

Selecting CD36 for a candidate gene study Steven Piazza¹, Renee Pasker², 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.

Search Gene	to Go Clear
Limite Proviou/Inde	
Display Full Report	Show 20 Send to
All: 1 Current Only: 1	Genes Genomes: 1 SNP GeneView: 1 🛠
□ 1: CD36 CD36 mol	lecule (thrombospondin receptor) [Homo sapiens]
GeneID: 948	updated
Official Symbol	CD36
Official Full Name	CD36 molecule (thrombospondin receptor)
Primary source	HGNC: 1663
See related	Ensembl:ENSG00000135218; HPRD:01430; MIM:173510
Gene type	protein coding
RefSeq status	Reviewed
Organism	<u>Homo sapiens</u>
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as	FAT; GP4; GP3B; GPIV; CHDS7; PASIV; SCARB3
Summary	The protein encoded by this gene is the fourth major glycoprotein of the platelet surface and serves as a r for thrombospondin in platelets and various cell lines. Since thrombospondins are widely distributed protei involved in a variety of adhesive processes, this protein may have important functions as a cell adhesion molecule. It binds to collagen, thrombospondin, anionic phospholipids and oxidized LDL. It directly mediate cytoadherence of Plasmodium falciparum parasitized erythrocytes and it binds long chain fatty acids and n function in the transport and/or as a regulator of fatty acid transport. Mutations in this gene cause platelet glycoprotein deficiency. Three alternatively spliced transcript variants encoding the same protein isoform been found for this gene.
-NCBI	Entrez Gene database information for CD3





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.

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.

• •	
<u>D</u> isplay <u>A</u> nalysis	<u>H</u> elp
	Tests
rs1527479	
rs6972923	
rs1194180	
rs3211867	
rs1405747	
rs1360741	
rs3211864	
rs3212001	
rs1527483	
rs5956	
rs3211876	
rs3211825	
rs3212012	
rc3211008	
Alleles captured	by Current Selection
rs1761665	
rs1413661	
rs3211885	
rs13236689	
rs1054516	
rs3211849	
rs1761661	
15212100	
rs1761662	
rs10499859	
rs3211870	
rs4545029	
rc7711847	
Captured 138 100 percent o Using 16 SNPs	of 138 alleles with mean r² of 0.958 f captured alleles with r² > 0.8 s in 17 tests.
(



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

assistance with this project.

5P20RR016466 from component of the National Institutes of Health (NIH).





Results

Haploviev	w 3.32 -	cd36_CEU_build	d35_87_39k	kb.pre				
								k
lot Haplo	otypes	Check Markers	Tagger	Association				
	Co	nfiguration Res	ults					
				Allele		Test	r ²	
			rs819437	Anere	rs1527479	0.902		6
			rs819436		rs1527479	0.905		
			rs7789369		rs6972923	0.892		
			rs1334518		rs1194180	1.0		
			rs7807607		rs1527479	0.905		
			rs10499857	7	rs3211867	0.841		- 11
			rs11771839	9	rs3211867	0.841		
			rs1524597		rs1194180	1.0		
			rs10245610)	rs1194180	1.0		
			rs1194182	-	rs1527479	0.933		
			rs1194181		rs1360741	0.866		4
			rs1194180		rs1194180	1.0		- 1
			rs1205467		rs1194180	1.0		- 1
	\sim		rs17154155	5	rs1527479	0.818		
			rs1194178		rs1527479	0.903		- 1
	Ψ.		rs1413659		rs1194180	1.0		- 1
			rs13236689)	rs1527479	0.818		- 1
	121		rs1949818		rs6972923	1.0		- 1
			rs1761661		rs1527479	0.935		- 1
			rs6968407		rs6972923	0.892		- 1
			rs1761662		rs1527479	0.934		- 1
			rs1722509		rs1194180	1.0		- 1
			rs1722507		rs6972923	1.0		
			rs6973242		rs6972923	0.892		
			rs1722506		rs6972923	1.0		
			rs6972923		rs6972923	1.0		
			rs1761663		rs6972923	1.0		- 1
			rs2781844		rs1194180	1.0		- 1
			rs10268404	ţ.	rs6972923	1.0		- 1
			rs10268417	7	rs6972923	1.0		- 1
			rs1722505		rs1527479	0.967		- 1
	Ψ.		rs13225492	2	rs6972923	0.848		- 1
			rs1722504		rs1194180	1.0		- 1
			rs1761664		rs1194180	1.0		- 1
			rs1722503		rs1194180	1.0		- 1
			rs12532196	5	rs6972923	0.882		- 1
			rs1722502		rs1527479	0.966		
			rs1761665		rs1527479	0.967		
			rs1413661		rs1527479	0.967		Ŧ

-Haploview screen with the list of SNPs

Acknowledgements

- Dr. Sue Hills and Ian Herriott for all of their teaching and
- Aly Jeannet for her guidance and supervision.
- Dr. George Happ for providing research funding.
- This publication was made possible by Grant Number
- the National Center for Research Resources (NCRR), a