UNIVERSITY OF CALIFORNIA, SAN DIEGO

# Search for an Association between Single Nucleotide Polymorphisms of *NPY2R* and Metabolic Syndrome Traits

A thesis submitted in partial satisfaction of the

requirements for the degree of Master of Science

in

Biology

by

Karthika Balasubramanian

Committee in charge:

Professor Daniel T. O'Connor, Chair
Professor Gen-Sheng Feng, Co-Chair
Professor Christopher Wills

2010

The thesis of Karthika Balasubramanian is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

---

---

Co-Chair

---

Chair

University of California, San Diego

2010

# DEDICATION

I would like to dedicate my Master of Science thesis to my mother, my father, and my sister. I would not be where I am today without all of the encouragement, help and support that I receive from them each and every day.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

First and foremost, I would like to acknowledge and thank my advisor Dr. Daniel O'Connor for accepting me as his student and for mentoring me with patience for the last two years. Dr. O'Connor has helped me gain more insight into the scientific process and enabled me to gain more knowledge and has inspired me to approach scientific research from a new perspective. It is because of the confidence that I gained under Dr. O'Connor's guidance that I feel prepared to meet the challenges that await me at UCLA's PhD program. I would also like to thank Dr. Pei-An Betty Shih for teaching me and assisting me with all of my research work. I would not have succeeded in completing my Master's in Science at UCSD without the advise, the support and mentoring they have offered to me.

ABSTRACT OF THE THESIS

Search for an Association between Single Nucleotide Polymorphisms of *NPY2R* and Metabolic Syndrome Traits

by

Karthika Balasubramanian

Master of Science in Biology

University of California, San Diego, 2010

Professor Daniel T. O'Connor, Chair

Professor Gen-Sheng Feng, Co-Chair

Metabolic syndrome, which characterizes an individual's risk for cardiovascular disease and type II diabetes mellitus, can be characterized by risk factors such as obesity and hypertension. These risk factors are predicted by traits like body mass index (BMI) and both systolic and diastolic blood pressure (SBP & DBP). Single nucleotide polymorphisms in the gene coding for the type 2 neuropeptide receptor (*NPY2R*) may affect BMI, SBP, and DBP since NPY2R and its ligand, peptide YY (PYY) are involved in the suppression of appetite. 703 individuals of Caucasian, Hispanic, and African-American biogeographical ancestry were genotyped for three candidate SNPs, rs6851222 (G-1606A), rs6857715 (C-599T), and rs1047214

(T+5895C), in the *NPY2R* promoter and open reading frame. No significant associations were found between these SNPs and BMI, SBP, and DBP ($p>0.05$). However, significant associations were found between haplotypes of these SNPs and these metabolic syndrome traits. In the aforementioned order of SNPs, haplotype GTT demonstrated a pleiotropic, positive directional effect on these three traits (BMI: $p=0.000375$, SBP: $p=0.0187$, and DBP: $p=0.0318$). Further study of this gene locus will help determine the relationship between these SNPs and the metabolic syndrome traits of body mass index, systolic blood pressure, and diastolic blood pressure.

.

## Introduction

Metabolic syndrome is characterized by a series of risk factors that contribute to both cardiovascular disease and type II diabetes mellitus[18]. Specifically, these risk factors are known as metabolic risk factors. Each metabolic risk factor is further defined by metabolic syndrome traits that contribute to the progression of the metabolic risk factor[19]. Examples of the metabolic risk factors include hypertension, obesity, atherosclerosis, and dyslipidaemia[11]. Some examples of the underlying risk factors include measurements of waist circumference, body mass index, blood pressure, triglycerides, high-density lipoproteins, and blood glucose levels[17]. Additionally, metabolic syndrome is also affected by other factors that include genetic and racial factors, aging, and endocrine disorders[18]. All of these conditions contribute to the progression of metabolic syndrome, and the metabolic syndrome traits may be affected by polymorphisms in the *NPY2R* gene.

The type 2 human neuropeptide Y receptor (Y2) is coded by the gene *NPY2R*. This gene is 8.5 kilo base pairs long, and it is located on the short arm of chromosome 4 (4q31)[1]. The actual receptor produced by this gene is 381 amino acids long with a mass of 42 kDa. Additionally, the Y2 receptor possesses 7-transmembrane domains that are characteristic of G-Protein coupled receptors[28]. These type two neuropeptide Y receptors are most commonly found on neuropeptide Y neurons in the hypothalamus as well as other regions of the brain, like the hippocampus[6,14]. Neuropeptide Y (NPY), peptide YY (PYY), and pancreatic polypeptide (PP) are all known agonists of the neuropeptide Y2 receptor, but PYY preferentially binds to

NPY2R[9,25]. All of these ligands are 36 amino acids long, possess a conserved tertiary structure, the pancreatic polypeptide (PP) fold, and are members of the polypeptide family of neuronal hormones[12,25].

The primary agonist of NPY2R, PYY, is named after the tyrosine residues found at both the C-terminus and the N-terminus of the full-length peptide molecule. The Y represents the single-letter amino acid code for tyrosine[20,26,32]. PYY is predominantly found circulating throughout the human body in two forms, the full length, 36 amino acid neurochemical and a truncated, 33 amino acid neurochemical. PYY is released into circulation from the gut after the consumption of meals like lunch and dinner[3,4,32]. Enteroendocrine L cells in the gut sense food and induce the production and release of PYY into circulation[26,32]. When this peptide binds to NPY2R, appetite is suppressed by increasing satiety in humans as well as other organisms, like dogs, mice and fish[16, 24,25]. Knockouts of PYY in mice results in an increase in weight gain due to uncontrolled consumption of food after both feeding and fasting[4]. Similarly, knocking out *NPY2R* in mice also results in uninhibited consumption of food[3,4]. Both PYY and NPY2R together are involved in regulating the ingestion of food, and the absence of this ligand and its receptor result in increased weight gain and obesity.

Since knockouts of PYY and its receptor, NPY2R, alter the physiological control of appetite, single nucleotide polymorphisms in the genes coding for PYY and NPY2R may yield similar results. In humans, SNPs in the PYY gene are significantly associated with body mass index and other metabolic syndrome traits[29]. It stands to

reason that SNPs in the *NPY2R* gene may also be associated with metabolic syndrome traits like body mass index and systolic blood pressure. This study aims to determine if variants in *NPY2R* and its 5' flanking regions are associated with the metabolic syndrome traits of body mass index and systolic blood pressure.

In *NPY2R*, several rare genetic variations have been found in the promoter, 5' untranslated region, and the open reading frame of the gene. Resequencing of this gene in 80 individuals of Caucasian, African-American, Hispanic, and Asian biogeographic ancestry yielded the presence of 10 common variants with a minor allele frequency greater than 5% and 11 rare variants with a minor allele frequency less than 5% (See Table 1 and Figure 1). Three candidate SNPs, rs6851222, rs6857715, and rs1047214, were chosen for genotyping based on their linkage disequilibrium with other common variants. Both rs6851222 and rs6857715 are found in the promoter region of the gene, and rs1047214 is found in the open reading frame of the second exon.

With rs6851222, the two variants of this SNP are A, the minor allele, and G, the major allele. This SNP may disrupt the binding of the transcription factor Spi-B, which is a member of the ETS family of transcription factors. This family of transcription factors is involved in effecting the gene expression of many biological processes like cellular growth and lymphoid development[34]. These transcription factors bind to a conserved purine rich binding motif with a GGA sequence[2], which is altered by this SNP. Spi-B is specifically involved in the development of mature B-cell lymphocytes and germinal center formation[35]. Further study of the effects of this

SNP will yield a greater understanding of the way in which the altered transcription factor binding motif affects the expression of *NPY2R*.

In addition to rs6851222, SNP rs6857715, which is also located in the promoter region of the gene, was genotyped. The major allele for this SNP was C, and the minor allele was T. In contrast to both of the promoter SNPs, rs1047214 is located in the open reading frame of exon 2. This SNP causes a silent mutation in codon 125 of the receptor changing an isoleucine into another isoleucine. However, previous studies have found associations between this SNP and metabolic syndrome traits, so further investigation of this genetic variant will increase the understanding of the role of this SNP in the development of metabolic syndrome.

## Materials and Methods

### *Subjects Genotyped*

Resequencing of the *NPY2R* gene was completed using the DNA of 80 individuals of the Caucasian, Hispanic, African-American, and Asian biogeographic ancestries. These individuals resided in urban areas of Southern California. All phenotypic data was self-reported by these individuals[27].

SNP genotyping was completed using samples from the UCSD Twin cohort. This cohort consists of 569 DNA samples from twins, siblings and pedigrees. The twins are both monozygotic twins and dizygotic twins. In many pairs of the monozygotic twins, the DNA of only one twin was available to genotype. Therefore, the genotype of the available twin's DNA was duplicated for the second twin in the pair. This brought the total sample size of individuals in the association study to n=703. The pedigrees are classified as siblings of twins and are members of the same families as the twins. The siblings belong to completely different families from the families of the twins.

These twins, siblings and pedigrees belong to three different biogeographic ancestries: Caucasian, Hispanic and African-American. A total of 146 monozygotic twin and 81 dizygotic twins were genotyped. Amongst the monozygotic twins, 123 pairs were Caucasian, 16 pairs were Hispanic, and 7 pairs were African-American. Amongst the dizygotic twins, 64 pairs were Caucasian, 9 pairs were Hispanic, and 8 pairs were African-American. Between both the monozygotic and dizygotic twin pairs, 92 males and 304 females were Caucasian, 17 males and 33 females were

Hispanic, and 11 males and 19 females were African-American. The zygosity of the twin pairs was self-reported and confirmed by sequencing. All subjects gave informed, written consent; the protocol was approved by the University of California at San Diego Human Research Protection Program[29].

***SNP Genotyping***

In order to determine whether an association exists between the SNPs of *NPY2R* and the metabolic syndrome traits of blood pressure and body mass index, three SNPs were genotyped using Real-Time Polymerase Chain Reaction by Applied Biosystems. The three SNPs that were genotyped contain the following rs numbers: rs6851222, rs6857715, and rs1047214. Rs6851222 and rs6857715 are located in the promoter region of the gene, and rs1047214 is located in the coding region of the gene.

The PCR for each of these SNPs was carried out in a microamp optical 384-well reaction plate which prevents well-to-well fluorescent contamination. The following protocol was used to genotype each of the above-mentioned SNPs using Real-Time PCR. Each well contained one microliter of human DNA. The DNA was dispensed into the well and dried down at the bottom of each well over night. After the DNA was dried, the following reagents were added to each well: 2.5 µL of Taqman Genotyping Master Mix, 0.25 µL of the SNP Genotyping Assay Mix and 2.25 µL of RNase free water. The main component of the Taqman master mix is the AmpliTaq Gold DNA Polymerase, which is inactive at room temperature and activated only at the high temperatures that are ideal for thermal cycling.

Each SNP was genotyped using a different assay with a different set of forward and reverse primers. The forward primer for rs6851222 was ACAAACAAAACAAAACAAAACAAAAACAACTACT and the reverse primer was ACCTAGAAACAAAGGTAAACAGAAATGGAA. The forward primer for rs6857715 was AGGATCTGAACTCGCTTTACCTTCT and the reverse primer was GTTTGGAGCACAGGGACCGCCCAGC. The forward primer for rs1047214 was GGGAGTATTCGCTGATTGAGATCAT and the reverse primer was CCGGACTTTGAGATTGTGGCCTGTA.

For each plate that underwent Real-Time PCR, the following thermal cycling conditions were utilized. The samples were first incubated at 95 °C for 10 minutes in order to activate the AmpliTaq Gold enzyme. Next, the DNA was denatured at 92 °C for 15 seconds. Lastly, annealing and extending of the DNA took place at 60 °C for 90 seconds. The denaturing, annealing, and extending of the DNA was repeated 39 times for a total of 40 cycles of PCR. The genotyping of the SNPs was completed by using the ABI 9700HT Thermal Cycler at the Salk Institute's microarray core facility.

***Analysis of SNP Genotyping Data***

After the thermal cycling of the DNA was completed, the genotype of each individual was determined using the SDS 2.0 software by Applied Biosystems. By utilizing the allelic discrimination function, the genotype of each individual was automatically called based on the fluorescence levels of the VIC and FAM reporter dyes (See Figure 2). For rs6851222, VIC represented the minor allele, A, and FAM represented the major allele, G. For rs6857715, VIC represented the major allele, C,

and FAM represented the minor allele, T. For rs1047214, VIC represented the minor allele, C, and FAM represented the major allele, T.

In order to verify that the genotypes were accurately assigned to each sample, the amplification curves of each sample were analyzed. The genotypes of samples that were homozygous for the minor allele for each SNP were confirmed by a strong binding of the VIC probe to the DNA samples (See Figure 3). The genotypes of the heterozygous individuals showed both a moderate binding to both the FAM and VIC probes with a slightly stronger binding to the FAM probe (See Figure 4). Samples that were identified as homozygous for the major allele were represented by a strong binding of the FAM probe to the DNA with a very weak affinity to the VIC probe (See Figure 5). In this way the accuracy of the automatic assignment of the genotypes to the samples was confirmed.

***Statistical Analysis of Genotyping Data***

After the genotype of each sample was determined, this data was statistically analyzed to determine the allelic frequencies of each SNP based on biogeographical ancestry; the dependence of the SNP genotypes on biogeographical ancestry; the genotype frequencies of the SNPs based on biogeographical ancestry; the phenotypic association between the SNPs and metabolic syndrome traits; and the haplotype association of these SNPs with metabolic syndrome traits. The allele frequencies and dependence of the genotypes of biogeographical ancestry were determined using Microsoft excel; the SNP Genotype frequencies were calculated using Statistical

Analysis System (SAS) software, and the phenotypic and haplotype associations were calculated using the R Project for Statistical Computing.

The allelic frequencies of each of these SNPs based on biogeographical ancestry were determined using a Microsoft Excel macro by Chris Carlson specific for calculating Hardy-Weinberg equilibrium. The raw, diploid genotype data was inputted into the macro, which subsequently calculated the frequency of each allele as well the significance of these alleles. P-values greater than 0.05 indicated that the allelic frequencies were in Hardy-Weinberg Equilibrium.

The dependence of the genotypes on biogeographical ancestry was determined using the chi-square test for independence in Microsoft Excel. The observed and expected frequencies of the three genotypes for each SNP (homozygotes for the major allele, heterozygotes, and homozygotes for the minor allele) and the three biogeographic ancestries were compared in a 3 x 3 contingency table. Using the chi-square function in Microsoft Excel, the independence of the genotypes and biogeographic ancestries was determined. P-values greater than 0.05 indicated that the two factors were independent.

The frequencies of the genotypes for each of the SNPs were determined using the phenotypic data for biogeographic ancestry and the raw genotyping results. This data was compared in SAS, and the genotype frequencies for Caucasians, Hispanics, and African-Americans were determined.

The phenotypic associations with BMI, SBP, and DBP were determined using one-way analysis of variance (ANOVA) the R Statistical Project. In order to account

for variation due to sex, age, and biogeographic ancestry, these three factors were included in the analysis as covariates. Additionally, the phenotypic data for BMI, SBP, and DBP were transformed using a natural log in order to normalize any skewed or pointed data. Since the samples genotyped included monozygotic twins, which possess duplicate genotypes, the one-way ANOVA was run using the generalized estimating equations (GEE) function. This adjusted the data for the duplicate genotypes that arose from monozygotic twins. Significant associations with BMI, SBP and DBP were determined by p-values less than 0.05.

Lastly, the haplotype associations were determined using haplo.glm, a function for determining linear regression based on the generalized linear model in the R Statistical Project. Linear regression models using the haplotypes as the independent variable and the metabolic syndrome traits as dependent variables were generated by comparing each haplotype to the haplotype present most frequently in the population of samples genotyped. Again, sex, age, and biogeographic ancestry were used as covariates in this analysis. Significant regression coefficients were determined p-values less than 0.05.

The determination of genotype frequencies based on biogeographic ancestry, the determination of phenotypic associations with the metabolic syndrome traits, and the determination of the haplotype associations with BMI, SBP, and DBP were all completed by Dr. Pei-An Betty Shih, PhD.

## Results

### *SNP Alleles are in Hardy-Weinberg Equilibrium*

Genotyping of the three SNPs, rs6851222, rs6857715, and rs1047214 yielded the incidence of the three diploid genotypes produced by each of these genetic variants. Polymorphisms in rs6851222 produced the three genotypes of AA, AG, and GG. Both the minor allele, A, and the major allele, G, exist in Hardy-Weinberg equilibrium for each of the biogeographic ancestries tested. Amongst Caucasians, the frequency of the minor allele, A, is 31%, and the frequency of the major allele, G, is 69%. Amongst Hispanics, the frequency of the minor allele, A, is 39%, and the frequency of the major allele, G, is 61%. Lastly, amongst African-Americans, the frequency of the minor allele, A, is 16%, and the frequency of the major allele, G, is 84%. All of these allele frequencies are maintaining Hardy-Weinberg equilibrium with $p>0.05$. (See Table 2)

In comparison to SNP rs6851222, rs6857715 produces the diploid genotypes of CC, CT, and TT. Based on the Caucasian reference group, the major allele in this population in the C allele, and the minor allele is the T allele. Amongst Caucasians, the frequency of the major allele, C, is 59%, and the frequency of the minor allele, T, is 41%. Amongst Hispanics, the frequency of the major allele, C, is 55%, and the frequency of the minor allele, T, is 45%. When the incidence of this SNP was examined amongst African-American, the major and minor allele changed. For African-Americans, the minor allele of this SNP is C, and the major allele is T. The frequency of the minor allele, C, is 32%, and the frequency of the major allele, T, is

68%. The best explanation for this deviation from the reference group, the Caucasians, is that genetic drift has led to a higher prevalence of the T allele amongst African-Americans. All of these allele frequencies are in Hardy-Weinberg equilibrium with $p>0.05$ (See Table 2).

Lastly, the allele frequencies of SNP rs1047214 were determined from the diploid genotypes of CC, CT, and TT. Using Caucasians as the reference group, C was designated the minor allele, and T was designated the major allele. Amongst Caucasians, the frequency of the minor allele, C, is 44%, and the frequency of the major allele, T, is 56%. Similar to the frequency of rs6857715 amongst African-Americans, there is a change in the minor and major allele within the Hispanic population. The major allele is C, and the minor allele is T for individuals of this biogeographic ancestry. Amongst Hispanics, the frequency of the major allele, C, is 61%, and the frequency of the minor allele, T, is 39%. Lastly, amongst African-Americans, the frequency of the minor allele, C, is 9%, and the frequency of the major allele, T, is 91%. Although there is a change in the designation of the major and minor alleles for Hispanic individuals, the allele frequencies for each biogeographic ancestry are all in Hardy-Weinberg equilibrium with $p>0.05$ (See Table 2).

***SNP Genotypes Dependent on Biogeographic Ancestry***

In order to determine if the frequencies of these three SNPs, rs6851222, rs6857715, and rs1047214, are independent of biogeographic ancestry, the chi-square test for independence was conducted. Comparisons of the observed frequencies and the expected frequencies of the genotypes AA, AG, and GG, which represent SNP

rs6851222, amongst individuals of Caucasian, Hispanic, and African-American origin demonstrated that these genotypes are dependent of biogeographic ancestry ($p=0.001276$) (See Table 3). Comparisons of the genotypes CC, CT, and TT, which represent SNP rs6857715, amongst individuals of the above-mentioned biogeographic ancestries, also showed that variants of this SNP are dependent on biogeographic ancestry ($p=4.83 \times 10^{-7}$) (See Table 4). Lastly, the genotypes of rs1047214, CC, CT, and TT, are all affected by biogeographic ancestry as well ($p=6.332 \times 10^{-21}$) (See Table 5). The chi-square analysis confirms that biogeographic ancestry does contribute to the genotype of an individual.

***Frequency of the SNPs Present in the Population***

Since the alleles of each SNP are present in Hardy-Weinberg equilibrium and the chi-square analysis demonstrated that the genotypes of each SNP are dependent on the biogeographic ancestry of the individual genotyped, the frequency of each of these genotypes based on biogeographic ancestry can be deduced. Looking at SNP rs6851222 and only Caucasians, 11.5% of the subjects were homozygous for the minor allele, AA, 39.96% of the subjects were heterozygous for the minor allele, AG, and 48.54% of the subjects were homozygous for the major allele, GG. Amongst Hispanics, 15.66% of the subjects were homozygous for the minor allele, AA, 45.78% of the subjects were heterozygous for the minor allele, AG, and 38.55% of the subjects were homozygous for the major allele, GG. Lastly, amongst African-Americans, 1.41% of the subjects were homozygous for the minor allele, AA, 29.58% of the subjects were heterozygous for the minor allele, AG, and 69.01% of the subjects were

homozygous for major allele, GG. All of these genotypic frequencies agree with the Hardy-Weinberg frequencies of the A and G alleles calculated above and confirm that the distribution of genotypes varies by biogeographic ancestry (See Table 6).

Next, the distribution of the genotypes of the second promoter SNP, rs6857715, also represented the Hardy-Weinberg frequencies of the major allele C and the minor allele T for each biogeographic ancestry. Amongst Caucasians, the reference group, 35.41% of the subjects were homozygous for the major allele, CC, 47.52% of the subjects were heterozygous for the major allele, CT, and 17.06% of the subjects were homozygous for the minor allele, TT. Amongst Hispanics, 29.41% of the subjects were homozygous for the major allele, CC, 50.59% of the subjects were heterozygous for the major allele, CT, and 20.00% of the subjects were homozygous for the minor allele, TT. As was seen in the allele frequencies, amongst African-Americans, there is a change in the minor and major allele. For individuals of this biogeographic ancestry, the minor allele is C, and the major allele is T. 5.97% of the subjects were homozygous for the minor allele, CC, 52.24% of the subjects were heterozygous for the minor allele, CT, and 41.79% of the subjects were homozygous for the major allele, TT. Although the minor and major allele deviates from that of the reference group amongst African-Americans, these values still agree with the allelic frequencies deduced above (See Table 7).

Finally, the frequency of the genotypes CC, CT, and TT were deduced for the SNP in the open reading frame of the gene, rs1047214. Amongst the reference group, Caucasians, the minor allele was C, and the major allele was T. 19.13% of the subjects

were homozygous for the minor allele, CC, 50.46% of the subjects were heterozygous for the minor allele, CT, and 30.42% of the subjects were homozygous for the major allele, TT. Like the change in minor allele and major allele seen above with SNP rs6857715, the major allele changes to C and the minor allele changes to T amongst Hispanic individuals. 37.35% of the subjects were homozygous for the major allele, CC, 48.19% of the subjects were heterozygous for the major allele, CT, and 14.46% of the subjects were homozygous for the minor allele, TT. This data agrees with the Hardy-Weinberg allelic frequency calculated above. Lastly, amongst African-Americans, 0% of the subjects were homozygous for the minor allele, CC, 18.31% of the subjects were heterozygous for the minor allele, CT, and 81.69% of the subjects were homozygous for the major allele, TT (See Table 8). These genotypic distributions confirm that the allelic frequencies of each SNP are true and that the genotypes vary by biogeographic ancestry (See Figure 6).

***Phenotypic Association between SNPs and Metabolic Syndrome Traits***

Once the distribution of the genotypes for each of the three SNPs were verified, the data was then statistically analyzed for a phenotypic association with the metabolic syndrome traits of body mass index (BMI), systolic blood pressure (SBP) and diastolic blood pressure (DBP). All of the phenotypic data for the metabolic syndrome traits were normalized using a natural logarithmic function and then analyzed using one-way analysis of variance. This analysis included all of the samples genotyped, so the analysis was adjusted for gene interactions based on sex, age, and

biogeographic ancestry. The following results were used to adjust the BMI, SBP, and DBP for variation due to biogeographic ancestry.

A total of 504 Caucasian individuals, 82 Hispanic individuals, and 81 African-American individuals were all genotyped. The average body mass index for each of these biogeographic ancestral groups in the above order was 25.14 $kg/m^2$, 27.70 $kg/m^2$, and 29.95 $kg/m^2$. A difference in these values does exist, so the adjustment for race in the ANOVA analysis accounts for this variation. The average systolic blood pressure for each biogeographic ancestral group in the above order was 130.15 mmHg, 128.97 mmHg, and 140.95 mmHg. The value for SBP amongst African-Americans is much higher than the SBP for both Caucasians and Hispanics, which could have an effect on the ANOVA analysis. Lastly, the average diastolic blood pressure for each of the biogeographical ancestral groups was 70.65 mmHg, 70.54 mmHg, and 77.28 mmHg. Again, the DBP for African-American individuals was nearly ten percent higher than the DBP for Caucasians and Hispanics, so the ANOVA analysis was adjusted to account for this variation.

Next, each SNP was analyzed for an association with each of the above mentioned metabolic syndrome traits. The first trait examined was the average systolic blood pressure. The average SBP was compared for the individuals with each of the three distinct genotypes, and the following results were obtained.

For SNP rs6851222, the average systolic blood pressure for individuals homozygous for the minor allele, AA, was 135.12 mmHg with a standard error of 1.02. The average SBP for individuals heterozygous for the minor allele, AG, was

134.29 mmHg with a standard error of 1.01. Lastly, the average SBP for individuals homozygous for the major allele, GG, was 134.03 mmHg with a standard error 1.01. Based on one-way ANOVA analysis of this data, the variances of the systolic blood pressure for each genotype occur at equal frequencies (p=0.9004) (See Table 9). Therefore, no significant association between rs6851222 and systolic blood pressure was found.

For the second promoter SNP, rs6857715, the average systolic blood pressure for individuals homozygous for the minor allele, CC, was 134.13 mmHg with a standard error of 1.01. The average SBP for individuals heterozygous for the minor allele, CT, was 134.26 mmHg with a standard error of 1.01. Lastly, the average SBP for individuals homozygous for the major allele, TT, was 134.14 mmHg with a standard error of 1.01. The one-way ANOVA analysis of the data demonstrated that the average systolic blood pressure was equal for each genotype (p=0.993), indicating that no association exists between rs6857715 and systolic blood pressure (See Table 10).

For the third and last SNP, rs1047214, the average systolic blood pressure for individuals homozygous for the minor allele, CC, was 134.63 mmHg with a standard error of 1.01. The average SBP for individuals heterozygous for the minor allele, CT, was 133.67 mmHg with a standard error of 1.01. Lastly, for individuals homozygous for the major allele, TT, the average SBP was 134.50 mmHg with a standard error of 1.01. The ANOVA analysis yielded the results that the average systolic blood pressure

was equal for each genotype (p=0.7615) (See Table 11). Overall, no significant associations were found between rs1047214 and systolic blood pressure.

In addition to studying possible associations between the SNPs and systolic blood pressure, the association between these SNPs and body mass index was also studied. For the first promoter SNP, rs6851222, the average BMI for individuals homozygous for the minor allele, AA, was 26.77 $kg/m^2$ with a standard error of 1.03. For individuals heterozygous for the minor allele, AG, the average BMI was 26.86 $kg/m^2$ and the standard error was 1.02. Lastly, the average BMI for individuals homozygous for the major allele, GG, was 27.42 $kg/m^2$ with a standard error of 1.02. Based on the ANOVA analysis of the genotyping data, the average body mass index was equal for all three of the genotypes (p=0.4891) (See Table 9). All in all, no significant association was found between the promoter SNP rs6851222 and body mass index.

Next, for the second promoter SNP, rs6857715, the average normalized BMI for individuals homozygous for the minor allele, CC, was 27.77 $kg/m^2$ with a standard error of 1.02. For individuals heterozygous for the minor allele, CT, the average BMI was 27.05 $kg/m^2$ with a standard error of 1.02. Lastly, for individuals homozygous for the major allele, TT, the average BMI was 26.88, and the standard error was 1.02. The ANOVA analysis of the data showed that the variances and subsequently the average body mass indexes for these three genotypes are the same (p=0.3716) (See Table 10). Altogether, no significant association was found between this SNP and body mass index.

Finally, for the third SNP, rs1047214, the average normalized BMI for individuals homozygous for the minor allele, CC, was 26.97 $kg/m^2$ with a standard error of 1.02. For individuals heterozygous for the minor allele, CT, the average BMI was 26.82 $kg/m^2$ with a standard error of 1.02. Lastly, for individuals homozygous for the major allele, TT, the average BMI was 27.54 with a standard error of 1.02. The ANOVA analysis of the phenotypic data demonstrated that the average body mass indexes were equal for all three genotypes (p=0.4836) (See Table 11). Overall, no association was found between SNP rs1047214 and body mass index.

ANOVA analysis was also completed in order to determine if an association exists between these SNPs and diastolic blood pressure. No statistically significant association was found between rs6851222 (p=0.25), rs6857715 (p=0.978), and rs1047214 (p=0.727) and diastolic blood pressure (data not shown). Overall, no phenotypic association was found between these three SNPs and the metabolic syndrome traits of body mass index, systolic blood pressure and diastolic blood pressure.

***Haplotype Association with Metabolic Syndrome Traits***

In addition to studying the phenotypic association between these SNPs, and metabolic syndrome traits, the association between haplotypes of *NPY2R* SNPs and the metabolic syndrome traits was also studied. A haplotype is a set of alleles that are inherited together on one chromosome of an individual's diploid set of genetic material. First, eight haplotypes for these three SNPs exist, but only six haplotypes are present at a recordable frequency. Each haplotype represents the three observed

polymorphisms in the following order: rs6851222, rs6857715, and rs1047214. The haplotype most prevalent in the population of samples genotyped is GCC, which was present in 31.23% of the samples. The next most frequent haplotype was GCT, which was present amongst 24.89% of the samples. ATT was the next most frequent haplotype with a prevalence of 20.78%. GTT was the fourth most frequent haplotype, and 11.08% of the samples possessed this haplotype. Haplotype ATC was present with a frequency of 9.38%, and haplotype GTC was present with a frequency of 2.32%. Lastly, the two least frequent haplotypes were ACC and ACT with a combined frequency of 0.31%. These frequencies represent the prevalence of each haplotype amongst all of the samples for all three biogeographical ancestries combined (See Figure 7 and Table 12).

Next, each of these haplotypes was analyzed for an association with the metabolic syndrome trait of body mass index. Analysis of this data using the generalized linear model compares each of the haplotypes to the most frequent haplotype, GCC. This results in linear regression models for each of the haplotypes except for the haplotype present most frequently in the population. For each of the subsequent haplotypes, the following regression coefficients were found. Haplotype GCT was marginally associated with BMI with a positive coefficient of regression, 0.02904, and $p=0.0481$. No association was found between haplotype ATT and BMI based on the coefficient of regression -0.00235 and $p=0.8710$. Haplotype GTT was significantly associated with BMI with the coefficient of regression 0.05835 and $p=0.000375$. Haplotype ATC was not significantly associated with BMI based on the

coefficient of regression 0.03011 and p=0.1700. Haplotype GTC was not associated with BMI based on the coefficient of regression -0.02463 and p=0.5220. Lastly the two rare haplotypes, ACC and ACT, were significantly associated with BMI with the coefficient of regression 0.51201 and p=0.00000004. This study demonstrates that four haplotypes, GCT, GTT, ACC, and ACT are all associated with body mass index (See Table 13).

In addition to analyzing the haplotype association with body mass index, the haplotype association with systolic blood pressure was also studied. In order to determine if an association exists between these haplotypes and systolic blood pressure, this data was also analyzed using the generalized linear model. For each of the subsequent haplotypes, the following regression coefficients were found. Haplotype GCT was not significantly associated with SBP with the coefficient of regression, 0.009235, and p=0.0807. No association was found between haplotype ATT and SBP based on the coefficient of regression 0.008855 and p=0.4050. Haplotype GTT was marginally associated with SBP with the coefficient of regression 0.009668 and p=0.0187. Haplotype ATC was not significantly associated with SBP based on the coefficient of regression 0.014456 and p=0.9510. Haplotype GTC was not associated with SBP based on the coefficient of regression 0.025308 and p=0.2110. Lastly the two rare haplotypes, ACC and ACT, were significantly associated with SBP with the coefficient of regression 0.070706 and p=0.00967. This study demonstrates that three haplotypes, GTT, ACC, and ACT, are all associated with systolic blood pressure (See Table 14).

Lastly, the haplotype association with the metabolic syndrome trait diastolic blood pressure was also examined. This haplotype data was also analyzed using the generalized linear model. For each of the subsequent haplotypes, the following regression coefficients were found. Haplotype GCT was not significantly associated with DBP with the coefficient of regression, -0.00571, and p=0.6020. No association was found between haplotype ATT and DBP based on the coefficient of regression 0.00862 and p=0.4190. Haplotype GTT was marginally associated with DBP with the coefficient of regression 0.02528 and p=0.0318. Haplotype ATC was not significantly associated with DBP based on the coefficient of regression -0.00918 and p=0.5890. Haplotype GTC was significantly associated with DBP based on the coefficient of regression -0.09739 and p=0.000632. Lastly the two rare haplotypes, ACC and ACT, were not associated with DBP with the coefficient of regression 0.14098 and p=0.121. This study demonstrates that two haplotypes, GTT and GTC, are all associated with diastolic blood pressure (See Table 15).

## Discussion

### *SNP Alleles are in Hardy-Weinberg Equilibrium*

In genetics, determining whether alleles are in Hardy-Weinberg equilibrium is an important step necessary for the completion of statistical analysis of genotyping data. Knowing that alleles exist in Hardy-Weinberg equilibrium ensures that the data collected is accurate. In this study, the alleles of each of the SNPs, rs6851222, rs6857715, and rs1047214, tested positive for Hardy-Weinberg equilibrium with $p>0.05$ for each biogeographic ancestry. This verifies that genotyping data can be accurately analyzed for phenotypic and haplotypic associations with the metabolic syndrome traits of body mass index, systolic blood pressure, and diastolic blood pressure.

### *SNP Genotypes Dependent on Biogeographic Ancestry*

Similar to the tests for Hardy-Weinberg equilibrium, the chi-square test for independence must be completed in order to determine if the biogeographic ancestry of an individual may alter the results of a phenotypic and haplotypic association study. Since the genotypes of each SNP were dependent on the biogeographic ancestry of the individuals genotyped, this variability must be accounted for during statistical analyses that combine the data of all three biogeographic races. When the one-way analysis of variance was used to determine if an association exists between the SNPs and metabolic syndrome traits, the data was adjusted for the biogeographic ancestry. In this way, the dependence of the genotypes on race does not alter the results of the phenotypic association study.

### *Frequency of the SNPs Present in the Population*

Based on the statistical analysis of the frequencies of the single nucleotide polymorphisms of *NPY2R* in both the promoter and exon region of the gene, the prevalence of each diploid genotype for each SNP was determined. For the promoter SNP rs6851222, amongst the entire sample size of n=702 individuals, the majority of the subjects were homozygous for the major allele, GG, which confirms the calculated frequency of the major allele by Hardy-Weinberg. When this SNP is examined based on race, the majority of both Caucasians and African-Americans are homozygous for the major allele, but the majority of Hispanics possess the heterozygous genotype. This can be attributed to the overall higher frequency of the minor allele amongst Hispanics in comparison to the lower frequencies discovered in both Caucasians and African-Americans. Since all of the genotypes are dependent on the biogeographic, these results are consistent with the results of the chi-square test for independence.

While the pattern of inheritance of SNP rs6851222 was similar for the three biogeographic ancestral groups studied, the genotype frequency of SNP rs6857715 varied amongst Caucasians, Hispanics and African-Americans. In general, the heterozygous genotype was the most prevalent in the population of samples genotyped, but the allele frequencies did vary between these biogeographical ancestral groups. Between both Caucasians and Hispanics, the major allele for this SNP was C, and the minor allele was T. However, the major allele was T, and the minor allele was C for African-Americans. This demonstrated that different genotype frequencies will be observed for individuals of varying biogeographic ancestry.

When the genotyping data for SNP rs1047214 was analyzed, an effect similar to the genotype distribution of rs6857715 was observed as well. Overall the majority of the subjects were heterozygous for the two alleles, but the biogeographic ancestral data showed that Caucasians, Hispanics, and African-Americans all inherited these alleles in different ways. Amongst both Caucasians and African-Americans, the major allele was the T allele, and the minor allele was the C allele. However, for Hispanics, the major allele was C, and the minor allele was T. This difference in minor and major allele can be attributed to the dependence of the genotypes on biogeographical ancestry. In addition to the changes in the minor and major alleles, variations in the most frequent genotype for Caucasians, Hispanics and African-Americans existed as well. While both the majority of Caucasians and Hispanics were heterozygous for these two alleles, the majority of the African-Americans were homozygous for the major allele. This coincides with the very low allele frequency of the minor allele, C. This very low allele frequency is valid for this group of individuals because the Hardy-Weinberg equilibrium confirms that the data is accurate. All together, variation existed amongst the frequency of the genotypes for rs1047214 because the genotypes are dependent of biogeographical ancestry.

Although the chi-square test for independence indicates that the genotypes of the test subjects are dependent on biogeographic ancestry, the differences in the SNP frequencies can also be attributed to the relatedness of the samples tested. All of the samples genotyped originated from families of monozygotic twins, dizygotic twins, siblings, and pedigrees. Since all of these individuals are related, the results will be

biased towards the familial inheritance of the alleles. To better characterize the frequency of these polymorphisms in humans, a more diverse range of unrelated samples should be genotyped. This will help determine the true frequencies of these SNPs amongst individuals of varying biogeographical ancestries.

***Phenotypic Association between SNPs and Metabolic Syndrome Traits***

After determining the frequency of these *NPY2R* polymorphisms based on biogeographical ancestry, statistical analysis of the genotyping data was completed in order to determine if a phenotypic association. The one-way analysis of variance indicated that no significant association exists between these three SNPs and the metabolic syndrome traits of body mass index, systolic blood pressure, and diastolic blood pressure. This result agrees with prior association studies that examined the relationship between polymorphic regions of *NPY2R* and obesity, a metabolic syndrome risk factor. In this study, six *NPY2R* SNPs, including the coding region SNP rs1047214, were genotyped in white Danish subjects. This research yielded the results that this SNP specifically was not associated with metabolic syndrome traits like body mass index and waist circumference[31]. Additionally, in another research study, the relationship between rs1047214 and the early onset of obesity in Germans was examined. This research also demonstrated that no association existed between this exon SNP and obesity when the body mass index of parents and their obese offspring were studied[33]. Both of these previous studies confirm the results found in this study that no associations were found between these three SNPs of *NPY2R* and body mass index, systolic blood pressure, and diastolic blood pressure.

***Haplotype Frequency***

Although no phenotypic association was found between the three *NPY2R* SNPs and the metabolic syndrome traits, several significant associations between haplotypes of these three SNPs and body mass index, systolic blood pressure, and diastolic blood pressure were found. By using the generalized linear model for linear regression, analysis of the phenotypic data demonstrated that four haplotypes were significantly associated with BMI. The second most frequent haplotype, GCT, was marginally associated with BMI, and it induces a directional effect on this trait. BMI does increase with an increased incidence of haplotype GCT. Haplotype GTT and the two rare haplotypes, ACC and ACT, were all significantly associated with BMI, and individuals with this haplotype do possess higher body mass indexes. Both of the rare haplotypes are very strongly associated with BMI, but since they appear so infrequently and amongst very few individuals, it is very difficult measure the effects of this haplotype. Overall, four haplotypes were associated with the metabolic syndrome trait of body mass index.

Next, the relationship between these haplotypes and systolic blood pressure was examined as well. Results very similar to the BMI associations were found with SBP as well. Three haplotypes were associated with SBP, and these were the two rare haplotypes, ACC and ACT, and haplotype GTT. All three of these haplotypes were correlated with an increase in systolic blood pressure as well. Again, the same drawback arises with the rare haplotypes. Although they are very significantly

associated with SBP, their recordable frequency is so low that it is difficult to measure the effect of these haplotypes.

Lastly, analysis was completed in order to determine if an association exists between these haplotypes and diastolic blood pressure. Two haplotypes, GTC and GTT, were both marginally associated with DBP. The association between GTC and DBP is very interesting because this haplotype does not predict an increase in diastolic blood pressure but rather a decrease in DBP. However, haplotype GTT is correlated with an increase in DBP. Not only does GTT predict an increase in DBP, but it also predicts an increase in BMI and SBP. This demonstrates that haplotype GTT has a pleiotropic effect on these traits. Pleiotropy is defined by the ability of one gene to affect multiple traits. Further study of these haplotypes may demonstrate the biological mechanism responsible for this pleiotropic effect.

***Future Studies***

Since no phenotypic association was found between these *NPY2R* SNPs and metabolic syndrome traits but an association was found between these traits and the haplotypes of these three polymorphisms, a study of the gene expression of the *NPY2R* gene may clarify the relationship between these SNPs and the metabolic syndrome traits. It is possible that another SNP in the region encompassed by the significant haplotypes may be responsible for the haplotype associations, or another biological process is responsible for these increases in BMI, SBP and DBP.

Also, in the past, significant associations have been found between these SNPs and BMI, so it would stand to reason these SNPs do alter the gene expression of

*NPY2R*. Previously, a marginal association was found between SNPs rs6857115 and rs1047214 and obesity in males and females of the Pima Indians, and a significant association existed between rs1047214 and obesity in males specifically. The individuals studied were extremely obese and their body mass indexes were compared to each of the SNPs to determine if an association exists[22]. This study looked at two of the candidate SNPs that were genotyped in this study, and significant associations were found between the SNPs and the metabolic syndrome traits. Also, another study found a significant association between the *NPY2R* SNP rs1047214 and the waist-to-hip ratios of severely obese children, and indicator of obesity[30]. Both of these studies have demonstrated that an association does exist between *NPY2R* SNPs and metabolic syndrome traits like body mass index and waist-to-hip ratios.

Since these SNPs may alter gene expression, this may be effected through the disruption of the Spi-B transcription factor binding motif. Since SPi-B binds to the conserved purine-rich binding motif of GGA, which is altered at the second G by rs6851222, expression of *NPY2R* may be affected. In order to determine if this actually occurs, a luciferase promoter reporter assay will be used to measure the level of *NPY2R* expression with the polymorphic SNPs. A 2,190 kilo base pair amplicon will be used to measure the gene expression since this fragment of DNA will encompass the entire region containing the common and rare SNPs in the *NPY2R* promoter and the 5' untranslated region. This experiment will truly help unravel the puzzle of *NPY2R* and its role in development of metabolic syndrome.

## Appendix

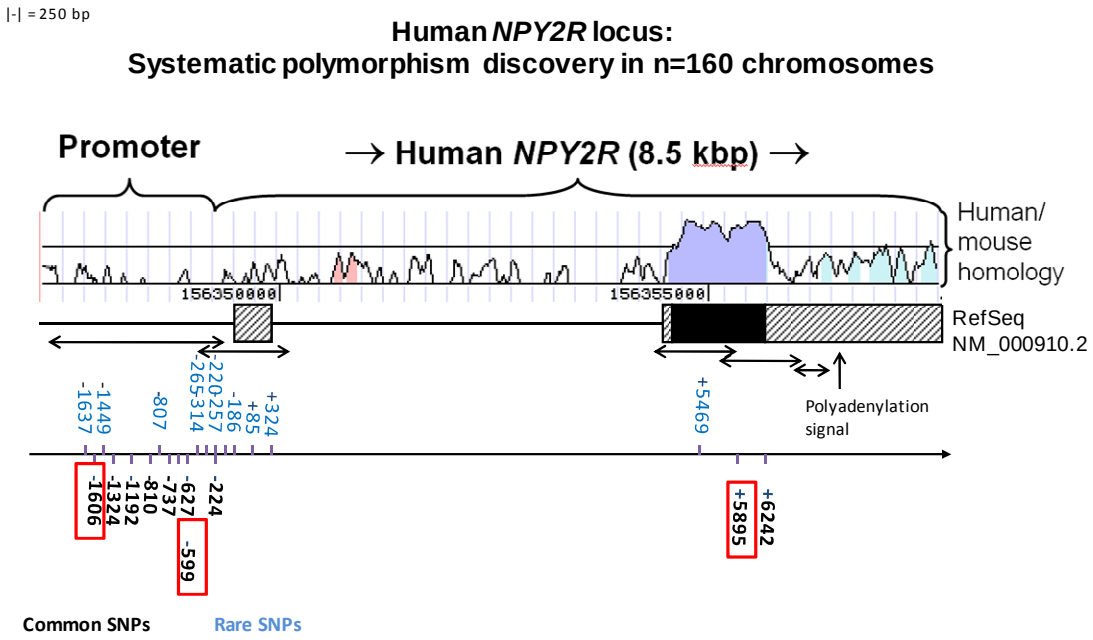


**Figure 1. *NPY2R* resequencing strategy and identified variants.** Sequences conserved between mouse and human *NPY2R* were visualized with VISTA <http://genome.lbl.gov/vista/index.shtml>. Locations of common (minor allele frequency >5%) and rare SNPs are indicated. Positions are numbered upstream or downstream of the CAP site. Solid blocks: Open Reading Frame. Hatched blocks: UTRs. SNPs selected for genotyping are boxed in red.

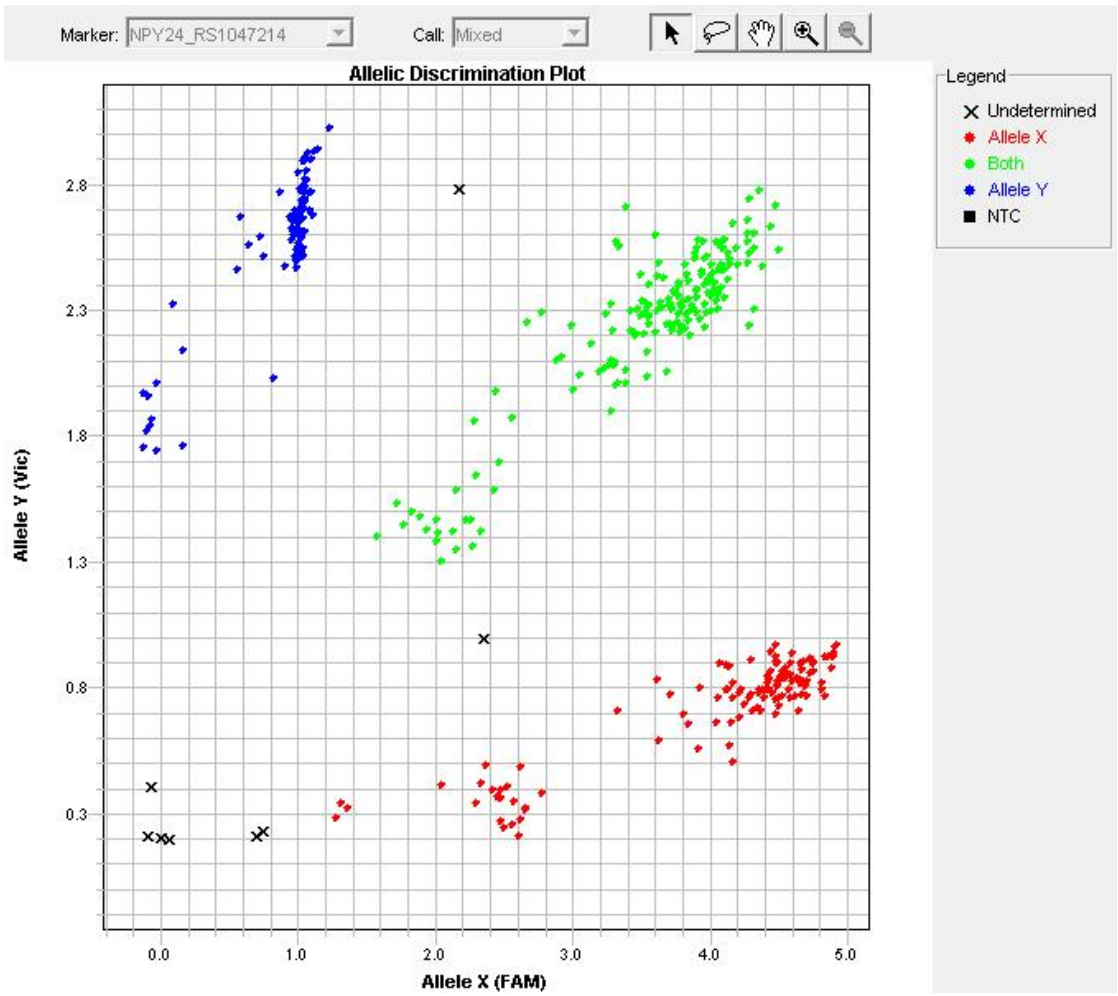


**Figure 2. Determination of SNP Genotypes based on Allelic Discrimination.** In this diagram, the genotypes subjects are determined for SNP rs1047214. The minor allele, C, is represented by allele Y, and the major allele, T, is represented by allele X. Blue marks represent individuals homozygous for the minor allele, green marks represent individuals heterozygous for the minor allele, and red marks represent individuals homozygous for the major allele, T. Genotypes were determined based on fluorescence levels of reporters representing each allele. Genotypes for SNPs rs6851222 and rs6857715 were also determined using the allelic discrimination function in the SDS 2.0 software.

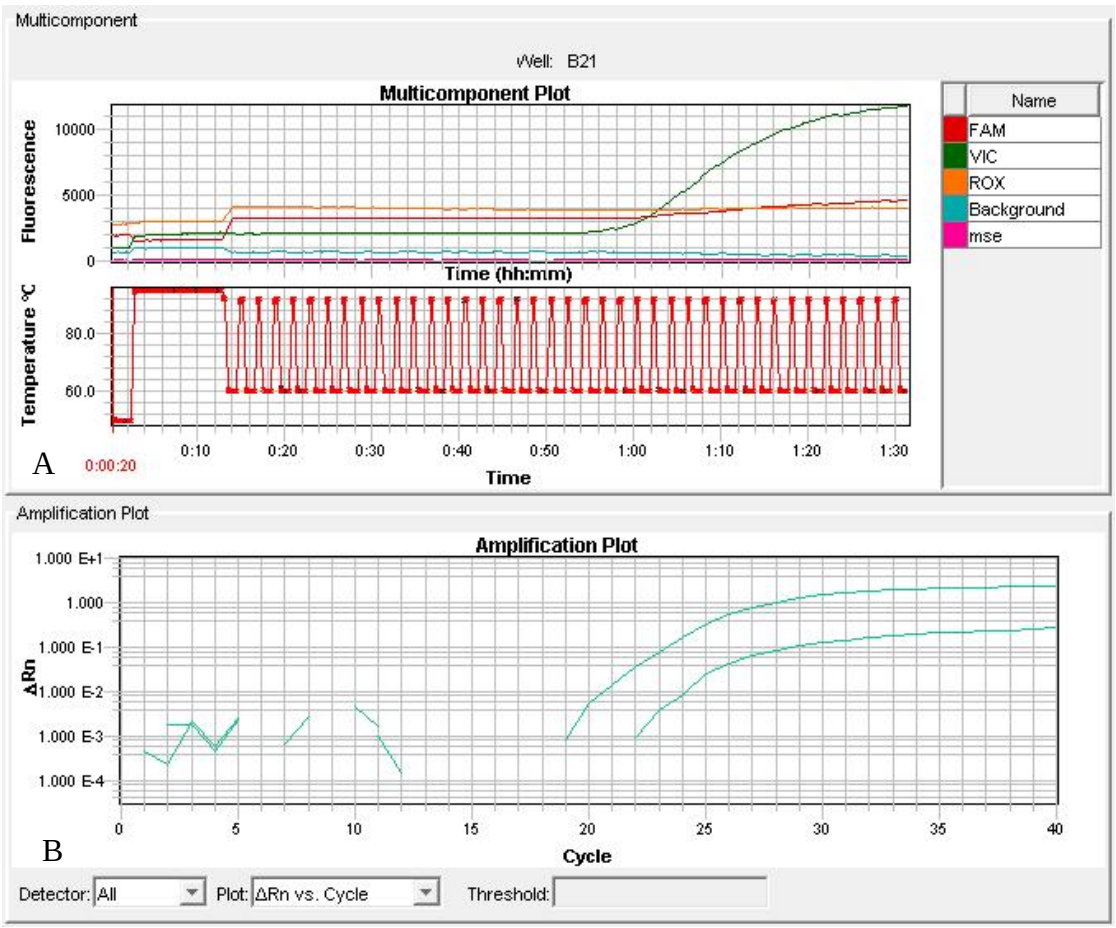


**Figure 3. A. Multicomponent Plot of Fluorescence Levels of Reporter Dyes for Genotypes Homozygous for the Minor Allele. 3B. Amplification Plot of Genotypes Homozygous for the Minor Allele.** (A) Genotypes determined using the allelic discrimination function are confirmed by checking the fluorescence levels in the multicomponent plot. Samples homozygous for the minor allele will show high fluorescence levels of DNA binding to the VIC probe. (B) The multicomponent plot demonstrates the level of amplification of PCR product in comparison to the number of cycles completed. The Rn values are the fluorescence levels of the reporter dyes normalized based on the passive reference fluorescence level. The amplification plot also verifies that the assigned genotype is true.

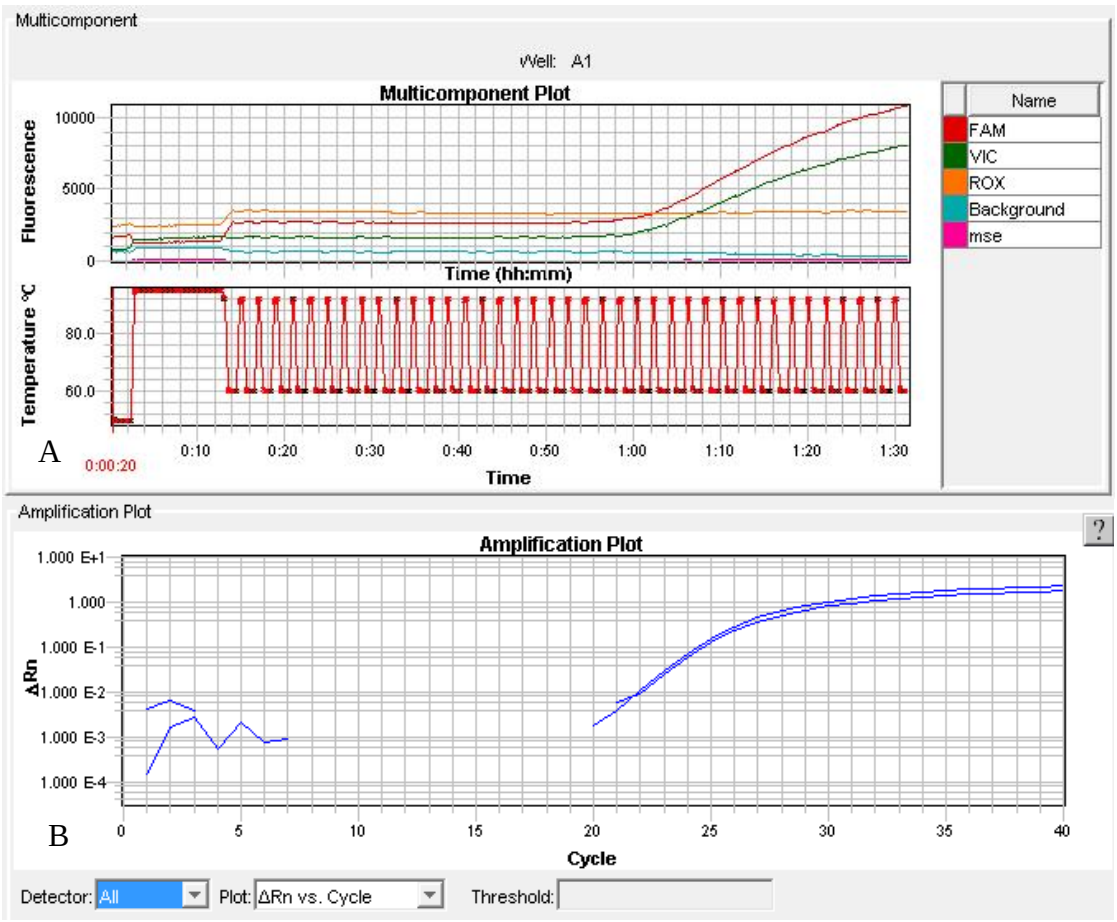


**Figure 4. A. Multicomponent Plot of Fluorescence Levels of Reporter Dyes for Genotypes Heterozygous for the Minor Allele. 3B. Amplification Plot of Genotypes Heterozygous for the Minor Allele.** (A) Genotypes determined using the allelic discrimination function are confirmed by checking the fluorescence levels in the multicomponent plot. Samples heterozygous for the minor allele will show high fluorescence levels of both VIC and FAM due to the DNA binding to both the VIC and FAM probe. (B) The multicomponent plot demonstrates the level of amplification of PCR product in comparison to the number of cycles completed. The Rn values are the fluorescence levels of the reporter dyes normalized based on the passive reference fluorescence level. This plot demonstrates that there is almost equal amplification of DNA bound to both the VIC and FAM probes.

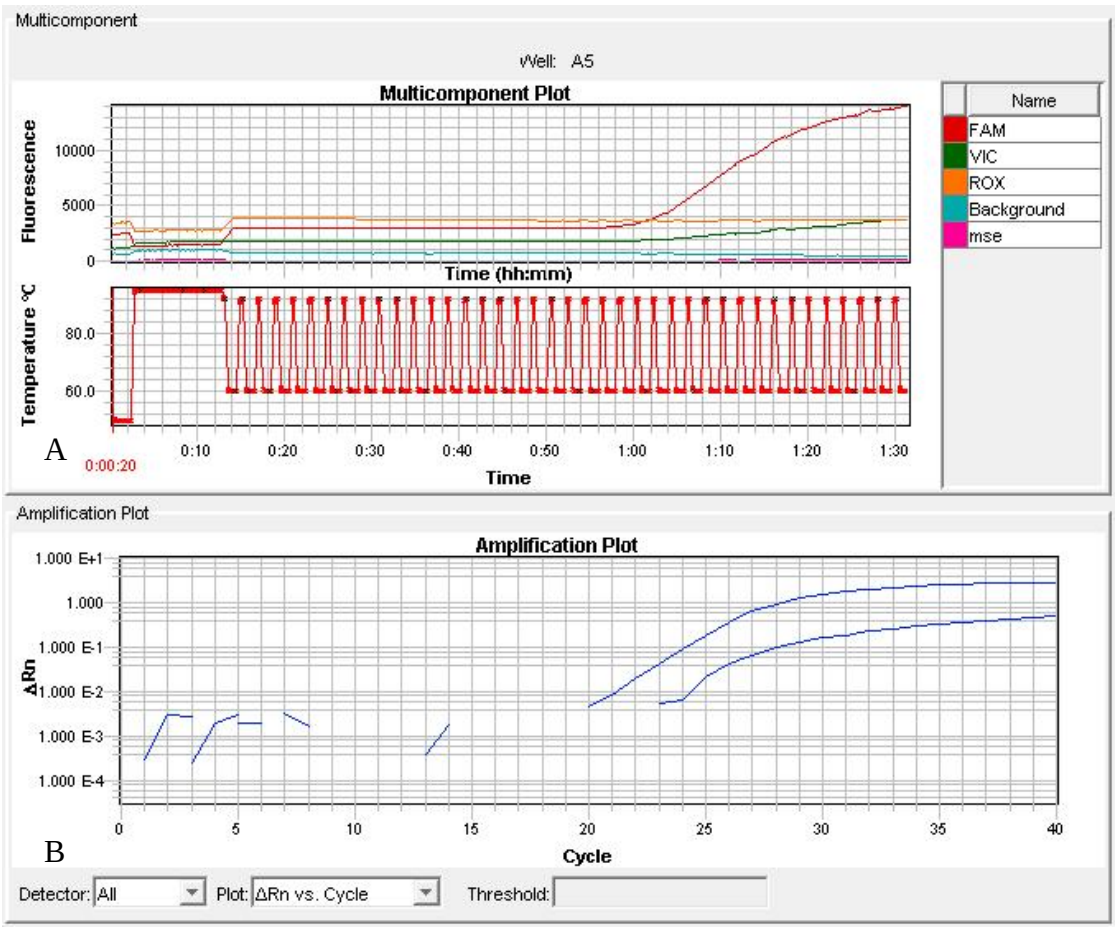


**Figure 5. A. Multicomponent Plot of Fluorescence Levels of Reporter Dyes for Genotypes Homozygous for the Major Allele. 3B. Amplification Plot of Genotypes Homozygous for the Major Allele.** (A) Genotypes determined using the allelic discrimination function are confirmed by checking the fluorescence levels in the multicomponent plot. Samples homozygous for the major allele will show high fluorescence levels FAM due to the DNA binding to the FAM probe. (B) The multicomponent plot demonstrates the level of amplification of PCR product in comparison to the number of cycles completed. The Rn values are the fluorescence levels of the reporter dyes normalized based on the passive reference fluorescence level. This plot demonstrates that there are higher levels of fluorescence of the FAM probe, indicating increased amounts of amplified DNA.

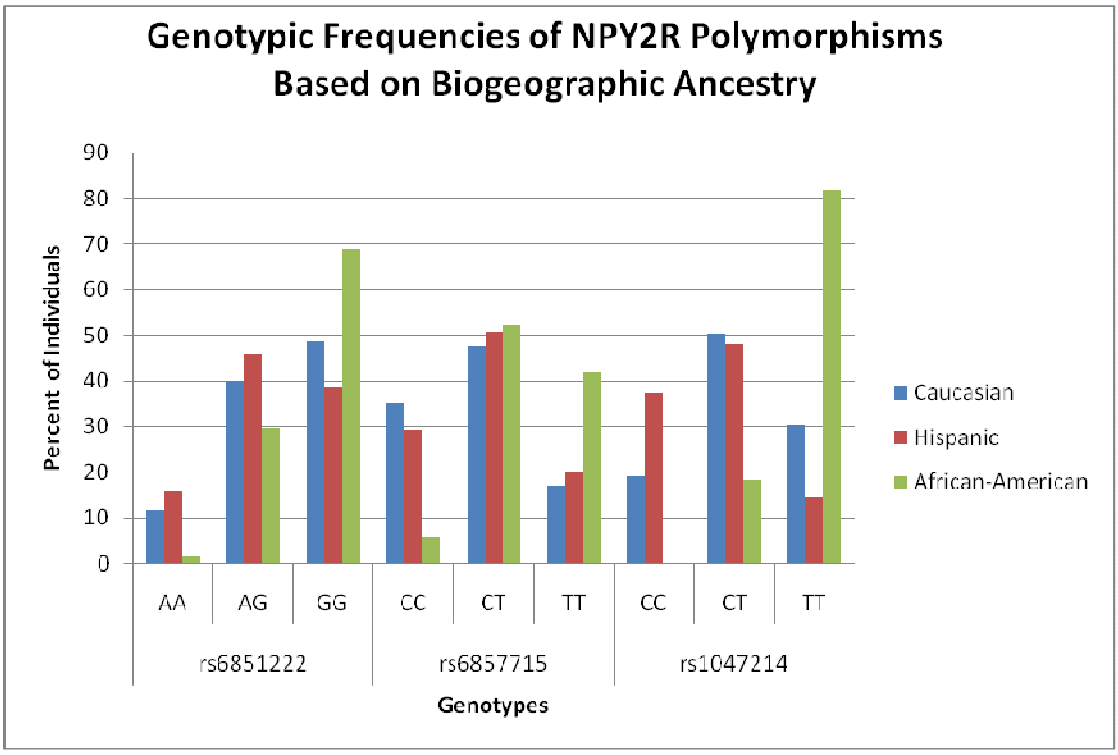


**Figure 6. Genotypic Frequencies of *NPY2R* Polymorphisms Based on Biogeographic Ancestry.** The frequency of the two promoter SNPs, rs6851222 and rs6857715, and the exon SNP, rs1047214, based on biogeographic ancestry are shown above. With respect to rs6851222, the major allele was G, and the minor allele A for all three biogeographic ancestries. With respect to rs6857715, the major allele was C , and the minor allele was T for Caucasians and Hispanics. The major allele was T, and the minor allele was C for African-Americans. With respect to rs1047214, the major allele was T, and the minor allele was C for Caucasians and African-Americans. The major allele was C, and the minor allele was T for Hispanics.

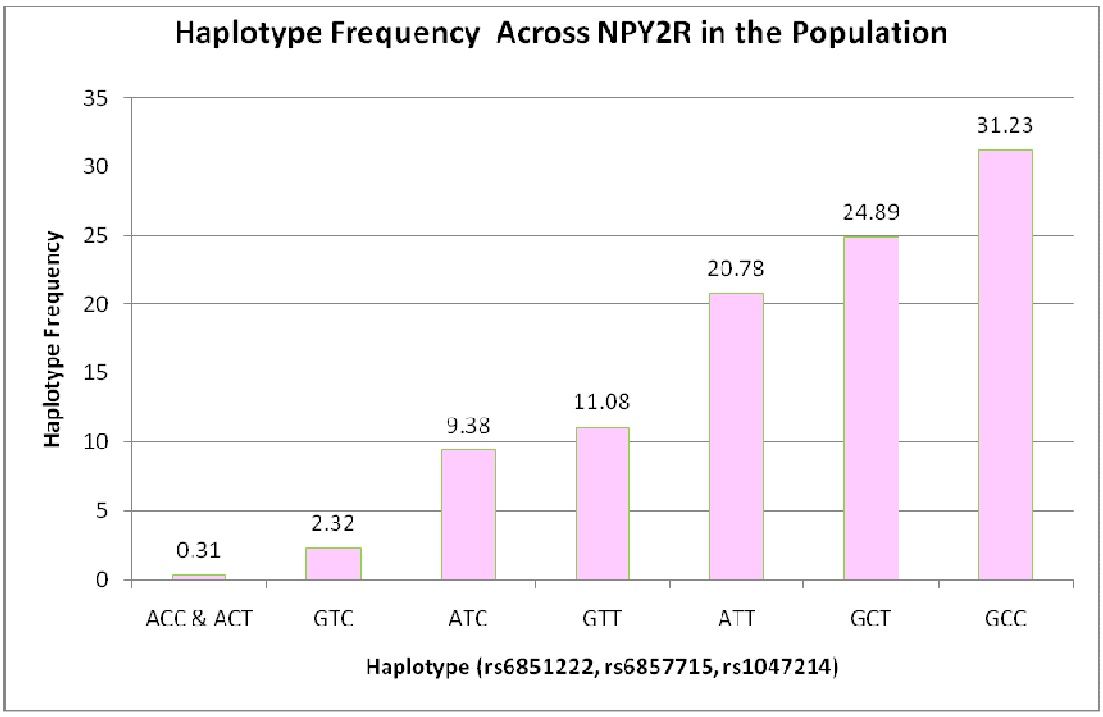


**Figure 7. Haplotype Frequency Across *NPY2R* in the Population.** In this figure, the frequency of each of the eight haplotypes produced from the three *NPY2R* SNPs is depicted above. The alleles in the haplotypes represent the SNPs in the following order: rs6851222, rs6857715, and rs1047214. The two rare haplotypes with a frequency <0.5% are clustered together. The most frequent haplotype is GCC, and the least frequent haplotypes are ACC and ACT.

**Table 1. *NPY2R* Polymorphism Discovery in 80 Individuals of 4 Biogeographical Ancestries.** 21 *NPY2R* SNPs were discovered in 80 individuals of the Caucasian, African-American, Hispanic, and Asian biogeographic ancestries. The 10 common SNPs are bolded and the 11 rare SNPs are unbolded. Common SNPs have a minor allele frequency greater than 5%. SNPs were found in the promoter region of the gene, the 5' untranslated region, and the open reading frame. Amino acid changes for SNPs in the coding region of the gene are listed below.

| *NPY2R* SNPs | Alleles | SNP position | RefSNP number | AA Change | Minor allele frequency | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | White n = 22 | Black n = 26 | Hispanic n = 16 | Asian n = 16 | All n = 80 |
| 1 | A/G | -1637, promoter | rs57869523 | none | - | - | - | - | 0.018 |
| **2** | G/A | -1606, promoter | rs6851222 | none | 0.25 | 0.16 | 0.25 | 0.47 | 0.270 |
| 3 | T/C | -1449, promoter | rs10212938 | none | - | - | - | - | 0.029 |
| **4** | */T | -1324, promoter | rs36032070 | none | 0.25 | 0.22 | 0.25 | 0.47 | 0.290 |
| **5** | G/A | -1192, promoter | rs33977152 | none | 0.11 | 0 | 0.12 | 0.19 | 0.090 |
| **6** | -/GA | -810, promoter | rs35987718 | none | 0.45 | 0.12 | 0.38 | 0.25 | 0.287 |
| 7 | -/AGAG | -807, promoter | rs34874489 | none | - | - | - | - | 0.017 |
| **8** | A/T | -737, promoter | rs12507396 | none | 0.12 | 0.08 | 0.22 | 0.2 | 0.140 |
| **9** | A/G | -627, promoter | rs6857530 | none | 0.29 | 0.65 | 0.41 | 0.47 | 0.470 |
| **10** | C/T | -599, promoter | rs6857715 | none | 0.29 | 0.36 | 0.41 | 0.47 | 0.470 |
| 11 | C/A | -314, promoter | NA | none | - | - | - | - | 0.011 |
| 12 | C/G | -265, promoter | NA | none | - | - | - | - | 0.018 |
| 13 | C/T | -257, promoter | rs73855386 | none | - | - | - | - | 0.018 |
| **14** | A/G | -224, promoter | rs2234759 | none | 0.24 | 0.2 | 0.33 | 0.53 | 0.310 |
| 15 | G/A | -220, promoter | NA | none | - | - | - | - | 0.047 |
| 16 | G/C | -186, promoter | NA | none | - | - | - | - | 0.005 |
| 17 | C/T | +85, exon1 (5'UTR) | NA | none | - | - | - | - | 0.005 |
| 18 | C/T | +324, exon1 (5'UTR) | rs72972775 | none | - | - | - | - | 0.029 |
| 19 | T/C | +5469, exon 2 | rs2342674 | L-53-L | - | - | - | - | 0.017 |
| **20** | T/C | +5895, exon 2 | rs1047214 | I-195-I | 0.48 | 0.12 | 0.34 | 0.07 | 0.250 |
| **21** | T/C | +6242, exon 2 | rs2880415 | I-312-I | 0.5 | 0.33 | 0.43 | 0.13 | 0.360 |

**Table 2. Allele Frequencies of SNPs rs6851222, rs6857715, and rs1047214.** The allele frequencies for each of the three SNPs genotyped are listed below in this table based on biogeographical ancestry. The results of tests for Hardy-Weinberg equilibrium are represented by the p-values of the last columns. All SNPs exist in Hardy-Weinberg equilibrium for each biogeographical ancestry.

| | **rs6851222** | | | **rs6857715** | | | **rs1047214** | | |
|---|---|---|---|---|---|---|---|---|---|
| | A | G | P-value | C | T | P-value | C | T | P-value |
| **Caucasian** | 0.31 | 0.69 | 0.085 | 0.59 | 0.41 | 0.701 | 0.44 | 0.56 | 0.604 |
| **Hispanic** | 0.39 | 0.61 | 0.759 | 0.55 | 0.45 | 0.848 | 0.61 | 0.39 | 0.876 |
| **African-American** | 0.16 | 0.84 | 0.451 | 0.32 | 0.68 | 0.104 | 0.09 | 0.91 | 0.396 |

**Table 3. Observed Genotype Frequencies for SNP rs6851222 based on Biogeographic Ancestry.** These observed frequencies were compared to the expected frequencies of each genotype in a chi-square test for independence. The indicated p-value demonstrates that this SNP is dependent on biogeographic ancestry since $p<0.05$.

| | AA | AG | GG | Total |
|---|---|---|---|---|
| **Caucasian** | 63 | 219 | 266 | 548 |
| **Hispanic** | 13 | 38 | 32 | 83 |
| **African-American** | 1 | 21 | 49 | 71 |
| **Total** | 77 | 278 | 347 | 702 |
| **P-value** | 0.001276 | | | |

**Table 4. Observed Genotype Frequencies for SNP rs6857715 based on Biogeographic Ancestry.** These observed frequencies were compared to the expected frequencies of each genotype in a chi-square test for independence. The indicated p-value demonstrates that this SNP is dependent on biogeographic ancestry since $p<0.05$.

| | CC | CT | TT | Total |
|---|---|---|---|---|
| **Caucasian** | 193 | 259 | 93 | 545 |
| **Hispanic** | 25 | 43 | 17 | 85 |
| **African-American** | 4 | 35 | 28 | 67 |
| **Total** | 222 | 337 | 138 | 697 |
| **P-value** | $4.183 \times 10^{-7}$ | | | |

**Table 5. Observed Genotype Frequencies for SNP 1047214 based on Biogeographic Ancestry.** These observed frequencies were compared to the expected frequencies of each genotype in a chi-square test for independence. The indicated p-value demonstrates that this SNP is dependent on biogeographic ancestry since $p<0.05$.

| | CC | CT | TT | Total |
|---|---|---|---|---|
| **Caucasian** | 105 | 277 | 167 | 549 |
| **Hispanic** | 31 | 40 | 12 | 83 |
| **African-American** | 0 | 13 | 58 | 71 |
| **Total** | 136 | 330 | 237 | 703 |
| **P-value** | $6.332 \times 10^{-21}$ | | | |

**Table 6. Frequency of SNP rs6851222 Amongst Caucasians, Hispanics and African-Americans**. This table provides the number of individuals that were homozygous for the minor allele, AA, heterozygous for the minor allele, AG, and homozygous for the major allele, GG, based on biogeographic ancestry.

| Genotype | Caucasian | | Hispanic | | African-American | | Total | |
|---|---|---|---|---|---|---|---|---|
| | No. | % | No. | % | No. | % | No. | % |
| AA | 63 | 11.50 | 13 | 15.66 | 1 | 1.41 | 77 | 10.97 |
| AG | 219 | 39.96 | 38 | 45.78 | 21 | 29.58 | 278 | 39.60 |
| GG | 266 | 48.54 | 32 | 38.55 | 49 | 69.01 | 347 | 49.43 |

**Table 7. Frequency of SNP rs6857715 Amongst Caucasians, Hispanics, and African-Americans.** The above table provides the number of individuals that were homozygous for the minor allele, CC, heterozygous for the minor allele, CT, and homozygous for the major allele, TT, based on biogeographic ancestry.

| Genotype | Caucasian | | Hispanic | | African-American | | Total | |
|---|---|---|---|---|---|---|---|---|
| | No. | % | No. | % | | No. | % | No. |
| CC | 193 | 35.41 | 25 | 29.41 | 4 | 5.97 | 222 | 31.85 |
| CT | 259 | 47.52 | 43 | 50.59 | 35 | 52.24 | 337 | 46.94 |
| TT | 93 | 17.06 | 17 | 20.00 | 28 | 41.79 | 138 | 33.71 |

**Table 8. Frequency of SNP rs1047214 Amongst Caucasians, Hispanics, and African-Americans.** The above table provides the number of individuals that were homozygous for the minor allele, CC, heterozygous for the minor allele, CT, and homozygous for the major allele, TT, based on biogeographic ancestry.

| Genotype | Caucasian | | Hispanic | | African-American | | Total | |
|---|---|---|---|---|---|---|---|---|
| | No. | % | No. | % | | No. | % | No. |
| CC | 105 | 19.13 | 31 | 37.35 | 0 | 0.00 | 136 | 19.35 |
| CT | 277 | 50.46 | 40 | 48.19 | 13 | 18.31 | 330 | 46.94 |
| TT | 167 | 30.42 | 12 | 14.46 | 58 | 81.69 | 237 | 33.71 |

**Table 9. Average Body Mass Index and Systolic Blood Pressure for SNP rs6851222**. The average systolic blood pressure and body mass indexes with standard errors are presented in this table. No significant association was found between the rs6851222 and both body mass index and systolic blood pressure.

| Genotype | Systolic Blood Pressure | | Body Mass Index | |
|---|---|---|---|---|
| | Mean | Standard Error | Mean | Standard Error |
| AA | 135.12 | 1.02 | 26.77 | 1.03 |
| AG | 134.29 | 1.01 | 26.86 | 1.02 |
| GG | 134.03 | 1.01 | 27.42 | 1.02 |
| P-Value | 0.9004 | | 0.4891 | |

**Table 10. Average Body Mass Index and Systolic Blood Pressure for SNP rs6857715**. The average systolic blood pressure and body mass indexes with standard errors are presented in this table. No significant association was found between the rs6857715 and both body mass index and systolic blood pressure.

| Genotype | Systolic Blood Pressure | | Body Mass Index | |
|---|---|---|---|---|
| | Mean | Standard Error | Mean | Standard Error |
| CC | 134.13 | 1.01 | 27.77 | 1.02 |
| CT | 134.26 | 1.01 | 27.05 | 1.02 |
| TT | 134.14 | 1.01 | 26.88 | 1.02 |
| P-Value | 0.9930 | | 0.3716 | |

**Table 11. Average Body Mass Index and Systolic Blood Pressure for SNP rs1047214**. The average systolic blood pressure and body mass indexes with standard errors are presented in this table. No significant association was found between the rs1047214 and both body mass index and systolic blood pressure.

| Genotype | Systolic Blood Pressure | | Body Mass Index | |
|---|---|---|---|---|
| | Mean | Standard Error | Mean | Standard Error |
| CC | 134.63 | 1.01 | 26.97 | 1.02 |
| CT | 133.67 | 1.01 | 26.82 | 1.02 |
| TT | 134.50 | 1.01 | 27.54 | 1.02 |
| P-Value | 0.7615 | | 0.4836 | |

**Table 12.Frequency of *NPY2R* Haplotypes**. The following haplotypes were derived from the three SNPs genotyped. The order of the three alleles represents each SNP in the following order: rs6851222, rs6857715, and rs1047214. The fractional and percent frequencies are listed below. The Haplotypes are ordered from the least frequent haplotype to the most frequent haplotype. ACC and ACT represent the two rarest haplotypes, which were present at very low frequencies.

| Haplotype | Frequency | Percent |
|---|---|---|
| **ACC & ACT** | 0.00317 | 0.31 |
| **GTC** | 0.02321 | 2.32 |
| **ATC** | 0.09383 | 9.38 |
| **GTT** | 0.11089 | 11.08 |
| **ATT** | 0.20779 | 20.78 |
| **GCT** | 0.24886 | 24.89 |
| **GCC** | 0.31226 | 31.23 |

**Table 13. Linear Regression Results for Haplotype Association Study with Body Mass Index.** Each haplotype was analyzed for an association with body mass index based on the generalized linear model. The coefficient of regression, standard error, and p-value for each haplotype is listed in the above table. GCT, GTT, ACC, and ACT all were associated with body mass index.

| Haplotype | Frequency | Coefficient of Regression | Standard Error | P-value |
|---|---|---|---|---|
| **ATC** | 9.38 | 0.03011 | 0.021894 | 0.1700 |
| **ATT** | 20.78 | -0.00235 | 0.014458 | 0.8710 |
| **GCT** | 24.89 | 0.02904 | 0.014671 | 0.0481 |
| **GTC** | 2.32 | -0.02463 | 0.038408 | 0.5220 |
| **GTT** | 11.08 | 0.05835 | 0.016332 | 0.00037500 |
| **ACC & ACT** | 0.31 | 0.51201 | 0.092454 | 0.00000004 |

**Table 14. Linear Regression Results for Haplotype Association Study with Systolic Blood Pressure.** Each haplotype was analyzed for an association with systolic blood pressure based on the generalized linear model. The coefficient of regression, standard error, and p-value for each haplotype is listed in the above table. GTT, ACC, and ACT all were associated with systolic blood pressure.

| Haplotype | Frequency | Coefficient of Regression | Standard Error | P-value |
|---|---|---|---|---|
| **ATC** | 9.38 | 0.014456 | -0.0612 | 0.9510 |
| **ATT** | 20.78 | 0.008855 | 0.8341 | 0.4050 |
| **GCT** | 24.89 | 0.009235 | 1.7492 | 0.0807 |
| **GTC** | 2.32 | 0.025308 | 1.2529 | 0.2110 |
| **GTT** | 11.08 | 0.009668 | 2.3578 | 0.0187 |
| **ACC & ACT** | 0.31 | 0.070706 | 2.5949 | 0.00967 |

**Table 15. Linear Regression Results for Haplotype Association Study with Diastolic Blood Pressure.** Each haplotype was analyzed for an association with diastolic blood pressure based on the generalized linear model. The coefficient of regression, standard error, and p-value for each haplotype is listed in the above table. GTT, ACC, and ACT all were associated with diastolic blood pressure.

| Haplotype | Frequency | Coefficient of Regression | Standard Error | P-value |
|---|---|---|---|---|
| **ATC** | 9.38 | -0.00918 | 0.01699 | 0.5890 |
| **ATT** | 20.78 | 0.00862 | 0.010654 | 0.4190 |
| **GCT** | 24.89 | -0.00571 | 0.010961 | 0.6020 |
| **GTC** | 2.32 | -0.09739 | 0.02836 | 0.000632 |
| **GTT** | 11.08 | 0.02528 | 0.011747 | 0.0318 |
| **ACC & ACT** | 0.31 | 0.14098 | 0.09079 | 0.121 |

**References**


1. Ammar, D. A., D. M. Eadie, D. J. Wong, Y. Y. Ma, L. F. Kolakowski, Jr., T. L. Yang-Feng and D. A. Thompson (1996). "Characterization of the human type 2 neuropeptide Y receptor gene (NPY2R) and localization to the chromosome 4q region containing the type 1 neuropeptide Y receptor gene." Genomics 38(3): 392-398.

2. Bartholdy, B., C. Du Roure, A. Bordon, D. Emslie, L. M. Corcoran and P. Matthias (2006). "The Ets factor Spi-B is a direct critical target of the coactivator OBF-1." Proceedings of the National Academy of Sciences 103(31): 11665-11670.

3. Batterham, R. L., M. A. Cowley, C. J. Small, H. Herzog, M. A. Cohen, C. L. Dakin, A. M. Wren, A. E. Brynes, M. J. Low, M. A. Ghatei, R. D. Cone and S. R. Bloom (2002). "Gut hormone PYY3-36 physiologically inhibits food intake." Nature 418(6898): 650-654.

4. Batterham, R. L., H. Heffron, S. Kapoor, J. E. Chivers, K. Chandarana, H. Herzog, C. W. Le Roux, E. L. Thomas, J. D. Bell and D. J. Withers (2006). "Critical role for peptide YY in protein-mediated satiation and body-weight regulation." Cell Metabolism 4(3): 223-233.

5. Batterham, R. L., C. W. Le Roux, M. A. Cohen, A. J. Park, S. M. Ellis, M. Patterson, G. S. Frost, M. A. Ghatei and S. R. Bloom (2003). "Pancreatic Polypeptide Reduces Appetite and Food Intake in Humans." Journal of Clinical Endocrinology Metabolism 88(8): 3989-3992.

6. Broberger, C., M. Landry, H. Wong, J. N. Walsh and T. Hokfelt (1997). "Subtypes Y1 and Y2 of the neuropeptide Y receptor are respectively expressed in pro-opiomelanocortin- and neuropeptide-Y-containing neurons of the rat hypothalamic arcuate nucleus." Neuroendocrinology 66(6): 393-408.

7. Camilleri, M. (2009). "Peripheral mechanisms in the control of appetite and related experimental therapies in obesity." Regulatory Peptides 156(1-3): 24-27.

8. Campbell, C. D., H. N. Lyon, J. Nemesh, J. A. Drake, T. Tuomi, D. Gaudet, X. Zhu, R. S. Cooper, K. G. Ardlie, L. C. Groop and J. N. Hirschhorn (2007). "Association studies of BMI and type 2 diabetes in the neuropeptide Y pathway: a possible role for NPY2R as a candidate gene for type 2 diabetes in men." Diabetes 56(5): 1460-1467.

9. Clark, J. T., P. S. Kalra, W. R. Crowley and S. P. Kalra (1984). "NEUROPEPTIDE Y AND HUMAN PANCREATIC POLYPEPTIDE STIMULATE FEEDING BEHAVIOR IN RATS." Endocrinology 115(1): 427-429.

10. Connie Hung, C.-C., F. Pirie, J. a. Luan, E. Lank, A. Motala, G. S. H. Yeo, J. M. Keogh, N. J. Wareham, S. O'Rahilly and I. S. Farooqi (2004). "Studies of the Peptide YY and Neuropeptide Y2 Receptor Genes in Relation to Human Obesity and Obesity-Related Traits." Diabetes 53(9): 2461-2466.

11. Duvnjak, L., and M. Duvnjak. (2009). "The Metabolic Syndrome—An Ongoing Story." Journal of Physiology and Pharmacology 60(6): 19-24.

12. Fuhlendorff, J., N. L. Johansen, S. G. Melberg, H. Thøgersen and T. W. Schwartz (1990). "The antiparallel pancreatic polypeptide fold in the binding of neuropeptide Y to Y1 and Y2 receptors." Journal of Biological Chemistry 265(20): 11706-11712.

13. Gehlert, D. R. (1998). "Multiple Receptors for the Pancreatic Polypeptide (PP-Fold) Family: Physiological Implications (44263). [Review]." Proceedings of the Society for Experimental Biology & Medicine May 218(1): 7-22.

14. Gerald, C., M. W. Walker, P. J. J. Vaysse, C. He, T. A. Branchek and R. L. Weinshank (1995). "Expression Cloning and Pharmacological Characterization of a Human Hippocampal Neuropeptide Y/Peptide YY Y2 Receptor Subtype." Journal of Biological Chemistry 270(45): 26758-26761.

15. Glenn Stanley, B., S. E. Kyrkouli, S. Lampert and S. F. Leibowitz "Neuropeptide Y chronically injected into the hypothalamus: A powerful neurochemical inducer of hyperphagia and obesity." Peptides 7(6): 1189-1192.

16. Gonzalez, R. and S. Unniappan (2010). "Molecular characterization, appetite regulatory effects and feeding related changes of peptide YY in goldfish." General and Comparative Endocrinology 166(2): 273-279.

17. Grundy, S. M. (2007). "Metabolic Syndrome: A Multiplex Cardiovascular Risk Factor." Journal of Clinical Endocrinology Metabolism 92(2): 399-404.
18. Grundy, S. M. (2008). "Metabolic Syndrome Pandemic." Arteriosclerosis, Thrombosis, and Vascular Biology 28(4): 629-636.

19. Grundy, S. M., J. I. Cleeman, S. R. Daniels, K. A. Donato, R. H. Eckel, B. A. Franklin, D. J. Gordon, R. M. Krauss, P. J. Savage, S. C. Smith, Jr., J. A. Spertus and F. Costa (2005). "Diagnosis and Management of the Metabolic

Syndrome: An American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement." Circulation 112(17): 2735-2752.

20. Karra, E. and R. L. Batterham (2010). "The role of gut hormones in the regulation of body weight and energy homeostasis." Molecular and Cellular Endocrinology 316(2): 120-128.

21. Karra, E., K. Chandarana and R. L. Batterham (2009). "The role of peptide YY in appetite regulation and obesity." The Journal of Physiology 587(1): 19-25.

22. Ma, L., P. A. Tataranni, R. L. Hanson, A. M. Infante, S. Kobes, C. Bogardus and L. J. Baier (2005). "Variations in Peptide YY and Y2 Receptor Genes Are Associated With Severe Obesity in Pima Indian Men." Diabetes 54(5): 1598-1602.

23. Morimoto, R., F. Satoh, O. Murakami, K. Totsune, M. Saruta, T. Suzuki, H. Sasano, S. Ito and K. Takahashi (2008). "Expression of peptide YY in human brain and pituitary tissues." Nutrition 24(9): 878-884.

24. Moriya, R., S. Mashiko, A. Ishihara, T. Takahashi, T. Murai, J. Ito, Y. Mitobe, Z. Oda, H. Iwaasa, F. Takehiro and A. Kanatani (2009). "Comparison of independent and combined chronic anti-obese effects of NPY Y2 receptor agonist, PYY(3-36), and NPY Y5 receptor antagonist in diet-induced obese mice." Peptides 30(7): 1318-1322.

25. Neary, M. T. and R. L. Batterham (2009). "Gut hormones: Implications for the treatment of obesity." Pharmacology & Therapeutics 124(1): 44-56.

26. Neary, M. T. and R. L. Batterham (2009). "Peptide YY: Food for thought." Physiology & Behavior 97(5): 616-619.

27. O'Connor, D. T. a., M. T. a. Kailasam, B. P. a. Kennedy, M. G. a. Ziegler, N. b. Yanaihara and R. J. a. Parmer (2002). "Early decline in the catecholamine release-inhibitory peptide catestatin in humans at genetic risk of hypertension. [Article]." Journal of Hypertension July 20(7): 1335-1345.

28. Rose, P. M., P. Fernandes, J. S. Lynch, S. T. Frazier, S. M. Fisher, K. Kodukula, B. Kienzle and R. Seethala (1995). "Cloning and Functional Expression of a cDNA Encoding a Human Type 2 Neuropeptide Y Receptor." Journal of Biological Chemistry 270(39): 22661-22664.

29. Shih, P.-a. B., L. Wang, S. Chiron, G. Wen, C. Nievergelt, M. Mahata, S. Khandrika, F. Rao, M. M. Fung, S. K. Mahata, B. A. Hamilton and D. T.

O'Connor (2009). "Peptide YY (PYY) Gene Polymorphisms in the 3'-Untranslated and Proximal Promoter Regions Regulate Cellular Gene Expression and PYY Secretion and Metabolic Syndrome Traits in Vivo." Journal of Clinical Endocrinology Metabolism 94(11): 4557-4566.

30. Siddiq, A., M. Gueorguiev, C. Samson, S. Hercberg, B. Heude, C. Levy-Marchal, B. Jouret, J. Weill, D. Meyre, A. Walley and P. Froguel (2007). "Single nucleotide polymorphisms in the neuropeptide Y2 receptor (NPY2R) gene and association with severe obesity in French white subjects." Diabetologia 50(3): 574-584.

31. Torekov, S. S., L. H. Larsen, G. Andersen, A. Albrechtsen, C. Glumer, K. Borch-Johnsen, T. Jorgensen, T. Hansen and O. Pedersen (2006). "Variants in the 5' region of the neuropeptide Y receptor Y2 gene (NPY2R) are associated with obesity in 5,971 white subjects." Diabetologia 49(11): 2653-2658.

32. Vincent, R. P. and C. W. le Roux (2008). "The satiety hormone peptide YY as a regulator of appetite." Journal of Clinical Pathology 61(5): 548-552.

33. Wang, H. J., A. K. Wermter, T. T. Nguyen, A. Scherag, K. Reichwald, B. Waldenmaier, P. Lichtner, T. Bettecken, J. Hebebrand and A. Hinney (2007). "No association of sequence variants in the neuropeptide Y2 receptor (NPY2R) gene with early onset obesity in Germans." Horm Metab Res 39(11): 840-844.

34. Wasylk, B., S. L. Hahn and A. Giovane (1993). "The Ets family of transcription factors." European Journal of Biochemistry 211(1-2): 7-18.

35. Yamamoto, H., F. Kihara-Negishi, T. Yamada, M. Suzuki, T. Nakano and T. Oikawa (2002). "Interaction between the Hematopoietic Ets Transcription Factor Spi-B and the Coactivator CREB-binding Protein Associated with Negative Cross-talk with c-Myb." Cell Growth Differentiation 13(2): 69-75.