

ARTICLE

A Single SNP in an Evolutionary Conserved Region within Intron 86 of the *HERC2* Gene Determines Human Blue-Brown Eye Color

Richard A. Sturm,^{1,3} David L. Duffy,^{2,3} Zhen Zhen Zhao,² Fabio P.N. Leite,² Mitchell S. Stark,² Nicholas K. Hayward,² Nicholas G. Martin,² and Grant W. Montgomery^{2,*}

We have previously demonstrated that haplotypes of three single nucleotide polymorphisms (SNPs) within the first intron of the *OCA2* gene are extremely strongly associated with variation in human eye color. In the present work, we describe additional fine association mapping of eye color SNPs in the intergenic region upstream of *OCA2* and within the neighboring *HERC2* (hect domain and RLD2) gene. We screened an additional 92 SNPs in 300–3000 European individuals and found that a single SNP in intron 86 of *HERC2*, rs12913832, predicted eye color significantly better (ordinal logistic regression $R^2 = 0.68$, association LOD = 444) than our previous best *OCA2* haplotype. Comparison of sequence alignments of multiple species showed that this SNP lies in the center of a short highly conserved sequence and that the blue-eye-associated allele (frequency 78%) breaks up this conserved sequence, part of which forms a consensus binding site for the helicase-like transcription factor (HLTF). We were also able to demonstrate the *OCA2* R419Q, rs1800407, coding SNP acts as a penetrance modifier of this new *HERC2* SNP for eye color, and somewhat independently, of melanoma risk. We conclude that the conserved region around rs12913832 represents a regulatory region controlling constitutive expression of *OCA2* and that the C allele at rs12913832 leads to decreased expression of *OCA2*, particularly within iris melanocytes, which we postulate to be the ultimate cause of blue eye color.

Introduction

Human eye color is a polymorphic phenotype under strong genetic control.¹ The gene responsible for oculocutaneous

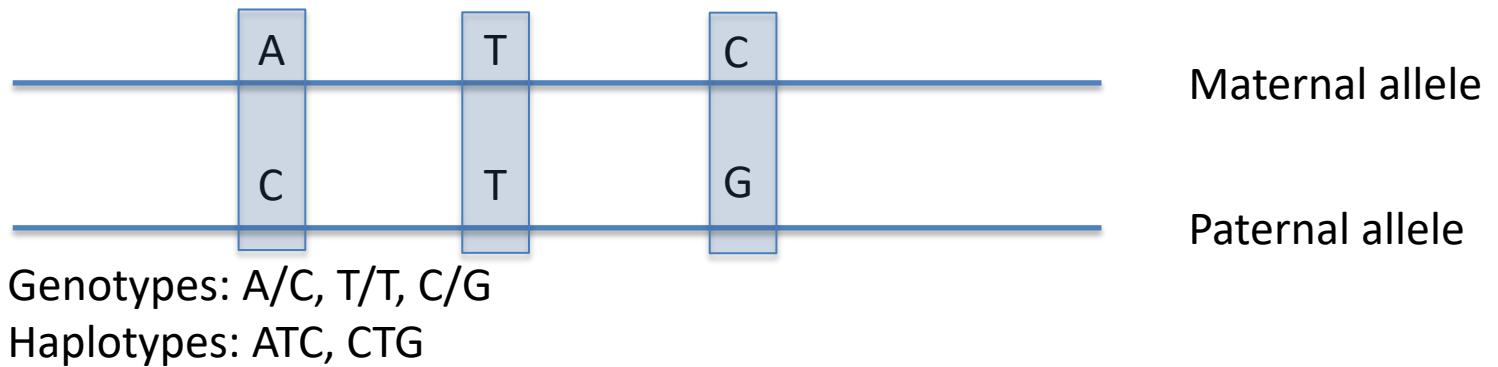
with three SNPs in intron 1: rs7495174 T/C, rs6497268 (now rs4778241) G/T, and rs11855019 (now rs4778138) T/C. We found the TGT/TGT diplotype in 62.2% of samples, and this was the major blue-eye genotype, with a frequency of 0.104 in blue-eyed individuals, com-

Terminology

- Linkage
 - “Genetic linkage is the tendency of alleles that are close together on a chromosome to be inherited together during the meiosis phase of sexual reproduction. Genes whose loci are nearer to each other are less likely to be separated onto different chromatids during chromosomal crossover, and are therefore said to be genetically linked. In other words, the nearer two genes are on a chromosome, the lower is the chance of a swap occurring between them, and the more likely they are to be inherited together.” (wikipedia)

Terminology

- Haplotype
 - A haplotype is, in the simplest terms, a specific group of genes or alleles that progeny inherited from one parent.
 - A tag SNP is a representative single nucleotide polymorphism (SNP) in a region of the genome with high linkage disequilibrium that represents a group of SNPs called a haplotype.
 - This illustration might be helpful to disentangle haplotype from genotype:



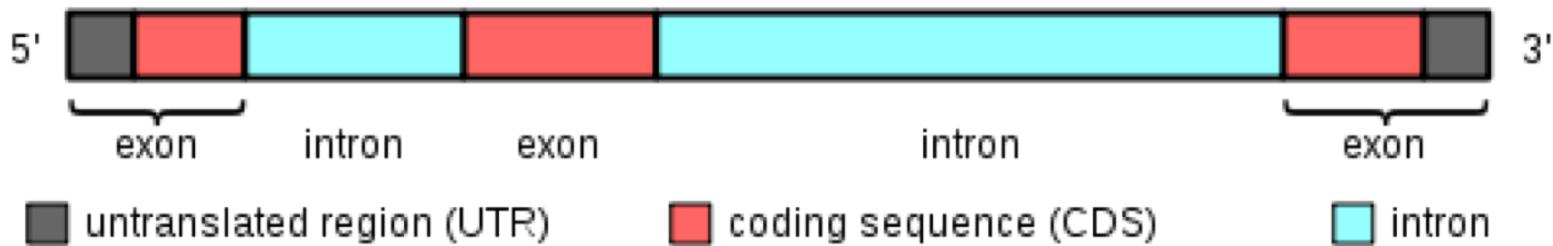
- Phasing
 - Linking two or more alleles into a haplotype is called phasing.
 - Above, the A at the first variant is phased with the C in the third into the haplotype ATC. They co-occur on the same chromosome copy!

Terminology

- LOD score
 - logarithm (base 10) of odds – a measure of linkage
- Penetrance
 - Penetrance in genetics is the proportion of individuals carrying a particular variant of a gene (allele or genotype) that also expresses an associated trait (phenotype).

Terminology

- Genes: Introns and exons



- Only exons get translated into amino acids.
 - The process of transcription in humans involves *splicing*, which skips over the introns (and sometimes exons) when generating RNA.

Abstract

We have previously demonstrated that haplotypes of three single nucleotide polymorphisms (SNPs) within the first intron of the *OCA2* gene are extremely strongly associated with variation in human eye color. In the present work, we describe additional fine association mapping of eye color SNPs in the intergenic region upstream of *OCA2* and within the neighboring *HERC2* (hect domain and RLD2) gene. We screened an additional 92 SNPs in 300–3000 European individuals and found that a single SNP in intron 86 of *HERC2*, rs12913832, predicted eye color significantly better (ordinal logistic regression $R^2 = 0.68$, association LOD = 444) than our previous best *OCA2* haplotype. Comparison of sequence alignments of multiple species showed that this SNP lies in the center of a short highly conserved sequence and that the blue-eye-associated allele (frequency 78%) breaks up this conserved sequence, part of which forms a consensus binding site for the helicase-like transcription factor (HLTF). We were also able to demonstrate the *OCA2* R419Q, rs1800407, coding SNP acts as a penetrance modifier of this new *HERC2* SNP for eye color, and somewhat independently, of melanoma risk. We conclude that the conserved region around rs12913832 represents a regulatory region controlling constitutive expression of *OCA2* and that the C allele at rs12913832 leads to decreased expression of *OCA2*, particularly within iris melanocytes, which we postulate to be the ultimate cause of blue eye color.

Introduction (1)

Human eye color is a polymorphic phenotype under strong genetic control.¹ The gene responsible for oculocutaneous albinism type II (*OCA2*)^{2, 3 and 4} has hitherto seemed the best candidate to explain the genetic linkage of blue-brown eye (BEY2/EYCL3 [MIM [227220](#)]) and brown hair (HCL3 [MIM [601800](#)]) color to chromosome 15q11.2–q12.^{5, 6, 7, 8, 9, 10 and 11} *OCA2* is the human homolog of the mouse pink-eyed dilution gene (*p*).¹² The human *OCA2* gene is divided into 24 exons covering >345 kbp of DNA; 23 of these exons span the 836 amino acid coding region, with exon 1 representing exclusively a noncoding 5'UTR.¹³ The resulting gene product, the P protein, is an integral membrane protein containing 12 transmembrane spanning regions that helps regulate melanogenesis.¹⁴

Introduction (2)

Many polymorphisms in *OCA2* occur in different populations with at least 13 nonsynonymous amino acid substitutions reported.^{2, 13, 15, 16, 17} and¹⁸ Only two of these were found to be present at significant frequency in our previous studies of adolescent twins and family members of mainly European descent from southeast Queensland, Arg305Trp (rs1800401) and Arg419Gln (rs1800407) at 0.05 and 0.09, respectively.¹⁸ They exhibited only a minor impact on eye color, leading us to conclude that *OCA2* coding alleles account for only a small proportion of the variation in iris pigmentation, at least within fair-skinned populations.

Introduction (3)

We proceeded therefore to screen tagging SNPs throughout the entire *OCA2* locus. The highest association for blue:nonblue (green/hazel or brown) eye color was found with three SNPs in intron 1: rs7495174 T/C, rs6497268 (now rs4778241) G/T, and rs11855019 (now rs4778138) T/C. We found the TGT/TGT diplotype in 62.2% of samples, and this was the major blue-eye genotype, with a frequency of 91% in blue- or green-eyed individuals, compared with only 9.5% in those with brown eyes.¹⁸ The position of this major diagnostic haplotype for eye color suggested that differences within the 5' proximal regulatory control region of *OCA2* alter temporospatial expression of the gene and may be responsible for these associations.

Introduction (4)

To further refine the elements controlling eye color, we have tested for association with haplotype-tagging SNPs proximal to intron 1 of *OCA2*, which span the intergenic region and encompass the 3' end of the upstream gene, *HERC2*. We describe the fine mapping of SNPs that are better predictors of blue-brown eye color than the existing *OCA2* intron 1 TGT haplotype block. Moreover, we propose a mechanism whereby a single base change (our best predictive SNP), contained within a highly conserved region of the *HERC2* gene intron 86, may abrogate accessibility of the *OCA2* chromosomal region by transcription factors necessary for expression of *OCA2* in people of different eye color.

Materials and Methods

- **Structure of the Study Population and Pigmentation Characteristics**
- ***OCA2* and *HERC2* SNP Genotyping**
- **Statistical Methods**

Results (1)

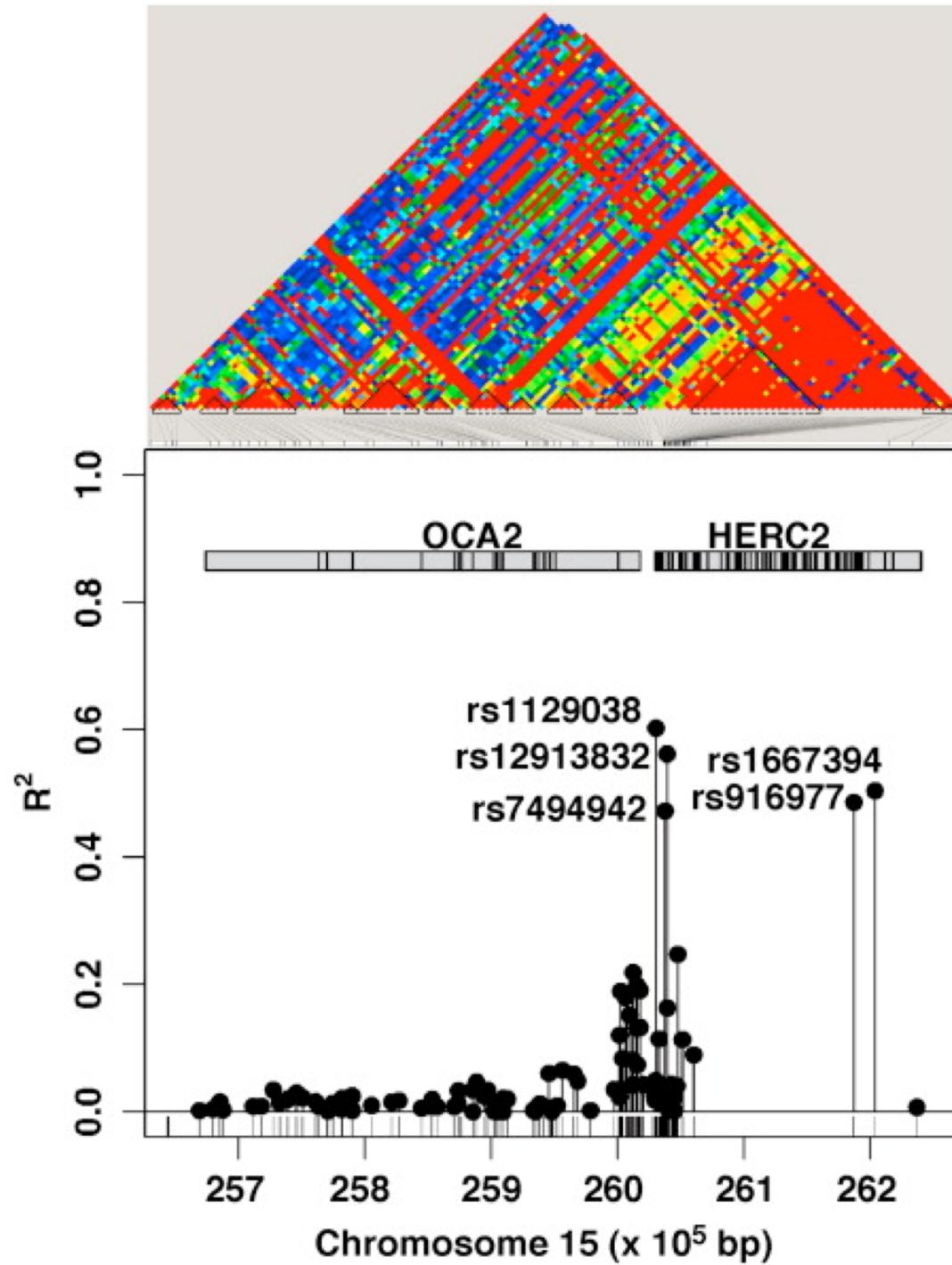
Association of *OCA2-HERC2* SNP Alleles and Haplotypes with Blue Eye Color in an Adolescent Twin Collection

The eye color grade distributions were similar to our earlier reports,¹⁸ and²⁷ with 46.1% blue/gray, 27.7% green/hazel, and 26.3% brown in the total sample collection. The genotyped subjects were approximately 52% female and 48% male and there were no significant gender differences in eye color distribution.

In our first round of follow-up genotyping of 8 SNPs 5' of the *OCA2* gene, the most strongly associated of the SNPs was rs12913832 ($p = 2 \times 10^{-78}$). Surprisingly, this SNP was 21.1 kb upstream of the *OCA2* first exon, in intron 86 of the *HERC2* gene (Figure 1). Individuals carrying the C/C genotype had only a 1% probability of having brown eyes. By contrast, T/T carriers had an 80% probability of being brown eyed.

Figure 1

Association of *OCA2* and *HERC2* SNPs with eye color (three-point scale) measured as R^2 in ordinal logistic regression analysis in the lower panel. The upper panel shows the linkage disequilibrium (r^2) between the SNPs.



Results (2)

Assay of *HERC2* SNP Alleles and Haplotype Association in Individuals Selected for Blue or Brown Eye Color

We then carried out genotyping at an additional 92 SNPs within the *OCA2* and *HERC2* genes in a subsample of 384 individuals from our collection (192 with blue eyes, 192 with brown eyes), choosing only one person from any family. Only one SNP, rs1129038, was found to slightly more accurately predict eye color than rs12913832 in this small subsample ([Figure 1](#)). This was located in the 3'UTR of the *HERC2* gene exon 93 and 12.4 kb upstream of the *OCA2* first exon. This had similar genotype frequencies to rs12913832 among blue-eyed and brown-eyed individuals. In addition, two more distant SNPs exhibited similar though slightly smaller effects on eye color: rs1667394 and rs916977, in introns 4 and 12 of *HERC2*, respectively. Further examination of these four eye color SNPs confirmed that they were all in strong linkage disequilibrium with one another (see [Table 2](#)). However, it was possible to show that rs1129038 and rs12913832 were more likely to be causative or in strong LD with the causative variant (see [Table 3](#)). Specifically, the rs1129038-rs12913832-rs916977-rs1667394 **ACGA/ACGA** genotype had a blue-eye phenotype whereas the **ACGA/GTGA** genotypes were brown-eyed ($p = 2 \times 10^{-16}$).

Table 3.

Phased Genotypes at Four Key HERC2 Locus SNPs versus Eye Color

	rs1129038	rs12913832	rs916977	rs1667394	Phased ^a	Blue	Brown
1 ^b	A/A	C/C	G/G	A/A	ACGA/ACGA	169	1
2, and	A/G	C/T	G/G	A/A	ACGA/GTGA	1	47
3 and	A/G	C/T	A/G	A/G	ACGA/GTAG	10	81
4 ^d	A/G	C/T	G/G	A/G	ACGA/GTGG	0	8
5	G/G	T/T	A/A	G/G	GTAG/GTAG	0	18
6	G/G	T/T	A/G	A/G	GTAG/GTGA	0	14
7	A/A	C/C	A/G	A/G	ACGA/ACAG	1	1
8	A/G	C/C	A/G	A/G	ACGA/GCAG	1	1
9	A/G	C/C	G/G	A/A	ACGA/GCGA	1	1
10	G/G	T/T	A/G	G/G	GTAG/GTGG	0	1
11	G/G	T/T	G/G	A/A	GTGA/GTGA	0	3

a AC phased rs1129038 and 12913832 in bold.

b The Fisher exact test comparing counts for row 1 versus 2, $p = 2 \times 10^{-16}$.

c Row 2 versus 3, $p = 0.097$.

d Row 2 versus 3 versus 4, $p = 0.195$.

Results (3)

- **Assay of *HERC2* SNP Alleles and Haplotype Association with Eye Color in an Adolescent Twin Collection**
- **Sequence Alignment of *HERC2* Exon 86 to Exon 93 across Species**

Discussion (1)

There is overwhelming evidence implicating the *OCA2* gene region in regulation of human pigmentation. Mutations in the *OCA2* gene lead to oculocutaneous albinism. Deletion of the region encompassing the *OCA2* gene on chromosome 15 as observed in Prader-Willi and Angelman syndromes is associated with hypopigmentation of the skin, hair, and eyes, and extra copies of this chromosomal region result in generalized hyperpigmentation of the skin. Normal variation in eye color shows strong genetic linkage ⁸ and association ¹⁸ to markers in the *OCA2* locus. In our previous combined segregation-linkage analysis in this twin sample, we estimated the frequencies of a dominant brown eye *B* allele as 21% and recessive *b* allele as 79%, which was close to the 26% *B* and 74% *b* allele modeled for the US white population by Hasstedt. ³⁶ The allele frequencies and penetrances estimated for rs12913832 in our Anglo-Celtic population match these predictions almost exactly. The European population frequencies in Hapmap for this SNP are in agreement (MAF 21%), and the variant is absent from other population groups. ³⁷ There is strong and extensive linkage disequilibrium across the *OCA2-HERC2* region, as well as marked differences in SNP allele frequencies between different ethnic groups. Voight et al. ³⁸ used HapMap data to show that this can be interpreted as a signal of positive genetic selection in Europeans, similar to that seen around other pigmentation loci such as *TYRP1*, *DTNBP1*, and *SLC24A5*. The hitchhiking seen around such loci can decrease the power of association-based fine mapping.

Discussion (2)

Although rs12913832 lies within a distal intron of *HERC2* (intron 86 of a 93 exon encoded gene spanning 211 kb), we do not believe that the *HERC2* protein is involved in regulating the pigmentation pathway. Three mouse deletions that map within this distal region of *HERC2*, which do not include the *OCA2* locus, were originally labeled as *p* gene mutants because they exhibit pigmentation changes that are intermediate to those seen in canonical *p* mutant homozygotes: dark pink eyes with mottling or partially decreased eumelanin quantities in hair.^{39, 40} and ⁴¹ Effects on pigmentation are thought to result from sequences within this region of *HERC2* controlling expression of *OCA2*. One of the radiation-induced mutant mouse alleles characterized by Lehman et al.,⁴⁰ *p*^{bs} (black eyed, sterile; originally *p*^{24H} ⁴²), involves a deletion of only 8 kb in *Herc2*, which spans the orthologous region to human *HERC2* exons 86–93. Lehman et al. comment in parentheses that “[t]he relatively mild hypopigmentation associated with the *p*^{bs} deletion apparently reflects an effect of the deletion on expression of the closely linked *p* gene.” Walkowicz et al.⁴¹ describe characterization of the *p*^{12DTR}, *p*^{103G}, and *p*^{39DSD} alleles.⁴³ These all involve deletions or rearrangements of *Herc2*. Those authors concluded that elements regulating the expression of *p* probably lie within *Herc2*.

Discussion (3)

Recently, strong association between eye color and rs1667394 in *HERC2* was reported in Icelandic population,⁴⁴ but we excluded this as the causal SNP based on haplotype analysis. One of our two most strongly associated SNPs, rs12913832, lies within an evolutionarily conserved 406 bp region. Moreover, database searches of this region revealed transcription factor binding sites for LEF1, HLTF, and MITF. Both LEF1 and MITF are critically important to gene regulation in melanocyte cell development, differentiation, and tissue-specific transcription.⁴⁵ and⁴⁶ Helicase-Like Transcription Factor (HLTF) is a member of the SWI2/SNF2 family, DNA-dependent chromatin-remodelling ATPases that have been implicated in a wide variety of processes involving the modification of chromatin configuration to allow access of the transcriptional machinery. HLTF is the only member of the family in humans that incorporates a specific DNA recognition site that binds a variety of gene promoters, including PAI-1 and beta-globins (where it seems to be involved in controlling levels of constitutive expression),⁴⁷ and⁴⁸ uteroglobin (transducing of the effects of prolactin) on uteroglobin,⁴⁹ and an enhancer region of the myosin light chain locus.⁵⁰ and⁵¹ The binding site in intron 86 matches the consensus recognized by the HLTF variant known as RUSH-1- α as determined by CASTing.⁴⁹

Discussion (4)

Another SWI/SNF family member, BRG1, has recently been shown to be involved in pigmentation pathways, being recruited by MITF to melanocyte-specific promoter regions to induce changes in chromatin structure at endogenous loci, and thus initiating the process of melanocyte differentiation.⁵² We therefore suggest that the presence of both MITF and HLTF enhancer sequence elements in the evolutionary conserved region of *HERC2* intron 86 is a further pointer that this is a locus control region that determines the expression of the *OCA2* gene product.

Discussion (5)

We have now genotyped most of the reported SNPs in the 3' end of *HERC2* and the 5' end of *OCA2*, finding a large number (39 out of 93 or 42% in the second round of genotyping) to be monomorphic in our sample. The blue-eye-color-associated allele must be reasonably common if a single SNP is to explain most of the variation in eye color in European-descended populations. This makes it less likely that another SNP as yet unidentified in the region is the true causative variant. Based on the foregoing, we conclude that the conserved region around rs12913832 represents a regulatory region controlling constitutive expression of *OCA2*, and that the C allele at rs12913832 leads to decreased expression of *OCA2*, particularly within iris melanocytes. We speculate that the regulatory mechanism is abrogation of the binding site for HLTF that regulates transcription of the neighboring *OCA2* gene. We also confirmed that the common coding variant *OCA2**R419Q acts to modify the penetrance of this locus and that this effect includes modification of the risk of malignant melanoma.

dbSNP Entry

Reference SNP (refSNP) Cluster Report: rs12913832

**** With other allele ****

RefSNP		Allele
Organism:	human (<i>Homo sapiens</i>)	Variation Class: SNV: single nucleotide variation
Molecule Type:	Genomic	RefSNP Alleles: A/G (FWD)
Created/Updated in build:	121/151	Allele Origin: A:germline G:germline
Map to Genome Build:	108/Weight 1	Ancestral Allele: A
Validation Status:		Variation Viewer: VarView
Citation:	PubMed LitVar <small>NEW</small>	Clinical Significance: With other allele [ClinVar]
Association:	NHGRI GWAS PheGenI	MAF/MinorAlleleCount: G=0.1773/888 (1000 Genomes) G=0.4533/56919 (TOPMED)