Science **301**(5631): 386-9.

Chen, D., C. Assad-Kottner, et al. (2008). "Cytokines and acute heart failure." Crit Care Med **36**(1 Suppl): S9-16.

Davidson, K. W., E. Mostofsky, et al. (2010). "Don't worry, be happy: positive affect and reduced 10-year incident coronary heart disease: The Canadian Nova Scotia Health Survey." Eur Heart J.

Deen, D. (2004). "Metabolic syndrome: time for action." Am Fam Physician **69**(12): 2875-82.

DiMatteo, M. R., H. S. Lepper, et al. (2000). "Depression is a risk factor for noncompliance with medical treatment: meta-analysis of the effects of anxiety and depression on patient adherence." Arch Intern Med **160**(14): 2101-7.

Evans, A., G. C. Van Baal, et al. (2003). "The genetics of coronary heart disease: the contribution of twin studies." Twin Res **6**(5): 432-41.

Friedlander, Y., D. S. Siscovick, et al. (1998). "Family history as a risk factor for primary cardiac arrest." Circulation **97**(2): 155-60.

Goldbacher, E. M. and K. A. Matthews (2007). "Are psychological characteristics related to risk of the metabolic syndrome? A review of the literature." Ann Behav Med **34**(3): 240-52.

Halaris, A. (2009). "Comorbidity between depression and cardiovascular disease." Int Angiol **28**(2): 92-9.

Jacobs, B. L. and E. C. Azmitia (1992). "Structure and function of the brain serotonin system." Physiol Rev **72**(1): 165-229.

Jonnakuty, C. and C. Gragnoli (2008). "What do we know about serotonin?" J Cell Physiol **217**(2): 301-6.

Joynt, K. E. and C. M. O'Connor (2005). "Lessons from SADHART, ENRICHD, and other trials." Psychosom Med **67 Suppl 1**: S63-6.

Katon, W., E. H. Lin, et al. (2007). "The association of depression and anxiety with medical symptom burden in patients with chronic medical illness." Gen Hosp Psychiatry **29**(2): 147-55.

Kendler, K. S., M. C. Neale, et al. (1992). "A population-based twin study of major depression in women. The impact of varying definitions of illness." Arch Gen Psychiatry **49**(4): 257-66.

Kendler, K. S. and C. A. Prescott (1999). "A population-based twin study of lifetime major depression in men and women." Arch Gen Psychiatry **56**(1): 39-44.

Kloiber, S., M. A. Kohli, et al. (2009). "Variations in tryptophan hydroxylase 2 linked to decreased serotonergic activity are associated with elevated risk for metabolic syndrome in depression." Mol Psychiatry.

Koponen, H., J. Jokelainen, et al. (2008). "Metabolic syndrome predisposes to depressive symptoms: a population-based 7-year follow-up study." J Clin Psychiatry **69**(2): 178-82.

Lander, E. S. and N. J. Schork (1994). "Genetic dissection of complex traits." Science **265**(5181): 2037-48.
Lesch, K. P., D. Bengel, et al. (1996). "Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region." Science **274**(5292): 1527-31.
Lesperance, F., N. Frasure-Smith, et al. (2007). "Effects of citalopram and interpersonal psychotherapy on depression in patients with coronary artery disease: the Canadian Cardiac Randomized Evaluation of Antidepressant and Psychotherapy Efficacy (CREATE) trial." JAMA **297**(4): 367-79.
Lyons, M. J., S. A. Eisen, et al. (1998). "A registry-based twin study of depression in men." Arch Gen Psychiatry **55**(5): 468-72.
Maurer-Spurej, E., C. Pittendreigh, et al. (2004). "The influence of selective serotonin reuptake inhibitors on human platelet serotonin." Thromb Haemost **91**(1): 119-28.
Mosovich, S. A., R. T. Boone, et al. (2008). "New insights into the link between cardiovascular disease and depression." Int J Clin Pract **62**(3): 423-32.
Nakatani, D., H. Sato, et al. (2005). "Influence of serotonin transporter gene polymorphism on depressive symptoms and new cardiac events after acute myocardial infarction." Am Heart J **150**(4): 652-8.
Nemeroff, C. B., D. L. Musselman, et al. (1998). "Depression and cardiac disease." Depress Anxiety **8 Suppl 1**: 71-9.
Otte, C., J. McCaffery, et al. (2007). "Association of a serotonin transporter polymorphism (5-HTTLPR) with depression, perceived stress, and norepinephrine in patients with coronary disease: the Heart and Soul Study." Am J Psychiatry **164**(9): 1379-84.
Rugulies, R. (2002). "Depression as a predictor for coronary heart disease. a review and meta-analysis." Am J Prev Med **23**(1): 51-61.
Scherrer, J. F., H. Xian, et al. (2003). "A twin study of depression symptoms, hypertension, and heart disease in middle-aged men." Psychosom Med **65**(4): 548-57.
Schleifer, S. J., M. M. Macari-Hinson, et al. (1989). "The nature and course of depression following myocardial infarction." Arch Intern Med **149**(8): 1785-9.
Schmalz, D. L. (2010). "'I Feel Fat': Weight-Related Stigma, Body Esteem, and BMI as Predictors of Perceived Competence in Physical Activity." Obes Facts **3**(1): 15-21.
Stewart, J. C., K. L. Rand, et al. (2009). "A prospective evaluation of the directionality of the depression-inflammation relationship." Brain Behav Immun.
Sutherland, J. P., B. McKinley, et al. (2004). "The metabolic syndrome and inflammation." Metab Syndr Relat Disord **2**(2): 82-104.
Van der Kooy, K., H. van Hout, et al. (2007). "Depression and the risk for cardiovascular diseases: systematic review and meta analysis." Int J Geriatr Psychiatry **22**(7): 613-26.
Whooley, M. A. (2006). "Depression and cardiovascular disease: healing the broken-hearted." JAMA **295**(24): 2874-81.


MICHIGAN