# Selecting CD36 for a candidate gene study

Steven Piazza[1], Renee Pasker[2], Bert Boyer[2,3]

[1] Rural Alaska Honors Institute, UAF; [2] University of Alaska Fairbanks

## Abstract

After six weeks of research, up to 17 SNPs in gene CD36 have been selected to be genotyped for the CANHR study.

Using the NCBI online databases, research was done to find previous evidence linking SNPs in CD36 with obesity and related phenotypes.

Two SNPs, rs1527479 and rs1527483, showed good support in a connection to the phenotypes. 15 other tagging SNPs were also chosen.

With these SNPs selected for genotyping, the genetic testing can begin.

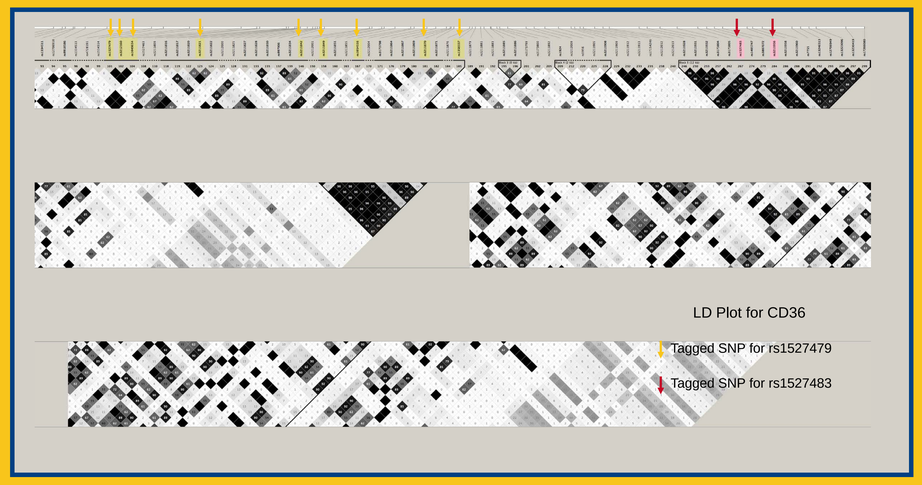
## Introduction

Humans are 99.9% identical on a genetic level. The 0.1% difference is caused by insertions, deletions and substitutions in the DNA sequence. These substitutions are known as Single Nucleotide Polymorphisms (SNPs). They occur about every 1000 base pairs. This still means there are about 3 million SNPs that make up differences in people.

In a candidate gene study, researchers look for associations between the SNPs, and peoples' phenotypes, descriptive factors, such as eye color or a disease.

Unfortunately, it is very expensive to genotype all SNPs of the subjects, and even more expensive to sequence all 3 billion base pairs of an individuals genome. Instead, only a few genes are picked and only a few SNPs are genotyped for each of those genes. It takes a lot of research before any testing is done

In the Center for Alaska Native Health Research (CANHR) study, researchers are trying to find connections between the identified SNPs and obesity in the Alaska Native Yup'ik population in the Yukon-Kuskokwim Delta.




## Methods

The first task in the process is to find a gene where there has been some association with the observed phenotype. Using the National Center for Biotechnology Information (NCBI) databases, I searched through online scientific journals for articles about genes that were associated with obesity.

After researching several genes, I decided to pick the CD36 gene for further research. The CD36 gene produces a protein that transports fatty acids in cells, thus has a good connection with the obesity phenotype.

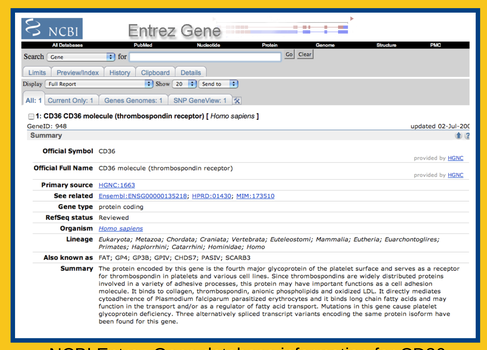
After choosing the gene, I obtained the sequence of base pairs through the Entrez Gene database on the NCBI website.

The next step was to find all the SNPs in that gene. The International HapMap Project has made a public database containing genotype information from four different populations: Chinese, Japanese, African and Caucasian. Using the site, I downloaded the SNP Genotype Data for CD36 from both the Chinese and Caucasian populations. The information was then put into a Microsoft Excel file.

The Excel file had to be converted using the SAS program. Using the converted file, the SNP data could then be used in the Haploview program. Haploview then takes the information and outputs a list of SNPs. Now, I had a list of over 150 SNPs for the CD36 gene. This program also created a Linkage Disequilibrium Plot for all the SNPs.

I used the NCBI SNP Database to look up each SNP in the list. On the site, there is a flanking sequence, a series of base pairs before the SNP, and I searched for that flanking sequence in the whole gene sequence. I then made a comment for that base pair with the SNP number. In the end, I had a Microsoft Word file that outlined all the SNPs found with Haploview.

The last step was to use the NCBI PubMed database to look for associations between SNPs in CD36 and obesity or other related phenotypes, such as metabolic syndrome. SNPs that are associated with the phenotypes, tag other SNPs, have a Minor Allele Frequency (MAF) greater than 1%, or alter the protein are good ones to choose for candidate genes implicated in multifactorial diseases.




-NCBI Entrez Gene database information for CD36

## Results

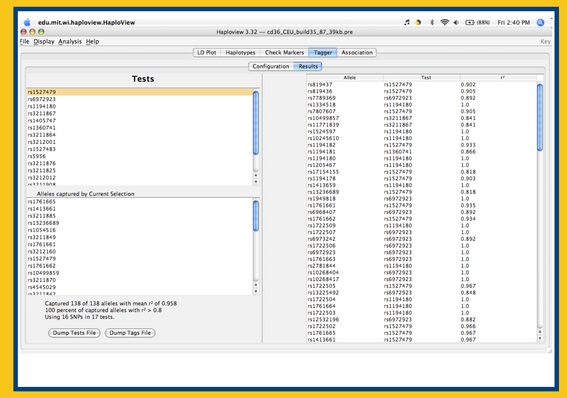
Two SNPs in CD36 were previously shown to be associated with obesity related phenotypes and were thus included in the candidate gene study.

rs1527479
-Minor Allele Frequency 49.2%
- -Tags 8 other SNPs in a large LD block
- -Associated with insulin resistance and Type 2 Diabetes

rs1527483
- -Minor Allele Frequency 15.8%
- -Tags 1 other SNP in a different LD block
- -Associated with high fatty acid levels and an increased cardiovascular risk

Pending available funding, 15 additional tagging SNPs will be used.

Now that the SNPs have been chosen for study, the genetic testing can be done to determine which SNPs are associated with obesity. Specially designed probes for each SNP are tested one at a time. The probes are used with a real time PCR machine. The SNP results will then be compared with phenotype information to see if an association exists.




-Haploview screen with the list of SNPs

## Acknowledgements

I would like to thank the University of Alaska Fairbanks and the Rural Alaska Honors Institute for providing me with this opportunity.

Dr. Sue Hills and Ian Herriott for all of their teaching and assistance with this project.

Aly Jeannet for her guidance and supervision.

Dr. George Happ for providing research funding.

This publication was made possible by Grant Number 5P20RR016466 from
the National Center for Research Resources (NCRR), a component of the
National Institutes of Health (NIH).

[3]Contact: bert.boyer@uaf.edu